Report No. UT-26.08

EVALUATING THE INTERACTION OF THE TRANSIT SYSTEM ON VULNERABLE ROAD USER SAFETY IN UTAH

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Final Report
June 2026

RESEARCH

UDOT Keeping Utah Moving

Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT and the Utah Transit Authority (UTA) on the Technical Advisory Committee for helping to guide the research:

- Travis Evans – UDOT Active Transportation Safety Program Manager
- Heather Siddoway – UDOT Region 2 Traffic Engineer
- Jim Golden – UDOT Transit Division Director
- Philip Castro – UDOT Transportation Technology Project Manager
- Cory Ward – UDOT Region 2 Traffic Engineer
- Travis Jensen – UDOT Research Project Manager (Consultant)
- Travis King – UTA Director of Safety and Security

TECHNICAL REPORT ABSTRACT

<table>
<tr><td>1. Report No.
UT-26.08</td><td>2. Government Accession No.
N/A</td><td colspan="2">3. Recipient's Catalog No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Evaluating the Interaction of the Transit System on Vulnerable Road User Safety in Utah</td><td colspan="2">5. Report Date
June 2026</td></tr>
<tr><td colspan="2">6. Performing Organization Code
GR01768</td></tr>
<tr><td colspan="2">7. Author(s)
Grant G. Schultz, Ph.D., P.E., PTOE; Becky Stevens, EIT; Mara Osumo; Ellie Mills; Greg Snow, Ph.D.</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Brigham Young University
Department of Civil and Construction Engineering
430 Engineering Building
Provo, UT 84602</td><td colspan="2">10. Work Unit No.
5H094 58H</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
25-8354</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Utah Department of Transportation
4501 South 2700 West
P.O. Box 148410
Salt Lake City, UT 84114-8410</td><td colspan="2">13. Type of Report & Period Covered
Final
Sep 2024 to June 2026</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
UT24.310</td></tr>
<tr><td colspan="4">15. Supplementary Notes
Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration</td></tr>
<tr><td colspan="4">16. Abstract
The general safety of vulnerable road users (VRUs) has been shown by some researchers to decrease in the vicinity of transit stops. Previous studies have shown correlations between pedestrian-vehicle crashes and pedestrian generators, specifically transit stops. Furthermore, previous researchers have identified a discrepancy between declining crash frequencies and rising VRU crash frequencies, a trend that helped catalyze the Vision-Zero initiative. Building on previous research that established a spatial correlation between transit locations and pedestrian-involved crashes, this study evaluates the impact of the transit system in Utah on road safety and operational efficiency.
The primary objective of this research is to evaluate, understand, and quantify the interaction of the transit system in the state of Utah as a function of traffic safety, specifically as it relates to VRUs. Geospatial tools, specifically within ArcGIS Pro software, were employed to locate transit stops and analyze surrounding crashes across the state. Crashes were further analyzed using statistical modeling to determine VRU crash probabilities while accounting for surrounding land use (land use classification and intersection involvement), transit stop characteristics (stop location, stop relation to intersection, and transit stop volume), and roadway characteristics (speed limit and roadway volume) variables. This report uses a quantified relationship between transit and VRUs to offer evidence-based recommendations for improving the relationship between public transportation and the physical safety of VRUs in Utah.</td></tr>
<tr><td colspan="2">17. Key Words
Safety, vulnerable road users, pedestrians, transit, transit stops, logistic regression, crashes</td><td>18. Distribution Statement
Not restricted. Available through:
UDOT Research Division
4501 South 2700 West
P.O. Box 148410
Salt Lake City, UT 84114-8410
www.udot.utah.gov/go/research</td><td rowspan="2">23. Registrant's Seal
N/A</td></tr>
<tr><td>19. Security Classification (of this report)
Unclassified</td><td>20. Security Classification (of this page)
Unclassified</td><td>21. No. of Pages
149
22. Price
N/A</td></tr>
</table>

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
AADT	Average Annual Daily Traffic
ADA	Americans with Disabilities Act
ADT	Average Daily Traffic
BRT	Bus Rapid Transit
CAR	Conditional Autoregressive
Cen	Center
CMF	Crash Modification Factors
CSV	Comma-Separated Values
DIC	Deviance Information Criteria
EPDO	Equivalent Property Damage Only
FHWA	Federal Highway Administration
FS	Far Side
GIS	Geographic Information Systems
ICAR	Intrinsic Conditional Autoregressive
kph	Kilometers Per Hour
LPI	Leading Pedestrian Interval
MB	Mid-Block
MCMC	Markov Chain Monte Carlo
mph	Miles Per Hour
NB	Negative Binomial
NB-CAR	Negative Binomial Conditional Autoregressive
NS	Near Side
NTD	National Transit Database
PHB	Pedestrian Hybrid Beacon
ppd	Passengers Per Day
SSI	Safety Stop Index
TSP	Transit Signal Priority
UDOT	Utah Department of Transportation

UGRC	Utah Geospatial Resource Center
UTA	Utah Transit Authority
UTM	Universal Transverse Mercator
VRU	Vulnerable Road User
vpd	Vehicles Per Day
WAIC	Watanabe-Akaike Information Criterion
WFRC	Wasatch Front Regional Council
WGS	World Geodetic System

EXECUTIVE SUMMARY

The Utah Department of Transportation (UDOT) has taken an active role in evaluating and improving traffic safety within the state of Utah. As part of the goal of Zero Fatalities and an interest especially in safety for the most vulnerable road users (VRUs), UDOT funded this research to investigate relationships between VRUs and transit. Traveling by public transit requires a "last mile," which involves a final travel leg to or from the transit stop. Transportation modes in the last mile include some element of walking or bicycle/scooter riding that qualifies the user as a VRU. The "last mile" matters because it includes VRUs' interaction with vehicular traffic (i.e., crossing at a crosswalk or walking alongside of a roadway to reach a transit stop), which increases the chances of VRU-vehicle crash occurrences.

Previous research found a spatial correlation between transit corridors and crashes involving VRUs and an increase in pedestrian crash frequency despite declining vehicular crashes. The objective of this research was to evaluate, understand, and quantify the interaction of the transit system in the state of Utah as a function of traffic safety, specifically as it relates to VRUs. The objective was accomplished by evaluating crash frequencies near transit infrastructure to identify specific causes of VRU crashes and propose specific countermeasures. Relationships between crash, roadway, and transit data were analyzed using geospatial software and logistic regression to control for confounding factors.

The integration of geographic information systems (GIS), visual diagnostics, and statistical modeling indicated that a crash's environmental context (i.e., land use, speed limit, and intersection relationship) was far more influential than simple proximity to a transit stop in predicting VRU-related crash odds. Intersection involvement emerged as the most significant risk factor, with crashes at intersections significantly (36.8 percent) more likely to involve a VRU than those at mid-block.

While spatial modeling suggested that boarding intensity and land use context are primary statistical predictors, their practical effect on crash odds was marginal. Similarly, despite statistical significance, fluctuations in predicted probability remained narrow across speed limit and transit stop location. Even at peak risk levels, the probability of a crash being VRU-related remained low, with a maximum predicted probability of approximately 5.0 percent. Similarly, the physical presence of a transit stop, represented by the specific distance of a crash from that stop, exhibited a marginal effect.

Ultimately, the results suggested that while transit stops are markers for VRU activity, the specific physical and operational characteristics of the stop environment and the existence of an intersection were more predictive of risk than the physical distance between a pedestrian and the transit stop. VRU safety in relation to transit is of relatively low concern.

Future research investigating VRU safety around intersections is highly recommended based on the correlation between VRU crashes and intersection involvement found in this research. General VRU improvements including pedestrian refuge island installation, midblock crossings, pedestrian signals, pedestrian barriers, raised pedestrian crossings, sidewalk installation, and center island installation are recommended as determined locally appropriate.

1.0 INTRODUCTION

1.1 Problem Statement

Pedestrian crashes have been steadily increasing in the U.S., reaching their 28-year peak in 2018 (Ulak et al., 2021); pedestrian fatalities reached a 41-year high in 2022 (NHTSA, 2024). While data has shown that traveling by public transportation is safer than travelling by personal vehicle based on deaths per 100 million passenger miles (National Safety Council, 2024), studies have shown a correlation between transit stops and pedestrian-vehicle crashes (Pulugurtha and Penkey, 2010; Truong and Somenahalli, 2011; Ulak et al., 2021; Voorhees Transportation Center, 2012). The Utah Department of Transportation (UDOT) has taken an active role in evaluating transit stops and safety in relation to vulnerable road users (VRUs) in line with the goal of Zero Fatalities (UDOT, 2026). In this study, VRUs are defined as pedestrians, bicyclists, and users of other micromobility modes. Combined crash, roadway, and transit data along the Wasatch Front were collected to analyze the interaction between transit and VRU crashes to improve VRU safety across the state.

Geographic Information Systems (GIS) tools have been adopted by many agencies, state departments, and researchers spatially analyzing relationships between transit and VRUs (Pulugurtha and Penkey, 2010; Ulak et al., 2021, Voorhees Transportation Center 2012). In this study, GIS, in conjunction with Python coding and a statistical model, were employed to spatially map transit stops and analyze surrounding crashes in the Utah Transit Authority (UTA) service area. Proposed implementable engineering recommendations are included in this report and are intended to assist UDOT in mitigating VRU crashes possibly influenced by the transit system.

1.2 Objectives

The objective of the research is to evaluate, understand, and quantify the interaction of the transit system in the state of Utah as a function of traffic safety, specifically as it relates to crashes involving VRUs. The objective was achieved through a multi-stage methodology involving geospatial data analysis, followed by statistical analysis, and synthesized into implementable engineering countermeasures.

While previous research has established spatial correlations between transit corridors and VRU incidents (Anis et al., 2024; Hess et al., 2004; Pulugurtha and Penkey, 2010; Pulugurtha and Vanapalli, 2008; Singh et al., 2022; Ulak et al., 2021; Yu, 2024), this study contributes to the body of science by introducing a unique, high-resolution buffering technique. Unlike traditional circular buffers, this method captures crashes along the roadway of the transit stop, enabling more precise data associations at greater distances. Furthermore, this research fills a critical gap by simultaneously evaluating additional stop and crash characteristics including speed limit, roadway volumes, transit stop volumes, transit stop location, and transit-stop land use that are rarely integrated into a single predictive statistical model.

1.3 Scope

This study is organized into four distinct tasks:

1. Literature Review: Review literature on the correlation between transit corridors, transit lines and transit stops, roadway characteristics, speed, traffic, and socioeconomic factors regarding pedestrian crashes. Identify any factors that pose an increased risk to road user safety, particularly in relation to transit stops. Investigate GIS and appropriate geospatial statistical methods.
2. Data Collection: Collect data from online resources including the UDOT AASHTOWare crash database, UTA open data portal, and Utah Geospatial Resource Center (UGRC) to establish crash, intersection, and transit stop locations. Other metadata collected includes speed limit, roadway type, crash type, crash date, crash geolocation data, transit stop geolocation, intersection geolocation, roadway volumes, and transit stop volumes.
3. Data Evaluation: Perform geospatial and statistical analysis of collected data. Analyses include intersection distance from crashes and transit stop distance from crashes using ArcGIS Pro, with additional Python analysis for visual comparison of geospatial results. The statistical analysis identifies contributing factors to

determining VRU crash odds and probabilities and reports on the specific impacts those contributing factors have.

4. Provide Conclusions and Recommendations: Based on the findings and performed analyses regarding safety surrounding intersections and transit stops, identify potential safety measures, designs, and specific locations that can be changed or upgraded to improve VRU safety.

1.4 Outline of Report

The report is organized into the following chapters:

- Chapter 1.0 Introduction: Proposes the research problem statement, project objectives, project scope, and outline of the report.
- Chapter 2.0 Literature Review: Includes research from studies investigating the relationship between crash and transit stop frequencies, geospatial statistical methodologies, and contributions to VRU safety.
- Chapter 3.0 Data Collection: Includes data from a variety of public and governmental databases and how the data were then analyzed and used to further the project objective.
- Chapter 4.0 Data Evaluation: Includes information on how the crash data were evaluated, formatted, and used within the scope of the project, as well as how tools such as ArcGIS Pro, Python, and Excel were used to analyze the crash data visually.
- Chapter 5.0 Statistical Analysis: Reviews the statistical results of the research and reports trends and significant factors.
- Chapter 6.0 Conclusions: Summarizes the project methodology, overall findings, and limitations and challenges.
- Chapter 7.0 Recommendations And Implementation: Provides recommendations on how the results of the research could be implemented within the state of Utah.

References and appendices follow the report chapters.

2.0 LITERATURE REVIEW

2.1 Overview

This literature review seeks to summarize the information currently available in the scope of transit and pedestrian interaction safety. The review covers correlations between transit corridors and crashes (both pedestrian and all crashes), transit stops and their safety impacts, data analysis, and a review of previously tested countermeasures and their effectiveness. For clarity, unless explicitly specified as "VRU" crashes, crashes include both vehicular and VRU crashes.

2.2 Correlation Between Transit Corridors and Crashes

The presence of a transit line along a multi-use corridor has been shown to increase crash frequency (Chen and Zhou, 2016; Craig et al., 2019; Guadamuz et al., 2020; Li et al., 2021; Pulugurtha and Vanapalli, 2008; Salum et al., 2024; Ulak et al., 2021). This section explores the relationship between transit corridors and crashes, crashes as a function of speed and roadway characteristics, the effect of traffic and rider volume, and the effect of socioeconomic factors and demographic factors on crashes.

2.2.1 Relationship Between Transit Corridors and Crash Frequency

Chen and Zhou (2016) report that "a higher transit route density and a better transit service correlate with higher crash rates." Researchers in Seattle found that 89 percent of high pedestrian crash locations were within 150 ft. of a transit stop and 90 percent were within 70 ft. of a crosswalk (Craig et al., 2019). This shows a strong correlation between crash frequency and transit stops. It also demonstrates how other factors, like the presence of a crosswalk or intersection, contribute to crash frequency. Ulak et al. (2021) used a Cross K-function to show that pedestrian crash locations are not randomly distributed in proximity to transit stops but are correlated to the existence of a transit stop and that transit stops have an impact on general pedestrian activity.

Researchers in Los Angeles likewise found that 66 percent of their high-pedestrian-crash intersections were close to one or more transit stops (Craig et al., 2019; Loukaitou-Sideris et al., 2007). Walgren (1998) found that high crash intersections were within 150 ft. of a transit stop, while Harwood et al. (2008) found high-pedestrian crash intersections were within 300 meters (1,000 ft.) of a transit stop. While not as strong a connection, the correlation was significant.

Craig et al. (2019) hypothesized that transit vehicles visually block pedestrians, so pedestrian crashes at intersections were theorized to be caused by pedestrians stepping out behind a transit vehicle (making near side (NS) transit stops more dangerous than far side (FS) transit stops). Singleton et al. (2024) found that more pedestrians cross mid-block (MB) near FS transit stops than they do at NS transit stops. Similarly, in Toronto, ON and Charlotte, NC, transit stops within 30 meters (about 100 ft.) of an intersection were found to have an increased risk of pedestrian crashes (Craig et al., 2019). Guadamuz et al. (2020) found that a transit corridor was shown to have a 27 percent increase in expected crash frequency when compared to similar routes with no transit stops. Their non-parametric analysis predicted a 31 percent increase in crashes along a roadway segment with the presence of at least one transit route. However, transit line removal is not so simple. Studies have shown that removing a transit route can be damaging. Guadamuz et al. (2020) predicted a 12 percent increase in crashes if a transit route was removed.

Transit route removal would generally lead to additional average annual daily traffic (AADT), which is predicted to increase crashes (Guadamuz et al., 2020; Li et al., 2021). Chen and Zhou (2016) reported that transit stop density showed a significant correlation to pedestrian crash frequency, but not risk (Chen and Zhou, 2016), suggesting that pedestrian crashes around transit stops may be a coincidental function of pedestrian volumes around transit stops. Kim et al. (2010) agree with this, stating that transit stops are more often found to be "guilty by association," in terms of pedestrian crashes. This means that these crashes likely would have happened with or without the transit stops.

Transit corridors may be a function of pedestrian traffic volumes as a trip generator and attractor but may not necessarily cause pedestrian crashes. Zegeer et al. (2001) and Pulugurtha and Vanapalli (2008) found that higher pedestrian volumes correlated to higher pedestrian crashes. Miranda-Moreno et al. (2011) also found a significant positive correlation between pedestrian volume and collision frequency. Ulak et al. (2021) reported that the number of generic crashes has steadily declined in Palm Beach County, FL since 1990, but pedestrian crashes have steadily increased during that time, reaching their peak in 2018. Additionally, Ulak et al. (2021) found there is "a significant spatio-statistical correlation between the transit stop locations and pedestrian-involved crashes."

2.2.2 Crashes as a Function of Speed and Roadway Characteristics

Several researchers have found a correlation between roadway speed and crash frequency (Chen and Zhou, 2016; Craig et al., 2019; FHWA, 2008b; Li et al., 2021; Phillips et al., 2021) as well as pedestrian crash severity increasing with speed (Miranda-Moreno et al., 2011). Pedestrian crash risk has been shown to increase with speed, the number of lanes, and road width (Craig et al., 2019; Ukkusuri et al., 2012). In a study on enforcement and yielding percentages, Craig et al. (2019) found that the percentage of vehicles yielding to VRUs decreased when a transit stop was present. Similarly, the presence of a transit stop near an unsignalized, marked crosswalk was associated with lower yielding rates (Craig et al., 2019).

Other roadway characteristics seemed to contribute to increased crashes. In Philadelphia and Pittsburgh, PA, the installation of barriers in the roadway demonstrated a statistically significant decrease in crash frequency, though results vary depending on barrier type and traffic volume (Guadamuz et al., 2020). This shows that open roads lacking protection or signage warning drivers of the presence of VRUs can lead to an increase in crashes involving VRUs.

Yang (2007) found that more than 80 percent of major transit collisions reported in the National Transit Database (NTD) occurred at or near intersections controlled by signal or stop signs. The NTD is the primary database maintained by United States Department of Transportation Federal Transit Administration. They argued that since most transit routes run through densely populated urban areas where many intersections have some traffic control measures in place, it is expected that a high percentage of transit collisions would occur at or near traffic signals and stop signs (Yang, 2007).

Further, urban intersections with pedestrians crossing streets and vehicles maneuvering in and out of lanes are likely to increase the probability of being involved in crashes. Transit stops are frequently located near intersections where a stopped transit vehicle in a traffic lane may cause a crash as cars maneuvering around the transit vehicle are visually blocked from turning vehicles and pedestrians (Phillips et al., 2021).

Sidewalk presence has shown to have a positive effect on VRU safety. Chen and Zhou (2016) found that areas with more sidewalks and higher pedestrian volumes had fewer crashes when compared with the opposite conditions. Chen and Zhou (2016) also found that sidewalk density and the proportion of areas with steep grades are negatively correlated to each other, concluding that pedestrian volumes are low in areas with steep grades. A negative association with sidewalks suggests lower pedestrian volumes. The authors theorized that areas without sidewalks dissuaded pedestrians from walking there (Chen and Zhou, 2016). Supporting the same argument about pedestrian safety and pedestrian volumes, sidewalk accessibility and connectivity were identified as factors that enhance a transit user's experience and accessibility (Brodeur-Ouimet and Cloutier, 2024; FHWA, 2008a) as roads without a sidewalk are 1.73 times more hazardous than roads with sidewalks (Barabino et al., 2021).

Several studies have analyzed the effect crosswalks and various crosswalk types had on VRU safety. For multi-lane roads, marked crosswalks with no other enhancements (e.g., warning lights, yield signs, etc.) were shown to have a significantly higher crash rate per pedestrian (Craig et al., 2019) when compared to unmarked crosswalks (Zegeer et al., 2001; Zegeer et al., 2004). On two-lane, low volume roads, unsignalized, marked crosswalks had no difference in crash rate compared to an unmarked crosswalk (Zegeer et al., 2001; Zegeer et al., 2004), which is consistent with other findings that marked crosswalks on high volume roads (AADT greater than 10,000 vehicles per day (vpd)) have much higher crash rates than at unmarked crosswalks (Zegeer et al., 2001).

Other countries (e.g., Germany, Finland, and Norway) prohibit uncontrolled crosswalks on roads with speed limits higher than 40 miles per hour (mph) for VRU safety interests (Zegeer et al., 2001) and are not recommended for roads with high speeds (Zegeer et al., 2004). Unsignalized, marked crosswalks near transit stops were associated with low yield rates (Craig et al., 2019), further showing that a crosswalk alone is not sufficient protection for pedestrians crossing there. Additional pedestrian safety protection measures (e.g., bulb outs at crosswalks, crosswalk signs, advanced stop lines in front of a crosswalk, pedestrian refuge islands, advanced warning signs/lights, pedestrian hybrid beacons, etc.) may be necessary to protect VRUs crossing streets, especially near transit stops (Craig et al., 2019; Fitzpatrick et al., 2021; Zegeer et al., 2004).

The relationship between crosswalk density and pedestrian crashes is less straightforward. High crosswalk densities were significantly and positively correlated with pedestrian crashes (Chen and Zhou, 2016). In Hawaii, only 38 percent of pedestrian crashes occurred at crosswalk locations, suggesting that the majority of crashes occur when a pedestrian crosses at a MB location or at an intersection without a crosswalk (Kim et al., 2010). A study in Seattle found that roughly 80 percent of high pedestrian crash locations were within 70 ft. of a crosswalk (Craig et al., 2019). A 1972 study found an incidence of increased pedestrian-involved collisions in marked crosswalks when compared with unmarked crosswalks (Zegeer et al., 2001).

Related to walkways, it is also reported that the most walkable transit stops are closer to major roads, have a higher road density, and high collisions per road length. In a study conducted by Brodeur-Ouimet and Cloutier (2024), the combination of these factors was termed road risk. Further, that same study concludes that stops located closer to central neighborhoods also have an increased road risk. This may be due to higher walkability stops having higher crosswalk densities and ultimately having more crashes.

2.2.3 Effect of Traffic and Rider Volume

Vehicular and transit passenger volumes appear to play a role in VRU crash frequency as well. The literature suggests that AADT and rider volumes are directly correlated to higher numbers of crashes (Craig et al., 2019; Li et al., 2021; Miranda-Moreno et al., 2011; Phillips et al., 2021), especially at intersections (Miranda-Moreno et al., 2011). Craig et al. (2019) found that a high Average Daily Traffic (ADT) led to poorer yielding for public transit vehicles. Miranda-Moreno et al. (2011) found that traffic volume was the primary cause of pedestrian collision frequency at intersections. The researchers concluded that a reduction in traffic volume would positively contribute to improvements in pedestrian safety (Miranda-Moreno et al., 2011).

Craig et al. (2019) and Zegeer et al., (2001) found that higher pedestrian volumes in an area correlated to higher pedestrian crashes. Volume increases (in both vehicle and pedestrian traffic) correlate to higher crash rates. Miranda-Moreno et al. (2011) further found that the relationship between pedestrian volumes and crash rates is not strictly linear (i.e., as the number of pedestrians increases, the risk faced by each pedestrian increases to a certain pedestrian volume). After reaching a certain pedestrian volume threshold, crash risk may begin to fall, supporting a theory of safety in numbers (Jacobsen, 2003; Miranda-Moreno et al., 2011).

2.2.4 Effect of Socioeconomic and Demographic Factors

Socioeconomic and demographic factors, including land use in connection to transit user crashes, are also important to consider. Four- and five-way intersections have been shown to be positively correlated with pedestrian crash frequency (Chen and Zhou, 2016; Ukkusuri et al., 2012). Intersections in general have been shown to be prone to pedestrian crashes due, in part, to the number of potential conflict points. Transit stop density has been shown to be correlated with high-pedestrian-crash frequency, as are high speed areas (i.e., arterials), population density, intersection density, pedestrian volume, commercial and office land uses, and signal densities. Traffic volume or vehicle miles traveled were shown to be positively correlated with the number of pedestrian crashes. School densities showed a positive relationship with the number of crashes but higher proximity to each school showed a negative correlation with pedestrian crashes. Alternatively, sidewalk density had a negative correlation with pedestrian crashes, meaning fewer crashes occurred in areas with a sidewalk, implying that an increase in sidewalk density can improve pedestrian safety. The total number of trips is positively correlated with pedestrian crash frequency, but negatively associated with crash risk (i.e., number of pedestrian crashes/forecasted number of trips) (Chen and Zhou, 2016).

Industrial and mixed land uses yielded mixed results in terms of pedestrian safety. Chen and Zhou (2016) and Miranda-Moreno et al. (2011) found a low or negative correlation between industrial areas and pedestrian crashes, hypothesizing that industrial areas have a low pedestrian volume because the environment in industrial areas is not pedestrian friendly. Ukkusuri et al. (2012) found a positive association between pedestrian crashes and industrial, open space, and commercial areas. The researchers found that areas with a greater percentage of industrial and commercial land uses had a higher likelihood of fatal pedestrian crashes. In other words, even if industrial and commercial areas did not contribute a significant number of crashes, they likely contribute to high severity crashes. However, Pulugurtha and Sambhara (2011) found that urban residential commercial and neighborhood transit services lowered pedestrian crashes. Ukkusuri et al. (2012) also found a positive correlation between transit volume (i.e., ridership) and the number of pedestrian crashes. The higher the ridership or number of subway stations, the higher the pedestrian crashes (Ukkusuri et al., 2012).

Employment density has also yielded mixed results (Chen and Zhou, 2016; Pulugurtha and Sambhara, 2011; Singh et al., 2022). Researchers have found that population density (Kim et al., 2010; Moudon et al., 2011; Pulugurtha and Sambhara, 2011; Singh et al., 2022; Ukkusuri et al., 2012; Wang and Kockelman, 2013; Wier et al., 2009), employment density (Kim et al., 2010; Pulugurtha and Sambhara, 2011; Singh et al., 2022), and transit stop density (Singh et al., 2022) contributed to pedestrian or bicycle crash rates, suggesting that an increase in population volume, employment, or transit stop densities leads to more crashes. Further, population density, employment density, and availability of intermodal connectivity positively affect transit ridership (Ulak et al., 2021) and are positively and significantly related (Kim et al., 2010). However, Chen and Zhou (2016) found that areas with a higher employment density have a lower pedestrian crash risk. Miranda-Moreno et al. (2011) found that population density, commercial land use, number of jobs, number of schools, presence of metro station, number of transit stops, percentage of major arterials, and average street length all have a powerful association with pedestrian activity; furthermore, as pedestrian volume increases, crash risks increase.

Ulak et al. (2021) found that the average walking distance to a transit stop is about 0.5 miles (2,640 ft.), and pedestrians in crashes related to a transit stop are likely within 0.25-1.88 miles (1,320 – 10,000 ft.) of a transit stop. This supports the assertion that higher population and employment density lead to higher transit stop densities, which contribute to higher pedestrian crash rates because of opportunity and exposure. The researchers also found a cluster of pedestrian fatalities in areas with the highest population densities, especially those areas of Palm Beach, FL that had high transit density, low income, high labor force density, and zero-vehicle areas. The percentage of workers commuting by walking was not considered because it corresponds to only 1 percent of total trips made in the area. However, Chen and Zhou (2016) asserted that high traffic and pedestrian volumes increase pedestrian safety because drivers are expecting to see a pedestrian. Doubling pedestrian volume in an area showed only a 52 percent increase in pedestrian-vehicle crashes (Geyer et al., 2006), which is consistent with other studies that found that the relationship between pedestrian volumes and pedestrian-vehicles crashes was not strictly linear (Jacobsen, 2003; Miranda-Moreno et al., 2011), and that pedestrian risk decreased (after a certain point) with increasing pedestrian and traffic volumes (Leden, 2002). Estimated block length appeared to have a negative influence on bicycle-involved crashes, possibly because bicyclists who travel farther distances have increased environmental adaptability and better reaction time (Singh et al., 2022).

Miranda-Moreno et al. (2011) found that land use near an intersection had a strong correlation with pedestrian activity, but a small direct effect on pedestrian crash rates, concluding that strategies that encourage or increase pedestrian volumes (e.g., mixed land use and transit stops) without other safety measures contribute to pedestrian crash increases. They found that the number of motor vehicles entering an intersection is a primary determinant of pedestrian-vehicle crashes (Miranda-Moreno et al., 2011).

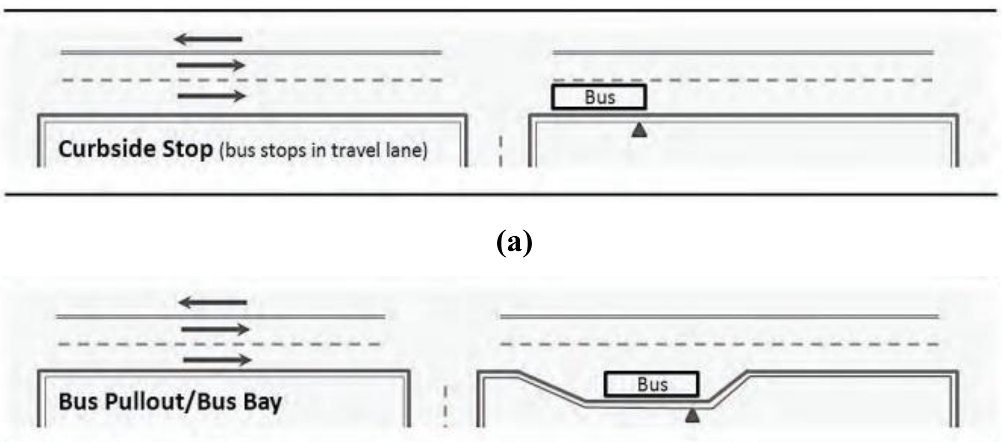
2.3 Transit Stops and Their Safety Impacts

Transit stops are one of the key factors relating crashes and transit. Transit stops aid VRUs in their transit experience and can also influence their safety. This section investigates transit-stop safety features and their impacts, and the Safety Stop Index (SSI) that is used to rank the transit stops for safety.

2.3.1 Transit Stops Safety Features and Impact

The Federal Highway Administration (FHWA) states that benches, shelters, and lighting at and around transit stops are important for comfort and safety (FHWA, 2008a). Moreover, a survey of passengers using the Kowloon Motor Bus Company Limited in Hong Kong, China found that a poor transit stop environment (i.e., lack of shelter, seats, and protection facilities) contributed to a higher level of transportation stress (Loo and Tsoi, 2024). The study also mentioned that the number of vehicle transfers and travel time are other significant stressors.

There are also many ways a transit stop can be built. Phillips et al. (2021) specifically investigates pull-out and curbside stops, like those shown in Figure 2.1, and how that distinction affects collision rates. Contrary to the authors' prediction that pull-out stops are safer than curbside stops, it was found that transit stop designs do not strongly influence collision rates. Despite the higher risk of nearby collisions at curbside stops, the authors concluded this is likely because they are often located near junctions and side roads, where collisions are more common, rather than the specific design (Phillips et al.,, 2021).



(b)

Figure 2.1 (a) Curbside and (b) pull-out stops (Larimer County Urban Area Street Standards 2015)

Dwell time for a transit vehicle stopped in a traffic lane has also been shown to increase crash risk. Crashes are prone to happen for a stopped transit vehicle when dwell time exceeds 30 seconds. Transit vehicles with long dwell times are recommended to stop outside of the travel lane (Zhou and Bromfield, 2007).

<u>2.3.2 Safety Stop Index</u>

Several analyses used the SSI to rank transit stops by their relative safety (Salum et al., 2024; Ulak et al., 2021). Ulak et al. (2021) ranked transit stops on an SSI scale using KABCO severity levels. KABCO is a crash severity scale that indicates the severity of the crash based on the crash effect on the people involved (AASHTO, 2010). The American Association of State Highway and Transportation Officials (AASHTO) definitions are outlined in Table 2.1. Ulak et al. (2021) used KABCO ratings to give each transit stop an SSI value based on weighted values for each KABCO severity value, similar to the UDOT values shown in Table 2.2 (UDOT, 2021). Table 2.2 converts crash costs to equivalent property damage only (EPDO) rankings.

Table 2.1 KABCO and UDOT Crash Severity Scales (AASHTO, 2010)

KABCO	UDOT	Severity
K	5	Fatal Injury
A	4	Suspected Serious Injury
B	3	Suspected Minor Injury
C	2	Possible Injury
O	1	Property Damage Only

Table 2.2 Updated Crash Costs for Use in Benefit Calculations (2020 Dollars) (UDOT, 2021)

Severity	Severity No.	Crash Cost Weighted	Crash Cost Unweighted	Unweighted EPDO
K (Fatal)	5	$3,078,500	$14,010,300	828
A (Sus. Serious Injury)	4	$3,078,500	$805,800	47
B (Sus. Minor Injury)	3	$264,000	$264,200	16
C (Possible Injury)	2	$148,000	$148,000	9
O (PDO)	1	$17,000	$17,000	1
KA (Severe)	5,4	$3,078,500	N/A	182
KAB (Injury)	5,4,3	$806,700	N/A	48
KABC (Anticipated Injury)	5,4,3,2	$415,100	N/A	24
KABCO (Total Crashes)	5,4,3,2,1	$134,400	N/A	8

In subsequent spatial and mathematical analyses, Ulak et al. (2021) found KABCO scoring of transit stops to be helpful in determining high crash areas, especially for transit stops with high VRU fatality rates. In their area of study (Palm Beach, FL), they found fatalities were congregated in areas of low-income and high population density. Salum et al. (2024) formulated a similar program, creating safety scores to rank and prioritize transit stops based on a numerical safety scoring system. Salum et al. (2024) weighted several safety factors to rank the 14 most dangerous transit stops in each municipality in the study area (three area transportation authorities in Florida). Ulak et al. (2021) took a spatial approach to the problem while Salum et al. (2024) took a numerical approach.

SSIs can be formulated to be user friendly for organizations that lack resources to conduct formal safety analyses. Ulak et al. (2021) formulated a regression tree model that can be used by practitioners who do not have the resources to conduct a formal SSI study. Salum et al. (2024) provided safety criteria to help agencies prioritize transit stops by safety features. An SSI analysis can be an effective measurement tool to monetarily quantify crash data impacts and prioritize crash-prone areas.

2.4 Data Analysis

Studying pedestrian-transit crash relations includes both spatial and numeric analyses. GIS programs (especially ArcGIS) can be used to perform spatial analysis. This section will outline relevant GIS applications and workflows for analysis found in the literature, followed by spatial analysis tools and crash radius parameters used to study crashes. Then various statistical approaches for both transit-specific and general safety analyses are presented.

2.4.1 Geographic Information Systems Applications and Workflows for Analysis

GIS analysis for spatial determination of the relationship between transit stops and VRU crashes depends on the desired results and the available data. Although some differences exist across studies depending on their parameters, most of the literature follows a general flow similar to the one used by Pulugurtha and Vanapalli (2008):

1. Geocode auto-pedestrian collision data
2. Create a crash concentration map
3. Overlay transit stop locations
4. Extract the number of collisions per transit stop using the "Spatial Join" tool
5. Identify traffic volumes and obtain ridership data
6. Compute collision frequency, crash rates, and rank high-collision transit stops

Ulak et al. (2021) involved KABCO severity types and distances from transit stops in their analysis using the following process:

1. Classify crashes by severity type
2. Overlay crashes on transit stops
3. Calculate distances from crashes to transit stops using shortest path network distance
4. Determine a numerical weighting system for severity scale

5. Calculate SSI values for each crash severity type
6. Assign SSI values to each transit stop and demonstrate spatially

2.4.2 Spatial Analysis Tools

The following GIS or spatial analysis tools were used to demonstrate spatial relationships between known crash locations and transit locations. "Kernel density" analysis can be used to create heat maps or crash clusters to spatially demonstrate crash densities at various intersections (Pulugurtha and Vanapalli, 2008; Truong and Somenahalli, 2011). Ulak et al. (2021) used a decay function to find an average distance that pedestrians would walk to reach a transit stop. The "buffer" tool in ArcGIS (or something interchangeable) is commonly used to limit crashes to those within a specified vicinity of a transit stop (Brodeur-Ouimet and Cloutier, 2024; Kim et al., 2010; Li et al., 2021; Phillips et al., 2021; Pulugurtha and Vanapalli, 2008; Pulugurtha and Sambhara, 2011; Salum et al., 2024; Singh et al., 2022; Ulak et al., 2021). Buffers can be used in conjunction with the "spatial join" tool to connect numerical data to geographical features (e.g., connecting the number of crashes within a transit stop radius to that transit stop) (Pulugurtha et al., 2008; Salum et al., 2024). Barabino et al. (2021) also optimized Google maps to calculate transit stop lengths.

Other approaches to spatial correlation are the K-function and G-function which use data clustering to find crash clusters. K- and G-functions are point pattern analysis tools that detect crash clusters. Global Moran's I statistical analysis can be used in conjunction with the K- and G-functions. Moran's I is an analysis that measures multi-dimensional spatial relationships based on point clusters. It is scale-dependent because of the way the analysis is approached while K- and G-functions can be used on data at a large scale (Habib et al., 2023). Similarly, a Cross K-function can be used to formulate spatio-statistical relationships between known points (e.g., crash locations) (Okabe et al., 2006; Okabe and Sugihara, 2012; Ulak et al., 2021). While K- and G-functions simply analyze point distribution patterns, a Cross K-function would draw relationships between the points (Habib et al., 2023). The ArcGIS Pro "incremental spatial correlation" tool is the only tool available in ArcGIS able to calculate a threshold distance, but it is limited in its size capacity because it uses Global Moran's I (Habib et al., 2023). The "generate network spatial weights" ArcGIS tool can be used to generate spatial weights based on threshold distances. The “incremental spatial autocorrelation” tool can be used to define threshold distances. Spatial weights and threshold distances can be used in the “Getis-Ord Gi* hot spot analysis” as an alternative to the Global Moran’s I tool to generate crash clusters (Habib et al., 2023).

2.4.3 Crash Radius Parameters

An important part of the GIS analysis is using the right parameters. Analysis of crashes related to transit stops includes defining a buffer zone to aid in determining if the pedestrian crashes were related to the transit stop and identifying which transit stops have the most associated pedestrian crashes. A radius of 100-200 ft. from the transit stop is generally accepted (Pulugurtha and Vanapalli, 2008; Salum et al., 2024), however, a wide range of values have been used in the literature. Salum et al. (2024) and Craig et al. (2019) used a 100 ft. radius, Pulugurtha and Vanapalli (2008) used radii of 100 and 200 ft. but determined that radii of 100-500 ft. were acceptable. Li et al. (2021) used buffers of 100 and 200 meters (330 ft. and 655 ft.) because they found that buffer sizes larger than 200 meters (655 ft.) caused the correlations between transit-critical events (i.e., hard braking, hard deceleration, or long dwell time) and crashes at the transit stop to decrease. Ulak et.al (2021) proposed that the average walking distance to a transit stop was about 0.5 miles (2,640 ft.), and no more than 2 miles (10,560 ft.), so an appropriate buffer would be to filter by pedestrian crashes within 0.5 miles (2,640 ft.) of a transit stop. In other words, a pedestrian crash more than 0.5 miles (2,640 ft.) away from a transit stop is unlikely to be associated with the transit presence. Singh et al. (2022) also used a 0.5-mile (2,640 ft.) buffer.

Ulak et al. (2021) proposed a 600 meter (2,000 ft.) radius was an appropriate measure for subways. Pulugurtha and Vanapalli (2008) argued that using a buffer radius of more than 100 ft. may complicate the results because it has the potential to include pedestrian crashes of pedestrians not involved with transit and may overlap with data from other transit stops (essentially double counting crash data). In the City of Las Vegas, NV, most non-MB transit stops are within 150 ft. of an intersection, so a buffer of more than 100 ft. would likely intersect with another transit stop. Furthermore, the researchers found that 95 percent of pedestrian trips within 100 ft. of a transit stop are transit related, implying that 95 percent of the pedestrian crashes within a 100-foot radius of a transit stop would be transit related. Phillips et al. (2021)

used a multi-ring buffer to make buffers of 10, 20, 30, 40, 50, and 60 meters (35, 70, 100, 130, 165, and 200 ft., respectively). Brodeur-Ouimet and Cloutier (2024) used a buffer of 500 meters (1,640 ft., close to 0.25 miles) in a high population density major city and 800 meters (2,625 ft., close to 0.5 miles) in other analysis areas. Judgment should be used to apply an appropriate buffer for the data involved in the analysis.

Other parameters used included ArcGIS cell sizes. Pulugurtha and Vanapalli (2008) used a cell size of 0.25 mile (1,320 ft.). No other resources were found related to ArcGIS cell sizes.

2.4.4 Statistical Approaches Used

A variety of statistical approaches have been used in the literature to analyze VRU-involved crash data relationships, including Spearman's rank correlation analysis, negative binomial (NB) distribution, various Bayesian analyses, Bayesian goodness-of-fit measures, Markov Chain Monte Carlo (MCMC) simulation, Cross K-Function, and bivariate risk model. Li et al. (2021) used Spearman's rank correlation coefficient to examine correlations between their defined transit critical events (i.e., hard braking, hard acceleration, and transit-vehicle dwell time) and crashes around transit stops. Spearman's rank correlation is generally accepted in the literature as a method for quantifying the relationship between crashes and surrogate safety factors (Li et al., 2021; Strauss et al., 2017). The Spearman's rank correlation estimate is useful for determining how strongly related other safety factors are to crash rates. Analysis using Spearman's rank correlation can help determine crash causation so effective countermeasures can be implemented.

A simple NB distribution can be used in crash data to evaluate crash density. If crash data follows a Poisson distribution, distribution characteristics can be analyzed with a probability function to determine crash clusters (i.e., high-risk crash areas) (Cui et al., 2022).

Bayesian models are commonly used in highway safety analyses (Aguero-Valverde, 2013; Ma et al., 2008; Park and Lord, 2007; Wang and Kockelman, 2013) to account for unseen differences or unknown spatial correlations. A full Bayesian before-after evaluation can be used to evaluate treatment effectiveness using before-and-after data (Ali et al., 2021; Kodi et al., 2021). This evaluation method was specifically used to evaluate safety effectiveness for transit signal priority (TSP) and signal timing changes by developing crash modification factors (CMFs) which quantified the expected changes in crash rates (Ali et al., 2021; Kodi et al., 2021). CMF values determined the safety effectiveness of a TSP system (Ali et al., 2021). Log-normal models have been implemented more recently than other statistical analyses in highway safety and are typically employed using a Bayesian approach (Aguero-Valverde, 2013; Wang and Kockelman, 2013). Empirical Bayes Methods can be used as a crash cluster estimation that measures severity of crashes (Elvik, 2008; Li and Zhang, 2007; Truong and Somenahalli, 2011).

Various forms of Bayesian conditional autoregressive (CAR) models are supported in the literature (Chen and Zhou, 2016; Singh et al., 2022; Wang and Kockelman, 2013). Li et al. (2021) used a Bayesian NB-CAR model to account for the spatial correlation between neighboring transit stops and their proximity to each other. The NB-CAR also accounted for outliers by relaxing the mean and variance. Chen and Zhou (2016) used a Bayesian hierarchical intrinsic CAR (ICAR) model to evaluate crash frequency and risk.

Goodness-of-fit measures can be used to quantify how well a Bayesian model matches the data. The Bayesian NB-CAR model performance was evaluated using two goodness-of-fit measures: Watanabe-Akaike Information Criterion (WAIC) and Deviance Information Criteria (DIC) (Li et al., 2021). WAIC is a Bayesian approach, used in the Li et al. (2021) study to estimate error prediction (Gelman et al., 2014). DIC was used to show variation in the data, which comments on model complexity (Li et al., 2021). Li et al. (2021) used an MCMC simulation to estimate the model parameters in preparation for using Bayesian NB and NB-CAR models. Ali et al. (2021) also used an MCMC simulation to collaborate parameters in a Poisson log-normal model used in their study to analyze crash counts on TSP-treated corridors. Li and Zhang (2007) used an MCMC analysis, to derive spatial parameters, account for errors, and determine spatially correlated effects. MCMC is popular in highway safety analysis (Ali et al., 2021; Li and Zhang, 2007,;Wang and Kockelman, 2013).

A Cross K-function is another method for formulating spatio-statistical relationships (Okabe et al., 2006; Okabe and Sugihara, 2012; Ulak et al., 2021). One study used a Cross K-function to determine if pedestrian crashes were statistically independent from the spatial distribution of transit stops. The researchers found that pedestrian crashes are spatially correlated with transit stops (Ulak et al., 2021).

A bivariate risk model can be used to relate crash risks to crash factors. Barabino et al. (2021) used a bivariate risk model focusing on the frequency and severity of the crashes to determine the crash risk of transit routes. In addition to those factors, an exposure factor, determined by the number of people or services that may be affected by the occurrence of the crashes, was also considered in this calculation. It was then followed by logical regression, using the aforementioned factors. Singh et al. (2022) used a bivariate spatial model to find the factors that influence bicyclist-involved crashes. Specifically, joint models based on the multivariate conditional autoregressive predictions with distance-oriented neighboring weight matrix were used.

2.5 Countermeasures

Several studies have been conducted in the literature to evaluate the effectiveness of countermeasures to proactively increase safety. These proposed countermeasures fall into the general categories of roadway design, education and enforcement measures, and roadway features.

2.5.1 Roadway Design

Roadway design has been shown to have significant impacts on pedestrian crash rates and severity (Miranda-Moreno et al., 2011). Design elements including speed and road width, sidewalk presence, and roundabouts play a role in pedestrian safety and crash rates.

Multiple studies have noted a strong association between increased pedestrian crash risk and wider, higher speed roadways (Brodeur-Ouimet and Cloutier, 2024; Craig et al., 2019; Kim et al., 2010). The presence of sidewalks is correlated with a decrease in crash severity (Barabino et al., 2021). Specifically, Barabino et al. (2021) calculated that roads without a sidewalk are 1.73 times more hazardous for pedestrians than roads with sidewalks. Fewer crashes were observed in areas with sidewalks, implying that an increase in sidewalk density can improve pedestrian safety (Chen and Zhou, 2016). Sidewalk accessibility and connectivity were identified as factors that enhance transit users' experience and accessibility (Brodeur-Ouimet and Cloutier, 2024; FHWA, 2008a). Roadways with high crash rates or transit presence should consider sidewalk installation or rehabilitation.

The literature includes several research studies on the impact of roundabouts on VRU crashes. Signalized intersections converted to roundabouts contribute to a 27-percent reduction in pedestrian crash types (De Brabander and Vereeck, 2007; FHWA, 2008b). Roundabout intersection conversion is more effective in reducing injury crashes; hence, it increases VRU safety for roadways with higher speeds (De Brabander and Vereeck, 2007).

Safety impacts caused by roundabout installation are dependent on the history of the intersection. According to research conducted in Belgium, VRU crash rates on previously signalized intersections increased by 28 percent but decreased 27 percent for non-signalized intersections converted to a roundabout (De Brabander and Vereeck, 2007). After the construction of a roundabout, the number of VRU fatalities per crash doubles or quintuples for intersections without or with, respectively, previous signalization. De Brabander and Vereeck, (2007) concluded that traffic signals protect VRUs more effectively than roundabouts, which are more effective than intersections without signalization. Although the pedestrian crash rate decreases with the construction of a roundabout, the severity (i.e., percentage of fatal casualties) increases (Daniels et al., 2008; Daniels et al., 2009; De Brabander and Vereeck, 2007). This assertion is supported by other research conducted in Belgium: "Previous research in Flanders has shown that the provision of roundabouts results in fewer injury crashes overall but increases in the number and severity of injury crashes among cyclists" (Van Cauwenberg et al., 2018).

The conclusions about roundabout effects are layered in complexity because other studies have shown a significant reduction in the number of collisions involving cyclists and pedestrians. Schoon and van Minnen (1993) found a 30 percent reduction for cyclists and 89 percent reduction for pedestrians while Hydén and Várhelyi (2000) found a 60-percent reduction for bicyclists and an 80-percent reduction for pedestrians. Brüde and Larsson (2000) found a 21-percent reduction in the number of injury crashes involving cyclists on single-lane roundabouts, but a 12 percent increase in crashes on multi-lane roundabouts. For pedestrians, crash rates decreased by 79 percent on single-lane roundabouts but increased by 12 percent on multi-lane roundabouts. Stone et al. (2002) saw a 7-percent reduction in pedestrian-vehicle crashes on roundabouts. The data suggest that intersection conversion to a roundabout is most effective for single-lane roundabouts at low speeds (less than 50 kilometers per hour (kph) or 30 mph) and in intersections with no prior signalization (Craig et al., 2019; De Brabander and Vereeck, 2007).

2.5.2 Education and Enforcement Measures

The Harborview Injury Prevention Research Center in Seattle, WA has studied the impact of education and enforcement measures on pedestrian safety over a 14-year period. Their study showed that strict enforcement of laws protecting pedestrians in crosswalks (intended to influence driver behavior) and teaching pedestrian safety skills to school children, yielded modest results (Chen and Zhou, 2016). The researchers found insignificant changes in driver or pedestrian behavior in response to their measures (Bergman et al., 2002; Britt et al., 1995; Rivara et al., 1991). Some dramatic cases even showed that driver compliance with pedestrian yielding decreased in the presence of focused enforcement (Britt et al., 1995). Craig et al. (2019) found that the percentage of vehicles yielding to VRUs was lower near public transit stops than when no transit stop was present. Other studies showed more promising results. Craig et al. (2019) found that police enforcement increased the percentage of vehicles yielding to pedestrians in a walkway from 33 to 44 percent. Another study showed enforcement to be effective, even up to 4 years later as yielding rates were shown to stay high. Their crash analysis showed a statistically significant decrease in yearly crash rates (Van Houten et al,. 2017).

2.5.3 Roadway Features

Various roadway countermeasures were proposed and studied in the literature. Prominent roadway features included MB crosswalks, raised medians, and Pedestrian Hybrid Beacons (PHB). PHBs showed a statistically significant difference in pedestrian-related and fatal-pedestrian-related crashes in signalized intersections in Arizona (Fitzpatrick et al., 2021). Further study results revealed that crosswalks alone (i.e., without additional PHBs or signaling) showed no difference in crash rates between marked crosswalks and unmarked crosswalks on two-lane roads (Zegeer et al., 2001). On multi-lane roads with traffic volumes above 12,000 vpd, having a marked crosswalk with no other traffic control was associated with a higher pedestrian crash rate (after controlling for other site factors) when compared with an unmarked crosswalk (Zegeer et al., 2001). Crosswalks are typically best used as a countermeasure in conjunction with other treatments (e.g., bulb outs, crosswalk islands, enhanced overhead lighting, pedestrian signals, roadway narrowing, etc.) (Zegeer et al., 2001; Zegeer et al., 2004).

Other roadway features have shown promise in increasing pedestrian safety. Median presence showed a decrease in pedestrian crashes (Zegeer et al., 2001). Center refuge islands in crosswalks have also been shown to be effective in improving pedestrian safety (Bergman et al., 2002; Brilon and Blanke, 1993; Craig et al., 2019). MB crossings account for more than 70 percent of pedestrian fatalities (FHWA, 2001), though the danger may be most applicable on large roads. Zegeer et al. (2005) found that MB crossings by pedestrians were significantly less risky on two- and three-lane roads when compared to roads with four or more lanes. MB crosswalks may be most effective on high volume or roads with four or more lanes. This is consistent with other findings that showed a correlation between pedestrian crash rates and the number of lanes (Zegeer et al., 2001). Raised pedestrian crossings have been shown to contribute to a 30- to 36-percent reduction in pedestrian crashes (Bahar et al., 2007; FHWA, 2008b; Zegeer et al., 2001).

Increasing sidewalk density has also been theorized to increase pedestrian safety as sidewalk density has a negative correlation with pedestrian crashes (Chen and Zhou, 2016). In Philadelphia and Pittsburgh, PA, Guadamuz et al. (2020) installed barriers in the roadway and saw a statistically significant decline in pedestrian-involved crashes. Chen and Zhou (2016) also suggested separation barriers between transit lines and sidewalks to protect VRUs from traffic. Dedicated transit lanes or transit stop bulb ,outs separate stopping transit vehicles and VRUs from the regular flow of traffic (Zegeer et al., 2001; Zegeer et al., 2004).

Traffic-calming measures have also proven to be successful in increasing pedestrian safety. In a study analyzing 200 traffic-calming measures in the United Kingdom, the average annual vehicle and pedestrian crash rate fell by about 60 percent (Webster and Mackie, 1996). The Brilon and Blanke (1993) study on traffic-calming measures in six German cities found street geometry modification to be more effective than introducing a 30-kph (20 mph) speed limit, especially for its impact on VRUs, by decreasing crash rates and crash severity. The researchers found traffic-calming measures decreased seriously injured persons by an average of

63 percent, slightly injured persons by 49 percent, crash costs by 40 percent, motorbikes involved in a crash by 78 percent, cyclists involved in a crash by 17 percent, and pedestrians involved in a crash by 25 percent. The authors also found that street modifications decreased crash severity and, consequently, crash costs fell as costs shifted largely to property damage (Brilon and Blanke, 1993).

Another proposed countermeasure is the use of TSP to give transit vehicles priority at a traffic signal. The argument for TSP is that it makes transit faster and more attractive to riders. Roadways with TSP are associated with decreases in vehicular crash rates, including up to a 22-percent reduction in angle crashes in Florida (Ali et al., 2021), 10-percent reduction in property damage crashes in Portland, OR (Song and Noyce, 2019), and a 14-percent reduction in total crashes in Melbourne, Australia (Ali et al., 2021). Conversely, a few studies found small statistical increases in crashes (1-2 percent) on TSP corridors (Ali et al., 2021). Another study revealed a correlation between TSP corridors and an increase in VRU-involved crashes (Song and Noyce, 2019). Transit ridership also grew with the installation of TSP and researchers suggested pedestrian crash increases are likely due to total pedestrian volume growth. As bicyclists follow vehicular traffic laws, increases in bicycle crashes are hypothesized to be related to TSP signal adjustments and other factors such as the promotion of cycling (Song and Noyce, 2019). Ali et al. (2021) theorized that arterials with the transit line had longer green times, leading to shorter green times on the cross streets. Pedestrians intending to cross the arterial, especially to catch an approaching transit vehicle, are more likely to be tempted to jaywalk, putting them in the way of an oncoming vehicle.

In addition to TSP, a Leading Pedestrian Interval (LPI) can be used to decrease pedestrian crashes. LPI gives the pedestrian an advantage when crossing the intersection (Arun et al., 2023; FHWA, 2021; Saneinejad and Lo, 2015). LPI is recommended for intersections with high turning volumes and has been shown to decrease pedestrian-vehicle crashes at intersections by 13 percent (FHWA, 2021). Because intersections have the highest crash rates, significant crash reductions at intersections are important for crash rate reductions systemwide.

2.6 Summary

The relationship between transit stops and VRUs is complex, and while there is a strong correlation between pedestrian crashes and transit presence, the reasons may not be straightforward. There is a clear correlation between pedestrian volumes and crash rates while certain land use types contribute to pedestrian crash rates more than others. Land use types that draw more pedestrian activity see higher crash rates, including commercial land areas, schools, and others. Spatial tools like "buffer," "kernel density," "spatial join," and "K"- and "G-functions" can be used to identify crash-prone areas. The prevailing statistical models used in highway safety research include Bayesian models, Spearman's rank correlation, MCMC analysis, and Cross K-function. The literature shows a strong correlation between roadway characteristics (e.g., speed, road width, and signalized intersections) and pedestrian crashes. Intersections tend to have the highest pedestrian crash rates. Effective countermeasures have included sidewalks, center islands, MB crossings, pedestrian signals, pedestrian barriers, dedicated transit lanes/bulb outs, and raised pedestrian crossings. The information from the literature review will provide a basis for data collection parameters, such as appropriate buffer radii and statistical models to use, land use features to investigate, and roadway features to focus on.

3.0 DATA COLLECTION

3.1 Overview

This chapter describes where the input data were obtained from for the transit stop and crash analyses, along with the purpose for using the various datasets. Data for this project were collected as either shapefiles or comma-separated values (CSV) files from the AASHTOWare Safety database, UTA Open Data Portal, UGRC, and the Wasatch Front Regional Council (WFRC) open database. Sections in this chapter are organized by database.

No data were excluded from an analysis unless a specific feature was uniformly filtered out for comparison purposes as described in Chapter 4.0. All data were projected into the World Geodetic System (WGS) 1984 Universal Transverse Mercator (UTM) Zone 12 North (12N) projection, optimized for Utah.

3.2 AASHTOWare Safety Database

The crash data came from the AASHTOWare Safety database (Numetric, 2025). Details about each crash, including location estimate, are recorded by the police officer who files the crash report, which is then stored in the AASHTOWare Safety Database. The crashes (point data type) file includes additional information. Each crash has a latitude and longitude value as shown in Figure 3.1, a unique crash ID (automatically generated), milepoint of the crash, and crash date. Crash severity is also recorded and is based on a KABCO severity rating as defined previously in Table 2.1. The manner of crash is recorded as head on, angle, sideswipe same direction, sideswipe opposite direction, front to rear, rear to rear, rear to side, parked vehicle, not applicable/single vehicle, other, or unknown. The roadway junction description is defined as 4-leg intersection, 5-leg or more intersection, T-intersection, Y-intersection, business drive, bridge (overpass/underpass), roundabout/traffic circle, multi-use path/trail intersection, railroad crossing, alley, crossover in median, farm/residential drive, on-ramp merge area, off-ramp diverge area, on-ramp, off-ramp, ramp intersection with crossroad, no feature/junction, other, or unknown. Lighting conditions are classified as daylight, dusk, dawn, dark-lighted, dark-not lighted, dark-unknown lighting, or unknown. Weather conditions are rated as clear, cloudy, rain, snowing, blowing snow, sleet/hail, fog/smog, severe crosswinds, or unknown. Roadway conditions are classified as wet, slush, dry, snow, ice/frost, sand, dirt, gravel, water (standing, moving), oil, mud, unknown, or other. The number of vehicles involved is also recorded. Crashes that are animal involved, distracted driving, drowsy driving, heavy truck involved, intersection involved, left or U-turn involved, motorcycle involved, pedestrian involved, bicycle involved, older driver, right turn involved, speeding involved, teenage driver, unrestrained passengers, and wrong way involved are classified with a Y (yes) or N (no) value.

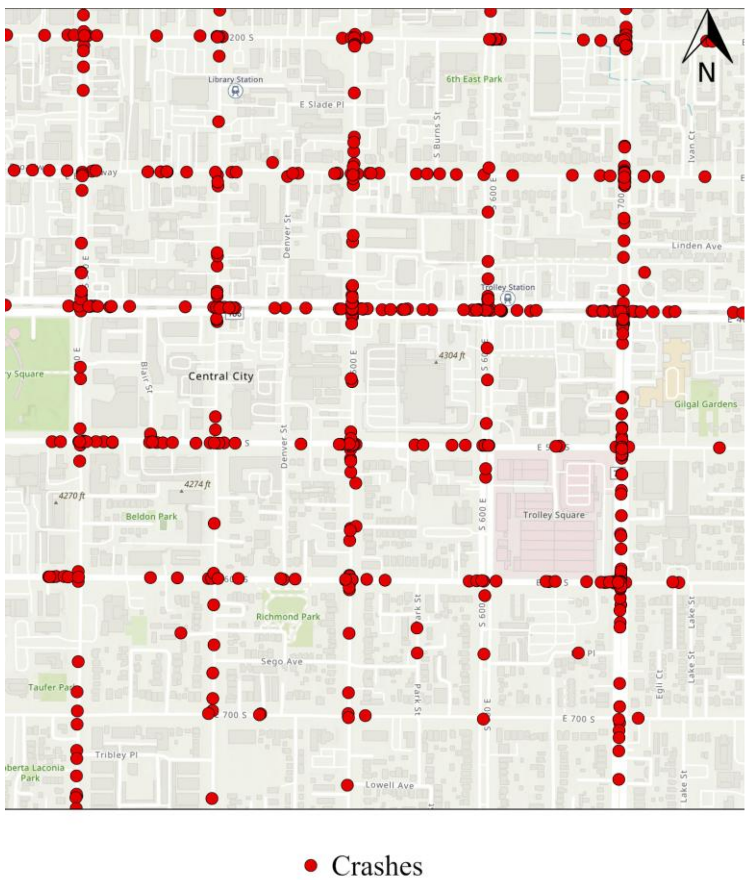


Figure 3.1 Crashes shapefile example from AASHTOWare safety database (Esri, 2023; Numetric, 2025)

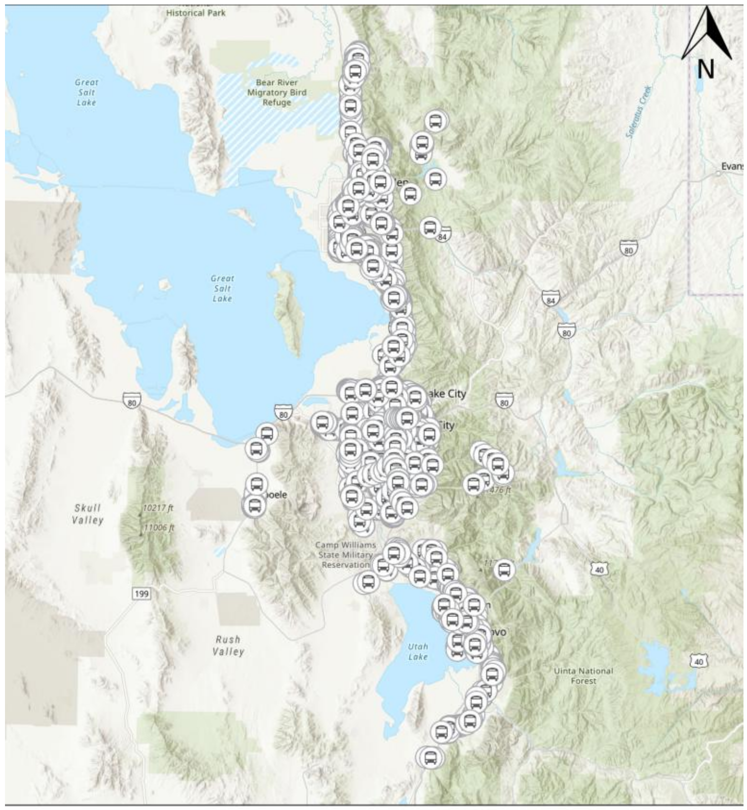
The AASHTOWare Safety database crashes from January 1, 2019, to December 31, 2023 were used for this research. All crash entries collected during that time have been verified by the UDOT data verification process using an online system. The quality-control verification process includes verification of vehicle and person data, accurate geolocation and vehicle direction, and verification of crash characteristics, such as crash severity, manner of collision and road junction type. Crash information was recorded by police officers at the time of a crash report. This dataset was used as the basis for the analysis.

A separate crash rollup file provided by UDOT contained designations indicating if a crash was pedestrian involved, bicycle involved, or transit involved. The rollup data were combined with the crash shapefile.

3.3 Utah Transit Authority Open Data Portal

The transit stop and most recent ridership shapefile (point data type) came from the UTA Open Data Portal (UTA, 2025). The shapefile was last updated in May 2025. UTA operates transit stops in Salt Lake, Utah, Weber, Morgan, and Davis Counties, as well as portions of Box Elder and Tooele counties as shown in Figure 3.2. These transit stops defined the service area for the research. Along with the geolocations of the points, the shapefile also contains each stop name, UTA Stop ID, and the average number of passengers boarding and alighting at each transit stop per day. A column for transit mode is also included with options of bus, flex bus, TRAX, FrontRunner, microtransit, or blank.

A CSV file of additional transit stop data produced by the internal UTA transit stop manager was later combined with the original transit stop shapefile by the research team. The CSV file includes a variety of transit stop characteristics. Bike lane designations are characterized as yes-single stripe, yes-sharrows, yes-buffered, yes-protected, or blank. Lighting is designated as direct, indirect, or absent. A designation indicates if the stop is Americans with Disabilities Act (ADA) compliant. The data also included the transit-stop point position in relation to an intersection defined as NS, FS, or MB. A transit stop is defined as MB if the stop is more than approximately 300 ft. from the nearest stop bar of the intersection, as determined by the transit-stop management planner's judgment. Stop direction tells if a transit stop faces north, south, east, or west. The stop point designates the on-street location of the transit stop, indicating whether the transit stop is in a travel lane, on or off street, in a parking lane, transit bay, or in a TRAX, FrontRunner, or bus rapid transit (BRT) platform. The surrounding land type is classified as business, commercial, industrial, farming, government, recreational, single home, apartment, library, hospital, church, transit station, school, mixed use, undeveloped, or blank. School zone presence is indicated as present or absent. Additional information including the sidewalk width around the transit stops and the pathway description were also included in the file. A portion of the CSV file is shown in Figure 3.3.



Transit Stops

Figure 3.2 Transit stop data from UTA portal (Esri, 2023; UTA, 2025)

STOP:LandType	STOP:Pathway	STOP:Point_Location
Apartment	Sidewalk-Not Connected	Travel Lane
Single Home	Sidewalk-Not Connected	Travel Lane
Business	Yes	On Street
Commercial	Yes	On Street
Business	Yes	On Street
Business	Sidewalk-Not Connected	On Street
Apartment	Sidewalk-Not Connected	On Street
Single Home	Sidewalk-Not Connected	On Street
Single Home	Sidewalk-Not Connected	On Street
Church	ADA Path	On Street
Business	Sidewalk-Not Connected	On Street
Business	ADA Path	On Street
Single Home	Sidewalk-Not Connected	On Street
Single Home	Sidewalk-Not Connected	On Street
Single Home	Sidewalk-Not Connected	On Street
Single Home	Yes	On Street

Figure 3.3 Transit stop data example from CSV file (UTA, 2025)

3.4 Utah Geospatial Resource Center

The Utah counties and roads shapefiles came from the UGRC (UGRC, 2025). Utah county boundaries (polygon data type) were used to clip statewide datasets (roads, intersections, and crashes) to exclude extraneous data outside the UTA transit service area. The counties used were Salt Lake, Utah, Weber, Morgan, and Davis. Relevant portions of Box Elder and Tooele counties were also included as shown in Figure 3.4. The counties shapefile was last updated in June 2025.

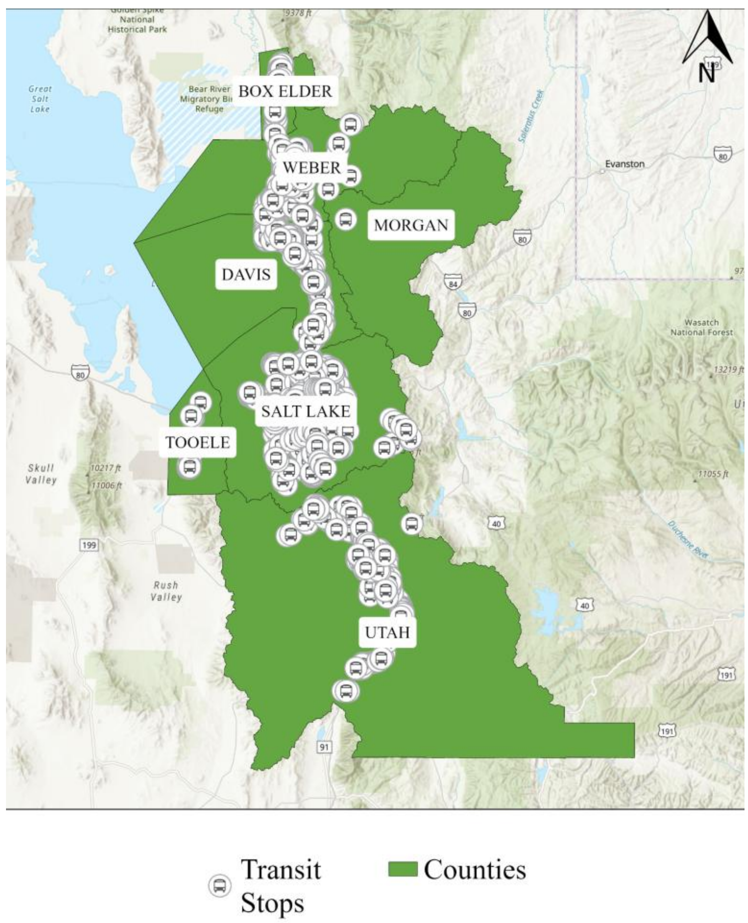


Figure 3.4 County boundaries used (Esri, 2023; UGRC, 2025; UTA, 2025)

The roads data shapefile (line data type) was used to create buffers that followed the length of the roadway and inform crash data as shown in Figure 3.5. Additional details about the analysis are recorded in Chater 4.0. The shapefile metadata contains roadway classifications including interstates, US highways, major state highways, other state highways, ramps, collectors, major local roads, other federal-aid-eligible local roads, other local, neighborhood, rural roads, other, non-road features, driveway, proposed, 4WD and/or high clearance may be required, service access roads, and general access roads. Roadway classifications are organized by a Cartographic Code. The Cartographic Codes associated with each roadway classification are listed in Table 3.1. The only classification codes excluded from the data were those related to interstates (cartographic codes 1 and 7 in Table 3.1). Crashes on interstates and ramps were excluded from the data because crashes on interstates were unlikely to be related to transit stops even though they could be in proximity in two-dimensional geographic space. A significant number of interstate-related crashes were originally included in the analysis. To systematically remove crashes related to interstates, road segments along interstates and ramps were removed from the roads shapefile. The shapefile was last updated in June 2025. The roads shapefile also contains data describing 2017 AADT for UDOT-owned routes, speed limit, and a cardinal direction that describes the primary cardinal direction of the road.

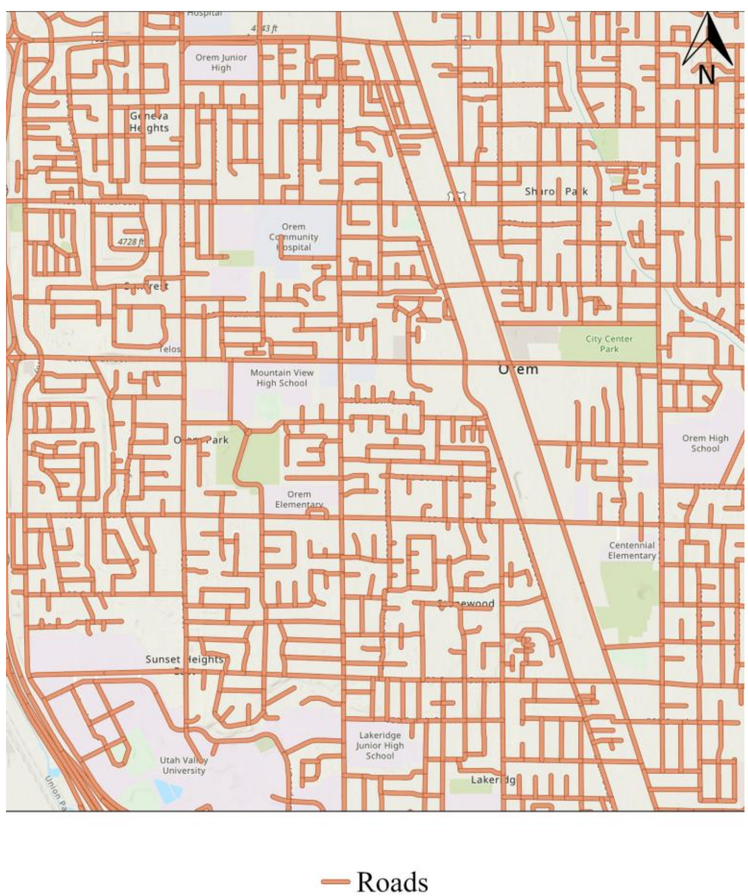


Figure 3.5 Roads shapefile example from UGRC (Esri, 2023; UGRC, 2025)

Table 3.1 Roadway Cartographic Code Classifications

Cartographic Code	Definition
1	Interstates
2	US Highways, Separated
3	US Highways, Unseparated
4	Major State Highways, Separated
5	Major State Highways, Unseparated
6	Other State Highways (Institutional)
7	Ramps, Collectors
8	Major Local Roads, Paved
9	Major Local Roads, Not Paved
10	Other Federal-Aid-Eligible Local Roads
11	Other Local, Neighborhood, Rural Roads
12	Other
13	Non-road feature
14	Driveway
15	Proposed
16	4WD and/or high clearance may be required
17	Service Access Roads
18	General Access Roads

3.5 Wasatch Front Regional Council Open Database

The intersection centers shapefile (point data type) came from the WFRC open database (WFRC, 2025). The intersection centers shapefile was last updated in February 2025. Each point in the intersection center shapefile represents the geographic location of the center of each intersection within the WFRC area, as shown in Figure 3.6.

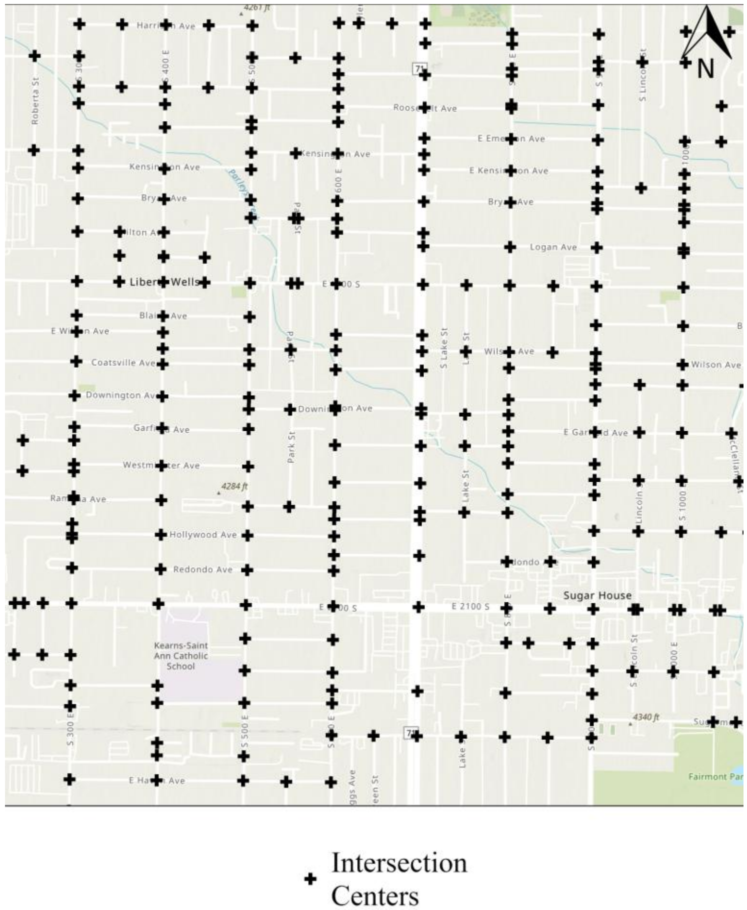


Figure 3.6 Intersection centers example from WFRC database (Esri, 2023; WFRC, 2025)

3.6 Summary

Data collected in this project came from existing databases or open data portals. Data includes shapefiles for UTA transit stops, crashes, Utah county boundaries, Census2020 blocks, elevation, Utah roads, intersection centers, and sidewalks. No data were excluded from the analysis unless a certain feature was uniformly filtered out for comparison purposes. All shapefile projections were defined in ArcGIS Pro (Esri, 2023) as WGS 1984 UTM Zone 12N. Data evaluation, described in Chapter 4.0, used a combination of Python and ArcGIS Pro (Esri, 2023) to analyze relationships in the data.

4.0 DATA EVALUATION

4.1 Overview

Chapter 4.0 details the analysis of the data collected in Chapter 3.0. This involved measuring the distance from each crash to the nearest intersection and the nearest transit stop. To control variables, both sets of data were divided into several sub-analyses. The following sections provide a detailed data evaluation encompassing intersection density analysis, transit stop distance analysis, and a Python data evaluation to compare spatial results.

4.2 Intersection Density Analysis

ArcGIS Pro (Esri, 2023) software was used to conduct all analysis. An important first step to ensure consistency and maintain accuracy in geospatial analysis was to verify that all input datasets used (intersection point locations, roads, etc.) were projected to the same geographical projection. This analysis used WGS 1984 UTM Zone 12N, as that is the projection that generally produces the most accurate results for the Wasatch Front in Utah (the chosen study area).

The intersection density analysis measured the geometric distance from intersections as a center point to transit stops, all crashes, and VRU crashes. VRU crashes are defined in this project as a crash involving either a bicycle or pedestrian. Intersections are common high incident locations (Chen and Zhou, 2016; Craig et al., 2019; Jovanis et al., 1991; Kim et al., 2010; Yang, 2007), so this analysis is intended to compare intersection crashes with crashes near transit stops, to better distinguish where crashes occurred between those two areas. Further analysis was completed to demonstrate the relationships between VRU crashes and transit. The intersection distance methodology and buffer width testing will be discussed in this section.

4.2.1 Intersection Distance Methodology

ArcGIS Pro (Esri, 2023) includes a buffer tool that uses polygons to capture one dataset related to another. In this analysis, a fixed-distance buffer was generated around each intersection using the ArcGIS Pro “buffer” tool (Esri, 2023), creating a circular buffer surrounding each intersection point to capture all crashes and transit stops within a specified distance of the intersection. This was used to filter extraneous data (crashes or transit stops clearly outside of the intersection influence area) and to reduce file sizes. The value inputs are discussed in Section 4.2.2. The intersection influence area was defined by UDOT as 400 ft.

Buffer radii of 250 ft., 400 ft., or 1,400 ft. were used. The buffer width parameters decision is outlined in Section 4.2.2. Buffers used “geodesic” type and dissolve settings to prevent repeated records and ensure the highest degree of accuracy. Buffers were measured with the ArcGIS Pro “measure” tool (Esri, 2023) as a measure of quality assurance.

The intersection buffers were then overlaid with the road network using the ArcGIS Pro “intersect” tool (Esri, 2023). By specifying the output geometry as “line,” the tool generates a layer of road segments clipped to the extent of the buffers. This process restricted the road features to a maximum length of 250 ft., 400 ft. or 1,400 ft. from the intersection, depending on the initial circular radius.

Clipped roads from the previous step were then buffered 100 ft. on either side of the centerline of the roadway for a total width of 200 ft. Road buffers also used the “geodesic” type, “flat” end type, and “dissolve features using the listed fields’ unique values or combination of values” settings.

Crashes were then intersected with the roads buffer using the “intersect” tool (Esri, 2023) to obtain a final output of crashes that were within a specified road length from the intersection (250 ft., 400 ft., or 1,400 ft.). In a separate step, transit stops were also intersected with the roads buffer. For VRU-specific analyses, bicycle and pedestrian crashes were isolated using the “select layer by attribute” tool (Esri, 2023) prior to performing the intersection buffers.

The next step (completed separately for transit stops and crashes, though described here only once) was to use the “near” tool in ArcGIS Pro (Esri, 2023) to calculate the distance of each crash or transit stop to the nearest intersection point. The object identification number of the nearest intersection was generated as a column in the shapefile table. The calculated distance was generated as another column. The output from this tool was a table exported to an Excel file of all the crashes or transit stops and their associated metadata within the defined distance (250 ft., 400 ft., or 1,400 ft.) from an intersection. These Excel files were then used in the Python analysis (described in Section 4.4.1) and statistical analysis (described in Chapter 5.0). A full methodology with all associated settings is included in Appendix A.

The methodology was repeated with crashes limited to only VRU-involved crashes. The final output of the intersection analysis was a series of datasets that were used for comparison purposes as will be outlined in Chapter 5.0. The datasets are:

- Transit stops within 250 ft. of the nearest intersection
- Transit stops within 1,400 ft. of the nearest intersection
- Crashes within 250 ft. of the nearest intersection
- Crashes within 1,400 ft. of the nearest intersection
- VRU crashes within 250 ft. of the nearest intersection
- VRU crashes within 1,400 ft. of the nearest intersection

4.2.2 Buffer Width Testing

The intersection density analysis used an initial circular buffer to capture crashes a specified distance from the intersection (classified by the radius of the circular buffer). The roads were clipped to exclude any section of road that does not fall within the intersection buffer. Each road was then buffered to a variable width of between 24 ft. and 100 ft. to capture all crashes and transit stops on either side of the road. Intersection and road buffer dimensions were iteratively adjusted to compare different spatial scales. Radii of 150 ft., 200 ft., and 250 ft. were tested. Since the initial circular buffer radius length becomes the length of the segment of road captured, buffer lengths are described by the radius length of the initial intersection buffer (e.g., 250 ft.) as shown in Figure 4.1.

A small proportion of transit stops were outside of the road buffers because they were transit maintenance stations or ski transit stops that were too far from the centerline of the road. It was expected that some proportion of all transit stops may not be captured, so the intention was to capture the highest percentage achievable, which was determined to be about 75 percent.

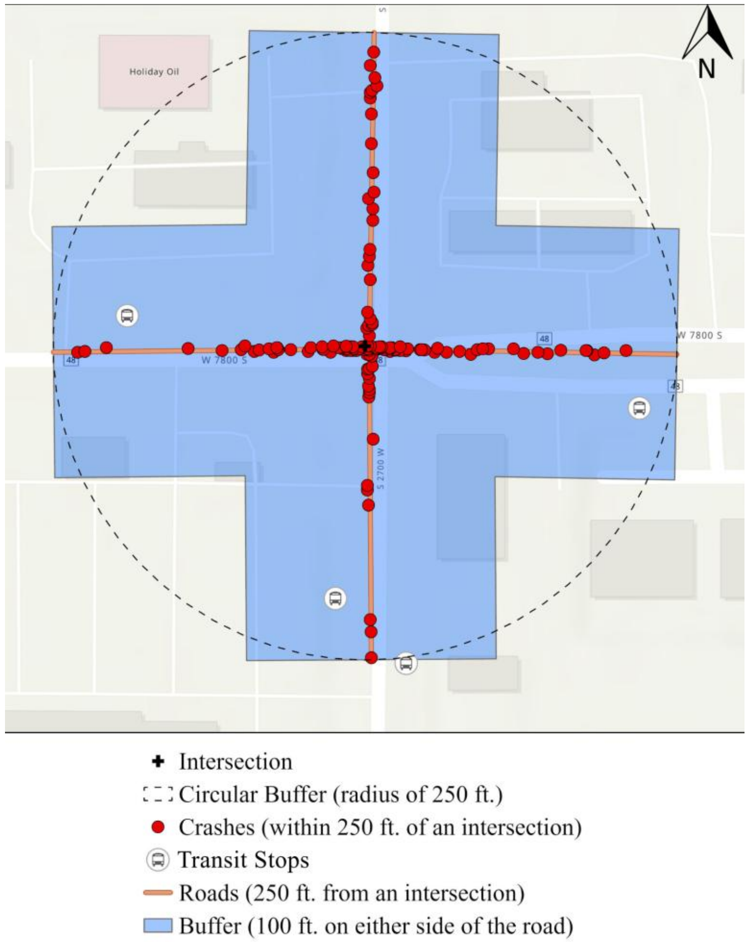


Figure 4.1 Road buffer lengths and widths as shown in ArcGIS Pro (Esri, 2023)

The first potential road buffer test used an intersection radius of 150 ft. and counted the number of transit stops within road buffer widths of 24 ft., 36 ft., 48 ft., 60 ft., and 100 ft. (on both sides of the roadway for a total buffer width of 48 ft., 72 ft., 96 ft., 120 ft., or 200 ft.). The number of transit stops counted within the buffer area compared to the total number of transit stops within the system was calculated as a percentage and is summarized in Figure 4.2. The percentage of transit stops captured by the buffer effectively plateaus at 60 ft., with larger radii yielding only marginal gains. While the data density stabilized at 60 ft., the final buffer was extended to 100 ft. as a conservative measure. This width was selected to capture nearly all relevant crash data while avoiding the inclusion of extraneous noise from excessively large buffers.

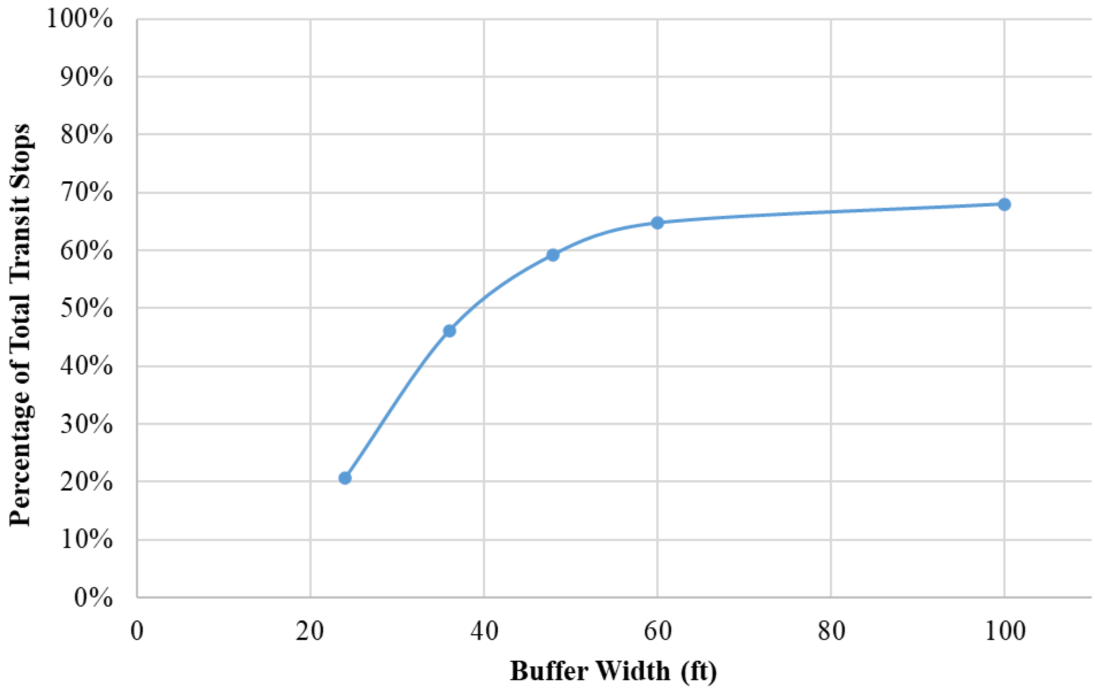


Figure 4.2 Percentage of transit stops per road buffer width on 150 ft. road segments

The same analysis was conducted for 200 ft. radius buffers as shown in Figure 4.3, with similar trends observed at the 250 ft. radius as shown in Figure 4.4. The 250 ft. width test revealed that after 48 ft., more than 100 percent of all stops were included in the analysis (up to 115 percent), as shown in Figure 4.4. Further inspection of the data revealed that transit stops were counted every time they appeared in a buffer. This meant about 2,000 transit stops were being counted more than once where buffers overlapped.

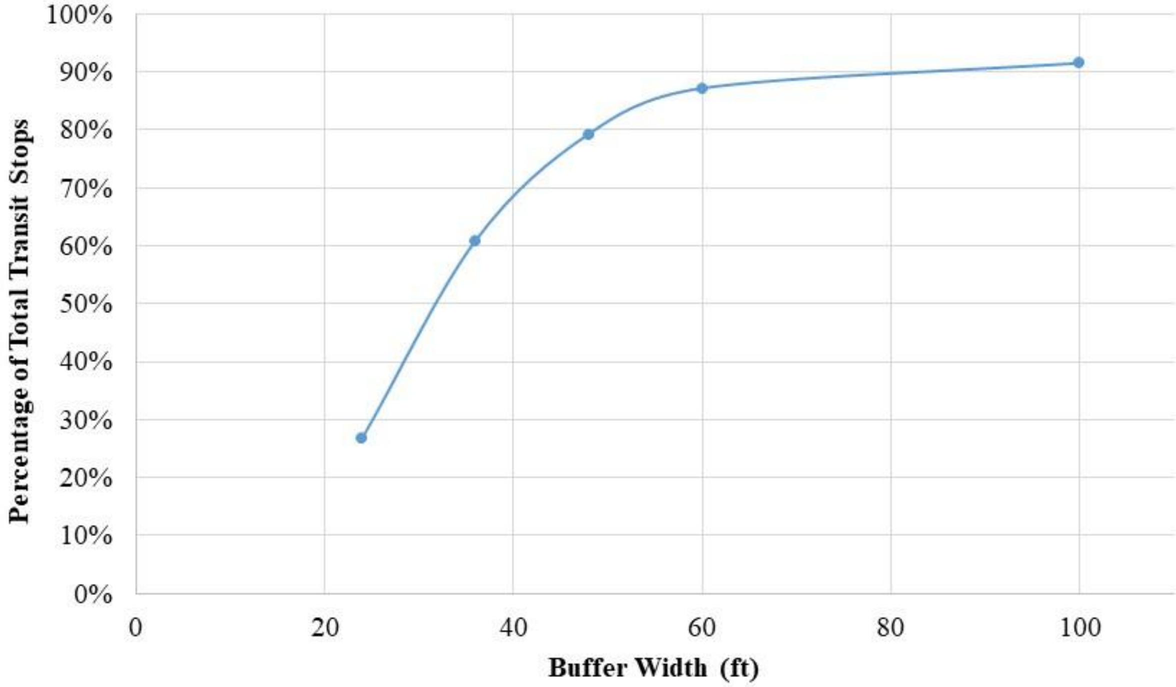


Figure 4.3 Percentage of transit stops per road buffer width on 200 ft. road segments

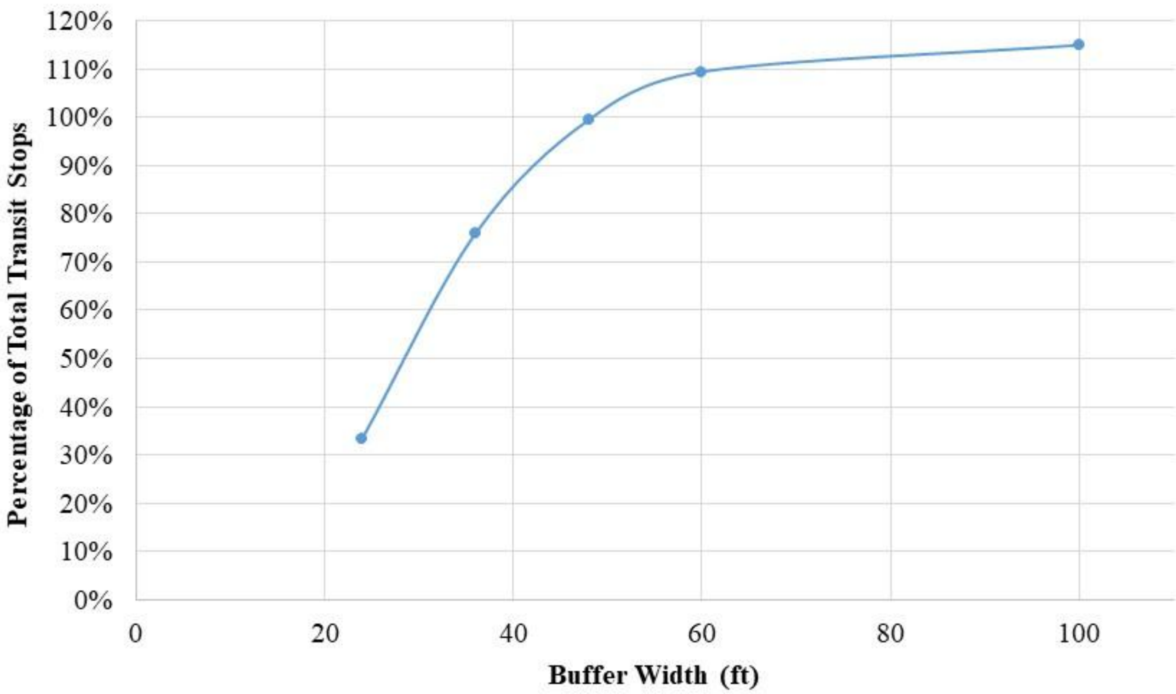


Figure 4.4 Percentage of transit stops per road buffer width on 250 ft. road segments

To prevent individual data points from appearing more than once, the roads buffers were all dissolved and the “geodesic” buffer method was used. Another test was run using the intersection buffer radius of 250 ft. with the “geodesic” buffer method and the dissolved buffer setting applied, as shown in Figure 4.5. This test revealed that 75 percent was the highest percentage of transit stops captured without counting data points more than once. A similar plateau was observed above 60 ft., so the maximum road width of 100 ft. with 250 ft. length was chosen as a conservative estimate and was used as the road buffers for the remainder of the project. Later analysis included intersection initial buffers up to 1,400 ft.

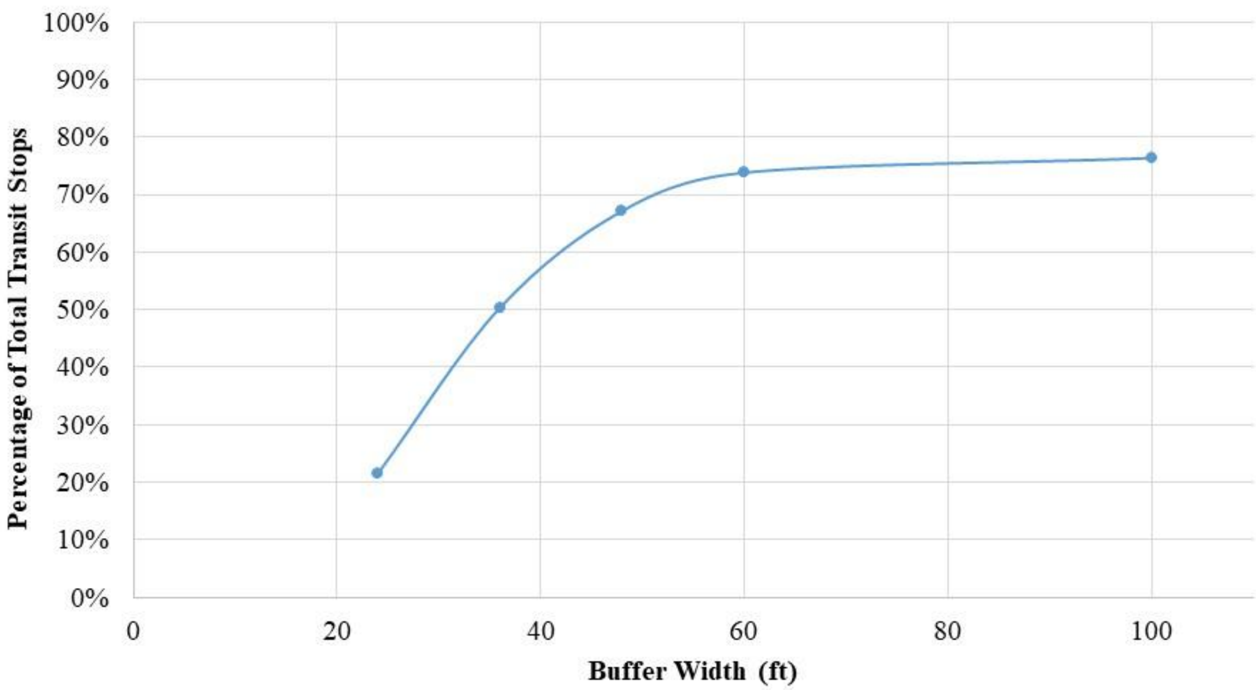


Figure 4.5 Percentage of transit stops per road buffer width on 250 ft. road segments

Figure 4.6 shows the ArcGIS Pro view of the roads with crashes, transit stops, and final buffer with a 250 ft. radius and 100 ft. wide buffer. Crashes and transit stops inside the buffer are counted in the analysis. If crashes or transit stops fall outside of the transit stop, they are not included.

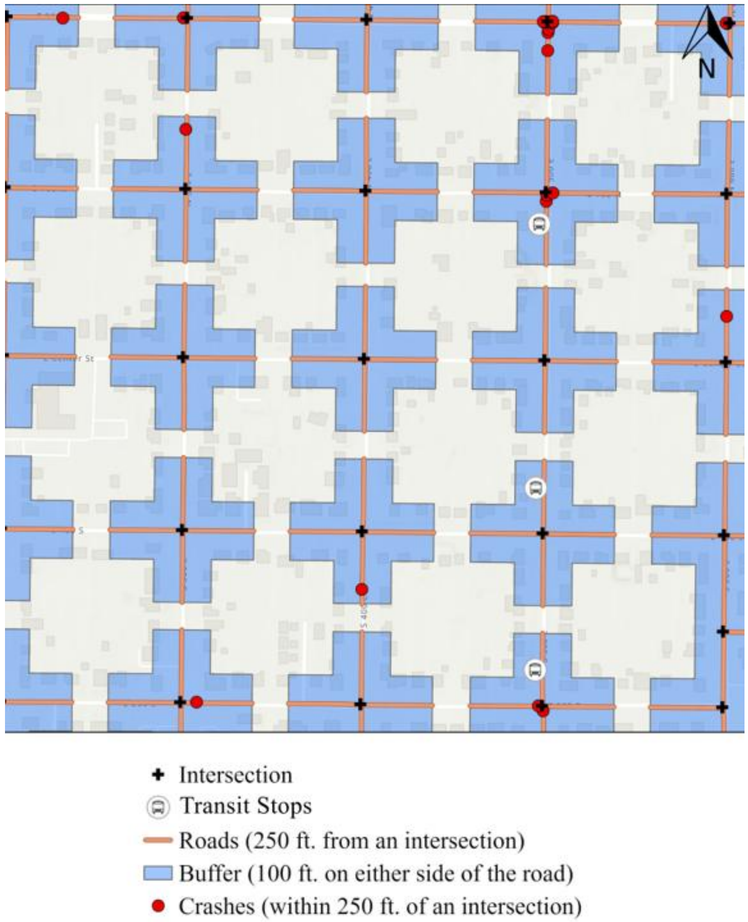


Figure 4.6 Intersection buffers view as shown in ArcGIS Pro (Esri, 2023)

4.3 Transit Stops Distance Analysis

Transit-stop distance analysis measured the geometric distances of all crashes and VRU crashes to the nearest transit stop. This serves as an extension of the intersection distance analysis, where patterns of the crashes around transit stops were studied. The “near” tool in ArcGIS Pro was used to calculate distances. The methodologies for the transit-stop-distance geospatial analyses will be discussed in this section.

Similar to the intersection distance analysis, a fixed-distance buffer was generated around each transit stop to capture all crashes within a specified distance of the transit stop. This was used to filter extraneous data (crashes or transit stops clearly outside of the transit stop influence area) and reduce the file size. The “buffer” tool was used to buffer the transit stops, with the “geodesic” type chosen. The buffers were dissolved, and buffer radii of 250 ft., 1,000 ft., or 1,400 ft. were used. Consistent with Ulak et al. (2021), a maximum radius of 1,400 ft (0.25 mi) was used to reflect average pedestrian walking distances.

The road network was intersected with the transit stop buffers using the “intersect” tool (Esri, 2023). In this context, intersecting refers to extracting road segments that geographically overlap with the buffer polygons, effectively clipping the network while preserving the attributes of both datasets. By specifying the output geometry as a “line,” the tool generated a layer of road segments clipped to the extent of the buffers. This process restricted the road features to the radius of the buffers.

Clipped roads from the previous step were then buffered 100 ft. on either side of the centerline of the road, for a total width of 200 ft. Road buffers also used the “geodesic” type, “flat” end type, and “dissolve features using the listed fields’ unique values or combination of values” settings. Figure 4.7 illustrates the clipped roads in ArcGIS Pro.

Crashes were then intersected with the roads buffer using the “intersect” tool (Esri, 2023) to obtain a final output of crashes that were within a specified road length (250 ft., 1,000 ft., or 1,400 ft.), from the transit stop. Figure 4.7 shows the buffers at 250 ft. Additional figures with buffers up to 1,400 ft. are included in Appendix B. In a separate step, transit stops were also intersected with the roads buffer. For analyses that focused on VRU crashes, the “select layer by attribute” tool (Esri, 2023) was used to select and separate bicycle and pedestrian-related crashes before they were intersected.

The next step was to use the “near” tool in ArcGIS Pro (Esri, 2023) to calculate the distance of each crash to the nearest transit stop. The object identification number (NEAR_FID) of the nearest transit stop is generated as a column in the shapefile table. The calculated distance is generated as another column. The output from this tool was a table exported to an Excel file of all the crashes and their associated metadata within the defined distance (250 ft., 1,000 ft., or 1,400 ft.) from a transit stop. A full methodology with all associated settings is included in Appendix B.

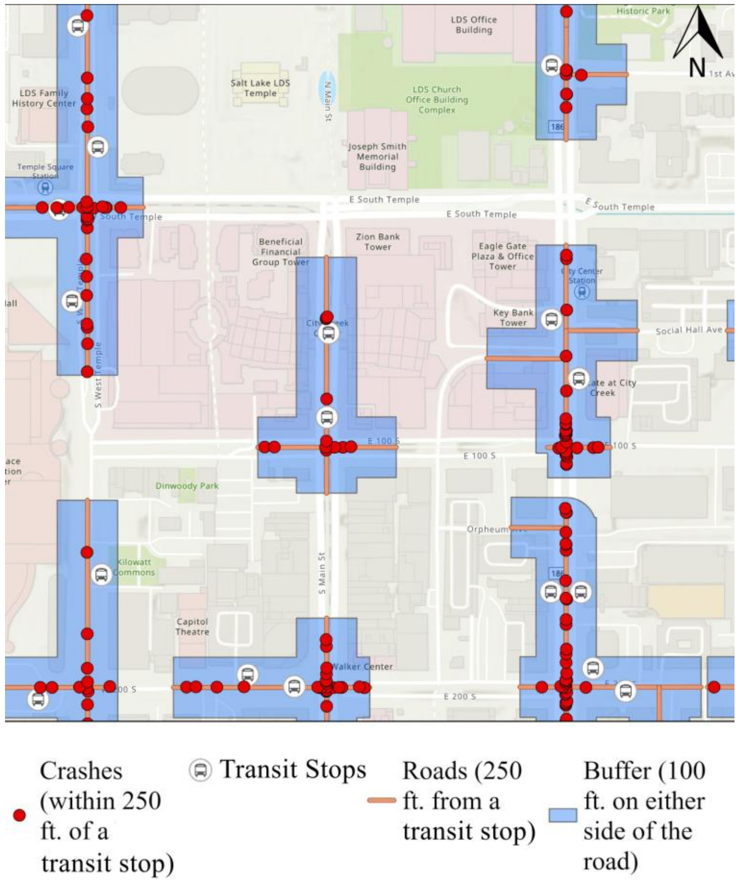


Figure 4.7 View of transit stop buffers as shown in ArcGIS Pro (Esri, 2023)

The methodology was repeated with crashes limited to VRUs. The final output of the transit stop analysis was a series of datasets that were used for comparison purposes as will be outlined in Chapter 5.0. The datasets are:

- Crashes 250 ft. from the nearest transit stop
- Crashes 1,400 ft. from the nearest transit stop
- VRU crashes 250 ft. from the nearest transit stop
- VRU crashes 1,400 ft. from the nearest transit stop

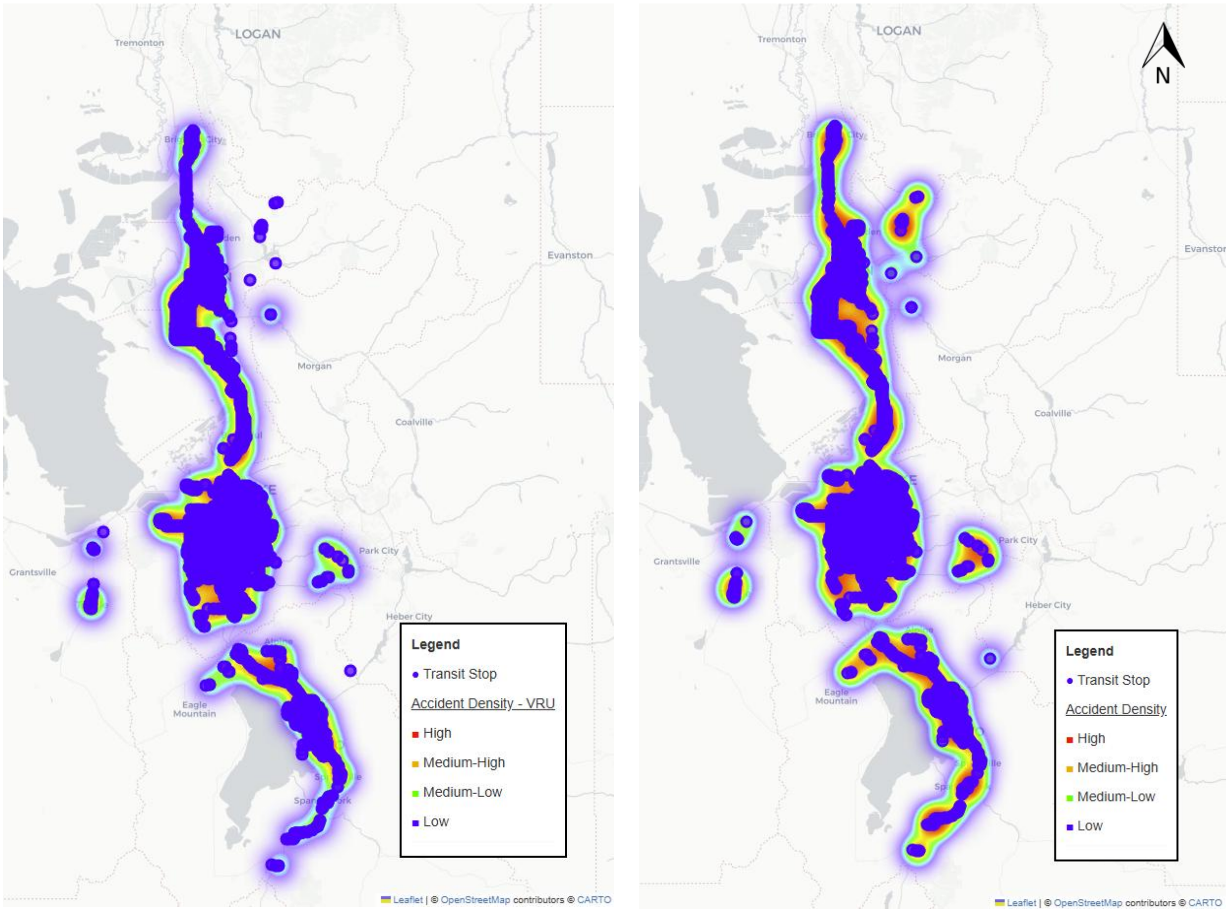
4.4 Python Data Evaluation

This section describes how Python code was used to aggregate, sort, and visually present the data generated in Section 4.2 and Section 4.3 for further analysis. This includes subsections of heatmap presentations and most frequently occurring transit stop and intersections comparison.

4.4.1 Heatmap Presentation

Python code was used to generate heatmaps from the output Excel files from the Intersection and Transit Stops distance analyses. The python code is included in Appendix C. The heatmaps aid in visually demonstrating the relationship between VRU crashes and the presence of a transit stop, while also locating high-crash-density areas or transit stops. Figure 4.8(a) and Figure 4.8(b) show the final output maps of the VRU crashes and total car crashes in a side-by-side comparison.

A general trend of crashes correlating to transit stop locations is apparent from the analysis. Few areas with high transit stop densities have low crash densities and few areas with low transit stop densities have high crash densities. In general, most VRU crashes occur in high VRU-friendly areas, which also coincide with high-density transit stop areas. This is evident most prominently in areas in Tooele County and northwest of Brigham City (the two clusters of transit stops to the left and right of the conglomerations of the main body of transit stops) as pictured in Figure 4.8(a), where there are clusters of transit stops, but few VRU crashes. Those same areas have ample vehicle crashes, so this suggests that those areas have little VRU traffic, contributing to fewer VRU crashes.



(a) **(b)**

Figure 4.8 (a) VRU crashes 250 ft. from transit stops and (b) all crashes 250 ft. from transit stops

4.4.2 Most Frequently Occurring Transit Stops and Intersections Comparison

Python code was used to organize each of the crash datasets from Section 4.2 and Section 4.3 to highlight transit stops or intersections that occur most frequently in the crash data. In other words, the Python code sorted the data to find the transit stops and intersections with the most crashes. The code is included in Appendix D.

The study established that the top 45 transit stops and intersections with the highest crash frequencies provided an adequate sample size. Histograms were generated to visualize the following analysis types with Figure 4.9 serving as an example:

- Transit stops with crashes up to 250 ft. away
- Transit stops with crashes up to 1,400 ft. away
- Intersections with crashes up to 250 ft. away
- Intersections with crashes up to 1,400 ft. away
- Transit stops with VRU crashes up to 250 ft. away
- Transit stops with VRU crashes up to 1,400 ft. away
- Intersections with VRU crashes up to 250 ft. away
- Intersections with VRU crashes up to 1,400 ft. away

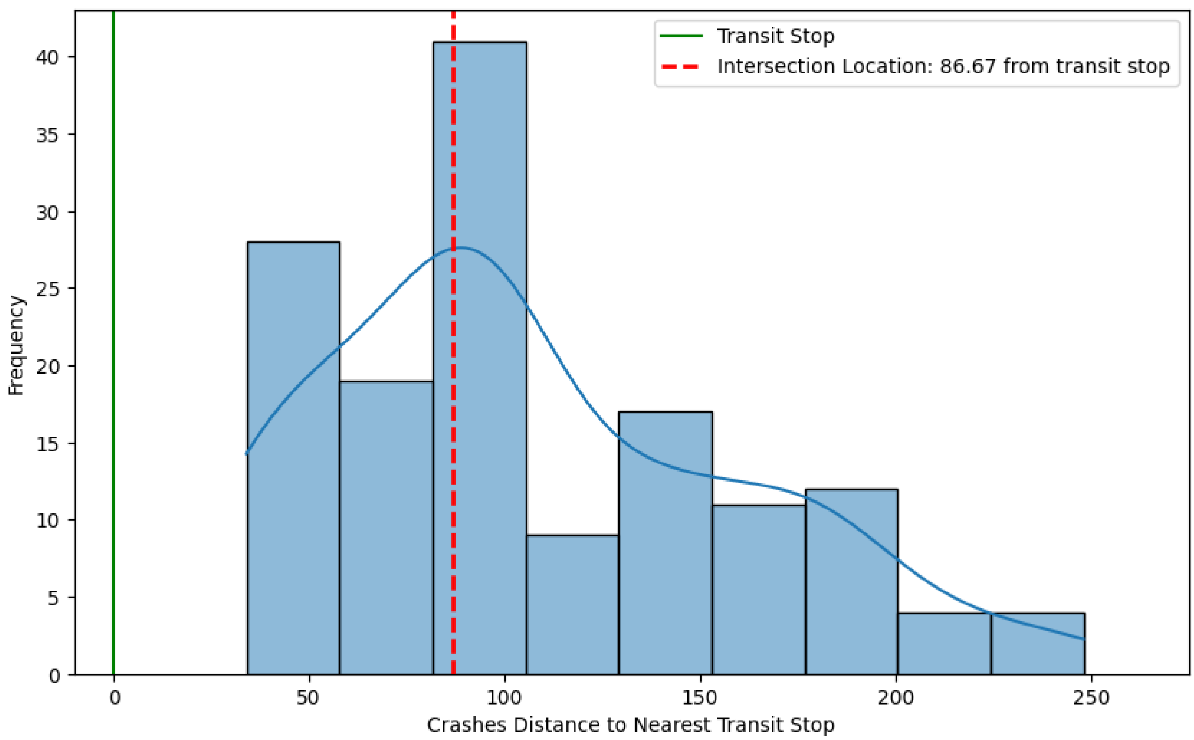


Figure 4.9 Histogram of crashes occurring within a 250 ft. buffer of a selected transit stop location

Analysis of transit-stop crash data reveals that crash clusters are more closely correlated with intersection proximity than with the transit stops themselves. Since most transit stops are adjacent to intersections, the latter serves as the more significant contributor to traffic crashes. This spatial relationship suggests that the primary risk stems from intersection dynamics, and reinforces the observation that pedestrians are choosing to cross at intersections rather than MB.

These transit stops were then mapped for a visual comparison. The transit stops and intersections with the most crashes were mapped together for direct comparison. The code for each of these analyses is included in Appendix D. A final map comparing the top 45 intersections, transit stops, and transit stops with VRU crashes (all with the crashes up to 1,400 ft. from the reference point), is shown in Figure 4.10.

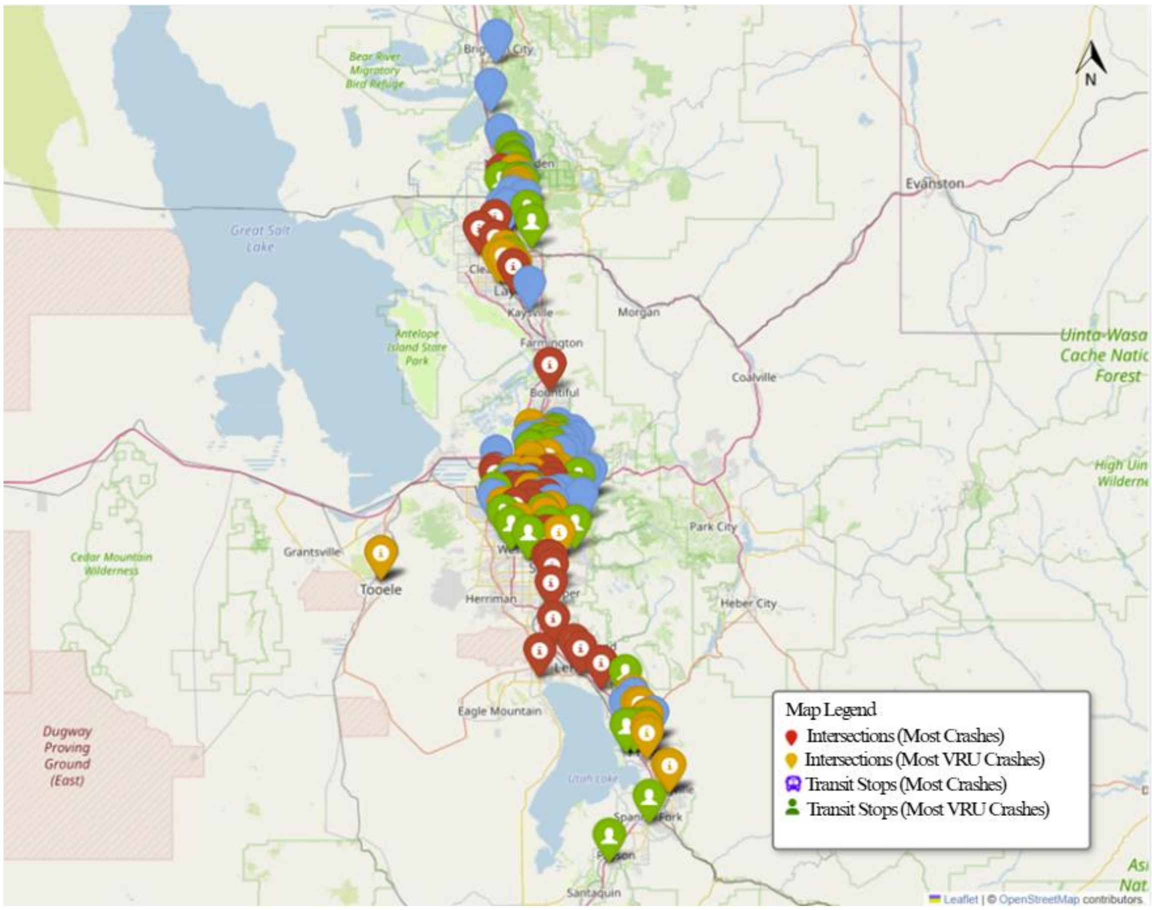


Figure 4.10 Combined top 45 transit stops, VRU transit stops, and intersections map

Visual inspection revealed minimum overlap between high-crash intersections and transit stops. Furthermore, VRU-specific high-crash transit stops rarely overlap with either top-tier intersections or transit stops.

4.5 Summary

The data analysis in this project came from a variety of methodology and analysis tools, including ArcGIS Pro and Python coding, to properly inspect and quantify the data results. This project aims to compare crash occurrences at and around intersections and transit stops, as well as to define correlations between high crash areas. ArcGIS Pro was utilized to generate spatial buffers, allowing for the mapping and analysis of all crashes and transit stops with a defined radius of each intersection or transit stop. Additionally, the intersection density and transit-stop distance analysis adds to the existing crash data measurement from each crash to the nearest intersection or transit stop. Measurements were also computed for the distance from each transit stop to the nearest intersection, and separate models were generated for VRU-specific crash distances from the nearest transit stop. Initial radii of 250 ft. and 1,400 ft. were chosen and used throughout the analysis. Crashes were intersected with road buffers to obtain a final output of crashes that were within either 250 ft. or 1,400 ft. from the transit stop. Crashes, transit stops, intersections, and VRU crashes datasets were then output into an Excel file.

Python code was used to analyze the Excel files by extracting the top 45 transit stops and top 45 intersections with the highest crash rates. Top 45 transit stops and intersections were mapped for visual comparison. This analysis confirmed previous numerical results, as it provided an opportunity to visually highlight intersections and transit stops with higher crash frequencies. Using the same datasets and analyses results, a statistical analysis was also performed and will be discussed in the next chapter.

5.0 STATISTICAL ANALYSIS

5.1 Overview

This chapter describes the statistical analysis component of the research, specifically statistical test methodologies and their results. A logistic regression model, single-term deletion test, and chi-squared-based likelihood ratio tests were employed to formulate a model that would ultimately describe the effect any predictor variable had on the odds that a crash is VRU related. The logistic regression model was run again using the reduced model. Calculated odds ratios and their 95-percent confidence intervals and crash probabilities were used to describe the impact each predictor variable had on predicting the odds a crash is VRU related and to compare the effects of various predictor variables. The overall purpose of the analysis especially focuses on determining if the proximity of a crash to a transit stop impacts the probability that a crash is VRU related, accounting for all other factors. The response variable in the logistic regression model is whether a crash is VRU related or not. To evaluate spatial sensitivity, the analysis was conducted at two distinct spatial thresholds: one for crashes within 1,400 ft. of a transit stop and a second for crashes within 250 ft. of a transit stop. The following sections describe the logistic regression model, derived odds ratios, and crash probability estimates.

5.2 Logistic Regression Model

Logistic regression is a statistical method for predicting the probability of an event having one of two possible outcomes (Ramsey and Schafer, 2013a): For the purposes of this study, whether a crash is VRU related or not. Logistic regression calculates both odds and probability and both are reported in this chapter in Section 5.3 and Section 5.4. To evaluate spatial sensitivity, the analysis was performed on two data subsets: crashes occurring within 1,400 ft. of a transit stop and crashes occurring within 250 ft. of a transit stop. The logistic regression model used 116,386 (1,400 ft. analysis) or 54,427 (250 ft. analysis) crash observations generated from the crash analyses detailed in Section 4.2 and Section 4.3. Observations with missing spatial data were excluded from the logistic regression model to ensure the model only utilized verified infrastructure configurations. The code for the model at each spatial threshold is reported in Appendix E. The full logistic regression and reduced models are described in this chapter.

5.2.1 Full Model

The logistic regression model estimates the probability of a crash involving a VRU. The full model for both subsets incorporated the following as predictor variables:

- Crash to Stop Distance: The distance from the crash to the nearest transit stop, modeled with a natural spline with two degrees of freedom (noted in the results tables as *ns*(Crash to Stop Distance, 2))
- Speed Limit: The posted speed limit (mph) of the roadway segment where the transit stop is located, modeled with a natural spline with two degrees of freedom (noted in the results tables as *ns*(Speed Limit, 2))
- AADT: The AADT of the roadway where the transit stop is located
- Average Boarding: The average daily boarding volume of the transit stop
- Average Alighting: The average daily alighting volume at the transit stop
- Intersection Involved: If the crash is intersection related (binary indicator)
- Stop Location: The transit stop location relative to the nearest intersection (NS, FS, center (Cen), or MB)
- Stop Land Use: The transit stop land use (residential/mixed, commercial, agricultural, institutional, or recreational)
- Stop to Intersection Distance: The distance (measured in ft.) from the nearest transit stop to the nearest intersection
- Crash-to-Stop Distance: Stop Location: An interaction term capturing the effect of the crash-to-stop distance and the transit stop's location relative to the nearest intersection (noted in the table as *ns*(Crash to Stop Distance, 2): Stop Location)

To account for non-linear relationships between the dependent variable and continuous, numeric values in the crash-to-stop distance and speed limit variables, a natural spline (*ns)* with two degrees of freedom was employed. A natural spline is a piecewise polynomial function used to estimate the probability of the response (whether a crash is VRU related or not) for non-linear relationships in continuous predictor variables. Natural splines allow for a flexible functional form while maintaining global smoothness. The degrees of freedom of the spline controls the smoothness of the fit. Lower degrees of freedom produce smoother fits; larger degrees of freedom allow more variability in the curve but also increase the possibility of over-fitting and are less interpretable. Preliminary diagnostics indicated that linear modeling crash-to-stop distance and speed limit variables resulted in numerical instability and poor fit. Transitioning to the use of natural splines stabilized the model and provided a more robust estimation of the probability of VRU-related crashes. Variables transformed by a natural spline are noted in the results tables as *ns*(Variable, 2).

While five, three, and two degrees of freedom were evaluated, three degrees of freedom was initially preferred for its balance of simplicity and trend capture. However, due to model convergence issues at higher complexities, two degrees of freedom was ultimately selected. Two degrees of freedom provided the most robust fit while maintaining interpretability and producing results consistent with the results from other degrees of freedom.

To improve model stability, transit-stop land use types were consolidated into five categories based on functional similarity as follows:

- Residential/Mixed: includes single home, apartment, and mixed-use land use types
- Commercial: includes commercial and business land use types
- Agricultural: includes undeveloped, farming, and industrial land use types
- Institutional: includes government, school, church, hospital, library, and transit station land use types
- Recreational: includes recreational land use type

Results between the two distance thresholds were similar, so generally results are reported for the 1,400 ft. analysis, while the 250 ft. analysis results are included in Appendix F. The results of the logistic regression for the 1,400 ft. analysis are shown in Table 5.1.

The reference category from each group was omitted from the display of the model. The reference category for intersection involved was no. The reference category for transit stop location was NS. The reference category for transit-stop land use was commercial.

Table 5.1 Logistic Regression Results (1,400 ft. Analysis)

Variable	Estimate	Standard Error	P-Value
(Intercept)	-4.216	0.298	< 0.001 ***
ns(Crash to Stop Distance, 2)1	-0.624	0.179	< 0.001 ***
ns(Crash to Stop Distance, 2)2	-0.174	0.177	0.326
ns(Speed Limit, 2)1	1.745	0.503	< 0.001 ***
ns(Speed Limit, 2)2	-1.367	0.230	< 0.001 ***
AADT	< 0.001	< 0.001	< 0.001 ***
Average Boarding	0.003	< 0.001	< 0.001 ***
Average Alighting	0.000	< 0.001	0.858
Intersection Involved - Yes	0.313	0.036	< 0.001 ***
Stop Location - Cen	-0.145	0.408	0.722
Stop Location - FS	-0.110	0.082	0.181
Stop Location - MB	0.143	0.107	0.181
Stop Land Use - Residential/Mixed	-0.008	0.041	0.840
Stop Land Use - Institutional	-0.066	0.059	0.264
Stop Land Use - Agricultural	-0.427	0.089	< 0.001 ***
Stop Land Use - Recreational	0.199	0.117	0.088 ^
Stop to Intersection Distance	-0.001	0.000	< 0.001 ***
ns(Crash to Stop Distance, 2)1: Stop Location - Cen	-0.848	1.119	0.448
ns(Crash to Stop Distance, 2)2: Stop Location - Cen	-0.078	0.944	0.934
ns(Crash to Stop Distance, 2)1: Stop Location - FS	0.212	0.231	0.358
ns(Crash to Stop Distance, 2)2: Stop Location - FS	0.053	0.223	0.812
ns(Crash to Stop Distance, 2)1: Stop Location - MB	-0.789	0.282	0.005**
ns(Crash to Stop Distance, 2)2: Stop Location - MB	-0.527	0.264	0.046*

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Significant variables indicated a statistical difference in VRU crash probability of that variable compared to the reference category (or baseline) of that indicator variable. Furthermore, the significance of terms in the logistic regression results suggested that each factor provided unique predictive power above and beyond the other terms in the model, accounting for potential redundancy between indicator variables. Crash-to-stop distance, speed limit, roadway volume (measured as AADT), average boarding, intersection involved, agricultural transit-stop land use, and the interaction terms between crash-to-transit stop distance and MB transit stop locations were found to be significant based on the full model logistic regression. Notably, the interaction between distance and MB locations was significant, indicating that the spatial risk profile for MB transit stops differed from the NS reference category. To confirm the relative contribution of each factor to the model's explanatory power, a single-term deletion test was performed.

A single-term deletion test was utilized to assess the significance of each indicator variable in the model. The test evaluates which factors would make a statistically significant decrease in model fit if any one factor was removed from the model. A reduced model eliminates any insignificant contributors and protects against overfitting (Ramsey and Schafer, 2013a). The results of the single-term deletion model for the 1,400 ft. analysis are shown in Table 5.2.

Table 5.2 Single-Term Deletion Test Results (1,400 ft. Analysis)

Variable	Degrees of Freedom	AIC	P-Value
<none>		33578	
ns(Speed Limit, 2)	2	33618	< 0.001 ***
AADT	1	33590	< 0.001 ***
Average Boarding	1	33605	< 0.001 ***
Average Alighting	1	33576	0.857
Intersection Involved	1	33656	< 0.001 ***
Stop Land Use	4	33601	< 0.001 ***
Stop to Intersection Distance	1	33598	< 0.001 ***
ns(Crash-to-Stop Distance, 2): Stop Location	6	33589	0.001 ***

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

The results of this test indicated that all independent variables were significant apart from average alighting. Given the high correlation between average boarding and average alighting volumes at individual transit stops, the lack of significance for alighting likely stems from multicollinearity. Consequently, removing average alighting from the model yielded negligible changes in the overall fit and predictive power.

To justify the inclusion of specific indicator variables and assess the model's goodness of fit, the study employed an analysis of deviance resulting in the likelihood ratio test (also called a drop-in-deviance chi-squared test). This procedure measures the marginal contribution of individual predictors on the model's log-likelihood. The results indicate the statistical viability to use a simplified reduced model in place of a full model (Ramsey and Schafer, 2013a). The analysis of variance test compared the reduced model (which removes average alighting) against the full model (which includes all variables). The analysis of variance test results for both the

1,400 ft. and 250 ft. analyses are reported in Appendix F. In both the 1,400 ft. (p = 0.857) and the 250 ft. (p = 0.385) analyses tests, average alighting was not a significant predictor of crash probability. These high p-values indicate that, within the context of this model, alighting volumes do not provide unique explanatory power beyond what is captured by other variables. The reduced model, which excludes alighting volumes, can be used to determine the significance the remaining variables have on predicting VRU crash odds.

5.2.2 Reduced Model

The logistic regression test was repeated using the reduced model to produce the final coefficient estimates. The results for the 1,400 ft. analysis are shown in Table 5.3.

Table 5.3 Reduced Logistic Regression Model Results (1,400 ft. Analysis)

Variable	Estimate	Standard Error	P-Value
(Intercept)	-4.218	0.298	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1	-0.625	0.179	< 0.001 ***
ns(Crash-to-Stop Distance, 2)2	-0.174	0.177	0.326
ns(Speed Limit, 2)1	1.749	0.503	< 0.001 ***
ns(Speed Limit, 2)2	-1.366	0.230	< 0.001 ***
AADT	< -0.001	< 0.001	< 0.001 ***
Average Boarding	0.003	< 0.001	< 0.001 ***
Intersection Involved - Yes	0.313	0.036	< 0.001 ***
Stop Location - Cen	-0.151	0.407	0.710
Stop Location - FS	-0.110	0.082	0.181
Stop Location - MB	0.143	0.107	0.181
Stop Land Use - Residential/Mixed	-0.008	0.041	0.847
Stop Land Use - Institutional	-0.066	0.059	0.258
Stop Land Use - Agricultural	-0.427	0.089	< 0.001 ***
Stop Land Use - Recreational	0.199	0.117	0.088
Stop to Intersection Distance	-0.001	< 0.001	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	-0.848	1.119	0.449
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	-0.077	0.944	0.935
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	0.213	0.231	0.357
ns(Crash-to-Stop Distance, 2)2: Stop Location - FS	0.053	0.223	0.813
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	-0.789	0.282	0.005 **
ns(Crash-to-Stop Distance, 2)2: Stop Location - MB	-0.527	0.264	0.046 *

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Distance from crash to nearest transit stop, speed limit, roadway AADT, average boarding, intersection involvement, agricultural land use, transit-stop to intersection distance, and the interaction term between crash-to-stop distance and MB transit stop location were all significant compared to the reference category and other terms in the model. The reference category for intersection involved was no, the reference category for transit stop location was NS, and the reference category for transit stop land use was commercial.

The single-term deletion test was run again to verify whether every variable in the reduced model is significant. The results for the 1,400 ft. analysis are reported in Table 5.4 and the results for the 250 ft. analysis are reported in Appendix F. Every term in the reduced model is statistically significant at the 95-percent level, so the reduced model was the final model used. The final interpretations of the models were analyzed using odds ratios and probabilities.

Table 5.4 Single-Term Deletion Test Results – Reduced Model (1,400 ft. Analysis)

Variable	Degrees of Freedom	AIC	P-Value
<none>		33576	
ns(Speed Limit, 2)	2	33616	< 0.001 ***
AADT	1	33588	< 0.001 ***
Average Boarding	1	33639	< 0.001 ***
Intersection Involved	1	33654	< 0.001 ***
Stop Land Use	4	33599	< 0.001 ***
Stop to Intersection Distance	1	33596	< 0.001 ***
ns(Crash-to-Stop Distance, 2): Stop Location	6	33587	0.001 ***

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

5.3 Odds Ratios

The odds ratio indicates the proportional change in odds of a crash being VRU related compared to the reference category, holding all other variables in the model constant (Ramsey and Schafer, 2013b). A confidence interval that includes the value of 1 is statistically insignificant in the difference from the reference category. A confidence interval with values greater than 1 indicates a value increase in odds a crash is VRU related relative to the reference category. A confidence interval with values less than 1 indicates a value decrease in odds a crash is VRU related relative to the reference category. The odds ratios and their confidence intervals for the 1,400 ft. analysis are shown in Table 5.5. The results for the 250 ft. analysis are reported in Appendix F.

Table 5.5 Odds Ratios (1,400 ft. Analysis)

Variable	Odds Ratio	Confidence Interval 2.5%	97.5%
(Intercept)	0.015	0.008	0.026
ns(Crash-to-Stop Distance, 2)1	0.535	0.377	0.760
ns(Crash-to-Stop Distance, 2)2	0.840	0.589	1.180
ns(Speed Limit, 2)1	5.746	2.173	15.579
ns(Speed Limit, 2)2	0.255	0.161	0.398
AADT	1.000	1.000	1.000
Average Boarding	1.003	1.002	1.003
Intersection Involved - Yes	1.368	1.276	1.467
Stop Location - Cen	0.860	0.372	1.842
Stop Location - FS	0.895	0.762	1.053
Stop Location - MB	1.153	0.935	1.420
Stop Land Use - Residential/Mixed	0.992	0.915	1.075
Stop Land Use - Institutional	0.936	0.833	1.048
Stop Land Use - Agricultural	0.652	0.546	0.774
Stop Land Use - Recreational	1.221	0.964	1.523
Stop to Intersection Distance	0.999	0.999	1.000
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	0.428	0.046	3.819
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	0.926	0.119	5.117
ns(Crash-to-top Distance, 2)1: Stop Location - FS	1.237	0.787	1.945
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	1.054	0.683	1.638
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	0.454	0.262	0.790
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	0.591	0.351	0.990

The relationships involving crash-to-transit-stop distance and speed limit were nonlinear and used a natural spline to model their relationships. The use of a natural spline transformation complicates the direct interpretation of odds ratios; impacts of crash-to-stop distance and speed limit are instead presented in Section 5.4 using probabilities rather than odds ratios.

The odds ratios and their confidence intervals indicated that as AADT changes, the odds that a crash is VRU related remain the same. For every person per day increase in average boarding, odds that a crash is VRU related increased by 0.3 percent or 1.003 times. Crashes at an intersection were 1.368 times or 36.8 percent more likely to be VRU related than a crash away from an intersection. Relative to NS transit stops, Cen, FS, and MB stop locations were not found to be significant predictors of VRU crashes. In all three cases, the 95-percent confidence intervals for the odds ratio included 1.0, suggesting no statistically observable difference in crash odds compared to the reference category. Agricultural land was 0.652 times as likely to be VRU related or a 34.8 percent decrease in odds a crash is VRU related compared to a commercial land-use transit stop. For every 1-ft. increase in transit-stop to intersection distance, there was a 0.1 percent decrease in odds a crash is VRU related. Overall, the odds ratios indicated that VRU crash prediction is most influenced by the relationship of a crash to an intersection. Crashes are most likely to be VRU related when at an intersection. Similarly, the intensity of use (boarding) and location context (land use) matter more in predicting odds a crash is VRU related than the presence of a transit stop.

For the crash-to-transit-stop distance, the first spline term showed a significant negative association (odds ratio = 0.535, 95-percent confidence interval: 0.377-0.760), while the second term was not strictly significant (odds ratio = 0.840, 95-percent confidence interval: 0.589-1.180), together indicating a non-constant, decreasing trend in odds as distance increases. As the distance from the transit stop increases, the odds a crash is VRU related decreases, but not at a constant rate. The significance of a difference in odds a crash is VRU related is most pronounced in the interaction term with MB transit stops. This implies that for MB stops, the effect of distance is more pronounced as you move away from MB stops compared to other stop types. Crash-to-stop distance odds for MB stops is significantly lower when compared to NS stops. For the speed limit, odds initially increase (odds ratio value of 5.746), and then decrease (odds ratio value of 0.255), indicating a curved relationship.

A final summary table of the numeric results of each variable is reported in Table 5.6. Bolded variables indicate the odds are statistically significant from the variable reference category and other terms in the model. Differences between the distance thresholds occur in the crash-to-stop distance and the interaction term between crash-to-stop distance and transit stop location. This suggests that distance effects at the 1,400 ft. level were more prominent farther away from MB stops than other stop types. A difference in significance for crash-to-stop distance between the 250 ft. and 1,400 ft. analyses suggest the influence area of a transit stop is larger than the immediate 250 ft. vicinity and the 1,400 ft. analysis is preferred as transit stop proximity to a crash within 250 ft. is negligible.

Table 5.6 Summary Odds Ratio Results

Variable	Odds Ratio 250 ft.	Confidence Intervals 250 ft. 2.5%	Confidence Intervals 250 ft. 97.5%	Odds Ratio 1,400 ft.	Confidence Intervals 1,400 ft. 2.5%	Confidence Intervals 1,400 ft. 97.5%
(Intercept)	0.039	0.018	0.083	0.015	0.008	0.026
ns(Crash-to-Stop Distance, 2)1	0.970	0.532	1.786	0.535	**0.377**	**0.760**
ns(Crash-to-Stop Distance, 2)2	0.619	0.306	1.209	0.840	0.589	1.180
ns(Speed Limit, 2)1	1.049	0.309	3.720	5.746	**2.173**	**15.579**
ns(Speed Limit, 2)2	0.268	**0.135**	**0.515**	0.255	**0.161**	**0.398**
AADT	1.000	1.000	1.000	1.000	1.000	1.000
Average Boarding	1.004	**1.003**	**1.005**	1.003	**1.002**	**1.003**
Intersection Involved - Yes	1.272	**1.147**	**1.414**	1.368	**1.276**	**1.467**
Stop Location - Cen	0.179	0.010	1.566	0.860	0.372	1.842
Stop Location - FS	0.573	**0.371**	**0.886**	0.895	0.762	1.053
Stop Location - MB	0.773	0.450	1.309	1.153	0.935	1.420
Stop Land Use - Residential/Mixed	0.938	0.838	1.048	0.992	0.915	1.075
Stop Land Use - Institutional	0.858	0.727	1.007	0.936	0.833	1.048
Stop Land Use - Agricultural	0.581	**0.449**	**0.741**	0.652	**0.546**	**0.774**
Stop Land Use - Recreational	1.165	0.852	1.553	1.221	0.964	1.523
Stop to Intersection Distance	0.999	0.999	1.000	0.999	0.999	1.000
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	0.469	0.007	49.091	0.428	0.046	3.819
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	0.001	< 0.001	2.429	0.926	0.119	5.117
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	2.165	0.936	5.023	1.237	0.787	1.945
ns(Crash-to-Stop Distance, 2)2: Stop Location - FS	0.554	0.240	1.299	1.054	0.683	1.638
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	1.528	0.542	4.383	0.454	**0.262**	**0.790**
ns(Crash-to-Stop Distance, 2)2: Stop Location - MB	0.668	0.213	2.036	0.591	**0.351**	**0.990**

5.4 Crash Probability

The results of the logistic regression analysis were used to calculate probability based on specific indicator variables, holding all other variables in the model constant (Ramsey and Schafer, 2013b). The predicted probability a given crash is VRU related based on the distance a crash is from a transit stop is shown in Figure 5.1. As shown in Figure 5.1, the overall probability that a crash is VRU related based on distance from a transit stop is between 4.0 percent and 5.0 percent for the 1,400 ft. analysis. The trend generally shows VRU probability decreasing as distance increases, and levels off after about 1,000 ft. While crash-to-stop distance was statistically significant, the marginal change in predicted probability suggests a lack of practical significance for this variable at this distance.

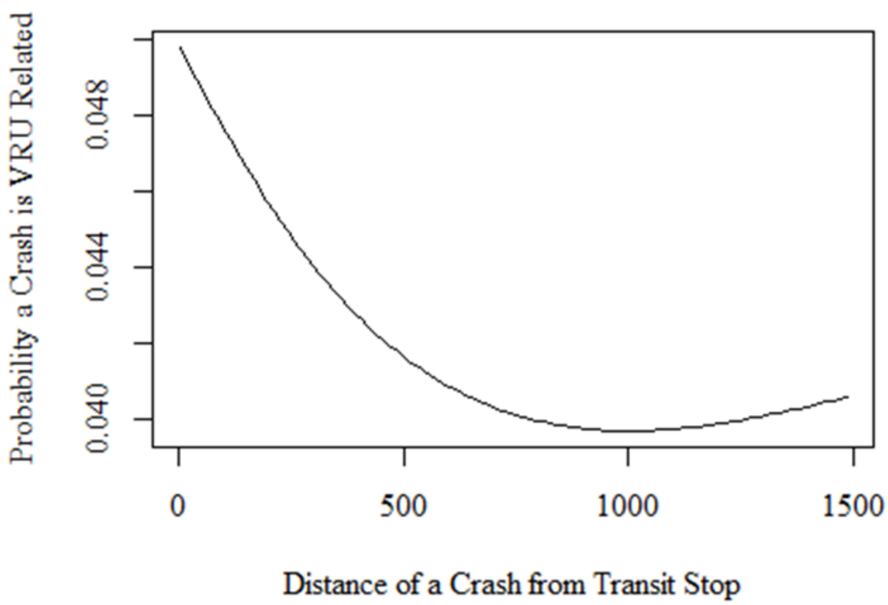


Figure 5.1 VRU crash probability – Stop distance (1,400 ft. analysis)

To create the probability graph, all other variables were held at a constant. The reference category or median values for the respective indicator were used for both distance thresholds. The variables for the 1,400 ft. analysis were set as:

- Speed Limit = 35 mph
- AADT = 20,000 vpd
- Average Boarding = 5 passengers per day (ppd)

- Average Alighting = 5 ppd
- Intersection Involved = Y
- Stop Location = FS
- Stop Land Use = Commercial
- Stop to Intersection Distance = 158.7 ft.

The overall probability that a crash is VRU related based on distance from a transit stop is approximately between 1.5 percent and 4.5 percent for the 250 ft. analysis as shown in Figure 5.2. The trend generally shows VRU probability increasing as distance increases, peaking at 150 ft., and dropping off as distance increases. A net change of 3.0 percent over 250 ft. is slightly more practically significant than the 0.8 percent change observed at the 1,400 ft. analysis; however, the marginal change in predicted probability still suggests an overall lack of practical significance for this variable.

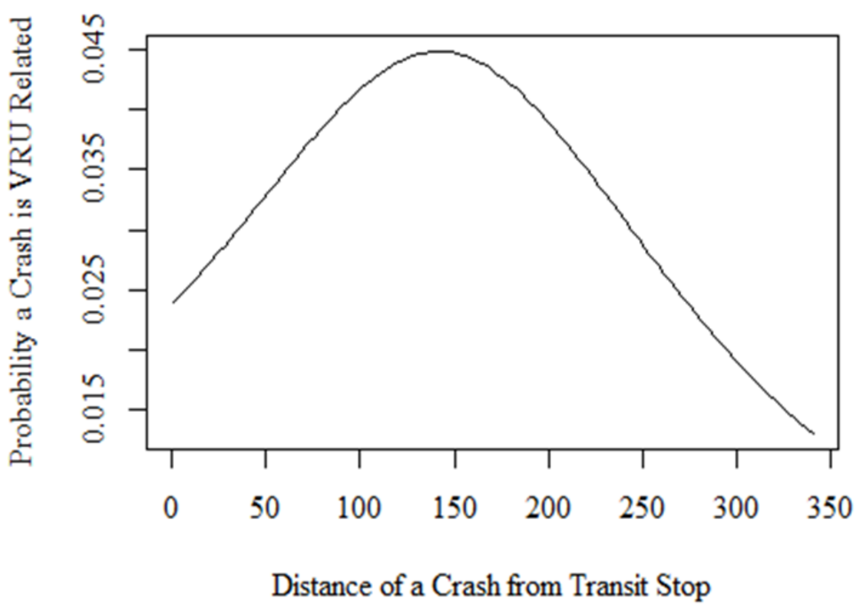


Figure 5.2 VRU crash probability – Stop distance (250 ft. analysis)

The reference category or median values for the respective indicator were used for both distance thresholds. The variables for the 250 ft. analysis were set as:

- Speed Limit = 40 mph
- AADT = 25,000 vpd

- Average Boarding = 5 ppd
- Average Alighting = 6 ppd
- Intersection Involved = Y
- Stop Location = FS
- Stop Land Use = Commercial
- Stop to Intersection Distance = 154.3 ft.

The overall probability that a crash is VRU related based on speed limit is approximately between 1.0 percent and 4.5 percent for the 1,400 ft. analysis as shown in Figure 5.3. The trend generally shows VRU probability increasing as speed limit increases, peaking at approximately 30 mph, and dropping off as speed limit increases. While speed limit was statistically significant, the marginal change in predicted probability suggests a lack of practical significance for this variable.

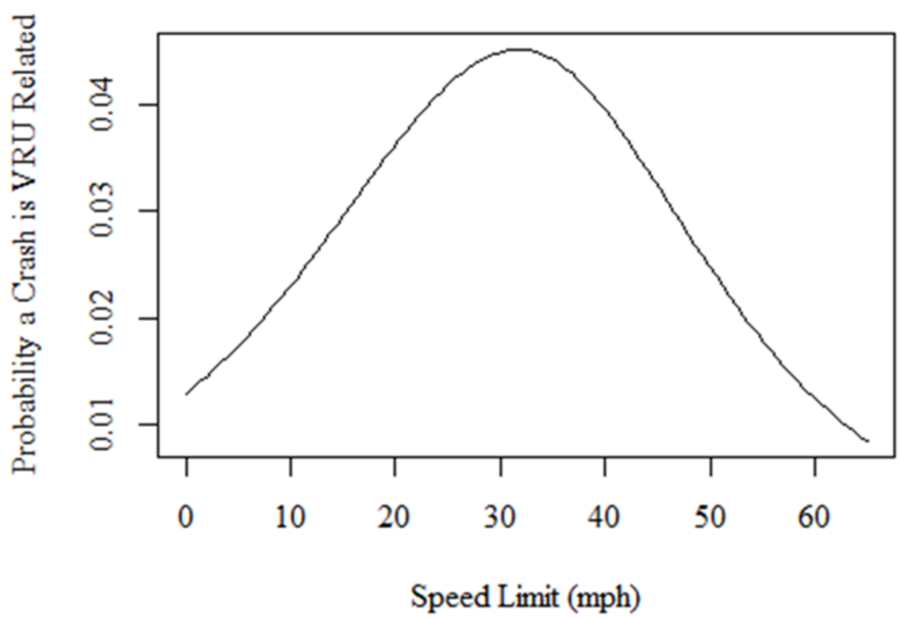


Figure 5.3 VRU crash probability – Speed limit (1,400 ft. analysis)

The reference category or median values for the respective indicator were used for both distance thresholds. The variables for the 1,400 ft. analysis were set as:

- Crash-to-Stop Distance = 268.8 ft.
- AADT = 20,000 vpd

- Average Boarding = 5 ppd
- Average Alighting = 5 ppd
- Intersection Involved = Y
- Stop Location = FS
- Stop Land Use = Commercial
- Stop to Intersection Distance = 158.7 ft.

The overall probability that a crash is VRU related based on speed limit is approximately between 2.0 percent and 5.0 percent for the 250 ft. analysis as shown in Figure 5.4. The trend generally shows VRU probability gradually increasing as speed limit increases, peaking at approximately 30 mph and dropping off as distance increases. While speed limit was statistically significant, the marginal change in predicted probability suggests a lack of practical significance for this variable.

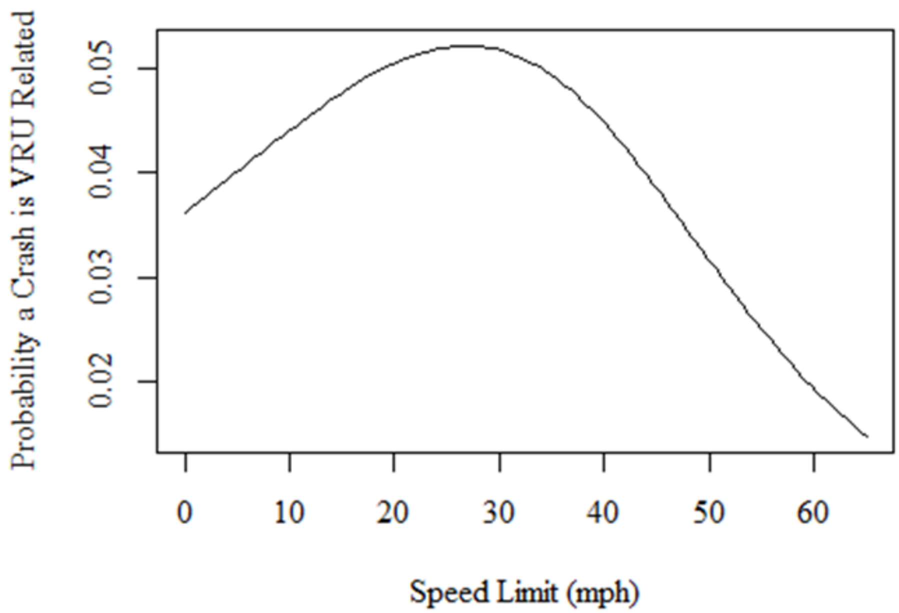


Figure 5.4 VRU crash probability – Speed limit (250 ft. analysis)

The reference category or median values or for the respective indicator were used for both distance thresholds. The variables for the 250 ft. analysis were set as:

- Crash-to-Stop Distance = 140.4 ft.
- AADT = 25,000 vpd

- Average Boarding = 5 ppd
- Average Alighting = 6 ppd
- Intersection Involved = Y
- Stop Location = FS
- Stop Land Use = Commercial
- Stop to Intersection Distance = 154.3 ft.

The probability a crash is VRU related based on transit stop location is shown in Figure 5.5 for the 1,400 ft. analysis. The mean probability a crash is VRU related for NS and FS locations is approximately 3.5 percent, whereas the mean probability for MB locations is slightly lower at approximately 3.0 percent. The mean probability a crash is VRU related for a Cen transit stop is approximately 2.0 percent.

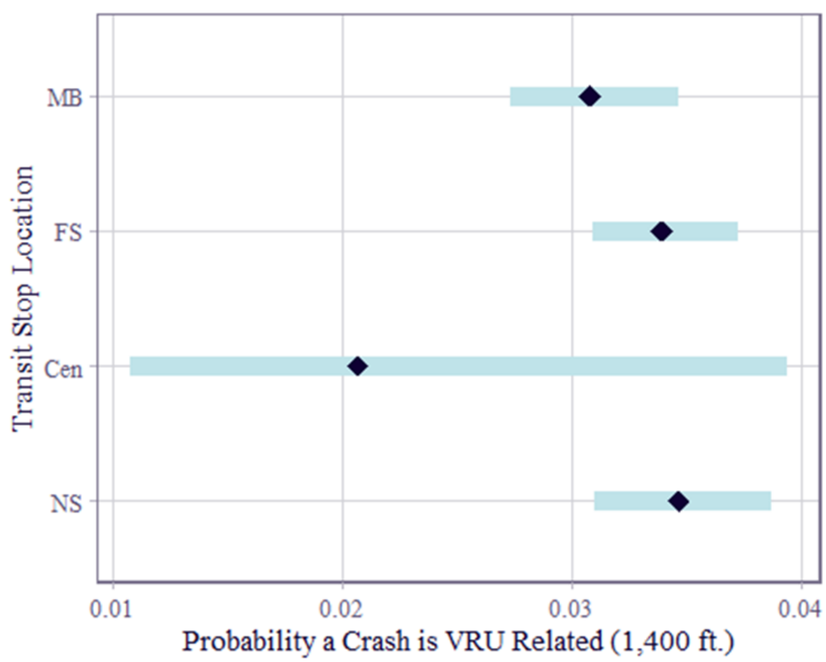


Figure 5.5 VRU-related crash probability – Transit stop location (1,400 ft. analysis)

Since the confidence intervals of all four location types overlap with each other, there is no statistically significant difference in probability a crash is VRU related based on transit stop location. The overall probabilities and marginal changes are low enough to suggest a lack of practical significance for this variable.

The probability a crash is VRU related based on transit stop location is shown in Figure 5.6 for the 250 ft. analysis. Figure 5.6 This figure shows the mean probability for MB, NS, and FS locations is approximately 4.0 percent. The mean probability a crash is VRU related for a Cen transit stop is approximately 2.5 percent. Since the confidence intervals of all four location types overlap with each other, there is no significant difference in probability a crash is VRU related based on transit stop location. The overall probabilities and marginal changes are low enough to suggest a lack of practical significance for this variable.

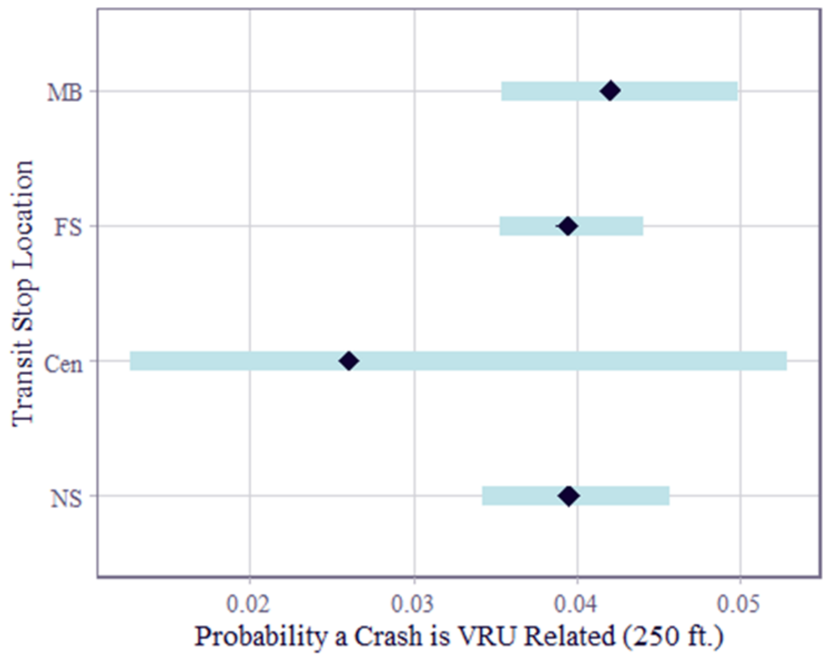


Figure 5.6 VRU-related crash probability – Transit stop location (250 ft. analysis)

The probability a crash is VRU related based on transit-stop land use for the 1,400 ft. analysis is shown in Figure 5.7. Figure 5.7. This figure shows the mean probability for institutional, residential/mixed use, and commercial land uses is approximately 3.0 percent. The mean probability a crash is VRU related for recreational land use is close to 4.0 percent, while the mean probability a crash is VRU related for agricultural land use is close to 2.0 percent. Since the confidence intervals of all land uses except for agricultural overlap with the reference category (commercial), there is no significant difference in probability a crash is VRU related from the reference category except for agricultural land use, which is about 1.0 percent lower.

The overall probabilities and marginal changes are low enough to suggest a lack of practical significance for this variable.

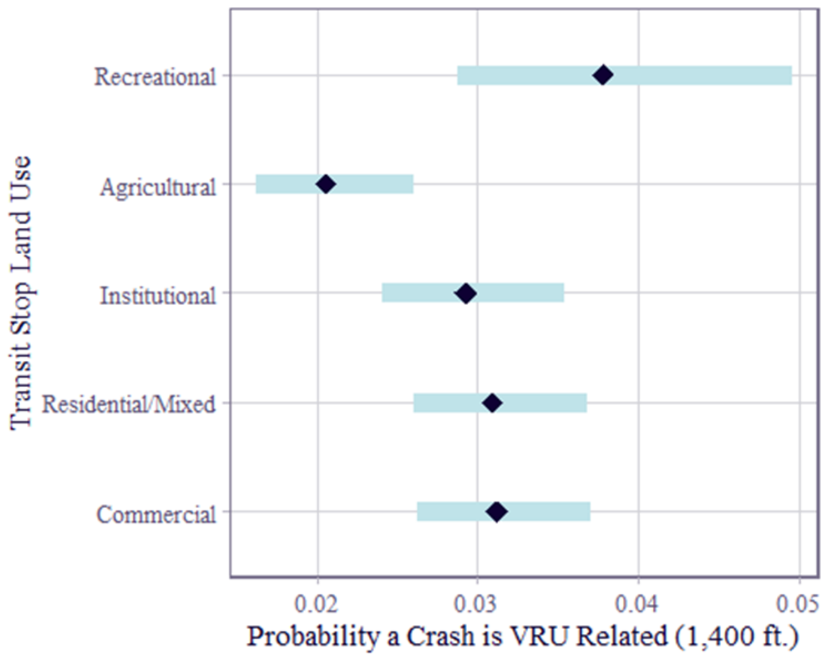


Figure 5.7 VRU-related crash probability – Transit-stop land use (1,400 ft. analysis)

The probability that a crash is VRU related based on transit-stop land use for the 250 ft. analysis is shown in Figure 5.8. Figure 5.8. This figure shows the mean probability for institutional, residential/mixed use, and commercial land uses is between 3.5 and 4.0 percent. The mean probability that a crash is VRU related for recreational land use is approximately 5.0 percent, while the mean probability that a crash is VRU related for agricultural land use is close to 2.5 percent. Since the confidence intervals of all land uses except for agricultural overlap with the reference category (commercial), there is no significant difference in probability a crash is VRU related from the reference category except for agricultural land use, which is about 1.5 percent lower. The overall probabilities and marginal changes are low enough to suggest a lack of practical significance for this variable.

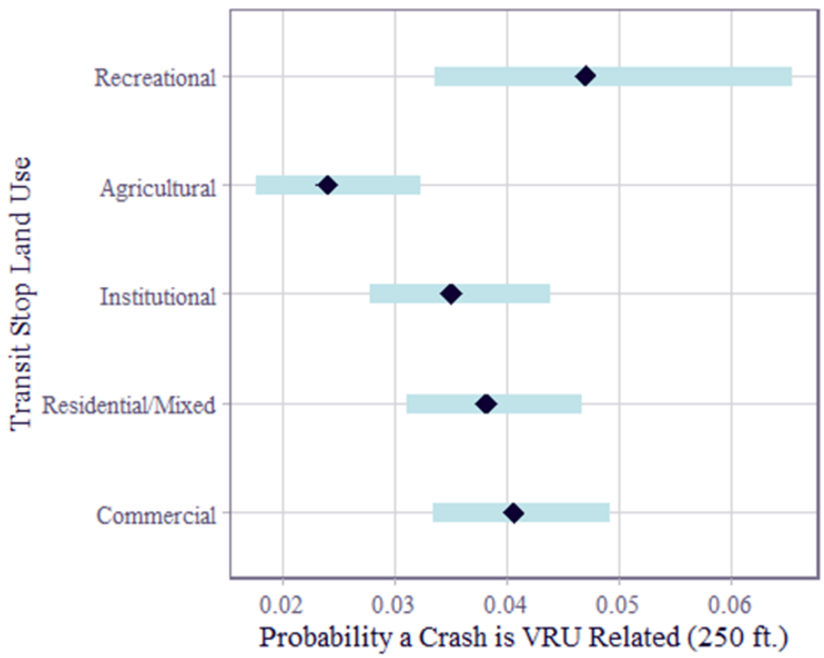


Figure 5.8 VRU-related crash probability – Transit-stop land use (250 ft. analysis)

A summary table of crash probabilities by their numeric values for transit stop location and transit-stop land use is reported in Table 5.7.

Table 5.7 Summary of VRU Crash Probabilities

Variable	Probability (250 ft.)	Probability (1,400 ft.)
Stop Location - NS	0.040	0.035
Stop Location - Cen	0.026	0.021
Stop Location - FS	0.039	0.034
Stop Location - MB	0.042	0.031
Stop Land Use - Commercial	0.041	0.031
Stop Land Use - Residential/Mixed	0.038	0.031
Stop Land Use - Institutional	0.035	0.029
Stop Land Use - Agricultural	0.024	0.021
Stop Land Use - Recreational	0.047	0.038

5.5 Summary

VRU crash risk is most heavily influenced by intersection involvement, with crashes occurring at intersections significantly more likely to involve a VRU. While spatial modeling at various thresholds confirms that the intensity of use (boarding) and location context (land use) are primary predictors, the physical presence of a transit stop, and the specific distance of a crash from that stop, exhibits a comparatively marginal effect on crash odds. Marginal changes in predicted probability suggest a lack of practical significance for crash-to-transit-stop distance, speed limit, transit stop location, or transit-stop land use. The maximum probability of a crash being VRU related remains relatively low, with a maximum probability of approximately 5.0 percent. **S**tatistical modeling results indicate that the specific characteristics of a stop are more influential than the proximity to a stop in predicting the odds of a crash being VRU related.

6.0 CONCLUSIONS

6.1 Summary

The purpose of this study was to evaluate, understand, and quantify the interaction of the transit system in the state of Utah as a function of traffic safety. Traffic safety is evaluated based on the observed discrepancy between total crash frequencies and VRU-related crash frequencies (Ulak et al., 2021). To evaluate a potential relationship between the presence of transit stops and VRU crashes, GIS analysis followed by logistic regression modeling was employed. The methodology, findings, and limitations and challenges encountered during the study are presented in this chapter.

6.2 Methodology

Data were collected from online geospatial databases. Crashes were collected from the AASHTOWare Safety database (Numetric, 2025). Transit stop data were collected from the UTA Open Data Portal (UTA, 2025). County boundaries that defined the study area and roads data were collected from the UGRC database (UGRC, 2025). Intersection data were collected from the WFRC open database (WFRC, 2025).

Crash data from January 1, 2019 to December 31, 2023 within the Wasatch Front transit area collected from the AASHTOWare Safety Database were used as the basis of the datasets. Crash data were filtered using GIS to capture datasets of crashes within two spatial thresholds for comparison: crashes within 250 ft. and crashes within 1,400 ft. of a transit stop.

Collected data were modeled in ArcGIS Pro. Road segments proximal to transit stops were isolated by clipping the network layer using a circular buffer centered on each transit stop. These segments were then buffered using rectangular geometry. Proximity analysis was performed to calculate the distance between crashes within the buffers and the nearest transit stop. The measured value (in ft.) was appended to the crash attribute table. This procedure was repeated for intersections by recording the distance from each crash and each transit stop to the nearest intersection. Crashes, transit stops, and intersections were spatially analyzed within a map environment to identify geographic patterns.

A logistic regression model was developed to analyze the relationships between predictor variables in the metadata of the crash datasets and VRU crash probability for each dataset. Logistic regression was employed to analyze the spatial data and determine the odds a crash was VRU related. The logistic regression model included several independent variables: crash-to-stop distance, speed limit, AADT, average boarding, average alighting, intersection involvement, stop location, stop land use, stop to intersection distance, and an interaction term between crash-to-stop distance and stop location. A single-term deletion test was utilized to assess the significance of each indicator variable in the model. Subsequently, a likelihood ratio test was used to confirm the statistical validity of a reduced model. The final reduced model was used to calculate the odds and probability of a crash being VRU related, controlling for all other factors.

6.3 Findings

The integration of GIS, visual diagnostics, and statistical modeling indicated that the environmental characteristics of a crash were far more influential than the proximity to a transit stop in predicting VRU-related crashes. Intersection involvement emerged as the most significant risk factor, with crashes at intersections significantly more likely to involve a VRU than those that occur MB.

While spatial modeling at various thresholds confirmed that boarding intensity and land use context were primary statistical predictors, their practical effect on crashes was marginal. Despite their statistical significance, the actual fluctuations in predicted probability remained narrow across speed limit and transit stop location. Even at peak risk levels, the probability of a crash being VRU related remained low, with a maximum predicted probability of approximately 5.0 percent. Similarly, the physical presence of a transit stop, and the specific distance of a crash from that stop, exhibited a marginal effect. Ultimately, these results suggest that while transit stops were markers for VRU activity, the specific physical and operational characteristics of the stop environment were more predictive of risk than the physical distance between a pedestrian and the transit stop.

A summary table of crash probabilities for transit stop location and transit-stop land use is reported in Table 6.1. The results demonstrate the low probability that a crash is VRU related based on stop location and land use. A maximum probability of less than 5.0 percent indicates that stop location and land use are practically insignificant in demonstrating a serious impact on VRU safety in the presence of transit.

Table 6.1 Summary of VRU Crash Probabilities

Variable	Probability (250 ft.)	Probability (1,400 ft.)
Stop Location - NS	0.040	0.035
Stop Location - Cen	0.026	0.021
Stop Location - FS	0.039	0.034
Stop Location - MB	0.042	0.031
Stop Land Use - Commercial	0.041	0.031
Stop Land Use - Residential/Mixed	0.038	0.031
Stop Land Use - Institutional	0.035	0.029
Stop Land Use - Agricultural	0.024	0.021
Stop Land Use - Recreational	0.047	0.038

The odds ratios of each of the indicator variables are summarized in Table 6.2, with predictors that differ significantly from their reference category in bold. While land use remains a statistically significant variable in the model, its practical impact varies by type. Recreational uses demonstrated a marginally higher probability of VRU-related crashes compared to other categories, while agricultural land use demonstrated a marginally lower probability of VRU-related crashes. Similarly, the presence of MB transit stops and closer proximity to a stop showed a positive correlation with VRU risk, though these increases in probability were notably low.

Table 6.2 Summary Odds Ratio Results

Variable	Odds Ratio 250 ft.	Confidence Intervals 250 ft. 2.5%	97.5%	Odds Ratio 1,400 ft.	Confidence Intervals 1,400 ft. 2.5%	97.5%
(Intercept)	0.039	0.018	0.083	0.015	0.008	0.026
ns(Crash-to-Stop Distance, 2)1	0.97	0.532	1.786	0.535	**0.377**	**0.76**
ns(Crash-to-Stop Distance, 2)2	0.619	0.306	1.209	0.84	0.589	1.18
ns(Speed Limit, 2)1	1.049	0.309	3.72	5.746	**2.173**	**15.579**
ns(Speed Limit, 2)2	0.268	**0.135**	**0.515**	0.255	**0.161**	**0.398**
AADT	1.000	1.000	1.000	1.000	1.000	1.000
Average Boarding	1.004	**1.003**	**1.005**	1.003	**1.002**	**1.003**
Intersection Involved - Yes	1.272	**1.147**	**1.414**	1.368	**1.276**	**1.467**
Stop Location - Cen	0.179	0.01	1.566	0.86	0.372	1.842
Stop Location - FS	0.573	**0.371**	**0.886**	0.895	0.762	1.053
Stop Location - MB	0.773	0.45	1.309	1.153	0.935	1.42
Stop Land Use - Residential/Mixed	0.938	0.838	1.048	0.992	0.915	1.075
Stop Land Use - Institutional	0.858	0.727	1.007	0.936	0.833	1.048
Stop Land Use - Agricultural	0.581	**0.449**	**0.741**	0.652	**0.546**	**0.774**
Stop Land Use - Recreational	1.165	0.852	1.553	1.221	0.964	1.523
Stop to Intersection Distance	0.999	0.999	1.000	0.999	0.999	1.000
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	0.469	0.007	49.091	0.428	0.046	3.819
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	0.001	< 0.001	2.429	0.926	0.119	5.117
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	2.165	0.936	5.023	1.237	0.787	1.945
ns(Crash-to-Stop Distance, 2)2: Stop Location - FS	0.554	0.24	1.299	1.054	0.683	1.638
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	1.528	0.542	4.383	0.454	**0.262**	**0.790**
ns(Crash-to-Stop Distance, 2)2: Stop Location - MB	0.668	0.213	2.036	0.591	**0.351**	**0.990**

6.4 Limitations and Challenges

Several challenges regarding data precision and roadway geometry, geospatial analysis and buffer constraints, attribution and transit relationships, and statistical modeling precision were encountered during the data collection and analysis portions of the research. While best efforts were made to mitigate these issues, the following limitations remain and should be considered when interpreting the results.

6.4.1 Data Precision and Roadway Geometry

The primary limitation of the project involved the spatial accuracy of the crash records. Crash location data were recorded by the responding police officer, frequently recording mile points rather than a GPS coordinate, resulting in an estimated error of up to 0.1 miles (528 ft). This margin of error was significant when compared to the high-precision shapefiles used for other geospatial layers, which maintained an error tolerance of only a few feet.

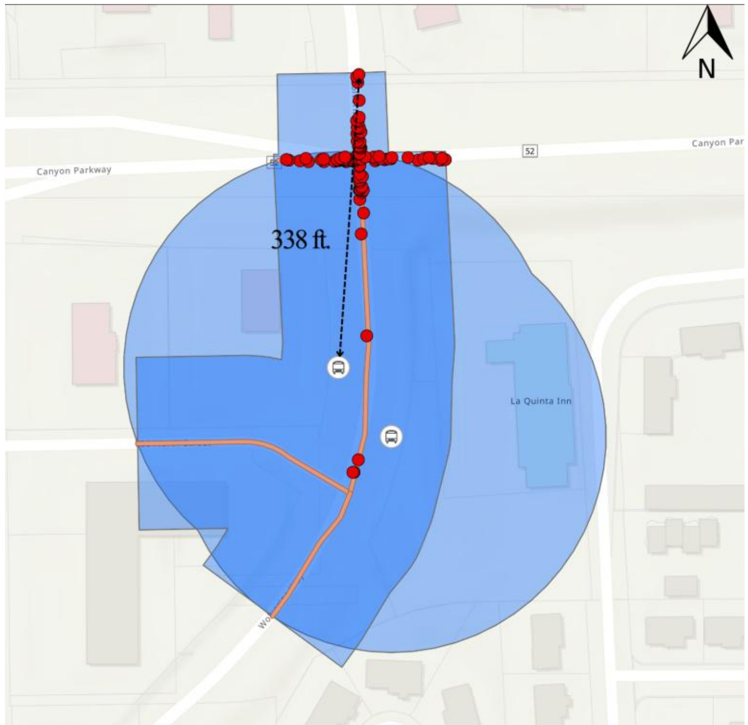
Furthermore, the data lacked directional travel indicators or details about the crash in relation to the roadway geometry. Consequently, it was not possible to determine if a crash occurred on a sidewalk, in a crossing, or within a travel lane. Additionally, the roadway shapefile was missing minor roads on the outskirts of the study area and intersection definitions were inconsistent, sometimes erroneously including driveways as valid intersections.

6.4.2 Geospatial Analysis and Buffer Constraints

The use of the "near" tool and buffer analysis introduced specific geometric inaccuracies including distance overruns, nearest neighbor conflicts, and network connection over attribution. Due to the nature of the geometric distance calculations used by the "near" tool and the width of buffers on adjacent streets, the 250 ft. analysis included crashes up to 340 ft. from a transit stop and the 1,400 ft. analysis captured crashes up to 1,500 ft. as demonstrated in Figure 6.1.

Nearest neighbor conflicts came from using the near tool in densely clustered data points. The "near" tool assigned each crash to the single closest transit stop. In areas where stops were densely clustered, this may artificially dilute crash counts for a specific stop by distributing them among neighboring points, even if the crashes were functionally related to a primary location.

Finally, because all buffers were dissolved to avoid duplicate counts, the analysis captured peripheral crashes. In the 1,400 ft. analysis, a crash could be attributed to a transit stop up to three streets away, even if physical obstacles like buildings or fences made a real-world pedestrian connection between the two locations impossible.



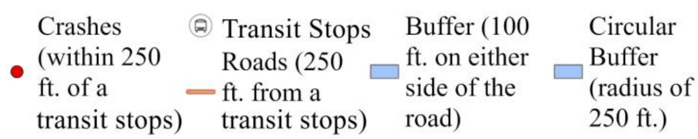


Figure 6.1 Crash-to-stop measurements (250 ft. analysis)

6.4.3 Attribution and Transit Relationship

Distinguishing between transit dependent and pass-by VRUs remains a methodological challenge. In the state of Utah, crash reports only categorize incidents as “transit-related” if a transit vehicle is physically involved; consequently, crashes involving boarding or alighting passengers were omitted from the “transit-related” category if no transit vehicle was struck. However, as noted by Rewalt, et al. (2025) the greatest risk to VRUs often occurs during travel to or from the stop, particularly at MB locations, which represented a relatively small portion of this dataset.

Because the data lacks this granular detail, the findings of this research represent the probability of a crash being VRU related based on proximity to transit, rather than a confirmed causal relationship. This limitation is widely acknowledged in existing literature (Anis et al., 2024; Hess et al., 2004; Pessaro et al., 2017; Pulugurtha and Penkey, 2010; Pulugurtha and Vanapalli, 2008; Salum et al., 2024; Singh et al., 2022; Ulak et al., 2021; Voorhees Transportation Center, 2012; Yu, 2024) as proximity does not always equal involvement.

6.4.4 Statistical Modeling Precision

Finally, the precision of the model fit was impacted by the requirements of model stability (especially affected by the quantity of data analyzed by the model: 116,386 crash records for the 1,400 ft. analysis or 54,427 crash records for the 250 ft. analysis). To ensure a stable output, the natural spline was limited to two degrees of freedom. While this improved the reliability of the logistic regression, it inherently reduced the precision of the fit compared to higher-degree models.

7.0 RECOMMENDATIONS AND IMPLEMENTATION

7.1 Overview

This chapter outlines recommendations and implementation suggestions based on the results and conclusions of this research. Proposed implementable engineering recommendations are included in this chapter to assist UDOT in evaluating transit stops and safety in relation to VRUs in line with the goal of Zero Fatalities (UDOT, 2026).

7.2 Recommendations

Recommendations in this section are based on the final conclusions of the project. Recommendations include specific VRU safety countermeasures, additional predictors, and future alternative research approaches.

7.2.1 VRU Safety Countermeasures

VRU crash risk was most heavily influenced by intersection involvement. The focus of countermeasures to improve pedestrian safety improvements should be at intersections and other VRU crossings. VRU intersection improvements may include pedestrian median refuge space (Bergman et al., 2002; Brilon and Blanke, 1993; Craig et al., 2019), aligning push buttons and crosswalks with desired pedestrian paths, and discouraging intersection-adjacent driveways (Singleton et al., 2023). Additional research on VRU safety at intersections is recommended. In and around transit stops, adequate MB crossings should be provided. Additional countermeasures like PHBs to assist with MB crossings should be considered as well. Pedestrian crossing improvements may benefit VRUs in the area, regardless of their transit usage (Rewalt, 2025).

7.2.2 Additional Predictors

Additional predictors in a regression model could add insight to the research. These predictors were previously left out of the statistical model to avoid overfitting. However, additional investigation of these predictors could gain additional insight. Crash-to-intersection distances and crash severity relationships could be added to a logistic regression analysis. Additional research relating previously proposed countermeasures (sidewalk installation, center islands, MB crossings, pedestrian signals, pedestrian barriers, dedicated transit lanes/bulb outs, and raised pedestrian crossings) impacting general VRU safety could also be analyzed to determine a best-fit countermeasure. Future research investigating VRU safety around intersections could be conducted given the high correlation between VRU crashes and intersection involvement.

7.2.3 Future and Alternative Research Directions

Future research directions could include alternative time periods and alternative research approaches. Crash data in this research was taken for the years 2019-2023. A more expansive dataset including a wider time period, a dataset excluding COVID-19 data, or a more recent dataset could be investigated in future research. The research objective is rooted in crash and VRU crash statistics from the last 41 years (NHTSA, 2024), so long-range time comparisons could be useful for validating those claims.

Alternative research approaches could study near-miss crash data near transit stops, hierarchical crash clustering analyses, and ArcGIS Pro geospatial statistical analysis tools including “Getis-Ord Gi*,” “Moran’s I,” “network analyst,” “density-based clustering,” and “kernel density estimation.” Additional research could use machine learning architectures or k-fold cross validation (Ramsey and Schafer, 2013c) techniques on the existing data set and other crash data sets to improve model fit and predictive power.

7.3 Implementation

The primary implementations suggested include further investigation of intersection safety, identifying intersection and transit stop safety improvement priorities, general VRU safety roadway improvements, and crash-data quality improvements.

The research findings indicated that general VRU safety improvements at intersections around transit stops yielded the highest VRU safety impact. The primary recommendation to

UDOT is to investigate VRU safety around intersections based on the correlation between VRU crashes and intersection involvement found in this research.

It is recommended that the UDOT Traffic and Safety Division sort transit stops and intersections by crash frequency, especially focusing on high-boarding stops to determine priorities for further VRU pedestrian safety improvements at nearby intersections and frequent crossing locations. GIS data collected as part of this project along with Python coding can be used to rank transit stops based on crash frequencies as outlined in Section 4.4.2 and Appendix D.

Other improvements based on the findings could include roadway improvements that impact VRU safety. Ensuring that adequate MB crossings were available near transit stops appeared significant in the data analysis and could improve safety. General VRU improvements that demonstrated the most impact in the literature included sidewalk installation, center island installation, pedestrian refuge island installation, MB crossings, pedestrian signals, pedestrian barriers, and raised pedestrian crossings. These measures are recommended as determined locally appropriate by the UDOT Traffic and Safety Division.

The final recommendation is to improve the precision of crash records. Future research involving crash data would benefit from any crash-data-quality remediation resulting from working with the police officers to improve crash location accuracy, including gathering GPS data rather than milepoint data.

REFERENCES

Aguero-Valverde, J. 2013. “Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates.” *Accident Analysis & Prevention*, 50, 289–297.

Ali, M.S., A.E. Kitali, J.H. Kodi, P. Alluri, and T. Sando. 2021. “Safety impacts of transit signal priority using a full Bayesian approach.” *Transportation Research Record*, 2675(11), 1189–1204.

AASHTO (American Association of State Highway and Transportation Officials). 2010. *Highway safety manual*. 1st ed. Washington, D.C.

Anis, M., S.R. Geedipally, and D. Lord. 2024. “Pedestrian crash causation analysis near bus stops: Insights from random parameters NB-Lindley models.” *arXiv preprint arXiv:2410.22253*.

Arun, A., C. Lyon, T. Sayed, S. Washington, F. Loewenherz, D. Akers, G. Ananthanarayanan, Y. Shu, M. Bandy, and M. Haque. 2023. “Leading pedestrian intervals – yay or nay? A before-after evaluation of multiple conflict types using an enhanced non-stationary framework integrating quantile regression into Bayesian hierarchical extreme value analysis.” *Accident Analysis & Prevention*, 181.

Bahar, G., M. Parkhill, E. Hauer, F. Council, B. Persaud, C. Zegeer, R. Elvik, A. Smiley, and B. Scott. 2007. “Prepare parts I and II of a Highway Safety Manual: Knowledge base for part II.” Unpublished material from NCHRP Project 17-27.

Barabino, B., M. Bonera, G. Maternini, A. Olivo, and F. Porcu. 2021. “Bus crash risk evaluation: An adjusted framework and its application in a real network.” *Accident Analysis & Prevention*, 159 106258.

Bergman, A.B., B. Gray, J.M. Moffat, E.S. Simpson, and F.P. Rivara. 2002. “Mobilizing for pedestrian safety: an experiment in community action. (Special Report).” *Injury Prevention*, 8(4).

Brilon, W., and H. Blanke. 1993. “Extensive traffic calming: Results of the accident analyses in 6 model towns.” *ITE 1993 Compendium of Technical Papers*, Institute of Transportation Engineers, Washington D.C., 119-123.

Britt, J.W., A.B. Bergman, and J. Moffat. 1995. “Law enforcement, pedestrian safety, and driver compliance with crosswalk laws: Evaluation of a four-year campaign in Seattle.” *Transportation Research Record*, 1485 160–167.

Brodeur-Ouimet, P. and M.S. Cloutier. 2024. “Walkable and safe route to transit for pedestrians in greater Montréal: An environmental justice approach.” *Transportation Research Record*, 2678(9), 562–573.

Brüde, U., and J. Larsson. 2000. “What roundabout design provides the highest possible safety?” *Nordic Road Transport Res*., 2 (2000), pp. 17-21

Chen, P. and J. Zhou. 2016. “Effects of the built environment on automobile-involved pedestrian crash frequency and risk.” *Journal of Transport & Health*, 3(4), 448–456.

Craig, C.M., N.L. Morris, V.H. Ron, and D. Mayou. 2019. “Pedestrian safety and driver yielding near public transit stops.” *Transportation Research Record*, 2673(1), 514–523.

Cui, H., J. Dong, M. Zhu, X. Li, and Q. Wang. 2022. “Identifying accident black spots based on the accident spacing distribution.” *Journal of Traffic and Transportation Engineering (English Edition)*, 9(6), 1017–1026.

Daniels, S., T. Brijs, E. Nuyts, and G. Wets. 2009. “Injury crashes with bicyclists at roundabouts: influence of some location characteristics and the design of cycle facilities.” *Journal of Safety Research*, 40(2), 141–148.

Daniels, S., E. Nuyts, and G. Wets. 2008. “The effects of roundabouts on traffic safety for bicyclists: An observational study.” *Accident Analysis & Prevention*, 40(2), 518–526.

De Brabander, B., and L. Vereeck. 2007. “Safety effects of roundabouts in Flanders: Signal type, speed limits and vulnerable road users.” *Accident Analysis & Prevention*, 39(3), 591–599.

Elvik, R. 2008. “A survey of operational definitions of hazardous road locations in some European countries.” *Accident Analysis & Prevention*, 40(6), 1830–1835.

Esri. 2023. *ArcGISPro* (Version 3.6) [Computer software]. Esri, Redlands, CA.

FHWA (Federal Highway Administration). 2001. “Pedestrian facilities users guide-providing safety and mobility.” FHWA-RD-01-102, Washington, DC. Accessed May 5, 2026. https://www.fhwa.dot.gov/publications/research/safety/01102/01102.pdf.

FHWA (Federal Highway Administration). 2008a. “Pedestrian safety guide for transit agencies.” *National Transportation Library*. FHWA-SA-07-017. Accessed May 5, 2026. https://staging-dot.cdc.gov/view/dot/37657.

FHWA (Federal Highway Administration). 2008b. “Toolbox of countermeasures and their potential effectiveness for pedestrian crashes.” Accessed May 5, 2026. https://safety.fhwa.dot.gov/ped_bike/tools_solve/ped_tctpepc/.

FHWA (Federal Highway Administration). 2021. “Leading pedestrian interval.” Accessed May 8, 2026. https://highways.dot.gov/safety/proven-safety-countermeasures/leading-pedestrian-interval.

Fitzpatrick, K., E.S. Park, M.J. Cynecki, M.P. Pratt, and M. Beckley. 2021. “Investigation of crashes at pedestrian hybrid beacons: Results of a large-scale study in Arizona.” *J. Saf. Res*., 78 59–68.

Gelman, A., J. Hwang, and A. Vehtari. 2014. “Understanding predictive information criteria for Bayesian models.” *Stat Comput* 24, 997–1016. Accessed May 8, 2026. https://doi.org/10.1007/s11222-013-9416-2.

Geyer, J., N. Raford, T. Pham, and D. Ragland. 2006. “Safety in numbers: Data from Oakland, California.” *Transportation Research Record*, 1982, 150-154.

Guadamuz, R., V.V. Gayah, and R. Paleti. 2020. “Impact of bus routes on crash frequency in metropolitan areas.” *Transportation Research Record*, 2674(3), 305–316.

Habib, M.F., R. Bridgelall, D. Motuba, and B. Rahman. 2023. “Exploring the robustness of alternative cluster detection and the threshold distance method for crash hot spot analysis: A study on vulnerable road users.” *Safety*, 9 (3): 57. Accessed May 8, 2026. https://doi.org/10.3390/safety9030057.

Harwood, D.W., K.M. Bauer, K.R. Richard, D.K. Gilmore, J.L. Graham, I.B. Potts, D.J. Torbic, and E. Hauer. 2008. “Pedestrian safety prediction methodology.” *NCHRP Project 17-26.* Transportation Research Board, Washington, D.C.

Hess, P. M., A.V. Moudon, and J.M. Matlick. 2004. “Pedestrian safety and transit corridors.” *Journal of Public Transportation,* 7(2), 73–93.

Hydén, C., and A. Várhelyi. 2000. “The effects on safety, time consumption and environment of large scale use of roundabouts in an urban area: a case study.” *Accident Analysis & Prevention*, 32(1), 11–23.

Jacobsen, P.L. 2003. “Safety in numbers: more walkers and bicyclists, safer walking and bicycling.” *Injury Prevention*, 9(3), 205-209.

Jovanis, P. P., J.L. Schofer, P. Prevedouros, and K. Tsunokawa. (1991). “Analysis of bus transit accidents: empirical, methodological, and policy issues.” *Transportation Research Record*, 1322, Large Vehicle Safety: Transit and Trucks 1991, (1322).

Kim, K., I.M. Brunner, E.Y. Yamashita, G. Nagamine, and J. Burke. 2010 “Guilty by association: A spatial analysis of bus stop locations and pedestrian accidents.” Transportation Research Board 89th Annual Meeting, (CD-ROM), Washington DC, United States.

Kodi, J.H., A.E. Kitali, M.S. Ali, P. Alluri, and T. Sando. 2021. “Estimating safety impacts of adaptive signal control technology using a Full Bayesian approach.” *Transportation Research Record*, 2675(11), 1168-1179. Accessed May 8, 2026. https://doi.org/10.1177/03611981211025281.

Larimer County Urban Area Street Standards. 2015. “Transfort bus stop design standards and guidelines Appendix J.” Transfort Bus Stop Design Standards and Guidelines, 16.

Leden, L. 2002. “Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario.” *Accident Analysis & Prevention*, 34(4), 457–464.

Li, L., and Y. Zhang. 2007. “Bayesian approach based on Geographic Information System to identify hazardous roadway segments for traffic crashes.” *Transportation Research Record*, 2024(1), 63–72.

Li, P., M. Abdel-Aty, and J. Yuan. 2021. “Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on GPS trajectory data.” *Accident Analysis & Prevention*, 150, 105924.

Loo, B.P.Y., and K.H. Tsoi. 2024. “Stressors for bus commuters and ways of improving bus journeys.” *Transportation Research Part A: Policy and Practice*, 187, 104164.

Loukaitou-Sideris, A., R. Liggett, H.G. Sung. 2007. “Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles.” *Journal of Planning Education and Research*. 26, 338-351.

Ma, J., Kockelman, K.M., and P. Damien. 2008. “A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods.” *Accident Analysis & Prevention*, 40(3), 964–975.

Miranda-Moreno, L.F., P. Morency, and A.M. El-Geneidy. 2011. “The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections.” *Accident Analysis & Prevention*, 43(5), 1624–1634.

Moudon, A.V., L. Lin, J. Jiao, P. Hurvitz, and P. Reeves. 2011. “The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington.” *Accident Analysis & Prevention*, 43(1), 11–24.

National Safety Council. 2024. “Deaths by transportation mode.” Accessed April 30, 2026. https://injuryfacts.nsc.org/home-and-community/safety-topics/deaths-by-transportation-mode/.

NHTSA (National Highway Traffic Safety Administration). 2024. “Traffic safety facts: 2022 data - pedestrians.” Rep. No. DOT HS 813 590, U.S. Dept. of Transportation, Washington, DC. Accessed May 25, 2025. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813590.

Numetric. 2025. *AASHTOWare Safety Database*. Accessed April 28, 2025. https://udot.aashtowaresafety.com/.

Okabe, A., K. Okunuki, and S. Shiode. 2006. “The SANET toolbox: New methods for network spatial analysis.” *Transactions in GIS*, 10(4), 535–550.

Okabe, A., and K. Sugihara. 2012. “Spatial analysis along networks.” John Wiley & Sons, Chichester, United Kingdom. Accessed May 8, 2026. http://doi.org/10.1002/9781119967101.ch3.

Park, E.S., and D. Lord. 2007. “Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity.” *Transportation Research Record*, 2019(1), 1–6.

Pessaro, B., M. Catalá, Z. Wang, and M. Spicer. 2017. “Impact of transit stop location on pedestrian safety.” Final Rep. BDV25-977-22. Florida Dept. of Transportation, Tallahassee, FL. Accessed May 8, 2026. https://rosap.ntl.bts.gov/view/dot/34939/dot_34939_DS1.pdf.

Phillips, R.O., O.H. Hagen, and S.H. Berge. 2021. “Bus stop design and traffic safety: An explorative analysis.” *Accident Analysis & Prevention*, 153, 105917.

Pulugurtha, S.S., and E.N. Penkey. 2010. “Assessing use of pedestrian crash data to identify unsafe transit service segments for safety improvements.” *Transportation Research Record*, 2198(1), 93–102.

Pulugurtha, S.S., and V.R. Sambhara. 2011. “Pedestrian crash estimation models for signalized intersections.” *Accident Analysis & Prevention*, 43(1), 439–446.

Pulugurtha, S.S., and V.K. Vanapalli. 2008. “Hazardous bus stops identification: An illustration using GIS.” *Journal of Public Transportation*, 11(2), 65–83.

R Foundation for Statistical Computing. 2025. *R: A language and environment for statistical computing* (Version 2026.01.1+403). [Computer Program]. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey, F.L., and D.W. Schafer. 2013a. "Chapter 20: Logistic regression for binary response variables." *The Statistical Sleuth: A Course in Methods of Data Analysis*, 3rd Ed., Brooks/Cole, Cengage Learning, Boston, MA, 601-626.

Ramsey, F.L., and D.W. Schafer. 2013b. "Chapter 18: Comparisons of proportions or odds." *The Statistical Sleuth: A Course in Methods of Data Analysis*, 3rd Ed., Brooks/Cole, Cengage Learning, Boston, MA, 549-566.

Ramsey, F.L., and D.W Schafer. 2013c. "Chapter 12: "Strategies for variable selection." *The Statistical Sleuth: A Course in Methods of Data Analysis*, 3rd Ed., Brooks/Cole, Cengage Learning, Boston, MA, 345-382.

Rewalt, A. 2025. "Using fatal crash data to analyze nationwide trends in pedestrian safety at bus stops." Master's thesis, Univ. of Tennessee, Knoxville, TN. Accessed May 8, 2026. https://trace.tennessee.edu/utk_gradthes/13901.

Rewalt, A., A. Thomson, C. Brakewood, and C. Cherry. 2025. "Identifying research priorities to improve safety for pedestrians and bicyclists accessing bus stops." Rep. No. 23UTK01, *Center for Pedestrian and Bicyclist Safety*, Univ. of Tennessee, Knoxville, TN.

Rivara, F.P., C.L. Booth, A.B. Bergman, L.W. Rogers, and J. Weiss. 1991. "Prevention of pedestrian injuries to children: effectiveness of a school training program." *Pediatrics*, 88(4), 770–775.

Salum, J.H., N. Kisabanzira, P. Alluri, and T. Sando. 2024. "Toolkit for the assessment of safety at bus stops." *Transportation Research Record*, 2678(2), 279–292.

Saneinejad, S., and J. Lo. 2015. "Leading pedestrian interval: Assessment and implementation guidelines." *Transportation Research Record*, 2519(1), 85–94.

Salt Lake City. N.d. *Salt Lake City GIS Open Database*. Accessed June 20, 2025. https://gis-slcgov.opendata.arcgis.com/.

Schoon, C., and J. van Minnen. 1993. “Ongevallen op rotondes II: tweede onderzoek naar de onveiligheid van rotondes vooral voor fietsers en bromfietsers.” *Stichting Wetenschappelijk Onderzoek Verkeersveiligheid* (SWOV), rapport R-93-16, Leidschendam.

Singh, M., Y. Zhang, W. Cheng, Y. Li, and E. Clay. 2022. “Effect of transit-oriented design on pedestrian and cyclist safety using bivariate spatial models.” *Journal of Safety Research*, 83, 152–162.

Singleton, P., A. Mandolakani, A. Subedi, and M. Mekker. 2024. “Pedestrian safety and traffic operations near transit stops.” Mountain Plains Consortium, MPC 24-550. Accessed May 8, 2026. https://www.mountain-plains.org/research/details.php?id=590.

Song Y. and D. Noyce. 2019 “Effects of transit signal priority on traffic safety: Interrupted time series analysis of Portland, Oregon, implementations.” *Accident Analysis & Prevention*, Vol. 123. Accessed May 8, 2026. https://doi.org/10.1016/j.aap.2018.12.001.

Stone, J.R., K. Chae, and S. Pillalamarri. 2002. “The effects of roundabouts on pedestrian safety.” *Southeastern Transportation Center*, working paper, Tennessee.

Strauss, J., S. Zangenehpour, L.F. Miranda-Moreno, and N. Saunier. 2017. “Cyclist deceleration rate as surrogate safety measure in Montreal using smartphone GPS data.” *Accident Analysis & Prevention*, 99, 287–296.

Truong, L.T., and S.V.C. Somenahalli. 2011. “Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops.” *Journal of Public Transportation*, 14(1), 99–114.

UGRC (Utah Geospatial Resource Center). 2025. *Utah Geospatial Resource Center Open Data Portal.* Accessed 25 April 2025. https://gis.utah.gov/.

Ukkusuri, S., L.F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. 2012. “The role of built environment on pedestrian crash frequency.” *Safety Science*, 50(4), 1141–1151.

Ulak, M.B., A. Kocatepe, A. Yazici, E.E. Ozguven, and A. Kumar. 2021. "A stop safety index to address pedestrian safety around bus stops." *Safety Science*, 133 105017.

UDOT (Utah Department of Transportation). 2021. *2021 update of crash costs based on severity*. Salt Lake City, Utah.

UDOT (Utah Department of Transportation). 2026. *Utah Strategic Highway Safety Plan* (SHSP). Accessed May 4, 2026. https://shsp.udot.utah.gov/.

UTA (Utah Transit Authority). 2025. *UTA Open Data Portal*. Accessed April 25, 2025. https://maps.rideuta.com/portal/apps/sites/#/uta-open-data/search?collection=Dataset&q=UTA%20SERVICE%20DATA.

Van Cauwenberg, J., P. Clarys, I. De Bourdeaudhuij, A. Ghekiere, B. de Geus, N. Owen, and B. Deforche. 2018. "Environmental influences on older adults' transportation cycling experiences: A study using bike-along interviews." *Landscape Urban Plann*., 169, 37–46.

Van Houten, R., J.L. Malenfant, R. Blomberg, B. Huitema, and J. Hochmuth. 2017. "High-visibility enforcement on driver compliance with pedestrian right-of-way laws: 4-year follow-up." *Transportation Research Record*. 2660: 58–65.

Voorhees Transportation Center. 2012. "Pedestrian safety at bus stops." Rutgers Univ., New Brunswick, NJ. Accessed May 8, 2026. https://vtc.rutgers.edu/wp-content/uploads/2014/07/Pedestrian-Safety-at-Bus-Stops_FINAL09192012.pdf.

Walgren, S. 1998. "Using geographic information system (GIS) to analyze pedestrian accidents." *68th Annual Meeting of the Institute of Transportation Engineers*, Toronto, ON, Canada.

Wang, Y., and K.M. Kockelman. 2013. "A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods." *Accident Analysis & Prevention*, 60 71–84.

WFRC (Wasatch Front Regional Council). 2025. *Wasatch Front Regional Council Open Database.* Accessed 25 April 2025. https://wfrc.maps.arcgis.com/home/gallery.html?sortField=relevance&sortOrder=desc.

Webster, D.C. and A.M. Mackie. 1996. "Review of traffic calming schemes in 20 mph zones." *Transportation Research Laboratory Report*. Berkshire, UK. 215 (41).

Wier, M., J. Weintraub, E.H. Humphreys, E. Seto, and R. Bhatia. 2009. "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning." *Accident Analysis & Prevention*, 41(1), 137–145.

Yang, C.Y.D. 2007. "Trends in transit bus accidents and promising collision countermeasures." *Journal of Public Transportation*, 10(3), 119–136.

Yu, C.Y. 2024. "The impact of built environments on pedestrian safety around bus stops." *J. Urban Affairs*, 46(2), 267–280. Accessed May 8, 2026. https://doi.org/10.1080/07352166.2022.2049447.

Zegeer, C.V., C.T. Esse, J.R. Stewart, H.F. Huang, and P. Lagerwey. 2004. "Safety analysis of marked versus unmarked crosswalks in 30 cities." *ITE Journal*, 74, 34–41.

Zegeer, C., J. Stewart, H. Huang, and P. Lagerwey. 2001. "Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Analysis of pedestrian crashes in 30 cities." *Transportation Research Record*. 1773: 56–68.

Zegeer, C., J. Stewart, H. Huang, P. Lagerwey, J. Feaganes, and B. Campbell. 2005. "Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Final report and recommended guidelines." FHWA-HRT-04-100.

Zhou, H., and S. Bromfield. 2007. "Moving the bus back into traffic safely – Signage and lighting configuration phase I." National Center for Transit Research (U.S.).

APPENDIX A: INTERSECTION DISTANCE METHODOLOGY

This appendix contains the intersection distance methodology documents, using ArcGIS Pro Version 3.6. These are step-by-step explanations of the steps followed for the analysis in this report that any user can implement with their own data. Specific settings used in this analysis are described in each step.

This document is intended to provide directions for how to perform a linear distance from intersections analysis in ArcGIS Pro. The datasets used for the analysis are from the AASHTOWare Safety Database, UGRC, WFRC, and UTA; although the template is written to be used as a general guide, where the user can input their own data.

This document is written under the assumption that the user is familiar with ArcGIS Pro and the ModelBuilder function within the application. The outline for this appendix will list the datasets used including their shapefile types, followed by the data analysis with step-by-step instructions.

A.1 Datasets Used

The following list contains the datasets used in this analysis, along with their respective shapefile type in the parentheses:

- Crashes (point)
- Roads (line)
- Intersections (point)
- Transit Stops (point)

Crashes and roads datasets were previously clipped to a general study area, defined as the relevant portions of the counties in Utah containing a UTA transit stop. All datasets were also projected to the same geographic coordinate system (GCS). WGS 1984 UTM Zone 12N for Central and Northern Utah was depicted in these instructions.

A.2 Data Analysis

Seven steps were conducted for the data analysis:

- Step 1: Project All Datasets
- Step 2: Buffer Intersections
- Step 3: Intersect Roads and Intersections
- Step 4: Buffer Roads
- Step 5: Intersect Crashes and Roads Buffer
- Step 6: Intersect Transit Stops and Roads Buffer
- Step 7: Distance Calculations

Details for each of these steps are provided in the following sections. The workflow for the analysis is shown in Figure A.1.

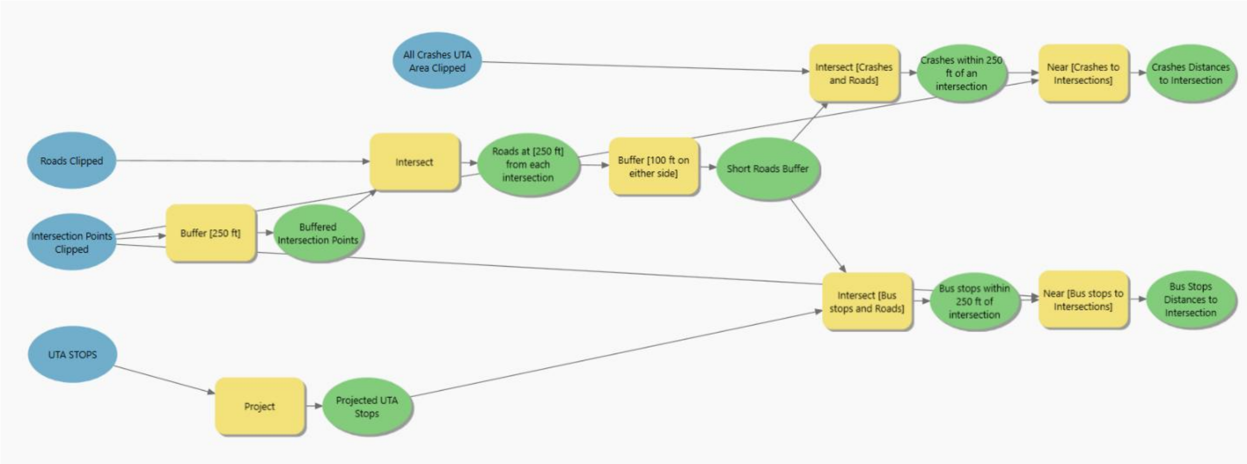


Figure A.1 Workflow for intersection analysis (Esri, 2023)

<u>Step 1: Project All Datasets</u>

The datasets need to be in a consistent projection, if not, the “project” tool can be used to project all datasets into an appropriate coordinate system. The tool was used in this report to project the UTA transit stops to the WGS 1984 UTM Zone 12N projection. Figure A.2 displays the project tool setting.

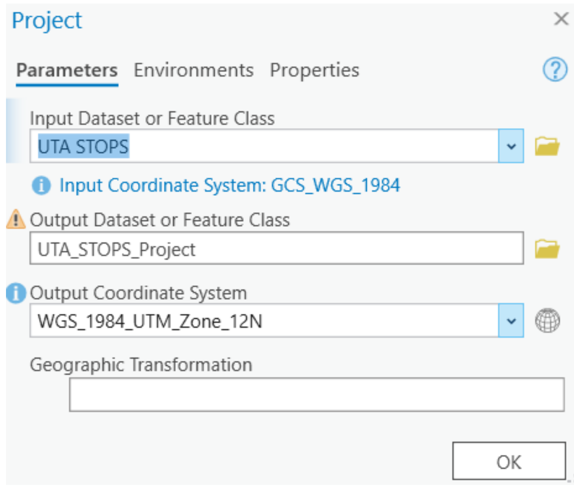


Figure A.2 Project tool settings (Esri, 2023)

<u>Step 2: Buffer Intersections</u>

The next step was to "buffer" all intersections. The buffer radius chosen should be the distance along the roadway to be investigated. A buffer of 250 ft. was used in this report. The *Distance* should be set to the distance specified by the investigation (e.g., 250 US Survey ft.). The *Type* was set to "geodesic" and *Dissolve* was set to "dissolve features using the listed fields' unique values or combination of values." All other fields were left at default settings, as shown in Figure A.3.

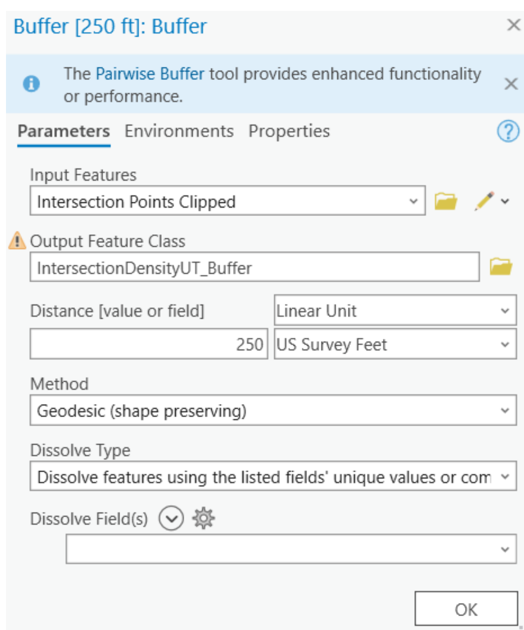


Figure A.3 Buffer tool settings (Esri, 2023)

Step 3: Intersect Roads and Intersections

The next step was to "intersect" the roads shapefile and the buffered intersections. This gave a road length of 250 ft. in either direction from each transit stop. For the *Input Feature* field, the roads shapefile was selected, and as the *Extent* feature, the buffered intersection points from the previous step were used. Select "line" as the *Output Type*. All other settings were set to the settings shown in Figure A.4. The output to this step was a line type feature class.

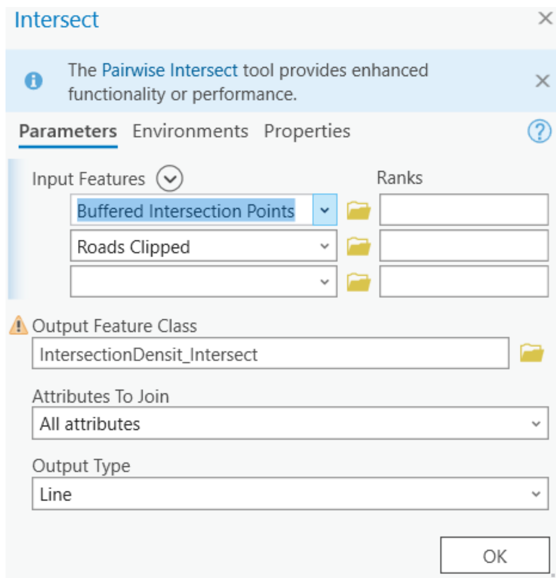


Figure A.4 Intersect roads and intersection settings (Esri, 2023)

Step 4: Buffer Roads

The next step was to "buffer" the roads a specific width from the road centerline to capture the data. A width of 100 ft. was used in this report. The *Distance* should be set to the desired distance away from the centerline (e.g., 100 US Survey ft.). The *Type* was set to "geodesic," and the *Dissolve* setting was set to "dissolve features using the listed fields' unique values or combination of values." The *End Type* was set to "flat." All other fields were left at default settings, as shown in Figure A.5. The output of this step is a polygon buffer around each road along each transit stop.

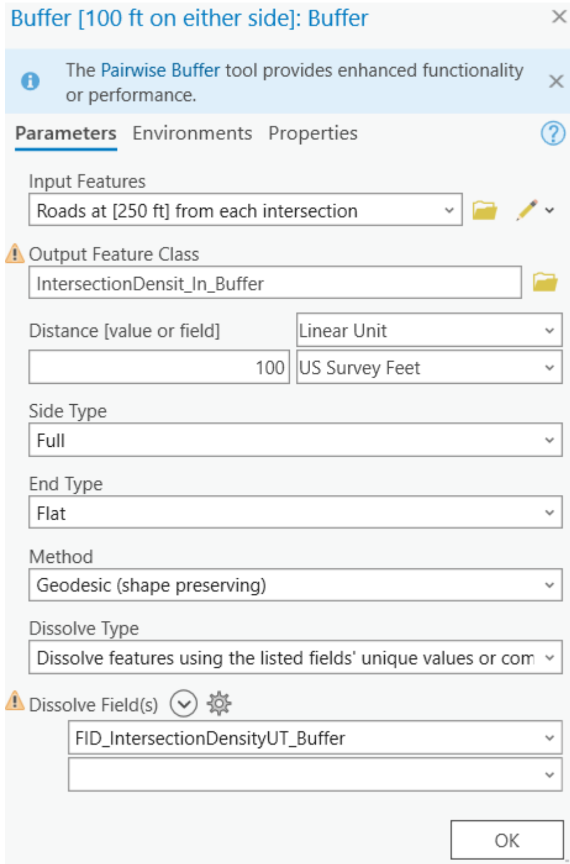


Figure A.5 Roads buffer settings (Esri, 2023)

<u>Step 5: Intersect Crashes and Roads Buffer</u>

The next step was to "intersect" the crashes with the roads buffer from the previous step. The *Input Feature* was set to the crashes shapefile, and the *Extent* feature was set to the roads buffer from the previous step. The "point" option was selected as the *Output Type*. All other settings were set to the settings shown in Figure A.6.

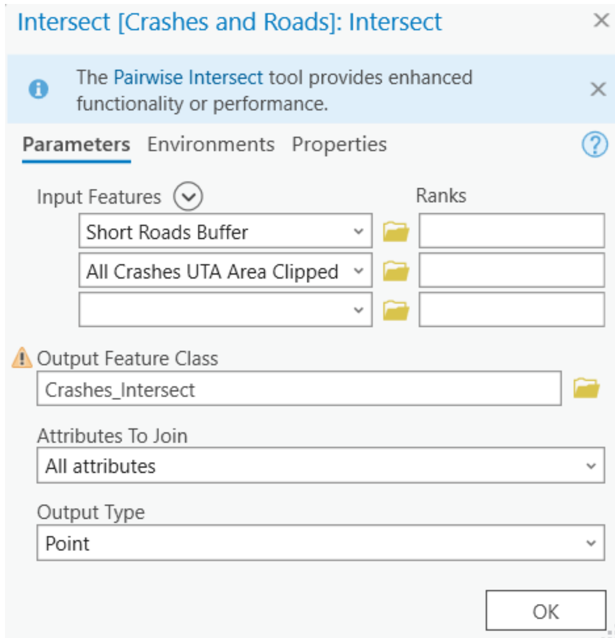


Figure A.6 Intersection tool settings for crashes and roads intersection (Esri, 2023)

After setting adjustments were completed and the dialog box closed, the resulting crashes feature class was added to the map display by the "add to display" setting. This feature class was the one shown on the map and further analyzed through its attribute table.

<u>Step 6: Intersect Transit Stops and Roads Buffer</u>

The next step was using the "intersect" tool with the transit stops shapefile from Step 6: Intersect Transit Stops and Roads Buffer and the roads buffer from the previous step. As the *Input Feature*, the transit stop shapefile was selected, and as the *Extent* feature, the roads buffer from the previous step was selected. The option "point" was selected as the *Output Type*. All other settings were set to match Figure A.7. Figure A.8 shows a screenshot of a map with the crashes and transit stops 1,400 ft. away from an intersection.

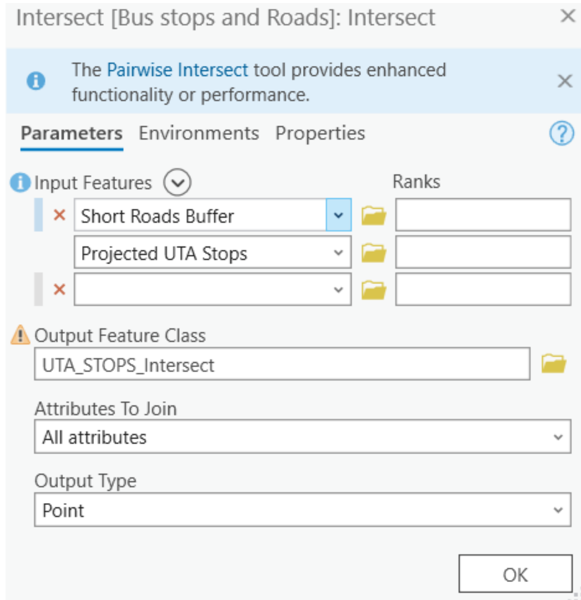


Figure A.7 Transit stops and roads intersection (Esri, 2023)

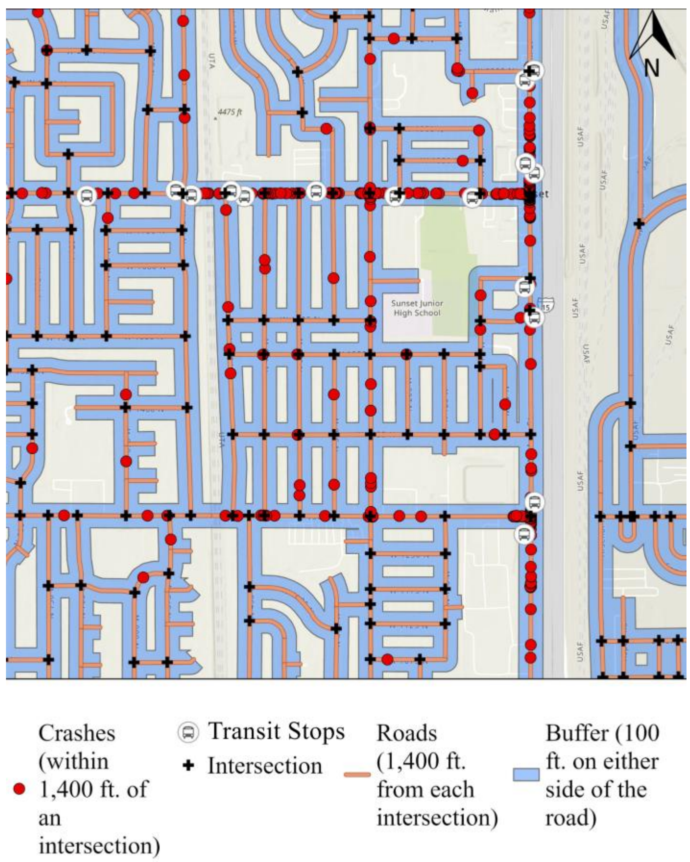


Figure A.8 View of intersection buffers for 1,400 ft. as shown in ArcGIS Pro (Esri, 2023)

After setting adjustments were completed and the dialog box closed, the resulting transit stops feature class was added to the map display by the "add to display" setting. This feature class was the one also shown on the map and analyzed from its attribute table.

Step 7: Distance Calculations

The "near" tool was used to calculate the distances of each crash from the nearest intersection and each transit stop from the nearest intersection. As the *Input Feature*, the crashes feature class from the previous step was selected, and as the *Near* feature, the intersections from Step 1: Project All Datasets were selected. If the shapefiles did not need to be projected, the original shapefiles were used instead. All other settings were set to match Figure A.9.

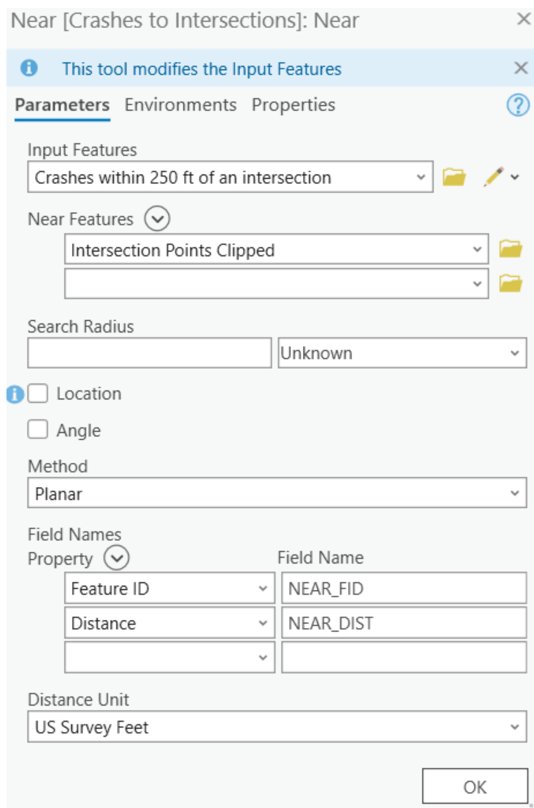


Figure A.9 Near tool settings for crashes (Esri, 2023)

The same was done in the "near" tool for the transit stops from Step 6: Intersect Transit Stops and Roads Buffer. As the *Input Feature*, the transit stops feature from the previous step was selected, and as the *Near* feature, the intersections from Step 1: Project All Datasets were selected. Similarly, if the shapefiles did not need to be projected, the original shapefiles were used instead. All other settings were set to match Figure A.10. The sample of the final output map is shown in Figure A.11.

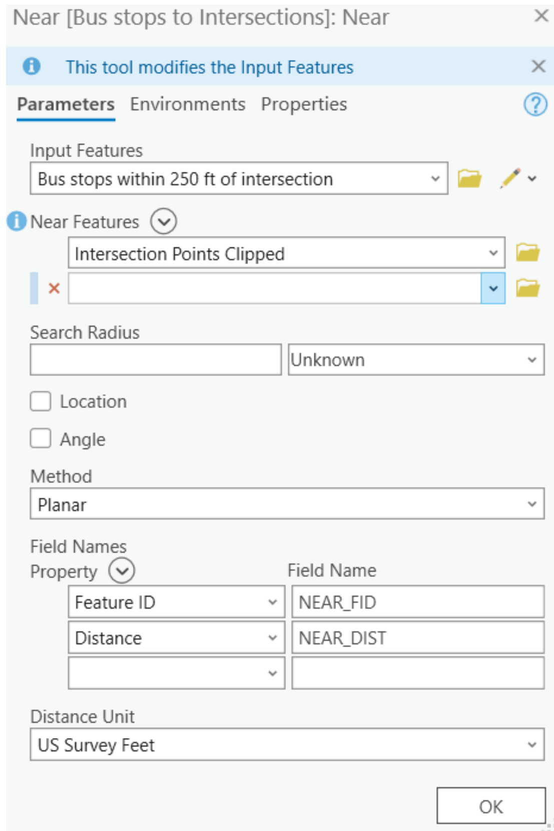


Figure A.10 Near tool settings for transit stops (Esri, 2023)

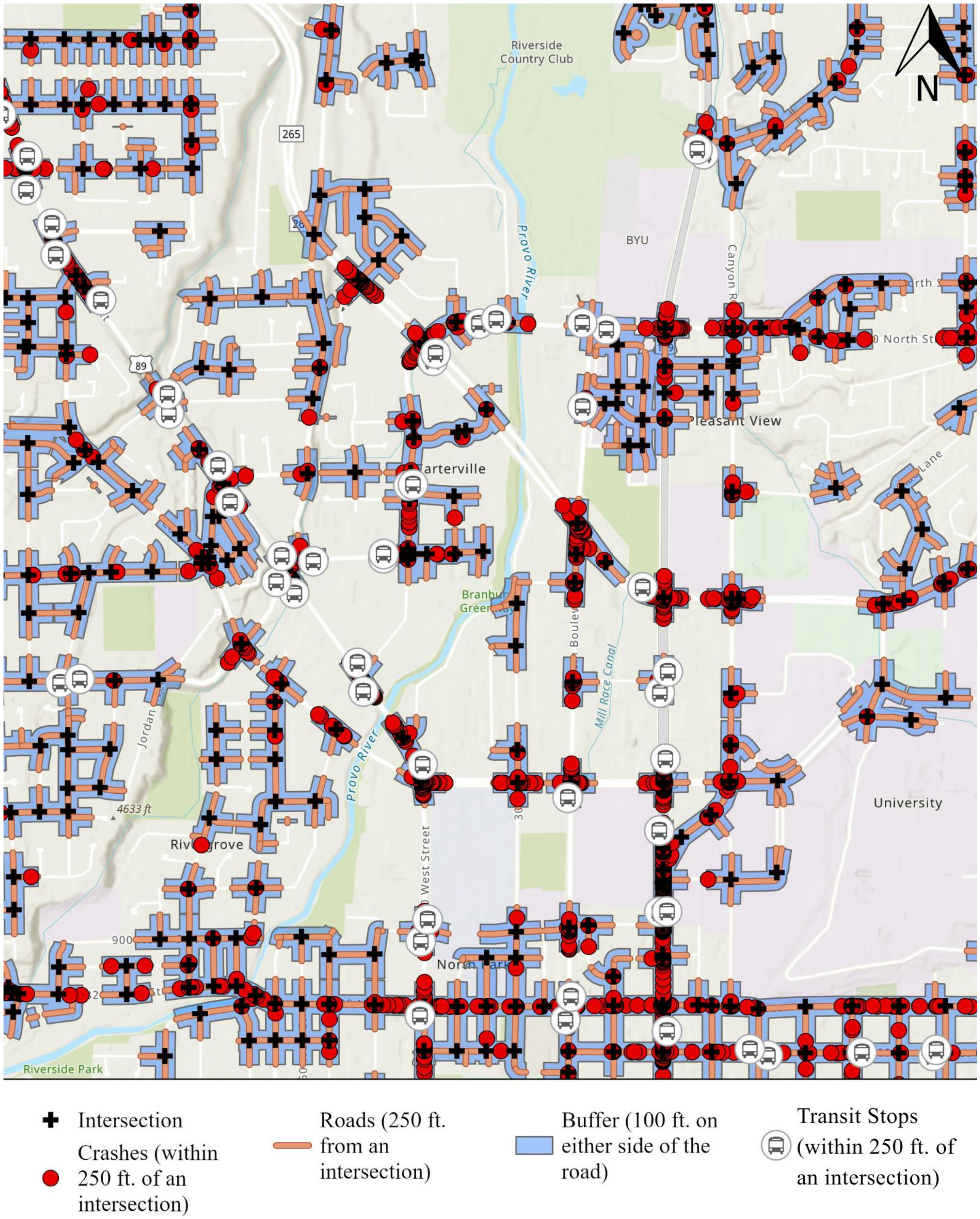


Figure A.11 Sample map after the intersection analysis is completed (Esri, 2023)

APPENDIX B: TRANSIT STOPS DISTANCE METHODOLOGY

This appendix contains the transit stop distance methodology documents, using ArcGIS Pro Version 3.6. These are step-by-step explanations of the steps followed for the analysis in this report that any user can implement with their own data. Specific settings used in this analysis are described in each step.

This document is intended to provide directions for how to perform a linear distance from transit stops analysis in ArcGIS. The datasets used for the analysis are from the AASHTOWare Safety Database, UGRC, and UTA; although the template is written to be used as a general guide, where the user can input their own data.

This document is written under the assumption that the user is familiar with ArcGIS Pro and the ModelBuilder function within the application. The outline for this appendix will list the datasets used including their shapefile types, followed by the data analysis with step-by-step instructions.

B.1 Datasets Used

The following list contains the datasets used in this analysis, along with their respective shapefile type in the parentheses. The following datasets are used:

- Crashes (point)
- Roads (line)
- Transit Stops (point)

Crashes and roads datasets were previously clipped to a general study area, defined as the relevant portions of the counties in Utah containing a UTA transit stop. All datasets were also projected to the same GCS. WGS 1984 UTM Zone 12N for Central and Northern Utah was depicted in these instructions.

B.2 Data Analysis

Six steps were conducted for the data analysis:

- Step 1: Project All Datasets
- Step 2: Buffer Transit Stops
- Step 3: Intersect Roads and Transit Stops
- Step 4: Buffer Roads
- Step 5: Intersect Crashes and Roads Buffer
- Step 6: Distance Calculations

Details for each of these steps are provided in the following sections. The workflow for the analysis is shown in Figure B.1.

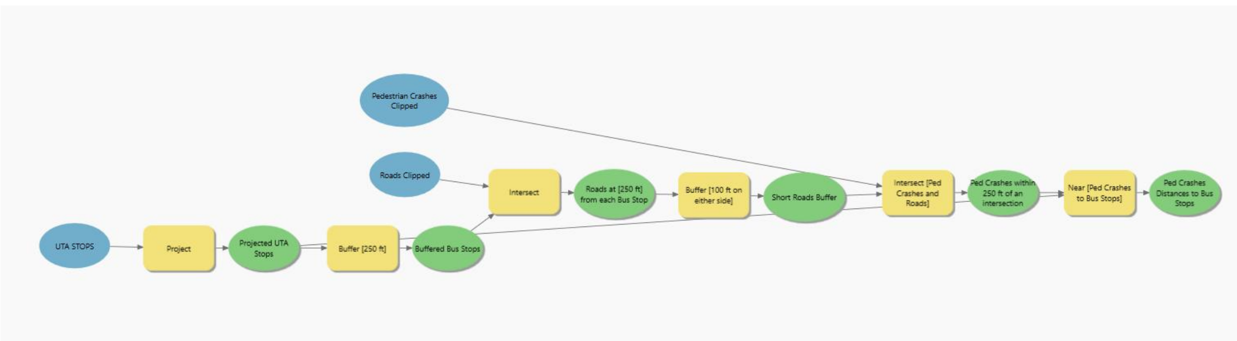


Figure B.1 Workflow for transit stop analysis (Esri, 2023)

<u>Step 1: Project All Datasets</u>

The datasets need to be in a consistent projection, if not, the "project" tool can be used to project all datasets into an appropriate coordinate system. The tool was used in this report to project UTA transit stops to the WGS 1984 UTM Zone 12N projection. Figure B.2 displays the project tool setting.

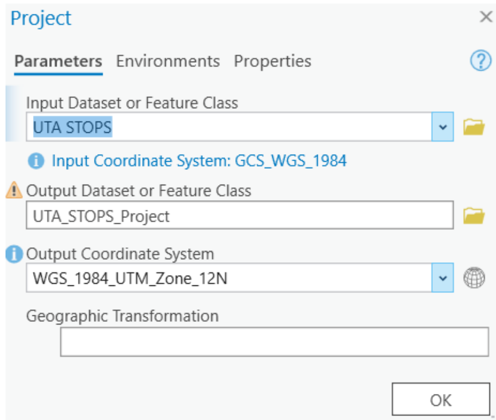


Figure B.2 Project settings (Esri, 2023)

<u>Step 2: Buffer Transit Stops</u>

The next step was to “buffer” all transit stops. The buffer radius chosen should be the distance along the roadway to be investigated. A buffer of 250 ft. was used in this report. The *Distance* should be set to the distance specified by the investigation (e.g., 250 US Survey ft.). The *Type* was set to “geodesic” and *Dissolve* was set to “dissolve features using the listed fields’ unique values or combination of values.” All other fields were left at default settings, as shown in Figure B.3.

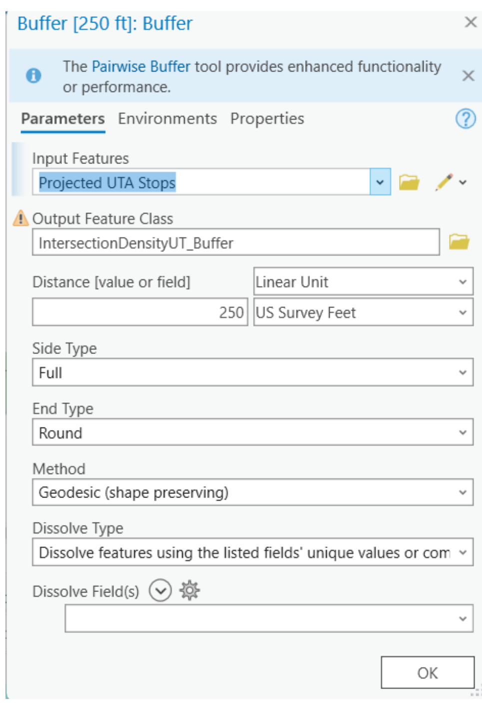


Figure B.3 Buffer settings (Esri, 2023)

Step 3: Intersect Roads and Transit Stops

The next step was to “intersect” the roads shapefile and the buffered transit stops. This gave a road length of 250 ft. in either direction from each transit stop. For the *Input Feature* field, the roads shapefile was selected, and as the *Extent* feature, the buffered transit stops from the previous step were used. Select “line” as the *Output Type*. All other settings were set to the settings shown in Figure B.4. The output to this step was a line type feature class.

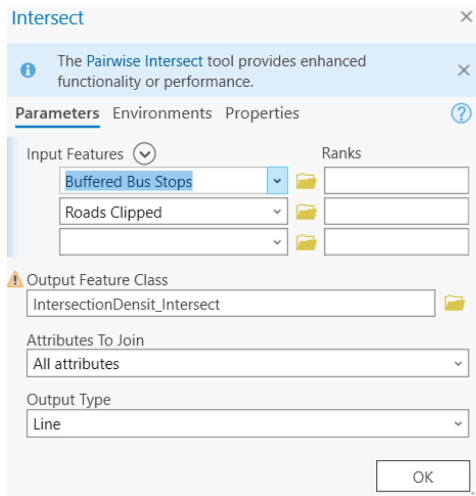


Figure B.4 Intersect roads and transit stops settings (Esri, 2023)

Step 4: Buffer Roads

The next step was to “buffer” the roads some width from the road centerline to capture the data. A width of 100 ft. was used in this report. The *Distance* should be set to the desired distance away from the centerline (e.g., 100 US Survey ft.). The *Type* was set to “geodesic,” and the *Dissolve* was set to “dissolve features using the listed fields’ unique values or combination of values.” The *End Type* was set to “flat.” All other fields were left at default settings, as shown in Figure B.5. The output of this step is a polygon buffer around each road along each transit stop.

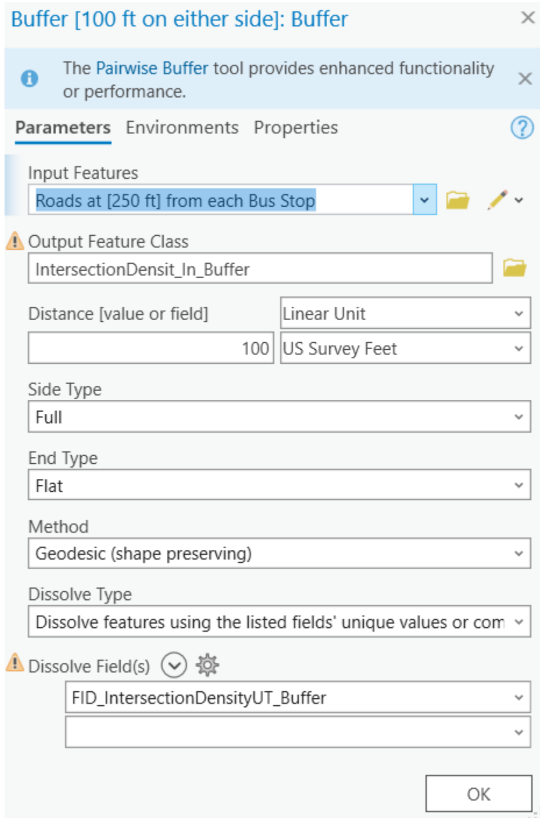


Figure B.5 Roads buffer settings (Esri, 2023)

<u>Step 5: Intersect Crashes and Roads Buffer</u>

The next step was to "intersect" the crashes with the roads buffer from the previous step. The *Input Feature* was set to the crashes shapefile, and the *Extent* feature was set to the roads buffer from the previous step. The "point" option was selected as the *Output Type*. All other settings were set to the settings shown in Figure B.6. Figure B.7 shows a screenshot of a map with the crashes and transit stops within 1,400 ft. from a transit stop.

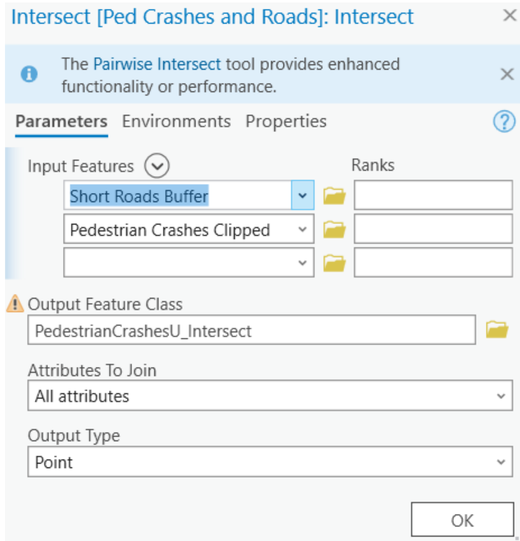


Figure B.6 Intersect crashes and roads settings (Esri, 2023)

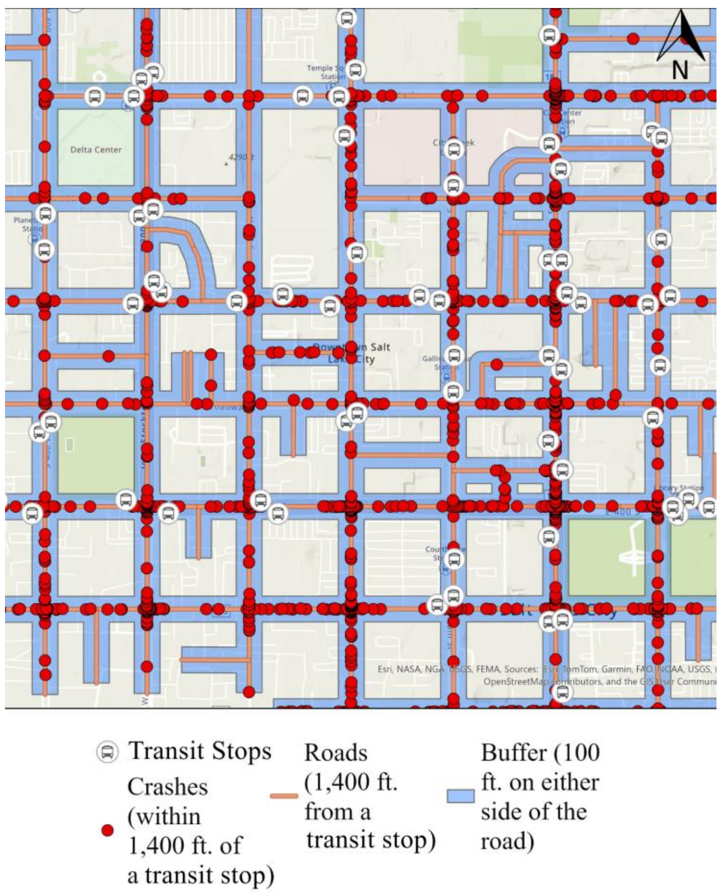


Figure B.7 View of transit stop buffers for 1,400 ft. as shown in ArcGIS Pro (Esri, 2023)

After setting adjustments were completed and the dialog box closed, the resulting crashes feature class was added to the map display by the "add to display" setting. This feature class was the one shown on the map and further analyzed through its attribute table.

<u>Step 6: Distance Calculations</u>

The "near" tool was used to calculate the distances of each crash from the nearest transit stop. As the *Input Feature*, the crashes feature class from the previous step was selected, and as the *Near* feature, the transit stops from Step 1: Project All Datasets were selected. If the shapefiles did not need to be projected, the original shapefiles were used instead. All other settings were set to match Figure B.8. A sample of the final output map is shown in Figure B.9.

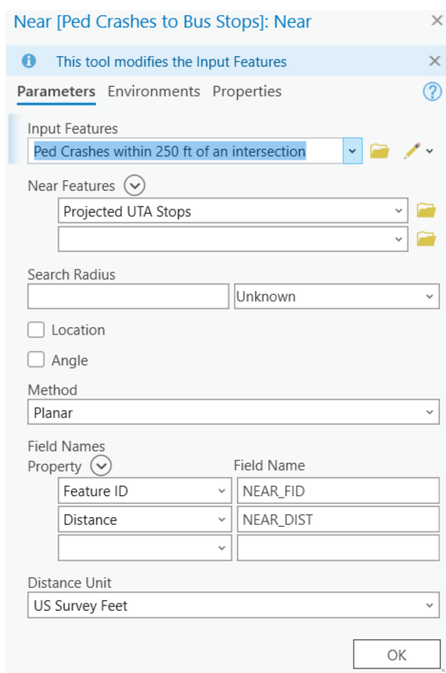


Figure B.8 Near tool settings (Esri, 2023)

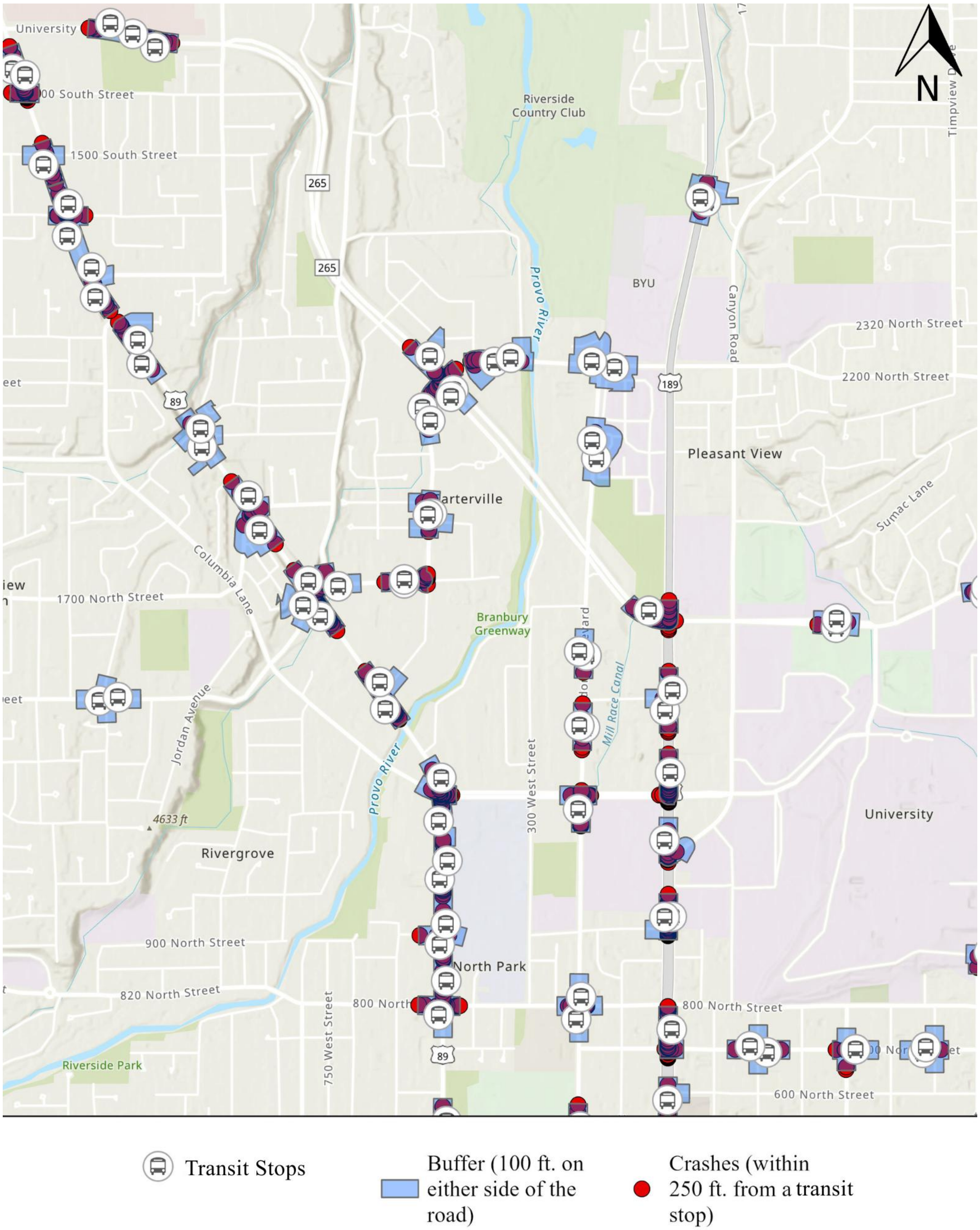


Figure B.9 Sample map after transit stop distance analysis is completed (Esri, 2023)

APPENDIX C: HEATMAP PYTHON CODE

Two sets of python code (one for all crashes and one for VRU crashes) were developed to create a heatmap that shows clusters of crashes as well as the transit stop locations for visual representation of the system. The code takes .csv files for transit stops and crashes based on their location as input. This code projected all the latitude and longitude data into WGS84 projection for cohesion. The python code for all crashes is listed in the first section, followed by the python code for VRU crashes.

C.1 All Crash Analysis Python Code

```
#All Crashes Analysis
import geopandas as gpd
import folium
from folium.plugins import HeatMap
import numpy as np
import pandas as pd # Import pandas

# Load data
# For bus_stops from CSV, first load as pandas DataFrame, then convert to
GeoDataFrame
bus_stops_df = pd.read_csv("BusStopsCC.csv")
# Assuming 'longitude' and 'latitude' columns exist in the CSV for bus stop
coordinates
# And assuming these coordinates are in WGS84 (EPSG:4326)
# If your CSV uses different column names for longitude/latitude (e.g.,
'X', 'Y', 'lon', 'lat'),
# please update 'longitude' and 'latitude' below to match.
geometry_bus_stops = gpd.points_from_xy(bus_stops_df['longitude'],
bus_stops_df['latitude'])
bus_stops = gpd.GeoDataFrame(bus_stops_df, geometry=geometry_bus_stops,
crs="EPSG:4326")

# Accidents data from CSV
accidents_df = pd.read_csv("Crashes250ftBusStopNICC.csv")
# Assuming 'Longitude' and 'Latitude' columns in
Crashes1400ftBusStopNICC.csv represent coordinates
# And assuming these coordinates are in WGS84 (EPSG:4326)
# If your CSV uses different column names for longitude/latitude, please
update them.
geometry_accidents = gpd.points_from_xy(accidents_df['Longitude'],
accidents_df['Latitude'])
accidents = gpd.GeoDataFrame(accidents_df, geometry=geometry_accidents,
crs="EPSG:4326")
```

```
# Reproject bus_stops to match accidents' CRS (UTM Zone 12N)
# Now both are GeoDataFrames and 'to_crs' should work
bus_stops = bus_stops.to_crs(epsg=32612)
accidents = accidents.to_crs(epsg=32612)

# Drop rows with missing geometries
bus_stops = bus_stops.dropna(subset=['geometry'])
accidents = accidents.dropna(subset=['geometry'])

# Filter out points near (0, 0) in UTM (meters)
tolerance = 10  # 10 meters
bus_stops = bus_stops[~((bus_stops.geometry.x.abs() < tolerance) &
(bus_stops.geometry.y.abs() < tolerance))]
accidents = accidents[~((accidents.geometry.x.abs() < tolerance) &
(accidents.geometry.y.abs() < tolerance))]

# Convert accidents to lat/lon (EPSG:4326) for Folium
accidents = accidents.to_crs(epsg=4326)
bus_stops = bus_stops.to_crs(epsg=4326)

# Calculate the center of the map (using the mean of accidents'
coordinates)
center_lat = accidents.geometry.y.mean()
center_lon = accidents.geometry.x.mean()

# Create a Folium map centered on the UTA area
m = folium.Map(
    location=[center_lat, center_lon],
    zoom_start=10,
    tiles='CartoDB Positron'  # Try a reliable basemap provider
)

# Alternative basemap options if CartoDB fails:
# tiles='OpenStreetMap'
# tiles='Esri_WorldStreetMap' (requires an API key in some cases)

# Prepare heatmap data (list of [lat, lon] points)
heat_data = [[point.y, point.x] for point in accidents.geometry]

# Add heatmap
HeatMap(
    heat_data,
    radius=15,
    blur=20,
    max_zoom=18,
    min_opacity=1,
    max_val=5
).add_to(m)
```

```
# Add bus stops as markers
for idx, row in bus_stops.iterrows():
    folium.CircleMarker(
        location=[row.geometry.y, row.geometry.x],
        radius=5,
        color='blue',
        fill=True,
        fill_color='blue',
        fill_opacity=0.7,
        popup=f"Bus Stop: {row.get('stopname', 'Unknown')}"  # Adjust
'stopname' to your column name
    ).add_to(m)

from folium.features import DivIcon

# Add a custom legend for the heatmap
legend_html = '''
<div style="position: fixed; bottom: 50px; right: 50px; z-index: 1000;
padding: 10px; background-color: white; border: 2px solid black;">
    <p><strong>Legend</strong></p>
    <p><span style="color: blue;">&#9679;</span> Bus Stop</p>
    <p><u>Accident Density</u></p>
    <p><span style="color: red;">&#9632;</span> High</p>
    <p><span style="color: orange;">&#9632;</span> Medium-High</p>
    <p><span style="color: lime;">&#9632;</span> Medium-Low</p>
    <p><span style="color: blue;">&#9632;</span> Low</p>
    <hr style="margin: 4px 0;">
</div>
'''

m.get_root().html.add_child(folium.Element(legend_html))
# Save again after adding the legend

# Save the map to an HTML file
m.save("uta_accident_heatmap_j26.html")
print("Map saved as uta_accident_heatmap_j26.html. Open it in a web browser
to view.")
```

C.2 VRU Crash Analysis Python Code

```
#Pedestrian Crashes Analysis
import geopandas as gpd
import folium
from folium.plugins import HeatMap
import numpy as np
import pandas as pd # Import pandas to read CSVs initially

# Load data
```

```
# For bus_stops from CSV, first load as pandas DataFrame, then convert to
GeoDataFrame
bus_stops_df = pd.read_csv("BusStopsCC.csv")
# Assuming 'LONGITUDE' and 'LATITUDE' columns exist in the CSV for bus stop
coordinates
# And assuming these coordinates are in WGS84 (EPSG:4326)
# If your CSV uses different column names for longitude/latitude (e.g.,
'X', 'Y', 'lon', 'lat'),
# please update 'LONGITUDE' and 'LATITUDE' below to match.
geometry_bus_stops = gpd.points_from_xy(bus_stops_df['longitude'],
bus_stops_df['latitude'])
bus_stops = gpd.GeoDataFrame(bus_stops_df, geometry=geometry_bus_stops,
crs="EPSG:4326")

# Accidents data from CSV
accidents_df = pd.read_csv("VRUCrashes250ftBusStopNICCNM.csv")
# Assuming 'LONGITUDE' and 'LATITUDE' columns in
VRUCrashes1400ftBusStopNICCNM.csv represent coordinates
# And assuming these coordinates are in WGS84 (EPSG:4326)
# If your CSV uses different column names for longitude/latitude, please
update them.
geometry_accidents = gpd.points_from_xy(accidents_df['Longitude'],
accidents_df['Latitude'])
accidents = gpd.GeoDataFrame(accidents_df, geometry=geometry_accidents,
crs="EPSG:4326")

# Reproject bus_stops to match accidents' CRS (UTM Zone 12N)
# Now both are GeoDataFrames and 'to_crs' should work
bus_stops = bus_stops.to_crs(epsg=32612)
accidents = accidents.to_crs(epsg=32612)

# Drop rows with missing geometries
bus_stops = bus_stops.dropna(subset=['geometry'])
accidents = accidents.dropna(subset=['geometry'])

# Filter out points near (0, 0) in UTM (meters)
tolerance = 10  # 10 meters
bus_stops = bus_stops[~((bus_stops.geometry.x.abs() < tolerance) &
(bus_stops.geometry.y.abs() < tolerance))]
accidents = accidents[~((accidents.geometry.x.abs() < tolerance) &
(accidents.geometry.y.abs() < tolerance))]

# Convert accidents to lat/lon (EPSG:4326) for Folium
accidents = accidents.to_crs(epsg=4326)
bus_stops = bus_stops.to_crs(epsg=4326)

# Calculate the center of the map (using the mean of accidents'
coordinates)
center_lat = accidents.geometry.y.mean()
center_lon = accidents.geometry.x.mean()
```

```
# Create a Folium map centered on the UTA area
m = folium.Map(
    location=[center_lat, center_lon],
    zoom_start=10,
    tiles='CartoDB Positron'  # Try a reliable basemap provider
)

# Alternative basemap options if CartoDB fails:
# tiles='OpenStreetMap'
# tiles='Esri_WorldStreetMap' (requires an API key in some cases)

# Prepare heatmap data (list of [lat, lon] points)
heat_data = [[point.y, point.x] for point in accidents.geometry]

# Add heatmap
HeatMap(
    heat_data,
    radius=15,
    blur=20,
    max_zoom=18,
    min_opacity=1,
    max_val=5
).add_to(m)

# Add bus stops as markers
for idx, row in bus_stops.iterrows():
    folium.CircleMarker(
        location=[row.geometry.y, row.geometry.x],
        radius=5,
        color='blue',
        fill=True,
        fill_color='blue',
        fill_opacity=0.7,
        popup=f"Bus Stop: {row.get('stopname', 'Unknown')}"  # Adjust
'stopname' to your column name
    ).add_to(m)

from folium.features import DivIcon

# Add a custom legend for the heatmap
legend_html = '''
<div style="position: fixed; bottom: 50px; right: 50px; z-index: 1000;
padding: 10px; background-color: white; border: 2px solid black;">
    <p><strong>Legend</strong></p>
    <p><span style="color: blue;">&#9679;</span> Bus Stop</p>
    <p><u>Accident Density</u></p>
    <p><span style="color: red;">&#9632;</span> High</p>
    <p><span style="color: orange;">&#9632;</span> Medium-High</p>
    <p><span style="color: lime;">&#9632;</span> Medium-Low</p>
```

```
    <p><span style="color: blue;">&#9632;</span> Low</p>
    <hr style="margin: 4px 0;">
</div>
'''

m.get_root().html.add_child(folium.Element(legend_html))
# Save again after adding the legend
# Save the map to an HTML file
m.save("accident_heatmap_pedcrashes.html") #Update these names
print("Map saved as accident_heatmap_pedcrashes.html. Open it in a web
browser to view.") #Update these names
```

APPENDIX D: MOST FREQUENT TRANSIT STOPS AND INTERSECTIONS PYTHON CODE

This python code was developed to find the top 45 transit stops and intersections with the highest crash frequencies. The code takes .csv files with latitude and longitude values of all crashes, VRU crashes, transit stops, and intersections as input. This code projected all the latitude and longitude data into WGS84 projection for cohesion.

D.1 Top 45 Most Frequent Transit Stops Python Code

This python code was developed to find the top 45 transit stops with the highest crash frequencies and create maps and histograms that depict the relationships between the crashes and the transit stops. The histogram output also shows the nearest intersection location

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the crash data, change the spatial threshold or dataset
here
try:
    df = pd.read_csv('Crashes1400ftBusStopNICC.csv')
except FileNotFoundError:
    print("Error: Crashes1400ftBusStop.csv not found. Please make sure the
file is in the correct directory.")
    exit()

# Load the bus stop intersection data
try:
    df_stops = pd.read_csv('BusStops1400ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: BusStops1400ftIntersection.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Sort the crash data by NEAR_FID (optional based on the task, but good
practice for grouping)
df_sorted = df.sort_values(by='NEAR_FID')

# Count the number of entries for each unique NEAR_FID
near_fid_counts = df_sorted['NEAR_FID'].value_counts()

# Take the top 45 most frequent NEAR_FID values
```

```
top_45_near_fids = near_fid_counts.head(45).index.tolist()

print("Top 45 most frequent NEAR_FID values:")
print(top_45_near_fids)

# Create a histogram for each of the top 10 NEAR_FID values
for fid in top_45_near_fids:
    # Filter the crash DataFrame for the current NEAR_FID
    df_fid = df_sorted[df_sorted['NEAR_FID'] == fid]

    # Calculate the number of entries and the mean NEAR_DIST from crash
data
    num_entries = len(df_fid)
    mean_near_dist = df_fid['NEAR_DIST'].mean()

    # Find the bus stop distance to the nearest intersection and STOP_LandT
    stop_info_row = df_stops[df_stops['OBJECTID'] == fid]
    if not stop_info_row.empty:
        bus_stop_distance = stop_info_row['NEAR_DIST'].iloc[0]
        stop_landt = stop_info_row['STOP_LandT'].iloc[0]
    else:
        bus_stop_distance = "N/A" # Handle cases where the FID is not found
in the stops data
        stop_landt = "N/A"

    print(f"Bus Stop ID: {fid}")
    print(f"Number of entries: {num_entries}")
    print(f"Average Crash Distance: {mean_near_dist:.2f}")
    if isinstance(bus_stop_distance, (int, float)):
        print(f"Bus Stop distance to nearest intersection:
{bus_stop_distance:.2f}")
    else:
        print(f"Bus Stop distance to nearest intersection:
{bus_stop_distance}")
    print(f"Land Use Type: {stop_landt}")

    # Create the histogram
    plt.figure(figsize=(10, 6))
    sns.histplot(data=df_fid, x='NEAR_DIST', kde=True)
    plt.xlabel('Crashes Distance to Nearest Bus Stop')
    plt.xlim(-50, 1400)   # sets x-axis from 0 to 1400 ft
    plt.ylabel('Frequency')

    # Add a vertical line at x=0 for the bus stop location
    plt.axvline(x=0, color='green', linestyle='-', label='Bus Stop')

    # Add a vertical line for the bus stop distance to the nearest
intersection
    if isinstance(bus_stop_distance, (int, float)):
```

```
        plt.axvline(bus_stop_distance, color='red', linestyle='dashed',
linewidth=2, label=f'Intersection Location: {bus_stop_distance:.2f} ft from
bus stop')
        # Add a 400ft buffer around the intersection location
        # Add If statement if bus stop distance is less than 400 (so xmin-
400 will be lesser than 0)
        if bus_stop_distance < 400:
          plt.axvspan(xmin=0, xmax=bus_stop_distance + 400, color='red',
alpha=0.2, label='Intersection Influence Area')
        else:
          plt.axvspan(xmin=bus_stop_distance - 400, xmax=bus_stop_distance
+ 400, color='red', alpha=0.2, label='Intersection Influence Area')
        plt.legend()

    # plt.title(f'Histogram of distance to nearest bus stop for Bus Stop
ID: {fid}')
    plt.xlabel('Crash Distance to Nearest Bus Stop (ft)')
    plt.ylabel('Crash Frequency')
    plt.show()
# Install necessary libraries for mapping
%pip install geopandas folium
import geopandas as gpd
import folium

# Merge df_stops with the top_45_near_fids to get the locations of the top
bus stops
top_stops_df = df_stops[df_stops['OBJECTID'].isin(top_45_near_fids)].copy()

# Convert to GeoDataFrame (assuming 'LATITUDE' and 'LONGITUDE' columns
exist in df_stops)
# Replace 'LATITUDE' and 'LONGITUDE' with the actual column names in your
df_stops if they are different
gdf_stops = gpd.GeoDataFrame(
    top_stops_df, geometry=gpd.points_from_xy(top_stops_df['longitude'],
top_stops_df['latitude']))

# Create a base map centered around the mean location of the bus stops
map_center = [gdf_stops['latitude'].mean(), gdf_stops['longitude'].mean()]
m = folium.Map(location=map_center, zoom_start=10)

# Add markers for each of the top bus stops
for idx, row in gdf_stops.iterrows():
    folium.Marker(
        location=[row['latitude'], row['longitude']],
        popup=f"Bus Stop ID: {row['OBJECTID']}<br>Land Use Type:
{row['STOP_LandT']}<br>Crashes: {near_fid_counts.get(row['OBJECTID'],
'N/A')}",
        icon=folium.Icon(color='blue', icon='info-sign')
    ).add_to(m)
```

```
# Display the map
m
```

D.2 Top 45 Most Frequent Intersections Python Code

This python code was developed to find the top 45 intersection with the highest crash frequencies and create maps and histograms that depict the relationships between the crashes and the transit stops. The histogram output also shows the nearest transit stop location.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the crash data, change the spatial threshold or dataset here
try:
    df = pd.read_csv('Crashes250ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: Crashes250ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Load the bus stop intersection data
try:
    df_stops = pd.read_csv('BusStops250ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: BusStops250ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Sort the crash data by NEAR_FID (optional based on the task, but good
practice for grouping)
df_sorted = df.sort_values(by='NEAR_FID')

# Count the number of entries for each unique NEAR_FID
near_fid_counts = df_sorted['NEAR_FID'].value_counts()

# Take the top 45 most frequent NEAR_FID values
top_45_near_fids = near_fid_counts.head(45).index.tolist()

print("Top 45 most frequent NEAR_FID values:")
print(top_45_near_fids)

# Create a histogram for each of the top 10 NEAR_FID values
for fid in top_45_near_fids:
    # Filter the crash DataFrame for the current NEAR_FID
    df_fid = df_sorted[df_sorted['NEAR_FID'] == fid]
```

```
    # Calculate the number of entries and the mean NEAR_DIST from crash
data
    num_entries = len(df_fid)
    mean_near_dist = df_fid['NEAR_DIST'].mean()

    # Find the bus stop distance to the nearest intersection and STOP_LandT
    stop_info_row = df_stops[df_stops['NEAR_FID'] == fid]
    if not stop_info_row.empty:
        bus_stop_distance = stop_info_row['NEAR_DIST'].iloc[0]
        stop_landt = stop_info_row['STOP_LandT'].iloc[0]
    else:
        bus_stop_distance = "N/A" # Handle cases where the FID is not found
in the stops data
        stop_landt = "N/A"

    print(f"\nIntersection ID: {fid}")
    print(f"Number of entries: {num_entries}")
    print(f"Average Crash Distance: {mean_near_dist:.2f}")
    if isinstance(bus_stop_distance, (int, float)):
        print(f"Bus Stop distance to nearest intersection:
{bus_stop_distance:.2f}")
    else:
        print(f"Bus Stop distance to nearest intersection:
{bus_stop_distance}")
    print(f"Land Use Type: {stop_landt}")

    # Create the histogram
    plt.figure(figsize=(10, 6))
    sns.histplot(data=df_fid[df_fid['NEAR_DIST'] <= 250], x='NEAR_DIST',
kde=True) #limit the crashes to 250ft
    plt.xlim(-10, 275)   # sets x-axis from 0 to 275 ft

    # Add a vertical line for the bus stop distance to the nearest
intersection
    if isinstance(bus_stop_distance, (int, float)):
        plt.axvline(bus_stop_distance, color='green', linestyle='-',
linewidth=2, label=f'Bus Stop Location: {bus_stop_distance:.2f} ft from
intersection')
        plt.legend()

    # Add a vertical line at x=0 for the intersection
    plt.axvline(x=0, color='red', linestyle='dashed', label='Intersection')
    plt.legend()

    # plt.title(f'Histogram of Distance to nearest intersection for
Intersection ID: {fid}')
    plt.xlabel('Crashes Distance to Intersection (ft)')
    plt.ylabel('Frequency')
    plt.show()
# Install necessary libraries for mapping
```

```
%pip install geopandas folium
import geopandas as gpd
import folium
import pandas as pd

# Load the crash data
try:
    df = pd.read_csv('Crashes250ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: Crashes250ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Load the bus stop intersection data
try:
    df_stops = pd.read_csv('BusStops1400ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: BusStops1400ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Calculate the mean Latitude and Longitude for each NEAR_FID
# This gives a more accurate center point for each intersection
intersection_locations =
df[df['NEAR_FID'].isin(top_45_near_fids)].groupby('NEAR_FID')[['Latitude',
'Longitude']].mean().reset_index()

# Merge with the crash counts for popup info
top_intersections_df =
intersection_locations.merge(near_fid_counts.rename('Crash_Count'),
left_on='NEAR_FID', right_index=True)

# Convert to GeoDataFrame (assuming 'LATITUDE' and 'LONGITUDE' columns
exist in df)
# Replace 'LATITUDE' and 'LONGITUDE' with the actual column names in your
df if they are different
gdf_intersections = gpd.GeoDataFrame(
    top_intersections_df,
geometry=gpd.points_from_xy(top_intersections_df['Longitude'],
top_intersections_df['Latitude']))

# Create a base map centered around the mean location of the intersections
map_center = [gdf_intersections['Latitude'].mean(),
gdf_intersections['Longitude'].mean()]
m = folium.Map(location=map_center, zoom_start=10)

# Add markers for each of the top intersections
for idx, row in gdf_intersections.iterrows():
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
```

```
        popup=f"Intersection ID: {row['NEAR_FID']}<br>Crashes:
{row['Crash_Count']}",
        icon=folium.Icon(color='red', icon='info-sign') # Changed color to
red for intersections
    ).add_to(m)

# Display the map
m
```

D.3 Combined Crashes, Transit Stops, and Intersections Map Python Code

This python code was developed to create a map that visually combines the top 45 transit stops and intersections with the highest crash frequencies. The code takes .csv files with latitude and longitude values of all crashes, VRU crashes, transit stops, and intersections as input. This code projected all the latitude and longitude data into WGS84 projection for cohesion.

```
#Map Crashes, VRU Crashes, Intersections
import geopandas as gpd
import folium
import pandas as pd
import io
from branca.element import Element
from PIL import Image

# Load the intersection crash data (used to find top intersections)
try:
    df_intersection_crashes =
pd.read_csv('Crashes1400ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: Crashes1400ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Load the all crashes bus stop data (used to find top bus stops for all
crashes)
try:
    df_all_crashes = pd.read_csv('Crashes1400ftBusStopNICC.csv')
except FileNotFoundError:
    print("Error: Crashes1400ftBusStopNICC.csv not found. Please make sure
the file is in the correct directory.")
    exit()

# Load the VRU crashes bus stop data (used to find top bus stops for VRU
crashes)
try:
    df_vru_crashes = pd.read_csv('VRUCrashes1400ftBusStopNICCNM.csv')
```

```
except FileNotFoundError:
    print("Error: VRUCrashes1400ftBusStopNICCNM.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Load the bus stop intersection data (used to get location and info for
bus stops)
try:
    df_stops = pd.read_csv('BusStops1400ftIntersectionNI.csv')
except FileNotFoundError:
    print("Error: BusStops1400ftIntersectionNI.csv not found. Please make
sure the file is in the correct directory.")
    exit()

# Load the VRU crashes bus stop data (used to find top bus stops for VRU
crashes)
try:
    df_vru_intersection_crashes =
pd.read_csv('VRUCrashes1400ftIntersectionNINM.csv')
except FileNotFoundError:
    print("Error: VRUCrashes1400ftIntersectionNINM.csv not found. Please
make sure the file is in the correct directory.")
    exit()

# Find top 45 intersections based on crash counts
near_fid_counts_intersections =
df_intersection_crashes['NEAR_FID'].value_counts()
top_45_near_fids_intersections =
near_fid_counts_intersections.head(45).index.tolist()
# Use the original DataFrame to get location information for intersections
top_intersections_df =
df_intersection_crashes[df_intersection_crashes['NEAR_FID'].isin(top_45_nea
r_fids_intersections)].drop_duplicates(subset=['NEAR_FID']).copy()

# Find top 45 intersections based on VRU crash counts
near_fid_counts_vru_intersections =
df_vru_intersection_crashes['NEAR_FID'].value_counts()
top_45_near_fids_vru_intersections =
near_fid_counts_vru_intersections.head(45).index.tolist()
# Use the original DataFrame to get location information for intersections
top_intersections_vru_df =
df_vru_intersection_crashes[df_vru_intersection_crashes['NEAR_FID'].isin(to
p_45_near_fids_vru_intersections)].drop_duplicates(subset=['NEAR_FID']).cop
y()

# Find top 45 bus stops based on all crash counts
near_fid_counts_all_crashes = df_all_crashes['NEAR_FID'].value_counts()
top_45_near_fids_all_crashes =
near_fid_counts_all_crashes.head(45).index.tolist()
# Get the location and info for these bus stops from df_stops
```

```
top_all_crashes_stops_df =
df_stops[df_stops['OBJECTID'].isin(top_45_near_fids_all_crashes)].copy()

# Find top 45 bus stops based on VRU crash counts
near_fid_counts_vru_crashes = df_vru_crashes['NEAR_FID'].value_counts()
top_45_near_fids_vru_crashes =
near_fid_counts_vru_crashes.head(45).index.tolist()
# Get the location and info for these bus stops from df_stops
top_vru_crashes_stops_df =
df_stops[df_stops['OBJECTID'].isin(top_45_near_fids_vru_crashes)].copy()

#Print Bus Stop and Intersection IDs
print("Intersections with the most crashes")
print(top_45_near_fids_intersections)
print("Intersections with the most VRU crashes")
print(top_45_near_fids_vru_intersections)
print("Bus Stops with the most crashes")
print(top_45_near_fids_all_crashes)
print("Bus Stops with the most VRU crashes")
print(top_45_near_fids_vru_crashes)

# Convert to GeoDataFrame (assuming 'Latitude' and 'Longitude' columns
exist in top_intersections_df, and 'latitude' and 'longitude' in bus stop
dataframes)
# Explicitly extract Latitude and Longitude Series for GeoDataFrame
creation
latitude_series = top_intersections_df['Latitude']
longitude_series = top_intersections_df['Longitude']

#Create variables for GeoDataFrame for VRU Intersections
latitude_vru_series = top_intersections_vru_df['Latitude']
longitude_vru_series = top_intersections_vru_df['Longitude']

gdf_intersections = gpd.GeoDataFrame(
    top_intersections_df,
geometry=gpd.points_from_xy(top_intersections_df['Longitude'],
top_intersections_df['Latitude']))

gdf_vru_intersections = gpd.GeoDataFrame(
    top_intersections_vru_df,
geometry=gpd.points_from_xy(top_intersections_vru_df['Longitude'],
top_intersections_vru_df['Latitude']))

gdf_all_crashes_stops = gpd.GeoDataFrame(
    top_all_crashes_stops_df,
geometry=gpd.points_from_xy(top_all_crashes_stops_df['longitude'],
top_all_crashes_stops_df['latitude']))

gdf_vru_crashes_stops = gpd.GeoDataFrame(
```

```
    top_vru_crashes_stops_df,
geometry=gpd.points_from_xy(top_vru_crashes_stops_df['longitude'],
top_vru_crashes_stops_df['latitude']))

# Create a base map centered around the mean location of all points
all_points = pd.concat([gdf_intersections[['Latitude', 'Longitude']],
gdf_vru_intersections[['Latitude', 'Longitude']],
gdf_all_crashes_stops[['latitude',
'longitude']].rename(columns={'latitude': 'Latitude', 'longitude':
'Longitude'}), gdf_vru_crashes_stops[['latitude',
'longitude']].rename(columns={'latitude': 'Latitude', 'longitude':
'Longitude'})])
map_center = [all_points['Latitude'].mean(),
all_points['Longitude'].mean()]
m = folium.Map(location=map_center, zoom_start=10)
# 1. Add Scale Bar
# ScaleBar(position='bottomleft').add_to(m) # Removed due to ImportError

# 2. Add North Arrow (Injected HTML)
north_arrow_html = '''
<div style="
    position: fixed;
    top: 30px; right: 30px; width: 40px; height: 40px;
    z-index: 9999; pointer-events: none;
    ">
    <img src="https://freesvg.org/img/north-arrow-2.png"
         style="width:100%; height:100%;">
</div>
'''
m.get_root().html.add_child(Element(north_arrow_html))
# Create the HTML for the legend
legend_html = '''
{% macro html(this, kwargs) %}
<div style="
    position: fixed;
    bottom: 50px; right: 50px; width: 280px; height: 140px;
    background-color: white; border:2px solid grey; z-index:9999; font-
size:14px;
    padding: 10px;
    border-radius: 5px;
    box-shadow: 2px 2px 5px rgba(0,0,0,0.3);
    ">
    <b>Map Legend</b><br>
    <i class="fa fa-map-marker" style="color:red"></i>  Intersections
(Most Crashes)<br>
    <i class="fa fa-map-marker" style="color:orange"></i> 
Intersections (Most VRU Crashes)<br>
    <i class="fa fa-bus" style="color:blue"></i>  Bus Stops (Most
Crashes)<br>
```

```
    <i class="fa fa-user" style="color:green"></i>  Bus Stops (Most
VRU Crashes)
</div>
{% endmacro %}
'''

# Create a MacroElement and add it to the map
from branca.element import Template, MacroElement

legend = MacroElement()
legend._template = Template(legend_html)

m.get_root().add_child(legend)

# Create FeatureGroups for each category to add to LayerControl
intersections_fg = folium.FeatureGroup(name='Intersections with most
crashes (Red)').add_to(m)
vru_intersections_fg = folium.FeatureGroup(name='Intersections with most
VRU crashes (Orange)').add_to(m)
all_crashes_stops_fg = folium.FeatureGroup(name='Bus Stops with most
crashes (Blue)').add_to(m)
vru_crashes_stops_fg = folium.FeatureGroup(name='Bus Stops with most VRU
crashes (Green)').add_to(m)

# Add markers for each of the top intersections (Red)
for idx, row in gdf_intersections.iterrows():
    popup_content = f"Intersection ID: {row['NEAR_FID']}<br>Crashes:
{near_fid_counts_intersections.get(row['NEAR_FID'], 'N/A')}"
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=folium.Popup(f"<div
style='width:150px;'>{popup_content}</div>"),
        icon=folium.Icon(color='red', icon='info-sign')
    ).add_to(intersections_fg)

# Add markers for each of the top VRU intersections (Orange)
for idx, row in gdf_vru_intersections.iterrows():
    popup_content = f"Intersection (VRU Crashes) ID:
{int(row['NEAR_FID'])}<br>Crashes:
{near_fid_counts_intersections.get(row['NEAR_FID'], 'N/A')}"
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=folium.Popup(f"<div
style='width:200px;'>{popup_content}</div>"),
        icon=folium.Icon(color='orange', icon='info-sign')
    ).add_to(vru_intersections_fg)

# Add markers for each of the top bus stops for all crashes (Blue)
for idx, row in gdf_all_crashes_stops.iterrows():
     # Get the crash count for this bus stop
```

```
    crash_count = near_fid_counts_all_crashes.get(row['OBJECTID'], 'N/A')
    popup_content = f"Bus Stop ID: {row['OBJECTID']}<br>All Crashes: 
{crash_count}<br>Land Use: {row['STOP_LandT']}"
    folium.Marker(
        location=[row['latitude'], row['longitude']],
        popup=folium.Popup(f"<div 
style='width:150px;'>{popup_content}</div>"),
        icon=folium.Icon(color='blue', icon='bus')
    ).add_to(all_crashes_stops_fg)

# Add markers for each of the top bus stops for VRU crashes (Green)
for idx, row in gdf_vru_crashes_stops.iterrows():
    # Get the VRU crash count for this bus stop
    vru_crash_count = near_fid_counts_vru_crashes.get(row['OBJECTID'], 
'N/A')
    popup_content = f"Bus Stop ID: {row['OBJECTID']}<br>VRU Crashes: 
{vru_crash_count}<br>Land Use: {row['STOP_LandT']}"
    folium.Marker(
        location=[row['latitude'], row['longitude']],
        popup=folium.Popup(f"<div 
style='width:150px;'>{popup_content}</div>"),
        icon=folium.Icon(color='green', icon='user') # Using a user icon 
for VRU (Vulnerable Road User)
    ).add_to(vru_crashes_stops_fg)

# Add a layer control to the map to toggle layers
#folium.LayerControl(collapsed=False).add_to(m)

print("Red=Intersections with the most crashes")
print("Orange=Intersections with the most VRU crashes")
print("Blue=Bus Stops with the most crashes")
print("Green=Bus Stops with the most VRU crashes")

# Display the map
m.save("Crash_Analysis_Map.html")

# Export to PNG (Requires Selenium and a Webdriver like ChromeDriver)
#try:
    #img_data = m._to_png(5) # 5 second delay to let tiles load
    #with open('Crash_Analysis_Map.png', 'wb') as f:
        #f.write(img_data)
    #print("Map successfully exported as PNG.")
#except Exception as e:
    #print(f"PNG export failed: {e}")
    #print("Note: PNG export requires Selenium and a Webdriver 
(Chrome/Firefox) installed.")

# Display the Map
m
```

APPENDIX E: STATISTICAL MODEL R CODE

This appendix contains the R code that runs the logistic regression analysis for the 1,400 ft. and 250 ft. analysis. The R code is used to examine what factors had a significant impact on VRU-related crashes. R Studio (R Foundation for Statistical Computing 2025) was used to run all models. The code takes a dependent variable (VRU crashes), independent variables and crash datasets, then calculates the relationship between different variables and the likelihood of a crash involving a VRU, thus showing which factors strongly contribute to an increased VRU crash risk.

E.1 1,400 ft. Analysis R Code

```
library(tidyverse)
library(emmeans)
library(splines)
library(readxl)
library(lme4)
library(dplyr)

# Read Data (1,400 ft.)
crash <- read.csv("Crashes1400ftBusStopNICC.csv", na = c("#N/A", "NA"))

bus_ref <- read.csv("Bus Stops Data - Reference.csv")  # adjust path/sheet
if needed

bus_int <- read.csv("BusStops1400ftIntersectionNI.csv")   # adjust path

# 2. Attach stopabbr to each crash via NEAR_FID ----
crash2 <- crash %>%
  left_join(
    bus_ref %>%
      filter(!is.na(FID)) %>%                  # Removes the 51 NAs
      distinct(FID, .keep_all = TRUE) %>%     # Ensures every FID is unique
      mutate(stopabbr = as.character(stopabbr)),
    by = c("NEAR_FID" = "FID")
  )

bus_int_single <- bus_int %>%
  filter(uta_stopid != 0 & !is.na(uta_stopid)) %>% # Removes ID 0 and any
NAs
  group_by(uta_stopid) %>%
  slice_min(order_by = NEAR_DIST, n = 1, with_ties = FALSE) %>%
  ungroup()
```

```
crash3 <- crash2 %>%
  left_join(
    bus_int_single %>%
      mutate(uta_stopid = as.character(uta_stopid)),
    by = c("stopabbr" = "uta_stopid"),
    suffix = c("", "_stop_int")
  ) %>%
  rename(
    CrashStopDist = NEAR_DIST,          # if this is the crash→stop
distance
    StopIntDist   = NEAR_DIST_stop_int  # stop→intersection distance
  )

crash3 <- crash3 %>%
  mutate(
    StopAtIntersection = if_else(StopIntDist <= 50, 1L, 0L),
    # Create Pedestrian by checking if either column is Y
    Pedestrian = if_else(Pedestri_1 == "Y" | Bicycle__1 == "Y", 1L, 0L)
  ) %>%
  rename(
    Bus_Stop_Location = Bus.Stop.Location,
    Bus_Stop_Land_Use = Bus.Stop.Land.Use
  )

crash3 <- crash3 %>%
  mutate(Bus_Stop_Land_Use = fct_collapse(as.character(Bus_Stop_Land_Use),
                                          "Residential/Mixed" = c("Single
Home", "Apartment", "Mixed Use"),
                                          "Commercial"        =
c("Commercial", "Business"),
                                          "Agricultural"      =
c("Undeveloped", "Farming", "Industrial"),
                                          "Institutional"     =
c("Government", "School", "Church", "Hospital", "Library", "Transit
station"),
                                          "Recreational"      =
c("Recreational")))

#Re-level Land Use to Commercial
crash3$Bus_Stop_Land_Use <- relevel(factor(crash3$Bus_Stop_Land_Use), ref =
"Commercial")

# Re-level Location to Nearside
crash3$Bus_Stop_Location <- relevel(factor(crash3$Bus_Stop_Location), ref =
"NS")

#Convert categories to numeric for glm fit
crash3$DOT_AADT <- as.numeric(as.character(crash3$DOT_AADT))
```

```
crash3$AvgBoard <- as.numeric(as.character(crash3$AvgBoard))

crash3$SPEED_LMT <- as.numeric(as.character(crash3$SPEED_LMT))

crash3$CrashStopDist <- round(crash3$CrashStopDist, 1)

crash3$StopIntDist <- round(crash3$StopIntDist, 1)

#Resolves empty strings
# This replaces "" and " " with NA for every column in the dataframe
crash3[crash3 == "" | crash3 == " "| crash3 == "Empty"] <- NA

# Then, drop the empty factor levels from all variables at once
crash3 <- droplevels(crash3)

# Modeling
crash3$Pedestrian <- factor(crash3$Pedestrian)
model4 <- glm(Pedestrian ~ ns(CrashStopDist, 2) +
                           ns(SPEED_LMT, 2) +
                           DOT_AADT +
                           AvgBoard +
                           AvgAlighting +
                           Intersecti +
                           Bus_Stop_Location +
                           Bus_Stop_Land_Use +
                           StopIntDist +
                           ns(CrashStopDist, 2)*Bus_Stop_Location,
              data = crash3, family = binomial)
summary(model4)

drop1(model4, test='Chisq')

tmp.w <- as.numeric(names(model4$y))
model4.1 <- update(model4, . ~ ns(CrashStopDist, 2) +
                           ns(SPEED_LMT, 2) +
                           DOT_AADT +
                           AvgBoard +
                           Intersecti +
                           Bus_Stop_Location +
                           Bus_Stop_Land_Use +
                           StopIntDist +
                           ns(CrashStopDist, 2)*Bus_Stop_Location, subset=t
mp.w)
summary(model4.1)

anova(model4.1, model4)

drop1(model4.1, test='Chisq')
```

```
#summary(crash3)
newdat<-data.frame(SPEED_LMT = 35, DOT_AADT = 20000, AvgBoard = 5,
                   AvgAlighting = 5, Intersecti = "Y",
                   Bus_Stop_Location = "FS", Bus_Stop_Land_Use =
"Commercial",
                   StopIntDist = 158.663, CrashStopDist = seq(0.58,
1487.834,length.out = 100))
#Use the median values of each predictor category to hold them all constant

newdat$pred1 <- predict(model4,newdat,type = "response")
plot(pred1~CrashStopDist,data = newdat, type = "l", xlab = "Distance of a
Crash from Bus Stop", ylab = "Probability (%) a Crash is VRU Related",main
= "VRU Crash Probability 1,400 ft.")
#Use the median values of each predictor category to hold them all constant

newdata<-data.frame(CrashStopDist = 286.8, DOT_AADT = 20000, AvgBoard = 5,
                   AvgAlighting = 5, Intersecti = "Y",
                   Bus_Stop_Location = "FS", Bus_Stop_Land_Use = "Commercia
l",
                   StopIntDist = 158.663, SPEED_LMT = seq(0, 65, length.out
= 100))
#Use the median values of each predictor category to hold them all constant


newdata$pred1 <- predict(model4,newdata,type = "response")
plot(pred1~SPEED_LMT,data = newdata, type = "l", xlab = "Speed Limit (mph)"
, ylab = "Probability (%) a Crash is VRU Related",main = "VRU Crash Probabi
lity 1,400 ft.")
#Use the median values of each predictor category to hold them all constant

library(emmeans)
em4.1 <- emmeans(model4.1, ~Bus_Stop_Land_Use)
plot(em4.1, type="response", xlab = "Probability (%) a Crash is VRU Related
(1,400 ft.)", ylab = "Bus Stop Land Use", main = "VRU Crash Probability 1,4
00 ft.")

plot(pairs(em4.1, type="response"))

em4.2 <- emmeans(model4.1, ~Bus_Stop_Location)

plot(em4.2, type="response", xlab = "Probability (%) a Crash is VRU Related
(1,400 ft.)", ylab = "Bus Stop Location", main = "VRU Crash Probability 1,4
00 ft.")

# Get the estimated marginal means (predicted probabilities) for each locat
ion
location_means <- emmeans(model4.1, ~ Bus_Stop_Location, type = "response")
# See the predicted probability for each specific location
print(location_means)
```

```
# Perform pairwise comparisons to see which specific locations differ from
one another, fdr adjustment
pairs(location_means, adjust = "none")

exp(coef(model4.1))

exp(confint(model4.1))
```

E.2 250 ft. Analysis R Code

```
library(tidyverse)
library(emmeans)
library(splines)
library(readxl)
library(lme4)
library(dplyr)

#Read Data (250 ft. Bus Stops)
crash <- read.csv("Crashes250ftBusStopNICC.csv", na = c("#N/A", "NA"))

bus_ref <- read.csv("Bus Stops Data - Reference.csv")  # adjust path/sheet
if needed

bus_int <- read.csv("BusStops1400ftIntersectionNI.csv")   # adjust path

# 2. Attach stopabbr to each crash via NEAR_FID ----
crash2 <- crash %>%
  left_join(
    bus_ref %>%
      mutate(stopabbr = as.character(stopabbr)),
    by = c("NEAR_FID" = "FID")
  )

bus_int_single <- bus_int %>%
  group_by(uta_stopid) %>%
  slice_min(order_by = NEAR_DIST, n = 1, with_ties = FALSE) %>%
  ungroup()

crash3 <- crash2 %>%
  left_join(
    bus_int_single %>%
      mutate(uta_stopid = as.character(uta_stopid)),
    by = c("stopabbr" = "uta_stopid"),
    suffix = c("", "_stop_int")
  ) %>%
  rename(
    CrashStopDist = NEAR_DIST,          # if this is the crash→stop distanc
e
```

```
    StopIntDist    = NEAR_DIST_stop_int  # stop→intersection distance
  )

crash3 <- crash3 %>%
  mutate(
    StopAtIntersection = if_else(StopIntDist <= 50, 1L, 0L),
    # Create Pedestrian by checking if either column is Y
    Pedestrian = if_else(Pedestri_1 == "Y" | Bicycle__1 == "Y", 1L, 0L)
  ) %>%
  rename(
    Bus_Stop_Location = Bus.Stop.Location,
    Bus_Stop_Land_Use = Bus.Stop.Land.Use
  )

crash3 <- crash3 %>%
  mutate(Bus_Stop_Land_Use = fct_collapse(as.character(Bus_Stop_Land_Use),
                                          "Residential/Mixed" = c("Single H
ome", "Apartment", "Mixed Use"),
                                          "Commercial"        = c("Commerci
al", "Business"),
                                          "Agricultural"      = c("Undevelo
ped", "Farming", "Industrial"),
                                          "Institutional"     = c("Governme
nt", "School", "Church", "Hospital", "Library", "Transit station"),
                                          "Recreational"      = c("Recreati
onal"),
                                          "Empty"             = c(" ")
  ))
#Re-level Land Use to Commercial
crash3$Bus_Stop_Land_Use <- relevel(factor(crash3$Bus_Stop_Land_Use), ref =
"Commercial")

# Re-level Location to Nearside
crash3$Bus_Stop_Location <- relevel(factor(crash3$Bus_Stop_Location), ref =
"NS")

crash3$DOT_AADT <- as.numeric(as.character(crash3$DOT_AADT))

crash3$AvgBoard <- as.numeric(as.character(crash3$AvgBoard))

crash3$SPEED_LMT <- as.numeric(as.character(crash3$SPEED_LMT))

crash3$CrashStopDist <- round(crash3$CrashStopDist, 1)

crash3$StopIntDist <- round(crash3$StopIntDist, 1)

#Resolves empty strings
# This replaces "" and " " with NA for every column in the dataframe
crash3[crash3 == "" | crash3 == " "| crash3 == "Empty"] <- NA
```

```
# Then, drop the empty factor levels from all variables at once
crash3 <- droplevels(crash3)

#Modeling
crash3$Pedestrian <- factor(crash3$Pedestrian)
model4 <- glm(Pedestrian ~ ns(CrashStopDist, 2) +
                ns(SPEED_LMT, 2) +
                DOT_AADT +
                AvgBoard +
                AvgAlighting +
                Intersecti +
                Bus_Stop_Location +
                Bus_Stop_Land_Use +
                StopIntDist +
                ns(CrashStopDist, 2)*Bus_Stop_Location,
              data = crash3, family = binomial)
summary(model4)

drop1(model4, test='Chisq')

#summary(crash3)
newdat<-data.frame(SPEED_LMT = 40, DOT_AADT = 25000, AvgBoard = 5,
                   AvgAlighting = 6, Intersecti = "Y",
                   Bus_Stop_Location = "FS", Bus_Stop_Land_Use = "Commercia
l",
                   StopIntDist = 154.303, CrashStopDist = seq(0.58, 340.27,
length.out = 100))
#Use the median values of each predictor category to hold them all constant

newdat$pred1 <- predict(model4,newdat,type = "response")
plot(pred1~CrashStopDist,data = newdat, type = "l", xlab = "Distance of a C
rash from Bus Stop", ylab = "Probability (%) a Crash is VRU Related",main =
"VRU Crash Probability 250 ft.")
#Use the median values of each predictor category to hold them all constant

newdata<-data.frame(CrashStopDist = 140.4, DOT_AADT = 25000, AvgBoard = 5,
                   AvgAlighting = 6, Intersecti = "Y",
                   Bus_Stop_Location = "FS", Bus_Stop_Land_Use = "Commercia
l",
                   StopIntDist = 154.303, SPEED_LMT = seq(0, 65,length.out
= 100))
#Use the median values of each predictor category to hold them all constant

newdata$pred1 <- predict(model4,newdata,type = "response")
plot(pred1~SPEED_LMT,data = newdata, type = "l", xlab = "Speed Limit (mph)"
, ylab = "Probability (%) a Crash is VRU Related",main = "VRU Crash Probabi
lity 250 ft.")
#Use the median values of each predictor category to hold them all constant
```

```
tmp.w <- as.numeric(names(model4$y))
model4.1 <- update(model4, . ~ ns(CrashStopDist, 2) +
                     ns(SPEED_LMT, 2) +
                     DOT_AADT +
                     AvgBoard +
                     Intersecti +
                     Bus_Stop_Location +
                     Bus_Stop_Land_Use +
                     StopIntDist +
                     ns(CrashStopDist, 2)*Bus_Stop_Location, subset=tmp.w)
anova(model4.1, model4)

summary(model4.1)

drop1(model4.1, test='Chisq')

library(emmeans)
em4.1 <- emmeans(model4.1, ~Bus_Stop_Land_Use)
plot(em4.1, type="response", xlab = "Probability (%) a Crash is VRU Related
(250 ft.)", ylab = "Bus Stop Land Use", main = "VRU Crash Probability 250 f
t.")

plot(pairs(em4.1, type="response"))

em4.2 <- emmeans(model4.1, ~Bus_Stop_Location)

plot(em4.2, type="response", xlab = "Probability (%) a Crash is VRU Related
(250 ft.)", ylab = "Bus Stop Location", main = "VRU Crash Probability 250 f
t.")

# Get the estimated marginal means (predicted probabilities) for each locat
ion
location_means <- emmeans(model4.1, ~ Bus_Stop_Location, type = "response")

# See the predicted probability for each specific location
print(location_means)

# Perform pairwise comparisons to see which specific locations differ from
one another, fdr adjustment
pairs(location_means, adjust = "none")

exp(coef(model4.1))

exp(confint(model4.1))
```

APPENDIX F: LOGISTIC REGRESSION RESULTS TABLES

This appendix contains the logistic regression results tables displaying the relationship between VRU-related crashes and tested independent variables. The tables report the number of variables, estimates, deviance, and statistical relationship between VRU crashes and a given variable. These results suggest that while there are some variables that impact VRU safety, the presence of a transit stop is not a significant factor. Table F.1 provides a summary of the statistical analysis results tables.

Table F.1 Statistical Analysis Results Tables Summary

Results	Table
Logistic Regression Results (250 ft. Analysis)	Table F.2
Single-Term Deletion Test Results (250 ft. Analysis)	Table F.3
Full and Reduced Test Results (1,400 ft. Analysis)	Table F.4
Full and Reduced Test Results (250 ft. Analysis)	Table F.5
Logistic Regression Results – Reduced Model (250 ft. Analysis)	Table F.6
Single-Term Deletion Test Results – Reduced Model (250 ft. Analysis)	Table F.7
Odds Ratios (250 ft. Analysis)	Table F.8

F.1 Logistic Regression Modeling

F.1.1 Full Model

Table F.2 Logistic Regression Results (250 ft. Analysis)

Variable	Estimate	Standard Error	P-Value
(Intercept)	-3.260	0.395	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1	-0.038	0.309	0.903
ns(Crash-to-Stop Distance, 2)2	-0.482	0.351	0.169
ns(Speed Limit, 2)1	0.066	0.635	0.918
ns(Speed Limit, 2)2	-1.306	0.342	< 0.001 ***
AADT	< 0.001	< 0.001	< 0.001 ***
Average Boarding	0.004	0.001	< 0.001 ***
Average Alighting	0.001	0.001	0.374
Intersection Involved - Yes	0.241	0.053	< 0.001 ***
Stop Location - Cen	-1.783	1.264	0.158
Stop Location - FS	-0.56	0.222	0.012 *
Stop Location - MB	-0.263	0.272	0.334

Table F.2 Continued

Variable	Estimate	Standard Error	P-Value
Stop Land Use - Residential/Mixed	-0.062	0.057	0.275
Stop Land Use - Institutional	-0.155	0.083	0.062
Stop Land Use - Agricultural	-0.541	0.128	< 0.001 ***
Stop Land Use - Recreational	0.155	0.153	0.312
Stop to Intersection Distance	-0.001	< 0.001	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	-0.889	2.205	0.687
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	-7.563	4.911	0.124
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	0.784	0.429	0.068
ns(Crash-to-Stop Distance, 2)2: Stop Location - FS	-0.586	0.431	0.173
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	0.443	0.533	0.407
ns(Crash-to-Stop Distance, 2)2: Stop Location - MB	-0.394	0.575	0.493

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Table F.3 Single-Term Deletion Test Results (250 ft. Analysis)

Variable	Degrees of Freedom	AIC	P-Value
<none>		17766	
ns(Speed Limit, 2)	2	17779	< 0.001 ***
AADT	1	17789	< 0.001 ***
Average Boarding	1	17790	< 0.001 ***
Average Alighting	1	17764	0.385
Intersection Involved	1	17785	< 0.001 ***
Stop Land Use	4	17782	< 0.001 ***
Stop to Intersection Distance	1	17776	< 0.001 ***
ns(Crash-to-Stop Distance, 2): Stop Location	6	17761	0.277

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Table F.4 Full and Reduced Test Results (1,400 ft. Analysis)

Model	Residual Degrees of Freedom	Deviance Degrees of Freedom	Degrees of Freedom	Deviance	P-value
1 (Reduced)	110037	33532			
2 (Full)	110036	33532	1	0.0323202	0.8574

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Table F.5 Full and Reduced Test Results (250 ft. Analysis)

Model	Residual Degrees of Freedom	Deviance Degrees of Freedom	Degrees of Freedom	Deviance	P-value
1 (Reduced)	53005	17720			
2 (Full)	53004	17720	1	0.75475	0.385

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

F.1.2 Reduced Model

Table F.6 Logistic Regression Results – Reduced Model (250 ft. Analysis)

Variable	Estimate	Standard Error	P-Value
(Intercept)	-3.251	0.395	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1	-0.030	0.309	0.922
ns(Crash-to-Stop Distance, 2)2	-0.480	0.351	0.171
ns(Speed Limit, 2)1	0.048	0.635	0.940
ns(Speed Limit, 2)2	-1.315	0.342	< 0.001 ***
AADT	< 0.001	< 0.001	< 0.001 ***
Average Boarding	0.004	< 0.001	< 0.001 ***
Intersection Involved - Yes	0.241	0.053	< 0.001 ***
Stop Location - Cen	-1.720	1.264	0.174
Stop Location - FS	-0.556	0.222	0.012 *
Stop Location - MB	-0.258	0.272	0.343
Stop Land Use - Residential/Mixed	-0.064	0.057	0.258
Stop Land Use - Institutional	-0.153	0.083	0.067
Stop Land Use - Agricultural	-0.543	0.128	< 0.001 ***
Stop Land Use - Recreational	0.152	0.159	0.319
Stop to Intersection Distance	-0.001	< 0.001	< 0.001 ***
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	-0.757	2.201	0.731
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	-7.363	4.866	0.130
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	0.772	0.429	0.071
ns(Crash-to-Stop Distance, 2)2: Stop Location - FS	-0.591	0.430	0.169
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	0.424	0.533	0.426
ns(Crash-to-Stop Distance, 2)2: Stop Location - MB	-0.404	0.575	0.482

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Table F.7 Single-Term Deletion Test Results – Reduced Model (250 ft. Analysis)

Variable	Degrees of Freedom	AIC	P-Value
<none>		17764	
ns(Speed Limit, 2)	2	17778	< 0.001***
AADT	1	17787	< 0.001***
Average Boarding	1	17824	< 0.001***
Intersection Involved	1	17783	< 0.001***
Stop Land Use	4	17781	< 0.001***
Stop to Intersection Distance	1	17775	< 0.001***
ns(Crash-to-Stop Distance, 2): Stop Location	6	17760	0.294

*** < 0.001, ** < 0.01, * 0.05, ^ <0.01

Although the interaction term between crash-to-stop distance and stop location was statistically insignificant at the 250 ft. distance threshold, the term was included in the reduced model because the interaction term was significant at the 1,400 ft. distance threshold and model consistency was desired.

F.2 Odds Ratios

Table F.8 Odds Ratios (250 ft. Analysis)

Variable	Odds Ratio	Confidence Interval 2.5%	97.5%
(Intercept)	0.039	0.018	0.083
ns(Crash-to-Stop Distance, 2)1	0.970	0.532	1.786
ns(Crash-to-Stop Distance, 2)2	0.619	0.306	1.209
ns(Speed Limit, 2)1	1.049	0.309	3.720
ns(Speed Limit, 2)2	0.268	0.135	0.515
AADT	1.000	1.000	1.000
Average Boarding	1.004	1.003	1.005
Intersection Involved - Yes	1.272	1.147	1.414
Stop Location - Cen	0.179	0.010	1.566
Stop Location - FS	0.573	0.371	0.886
Stop Location - MB	0.773	0.450	1.309
Stop Land Use - Residential/Mixed	0.938	0.838	1.048
Stop Land Use - Institutional	0.858	0.727	1.007
Stop Land Use - Agricultural	0.581	0.449	0.741
Stop Land Use - Recreational	1.165	0.852	1.553

Table F.8 Continued

Variable	Odds Ratio	Confidence Interval 2.5%	97.5%
Stop to Intersection Distance	0.999	0.999	1.000
ns(Crash-to-Stop Distance, 2)1: Stop Location - Cen	0.469	0.007	49.091
ns(Crash-to-Stop Distance, 2)2: Stop Location - Cen	0.001	< 0.001	2.429
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	2.165	0.936	5.023
ns(Crash-to-Stop Distance, 2)1: Stop Location - FS	0.554	0.240	1.299
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	1.528	0.542	4.383
ns(Crash-to-Stop Distance, 2)1: Stop Location - MB	0.668	0.213	2.036