TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SPR-1741	N/A	N/A

4. Title and Subtitle **Leveraging Crowd-sourced Data in Planning, Design, Analysis, and Evaluation of Pedestrian and Bicycle Traffic**	**5. Report Date** 08/30/2024 **6. Performing Organization Code** N/A
7. Author(s) Aditi Misra, Ph.D. https://orcid.org/0000-0002-5600-5973, Krista Nordback, Ph.D., Shubhayan Ukil, Mike Vann, Garrett Fardon, Meghna Chakraborty, Ph.D., Wesley Marshall, Ph.D.	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address University of Colorado – Denver Office of Grants and Contracts Research Administration Mail Stop F428 13001 East 17th Place, Room W1124 Aurora, CO 80045	**10. Work Unit No.** N/A **11. Contract or Grant No.** OR#022-006
12. Sponsoring Agency Name and Address Michigan Department of Transportation (MDOT) Research Administration 8885 Ricks Road P.O. Box 33049 Lansing, Michigan 48909	**13. Type of Report and Period Covered** Final Report, 10/1/2021 – 8/31/2024 **14. Sponsoring Agency Code** N/A
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration. MDOT research reports are available at www.michigan.gov/mdotresearch. Enter information not included elsewhere, such as translation of (or by), report supersedes, old edition number, alternate title (e.g. project name), hypertext links to documents or related information in the form of URLs, PURLs (preferred over URLs - https://purl.org/docs/index.html), DOIs (http://www.doi.org), insertion of QR codes, copyright or disclaimer statements, etc. Edit boilerplate FHWA statement above if needed.	

16. Abstract

Data on active transportation are difficult to collect for DOTs, MPOs and local agencies because of the short duration of trips, non-traditional time and routes used by the bicyclists and pedestrians as well as lack of extensive resources needed to track, count and map such movements. Crowdsourced data can be a great resource in this case to fill in the gap. However, crowdsourced data comes with its own data quality and coverage challenges which makes it difficult to use it as available. Through this project, the project team developed an analytical data-driven framework, resulting in reliable, reproducible and transferable statistical models that can be used to predict bicycling and pedestrian volumes in road networks using crowdsourced data. The project team first surveyed multiple state DOTs through an online survey to identify relevant crowdsourced data sources and their advantages and disadvantages. Then the project team acquired and processed data from 43 permanent Using volume data from 43 counters in Michigan and crowdsourced data from Strava along with several other land use, network and weather-related variables from multiple open access resources, the project team developed two types of models separately for bicycling volume, pedestrian volume, and bicycling and pedestrian volume for Michigan. The first set of models, inspired by the gravity models from traditional travel demand forecasting models, developed the 'Active Expansion Factor' models with prediction accuracy for bicycling volume ranging between -20% to +50%, while that for pedestrian volume ranging between 100 to 200%. The second set of models used mixed effect zero-inflated count models to identify more nuanced causal relationships between volumes and land-use, weather and infrastructure-related variables. Findings indicate that Strava data are significantly correlated with active transportation volume data, especially bicycling volume data, but better model fits and prediction can be achieved when Strava data is augmented by land use and in case of bicycling, weather-related variables.

17. Key Words	**18. Distribution Statement**
Crowdsourced data, Active transportation volume, Strava, Count data, Expansion factor, Mixed-effect models, Variable selection models, Decision trees.	No restrictions. This document is also available to the public through the Michigan Department of Transportation.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	71	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

List of Tables

List of Figures

Executive Summary

Data on active transportation are difficult to collect for DOTs, MPOs and local agencies because of the short duration of trips, non-traditional time and routes used by the bicyclists and pedestrians as well as lack of extensive resources needed to track, count and map such movements. Traditional methods of data collection on active transportation include manual and automatic count stations where volume of bicyclists passing through that station is counted by volunteers (in case of manual counts) or by a mechanical counter (in case of automatic counts). However, the extensive nature of Michigan road network makes it prohibitively resource intensive to collect data on bicyclists and pedestrians on every road on the network using count stations, spatially and temporally. In addition, the traditional count method, whether manual or automatic, stops short of providing information on travel patterns and characteristics that could help planners and engineers understand the role of infrastructure or other related factors influencing bicycling or walking. Crowdsourced data has been promoted as a great resource in such cases to fill in the gap between what data are collected through count stations and manual counts and what needs to be known for planning and road safety analysis purposes. However, crowdsourced data comes with its own issues of data quality and representativeness and may not be used directly as an alternative to traditional data.

The aim of this project was ***to develop a framework for identifying and utilizing the best resource available in crowdsourced data for active transportation through a generalizable model mapping crowdsourced data to count data and related contextual features that influence bicycling and walking.*** Towards that, the objectives of this study was to (i) review crowdsourced bicycle and pedestrian data resources and crowdsourcing tools; (ii) discuss potential planning implementations of crowdsourced data for a variety of bicycle and pedestrian project types; (iii) provide examples of how crowdsourcing is currently being used by the planning community; (iv) acquire/collect and process crowdsourced and count data at select locations; (v) develop QAQC procedure for data types and identify factors contributing to differences in count data and crowdsourced data at those select locations; and (vi) finally develop a generalizable model and adjustment factors mapping crowdsourced data with count data and other contextual features that will enable Michigan Department of Transportation (MDOT) to use crowdsourced data for different purposes.

The project team searched multiple databases and found that Transportation Research Thesaurus (TRT) was the most comprehensive source. Multiple keywords were used for the search*: pedestrian counts crowdsource, pedestrian & crowdsource, pedestrian crowdsource, bicyclist crowdsource, bicycling crowdsource, bicycling count* and finally, *crowdsource, crowd-source and crowdsourcing*. At the beginning of the project, the project team reviewed over 2000 records matching the keywords and deemed 52 of them to be related and informational for this project. Towards the end of the project, a supplementary literature search was done to update the previous findings. Through this search 7 additional studies were included that have been completed and published between 2022 and May 2024. The project team also completed a survey of multiple state department of transportation (DOT) and metropolitan planning organization (MPO) officials in the US. Findings from the survey indicated that most DOTs and MPOs faced the same problem of lack of data for bicycling and pedestrian planning. Some of the DOTs had used crowdsourced data mainly from two third party data vendors – Streetlight and Strava. People who had used crowdsourced data consistently mentioned the data quality, coverage and representativeness. Some smaller agencies indicated prohibitive cost of acquiring data from third party vendors as reason for not using crowdsourced data. Across the board, however, everyone reported a lack of crowdsourced data for pedestrian volume.

Since Streetlight and Strava were the most used crowdsourced data sources, the project team then developed collaboration with these two vendors and obtained data from them for the permanent and temporary counter locations in Michigan. There were three types of counter data – bicycling only, pedestrian only and bicycling and pedestrian only. Both the counter data and the crowdsourced data were processed and aggregated to make them comparable. Basic SWOT analysis showed Streetlight data to be closer to ground truth (counter) data than any other sources and that crowdsourced data could only be used for comparison after a temporal aggregation e.g., monthly average volume instead of hourly volume because of the sparse nature of the data. In 2022 Streetlight decided to stop providing bicycling and pedestrian data at road segment level and moved to providing data only at an aggregated census block group level, which wasn't aligned to the goals of the project. At the same time, while originally being an application for tracking bicycling, Strava added functionality to track running/walking exercises which could be somewhat a proxy for pedestrians. In addition, Strava showed a high user volume, a track record of being used by multiple cities for understanding their bicycling and walking patterns and offered the data free to planning and government organizations like DOTs. So, for the final comparisons and models, Strava was chosen to be the crowdsourced data source.

A set of generalized models were then developed using Strava data that could be used for the entire state of Michigan. The first set of models were based on adjustment factor method, borrowed from the gravity model of traditional 4-step travel demand models, to understand if there was a standard calibration factor between crowdsourced data and counter data. While the adjustment factors were reasonably stable temporally, because of the low number of counters (~40) used to derive these factors and the lack of variability in the type of facilities covered by these counters, using the adjustment factors single-handedly to map Strava data to counter data was deemed unreliable. Contextual variables were chosen based on literature to explain the variation of the adjustment factors at different locations and modeled accordingly. These models were less resource intensive but provided prediction accuracy of bicycling volumes with ranging between -19% to 130% while pedestrian volumes could be predicted with accuracy ranging between 100% to 160% of the counter data. Next, two sets of models, one using Strava data and another without Strava data, were developed for each mode type: bicyclist only counters, pedestrian only counters and bicyclist and pedestrian combined counters (a total of six models). To account for some of the many factors that may impact pedestrian and bicycle traffic and the fit of the Strava data to the ground truth count data, about 166 additional data elements were considered based on existing literature. Because of the panel structure of the data both spatially and temporally, mixed effect models were used for these six models. Results indicate that Strava data was strongly correlated with counter data for bicycling volume and aggregated bicycling and pedestrian volume but had low correlation with pedestrian volume. Land use and population variables were useful in improving the prediction of non-motorized traffic volume in conjunction with Strava data.

In conclusion, while this research, like similar other contemporary research on usefulness of crowdsourced data, could not recommend use of crowdsourced data as a replacement of counter data, it developed tools that could be implemented in spreadsheets and used to predict bicycling and pedestrian counts with a reasonable accuracy. It is expected that over time more counters will be installed covering different types of roadways and facilities across the state and thus, more data will be available to develop more standardized and reliable adjustment factors for different types of locations. Future research in this area should focus on collecting and using data from a larger and more diverse set of locations to test the sensitivity of the models and adjustment factors. Finally, given the increasing concerns about privacy, it is unlikely that quality of crowdsourced data will get better or have more coverage. Alternative data sources like video data or infrastructure-based sensor data may become more viable options to address data needs in the future. Future research should also consider these data sources for their cost effectiveness and accuracy.

Phase 1: Oct 2021 – Jul 2022

Chapter 1. Literature Review

1.0 Introduction

Counts provide the foundation for measuring nonmotorized travel along a link or a network and are also useful for monitoring trends, planning new infrastructure, and for conducting safety, health, and economic analyses. For safety analysis, they are critical in assessing the exposure to risk. Over the last decade, several automated technologies have been developed to count bicyclists and pedestrians. Despite advances in counting technology, cost and other considerations will continue to limit direct observation to small subsets of entire networks. The emergence of crowdsourced data such as Strava and Streetlight has allowed for the collection of large-scale datasets over broad areas of the network. However, crowdsourced data comes with its own issues of data quality and representativeness. Particularly for multimodal travel data that are collected passively from cellphone/GPS tracking, determining trips as bicycling is extremely hard – bicyclists can go as fast as a slow bus and as slow as a fast pedestrian (Nelson et al. 2021). Attempts at fusing emerging data sources with count data have also indicate that even when data are bought from well- known commercial third part aggregators of cellphone GPS data, the quality of data on bicycling and walking are questionable and needs further processing at the very least (Broach et al. 2023). On the other hand, cycling/walking (running) focused smartphone applications and trackers provide the certainty that we are considering only relevant trips, but may or may not be representative of all types of cyclists and pedestrians.

With evolving technologies, new ways are being used to collect information and communicate in the field of active transportation planning to make travel easier and safety for active transportation users. These approaches, often aggregated under an umbrella term of 'Crowdsourcing', can take many different forms and serve a variety of transportation planning needs. Crowdsourcing can be defined as a strategic model to attract an interested, motivated crowd of individuals capable of providing solutions superior in quality and quantity to those that even traditional forms of business can. While its interpretation is flexible and varies by field, crowdsourcing broadly involves incorporating the value of knowledge compiled from a large group of diverse perspectives compared to more traditions means of data collection (Smith, 2015b). Crowdsourced data is a more nuanced form of crowdsourcing in which participation of crowd depends on the method used to crowdsource data. With advancements in the big data aggregation, data generated through implicit crowdsourcing techniques have become available that repurpose large user-generated datasets collected for other intents. Crowdsourcing methods not only provide access to high-quality data and at a finer spatial resolution, as well as directly engaging with the community members, but also offers options to increased public participation on bike and pedestrian planning. This helps to understand the relationships of bicyclists and pedestrians with the built environment, their travel decisions, and their needs in a bottom-up planning process, in contrast to the top-down approach where the needs of non-motorized travelers may be unmet (Smith, 2015a).

Federal Highway Administration as part of its program to advance innovation in the transportation community explored the use of crowdsourced data to improve traffic operations. Under it they explored the use of transportation systems users as real-time sensors to obtain low-cost, high-quality data on traffic operations, conditions, and patterns. These datasets can be used to optimize the use of roadway facilities through traveler information, incident management, road weather management, arterial management, and other strategies targeting the causes of congestion (Every Day Counts: Innovation for a Nation on the Move, 2021). State departments of transportation (DOTs) and other transportation organizations are also evaluating the use of commercially available counting technologies to complement their traditional methods of collecting non-motorized traffic counts for nonmotorized travel monitoring. In Vermont, nonmotorized traffic counts are collected by the University of Vermont Transportation Research Center, VTrans, and several of the state's regional planning commissions (RPC). These datasets are complemented with data from Strava (Karen Sentoff & James Sullivan, 2017). Local institutions are also interested to improve their walking and bicycle infrastructure and need pedestrian and bicyclists counts but may not have the capacity to do so. While investigating the feasibility of a pedestrian and bicycle count program in Virginia, Ohlms et al. (2018) found that some local communities are interested in pedestrian and bicycle volume counts and want to partner with partnering with Virginia DOT for counts. State organizations at different levels need to provide assistance to local organizations to establish a pilot nonmotorized count program as there are many considerations to start a count program. Ohlms et al.(2019) reviewed the bicycle and pedestrian count program in the United States and found that there are several considerations beyond just purchase and installation of automatic count equipment or buying a crowdsourced big data. Some of the key components of a program enabling data to be useful are maintenance costs; data validation, formatting, quality, and storage; and analysis and/or modeling, training and outreach.

The aim of this project has been to create a generalizable framework that will enable us to use crowdsourced data efficiently, filling the gap in traditional count methods, while taking into account the drawbacks of crowdsourced data. To create a benchmark from which the project could take-off, our first task was to review the existing body of literature and projects that have either (i) assessed the quality of crowdsourced data for bicycling and/or pedestrian volume or safety, or (ii) used crowdsourced data to model bicycling and pedestrian volumes using novel, state-of-the-art methodologies. In doing so, we also uncovered some literature on using sensing technology to count bicyclists and pedestrians and some literature on using passive technology/crowdsourcing to collect data on bicycling and pedestrian related infrastructure, like sidewalks. As part of the task, we searched the following databases: Transportation Research International Database (TRID), ScienceDirect, Web of Science, and Google Scholar. We also scanned appropriate websites from different organizations, such as FHWA, National Highway Safety Traffic Administration (NHTSA), as well as the American Association of State Highway and Transportation Officials (AASHTO), the Institute of Transportation Engineers (ITE), the National Association of City Transportation Officials (NACTO), the Transportation Research Board (TRB), and the National Cooperative Highway Research Program (NCHRP). However, most comprehensive coverage of the literature on use of crowdsourced data for topics related to bicyclists and pedestrians was found using Transportation Research Thesaurus or TRT, an indexing tool developed as part of NCHRP project 20-32 (https://trt.trb.org/about-the-trt). Although the initial idea for the research thesaurus was to develop a common terminology for various transportation related applications, it morphed into a comprehensive indexing system of federal, state and university generated contents. TRT is regularly updated, and its content is vetted by a committee to make sure it is relevant and useful for the transportation community.

The abovementioned databases and websites were used to set up the initial list of documents and then used TRT to retrieve summary and indexing information for future use. The TRT database was then searched with multiple keywords*: pedestrian counts crowdsource, pedestrian & crowdsource, pedestrian crowdsource, bicyclist crowdsource, bicycling crowdsource, bicycling count* and finally, *crowdsource, crowd-source and crowdsourcing*. This review was conducted between November 19, 2021, and December 3, 2021, and then again between March 27, 2024, and May 7, 2024. Over 2000 records matching the keywords were reviewed and 52 of them deemed to be related and informational for this project in the first phase. In the second phase 7 additional studies were included that were completed and published between 2022 and May 2024. Most of the records presented in this report were completed between 2018 and 2024, but a few older records were included as well, which were deemed foundational and important.

As shown in the bibliography presented as Appendix A, multiple projects and studies had tried using crowdsourced data for volume estimation, as well as to explore safety concerns and conflict situations for bicyclists and pedestrians (more so for bicyclists than pedestrians). Crowdsourced data collection platforms range from established third-party vendors like Strava and Streetlight (Broach et. al 2023) to smartphone applications created by research/project groups for the particular purpose of the project (e.g., Minnich 2023 created an app to gamify participation in active transportation modes). Broach et. al. (2023) has recently completed a study that used crowdsourced data from different sources together to estimate bicyclist volume. However, one of the crowdsourced data sources, Streetlight, has stopped sharing segment level data starting 2022 on account of data quality and coverage, underlining the inherent uncertainty of availability, quality and reliability of crowdsourced data. The bibliography also includes some literature on data collected using passive sensing technology like LiDAR (Lesani et. al. 2021) as well as Bluetooth and Wifi (Lesani and Miranda-Moreno, 2019) - these technologies can supplement the data quality issues arising from smartphone-based data but are expensive to install and operate and requires extremely high computational capabilities. Some research projects have used crowdsourced data to understand safety scenarios (Kwayu et. al. 2022, Rahman 2019, Carlson et. al. 2018), while some have used Twitter to understand perception (Rahman et. al. 2021, Chandra et. al. 2019). Finally, we also present a study using sensing and crowdsourced data for pedestrian infrastructure accessibility – this is an emerging research area with potential to improve overall performance of bicyclist and pedestrian volume estimation along segments and in route planning. Below, a comprehensive summary of the literature reviewed is presented, organized by the thematic areas of types of crowdsourced data, methods of collecting crowdsourced data, uses of crowdsourced data, methods of collecting and analyzing crowdsourced big data, issues with crowdsourced data and finally crowdsourced data ethics.

2.0 Literature Review

1. Types of Crowdsourced Data

Crowdsourced data can be of different types. 1) In-situ data are data sources that include mobile applications tracking real-time, geotagged data representing the travel patterns of individuals (e.g. Strava, Map by Fitness, Moves, etc.). 2) Thematic Data sources include data aggregated, categorized, and/or summarized within pre-defined geographic area (e.g. Decennial Census, American Community Survey, National Household Travel Survey). These data sources can be text-based with active transportation characteristics linked to a particular place, such as city, county, Census tract, or transportation analysis zone. 3) Thumbtuck data sources include point locations added to a map with associated attribute information (e.g. map developed by Divvy, Chicago’s bike sharing system). 4) Spatial Inventory data includes digitized representations of ground features, often with associated attributes. Crowdsourced data can also be collected directly using smartphone mobile apps or through other forms of self-reported data.

Specifically for bicycle and pedestrian travel demand modeling and transportation planning and safety studies, there are several commercially available in-situ data sources, such as Streetlight, Strava, INRIX, etc. These datasets vary in terms of the format they are available, the magnitude of resolution at which they are available and can be put to different uses. Streetlight data mains provide origin-destination (OD) related metrics. It can be used to quickly estimate origin-destination trip tables. However, Streetlight metrics heavily rely on the data points sampled from smartphone applications and global positioning services (GPS) devices, which may be subject to potential bias and coverage issues. Yang et al. (2020)developed a set of guidelines for the Virginia Department of Transportation (VDOT) to understand the performance of the SL metrics in different application contexts and use Streetlight data, in general. They also tested the accuracy of Streetlight metrics comparing it to ground-truth data from different sources such as continuous count stations, toll transaction data, VDOT's internal traffic estimations, etc. The evaluation results were mixed. The latest AADT estimates showed relatively small absolute percentage errors, whereas using the SL metrics to estimate OD trips, traffic counts on roadway segments and at intersection. However, large percentage errors were often found to be associated with lower volume levels estimated based on the SL metrics. In addition, using the SL metrics from individual periods as the input for estimating these traffic measures resulted in larger errors. Instead, the aggregation of data from multi-periods helped reduce the errors, especially for low volume conditions. Another data source is Strava. Its mobile app and its desktop website interface allow athletes to track, analyze, plan, and share their training rides and runs. The Strava data is anonymized and aggregated. The final data to use for transportation planning is available at a street level, at a geographic level, at origin and destination level, and provide travel time between locations. Miovision provides real-time data and analytics to optimize traffic flow, improve traffic safety. They provide speed data, onboard Average Traffic Rates (ATRs), etc. One of their tools is Miovision's TrafficLink Multimodal Detection for traffic detection at intersections. Minh (2019) used the tool in a study to count pedestrians from forty hours of video. The data represents one week of selected hourly weekday and weekend pedestrian counts at two intersections in Austin, Texas. Manual counts were compared to Miovision's count data across different combinations of lighting conditions and pedestrian volumes. The results showed that Miovision system performed well with accuracy results of 15% error for daytime and 24% for nighttime for the combined intersection legs.

Datasets from different sources also give different accuracy compared to counter data. In an exploratory study to analyze the use of different types of crowdsourced data for roadway safety analysis, Turner et al. (2020) compared four different sources of data to counts from counter. Firstly, they compared passively collected crowdsourced bicyclist activity data from Streetlight with bicyclist counts from 32 locations in eight Texas cities. The results found good correlations of 62% and 69% for monthly weekday and weekend daily averages. The correlations improved to 94% when compared with countywide Strava data. Next, they evaluated the pedestrian counting accuracy of the Miovision system and found 15% error for daytime and 24% error for nighttime conditions. They also analyzed INRIX trip trace data to determine origin-destination patterns and developed 40 decision rules to define the origin-destination patterns. Finally, they analyzed crowdsourced Waze data (i.e., traffic incidents) and found it to be a useful alternative to observed and predicted crashes, with the ability to identify high-risk locations. 77% of high-risk locations identified from police-reported crashes were also identified as high-risk in Waze data. The degree of accuracy with these datasets and the diverse ways in which they can be used shows the utility of these datasets.

2. *Methods to Collect Crowdsourced Data*

Although digitally produced big data sources are being highly used in transportation planning with methodological advancements to leverage them, crowdsourced data can be collected more actively with different levels of public engagement and each tailored for separate processes. Below are three methods that can be used to crowdsource data:

a. In-person meetings is a form of crowdsourced data collection that has been used for community engagement for a long time. They help connect public directly to the planners and get more detailed feedback through in-person engagement. However, conducting them is also labor intensive. In Austin, for the development of the 2014 Austin Bicycle Master Plan Update, in-person meetings were conducted and were one of the several methods to engage public in the planning process. Other methods were a telephone survey, an urban trail intercept survey, an online survey, a virtual open house, and discussion at multiple City of Austin boards and commissions meetings (Austin Transportation Department 2014). Though less in number, the meetings connected interested persons with city staff directly—a rich engagement approach not afforded by online methods (Greg P. Griffin & Junfeng Jiao, 2018).

b. Public Participation Geographic Information System (PPGIS) is another way to crowdsource data using online tools. The Capital Area Metropolitan Planning Organization (CAMPO)—the regional transportation planning agency in Austin—used a public participation geographic information system (PPGIS) called "WikiMaps" in the development of the 2045 Regional Active Transportation Plan (Capital Area Metropolitan Planning Organization 2017). Public contribution of knowledge to the planning process was done using PPGIS that allowed people to click a location on a computer-based map and identify barriers and other issues for bicycling and walking (Greg P. Griffin & Junfeng Jiao, 2018). Using online tool for public participation to collect crowdsourced data increased the participation rate. Using online tools helped to get feedback from a wider geographic area, which helped the organization since they focus on region-wide development, a seven-county area surrounding Austin.

c. Smartphone applications are being widely used to collect data. They provide flexibility to collect travel data as people travel without the crowdsourcer having to spend a lot of time to provide feedback unlike using an online tool or participating in a meeting. Austin Transportation Department used Ride Report, a smartphone application, that records contributors' bicycle trips, detected automatically using the phone's accelerometer and GPS (City of Austin 2018; Ride Report 2016). It also detects the conclusion of a bicycle trip, and prompts users to rate a ride as positive or negative. The platform aggregates multiple overlaid trips by all participants to compute an average rating, in addition to recording the total count of users for each roadway and trail segment (Griffin & Jiao, 2018). Smartphone Apps can also be used to collect user perception data. Hopkin et al. (2014) developed a smartphone application as part of a pilot study to collect data on how app users perceived their journey quality and value of time at different stages of their journey by different modes. They found that the feedback from users can be incorporated by transport operators into real time information services to enhance the information available to passengers. Also, users could share their journey experiences directly with other travelers, so that other travelers could benefit from their good or bad experiences. Using such apps also increases the chances of having noise in data when mobility data is collected in the form of GPS data and may require adjustments. Also, data collection using such apps happens over a period and may see drop out from participants or lack of commitment, which may affect the quality of the generated crowdsourced data.

3. Uses of Crowdsourced Approaches and Crowdsourced Data

a. Forecasting Pedestrian and Bicycle Count

In pedestrian and bicycle research, crowdsourced data can be used in multiple ways: Crowdsourced data sources are useful for State DOTs to complement their data needs for bike and pedestrian demand forecasting. Even with advancements in counting technology, direct observations of network is limited to small subset of locations, which restrict the available information about the activity happening on the network around the counter locations. This lack of pedestrian count data availability limits studying any kind pedestrian safety hazard, which has increased rampantly over the past decade. Crowdsourced data sources such as Strava, Streetlight, etc. fills in the data gap with large-scale datasets available across broad areas of network. This can be used for microscopic pedestrian count forecasting, which has direct application in improving safety for pedestrians. Ongoing studies, such as study titled 'Exploring the Use of Crowdsourced Data Sources for Pedestrian Count Estimations' by National Institute for Transportation and Communities are exploring the use of crowdsource data sources for count estimations. More details can be found in the bibliography section. Crowdsourced data can be used to estimate sectional volume of travelers, i.e. the number of travelers crossing a section boundary, within a certain time period. In a study to estimate sectional volume of travelers using mobile phone data, Liu et al. (2020) used a three-stage framework. In the first two stages, the spatial and temporal uncertainties of trajectories were explicitly addressed by a hybrid filtering algorithm and a cell-to-cell trajectory inference algorithm, respectively. Finally, the sectional volume of travelers was estimated using aggregated trajectories. The proposed framework was validated using a sampled dataset with annotated ground truth and a city-scale dataset. The results suggest that the proposed framework is effective to deal with spatial and temporal uncertainties of trajectories and gives sectional volumes with a low average error rate.

Even as the utility of crowdsourced data for forecasting non-motorized traffic demand for monitoring and safety of bicyclists and pedestrians is being explored, learnings from still widely used counter data can be applied to the case of crowdsourced data. Nordback et al. (2019) provided new guidance for monitoring and volume estimation of nonmotorized traffic using continuous count data from 102 sites across six cities. They found that mean absolute percent error (MAPE) in estimated annual average daily nonmotorized traffic (AADNT) is minimized when seven-day short duration counts are based on data collected between June and September for 24-h counts, when data are collected Tuesdays through Thursdays (except for pedestrian-only counts). MAPE across all days (except holidays) and seasons was 34% for 24-h and 20–22% for seven-day short duration counts. The magnitude of bicycle and pedestrian volumes did not significantly affect estimation errors. They also found that the length of short duration samples may influence accuracy of AADNT estimates more than the number of counters per group, all else equal for factor groups larger than one counter. They suggested using four or more counters per factor group for bicycle and five or more for pedestrian travel monitoring to maximize precision of estimates of AADNT (Turner et al., 2021) have guidance for collecting additional count data that can be used for non-motorized data collection. Availability of long-duration count data is important to holistically evaluate safety for pedestrian and bicyclists, where a combination of traditional and emerging technologies to count non-motorized traffic data can be done. Existing traffic monitoring activities in coordination with crowdsourced data and machine learning methods can lead to an incremental development of systematic active transportation (Tolford et al., 2019).

b. Travel Behavior Analysis

Crowdsourced approaches are also being used to analyze shift in travel behavior, such as mode-shift, understand mobility patterns, Vehicle Miles Traveled (VMT), etc. Marzano et al. (2019) provide a detailed review of the applications of crowdsourced data to understand urban mobility. To understand shifts in travel mode, Chandra et al. (2020) developed a crowdsourcing-based perception to estimate any changes in mode-shift behavior of college students in California State University, Long Beach. An empirical experiment was conducted with a sample of 30 participants spanning over two phases. Participants used one of the five modes transit bus, bicycling, walking, car and carpool to arrive at the university campus. In the first phase, a control was created by identifying the mode choice of participants and their numeric value of perception of each specific mode. In Phase II, the participants were asked to post their comments publicly anonymously on modes on a “Twitter” address used for this study each time they arrived at the campus. The crowdsourcing platform was utilized to observe mode choice of other participants without knowing their identity. Results showed an overall shift of users from private car to other modes of transportation.

Crowdsourced data from public transportation sources can also be used to understand the use of public transport as well as the barriers to its use and for route optimization of non-fixed guideway transit system based on data received adaptive vehicle navigator systems (Marzano et al., 2019). To understand how transit transfers are a safety issue that act as an impediment to transit use, Traut & Steinfeld (2019) used crowdsourced public transit ridership data to analyze transit transfers in Pittsburgh using the Tiramisu Transit app. Poor transit transfers can lead to both a real and perceived reduction in convenience and safety, and expose riders to bad weather and crime. This can reduce transit ridership by motivating riders who have the option of driving or using paratransit to elect a more expensive and inefficient travel mode. The Tiramisu Transit app merges open transit data with information contributed by users about which trips they take. They used the Tiramisu data to conduct origin-destination analysis and identify connecting trips to understand where and when poor transfers occurred in the Pittsburgh region. The results with data from other open public data sources (such as, crime data) were merged to create a data resource that can be used for planning and identification of locations where infrastructure improvements may lead to safer and more comfortable waits and more accessible transfers. The results found that 66.6% of transfers were within 0.4 km (0.25 mi.) and 44.1% of transfers were less than 10 min. They found several highly utilized transfer locations that were not identified by the Port Authority of Allegheny County as recommended transfer points, and so might need more planning attention.

c. Application to Traffic Operations

Apart from transportation planning, crowdsourced data is also being used to improve traffic operations due to the low-cost of collecting data using crowdsourced methods, unlike physical sensors that incur significant capital and maintenance costs. Dixit et al. (2020) developed a model to utilize real-time crowdsourced delay data to allocate the length of green time to a phase using real-time crowdsourced delay data. This approach is useful as it does not require any real-time traffic volume or queue length data, which require installation and frequent maintenance of multiple loop detectors or video detectors at each intersection. Also, physical sensors do not work properly in mixed modes and shared lane traffic. The model was analyzed for seven different intersections across three cities and two countries and worked well showcasing the benefit of shifting from physical sensors to low-cost, reliable crowdsourced data.

d. Safety Analysis

Using the crowdsourced data, several monitoring and decision support system for pedestrian safety have been developed. Although, decision support tools for pedestrian and bicyclist safety, in comparison to motorists, have been developed in the past as well, their impact was reduced due to limited availability of disaggregate data to measure impact at a more granular level (Torbic et al., 2019). Crowdsourced data expands the scope of decision making tools for pedestrian and bicyclist safety. Hamilton et al. (2021) developed a Highway Safety Information System (HSIS) using crowdsourced data for pedestrian safety in Charlotte, North Carolina (NC). They spatially integrated HSIS data with multi-jurisdictional and crowdsourced datasets to analyze two measures of pedestrian safety performance: the severity of a pedestrian crash that has occurred, and the probability that a pedestrian crash will occur. A pedestrian count model was developed to predict pedestrian volumes at locations without pedestrian counts and integrated speed information from probe data to supplement other roadway and contextual transportation data from several agencies. Results showed that higher pedestrian volumes resulted in both lower crash severities and probabilities, but the safety benefit was reduced by higher vehicle volumes.

From a pedestrian safety and justice perspective, crowdsourced data is also useful when traffic safety data collected through formal institutions may have undercounted data or underreported incidents. Medury et al. (2019) analyzed non-motorized concerns in and around three universities. They compared the police-reported crash data with traffic safety information crowdsourced from the campus communities themselves. The crowdsourced traffic safety included both self-reported crashes and perceived hazardous locations. The results suggested that police-reported crashes underrepresent non-motorized safety concerns in and around the campus regions. The spatial distribution of police-reported crashes showed that crashes are predominantly unavailable inside the main campus areas, and the off-campus crashes over-represent automobile involvement. In contrast, the self-reported crash results reported a wide variety of off-campus collisions not involving automobiles. They also showed the issue of high crash concentrations along campus boundaries. The perceived hazardous locations (PHLs) assessment indicated that high concentrations of such observations at/near a given location were statistically associated with both survey-reported crashes as well as future police-reported crashes. The findings suggested that existing knowledge of traffic safety can be improved through crowdsourcing to better estimate existing as well as emerging traffic safety concerns and not underestimate non-motorized safety concerns.

e. User perception Analysis with Crowdsourced Data

Beyond objective measures, crowdsourced data are also used to understand more subjective attributes related to perceptions of travelers, which can also be a part of decision support system for planning and safety of pedestrians and bicyclists. Rahman et al. (2021) used twitter geotagged data to evaluate the level of service of shared transportation facilities and analyze the perceptions of road users. Using text mining relevant information related to users' perceptions toward active mobility was filtered out. By analyzing the sentiments of the filtered data, the existing condition of biking and walking facilities could be inferred for a location. The results can also be used to produce relevant information on walking and biking facilities as well as safety concerns. Leveraging digital technologies, makes it easy to crowdsource data on user perception, such as perceived safety and realized travel together, that are often difficult to collect using traditional survey methods. Specifically, data related to near-misses or infrastructure problems that may affect bicycle safety are not systematically collected or analyzed. Blanc & Figliozzi (2017) developed a smartphone application, called ORcycle, to crowdsource bicycle travel and safety data in Oregon to explore which factors affect the urgency of a perceived safety problem. They found that the demographic variables, cyclists' gender, and income levels influence safety reports' urgency and type. Also, higher traffic volumes and speeds increase the urgency of safety reports. However, other variables such as long waiting times at traffic signals are associated with less urgent safety reports. In addition, a thorough quality control revealed a very high level of data accuracy, and the statistical models produces indicated that users are generally reliable when reporting the urgency. Also, crowdsourced data collection expands the scope about the type of data that can be collected for safety analysis. Rahman et al. (2019) developed an android-based crowdsourced app to collect perception data to better understand conflicts and their severity. They were particularly focused to understand the perceptions around less severe conflict locations where paths cross, but no evasive maneuver is required but may affect public perception of safety with using vulnerable modes. They concluded such crowdsourced data can enable communities to create their own data collection efforts and identify fatality hotspots within their neighborhoods. At a low-cost, agencies can help inform decision making related to bicycle and pedestrian education, encouragement, enforcement, programs, policies, and infrastructure design and planning.

f. Activity locations – spatial

Crowdsourced data with their granular temporal resolution allow to study both, mobility and activity patterns of people. Particularly with mobile-phone collecting data on movement of people, data mining techniques can be leveraged to useful activity and travel information. Yin & Leurent (2021) analyzed individual activity-travel patterns from a sample of mobile-phone users using a two-week geolocation data set from the Paris region in France. The goal was to understand the individual mobility patterns and reveal home-based differences in spatial distribution for individuals in the study. They classified activity spaces in primary anchor place and the secondary place and reconstructed activity-travel program with the detected activity places and the trips in-between. Using on user-day timelines, they did mobility pattern analysis using a three-stage clustering technique. Firstly, activity types were identified by clustering analysis. Next, daily mobility patterns were obtained after clustering the daily mobility features. Finally, the individual mobility patterns were analyzed for all samples over 14 days. They also identified similar travel behaviors across individual samples are divided into several groups.

g. Infrastructure – Accessibility

Crowdsourcing approaches are also being used to collect data on the condition of pedestrian and bike facilities infrastructure. The collected data are used with deep learning techniques to improve infrastructure that can increase walking and biking activity. Froehlich (2021) developed a remote crowdsourcing data collection technique to collect data on pedestrian-related accessibility as part of their project called Sidewalk. Online crowd workers remotely labelled pedestrian-relation accessibility-problems by virtually walking through city streets in Google Street View. Using this low-cost technique, 1,150+ users provided over 200,000 geo-located sidewalk accessibility labels and audited 3,000 miles of D.C. streets. They completed an 18-month deployment in Washington, D.C in 2019. The results showed that with simple quality control mechanisms, minimally trained remote crowd worker could find and label 92 percent of accessibility problems in street view scenes, including missing curb ramps, obstacles in the path, surface problems, and missing sidewalks. They plan to scale the project further to other cities and use deep learning techniques to automatically identify and classify sidewalk problems. Results from such studies can be developed as interactive visualization tools to give stakeholders—from citizens to transit authorities—new understanding of their city's accessibility.

Understanding of pedestrian-related accessibility can be used to classify streets based on a measure of local destination accessibility, complemented by counter data, which can be used to estimate area-wide seasonal average daily pedestrian counts and average daily pedestrian miles traveled. Gehrke et al. (2019) developed a pedestrian-oriented approach to classifying streets based on a measure of local destination accessibility along a given street segment, or its network utility, along with pedestrian count data collected from multiple randomly selected sites in four neighborhoods across Massachusetts. The approach is useful to measure active travel at a neighborhood scale to estimate the population-level impact of policy, systems, or environmental changes on transportation-related physical activity, since most studies focus on measuring area-wide levels of active travel.

Social media tools can also be leveraged to collect accessibility information. Tarkiainen et al. (2011) crowdsourced accessibility information from Points of Interest, POIs (e.g. restaurants, shops) especially in the Helsinki capital region in Finland as part of Supremo project. The objective was to support mobility of people that have restricted movement due to some reason. The collected information is useful for personalised journey planning, especially for disabled or elderly population or people with temporary mobility restrictions, such as people with infants that need to be drive in a push chair.

h. Others

Crowdsourced approaches are not only being used for bike and pedestrian count forecasting, but also in the case of micro-mobility services to better plan for these services. Elhenawy et al. (2021) developed a model to solve the problem of charging and maintaining a large number of light vehicles with this work done by the crowd of suppliers. The proposed model consists of three entities: suppliers, customers, and a management party responsible for receiving, renting, booking, and demand matching with offered resources. Suppliers can define the location of their private e-scooters/e-bikes and the period they are available for rent. The model was applied to 9 million e-scooter trips in Austin, Texas showed that the proposed model can be advantageous to shift the charging and maintenance efforts to a crowd of suppliers.

4. Issues with Crowdsourced Data

Crowdsourced data also have some issues. These are related to geographic coverage gaps, lags in information timeliness, life-cycle costs for field equipment, and jurisdictional stovepipes associated with fixed sensor and camera monitoring can limit agencies' abilities to proactively operate transportation systems (Every Day Counts: Innovation for a Nation on the Move, 2021).

One of the major issues with crowdsourced transportation data is the lack of information on the travel mode. Even if mode-unspecified is big with granular spatial and temporal resolutions without sorting out non-motorized travelers its use is limited. Commercial vendors often provide a vast volume of mode-unspecified data, but they are predominantly used for motorized trips analysis. Mode-specified data for non-motorized travel are mostly focused on bicycling. Despite the potential of emerging crowdsourced data, their use also has challenges, such as limited mode inference, sample bias, and lack of detailed trip/traveler information due to privacy issues. Data accuracy needs to be improve and robust data fusion techniques need to be developed to fully utilize the emerging data sources (Lee & Sener, 2020).

Another big issue when it comes to digitally produced crowdsourced data is data privacy. Differential privacy (DP) can be used to ensure privacy to crowdsourcers. Classical DP requires a centralized trusted data curator (DC) who collects all the responses from users and publishes anonymized statistical information. In addition, a "local differential privacy" (LDP) paradigm has been proposed to prevent an untrusted DC from learning and using the personal information of data providers. In LDP, each data provider randomizes their data locally before sending it to the (untrusted) DP who aggregates the data without having access to the personal information of the data providers (Marzano et al., 2019).

Apart from crowdsourced big data sources that are often digitally produced, transportation planners are increasingly using more participatory approaches, both in-person and online forms, in planning. While traditional participation methods focus on the use of language to involve people in planning processes, digital methods rely on broadband and smartphone access. It is important to recognize that some disadvantaged groups may not have to access these technologies, which may restrict opportunities for them. To analyze geography and equity outcomes of different participation methods for crowdsourced data collection, Griffin & Jiao (2019) conducted a study in Austin, Texas to co-produce informed plans for active transportation (bicycling and pedestrian) modes. The three approaches were in-person meetings, public participation geographic information system (PPGIS), and an emerging smartphone platform that logs trips and encourages input on route quality. They also conducted qualitative case analysis to contextualize the geographic and equity implications of different participation approaches. The results showed that both online techniques resulted in a larger geography for participation than in-person meetings, with the regional PPGIS covering the most area. Also, digital methods can be useful to plan at a megaregion scale due to its ability to capture crowdsourced response from a wider geographic area. However, reviewing the income levels in each area revealed the use of the smartphone-based crowdsourcing platform was aligned with lowest-income areas. Hence, online participation methods to crowdsource data are not homogeneous regarding geography or equity. Smartphone applications can help reach lower-income communities, even when compared with in-person meetings.

Managing both crowdsourced data and more traditional data sources in an integrated manner to guide planning for non-motorized modes is difficult. While, variety in data collection efforts creates a diverse dataset, it makes compilation of a single data archive difficult, especially when the geographic scale is at a state level or a region level. Karen Sentoff & James Sullivan (2017) create a unified bicycle and pedestrian count database for the state of Vermont to be able to communicate the situation of non-motorized travel statewide and make recommendations for future data collection and management. Some of their suggestions were:

- Creation of a data input tool that standardizes the data formats and response options based on national protocols and tailor them to the needs of the statewide count program.

- Creation of a new database with a linked Site ID. This will prevent data duplication and loss.
- Creation of a new web portal to view the existing count data in a site summary form or to download raw data. The new web portal will also have a fixed link to the new data input tool allowing for easy navigation to data input and output by all other entities statewide.
- They also recommended new count sites to take a more representative sampling approach and not focus just on sidewalks and multiuse paths.
- They suggested exploring correction factors for existing counts collected with automated counters throughout the state. Automated infrared counts can be multiplied by a correction factor of 1.16 to account for occlusion, but this factor is affected by the social context of the pedestrian activity at the site.
- They also suggested exploring the use of Strava data resources to complement the nonmotorized count data program. It can be a useful source to complete-screen line data when sidewalk or on-network multiuse path counts need to be supplemented with roadway volumes.

5. *Crowdsource Methods to Collect and Analyze Big Data*

d. Forecasting Methods

To leverage the crowdsourced big data sources, several methods have been developed using machine learning techniques to address the issues related to data quality, gaps in geographic availability of data, etc. Daily count can be imputed using several methods to monitor non-motorized traffic, when permanent count site data is unavailable. Random forest and day-of-year (DOY) factor approaches could be used to impute daily counts for nonmotorized traffic, but each approach comes with tradeoffs. Though for many missing data scenarios random forest performed best, this method is complicated to estimate and apply. DOY factor-based methods are simpler to create and apply, and though more accurate in scenarios with significant amounts of missing data, they were less flexible given the need for data from neighboring count sites. Negative binomial regression was also found to work well in scenarios with moderate to low amounts of missing data (Roll, 2021). Data Mining techniques can also be used in crash studies to estimate pedestrian and bicycle crash counts to find the most important variables influencing chances of a crash. M. S. Rahman et al. (2019) developed a decision tree regression (DTR) model to predict crashes for pedestrians and bicyclists for the state of Florida using the data from 2010 to 2012. They found that for pedestrian and bicycle crash count traffic, roadway, and socio demographic characteristics are the most significant factors influences crashes. In addition, using spatial variables of neighboring traffic analysis zones in the DTR model led to improved prediction accuracy compared to aspatial DTR model. They also compared applying three ensemble techniques (Bagging, Random Forest, and Boosting) to improve the prediction accuracy of weak learner (DTR models) for macro-level crash count. All the ensemble technique performed better than the DTR model and the gradient boosting technique outperformed amongst the three in macro-level crash prediction model. Pedestrian safety can be positively correlated with increased pedestrian traffic in a given area, which is known as Safety in Numbers (SIN). A study by Kristin Carlson et al. (2018) use alternative datasets to model crash frequencies at the intersection level as a function of modal traffic inputs in Minneapolis to check if SIN effect is observable using the available datasets. Detailed historical multimodal crash and traffic volume data are often not available at a granular level, especially for non-motorized transport flow levels. Pedestrian and cyclist traffic counts, average automobile traffic, and crash data from the city of Minneapolis are used for model development. The developed models help to analyze where the SIN effect is observable within the available datasets for pedestrians, cyclists, and cars. It also helps to know the locations within Minneapolis where non-motorized travelers experience elevated levels of risk of crashes with automobiles and need improvement of non-motorized facilities.

Even though big data sources typically have high temporal resolution, data may not be available for the entire day if they are crowdsourced from people based on their daily activity. These datasets could only be used to estimate short-term counts, which can then be used to estimate daily, weekly, or annual volumes using expansion factors. A problem is that the count may differ by location based on the activity pattern at the site. Griswold et al. (2018) proposed a method to develop factor groups for hour-to-week pedestrian count expansion factors with two different approaches. The land use (LU) classification approach assumes that surrounding land use affect the pedestrian activity at a location, and it is easy to apply to short-term count locations based on identifiable attributes of the site. The empirical clustering (EC) approach uses statistical methods to match locations based on the actual counts, which may produce more accurate volume estimates, but it is difficult to determine which factor group to apply to a location. However, both the LU and EC approaches provided better weekly pedestrian volume estimates compared to single factor approach of taking the average of all locations. Also, the differences between LU and EC estimation errors are minimal. The authors suggest using LU approach as they are easy to apply. Also, LU groupings can be modified with insights from the EC results, improving estimates. Also, the ideal times for short-term counts, which are used for estimation of long-term counts, are during peak activity periods, as they produce estimates with fewer errors than off-peak periods.

e. Methods to address Quality Issues in Count Data

Apart from it, crowdsourced data sources have quality issues even though typically have wide geographic coverage. One of them is highly variable measurement noise in the data due to a variety of users and sample size. If this noise is not accounted for during analysis, the application of the predictive models/ decision support system is severely compromised. To minimize the noise, studies have explored different methods. Rodrigues & Pereira (2018) propose the use of heteroscedastic Gaussian processes (HGP) to model the time-varying uncertainty in large-scale crowdsourced traffic data. They have developed a HGP conditioned on sample size and traffic regime (SSRC-HGP), which makes use of sample size information (probe vehicles per minute) as well as previous observed speeds, to more accurately model the uncertainty in observed speeds. The results show significantly better predictive distributions when compared to current state-of-the-art methods for both speed imputation and short-term forecasting tasks for the case of Copenhagen, Denmark.

Another issue is the counting multiple pedestrians walking together. To resolve this issue, Shi et al. (2018) developed a modeling algorithm to count multi-pedestrian candidates. This approach, firstly a background modeling algorithm is applied to actively obtain multi-pedestrian candidates. This is followed by a confirmation step with classification. Next, each pedestrian patch is handled by real-time TLD (Tracking-Learning-Detection) to get a new predication position according to similarity measure. The TLD results are also compared with classification list to determine a new, disappeared, or existing pedestrian. Finally, the single line counting with buffer zone is employed to count pedestrians.

f. Image and Video Processing Methods

In addition to GPS data to predict spatial patterns, machine learning tools have also enabled using images as a tool for predicting spatial patterns of travel for bicycle and pedestrian travel. Hankey et al. (2021) used destinations from Google Point of Interest data (e.g., restaurants, schools) and pixel classification from Google Street View imagery (e.g., sidewalks, trees, streetlights) to model bicycle and pedestrian traffic at 4145 count locations across 20 U.S. cities using new micro-scale variables. Deep learning methods were used to evaluate how well street-level variables predict bicycling and walking rates. Adding street-level variables improved out-of-sample prediction accuracy of bicycling and walking activities. Street-level variables can be a useful alternative to Census data.

A limitation of video-based traffic counting technology is the reliability and accuracy of the information extracted. The level of errors in computer vision-based sensing technology can be relatively large as it is highly sensitive to environmental factors, such as illumination, weather conditions, and occlusion. In addition, traffic counts can be inaccurate when video-based technologies are used to estimate counts in multi-modal traffic. In an ongoing project led by Center for Connected Multimodal Mobility in Clemson University, data fusion techniques are being explored to improve accuracy of multi-modal traffic counts. The information extracted from video data can be complemented from other data sources, such as tube counters, magnetic loops, radar, vibration, and laser measurements. In the project, the team is using the combined raw data from the tube-based vehicle counting/classification method and an integrated artificial neural network (ANN) developed using computer vision-based sensing technology to classify vehicle types with better accuracy than existing methods using data from one type of sensor. Data fusion methods are being explored to integrate not only data of different types but also multiple sources of video data as well. Huang et al. (2018) proposed a method to estimate pedestrian counts based on multisource video data. Firstly, partial least squares regression (PLSR) model is developed to estimate the number of pedestrians from single-source video (either visible light video or infrared video). The temporal feature of the scenario (daytime or nighttime) is identified based on visible light video as well. Using the recognized time periods, pedestrian count detection results from the visible light and infrared video data are obtained with preset corresponding confidence levels. The empirical results showed that this fusion method substantially improved accuracy of pedestrian counting and can be helpful in monitoring 24-hours, especially when the pedestrian waiting area is outdoors.

Deep learning for image processing can also be used for pedestrian surveillance to improve their safety (Baqui et al.2020) developed an automatic and improved high-density pedestrian traffic (HDPT) surveillance system by integrating and optimizing multiple computational steps to predict pedestrian distribution from input video frames. A fast and efficient particle image velocimetry (PIV) technique was used to yield pedestrian velocities. Boosted Ferns, a machine learning regressor model, was used to improve pedestrian count and density estimation: an essential metric for HDPT analysis. A camera perspective model was used to improve the speed and position estimates of HDPT by projecting 2D image pixels to 3D world-coordinate dataa. These functional improvements in HDPT velocity and displacement estimations were used as inputs to a pedestrian flow evolution model, PEDFLOW to predict HDPT distribution at a future time point. The results show that predicted and simulated HDPT properties (density, velocity) obtained using the proposed framework led to low errors when compared to the ground truth data.

g. Other technologies to Crowdsource Count and Other Data Related to Pedestrian and Bicyclists

Alternate technologies can also be used to crowdsource pedestrian count data. Wi-Fi channel state information (CSI) can be used to analyze human movements. Wi-Fi CSI represents the amplitudes and phases information for orthogonal frequency-division multiplexing (OFDM) subcarriers, which is mainly impacted by the static environment and moving object in surrounding areas. It is a step ahead from traditional sensing technologies usually sense pedestrians, based on the reflected signal of the transmitted infrared ray, sound wave, or electromagnetic wave which only can count the number of times that pedestrians passing a line of sight (LoS). Also, it eliminates the issues of errors with image processing due to environmental factors. Pu et al. (2020) demonstrated the use of Wi-Fi CSI-based sensing method for pedestrian existence and moving direction recognition. They conducted experiments in both indoor and outdoor environments. According to the results, the accuracy of pedestrian existence detection based on the data of the 100 Hz sampling ratio achieved 99.23% accuracy and 0.26% fast positive rate. For the moving direction recognition, the detection accuracy in the indoor environment achieved 100% and

96.92% for two directions and got 92.21% and 93.51% in the outdoor environment. Alternative low-cost systems for counting pedestrians are also being developed using embedded systems capable of performing in real-time under high volume flow. Lesani et al. (2020) developed a real-time counting system to monitor high pedestrian flow using two-dimensional LiDAR sensor. The system used the distance measurements from a two-dimensional LiDAR sensor with a set of distinct laser channels and a given angular resolution between each channel. The measurements were processed using a clustering algorithm to detect, count, and identify the direction of travel of each pedestrian. The results showed that the system accurately counts more than 97% of the pedestrians at the disaggregate level, with a false direction detection rate of 1.1% when compared to manual counts. The over-counting error is 0.7% and the under-counting errors are 1.3% and 2.7% for the two selected sites. At the aggregate level (15-minutes interval), the average absolute percentage deviations (AAPDs) are 1.6% and 4.3% while the weighted AAPDs are 1.5% and 3.5% for the first and second sites, respectively. Blue-tooth censors are also being used to counter pedestrian flow to develop real-time pedestrian monitoring system. In another study by Lesani & Miranda-Moreno (2019), they used unique media access control (MAC) addresses of mobile devices carried by pedestrians, captured from Bluetooth (BT) sensors, using WiFi signals. This method is advantageous over just using Bluetooth sensors as it may suffer from low-detection rates. It provides information about traffic flow, speeds, travel times, and time spent in areas or transportation facilities of interest to generate origin-destination information, trip paths, travel times, or time spent, which cannot be provided by fixed counters. The results showed that high detection rates for the developed WiFi system in comparison with BT sensors.

In addition to monitoring pedestrian activity on footpaths, technological advancements are also being used to monitor pedestrians flow during special events. Olfert et al. (2018) presented results from a pedestrian monitoring study done to identify counter sites using infrared pedestrian counters in downtown Winnipeg, Canada. Count sites were allocated to traffic pattern groups (TPGs) based on their response to special events occurring in the study area. These groups enable the spatial variation of short duration counts to be adjusted to annual statistics by the temporal variation of similarly behaving continuous counts. Once groups were defined, eight continuous count sites were installed to initiate an ongoing pedestrian traffic monitoring program for the city. Short-duration count sites were characterized by daily and hourly trends to be in line with existing pedestrian traffic monitoring practices. A metric the evening proportion ratio (EPR) was developed to quantify the effect of special events. For downtown Winnipeg, two TPGs were developed. These were the “urban utilitarian” and “urban utilitarian – event” groups. These groups were used to select continuous count locations to have an ongoing pedestrian traffic data collection.

6. *Crowdsourced Data Ethics*

The rapidly generated crowdsourced big data sources have increased access to new types of transportation data with characteristics that include improved quality, increased temporal and wide geographic coverage compared to traditional datasets. However, such datasets are often proprietary in nature. State DOT’s and Metropolitan Organizations (MPO) face difficulties with obtaining these datasets. which can fill in the gaps in knowledge of travel behavior and mobility, in general due to limitations of traditional datasets. For instance, speed data are being used by transportation agencies across the United States for a variety of applications. Also, O-D data produced by highly precise GPS data from in-vehicle systems and mobile phones are being used for demand forecasting. However, the agencies face some barriers with these proprietary datasets in terms of data and service quality, cost, staff expertise and information technology resources, finding the right product, and legal issues (Chen et al., 2019).

Chapter 2. State Agency Survey Data Collection/Analysis

1.0 Introduction

As part of the project, a survey was designed to collect information on the current state of the practice on bicyclist and pedestrian data collection for volume and safety/risk assessment. The survey was developed by University of Michigan Transportation Research Institute (UMTRI) with input from Highway Safety Research Center, UNC-Chapel Hill (HSRC) and in consultation with MDOT. Qualtrics, an online third-party survey software provider was used to implement and deploy the survey. The license to the software was provided by the University of Michigan (UM). The survey questionnaire and protocol were submitted to the UM Institutional Review Board (IRB) and was deemed to be exempted from IRB review because of the anonymity and non-sensitive nature of the data to be collected. The survey was then distributed through anonymous links sent to contacts within government agencies and state department of transportation (DOTs). These contacts were obtained through UMTRI's ongoing collaboration with state DOTs for safety data collection and through snowballing effect i.e., every contact was requested to provide us with emails of people and colleagues they thought would have information pertinent to the survey.

The survey has been open for collecting responses on November 23, 2021, after two sets of beta testing internally and with MDOT. This report analyzes the seventeen (17) responses received until December 26, 2021. The report is organized in three sections following the survey design – the first section provides information on the participant; the second section provides information on the current state of the practice as reported by the participants and the third section provides information on use of crowdsourced data and the opinion of the participants related to quality and use of crowdsourced data for volume and safety/risk estimation. Detailed analysis of the survey responses is presented in Appendix B for brevity and conciseness of the report.

2.0 Survey Findings

Section 1: The Participant: Affiliation, Experience and Role

The first section of the survey asked questions about the respondent's professional experience and role to understand the validity and strength of their responses. As the primary candidates for the survey were DOT representatives, the response distribution in terms of affiliation reflects the same – 14 out of 17 responses are from state DOTs, one each from a city and an MPO and one respondent was from Office of Public Safety, Traffic Safety. In terms of role within their respective organization, seven (7) respondents were bicyclist and pedestrian coordinator, five (5) respondents were in planning related roles and five (5) respondents were in safety related roles. The majority of the respondents had bicyclist and pedestrian safety and risk as their primary responsibility followed by respondents with responsibilities in volume and data collection. There were also respondents with responsibilities in project planning, research, project implementation etc. Most of the respondents (8 out of 17) had experience in the range of 1-3 years. There were three (3) respondents each in the 4 to7 year, 8 to10 year and 10+ years of experience categories.

Section 2. Current practices and experiences with data collection

In this section, the respondents were asked questions about their current practice of data collection and their experiences and opinions about the practices. These questions were asked based on their response to their choices in responsibilities – for example, the respondents who chose only bicyclist volume as main responsibility, was presented the questions about bicyclist volume data sources while respondents who chose pedestrian volume as the only responsibility, were asked the questions for pedestrian volume data sources. Respondents who chose both pedestrian and bicyclist volume were asked questions separately for bicyclist and pedestrian volume data. Similarly, respondents who indicated volume and safety data to be their responsibilities were asked questions about both volume data and safety data.

For bicyclist volume, three (3) respondents ranked manual count at the first place while three (3) other respondents ranked permanent count stations as the first. Two (2) respondents ranked third party data as first. Short duration counts are overwhelmingly popular as second choice with six (6) respondents giving it second rank. From the number of respondents and ranking, it appears that permanent count stations, short duration count stations, manual counts and third-party data are the most prevalent data types used for bicyclist volume estimation. For bicyclist safety, the overwhelming majority listed state crash data as the most important source of data. National crash data and other data sources were chosen as second choices by some of the respondents. Third party data also appear to be a popular choice, although not the first choice.

For pedestrian volume, the preferred source is manual count followed by automatic counters. However, unlike bicyclist volume, none of the respondents chose third party data for pedestrian volume data. Instead, estimation models and travel surveys seem to be more relevant for pedestrian volume data. For pedestrian safety and risk, similar to bicyclist safety and risk, state crash data appears to be the most important source followed by national crash data and third-party data.

For data quality questions, most respondents appear to be somewhat or moderately satisfied with the data sources that they are using for volume estimation and mostly satisfied with the data sources that they are using for safety/risk estimation on account of usefulness, accuracy and data accessibility. These questions were asked using a Likert scale where respondents had to indicate how satisfied (not at all, somewhat, moderately, mostly, extremely, cannot say) they were with the current data sources for accuracy, usefulness and data accessibility. None of the respondents indicated being not at all satisfied or extremely satisfied with the data sources for volume estimation whereas some respondents were extremely satisfied with the data source that they use for safety/risk estimation, and none noted to be not at all satisfied. Overall, the respondents appear to be more satisfied with the safety/risk data sources than volume data sources in all the three aspects of accuracy, usefulness and data accessibility.

Section 3. Crowdsourced Data

In this section, the respondents were asked questions related to their use of and experience with crowdsourced data. Specifically, the respondents were asked if their organization had used crowdsourced data for volume or safety/risk data, which data sources they had used or considered using, the reasons behind using those data sources as well as their reasons for not using crowdsourced data. As with the previous section, the respondents who indicated that they had used crowdsourced data, were asked to rate how satisfied they were with the crowdsourced data in terms of accuracy, coverage, relevance, representativeness, cost, timeliness, scalability, quality, technical support.

Of the participants responding to these questions, five (5) respondents had used crowdsourced data for both bicycling and pedestrian planning purposes while five (5) other respondents had used crowdsourced data for bicycling planning only. 3 of the respondents did not use crowdsourced data and one respondent could not say for certain. Respondents who had used crowdsourced data had used Strava and Streetlight. One respondent mentioned the University of South Florida Center for Urban Transportation Research (CUTR) and University of North Florida Public Opinion Research Lab (PORL). The reasons for using the particular data provider were mostly cost – the data were either provided free of cost for research purposes or at reduced cost for evaluation or the agency already had contract with the data provider. One respondent mentioned representativeness and better estimation of demand and another respondent mentioned survey data from target audience from research institutes like CUTR and PORL helped them validate data collected from public opinion surveys.

On satisfaction with volume estimation for respondents who had used crowdsourced data for only bicycling planning, crowdsourced data sources had a mean score of 2 on a 3-point Likert scale for accuracy, coverage, relevance and representativeness indicating most respondents were neither satisfied nor dissatisfied with the data. For cost and timeliness, the mean score was 1.6 and for scalability, quality and technical support the mean score was 1.8 indicating respondents were mostly satisfied with these features of the data sources. On satisfaction with the data for safety/risk estimation, the same group of respondents gave a mean score of 2 for accuracy, coverage and relevance but 2.4 for representativeness, indicating dissatisfaction on the data for that aspect. The mean score of cost and timeliness was 1.6 and 1.8 for scalability and technical support. Quality had a mean score of 2 indicating respondents were mostly neither satisfied nor dissatisfied with that aspect of the data.

Respondents who had used crowdsourced data for both bicycling and pedestrian planning purposes consistently indicated dissatisfaction on all aspects of the data sources for pedestrian volume estimation. On the contrary, for bicycling, they were either satisfied or neutral for most of the aspects of the data sources. For bicycling volume estimation, accuracy, relevance, cost, quality and technical support got a mean score of 1.75, indicating mostly satisfied users while coverage, timeliness and scalability received a mean score of 2, indicating neutral users. Representativeness, however, received a mean score of 2.5 indicating respondents were dissatisfied with this aspect of the data. For safety/risk estimation, accuracy, relevance, timeliness, quality received an average score of 1.75 indicating respondents were mostly satisfied with this aspect while all other aspects received a mean score of 2 indicating neither satisfaction nor dissatisfaction. As mentioned earlier, for pedestrian volume estimation, only cost, quality and technical support received a neutral mean score and all other aspects received a score greater than 2, indicating respondents were dissatisfied with the data sources on those aspects, especially coverage and representativeness. For pedestrian safety/risk estimation though, respondents mostly were neutral on most aspects of the data sources.

Of the four (4) respondents who indicated that they had not used crowdsourced data previously, two (2) respondents had considered using crowdsourced data and two (2) respondents did not. The crowdsourced data sources mostly considered were Strava and Streetlight, but also included Ford Insight, ESRI, HERE and public opinion and survey software like Qualtrics. The major reason for not using crowdsourced data was cost followed by concerns about coverage and quality. In free form answers, respondents mentioned lack of benchmarking or study comparing crowdsourced data with other data sources, concerns about good coverage in urban or densely populated areas but not so much in rural areas as well as lack of non-recreational users, BIPOC and low-income bicyclists and pedestrians.

Overall, it appears that the respondents who had used crowdsourced data, had used it because it was made available to them at a low cost, for evaluation purposes. It also appears that they found the data quality, accuracy and technical support to their satisfaction level. However, combining with comments from respondents who have not used crowdsourced data, it seems that coverage and representativeness are issues that need further research or support using other data sources. Strava and Streetlight are the most commonly used or investigated data sources, which partially explains why respondents using crowdsourced bicycling volume data are mostly satisfied – Strava predominantly caters to that data need and provides data free of cost to government agencies. Most representatives also seem to be fairly neutral about their current data sources like permanent counters or manual counts, i.e., they are not dissatisfied with the accuracy, relevance or usefulness of the data sources. Our actionable takeaways from this survey responses are: (i) to investigate Strava and Streetlight data for coverage and representation in Michigan, (ii) to investigate additional crowdsourced data sources like Ford Insight and ESRI for usability and compatibility, (iii) to reach out to Oregon DOT to gather further insight into their recent use of crowdsourced data, and (iv) identify resources for pedestrian volume data.

Phase 2: August 2023- August 2024

Chapter 3. Ground truth Data and Crowdsourced Data Collection, Processing and Validation

1.0 Introduction

This chapter provides an overview of the counter and crowdsourced data collected and obtained, preparation and processing of such data for further analysis, a comparison of these two data sources and preliminary information on other data sources that could be used as crowdsourced data sources. The chapter is organized as follows: in Section 2, details of ground truth or counter data collection and preparation is presented; in Section 3, an overview of the different crowdsourced data sources is presented along with a comparison between these different sources, followed by detail review of the crowdsourced data used in this project; in Section 4, crowdsourced and counter data comparison and relationship analysis is presented, followed by spatial relationship analysis in Section 5.

2.0 Ground Truth Counter Data Collection

In order to establish some sense of ground truth, data from existing bicycle and pedestrian counters were gathered from around the state from 2018 through 2022. Sites where these data were collected are listed in Table 1 below. Most of the sites are on paved or unpaved paths, but some are on protected or standard bike lanes or sidewalks. Most of the data are from urban or suburban areas, but a few are from rural areas. Additional rural and suburban sites were collected as part of summer 2022 data collection conducted by the project team (Table 2).

The data were aggregated to the monthly level using the monthly average daily traffic (MADT) metric as described in Equation 1 below for the warmer months: May through September. Colder months were not included because bicycle and pedestrian traffic in these months is low due to snow and cold temperatures and are thus highly variable.

The MADT (at a given site) is an average of averages, of averages. The first average is the total counts for each day of the week, which gives seven (7) values, and the second average is the average of those seven (7) values together to get the Average Daily Traffic. An average of average daily traffic across months is taken to get the Average MADT value using the following equation:

$$MADT = \frac{1}{7}\sum_{j=1}^{7}\left[\frac{1}{n_{jm}}\sum_{i=1}^{n_{jm}} Vol_{ijm}\right] \qquad \text{---------------------------- (1)}$$

where

$MADT$ = Monthly Average Daily Pedestrian or Bicycle or both Pedestrian and Bicycle Traffic

Vol_{ijm} = traffic volume for the i^{th} occurrence during the j^{th} day of the week within the m^{th} month

i = occurrence of particular day of the week in a particular month ($I = 1 \ldots\ldots n_{jm}$) for which traffic volumes are available

j = day of the week (*j* = 1, 2….7)

m = month of the year (*m* = 1, 2…12)

n_{jm} = count of the j^{th} day of the week during the m^{th} month of the year for which traffic volume is available

The average MADT for bicycle or pedestrian or both combined was also computed and shown in the maps in Figures 1-3.

Table 1. Count Sites Used for Ground Truth

Site Name	Municipality / County	Facility Type	Mode	Months with MADT
Dexter-Huron	Washtenaw County	Path	Pedestrian	10
			Bicycle	10
Kensington	Harrison Township	Path	Pedestrian	10
			Bicycle	10
Lake St. Clair	Harrison Township	Path	Pedestrian	3
			Bicycle	3
Willow	Huron Charter Township	Path	Pedestrian	10
			Bicycle	10
Adams - Cass	Detroit	Sidewalk / Separated Bike Lane	Pedestrian	14
			Bicycle	15
Canfield - Cass	Detroit	Sidewalk / Separated Bike Lane	Pedestrian	10
			Bicycle	15
Dequindre Cut @ Gratiot	Detroit	Paved Rail Trail	Pedestrian	15
			Bicycle	15
Dequindre Cut @ Mack	Detroit	Paved Rail Trail	Pedestrian	15
			Bicycle	15
Dequindre Cut @ Wilkins	Detroit	Paved Rail Trail	Pedestrian	8
			Bicycle	10
Cass/Warren	Detroit	Separated Bike Lane	Bicycle	7
Cass/Michigan	Detroit	Separated Bike Lane	Bicycle	15
Kirby - Cass	Detroit		Pedestrian	14

Site Name	Municipality / County	Facility Type	Mode	Months with MADT
		Sidewalk / Separated Bike Lane	Bicycle	15
Lafayette - Cass	Detroit	Sidewalk / Separated Bike Lane	Pedestrian	13
			Bicycle	14
Milwaukee - Cass	Detroit	Sidewalk / Separated Bike Lane	Pedestrian	10
			Bicycle	12
Temple - Cass	Detroit	Sidewalk / Separated Bike Lane	Pedestrian	11
			Bicycle	13
Dexter - Fire Station	Washtenaw County	Path	Pedestrian	19
			Bicycle	24
Dexter DPW	Washtenaw County	Path	Pedestrian	25
			Bicycle	25
Pittsfield - Multi	Washtenaw County	Path	Pedestrian	19
			Bicycle	21
Chocolay Bayou Bridge	Marquette	Path	Pedestrian / Bicycle	14
Negaunee Jackson Mine	Negaunee	Path	Pedestrian / Bicycle	22
Pellet Pavilion	Marquette	Path	Pedestrian / Bicycle	20
Galesburg	Kalamazoo	Path	Pedestrian / Bicycle	20
Comstock (East)	Kalamazoo	Path	Pedestrian / Bicycle	25
M-96	Kalamazoo	Path	Pedestrian / Bicycle	10
Northside	Kalamazoo	Path	Pedestrian / Bicycle	20
Drake	Kalamazoo	Path	Pedestrian / Bicycle	20

Site Name	Municipality / County	Facility Type	Mode	Months with MADT
D Ave	Kalamazoo	Path	Pedestrian / Bicycle	2
Chevy Commons	Flint	Path	Pedestrian	14
			Bicycle	14
Irish Road	Flint	Path	Pedestrian	6
			Bicycle	6
Genesee Road	Flint	Path	Pedestrian	10
			Bicycle	10
Linden	Linden	Path	Pedestrian	2
			Bicycle	2

In the summer of 2022, data were collected at four additional temporary short duration count sites for at least one month using a MobileMULTI Eco-Counter brand device that combines passive infrared and pneumatic tube technologies in order to separate bicycle from pedestrian volumes at a given site. Table 2 lists these additional sites.

Table 2. Short Duration Count Sites

Site Name	Municipality	Facility Type	Mode	Months with MADT
Bonisteel Blvd	Ann Arbor	Sidewalk / Road	Pedestrian	1
			Bicycle	1
Gallup Park Pathway	Ann Arbor	Path	Pedestrian	1
			Bicycle	1
I-275 Metro Trail	Romulus	Path	Pedestrian	3
			Bicycle	2
Mike Levine Lakelands Trail	Stockbridge	Path	Pedestrian	1
			Bicycle	1

2.1 Data Preparation

The raw data from the counters was analyzed to remove days with suspicious counts. These include more than three days with 0 counts, anomalously high or low counts, any period of days with sharp differences between daily counts and days with fewer than 23 hours of counts (23 hours was chosen because the first day of daylight saving time in the spring has only 23 hours). Days with particularly high or low counts were further checked for weather conditions that might have affected trail use (the team used historic weather data from the National Weather Service to verify), any event that might have occurred on the trail using a web search, and hourly counts were checked to determine if there was a specific time for any high volume. Suspicious counts were flagged and removed from the MADT calculation. Any month with fewer than three weeks of counts for any day of the week was excluded from the process.

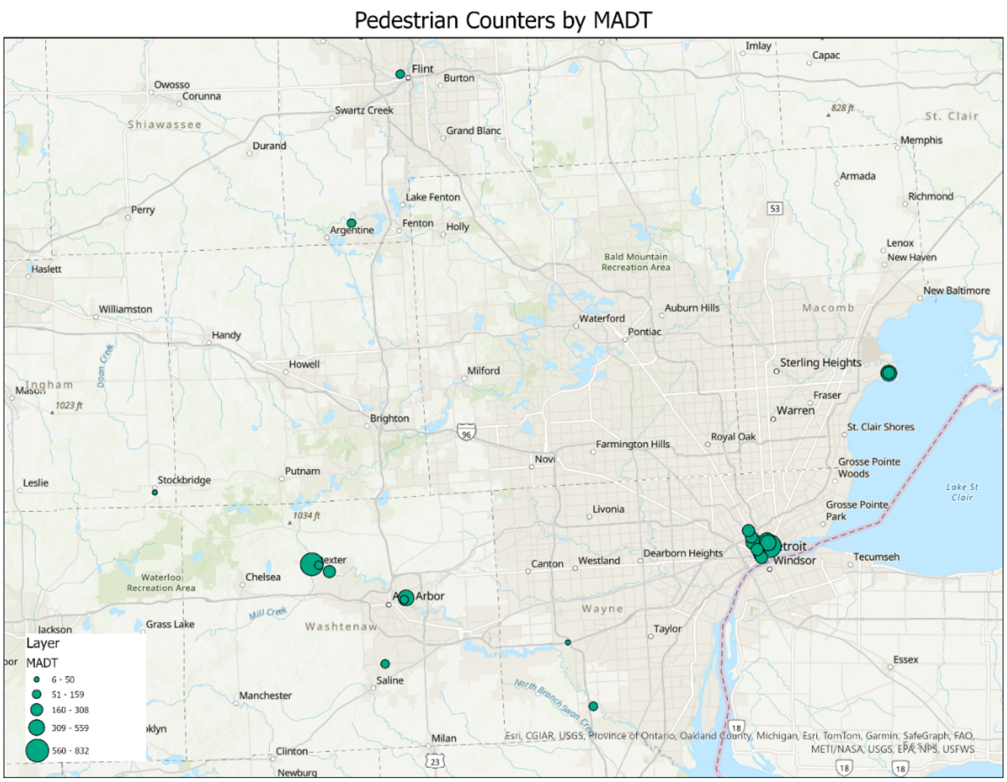


Figure 1. Pedestrian Counters by MADT Values

Bicycle Counters by MADT

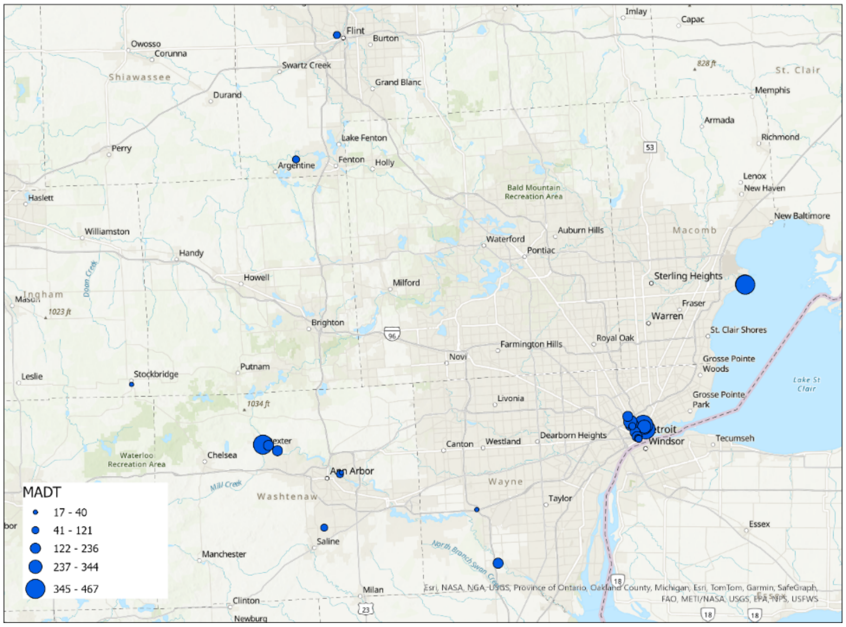


Figure 2. Bicycle Counters by MADT Values

Counters with Combined Bicycle and Pedestrian MADT

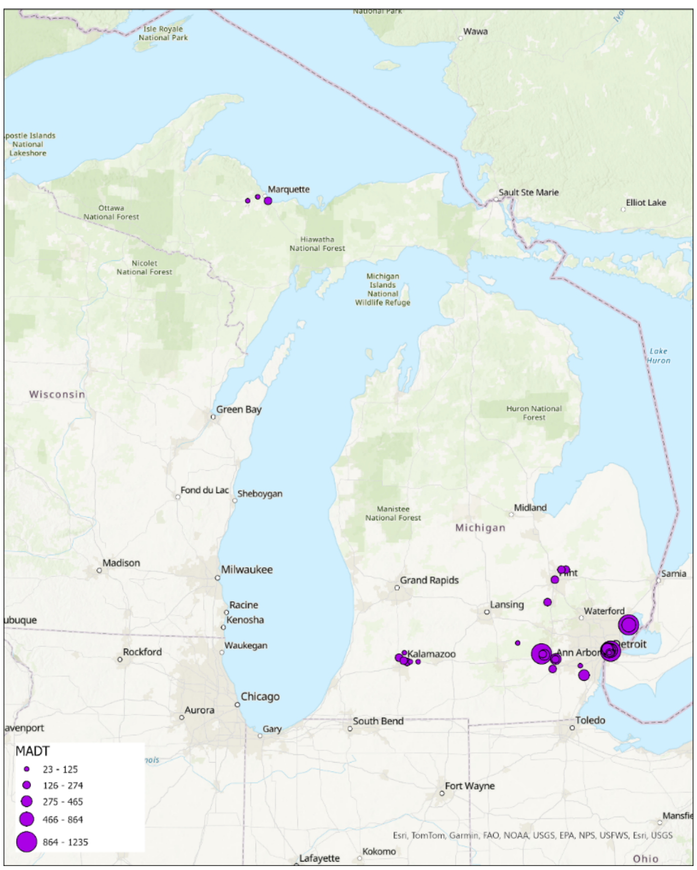


Figure 3. Combined Bicycle-pedestrian Counters by MADT Values

3.0 Comparison of Crowdsourced Data

Crowdsourced data are widely being used in transportation studies because of the ease of capturing real time data with numerous and widespread observations at a lower cost than traditional sources. StreetLight, Strava, and SafeGraph are the three most common sources of crowdsourced data used in transportation studies. Each of these data sources have a different approach to data collation. Below is an analysis of the strengths and weaknesses of the three types of crowdsourced data and an example of them in comparison to traditionally collected data.

3.1 Crowdsourced Data Sources

SafeGraph

SafeGraph products are created from a combination of machine learning, web crawling, and third-party licensing. The foot traffic dataset of 'Places Patterns' provides visit counts and dwell-time data for Places of Interest. It is built by licensing aggregated and anonymized mobility data, sourced from mobile applications of users that have opted-in to share their location. The Core Places and Geometry datasets are used to derive visit attribution to specific places.

The Places Patterns is a robust dataset that includes:

- How often do people visit a location or Census Block Group
- How long do they stay
- Where they come from
- Where else do they go

The dataset also provides insights into where people travel from to get to a specific place, and where else they go. The trip origin information is aggregated at the Census Block Group (CBG) level and differential privacy is applied to enable analytics at an optimal geographic scale.

Similar to Places Patterns, the Neighborhood Patterns dataset also gives aggregated and anonymized mobility data sourced from third-party applications. The difference with Places patterns is the level at which SafeGraph aggregates the foot traffic counts. Neighborhood Patterns focuses exclusively on CBG-to-CBG mobility, providing device counts, dwell times, and origin locations for specific timeframes. The same methodology used for visit attribution to create Places Patterns is used but at the CBG-level.

Strengths:

- High Quality of data
- Reliable Outcomes with Little geographical bias: A test for bias is done regularly by comparing the panel to the true proportions reported in the US Census.
- The pattern data is available in various time formats -
    - Weekly Patterns: The data provides the same foot traffic data insights from Patterns, updated weekly.
    - Neighbourhood Patterns: Is Anonymized and aggregated foot traffic and mobility data to census block groups (CBG) to help understand consumer behaviour.

- The pattern data at the destination is available at different geographic levels of individual places and an aggregated level of CBG. The data is available across all of the US.

Weaknesses:

- Misleading patterns can be captured in trips where walk trips are made from parking destinations or transit stops to the place

Opportunities:

- Can find other details such as related same-day stores or other shops in a mall by adding a couple of lines to the original query.

Strava

Strava is used by millions of people to track their rides, runs, and walks. Strava Metro provides aggregated information on travel patterns captured by Strava.

Strengths:

- The data provided through the Strava Metro platform has been aggregated and deidentified, consistent with the European Union's GDPR and the California Consumer Privacy Act (CCPA) to maintain the anonymity of users.

- The data is based on all non-private activities, excluding the ones that have been made private by Strava members and for members who have requested to have their accounts deleted.

- Commute purpose is identified based on a model developed by Strava and uses the commute tag used by Strava members as a ground truth. It is not necessary for Strava members to mark their activity as a commute in the app for it to be included in commuting analyses on the Metro platform.

- Strava Metro provides three licenses that can be purchased based upon data aggregation, which increases flexibility for data analysis and management units:

    - node (point)

    - street (segment)

    - Origin - Destination (polygon)

The shapefiles are useful for customizing analyses in GIS software, and finer spatial/temporal resolution—all of which are broadly applicable from a small scale to large scale and compatible with other data sources. While trip purpose filtering is possible (commute and non-commute) at the aggregate level, trip, and demographic information is not available at the discrete level due to privacy issues.

- The data product provides minute-to-minute data, rolled-up summary data, geometry files, and demographic files.

- The data sets do not require completion of initial data mining processes, i.e. access to data sets already cleaned, smoothed, and matched to network geometry by analytic teams is enabled. It saves time and resources when preprocessing raw GPS trajectories.

- It offers extensive data coverage in time and space at a relatively reasonable cost.
- Strava provides roll-up files, customized to the needs of the user, which provide a set of summarized Strava counts at requested temporal scales. It may be useful for correlation analysis with counter data. The dataset also contains a demographic summary file, including average distance, median distance, average time, and median time.

Weaknesses:

- Even if the Strava sample is not insignificant when compared to the total population, the Strava sample population can underrepresent general populations, and overrepresent certain populations. This is the biggest issue related to sampling drawbacks that must be considered when framing research questions and interpreting results.
- Strava data cannot guarantee that users of the application have not mistakenly recorded a vehicle trip as a part of a bicycling trip. Median speeds need to be checked to ensure that the trip is a bike trip.
- Sometimes mobile device GPS points do not match precisely with the roadway network, which can lead to incorrect facility assignments.

Opportunities:

- As Strava is one of the most popular fitness trackers globally, a number of app users contribute to an accumulated database and there are many customers around the globe.
- Strava members' travel patterns are representative of the overall population and can be combined with additional sources of fixed locations count data, generate adjustment factors, and produce very robust insights about the entire network. It does not use cellular tower and ad-network derived data.

Streetlight

Streetlight Data provides access to data sets that have significantly comparable attributes to Strava Metro data overall. Streetlight Data's Metrics are currently derived from two types of locational "Big Data": navigation-GPS data (INRIX, used to differentiate commercial truck trips from personal vehicle trips) and Location-Based Services (LBS) data (Cubeiq). However, bicycle and pedestrian data are reliant on the LBS data, the availability of which were greatly reduced after April 2022 when privacy setting changes allowed smart phone app users to more easily opt out of automated location detection. Streetlight data is processed through multiple steps including data extraction, transformation and loading onto the algorithm for cleaning, quality assurance, creating and contextualizing trips and activities and finally normalizing (Streetlight 2018). The cleaned data is then stored in a secured data repository and provided to end users in aggregated form based on their queries. Prior to April 2022, penetration rates for individual analyses can range from as small as 1% to as large as 35%, and typical daily trip penetration rates are between 1 and 5% of all trips on any one specific day.

After April 2022, Streetlight no longer provides road segment level bicycle and pedestrian data due to the decrease in availability of LBS data from which volume estimates and indices had been derived prior to that date. After April 2022, Streetlight is working on providing bicycle and pedestrian volume metrics at the census tract level or larger area wide estimates.

Strengths:

- It provides multi-app location-based data in the form of origin-destination (OD) based travel demand, aggregated or averaged traffic parameters (e.g., volume, distance, time, and speed) for selected settings (time and geometry), and deduced contextual information (e.g., trip type and income levels).
- It is better than other data sources in terms of sampling bias since it is integrated and validated with various other sources of data (e.g., active mode app, in-road sensor, video reader, and traditional travel survey).
- Streetlight has algorithms developed for mode recognition and data fusion.
- The on-demand analytic service provides OD travel demand (trip volume between OD and within OD), traffic attributes (e.g., volume, distance, time, and speed) for the selected time frame (e.g., day of week and time of day), and geometry (e.g., zone, link, or city), and inferred context information (sociodemographic and trip purpose).
- It provides comprehensive sample size information for analyses.

Weaknesses:

- More detailed information (e.g., information on reasons behind a mode or trip route choice and how individual sociodemographic characteristics affect the decision) cannot be solved because such level of details is not allowed due to privacy invasion issues.
- Sampling issues as the sample size is limited to distinguish casual bike users vs regular membership users cannot be done.
- The app does not provide data at the individual person/trip level like Strava Metro.
- The StreetLight Index for GPS data is normalized to adjust for changes in sample size. It is not normalized for population sampling bias (because home blocks for GPS data cannot be deduced based on data). Hence, it is advised to use LBS data for all personal travel analytics.

Opportunities:

- Data fusion of emerging data with traditional sources improves the quality of data (e.g., observed counts and travel survey results). For example, bicycle flows collected via fitness-tracking apps can be validated by field counts or complemented by intercept surveys.
- The fusion of multiple data sets generates more comprehensive and reliable insights as different sources cover different types of travel activities, journey purposes, and spatial variations
- If a trip appears to have issues with metrics, such as speed, distance, etc it is flagged as "bad." Flagged trips and activities are not deleted from databases altogether, but they are filtered out from StreetLight queries and Metrics.

3.2 Crowdsource Data Used in this Study

For this study, mainly Strava data was considered of sufficient quality for comparison with the ground truth data, but first a preliminary investigation of Streetlight and Strava data was conducted at two sites to understand what the data looked like and the feasibility of working with it.

Preliminary investigation of Streetlight and Strava

Prior to the reduction in LBS data availability in 2022, a preliminary investigation of ground truth with Streetl7ight and Strava data was conducted for two sites. Average hourly zone bicycle traffic by month (StL Index) data produced by Streetlight was obtained for the months from June to September 2021. The average hourly zone bicycle traffic data is classified by weekdays/weekends. This data was used to calculate monthly average daily zone traffic, separately for weekdays and weekends. Further, using the monthly values of average daily zone traffic, the average across the months was calculated, which is the Streetlight Average of Monthly Average Daily Traffic (MADT) bike traffic values given in the table below.

The Average of Monthly Average Daily Traffic (MADT) for the two locations (Dexter at Fire Station and Kirby-Cass in Detroit) of bike trips are given below:

Table 3. Comparison of Strava, Streetlight and Counter data at Two Locations

Locations	Counter Location*	Strava#	Streetlight Weekday	Streetlight Weekend	Notes
Dexter at Fire Station	375	47	229	389	* Counter Location data is for the months of May to October in 2020, 2021. Also, it is a sum of counts in both directions. # Strava counts are a sum of traffic in both directions. ‘Street Light data is an indexed value. Monday to Thursday is considered ass the weekday value.
Kirby - Cass Site	296	22	255	307	

Based on this preliminary analysis of the two sites, the crowd-sourced data from Strava is much lower than Counter Location data, which is expected as the crowd-sourced data does not capture all the bike trips. Hence, they need to be supplemented by the counter data. Streetlight seems much more reliable as it is closer to the observed counter location counts, but this Streetlight metric is not available after April 2022.

3.3 Calculating MADT for Strava Data

For Strava, the daily bike trips count was obtained for the locations from Strava Metro for the months from June to September 2021. It was used to calculate the average daily bike traffic separately for all four (4) months. First, an average of the total counts for each day of the week in a month is calculated, which gives seven (7) values per month and the second average is the average of those seven (7) values together to get the Average Daily Traffic (ADT) by month. Using the ADT values, the Average of Monthly Average Daily Traffic (MADT) value was calculated, shown in the table below.

Strava ADT coded as zeros were left as zeros, and blank values were coded as zeros. This is because sites coded as blank have no Strava users, and Strava's 5 value represents 4 and higher, while 3 and lower is already rounded down to zero. Table 4 and Table 5 show examples of the data used for the comparison study for one site (Dequindre Cut at Gratiot in Detroit). This illustrates that multiple MADT values were compared with multiple Strava values. Both the ground truth counts and the Strava MADT had to be summed for both directions and if multiple parallel facilities were present, had to be summed for those too. For example, if there were bike lanes on both sides of a street and both were counted separately, the volumes from both of those bicycle lanes had to be summed.

Table 4. Pedestrian MADT for the Dequindre Cut @ Gratiot Counter

Site Number	Site Name	Year	Month	Count Average MADT	Strava MADT
7	Dequindre Cut @ Gratiot	2019	May	563	24
			June	728	28
			July	760	26
			August	766	37
			September	601	44
		2020	May	878	34
			June	1,163	40
			July	1,100	34
			August	1,152	44
			September	997	49
		2021	May	798	36
			June	737	35
			July	684	37
			August	629	36
			September	603	44

Table 5. Bicycle MADT for the Dequindre Cut @ Gratiot Counter

Site Number	Site Name	Year	Month	Count Average MADT	Strava MADT
7	Dequindre Cut @ Gratiot	2019	May	225	35
			June	377	50
			July	429	54
			August	406	49
			September	307	40
		2020	May	399	57
			June	548	76
			July	542	90
			August	536	90

Site Number	Site Name	Year	Month	Count Average MADT	Strava MADT
			September	416	76
		2021	May	430	65
			June	465	65
			July	464	67
			August	446	64
			September	386	60

The last table, Table 6, illustrates how ground truth combined pedestrian and bicycle volumes collected using infrared counters only, which don't separate bicyclists from pedestrians, were compared to Strava data, at the Drake path count site in Kalamazoo. Here the Strava walking and bicycling trip counts were summed for comparison with these combined counters. This may not be advisable, because there may be a larger number of bicyclists who use the app, or a larger number of walkers or runners who use the app, which might influence the Strava numbers, so it may not make sense to simply add these together. However, for lack of a better methodology, and for lack of sufficient count sites to study, the Strava data was summed up for comparison with these combined sites.

Table 6. Combined Pedestrian and Bicycle MADT for the Drake Counter

Site Number	Site Name	Year	Month	Count Average MADT	Strava MADT
26	Drake	2019	May	162	25
			June	188	27
			July	171	30
			August	187	30
			September	151	28
		2020	May	366	40
			June	357	45
			July	300	42
			August	314	40
			September	278	38
		2021	May	225	37
			June	261	37
			July	310	37
			August	304	32
			September	194	33

4.0 Relation Between Strava and Counter Data

In general, the Strava data range is well under the data range obtained from the counter data for all the sites. Figure 4 shows the histograms for MADT for bike counters and MADT from Strava for the same locations. Figure 5 shows the histograms for MADT for pedestrian counters and MADT for pedestrians using Strava data for those locations. Figure 6 shows the histograms for the bike and pedestrian combined counters. As is evident, especially for pedestrian only counter sites, Strava data significantly underestimates the count of pedestrians as compared to the counter data.

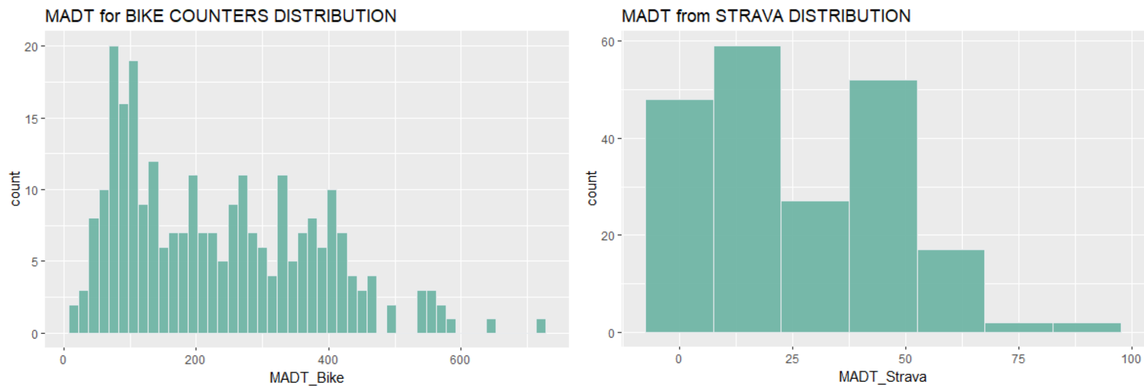


Figure 4. Histograms for MADT from Bike Counters and MADT from Strava for the same locations

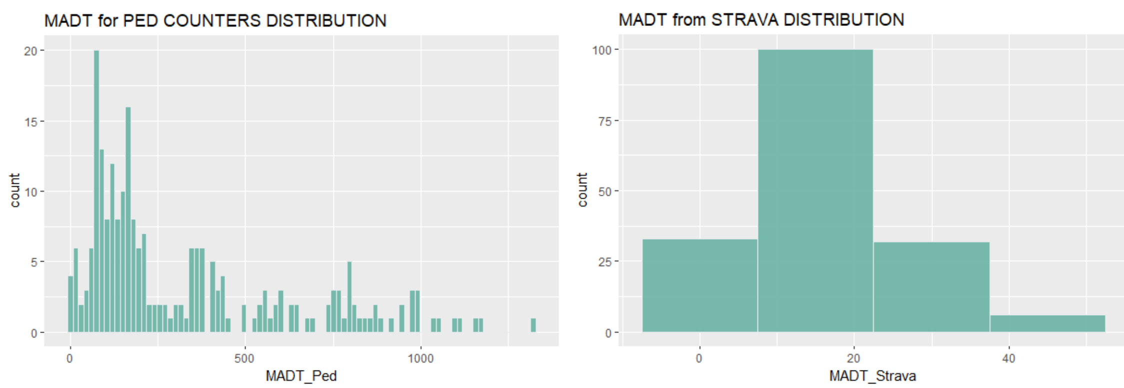


Figure 5. Histograms for MADT from Pedestrian Counters and MADT from Strava for the same locations

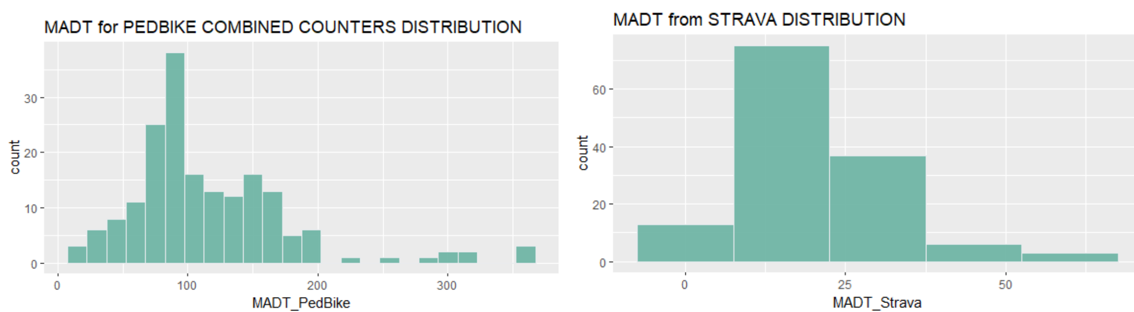


Figure 6. Histograms for MADT from Pedestrian Counters and MADT from Strava for the same locations

Figures 7-9 show MADT distributions by month for both the counter and the Strava data. In addition to underestimating the counts, Strava data are also more dispersed than counter data as is seen in these figures. These distributions also indicate a month-to-month temporal variation of MADT both for counter data and Strava data. However, seasonal variation is much more evident in bike volume data than pedestrian volume data.

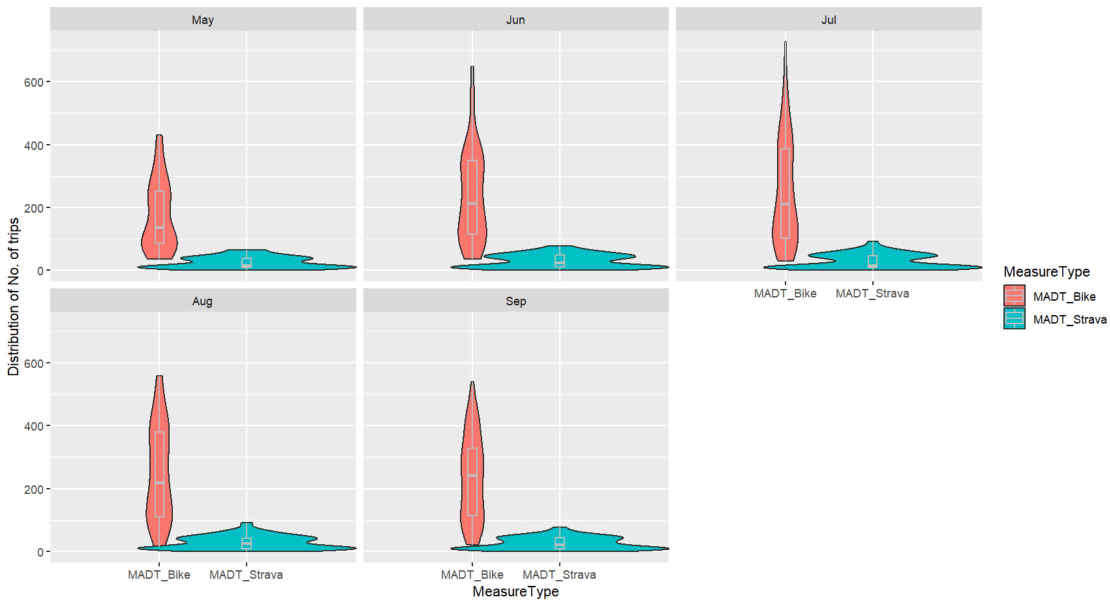


Figure 7. Distribution of Bike Volume from Counter and Strava Data across Months

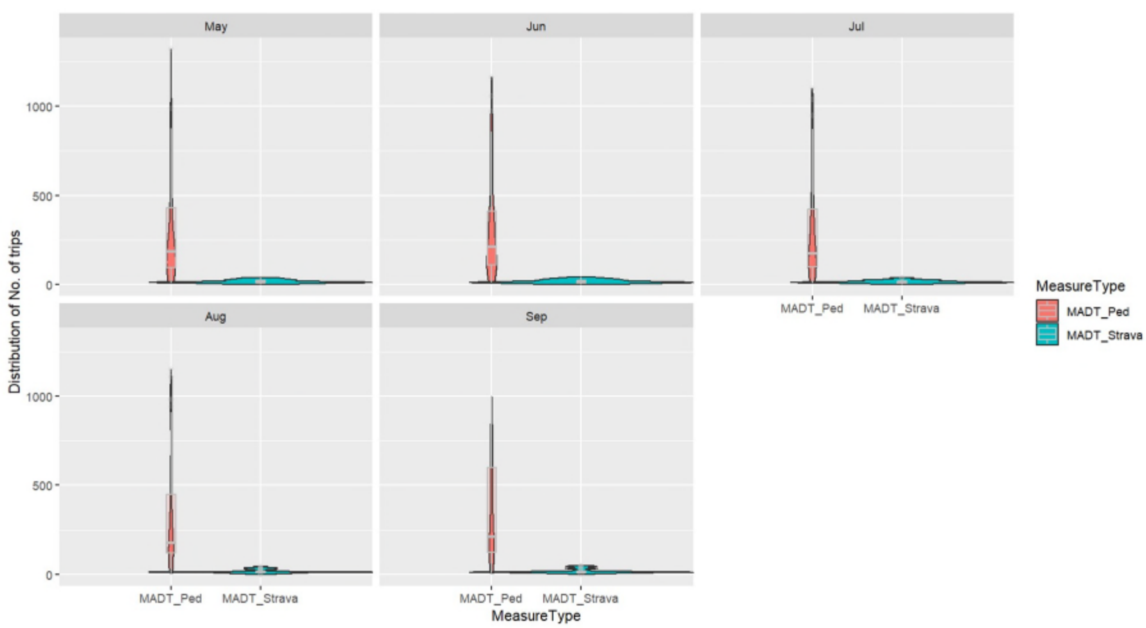


Figure 8. Distribution of Pedestrian Volume in Counter and Strava Data across Months

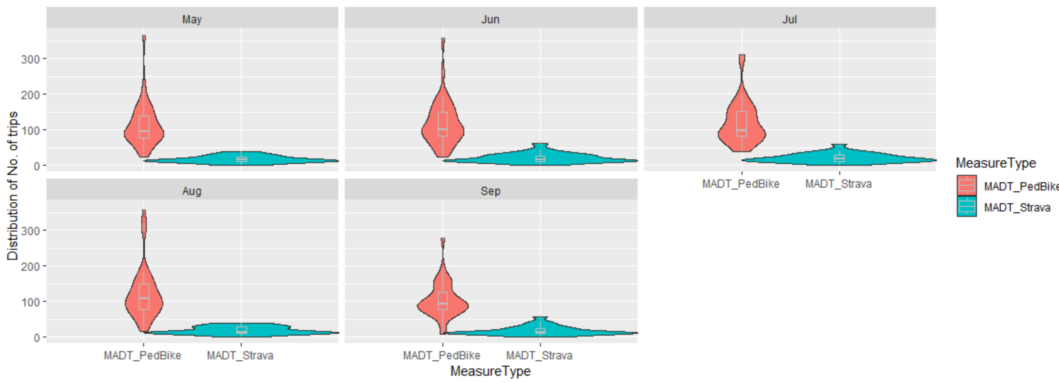


Figure 9. Distribution of Bike+Ped Volume in Counter and Strava Data across Months

Figures 10-11 show the relationship between MADT Strava data and MADT counter data for bike and pedestrian counters respectively. While a loess curve is a better fit for the relationship, linear regression models also provide reasonably good fits. The goodness of fit metric, R^2 values, are typically low for 2019 data for both bike and pedestrian volumes but improve significantly from 2020, ranging from 0.3 to 0.75. This may be an effect of the pandemic with more people taking up walking and bicycling during and after the pandemic as well as using Strava app for fitness tracking, resulting in better coordination between the two sources.

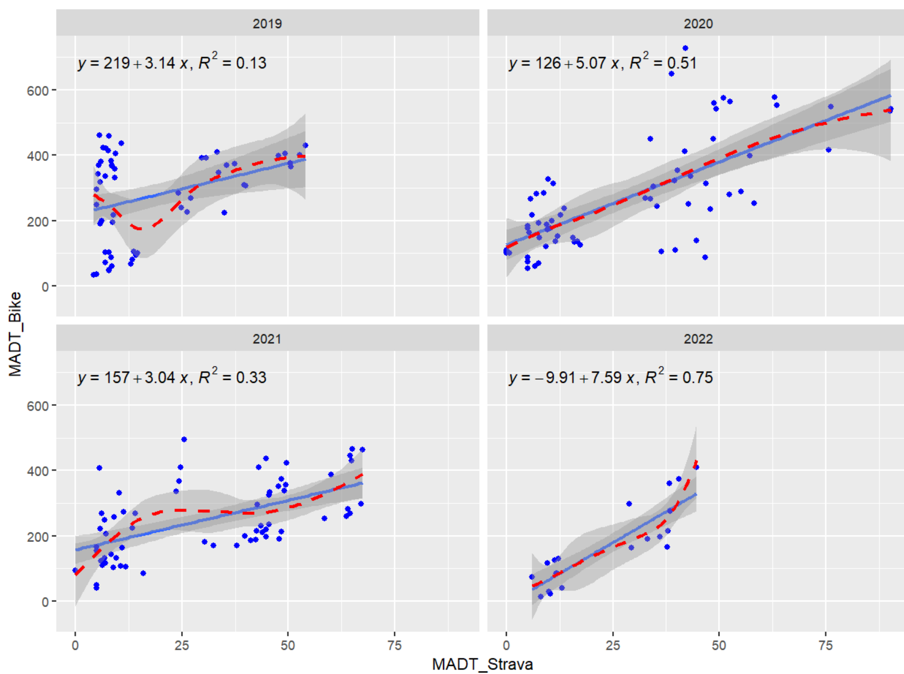


Figure 10. Relationship between Strava and Bike Counter MADT

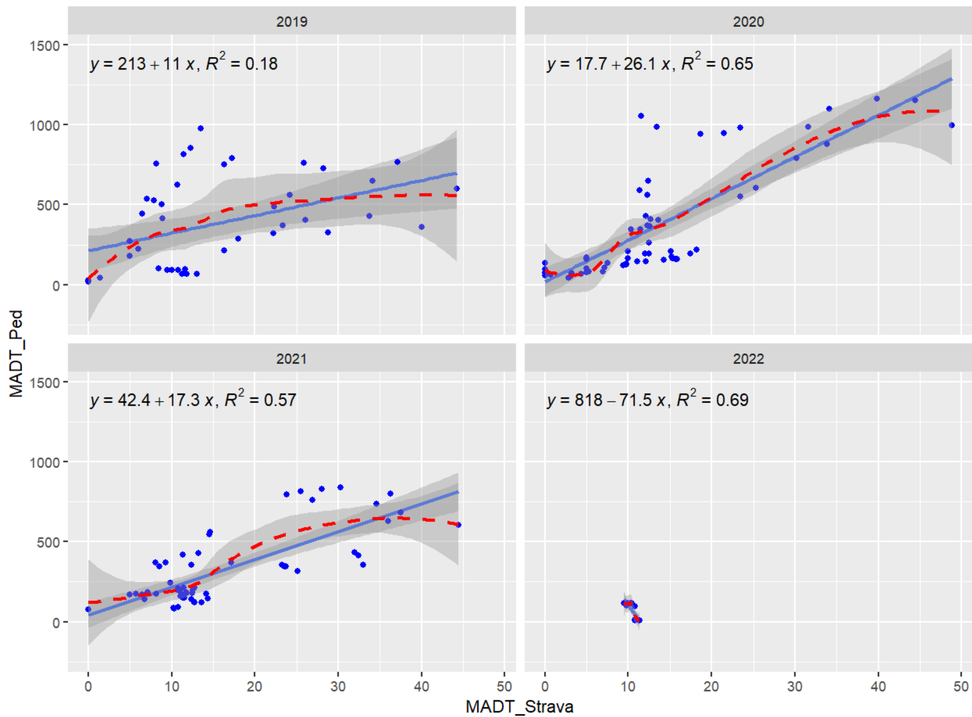


Figure 11. Relationship between Strava and Pedestrian Counter MADT

However, it should be noted that these modes are systemwide fits meaning that any volume predicted using these linear models will provide a reasonable estimate of bike and/ped volume across all of Michigan but might be quite different when any individual site is considered. Thus, the hypothesis is that to have better individual counter basis estimate there is need to explore spatial models that utilize locational attributes of the places. Hence, in the next step the spatial relationship between the data points is assessed using spatial autocorrelation fit statistics.

5.0 Spatial Relationships

Spatial autocorrelation is used to describe the extent to which a variable is correlated with itself through space. Spatial autocorrelation can be assessed using indices that summarize the degree to which similar observations tend to occur near each other over the study area. Positive spatial autocorrelation occurs when observations with similar values are closer together (i.e., clustered). Negative spatial autocorrelation occurs when observations with dissimilar values are closer together (i.e., dispersed).

Moran's I is a measure of spatial autocorrelation–how related the values of a variable (ped/bike volume) are based on the locations where they were measured. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random. For this project, there is a total of 385 observations (removing the missing volume observations) – each observation being one location per year and month. The results for Moran's I calculation are as follows:

MADT Counter

Moran I statistic standard deviate = 33.134, p-value < 2.2e-16

Moran I statistic Expectation Variance

0.7264628668 -0.0026246719 0.0004841842

MADT Strava

Moran I statistic standard deviate = 38.948, p-value < 2.2e-16

Moran I statistic Expectation Variance

0.8561610118 -0.0026246719 0.0004861937

Figure 12 shows the results of the analysis for the two data sets.

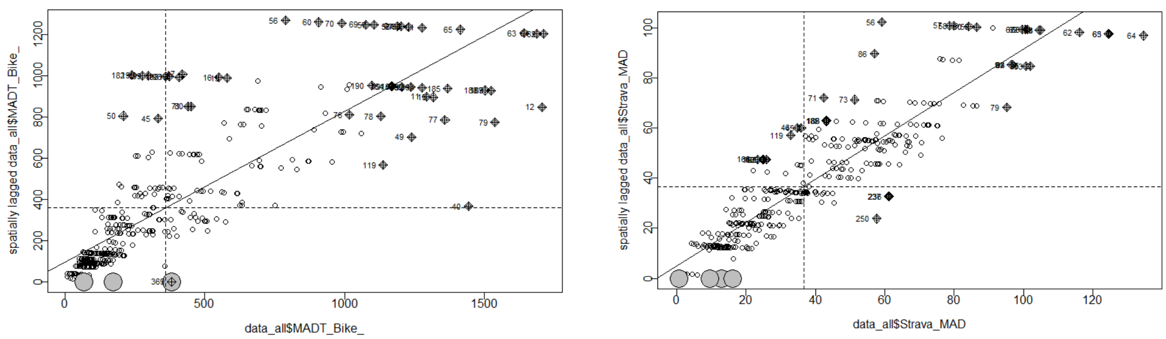


Figure 12. Moran's I statistic for MADT from Counter Data and MADT from Strava Data

The results indicate spatial autocorrelation between data points with Strava MADT having slightly higher Moran's I than counter MADT. However, since data points are for same location over time, a significant portion of the autocorrelation can be attributed to that aspect of the data. Hence, in the next step, location specific attributes will be added to develop causal and predictive models that could be more generalizable and transferable. Temporal variability of the data was also considered as there was data across multiple summer months but also across years for the same location.

6.0 Summary Findings

The findings from the survey of state agency personnel and the literature review presented in the previous chapters noted crowdsourced data quality to be a challenge, so in this chapter two popular crowdsourced data sources, Streetlight and Strava were compared to ground truth counter data from Michigan sites. In comparing Michigan specific counter data to crowdsourced data, similar data quality issues were observed, including lack of data because of privacy reasons where counts are low and overall underestimation of counts across corresponding counter locations. Although Streetlight data was found to be most effective and useful, Streetlight stopped providing non-motorized volume data at road segment level from 2022 due to lack of sufficient volume and spatial coverage. Thus, the only crowdsourced data available for comparing with count data was Strava Metro data. In an effort to find a relationship between counter data and Strava data, scatterplots were created to visualize the relationship pattern. While it was possible to fit a simple linear regression model between Strava and counter data for bike and pedestrian sites, the estimates from such models could only be comparable at a system level instead of individual counter level because of the number and range of outliers. Spatial autocorrelation models were tested but were found to be not effective because of the low sample size of counter locations.

Chapter 4. A Generalized Model for Using Crowdsourced Data in Estimating Bicycle and Pedestrian Volume

1.0 Introduction

One of the persistent problems in planning for bicycling and pedestrian facilities and safety is lack of consistent and good quality data on their volume. While counter data is often considered the best quality data available, installing counters is hindered by budget, understanding appropriate technology needed as well as maintenance needs. In addition, there is the problem of incompatible data from counters that count bicyclists and pedestrians separately as opposed to counters that collect data on combined bicycle and pedestrian volume. Some attempts have been made at predicting bicycling and pedestrian volumes using modified travel demand models as it fits the constraints of bicycling and pedestrian travel (e.g., smaller analysis zones than standard traffic analysis zones TAZs). Noteworthy among them are Singleton et. al. and Hampshire et. al. of which Hampshire et. al. is a model for Michigan and based off Michigan travel diary data. However, the nature of high-resolution data needed for good estimation power and the high computational burden associated with these models make them difficult to use.

In recent times, crowdsourced data has become one of the alternative data sources for estimating bicycling and pedestrian volumes, especially as use of GPS enabled fitness applications (apps) proliferated. Crowdsourced data often results in both non-standardized and large datasets and can be biased because of including data from people who are either users of a particular app or are of a particular sociodemographic group who have access to smartphones and are able to use it. Thus, it is not appropriate to use these data directly as a substitute for counter data when planning or estimating volumes.

The purpose of this task of the project is to examine crowdsourced data and its applicability for volume estimation through a generalized model that can be used for the entire state of Michigan. The first set of models examine the adjustment factor (k-factor) method, borrowed from the gravity model of traditional 4 step travel demand models (McNally 2008). The purpose of these models is to understand if there is a standard adjustment factor between crowdsourced data and counter data that is generalizable across facility types or locations. In this case, because of the low sample size of counters and lack of coverage across all roadway/facility types, standardization of the factor can be deemed unreliable and so, a model is developed to understand contextual variables commonly contributing to these factors. The goal of this set of models is to be spreadsheet implementable – if data on the contextual variable are available along with Strava data, the volume of bicyclists and pedestrians can be predicted within a certain range of accuracy just by inputting the values of the contextual variables.

Next, two sets of more statistically sophisticated models are developed for each mode type: bicyclist only counters, pedestrian only counters and bicyclist and pedestrian combined counters (a total of six models). The first set of models relate bicycling/pedestrians/bicycling+pedestrian volume from counter data to crowdsourced data using location specific attributes. The purpose of this set of models is to identify factors that contribute to the difference between counter data and crowdsourced data, while correcting for panel data structure (multi-year data for same location) and location specific unobserved heterogeneity. The final set of models (Appendix C) is to estimate volume using only the location specific attribute data – these set of models were developed for situations when crowdsourced data may not be an available or viable option, as seen with Streetlight data.

2.0 Models Relating Strava Data and Counter Data

2.1 Data

Crowdsourced Data

For crowdsourced data, although the intention has been to use multiple data sources, based on data availability and quality, Strava (Strava.com) has been found to be the most appropriate source because of its high user volume, a track record of being used by multiple cities for understanding their bicycling and walking patterns and for the data being offered free to planning and government organizations like DOTs. While originally an app for tracking bicycling, Strava recently has added functionality to add running/walking which can be somewhat a proxy for pedestrians. However, Strava data is biased toward people who use the Strava app. Strava users are disproportionately young adults (25–35 years in age) and male (Roy et al., 2019). Women, children, older adults, and low-income bicyclists are under sampled by Strava data.

Location Specific Data

To account for some of the many factors that may impact pedestrian and bicycle traffic and the fit of the Strava data to the ground truth count data, about 166 additional data elements were considered in the categories listed below. These were based on variables used in previous bicycle data fusion studies (Broach et al, 2023).

- Monthly Weather data (Temperature, humidity, and precipitation)
- Facility Type (bike lane/bike path etc.)
- Sociodemographics
  - Median Household Income
  - Age
  - Population Density
  - Employment Density
- Distance to central business district
- Proximity to university/college and other destinations
- Distance to edge of park and water bodies as well as area of park and water bodies

Appendix D shows the details of the data elements and how they were calculated. The analysis was done using Python and the code can be made available upon request.

2.2 Adjustment-factor Models

Calibration factors or calibration constants are common in models across multiple disciplines. Calibration factors are used to map real world measured values to ground truth values and are generally multiplicative in nature. In transportation, the two most popular uses of such factors are the use of k factor in the trip distribution step of the four-step travel demand model and in determining design hour traffic volume on a segment from annual average daily traffic volume (AADT). In fact, seasonal and time of day adjustment factors are also in some sense calibration factors, capturing the seasonality effect of travel that may be different than average traffic volume.

Sometimes, especially in experimental sciences, the calibration factors are just ratios of observed value to actual expected value, measured multiple times over time, under similar conditions and hence, almost of a constant value. In naturalistic settings, however, the adjustment factors almost always represent unobserved factors related either to contextual elements or behavior. For example, the k-factor in trip distribution is assumed to account for socioeconomic factors associated with trip distribution which are not part of the gravity model of trip distribution itself but has the potential to affect people's trip making patterns.

The adjustment factor models are inspired from both experimental setting and the k-factor concept traditionally used in the gravity model in the trip distribution step of travel demand models. In the trip distribution step of the travel demand models, as the name suggests, trips are distributed between origins and destinations, where destinations are assigned based on their attractiveness for travel and impedance factor. Of the different models used to complete this step, the gravity model is the most popular one, modeled after Newton's law of gravity. The gravity model is represented as:

$$T_i j = P_i \; [(A_j \; F_i j \; K_i j)/(\textstyle\sum_l \; A_j \; F_i j \; K_i j \;)] \qquad \text{-------------------------(2)}$$

where:
T_{ij} = number of trips that are produced in zone i and attracted to zone j
P_i = total number of trips produced in zone i
A_j = number of trips attracted to zone j
F_{ij} = a value which is an inverse function of travel time
K_{ij} = socioeconomic adjustment factor for interchange ij

For this study, the formula was simplified as:

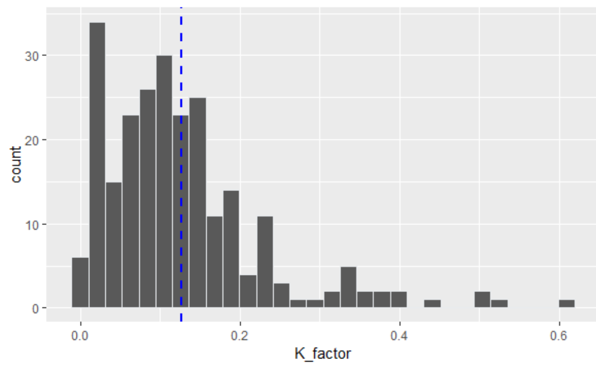
$$MADT_{counter\ i} = \left(\frac{1}{k_i}\right) \times MADT_{Strava\ i} \qquad \text{-----------------------------(3)}$$

such that

$$k_i = {}^{MADT_{Strava\ i}}/_{MADT_{counter\ i}} \qquad \text{---(4)}$$

where:
$MADT_{counter\ i}$ = monthly average daily bicycle/pedestrian count at counter location i
$MADT_{Strava\ i}$ = monthly average daily bicycle/pedestrian volume from Strava data at location i
k_i = the socioeconomic adjustment factor for location ***i,*** named active expansion factor (AEF) hereafter
$\overline{K}$ = mean active expansion factor calculated by taking average of the active expansion factors (k_i) across the state. Given that Strava is representative of a subgroup of the bicycling and walking population, it is hypothesized that K <=1.

Figures 13-14 shows the distribution of the k_i for bicyclists and pedestrians across all the locations, Figures 15-16 show the distribution of AEF across months for bicyclists and pedestrians and Figures 17-18 show the distribution of AEF for bicyclists and pedestrians across different locations. The x-axis is the value of the AEF while the y-axis represents the frequency or the number of observations having an AEF within a particular bin. The blue dotted line represents the mean AEF across all observations. For the pedestrian data, some of the AEFs were above 1, meaning Strava data showed more pedestrians than the counter data. These were considered outliers and removed from further analysis. The AEF ranges from 0 to 0.6 for bicycles and from 0 to 0.77 for pedestrians, after excluding the extreme outliers. The mean AEF for bicycling is 0.12 while the median is 0.11. The same for pedestrians is 0.06 (mean) and 0.05 (median) respectively.



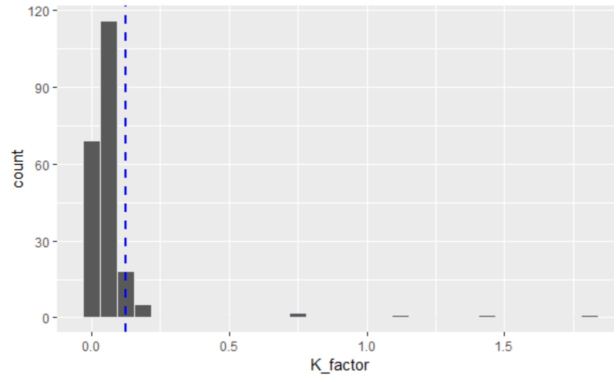


Figure 13. AEF **distribution for Bikes**

Figure 14. AEF **distribution for Pedestrians**

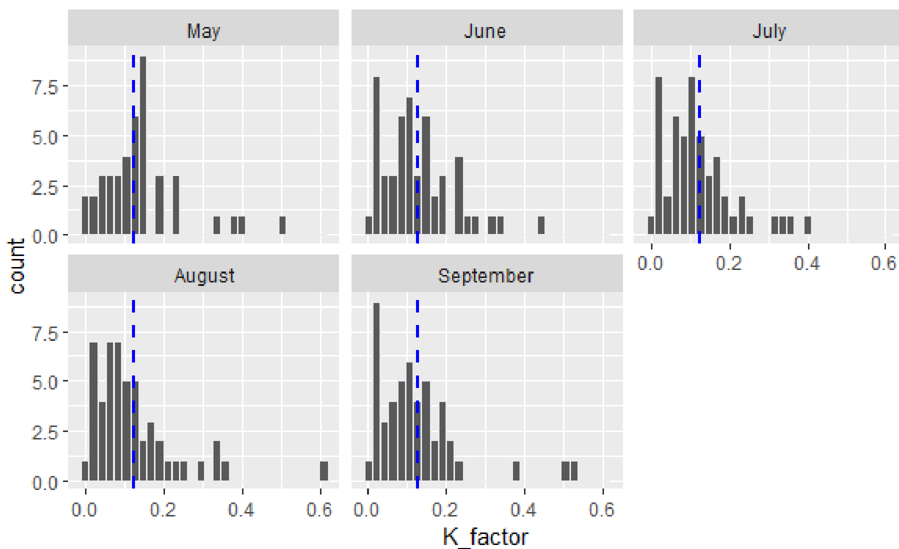


Figure 15. **Distribution of Bike AEF across months**

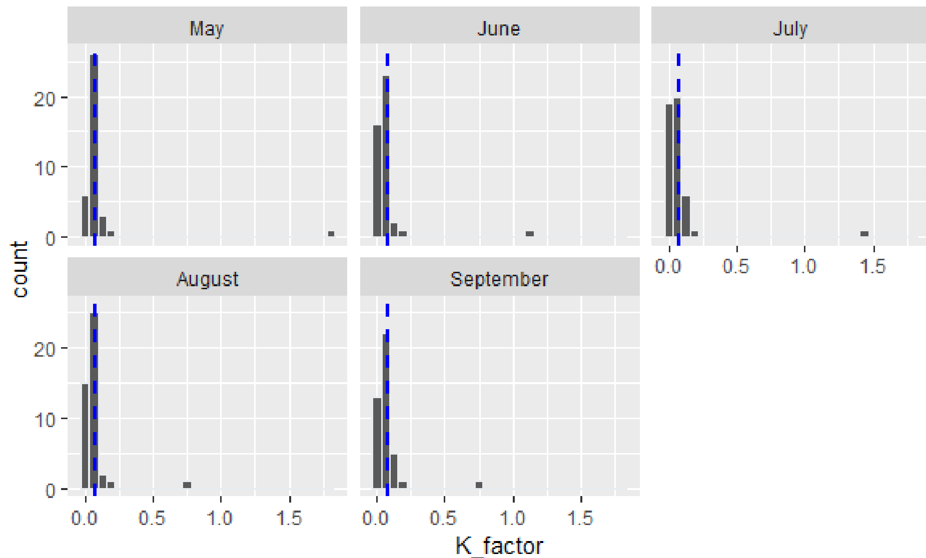


Figure 16. Distribution of Pedestrian AEF **across months**

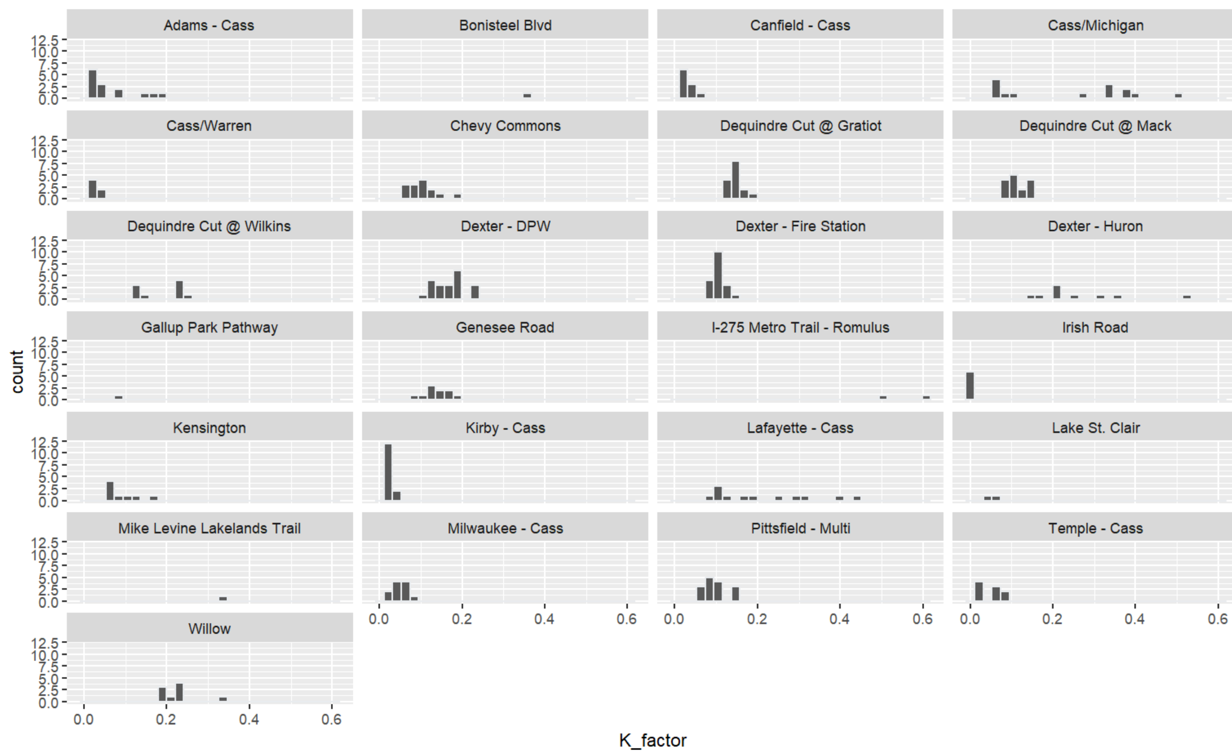


Figure 17. Distribution of Bike AEF **across locations**

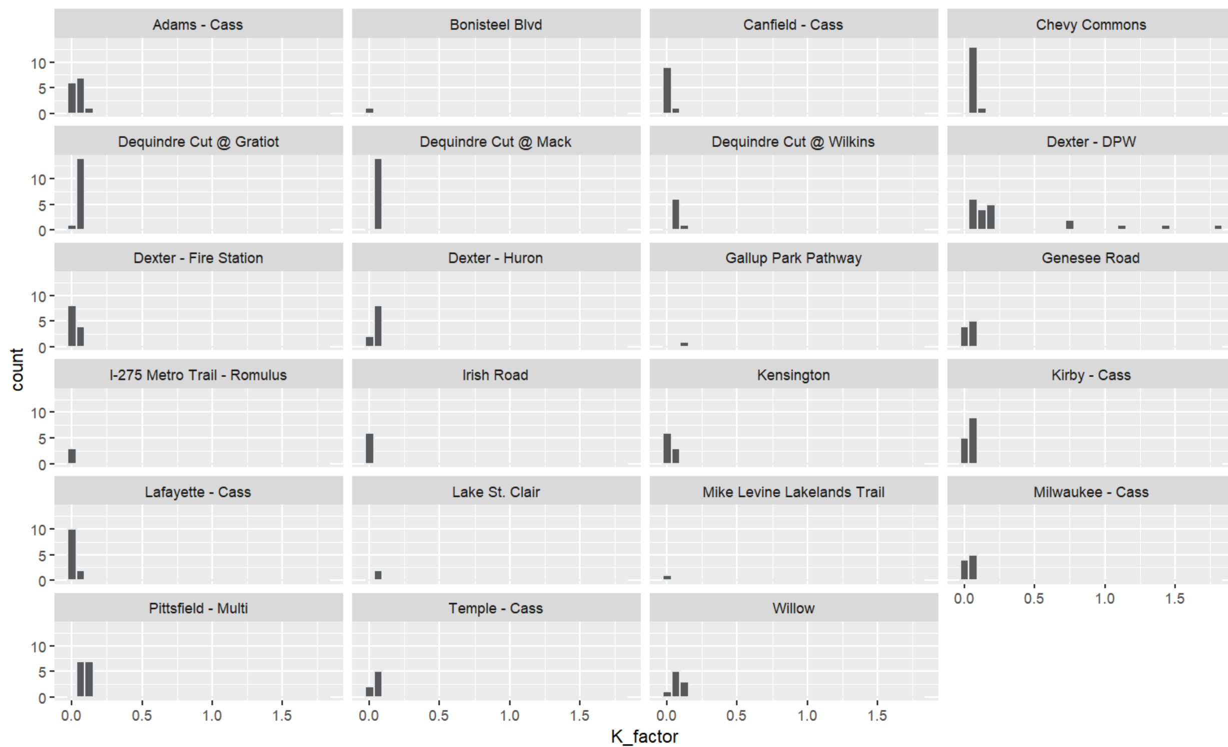


Figure 18. Distribution of Pedestrian AEF **across locations**

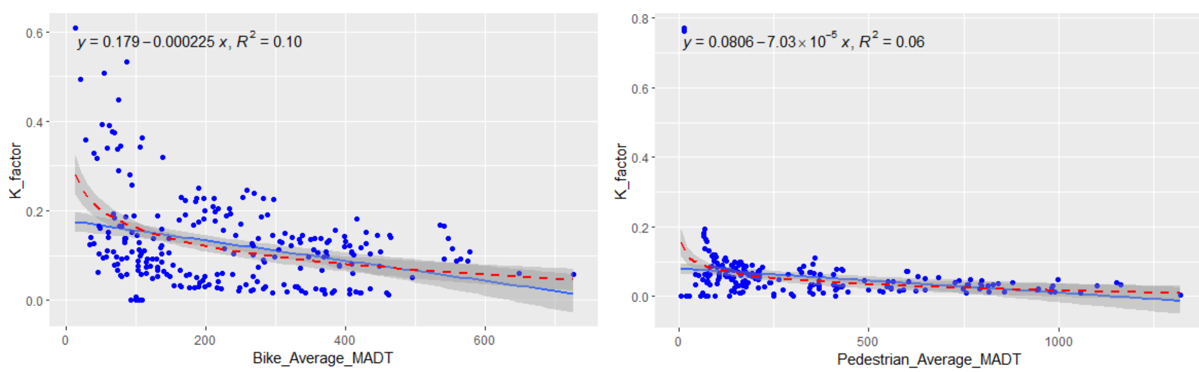


Figure 19. Relationship between AEF **and Bike and Ped MADT**

Figure 19 shows the relationship between MADT and the AEF for bicyclists and pedestrians. As is seen from the figures, for bicycles, there is a non-linear relationship (log) with low fit statistics while there is almost no relationship between the AEF and the pedestrian volume.

Modeling Individual K factors

Given the findings from the previous section that the bike and pedestrian MADT is very loosely related to AEFs the next step was to regress AEFs with census block level socioeconomic variables, excluding bike and pedestrian MADTs. The advantage of modeling AEF instead of counts is that any variability related to a particular location is accounted for in the ratio itself, so there is no need to include location specific stochastic or random effect. Further, since there is very little variability in monthly averages as seen in the previous section, month was not considered as a factor in these models. However, since AEFs are fractions, a special class of regressions called fractional regression models are used where the outcome variable ranges between 0 and 1. The glm function within R with the quasibinomial family, link = logit is used for estimating these models. The variables for the models are selected based on significance level and model fit or change in model fit. Table 7 shows the bicycle AEF model while Table 8 shows the pedestrian AEF model.

Table 7. Model Estimate Results for Bicycle AEF

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.92	0.34	-8.519	2.02E-15	***
Distance to nearest park center	1.66	0.21	8.031	4.77E-14	***
Distance to nearest grass area center	-0.18	0.05	-3.199	0.001571	**
Distance to nearest university	0.28	0.05	5.6	6.00E-08	***
Distance to nearest college	-0.19	0.04	-4.567	8.00E-06	***
Distance to CBD	-0.09	0.03	-3.517	0.000524	***
Residential Roadmiles	-0.01	0.00	-5.602	5.96E-08	***
Household density per acre	-0.44	0.07	-6.024	6.58E-09	***
Median household income/$10k	0.07	0.02	3.333	0.000998	***
Percentage Bike commuters	0.09	0.03	2.767	0.006112	**
Number of bus stops	0.01	0.00	2.495	0.013282	*
Percentage Education level at least college	0.01	0.00	2.161	0.031741	*
Signif. codes	: 0 ‘***’	0.001 ‘**’	0.01 ‘*	’ 0.05 ‘.	’ 0.1 ‘ ’ 1

While interpreting the results, it should be noted that a positive coefficient for a variable indicates an increasing effect of AEF i.e., the Strava data needs to be divided by a larger adjustment factor as those variables increase. For example, as the distance from park increases, Strava bicycle volumes need to be adjusted downward while as the distance from grass area increases, Strava bicycle volumes need to be adjusted upward to map to counter volume. In other words, away from park areas Strava over predicts the volume while away from grass areas Strava under predicts the volume. This can be due to the bicycling pattern of Strava users who are likely to be using trails and paths, while pedestrians are more likely to frequent parks and recreational areas. Reflecting the same user riding pattern, Strava data systematically under reports volume as the distance from CBD increases, in areas with higher residential roadmiles and high household density. Strava better represents riders in areas with a high median income, higher percentage of bike commuters, higher percentage of people with college education and higher number of bus stops. It should also be noted that the linear version of the model gives an R2 ~ 0.5 indicating these variables can predict about 50% of the AEF and there is still need for calibration of the AEF using more data from different types of facilities and locations.

Table 8. Model Estimate Results for Pedestrian AEF

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.27	0.80	-0.341	0.733271	
Distance to nearest grass area center	-0.30	0.09	-3.304	0.001127	**
Distance to nearest university	-0.03	0.02	-1.338	0.182371	
Median household income/$10k	0.10	0.02	6.232	2.66E-09	***
Percentage Bike commuters	-0.09	0.05	-1.944	0.053294	.
Number of bus stops	-0.01	0.00	-3.77	0.000215	***
Median Age	0.02	0.01	1.488	0.138223	
Percentage of White population	-0.04	0.01	-3.931	0.000116	***
Population Density in Acres	-0.19	0.06	-3.079	0.002371	**
Signif. codes:	0 '***'	0.001 '**'	0.01 '*	' 0.05 '.	' 0.1 ' ' 1

From the pedestrian AEF models, Strava data systematically needs higher adjustment factor as distance from grass area increases as does distance from university, in areas of high population density, high percentage of White population, higher percentage of bike commuters and higher number of bus stops. It needs lower adjustment factor for high median income and higher population median age. Given that Strava app is mostly used for running rather than walking (non-exercise), the conclusions mostly are in line with expectations. It should also be noted that the linear version of the model has R2 of only 0.31 indicating that this model can explain only about 30% of the variability in the adjustment factors.

Model Prediction Validation

- The data was split in the ratio of 75:25 (train:test) and models were developed based on 75% of the data.
- Validation was a problem because with some Strava volume being zero, AEF for those locations were zero
- The validation was done for predicted MADT by using the calculated AEF from the models
- In predicting the volume for the 25% test data, error percentage ranged from -19% to 131% for bicycles
- Error percentage ranged from 100% to 160% for pedestrians

2.3 Generalized count models

Feature Selection using Random Forest

In data-driven problems, high-dimensional data, especially in terms of many features, is prevalent these days. Many researchers explore the issues with high-dimensional data with numerous independent variables by extracting important features from these high-dimensional data (Wang et. al. 2019). There are several reasons as to why feature selection is important in high-dimensional data analysis including those (1) to spare the model to decrease the number of parameters, (2) to decrease the model development time, (3) to lower the issue with overfitting by increasing generalization, and (4) to "avoid the curse of dimensionality" (Chen et. al. 2020).
Random Forest is a supervised ensemble machine learning method that is used for classification and regression. It uses decision trees, using a bagging algorithm and feature randomness, for training purposes and for classification problem, the output of the random forest is the class selected by most trees. Decision trees are very sensitive to the data they are trained on — small changes to the training set can result in significantly different tree structures. Random forest takes advantage of this by allowing each individual tree to randomly sample from the dataset with replacement, resulting in different trees. This phenomenon is called bagging. To treat the bias-variance trade-off, multiple decision trees are trained separately, and the output of the random forest is obtained as an average of the outputs of individual decision trees (Chakraborty et. al. 2023). In extant research, Random Forest has been widely used in several domains for feature selection and has been considered as a robust technique to identify the most important variables in the data based on increasing node purity (Khoshgaftar 2007, Li et. al. 2012). Particularly, each tree of the random forest can calculate the importance of a feature according to its ability to increase the pureness of the leaves. The higher the increment in leaves purity, the higher the importance of the feature. This is done for each tree, then is averaged among all the trees, thereby each predictor having a separate variable of importance for each class. In this study, the feature selection by Random Forest was accomplished by using "randomForest" package along with the varImp()function on R (version 4.1.2) statistical tool. Appendix D shows the node purity measure for each of the variables.

Feature Selection Based on Correlation Coefficient

After all independent variables were ranked by Random Forest in the order of increasing node purity, we generated the correlation coefficient matrix in order to understand the relationships individual variables have with one another. The correlation coefficient is a statistical measure of the strength of a linear relationship between two variables. A correlation coefficient is a number between -1 and 1 that provides us with the strength and direction of a relationship between variables. Particularly, correlation coefficient of ±1 indicates the strongest possible correlation and 0 indicates no correlation.

Mixed-Effects Negative Binomial Models

Traditional linear regression techniques are generally inappropriate for count data which do not follow the assumptions of a normal distribution. As an alternative, the Poisson distribution provides a starting point for the analyses. In the context of this analysis, in Poisson model, the probability of counter location *i* experiencing y_i pedestrian/bike volumes in one year can be expressed as

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i{}^{y_i}}{y_i!} \qquad \text{--------------------- (1)}$$

where P(y_i) is the probability of counter location *i* experiencing y_i pedestrian/bike volumes, and λ_i is the Poisson parameter or the expected number of pedestrian/bike volumes for location *i*, E[y_i]. The Poisson regression model relates the expected number of volumes at a location, λ_i , to a function of explanatory variables, expressed as:

$$\lambda_i = exp(\beta X_i) \qquad \text{--------------------------(2)}$$

where X_i is a vector of explanatory variables and β is a vector of estimable parameters. A limitation with Poisson distribution is the assumption that the mean and variance are equal, which is not the case with our data in hand. In this data, variance exceeds mean, leading to an overdispersion. The negative binomial model addresses this overdispersion by adding an unobserved heterogeneity term as,

$$\lambda_i = exp(\beta X_i + \varepsilon_i) \qquad \text{----------------------------(3)}$$

where $exp(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α. The inclusion of this term essentially allows the variance to differ from mean as

$$VAR[y_i] = E[y_i] + \alpha E[y_i]^2 \qquad \text{----------------------------(4)}$$

This α is termed as the overdispersion parameter. In the transportation safety analysis, negative binomial regression models have been widely used (*6*–*8*) and accepted as the current practice for modelling count data, as such models account for overdispersion.
Recently, mixed-effects negative binomial models have gained popularity due to their capability of accounting for spatial effects and heterogeneity across observations (9). Unobserved heterogeneity can be defined as unknown variability in the effect of variables across the sample population. It is imperative to address this issue of unobserved heterogeneity to avoid erroneous predictions resulting from the biased estimated parameters (*10*). The issue with non-random sampling and unobserved heterogeneity in the data is addressed by including a combination of location- (i.e., location ID) and year-specific random effects (intercepts) in the negative binomial models, effectively developing mixed-effects models. In a mixed-effects model, each intercept is drawn at random from the intercept distribution and is independent of the error term for any particular observation and uncorrelated with the independent variables. The regression analyses in this study were conducted using R statistical software version 4.1.2.

Results and Discussion of the Regression Models with Strava Volume as the Independent Variable and Counter Volume as the Dependent Variable

In order to better understand what features most impact MADT when Strava volumes are included, we ran another set of models including Strava variables and other features as suggested by random forest and the correlation matrices previously presented. Table 9 shows results of the negative binomial model of bicycle MADT with Strava bicyclist volume and other selected variables.

Table 9. Bicycle MADT Model Results with Strava Bicyclist Volume

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	4.0212	0.5600	7.1800	<0.001
Strava bicyclist volume	0.0152	0.0024	6.3990	<0.001
Presence of bikeway	0.6838	0.4064	1.6830	0.0924
Presence of bus stops	-0.0167	0.0067	-2.4790	0.0132
Number of lanes	0.3268	0.1649	1.9820	0.0475
Grass area	0.0145	0.0049	2.9330	0.0034

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
Population density	0.7612	0.2037	3.7360	<0.001
Employment density	-0.6688	0.2251	-2.9710	0.0030
Humidity	-0.0154	0.0050	-3.0940	0.0020
May	Baseline			
June	0.2316	0.0526	4.4030	<0.001
July	0.3249	0.0560	5.8010	<0.001
August	0.3207	0.0569	5.6330	<0.001
September	0.3482	0.0690	5.0460	<0.001
Overdispersion	0.0525			
Random Intercept	**Variance**	**Std. Dev.**		
Location	0.2466	0.4966		
Year	0.0192	0.1386		

Strava volume is positively associated with counter volume meaning, the higher the Strava volume, the higher the counter volume. And Strava volume was identified as the most important variable also by the Random Forest algorithm. Among other variables, while presence of bikeway and number of lanes were positively associated with higher counter volume, presence of bus stops was negatively associated with the counter volume. Among other environmental and sociodemographic factors, while counter volumes show a positive association with nearby grass area, expectedly, humidity is negatively associated with counter volume. Additionally, while population density is positively associated with counter volume, employment density shows a negative association with counter bike volumes. Lastly, keeping the month of May as the baseline, all other summer months show a higher counter volume, with July and August being somewhat comparable, and the highest volume in the month of September. The variance and standard deviations of the random parameters confirms that there are an underlying heterogeneity among years and within sites.

Similarly, Table 10 shows the results of the negative binomial model of pedestrian MADT with Strava pedestrian volume and other selected variables. Strava pedestrian volume shows a positive association with the counter volume. It was somewhat counter-intuitive that the distance to paths was positively associated with counter volume. Unlike the bicycle model, the presence of bus stops is positively associated with the counter pedestrian volume. Additionally, white population and household density were both positively associated with counter volume, while both median age and humidity were negatively associated with counter pedestrian volume. Note that unlike the bicycle-only model, the month factor was not statistically significant in the pedestrian-only model. Looking at the random parameters, it can be said that the variability in both these factors were higher for pedestrian volumes compared to the bike volumes.

Table 10. Pedestrian MADT Model Results with Strava Pedestrian Volume

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	3.3401	1.0034	3.3290	<0.001
Strava pedestrian volume	0.0350	0.0098	3.5900	<0.001
Distance to paths	0.1998	0.1145	1.7450	0.0810
Presence of bus stops	0.0172	0.0052	3.3130	<0.001
Ethnicity - White	0.0345	0.0095	3.6360	<0.001

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
Household density	0.4085	0.1226	3.3320	<0.001
Median age	-0.0395	0.0155	-2.5430	0.0110
Humidity	-0.0118	0.0071	-1.6690	0.0951
Overdispersion	0.2132			
Random Intercept	**Variance**	**Std. Dev.**		
Location	0.3567	0.5972		
Year	0.1617	0.4021		

The final model to be discussed is the bicyclist and pedestrian volume combined model. It models bicycle and pedestrian combined MADT at both sites where pedestrians and bicyclists are counted separately at a given site and where they are counted together and not differentiated from each other. Table 11 shows the results of the negative binomial model of these combined bicycles and pedestrian MADT values with Strava bicyclist-plus-pedestrian volumes.

Table 11. Bicyclist and Pedestrian MADT Model Results with Strava Bike-Ped Volumes

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	3.2341	0.3592	9.0030	<0.001
Strava bicyclist-plus-pedestrian volume	0.0182	0.0017	10.5220	<0.001
Presence of bikeway lane	0.4486	0.2667	1.6820	0.0925
Presence of bus stops	-0.0134	0.0071	-1.9040	0.0569
Distance to footways	0.0736	0.0192	3.8440	<0.001
Intersection density	-0.1271	0.0701	-1.8140	0.0697
Park area	0.0023	0.0007	3.3670	<0.001
Population density	0.3694	0.1840	2.0080	0.0447
Employment density	-0.3473	0.1956	-1.7760	0.0758
Ethnicity - White	0.0104	0.0036	2.8850	0.0039
Population percent bike commuting	0.0915	0.0340	2.6880	0.0072
Precipitation	-0.6054	0.2933	-2.0640	0.0390
Overdispersion	0.1006			
Random Intercept	**Variance**	**Std. Dev.**		
Location	0.1933	0.4396		
Year	0.0242	0.1557		

In the bicycle-pedestrian combined model, as in the case with other models, Strava volume was positively associated with counter volume. Also, both presence of bikeways and distance to footways were positively associated with the counter volume, while presence of bus stops, and intersection density were negatively associated with counter volume. Additionally, park area, population density, and percent of population using bike were shown to have positive associations with counter volume, while employment density, and precipitation were found to have negative associations with counter volume. Lastly, similar to the pedestrian-only model, white population is shown to be positively associated with counter volume. Looking at the random parameters, it can be seen that, while the variability among the locations were lowest in the ped-bike combined model, the year-specific variability in this model was higher than the bike-only model but lower than the ped-only model.

Chapter 5 Conclusion

Crowdsourced data can be a great resource to fill in the gap between what data are collected through count stations and manual counts and what needs to be known for planning and road safety analysis purposes. However, crowdsourced data has its own challenges especially in areas of data quality and generalizability. Very little effort has been put towards validating the crowdsourced data sources using traditional count station data and the research that has been done, is fraught with contradictory findings – for example, Watkins et al. (2016) finds Strava data to be not representative of all types of cyclists (commuters, recreational, seasonal, experienced, new etc.) while Jestico et al. (2016) finds the same data source to be representative of urban cyclists. Similarly, efforts in fusing these emerging data sources with traditional count data are also few and far between. Only recently a few research teams have started to look into mapping crowdsourced data to count data. The aim of this project thus, was to ***develop a framework for identifying and utilizing the best resource available in crowdsourced data for active transportation through a generalizable model mapping crowdsourced data to count data and related contextual features that influence bicycling and walking.*** The scope of the project is extremely timely and needed for the broader research and practitioner community, in addition to addressing the lack of data on bicycling and walking within Michigan.

Towards fulfilling the aim of the project, first a detailed literature search was done to understand the state of the practice. Then an online survey was conducted for practitioners and DOT personnel actively involved in bicyclist and/or pedestrian planning and safety across the country. The survey was intended to understand what crowdsourced data were being used (if any) and the end user experience with different crowdsourced data sources. The majority of the respondents reported being familiar with crowdsourced data but fewer actually used that data because of lack of trust on quality, coverage and sometimes cost of these data sources. Most popular and well-known data sources were Streetlight and Strava and almost no crowdsourced data were available or used for pedestrian volume or risk assessment. In the next step, Michigan specific data was obtained from both vendors for test sites for initial data comparison. Streetlight data was found to be more effective and useful, however, when the counter data from Michigan were being processed for all the counter sites, Streetlight had stopped sharing segment level data and hence, Streetlight could not be used for mapping the crowdsourced data to counter data. Strava data was chosen as an alternative given it was available to planning agencies and DOTs free of cost and its popularity among the practitioners and bicycling community. Strava also started providing pedestrian data around the same time when they were chosen as the crowdsourced data source to use for mapping, which provided somewhat of a proxy for mapping pedestrian volume.

A set of generalized models were then developed using Strava data that could be used for the entire state of Michigan. The first set of models were based on adjustment factor method, borrowed from the gravity model of traditional 4-step travel demand models, to understand if there was a standard calibration factor between crowdsourced data and counter data. While the adjustment factors, named active expansion factors (AEFs) in this project were reasonably stable temporally, because of the low number of counters (~40) used to derive these factors and the lack of variability in the type of facilities covered by these counters, using the adjustment factors single-handedly to map Strava data to counter data could be deemed unreliable. Contextual variables were chosen based on literature to explain the variation of the AEFs at different locations and modeled accordingly. These models were less resource intensive, with the purpose of being implementable on a spreadsheet, but provided prediction accuracy of bicycling volumes with ranging between -19% to 130% while pedestrian volumes could be predicted with accuracy ranging between 100% to 160% of the counter data. Next, two sets of models, one using Strava data and another without Strava data, were developed for each mode type: bicyclist only counters, pedestrian only counters and bicyclist and pedestrian combined counters (a total of six models). To account for some of the many factors that may impact pedestrian and bicycle traffic and the fit of the Strava data to the ground truth count data, about 166 additional data elements were considered based on existing literature. Because of the panel structure of the data both spatially and temporally, mixed effect models were used for these six models. Results indicate that Strava data was strongly correlated with counter data for bicycling volume and aggregated bicycling and pedestrian volume but had low correlation with pedestrian volume. Land use and population variables were useful in improving the prediction of non-motorized traffic volume in conjunction with Strava data.

In conclusion, while this research, like similar other contemporary research on usefulness of crowdsourced data, could not recommend use of crowdsourced data as a replacement of counter data, it developed tools that could be implemented in spreadsheets and used to predict bicycling and pedestrian counts with a reasonable accuracy. It is expected that over time more counters will be installed covering different types of roadways and facilities across the state and thus, more data will be available to develop more standardized and reliable AEFs for different types of locations. Future research in this area should focus on collecting and using data from a larger and more diverse set of locations to test the sensitivity of the models and adjustment factors. Finally, given the increasing concerns about privacy, it is unlikely that quality of crowdsourced data will get better or have more coverage. Alternative data sources like video data or infrastructure-based sensor data may become more viable options to address data needs in the future. Future research should also consider these data sources as alternative and complementary data to counter and/or crowdsourced data for increasing their cost effectiveness and accuracy.

References

- Broach, J., Kothuri, S., Miah, M. M., McNeil, N., Hyun, K., Mattingly, S., Nordback, K., & Proulx, F. (2023). Evaluating the Potential of Crowdsourced Data to Estimate Network-Wide Bicycle Volumes. *Transportation Research Record*, 03611981231182388. https://doi.org/10.1177/03611981231182388

- Lee, K., & Sener, I. N. (2020). Emerging data for pedestrian and bicycle monitoring: Sources and applications. Transportation Research Interdisciplinary Perspectives, 4, 100095. https://doi.org/10.1016/j.trip.2020.100095

- Lin, Z., & Fan, W. (David). (2020). Modeling bicycle volume using crowdsourced data from the Strava smartphone application. International Journal of Transportation Science and Technology, 9(4), 334–343. https://doi.org/10.1016/j.ijtst.2020.03.003

- StreetLight Insight: Our Methodology and Data Sources. (2018). StreetLight Data.

- Wang, X.-D., R.-C. Chen, F. Yan, Z.-Q. Zeng, and C.-Q. Hong. Fast Adaptive K-Means Subspace Clustering for High-Dimensional Data. IEEE Access, Vol. 7, 2019, pp. 42639–42651. https://doi.org/10.1109/ACCESS.2019.2907043.

- Chen, R.-C., C. Dewi, S.-W. Huang, and R. E. Caraka. Selecting Critical Features for Data Classification Based on Machine Learning Methods. Journal of Big Data, Vol. 7, No. 1, 2020, p. 52. https://doi.org/10.1186/s40537-020-00327-4.

- Chakraborty, M., T. J. Gates, and S. Sinha. Causal Analysis and Classification of Traffic Crash Injury Severity Using Machine Learning Algorithms. Data Science for Transportation, Vol. 5, No. 2, 2023, p. 12. https://doi.org/10.1007/s42421-023-00076-9.

- Khoshgoftaar, T. M., M. Golawala, and J. V. Hulse. An Empirical Study of Learning from Imbalanced Data Using Random Forest. Patras, Greece, 2007.

- Li, Y., J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai. An Efficient Intrusion Detection System Based on Support Vector Machines and Gradually Feature Removal Method. Expert Systems with Applications, Vol. 39, No. 1, 2012, pp. 424–430. https://doi.org/10.1016/j.eswa.2011.07.032.

- Hauer, E., C. N. N. Jerry, and J. Lovell. Estimation of Safety at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1185, 1988, pp. 48–61.

- Persaud, B., and L. Dzbik. Accident Prediction Models for Freeways. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1401, 1993, pp. 55–60.

- Oh, J., C. Lyon, S. Washington, B. Persaud, and J. Bared. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1840, No. 1, 2003, pp. 41–49. https://doi.org/10.3141/1840-05.

- Shankar, V. N., R. B. Albin, J. C. Milton, and F. L. Mannering. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects

Negative Binomial Model. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1635, No. 1, 1998, pp. 44–48. https://doi.org/10.3141/1635-06.

- Mannering, F. L., V. Shankar, and C. R. Bhat. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. Analytic Methods in Accident Research, Vol. 11, 2016, pp. 1–16. https://doi.org/10.1016/j.amar.2016.04.001.
- McNally, M. G. (2008). The Four Step Model. *UC Irvine: Center for Activity Systems Analysis*. Retrieved from https://escholarship.org/uc/item/0r75311t
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of transport geography*, *52*, 90-97.
- Watkins, K., Ammanamanchi, R., LaMondia, J., & Le Dantec, C. A. (2016). *Comparison of smartphone-based cyclist GPS data sources* (No. 16-5309).
- Baqui, M., Samad, M. D., & Löhner, R. (2020). A novel framework for automated monitoring and analysis of high density pedestrian flow. *Journal of Intelligent Transportation Systems*, *24*(6), 585–597. https://doi.org/10.1080/15472450.2019.1643724
- Blanc, B., & Figliozzi, M. (2017). *Safety Perceptions, Roadway Characteristics, and Cyclists' Demographics: A Study of Crowdsourced Smartphone Bicycle Safety Data*. Transportation Research Board 96th Annual Meeting, Washington DC.
- Chandra, S., Naik, R. T., & Jimenez, J. (2020). Crowdsourcing for Mode Shift: An Empirical Evidence of its Success among College Students. *Transportation Research Procedia*, *48*, 1430–1434. https://doi.org/10.1016/j.trpro.2020.08.173
- Chen, X. Z., Van Dyke, C., Erhardt, G., & Chen, M. (2019). *Practices on Acquiring Proprietary Data for Transportation Applications: A synthesis of Highway Practice*. National Cooperative Highway Research Program. https://doi.org/10.17226/25519
- Dixit, V., Nair, D. J., Chand, S., & Levin, M. W. (2020). A simple crowdsourced delay-based traffic signal control. *PLOS ONE*, *15*(4), e0230598. https://doi.org/10.1371/journal.pone.0230598
- Elhenawy, M., Komol, M. R., Masoud, M., Liu, S. Q., Ashqar, H. I., Almannaa, M. H., Rakha, H. A., & Rakotonirainy, A. (2021). A Novel Crowdsourcing Model for Micro-Mobility Ride-Sharing Systems. *Sensors*, *21*(14), Article 14. https://doi.org/10.3390/s21144636
- *Every Day Counts: Innovation for a Nation on the Move* (EDC-5). (2021). Federal Highway Administration.
- Froehlich, J. (2021). *Combining Crowdsourcing and Machine Learning to Collect Sidewalk Accessibility Data at Scale*. Pacific Northwest Transportation Consortium (PacTrans). http://hdl.handle.net/1773/47853
- Gehrke, S. R., James, P., Reeves, H., Ron, S., Reardon, T. G., Keppard, B., & Ursprung, W. W. S. (2019). *A Pedestrian-Oriented Framework for Measuring Area-Wide Pedestrian Activity* (19–02908). Article 19–02908. Transportation Research Board 98th Annual MeetingTransportation Research Board. https://trid.trb.org/View/1572783

- Greg P. Griffin & Junfeng Jiao. (2018). *Can Crowdsourcing Support Co-productive Transportation Planning in Megaregion? Evidence from Local Practice* (USDOT/69A3551747135). Cooperative Mobility for Competitive Megaregions at The University of Texas at Austin.

- Griffin, G. P., & Jiao, J. (2019). The Geography and Equity of Crowdsourced Public Participation for Active Transportation Planning. *Transportation Research Record*, *2673*(1), 460–468. https://doi.org/10.1177/0361198118823498

- Griswold, J., Medury, A., Schneider, R., & Grembek, O. (2018). Comparison of Pedestrian Count Expansion Methods: Land Use Groups versus Empirical Clusters. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*, 036119811879300. https://doi.org/10.1177/0361198118793006

- Hamilton, I., Kersavage, K., Porter, R. J., Smith, K., Sanchez, J., Gayah, V., & Eccles. (2021). *An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC*. Federal Highway Administration.

- Hankey, S., Zhang, W., Le, H. T. K., Hystad, P., & James, P. (2021). Predicting bicycling and walking traffic using street view imagery and destination data. *Transportation Research Part D: Transport and Environment*, *90*, 102651. https://doi.org/10.1016/j.trd.2020.102651

- Hopkin, J., Ball, S., Hannay, A., Hutchins, R., Palmer, D., Rahman, S., Naberezhnykh, D., & Longhi, D. (2014). *Pilot trial of a transport crowdsourcing smartphone app: Final report - (PPR 719).*

- Huang, S., Chen, W., Yu, R., Yang, X., & Dong, D. (2018). Predicting Pedestrian Counts for Crossing Scenario Based on Fused Infrared-Visual Videos. *Journal of Advanced Transportation*, *2018*, 1–11. https://doi.org/10.1155/2018/8703576

- Karen Sentoff & James Sullivan. (2017). *Vermont Bicycle and Pedestrian Counting Program*. Transportation Research Center.

- Kristin Carlson, Brendan Murphy, Alireza Ermagun, David Levinson, & Andrew Owen. (2018). *Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis | Center for Transportation Studies* (CTS 18-05). Roadway Safety Institute Center for Transportation Studies University of Minnesota. https://www.cts.umn.edu/publications/report/safety-in-numbers-pedestrian-and-bicyclist-activity-and-safety-in-minneapolis

- Lesani, A., & Miranda-Moreno, L. (2019). Development and Testing of a Real-Time WiFi-Bluetooth System for Pedestrian Network Monitoring, Classification, and Data Extrapolation. *IEEE Transactions on Intelligent Transportation Systems*, *20*(4), 1484–1496. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2018.2854895

- Lesani, A., Nateghinia, E., & Miranda-Moreno, L. F. (2020). Development and evaluation of a real-time pedestrian counting system for high-volume conditions based on 2D LiDAR. *Transportation Research Part C: Emerging Technologies*, *114*, 20–35. https://doi.org/10.1016/j.trc.2020.01.018

- Liu, Z., Fu, X., Liu, Y., Tong, W., & Liu, Z. (2020). Estimating Sectional Volume of Travelers Based on Mobile Phone Data. *Journal of Transportation Engineering, Part A: Systems*, *146*(10), 04020110. https://doi.org/10.1061/JTEPBS.0000429

- Marzano, G., Lizut, J., & Siguencia, L. O. (2019). Crowdsourcing solutions for supporting urban mobility. *Procedia Computer Science*, *149*, 542–547. https://doi.org/10.1016/j.procs.2019.01.174
- Medury, A., Grembek, O., Loukaitou-Sideris, A., & Shafizadeh, K. (2019). Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses − a crowdsourcing approach. *Accident Analysis & Prevention*, *130*, 99–107. https://doi.org/10.1016/j.aap.2017.08.014
- Minh, L. (2019). *Exploring Crowdsourced Monitoring Data for Safety—Evaluation of Miovision Pedestrian Count Data*. Safety Through Disruption University Transportation Center (Safe-D) Texas A&M Transportation Institute. https://rosap.ntl.bts.gov/view/dot/50717
- Nordback, K., Kothuri, S., Johnstone, D., Lindsey, G., Ryan, S., & Raw, J. (2019). Minimizing Annual Average Daily Nonmotorized Traffic Estimation Errors: How Many Counters Are Needed per Factor Group? *Transportation Research Record*, *2673*(10), 295–310. https://doi.org/10.1177/0361198119848699
- Ohlms, P. B., Dougald, L. E., & MacKnight, H. E. (2019). Bicycle and Pedestrian Count Programs: Scan of Current U.S. Practice. *Transportation Research Record*, *2673*(3), 74–85. https://doi.org/10.1177/0361198119834924
- Ohlms, P. B., Dougald, L. E., MacKnight, H. E., & Virginia Transportation Research Council. (2018). *Assessing the Feasibility of a Pedestrian and Bicycle Count Program in Virginia* (FHWA/VTRC 19-R4). Virginia Transportation Research Council. https://rosap.ntl.bts.gov/view/dot/37155
- Olfert, C., Poapst, R., & Montufar, J. (2018). Incorporating the Effect of Special Events into Continuous Count Site Selection for Pedestrian Traffic. *Transportation Research Record*, *2672*(43), 65–74. https://doi.org/10.1177/0361198118788188
- Pu, Z., Zhang, Q., Zhuang, Y., Lv, Y., & Wang, Y. (2020). *A Device-Free Wi-Fi Sensing Method for Pedestrian Monitoring Using Channel State Information* (p. 220). https://doi.org/10.1061/9780784483138.019
- Rahman, M. S., Abdel-Aty, M., Hasan, S., & Cai, Q. (2019, January 17). *Applying Data Mining Techniques to Analyze the Pedestrian and Bicycle Crashes at the Macroscopic Level*. Transportation Research Board 98th Annual Meeting, Washington DC.
- Rahman, R., Redwan Shabab, K., Chandra Roy, K., Zaki, M. H., & Hasan, S. (2021). Real-Time Twitter Data Mining Approach to Infer User Perception Toward Active Mobility. *Transportation Research Record*, *2675*(9), 947–960. https://doi.org/10.1177/03611981211004966
- Rahman, Z., Mattingly, S. P., Kawadgave, R., Nostikasari, D., Roeglin, N., Casey, C., & Johnson, T. (2019). Using crowd sourcing to locate and characterize conflicts for vulnerable modes. *Accident Analysis & Prevention*, *128*, 32–39. https://doi.org/10.1016/j.aap.2019.03.014
- Rodrigues, F., & Pereira, F. C. (2018). Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data. *Transportation Research Part C: Emerging Technologies*, *95*, 636–651. https://doi.org/10.1016/j.trc.2018.08.007
- Roll, J. (2021). Daily Traffic Count Imputation for Bicycle and Pedestrian Traffic: Comparing Existing Methods with Machine Learning Approaches. *Transportation Research Record*, *2675*(11), 1428–1440. https://doi.org/10.1177/03611981211027161

- Shi, J., Wang, X., & Xiao, H. (2018). Real-Time Pedestrian Tracking and Counting with TLD. *Journal of Advanced Transportation*, *2018*, 1–7. https://doi.org/10.1155/2018/8486906
- Smith, A. (2015a). Crowdsourcing for Active Transportation. *Ite Journal*, *85*, 30–35.
- Smith, A. (2015b). *Crowdsourcing Pedestrian and Cyclist Activity Data* (White Paper DTFHGI-11-H-00024). Federal Highway Administration.
- Tarkiainen, M., Bäck, A., Hulkkonen, J., & Könkkölä, K. (2011). Crowd Sourcing Accessibility Related Information From POI-Destinations in Finland: 18th ITS World Congress, Orlando. *Proceedings of the 18th ITS World Congress*. 18th ITS World Congress, Orlando.
- Tolford, T. M., Izadi, M., Ash, C., Codjoe, J., & University of New Orleans. Transportation Institute. (2019). *Pedestrians and Bicyclists Count: Developing a Statewide Multimodal Count Program* (FHWA/LA.17/599). https://rosap.ntl.bts.gov/view/dot/63547
- Torbic, D. J., Cook, D. J., & Hutton, J. M. (2019). *Traffic Safety Counter measures: Impacts on Pedestrian and Bicyclist Safety*. National Cooperative Highway Research Program.
- Traut, E. J., & Steinfeld, A. (2019). Identifying commonly used and potentially unsafe transit transfers with crowdsourcing. *Transportation Research Part A: Policy and Practice*, *122*, 99–111. https://doi.org/10.1016/j.tra.2019.02.005
- Turner, S., Lasley, P., Benz, R., Martin, M., & Texas A&M Transportation Institute. (2021). *Training and Implementation Resources for Pedestrian and Bicyclist Count Database and Monitoring Process* (FHWA/TX-20/5-6927-01-IPR1). Texas A&M Transportation Institute. https://rosap.ntl.bts.gov/view/dot/60131
- Turner, S., Martin, M., Griffin, G., Le, M., Das, S., Wang, R., Dadashova, B., & Li, X. (2020). *Exploring Crowdsourced Monitoring Data for Safety*. Safety through Disruption (Safe-D) University Transportation Center (UTC); Texas A&M Transportation Institute; United States. Dept. of Transportation. Office of the Assistant Secretary for Research and Technology. https://rosap.ntl.bts.gov/view/dot/50717
- Yang, H., Cetin, M., Ma, Q., & Virginia Transportation Research Council (VTRC). (2020). *Guidelines for Using StreetLight Data for Planning Tasks* (FHWA/VTRC 20-R23). Virginia Transportation Research Council (VTRC). https://rosap.ntl.bts.gov/view/dot/55501
- Yin, B., & Leurent, F. (2021). Exploring Individual Activity-Travel Patterns Based on Geolocation Data from Mobile Phones. *Transportation Research Record, 2675*(12), 771–783. https://doi.org/10.1177/03611981211031234

Appendix A: Bibliography

Bibliography

Crowdsourcing Pedestrian and Cyclist Activity Data

This paper considers how crowdsourcing applications and crowdsourced data are currently being applied, as well as potential new uses for active transportation research and planning efforts of various types. The objectives of this white paper are to review crowdsourced bicycle and pedestrian data resources and crowdsourcing tools; discuss potential planning implementations of crowdsourced data for a variety of bicycle and pedestrian project types; and provide examples of how crowdsourcing is currently being used by the planning community. Due to software application turnover, many of the examples provided describe tools that may no longer be in use, have evolved significantly, or have been/will eventually be depreciated with the advance of new technologies. This paper is not intended to be a comprehensive outline of crowdsourcing applications in the transportation planning profession or a dictionary of crowdsourcing system types, but rather a resource for those interested in using crowdsourcing systems in active transportation planning and research.

- **Record URL:**
  http://www.pedbikeinfo.org/cms/downloads/PBIC_WhitePaper_Crowdsourcing.pdf

- **Corporate Authors:**
  Fehr & Peers
  University of North Carolina, Chapel Hill
  Highway Safety Research Center, Pedestrian and Bicycle Information
  730 Martin Luther King Jr. Boulevard, CB 3430
  Chapel Hill, NC United States 27599-3430

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590

Authors: Smith, Amy

Publication Date: 2015-1

Language   English

Media Info

- **Media Type:** Digital/other
- **Edition:** White Paper
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** 35p

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Cyclists; Data collection; Data quality; Decision making; Pedestrians; Research projects; Transportation planning; Travel behavior

- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Planning and Forecasting; I72: Traffic and Transport Planning;

Filing Info

- **Accession Number:** 01557291
- **Record Type:** Publication
- **Contract Numbers:** DTFHGI-11-H-00024
- **Files:** TRIS, ATRI, USDOT
- **Created Date:** Mar 11 2015 8:38AM

TWO

This is an active and ongoing project of interest – the final reports will be available after project completion in July 2026.

State DOT Usage of Bicycle and Pedestrian Data: Practices, Sources, Needs, and Gaps

States would benefit from research that summarizes the existing literature on active transportation data and catalogs relevant sources and data sets related to active transportation. Furthermore, innovative, cost-effective data use cases could provide scalable examples among state DOT practitioners. The research should also capture any untraditional or unusual sources or applications of data that may be primarily for other purposes but could be adapted or integrated into active transportation analysis. This research would inform practitioners on the expanse of available data, which may be unconventional, such as police and hospital reports; capture information on how peer agencies are identifying and using data, identify gaps for future research, and provide recommendations (identification, collection, cleaning, utilizing, analyzing, standardizing, storing, funding, privacy and legal concerns, etc.) Submitting a data-related research problem statement is part of AASHTO's Council on Active Transportation's work plan and the Council's number one priority. The research proposed in this problem statement complements ongoing research. In particular, this problem statement expands the scope of NCHRP Project 20-05/Topic 50-10 to include bicycling data and conduct research aimed at identifying the data needs and wants of state DOTs and the gap between them and what is available. NCHRP Project 08-108, Developing National Performance Management Data Strategies to Address Data Gaps, Standards, and Quality, does not include any information on data related to active transportation. This proposed research will dive deeper into data than the more general NCHRP Project 20-123(02), Research Roadmap for the AASHTO Council on Active Transportation. It will also build upon FHWA's Roadway Data Improvement Program (RDIP), which seeks to improve the quality of states' roadway data. Under the Roadway Safety Program, FHWA also prepares Roadway Safety Data Capabilities Assessments for states that include data components. The objectives of this research are to determine how state DOTs are using data and to identify data sources, gaps, and recommendations on the next steps to develop the data and tools state DOTS need. To fulfill these objectives, the research contractor will need to complete the following: (1) Summarize/synthesize existing research on active transportation data. (2) Survey state DOTs to understand the current state of data sources and uses, as well as unmet needs. (3) Catalog active transportation data sets, common attributes, uses, including both well-known sources (e.g., Strava Metro) and less utilized sources (e.g., police reports, hospital reports, etc.). (4) Conduct a gap analysis between the data that state DOTs need/want versus what is currently available/being used. (5) Develop recommendations on next steps for developing, standardizing, maintaining, and storing the identified data, information, models, and/or tools. The research will present an urgently needed inventory of available data sources and identify any gaps based on direct feedback from state DOTs. Practitioners will gain valuable insight into how their peers are utilizing data and receive recommendations to bolster data in their agencies. This research will advance the technical expertise of state DOTs.

- **Record URL:**
  http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4946

- **Supplemental Notes:**
  - Ongoing.

Language: English

Project

- **Status:** Awarded
- **Funding:** $800000
- **Contract Numbers:**
  Project 07-31
- **Sponsor Organizations:**
  National Cooperative Highway Research Program
  Transportation Research Board
  500 Fifth Street, NW
  Washington, DC United States 20001

  American Association of State Highway and Transportation Officials (AASHTO)
  444 North Capitol Street, NW
  Washington, DC United States 20001

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590

- **Project Managers:**
  Parker, Stephan
- **Start Date:** 20211012
- **Expected Completion Date:** 07/31/2026
- **Actual Completion Date:** 0

Subject/Index Terms

- **TRT Terms:** Best practices; Bicycling; Cyclists; Data collection; Data management; Injuries; Pedestrian safety; Pedestrians; State departments of transportation; State of the practice; Surveys
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Planning and Forecasting; Safety and Human Factors;

Filing Info

- **Accession Number:** 01739644
- **Record Type:** Research project
- **Source Agency:** Transportation Research Board
- **Contract Numbers:** Project 07-31
- **Files:** TRB, RIP

- **Created Date:** May 18 2020 3:05PM

THREE

Daily Traffic Count Imputation for Bicycle and Pedestrian Traffic: Comparing Existing Methods with Machine Learning Approaches

Monitoring nonmotorized traffic is becoming increasingly common practice at local and state departments of transportation. These travel activity data are necessary to monitor the system and track progress toward active transportation policy and program goals. A common problem is that permanent count site data are often missing, making those sites less useful. Being able to accurately estimate those missing data records functionally increases the amount of data available to use by themselves as metrics for monitoring traffic but also makes available more data for factoring short-term sites. Using nonmotorized traffic counts from several cities in Oregon, this research compared the ability of day-of-year (DOY) factors, a statistical model, and machine learning algorithms to accurately impute daily traffic records for annual traffic estimation. Based on exhaustive cross-validation experiments using data not missing at random scenarios, this research concluded that random forest and DOY factor approaches could be used to impute daily counts for nonmotorized traffic but each approach comes with tradeoffs. Though for many missing data scenarios random forest performed best, this method is complicated to estimate and apply. DOY factor-based methods are simpler to create and apply, and though more accurate in scenarios with significant amounts of missing data, they were less flexible given the need for data from neighboring count sites. Negative binomial regression was also found to work well in scenarios with moderate to low amounts of missing data. This work can inform nonmotorized traffic count programs needing vetted solutions for traffic data imputation.

- **Record URL:**
  https://doi.org/10.1177/03611981211027161

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Supplemental Notes:**
  -
  - **Authors:** Roll, Josh
- **Publication Date:** 2021-11

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** pp 1428-1440
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2675
  - Issue Number: 11
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Bicycle counts; Machine learning; Nonmotorized transportation; Pedestrian traffic; Traffic counting; Urban areas
- **Geographic Terms:** Oregon
- **Subject Areas:** Operations and Traffic Management; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01778687
- **Record Type:** Publication
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Jul 28 2021 3:13PM

FOUR

Exploring the Use of Crowdsourced Data Sources for Pedestrian Count Estimations

Counts provide the foundation for measuring nonmotorized travel along a link or a network and are also useful for monitoring trends, planning new infrastructure, and for conducting safety, health, and economic analyses. For safety analysis, they are critical in assessing the exposure to risk. Over the last decade, several automated technologies have been developed to count bicyclists and pedestrians. Despite advances in counting technology, cost and other considerations will continue to limit direct observation to small subsets of entire networks, as is the case for motorized traffic. A primary limitation with these counters is that they can only provide information about the activity that is directly on or near them but nothing about the activity on the network. The lack of widely available pedestrian count data precludes safety studies and analysis of trends, which has become critically important especially with the nationwide increase in pedestrian crashes over the last decade. The emergence of crowdsourced data such as Strava and StreetLight has allowed for the collection of large-scale datasets over broad areas of the network. While several research studies have evaluated and applied bicycle data from these datasets, no study has yet looked at pedestrian count estimates from these data sources or assessed how these compare to traditional pedestrian counts and other measures of pedestrian activity such as pedestrian actuations from traffic signals. The current study will evaluate pedestrian data estimates from the crowdsourced data sets and explore how these can be used along with traditional count data and sociodemographic data to derive count estimates.

- **Record URL:**
  https://nitc.trec.pdx.edu/research/project/1489

Language English

Project

- **Status:** Active
- **Funding:** $165000
- **Contract Numbers:** NITC-1489
- **Sponsor Organizations:**

Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC United States 20590

- **Managing Organizations:**
  TREC at Portland State University
  1900 SW Fourth Ave, Suite 175
  P.O. Box 751
  Portland, Oregon United States 97201

- **Performing Organizations:**
  Portland State University
  Department of Civil and Environmental Engineering
  Engineering Bldg, 301D, 1930 SW 4th Ave.
  Portland, OR United States 97201

  University of Texas at Arlington
  Department of Civil Engineering
  Box 19308
  Arlington, TX United States 76019

- **Principal Investigators:**
  Kothuri, Sirisha
  Hyun, Kate
  Mattingly, Stephen
  McNeil, Nathan
- **Start Date:** 20211001
- **Expected Completion Date:** 20221231
- **Actual Completion Date:** 0
- **USDOT Program:** University Transportation Centers

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Estimates; Pedestrian counts
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01781407
- **Record Type:** Research project
- **Source Agency:** National Institute for Transportation and Communities
- **Contract Numbers:** NITC-1489
- **Files:** UTC, RIP
- **Created Date:** Sep 7 2021 5:54PM

FIVE

An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC

This report built on a geospatial pilot effort by the Highway Safety Information System (HSIS) using data from Charlotte, NC. The main objective of this study was to spatially integrate HSIS data with multi-jurisdictional and emerging datasets to analyze two measures of pedestrian safety performance: the severity of a pedestrian crash that has occurred, and the probability that a pedestrian crash will occur. The study explored several high-priority research topics in safety data and analysis, including pedestrian crash analysis, probe data integration and analysis, and geospatial HSIS data integration. The project team developed a pedestrian count model to predict pedestrian volumes at locations without pedestrian counts and integrated speed information from probe data to supplement other roadway and contextual transportation data from several agencies. Demographic and socioeconomic characteristics, employment, land use, sidewalk presence, transit access, and roadway and intersection characteristics all significantly contributed to pedestrian volume predictions. The project team identified numerous significant factors that influenced pedestrian crash severity and probability. These factors included those identified in previous research, as well as new relationships between pedestrian volumes and vehicular traffic that have implications for pedestrian safety-in-numbers concepts. Results showed that higher pedestrian volumes resulted in both lower crash severities and probabilities, but the safety benefit was reduced by higher vehicle volumes. Higher speeds, higher traffic volumes, larger vehicles striking the pedestrian, pedestrian impairment, and older pedestrian ages were all indicative of higher probabilities of a pedestrian crash resulting in a fatality or serious injury. By adding a direct measure of speed from probe data (and given the known importance of speed to crash injury severity), the pedestrian crash severity model excluded commonly used speed surrogates without sacrificing model fit. The probability of a pedestrian crash occurring on a road segment was affected by segment length, interactions of pedestrian volumes and traffic volumes, and interactions of posted speed limit, median presence, and number of lanes. This study highlights the applicability of integrating HSIS with emerging safety data resources to inform data-driven and performance-based approaches to road safety management.

- **Record URL:**
  https://www.fhwa.dot.gov/publications/research/safety/21087/21087.pdf

- **Summary URL:**
  https://www.fhwa.dot.gov/publications/research/safety/21087/index.cfm

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/57960

- **Corporate Authors:**
  Vanasse Hangen Brustlin, Incorporated
  Raleigh, NC United States

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590

- **Authors:** Hamilton, Ian
  - Kersavage, Kristin
  - Porter, R J
  - Smith, Keith
  - Sanchez, Josie
  - Gayah, Vikash
  - Eccles, Kimberly
- **Publication Date:** 2021-10

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Report
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** 67p
-

Subject/Index Terms

- **TRT Terms:** Crash data; Crash risk forecasting; Crash severity; Pedestrian safety; Pedestrian vehicle crashes
- **Identifier Terms:** Highway Safety Information System
- **Geographic Terms:** Charlotte (North Carolina)
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01782472
- **Record Type:** Publication
- **Report/Paper Numbers:** FHWA-HRT-21-087
- **Contract Numbers:** DTFH61-11-C-00050
- **Files:** NTL, TRIS, ATRI, USDOT
- **Created Date:** Sep 8 2021 2:43PM

SIX

Motor Vehicle Safety Countermeasures: Impacts on Pedestrian and Bicyclist Safety

The objective of this research is to (1) develop a decision tool (e.g., checklists, matrices) to identify tradeoffs in safety and mobility in different types of facilities between motorists and pedestrians and bicyclists arising from motor vehicle safety countermeasures; and (2) develop future research needs. The research should include, but not be limited to, the following: (1) Identify candidate countermeasures: review literature (including the Crash Modification Factors Clearinghouse) and consult with subject matter experts to identify the highest priority motor vehicle safety countermeasures to explore in this project. - (2) Conduct in-depth reviews of countermeasures: synthesize the literature, guidelines, and other materials to identify tradeoffs related to motorist, pedestrian, and bicyclist safety, and other issues such as convenience (e.g., delay and travel distance for pedestrians), and Level of Service. (3) Develop the decision tool and research recommendations: develop checklists/matrices to reflect tradeoffs found in #2;

develop a prioritized list of research for the AASHTO Technical Committee on Non-motorized Transportation.

- **Record URL:**
  http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4092

Language English

Project

- **Status:** Completed
- **Funding:** $100000
- **Contract Numbers:**
  Project 20-07, Task 393
- **Sponsor Organizations:**
  National Cooperative Highway Research Program
  Transportation Research Board
  500 Fifth Street, NW
  Washington, DC United States 20001

  American Association of State Highway and Transportation Officials (AASHTO)
  444 North Capitol Street, NW
  Washington, DC United States 20001

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590
- **Performing Organizations:**
  MRIGlobal
- **Principal Investigators:** Torbic, Darren
- **Start Date:** 20210908
- **Expected Completion Date:** 20190329
- **Actual Completion Date:** 0

Subject/Index Terms

- **TRT Terms:** Crash modification factors; Cyclists; Decision support systems; Literature reviews; Mobility; Pedestrian safety; Traffic safety
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting; Safety and Human Factors;

Filing Info

- **Accession Number:** 01781474
- **Record Type:** Research project
- **Source Agency:** Transportation Research Board

- **Contract Numbers:** Project 20-07, Task 393
- **Files:** TRB, RIP
- **Created Date:** Sep 6 2021 3:17PM

SEVEN

Real-Time Twitter Data Mining Approach to Infer User Perception Toward Active Mobility

This study evaluates the level of service of shared transportation facilities through mining geotagged data from social media and analyzing the perceptions of road users. An algorithm is developed adopting a text classification approach with contextual understanding to filter out relevant information related to users' perceptions toward active mobility. Using a heuristic-based keyword matching approach produces about 75% tweets that are out of context, so that approach is deemed unsuitable for information extraction from Twitter. This study implements six different text classification models and compares the performance of these models for tweet classification. The model is applied to real-world data to filter out relevant information, and content analysis is performed to check the distribution of keywords within the filtered data. The text classification model "term frequency-inverse document frequency" vectorizer-based logistic regression model performed best at classifying the tweets. To select the best model, the performances of the models are compared based on precision, recall, F1 score (geometric mean of precision and recall), and accuracy metrics. The findings from the analysis show that the proposed method can help produce more relevant information on walking and biking facilities as well as safety concerns. By analyzing the sentiments of the filtered data, the existing condition of biking and walking facilities in the DC area can be inferred. This method can be a critical part of the decision support system to understand the qualitative level of service of existing transportation facilities.

- **Record URL:**
  https://doi.org/10.1177/03611981211004966

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Supplemental Notes:**
  -
- **Authors:**
  - Rahman, Rezaur
  - Redwan Shabab, Kazi
  - Chandra Roy, Kamol
  - Zaki, Mohamed H
  - Hasan, Samiul
- **Publication Date:** 2021-9
-

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; References; Tables;

- **Pagination:** pp 947-960
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2675
  - Issue Number: 9
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Bicycle facilities; Data mining; Evaluation; Level of service; Logistic regression analysis; Nonmotorized transportation; Pedestrian areas; Real time information; Social media
- **Identifier Terms:** Twitter
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Terminals and Facilities;

Filing Info

- **Accession Number:** 01764224
- **Record Type:** Publication
- **Report/Paper Numbers:** TRBAM-21-02955
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Dec 23 2020 11:22AM

EIGHT

Practices on Acquiring Proprietary Data for Transportation Applications

This synthesis gathers information about how state departments of transportation (DOTs) and metropolitan planning organizations (MPOs) acquire proprietary data for transportation applications. The focus is on those data generated by technologies such as GPS, mobile phones, or crowdsource travel alerts. Recent technological advancements have led to new types of transportation data with characteristics that include improved quality and greater temporal and wider geographical coverage than traditional data sets. State DOTs and MPOs face challenges associated with obtaining new proprietary data. The information contained in this synthesis was obtained using three sources. First, a literature review compiled relevant existing research about the topic. Second, the consultant surveyed state DOTs and large MPOs. Finally, the consultant conducted interviews with five agencies that identified how agencies acquire proprietary data, which resulted in case examples and lessons learned that describe how state DOTs and MPOs assess licensing options, caveats and risks, and data negotiations. The study found that unmet needs for data and new insights offered by proprietary data are the main driving factors that prompt transportation agencies to acquire proprietary data. Among the data that have been acquired, speed data are being used widely by transportation agencies around the United States for a variety of applications and have been integrated into mainstream agency business areas by some agencies. Numerous uses have also been found for O-D data enabled by highly precise GPS data from in-vehicle systems and mobile phones. The study also found that most procurements were directly handled by transportation agencies, while some were handled by consultants (including universities). The survey respondents and interviewees identified several barriers and concerns associated with these proprietary data and shared their perspectives and practices as they relate to these concerns. These concerns include: data and service quality, cost, staff expertise and information technology resources, finding the right product, and legal issues.

- **Record URL:**
  http://www.trb.org/Main/Blurbs/179336.aspx

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/isbn/9780309480512
- **Authors:**
  - Zhang, Xu
  - Van Dyke, Chris
  - Erhardt, Greg
  - Chen, Mei
- **Publication Date:** 2019

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Appendices; Figures; Glossary; References; Tables;
- **Pagination:** 227p
- **Serial:**
  - NCHRP Synthesis of Highway Practice
  - Issue Number: 541
  - Publisher: Transportation Research Board
  - ISSN: 0547-5570
- **Publication flags:** Open Access (libre)

Subject/Index Terms

- **TRT Terms:** Advanced driver information systems; Automatic data collection systems; Crowdsourcing; Data collection; Global Positioning System; Mobile telephones; Private enterprise; Technological innovations
- **Subject Areas:** Data and Information Technology; Highways; Operations and Traffic Management; Planning and Forecasting;

Filing Info

- **Accession Number:** 01713275
- **Record Type:** Publication
- **ISBN:** 9780309480512
- **Report/Paper Numbers:** Project 20-05, Topic 49-14
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Aug 5 2019 1:07PM

NINE

Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data

Accurately modeling traffic speeds is a fundamental part of efficient intelligent transportation systems. Nowadays, with the widespread deployment of global positioning system (GPS)-enabled devices, it has become possible to crowdsource the collection of speed information to road users (e.g. through mobile applications or dedicated in-vehicle devices). Despite its rather wide spatial coverage, crowdsourced speed data also brings very important challenges, such as the highly variable measurement noise in the data due to a variety of driving behaviors and sample sizes. When not properly accounted for, this noise can severely compromise any application that relies on accurate traffic data. In this article, the authors propose the use of heteroscedastic Gaussian processes (HGP) to model the time-varying uncertainty in large-scale crowdsourced traffic data. Furthermore, the authors develop a HGP conditioned on sample size and traffic regime (SSRC-HGP), which makes use of sample size information (probe vehicles per minute) as well as previous observed speeds, in order to more accurately model the uncertainty in observed speeds. Using 6 months of crowdsourced traffic data from Copenhagen, the authors empirically show that the proposed heteroscedastic models produce significantly better predictive distributions when compared to current state-of-the-art methods for both speed imputation and short-term forecasting tasks.

- **Record URL:**
  https://doi.org/10.1016/j.trc.2018.08.007

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S0968090X18300147

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/0968090X
- **Supplemental Notes:**

- **Authors:**
  - Rodrigues, Filipe
  - Pereira, Francisco C
- **Publication Date:** 2018-10

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; References; Tables;
- **Pagination:** pp 636-651
- **Serial:**
  - Transportation Research Part C: Emerging Technologies
  - Volume: 95

- Issue Number: 0
- Publisher: Elsevier
- ISSN: 0968-090X
- Serial URL: http://www.sciencedirect.com/science/journal/0968090X

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Data collection; Floating car data; Forecasting; Global Positioning System; Heteroscedasticity; Traffic data; Traffic speed; Uncertainty
- **Uncontrolled Terms:** Gaussian processes
- **Subject Areas:** Data and Information Technology; Highways; Planning and Forecasting;
-

Filing Info

- **Accession Number:** 01681831
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Sep 4 2018 3:02PM

TEN

Safety Perceptions, Roadway Characteristics, and Cyclists' Demographics: A Study of Crowdsourced Smartphone Bicycle Safety Data

Safety is one of the most important factors that affects how much, where, and when people cycle. Bicycle safety has become a primary concern across many communities, especially in the context of Vision Zero programs.However, datasets about both perceived and actual cyclists' safety are difficult to collect.More specifically, data related to near-misses or infrastructure problems that may affect bicycle safety are not systematically collected or analyzed. The Oregon Department of Transportation (ODOT) has sponsored a study to evaluate whether crowdsourcing methods could be useful to collect data related to bicycle safety and infrastructure.A result of this project was a smartphone application, called ORcycle, which was developed to crowdsource bicycle travel and safety data in Oregon. There is scant research related to the quality and usefulness of crowdsourced transportation data and to the relationships among safety reports, cyclists' demographics, and roadway environment. This research explores which factors affect the urgency of a perceived safety problem. Results are encouraging; a thorough quality control reveals a very high level of data accuracy. Statistical models produce intuitive results and indicate that users are generally reliable when reporting the urgency of safety and infrastructure issues. Among the demographic variables, cyclists' gender and income levels seem to influence safety reports'urgency and type. Among the traffic and infrastructure related variables, higher traffic volumes and speeds increase the urgency of safety reports,whereas other variables such as long waiting times at traffic signals are associated with less urgent safety reports.

- **Supplemental Notes:**
  - This paper was sponsored by TRB committee ABJ35 Standing Committee on Highway Traffic Monitoring. Alternate title: Safety Perceptions, Roadway Characteristics, and Cyclists' Demographics: Study of Crowdsourced Smartphone Bicycle Safety Data
- **Corporate Authors:**
  Transportation Research Board
  500 Fifth Street, NW
  Washington, DC United States 20001

- **Authors:**
    - Blanc, Bryan
    - Figliozzi, Miguel
- **Conference:**
    - Transportation Research Board 96th Annual Meeting
    - Location: Washington DC, United States
    - Date: 2017-1-8 to 2017-1-12
- **Date:** 2017

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Photos; References; Tables;
- **Pagination:** 20p
- **Monograph Title:** TRB 96th Annual Meeting Compendium of Papers

Subject/Index Terms

- **TRT Terms:** Bicycle travel; Crowdsourcing; Cyclists; Data collection; Demographics; Infrastructure; Perception; Smartphones; Traffic safety
- **Identifier Terms:** Oregon Department of Transportation; Vision Zero
- **Geographic Terms:** Oregon
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01623107
- **Record Type:** Publication
- **Report/Paper Numbers:** 17-03262
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Dec 8 2016 11:13AM

ELEVEN

Using Crowd Sourcing to Locate and Characterize Conflicts for Vulnerable Modes

Most agencies and decision-makers rely on crash and crash severity (property damage only, injury or fatality) data to assess transportation safety; however, in the context of public health where perceptions of safety may influence the willingness to adopt active transportation modes (e.g. bicycling and walking), pedestrian-motor vehicle and other similar conflicts types may define a better performance measure for safety assessment. In the field of transportation safety, an absolute conflict occurs when two parties' paths cross and one of the parties must undertake an evasive maneuver (e.g. change direction or stop) to avoid a crash. Other less severe conflicts where paths cross but no evasive maneuver is required may also impact public perceptions of safety especially for vulnerable modes. Most of the existing literature focuses on vehicle conflicts. While in the past several years, more research has investigated bicycle and pedestrian conflicts, most of this has focused on the intersection environment. A comprehensive analysis of conflicts appears critical. The major objective of this study is two-fold: 1) Development of an innovative and cost

effective conflict data collection technique to better understand the conflicts (and their severity) involving vulnerable road users (e.g. bicycle/pedestrian, bicycle/motor vehicle, and pedestrian/motor vehicle) and their severity. 2) Test the effectiveness and practicality of the approach taken and its associated crowd sourced data collection. In an endeavor to undertake these objectives, the researchers developed an android-based crowd-sourced data collection app. The crowd-source data collected using the app is compared with traditional fatality data for hot spot analysis. At the end, the app users provide feedback about the overall competency of the app interface and the performance of its features to the app developers. If widely adopted, the app will enable communities to create their own data collection efforts to identify dangerous sites within their neighborhoods. Agencies will have a valuable data source at low-cost to help inform their decision making related to bicycle and pedestrian education, encouragement, enforcement, programs, policies, and infrastructure design and planning.

- **Record URL:**
  - https://doi.org/10.1016/j.aap.2019.03.014

- **Record URL:**
  - http://www.sciencedirect.com/science/article/pii/S0001457519304749

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/00014575
- **Supplemental Notes:**
  -
- **Authors:**
  - Rahman, Ziaur
  - Mattingly, Stephen P
  - Kawadgave, Rahul
  - Nostikasari, Dian
  - Roeglin, Nicole
  - Casey, Colleen
  - Johnson, Taylor
- **Publication Date:** 2019-7

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** pp 32-39
- **Serial:**
  - Accident Analysis & Prevention
  - Volume: 128
  - Issue Number: 0
  - Publisher: Elsevier
  - ISSN: 0001-4575
  - Serial URL: http://www.sciencedirect.com/science/journal/00014575

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Data collection; Mobile applications; Pedestrian vehicle crashes; Traffic conflicts; Vulnerable road users
- **Candidate Terms:** Bicycle vehicle interface
- **Subject Areas:** Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01707840
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** May 21 2019 3:09PM

TWELVE

Exploring Individual Activity-Travel Patterns Based on Geolocation Data from Mobile Phones

Data mining techniques can extract useful activity and travel information from large-scale data sources such as mobile phone geolocation data. This paper aims to explore individual activity-travel patterns from samples of mobile phone users using a two-week geolocation data set from the Paris region in France. After filtering the data set, we propose techniques to identify individual stays and activity places. Typical activity places such as the primary anchor place and the secondary place are detected. The daily timeline (i.e., activity-travel program) is reconstructed with the detected activity places and the trips in-between. Based on user-day timelines, a three-stage clustering method is proposed for mobility pattern analysis. In the method framework, activity types are first identified by clustering analysis. In the second stage, daily mobility patterns are obtained after clustering the daily mobility features. Activity-travel topologies are statistically investigated to support the interpretation of daily mobility patterns. In the last stage, we analyze statistically the individual mobility patterns for all samples over 14?days, measured by the number of days for all kinds of daily mobility patterns. All individual samples are divided into several groups where people have similar travel behaviors. A kmeans++ algorithm is applied to obtain the appropriate number of patterns in each stage. Finally, we interpret the individual mobility patterns with statistical descriptions and reveal home-based differences in spatial distribution for the grouped individuals.

- **Record URL:**
  https://doi.org/10.1177/03611981211031234

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Supplemental Notes:**
  - Biao Yin https://orcid.org/0000-0001-8087-5939
- **Authors:**
  - Yin, Biao

- Leurent, Fabien
- **Publication Date:** 2021-12

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** pp 771-783
- **Serial:**
    - Transportation Research Record: Journal of the Transportation Research Board
    - Volume: 2675
    - Issue Number: 12
    - Publisher: Sage Publications, Incorporated
    - ISSN: 0361-1981
    - EISSN: 2169-4052
    - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Location data; Mobility; Smartphones; Travel behavior; Travel demand
- **Geographic Terms:** Paris (France)
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01764178
- **Record Type:** Publication
- **Report/Paper Numbers:** TRBAM-21-01547
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Dec 23 2020 11:21AM

THIRTEEN

Exploring the Use of Crowdsourced Data Sources for Pedestrian Count Estimations

Counts provide the foundation for measuring nonmotorized travel along a link or a network and are also useful for monitoring trends, planning new infrastructure, and for conducting safety, health, and economic analyses. For safety analysis, they are critical in assessing the exposure to risk. Over the last decade, several automated technologies have been developed to count bicyclists and pedestrians. Despite advances in counting technology, cost and other considerations will continue to limit direct observation to small subsets of entire networks, as is the case for motorized traffic. A primary limitation with these counters is that they can only provide information about the activity that is directly on or near them but nothing about the activity on the network. The lack of widely available pedestrian count data precludes safety studies and analysis of trends, which has become critically important especially with the nationwide increase in pedestrian crashes over the last decade. The emergence of crowdsourced data such as Strava and StreetLight has allowed for the collection of large-scale datasets over broad areas of the network. While several research studies have evaluated and applied bicycle data from these datasets, no study has yet looked at pedestrian count estimates from these data sources or assessed how these compare to traditional pedestrian counts and other measures of pedestrian activity such as pedestrian actuations from traffic signals. The current study will evaluate pedestrian data estimates from the crowdsourced data sets and explore how these can be used along with traditional count data and sociodemographic data to derive count estimates.

- **Record URL:**
  https://nitc.trec.pdx.edu/research/project/1489

Language English

Project

- **Status:** Active
- **Funding:** $165000
- **Contract Numbers:**
  NITC-1489
- **Sponsor Organizations:**
  Office of the Assistant Secretary for Research and Technology
  University Transportation Centers Program
  Department of Transportation
  Washington, DC United States 20590
- **Managing Organizations:**
  TREC at Portland State University
  1900 SW Fourth Ave, Suite 175
  P.O. Box 751
  Portland, Oregon United States 97201
- **Performing Organizations:**
  Portland State University
  Department of Civil and Environmental Engineering
  Engineering Bldg, 301D, 1930 SW 4th Ave.
  Portland, OR United States 97201

  University of Texas at Arlington
  Department of Civil Engineering
  Box 19308
  Arlington, TX United States 76019
- **Principal Investigators:**
  Kothuri, Sirisha
  Hyun, Kate
  Mattingly, Stephen
  McNeil, Nathan

- **Start Date:** 20211001
- **Expected Completion Date:** 20221231
- **Actual Completion Date:** 0
- **USDOT Program:** University Transportation Centers

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Estimates; Pedestrian counts
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01781407
- **Record Type:** Research project
- **Source Agency:** National Institute for Transportation and Communities
- **Contract Numbers:** NITC-1489
- **Files:** UTC, RIP
- **Created Date:** Sep 7 2021 5:54PM

FOURTEEN

Combining Crowdsourcing and Machine Learning to Collect Sidewalk Accessibility Data at Scale

The authors are developing new data collection approaches that use a combination of remote crowdsourcing, machine learning, and online map imagery. Their newest effort, called Project Sidewalk, enables online crowdworkers to remotely label pedestrian-related accessibility problems by virtually walking through city streets in Google Street View. In 2019, the authors completed an 18-month deployment in Washington, D.C.: 1,150+ users provided over 200,000 geo-located sidewalk accessibility labels and audited 3,000 miles of D.C. streets. With simple quality control mechanisms, the authors found that minimally trained remote crowd workers could find and label 92 percent of accessibility problems in street view scenes, including missing curb ramps, obstacles in the path, surface problems, and missing sidewalks. For their PacTrans project, the authors proposed three threads of additional work. (1) First, the authors are deploying Project Sidewalk into three more cities, including two in the Pacific Northwest: Seattle, Washington, and Newberg, Oregon, to enable them to study and compare sidewalk accessibility factors across cities. (2) Second, to further scale their approach, the authors proposed new methods to automatically identify and classify sidewalk problems using deep learning techniques, which would be uniquely enabled by their large dataset. (3) Finally, the authors proposed new sidewalk accessibility models and interactive visualization tools to give stakeholders—from citizens to transit authorities—new understanding of their city's accessibility.

- **Dataset URL:**
  https://doi.org/10.7910/DVN/YOTY6A

- **Record URL:**
  http://hdl.handle.net/1773/47853

- **Supplemental Notes:**

    - This document was sponsored by the U.S. Department of Transportation, University Transportation Centers Program. Supporting datasets available at: https://doi.org/10.7910/DVN/YOTY6A
- **Corporate Authors:**

Pacific Northwest Transportation Consortium
University of Washington
More Hall Room 112
Seattle, WA United States 98195-2700

University of Washington, Seattle
Department of Computer Science and Engineering
Seattle, WA United States 98195

Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC United States 20590

- **Authors:**
    - Froehlich, Jon E
- **Publication Date:** 2021-6-9

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Figures; Photos; References; Tables;
- **Pagination:** 44p

Subject/Index Terms

- **TRT Terms:** Accessibility; Crowdsourcing; Data collection; Digital maps; Machine learning; Sidewalks
- **Identifier Terms:** Google Street View
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01789205
- **Record Type:** Publication
- **Report/Paper Numbers:** 2019-S-UW-2
- **Contract Numbers:** 69A3551747110
- **Files:** UTC, TRIS, ATRI, USDOT
- **Created Date:** Nov 2 2021 10:45AM

FIFTEEN

A Novel Crowdsourcing Model for Micro-Mobility Ride-Sharing Systems

Substantial research is required to ensure that micro-mobility ride sharing provides a better fulfilment of user needs. This study proposes a novel crowdsourcing model for the ride-sharing system where light vehicles such as scooters and bikes are crowdsourced. The proposed model is expected to solve the problem of charging and maintaining a large number of light vehicles where these efforts will be the responsibility of the crowd of suppliers. The proposed model consists of three entities: suppliers, customers, and a management party responsible for receiving, renting, booking, and demand matching with offered resources. It can allow suppliers to define the location of their private e-scooters/e-bikes and the period of time they are available for rent. Using a dataset of over 9 million e-scooter trips in Austin, Texas, the authors ran an agent-based simulation six times using three maximum battery ranges (i.e., 35, 45, and 60 km) and different numbers of e-scooters (e.g., 50 and 100) at each origin. Computational results show that the proposed model is promising and might be advantageous to shift the charging and maintenance efforts to a crowd of suppliers.

- **Record URL:**
  http://dx.doi.org/10.3390/s21144636

- **Record URL:**
  https://www.mdpi.com/1424-8220/21/14/4636

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/14248220
- **Supplemental Notes:**

- **Authors:**
  - Elhenawy, Mohammed
  - Komol, Mostafizur R
  - Masoud, Mahmoud
  - Liu, Shi Qiang
  - Ashqar, Huthaifa I
  - Almannaa, Mohammed Hamad
  - Rakha, Hesham A
  - Rakotonirainy, Andry
- **Publication Date:** 2021-7

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; References;
- **Pagination:** 4636
- **Serial:**
  - Sensors
  - Volume: 21
  - Issue Number: 14

- Publisher: MDPI AG
    - ISSN: 1424-8220
    - Serial URL: http://www.mdpi.com/journal/sensors
- **Publication flags:** Open Access (libre)

Subject/Index Terms

- **TRT Terms:** Business models; Computer algorithms; Crowdsourcing; Ridesourcing; Shared mobility; Vehicle sharing
- **Geographic Terms:** Austin (Texas)
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting; Public Transportation; Vehicles and Equipment;

Filing Info

- **Accession Number:** 01779549
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Aug 3 2021 4:10PM

SIXTEEN

Every Day Counts: Innovation for a Nation on the Move

Every Day Counts (EDC) is the Federal Highway Administration's (FHWA's) program to advance a culture of innovation in the transportation community in partnership with public and private stakeholders. Through this State-based effort, FHWA coordinates rapid deployment of proven strategies and technologies to shorten the project delivery process, enhance roadway safety, reduce traffic congestion, and integrate automation. This report summarizes the December 2020 status of deployment for the 10 innovations in the fifth round of EDC. The innovations are: advanced geotechnical methods in exploration; collaborative hydraulics: advancing to the next generation of engineering (CHANGE); crowdsourcing for operations; project bundling; focus on reducing rural roadway departures; safe transportation for every pedestrian (STEP); unmanned aerial systems (UAS); value capture; virtual public involvement; and weather-responsive management strategies. The report is intended to be a resource for transportation stakeholders as they implement their innovation deployment plans and to encourage innovation in managing highway project delivery to better serve the Nation.

- **Record URL:**
  https://www.fhwa.dot.gov/innovation/everydaycounts/reports/edc5_finalreport.pdf

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/56287

- **Corporate Authors:**
  Federal Highway Administration

1200 New Jersey Avenue, SE
Washington, DC United States 20590

- **Publication Date:** 2021-4

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Appendices; Figures; Maps; Photos; Tables;
- **Pagination:** 28p

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Drones; Geotechnical engineering; Hydraulics; Implementation; Public participation; Technological innovations; Traffic safety; Value capture
- **Identifier Terms:** Every Day Counts; U.S. Federal Highway Administration
- **Subject Areas:** Highways; Operations and Traffic Management; Safety and Human Factors;

Filing Info

- **Accession Number:** 01771685
- **Record Type:** Publication
- **Report/Paper Numbers:** FHWA-21-CAI-018, EDC-5
- **Files:** NTL, TRIS, ATRI, USDOT
- **Created Date:** May 6 2021 12:58PM

SEVENTEEN

Estimating Sectional Volume of Travelers Based on Mobile Phone Data

The sectional volume of travelers refers to the number of travelers crossing a section boundary (e.g., river, mountain, railway line, etc.) within a certain time period. Mobile phone data provides continuous and large-scale mobility pattern information without compromising the comprehensiveness of travel modes. The authors propose a three-stage framework to estimate the sectional volume of travelers using the base station trajectory from massive mobile phone data. In the first two stages, the spatial and temporal uncertainties of trajectories are explicitly addressed by a hybrid filtering algorithm and a cell-to-cell trajectory inference algorithm, respectively. In the third stage, the sectional volume of travelers is estimated using aggregated trajectories. The proposed framework is validated using a sampled dataset with annotated ground truth and a city-scale dataset. The results show that the proposed framework is effective in dealing with spatial and temporal uncertainties of trajectories. The sectional volume estimation method performs stably with a low average error rate and is applicable to section boundaries of different scales.

- **Record URL:**
  https://doi.org/10.1061/JTEPBS.0000429

- **Availability:**
    - Find a library where document is available. Order URL: http://worldcat.org/issn/24732907
- **Supplemental Notes:**
    -
- **Authors:**
    - Liu, Zhichen
    - Fu, Xiao
    - Liu, Yang
    - Tong, Weiping
    - Liu, Zhiyuan
- **Publication Date:** 2020-10

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** 04020110
- **Serial:**
    - Journal of Transportation Engineering, Part A: Systems
    - Volume: 146
    - Issue Number: 10
    - Publisher: American Society of Civil Engineers
    - ISSN: 2473-2907
    - EISSN: 2473-2893
    - Serial URL: http://ascelibrary.org/journal/jtepbs

Subject/Index Terms

- **TRT Terms:** Algorithms; Cellular telephones; Crowdsourcing; Data mining; Mobility; Traffic volume; Trajectory; Travel patterns; Travelers
- **Subject Areas:** Data and Information Technology; Highways; Planning and Forecasting;

Filing Info

- **Accession Number:** 01748475
- **Record Type:** Publication
- **Files:** TRIS, ASCE
- **Created Date:** Jul 20 2020 3:06PM

EIGHTEEN

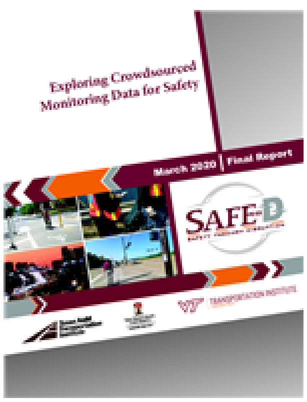
Exploring Crowdsourced Monitoring Data for Safety

This project included four distinct but related exploratory studies of data sources that could improve roadway safety analysis. The first effort evaluated passively gathered crowdsourced bicyclist activity data from StreetLight Data and found promising correlations (R2 of 62% and 69% for monthly weekday and weekend daily averages) when the StreetLight data were compared to bicyclist counts from 32 locations in eight Texas cities, and even better correlation (R2 of 94%) when compared with countywide Strava data expanded to represent total bicycling activity. The second effort evaluated the pedestrian counting accuracy of the Miovision system and found 15% error for daytime and 24% error for nighttime conditions. The third effort used INRIX trip trace data to determine origin-destination patterns and developed 40 decision rules to define the origin-destination patterns. The fourth effort analyzed crowdsourced Waze data (i.e., traffic incidents) and found it to be a reliable alternative to observed and predicted crashes, with the ability to identify high-risk locations: 77% of high-risk locations identified from police-reported crashes were also identified as high-risk in Waze data. The researchers propose a method to treat the redundant Waze reports and to match the unique Waze incidents with police crash reports.

- **Dataset URL:**
  https://doi.org/10.15787/VTT1/OBV82F

- **Dataset URL:**
  https://doi.org/10.15787/VTT1/351GZJ

- **Dataset URL:**
  https://doi.org/10.15787/VTT1/81SKJW

- **Record URL:**
  https://www.vtti.vt.edu/utc/safe-d/wp-content/uploads/2020/04/TTI-Student-05_Final-Research-Report_Final.pdf
- 

- **Summary URL:**
  https://safed.vtti.vt.edu/projects/exploring-crowdsourced-monitoring-data-for-safety/

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/50717

- **Dataset URL:**
  https://rosap.ntl.bts.gov/view/dot/53593

- **Dataset URL:**
  https://rosap.ntl.bts.gov/view/dot/53577

- **Dataset URL:**
  https://rosap.ntl.bts.gov/view/dot/54626

- **Supplemental Notes:**
  - This document was sponsored by the U.S. Department of Transportation, University Transportation Centers Program. Supporting datasets available at: https://doi.org/10.15787/VTT1/OBV82F; https://doi.org/10.15787/VTT1/351GZJ; https://doi.org/10.15787/VTT1/81SKJW; https://rosap.ntl.bts.gov/view/dot/53593; https://rosap.ntl.bts.gov/view/dot/53577; https://rosap.ntl.bts.gov/view/dot/54626

- **Corporate Authors:**

  Safety through Disruption University Transportation Center (Safe-D)
  Virginia Tech Transportation Institute
  Blacksburg, VA United States 24060

  Texas A&M Transportation Institute
  Texas A&M University System
  3135 TAMU
  College Station, TX United States 77843-3135

  Office of the Assistant Secretary for Research and Technology
  University Transportation Centers Program
  Department of Transportation
  Washington, DC United States 20590

- **Authors:**
  - Turner, Shawn
  - Martin, Michael
  - Griffin, Greg
  - Le, Minh
  - Das, Subasish
  - Wang, Ruihong
  - Dadashova, Bahar
  - Li, Xiao
- **Publication Date:** 2020-3

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Research Report
- **Features:** Appendices; Figures; Maps; References; Tables;
- **Pagination:** 45p

Subject/Index Terms

- **TRT Terms:** Accuracy; Bicycle counts; Crash data; Crash risk forecasting; Crowdsourcing; Data analysis; High risk locations; Origin and destination; Pedestrian counts; Travel patterns
- **Geographic Terms:** Texas
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01739009
- **Record Type:** Publication
- **Report/Paper Numbers:** TTI-Student-05
- **Contract Numbers:** 69A3551747115/TTI-Student-05
- **Files:** UTC, NTL, TRIS, ATRI, USDOT
- **Created Date:** Apr 30 2020 12:36PM

NINETEEN

Emerging data for pedestrian and bicycle monitoring: Sources and applications

Growing attention on the benefits of non-motorized travel has increased the demand for accurate and timely pedestrian and bicycle travel data. Advancements in technologies and the proliferation of smartphones have created new data sources that can help eliminate limitations related to small sample size and infrequent updates due to limited resources. This study reviews the emerging data sources and their current use, focusing on non-motorized travel monitoring. In this study, the emerging data are categorized into mode-unspecified and mode-specified data based on whether the mode used can be detected with no or little effort. While mode-unspecified data are collected without sorting out non-motorized travelers, mode-specified data at least know who (which mode) is being monitored. So far, commercial vendors provide a vast volume of mode-unspecified data, but their products have been mainly used for motorized trips or are in initial stages of development. Meanwhile, readily available data sources and their applications are more concentrated on mode-specified data, which have enabled varying non-motorized travel studies—including travel pattern identification, route-choice modeling, crash/air pollution exposure estimation, and new facility provision evaluation—but are mostly focused on bicycling. Despite the potential of emerging data, their use also has several challenges, such as limited mode inference, sample bias, and lack of detailed trip/traveler information due to privacy issues. More efforts are needed, such as improving data accuracy and developing robust data fusion techniques, to be able to fully utilize the emerging data sources.

- **Record URL:**
  https://doi.org/10.1016/j.trip.2020.100095

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S2590198220300063

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/53624

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/25901982
  -
- **Supplemental Notes:**
  - Published by Elsevier Ltd. Abstract reprinted with permission of Elsevier.
- **Authors:**
  - Lee, Kyuhyun
  - Sener, Ipek Nese
- **Publication Date:** 2020-3

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; References;
- **Serial:**
  - Transportation Research Interdisciplinary Perspectives
  - Volume: 4
  - Issue Number: 0
  - Publisher: Elsevier
  - ISSN: 2590-1982
  - Serial URL: https://www.journals.elsevier.com/transportation-research-interdisciplinary-perspectives

- **Publication flags:** Open Access (libre)
- You have access to this record's full text

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Cyclists; Data collection; Monitoring; Pedestrians; Smartphones
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01736304
- **Record Type:** Publication
- **Files:** NTL, TRIS
- **Created Date:** Mar 2 2020 3:04PM

TWENTY

Guidelines for Using StreetLight Data for Planning Tasks

The Virginia Department of Transportation (VDOT) has purchased a subscription to the StreetLight (SL) Data products that mainly offer origin-destination (OD) related metrics through crowdsourcing data. Users can manipulate a data source like this to quickly estimate origin-destination trip tables. Nonetheless, the SL metrics heavily rely on the data points sampled from smartphone applications and global positioning services (GPS) devices, which may be subject to potential bias and coverage issues. In particular, the quality of the SL metrics in relation to meeting the needs of various VDOT work tasks is not clear. Guidelines on the use of the SL metrics are of interest to VDOT. This study aimed to help VDOT understand the performance of the SL metrics in different application contexts. Specifically, existing studies that examined the potential of SL metrics have been reviewed and summarized. In addition, the experiences, comments, and concerns of existing users and potential users have been collected through online surveys. The developed surveys were primarily distributed to VDOT engineers and planners as well as other professionals in planning organizations and consultants in Virginia. Their typical applications of the SL metrics have been identified and feedback has been used to guide and inform the design of the guidelines. To support the development of a set of guidelines, the quality of the SL metrics has been independently evaluated with six testing scenarios covering annual average daily traffic (AADT), OD trips, traffic flow on road links, turning movements at intersections, and truck traffic. The research team has sought ground-truth data from different sources such as continuous count stations, toll transaction data, VDOT's internal traffic estimations, etc. Several methods were used to perform the comparison between the benchmark data and the corresponding SL metrics. The evaluation results were mixed. The latest SL AADT estimates showed relatively small absolute percentage errors, whereas using the SL metrics to estimate OD trips, traffic counts on roadway segments and at intersections, and truck traffic did not show a relatively low and stable error rate. Large percentage errors were often found to be associated with lower volume levels estimated based on the SL metrics. In addition, using the SL metrics from individual periods as the input for estimating these traffic measures resulted in larger errors. Instead, the aggregation of data from multi-periods helped reduce the errors, especially for low volume conditions. Depending on project purposes, the aggregation can be based on metrics of multiple days, weeks, or months. The results from the literature review, surveys, and independent evaluations were synthesized to help develop the guidelines for using SL data products. The guidelines focused on five main aspects: (1) a summary for using SL data for typical planning work tasks; (2) general guidance for data extraction and preparation; (3) using the SL metrics in typical application scenarios; (4) quality issues and calibration of the SL metrics; and (5) techniques and tools for working with the SL metrics. The developed guidelines were accompanied with illustrative examples to allow users to go through the given use cases. Based on the results, the study recommends that VDOT's Transportation and Mobility Planning Division (TMPD) should encourage and support the use of the guidelines in projects involving SL data, and that TMPD should adopt a checklist (table) for reporting performance, calibration efforts, and benchmark data involved in projects that use the SL metrics.

- **Record URL:**

http://www.virginiadot.org/vtrc/main/online_reports/pdf/20-R23.pdf

- 

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/55501

- **Corporate Authors:**
  Old Dominion University
  Department of Civil and Environmental Engineering, 135 Kaufman Hall
  Norfolk, VA United States 23529

  Virginia Transportation Research Council
  530 Edgemont Road
  Charlottesville, VA United States 22903

  Virginia Department of Transportation
  1401 East Broad Street
  Richmond, VA United States 23219

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590
- **Authors:**
  - Yang, Hong
  - Cetin, Mecit
  - Ma, Qingyu
- **Publication Date:** 2020-3

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Appendices; Figures; Maps; References; Tables;
- **Pagination:** 122p

Subject/Index Terms

- **TRT Terms:** Annual average daily traffic; Crowdsourcing; Data files; Data quality; Global Positioning System; Guidelines; Origin and destination; Traffic flow; Transportation planning; Trip tables; Truck traffic; Turning traffic
- **Identifier Terms:** Virginia Department of Transportation
- **Subject Areas:** Data and Information Technology; Highways; Operations and Traffic Management; Planning and Forecasting;

Filing Info

- **Accession Number:** 01735841
- **Record Type:** Publication
- **Report/Paper Numbers:** FHWA/VTRC 20-R23, VTRC 20-R23
- **Contract Numbers:** 112110
- **Files:** NTL, TRIS, ATRI, USDOT, STATEDOT
- **Created Date:** Mar 31 2020 4:27PM

TWENTY ONE

Crowdsourcing for Mode Shift: An Empirical Evidence of its Success among College Students

In this paper, a crowdsourcing-based perception framework is developed to estimate any changes in mode-shift behavior of college students. An empirical experiment was conducted with a sample of 30 participants from California State University, Long Beach, spanning over two phases, Phase I followed by Phase II. Participants used one of the five modes transit bus, bicycling, walking, car and carpool to arrive at the university campus. During Phase I, a control was created by individually acquiring the mode choice of participants and their numeric value of perception of each specific mode with the identity of a participant being kept anonymous to other participants throughout this research. The participants in Phase II were asked to post their comments publicly anonymously on modes on a "Twitter" address used for this study each time they arrived at the campus, thus utilizing a crowdsourcing platform to observe mode choice of other participants without knowing their identity. Data compilation showed an overall shift of users from private car to other modes of transportation.

- **Record URL:**
  https://doi.org/10.1016/j.trpro.2020.08.173

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S2352146520305901

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/23521465

- **Supplemental Notes:**

- **Authors:**
  - Chandra, Shailesh
  - Naik, R Thirumaleswara
  - Jimenez, Jose
- **Conference:**
  - World Conference on Transport Research (WCTR) 2019
  - Location: Mumbai , India
  - Date: 2019-5-26 to 2019-5-31
- **Publication Date:** 2020

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; References; Tables;
- **Pagination:** pp 1430-1434
- **Serial:**
  - Transportation Research Procedia
  - Volume: 48
  - Issue Number: 0
  - Publisher: Elsevier
  - ISSN: 2352-1465
  - Serial URL: http://www.sciencedirect.com/science/journal/23521465/
- **Publication flags:** Open Access (libre)
- You have access to this record's full text

Subject/Index Terms

- **TRT Terms:** College students; Crowdsourcing; Modal shift; Perception; Travel behavior
- **Identifier Terms:** California State University; Twitter
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Public Transportation;

Filing Info

- **Accession Number:** 01753857
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Sep 17 2020 3:28PM

TWENTY TWO

A simple crowdsourced delay-based traffic signal control

Current transportation management systems rely on physical sensors that use traffic volume and queue-lengths. These physical sensors incur significant capital and maintenance costs. The ubiquity of mobile devices has made possible access to accurate and cheap traffic delay data. However, current traffic signal control algorithms do not accommodate the use of such data. In this paper, the authors propose a novel parsimonious model to utilize real-time crowdsourced delay data for traffic signal management. The authors demonstrate the versatility and effectiveness of the data and the proposed model on seven different intersections across three cities and two countries. This signal system provides an opportunity to leapfrog from physical sensors to low-cost, reliable crowdsourced data.

- **Record URL:**
  - https://doi.org/10.1371/journal.pone.0230598

- **Supplemental Notes:**
  -

- **Authors:**
  - Dixit, Vinayak
  - Nair, Divya Jayakumar
  - Chand, Sai
  - Levin, Michael W

- **Publication Date:** 2020

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** e0230598
- **Serial:**
  - PLoS One
  - Volume: 15
  - Issue Number: 4
  - Publisher: Public Library of Science
  - EISSN: 1932-6203
  - Serial URL: https://journals.plos.org/plosone/
- **Publication flags:** Open Access (libre)

Subject/Index Terms

- **TRT Terms:** Alternatives analysis; Cities; Costs; Crowdsourcing; Data sharing; Mobile telephones; Sensors; Signalized intersections; Traffic delays; Traffic signal control systems
- **Subject Areas:** Data and Information Technology; Highways; Operations and Traffic Management;

Filing Info

- **Accession Number:** 01741853
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Apr 14 2020 9:59AM

TWENTY THREE

Exploring Crowdsourced Monitoring Data for Safety - Evaluation of Miovision Pedestrian Count Data (TTI-Student-05) [supporting dataset]

Project Description: The data represent one week of selected hourly weekday and weekend pedestrian counts at two intersections in Austin, Texas. The pedestrian counts were produced from manually reducing video files from Miovision's TrafficLink Multimodal Detection and Counts system. Eighty hours of video were gathered at each intersection between June 18 and July 14, 2019. However, only 40 hours at each intersection were reduced and evaluated. The manual counts were compared to Miovision's count data across different combinations of lighting conditions and pedestrian volumes. Overall, Miovision system performed fairly well with accuracy results of 15% error for daytime and 24% for nighttime for the combined intersection legs. Data Scope: Selected hourly weekday and weekend pedestrian counts. 2 unique count locations. 240 hourly observations from 2 student workers and the Miovision system.

- **Dataset URL:**
  http://doi.org/10.15787/VTT1/351GZJ

- **Dataset URL:**
  https://rosap.ntl.bts.gov/view/dot/54626

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/50717

- **Supplemental Notes:**
  - The dataset supports report: Exploring Crowdsourced Monitoring Data for Safety, available at the URL above. This document was sponsored by the U.S. Department of Transportation, University Transportation Centers Program.
- **Corporate Authors:**

  Safety Through Disruption University Transportation Center (Safe-D)
  Texas A&M Transportation Institute
  College Station, TX United States

  Office of the Assistant Secretary for Research and Technology
  University Transportation Centers Program
  Department of Transportation
  Washington, DC United States 20590

- **Authors:**
  - Le, Minh
- **Publication Date:** 2019-11-5

Language English

Media Info

- **Media Type:** Dataset
- **Dataset:** Version: 1.0 Integrity Hash:
- **Dataset publisher:**
  Virginia Tech Transportation Institute Dataverse

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Intersections; Pedestrian counts; Periods of the day; Traffic data; Traffic volume; Video
- **Geographic Terms:** Austin (Texas)
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01775946
- **Record Type:** Publication
- **Contract Numbers:** 69A3551747115/TTI-Student-05
- **Files:** UTC, NTL, TRIS, ATRI, USDOT
- **Created Date:** Jun 11 2021 4:45PM

TWENTY FOUR

Combining Crowdsourcing and Machine Learning to Collect Sidewalk Accessibility Data at Scale

Sidewalks significantly impact the mobility and quality of life of millions of Americans. In the proposal, the research team described new, scalable methods for collecting data on sidewalk accessibility using machine learning, crowdsourcing, and online map imagery as well as new interactive visualizations aimed at providing novel insights into urban accessibility. As with the team's prior research, the team will work closely with key stakeholders, including local governments and transit departments, mobility-impaired individuals and caretakers, and walkability advocates to help shape and evaluate the design of the team's tools.

https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/47853/Froehlich_ProjectSidewalk_FinalReport.pdf?sequence=1&isAllowed=y

Language English

Project

- **Status:** Active
- **Funding:** $100000
- **Contract Numbers:**
  69 A3551747110
- **Sponsor Organizations:**
  United States Department of Transportation - FHWA - LTAP
  1200 New Jersey Avenue, SE
  Washington, DC 20590

Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC United States 20590

- **Managing Organizations:**
  University of Washington, Seattle
  433 Brooklyn Ave. NE
  Box 359472
  Seattle, WA United States 98195-9472

- **Performing Organizations:**

  University of Washington, Seattle
  433 Brooklyn Ave. NE
  Box 359472
  Seattle, WA United States 98195-9472

- **Principal Investigators:**
  Froehlich, Jon
- **Start Date:** 20190916
- **Expected Completion Date:** 20210915
- **Actual Completion Date:** 0
- **USDOT Program:** University Transportation Centers Program

Subject/Index Terms

- **TRT Terms:** Accessibility; Crowdsourcing; Data collection; Digital maps; Machine learning; Sidewalks; Stakeholders; Urban areas
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01723937
- **Record Type:** Research project
- **Source Agency:** Pacific Northwest Transportation Consortium
- **Contract Numbers:** 69 A3551747110
- **Files:** UTC, RIP
- **Created Date:** Nov 27 2019 7:04PM

TWENTY FIVE

Investigating the Underreporting of Pedestrian and Bicycle Crashes in and Around University Campuses − A Crowdsourcing Approach

In this paper, the non-motorized traffic safety concerns in and around three university campuses are evaluated by comparing police-reported crash data with traffic safety information sourced from the

campus communities themselves. The crowdsourced traffic safety data comprise of both self-reported crashes as well as perceived hazardous locations. The results of the crash data analysis reveal that police-reported crashes underrepresent non-motorized safety concerns in and around the campus regions. The spatial distribution of police-reported crashes shows that police-reported crashes are predominantly unavailable inside the main campus areas, and the off-campus crashes over-represent automobile involvement. In comparison, the self-reported crash results report a wide variety of off-campus collisions not involving automobiles, while also highlighting the issue of high crash concentrations along campus boundaries. An assessment of the perceived hazardous locations (PHLs) reveals that high concentrations of such observations at/near a given location have statistically significant association with both survey-reported crashes as well as future police-reported crashes. Moreover, the results indicate the presence of a saturation point in the relationship between crashes and PHLs wherein beyond a certain limit, an increasing number of traffic safety concerns may not necessarily correlate with a proportional increase in the number of crashes. These findings suggest that augmenting the existing knowledge of traffic safety through crowdsourcing techniques can potentially help in better estimating both existing as well as emerging traffic safety concerns.

- **Record URL:**
  https://doi.org/10.1016/j.aap.2017.08.014

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S0001457517302920

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/00014575
- **Supplemental Notes:**

- **Authors:**
  - Medury, Aditya
  - Grembek, Offer
  - Loukaitou-Sideris, Anastasia
  - Shafizadeh, Kevan
- **Publication Date:** 2019-9

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** pp 99-107
- **Serial:**
  - Accident Analysis & Prevention
  - Volume: 130
  - Issue Number: 0
  - Publisher: Elsevier

- ISSN: 0001-4575
- Serial URL: http://www.sciencedirect.com/science/journal/00014575

Subject/Index Terms

- **TRT Terms:** Bicycle crashes; Campuses; Crash reports; Crowdsourcing; High risk locations; Nonmotorized transportation; Pedestrian vehicle crashes; Traffic safety
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01717020
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Sep 4 2019 3:05PM

TWENTY SIX

Using Crowd Sourcing to Locate and Characterize Conflicts for Vulnerable Modes

Most agencies and decision-makers rely on crash and crash severity (property damage only, injury or fatality) data to assess transportation safety; however, in the context of public health where perceptions of safety may influence the willingness to adopt active transportation modes (e.g. bicycling and walking), pedestrian-motor vehicle and other similar conflicts types may define a better performance measure for safety assessment. In the field of transportation safety, an absolute conflict occurs when two parties' paths cross and one of the parties must undertake an evasive maneuver (e.g. change direction or stop) to avoid a crash. Other less severe conflicts where paths cross but no evasive maneuver is required may also impact public perceptions of safety especially for vulnerable modes. Most of the existing literature focuses on vehicle conflicts. While in the past several years, more research has investigated bicycle and pedestrian conflicts, most of this has focused on the intersection environment. A comprehensive analysis of conflicts appears critical. The major objective of this study is two-fold: 1) Development of an innovative and cost effective conflict data collection technique to better understand the conflicts (and their severity) involving vulnerable road users (e.g. bicycle/pedestrian, bicycle/motor vehicle, and pedestrian/motor vehicle) and their severity. 2) Test the effectiveness and practicality of the approach taken and its associated crowd sourced data collection. In an endeavor to undertake these objectives, the researchers developed an android-based crowd-sourced data collection app. The crowd-source data collected using the app is compared with traditional fatality data for hot spot analysis. At the end, the app users provide feedback about the overall competency of the app interface and the performance of its features to the app developers. If widely adopted, the app will enable communities to create their own data collection efforts to identify dangerous sites within their neighborhoods. Agencies will have a valuable data source at low-cost to help inform their decision making related to bicycle and pedestrian education, encouragement, enforcement, programs, policies, and infrastructure design and planning.

- **Record URL:**
  https://doi.org/10.1016/j.aap.2019.03.014

- **Record URL:**

http://www.sciencedirect.com/science/article/pii/S0001457519304749

- **Availability:**
    - Find a library where document is available. Order URL: http://worldcat.org/issn/00014575
- **Supplemental Notes:**
    -
- **Authors:**
    - Rahman, Ziaur
    - Mattingly, Stephen P
    - Kawadgave, Rahul
    - Nostikasari, Dian
    - Roeglin, Nicole
    - Casey, Colleen
    - Johnson, Taylor
- **Publication Date:** 2019-7

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** pp 32-39
- **Serial:**
    - Accident Analysis & Prevention
    - Volume: 128
    - Issue Number: 0
    - Publisher: Elsevier
    - ISSN: 0001-4575
    - Serial URL: http://www.sciencedirect.com/science/journal/00014575

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Data collection; Mobile applications; Pedestrian vehicle crashes; Traffic conflicts; Vulnerable road users
- **Candidate Terms:** Bicycle vehicle interface
- **Subject Areas:** Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01707840
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** May 21 2019 3:09PM

TWENTY SEVEN

Identifying commonly used and potentially unsafe transit transfers with crowdsourcing

Public transit is an important contributor to sustainable transportation as well as a public service that makes necessary travel possible for many. Poor transit transfers can lead to both a real and perceived reduction in convenience and safety, especially for people with disabilities. Poor transfers can expose riders to inclement weather and crime, and they can reduce transit ridership by motivating riders who have the option of driving or using paratransit to elect a more expensive and inefficient travel mode. Unfortunately, knowledge about inconvenient, missed, and unsafe transit transfers is sparse and incomplete. The authors show that crowdsourced public transit ridership data, which is more scalable than conducting traditional surveys, can be used to analyze transit transfers. The Tiramisu Transit app merges open transit data with information contributed by users about which trips they take. The authors use Tiramisu data to do origin-destination analysis and identify connecting trips to create a better understanding of where and when poor transfers are occurring in the Pittsburgh region. The authors merge the results with data from other open public data sources, including crime data, to create a data resource that can be used for planning and identification of locations where bus shelters and other infrastructure improvements may facilitate safer and more comfortable waits and more accessible transfers. The authors use generalizable methods to ensure broader value to both science and practitioners. They present a case study of the Pittsburgh region, in which they identified and characterized 338 transfers from 142 users. The authors found that 66.6% of transfers were within 0.4 km (0.25 mi.) and 44.1% of transfers were less than 10 min. They identified the geographical distribution of transfers and found several highly-utilized transfer locations that were not identified by the Port Authority of Allegheny County as recommended transfer points, and so might need more planning attention. The authors cross-referenced transfer location and wait time data with crime levels to provide additional planning insight.

- **Record URL:**
  - https://doi.org/10.1016/j.tra.2019.02.005

- **Record URL:**
  - http://www.sciencedirect.com/science/article/pii/S0965856417314040

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/09658564
- **Supplemental Notes:**
  -
- **Authors:**
  - Traut, Elizabeth J
  - Steinfeld, Aaron
- **Publication Date:** 2019-4

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** pp 99-111

- **Serial:**
    - [Transportation Research Part A: Policy and Practice](#)
    - Volume: 122
    - Issue Number: 0
    - Publisher: Elsevier
    - ISSN: 0965-8564
    - Serial URL: http://www.sciencedirect.com/science/journal/09658564

Subject/Index Terms

- **TRT Terms:** Case studies; Crowdsourcing; Data mining; Origin and destination; Persons with disabilities; Ridership; Safety; Transfers
- **Geographic Terms:** Pittsburgh (Pennsylvania)
- **Subject Areas:** Data and Information Technology; Operations and Traffic Management; Public Transportation; Safety and Human Factors;

Filing Info

- **Accession Number:** 01700018
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Mar 5 2019 3:05PM

TWENTY EIGHT

Crowdsourcing solutions for supporting urban mobility

Recently, several urban crowdsourcing investigations and various experiments have been conducted with the aim of engaging citizens in order to produce information about their cities and their communities. This article reports on the results of a research based on a systematic analysis of the current literature on "urban crowdsourcing" and "citizensourcing" conducted by the authors (September 2017- September 2018) analyzing and discussing the applications that have been proposed and experimented to support urban mobility.

- **Record URL:**
  https://doi.org/10.1016/j.procs.2019.01.174

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S1877050919301826

- **Availability:**
    - Find a library where document is available. Order URL: http://worldcat.org/issn/18770509
- **Supplemental Notes:**

    -
- **Authors:**
    - Marzano, Gilberto
    - Lizut, Joanna
    - Siguencia, Luis Ocha
- **Conference:**
    - ICTE in Transportation and Logistics 2018 (ICTE 2018)
    - Location: Budapest , Hungary
    - Date: 2018-4-8 to 2018-4-10
- **Publication Date:** 2019

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; References;
- **Pagination:** pp 542-547
- **Serial:**
    - Procedia Computer Science
    - Volume: 149
    - Issue Number: 0
    - Publisher: Elsevier
    - ISSN: 1877-0509
    - Serial URL: http://www.sciencedirect.com/science/journal/18770509
- **Publication flags:**
  Open Access (libre)
- You have access to this record's full text

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Mobility; Urban transportation
- **Subject Areas:** Highways; Operations and Traffic Management; Planning and Forecasting;

Filing Info

- **Accession Number:** 01708866
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Mar 14 2019 3:16PM

TWENTY NINE

The Geography and Equity of Crowdsourced Public Participation for Active Transportation Planning

Transportation planners increasingly use new forms of online public participation alongside traditional in-person approaches, including crowdsourcing tools capable of encouraging geographically specific input. Digital involvement may be particularly valuable in exploring methods to plan at a megaregional scale.

Research is beginning to address digital inequalities, recognizing that broadband and smartphone access may restrict opportunities for disadvantaged groups. However, the geography and equity of participation remain pragmatic issues for practice and research. This paper reviews the geography and equity of the participation methods in Austin, Texas for active transportation (bicycling and pedestrian) through three approaches to co-produce informed plans: in-person meetings, public participation geographic information system (PPGIS), and an emerging smartphone platform that logs trips and encourages input on route quality. In addition to spatial analysis with standard deviational ellipses, we include qualitative case analysis to contextualize the geographic and equity implications of different participation approaches. Results show that both online techniques resulted in a larger geography for participation than in-person meetings, with the regional PPGIS covering the most area. However, review of the income levels in each area shows that use of the smartphone-based crowdsourcing platform was aligned with lowest-income areas. This study shows that online participation methods are not homogeneous regarding geography or equity. In some contexts, smartphone applications can help reach lower-income communities, even when compared with in-person meetings. Crowdsourcing tools can be valuable approaches to increase geography and equity of public participation in transportation planning.

- **Record URL:**
    https://doi.org/10.1177/0361198118823498
- **Availability:**
    - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Authors:**
    - Griffin, Greg P
    - Jiao, Junfeng
- **Publication Date:** 2019

Language English

Media Info

- **Media Type:** Print
- **Features:** Figures; References; Tables;
- **Serial:**
    - Transportation Research Record: Journal of the Transportation Research Board
    - Issue Number: 0
    - Publisher: Sage Publications, Incorporated
    - ISSN: 0361-1981
    - EISSN: 2169-4052
    - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Bicycle travel; Crowdsourcing; Equity (Justice); Geographic information systems; Geography; Pedestrians; Public participation; Transportation planning
- **Geographic Terms:** Austin (Texas)
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting; Policy; Society;

Filing Info

- **Accession Number:** 01690760

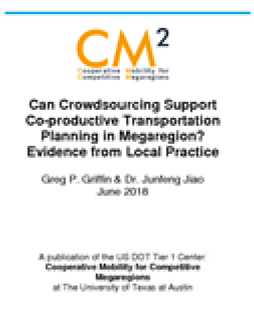
- **Record Type:** Publication
- **Report/Paper Numbers:** 19-00284
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Jan 3 2019 11:46AM

THIRTY

Can Crowdsourcing Support Co-productive Transportation Planning in Megaregion? Evidence from Local Practice

This project explores how different technologies (smart phone, social media, web scrapping) can facilitate co-productive transportation planning process in Megaregion. As case studies, GRAs explored bike-sharing planning in NYC, Chicago, Austin, San Antonio and Houston.

- **Record URL:**
  http://sites.utexas.edu/cm2/files/2018/06/Year-1-JiaoGriffin_MegaRegionCrowdsourcing.pdf



- **Summary URL:**
  https://library.ctr.utexas.edu/Presto/catalogid=36048

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/36260

- **Supplemental Notes:**
  - This document was sponsored by the U.S. Department of Transportation, University Transportation Centers Program. Cover date: June 2018.
- **Corporate Authors:**
  University of Texas, Austin
  Urban Information Lab
  School of Architecture
  Austin, TX United States

  Cooperative Mobility for Competitive Megaregions (CM2)
  University of Texas at Austin
  Austin, TX United States 78712

Office of the Assistant Secretary for Research and Technology
University Transportation Centers Program
Department of Transportation
Washington, DC United States 20590

- **Authors:**
  - Griffin, Greg P
  - Jiao, Junfeng
- **Publication Date:** 2018-7

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Maps; Photos; References; Tables;
- **Pagination:** 37p

Subject/Index Terms

- **TRT Terms:** Bicycle facilities; Bicycles; Case studies; Crowdsourcing; Location; Public participation; Transportation planning; Vehicle sharing
- **Uncontrolled Terms:** Megaregions
- **Geographic Terms:** Austin (Texas); Chicago (Illinois); Houston (Texas); New York (New York); San Antonio (Texas)
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Planning and Forecasting; Terminals and Facilities;

Filing Info

- **Accession Number:** 01675088
- **Record Type:** Publication
- **Report/Paper Numbers:** USDOT/69A3551747135, CM2-2017-2
- **Files:** UTC, NTL, TRIS, ATRI, USDOT
- **Created Date:** Jun 29 2018 9:52AM

THIRTY ONE

Crowdsourcing for Active Transportation

There is a new approach to collaboration in motion in the field of active transportation planning: one that harnesses many of the quickly evolving technologies changing how planners collect information and communicate. It's called "crowdsourcing," which can be described as the process of obtaining information, insight, and knowledge from user-generated data provided though web and mobile applications. Crowdsourcing has already taken many different forms and served a wide variety of transportation planning purposes, from identifying new bike share station locations and collecting personal travel data using GPS to mining data from personal fitness apps for travel patterns. Access to high-quality data in greater quantities and at finer spatial resolutions, as well as new capabilities for direct communication with community members, offers important new options for listening to bicyclists and pedestrians and working with them to better understand their relationships with the built environment, their travel decisions, and their needs.

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/oclc/614107147
- **Supplemental Notes:**
  - Abstract reprinted with permission from the Institute of Transportation Engineers.
- **Authors:**
  - Smith, Amy
- **Publication Date:** 2015-5

Language English

Media Info

- **Media Type:** Print
- **Features:** Figures; Maps; Photos; References;
- **Pagination:** pp 30-35
- **Serial:**
  - ITE Journal
  - Volume: 85
  - Issue Number: 5
  - Publisher: Institute of Transportation Engineers (ITE)
  - ISSN: 0162-8178
  - Serial URL: https://www.ite.org/publications/ite-journal/

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Data collection; Outreach; Public participation; Technological innovations; Transportation planning
- **Subject Areas:** Data and Information Technology; Highways; Pedestrians and Bicyclists; Planning and Forecasting; I72: Traffic and Transport Planning;

Filing Info

- **Accession Number:** 01577323
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Sep 4 2015 1:48PM

THIRTY TWO

Pilot Trial of a Transport Crowdsourcing Smartphone App

In this project a smartphone app was developed to collect real time information on journey quality and value of time during journey. A two-week pilot trial of the app was conducted to determine the feasibility and validity of obtaining information from participants at different journey stages. In addition, the results of literature reviews on crowdsourcing, social media, value of time, and mode detection using smartphone data are discussed.

- **Record URL:**
  http://trl.co.uk/reports-publications/trl-reports/report/?reportid=6985

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/isbn/9781910377147
- **Corporate Authors:**
  Transport Research Laboratory
  Crowthorne House, Nine Mile Ride
  Wokingham, Berkshire United Kingdom RG40 3GA
- **Authors:**
  - Hopkin, J
  - Ball, S D
  - Hannay, A
  - Hutchins, R
  - Palmer, D
  - Rahman, S
  - Naberezhnykh, D
  - Longhi, D
- **Publication Date:** 2014-10-30

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Figures; References; Tables;
- **Pagination:** 24p

Subject/Index Terms

- **TRT Terms:** Crowdsourcing; Feasibility analysis; Literature reviews; Public transit; Quality of service; Real time information; Smartphones; Travel time
- **Subject Areas:** Data and Information Technology; Transportation (General); I72: Traffic and Transport Planning;

Filing Info

- **Accession Number:** 01544642
- **Record Type:** Publication
- **ISBN:** 9781910377147
- **Contract Numbers:** PPR719
- **Files:** TRIS
- **Created Date:** Nov 11 2014 2:59PM

THIRTY THREE

Crowdsourcing Accessibility Related Information From POI-Destinations in Finland

The use of Social media is growing fast. The objective of the SosPromo project was to investigate the possibilities of Social Media tools and processes as a rapid way of information collection and sharing to enable and support the mobility of the selected user group, people with some movement reducing factor. The actual implementation in SosPromo-project focused on crowdsourcing accessibility information from Points Of Interest, POIs (e.g. restaurants, cafes, shops) especially in the Helsinki capital region in Finland. This information can be used for personalised journey planning especially with disabled or elderly users or persons with temporary mobility restrictions, e.g. when moving in a city with a small child in a push chair. The introduced SosPromo concept enables further information sharing between different user groups.

- **Record URL:**
  http://itswc.confex.com/itswc/WC2011/webprogram/Paper1771.html
- **Availability:**
  - Find a library where document is available. Order URL: http://itswc.confex.com/itswc/WC2011/webprogram/start.html
- **Supplemental Notes:**
  - Abstract reprinted with permission from Intelligent Transportation Society of America. Index title: Crowd Sourcing Accessibility Related Information From POI-Destinations in Finland.
- **Corporate Authors:**
  ITS America
  1100 17th Street, NW, 12th Floor
  Washington, DC United States 20036
- **Authors:**

  - Tarkiainen, Mikko
  - Back, Asta
  - Hulkkonen, Juha
  - Konkkola, Kalle

- **Conference:**
  - 18th ITS World Congress
  - Location: Orlando Florida, United States
  - Date: 2011-10-16 to 2011-10-20
- **Publication Date:** 2011

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Maps; References;
- **Pagination:** 10p
- **Monograph Title:** 18th ITS World Congress, Orlando, 2011. Proceedings

Subject/Index Terms

- **TRT Terms:** Accessibility; Aged; Crowdsourcing; Information dissemination; Mobility; Persons with disabilities; Social media; Transportation disadvantaged persons; Traveler information and communication systems
- **Uncontrolled Terms:** Points of interest
- **Geographic Terms:** Helsinki (Finland)
- **Subject Areas:** Data and Information Technology; Passenger Transportation; Pedestrians and Bicyclists; Society; I72: Traffic and Transport Planning;

Filing Info

- **Accession Number:** 01485095
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Jun 26 2013 10:57AM

THIRTY FOUR

Predicting bicycling and walking traffic using street view imagery and destination data

Few studies predict spatial patterns of bicycling and walking across multiple cities using street-level data. This study aims to model bicycle and pedestrian traffic at 4145 count locations across 20 U.S. cities using new micro-scale variables: (1) destinations from Google Point of Interest data (e.g., restaurants, schools) and (2) pixel classification from Google Street View imagery (e.g., sidewalks, trees, streetlights). The authors applied machine learning algorithms to assess how well street-level variables predict bicycling and walking rates. Adding street-level variables improved out-of-sample prediction accuracy of bicycling and walking activities. The authors also found that street-level variables (10-fold CV R^2: 0.82–0.88) may be a useful alternative to Census data (0.85–0.88). Macro-scale factors (e.g., zoning) captured by Census data and micro-scale factors (e.g., streetscapes) captured in the authors' street-level data are both useful for predicting active travel. The authors' models provide a new tool for estimating and understanding the spatial patterns of active travel.

- **Record URL:**
  https://doi.org/10.1016/j.trd.2020.102651

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S1361920920308361

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/13619209

- **Supplemental Notes:**

- **Authors:**
  - Hankey, Steve
  - Zhang, Wenwen
  - Le, Huyen T K
  - Hystad, Perry
  - James, Peter

- **Publication Date:** 2021-1

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** 102651
- **Serial:**
  - Transportation Research Part D: Transport and Environment
  - Volume: 90
  - Issue Number: 0
  - Publisher: Elsevier
  - ISSN: 1361-9209
  - Serial URL: http://www.sciencedirect.com/science/journal/13619209

Subject/Index Terms

- **TRT Terms:** Bicycling; Census; Imagery; Machine learning; Mathematical prediction; Nonmotorized transportation; Pedestrian traffic; Spatial analysis; Streets; Streetscape; Traffic; Walking
- **Identifier Terms:** Google Street View
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01767806
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Dec 3 2020 3:16PM

THIRTY FIVE

A novel framework for automated monitoring and analysis of high density pedestrian flow

Pedestrian traffic is an important subject of surveillance to ensure public safety and traffic management, which may benefit from intelligent and continuous analysis of pedestrian videos. State-of-the-art methods for intelligent pedestrian surveillance have a number of limitations in automating and deriving useful

information of high-density pedestrian traffic (HDPT) using closed circuit television (CCTV) images. This work introduces an automatic and improved HDPT surveillance system by integrating and optimizing multiple computational steps to predict pedestrian distribution from input video frames. A fast and efficient particle image velocimetry (PIV) technique is proposed to yield pedestrian velocities. A machine learning regressor model, boosted Ferns, is used to improve pedestrian count and density estimation: an essential metric for HDPT analysis. A camera perspective model is proposed to improve the speed and position estimates of HDPT by projecting 2D image pixels to 3D world-coordinate dat. All these functional improvements in HDPT velocity and displacement estimations are used as inputs to a sophisticated pedestrian flow evolution model, PEDFLOW to predict HDPT distribution at a future time point, which is a crucial information for pedestrian traffic management. The predicted and simulated HDPT properties (density, velocity) obtained using the proposed framework show low errors when compared to the ground truth data. The proposed framework is computationally efficient, suitable for multiple camera feeds with HDPT videos, and capable of rapidly analyzing and predicting flows of thousands of pedestrians. The paper shows one of the first steps towards fully integrated CCTV-based automated HDPT management system.

- **Record URL:**
    - https://doi.org/10.1080/15472450.2019.1643724

- **Availability:**
    - Find a library where document is available. Order URL: http://worldcat.org/issn/15472450
- **Supplemental Notes:**
    -

- **Authors:**
    - Baqui, Muhammad
    - Samad, Manar D
    - Löhner, Rainald
- **Publication Date:** 2020-11

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Appendices; Figures; Photos; References; Tables;
- **Pagination:** pp 585-597
- **Serial:**
    - Journal of Intelligent Transportation Systems
    - Volume: 24
    - Issue Number: 6
    - Publisher: Taylor & Francis
    - ISSN: 1547-2450
    - EISSN: 1547-2442
    - Serial URL: http://www.tandfonline.com/loi/gits20

Subject/Index Terms

- **TRT Terms:** Automation; Crowds; Intelligent transportation systems; Pedestrian counts; Pedestrian flow; Traffic density; Traffic surveillance; Video
- **Subject Areas:** Operations and Traffic Management; Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01759163
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Oct 31 2020 3:00PM

THIRTY SIX

A Device-Free Wi-Fi Sensing Method for Pedestrian Monitoring Using Channel State Information

Pedestrian detection accuracy strongly impacts the effectiveness and reliability of intelligent pedestrian-related control systems. Traditional sensing technologies usually sense pedestrians based on the reflected signal of the transmitted infrared ray, sound wave, and electromagnetic wave which only can count the number of times that pedestrians passing a line of sight (LoS) but the moving feature monitoring, e.g., moving direction, speed, etc. For pedestrian monitoring based on computer vision-based sensing technology, the level of errors is relatively large and highly sensitive to environmental factors, such as illumination, weather conditions, and occlusion. Wi-Fi channel state information (CSI) represents the amplitudes and phases information for orthogonal frequency-division multiplexing (OFDM) subcarriers, which is mainly impacted by the static environment and moving object in surrounding areas. Previously, scholars utilized Wi-Fi CSI to analyzed multiple microscopic human movements, e.g., gesture, gait, and fall action in the indoor environment, but no application in the outdoor environment for pedestrian monitoring. The main objective of this research is to demonstrate the feasibility and reliability of the Wi-Fi CSI-based sensing method for pedestrian existence and moving direction recognition. The impacts of the CSI signal sampling ratio on the detection accuracy was investigated as well. The experiments were conducted in both indoor and outdoor environments. According to the results, the accuracy of pedestrian existence detection based on the data of the 100 Hz sampling ratio achieved 99.23% accuracy and 0.26% fast positive rate. For the moving direction recognition, the detection accuracy in the indoor environment achieved 100% and 96.92% for two directions, and got 92.21% and 93.51% in the outdoor environment. The findings of this research demonstrate the proposed Wi-Fi CSI signal is highly effective for pedestrian existence detection and moving direction recognition. The future research will continue in pedestrian moving speed estimation, overlapped pedestrian identification, and pedestrian, bicyclists, and wheelchair classification.

- **Record URL:**
  https://doi.org/10.1061/9780784483138.019

- **Availability:**

- Find a library where document is available. Order URL: http://worldcat.org/isbn/9780784483138
- **Supplemental Notes:**
  -
- **Corporate Authors:**
American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, VA United States 20191-4400

American Society of Civil Engineers
Transportation and Development Institute
1801 Alexander Bell Drive
Reston, VA United States 20191-4400

- **Authors:**
  - Pu, Ziyuan
  - Zhang, Qiannan
  - Zhuang, Yifan
  - Lv, Yongqiang
  - Wang, Yinhai
- **Conference:**
  - International Conference on Transportation and Development 2020
  - Date: 2020-5-26 to 2020-5-29
- **Publication Date:** 2020-8

Language English

Media Info

- **Media Type:** Web
- **Pagination:** pp 207-220
- **Monograph Title:** International Conference on Transportation and Development 2020: Emerging Technologies and Their Impact

Subject/Index Terms

- **TRT Terms:** Pedestrian detectors; Pedestrian safety; Sensors; Wireless communication systems
- **Subject Areas:** Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01752112
- **Record Type:** Publication
- **ISBN:** 9780784483138
- **Files:** TRIS, ASCE
- **Created Date:** Aug 31 2020 3:01PM

THIRTY SEVEN

Development and evaluation of a real-time pedestrian counting system for high-volume conditions based on 2D LiDAR

Automated monitoring of pedestrians on non-motorized facilities with high pedestrian flows is challenging. Several automated sensor solutions are commercially available that have been evaluated in the literature including traditional point-based sensors, such as inductive loop detectors for bicycles and infrared sensors for pedestrians. More recently, image-based systems, based on video cameras or thermal video cameras, have been developed. Despite the various options, some key limitations of existing solutions exist, in particular, the lack of low-cost solutions using embedded systems capable of performing in real-time under high volume (flow) conditions. This work aims at developing and evaluating the performance of a novel, real-time counting system, developed for environments with high pedestrian flows. The proposed system is based on emerging LiDAR (Light Detection and Ranging) technology. As an input, the system uses the distance measurements from a two-dimensional LiDAR sensor with a set of distinct laser channels and a given angular resolution between each channel. The developed system processes those measurements using a clustering algorithm to detect, count, and identify the direction of travel of each pedestrian. The system's performance is evaluated by comparing its directional counting outputs with manual counts (ground truth) using disaggregate and aggregate (15-minutes interval) counts at two different monitoring locations. The results demonstrate that the system accurately counts more than 97% of the pedestrians at the disaggregate level, with a false direction detection rate of 1.1%. The over-counting error is 0.7% and the under-counting errors are 1.3% and 2.7% for the two selected sites. At the aggregate level (15-minutes interval), the average absolute percentage deviations (AAPDs) are 1.6% and 4.3% while the weighted AAPDs are 1.5% and 3.5% for the first and second sites, respectively. The accuracy of the proposed system is higher than the traditional technologies used for the same purpose.

- **Record URL:**
  https://doi.org/10.1016/j.trc.2020.01.018

- **Record URL:**
  http://www.sciencedirect.com/science/article/pii/S0968090X1930083X

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/0968090X
- **Supplemental Notes:**
  -
- **Authors:**

  - Lesani, Asad
  - Nateghinia, Ehsan
  - Miranda-Moreno, Luis F

- **Publication Date:** 2020-5

Language English

Media Info

- **Media Type:** Web
- **Features:** Figures; References; Tables;
- **Pagination:** pp 20-35
- **Serial:**
  - Transportation Research Part C: Emerging Technologies
  - Volume: 114
  - Issue Number: 0
  - Publisher: Elsevier
  - ISSN: 0968-090X
  - Serial URL: http://www.sciencedirect.com/science/journal/0968090X

Subject/Index Terms

- **TRT Terms:** Laser radar; Pedestrian counts; Pedestrian flow; Real time information; Traffic volume
- **Subject Areas:** Data and Information Technology; Operations and Traffic Management; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01735935
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Feb 17 2020 3:16PM

THIRTY EIGHT

Minimizing Annual Average Daily Nonmotorized Traffic Estimation Errors: How Many Counters Are Needed per Factor Group?

Accurate estimates of bicycle and pedestrian volume inform safety studies, trend monitoring, and infrastructure improvements. The Federal Highway Administration's Traffic Monitoring Guide advises current practice for estimation of nonmotorized traffic. While methodologies have been developed to minimize error in estimation of annual average daily nonmotorized traffic (AADNT), challenges persist. This study provides new guidance for monitoring and volume estimation of nonmotorized traffic. Using continuous count data from 102 sites across six cities, the findings confirm that mean absolute percent error (MAPE) in estimated AADNT is minimized when seven-day short duration counts are collected in June through September and for 24-h counts, when data are collected Tuesdays through Thursdays (except for pedestrian-only counts). MAPE across all days (except holidays) and seasons was 34% for 24-h and 20–22% for seven-day short duration counts. The magnitude of bicycle and pedestrian volumes did not significantly affect estimation errors. For factor groups larger than one counter, the length of short duration samples may influence accuracy of AADNT estimates more than the number of counters per group, all else equal. To maximize precision of estimates of AADNT, four or more counters per factor group for bicycle and five or more for pedestrian travel monitoring are recommended. These findings provide guidance for practitioners seeking to establish or improve nonmotorized traffic monitoring programs.

- **Record URL:**
  https://doi.org/10.1177/0361198119848699

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981

- **Authors:**
  - Nordback, Krista
  - Kothuri, Sirisha
  - Johnstone, Dylan
  - Lindsey, Greg
  - Ryan, Sherry
  - Raw, Jeremy
- **Publication Date:** 2019-10

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; References; Tables;
- **Pagination:** pp 295-310
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2673
  - Issue Number: 10
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Annual average daily traffic; Bicycle counts; Estimating; Nonmotorized transportation; Pedestrian counts; Traffic volume; Travel behavior
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01704727
- **Record Type:** Publication
- **Report/Paper Numbers:** 19-05935
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Apr 17 2019 12:07PM

THIRTY NINE

Implementation of Pedestrian and Bicyclist Count Database and Monitoring Process

Project 0-6927 developed a statewide pedestrian and bicyclist count database, as well as guidance for collecting additional count data and requirements for submitting additional count data to the statewide database,in anticipation of eventual mandatory reporting to the federal Travel Monitoring Analysis System (TMAS). Training on data collection and analysis is needed to effectively disseminate the information developed in this project to TxDOT District and regional/local agency staff and to utilize bicycle/pedestrian data in day-to-day operations.

- **Record URL:**
  https://library.ctr.utexas.edu/Presto/project=5-6927-01

- **Supplemental Notes:**
  - The project is a continuation of TxDOT Research Project 0-6927, "Evaluation of Bicycle and Pedestrian Monitoring Equipment to Establish Collection Database Methodologies for Estimating Non-Motorized Transportation."

Language English

Project

- **Status:** Completed
- **Funding:** $150,778
- **Contract Numbers:** 5-6927-01
- **Sponsor Organizations:**
  Texas Department of Transportation
  125 E. 11th Street
  Austin, TX United States 78701-2483

- **Managing Organizations:**
  Texas Department of Transportation
  125 E. 11th Street
  Austin, TX United States 78701-2483

- **Project Managers:**
  Glancy, Chris
- **Performing Organizations:**
  Texas A&M Transportation Institute
  Texas A&M University System
  3135 TAMU
  College Station, TX United States 77843-3135

- **Principal Investigators:**

Turner, Shawn
- **Start Date:** 20190901
- **Expected Completion Date:** 20200630
- **Actual Completion Date:** 0

Subject/Index Terms

- **TRT Terms:** Cyclists; Data collection; Nonmotorized transportation; Pedestrians; Traffic data; Traffic surveillance
- **Subject Areas:** Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01766433
- **Record Type:** Research project
- **Source Agency:** Texas Department of Transportation
- **Contract Numbers:** 5-6927-01
- **Files:** RIP, STATEDOT
- **Created Date:** Mar 3 2021 4:24PM

FORTY

Pedestrians and Bicyclists Count: Developing a Statewide Multimodal Count Program

The purpose of this study was to research best practices and available methods and technologies for measuring active transportation activity, in order to provide the Louisiana Department of Transportation and Development (DOTD) with needed information in support of the development of an efficient, cost-effective bicycle and pedestrian count program. Measuring progress toward Complete Streets policy implementation, as well as measuring the performance of individual projects in terms of safety outcomes, requires understanding patterns of and changes in active transportation demand so as to a) evaluate safety outcomes relative to rates of exposure, b) identify appropriate, context-sensitive complete streets infrastructure interventions, and c) understand overall statewide and location-specific transportation trends which will impact long-range planning and investment. To this end, the research team conducted a comprehensive review of academic and applied literature pertaining to collecting pedestrian and bicycle data collection and benchmarking, with a focus on techniques for using count data to evaluate exposure rates and safety outcomes or trends, researched methods of counting bicycles and pedestrians including both manual counts and automated electronic counts using various technologies (including automated video-based counts), and identified potential funding sources and potential partners for systematic as well as incidental data collection. Finally, the research team conducted pilot data collection and analysis at three case study locations in New Orleans and Baton Rouge to test recommended count equipment and count methodology and advance fundamental elements of comprehensive evaluation of the safety impacts of complete streets-oriented infrastructure. The results of this research indicate that the incremental development of systematic active transportation monitoring, in coordination with existing traffic monitoring activities and in cooperation with local and regional agencies interested in or already engaged in data collection and analysis, is feasible and scalable (geographically and fiscally) using a combination of traditional and emerging technologies. Moreover, significant expansion of long-duration count data availability is critical to all efforts to holistically evaluate safety impacts at the project level, and an area where state leadership and investment will have the greatest impact.

- **Record URL:**
  https://www.ltrc.lsu.edu/pdf/2019/FR_599.pdf


Louisiana Transportation Research Center

-
- **Corporate Authors:**
  University of New Orleans
  Transportation Institute, 2000 Lakeshore Drive
  New Orleans, LA United States 70148

  Louisiana Department of Transportation and Development
  1201 Capitol Access Road, P.O. Box 94245
  Baton Rouge, LA United States 70804-9245

  Federal Highway Administration
  1200 New Jersey Avenue, SE
  Washington, DC United States 20590

- **Authors:**
  - Tolford, Tara M
  - Izadi, Maryam
  - Ash, Colin
  - Codjoe, Julius
- **Publication Date:** 2019-8

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Appendices; Figures; Photos; References; Tables;
- **Pagination:** 134p

Subject/Index Terms

- **TRT Terms:** Best practices; Bicycle counts; Case studies; Complete streets; Financing; Literature reviews; Long range planning; Monitoring; Nonmotorized transportation; Pedestrian counts

- **Geographic Terms:** Baton Rouge (Louisiana); New Orleans (Louisiana)
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01719515
- **Record Type:** Publication
- **Report/Paper Numbers:** FHWA/LA.17/599
- **Contract Numbers:** LTRC 16-4SA
- **Files:** TRIS, ATRI, USDOT, STATEDOT
- **Created Date:** Sep 25 2019 4:04PM

FORTY ONE

Development and Testing of a Real-Time WiFi-Bluetooth System for Pedestrian Network Monitoring, Classification, and Data Extrapolation

A real-time pedestrian monitoring system provides information about traffic flow, speeds, travel times, and time spent in areas or transportation facilities of interest. This is useful in travel information systems and crowd management strategies, as well as in planning and emergencies in public spaces, such as airports, parks, malls, and university campuses. While there are technologies that can obtain count data for non-motorized transportation at specific locations, most technologies cannot provide origin-destination information, trip paths, travel times, or time spent. To overcome these shortcomings, some studies have explored the use of Bluetooth (BT) sensors to capture the unique media access control (MAC) addresses of mobile devices carried by pedestrians. However, this collection method may suffer from low-detection rates. As an alternative, collecting MAC data from WiFi signals has emerged. The objective of this paper is three-fold: 1) develop and evaluate the performance of an integrated WiFi-BT system to monitor pedestrian-cyclists activity traffic; 2) develop and validate a classification method for differentiating pedestrians from bicycles; and 3) propose a simple extrapolation method that combines counts and MAC data. Among other results, relatively high detection rates were obtained for the developed WiFi system in comparison with BT sensors. In addition, high correlation between estimated and ground truth speeds and low classification errors are observed. Finally, the extrapolated WiFi counts and ground truth counts were found to be highly correlated. These results demonstrate the feasibility of the proposed system and methods to estimate travel times (speeds), to classify bicycle-pedestrian WiFi signals, and to extrapolate pedestrian MAC counts.

- **Record URL:**
  - https://doi.org/10.1109/TITS.2018.2854895
- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/oclc/41297384
- **Supplemental Notes:**
  -
- **Authors:**
  - Lesani, Asad
  - Miranda-Moreno, Luis

- **Publication Date:** 2019-4

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; References; Tables;
- **Pagination:** pp 1484-1496
- **Serial:**
    - IEEE Transactions on Intelligent Transportation Systems
    - Volume: 20
    - Issue Number: 4
    - Publisher: Institute of Electrical and Electronics Engineers (IEEE)
    - ISSN: 1524-9050
    - Serial URL: http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6979

Subject/Index Terms

- **TRT Terms:** Bluetooth technology; Computer network protocols; Monitoring; Pedestrian flow; Pedestrian traffic; Pedestrians; Wireless communication systems
- **Subject Areas:** Data and Information Technology; Highways; Operations and Traffic Management; Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01706642
- **Record Type:** Publication
- **Files:** TLIB, TRIS
- **Created Date:** Apr 5 2019 2:18PM

FORTY TWO

Bicycle and Pedestrian Count Programs: Scan of Current U.S. Practice

As bicycling and walking have become more integrated into transportation agencies' processes of planning, design, and operations, some state, regional, and local agencies have established nonmotorized data collection programs of varying scopes and with varying methods. The purpose of this study was to identify ways to plan and implement a nonmotorized count program in Virginia, and the scope included reviewing existing U.S. national-level guidance and examples from state departments of transportation (DOTs) other than Virginia's to determine the most effective ways of implementing such a program. Study tasks included synthesizing the literature to obtain relevant information with regard to nonmotorized travel monitoring programs, practices, and technologies, as well as obtaining information from representatives of three states through interviews of public agency staff and researchers involved in each state's program. The study found a large volume of recent research on the topic of nonmotorized travel monitoring. The study concluded that the practice of nonmotorized travel monitoring has evolved and expanded in recent years; that many commercially available counting technologies exist and have been evaluated; that the practice of nonmotorized travel monitoring, as with motorized travel monitoring, has several aspects beyond purchase and installation of automatic count equipment; and that several states are developing nonmotorized count programs and have begun putting their data to use. The findings provide a foundational resource for state DOTs that are considering developing state-level counting programs.

- **Record URL:**
  - https://doi.org/10.1177/0361198119834924

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Authors:**
  - Ohlms, Peter B
  - Dougald, Lance E
  - MacKnight, Hannah E
- **Publication Date:** 2019-3

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** References;
- **Pagination:** pp 74-85
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2673
  - Issue Number: 3
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Data collection; Nonmotorized transportation; Research; State of the practice; Traffic counting
- **Geographic Terms:** United States
- **Subject Areas:** Data and Information Technology; Operations and Traffic Management; Pedestrians and Bicyclists; Research;

Filing Info

- **Accession Number:** 01698882
- **Record Type:** Publication
- **Report/Paper Numbers:** 19-00615
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Feb 12 2019 4:41PM

FORTY THREE

Applying Data Mining Techniques to Analyze the Pedestrian and Bicycle Crashes at the Macroscopic Level

This paper presents different data mining techniques to analyze the vulnerable road user (i.e., pedestrian and bicycle) crashes by developing crash prediction models at macro-level. In this study, the authors developed data mining approach (i.e., decision tree regression (DTR) models) for both pedestrian and bicycle crash counts. To author knowledge, this is the first application of DTR models in the growing traffic safety literature at macro-level. The empirical analysis is based on the Statewide Traffic Analysis Zones (STAZ) level crash count data for both pedestrian and bicycle from the state of Florida for the year of 2010 to 2012. The model results highlight the most significant predictor variables for pedestrian and bicycle crash count in terms of three broad categories: traffic, roadway, and socio demographic characteristics. Furthermore, spatial predictor variables of neighboring STAZ were utilized along with the targeted STAZ variables in order to improve the prediction accuracy of both DTR models. The DTR model considering spatial predictor variables (spatial DTR model) were compared without considering spatial predictor variables (aspatial DTR model) and the models comparison results clearly found that spatial DTR model is superior model compared to aspatial DTR model in terms of prediction accuracy. Finally, this study contributed to the safety literature by applying three ensemble techniques (Bagging, Random Forest, and Boosting) in order to improve the prediction accuracy of weak learner (DTR models) for macro-level crash count. The model's estimation result revealed that all the ensemble technique performed better than the DTR model and the gradient boosting technique outperformed other competing ensemble technique in macro-level crash prediction model.

- **Supplemental Notes:**
    - This paper was sponsored by TRB committee ANB20 Standing Committee on Safety Data, Analysis and Evaluation.
- **Corporate Authors:**
  Transportation Research Board
- **Authors:**
    - Rahman, Md Sharikur
    - Abdel-Aty, Mohamed
    - Hasan, Samiul
    - Cai, Qing
- **Conference:**
    - Transportation Research Board 98th Annual Meeting
    - Location: Washington DC, United States
    - Date: 2019-1-13 to 2019-1-17
- **Date:** 2019

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** References; Tables;
- **Pagination:** 7p

Subject/Index Terms

- **TRT Terms:** Bicycle crashes; Data mining; Pedestrian vehicle crashes; Traffic analysis zones; Vulnerable road users
- **Geographic Terms:** Florida
- **Subject Areas:** Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01697901
- **Record Type:** Publication
- **Report/Paper Numbers:** 19-00055
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Dec 7 2018 9:41AM

FORTY FOUR

A Pedestrian-Oriented Framework for Measuring Area-Wide Pedestrian Activity

Research continuously suggests the built environment provides opportunities as well as barriers to active travel and physical activity. Individuals residing in densely-populated neighborhoods with a mix of land uses tend to walk more. While insights into how neighborhood-level built environment features impact individual-level behavior continue to inform urban policies, research centered on measuring active travel at a neighborhood scale is needed to estimate the population-level impact of policy, systems, or environmental changes on transportation-related physical activity. To date, a nascent body of research has sought to create the requisite tools for measuring area-wide levels of active travel, albeit by applying existing vehicle-based methods. The authors' study advances current practice by introducing a pedestrian-oriented approach to classifying streets based on a measure of local destination accessibility along a given street segment, or its network utility, and pedestrian count data collected from multiple randomly-selected sites in four neighborhoods across Massachusetts. As a proof of concept, in one study area, data collected with automated counters using the authors' pedestrian-based street stratification method were expanded in order to create area-wide estimates of seasonal average daily pedestrian counts that were in turn used to estimate average daily pedestrian miles traveled. An area-wide pedestrian activity estimate, which resulted from a method that can establish pre/post-intervention statistics of neighborhood-level pedestrian travel and physical activity needed to inform evidence-based policies.

- **Supplemental Notes:**
    - This paper was sponsored by TRB committee ABJ35 Standing Committee on Highway Traffic Monitoring.
- **Corporate Authors:**
  Transportation Research Board
- **Authors:**
    - Gehrke, Steven R
    - James, Peter
    - Reeves, Halley
    - Ron, Sharon
    - Reardon, Timothy G
    - Keppard, Barry
    - Ursprung, W W Sanouri

- **Conference:**
    - Transportation Research Board 98th Annual Meeting
    - Location: Washington DC, United States
    - Date: 2019-1-13 to 2019-1-17
- **Date:** 2019

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** References;
- **Pagination:** 3p

Subject/Index Terms

- **TRT Terms:** Built environment; Classification; Measurement; Neighborhoods; Nonmotorized transportation; Pedestrian counts; Pedestrians; Policy; Streets; Walkability; Walking
- **Geographic Terms:** Massachusetts
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01697760
- **Record Type:** Publication
- **Report/Paper Numbers:** 19-02908
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Dec 7 2018 9:36AM

FORTY FIVE

Data Fusion to Improve the Accuracy of Multi-Modal Traffic Counts

Description: Current traffic counting systems often only measure one transportation mode accurately. In this project, the research team will improve the reliability and accuracy of video-based traffic counting technology by augmenting the video data with information extracted from other sensing technology. Additional data can originate from tube counters, magnetic loops, radar, vibration, and laser measurements. The project will use the raw data from the augmented sensors (with transient tube pressure signals) to count and classify vehicle types (FHWA 13 types, bicycles, and pedestrian traffic). Intellectual Merit: This project will evaluate the use of combined raw data from the tube-based vehicle counting/classification method and an integrated artificial neural network (ANN) to classify vehicle types with better accuracy than existing methods using data from one type of sensor. Broader Impacts: Improved data on the multi-modal movement of people and freight will provide transportation planners with better quantitative information on use of the existing system. Technology Transfer Plan: This research will be generating an implementation-ready hybrid traffic data collection tool for DOTs.

Project

- **Status:** Active
- **Funding:** $108524
- **Contract Numbers:**
  69A3551747117

- **Sponsor Organizations:**
  Office of the Assistant Secretary for Research and Technology
  University Transportation Centers Program
  Department of Transportation
  Washington, DC United States 20590

  Center for Connected Multimodal Mobility
  Clemson University
  Clemson, SC United States 29634

  University of South Carolina, Columbia
  502 Byrnes Building
  Columbia, SC United States 29208

  Benedict College
  1600 Harden Street
  Columbia, South Carolina United States 29204

- **Managing Organizations:**
  University of South Carolina, Columbia
  502 Byrnes Building
  Columbia, SC United States 29208

- **Project Managers:**
  Mullen, Robert
- **Performing Organizations:**
  University of South Carolina, Columbia
  502 Byrnes Building
  Columbia, SC United States 29208

  Benedict College
  1600 Harden Street
  Columbia, South Carolina United States 29204

- **Principal Investigators:**
  Mullen, Robert
  Comert, Gurcan
  Huynh, Nathan
- **Start Date:** 20181201
- **Expected Completion Date:** 20200831
- **Actual Completion Date:** 0
- **USDOT Program:** University Transportation Center

Subject/Index Terms

- **TRT Terms:** Data analysis; Planning; Traffic; Video imaging detectors
- **Subject Areas:** Data and Information Technology; Planning and Forecasting; Transportation (General);

Filing Info

- **Accession Number:** 01690758
- **Record Type:** Research project
- **Source Agency:** Center for Connected Multimodal Mobility
- **Contract Numbers:** 69A3551747117
- **Files:** UTC, RIP
- **Created Date:** Jan 11 2019 4:06PM

FORTY SIX

Comparison of Pedestrian Count Expansion Methods: Land Use Groups versus Empirical Clusters

Expansion factors based on the trends in long-term count data are useful tools for estimating daily, weekly, or annual volumes from short-term counts, but it is unclear how to differentiate locations by activity pattern. This paper compares two approaches to developing factor groups for hour-to-week pedestrian count expansion factors. The land use (LU) classification approach assumes that surrounding LUs affect the pedestrian activity at a location, and it is easy to apply to short-term count locations based on identifiable attributes of the site. The empirical clustering (EC) approach uses statistical methods to match locations based on the actual counts, which may produce more accurate volume estimates, but presents a challenge for determining which factor group to apply to a location. We found that both the LU and EC approaches provided better weekly pedestrian volume estimates than the single factor approach of taking the average of all locations. Further, the differences between LU and EC estimation errors were modest, so it may be beneficial to use the intuitive and practical LU approach. LU groupings can also be modified with insights from the EC results, thus improving estimates while maintaining the ease of application. Ideal times for short-term counts are during peak activity periods, as they generally produce estimates with fewer errors than off-peak periods. Weekly volume estimated from longer-duration counts (e.g., 12?h) is generally more accurate than estimates from shorter-duration counts (e.g., 2?h). Practitioners can follow this guidance to improve the quality of weekly pedestrian volume estimates.

- **Record URL:**
  - https://doi.org/10.1177/0361198118793006

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981

- **Authors:**
  - Griswold, Julia B
  - Medury, Aditya
  - Schneider, Robert J
  - Grembek, Offer
- **Publication Date:** 2018-12

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** pp 87-97
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2672
  - Issue Number: 43
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Data analysis; Land use; Location; Pedestrian counts; Periods of the day; Traffic volume
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01664256
- **Record Type:** Publication
- **Report/Paper Numbers:** 18-05579
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Jan 8 2018 11:25AM

FORTY SEVEN

Incorporating the Effect of Special Events into Continuous Count Site Selection for Pedestrian Traffic

This paper presents results from pedestrian monitoring research conducted in a dense urban environment in Winnipeg, Canada. Pedestrian counts were conducted in downtown Winnipeg using infrared pedestrian counters. Count sites were assigned to traffic pattern groups (TPGs) based on their response to special events occurring in the study area. Once these groups were established, eight continuous count sites were installed to initiate an ongoing pedestrian traffic monitoring program for the city. Traffic monitoring efforts have primarily focused on motorized travel. As more jurisdictions prioritize active transportation, addressing the need for network-level pedestrian data is essential to optimize engineering decisions. The first step to developing any system-wide traffic monitoring program is to define TPGs. These groups enable the spatial variation of short-duration counts to be adjusted to annual statistics by the temporal variation of similarly behaving continuous counts. Short-duration count sites were characterized by daily and hourly trends consistent with existing pedestrian traffic monitoring practices. Recognizing the influence of large evening events on pedestrian traffic, a metric was developed called the evening proportion ratio (EPR) to quantify the effect of special events. Based on the spatial distribution of EPR

values, two TPGs were developed for downtown Winnipeg. These are the “urban utilitarian” and “urban utilitarian – event” groups. These groups were used to select continuous count locations for ongoing pedestrian traffic data collection. The importance of this research lies in its future applicability to other jurisdictions in developing a standard approach for urban transportation authorities to strategically implement pedestrian traffic monitoring programs.

- **Record URL:**
  - https://doi.org/10.1177/0361198118788188
- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/issn/03611981
- **Authors:**
  - Olfert, Caleb
  - Poapst, Rob
  - Montufar, Jeannette
- **Publication Date:** 2018-12

Language English

Media Info

- **Media Type:** Digital/other
- **Features:** Figures; Maps; References; Tables;
- **Pagination:** pp 65-74
- **Serial:**
  - Transportation Research Record: Journal of the Transportation Research Board
  - Volume: 2672
  - Issue Number: 43
  - Publisher: Sage Publications, Incorporated
  - ISSN: 0361-1981
  - EISSN: 2169-4052
  - Serial URL: http://journals.sagepub.com/home/trr

Subject/Index Terms

- **TRT Terms:** Evening; Groups; Pedestrian counts; Pedestrian flow; Pedestrian traffic; Special events; Traffic data; Traffic surveillance; Travel patterns; Urban travel
- **Geographic Terms:** Winnipeg (Canada)
- **Subject Areas:** Data and Information Technology; Operations and Traffic Management; Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

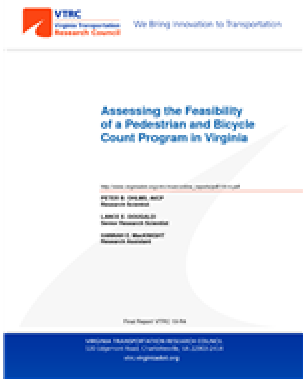
- **Accession Number:** 01657472
- **Record Type:** Publication
- **Report/Paper Numbers:** 18-01261
- **Files:** TRIS, TRB, ATRI
- **Created Date:** Jan 8 2018 10:19AM

FORTY EIGHT

Assessing the Feasibility of a Pedestrian and Bicycle Count Program in Virginia

In recent years, there has been a paradigm shift within transportation agencies to account for and incorporate nonmotorized travel in business and strategic highway safety plans. Several federal initiatives have been developed to encourage the creation of safer, more comfortable, and more connected bicycling and walking environments. In addition, local and regional agencies have established data collection programs of varying scopes and with varying methods. Some local governments and other organizations have implemented automatic counting equipment that provides short-duration or continuous count data. With some exceptions in urban areas and on major off-street trails, the Virginia Department of Transportation (VDOT) has not typically collected or made use of these data, which vary in terms of quality and availability. Based on discussions with staff of VDOT's Transportation and Mobility Planning Division and Traffic Engineering Division, no formal approach or program had been established to collect or incorporate count data for bicycle and pedestrian modes throughout Virginia, thus making it difficult to plan projects, prioritize improvements, and justify investments. The purpose of this study was to identify ways to plan and implement a nonmotorized count program in Virginia including an understanding of whom such a program would serve and how frequently data would need to be collected and for what purposes. The study tasks included (1) reviewing existing national-level guidance and examples from other state departments of transportation to determine effective ways to implement bicycle and pedestrian counting programs; (2) obtaining Virginia-specific information from localities and organizations including data collection locations and methods; and (3) developing a framework for VDOT to initiate a pilot count program in Virginia. The study found a large volume of recent research on the topic of nonmotorized travel monitoring; several states were developing count programs and had begun putting their data to use. In Virginia, many localities were interested in some level of pedestrian and bicycle volume data collection, although relatively few already engaged in the practice. To assist with counting efforts, localities in VDOT's Salem and Northern Virginia districts expressed a high level of interest in partnering with VDOT using partnership models currently employed by the North Carolina Department of Transportation and/or the Minnesota Department of Transportation. The study recommends that VDOT's Transportation and Mobility Planning Division, with assistance from the Virginia Transportation Research Council, establish a pilot nonmotorized count program in one or more VDOT districts. Recommended program elements include purchasing and installing count equipment; identifying opportunities for training and outreach; and working with VDOT's Traffic Engineering Division to identify an acceptable data storage mechanism. The study also recommends that the Virginia Transportation Research Council assist in evaluating the pilot program and documenting lessons learned. Providing count data that could be of use to localities and VDOT as described in this report and incrementally expanding VDOT's capabilities in this area will inform future actions including maximizing the value of efforts (by using compatible data formats and methodologies), simplifying data analysis and use, and facilitating reporting of such data to the federal data repository

- **Record URL:**
  http://www.virginiadot.org/vtrc/main/online_reports/pdf/19-R4.pdf

- 
Assessing the Feasibility of a Pedestrian and Bicycle Count Program in Virginia

- **Corporate Authors:**

Virginia Transportation Research Council
530 Edgemont Road
Charlottesville, VA United States 22903

Virginia Department of Transportation
1221 East Broad Street
Richmond, VA United States 23219

Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC United States 20590

- **Authors:**
  - Ohlms, Peter B
  - Dougald, Lance E
  - MacKnight, Hannah E
- **Publication Date:** 2018-8

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Appendices; Figures; References; Tables;
- **Pagination:** 132p

Subject/Index Terms

- **TRT Terms:** Bicycle counts; Data collection; Feasibility analysis; Nonmotorized transportation; Pedestrian counts
- **Uncontrolled Terms:** Data compatibility; Pilot programs
- **Geographic Terms:** Virginia
- **Subject Areas:** Highways; Pedestrians and Bicyclists; Planning and Forecasting;

Filing Info

- **Accession Number:** 01680023
- **Record Type:** Publication
- **Report/Paper Numbers:** FHWA/VTRC 19-R4, VTRC 19-R4
- **Contract Numbers:** 109948
- **Files:** TRIS, ATRI, USDOT, STATEDOT
- **Created Date:** Aug 16 2018 7:34AM

FORTY NINE

Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis

This investigation aims to evaluate whether the Safety in Numbers phenomenon is observable in the midwestern U.S. city of Minneapolis, Minnesota. Safety in Numbers (SIN) refers to the phenomenon that pedestrian safety is positively correlated with increased pedestrian traffic in a given area. Walking and bicycling are increasingly becoming important transportation modes in modern cities. Proper placement of non-motorized facilities and improvements has implications for safety, accessibility, and mode choice, but proper information regarding estimated non-motorized traffic levels is needed to locate areas where investments can have the greatest impact. Assessment of collision risk between automobiles and non-motorized travelers offers a tool that can help inform investments to improve non-motorized traveler safety. Models of non-motorized crash risk typically require detailed historical multimodal crash and traffic volume data, but many cities do not have dense datasets of non-motorized transport flow levels. Methods of estimating pedestrian and bicycle behavior that do not rely heavily on high-resolution count data are applied in this study. Pedestrian and cyclist traffic counts, average automobile traffic, and crash data from the city of Minneapolis are used to build models of crash frequencies at the intersection level as a function of modal traffic inputs. These models determine whether the SIN effect is observable within the available datasets for pedestrians, cyclists, and cars, as well as determine specific locations within Minneapolis where non-motorized travelers experience elevated levels of risk of crashes with automobiles.

- **Record URL:**
  http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=2656

- **Record URL:**
  http://www.roadwaysafety.umn.edu/publications/researchreports/reportdetail.html?id=2656

- **Record URL:**
  https://rosap.ntl.bts.gov/view/dot/37252

- **Corporate Authors:**
  University of Minnesota, Twin Cities

Department of Civil, Environmental and Geo-Engineering
500 Pillsbury Drive SE
Minneapolis, MN United States 55455

University of Minnesota, Twin Cities
Roadway Safety Institute, Center for Transportation Studies
200 Transportation and Safety Building, 511 Washington Ave. SE
Minneapolis, MN United States 55455

- **Authors:**
    - Carlson, Kristin
    - Murphy, Brendan
    - Ermagun, Alireza
    - Levinson, David M
    - Owen, Andrew
- **Publication Date:** 2018-3

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Figures; References; Tables;
- **Pagination:** 72p

Subject/Index Terms

- **TRT Terms:** Bicycling; Cyclists; Nonmotorized transportation; Pedestrian safety; Pedestrians; Traffic counts; Traffic safety
- **Geographic Terms:** Minneapolis (Minnesota)
- **Subject Areas:** Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01667116
- **Record Type:** Publication
- **Report/Paper Numbers:** CTS 18-05
- **Contract Numbers:** DTRT13-G-UTC35
- **Files:** NTL, TRIS, ATRI, STATEDOT
- **Created Date:** Mar 27 2018 4:08PM

FIFTY

Predicting Pedestrian Counts for Crossing Scenario Based on Fused Infrared-Visual Videos

Estimating the number of pedestrians based upon surveillance videos and images has been a critical task while implementing intelligent signal controls at intersections. However, this has been a difficult task considering the pedestrian waiting area is an outdoor scenario with complex and time-varying

surrounding environment. In this study, a method for estimating pedestrian counts based on multisource video data has been proposed. First, the partial least squares regression (PLSR) model is developed to estimate the number of pedestrians from single-source video (either visible light video or infrared video). Meanwhile, the temporal feature of the scenario (daytime or nighttime) is identified based on visible light video. According to the recognized time periods, pedestrian count detection results from the visible light and infrared video data can be obtained with preset corresponding confidence levels. The empirical experiments showed that this fusion method based on environment perception holds the benefits of 24-hour monitoring for outdoor scenarios at the pedestrian waiting area and substantially improved accuracy of pedestrian counting.

- **Record URL:**
  - https://doi.org/10.1155/2018/8703576

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/oclc/5121625
- **Supplemental Notes:**
  - © 2019 Shize Huang, et al.
- **Authors:**
  - Huang, Shize
  - Chen, Wei
  - Yu, Rongjie
  - Yang, Xiaolu
  - Dong, Decun
- **Publication Date:** 2018

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** 11p
- **Serial:**
  - Journal of Advanced Transportation
  - Volume: 2018
  - Issue Number: Article ID 8703576
  - Publisher: John Wiley & Sons, Incorporated
  - ISSN: 0197-6729
  - EISSN: 2042-3195
  - Serial URL: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2042-3195
- **Publication flags:** Open Access (libre)

Subject/Index Terms

- **TRT Terms:** Image analysis; Intelligent transportation systems; Monitoring; Pedestrian clearance interval (Traffic signals); Pedestrian counts; Traffic signals; Traffic surveillance; Video
- **Subject Areas:** Highways; Pedestrians and Bicyclists; Safety and Human Factors;

Filing Info

- **Accession Number:** 01702952
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Mar 14 2019 3:45PM

FIFTY ONE

Real-Time Pedestrian Tracking and Counting with TLD

This paper describes a solution to solve the issue of automatic multipedestrian tracking and counting. First, background modeling algorithm is applied to actively obtain multipedestrian candidates, followed by a confirmation step with classification. Then each pedestrian patch is handled by real-time TLD (Tracking-Learning-Detection) to get a new predication position according to similarity measure. Further TLD results are compared with classification list to determine a new, disappeared, or existing pedestrian. Finally single line counting with buffer zone is employed to count pedestrians. Experiments results on the public database, PETS, demonstrate the validity of the solution.

- **Record URL:**
  - https://doi.org/10.1155/2018/8486906

- **Availability:**
  - Find a library where document is available. Order URL: http://worldcat.org/oclc/5121625
- **Supplemental Notes:**
  -
- **Authors:**
  - Shi, Jiawei
  - Wang, Xianmei
  - Xiao, Huer
- **Publication Date:** 2018

Language English

Media Info

- **Media Type:** Web
- **Features:** References;
- **Pagination:** 7p
- **Serial:**
  - Journal of Advanced Transportation
  - Volume: 2018
  - Issue Number: Article ID 8486906

- Publisher: John Wiley & Sons, Incorporated
  - ISSN: 0197-6729
  - EISSN: 2042-3195
  - Serial URL: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2042-3195
- **Publication flags:** Open Access (libre)

Subject/Index Terms

- **TRT Terms:** Classification; Data analysis; Detection and identification systems; Pedestrian counts; Pedestrian detectors; Real time information
- **Subject Areas:** Data and Information Technology; Pedestrians and Bicyclists;

Filing Info

- **Accession Number:** 01685977
- **Record Type:** Publication
- **Files:** TRIS
- **Created Date:** Oct 29 2018 10:54AM

FIFTY TWO

Vermont Bicycle and Pedestrian Counting Program

The goals of this project were to create a bicycle and pedestrian count database for the state of Vermont, communicate the state of non-motorized travel statewide, and make recommendations for future data collection and management. The specific objectives of this project were to:1. review best practices in non-motorized count data management and count planning; 2. gather all existing counts into a single database and create or implement a bike and pedestrian count data web portal; 3. validate the existing count data and the methods of collection; 4. identify gaps in the statewide data set; and 5. develop a continued counting plan to be implemented by the Vermont Agency of Transportation (VTrans}.

- **Record URL:**
  http://www.uvm.edu/~transctr/research/trc_reports/UVM-TRC-17-006.pdf

- 

- **Corporate**
- **:** 2017-11

Language English

Media Info

- **Media Type:** Digital/other
- **Edition:** Final Report
- **Features:** Appendices; Figures; Maps; Photos; References; Tables;
- **Pagination:** 130p

Subject/Index Terms

- **TRT Terms:** Bicycle counts; Data collection; Data management; Nonmotorized transportation; Pedestrian counts
- **Geographic Terms:** Vermont
- **Subject Areas:** Pedestrians and Bicyclists; Planning and Forecasting;

- Filing Info **Authors:**
University of Vermont, Burlington
Transportation Research Center
210 Colchester Avenue
Burlington, VT United States 05405-1757

Vermont Agency of Transportation
1 National Life Drive, Drawer 33
Montpelier, VT United States 05633-0001

- **Authors:**
    - Sentoff, Karen
    - Sullivan, James L

Publication Date

- **Accession Number:** 01698909
- **Record Type:** Publication
- **Report/Paper Numbers:** TRC Report 17-006
- **Files:** TRIS, ATRI, STATEDOT
- **Created Date:** Feb 27 2019 8:03AM

Appendix B: State Agencies Survey Report

The survey has been open for collecting responses on November 23, 2021, after two sets of beta testing internally and with MDOT. This report analyzes the seventeen (17) responses received until December 26, 2021. The report is organized in three sections following the survey design – the first section provides information on the participant; the second section provides information on the current state of the practice as reported by the participants and the third section provides information on use of crowdsourced data and the opinion of the participants related to quality and use of crowdsourced data for volume estimation and safety/risk estimation.

Section 1: The Participant: Affiliation, Experience and Role

1. Your current organization is:

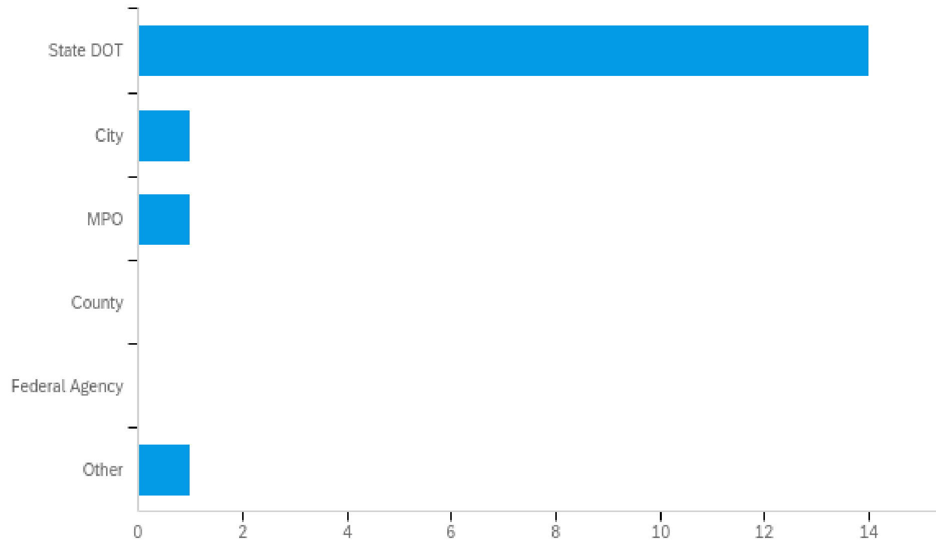


Figure 1. Distribution of respondents across organization type

Table 1. Percentage distribution of respondents across organization types

Organizations	%	Count
State DOT	82.35%	14
City	5.88%	1
MPO	5.88%	1
County	0.00%	0
Federal Agency	0.00%	0
Other	5.88%	1
Total	100%	17

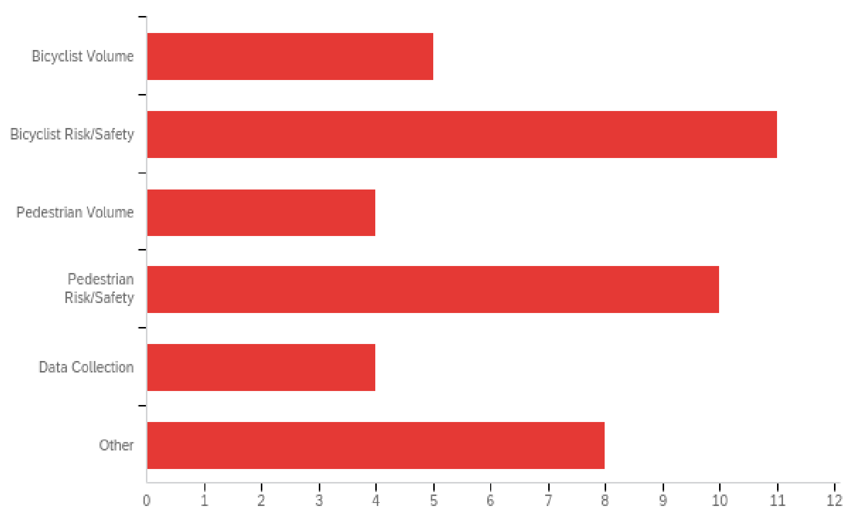
2. Please indicate your role in your current organization (for example, Bicyclist and Pedestrian planner):

Table 2. List of respondent roles as reported by the respondents

PennDOT Bike / Ped Coordinator
Transportation Planner

Transportation Planner - Active Transportation
Highway Safety Engineer
Planning Manager and Bicycle Pedestrian Coordinator
Bike Ped Coordinator
State Traffic Safety Engineer
Bicycle and Pedestrian Coordinator
Traffic Safety Analyst
Assistant Bike/Ped Coordinator
Traffic Engineer including Bike/Ped Coordinator
Transit and Active Transportation Planning Supervisor
Bicycle and Pedestrian Planner
Program Administrator for Minnesota's traffic crash report system
Bicycle and Pedestrian Planner Consultant
Pedestrian and Bicyclist Data Coordinator
Traffic Safety Marketing Coordinator/Strategic Highway Safety Plan Administrator

3. **In your day-to-day activities, which of the following is in your area of responsibility? Please choose all that apply.**



Other - Text
Project Planning
I run the program and reviews and impliments bike lanes and pedestrian trails
Data Analysis
oversight and direction for transit & active transportation planning and programming
Bicycle and Ped funding
Crash data management, improvement and analysis
Related research and application

Figure 2. Distribution of respondents across responsibilities

4. How many years of experience do you have in your bicyclist-pedestrian related area of responsibility?

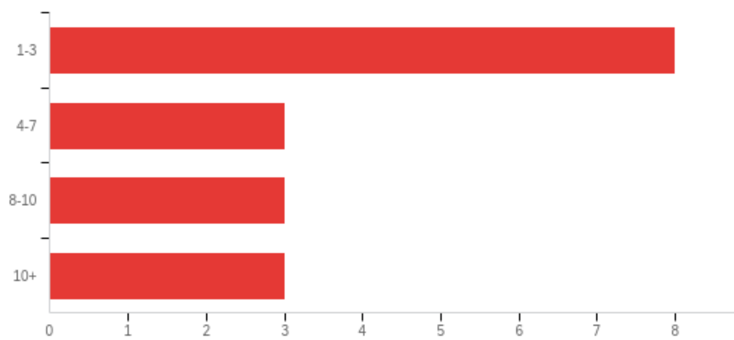


Figure 3. Distribution of respondents across experience in years

Section 2. Current practices and experiences with data collection

5. **What is/are your primary data source(s) for estimating bicyclist volume? Please rank the following choices in order of importance with '1' being the most important data source. If you do not use a particular data source, please use '0' to indicate its rank. If multiple data sources are equally important, please give them the same rank. If you are unsure, please use the option 'Other' to elaborate.**

Table 3. Ranking of different data sources for bicyclist volume estimation

Data Sources	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5		Rank 6		Rank7		Total
Manual counts	42.86%	3	0.00%	0	14.29%	1	42.86%	3	0.00%	0	0.00%	0	0.00%	0	7
Permanent or long term automated counters	37.50%	3	25.00%	2	37.50%	3	0.00%	0	0.00%	0	0.00%	0	0.00%	0	8
Short duration (<3 months) automated counters	0.00%	0	75.00%	6	25.00%	2	0.00%	0	0.00%	0	0.00%	0	0.00%	0	8
Travel surveys	25.00%	1	0.00%	0	0.00%	0	50.00%	2	25.00%	1	0.00%	0	0.00%	0	4
Estimation models	0.00%	0	25.00%	1	50.00%	2	0.00%	0	25.00%	1	0.00%	0	0.00%	0	4
Third party data (e.g. Strava, Streetlight)	33.33%	2	0.00%	0	50.00%	3	16.67%	1	0.00%	0	0.00%	0	0.00%	0	6
Other (please specify)	50.00%	1	0.00%	0	0.00%	0	0.00%	0	0.00%	0	50.00%	1	0.00%	0	2

6. **What is/are your primary data source(s) for estimating bicyclist safety/risk? Please rank the following choices in order of importance with '1' being the most important data source. If you do not use a particular data source, please use '0' to indicate its rank. If multiple data sources are equally important, please give them the same rank. If you are unsure, please use the option 'Other' to elaborate.**

Table 4. Ranking of different data sources for bicyclist safety/risk estimation

Data Sources	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5		Total
State crash data	100.00%	12	0.00%	0	0.00%	0	0.00%	0	0.00%	0	12
National crash data	0.00%	0	50.00%	3	16.67%	1	33.33%	2	0.00%	0	6
Crash prediction models	0.00%	0	33.33%	1	33.33%	1	33.33%	1	0.00%	0	3
Third party data (e.g. Twitter, bikemap.org)	0.00%	0	28.57%	2	42.86%	3	0.00%	0	28.57%	2	7
Other (please specify)	0.00%	0	100.00%	4	0.00%	0	0.00%	0	0.00%	0	4

7. **What is/are your primary source(s) of data for pedestrian volume estimation? Please rank the following choices in order of importance with '1' being the most important data source. If you do not use a particular data source, please use '0' to indicate its rank. If multiple data sources are equally important, please give them the same rank. If you are unsure, please use the option 'Other' to elaborate.**

Table 5. Ranking of different data sources for pedestrian volume estimation

Data Sources	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5		Rank 6		Total
Manual Counts	60.00%	3	20.00%	1	20.00%	1	0.00%	0	0.00%	0	0.00%	0	5
Automated Counters	28.57%	2	42.86%	3	28.57%	2	0.00%	0	0.00%	0	0.00%	0	7
Travel Surveys	66.67%	2	0.00%	0	0.00%	0	33.33%	1	0.00%	0	0.00%	0	3
Estimation models	0.00%	0	50.00%	2	50.00%	2	0.00%	0	0.00%	0	0.00%	0	4
Third party data	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0

Data Sources	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5		Rank 6		Total
(e.g. Safegraph)													
Other (please specify)	66.67%	2	33.33%	1	0.00%	0	0.00%	0	0.00%	0	0.00%	0	3

8. **What is/are your primary source(s) of data for pedestrian risk estimation? Please rank the following choices in order of importance with '1' being the most important data source. If you do not use a particular data source, please use '0' to indicate its rank. If multiple data sources are equally important, please give them the same rank. If you are unsure, please use the option 'Other' to elaborate.**

Table 6. Ranking of different data sources for pedestrian safety/risk estimation

Data Sources	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5		Total
State crash data	92.86%	13	7.14%	1	0.00%	0	0.00%	0	0.00%	0	14
National crash data	0.00%	0	57.14%	4	14.29%	1	28.57%	2	0.00%	0	7
Crash prediction models	0.00%	0	25.00%	1	50.00%	2	25.00%	1	0.00%	0	4
Third party data (e.g. Streetlight)	0.00%	0	33.33%	2	33.33%	2	0.00%	0	33.33%	2	6
Other (please specify)	20.00%	1	80.00%	4	0.00%	0	0.00%	0	0.00%	0	5

9. **How frequently do you collect/update your primary data source(s)?**

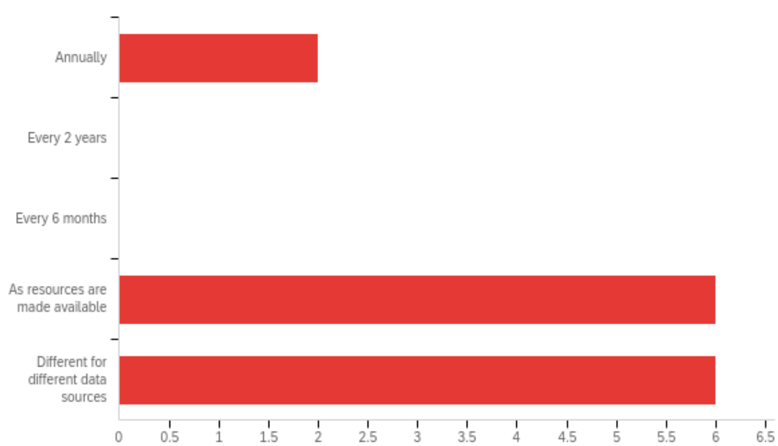


Figure 4. Distribution of respondents for data source update frequency

Additional Comments from Respondents on 'Different for Different Data Sources':

State crash data is updated as soon as reported and processed by state DOT. National crash data is updated annually.
as needed
Traffic Crash data are updated in real time as crash reports are created by LEOs. However, we clean and freeze annually for reporting purposes.
Permanent counter data is reviewed and cleaned once a year. temporary count data is assumed to be truth when it's removed from a count site.
We consistently review data and update our data sources

10. Please answer the following questions as it applies to your current data source(s) for volume estimation.

Table 7. Satisfaction level with volume estimation data sources

Questions	Not at all		Somewhat		Moderately		Mostly		Extremely		Cannot Say		Total
How confident are you in the accuracy of the information this source provides?	0.0%	0	33.3%	3	33.3%	3	22.2%	2	0.0%	0	11.1%	1	9
How useful is the information from this source?	0.0%	0	33.3%	3	22.2%	2	33.3%	3	0.0%	0	11.1%	1	9
How accessible is this information and/or the source of information? (For example, are the data available in public domain?)	0.0%	0	44.4%	4	11.1%	1	11.1%	1	22.2%	2	11.1%	1	9

11. **Please answer the following questions as it applies to your current data source(s) for safety/risk estimation.**

Table 8. Satisfaction level for safety/risk estimation data sources

Question	Not at all		Somewhat		Moderately		Mostly		Extremely		Cannot Say		Total
How confident are you in the accuracy of the information this source provides?	0.0%	0	14.29%	2	14.29%	2	57.14%	8	7.14%	1	7.14%	1	14
How useful is the information from this source?	0.0%	0	7.14%	1	21.43%	3	57.14%	8	7.14%	1	7.14%	1	14
How accessible is this information and/or the source of information? (For example, are the data available in public domain?)	0.0%	0	28.57%	4	35.71%	5	14.29%	2	21.43%	3	0.00%	0	14

Section 3. Crowdsourced Data

12. **Crowd-sourced data are data voluntarily contributed by people directly (e.g. self reported near misses) or indirectly (e.g. trip data from smartphone based GPS tracking) which are often aggregated and shared or sold by third party data vendors. Have you or your organization used crowd-sourced data for bicycle or pedestrian planning purposes?**

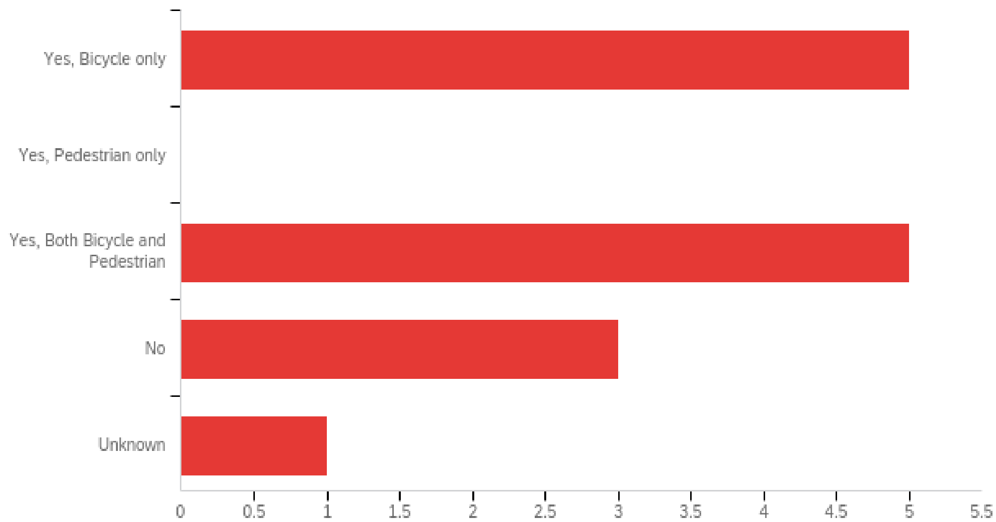


Figure 5. Distribution of respondents across experience with crowdsourced data

13. **The crowd-sourced data used by your organization was obtained/provided by:**

Table 9. List of crowdsourced data provider as reported by respondents

Strava
Strava | Metro
Research - Strava and Streetlight
Public Meeting Discussions
Stravos
STRAVA
Strava and Streetlight
Streetlight and Strava so far

University of South Florida Center for Urban Transportation Research (CUTR) and University of North Florida Public Opinion Research Lab (PORL)

14. Reason(s) for choosing your crowd-sourced data provider

Table 10. List of reasons for choosing a data provider

It was offered to MPOs for free for internal analysis
They provided the service for free to public agencies
Research- investigating uses for datasets within our processes/procedures
Public Meetings are part of our project delivery process.
Was chosen by our division, unsure of reason for selection
First to provide data, cost
Crowd-source data seen as a broad geographic representation of user trends and reasonably representative actual user demand.
Free or low cost or agency was already contracting with the provider
CUTR is an internationally recognized transportation research, education and technology transfer/training/outreach center with the ability to collect both observational data as well as public opinion survey data. This provides the opportunity to verify what percentage of accuracy we are getting in the public opinion data. Similarly, the PORL provides us with survey data from specific target audiences in specified zip codes.

15. **Thinking of volume estimation, how satisfied are you with the crowd-sourced data in the following aspects?**

Table 11. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for volume estimation

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	1.00	3.00	2.00	0.63	0.40	5
Coverage	1.00	3.00	2.00	0.89	0.80	5
Relevance	1.00	3.00	2.00	0.63	0.40	5
Representativeness	1.00	3.00	2.00	0.89	0.80	5
Cost	1.00	3.00	1.60	0.80	0.64	5
Timeliness	1.00	3.00	1.60	0.80	0.64	5
Scalability	1.00	3.00	1.80	0.75	0.56	5
Quality	1.00	3.00	1.80	0.75	0.56	5
Technical support	1.00	3.00	1.80	0.75	0.56	5

Table 12. Frequency Distribution of responses on satisfaction with different aspects of crowdsourced data for volume estimation

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	20.00%	1	60.00%	3	20.00%	1	5
Coverage	40.00%	2	20.00%	1	40.00%	2	5
Relevance	20.00%	1	60.00%	3	20.00%	1	5
Representativeness	40.00%	2	20.00%	1	40.00%	2	5
Cost	60.00%	3	20.00%	1	20.00%	1	5
Timeliness	60.00%	3	20.00%	1	20.00%	1	5
Scalability	40.00%	2	40.00%	2	20.00%	1	5
Quality	40.00%	2	40.00%	2	20.00%	1	5
Technical support	40.00%	2	40.00%	2	20.00%	1	5

16. Thinking of safety/risk estimation, how satisfied are you with the crowd-sourced data in the following aspects?

Table 13. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for safety/risk estimation

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	2.00	2.00	2.00	0.00	0.00	5
Coverage	1.00	3.00	2.00	0.63	0.40	5
Relevance	2.00	2.00	2.00	0.00	0.00	5
Representativeness	2.00	3.00	2.40	0.49	0.24	5
Cost	1.00	2.00	1.60	0.49	0.24	5
Timeliness	1.00	2.00	1.60	0.49	0.24	5
Scalability	1.00	2.00	1.80	0.40	0.16	5
Quality	2.00	2.00	2.00	0.00	0.00	5
Technical support	1.00	2.00	1.80	0.40	0.16	5

Table 14. Frequency distribution of responses on satisfaction with different aspects of crowdsourced data for safety/risk estimation

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	0.00%	0	100.00%	5	0.00%	0	5
Coverage	20.00%	1	60.00%	3	20.00%	1	5
Relevance	0.00%	0	100.00%	5	0.00%	0	5
Representativeness	0.00%	0	60.00%	3	40.00%	2	5
Cost	40.00%	2	60.00%	3	0.00%	0	5
Timeliness	40.00%	2	60.00%	3	0.00%	0	5
Scalability	20.00%	1	80.00%	4	0.00%	0	5
Quality	0.00%	0	100.00%	5	0.00%	0	5
Technical support	20.00%	1	80.00%	4	0.00%	0	5

17. **Thinking of bicycle volume estimation, how satisfied are you with the crowd-sourced data in the following aspects?**

Table 15. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for bicycle volume estimation

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	1.00	2.00	1.75	0.43	0.19	4
Coverage	1.00	3.00	2.00	0.71	0.50	4
Relevance	1.00	2.00	1.75	0.43	0.19	4
Representativeness	2.00	3.00	2.50	0.50	0.25	4
Cost	1.00	3.00	1.75	0.83	0.69	4
Timeliness	1.00	3.00	2.00	0.71	0.50	4
Scalability	1.00	3.00	2.00	0.71	0.50	4
Quality	1.00	2.00	1.75	0.43	0.19	4
Technical support	1.00	2.00	1.75	0.43	0.19	4

Table 16. Frequency distribution of responses on satisfaction with different aspects of crowdsourced data for bicycle volume estimation

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	25.00%	1	75.00%	3	0.00%	0	4
Coverage	25.00%	1	50.00%	2	25.00%	1	4
Relevance	25.00%	1	75.00%	3	0.00%	0	4
Representativeness	0.00%	0	50.00%	2	50.00%	2	4
Cost	50.00%	2	25.00%	1	25.00%	1	4
Timeliness	25.00%	1	50.00%	2	25.00%	1	4
Scalability	25.00%	1	50.00%	2	25.00%	1	4
Quality	25.00%	1	75.00%	3	0.00%	0	4
Technical support	25.00%	1	75.00%	3	0.00%	0	4

18. Thinking of bicycle safety/risk estimation, how satisfied are you with the crowd-sourced data in the following aspects?

Table 17. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for bicycle safety/risk

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	1.00	2.00	1.75	0.43	0.19	4
Coverage	2.00	2.00	2.00	0.00	0.00	4
Relevance	1.00	2.00	1.75	0.43	0.19	4
Representativeness	2.00	2.00	2.00	0.00	0.00	4
Cost	2.00	2.00	2.00	0.00	0.00	4
Timeliness	1.00	2.00	1.75	0.43	0.19	4
Scalability	2.00	2.00	2.00	0.00	0.00	4
Quality	1.00	2.00	1.75	0.43	0.19	4
Technical support	2.00	2.00	2.00	0.00	0.00	4

Table 18. Frequency distribution of responses on satisfaction with different aspects of crowdsourced data for bicycle safety/risk

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	25.00%	1	75.00%	3	0.00%	0	4
Coverage	0.00%	0	100.00%	4	0.00%	0	4
Relevance	25.00%	1	75.00%	3	0.00%	0	4
Representativeness	0.00%	0	100.00%	4	0.00%	0	4
Cost	0.00%	0	100.00%	4	0.00%	0	4
Timeliness	25.00%	1	75.00%	3	0.00%	0	4
Scalability	0.00%	0	100.00%	4	0.00%	0	4
Quality	25.00%	1	75.00%	3	0.00%	0	4
Technical support	0.00%	0	100.00%	4	0.00%	0	4

19. **Thinking of pedestrian volume estimation, how satisfied are you with the crowd-sourced data in the following aspects?**

Table 19. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for pedestrian volume estimation

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	2.00	3.00	2.25	0.43	0.19	4
Coverage	2.00	3.00	2.50	0.50	0.25	4
Relevance	2.00	3.00	2.25	0.43	0.19	4
Representativeness	2.00	3.00	2.50	0.50	0.25	4
Cost	1.00	3.00	2.00	0.71	0.50	4
Timeliness	2.00	3.00	2.25	0.43	0.19	4
Scalability	2.00	3.00	2.25	0.43	0.19	4
Quality	2.00	2.00	2.00	0.00	0.00	4
Technical support	2.00	2.00	2.00	0.00	0.00	4

Table 20. Frequency distribution of responses on satisfaction with different aspects of crowdsourced data for pedestrian volume estimation

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	0.00%	0	75.00%	3	25.00%	1	4
Coverage	0.00%	0	50.00%	2	50.00%	2	4
Relevance	0.00%	0	75.00%	3	25.00%	1	4
Representativeness	0.00%	0	50.00%	2	50.00%	2	4
Cost	25.00%	1	50.00%	2	25.00%	1	4
Timeliness	0.00%	0	75.00%	3	25.00%	1	4
Scalability	0.00%	0	75.00%	3	25.00%	1	4
Quality	0.00%	0	100.00%	4	0.00%	0	4
Technical support	0.00%	0	100.00%	4	0.00%	0	4

20. Thinking of pedestrian safety/risk estimation, how satisfied are you with the crowd-sourced data in the following aspects?

Table 21. Basic statistics of responses on satisfaction with different aspects of crowdsourced data for pedestrian safety/risk estimation

Data Aspects	Minimum	Maximum	Mean	Std Deviation	Variance	Count
Accuracy	1.00	2.00	1.75	0.43	0.19	4
Coverage	1.00	2.00	1.75	0.43	0.19	4
Relevance	1.00	2.00	1.75	0.43	0.19	4
Representativeness	1.00	2.00	1.75	0.43	0.19	4
Cost	2.00	2.00	2.00	0.00	0.00	4
Timeliness	1.00	2.00	1.75	0.43	0.19	4
Scalability	2.00	2.00	2.00	0.00	0.00	4
Quality	1.00	2.00	1.75	0.43	0.19	4
Technical support	2.00	2.00	2.00	0.00	0.00	4

Table 22. Frequency distribution of responses on satisfaction with different aspects of crowdsourced data for pedestrian safety/risk

Data Aspects	Satisfied		Neither satisfied nor dissatisfied		Dissatisfied		Total
Accuracy	25.00%	1	75.00%	3	0.00%	0	4
Coverage	25.00%	1	75.00%	3	0.00%	0	4
Relevance	25.00%	1	75.00%	3	0.00%	0	4
Representativeness	25.00%	1	75.00%	3	0.00%	0	4
Cost	0.00%	0	100.00%	4	0.00%	0	4
Timeliness	25.00%	1	75.00%	3	0.00%	0	4
Scalability	0.00%	0	100.00%	4	0.00%	0	4
Quality	25.00%	1	75.00%	3	0.00%	0	4
Technical support	0.00%	0	100.00%	4	0.00%	0	4

21. Have you considered using crowd-sourced data for bicyclist and pedestrian planning?

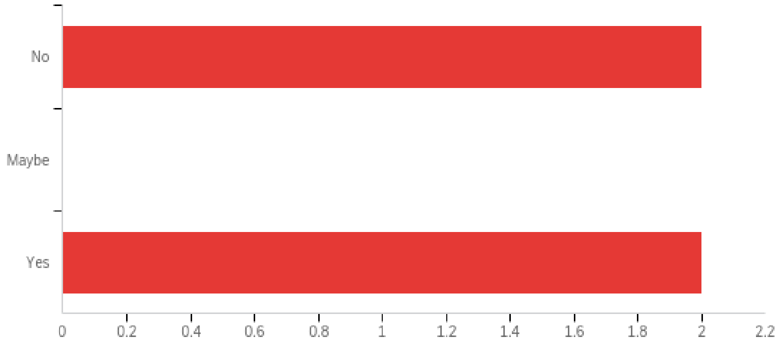


Figure 6. Distribution of responses on use of crowdsourced data for participants not currently using crowdsourced data

22. Which crowd-sourced data source(s) did you consider?

Table 23. List of crowdsourced data sources considered by respondents

Streetlight, Ford Insight
Streetlight & Strava
Streetlight and Strava
Strava and Streetlight
Public Comments in General
Streetlight, HERE
Strava
Not chosen by me
Strava and Streetlight
Streetlight and Strava
ESRI, GIS, Tableau, Qualtrix, Social Networking/Social Media Metrics, Collaborative Websites, Partnerships/Coalitions, etc.

23. **Please indicate reasons for which you may not have used or would not recommend using crowd-sourced data based on your experience. Please choose all that apply and use the text boxes to elaborate if needed.**

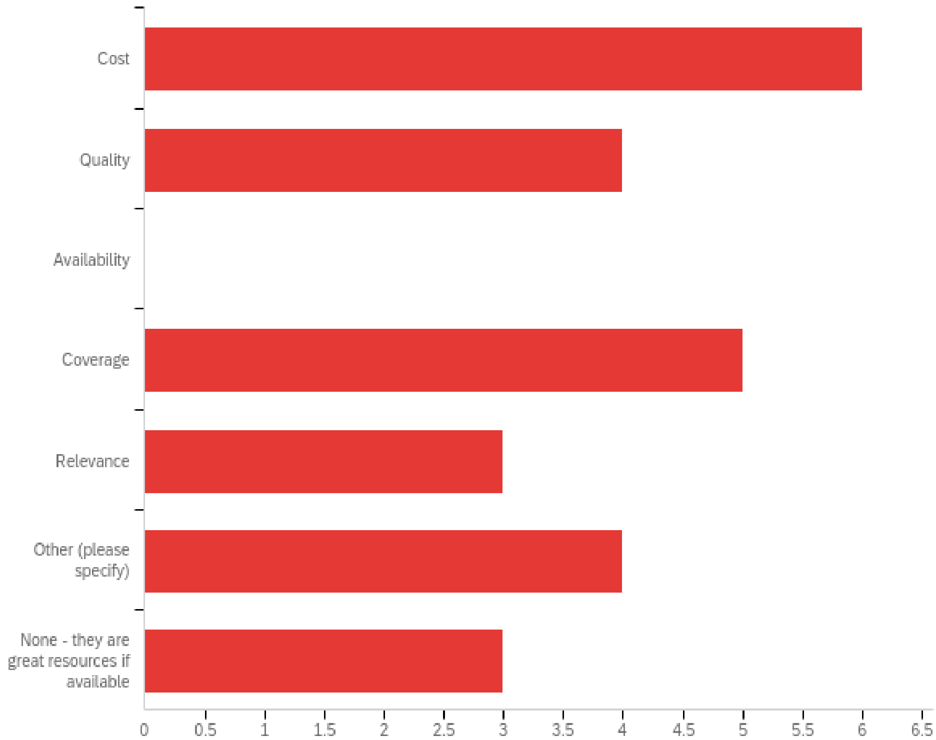


Figure 7. Distribution of responses across different reasons for not using crowdsourced data

Table 24. Frequency distribution of respondents across reasons for not using crowdsourced data

Reasons for not using crowdsourced data	%	Count
Cost	24.00%	6
Quality	16.00%	4
Availability	0.00%	0
Coverage	20.00%	5
Relevance	12.00%	3
Other (please specify)	16.00%	4
None - they are great resources if available	12.00%	3
Total	100%	25

Details on reasons:

Cost - Text

The cost for Streetlight is more than we as a municipality can offered.

Quote costs have been near $500K per year for crowd-sourced bike/ped data

Our studies were conducted some time ago, not sure about cost currently

Resources are limited, so manual counts are used based on funding availability

Quality - Text

These emerging data sets are still be studied and benchmarked against other data sets to verify validity. We know that the data set is only a sample and can only be used in a self-comparative manner and not an absolute volume method.

Self reporting will not ensure comprehensive counts

does count all users

Coverage - Text

Crowd-sourced data has more data points in more populous areas and limits the utility of these data sets mostly to urban areas.

Seemed mostly available in larger urban areas

often great in densely populated areas, but you'll run into issues as you look in more rural areas where the companies try to protect anonymity resulting in very little useable data.

Relevance - Text

Although we got Strava for free. We recognize that there are limitations to the data. First, the users of the app are probably not representative of Lincoln's user base. Also BIPOC populations and low-income populations may not be shown in the data.

Concerns with static counts, pedestrian/bicycle volumes are largely dependent on landuse

misses non recreational users

Other (please specify) - Text

Ford Insight was a external selection process

Not sure if there is enough usage to be beneficial in my area

We don't collect volume data.

Streetlight- Due to pricing structure, Streetlight must be used only for targeted/project-specific use cases

Appendix C Data Description and Python Code

Field Name	Variable Type	NITC Variable Name	Description
id	N/A	N/A	Unique ID- corresponds to Location Number in AADT/Strava data that Mike put together
location	N/A	N/A	Location name
latitude	N/A	N/A	Latitude of the location
longitude	N/A	N/A	Longitude of the location
region	N/A	N/A	Usually the county of the location, though sometimes multiple counties (in the case the location was near a county border)- largely used to determine the scope of OSM data to download
weathercty	N/A	N/A	The city used to gather weather-related variables for the location
primary	Environmental	Primary	Total length of primary road segments within the buffer around each count station
secondary	Environmental	Secondary	Total length of secondary road segments within the buffer around each count station
tertiary	Environmental	Tertiary	Total length of tertiary road segments within the buffer around each count station
resi_road	Environmental	Residential	Total length of residential road segments within the buffer around each count station
path	Environmental	Path	Total length of path segments within the buffer around each count station
cycleway	Environmental	Cycleway	Total length of cycle way segments within the buffer around each count station
cwy_lneall	Environmental	Cycleway_lane_all	Total length of cycleway lane, left and right segments within the buffer around each count station
cwy_binary	Environmental	Cycleway_lane_all	Link type (0=absence and 1= presence) of cycleway lane, left and right segments within the buffer around each count station
cwy_trkall	Environmental	Cycleway_track_all	Total length of cycleway track, left and right segments within the buffer around each count station
cwy_trkbin	Environmental	Cycleway_track_all	Link type (0=absence and 1= presence) of cycleway track, left and right segments within the buffer around each count station
footway	Environmental	Footway	Total length of footway segments within the buffer around each count station
meanspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
pointspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
bikeprking	Environmental	Number of Bicycle Parking Spaces	Count of bicycle parking spots within the buffer around the count station
bus stops	Environmental	Number of Bus stops	Count of bus or rail stops within the buffer around the count station
intd_hlfmi	Environmental	Intersection Density	Number of intersections per square mile, within a half mile buffer of the count station
intd_1mi	Environmental	Intersection Density	Number of intersections per square mile, within a one mile buffer of the count station
lanes	Environmental	Number of lanes	Number of traffic lanes along corresponding count station street segment
bridge	Environmental	Presence of Bridges	Binary variable: 1=presence and 2=absence of bridges within the buffer around the count station
water area	Environmental	Water Body area or Distance to water body	Water body area within the buffer around the count station
water dist	Environmental	Water Body area or Distance to water body	nearest distance to edge of water body from the count station
wtr cntdis	Environmental	Water Body area or Distance to water body	nearest distance to centroid of water body from the count station
park area	Environmental	Park Area or Distance to park	Park or open space area within the buffer around the count station
park dist	Environmental	Park Area or Distance to park	nearest distance to edge of park or open space from the count station\
prk cntdis	Environmental	Park Area or Distance to park	nearest distance to centroid of park or open space from the count station
frest area	Environmental	Forest Area	Forest area within the buffer around the count station
forest dis	Environmental	Forest Area	nearest distance to edge of forest from the count station (0 indicates the counter is within a forest)
frst cntdi	Environmental	Forest Area	nearest distance to centroid of forest from the count station
grass area	Environmental	Grass area or Distance to grass area	Grass area within the buffer around the count station
grass dist	Environmental	Grass area or Distance to grass area	nearest distance to edge of grass from the count station
grs cntdis	Environmental	Grass area or Distance to grass area	nearest distance to centroid of grass from the count station
comm area	Environmental	Commercial area or Distance to commercial	Commercial area within the buffer around the count station
comm dist	Environmental	Commercial area or Distance to commercial	nearest distance to edge of commercial area from the count station
comm cntdi	Environmental	Commercial area or Distance to commercial	nearest distance to centroid of commercial area from the count station
ind area	Environmental	Industrial Area or Distance to industrial area	Industrial area within the buffer around the count station
ind dist	Environmental	Industrial Area or Distance to industrial area	nearest distance to edge of industrial area from the count station
ind cntdis	Environmental	Industrial Area or Distance to industrial area	nearest distance to centroid of industrial area from the count station
resi area	Environmental	Residential Area or Distance to residential area	Residential area within the buffer around the count station
resi dist	Environmental	Residential Area or Distance to residential area	nearest distance to edge of residential area from the count station
resi cntdi	Environmental	Residential Area or Distance to residential area	nearest distance to centroid of residential area from the count station

Field Name	Variable Type	NITC Variable Name	Description
ret area	Environmental	Retail Area or Distance to Retail area	Retail area within the buffer around the count station
Field Name	Variable Type	NITC Variable Name	Description
id	N/A	N/A	Unique ID- corresponds to Location Number in AADT/Strava data that Mike put together
location	N/A	N/A	Location name
latitude	N/A	N/A	Latitude of the location
longitude	N/A	N/A	Longitude of the location
region	N/A	N/A	Usually the county of the location, though sometimes multiple counties (in the case the location was near a county border)- largely used to determine the scope of OSM data to download
weathercty	N/A	N/A	The city used to gather weather-related variables for the location
primary	Environmental	Primary	Total length of primary road segments within the buffer around each count station
secondary	Environmental	Secondary	Total length of secondary road segments within the buffer around each count station
tertiary	Environmental	Tertiary	Total length of tertiary road segments within the buffer around each count station
resi_road	Environmental	Residential	Total length of residential road segments within the buffer around each count station
path	Environmental	Path	Total length of path segments within the buffer around each count station
cycleway	Environmental	Cycleway	Total length of cycle way segments within the buffer around each count station
cwy_lneall	Environmental	Cycleway_lane_all	Total length of cycleway lane, left and right segments within the buffer around each count station
cwy_binary	Environmental	Cycleway_lane_all	Link type (0=absence and 1= presence) of cycleway lane, left and right segments within the buffer around each count station
cwy_trkall	Environmental	Cycleway_track_all	Total length of cycleway track, left and right segments within the buffer around each count station
cwy_trkbin	Environmental	Cycleway_track_all	Link type (0=absence and 1= presence) of cycleway track, left and right segments within the buffer around each count station
footway	Environmental	Footway	Total length of footway segments within the buffer around each count station
meanspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
pointspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
bikeprking	Environmental	Number of Bicycle Parking Spaces	Count of bicycle parking spots within the buffer around the count station
bus stops	Environmental	Number of Bus stops	Count of bus or rail stops within the buffer around the count station
intd_hlfmi	Environmental	Intersection Density	Number of intersections per square mile, within a half mile buffer of the count station
intd_1mi	Environmental	Intersection Density	Number of intersections per square mile, within a one mile buffer of the count station
lanes	Environmental	Number of lanes	Number of traffic lanes along corresponding count station street segment
bridge	Environmental	Presence of Bridges	Binary variable: 1=presence and 2=absence of bridges within the buffer around the count station
water area	Environmental	Water Body area or Distance to water body	Water body area within the buffer around the count station
water dist	Environmental	Water Body area or Distance to water body	nearest distance to edge of water body from the count station
wtr cntdis	Environmental	Water Body area or Distance to water body	nearest distance to centroid of water body from the count station
park area	Environmental	Park Area or Distance to park	Park or open space area within the buffer around the count station
park dist	Environmental	Park Area or Distance to park	nearest distance to edge of park or open space from the count station\
prk cntdis	Environmental	Park Area or Distance to park	nearest distance to centroid of park or open space from the count station
frest area	Environmental	Forest Area	Forest area within the buffer around the count station
forest dis	Environmental	Forest Area	nearest distance to edge of forest from the count station (0 indicates the counter is within a forest)
frst cntdi	Environmental	Forest Area	nearest distance to centroid of forest from the count station
grass area	Environmental	Grass area or Distance to grass area	Grass area within the buffer around the count station
grass dist	Environmental	Grass area or Distance to grass area	nearest distance to edge of grass from the count station
grs cntdis	Environmental	Grass area or Distance to grass area	nearest distance to centroid of grass from the count station
comm area	Environmental	Commercial area or Distance to commercial	Commercial area within the buffer around the count station
comm dist	Environmental	Commercial area or Distance to commercial	nearest distance to edge of commercial area from the count station
comm cntdi	Environmental	Commercial area or Distance to commercial	nearest distance to centroid of commercial area from the count station
ind area	Environmental	Industrial Area or Distance to industrial area	Industrial area within the buffer around the count station
ind dist	Environmental	Industrial Area or Distance to industrial area	nearest distance to edge of industrial area from the count station
ind cntdis	Environmental	Industrial Area or Distance to industrial area	nearest distance to centroid of industrial area from the count station
resi area	Environmental	Residential Area or Distance to residential area	Residential area within the buffer around the count station

Field Name	Variable Type	NITC Variable Name	Description
resi dist	Environmental	Residential Area or Distance to residential area	nearest distance to edge of residential area from the count station
resi cntdi	Environmental	Residential Area or Distance to residential area	nearest distance to centroid of residential area from the count station
ret area	Environmental	Retail Area or Distance to Retail area	Retail area within the buffer around the count station
Field Name	Variable Type	NITC Variable Name	Description
id	N/A	N/A	Unique ID- corresponds to Location Number in AADT/Strava data that Mike put together
location	N/A	N/A	Location name
latitude	N/A	N/A	Latitude of the location
longitude	N/A	N/A	Longitude of the location
region	N/A	N/A	Usually the county of the location, though sometimes multiple counties (in the case the location was near a county border)- largely used to determine the scope of OSM data to download
weathercty	N/A	N/A	The city used to gather weather-related variables for the location
primary	Environmental	Primary	Total length of primary road segments within the buffer around each count station
secondary	Environmental	Secondary	Total length of secondary road segments within the buffer around each count station
tertiary	Environmental	Tertiary	Total length of tertiary road segments within the buffer around each count station
resi_road	Environmental	Residential	Total length of residential road segments within the buffer around each count station
path	Environmental	Path	Total length of path segments within the buffer around each count station
cycleway	Environmental	Cycleway	Total length of cycle way segments within the buffer around each count station
cwy_lneall	Environmental	Cycleway_lane_all	Total length of cycleway lane, left and right segments within the buffer around each count station
cwy_binary	Environmental	Cycleway_lane_all	Link type (0=absence and 1= presence) of cycleway lane, left and right segments within the buffer around each count station
cwy_trkall	Environmental	Cycleway_track_all	Total length of cycleway track, left and right segments within the buffer around each count station
cwy_trkbin	Environmental	Cycleway_track_all	Link type (0=absence and 1= presence) of cycleway track, left and right segments within the buffer around each count station
footway	Environmental	Footway	Total length of footway segments within the buffer around each count station
meanspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
pointspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
bikeprking	Environmental	Number of Bicycle Parking Spaces	Count of bicycle parking spots within the buffer around the count station
bus stops	Environmental	Number of Bus stops	Count of bus or rail stops within the buffer around the count station
intd_hlfmi	Environmental	Intersection Density	Number of intersections per square mile, within a half mile buffer of the count station
intd_1mi	Environmental	Intersection Density	Number of intersections per square mile, within a one mile buffer of the count station
lanes	Environmental	Number of lanes	Number of traffic lanes along corresponding count station street segment
bridge	Environmental	Presence of Bridges	Binary variable: 1=presence and 2=absence of bridges within the buffer around the count station
water area	Environmental	Water Body area or Distance to water body	Water body area within the buffer around the count station
water dist	Environmental	Water Body area or Distance to water body	nearest distance to edge of water body from the count station
wtr cntdis	Environmental	Water Body area or Distance to water body	nearest distance to centroid of water body from the count station
park area	Environmental	Park Area or Distance to park	Park or open space area within the buffer around the count station
park dist	Environmental	Park Area or Distance to park	nearest distance to edge of park or open space from the count station\
prk cntdis	Environmental	Park Area or Distance to park	nearest distance to centroid of park or open space from the count station
frest area	Environmental	Forest Area	Forest area within the buffer around the count station
forest dis	Environmental	Forest Area	nearest distance to edge of forest from the count station (0 indicates the counter is within a forest)
frst cntdi	Environmental	Forest Area	nearest distance to centroid of forest from the count station
grass area	Environmental	Grass area or Distance to grass area	Grass area within the buffer around the count station
grass dist	Environmental	Grass area or Distance to grass area	nearest distance to edge of grass from the count station
grs cntdis	Environmental	Grass area or Distance to grass area	nearest distance to centroid of grass from the count station
comm area	Environmental	Commercial area or Distance to commercial	Commercial area within the buffer around the count station
comm dist	Environmental	Commercial area or Distance to commercial	nearest distance to edge of commercial area from the count station
comm cntdi	Environmental	Commercial area or Distance to commercial	nearest distance to centroid of commercial area from the count station
ind area	Environmental	Industrial Area or Distance to industrial area	Industrial area within the buffer around the count station
ind dist	Environmental	Industrial Area or Distance to industrial area	nearest distance to edge of industrial area from the count station

Field Name	**Variable Type**	**NITC Variable Name**	**Description**
ind cntdis	Environmental	Industrial Area or Distance to industrial area	nearest distance to centroid of industrial area from the count station
resi area	Environmental	Residential Area or Distance to residential area	Residential area within the buffer around the count station

Field Name	Variable Type	NITC Variable Name	Description
id	N/A	N/A	Unique ID- corresponds to Location Number in AADT/Strava data that Mike put together
location	N/A	N/A	Location name
latitude	N/A	N/A	Latitude of the location
longitude	N/A	N/A	Longitude of the location
region	N/A	N/A	Usually the county of the location, though sometimes multiple counties (in the case the location was near a county border)- largely used to determine the scope of OSM data to download
weathercty	N/A	N/A	The city used to gather weather-related variables for the location
primary	Environmental	Primary	Total length of primary road segments within the buffer around each count station
secondary	Environmental	Secondary	Total length of secondary road segments within the buffer around each count station
tertiary	Environmental	Tertiary	Total length of tertiary road segments within the buffer around each count station
resi_road	Environmental	Residential	Total length of residential road segments within the buffer around each count station
path	Environmental	Path	Total length of path segments within the buffer around each count station
cycleway	Environmental	Cycleway	Total length of cycle way segments within the buffer around each count station
cwy_lneall	Environmental	Cycleway_lane_all	Total length of cycleway lane, left and right segments within the buffer around each count station
cwy_binary	Environmental	Cycleway_lane_all	Link type (0=absence and 1= presence) of cycleway lane, left and right segments within the buffer around each count station
cwy_trkall	Environmental	Cycleway_track_all	Total length of cycleway track, left and right segments within the buffer around each count station
cwy_trkbin	Environmental	Cycleway_track_all	Link type (0=absence and 1= presence) of cycleway track, left and right segments within the buffer around each count station
footway	Environmental	Footway	Total length of footway segments within the buffer around each count station
meanspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
pointspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
bikeprking	Environmental	Number of Bicycle Parking Spaces	Count of bicycle parking spots within the buffer around the count station
bus stops	Environmental	Number of Bus stops	Count of bus or rail stops within the buffer around the count station
intd_hlfmi	Environmental	Intersection Density	Number of intersections per square mile, within a half mile buffer of the count station
intd_1mi	Environmental	Intersection Density	Number of intersections per square mile, within a one mile buffer of the count station
lanes	Environmental	Number of lanes	Number of traffic lanes along corresponding count station street segment
bridge	Environmental	Presence of Bridges	Binary variable: 1=presence and 2=absence of bridges within the buffer around the count station
water area	Environmental	Water Body area or Distance to water body	Water body area within the buffer around the count station

Field Name	Variable Type	NITC Variable Name	Description
id	N/A	N/A	Unique ID- corresponds to Location Number in AADT/Strava data that Mike put together
location	N/A	N/A	Location name
latitude	N/A	N/A	Latitude of the location
longitude	N/A	N/A	Longitude of the location
region	N/A	N/A	Usually the county of the location, though sometimes multiple counties (in the case the location was near a county border)- largely used to determine the scope of OSM data to download
weathercty	N/A	N/A	The city used to gather weather-related variables for the location
primary	Environmental	Primary	Total length of primary road segments within the buffer around each count station
secondary	Environmental	Secondary	Total length of secondary road segments within the buffer around each count station
tertiary	Environmental	Tertiary	Total length of tertiary road segments within the buffer around each count station
resi_road	Environmental	Residential	Total length of residential road segments within the buffer around each count station
path	Environmental	Path	Total length of path segments within the buffer around each count station
cycleway	Environmental	Cycleway	Total length of cycle way segments within the buffer around each count station
cwy_lneall	Environmental	Cycleway_lane_all	Total length of cycleway lane, left and right segments within the buffer around each count station
cwy_binary	Environmental	Cycleway_lane_all	Link type (0=absence and 1= presence) of cycleway lane, left and right segments within the buffer around each count station
cwy_trkall	Environmental	Cycleway_track_all	Total length of cycleway track, left and right segments within the buffer around each count station
cwy_trkbin	Environmental	Cycleway_track_all	Link type (0=absence and 1= presence) of cycleway track, left and right segments within the buffer around each count station
footway	Environmental	Footway	Total length of footway segments within the buffer around each count station
meanspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.
pointspeed	Environmental	Speed limit	Speed limit on the link where the counter is situated; if unavailable the speed of the nearest link with the same functional class is extracted. For average speed within the buffer, the mode of speed for the functional class was obtained.

Field Name	Variable Type	NITC Variable Name	Description
bikeprking	Environmental	Number of Bicycle Parking Spaces	Count of bicycle parking spots within the buffer around the count station
bus stops	Environmental	Number of Bus stops	Count of bus or rail stops within the buffer around the count station
intd_hlfmi	Environmental	Intersection Density	Number of intersections per square mile, within a half mile buffer of the count station
intd_1mi	Environmental	Intersection Density	Number of intersections per square mile, within a one mile buffer of the count station
lanes	Environmental	Number of lanes	Number of traffic lanes along corresponding count station street segment
bridge	Environmental	Presence of Bridges	Binary variable: 1=presence and 2=absence of bridges within the buffer around the count station
water area	Environmental	Water Body area or Distance to water body	Water body area within the buffer around the count station
water dist	Environmental	Water Body area or Distance to water body	nearest distance to edge of water body from the count station
wtr cntdis	Environmental	Water Body area or Distance to water body	nearest distance to centroid of water body from the count station
park area	Environmental	Park Area or Distance to park	Park or open space area within the buffer around the count station
park dist	Environmental	Park Area or Distance to park	nearest distance to edge of park or open space from the count station\
prk cntdis	Environmental	Park Area or Distance to park	nearest distance to centroid of park or open space from the count station
frest area	Environmental	Forest Area	Forest area within the buffer around the count station
forest dis	Environmental	Forest Area	nearest distance to edge of forest from the count station (0 indicates the counter is within a forest)
frst cntdi	Environmental	Forest Area	nearest distance to centroid of forest from the count station
grass area	Environmental	Grass area or Distance to grass area	Grass area within the buffer around the count station
grass dist	Environmental	Grass area or Distance to grass area	nearest distance to edge of grass from the count station
grs cntdis	Environmental	Grass area or Distance to grass area	nearest distance to centroid of grass from the count station
comm area	Environmental	Commercial area or Distance to commercial	Commercial area within the buffer around the count station
comm dist	Environmental	Commercial area or Distance to commercial	nearest distance to edge of commercial area from the count station
comm cntdi	Environmental	Commercial area or Distance to commercial	nearest distance to centroid of commercial area from the count station
ind area	Environmental	Industrial Area or Distance to industrial area	Industrial area within the buffer around the count station
ind dist	Environmental	Industrial Area or Distance to industrial area	nearest distance to edge of industrial area from the count station
ind cntdis	Environmental	Industrial Area or Distance to industrial area	nearest distance to centroid of industrial area from the count station
resi area	Environmental	Residential Area or Distance to residential area	Residential area within the buffer around the count station

A PDF of the Python coding can be found at this link:

https://mdotjboss.state.mi.us/SpecProv/getDocumentById.htm?docGuid=d3c3233a-874a-4c85-a563-485f7fecd61a

Appendix D: Additional Model Results, Correlation Table, Random Forest Node Purity Measures

Results of MADT Models without Strava Data

Table 1. Model Results for MADT Bicycle without Strava Data

Random effects	**:**				
Groups	Name	Variance	Std. Dev		
Location Number	(Intercept)	0.3325	0.5767		
Year	(Intercept)	0.0164	0.1281		
Fixed effects:					
Coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	**4.53**	0.70	6.44	1.22E-10	***
Residential Roadmiles	**0.01**	0.00	1.67	0.094824	.
Distance to nearest park center	**-1.45**	0.53	-2.75	0.005913	**
Distance to nearest grass area center	-0.07	0.12	-0.59	0.55574	
Household density per acre	**0.52**	0.12	4.32	1.56E-05	***
Distance to nearest university	-0.16	0.10	-1.63	0.104025	
Distance to nearest college	**0.15**	0.08	1.84	0.065622	.
Median household income/$10k	-0.01	0.03	-0.55	0.58514	
Percentage Bike commuters	-0.01	0.03	-0.21	0.837477	
Number of bus stops	0.00	0.01	-0.46	0.64302	
Distance to CBD	0.07	0.05	1.26	0.209524	
Percentage Education level at least college	0.00	0.01	-0.10	0.919288	
Month June	**0.28**	0.06	4.90	9.50E-07	***
Month July	**0.34**	0.06	5.77	7.76E-09	***
Month August	**0.31**	0.06	5.25	1.54E-07	***
Month September	**0.22**	0.06	3.75	0.000177	***
Signif. codes:	0 ‘***’ 0.	001 ‘**’ 0	.01 ‘*’	0.05 ‘.’	0.1 ‘ ’ 1

Table 2. Model Results for MADT Pedestrian without Strava Data

Random effects					
Groups	Name	Variance	Std.Dev.		
Location_Number	(Intercept)	0.7168	0.8467		
Year	(Intercept)	0.2119	0.4604		
Fixed effects					
Coefficients	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	2.85	1.35	2.11	0.03	*
Distance to nearest grass area center	-0.29	0.13	-2.33	0.02	*
Distance to School	-0.93	0.55	-1.68	0.09	.
Number of Schools	-0.23	0.14	-1.64	0.10	
Percentage Bike commuters	0.25	0.06	4.38	1.17E-05	***
Number of bus stops	0.01	0.01	0.90	0.370	
Median Age	-0.04	0.02	-2.30	0.022	*
Percentage White	0.05	0.01	4.68	0.000	***
Population Density/acre	0.20	0.08	2.509	0.012	*
Signif. codes:	0 ‘***’ 0.	001 ‘**’ 0	.01 ‘*’	0.05 ‘.’	0.1 ‘ ’ 1

	Strava_MADT	primary_mi	secondary_mi	tertiary_mi	resi_road_mi	path_mi	cycleway_mi	cwy_lneall_mi	cwy_binary	cwy_trkall_mi	cwy_trkbin	footway_mi	meanspeed	bikeprking	bus_stops
Strava_MADT	1														
primary_mi	-0.256446001	1													
secondary_mi	-0.138513117	0.460240853	1												
tertiary_mi	-0.475833296	0.771430089	0.622369856	1											
resi_road_mi	-0.178475706	0.55806637	0.758119907	0.711433627	1										
path_mi	0.468934633	-0.216323074	-0.090442782	-0.397259404	-0.126150134	1									
cycleway_mi	0.06547741	-0.545390311	-0.440552584	-0.230614489	-0.321014875	0.076358825	1								
cwy_lneall_mi	-0.135621259	0.803747544	0.285553061	0.677971168	0.329256207	-0.109414354	-0.143561261	1							
cwy_binary	0.04144274	0.429163794	0.228013049	0.45203678	0.276557807	-0.281260534	-0.048659827	0.602237771	1						
cwy_trkall_mi	-0.458872168	0.851331874	0.29477522	0.6809039	0.316284384	-0.327579513	-0.507950357	0.697168138	0.320220226	1					
cwy_trkbin	-0.270481916	0.84114183	0.459915485	0.726091243	0.451896411	-0.073184285	-0.490776794	0.698163562	0.371054119	0.863001405	1				
footway_mi	-0.240401027	0.90160237	0.50806234	0.840404798	0.529013149	-0.218208305	-0.30956368	0.944559226	0.568260258	0.769278835	0.805522914	1			
meanspeed	0.124095763	-0.332510635	-0.207019566	-0.271488699	-0.15857161	-0.281966014	0.313135583	-0.328080916	0.176382024	-0.239407035	-0.229174365	-0.35896492	1		
bikeprking	-0.269500763	0.709153249	0.307048835	0.666014648	0.673646156	-0.09147218	-0.251950694	0.527652071	0.336177285	0.647368809	0.729970754	0.618142997	-0.138897887	1	
bus_stops	-0.296501155	0.890223048	0.196343049	0.716221857	0.251293303	-0.214042184	-0.305005336	0.859636855	0.423583744	0.811839824	0.815504558	0.888792585	-0.348899895	0.591265013	1
intd_1mi_acres	-0.242842022	0.740412547	0.347989214	0.761675323	0.351414053	-0.198151359	-0.0437064	0.954092958	0.590780016	0.642011728	0.668305013	0.937298106	-0.289126321	0.453261939	0.839593504
lanes	-0.163052562	0.380855821	-0.022065791	0.17804626	-0.235196968	0.02709401	-0.270321051	0.284969286	0.19295973	0.388270535	0.451118406	0.307631674	-0.152286889	0.047005106	0.548136943
water_area_acres	0.084632675	-0.456866953	-0.26725057	-0.467798867	-0.340411899	0.341472595	0.229937799	-0.556782742	-0.734694054	-0.346134712	-0.3955941	-0.550189591	-0.05157888	-0.339271603	-0.426483393
wtr_cntdis_mi	0.159979373	0.204017678	0.568044931	0.131588897	0.361820405	-0.023107333	-0.484216262	-0.012253745	-0.012082228	0.010247143	0.130653889	0.120226762	-0.060378701	-0.090830498	-0.051869203
park_area_acres	0.090129831	-0.560679498	-0.294868182	-0.528511193	-0.370348758	0.449842706	0.168803394	-0.639593921	-0.845495266	-0.427169662	-0.46079366	-0.642555269	-0.247312456	-0.410642089	-0.50709962
prk_cntdis_mi	0.116510409	-0.30958562	-0.161452193	-0.374891087	-0.181719803	0.432368865	-0.092680938	-0.375628762	-0.63130051	-0.324588253	-0.384312895	-0.399866686	-0.375300114	-0.298679167	-0.328829911
frest_area_acres	0.074575654	-0.321592379	-0.181883238	-0.348036865	-0.227727715	0.141158693	0.329086967	-0.319759251	-0.359384952	-0.242113749	-0.276763764	-0.36120962	0.372339158	-0.118053045	-0.279158675
frst_cntdi_mi	-0.326747818	-0.264987124	-0.265187456	0.061498029	-0.200764288	-0.150166457	0.480600071	-0.031801582	-0.373023578	-0.139226433	-0.232590876	-0.079259663	-0.298647269	-0.183575131	-0.0300342
grass_area_acres	0.27694484	0.453243008	0.374926483	0.372899809	0.569296364	0.349617739	0.009443382	0.649556619	0.340429626	0.152181088	0.356319951	0.603185848	-0.313614799	0.446016938	0.352987165
grs_cntdis_mi	-0.196515323	-0.394159464	-0.250130317	-0.388728374	-0.262157901	-0.000904019	0.103370861	-0.513605753	-0.627280441	-0.299064735	-0.347630388	-0.504760884	-0.033007216	-0.317221672	-0.398003941
comm_area_acres	-0.183554425	0.715590655	0.342647368	0.539849083	0.085147332	-0.191153356	-0.44244536	0.648756269	0.327978254	0.61352997	0.690669645	0.726814356	-0.296824522	0.218423838	0.835083622
comm_cntdi_mi	-0.095656014	-0.480154194	-0.266268738	-0.508476959	-0.314901905	0.302273167	0.053029086	-0.641392363	-0.908501713	-0.362869144	-0.425523399	-0.627206534	-0.229293753	-0.363781407	-0.484640945
ind_area_acres	0.491207759	0.110113479	0.205436365	-0.058198595	0.294983465	0.268166694	-0.002539728	0.238379097	0.184582413	-0.31785677	-0.239571748	0.169905596	-0.161619455	-0.12163736	-0.061507292
ind_cntdis_mi	-0.208897174	-0.389696104	-0.25365032	-0.4190448	-0.320413066	0.273815321	-0.037675975	-0.54671967	-0.754535851	-0.248494425	-0.328348049	-0.52845064	-0.400095807	-0.299497736	-0.361719722
resi_area_acres	-0.085246803	-0.185726362	-0.048402458	-0.146699424	-0.047475851	0.264608885	0.176224806	-0.101445532	0.23021932	-0.153449977	-0.078594823	-0.176149304	0.443676356	-0.066273789	-0.206593968
resi_cntdi_mi	-0.059047192	-0.342267183	-0.246820964	-0.392461619	-0.272108421	0.001890948	0.027633329	-0.494646385	-0.655403784	-0.293873213	-0.352433801	-0.482602882	-0.07048165	-0.33311171	-0.34718198
ret_area_acres	0.24221476	0.583246955	0.359626733	0.425344627	0.532861273	0.288543771	-0.101754401	0.657614624	0.374980912	0.320501158	0.456398763	0.620711469	-0.242716385	0.531088289	0.421283137
ret_cntdis_mi	-0.018560413	-0.540352625	-0.315034029	-0.564056858	-0.36906837	0.288363241	0.157336159	-0.698872271	-0.89001777	-0.402832602	-0.472127239	-0.691311101	-0.031489459	-0.414437431	-0.533011413
school	-0.347164889	0.832474392	0.773124566	0.842550326	0.755428059	-0.263009485	-0.521280534	0.669087506	0.39625752	0.781850393	0.824499983	0.834426233	-0.258783259	0.683604589	0.641263571
distschool_mi	0.064537927	-0.786755696	-0.457712732	-0.710350896	-0.563744323	0.245329491	0.286368124	-0.812585099	-0.580231889	-0.562215304	-0.639533593	-0.854697746	0.136041756	-0.547891195	-0.713039375
college	-0.250867485	0.613427037	0.091488082	0.442477922	-0.037169705	-0.265246329	-0.390964427	0.412866151	0.205889248	0.522255767	0.5548766	0.507716408	-0.212634037	0.204533968	0.771181285
distcolleg_mi	0.263771348	-0.778477501	-0.456281669	-0.742923678	-0.52691218	0.119566641	0.098875649	-0.837594391	-0.418280903	-0.594936111	-0.688141567	-0.871158822	0.220414324	-0.569672307	-0.78098445
university	-0.141597116	0.841609422	0.505495008	0.782604351	0.583309423	-0.054727227	-0.17700995	0.929759185	0.515929816	0.620666254	0.719194952	0.957741116	-0.384222178	0.63324053	0.821010399
distuniver_mi	0.24162929	-0.771609607	-0.448893552	-0.751226652	-0.541572492	0.138932289	0.171676567	-0.879349574	-0.647687487	-0.592368191	-0.672235234	-0.891360078	0.282223253	-0.598357077	-0.761718963
dist_cbd_mi	0.196529571	-0.66045642	-0.313626224	-0.599178156	-0.360962915	0.136730322	0.202033171	-0.801033228	-0.660562087	-0.496520089	-0.556615615	-0.78053076	0.316767467	-0.450533359	-0.683861044
slope	0.040760891	0.029992248	-0.006832628	-0.187586205	0.034607386	0.246931013	-0.060071651	0.004582586	-0.090485453	-0.132717065	-0.158586519	-0.043399365	-0.399405018	0.032019775	-0.078275379
white	0.27147415	-0.253854353	-0.319055557	-0.448795932	-0.239446159	0.080796784	0.079961618	-0.27767819	0.038706007	-0.341758569	-0.23397908	-0.345325227	0.345156784	-0.222331222	-0.223015915
afam	-0.172463786	0.176992633	-0.02437969	0.279905709	0.096562661	0.102390078	0.062859574	0.2615936	-0.107023738	0.229798541	0.113402007	0.243067637	-0.430186218	0.273617368	0.229672523
male	-0.27450173	-0.046697987	0.31744131	0.171590966	0.043391926	-0.048572999	0.02732133	-0.070089792	-0.36541991	0.032970867	-0.070032729	0.038479824	-0.268573776	-0.20260819	-0.037422602
female	0.27450173	0.046697987	-0.31744131	-0.171590966	-0.043391926	0.048572999	-0.02732133	0.070089792	0.36541991	-0.032970867	0.070032729	-0.038479824	0.268573776	0.20260819	0.037422602
lst_col	0.338982644	-0.237340057	0.284829483	-0.130213799	0.105234535	-0.01861063	-0.186564751	-0.292468838	0.281786206	-0.228802483	-0.100807257	-0.222594456	0.382620777	-0.196581216	-0.419566258
stu_acc_100	-0.277901344	0.234689307	0.740778891	0.461478254	0.371472377	-0.202886727	-0.259254338	0.114695253	0.038802861	0.303404466	0.309512954	0.287429002	-0.069853442	0.033339983	0.066245599
hh_den_acres	-0.013094616	-0.062148857	0.102131667	0.042061472	0.038660838	0.222749817	-0.171232679	-0.127497004	-0.350802952	0.052652727	-0.003679819	-0.086576564	-0.285419695	0.052782712	-0.111488841
med_inc_10k	0.318686004	-0.428004693	-0.360952753	-0.4803153	-0.315964727	-0.113110802	-0.070107915	-0.414598862	0.157200886	-0.421788897	-0.446529091	-0.471400076	0.336869012	-0.377776301	-0.450252837
pop_den_acres	-0.157732988	0.067376121	0.414844028	0.287154645	0.211480817	0.025815157	-0.199293618	-0.017936826	-0.208759563	0.20348566	0.142376272	0.088102516	-0.247829092	0.076474961	-0.043111569
med_age	0.291124774	-0.25058064	-0.509851186	-0.507231012	-0.357039722	0.295436954	-0.004118947	-0.28255519	0.046176395	-0.27293662	-0.142596238	-0.387662945	0.281427362	-0.099636226	-0.171114162
emp_den_acres	-0.179488754	0.08442399	0.472178651	0.307950414	0.230329997	0.010967779	-0.222164623	-0.008969237	-0.206253664	0.21186814	0.163047541	0.110003133	-0.238878169	0.057818739	-0.033219728
bik_pct	0.068843211	0.033493175	0.244977269	0.090955398	0.095562066	-0.122789395	-0.141731091	-0.077351983	0.163163831	0.174076597	0.146534466	-0.004498981	0.099179527	0.10223932	-0.075059548
bik_den_acres	-0.170689256	0.198240108	0.490981869	0.337344228	0.290255225	0.005765722	-0.332069964	0.022524419	-0.084588079	0.336217835	0.268263038	0.145301985	-0.21300909	0.186254321	0.022951767
temp	0.103011314	0.086764308	0.046279229	0.020816115	0.026616504	-0.028628725	-0.063779885	0.064878261	0.074204813	0.076052831	0.068861194	0.066759894	0.069508724	0.067385155	0.062175847
hum	-0.05601286	-0.136141049	-0.089685393	-0.06708218	-0.0857549	-0.015824805	0.122338809	-0.062459236	0.011458625	-0.103727818	-0.127433824	-0.091077321	0.048900438	-0.097551519	-0.100498554
prec	0.110715886	-0.001794674	-0.008664784	0.02013474	-0.000803747	-0.005741052	0.037434575	0.01148524	0.011972061	-0.015756645	0.006666385	0.014732443	-0.012875508	-0.031928394	0.021950149
Bike_Average_MADT	0.571984644	0.012969003	0.126713748	-0.108650126	0.26321673	0.397737044	-0.158414181	0.027976967	0.088979112	-0.08070303	0.021168271	0.006868416	-0.01881093	0.168578662	-0.176694357

intd_1mi_acres	lanes	water_area_acres	wtr_cntdis_mi	park_area_acres	prk_cntdis_mi	frest_area_acres	frst_cntdi_mi	grass_area_acres	grs_cntdis_mi	comm_area_acres	comm_cntdi_mi	ind_area_acres	ind_cntdis_mi
1													
0.274082855	1												
-0.545635213	-0.077202417	1											
-0.05676876	0.003935008	-0.253564233	1										
-0.625058564	-0.120973049	0.793670977	-0.145945719	1									
-0.381584246	-0.274559056	0.338967213	-0.065026307	0.550582469	1								
-0.326806194	-0.178758622	0.591405052	-0.128619026	0.380332748	0.043612431	1							
0.08952134	-0.16070541	0.060705084	-0.309581526	0.288314452	0.282601063	-0.239707396	1						
0.600990041	-0.196218086	-0.32611176	0.1028417	-0.35090413	-0.068703402	-0.205186285	-0.077902102	1					
-0.522564683	0.012448896	0.308340083	0.148956091	0.445115489	0.138444541	-0.072767728	0.421707211	-0.415779074	1				
0.671539621	0.691502508	-0.336176924	0.176931928	-0.399946691	-0.296790289	-0.241515499	-0.152795638	0.138980274	-0.298313174	1			
-0.646294756	-0.112625522	0.675461769	0.009574119	0.825872413	0.519627466	0.222833988	0.373206671	-0.400835108	0.789986535	-0.386235619	1		
0.163428555	-0.293790341	-0.174793721	0.392856035	-0.225880619	0.1254183	-0.087115439	-0.219282739	0.649731037	-0.194863362	-0.044265639	-0.22523021	1	
-0.551779622	0.067966052	0.507209798	-0.092108393	0.734591233	0.531544325	-0.041953331	0.421318707	-0.426176189	0.780483692	-0.270103561	0.910973824	-0.315841067	1
-0.038332662	0.052268057	-0.162392028	-0.122487025	-0.14467816	-0.035841807	0.115455888	-0.278392668	-0.024802572	-0.102603493	-0.203297969	-0.134006003	-0.029762562	-0.119999375
-0.49700409	-0.139307257	0.336725292	0.171045254	0.430496585	0.298311876	0.030798331	0.347465842	-0.368704797	0.79128559	-0.248933857	0.798573691	-0.142782764	0.710465915
0.549568911	-0.078613229	-0.383095344	0.278277612	-0.431659355	-0.119640481	-0.273177781	-0.1873936	0.809883602	-0.356901493	0.167337666	-0.419898578	0.5618286	-0.443717757
-0.686230111	-0.119423446	0.814578031	-0.043019493	0.898476882	0.390539851	0.483636242	0.206937312	-0.457102549	0.648703461	-0.409516567	0.91952264	-0.248507234	0.751493654
0.673741657	0.161220284	-0.420157784	0.273883428	-0.506854175	-0.352584819	-0.290329944	-0.223021949	0.474194995	-0.367244446	0.533590476	-0.445876279	0.004569197	-0.376419212
-0.791959622	-0.13116336	0.774723933	-0.307599036	0.734571061	0.351074314	0.441091121	0.100911299	-0.621418724	0.41289442	-0.55453216	0.644933988	-0.335349746	0.590345049
0.439426087	0.779667544	-0.218725058	0.085276748	-0.270690685	-0.218830429	-0.153383607	-0.117470456	-0.143727387	-0.18647034	0.881298876	-0.249466653	-0.217093165	-0.129952849
-0.862618361	-0.243534473	0.437412012	-0.129882738	0.552676878	0.326740581	0.133638995	0.027728968	-0.642215367	0.389459888	-0.615966196	0.530229266	-0.257348409	0.501950716
0.912591581	0.212272166	-0.507510204	0.130276823	-0.593312754	-0.314602483	-0.278095926	-0.047880927	0.758514958	-0.484541812	0.626012583	-0.575306148	0.33160382	-0.509243715
-0.877250568	-0.250975867	0.713373075	-0.112592354	0.698504916	0.341137126	0.419704734	0.017433234	-0.650757476	0.433689182	-0.571094655	0.653737766	-0.260824688	0.503496446
-0.781085227	-0.306001755	0.79055899	-0.083481031	0.729454138	0.299660871	0.637889639	-0.075786215	-0.531472018	0.260752424	-0.536722888	0.533108834	-0.216862722	0.325337628
-0.109955311	0.017799842	0.207862934	0.041309092	0.123273862	0.044058588	-0.026446883	-0.068046568	0.186266271	0.313346048	-0.111910438	0.284164138	0.354390689	0.327052411
-0.328221847	0.227035407	0.087810564	-0.024152404	0.030139979	-0.269846555	0.149421273	-0.286021661	-0.102312684	0.271219283	-0.092323534	0.090256934	0.083692113	0.072080354
0.269952097	-0.24455126	-0.01183998	-0.302153288	0.108727231	0.396562058	-0.165181506	0.403474509	0.164627595	-0.231222298	-0.017980676	0.016934043	-0.060093592	0.062675388
0.056656227	-0.048679021	0.26454239	0.091269356	0.319454577	0.298294888	0.136584182	0.350720061	-0.141766001	0.069866353	0.073635636	0.290224701	-0.127041273	0.251601123
-0.056656227	0.048679021	-0.26454239	-0.091269356	-0.319454577	-0.298294888	-0.136584182	-0.350720061	0.141766001	-0.069866353	-0.073635636	-0.290224701	0.127041273	-0.251601123
-0.291933465	-0.105658145	-0.132288794	0.4143451	-0.244795742	-0.126495861	-0.049408403	-0.470571733	-0.102784872	-0.18328027	-0.256339344	-0.257480798	0.081044261	-0.29049585
0.198088758	0.011157536	-0.063578406	0.505385323	-0.179884146	-0.024695261	-0.028919582	-0.088427015	-0.021061088	-0.116694826	0.204118425	-0.13677089	-0.064452467	-0.141251612
-0.141913561	-0.249957444	0.124691575	-0.076106301	0.28812749	0.674936176	-0.118374318	0.273226149	-0.071857459	-0.103845913	-0.175476141	0.217723029	-0.169601661	0.238417829
-0.482533695	-0.123925325	-0.072537553	0.145848802	-0.053860319	-0.158003325	0.027225992	-0.318170709	-0.355466091	0.065171486	-0.330521167	-0.03857382	0.092964622	-0.092644932
0.012525682	-0.215398513	0.04913635	0.164521902	0.118662107	0.446361672	-0.107458049	0.186318465	-0.071635913	-0.145906372	-0.053453975	0.067471109	-0.173999315	0.08686946
-0.348681709	0.17065788	-0.028102761	-0.226647782	0.09200502	-0.046520148	0.030321897	-0.282946476	-0.144118801	0.10335027	-0.145535894	0.078061092	-0.096586869	0.105510184
0.032609205	-0.186759592	0.044806546	0.212565918	0.095813556	0.428132379	-0.097135779	0.168383528	-0.066473648	-0.14061984	-0.01449716	0.06103198	-0.162651019	0.074105768
-0.099083827	-0.061189727	-0.022039474	-0.006468847	-0.1199641	-0.004979704	-0.190926979	-0.125069822	-0.145688137	-0.152610208	-0.03187648	-0.164778784	-0.252129904	-0.050999918
0.01615526	-0.120527619	0.004989073	0.215581687	0.028312429	0.314879925	-0.138292319	-0.020309859	-0.086154313	-0.169886653	0.022033129	-0.004293279	-0.202921877	0.056560506
0.042894514	0.016043929	-0.015876251	0.001330796	-0.09826392	-0.046199086	0.080553795	-0.143079575	0.033835008	-0.124453286	0.05972673	-0.097982414	0.018583091	-0.119221876
-0.043506418	-0.078251476	-0.001905809	-0.096036154	0.025699561	0.045406223	-0.02385722	0.097934621	-0.04720879	-0.011519633	-0.108445644	-0.006706272	-0.020482312	0.001480605
0.030626418	0.010432634	0.025428875	-0.008305808	0.037858815	-0.051382899	-0.023381421	0.039542647	0.031096544	-0.028408791	0.033439229	-0.010763337	0.001994916	-0.029312822
-0.095888607	-0.281809227	-0.058838733	0.0699861	-0.01282222	0.146756833	-0.174232436	-0.253109117	0.414968781	-0.228085593	-0.258931855	-0.144680755	0.336742911	-0.158473344

resi_area_acres	*resi_cntdi_mi*	*ret_area_acres*	*ret_cntdis_mi*	*school*	*distschool_mi*	*college*	*distcolleg_mi*	*university*	*distuniver_mi*	*dist_cbd_mi*	*slope*	*white*	*afam*	*male*	*female*	*lst_col*
1																
-0.240690191	1															
0.023793168	-0.357474959	1														
-0.05666835	0.652733306	-0.487880137	1													
-0.116063486	-0.381024972	0.482423678	-0.498932778	1												
0.085767625	0.333719913	-0.66315383	0.735486335	-0.691200958	1											
-0.213149337	-0.112220416	-0.050850837	-0.25584178	0.310146467	-0.379279162	1										
-0.021858915	0.484596623	-0.6257406	0.543543054	-0.721371567	0.731855882	-0.420687079	1									
-0.14397791	-0.45170428	0.727666363	-0.650181595	0.768045397	-0.84333099	0.399064449	-0.913859766	1								
-0.076803373	0.546635525	-0.653094646	0.758166002	-0.716247516	0.816065752	-0.369931362	0.902043014	-0.91438753	1							
-0.054399696	0.2791909	-0.546158131	0.755553937	-0.554751197	0.784305903	-0.362815243	0.655842447	-0.782004721	0.8792707	1						
-0.032953291	0.10782725	0.21766694	0.185799961	-0.044996828	0.213397487	-0.157653004	-0.076681111	0.058179857	-0.089690513	-0.108638709	1					
0.126012362	0.172236886	-0.349658075	0.177027246	-0.313050835	0.246628135	0.023866791	0.219919003	-0.297404805	0.231132463	0.167538012	0.194662565	1				
-0.119490729	-0.12702647	0.343267943	-0.086965688	0.087746001	-0.152042714	-0.064949815	-0.126344788	0.242349652	-0.177934684	-0.171285589	-0.066576583	-0.866787661	1			
-0.102376208	0.114654786	-0.131104964	0.260385512	0.10280697	0.08794217	0.019724157	-0.032616602	0.019734363	0.069796994	0.111427643	-0.110324223	-0.572939617	0.406089724	1		
0.102376208	-0.114654786	0.131104964	-0.260385512	-0.10280697	-0.08794217	-0.019724157	0.032616602	-0.019734363	-0.069796994	-0.111427643	0.110324223	0.572939617	-0.406089724	-1	1	
0.18155536	-0.112747039	-0.033710767	-0.232248082	0.002164767	0.061884139	-0.264354364	0.333786206	-0.265061458	0.225237186	0.152763752	-0.227059805	0.12494327	-0.413045536	-0.134884659	0.134884659	1
-0.02427062	-0.123162371	0.152357833	-0.168870087	0.526666046	-0.191124519	0.054602329	-0.261067155	0.224431841	-0.185378556	-0.058199214	-0.189592211	-0.513417285	0.055645026	0.494551258	-0.494551258	0.414134306
-0.142375602	0.039729764	0.077445317	0.072385897	0.007028689	0.079855653	-0.170188615	0.22993205	-0.099203238	0.160881085	0.106166723	-0.25510654	-0.683640099	0.69351449	0.392938156	-0.392938156	0.028684861
0.062264681	0.106112252	-0.366878957	0.013834563	-0.478880827	0.262022701	-0.173386455	0.589230995	-0.532396919	0.443918665	0.264124399	-0.00088026	0.462084083	-0.431761926	-0.36762011	0.36762011	0.510503484
-0.142518848	-0.040961751	0.150039615	-0.043209917	0.257670823	-0.041274905	-0.121758253	0.046248909	0.042683276	0.024744341	0.054781903	-0.300498956	-0.803446098	0.591422588	0.528387706	-0.528387706	0.202202251
0.360882962	0.072622184	-0.271525014	0.13306698	-0.423059077	0.244772106	0.026682375	0.320900978	-0.37047481	0.232688469	0.116391798	0.088681921	0.710639818	-0.445886792	-0.544359668	0.544359668	0.020141521
-0.126392723	-0.044136181	0.141427901	-0.049472095	0.29550433	-0.056094806	-0.095477431	0.007448644	0.06368456	0.002240122	0.041880304	-0.29386535	-0.784816721	0.537254345	0.557671068	-0.557671068	0.228248038
-0.176224935	-0.098525001	0.030740081	-0.207123224	0.179563732	0.096543405	-0.093771551	0.159127883	-0.068186588	0.05440006	0.02378867	-0.165781228	-0.242084289	0.087179633	0.008538611	-0.008538611	0.318980678
-0.101358204	-0.086884392	0.187719174	-0.098082059	0.378668748	-0.054182477	-0.055083106	0.002993644	0.082537042	-0.040218922	0.007083528	-0.2207642	-0.691866342	0.455608146	0.396743017	-0.396743017	0.238807261
-0.004914683	-0.061237696	0.040816921	-0.076581922	0.074253064	-0.061554996	0.046157688	-0.04414654	0.061599005	-0.029154771	-0.019913603	-0.027296432	0.007853816	-0.030661393	-0.027665373	0.027665373	0.064164781
0.02362099	-0.001831095	-0.063872013	-0.007720903	-0.116166665	0.066008571	-0.098379136	0.107618908	-0.091981666	0.07096428	0.036539811	-0.067650686	-0.024192262	0.047632735	-0.059326717	0.059326717	0.024534555
-0.064214456	-0.013359094	-0.010935902	0.017807181	-0.001412773	0.015245277	0.029309649	-0.019496571	0.015689124	-0.004779512	0.013204943	-0.013142106	0.048365648	-0.027723974	0.011688069	-0.011688069	-0.040988531
-0.05149128	-0.223360864	0.408405987	-0.172123917	0.076231315	-0.129769536	-0.39988876	0.121043899	0.06095074	0.007400769	-0.002845011	-0.042402821	-0.06617569	0.141679036	-0.213641855	0.213641855	0.2635636

stu_acc_100	hh_den_acres	med_inc_10k	pop_den_acres	med_age	emp_den_acres	bik_pct	bik_den_acres	temp	hum	prec	Bike_Average_MADT
1											
0.282569854	1										
-0.335275572	-0.269558805	1									
0.696267897	0.870377066	-0.36769962	1								
-0.697240615	-0.360267565	0.501778864	-0.654253932	1							
0.754851633	0.836821578	-0.380668545	0.994252	-0.667698829	1						
0.37618693	0.333349679	-0.189403318	0.454147492	-0.257463724	0.441556279	1					
0.717792096	0.751582962	-0.376175298	0.92452766	-0.586820331	0.921050158	0.651900182	1				
0.036497494	-0.006628419	0.006476442	0.006088609	-0.009794843	0.009887068	0.042854068	0.020788559	1			
-0.034097649	0.063571347	0.040282103	0.041921969	-0.011568486	0.029844858	0.017424041	-0.002582678	0.035654452	1		
-0.062502113	-0.07605106	-0.025244676	-0.081935237	0.0433331	-0.081737311	0.020244496	-0.065843738	0.308144484	0.208540266	1	
-0.067633837	0.338669524	0.038455046	0.212190287	0.013343606	0.175665788	0.277973877	0.245185785	0.190058717	-0.028314453	0.003547832	1

	Strava_MADT	primary_mi	secondary_mi	tertiary_mi	resi_road_mi	path_mi	cycleway_mi	cwy_lneall_mi	cwy_binary	cwy_trkall_mi	cwy_trkbin	footway_mi	meanspeed	bikeprking	bus_stops
Strava_MADT	1														
primary_mi	-0.034443084	1													
secondary_mi	0.034265108	0.565807743	1												
tertiary_mi	-0.238682917	0.76063059	0.671570755	1											
resi_road_mi	0.089294864	0.62394538	0.795214916	0.740259506	1										
path_mi	0.535724858	-0.166069667	-0.104550476	-0.375224594	-0.142890147	1									
cycleway_mi	0.006019815	-0.484056578	-0.473439065	-0.147008986	-0.34604112	0.011415378	1								
cwy_lneall_mi	0.1551826	0.784617982	0.350508322	0.67841069	0.375883468	-0.073961951	-0.045539645	1							
cwy_binary	0.294604155	0.431571746	0.258768222	0.456502618	0.297461082	-0.276648579	-0.011509999	0.618762707	1						
cwy_trkall_mi	-0.292926511	0.840114228	0.388692892	0.665960265	0.355709389	-0.286671268	-0.458481595	0.669752946	0.312991561	1					
cwy_trkbin	-0.020666571	0.82585713	0.547415303	0.695846649	0.488264686	-0.005687992	-0.432149798	0.680062457	0.370862758	0.843955223	1				
footway_mi	0.041049897	0.887529661	0.582533289	0.835381278	0.570320339	-0.179238049	-0.218465643	0.945374624	0.581038888	0.744061302	0.784972259	1			
meanspeed	0.087717205	-0.268806685	-0.211320201	-0.22418867	-0.150170196	-0.337435498	0.283952689	-0.260078683	0.199741876	-0.186720592	-0.173095891	-0.295922735	1		
bikeprking	0.00498442	0.721073383	0.409890128	0.669815954	0.687763938	-0.062172648	-0.23653678	0.538168891	0.340891579	0.646857998	0.740009962	0.618740061	-0.109261424	1	
bus_stops	-0.128960707	0.879062149	0.317896987	0.724865761	0.329607761	-0.150370699	-0.189795688	0.869127658	0.443130764	0.796743857	0.804813816	0.90052837	-0.293442013	0.609999589	1
intd_1mi_acres	0.035480806	0.710826567	0.399994069	0.759055903	0.386904306	-0.170546147	0.06694074	0.951316783	0.60434298	0.605484425	0.63812477	0.934186826	-0.218660826	0.449558782	0.847281638
lanes	-0.080775843	0.285434724	0.085956569	0.097995283	-0.168855088	0.219108783	-0.153868225	0.207005925	0.185631954	0.309028474	0.420439259	0.238726576	-0.133196618	0.040926571	0.410704059
water_area_acres	-0.194409506	-0.446683993	-0.296718021	-0.461125865	-0.354908179	0.31837394	0.195204748	-0.560692734	-0.753440766	-0.329118056	-0.384086102	-0.550569284	-0.07826961	-0.337236446	-0.430012509
wtr_cntdis_mi	0.175083181	0.191912134	0.532141934	0.08504897	0.3490835	-0.008672025	-0.47142245	-0.034270893	0.007004715	-0.014275759	0.113246234	0.090102106	-0.015394048	-0.093369512	-0.086815777
park_area_acres	-0.159931071	-0.538149973	-0.316448756	-0.509507432	-0.37694017	0.455320881	0.103914097	-0.62540665	-0.831452101	-0.39903301	-0.435618091	-0.623800239	-0.31953479	-0.396945743	-0.498223926
prk_cntdis_mi	-0.017629895	-0.331629485	-0.184394264	-0.380881675	-0.19647607	0.42111803	-0.123399444	-0.415415823	-0.645208955	-0.328721891	-0.391314461	-0.429543787	-0.386698819	-0.302331227	-0.366279468
frest_area_acres	-0.113774166	-0.320327685	-0.211638521	-0.348425115	-0.243807276	0.10355376	0.328435346	-0.322096415	-0.363979055	-0.233576711	-0.273283833	-0.363934276	0.38853375	-0.106436049	-0.283318406
frst_cntdi_mi	-0.374659418	-0.271150397	-0.265031972	0.088768151	-0.202753102	-0.148877577	0.48689057	-0.014055321	-0.344853389	-0.146744145	-0.244692602	-0.059725786	-0.3236302	-0.198008417	-0.012741663
grass_area_acres	0.569005678	0.501744744	0.398990648	0.406295626	0.568221582	0.339405499	0.012148825	0.696495537	0.367371538	0.177315052	0.410322279	0.65208094	-0.290805977	0.464864192	0.444632689
grs_cntdis_mi	-0.318187117	-0.365565429	-0.2669523	-0.366670195	-0.259356859	-0.013746837	0.068285689	-0.490496007	-0.601388345	-0.272044407	-0.323130756	-0.480467604	-0.064257145	-0.300414322	-0.38815324
comm_area_acres	-0.121531455	0.706933274	0.444273329	0.528340644	0.171141853	-0.124279695	-0.356241692	0.645055403	0.341637042	0.611165768	0.695361311	0.738672347	-0.24707051	0.251455599	0.823526939
comm_cntdi_mi	-0.311287036	-0.471771539	-0.29857057	-0.507622824	-0.332679263	0.301881783	0.002099601	-0.647340359	-0.894345248	-0.347199098	-0.417189243	-0.629751752	-0.275306408	-0.359628214	-0.496062046
ind_area_acres	0.584303595	0.163877737	0.186730828	-0.022690467	0.302525503	0.232195759	-0.019629934	0.280742362	0.204510952	-0.286413857	-0.195726774	0.217815713	-0.142470662	-0.098286016	0.001700629
ind_cntdis_mi	-0.348053542	-0.38948819	-0.258313456	-0.417466916	-0.319186361	0.293350273	-0.08322304	-0.554216765	-0.742338518	-0.244614924	-0.3288209	-0.532183291	-0.456838721	-0.297069723	-0.383927091
resi_area_acres	0.155843439	-0.150953954	-0.066044086	-0.124788811	-0.062379481	0.236871129	0.156573762	-0.062349849	0.253970103	-0.126158377	-0.040889924	-0.142953946	0.443315622	-0.052257812	-0.165238796
resi_cntdi_mi	-0.293953751	-0.343853536	-0.27116406	-0.397171021	-0.279123373	-0.003968659	0.003107931	-0.505937845	-0.632635065	-0.292353575	-0.356851598	-0.491685957	-0.08468314	-0.333172027	-0.372088531
ret_area_acres	0.552446285	0.598947916	0.403216268	0.427031771	0.502006041	0.318769517	-0.079628187	0.691254929	0.39087255	0.305628864	0.475818191	0.637739316	-0.210423533	0.488319577	0.470460035
ret_cntdis_mi	-0.285191176	-0.530635423	-0.346669283	-0.559944434	-0.383659923	0.28149299	0.111561529	-0.701101985	-0.879208862	-0.384570753	-0.461218458	-0.690621852	-0.072198208	-0.410663296	-0.546237875
school	-0.072972584	0.850677958	0.810122128	0.837331347	0.793985194	-0.241034288	-0.476264022	0.665643209	0.401938899	0.795683436	0.830075054	0.833484806	-0.207338629	0.728941944	0.668446956
distschool_mi	-0.225902319	-0.76980683	-0.516032537	-0.701150667	-0.59331604	0.2085408	0.218109173	-0.808334605	-0.589447457	-0.530414293	-0.619723048	-0.848180728	0.076278826	-0.539353671	-0.714484615
college	-0.343108473	0.580918564	0.199677082	0.421713082	0.042667059	-0.20306309	-0.301858803	0.35434016	0.195805617	0.493182864	0.527973254	0.476508739	-0.163168482	0.213414405	0.719459865
distcolleg_mi	-0.014744453	-0.749246356	-0.498573451	-0.724280068	-0.545715123	0.10649172	-0.016048618	-0.823939524	-0.42587159	-0.55893079	-0.659582383	-0.853658501	0.117483229	-0.563362264	-0.78127722
university	0.158901086	0.828106965	0.563429959	0.774991908	0.611516738	-0.020420288	-0.086092489	0.931939112	0.52763899	0.593192632	0.702872162	0.957001436	-0.319526429	0.636326656	0.839837032
distuniver_mi	-0.119706458	-0.751751451	-0.498015081	-0.740256506	-0.562381051	0.115301286	0.087625045	-0.878696885	-0.672277053	-0.562822045	-0.650636176	-0.885858563	0.215730234	-0.595219203	-0.772413542
dist_cbd_mi	-0.142433188	-0.62850088	-0.357236763	-0.579883094	-0.385445629	0.100042552	0.134177123	-0.785422781	-0.679562922	-0.457543899	-0.52665752	-0.763789362	0.27086626	-0.444730813	-0.673229511
slope	0.04757741	0.053998346	-0.034791934	-0.17404785	0.025037932	0.210418439	-0.077491339	0.006583394	-0.061899441	-0.093377495	-0.124408099	-0.033993068	-0.384556023	0.064043425	-0.050485434
white	0.228397661	-0.207083438	-0.313530847	-0.428578725	-0.10827094	0.065277365	0.082803276	-0.258381983	0.097547969	-0.308997892	-0.181074693	-0.307824675	0.378387562	-0.078415478	-0.245565066
afam	-0.123988154	0.130487725	-0.011820476	0.265216723	-0.025044625	0.137445673	0.044626933	0.249334031	-0.162136656	0.196946895	0.065126282	0.21197041	-0.493715289	0.129201137	0.250077673
male	-0.394199114	-0.063557738	0.262602131	0.16143	-0.023994119	-0.037246117	0.032647658	-0.082264587	-0.400056812	0.029011331	-0.095393053	0.022136135	-0.298177157	-0.249627452	-0.008377743
female	0.394199114	0.063557738	-0.262602131	-0.16143	0.023994119	0.037246117	-0.032647658	0.082264587	0.400056812	-0.029011331	0.095393053	-0.022136135	0.298177157	0.249627452	0.008377743
lst_col	0.343039459	-0.236125807	0.245962945	-0.139963403	0.113258654	-0.025059497	-0.207554079	-0.296645327	0.267300339	-0.239539664	-0.089843193	-0.230433516	0.377868756	-0.166650036	-0.447774688
stu_acc_100	-0.249337491	0.249037001	0.689699192	0.456164052	0.352163679	-0.216312945	-0.235228889	0.100305411	0.003606817	0.311919074	0.308980791	0.273947906	-0.041040456	0.059022582	0.098852399
hh_den_acres	-0.040834236	-0.053604271	0.07399854	0.036431162	0.003873378	0.242107169	-0.21748384	-0.133828213	-0.377956621	0.060676572	-0.002972899	-0.094387183	-0.336635541	0.044582168	-0.088550711
med_inc_10k	0.186599606	-0.40327027	-0.364064637	-0.454773051	-0.282548113	-0.145205164	-0.11532083	-0.3919461	0.190415193	-0.402811365	-0.423787745	-0.442178496	0.321557641	-0.345660425	-0.467031998
pop_den_acres	-0.162610045	0.073419954	0.357832883	0.271318351	0.157577678	0.045701526	-0.224866012	-0.030637876	-0.258623241	0.206061419	0.131488801	0.06846536	-0.286810902	0.068330067	-0.009959546
med_age	0.267473884	-0.190698923	-0.492083144	-0.469990766	-0.294451424	0.280298833	-0.062393151	-0.248976247	0.10981264	-0.216482922	-0.069234162	-0.339011886	0.265397959	-0.014454454	-0.143424867
emp_den_acres	-0.181803622	0.093171226	0.409245933	0.293686928	0.183371459	0.027850063	-0.241376012	-0.022799218	-0.255097046	0.21858403	0.155321082	0.091069541	-0.26998018	0.063093146	0.001358675
bik_pct	-0.001977562	0.070752206	0.22029066	0.101285936	0.058973035	-0.137798199	-0.18821078	-0.063596897	0.154887194	0.211971156	0.182671763	0.007415269	0.073755391	0.10613458	-0.014838557
bik_den_acres	-0.166465988	0.223267422	0.44518323	0.329918977	0.247885356	0.014593602	-0.365537102	0.012943982	-0.124002001	0.365458301	0.279018854	0.133462823	-0.240314558	0.197365586	0.069661153
temp	0.015124345	0.062352288	0.039854017	0.017805224	0.035589707	-0.029697973	-0.049364536	0.027651355	0.034023281	0.065683588	0.066503241	0.03823403	0.098193103	0.072338812	0.041455858
hum	0.038094652	-0.125460663	-0.052682387	-0.036621001	-0.068424338	-0.01713448	0.089903885	-0.055525561	0.006838977	-0.095633859	-0.109959327	-0.069073353	0.006170595	-0.099624141	-0.089594809
prec	0.036969828	-0.009275482	-0.002590749	0.016681057	0.005955443	-0.008214182	0.041118361	0.003077795	0.014640905	-0.023441982	-6.76634E-05	0.013576592	-0.018113113	-0.040701097	0.012849415
Pedestrian_Average_MADT	0.642839082	0.089829472	0.05931331	-0.062774255	0.102748738	0.46427302	-0.060784589	0.152325044	0.171420121	-0.016105739	0.172794619	0.099468798	-0.065626694	0.189697386	0.037741334

intd_1mi_acres	lanes	water_area_acres	wtr_cntdis_mi	park_area_acres	prk_cntdis_mi	frest_area_acres	frst_cntdi_mi	grass_area_acres	grs_cntdis_mi	comm_area_acres	comm_cntdi_mi	ind_area_acres	ind_cntdis_mi
1													
0.195567998	1												
-0.543950564	-0.043906794	1											
-0.095240018	0.001458841	-0.249986127	1										
-0.610518421	-0.042924663	0.771785796	-0.145168204	1									
-0.414859593	-0.333541978	0.366139493	-0.09367253	0.552729001	1								
-0.328413311	-0.195602956	0.596760429	-0.130579997	0.347036825	0.047524673	1							
0.115258915	-0.171218832	0.051210404	-0.328013692	0.28007223	0.269902605	-0.257087747	1						
0.643994703	-0.11674285	-0.344205844	0.095102488	-0.357206754	-0.113749219	-0.226865439	-0.073255632	1					
-0.500035877	0.100825053	0.301876102	0.163205937	0.411594965	0.11714446	-0.077719965	0.400913237	-0.41516273	1				
0.668524916	0.572556936	-0.339354353	0.156293554	-0.392977972	-0.334590496	-0.248457273	-0.150865797	0.23151222	-0.291813473	1			
-0.653887569	-0.035097239	0.674854413	0.001177876	0.82389992	0.511945959	0.201018668	0.356572792	-0.425266194	0.777388363	-0.394929425	1		
0.203636364	-0.271032897	-0.188086742	0.408281227	-0.259775672	0.08442703	-0.093729405	-0.223817126	0.640002947	-0.203926928	0.009422193	-0.261424564	1	
-0.559058427	0.134091021	0.501979417	-0.097402253	0.738871811	0.538009742	-0.060229194	0.409226398	-0.426638793	0.759521071	-0.288377005	0.913769749	-0.332233699	1
-0.002373884	0.201193376	-0.185125133	-0.11638866	-0.17491504	-0.064245152	0.093998504	-0.288647878	-0.030666072	-0.11011199	-0.159563682	-0.165363474	-0.039679796	-0.137746043
-0.507850881	-0.146280193	0.338816108	0.168794861	0.413854377	0.285118847	0.021503877	0.324881996	-0.383491567	0.781514339	-0.274253295	0.790275452	-0.161864295	0.699204243
0.571669103	-0.004177178	-0.388791133	0.294994949	-0.426526342	-0.145716828	-0.282547689	-0.186723713	0.814948933	-0.342174076	0.247141167	-0.431166813	0.59287098	-0.4460207
-0.688130713	-0.05410353	0.813626792	-0.043167664	0.891813246	0.382829002	0.469209657	0.186952437	-0.476936854	0.632286417	-0.4188356	0.916702046	-0.280068401	0.746603167
0.660108158	0.13064332	-0.414298978	0.228300953	-0.490165577	-0.367047116	-0.291951867	-0.217377623	0.500985124	-0.345220984	0.546217054	-0.444903327	0.022233083	-0.370839895
-0.784510871	-0.049413502	0.775113058	-0.298232231	0.711977062	0.376729833	0.438171127	0.086379287	-0.652838146	0.392010925	-0.557575412	0.642219201	-0.371530176	0.589484512
0.390689417	0.629448691	-0.201041763	0.07411826	-0.244334641	-0.243294442	-0.143403759	-0.121727449	-0.099619636	-0.168170074	0.84423481	-0.233211209	-0.185558005	-0.130769359
-0.847863248	-0.189848074	0.409665789	-0.086660823	0.553080498	0.362817186	0.108196671	0.024490347	-0.665496772	0.359979694	-0.613213921	0.550485709	-0.298811538	0.522401673
0.909688234	0.167108878	-0.502854843	0.097906018	-0.574181957	-0.347651867	-0.277659487	-0.029003082	0.795428594	-0.458598165	0.639034143	-0.577962049	0.368153981	-0.508144545
-0.871115525	-0.235130961	0.717613509	-0.083417825	0.685955404	0.381748381	0.428899708	0.000621254	-0.67664524	0.407846188	-0.578362615	0.664134126	-0.289967057	0.504948528
-0.759993547	-0.294921028	0.79762139	-0.067421282	0.694261409	0.340984786	0.656165019	-0.10050905	-0.554201779	0.226558813	-0.522011044	0.517747848	-0.230890484	0.307373463
-0.114206132	0.156053038	0.200822494	0.033852242	0.107965435	-0.02006605	-0.049456533	-0.070504028	0.138336851	0.345478855	-0.093687821	0.276994894	0.293045904	0.33737803
-0.3151645	0.164684196	0.041984398	0.022983518	-0.035829485	-0.360504249	0.145723286	-0.340077701	-0.008727641	0.266977082	-0.197496689	0.055745056	0.111074729	0.019538537
0.262624284	-0.180819191	0.021630791	-0.33648213	0.191261057	0.477640796	-0.179767581	0.463483371	0.083851614	-0.224381166	0.076097492	0.068234063	-0.093954192	0.131715833
0.05102867	0.052636574	0.291726767	0.030814694	0.367944761	0.328836445	0.137105026	0.393852795	-0.19433349	0.08168363	0.129245165	0.324514336	-0.172443627	0.295083929
-0.05102867	-0.052636574	-0.291726767	-0.030814694	-0.367944761	-0.328836445	-0.137105026	-0.393852795	0.19433349	-0.08168363	-0.129245165	-0.324514336	0.172443627	-0.295083929
-0.297160592	-0.156772962	-0.137392198	0.427025438	-0.264153333	-0.091284632	-0.033900408	-0.473598716	-0.088890435	-0.196351851	-0.299669227	-0.260389719	0.095695819	-0.297028962
0.183481276	0.030298132	-0.023047062	0.454070391	-0.189797067	0.025459652	0.008277481	-0.086333202	-0.045062968	-0.111774814	0.23256019	-0.144883776	-0.057150235	-0.142905788
-0.14951293	-0.242402745	0.133173572	-0.09703535	0.317438713	0.710962852	-0.122615824	0.283054302	-0.095986543	-0.115493685	-0.149995902	0.239523611	-0.186264843	0.271764646
-0.462141245	-0.154726825	-0.1139572	0.200317933	-0.092531979	-0.188706941	0.007113991	-0.337009793	-0.341641794	0.044981712	-0.361017858	-0.064859102	0.095501623	-0.119475144
-0.003439581	-0.205036843	0.079755525	0.114687972	0.150576132	0.524453336	-0.09404287	0.204849385	-0.105532998	-0.153903949	-0.018789436	0.091373759	-0.181776655	0.122955964
-0.322491787	0.258432018	-0.08855165	-0.177587161	0.051505919	-0.110832924	-0.021530673	-0.320954576	-0.110571124	0.071830857	-0.150709564	0.039313947	-0.119237436	0.075149556
0.015117677	-0.178966556	0.077152984	0.15565651	0.123352508	0.503907196	-0.082814176	0.186815568	-0.098513006	-0.148662615	0.014198131	0.080682733	-0.172269279	0.106557394
-0.084935786	0.036471616	-0.028027137	-0.015575199	-0.127168516	0.031837061	-0.189842009	-0.121775438	-0.174408378	-0.166277826	0.052800847	-0.163928072	-0.280650363	-0.039411989
0.001155806	-0.071918083	0.029173585	0.173491513	0.050781951	0.378548897	-0.129523212	-0.008536109	-0.12465416	-0.17520383	0.071309606	0.01271228	-0.224135554	0.088203205
0.010933118	-0.013706186	0.00476336	0.003408334	-0.080374099	-0.033486636	0.101432608	-0.127280324	0.00809373	-0.091805331	0.043907678	-0.075549149	-0.004691008	-0.100920541
-0.027923801	-0.051042327	-0.025526375	-0.061966184	0.046201	0.043454257	-0.053971006	0.118583714	-0.050158657	-0.009383877	-0.082772679	0.007065496	-0.042275264	0.023369461
0.020887961	0.025918769	0.017156374	0.001515262	0.058747883	-0.071051706	-0.054767269	0.055567685	0.031974248	-0.012915831	0.030568258	-0.005557875	-0.004486819	-0.015894015
0.039597479	0.096393914	-0.094881241	-0.030620471	-0.033071382	0.057678234	-0.19175787	-0.214899879	0.381147858	-0.236790812	-0.007885777	-0.199281361	0.206733006	-0.124322094

resi_area_acres	*resi_cntdi_mi*	*ret_area_acres*	*ret_cntdis_mi*	*school*	*distschool_mi*	*college*	*distcolleg_mi*	*university*	*distuniver_mi*	*dist_cbd_mi*	*slope*	*white*	*afam*	*male*	*female*	*lst_col*
1																
-0.256057439	1															
0.036399083	-0.360674957	1														
-0.085970337	0.64199489	-0.496481324	1													
-0.099249125	-0.384723633	0.489651108	-0.492508683	1												
0.044156457	0.333519771	-0.669998673	0.729773628	-0.68428748	1											
-0.172703016	-0.113238031	-0.02556584	-0.241080702	0.300228876	-0.349553687	1										
-0.055569952	0.508313332	-0.63112852	0.548367319	-0.704191162	0.703401606	-0.385954847	1									
-0.115427347	-0.459403676	0.742405902	-0.64829314	0.76176876	-0.833962904	0.371097703	-0.898399801	1								
-0.110247847	0.560410225	-0.659228049	0.765735769	-0.705908758	0.804370097	-0.3415488	0.882029671	-0.906955853	1							
-0.087144007	0.269857711	-0.554372008	0.744092258	-0.536636173	0.769820657	-0.317527251	0.603532114	-0.763464483	0.869882813	1						
-0.051022236	0.104989514	0.198362403	0.178009563	-0.03489456	0.229071993	-0.134280003	-0.07257795	0.054979802	-0.096485648	-0.121904964	1					
0.169396134	0.150055376	-0.217131257	0.147838358	-0.260576458	0.194912988	-0.089537827	0.154840471	-0.24360132	0.161788439	0.121472528	0.264589122	1				
-0.158989625	-0.099087979	0.213558209	-0.048463904	0.055186444	-0.104923332	0.02707983	-0.054840463	0.198065509	-0.120070098	-0.147989013	-0.123181961	-0.859205804	1			
-0.128574227	0.130587171	-0.206156901	0.295150232	0.055990221	0.111331025	0.119189836	0.000224938	-0.003708194	0.10478615	0.126733409	-0.155793624	-0.599689918	0.4585896	1		
0.128574227	-0.130587171	0.206156901	-0.295150232	-0.055990221	-0.111331025	-0.119189836	-0.000224938	0.003708194	-0.10478615	-0.126733409	0.155793624	0.599689918	-0.4585896	-1	1	
0.188743784	-0.09776875	-0.005177375	-0.239031414	-0.015139093	0.055553085	-0.311985613	0.316219568	-0.264263691	0.215942989	0.149127583	-0.216408449	0.1233142	-0.415898049	-0.194580706	0.194580706	1
-0.035943797	-0.121287327	0.152192407	-0.169956168	0.480975959	-0.172718016	0.117664502	-0.278288348	0.210937879	-0.161961736	-0.009317222	-0.206867453	-0.548181789	0.082750266	0.45356694	-0.45356694	0.391952513
-0.16360928	0.051793298	0.053639064	0.087260426	-0.013589163	0.083513065	-0.125894784	0.251415387	-0.108172808	0.176057982	0.108122797	-0.280396984	-0.74544703	0.769674442	0.41213268	-0.41213268	0.005323284
0.059852887	0.093944023	-0.326734961	-0.015468135	-0.448693343	0.214966789	-0.196727774	0.569845879	-0.506686564	0.405380885	0.218366056	0.012062288	0.440621449	-0.420131634	-0.380343838	0.380343838	0.526305137
-0.169151177	-0.024132672	0.122173566	-0.024593294	0.209230388	-0.022781054	-0.05545569	0.064665337	0.023117787	0.055073526	0.081237113	-0.325470202	-0.864969469	0.669309438	0.522782431	-0.522782431	0.163613164
0.367183652	0.041718981	-0.199334275	0.089998869	-0.363883001	0.182313328	0.047038229	0.298953067	-0.3264233	0.174687106	0.049473576	0.093592796	0.706945342	-0.430545682	-0.545432457	0.545432457	0.049052584
-0.152767826	-0.030949748	0.122454653	-0.034296503	0.244983483	-0.038546825	-0.031693044	0.023607587	0.044588832	0.032600155	0.071920687	-0.322492658	-0.856148481	0.625384956	0.549307993	-0.549307993	0.188825259
-0.198216588	-0.080551702	-0.002355672	-0.212085663	0.177939379	0.10147501	0.015886432	0.140285879	-0.05982897	0.039318206	0.004549902	-0.166951979	-0.210202608	0.060047563	-0.037366125	0.037366125	0.30145403
-0.127669013	-0.071220345	0.162778086	-0.085612092	0.344381933	-0.037039622	0.031325626	0.017020344	0.066208413	-0.017019435	0.02815286	-0.245487887	-0.735008746	0.511414433	0.37853436	-0.37853436	0.20818865
-0.008624568	-0.037339883	0.002769478	-0.053184288	0.058849599	-0.042156354	0.042006292	-0.027193733	0.033291451	0.003321682	0.022367501	-0.082517341	0.017242657	-0.049767163	-0.020793026	0.020793026	0.063844149
0.004570681	0.006317859	-0.06609808	-0.002009492	-0.091870835	0.035261574	-0.080151503	0.109748541	-0.075344733	0.055802483	0.008231411	-0.091949893	-0.061358368	0.077024049	6.17943E-05	-6.17943E-05	0.015935597
-0.103953979	-0.008755129	-0.01053324	0.023527547	-0.004046902	0.003370532	0.024723695	0.009324461	0.009627093	0.00695818	0.011824469	-0.037485715	0.056518098	-0.025744399	0.009998485	-0.009998485	-0.031412367
-0.028785856	-0.234352059	0.404452573	-0.222371029	0.05534222	-0.169310032	-0.145582701	0.024308665	0.160433397	-0.094495059	-0.125231559	-0.087505518	0.067759069	0.05393693	-0.29414264	0.29414264	0.146710408

stu_acc_100	hh_den_acres	med_inc_10k	pop_den_acres	med_age	emp_den_acres	bik_pct	bik_den_acres	temp	hum	prec	Pedestrian_Average_MADT
1											
0.263559627	1										
-0.337573496	-0.27819284	1									
0.669388034	0.879924436	-0.372340572	1								
-0.715011081	-0.355385249	0.484389208	-0.647529439	1							
0.726828741	0.848622689	-0.388581187	0.99518885	-0.668788684	1						
0.352760547	0.305121229	-0.18896008	0.416245952	-0.225863941	0.407473209	1					
0.696415668	0.754165642	-0.380380798	0.918454505	-0.577330189	0.915867791	0.634033919	1				
0.047224029	0.000974239	0.015519952	0.016078401	-0.011924356	0.020702065	0.057735084	0.037288532	1			
-0.020831745	0.08431605	0.027639582	0.067167531	-0.034244161	0.055649154	-0.001555436	0.002102396	0.029503016	1		
-0.07963958	-0.064680446	-0.029503916	-0.075325212	0.02315995	-0.081423918	0.020177858	-0.076074393	0.282134569	0.227328812	1	
-0.123066349	0.211859445	-0.100466005	0.10095697	0.111816971	0.070179293	0.367726051	0.197593065	0.018129599	-0.028105936	0.012008375	1

	primary_mi	secondary_mi	tertiary_mi	resi_road_mi	path_mi	cycleway_mi	cwy_lneall_mi	cwy_binary	cwy_trkall_mi	cwy_trkbin	footway_mi	meanspeed	bikeprking	bus_stops
Strava_MADT														
primary_mi	1													
secondary_mi	0.523408939	1												
tertiary_mi	0.824922047	0.637488793	1											
resi_road_mi	0.645419163	0.77167692	0.767868267	1										
path_mi	-0.128328754	-0.04253292	-0.209819792	-0.089179456	1									
cycleway_mi	-0.466634729	-0.424748087	-0.205083588	-0.296308163	0.021417036	1								
cwy_lneall_mi	0.850210588	0.380093968	0.782585962	0.490163978	-0.078725765	-0.130332371	1							
cwy_binary	0.546511366	0.311170616	0.57133152	0.397691472	-0.220609321	-0.068494081	0.674642415	1						
cwy_trkall_mi	0.864593598	0.360423058	0.743348567	0.433038222	-0.161990633	-0.450954275	0.741087924	0.399157649	1					
cwy_trkbin	0.862352457	0.5067131	0.790520821	0.55450572	-0.043488382	-0.428173927	0.758684827	0.449489441	0.888024527	1				
footway_mi	0.918763391	0.554298513	0.893095464	0.636354462	-0.121671323	-0.254990235	0.956624822	0.637479676	0.800537586	0.841961165	1			
meanspeed	-0.001444671	-0.016879983	0.069226254	0.055419109	-0.36537853	0.156852428	0.054152169	0.489277285	-0.015541457	0.006596638	0.019762427	1		
bikeprking	0.761699376	0.375204922	0.740908131	0.731290788	-0.051738723	-0.22943706	0.636521985	0.422751342	0.712961762	0.783377141	0.703916144	0.044619296	1	
bus_stops	0.904659789	0.296741007	0.795165599	0.409533697	-0.105643505	-0.267554621	0.883581917	0.49922977	0.844991616	0.857022456	0.913101667	-0.03648942	0.680800133	1
intd_1mi_acres	0.810945586	0.442253645	0.833612994	0.50636912	-0.130280182	-0.063543127	0.958449627	0.690416092	0.695687499	0.732009156	0.951776662	0.113913645	0.575320626	0.863181622
lanes	0.363397069	0.016678359	0.212186757	-0.107392778	-0.07339208	-0.191576748	0.330271238	0.238615134	0.361910953	0.411668735	0.303251256	-0.057013554	0.107328041	0.478952766
bridge	0.31263901	0.176266721	0.360220547	0.268156681	-0.31516912	0.039520739	0.382949736	0.567633649	0.226575312	0.255145332	0.365091504	0.679235693	0.243563079	0.283379616
water_area_acres	-0.483232521	-0.286672317	-0.508347941	-0.380718028	0.247464797	0.012138133	-0.541579462	-0.594920397	-0.375752543	-0.420943003	-0.570917452	-0.160042065	-0.3863648	-0.457775809
wtr_cntdis_mi	0.132977965	0.355363499	0.088907254	0.234697122	-0.287259558	-0.29358965	0.011510855	0.17106596	-0.012261533	0.064908945	0.066694422	0.246016276	-0.076148064	-0.054152534
park_area_acres	-0.404225812	-0.223978326	-0.344485136	-0.274441634	0.520382963	0.138147038	-0.403634805	-0.511034721	-0.306847688	-0.319243072	-0.404851323	-0.271381698	-0.288046666	-0.340335339
prk_cntdis_mi	-0.34355653	-0.228677074	-0.378981112	-0.261723276	-0.098264329	-0.006658155	-0.401919825	-0.429134033	-0.298932427	-0.34063237	-0.402111978	-0.282141602	-0.296276116	-0.331611821
frest_area_acres	-0.255707985	-0.138888811	-0.252139223	-0.175840575	-0.016183822	0.260729591	-0.223748618	-0.21283292	-0.196371972	-0.218333898	-0.26794351	0.411729183	-0.103874939	-0.216043021
frst_cntdi_mi	-0.322419848	-0.231284335	-0.252242045	-0.272483327	0.636428495	0.227466117	-0.29470029	-0.508427073	-0.214538999	-0.266767435	-0.295943325	-0.553991894	-0.24247062	-0.217583869
grass_area_acres	0.582954421	0.445541799	0.554607259	0.656794177	0.144031897	-0.015394364	0.745302099	0.460521825	0.327490725	0.503832062	0.721483358	-0.017846995	0.566708457	0.523639596
grs_cntdis_mi	-0.509259916	-0.299879644	-0.536646607	-0.386574341	0.301650062	0.054306873	-0.611702163	-0.771472534	-0.382510675	-0.431285292	-0.597722888	-0.502760564	-0.407489232	-0.479641496
comm_area_acres	0.754908424	0.407252723	0.633459587	0.2399348	-0.098423374	-0.396841579	0.70605775	0.406173888	0.67158536	0.744006153	0.770998805	-0.044461329	0.353149468	0.861864662
comm_cntdi_mi	-0.382644891	-0.225926565	-0.447551359	-0.330918573	-0.271838144	0.01678738	-0.49422132	-0.285548193	-0.35441421	-0.399852936	-0.496661089	0.154362228	-0.373152622	-0.443959396
ind_area_acres	0.183064759	0.226345371	0.066372957	0.332759879	0.083412395	-0.00220672	0.298760129	0.197639177	-0.195430445	-0.114283876	0.244393954	0.019961807	-0.016342018	0.051205616
ind_cntdis_mi	-0.432137051	-0.277462833	-0.464722671	-0.374495269	0.533755506	0.032872146	-0.544964593	-0.578056926	-0.303195072	-0.366091742	-0.522460631	-0.490846974	-0.344993392	-0.399059941
resi_area_acres	-0.018521758	0.027262688	0.047966118	0.077222657	0.108616711	0.186553856	0.094024864	0.269790486	-0.013077067	0.066814683	0.039202394	0.284268627	0.069356105	-0.021180902
resi_cntdi_mi	-0.497435071	-0.334299942	-0.551611479	-0.4160158	-0.137497889	0.019583303	-0.614508776	-0.705118053	-0.400887013	-0.4564737	-0.60612267	-0.32024991	-0.435856563	-0.483304618
ret_area_acres	0.671405678	0.425453734	0.573588974	0.622400104	0.119663788	-0.101979484	0.735726264	0.456314772	0.448980448	0.567106677	0.716217622	0.000930398	0.6291927	0.558603891
ret_cntdis_mi	-0.561018039	-0.319148725	-0.608777907	-0.443766376	0.275887327	0.0251286	-0.680204204	-0.737741581	-0.438971628	-0.495972311	-0.670145166	-0.36755722	-0.462896759	-0.548938971
school	0.862195965	0.775813816	0.865648362	0.795939774	-0.155433277	-0.455187868	0.737218738	0.476280917	0.812116457	0.850587824	0.858380443	0.013975688	0.737443284	0.712686765
distschool_mi	-0.636757204	-0.416322377	-0.578239472	-0.492767617	-0.018448037	0.223981058	-0.636885821	-0.567272158	-0.457869045	-0.509080842	-0.652101349	-0.224960901	-0.45750007	-0.55691513
college	0.63509245	0.156583371	0.500703137	0.078977403	-0.137298059	-0.366114812	0.47097357	0.268411785	0.565171663	0.597148142	0.549194345	-0.043881712	0.290815094	0.774742411
distcolleg_mi	-0.533800485	-0.340255271	-0.428434978	-0.34939111	-0.055015395	0.010462046	-0.487205752	-0.16583781	-0.407131984	-0.457823468	-0.5111949	0.147998411	-0.376665168	-0.504620341
university	0.768506074	0.486147046	0.685624902	0.556724624	-0.141311312	-0.141631422	0.799807824	0.548688309	0.559892445	0.630492095	0.800346496	0.089004424	0.574139947	0.700262108
distuniver_mi	-0.670657634	-0.410724574	-0.614311406	-0.490461116	0.377377808	0.113823951	-0.696654262	-0.551875179	-0.511207375	-0.564634253	-0.701270209	-0.143189983	-0.514984628	-0.622433184
dist_cbd_mi	-0.553966374	-0.282654262	-0.461908079	-0.31774474	0.341995076	0.131027209	-0.599289306	-0.506794496	-0.406810628	-0.443945127	-0.583738016	-0.061800396	-0.368699269	-0.529506051
slope	-0.085902708	-0.096990485	-0.224612353	-0.068042278	0.087422229	0.016809606	-0.116295714	-0.258178723	-0.159216901	-0.183222323	-0.130701977	-0.412106703	-0.045864435	-0.13057934
white	0.124435793	0.013435337	0.097792727	0.110113819	0.013568709	-0.055642138	0.195063692	0.321919608	0.040621555	0.129637407	0.160940622	0.322237532	0.114737816	0.166598725
afam	-0.183744529	-0.204560074	-0.20993109	-0.199775938	0.012291287	0.123224229	-0.226392755	-0.372523945	-0.125854452	-0.209621808	-0.236488047	-0.326858593	-0.118642321	-0.192015133
male	-0.042583032	0.238532168	0.012650776	-0.019393819	0.162255589	-0.051049099	-0.114664919	-0.111880885	-0.036806186	-0.101182258	-0.039758179	-0.241705788	-0.178030759	-0.090102184
female	0.042583032	-0.238532168	-0.012650776	0.019393819	-0.162255589	0.051049099	0.114664919	0.111880885	0.036806186	0.101182258	0.039758179	0.241705788	0.178030759	0.090102184
lst_col	-0.217519495	0.255016058	-0.131183581	0.06376086	0.157331492	-0.173566418	-0.256769726	0.087097525	-0.210529775	-0.103666559	-0.193286741	0.09781585	-0.186914821	-0.353693777
stu_acc_100	0.045410618	0.428619185	0.180841774	0.163106808	-0.120872031	-0.082612296	-0.065434558	0.04657584	0.128524101	0.117817186	0.051956633	-0.013754235	-0.057104698	-0.0509372
hh_den_acres	-0.296653482	-0.141856916	-0.328997271	-0.254180145	0.18189366	0.042238883	-0.414768648	-0.263563951	-0.233541919	-0.283921276	-0.377723215	-0.264717289	-0.25027924	-0.347436775
med_inc_10k	-0.265803593	-0.223372527	-0.276338198	-0.190518541	0.014082893	-0.136629115	-0.213668603	0.110315376	-0.286913216	-0.289537464	-0.256429811	0.220067088	-0.247680438	-0.275825048
pop_den_acres	-0.213291019	0.017757957	-0.1786092	-0.139232403	0.074912921	0.024161653	-0.33462022	-0.175076729	-0.128668536	-0.183883414	-0.270593923	-0.192583357	-0.194677992	-0.285043935
med_age	-0.177813967	-0.430089367	-0.343223477	-0.270701301	0.059369396	-0.038407013	-0.158149289	-0.010780961	-0.200562159	-0.085637522	-0.244832229	0.169183095	-0.055814194	-0.100901498
emp_den_acres	-0.205928714	0.046936577	-0.173381865	-0.132146617	0.091480283	0.013930913	-0.33321093	-0.177175198	-0.126461682	-0.176939417	-0.263759434	-0.202552544	-0.204067521	-0.28296642
bik_pct	-0.009863022	0.157274405	0.018210803	0.031558522	-0.032533152	-0.055207614	-0.10296473	0.083324375	0.106510836	0.074865362	-0.052431995	0.003783584	0.044298807	-0.106741194
bik_den_acres	-0.057165408	0.188104794	-0.025800357	0.01080679	0.019818613	-0.119312619	-0.223473605	-0.146738229	0.053544959	-0.016920823	-0.149779415	-0.201809504	-0.052009293	-0.191448817
temp	0.183790134	0.112427287	0.18264546	0.132844372	-0.211012276	-0.094258351	0.200345497	0.255443272	0.149858013	0.161128987	0.20301123	0.3100316	0.145048809	0.166366012
hum	-0.256011773	-0.160268046	-0.261214213	-0.209114085	0.21951466	0.114430779	-0.249203567	-0.280585788	-0.207217594	-0.237225026	-0.268414315	-0.317817362	-0.209292977	-0.234310244
prec	0.099013441	0.055817382	0.13136876	0.086771495	-0.239205906	-0.037106898	0.12802034	0.19524967	0.057606358	0.07787123	0.124093622	0.254871924	0.056587242	0.095470131
Count_Average_MADT	0.210785423	0.207583891	0.191362827	0.313579127	0.221061706	-0.057978452	0.326956456	0.349433283	0.107850838	0.264421248	0.295947902	0.131242166	0.298889605	0.15232566

intd_1mi_acres	lanes	bridge	water_area_acres	wtr_cntdis_mi	park_area_acres	prk_cntdis_mi	frest_area_acres	frst_cntdi_mi	grass_area_acres	grs_cntdis_mi	comm_area_acres	comm_cntdi_mi	ind_area_acres	ind_cntdis_mi
1														
0.283839649	1													
0.433587868	0.154892808	1												
-0.556464025	-0.030411405	-0.355251381	1											
-0.015944237	0.043157269	0.307585993	-0.062345697	1										
-0.417001644	-0.135099624	-0.173551525	0.280693624	-0.216683065	1									
-0.420072148	-0.279875747	-0.332953573	0.047555421	-0.04067389	0.101406808	1								
-0.226605575	-0.108646016	0.149481877	0.308064734	-0.14463781	0.242574623	-0.036194861	1							
-0.280540689	-0.199548161	-0.606639892	0.358974046	-0.30123107	0.375651364	0.064505992	-0.227639854	1						
0.70562842	-0.054836354	0.291617262	-0.42555947	0.044411879	-0.216668308	-0.223673943	-0.163662575	-0.240397609	1					
-0.641522605	-0.066426871	-0.675352998	0.526703162	-0.211503952	0.328731937	0.169459879	-0.037525079	0.628503246	-0.496057728	1				
0.718631861	0.59830576	0.230557966	-0.37478667	0.103164631	-0.282296639	-0.289030285	-0.194811495	-0.223565007	0.32224006	-0.384795779	1			
-0.456128216	-0.177177284	-0.11592151	0.46451684	0.375585608	-0.045531846	0.232552074	0.095905968	0.016180952	-0.440819544	0.186808594	-0.362866258	1		
0.242180161	-0.195554977	0.193919073	-0.103936476	0.347572662	-0.209597578	-0.01221316	-0.070112164	-0.182841602	0.62664908	-0.272035178	0.045559756	-0.135643223	1	
-0.560673835	-0.084294918	-0.433664341	0.277517114	-0.203332742	0.642680834	0.316569965	-0.120733865	0.676366705	-0.438206631	0.697236052	-0.317535931	0.137345072	-0.364433775	1
0.134744823	0.136028425	0.187652975	-0.261191166	-0.099054929	-0.109962724	-0.135325506	0.0769714	-0.214645707	0.126851942	-0.197011475	-0.056144522	-0.272463277	0.025026306	-0.161985583
-0.646547671	-0.23384719	-0.625682693	0.404500626	0.039608372	0.140442462	0.60861124	0.025831996	0.271141809	-0.499816252	0.64908578	-0.383782331	0.543945331	-0.193163709	0.418929736
0.652481266	0.014309939	0.259986285	-0.42089118	0.16095673	-0.296607494	-0.22806459	-0.215677885	-0.256454403	0.852798882	-0.440994186	0.322276534	-0.405472648	0.561161605	-0.441921696
-0.691448162	-0.165282203	-0.673660046	0.777514127	-0.054336285	0.372849644	0.117618269	0.146052527	0.558269018	-0.534256369	0.758504911	-0.445499238	0.590935892	-0.262702506	0.547337409
0.744259324	0.187776548	0.295070678	-0.417888812	0.206640205	-0.382733508	-0.337029018	-0.248893142	-0.276656231	0.5835232	-0.475458864	0.6085917	-0.343281636	0.110156029	-0.42884929
-0.67698915	-0.19391236	-0.519506046	0.218314775	-0.245892539	0.459709227	0.592697003	0.250210211	0.149814658	-0.509127152	0.418262656	-0.45458222	0.22524178	-0.314449393	0.439136845
0.489054016	0.675497864	0.152359561	-0.250997783	0.049709713	-0.202896836	-0.206744367	-0.130199829	-0.155589894	0.022588618	-0.253784548	0.884796131	-0.240887858	-0.142331922	-0.182692878
-0.516924352	-0.133532097	0.040367493	0.162223907	-0.071155975	0.462476907	0.009653989	0.111242138	-0.167190287	-0.388455409	0.137959501	-0.420242058	0.011716696	-0.194649475	0.173233321
0.816017454	0.233233297	0.404676496	-0.306238463	0.231367583	-0.486474362	-0.302205216	-0.203048848	-0.256465976	0.657117841	-0.58551382	0.562817697	-0.117288241	0.398941025	-0.566322751
-0.717568445	-0.226155301	-0.350428406	0.447516845	-0.270348688	0.689368586	0.114694504	0.263617489	0.352854508	-0.535187342	0.521215881	-0.491343817	-0.046830268	-0.277444279	0.550736099
-0.608038937	-0.262491483	-0.282626531	0.451184928	-0.24056495	0.700840161	0.078023293	0.453239855	0.247090847	-0.411729719	0.370227849	-0.43717134	-0.099615814	-0.219559617	0.384835697
-0.230861254	-0.128317223	-0.401597383	-0.016725277	0.025668105	0.178990999	0.329536558	-0.116704716	0.150658614	0.044977943	0.280427549	-0.142931805	0.129854565	0.208275455	0.330471677
0.17560893	0.301950809	0.298737458	-0.145640106	-0.2206901	0.034772945	-0.466548408	0.251227329	-0.387987767	0.23229883	-0.077669382	0.181280756	-0.449835649	0.085865905	-0.210552228
-0.231356826	-0.293871813	-0.310900626	0.202709963	0.11881575	0.024168943	0.494737393	-0.237414388	0.394674684	-0.2252928	0.103395748	-0.24668519	0.499290452	-0.049637566	0.223369148
0.005930951	-0.124803117	-0.203557291	-0.029802688	-0.253500798	0.178406981	0.218108938	0.03209655	0.208155371	-0.147213933	0.172262596	-0.012489396	-0.053953306	-0.223859564	0.313903372
-0.005930951	0.124803117	0.203557291	0.029802688	0.253500798	-0.178406981	-0.218108938	-0.03209655	-0.208155371	0.147213933	-0.172262596	0.012489396	0.053953306	0.223859564	-0.313903372
-0.247051928	-0.12525233	-0.149519038	-0.05233967	0.111129331	-0.104194118	-0.160109721	-0.002889639	-0.029503091	-0.107744394	0.063200642	-0.229533762	-0.117680405	-0.028163487	-0.037749019
-0.008714449	-0.004500368	0.053528857	-0.066073432	0.433140711	-0.191667323	0.148505434	-0.092085803	0.025607902	-0.11031062	-0.039887068	0.059371673	0.137193393	-0.12387071	0.127659608
-0.39884286	-0.294759949	-0.39920934	0.082719432	0.11498499	0.135784742	0.446430396	-0.222296249	0.45533368	-0.342750769	0.227338067	-0.312940704	0.52295981	-0.260376614	0.545338683
-0.251283861	-0.03044917	0.012445307	0.045847471	-0.079768046	0.004185708	-0.327278931	0.164181776	-0.168297965	-0.201292328	0.096948294	-0.207816595	-0.125084247	0.054094722	-0.133506419
-0.301864445	-0.259330423	-0.246558681	0.032295468	0.24447204	0.050791183	0.396096266	-0.216659196	0.344716425	-0.302363043	0.121101742	-0.243863552	0.47361211	-0.249933642	0.448711819
-0.226963625	0.129112306	-0.037077771	0.002726777	-0.154277477	0.098505649	-0.026240036	0.070555244	-0.186722907	-0.077238081	0.034694142	-0.090463291	0.016140419	-0.050867365	-0.049894492
-0.294084395	-0.248192924	-0.258321863	0.02961352	0.238148053	0.046940055	0.384429784	-0.211117518	0.357462226	-0.302327084	0.136199271	-0.228960095	0.457765967	-0.251353065	0.457071561
-0.114531583	-0.104033086	0.001954776	-0.056847311	0.130639959	-0.100182859	0.118969855	-0.285990408	0.070428468	-0.164764461	-0.129840535	-0.068922781	0.063891357	-0.200473362	0.057388374
-0.215871495	-0.185318829	-0.141610608	0.043385033	0.306583972	-0.020965792	0.384728626	-0.254840303	0.224591655	-0.263811271	0.011795871	-0.158712387	0.342862093	-0.214313037	0.289613139
0.206376512	0.030841484	0.298152579	-0.107077073	0.154239884	-0.089959274	-0.181036156	0.107823596	-0.35511684	0.148449724	-0.33738181	0.149986928	0.031836856	0.088654807	-0.316771061
-0.25934576	-0.086432664	-0.382609537	0.150023275	-0.19839141	0.066663578	0.244058159	-0.08941769	0.393252445	-0.202072263	0.335097298	-0.208918669	0.016992356	-0.112596582	0.307789538
0.147466812	0.040867952	0.257718459	-0.021636004	0.189190909	-0.031657828	-0.187258187	0.036296483	-0.303405112	0.105678163	-0.254847721	0.086422934	0.149994678	0.069538846	-0.271981166
0.253174486	-0.038048056	0.301216728	-0.303784784	0.001287911	0.058972575	-0.143392786	-0.135678132	-0.267601414	0.549395285	-0.395914304	0.074499821	-0.363024665	0.306002136	-0.22132146

resi_area_acres	*resi_cntdi_mi*	*ret_area_acres*	*ret_cntdis_mi*	*school*	*distschool_mi*	*college*	*distcolleg_mi*	*university*	*distuniver_mi*	*dist_cbd_mi*	*slope*	*white*	*afam*	*male*	*female*	*lst_col*
1																
-0.334954127	1															
0.152280708	-0.464425528	1														
-0.252729906	0.635060993	-0.507311736	1													
0.022416994	-0.473871669	0.581005948	-0.514049516	1												
-0.055461188	0.664879047	-0.527017	0.41893913	-0.583165067	1											
-0.114688389	-0.232470953	0.079653402	-0.293989026	0.381838137	-0.319033665	1										
0.018509907	0.15581602	-0.405972829	0.156058387	-0.51238586	0.405522434	-0.307135679	1									
-0.041369318	-0.447776478	0.63877616	-0.527893218	0.729152003	-0.701647954	0.376497183	-0.663717565	1								
-0.090929002	0.307884248	-0.550252609	0.495274015	-0.640565055	0.539691869	-0.334297996	0.722457036	-0.831485616	1							
-0.054491327	0.143627269	-0.435968829	0.428528845	-0.486148557	0.499360584	-0.310769363	0.554998079	-0.719397963	0.897944924	1						
-0.089097953	0.420540734	0.08374775	0.213244071	-0.12278978	0.532596012	-0.160968134	-0.066960358	-0.135227738	0.004324819	-0.034053998	1					
0.225198641	-0.381279963	0.074401808	-0.272591458	0.032838618	-0.159887767	0.165184976	0.316228399	-0.075374141	0.129725245	0.152642279	-0.194366246	1				
-0.228742497	0.4515462	-0.113141206	0.308657303	-0.154098434	0.225204094	-0.18871482	-0.25815165	0.069736335	-0.110729671	-0.167003055	0.277373937	-0.947159238	1			
-0.112128527	0.035721521	-0.140820772	0.071957496	-0.007393728	0.028764325	-0.023512776	-0.1408457	-0.062103842	0.109251848	0.111791907	-0.139611472	-0.16376037	0.0323628	1		
0.112128527	-0.035721521	0.140820772	-0.071957496	0.007393728	-0.028764325	0.023512776	0.1408457	0.062103842	-0.109251848	-0.111791907	0.139611472	0.16376037	-0.0323628	-1	1	
0.148960428	-0.080759664	-0.057900646	0.055750868	-0.050134451	0.051828023	-0.2414553	0.285722002	-0.347679455	0.31069995	0.249429195	-0.170119315	0.175942814	-0.326330286	0.073158887	-0.073158887	1
-0.006444703	-0.057722008	0.01067788	0.018059258	0.240623628	-0.077322778	-0.0116538	-0.247602726	-0.037662045	-0.11066809	-0.034362992	-0.100327985	-0.368028587	0.100356985	0.249724399	-0.249724399	0.304952587
-0.211816924	0.35834264	-0.264546157	0.449128797	-0.271592743	0.282613781	-0.226548961	-0.126499465	-0.283068169	0.129382101	0.047544956	0.200293566	-0.670068414	0.610346128	0.29376033	-0.29376033	0.080920489
0.068603375	-0.072295088	-0.227008408	0.046246203	-0.340797294	0.042572138	-0.109055358	0.560411369	-0.380127551	0.394452558	0.272285347	-0.144479864	0.552943191	-0.509147175	-0.133417871	0.133417871	0.510873239
-0.182184254	0.24006803	-0.185541804	0.33542046	-0.131112361	0.174338843	-0.19934194	-0.153542947	-0.215283348	0.053749708	0.014873322	0.093500678	-0.68391825	0.55215049	0.2994946	-0.2994946	0.133631175
0.292226169	0.132300286	-0.18805925	0.039981359	-0.314487399	0.237226318	0.043197042	0.340043912	-0.238477768	0.159307669	0.06853176	0.146152178	0.419456251	-0.201241603	-0.43735775	0.43735775	-0.015729051
-0.174954322	0.231570414	-0.19062861	0.337149812	-0.117554569	0.16501406	-0.188780569	-0.177031617	-0.21019716	0.053830907	0.017787523	0.087031591	-0.669948846	0.528666349	0.335413491	-0.335413491	0.15669135
-0.164131493	-0.009475122	-0.024070576	-0.085408992	0.129078481	0.071413609	-0.108068912	0.031072718	-0.009262636	0.01275562	-0.025738939	-0.033334187	-0.390170323	0.307617301	-0.044771401	0.044771401	0.200609841
-0.155016595	0.17143761	-0.073692318	0.168890703	0.081206179	0.090329463	-0.150495688	-0.171821341	-0.040742641	-0.034344021	-0.043242828	0.02290385	-0.739090721	0.599772078	0.205685757	-0.205685757	0.116570928
0.040346934	-0.181071217	0.127753443	-0.204450358	0.173721794	-0.149915571	0.107623892	0.109235418	0.186178584	-0.135906603	-0.086021939	-0.101928347	0.162917778	-0.160957473	-0.155900419	0.155900419	-0.017114581
-0.063174168	0.242713616	-0.19442823	0.254233166	-0.235804769	0.197070259	-0.159026214	-0.086002735	-0.230591361	0.179973	0.105982128	0.081637133	-0.243667456	0.251250551	0.138572435	-0.138572435	0.064606802
-0.021089335	-0.104876228	0.075538517	-0.08775403	0.092717179	-0.110789172	0.061084487	0.134142802	0.153096756	-0.136035268	-0.097159005	-0.074463742	0.134383683	-0.105835033	-0.166453521	0.166453521	-0.097365237
0.13074526	-0.439195786	0.48660939	-0.418969643	0.226546406	-0.230690961	-0.121506978	0.174294114	0.180885033	-0.053398069	-0.035359116	-0.124784819	0.28967048	-0.264135225	-0.199223451	0.199223451	0.129050823

stu_acc_100	*hh_den_acres*	*med_inc_10k*	*pop_den_acres*	*med_age*	*emp_den_acres*	*bik_pct*	*bik_den_acres*	*temp*	*hum*	*prec*	*Count_Average_MADT*
1											
0.482406665	1										
-0.380890508	-0.337971836	1									
0.70612726	0.952289538	-0.410368642	1								
-0.655998768	-0.280067949	0.443962381	-0.462533067	1							
0.722550465	0.949252967	-0.406693251	0.997732093	-0.475493421	1						
0.370009051	0.325011613	-0.334802828	0.406551157	-0.27114459	0.397617259	1					
0.69526697	0.744718983	-0.506508018	0.84575694	-0.523452593	0.841291159	0.674320287	1				
-0.064009201	-0.213246401	0.063501336	-0.170390473	0.060940528	-0.179106094	-0.02529639	-0.131400441	1			
0.058484565	0.285607705	-0.062714739	0.228283995	-0.070829006	0.234385811	0.082947063	0.196482019	-0.209044831	1		
-0.112770983	-0.186500725	0.043517324	-0.159239116	0.099297001	-0.171720937	-0.045365513	-0.145698692	0.322859405	-0.079901824	1	
-0.105138509	-0.179890038	0.030938226	-0.151896939	0.112276164	-0.165059836	0.218923681	-0.045352757	0.194708479	-0.196246828	0.095081652	1

LASSO	
Var	Mod_Importance
bik_den_acres	854.903
prk_cntdis_mi	193.882
wtr_cntdis_mi	184.100
prec	144.383
slope	67.427
college	63.038
resi_cntdi_mi	53.395
hh_den_acres	48.391
cycleway_mi	48.125
ind_dist_mi	33.585
grass_dist_mi	23.154
lanes	18.106
secondary_mi	8.600
tertiary_mi	6.768
path_mi	5.406
Strava_MADT	3.480
temp	3.435
male	3.353
bik_pct	3.276
resi_road_mi	1.633
ind_area_acres	1.587
dist_cbd_mi	0.914
ret_area_acres	0.876
grass_area_acres	0.785
lst_col	0.587
med_age	0.476
hum	0.116
water_dist_mi	0.058
female	0.000
primary_mi	0.000
cwy_lneall_mi	0.000
cwy_binary	0.000
cwy_trkall_mi	0.000
cwy_trkbin	0.000
footway_mi	0.000
meanspeed	0.000
bikeprking	0.000
bus_stops	0.000
intd_1mi_acres	0.000
water_area_acres	0.000
park_area_acres	0.000
park_dist_mi	0.000
frest_area_acres	0.000
forest_dis_mi	0.000
frst_cntdi_mi	0.000
grs_cntdis_mi	0.000
comm_area_acres	0.000
comm_dist_mi	0.000
comm_cntdi_mi	0.000
ind_cntdis_mi	0.000
resi_area_acres	0.000
resi_dist_mi	0.000
ret_dist_mi	0.000
ret_cntdis_mi	0.000
school	0.000
distschool_mi	0.000
distcolleg_mi	0.000
university	0.000
distuniver_mi	0.000
white	0.000
afam	0.000
stu_acc_100	0.000
med_inc_10k	0.000
pop_den_acres	0.000
emp_den_acres	0.000

RF				
Var	IncNodePurity	1st quartile	20th percentile	15th percentile
Strava_MADT	773153.432	15656.816	14418.273	9792.871
ret_area_acres	355543.873			
bik_pct	245756.118			
slope	202866.741			
distcolleg_mi	162669.200			
resi_road_mi	159260.817			
white	146033.052			
grass_area_acres	140153.226			
bikeprking	138972.318			
bik_den_acres	135312.648			
secondary_mi	131961.806			
afam	118347.192			
temp	114903.711			
hh_den_acres	110472.078			
male	108872.592			
med_age	100715.063			
lst_col	92938.729			
female	89455.259			
med_inc_10k	86376.835			
prec	83950.767			
pop_den_acres	82474.365			
path_mi	73657.552			
hum	71247.310			
resi_cntdi_mi	69736.969			
stu_acc_100	69532.885			
prk_cntdis_mi	68960.765			
comm_dist_mi	68825.853			
ind_dist_mi	66297.792			
emp_den_acres	64952.455			
distuniver_mi	63987.806			
comm_cntdi_mi	62533.532			
ind_cntdis_mi	62345.731			
dist_cbd_mi	57741.210			
ret_dist_mi	51928.508			
grs_cntdis_mi	48131.020			
comm_area_acres	43378.929			
ret_cntdis_mi	36812.139			
ind_area_acres	36117.531			
frst_cntdi_mi	35529.432			
resi_dist_mi	32555.428			
footway_mi	30515.470			
forest_dis_mi	29084.628			
tertiary_mi	28275.471			
wtr_cntdis_mi	23883.805			
cycleway_mi	21770.634			
water_area_acres	20652.024			
park_area_acres	19550.973			
school	17174.969			
distschool_mi	15656.816			
intd_1mi_acres	15641.135			
frest_area_acres	14910.763			
water_dist_mi	14853.745			
grass_dist_mi	12676.385			
park_dist_mi	10514.100			
resi_area_acres	10111.041			
primary_mi	9315.617			
lanes	6458.178			
cwy_lneall_mi	4942.736			
college	3741.111			
bus_stops	3165.293			
meanspeed	3091.501			
cwy_trkall_mi	2701.300			
university	804.202			
cwy_trkbin	728.245			
cwy_binary	662.142			

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Presence of bridge was excluded as there was no more than 1 category was available
 # Number of jobs variable was excluded, and employment density was included instead

LASSO	
Var	**Mod_Importance**
bik_den_acres	8939.803
distschool_mi	958.775
prk_cntdis_mi	562.823
slope	175.791
wtr_cntdis_mi	167.842
resi_cntdi_mi	152.856
cycleway_mi	109.499
water_dist_mi	80.965
secondary_mi	78.365
forest_dis_mi	70.862
cwy_trkall_mi	66.477
path_mi	64.972
lanes	51.915
stu_acc_100	45.508
primary_mi	43.525
park_dist_mi	43.483
distcolleg_mi	37.506
Year	31.752
prec	30.066
ret_area_acres	28.102
bik_pct	24.705
white	21.951
distuniver_mi	17.677
male	17.648
pop_den_acres	15.764
afam	15.207
med_inc_10k	12.151
Strava_MADT	9.833
lst_col	9.827
Location_Number	9.465
grass_area_acres	7.155
hum	3.954
grass_dist_mi	3.464
meanspeed	3.233
cwy_lneall_mi	3.078
bikeprking	2.643
school	1.717
med_age	1.598
temp	0.576
park_area_acres	0.502
comm_area_acres	0.094
cwy_binary	0.027
tertiary_mi	0.000
female	0.000
Location	0.000
Month	0.000
resi_road_mi	0.000
cwy_trkbin	0.000
footway_mi	0.000
bus_stops	0.000
intd_1mi_acres	0.000
water_area_acres	0.000
frest_area_acres	0.000
frst_cntdi_mi	0.000
grs_cntdis_mi	0.000
comm_dist_mi	0.000
comm_cntdi_mi	0.000
ind_area_acres	0.000
ind_dist_mi	0.000
ind_cntdis_mi	0.000
resi_area_acres	0.000
resi_dist_mi	0.000
ret_dist_mi	0.000
ret_cntdis_mi	0.000
college	0.000
university	0.000
dist_cbd_mi	0.000
hh_den_acres	0.000
emp_den_acres	0.000

RF				
Var	**IncNodePurity**	**1st quartile**	**20th percentile**	**15th percentile**
Strava_MADT	3121266.551	34195.595	28360.275	24173.405
slope	1234789.113			
path_mi	966232.324			
distcolleg_mi	794310.757			
bik_pct	713258.483			
bik_den_acres	616422.515			
ret_area_acres	524529.586			
male	480743.360			
grass_area_acres	472668.444			
lst_col	428095.758			
afam	420859.477			
female	416621.387			
bikeprking	413295.026			
hum	412492.365			
Location_Number	345860.194			
white	345130.590			
med_age	334883.143			
prec	290002.039			
hh_den_acres	274194.375			
ret_cntdis_mi	224076.459			
cycleway_mi	212036.505			
med_inc_10k	209657.316			
temp	185153.041			
stu_acc_100	175351.906			
pop_den_acres	160429.005			
secondary_mi	155681.953			
ret_dist_mi	154707.383			
ind_cntdis_mi	145897.105			
emp_den_acres	133741.331			
lanes	132983.679			
prk_cntdis_mi	125024.545			
Year	123449.493			
ind_dist_mi	120330.229			
grass_dist_mi	115730.023			
park_area_acres	115573.968			
distuniver_mi	109664.241			
grs_cntdis_mi	108289.851			
Month	102676.708			
tertiary_mi	95843.827			
Location	78517.552			
resi_road_mi	60276.803			
wtr_cntdis_mi	56028.187			
frest_area_acres	50565.757			
dist_cbd_mi	47473.787			
ind_area_acres	45978.122			
resi_cntdi_mi	44242.418			
footway_mi	43654.276			
distschool_mi	43367.435			
primary_mi	39071.978			
frst_cntdi_mi	35667.994			
comm_dist_mi	34318.262			
forest_dis_mi	34195.595			
resi_area_acres	30591.732			
comm_area_acres	28726.047			
water_area_acres	28647.631			
meanspeed	27929.240			
water_dist_mi	25734.352			
bus_stops	25199.080			
comm_cntdi_mi	23916.986			
resi_dist_mi	22014.957			
intd_1mi_acres	21932.302			
school	18832.948			
cwy_trkall_mi	17740.560			
cwy_lneall_mi	13907.442			
park_dist_mi	13732.503			
cwy_trkbin	3927.567			
cwy_binary	1428.617			
college	836.328			
university	376.443			

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Presence of bridge was excluded as there was no more than 1 category was available
 # Number of jobs variable was excluded, and employment density was included instead

LASSO	
Var	Mod_Importance
bik_den_acres	1329.468
wtr_cntdis_mi	427.814
cwy_binary	268.787
bridge	256.796
college	125.908
slope	93.795
grass_dist_mi	74.546
prec	71.781
lanes	67.920
path_mi	63.892
bik_pct	55.443
ret_dist_mi	49.287
frst_cntdi_mi	45.316
resi_cntdi_mi	44.863
distcolleg_mi	43.474
ind_cntdis_mi	31.480
distuniver_mi	24.376
emp_den_acres	23.183
cwy_trkbin	22.524
distschool_mi	19.298
comm_dist_mi	15.654
footway_mi	13.264
male	11.478
Strava_MADT	9.373
ret_area_acres	7.008
med_inc_10k	6.930
lst_col	6.631
cycleway_mi	6.099
ret_cntdis_mi	4.942
university	3.095
resi_road_mi	2.461
tertiary_mi	2.047
hum	2.024
afam	1.958
temp	1.522
meanspeed	1.371
ind_area_acres	1.274
secondary_mi	1.233
comm_area_acres	1.125
white	1.123
water_area_acres	1.026
park_area_acres	0.188
resi_area_acres	0.146
frest_area_acres	0.080
female	0.000
primary_mi	0.000
cwy_lneall_mi	0.000
cwy_trkall_mi	0.000
bikeprking	0.000
bus_stops	0.000
intd_1mi_acres	0.000
water_dist_mi	0.000
park_dist_mi	0.000
prk_cntdis_mi	0.000
forest_dis_mi	0.000
grass_area_acres	0.000
grs_cntdis_mi	0.000
comm_cntdi_mi	0.000
ind_dist_mi	0.000
resi_dist_mi	0.000
school	0.000
dist_cbd_mi	0.000
stu_acc_100	0.000
hh_den_acres	0.000
pop_den_acres	0.000
med_age	0.000

RF	
Var	IncNodePurity
Strava_MADT	13403269.333
bikeprking	2332578.001
slope	2295916.578
distcolleg_mi	2149963.767
grass_area_acres	1854308.435
path_mi	1579474.395
bik_pct	1409490.370
ind_cntdis_mi	1312311.982
grs_cntdis_mi	1272142.668
ret_area_acres	1236126.618
resi_dist_mi	1131153.155
secondary_mi	1106510.494
male	1076118.175
female	1072665.617
resi_cntdi_mi	957274.245
prec	922648.537
ret_cntdis_mi	857733.882
bik_den_acres	854332.024
ret_dist_mi	841596.629
hum	791881.890
resi_road_mi	790989.842
lst_col	672084.940
med_age	651769.691
ind_dist_mi	643646.777
white	612383.880
afam	605607.131
temp	594059.395
tertiary_mi	585807.752
hh_den_acres	564589.173
grass_dist_mi	511607.170
distuniver_mi	483015.508
med_inc_10k	471206.833
pop_den_acres	458791.606
emp_den_acres	442706.968
park_area_acres	395365.295
cycleway_mi	371302.268
stu_acc_100	313958.521
footway_mi	290747.616
comm_dist_mi	276602.914
prk_cntdis_mi	271045.729
lanes	252196.428
dist_cbd_mi	241415.192
school	219691.561
comm_area_acres	217118.308
primary_mi	166583.630
water_area_acres	164055.913
frst_cntdi_mi	159835.772
distschool_mi	157837.604
forest_dis_mi	157490.318
comm_cntdi_mi	132346.311
wtr_cntdis_mi	112201.531
resi_area_acres	111862.721
ind_area_acres	78054.691
water_dist_mi	75085.752
intd_1mi_acres	71884.422
meanspeed	59609.471
frest_area_acres	54113.011
cwy_lneall_mi	49565.919
park_dist_mi	31540.882
cwy_trkall_mi	25738.050
university	20848.880
bus_stops	16757.723
cwy_trkbin	16728.311
cwy_binary	16445.103
college	6577.993
bridge	4137.957

1st quartile	20th percentile	15th percentile
138632.313	78054.691	58235.356

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Number of jobs variable was excluded, and employment density was included instead

LASSO	
Var	Mod_Importance
bik_den_acres	4044.143
prk_cntdis_mi	463.143
wtr_cntdis_mi	170.806
prec	159.426
lanes	80.597
distschool_mi	65.964
cycleway_mi	62.963
secondary_mi	43.456
frst_cntdi_mi	40.287
tertiary_mi	36.055
pop_den_acres	32.282
resi_cntdi_mi	26.326
cwy_lneall_mi	17.897
ret_area_acres	16.293
bik_pct	13.140
male	8.988
grs_cntdis_mi	7.449
grass_area_acres	5.722
resi_road_mi	5.358
Strava_MADT	4.192
temp	3.603
bikeprking	3.565
dist_cbd_mi	3.458
white	2.870
lst_col	2.796
stu_acc_100	2.457
ind_area_acres	1.275
med_age	1.017
meanspeed	0.519
footway_mi	0.516
hum	0.510
distcolleg_mi	0.345
resi_area_acres	0.144
comm_area_acres	0.051
park_area_acres	0.013
school	0.000
afam	0.000
female	0.000
primary_mi	0.000
path_mi	0.000
cwy_binary	0.000
cwy_trkall_mi	0.000
cwy_trkbin	0.000
bus_stops	0.000
intd_1mi_acres	0.000
water_area_acres	0.000
frest_area_acres	0.000
comm_cntdi_mi	0.000
ind_cntdis_mi	0.000
ret_cntdis_mi	0.000
college	0.000
university	0.000
distuniver_mi	0.000
slope	0.000
hh_den_acres	0.000
med_inc_10k	0.000
emp_den_acres	0.000

RF					
Var	IncNodePurity	1st quartile	20th percentile	15th percentile	
Strava_MADT	709812.577	20839.747	14233.791	10026.946	
ret_area_acres	409346.608				9 positive coeff
slope	271237.202				18 positive coeff
bik_pct	254718.561				
distcolleg_mi	181419.933				14 negative coeff
resi_road_mi	176748.215				
grass_area_acres	165792.338				
secondary_mi	137907.815				
white	127777.782				
bik_den_acres	125182.256				
afam	123317.484				19 positive coeff
temp	121710.168				
bikeprking	118831.077				
hh_den_acres	113958.039				
female	105174.709				
med_age	102034.654				
comm_cntdi_mi	98778.466				4 negative coeff
lst_col	95978.906				21 positive coeff
prec	92452.089				25 negative coeff
male	89291.451				
med_inc_10k	86941.712				1 positive coeff
path_mi	85425.002				
prk_cntdis_mi	83152.028				
dist_cbd_mi	80836.427				17 positive coeff
distuniver_mi	78877.295				16 negative coeff
ind_cntdis_mi	74469.584				6 negative coeff
hum	72238.563				
emp_den_acres	71627.094				
ret_cntdis_mi	69020.351				10 positive coeff
resi_cntdi_mi	67378.378				23 negative coeff
stu_acc_100	62405.776				22 positive coeff
pop_den_acres	57755.030				20 positive coeff
ind_area_acres	50187.806				5 negative coeff
grs_cntdis_mi	48779.720				
frst_cntdi_mi	43845.500				
comm_area_acres	40118.042				3 negative coeff
footway_mi	32306.762				
cycleway_mi	28849.025				
wtr_cntdis_mi	28034.562				
park_area_acres	25426.298				26 negative coeff
water_area_acres	24596.368				2 negative coeff
tertiary_mi	22197.005				24 positive coeff
frest_area_acres	20839.747				
distschool_mi	18119.975				12 negative coeff
intd_1mi_acres	16391.233				
primary_mi	13694.431				
school	12380.880				11 negative coeff
resi_area_acres	11646.872				7 positive coeff
lanes	8946.996				
cwy_lneall_mi	8554.098				8 positive coeff
cwy_trkall_mi	5837.965				
meanspeed	5821.095				
bus_stops	3429.471				
college	2645.295				13 negative coeff
cwy_binary	1124.398				
cwy_trkbin	560.433				
university	407.660				15 positive coeff

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Presence of bridge was excluded as there was no more than 1 category was available
 # Number of jobs variable was excluded, and employment density was included instead

 # Water_dist or nearest distance to edge of water body from the count station has been excluded
 # Park_dist or nearest distance to edge of park or open space from the count station has been excluded
 # Forest_dist or nearest distance to edge of forest from the count station has been excluded
 # Grass_dist or nearest distance to edge of grass from the count station has been excluded
 # Comm_dist or nearest distance to edge of commercial area from the count station has been excluded
 # Ind_dist or nearest distance to edge of industrial area from the count station has been excluded
 # Resi_dist or nearest distance to edge of residential area from the count station has been excluded
 # Ret_dist or nearest distance to edge of retail area from the count station has been excluded

LASSO	
Var	Mod_Importance
bik_den_acres	6507.898
distschool_mi	531.320
wtr_cntdis_mi	224.709
prk_cntdis_mi	205.352
resi_cntdi_mi	154.611
cwy_binary	85.322
slope	51.581
lanes	49.339
primary_mi	46.537
bik_pct	33.022
distcolleg_mi	30.377
secondary_mi	25.714
emp_den_acres	21.682
white	20.228
distuniver_mi	17.743
grs_cntdis_mi	16.878
cwy_trkall_mi	15.291
path_mi	13.800
stu_acc_100	13.552
afam	10.812
Strava_MADT	9.623
male	7.832
lst_col	4.901
ret_area_acres	4.808
grass_area_acres	4.656
hum	3.713
med_age	1.976
med_inc_10k	1.872
comm_area_acres	1.805
hh_den_acres	1.022
temp	0.513
cwy_lneall_mi	0.473
park_area_acres	0.189
female	0.000
tertiary_mi	0.000
resi_road_mi	0.000
cycleway_mi	0.000
cwy_trkbin	0.000
footway_mi	0.000
meanspeed	0.000
bikeprking	0.000
bus_stops	0.000
intd_1mi_acres	0.000
water_area_acres	0.000
frest_area_acres	0.000
frst_cntdi_mi	0.000
comm_cntdi_mi	0.000
ind_area_acres	0.000
ind_cntdis_mi	0.000
resi_area_acres	0.000
ret_cntdis_mi	0.000
school	0.000
college	0.000
university	0.000
dist_cbd_mi	0.000
pop_den_acres	0.000
prec	0.000

RF					
Var	IncNodePurity	1st quartile	20th percentile	15th percentile	Month goes away first
Strava_MADT	3149033.873	48008.255	37522.660	32913.582	
path_mi	1065063.276				
slope	1062911.155				22 negative coeff
distcolleg_mi	879762.333				18 positive coeff
bik_pct	773904.446				1 negative coeff
bik_den_acres	668426.172				
ret_area_acres	587436.262				13 negative coeff
hum	523620.294				
grass_area_acres	502145.839				5 positive coeff
male	481430.610				
female	479385.650				
lst_col	475617.957				
ret_cntdis_mi	453154.550				14 negative coeff
bikeprking	404699.404				
afam	397016.940				24 negative coeff
white	341633.833				
hh_den_acres	308051.434				
med_age	305750.489				
prec	274861.213				2 negative coeff
med_inc_10k	235809.813				3 negative coeff
secondary_mi	231092.780				
pop_den_acres	228579.838				
cycleway_mi	225349.796				
ind_cntdis_mi	212931.091				
temp	210250.642				23 negative coeff
stu_acc_100	185493.807				
distuniver_mi	166612.420				20 positive coeff
prk_cntdis_mi	140458.227				
emp_den_acres	134485.298				
park_area_acres	128152.826				6 positive coeff
resi_road_mi	116627.236				25 negative coeff
grs_cntdis_mi	113860.108				8 negative coeff
wtr_cntdis_mi	97360.677				
frest_area_acres	95493.242				
tertiary_mi	90590.754				
distschool_mi	80169.983				16 negative coeff
lanes	79211.017				
dist_cbd_mi	76482.444				21 negative coeff
resi_cntdi_mi	69088.533				12 positive coeff
primary_mi	63703.493				
comm_cntdi_mi	51729.374				9 negative coeff
meanspeed	48967.729				
frst_cntdi_mi	48008.255				
comm_area_acres	46532.925				7 positive coeff
ind_area_acres	38007.677				
footway_mi	37401.406				4 positive coeff
intd_1mi_acres	34000.871				
bus_stops	33401.272				
water_area_acres	32588.456				
resi_area_acres	22882.379				11 negative coeff
cwy_lneall_mi	16785.214				
school	14706.124				15 negative coeff
cwy_trkall_mi	12168.492				
college	7954.669				17 positive coeff
cwy_binary	6444.690				10 positive coeff
cwy_trkbin	3986.048				
university	1932.203				19 positive coeff

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Presence of bridge was excluded as there was no more than 1 category was available
 # Number of jobs variable was excluded, and employment density was included instead

 # Water_dist or nearest distance to edge of water body from the count station has been excluded
 # Park_dist or nearest distance to edge of park or open space from the count station has been excluded
 # Forest_dist or nearest distance to edge of forest from the count station has been excluded
 # Grass_dist or nearest distance to edge of grass from the count station has been excluded
 # Comm_dist or nearest distance to edge of commercial area from the count station has been excluded
 # Ind_dist or nearest distance to edge of industrial area from the count station has been excluded
 # Resi_dist or nearest distance to edge of residential area from the count station has been excluded
 # Ret_dist or nearest distance to edge of retail area from the count station has been excluded

LASSO	
Var	Mod_Importance
bik_den_acres	1369.208
wtr_cntdis_mi	418.157
bridge	275.790
cwy_binary	274.875
college	130.695
slope	94.712
prec	72.363
grs_cntdis_mi	69.275
lanes	67.819
path_mi	62.346
bik_pct	55.441
ret_cntdis_mi	53.730
resi_cntdi_mi	47.234
frst_cntdi_mi	46.097
distcolleg_mi	42.761
distuniver_mi	24.196
ind_cntdis_mi	23.836
cwy_trkbin	22.266
emp_den_acres	22.123
distschool_mi	17.747
comm_cntdi_mi	15.111
footway_mi	13.106
male	10.870
cycleway_mi	9.468
Strava_MADT	9.366
med_inc_10k	7.660
ret_area_acres	6.669
lst_col	6.646
secondary_mi	2.414
resi_road_mi	2.385
tertiary_mi	2.243
hum	2.008
afam	1.675
meanspeed	1.622
temp	1.510
white	1.359
ind_area_acres	1.214
comm_area_acres	1.081
water_area_acres	0.872
female	0.440
park_area_acres	0.163
resi_area_acres	0.125
med_age	0.075
primary_mi	0.000
cwy_lneall_mi	0.000
cwy_trkall_mi	0.000
bikeprking	0.000
bus_stops	0.000
intd_1mi_acres	0.000
prk_cntdis_mi	0.000
frest_area_acres	0.000
grass_area_acres	0.000
school	0.000
university	0.000
dist_cbd_mi	0.000
stu_acc_100	0.000
hh_den_acres	0.000
pop_den_acres	0.000

RF						
Var	IncNodePurity	1st quartile	20th percentile	15th percentile		
Strava_MADT	12838538.481	173331.913	124121.042	79362.121	Month # 10	negative coeff
slope	2681738.263					
grass_area_acres	2617620.218					
distcolleg_mi	2460498.036				8	positive coeff
bikeprking	2384095.469					
path_mi	2060830.490					
ind_cntdis_mi	1843631.788					
resi_cntdi_mi	1547760.974					
bik_pct	1490086.979					
grs_cntdis_mi	1263222.696					
ret_cntdis_mi	1182060.024					
ret_area_acres	1083599.023				16	negative coeff
female	1067927.480					
secondary_mi	1021111.465				5	positive coeff
male	984116.889					
prec	916366.778					
hum	887777.916				17	negative coeff
resi_road_mi	852470.837					
bik_den_acres	848385.960					
lst_col	752345.648					
med_age	725331.895				3	positive coeff
distuniver_mi	668916.618				12	negative coeff
tertiary_mi	626132.197				7	negative coeff
temp	599287.967					
med_inc_10k	550087.231					
afam	536852.312					
white	528382.931				15	negative coeff
hh_den_acres	515963.455					
emp_den_acres	486751.353					
pop_den_acres	458928.138					
stu_acc_100	448349.977				14	negative coeff
footway_mi	409438.970					
prk_cntdis_mi	401491.861					
comm_area_acres	385993.983					
comm_cntdi_mi	374553.023					
park_area_acres	322648.420					
lanes	304433.461					
cycleway_mi	286342.143					
school	261012.258					
dist_cbd_mi	202917.092				6	negative coeff
resi_area_acres	197622.345					
distschool_mi	176196.604				9	negative coeff
primary_mi	173577.182				13	negative coeff
wtr_cntdis_mi	173250.156					
frst_cntdi_mi	168565.610					
water_area_acres	138530.856					
frest_area_acres	114514.499				2	negative coeff
ind_area_acres	96988.865					
cwy_lneall_mi	81508.732				11	negative coeff
intd_1mi_acres	76738.485					
cwy_trkall_mi	71003.080					
bus_stops	56392.700					
meanspeed	50491.636					
cwy_binary	28111.983					
college	19455.478				1	negative coeff
university	3217.012				4	negative coeff
bridge	2699.632					
cwy_trkbin	1466.005					

 # Non numeric variables were excluded
 # Pointspeed excluded, only meanspeed was included instead
 # Intersection density within half a mile was exluded, that within 1 mile was included
 # Number of jobs variable was excluded, and employment density was included instead

 # Water_dist or nearest distance to edge of water body from the count station has been excluded
 # Park_dist or nearest distance to edge of park or open space from the count station has been excluded
 # Forest_dist or nearest distance to edge of forest from the count station has been excluded
 # Grass_dist or nearest distance to edge of grass from the count station has been excluded
 # Comm_dist or nearest distance to edge of commercial area from the count station has been excluded
 # Ind_dist or nearest distance to edge of industrial area from the count station has been excluded
 # Resi_dist or nearest distance to edge of residential area from the count station has been excluded
 # Ret_dist or nearest distance to edge of retail area from the count station has been excluded