CPBS
Center for Pedestrian and Bicyclist Safety

A Tier-1 University Transportation Center

The Role of Urban Form and Demographics in Pedestrian and Bicycle Safety

July 2024

A Report From the
Center for Pedestrian and Bicyclist Safety

Asad Khattak
The University of Tennessee-Knoxville

Zeinab Bayati
The University of Tennessee-Knoxville

About the Center for Pedestrian and Bicyclist Safety (CPBS)

The Center for Pedestrian and Bicyclist Safety (CPBS) is a consortium of universities committed to eliminating pedestrian and bicyclist fatalities and injuries through cutting-edge research, workforce development, technology transfer, and education. Consortium members include: The University of New Mexico; San Diego State University; The University of California Berkeley; The University of Tennessee Knoxville; and The University of Wisconsin Milwaukee. More information can be found at: https://pedbikesafety.org

Disclaimer

Acknowledgments

This study was funded, partially or entirely, by a grant from the Center for Pedestrian and Bicyclist Safety (CPBS), supported by the U.S. Department of Transportation (USDOT) through the University Transportation Centers program. The authors would like to thank CPBS and the USDOT for their support of university-based research in transportation and especially for the funding provided in support of this project. We sincerely thank Dr. Latif Patwary, Dr. Iman Mahdinia, and Ms. Nastaran Moradloo for their assistance with identifying databases and input in conducting the analysis.

TECHNICAL DOCUMENTATION

<table>
<tr><td>1. Project No.
23UTK04</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
The Role of Urban Form and Demographics in Pedestrian and Bicycle Safety</td><td>5. Report Date
July 2024</td></tr>
<tr><td>6. Performing Organization Code
N/A</td></tr>
<tr><td colspan="2">7. Author(s)
Asad Khattak https://orcid.org/0000-0002-0790-7794
Zeinab Bayati https://orcid.org/0000-0002-5815-4657</td><td>8. Performing Organization Report No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Center for Pedestrian and Bicyclist Safety
Centennial Engineering Center 3020
The University of New Mexico
Albuquerque, NM 87131</td><td>10. Work Unit No. (TRAIS)</td></tr>
<tr><td>11. Contract or Grant No.
69A3552348336</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
United States of America
Department of Transportation
Office of Research, Development, and Technology (RD&T)</td><td>13. Type of Report and Period Covered
Final Report – June 2023 to May 2024</td></tr>
<tr><td>14. Sponsoring Agency Code
USDOT OST-R</td></tr>
<tr><td colspan="3">15. Supplementary Notes
Report available at the CPBS website [https://pedbikesafety.org] and DOI https://doi.org/10.21949/mn1x-v461</td></tr>
<tr><td colspan="3">16. Abstract:
The rising number of fatal crashes involving pedestrians and cyclists underscores the urgent need for enhanced safety interventions, especially in burdened communities. This study, aligned with Vision Zero goals, explores the temporal and geographic patterns of fatal pedestrian and bicycle crashes, specifically focusing on burdened communities. It identifies trends and crash sites across the U.S., examining the correlation between communities' transportation burden, health vulnerability, environmental burden, social vulnerability, disaster risk burden, and the rise in pedestrian and bicycle fatalities in 84,443 census tracts. Using Florida as a case study due to its significant mix of burdened and non-burdened communities, the research uses a comprehensive methodology, including time series Difference-in-Difference analysis, spatial analysis, and Zero-Inflated Negative Binomial regression to uncover relationships. The study integrates data from the US Census, the Fatality Analysis Reporting System, and other datasets. Findings reveal a substantial increase in pedestrian/bicycle fatalities, especially in burdened communities, over the past decade, and a strong correlation between higher street network density and increased pedestrian/bicycle crashes. The insights are vital for policymakers and transportation engineers in developing long-term safety interventions.</td></tr>
<tr><td colspan="2">17. Key Words
Pedestrians and Bicyclists, Burdened Communities, Geographically Weighted Regression, Time Series Modeling, Fatal Crashes</td><td>18. Distribution Statement
No restrictions. This document is available through the University Transportation Centers program of the U.S. Department of Transportation.</td></tr>
</table>

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	55	

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

The Role of Urban Form and Demographics in Pedestrian and Bicycle Safety

A Center for Pedestrian and Bicyclist Safety Research Report

May 2024

Asad J. Khattak
Department of Civil and Environmental Engineering
The University of Tennessee-Knoxville

Zeinab Bayati
Department of Civil and Environmental Engineering
The University of Tennessee-Knoxville

Table of Contents

List of Figures

List of Tables

Acronyms, Abbreviations, and Symbols

AIC	Akaike's Information Criterion
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
DiD	Difference in Difference
FARS	Fatality Analysis Reporting System
GWZINBR	Geographically weighted Zero-Inflated Negative Binomial Regression
GWR	Geographically weighted Regression
KDE	Kernel Density Estimate
NaNDA	National Neighborhood Data Archive
NRSS	National Roadway Safety Strategy
STL	Trend decomposition using Loess
USDOT	U.S. Department of Transportation
VMT	Vehicle Miles Traveled
ZINB	Zero-Inflated Negative Binomial

Abstract

The increasing number of fatal crashes involving pedestrians and bicycles highlights the critical need for improved safety interventions. The U.S. Department of Transportation (USDOT) is particularly interested in addressing safety issues in burdened communities to advance Vision Zero goals. This study investigates both temporally and geographically fatal pedestrian/bicycle crashes. Initially, it identifies trends in pedestrian/bicycle crashes and frequent crash locations in the U.S. Importantly, the study examines the correlation between various indices, including transportation insecurity, health vulnerability, environmental burden, social vulnerability, and disaster risk burden, and the rising pedestrian/bicycle fatal crashes (N=84,443 census tracts). Florida was selected as a case study because it has many such crashes and a mix of burdened and non-burdened communities. The study applies a comprehensive methodology, including temporal analysis, i.e., time series Difference-in-Difference analysis, and spatial analysis using kernel density and a Zero-Inflated Negative Binomial (ZINB) global and local geographically weighted regression (GWR). The approach integrates multiple datasets, creating a unique database from Census and the Fatality Analysis Reporting System (FARS) data supplemented by datasets that provide critical variables. The study documents a significant increase of over 50% in pedestrian/bicycle fatal crashes in the U.S. over the past decade. It highlights that burdened areas and higher street network density correlate with increased pedestrian and bicycle fatal crash rates. Additionally, spatial research on the indices highlights strong relationships between crashes and the measures, highlighting the need for improved transportation interventions in certain areas. The findings offer valuable insights for policymakers and transportation engineers, enabling them to develop effective transportation systems and implement policies to enhance transportation safety.

Keywords: Burdened Communities, Pedestrians and Bicyclists, Geographically Weighted Regression, Time Series Modeling, Fatal Crashes

Executive Summary

Pedestrians and cyclists are facing a high and increasing rate of crashes. To address this, the USDOT is investing resources to eliminate road traffic fatalities and severe injuries. The study explores how various spatial indices at the Census tract level, characterized by transportation insecurity, health vulnerability, social vulnerability, disaster risk, and environmental burden area, are associated with pedestrian and bicycle crashes. This comprehensive approach enables policies that address vulnerabilities, creating a safer and more accessible transportation system.

To implement interventions, this study explores risky locations. The study creates a unique database and applies several analytical approaches: first, descriptive analysis provides a concise summary of the data, enabling an understanding of characteristics, patterns, and safety distributions in burdened and non-burdened locations. Key trends are identified, and hypotheses regarding safety are tested, ensuring a robust foundation for further research. Second, time-series analysis delves into national and regional trends in fatal crashes over a decade, utilizing various methods to pinpoint frequent crash locations. Techniques applied include kernel density mapping, trend decomposition using the Loess (STL) method combined with Autoregressive Integrated Moving Average (ARIMA), and Difference-in-Difference (DiD) analysis. The scope of analysis spans multiple scales, encompassing state-wide and regional (USDOT regions) assessments.

Spatial variation (heterogeneity) is expected when exploring the relationships between the safety and burdened communities. The spatial analysis phase underscores the importance of differentiating between global and local analytical models to tackle spatial heterogeneity in safety outcomes. The geographically weighted zero-inflated negative binomial regression (GWZINBR) method is applied, providing a new application for this model. This approach leverages a newly developed code tailored for GWR, allowing for a nuanced analysis of the factors associated with pedestrian and bicycle crashes. The study focuses on assessing the associations of spatial indices alongside land use and traffic variables across various census tracts. To demonstrate, this method is applied to Florida, which is recognized for its high pedestrian and bicycle fatalities and substantial variations in socio-demographics.

Developing a unique database includes integrating four datasets that comprehensively analyze fatal pedestrian and bicycle crashes over ten years. The datasets contain a GIS shape file from USDOT, the FARS database, land cover data from the NaNDA dataset, and annual Vehicle Miles Traveled (VMT) from the REPLICA dataset. Using ArcGIS, crash locations from FARS are accurately aligned with the centroids of the census tracts to ensure precise spatial locations. This integration allows for comprehensive descriptive and time-series analyses across all census tracts, while the modeling phase concentrates specifically on Florida. This methodological approach enables a detailed exploration of pedestrian and bicycle safety dynamics, facilitating the development of targeted safety initiatives.

The analysis reveals that the spatial indicators are positively associated with fatal pedestrian/bicycle crashes (N=84,443 census tracts). Furthermore, fatal crashes are positively associated with high-intensity development in census tracts. Temporally, the study documents an increasing trend in crashes over the years, with certain regions showing more pronounced increases, as evidenced by time series analysis results. The investigation utilizes DiD methods to illustrate how increased street network density is associated with higher fatal pedestrian and bicyclist crashes. This suggests that significant infrastructure additions, e.g., arterials or local roads, may increase exposure and worsen pedestrian and bicycle safety. Furthermore, the analysis across USDOT regions reveals a particularly steep increase in Region 4, supported by kernel density mapping results. Notably, the study finds greater increases in fatal crashes for burdened communities, suggesting a more substantial deterioration of pedestrian/bicycle safety.

The analysis reveals heterogeneity between safety and community factors across spatial contexts in Florida. Specifically, though significant globally, the cost index has different local relationships with the safety of pedestrians and bicyclists. The findings indicate that transportation costs and health indices exhibit significant spatial heterogeneity. For example, the transportation cost burden is more strongly associated with crashes in some parts of Florida (Southern regions) than others, underscoring the importance of understanding the nuanced spatial distribution of safety correlates to designing local interventions.

Overall, this study's findings offer a comprehensive view of fatal pedestrian and bicycle crashes over the decades, providing both temporal and spatial analyses. These insights are crucial for creating a safer environment where safety measures prioritize all individuals. The detailed understanding gained from this research can inform more effective policymaking and prompt action in areas identified as particularly dangerous, helping to prevent severe accidents and enhance overall road safety.

Introduction

The number of fatal pedestrian and bicycle crashes in the U.S. has increased significantly from 5,589 to 8,427 between 2012 and 2021. Each number represents a life lost and the affected community, emphasizing the need to take quick action to improve road safety and stop more tragedies. As a result, the USDOT is working to reduce this concern by developing the National Road Safety Strategy (NRSS), which includes several steps to minimize these fatalities and is a crucial project in this endeavor. This fits with the long-term objectives of Vision Zero, which aims to achieve zero road traffic fatalities and serious injuries by developing more egalitarian and safe transportation systems for all users of the roads. (Evenson et al., 2023).

As transportation planners and engineers emphasize safety, ensuring access to transportation networks for all population segments is crucial. With this strategy, USDOT attempts to involve all groups in the planning process. In this regard, spatial indices that include transportation insecurity, health vulnerability, social vulnerability, disaster risk burden, and environmental burden can be used as spatial indices. If a census tract fulfills or surpasses the predetermined standards in four or more categories, it can be classified as a burdened community.

By utilizing the census tract-level data, the main objective of this study is to solve transportation safety issues. This study uses spatial indices and other critical factors, such as land use, traffic-related features, and sociodemographic characteristics, to investigate their associations with fatal pedestrian/bicycle crashes. Notably, the significant component of this study is the investigation of geographical heterogeneity in safety outcomes, emphasizing the variation of safety metrics across different regions. In this regard, the analyses undertaken in this study include:

Descriptive analysis: This analysis provides a concise summary of the data, enabling an understanding of its characteristics, patterns, and distributions. This analysis identifies key trends and helps refine the initial hypothesis, ensuring a robust foundation for further research.

Time-series analysis: This section examines national and regional trends in fatal crashes, focusing on the fourth region of USDOT (excluding Puerto Rico and the Virgin Islands) to identify crash locations over ten years.

Modeling Phase: This section uses a global and local model to address spatial heterogeneity and non-stationarity in safety outcomes. The GWZINBR method is applied to achieve this. This innovative approach, which has not been previously applied, involves a newly developed code for this specific application of GWR. Our analysis specifically examined whether higher fatal crash rates are associated with certain urban forms, demographic profiles, and spatial indices across census tracts. This method is applied to Florida, which has a notably high pedestrian/bicycle fatality rate.

The analysis addresses the following questions:

1. What is the association between pedestrian and bicycle crashes, spatial indices, land use/urban form, and demographics?
2. Is spatial heterogeneity present when analyzing pedestrian and bicycle safety?
3. What safe systems interventions can effectively address pedestrian and bicycle safety?

By answering these concerns, this study helps improve safety in transportation systems, guaranteeing safer and more accessible travel for pedestrians and bicyclists, particularly in burdened regions. This research hopes to contribute toward the vision zero goal, giving optimism for a day when pedestrian/bicycle fatal crashes will be eliminated or significantly reduced.

Literature Review

Worldwide, pedestrians and bicyclists are more likely than car occupants to be involved in fatal crashes, according to the World Health Organization, which is a global concern. This is consistent with many studies in this area, one of which represents the growing trend of pedestrian/bicycle fatalities in the U.S. between 2009 and 2016, highlighting the critical need for improved safety protocols for pedestrians and bicyclists worldwide (Al-Mahameed et al., 2019; Chandran et al., 2012; Edirisinghe et al., 2014; Goel et al., 2018; Olszewski et al., 2019; Sosik-Filipiak & Osypchuk, 2023; Vanlaar et al., 2016). To lower traffic-related injuries and fatalities, comprehensive strategies are outlined in the USDOT's NRSS to make all road users' experiences safer (Evenson et al., 2018; Fleisher et al., 2016). Similarly, global efforts to increase the safety of transportation systems for pedestrians and bicyclists align with programs like San Francisco's policy framework for reducing traffic deaths through infrastructure reform (Kronenberg et al., 2019). Many studies that advocate for better pedestrian/bicycle infrastructure endorse these projects as critical initiatives in combating the rising trend of bicycle and pedestrian fatalities.

This study aligns with related research showing that some areas are more prone to fatal crashes due to inadequate infrastructure for pedestrians/bicyclists (Al-Mahameed et al., 2019). Furthermore, (Cottrill & Thakuriah, 2010) confirms that the frequency of pedestrian crashes is four times higher in poor neighborhoods—a statistic that defies explanation based on factors like age, education, English proficiency, population density, and Lower car ownership rates. This emphasizes the need for a comprehensive strategy to address road safety that considers the underlying socio-demographics. To this end, trends and correlates of pedestrian and bicyclist crashes have been identified through spatial and temporal modeling, as detailed below.

Spatial Dimensions of pedestrian and bicyclist Safety

Safety initiatives can significantly reduce pedestrian mortality, as evidenced by a 54% reduction in pedestrian fatalities in New York City from 1990 to 2006 (Ukkusuri et al., 2011). Global geographical techniques support these findings. For instance, a multivariate spatial analysis in Houston, Texas, identified critical factors affecting pedestrian collisions, particularly in communities with a high concentration of Black residents (Haddad et al., 2023). Additionally, sophisticated spatial models have pinpointed regions with consistently high injury rates due to heavy traffic and social dispersion (DiMaggio, 2015).

Various local models, such as spatially weighted panel logistic regression, have been used to assess the associations of environmental conditions, vehicles, and human factors on traffic injury severity. These models highlight the importance of design safety, especially in urban areas with unique challenges like pedestrian access on islands and bus stops (Xiao et al., 2019). The GWR model was also used to investigate Novi Sad, a city in northern Serbia. The dependent variables observed in this study include all types of pedestrian crashes over five years, from 2015 to 2019 (Pljakic et al., 2022).

Temporal Dimensions of pedestrian and bicyclist Safety

Temporal analysis in pedestrian/bicycle safety studies examines both long-term trends and short-term fluctuations. While long-term trends provide insights into the effects of urban development, policy changes, and socioeconomic shifts on safety over the years (Robert B. Noland & Quddus, 2004), studies also explore short-term fluctuations like daily variations. This helps identify specific patterns and shifts in risk factors, offering targeted insights for immediate interventions (Mokhtarimousavi, 2019; Mokhtarimousavi et al., 2020). Additionally, time-series analyses shed light on seasonal variations and peak times for crashes, further enhancing our understanding of risk factors and intervention effectiveness. However, it is important to note that temporal analysis is not the primary focus of these studies but rather a component used to confirm broader safety findings.

Despite extensive research on pedestrian/bicycle crashes, significant gaps and unexplored areas persist, underscoring the need for further study. In particular, exploring the association between pedestrian/bicycle crashes in burdened communities and correlations with urban form and demographics can reveal complex interdependencies. Previous studies have provided broader initial insights, focusing on the positive correlation of spatial indices and high-density land use with pedestrian safety globally. However, transportation differences in safe mobility need further investigation, and how differences in urban infrastructure and access across regions heighten them. In this regard, exploring spatial heterogeneity within burdened communities is essential. Furthermore, time series analysis is needed to obtain more detailed and comprehensive insights into the trends and patterns of pedestrian and bicycle crashes over time. This study will address these gaps while realizing that additional work is needed to fully understand pedestrian and bicyclist safety risks, especially in burdened communities, and explore appropriate solutions.

Data and Methodology

This study integrates two main datasets to analyze factors influencing pedestrian/bicycle fatal crashes: a GIS shape file from the USDOT and the FARS (National Highway Traffic Safety Administration, 2022), which provides fatal crashes for pedestrians/bicycles over ten years from 2012 to 2021. These datasets are meticulously combined using the crash locations from FARS, aligned with the centroids of each census tract using ArcGIS to ensure accurate spatial correlation. Two additional datasets are included: NaNDA (Ailshire et al., 2023), adding land cover data, and REPLICA (REPLICA, 2022), providing essential traffic variables. Each dataset is described below.

Spatial Index Dataset

This study uses indices from the 2020 census tract dataset that characterize space across five critical areas: Transportation Insecurity, Disaster Risk Burden, Environmental Burden, Health Vulnerability, and Social Vulnerability. Each index is created by multiple sub-variables, as demonstrated in Figure 1. The links between these indices and fatal pedestrian/bicycle crashes are provided in Table 1. A census tract achieving a score in the 65th percentile is designated as burdened. This dataset also details other features for each census tract, such as rural and urban attributes and total population.

The transportation index includes three main categories: Transportation safety, Transportation cost, and Transportation access. In our analysis, the Transportation Safety sub-variable, which consists of the number of traffic fatalities per 100,000 people, was excluded from the transportation insecurity index to prevent using safety measures as explanatory variables, given that the dependent variable is the number of fatal pedestrian/bicycle crashes in the model. However, the descriptive analysis is based on the complete index. Figure 2 also represents the variables creating the transportation access and cost indices.

FARS Dataset

This dataset derives the overall number of fatal crashes involving pedestrians/bicycles for each year and census tract. Two additional variables are also retrieved from the FARS dataset using Python, emphasizing Florida. These are the percentages of crashes involving drinking alcohol that occurred in work zones detailed in the descriptive analysis.

NaNDA Dataset

The NaNDA dataset, which includes precise land use information based on the most recent census, is used in this study. To avoid correlation issues in the modeling phase, this study only uses one variable from the dataset: the total length of the street segments within each census tract. Additionally, the variable representing street network density across all census tracts for the years 2020 and 2010, derived from this data, is utilized in the DiD methodology section. It offers insightful information about how transportation infrastructure affects pedestrian/bicycle safety.

REPLICA dataset

By adding up all the weekly VMTs for 2022, this dataset calculates the annual VMT for every census tract to use in the modeling phase. Collecting information on this variable is essential because it estimates the amount of traffic exposure, facilitates the creation of focused safety initiatives and infrastructure improvements, and provides a more accurate assessment of the risks associated with traffic.

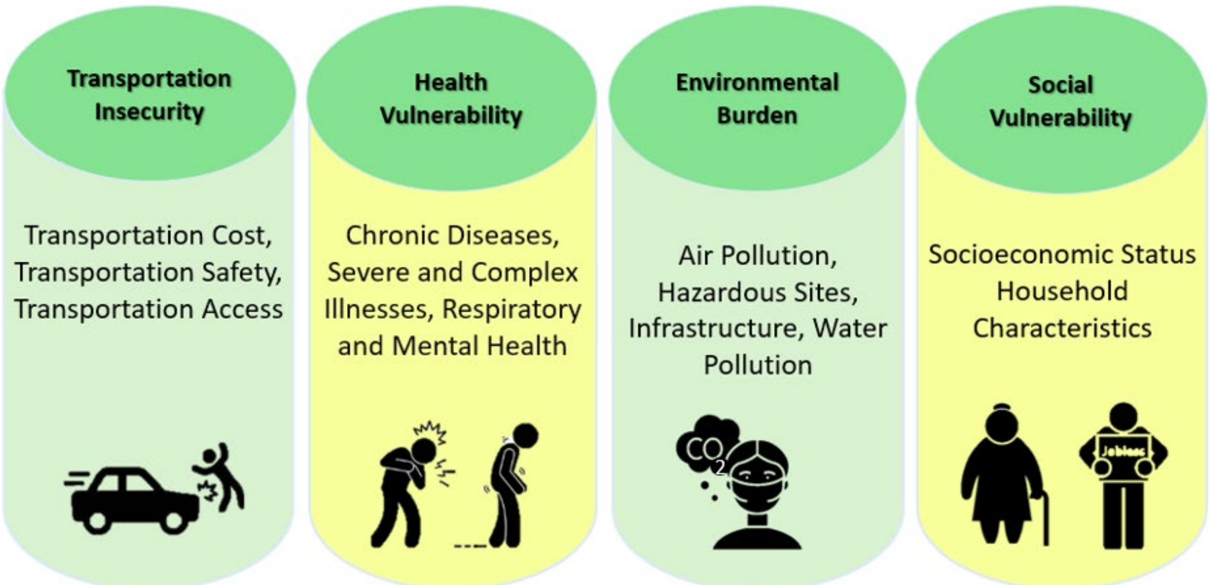


Figure 1. List of variables used to create the final spatial indices

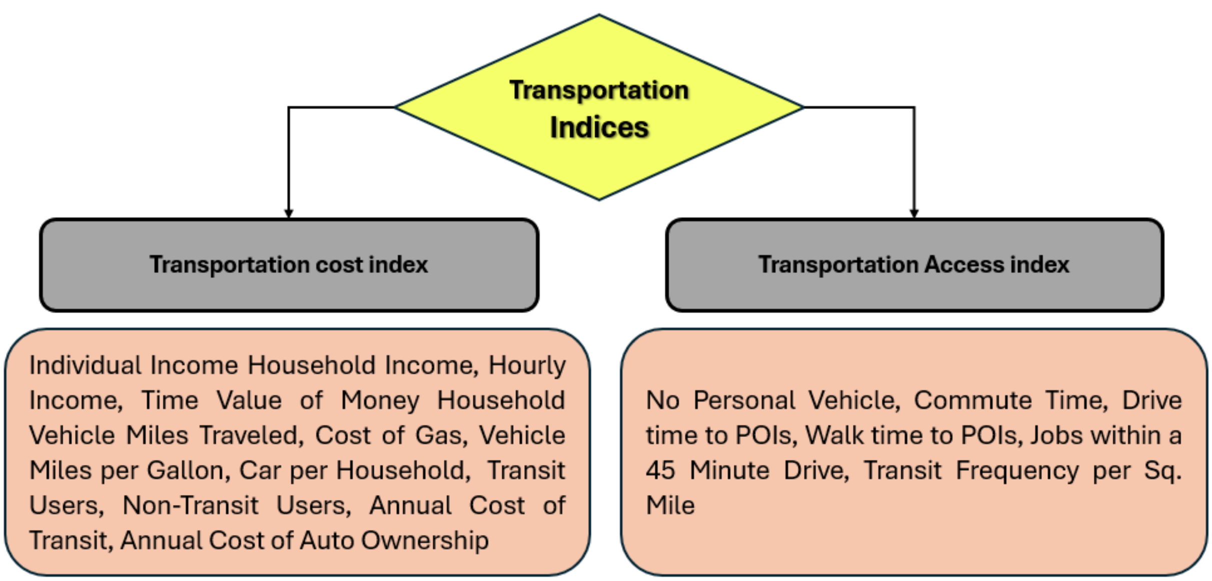


Figure 2. List of variables creating the transportation Indices used in this study

Table 1. Indices Linked to fatal pedestrian/bicycle crashes

Index		Sub variables	Plausible links with pedestrian/bicycle fatal crashes
	Transportation Insecurity	Transportation Cost, Transportation Safety, Transportation Access	Burdened communities may lack sidewalks, pedestrian crossings, and bicycle lanes (infrastructure). When pedestrians and bicyclists are forced to share the road space with vehicles, they are at a higher risk of crashes and fatalities.
	Disaster Risk Burden	Air Pollution, Hazardous Sites, Infrastructure, Water Pollution	The disaster risk burden due to extreme weather events (e.g., hurricanes, flooding) can impact the hazards and risks pedestrians and bicyclists face, especially during such events.
	Environmental Burden	Hazard Risk, Impervious Surfaces	Communities face environmental burdens, e.g., reduced visibility on roadways and higher crash risks for pedestrians and bicyclists. Additionally, more hazardous sites might exist in these regions, further elevating the risk for pedestrians and bicyclists.
	Health Vulnerability	Chronic Diseases, Severe and Complex Illnesses, Respiratory and Mental Health	Prevalence of poor health conditions in some communities, e.g., visual and cognitive disabilities, can limit the ability to use transportation infrastructure and result in higher risks for pedestrians and bicyclists.
	Social Vulnerability	Socioeconomic Status Household Characteristics	Burdened communities often have low education and income levels and fewer social services, which can increase travel risk by pedestrian and cycling modes. Furthermore, age and disability are examples of household characteristics that can raise risk as they can limit movement and awareness, making navigation more challenging.

The methodology of this study is structured into three distinct sections, each leveraging different software tools to enhance the analysis. These tools include R language for statistical analysis, Python for data processing, ArcGIS for spatial analysis, and Excel for data organization and preliminary analysis. Figure 3 provides a detailed illustration of the study's framework, outlining the methodologies and variables involved in the analysis. The sections are as follows.

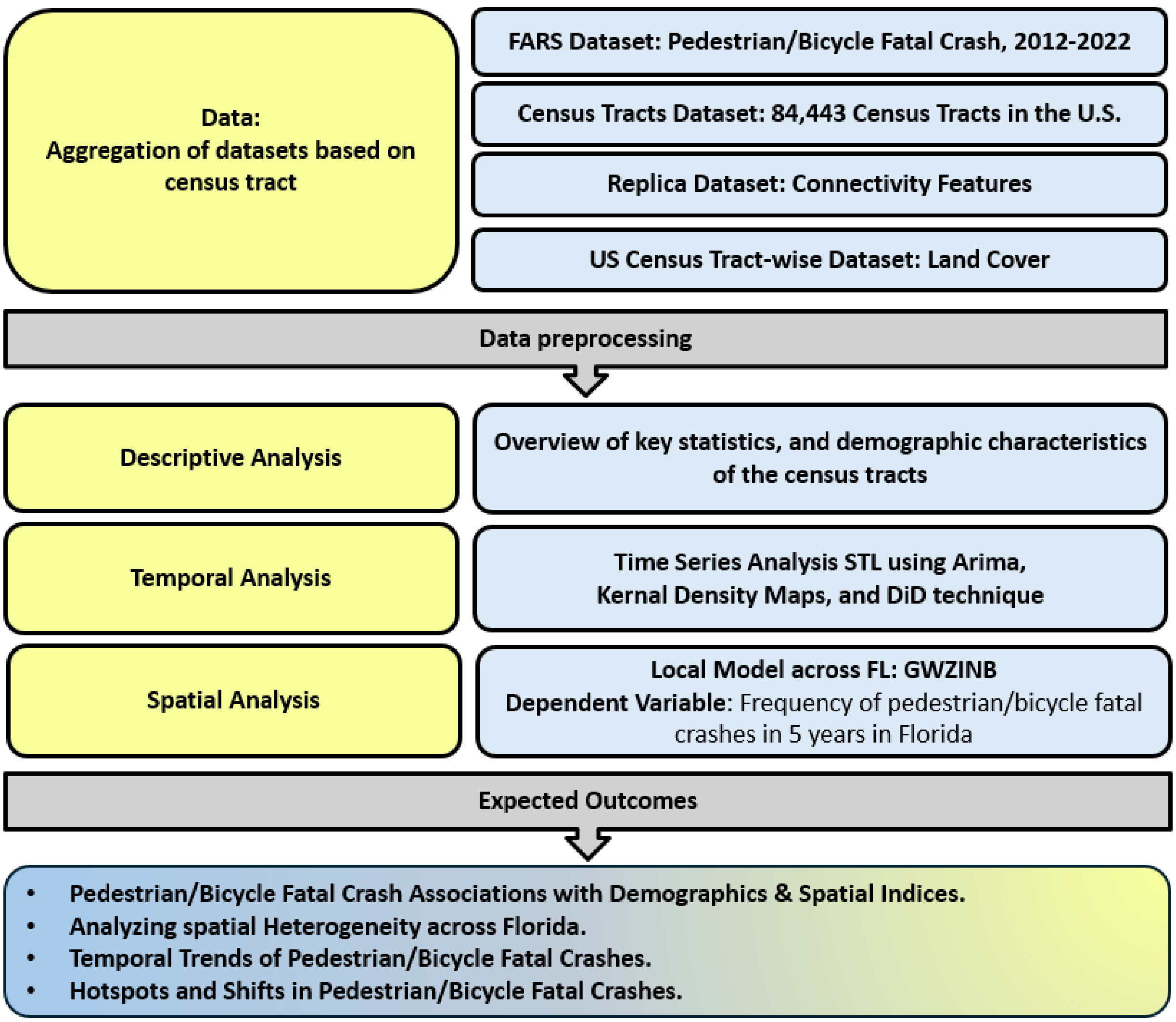


Figure 3. Study framework

Descriptive Analysis

Descriptive analysis condenses the key characteristics of the information on fatal pedestrian/bicycle crashes and reveals distributions of variables. Histograms and scatter plots can provide insights and help identify/remove potential data anomalies. This analysis is crucial for understanding data characteristics across the entire U.S., specifically in Florida, identifying trends, assessing data quality, and establishing a baseline for more complex analyses. Therefore, this methodology section helps refine hypotheses and guide subsequent analytical phases, including detailed spatial analyses.

Time Series Analysis

This section focuses on time series analysis, typically used to identify patterns and trends in data over time and to forecast future events based on historical patterns. This section is conducted at the national level, encompassing the entire U.S. Time series analysis, which employs hierarchical data to analyze variables over specific periods, assessing dependencies in the dataset. A high volume of data is needed to ensure reliable estimations and to capture seasonal effects accurately, and in this study, there are ten years of data. The method used in this study is Seasonal and STL using ARIMA, which requires at least 25 monthly observations. It accounts for autocorrelations within the dataset and is mathematically represented as (Sekadakis et al., 2021):

$$\boldsymbol{Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_q \epsilon_{t-q}} \quad (1)$$

Here,
Y_t is the predicted value at time t,
α is a constant,
β terms are the coefficients of lags of the series up to p lags, and
ϕ terms are the coefficients of lagged forecast errors up to q lags.

For the STL section, the formula is as follows:

$$\boldsymbol{Y_t = S_t + A_t} \quad (2)$$

Here,
S_t represents the seasonal component and
A_t is the seasonally adjusted component.
Model accuracy is assessed using Mean Absolute Percentage Error (MAPE), calculated as:

$$\boldsymbol{M = \frac{1}{n}\sum_{t=1}^{n} \frac{[(actual\ value_t - \textbf{forecasted value}_t)}{actual\ value_t}] \times 100} \quad (3)$$

Here,
n is the number of observations

The model is selected based on the lowest MAPE, which indicates the best predictive accuracy.

In spatial analysis, KDE also estimates geographic feature densities over a location. It places a smooth surface over each point that decreases in value with distance, using a selected bandwidth based on the KDE function, then summed to create a density map (Mohaymany et al., 2013):

$$f(x,y) = \frac{1}{n\tau^2}\sum_{i=1}^{n} K\left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{\tau}\right) \quad (4)$$

Here,
$f(x, y)$ is the total estimated density at point (x, y)
N is the total number of data points
(x_i, y_i) is the coordination of the i[th] data point
K is the Kernel density function
τ is the bandwidth.
KDE provides an insightful visualization of the changes in fatal crash occurrences involving pedestrians/bicycles over 10 years.

DiD Approach

DiD is a statistical technique used to estimate the effect of a specific intervention or treatment by comparing the changes in outcomes over time between a population subjected to the intervention (treatment group) and a nonequivalent population that is not (control profile). DiD is used to infer relationships in observational data where random assignment is not feasible. This method helps control confounding variables that could influence observed changes, assuming both groups followed similar trends without intervention. DiD is frequently used to assess policies or specific changes in the population. In the context of a DiD analysis, the regression model can be expressed with the following equation (Grafova et al., 2014):

$$Y_{it} = \beta_0 + \beta_1 Treatment_i + \beta_2 Post_t + (\beta_3 Treatment_i \times Post_t) + \epsilon_{it} \quad (5)$$

Here,
Y_{it} is the outcome variable (crashes) for unit i at time t
$Treatment_i$ is a dummy variable indicating whether the unit ii is in the treatment group (1 if treatment, 0 otherwise)
$Post_t$ is a time dummy variable indicating the post-treatment period (1 if the intervention is applied, 0 otherwise)
β_3 is the DiD estimator, measuring the effect of the treatment on the treated group over time relative to the control group

This method is used in this study to consider two factors. First, the population is defined by states with a relatively lower increase in street network density from 2010 to 2020, and the intervention states that experienced a greater rise. Second, the analysis includes a comparison between burdened and non-burdened communities.

Modeling Phase

The main objective of this section is to evaluate the correlations of spatial indices with the number of fatal pedestrian/bicycle crashes in the test case of Florida. Over five years, 4,600 pedestrian/bicycle fatal crashes have been documented in the state. The study focuses on land use analysis at the census tract level, recording fatal pedestrian/bicycle crashes for each tract. Notably, about 53% of census tracts report zero values, indicating that no crashes were documented in more than half of the areas studied. As shown in Figure 4, this skewed distribution necessitates cautious model selection, especially when addressing the large number of zeros in the data.

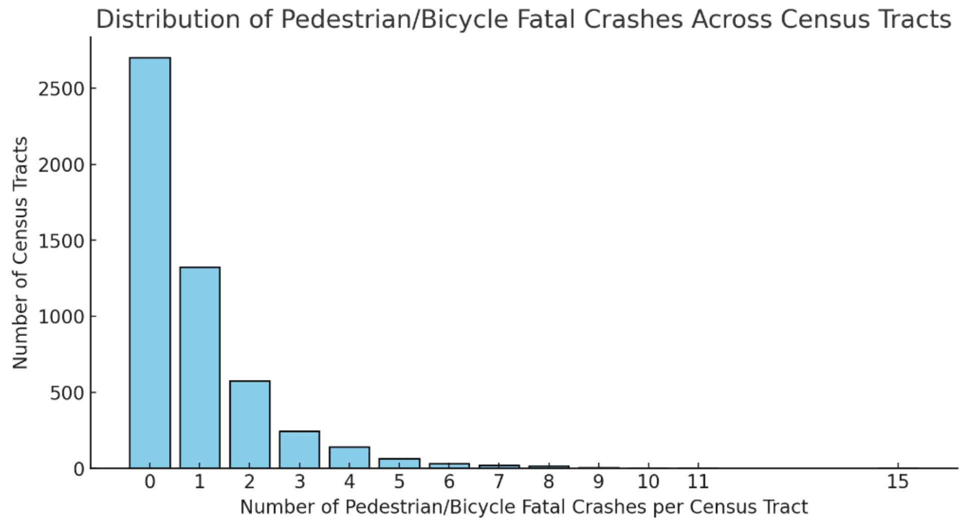


Figure 4. Frequency of pedestrian/bicycle fatal crashes by census tract over 5 years

As a part of the modeling phase, spatial indices are first assessed for significance along with other variables related to land use, traffic, and socio-demographics using a global model. The spatial heterogeneity of these parameters is then examined overall in Florida census tracts using a GWZINBR model, offering a comprehensive geographical insight. The explanations are given below.

Global Model: ZINB

The ZINB model is selected to address the skewed distribution with excess zeros observed in the data. This model combines a distinct model for the probability of excess zeros with a negative binomial distribution for count data. It allows it to evaluate count data with excess zero counts and overdispersion. As such, it can deal with the inflated zero count and overdispersion. Significant variables from this model are then used in the local model, described in the next section. The ZINB model is represented as follows (Sirinapa Aryuyuen, 2014):

$$\boldsymbol{P(X = x|\phi) = \phi\omega_0(x) + (1 - \phi)\mathrm{f}(x; \theta)} \quad (6)$$

Here,
X represents the count variable,
Φ is the probability mass function of X with the parameter θ
$\omega_0(x)$=1 if x=0; otherwise $\omega_0(x) = 0$

Local Model: GWZINBR

The study examines how the associations of independent variables vary between census tracts, providing insights into potential regional heterogeneity. Since the dependent variable is assumed to have a ZINB distribution, GWZINBR—in which the ZINB model is applied in each tract—is estimated. The dataset with N observations is subjected to N different regression models, each associated with a census tract. The model for a census tract is based on this stated number of observations. In this model, the bandwidth is 2,000, determining the number of surrounding observations examined for a specific observation. This is the best value after experimenting with various bandwidths, determined by the goodness of fit. A local model is unique to each census tract (Mohammadnazar et al., 2021).

For the surrounding observations in a sample, this model requires a geographical weight corresponding to the observation's weight placed in coordination (u_j, v_j). The Bi-square Kernel weighing function serves as the foundation for this weighing strategy (A P Handayani et a, 2019):

$$\boldsymbol{w_j^s(u_i, v_i) = \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2} \tag{7}$$

Here,

(u_i, v_i) is the coordination of the centroid i

$w_j^s(u_i, v_i)$ is a geographical weight assigned to the observation j with centroid i

d_{ij} is the spatial distance between observation j and centroid i

b is the maximum distance between the centroid j and the most distant observation within the group.

A non-stationarity test is used once the coefficient values for census tracts have been obtained to find the spatial heterogeneity of variables. Any variable that passes this test has spatial heterogeneity in the areas where it is tested. This is how the test ought to look (Liu et al., 2020):

$$\mathbf{Delta = Q_{3,\beta} - Q_{1,\beta}} \begin{cases} \mathbf{> 1.96 \times SE(\text{ global model}), \text{and } Max|Z| > 1.96} \\ \mathbf{\text{otherwise, fail the non} - \text{stationarity test}} \end{cases} \tag{8}$$

Here,

$Delta$ the difference between 25th and 75th percentile of coefficients of β

SE is the standard error of the coefficients from the global model

$|Z|$ z value of the coefficients in the local model

Metrics like the log-likelihood of the estimated models, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC)$M^{R^2}_{McFadden'sPseudo}$ are used to compare local and global models. Better model performance is indicated by a larger McFadden's Pseudo R-squared,$M^{R^2}_{McFadden'sPseudo}$ value, whereas a lower AIC or BIC value signals a more effective model. The following formulas are utilized to calculate these measurements:

$$\boldsymbol{AIC = 2K - 2\ln(L)} \tag{9}$$

Here,

k is the number of parameters in the statistical model.

$\ln(L)$ is the natural logarithm of the maximum likelihood estimate of the model.

$$\boldsymbol{BIC = ln(n)k - 2ln(L)} \tag{10}$$
Here:
n is the number of observations
k is the number of estimations in the model
$ln(L)$ is the natural logarithm of the likelihood of the model
$$\boldsymbol{M^{R^2}_{McFadden's Pseudo} = 1 - \frac{\ln L_M}{\ln L_0}} \tag{11}$$
Here,
L_M is the natural logarithm of the likelihood of the model
L_0 is the natural logarithm of the likelihood of the model with no predictors except for the intercept.

Results

The findings are arranged into three sections, each corresponding to the methodology used. Data on all census tracts in the U.S. are first summarized using descriptive analysis, followed by a case-study analysis of Florida. Time series analyses are then shown to monitor trends across time. The modeling phase results are presented in detail, highlighting spatial heterogeneity by comparing global and local models. This systematic methodology enables thorough data comprehension and provides insights into the temporal and geographical fluctuations in fatal pedestrian/bicycle crashes.

Descriptive Analysis

The following represents a more in-depth descriptive study of the entire U.S. and Florida. The detailed analysis for Florida is specifically utilized in the modeling phase section to investigate further spatial heterogeneity and the association of local factors with Pedestrian/bicycle crash fatalities.

Census Tract Level Analysis of the United States

Table 2 provides an overview of census tracts' status for different variables, i.e., spatial indices, sociodemographic features, and land use. Overall, the average total population per tract stands at 3,867.38, but this varies widely, reaching as high as 39,373 in some areas, indicating a significant variance in population density. Notably, the spatial indices reveal that among 84,443 census tracts, approximately 34% are identified as burdened in various aspects such as transportation, health, social vulnerability, environmental vulnerability, and disaster risks. Each of these binary indices is defined based on whether they meet or exceed the 65th percentile of their respective scores. Although the percentages of these indices are similar, they apply to different geographical census tracts.

In addition, the "Areas of Persistent Poverty Index" variable shows that 18,900 census tracts, representing 22.38% of the total, have regions where poverty is a long-standing issue. Sociodemographic features reveal details about the population characteristics within these tracts. The percentage of the population with income below 200% of the poverty level is an average of 30.05%, underscoring the economic challenges many residents face. Additionally, the table includes percentages for population characteristics like those uninsured and households without internet, with an average of 8.96% and about 15.3%, respectively, in these census tracts. These census tracts have an average of 16.59% elderly and 21.65% youth, which is crucial for planning community services and infrastructure. Furthermore, on average, in these census tracts, 12.03% of the population aged 25+ have less than a high school diploma, and 3.51% of those aged 16+ are unemployed.

Land use variables, such as the "Average Percent Land classified as Impervious Surface," standing as 30.87%, offer insights into urbanization levels and environmental concerns. The significant variation in this metric, with values reaching as high as 98.56%, indicates a wide range of urban development levels. It is notable that, on average, about 73.48% of the census tracts are identified as urban. Additionally, the data shows that the average cost of transportation across the census tracts is approximately $11,153, with costs reaching as high as $12597 and as low as $0. Another key metric is commuting time, averaging 26.43 minutes and going up to 98.66 minutes (about 3 hours). This comprehensive overview from Table 2 thus serves as a critical tool for understanding the multifaceted issues facing different census tracts and planning appropriate interventions.

Table 2. Descriptive statistics for the entire U.S. (N=84,443)

Count Variables	Frequency		Percent	
Areas of Persistent Poverty Index	18,900		22.38	
Transportation Insecurity Index	29,257		34.65	
Health Vulnerability Index	29,629		35.09	
Environmental Burden Index	29,555		35.00	
Social Vulnerability Index	28,669		33.95	
Disaster Risk Burden Index	29,584		35.03	
Binary Index for Burdened Communities	29,257		34.65	
Urban Tract Index	62,049		73.48	
Count Variables	**Mean**	**Max**	**Min**	**SD**
Total Population	3,867.38	39,373	0	16,71.47
Estimated Cost of Transportation	11,152.87	12597	0	1739.34
Average commute time to work	26.43	98.66	0	7.97
Percent of the population with Income below 200% of poverty level	30.05	100	0	17.73
Percent of people age 25+ with less than a high school diploma	12.03	100	0	10.44
Percent of people age 16+ unemployed	3.51	100	0	2.89
Percent of the population uninsured	8.96	100	0	7.05
Percent of households with no Internet subscription	15.3	100	0	10.91
Percent of population 65 years or older	16.59	100	0	9.12
Percent of the population 17 years or younger	21.65	100	0	7.43
Average Percent Land classified as Impervious Surface	30.87	98.56	0	25.49
Estimate Percent Population Below Poverty Level	13.67	100	0	11.54

Census Tract Level Analysis of Florida

In this section, the descriptive statistical analysis of the most important variables used in the modeling phase, highlighting the spatial heterogeneity across Florida, is described in Table 2 (excluding census tracts in American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, Puerto Rico, and the United States Virgin Islands.). An overview of the sociodemographic characteristics of the census shows that the average population per census tract, which ranges from zero to 21,117, is 4,142. The average proportion of the population that is 65 years of age or older is 22.3%, and the average percentage of people who are 17 years of age or younger is 18.6%. The average unemployment rate for individuals sixteen years of age and above is 3.2%.

Analysis of NaNDA also provides valuable insights into traffic dynamics at the census tract level in this state. Table 3 shows that the average street length per tract is approximately 43.68 mi, with an average of 62% of these roads experiencing heavy traffic. The average of VMT stands at 4,688,996. Furthermore, two variables—the percentage of crashes involving alcohol drinking and the rate of crashes that occurred in work zones—are extracted from the FARS datasets. These variables focus on census tracts where at least one pedestrian/bicycle fatal crash has happened over five years. The results indicate that roughly 58% of these crashes involve alcohol, and 77% of them occur in work zone areas.

This table also shows that 19.93% of Florida's census tracts have severe environmental burdens, 35.42% have health vulnerabilities, and 34.34% are burdened because of insecure transportation systems. 42.85% of tracts have high levels of social vulnerability, and 39.57% are burdened due to threats of natural disasters. Overall, 40% of the census tracts in Florida are identified as burdened based on an aggregated score that includes all criteria, shown in Figure 5. Additionally, the majority of the tracts—91.37 percent—are classified as urban tracts. These statistics provide a detailed overview of the census tract's sociodemographic and environmental challenges.

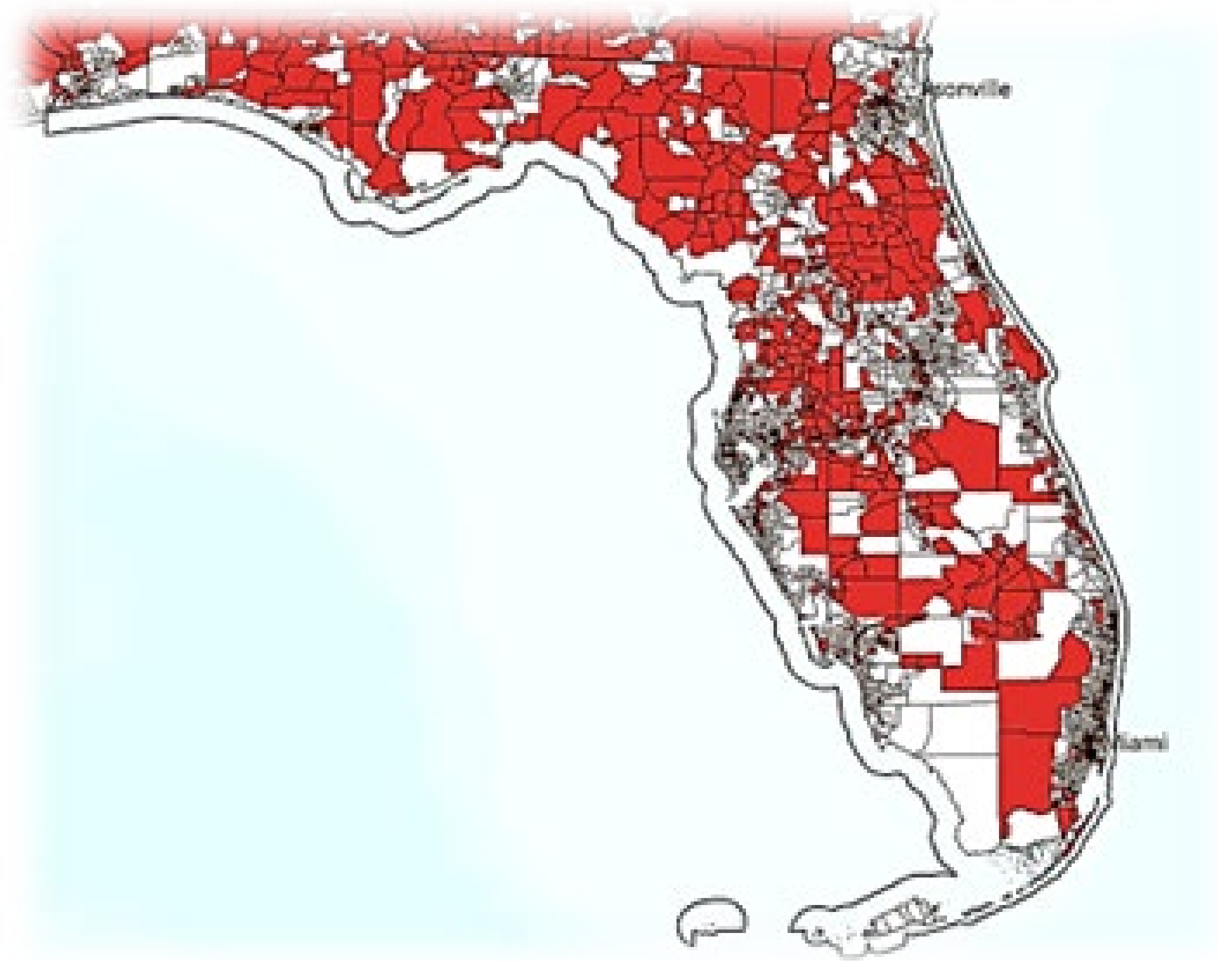


Figure 5. Spatial distribution of communities in Florida is highlighted in red

Table 3. Descriptive statistics for Florida (N=5122)

Categorical Variable	Frequency		Percent	
Transportation Insecurity Index	1,759		34.34	
Health Vulnerability Index	1,814		35.42	
Environmental Burden Index	1,021		19.93	
Social Vulnerability Index	2,195		42.85	
Disaster Risk Burden Index	2,027		39.57	
Urban Tract Index	4,680		91.37	
Count Variable	**Mean**	**Max**	**Min**	**SD**
Percent of Population 65 years or older	22.3	96.28	0	13.91
Percent of Population 17 years or younger	18.6	48.14	0	8
Percent of People age 16+ unemployed	3.2	34.4	0	2.8
Length of Street within a Tract (mile)	43.68	1236.44	0.45	79.71
Percent of Tracts Within 1 Mile of High-Volume Road	62.3	100	25	38.56
VMT	4,688,996	26,141,165	2903	2,773,439
Total Population per Census Tract	4142	21117	0	1939.0
Count Variable Focusing on Non-zero Regions	**Mean**	**Max**	**Min**	**St.**
Percent of Crashes Involving Drinking of Alcohol	58.47	100	0	31.58
Percent of Crashes Happened in Work Zone	77.9	100	0	53.07

Time Series Analysis

The results of the time series analysis are provided in this section for the entire U.S. Figure 6 illustrates the frequency of fatal pedestrian/bicycle crashes over the past years, highlighting a noticeable increase. Figure 7 also shows the locations of 8427 crashes in the year 2021. In addition, Figures 8 and 9 result from a time series forecast graph of fatal crashes involving pedestrians and bicycles in the U.S. predicted using the historical information from 2012 to 2021. The actual data on the frequency of fatal crashes involving pedestrians/bicycles in 240 months is shown with the gray line representing the fluctuations in crash numbers over this period. The blue dashed line demonstrates the trend of fatal crashes in 10 years. The overall trend shows an increase in fatal crashes over these years. However, the figures also reveal certain points where a decrease is observed, possibly due to unpredictable events or other irregular influences on this type of crash.

The shaded area beyond the last actual data point in 2021 indicates the forecasted values in two years, with the boundaries of the shaded area representing confidence intervals. This forecast extends into the future, predicting the number of fatal crashes involving pedestrians/bicycles while accounting for the observed trend. The dashed trend line suggests an overall increase in this type of fatal crash two years after the last year of data, 2021. MAPE represents the average magnitude of errors between the forecasted and actual values, which is 4.7%. Thus, on average, the forecasts deviate from the actual values by 4.7%, suggesting a relatively high level of accuracy in the forecasting model.

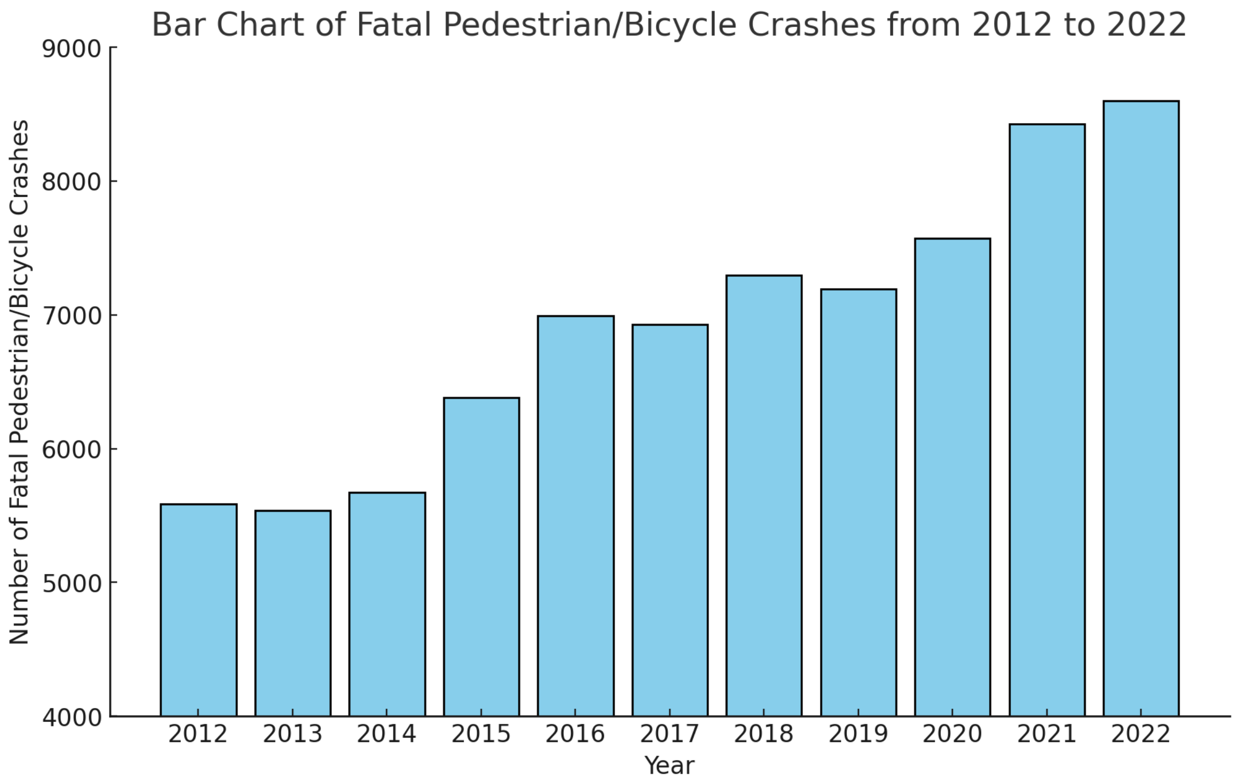


Figure 6. Frequency of pedestrian/bicycle fatal crashes over ten years

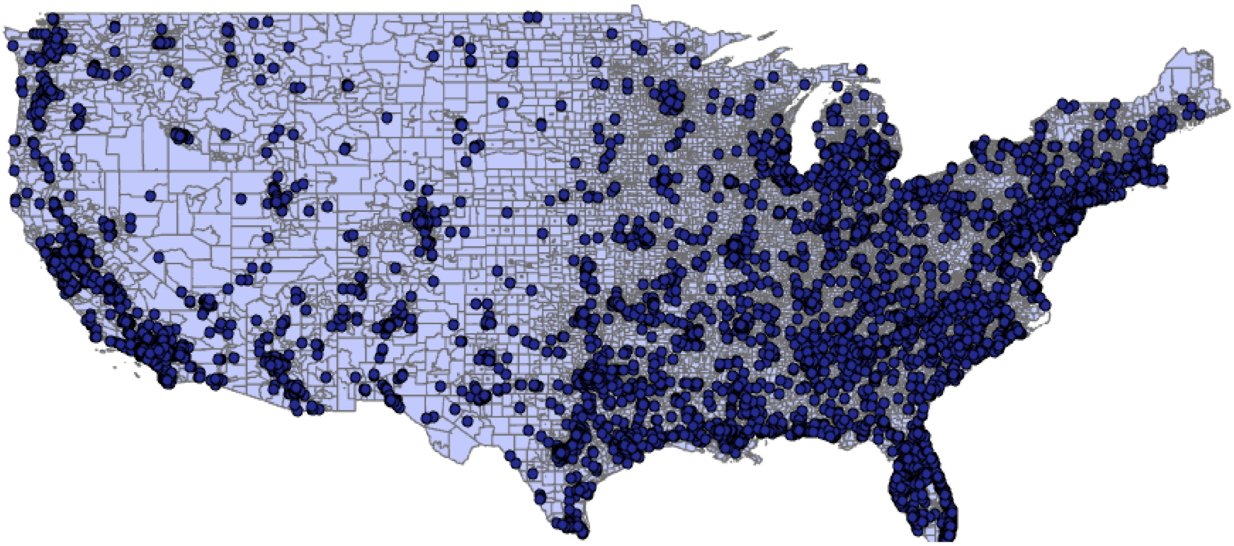


Figure 7. Pedestrian/bicycle fatal crash locations over the U.S. in 2021

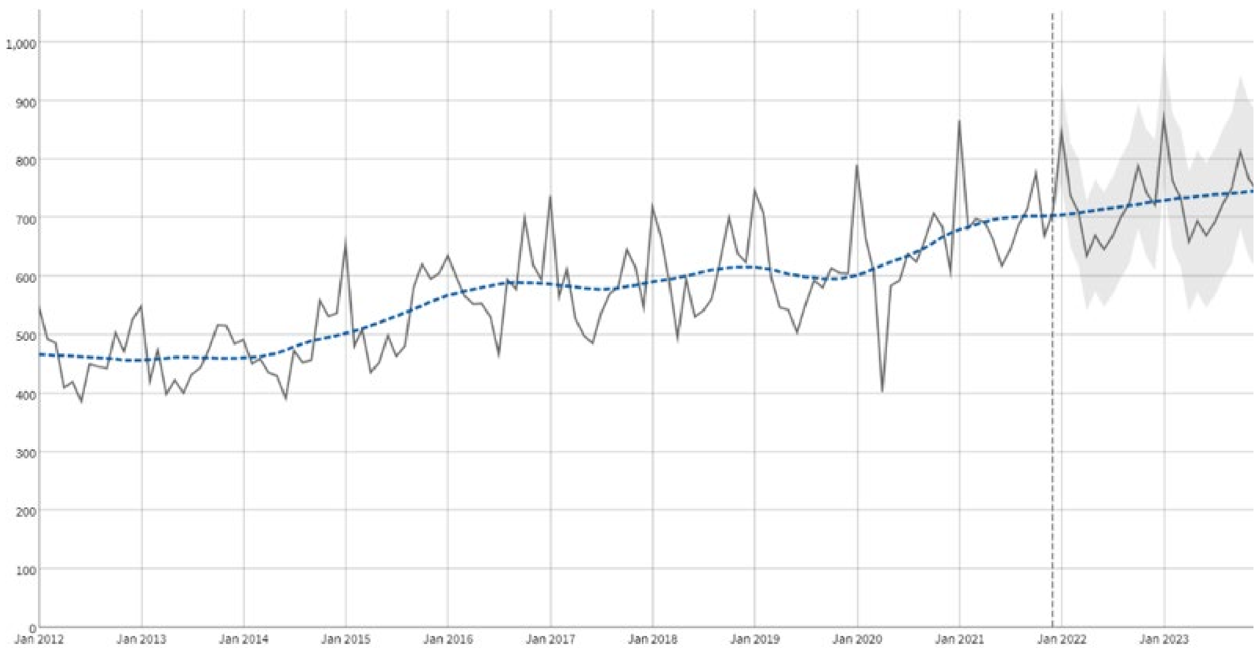


Figure 8. Time series analysis, STL using Arima

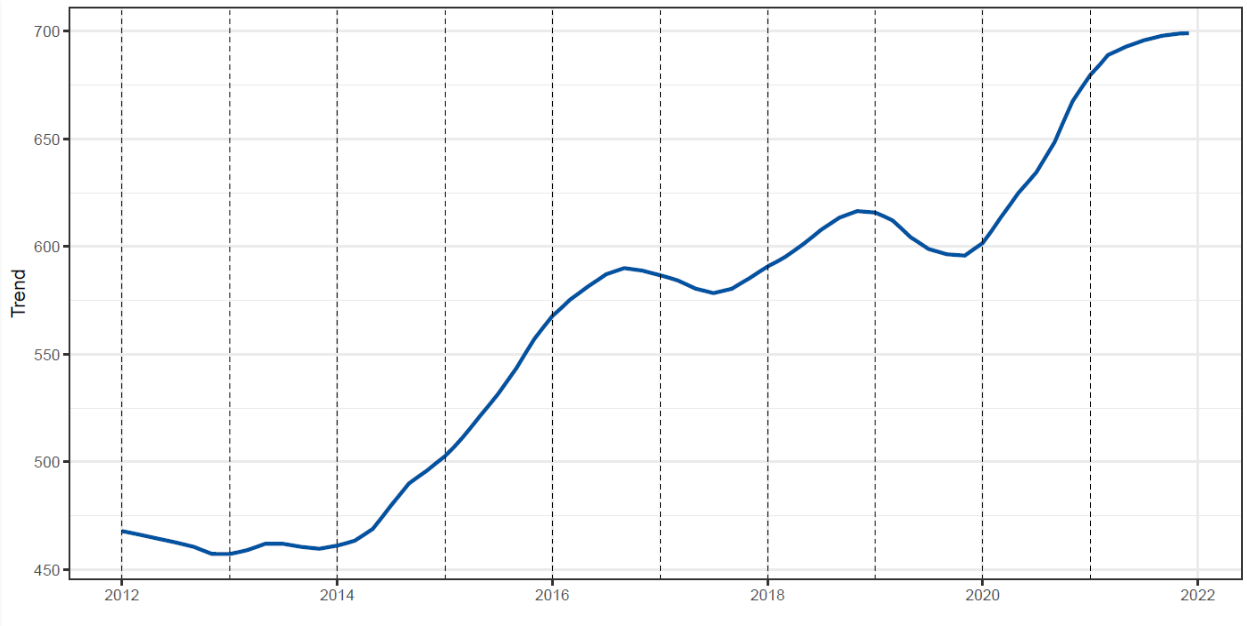


Figure 9. Trends in fatal pedestrian/bicycle crash data

The result of KDE is shown in Figure 10. Lighter colors show fewer fatal crashes, whereas darker colors represent the higher frequency of fatal crashes in a targeted location. This illustrates an increasing trend of this type of crash. This rising trend is consistent with the results of the time series analysis, highlighting a gradual nationwide increase in fatal pedestrian/bicycle crashes, with specific locations emerging. In the 2021 map, three significant hazardous locations are identified; one is in Region Four of the DOT, specifically in Florida. The monthly trends of pedestrian/bicycle fatal crashes are highlighted in Figure 11, underscoring their increasing trajectory in 2021 compared to 2012.

Moreover, the data reveals the trend within each year. Additionally, Figure 12 shows the distribution of fatal crashes involving pedestrians/bicycles across the different regions of the USDOT in 10 years, which confirms that the highest frequency of such fatal crashes is in Region 4 compared to the others. It highlights the justification for its selection as the focus for spatial analysis in the next step of methodology. For more clarification, the regions are shown in Figure 13.

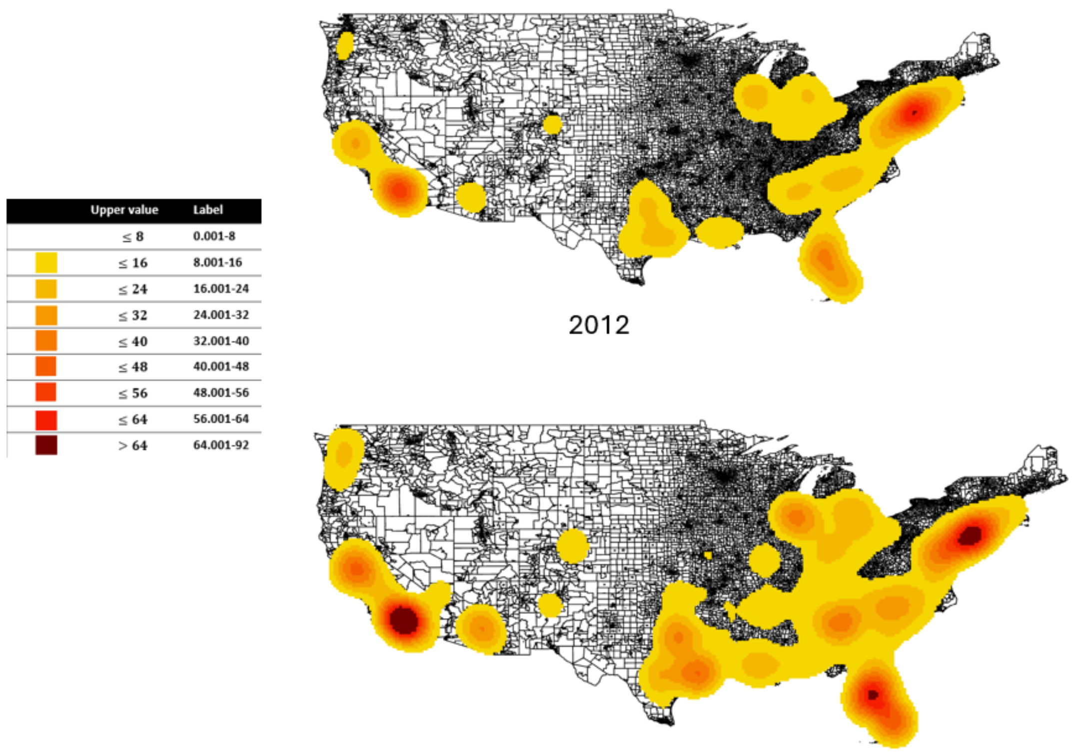


Figure 10. Kernel density of pedestrian/bicycle fatal crashes over a decade in the U.S.

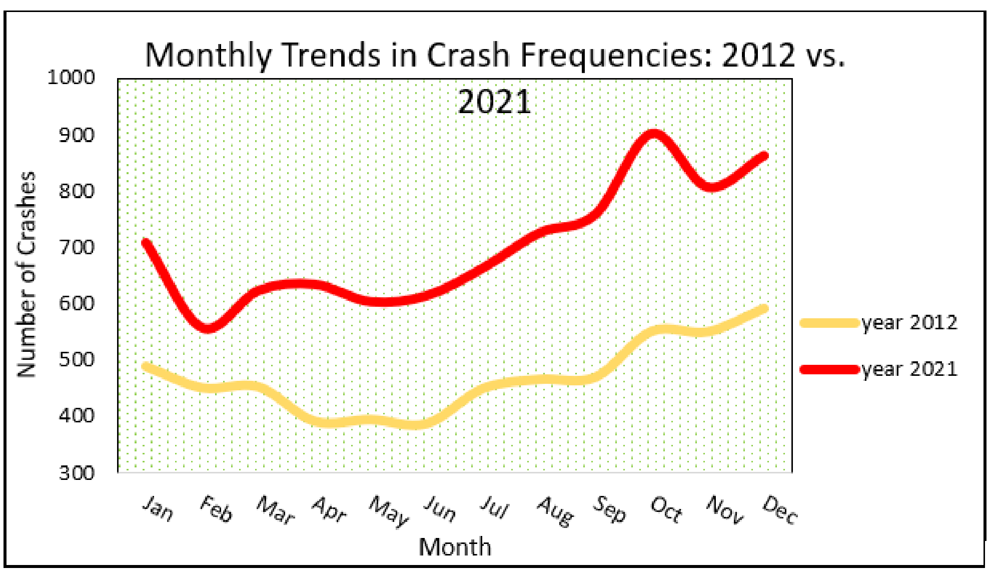


Figure 11. Monthly trends in pedestrian/bicycle fatal crashes: 2012 vs. 2021

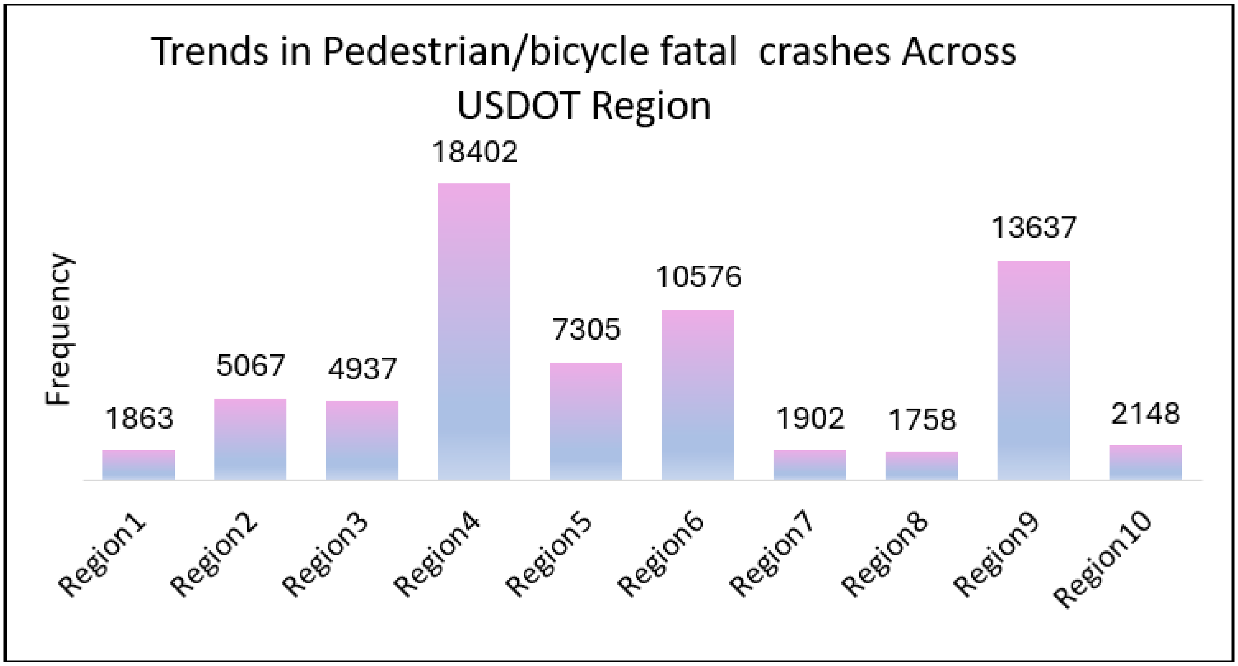


Figure 12. Pedestrian/bicycle fatal crash trends across USDOT regions, 2012-2021

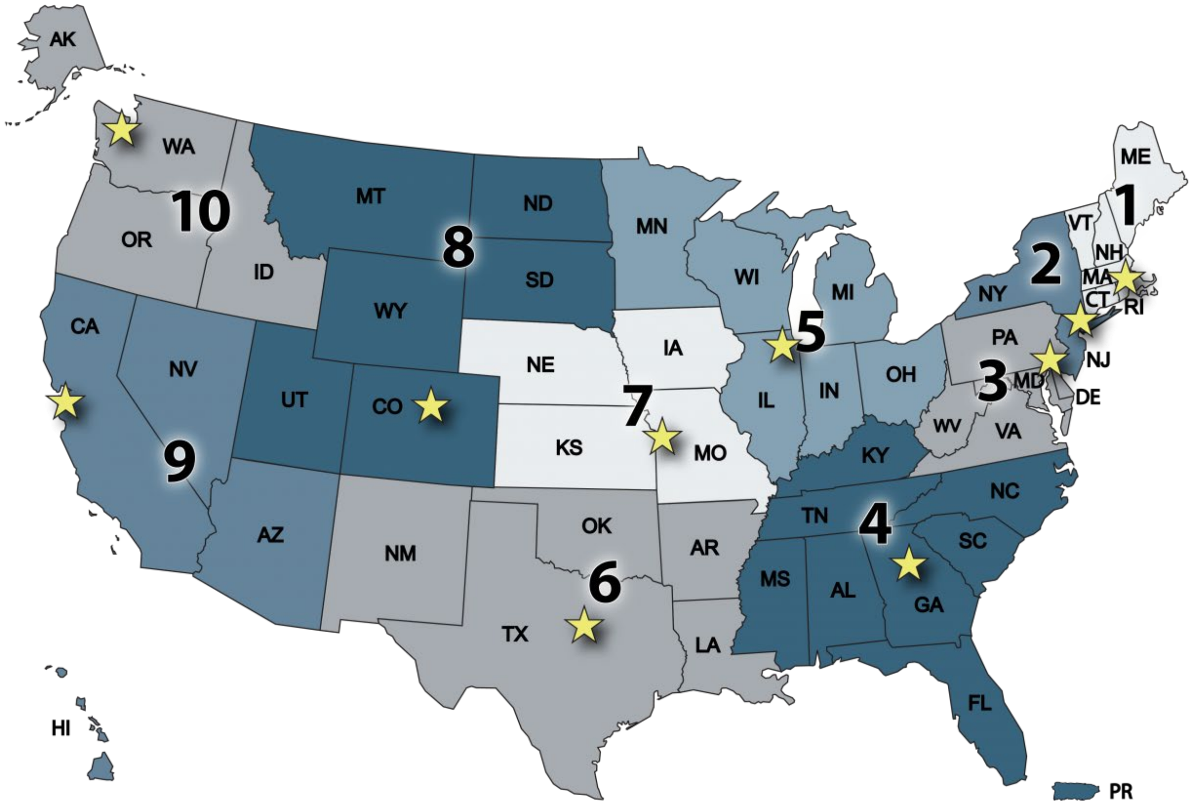


Figure 13. USDOT regions[1]

DiD Results

This section analyzes two key features: spatial indices and urban form. The analysis explores how changes in street network density between 2010 and 2020 correlate with fatal pedestrian and bicycle crashes in 2012 and 2021. The subsequent part details the findings from the DiD analysis focusing on spatial indices.

Pedestrian/bicycle fatal crashes change based on urban form over time:

This section divides states into two groups as follows:

- The "Treated group of states." These states saw a higher increase in street network density, above the median value observed across all states, including Alabama, Arizona, Arkansas, Colorado, Delaware, District of Columbia, Florida, Georgia, Idaho, Maine, Mississippi,

[1] https://www.transit.dot.gov/

Missouri, Montana, Nevada, New Mexico, North Carolina, North Dakota, Oregon, South Carolina, South Dakota, Texas, Utah, Virginia, Washington, Wyoming.

- The "non-treated or control group of states." These states experienced a relatively lower increase in street network density from 2010 to 2020, specifically below the median increase observed across all states, including Alaska, California, Connecticut, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Nebraska, New Hampshire, New Jersey, New York, Ohio, Oklahoma, Pennsylvania, Rhode Island, Tennessee, Vermont, West Virginia, Wisconsin.

Figure 14 illustrates the result of DiD according to the control and treated groups. The control group shown by the blue line includes states with a relatively lower increase in street network density from 2010 to 2020, below the median increase observed across all states. The data shows that the average number of pedestrian/bicycle fatal crashes in these states increased moderately over the decade. This suggests that changes in the number of these crashes in these states might be influenced more by factors like traffic volume, law enforcement, vehicle safety, or socioeconomic conditions rather than significant changes in road infrastructure.

Conversely, states with a greater rise in street network density above the median value are included in the treated group shown in the red line. According to the data, the average number of pedestrian/bicycle fatal crashes increased more sharply in these states throughout the same period. This implies that significant modifications to roadway networks are correlated with a higher frequency of fatal crashes. It represents that while improved road networks can facilitate better vehicle access and flow, they may also spread the risk of crashes for pedestrians and bicyclists due to factors like denser traffic, faster speeds, or more complex road systems.

Comparing these two groups over time helps isolate the effect of increased street network density on fatal pedestrian and bicyclist crashes. The steeper rise in crashes in the treated group suggests that more significant infrastructure changes might have unintended negative consequences on safety—although it is recognized that pedestrian exposure and other variables were not controlled for in this analysis. This insight is crucial for policymakers and urban planners, emphasizing the need for careful planning and integration of safety measures when undertaking major road infrastructure projects to mitigate potential adverse effects on road safety.

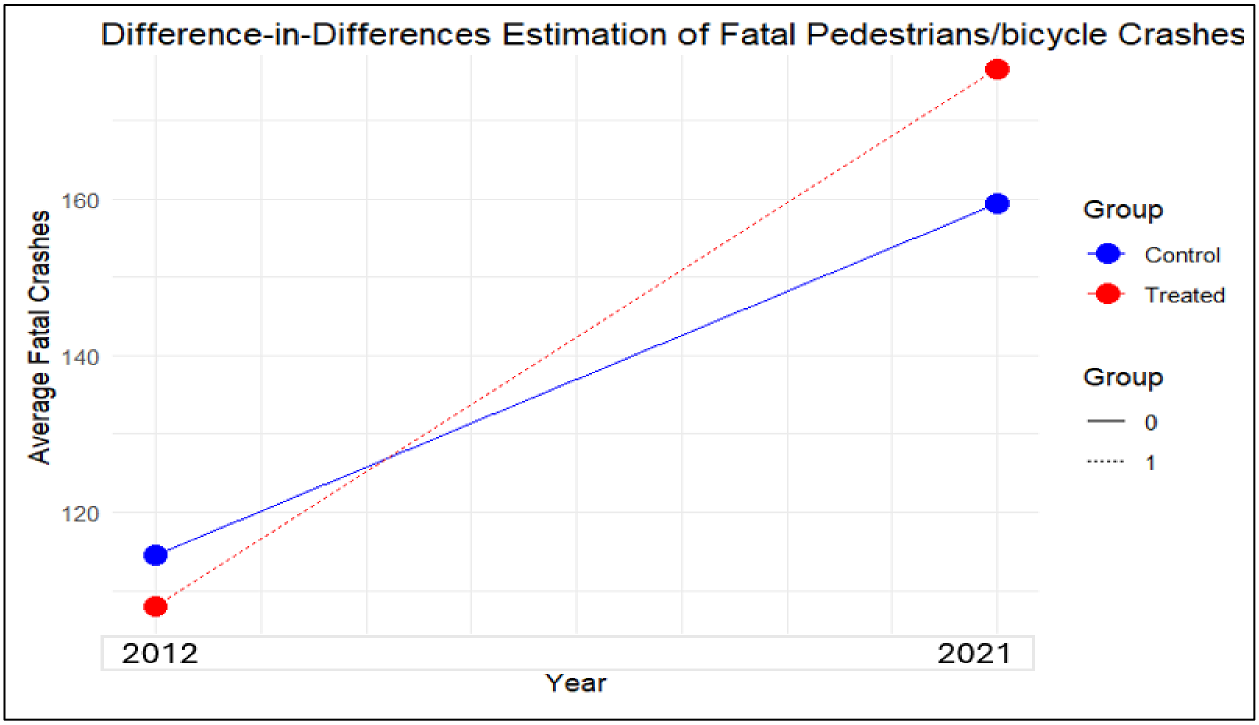


Figure 14. Street network density association with pedestrian/bicycle fatal crashes

Pedestrian/bicycle fatal crashes change over time

The study assumes that the status of burdens may have remained consistent over the ten years; if a census tract was classified as burdened in 2020, it was also considered burdened in 2012. The census tracts in the U.S. are divided into two groups as follows:

- The treated group is assumed to be the burdened communities
- The un-treated/control group is the non-burdened community

The results will remain the same if we reverse the burdened vs. non-burdened coding. Figure 15 shows that although both burdened and non-burdened groups saw an increase in fatal crashes from 2012 to 2021, the increase was larger for the burdened group than the non-burdened groups. This highlights the need for targeted interventions and policies to address the challenges that burdened communities face to improve safety and reduce pedestrian and bicyclist crashes.

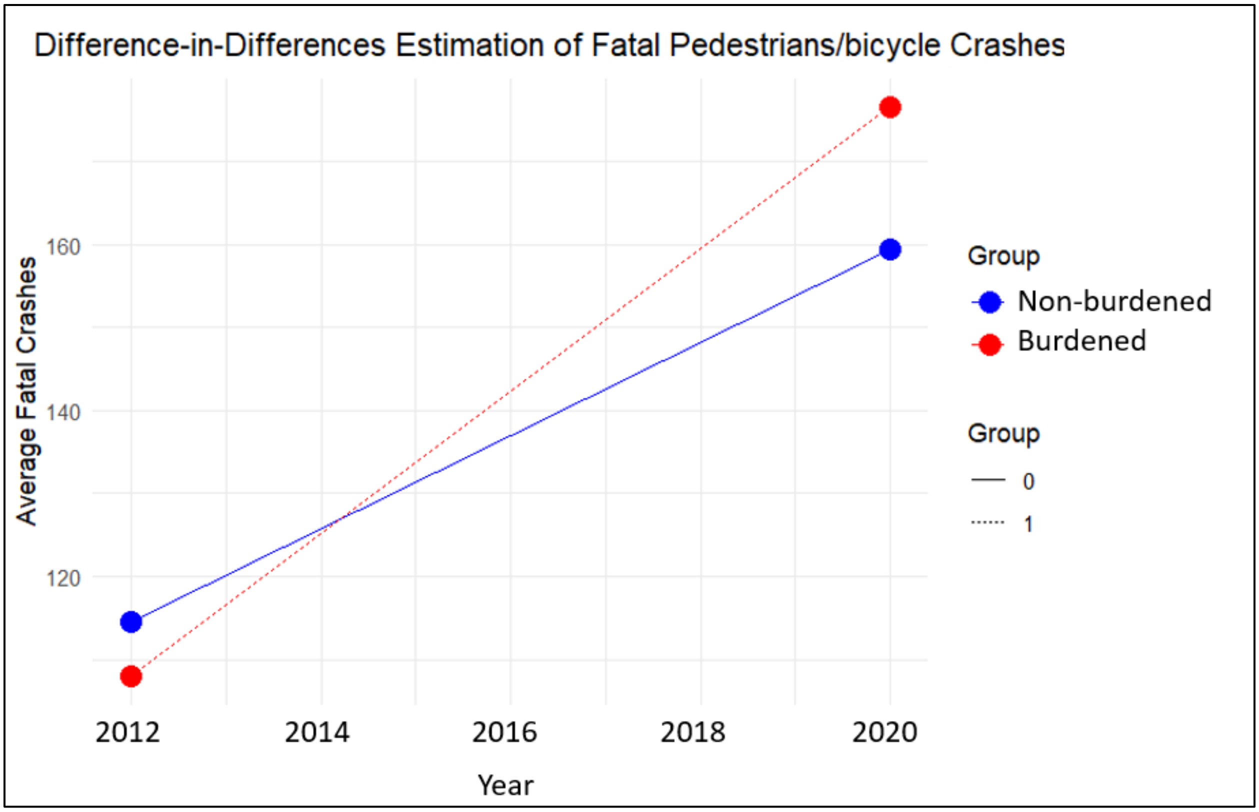


Figure 15. Burdened and non-burdened increases in pedestrian/bicycle fatal crashes

Modeling Results

This section investigates the spatial heterogeneity in safety outcomes, focusing on a state with high fatal pedestrian/bicycle crash rates. Finding the variation in safety metrics across this state facilitates focused inquiries and actions, improving safety for all road users. Initially, a global model will be used to find the significant factors in fatal pedestrian/bicycle crashes. Subsequently, a local model investigates spatial heterogeneity across these significant variables. Addressing local needs with safer countermeasures is desirable. In this regard, the result of each model is explained fully below.

Global Model

The result of the ZINB model is shown in Table 4. This result is divided into two sections: count and zero. The ‘count’ section refers to the number of fatal pedestrian/bicycle crashes at the census tract level, capturing the frequency of these events. At the same time, the ‘zero’ parts address the probability of census tracts having no such crashes. As shown in the count section, all valid indices used in this model are globally significant and positively correlated with the frequency of fatal pedestrian/bicycle crashes. These observations mean that a higher score indicates more significant challenges for that indicator.

The analysis incorporates sub-variables of the Social Vulnerability Index that show low correlation, such as the percentage of younger people and the unemployment rate in each census tract. Each feature is directly or indirectly associated with increased fatal crashes of pedestrians/bicycles. For example, a higher percentage of younger people correlates with a decrease in the potential of regions with zero fatal crashes. In other words, while these variables may not directly lead to an increase in fatal crashes, their presence is associated with a lower probability of tracts having zero crashes.

As the model shows, some traffic and urban form-related factors, such as VMT, length of the street in each tract, and urban indicators, are also associated with increasing fatal crashes of pedestrians/bicycles. Moreover, the model demonstrates robust performance, as evidenced by its goodness of fit. Additionally, the Log(theta) parameter in the model is statistically significant, indicating a notable association with the dispersion of the data, suggesting that the variable effectively contributes to heterogeneity in the dependent variable within the model. Not all indices were included in the model due to correlation issues. Figure 16 highlights that the social vulnerability index exhibits a correlation exceeding 50% with the transportation cost index, which led to its exclusion from the modeling.

Table 4. Result of the ZINB model (Global Model)

Variable	Estimate	Std. Error	z value	Pr(>|z|)
Count Section				
Constant	-3.85	0.04	-6.39	0.00***
Transportation Cost Index	0.51	0.05	12.52	0.00***
Environmental Index	0.51	0.04	10.46	0.00***
Health Index	0.18	0.05	4.13	0.00***
Disaster Risk Burden Index	0.13	0.04	2.75	0.01**
Logarithm VMT	0.22	0.04	5.73	0.00***
Zero Section				
Constant	10.78	1.55	6.96	0.00***
Percent of People age 16+ unemployed	-0.22	0.08	-2.84	0.00**
Logarithm of Length of a Street within a Tract	-1.05	0.15	-7.00	0.00***
Percent of Population 17 years or younger	-0.04	0.02	-2.53	0.01**
Urban Tract Indicator	-0.78	0.36	-2.15	0.03*

Note: Significance codes: 0 '***', 0.001 '**', R2Pseudo = 0.04, AIC = 12942.26, LL = -6459, Log(theta) z value: 6.4, Pr(>|z|): 8.82e-11 ***

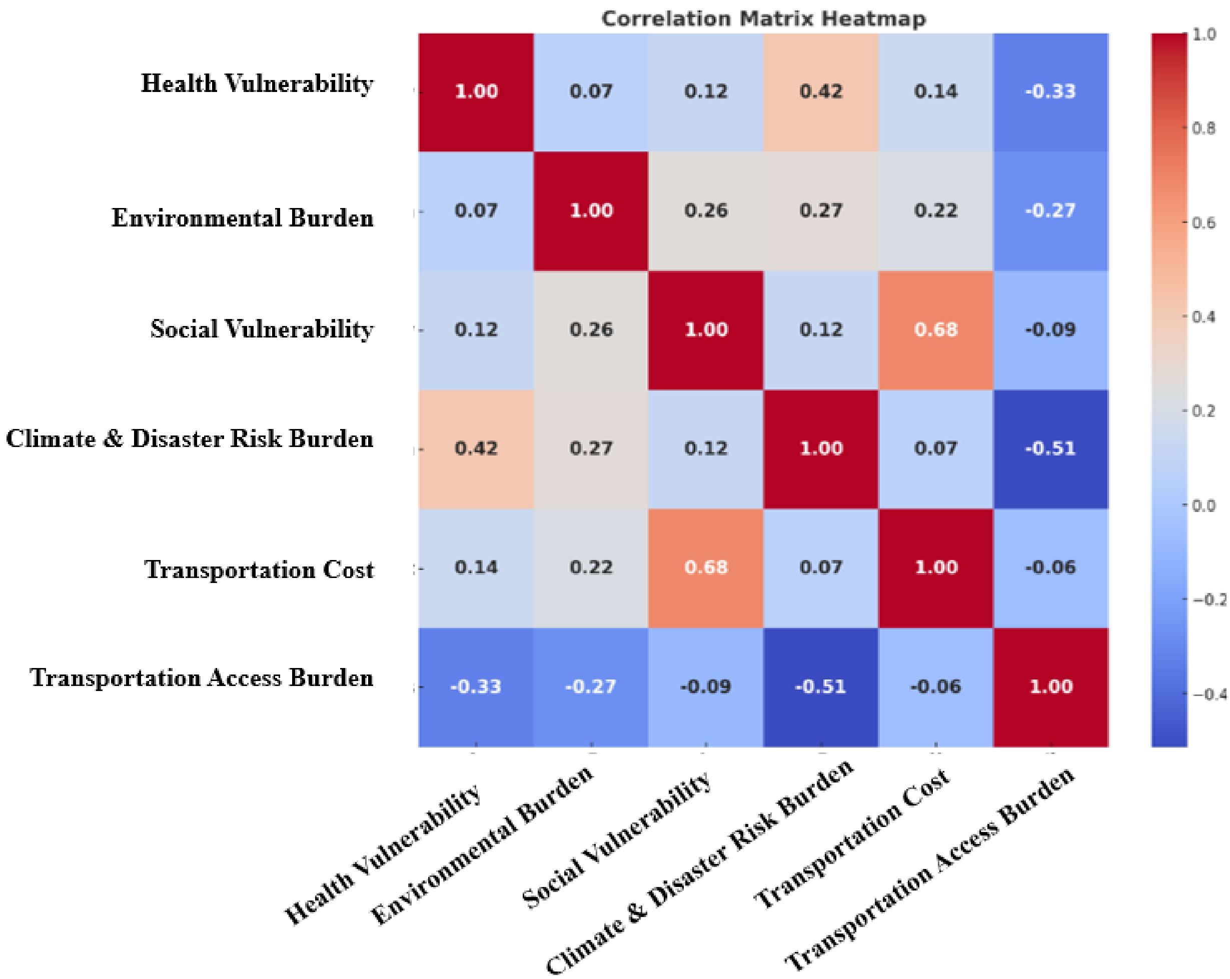


Figure 16. Correlations among spatial indices

Local Model

Table 5 presents the results of the GWINBR model, highlighting the significant variables from the global model that also passed the non-stationary test described in the methodology. The analysis reveals that most variables exhibit spatial heterogeneity across Florida, including two valid spatial indices and several other variables. However, the percentage of people age 16+ unemployed, and the logarithm of VMT do not show any spatial heterogeneity. Variables demonstrating spatial heterogeneity are depicted on a map using ArcGIS to represent these differences visually in Figures 17-19. In these maps, regions are color-graded from lighter to darker, showing a progression from lower to higher significance of each variable. The color gradient from yellow to red indicates variables positively associated with fatal pedestrian and bicycle crashes, while the green trend signifies a negative association.

Among the spatial indices, the Transportation Cost and Health Vulnerability Index shows spatial heterogeneity in Florida. For instance, Figure 17 shows the correlation of the Transportation Costs

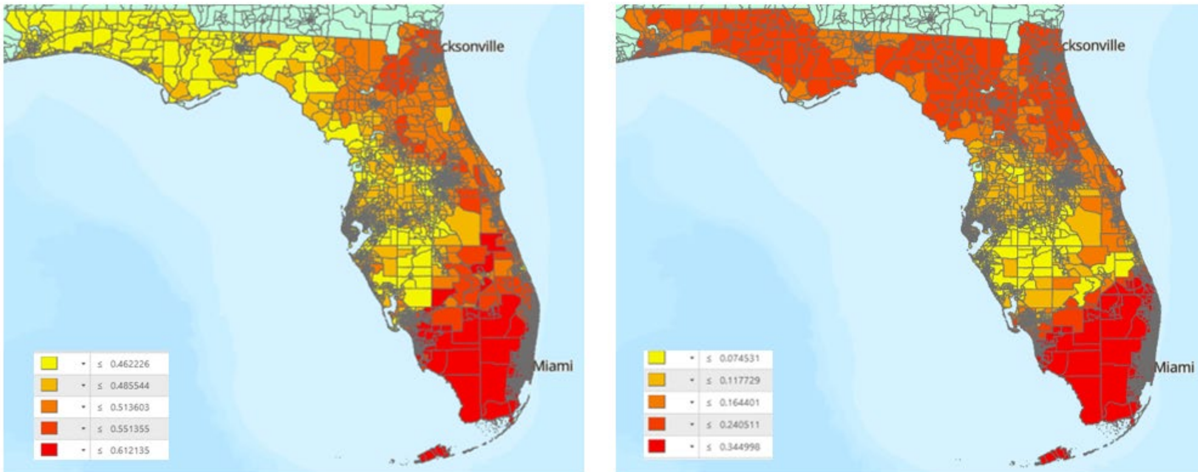
Index with crashes, which is lower in the Panhandle regions and higher in the southern regions. Meanwhile, the Health Vulnerability Index is less significant in central Florida, as shown in Figure 17. The Environmental Burden Index and Disaster Burden Index do not show spatial heterogeneity in this state. Figures 18 and 19 illustrate additional variables associated with fatal pedestrian/bicycle crashes.

Notably, younger populations have a greater significance as one moves toward the South, and there are areas in the North and Panhandle regions where these demographics positively contribute to reducing fatal crashes. The urban variable, typically linked to an increase in fatal crashes globally, exhibits both negative and positive associations in a spatial heterogeneity analysis. In certain census tracts of South Florida, it is associated with lower fatal pedestrian/bicycle crashes. Furthermore, the variable representing the logarithm of street length within a tract shows decreasing significance in Southern regions. Importantly, Figure 20 illustrates the spatial heterogeneity of significant spatial indices, focusing exclusively on burdened areas and revealing differences within burdened communities. This model shows a reasonable fit considering its complexity.

Table 5. Result of the GWZINBR model (Local Model)

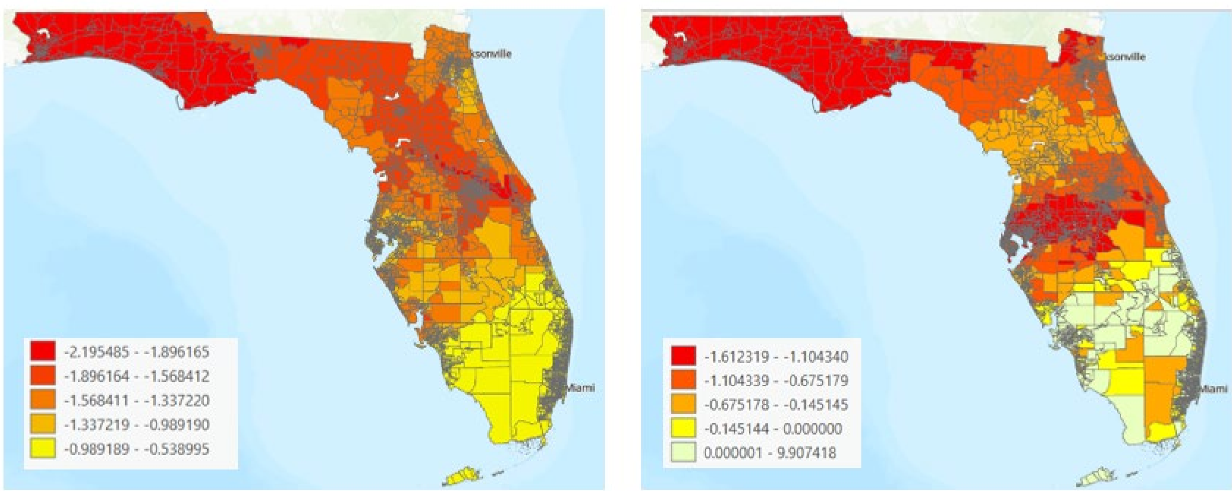
Variables	$Q3, \beta$	$Q1, \beta$	Delta	1.96*SE	Max |Z|	Heterogeneity
Transportation Cost Index	0.56	0.46	0.10	0.08	8.83	Yes
Environmental Index	0.57	0.48	0.09	0.10	9.67	No
Health Index	0.30	0.09	0.21	0.09	4.39	Yes
Disaster Risk Burden Index	0.18	0.13	0.06	0.09	3.55	No
Logarithm VMT	0.21	0.18	0.03	0.08	3.89	No
Percent of People age 16+ unemployed _Z	-0.19	-0.30	0.10	0.15	2.91	No
Logarithm of Length of a Street within a Tract _Z	-0.80	-1.59	0.79	0.29	7.30	Yes
Percent of Population 17 years or younger_Z	-0.01	-0.06	0.05	0.03	3.35	Yes
Urban Tract Indicator_Z	0.07	-1.06	1.13	0.71	3.26	Yes

Note: Pseudo-R^2 = 0.042, AIC =5077, LL = -6440.6, Z represents the variables used in the zero section of ZINB



a) **b)**

Figure 17. Transportation Cost (a) & Health Vulnerability (b) Indices showing spatial heterogeneity



a) Logarithm of Length of a Street within a Tract **b) Urban Indicator**

Figure 18. Urban form spatial heterogeneity

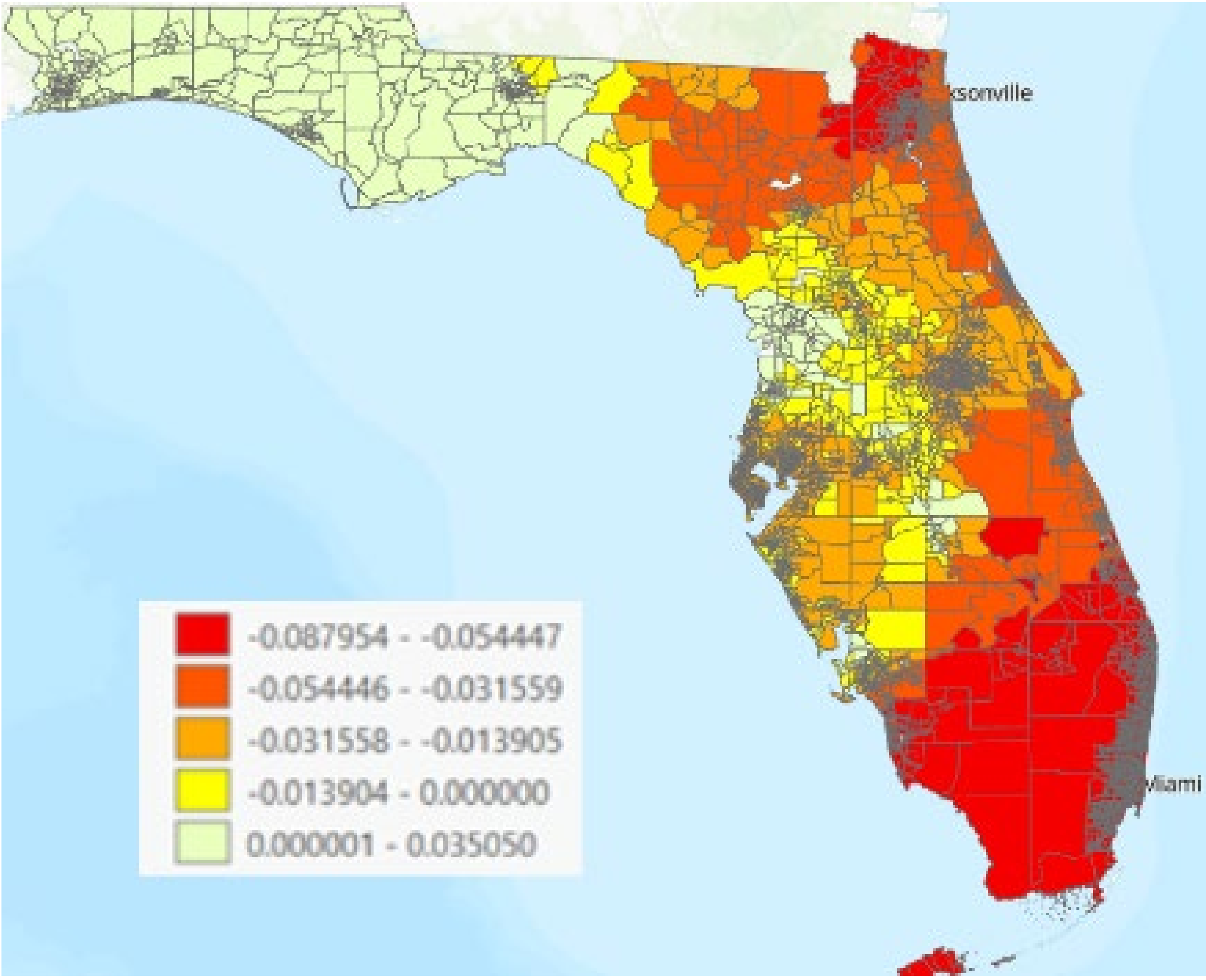


Figure 19. Percent of the Population 17 years or younger—Spatial Heterogeneity

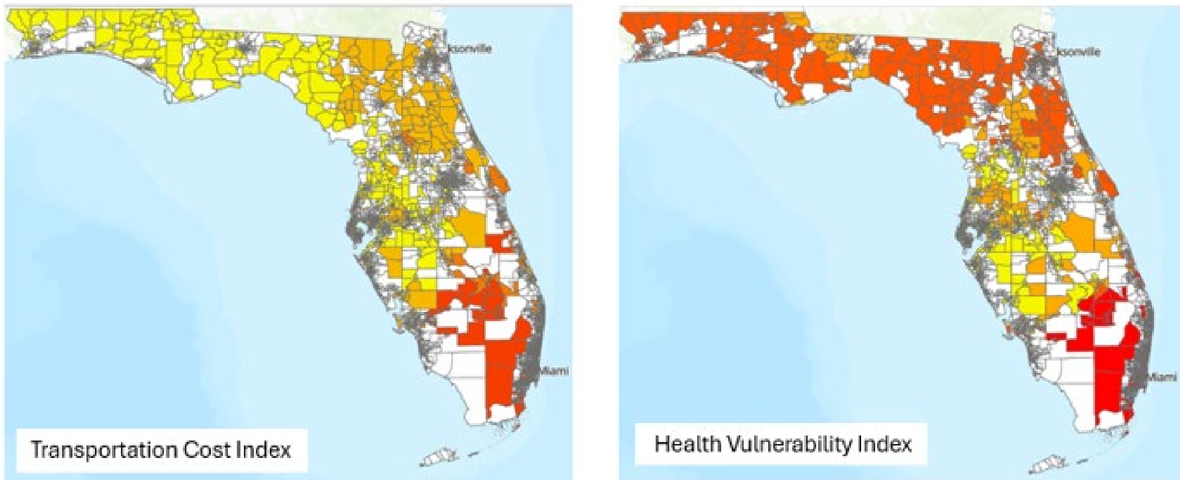


Figure 20. Burdened communities and spatial heterogeneity

Discussion

This study provides a comprehensive overview of the increasing trend in fatal pedestrian/bicycle crashes, highlighting the urgency for immediate preventative measures. The entire U.S. has experienced a 50% increase in crashes over the past decade. Florida is an example of a state with substantial pedestrian and bicyclist crashes. Analysis in such contexts can guide safety prevention efforts. Emphasizing groups who suffer from inadequate facilities is crucial in these efforts.

In addition, research shows a correlation between higher roadway network density and increased crash rates. The complexity and volume of the street network contribute to more frequent interactions between vehicles, pedestrians, and bicyclists, leading to a higher likelihood of crashes. This further emphasizes the crucial role of urban design in enhancing safety, highlighting the need for well-planned urban forms that prioritize the protection of pedestrians and cyclists.

In addition to identifying frequent crash locations, a more detailed analysis can be conducted to explore the spatial heterogeneity of key factors within these hazardous locations. This local observation could reveal geographic variations in the main factors influencing these crashes. The findings support previous research (Polus et al., 2005) spatial indices have a major global correlation with pedestrian/bicycle fatal crashes; nevertheless, local differences in these indices are important. To create focused, effective solutions, it is necessary to quantify these variances. By examining the spatial index across Florida, the overall trend and the census tract most affected are identified. These insights confirm that spatial heterogeneity aligns with the overall trends observed in the state, underscoring the need to target interventions effectively.

The spatial heterogeneity analysis of Transportation Cost also underscores its importance for pedestrian and bicyclist safety in Southern regions. This index includes sub-variables like income and gasoline cost, which is higher in southern Florida. The rising cost of gasoline could deter lower-income individuals from owning cars, leading them to rely more on walking or bicycling. In addition, higher poverty rates are found predominantly in the South. Notably, cities such as Opa-Locka, Belle Glade, and Brownsville, which have about 40% of their populations living below the poverty line, place them among the top five most impoverished areas (Stacker, 2022). Consequently, cost significantly influences lifestyle choices in southern regions, posing substantial challenges for low-income families. These economic pressures often force changes in mobility behaviors, potentially increasing their vulnerability to traffic accidents. The Health Index also demonstrates significant regional differences across Florida, particularly in the South. These differences may indirectly affect walking and driving behaviors, influencing road safety conditions. This underscores the critical association of health issues with daily activities and road safety in specific regions.

Additionally, the Environmental Burden Index, defined by factors such as air pollution, hazardous sites, infrastructure, and water pollution, shows global significance in the state. The main association of this index with fatal pedestrian/bicycle crashes could be attributed primarily to air pollution, which may cause visibility issues and hazardous sites. Furthermore, infrastructure development and management play critical roles; poorly managed or inadequate infrastructure can exacerbate these risks (Zegeer & Bushell, 2012). The importance of the disaster risk burden index in the global model could be related to extreme weather events, such as storms and floods, which increase the risks and dangers for pedestrians and bicyclists, as people may be at a higher risk, e.g., more likely to jaywalk on the road during rainy periods or drivers may engage in more aggressive behavior on warmer days. The environmental challenges may affect the safety and mobility of pedestrians and bicyclists in adverse weather conditions (Zhai et al., 2019).

The findings of this study reveal that age has a differential association with fatal pedestrian/bicycle crashes. Younger people, particularly teenagers, are more prone to such crashes, which are more pronounced in southern regions of Florida. Numerous studies attribute this behavior to teenagers' reduced attention and riskier crossing behaviors (Das et al., 2021; Tuckel, 2021). However, in the Panhandle region, a younger population seems to contribute to a decrease in fatal crashes. This suggests that age-related associations with pedestrian/bicycle safety can vary significantly by region, potentially due to differences in supervision, education programs, and community engagement (Onieva-Garcia et al., 2016).

The urban indicator variable is positively associated with fatal pedestrian/bicycle crashes globally. In spatial heterogeneity analysis, some regions in South Florida are negatively associated with the urban indicator, particularly in densely populated areas like Palm Beach and Monroe. This can be attributed to the high level of urbanization in these areas, which, in contrast to other regions, may reduce the number of fatal crashes. These results highlight the effectiveness of advanced urbanization in enhancing safety in some parts of southern Florida. In addition, the length of street overlay could decrease the possibility of seeing zero pedestrian/bicycle fatal crash regions. However, this effect is less pronounced in Southern regions of Florida.

This study's results align with Florida's existing traffic conditions, transportation infrastructure, and sociodemographic characteristics. A detailed understanding of the association of each variable can inform policy changes and guide the implementation of targeted initiatives across various census tracts. By adapting strategies to the specific needs of the most affected areas, these interventions can be more cost-effective, efficiently implemented, and yield quicker improvements. This adapted approach promises to enhance the state's overall status and contributes to broader efforts to improve pedestrian/bicycle safety across the US.

Conclusions and Recommendations

The increasing trend in fatal pedestrian and bicyclist crashes highlights the urgent need for enhanced safety measures. This study examines the geographical and temporal distribution of fatal pedestrian/bicycle crashes by creating a unique database and finding temporal patterns and spatial distributions of crashes in the U.S. A unique aspect of this study is examining exploring the possible relationship between spatial indices—such as Health Vulnerability, Environmental Burden, Social Vulnerability, Transportation Insecurity, and Disaster Risk Burden—and the association with fatal pedestrian/bicycle crashes. A case study focusing on Florida demonstrates spatial heterogeneity between spatial indicators and safety. The study uses a rigorous technique that includes GWZINBR and time series analysis on integrated FARS and spatial indicators with other datasets that offer important variables.

The study found a significant correlation between fatal pedestrian and bicycle crashes and 1) spatial indicators, 2) urban form, and 3) demographics at the census tract level. This indicates that the relationships between pedestrian and bicyclist safety and spatial indicators are nuanced and complex, with transportation cost, health vulnerability, and environmental burden playing out as having strong positive associations. Furthermore, denser networks, e.g., arterial roads, may play a role in higher fatal pedestrian and bicycle crashes, highlighting the need to investigate the issue further in the context of pedestrian and bicyclist safety. This study further highlights differences in transportation safety, notably a significant increase in pedestrian and bicycle fatalities over the past decade. It demonstrates that burdened communities have a higher incidence of fatal pedestrian and bicycle crashes, and this trend seems to be increasing.

The study reveals significant spatial heterogeneity associated with differences in various factors, including spatial indices, land use/urban form, and demographics, for the Florida test case. Substantial heterogeneity was found for transportation cost and health vulnerability index, pointing out that the relationships vary in space, even within burdened communities. Additionally, while 40% of census tracts in Florida are identified as burdened regions, they are dispersed throughout the state. This distribution highlights spatial differences even within burdened areas.

The study shows that safety in all communities needs to be quantified locally and points to the need to prioritize prompt interventions and focus on high-hazard regions, specifically addressing key factors such as transportation cost and accessibility. Identifying the burdened communities and the correlation of safety is crucial to formulating new policies and safety initiatives, especially the spatial indices and their components. For instance, transportation costs may be high, and infrastructure may be deficient, e.g., sidewalks, pedestrian crossings, and bicycle lanes, forcing pedestrians and bicyclists to share the road space with vehicles. The health differences may be improved by implementing active living strategies and healthier lifestyles through improved walking and bicycling infrastructure, especially in tracts where health and safety may have particularly high correlations. Furthermore, addressing environmental issues in some communities may have safety benefits, e.g., through improved visibility. Regarding socio-demographics, younger drivers may benefit from customized and effective training programs and stricter driving regulations (e.g., graduated driving licensing). These considerations and exploration of customized strategies can create a more efficient transportation system and enact safety measures for all road users.

This research was not without limitations. One significant challenge was that spatial indices were defined only recently, making it difficult to analyze them temporally and spatially simultaneously. Moreover, the indices may need additional refinement and validation to improve their reliability and utility in future studies. Additionally, census tracts change every few years, complicating the retrieval of certain variables since they are reported based on outdated census tract numbers. Future research could broaden its geographic focus to include more areas, especially those with more fatal pedestrian/bicycle crashes. Adding more variables and applying explainable machine learning methods can enrich the understanding of contributing factors and improve targeted safety interventions.

References

Ailshire, J., Melendez, R., Chenoweth, M., & Gypin, L. (2023). *National Neighborhood Data Archive (NaNDA): Street Connectivity by Census Tract and ZIP Code Tabulation Area, United States, 2010 and 2020* (Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR38580.v2

Al-Mahameed, F. J., Qin, X., Schneider, R. J., & Shaon, M. R. R. (2019). Analyzing Pedestrian and Bicyclist Crashes at the Corridor Level: Structural Equation Modeling Approach. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(7), 308-318. https://doi.org/10.1177/0361198119845353

Chandran, A., Sousa, T. R., Guo, Y., Bishai, D., Pechansky, F., & Vida No Transito Evaluation, T. (2012). Road traffic deaths in Brazil: rising trends in pedestrian and motorcycle occupant deaths. *Traffic Injury Prevention*, *13 Suppl 1*, 11-16. https://doi.org/10.1080/15389588.2011.633289

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

Das, S., Tamakloe, R., Zubaidi, H., Obaid, I., & Alnedawi, A. (2021). Fatal pedestrian crashes at intersections: Trend mining using association rules. *Accident Analysis & Prevention*, *160*, 106306.

DiMaggio, C. (2015). Small-area spatiotemporal analysis of pedestrian and bicyclist injuries in New York City. *Epidemiology*, *26*(2), 247-254. https://doi.org/10.1097/EDE.0000000000000222

Edirisinghe, P. A., Kitulwatte, I. D., & Senarathne, U. D. (2014). Injuries in the vulnerable road user fatalities; a study from Sri Lanka. *Journal of Forensic and Legal Medicine*, *27*, 9-12. https://doi.org/10.1016/j.jflm.2014.07.002

Evenson, K. R., LaJeunesse, S., & Heiny, S. (2018). Awareness of Vision Zero among United States' road safety professionals. *Injury Epidemiology*, *5*(1), 21. https://doi.org/10.1186/s40621-018-0151-1

Evenson, K. R., LaJeunesse, S., Keefe, E., & Naumann, R. B. (2023). Mixed-methods approach to describing Vision Zero initiatives in United States' municipalities. *Accident Analysis & Prevention*, *184*, 107012. https://doi.org/10.1016/j.aap.2023.107012

Fleisher, A., Wier, M. L., & Hunter, M. (2016). A Vision for Transportation Safety: Framework for Identifying Best Practice Strategies to Advance Vision Zero. *Transportation Research Record: Journal of the Transportation Research Board*, *2582*(1), 72-86. https://doi.org/10.3141/2582-09

Goel, R., Jain, P., & Tiwari, G. (2018). Correlates of fatality risk of vulnerable road users in Delhi. *Accident Analysis & Prevention*, *111*, 86-93. https://doi.org/10.1016/j.aap.2017.11.023

Grafova, I. B., Freedman, V. A., Lurie, N., Kumar, R., & Rogowski, J. (2014). The difference-in-difference method: assessing the selection bias in the effects of neighborhood environment on health. *Econ Hum Biol*, *13*, 20-33. https://doi.org/10.1016/j.ehb.2013.03.007

Haddad, A. J., Mondal, A., Bhat, C. R., Zhang, A., Liao, M. C., Macias, L. J., Lee, M. K., & Watkins, S. C. (2023). Pedestrian crash frequency: Unpacking the effects of contributing factors and racial disparities. *Accident Analysis & Prevention*, *182*, 106954. https://doi.org/10.1016/j.aap.2023.106954

Handayani, A., et al. (2019). *Inverse Distance Weighting interpolation on the optimum distribution of kernel -Geographically Weighted Regression for land price* Earth and Environmental Science, San Fransisco, CL.

Kronenberg, C., Woodward, L., DuBose, B., & Weissman, D. (2019). Achieving Vision Zero: Data-Driven Investment Strategy to Eliminate Pedestrian Fatalities on a Citywide Level. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*(1), 146-156. https://doi.org/10.3141/2519-16

Liu, J., Li, X., & Khattak, A. J. (2020). An integrated spatio-temporal approach to examine the consequences of driving under the influence (DUI) in crashes. *Accident Analysis & Prevention*, *146*, 105742. https://doi.org/10.1016/j.aap.2020.105742

Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A. J., & Liu, J. (2021). Understanding how relationships between crash frequency and correlates vary for multilane rural highways: Estimating geographically and temporally weighted regression models. *Accident Analysis & Prevention*, *157*, 106146. https://doi.org/10.1016/j.aap.2021.106146

Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-spatial Information Science*, *16*(2), 113-119. https://doi.org/10.1080/10095020.2013.766396

Mokhtarimousavi, S. (2019). A Time of Day Analysis of Pedestrian-Involved Crashes in California: Investigation of Injury Severity, a Logistic Regression and Machine Learning Approach Using HSIS Data. *Institute of Transportation Engineers. ITE Journal*, *89*(10), 25-33.

Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, 1-16. https://doi.org/10.1016/j.ijtst.2020.01.001

National Highway Traffic Safety Administration. (2022). *Fatality Analysis Reporting System (FARS)* (U.S. Department of Transportation. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars

Olszewski, P., Szagala, P., Rabczenko, D., & Zielinska, A. (2019). Investigating safety of vulnerable road users in selected EU countries. *Journal of safety research*, *68*, 49-57. https://doi.org/10.1016/j.jsr.2018.12.001

Onieva-Garcia, M. A., Martinez-Ruiz, V., Lardelli-Claret, P., Jimenez-Moleon, J. J., Amezcua-Prieto, C., de Dios Luna-Del-Castillo, J., & Jimenez-Mejias, E. (2016). Gender and age differences in components of traffic-related pedestrian death rates: exposure, risk of crash and fatality rate. *Inj Epidemiol*, *3*(1), 14. https://doi.org/10.1186/s40621-016-0079-2

Pljakic, M., Jovanovic, D., & Matovic, B. (2022). The influence of traffic-infrastructure factors on pedestrian accidents at the macro-level: The geographically weighted regression approach. *Journal of safety research*, *83*, 248-259. https://doi.org/10.1016/j.jsr.2022.08.021

Polus, A., Pollatschek, M. A., & Farah, H. (2005). Impact of infrastructure characteristics on road crashes on two-lane highways. *Traffic Inj Prev*, *6*(3), 240-247. https://doi.org/10.1080/15389580590969210

REPLICA. (2022). *AADT, TMC & Speeds* https://www.replicahq.com/traffic-datasets

Robert B. Noland, & Quddus, M. A. (2004). Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data. *Transportation Research Record: Journal of the Transportation Research Board*, *1897*(1). https://doi.org/10.3141/1897-04

Sekadakis, M., Katrakazas, C., Michelaraki, E., Kehagia, F., & Yannis, G. (2021). Analysis of the impact of COVID-19 on collisions, fatalities and injuries using time series forecasting: The case of Greece. *Accident Analysis & Prevention*, *162*, 106391. https://doi.org/10.1016/j.aap.2021.106391

Sirinapa Aryuyuen, W. B., and Thidaporn Supapakorn. (2014). Zero inflated negative binomial-generalized exponential distribution and its applications. *Songklanakarin Journal of Science and Technology*, *36*(4), 483-491.

Sosik-Filipiak, K., & Osypchuk, O. (2023). Identification of Solutions for Vulnerable Road Users Safety in Urban Transport Systems: Grounded Theory Research. *Sustainability*, *15*(13). https://doi.org/10.3390/su151310568

Stacker. (2022). *Cities in Florida with the most living in poverty*. https://stacker.com/

Tuckel, P. (2021). Recent trends and demographics of pedestrians injured in collisions with cyclists. *Journal of safety research*, *76*, 146-153.

Ukkusuri, S., Hasan, S., & Aziz, H. M. A. (2011). Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, *2237*(1), 98-106. https://doi.org/10.3141/2237-11

Vanlaar, W., Mainegra Hing, M., Brown, S., McAteer, H., Crain, J., & McFaull, S. (2016). Fatal and serious injuries related to vulnerable road users in Canada. *Journal of safety research*, *58*, 67-77. https://doi.org/10.1016/j.jsr.2016.07.001

Xiao, D., Xu, X., & Duan, L. (2019). Spatial-Temporal Analysis of Injury Severity with Geographically Weighted Panel Logistic Regression Model. *Journal of Advanced Transportation*, *2019*, 1-15. https://doi.org/10.1155/2019/8521649

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

Zhai, X., Huang, H., Sze, N. N., Song, Z., & Hon, K. K. (2019). Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accident Analysis & Prevention*, *122*, 318-324. https://doi.org/10.1016/j.aap.2018.10.017