2024

U.S. Department of Transportation
Federal Highway Administration

NEBRASKA
DEPARTMENT OF TRANSPORTATION

NEBRASKA
TRANSPORTATION CENTER

Final Report SPR-FY23(025)

WBS Report #
26-0206-0081-001

Modeling Pedestrian and Bicyclist Crash Exposure with Location-Based Service Data

Yunwoo Nam, Ph.D.
Professor
Community and Regional Planning
University of Nebraska-Lincoln

Don Butler, P.E.
Highway Safety Engineer
Nebraska Department of Transportation

Mohammad Elayan, M.Sc.
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Nebraska-Lincoln

Jason Hawkins, Ph.D.
Assistant Professor
Department of Civil and Environmental Engineering
University of Nebraska-Lincoln

Nicholas Aldridge, M.Sc.
Engineer
Nebraska Department of Transportation

Kwang il (Jason) Yoo, P.S.M.
Graduate Research Assistant
School of Global Integrated Studies
University of Nebraska-Lincoln

Nebraska Department of Transportation Research
Headquarters Address
1400 Nebraska Parkway
Lincoln, NE 68509
ndot.research@nebraska.gov
(402) 479-4697
https://dot.nebraska.gov/business-center/research/

Nebraska Transportation Center
262 Prem S. Paul Research Center at Whittier School
2200 Vine Street
Lincoln, NE 68583-0851
(402) 472-1932
http://ntc.unl.edu

This report was funded in part through grant from the U.S. Department of Transportation Federal Highway Administration. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the U.S. Department of Transportation.

Modeling Pedestrian and Bicyclist Crash Exposure with Location-Based Service Data

Yunwoo Nam, Ph.D.
Professor
Community and Regional Planning
University of Nebraska Lincoln

Jason Hawkins, Ph.D.
Assistant Professor
Civil and Environmental Engineering
University of Nebraska Lincoln

Don Butler, P.E.
Highway Safety Engineer
Nebraska Department of Transportation

Nicholas Aldridge, M.Sc.
Engineer
Nebraska Department of Transportation

Mohammad Elayan, M.Sc.
Graduate Research Assistant
Civil and Environmental Engineering
University of Nebraska Lincoln

Kwang il (Jason) Yoo, P.S.M.
Graduate Research Assistant
School of Global Integrated Studies
University of Nebraska Lincoln

Final Report on Research Sponsored by

Nebraska Department of Transportation

and

Nebraska Transportation Center

University of Nebraska–Lincoln

May 2024

Technical Report Documentation Page

1. Report No. SPR-FY23(025)	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Modeling Pedestrian and Bicyclist Crash Exposure with Location-Based Service Data		5. Report Date May 2024
		6. Performing Organization Code
7. Author(s) Yunwoo Nam, Jason Hawkins, Don Butler, Nicholas Aldridge, Mohammad Elayan, Kwang il (Jason) Yoo		8. Performing Organization Report No. WBS # : 26-0206-0081-001
9. Performing Organization Name and Address Nebraska Research Center, University of Nebraska-Lincoln 2200 Vine Street, Prep Prem S. Paul Research Center Suite 262, Lincoln, NE 68583-0851		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. SPR-FY23(025)
12. Sponsoring Agency Name and Address Nebraska Department of Transportation 1500 Nebraska Parkway Lincoln, NE 68502		13. Type of Report and Period Covered Final Report, July 2022 – May 2024
		14. Sponsoring Agency Code MATC TRB RiP No. 34760
15. Supplementary Notes		

16. Abstract
The rising popularity of non-motorized transportation has brought about safety risks. The inclusion of accurate traffic volume information is one of the key elements to produce robust outcomes when researching safety. The project investigates and enhances transportation safety for pedestrians and bicyclists, emphasizing the integration of Location-Based Services (LBS) data, particularly StreetLight data, to analyze traffic volume and associated risks. The calibration process demonstrates the significance of traffic volume as a key variable influencing prediction accuracy across pedestrian, bicyclist, and vehicle models. The crash analysis reveals a strong correlation between crash counts and traffic activity, with noteworthy findings emerging at the facility scale. The safety ranking analysis identifies higher and lower risk areas in the city. The spatial and temporal analysis at the street segment level highlights changes in traffic volumes before and during the COVID-19 period, along with distinct geographic patterns of activities at downtown and recreational locations. The project contributes to the knowledge and methodology surrounding transportation safety for non-motorized road users. By leveraging LBS data, the research provides comprehensive analyses of traffic patterns and safety considerations, aiming to create a safer environment for pedestrians and bicyclists.

17. ORCID No. of each Researcher Jason Hawkins 0000-0002-9655-4846 Mohammad Elayan 0009-0001-2562-5694 Kwang il (Jason) Yoo 0009-0009-6835-5135 Yunwoo Nam 0000-0002-0143-5132	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 116	22. Price

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Annual Average Daily Traffic (AADT)
Akaike information criterion (AIC)
Application programming interface (API)
Central business district (CBD)
Census block group (CBG)
City of Lincoln (CoL)
Composite Normalized Score (CNS)
Composite Safety Scores (CSS)
EPA Smart Location Database (EPA-SLD)
Federal Highway Administration (FHWA)
Geographic Information System (GIS)
Global positioning system (GPS)
Greenhouse gases (GHG)
Housing Unit (HU)
Insurance Institute for Highway Safety (IIHS)
Location-based service (LBS)
Mean absolute error percentage (MAPE)
National Highway Traffic Safety Administration (NHTSA)
Nebraska Department of Transportation (NDOT)
Nebraska Transportation Center (NTC)
OpenStreetMap (OSM)
Root mean squared error (RMSE)
StreetLight Inc. (StL)
Support Vector Machine (SVM)
Traffic Analysis Zones (TAZ)
United States Department of Transportation (USDOT)
University of Nebraska-Lincoln (UNL)

Acknowledgments

The authors express their gratitude to the Nebraska Department of Transportation for sponsoring this research and providing invaluable support. Appreciation is extended to the Nebraska Department of Transportation, the City of Lincoln, and the StreetLight Data for their assistance in data collection. Special recognition is reserved for the Technical Advisory Committee, whose dedicated efforts have significantly enriched this work through constructive feedback.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily reflect the official views or policies neither of the Nebraska Department of Transportations nor the University of Nebraska-Lincoln. This report does not constitute a standard, specification, or regulation. Trade or manufacturers' names, which may appear in this report, are cited only because they are considered essential to the objectives of the report.

The United States (U.S.) government and the State of Nebraska do not endorse products or manufacturers. This material is based upon work supported by the Federal Highway Administration under SPR-FY(025). Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

This report has been reviewed by the Nebraska Transportation Center for grammar and context, formatting, and compliance with Section 508 of the Rehabilitation Act of 1973.

Executive Summary

The rising popularity of non-motorized transportation modes, such as walking and biking, can be attributed to the associated benefits of physical activity, improved health, enhanced psychological well-being, leisure, and a positive impact on environmental quality. However, the increase in walking and biking activities has brought about notable safety risks. The disproportionately higher risk of fatalities and serious injuries in incidents involving pedestrians and cyclists underscores the critical need for proactive measures to mitigate and reduce such crashes.

When developing traffic risk models and analyzing associated safety scenarios, the inclusion of accurate traffic volume information is one of the key elements in producing robust outcomes. This project introduced and utilized Location-Based Services (LBS) data to analyze pedestrian and bicyclist traffic volume and patterns.

This research report systematically pursues the project objectives in subsequent chapters, covering literature review, data collection, calibration processes, safety analysis, and spatiotemporal pattern analysis.

Chapter 2 provides a thorough literature review and discusses the connectedness among traffic volume, vehicle speed, vehicle characteristics, urban form, and crash risk. The chapter also discusses the importance of LBS data, with a focus on the data provided by StreetLight Data Inc. It underscores the enrichment of spatial and temporal traffic volume information using such data.

Chapter 3 details the challenges and methods involved in collecting and managing diverse datasets, showcasing the complexity of the research process. It provides insights into the

intricate nature of handling varied data sources and highlights the hurdles encountered in the research process.

In Chapter 4, the calibration process is examined using advanced techniques such as random forest models to effectively manage complex datasets and capture non-linear relationships. The results indicate that traffic volume stands out as the most crucial variable influencing prediction accuracy across all three models—pedestrian, bicyclist, and vehicle. Overfitting analysis reveals a reasonable model fit, with minimal disparities between training and testing R-squared values for all modes. Overall, the chapter presents the calibration process, model selection, and performance evaluation for predicting traffic volumes in various travel modes.

In Chapter 5, safety analysis and result interpretation are presented using census block group, link, and intersection data to incorporate both aggregate and facility-specific features. The research reveals variations in results based on the exposure measure employed to normalize crash counts. While factors like population and network density influence the safety of non-motorized road users on an aggregate level, the most noteworthy findings appear when examining infrastructure at the facility scale. The safety ranking analysis identifies higher and lower risk areas in the city.

Chapter 6 presents the analysis and visualization of spatial and intertemporal patterns in traffic volumes for pedestrians and bicyclists at the street segment level. A comparative examination of bicyclist and pedestrian traffic volumes between 2019 and 2020 reveals the change of volumes and the disparities in geographic patterns during the COVID-19 period. The analysis extends to include pedestrian activity patterns on weekdays and weekends, revealing distinct traffic variations at downtown and recreational locations. Spatial-temporal analysis

indicates that downtown and business districts show higher activities during peak hours. Additionally, the seasonal pattern analysis identifies a significant reduction in winter activities for pedestrians and bicyclists. Chapter 7 encompasses the project's benefits and concluding remarks.

In conclusion, this project contributes insights into the ongoing discussion on the safety of pedestrians and bicyclists, particularly in relation to the use of LBS data for transportation research. The outcomes play a pivotal role in shaping policies and strategies geared towards establishing a safer and more sustainable environment.

Chapter 1 Introduction

Non-motorized transportation modes, such as walking and biking, have gained increased popularity due to the benefits of physical activity, health, psychological well-being, leisure, and improved environmental quality. Studies highlight a diminished risk of cardiovascular and cancer mortality among those who commute using non-motorized modes (Panter et al., 2018). These travel modes have also demonstrated a positive impact on mental wellbeing (Martin et al., 2014); additionally, their environmental benefits include decreased greenhouse gas (GHG) and air pollutant emissions. While many trips are too long to be accomplished by non-motorized modes, about 21% of trips are less than one mile (Federal Highway Administration, 2017). Furthermore, non-motorized modes provide a more economical alternative to pricier transportation options, like transit and private vehicles. Walking can result in savings of 25 to 50 cents per vehicle-mile traveled (Litman, 2022). In comparison, as of 2010, constructing bicycle lanes can range from $10,000 to $50,000 per mile, whereas an urban area 4-lane road can cost between $8 million to $10 million (Litman, 2011).

Walking and biking can be facilitated or constrained by the physical form of communities in general as well as transportation system in particular. Therefore, fostering a walkable and bikeable environment emerges as a pivotal strategy to promote active living and cultivate a healthy community. In response to this, local governments are progressively considering how the design of communities and transportation systems will influence the active and healthy lifestyle of their residents.

Along with the popularity of walking and biking, the negative impact of crashes has also garnered attention. Total number of police-reported injured persons resulting from crashes on U.S. highways increased from 2.2 million in 2009 to 2.7 million in 2019 (BTS, 2019). Notably,

4.6% of these injuries were pedestrians or bicyclists, despite these modes only comprising 1.0% of total person-miles traveled (Federal Highway Administration, 2017). Furthermore, 19.2% of all fatalities were pedestrians or cyclists. From 2019-2022, 81 of the 946 fatalities on Nebraska roadways were pedestrians or cyclists. The elevated risk of fatalities and serious injuries in incidents involving pedestrians or bicyclists emphasizes the critical need for proactive measures to mitigate and prevent such crashes.

To enhance our understanding of pedestrian and bicyclist safety and design effective policy solutions, information about traffic movements that may lead to safety or crash issues is critically important. Upon reviewing research on transportation safety, it becomes evident that the inclusion of traffic volume information is one of the key elements to produce robust outcomes when developing transportation safety models and simulating associated risk scenarios, as there is a strong positive correlation between crash counts and traffic volumes (Høye & Hesjevoll, 2020; Retallack & Ostendorf, 2020). This correlation is an intuitive relationship; higher volumes of motorized traffic lead to increased interactions and potential conflicts with pedestrians and bicyclists, subsequently raising the risk of crashes. Therefore, both motorized and non-motorized traffic volumes emerge as crucial variables when assessing the safety of the transportation system. However, the availability of adequate data has been a major challenge in the research efforts about measuring pedestrian and bicyclist's travel volume and patterns. Traditionally, traffic volumes are approximated from manually collected traffic counts at a small number of specific locations. While this method offers a high level of accuracy for the selected locations, estimating traffic volumes at other locations without count data can be challenging. In recent years, the advancements of location-aware telecommunications

technologies have provided transport planners and engineers with new opportunities to study traffic volumes at a street level as a solution to this challenge.

Over the last decade, the advancement in information and communication technologies, coupled with geospatial technology, have introduced various new sources of transportation data. Many of these data sources use cellular phone locations, derived from a combination of Location-Based Service (LBS), Bluetooth, and global positioning system (GPS) pings. The characteristics and coverage of these data vary widely.

LBS data is typically categorized into two groups: mode-specified data and mode-unspecified data. Mode-specified data is the type of data derived from phone apps, such as Strava and Ride Report. It is called mode-specified data because the data provider receives the data with the travel mode specified by the user. These data sources have the benefit of detailed trip information because the user specifies the trip start and end times, as well as whether they are biking or walking. However, these data require the user to be using the app while making the trip, so they will not capture trips made by individuals outside their customer base. This feature means that the sample sizes are relatively small, and the data tend to be biased toward middle-aged individuals (Venter et al., 2023). Mode-unspecified data, such as LBS, can address these shortcomings.

Mode-unspecified data is usually more representative of the population than mode-specified data because there is no requirement for user interaction. Examples of mode-unspecified data providers include INRIX, Miovision, and StreetLight (StL). While datasets are large, a limitation of mode-unspecified data and many emerging data sources is that the accuracy of the total volume measurements on a road segment is generally unknown. Many trips do not appear in the data, whether because a road user does not own a phone with LBS capabilities, they

have turned off all location tracking services, or they use a service provider not purchased by the data vendor. Traffic volumes gathered by mode-unspecified data providers are only a sample and thus require calibration to reconcile differences between the LBS traffic indicator and measured traffic volumes. While commercial vendors provide simple calibration tools in their products, more rigorous calibration is often necessary to perform detailed analyses.

One significant application of LBS traffic data is safety analysis, particularly for non-motorized modes such as pedestrians and bicyclists. Traffic data for these modes has historically been unreliable due to scarcity. This data is important because higher traffic volumes, specifically higher pedestrian/cyclist volumes, can lead to more conflicts with motorized travel modes. As such, transportation engineers must have an accurate understanding of traffic patterns and volumes of different travel modes. Municipal and state governments collect traffic volumes through manual and automatic traffic counts, typically at intersections, to accomplish various planning and operations tasks. While these datasets offer considerable benefits, their applicability is often limited. Traffic counts recorded at specific intersections may not be universally transferable to other locations and timeframes, leading to challenges in making broad inferences. Mode-unspecified data can address this limitation by increasing the sample size and the geographic extent for which data is available.

To utilize mode-unspecified LBS data for safety research, it is essential to measure the accuracy of the data. This project seeks to verify the accuracy of a traffic volume dataset from StreetLight (StL) by creating traffic prediction models and calibrating the data against traditional traffic volume data gathered by the City of Lincoln (CoL).

1.1 Objectives

The research is driven by the motivation to enhance pedestrian and bicyclist safety through the utilization of LBS data. The primary objective of the project is to develop a reliable methodology for measuring pedestrian and bicyclist exposure and analyzing the associated safety.

To obtain the objective of the project, the following specific aims are outlined.

1. Review and Document Existing Methods: the project conducts a comprehensive review and documentation of existing methods employed for estimating the risk exposure of pedestrians and bicyclists, as well as factors contributing to crash and severity.
2. Creation of an Integrated Geodatabase: the project develops an integrated geodatabase designed for pedestrian and bicyclist safety research in Lincoln, Nebraska. This geodatabase incorporates calibrated LBS-based movement information categorized by street segment, intersection, and census block group (CBG) alongside place-based characteristics.
3. Traffic Data Calibration: the project develops calibration models. StreetLight data requires calibration against traffic volume to adjust their magnitude and temporal distribution.
4. Safety Analysis: The project conducts a comprehensive analysis of traffic crashes, using a multifaceted approach to investigate factors influencing the safety of pedestrians and bicyclists.
5. Spatial-temporal Analysis: The project develops reliable measures of pedestrian and bicyclist movements and conducts a spatial-temporal analysis of travel patterns and associated risks in Lincoln, Nebraska. The project produces hotspot maps that visually

highlight spatial variations in active transportation at the street-segment level. The GIS (Geographic Information System) models are utilized to identify locations of higher exposures.

1.2 Report Structure

The project focuses on analyzing pedestrian and bicyclist safety in Lincoln, Nebraska, leveraging LBS data obtained from StreetLight Data. The structure of the document is organized as follows: Chapter 2 provides an in-depth examination of the current practices in active transportation safety analysis and the utilization of LBS data by state Departments of Transportation (DOTs). Chapter 3 describes the data collection process and extraction methods along with detailing the various sources utilized in the process. Chapter 4 outlines the calibration process for LBS data to align it with observed multi-modal traffic counts sourced from the City of Lincoln. Chapter 5 presents safety analysis methods and the corresponding results. Chapter 6 presents spatial-temporal analysis and visualization as well as hotspot mapping. It also provides the comparison of weekdays and weekends, the analysis of peak hour volume and seasonal changes, and the examination of interconnectedness between walkable environment and walking volumes. Chapter 7 provides conclusive insights and recommendations to the DOT based on the project team's experience using LBS data to conclude this project.

Chapter 2 Literature Review

This project integrates two analysis streams. The first stream is non-motorized road user safety and its contributing factors. The second stream is LBS data used by state DOTs. Evaluated together, the research team can identify where LBS data may enhance safety analyses based on conventional data sources.

2.1 Non-Motorized Road User Safety Contributing Factors

Many factors influence non-motorized traveler safety, such as roadway design characteristics, demographics, urban form features, and environmental and weather conditions (Lee & Abdel-Aty, 2005; Rahman & Kockelman, 2020; Ukkusuri et al., 2012; Wang & Kockelman, 2013; Yue et al., 2020). Urban form features including population density, employment density, and transportation facility design have important roles in increasing or decreasing non-motorized travel (Zhao et al., 2018). Urban areas that contain substantial amounts of urban sprawl and set aside large portions of their transportation right-of-way for automobile infrastructure generally have fewer daily trips taken by non-motorized modes (Eldeeb et al., 2021). These urban form features affect the total travel time and safety of pedestrians and bicyclists. While total travel time is more influential for users' decision making (Witchayaphong et al., 2020), safety concerns for non-motorized road users require careful attention from transportation planners due to the higher fatality risk.

Another factor influencing non-motorized road user safety is roadway length. For every percent increase in roadway length, a greater than one percent increase in crashes is expected. The number of lanes on a roadway can also affect crash frequency (Ukkusuri et al., 2012). In addition to these local variables, macroscopic variables, such as socio-demographic composition, show significant correlations with non-motorized road user safety. Merlin et al. (2019) found that

for every percent increase in poverty, there is a 0.08 to 0.23 percent increase in crash rate. This result is intuitive as people in poverty and in areas with more mixed land uses are more likely to use non-motorized modes of travel.

Transportation safety inequity is exacerbated by disparities in access to safe and convenient pedestrian infrastructure frequently observed in economically disadvantaged communities. Examples of decreased access to these facilities include inadequate sidewalk maintenance, limited crosswalks, and poorly designed pedestrian pathways, which can also pose safety risks (Rajaee et al., 2021). Addressing these disparities is a critical component of promoting equitable access to transportation options and ensuring these populations benefit from safe and accessible pedestrian facilities.

Vehicle travel speed is among the principal factors influencing non-motorized fatalities. Research using 2007-2009 data found that the median impact speed for pedestrian injuries was 14 mph compared with 35 mph for pedestrian fatalities (Tefft, 2013). Within the 25 to 40 mph range, each one mph speed increase produced a roughly three percent increase in the fatality likelihood. Fatality likelihood rose to 90% at vehicle travel speeds of 55 mph or greater. Other research from Philadelphia found that roads with vehicle speeds of 55 to 65 mph generated roughly 1.6 times more crashes than would be expected on a 25 mph road (Merlin et al., 2019). Figure 2.1 illustrates the field of vision for drivers at various speeds. Increasing vehicle speed tends to diminish the effective area that a motorist perceives. Also, higher vehicle speeds increase the required perception reaction distance for drivers in cases of emergency braking or maneuvering. In summary, vehicle speed increases both the likelihood and severity of crashes involving non-motorized road users.

(a)10-15 mph

(b) 30-35 mph

(c) 40+ mph

Figure 2.1 Driver's Peripheral Vision at Standard Travel Speeds (National Association of City Transportation Officials, 2013)

A second major factor to consider is vehicle size. Larger vehicles tend to cause more severe outcomes for non-motorized road users when involved in crashes (Governors Highway Safety Association, 2019). This phenomenon is explained by these vehicles' higher momentum and center of gravity, which will place more stress on smaller automobiles and non-motorized road users. A related regulatory point is that neither the National Highway Traffic Safety Administration (NHTSA) nor the Insurance Institute for Highway Safety (IIHS) provide vehicle safety ratings for vehicle-pedestrian crashes—only those with other vehicles and fixed objects (Lee & Kockelman, 2022). While state DOTs do not have the oversight to regulate vehicle sales, several jurisdictions are addressing the above concerns through registration fees that scale with vehicle weight (Zipper, 2022).

2.2 Exposure Measures

Because many leading factors in non-motorized crashes and fatalities are related to vehicular characteristics, movements, and volumes, understanding how non-motorized and motorized road users interact with each other is essential. This interaction is commonly studied by analyzing pedestrian crash exposure, which is defined as a pedestrian/bicyclist's relative risk of experiencing a non-motorized crash (Pucher et al., 2011). The most common definition of risk is a measure of the probability of a crash to occur given exposure to potential crash events. There are a wide variety of exposure measures used in practice. Geographic scale is a critical element in most exposure analyses, and most analyses can be grouped into one of four scales: 1) regional (e.g., city, county, state); 2) network (e.g., traffic analysis zone, Census tract, Census block group); 3) road segment; and 4) point (e.g., mid-block or intersection street crossing).

Most transportation agencies collect consistent and reliable estimates of vehicular traffic volume, which may be used to define crash exposure measures. However, pedestrian exposure

data is rarely available. Surrogate measures, such as population, are typically used in place of pedestrian traffic volumes (Cottrill & Thakuriah, 2010; MRIGlobal et al., 2023; Wier et al., 2009). One approach to non-motorized road user exposure estimation, commonly referenced as the Keall method after its originator (Keall, 1995), uses travel diary information to impute pedestrian volume data. This method begins from origin-destination data for pedestrian-mode trips. These trips are expanded using survey weights and assigned to the transportation network using a shortest path algorithm. There are many limitations associated with this approach. First, such an approach misses pedestrian travel when accessing another mode (e.g., a person returning to their private vehicle from a store or accessing a bus stop from their home) (Li et al., 2020). Second, pedestrian trips are known to be underreported in travel surveys (Sammer et al., 2018). Finally, it is common for the safest pedestrian route to differ from the shortest path (e.g., a route without sidewalks is avoided even if it has the shortest travel time) (Li et al., 2020). In several cases (Dong et al. 2020; Lee and Abdel-Aty 2005; Sze, Su, and Bai 2019; Keall 1995), travel diary data are used to impute exposure and origin and destination data are assigned to the transport network using Dijkstra's shortest path algorithm (or a similar approach).

The "safety in numbers" hypothesis says that pedestrian (and bicycle) crash rates tend to decrease as a function of increasing volume. The assumption is that motorized travelers are more cautious when traveling in areas known to have high pedestrian volumes. For example, drivers may be more cautious in Lincoln's Haymarket or Omaha's Old Market because they anticipate pedestrian interaction. However, most tests of this hypothesis have relied on surrogate pedestrian exposure measures—a "space syntax" modeling approach that assumes pedestrian volume is a function of route directness and land use features (Geyer et al., 2006; Raford & Ragland, 2006), travel survey data (as detailed above) (Jacobsen et al. 2009), random number generation (Elvik,

2013), and machine learning (Mahmoud et al., 2021). The relationship between crash frequency and exposure is likely non-linear (Tao et al., 2021) and the relationship tends to decline as a function of exposure (Elvik & Goel, 2019).

Islam et al. (2022) developed a novel pedestrian exposure measured based on traffic signal logs. They combined pedestrian button events with 10,000 hours of video recorded in 2019 at 90 randomly selected signals throughout Utah to predict pedestrian volumes at intersections. They found a correlation of 0.84 between their observed and predicted hourly pedestrian crossing volumes, with a mean absolute error of 3.0 (Singleton et al., 2020). The traffic signal log requirement limits the applicability of such an approach to intersections with pedestrian actuation. They found a 4 percent increase in crash frequency for a 10 percent increase in pedestrian crossing volume.

2.2.1 Macro Crash Exposure Measures

Pedestrian and bicyclist crashes are (fortunately) rare events. Using area-wide analyses helps reduce zero observations, which can be challenging to accommodate in statistical models (Tao et al., 2021). Macro-level analysis is a simple method that researchers often use to examine crash exposure. This usually involves looking at crash counts by region (city, state, or country) or network zone (traffic analysis zone, Census tract, Census block group). The traffic volumes and pedestrian infrastructure within a given area or city are compared with other cities or areas of a city. This comparison is used to make inferences about the variables that are influential in creating pedestrian-vehicle crashes. Analyzing crashes at this level can be useful for policy makers when deciding to fund different projects. For example, policymakers could be looking to improve pedestrian infrastructure to lower motor vehicle-pedestrian crash frequency. If it can be shown that better infrastructure in a city or in an area lowers the frequency of crashes, the policy

makers can be persuaded to direct funds at these types of projects. This can be done by examining different regions or networks that vary in pedestrian infrastructure and trying to find relationships between these two variables. However, the drawback of this analysis is that it does not give much insight as to which specific types of infrastructure are the most successful in reducing crash frequencies.

2.2.2 Micro Crash Exposure Measures

Micro level analysis is most useful for researchers who are trying to determine which aspects of the transportation system most affect crash counts. When analyzing pedestrian-vehicle crashes at the micro level, two geographic scales are normally used: road segments and points (i.e., crash counts per intersection). Intersection-level is common due to the high number of pedestrian and vehicle conflict points at intersecting roadways. Furthermore, focusing on intersections allows the safety analyst to examine which aspects of infrastructure geometry or other intersection features are influential in increasing crash risk and severity. A drawback to only analyzing intersections is that it ignores midblock crashes.

Crash exposure data are generally used in two complementary applications. The first application is the construction of hot spot maps that visually indicate higher risk locations. Hot spot maps can be beneficial because they are readily understood by the public when justifying a safety improvement project. The second application is statistical modeling of person, geometry, and environment conditions for their effect on crash frequency and severity. The most common approaches are Poisson and negative binomial regression models. In recent years, several studies explored geographically weighted regression (GWR) applications (Mohammadnazar et al., 2021; Xu & Huang, 2015). GWR is a flexible method that incorporates spatial variation and has the

benefit of generating interpretable maps of the parameter space. However, GWR should be used with caution as it is principally a data fitting algorithm that lacks causal interpretation.

Tao et al. (2021) reported a pedestrian exposure elasticity of 0.66 at intersections and 0.47 at midblock locations. That is, a 10 percent increase in pedestrian volume produced a 6.6 percent and 4.7 percent crash frequency increase at intersection and midblock locations, respectively. These effects were only surpassed at intersection locations by number of intersection legs (-1.51), vehicular AADT (1.19), and average household size (-1.02). Li et al. (2020) found that the walking distance and number of crossed roads tended to be better metrics than total trips because they more closely measured potential pedestrian-vehicle interactions. Additional studies reviewed during the development of this project are summarized in Table 2.1.

Table 2.1 Summary of Crash Exposure Studies

Authors	Study Region	Metric	Geographic Scale	Study Period	Research Methods	Exposure Measures	
Gong and Abdel-Aty 2022)	Florida, USA	Micro	Intersection	October 2019	One-class SVM	Yes	Yes
Rahman et al. (2022)	Texas, USA	Micro	Road segment	2010-2019	Negative Binomial	No	Yes
Raihan et al. (2019)	Florida, USA	Micro	Road segment and intersection	2011-2014	Zero Inflated Negative Binomial	No	Yes
Sarkar et al. (2011)	Bangladesh	Micro	Road segment	1998-2006	Binary Logistic Regression	No	Yes
Wang et al. (2018)	Oregon, USA	Micro	Intersection	2009-2013	Logistic Regression	No	Yes
Almasi et al. (2021)	Tehran, Iran	Macro	Traffic Analysis Zone	2017-2019	Geographically Weighted Regression	No	Yes
Guo et al. (2018)	Vancouver, Canada	Macro	Traffic Analysis Zone	2009-2013	Full Bayesian Poisson Lognormal	No	Yes
Osama and Sayed (2017)	Vancouver, Canada	Macro	Traffic Analysis Zone	2009-2013	Area-based Poisson Lognormal	No	Yes
Pljakić et al. (2022)	Novi Sad, Serbia	Macro	Traffic Analysis Zone	2015-2019	Geographically Weighted Regression	No	Yes
Saha et al. (2018)	Florida, USA	Macro	Census block	2011-2014	Conditional Autoregressive	No	Yes
Wang and Kockelman (2013)	Austin, USA	Macro	Census tract	2007-2009	Multivariate Conditional Autoregressive	No	Yes

2.3 DOT Use of LBS

Given the focus on traffic volume in crash exposure measures, the emergence of mode-unspecified LBS data offers transportation engineers promising new avenues for safety analysis. However, it is important to understand the accuracy of trip details derived from LBS data, such as trip origin and destination locations, to assess its usefulness in transportation decision-making. To determine the reliability of the data source, researchers often compare the traffic volumes obtained from LBS data providers (termed by StL as a *traffic index*) to intersection or streetsegment traffic counts. One study analyzed StL traffic volumes at around 450 locations and compared them with traffic counts from Minnesota Department of Transportation sources. The researchers found that the mean absolute error of StL data at locations with an AADT of 10,000 or greater is roughly 8-10% (Turner et al., 2020). For low-volume roads (<1,000 AADT), the error was 42% and was often overestimated. Low-volume roads typically have higher error percentages in traffic counts because a single miscounted road user will represent a higher proportion of the total volume. An earlier report from the same Minnesota study evaluated StL accuracy using mean absolute error percentage (MAPE). They found the MAPE decreased as a function of increasing roadway AADT (Turner & Koeneman, 2017). A second potential cause of higher error on low volumes urban roads is that motor vehicle speeds are generally lower than on high urban volume roads. Slow moving vehicles represent a challenge for volume estimation using StL because it can be difficult to differentiate between slow moving vehicles and bicycles (Lee & Sener, 2020). This challenge also offers one explanation for StL's tendency to overestimate bicycle and pedestrian counts at these locations.

Given that bicycle trips comprise a low percentage of total trips, especially in U.S. cities, it can be difficult to estimate bicycle volumes using StL data alone. Li et al. (2020) compared

StL data with bicycle counts at 32 locations in Texas and found a positive correlation between the two measurements (R^2 values of 0.62 for weekday traffic and 0.69 for weekends). Weekend bicycle traffic was found to be greater than weekday traffic in this study, which supports a pattern of recreational over commuting cycling.

Research on pedestrian volumes has yielded similar results. Turner et al (2019) studied pedestrian volumes from Miovision at multiple intersections throughout the City of Austin, Texas. They found that the average daytime error between the StL index and measured pedestrian volumes was 15%, while nighttime conditions yielded an error of 24%. While left unstated, potential reasons for a nighttime drop in accuracy may have been due to lower pedestrian volumes at night. This pattern aligned with the findings of Li et al. (2020) that the StL index more accurately predicted weekend bicycle volumes, due to the higher volumes relative to weekdays. In summary, the accuracy of LBS-based non-motorized road user volumes is anticipated to be lower than that of motorized volumes at the same location. Transportation engineers and planners should consider these factors when using LBS data in non-motorized road user planning and safety analysis.

LBS data is best leveraged through fusion with other datasets. Sener et al. (2021) fused traditional traffic count, bikeshare, Strava, and StL data to model average annual daily bicycle traffic in Austin, Texas. This study found that the accuracy of the fused data depended on the fusion technique. The simplest fusion techniques took the form of generalized linear models. Machine learning algorithms show great promise to improve prediction accuracy. In one study, Kothuri et al. (2021) used random forest models to model a fused dataset of Strava, StL, and traditional traffic counts in Portland and Eugene, Oregon. The models had root mean squared errors (RMSE) of 20% and 14% for Portland and Eugene, respectively.

In addition to accuracy of the StL index relative to observed traffic, it is important to understand the measurement accuracy associated with the underlying data. LBS data are typically much more spatially precise (5-25 meters) than cellular tower data (100-200 meters), but less precise than in-vehicle navigation (3-5 meters) (StreetLight Data Inc., 2019). Given that LBS data can deviate by up to 25 meters from the true cellular device location, the exact counts on certain street networks can be uncertain. Accurate assignment to a road depends on the separation distance between parallel roads. The spatial precision of LBS data also decreases subject to multipath errors, which result from signals being deflected by surrounding objects, including trees and buildings (Mok et al., 2012). Finally, people travelling on subways are more difficult for LBS data providers to track because signals are often obstructed by ground cover. Fortunately, Lincoln does not contain subway lines.

It must also be noted that LBS data exhibit socio-demographic biases. Yang et al. (2020) reviewed several studies on the topic and found the StL penetration ratio can be biased by income and age distributions. This bias can be explained by lower cellular device ownership among low income and older age groups. These findings provide crucial insights for planners and engineers to consider when using these data. Related, traffic safety outcomes correlate negatively with income and age (Merlin et al., 2019; Toran pour et al., 2017). Given these populations are also more likely to use non-motorized travel modes, inaccuracies in the representativeness of LBS data should be corrected through careful calibration to observed multi-modal traffic counts.

2.4 Chapter Summary

The literature review found that total traffic volume and motorized vehicle speed are highly correlated with crash risk. Vehicle characteristics, such as weight, influence the severity

of a given crash. Traffic volume composition (i.e., the split between motorized and non-motorized modes) affects safety through its influence on conflict opportunities and roadway facilities. LBS data, such as that provided by StL, can be useful to expand the available spatial and temporal traffic volume coverage for a region. However, accuracy depends on the true traffic volume, with accuracy declining as a function of mode-specific volume. Data accuracy is also affected by the ability to infer mode and road segment, which are functions of relative travel speeds and transportation network density, respectively.

Chapter 3 Data Collection

While no primary data collection was conducted for this project, secondary data collection was a major task. This chapter summarizes two phases of data collection. In Phase 1 data collection, StL data were compiled from the StreetLight Insights platform. Supplemental data necessary for traffic volume calibration and safety analysis were also collected during Phase 1. Phase 2 data collection was informed by challenges encountered by the project team during Phase 1 (as detailed herein). In Phase 2, a Python-based approach was developed that directly queried the StreetLight Insights database through its application programming interface (API).

3.1 Phase 1 Data Collection

Three data sources were used in the Phase 1 traffic volume calibration procedure and an additional data source was necessary for the crash analysis. The first data source was City of Lincoln (CoL) traffic counts. Counts are available in a standard format, with turning and through movements at intersections tallied in 15-minute intervals. Most traffic counts span from 6 am to 6 pm, although a small number of intersections have counts taken from 12 am to 6 am, or 6 pm to 12 pm. Count data include bicycle, pedestrian, and vehicular traffic movements at 422 different intersections throughout CoL. In addition, traffic counts were taken on various days and in various years spanning from 2014 to 2022, although most counts occur in the years 2018, 2019, and 2021, as well as on Tuesdays, Wednesdays, and Thursdays. In total, there are 768 traffic counts available for calibration.

The next data source was from StreetLight (StL). StL takes cellular LBS data and imputes trip origin, destination, and travel mode based on cellular location ping locations, speed, acceleration, and other features. The data gathered from StL tracks transportation movements throughout specified areas or zones and is collected from several cellular service providers'

cloud environments. Once the data is gathered, it is de-identified by the suppliers and StL runs the data through numerous quality assurance tests that perform proprietary checks. After this processing, trips are constructed based on a set of assumptions. First, origin and destination locations are set based on the idle time for a cellular device. That is, if a device does not change its location for a specified amount of time (e.g., 5 minutes), the person is assumed to be at the end of a trip. Home and work locations are inferred from the frequency a device travels to a particular location, combined with land use information for that location. The most frequently visited residential location is coded as the home and the most frequently visited non-residential location is coded as the work or school location. Trip mode is inferred from travel speed. A low speed (calculated from the distance and time between successive device pings) denotes a walking trip, while a high speed denotes a motorized trip. Finally, StL normalizes the data monthly to account for certain parameter changes, such as changes in supplier sample size and population. This creates the StL Index. In this study, we used the OpenStreetMap road network layer provided by StL (i.e., zones are line segments that overlay the streets, sidewalks, and bicycle lanes within Lincoln city limits). This data is available for the day types and day parts defined in Table 3.1 below.

Table 3.1 StL Day Types and Day Parts Used to Extract Data

Day Types	Day Parts
All Days	All Day (12 am - 12 am)
Monday	Early AM (12 am - 6 am)
Tuesday	Peak AM (6 am - 10 am)
Wednesday	Mid-Day (10 am - 3 pm)
Thursday	Peak PM (3 pm - 7 pm)
Friday	Late PM (7 pm - 12 am)
Saturday	--------
Sunday	--------

The last data source used for the model calibration was the EPA Smart Location Database (EPA-SLD). This database includes roughly 90 variables related to land use, built environment, and sociodemographic characteristics of census block groups (CBGs) for the United States. The EPA-SLD variables include total population and employment, road network density measured as facility miles per square mile, and industry entropy as a measure of land use diversity. Version 3.0 of the Smart Location Database was used, which was released by the EPA in 2021 (Chapman et al., 2021).

The data collection process is straightforward for the CoL counts and EPA-SLD dataset; engineers at the CoL provided the count data and the EPA-SLD is accessible online. Generating StL data in the necessary format is a more difficult process. StL provides an interactive online platform that allows the user to select data by location, day type, day part, and travel mode. Users can run many different types of analyses, but the relevant analysis for this case is *Zone Activity.* Running a *Zone Activity* analysis allows the user to extract traffic attributes such as year, day of the week, and hour of the day, as well as trip attributes such as average traffic volume and average traffic speed imputed by StL based on raw LBS data inputs.

For this study, a comprehensive *Zone Activity* analysis was conducted covering the City of Lincoln for the year 2021. This specific year was chosen due to the data availability provided by StL, which includes monthly data for the years 2019, 2020, 2021, and a portion of 2022. Additionally, the StL platform limits the extraction of data to a maximum of 12 months in a single analysis. Therefore, 2021 was selected as the most suitable year for data collection. This decision was based on it being the most recent full calendar year and the diminished effect of the COVID-19 pandemic relative to 2020, which exhibited significant changes in travel patterns (Shamshiripour et al., 2020).

As part of the *Zone Activity* analysis, another critical part of the StL data extraction process is how data is gathered on smaller time scales. StL offers users the choice between *daily* and *week part* data, as well as *hourly* and *day part* data. In StL, a *week part* separates data between weekdays and weekends, while daily data provides traffic volumes for each individual day of the week. Similarly, "*hourly*" data captures volumes on an hourly basis, while "*day part*" data records them during specified time intervals, as detailed in Table 3.1. For this analysis, data collection focused on daily data within the *day parts* defined by StL. Daily and hourly data was too cumbersome to download, and *week part* data was too aggregated to be effectively compared to the CoL data, which was gathered on specific days.

It is essential to note that during the creation of the StL index provided to users for download, the StL software performs a *Zone Activity* analysis on zones in Lincoln, NE for each day of the week, across all StL day parts, for each month in 2021. This means that multiple different days are averaged together to generate a single StL index for a given day of the week. Consequently, if a user seeks traffic volume data at a highly specific date and time (e.g., 12:30 pm on Friday, May 7th, 2021), the information provided by StL may not be as precise as ground truth counts collected at that exact moment. StL data is designed to offer users a reliable estimate of travel patterns in general settings, while ground truth counts are better suited for capturing data in more specific and precise environments.

Extracting all the necessary zones presents a challenge because StL only allows the user to include 1,000 zones in each analysis. Using line segments as the zones, the CoL has close to 80,000 zones for each of the three travel modes (vehicles, bicycles, and pedestrians). This means that many different analyses must be run on the platform to collect all the necessary data. A tracking system was set up so that all the zones within the city limits could be extracted;

however, this system resulted in some zones being extracted in multiple analyses due to the extraction process being performed manually. Furthermore, the initial extraction process was done on the bicycle zones, as the zones that make up the bicycle layer in StL comprise the sum of vehicle and pedestrian zones. This is because bicycles are legally allowed to travel on more transportation facilities than both pedestrians and vehicles. Once *Zone Activity* analysis could be performed on all bicycle zones within Lincoln city limits, the extraction could be easily copied and repeated for the other two modes. Finally, all StL analysis data were downloaded onto a local computer.

3.2 Phase 2 Data Collection

Phase 1 data collection was performed to do data calibration. StreetLight (StL) set daily data points as Early AM, Peak AM, Mid-Day, etc. and among StL data from 2019 to 2022, we only selected the 2021 *Zone Activity* analysis from StL for calibration. The StL data was downloaded manually and the time frame and scale of StL data matched the traffic counts of the City of Lincoln (CoL). Phase 2 of data collection aimed to download all available data from January 2019 to April 2022 in more detailed time scales. Therefore, we systematically and efficiently downloaded StL data using Python environments to improve performance.

3.2.1 Data Access Using API

The StL data can be accessed using the StL Insights API. We utilized OpenStreetMap (OSM) segment zones to perform *Zone Activity* analysis and downloaded StL volume data for the period January 2019 to April 2022. The raw StL data is approximately 1.5 TB and comprises about 150,000 comma-separated value files. The data includes six different attribute files for each mode and the StL volume file. The attribute files contain information on equity, household,

education and income, trip purpose, employment, and trip information derived from Census data. These traveler attributes are calculated from the aggregate likely home locations of devices.

Table 3.2 summarizes the datasets collected from the StL Insights using the API approach. The “Geographical Scale” and “Temporal Scale” columns show the scale of an actual dataset, and the “Downloading Scale” column lists which scale those datasets are downloaded from. For example, if the geographical scale is OSM segment, the temporal scale is every day/hourly, and the downloading scale is CBG/monthly; this means the hourly OSM segment scale of traffic volume for every day from January 2019 to April 2022 was individually downloaded by the CBG and Monthly scales.

Table 3.2 Summary of StL Insights API Datasets Downloaded

Mode	Type	Data Periods	Geographical Scale	Temporal Scale	Downloading Scale
Bicycle	Zone Activity	Jan 2019 – Apr 2022	OSM Segment	Everyday / Hourly	CBG / Monthly
Pedestrian	Zone Activity	Jan 2019 – Apr 2022	OSM Segment	Everyday / Hourly	CBG / Monthly
Vehicle	Zone Activity	Jan 2019 – Apr 2022	OSM Segment	Everyday / Hourly	CBG / Monthly
Bicycle	Zone Activity	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Pedestrian	Zone Activity	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Vehicle	Zone Activity	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Vehicle	Zone Activity	Jan 2019 – Apr 2022	CBG	Weekly / Hourly	Lincoln City Limit / Weekly
Bicycle	Origin-Destination	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Pedestrian	Origin-Destination	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Vehicle	Origin-Destination	Jan 2019 – Apr 2022	CBG	Everyday / Hourly	Lincoln City Limit / Monthly
Vehicle	Origin-Destination	Jan 2019 – Apr 2022	CBG	Weekly / Hourly	Lincoln City Limit / Weekly

The StL Insights API was used to download this large dataset to improve efficiency and reduce download time. The StL Insights API simplifies data access and integration—allows Python packages and applications to communicate and download data directly from the StL server—and streamlines operations to reduce manual intervention, minimize errors, accelerate job execution, and improve automation. The StL developer hub also contributes to innovation

and development speed by allowing developers to build on existing scripts, which speeds up the development process and encourages the creation of new features and applications. Additionally, access to other Python packages can extend and improve data processing and analysis capabilities. Moreover, the StL Insights API includes some features unavailable in the web application. The API has a significantly larger zone limit per data analysis, reducing the data analysis time. Figure 3.1 shows each step for creating zone sets and running analysis. OSM segments were first divided into CBG, allowing the download of street segments for subscription zones by CBG sequentially from StL Insights. This implementation allows for future StL data extractions that capture changes in street configurations. The next step is creating zone sets based on census boundaries and OSM segments (as shown in Figure 3.1 on the right). After creating the OSM segment shapefile, the necessary StL Insights analysis can be run, and associated data can be downloaded to a local computer as CSV files.

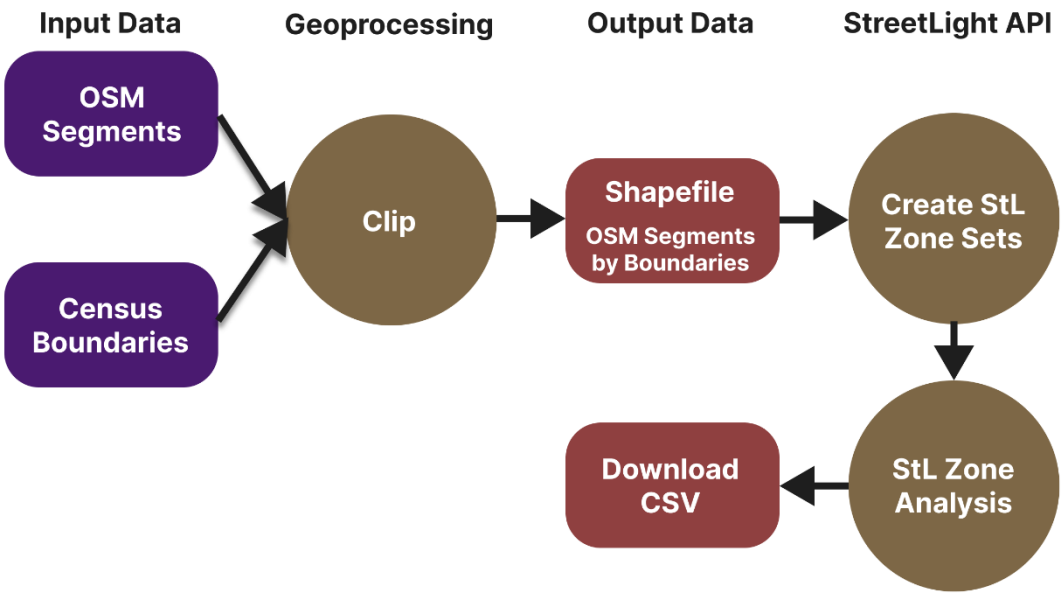

Figure 3.1 StL Insights API modeling framework

3.2.2 Data Extraction

This task involved joining and cleaning separate CSV files that contain segment volume data for each CBG in Lincoln. Managing segment-based volume was challenging due to the large file sizes and numerous rows in the CSV files. Parallel processing was used as a faster method and a solution to these challenges. As a result, all Lincoln segments were merged and cleaned into a single shapefile within 3-5 minutes (depending on computer capabilities). The entire procedure includes various data processing and aggregation processes. We first assigned folder paths to the CSV files and then efficiently parallelized the file-reading operation. The generated data frames were merged and are filterable based on specific criteria defined by the user. For our analysis, we aggregated the data to daily averages for each month by mode (vehicle, pedestrian, and bicycle), segment, and CBG geography.

3.2.3 Geodatabase

An ArcGIS Pro map package was prepared for spatial-temporal analysis. The package includes geodatabase and map visualizations. The geodatabase contains monthly bicycle, pedestrian, and vehicle volume layers at the OSM segment and CBG scales. Table 3.3 shows the list of files in the geodatabase.

Table 3.3 Files Included in Geodatabase for spatial-temporal analysis

File Name (File Type: Feature Class)	Description
bike_blkgrp_2019_agg	Monthly bicycle volume by CBG
bike_blkgrp_2020_agg	Monthly bicycle volume by CBG
bike_blkgrp_2021_agg	Monthly bicycle volume by CBG
bike_blkgrp_2022_agg	Monthly bicycle volume by CBG

Table 3.3 cont. Files Included in Geodatabase for spatial-temporal analysis

File Name (File Type: Feature Class)	Description
bike_seg_2019_agg	Monthly bicycle volume by OSM segment
bike_seg_2020_agg	Monthly bicycle volume by OSM segment
bike_seg_2021_agg	Monthly bicycle volume by OSM segment
bike_seg_2022_agg	Monthly bicycle volume by OSM segment
ped_blkgrp_2019_agg	Monthly pedestrian volume by CBG
ped_blkgrp_2020_agg	Monthly pedestrian volume by CBG
ped_blkgrp_2021_agg	Monthly pedestrian volume by CBG
ped_blkgrp_2022_agg	Monthly pedestrian volume by CBG
ped_seg_2019_agg	Monthly pedestrian volume by OSM segments
ped_seg_2020_agg	Monthly pedestrian volume by OSM segments
ped_seg_2021_agg	Monthly pedestrian volume by OSM segments
ped_seg_2022_agg	Monthly pedestrian volume by OSM segments
veh_blkgrp_2019_agg	Monthly vehicle volume by CBG
veh_blkgrp_2020_agg	Monthly vehicle volume by CBG
veh_blkgrp_2021_agg	Monthly vehicle volume by CBG
veh_blkgrp_2022_agg	Monthly vehicle volume by CBG
veh_seg_2019_agg	Monthly vehicle volume by OSM segments
veh_seg_2020_agg	Monthly vehicle volume by OSM segments
veh_seg_2021_agg	Monthly vehicle volume by OSM segments
veh_seg_2022_agg	Monthly vehicle volume by OSM segments

3.3 Chapter Summary

Three datasets were collected during Phase 1 data collection to perform traffic volume calibration and crash analysis. In this phase, CoL traffic counts, StL segment scale traffic volume, and EPA-SLD datasets were collected. CoL traffic counts and EPA_SLD datasets were straightforward to download. However, downloading the StL segment scale traffic volume was more complex because StL zone sets and different time variables needed to be selected from the StL interactive online platform to create a comprehensive Zone Activity analysis. Also, StL had a number of data limits when analyzing and downloading the data. During the second phase of data collection, we utilized the StL Insights API to access the StL data. Our analysis involved segment zones from OpenStreetMap (OSM) and the download of StL volume data from January 2019 to April 2022. Managing StL segment volume data was challenging due to the large file sizes and the countless rows in the multiple CSV files. Parallel processing was utilized to overcome these limitations. Additionally, spatial-temporal analysis was completed with an ArcGIS Pro map package, which includes geodatabase and map visualizations. The geodatabase contains monthly layers of bicycle, pedestrian, and vehicle volumes at the OSM segment and CBG scales.

Chapter 4 Traffic Data Calibration

This chapter focuses on the traffic data calibration process. The calibration required the use of City of Lincoln (CoL) intersection counts, as well as land use and roadway geometric features, to adjust the StL traffic indicators to better represent multi-modal traffic on the network.

4.1 Data Preparation

An overview of the calibration procedure is provided in Figure 4.1. The StL traffic indices and EPA-SLD land use features are calibration inputs that explain the variation in observed traffic volumes, as reported in the CoL intersection counts. Any number of continuous outcome models could be used for calibration, such as linear regression, random forest, or neural networks. The trained model can then be used to predict segment-level traffic volumes by travel mode. With the three core datasets in hand, the first step is to prepare these data for calibration by reconciling differences in their reporting units. The CoL traffic counts were collected for approach movements at intersections (points). The StL data were collected for facility segments (lines). The EPA-SLD data were collected for census block groups (areas). Subsequently, we adjusted the model using roadway features.

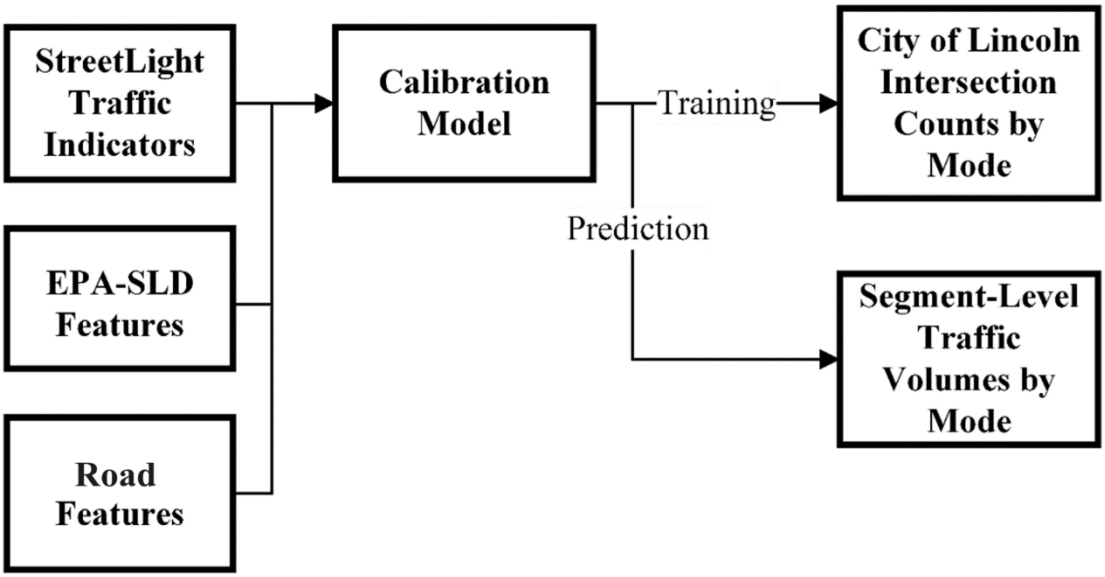

Figure 4.1 Overview of data calibration process

Beginning with the StL and CoL data, we associated StL segments with intersections containing available traffic counts. The summation of StL volumes at an intersection represents the total entering and exiting volumes (see Figure 4.2). That is, each approach volume was counted twice—once as an upstream volume and a second time as a downstream volume. The 15-minute CoL volumes were also aggregated across the 5 StL day parts. Then, those totals were multiplied by two to account for StL measuring traffic volumes in both directions, which results in double counting movements going upstream and downstream. To complete this step, the intersections where traffic counts recorded by CoL were geocoded and converted into a spatial data format for use in GIS software. Then, a buffer outlining all the StL segments that enter each intersection was drawn. Finally, the StL facility segments that enter the intersections were extracted using a spatial join function.

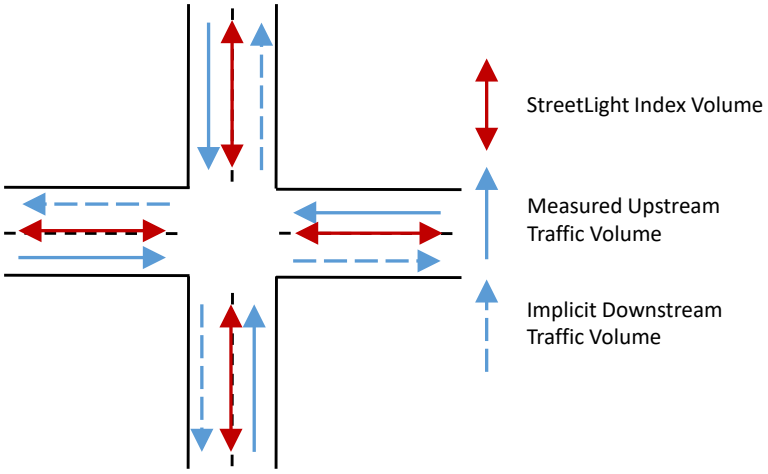

Figure 4.2 StreetLight index and traffic count fusion process

Once these facility segments were extracted, the CoL data and the StL data needed to be combined into one table for further use. Each CoL traffic count record was joined with its corresponding StL record by intersection location, day of week, and day part. The result is a table with only CoL traffic counts and StL records with locations corresponding to days of the week.

With a fused dataset, the final step of the calibration data preparation was to add the EPA-SLD and roadway features data. A spatial join was defined between the EPA-SLD census block group (CBG) polygons and StL line segments. This join resulted in multiple CBGs being associated with StL segments entering a CoL intersection of interest. There are two possible methods to reconcile this many-to-one join. The first method is to use average statistics across the joined CBG. This approach has the benefit of creating a one-to-one join but reduces the feature variation and sample size for calibration model training. The second method (and the one used herein) is to maintain the many-to-one join and assume that predicted dependent variable values (i.e., traffic volumes) are a function of the combination of land use features in the adjacent CBGs. The second method was chosen for this analysis due to the drawbacks of Method 1 (i.e.,

reduction of sample size and feature variation). Method 2, averaging predictions across results assuming multiple CBG feature combinations, captures the influence of feature variation on traffic counts.

As for the roadway attributes, a common ID or defining attribute could not be found between the bicycle, pedestrian, and vehicle networks and the roadway attribute shapefiles. Therefore, the following approach was used to spatially join the roadway attributes, namely median type, shoulder type, shoulder width, presence of shoulder rumble strips, number of lanes, speed, and one-way/two-way configuration.

1. Create a 70 ft buffer around the NDOT lines (buffer 1).
2. Create a 10 ft buffer around the StL line segments (buffer 2).
3. Intersect the 10 ft buffer layer with itself to split minor road parts where they intersect with major roads (buffer 2 intersect).
4. Intersect the two layers (buffer 1 and buffer 2 intersect) and add the street attribute(s) of interest to the buffered StL line segments (buffer 3).
5. Remove the polygons in buffer 3 with an area less than 300 sqft. This way, the perpendicular segments where minor roads intersect major roads are removed (filtered buffer 3).
6. Intersect the original StL line segments with filtered buffer 3 (StL line intersect).
7. Remove any line segments with lengths less than 6 ft. Since the buffer in filtered buffer 3 is 5 ft on each side, this accounts for when minor roads intersect major roads at non-perpendicular angles (filtered StL line intersect).
8. Add street attribute(s) from the buffer to filtered StL line intersect by location.
9. Add these attributes to the original vehicle network based on ObjectID.

4.2 Calibration Models

A random forest model was chosen to analyze the features considered for each travel mode. Random forest is an ensemble method that combines decision trees, which use dataset-splitting rules to partition the feature space based on decision metrics, such as information gain (Ding et al., 2016). Random forest uses bagging and feature randomization to combine decision trees and minimize overfitting bias. They have some advantages over classical linear regression. The first advantage random forest models offer is that they excel at catching non-linear relationships between variables, which can help accommodate the complexities in transportation data. Also, the "Feature Importance" analysis provided by random forest models allows users to identify the variables most influential to the model. Additionally, random forest models are not as sensitive to outliers and the inclusion of irrelevant features as other models, which increases model robustness. Finally, a random forest model's ability to account for interactions between variables is an important asset for creating models with complex datasets.

Because numerous features influence traffic patterns, random forest models are an ideal choice for this analysis. Liu and Wu (2017) used a random forest model to predict traffic congestion on the Shanghai road network. They used attributes such as time of day, weather, and road quality gathered from the Shanghai Traffic Management Department and found the prediction accuracy of the model was approximately 88%. The advantages of random forest models are further demonstrated in research performed by Cheng et al. (2019). They applied a random forest model to travel mode choice behavior and found that random forest outperformed the three other models (Adaboost, Support Vector Machine (SVM), and multinomial logit) based on mean absolute error percentage.

4.3 Calibration Model Results

4.3.1 Linear Regression Results

As a starting point, linear regression models were chosen to explore how well the StL indices predict bicycle, pedestrian, and vehicle traffic volumes. In each linear regression model, EPA-SLD variables were included to improve model fit. The EPA-SLD database has roughly 90 different variables including total population and road network density for each census block group (CBG). Variables were added or removed from the models based on the AIC and using a systematic process that tested all variable combinations. Results were benchmarked against results from past research (H. Li et al., 2020; Turner et al., 2020); a comprehensive review of the literature suggested that a model with an R-squared score equal to or exceeding 0.75 indicates acceptable accuracy.

Based on this benchmark, only the vehicle model was anticipated to attain the previously defined acceptable level of accuracy. This anticipated result was bolstered by initial correlative analyses finding that StL vehicle volumes correlated highly ($R^2 \approx 0.90$) with aggregated CoL counts. However, StL pedestrian volumes have a weaker correlation ($R^2 \approx 0.60$) than vehicle volumes and StL bicycle volumes performed even worse in this setting ($R^2 \approx 0.25$).Linear regression model results are provided in Table 4.1, followed by a discussion of relevant model results.

Table 4.1 Linear Regression Model Results for Linear Regression Models by Mode

Mode	Adj-R-squared	No. of Observations
Vehicle	0.91	2263
Pedestrian	0.78	3348
Bicycle	0.37	3559

The vehicle model included four EPA-SLD variables: *Total population of each CBG in 2018*, *Number of occupied housing units in 2018*, *Percentage of two-plus-car households in 2018*, and *Total employment in 2017*. The adjusted R^2 value for this model was higher than the predetermined benchmark of 0.75. Furthermore, the adjusted R^2 value of 0.91 represents the highest of all three models, which aligns with prior expectations. The number of occupied housing units correlated negatively with CoL counts. The pedestrian model included four EPA-SLD variables: *Total population in 2018*, *Number of occupied housing units in 2018*, *Percentage of zero-car households in each CBG in 2018*, and *Total employment in 2017*. The bicycle model included fourteen additional EPA-SLD variables, some of which are *Total road network density in the CBG*, *Percentage of two-plus-car households in 2018*, *Percentage of one-car households in 2018*, and *Percentage of high wage workers*. Additional EPA-SLD variables were included to this model due to its lower accuracy. The StL index was identified as strongly predictive of traffic volume across all modes.

4.3.2 Random Forest Model Results

Random forest models were trained to improve prediction accuracy. In each random forest model, variables identified in the linear regression models were retained. All the random forest models had R^2 scores greater than or equal to 0.97. The bicycle and pedestrian models

performed the best with an R^2 score of 0.997. The vehicle model's R^2 score was slightly lower than the two other models at 0.971.

In the bicycle model, the most important variable for influencing prediction accuracy was the *Average Daily Zone Traffic (StL Volume)*. Each feature's importance index to the model can be found in Table 4.2 below. The importance index provided for the StL volume variable was lower for the bicycle model than for the pedestrian and vehicle models. This could be explained by the bicycle dataset having lower volumes, so the additional variables added to the model provided more prediction accuracy than in the two other models. Additionally, the EPA-SLD variables available for the pedestrian and vehicle models could have been insufficient. Finally, there may have been other variables correlated with traffic volume variations, such as weather conditions.

Table 4.2 Feature Importance for Bicycle Model

Feature	**Importance**
Average Daily Zone Traffic (StL Volume)	0.379
Number of lanes	0.253
Total road network density	0.165
Percent of high wage workers (workplace)	0.078
Percent medium wage workers (workplace)	0.034
Gross residential density (HU/acre)	0.024
One way	0.018
National functional classification	0.009
Gross_industrial (5-tier) employment density (jobs/acre)	0.008
Median type	0.007
8-tier employment entropy	0.005
Employment and household entropy	0.004
Gross industrial (8-tier) employment_density (jobs/acre)	0.003
Percent of one-car households	0.003
Housing units	0.003
Percent of two-plus-car households	0.003
Total land area (acres)	0.002
Total land area (acres) that is not protected from development	0.001
Shoulder	0.001

Like the bicycle model, the *Average Daily Zone Traffic (StL Volume)* variable was the most influential variable for the pedestrian model, as shown in Table 4.3 below. This variable impacted this model more than the bicycle model. This can be explained by the pedestrian dataset having higher volumes per segment.

Table 4.3 Feature Importance for Pedestrian Model

Feature	**Importance**
Average Daily Zone Traffic (StL Volume)	0.660
Number of lanes	0.128
Percent of zero-car households	0.078
Total employment	0.034
Population	0.032
Median type	0.028
National functional classification	0.021
One way	0.016
Households (occupied housing units)	0.004

The feature importance for each variable in the vehicle model can be located in Table 4.4 below. As with the two prior models, the StL volume variable provided the most statistically significant contribution toward the accuracy of this model. An interesting observation from this model and the two previous models is the placement of the *Number of lanes* variable consistently ranking the second most important. The variables included in the vehicle model were more

relevant to vehicle travel than in the previous model. However, none of these variables has an importance value that exceeds 0.05.

Table 4.4 Feature Importance for Vehicle Model

Feature	**Importance**
Average Daily Zone Traffic (StL Volume)	0.785
Number of lanes	0.048
Speed limit	0.041
Jobs within 45 minutes auto travel time (time-decay weighted)	0.031
Percent of high-wage workers (workplace)	0.020
Total road network density (facility miles per square mile)	0.018
Percent of two-plus-car households	0.015
Intersection density (auto-oriented intersections per square mile)	0.014
One way	0.010
Shoulder	0.008
Median type	0.006
National functional classification	0.005

4.3.3 ANOVA Test Results

Note that CoL traffic count locations are not a random sample. Traffic counts are usually gathered on major roadways and in commercial areas, with few counts taken on low-volume residential streets. ANOVA tests were performed for all model features to quantify the variation between the training and prediction datasets and understand count location sampling bias. The

results of the ANOVA tests of each variable of the three different models can be found in Table 4.5 for vehicle volumes, Table 4.6 for pedestrian volumes, and Table 4.7 for bicycle volumes.

The ANOVA test used a one-way blocking on whether a facility segment was included in the training dataset. In nearly all cases, there is a statistically significant difference in the feature variance between the training and prediction datasets (at a 0.05 significance level). These results indicate that the training dataset is not drawn from a random sample of the facility segment population. The high R^2 scores for each model may be partially explained by this intersection selection bias. As such, prediction accuracy is unlikely to be as high as that reported in the calibration results. However, this finding tempers the interpretation of prediction accuracy rather than invalidating it. It is commonly the case in traffic prediction that model validation is performed on high-traffic volume facilities for which the analyst has traffic counts available. Additionally, this finding aligns with the findings from Turner et al. (2020), where the prediction accuracy of StL data decreased significantly on low-volume roads. Another factor to consider when interpreting these results is the final research objective. Predicted traffic volumes are used in areal crash exposure metrics, which aggregate volumes across multiple facilities.

Table 4.5 ANOVA Results for StL Vehicle Zones Entering Intersections vs. Zones Not Entering Intersections

Feature	**F-statistic**
Jobs within 45 minutes auto travel time (time-decay weighted)	143.67*
Percent of two-plus-car households in CBG	297.19*
Total road network density (facility miles per square mile)	248.12*
Percent of high-wage workers (workplace)	5.42
Intersection density (auto-oriented intersections per square mile)	654.08*

* F-statistic statistically significant at 0.05 level

Table 4.6 ANOVA Results for StL Pedestrian Zones Entering Intersections vs. Zones Not Entering Intersections

Feature	F-statistic
Total employment	1314.93*
Total population	134.88*
Total households	81.34*
Percent of zero-car households	908.16*

* F-statistic statistically significant at 0.05 level

Table 4.7 ANOVA Results for StL Bicycle Zones Entering Intersections vs. Zones Not Entering Intersections

Feature	F-statistic
Percent of high-wage workers (workplace)	11.26*
Total land area (acres) that is not protected from development (i.e., not a park, natural area, or conservation area)	147.61*
Total road network density (facility miles per square mile)	332.64*
Percent of one-car households	347.51*
Total land area (acres)	147.89*
Total housing units	58.36*
Employment and household entropy	40.44*
Gross industrial (5-tier) employment density (jobs/acre) on unprotected land	38.09*
Percent of two-plus-car households	534.51*
Gross residential density (HU/acre) on unprotected land	212.64*
Percent medium wage workers (workplace)	0.52
8-tier employment entropy	123.86*
Gross industrial (8-tier) employment density (jobs/acre)	38.09*

* F-statistic statistically significant at 0.05 level

4.3.4 Overfitting Analysis

Another important aspect of verifying the accuracy of the random forest model is checking for overfitting. Overfitting is a common issue in statistical modeling where a model becomes overly complex, capturing random fluctuations in the training data rather than the underlying patterns or relationships. In an overfit model, the algorithm fits the training data so

closely that it fails to generalize well to unseen or new data. This results in poor performance when the model encounters other datasets that differ from the training data, essentially memorizing the training data rather than learning from it. Overfit models typically exhibit high accuracy on the training data but poor accuracy on validation or test datasets. Table 4.8 summarizes the results for the traffic calibration models. The difference between the testing and training datasets are small for all modes, indicating the three models are not overfitting the data and thus represent a reasonable model fit.

Table 4.8 Testing and Training R^2 values for each Random Forest Model

Bicycle Model	
R^2 for Training Data	0.997
R^2 for Testing Data	0.981
Pedestrian Model	
R^2 for Training Data	0.992
R^2 for Testing Data	0.983
Vehicle Model	
R^2 for Training Data	0.992
R^2 for Testing Data	0.954

4.4 Chapter Summary

Chapter 4 described the traffic volume calibration developed in this research project. StL indices require calibration against traffic volumes to adjust their magnitude and temporal distribution. A large historical dataset comprising multi-modal traffic counts at major intersections was used in the calibration process. StL indices were combined with land use and roadway features to train random forest models on the historical traffic counts. High predictive accuracies were obtained for all models and statistical tests did not suggest model overfit. ANOVA tests were performed to test for sampling bias in the locations chosen for traffic counts.

While bias was identified in this sampling, it is believed that out-of-sample prediction was primarily used for low volume roads for which high prediction error is anticipated as a priori.

Chapter 5 Safety Analysis

This chapter includes a comprehensive analysis of traffic crashes, using a multifaceted approach to study the factors influencing road safety. The chapter begins with an explanation of data sources, which sets the foundation for subsequent analytical procedures. The crash analysis procedure is comprised of several components: crash analysis, crash rate analysis, spatial lag model analysis, and relative rank analysis. The results section presents findings in detail, and each subsection is accompanied by a dedicated discussion that delves into the nuances of the results, fostering a holistic understanding of the factors contributing to crashes involving non-motorized road users.

5.1 Data Sources

Many of the data sources in this chapter aligned with Chapter 3. Data included vehicle, bicycle, and pedestrian volumes on each transportation network segment, as determined by the calibration models described in Chapter 4, which facilitated the calculation of crash exposure measures. EPA-SLD land use variables by census block group (CBG) and roadway features were used in the analysis. Finally, a motorized-to-non-motorized crash dataset was provided by the Nebraska Department of Transportation. The dataset included 2,171 crashes by date and time, weather and road surface conditions, total fatalities, and crash severity for the period 2011 to 2020 within the City of Lincoln (CoL) corporate limits.

5.2 Analysis Procedures

5.2.1 Crash Analysis Procedure

Crash analysis was performed using a combination of graphical and quantitative analyses. Graphical analysis used maps in ArcGIS and quantitative analysis used Python scripts. Graphical analysis assigned crashes to the nearest intersection and sized each intersection point

proportional to the crash count at that location over the analysis period. This point layer was combined with choropleth maps for land use and demographic variables of interest (e.g., crash count vs. poverty rate). Census block groups (CBGs) were used as the geographic unit of analysis because they were the smallest geographic unit available for most demographic variables and they aligned with the EPA-SLD. The variables used in generating the crash rate maps in this study are summarized below and represent a mix of crash exposure and demographic variables that provided diversity in the preliminary crash evaluation.

- **Crashes:** the number of crashes involving pedestrians and/or bicycles. A total of 2,171 incidents were included in the analysis among which 1,653 (76%) were labeled as "Intersection Involved". These crashes were then assigned to the nearest intersection.
- **Vehicle miles:** the predicted vehicle traffic activity on vehicle segments, which were generated from the random forest model. Vehicle volumes on each segment were multiplied by the corresponding segment length to calculate vehicle travel miles. Then, all the vehicle miles within a given CBG were aggregated to obtain the total vehicle miles per CBG.
- **Pedestrian miles:** the predicted pedestrian traffic activity on pedestrian segments, which were generated from the random forest models. Pedestrian volumes on each segment were multiplied by the corresponding segment length to calculate pedestrian travel miles. Then, all the pedestrian miles within a given CBG were aggregated to obtain the total pedestrian miles per CBG. Pedestrian segments located beyond 35 feet from the road center line were excluded from the aggregation due to the lack of pedestrian-vehicle interaction on such segments.

- **Bicycle miles:** the predicted bicycle traffic activity on pedestrian segments, which were generated from the random forest models. Bicycle volumes on each segment were multiplied by the corresponding segment length to calculate bicycle travel miles. Then, all the bicycle miles within a given CBG were aggregated to obtain the total bicycle miles per CBG. Bicycle segments located beyond 35 feet from the road center line were excluded from the aggregation due to the lack of bicycle-vehicle interaction on such segments.
- **Active miles:** the active miles were calculated by a simple addition of pedestrian and bicycle miles per CBG.
- **Poverty:** the poverty rate within the CBG was generated using the EPA-SLD variables.
- **Non-white population:** the non-white population was calculated by the ratio of the non-white population to the total population within the CBG, as given in the EPA-SLD.

5.2.2 Crash Rate Analysis Procedure

Like the crash analysis, the crash rate analysis used both visual and quantitative analyses, with visual analysis done using maps in ArcGIS and quantitative analysis done with Python scripts. Crash rate analysis was performed at the macro level, as performing micro-level analysis was difficult due to the small number of crashes recorded on many StL links (addressed in subsequent analysis detailed in Section 5.2.5 below). Three crash rate variables were defined in the analysis, which capture different aspects of crash risk. The first variable, *multiplicative crash rate* measure, accounts for the total traffic activity within each CBG. The *multiplicative crash rate* measure multiplies the active and vehicle miles but does not capture the difference in relative proportions between active and vehicle miles across CBGs. For example, two CBGs (CBG1 and CBG2) could have the same multiplicative exposure (i.e., active miles × vehicle

miles), but CBG1 could have twice the number of active miles as CBG2. In this example, if each CBG contains the same number of crashes, their *multiplicative crash rate* measures would be the same.

$$\text{Multiplicative Crash Rate} = \frac{\text{No. of Crashes}}{\text{Active miles} \times \text{Vehicle miles}} \tag{5.1}$$

The *ratio crash rate* measure was defined as an alternative to the *multiplicative crash rate* measure. Using the example from above, if two CBGs have the same number of total miles but CBG1 has twice the number of active miles as CBG2, the ratio of active-to-vehicle-miles would be larger for CBG2. Furthermore, if these CBGs had the same crash total, they would have different *ratio crash rate* values. For CBGs with the same ratio, there is also the possibility for significantly different total volumes for the same crash count. For example, CBG1 with one active mile and 100 vehicle miles would have the same *ratio crash rate* as CBG2 with 1,000 active miles and 100,000 vehicle miles if each CBG contains the same number of crashes.

$$\text{Ratio Crash Rate} = \frac{\text{No. of Crashes}}{\frac{\text{Active miles}}{\text{Vehicle miles}}} \tag{5.2}$$

The multiplicative and ratio rates often show conflicting raking results. The active miles crash rate measure was created to focus on the non-motorist crash exposure per mile of non-motorist travel. If two CBGs have the same number of active miles and crashes but different vehicle miles, the active miles crash rate values will be the same.

$$\text{Active Miles Crash Rate} = \frac{\text{No. of Crashes}}{\text{Active Miles}} \tag{5.3}$$

5.2.3 Spatial Lag Model Procedure

Spatial lag models were estimated, setting the crash rates as the dependent variables, to understand the variations of each crash rate due to demographic and land use drivers. The spatial lag model specification controls for dependency in crash rate arising from unobserved but spatially related variables. The model is defined by the following equation.

$$y = \beta X + \rho W y + \varepsilon \tag{5.4}$$

where βX represents an ordinary least squares regression model structure, W is a term representing a global spatial weight, ρ is a spatial correlation parameter, and ε is a normally distributed error term. The model was estimated by maximum likelihood, so uses a pseudo-R^2 value, which does not represent a proportionate reduction in error. Variables were retained in the model with p-values less than 0.20 to avoid excluding variables with policy relevance that may have weak correlation due to low statistical power.

5.2.4 Relative Rank Analysis Procedure

Making comparisons between the two crash rates' variables, as well as crash totals, requires a normalization to put each variable on the same scale. The relative rank measure is defined in the following equation.

$$\text{Relative Rank} = \frac{\text{Rate Measure}_{x} - \text{Rate Measure}_{\text{lowest}}}{\text{Rate Measure}_{\text{highest}} - \text{Rate Measure}_{\text{lowest}}} \tag{5.5}$$

Where Rate Measure$_x$ represents the rate measure for each CBG. This process assigns a value between zero and one for each crash rate or total measure. The CBG with the highest rate measure will return a value of 1, while a CBG with a measure of half the maximum magnitude will have a value of 0.5. However, it is worth noting that having a relative rank of 0.5 does not mean the CBG is at the "50th-percentile", as the value of rate measures are not evenly distributed across CBGs.

5.2.5 Safety Assessment and Ranking of Links and Intersections

A final set of analyses provides link and intersection-level results. The Composite Safety Scores (CSS) used in this section is based on an approach developed by Iowa DOT (2020). It ranks elements (i.e., links or intersections) in terms of crash rate based on their attribute values relative to those for all elements in the Lincoln road network. Each element is assigned a set of values (numeric or categorical) corresponding to the attributes used in the assessment. Through this ranking method, elements in the same group (i.e., possess the same values for all ranking attributes) are expected to have the same composite safety score. To ensure consistency, the score of every attribute is normalized on a scale from 1 to 10. Then, each attribute-based normalized score is adjusted against the attribute weight before the composite score of the group (or attribute set) is determined as the sum of the weighted normalized scores.

5.2.5.1 Safety Assessment and Ranking of Links

The following describes the road-link ranking method in detail.

1. Crashes are assigned to the closest link. The number of crashes per link is then calculated.
2. Mileage is calculated for each link.
3. The attributes/safety factors are selected for this analysis.

4. The bins (categories) are determined for each attribute.
5. Links are aggregated into bins by attribute and total crashes per mile is calculated for each bin.
6. The weight of each attribute/factor (W) is determined such that the total weight is 100. In our analysis, the weights for the volume and speed limit attributes were assumed twice the individual weights of the remaining attributes, based on literature suggesting these attributes are particularly relevant to non-motorized road user safety. That is, considering a total of eight attributes, both volume and speed limit were assigned a weight of 20, whereas the remaining attributes were assigned a weight of 10 each.
7. The Weighting Factor (WF) for each attribute is determined. Since a 10-tier uniform Normalized Score (NS) scale is adopted in this method, the weight of each attribute is divided by 10. The weighting factor converts the attribute-based normalized scores to produce the composite score for the link.

$$\mathrm{WF} = \frac{100}{\mathrm{W}} \tag{5.6}$$

8. The crashes and miles are aggregated to each bin within each attribute. This normally yields the same total miles and total crashes across all attributes (and the study area).
9. The Crash Rate (CR) is calculated for each bin as the ratio of crashes to mileage for the bin. This will be the basis for defining the normalized score intervals, which will determine the attribute-based Normalized Scores (NS).
10. The bounds of the NS intervals for each attribute are determined. For a particular attribute, it is calculated as follows.

$$\text{NS Interval Width (NSIW)} = \frac{\text{Max bin CR} - \text{Min bin CR}}{10} \tag{5.7}$$

$$\text{NS}_\text{n}\ \text{Interval} = [\text{Min bin CR}, \text{Min bin CR} + (\text{n} - 1)\text{NSIW}) \quad ; \text{n} = 1, 2, \dots, 10 \tag{5.8}$$

Note that the intervals used in this method are left-closed.

11. The Normalized Score (NS) is calculated for each bin as the order of the attribute NS interval within which the bin-specific Crash Rate (CR) falls. That is, a Speed_limit bin for which the CR value falls within the 3rd interval [Min bin rate, Min bin CR + (3 – 1) NSIW) is given a NS of 3 for the Speed_limit attribute.
12. The Composite Normalized Score (CNS) of the link (or attribute set) is calculated as the sum of the attribute NS values each multiplied by its WF as follows.

$$CNS = \sum_{n=1}^{10} WF_n \times NS_n \tag{5.9}$$

The table below presents the attributes, weights, weighting factors, and bins used in the links assessment and ranking.

Table 5.1 Selected Attributes, Bins and Weights for Link Safety Ranking

Attribute (Risk Factor)	Attribute Weight	Weighting Factor	Bins
Volume	20	2	0 - 250 vph
			250 - 500 vph
			500 - 1000 vph
			1000 - 2000 vph
			2000 vph or more
Shoulder width	10	1	0 ft (no shoulder)
			1 - 3 ft
			3 - 6 ft
			6 ft or greater
Shoulder presence	10	1	No shoulder
			Shoulder
Median type	10	1	None
			Painted
			Open
			Barrier
			Raised
Facility type	10	1	One-way
			Two-way
			Ramp
Number of lanes	10	1	1 lanes
			2 lanes
			3 lanes
			4 lanes or more
Speed limit	20	2	20 - 30 mph
			30 - 40 mph
			40 - 50 mph
			50 mph or greater
Road classification	10	1	Remote residential
			Local
			Collector
			Minor arterial
			Major arterial and above

5.2.5.2 Safety Assessment and Ranking of Intersections

The procedure used in the assessment and ranking of intersections differs slightly from that used for links. The differences include the denominator of the crash rates formula, as well as the treatment of the street attributes. Since multiple links converge at an intersection, a parameter (mean, minimum, maximum, etc.) that represents the attribute (e.g. Number of lanes) for the whole intersection needs to be selected for each attribute. The following describes the ranking method of intersections in detail.

1. Crashes are filtered such that only intersection-related crashes (ITR_IVL_S = 'Y') are included in the analysis. The relevant crashes are then assigned to the closest intersection. The number of crashes per intersection is then calculated.
2. The total predicted pedestrian and bicycle (non-motorized) volumes is determined for each intersection.
3. The attributes/safety factors are selected for this analysis.
4. The bins (categories) are determined for each attribute.
5. Intersections are aggregated into bins by attribute and total crashes per non-motorized traffic volume is calculated for each bin.
6. The weight of each attribute/factor (W) is determined such that the total weight is 100. The individual weights of AADT and maximum speed limit attributes were assumed to be twice the individual weights of the remaining attributes based on prior research findings. That is, for a total of five attributes, both AADT and max speed were assigned a weight of 28.6, whereas the remaining attributes were given a weight of 14.3 each.

7. The Weighting Factor (WF) for each attribute is determined. For the same 10-tier uniform Normalized Score (NS) scale adopted in the link analysis, the weight of each attribute is divided by 10.

$$\mathrm{WF} = \frac{100}{\mathrm{W}} \tag{5.10}$$

8. The assigned crashes and non-motorized traffic volumes are aggregated to each bin within each attribute.
9. The Crash Rate (CR) is calculated for each bin as the ratio of crashes to pedestrian and bicycle AADT for the bin.
10. The bounds of the NS intervals for each attribute are determined. For a particular attribute, the NS interval is determined as follows.

$$\text{NS Interval Width (NSIW)} = \frac{\text{Max bin CR} - \text{Min bin CR}}{10} \tag{5.11}$$

$$\mathrm{NS_n}\ \text{Interval} = [\text{Min bin CR}, \text{Min bin CR} + (\mathrm{n}-1)\text{NSIW}) \quad ; \mathrm{n} = 1, 2, \dots, 10 \tag{5.12}$$

11. The Normalized Score (NS) is calculated for each bin as the order of the attribute NS interval within which the bin-specific Crash Rate (CR) falls. That is, if the highest road class bin for which a CR value falls is within the 7th interval [Min bin rate, Min bin CR + (7 – 1) NSIW] is given an NS of 7.
12. The Composite Normalized Score (CNS) of the link (or attribute set) is calculated as the sum of the attribute NS values multiplied by its WF as follows:

$$\mathrm{CNS} = \sum_{n=1}^{5} \mathrm{WF_n} \times \mathrm{NS_n} \tag{5.13}$$

The table below presents the attributes, weights, weighting factors, and bins used in the intersections assessment and ranking.

Table 5.2 Selected Attributes, Bins and Weights for Intersection Safety Ranking

Attribute (Risk Factor)	**Attribute Weight**	**Weighting Factor**	**Bins**
Vehicle AADT	28.6	2.86	0 - 250 vph
			250 - 500 vph
			500 - 1000 vph
			1000 - 2000 vph
			2000 vph or more
Max speed limit	28.6	2.86	20 - 30 mph
			30 - 40 mph
			40 - 50 mph
			50 mph or greater
Highest road class	14.3	1.43	Remote residential
			Local
			Collector
			Minor arterial
			Major arterial and above
Median type	14.3	1.43	None
			Painted
			Open
			Barrier
			Raised
Average number of lanes	14.3	1.43	2 lanes or fewer
			2 - 3 lanes
			3 - 4 lanes
			4 lanes or more

5.3 Results

5.3.1 Crash Analysis Results

Figure 5.1 through Figure 5.3 show the results of the spatial analysis performed through the creation of maps in ArcGIS. Each of these three maps shows the number of crashes that occurred at various intersections overlaid with the quantity of a variable of interest by CBG: *vehicle miles* in Figure 5.1, *active miles* in Figure 5.2, and *percentage of non-white citizens* and *poverty rate* Figure 5.3a and Figure 5.3b, respectively.

As demonstrated by Figure 5.1, there is a moderately high spatial correlation between motorized-to-non-motorized crashes and vehicles miles by CBG. Vehicle miles are very high toward the central business district (CBD) of Lincoln because there is a high concentration of commercial and residential buildings in this area. The CBD also contains the most crashes of any CBG in Lincoln, totaling 239 motorized-to-non-motorized crashes during the data collection period of 2011 to 2020. A notable aspect of this analysis outside this observation is the high number of crashes taking place on 27th Street. This street is located roughly one mile east of the Lincoln CBD and, as Figure 5.1 shows, the number of crashes in this area is moderately correlated with vehicle miles. This result is supported by the quantitative analysis shown in Table 5.3; the *vehicle miles* variable has a large coefficient of 0.58 and is statistically significant at the 0.05 level.

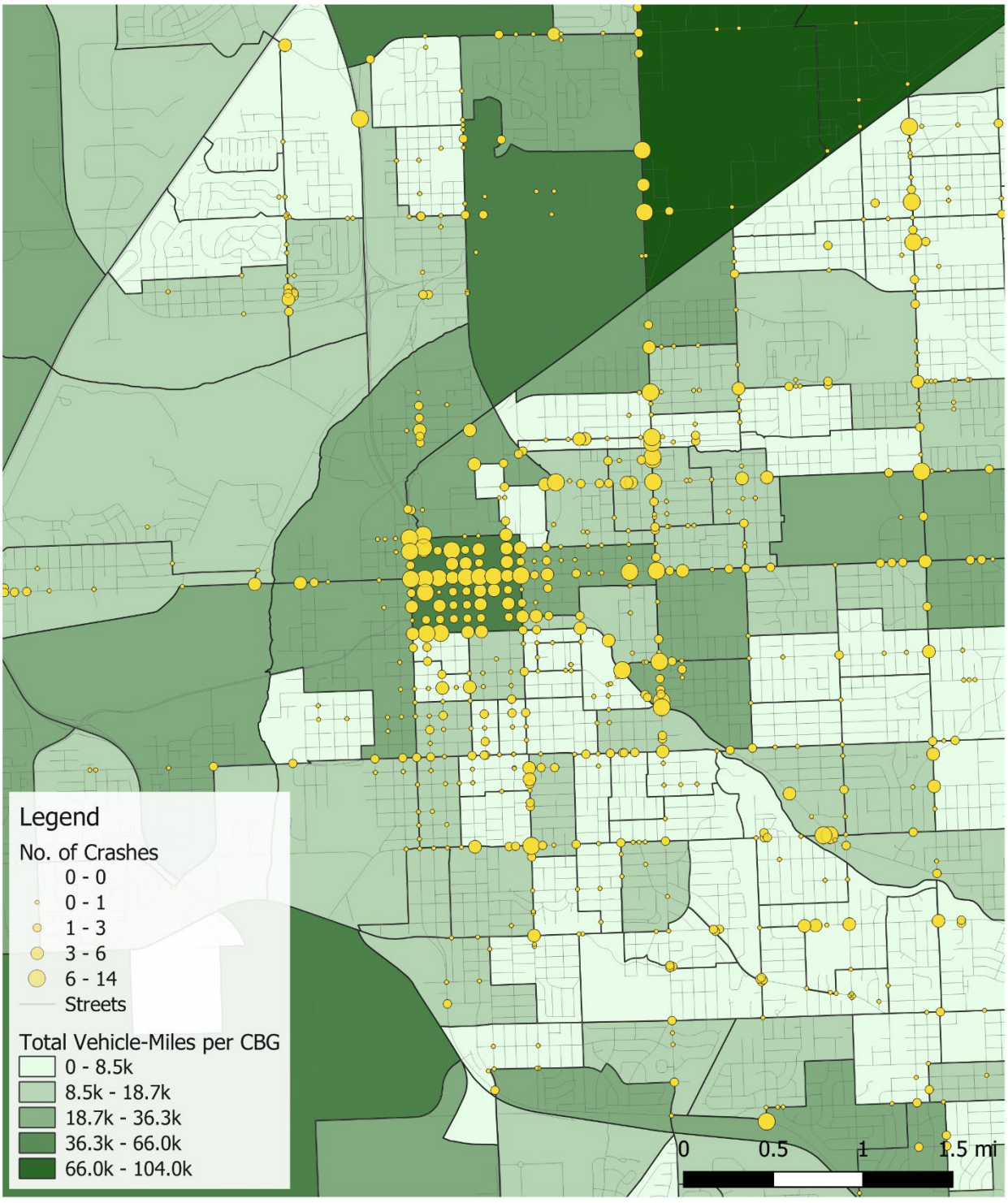

Figure 5.1 Number of crashes per intersection vs. vehicle miles per CBG

There is an even higher spatial correlation between the number of motorized-to-non-motorized crashes and active miles per CBG, as shown in Figure 5.2. Similar to vehicle miles in Figure 5.1, many CBGs surrounding the Lincoln CBD contain a high number of active miles. In fact, these CBGs comprise a proportionality higher share of the active miles for the study area than vehicle miles. This phenomenon is due to the many commercial, residential, and educational land uses within the Lincoln CBD and its surrounding areas. Notable CBGs that contained higher levels of active miles compared to their vehicle miles include the CBGs northwest of the Lincoln CBD. Table 5.3 provides quantitative confirmation of this finding.

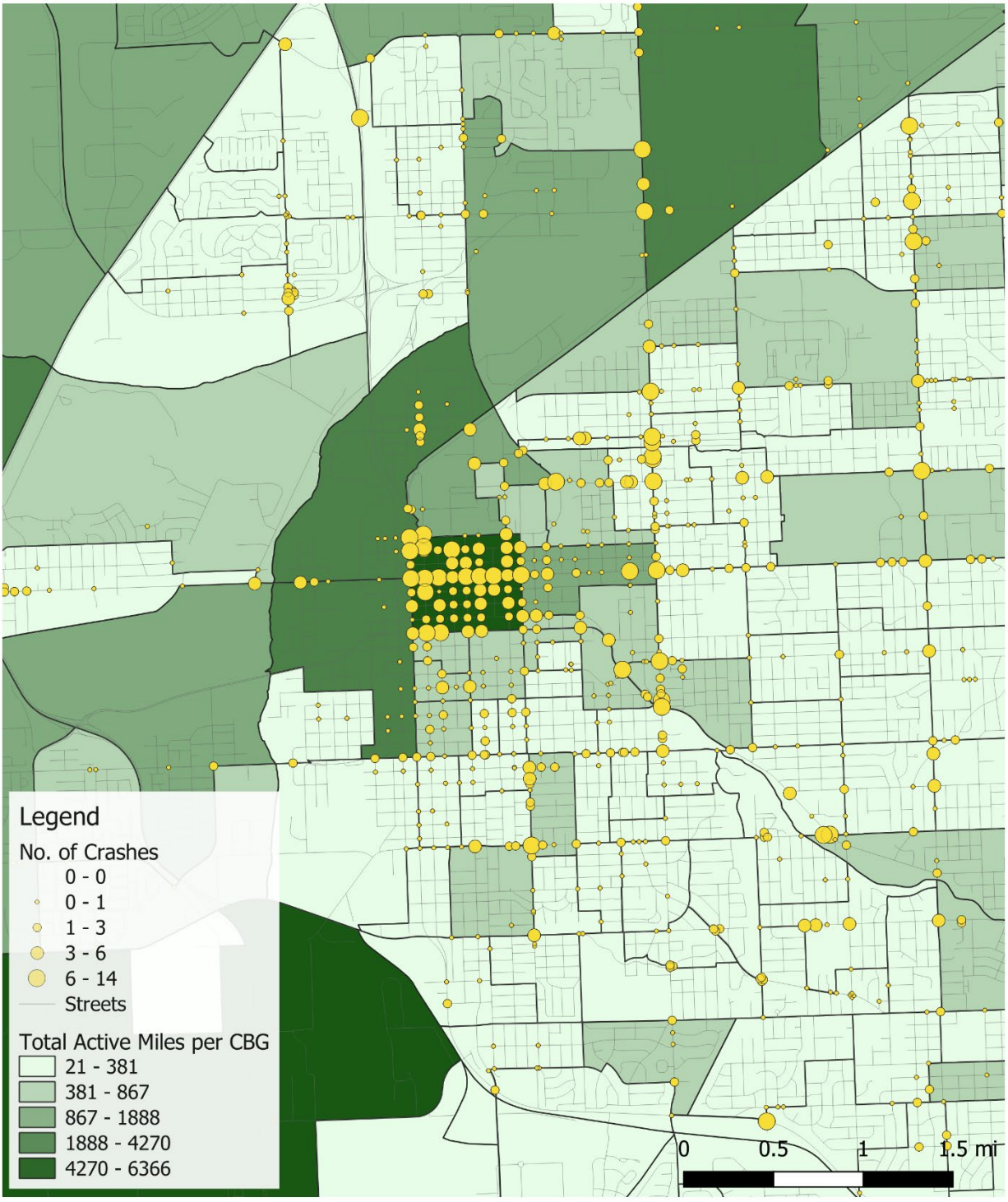

Figure 5.2 Number of crashes per intersection vs. active miles per CBG

Figure 5.3a (proportion non-white) exhibits the lowest correlation with total crash count. As previously discussed, the Lincoln CBD contains the most crashes of any CBG. However, this CBG does not contain a high proportion of non-white citizens, which affects the correlation strength between these two variables. Additionally, while there are many crashes on 27th Street, only a few CBGs along this corridor have a high proportion of non-white citizens. Quantitative correlations show that the percentage of non-white citizens had the lowest statistical significance of all four variables included in the figures.

Figure 5.3b shows the correlation between the number of crashes per intersection and poverty rate per CBG. There is a stronger spatial correlation between these two variables than for the proportion non-white variable because the CBD and the CBGs surrounding it having a higher proportion of citizens in poverty than other areas of Lincoln.

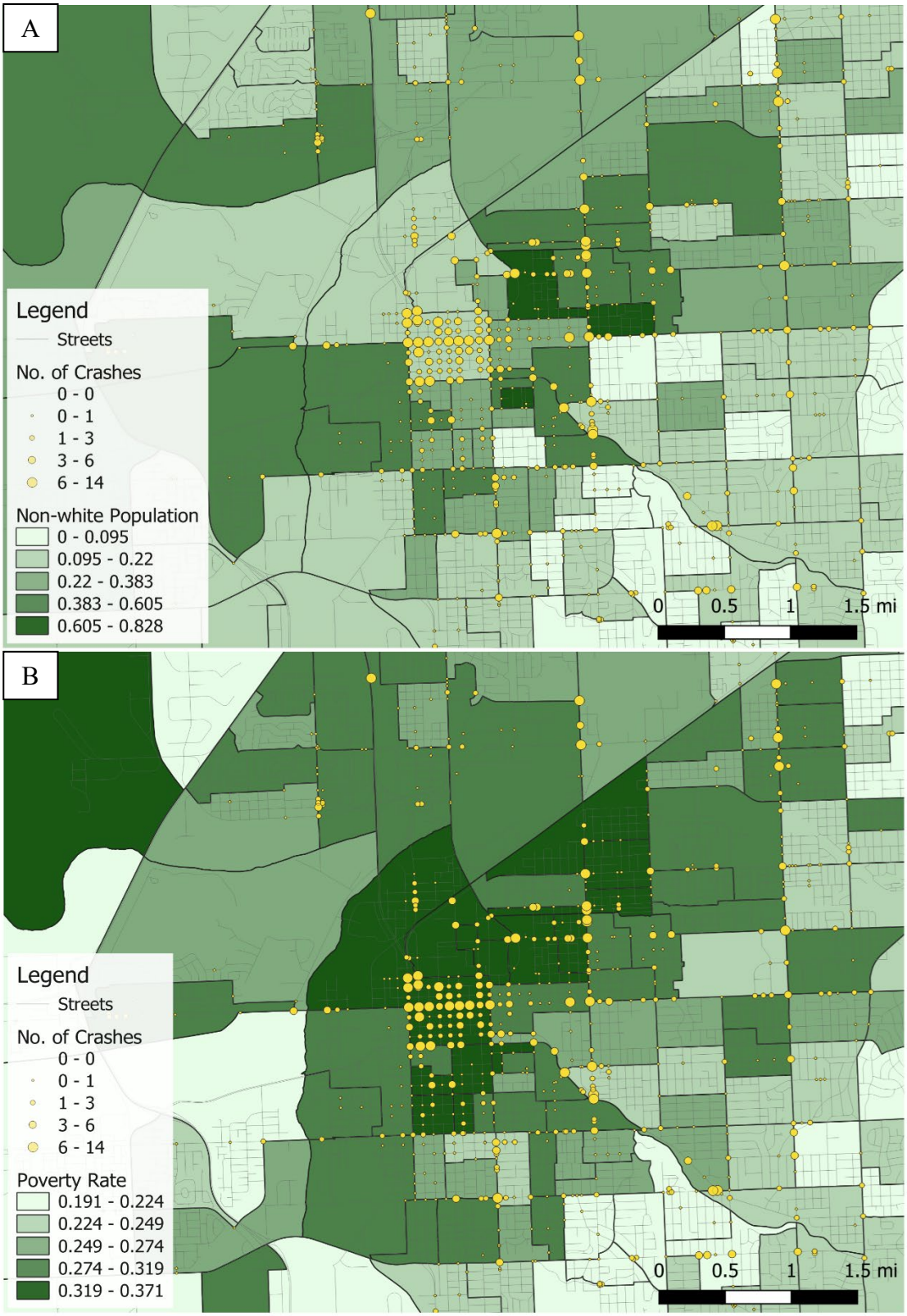

Figure 5.3 Number of crashes per intersection vs proportion of non-white population (a) and poverty rate (b) in CBG

Table 5.3 Quantitative Analysis of Crash Analysis

Variable	Correlation	P-Value
Vehicle Miles	0.58	0.00
Active Miles	0.76	0.00
Non-White Population	0.13	0.09
Poverty	0.29	0.00

5.3.2 Crash Rate Analysis Results

As with the general crash analysis, each of the two crash rate analyses had visual and quantitative components. The crash analysis from Section 5.3.1 shows that the most statistically significant correlations with crash counts that arise come from traffic volumes. This result highlights the importance of the creation of the three crash rate measures used for analysis in Section 5.3.2. Figure 5.4 shows a map of the *multiplicative crash rate* by CBG. Using the *multiplicative crash rate* measure, characteristics of CBGs that have high *multiplicative crash rates* are areas with a low amount of vehicle and active miles relative to the number of crashes. The most notable difference between this map and the maps shown in Figure 5.1 and Figure 5.2 is that the CBD of Lincoln and the surrounding CBGs show a lower concentrated risk, meaning that the crash counts are low relative to the total volume of traffic activity. The higher risk areas appear in CBGs outside of the Lincoln CBD and its surrounding areas. Additionally, there are higher risk areas in the south-central portion of the city, just south of Nebraska Parkway. Another notable aspect of the spatial analysis is that the CBG that encompasses the city campus of the University of Nebraska-Lincoln has a low risk based on this rate measure. The same result is not found on the two other major postsecondary campuses within the city: the east campus of

the University of Nebraska-Lincoln and Nebraska Wesleyan University. All the CBGs that overlap these campuses have much higher rate values. It is important to note that the fifth class in the GIS map in Figure 5.4 shows a considerably broader range of values compared to the preceding four classes. This discrepancy can be attributed to a few CBGs with substantially higher risk levels than all other CBGs. These outliers are discussed further in Section 5.3.4.

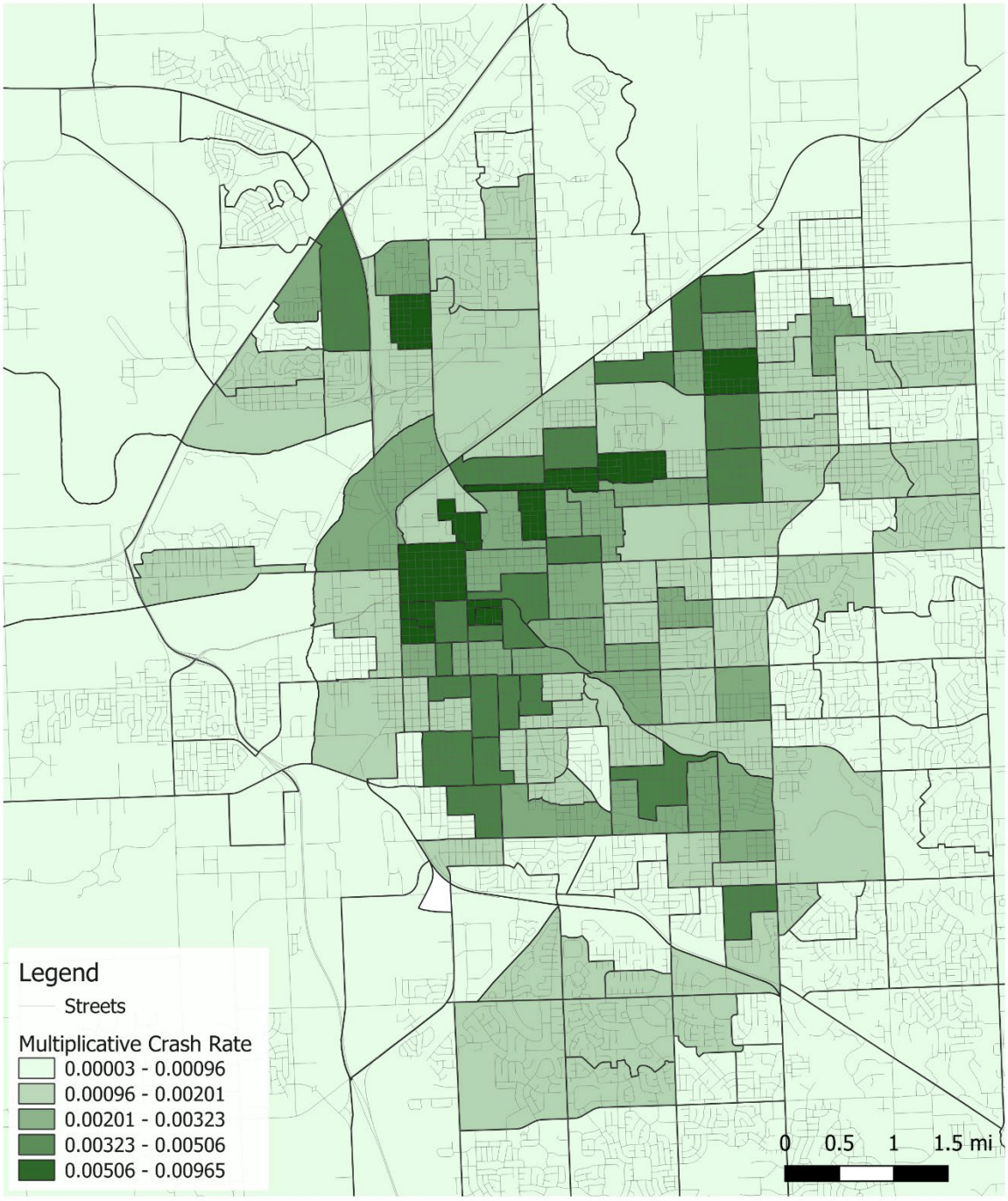

Figure 5.4 Multiplicative crash rate measure by CBG

Figure 5.5 shows the *ratio crash rate* level for each CBG. Spatial analysis was conducted using this map. The most notable difference between results from this rate measure compared to the *multiplicative crash rate* results is that the Lincoln CBD and the surrounding CBGs show significantly higher risk. This is primarily due to the high number of vehicle miles and crashes in these areas. However, there are a few higher risk areas just outside the CBD and many more in the outer portions of the city, compared with the results for the *multiplicative crash rate*. Another notable aspect of the spatial analysis is the low-risk areas in the city campus of UNL. Like the UNL city campus, the CBG containing the campus of Nebraska Wesleyan University also shows a low *ratio crash rate*. Finally, this spatial analysis had similar issues to the previous analysis, where the fifth class in the map in Figure 5.5 displays a broader range of values compared to the other classes, due to outliers in the data. These outliers are discussed further in the relative rank analysis in Section 5.3.4.

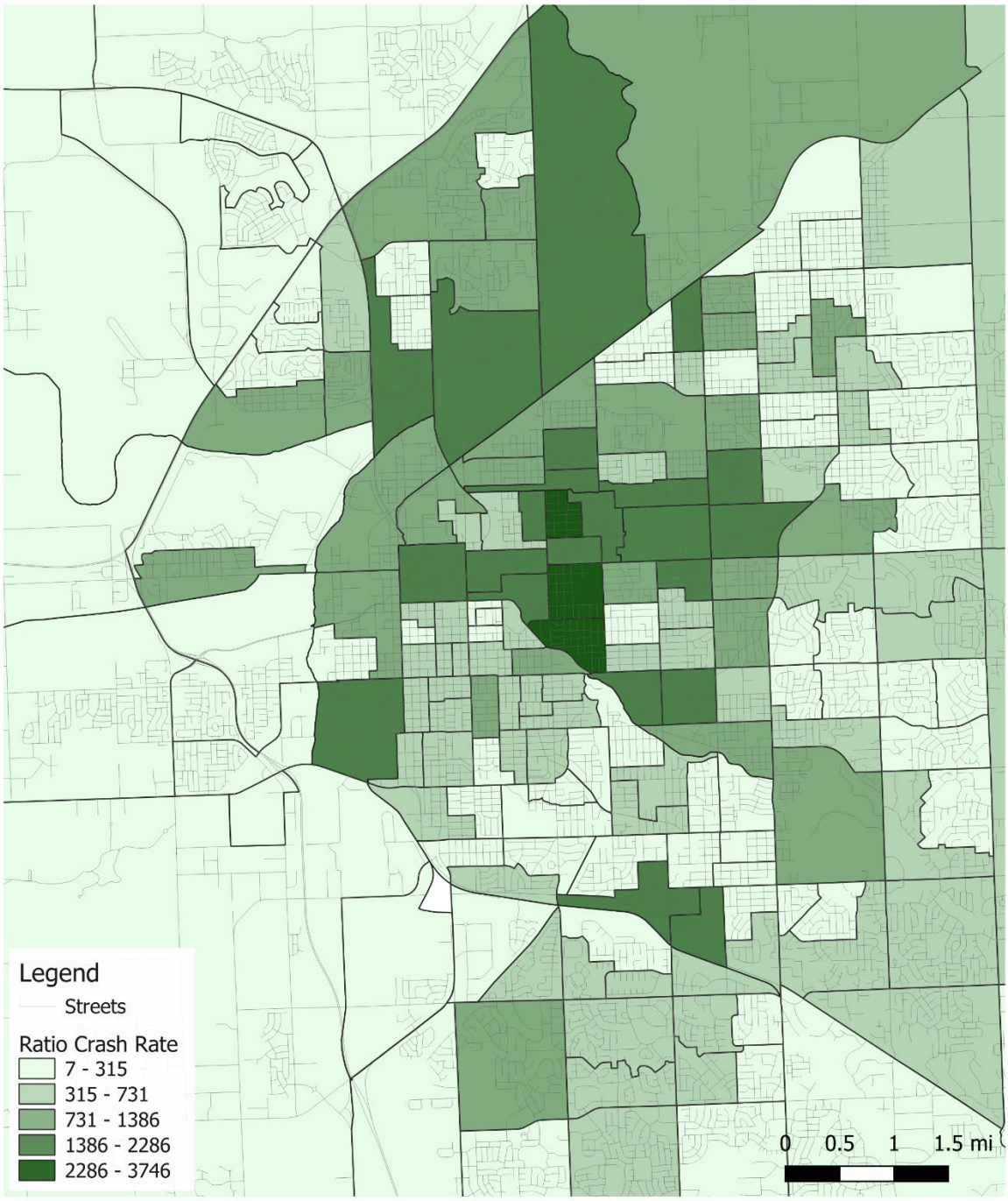

Figure 5.5 Ratio crash rate measure by CBG

Figure 5.6 provides an alternative summary based solely on active miles as the exposure measure. The results are a blend of the results from the multiplicative crash rate and the ratio crash rate. Similar to the ratio crash rate results, the regions to the east of the Lincoln CBD show elevated crash risk. On the other hand, the active travel crash rate reflected the multiplicative crash rate results in the regions north of Cornhusker Highway or near Nebraska Parkway, where reduced crash risk is observed. The most notable difference between the results from this rate measure compared to the other two rate results is the Lincoln CBD, showing reduced crash risk.

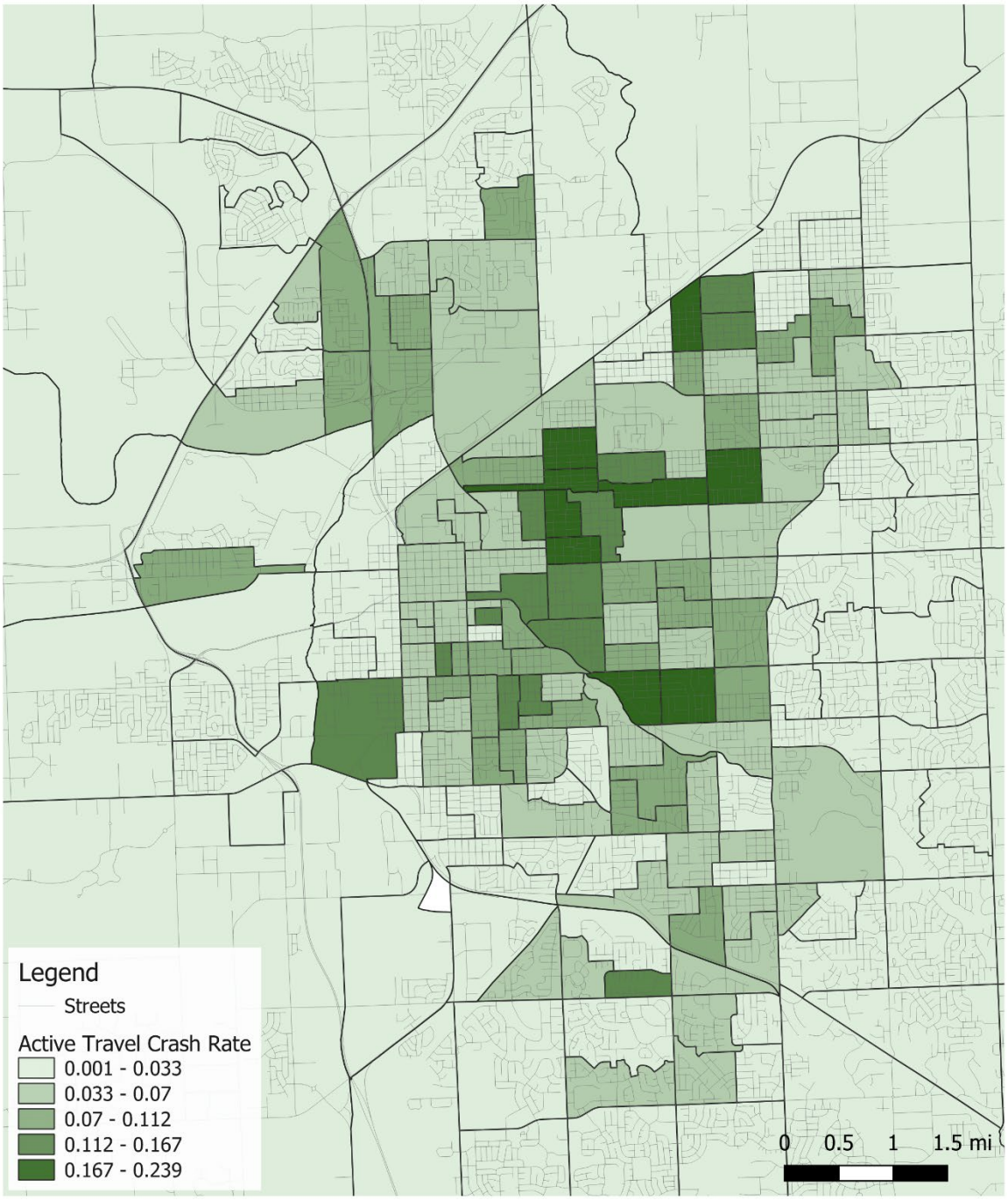

Figure 5.6 Active miles crash rate measure by CBG

5.3.3 Spatial Lag Model Analysis Results

Table 5.4 summarizes the results from the *multiplicative crash rate* spatial lag linear model. The spatial weight matrix used in this analysis used the Queen contiguity criterion. Under this criterion, spatial units were considered neighbors if they shared any part of their boundary, including corners. The resulting spatial weight matrix captured adjacent relationships among observations, reflecting the spatial connectivity in the dataset. The variables *density*, *street intersection density (weighted, auto-oriented intersections eliminated)*, and *household workers per job by CBG* were downscaled by a factor of 10,000. This scaling made model interpretation more difficult because each variable was on a different scale. To account for differences in the explanatory variable scale, an effect-size estimate was calculated by multiplying the variable parameter and standard deviation, dividing the product of these two values by the mean of the dependent variable (i.e., crash rate), and multiplying this value by 100 to obtain a percentage (Gelman et al., 2020). While the parameter value measures a change in the dependent variable for a one-unit change in the explanatory variable, the effect size measures a percent change in the dependent variable for a one-standard- deviation change in the explanatory variable. This standardization helps to demonstrate the effect each variable has on the dependent variable, regardless of scale.

Both demographic variables, *percent of one-car households in CBG* and *percentage of low wage workers in CBG*, were positively correlated with *multiplicative crash rate*. These two variables also had two of the top five largest effect sizes. Additionally, there was strong evidence that the effects were statistically significant. All diversity variables were negatively correlated with the rate measure. While *5-tier employment entropy* displayed high statistical significance an

a large effect size, *trip productions and trip attractions equilibrium index* and *household workers per job by CBG* had much weaker effects on *multiplicative crash rate*.

Density variables exhibited varying results. *Gross employment density (jobs/acre) on unprotected land* and *street intersection density (weighted, auto-oriented intersections eliminated)* were both negatively correlated with *multiplicative crash rate*. The two density variables that positively correlated with *multiplicative crash rate* (*density* and *network density in terms of facility miles of multi-modal links per square mile*), exhibited much different effect size estimates. *Network density in terms of facility miles of multi-modal links per square mile* was highly statistically significant and had the largest effect size estimate. The three other density variables exhibited low effect sizes for the *multiplicative crash rate* measure. The *walkability index* was negatively correlated with *multiplicative crash rate*, with a moderate effect size and statistical significance. Finally, the spatial weight term for this model had very low statistical significance, suggesting there was minimal residual spatial dependency for the *multiplicative crash rate* measure after controlling for other covariates.

Table 5.4 Results for Multiplicative Crash Rate Spatial Lag Linear Model

Variable	**Coefficient**	**Effect Size**	**P-value**
Demographic Variables			
Percent of one-car households in CBG	0.16	79.18	0.01
Percentage of Low Wage Workers in CBG	0.15	79.45	0.01
Diversity Variables			
5-tier employment entropy	-0.11	-88.62	0.00
Trip productions and trip attractions equilibrium index	-0.040	-34.35	0.26
Household Workers per Job, by CBG	-7.94	-44.95	0.15
Density Variables			
Gross employment density (jobs/acre) on unprotected land	-2.30×10^{-3}	-76.12	0.00
Density	2.55×10^{-3}	44.45	0.19
Street intersection density (weighted, auto-oriented intersections eliminated)	-1.95	-59.04	0.06
Network density in terms of facility miles of multi-modal links per square mile	0.010	112.27	0.00
Transport Variables			
Walkability Index	-0.010	-67.64	0.07
Model Related Variables			
Multiplicative Crash Rate Spatial Weight	3.34×10^{-3}	----------	0.94
Constant	0.11	----------	0.03
Pseudo-R^2	0.23		
Spatial Pseudo-R^2	0.23		
Number of Observations	187		
Number of Variables	12		

Table 5.5 summarizes the results for the *ratio crash rate* spatial lag linear model. The demographic variables included in this model (*total households* and *percent of two-plus-car households in CBG*) were both positively correlated with *ratio crash rate*. The effect sizes for these variables were among the lowest among all included variables. This model contained fewer diversity variables than the *multiplicative crash rate* model but added two new variables, *employment* and *destination accessibility*. These variables were positively correlated with *ratio crash rate*. *Regional diversity* and *office jobs within a 5-tier employment classification scheme* had large effect sizes and showed high statistical significance. *Destination accessibility* had a moderate statistical significance and a low effect size.

Much like the *multiplicative crash rate* model, the density variables results were mixed. The density variables include *network density in terms of facility miles of auto-oriented links per square mile*, *gross entertainment (5-tier) employment density (jobs/acre) on unprotected land*, *and intersection density in terms of multi-modal intersections having three legs per square mile*, all of which were statistically significant. *Network density in terms of facility miles of auto-oriented links per square mile* and *intersection density in terms of multi-modal intersections having three legs per square mile* were positively correlated with crash rate, while *gross entertainment (5-tier) employment density (jobs/acre) on unprotected land* was negatively correlated. The effect sizes for *intersection density in terms of multi-modal intersections having three legs per square mile* and *gross entertainment (5-tier) employment density (jobs/acre) on unprotected land* were high, while *network density in terms of facility miles of auto-oriented links per square mile* had a moderate effect size. Lastly, the spatial dependence was much higher for the *ratio crash rate* model results than for the *multiplicative crash rate* model.

Table 5.5 Results for Ratio Crash Rate Spatial Lag Linear Model

Variable	**Coefficient**	**Effect Size**	**P-Value**
Total households	0.33	17.68	0.12
Percent of two-plus-car households in CBG	0.080	21.98	0.04
Regional diversity of employment to population	0.070	30.33	0.00
Office jobs within a 5-tier employment classification scheme	0.39	38.25	0.00
Jobs within 45 minutes auto travel time, time- decay (network travel time) weighted	2.20×10^{-4}	23.17	0.06
Network density in terms of facility miles of auto-oriented links per square mile	0.010	28.40	0.00
Intersection density in terms of multi-modal intersections having three legs per square mile	1.60×10^{-3}	57.41	0.00
Gross entertainment (5-tier) employment density (jobs/acre) on unprotected land	-0.030	-34.37	0.01
Ratio Crash Rate Spatial Weight	0.060	----------	0.08
Constant	-0.18	----------	0.00
Pseudo R-squared	0.29		
Spatial Pseudo R-squared	0.28		
Number of Observations	187		
Number of Variables	10		

Table 5.6 summarizes the results for the *active miles crash rate* spatial lag linear model. This model contained a smaller number of vatiable compared to the two previous models. Both demographic variables (*percent of two-plus-car households in CBG* and *percent low wage workers*) were positively correlated with the *active miles crash rate with percent low wage workers* having a larger effect size. As for diversity variables, *Regional diversity of employment to population* showed the lowest statistical significance and smallest effect size of all the included variables. Much like the *multiplicative crash rate* and *ratio crash rate* models, the density variables results were mixed. The density variables included *residential density in HU/acre*, *Street intersection density*, and *Intersection density in terms of multi-modal intersections having three legs per square mile*. All of these variables demonstrated statistical significance and exhibited comparable effect sizes. Lastly, the spatial dependence was notably the highest in terms of coefficient and statistical significance for the *active miles crash rate* model compared to the other models.

Table 5.6 Results for Active Miles Crash Rate Spatial Lag Linear Model

Variable	**Coefficient**	**Effect Size**	**P-Value**
Percent of two-plus-car households in CBG	3.657	13.510	0.051
Percent low wage workers	27.619	19.321	0.003
Regional diversity of employment to population	-1.506	-7.702	0.165
Residential density in HU/acre	0.257	15.703	0.013
Street intersection density (auto-oriented removed)	-0.008	-14.109	0.029
Intersection density in terms of multi-modal intersections having three legs per square mile	0.048	17.361	0.005
Active Crash Rate Spatial Weight	0.098	----------	0.000
Constant	-7.368	----------	0.018
Pseudo R-squared	0.413		
Spatial Pseudo R-squared	0.185		
Number of Observations	194		
Number of Variables	8		

A final aspect of the spatial lag results involved calculating the Moran's I statistic for the dependent variables of each model. This was done to confirm the spatial autocorrelation in these variables. A Moran's I statistic close to one suggests spatial clustering and a value close to negative one suggests spatial dispersion. The Moran's I statistic was computed for both crash rate measures, giving values of 0.47, 0.39 and 0.457 for the *multiplicative*, *ratio* and *active miles* models, respectively. The Moran's I statistic for all three models indicates a moderate spatial autocorrelation.

5.3.4 Relative Rank Analysis Results

The results for the relative rank analysis for *multiplicative crash rate* are displayed in Table 5.7. Most of the highest ranking CBGs by *multiplicative crash rate* are located just outside the Lincoln CBD. This result is bolstered by analyzing the spatial analysis for crashes, detailed in Figure 5.1 and Figure 5.2, which illustrate that vehicle and active miles are lower in the CBGs surrounding the CBD; however, there are a relatively large number of crashes in these areas, resulting in a high crash rate. Additionally, most of the CBGs included in Table 5.7 also have relatively high ranking CBGs surrounding them (with some support from the spatial lag model results for this spatial correlation). CBGs boundaries sometimes do not perfectly overlap the boundaries for neighborhoods, so different CBGs may contain similar traffic patterns. Table 5.7 also shows the crash counts in these high ranking CBGs; the crash counts shown are not notably high, excepting the Lincoln CBD that is ranked as the highest CBG for crash count over the analysis period.

Table 5.7 Sample of Results from Multiplicative Crash Rate Relative Rank Analysis

Major Neighborhood inside CBG	Multiplicative Crash Rate Measure	Multiplicative Crash Rate Rank	Crash Counts	Crash Count Rank
Near South	1015.53	1.00	9	0.04
40th & A	813.26	0.80	2	0.01
West Lincoln	778.16	0.77	4	0.02
40th & A	485.31	0.48	3	0.01
Near South	275.17	0.27	10	0.04

The results for the relative rank analysis for *ratio crash rate* are shown in Table 5.8. Four of the five highest ranking CBG by *ratio crash rate* are just east of the Lincoln CBD, in the Woods Park and Hartley neighborhoods.

Table 5.8 Sample of Results from Ratio Crash Rate Relative Rank Analysis

Major Neighborhood inside CBG	Ratio Crash Rate	Ratio Crash Rate Rank	Crash Counts	Crash Count Rank
Woods Park	16503.79	1.00	43	0.18
Hartley	7222.45	0.44	32	0.13
Oak Hills	6683.26	0.40	12	0.05
Edenton South	4863.86	0.29	14	0.06
Havelock	4849.73	0.29	7	0.03

The results for the relative rank analysis for *active miles crash rate* are shown in Table 5.9. Adjusting for active miles brings several CBGs to the top of the ranking list that are ranked lower based only on crash count. Similar to the ratio crash rate results, Woods Park arises as a top location.

Table 5.9 Sample of Results from Active Miles Crash Rate Relative Rank Analysis

Major Neighborhood inside CBG	Active Crash Rate	Active Crash Rate Rank	Crash Counts	Crash Count Rank
Woods Park	0.64	1.00	43	0.18
Havelock	0.61	0.95	19	0.08
Witherbee	0.59	0.92	18	0.08
40th & A	0.58	0.91	20	0.08
Clinton	0.57	0.89	28	0.12

Analysis was also conducted to compare how the two crash rate measures differed within the same CBG. The "movement" of ranks in Table 5.10 represents the difference between the relative rank according to each pair of rate measures. For example, a negative value in column 2 is interpreted to mean that the *ratio crash rate* for a CBG is higher than the *multiplicative crash rate*, whereas a positive value means that the *multiplicative crash rate* is higher than the *ratio crash rate*. The CBGs representing the top and bottom 5 differences between *multiplicative* and *ratio exposure measures* are used for all differences. There is significant variation between the crash ranks for Near South (higher for multiplicative than ratio crash rate) and Woods Park (lower for multiplicative than ratio crash rate). Similar patterns are found for active miles rate, suggesting that the ratio and active miles rate measures give similar results for these CBGs.

Table 5.10 Sample of Results for Movement of Ranks across Models

Major Neighborhood inside/near CBG	Movement of Ranks between Multiplicative and Ratio Rate Measures	Movement of Ranks between Multiplicative and Active Miles Rate Measures	Movement of Ranks between Ratio and Active Miles Rate Measures
Near South	0.99	0.79	-0.20
40th & A	0.80	0.67	-0.13
West Lincoln	0.76	0.55	-0.21
40th & A	0.46	0.34	-0.13
Near South	0.26	-0.01	-0.27
Edenton South	-0.29	-0.54	-0.25
Havelock	-0.29	-0.35	-0.06
Oak Hills	-0.40	-0.25	0.15
Hartley	-0.43	-075	-0.32
Woods Park	-0.99	-0.99	0.00

CBGs with high positive movement values have a high number of total miles (i.e., vehicle plus active miles), but a very high proportion of those total miles comprise vehicle miles. The opposite is expected for CBGs with a high negative value; however, few CBGs in Lincoln have a high proportion of total miles accounted for by active miles. As such, the ratio between active miles and vehicle miles in these CBGs will be some of the lowest throughout the city.

The final analysis on this topic is the difference between the *relative crash rank* of each CBG and the rank of each rate measure for each CBG. The *relative crash rank* was determined through a similar equation to the relative rank of the rate measures, given by

$$\text{Relative Crash Rank} = \frac{\text{Crash Count}_{\text{x}} - \text{Crash Count}_{\text{low}}}{\text{Crash Count}_{\text{high}} - \text{Crash Count}_{\text{low}}} \quad (5.14)$$

Where Crash Count_{x} represents the crash count in a single CBG.

Table 5.11 below shows the top and bottom 5 CBGs with the most movement between *relative crash rank* and the three crash rate ranks (i.e., multiplicative, ratio, and active miles). Positive values in Table 5.11 indicate a movement from a higher *crash rank* to a lower *crash rate rank*. A negative value indicates a movement from a lower *crash rank* to a higher *crash rate rank*. These results indicate that normalization by traffic volume consistently decreases the crash ranking for downtown Lincoln relative to the unnormalized crash count. Similarly, Near South, Woods Park, and Clinton neighborhoods represent higher risk (i.e., higher crash rate) areas depending on the normalization assumption.

Table 5.11 Sample of Results from the Movement of Ranks between Crash Rank and Rate Rank

Major Neighborhood inside CBG	Crash to Multiplicative Rank Movement	Major Neighborhood inside CBG	Crash to Ratio Rank Movement	Major Neighborhood inside CBG	Crash to Active Miles Rank Movement
Downtown Lincoln	1.00	Downtown Lincoln	0.83	Downtown Lincoln	0.79
Downtown Lincoln	0.26	The Haymarket	0.17	North Bottom	0.10
Sunset Acres	0.25	Sunset Acres	0.14	North Bottom	0.07
South Salt Creek	0.21	UNL City Campus	0.08	UNL	0.05
The Haymarket	0.20	Near South	0.08	Fallbrook	0.00
Near South	-0.23	Edenton South	-0.24	Havelock	-0.77
40th & A	-0.47	Havelock	-0.26	Witherbee	-0.82
West Lincoln	-0.75	Hartley	-0.30	40th & A	-0.82
40th & A	-0.79	Oak Hills	-0.35	Woods Park	-0.84
Near South	-0.96	Woods Park	-0.82	Clinton	-0.87

5.3.5 Link and Intersection Safety Ranking Results

The approach described in Section 5.2.5 is used to calculate the composite normalized scores for the attribute groups of links and intersections. The calculated scores were then applied to all the links and intersections in Lincoln based on their attribute groups. For instance, all links that share the same 8 attributes considered in the link ranking (same bins for volume, Number of lanes, median type, shoulder presence, shoulder width, speed limit, road classification, and facility type) would share the same Composite Normalized Score (CNS). The Normalized Scores

for all bins (within attributes) are presented in Figure 5.7 for links and Figure 5.8 for intersections. The Ranking of all the links and intersections in Lincoln is visualized in Figure 5.8 and Figure 5.8, respectively.

Attribute (Risk Factor)	Attribute Weight	Weighting Factor (WF)	Bins	Mileage	Crashes	Crash Rate (CR)	Normalized Score (NS)	Weighted Normalized Score = WF x NS									
								1	2	3	4	5	6	7	8	9	10
Volume	20	2	0-250	1,215.13	265	0.218	1	2	4	6	8	10	12	14	16	18	20
			250-500	236.32	174	0.736	2										
			500-1000	410.72	446	1.086	3										
			1000-2000	101.53	327	3.221	9										
			2000_or_more	235.93	923	3.912	10										
Shoulder_width	10	1	0	1,898.82	2,083	1.097	10	1	2	3	4	5	6	7	8	9	10
			1-3	88.72	6	0.068	1										
			3-6	149.63	36	0.241	2										
			6_or_more	62.46	10	0.160	1										
Shoulder_presence	10	1	No_shoulder	1,899.43	2,083	1.097	10	1	2	3	4	5	6	7	8	9	10
			Shoulder	300.21	52	0.173	1										
Median Type	10	1	None	1,915.45	1,579	0.824	5	1	2	3	4	5	6	7	8	9	10
			Painted	14.40	16	1.111	6										
			Open	7.20	10	1.390	7										
			Barrier	0.15	0	0.000	1										
			Raised	262.43	530	2.020	10										
Facility_type	10	1	One-way	27.23	262	9.621	10	1	2	3	4	5	6	7	8	9	10
			Two-way	2,172.07	1,873	0.862	1										
			Ramp	0.34	0	0.000	1										
Number_of_lanes	10	1	1	40.17	18	0.448	1	1	2	3	4	5	6	7	8	9	10
			2	1,854.98	985	0.531	1										
			3	230.40	547	2.374	3										
			4_or_more	74.08	585	7.897	10										
Speed_limit	20	2	20-30	1,410.25	882	0.625	2	2	4	6	8	10	12	14	16	18	20
			30-40	253.36	772	3.047	10										
			40-50	322.08	463	1.438	5										
			50_or_more	213.95	18	0.084	1										
Road_classification	10	1	Remote residential	117.80	5	0.042	1	1	2	3	4	5	6	7	8	9	10
			Local	1,413.87	653	0.462	2										
			Collector	184.53	222	1.203	4										
			Minor_arterial	352.41	857	2.432	8										
			Major_arterial_and_above	131.02	398	3.038	10										
Composite Normalized Score (CNS) = 6 + 10 + 10 + 6 + 1 + 1 + 20 + 4 = 58																	

Figure 5.7 Calculation of Composite Normalized Score (CNS) for Links

Attribute (Risk Factor)	Attribute Weight	Weighting Factor (WF)	Bins	Ped and Bike AADT	Crashes	Crash Rate (CR)	Normalized Score (NS)	Weighted Normalized Score = WF x NS									
								1	2	3	4	5	6	7	8	9	10
Vehicle_AAADT	28.6	2.86	0-1000	97,460	178	0.00183	1	2.86	5.72	8.58	11.44	14.3	17.16	20.02	22.88	25.74	28.6
			1000-2000	22,632	51	0.00225	2										
			2000-4000	138,874	228	0.00164	1										
			4000-8000	57,393	195	0.00340	4										
			8000_or_more	144,695	1,036	0.00716	10										
Median Type	10	1.43	None	345,549	882	0.00255	1	1.43	2.86	4.29	5.72	7.15	8.58	10.01	11.44	12.87	14.3
			Painted	7,647	57	0.00745	10										
			Open	1,644	11	0.00669	9										
			Raised	106,213	738	0.00695	9										
Average_number_of_lanes	10	1.43	Less_than_2	19,155	45	0.00235	1	1.43	2.86	4.29	5.72	7.15	8.58	10.01	11.44	12.87	14.3
			2-3	334,614	1,032	0.00308	2										
			3-4	63,735	302	0.00474	6										
			4_or_more	43,550	309	0.00710	10										
Max_speed_limit	20	2.86	20-30	271,996	428	0.00157	1	2.86	5.72	8.58	11.44	14.3	17.16	20.02	22.88	25.74	28.6
			30-40	121,550	813	0.00669	9										
			40-50	64,471	444	0.00689	10										
			50_or_more	3,037	3	0.00099	1										
Highest_road_class	10	1	Remote residential	334	0	0.00000	1	1.43	2.86	4.29	5.72	7.15	8.58	10.01	11.44	12.87	14.3
			Local	191,440	144	0.00075	1										
			Collector	52,701	134	0.00254	3										
			Minor_arterial	174,905	994	0.00568	6										
			Major_arterial_and_above	41,674	416	0.00998	10										
Composite Normalized Score (CNS) = 2.86+14.3+2.86+25.74+4.29 = 50																	

Figure 5.8 Calculation of Composite Normalized Score (CNS) for Intersections

To this point, analysis has focused on CBG areal statistics and aggregate land use and transportation system features. The safety ranking analysis provided a facility-level complement from which Nebraska DOT can identify appropriate infrastructural countermeasures. These results supported the CBG findings. Outside the Lincoln CBD, the streets to its south and near the Nebraska State Capitol were identified as higher risk facilities for non-motorized road users. Vine Street, 27th Street, and the Capitol Parkway also arise as foci for safety improvements.

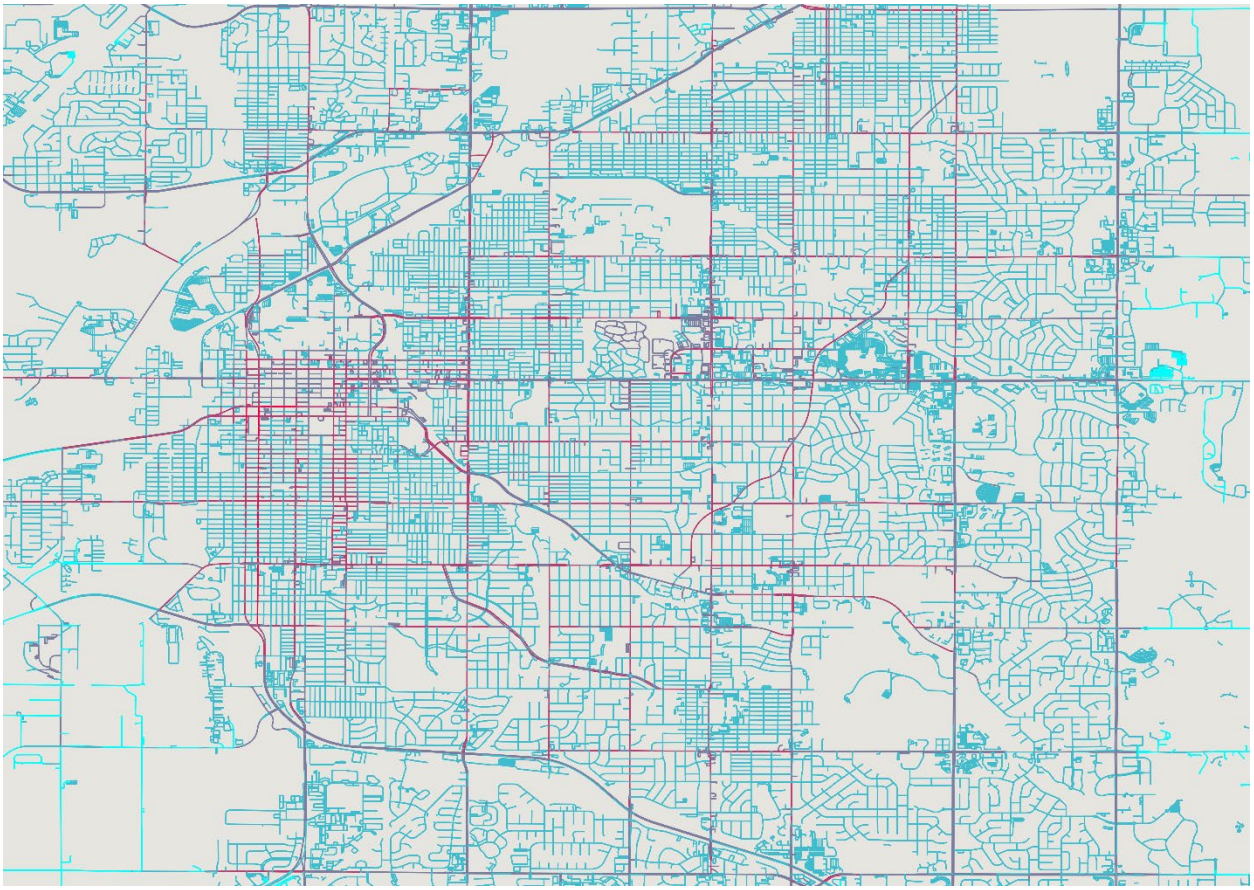

Figure 5.9 Composite Normalized Score (CNS) for links (red = higher risk)

Based on Figure 5.10, Vine Street and 27th Street are also consistent foci for safety improvements at their intersections. Fifty-sixth street near Holmes Lake was also identified as a location of high non-motorized traveler risk.

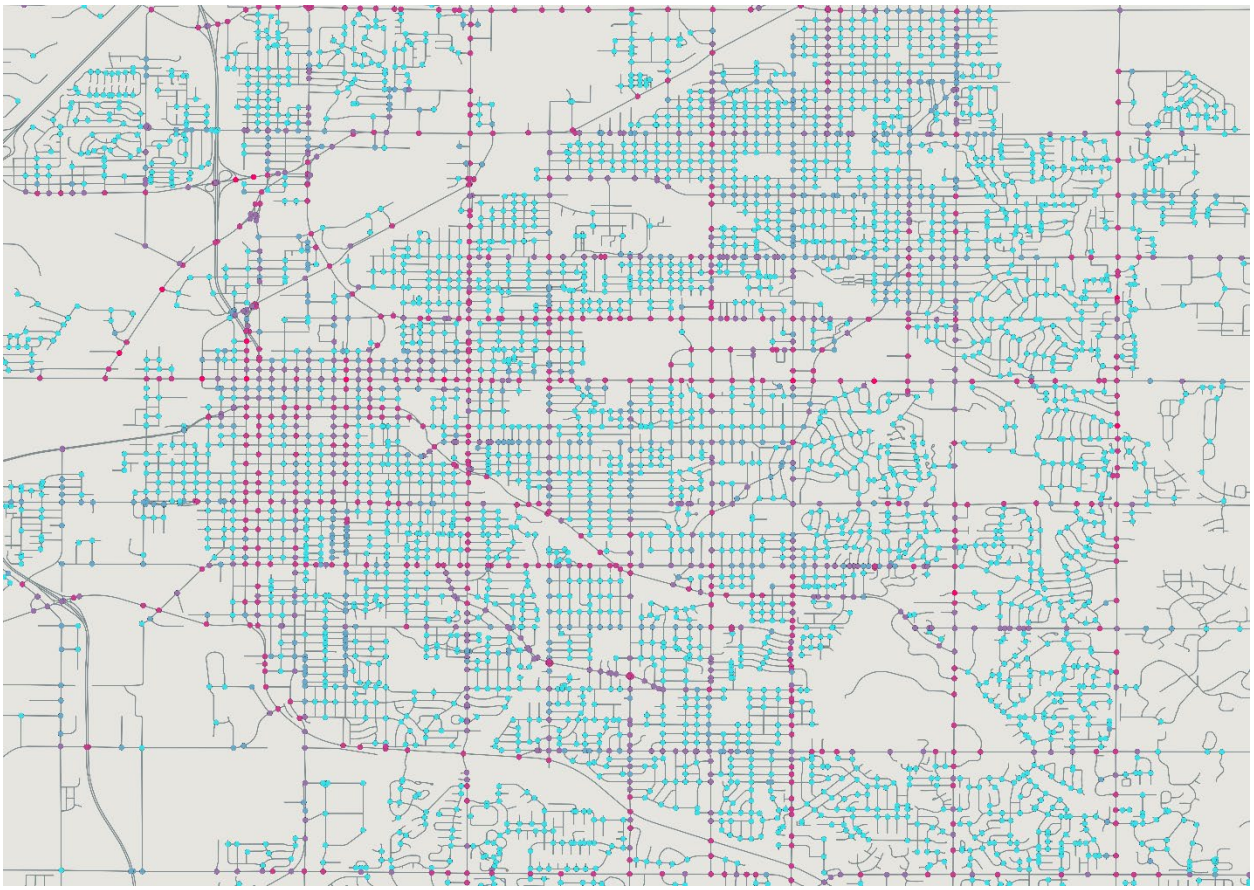

Figure 5.10 Composite Normalized Score (CNS) for intersections (red = higher risk)

5.4 Chapter Summary

Chapter 5 provided the safety analysis results using CBG, link, and intersection results. Both aggregate and facility-specific features were considered in these analyses. It was found that results differ depending upon the exposure measure used to normalize crash counts. While aggregate features, such as population and network density, affect non-motorized road user

safety, the clearest results arise when considering infrastructure at the facility scale. Several communities surrounding the Lincoln CBD were identified as areas for further study. Safety ranking analysis identified specific corridors requiring more in-depth safety analysis to identify countermeasures.

Chapter 6 Spatiotemporal Pattern Analysis and Visualization

This chapter presents the analysis and visualization of spatial and intertemporal patterns in traffic volumes for pedestrians and bicyclists at the street segment level. While area-based analyses, such as those using Traffic Analysis Zones (TAZ), census tracts, or census block groups, offer a broader perspective on safety trends across larger regions, a street segment-level analysis provides a more precise identification of volume patterns and potential risk locations. This approach allows for the pinpointing of specific intersections, crossings, or stretches of road where safety improvement opportunities may be concentrated.

A noteworthy observation arose when comparing block group-level data with street segments. This comparison revealed that street segment data offer better detail, as block group data were derived from aggregated segment data. Consequently, block group data may indicate a high volume, even when only specific street segments within that block group contribute, and the majority of the area in that block group lacks streets (Figure 6.1). The lack of granularity in block group-scale data raises concerns about potential biases and inaccuracies, underscoring the importance of selecting an appropriate geographical scale for safety analysis.

Moreover, map visualization uncovered a distinct trend in the presentation of hot spot maps for bicycle volumes, which appeared linear, while pedestrian volumes formed clusters across the city. This visual distinction may offer insights into unique patterns related to different transportation modes and varying areas or demographics. Overall, map visualization proved to be a powerful tool, effectively conveying complex data in a simple and easily understandable format. This capability enables users to discern patterns, trends, and relationships that might be challenging to extract from raw numerical data alone.

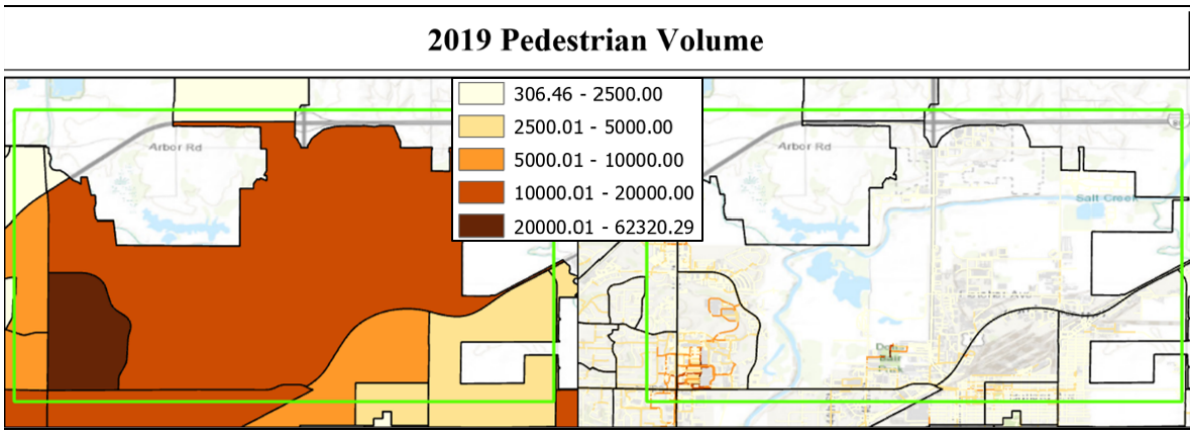

Figure 6.1 Maps showing the difference between the census block group (CBG) scale and segment scale visualization

6.1 Pre and During COVID-19

We compared bicycle and pedestrian traffic volume from 2019 and 2020 to understand the differences in traffic patterns between Non-COVID-19 and COVID-19 periods. Several studies indicated decreased pedestrian and bicycle volumes in 2020 compared to 2019 (Y. Li & Xu, 2021; Möllers et al., 2022). However, their geographic scale was area-based. In contrast, our analysis was done with segment scale data, giving us detailed information. There was a noticeable shift in pedestrian and bicycle activity, particularly near parks and residential areas, with an increase in traffic volume. In contrast, downtown activity decreased significantly in 2020 compared to 2019.

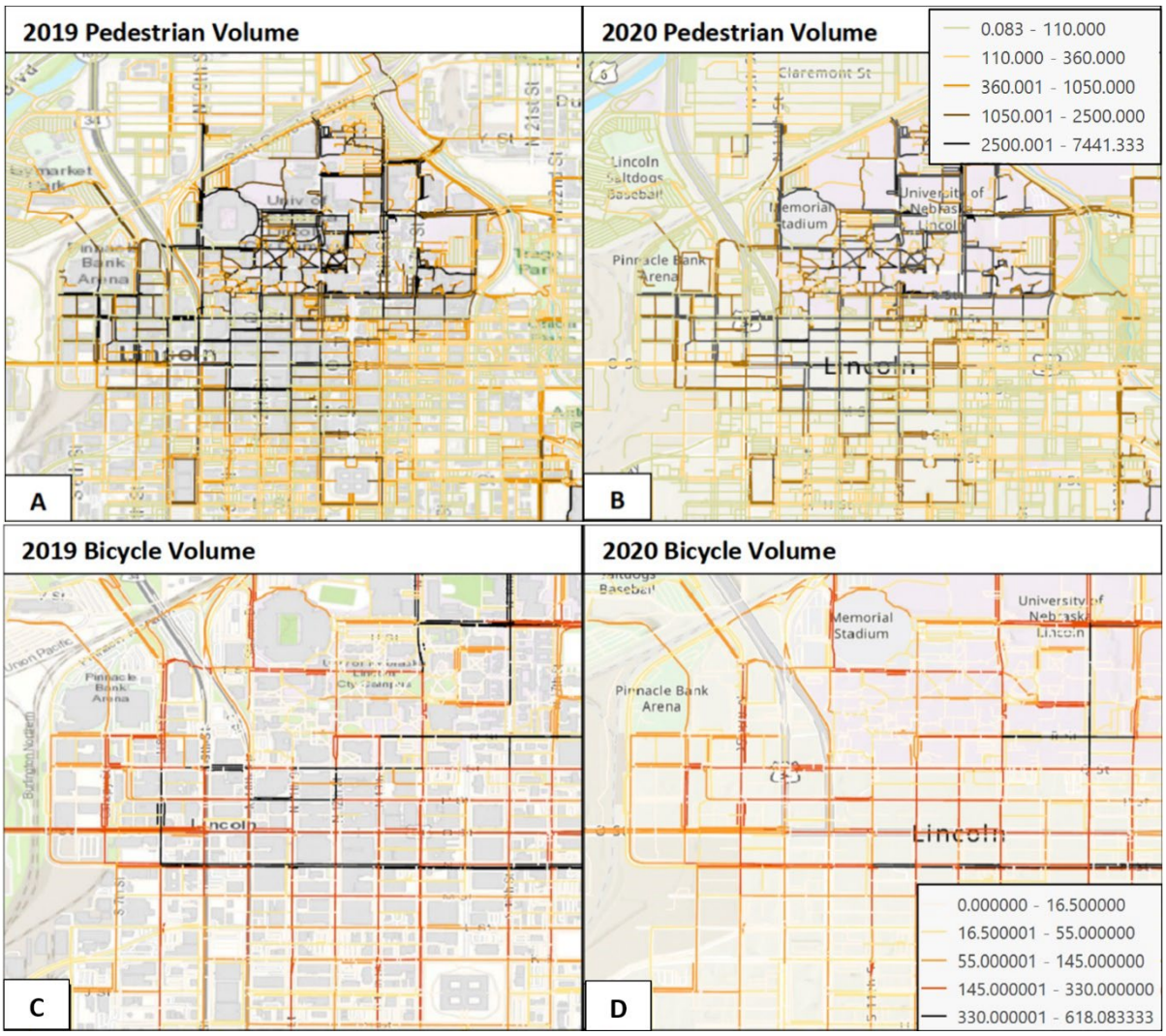

Figure 6.2 Comparing 2019-2020 pedestrian and bicycle volume – CBD, UNL, Haymarket

Figure 6.2 shows the CBD and Haymarket areas. Pedestrian and bicycle volume in 2020 significantly decreased compared when compared to the previous year, 2019. This decrease is associated with the impact of the COVID-19 pandemic, reflecting an increase in remote work, as well as a decrease in business activities and commuting traffic during this period.

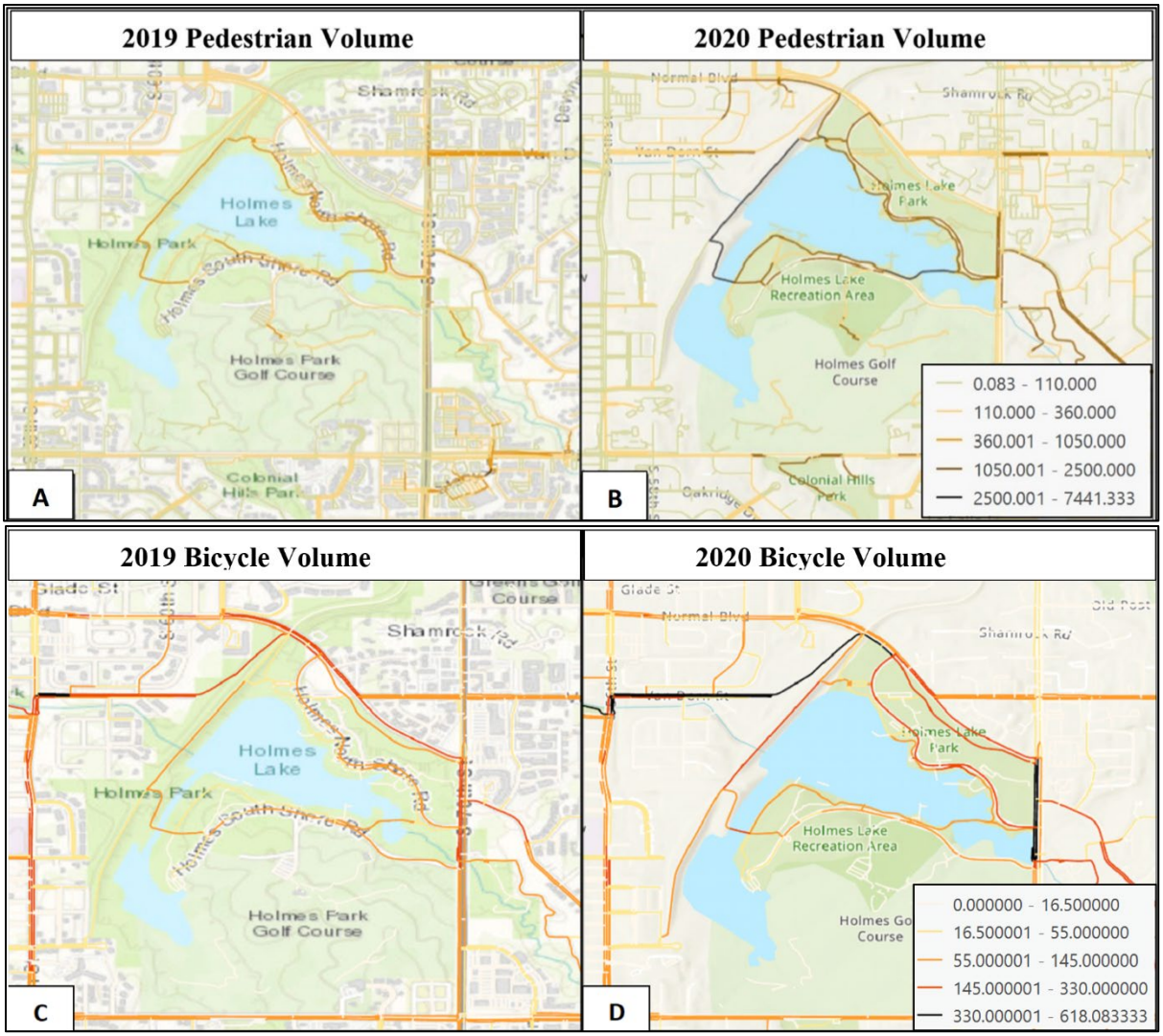

Figure 6.3 Comparing 2019-2020 pedestrian and bicycle volume – Holmes Lake Park

Interestingly, pedestrian and bicycle traffic increased in parks around Lincoln. For example, as shown in Figure 6.3, Holmes Lake Park shows increased pedestrian and bicycle traffic volume, indicating that people did more outdoor activities in the parks where they can maintain social distancing due to COVID-19.

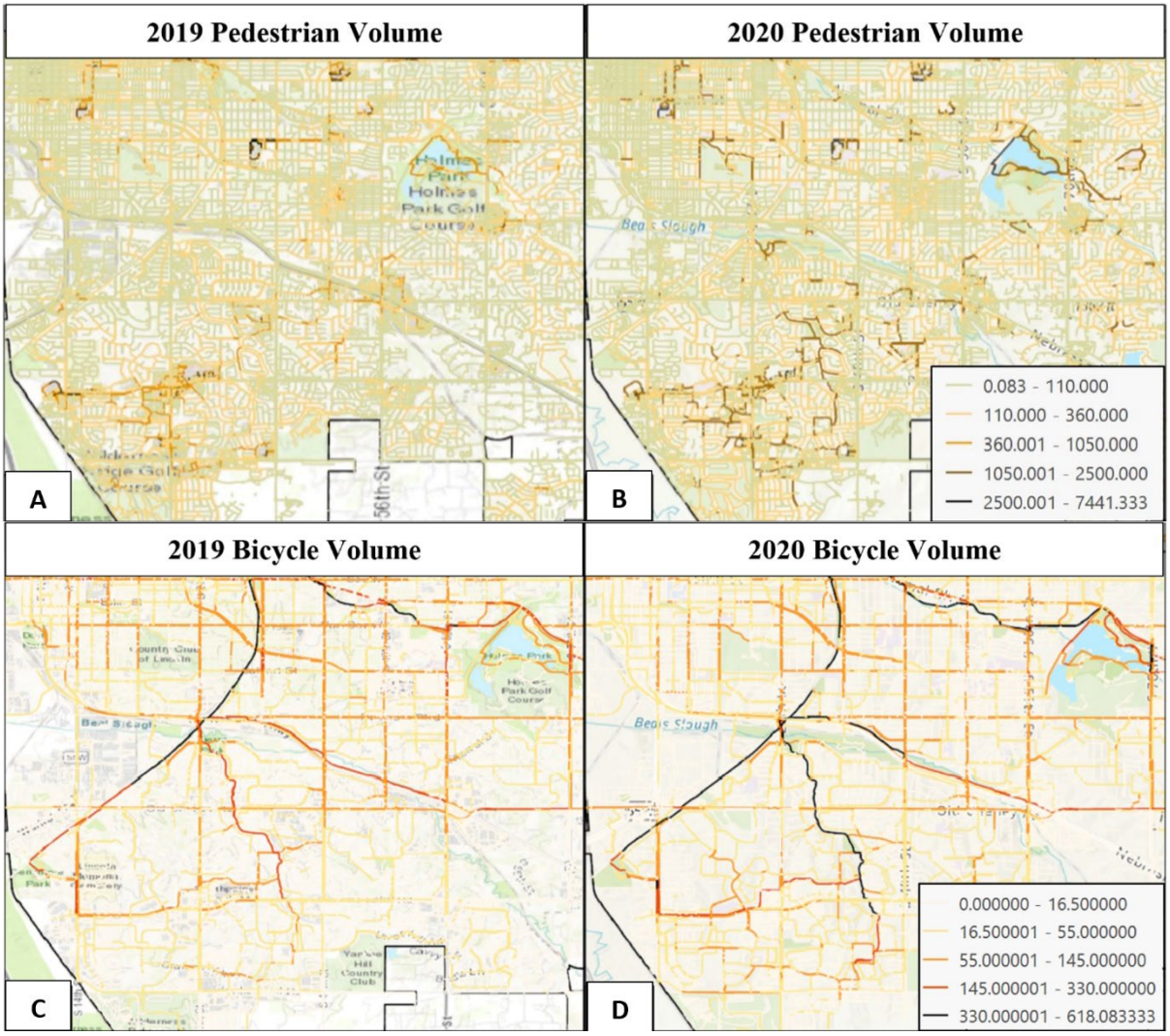

Figure 6.4 Comparing 2019-2020 pedestrian and bicycle volume - residential area

Furthermore, Figure 6.4 shows that pedestrian and bicycle traffic in residential areas increased in 2020 during COVID-19. This pattern indicates that people preferred activities around their home to avoid contact with other people.

As shown in Figure 6.4, these trends can be explained by a variety of reasons, including the closure of public spaces, the need for exercise, and the desire to spend time outside in a safe and socially isolated environment. Walking is a low-impact exercise that helps reduce stress, enhance cardiovascular health, and strengthen the immune system (Bonnell et al., 2022). It is

also an activity that can be easily included in daily routines, such as walking to the grocery store or walking around the neighborhood. The increase in walking activity during COVID-19 can be seen as a positive outcome because it promotes physical activity and can enhance overall health.

6.2 Weekday/Weekend Comparison

When comparing pedestrian and bicycle traffic patterns on maps showing volumes on weekdays and weekends, there were significant differences. The analyzed data highlighted trends in different areas that helped us understand pedestrian traffic patterns. First, public school areas showed a high pedestrian activity during weekdays. However, private school areas did not show much increase in pedestrian traffic during the weekdays. Second, parks showed a distinct pattern, with a significant increase in pedestrian volume during weekends, indicating increased leisure and recreational activities. Third, the University of Nebraska-Lincoln (UNL) city campus and Haymarket areas showed weekday/weekend dynamics. UNL city campus experienced more pedestrian traffic on weekdays, which was likely due to school activities and classes being in-session. In contrast, the Haymarket area had more weekend activities, indicating a shift toward entertainment, local restaurant, and market activities during those days. These visualizations provide insights into the pedestrian walking trend and highlight the variations across different areas between weekdays and weekends. Unlike pedestrian volume patterns, bicycle patterns did not exhibit significant variation between weekdays and weekends. However, there was a slight increase in bicycle volumes around park areas.

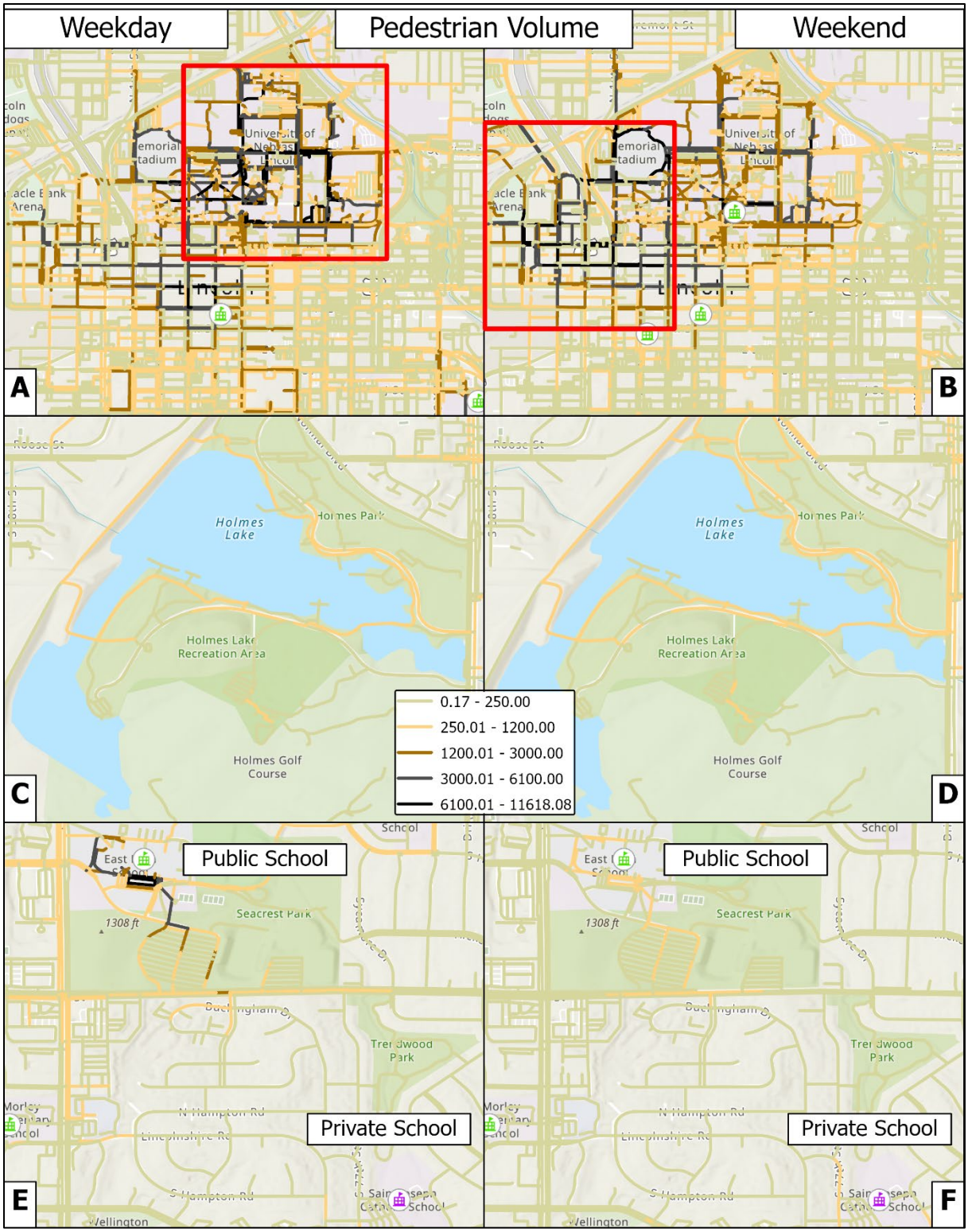

Figure 6.5 Comparing Weekday-Weekend Pedestrian Volume – Downtown, Haymarket, Holmes Lake Park, School Areas

6.3 Walkability and Pedestrian Volume

To explore whether people walk in a walkable environment, we created a bivariate map that divides areas into four quadrants based on Lincoln's walkability and daily pedestrian volume. The National Walkability Index from the EPA was used in this analysis (US EPA, 2014). The walkability index is at the census block group level, considering factors such as intersection density, proximity to transit stops, and the mix of employment types and housing. The four categories are: high pedestrian volume with high walkability, low volume-low walkability, low volume-high walkability, and high volume-low walkability. This mapping strategy comprehensively examines walkability and volume intersections at a block group level. It provides insights into urban planning and development initiatives by identifying areas in need of improvement and those excelling in both walkability and volume. Such bivariate maps facilitate visual understanding of walkability patterns, serving as strategic tools for policymakers and city planners to target specific areas for enhancements and interventions.

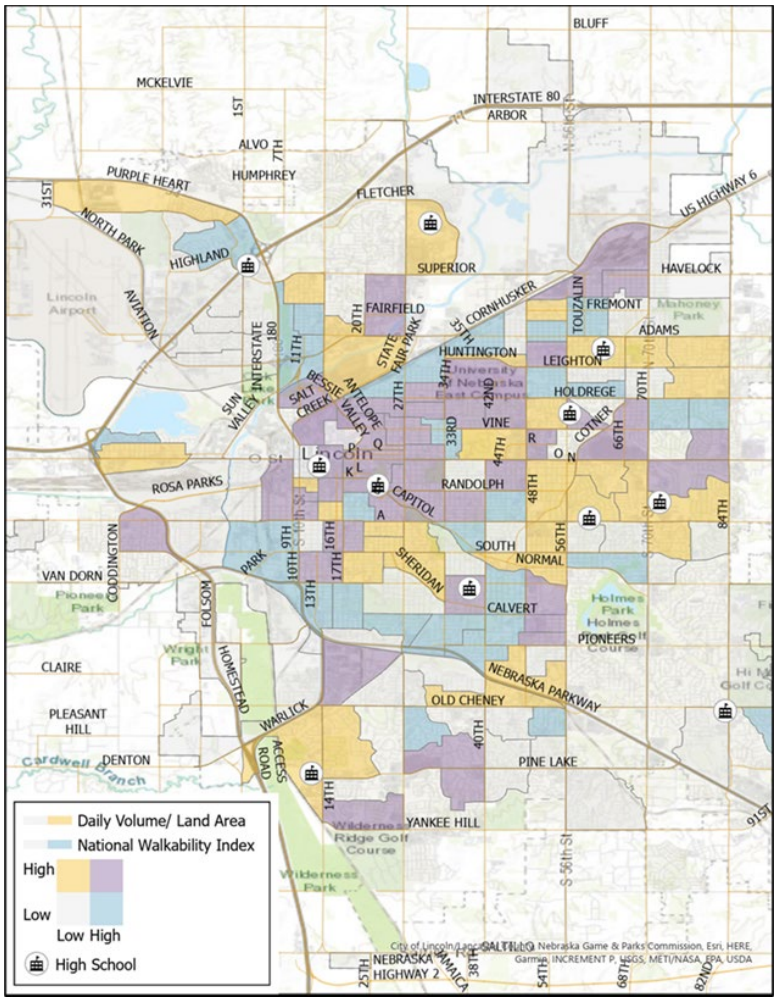

Figure 6.6 Map of EPA National Walkability Index and StL pedestrian volume (by CBG)

6.4 Peak Hour Traffic Volume

We compared morning and evening peak hours using 2019 pedestrian traffic volume data. The morning peak hours from StL were defined as 6-10 am, and evening peak hours as 3-7 pm. We analyzed peak-hour traffic volume and found interesting patterns in pedestrian volume maps based on these hours. During the morning peak hours, most pedestrian traffic volumes were observed around the UNL city campus and the downtown area, likely due to school and business activities during these hours. However, in the evening peak hours, we observed increased pedestrian activity patterns around the Haymarket area, parks around Lincoln, and near

Gateway Mall. This result fits with the likelihood of people conducting recreational activities such as shopping, dining out, and visiting parks during this time period. Notably, the Haymarket area is a popular destination for numerous bars, dining, and entertainment options, attracting many evening visitors. Gateway Mall shows high pedestrian volume during evening peak hours around the year, which confirms that it is a major shopping destination in the city.

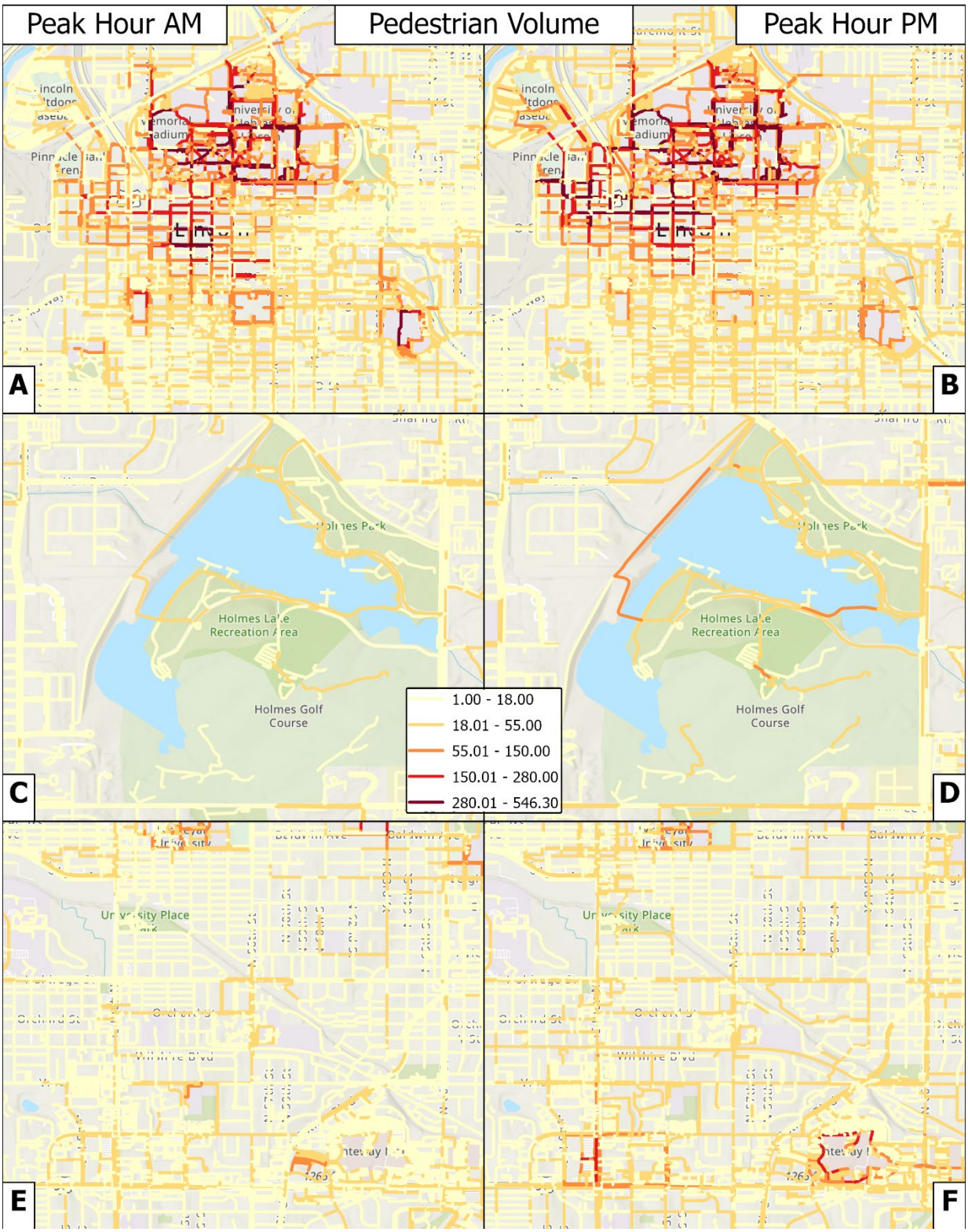

Figure 6.7 Comparing peak AM-peak PM pedestrian volume - CBD, UNL, Haymarket, Holmes Lake Park, shopping malls

6.5 Seasonal Variation

Pedestrian and bicycle traffic volume is impacted by the season (Aultman-Hall et al., 2009). A survey by Sears et al. (2012) found that cyclists prefer not to bike during winter. To confirm these findings and see a detailed seasonal variation of pedestrian and bicycle traffic volume in Lincoln, a segment-scale seasonal comparison was made between winter and summer volumes. We divided the months of the year into seasons as shown in Table 6.1.

Table 6.1 Months by Season

Season	Months
Winter	12, 1, 2
Spring	3, 4, 5
Summer	6, 7, 8
Fall	9, 10, 11

The seasonal comparison showed that overall pedestrian and bicycle volumes are significantly higher during the spring, summer, and fall compared to winter. However, a few places maintained similar traffic volume levels in winter compared to summer. As exemplified in the green highlighted areas in Figure 6.9, the North Sams Club, Gateway Mall, South Pointe mall, Costco, and around Wilderness Hills mall areas did not show much difference between winter and other seasons because, unlike parks and the Haymarket area, most of the shopping facilities are a mix of indoor and outdoor environment. Holiday season shopping may also contribute to the pedestrian volume. The comparison between summer and fall revealed that pedestrian volume is higher in many areas during summer compared to fall. Parks around Lincoln had especially heavy pedestrian traffic for outdoor activities. However, in the fall season, there is a significantly high volume in downtown, UNL, and the Haymarket area, making Lincoln's overall fall volume higher than summer. Figure 6.8 also shows that bicycle volumes

decreased during winter but did not show the same pattern as pedestrian volumes. Reduced bicycle volume may be due to safety concerns during cold weather.

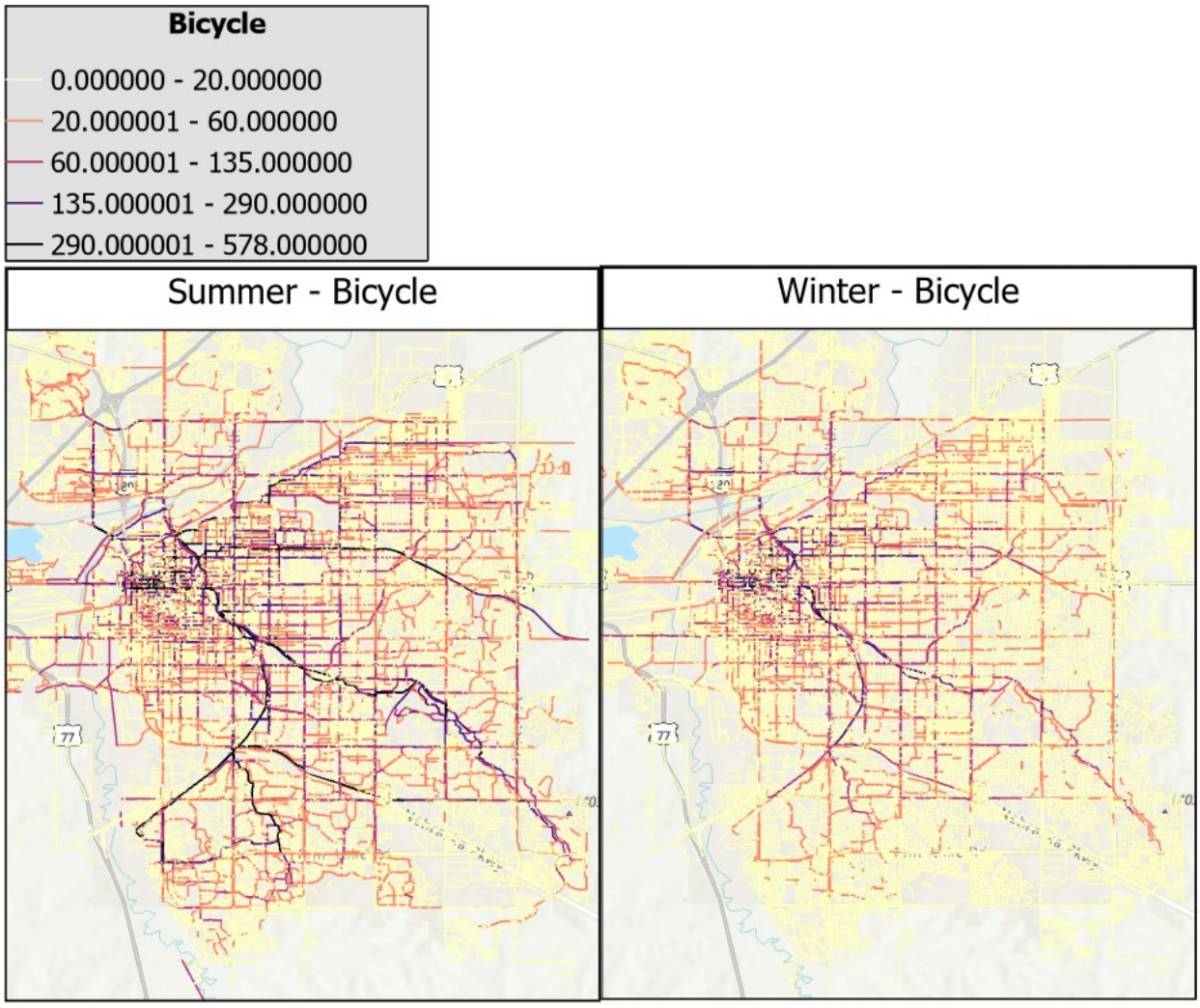

Figure 6.8 Comparison of bicycle volume seasonal variation (summer/winter)

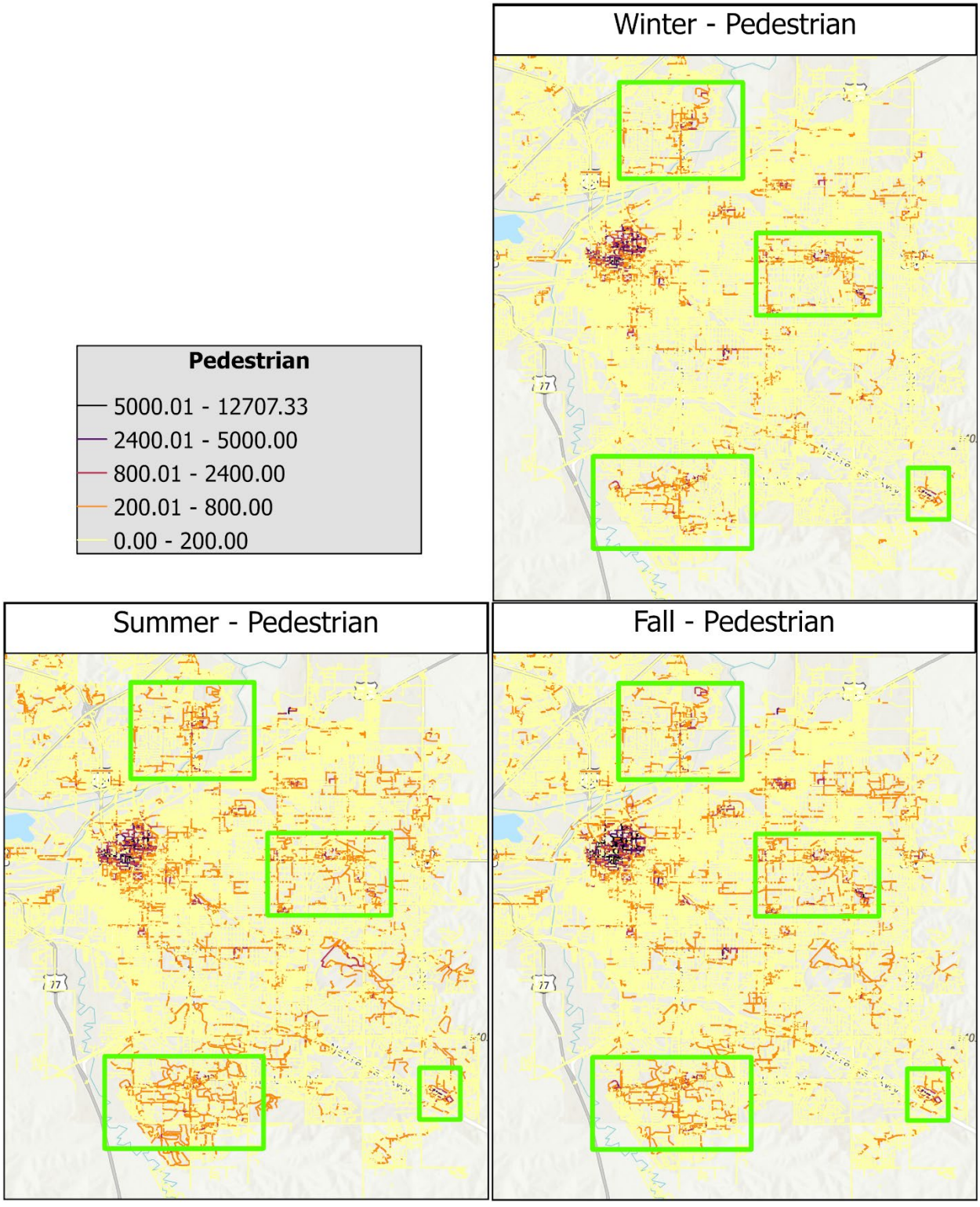

Figure 6.9 Comparison of pedestrian volume seasonal variation (summer/fall/winter)

6.6 Chapter Summary

Chapter 6 involved conducting five different analyses to visualize and analyze spatiotemporal patterns. First, we compared bicycle and pedestrian traffic volume from 2019 and 2020 to see differences during the non-COVID-19 and COVID-19 periods. The analysis revealed changes in traffic volume during 2020, which could be attributed to the closure of public spaces, the need for exercise, and the desire to spend time outside in a safe and socially isolated environment. We also examined map visualization to uncover interesting pedestrian traffic patterns by comparing pedestrian activity volume on weekdays and weekends. Our analysis showed different traffic patterns for weekdays and weekends. Additionally, we created a bivariate map that divided areas into four quadrants based on walkability and daily pedestrian volume. Furthermore, we analyzed peak-hour traffic volume and found that downtown and UNL-city campuses showed high activities during morning peak hours, and there was an increased volume at locations related to recreational activities. Lastly, our analysis of seasonal patterns revealed significantly reduced winter activities for pedestrians and bicycles.

Chapter 7 Conclusion

7.1 Summary of Findings

The project investigated transportation safety of pedestrians and bicyclists, who are disproportionately affected by injuries and fatalities. A key focus was the integration of LBS data to analyze pedestrian and bicyclist traffic volume and patterns. The importance of accurate traffic volume analysis and the limitations of traditional methods were discussed, paving the way for the introduction of LBS data, StreetLight data in particular.

The calibration of StreetLight data indices against traffic volumes was essential to adjust their magnitude and temporal distribution accurately. According to the calibration results, traffic volume stands out as the most crucial variable influencing prediction accuracy across all three models—pedestrian, bicyclist, and vehicle models. The calibration models demonstrated high predictive accuracies. The R-squared values of all three models were greater than 0.5. Overfitting analysis revealed a reasonable model fit, with minimal disparities between training and testing R-squared values for all modes.

The crash analysis procedure was comprised of several components: crash analysis, crash rate analysis, spatial lag model analysis, and relative rank analysis. According to crash analysis results, crash counts were highly correlated with traffic activity (volume). The research also revealed variations in results based on the exposure measure employed to normalize crash counts. While factors like population and network density influenced the safety of non-motorized road users on an aggregate level, the most noteworthy findings appeared when examining infrastructure at the facility scale. The safety ranking analysis identified higher and lower risk areas in the city.

The project also conducted spatial and temporal analysis of traffic volumes for pedestrians and bicyclists at the street segment level. Through a comparative examination of bicyclist and pedestrian traffic volumes between 2019 and 2020, the research unveiled changes in volumes and spatial disparities during the COVID-19 period. The analysis extended to include pedestrian activity patterns on weekdays and weekends, highlighting distinct traffic variations at downtown and recreational locations. Spatial-temporal analysis indicated that downtown and business districts exhibit higher activities during peak hours. Furthermore, the seasonal pattern analysis identified a noteworthy reduction in winter activities for pedestrians and bicyclists.

Overall, this project contributes to the growing body of knowledge aimed at creating a safer environment for non-motorized road users.

7.2 Merits of Project

The project provided the following merits. First, this project contributed to the ongoing enhancement of traffic prediction models, crash risk assessments, and pedestrian/bicycle related designs in a city, as local planners and traffic engineers will have a better understanding of risk exposure patterns of roadway users. The project outcome can be utilized to identify and prioritize locations of higher risk where safety improvements would be most effective.

Second, by analyzing safety at the street segment level, it becomes easier to monitor the effectiveness of implemented safety measures. This allows for ongoing adjustments and improvements based on data and feedback.

Third, the pedestrian and bicyclist travel patterns developed in this model can be used as a foundation for future pedestrian and bicyclist modeling research. The current problem of inadequate pedestrian and bicyclist volume data restricts research and practical analysis on multiple pedestrian and bicyclist topics. The LBS data and development of a quality exposure

model would be a foundational platform to develop further research and practical tools to address these adjacent issues.

7.3 Limitations and Recommendations

While the project yielded promising results, there remains room for improvement in future traffic volume modeling efforts. The abundance of information generated by LBS data facilitates the need for the development of a time and cost-effective data extraction method to enhance the utilization of StreetLight datasets by local planners and traffic engineers. The use of APIs and Python programming, as demonstrated in this project, has proven effective and efficient for such purposes.

It is also noted that LBS data comes with inherent limitations. First, LBS data may be biased towards individuals who use location-aware devices, such as smartphones. This potential bias could lead to the underrepresentation of certain demographic groups or those who do not carry such devices. Second, the accuracy of StreetLight data varies based on the size of traffic volume. Notably, a higher number of cases yield higher accuracy levels in the data. Third, Data accuracy is also subject to the functions of relative travel speeds and transportation network density. These factors can introduce variability in the accuracy of the data. Finally, there is a trend in the transportation data vendor space towards using connected vehicle data. StreetLight Data started transitioning their platform to connected vehicle data during the project period. Connected vehicle data is more accurate and comprehensive than LBS data, but it does not capture the bicycle and pedestrian volumes critical to this project. State DOTs should stay abreast of these changes and understand their implications for multi-modal transportation planning.

Transportation engineers and planners should approach the use of LBS data for transportation safety research with a clear understanding of these limitations. Combining LBS data with other data sources and employing appropriate methodologies can help mitigate some of these challenges and enhance the reliability of safety analyses.

References

Almasi, S. A., Behnood, H. R., & Arvin, R. (2021). Pedestrian Crash Exposure Analysis Using Alternative Geographically Weighted Regression Models. *Journal of Advanced Transportation*, *2021*, e6667688. https://doi.org/10.1155/2021/6667688

Aultman-Hall, L., Lane, D., & Lambert, R. R. (2009). Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record*, *2140*(1), 35–43. https://doi.org/10.3141/2140-04

Bonnell, L. N., Clifton, J., Wingood, M., Gell, N., & Littenberg, B. (2022). *The Relationship Between Mental and Physical Health and Walking During the COVID-19 Pandemic*.

BTS. (2019). *National Transportation Statistics (NTS)*. US Department of Transportation. https://tinyurl.com/rosapntlbtsNTS

Chapman, J., Fox, E. H., Bachman, W., Frank, L. D., Thomas, J., & Rourk Reyes, A. (2021). *Smart Location Database*. Environmental Protection Agency.

Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, *14*, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002

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

Ding, C., Wang, Y., Tang, T., Mishra, S., & Liu, C. (2016). Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice. *Transportation Research Part D: Transport and Environment*. https://doi.org/10.1016/j.trd.2016.08.004

Dong, N., Meng, F., Zhang, J., Wong, S. C., & Xu, P. (2020). Towards activity-based exposure measures in spatial analysis of pedestrian–motor vehicle crashes. *Accident Analysis & Prevention*, *148*, 105777. https://doi.org/10.1016/j.aap.2020.105777

Eldeeb, G., Mohamed, M., & Páez, A. (2021). Built for active travel? Investigating the contextual effects of the built environment on transportation mode choice. *Journal of Transport Geography*, *96*, 103158. https://doi.org/10.1016/j.jtrangeo.2021.103158

Elvik, R. (2013). Can a safety-in-numbers effect and a hazard-in-numbers effect co-exist in the same data? *Accident; Analysis and Prevention*, *60*, 57–63. https://doi.org/10.1016/j.aap.2013.08.010

Elvik, R., & Goel, R. (2019). Safety-in-numbers: An updated meta-analysis of estimates. *Accident Analysis & Prevention*, *129*, 136–147. https://doi.org/10.1016/j.aap.2019.05.019

Exposure to pedestrian crash based on household survey data: Effect of trip purpose. (2019). *Accident Analysis & Prevention*, *128*, 17–24. https://doi.org/10.1016/j.aap.2019.03.017

Federal Highway Administration. (2017). *National Household Travel Survey*. U.S. Department of Transportation. https://nhts.ornl.gov/faq

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781139161879

Geyer, J., Raford, N., Pham, T., & Ragland, D. (2006). Safety in Numbers: Data from Oakland, California. *Transportation Research Record*, *1982*(1). https://journals.sagepub.com/doi/10.1177/0361198106198200119

Gong, Y., & Abdel-Aty, M. (2022). Using machine learning to estimate pedestrian and bicyclist count of intersection by Bluetooth low energy. *Journal of Transportation Engineering, Part A: Systems*, *148*(1), 04021101.

Governors Highway Safety Association. (2019). *Pedestrian Traffic Fatalities by State*.

Guo, Q., Xu, P., Pei, X., Wong, S. C., & Yao, D. (2017). The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. *Accident Analysis & Prevention*, *99*, 114–124. https://doi.org/10.1016/j.aap.2016.11.002

Høye, A. K., & Hesjevoll, I. S. (2020). Traffic volume and crashes and how crash and road characteristics affect their relationship – A meta-analysis. *Accident Analysis & Prevention*, *145*, 105668. https://doi.org/10.1016/j.aap.2020.105668

Iowa DOT. (2020). *Statewide Bicycle and Pedestrian Systemic Safety Analysis 2020*. Iowa DOT.

Islam, A., Mekker, M., & Singleton, P. A. (2022). Examining Pedestrian Crash Frequency, Severity, and Safety in Numbers Using Pedestrian Exposure from Utah Traffic Signal Data. *Journal of Transportation Engineering, Part A: Systems*, *148*(10), 04022084. https://doi.org/10.1061/JTEPBS.0000737

Jacobsen, P. L., Racioppi, F., & Rutter, H. (2009). Who owns the roads? How motorised traffic discourages walking and bicycling. *Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention*, *15*(6), 369–373. https://doi.org/10.1136/ip.2009.022566

Keall, M. D. (1995). Pedestrian exposure to risk of road accident in New Zealand. *Accident Analysis & Prevention*, *27*(5), 729–740. https://doi.org/10.1016/0001-4575(95)00019-V

Kothuri, S., Broach, J., McNeil, N., Hyun, K., & Mattingly, S. (2021). *Exploring Data Fusion Techniques to Estimate Network-Wide Bicycle Volumes*. National Institute for Transportation and Communities. https://nitc.trec.pdx.edu/research/project/1269/Exploring_Data_Fusion_Techniques_to_Estimate_Network-Wide_Bicycle_Volumes

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

Lee, J., & Kockelman, K. (2022). *Prioritizing Investments to Reduce Vehicle-Pedestrian Crashes in Texas*. https://www.caee.utexas.edu/prof/kockelman/public_html/TRB23PedCrashBCRatios.pdf

Lee, K., & Sener, I. N. (2020). Emerging data for pedestrian and bicycle monitoring: Sources and applications. *Transportation Research Interdisciplinary Perspectives*, *4*, 100095. https://doi.org/10.1016/j.trip.2020.100095

Li, H., Wu, D., Graham, D. J., & Sze, N. N. (2020). Comparison of exposure in pedestrian crash analyses: A study based on zonal origin-destination survey data. *Safety Science*, *131*, 104926. https://doi.org/10.1016/j.ssci.2020.104926

Li, Y., & Xu, L. (2021). The Impact of COVID-19 on Pedestrian Flow Patterns in Urban POIs—An Example from Beijing. *ISPRS International Journal of Geo-Information*, *10*(7), Article 7. https://doi.org/10.3390/ijgi10070479

Litman, T. (2011). Transit Price Elasticities and Cross-Elasticities. In *Transportation* (Vol. 7, pp. 37–58). https://doi.org/10.1007/s11116-011-9369-2

Litman, T. (2022). *Understanding Transport Demands and Elasticities*. VTPI.

Liu, Y., & Wu, H. (2017). *Prediction of Road Traffic Congestion Based on Random Forest*. 10th International Symposium on Computational Intelligence and Design (ISCID). https://doi.org/10.1109/ISCID.2017.216

Mahmoud, N., Abdel-Aty, M., Cai, Q., & Zheng, O. (2021). Vulnerable road users' crash hotspot identification on multi-lane arterial roads using estimated exposure and considering context classification. *Accident Analysis & Prevention*, *159*, 106294. https://doi.org/10.1016/j.aap.2021.106294

Martin, A., Goryakin, Y., & Suhrcke, M. (2014). Does active commuting improve psychological wellbeing? Longitudinal evidence from eighteen waves of the British Household Panel Survey. *Preventive Medicine*, *69*, 296–303. https://doi.org/10.1016/j.ypmed.2014.08.023

Merlin, L., Guerra, E., & Dumbaugh, E. (2019). Crash risk, crash exposure, and the built environment: A conceptual review. *Accident Analysis & Prevention*, *134*, 105244. https://doi.org/10.1016/j.aap.2019.07.020

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

Mok, E., Retscher, G., & Wen, C. (2012). Initial test on the use of GPS and sensor data of modern smartphones for vehicle tracking in dense high rise environments. *2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS)*, 1–7. https://doi.org/10.1109/UPINLBS.2012.6409789

Möllers, A., Specht, S., & Wessel, J. (2022). The impact of the Covid-19 pandemic and government intervention on active mobility. *Transportation Research Part A: Policy and Practice*, *165*, 356–375. https://doi.org/10.1016/j.tra.2022.09.007

MRIGlobal, Torbic, D. J., Potts, I. B., Guler, S. I., Gayah, V. V., Harwood, D. W., Grembek, O., Griswold, J. B., Turner, S. A., National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2023). *Pedestrian and Bicycle Safety Performance Functions* (p. 27294). Transportation Research Board. https://doi.org/10.17226/27294

National Association of City Transportation Officials. (2013). *Urban Street Design Guide*. https://nacto.org/publication/urban-street-design-guide/design-controls/design-speed/

Osama, A., & Sayed, T. (2017). Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. *Accident Analysis & Prevention*, *107*, 117–125. https://doi.org/10.1016/j.aap.2017.08.001

Panter, J., Mytton, O., Sharp, S., Brage, S., Cummins, S., Laverty, A. A., Wijndaele, K., & Ogilvie, D. (2018). Using alternatives to the car and risk of all-cause, cardiovascular and cancer mortality. *Heart*, *104*(21), 1749–1755. https://doi.org/10.1136/heartjnl-2017-312699

Pljakić, M., Jovanović, D., & Matović, 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

Pucher, J., Buehler, R., Merom, D., & Bauman, A. (2011). Walking and Cycling in the United States, 2001–2009: Evidence From the National Household Travel Surveys. *American Journal of Public Health*, *101*(S1), S310–S317. https://doi.org/10.2105/AJPH.2010.300067

Raford, N., & Ragland, D. (2006). *Pedestrian Volume Modeling for Traffic Safety and Exposure Analysis: The Case of Boston, Massachusetts*. UC Berkeley Traffic Safety Center. https://escholarship.org/uc/item/61n3s4zr

Rahman, M., & Kockelman, K. (2020). Predicting Pedestrian Crash Occurrences and Injury Severity in Texas. *Presented at the 100th Annual Meeting of the Transportation Research Board*.

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

Raihan, M. A., Alluri, P., Wu, W., & Gan, A. (2019). Estimation of bicycle crash modification factors (CMFs) on urban facilities using zero inflated negative binomial models. *Accident; Analysis and Prevention*, *123*, 303–313. https://doi.org/10.1016/j.aap.2018.12.009

Rajaee, M., Echeverri, B., Zuchowicz, Z., Wiltfang, K., & Lucarelli, J. F. (2021). Socioeconomic and racial disparities of sidewalk quality in a traditional rust belt city. *SSM - Population Health*, *16*, 100975. https://doi.org/10.1016/j.ssmph.2021.100975

Retallack, A. E., & Ostendorf, B. (2020). Relationship Between Traffic Volume and Accident Frequency at Intersections. *International Journal of Environmental Research and Public Health*, *17*(4). https://doi.org/10.3390/ijerph17041393

Saha, D., Alluri, P., Gan, A., & Wu, W. (2018). Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. *Accident Analysis & Prevention*, *118*, 166–177. https://doi.org/10.1016/j.aap.2018.02.014

Sammer, G., Gruber, C., Roeschel, G., Tomschy, R., & Herry, M. (2018). The dilemma of systematic underreporting of travel behavior when conducting travel diary surveys – A meta-analysis and methodological considerations to solve the problem. *Transportation Research Procedia*, *32*, 649–658. https://doi.org/10.1016/j.trpro.2018.10.006

Sarkar, S., Tay, R., & Hunt, J. D. (2011). Logistic Regression Model of Risk of Fatality in Vehicle–Pedestrian Crashes on National Highways in Bangladesh. *Transportation Research Record*, *2264*(1), 128–137. https://doi.org/10.3141/2264-15

Sears, J., Flynn, B. S., Aultman-Hall, L., & Dana, G. S. (2012). To Bike or Not to Bike: Seasonal Factors for Bicycle Commuting. *Transportation Research Record*, *2314*(1), 105–111. https://doi.org/10.3141/2314-14

Sener, I., Munira, S., & Zhang, Y. (2021). *Data Fusion for Non-Motorized Safety Analysis*. Texas A&M Transportation Institute. https://rosap.ntl.bts.gov/view/dot/60229#:~:text=This%20project%20explored%20an%20emerging,applied%20in%20two%20crash%20analyses.

Shamshiripour, A., Rahimi, E., Shabanpour, R., & Mohammadian, A. (Kouros). (2020). How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transportation Research Interdisciplinary Perspectives*, *7*(September), 100216. https://doi.org/10.1016/j.trip.2020.100216

Singleton, P. A., Runa, F., Humagain, P., & Utah State University. Department of Civil and Environmental Engineering. (2020). *Utilizing Archived Traffic Signal Performance Measures for Pedestrian Planning and Analysis* (UT-20.17). https://rosap.ntl.bts.gov/view/dot/54924

StreetLight Data Inc. (2019). *AADT 2018 Methodology and Validation White Paper*. https://learn.streetlightdata.com/aadt-white-paper

Tao, T., Lindsey, G., Cao, J., & Wang, J. (2021). The effects of pedestrian and bicycle exposure on crash risk in Minneapolis. *Journal of Transport and Land Use*, *14*(1), 1187–1208.

Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention*, *50*, 871–878. https://doi.org/10.1016/j.aap.2012.07.022

Toran pour, A., Moridpour, S., Tay, R., & Rajabifard, A. (2017). *Influencing Factors on Vehicle-Pedestrian Crash Severity of School-Aged Pedestrians* (p. 475). https://doi.org/10.2495/UT170401

Turner, S., & Koeneman, P. (2017). *Using Mobile Device Samples to Estimate Traffic Volumes*. Texas A&M Transportation Institute.

Turner, S., Martin, M., Griffin, G., Le, M., Das, S., Wang, R., Dadashova, B., & Li, X. (2020). *Exploring Crowdsourced Monitoring Data for Safety. Exploring Crowdsourced Monitoring Data for Safety*. SAFE-D UTC. https://rosap.ntl.bts.gov/view/dot/50717

Ukkusuri, S., Miranda-Moreno, L., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, *50*. https://doi.org/10.1016/j.ssci.2011.09.012

US EPA, O. (2014). *Smart Location Mapping-National Walkability Index* [Data and Tools]. https://www.epa.gov/smartgrowth/smart-location-mapping

Venter, Z. S., Gundersen, V., Scott, S. L., & Barton, D. N. (2023). Bias and precision of crowdsourced recreational activity data from Strava. *Landscape and Urban Planning*, *232*, 104686. https://doi.org/10.1016/j.landurbplan.2023.104686

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

Wang, Y., Monsere, C., Chen, C., & Wang, H. (2018). Development of a Crash Risk-Scoring Tool for Pedestrian and Bicycle Projects in Oregon. *Transportation Research Record*, *2672*(32), 30–39. https://doi.org/10.1177/0361198118794285

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

Witchayaphong, P., Pravinvongvuth, S., Kanitpong, K., Sano, K., & Horpibulsuk, S. (2020). Influential Factors Affecting Travelers' Mode Choice Behavior on Mass Transit in Bangkok, Thailand. *Sustainability*, *12*, 9522. https://doi.org/10.3390/su12229522

Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis & Prevention*, *75*, 16–25. https://doi.org/10.1016/j.aap.2014.10.020

Yang, H., Cetin, M., Ma, Q., & Virginia Transportation Research Council (VTRC). (2020). *Guidelines for Using StreetLight Data for Planning Tasks* (FHWA/VTRC 20-R23). https://rosap.ntl.bts.gov/view/dot/55501

Yue, L., Abdel-Aty, M., Wu, Y., Zheng, O., & Yuan, J. (2020). In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. *Journal of Safety Research*, *73*, 119–132. https://doi.org/10.1016/j.jsr.2020.02.020

Zhao, C., Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., & Meng, X. (2018). Urban form, demographic and socio-economic correlates of walking, cycling, and e-biking: Evidence from eight neighborhoods in Beijing. *Transport Policy*, *64*, 102–112. https://doi.org/10.1016/j.tranpol.2018.01.018

Zipper, D. (2022, May 26). A City Fights Back Against Heavyweight Cars. *Bloomberg.Com*. https://www.bloomberg.com/news/articles/2022-05-26/a-new-way-to-curb-the-rise-of-oversized-pickups-and-suvs

Appendix A Sample Python Code

```
# zone_activity_analysis
mode = "Bicycle" # change this based on mode

for year in (2019, 2020, 2021, 2022):
    for month in range(1, 13):
        if year != 2022 and month < 10:
            start_month = (str(0)+str(month))
        elif year != 2022 and month >= 10:
            start_month = (str(month))
        elif year == 2022 and month < 5:
            start_month = (str(0)+str(month))
        elif year == 2022 and month >= 5:
            break

        analysis_name = "za_lancaster_co_"+mode+"_"+str(month)+"_"+str(year)

        import requests

        url = "https://insight.streetlightdata.com/api/v2/analyses?key=mSO3R52yrjW08V6nFCzQtJ0C4uyFtobV"

        payload = {
            "analysis_type": "Zone_Activity_Analysis",
            "travel_mode_type": mode,
            "output_type": "volume",
            "day_parts": "All Day|0023,1AM|0101,2AM|0202,3AM|0303,4AM|0404,5AM|0505,6AM|0606,7AM|0707,8AM|0808,9AM|0909,10AM|1010
            "day_types": "Weekday|15,Weekend Day|67,All Days|17,Mo|11,Tu|22,We|33,Th|44,Fr|55,Sa|66,Su|77",
            "insight_login_email": "kyoo2@huskers.unl.edu",
            "analysis_name": analysis_name,
            "oz_sets": [{ "name": "lnk_block_group_zones" }],
            "unit_of_measurement": "miles",
            "geography_type": "blkgrp",
            "traveler_attributes": True,
            "trip_attributes": True,
            "date_ranges": [
                {
                    "start_date": str(start_month)+"/01/"+str(year),
                    "end_date": get_last_date_of_month(year, month)
                }
            ]
        }
        headers = {
            "accept": "application/json",
            "content-type": "application/json"
        }

        response = requests.post(url, json=payload, headers=headers)

        print(response.text)

        sleep(6) # will have [429 Too Many Requests Error] without this
```

Figure A.1 Sample StL Insights API call