CPBS
Center for Pedestrian and Bicyclist Safety

A Tier-1 University Transportation Center

Enhancing Collaboration through Web-based Visualization and Analysis of Traffic Crash Data

July 2024

A Report From the Center for Pedestrian and Bicyclist Safety

Su Zhang
University of New Mexico

Tyler A. Eshelman
University of New Mexico

Lisa L. Sinclair
University of New Mexico

Nicholas N. Ferenchak
University of New Mexico

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.

TECHNICAL DOCUMENTATION

1. Project No. 23UNM03	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Enhancing Collaboration through Web-based Visualization and Analysis of Traffic Crash Data		**5. Report Date** July 2024 **6. Performing Organization Code** N/A
7. Author(s) Su Zhang https://orcid.org/0000-0002-0396-2518 Tyler A. Eshelman https://orcid.org/0000-0002-5994-8707 Lisa L. Sinclair https://orcid.org/0009-0006-1164-1083 Nicholas N. Ferenchak https://orcid.org/0000-0002-3766-9205		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Center for Pedestrian and Bicyclist Safety Centennial Engineering Center 3020 The University of New Mexico Albuquerque, NM 87131		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** 69A3552348336
12. Sponsoring Agency Name and Address United States of America Department of Transportation Office of Research, Development, and Technology (RD&T)		**13. Type of Report and Period Covered** Final Report – June 2023 to May 2024 **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes

Report accessible via the CPBS website https://pedbikesafety.org and DOI https://doi.org/10.21949/qz45-3057

16. Abstract

It is widely accepted that road traffic safety is a significant public health issue. One of the effective ways to improve road traffic safety is analyzing crash data to understand where traffic accidents occur, identify associated spatial and temporal patterns, and determine causation. In the State of New Mexico, locations of traffic accidents are currently visualized using a variety of static maps. Although these statics maps are easier to create and producers can control how users view the data, they are not able to visualize crash density information because users cannot zoom in or zoom out, and hence cannot identify any associated spatial and temporal patterns. To solve the problems inherent with the current static maps, this research project focused on exploring the utility of dynamic and interactive web mapping and visualization techniques to visualize and analyze traffic crash data with the aim of helping transportation professionals determine the causes of traffic crashes and identify high-crash locations and other associated spatial and temporal patterns, and ultimately, achieving improved safety, enhanced resiliency, and increased efficiency for road users. This was achieved exclusively with open source tools and the implementation of well-known geospatial statistical analysis tools proven effective in traffic safety analysis.

17. Key Words	**18. Distribution Statement**
Traffic Crashes; Safety; Web Applications; Mapping; Analysis	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	41	

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

Enhancing Collaboration through Web-based Visualization and Analysis of Traffic Crash Data

A Center for Pedestrian and Bicyclist Safety Research Report

July 2024

Su Zhang
Earth Data Analysis Center
Department of Geography and Environmental Studies
University of New Mexico

Tyler A. Eshelman
Earth Data Analysis Center
University of New Mexico

Lisa L. Sinclair
Earth Data Analysis Center
University of New Mexico

Nicholas N. Ferenchak
Department of Civil Engineering
University of New Mexico

TABLE OF CONTENTS

List of Figures

List of Tables

Acronyms, Abbreviations, and Symbols

API	Application Programming Interface
AOI	Area of Interest
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CRS	Coordinate Reference System
DBMS	Database Management System
GCS	Geographic Coordinate System
GDAL	Geospatial Data Abstraction Library
GUI	Graphical User Interface
HTML	Hypertext Markup Language
JSON	JavaScript Object Notation
KDE	Kernel Density Estimation
NMDOT	New Mexico Department of Transportation
PCS	Projected Coordinate System
UTM	Universal Transverse Mercator

Abstract

It is widely accepted that road traffic safety is a significant public health issue. One of the effective ways to improve road traffic safety is analyzing crash data to understand where traffic accidents occur, identify associated spatial and temporal patterns, and determine causation. In the State of New Mexico, locations of traffic accidents are currently visualized using a variety of static maps. Although these static maps are easier to create and producers can control how users view the data, users cannot customize these maps to meet their special needs. New maps need to be created for any update or modification. More importantly, these static maps are not able to visualize crash density information because users cannot zoom in or zoom out, and hence they cannot be used to identify any associated spatial and temporal patterns. Subsequently, it is challenging if not impossible for users to conduct additional analyses to determine the causes of traffic crashes in an efficient, effective, and accurate manner. To solve the problems inherent with the current static maps, this research project focused on exploring the utility of dynamic and interactive web mapping and visualization techniques to visualize and analyze traffic crash data with the aim of helping transportation planners, engineers, and policymakers determine the causes of traffic crashes and identify high-crash locations and other associated spatial and temporal patterns, and ultimately, achieving improved safety, enhanced resiliency, and increased efficiency for road users. This was achieved exclusively with open source tools and the implementation of well-known geospatial statistical analysis tools proven effective in traffic safety analysis.

Executive Summary

Analyzing crash data is considered one of the most efficient methods to enhance road traffic safety. By analyzing the data, transportation professionals can better understand where traffic crashes occur, recognize associated patterns and trends, and develop countermeasures to prevent future crashes. It helps to improve understanding of the underlying causes of accidents and allows professionals to take the necessary steps to mitigate them. Crash data analysis is crucial in creating effective road safety policies and implementing appropriate traffic enforcement strategies.

The State of New Mexico currently relies on conventional and static maps to visualize the locations of traffic crashes. While static maps are easier to create and allow producers to control how users view the data, they have limitations in terms of customization and cannot be used to visualize crash density information. Additionally, users cannot dynamically and interactively zoom in or zoom out on these maps, making it difficult to identify associated spatial and temporal patterns. That said, this approach is sufficient to understand and analyze the complex relationships between different factors that contribute to road traffic accidents, such as the behavior of drivers, road conditions, weather, and other environmental factors.

This study aims to explore the utility of dynamic, interactive web mapping and visualization techniques to address the limitations of static maps. Implementing these techniques, would enable the creation of interactive web maps that could provide insights into traffic crash patterns, trends, and hotspots in a more customizable and efficient manner. In addition, these techniques can provide near real-time information on crash density, allowing for a more proactive approach to road safety management. By identifying high-crash locations and associated spatial and temporal patterns, transportation professionals can allocate resources more effectively and develop targeted interventions to reduce the occurrence of traffic crashes. The goal is to help transportation professionals determine the causes of traffic crashes and identify high-crash locations and other associated spatial and temporal patterns. Ultimately, this can lead to improved safety, enhanced resiliency, and increased efficiency for road users, benefiting all road users.

The creation of the Crash Mapping Prototype required a diverse range of technical skills and tools, including proficiency in different programming languages and relational databases. Through the application of these technologies, an interactive web mapping tool was designed and developed to provide users with a visual representation of traffic crash data collected within the state of New Mexico. The use of these tools enabled the creation of a customized and efficient application that allows for more in-depth and interactive analysis of traffic crash data.

Introduction

It is generally acknowledged that road traffic safety is a serious public health issue. According to the World Health Organization (WHO), motor vehicle crashes kill more than 1 million people and seriously injure approximately 20 to 50 million people around the world each year, affecting all road users such as vehicle drivers and passengers, pedestrians, bicyclists, and transit users (World Health Organization, 2023; U.S. Department of Transportation, 2017). In the United States, approximately 37,000 people on average are killed and an estimated 2.3 million injured each year in motor vehicle accidents over the last ten years (U.S. Department of Transportation, 2017). Transportation agencies, legislators, and advocacy organizations in the United States have been prioritizing efforts to improve road safety through a variety of strategies, such as road safety audits, speed management, geometric design, and safety performance measurement and evaluation, among others, however, there is still an ongoing commitment to prioritize and improve road safety even further.

According to the latest statistics from the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation, the State of New Mexico has the fifth highest motor-vehicle fatality rate in the United States at 18.8 fatalities per 100,000 residents (67.0% higher than the national average). In addition, New Mexico has the highest pedestrian fatality rate in the United States at 3.8 pedestrian fatalities per 100,000 residents (89.4% higher than the national average) (National Highway Traffic Safety Administration, 2021). Moreover, New Mexico is often ranked among the most dangerous states for bicyclists (4-6). Subsequently, there is a significant need for improved traffic safety in the State of New Mexico. The Federal Highway Administration's (FHWA's) Focused Approach to Safety Program has designated the State of New Mexico as one of fifteen "Pedestrian Bicycle Safety Focus States" and the City of Albuquerque, New Mexico's largest city with 26.1% of the state's population, as one of twenty-six "Continuing Safety Focus Cities

Analyzing crash data is an effective method for improving road safety. Crash data helps to expound the underlying causes of crashes, such as driver error, weather conditions, vehicle malfunctions, or infrastructure deficiencies. With this information, transportation professionals can develop more comprehensive and practical safety programs that address the root causes of crashes and help to reduce their frequency and severity. Additionally, by examining the patterns and trends present in crash data, transportation experts can identify locations that are particularly prone to crashes, such as busy intersections or stretches of highways with sharp turns. With this knowledge, they can practice specific safety measures, such as adding traffic signals, improving road signage and street lighting, or making changes to the road surface to proactively help prevent crashes from occurring in the first place.

Among all the factors involved, understanding where traffic crashes occur and identifying associated spatial and temporal patterns are crucial first steps in implementing effective road safety management and allocating appropriate traffic enforcement. Currently, the adopted method for visualizing where traffic accidents occur in the State of New Mexico uses a variety of static maps, primarily in PDF format (see Figure 1). Although these maps are easy to create and producers can control how users view the data, users cannot customize these maps to meet their special needs. New maps need to be created for any update or modification and once distributed, it becomes impossible to update these maps. More importantly, these static maps are not able to visualize crash density information because users are unable to zoom in or out and view different areas at varying scales, requiring separate maps for each level of detail. As a result, these maps fail to illustrate spatial and temporal patterns associated with crashes. Additionally, to maintain clarity and focus, the inclusion of data layers must be limited. Consequently, users face significant challenges, if not impossibilities, when attempting to efficiently, effectively, and accurately analyze the causes of traffic crashes.

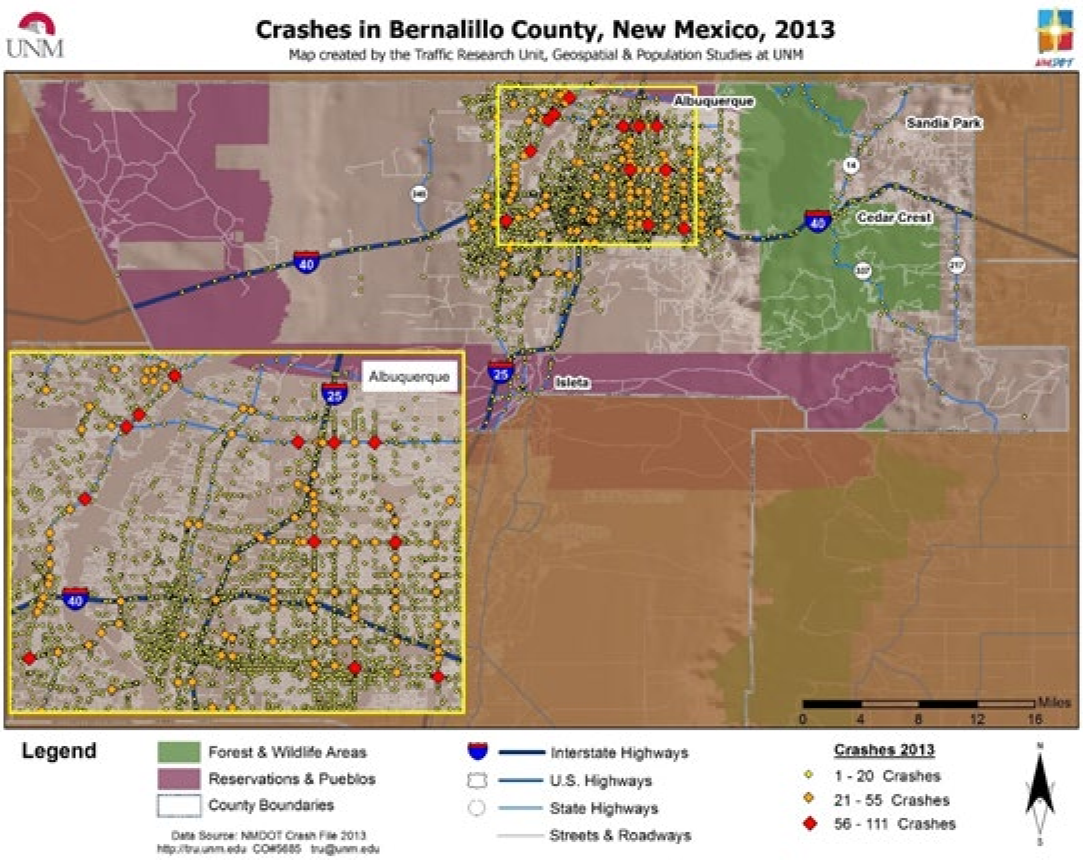

Figure 1. A static traffic map in PDF format (source: Geospatial and Population Studies at the University of New Mexico).

To prevent traffic crashes and ultimately reduce injuries and fatalities, particularly those involving pedestrians and bicyclists, it is imperative to perform a comprehensive analysis of recorded data to identify patterns, trends, and key risk factors that contribute to these crashes. However, analyzing recorded data is challenging because crashes typically involve multiple factors such as time, roadway and weather conditions, driver maneuvers, user characteristics, and varying modes and severities. While there has been a steady march of progress in the field of traffic safety analysis over the last several decades, which leads to the development of approaches ranging from safety performance functions to more complex statistical models that account for underlying spatial and temporal correlation derived from the shared effects of unobserved factors, many of those methodologies are complex and out of reach for the majority of small to medium transportation agencies and community organizations. In contrast, conventional tools such as static maps or online crash portals lack the dynamic and interactive capabilities to effectively communicate comprehensive crash trends across different scales and relevant variables, failing to provide a holistic understanding of the underlying patterns and causations. To address these aforementioned limitations, it is crucial to explore and implement advanced, dynamic, and interactive web-based mapping and visualization techniques for visualizing and analyzing recorded traffic crash data.

Users of the web application will be able to visualize maps of hot/cold spots and crash density and by providing easy-to-understand visual representations of crash data, users can gain a better understanding of where and how crashes are occurring and can use this information to advocate for safety improvements in their local area. Application programming interfaces (APIs) of the developed web application will be made available to transportation agencies at all levels (e.g., federal, state, local, and tribal). This will provide them with the means to create their own visualization and analysis web applications, diagnose their traffic safety issues, implement evidence-based strategies, and develop their own safety culture.

Objectives

To solve the problems inherent with current static maps and proprietary web portals, this proposed project will be focused on exploring the utility of dynamic, interactive, and free programming techniques to visualize and analyze traffic crash data. Techniques that will be explored and implemented with the portal include, but are not limited to, spatial data management and visualization, spatial analysis, and internet mapping. Many spatial analysis tools will be investigated and developed for the proposed web portal. These spatial analysis tools will be developed with the Python programming language and Geospatial Data Abstraction Library (GDAL). Specifically, these spatial analysis tools include point pattern analysis, High/Low Clustering (Getis-ord General G), Hot/Cold Spot Analysis (Getis-ord Gi*), Global and Local Moran's I, and Kernel Density. Point pattern analysis will be leveraged to identify the central tendency, dispersion, and directional trends of crashes in an area of interest (AOI). High/Low Clustering will be leveraged to examine if a specific attribute associated with traffic crashes exhibits a pattern of either high values or low values being clustered. Hot/Cold Spot Analysis will be leveraged to identify statistically significant hot spots or colds spots of a specific traffic crash typology. Global and Local Moran's I will be leveraged to examine if traffic crashes are spatially auto-correlated on a global or local scale. Kernel Density will be leveraged to examine the density of traffic crashes in a neighborhood around those crashes. Noting that it is likely that not all these tools will be used in the final version of the web portal, but input from stakeholders will be sought to determine which tools should be incorporated in the final web portal. The web portal, spatial analysis tools, and all associated APIs will be developed in a Python framework (e.g., Flask). The portal's web applications graphical user interface (GUI) will be developed with Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript, and Python. The backend programming interface (i.e., APIs) will be developed with the Python programming language.

The expected outcomes of this project are a web portal for crash visualization and analysis (of interest to users within New Mexico) and the APIs for further adoption, development, and innovation of additional tools (of interest to users across the United States). A web mapping and analysis portal will be developed to enable: (1) online visualization – displaying the locations and other associated attributes of traffic crashes; (2) online analytics – analyzing and displaying crash hot/cold spots, crash density, and crash counts in an AOI; and (3) map export – downloading the hot/cold spots maps and crash density maps. APIs for the proposed web portal's applications will be developed and freely shared with transportation agencies and community organizations at all levels (e.g., local, state, federal, and tribal) to enable them to develop similar web applications, further spreading a culture of safety. The proposed project also aims to produce two peer-reviewed journal articles. The first journal article will delve into the architecture and design of the web mapping and analysis portal developed as part of the project. The second journal article will shift the focus towards discussing the practical utility and benefits of the web mapping and analysis portal. A guidebook will be developed to provide detailed instructions on utilizing the web mapping and analysis portal, along with its associated APIs, which will be focused on workforce development and technology transfer (T2). The project team will follow a comprehensive program to disseminate the project outcomes. The project's publicity program includes: (1) developing a project website; (2) disseminating through peer-exchange network (e.g., journals, webinars, conferences, etc.); (3) hosting training workshops to educate prospective users (both inside and outside New Mexico) on the use of the web portal and its APIs for workforce development and T2; (4) social and public media; and (5) disseminating the developed web portal and its APIs through the New Mexico Local Technical Assistance Program (NM LTAP) Center, which has connections with all communities and government agencies across New Mexico.

Literature Review

Road Traffic Safety

The United States held a prominent position in traffic safety throughout most of the 20th century; however, during recent decades, this has not been the case any longer (7). By 2002, the United States had dropped to 16th place in deaths per registered vehicle and 10th place in deaths per distance traveled (Evans, 2004). In recent decades, nearly every high-income country has made more rapid progress in reducing road traffic deaths and death rates per kilometer of vehicle travel compared to the United States. Consequently, the United States can no longer claim a high rank in road safety on a global scale. Between 1994 and 2004, road fatalities in the United States increased by 5%, reaching 42,636 compared to 40,716. In contrast, most other developed countries witnessed substantial reductions in fatalities during the same period. In 2003, speeding-related crashes accounted for the deaths of 13,380 individuals (31.4% of all road traffic fatalities) in the United States (Evans, 2003). The growing population and number of vehicles contribute to an increase in road accidents and a significant loss of lives each year. To prevent road crashes in the United States, effective measures need to be implemented across all domains, including engineering, education, and enforcement. However, one significant barrier to producing these measures lies in the difficulty of analyzing the contributions of road crashes. To address this, map-based crash data system based on available spatial data are proposed as a potential solution to enhance the comprehension of road accident data and facilitate the development of better preventive measures (Leelakajonjit, 2012).

Traffic safety analysis utilizing Geographic Information Systems (GIS) has emerged as a critical tool for understanding and mitigating road accidents and pedestrian fatalities. GIS technology allows researchers and policymakers to analyze spatial data related to traffic incidents, identify high-risk areas, and implement targeted interventions to improve safety.

Traditionally, traffic safety analysis relied on manual methods and basic statistical analyses. However, the integration of GIS technology has revolutionized the field by enabling spatial analysis of traffic data. GIS software is commonly used for processing and visualizing spatial data related to crashes, road infrastructure, traffic flow, and environmental factors. These tools allow researchers to perform spatial queries, hotspot analysis, network analysis, and spatial interpolation to identify patterns and correlations in traffic safety data. Proprietary GIS software platforms offer intuitive graphical user interfaces (GUIs) that facilitate these processes. A wide range of built in tools can streamline common traffic safety analysis tasks and workflows. While proprietary GIS software offers comprehensive functionality and support, it typically involves costly licensing fees and ongoing maintenance costs. For organizations with limited budgets open-source alternatives present an attractive option.

Tools

Python is a high-level programming language known for its simplicity and versatility. It's widely used across various industries and disciplines, including data science, web development, automation, scientific computing, and geospatial analysis (Shamroukh & Aziz, 2023). Python's popularity stems from its readability, extensive standard library, and large ecosystem of third-party libraries and frameworks.

GDAL is an open-source library for reading and writing raster and vector geospatial data formats. Developed by the Open Source Geospatial Foundation (OSGeo), GDAL provides a set of tools and utilities for working with geospatial data in various formats such as GeoTIFF, Shapefile, and GeoJSON. GDAL supports data manipulation, transformation, reprojection, and analysis, making it a fundamental component of many geospatial workflows and applications.

In the context of traffic safety analysis, Python and GDAL are commonly used together to perform spatial analysis, data processing, and automation tasks (Watson & Ryan, 2024). Python provides a flexible and powerful programming environment, while GDAL offers a comprehensive set of geospatial tools for working with traffic-related datasets such as crash data and road networks. By leveraging Python's scripting capabilities and GDAL's geospatial functionality, researchers can develop custom workflows to analyze traffic patterns, identify high-risk areas, and generate actionable insights for improving road safety.

Static vs. Interactive Maps

Static maps, commonly produced using GIS software, provide a snapshot of traffic safety data at a specific point in time. They offer several advantages, including simplicity of creation, ease of distribution, and compatibility with various devices. Static maps are useful for presenting aggregated data, such as crash density maps or spatial distributions of pedestrian fatalities. However, they have limitations in terms of interactivity and dynamic visualization. Static maps may not effectively convey temporal trends or allow users to explore data interactively, limiting their utility for in-depth analysis.

Interactive web maps, powered by GIS technology and web-based mapping platforms like Google Maps API, Leaflet, and OpenLayers, offer dynamic visualization and user interaction capabilities (Lindell, 2020). They allow users to explore traffic safety data in real-time, customize map layers, and perform on-the-fly spatial analysis. Interactive web maps are particularly effective for presenting complex spatial relationships, temporal patterns, and interactive dashboards. They enhance user engagement and facilitate data-driven decision-making. However, developing and maintaining interactive web maps requires technical expertise, and they may be resource-intensive in terms of data processing and server requirements. Additionally, user experience may vary depending on internet connectivity and device compatibility.

Geospatial Analysis

Several approaches have been established in the literature for geospatial crash analysis. These spatial statistics methods are employed to comprehend the features of both the spatial and temporal distribution of road traffic accidents (Alam et al., 2023). Spatial statistics for spatial autocorrelation analysis can be divided into global and local indices. A global index analysis enables understanding the distribution of incidents across the network, distinguishing between clustered, scattered, or randomly distributed data. A local index analysis is employed to precisely identify the location and size of each detected cluster (Cheng et al., 2019). Global indices include Global Moran’s I and Getis-Ord G, while local indices include Anselin Moran’s I, Kernel Density Estimation, and Getis Ord Gi*.

Global Indices

Global Moran's I is a statistical measure used in spatial analysis to assess the overall spatial autocorrelation or the degree of spatial clustering or dispersion in a dataset (Shahzad, 2020). Global Moran’s I is represented as

$$I = \frac{N \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} w_{ij} (X_i - \underline{X})(Xj - \underline{X})}{(\Sigma_{i \neq 1}^{N} \Sigma_{j=1}^{N} w_{ij})(\Sigma_{i=1} (X_i - \underline{X})^2)}$$

The null hypothesis of Global Moran’s I states the attribute being analyzed is randomly distributed among the features in the study area. It quantifies the similarity between attribute values at different locations across a geographic area, taking into account both the values themselves and their spatial relationships. Global Moran's I ranges from -1 to 1, where values close to 1 indicate strong positive spatial autocorrelation (clustering), values near -1 indicate strong negative spatial autocorrelation (dispersion), and values around 0 indicate spatial randomness. The statistical test excels in measuring spatial clustering when both high and low attribute values cluster together and is useful to measure broad trends in accidents. Global Moran's I has been increasingly utilized in traffic safety and crash severity research to identify and analyze the spatial patterns and hotspots of road crashes (Zandi et al., 2022). This spatial autocorrelation measure helps in understanding how crash severity is distributed across different regions and is pivotal for developing targeted interventions for enhancing road safety.

The Getis-Ord General Index, also known as the Gi statistic or Getis-Ord General G statistic, measures the degree of clustering for either high or low values. Its conceptual foundation was initially laid by Getis and Ord in 1992, with subsequent refinement from Ord and Getis in 1995. Originally rooted in point pattern analysis, its earliest iteration comprised a ratio denoting the count of observations within a specified range of a point against the total count of points. However, in its broader manifestation, the index is employed in appraising values at neighboring locations, as delineated by spatial weights.

Getis-Ord General entails computing the ratio of the weighted average of values across neighboring locations to the summation of all values, including the value at the focal location (x_i). The global equation is represented as:

$$G = \frac{\Sigma_i \Sigma_j w_{ij}(X_i X_j)}{\Sigma_i \Sigma_j (X_i \Sigma_j)}$$

The statistic is most appropriate when data have a fairly even distribution and analysis is looking for unexpected spatial spikes of high or low values. However, it's important to note that in situations where both high and low values exhibit clustering tendencies, they might counterbalance each other, thus mitigating the effectiveness of this tool. The null hypothesis states that there is no spatial clustering of feature values. Consequently, upon obtaining a small and statistically significant p-value from the analysis, the null hypothesis is rejected. The sign of the z-score assumes significance in this context. A positive z-score is indicative of clustered high attribute values within the study area. Conversely, a negative z-score signifies the clustering of low attribute values within the study area. A z-score near zero indicates no observed clustering of high or low values.

Getis-Ord General is used in traffic safety studies to identify hotspots and low-risk areas for traffic crashes. This statistic is commonly applied in studies aiming to determine high-risk areas and spatially analyze traffic crashes resulting in fatalities. For example, Moradi et al. (2016) utilized Getis-Ord G to identify hotspots and low-risk areas for traffic crashes resulting in pedestrian deaths in Tehran. Similarly, Requia et al. (2015) employed Getis-Ord General to characterize spatial patterns of vehicle emissions along main traffic routes in the Federal District of Brazil. Getis-Ord General was applied in a study of road safety where the statistic combined with crash rate identified significant hot spots of traffic accidents on specified road segments (Berhanu et al., 2023). Additionally, the Getis-Ord method's application underscored its utility in validating hotspot identification and enhancing the accuracy and reliability of road traffic accident analysis (Berhanu et al., 2023).

While both indices share similarities, they diverge fundamentally in their calculation approaches. Moran's I evaluates spatial autocorrelation by comparing the value of each feature with those of neighboring features, computing the average of the values in the neighborhood while excluding the reference or core feature (Mohammed et al., 2023). This index aims to determine whether similar values tend to be proximate within the study area. Conversely, the Getis-Ord General calculates the average based on all features in the neighborhood, including the reference feature. It assesses whether a particular feature exhibits significantly high or low values compared to its neighboring features, thus identifying local clustering (Bombom et al., 2022). In interpretation, Moran's Index identifies clusters based on similarity within neighborhoods, with a high positive index indicating clustering of high values and a low negative index suggesting clustering of low values. Conversely, the Getis-Ord Index highlights significant hot spots or cold spots in the dataset, with a high positive index indicating a hot spot where the feature and its neighbors have significantly high values compared to the rest of the dataset.

Local Indices

Employed with a set of weighted features, the Anselin Local Moran's I statistic detects significant hot spots, cold spots, and spatial outliers, revealing local variations in spatial patterns (Anselin, 1995). It measures the degree of spatial autocorrelation in a local context and is sensitive to spatial outliers within the data, identifying areas where a particular value is significantly higher or lower than its neighbors. This sensitivity provides insights into localized patterns of similarity or dissimilarity, making it particularly useful for identifying spatial clusters and outliers within a dataset. The Local Moran's I statistic is expressed as:

$$I_i = \frac{x_i - \underline{X}}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_i - \bar{X})$$

Where:

$$S_i^2 = \frac{\sum_{j=1,j\neq i}^{n} w_{i,j}(x_j - \bar{X})}{n-1}$$

x_i is the value of an attribute for local feature i, $\underline{X}$ is the mean of the corresponding attribute, w_{ij} is the spatial weight between feature i and its neighbor j, x_j is the value of an attribute for local feature j and n is the total number of features.

After computing Local Moran's I for each feature, a map or list of values is generated, with interpretation based on both sign and magnitude. Positive values denote spatial clustering, indicating that features with similar values tend to be close together, while negative values signify dispersion, where features with dissimilar values cluster (Blazquez et al., 2020). Values near zero suggest randomness or spatial independence. Significance testing, often done through permutation tests, helps determine if observed spatial patterns are unlikely to have occurred by random chance.

To assess statistical significance, z scores of local Moran's I at each observation can be calculated. A large positive z score (> +1.96) suggests significant similarity among observations and their neighbors, forming spatial clusters of either high or low values. Conversely, a large negative z score (< -1.96) indicates significant dissimilarity, identifying spatial outliers.

Local Moran's I has been employed in numerous traffic-related spatial studies. Mohaymany et al. (2017) utilized Moran's I in developing crash prediction models based on Traffic Analysis Zone (TAZ)-level crashes in Mashhad, Iran. The study found significant Moran's I values, indicating spatial autocorrelation in crash frequencies within the TAZs. Additionally, the study demonstrated the reliability of spatial models over conventional Generalized Linear Models (GLMs) by analyzing residuals using Moran's I. An increase in population in Turkey has resulted in an increase in the number of vehicles on the road and the number of resulting traffic accidents (Haybat et al., 2022). A study completed in the five central districts of Bursa, Turkey utilized Local Moran's I to find that only some areas showed statistically significant clustering of traffic accidents in relation to surrounding highways which also had high-clustering (Haybat et al., 2022). In their 2014 study, Chaney and Kim analyzed the spatial distribution of bicycle collisions in Cincinnati, Ohio, employing both global and local Moran's I statistics. Their findings revealed substantial clustering of these accidents in the downtown and southwest regions of the city. The authors concluded that out of the 51 neighborhoods examined, 10 exhibited a significant clustering effect. In another study completed in the Maule region of Chile, researchers employed the local Moran's I index to pinpoint statistically significant clustering of high bicycle crash-related attribute values, specifically focusing on High-High (HH) clusters to identify zones with elevated crash risk (Blazquez et al., 2020).

Kernel Density Estimation (KDE) is a widely recognized technique that creates a smooth, continuous surface to depict spatial patterns, effectively addressing data scarcity by interpolating a discrete density surface. The process involves calculating the density of events at each point, adjusted for the distance to each event, thereby generating a continuous risk map.

The fundamental equation for KDE is:

$$f(x) = \frac{1}{nh} \Sigma_{i=1}^{n} \frac{K(x - x_i)}{h}$$

Here, x_i represents the value of the variable x at location i, n indicates the number of locations, h the search radius or bandwidth, and K the kernel function, which adjusts based on distance and bandwidth, as detailed by Silverman (2018).

Despite the variety of kernel functions available, such as Gaussian, Quartic, and Triangular, studies like those by Yu et al. (2014) suggest that the choice of kernel function has minimal impact on outcomes. However, the selection of bandwidth is crucial, significantly influencing results, with no perfect method for its determination.

KDE is particularly effective in identifying areas with high occurrences of traffic accidents. It divides the study area into cells, overlays a symmetrical, curved surface on each accident location, and aggregates these values within a given radius to estimate density. The density is highest at the accident location and decreases with distance, reflecting a distance decay effect.

Selecting the right grid size and bandwidth is vital for accurate results. A narrow bandwidth can reveal local fluctuations, useful for detailed analyses, while a larger bandwidth smooths out variations, providing a general overview. The choice depends on factors such as computational resources, sample size, and the nature of the data analyzed.

KDE is a widely used method in traffic safety studies to estimate the density of spatial point events such as traffic accidents. Xie et al. (2008) introduced a network KDE approach to estimate the density of traffic accidents, considering the impacts of different kernel functions, lixel lengths, and search bandwidths. Erdogan et al. (2008) developed a GIS-aided traffic accident analysis system to transform textual data into tabular form and georeference them onto highways for the management of accident analysis and identification of hot spots. Anderson (2009) utilized GIS and KDE to study the spatial patterns of injury-related road accidents in London, UK, and created a classification of road accident hotspots using clustering methodology. Mohaymany et al. (2013) proposed a GIS-based method for detecting high-crash-risk road segments using network KDE, which helps traffic engineers and safety specialists identify segments requiring more safety attention. Nie et al. (2015) focused on detecting spatial cluster patterns and riskier road segments of traffic crashes in Wuhan, China, using a network-constrained integrated method combining density estimation and spatial autocorrelation. Lee et al. (2019) incorporated crash severity into hot spot analysis to enable more informed decision-making regarding highway safety. Kazmi et al. (2020) suggested the use of KDE technique and GIS technology to automatically identify accident hotspots in the UK, emphasizing the increasing research interest in integrating GIS for accident analysis and safety management. Audu et al. (2021) applied Geographic Information System (GIS) as an intelligent system for emergency responses in road traffic accidents in Ibadan, utilizing spatial and non-spatial data to determine dynamic distance variations for optimal route planning. Overall, the integration of KDE with GIS technology has proven to be valuable in analyzing traffic accidents, identifying hotspots, and improving road safety measures.

The Getis-Ord $Gi*$ statistic is a local measure of spatial autocorrelation designed for identifying spatial clusters or hot spots and cold spots within a dataset. Unlike global measures that summarize spatial autocorrelation with a single value for the entire dataset, $Gi*$ provides individual values for each location, allowing the detection of localized patterns. This statistic is calculated by comparing the sum of values within a defined neighborhood of each feature, including the feature's own value, to the overall mean of the dataset. This comparison helps determine whether the area exhibits a higher or lower concentration of a particular characteristic relative to the dataset as a whole. The outcome of the $Gi*$ statistic is usually presented as a z-score, which measures how many standard deviations the local sum is from the expected value under the assumption of spatial randomness. A high positive z-score indicates a hot spot, where high values cluster, whereas a high negative z-score points to a cold spot, signaling a cluster of low values. Additionally, a p-value is calculated to assess the significance of the z-score, making $Gi*$ a powerful tool in fields like epidemiology, crime analysis, and environmental studies, where understanding geographical distribution and clustering of data is crucial.

The general equation of Getis Ord Gi* is expressed as shown below, where x_i is the attribute value for feature j, w_{ij} is the spatial weight between feature i and j, indicating the spatial relationship or proximity. This weight is often 0 if j is outside the neighborhood of i. $\underline{X}$ is the mean of all attribute values of x_j, σ is the standard

$$G_i^* = \frac{\Sigma_{j=1}^{n} w_{ij} x_j - \underline{X} \Sigma_{j=1}^{n} w_{ij}}{\sigma \sqrt{[\frac{n \Sigma_{j=1}^{n} w_{ij}^2 - (\Sigma_{j=1}^{n} w_{ij})^2}{n-1}}}$$

Spatial analysis techniques such as Getis-Ord Gi* have been widely used in traffic safety studies to identify hotspots and patterns of road accidents. Nie et al. (2015) utilized a network-constrained integrated method to detect spatial cluster patterns and identify riskier road segments of traffic crashes in Wuhan, China. Soltani et al. (2017) explored spatial autocorrelation of traffic crashes based on severity at the traffic analysis zonal level in urban environments. Additionally, Hazaymeh et al. (2022) conducted a spatiotemporal analysis of traffic accident hotspots in the Irbid Governorate, Jordan, using the Getis-Ord Gi* technique within a GIS environment. Furthermore, Jackson et al. (2016) focused on rainfall impacts on traffic safety by examining the temporal and spatial distribution of rain-related fatal crashes in Texas. Rahman et al. (2018) analyzed road traffic accident fatalities in Bangladesh through a spatio-temporal characterization of fatality rates, integrating newspaper accounts and gridded population data. Mohebbi et al. (2019) investigated the impacts of dust storms on freeway safety and operations in Arizona using a modeling approach. Achu et al. (2019) conducted a spatio-temporal analysis of road accident incidents in Thrissur district, Kerala, India, using geospatial tools to delineate hotspots. Lastly, Saadat et al. (2019) focused on spatial analysis of driving accidents leading to deaths related to motorcyclists in Tehran. Overall, the use of spatial analysis techniques such as Getis-Ord Gi* has proven to be valuable in identifying hotspots, patterns, and riskier road segments in traffic safety studies, contributing to the development of effective safety countermeasures and policies.

Data and Methodology

Data and Storage

The project utilizes a point dataset of vehicle crashes collected by the New Mexico Department of Transportation (NMDOT). The data, ranging from 2012 to 2021, was provided in a Comma Separated Value (CSV) format, and contained coordinates in both the Geographic Coordinate System (GCS) and the Projected Coordinate System (PCS) Universal Transverse Mercator (UTM) Zone 13N. To prepare the data for use by the analysis tools and web application, a series of preprocessing steps were necessary. This included data normalization and the omission of irrelevant data before ingestion into a geospatially enabled database. There were exactly 421,765 usable data points that were extracted from the file with each year containing between 37,000 and 47,000 data points.

After carefully reviewing the CSV data, the following list of attributes was determined to be the most relevant to use in the tool calculations.

Table 1. Crash data fields used in geospatial analysis tools.

Data Field ID	Data Field Name
1	WHETHER ALCOHOL INVOLVED OR NOT
2	CRASH SEVERITY
3	DAY OF WEEK
4	NUMBER OF PEOPLE KILLED IN CRASH
5	NUMBER OF PEOPLE WITH INCAPACITATING INJURIES (CLASS A) IN CRASH
6	NUMBER OF PEOPLE WITH VISIBLE INJURIES (CLASS B) IN CRASH
7	NUMBER OF PEOPLE WITH POSSIBLE INJURIES (CLASS C) IN CRASH
8	NUMBER OF PEOPLE INJURED (CLASS A+B+C) IN CRASH
9	NUMBER OF PEOPLE NOT INJURED (CLASS O) IN CRASH
10	NUMBER OF VEHICLES, BICYCLES, AND PEDESTRIANS INVOLVED
11	NUMBER OF PEOPLE IN MOTOR VEHICLES
12	NUMBER OF PEOPLE NOT IN MOTOR VEHICLES
13	NUMBER OF MOTOR VEHICLES INVOLVED
14	TOTAL NUMBER OF PEOPLE IN CRASH

The attributes *Day Of Week*, *Alcohol Involvement*, and *Crash Severity* all required conversion from categorical to numerical data to enable their utilization in tool calculations. *Day Of Week* was converted from a string data type, the name of the day (e.g. 'Monday', 'Tuesday', etc.), to integer values of one through seven. *Alcohol Involvement* was converted from string values *'No'*, *'Not Involved'*, *'Yes'*, and *'Involved'* to Boolean values *True* and *False*. Lastly, *Crash Severity* was converted from string values *'Property Damage Only Crash'*, *'Injury Crash'*, and *'Fatal Crash'* to integer values 1, 2, and 3 respectively. Other categorical data was identified to be useful, such as *Lighting* and *Weather*, but requires further investigation and collaboration with the DOT to determine a meaningful numeric scale. Additionally, some data points were excluded due to a lack of spatial reference or missing attribute data required for meaningful calculations.

Once the data was correctly formatted, it was inserted into a PostgreSQL database. The PostgreSQL Database Management System (DBMS) was chosen to store the data as it is a widely used database with extensive documentation and can leverage the PostGIS extension, designed for storing and querying geospatial data. The data was inserted using the Python SQLAlchemy library, a powerful SQL query toolkit, which allows us to define the database schema in code and enables programmatic database interactions. These tools not only integrated well into our development workflow but also ensured data standardization, efficient retrieval and manipulation of large datasets, and the flexibility to scale and adapt the database schema as needed (Elmasri & Navathe, 2015).

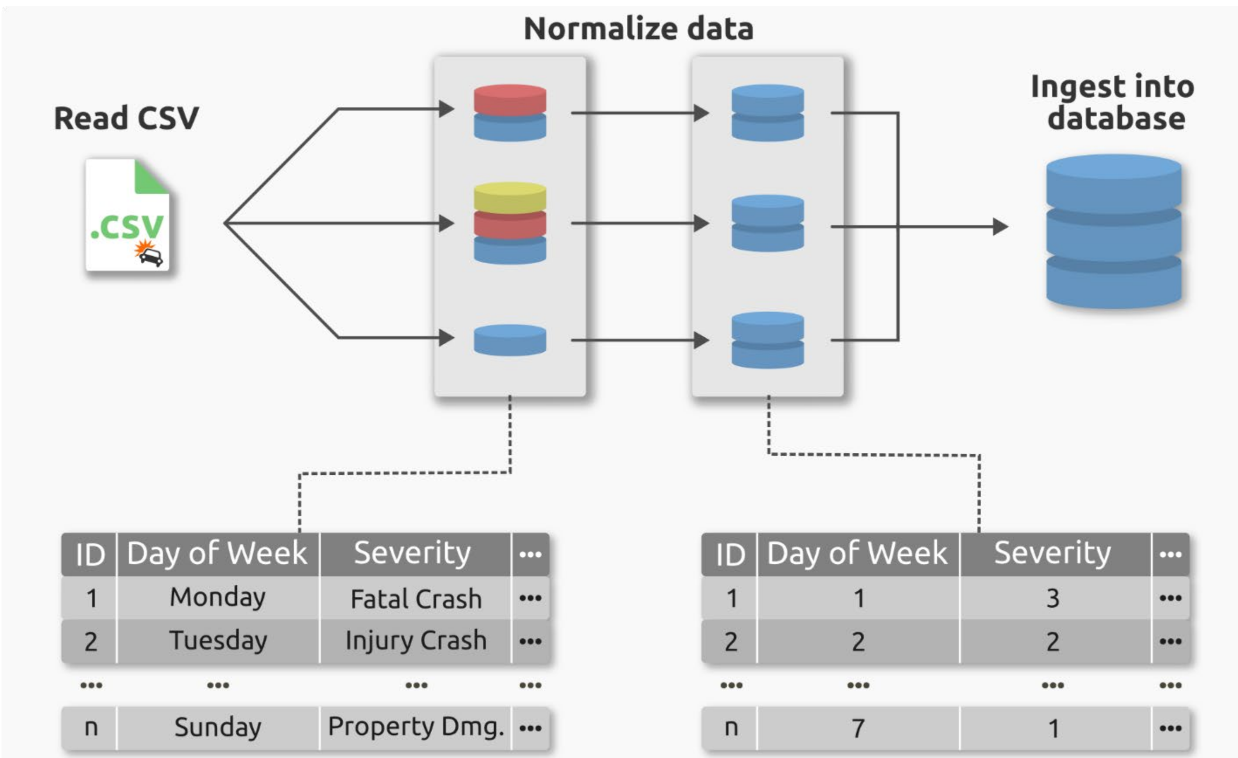

Figure 2. Depicts an overview of the normalization process when reading from a CSV. The figure illustrates how the categorical data is converted to numerical data for use in calculations before inserting into the database.

Data utilized in this research can be accessed directly from the NMDOT website. At the time of writing the request form is available at the following address: https://www.dot.nm.gov/planning-research-multimodal-and-safety/modal/traffic-safety/traffic-records/.

Web Portal

The web portal is a Graphic User Interface (GUI) that is comprised of a web map and several analysis tools that were developed to allow users to access and analyze the yearly crash data collected in New Mexico through an accessible web interface. The portal is built using JavaScript, HTML, and CSS programming languages and utilizes OpenLayers' open-source libraries to integrate mapping functionality seamlessly. OpenLayers was chosen for its compatibility with a wide range of spatial data formats and its flexibility to customize map styling and functionality. When accessing the web portal, the user is presented with a year select box, tools for querying, and a map with a hexbin layer summarizing the density of crashes for the first year in the dataset. These elements serve as inputs and outputs for interacting with the tools on the server. When a user runs a tool, the web portal generates a JavaScript Object Notation (JSON) object with the arguments for the tool and makes a request to the server. This hands off the processing task to the API which passes the same function arguments from the web portal to the corresponding tool functions, rather than running the process locally on the user's machine. This approach ensures consistent processing and outputs across all client environments and eliminates the need for users to own high-performance equipment to run the tools effectively (Kulawiak et al., 2019).

Application Programming Interface (API)

The API is the intermediary between the web portal, the database, and various functions and modules comprising the system. The API was developed using Python Flask, a micro web framework providing a robust foundation for web development. Flask was selected as the base for our API because of its simplicity, flexibility, and extensive documentation, making it ideal for rapid web development using the Representational State Transfer (REST) architectural software model. This model enables a user's web browser to send a request to the server, the server processes the request, and a server-generated response is returned to the browser. For example, when a user loads the main page a GET request is sent to the server to fetch the root endpoint (crash-mapping.edacnm.org/). The server then processes the request and returns a response with the data to render the page. By adhering to the REST principles, the API offers a standardized interface for accessing and manipulating the crash data stored in the PostgreSQL database. Additionally, Flask integrates easily with SQLAlchemy which streamlines database interactions, allowing for efficient data retrieval and manipulation which enhances the functionality and performance of the API. A range of public API routes were established to provide access to the system's tools and functionalities. These routes enable communication between the web portal and the backend services allowing users to retrieve, analyze, and visualize crash data with ease. These routes are accessible directly from the web portal or can be integrated into custom scripts and applications, offering flexibility to users of all skill levels. Documentation for the API is available at https://crash-mapping.edacnm.org/api/documentation/, which provides detailed information on the available endpoints, request parameters, and response formats. The documentation serves as a valuable resource for developers, facilitating integration with our system and promoting collaboration and innovation.

The API forms the backbone of our system, facilitating seamless communication between the web portal, the database, and accompanying modules. By leveraging Flask, SQLAlchemy, and RESTful principles, our API offers a robust and scalable solution for accessing and analyzing crash data in New Mexico. With its intuitive interface, extensive documentation, and flexible architecture, the API empowers users to explore and understand transportation safety trends, ultimately contributing to informed decision-making and improved road safety initiatives.

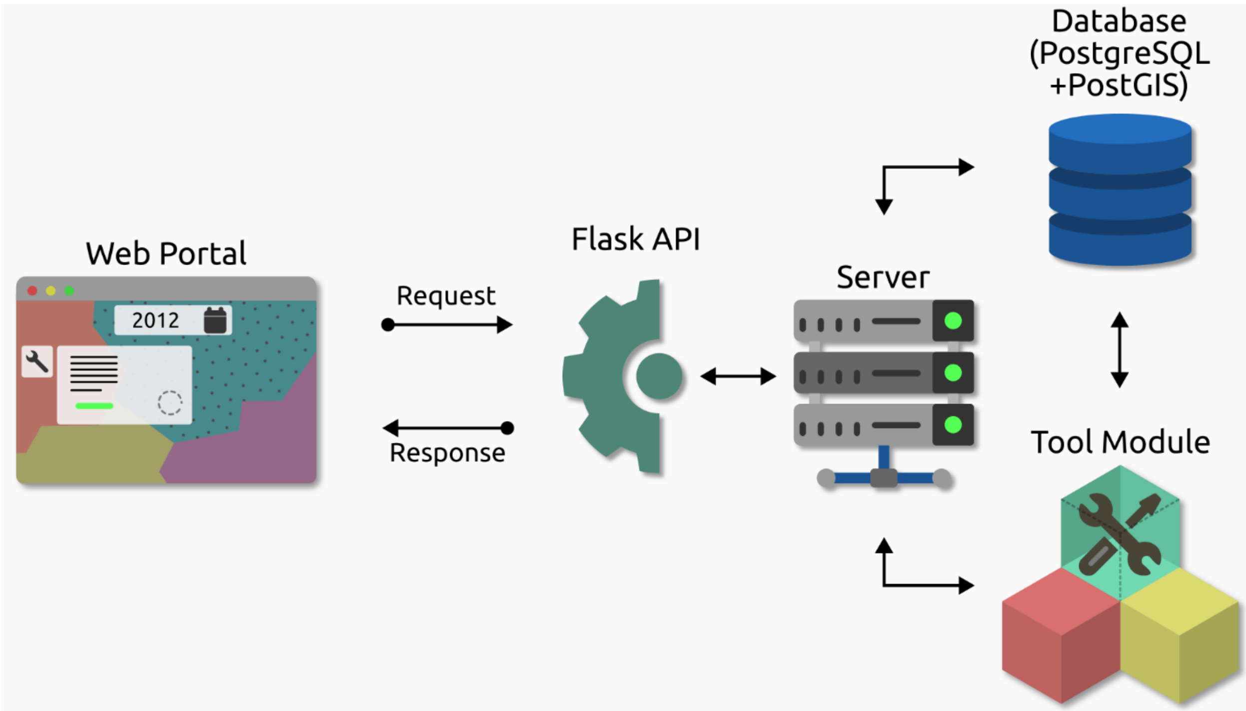

Figure 3. Illustrations of the connections between the Web Portal (client), the API, the Tool Module, and the Database.

Tool Module

The tool module is the core set of analysis tools that were identified to be most useful for crash mapping analysis. The tools include Getis Ord General G, Getis Ord GI*, Point Density Estimation, Kernel Density Estimation, Global and Local Moran's I, Mean Center, and Median Center. The tools were implemented in a stand-alone module and imported into the API code base so that the tools can be used independently of both the web portal and API. The production version of the module was uploaded to the Python Package Index (PyPI) and is publicly accessible to present a level of transparency to stoke conversation and collaboration among the community to build and improve upon the tools. The tool module was implemented using the Python programming language and various supporting libraries including the Geospatial Data Abstraction Library (GDAL) which allows for the creation and manipulation of raster and vector data. GDAL is leveraged by the tools because of the level of control it gives over the data and metadata. The tool module can be found at https://pypi.org/project/crash-mapping-tools/0.0.2/.

Getis Ord General G

The Getis Ord General G tool calculates the General G statistic, which measures the degree of high and low clustering (hot and cold spots) of spatial features. It takes a GeoJSON object, a variable name, a spatial relationship type such as k-nearest neighbors or a distance threshold, and a value corresponding to the relationship as inputs. After converting the GeoJSON to a GeoDataFrame and extracting point coordinates, it creates a spatial weights matrix using the libpysal library based on the specified spatial relationship. If the data exceeds 10,000 rows, it utilizes scikit-learn's NearestNeighbors function for efficient computation of the numerator. Otherwise, it employs libpysal's G function directly. The function calculates the observed General G statistic, its expected value, variance, z-score, and p-value, indicating the significance of clustering. It returns these statistics for both the native and libpysal methods in a dictionary, or an error message if exceptions occur during calculation.

<u>Function Definition</u>
Function: General_G
Parameters:

- *this_json* (GeoJSON): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in Coordinate Reference System (CRS) EPSG: 26913 (UTM Zone 13N).
- *variable* (str): The name of the variable/column in the GeoJSON data to be used for the analysis.
- *spatial_relationship* (str): The type of spatial relationship to consider, either 'knn' (k-nearest neighbors) or 'distance' (distance threshold).
- *spatial_relationship_value* (int for float): The value for the specified spatial relationship, representing either the number of neighbors for KNN or the distance threshold.

Return Value: Returns a dictionary containing the calculated General G statistics, including the observed value, expected value, p-value, and z-score. If the calculation is successful, the dictionary will have keys 'native' and 'libpysal' (if applicable), each containing the respective statistics. If an error occurs during the calculation, the dictionary will have a 'status' key with the value 'error' and a 'message' key describing the error.

Getis Ord GI*

The Getis Ord Gi* tool calculates the Getis-Ord Gi* statistic for spatial clustering of given attribute values in a GeoJSON dataset. It takes a GeoJSON object, a variable name, a spatial relationship type ('knn' or 'distance'), and a value for that relationship as input. It extracts the points and corresponding variable values from the GeoJSON features. If the spatial relationship given is 'knn', it uses the query function from the scipy.spatial.cKDTree library to find the k-nearest neighbors, which is capped at 400 for performance limitations, for each point and calculates the Gi* statistic.

If the spatial relationship is 'distance', it uses the the query_ball_point function from the same scipy library to find the points within a specified radius, capped at 400 meters, and calculates the Gi* statistic similarly. The function then updates the GeoJSON features with the calculated z-score, p-value, and confidence level for the Gi* statistic and returns the a GeoJSON object.

<u>Function Definition</u>
Function: Gi_star
Parameters:

- *this_geojson* (GeoJSON): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).
- *variable* (str): The name of the attribute within the GeoJSON's features to be used for the Getis-Ord Gi* calculation.
- *spatial_relationship* (str): The type of spatial relationship to define the neighborhoods. Acceptable values include 'knn' (k-nearest neighbors) or 'distance' (based on a specified radius distance).
- *spatial_relationship_value* (int or float): The value that defines the neighborhood based on the spatial_relationship - if 'knn', it represents the number of nearest neighbors; if 'distance', it represents the radius (e.g., in meters) within which other points are considered as neighbors.

Return Value: Returns a modified version of the original GeoJSON dictionary containing the calculated Gi* statistic added to each feature's properties.

Point Density

The point density tool calculates the density of points found within each grid cell. This is the count of the number of points within a given cell divided by pixel resolution which produces the density per unit area returned in TIFF format. To calculate point density, the function requires a GeoJSON object, the resolution for the raster, and a directory path to save the file. The function first unpacks the points from the GeoJSON and calculates the min and max coordinates for all the points to get the bounds of the data. Next, the cell resolution is added to the x and y min values and subtracted from the min values. Once correctly formatted, the min and max values are used to calculate the range of x and y values. The ranges are then used to create a grid using the numpy meshgrid function. To get the counts for each cell, we iterate over all the crash points and subtract the minimum coordinates from the current point’s coordinates, round down using a floor function, and divide by the resolution. This gives us a cell index which is then incremented in an empty array, illustrated in Figure 4. Once all points have been accounted for, the array is written to a TIFF file and saved.

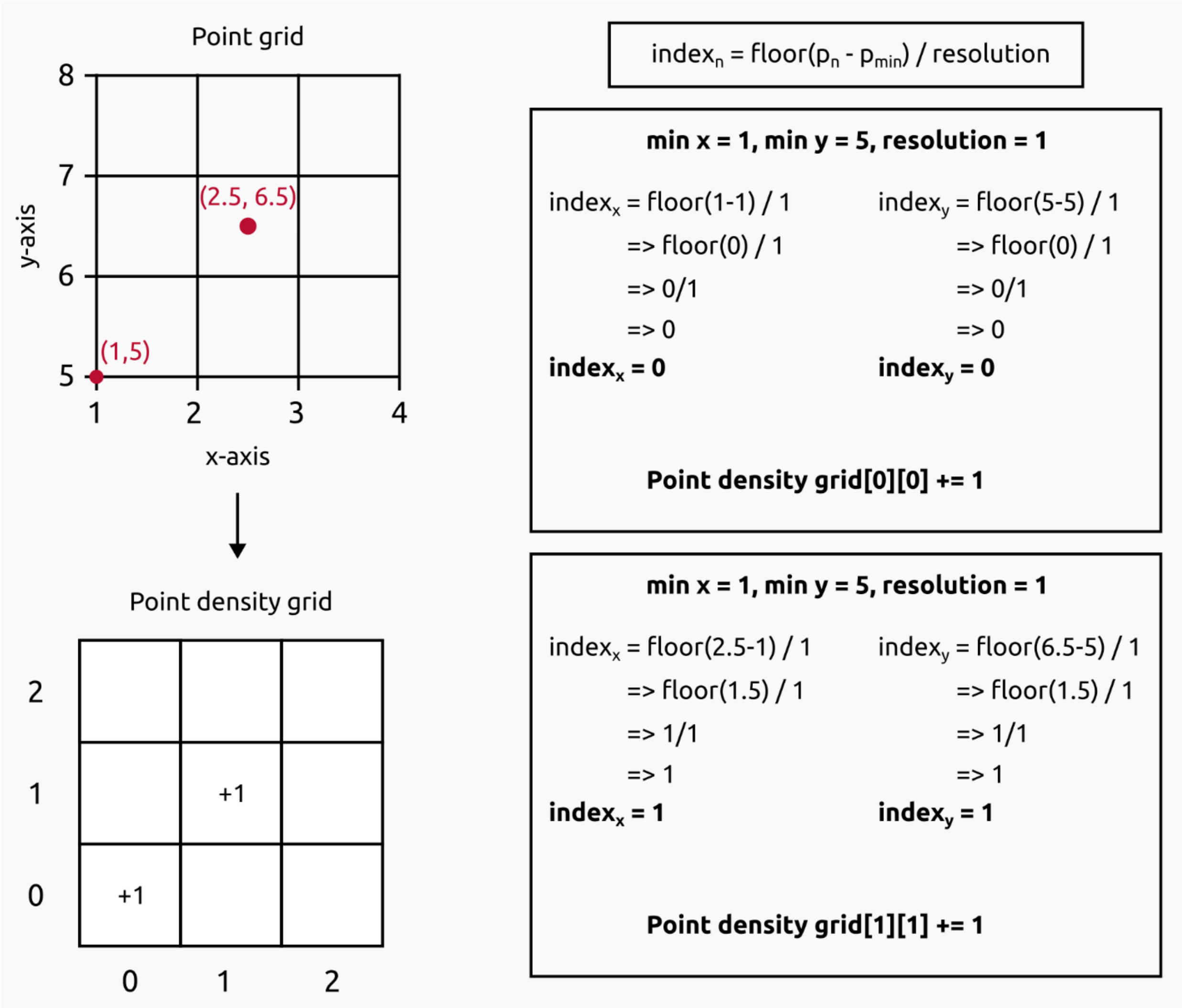

Figure 4. This figure illustrates how the point density formula used in the code is applied to points from a set of features to find the array indices to be incremented.

<u>Function Definition</u>
Function: 'point_denstiy'
Parameters:

- *geo_json* (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).
- *resolution* (float): Resolution of the cells to be output by the function. If given 3 the resolution will be 3x3 meter resolution.
- *data_dir* (str): Directory that the raster will be saved to (e.g. '/your/path/here/')

Return Value: Returns a point density image in TIFF format in CRS EPSG:3857 which is expected by OpenLayers map in web portal.

Kernel Density Estimation

The KDE tool calculates the density of crashes using a kernel function and outputs the result to a raster dataset as a TIFF. The kernel function used to calculate KDE is the 'KernelDensity' function from Sklearn's neighbor's library. This function allows you to choose from a list of pre-defined kernels to use including 'gaussian', 'tophat', 'epanechnikov', 'exponential', 'linear', and 'cosine'. The epanechnikov kernel was chosen because it has been shown to produce more accurate results using fewer data points when compared to a Gaussian kernel (Moraes et al., 2021). The function also allows you to set your own bandwidth radius. The bandwidth radius that we used was calculated using the Silverman's rule of thumb formula which has been shown to be less affected by outliers in the data (Hurley & Leslie, 2024). Silverman's formula is utilized in this research according to the following formula:

$$radius = 0.9 \cdot min(SD, \sqrt{\frac{1}{\ln(2)}} \cdot D_m) \cdot n^{-0.2}$$

To calculate the KDE, we use the crash data points from a GeoJSON and fit them to the Kernel Density function model. Then we create a grid using the resolution passed from the web portal and calculate the center points for each cell. The kernel is calculated on these coordinates. Using the Kernel Density function, we calculate the log densities for all the grid points and return log densities. To be usable in context to the original data, we convert the log densities to probability densities by taking the exponential of the data and then normalize the data by multiplying by the number of points of the original crash data.

<u>Function Definition</u>
Function: 'kde'
Parameters:

- *geo_json* (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).
- *resolution* (float): Resolution of the cells to be output by the function. If given 3 the resolution will be 3x3 meter resolution.
- *data_dir* (str): Directory that the raster will be saved to (e.g. '/your/path/here/')

Return Value: Returns a kernel density image in TIFF format in CRS EPSG:3857 which is expected by OpenLayers map in web portal.

Global Moran's I

The Global Moran's I tool calculates the Global Moran's I statistics for a set of points using a specified attribute and the number of nearest neighbors. To calculate these statistics the points and the number of nearest neighbors are first passed to the KNN function from the libpysal.weights.distance library to calculate the point weights. Once weights have been calculated, the points and their weights are given to the Moran function from the esda.moran library to generate the Global Moran's I statistics; the I value, expected I value, P value, and Z value. The statistics are then returned to the user in a python dictionary.

<u>Function Definition</u>
Function: 'global_morans_i'
Parameters:

- *geo_json* (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).
- *key* (string): The string definition of the attribute to use for calculation (e.g. 'alcohol_involvement')
- *k* (int): Number of nearest neighbors to calculate weights.

Return Value: Returns a dictionary with the Moran's I value, expected I value, z value, and p value.

Local Moran's I

The Local Moran's I tool calculates the local Moran's I statistics for a set of points using a specified attribute and the number of nearest neighbors. Similarly to the Global Moran's I tool, the points and the number of nearest neighbors are passed to the KNN function from the libpysal.weights.distance library to calculate the point weights. After the weights are calculated, the points and their weights are given to the Moran_Local function from the esda.moran library to generate the local I values, P values, and quadrants. These values are then mapped to their corresponding data point in an array and then filtered by whether they are significant. Significance is determined by whether their p-value is less than 0.05. The filtered list of points is then returned as a GeoJSON object. The points can then be visualized on the map to show which quadrant they belong; High-High, Low-High, High-Low, Low-Low.

<u>Function Definition</u>
Function: 'local_morans_i'
Parameters:

- *geo_json* (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).
- *key* (string): The string definition of the attribute to use for calculation (e.g. 'alcohol_involvement')
- *k* (int): Number of nearest neighbors to calculate weights.

Return Value: Returns a geojson of significant points with mappings to their respective quadrants.

Mean Center

The Mean Center tool is used to calculate the mean center of the point distribution. To calculate the mean center, points are extracted from the provided GeoJSON, and numpy's mean function is run on the list of point coordinates. Numpy was used for this tool because it is able to quickly calculate both the x and y coordinates by vectorizing the calculation. Vectorization is a method of performing calculations on an array without the need to use for loops, which can be resource intensive. Once the mean has been generated, the point is added to a GeoJSON object and returned.

Function Definition
Function: 'mean_center'
Parameters:

- geo_json (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).

Return Value: Returns a geojson with the point geometry of the mean center.

Median Center

The Median Center tool is used to calculate the geometric median of the point distribution by minimizing the sum of distances to each point. After extracting the points from the GeoJSON, the initial median value is set by calculating the mean of the points. This is used as a starting point for the rest of the calculation. We then iterate and calculate the distances from the point to the existing median. In the same iteration, median estimate is updated using a weighted mean of the points, where the weights are inversely proportional to the distances (closer points have higher weights). Iteration continues to update the median estimate until the median converges. Once iterations conclude, the median point is inserted into a GeoJSON object and returned.

Function Definition
Function: 'geometric_median'
Parameters:

- *geo_json* (geojson): A GeoJSON object containing spatial coordinates of the points to be analyzed. Points should be in CRS EPSG: 26913 (UTM Zone 13N).

Return Value: Returns a geojson with the point geometry of the mean center.

Results

The results of this study show that there are comprehensive, and relatively inexpensive, solutions available for analyzing data to help make informed decisions about traffic and pedestrian safety in place of static mapping mediums. Static maps work well to give an overview of a dataset but fall short when working with a large number of features over a temporal range. These issues are addressed by the integration of open-source software, tools, and programming techniques to build a web portal and community toolset capable of analyzing over a decade or crash data. With this study we have been able to address the inherent issue with static maps by providing a dynamic web portal that allows users to explore the data in an intuitive way.

The web portal provides the point data for a given year and a hexbin layer to show local densities calculated based on the current viewports zoom level. The hex bin sizes include state level, county level, city level, and block level bin sizes of 10 miles, 5 miles, 0.5 miles, and 50 feet radius' respectively which are recalculated each time a user zooms in or out of the map. This is done so that the dynamic map represents a more accurate depiction of the data as zoom levels, area size, and number of points within the viewport change. Along with this data overview, eight tools are available to users. Getis Ord General G, which allows users to identify clustering of high or low values; Getis Ord GI*, for Hot/Cold spot analysis; Point Density, for identifying cell density; Kernel Density, for finding density using a kernel function over a continuous surface; Global Moran's I, for assessing spatial autocorrelation; Local Moran's I, for identifying the degree of spatial autocorrelation and significant hot spots, cold spots, and outliers; Mean Center, to find the mean center for the point distribution; and finally the Median center, for identifying the median point distribution of the dataset. All of the tools can be accessed through the web portal, the API, or by downloading the tool set directly. This makes data analysis even more accessible and allows it to be integrated into different workflows by allowing users, of all different skill levels and backgrounds, the opportunity to be able to analyze crash data effectively.

Discussion

The development of the dynamic, interactive web portal for visualizing and analyzing traffic crash data marks a significant advancement over traditional static maps and proprietary web portals. This project effectively harnessed free programming techniques and sophisticated spatial analysis tools to provide comprehensive insights into traffic crash patterns and trends.

One of the core strengths of this project lies in its use of advanced spatial data management and visualization techniques. Using data handling techniques allows for optimizing data retrieval, improving performance, and solidifying a standardized data format to be used by the tools. Each spatial analysis tool demonstrated significant potential in understanding traffic crash data. For instance, point pattern analysis helps identify central tendencies, dispersion, and directional trends, which are crucial for identifying high-risk areas and informing traffic safety measures. High/Low Clustering and Hot/Cold Spot Analysis provide insights into areas with statistically significant concentrations of high or low crash values, aiding in the identification of problematic zones. Similarly, Global and Local Moran's I and Kernel Density analysis offer detailed views on spatial autocorrelation and crash densities, respectively, highlighting areas that require targeted interventions.

The technical foundation of the web portal is robust, leveraging the Python programming language and the GDAL library. The use of Python frameworks such as Flask for the backend, along with HTML, CSS, and JavaScript for the graphical user interface, ensured a seamless and user-friendly experience. This choice of technology stack facilitated the development of efficient and scalable tools and APIs, critical for the portal's performance and future scalability. Python, specifically, was chosen because of its vast range of libraries and its widely adopted use in the GIS community and other industries as using familiar tools helps to promote the project for use and development of custom tools and interfaces. By leveraging open-source tools to create this project, we have effectively managed to prove that there are cost-effective solutions for advanced spatial analysis.

Conclusions and Recommendations

In this study, it was concluded that analyzing crash data is an effective method for improving road safety. However, the current method for visualizing crash data in New Mexico using static maps has limitations physically and temporally. Therefore, this study aimed to overcome these limitations by developing a dynamic and interactive web application for visualizing crash data, enabling transportation professionals to develop analysis workflows and evidence-based strategies to reduce the frequency and severity of crashes. The methodology used in this study effectively analyzed and visualized the motor vehicle traffic crash data in the state of New Mexico. The GIS dataset containing coordinates and attributes of reported crashes from 2012 to 2021 was a comprehensive source of information for this investigation. Spatial statistics methods were employed to comprehend the features of both the spatial and temporal distribution of road traffic accidents. Moreover, tools were developed in an open source environment to perform analysis with global and local spatial statistics indices. It is recommended to further investigate and test the available fields to identify unforeseen useful information in the crash dataset to fully exploit its richness and the questions which can be answered with the tools developed.

The crash mapping prototype developed in this study provided an interactive web-based mapping tool enabling the users to visualize the yearly crash data gathered in New Mexico. The application allowed the users to choose from a list of years to be displayed, loaded the corresponding dataset into the map as a point layer, and generated hex bins and an accompanying legend with class density ranges. This study analyzed and created a web-based software to visualize crash data in the state of New Mexico, providing valuable insights into the patterns and trends of traffic accidents in the area. By splitting the portal into independent parts, the study provides easily accessible analysis tools for less technical users and a clear starting point for users who wish to further explore implementing their own solutions. This also opens the door for collaboration and tool optimization in future revisions of the code base. The crash mapping prototype developed in this study can be used by policymakers and city planners to identify high-risk areas and implement targeted interventions to improve traffic safety in the city. The methodology can also be applied to other regions to provide valuable insights into traffic safety and inform evidence-based policy decisions.

It is recommended that further investigation be conducted into creating data handling procedures and analysis workflows such that a dynamic mapping system can be updated and enabled to run analyses efficiently when new data is presented.

References

Achu, A. L., Aju, C. D., Suresh, V., Manoharan, T. P., & Reghunath, R. (2019). Spatio-temporal analysis of road accident incidents and delineation of hotspots using geospatial tools in Thrissur District, Kerala, India. *KN-Journal of Cartography and Geographic Information, 69*(4), 255-265.

Alam, M. S., & Tabassum, N. J. (2023). Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio. *Heliyon, 9*(5).

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention, 41*(3), 359-364.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93-115.

Audu, A. A., Iyiola, O. F., Popoola, A. A., Adeleye, B. M., Medayese, S., Mosima, C., & Blamah, N. (2021). The application of geographic information system as an intelligent system towards emergency responses in road traffic accident in Ibadan. *Journal of Transport and Supply Chain Management, 15*, 17.

Berhanu, Y., Schroeder, D., Teklu, B., & Alemayehu, E. (2023). Spatial analysis of road traffic accidents: Identifying hotspots for improved road safety in Addis Ababa, Ethiopia. *Cogent Engineering, 10*(2), 2269655.

Blazquez, C. A., Calderón, J. F., & Puelma, I. (2020). Towards a safe and sustainable mobility: Spatial-temporal analysis of bicycle crashes in Chile. *Journal of Transport Geography, 87*, 102802.

Bombom, L. S., Saidu, B. J., Akintunde, E. A., & Nyango, K. C. (2022). Explorative geospatial analysis of road traffic fatalities in Nigeria for the period 2005-2018. *FUDMA Journal of Sciences, 6*(3), 167-176.

Chaney, R. A., & Kim, C. (2014). Characterizing bicycle collisions by neighborhood in a large Midwestern city. *Health Promotion Practice, 15*(2), 232-242.

Cheng, Z., Zu, Z., & Lu, J. (2019). Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability, 11*(1), 160.

Elmasri, R., & Navathe, S. B. (2015). *Fundamentals of database systems* (7th ed.). Pearson Education.

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention, 40*(1), 174-181.

Evans, L. (2003). A new traffic safety vision for the United States. *American Journal of Public Health, 93*(9), 1384-1386.

Evans, L. (2004). *Traffic safety*. Science Serving Society.

Gedamu, W. T., Plank-Wiedenbeck, U., & Wodajo, B. T. (2024). A spatial autocorrelation analysis of road traffic crash by severity using Moran's I spatial statistics: A comparative study of Addis Ababa and Berlin cities. *Accident Analysis & Prevention, 200*, 107535.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189-206.

Haybat, H., Zerenoğlu, H., & Özlü, T. (2022). Temporal and spatial analysis of traffic accidents: The case of Bursa city. *International Journal of Geography and Geography Education,* (45), 404-423.

Hazaymeh, K., Almagbile, A., & Alomari, A. H. (2022). Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. *ISPRS International Journal of Geo-Information, 11*(4), 260.

Hurley, B. J., & Leslie, T. F. (2024). Spatiotemporal variograms as neighborhood definers. *Geographical Analysis, 56*(2), 404-424.

Jackson, T. L., & Sharif, H. O. (2016). Rainfall impacts on traffic safety: Rain-related fatal crashes in Texas. *Geomatics, Natural Hazards and Risk, 7*(2), 843-860.

Kazmi, S. S. A., Ahmed, M., Mumtaz, R., & Anwar, Z. (2022). Spatiotemporal clustering and analysis of road accident hotspots by exploiting GIS technology and kernel density estimation. *The Computer Journal, 65*(2), 155-176.

Kulawiak, M., Dawidowicz, A., & Pacholczyk, M. E. (2019). Analysis of server-side and client-side Web-GIS data processing methods on the example of JTS and JSTS using open data from OSM and geoportal. *Computers & Geosciences, 129*, 26–37.

Lee, M., & Khattak, A. J. (2019). Case study of crash severity spatial pattern identification in hot spot analysis. *Transportation Research Record, 2673*(9), 684-695.

Leelakajonjit, A. (2012, August 24-25). Advantages of map-based accident data [Conference presentation]. 5th ATRANS Symposium, Bangkok, Thailand.

Lindell, M. K. (2020). Improving hazard map comprehension for protective action decision making. *Frontiers in Computer Science, 2*, 27.

Mohammed, S., Alkhereibi, A. H., Abulibdeh, A., Jawarneh, R. N., & Balakrishnan, P. (2023). GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation planning. *Transportation Research Interdisciplinary Perspectives, 20*, 100836.

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.

Mohaymany, A., & Shahri, M. (2017). Crash prediction modeling using a spatial semi-local model: A case study of Mashhad, Iran. *Applied Spatial Analysis and Policy, 10*, 565-584.

Mohebbi, A., Green, G. T., Akbariyeh, S., Yu, F., Russo, B. J., & Smaglik, E. J. (2019). Development of dust storm modeling for use in freeway safety and operations management: An Arizona case study. *Transportation Research Record, 2673*(5), 175-187.

Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., Jamshidi, E., & Zeini, S. (2016). Spatial analysis to identify high risk areas for traffic crashes resulting in death of pedestrians in Tehran. *Medical Journal of the Islamic Republic of Iran, 30*, 450.

Moraes, C. P., Fantinato, D. G., & Neves, A. (2021). Epanechnikov kernel for PDF estimation applied to equalization and blind source separation. *Signal Processing, 189*, 108251.

National Highway Traffic Safety Administration. (2021). *Traffic safety facts annual report tables*. U.S. Department of Transportation.

Nie, K., Wang, Z., Du, Q., Ren, F., & Tian, Q. (2015). A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: A case study from Wuhan, China. *Sustainability, 7*(3), 2662-2677.

Rahman, M. K., Crawford, T., & Schmidlin, T. W. (2018). Spatio-temporal analysis of road traffic accident fatality in Bangladesh integrating newspaper accounts and gridded population data. *GeoJournal, 83*, 645-661.

Réquia Jr, W. J., Koutrakis, P., & Roig, H. L. (2015). Spatial distribution of vehicle emission inventories in the Federal District, Brazil. *Atmospheric Environment, 112*, 32-39.

Saadat, S., Rahmani, K., Moradi, A., & Darabi, F. (2019). Spatial analysis of driving accidents leading to deaths related to motorcyclists in Tehran. *Chinese Journal of Traumatology, 22*(3), 148-154.

Shahzad, M. (2020). Review of road accident analysis using GIS technique. *International Journal of Injury Control and Safety Promotion, 27*(4), 472-481.

Shamroukh, M., & Aziz, M. A. (2023). Revolutionizing geospatial analysis through the integration of GIS, GeoAI, and geospatial programming: A comprehensive review.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury, 48*(3), 637-647.

U.S. Department of Transportation. (2017, June). *Road safety fundamentals*. Federal Highway Administration. https://rspcb.safety.fhwa.dot.gov/RSF/unit1.pdf

Watson, R. B., & Ryan, P. J. (2024). Geospatial factors applied to road accidents: A review. *Journal of Advances in Information Technology, 15*(3).

World Health Organization. (2023, December 13). *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems, 32*(5), 396-406.

Yu, H., Liu, P., Chen, J., & Wang, H. (2014). Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis & Prevention, 66*, 80-88.

Zandi, K., Tavakoli Kashani, A., & Okabe, A. (2022). Influence of traffic parameters on the spatial distribution of crashes on a freeway to increase safety. *Sustainability, 15*(1), 493.