PEDESTRIAN TRAVEL DISTANCES AND EXPOSURE-BASED CRASH RATES ACROSS U.S. SETTINGS

Keya Li
Graduate Research Assistant
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
keya_li@utexas.edu

Kara M. Kockelman, Ph.D., P.E. (corresponding author)
Dewitt Greer Centennial Professor in Engineering
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
kkockelm@mail.utexas.edu 512-471-0210

For presentation at the Bridging Transportation Researchers conference, August 2025

ABSTRACT

This paper examines the demographic and land-use variables that influence walk-trip frequency and walking distances in the U.S. Using a hurdle regression model to predict walk-miles traveled (WMT) per person per day, the study finds that older males, individuals with higher educational attainment, smaller household sizes, and households without vehicles are more likely to walk and to walk longer distances within their home tracts. Households with one vehicle are 53.14% less likely to walk, and those with two or more vehicles are 62.66% less likely. Each standard deviation (SD) increase in age is associated with a 58.18% increase in walking distance. Land use factors – such as population and job density, roadway-network density are strong predictors of one's walking (on the travel survey day) but have little effect on walking distances. These model results were applied to estimate the spatial distribution of WMT across all 60,000 U.S. census tracts to better understand pedestrian activities and their relationship to crash rates. Findings show that Americans average 0.236 WMT per capita per day. A one-week sample of traffic counts on nearly all Texas roadways (by INRIX) was also used to estimate vehicle-miles traveled (VMT) across tracts. Results indicate an average pedestrian crash rate of 0.491 per 1 million WMT and 0.027 per 1 million VMT in Texas. These findings provide a spatial basis for linking walking activity, driving exposure, and pedestrian safety across diverse settings and improve pedestrian safety.

KEYWORDS: Walking trips, Pedestrian deaths, Walking distances, Crash rates

INTRODUCTION

Roadway safety is a serious global challenge, with nearly 1.2 million deaths and 50 million injuries worldwide every year (WHO, 2023). And the U.S. appears to be falling behind its peer nations, with a stubbornly high fatality rate of 1.2 deaths per 100 million vehicle-miles traveled (VMT) (NHTSA, 2025). While this rate was typical 3 decades ago, in 1995, Americans now die at three times the rate per VMT as those in France (Freemark and Jenkins, 2022) and 1.74 times the rate as those in Canada (Transport Canada, 2025). In addition to the death toll, road crashes deliver other serious losses, with the NHTSA estimating economic-only and comprehensive (total) costs of $1,300 and $4,100 per American per year (Blincoe et al., 2019). Among all road users, pedestrians are especially vulnerable to fatalities and injuries in traffic crashes, but they are often ignored when motorists dominate the transportation landscape. In 2023, 7,318 pedestrians were killed and 68,244 were injured in U.S. traffic crashes (NHTSA, 2024). Over the past decade, U.S. pedestrian deaths have risen 54.6% and their share (among total crash deaths) has risen from 14.5% in 2013 to 17.9% in 2023 (NHTSA, 2024; NHTSA, 2014), underscoring the need to investigate pedestrian crash rates.

Various studies (e.g., Lee and Abdel-Aty, 2005; Amoh-Gyimah et al., 2016; Hussain et al., 2019) have identified multiple factors that affect pedestrian crash rates. And instance, Zegeer and Bushell (2012) categorized the influencing factors into five categories: roadway/environmental, driver, vehicle, pedestrian, and demographic/social factors. By examining walking distances across the U.S. across times of day and days of the year, Vellimana and Kockelman (2023) found that southern US states, with longer daylight hours, demonstrate less walking and higher pedestrian death rates (per VMT and per walk-miles traveled (WMT)). But the decision to walk and distances walked vary significantly with demographic attributes, time of year, latitude, and state of residence. Hu and Cicchino (2018) used Poisson and ordinary least-squares (OLS) regression models to analyze U.S. pedestrian fatalities (from motor vehicle crashes) between 2009 to 2016 and found that pedestrian deaths rose most in urban areas, on arterial roads, at midblock locations, and under dark (unlit, nighttime) conditions. Similarly, Zuniga-Garcia et al. (2022) pointed out that signalized intersections and arterial roads tend to have higher pedestrian crash rates. Midblock segments have a three times higher increase of crashes than intersections, with each standard deviation (SD) increase in daily VMT. Along with roadway and environment variables, a corridor's or neighborhood's walking and transit access levels, crime rates, and population demographics (including incomes and presence of children), population and jobs densities, traffic volumes, and land use attributes (like the presence of entertainment districts and food stores) have an important influence on pedestrian crash rates (Cottrill and Thakuriah, 2010; Rahman et al. 2022; Loukaitou-Sideris et al., 2007; Dumbaugh et al., 2024).

Taking all these factors into account, several models (e.g., Turner et al. 2006; Caliendo et al., 2007; Zhao et al., 2024) have been developed to predict pedestrian crash rates based on built environment context (including roadway and land use characteristics) and demographic attributes. For example, Wier et al. (2009) used a multivariate, areal-level regression model to estimate vehicle-pedestrian injury crashes using data from San Francisco census tracts. The model incorporated street, land use, and population characteristics and accounted for nearly 72% of the systematic variation in crash occurrence. Results indicated that a 15% increase in census-tract traffic volume was associated with a 11% rise in vehicle-pedestrian injury collisions, while a 15% increase in area employees led to a 3% increase in vehicle-pedestrian injury collisions. Considering spatial correlation, Wang and Kockelman (2009) applied a Poisson-lognormal multivariate conditional autoregressive (CAR) model to analyze pedestrian crash counts in Travis County, Texas, from 2007 to 2009, and they found that pedestrian crash risk across all severity levels occurred in areas with greater mixing of residents and commercial land uses. Working at a disaggregate level, Rahman et al. (2022) adopted a negative binomial model and a heteroskedastic ordered probit model to investigate pedestrian counts and injury severity over 0.7 million segments across Texas. They concluded that higher speed limits were associated with lower crash frequencies but more fatalities, and that the use of light-duty trucks also increased the risk of pedestrians being severely injured or killed.

Based on these modeling approaches, incorporating exposure-based crash rates (e.g., per WMT, per VMT) is essential for accurately assessing crash risk and identifying high-risk locations. Many studies (e.g., Agrawal and Schimek, 2007; Iosa et al., 2012; Yang and Diez-Roux, 2012; Buehler et al., 2020) have investigated walk distances across the U.S. and/or other nations. For instance, Buehler and Pucher (2021) estimated Europeans' walk distances to average 45% to 118% more than those of Americans. As a result of this, plus higher speeds, less driver training, and myriad other reasons, Americans end up experiencing 5 to 10 times more pedestrian deaths per million WMT than their EU counterparts. Understanding exposure-based pedestrian crash rates is crucial for identifying high-risk locations and implementing safety countermeasures or new technologies. As autonomous vehicles (AVs) continue to develop rapidly, they have demonstrated substantial safety benefits in decreasing crashes (Combs et al., 2019; Utriainen, 2020; Sohrabi et al., 2021; Alozi and Hussein, 2022). For example, Kusano et al. (2024) performed a safety assessment of Waymo's Rider-Only (without a driver behind the steering wheel) crash rates compared to human benchmarks using over 7 million Rider-Only miles in Phoenix, San Francisco, and Los Angeles, and they found that there was a 55% reduction in police-reported crashes and a 80% in any-injury-reported crashes. Susilawati et al. (2023) simulated various scenarios with different levels of AV and CAV deployment to test their safety effectiveness. Results indicated that, compared with non-AV scenarios, assuming all vehicles are AVs or CAVs could reduce pedestrian crashes by 46% and 59%, respectively. When AVs and CAVs were restricted to arterial roads, pedestrian crashes decreased by 59% and 69%, respectively.

However, most exposure-based pedestrian crash-rate estimates (e.g., Vellimana and Kockelman, 2023; Bernhardt and Kockelman, 2021; NHTSA, 2024) are limited to national or state levels, making it difficult to effectively target and operate AV deployment. To address this limitation, this study estimates WMT and VMT at the tract level in three states (Arizona, California, and Texas) where Waymo is actively operating and derives pedestrian exposure-based crash rates. The analysis targets regions with high pedestrian exposure risk and provides insights to inform future AV deployment. The following sections describe predictive methods and all datasets used, followed by estimates of pedestrian crash rates (per WMT, per VMT, and per capita) across these three states, with conclusions.

DATA DESCRIPTION

This section describes the datasets used to derive tract-level WMT estimates, tract-level VMT estimates, and pedestrian crash rates, along with preliminary analyses.

WMT Estimates

This paper uses the 2016/17 NHTS dataset to develop a hurdle regression model that predicts WMT of each respondent per day. The 2016/17 NHTS dataset (FHWA, 2017) obtained travel and other data from nearly 130,000 U.S. households on an assigned survey day (one day per household). The resulting dataset includes 923,572 person-trips over 13 months (March 2016 through March 2017). Personal vehicles (i.e., passenger cars, SUVs, vans, and pickup trucks) were the primary mode for 85.4% of all trips, while walking accounted for 8.8% of the total (or 81,288 walk trips). Additionally, 16,073 trips include people walking to or from transit stops, and those are included in the analysis as well. To exclude hiking and unusually long walks, which tend to be away from public roadways (limiting actual pedestrian exposure to police-recorded crashes), longer walk-trip distances are set to a maximum of 3 miles (using shortest-path calculations between start and end addresses) and included a total of 95,062 walk trips, with a mean distance of 0.53 miles and a standard deviation of 0.56 miles.

In the NHTS data, 19.8% of walk trips occurred between different census tracts, accounting for 15.8% of the total 83,837 surveyed WMT. For simplification, inter-tract walk trips were evenly distributed between the origin and destination tracts. Additionally, 47.3% of all walk trips occurred outside the respondent's home tract, representing 41.5% of total surveyed WMT. In addition to trip and demographic data, land use and pedestrian-related attributes, such as network density, intersection density, and the percentage of parks and lawns, which tend to attract many walk trips, were derived from the American Community Survey (ACS) (U.S. Census Bureau, 2020), the EPA Smart Location Database v3.0 (EPA, 2021), and the National Land Cover Database (NLCD) (Dewitz, 2020). Table 1 summarizes the statistics for all NHTS trips across all modes.

Table 1: Summary Statistics of 2016/17 NHTS Person Records (n = 211,354 respondents)

Variables	Mean	Median	Std Dev	Min	Max
Respondent WMT on sample day	0.35	0	1.30	0	36.78
Age (in years)	47.6	51	21.3	5	92
Household Size	2.71	2	1.39	1	13
Male	0.48	0	0.50	0	1
White	0.82	1	0.39	0	1
African American	0.07	0	0.26	0	1
Asia	0.04	0	0.21	0	1
Other Race	0.06	0	0.24	0	1
No Household Vehicle	0.03	0	0.16	0	1
One Household vehicle	0.22	0	0.41	0	1
Two or More Household Vehicles	0.76	1	0.43	0	1
No High School or College Degree	0.32	0	0.47	0	1
Some College or Associates Degree	0.26	0	0.44	0	1
Bachelor's Degree	0.22	0	0.42	0	1
Graduate Degree	0.20	0	0.40	0	1
Worker	0.53	1	0.50	0	1
Monday	0.15	0	0.36	0	1
Tuesday	0.16	0	0.37	0	1
Wednesday	0.16	0	0.37	0	1
Thursday	0.16	0	0.37	0	1
Friday	0.16	0	0.37	0	1
Saturday	0.10	0	0.30	0	1
Sunday	0.10	0	0.30	0	1
Household Income less than $50k	0.33	0	0.47	0	1
Household Income less than $100k	0.33	0	0.47	0	1
Household Income less than $200k	0.26	0	0.44	0	1
Household Income more than $200k	0.08	0	0.27	0	1

Hispanic Origin	0.09	0	0.29	0	1
Population Density of Home Tract (per acre)	5.45	2.53	13.03	0	389.95
Job Density of Home Tract (per acre)	2.57	0.54	16.77	0	1306.41
Intersection Density of Home Tract (per acre)	56.56	39.19	60.88	0	1266.15
Road Density of Home Tract (miles per acre)	12.52	10.97	9.04	0.2	78.45
Share of Parks of Home Tract (out of 1)	0.11	0.08	0.09	0	0.74
Note: Population data are from the ACS (U.S. Census Bureau, 2020); job, intersection, and road data are from the EPA Smart Location Database v3.0 (EPA, 2021); and the share of parks is from the NLCD (Dewitz, 2020).					

Tract-level WMT per day estimates were derived using the synthetic population developed by Kockelman et al. (2022) via PopGen 2, based on the Public Use Microdata Sample (PUMS) dataset. The PUMS dataset, released by the U.S. Census Bureau (2023), contains detailed demographic and housing information from individual-level responses and is commonly used to produce estimates at the Public Use Microdata Area (PUMA) level. Kockelman et al. (2022) created a 10% sample of the U.S. household and person synthetic data at the census tract level, based on marginals from the 2019 ACS dataset, which can be used to reconstruct demographic information for each tract. It includes nearly 31 million individuals and 13 million households over 73,056 U.S. tracts. The detailed methodology is described in the following section.

VMT Estimates

Traditionally, VMT data is publicly available at the national and state levels through the USDOT and state governments (e.g., USDOT, 2018; California DOT, 2023), based on the Highway Performance Monitoring System (HPMS) database. However, this database uses a stratified sampling method to collect traffic volume data for all public roads, excluding rural minor collector and local roads. Therefore, several studies (e.g., Klatko et al., 2016 and 2017; Williams et al., 2016; Alexander et al., 2024) have been carried out to infer local road performance and improve the accuracy of VMT estimates at a finer level. For example, Klatko et al. (2017) clustered local roads, applied spatial interpolation (including Kriging, natural neighbor, inverse distance weighting, and trend methods), and used sparse traffic volume data for segments within each cluster. Similarly, Chen et al. (2024) used small grid cells to address the lack of granular HPMS data on local roads and produced crash rates at the grid-cell level. In addition to VMT, many researchers (Sharma et al., 2001; Castro-Neto et al., 2009; Sun and Das, 2015; Baffoe-Twum et al., 2023) have employed various models to estimate AADT, including regression analysis and neural networks. For instance, Zhao and Park (2004) applied geographically weighted regression (GWR) using data from 857,775 count stations and tested the modeled on 82 stations. Results showed that 63.41% of the testing points had errors of less than 20% in AADT values.

This paper uses TxDOT Roadway Inventory dataset (TxDOT, 2025), which includes Texas road details such as road class, number of lanes, and other characteristics, along with Annual Average Daily Traffic (AADT) estimates for each road, to construct a Weighted Least Square (WLS) model for predicting AADT in other states. As California and Arizona do not maintain comparable roadway inventory datasets, OpenStreetMap data (OSM, 2025) were retrieved for these two states. Land use variables were mapped using QGIS with the EPA Smart Location Database. Table 2 presents a summary of statistics for the roads in these three states.

Table 2: Summary Statistics of AZ, CA, and TX Road Attributes

Variable	Description	Mean	Media n	Std. Dev.	Min	Max
Texas Roads: n=908,695						
Interstate	1 if the segment is an interstate; 0 otherwise	0.04	0	0.20	0	1
Freeway	1 if the segment is a freeway; 0 otherwise	0.02	0	0.15	0	1
Pr_arterial	1 if the segment is a principal arterial; 0 otherwise	0.10	0	0.29	0	1
Min_arterial	1 if the segment is a minor arterial; 0 otherwise	0.07	0	0.25	0	1
Maj_collector	1 if the segment is a major collector; 0 otherwise	0.13	0	0.33	0	1
Min_collector	1 if the segment is a minor collector; 0 otherwise	0.02	0	0.15	0	1
Local	1 if the segment is a local road; 0 otherwise	0.62	0	0.48	0	1
Oneway	1 if the segment is one-way; 0 two-way	0.03	0	0.17	0	1
Median	1 if the segment has a median (a TWLTL) is not considered as a median); 0 otherwise	0.12	0	0.33	0	1
# Lanes	Number of lanes per direction on the segment	2.26	2	0.83	1	17
Lane_width	Width of each lane (ft)	3.68	0	5.58	0	47.00
Length	Length of the segment (mi)	0.38	0.16	0.74	0	44.24
Pop_density	Population per acre (residents/acre) in the nearest block group	2.50	0.74	3.64	0	90.34
Job_density	Jobs per acre in the nearest block group	1.29	0.13	5.96	0	270.62
Road_density	Total road network density per square miles	10.02	6.58	8.78	0.08	69.85
Arizona Roads: n=341,093						
Interstate	1 if 'highway' tag is 'motorway' or 'motorway_link'; 0 otherwise	0.03	0	0.18	0	1
Freeway	1 if 'highway' tag is 'trunk' or 'trunk_link'; 0 otherwise	0.02	0	0.14	0	1
Pr_arterial	1 if 'highway' tag is 'primary' or 'primary_link'; 0 otherwise	0.05	0	0.22	0	1
Min_arterial	1 if 'highway' tag is 'secondary' or 'secondary_link'; 0 otherwise	0.13	0	0.33	0	1
Maj_collector	1 if 'highway' tag is 'tertiary' or 'tertiary_link'; 0 otherwise	0.10	0	0.30	0	1
Min_collector	1 if 'highway' tag is 'unclassified'; 0 otherwise	0.05	0	0.22	0	1
Local	1 if 'highway' tag is 'residential'; 0 otherwise	0.62	1	0.49	0	1
Oneway	1 if 'oneway' tag is 'yes'; 0 otherwise	0.27	0	0.44	0	1
Median	0 if 'lane' tag is 1 or 'lane_markings' tag is 'no median'; 1 otherwise	0.38	0	0.49	0	1
# Lanes	Number of lanes per direction on the segment retrieved from 'lane' tag	1.75	1	1.15	1	11
Lane_width	Width of each lane (ft) calculated based on 'width' tag	0.03	0	1.05	0	144.36
Length	Length of the segment calculated using QGIS (mi)	0.25	0.09	0.75	0	51.05
Pop_density	Same as TX dataset	3.81	2.32	4.40	0	53.57
Job_density		1.82	0.24	6.49	0	151.17

Road_density		13.32	13.52	9.43	0.04	57.26
California Roads: n=1,100,034						
Interstate	Same as AZ dataset	0.06	0	0.24	0	1
Freeway		0.01	0	0.11	0	1
Pr_arterial		0.11	0	0.31	0	1
Min_arterial		0.14	0	0.35	0	1
Maj_collector		0.10	0	0.30	0	1
Min_collector		0.03	0	0.17	0	1
Local		0.55	1	0.50	0	1
Oneway		0.29	0	0.46	0	1
Median		0.45	0	0.50	0	1
# Lanes		1.72	1	1.26	1	55
Lane_width		0.05	0	1.68	0	209.97
Length		0.27	0.09	0.77	0	66.46
Pop_density		8.03	5.38	9.76	0	261.67
Job_density		4.69	1.05	1.98	0	767.31
Road_density		17.45	18.34	10.89	17.87	166.89
Note: Texas data are from TxDOT Roadway Inventory dataset (2025). Arizona and California data are extracted from OSM (2025). Roads that cannot carry traffic (e.g., steps, footway, construction) were removed from the OSM data. Other road attributes in the Texas dataset (including school zones, median width, and curve) were not included in this analysis due to missing data in the OSM dataset.						

As detailed in the following section, the WLS model is primarily trained using Texas AADT data, along with a portion of Arizona and California AADT data, and tested on the remaining AADT data from these two states. Figure 1 presents the AADT data provided by Arizona DOT (2024) and California DOT (Caltrans) (2024). Both data were collected from 2023 HPMS database. In contrast to Arizona's line segment AADT dataset (Figure 1 (a)), California's AADT dataset is a point-based GIS dataset (yellow dots in Figure 1 (b)) that contains *Ahead_AADT* and *Back_AADT* attributes. To assign AADT values to the corresponding road segments, California's State Highway Network Lines dataset (blue lines in Figure 1 (b)) was used to determine roadway direction (e.g., southbound). Using QGIS, each data point was connected with its ahead segment and back segment IDs, and the AADT values were then assigned to the associated road segments. In total, 111,057 and 72,524 AADT segment records were collected in Arizona and California, respectively.

Pedestrian Crash Data in Texas

Pedestrian crash data were collected from California's Transportation Injury Mapping System (TIMS, 2025), Arizona DOT's Crash Query System (2025), and TxDOT's Crash Records Information System (CRIS) Query Tool (2025). These statewide databases include all reported motor vehicle traffic crashes, along with detailed information such as crash time, location, weather conditions, and person characteristics. However, at least half of all pedestrian and bicyclist crashes are underreported to the police (Stutts and Hunter, 1998). In this study, these missing data—which are difficult to obtain—are not included. Pedestrian-related crashes reported from 2020 to 2024 (the most recent five available years) were queried for analysis, and crashes with missing or out-of-boundary location information were also excluded. Figure 2 plots the distribution of pedestrian crash data over five years in the three states, and the summary of crash counts is provided in Table 3.

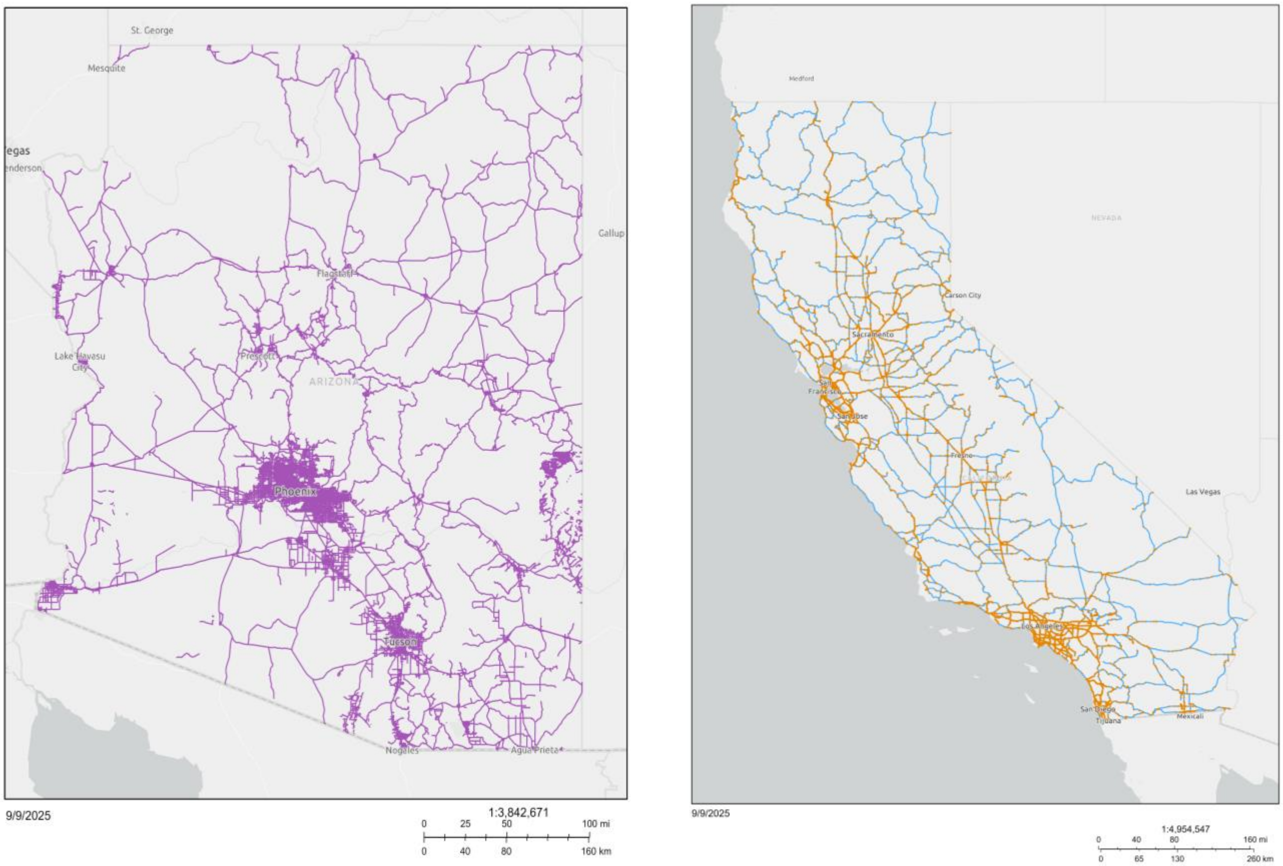
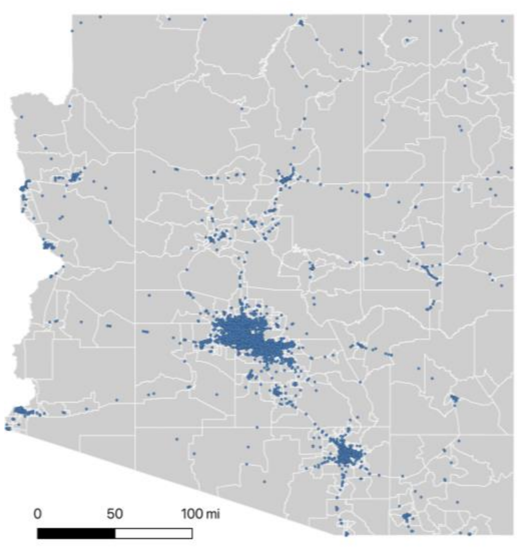
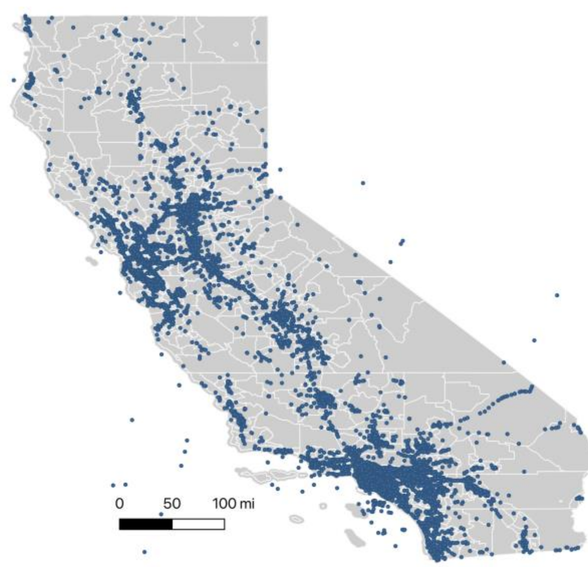


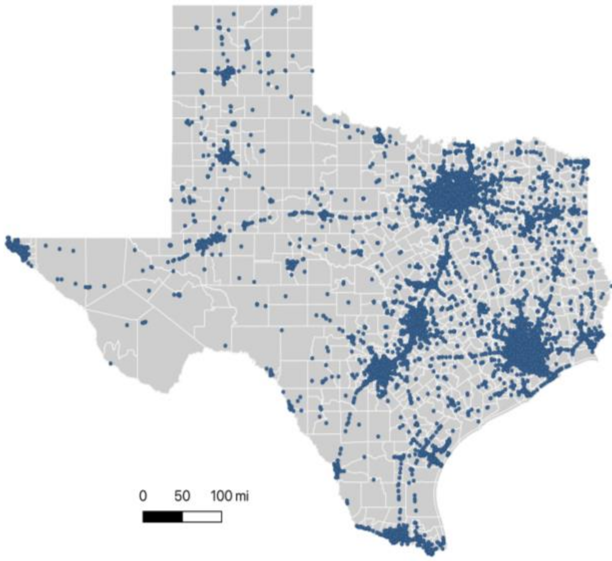
Figure 1: AADT Data for Arizona (left) **and California** (right)



(a) Arizona



(b) California



(a) Texas

Figure 2: Pedestrian Crashes across Census Tracts in Arizona, California, and Texas (Data: 2020 through 2024)

Table 3: Summary Statistics of Pedestrian Crashes

	Arizona	California	Texas
# Pedestrian crashes per yr	1,858	4,795	7,502
# Pedestrian fatalities per yr	255	736	814
# Pedestrian serious injuries per yr	378	1,055	1,607
# Pedestrian minor injuries per yr	774	1,786	3,015
# Pedestrian possible injuries per yr	451	1,218	2,066
Source: Crash data (2020-2024) are from California's Transportation Injury Mapping System (TIMS), Arizona Department of Transportation Crash Query, and Texas's Crash Records Information Systems (CRIS).			

Using Texas as an example, Figure 3 plots the number of pedestrian crashes by time of day and day of week over five years across the state. Pedestrian crashes, especially fatalities, mostly occur during late nights, particularly on Fridays and Saturdays. The number of crashes peaks between 5 and 7 a.m. and between 6 and 9 p.m. on weekdays, and between 6 and 9 p.m. on weekends. Two-thirds of crashes occur at non-intersections, and more than 80% take place in urban areas. Male pedestrians involved in crashes are twice as common as female pedestrians, and those between 20 and 40 years old are more likely to be injured.

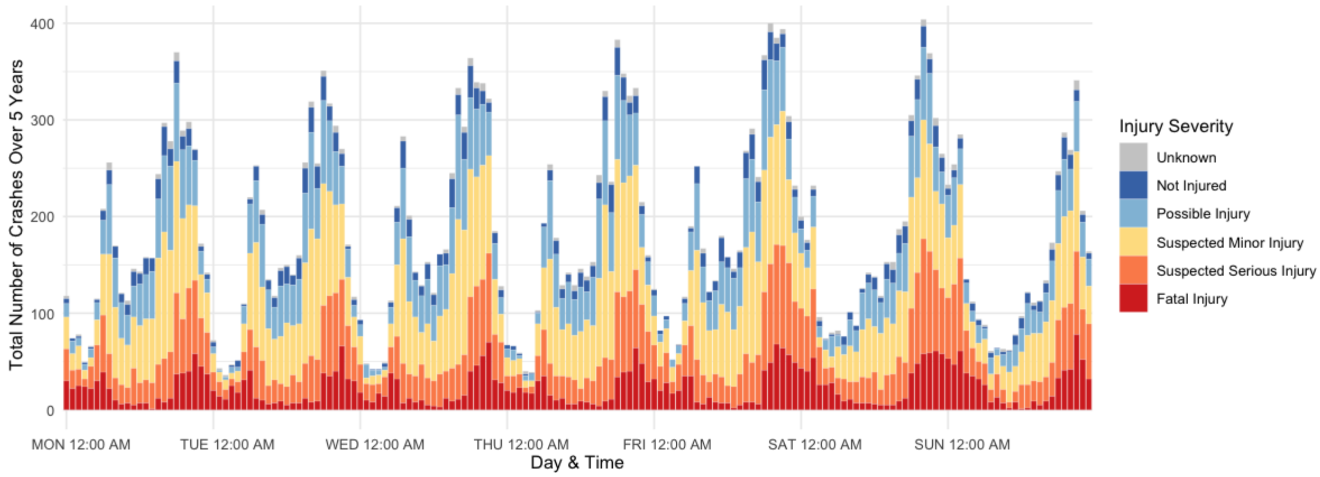


Figure 3: Texas' Pedestrian Crashes by Time of Day and Day of Week over 5 Years
(Source: CRIS data, 2020 through 2024)

METHODS

To derive pedestrian exposure-based crash rates, this study estimates tract-level WMT and VMT. Figure 4 presents the datasets and outlines the steps of the framework, coded in R and Python. In estimating WMT, given that nearly half of walk trips occur outside the respondent's home tract, mainly for work, study, and grocery purposes, this study divides the WMT model into two parts: a production model and an attraction model. The production model estimates walk trips originating in the tract where residents live, using the synthetic dataset developed by Kockelman et al. (2022) to produce tract-level estimates of WMT per resident per day. The attraction model, by contrast, estimates walk trips drawn to each tract by non-residents. Since demographic information about workers, students, or customers within each tract is unavailable and difficult to collect, this model combines the average WMT for workers walking outside their home tract with tract job density data to provide a simplified estimate of walk trips attracted to each tract.

Since only 8.8% of all NHTS 2016/2017 person-trips are walking trips, a hurdle regression model was used to reflect the high frequency of zero walk distances by respondents on their survey day. This two-part model uses a logistic model to predict zeros in the data, followed by a continuous log-linear model with normally distributed error terms for all non-zero distances (Cragg 1971; Gurmu, 1988; Gurmu and Trivedi, 1996; Lahiri and Xing, 2004). Covariates include the respondent's age and age squared; gender, race and ethnicity; education level and employment status; household size and income; home ownership; population density, job density, (public) roadway network and intersection density, and the share of parks and lawns in the home tract. The practical significance of each covariate is inferred here by increasing that covariate by one SD, across the NHTS sample, and comparing the resulting (population-average) predicted WMT/person/day value to the (population-weighted) sample average before the modification. Those variables having the greatest effect on average predicted walking distance are considered the most practically significant, as discussed in the next section. After developing the production model (WMT per person per day produced by residents), the estimated parameters were applied to predict tract-level daily WMT using the synthetic dataset. On the other hand, WMT occurring outside home tracts (e.g., trips attracted to a tract for work, school, or other purposes) were estimated using the average walking distance outside home tracts of sampled workers in the NHTS dataset. By combining the production and attraction models, this study derives estimates of daily WMT for each tract across the United States.

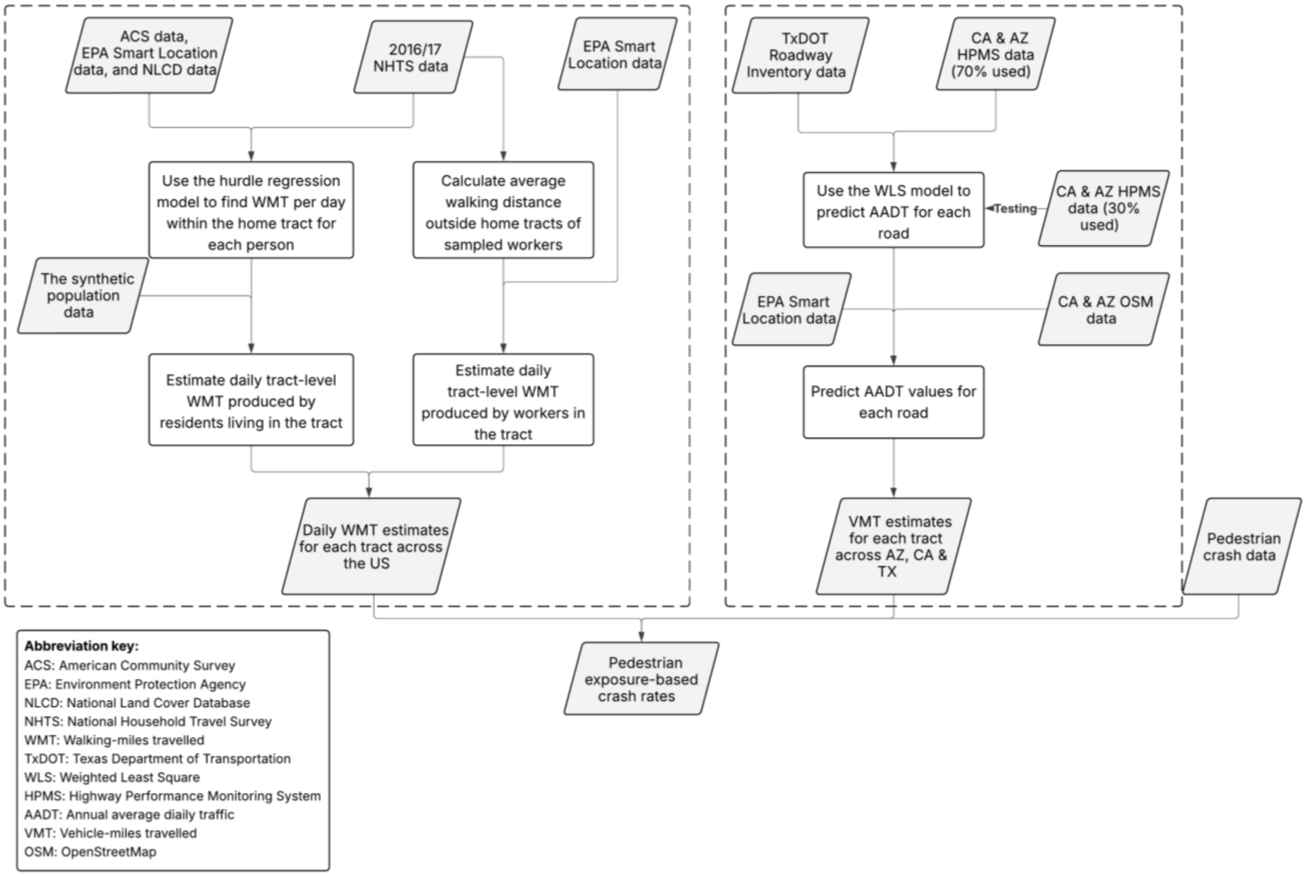


Figure 4: Steps for Deriving Pedestrian Exposure-based Crash Rates

To estimate VMT at the road level and aggregate it into tract-level values, this study applies a WLS model to predict AADT across road segments, using TxDOT's Roadway Inventory combined with HPMS data from California and Arizona. While the HPMS datasets in California and Arizona primarily represent higher functional classes such as interstates and principal arterials, the Texas dataset provides more comprehensive AADT values across all roadway types. To account for systematic differences between states, the model incorporates a state indicator variable and assigns greater weights to California and Arizona observations to enhance transferability, since predictions are ultimately required for these states. The weighting scheme is defined as the product of road length and functional class, with functional class weights set as follows: interstate and other freeways (=3), principal arterials (=2), minor arterials (=1.5), major collectors (=1), minor collectors (=0.75), and locals (=0.5), considering the fact that longer, higher-class roads generally carry more traffic and are measured with greater accuracy. Model training uses 70% of the HPMS data, with the remaining 30% of HPMS records are reserved for validation to evaluate model performance prior to statewide prediction. Finally, AADT estimates are integrated with OSM road network data and tract-level land use attributes to derive VMT estimates, which are subsequently used to calculate pedestrian exposure-based crash rates.

RESULTS

The following section presents key highlights from three parts: (1) the WMT/person/day hurdle model, (2) the AADT WLS model and its prediction performance, and (3) pedestrian exposure-based crash rates in three states.

Hurdle Model of WMT/Person/Day in Home Tract

Table 4 presents parameter estimates and practical significance for the *logistic* and *exponential portion* of the production model. In the logistic model, results indicate that older individuals, males, those with higher education levels, and individuals in smaller households without vehicles were more likely to walk within their home tract – everything else constant. Among demographic and household characteristics, Hispanic individuals were less likely to walk, while education levels above high school were positively associated with walking behavior. Land use attributes, including higher population and job density in the home tract, were also significantly associated with an increased likelihood of walking, suggesting that denser neighborhoods with more parks encourage pedestrian activity. In the exponential part, similar to logistic results, older individuals, males, and those with higher education levels tended to walk longer distances, whereas workers generally walked shorter distances. Household vehicle ownership strongly decreased walking distance as well. Unlike the logistic portion, land use variables had limited influence on walking distance, indicating that while neighborhood density affects whether individuals walk at all, it has less effect on how far they walk once they decide to walk.

Among all factors, education levels, vehicle ownership, and work status emerged as the strongest predictors of walking behavior, as shown in Figure 5. Each SD increase in education levels were associated with a 67.9% to 117.6% increase in walking distance. Conversely, a SD increase in household vehicle ownership was linked to up to a 91.3% reduction in walking distance, highlighting the substitutive effect of vehicles on walking.

Table 4: Model Estimates for WMT/day/person (Source: NHTS 2016/17 Data)

Variable	Logistic Model for Pr(WMT>0)			Exponential Model for Pr(WMT = d)			Pract. Sign
	Estimate	t-stat	P-value	Estimate	t-stat	P-value	
Intercept	-0.003	-0.1	0.957	-0.108	-2.3	0.019	
Age	0.005	3.0	0.003	0.022	13.7	0.000	78.3%
Age Squared	-8.6e-05	-4.9	0.000	-2.5e-04	-14.3	0.000	
Household Size	-0.016	-2.6	0.010	0.002	0.3	0.738	-2.0%
Male	0.020	1.5	0.129	0.078	6.7	0.000	5.0%
African American	-0.269	-10.1	0.000	-0.057	-2.4	0.016	-8.1%
Asian	-0.080	-2.6	0.009	-0.019	-0.7	0.455	-2.0%
Other Race	0.014	0.5	0.622	0.015	0.6	0.540	0.7%
One Car in HH	-1.734	-54.2	0.000	-0.121	-6.0	0.000	-53.4%
Two or More Cars in HH	-2.166	-65.8	0.000	-0.167	-7.5	0.000	-63.3%
Some College or Associates Degree	0.027	1.3	0.190	0.078	4.1	0.000	4.7%
Bachelor's Degree	0.320	15.0	0.000	0.105	5.3	0.000	19.3%
Graduate Degree	0.503	22.7	0.000	0.158	7.9	0.000	30.0%

Worker	-0.515	-31.4	0.000	-0.123	-8.8	0.000	-27.2%
Tuesday	-0.026	-1.1	0.255	-0.006	-0.3	0.785	-1.2%
Wednesday	-0.039	-1.7	0.096	-0.003	-0.2	0.868	-1.5%
Thursday	-0.064	-2.7	0.006	0.010	0.5	0.640	-2.0%
Friday	-0.108	-4.6	0.000	-0.007	-0.3	0.754	-4.1%
Saturday	-0.113	-4.2	0.000	0.006	0.2	0.808	-3.2%
Sunday	-0.088	-3.3	0.001	0.055	2.5	0.014	-1.0%
Household Income <= $100k	-0.102	-5.7	0.000	-0.004	-0.3	0.798	-4.9%
Household Income <= $200k	-0.039	-2.0	0.050	0.047	2.6	0.009	0.4%
Household Income > $100k	-0.015	-0.5	0.603	0.110	4.7	0.000	2.6%
Hispanic Origin	-0.200	-7.9	0.000	2.5e-04	0.0	0.991	-5.6%
Population Density of Home Tract (per acre)	0.014	21.8	0.000	-6.9e-04	-2.7	0.008	18.9%
Job Density of Home Tract (per acre)	0.003	6.0	0.000	9.2e-05	0.7	0.492	5.0%
Intersection Density of Home Tract (per acre)	0.002	9.5	0.000	1.2e-04	0.9	0.344	14.2%
Road Density of Home Tract (miles per acre)	0.012	7.1	0.000	0.004	3.2	0.001	15.2%
Share of Parks of Home Tract (out of 1)	-0.160	-2.1	0.034	0.158	2.5	0.014	-0.01%
Notes: N=211,354, Log-likelihood=-5.129e+04. Base case is White respondent, no car, high school degree or lower, monday, and household income less than $50k. Population and job data are from the ACS (U.S. Census Bureau, 2020) and the EPA Smart Location Database v3.0 (EPA, 2021), and and the percentage of parks is from the NLCD (Dewitz, 2020).							

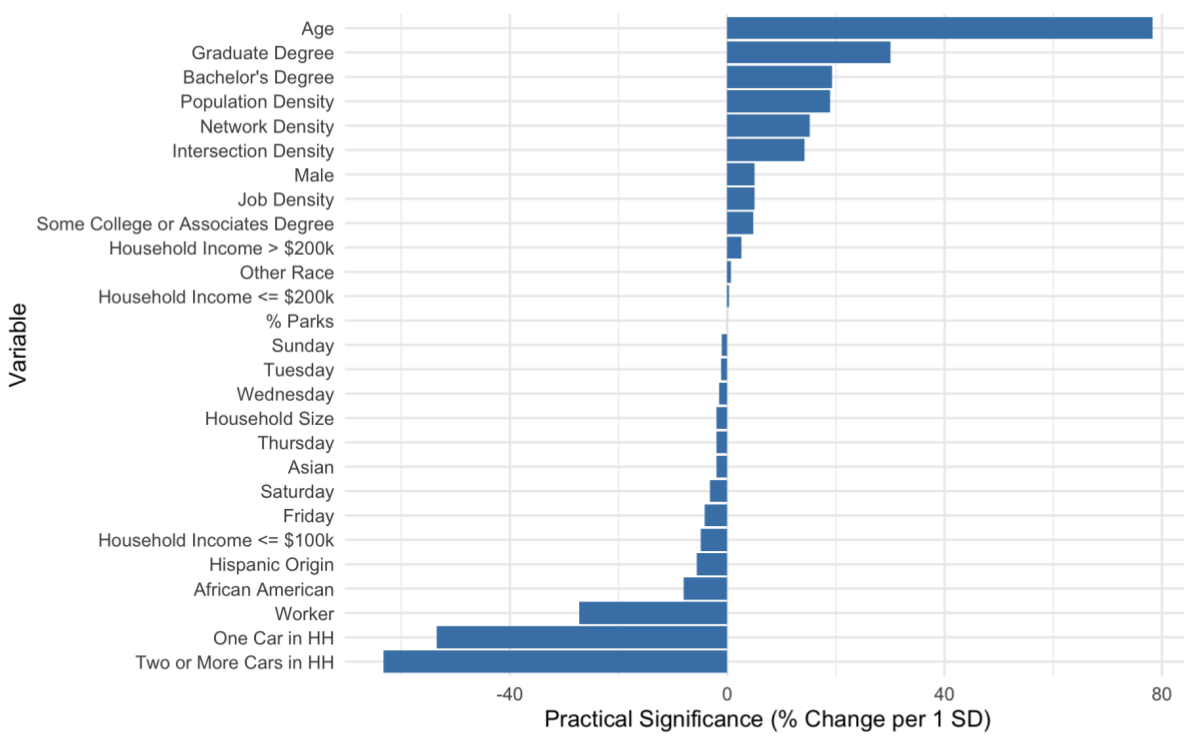


Figure 5: Sensitivity Analysis for WMT/person/day Model

WLS Model for AADT Prediction

Parameter estimates and practical significance for the WLS model predicting log-transformed AADT are shown in Table 5. Results indicated that functional classification strongly influenced AADT, with a one-SD increase in interstate associated with a 53.3% higher traffic volume, while minor collectors and local roads were negatively associated with AADT. Roadway design and land-use attributes also had notable effects. The number of lanes was associated with a 12.8% increase in AADT per SD, while wider lanes showed an exceptionally high 170% increase, mainly due to the large standard deviation of lane width in the trainning dataset. Median presence and two-way design also increased AADT. Additionally, both job density and road density were positively related to AADT, with the latter showing the strongest impact: a one-SD increase in road density corresponded to an 86% rise in AADT.

The model's predictive performance was evaluated on 51,969 HPMS records, yielding a MAE of 1.08, RMSE of 1.37, and an R^2 of 0.476. These results indicate that the model explained nearly half of the variation in log(AADT), with predictions deviating from observed log(AADT) values by an average of 1.08. Since the errors remain within a reasonable range, the model was applied to predict AADT values for the remaining road segments.

Table 5: WLS Model Estimates

Variable	Summary Statistics		Model Estimates for log(AADT)			
	Mean	Std. Dev	Estimate	t-stat	P-value	% Change in AADT per 1 SD
Intercept	-	-	4.746	1160	0.000	-
Interstate	0.060	0.237	1.793	393.4	0.000	53.0%
Freeway	0.026	0.158	1.326	301.7	0.000	23.3%
Pr_arterial	0.097	0.297	0.984	266.4	0.000	33.9%
Min_arterial	0.077	0.267	0.954	263.0	0.000	29.0%
Maj_collector	0.127	0.333	0.191	56.2	0.000	6.6%
Min_collector	0.024	0.152	-0.235	-36.6	0.000	-3.5%
Local	0.589	0.492	-0.268	-49.8	0.000	-12.4%
Oneway	0.077	0.266	-0.139	-30.3	0.000	-3.6%
Median	0.187	0.390	0.279	44.9	0.000	11.5%
# Lanes	2.230	0.906	0.133	106.5	0.000	12.8%
Lane_width	3.247	5.389	0.184	373.8	0.000	169.5%
Length	0.378	0.771	-0.063	-220.9	0.000	-4.7%
Pop_density	2.750	4.011	-0.004	-9.5	0.000	-1.6%
Job_density	1.523	7.101	0.003	23.1	0.000	2.2%
Road_density	10.598	9.159	0.068	317.8	0.000	86.4%
AZ indicator	-	-	1.603	581.4	0.000	51.8%
CA indicator	-	-	3.142	963.4	0.000	91.1%
TX indicator	-	-	0.001	0.3	0.000	0.03%

Notes: N=1,029,955, R^2=0.708, Log-likelihood=-2.777e+06. Training AADT data are from TxDOT Roadway Inventory dataset (2025), Arizona DOT (2024), and California DOT (2024). Road attributes are from TxDOT Roadway Inventory dataset (2025) and OSM (2025). Land use attributes are from EPA Smart Location Database v3.0 (EPA, 2021).

Prediction Performance of the WLS Model (N=51,969)		
Mean Absolute Error (log(AADT))	Root Mean Square Error (log(AADT))	R^2
1.08	1.37	0.476

Pedestrian Exposure-based Crash Rates

	State-level Crash Rates			U.S. Crash Rates
	California	**Arizona**	**Texas**	**Nation**
VMT/capita/year	7974	10529	10345	9641
WMT/capita/year	88.5	69.4	45.9	63.3
# Ped Crashes	4795	1858	7502	65317
per 1 million WMT	1.37	3.74	5.61	3.03
per 1 billion VMT	15.21	24.68	24.88	23.84
per 10,000 Capita	1.21	2.60	2.57	1.92
# Ped Fatalities	736	255	814	7043
per 1 million WMT	0.21	0.51	0.61	0.33
per 1 billion VMT	2.33	3.39	2.70	2.57
per 10,000 Capita	0.19	0.36	0.28	0.21
# Ped Serious Injury	1055	378	1607	
per 1 million WMT	0.30	0.76	1.20	
per 1 billion VMT	3.35	5.03	5.33	
per 10,000 Capita	0.27	0.53	0.55	
# Ped Minor Injury	1786	774	3015	
per 1 million WMT	0.51	1.56	2.25	
per 1 billion VMT	5.67	10.28	10.00	
per 10,000 Capita	0.45	1.08	1.04	
# Ped Possible Injury	1218	451	2066	
per 1 million WMT	0.35	0.91	1.54	
per 1 billion VMT	3.86	5.99	6.85	
per 10,000 Capita	0.31	0.63	0.71	
# Total Crashes per 1 billion VMT	1272	1764	2183	2203
# Fatalities per 1 billion VMT	12.2	15.6	14.2	14.3
# Serious Injuries per 1 billion VMT	32.9	48.0	61.1	
# Minor Injuries per 1 billion VMT	123.0	277.8	298.9	
# Possible Injuries per 1 billion VMT	234.7	350.3	451.5	
# PDO Crashes per 1 billion VMT	869	1073	1357	1246

Source:
1. State-level crash data for the most recent five available years (2020-2024) are from California's Transportation Injury Mapping System (TIMS), Arizona Department of Transportation Crash Query, and Texas's Crash Records Information Systems (CRIS); national crash data for the most recent five available years (2019-2023) is from NHTS.

2. WMT is calculated via WMT hurdle regression model based on 2016/17 NHTS sampled data (Li and Kockelman, 2025).
3. VMT is from California DOT, Arizona DOT, Texas DOT, and Department of Energy (2024).
State population is from U.S. Census Bureau (2024).

This study covers nearly 60,000 census tracts across the United States, excluding those with missing data on area, job density, population density, or other land use attributes. Table 3 presents summary statistics for all tracts included in the analysis. By applying the WMT/person/day production model to the synthetic dataset, estimates of WMT occurring within home tracts were generated for each census tract. A total of 41 million walk-miles occur daily within the home tracts where walkers reside across the U.S.

Table 3: Summary Statistics of U.S. Census Tracts (n = 59,920 tracts)

Variables	Mean	Median	Std Dev	Min	Max
Area (per sq. miles)	39.48	1.57	211.48	0.01	9489.81
Population Density (per sq. miles)	5706.18	2499.26	12331.01	0	257270.06
Job Density (per sq. miles)	2405.09	522.70	16604.99	0	1165419.68
Intersection Density (per sq. miles)	73.15	59.01	73.73	0	1558.38
Network Density (per sq. miles)	15	14.82	10.31	0	96.41
% Parks	0.1	0.07	0.1	0	0.78

Among nearly 950,000 total trips in the 2016/17 NHTS data, 12% (about 113,000 trips) were work-related. The average WMT per worker per day outside their home tract on the survey day was 0.054 miles. Due to data limitations, this study combines the average with the number of jobs in each tract to estimate the WMT generated by people working in that tract. As WMT is highly correlated with population, the final results of WMT per day per capita across tracts in Texas and Houston are presented in Figure 6. On average, WMT per capita per day is 0.245 miles in the U.S., compared to 0.7 miles in the UK (UK Department for Transport, 2024).

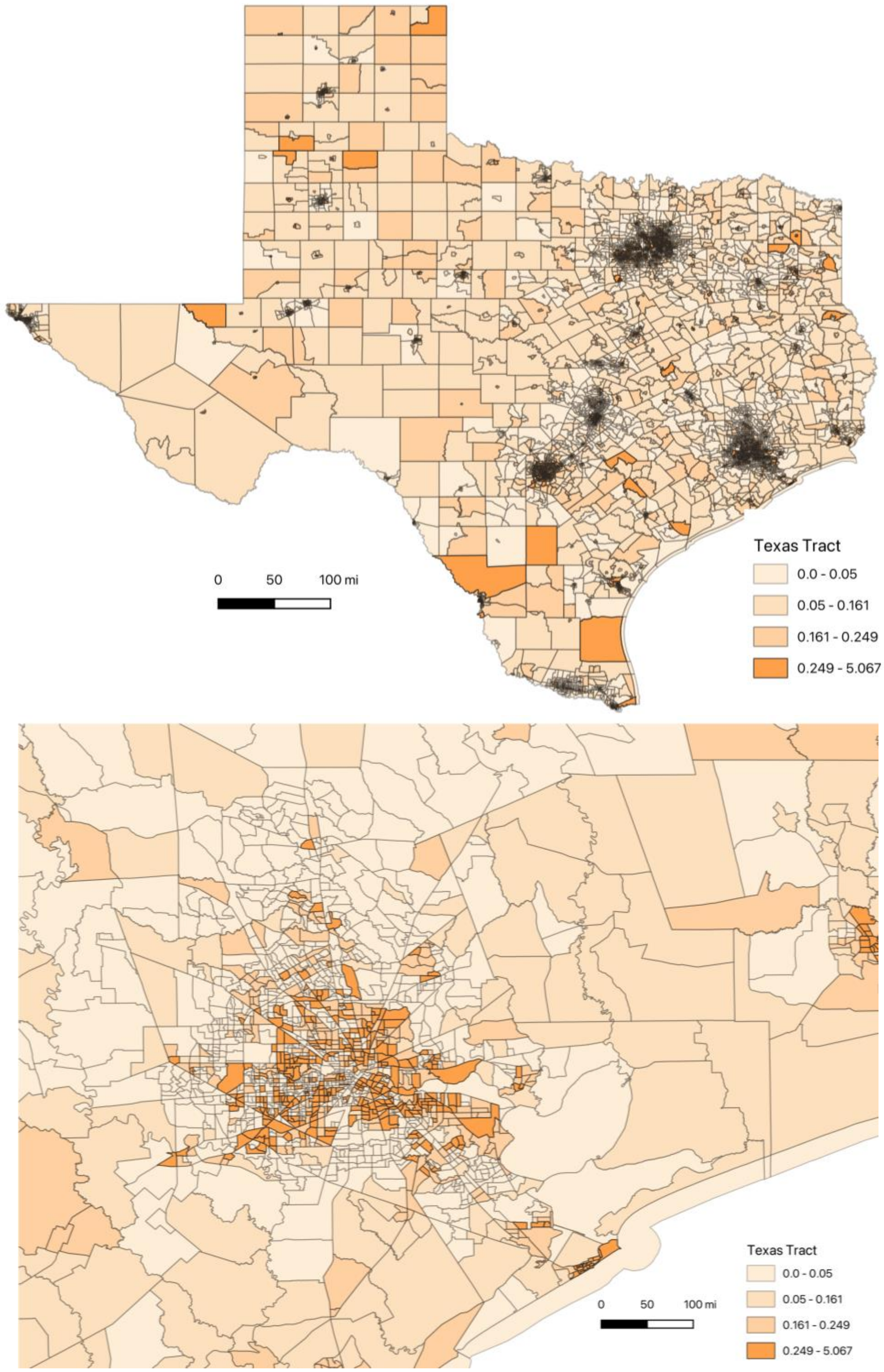
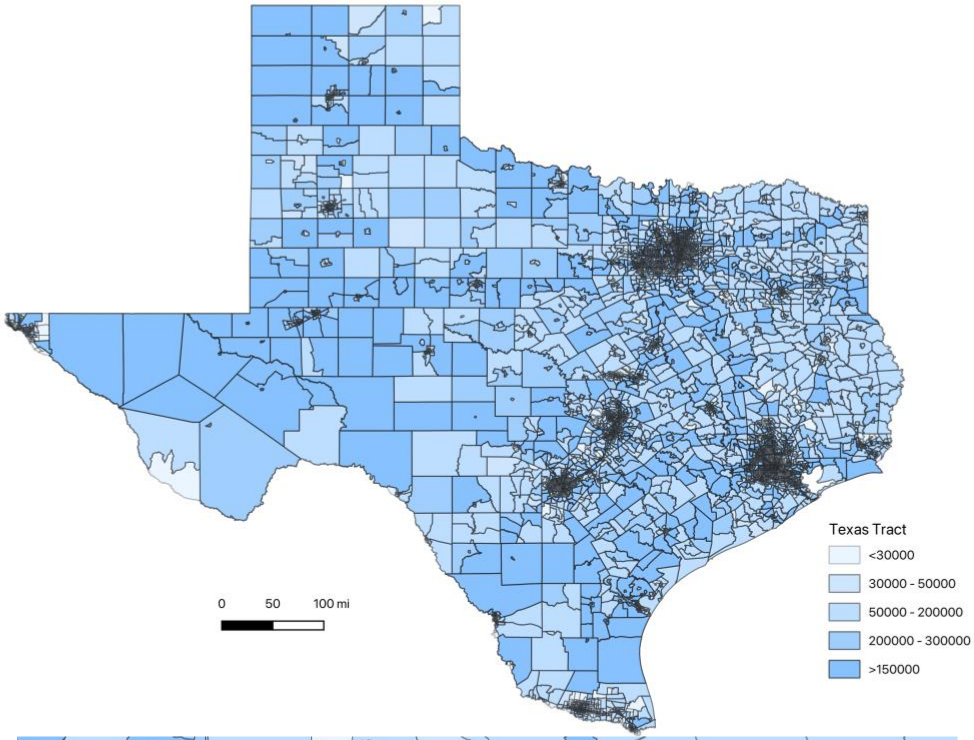


Figure 6: Tract-Level WMT/Day/Capita Estimates in Texas (upper) **and Houston** (bottom)

Tract-level VMT per Day across U.S.

A one-week sample of INRIX data, covering 5.4 million road segments, was used in this analysis. The total VMT in the INRIX sample was 17 million miles. According to TxDOT (2024), the average daily VMT on all roadways in Texas is 825.4 million miles. Therefore, a modification factor of 48.542 was applied, and Figure 7 presents the total VMT in 5,265 tracts across Texas.



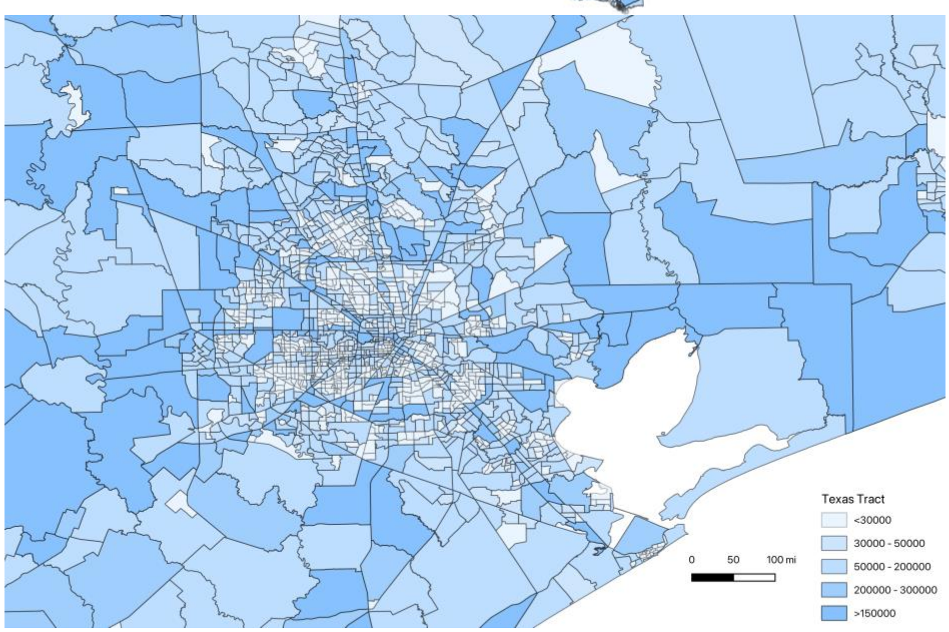


Figure 7: Tract-Level VMT/Day/Capita Estimates in Texas (upper) **and Houston** (bottom)

Pedestrian Crash Rates by Tract in Texas

Across Texas, the average pedestrian crash rate is 0.491 per 1 million WMT per day across all severity levels, and 0.027 per 1 million VMT per day. Table 4 summarizes the total annual number of crashes, as well as crash rates per 1 million WMT and per 1 million VMT per day in the state.

Table 4: Summary Statistics of Pedestrian Crash Rates

Severity Level	Total Number per Year	Crash Rate per 1 million WMT per day	Crash Rate per 1 million VMT per day
Fatal	814	0.270	0.048
Serious Injury	1607	0.533	0.096
Minor Injury	3015	1.000	0.179
Possible Injury	2066	0.686	0.123
Not Injured	595	0.197	0.035

SUMMARY AND CONCLUSION

This paper explores the demographic and land use-related factors affecting Americans' walk trips and walking distances. The results suggest that younger individuals, those with smaller household sizes, no vehicles, and females with higher education and income are more likely to walk, while older males without a vehicle and individuals with higher education levels and household incomes tend to walk longer distances. Demographic and land use variables influence the decision to walk, with demographic factors having a greater impact on walking distance. Pedestrian-related facilities, including green land coverage and network density, play a more significant role in the probability of taking a walk trip. Based on the WMT/person/day model, a U.S.-wide map for WMT/day/capita is generated, with the average WMT for the U.S. being 0.38 miles per person per day. This study also identifies areas where Americans walk more, showing that regions with dense populations and job opportunities tend to attract more walking. Another question addressed in this paper is the identification of the riskiest regions for pedestrians considering their walking frequency, with southern states exhibiting a significantly higher pedestrian death risk compared to northern states. Among all states, Texas is estimated to have the highest pedestrian death rate (ped deaths/WMT).

This study serves as a preliminary analysis of pedestrian safety. Future work will extend the current analysis by incorporating pedestrian crash severity and estimating Bike-Miles Traveled (BMT) to assess cyclist safety. Since cyclists may travel across multiple counties, weighted factors will be included to reflect cross-boundary movement. These additions will enable a more comprehensive understanding of safety concerns for vulnerable road users and support more effective policy implementation.

ACKNOWLEDGEMENTS

The authors thank Waymo for its financial support of this work, and Drs. Kris Kusano and John Scanlon for their insights on this topic.

REFERENCES

Agrawal, A. W., & Schimek, P. (2007). Extent and correlates of walking in the USA. *Transportation Research Part D: Transport and Environment*, *12*(8), 548-563. https://doi.org/10.1016/j.trd.2007.07.005

Alozi, A. R., & Hussein, M. (2022). Evaluating the safety of autonomous vehicle–pedestrian interactions: An extreme value theory approach. *Analytic methods in accident research*, *35*, 100230. https://doi.org/10.1016/j.amar.2022.100230

Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident Analysis & Prevention*, *93*, 147-159. https://doi.org/10.1016/j.aap.2016.05.001

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, *41*(3), 359-364. https://doi.org/10.1016/j.aap.2008.12.014

Arizona Department of Transportation (ADOT). (2024). ADOT 2023 Average Annual Daily Traffic (AADT). Retrieved from https://azgeo-open-data-agic.hub.arcgis.com/maps/51fd91145d034bcd812eb62fd9cf82b2/about

Arizona Department of Transportation (ADOT). (2025). Motor Vehicle Crash Data. Retrieved from https://azdot.gov/tags/crash-data

Baffoe-Twum, E., Asa, E., & Awuku, B. (2023). Estimation of annual average daily traffic (AADT) data for low-volume roads: a systematic literature review and meta-analysis. *Emerald Open Research*, *1*(5). https://doi.org/10.1108/EOR-05-2023-0010

Bernhardt, M., & Kockelman, K. (2021). An analysis of pedestrian crash trends and contributing factors in Texas. *Journal of Transport & Health*, *22*, 101090. https://doi.org/10.1016/j.jth.2021.101090

Blincoe, L., Miller, T., Wang, J.-S., Swedler, D., Coughlin, T., Lawrence, B., Guo, F., Klauer, S., & Dingus, T. (2023, February). *The economic and societal impact of motor vehicle crashes, 2019 (Revised)* (Report No. DOT HS 813 403). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813403.pdf

Buehler, R., & Pucher, J. (2021). The growing gap in pedestrian and cyclist fatality rates between the United States and the United Kingdom, Germany, Denmark, and the Netherlands, 1990–2018. *Transport reviews*, *41*(1), 48-72. https://doi.org/10.1080/01441647.2020.1823521

Buehler, R., Pucher, J., & Bauman, A. (2020). Physical activity from walking and cycling for daily travel in the United States, 2001–2017: Demographic, socioeconomic, and geographic variation. *Journal of Transport & Health*, *16*, 100811. https://doi.org/10.1016/j.jth.2019.100811

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention*, *39*(4), 657-670. https://doi.org/10.1016/j.aap.2006.10.012

California Department of Transportation. (2024). Annual Average Daily Traffic. Retrieved from https://gis.data.ca.gov/datasets/d8833219913c44358f2a9a71bda57f76_0/about

California Department of Transportation. (2025). State Highway Network Lines. Retrieved from https://gis.data.ca.gov/datasets/77f2d7ba94e040a78bfbe36feb6279da_0/explore?location=38.135024%2C-116.362371%2C6.63

Castro-Neto, M., Jeong, Y., Jeong, M. K., & Han, L. D. (2009). AADT prediction using support vector regression with data-dependent parameters. *Expert Systems with Applications*, *36*(2), 2979-2986. https://doi.org/10.1016/j.eswa.2008.01.073

Chen, Y. H., Scanlon, J. M., Kusano, K. D., McMurry, T. L., & Victor, T. (2024). Dynamic benchmarks: spatial and temporal alignment for ADS performance evaluation. *arXiv preprint arXiv:2410.08903*. https://doi.org/10.48550/arXiv.2410.08903

Combs, T. S., Sandt, L. S., Clamann, M. P., & McDonald, N. C. (2019). Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection. *American journal of preventive medicine*, *56*(1), 1-7. https://doi.org/10.1016/j.amepre.2018.06.024

Cottrill, C. D., & Thakuriah, P. V. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention*, *42*(6), 1718-1728. https://doi.org/10.1016/j.aap.2010.04.012

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: journal of the Econometric Society*, 829-844. https://doi.org/10.2307/1909582

Dewitz, J. (2020). National land cover database (NLCD) 2019 products: US Geological Survey data release. https://www.mrlc.gov/data?f%5B0%5D=category%3ALand%20Cover

Dumbaugh, E., Stiles, J., Mitsova, D., & Saha, D. (2024). The most vulnerable user: examining the role of income, race, and the built environment on pedestrian injuries and deaths. *Transportation Research Record*, *2678*(2), 743-752. https://doi.org/10.1177/03611981231175888

Federal Highway Administration. (2017). *2017 National Household Travel Survey*, U.S. Department of Transportation, Washington, DC. Available online: https://nhts.ornl.gov.

Environmental Protection Agency (EPA) (2021). Smart Location Database. Retrieved from https://www.epa.gov/smartgrowth/smart-location-mapping

Freemark, Y., & Jenkins, W. (2022). With US Traffic Fatalities Rising, What Would It Take to Save Lives? *Urban Wire.* https://www.urban.org/urban-wire/us-traffic-fatalities-rising-what-would-it-take-save-lives

Gurmu, S. (1998). Generalized hurdle count data regression models. *Economics Letters*, *58*(3), 263-268. https://doi.org/10.1016/S0165-1765(97)00295-4

Gurmu, S., & Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, *14*(4), 469-477. https://doi.org/10.1080/07350015.1996.10524676

Hu, W., & Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of safety research*, *67*, 37-44. https://doi.org/10.1016/j.jsr.2018.09.009

Hussain, Q., Feng, H., Grzebieta, R., Brijs, T., & Olivier, J. (2019). The relationship between impact speed and the probability of pedestrian fatality during a vehicle-pedestrian crash: A systematic review and meta-analysis. *Accident Analysis & Prevention*, *129*, 241-249. https://doi.org/10.1016/j.aap.2019.05.033

Iosa, M., Fusco, A., Morone, G., & Paolucci, S. (2012). Walking there: environmental influence on walking-distance estimation. *Behavioural brain research*, *226*(1), 124-132. https://doi.org/10.1016/j.bbr.2011.09.007

Klatko, T. J., Agbelie, B. R., Labi, S., Fricker, J. D., & Sinha, K. C. (2016). Estimation and prediction of statewide Vehicle Miles Traveled (VMT) by highway category and vehicle classification. https://doi.org/10.5703/1288284316349

Klatko, T. J., Saeed, T. U., Volovski, M., Labi, S., Fricker, J. D., & Sinha, K. C. (2017). Addressing the local-road VMT estimation problem using spatial interpolation techniques. *Journal of Transportation Engineering, Part A: Systems*, *143*(8), 04017038. https://doi.org/10.1061/JTEPBS.0000064

Kim, K., & Yamashita, E. Y. (2007). Using ak-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii. *Journal of advanced transportation*, *41*(1), 69-89. https://doi.org/10.1002/atr.5670410106

Kim, S., & Coifman, B. (2014). Comparing INRIX speed data against concurrent loop detector stations over several months. *Transportation Research Part C: Emerging Technologies* 49, 59-72. https://doi.org/10.1016/j.trc.2014.10.002

Kockelman, K., Huang, Y., Fakhrmoosavi, F., Perrine, K. A., Paithankar, P., Hawkins, J., ... & Vellimana, M. (2022). *Understanding the Impact of Autonomous Vehicles on Long-Distance Passenger and Freight Travel in Texas* (No. FHWA/TX-23/0-7081-1). University of Texas at Austin. Center for Transportation Research. https://rosap.ntl.bts.gov/view/dot/68788

Lahiri, K., & Xing, G. (2004). An econometric analysis of veterans' health care utilization using two-part models. *Empirical Economics 29*, 431-449. https://doi.org/10.1007/s00181-003-0178-0

Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention 37*(4), 775-786. https://doi.org/10.1016/j.aap.2005.03.019

Lee, J., Abdel-Aty, M., Choi, K., & Huang, H. (2015). Multi-level hot zone identification for pedestrian safety. *Accident Analysis & Prevention*, *76*, 64-73. https://doi.org/10.1016/j.aap.2015.01.006

Loukaitou-Sideris, A., Liggett, R., & Sung, H. G. (2007). Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. *Journal of Planning Education and Research*, *26*(3), 338-351. https://doi.org/10.1177/0739456X06297008

National Highway Traffic Safety Administration. (2014). Motor vehicle traffic fatalities in 2013. *Publication DOT HS*, *812*, 169. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812169

National Highway Traffic Safety Administration. (2024). Early estimate of motor vehicle traffic fatalities in 2023. *Publication DOT HS*, *813*, 561. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813561

National Highway Traffic Safety Administration. (2025). Early estimate of motor vehicle traffic fatalities in 2024. *Publication DOT HS*, *813*, 710. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813710

OpenStreetMap. (2025). Retrieved from https://download.geofabrik.de/north-america/us/california.html

Olszewski, P., Dybicz, T., Jamroz, K., Kustra, W., & Romanowska, A. (2018). Assessing highway travel time reliability using Probe Vehicle Data. *Transportation Research Record*, *2672*(15), 118-130. https://doi.org/10.1177/0361198118796716

Rahman, M., Kockelman, K. M., & Perrine, K. A. (2022). Investigating risk factors associated with pedestrian crash occurrence and injury severity in Texas. *Traffic injury prevention*, *23*(5), 283-289. https://doi.org/10.1080/15389588.2022.2059474

Sabel, C. E., Bartie, P., Kingham, S., & Nicholson, A. (2006). Kernel density estimation as a spatial-temporal data mining tool: exploring road traffic accident trends. In *Proceedings GISRUK'06* (pp. 191-196). https://researchprofiles.canberra.edu.au/en/publications/kernel-density-estimation-as-a-spatial-temporal-data-mining-tool-

Schneider, R. J., Ryznar, R. M., & Khattak, A. J. (2004). An accident waiting to happen: a spatial approach to proactive pedestrian planning. *Accident Analysis & Prevention*, *36*(2), 193-211. https://doi.org/10.1016/S0001-4575(02)00149-5

Schneider, R. J., Proulx, F. R., Sanders, R. L., & Moayyed, H. (2021). United States fatal pedestrian crash hot spot locations and characteristics. *Journal of transport and land use*, *14*(1), 1-23. https://www.jstor.org/stable/48646174

Sharma, A., Ahsani, V., & Rawat, S. (2017). Evaluation of opportunities and challenges of using INRIX data for real-time performance monitoring and historical trend assessment. https://rosap.ntl.bts.gov/view/dot/54760

Sharma, S., Lingras, P., Xu, F., & Kilburn, P. (2001). Application of neural networks to estimate AADT on low-volume roads. *Journal of Transportation Engineering*, *127*(5), 426-432. https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426)

Sohrabi, S., Khodadadi, A., Mousavi, S. M., Dadashova, B., & Lord, D. (2021). Quantifying the automated vehicle safety performance: A scoping review of the literature, evaluation of methods, and directions for future research. *Accident Analysis & Prevention*, *152*, 106003. https://doi.org/10.1016/j.aap.2021.106003

Stutts, J. C., & Hunter, W. W. (1998). Police reporting of pedestrians and bicyclists treated in hospital emergency rooms. *Transportation Research Record*, *1635*(1), 88-92. https://doi.org/10.3141/1635-12

Sun, X., & Das, S. (2015). *Developing a method for estimating AADT on all Louisiana roads* (No. FHWA/LA. 14/548). Louisiana Transportation Research Center. https://www.researchgate.net/profile/Subasish-Das/publication/321316094_Developing_a_Method_for_Estimating_AADT_on_all_Louisiana_Roads/links/5a1c406e0f7e9bd1be5c57ee/Developing-a-Method-for-Estimating-AADT-on-all-Louisiana-Roads.pdf

Susilawati, S., Wong, W. J., & Pang, Z. J. (2023). Safety effectiveness of autonomous vehicles and connected autonomous vehicles in reducing pedestrian crashes. *Transportation research record*, *2677*(2), 1605-1618. https://doi.org/10.1177/03611981221108984

Texas Department of Transportation (2024). Pocket Facts. Retrieved from https://www.txdot.gov/content/dam/docs/division/cmd/txdot-pocket-facts.pdf

Texas Department of Transportation (TxDOT). (Accessed May 2025). TxDOT Roadway Inventory. Retrieved from https://gis-txdot.opendata.arcgis.com/datasets/txdot-roadway-Inventory/explore

Transportation Injury Mapping System (TIMS). (2025). Retrieved from https://tims.berkeley.edu/

Transport Canada. (2025). Canadian Motor Vehicle Traffic Collision Statistics: 2023. Retrieved from https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics/2023/canadian-motor-vehicle-traffic-collision-statistics-2023

Turner, S., Francis, T., Roozenburg, A., & Transport, N. L. (2006). *Predicting accident rates for cyclists and pedestrians* (p. 72). Wellington: Land Transport New Zealand. https://nzta.govt.nz/assets/resources/research/reports/289/docs/289-Predicting-accident-rates-for-cyclists-and-pedestrians.pdf

UK Department for Transport (2024). Walking and Cycling Statistics. Retrieved from https://www.gov.uk/government/statistics/walking-and-cycling-statistics-england-2023/walking-and-cycling-statistics-england-introduction-and-main-findings-national-travel-survey#:~:text=The%20latest%20data%20shows%20that,their%20distance%20travelling%20by%20walking

U.S. Census Bureau (2020). American Community Survey 5-Year Estimates: Comparison Profiles 5-Year. Retrieved from http://api.census.gov/data/2019/acs/acs5

U.S. Census Bureau (2023). ACS Public Use Microdata Sample (PUMS). Retrieved from https://www.census.gov/programs-surveys/acs/microdata/access.html

U.S. Census Bureau. (n.d.). *U.S. and World Population Clock*. U.S. Department of Commerce. Retrieved November 12, 2024, from https://www.census.gov/popclock/

U.S. Department of Transportation (2018). Highway Statistics Series. Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics.cfm

Utriainen, R. (2020). The potential impacts of automated vehicles on pedestrian safety in a four-season country. *Journal of Intelligent Transportation Systems*, *25*(2), 188-196. https://doi.org/10.1080/15472450.2020.1845671

Vellimana, M., & Kockelman, K. (2023). Darkness and Death in the US: Walking Distances Across the Nation by Time of Day and Time of Year. *Transport Findings*. https://findingspress.org/article/89418-darkness-and-death-in-the-u-s-walking-distances-across-the-nation-by-time-of-day-and-time-of-year

Walk Score. (2025) Accessed on April 2025. Retrieved from https://www.walkscore.com/score/

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, *60*, 71-84. https://doi.org/10.1016/j.aap.2013.07.030

Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, *41*(1), 137-145. https://doi.org/10.1016/j.aap.2008.10.001

Williams, T. A., Chigoy, B., Borowiec, J. D., & Glover, B. (2016). Methodologies used to estimate and forecast vehicle miles traveled (VMT). https://rosap.ntl.bts.gov/view/dot/32689

World Health Organization. (2023). *Road traffic injuries.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American journal of preventive medicine*, *43*(1), 11-19. https://doi.org/10.1016/j.amepre.2012.03.015

Zegeer, C. V., & Bushell, M. (2012). Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis & Prevention*, *44*(1), 3-11. https://doi.org/10.1016/j.aap.2010.12.007

Zhao, B., Zuniga-Garcia, N., Xing, L., & Kockelman, K. M. (2024). Predicting pedestrian crash occurrence and injury severity in texas using tree-based machine learning models. *Transportation Planning and Technology*, *47*(8), 1205-1226. https://doi.org/10.1080/03081060.2023.2216202

Zheng, F., & Van Zuylen, H. (2013). Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, *31*, 145-157. https://doi.org/10.1016/j.trc.2012.04.007