SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Comprehensive Performance Assessment of Passive Crowdsourcing for Counting Pedestrians and Bikes

Wen Cheng, PhD
Yongping Zhang, PhD, PE
Edward Clay

CSUTC California State University Transportation Consortium

CalPolyPomona

CSU TRANSPORTATION CONSORTIUM

transweb.sjsu.edu/csutc

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 22-03

Comprehensive Performance Assessment of Passive Crowdsourcing for Counting Pedestrians and Bikes

Wen Cheng, PhD

Yongping Zhang, PhD, PE

Edward Clay

January 2022

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 22-03	2. Government Accession No.	3. Recipient's Catalog No. 22-03	
4. Title and Subtitle Comprehensive Performance Assessment of Passive Crowdsourcing for Counting Pedestrians and Bikes		5. Report Date January 2022 6. Performing Organization Code	
7. Authors Wen Cheng ORCID: 0000-0001-7225-6169 Yongping Zhang ORCID: 0000-0002-5935-3834 Edward Clay ORCID: 0000-0002-7096-2323		8. Performing Organization Report CA-MTI-2025	
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		10. Work Unit No. 11. Contract or Grant No. ZSB12017-SJAUX	
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Long Beach, CA 90802		13. Type of Report and Period Covered 14. Sponsoring Agency Code	
15. Supplemental Notes			
16. Abstract Individuals who walk and cycle experience a variety of health and economic benefits while simultaneously benefiting their local environments and communities. It is essential to correctly obtain pedestrian and bicyclist counts for better design and planning of active transportation-related facilities. In recent years, crowdsourcing has seen a rise in popularity due to the multiple advantages relative to traditional methods. Nevertheless, crowdsourced data have been applied in fewer studies, and their reliability and performance relative to other conventional methods are rarely documented. To this end, this research examines the consistency between crowdsourced and traditionally collected count data. Additionally, the research aims to develop the adjustment factor between the crowdsourced and permanent counter data, and to estimate the annual average daily traffic (AADT) data based on hourly volume and other predictor variables such as time, day, weather, land use, and facility type. With some caveats, the results demonstrate that the StreetLight crowdsourcing count data for pedestrians and bicyclists appear to be a promising alternative to the permanent counters.			
17. Key Words Active Transportation; Pedestrian;Bicyclist; Count; Crowdsourcing.	18. Distribution Statement No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 63	22. Price

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2022.2025

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2025

Acknowledgments

This study is funded by the California State University Transportation Consortium (CSUTC). The authors are grateful to Mr. Benjamin Schuster and Mr. Agustin Cuello Leon, who served as the external project advisors and provided consistent support and insights. The authors would also like to thank Dr. Hilary Nixon for her valuable comments and continuous support. Finally, the efforts and comments of the anonymous reviewers are significantly appreciated as well.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive Summary

Walking and bicycling provide several health, environmental, and economic benefits to those performing these actions for transportation, communities, and local traffic areas alike. Due to the numerous benefits active transportation offers, it is essential to understand pedestrian and bicyclist traffic volumes at various times in order to better design active transportation-related infrastructures and establish associated policies. There are many methods available to record these volumes, including permanent bike and pedestrian detectors/counters. In recent years, crowdsourcing has seen a rise in popularity due to the ease of collecting data this way compared to traditional methods. Nevertheless, crowdsourced data have been applied in fewer studies, and in the limited research available, crowdsourced data appear to differ from the counts provided by other means. For this reason, it is necessary to further check the consistency between the crowdsourced and other count data and generate an adjustment factor if needed. These are the goals of the present study. Specifically, crowdsourced data were collected from StreetLight, and permanent counter data were obtained from the City of San José and the national archive for bicycle and pedestrian count data maintained by Portland State University. The data of interest originate from various cities in California, including Del Mar, San José, and San Diego, and they cover various types of facilities including collector and arterial roads, trails, and shared use paths. To understand the statistical difference between the two sets of data (i.e., the StreetLight and permanent counter data), we performed a two-tailed t-test and a Wilcoxon signed-rank test.

Following these tests, both R-squared and Pearson's correlation coefficient were calculated to determine the linear association between datasets. In addition, Spearman's correlation coefficient and Kendall's τ tests were also conducted to check the association between the two types of data in case of a nonlinear relationship. Moreover, the systematic adjustment factor between the StreetLight and counter data were determined using both fixed and random intercept models based on the Integrated Nested Laplace Approximation (INLA) package in R. Finally, to estimate a more useful annual average daily traffic (AADT) of the active transportation counts based on the readily available hourly volume of a specific period, the study provided various estimation models with different levels of complexity being considered. Such models provide additional insights to practitioners engaged in estimating active transportation-related AADT based on the associated hourly volume.

The proposed study dedicated to evaluating the counting performance of emerging technology (namely, SL crowdsourcing as a data collection method) is expected to benefit the research community and Californians in different ways. The results shed much light on the topics under consideration for researchers, planning practitioners, and policy makers, enhancing the understanding of non-motorized counting accuracy associated with the emerging passive crowdsourcing technology, which is very important information equipping various California jurisdictions to make the proper choice among the available counting technologies. For example, the Department of Transportation in the City of San José has piloted the StreetLight big data

program to track automobile volumes in the City, and now they need a timely research study to evaluate the StreetLight data performance for detecting the volume of active transportation-related modes. In addition, the better pedestrian and bike counts resulting from this project could aid Californians in: (a) accurate modeling of transportation networks and estimating annual volumes; (b) better evaluation of the effects of new infrastructure on pedestrian and bicycle activity; (c) reliable tracking of changes in pedestrian and bicycle activity over time; (d) precise non-motorist exposure information needed for relative risk analyses; and (e) enhanced prioritization of pedestrian and bicycle projects.

1. Introduction

Bicycle and pedestrian data collection play a critical role in many aspects of transportation, including prioritization and evaluation of new facility provision, calibration of various transportation demand models, development of multimodal safety performance functions, estimation of active transportation volumes under different conditions, and so on (Cheng et al., 2018a & 2018b). However, the data available for pedestrians and bicyclists are more limited compared to other transportation modes, even though the non-motorized modes preceded their motorized counterparts. The possible explanations for this discrepancy stem from the technological challenges unique to non-motorized traffic monitoring, which include, but are not limited to, more unexpected biking and walking movements, less predictable travel paths, more travel in groups, larger temporal variability in demand, greater sensitivity to weather conditions, lower speed compared to motorized trips, and difficulty of detecting active mode users, whose moving volume is smaller than motorized mode users' (Ohlms et al., 2019).

Despite all the obstacles mentioned above, the collection of pedestrians and bicyclist counts has enjoyed significant progress in the past decade due to technological advancements, a growing demand for detailed information on non-motorized modes, and the availability of a set of guidelines for such data collection (Birk et al., 2006; Ryus et al., 2014; FHWA, 2016). Many counting methods have been developed with different taxonomies: for example, manual vs. automated, passive vs. active, and traditional vs. emerging. In general, the different types of methods have their strengths and weaknesses. The manually collected data (directly from field observation or video recordings) appear to be more accurate, while automated counting (using sensors, tubes, or other devices) has the benefit of requiring fewer personnel hours and therefore being less expensive if properly calibrated and subsequently maintained (FHWA 2016). Even though the passive data collection methods are more convenient since there is little or no interaction required with the pedestrians and bicyclists, the active methods could yield more information from the respondents, such as the perceived level of service and socioeconomic input at the individual journal level. The traditional counting data (Cottrell & Dharminder, 2003; Griffin et al., 2014) are more historically available and include both passive data collection methods (e.g., the manual or automated counting) and active ones including various types of surveys such as the U.S. Census American Community Survey, national/regional household travel surveys, GPS-oriented surveys, and web-based and intercept surveys.

Compared with the traditional data collection methods, the emerging methodologies have fewer applications so far and mainly rely on crowdsourcing based on mobile devices such as mobile phones, wearable wristbands, tablets, and so forth. To explore another taxonomic pair, the crowdsourced data collection methods could also be divided into active and passive alternatives. The active technologies include regional bicycling tracking apps developed by public agencies, private companies' fitness/activity tracking apps, app-based bike-share programs, and map inventory apps based on user feedback. The passive data sources can include global positioning

systems, location-based services (LBS), and mobile phone positioning (MPP). A literature review report (Lee and Sener, 2017) identifies the details of all crowdsourcing types and associated commercial vendors: for example, CycleTracks, Strava, or AirSage. It is known from that literature review that emerging technologies have the main advantage of providing much broader and more diverse samples of the active transportation population with fewer resource constraints. In contrast, traditional monitoring methods have heavily relied on massive efforts from data collectors, which often lead to limited data collection locations and under-sampled data. Nonetheless, the conventional counting methods have widespread applications, and thus, for these methods, more historical data are available for various assessment purposes. Most transportation agencies have standard practices or guidance for collecting non-motorized data using the typical methods. In addition, they have developed "rules of thumb" to adjust the data collected to enhance the accuracy based on a multitude of evaluation/validation studies that cover various counting technologies (Nordback et al., 2011; Nordback et al., 2011; Kothuri et al., 2017; Mooney et al., 2016), the expansion of short-term counts to longer-period ones (Schneider et al., 2009; Beitel et al., 2017 & 2018; Hankey et al., 2014), estimating average daily volumes (El Esawey et al., 2013 & 2014; Figliozzi et al., 2014), the development of weather and other adjustment factors (Schmiedeskamp & Zhao, 2016; Nosal et al., 2014), calibration of non-motorized volume modeling (Raford & Ragland, 2004; Liu & Griswold, 2009), and so on. In contrast, little research is dedicated to the evaluation of emerging counting methods, despite the growing applications of such technologies (Barajas et al., 2017; Saad et al., 2019). Thus far, most of these evaluation studies have focused on crowdsourced active data (Griffin and Jiao, 2015; Jestico et al., 2016; Watkins et al., 2016). As for passively crowdsourced data, little research has been conducted based on small-scale GPS and MPP data (see Jahangiri & Rakha, 2015, and Mun et al., 2008, for details). To the authors' best knowledge, no evaluation has been performed based on LBS for walking and cycling count data accuracy.

The present study is dedicated to the comprehensive assessment of LBS data accuracy for non-motorized count data recently made available by StreetLight (SL) to fill the research gap. Moreover, we develop a system adjustment factor between the permanent counter and SL data. Finally, we offer models containing the temporal, weather, land use, and facility type variables to facilitate the estimation of the AADT of pedestrians and bicyclists.

2. Literature Review

Active transportation refers to human-powered methods of travel, such as walking, bicycling, or rolling to get from one place to another (ATSP, n.d.). There are many benefits associated with active transportation. One significant benefit is improved overall health, as active transportation can reduce the risk of obesity and other chronic conditions such as diabetes and various cardiovascular diseases (DoT, 2015). Another benefit of active transportation is neighborhood livability: when there is an increased number of people out and about in neighborhoods, the likelihood of crime is noticeably reduced (Schlossberg et al., 2012). Active transportation is also known to reduce the cost of transportation for individuals and reduce the road maintenance required, thus generating economic benefits (Litman, 2013; Litman, 2015). In addition to financial benefits, there are also several environmental benefits. Since active transportation relies on one's own power, there is no need for gasoline, diesel, or electricity (Hong, 2018). Finally, the wide variety of active transportation methods, primarily walking and biking, as well as their relatively low cost compared to vehicles, improves the overall mobility factor in specific areas where there is greater access to active transportation facilities. When facilities for active modes of transportation are available, people are less inclined to use cars that eventually contribute to traffic congestion (Lindblad, n.d.).

Several methods are available to determine the volume of bicyclists and pedestrians at specific intersections and trails. One of the more traditional methods of recording data is manual counting (Somasundaram et al., 2009). While manual counting can produce the count data in a more direct way, it is subject to some issues: for example, it is time-consuming, prone to human error, and poses a safety risk to those recording data on-site (John & Johnson, 2000; Toth et al., 2013). For these reasons, manual counting isn't an ideal method of data collection. Fortunately, automatic data collection methods exist, facilitated by technologies including electromagnetic loop detectors (Anderson, 1970) and cameras paired with computer vision algorithms (Uke & Thool, 2013); these methods are preferable, since they are capable of passive data collection following installation. For this reason, many cities have opted for automatic methods for recording vehicle volumes. However, electromagnetic loop detectors in some rare cases cannot accurately detect bicyclists or pedestrians, unlike manual counting, despite their ability to passively count larger vehicles within an intersection or along a road. Additionally, loop detectors must be installed within the pavement and require regular maintenance (Han et al., 2009). On the other hand, while cameras can be installed above ground, they are expensive, prone to poor performance in certain weather and lighting conditions, and need regular maintenance and proper placement in reference to line-of-sight obstructions (Fries et al., 2007). Crowdsourced data can fall into two distinct subgroups: active data and passive data. Active counting has been used more recently than passive counting, and both types of counting utilize crowdsourced data collection methods, specifically, data collected from personal phones, smart devices, and other devices that could provide location information. Active data are data collected from devices actively launched without explicit user consent, such as fitness apps and app-based bike-share programs. Active data are data requested

from individuals (Kuik, 2018) whose explicit consent is needed. This method ensures that data are gathered from those who wish to provide data, thus providing a natural filter that promotes the collection of relevant data (Hub, n.d.). Unfortunately, this method of data collection still requires voluntary input from the general public. To ensure that a complete set of data is collected, passive data collection is often preferred. Passive data are collected from programs that are passively running in the background of the device. These background programs include location services, global positioning systems (GPS), and mobile phone positioning (MPP). In the case of providing pedestrian and bicyclist counts, collecting and utilizing passive data allows for a more thorough understanding of the total volume of pedestrians and bicyclists, as the only major component required is a system to record the location and relative proximity of the devices running the background programs previously mentioned (Duff, n.d.). In addition, passive data could allow for a more instantaneous record of data compared to the manual data input given by active methods (Duff, n.d.). Some studies have already examined the benefits of these methods (Battaglia et al., 2008; Revilla et al., 2017; Keuschet al., 2019).

3. Data Description

To fulfill the previously mentioned research objective, namely, the comprehensive comparison between the traditional permanent counter data and emerging StreetLight data, the present study collected data originating from four separate sources. The various data sources support the different subtasks within the research objective, including checking consistency between SL and permanent counter data, development of the systematic adjustment factor, and estimation of AADT based on hourly volume and other popular predictor variables. The detailed description of each source is presented as follows.

The first set comes from the location-based services (LBS) crowdsourced database for pedestrian and bicyclist counts, StreetLight (SL). StreetLight compiles data collected from smartphones' location-based services, and the database cross-references to satellite locations based on the smartphones that are harnessed as sensors. Once the data are collected and cross-referenced, the product will contain counts for either pedestrians or bicyclists. The authors collected these data by first outlining the specific zones where pedestrian and bicyclist data were available from the Portland State University's Active Transportation Database and the City of San José. Then, the data requested were organized into pedestrian and bicyclist counts, which were further categorized by year, month, weekday, and hour. The final data provided by StreetLight are presented by either SL Index value or calibrated counts, which are obtained based on other sources of counts provided by users (e.g. manually collected counts) and presented in the form of average hourly volume within a given month/year.

The second database is the national archive for bicycle and pedestrian count data maintained by Portland State University (Bike-Ped Archive, 2021). The national archive collects pedestrian and bicyclist count data in a typical way using staffed (personnel), temporary, or permanent counters (e.g., loop detectors) along sidewalks and trails alike. These areas cover key cities in California, Oregon, Washington, and eastern Canada. Given time availability and various roadway facilities covered by the data, the present study focuses on selected permanent counter counts from cities in California, including San José, San Diego, and Imperial Beach, to name a few. Each flow detector has location information in the form of state, city, longitude and latitude, count type (e.g., pedestrian or bicyclist), and functional classification (e.g., a trail detector or a detector along a minor or major arterial). The data from each flow detector are organized into 15-minute intervals. Each interval has start and end time data as well as the count within each interval. These data are then arranged in the same manner as those collected from StreetLight for comparison purposes.

The third database was provided by the City of San José based on a recently purchased permanent bike counter located on the Three Creeks Trail between Coe and Broadway in the Willow Glen neighborhood of San José (see Figure 1). This detector, as shown in Figure 2, collected bicyclist counts in a similar manner as most permanent counters utilized to collect the data compiled by the national archive. This detector was carefully calibrated by the city staff to ensure the data produced

are maximally accurate. The bike counter was purchased and installed during the project period, so it can provide some recent data. Compared with the national archive data, such data from San José can be checked against the SL data quality for more recent periods. Overall, the data are organized in the same way as in the national archive.

Figure 1. Map of the Location of the Permanent Bike Counter in the City of San José

Source: Google Earth Pro

Figure 2. Illustration of the Permanent Bike Counter owned by the City of San José

Once the data from the above three sources were collected, they were compiled to explore the consistency between data from SL and the permanent counters. The compiled data were cleaned first by removing some outliers (in some cases, the SL counts are unreasonably large). After the data cleaning, 6,777 observations are used, with 6,403 for bicyclists and 374 for pedestrians. Detailed descriptive statistics for the cleaned dataset are shown in Table 1.

Finally, weather data from Weather Underground and permanent counter data from the national archive for bicycle and pedestrian count data were collected for the development of various adjustment factors associated with facility type, weather, time (i.e., hour), day, month, and land use, and for the estimation of Average Daily Traffic Volume (AADT) for bicyclists and pedestrians. The national archive organizes data by flow detector, and data associated with each detector report hourly volume counts, when data were recorded, and each permanent counter's exact location. The counter's location can be used to find historical weather data for that location from Weather Underground. Weather Underground provides weather information organized based on date and time, which were then combined with the count data collected from the national archive in terms of date and time. Detailed descriptive statistics for this dataset are shown in Table 2.

Table 1. Descriptive Statistics for the Data used for Comparisons between StreetLight Calibrated Counts and Counter Counts for Bicyclists and Pedestrians

Numerical Variables					
Variables	**Description**	**Minimum**	**Maximum**	**Mean**	**S.D.**
StreetLight Calibrated Counts	Average hourly volume for pedestrian and bicyclist of calibrated streetlight data of specific month and year	0	340	18.70	25.66
Counter Counts	Average hourly volume for pedestrian and bicyclist of permanent counter data of specific month and year	0	222	13.33	18.52
Categorical Variables					
Variables	**Description**	**Details of Categories (frequency, percentage)**			
Year	Year in which data were collected	2018 (2495, 36.81%); 2019 (2547, 37.58%); 2020 (1705, 25.15%); 2021 (30, 0.44%)			
Month	Month in which data were collected	January (449, 6.62%); February (418, 6.17%); March (529, 7.76%); April (495, 7.30%); May (652, 9.62%); June (701, 10.34%); July (761, 11.23%); August (736, 10.86%); September (589, 8.69%); October (495, 7.30%); November (505, 7.45%); December (450, 6.64%)			
Day	Day of the week when the data were collected	Monday (773, 11.40%); Tuesday (874, 12.89%); Wednesday (863, 12.73%); Thursday (905, 13.35%); Friday (944, 13.93%); Saturday (1,292, 19.06%); Sunday (1,126, 16.61%)			
Hour	Hour of the day when the data were collected	12 AM (20, 0.30%); 1 AM (14, 0.21%); 2 AM (10, 0.15%); 3 AM (4, 0.01%); 4 AM (17, 0.25%); 5 AM (32, 0.47%); 6 AM (131, 1.93%); 7 AM (258, 3.81%); 8 AM (347, 5.12%); 9 AM (448, 6.61%); 10 AM (501, 7.39%); 11 AM (511, 7.54%); 12 PM (554, 8.17%); 1 PM (540, 7.97%); 2 PM (573, 8.45%); 3 PM (609, 8.99%); 4 PM (522, 7.70%); 5 PM (507, 7.45%); 6 PM (452, 6.67%); 7 PM (291, 4.29%); 8 PM (203, 3.00%); 9 PM (136, 2.01%); 10 PM (65, 0.96%); 11 PM (32, 0.47%)			
Count Type	Whether data in question are for bicyclists or pedestrians	Bicyclist (6,403, 94.47%); Pedestrian (374, 5.52%)			

Note: Not all hours of the day were considered for the comparison between StreetLight and counter data. Only those hours when the counts were most likely available were included in the study.

Table 2. Descriptive Statistics for the Data used for the Estimation of AADT for Pedestrians and Bicyclists

Numerical Variables					
Variables	Description	Minimum	Maximum	Mean	S.D.
$AADT_P$	Average Annual Daily Traffic for pedestrians	4	95	25.78	31.20
$AADT_B$	Average Annual Daily Traffic for bicyclists	11	84	45.28	24.36
HV_P	Hourly pedestrian volume count	0	55	6.40	9.13
HV_B	Hourly bicyclist volume count	0	36	11.70	8.14
Temp	Hourly average temperature in degrees Fahrenheit	46	91	65.44	8.15
DewPoint	Hourly average dew point in degrees Fahrenheit	17	90	59.12	10.80
$Humidity_{\%}$	Hourly average percent humidity	5	361	61.30	22.54
WindSpeed	Hourly average wind speed in miles per hour	0	19	6.44	3.75
$Pressure_{Hg}$	Hourly average pressure in inches of Mercury	29	29394	61.83	923.82
Categorical Variables					
Variables	**Description**	**Details of Categories (frequency, percentage)**			
Time	Time of day	6 AM (121, 11.92%); 7 AM (100, 9.85%); 8 AM (80, 7.88%); 9 AM (118, 11.63%); 10 AM (82, 8.08%); 11 AM (34, 3.35%); 12 PM (41, 4.04%); 1 PM (21, 2.07%); 2 PM (63, 6.21%); 3 PM (47, 4.63%); 4 PM (63, 6.21%); 5 PM (25, 2.46%); 6 PM (36, 3.55%); 7 PM (51, 5.02%); 8 PM (65, 6.400%); 9 PM (68, 6.70%)			
Month	Month of the year	January (90, 8.87%); February (96, 9.46%); March (93, 9.16%); April (93, 9.16%); May (86, 8.47%); June (93, 9.16%); July (90, 8.87%); August (82, 8.08%); September (79, 7.78%); October (67, 6.60%); November (69, 6.80%); December (77, 7.59%)			
Day	Day of the month	1 (18, 1.77%); 2 (33, 3.25%); 3 (42, 4.14%); 4 (22, 2.16%); 5 (38, 3.74%); 6 (40, 3.94%); 7 (13, 1.28%); 8 (27, 2.66%); 9 (74, 7.29%); 10 (20, 1.97%); 11 (39, 3.84%); 12 (44, 4.34%); 13 (12, 1.18%); 14 (39, 3.84%); 15 (38, 3.74%); 16 (42,			

Numerical Variables					
Variables	Description	Minimum	Maximum	Mean	S.D.
		4.14%); 17 (11, 1.08%); 18 (37, 3.65%); 19 (57, 5.62%); 20 (33, 3.25%); 21 (16, 1.58%); 22 (28, 2.76%); 23 (61, 6.01%); 24 (12, 1.18%); 25 (28, 2.76%); 26 (67, 6.60%); 27 (10, 0.99%); 28 (37, 3.65%); 29 (32, 3.15%); 30 (43, 4.24%); 31 (2, 0.20%)			
Facility Type	Type of facility where data were collected	Major Collector (159, 15.67%); Minor Arterial (281, 27.68%); Principal Arterial – Other (160, 15.76%); Trail or Shared Use Path (415, 40.89%)			
Land Use	Land use type based on location of count detector	Commercial (156, 15.37%); Recreation (734, 72.32%); Transportation (125, 12.32%)			

4. Methodology

As previously mentioned, the primary goals of this study are (a) to understand the consistency between crowdsourced data collected from StreetLight and permanent counter data (particularly from the national archive and the City of San José) and (b) to generate various adjustment factors to convert count data in between different types efficiently. This is accomplished through the use of a set of statistical tools.

4.1 Consistency Checking between StreetLight and Permanent Counter Data

The first goal of this study is to comprehend the consistency between the calibrated SL and permanent counter counts. The statistical tools used to facilitate such a task are outlined as follows.

4.1.1 Statistical Difference

T-Test

The first approach to determining the statistical difference is a two-sample t-test (Statistics Solutions, 2021). Due to the unequal variances between the two datasets, it is necessary to implement Welch's t-test (Welch, 1947). Welch's t-test generates a t-value (t) and the degrees of freedom (df) in Equation 1 and Equation 2.

$$t = \frac{m_b - m_a}{\sqrt{\frac{S_b^2}{n_b} + \frac{S_a^2}{n_a}}} \tag{1}$$

$$df = \left(\frac{S_b^2}{n_b} + \frac{S_a^2}{n_a}\right) \Bigg/ \left(\frac{S_b^2}{n_b^2(n_b - 1)} + \frac{S_a^2}{n_a^2(n_a - 1)}\right) \tag{2}$$

In the above equations, m_b and m_a represent the sample means, S_b^2 and S_a^2 represent the variances of the two types of count data, respectively, and n_a and n_b are the associated sample sizes. In the present study, n_a is equal to n_b since the count data were generated for comparison purposes for the same locations and periods. Note that for the results from Welch's t-test to be effective, the degrees of freedom between the two sets of data must be greater than 5 (Allwood, 2008), which is the case for the present study (see Table 3).

Wilcoxon Test

The second method used to determine the statistical difference is the Wilcoxon signed-rank test (Hayes, 2021). The Wilcoxon test operates in the same manner as the t-test. However, it is more attuned to non-parametric data (Conover, 1999). This test can be calculated with Equation 3.

$$V = \sum_{i=1}^{N} \left[sgn(x_{2,i} - x_{1,i}) * R_i \right] \tag{3}$$

Here, R_i is the rank of the pair as calculated by the position of observation in an ordered list of $|x_{2,i} - x_{1,i}|$, with x repressing the counts. The subscript "i" denotes the observation ID, and the subscripts "1" and "2" denote SL and permanent counter data, respectively.

4.1.2 Linear Association

Linear Regression R-Squared

R-squared is a popular method used to calculate the linear association of two variables, and the value is obtained from the linear regression (Frost, 2021) and calculated as the ratio between the sum of squares of residuals and the total sum of squares:

$$SS_{res} = \sum_{i} (y_i - f_i)^2 \tag{4}$$

$$SS_{tot} = \sum_{i} (y_i - \bar{y})^2 \tag{5}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

where SS_{res} is the sum of squares of residuals, y_i is the individual y value for a given "i" observation, f_i is the result of the equation of the line of best fit for a given x value, SS_{tot} is the total sum of squares with $\bar{y}$ representing the average y-values of the line of best fit, and R^2 is the R-squared value (Steel, 1960).

Pearson's Correlation Coefficient

Pearson's correlation coefficient is another measurement of linear correlation between two datasets. In short, Pearson's correlation coefficient is the ratio between the covariance of two variables and the product of their respective standard deviations, resulting in a value between -1 (negatively

correlated) and +1 (positively correlated) (Glen, 2017). Pearson's correlation coefficient can be calculated using Equation 7:

$$r = \frac{\sum(x - m_x)\left(y - m_y\right)}{\sqrt{\sum(x - m_x)^2 \sum\left(y - m_y\right)^2}} \tag{7}$$

where m_x and m_y are the means of the x and y variables, respectively.

4.1.3. Ordinal Association

In addition to the linear association that assumes a normal distribution of the concerned variables, the study also explores the ordinal association between the two variables free from any assumptions of the data distribution.

Spearman Correlation

Charles Spearman proposed Spearman's rank correlation coefficient in 1904. This method is similar to Pearson's correlation. However, unlike Pearson, which assesses the linear relationship between variables, Spearman considers the ordinal relationship within two sets of data (Lehman, 2005). Spearman's correlation coefficient can be calculated with Equation 8.

$$\rho = \frac{\sum(x' - m_{x'})\left(y_i' - m_{y'}\right)}{\sqrt{\sum(x' - m_{x'})^2 \sum\left(y' - m_{y'}\right)^2}} \tag{8}$$

Here, m represents the mean, and x' and y' represent the ranks of x and y, respectively.

Kendall's Tau

Kendall's rank correlation coefficient, also known as Kendall's τ, is another popular metric used to measure the ordinal association between two datasets. Developed in 1938 by Maurice Kendall, this rank correlation will produce values between +1 (positively correlated) and -1 (negatively correlated) to show the correlation between two variables. This index is achieved by first calculating the total number of concordant pairs and discordant pairs. The following equations outline how Kendall's τ is calculated:

$$n_c = num(y_j > y_i) \tag{9}$$

$$n_d = num(y_j < y_i) \tag{10}$$

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \tag{11}$$

where $num()$ is the function used to count the observations satisfying specific conditions, and n represents the associated counts.

4.2 Systematic Adjustment Factor Development between StreetLight and Permanent Counter Data

The above tools are mainly used to check the consistency between the SL and permanent counter data. Once it is determined that these two types of data are not perfectly consistent with each other, the professionals may need some adjustment factor that enables an estimation of the permanent counter data based on the SL data, and vice versa. The modeling tool employed in the present study is the Integrated Nested Laplace Approximation (INLA) method, which acts as an alternative to the Markov Chain Monte Carlo (MCMC) approach that has previously served as a standard procedure for Gaussian distribution models due to INLA's capacity to handle complex model structures based on the simulation method. The INLA method is a Bayesian hierarchical framework that utilizes Laplace approximation to estimate the parameters following into the Gaussian Markov Random Field (GMRF), which can significantly reduce the computation time while having approximately the same level of accuracy (Martino & Rue, 2007). Hence, the INLA method was selected for the study due to its faster computation and ease of use for greater model complexity.

For the development of a count model that allows us to develop the adjustment factors between the two types of counts, the pedestrian or bicyclist counts of certain locations or observational units are usually assumed to follow a Poisson distribution (Singh et al., 2021). In the present study, two types of models are used for more reliable and accurate estimates for adjustment factors: fixed intercept model and random intercept model. The models with better evaluation performance will be used to develop the corresponding factors. Under the fixed intercept model, the calibrated SL hourly volume of pedestrians and bicyclists y_i can be expressed as follows:

$$y_i \sim Poisson\ (\lambda_i) \tag{12}$$

$$\ln(\lambda_i) = \beta_0 + offset(\ln(X_i)) \tag{13}$$

where subscript i represents an observation (hourly count), λ_i is the corresponding rate, β_0 represents a global intercept, and X is the independent variable representing the permanent counter counts. The fixed global intercept assumes all observations follow the same base condition. We employ $offset(\ln(X_i))$ to ensure there is no model-generated coefficient for $\ln(X_i)$, and hence force y_i and X_i to have the final relationship expressed as follows:

$$y_i = \alpha * X_i \tag{14}$$

where α is the adjustment factor that is calculated as:

$$\alpha = \exp(\beta_0). \tag{15}$$

Under the random intercept model, all equations mentioned above remain the same except the following expressions:

$$\ln(\lambda_i) = \beta_{0i} + offset(\ln(X_i)) \tag{16}$$

$$\beta_{0i} = \beta_0 + \varepsilon_i \tag{17}$$

where β_{0i} is the random intercept, which consists of the fixed global intercept β_0 and the white noises ε_i used to capture the unobserved heterogeneity associated with each observation.

4.3 Estimation of Annual Average Daily Traffic based on Hourly Volume and Other Predictor Variables

The hourly counts for pedestrians and bicyclists can be easily obtained via all kinds of collection methods (that is, manual collection, permanent counter, or LBS). However, such hourly counts can only represent the traffic conditions for a short time period, and they cannot display a reliable picture of traffic patterns for a long time period: the average over time is usually represented by the annual average daily traffic (AADT). Hence, in addition to the systematic adjustment factor between SL and permanent counter data, practitioners may also be interested in various adjustment factors for AADT of pedestrians and bicyclists, which can be utilized to adjust the estimated AADT based on different conditions such as hour, day, month, land use, facility type, etc. To accommodate such a need, the study also collected pertinent data to develop models for estimating pedestrian and bicyclist AADT. Therefore, in addition to the response variable of AADT, the data also contained predictor variables for hourly volume, time (i.e., hour), day, month, land use, facility type, and weather conditions (temperature, dew point, wind speed, pressure, and humidity).

With a large number of covariates, numerous combinations of diverse variables of interest are possible. For illustration purposes, the present study selected three sets of independent variables to develop models with varying levels of complexity. As with the systematic adjustment factor development between SL and permanent counter data, the AADTs are also assumed to follow the Poisson distribution. These models are formulated using the following expressions:

$$AADT_i \sim Poisson(\lambda_i) \tag{18}$$

$$\ln(\lambda_i) = \beta_0 + HV_i + Time_i \tag{19}$$

$$\ln(\lambda_i) = \beta_0 + HV_i + Time_i + Facility\_Type_i \quad (20)$$

$$\ln(\lambda_i) = \beta_0 + HV_i + Time_i + Facility\_Type_i + Temp_i + DewPoint_i + WindSpeed_i + Pressure_i + Humidity_i \quad (21)$$

Note that i, λ_i, and β_0 are the same as shown in previous equations, HV_i is the hourly volume, $Time_i$ is the time (or hours), $Facility\_Type_i$ is the type of facility, $Temp_i$ is the temperature (°F), $DewPoint_i$ is the dew point (°F), $WindSpeed_i$ is the wind speed (in mph), pressure is the air pressure (in Hg), and $Humidity_i$ is the amount of water in the air in relation to the maximum amount of water vapor (in %).

The predictor variables vary from model to model, allowing the practitioner to estimate the AADT based on the HV, considering various variables of interest.

4.4 Validation

Several criteria are used to determine the effectiveness and goodness of fit of Bayesian models (Muthukumarana & Tiwari, 2016). This study utilizes the mean absolute difference (MAD) to select models for the development of systematic adjustment factors. For models aiming to develop other adjustment factors, two criteria—Deviance Information Criterion (DIC) and Log Pseudo Marginal Likelihood (LPML)—are used to assess model performance, where the former is based on in-sample data and the latter is based on out-of-sample data.

4.4.1 Mean Absolute Difference (MAD)

MAD is a popular index used to evaluate model performance based on the discrepancy between the predicted and actual calibrated counts. In the present study, the MAD was calculated as follows:

$$\text{MAD} = \frac{1}{\text{n}}\sum_{\text{i=1}}^{\text{n}} |Y_i - O_i| \quad (22)$$

Y_i is the Bayesian-estimated SL calibrated count of observation i by a model, and O_i is the observed SL count of the same observations. The smaller the value, the better the concerned model tends to perform.

4.4.2 Deviance Information Criterion (DIC)

Developed in 2002, DIC is an informal method used to determine the effectiveness of models in explaining the observed data while also maintaining accuracy with new data (Spiegelhalter et al., 2002). DIC uses the following equation to assign an effectiveness rating to each model generated:

$$DIC = D(\bar{\theta}) + 2P_D \tag{23}$$

The posterior mean of the deviance is represented by $D(\bar{\theta})$. The DIC value also has a penalty applied for the greater model complexity that tends to over-fit the interest data. This penalty is represented by $2P_D$, where P_D is the effective number of parameters. Knowing this, the larger the DIC value, the less effective the model is (Spiegelhalter et al., 2002).

4.4.3 Log Pseudo Marginal Likelihood (LPML)

LPML was developed based on the Conditional Predictive Ordinate (CPO) (Geisser & Eddy, 1979). Compared with the previous criteria of MAD and DIC, which are mainly based on in-sample data, LMPL relies on cross-validation based on out-of-sample (OOS) data (in other words, the data used for model development are separated from those reserved for model validation). Due to this special property, LMPL has been used in a number of fields of study since its development (Cheng et at., 2020). The CPO is calculated as:

$$CPO_i = \int f(y_i|\theta, x_i)\pi\left(\theta|D^{(-i)}\right)d\theta \tag{24}$$

where: θ is the unknown parameter of interest; y_i and x_i are the response and covariate vectors, respectively; $D^{(-i)}$ is the dataset without the i^{th} observation; and $\pi(\theta|D^{(-i)})$ is the posterior density of θ based on data $D^{(-i)}$. Using CPO_i, LPML can then be calculated as follows:

$$LPML = \sum_{i=1}^{n} \log(CPO_i) \tag{24}$$

Contrary to DIC, LMPL uses larger values to represent models with higher prediction capabilities. In addition, since LMPL performs cross-validation using OOS data (based on the leave-one-out technique where n iterations are implemented with one data record being held out for validation in each iteration), there is no penalty dedicated to the complexity of models generated.

5. Model Results

The present study aims to check the consistency between the StreetLight (SL) counts and those collected from the permanent counters from various cities and to develop an adjustment factor that can be used to estimate one type of count based on the other type.. In addition, models for the estimation of AADT for pedestrians and bicyclists are developed based on a set of influential factors. The detailed results for each objective are presented in order in the following sections.

5.1 Consistency Checking between StreetLight and Permanent Counter Data

The subsection demonstrates the difference between LBS-based (or SL) active transportation counts and the permanent counter counts using distinct types of statistical techniques (parametric vs. non-parametric) and assumptions (linear vs. non-linear).

5.1.1 Statistical Difference

The most straightforward way to check the consistency between the two types of counts is to explore the discrepancy of the count magnitude. Both parametric (paired t-test) and non-parametric (Wilcoxon test) approaches are employed to assess the difference.

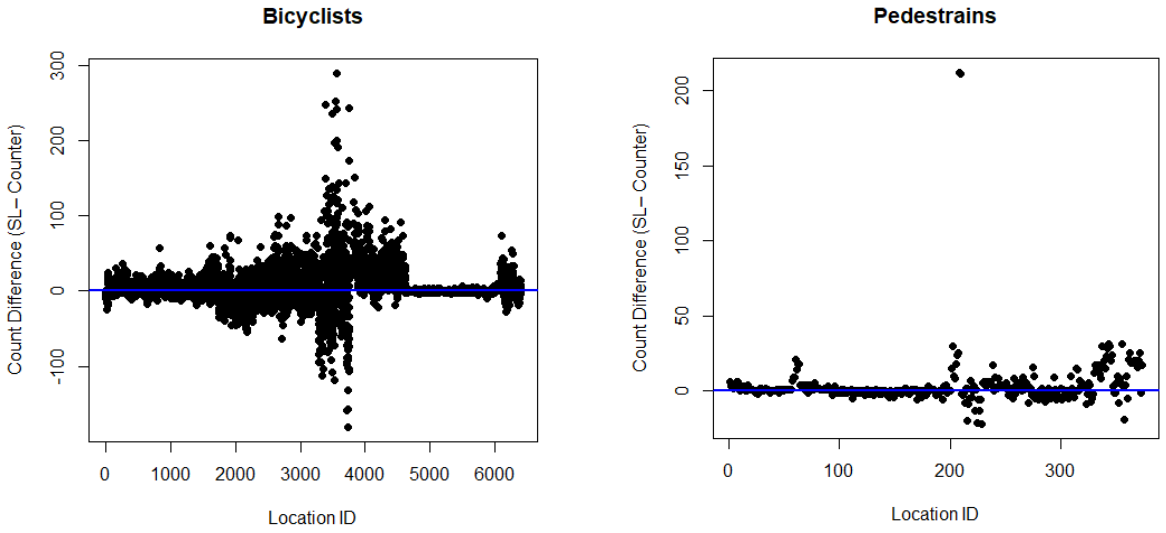
Table 3. Paired T-Test Results between StreetLight Calibrated Counts and Counter Counts for Bicyclists and Pedestrians

	t	df	p-value	95% CI	Mean of Difference
Bicyclists	19.296	6,402	< 2.2e-16	[4.903, 6.012]	5.457
Pedestrians	4.197	373	3.383e-05	[1.965, 5.430]	3.698

Notes: df is degree of freedom; CI is the confidence interval; t is defined in Equation 1.

Consulting Table 3, we see that the SL and counter counts are statistically different for both types of non-motorized modes (pedestrian and bicyclist). However, the means of difference are relatively small. Figure 3 further illustrates the count difference between the two types of data sources in a visual format. For bicyclists, the difference ranges from -200 to 300, with the greater values following into the location IDs between 3,000 and 4,000. The differences are much smaller for pedestrians, with two points showing a value of more than 200. Given the smaller overall means of difference, even with the existence of a few locations where the count differences are proportionally large, it can be concluded that LBS-based (or SL) active transportation counts could serve as an efficient alternative when the counts from the permanent counter are not available. However, it is also noteworthy that the smaller means of difference may result from removing some data outliers before the t-test was performed.

Figure 3. Plot of Difference between StreetLight and Counter Counts for Bicyclists and Pedestrians



Note: Points below the blue line indicate observations where SL values are greater than counter counts, where SL-Counter is negative.

Like the parametric paired t-test, the non-parametric paired Wilcoxon signed-rank test reveals the statistically significant difference between the two types of counts based on the p-values for pedestrians and bicyclists, as shown in Table 4. The V-value for bicyclists is much larger than for pedestrians, as the sample size for the former (6,403) is about 17 times that of the latter (374).

Table 4. Paired Wilcoxon Signed-Rank Test Results between StreetLight and Counter Counts for Bicyclists and Pedestrians

	V	p-value
Bicyclists	10,649,226	< 2.2e-16
Pedestrians	31,627	2.91e-10

Note: V is defined in Equation 3.

5.1.2 Linear Association

In addition to the magnitude difference checking, the similarity between SL and permanent counters counts can be evaluated via assessing the linear association in between.

Linear Regression R-Squared Value

The first popular linear association index is the coefficient of determination (i.e., R-squared value) of a linear model, in this case used to compare the two count types. As shown in Table 5, the R-

squared value for both modes of counts are 0.2765 and 0.2547, respectively. The R^2 value indicates the ratio of the variation explained by the predictor variables to the total variations among the response variable. Usually, the more predictors are included in the model, the larger the R^2 value tends to be. Since the linear model contains only one independent variable (the counter count), it can be concluded that the SL and counter counts demonstrate a notable linear association.

Table 5. Results of Linear Association between StreetLight and Counter Counts for Bicyclists and Pedestrians

Modes	R-squared of simple linear regression	Pearson's correlation coefficient
Bicyclists	0.2765	**0.5259 [0.5079, 0.5472]**
Pedestrians	0.2547	**0.5047 [0.4251, 0.5766]**

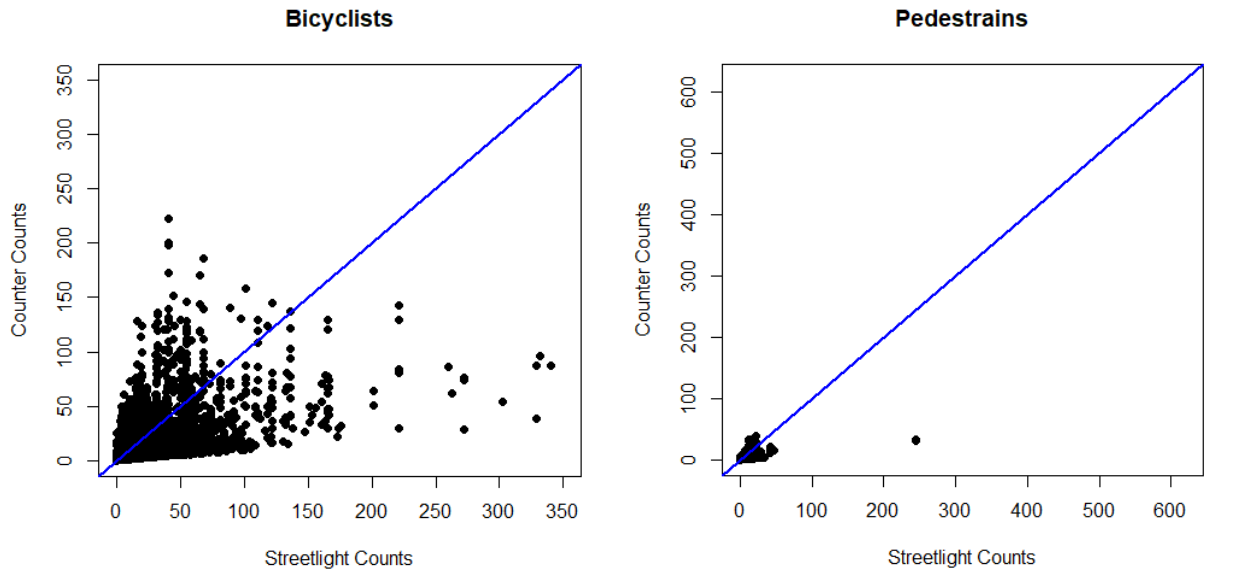
Notes: 1. The numbers in the square bracket represent the 95% confidence level for the correlation coefficient. 2. The bold font indicates the statistical significance at the level of 0.05.

Pearson's Correlation Coefficient

In addition to the R-squared value, Pearson's correlation coefficient is another prevalent measure used to assess the linear association of the variables of interest. It is essentially the R-squared value based on the standardized independent and dependent variables. Table 5 shows that the two types of counts are statistically positively correlated with somewhat larger coefficient values, 0.5259 and 0.5047.

The graphical correlation between the two types of counts is shown in Figure 4, where it is evident that the positive correlations are exhibited in both panels.

Figure 4. Plot of StreetLight vs. Counter Counts for Pedestrians and Bicyclists



Note: The blue line indicates the reference line that bisects SL index values and counter counts.

5.1.3 Ordinal Association

The previous results reveal that SL and counter counts have a notable positive linear correlation. However, these two counts may be related through a non-linear function as well. In this case, the above-mentioned Pearson's correlation coefficient would be low even with a strong association between the variables. Therefore, the study also employed two rank-based correlation approaches, Spearman correlation and Kendall's τ, to properly identify the association between counts from different collection methods, assuming a non-linear relationship in between.

Table 6. Results of Ordinal Association between StreetLight and Counter Counts for Bicyclists and Pedestrians

Modes	Spearman correlation coefficient (ρ)	Kendall correlation coefficient (τ)
Bicyclists	**0.691** (< 2.2e-16)	**0.517** (< 2.2e-16)
Pedestrians	**0.716** (< 2.2e-16)	**0.546** (< 2.2e-16)

Notes: 1. The numbers in parentheses represent the p-values for the correlation coefficients. 2. The black font indicates the statistical significance at the level of 0.05. 3. See the Methodology section for the definition of rho and tau.

As shown in Table 6, there appear to be statistically significant positive correlations for both pedestrians and bicyclists. Moreover, under the non-linear assumption, the Spearman correlation coefficient values are greater than the Pearson's correlation ones. Again, such results demonstrate the great consistency between the SL and counter counts.

The Kendall correlation also measures the non-linear association between the two numerical variables. Once more, the tau values are statistically significantly positive, indicating the strong association between the types of counts.

5.2 Systematic Adjustment Factor Development between StreetLight and Permanent Counter Data

The above results reveal the notable level of consistency between the two types of counts (SL and counter). Nonetheless, these values are still far from being equivalent to each other. For example, the maximum correlation coefficient is 0.716 among all situations (linear vs. non-linear, bicyclist vs. pedestrian). The t-test reveals that the two count types are statistically significantly different. Systematic error may be the culprit. First, the SL data are based on "pings" from cellular devices. Therefore, the non-motorist will not be counted if the mobile devices are not carried. Second, in some congested situations, the traveling speeds are very close among the vehicles and the active transportation mode users. Such phenomenon significantly increases the difficulty of mode classification. Therefore, it is imperative to adjust the SL counts before entirely replacing the counts from counters.

To this end, INLA Bayesian models were developed for both active transportation modes, assuming a Poisson distribution for the SL counts. The detailed model results are presented in Table 7. Based on the evaluation criteria, including DIC, $\overline{D}$, P_D, and LPML, the random intercept model appears to perform better than the fixed intercept one for pedestrians and bicyclists. Moreover, the statistically significant random effects indicate the necessity of including the random intercepts to capture the unobserved heterogeneity among all observations.

Table 7. Posterior Model Parameter Estimates for the Development of System Adjustment Factors between StreetLight and Counter Counts

	Bicyclists		Pedestrians	
Variables	Fixed Intercept Model	Random Intercept Model	Variables	Fixed Intercept Model
	Mean (2.5%, 97.5%)	Mean (2.5%, 97.5%)	Mean (2.5%, 97.5%)	Mean (2.5%, 97.5%)
Intercept	**0.333 (0.328, 0.339)**	**0.282 (0.258, 0.306)**	**0.514 (0.480, 0.547)**	**0.384 (0.290, 0.477)**
Random Effect	NA	**0.042 (-1.364, 1.713)**	NA	**0.049 (-0.980, 1.517)**
DIC	124377.9	37070.7	4082.1	1844.8
$\overline{D}$	124376.1	31891.6	4081.1	1591.7
P_D	1.8	5179.1	1.0	253.1
LPML	-62217.6	-26058.2	-2044.3	-133.2

Notes: 1. NA means not applicable. 2. The bolded cells represent the variables with statistical significance at the level of 0.05.

Even though the random intercept model demonstrates superior performance using the popular Bayesian-related evaluation criterion, both models will need to adjust the SL counts to counter counts. The reason is that the white noises (ε_i) in the random intercept model have to be dropped when developing the generalized adjustment equations that cannot include any random components. Given that the main goal is to get the adjusted SL (or predicted counter count) closer to the counter counts, the mean absolute difference (MAD) between the predicted and collected counter counts is employed to determine more accurate adjustment equations based on the model estimates as shown in the above table. The adjustment equations leading to the smaller MAD values are retained for the recommended adjustment equations.

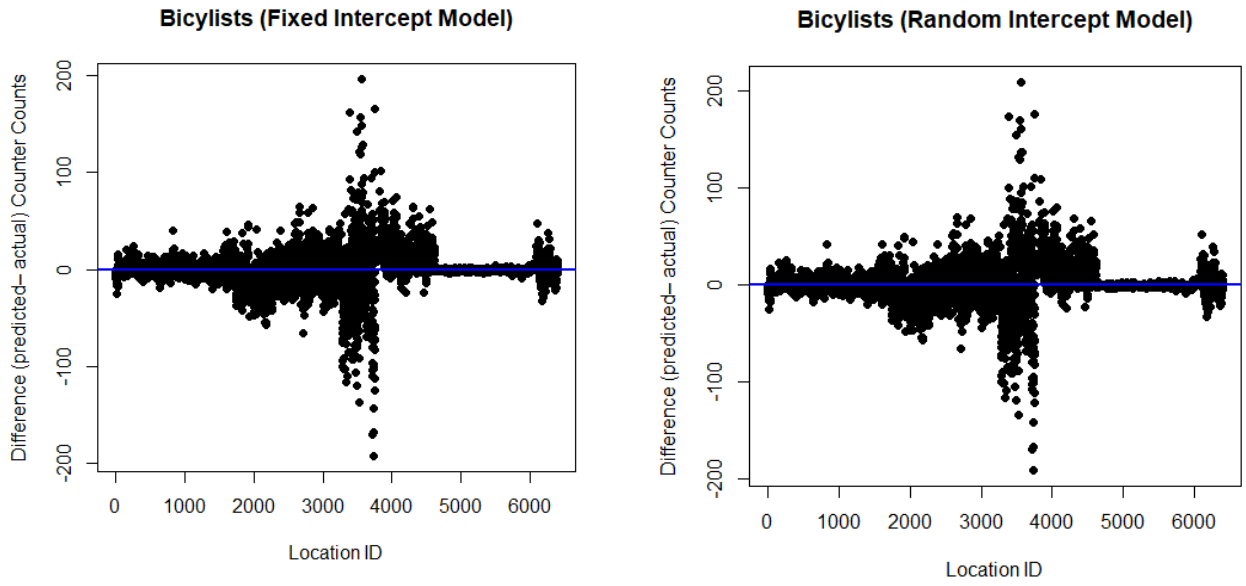
From Table 8, it is interesting to see that fixed intercept models yield better adjustment equations for both active transportation modes. This phenomenon indicates the superiority of the model performance (in terms of goodness-of-fit) may not be transferred to the adjustment purpose even though the heterogeneity can be considered in the random intercept models.

Table 8. The Adjustment Equations between StreetLight Index Values and StreetLight and Counter Counts for Bicyclists and Pedestrians

Model Types	MAD	Adjustment Equations
		Bicyclists
Fixed Intercept Model	9.8	$\boldsymbol{SL = \exp(0.333 + \ln(Counter)) = 1.395 * CC}$
Random Intercept Model	10.1	$SL = \exp(0.282 + \ln(Counter)) = 1.326 * CC$
		Pedestrians
Fixed Intercept Model	4.1	$\boldsymbol{SL = \exp(0.514 + \ln(Counter)) = 1.672 * CC}$
Random Intercept Model	4.3	$SL = \exp(0.384 + \ln(Counter)) = 1.468 * CC$

Notes: 1. CC means counter counts. 2. The bold cells represent the recommended adjustment equations based on MAD values. 3. MAD is the mean absolute difference.

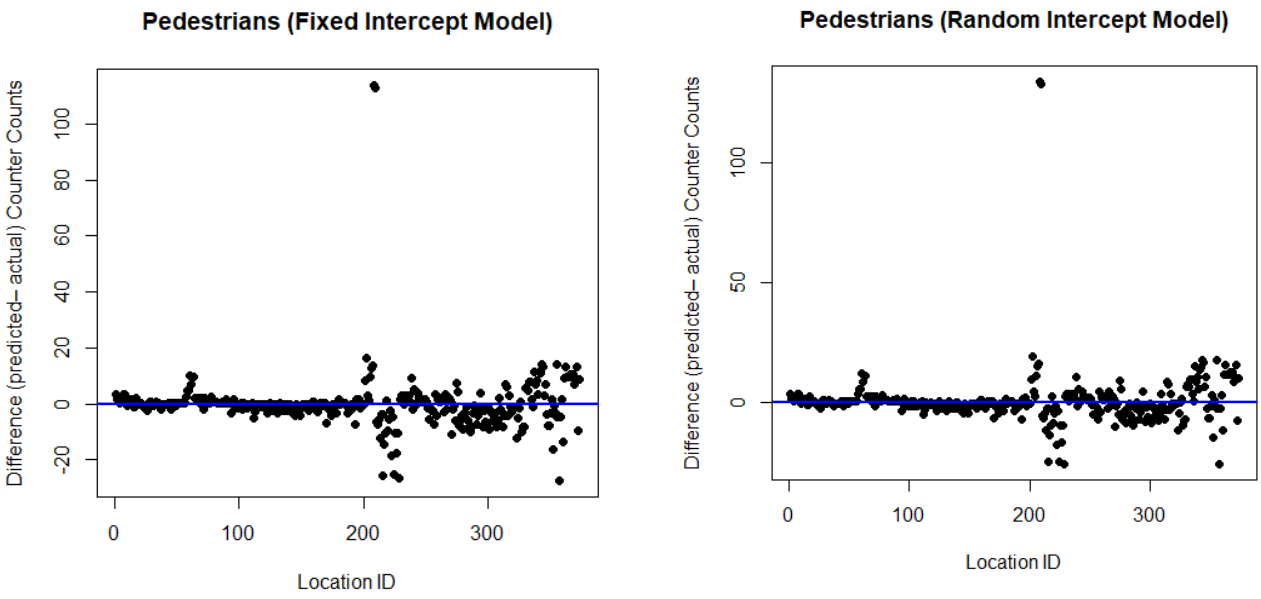
Figure 5. Plot of Difference between Predicted and Collected Counter Counts for Bicyclists



Note: Points below the blue line indicate observations where actual counter values are greater than predicted counter counts, where (predicted-actual) is negative.

In addition to the MAD values, the graphical illustration of the difference between the adjusted SL and counter counts is presented in Figure 5 for bicyclists and Figure 6 for pedestrians. Comparing the two figures with Figure 3 clearly demonstrates that the difference between the adjusted SL and counter counts is much smaller than the difference between the original SL and counter counts for all different situations (that is, pedestrian and bicyclist, as well as fixed and random intercept models). Such results imply the importance of adjusting the SL counts before they can be used for different purposes.

Figure 6. Plot of Difference between Predicted and Collected Counter Counts for Pedestrians



Note: Points below the blue line indicate observations where actual counter values are greater than predicted counter counts, where (predicted-actual) is negative.

5.3 Other Adjustment Factors based on Permanent Counter Data

Aside from the necessity for systematic adjustment in between the two count types, practitioners may also wish for models that enable the estimation of the AADT for pedestrians and bicyclists based on the hourly volume (HV) at specific time periods, while the adjustments for various conditions are allowed at the same time. To accommodate this need, the Bayesian INLA Poisson models are used to develop the models linking the AADT and HV, with standard influential variables being considered including the time (hours), day, month, land use, facility type, and weather conditions (temperature, dew point, wind speed, pressure, and humidity). With the availability of these models, practitioners can better estimate the AADT of the non-motorist counts based on the HVs for specific time periods, which are cheaper to collect compared with AADT that requires data lasting at least one year. Hence, these models are truly desirable, especially for those agencies with limited human resources. Since there are numerous possible conditions with the different combinations of these input variables, the project developed three different models for pedestrians and bicyclists, representing the different levels of model complexity and data availability. The detailed model results, containing formula, model performance, base conditions, and adjustment factors are exhibited in Tables 8–13. Thus, practitioners can select the most suitable models based on their specific data availability.

Table 9. Illustration of Prediction of AADT for Pedestrian based on Pedestrian HV and Time

Formula: $\boldsymbol{AADT_P = 23.760 * e^{0.069*HV_P} * k_{TIME}}$				
Model Performance: DIC=22042.8; Dbar=22025.6; pD=17.2; LPML=-11240.6				
Base Condition: Time= 6am				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.925;	8am:0.743;	9am:0.503;	10am:0.371;	11am: 0.278
12pm:0.208;	1pm:0.337;	2pm:0.385;	3pm:0.352;	4pm: 0.369
5pm:0.296;	6pm:0.454;	7pm:0.495;	8pm:0.739;	9pm:0.943
Numerical Example: For one location, assume the HV_P collected at 8 am is 5. Estimate $AADT_P$. **Solution:** $\boldsymbol{AADT_P = 23.760 * e^{0.069*HV_P} * k_{TIME} = 23.760 * e^{0.069*5} * 0.743 \approx 25}$				

Table 10. Illustration of Prediction of AADT for Pedestrian based on Pedestrian HV, Time, and Facility Type

Formula: $\boldsymbol{AADT_P = 23.547 * e^{0.042*HV_P} * k_{TIME} * k_{FACILITY}}$				
Model Performance: DIC=12469.8; Dbar=12449.2; pD=20.2; LPML=-6352.5				
Base Conditions: Time= 6am; Facility_Type=Major Collector				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.887;	8am:0.788;	9am:0.612;	10am:0.533;	11am: 0.345
12pm:0.347;	1pm:0.417;	2pm:0.538;	3pm:0.460;	4pm: 0.538
5pm:0.389;	6pm:0.525;	7pm:0.562;	8pm:0.748;	9pm:0.931
Adjustment Factors for Time ($\boldsymbol{k_{FACILITY}}$):				
Minor Arterial: 2.246;	Other Principal Arterial:1.054;	Trail or Shared Use Path:0.281;		
Numerical Example: For one location, assume the HV_P collected at 10 am is 6, and the facility type is other principal arterials. Estimate $AADT_P$. Solution: $\boldsymbol{AADT_P = 23.547 * e^{0.042*HV_P} * k_{TIME} * k_{FACILITY} = 23.547 * e^{0.042*6} * 0.533 * 1.054 \approx 17}$				

Table 11. Illustration of Prediction of AADT for Pedestrian based on Pedestrian HV, Time, Facility Type, and Various Weather Variables

Formula: $\boldsymbol{AADT_P = 39.330 * e^{0.030HV_P} * e^{-0.072Temp} * e^{0.068Dew_{Piont}} * e^{-0.003Humidity_{\%}} * e^{0.008Wind_{Speed}} * e^{0.000Pressure\_Hg} * k_{TIME} * k_{FACILITY}}$				
Model Performance: DIC=10640.8; Dbar=10615.6; pD=25.2; LPML=-5279.0				
Base Conditions: Time= 6am; Facility_Type=Major Collector				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.927;	8am:0.855;	9am:0.727;	10am:0.629;	11am: 0.467
12pm:0.532;	1pm:0.641;	2pm:0.670;	3pm:0.667;	4pm: 0.674
5pm:0.540;	6pm:0.660;	7pm:0.667;	8pm:0.856;	9pm:0.956
Adjustment Factors for Time ($\boldsymbol{k_{FACILITY}}$):				
Minor Arterial: 2.622;	Other Principal Arterial:1.078;	Trail or Shared Use Path:0.359;		
Numerical Example: For one location, assume the HV_P collected at 1 pm is 8, and the temperature, dew point, humidity, wind speed, and pressure are 51°F, 49°F, 53%, 5 mph, and 30.12Hg, respectively, while the facility type is other principal arterials. Estimate $AADT_P$. Solution: $\boldsymbol{AADT_P = 39.330 * e^{0.030HV_P} * e^{-0.072Temp} * e^{0.068Dew_{Point}} * e^{-0.003Humidity_{\%}} * e^{0.008Wind_{Speed}} * e^{0.000Pressure_{Hg}} * k_{TIME} * k_{FACILITY} = 39.33 * e^{0.030*8} * e^{-0.072*51} * e^{0.068*49} * e^{-0.003*53} * e^{0.008*5} * e^{0*30.12} * 0.641 * 1.078 \approx 22}$				

Table 12. Illustration of Prediction of AADT for Bicyclist based on Bicyclist HV and Time

Formula: $\boldsymbol{AADT_B = 41.763 * e^{0.051*HV_B} * k_{TIME}}$				
Model Performance: DIC=15067.8; Dbar=15050.5; pD=17.3; LPML=-7610.8				
Base Condition: Time= 6am				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.893;	8am:0.643;	9am:0.497;	10am:0.458;	11am: 0.427
12pm:0.402;	1pm:0.325;	2pm:0.385;	3pm:0.369;	4pm: 0.387
5pm:0.411;	6pm:0.473;	7pm:0.471;	8pm:0.671;	9pm:0.902
Numerical Example: For one location, assume the HV_B collected at 8 am is 10. Estimate $AADT_B$. Solution: $\boldsymbol{AADT_B = 41.763 * e^{0.051*HV_B} * k_{TIME} = 41.763 * e^{0.051*10} * 0.643 \approx 45}$				

Table 13. Illustration of Prediction of AADT for Bicyclist based on Bicyclist HV, Time, and Facility Type

Formula: $\boldsymbol{AADT_B = 77.634 * e^{0.025*HV_B} * k_{TIME} * k_{FACILITY}}$				
Model Performance: DIC=7412.5; Dbar=7392.2; pD=20.3; LPML=-3709.8				
Base Conditions: Time= 6am; Facility_Type=Major Collector				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.933;	8am:0.797;	9am:0.691;	10am:0.642;	11am: 0.621
12pm:0.646;	1pm:0.586;	2pm:0.645;	3pm:0.615;	4pm: 0.630
5pm:0.623;	6pm:0.662;	7pm:0.681;	8pm:0.811;	9pm:0.912
Adjustment Factors for Time ($\boldsymbol{k_{FACILITY}}$):				
Minor Arterial: 0.373;	Other Principal Arterial:0.877;	Trail or Shared Use Path:0.405;		
Numerical Example: For one location, assume the HV_B collected at 10 am is 8, and the facility type is other principal arterials. Estimate $AADT_B$. Solution: $\boldsymbol{AADT_B = 77.634 * e^{0.025*HV_B} * k_{TIME} * k_{FACILITY} = 77.634 * e^{0.025*8} * 0.642 * 0.877 \approx 53}$				

Table 14. Illustration of Prediction of AADT for Bicyclist based on Bicyclist HV, Time, Facility Type, and Various Weather Variables

Formula: $\boldsymbol{AADT_B = 79.758 * e^{0.022HV_B} * e^{-0.011Temp} * e^{0.012Dew_{Piont}} * e^{-0.001Humidity_{\%}} * e^{0.003Wind_{Speed}} * e^{0.000Pressure\_Hg} * k_{TIME} * k_{FACILITY}}$				
Model Performance: DIC=7210.2; Dbar=7184.9; pD=25.3; LPML=-3608.6				
Base Conditions: Time= 6am; Facility_Type=Major Collector				
Adjustment Factors for Time ($\boldsymbol{k_{TIME}}$):				
7am: 0.947;	8am:0.824;	9am:0.731;	10am:0.676;	11am: 0.658
12pm:0.698;	1pm:0.675;	2pm:0.682;	3pm:0.663;	4pm: 0.672
5pm:0.653;	6pm:0.703;	7pm:0.718;	8pm:0.834;	9pm:0.933
Adjustment Factors for Time ($\boldsymbol{k_{FACILITY}}$):				
Minor Arterial: 0.386;	Other Principal Arterial:0.886;	Trail or Shared Use Path:0.444;		
Numerical Example: For one location, assume the HV_B collected at 1 pm is 8, and the temperature, dew point, humidity, wind speed, and pressure are 51°F, 49°F, 53%, 5 mph, and 30.12 Hg, respectively, while the facility type is other principal arterials. Estimate $AADT_B$. **Solution:** $\boldsymbol{AADT_B = 79.758 * e^{0.022HV_B} * e^{-0.011Temp} * e^{0.012Dew_{Point}} * e^{-0.001Humidity_{\%}} * e^{0.003Wind_{Speed}} * e^{0.000Pressure_{Hg}} * k_{TIME} * k_{FACILITY} = 79.758 * e^{0.022*8} * e^{-0.011*51} * e^{0.012*49} * e^{-0.001*53} * e^{0.003*5} * e^{0*30.12} * 0.675 * 0.886 \approx 56}$				

The above examples demonstrate the application of the proposed models, under different conditions, to develop a systematic adjustment factor between SL and permanent counter data and estimate the active-transportation-related AADT based on hourly counts and other predictor variables related to weather, land use, time, day, and so on. Despite their ease of use, these statistical models represent a small proportion of tools available to fulfill the same goals. The practitioner can also choose their preferred statistical models (such as the typical non-Bayesian ones) and/or machine learning algorithms (e.g., artificial neural networks, tree-based models, deep learning methods) and detect the best-performing ones based on their specific data.

6. Conclusions and Recommendations

Active transportation has become more popular due to the numerous health, environmental, and economic benefits it provides. With the rise in popularity comes the need for adequate facilities to accommodate those who wish to walk or ride a bike for transportation instead of using a motor vehicle. In an effort to plan for the design and construction of proper facilities (i.e., sidewalks, bike lanes, trails, etc.), it is necessary to understand the respective volumes of transportation by mode. These volumes can be collected through several methods. However, most of them rely on typical approaches, such as manual counts by personnel or automatic counts by loop detectors or other devices. In recent years, crowdsourced data is becoming popular due to the potential ease of data collection of large areas compared to traditional methods. Yet there is a lack of comprehensive comparison between the crowdsourced data and those counts obtained from the typical methods. To this end, this project collected crowdsourced data from StreetLight and permanent counter data from the national archive maintained by Portland State University and the City of San José. To determine whether SL data are viable for determining non-motorized counts, a consistency analysis comprising statistical differences, linear association, and an ordinal association was performed to comprehend the statistical similarities and differences between the SL data and counter data. In addition, R-INLA, a package oriented in Integrated Nested Laplace Approximation, is used to generate both fixed and random effect models that would help adjust the SL to the typical counter data by accounting for the systematic error that may be associated with the SL data. Finally, a set of models was developed to demonstrate how to estimate the AADT data based on the easily collected hourly volumes with different influential factors being included. Subsequent to the distinct analyses, the following conclusions and recommendations can be made.

1. The SL count data for pedestrians and bicyclists appear to be a viable alternative to the permanent counters under different evaluation methods when the data outliers were removed (see point #2 below). The former demonstrates a notable consistency with the latter from different perspectives, including statistical difference, linear association, and non-linear association.

2. However, there are some caveats to the above conclusion. First, the results were obtained by removing data outliers satisfying the arbitrary criteria established by the authors. Second, a large volume of the more recent permanent counter data (e.g., most of the bicycle count data for the City of San José) was removed as outliers due to unreasonably large SL count values for these observations. Hence, it is highly recommended that SL data be carefully examined for accuracy given the somewhat new nature of this kind of data. For example, practitioners could use their prior experience or knowledge to determine whether the magnitudes of the SL counts are reasonable for certain types of active transportation facilities at certain locations.

3. The discrepancy between SL and counter data is much smaller after the SL data are adjusted by applying the developed adjustment factors using the different count models. Therefore, it is highly recommended to adjust the SL data before its use due to the systematic error associated with the SL data, which are reliant on cellular devices.

4. For agencies wishing for the AADT of the non-motorist counts but subject to limited resources, they may refer to the models developed in the study that link AADT with an hourly volume of specific time periods, or develop their preferred models for their particular needs. Such models can estimate the AADT based on the hourly volume, which is much easier and more economical to collect than AADT, which requires continuous data availability for a longer time period, say, at least one year. The estimated AADT can also be adjusted based on different conditions such as hours, month, weekday, temperature, land use, facility type, etc.

The abovementioned findings demonstrate that crowdsourced SL data are promising and could serve as a viable alternative to the conventional data sources. Even though the SL data accuracy still has some room for improvement compared with the permanent counter data, the former collection method can collect the large-scale data in a more economical way, making it a truly appealing method especially when a large volume of data with a certain level of accuracy is needed. This is exciting considering the other enormous benefits of crowdsourced data relative to the typical methods. However, it is essential to note that the results are presented along with some cautionary notes.

First, the data collected are based on active transportation counts from various cities in California only. More data from other locations or locations are needed for more reliable findings of the accuracy of the SL data, whose performance was explored based on the removal of certain data outliers satisfying somewhat arbitrary criteria established by the authors. Different criteria for identifying outliers may yield totally different results. Second, the study utilized the permanent counter data as the benchmark and focused on checking on the consistency of SL data with the former alternative. Due to time and resource limitations, the authors assume the permanent counter data collected from the national archive database and the City of San José correctly represent the real-world situations. Improperly calibrated or maintained permanent counters may lead to misleading results, as shown in the report. Third, there are numerous reports showing the successful application of machine learning methods to predict and/or classify different categorical instances, which is also used by StreetLight to distinguish the different transportation modes when providing the count data; however, some special situations may diminish the mode differences, making the classification of modes based on cellular data almost impossible. For instance, some serious runners prefer pavement to sidewalk facilities and may be mistaken for bicyclists. Likewise, in congested situations, bicyclists in the lane may be mistaken for vehicles. All these circumstances would complicate the mode classification and lead to additional biases.

Bibliography

Active Transportation Strategic Plan (ATSP). (n.d.). https://www.metro.net/projects/active-transportation-strategic-plan-atsp/

Allwood, M. (2008). AP® Statistics.

Anderson, R. L. (1970). Electromagnetic loop vehicle detectors. *IEEE Transactions on Vehicular Technology*, 19(1), 23–30.

Battaglia, J., Zollo, A., Virieux, J., & Iacono, D. D. (2008). Merging active and passive data sets in traveltime tomography: The case study of Campi Flegrei caldera (Southern Italy). *Geophysical Prospecting*, 56(4), 555–573.

Bike-Ped Archive. http://bikeped.trec.pdx.edu/bp/. (March 2021).

CDC. (n.d.). Healthy places. Centers for Disease Control and Prevention. https://www.cdc.gov/healthyplaces/transportation/promote_strategy.htm

Cheng, W., Gill, G. S., Zhou, J., Ensch, J. L., Kwong, J., & Jia, X. (2020). Alternative multivariate multimodal crash frequency models. *Journal of Transportation Safety & Security*, 12(5), 628–652.

Coifman, B. (2001). Improved velocity estimation using single loop detectors. *Transportation Research Part A: Policy and Practice*, 35(10), 863–880.

Conover, W. J. (1999). Practical nonparametric statistics (Vol. 350). John Wiley & Sons.

Conrow, L., Wentz, E., Nelson, T., & Pettit, C. (2018). Comparing spatial patterns of crowdsourced and conventional bicycling datasets. *Applied Geography*, 92, 21–30.

DoT. (2015, August 24). Active transportation. U.S. Department of Transportation. https://www.transportation.gov/mission/health/active-transportation#:~:text=Benefits%20of%20active%20transportation,as%20diabetes%20and%20cardiovascular%20disease

Duff, C. (n.d.). Passive Data Collection vs. Observational Research - The Difference Defined. Retrieved September 14, 2021, from https://blog.flexmr.net/passive-data-collection-vs-observational-research.

Fries, R., Chowdhury, M., & Ma, Y. (2007). Accelerated incident detection and verification: A benefit to cost analysis of traffic cameras. *Journal of Intelligent Transportation Systems*, 11(4), 191–203.

Frost, J. (2021, August 25). How to interpret R-squared in regression analysis. Statistics By Jim. Retrieved September 22, 2021, from https://statisticsbyjim.com/regression/interpret-r-squared-regression/.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365), 153–160.

Glen, S. (2017). Correlation coefficient: Simple definition, formula, easy steps. StatisticsHowTo.com. Retrieved August 3, 2020, from https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/.

Han, B., Yu, X., & Kwon, E. (2009). A self-sensing carbon nanotube/cement composite for traffic monitoring. *Nanotechnology*, 20(44), 445501.

Hayes, A. (2021, May 19). How the Wilcoxon test is used. Investopedia. Retrieved September 22, 2021, from https://www.investopedia.com/terms/w/wilcoxon-test.asp.

Hong, A. (2018). Environmental Benefits of Active Transportation. In *Children's Active Transportation* (pp. 21–38). Elsevier.

Hub, T. E. M. (n.d.). Why impressions won't cut it anymore: The importance of active data capture for live marketing. Limelight Platform. Retrieved September 14, 2021, from https://www.limelightplatform.com/blog/active-data-capture.

John, W. S., & Johnson, P. (2000). The pros and cons of data analysis software for qualitative research. *Journal of Nursing Scholarship*, 32(4), 393–397.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, 83(S1), 210–235.

Kuik, K. (2018, September 6). Active data vs Passive data | Lab1 insights. Retrieved September 13, 2021, from https://www.lab.one/insights/active-data-vs-passive-data.

Lehman, A. (2005). JMP for basic univariate and multivariate statistics: A step-by-step guide. SAS Institute.

Lindblad, B. (n.d.). More choices, less traffic. Climate Resolve. https://www.climateresolve.org/more-choices-less-traffic/.

Litman, T. (2015). Evaluating active transport benefits and costs: Guide to valuing walking and cycling improvements and encouragement programs (pp. 134–140). Victoria Transport Policy Institute.

Martino, S., & Rue, H. (2009). Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the INLA program. Department of Mathematical Sciences, NTNU, Norway.

Muñoz-Pichardo, J. M., Lozano-Aguilera, E. D., Pascual-Acosta, A., & Muñoz-Reyes, A. M. (2021). Multiple Ordinal Correlation Based on Kendall's Tau Measure: A Proposal. *Mathematics*, 9(14), 1616.

Muthukumarana, S., & Tiwari, R. C. (2016). Meta-analysis using Dirichlet process. *Statistical Methods in Medical Research*, 25(1), 352–365.

Portland State University. (n.d.). Portland State University. Transportation Data | Transportation Research and Education Center. Retrieved September 22, 2021, from https://trec.pdx.edu/transportation-data-research.

Revilla, M., Ochoa, C., & Loewe, G. (2017). Using passive data from a meter to complement survey data in order to study online behavior. *Social Science Computer Review*, 35(4), 521–536.

Schlossberg, M., Evers, C., Kato, K., & Brehm, C. (2012). Active Transportation, Citizen Engagement and Livability: Coupling Citizens and Smartphones to Make the Change. *Journal of the Urban & Regional Information Systems Association*, 24(2).

Singh, M., Cheng, W., Samuelson, D., Kwong, J., Li, B., Cao, M., & Li, Y. (2021). Development of pedestrian-and vehicle-related safety performance functions using Bayesian bivariate hierarchical models with mode-specific covariates. *Journal of Safety Research*.

Somasundaram, G., Morellas, V., & Papanikolopoulos, N. (2009, October). Counting pedestrians and bicycles in traffic scenes. In 12th International IEEE Conference on Intelligent Transportation Systems (pp. 1–6). IEEE.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.

Statistics Solutions. (2021, August 2). Paired sample T-test. Statistics Solutions. Retrieved September 22, 2021, from https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/paired-sample-t-test/.

Steel, R. G. (1960). Principles and procedures of statistics: With special reference to the biological sciences (No. 04; QA276, S82).

StreetLight Data. (2021, June 15). Why streetlight: Our data. Retrieved September 17, 2021, from https://www.streetlightdata.com/our-data/.

Toth, C., Suh, W., Elango, V., Sadana, R., Guin, A., Hunter, M., & Guensler, R. (2013). Tablet-based traffic counting application designed to minimize human error. *Transportation Research Record*, 2339(1), 39–46.

Uke, N., & Thool, R. (2013). Moving vehicle detection for measuring traffic count using opencv. *Journal of Automation and Control Engineering*, 1(4).

Welch, B. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28–35. doi:10.2307/2332510

Appendix

Weighted Box Plots of Variables Related to the Built Environment (Figures 7–12)

Weighted box plots are box plots that represent the number of observations within each category by the width of the box. The upper and lower sides of each box represent the third and first quartiles of the data, respectively. The horizontal line within each box represents the average value. The whiskers that extend past the box indicate one standard deviation past the first or third quartile. Any point visible past the end of the whisker is considered an outlier, however, for this study, these outliers were the result of post-filtered data; therefore, despite these values being considered outliers, they were still part of the analysis and results outlined above.

Figure 7. Box Plot of StreetLight Data and Year Distribution

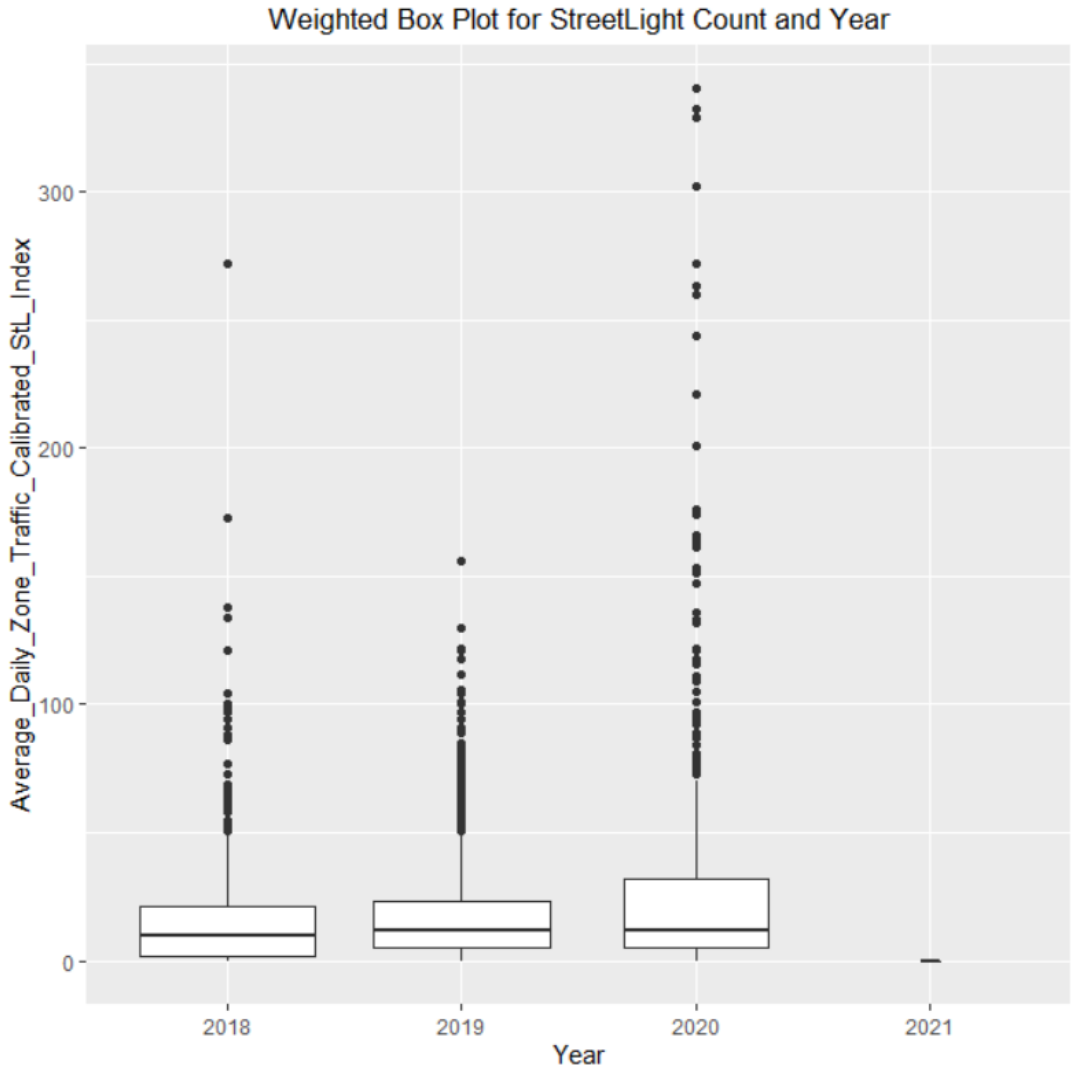


Figure 8. Box Plot of Portland Count Data and Year Distribution

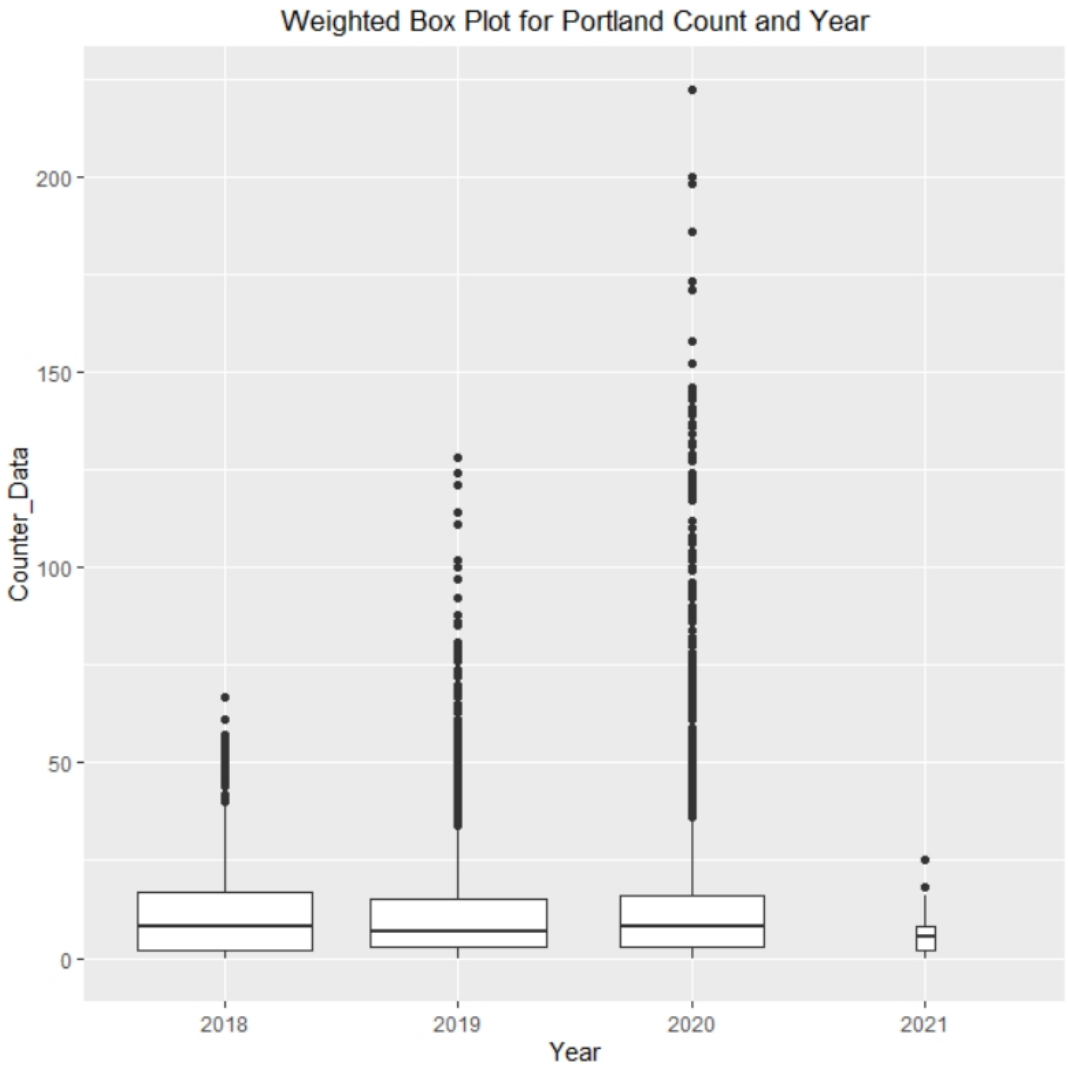


Figure 9. Box Plot of StreetLight Data and Month Distribution

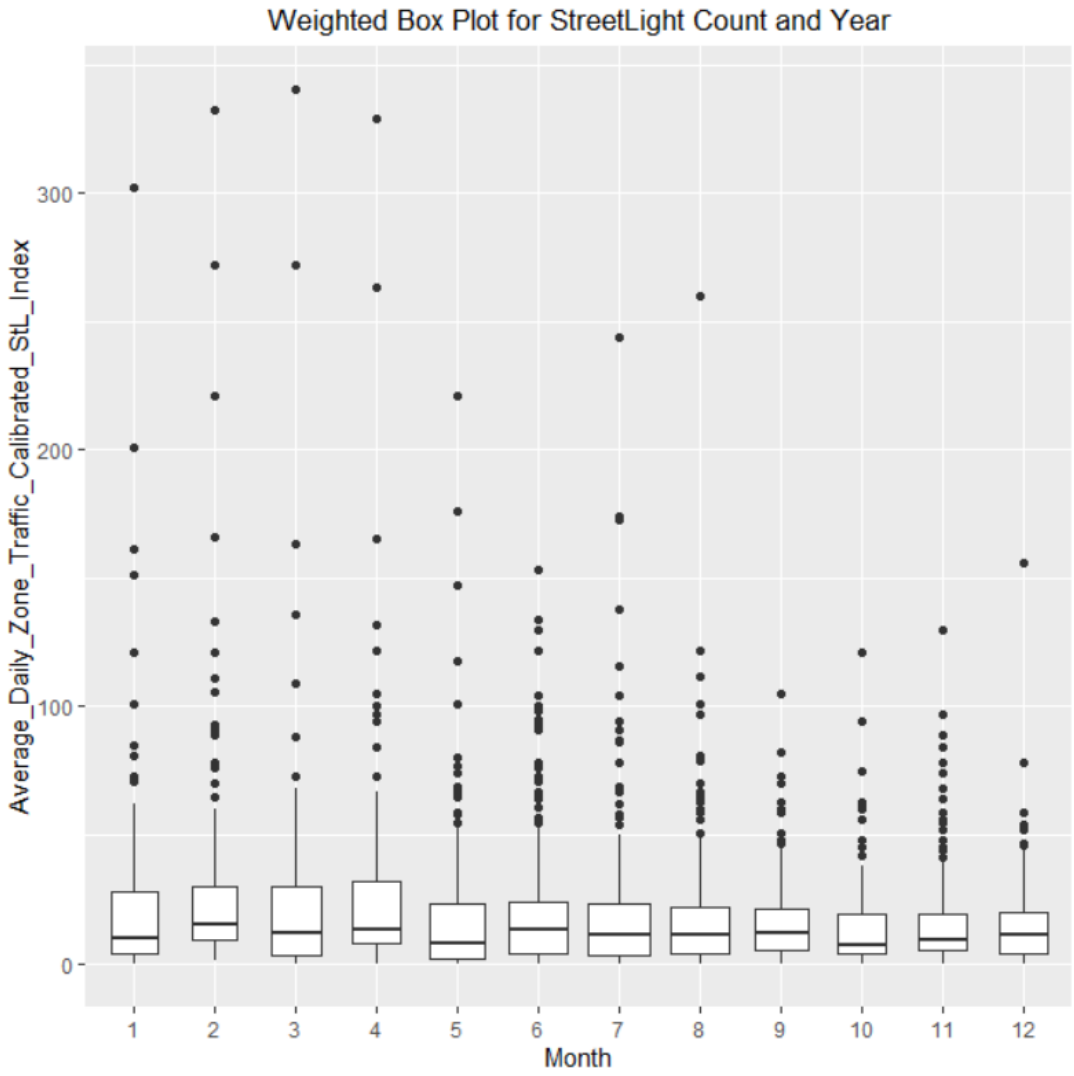


Figure 10. Box Plot of Portland Count Data and Month Distribution

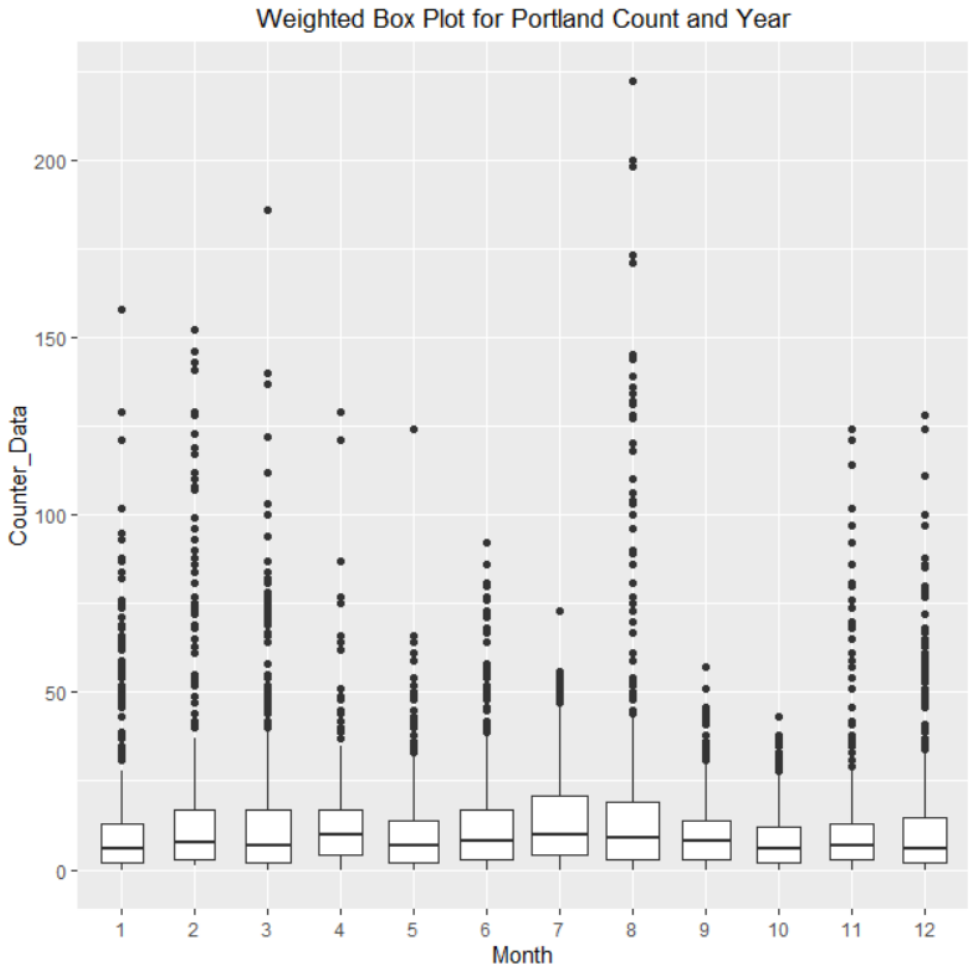


Figure 11. Box Plot of StreetLight Data and Day Distribution

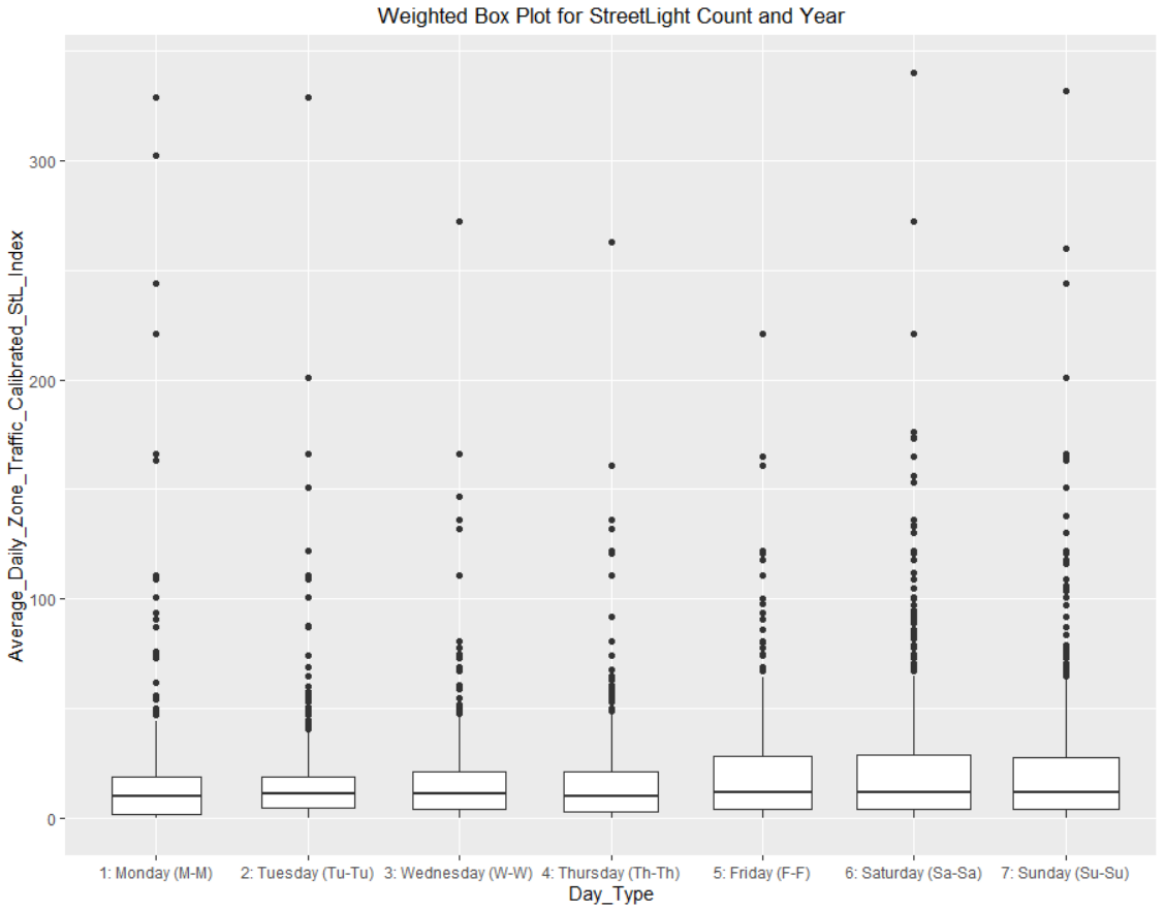


Figure 12. Box Plot of Portland Count Data and Day Distribution

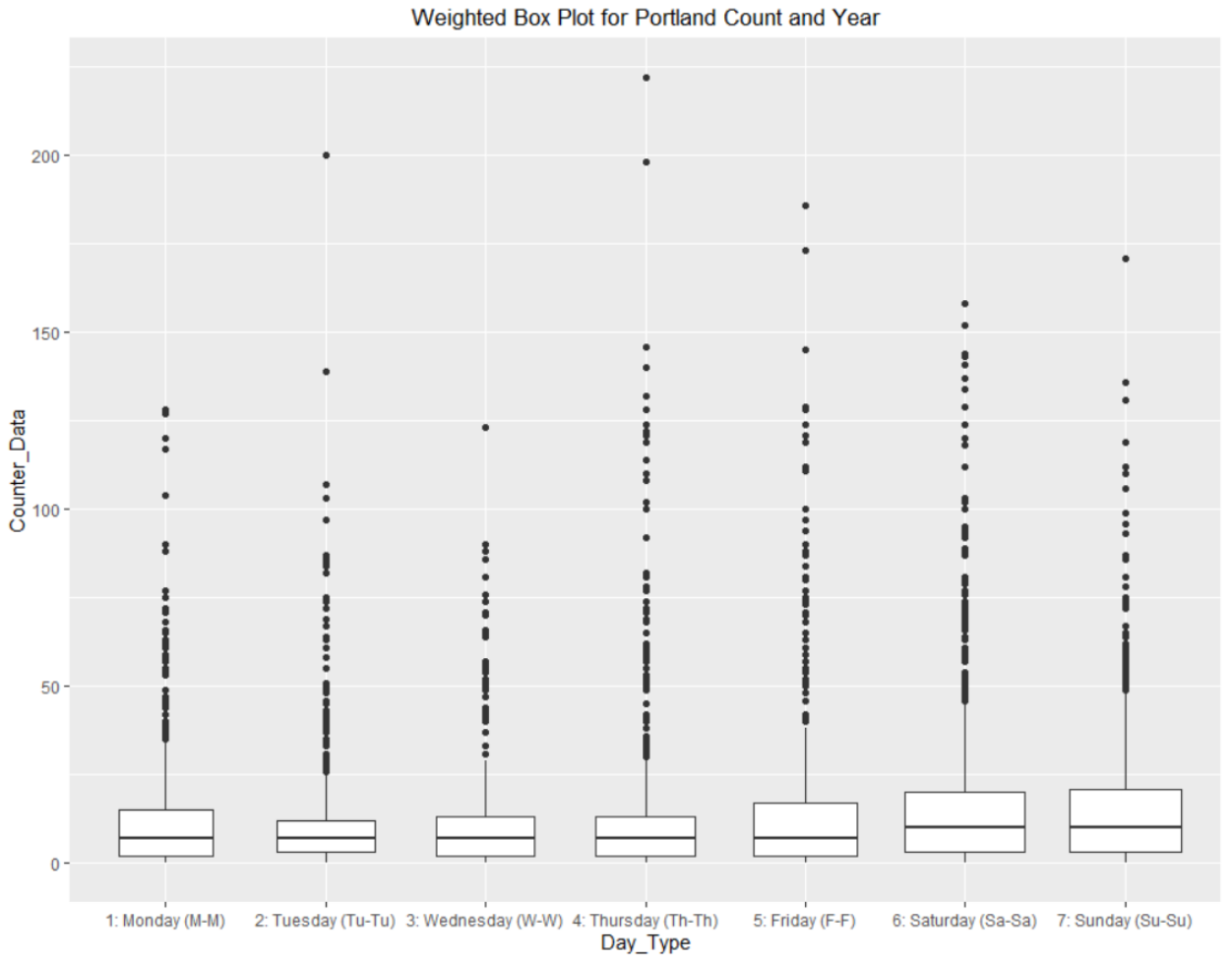


Table 15. List of Variables Present in Dataset Collected from Portland State University

Variables available from Portland State University Active Transportation Database			
segment_area_id	color	directions	long
segment_name	color_type	bicycle	lat
state	buffer	pedestrian	detector_id
city	overpass	equestrian	org_id
tmg_type_id	underpass	off_road	detector_description
facility_id	sharrows	motor_vehicles	detector_make
facility_description	bike_rte_signs	other_flow_type	detector_model
paved	bike_boulevard	flow_detector_id	detector_automated
side	intersection	flowdetector_startdate	functional_classification
facility_width	flow_id	flowdetector_enddate	organization_name

Figure 13. Generation of Zone in StreetLight

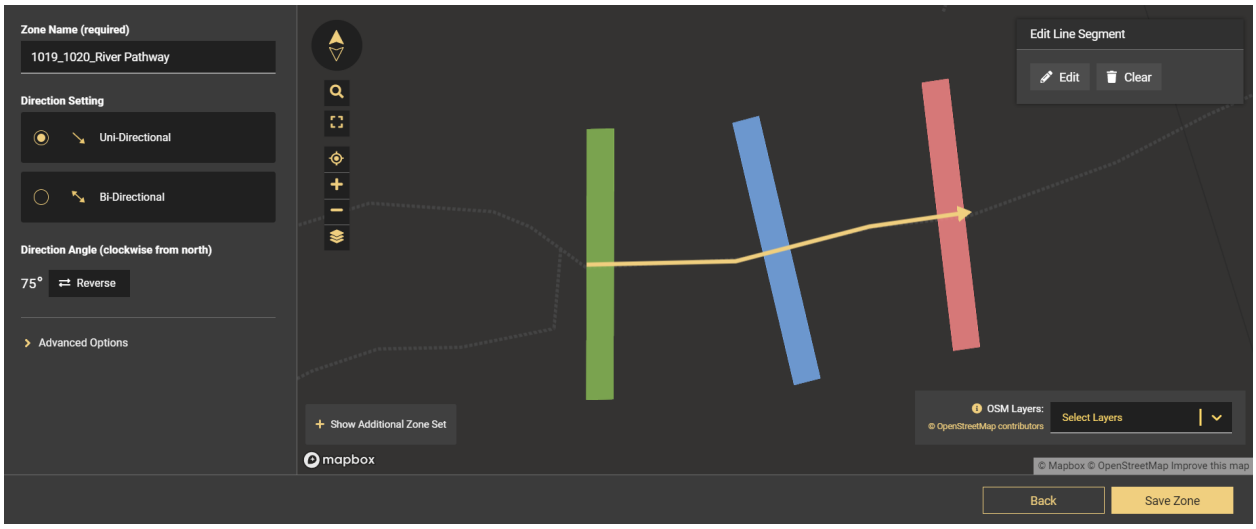


Figure 14. Generation of Calibration Zone in StreetLight

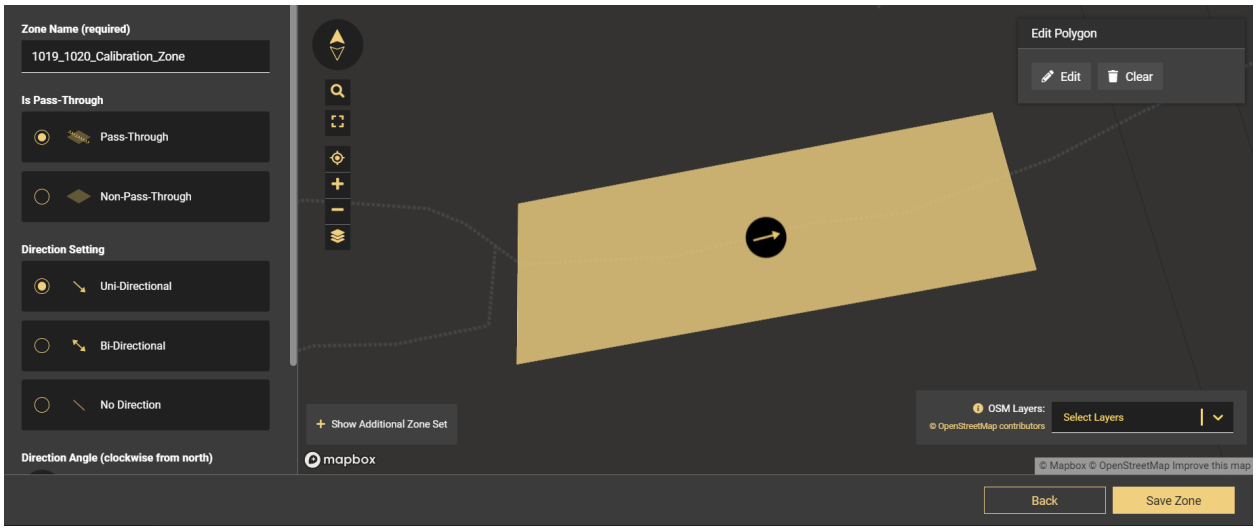


Figure 15. Generation of Analysis After Attaching Zone and Calibration Zone

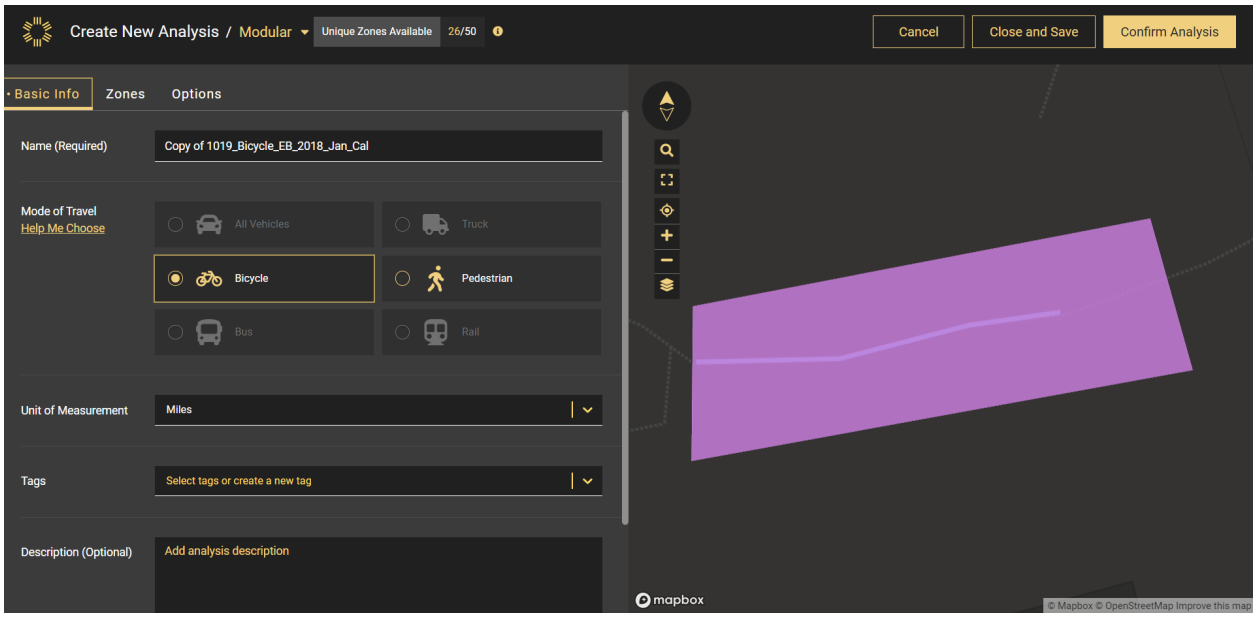


Figure 16. Example of Data Exported by StreetLight Analysis

Mode of T	Intersectic	Zone ID	Zone Nam	Zone Is Pa	Zone Direc	Zone is Bi-	Day Type	Day Part	Average D	Avg Trip D	Avg All Trip	Avg Trip Le	Avg All Trip L
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	00: All Day	29	2765	2765	8.2	8.2
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	07: 6am (6	4	2292	2292	4.7	4.7
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	12: 11am (	4	7959	7959	30.5	30.5
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	14: 1pm (1	4	2826	2826	8	8
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	15: 2pm (2	4	2160	2160	4	4
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	16: 3pm (3	7	2040	2040	6.8	6.8
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	18: 5pm (5	4	928	928	1.6	1.6
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	0: All Days	21: 8pm (8	4	1873	1873	3.3	3.3
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	1: Monday	00: All Day	22	1873	1873	3.3	3.3
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	1: Monday	21: 8pm (8	22	1873	1873	3.3	3.3
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	2: Tuesday	00: All Day	45	5126	5126	17.6	17.6
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	2: Tuesday	07: 6am (6	22	2292	2292	4.7	4.7
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	2: Tuesday	12: 11am (	22	7959	7959	30.5	30.5
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	3: Wednes	00: All Day	22	928	928	1.6	1.6
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	3: Wednes	18: 5pm (5	22	928	928	1.6	1.6
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	5: Friday (F	00: All Day	28	386	386	1.2	1.2
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	5: Friday (F	16: 3pm (3	28	386	386	1.2	1.2
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	6: Saturda	00: All Day	56	2493	2493	6	6
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	6: Saturda	14: 1pm (1	28	2826	2826	8	8
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	6: Saturda	15: 2pm (2	28	2160	2160	4	4
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	7: Sunday (	00: All Day	28	3694	3694	12.3	12.3
Bicycle - St	Trip Pass-Through		1019_102(	yes	75	no	7: Sunday (	16: 3pm (3	28	3694	3694	12.3	12.3

Figure 17. Average Annual Daily Traffic Pedestrian Count Data Distribution Plot According to Location

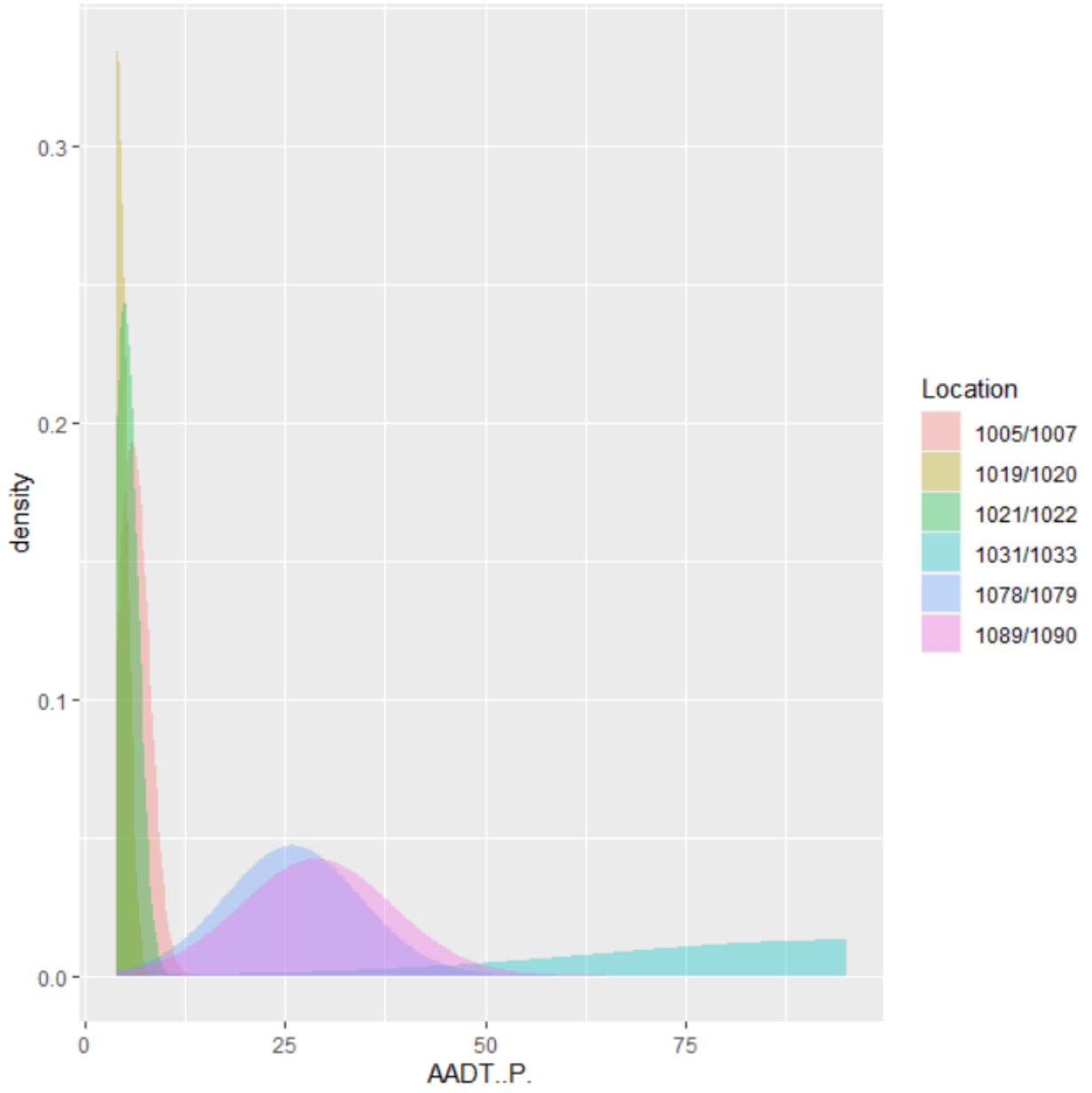


Figure 18. Average Annual Daily Traffic Bicyclist Count Data Distribution Plot According to Location

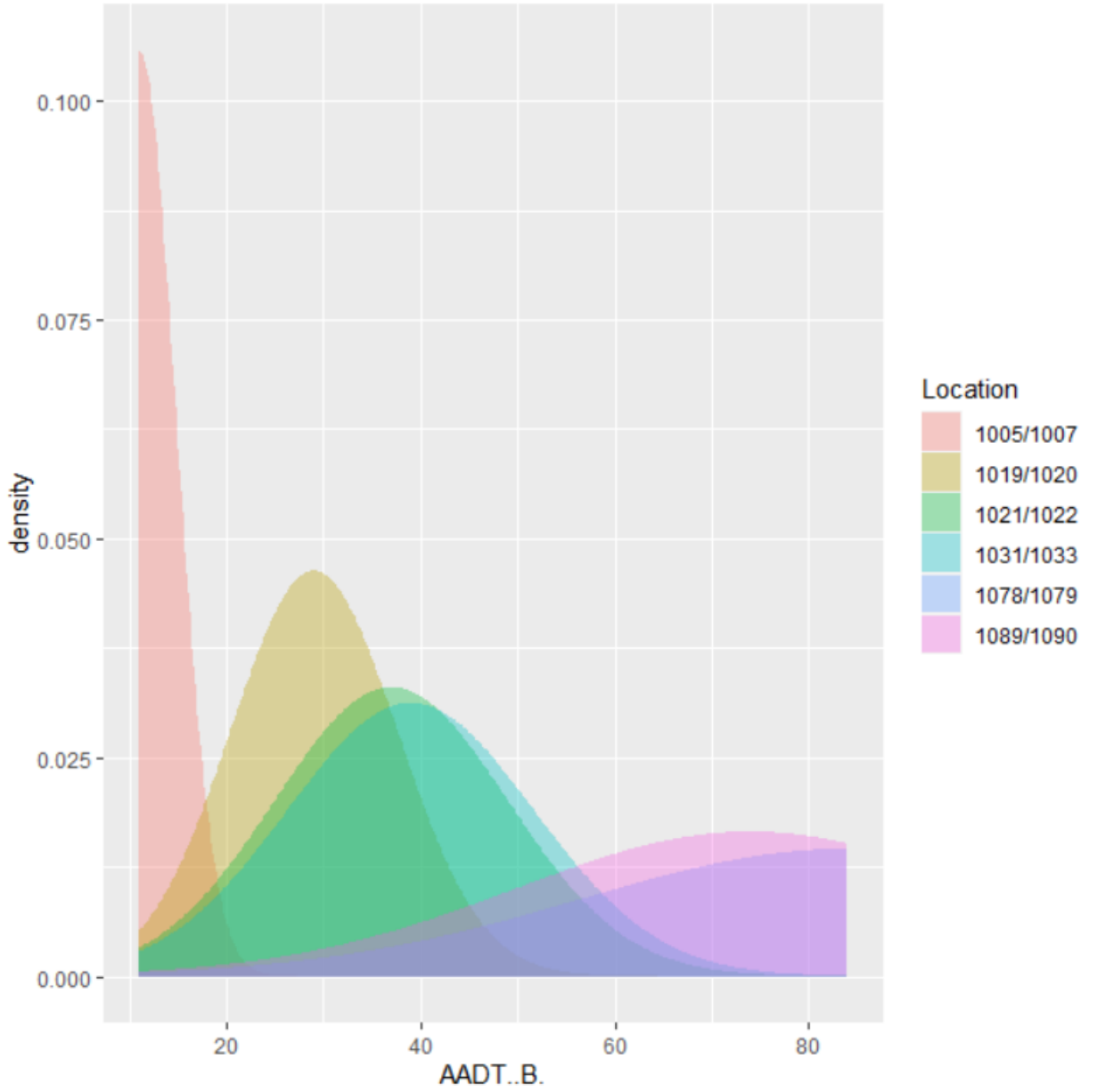


Figure 19. Hourly Pedestrian Volume Count Distribution Plot According to Location

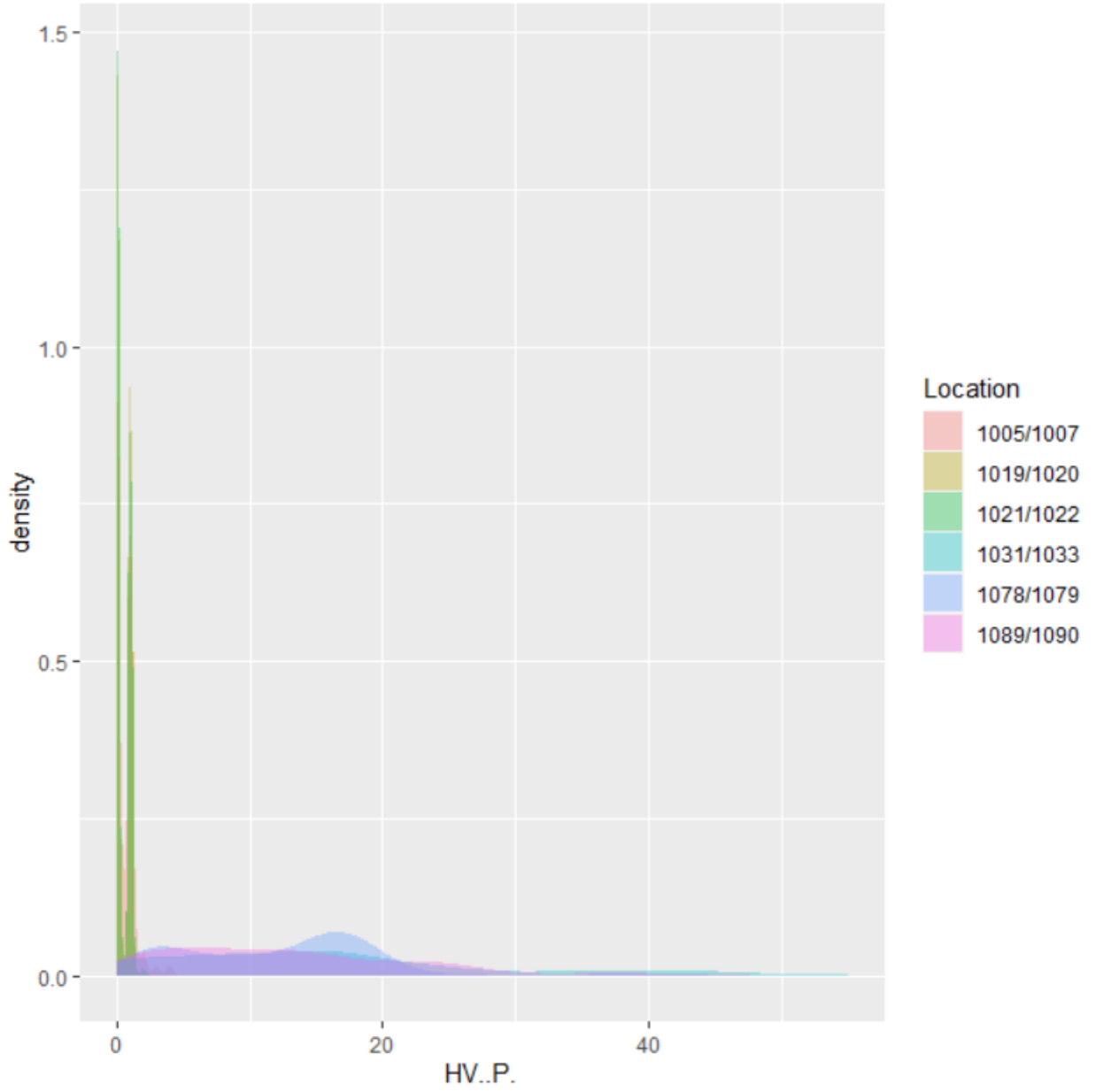


Figure 20. Hourly Bicyclist Volume Count Distribution Plot According to Location

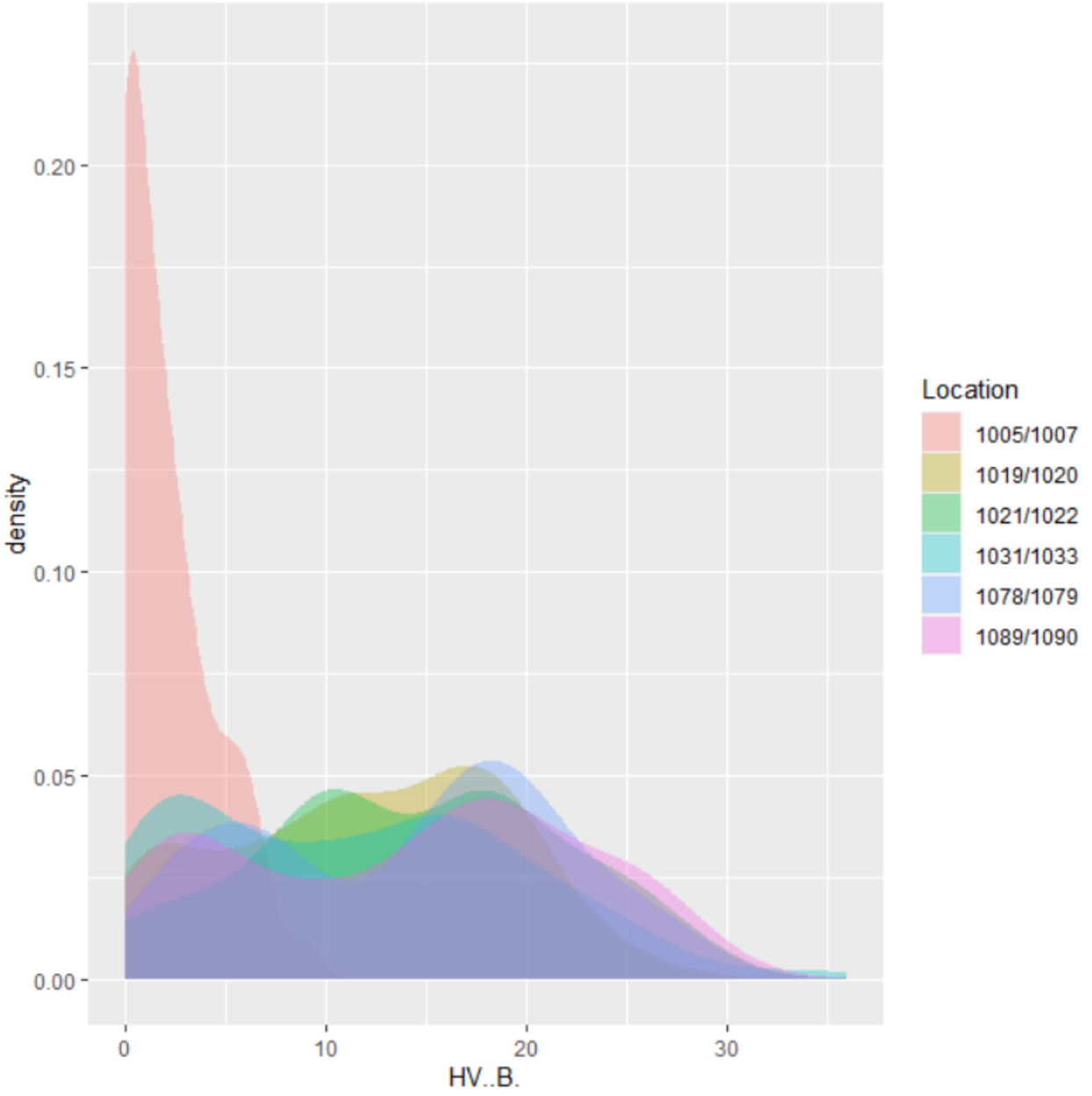


Figure 21. Temperature (°F) Distribution Plot According to Location

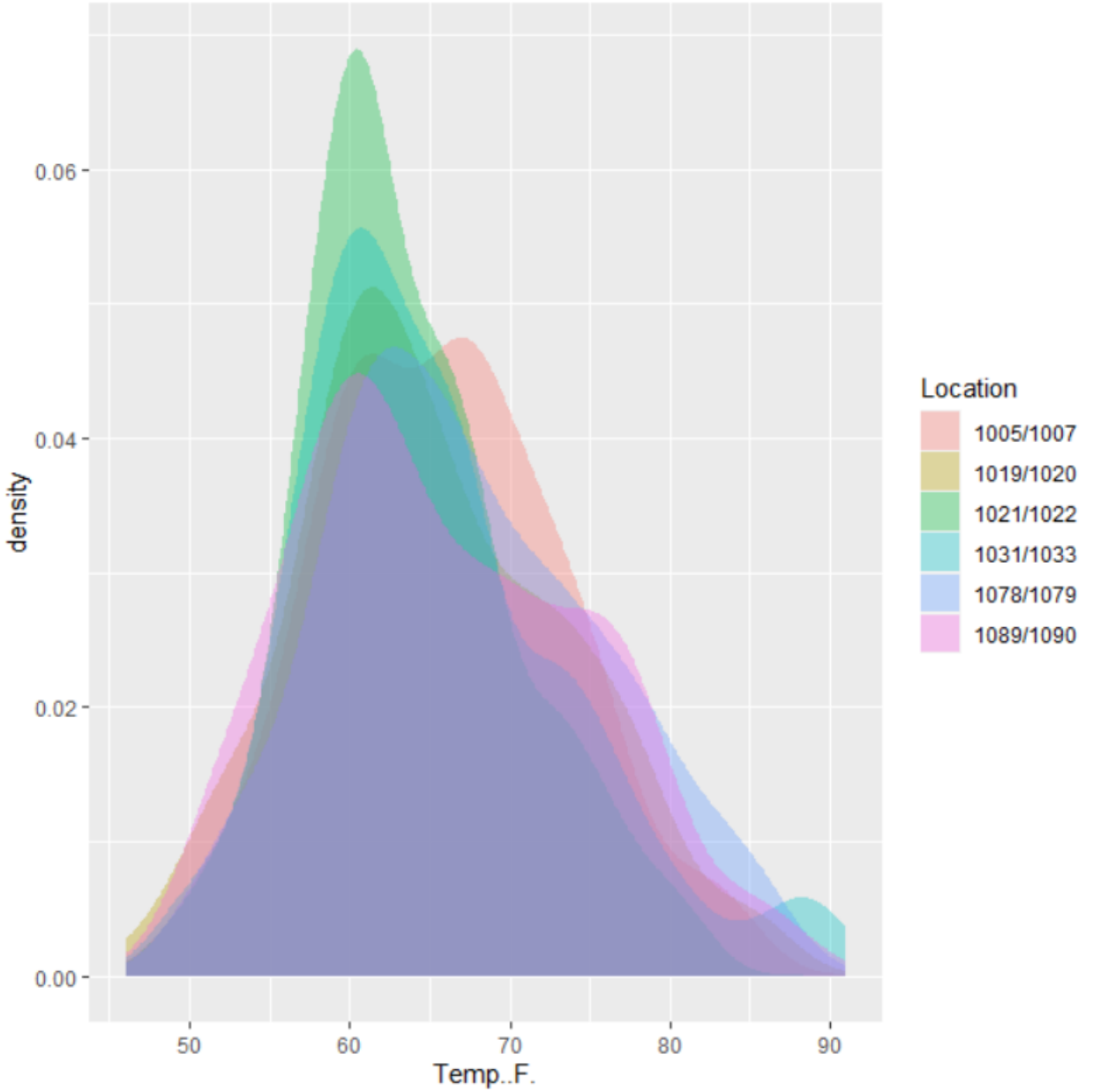


Figure 22. Dew Point (°F) Distribution Plot According to Location

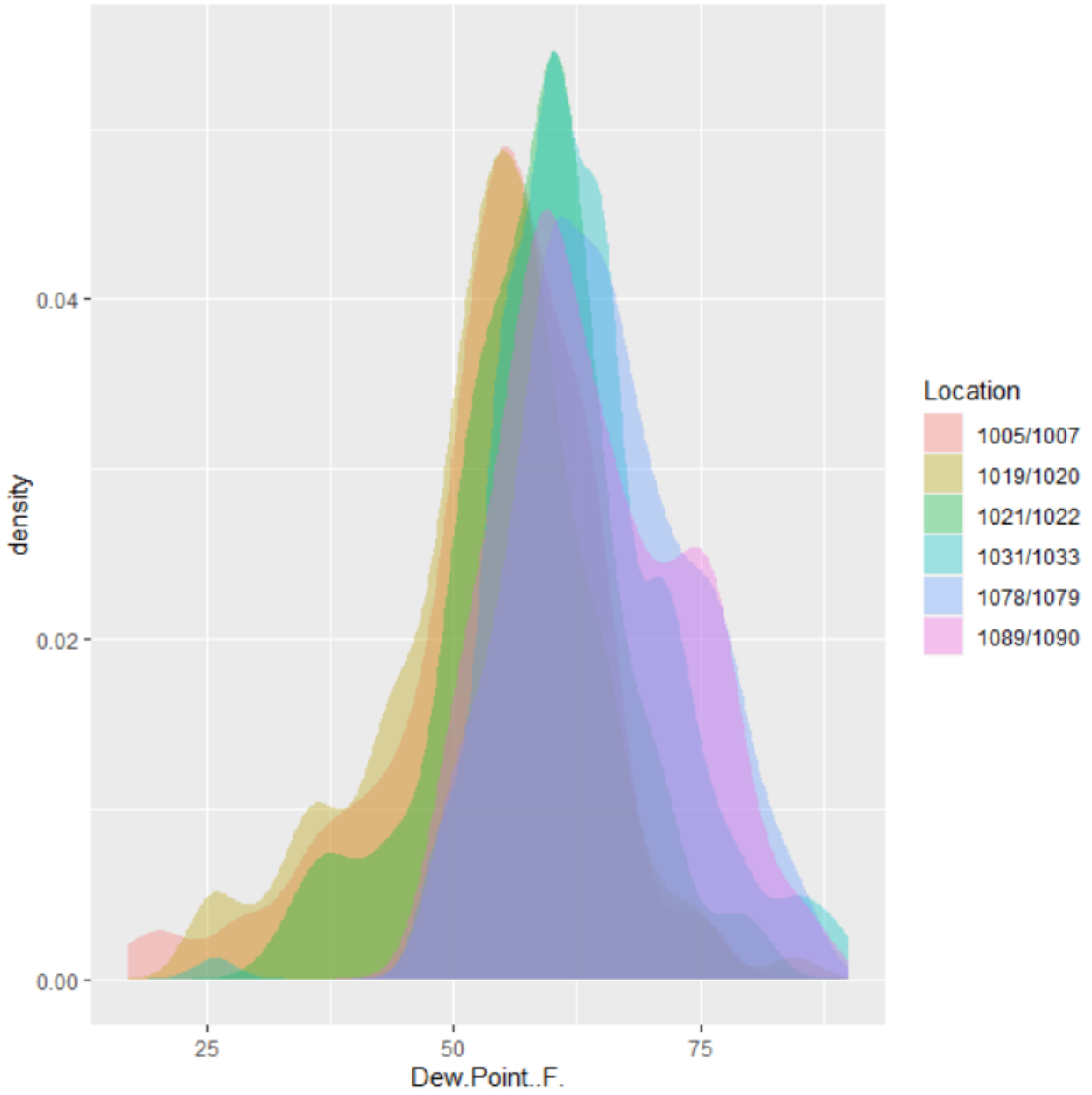


Figure 23. Humidity (%) Distribution Plot According to Location

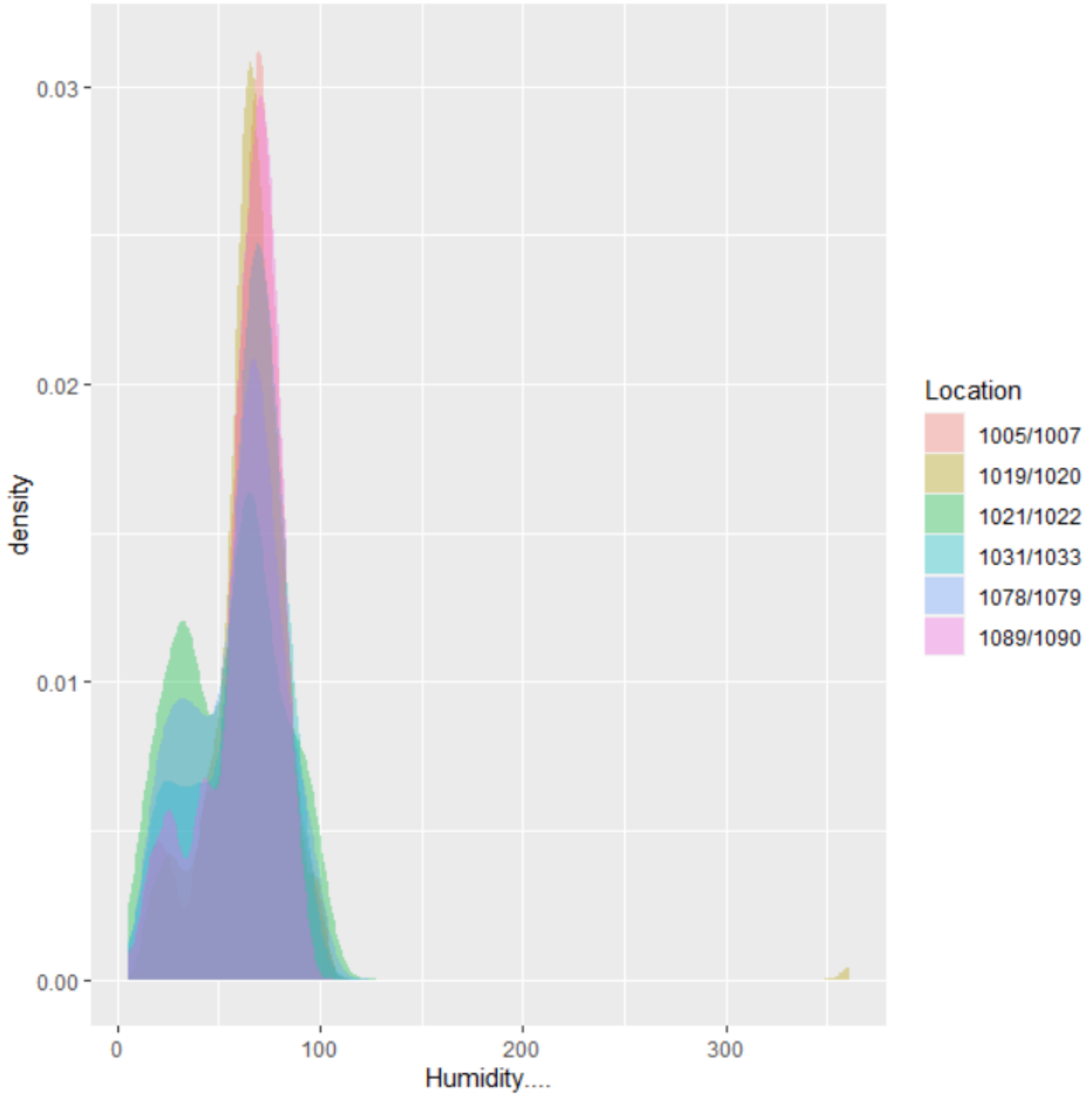
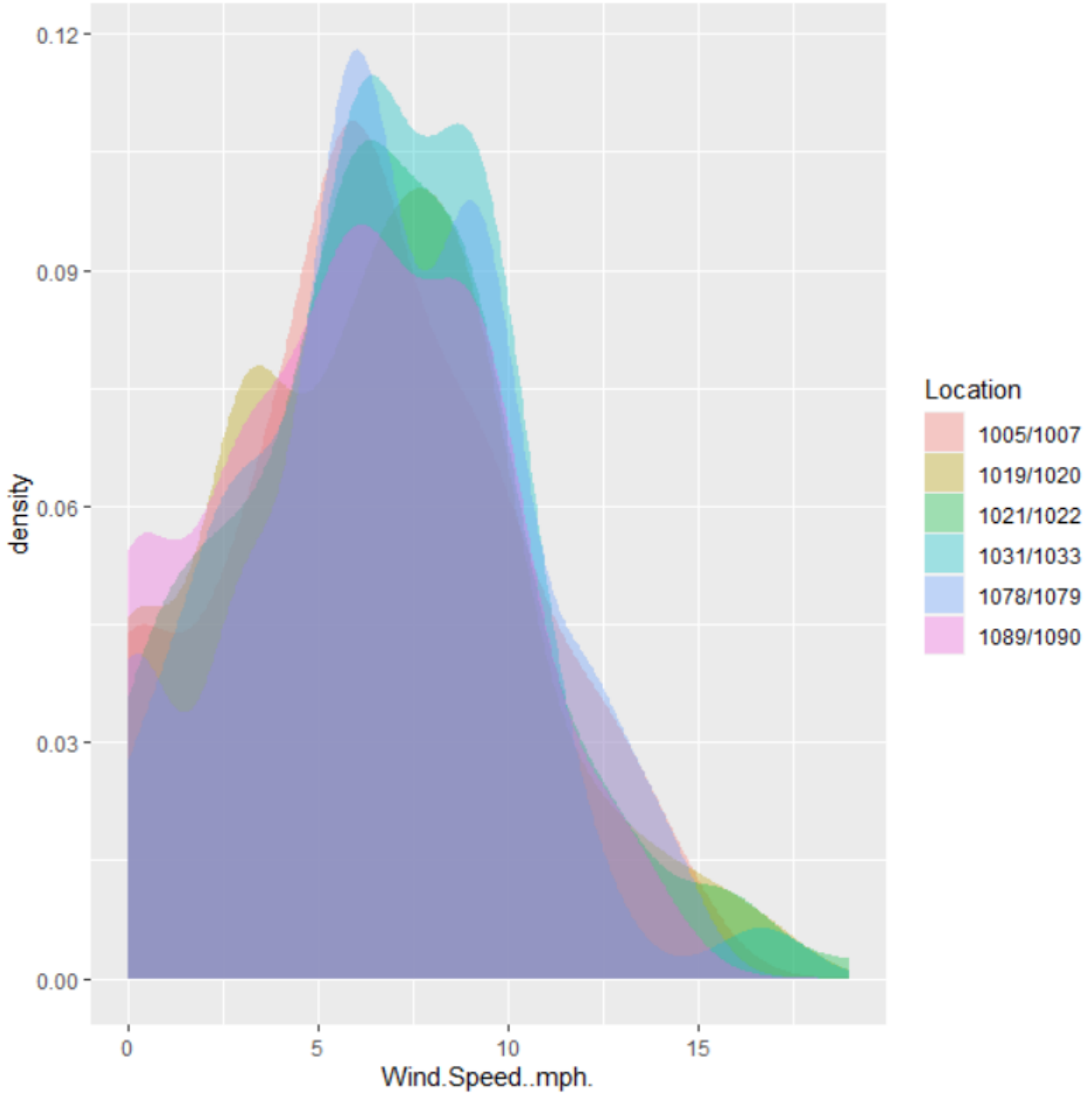


Figure 24. Wind Speed (mph) Distribution Plot According to Location



About the Authors

Wen Cheng, PhD, PE, TE, PTOE

Dr. Cheng is a professor from the civil engineering department at Cal Poly Pomona. His specialty areas include statistical modeling and traffic safety.

Yongping Zhang, PhD, PE

Dr. Zhang is an associate professor from the civil engineering department at Cal Poly Pomona. His specialty areas include transportation planning models and policy development.

Edward Clay

Mr. Clay is a research assistant from the civil engineering department at Cal Poly Pomona. His specialty areas include statistical modeling, data mining, and computer vision.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors