

Taking Steps Towards Pedestrian Safety: A Case Study of Charlotte, North Carolina

Cheyne Campbell

CASA0010 for MSc Dissertation

Supervised by Dr. Andrew MacLachlan

Word Count: 10,365

This dissertation is submitted in partial
fulfillment of the requirements for the degree of
MSc in the Centre for Advanced Spatial Analysis,
Bartlett Faculty of the Built Environment, UCL.

24 August 2020

Abstract

Through global and local policies, improved walkability and pedestrian safety are being promoted as important transportation objectives in cities around the world due to the corresponding benefits to public health, the economy, and the environment. Spatial data methodologies are opening new doors for pedestrian safety analysis, allowing for contextually appropriate, data-driven decision making and adjustments to current strategies. This study used the methodology proposed by Kumfer et al. (2019) to investigate if features of the built environment can be used to predict the locations of pedestrian crashes at the street level in Uptown Charlotte, North Carolina. Through the use of random forest and negative binomial regression, the number of traffic signals was determined to be one of the most significant features in identifying high-risk roads. The results show that current, city-wide pedestrian safety policies may not be helpful in Uptown Charlotte, and that a more targeted approach should be taken. For example, on three of the most dangerous segments highlighted by the model, reducing the roads to single-direction and single-lane could reduce pedestrian crashes by up to 50.4%. This paper advocates for enhanced pedestrian count programs in Charlotte, the data from which would enable pedestrian volume modeling and improved interpretative capabilities in analyses such as this one. Further, by using open data and contending with missing information, the study demonstrates that the methodology could be applied to a range of cities.

Declaration of Authorship

I, Cheyne Campbell, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 10,365 words in length.

A handwritten signature in black ink, appearing to read "Cheyne Campbell".

.....
Cheyne Campbell

24 August 2020

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Andrew MacLachlan, for his invaluable advice and assistance throughout this process. I would also like to thank Scott Curry and Keith Bryant at the City of Charlotte Department of Transportation for their time and insights. Thanks also to Dr. Farzaneh Bahrami for her guidance in the early stages of the project. Finally, I am extremely grateful to my family and friends for their support and encouragement.

Contents

List of Figures	7
List of Tables	8
1 Introduction	9
2 Literature Review	12
2.1 Benefits of Walkability	13
2.2 Pedestrian Safety Pandemic	14
2.2.1 Pedestrian Policies, Globally and Locally	15
2.2.2 Factors Influencing Pedestrian Safety	17
2.3 Pedestrian Data Dilemmas	20
2.4 Methods for Considering Pedestrian Safety	21
3 Methodology	25
3.1 Study Area	25
3.2 Data	27
3.2.1 Ethical Evaluation	32
3.3 Methodology Overview	34
3.3.1 Random Forest	36
3.3.2 Negative Binomial Regression	36
3.3.3 Prediction and Interpretation	38
4 Results	39

<i>Contents</i>	6
5 Discussion	47
5.1 Interpretation of Model	47
5.2 Assessment of Variables	48
5.3 The Need for Pedestrian Volume Data	49
5.4 Pedestrian Safety Policy Contextualization	50
5.5 Limitations	52
6 Conclusion	53
Bibliography	55
Appendices	71
A Pedestrian Crash Descriptive Stats	71
B Random Forest Variable Importance	74
C Research Log	75

List of Figures

2.1	Proposed ‘Complete Street’ Redesign of North Tryon Street.	17
2.2	Safe Systems Approach Components.	18
3.1	Heatmap of Pedestrian Crashes in City of Charlotte (2007-2018) . .	26
3.2	Heatmap of Pedestrian Crashes in Uptown Charlotte (2007-2018) . .	27
3.3	Variable Maps.	30
3.4	Variable Maps (continued).	31
3.5	Methodology Flowchart.	35
3.6	Correlation Matrix for Continuous and Discrete Variables.	37
4.1	Sample Decision Tree.	40
4.2	Variable Importance.	40
4.3	Actual Pedestrian Crashes.	43
4.4	Predicted Pedestrian Crashes.	43
4.5	Most Dangerous Roads.	44
4.6	Google Street View Images A-C.	45
4.7	Google Street View Images D-F.	46
5.1	Vision Zero Interactive Feedback Map.	51
A.1	Temporal Heatmap of Pedestrian Crashes (2007-2018)	71
A.2	Histogram of Pedestrian Crash Types.	72
A.3	Exploratory Bar Charts for Pedestrian Crash Attributes.	73

List of Tables

2.1	Features of the Built Environment and Pedestrian Safety.	19
3.1	Data Sources, Cleaning, and Processing.	28
3.2	Continuous and Discrete Variables.	32
3.3	Categorical Variables.	33
4.1	Random Forest Performance Metrics.	39
4.2	Negative Binomial Regression Model.	41
4.3	Negative Binomial Regression Performance Metrics.	42
4.4	Top 1% Most Dangerous Roads.	44
B.1	Variable Importance.	74

Chapter 1

Introduction

Traffic safety problems are causing incommensurate harm to pedestrians in cities around the world (Bartolomeos et al. 2013). In the United States, 49,340 pedestrians were killed by moving vehicles between 2008 and 2017 (Zaccaro 2019). As walkability becomes more central in the minds of transportation planners, government officials, and citizens due to the corresponding advantages in the realms of health, wealth, and happiness (see section 2.1), it is critical that pedestrian traffic fatalities come to be seen not as inevitable accidents, but as ‘preventable and unacceptable’ occurrences (City of Charlotte DOT 2019, p. 8). Further, it is essential that burgeoning cities like Charlotte, North Carolina, where this study is focused, are able to grow in ways that are both sustainable and appealing. In such places, effectively addressing pedestrian safety and walkability will be one of the keys to future success. Accordingly, this goal is an important component of various global and local policies (see section 2.2).

The issue of pedestrian traffic safety has not escaped the attention of researchers and has been studied from a variety of perspectives and with a range of techniques, both spatial and non-spatial (see sections 2.2.2 and 2.4). Past investigations have largely focused on hot spot identification and non-spatial assessments of causality. While the subject has been previously explored, there is more work to be done when it comes to using spatial data to determine features of the built environment that are related to pedestrian collisions, and subsequently to prioritize dangerous areas for correction using specific design alterations that will make

the urban environment more conducive for walking (Kumfer et al. 2019, Taquetchel 2009). The results of such analyses could be very helpful in supplementing or amending current strategies for pedestrian safety, which generally lack spatial, evidence-based support (see section 2.2.1).

It is important that this research is attempted in the context of Charlotte because the city is expanding quickly while simultaneously struggling with pedestrian safety (Henderson 2020) and a car-dominated transportation system (City of Charlotte DOT 2017), as explained in section 5.4. As argued in Charlotte's *Vision Zero* Action Plan, 'The transportation system should be designed so mistakes are not fatal '(City of Charlotte DOT 2019, p. 8). Spatial data analysis has the potential to help Charlotte reach its walkability and traffic safety goals (City of Charlotte DOT 2017, 2019). By adjusting the urban environment using data-driven decision-making, Charlotte's streets can be made more welcoming for pedestrians and safer for all types of users.

The research question addressed here is: *Can features of the built environment be used to predict the locations of pedestrian crashes at the street level in Uptown Charlotte, North Carolina?* This study hopes to answer this question using spatial data about pedestrian crashes and features of the built environment and follows the methodology proposed by Kumfer et al. (2019), which includes the combined use of random forest and negative binomial regression. In attempting to answer the research question, this paper also aims to address the following objectives:

1. What features of the built environment are thought to influence pedestrian safety, and can our understanding of these relationships be strengthened?
2. What are the challenges when it comes to studying pedestrian safety in terms of the lack of available pedestrian data? Does Charlotte collect and make public the data required for this type of analysis?
3. Most importantly, how can the findings of this research be used to improve upon existing policy or urban design strategies by highlighting targeted areas and adjustments for improved pedestrian safety?

There is hope that the hegemony of the automobile in United States planning and urban design is petering out (Speck 2013), but car-dominated policies have left their mark on everything from the physical streetscape, to the phrasing of laws about how a citizen can move in the city, to the types of transportation data that are commonly collected (City of Charlotte DOT 2017). Ultimately, this study intends to contribute to the field of traffic safety research by further investigating the possibilities for data-driven advocacy for pedestrian needs and solutions. In the words of the visionary Enrique Peñalosa, ‘The essence of the conflict today, really, is cars and people...We can have a city that is very friendly to cars, or a city that is very friendly to people. We cannot have both’ (NYC Streets Renaissance 2006).

Chapter 2

Literature Review

‘Through our feet, we are reminded that the planet is a whole thing, and that we are animals designed to traverse it with a sure step and elongated spines’ (Malchik 2019, p. 3). Many cities across the United States, and indeed across the world, are realizing that they should be designed to better facilitate humankind’s most natural form of travel, walking. When a place accommodates or even encourages walking, it has been shown that a host of economic, public health, and environmental advantages follow (see section 2.1). This means that cities, including London (TfL 2018), New York City (NYC DOT 2020), and even Charlotte, North Carolina (City of Charlotte DOT 2017) are increasingly seeking to make data-driven design choices that promote walking as a form of transportation and recreation. Of course, encouraging walking involves taking steps to prevent pedestrian traffic injuries and deaths, which are a huge problem in cities around the world (see section 2.2).

A lack of spatial data about pedestrian behaviours, crash incidents, and street or intersection design can make it difficult to decide what changes to make to the urban environment to increase pedestrian safety, or to measure the impacts of changes that have already been made (see section 2.3). Some cities, such as New York City and Seattle, maintain open data platforms that make the relevant information available to the public. Other cities, like Nairobi, do not, which complicates the research process and makes it difficult to suggest data-driven adjustments that may improve pedestrian safety. Section 2.4 discusses methods that have been implemented in the past, using the data that does exist, to study the spatial nature of pedestrian crashes.

This chapter builds upon the concepts introduced in Chapter 1, and is divided into four sections that cover the benefits of walkability (2.1), pedestrian safety issues including pertinent policies and factors influencing pedestrian safety (2.2), pedestrian data dilemmas (2.3), and methods for considering pedestrian safety (2.4).

2.1 Benefits of Walkability

The benefits of enabling safe active travel are numerous. For instance, the walkability benefits framework developed by Arup includes sixteen different categories (Claris & Scopelliti 2016, p. 30). While this section will focus on advantages that extend to the economy and to public health, improved walkability also has a positive impact on the environment, which will not be discussed at length here (Claris & Scopelliti 2016, pp. 68-83).

As explained by Speck (2013), the increased desire for walkability in the United States is the product of a cultural shift; feelings of personal freedom are no longer derived from motorized travel, but instead from living in close, walkable proximity with the people and activities that bring one joy (Claris & Scopelliti 2016, pp. 21-22). This cultural shift is particularly evident in the preferences of young, ‘creative class’ professionals and older ‘empty nesters’, who are attracted to the conveniences and excitements of more urban lifestyles and repelled by car dependency (Speck 2013, p. 21). These changing cultural preferences correspond with economic benefits when a city or town is able to offer walkable real estate, which Speck (2013) notes is in very short supply. Speck also discusses how the construction of public transit infrastructure, including sidewalks and bike paths, creates more jobs than highway construction projects (Speck 2013, p. 31), and that walkability fosters a collaborative atmosphere that leads to more innovation, and subsequently more economic prosperity (Speck 2013, p. 33). These concepts are echoed by Litman (2003), who described how walkability can lead to public and personal cost savings, economic development, and improved social equity. It is for these reasons that municipalities across the United States, including Boston, Detroit, and Pittsburgh (Loh et al. 2019, p. 25), are seeking to become more walkable,

and to avoid becoming ‘places that are not worth caring about’ (Kunstler 2004).

There are also copious health benefits of walkability. These stem from both the direct, positive effects of walking more, such as lowered obesity and heart disease rates, and the reduced negative effects of cars, such as pedestrian collisions and exhaust that pollutes the air and infects our lungs (Speck 2013, p. 38). Walking can have benefits not only for physical health, but for mental well-being and happiness as well. This idea is addressed at length in *Happy City* by Montgomery (2013) and *A Walking Life* by Malchik (2019). Further, the current COVID-19 pandemic has resulted in a shift in transport mode usage due to the necessity for social distancing. For the time being, active travel is becoming more prominent as it replaces crowded public transportation, and many cities, like Bogota (Hidalgo 2020) and Vancouver (City of Vancouver 2020), are in the process of expanding their active travel infrastructure to accommodate increasing demand and to prevent the spread of the disease (Bliss et al. 2020).

As the aforementioned results of walkability become more alluring, or in the case of COVID-19, absolutely necessary, cities across the world are investigating how to best support walking as a form of recreation and a mode of transportation (Hirst 2019). Advancing walkability involves addressing a different kind of pandemic, as described by Mohan (2003): the persistent problem of pedestrian traffic deaths and injuries. It is important to balance an increased desire, or need, for walkability with infrastructure that facilitates *safe* walking journeys. Further, it is critical that, in the spirit of equitable development (Khayesi et al. 2010), this infrastructure exists in places where walking is the only transportation option for some communities (Pojani & Stead 2015, p. 7792). The issue of pedestrian safety will be expanded upon in the following section.

2.2 Pedestrian Safety Pandemic

According to the World Health Organization, twenty-two percent of all road fatalities each year are pedestrians (Bartolomeos et al. 2013, p. vii). This unsettling statistic can perhaps be attributed to the rapid growth of urban centers, paired with

a lack of effort to expand pedestrian facilities (Halais 2020). Indeed, pedestrian safety is a problem all over the world. In cities in sub-Saharan Africa, more than half of all trips occur by foot (Pendakur 2005, p. 10), yet pedestrians are regularly and disproportionately involved in incidents with motor vehicles due to a lack of sufficient walkways, lighting, and other helpful urban design provisions (Halais 2020, Murguía 2018). Pedestrian safety is an enduring problem in the United States as well. According to Smart Growth America, ‘Between 2008 and 2017, drivers struck and killed 49,340 people who were walking on streets all across the United States...the equivalent of a jumbo jet full of people crashing — with no survivors — every single month’ (Zaccaro 2019, p. 2).

In the United States, seventy-six percent of pedestrian deaths happen in urban environments (Bartolomeos et al. 2013, p. 14). This statistic emphasizes the importance of considering pedestrian crash data spatially and at a range of spatial scales. To be more geographically specific, Smart Growth America lists the southern continental United States and low-income communities as being more heavily effected by pedestrian collisions (Zaccaro 2019, p. 2). This research uses Charlotte, North Carolina as a case study; according to Smart Growth America’s *Pedestrian Danger Index*, North Carolina is the thirteenth most dangerous state for walking and the Charlotte-Gastonia-Concord region is the thirty-third most dangerous metropolitan area (Zaccaro 2019, p. 12, 28).

2.2.1 Pedestrian Policies, Globally and Locally

Due to the many benefits and its inseparable relationship with pedestrian safety, improving walkability is integral to achieving international policy objectives. Thornton (2015) highlights the indispensable role that walking will play in reaching the United Nation’s Sustainable Development Goals for 2030, particularly the following targets:

- Goal 3: Good Health and Well-being
- Goal 9: Industry, Innovation and Infrastructure
- Goal 11: Sustainable Cities and Communities

While not specifically mentioned in the Sustainable Development Goals, walkability forms the core of several other important, international policy initiatives. For example, the United Nation's *Share the Road Programme* is intended to 'provide global leadership and support to encourage and advocate for systematic investment in Non-Motorized Transport (NMT) as one of the key sustainable solutions to global transport challenges' (de Jong et al. 2019, p. 11). Other initiatives like the World Bank's *Walkability Index* (Krambeck 2006), the Institute for Transport Development Policy's *Pedestrians First* toolkit (ITDP 2018), and *Vision Zero* (Vision Zero Network 2018) demonstrate that pedestrians' needs and safety are being prioritized at the global level, and gaining subsequent attention from national and local governments.

Charlotte was chosen for this study because, as a rapidly growing 'Sun Belt' city, it is positioned to both resolve existing pedestrian safety issues and anticipate the rising interest in walkability through urban design techniques. The City of Charlotte has exhibited an interest in becoming more walkable and safe for pedestrians as it expands. This is evident in development projects like the *Charlotte Center City 2020 Vision Plan* (MIG, Inc. 2011) and the *North Tryon Vision Plan* (MIG, Inc. 2016), as well as in more comprehensive policy documents like the *Charlotte WALKS Pedestrian Plan* (City of Charlotte DOT 2017), Charlotte's *Vision Zero Action Plan* (City of Charlotte DOT 2019), and the *Urban Street Design Guidelines* (City of Charlotte Staff 2007). All of these schemes prioritize improving pedestrian safety, connections, and infrastructure. For example, the *North Tryon Vision Plan* involves a 'complete streets' approach to making some roads safer for walking through design (see figure 2.1).

Acuto notes that, 'Cities and city networks are poorly linked to national and international policy frameworks' (Acuto 2016, p. 613). While the aforementioned initiatives encourage improved conditions for walking, such campaigns often lack a discussion of how the inclusion of geospatial data can allow for the targeted optimization of resources to further pedestrian safety in specific places. This research explores how the analysis of geospatial crash and street characteristic data can sup-



Figure 2.1: Proposed ‘Complete Street’ Redesign of North Tryon Street. From the *North Tryon Vision Plan* for Charlotte, prepared by MIG, Inc. (2016).

plement and improve upon urban planning policy standpoints.

2.2.2 Factors Influencing Pedestrian Safety

Increased interest in walkability and concern over pedestrian traffic crashes has coincided with an effort to discern what exactly causes these incidents and how they could best be prevented. Factors such as traffic regulations, driver and pedestrian behaviours, alcohol consumption, and time of day, week or year have been found to correspond with pedestrian-involved crashes (National Highway Traffic Safety Administration 2017, International Transport Forum 2012). While studying such potential causes can lead to interesting insights, it has been argued that human behaviour is difficult to change and that the built environment should instead be adjusted to accommodate behaviours (Taquechel 2009). In this vein, Khayesi et al. (2010) emphasize that it is critical that pedestrians are not blamed for the failures of government policies and planning, and that a holistic, ‘systems approach’ (see figure 2.2) is required to deal with traffic safety issues. The PEDSAFE tool (Zegeer et al. 2013), for example, suggests a range of countermeasures based on the specific type of pedestrian crash, reflecting the idea that dangerous pedestrian behaviours sometimes stem from certain urban conditions.

There have been sustained and contrasting efforts to discern and describe what

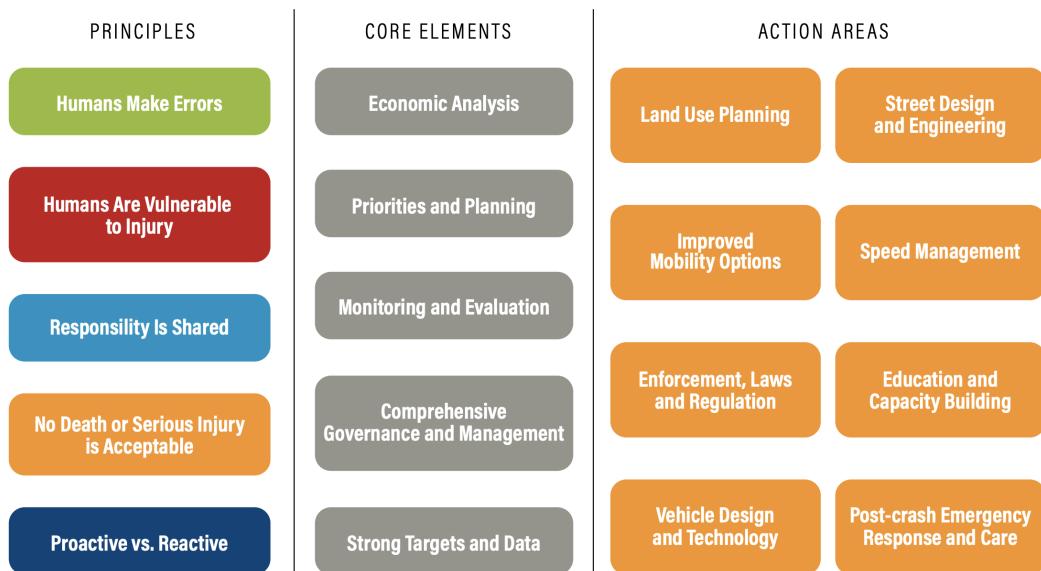


Figure 2.2: Safe Systems Approach Components. From *Sustainable & Safe: A Vision and Guidance for Zero Road Deaths*, published by the World Resources Institute (Welle et al. 2018, p. 6).

kinds of built environments are the most pedestrian-friendly (Southworth 2005, p. 250). The report by Arup outlines urban design actions that can be taken to improve the experience of walking, including techniques such as road diets that reduce the number of vehicular traffic lanes, infrastructure reuse like the High Line in Manhattan, and greenways or pocket parks (Claris & Scopelliti 2016, pp. 118-127). Most relevantly for the purpose of this research, Arup's report notes that mapping for pedestrian safety can be used to inform data-driven urban design (Claris & Scopelliti 2016, p. 127). Table 2.1 includes a list of built environment variables that have been shown to correspond with pedestrian safety, such as the presence of sidewalks, traffic calming measures, land use, and network connectivity.

As is one of the goals of this study, further statistical relationships should be identified to reinforce anecdotal and existing quantitative evidence that street-level features of the urban environment impact pedestrian safety. Some existing studies explore this topic from a spatial perspective (see section 2.4), but there is certainly room for further investigation. Before moving on to a discussion of relevant data sources and methodologies, it should be noted that many of the challenges facing pedestrians do not necessarily involve a vehicular crash (Bartolomeos et al. 2013,

Table 2.1: Features of the Built Environment and Pedestrian Safety.

Feature	Impact on Pedestrian Safety
Presence and Condition of Sidewalks	If sidewalks are missing or in disrepair, pedestrians may be forced to walk in the road. This heightens the chance of a collision (Albers et al. 2010, Murguía 2018).
Land Use	Commercial and mixed land uses correspond with more pedestrian crashes (Moradi et al. 2017), because there are more pedestrians in these areas (Southworth 2005). One study found that most walking journeys do not originate from the home (Millward et al. 2013), reflecting that there may be less pedestrians in residential areas to crash into.
Street Trees and Landscaping	Street trees, planting zones, and urban vegetation can protect pedestrians from vehicular traffic and make them feel safer on sidewalks (Murguía 2018).
Crosswalks	The quality and frequency of crosswalks is known to influence pedestrian injury rates. Crosswalks prevent pedestrians from having to cross multiple lanes of traffic when they do not have the right of way (Albers et al. 2010, Murguía 2018, International Transport Forum 2012).
Intersection Characteristics	Intersections should be designed such that pedestrians do not have to wait for long periods of time, and are clearly visible to turning drivers (Babinard et al. 2018).
Lighting	Street lighting makes walking safer because it makes pedestrians more visible at nighttime, when a lot of pedestrian accidents happen (Albers et al. 2010, de Andrade et al. 2014, Murguía 2018, Stoker et al. 2015).
Road Qualities	Road qualities such as length, curvature, width and number of lanes have been found to correlate with pedestrian crashes because these factors can influence both driver and pedestrian behaviour (de Andrade et al. 2014, Stoker et al. 2015).
Traffic Calming Measures	Traffic calming measures like speed bumps and slow speed limit zones are used to reduce driver speed, and subsequently reduce pedestrian fatalities (International Transport Forum 2012, Stoker et al. 2015).
Pedestrian Network Connectivity	Denser networks with shorter block lengths can reduce pedestrian crashes by giving pedestrians more chances to cross the road and by discouraging vehicles from accelerating too quickly (Babinard et al. 2018).

p. 12), and can include commonplace occurrences like falls or slips, which may even comprise up to seventy-five percent of injuries (International Transport Forum 2012, p. 37). While this research focuses on pedestrian safety as it pertains to pedestrian-vehicle collisions, there is a large body of research that deals with the overall experience of walking via the creation of walkability indexes and pedestrian level of service analyses, while contending with a troubling lack of data in this area (Zegeer et al. 2013, Otieno 2013, Albers et al. 2010).

2.3 Pedestrian Data Dilemmas

Accurate and detailed pedestrian data is notoriously difficult to obtain. Pedestrian data has been commonly derived from surveys, although Litman (2003) acknowledges that walking is often undercounted in such surveys, which tend not to recognize shorter trips or linkages between other transport modes (International Transport Forum 2012, Litman 2003). Importantly, Litman also purports that such discrepancies can actually shift decision-making in favor of the automobile, because it is difficult to prove the critical role that walkability plays in the larger transport network (Litman 2003, p. 5). Many studies have also used manual pedestrian count data, which comes with limitations as the counts are only conducted at a small number of locations and times, and are subject to a high degree of human error (Lai & Kontokosta 2017).

While many novel sources of pedestrian data are currently being explored, such as video surveillance footage (Leeson et al. 2014, Jahangiri et al. 2019, Petrasova et al. 2019), crowdsourced GPS data (Kapenekakis & Chorianopoulos 2017, Lee & Sener 2017), Wi-Fi traces (Kurkcu & Ozbay 2017), and sensor data (Benz et al. 2013, Kurkcu & Ozbay 2017), this is a developing field of research and the available data is sometimes not sufficient for accurately modeling human mobility or pedestrian movements (Leeson et al. 2014, Solmaz & Turgut 2019, Cesme et al. 2017). Privacy issues are one of the primary obstacles for these novel, automatic pedestrian data collection methods. Common privacy protection techniques are the anonymization of records and aggregation of route data, which, while necessary,

result in the loss of helpful information (Lee & Sener 2020).

Detailed pedestrian route data is not widely accessible, due in part to very legitimate privacy concerns, but data about traffic collisions involving pedestrians is more commonly available in the United States via open data websites, as is the case for Charlotte, North Carolina. However, not all countries collect this data or make spatial traffic crash data public, meaning it can be necessary to formally request the data from government agencies or mine and geolocate the data from crowdsourcing platforms (Qian 2016, Bedoya Arguelles et al. 2019, Resor 2015). Further, the World Health Organization points out that data on pedestrian crashes may under-represent the actual number of occurrences due to poor reporting of these events, which should be taken into account when considering the results of this analysis (Bartolomeos et al. 2013, p. 10).

When it comes to conducting research on specific elements of the streetscape, there can be additional challenges due to the spatial resolution of the data and its associated accuracy. For example, data on the locations and qualities of sidewalks and street furniture is most commonly collected via manual surveys of a specific site, which can of course be time consuming, resource intensive, and subject to error (Taquechel 2009, Albers et al. 2010). There are budding, innovative methodologies to collect such information more efficiently using crowdsourcing (Santani et al. 2015), combining machine learning, image processing, and triangulation (Mapillary 2020), and even remotely using Google Earth (Marshall & Garrick 2010, Taylor et al. 2011). When data about the built environment is accessible via an open data platform, the methods that were used for collecting this data should be considered.

Regardless of the obstacles concerning availability of and access to data, many studies have found ways to investigate pedestrian safety spatially. Some of these studies and the methods that they used are expanded upon in the following section.

2.4 Methods for Considering Pedestrian Safety

This final section of the literature review will discuss how relevant data, where it does exist, can generate findings that may inform local urban design choices that

better facilitate active travel and walkability. The methodologies described below were considered for their applicability to this paper's research question, and informed the choice of the methodology for this study.

Lai & Kontokosta (2017) used regression analysis to model pedestrian volumes at intersections based on contextual characteristics in New York City. They found that the streetscape influences pedestrian volumes differently at certain days and times, specifically weekdays as compared to weekends (Lai & Kontokosta 2017, p. 307). This study did not investigate pedestrian safety specifically, but its conclusions reaffirm that spatial, street-level characteristics influence pedestrian behaviours and can even be used to estimate pedestrian volumes.

Many studies related to pedestrian safety implement Moran's I and kernel density estimation to identify spatial autocorrelation and to map dangerous hot spots. The mapping of hot spots, also known as black spots, is a popular way of understanding the locations of pedestrian crashes, which can be particularly interesting in places where crash data is not openly reported (Resor 2015, Bedoya Arguelles et al. 2019, Bunnarong & Upala 2018, Moradi et al. 2016, Patel et al. 2016). These studies generally do not attempt to address causality and instead focus on identifying where and with what intensity pedestrian crashes happen, although they also tend to suggest that further research be conducted on what causes the identified hot spots to arise (Patel et al. 2016, Moradi et al. 2016).

Establishing causality in this sense has been attempted in an number of papers. Taquechel (2009) investigated the relationship between features of the built environment and pedestrian crashes in Atlanta, Georgia using only GIS visualization and descriptive statistics, specifically looking at lighting, driveways, street furniture, street direction, street conditions, signage, pedestrian signals, and public transit (Taquechel 2009). Building on this rather rudimentary approach, a study that looked at traffic fatalities on a single highway in Parana, Brazil implemented more advanced statistical methods such as kernel density estimation, wavelet analysis, built environment analysis and principle component analysis (PCA) to identify both traffic fatality hot spots and environmental factors causing these hot spots to occur

(de Andrade et al. 2014). However, a limitation of the use of PCA is that, as it aims to reduce dimensionality by combining variables into components, the original features become obscured and interpretation becomes more difficult (Jolliffe & Cadima 2016, p. 11).

In an attempt to advance previous methods, a study on downtown Tehran analyzed pedestrian traffic deaths using both ordinary least squares regression and geographically weighted regression (Moradi et al. 2017). In contrast with PCA, the use of regression allows for a clearer interpretation of how each variable relates to traffic deaths. Moradi et al. (2017) essentially concluded that there are more pedestrian-vehicle collisions in places where there are a lot of pedestrians and a lot of vehicles, but they did not directly control for pedestrian and traffic volumes. Doing so could have focused the findings on the effects of more interesting variables. This idea was attempted by Jahangiri et al. (2019), who used a variety of data sources to identify intersections in San Diego that were particularly dangerous for cyclists and pedestrians. The authors point out that a lack of active travel data can make it difficult to effectively prioritize areas for improvements, which they addressed by creating exposure models that allow for the consideration of active travel volumes, and subsequently determined the adjusted risk of crashes (Jahangiri et al. 2019). Significantly, the study found that normalizing for exposure and considering crash severity have notable impacts on which intersections were found to be the most dangerous (Jahangiri et al. 2019, p. 18).

To counter the limitations of regression analysis used by Moradi et al. (2017), Hezaveh et al. (2018) used Exhaustive Chi-square Automatic Interaction Detector (CHAID) to examine the relationship between pedestrian crash severity and a number of variables including pedestrian characteristics, road characteristics, and environmental factors. As explained in the study, CHAID is a non-parametric decision tree method that has a number of advantages when compared to logistic regression modeling that make it more suitable for this context, notably including its ability to handle categorical variables with many possible values (Hezaveh et al. 2018, p. 4). Based on the findings from Hezaveh et al. (2018), the use of CHAID had many ad-

vantages, but their methodology was not strongly applicable to evaluating the safety of the pedestrian environment from a spatial perspective.

The methodology utilized by Kumfer et al. (2019) was the most suitable for the research question being addressed here. By combining random forest to determine variable importance and negative binomial regression to model pedestrian crashes in the context of downtown Seattle, Washington, the authors aimed to develop ‘a framework that allows practitioners to identify and prioritize locations within a jurisdiction that are risky for pedestrians and to identify and implement effective, appropriate treatments at many such locations’ (Kumfer et al. 2019, p. 420). Kumfer et al. (2019) note that past crash history is not necessarily indicative of where future crashes will happen, explaining that it is for this reason that a systemic, network-wide evaluation should be used. Due to the similarity in objectives of that study and this one, it was decided to closely follow the methodology used by Kumfer et al. (2019), with some modifications. Specifically, this study adapts the original methodology to Charlotte, North Carolina, contends with different and missing data, expands upon model verification, and explores application processes.

Smart Growth America’s *Dangerous by Design* report recommends that ‘future research should explore how historic and ongoing street design practices contribute to [the high concentration of pedestrian fatalities in the Sun Belt]’ (Smart Growth America 2015, p. 10). This study attempts to follow the methodology proposed by Kumfer et al. (2019) to explore if features of the built environment can be used to predict the locations of pedestrian crashes in Charlotte, North Carolina. The results of the analysis will be followed by a discussion of how targeted urban design adjustments can make walking more safe.

Chapter 3

Methodology

Chapter 2 contextualized the issue of pedestrian safety and described previous research on the topic. This chapter will explain the methodology used in this research, the objective of which is to determine if features of the built environment can be used to predict unsafe areas for pedestrians, and in turn highlight targeted areas and specific urban design changes that should be made to protect pedestrians in Charlotte, North Carolina. The code used to conduct this analysis can be accessed here: <https://github.com/cheynecampbell/charlotte-pedestrian-safety>.

3.1 Study Area

This study focuses on the Uptown area of Charlotte, North Carolina. Despite recent efforts to advance walkability and address pedestrian-vehicle collisions (see section 2.2.1), the city continues to struggle with rising pedestrian deaths (Henderson 2020), suggesting that further research is required to identify where and why these crashes are occurring. The North Carolina Department of Transportation maintains a spatial database of crashes involving pedestrians from 2007-2018, so the first step was to generate heatmaps for data visualisation purposes. Examination of the heatmap for the City of Charlotte (see figure 3.1) showed that most of the incidents took place in the central business district, also known as Uptown, as indicated by the location of the red hot spot. Charlotte overall has a Walk Score, a walkability metric developed by real estate firm Redfin, of only 26 out of 100,

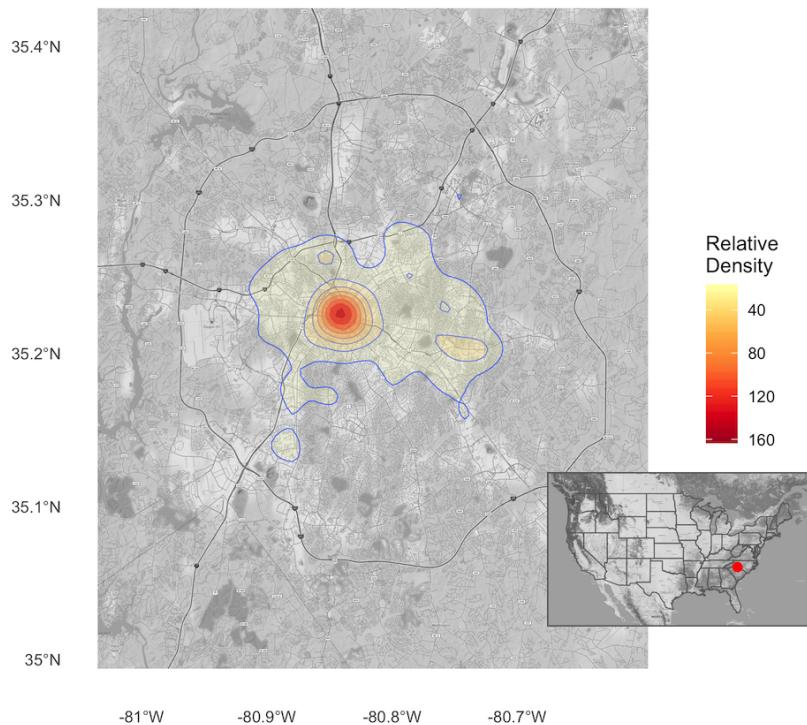


Figure 3.1: Heatmap of Pedestrian Crashes in City of Charlotte (2007-2018). On the color bar guide, a higher level contour corresponds with a higher density of pedestrian crashes in that area. Generated using kernel density estimation and without consideration of crash severity. Crash data from NCDOT DBPT (2020), basemap generated using ggmap (Kahle & Wickham 2013).

whereas properties in Uptown have Walk Scores of up to 95 (Walk Score 2020). As expected, the heatmap confirms that pedestrian crashes are spatially clustered in an area where walking is more feasible (City of Charlotte DOT 2017, p. 50). A second heatmap of Uptown Charlotte (see figure 3.2) shows that within this more walkable area, some places are more threatening to pedestrians than others.

For Charlotte to grow in a sustainable and appealing way, it is vital that effort is made to further walkability and pedestrian safety objectives across the entire city. This goal is reflected in the city-wide *Charlotte WALKS Pedestrian Plan* (City of Charlotte DOT 2017). However, since Charlotte was founded in 1768, Uptown was formed when ‘walking distance was...the natural yardstick for urban growth and design’ (City of Charlotte DOT 2017, pp. 28-29), whereas the peripheral areas were developed following the 1950s, in response to a surge in population growth and

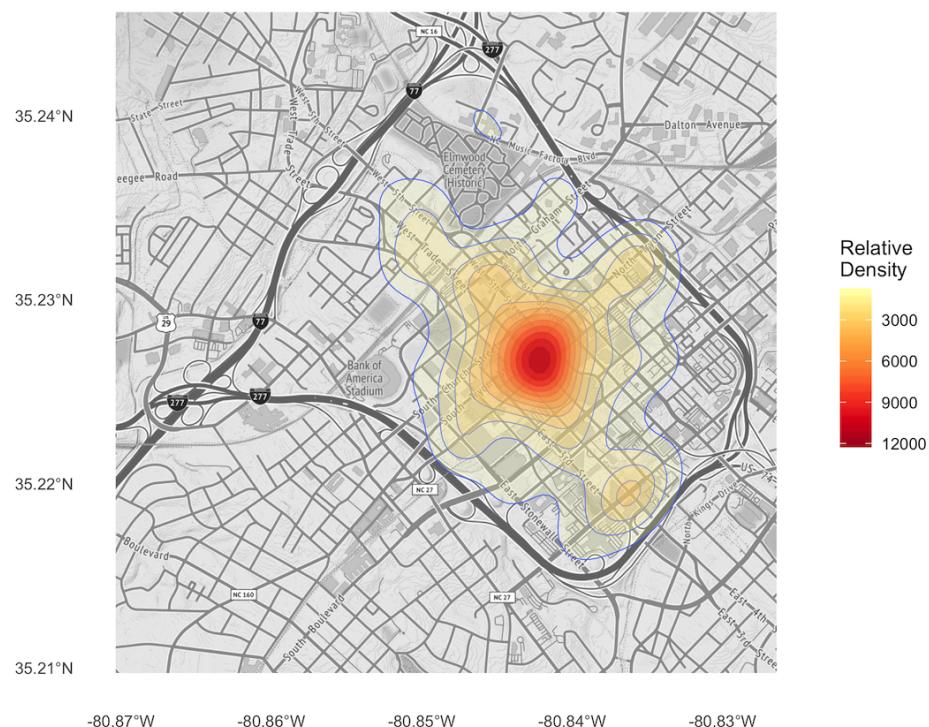


Figure 3.2: Heatmap of Pedestrian Crashes in Uptown Charlotte (2007-2018). Created using the same process described in figure 3.1.

alongside auto-oriented policies. Because of Charlotte's history, the Uptown area is distinct from the surrounding region and should be considered for different, more immediate walkability-oriented renovations. This concept is suggested in various Charlotte planning and policy documents, which will be expanded upon in section 5.4. While this analysis spotlights Uptown Charlotte, future research should include adjacent neighborhoods that have potential for improved walkability.

3.2 Data

As shown in table 3.1, the data was collected from a variety of open data platforms and included spatial boundaries, road segments, crash data, and features of the built environment. R (R Core Team 2019, Wickham et al. 2019) was used to reproject the data where needed, to clean each dataset accordingly, to extract only relevant information using the spatial boundaries, and to conduct point-in-polygon or other appropriate GIS overlay analyses to convert the crash data and features of the built

Table 3.1: Data Sources, Cleaning, and Processing.

Data	Description	Variables	Source
Charlotte Boundary	Used to geographically filter the other datasets.	-	Mecklenburg County (2020b)
Uptown Boundary	Used to geographically filter the other datasets.	-	Char-Meck Planning Dept. (2019)
Road Segments	Uptown Charlotte segments were extracted, excluding interstates and ramps. Segment length and edge centrality were calculated.	traveldirection lfttrnlntyp rttrnlntyp median leftshoulder rightshoulder thrulanecount surfacewidth funcclass length_ft edgebtw	NCDOT (2020)
Pedestrian Crashes	Pedestrian crashes 2007-2018. Only weekday crashes on non-interstate roads included.	crashes	NCDOT DBPT (2020)
Tax Parcel Building Data	Filtered and snapped to roadways based on address.	housingunits residential multifamily office commercial	Mecklenburg County (2020c)
AADT Volumes	Buffered; averaged where multiple polygons overlapped a road segment.	traffvols	Traffic Survey Group (2018)
Bike Lanes	Buffered to find overlaps with road segments.	bikelanes	Charlotte DOT (2019a)
Bus Stops	Filtered, snapped, counted.	busstops	Charlotte Area Transit (2019)
Light Rail Stations	Filtered, snapped, counted.	lightrail	Jenkins (2019a), Jenkins (2019b)
Traffic Signals	Filtered, snapped, counted.	trafficsignals	Charlotte DOT (2019c)
Trees	Filtered, snapped, counted.	trees	Porter (2018)
Parking Meters	Filtered, snapped, counted.	parkingmeters	Charlotte DOT (2019b)
Crosswalks	Filtered, snapped, counted.	crosswalks	OpenStreetMap Contributors (2020)

environment into attributes of the road network, which consisted of 683 segments. The process of searching for and preparing the relevant datasets took a significant

amount of time; 12 separate datasets were compiled to form the 25 variables used in the analysis.

To demonstrate the general procedure used for each dataset, the pedestrian crash data was processed as follows. Only crashes that occurred within the Uptown boundary were retained. Next, it was decided to filter out crashes that occurred on weekends. A heatmap of the days and times of crashes (see figure A.1) showed that crashes tended to be during peak commuting times on weekdays, and late at night on weekends, suggesting that the nature of these crashes might have been different. Further, this study excluded crashes that did not occur on local streets or on roadways (see figure A.3). This left 372 pedestrian crashes, which occurred from 2007 to 2018, remaining for analysis. The crash points were snapped to the nearest road segment, counted using buffered polygons of the road segments, and the resulting values were reassigned to the original line segments.

After data collection and processing, preliminary tables and maps were created to get a better understanding of the data through descriptive statistics and visualization (see tables 3.2 and 3.3 and figures 3.3 and 3.4).

Which variables to include were based on the findings of past studies discussed in the literature review (see section 2.2.2). Due to lack of availability or ubiquity, some variables that may have been helpful were not included. For example, the Charlotte DOT has a spatial dataset of sidewalks (Charlotte DOT 2018). As sidewalks are present on both sides of nearly every Uptown road, it was decided to exclude this variable. Additionally, street lighting data was not available through the Charlotte Open Data Portal, nor another easily accessible source. Since the study focuses on weekday crashes which largely happen during daylight hours, it was decided that this data could be left out as well. Most importantly, pedestrian volume data was not included. While there are several budding pedestrian count programs in Charlotte, the data collection points are either too sparse to estimate pedestrian volumes using proven methodologies (Institute for Transportation Research and Education 2020), or the data is not available publicly (CRTPO 2020). This exclusion allowed for a discussion about the need for pedestrian data.

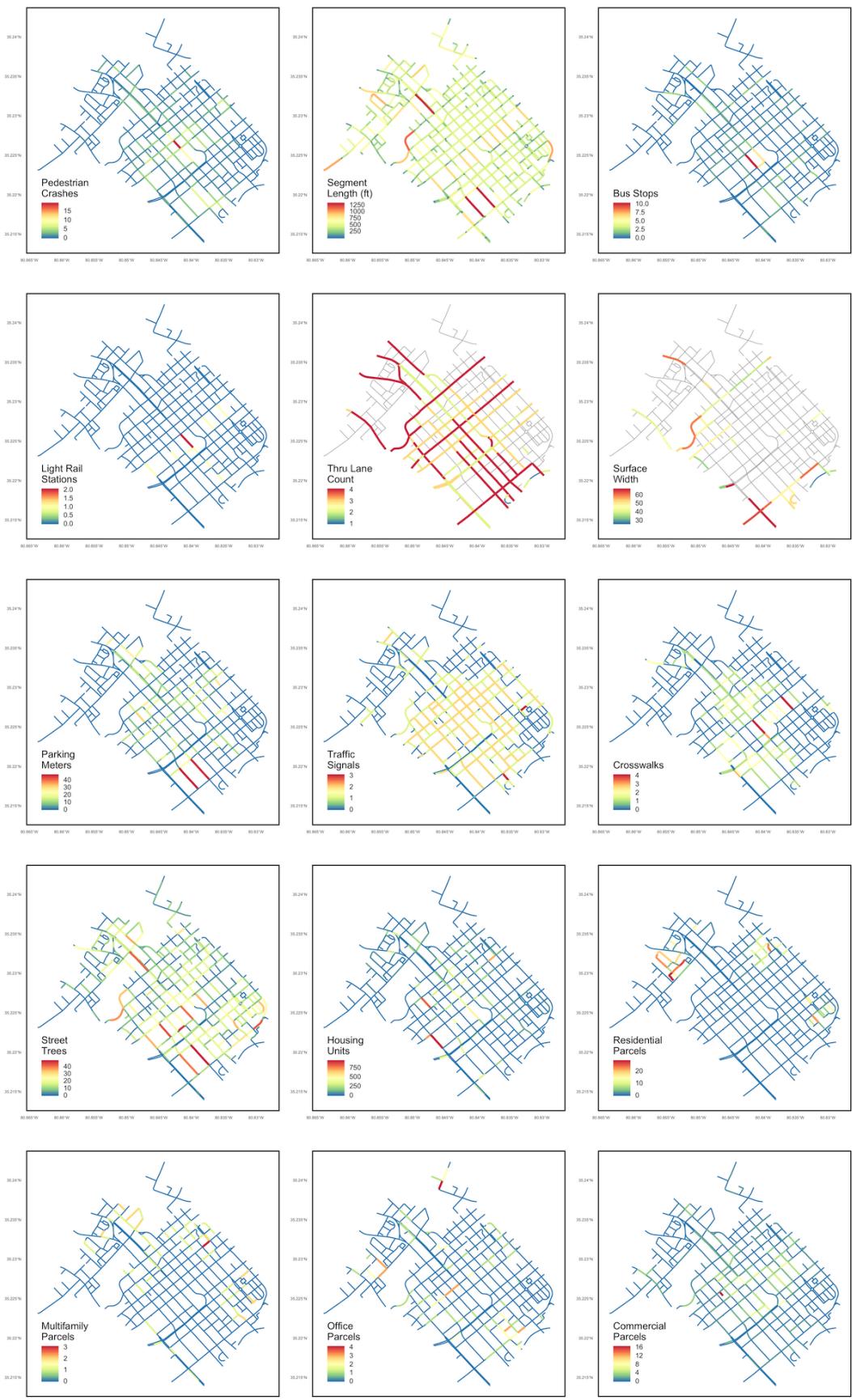


Figure 3.3: Variable Maps. If the variable was not an existing attribute of the road network data, road segments were assigned values using various GIS overlay techniques.

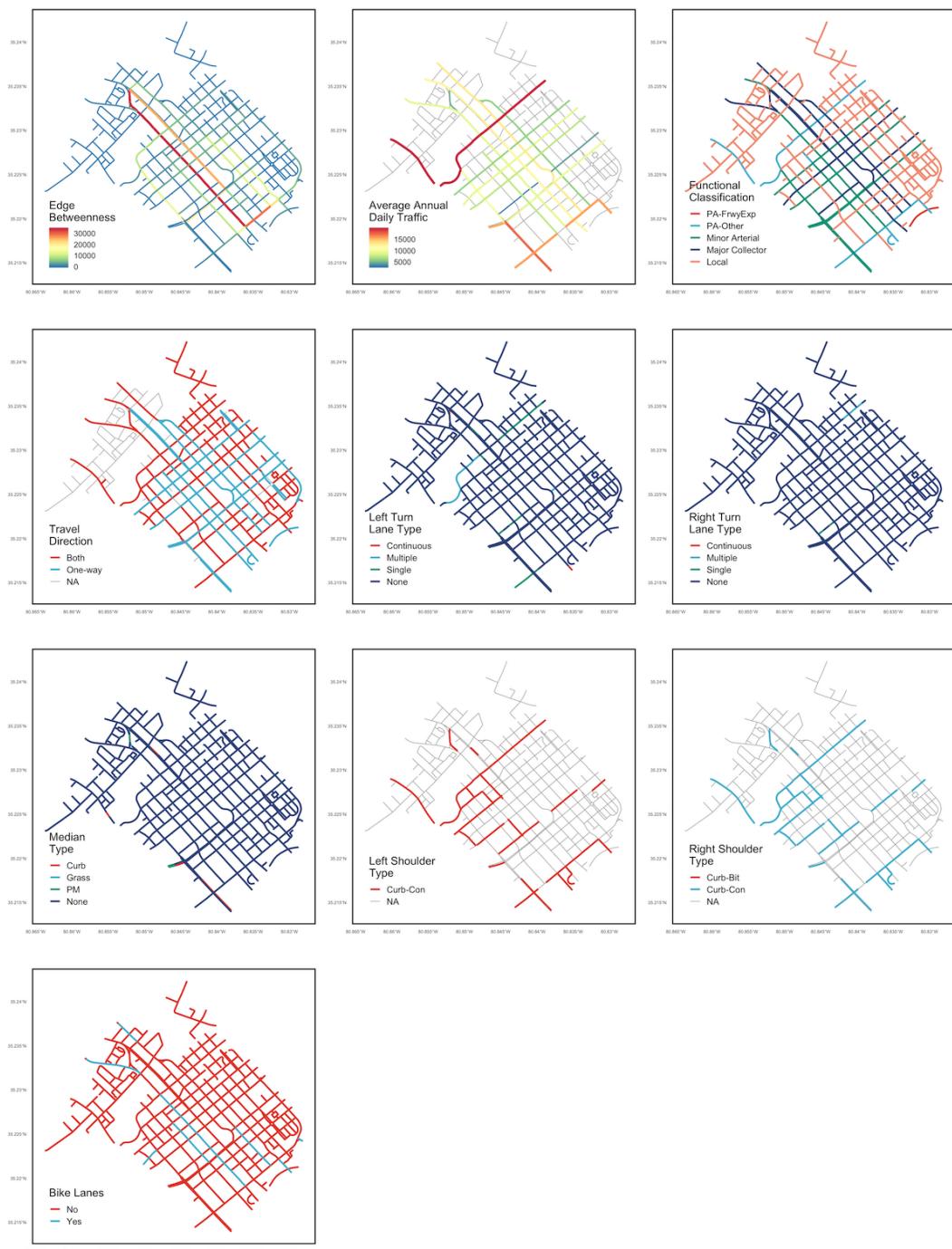


Figure 3.4: Variable Maps (continued). Continuous, discrete, and categorical data was included in the analysis.

Table 3.2: Continuous and Discrete Variables. Values rounded to 3 significant figures.

Variable	Description	Obs.	Max	Min	Mean	S.D.
crashes	pedestrian crashes	683	19	0	0.616	1.46
length_ft	length of segment in ft	683	1290	0.00161	296	217
busstops	bus stops	683	10	0	0.174	0.641
lightrail	light rail stations	683	2	0	0.0146	0.132
thrulanes	number of thru lanes	408	4	1	3.12	0.846
surfacewidth	surface width of pavement in ft	135	66	26	47.3	10.4
parkingmeters	parking meters	683	46	0	1.83	5.31
trafficsignals	traffic signals at intersections	683	3	0	0.548	0.716
crosswalks	crosswalks	683	5	0	0.261	0.617
trees	street trees	683	142	0	7.57	12.6
housingunits	housing units	683	918	0	26.6	89.7
residential	residential tax parcels	683	408	0	6.22	31.2
multifamily	multifamily tax parcels	683	7	0	0.184	0.705
office	office tax parcels	683	46	0	0.423	3.11
commercial	commercial tax parcels	683	19	0	0.635	1.63
traffvols	average annual daily traffic volumes	350	19700	2890	10600	4740
edgebetw	network edge betweenness	683	34200	0	4330	8910

3.2.1 Ethical Evaluation

As this study used data that was available from various online and easily accessible sources like Open Mapping and the Charlotte Open Data Portal, a secondary objective of this research became to test the limits of the data available on such platforms and to identify areas where Charlotte should focus attention for further collection and sharing. Open data platforms are considered an integral aspect of the ‘smart city’ (Barns 2018), although it should be noted that there is some debate regarding the underlying intentions behind the smart city concept (Allam & Newman 2018).

Table 3.3: Categorical Variables. Values rounded to 3 significant figures.

Variable	Description	Obs.	Value	Count	Perc.
traveldirection	travel direction	568	both	325	57.2
			one-way	243	42.8
funcclass	functional classification	683	PA-FrwyExp	10	1.46
			PA-Other	75	11
			Minor Art.	167	24.5
			Major Coll.	143	20.9
			Local	288	42.2
lfttrnlntyp	presence and type of left turn lane	683	continuous	6	0.878
			multiple	11	1.61
			single	20	2.93
			none	646	94.6
rttrnlntyp	presence and type of right turn lane	683	continuous	3	0.439
			multiple	3	0.439
			single	5	0.732
			none	672	98.4
median	presence and type of median	683	curb	11	1.61
			grass	2	0.293
			PM	10	1.46
			none	660	96.6
leftshoulder	left shoulder surface type	167	bitum	0	0
			concrete	167	100
rightshoulder	right shoulder surface type	167	bitum	5	2.99
			concrete	162	97
bikelanes	presence of bike lane	683	yes	69	10.1
			no	614	89.9

A full ethical evaluation of Charlotte's smart city vision (City of Charlotte 2016) is beyond the scope of this research, but some of broader implications of using open data will be briefly discussed here.

There are many benefits to open data, as explained by Janssen et al. (2012, p. 260): 'users can validate and verify whether the conclusions drawn from the data are correct and justified, and they can analyze the previously collected data to sharpen the focus of policy-making'. This explanation is in line with government policy from both Mecklenburg County and Charlotte (Mecklenburg County 2020a, City of Charlotte 2015). Therefore, it is safe to say that the use of open data to achieve the objective of this paper, ultimately to improve upon pedestrian safety policies, is

in congruence with the expectations of how open data will be used. Further, using open data for research saves time and money, which is hugely helpful when these resources are not abundant.

However, it is important to consider the larger-scale ethical implications of the sharing of and use of open data from such government platforms. Janssen et al. (2012) addressed five myths that surround the idea of open data. Three of their primary concerns are that 1) open data will further the ‘digital divide’ as only certain communities may have the knowledge, skills, and time for analysis and interpretation, 2) according to institutional theory, the data shared on open data platforms will not be information that has the potential to uproot ingrained, problematic systems, and 3) there is little standardization in terms of what data is shared, data quality, and what metadata is provided (Janssen et al. 2012). This research was conducted with the aforementioned ethical considerations in mind, and was not subject to UCL ethical approval.

3.3 Methodology Overview

As previously mentioned, this research attempted to follow and expand upon the methodology proposed by Kumfer et al. (2019), the goal of which was to establish a risk-based, systemic pedestrian safety analysis approach using pedestrian crash data and other environmental features at the street level in Seattle, Washington. Figure 3.5 shows a flowchart of the overall, modified process used here.

There are several important differences in the application of this methodology to Charlotte, which lend insight into its widespread applicability. First, this study attempts to replicate the methodology in a fundamentally dissimilar context. Charlotte has less global renown than Seattle and is not widely considered walkable or ‘smart’ (Berrone & Ricart 2019). As a result, adjustments had to be made in regards to what variables were included. The authors write: ‘Pedestrian activity data or demand is...critical for prioritizing locations for treatment’ (Kumfer et al. 2019, p. 422). For the reasons previously discussed (see section 2.3), thorough pedestrian volume data does not exist in many places, including Charlotte, so it was necessary

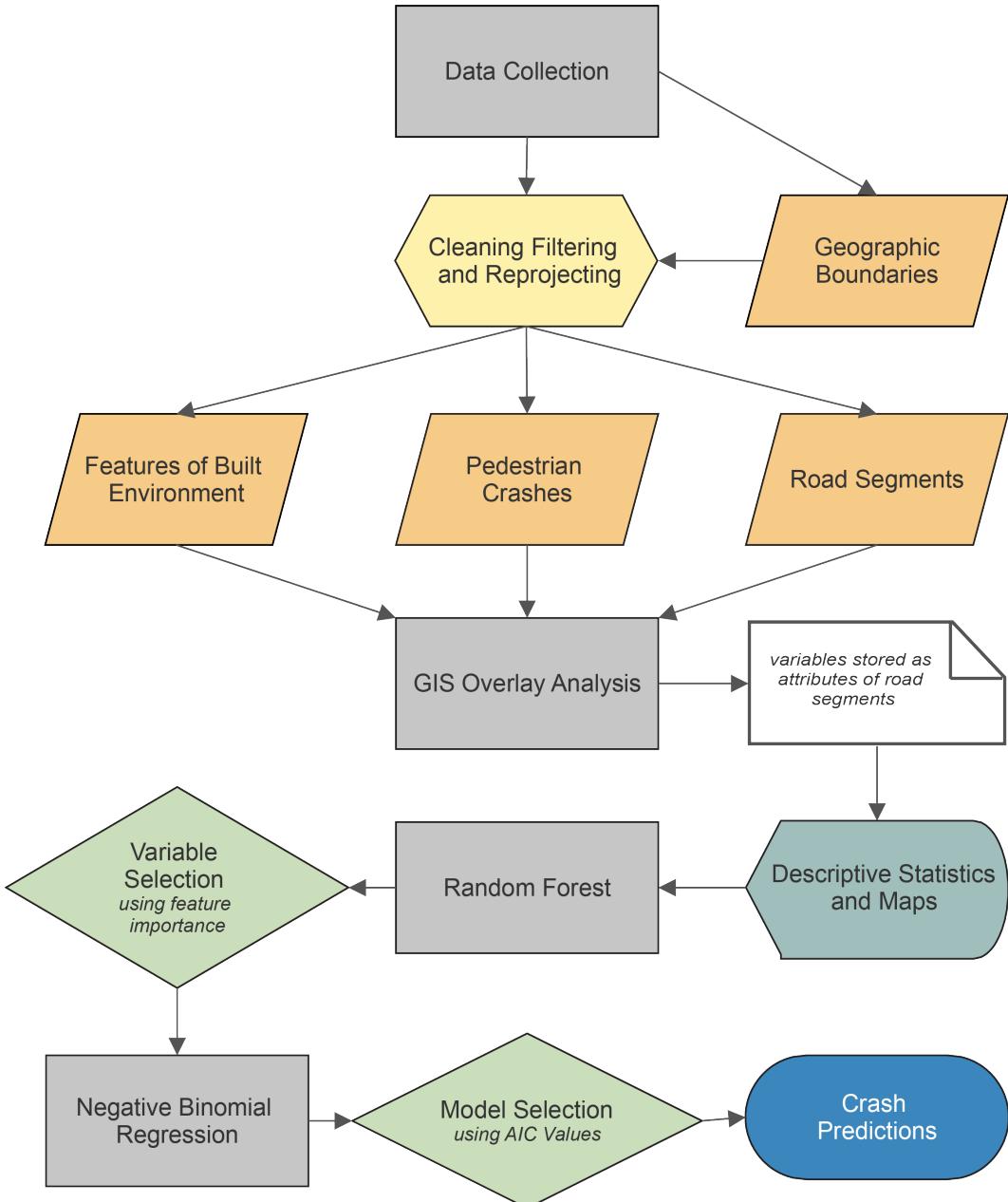


Figure 3.5: Methodology Flowchart. The crash predictions and the final negative binomial regression model were used to explore dangerous roads and possible solutions.

to explore alternative options. The number of housing units was included as a possible proxy measure here, with highly debatable substitutive ability. In this vein, this study only considers features of the built environment and excludes economic characteristics in order to focus the results on physical, urban design changes that could be carried out to improve pedestrian safety. Finally, the original study does not delve into validation of the methods used. Various validation techniques are uti-

lized and addressed in this study to better understand the legitimacy of the results. Each step of the process is expanded upon below:

3.3.1 Random Forest

Random forest regression was used to model the number of crashes along each road segment using the features of the built environment as predictors, in order to ultimately determine variable importance. Random forest is a ensemble machine learning method that greatly improves upon the accuracy of a single decision tree (Breiman 2001), and offers the following advantages: managing of missing data, handling of outliers, and avoidance of overfitting (Gupta 2020). However, a significant drawback is interpretability (Gupta 2020), which is why the next step implements negative binomial regression.

For validation purposes, the data was divided into training (90%) and testing (10%) sets, as in the example by Koehrsen (2018). Once the model was generated using the training data and predictions for the testing data set were determined, the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) were reported. Afterwards, the variable importance was found using the mean decrease in accuracy (MDA). Various measures of variable importance have been suggested when it comes to interpreting the predictors used in random forest. Due to the mixed data types, the varying scales of the predictors, and the issue of missing data, it was important to find a measure that could account for all of these obstacles without introducing bias. MDA fulfills these requirements, although it performs poorly when predictors are highly correlated (Strobl et al. 2009). This was not a problem in this instance (see correlation matrix in figure 3.6).

If the variable had a feature importance of greater than zero, meaning it was advantageous to the random forest process, it was considered for inclusion in the negative binomial regression, explained further in the next section.

3.3.2 Negative Binomial Regression

Next, negative binomial regression models were generated for all possible combinations of the important variables, and the Akaike Information Criterion (AIC) was

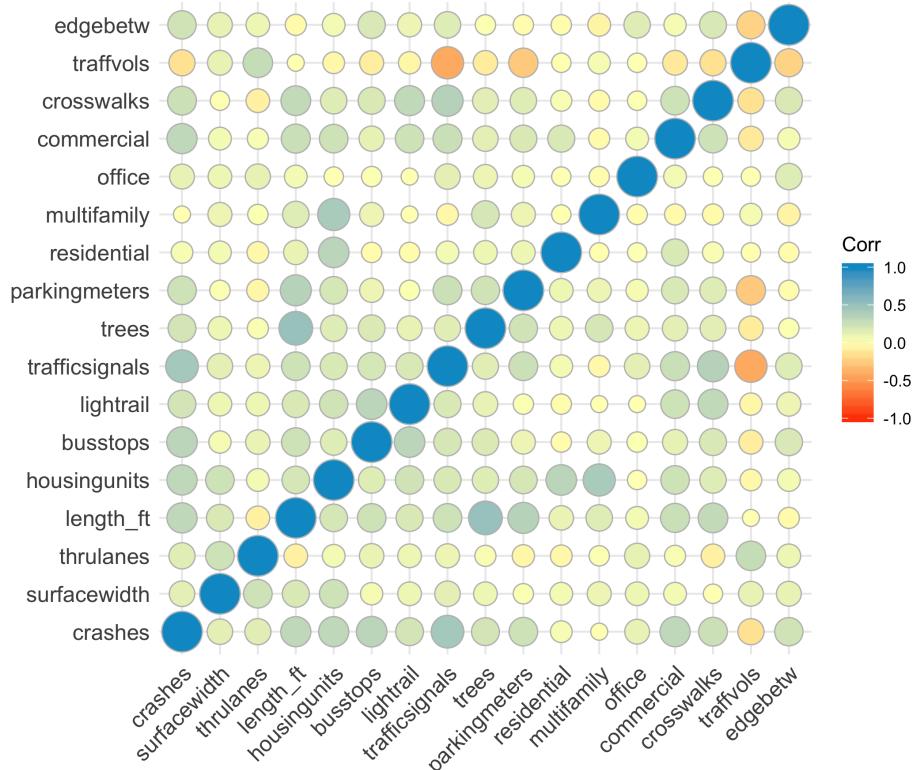


Figure 3.6: Correlation Matrix for Continuous and Discrete Variables. Uses imputed values where road segments had missing data. While there are some correlations, none are worthy of concern in terms of multicollinearity.

used as a relative indicator of model performance to select the best option. This step was completed to allow for clearer interpretation of the relationships between pedestrian crashes and the various features included in the chosen negative binomial regression model.

Negative binomial regression was used as opposed to Poisson regression because it is more suitable for over-dispersed count data, or when the variance of the outcome variable is not equal to the mean (Ford 2016); the pedestrian crash variable had a mean of 0.616 and a variance of 2.14. Assumptions of the negative binomial regression model include: ‘linearity in model parameters, independence of individual observations, and the multiplicative effects of independent variables’ (Yang & Berdine 2015). Further, AIC was chosen to compare the models because it does not show preference for simpler models like the Bayesian Information Criterion (BIC) does (Brownlee 2019). Negative binomial regression is not capable of handling

missing values, so for each combination of variables, observations with missing values were dropped.

3.3.3 Prediction and Interpretation

The number of crashes were predicted per road segment using the selected negative binomial regression model and the predictions were compared to the true values, again using MAE, MSE, and RMSE. Where an observation was missing data, the required values were imputed using the `rfImpute()` function (Breiman 2003). The predicted crashes were mapped, with emphasis on the top 1% and top 25% most dangerous roads. Kumfer et al. (2019) used Empirical Bayes estimates to weight the predictions with historical crash data, but this step was omitted here. The authors write that the non-weighted predictions offer ‘a more holistic measure of pedestrian crash risk, given the uncertainty about future crash locations inherent in any crash prediction method’ (Kumfer et al. 2019, p. 427). The top 1% were further examined using Google Street View and possible pedestrian safety interventions were discussed. To contextualize these findings within the focus of existing traffic safety programs in Charlotte, roads with crashes in the top 1% were mapped against the High Injury Network developed as part of Charlotte’s involvement with *Vision Zero* (City of Charlotte DOT 2019). Finally, the results from the analysis were discussed in the context of the policies suggested in various walkability and pedestrian safety documents produced by the City of Charlotte.

Chapter 4

Results

The methodology described in Chapter 3 was carried out, and the results are described here. The random forest, made up of 500 decision trees, was constructed using the training dataset to predict pedestrian crashes and included the use of all 24 independent variables. Although the examination of individual trees was a superfluous step, a sample decision tree from the random forest can be seen in figure 4.1. Performance metrics comparing the predicted values to the actual values for the testing dataset can be seen in table 4.1. The MAE indicates that, on average, the estimates varied from the actual crash values by less than a single crash. However, the higher RMSE score reveals the presence of some estimates with larger errors and that the model may have questionable predictive capability, as the dependent variable ranges from only 0-19 pedestrian crashes (JJ 2016).

Table 4.1: Random Forest Performance Metrics. Rounded to 3 significant figures.

MAE	MSE	RMSE
0.738	1.62	1.27

For each of the 24 variables, the corresponding mean decrease in accuracy (MDA) was determined as a metric of variable importance (see figure 4.2 and table B.1). 15 variables had a MDA of greater than 0, reflecting that they were advantageous predictors when used in the random forest regression model: traveldirection, funcclass, thrulanes, length_ft, housingunits, busstops, lightrail, trafficsignals, trees, parkingmeters, residential, multifamily, office, commercial, and edgebetw.

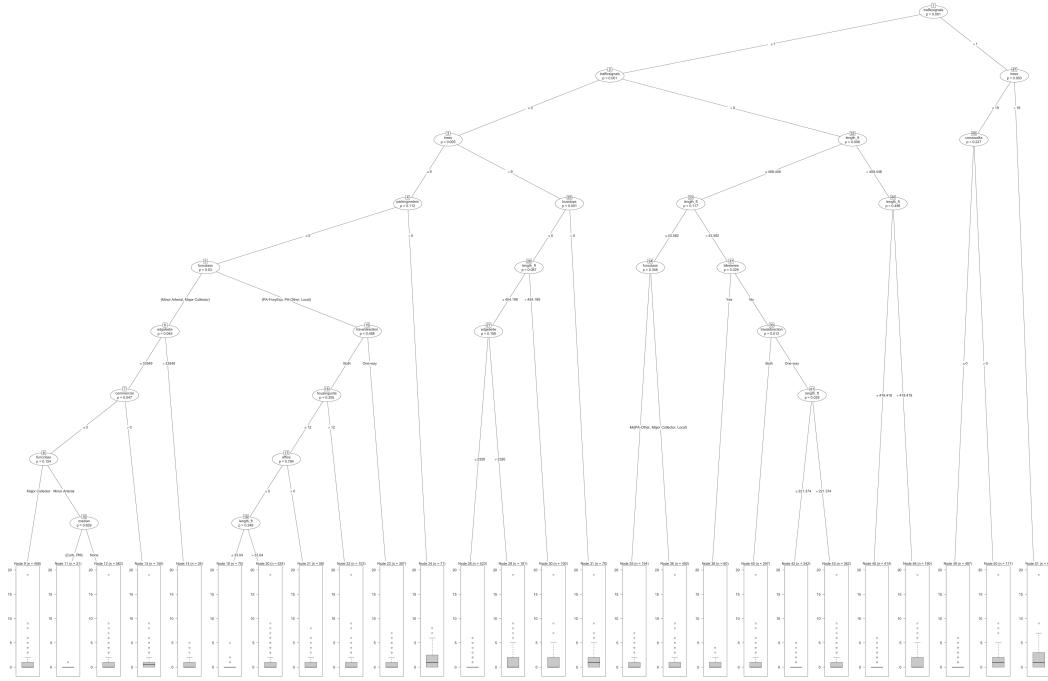


Figure 4.1: Sample Decision Tree. One of 500 decision trees comprising the random forest.

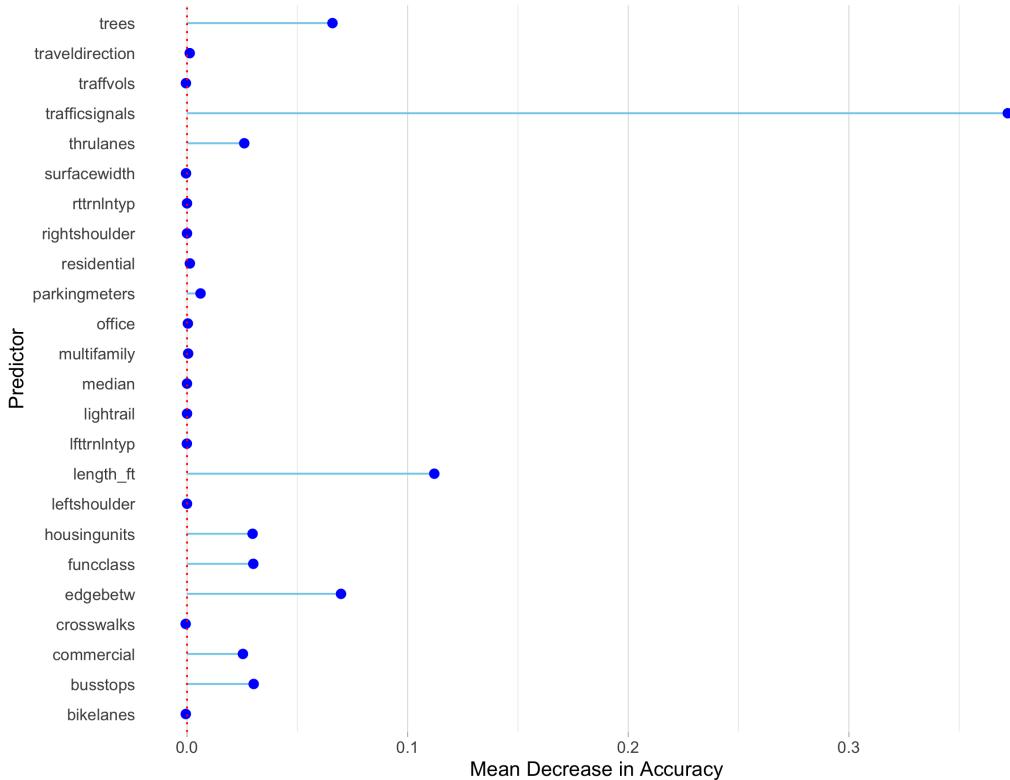


Figure 4.2: Variable Importance. 15 predictors had a MDA greater than 0, and were considered for inclusion in the negative binomial regression model. See table B.1 for values.

Table 4.2: Negative Binomial Regression Model. Values rounded to 3 significant figures.

Coefficients	Variable	Estimate	Std. Error	z value	p-value
	Intercept	-2.71	0.501	-5.42	5.92e-8
	traveldirect (one-way)	-0.0791	0.212	-0.372	0.710
	thrulanes	0.207	0.128	1.62	0.105
	length_ft	0.00256	0.000384	6.66	2.67e-11
	housingunits	0.00126	0.000658	1.91	0.0560
	lightrail	-0.690	0.408	-1.69	0.0905
	trafficsignals	0.665	0.116	5.75	8.71e-9
	commercial	0.176	0.0551	3.19	0.00142
	edgebetw	0.0000206	0.00000731	2.82	0.00482
Deviance Residuals	Min	1Q	Median	3Q	Max
	-1.96	-0.799	-0.589	0.0379	3.12

These 15 variables were considered for inclusion in the negative binomial regression model.

A negative binomial regression model was fit for all possible combinations of the variables identified, with a minimum number of 2 independent variables and a maximum of all 15. 32,752 models in total were generated and their performances were compared using the Akaike Information Criterion (AIC) values. The winning model (see table 4.2) had the lowest AIC value of approximately 833, and included 8 independent variables: traveldirection (one-way), thrulanes, length_ft, housingunits, lightrail, trafficsignals, commercial, and edgebetw. The p-values show that all variables, aside from traveldirection (one-way) and thrulanes, were statistically significant at a 95% confidence level.

Next, predictions of the number of pedestrian crashes for all road segments were generated using the model, and the resulting performance metrics can be seen in table 4.3. The negative binomial regression performed only slightly worse than the random forest, but offered the additional benefit of improved interpretability via the coefficients and p-values associated with a parametric approach. For example, the negative binomial regression model indicates that while the other variables are

Table 4.3: Negative Binomial Regression Performance Metrics. Values rounded to 3 significant figures.

MAE	MSE	RMSE
0.775	2.28	1.51

held constant, every unit increase in the number of commercial tax parcels coincides with a log 0.176 increase in the number of pedestrian crashes. This is because ‘the model models the log of the expected count as a function of the predictor variables’ (*Negative Binomial Regression: Stata Annotated Output* 2006).

Figures 4.3 and 4.4 display the actual crash data and the predicted number of crashes spatially. These visualizations show that the model predictions have captured roads that have been dangerous in the past, while also highlighting roads that should potentially be the focus of future concern. To further emphasize the unsafe areas highlighted by the model, the top 1% and top 25% most dangerous roads based on the number of predicted crashes were located and mapped (see figure 4.5). The actual and predicted crash statistics for the 6 roads in the top 1% most dangerous can be seen in table 4.4.

Each of these 6 roads was further examined using Google Street View in order to manually discern what the problems for pedestrian safety might be (see figures 4.6 and 4.7), and to compare these findings to the significant variables included in the model. The limitations of the model will be addressed in the following section, along with the methodology’s restricted ability to identify specific, targeted urban design changes. Essentially, without controlling for pedestrian volumes, some urban design elements that should theoretically reduce pedestrian-involved incidents appear to cause them to increase.

Nonetheless, based on the included variables and the manual inspection process, a possible solution to reduce pedestrian collisions along segments A, B, and C as seen in figure 4.5 could be to change these routes to one-direction, single-lane roads in attempt to reduce conflict between pedestrians and cars. The theoretical results of these changes were predicted using the model and are visible in table 4.4, resulting in up to a 50.4% decrease in pedestrian crashes. Finally, roads with crashes

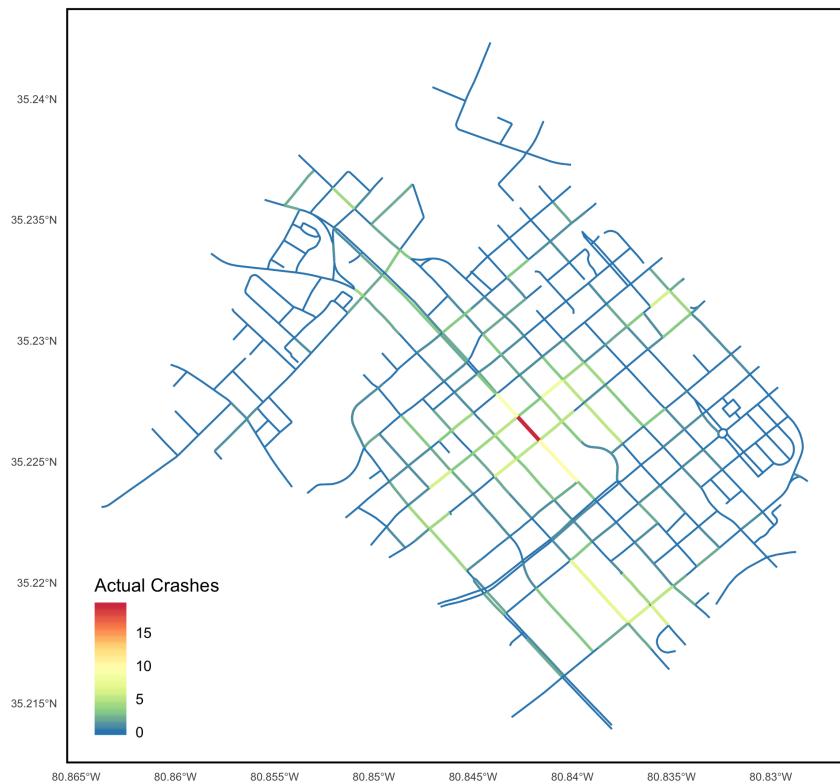


Figure 4.3: Actual Pedestrian Crashes. Original data.

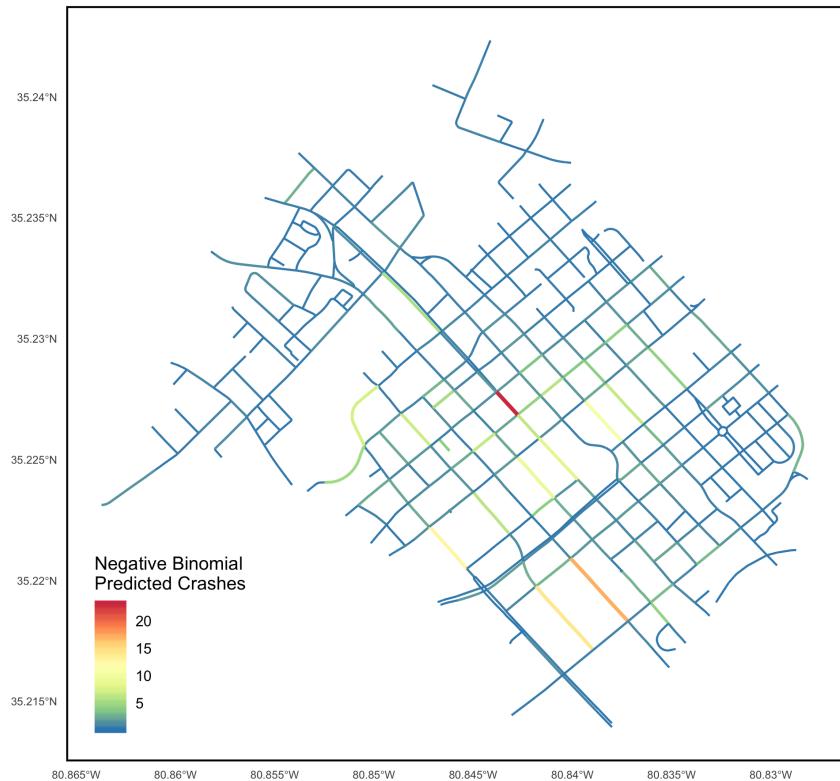


Figure 4.4: Predicted Pedestrian Crashes. Results of negative binomial regression model.

in the top 1% were mapped against the High Injury Network developed as part of Charlotte's involvement with *Vision Zero* (City of Charlotte DOT 2019). The model generated through this research identified two roads (D and E) that are not part of the High Injury Network. Potential reasons for this difference will be discussed in the next section, along with other interpretations and explanations of the results.

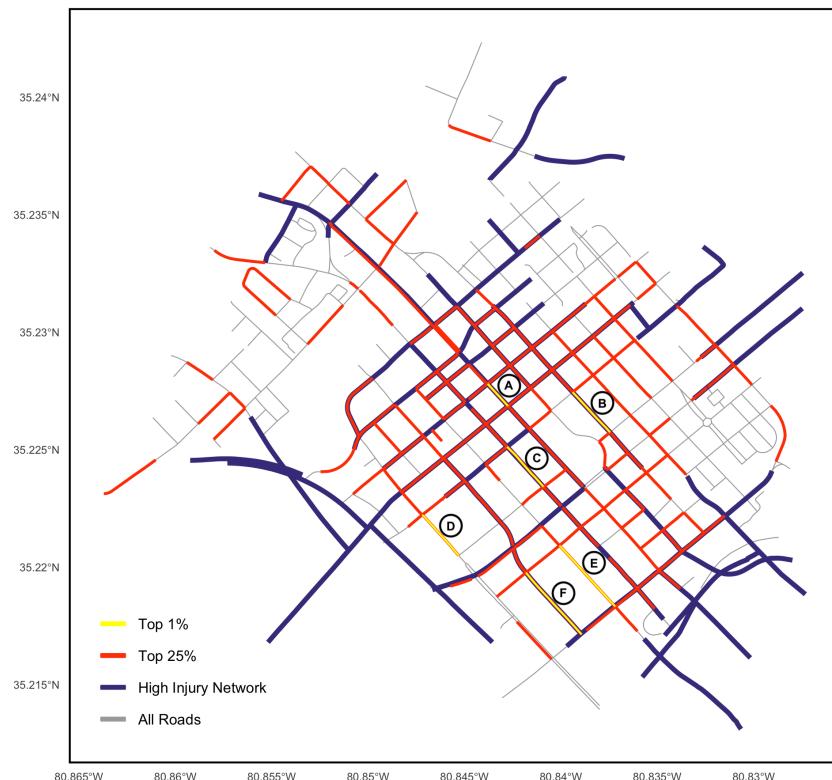


Figure 4.5: Most Dangerous Roads. Displays the top 1% and 25% most dangerous roads according to the model, and the *Vision Zero* High Injury Network (CharlotteNC 2020). Labels relate to table 4.4 and figures 4.6 and 4.7.

Table 4.4: Top 1% Most Dangerous Roads.

Map Label	Name	Crashes	Predicted Crashes	After Adjustments	Perc. Change
A	W Trade St	9	23	11	-50.4%
B	E 6th St	5	10	7	-33.9%
C	E 4th St	3	9	9	-46.3%
D	E Stonewall St	4	13	-	-
E	E 3rd St	7	17	-	-
F	E Martin Luther King Jr Bv	3	14	-	-

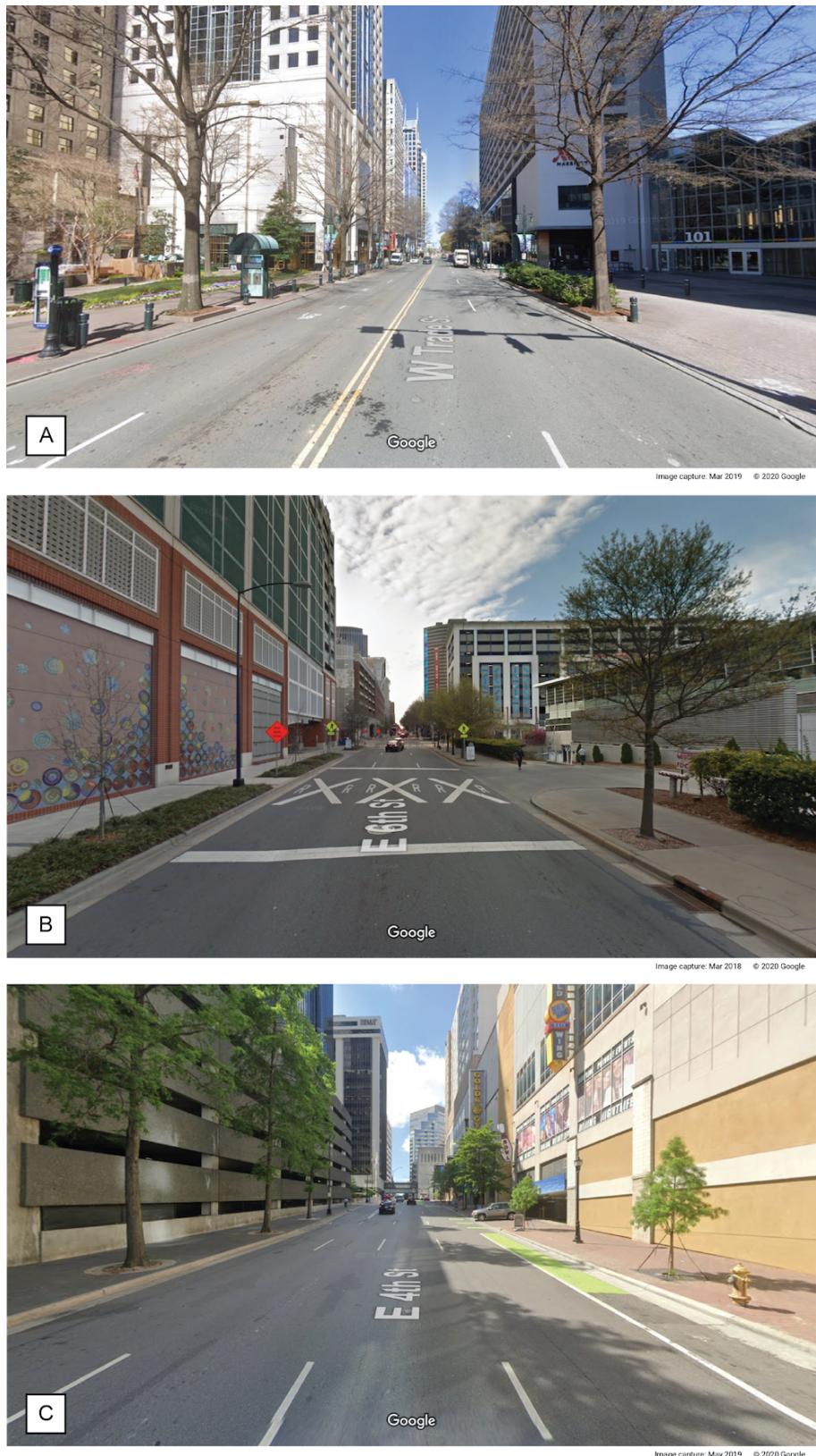


Figure 4.6: Google Street View Images A-C (Google 2019a, 2018, 2019c).

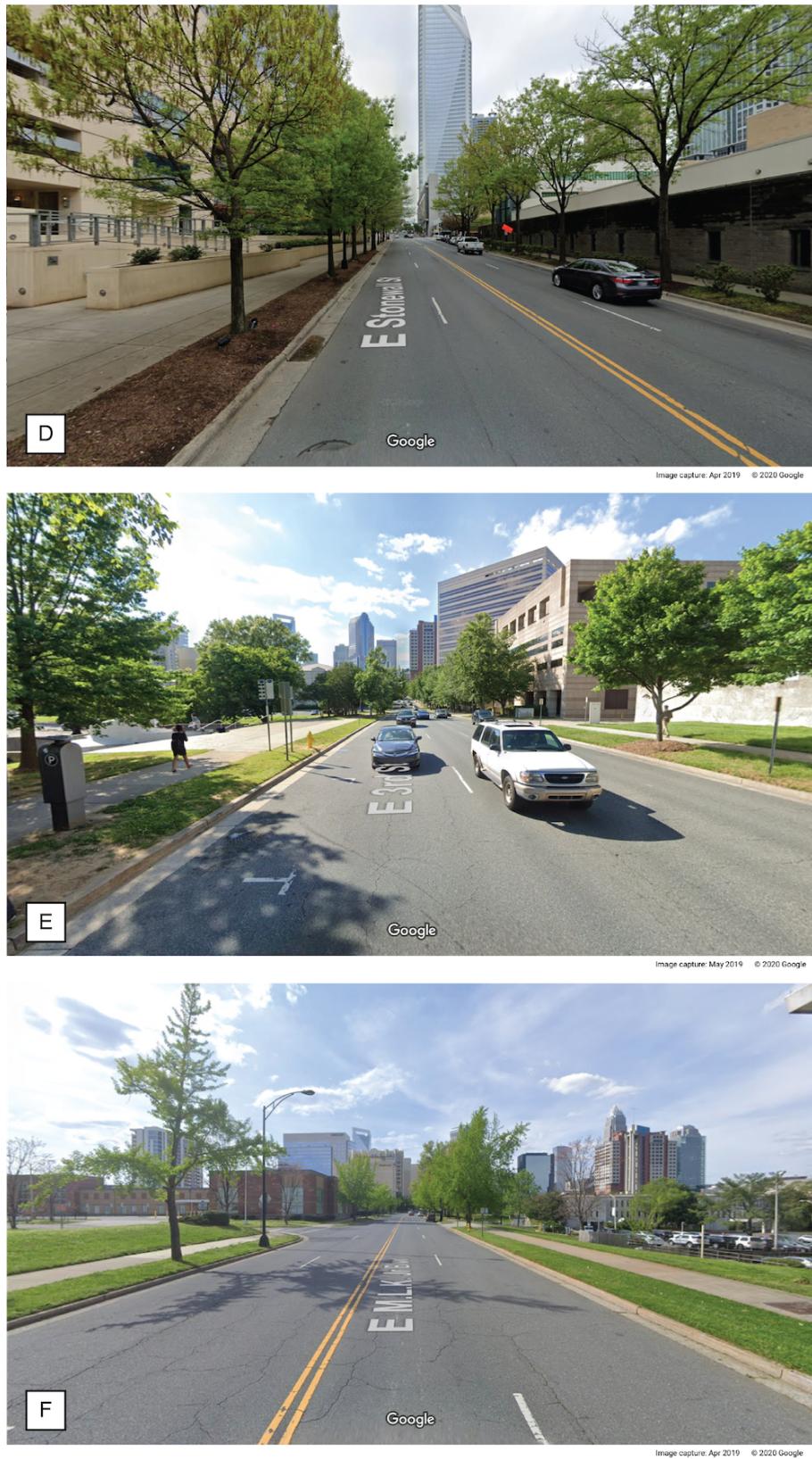


Figure 4.7: Google Street View Images D-F (Google 2019b,e,d).

Chapter 5

Discussion

This chapter explains and expands on the results of the analysis presented in Chapter 4. The study has attempted to answer if features of the built environment can be used to predict the locations of pedestrian collisions at the street level in Uptown Charlotte, North Carolina, and subsequently if this process can allow for the recommendation of targeted urban design changes that would make the streets safer for pedestrians. After following the methodology proposed by Kumfer et al. (2019), the results show that features of the built environment can be used to predict the locations of pedestrian crashes with a reasonable amount of accuracy. However, applying these findings to determine specific urban design adjustments is difficult without the inclusion of pedestrian volume data, even when potential proxy variables are utilized.

5.1 Interpretation of Model

In comparing figures 4.3 and 4.4, it can be seen that the negative binomial regression model was able to capture the general areas of historic pedestrian crashes within a comparable range of values to the original data. However, the model will never, and should never, achieve 100% accuracy. Some crashes will happen for reasons that have little to do with the physical environment, and in these instances, this model will not offer much explanatory power. Additionally, as pointed out by Kumfer et al. (2019), areas that have not had crashes in the past are not necessarily immune from experiencing them in the future. Batty writes, ‘Models are simplifications of reality

– theoretical abstractions that represent systems in such a way that essential features crucial to the theory and its application are identified and highlighted’ (Batty 2009, p. 52). In this sense, the purpose of the model is not to replicate past crashes with perfect accuracy, but rather to highlight streets that are at risk for crashes based on the presence of certain features or characteristics. The model succeeds in doing so.

5.2 Assessment of Variables

Along these lines, the first step of the analysis used mean decrease in accuracy values to disqualify some of the built environment variables. Surprisingly, the traffic volumes variable was eliminated at this stage due to its unfavorable impact on the model’s accuracy. Past studies have shown that higher vehicular volumes are generally associated with higher numbers of pedestrian crashes (Moradi et al. 2017, Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2018), but in Charlotte, some roads with high vehicular traffic volumes may also be roads where there are rarely any pedestrians. The presence of left turn lanes, medians, bike lanes, number of crosswalks, and surface width were also deemed to be unimportant. These variables were initially included for their theoretical usefulness as per the literature review (Chapter 2), but their elimination highlights the importance of 1) using accurate and complete data and 2) localized analysis, as some factors may be more or less important according to local conditions (Bartolomeos et al. 2013, p. 43).

Once the negative binomial regression model was generated, the coefficients and p-values offered insight as to which of the 8 included features best explained the variation in the number of pedestrian crashes. It was unexpected to see that the number of traffic signals at intersections was the best indicator of the number of pedestrian collisions. The positive coefficient could either mean that signalized intersections are located in areas with high pedestrian traffic (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2018, p. 60), or that traffic signals correlate with other issues such as pedestrian visibility to drivers who are turning left (Bartolomeos et al. 2013, p. 33).

Of the other variables included in the negative binomial regression model, travel direction and number of thru lanes offered the most possibility for feasible urban design intervention, although these variables were also the only ones that were not significant at a 95% confidence level. Three of the included variables, housing units, commercial parcels, and edge betweenness, could have led to interesting findings and suggestions, but since these variables are potentially correlated with pedestrian volumes, it is complicated to interpret their role in the model.

5.3 The Need for Pedestrian Volume Data

Without the ability to directly control for pedestrian volumes, factors that would theoretically reduce the number of pedestrian crashes may appear to correlate with their increase. For example, blank *streetwalls* are considered to be a common walkability blunder, whereas dense, interesting, mixed-use streets are thought to be much more conducive to pedestrian activity (Speck 2013). Five of the top six most dangerous roads identified by the model have a noticeable lack of pedestrian-scale intrigue (see figures 4.6 and 4.7). A theoretical solution for improved walkability would be to increase the number of commercial units along these roads, but the model suggests that this would cause pedestrian crashes to rise. Some past research has suggested that more commercial space correlates with more pedestrian crashes (Moradi et al. 2017, Southworth 2005). Other findings indicate a non-linear relationship in which more pedestrians mean more crashes until a threshold is reached, at which point pedestrians become a dominate force in the transport ecosystem and the number of crashes levels off (Greene-Roesel et al. 2007). Despite the findings of the model, increasing commercial units and allowing pedestrians to claim space in the Charlotte streetscape could eventually reduce crashes by making drivers more expectant of a pedestrian presence.

As such, pedestrian volume data is definitely needed to make this analysis productive. The number of residential parcels, multifamily parcels and office parcels were excluded from final model, indicating that they are not good substitutes for pedestrian volumes. As mentioned, commercial parcels, edge betweenness, and

housing units were included and may be suitable as proxy variables, but this would need to be confirmed. Regardless, expanding the collection of pedestrian volume data in Charlotte should be a priority and would reflect a genuine respect for walking as an important form of transportation. There are several young pedestrian count programs in Charlotte, but the data is collected at only a few points (Institute for Transportation Research and Education 2020) or is not shared publicly (CRTPO 2020). The *Charlotte WALKS Pedestrian Plan* mentions a need for expanded pedestrian counting, stating that: ‘As a participant in [the statewide Non-Motorized Volume Data Program], Charlotte will receive and install 18 continuous pedestrian/bicycle counters in key locations throughout the city’ (City of Charlotte DOT 2017, p. 72). Charlotte is 800.94 km² and pedestrian volume modeling requires collection points to be much more dense to get reliable results (Raford & Ragland 2006). Pedestrian data collection and volume modeling, specifically in Uptown Charlotte, should certainly be the focus of future research.

5.4 Pedestrian Safety Policy Contextualization

As part of Charlotte’s work with *Vision Zero*, an online interactive feedback map was developed to allow citizens the opportunity to share concerns about traffic safety within the city (see figure 5.1). Griffin & Jiao find that the use of such digital technologies may introduce accessibility challenges and bias, but that ultimately, ‘Crowdsourcing tools can be valuable approaches to increase geography and equity of public participation in transportation planning’ (Griffin & Jiao 2019, p. 460). The Charlotte *Vision Zero* report explains: ‘The top five comments submitted related to traffic safety concerns were street design, speeding, lack of pedestrian facilities, failing to yield to pedestrians and drivers running stop signs and red lights’ (City of Charlotte DOT 2019, p. 22). The analysis conducted here has the potential to quantify the impact of changes to the built environment in terms of the number of pedestrian collisions, which could be a powerful tool for advocating for pedestrian needs in the city when combined with existing efforts such as the interactive map.

The *Charlotte WALKS Pedestrian Plan* mentions sidewalks, street lighting,

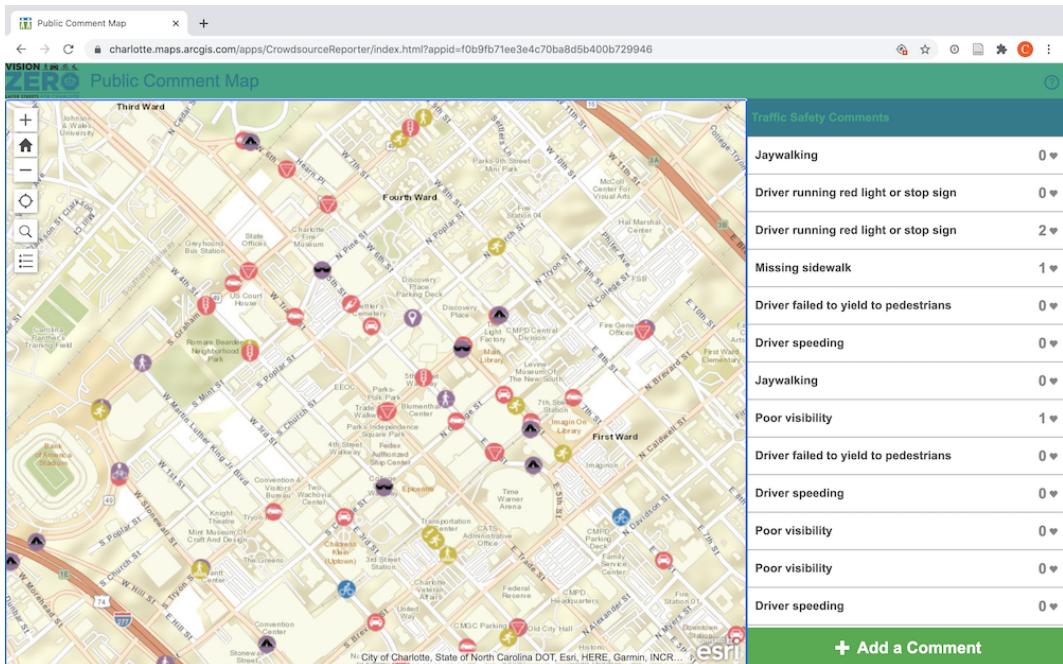


Figure 5.1: Vision Zero Interactive Feedback Map. Screenshot from website (City of Charlotte 2020).

planting strips, and crossings as the urban design elements that are the most crucial for pedestrian safety or perceptions of safety (City of Charlotte DOT 2017). Based on the variables included in the negative binomial regression model, this study did not find any of these features to be particularly important in Uptown Charlotte, confirming that a different approach to improved pedestrian safety needs to be taken in this area. One of the directed techniques suggested in the *Charlotte WALKS Pedestrian Plan* is: ‘...In places like Uptown and South End, the city should consider assigning specific modal priority to pedestrians in all public and private projects’ (City of Charlotte DOT 2017, p. 38). Indeed, Charlotte’s *Urban Street Design Guidelines* suggest that pedestrian needs and vehicular traffic priorities are often at odds with each other, and that it is difficult to cultivate an urban environment that makes both types of users happy (City of Charlotte Staff 2007, pp. 42-47). This research attempted to quantify the benefits that could occur by reducing certain, high-risk roads to one lane and one direction. Based on the predictions from the model, such actions could theoretically reduce the number of crashes by up to 50.4%. This is in line with findings from prior research, which suggest that so-called

‘road diets’ can lead to a 19-47% reduction in crashes and that they are a relatively easy and cost-effective solution (U.S. DOT FHA 2016). Including pedestrian volume data could help to confirm if this is a good idea by checking if these roads are high pedestrian traffic areas where walking should be given ‘modal priority’.

Further, there are some differences between the High Injury Network developed through Charlotte’s work with *Vision Zero* (City of Charlotte DOT 2019) and the top 1% most dangerous roads identified in this study. The High Injury Network includes all types of traffic collisions with an emphasis on those that resulted in death or serious injury. By including all, and only, pedestrian collisions regardless of severity, this research may have indicated places that have not been the site of fatal crashes in the past but that are likely places for future pedestrian-involved incidents to occur. Thus, with development, this pedestrian safety analysis process has the potential to both provide support for existing policies and to suggest areas for adjustment where the data does not support the current polices that are in place.

5.5 Limitations

There were several limitations to this research in addition to the omission of pedestrian volume data. First, the study did not distinguish between mid-block and intersection crashes. Dividing these crashes into two categories may have allowed for more specific insights into certain design problems. Next, in correspondence with the methodology used by Kumfer et al. (2019), this study included length as a separate independent variable. An alternative approach would have been to standardize the other variables by road length, although length did not highly correlate with any of the other features. Finally, missing data meant that imputation was necessary to generate predictions for all the street segments using the final negative binomial regression model. While more time would have allowed for manual data collection via in-person site visits or via Google Street View, it would also have been quite helpful for the datasets offered via Charlotte’s open data platforms to be complete.

Chapter 6

Conclusion

Improving walkability is gaining traction as an essential strategy for achieving economic, public health and sustainability goals in cities worldwide. The newfound appreciation for walkability has coincided with concern about effectively facilitating safe walking journeys. While data-driven analysis methods have long been used to study vehicular traffic, such methods are only recently being applied to pedestrian problems. Cities such as Charlotte, North Carolina can take advantage of spatial, data-driven pedestrian analysis methods to improve the walkability and safety of their streets, and as a result become better prepared for future urban growth and more attractive to a new generation that values non-motorized travel.

This study follows the methodology proposed by Kumfer et al. (2019) to explore if features of the built environment can be used to predict the locations of pedestrian crashes at the street level in Uptown Charlotte, North Carolina. Using spatial data from open data platforms, the process applied GIS overlay analysis, random forest regression for the purpose of selecting important features, and negative binomial regression to create a legible model of pedestrian crashes. 24 features of the built environment were initially considered as independent variables, and the final model included 8 of these features: travel direction, thru lanes, length, housing units, light rail stations, traffic signals, commercial parcels, and edge betweenness. 6 roads were identified as being in the top 1% most dangerous, and were further examined using Google Street View.

While contending with a troublesome lack of pedestrian volume data, the re-

sults of the analysis were followed by a discussion of how targeted urban design adjustments can make walking more safe. Several key factors were highlighted for their design intervention potential, specifically the number of thru lanes, travel direction, and the number of commercial parcels along each road.

When contextualized within the existing city-wide pedestrian safety policy framework, this research showed that many of the focuses of policy documents are not applicable to this specific area. Thus, the study further confirmed the need for pedestrian safety policies that are specific to Uptown Charlotte. With the aid of digital inspection and manual visits to the sites, models like the one developed in this paper have the potential to indicate certain urban design elements that may be either problematic or helpful for pedestrian safety, as well as to model the impact of hypothetical alterations in terms of the number of predicted pedestrian crashes. Such adjustments and prioritization of streets should also be considered in relation to citizen feedback, as is already collected through the *Vision Zero* interactive map (City of Charlotte 2020). The research conducted here shows that this methodology could be applied to other cities that lack detailed pedestrian data and still yield helpful results.

This research can be seen as a preliminary approach to data-driven pedestrian safety analysis in Charlotte, North Carolina, demonstrating the potential of using spatial data to highlight dangerous roads and statistical relationships between pedestrian crashes and features of the built environment. Future research should focus on incorporating pedestrian volume data into the models; this paper advocates for an expanded pedestrian count program in Charlotte, which would signal a value for walking as an important mode of transportation in the city and would allow for further interpretive possibilities. ‘The loss of walking as an individual and a community act has the potential to destroy our deepest spiritual connections, our democratic society, our neighborhoods, and our freedom’ (Malchik 2019, p. 4). Charlotte has long been subject to auto-dominated planning and policy, with little consideration given to non-motorized transportation. Safe and accessible walkability will be key to Charlotte’s sustainable, equitable, and appealing development.

Bibliography

- Acuto, M. (2016), ‘Give cities a seat at the top table’, *Nature* **537**(7622), 611–613.
- Albers, P. N., Wright, C. & Olwoch, J. (2010), ‘Developing a South African pedestrian environment assessment tool: Tshwane case study’, *South African Journal of Science* **106**, 1–8.
- Allam, Z. & Newman, P. (2018), ‘Redefining the smart city: Culture, metabolism and governance’, *Smart Cities* **1**(1), 4–25.
- Babinard, J., Carrigan, A., Kost, C., Eyken, C. V., Mason, J., Nadal, L., Jacobsen, M. & King, M. (2018), *Pedestrian Mobility for Urban Growth: Walking and Its Links to Transportation*, The International Bank for Reconstruction and Development, Washington, D.C.
- Barns, S. (2018), ‘Smart cities and urban data platforms: Designing interfaces for smart governance’, *City, Culture and Society* **12**, 5–12.
- Bartolomeos, K., Croft, P., Job, S., Khayesi, M., Kobusingye, O., Peden, M., Schwebel, D., Sleet, D., Tiwari, G., Turner, B. & van Waeg, G. (2013), *Pedestrian Safety: A Road Safety Manual for Decision-Makers and Practitioners*, World Health Organization, Geneva, Switzerland.
- Batty, M. (2009), Urban modeling, in N. Thrift & R. Kitchin, eds, ‘International Encyclopedia of Human Geography’, Elsevier, pp. 51–58.
- Bedoya Arguelles, G., Milusheva, S. P., Legovini, A. & Williams, S. E. (2019), *Smart and Safe Kenya Transport (SMARTTRANS)*, Development Impact Evaluation (DIME) Brief, World Bank Group, Washington, D.C. Available at:

- <http://documents.worldbank.org/curated/en/723411574361015073/Smart-and-Safe-Kenya-Transport-SMARTTRANS>.
- Benz, R. J., Turner, S., & Qu, T. (2013), ‘Pedestrian and bicyclist counts and demand estimation study’, *Texas A&M Transportation Institute*. Available at: <https://static.tti.tamu.edu/tti.tamu.edu/documents/TI-2013-3.pdf>.
- Berrone, P. & Ricart, J. E. (2019), *IESE Cities in Motion Index*, IESE Business School. Available at: <https://media.iese.edu/research/pdfs/ST-0509-E.pdf>.
- Bliss, L., Lin, J. C. & Patino, M. (2020), ‘Pandemic travel patterns hint at our urban future’. Bloomberg CityLab, available at: <https://www.bloomberg.com/graphics/2020-coronavirus-transportation-data-cities-traffic-mobility/> [Accessed 28 Jul. 2020].
- Breiman, L. (2001), ‘Random forests’, *Machine Learning* **45**, 5–32.
- Breiman, L. (2003), ‘Manual for setting up, using, and understanding random forest v4.0.’. Available at: https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf [Accessed 21 Aug. 2020].
- Brownlee, J. (2019), ‘Probabilistic model selection with AIC, BIC, and MDL’. Machine Learning Mastery, available at: <https://machinelearningmastery.com/probabilistic-model-selection-measures/> [Accessed 3 Aug. 2020].
- Bunnarong, S. & Upala, P. (2018), ‘Spatial analysis to identify pedestrian crash zones: A case study of school zones in Thailand’, *The Open Transportation Journal* **12**, 167–181.
- Cesme, B., Dock, S., Westrom, R. J., Lee, K. & Barrios, J. A. (2017), ‘Data-driven urban performance measures: Case study application in the District of Columbia’, *Transportation Research Record* **2605**, 45–53.

- Char-Meck Planning Dept. (2019), ‘Downtown wards’. Electronic dataset, Charlotte Open Data Portal, available at: https://data.charlottenc.gov/datasets/64dde23672004971a1b2dbf3a12f733e_13 [Accessed 21 Jul. 2020].
- Charlotte Area Transit (2019), ‘Bus stops’. Electronic dataset, Charlotte Open Data Portal, available at: https://data.charlottenc.gov/datasets/e0643b893ea24703830c3e41e231db69_0 [Accessed 21 Jul. 2020].
- Charlotte DOT (2018), ‘Sidewalks’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/datasets/sidewalks> [Accessed 2 Aug. 2020].
- Charlotte DOT (2019a), ‘Bike lanes’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/datasets/bike-lanes?geometry=-82.798%2C34.842%2C-78.843%2C35.627> [Accessed 21 Jul. 2020].
- Charlotte DOT (2019b), ‘Parking meters’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/datasets/parking-meters?geometry=-80.849%2C35.223%2C-80.834%2C35.226> [Accessed 21 Jul. 2020].
- Charlotte DOT (2019c), ‘Traffic signals’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/datasets/traffic-signals-1> [Accessed 21 Jul. 2020].
- CharlotteNC (2020), ‘High Injury Network’. Electronic dataset, Charlotte Open Data Portal, available at: https://data.charlottenc.gov/datasets/4a27df737ba24be7acf1314f38925a4f_87 [Accessed 31 Jul. 2020].
- City of Charlotte (2015), ‘City of Charlotte open data policy’, Open Mapping, available at: <https://charlotte.maps.arcgis.com/home/item.html?id=7c88b8633b034ddcbbd6badb1b7076fe> [Accessed 31 Jul. 2020].

- City of Charlotte (2016), *Beyond Traffic: Smart City Challenge*, U.S. Department of Transportation. Available at: <https://www.transportation.gov/sites/dot.gov/files/docs/NC%20Charlotte.pdf>.
- City of Charlotte (2020), ‘Vision Zero public comment map’. City of Charlotte, State of North Carolina DOT, Esri, HERE, Garmin, NGA, USGS, NPS, available at: <https://charlotte.maps.arcgis.com/apps/CrowdsourceReporter/index.html?appid=f0b9fb71ee3e4c70ba8d5b400b729946> [Accessed 2 Aug. 2020].
- City of Charlotte DOT (2017), *Charlotte WALKS Pedestrian Plan*, City of Charlotte. Available at: <https://charlottenc.gov/Transportation/Programs/Documents/Charlotte%20WALKS%20Adopted%20Plan%20-%20February%202017.pdf>.
- City of Charlotte DOT (2019), *Vision Zero: Safer Streets for Charlotte 2019-2030 Action Plan*, City of Charlotte. Available at: <https://charlottenc.gov/VisionZero/Documents/Vision%20Zero%20Action%20Plan.pdf>.
- City of Charlotte Staff (2007), *Urban Street Design Guidelines*, City of Charlotte. Available at: <https://charlottenc.gov/Transportation/PlansProjects/Documents/USDG%20Full%20Document.pdf>.
- City of Vancouver (2020), ‘Making streets for people: Temporary road closures and changes in response to COVID 19’, City of Vancouver, available at: <https://vancouver.ca/home-property-development/covid-19-coronavirus-road-closures.aspx#redirect> [Accessed 18 Aug. 2020].
- Claris, S. & Scopelliti, D. (2016), *Cities Alive: Towards a Walking World*, Arup, London. Available at: <https://www.arup.com/perspectives/publications/research/section/cities-alive-towards-a-walking-world>.

- CRTPO (2020), ‘Traffic counts’. Charlotte Regional Transportation Planning Organization, available at: <http://www.crtpo.org/resources/traffic-counts> [2 Aug. 2020].
- de Andrade, L., Vissoci, J. R. N., Rodrigues, C. G., Finato, K., Carvalho, E., Pietrobon, R., de Souza, E. M., Nihei, O. K., Lynch, C. & de Barros Carvalho, M. D. (2014), ‘Brazilian road traffic fatalities: A spatial and environmental analysis’, *PLOS ONE* 9(1), 1–10.
- de Jong, R., Koinange, C. & Odero, D. (2019), *Share the Road Programme Annual Report 2018*, UN Environment. Available at: <https://www.unenvironment.org/resources/annual-report/share-road-programme-annual-report-2018>.
- Ford, C. (2016), ‘Getting started with negative binomial regression modeling’. University of Virginia Library, available at: <https://data.library.virginia.edu/getting-started-with-negative-binomial-regression-modeling/> [Accessed 21 Jul. 2020].
- Google (2018), ‘311 E 6th St’. Google Street View [online], available at: <https://goo.gl/maps/UoQKyWFBMGjCat6p9> [Accessed 12 Aug. 2020].
- Google (2019a), ‘101 W Trade St’. Google Street View [online], available at: <https://goo.gl/maps/R5QCpe6BTNJDaBor9> [Accessed 12 Aug. 2020].
- Google (2019b), ‘263 E Stonewall St’. Google Street View [online], available at: <https://goo.gl/maps/qe36Greel7wZ8ZTZQ9> [Accessed 12 Aug. 2020].
- Google (2019c), ‘299 E 4th St’. Google Street View [online], available at: <https://goo.gl/maps/TqYNwwD3hY5nZCif7> [Accessed 12 Aug. 2020].
- Google (2019d), ‘667 E M.L.K. Jr Blvd’. Google Street View [online], available at: <https://goo.gl/maps/hPKSFzhZvnh7dmph6> [Accessed 12 Aug. 2020].

- Google (2019e), ‘767 E 3rd St’. Google Street View [online], available at: <https://goo.gl/maps/KTNEYoi1BkbRojdz9> [Accessed 12 Aug. 2020].
- Greene-Roesel, R., Diogenes, M. & Ragland, D. (2007), ‘Estimating pedestrian accident exposure: Protocol report’, *University of California Traffic Safety Center*. Available at: <https://escholarship.org/uc/item/8j8685jt>.
- Griffin, G. P. & Jiao, J. (2019), ‘The geography and equity of crowdsourced public participation for active transportation planning’, *Transportation Research Record* **2673**(1), 460–468.
- Gupta, S. (2020), ‘Pros and cons of various machine learning algorithms’. Towards Data Science, available at: <https://towardsdatascience.com/pros-and-cons-of-various-classification-ml-algorithms-3b5bfb3c87d6> [Accessed 3 Aug. 2020].
- Halais, F. (2020), ‘In the walking capitals of the world, drivers still rule the road’, Wired, available at: <https://www.wired.com/story/in-the-walking-capitals-of-the-world-drivers-still-rule-the-road/> [Accessed 28 Jul. 2020].
- Henderson, B. (2020), ‘Charlotte pedestrian deaths continue at record pace, despite more sidewalks, crossings’, The Charlotte Observer, available at: <https://www.charlotteobserver.com/news/local/crime/public-safety-blog/article239925293.html> [Accessed 30 Jul. 2020].
- Hezaveh, A. M., Azad, M. & Cherry, C. R. (2018), Pedestrian crashes in Tennessee: A data mining approach, in ‘Transportation Research Board 97th Annual Meeting’, Washington, D.C. Available at: <https://trid.trb.org/view/1494936#:~:text=Findings%20indicated%20that%20odds%20of,only%20in%20crashes%20involved%20pedestrians> .
- Hidalgo, D. (2020), ‘Colombia’s new national urban mobility policy is mostly a good thing’, The City Fix, available at: <https://thecityfix.com/blo>

- g/colombias-new-national-urban-mobility-policy-most-likely-good-thing-dario-hidalgo/ [Accessed 18 Aug. 2020].
- Hirst, D. (2019), ‘Active travel: Trends, policy and funding’, *House of Commons Research Briefing Paper*. Number 8615. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8615/>.
- Institute for Transportation Research and Education (2020), ‘North carolina non-motorized volume data program’. NC State University, available at: <https://itre.ncsu.edu/focus/bike-ped/nc-nmvdp/> [Accessed 2 Aug 2020].
- International Transport Forum (2012), *Pedestrian Safety, Urban Space and Health*, OECD Publishing. Available at: <http://dx.doi.org/10.1787/9789282103654-en>.
- ITDP (2018), *Pedestrians First: Tools for a Walkable City*, Institute for Transportation and Development Policy, New York. Available at: <https://www.itdp.org/publication/walkability-tool/>.
- Jahangiri, A., Hasani, M., Sener, I. N., Munira, S., Owens, J., Appleyard, B., Ryan, S., Turner, S. M. & Machiani, S. G. (2019), *Data Mining to Improve Planning for Pedestrian and Bicyclist Safety*, SAFE-D Report Number 01-003, SAFE-D: Safety Through Disruption National University Transportation Center. Available at: <https://vttechworks.lib.vt.edu/handle/10919/96338>.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012), ‘Benefits, adoption barriers and myths of open data and open government’, *Information Systems Management* **29**, 258–268.
- Jenkins, W. (2019a), ‘Lynx blue line stations’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/datasets/lynx-blue-line-stations> [Accessed 16 Jul. 2020].

- Jenkins, W. (2019b), ‘Lynx gold line stops’. Electronic dataset, Charlotte Open Data Portal, available at: <https://data.charlottenc.gov/dataset/s/lynx-gold-line-stops> [Accessed 14 Jul. 2020].
- JJ (2016), ‘MAE and RMSE — which metric is better?’. Human in a Machine World, available at: <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d> [Accessed 3 Aug. 2020].
- Jolliffe, I. T. & Cadima, J. (2016), ‘Principal component analysis: A review and recent developments’, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374**(20150202), 1–16.
- Kahle, D. & Wickham, H. (2013), ‘ggmap: Spatial visualization with ggplot2’, *The R Journal* **5**(1), 144–161. Available at: <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- Kapenekakis, I. & Chorianopoulos, K. (2017), ‘Citizen science for pedestrian cartography: Collection and moderation of walkable routes in cities through mobile gamification’, *Human-Centric Computing and Information Sciences* **7**(10), 1–23.
- Khayesi, M., Monheim, H. & Nebe, J. M. (2010), ‘Negotiating “streets for all” in urban transport planning: The case for pedestrians, cyclists and street vendors in Nairobi, Kenya’, *Antipode* **42**(1).
- Koehrsen, W. (2018), ‘An implementation and explanation of the random forest in python’. Towards Data Science, available at: <https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76> [Accessed 3 Aug. 2020].
- Krambeck, H. V. (2006), The Global Walkability Index, PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Kumfer, W., Thomas, L., Sandt, L. & Lan, B. (2019), ‘Midblock pedestrian crash

- predictions in a systemic, risk-based pedestrian safety process', *Transportation Research Record* **2673**(11), 420–432.
- Kunstler, J. H. (2004), 'The ghastly tragedy of the suburbs', [video], TED Conferences, LLC. Available at: https://www.ted.com/talks/james_howard_kunstler_the_ghastly_tragedy_of_the_suburbs#t-1003755 [Accessed 12 Jul. 2020].
- Kurkcu, A. & Ozbay, K. (2017), 'Estimating pedestrian densities, wait times, and flows with wi-fi and bluetooth sensors', *Transportation Research Record: Journal of the Transportation Research Board* **2644**(1), 72–82.
- Lai, Y. & Kontokosta, C. (2017), Analyzing the drivers of pedestrian activity at high spatial resolution, in 'International Conference on Sustainable Infrastructure', American Society of Civil Engineers, New York, pp. 303–314.
- Lee, K. & Sener, I. N. (2017), 'Emerging data mining for pedestrian and bicyclist monitoring: A literature review report', *Safety through Disruption (Safe-D) National University Transportation Center (UTC) Program* pp. 1–37. Available at: https://www.vtti.vt.edu/utc/safe-d/wp-content/uploads/2018/04/UTC-Safe-D_Emerging-Data-Mining-for-PedBike_TTI-Report_26Sep17_final.pdf.
- Lee, K. & Sener, I. N. (2020), 'Emerging data for pedestrian and bicycle monitoring: Sources and applications', *Transportation Research Interdisciplinary Perspectives* **4**(100095).
- Leeson, A., Alvarez, P. & Ghosh, S. (2014), Understanding how big data and crowd movements will shape the cities of tomorrow, in 'European Transport Conference', Association for European Transport, Frankfurt, Germany, pp. 1–13. Available at: <https://trid.trb.org/view/1341346>.
- Litman, T. A. (2003), 'Economic value of walkability', *Transportation Research Record* **1828**(1), 3–11.

- Loh, T. H., Leinberger, C. B. & Chafetz, J. (2019), *Foot Traffic Ahead: Ranking Walkable Urbanism in America's Largest Metros*, The George Washington School of Business and Smart Growth America.
- Malchik, A. (2019), *A Walking Life: A Walking Life: Reclaiming Our Health and Our Freedom One Step at a Time*, Da Capo Press, New York.
- Mapillary (2020), 'Make better maps'. Available at: <https://www.mapillary.com/> [Accessed 31 Jul. 2020].
- Marshall, W. E. & Garrick, N. W. (2010), 'Effect of street network design on walking and biking', *Transportation Research Record* **2198**(1), 103–115.
- Mecklenburg County (2020a), 'About', Open Mapping, available at: <http://maps.co.mecklenburg.nc.us/openmapping/about.html> [Accessed 31 Jul. 2020].
- Mecklenburg County (2020b), 'Mecklenburg County jurisdictions'. Electronic dataset, Open Mapping, available at: <http://maps.co.mecklenburg.nc.us/openmapping/data.html?search=Mecklenburg%20County%20Jurisdictions> [Accessed 21 Jul. 2020].
- Mecklenburg County (2020c), 'Tax parcel ownership and building data'. Electronic dataset, Open Mapping, available at: <http://maps.co.mecklenburg.nc.us/openmapping/data.html?search=Tax%20Parcel%20Ownership%20and%20Building%20Data> [Accessed 21 Jul. 2020].
- MIG, Inc. (2011), *Charlotte Center City 2020 Vision Plan*, Charlotte Center City Partners. Available at: <https://www.mecknc.gov/OpenMecklenburg/Documents/CenterCityVision2020.pdf>.
- MIG, Inc. (2016), 'The North Tryon vision plan'. Available at: <http://northtryon.org/vision-plan/> [Accessed 31 Jul. 2020].

- Millward, H., Spinney, J. & Scott, D. (2013), 'Active-transport walking behavior: Destinations, durations, distances', *Journal of Transport Geography* **28**, 101–110.
- Mohan, D. (2003), 'Road traffic injuries - a neglected pandemic', *Bulletin of the World Health Organization* **81**(9), 684–685.
- Montgomery, C. (2013), *Happy City: Transforming Our Lives Through Urban Design*, Farrar, Straus and Giroux, New York.
- Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., Jamshidi, E. & Zeini, S. (2016), 'Spatial analysis to identify high risk areas for traffic crashes resulting in death of pedestrians in Tehran', *Medical Journal of the Islamic Republic of Iran* **30**(450), 1–10.
- Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., Nematollahi, S. & Zeini, S. (2017), 'Effective environmental factors on geographical distribution of traffic accidents on pedestrians, downtown Tehran City', *International Journal of Critical Illness and Injury Science* **7**(2), 101–106.
- Murguía, R. O. (2018), *Share the Road: Design Guidelines for Non Motorised Transport in Africa*, United Nations Environment Programme. Available at <https://www.unenvironment.org/resources/toolkits-manuals-and-guides/share-road-design-guidelines-non-motorized-transport-africa> [Accessed 28 Jul. 2020].
- National Highway Traffic Safety Administration (2017), 'Traffic safety facts: 2015 data'. Available at: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812375#:~:text=In%202015%20there%20were%205%2C376,ans%20killed%20annually%20since%201996.&text=In%202015%2C%20pedestrian%20deaths%20accounted,percent%20of%20all%20traffic%20fatalities.>
- NCDOT (2020), 'NCDOT road characteristics'. Electronic dataset, NC OneMap, available at: <https://www.nconeemap.gov/datasets/NCDOT::>

ncdot-road-characteristics?geometry=-80.856%2C35.222%2C-80.825%2C35.228 [Accessed 21 Jul. 2020].

NCDOT DBPT (2020), ‘Bicyclist and pedestrian crashes that occurred in North Carolina between years 2007 through 2018’. Electronic dataset, ArcGIS, available at: <https://www.arcgis.com/home/item.html?id=801a34b6d3e64786a131d1fa2eca793d> [Accessed 21 Jul. 2020].

Negative Binomial Regression: Stata Annotated Output (2006). UCLA Statistical Consulting Group, available at: <https://stats.idre.ucla.edu/stat/a/output/negative-binomial-regression/#:~:text=We%20can%20interpret%20the%20negative,in%20the%20model%20are%20held> [Accessed 19 Aug. 2020].

NYC DOT (2020), ‘Vision Zero borough pedestrian safety action plans’, The City of New York, available at: <https://www1.nyc.gov/html/dot/html/pedestrians/ped-safety-action-plan.shtml> [Accessed 18 Aug. 2020].

NYC Streets Renaissance (2006), ‘Interview with enrique peñalosa’, [video], Streetfilms. Available at: <https://www.streetfilms.org/interview-with-enrique-penalosa-short/> [Accessed 8 Aug. 2020].

OpenStreetMap Contributors (2020), ‘Planet dump retrieved from <https://planet.osm.org>’.

Otieno, O. M. (2013), Evaluation of pedestrian walkways along urban roads: The case of Komarock, Nairobi, Master’s thesis, University of Nairobi, Nairobi, Kenya. Available at: <https://www.semanticscholar.org/paper/Evaluation-of-Pedestrian-Walkways-along-Urban-The-Orege/46f062fe4db2aecc1e1b3464fb467f3eb93c1c3c>.

Patel, A., Krebs, E., Andrade, L., Rulisa, S., Vissoci, J. R. N. & Staton, C. A. (2016), ‘The epidemiology of road traffic injury hotspots in Kigali, Rwanda from police data’, *BMC Public Health* **16**(697), 1–10.

- Pendakur, V. S. (2005), ‘Non-Motorized transport in African cities: Lessons from experience in Kenya and Tanzania’, *Sub-Saharan Africa Transport Policy Program SSATP World Bank Working Paper* (80).
- Petrasova, A., Hipp, J. A. & Mitasova, H. (2019), ‘Visualization of pedestrian density dynamics using data extracted from public webcams’, *International Journal of Geo-Information* 8(599), 1–16.
- Pojani, D. & Stead, D. (2015), ‘Sustainable urban transport in the developing world: Beyond megacities’, *Sustainability* 7, 7784–7805.
- Porter, T. (2018), ‘Charlotte, NC greenways trees’. Electronic dataset, Koordinates, available at: <https://koordinates.com/layer/96938-charlotte-nc-greenways-trees/> [Accessed 21 Jul. 2020].
- Qian, Z. S. (2016), Real-time incident detection using social media data, Technical report, Pennsylvania Department of Transportation and Carnegie Mellon University, Pennsylvania. Available at: <https://www.penndot.gov/ProjectAndPrograms/Planning/Research-And-Implementation/Documents/Real%20Time%20Incident%20Detection%20Using%20Social%20Media%20Data.pdf>.
- R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raford, N. & Ragland, D. R. (2006), ‘Pedestrian volume modeling for traffic safety and exposure analysis: The case of Boston, Massachusetts’, *UC Berkeley: Safe Transportation Research Education Center*.
- Resor, E. (2015), ‘Nairobi accident map’, Available at <https://nairobiaccidentmap.com/> [Accessed 21 Aug. 2020].
- Santani, D., Njuguna, J., Bills, T., Bryant, A. W., Bryant, R., Ledgard, J. & Gatica-Perez, D. (2015), Communisense: Crowdsourcing road hazards in Nairobi, *in*

- 'Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services', MobileHCI '15, Association for Computing Machinery, Copenhagen, Denmark, p. 445–456. Available at: <https://doi.org/10.1145/2785830.2785837>.
- Smart Growth America (2015), *(Re)Building Downtown: A Guidebook for Revitalization*, Smart Growth America, Washington, D.C. Available at: <https://smartgrowthamerica.org/resources/rebuilding-downtown/>.
- Solmaz, G. & Turgut, D. (2019), 'A survey of human mobility models', *IEEE Access* **7**, 125711–125731.
- Southworth, M. (2005), 'Designing the walkable city', *Journal of Urban Planning and Development* **131**(4), 246–257.
- Speck, J. (2013), *Walkable City: How Downtown Can Save America, One Step at a Time*, North Point Press, New York.
- Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M. N., Peden, M. & Ewing, R. (2015), 'Pedestrian safety and the built environment: A review of the risk factors', *Journal of Planning Literature* **30**(4), 377–392.
- Strobl, C., Hothorn, T. & Zeileis, A. (2009), 'Party on! A new, conditional variable importance measure for random forests available in the party package', *The R Journal* **1**, 14–17.
- Taquechel, E. P. (2009), A spatial analysis of the relationship between pedestrian crash events and features of the built environment in downtown Atlanta, Master's thesis, Georgia State University, Atlanta, Georgia.
- Taylor, B. T., Fernando, P., Bauman, A. E., Williamson, A., Craig, J. C. & Redman, S. (2011), 'Measuring the quality of public open space using Google Earth', *American Journal of Preventive Medicine* **40**(2), 105–112.

TfL (2018), *Walking Action Plan: Making London the World's Most Walkable City*, Transport for London. Available at: <http://content.tfl.gov.uk/mts-walking-action-plan.pdf>.

Thorton, B. (2015), ‘Where is walking in the UN sustainable development goals?’, Smart Cities Dive, available at: <https://www.smartcitiesdive.com/ex/sustainablecitiescollective/where-walking-un-sustainable-development-goals/1073491/> [Accessed 24 Jul. 2020].

Traffic Survey Group (2018), ‘2018 traffic segments shapefile’. Electronic dataset, Connect NCDOT, available at: <https://connect.ncdot.gov/resources/State-Mapping/Pages/Traffic-Survey-GIS-Data.aspx> [Accessed 21 Jul. 2020].

Transportation Research Board and National Academies of Sciences, Engineering, and Medicine (2018), *Systemic Pedestrian Safety Analysis*, The National Academies Press, Washington, DC.

U.S. DOT FHA (2016), *Road Diet Informational Guide*, Washington, D.C. Available at: https://safety.fhwa.dot.gov/road_diets/resources/fhwasa16100/fhwasa16100.pdf.

Vision Zero Network (2018), ‘What is Vision Zero?’. Available at: <https://visionzeronetwork.org/about/what-is-vision-zero/> [Accessed 29 Jul. 2020].

Walk Score (2020), ‘Living in Charlotte’, Available at <https://www.walkscore.com/NC/Charlotte> [Accessed 22 Jul. 2020].

Welle, B., Sharpin, A. B., Adriazola-Steil, C., Job, S., Shotten, M., Bose, D., Bhatt, A., Alveano, S., Obelheiro, M. & Imamoglu, C. T. (2018), *Sustainable & Safe: A Vision and Guidance for Zero Road Deaths*, World Resources Institute, Washington, D.C. Available at: <http://pubdocs.worldbank.org/en/912871516999678053/Report-Safe-Systems-final.pdf>.

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019), 'Welcome to the tidyverse', *Journal of Open Source Software* **4**(43), 1686.
- Yang, S. & Berdine, G. (2015), 'Negative binomial regression', *The Southwest Respiratory and Critical Care Chronicles* **3**(10).
- Zaccaro, H. (2019), *Dangerous by Design*, Smart Growth America, Washington, D.C. Available at: <https://smartgrowthamerica.org/dangerous-by-design/>.
- Zegeer, C. V., Nabors, D. & Lagerwey, P. (2013), *Pedestrian Safety Guide and Countermeasure Selection System*, Federal Highway Administration, Office of Safety, chapter Selecting Improvements for Pedestrians, pp. 11–34. Available at: http://www.pedbikesafe.org/pedsafe/guide_analysis.cfm.

Appendix A

Pedestrian Crash Descriptive Stats

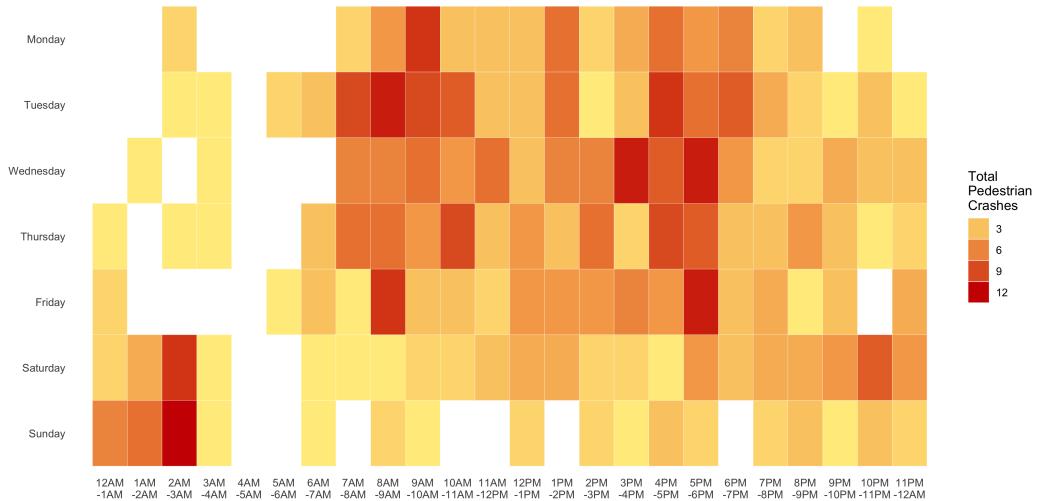


Figure A.1: Temporal Heatmap of Pedestrian Crashes in Uptown Charlotte (2007-2018). Pedestrian crashes appear to be concentrated around peak commute times during the week, and late at night during the weekend. Crashes that occurred on the weekend were removed for this study, as it is possible that the nature of weekday and weekend crashes is different.

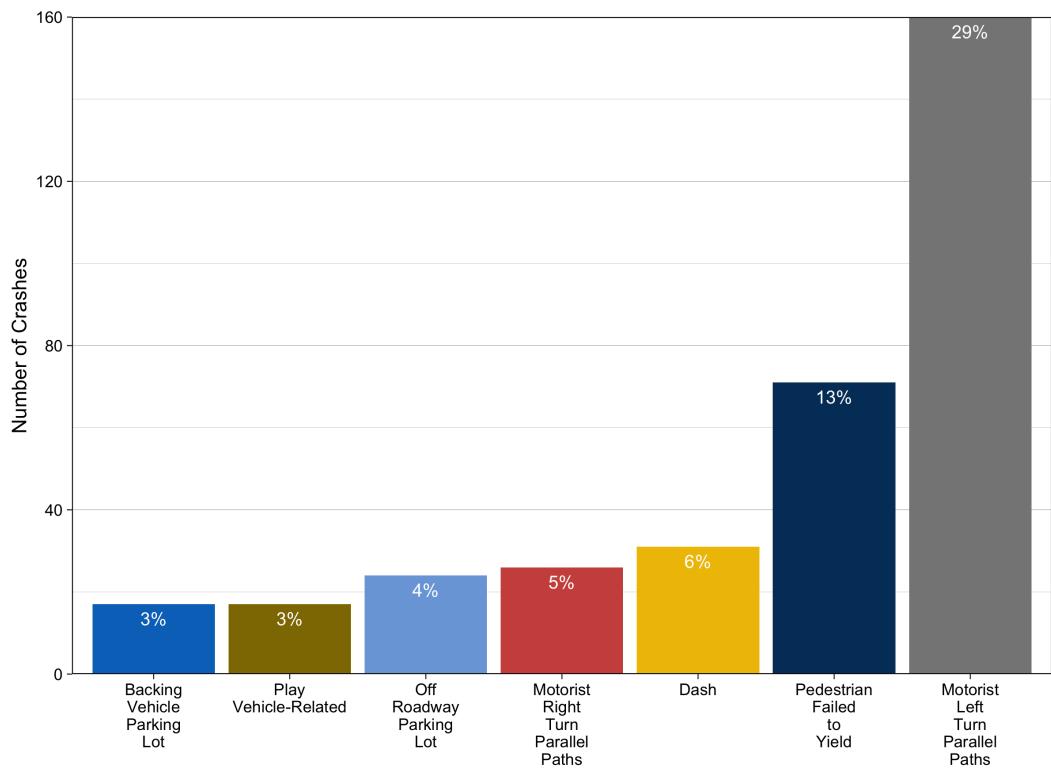


Figure A.2: Histogram of Pedestrian Crash Types. Includes all crashes in Uptown 2007-2018, and types with 15 or more occurrences. Reflects some problems that Uptown Charlotte may be facing in terms of the relationship between urban design and pedestrian safety, although the presence of left turn lanes was not included as a variable in the final negative binomial regression model.

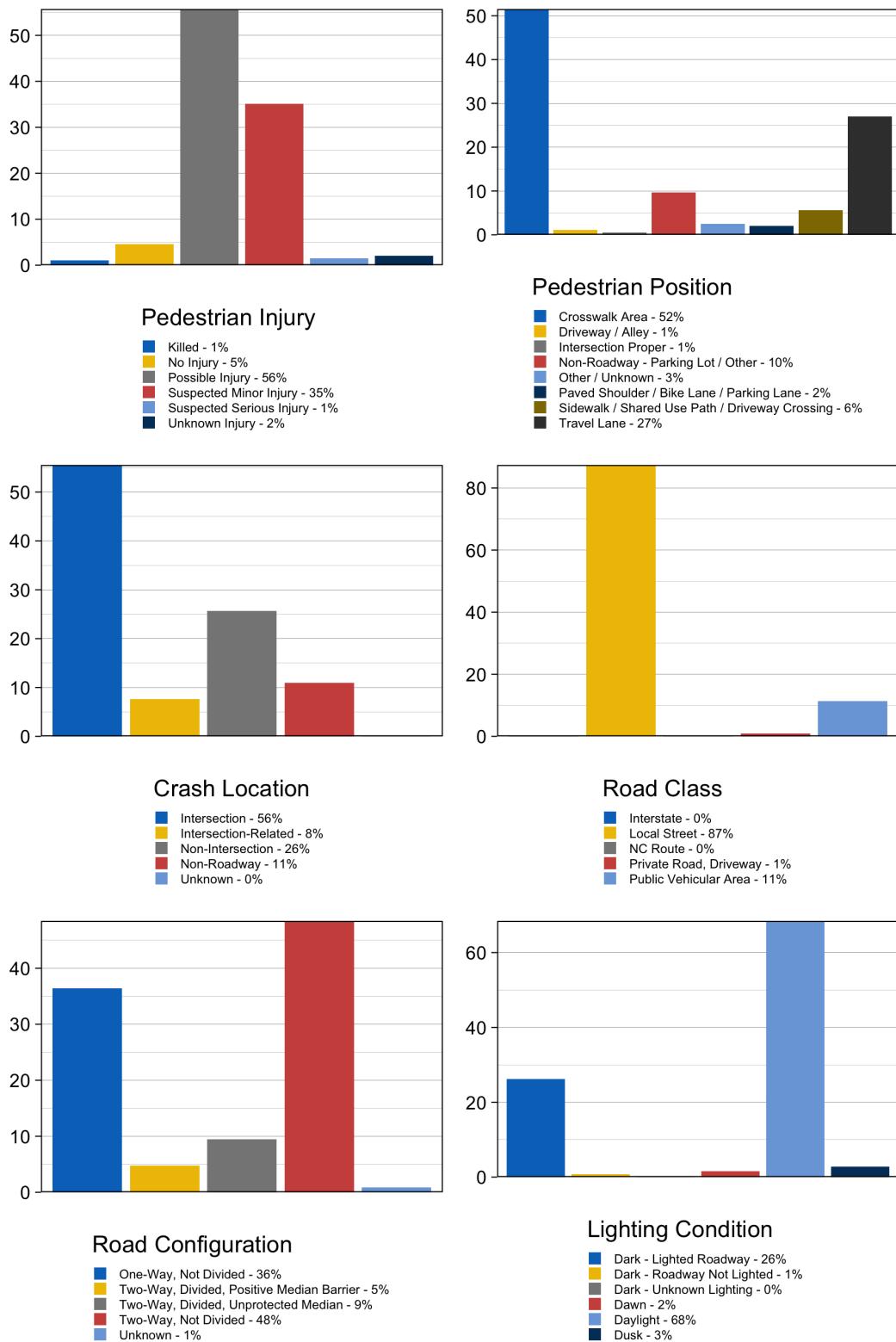


Figure A.3: Exploratory Bar Charts for Pedestrian Crash Attributes. Bar charts include all crashes that occurred in Uptown Charlotte from 2007-2018. Crashes with ‘non-roadway’ locations were removed for the study, as were crashes that did not occur on local streets.

Appendix B

Random Forest Variable Importance

Table B.1: Variable Importance. Predictors with mean decrease in accuracy values of greater than 0 were considered for inclusion in the negative binomial regression model.

Predictor	MDA
traveldirection	0.00123
funcclass	0.0301
lfttrnlntyp	-0.0000728
median	-0.00000666
rttrnlntyp	0
surfacewidth	-0.000446
thrulanes	0.026
leftshoulder	0
rightshoulder	0
length_ft	0.112
housingunits	0.0297
busstops	0.0302
lightrail	0.00000901
trafficsignals	0.372
trees	0.0659
bikelanes	-0.000547
parkingmeters	0.00613
residential	0.0013
multifamily	0.00052
office	0.000391
commercial	0.0253
crosswalks	-0.000607
traffvols	-0.00049
edgebetw	0.0698

Appendix C

Research Log

Date	Tasks
31 May 2020	looked for data sources, contacted industry partner, began literature review, set up Monday.com schedule
7 June 2020	revised topic and research question based on lack of data for former idea, began new literature review, located traffic incident data, contacted industry partner
14 June 2020	discussed new topic with industry partner, continued literature review, contacted sources to get data
21 June 2020	settled on final topic after changing (again due to lack of access to required data)
28 June 2020	continued literature review, investigated methodologies, located supplementary data sources, began descriptive statistics and preliminary figures on pedestrian crash data
5 July 2020	set up GitHub, further data collection, more descriptive statistics, began writing methodology section
12 July 2020	self-isolation, decided on direction for methodology, more work on specifics of analysis
19 July 2020	finished data preparation, continued research for methodology, random forest model and analysis
26 July 2020	spoke with supervisor and discussed results, adjusted code and methodology, began writing up results and considering interpretation
2 August 2020	edited figures, outputs, and results section, read various policy documents
9 August 2020	wrote introduction and edited formatting of paper according to handbook
16 August 2020	writing retreat, wrote discussion and conclusion, updated GitHub, sent draft to supervisor
23 August 2020	final edits