Data Mining to Improve Planning for Pedestrian and Bicyclist Safety

November 2019 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VT | TRANSPORTATION INSTITUTE
VIRGINIA TECH

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 01-003	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Data Mining to Improve Planning for Pedestrian and Bicyclist Safety		5. Report Date November 2019 6. Performing Organization Code:
7. Author(s) Arash Jahangiri [1] Mahdie Hasani [1] Ipek N. Sener [2] Sirajum Munira [2] Justin Owens [3] Bruce Appleyard [1] Sherry Ryan [1] Shawn M. Turner [2] Sahar Ghanipoor Machiani [1]		8. Performing Organization Report No. Report 01-003
9. Performing Organization Name and Address: Safe-D National UTC [1] San Diego State University [2] Texas A&M Transportation Institute [3] Virginia Tech Transportation Institute		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/Project 01-003
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT) State of Texas		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code
15. Supplementary Notes This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program, and, in part, with general revenue funds from the State of Texas.		
16. Abstract Between 2009 and 2016, the number of pedestrian and bicyclist fatalities saw a marked trend upward. Taken together, the overall percentage of pedestrian and bicycle crashes now accounts for 18% of total roadway fatalities, up from 13% only a decade ago. Technological advancements in transportation have created unique opportunities to explore and investigate new sources of data for the purpose of improving safety planning. This study investigated data from multiple sources, including automated pedestrian and bicycle counters, video cameras, crash databases, and GPS/mobile applications, to inform bicycle and pedestrian safety improvements. Data mining techniques, a new sampling strategy, and automated video processing methods were adopted to demonstrate a holistic approach that can be applied to identify facilities with highest need of improvement. To estimate pedestrian and bicyclist counts at intersections, exposure models were developed incorporating explanatory variables from a broad spectrum of data sources. Intersection-related crashes and estimated exposure were used to quantify risk, enabling identification of high-risk signalized intersections for walking and bicycling. The modeling framework and data sources used in this study will be beneficial in conducting future analyses for other facility types, such as roadway segments, and also at more aggregate levels, such as traffic analysis zones.		
17. Key Words high-risk signalized intersections, exposure modeling, direct demand models, pedestrian and bicyclist safety, data mining	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.	

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 53	22. Price $0

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

Abstract

Between 2009 and 2016, the number of pedestrian and bicyclist fatalities saw a marked trend upward. Taken together, the overall percentage of pedestrian and bicycle crashes now accounts for 18% of total roadway fatalities, up from 13% only a decade ago. Technological advancements in transportation have created unique opportunities to explore and investigate new sources of data for the purpose of improving safety planning. This study investigated data from multiple sources, including automated pedestrian and bicycle counters, video cameras, crash databases, and GPS/mobile applications, to inform bicycle and pedestrian safety improvements. Data mining techniques, a new sampling strategy, and automated video processing methods were adopted to demonstrate a holistic approach that can be applied to identify facilities with highest need of improvement. To estimate pedestrian and bicyclist counts at intersections, exposure models were developed incorporating explanatory variables from a broad spectrum of data sources. Intersection-related crashes and estimated exposure were used to quantify risk, enabling identification of high-risk signalized intersections for walking and bicycling. The modeling framework and data sources used in this study will be beneficial in conducting future analyses for other facility types, such as roadway segments, and also at more aggregate levels, such as traffic analysis zones.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program, and, in part, with general revenue funds from the State of Texas

We would like to extend gratitude to the project's subject matter expert, Dr. Krista Nordback, for her insightful review. We also much appreciate the work of our student Research Assistants in this project, namely Chenlei Zhang, James Davisson, and Christopher Galan at SDSU, and Kyuhyun Lee at TTI.

Table of Contents

List of Figures

List of Tables

Introduction

Statistics show that, over the last decade, more Americans have taken up walking and bicycling for commuting and recreation (*1*). While it is desirable to provide high levels of safety for these eco-friendly modes of travel, historical crash data do not appear to indicate that is happening. According to the fatality analysis reporting system (FARS; https://www-fars.nhtsa.dot.gov/Main/index.aspx) encyclopedia, between 2009 and 2016, the annual average pedestrian and bicyclist roadway fatalities numbers were around 4,800 and 720, respectively, while the annual average pedestrian and bicyclist injury numbers hovered around 65,000 and 50,000. Between 2009 and 2016, the number of pedestrian and bicyclist fatalities saw a marked trend upward. Taken together, pedestrian and bicycle crashes account for 18% of total fatalities in 2016, up from 13% in 2009. In 2016, there were over 6,000 pedestrian deaths and over 800 bicyclist deaths. In San Diego County, a similar trend is observed from FARS data: the overall percentage of pedestrian and bicycle fatalities increased from 27% in 2009 to 32% in 2016. This alarming trend requires researchers' and practitioners' urgent attention. While the figures presented above were extracted from the FARS database, it should be noted that other sources may report pedestrian and bicyclist fatality and injury data differently; for example, according to the Centers for Disease Control and Prevention (CDCP; https://www.cdc.gov/injury/wisqars/), 1,015 bicyclists and 7,330 pedestrians died in 2016 in the U.S.

Bicycle and pedestrian volumes, known as exposure data, are an essential part of safety assessments. However, most existing bicycle and pedestrian networks are not equipped to routinely collect count data in the manner it is typically collected for vehicular networks (e.g., via loop detectors). Because local agency staff lack bicycle and pedestrian exposure data, they are not able to accurately assess which facilities are in highest need of improvement. Specifically, identification of high crash risk locations is deemed important in prioritizing critical candidate locations for countermeasure implementation. Technological advancements and data mining algorithms are creating unique opportunities to investigate and analyze new sources of data to more accurately count pedestrian and bicycle activity and risk exposure, thereby improving safety modeling and planning. This study investigated data from multiple sources, including automated pedestrian and bicycle counters, video cameras, crash databases, and GPS/mobile applications (both active and passive monitoring), to inform bicycle and pedestrian safety improvements. The study goal is to integrate and use these data sources to produce useful insights in transportation safety planning. Utilizing these data, this study developed a process to identify high-risk intersections for walking and bicycling. The modeling framework in this study will be beneficial for conducting future analyses for other facility types, such as roadway segments, and also at more aggregate levels, such as in traffic analysis zones.

An important consideration in exposure and crash risk studies is the unit of analysis with regards to geographic scale. Some studies (*2*–*5*) focus on area-wide risk analysis, while others (*6*–*9*) focus on facility-specific analysis. The unit of analysis at the area-wide level includes traffic analysis

zones, census tracts, cities, counties, etc. On the other hand, roadway segments and intersections are usually considered the unit of analysis at the facility-specific level (*10*, *11*). The present study focuses on signalized intersections in the city of San Diego, which are by definition facility-specific. Several steps should be taken for a safety evaluation to identify high risk intersections, as outlined below.

1. *Identifying a sample of intersections for collecting short-term count data.* Ideally, pedestrian and bicyclist counts at all intersections are desirable for safety assessment. However, it is not feasible to collect data for all intersections due to budget, time, and equipment constraints. A sampling strategy is therefore required to obtain volume data at a subset of intersections with which a model is developed. Subsequently, this model is utilized to estimate the volumes at other intersections.
2. *Adopting a mechanism to collect short-term (e.g., 2 hrs, 12 hrs, days, or weeks) counts.* Most studies and agencies have used a manual approach (*2*, *12*, *13*) that includes having people collect counts in the field or having them extract counts by watching video data collected at the selected intersections. To facilitate the manual work, advanced video processing algorithms (*14–16*) can be applied to automatically count pedestrians and bicyclists. Different methods and technologies for collecting vulnerable road users are extensively discussed in (*17*).
3. *Applying methodologies to extrapolate short-term count data to yearly data*. The average annual daily bicyclist/pedestrian volume (AADB, AADP) are measures used in non-motorized transportation studies that can be estimated by averaging the daily bicyclist/pedestrian volume measured throughout the year. As it is not feasible to collect count data for a whole year at the selected sites, a common practice is to collect data for shorter periods of time (i.e., step 2) and apply an extrapolation method to convert short-term count data to yearly data in AADB and AADP forms. Continuous count data at several locations are also required to perform this estimation. Extrapolation has been used in several studies for estimating AADP and AADB volumes (*18–22*).
4. *Developing exposure models and quantifying crash risk.* Crash risk is generally estimated by dividing the number of unsafe events, such as crashes, by a total number of people who were likely to be involved in the unsafe events (i.e., exposure). Hence, crash risk is the probability of crash occurrence per unit of exposure. Focusing on pedestrians and bicyclists, exposure has been defined in different ways, including as "intersection pedestrian or bicycle counts," (*23*) "millions of cyclists or pedestrians," (*3*) and "10 million pedestrian crossings" (*4*). In most studies, the number of pedestrian/bicyclist crashes has been divided by a single exposure variable in order to calculate the crash risk. However, exposure can also be defined using multiple variables simultaneously. For example, the number of pedestrians and vehicles were used in (*24*); average daily pedestrians, average daily traffic, and distance crossed were used in (*25*); and hundred million

pedestrian/bicycle miles of roadway traveled were used in a study conducted in Washington, D.C. (*5*, *26*).

Literature Review

Exposure Modeling for Vulnerable Road Users

This section provides a brief review on exposure modeling, particularly direct demand models. For a comprehensive review of direct demand models, please see (*27*).

The direct-demand model relates walking and bicycling demand directly to various associated factors, such as sociodemographic and land use characteristics. Studies dating back as far as 50 years have used this modeling approach to forecast nonmotorized traffic, and it has been widely used in different areas of transportation. Recently, this approach has attracted attention because it benefits from the availability of a large volume of good-quality data and spatial database management software, such as geographic information systems, and results in comparatively simple tools that enable transport planners to predict nonmotorized traffic at relevant locations where count data are not available.

There have been various studies that used the direct-demand modeling approach to estimate pedestrian, bicycle, and trail traffic volume at different locations (see (*28–32*)). While the objectives of the studies are similar, researchers and transportation planners have used myriad approaches based on the magnitude of the available data and the characteristics of the study area. The generalized approach to develop a direct-demand model includes selection of a wide array of independent variables, often at various spatial scales, and choice of a suitable analysis method to estimate pedestrian, bicycle, or trail traffic in an area or location. Typically, the dependent variables of direct-demand models are pedestrian, bicycle, or trail traffic volumes for various time periods, such as during the peak period, hourly, daily, or annually. While some research has directly used data for the specific collection period, other studies have expanded short-period data to longer periods by using a scaling factor to be integrated and used in models. Studies have also explored a wide array of independent variables. Preference for explanatory variables often varies by location and time period, and might be categorized into nine groups: demographic, socioeconomic, network/interaction with vehicle traffic, pedestrian- or bicycle-specific infrastructure, transit facilities, major generators, weather and environmental, temporal or time related, and land use factors. Studies have highlighted that walking and bicycling trip behaviors differ substantially and need to be investigated separately. To identify the impact of land-use and built-environment characteristics on nonmotorized volume, a number of studies have considered a range of buffer widths. Investigating the influence of various independent variables by different buffer widths, studies have suggested that the best model may be obtained using different scales of buffer zones for different variables, as the variables are unlikely to be significant at the same buffer scale. A wide variety of approaches and methods (e.g., ordinary least squares, negative

binomial, and Poisson models) have been used in predicting nonmotorized activity using direct-demand models.

The major advantage of the direct-demand modeling approach is that it can be developed largely using existing data and common software packages. Because the model explains the impact of different factors that influence travel mode choice, it can provide important contributions to the decision-making process. However, direct demand modeling has some limitations, especially when transferred far into the future and for large areas, and researchers and practitioners need to be judicious in developing and applying these models.

Risk Quantification

This section summarizes the methods that have been used to quantify walking and bicycling crash risk at specific locations. Several studies have focused on pedestrian and bicyclist crash risk modeling to investigate the factors that increase or decrease that risk. However, the current study does not intend to examine the impacts of different variables, but rather focuses on determining a metric to quantify risks based on the literature. Crash risk is generally estimated by dividing the number of unsafe events (e.g., crashes and/or near crashes) by the total number of people likely to be involved in the unsafe events (i.e., exposure) as shown in Equation 1. Hence, the crash risk is the probability of crash occurrence per unit of exposure.

$$Estimated\ Crash\ Risk = \frac{Number\ of\ unsafe\ events}{Exposure} \qquad \text{Equation 1}$$

Focusing on pedestrians and bicyclists, exposure has been defined in different ways, such as pedestrian or bicyclist volume and estimated number of streets or travel lanes crossed (*33*). As the equation numerator and the denominator have been defined in different ways, the pedestrian and bicycle crash risk can also be calculated in several ways. Appendix A summarizes the methods that have been applied in a number of previous studies to calculate risk.

Vision-based Data Collection

Two main modules of object detection and object tracking comprise a vision-based monitoring system that can be employed to automatically count the number of pedestrians and bicyclists, or moving objects in general. Object detection is the process of identifying and classifying different objects (e.g., pedestrians, bicycles) in an image or in video frames. Once the objects are detected, an object tracking module traces different objects frame by frame to monitor their spatial and temporal characterization. Object detection and tracking through video cameras are considered critical topics in transportation safety monitoring, as they can significantly reduce manual work and enhance safety assessment. Objects are usually defined as any moving entity, such as pedestrians, bicyclists, passenger cars, trucks, etc., all interacting with one another at intersections.

Vision-based monitoring systems have several applications, such as speed evaluation (*34*, *35*), count estimation (*14*, *34*, *36*), waiting time assessment (*14*), lane change detection (*37*), violation evaluation (*35*, *38*), overtaking detection (*39*), and safety evaluations (*40*). For example, a

pedestrian-oriented system was presented in (*14*) to estimate pedestrian counts, waiting time, and crossing speed. The authors improved the contextual fusion system, which was previously proposed in (*41*). Their system added a local binary pattern to a Gaussian mixture model to detect pedestrians. A Gaussian mixture model (*42*) is an adaptive background subtraction method and is typically used to detect moving pedestrians in a scene by creating the adaptive background. A bicycle-oriented system was developed in (*34*) to count the number bicyclists and calculate their travel speed. Detection was accomplished by using a machine learning classifier, while the movement was recorded in sequence of frames. Researchers found that the addition of variables such as bicycle type, gender, and helmet use, which were manually collected, could be integrated to improve system capability.

Several object detection and tracking methods have been proposed that use video data from cameras installed on vehicles (i.e., moving cameras) and installed at infrastructure (i.e., stationary cameras). Methods for detecting and tracking pedestrians and bicyclists have been investigated through moving cameras mostly in autonomous vehicle systems. For example, a warning and control system was proposed, in which a support vector machine was employed to perform pedestrian detection (*43*). AdaBoost and cascaded classifiers with appearance and motion features were also used to identify pedestrians (*44*). An efficient method was proposed in (*45*), in which AdaBoost was combined with other algorithms to improve efficiency. Analyzing vulnerable road users has also been conducted using stationary cameras, as investigated in (*14*, *38*, *46*) for pedestrians, and in (*34*, *47*, *48*) for bicyclists. These studies were proposed mostly for monitoring road users at critical locations, such as intersections, to conduct safety analysis, speed evaluation, and count estimation. For example, a system which can count the number and speed of pedestrians at intersections was proposed in (*14*), and a system was developed for counting the number and speed of bicyclists in (*34*). In addition, researchers have developed open source software for tracking objects; a combination of blob and feature based tracking was used to develop Urban Tracker for tracking objects in urban mixed traffic (*49*), and a deep learning approach was adopted in the European Union's InDev project to develop STRUDL (Surveillance Tracking Using Deep Learning) (*50*).

Emerging Data Sources

While traditional pedestrian and bicyclist monitoring methods require active efforts from data collectors, advancements in technology have made it possible to use the proliferation of mobile phones to capture real-world walking and bicycling patterns. These emerging methods promise new opportunities, but much work remains to fully realize the potential of accessible data and practices. In an effort to explore available resources and approaches, the research team conducted a review of currently available crowdsourced data and how those data are used (see (*51*)).

Researchers also contacted several large companies about their ability and willingness to provide passively-collected GPS data for purposes of measuring pedestrian and bicyclist use/activity levels. However, the data were not available to be used within the current study's timeframe.

Data Description

The unit of analysis in the present study is signalized intersections in the City of San Diego. A total of 1,522 signalized intersections were identified using an ArcGIS shapefile. Short-term video data were collected by National Data and Surveying Services at a sample of 45 intersections. These intersections were equipped with cameras, and the videos were recorded for 12 hours (7:00 a.m. to 7:00 p.m., a period that is believed to best cover the morning and evening peaks based on local knowledge) on a Tuesday, Wednesday, or Thursday workday, which are the typical days for short term data collection (*52*, *53*) in May, June, or July 2018. Short-term data collection can be conducted for different lengths of time, from a few hours to a few weeks, and generally more data leads to smaller extrapolation errors. In this study, a data collection length of 12 hours was used for two reasons. First, data collection for long periods were not feasible due to budget constraints. Second, according to Nordback et al.'s study (*53*), it was found that extrapolation error rates do not decrease significantly from 12 to 24 hours. Following data collection, pedestrian and bicycle short-term counts were automatically obtained through machine-vision modeling. In addition to the short-term counters at selected intersections, 43 of San Diego's automated counters were also utilized. These counters are not located at intersections, but they have been continuously collecting pedestrian and bicyclists counts since 2012, which provides pedestrian and bicyclist activity patterns. Due to the vandalization of some equipment and issues with battery counters, data from 2015 were used, as this is the year believed to contain the most reliable data. For every intersection, demographic and socioeconomic characteristics and built environment variables were obtained by buffer analysis in ArcGIS. Specifically, this study gathered 396 explanatory variables, of which 129 were buffer variables (three different buffer sizes, including 0.1, 0.25, and 0.5 miles were used based on similar studies (*54–59*)) and 9 variables were point- or intersection-related variables. The data sources for these variables were as follows: 2015 five-year estimates from the U.S. Census Bureau, 2015 longitudinal employer-household dynamics data, the San Diego crime dataset, and San Diego's Regional Planning Agency. In addition, crash data involving pedestrians and bicyclists were obtained from the Statewide Integrated Traffic Records System through the Transportation Injury Mapping System. Crash data per victim for each intersection was extracted from 2006 to 2016.

Methodology

This section consists of several steps and methods utilized or developed to conduct safety evaluation with the goal of identifying high-risk intersections for walking and bicycling. The four major steps, as discussed below and also illustrated in Appendix B are: (1) site selection; (2) short-term data collection: vision-based monitoring system, (3) extrapolation, and (4) exposure modeling and risk quantification.

Site Selection

The first step in identifying high crash risk intersections is to determine a sample of intersections for short-term data collection. The site selection task applies methods to select intersections with a wide variety of characteristics to obtain a representative sample for the entire area of interest.

Two general groups of sampling methods for site selection include probabilistic and non-probabilistic sampling (*10*, *60*). Non-probabilistic sampling techniques are mostly based on non-random factors, such as engineering judgment. These techniques have low cost and are easy to implement. However, they may not lead to a representative sample. In contrast, probabilistic sampling techniques involve some random selection in the process and thus result in better generalization. As the population size grows, the simple random selection technique and other similar versions of this method may require larger observations to ensure a representative sample. Other probabilistic sampling techniques include stratified sampling, cluster analysis, and multi-stage random, which is basically a combination of clustering and stratification (*10*, *60*).

In stratified sampling, a few variables are used to create strata and each intersection is associated with a stratum. For example, population density, median income, and proximity to commercial properties, each with three levels of high, moderate, and low were used in (*61*, *62*) to create 27 strata (e.g., high population density, low median income, and moderate proximity to commercial properties make a single stratum). Stratified sampling is effective in that it ensures that the sample contains observations with different levels for the variables used. However, as the number of variables increase, the number of strata grows rapidly and thus selecting one intersection per stratum could make sample too large to use due to limited resources. Stratified sampling could restrict the number of variables used in the site selection process. Cluster analysis classifies all the intersections into different clusters with similar characteristics. More variables can be used in cluster analysis compared to stratified sampling without having to increase the sample size. Cluster analysis is suitable when the population is too large, but the sample resulting from a cluster analysis may not be as representative (*10*). In some studies, combinations of different methods with some variations have been used. For example, stratified sampling was used in (*12*) with six variables, including different number of categories per variables (e.g., day of week with seven categories). As another selection method, which is a non-probabilistic sampling technique, inputs from local stakeholders were also used in conjunction with other methods (*63*).

A multi-stage random approach was adopted in the present study to benefit from both stratification and cluster analysis. Since the population size in our case is fairly large (~ 1,500 signalized intersections), cluster analysis allowed us to use many variables without making the sample size too large. These variables were selected based on literature review as well as common sense. In addition, the stratified sampling reinforced our approach to make the sample as representative as possible. The data in this study contain both numerical and categorical variables and thus the Gower coefficient, proposed in (*64*), was used to handle both numerical and categorical variables when calculating distances or similarities between two observations. After determination of

pairwise distance among observations, a partitional clustering algorithm, partitioning around mediods (PAM) (*65*), was employed to identify the clusters. In the stratification step, depending on the number of clusters used, a stratified sampling method can be applied using one or more intersection characteristics to ensure that intersections with different levels for these characteristics are included in the sample. If the number of clusters turns out to be very high or very low, then the number of variables and/or levels of variables used for stratification can be adjusted to obtain the desired sample size. This flexibility was adopted due to resource limitations.

Short-term Data Collection: Vision-Based Monitoring System

A vision-based monitoring system was used to count the number of pedestrians and bicyclists crossing the intersections using short-term video data. The system consists of three steps: object detection, object tracking, and object counting, as illustrated in Figure 1.

Object detection was performed by utilizing Faster R-CNN (*15*) to detect pedestrians and bicyclists in video frames. Subsequently, detection results were used to perform object tracking using an Intersection-over-Union (IoU) (*66*) tracker, which associates each detection result with the highest IoU to the last detection result in the previous frame. The tracker will start a new trajectory and end the old trajectories if all detections are unassociated with any old trajectories. Finally, pedestrian and bicyclist counts were obtained from regions of interest, which were defined as areas typically used by pedestrians and bicyclists to cross intersections. To obtain the counts, any trajectory that entered the regions of interest was counted as a crossing pedestrian or bicyclist.

Figure 1. Vision-based monitoring system framework.

Extrapolation

To estimate AADP and AADB from short-term counts, similar pedestrian and bicyclist volume patterns for each short-term data collection site need to be identified. These volume patterns are utilized for extrapolating long-term counts from short-term counts. Permanent counters, even at a different location than the short-term counters, are typically used to identify similar demand patterns. Therefore, each short-term counter should be matched with one or more permanent counters. In this study, the matching process was performed in two steps as proposed by the authors in (*67*): in the first step, the PAM clustering method was applied to classify long-term counters into different clusters based on traffic distribution indexes, including the morning over the midday peak index (AMI), the weekend over weekday index (WWI) and the peak period to non-peak period index (PPI) for bicycle counters and AMI and WWI for pedestrian counters. (More details

about these indices can be found in (*68–70*).) In the second step, the classified long-term counters were used as the training data for developing a K-nearest neighbor (*65*) model to match short-term counters to appropriate clusters.

Several variables were calculated and utilized in the extrapolation process. These variables include population density, employment density, and land use density (commercial, residential, government, industrial, park and recreational) within a given 0.402-kilometer counter buffer (0.25-mile buffer), as well as traffic distribution indices such as AMI, WWI, and PPI, which are introduced in (*68*, *69*). These indices reflect bicyclist volume in morning peak hour over midday peak hour, weekend over weekday, and monthly variations, respectively.

Several extrapolation methods have been used in previous studies and a few—such as day-of-year (*70*), day-by-month (*68*), and weather model—have been shown to produce lower AADP and AADB estimation errors. The day-of-year method was not applicable in this study, as it required the short-term and long-term data to be collected in the same year; our short-term data collection was conducted in 2018 while long-term data were collected in 2015. In addition, the weather model was not deemed to be beneficial due to San Diego's year-round mild weather. Thus, the day-by-month method was applied with minor modification, as described below.

First, 12-hour counts were converted to 24-hour counts using Equation 2. Equation 3 shows how day-by-month factor was calculated for every day of week (d) and month (m). Subsequently, AADP and AADB counts were estimated by applying day-by-month adjustment factors to the 24-hour counts using Equation 4. The long-term counter data in the following equations refer to a counter that has been matched with the short-term counter of interest. However, it should be noted that if two or more permanent counters are matched to a short-term counter, the mean of adjustment factors across all matched counters was used. Please also note that the following equations are for estimating pedestrian counts (i.e., AADP), but they were also used for estimating bicyclist counts (i.e., AADB). To properly modify the equations for bicyclists, the letter *P* is changed to *B* in all equations, as noted in the variable descriptions following the equations.

$$P_{dms} = \sum_{h=7}^{18} P_{hdms} \times \frac{\sum_{h=1}^{24} P_{hdml}}{\sum_{h=7}^{18} P_{hdml}} \qquad \text{Equation 2}$$

$$PF_{dm} = \frac{AADP_l}{ADP_{dml}} \qquad \text{Equation 3}$$

$$AADP_s = P_{dms} \times PF_{dm} \qquad \text{Equation 4}$$

h: Hour of day. $h = \{1,2, \dots,24\}$

d: Day of week. $d = \{1 = Mon, 2 = Tue, \dots,7 = Sun\}$

m: Month of year. $d = \{1 = Jan, 2 = Feb, \dots,12 = Dec\}$

P_{dm}^{s} $(or\ B_{dm}^{s}\ for\ bicyclists)$: Pedestrian (bicyclists) count on day *d* of a week, in month *m* estimated for short-term counter *s*

P_{hdm}^{s} $(or\ B_{hdm}^{s}\ for\ bicyclists)$: Pedestrian (bicyclist) count in hour *h* of day *d* of a week, in month *m* from short-term counter *s*

P^l_{hdm} $(or\ B^l_{hdm}\ for\ bicyclists)$: Pedestrian (bicyclist) count in hour *h* of day *d* of a week, in month *m* from matched long-term counters

PF_{dm} $(or\ BF_{dm}\ for\ bicyclists)$: Pedestrian (bicyclists) day-by-month factor for day *d* of a week, in month *m*

$AADP^l$ $(or\ AADB^l\ for\ bicyclists)$: Average annual daily pedestrian (bicyclist) count obtained from matched long-term counters

ADP^l_{dm} $(or\ ADB^l_{dm}\ for\ bicyclists)$: Average daily pedestrian (bicyclist) volume on day *d*, in month *m* from matched long-term counters

$AADP^s$ $(or\ AADB^s\ for\ bicyclists)$: Average annual daily pedestrian (bicyclist) count estimated for short-term counter *s*

Exposure Modeling and Risk Quantification

A wide variety of approaches and methods have been used in predicting nonmotorized activity using direct-demand models. Given the nature of the dependent variable, exposure, which is discrete in nature and with a variance greater than the mean, the negative binomial model was selected as the exposure model for this study. The wide array of variables and the limited number of observations pose challenges for model development, which is often the case for both pedestrian and bicycle datasets. When many exploratory variables are available, identifying the best subset of those variables to include in a model is one of the hardest parts of model building (*71*) since evaluating models with a high number of variables is too computationally intensive. To develop a model with satisfactory performance, it is imperative to select a set of variables that can explain the variation of the pedestrian and bicycle annual average volumes for all locations.

As discussed earlier, many variables (a total of 396) were considered in the analysis. Univariate and bivariate correlation analyses were first conducted to explore variables' distribution or pattern and to investigate the relationship between the dependent and independent variables. Several variable forms and functions were examined to get the best data fit. The process also investigated interactions between several pairs of variables to see how the variables might interact with each other. A set of key variables were identified to be included in the final model-building process based on statistical tests, intuitive considerations, and insights from the previous literature.

After several model trials with different combinations of the key variables, the best models were evaluated based on their predictive accuracy regarding mean absolute error (MAE), and root mean squared error (RMSE). A cross-validation technique was employed for performance evaluation. Cross-validation is a resampling technique that helps identify a parameter value, ensuring a proper balance between bias and variance (*72*). For cross-validation, a subset of the data, known as the training set, is used to train the model, and the remaining data points serve as a test set or validation set. While fitting a model on a training set, it is desirable to have minimum MAE, so there is minimal difference between the prediction and the actual observation. This research used a 10-fold cross-validation method to evaluate and compare the performance of the developed models. This method split the feature vector sets into 10 approximately equally sized distinct partitions.

While one set was used for testing, the other nine sets were used for training. Then, the procedure was repeated 10 times (with each of the 10 sets used once for testing), and all accuracy rates over these 10 runs were averaged to improve the estimate. The performance evaluation criterion was the average accuracy. The final models were identified based on statistical, predictive and intuitive considerations as well as insights from the literature.

Utilizing the estimated pedestrian and bicycle exposure, risks associated with walking and bicycling at signalized intersections can be calculated. Considering Equation 1, several variables were examined in terms of their contribution in quantifying risk. Several studies have used the number of crashes in the risk equation as the number of unsafe events. As a result, two different locations with same number of crashes and exposure would lead to the same level of risk, though each crash may involve more than one victim. Taking number of victims into consideration instead, a crash with multiple victims should be associated with a higher risk compared to a crash with only one victim. In addition, crashes with higher levels of severity should be considered higher risk. Alternately, the number of fatalities could be used to provide an estimate of the relative lethality of intersections. Therefore, a combination of fatalities and injuries was utilized to provide a more holistic risk estimate. Crash severity was incorporated in risk quantification by utilizing crash costs associated with severity levels. Other factors, such as AADP (AADB), as the exposure and distance crossed, were also included in the risk equation, as presented in Equation 5. The unit of risk in this equation is crash cost per average annual daily pedestrian (or bicyclist) feet. In addition to the crash cost, the equation numerator includes a term N^k to produce more weight on the locations with a higher frequency of victims. Since crashes are rare events, it is important to magnify the number of occasions that led to fatalities and injuries. The tuning parameter k can also be used to provide the extent of the weight. For example, if zero is selected for this parameter, N^k becomes one, which means zero weight is given to the victim frequency. As k increases, it there is more weight on the victim frequency, resulting in higher risks.

Crash cost has been used for different purposes, such as analyzing the effectiveness of a specific roadway enhancement and measuring the effect of seatbelts based on injury severity in several studies (*73*–*77*). In 1993, Miller estimated motor-vehicle crash comprehensive costs by injury severity and body region (*73*). Another study estimated crash costs of medium and heavy trucks by seven injury severity levels (*74*). Miller et al. broke down pedestrian and bicyclist crash costs by age, injury severity, and body region in the U.S. (*75*). The Federal Highway Administration (*78*) also presented an estimation of crash cost based on maximum police-reported injury severity.

In crash cost studies, maximum abbreviated injury severity is defined as the maximum threat of a crash to a victim's life (*79*). Crash cost generally results from a combination of cost categories, including medical, emergency service, lost productivity, the monetized value of the pain and suffering, and lost quality of life costs. Collectively, these costs have been called comprehensive costs. Monetary or economic cost value of a crash can be obtained by subtracting lost quality of life from the comprehensive cost (*73*). In this study, lifetime costs were used; these included medical, work loss, and quality of life costs (*75*). It should be pointed out that vehicle volume may

also impact the quantified risk. However, this variable was not considered in this work as the exposure was defined as the multiplication of AADP (AADB) and distance crossed (i.e., equation denominator). Ideally, to capture the impact of vehicle volume, total number of interactions between pedestrians (bicyclists) and vehicles should be used as a measure of exposure. However, the historical data do not include number of interactions and it would also be very difficult to collect new data to measure number of interactions, which is outside the scope of this study.

$$Quantified\ Risk\ for\ an\ intersection = \frac{C \times N^k}{AADP(AADB) \times D}$$

$$C = \sum_s N_s \times C_s \qquad \text{Equation 5}$$

Where,

C: Total crash cost weighted by severity

N_s: Number of pedestrian or bicycle victims with injury severity level s

C_s: Cost per victim with injury severity level s

s: Severity level = {fatal, severe injury, other visible injury, and complaint of pain}

N: Total number of victims

k: Exponent of N, a tuning parameter to magnify the frequency of victims

$AADP(AADB)$: Average Annual daily pedestrian (bicyclist) count

D: Distance a pedestrian or bicyclist crossed

Results and Discussion

Site Selection

To select a representative sample of intersections, several intersection characteristics (i.e., variables) can be utilized in the site selection process, including those that influence pedestrian and bicyclist activity as well as their safety. The selection of variables was made based on previous studies and common sense (i.e., their potential impact on pedestrian and bicyclist exposure and safety). A total of 18 variables were examined and, after several trials with different subsets of variables, a subset of 12 variables were selected to perform site selection—population density, land use (parks and recreational, residential), presence of college, presence of school, transit stops density, mean traffic volume, pedestrian victims, bicyclist victims, proximity to Balboa Park, proximity to beaches, sidewalk density, and bikeway density. Using the selected variables, signalized intersections were grouped into clusters by applying the PAM clustering method. The silhouette metric and elbow method were applied to identify the best number of clusters. The silhouette metric provides a way to measure how similar a data point is to its assigned cluster compared to other clusters. This provides a way to evaluate clustering performance in relation to the number of clusters used. The highest silhouette value—that which shows the highest clustering performance—was obtained when using five clusters, as shown in Figure 2. The elbow method

plots the total within sum of squares against number of clusters used. A lower value of sum of squares mean data points within each cluster are very close to each other and thus contribute to a better clustering performance. The best number of clusters is typically identified at a point on the plot beyond which no significant decrease is observed, known as the elbow point. Based on the elbow plot in Figure 2, as the number of clusters increases, the total sum of squares decreases. However, no clear elbow point is visible, and thus the best number of clusters was selected to be five based on the silhouette method only. The geographic distribution of these five clusters is shown in Appendix C.

Within each cluster, stratified sampling was performed using two variables: the number of pedestrian victims and the number of bicyclist victims. The purpose was to ensure that the sample included intersections with high, moderate, and low numbers of victims. The number of pedestrian victims, ranging from 0 to 13, was divided into three levels (low: 0, 1, 2; moderate: 3, 4, 5; high: ≥ 6). Similarly, the number of bicyclists victims was divided into three levels (low: 0, 1; moderate: 2, 3, 4; high: ≥ 5). Consequently, nine strata for each cluster (3*3 = 9) resulted. Subsequently, a sample of 45 intersections was identified, as shown in Figure 2, by selecting one intersection per stratum (5*9 = 45). This selection was mainly random, but we opted not to select adjacent intersections or intersections close in proximity, as these intersections could be very similar to each other in some respects.

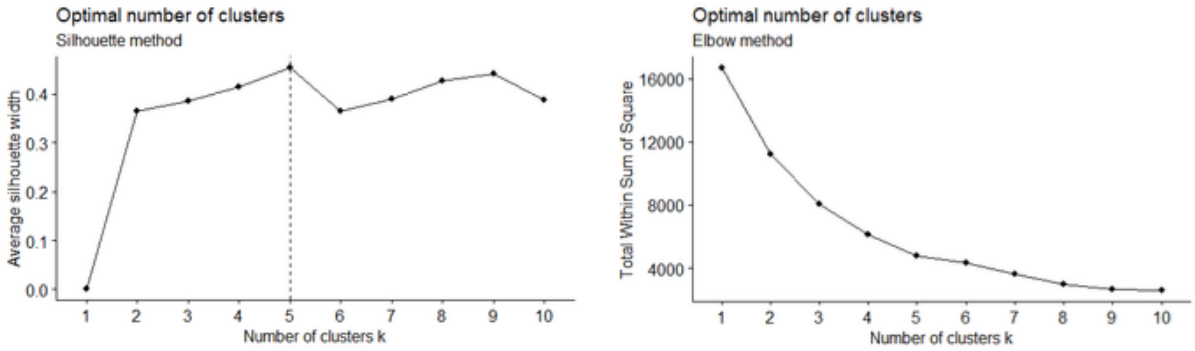


Figure 2. Identifying the best number of clusters: silhouette method on left and elbow method on right.

Vision-based Monitoring System

Video data were collected at the selected sample of intersections as described in the data description section. The video data collected from three intersections were utilized for training machine-vision models to automatically detect, track, and count pedestrians and bicyclists. These intersections were selected by manually reviewing and identifying high presence of both pedestrian and bicyclist activities. Several pedestrians and bicyclists were labeled to perform the training task. These models were tested in several scenarios and system performance was assessed using real-world video data at these intersections. Using test data, the average pedestrian and bicycle counting accuracies (correctly counted peds/bikes over total peds/bikes) were 85% and 81%, respectively. Several factors impacted the model performance, including the number of pedestrians and bicyclists labeled, intersection shape and size, lighting condition, occluded objects,

and video quality. The best counting accuracy of 95% was achieved for both pedestrians and bicyclists. It appears several different factors impact the counting accuracy, such as number of objects labeled, lighting conditions (i.e., different time of day, presence of shadows), object distance to the camera, and the way pedestrians cross (individually vs in groups) the intersection.

As expected, more labeled pedestrians and bicyclists led to better model performance, as the models were provided with more information in terms of positioning, angles, and lighting conditions. For example, the detection accuracy increased by 68% when the number of labeled pedestrians increased from 15 to 32. Model transferability was examined by using data from one intersection to train models and then testing them on a different intersection. The benefit was that manual labeling efforts were reduced, but the models performed poorly (counting accuracy ≤ 60%) since different intersections have different shapes and sizes. Focusing on one intersection for both training and testing, the way people cross the intersection (as individuals vs in groups) and lighting conditions due to time of day significantly affected model generalizability. For instance, the models had difficulty detecting and tracking pedestrians and bicyclists crossing the intersection in groups since some were occluded by others in multiple video frames. In addition, the object distances to the camera impacted the results. Cameras used in this study were set at a corner of each intersection. Detecting objects crossing the two farther intersection approaches from the camera was challenging, especially in large intersections where pedestrian and bicyclists were too small to distinguish, as can be inferred from Figure 2.

Extrapolation

Pedestrian and bicycle volume patterns at permanent counters were identified using the PAM clustering method. Appendix D illustrates pedestrian counters classified into three clusters (recreational, mixed, utilitarian) based on AMI and WWI. Similarly, bicyclist counters were grouped into four clusters—utilitarian, recreational, mixed recreational, mixed utilitarian—based on AMI, PPI, and WWI as shown in Appendix D. Pattern classification into three and four clusters has been used in past studies (*80–82*) and the only reason three clusters were used (instead of four) for pedestrian counters was the limited number of counters, which would have led to having only one counter in one of the clusters, which is not recommended (*53*, *83*).

Appendix D shows the clustering results for the four groups of long-term bicycle counters. Utilitarian counters have two distinct peak hours: mornings and evenings on weekdays. They also have a relatively uniform distribution throughout the week, as evident in their daily pattern. Recreational counters have higher weekend than weekday peaks, as expected. The daily patterns also represent the highest volume on Saturday and Sunday compared to other days of the week. Mixed, mixed recreational, or mixed utilitarian counters represent different combinations of utilitarian and recreational sites. These classifications for bicycle and pedestrian counters were consistent with the literature as identified in (*81*). Finally, each short-term counter was matched to one of the specified clusters by implementing a K-nearest neighbor model that identifies the volume patterns most similar to the identified long-term counters' clusters.

Exposure Modeling and Risk Quantification

After the estimation of AADP and AADB for the short-term sample, exposure modeling was applied to calculate AADB and AADP for the remainder of intersections (i.e., not in the short-term sample). The dependent variables for the pedestrian and bicycle exposure models are AADP and AADB, respectively. Tables 3–4 show the results of the negative binomial regression models of pedestrian and bicycle annual average daily volume. The tables present the explanatory variables and their estimates. The number following each variable represents the buffer area of influence.

Table 1 and Table 2. Negative Binomial Regression Model for Pedestrians

Variable	Estimates	t-stat	Sig.
(Intercept)	4.865	23.24	0.000
Transit stop density (0.805 kilometers ≈ 0.5 miles)	8.782	3.81	0.000
Percentage of regular transit rider, pedestrian, or bicyclist population (0.402 kilometers ≈ 0.25 miles)	3.717	1.95	0.051
Employment density (0.402 kilometers ≈ 0.25 miles)	0.051	2.41	0.016
Maximum speed limit within the intersection less than 64 kph (40 mph)	1.135	5.37	0.000
Percentage of vacant housing units (0.805 kilometers ≈ 0.5 miles)	−3.517	−2.99	0.003
Total commercial or mixed-use land area (0.160 kilometers ≈ 0.1 mile)	0.190	5.01	0.000
If the area contains a higher crime count than the average crime counts among the buffers (0.402 kilometers ≈ 0.25 miles)	−0.292	−1.65	0.098

Measure	Estimates
R-squared	0.70
RMSE	1633.24
MAE	1147.41

Table 3 and Table 4. Negative Binomial Regression Model for Bicyclists

Variable	Estimates	t-stat	Sig.
(Intercept)	4.265	12.56	0.000
Regular bicyclist population (0.402 kilometers ≈ 0.25 miles)	0.015	3.90	0.000
Transit stop density (0.160 kilometers ≈ 0.1 miles)	0.852	1.55	0.121
Maximum speed limit within the intersection less than 64 kph (40 mph)	0.457	2.90	0.004
Distance between the intersection and beachfront access point less than or equal to 16 kilometers (10 miles)	0.379	2.33	0.020
Presence of a school (0.805 kilometers ≈ 0.5 miles)	−0.483	−2.40	0.016
Bike facility density (0.805 kilometers ≈ 0.5 miles)	1.371	2.00	0.045
Total commercial or mixed-use land area (0.402 kilometers ≈ 0.25 miles)	0.021	2.18	0.030

Measure	Estimates
R-squared	0.67
RMSE	105.93
MAE	87.68

As shown in the first two tables, the pedestrian model has seven variables, and all of them were statistically significant at the 90% confidence level. As shown in the second two tables, the bicycle model also has seven variables, and all variables except one (transit stop density at 0.160 kilometers or 0.1 miles) were statistically significant at the 95% confidence level. The number of

observations in both models was 45. The R-squared values for the pedestrian and bicycle models were 0.7 and 0.67, respectively.

The negative binomial model results revealed that the variables were influential at multiple buffer areas and showed differences across pedestrian and bicycle activity. The results underscored the importance of location and community in characterizing nonmotorized demand, and targeted improvements to encourage nonmotorized activities. Appendix E provides a detailed discussion on the variables' effects for both bicycle and pedestrian models.

After estimating pedestrian and bicycle volumes (i.e., AADP and AADB) as the exposure measure, the risk was quantified by applying the proposed quantified risk equation (Equation 5). In this equation, number of victims and crash severity levels were obtained from the Statewide Integrated Traffic Records System data; distance pedestrians or bicyclists crossed at the intersection was calculated by multiplying the average number of lanes (across all approaches) by the lane width (3.65 m [12 ft] was assumed); cost per victim was obtained based on the victim's age and injury severity as estimated by Miller et al. (*75*). After experimenting with several values for the tuning parameter k, a value of three was chosen for the final model. It should be noted that there is no right or wrong value for this parameter and the researcher should determine the weight they wish to put on crash frequency and severity. Models with smaller values of k identified some intersections as high risk (i.e., top 50) with only one or two victims in the past 10 years. Although other factors such as a small AADP (AADB) and/or a high crash cost contributed to identification of high-risk intersections, it may not be practical to recognize an intersection with only one victim in the past 10 years as a high-risk intersection. This potential issue could apply to intersections with a small number of victims but with high severity levels (i.e., high cost). On the other hand, high values of k resulted in extreme values of risk for locations with higher number of victims compared to locations with lower number of victims. This significantly diminished the impact of crash severity in the numerator (i.e., crash cost) and therefore the risk. Consequently, risk calculated for an intersection with a high number of victims but with low severity levels could be significantly higher than that of an intersection with a medium to high number of victims with severe injuries. Accordingly, the value of three was selected as it provided a reasonable outcome. The risk for all signalized intersections was calculated to identify high-risk intersections for walking and bicycling, as mapped in Figure 3.

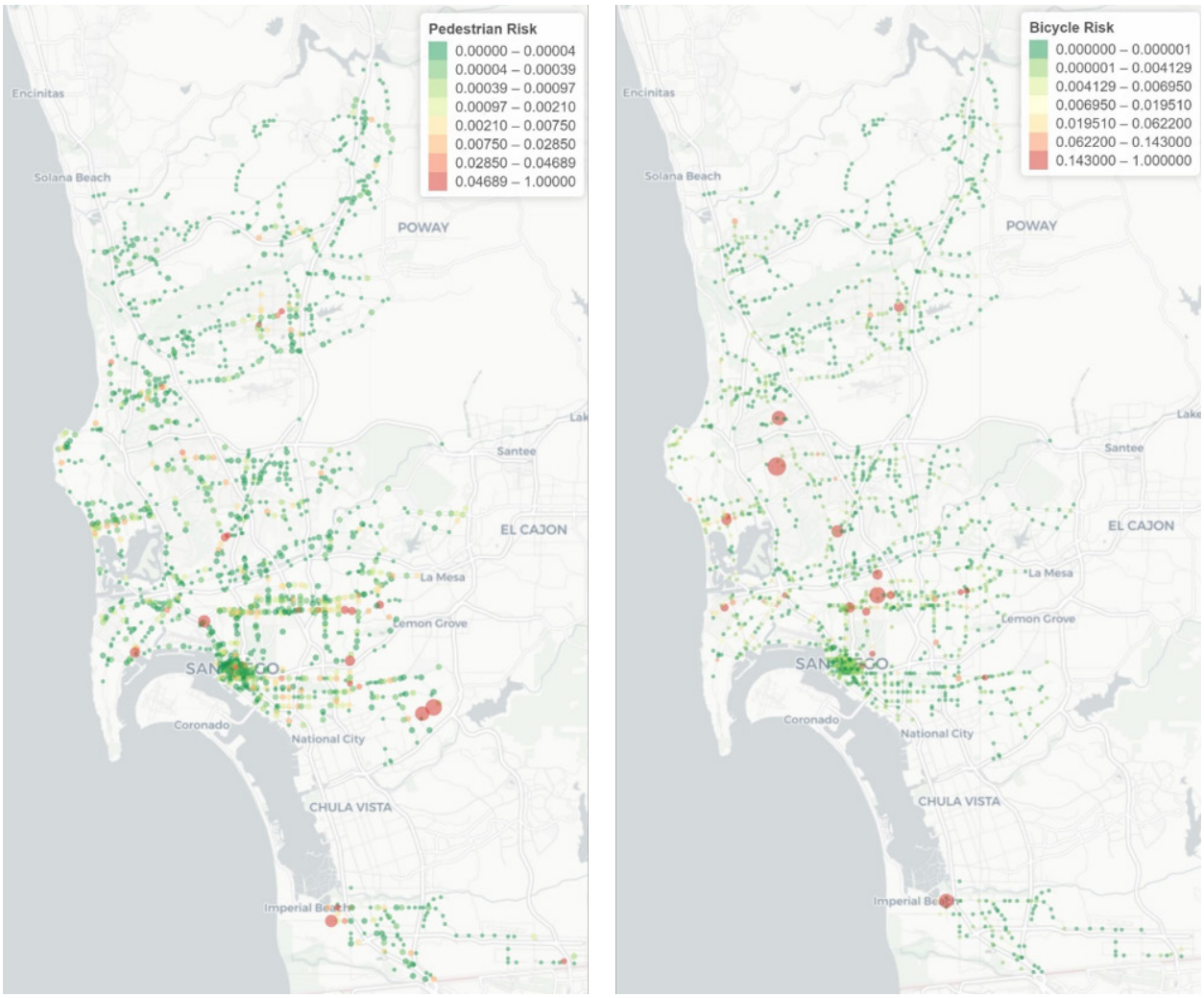


Figure 3. High-risk intersections for walking (left side) and bicycling (right side).

Previously, 15 intersections, known as the "fatal 15," were identified as the deadliest intersections for pedestrians in the city of San Diego (*84*). As expected, it was found that these intersections had a higher number of victims than other intersections. However, when exposure and other factors were taken into account using the quantified risk equation, not all intersections with the highest number of pedestrian and bicyclist victims were identified as high-risk. For example, out of 39 intersections with the highest number of pedestrian victims (victims ≥ 8), only 22 made it to the top 39 high-risk intersections based on quantified risk. Similarly, out of 36 intersections with the highest number of bicyclist victims (victims ≥ 5), only 26 remained in the top 36 high-risk intersections.

Conclusions and Recommendations

While statistics show an increasing trend in using eco-friendly modes of travel, such as walking and bicycling, historical crash data shows a growing trend in road crash victims involving pedestrians and bicyclists. Utilizing multiple data sources, such as automated pedestrian and bicycle counters and video cameras, this study estimated pedestrian and bicyclist exposure and identified signalized intersections with the highest risk for walking and bicycling within the city of San Diego, California.

A sampling strategy was used to identify a representative sample of intersections to collect short-term video data by applying cluster analysis and stratified sampling. A vision-based monitoring system was developed to automatically detect, track, and count pedestrians and bicyclists at the selected intersections. Situations where high-quality videos were used, a sufficient number of pedestrians and bicyclists were annotated, pedestrians and bicyclists were not too far from the camera, did not cross the intersection in groups, and good lighting was present, led to a high counting accuracy of 95%. Utilizing permanent counters, an extrapolation and a novel matching method were employed to estimate yearly counts that were used for estimating exposure by direct demand models. Exposure analysis identified transportation network, population, traffic generator, and land use variables as statistically significant in estimating pedestrian and bicyclist volume. Accounting for exposure as a normalization factor and considering other factors, such as frequency of victims and crash severity, in quantifying risk had a significant impact on the selection of high-risk intersections; not all intersections with the highest number of pedestrian and bicyclist victims were identified as high-risk. In addition, the variables were found to be influential at multiple buffer areas and showed differences across pedestrian and bicycle activity. The results underscored the importance of location and community in characterizing nonmotorized demand, and targeted improvements to encourage nonmotorized activities.

The modeling framework and data sources used in this study will be beneficial in conducting future analyses for other facility types, such as roadway segments, and also at more aggregate levels, such as traffic analysis zones. The approach is also beneficial to public agencies, as it can help quantify the risk of walking and bicycling at intersections, which in turn can aid in the development of procedures to identify high-risk facilities and prioritize them for countermeasure implementation. It should be pointed out that safety performance functions were not considered in this study, and thus potential future work might focus on how a combination of risk quantification and performance functions can assess safety. Since crashes are rare events, the identification of high-risk facilities will be lengthy, and a potential future direction is to proactively assess safety by discovering near-crash situations in video analysis. This would enable researchers and practitioners to quantify risk and evaluate safety in a much shorter period of time.

Additional Products

The Education and Workforce Development and Technology Transfer products created as part of this project are located on the project page of the Safe-D website, here. The final project dataset is located in the Safe-D Collection of the VTTI Dataverse.

Education and Workforce Development Products

The following resulted from project activities:

1. This research study contributed to a graduate master's thesis. SDUS's Mahdie Hasani's thesis is titled "Evaluation of Pedestrian and Bicycle Exposure and Crash Risk at Signalized Intersections in San Diego." Another master's student from SDSU, Christopher Galan, and a PhD student from TTI, Silvy Munira, were also involved in this research. While the project did not directly contribute to their thesis/dissertation, it increased their knowledge in the field.
2. The data collected and processed throughout the project were utilized in a group project assignment for *CIVE 160 Statistical Methods for the Built Environment* course at SDSU. The assignment tasks can be found in Appendix F.
3. A pedestrian and bicyclist vision-based counting tool was developed in MATLAB. The tool shows how machine-vision can be applied to automatically count the number of crossing pedestrians and bicyclists at intersections. More details can be found in Appendix G.
4. Demo video files of the counting tool were developed and used at the VT Science Festival on November 4, 2017 at VTTI headquarters in Blacksburg, VA.
5. The project team had a booth display and technology demonstration of Safe-D projects, including the current project, on March 17, 2018 for SDSU Explore Day event at SDSU, San Diego, CA. https://admissions.sdsu.edu/tours_events/explore
6. The project team convened a meeting with the City of San Diego and discussed project methodology and activities on September 15, 2017.

Technology Transfer Products

The following T2 products resulted from project activities:

1. One journal paper and seven presentations (peer reviewed conference papers and other presentations) were resulted from this project as listed on the project website.
2. A web-based tool (Utilized Shiny, an open source R package that provides an elegant and powerful web framework for building web applications using R.) was developed to visualize high-risk intersections for walking and bicycling. This interactive tool can be accessed here: https://g-hasani.shinyapps.io/RiskCalc/

Data Products

- Link to Dataset – https://doi.org/10.15787/VTT1/IUTNDS
- Project Description – The study goal was to identify high-risk signalized intersections for walking and bicycling in the City of San Diego using data mining methods. The data used in this study was collected from multiple sources, such as San Diego's automated pedestrian and bicyclist counting system in 2015, several video cameras through National Data and Surveying Services in June and July 2018, crash data from SWITERS from 2006 to 2016, and GIS shapefiles.
- Data Scope – The data from multiple sources, as mentioned above, were compiled to create a data table in CSV format. Total number of observations for this table is 1,520, with a total of 502 variables (i.e., columns).
- Data Specification – a detailed description of each variable in data set can be found in Appendix H.
- Citation Metadata:
    - Title of data set: "SafeD-01-003-Data.csv"
    - Author list with researcher ORCIDs
        - Mahdie Hasani, 0000-0003-3787-0547
        - Arash Jahangiri, 0000-0002-8825-961X
        - Christopher Johnathan Galan, 0000-0001-9437-7754
        - Ipek Nese Sener, 0000-0001-5493-8756
        - Sirajum Munira, 0000-0002-4953-2628
    - Contact information (email) for corresponding author: AJahangiri@sdsu.edu
    - Keywords: high-risk signalized intersections, exposure modeling, direct demand models, pedestrian and bicyclist safety planning, data mining

References

1. US Department of Transportation. Beyond Traffic 2045. *US Department of Transportation*. https://www.transportation.gov/sites/dot.gov/files/docs/BeyondTraffic_tagged_508_final.pdf.
2. Cameron, M. H. A Method of Measuring Exposure to Pedestrian Accident Risk. *Accident Analysis & Prevention*, Vol. 14, No. 5, 1982, pp. 397–405.
3. Strauss, J., L. F. Miranda-Moreno, and P. Morency. Multimodal Injury Risk Analysis of Road Users at Signalized and Non-Signalized Intersections. *Accident Analysis & Prevention*, Vol. 71, 2014, pp. 201–209.
4. Schneider, R., O. Grembek, and M. Braughton. Pedestrian Crash Risk on Boundary Roadways: University Campus Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2393, 2013, pp. 164–173.
5. Molino, J., J. Kennedy, P. Johnson, P. Beuse, A. Emo, and A. Do. Pedestrian and Bicyclist Exposure to Risk: Methodology for Estimation in an Urban Environment. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2140, 2009, pp. 145–156. https://doi.org/10.3141/2140-16.
6. Blaizot, S., F. Papon, M. M. Haddak, and E. Amoros. Injury Incidence Rates of Cyclists Compared to Pedestrians, Car Occupants and Powered Two-Wheeler Riders, Using a Medical Registry and Mobility Data, Rhône County, France. *Accident Analysis & Prevention*, Vol. 58, 2013, pp. 35–45.
7. Jonah, B. A., and G. R. Engel. Measuring the Relative Risk of Pedestrian Accidents. *Accident Analysis & Prevention*, Vol. 15, No. 3, 1983, pp. 193–206.
8. Beck, L. F., A. M. Dellinger, and M. E. O'neil. Motor Vehicle Crash Injury Rates by Mode of Travel, United States: Using Exposure-Based Methods to Quantify Differences. *American Journal of Epidemiology*, Vol. 166, No. 2, 2007, pp. 212–218.
9. Keall, M. D. Pedestrian Exposure to Risk of Road Accident in New Zealand. *Accident Analysis & Prevention*, Vol. 27, No. 5, 1995, pp. 729–740.
10. Greene-Roesel, R., M. C. Diogenes, and D. R. Ragland. *Estimating Pedestrian Accident Exposure*. Publication UCB-ITS-PRR-2010-32. Calofornia Path Program, 2010.
11. Turner, S., I. N. Sener, M. E. Martin, S. Das, R. C. Hampshire, K. Fitzpatrick, L. J. Molnar, M. Colety, S. Robinson, E. Shipp, and R. K. Wijesundera. Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities. No. FHWA-SA-17-041, 2017.
12. Molino, J., J. Kennedy, P. Johnson, P. Beuse, A. Emo, and A. Do. Pedestrian and Bicyclist Exposure to Risk: Methodology for Estimation in an Urban Environment. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2140, 2009, pp. 145–156.
13. Qin, X., and J. Ivan. Estimating Pedestrian Exposure Prediction Model in Rural Areas. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1773, 2001, pp. 89–96.
14. Shirazi, M. S., and B. T. Morris. Vision-Based Turning Movement Monitoring:Count, Speed Waiting Time Estimation. *IEEE Intelligent Transportation Systems Magazine*, Vol. 8, No. 1, 2016, pp. 23–34. https://doi.org/10.1109/MITS.2015.2477474.
15. Ren, S., K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, Vol. PP, No. 99, 2016, pp. 1–1. https://doi.org/10.1109/TPAMI.2016.2577031.
16. Hediyeh, H., T. Sayed, M. H. Zaki, and K. Ismail. Automated Analysis of Pedestrian Crossing Speed Behavior at Scramble-Phase Signalized Intersections Using Computer Vision Techniques. *International Journal of Sustainable Transportation*, Vol. 8, No. 5, 2014, pp. 382–397. https://doi.org/10.1080/15568318.2012.708098.
17. Ryus, P., E. Ferguson, K. M. Laustsen, F. R. Proulx, R. J. Schneider, T. Hull, and L. Miranda-Moreno. *NCHRP Web Document 205: Methods and Technologies for Pedestrian and Bicycle Volume Data Collection*. National Cooperative Highway Research Program, Washington, DC, 2015.
18. Roll, J. F. Bicycle Traffic Count Factoring: An Examination of National, State and Locally Derived Daily Extrapolation Factors. 2013.
19. Hankey, S., G. Lindsey, and J. Marshall. Day-of-Year Scaling Factors and Design Considerations for Nonmotorized Traffic Monitoring Programs. *Transportation Research Record*, Vol. 2468, No. 1, 2014, pp. 64–73.
20. Hankey, S., G. Lindsey, X. Wang, J. Borah, K. Hoff, B. Utecht, and Z. Xu. Estimating Use of Non-Motorized Infrastructure: Models of Bicycle and Pedestrian Traffic in Minneapolis, MN. *Landscape and Urban Planning*, Vol. 107, No. 3, 2012, pp. 307–316.
21. Schneider, R., L. Arnold, and D. Ragland. Methodology for Counting Pedestrians at Intersections: Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2140, 2009, pp. 1–12.
22. Nordback, K., W. E. Marshall, B. N. Janson, and E. Stolz. Estimating Annual Average Daily Bicyclists: Error and Accuracy. *Transportation research record*, Vol. 2339, No. 1, 2013, pp. 90–97.
23. Liggett, R., J. Cooper, H. Huff, R. Taylor-Gratzer, N. Wong, D. Benitez, T. Douglas, J. Howe, J. Griswold, D. Amos, and F. Proulx. *Bicycle Crash Risk: How Does It Vary, and Why?* 2016.
24. Cameron, M. H. A Method of Measuring Exposure to Pedestrian Accident Risk. *Accident Analysis & Prevention*, Vol. 14, No. 5, 1982, pp. 397–405. https://doi.org/10.1016/0001-4575(82)90019-7.
25. Radwan, E., H. Abou-Senna, A. Mohamed, A. Navarro, N. Minaei, J. Wu, and L. Gonzalez. *Assessment of Sidewalk/Bicycle-Lane Gaps with Safety and Developing Statewide Pedestrian Crash Rates*. Florida Department of Transportation Research Center, 2016.
26. Molino, J., J. F. Kennedy, P. Inge, M. A. Bertola, P. A. Beuse, N. L. Fowler, A. K. Emo, and A. Do. *A Distance-Based Method to Estimate Annual Pedestrian and Bicyclist Exposure in an Urban Environment*. United States. Federal Highway Administration. Office of Safety Research and Development, 2012.
27. Munira, S., and I. N. Sener. *Use of Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-Analysis*. Safety through Disruption (Safe-D) National University Transportation Center (UTC) Program, 2017.
28. Tabeshian, M., and L. Kattan. Modeling Nonmotorized Travel Demand at Intersections in Calgary, Canada: Use of Traffic Counts and Geographic Information System Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2430, 2014, pp. 38–46. https://doi.org/10.3141/2430-05.

29. Strauss, J., L. F. Miranda-Moreno, and P. Morency. Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modelling Approach. *Accident Analysis & Prevention*, Vol. 59, 2013, pp. 9–17. https://doi.org/10.1016/j.aap.2013.04.037.
30. Fagnant, D. J., and K. Kockelman. A Direct-Demand Model for Bicycle Counts: The Impacts of Level of Service and Other Factors. *Environment and Planning B: Planning and Design*, Vol. 43, No. 1, 2016, pp. 93–107. https://doi.org/10.1177/0265813515602568.
31. Schneider, R., T. Henry, M. Mitman, L. Stonehill, and J. Koehler. Development and Application of a Pedestrian Volume Model in San Francisco, California. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2299, 2012, pp. 65–78. https://doi.org/10.3141/2299-08.
32. Hankey, S., and G. Lindsey. Facility-Demand Models of Peak Period Pedestrian and Bicycle Traffic. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2586, 2016, pp. 48–58. https://doi.org/10.3141/2586-06.
33. Turner, S., I. Sener, M. Martin, S. Das, E. Shipp, R. Hampshire, K. Fitzpatrick, L. Molnar, R. Wijesundera, M. Colety, and S. Robinson. *Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities*. Publication FHWA-SA-17-041. Texas A&M Transportation Institute, 2017.
34. Ponte, G., Z. L. Szpak, J. E. Woolley, and D. J. Searson. Using Specialised Cyclist Detection Software to Count Cyclists and Determine Cyclist Travel Speed from Video. Presented at the Australasian Road Safety Research Policing Education Conference, 2014, Melbourne, Victoria, Australia, 2014.
35. Guo, Y., T. Sayed, and M. H. Zaki. Automated Analysis of Pedestrian Walking Behaviour at a Signalised Intersection in China. *IET Intelligent Transport Systems*, Vol. 11, No. 1, 2017, pp. 28–36. https://doi.org/10.1049/iet-its.2016.0090.
36. Tang, H. Development of a Multiple-Camera Tracking System for Accurate Traffic Performance Measurements at Intersections. http://www.its.umn.edu/Publications/ResearchReports/reportdetail.html?id=2254. Accessed May 1, 2017.
37. Kasper, D., G. Weidl, T. Dang, G. Breuel, A. Tamke, A. Wedel, and W. Rosenstiel. Object-Oriented Bayesian Networks for Detection of Lane Change Maneuvers. *IEEE Intelligent Transportation Systems Magazine*, Vol. 4, No. 3, 2012, pp. 19–31. https://doi.org/10.1109/MITS.2012.2203229.
38. Zaki, M., T. Sayed, K. Ismail, and F. Alrukaibi. Use of Computer Vision to Identify Pedestrians' Nonconforming Behavior at Urban Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2279, 2012, pp. 54–64. https://doi.org/10.3141/2279-07.
39. Garcia, F., P. Cerri, A. Broggi, A. de la Escalera, and J. M. Armingol. Data Fusion for Overtaking Vehicle Detection Based on Radar and Optical Flow. Presented at the 2012 IEEE Intelligent Vehicles Symposium, 2012.
40. Fu, T., L. Miranda-Moreno, and N. Saunier. *A Novel Framework to Evaluate Pedestrian Safety at Non-Signalized Locations Using Video-Based Trajectory Data*. 2017.
41. Shirazi, M. S., and B. Morris. Contextual Combination of Appearance and Motion for Intersection Videos with Vehicles and Pedestrians. Presented at the International Symposium on Visual Computing, 2014.

42. Zivkovic, Z., and F. van der Heijden. Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction. *Pattern Recognition Letters*, Vol. 27, No. 7, 2006, pp. 773–780. https://doi.org/10.1016/j.patrec.2005.11.005.
43. Cheng, H., N. Zheng, and J. Qin. Pedestrian Detection Using Sparse Gabor Filter and Support Vector Machine. Presented at the IEEE Proceedings. Intelligent Vehicles Symposium, 2005., 2005.
44. Viola, P., M. J. Jones, and D. Snow. Detecting Pedestrians Using Patterns of Motion and Appearance. Presented at the Proceedings Ninth IEEE International Conference on Computer Vision, 2003.
45. Abramson, Y., and B. Steux. Hardware-Friendly Pedestrian Detection and Impact Prediction. Presented at the IEEE Intelligent Vehicles Symposium, 2004, 2004.
46. Wu, B., and R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, No. 1, 2005, pp. 90- 97 Vol. 1.
47. Malinovskiy, Y., J. Zheng, and Y. Wang. Model-Free Video Detection and Tracking of Pedestrians and Bicyclists. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, No. 3, 2009, pp. 157–168. https://doi.org/10.1111/j.1467-8667.2008.00578.x.
48. Zaki, M. H., and T. Sayed. A Framework for Automated Road-Users Classification Using Movement Trajectories. *Transportation Research Part C: Emerging Technologies*, Vol. 33, 2013, pp. 50–73. https://doi.org/10.1016/j.trc.2013.04.007.
49. Jodoin, J.-P., G.-A. Bilodeau, and N. Saunier. Urban Tracker: Multiple Object Tracking in Urban Mixed Traffic. 2014.
50. Polders, E., and T. Brijs. *How to Analyse Accident Causation? A Handbook with Focus on Vulnarable Road Users*. Hasselt University, 2018.
51. Lee, K., and I. N. Sener. *Emerging Data Mining for Pedestrian and Bicyclist Monitoring: A Literature Review Report*. Safety through Disruption (Safe-D) National University Transportation Center (UTC) Program, 2017.
52. Schneider, R., L. Arnold, and D. Ragland. Methodology for Counting Pedestrians at Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2140, 2009a, pp. 1–12. https://doi.org/10.3141/2140-01.
53. Nordback, K., W. Marshall, B. Janson, and E. Stolz. Estimating Annual Average Daily Bicyclists. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2339, 2013, pp. 90–97. https://doi.org/10.3141/2339-10.
54. Schneider, R. J., L. S. Arnold, and D. R. Ragland. Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. *Transportation research record*, Vol. 2140, No. 1, 2009, pp. 13–26.
55. Griswold, J. B., A. Medury, and R. J. Schneider. Pilot Models for Estimating Bicycle Intersection Volumes. *Transportation research record*, Vol. 2247, No. 1, 2011, pp. 1–7.
56. Liu, X., and J. Griswold. Pedestrian Volume Modeling: A Case Study of San Francisco. *Yearbook of the Association of Pacific Coast Geographers*, 2009, pp. 164–181.
57. Miranda-Moreno, L. F., and D. Fernandes. Modeling of Pedestrian Activity at Signalized Intersections: Land Use, Urban Form, Weather, and Spatiotemporal Patterns. *Transportation research record*, Vol. 2264, No. 1, 2011, pp. 74–82.
58. Tabeshian, M., and L. Kattan. Modeling Nonmotorized Travel Demand at Intersections in Calgary, Canada: Use of Traffic Counts and Geographic Information System Data. *Transportation Research Record*, Vol. 2430, No. 1, 2014, pp. 38–46.

59. Radwan, E., H. Abou-Senna, A. Mohamed, A. Navarro, J. Wu, N. S. Minaei, and L. Gonzalez. Assessment of Sidewalk/Bicycle-Lane Gaps with Safety and Developing Statewide Pedestrian Crash Rates. 2016.
60. Aggarwal, Y. P. *Better Sampling: Concepts, Techniques, and Evaluation*. Stosius Inc/Advent Books Division, 1988.
61. Schneider, R., L. Arnold, and D. Ragland. Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2140, 2009, pp. 13–26.
62. Griswold, J., A. Medury, and R. Schneider. Pilot Models for Estimating Bicycle Intersection Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2247, 2011, pp. 1–7. https://doi.org/10.3141/2247-01.
63. Jones, M. G., S. Ryan, J. Donlon, L. Ledbetter, D. R. Ragland, and L. S. Arnold. Seamless Travel: Measuring Bicycle and Pedestrian Activity in San Diego County and Its Relationship to Land Use, Transportation, Safety, and Facility Type. *PATH Research Report*, 2010.
64. Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 1971, pp. 857–871.
65. Kaufman, L., and P. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.
66. Bochinski, E., V. Eiselein, and T. Sikora. High-Speed Tracking-by-Detection Without Using Image Information [Challange Winner IWOT4S]. 2017.
67. Hasani, M., A. Jahangiri, and S. G. Machiani. Developing Models for Matching of Short-Term and Long-Term Data Collection Sites to Improve the Estimation of Average Annual Daily Bicyclists. Presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018.
68. Beitel, D., and L. F. Miranda-Moreno. Methods for Improving and Automating the Estimation of Average Annual Daily Bicyclists. Presented at the Transportation Research Board 95th Annual Meeting Transportation Research Board, 2016.
69. Nosal, T. *Improving the Accuracy of Bicycle AADT Estimation: Temporal Patterns, Weather and Bicycle AADT Estimation Methods*. Master's Thesis. McGill University, 2014.
70. Hankey, S., G. Lindsey, and J. Marshall. Day-of-Year Scaling Factors and Design Considerations for Nonmotorized Traffic Monitoring Programs. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2468, 2014, pp. 64–73. https://doi.org/10.3141/2468-08.
71. Ratner, B. Variable Selection Methods in Regression: Ignorable Problem, Outing Notable Solution. *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 18, No. 1, 2010, pp. 65–75.
72. Chan-Lau, M. J. A. *Lasso Regressions and Forecasting Models in Applied Stress Testing*. International Monetary Fund, 2017.
73. Miller, T. R. Costs and Functional Consequences of U.S. Roadway Crashes. *Accident Analysis & Prevention*, Vol. 25, No. 5, 1993, pp. 593–607. https://doi.org/10.1016/0001-4575(93)90011-K.
74. Zaloshnja, E., and T. Miller. *Unit Costs of Medium and Heavy Truck Crashes*. Publication FMCSA-RRA-07-034. U.S. Department of Transportation Federal Motor Carrier Safety Administration, 2007.

75. Miller, T. R., E. Zaloshnja, B. A. Lawrence, J. Crandall, J. Ivarsson, and A. E. Finkelstein. Pedestrian and Pedalcyclist Injury Costs in the United States by Age and Injury Severity. *Annual Proceedings / Association for the Advancement of Automotive Medicine*, Vol. 48, 2004, pp. 265–284.
76. *Maximum Police-Reported Injury Severity Within Selected Crash Geometries, Chapter 4 - FHWA-HRT-05-051*. 2005.
77. Zaloshnja, E., T. Miller, F. Council, and B. Persaud. Comprehensive and Human Capital Crash Costs by Maximum Police-Reported Injury Severity Within Selected Crash Types. *Annual Proceedings / Association for the Advancement of Automotive Medicine*, Vol. 48, 2004, pp. 251–263.
78. Council, F. M., E. Zaloshnja, T. Miller, and B. N. Persaud. *Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometrics*. Turner-Fairbank Highway Research Center, 2005.
79. Association for the Advancement of Automotive Medicine. Abbreviated Injury Scale (AIS). *AAAM.org*. https://www.aaam.org/abbreviated-injury-scale-ais/.
80. Lindsey, G., S. Hankey, X. Wang, and J. Chen. *The Minnesota Bicycle and Pedestrian Counting Initiative: Methodologies for Non-Motorized Traffic Monitoring*. Publication MnDOT 2013-24. Minnesota Department of Transportation, 2013.
81. Miranda-Moreno, L., T. Nosal, R. Schneider, and F. Proulx. Classification of Bicycle Traffic Patterns in Five North American Cities. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2339, 2013, pp. 68–79. https://doi.org/10.3141/2339-08.
82. Lindsey, G., K. Nordback, and M. Figliozzi. Institutionalizing Bicycle and Pedestrian Monitoring Programs in Three States. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2443, 2014, pp. 134–142. https://doi.org/10.3141/2443-15.
83. Johnstone, D., K. Nordback, and M. Lowry. *Collecting Network-Wide Bicycle and Pedestrian Data: A Guidebook for When and Where to Count*. 2017.
84. The Fatal Fifteen Intersections - Vision Zero. *Circulate San Diego*. http://www.circulatesd.org/fatal15sd. Accessed Dec. 3, 2018.
85. Wachtel, A., and D. Lewiston. Risk Factors for Bicycle-Motor Vehicle Collisions at Intersections. *ITE Journal(Institute of Transportation Engineers)*, Vol. 64, No. 9, 1994, pp. 30–35.
86. Schneider, R., T. Henry, M. Mitman, L. Stonehill, and J. Koehler. Development and Application of a Pedestrian Volume Model in San Francisco, California. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2299, 2012, pp. 65–78.
87. Radwan, E., H. Abou-Senna, A. Mohamed, A. Navarro, N. Minaei, J. Wu, and L. Gonzalez. Assessment of Sidewalk/Bicycle-Lane Gaps with Safety and Developing Statewide Pedestrian Crash Rates. 2016.
88. Chu, X. The Fatality Risk of Walking in America: A Time-Based Comparative Approach. 2003.
89. Schneider, R., L. Arnold, and D. Ragland. Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2140, 2009b, pp. 13–26. https://doi.org/10.3141/2140-02.

90. Liu, X., and J. Griswold. Pedestrian Volume Modeling: A Case Study of San Francisco. *Yearbook of the Association of Pacific Coast Geographers*, Vol. 71, 2009, pp. 164–181.
91. Miranda-Moreno, L., and D. Fernandes. Modeling of Pedestrian Activity at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2264, 2011, pp. 74–82. https://doi.org/10.3141/2264-09.
92. Griswold, J. B., A. Medury, and R. J. Schneider. Pilot Models for Estimating Bicycle Intersection Volumes. *Transportation Research Record*, Vol. 2247, No. 1, 2011, pp. 1–7. https://doi.org/10.3141/2247-01.
93. Pulugurtha, S., and S. Repaka. Assessment of Models to Measure Pedestrian Activity at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2073, 2008, pp. 39–48. https://doi.org/10.3141/2073-05.
94. City of San Diego. *Vision Zero Saves Lives. City of San Diego Comprehensive Pedestrian Collision Analysis, Transportation and Stormwater Division.* 2014.
95. City of San Diego. *Zero Traffic Related Fatalities and Severe Injuries by 2025*. 2018.
96. Jones, M., S. Ryan, J. Donlon, L. Ledbetter, D. R. Ragland, and L. S. Arnold. Seamless Travel: Measuring Bicycle and Pedestrian Activity in San Diego County and Its Relationship to Land Use, Transportation, Safety, and Facility Type. *PATH Research Report*, 2010.
97. Bowen, A. New Bike Sharing Program Launching In San Diego. *KPBS Public Media*. http://www.kpbs.org/news/2018/feb/15/new-bike-sharing-launches-san-diego-limebike/. Accessed Dec. 20, 2018.
98. Strauss, J., and L. F. Miranda-Moreno. Spatial Modeling of Bicycle Activity at Signalized Intersections. *Journal of Transport and Land Use*, Vol. 6, No. 2, 2013, pp. 47–58.

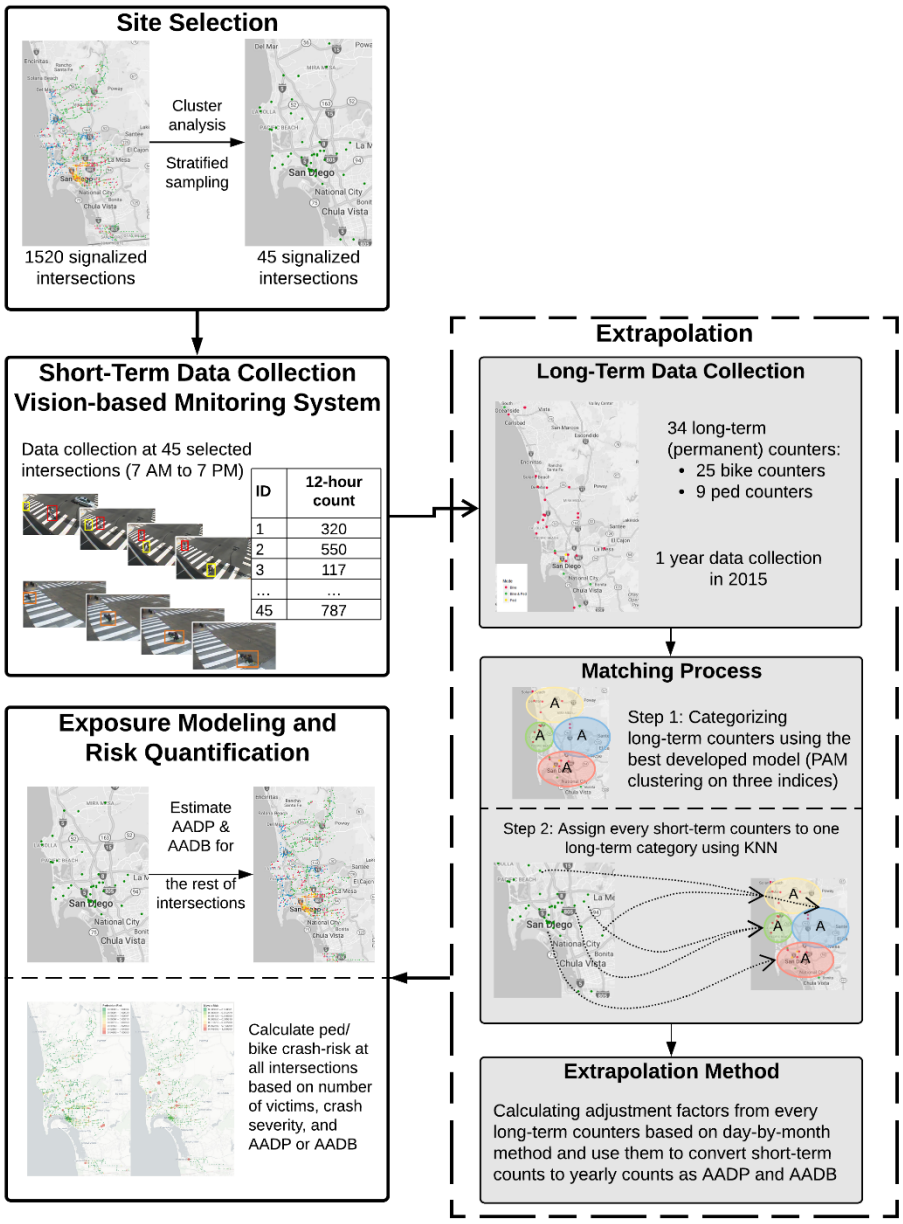
Appendices

Appendix A - Risk Quantification Methods

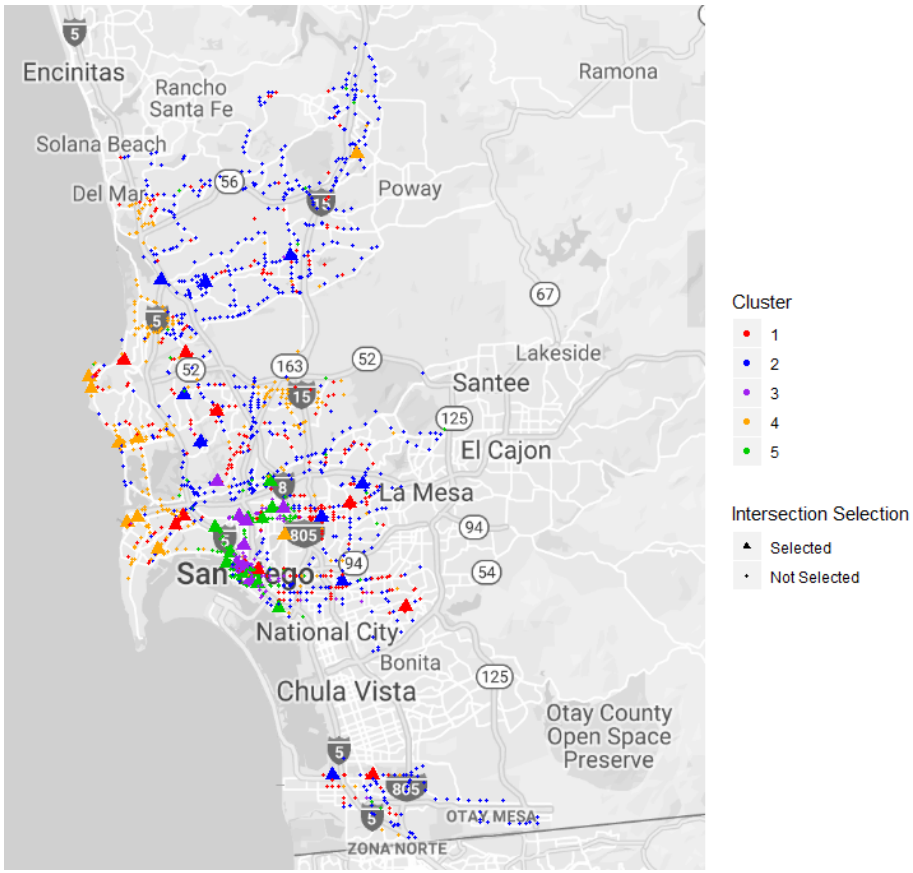
Crash risk (rate) calculation (Relative risk is indicated by r*)	Geographic unit of analysis; number of units	Reference
$r = \frac{Number\ of\ pedestrian\ casualties}{Million\ hours\ walking\ for\ an\ age\_sex\ group}$	Area-wide; 1 country (New Zealand)	(*9*)
$r = \frac{Number\ of\ pedestrian\ casualties}{Number\ of\ roads\ crossed\ (M)\ for\ an\ age\_sex\ group}$		
$r = \frac{Number\ of\ pedestrian\ casualties}{Thousand\ population\ for\ a\ age\_sex\ group}$		
$r = \frac{Number\ of\ pedestrian\ accidents}{product\ of\ the\ numbers\ of\ pedestrians\ and\ vehicles\ at\ a\ road\ section\ during\ a\ time\ interval}$	Facility-specific; 4 intersections	(*2*)
$r = \frac{Number\ of\ pedestrian\ or\ bicycle\ crashes}{hundred\ million\ pedestrian\ or\ bicycle\ miles\ of\ roadway\ (shared\ facility)\ traveled}$	Facility-specific; 122 [39 signalized intersections, 27 stop-controlled intersections, etc.]	(*12*)
$r^* = \left(\frac{Number\ of\ bicyle\ accidents\ that\ occur\ to\ the\ group}{Total\ number\ of\ bicyle\ accidents}\right) \Big/ \left(\frac{Number\ of\ bicylists\ in\ the\ group}{Total\ number\ of\ bicyclists}\right)$	Facility-specific; 9 [y signalized intersections and 2 stop-controlled intersections]	(*85*)
$r = \frac{Number\ of\ crashes\ for\ a\ certain\ mode\ of\ travel}{100\ million\ person\_trips}$	Area-wide; 1 country (United States)	(*8*)
$r = \frac{Number\ of\ injuries}{Millions\ of\ cyclists\ or\ pedestrians}$	Facility-specific; 1082 [647 signalized and 435 stop-controlled intersections]	(*3*)
$r = \frac{Number\ of\ pedestrian\ crashes}{10\ million\ pedestrian\ crossings}$	Facility-specific; 50 intersections	(*86*)
$\frac{r = Totlal\ number\ of\ pedestrian\ crashes\ at\ an\ intersection\ over\ a\ specific\ time\ period}{10\ million\ pedestrian\ crossings}$	Facility-specific; 22 intersections	(*4*)
$r = \frac{Number\ of\ Pedestrian\ Crashes\ over\ 5\ years}{Sum\ of\ cross\ product\ of\ the\ average\ daily\ pedestrians * distance\ crossed * average\ daily\ traffic\ for\ each\ directional\ approach}$	Facility-specific; 52 intersections	(*87*)
$r = \frac{Number\ of\ Pedestrian\ Crashes\ over\ 5\ years}{total\ length\ of\ the\ roadway\ category * Average\ daily\ traffic\ for\ the\ roadway\ category}$	Facility-specific; 135 roadway segments	(*87*)

$r = \frac{Number\ of\ bicycle\ crashes}{Intersection\ bicycle\ counts}$	Facility-specific; 232 intersections	(*23*)
$r = \frac{Number\ of\ bicycle\ crashes}{roadway\ screenline\ bicycle\ counts}$	Facility-specific; 816 roadway segments	(*23*)
$r = \frac{Number\ of\ pedestrian\ fatalities}{10\ million\ person\ hours\ traveled}$	Area-wide; 1 country (United States)	(*88*)
$r = \frac{Number\ of\ pedestrian\ crashes}{population}$ $r = \frac{Number\ of\ pedestrian\ crashes}{number\ of\ trips}$ $r = \frac{Number\ of\ pedestrian\ crashes}{trip\ distances}$ $r = \frac{Number\ of\ pedestrian\ crashes}{trip\ duration}$ $r = \frac{Number\ of\ pedestrian\ crashes}{number\ of\ streets\ crossed}$	Area-wide; 1 Ottawa-Carleton (community in Ontario, Canada)	(*7*)
$r = \frac{Number\ of\ pedestrian\ crashes}{number\ of\ trips}$ $r = \frac{Number\ of\ pedestrian\ crashes}{distance\ traveled}$ $r = \frac{Number\ of\ pedestrian\ crashes}{time\ spent\ travelling}$	Area-wide; 1 county (Rhône county, France)	(*6*)

Appendix B – Study workflow



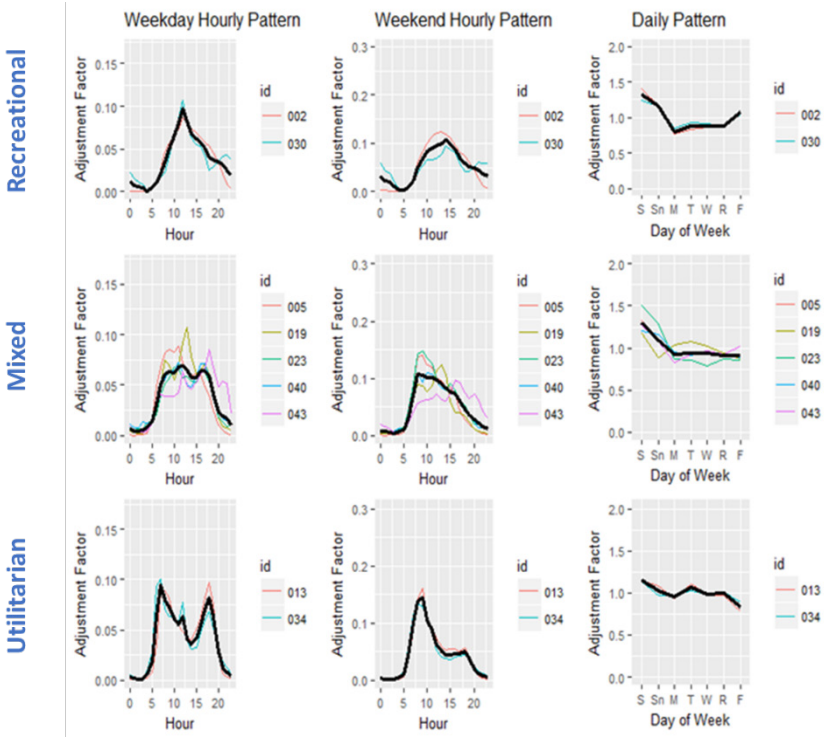
Appendix C – Site selection results by cluster analysis and stratification



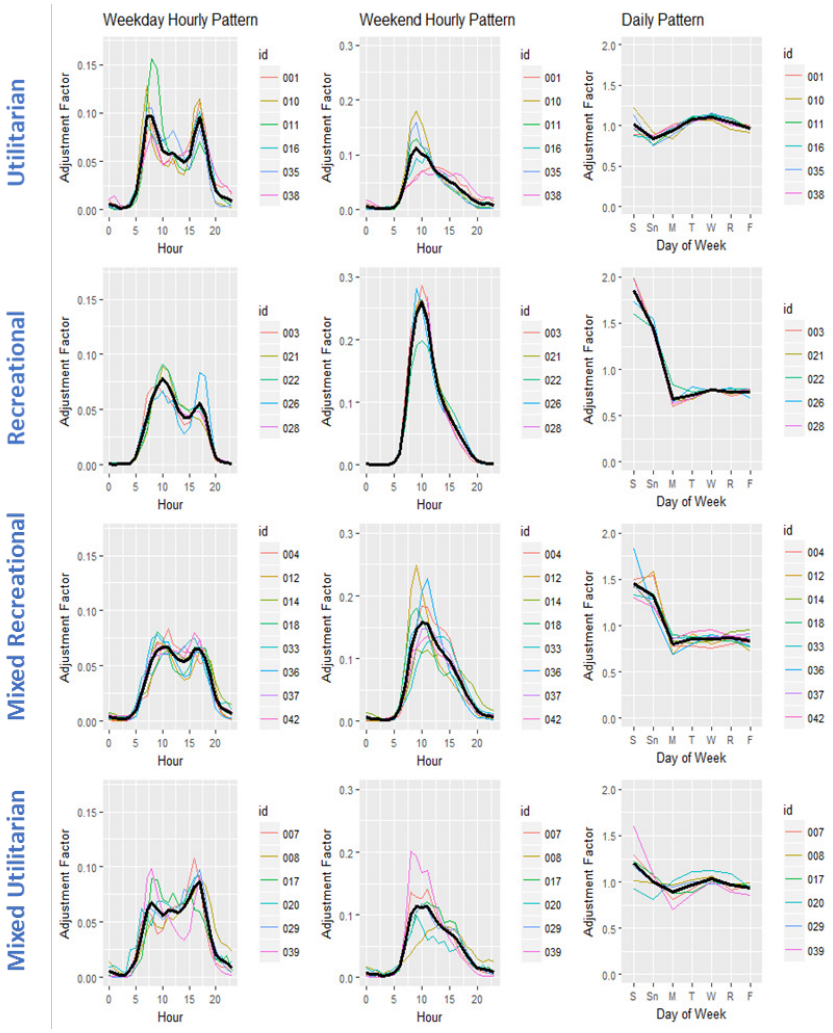
*** Triangles represent the final selection.**

Appendix D – Clustering result of long-term pedestrian and bicyclist counters

Clustering result of long-term pedestrian counters classified into three groups: recreational, mixed, and utilitarian.



Clustering result of long-term bicycle counters classified into four groups: utilitarian, recreational, mixed recreational, and mixed utilitarian.



Appendix E – Discussion of the Bicycle and Pedestrian Exposure Model Results

The pedestrian volume was characterized by transportation network (transit stop density and speed limit), population (employment density, and regular transit rider, pedestrian, or bicyclist population), and land use (vacant housing units, commercial or mixed-use land area, and high-crime area). The best model was obtained with variables of different spatial scales. This finding is consistent with previous studies (*89–91*) that suggested it is unlikely to have all the variables significant at the same buffer scale. The direct-demand model for bicycle traffic included variables that represent characteristics of the transportation network (density of bicycle facility, maximum posted speed within an intersection, and transit stops density), population (total regular bicyclist population), traffic generator (presence of a school and proximity to the beach), and land use (total commercial or mixed-use area).

Interestingly, some variables, such as commercial or mixed-use land area and transit stop density, influence both pedestrian and bicycle traffic for the study area, but the spatial scale of influence varies. The commercial or mixed-use land area influences pedestrian and bicycle volume within 0.160 kilometers (0.1 miles) and 0.402 kilometers (0.25 miles), respectively. This indicates that the commercial and mixed-use land area attracts bicyclist traffic for a larger catchment area than pedestrian traffic does. Previous studies have also indicated that commercial areas attract pedestrian (*28*, *91*) and bicycle (*90*, *28*, *92*, *89*) activity. However, the study by Tabeshian and Kattan (2014), conducted in Canada, found a significant impact of commercial areas on pedestrian and bicycle traffic within 0.402 kilometers (0.25 miles) and 0.160 kilometers (0.1 miles), respectively, which is contrary to this study's findings (*28*). Two separate studies in Alameda County, California, have also found a significant influence on commercial areas on pedestrians within 0.402 kilometers (0.25 miles) (*89*) and bicyclists within 0.160 kilometers (0.1 miles) (*92*). The comparison suggests that not only the explanatory variables but also their influence area for nonmotorized traffic activity vary with location and community.

Similarly, transit stop density influences pedestrians within 0.805 kilometers (0.5 miles), which is larger than the bicycle buffer of 0.160 kilometers (0.1 miles). The results indicate that pedestrians are likely to travel more to ride a transit facility than bicyclists are. Previous studies have also observed a significant association of transit facilities with pedestrians within 0.805 kilometers (0.5 miles) (*93*) and bicyclists within 0.805 kilometers (0.5 miles) (*29*). Given that mass transit facilities are bicycle-friendly in San Diego (San Diego Metropolitan Transit System, 2018), transit riders probably make up a large proportion of pedestrians and bicyclists in the city.

Pedestrian and bicycle volumes decrease when the maximum intersection speed limit exceeds 64 kph (40 mph). The finding confirms that pedestrians and bicyclists are more likely to avoid high-speed intersections and find an alternative route. Fagnant, D. J., and K. Kockelman, (*30*) also

observed a similar relationship for bicycle traffic in the Seattle, Washington, area. The finding is not surprising given the rise of traffic fatalities. A report (*94*) indicated that around 1,000 pedestrians and bicyclists are hit and seriously injured annually in San Diego, and in 2012, pedestrian collisions increased 20 percent compared to previous years. In 2017, there were 12 deaths and 71 serious injuries involving pedestrians and bicyclist (*94*, *95*). The high crash risk could discourage pedestrians and bicyclists from using high-speed intersections.

The pedestrian model had a strong positive association between the pedestrian volume and the percentage of regular transit rider, pedestrian, or bicyclist population within 0.402 kilometers (0.25 miles). As expected, the population inclined to use active modes and public transportation was more likely to contribute to the walking volume within their neighborhood. The pedestrian volume also increased with increasing employment density within 0.402 kilometers (0.25 miles). Previous studies also observed similar influence in San Francisco (*31*) and San Diego, California (*96*). The results suggest that with more people working in a neighborhood, intersections are more likely to observe higher pedestrian volume. Similarly, the negative association between pedestrian volume and total vacant housing units indicates that pedestrians are less likely to generate from neighborhoods with many vacant properties. The negative influence of crime on pedestrian volume, but not on bicyclists was also observed, which shows that people are more likely to avoid high-crime locations and conforms with previous research (83, 84).

As expected, the bicycle model indicated higher bicycle volume in areas with a larger population of regular bicyclists. The model also indicated that the intersections near beach access points (less than 1.6 kilometers (10 miles)) were more likely to observe high bicycle traffic. The density of bike facilities also had a positive influence on daily bicycle volume. The finding can be attributed to the recent surge of dockless bicycles in the city (*97*) as well as the 2.5 kilometers (16 miles) of separated bike paths around San Diego Bay, completed in February 2018 (San Diego Association of Governments, 2018). The Bay Shore bikeway was built with a vision to provide a scenic and convenient way for bicyclists to travel in the San Diego area. The dockless bike sharing facilities were first launched in February 2018 and added the convenience of using bicycles. The combined influence may contribute to a higher bicycle volume in locations near beach areas and with better bicycle facilities. Surprisingly, the model revealed a negative association between the presence of a school and bicyclist volume, which contradicts previous studies conducted in Canada (*29*, *98*). However, a study conducted in the United States suggested that the number of students (ages 5 to 18) who walk or bike to school decreased sharply in recent years due to increased traffic collisions, lack of sidewalks, and urban sprawl (McMillan, 2009). Perhaps the increasing collision rate in the city discouraged children from bicycling to schools, and adult bicyclists tend to avoid locations near schools.

Appendix F - Group Project Assignment for *CIVE 160: Statistical Methods for the Built Environment*

For this project, use the data set that is assigned to your group to perform the following tasks:

Task 1: Descriptive Statistics

- Task 1-1: What is the variable that you are trying to estimate? (e.g., AADP). Use descriptive statistics to summarize this variable.
- Task 1-2: See other variables that are included in your data set. Select two variables (preferably continuous) in your data set that you think have impact on the variable you identified in Task 1-1. Use descriptive statistics to summarize these two variables.

Task 2: Hypothesis Testing

- *Task 2-1:* From your data set, select the variable identified in Task 1-1.
    - Perform a hypothesis testing at 95% confidence level to find out if the criterion for this variable (e.g., AADP > 200) is met (apply the p-value approach).
    - Construct a 99% CI on the population mean based on the criterion applied earlier.
- *Task 2-2:* Choose another continuous variable (i.e. different from AADP) of your choice, define a hypothetical criterion that requires a two-sided hypothesis, and perform the hypothesis testing for your scenario (use sample size of 60).

Task 3: Regression Analysis

- *Task 3-1:* Develop a simple linear regression model to estimate the variable identified in Task 1-1. Try to find the best simple linear regression model. Plot the best model and report the outcomes (model parameters as well as the R^2 measure). Interpret the results.
- *Task 3-2:* Develop a multiple linear regression model to estimate the variable identified in Task 1-1. Find the best model and report the outcomes of the best model. You could include as many independent variables as you want in your model but see if adding variables would improve your model. Interpret the results.

Appendix G - Pedestrian/Bicyclist Counting Tool

**Only for windows systems*

First, take the following four steps to install required software/packages in order to use the counting tool.

Step 1:

- Install MATLAB Runtime software by clicking on the link below.

https://www.mathworks.com/products/compiler/matlab-runtime.html

- Must be 64-bit, each release should work.

Step 2:

- Download the app package attached (PedCounterMLRuntime.rar) and uncompress it. Install the app by clicking the "PedCounterMLRuntime\for_redistribution\MyAppInstaller_web.exe" file.

- You will be asked to fill in the Matlab Runtime software installation path, like "C:\Program Files\MATLAB\MATLAB Runtime".

- The app will be installed in a path like "C:\Program Files\PedCounterMLRuntime".

Step 3:

- Copy the "PedCounterMLRuntime\Frames" folder into the "application" folder of app installation path "C:\Program Files\PedCounterMLRuntime\application\".

Step 4:

- Run the system by clicking the PedCounterMLRuntime.exe in the "application" folder of app installation path "C:\Program Files\PedCounterMLRuntime\application\".

Second, follow the below instructions to use the tool:

After running the PedCounterMLRuntime.exe, a GUI pops up that shows three buttons: Select Region; Play; and Clear Region. For the first time, click on the "Play" button so the video images are loaded. Then, click on the Select Region to define an area for counting by drawing a rectangle. Then, hit play; the pedestrians entering the area will be counted and total number of pedestrians counted are shown on the screen. To define a new region, you can use "Clear Region" button and then use the "Select Region" button to define a new area.

Appendix H - Data Specification

X indicates three different buffer sizes including 0.160, 0.402, and 0.805 kilometers (0.1, 0.25, and 0.5 miles) = {1: 0.1 mile, 2: 0.25 mile, 5: 0.5 mile}

xx indicates year 2006 to 2016. (06: 2006, 07:2007, …, 16: 2016)

TOT_POP_X (Double)

-- The total sum of population within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

POP_DENS_X (Double)

-- Calculated by TOT_POP divided by BUFF_ACRE. This is the population density of each acre within the intersection buffer.

WHITE_X (Double)

--The total sum of population who self-identifies as white within a given intersection buffer, proportioned by total population by census group over buffer acreage.

WHITE_PCT_X (Double)

-- Calculated by WHITE divided by TOT_POP. This is the percentage of the intersection buffer that identifies as white.

BLK_X (DOUBLE)

--The total sum of population who self-identifies as African-American within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

BLACK_PCT_X (Double)

-- Calculated by BLACK divided by TOT_POP. This is the percentage of the intersection buffer that identifies as African-American.

AM_IND_X (DOUBLE)

--The total sum of population who self-identifies as American-Indian within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

AM_IND_PCT_X (Double)

-- Calculated by AM_IND divided by TOT_POP. This is the percentage of the intersection buffer that identifies as American-Indian

ASIAN_X (DOUBLE)

--The total sum of population who self-identifies as Asian within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

ASIAN_PCT_X (Double)

-- Calculated by ASIAN divided by TOT_POP. This is the percentage of the intersection buffer that identifies as Asian.

NAT_HI_X (DOUBLE)

--The total sum of population who self-identifies as Native-Hawaiian within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

NAT_HI_PCT_X (Double)

-- Calculated by NAT_HI divided by TOT_POP. This is the percentage of the intersection buffer that identifies as Native-Hawaiian.

OTHER_X (DOUBLE)

--The total sum of population who self-identifies as any other race, not previously mentioned, within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

OTHER_PCT_X (Double)

-- Calculated by OTHER divided by TOT_POP. This is the percentage of the intersection buffer that identifies as Other.

MIXED_X (DOUBLE)

--The total sum of population who self-identifies as Mixed within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

MIXED_PCT_X (Double)

-- Calculated by MIXED divided by TOT_POP. This is the percentage of the intersection buffer that identifies as Mixed.

MINOR_X (Double)

--The total sum of population who self-identifies as a minority, including BLK, AM_IND, ASIAN, NAT_HI, OTHER, and MIXED within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

MINOR_PCT_X (Double)

-- Calculated by MINOR divided by TOT_POP. This is the percentage of the intersection buffer that identifies as a minority.

HISPANIC_X (Double)

--The total sum of population who self-identifies as a Hispanic or Latino within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_HISP_X (Double)

-- Calculated by HISPANIC divided by TOT_POP. This is the percentage of the intersection buffer that identifies as a Hispanic or Latino.

MALE_X (Double)

-- The total sum of population who self-identifies as male within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_MALE_X (Double)

-- Calculated by MALE divided by TOT_POP. This is the percentage of the intersection buffer that identifies as a male.

FEMALE_X (Double)

-- The total sum of population who self-identifies as female within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_FEMALE_X (Double)

-- Calculated by FEMALE divided by TOT_POP. This is the percentage of the intersection buffer that identifies as a female.

A_18_LESS_X (Double)

-- The total sum of population 18 years old or less within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_18_LESS_X (Double)

-- Calculated by A_18_LESS divided by TOT_POP. This is the percentage of the intersection buffer that identifies as under 18 years old.

A_65_OLDER_X (Double)

-- The total sum of population 65 years old or older within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_65_OVER_X (Double)

-- Calculated by A_65_OLDER divided by TOT_POP. This is the percentage of the intersection buffer that identifies as over 65 years old.

A18_TO_65_X (Double)

-- The total sum of population between the ages of 18 and 65 years old within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_18_65_X (Double)

-- Calculated by A18_TO_65 divided by TOT_POP. This is the percentage of the intersection buffer that identifies within this age group.

A18_24_X (Double)

-- The total sum of population between the ages of 18 and 24 years old within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_A18_24_X (Double)

-- Calculated by A18_24 divided by TOT_POP. This is the percentage of the intersection buffer that identifies between the ages of 18 and 24 years old.

TBP_Ridership_X (Double)

-- The total sum of population that that are transit riders, pedestrians, or bicyclists within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_TBP_X (Double)

-- Calculated by TBP_Ridership divided by TOT_POP. This is the percentage of the intersection buffer that identifies as a regular transit rider, bicyclist, or pedestrian.

Transit_RS_X (Double)

-- The total sum of population that that are regular transit riders within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

Bicycle_RS_X (Double)

-- The total sum of population that that are regular bicyclists within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

Ped_RS_X (Double)

-- The total sum of population that that are regular pedestrians within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

College_Cnt_X (Double)

-- The total sum of population with some college education within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_COLLEGE_X (Double)

-- Calculated by College_Cnt divided by TOT_POP. This is the percentage of the intersection buffer that has some college education.

POVERTY_X (Double)

-- The total sum of population living in conditions of Poverty (as defined by Census.gov ID= B17010) within a given intersection buffer, proportioned by total population by census block group over buffer acreage.

PCT_POV_X (Double)

-- Calculated by Poverty divided by TOT_POP. This is the percentage of the intersection buffer that is living in poverty.

EMP_2015_X (Double)

-- The total sum of employees within a given intersection buffer, proportioned by total employees in 2015 by census block group over buffer acreage. (Used ID= B24080).

EMP_DENS_X (Double)

-- EMP_2015 divided by acres within the intersect created for apportioning.

TOT_JOBS_X (Double)

-- The total sum of jobs within a given intersection buffer.

EMP_POP_X (Double)

-- Calculated by EMP_2015 divided by TOT_POP. This is the ratio of employees to population within the intersection buffer.

MED_INC_X (Double)

--The average median income within a given intersection buffer.

TOT_HH_X (Double)

-- The total sum of households within a given intersection buffer, proportioned by total households by census block group over buffer acreage.

TOT_HU_X (Double)

-- The total sum of housing units within a given intersection buffer, proportioned by total housing units by census block group over buffer acreage.

TOT_VAC_X (Double)

-- The total sum of vacant housing units within a given intersection buffer, proportioned by total housing units by census block group over buffer acreage.

PCT_VAC_X (Double)

-- Calculated by TOT_VAC divided by TOT_HU. This is the percentage of vacant housing units within the intersection buffer.

NO_CAR_HH_X (Double)

--The total sum of households without an automobile within a given intersection buffer, proportioned by total households without an automobile by census tract over buffer acreage.

SW_FEET_X (Double)

-- The total sum of sidewalk length in feet within a given intersection buffer.

SW_DENS_X (Double)

-- Sidewalk feet within the buffer divided by roadway feet within the buffer

BIKE_FEET_X (Double)

-- The total sum of bicycle facility length in feet within a given intersection buffer.

BIKE_DENS_X (Double)

-- Bicycle facility length within the buffer divided by the roadway within the buffer.

R_INT_CNT_X

-- The total sum of intersections within a given buffer.

R_INT_DENS_X

--Calculated by dividing R_INT_CNT by the total acreage of the buffer.

AUTO_LU_X (Double)

-- The total sum of automobile-dependent acres within the buffer.

A_LU_DENS_X (Double)

-- AUTO_LU divided by buffer acres.

AG_LU_X (Double)

-- The total sum of Agriculture Land Use acres within the buffer.

AG_LU_DENS_X (Double)

-- AG_LU divided by buffer acres.

COMM_LU_X (Double)

--The total sum of Commercial or Mixed-Use Land use acreage within the buffer.

C_LU_DENSE_X (Double)

--C_LU divided by acreage within the buffer.

GOVT_LU_X (Double)

--The total sum of Government Land Use acreage within the buffer.

GOVT_LU_DENS_X (Double)

--GOVT_LU divided by acreage within the buffer.

IND_LU_X (Double)

--The total sum of Industrial Land Use acreage within the buffer.

IND_LU_DENS_X (Double)

--C_LU divided by acreage within the buffer.

PAR_REC_LU_X (Double)

--The total sum of Parks and Recreation Land Use acreage within the buffer.

P_R_LU_DENS_X (Double)

--PAR_REC_LU divided by acreage within the buffer.

RES_LU_X (Double)

--The total sum of Residential Land Use acreage within the buffer.

RES_LU_DENS_X (Double)

--RES_LU divided by acreage within the buffer.

TRVL_LU_X (Double)

--The total sum of Travel Land Use acreage within the buffer.

TRVL_LU_DENS_X (Double)

--TRVL_LU divided by acreage within the buffer.

VAC_UN_LU_X (Double)

--The total sum of Vacant and Undeveloped Land Use acreage within the buffer.

VA_UN_LU_DENS_X (Double)

--VAC_UN_LU divided by acreage within the buffer.

Office_LU_X (Double)

--The total sum of Office Land Use acreage within the buffer.

Pct_Office_X (Double)

--Office_LU divided by acreage within the buffer.

S_Fam_LU_X (Double)

--The total sum of Single-Family Housing Land Use acreage within the buffer.

Pct_S_Fam_X (Double)

--S_Fam_LU divided by acreage within the buffer.

M_Fam_LU_X (Double)

--The total sum of Multi-Family Housing Land Use acreage within the buffer.

Pct_M_Fam_X (Double)

--M_Fam_LU divided by acreage within the buffer.

Retail_LU_X (Double)

--The total sum of Retail Land Use acreage within the buffer.

Pct_Retail_X (Double)

--Retail_LU divided by acreage within the buffer.

Business_LU_X (Double)

--The total sum of Business Land Use acreage within the buffer.

Pct_Business_X (Double)

--Business_LU divided by acreage within the buffer.

RoadAll_Ft_X (Double)

-- The total sum footage of All Roads within the given intersection buffer.

Major_Arterial_X (Double)

-- The total sum footage of all Major Arterial Roads within the given intersection buffer.

Pct_MjrArt_X (Double)

-- Calculated by Major_Arterial divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Major Arterial.

MajorRD_4L_X (Double)

-- The total sum footage of all 4-Lane Major Roads within the given intersection buffer.

Pct_4L_MRd_X (Double)

-- Calculated by 4L_MajorRD divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a 4-Lane Major Road.

Collector_4L_X (Double)

-- The total sum footage of all 4-Lane Collector Roads within the given intersection buffer.

Pct_4L_Col_X (Double)

-- Calculated by 4L_Collector divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a 4-Lane Collector Road.

Local_St_X (Double)

-- The total sum footage of all Local Streets within the given intersection buffer.

Pct_LcalSt_X (Double)

-- Calculated by Local_St divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Local Street.

Local_Rd_X (Double)

-- The total sum footage of all Local Roads within the given intersection buffer.

Pct_LcalRd_X (Double)

-- Calculated by Local_Rd divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Local Road or Rural Light Collector.

Multi_UsePth_X (Double)

-- The total sum footage of all Pedestrianways or Bikeways within the given intersection buffer.

Pct_MU_Pth_X (Double)

-- Calculated by Multi_UsePth divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Pedestrianway or Bikeway

Collector_2L_X (Double)

-- The total sum footage of all 2-Lane Collector Roads within the given intersection buffer.

Pct_2L_COL_X (Double)

-- Calculated by 2L_Collector divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a 2-Lane Collector Road.

Rrl_Cllctr_X (Double)

-- The total sum footage of all Rural Collector Roads within the given intersection buffer.

Pct_RClctr_X (Double)

-- Calculated by Rrl_Cllctr divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Rural Collector Road.

Private_St_X (Double)

-- The total sum footage of all Private Streets within the given intersection buffer.

Pct_Priv_X (Double)

-- Calculated by Private_St divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Private Street.

Alley_Ft_X (Double)

-- The total sum footage of all Alleys within the given intersection buffer.

Pct_Alley_X (Double)

-- Calculated by Alley_Ft divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as an Alley.

MjrRd_2L_X (Double)

-- The total sum footage of all 2-Lane Major Roads within the given intersection buffer.

Pct_2L_MRd_X (Double)

-- Calculated by 2L_MjrRd divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a 2-Lane Major Road.

Milit_St_X (Double)

-- The total sum footage of all Military Streets within the given intersection buffer.

Pct_Milit_X (Double)

-- Calculated by Milit_St divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Military Street.

MjrSt_6L_X (Double)

-- The total sum footage of all 6-Lane Major Streets within the given intersection buffer.

Pct_6L_MSt_X (Double)

-- Calculated by 6L_MjrSt divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a 6-Lane Major Street

Unpaved_Rd_X (Double)

-- The total sum footage of all Unpaved Roads within the given intersection buffer.

Pct_Unpaved_X (Double)

-- Calculated by Unpaved_Rd divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as an Unpaved Road.

Freeway_Ft_X (Double)

-- The total sum footage of all Freeways within the given intersection buffer.

Pct_FW_X (Double)

-- Calculated by Freeway_Ft divided by RoadAll_Ft. This is the percentage of the intersection buffer's roadways that are classified as a Freeway.

MTS_COUNT_X (Single)

-- Sum of total transit stops within a given intersection buffer.

MTS_DENS_X (Double)

-- MTS_COUNT divided by acreage within the buffer.

Bus_Km_X (Double)

-- Total sum kilometers of bus routes within a given intersection buffer.

ROUTE_CNT_X (Single)

-- Sum of total transit routes within a given intersection buffer.

H_School_X (Single)

-- Sum of total high schools within a given intersection buffer.

M_School_X (Single)

-- Sum of total middle schools within a given intersection buffer.

E_School_X (Single)

-- Sum of total elementary schools within a given intersection buffer.

Schools_X (Single)

-- Sum of total schools, including colleges, within a given intersection buffer.

BUFF_ACRE_X (Double)

--Total acreage of each individual intersection buffer.

SCHOOL_Y_X (Binary)

--Binary code (1=yes, 0= no) distinguishes the presence of a school within a given intersection buffer.

COLLEGES_Y_X (Binary)

--Binary code (1=yes, 0= no) distinguishes the presence of a college within a given intersection buffer.

FREEWAY_Y_X (Binary)

--Binary code (1=yes, 0= no) distinguishes the presence of a freeway within a given intersection buffer.

CRIME_07_12_X (Double)

-- The total sum of reported incidents of crime between the years of 2007-2012 within a given intersection buffer, proportioned by total crime count by census block group over buffer acreage.

CRIME_AVG_X (Binary)

--Binary code (1=yes, 0= no) distinguishes whether the buffer contains a crime count that is higher than the average crime counts amongst the buffers.

INTERID

-- Internal Feature Number

LATITUDE

--provides the latitude geographic coordinates of a given intersection on a map.

LONGITUDE

--Provides the latitude geographic coordinates of a given intersection on a map

TOT_xx_X (Years 2006-16)

-- The total sum of automobile collisions involving either bicyclists and/or pedestrians within a given intersection's 9-meter (30-foot) buffer in a given year.

*TOT_F_xx_X (*Years 2006-16)

--The total sum of automobile collisions ending in a fatality that involved either bicyclists and/or pedestrians within a given intersection in a given year.

P_xx_X (Years 2006-2016)

--The total sum of automobile collisions involving a pedestrian within a given intersection in a given year.

P_F_xx_X (Years 2006-2016)

--The total sum of automobile collisions ending in a fatality and involving a pedestrian within a given intersection in a given year.

B_xx_X (Years 2006-2016)

--The total sum of automobile collisions involving a bicylist within a given intersection in a given year.

B_F_xx_X (Years 2006-2016)

--The total sum of automobile collisions ending in a fatality and involving a bicylist within a given intersection in a given year.

INT_CNT_xx_X (Years 2006-16)

-- The total sum of automobile collisions involving either bicyclists and/or pedestrians within a given intersection point in a given year.

TOTAL_X

-- The total sum of automobile collisions involving either bicyclists and/or pedestrians within a given intersection between the years 2006 and 2016. (Note: Data set for the years 2013-2016 remains provisional).

TOTAL_F_X

-- The total sum of automobile collisions involving either bicyclists and/or pedestrians, that included a fatality, within a given intersection between the years 2006 and 2016. (Note: Data set for the years 2013-2016 remains provisional).

MIN_LNS (Single)

-- The minimum number of roadway lanes within a given intersection.

MAX_LNS (Single)

-- The maximum number of roadway lanes within a given intersection.

MIN_VOL (Double)

-- Flow data extracted from hwy573 load data from SANDAG. If no data was available, the value was given a default 0

MAX_VOL (Double)

--Flow data extracted from hwy573 load data from SANDAG. If no data was available, the value was given a default value of 3500.

MIN_PT_SD (Single)

-- The minimum posted speed within an intersection measured by miles per hour.

MAX_PT_SD (Single)

-- The maximum posted speed within an intersection measured by miles per hour.

MILES_BP (Double)

-- Distance measured in miles from the intersection point to Balboa Park.

MILES_BCH (Double)

-- Distance measured in miles from the intersection point to beach front access points.

MILES_RAMP (Double)

-- Distance measured in miles from the intersection point to a highway on or off ramp.

MEAN_VOL (Double)

--Average of MIN_VOL and MAX_VOL

MEAN_LNS (Double)

--Average of MIN_LNS and MAX_LNS

BEACH (Double)

-- Categorized by the following:

"pedDist" = Miles to beach less than 0.805 kilometers (0.5 miles)

"bikeDist" = Miles to beach greater than 0.805 kilometers (0.5 miles) and less than 4.8 kilometers (3 miles)

"CarDist" = Miles to beach greater than 4.8 kilometers (3 miles)

BALBOAP (Double)

-- Categorized by the following:

"pedDist" = Miles to Balboa Park less than 0.805 kilometers (0.5 miles)

"bikeDist" = Miles to Balboa Park greater than 0.805 kilometers (0.5 miles) and less than 4.8 kilometers (3 miles)

"CarDist" = Miles to Balboa Park greater than 4.8 kilometers (3 miles)

PED_VIC

-- Number of pedestrian victims during year 2006 to 2016

BIKE_VIC

-- Number of bicycle victims during year 2006 to 2016

PED_CRA

-- Number of pedestrian crashes occurred during year 2006 to 2016

BIKE_CRA

-- Number of bicycle crashes occurred during year 2006 to 2016

PED_COST

-- Total pedestrian crash costs occurred at an intersection during year 2006 to 2016 weighted by severity level.

BIKE_COST

-- Total bicycle crash costs occurred at an intersection during year 2006 to 2016 weighted by severity level.

PED_SEV

-- Average pedestrian injury severity level during year 2006 to 2016

BIKE_SEV

-- Average bicyclist injury severity level during year 2006 to 2016

DIST_CROSSED

-- Distance a pedestrian or a bicyclist crossed at an intersection

aadp

--Annual Average Daily Pedestrian counts calculated for 45 selected intersections by converting short-term count data to yearly count data

aadb

--Annual Average Daily Bicyclist counts calculated for 45 selected intersections by converting short-term count data to yearly count data

AADP

--Annual Average Daily Pedestrian counts calculated for all intersections through exposure modeling

AADB

--Annual Average Daily bicyclist counts calculated for all intersections through exposure modeling

PED_RISK

--Pedestrian risk at an intersection calculated based on the proposed equation

BIKE_RISK

--Bicyclist risk at an intersection calculated