An Empirical Analysis of Bike Safety in Lawrence Using Road Geometry and Traffic Characteristics

By

Shofi Ull Azum Shofi

Submitted to the graduate degree program in Urban planning, Geography and Atmospheric Science the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chair: Ting Lei

Xingong Li

Ward Lyles

Date Defended: 31 August 2017

The thesis committee for Shofi Ull Azum Shofi certifies that this is
the approved version of the following thesis:

An Empirical Analysis of Bike Safety in Lawrence Using Road Geometry and Traffic Characteristics

Chair: Ting Lei

Date Approved: 31 August 2017

Abstract

This study focuses on analyzing bike route safety in Lawrence using road geometry, infrastructure, and traffic characteristics. Bike crash incidence has been considered as a measure of bike route safety in this study. The independent variables considered for the bike route safety analysis are the number of lanes, route slopes (average), traffic volume (dummy variable based on functional road classification), the availability of bike routes, and posted speed limits. Bike crash data (the dependent variable), and Digital Elevation Model (DEM) data were collected from city of Lawrence. For the study purpose, streetwise bike route facilities, and traffic lane numbers have been updated based on city of Lawrence database. Finally, the average slope for each street in Lawrence has been calculated from DEM raster using ArcMap. As the data were characterized by over-dispersion and zero inflation, conventional negative binomial and zero inflated negative binomial models generate statistically significant variable coefficients. Interestingly, coefficients from both model have produced near identical bike compatibility maps for Lawrence. The study has found that bike route safety decreases with the increase of the traffic volume and lane numbers. In other words, collector and arterial roads are not the safest option for bicyclists in Lawrence, but the local neighborhood level streets are more suitable for biking. The route slope has no significant impact on bike route safety and the speed is negatively related with bike crash incidence. The unavailability of actual bike count data and bike speed data result in some flaws in the outcome of bike compatibility map. In a nutshell, complex statistical analysis adds some values in the current understanding of bike safety with the data available.

Acknowledgments

First and foremost, I would like to thank my supervisor, **Dr. Ting Lei** for his guidance, corrections on the draft, and being supportive all the time. Without his suggestions and supervision, it would be very difficult for me to conduct this research and finish it on time. I am grateful to him and I do highly appreciate his support and benevolence.

I would like to thank **Dr. Ward Lyles** and **Dr. Xingong Li** for reviewing the thesis draft and providing necessary corrections and guidance.

I would like to thank **Micah Seybold**, GIS coordinator of City of Lawrence for his support to facilitate acquisition of necessary secondary data. **Jessica Mortinger**, senior transportation planner, also helped me by providing pertinent studies available at city of Lawrence.

I am also thankful to my parents who were always with me and gave me support in a hard time.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Biking is becoming a very popular mode of transport for commuters in modern cities, especially as the last/first mile connector. It not only offers physical and environmental benefits but also reduces the negative impact of motor vehicle travel on the society. There is a growing interest to promote the bicycle as an alternative form of transportation and a means to reduce the dependency on automobiles for shorter trips. The widespread use of automobiles for daily commutes in urban areas contributes to serious environmental, economic, and public health problems. The heavy dependency on automobiles can affect urban mobility and the quality of life. On the other hand, bicycle offers a sustainable active transportation option to city dwellers and it helps to improve public health conditions. In such a context, the bicycle may be a viable option for a more sustainable urban mobility.

However, the provision of efficient bicycle system in the cityscape is challenging, because city routes are generally characterized with high traffic volumes, speed, and wider routes. These routes are prone to bike crash unless sufficient bicycle route facilities are available. The common forms of bike route infrastructure are separate bike routes, bike lanes, and shared use paths. It is generally believed that the existence of continuous bicycle route infrastructure reduces bicycle crash incidence. Besides the bicycle infrastructure, the physical and geographical attributes of the route influence bicyclists' route choice. Therefore, these factors need to be considered in providing a safe and convenient bike route. It is worthwhile to investigate these factors to reduce bike crashes and make bicycling safe for all ages in Lawrence.

1.1. Research Question, Objective and Problem Statement

This section comprises the research question including the main inquiry of the study, a set of objectives to achieve the study goal and a problem statement. The research question of this study is following:

"What road geometry, infrastructure, and traffic characteristics influence bike route safety in the City of Lawrence, Kansas?"

The study investigates the following more specific questions to understand the bike route safety conditions in Lawrence.

- Is there any relationship between the route slope and the number of bike crashes?
- Does the presence of bike routes influence the safety for bicyclists?
- Do the increase of lane number, speed, and traffic volume positively affect bike crash incidence?
- Where are the existing bicycle crash zones located in Lawrence?
- Which routes are safer for bicyclists in Lawrence?

Three specific objectives have been identified in studying bike route safety in Lawrence:

- To explore the relationship between road geometry, traffic characteristics, and bike crash incidence
- To develop a bike compatibility map that would identify safer bike routes in Lawrence
- To investigate whether an empirical analysis makes any difference in the understandings of bike safety with the available data

The problem statement illustrates the need of research and provides evidence to support the study goal. Lawrence is a college town with a population of just over 92,500 and the bicycle is a very popular mode of transport in this city compared to other cities in Douglas County (United

States Census Bureau, 2014). The city of Lawrence already has some (maybe insufficient) bike routes, bike lanes and shared use paths in the existing city routes. Despite these existing bike infrastructure, the rate of bicycle and pedestrian crashes in the City of Lawrence is significantly higher than other cities in Douglas County, Kansas (Mortinger, 2015). About 97% bike and pedestrian crash incidents of Douglas County occurred in Lawrence between 2009 and 2013. (City of Lawrence, 2015). According to ETC Institute Citizen Survey in 2015, only 10% residents feel very safe riding bike in Lawrence, and 21% are satisfied with the connectivity of bicycle lanes (City of Lawrence, 2015). The University of Kansas (KU) is the major hub or activity center in the city of Lawrence, and KU is located on top of a hill- Mount Oread. An empirical research is needed to conclude whether the steep slopes of the connecting routes to the University of Kansas makes biking more challenging or some other road infrastructure and traffic characteristics such as lanes, traffic volumes, speed, bike routes etc. play a key role in bike crashes. To this end, it is important to identify the prevailing crash zones in the City of Lawrence and figure out the safe route for bicyclists. The provision of bike infrastructure can play a vital role in ensuring safety to the riders e.g. KU students, faculties, staff, city residents etc. In a nutshell, the objectives of this study are to figure out how and what road geometry, infrastructure, and traffic characteristics affect bike safety in the city of Lawrence, KS. This study identifies the existing bike crash zones, the spatial relationship between the route slope and the number of crash incidence and between the presence of bike facilities and crash incidence. Finally, a bike compatibility map is prepared to identify which routes are more comfortable and safer options for the bicyclists in Lawrence. The similarities and distinction of this compatibility map with the City's bike rideability map would be helpful in providing recommendations for the next moves of the City of Lawrence.

1.2. Status quo of Bike Analysis (Plans and Policies)

1.2.1. University of Kansas Bike Plan. The University of Kansas has updated its campus bike plan in spring 2016 and the League of American Bicyclists designated city of Lawrence as a bicycle friendly community at the bronze (the lowest) level in August 2016. The plan tried to achieve following goals:

- Enhance the multimodal network linking residential, academic, and recreational destinations on campus and in the community
- Promote a safe and healthy campus environment
- Increase bicycle and pedestrian mode share through the implementation of new policies, programs, and infrastructure
- Improve coordination with the City of Lawrence and create seamless transitions between university and city bike routes and infrastructure.
- Improve movement uphill by identifying policy, program, and infrastructure solutions that encourage people to overcome the real and perceived barrier of steep routes to campus (Center for Sustainability, 2016).

1.2.2. Countywide (Douglas County) Bikeway System Plan. The main goal of the Countywide Bikeway Plan was to identify bikeway recommendations for the portions of Douglas County that have not previously had bikeways planned. In addition to the provision of bike infrastructure, the plan has emphasized 5 E’s (Encouragement, Education, Enforcement, Evaluation, and Engineering). The major recommendations include building bicycle parking facilities, connecting bicycle facilities and routes to transit stops, continue installing bike racks on all buses, a bike share feasibility study which is now underway.

1.2.3. Bicycle Rideability map. The bicycle rideability map is designed to promote, encourage and educate bicyclists. The map assists riders in choosing routes most applicable to their skill level and alerts bicyclists about difficult intersections. It also shows major landmarks to help bicyclists navigate around town and identifies transit connections.

The creation of the rideability layer was not scientific. It evolved over time, the current rideability map is the third generation of it. There were 9 Bicycle Advisory Committee (BAC) members on the committee. BAC members were asked to draw colored lines to represent the categories and then they discussed particular sections if they had differences. No road geometry and traffic factors were considered. BAC members used their own experiences to rank the routes. This is very subjective judgment and a limitation of the current rideability map.

Bicycle planning has always been a priority for the city of Lawrence since last two decades and the increasing citizen concern about bicycling issues prompted the city commission to form Bicycle Advisory Committee (BAC) in the mid-1990s. The initial purpose of BAC was to promote bicycle transportation and safety and the future of bikeway planning in Douglas County. Over the years, BAC, currently known as Transportation Commission, has made recommendations on the location and design of bikeways, expenditures of public funds for bikeway development and maintenance, promotion of bicycle use and safety, street design issues related to bicycle use, and other bicycle and transportation issues.

1.3. Research Gap (Motivations of the Study)

The empirical study of bike safety analysis using road geometry, infrastructure, and traffic characteristics is non-existent. The City of Lawrence has a bike rideability map which was prepared by Bicycle Advisory Committee (BAC) members based on their riding experiences. This measurement of bicycle safety is subjective and not scientific. There might be a gap between actual safer routes and perceived safer routes by BAC members. That is why an empirical research is required to find out the safer routes for bicyclists and what road factors contribute to the bicyclist's safety. A statistical analysis (regression) has been conducted to best fit the data and the regression coefficients are then used to develop a bike compatibility map for Lawrence. To find the existing bike crash zones, a kernel density map has been prepared using ArcMap. The bike safety analysis would investigate what factors are responsible for an area prone to a bike crash. This information would be useful for city planners and decision makers to understand what measures need to be taken to improve bike safety for certain areas. Finally, the bike compatibility map would be helpful in identifying the safe and convenient routes for the bike users. To pursue this scope of study, the city of Lawrence has been chosen as a study area and it is important to know some basic geographic features and existing bike route facilities in Lawrence.

1.4. Study Area

Lawrence is the sixth largest city in the state of Kansas and it is a county seat of Douglas County, Kansas. It is located in northeastern Kansas next to Interstate 70, alongside the banks of the Kansas and Wakarusa Rivers. Lawrence has a diverse economy spanning education, scientific research, industrial, agricultural, finance and government. Most of these activities are tied to the University of Kansas which is the largest employer in Lawrence.

At present, the city of Lawrence has 49 miles on-road designated bike facilities i.e. bike lanes, bike routes, and marked shared lanes. The city also has 28 miles hard surface shared use paths, and 40 miles of off-road natural surface paths i.e. trails at Clinton Lake (City of Lawrence, 2015). It is noteworthy that Lawrence has been recognized as bicycle friendly community at the bronze level (the lowest level) since 2004 by the League of American Bicyclists. In 2016, Lawrence was still at the bronze level and the Lawrence still have much higher bike crashes than other cities in the Douglas County. As a college town, there is a higher propensity to use bikes in Lawrence and bike safety issues demand an empirical analysis to help reduce crash incidence (alternatively to achieve vision zero crash incidence) in the city. Therefore, this study chooses the city of Lawrence as the study area. The availability of required data for regression analysis is a plus.

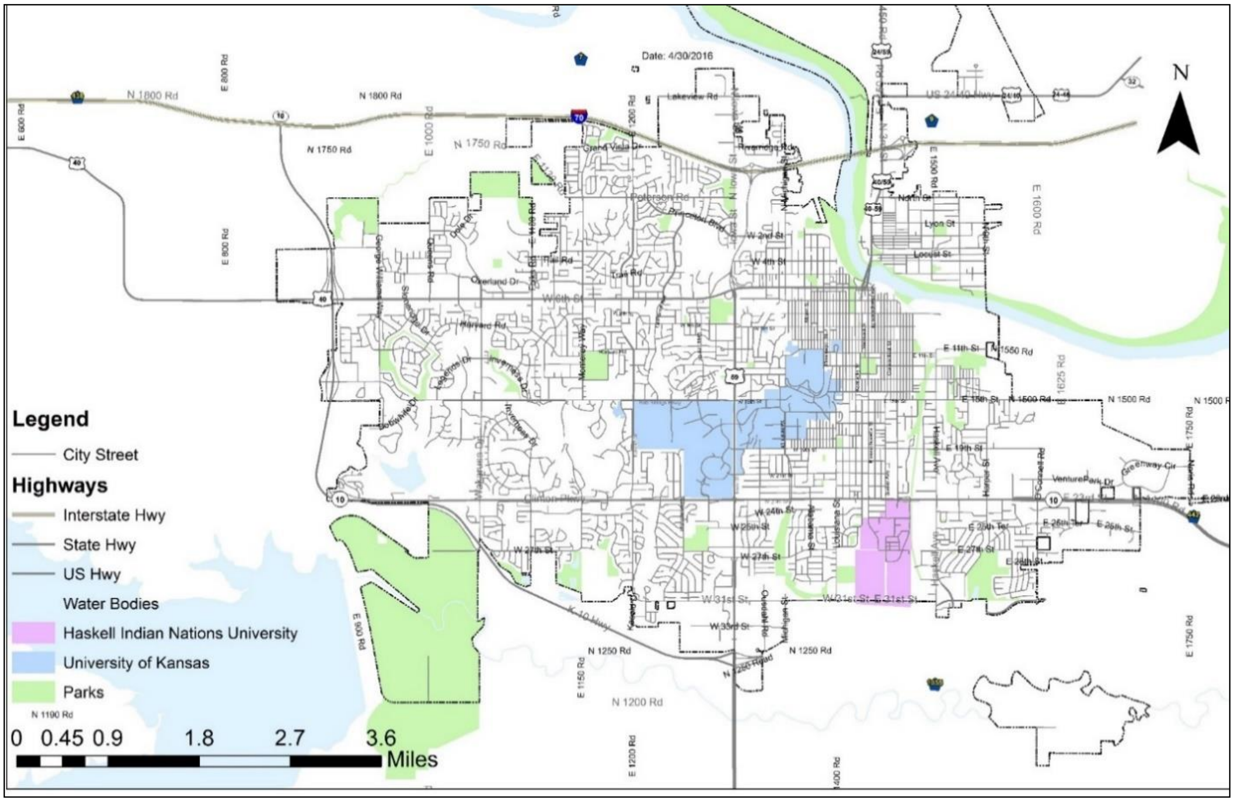


Figure 1: Map of the study area (GIS Data Source: City of Lawrence, KS)

1.5. Summary of the Chapter

The goal of the study is to identify what road geometry, infrastructure, and traffic characteristics play a key role in the bike route safety. The number of bicycle crash incidence has been used as an indicator for bicyclist's safety in this study. Table 1 summarizes the type of data required for the study, source of the data, spatial and statistical tools used to analyze the data, and the queries answered by the collected data and analysis.

Table 1: Sources of data and tools to be used in answering research question

Data	**Source**	**Tools**	**Questions to Answer**
Bike crash data	City of Lawrence	RStudio Excel ArcMap	How road geometry and traffic characteristics factors affect bike route safety?
Route slope	City of Lawrence		
Existing bike route	City of Lawrence		
Traffic volume	KDOT and City of Lawrence		
Number of lane	City of Lawrence		
Crash locations	KDOT	Spatial Analyst (ArcMap)	To identify crash zones
Slope vs Crash	City of Lawrence	RStudio / Excel	Impact on bicyclists safety
Factors of crashes		RStudio, Excel,	To analyze bike safety

The rest of the thesis is organized as is shown in Figure 2 below.

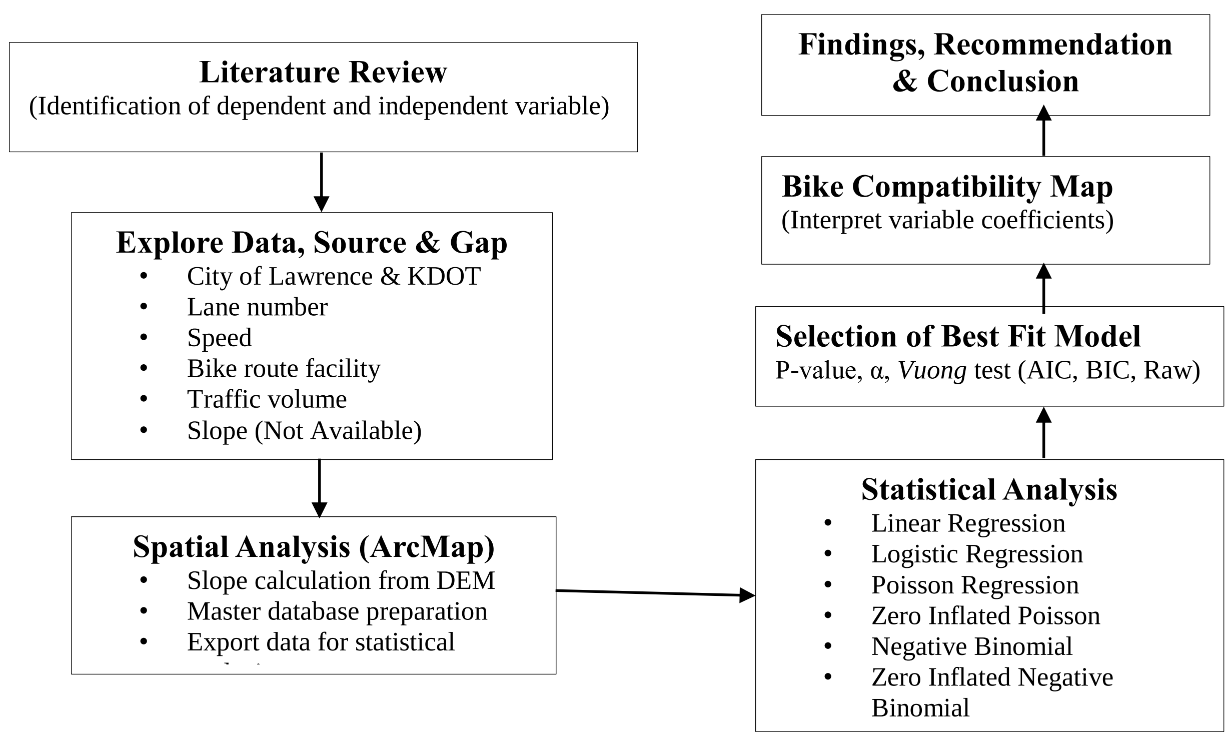


Figure 2: Workflow diagram of the study

Chapter 2: Literature Review

A literature survey has been conducted to review relevant existing studies, and the approaches taken by other researcher to solve similar or related issues. Even though bike safety is not a new phenomenon, the context and physical features of each study are different. Therefore, the literature review would help to identify what resources are available and what gaps in research exists. The identification of dependent and independent variables are the major findings of the literature survey. Once the relevant variables are identified, the next step is to check what statistical methods are available to find the relationship between independent and dependent variable. This chapter has three main components: factors that affect bike route safety, methods to conduct safety analysis, and strategies to provide bike route facilities.

2.1. Factors that Affect Bike Route Safety

The factors that affect bicycle route choice are three-fold, namely, the characteristics of roads, traffic characteristics, and environmental factors.

2.1.1. Characteristics of the Roads. For the bike route safety analysis, it is important to know the road geometry, width and traffic lane numbers, pavement type and condition, the slope of road segments. Pertinent research findings on road characteristics have been summarized as below.

Petritsch, Landis, Huang, and Challa (2006) found that a majority of bicyclists prefer to bike on streets with two lanes rather than pedaling on wider roads (with four lanes). According to these authors, the drivers tend to pay more attention to other vehicles on the wider roads than to bicyclists. This behavior increases the probability of bike crash incidence. On the other hand, Hyodo, Suzuki, and Takahashi (2000) found that usually, bicyclists plan their trips by directing them to main streets, with several traffic lanes. The reason given by the authors is that the wider roads are better known by users, which facilitates the planning of their trips.

The bad condition of the pavement or debris on a road may be a major impediment for the bicyclist to ride on it, because the lack of a suitable surface for cycling decreases the sense of security, forcing the bicyclist to choose other routes (Noland & Kunreuther, 1995). M. Stinson and Bhat (2004) found that bicyclists avoid biking on unpaved roads and prefer to use roads with the paved and smooth surface. The pavement type and condition are more important for experienced cyclists because these users are more capable in evaluating the quality of the pavement.

The presence of uphill stretches impede in the choice of the route and severe slopes are often avoided by bicyclists. M. A. Stinson and Bhat (2005) found that the tolerance with uphill stretches is directly related to the type of bicyclist. The preference for flat roads is higher among non-experienced bicyclists. But the more experienced bicyclists prefer to ride on roads with steep slopes because these roads require a greater level of physical exercise. The study did not mention what slope is a severe or steep slope. Sener, Eluru, and Bhat (2008) used three categories of slope (flat terrain, moderate slopes, and steep slope) and have drawn an interesting conclusion that bicyclists prefer routes with a moderate slope. Winters, Teschke, Grant, Setton, and Brauer (2010) claim that there is no consensus on the threshold above which the slope is considered unsuitable for biking, but the limit considered in the study was 10%. Broach, Dill, and Gliebe (2012) found that some bicyclists in Portland, Oregon were willing to go 37% longer distances on a flatter route to avoid slopes that are greater than 2%.

The existence of bike paths, bike lanes, and bike routes are of prime importance for bicyclist route choice. In addition to the existence of infrastructure for bicyclists, it is necessary that this infrastructure is continuous. Roads with uninterrupted bicycling infrastructure are much more attractive to bicyclists than roads with only a few stretches of bike lanes or bike paths. On the other hand, Dill (2009) concluded that most bicycle users will not use the biking infrastructure (even if it is very good) if this route implies a very large deviation from the shortest path between their points of origin and destination. With respect to preference for use of bike paths and lanes, Larsen and El-Geneidy (2011) found that there is no statistically significant difference between men and women in Canada, whereas Garrard et al. (2008) in Melbourne, Australia, found that the percentage of women who prefer to use cycling infrastructure is statistically higher than that of men.

2.1.2. Traffic Characteristics. Higher number of conflict points with motor vehicles, long time exposure to high-volume and high-speed road traffic result in mental stress for the bicyclists. Therefore, traffic volume is considered as a very important factor in route choice and bicycle users tend to prefer roads with low traffic volume. The vehicle flow (volume) is inversely proportional to the bike rider’s experience. Experienced bicyclists tend not to bother with the volume and speed of traffic sharing the road with them. Casello, Nour, Rewa, and Hill (2011) concluded that what really bothers bicyclists is the behavior of vehicle drivers to bicyclists instead of traffic flow. The relationship between speed, traffic volume, and the perceived risk of accidents is very important (Snizek, Nielsen, & Skov-Petersen, 2013). The availability of traffic volume data for each street is difficult, so Snizek et al. (2013) used the road hierarchy (arterial, collector or local roads) as a proxy for the volume and speed of traffic. It is a common practice in the study of bike route safety.

2.1.3. Environmental Characteristics. The perception of security (a possible risk of robbery and physical assault) is an important criterion in the bike route decision making. Sener et al. (2008) found that about 20% of the bicyclists were anxious about their personal safety while biking. Absence of streetlight in any route segment would force bicyclists to change the route while biking at night. Weather conditions can be incorporated in the bike route safety analysis. Presumably rainy or snowy weather condition are riskier as the road surface not favorable for the bicyclists.

2.2. Methods and Strategies for Bike Route Safety Analysis

This section has three major components, i.e. GIS based approach, statistical approach, and survey based (participatory) approach. Researchers have tried to explore different ways to analyze bike route safety. All three types of strategies are discussed below.

2.2.1. GIS Based Approach. Singleton and Lewis (2012) conducted a bike safety study in London using bicycle accident data. The authors used Geographic Information Systems (GIS) tool for analyzing and visualizing crash occurrences. Network routing algorithm was adopted to account for the frequency of bike accidents within a series of proposed journeys. This pilot routing application compared the quickest route with an accident avoidance weighted route between a series of origins and destinations. Finally, route safety attributes were applied.

Hu, Zhong, Cheng, and Wang (2012) conducted a study on road safety evaluation models for bicyclist in the campus of Higher Education Mega Center (HEMC) in China using GIS. In China, road accidental injuries, mostly bicyclists, to the students in HEMC are a large number. The authors have summarized key factors that impact the bicyclist safety in campus: the slope gradient, road curvature, distance to intersections and other special factors (like a rainy day) that contribute to the road risk. These factors were quantified, normalized, and finally, a comprehensive safety evaluation model (road risk distribution model) for bicyclist was developed. With the help of ArcMap software, all risk values of points in the road were interpolated through sample points by Digital Terrain Model (DTM) function, and the risk DTM was overlaid on the remote sensing image of the same area. Finally, a risk distribution map along the road for bicyclists of campus in sunny days and rainy days were created using ArcMap and Network Analyst.

Chu, Azer, Catalanotto, Ungar, and Goodman (1999) used a method to identify the high accident locations through clustering techniques that involve query by crash type, time of day and crash severity. The outcome of the study can be used to identify and rank crash locations for the most effective use of safety funds. NCCGIA (2000) used route planning and cluster analysis to identify high bike crash zones and it could be used for practical decision making in the provision of bike route facilities. Smaller search radii can be used to show crash clusters on intersections and route segments whereas larger search radii can be used to recognize large zonal clusters. The perception of bike and pedestrian crash risk can be used for crash prevention. Butchart, Kruger, and Lekoba (2000) conducted survey of households in six neighborhoods in a low-income area of Johannesburg, South Africa (Sample size, N = 1075). The authors have concluded that inadequate signage, traffic lights, and alcohol involvement were perceived as pedestrian risk factors. The preventive measures include increased law enforcement and traffic calming. The most pronounced methods to measure and rank high crash zone are individual methods such as crash frequency, crash rate, crash density, and composite method such as the sum-of-the-ranks method, which can be used to rank high pedestrian crash zones. The crash frequency (CF) is the number of reported pedestrian/bike fatal and injury crashes. All types of fatal and injury crashes are given equal weights. The crash density (CD_A) method is a strategy to rank zones based on crash frequency, CF or CF_S per length or area. The crash rate (CR) method is used to rank zones based on the CF or CF_S in relation to a measure of exposure. Typical measures of exposure are pedestrian and vehicular volume, and the population in the proximal area. The crash rate based on vehicular volumes (CR_{VV}) is the ratio of CF to the vehicular volume. The vehicular volumes are measured either as the number of vehicles crossing a point in a given time period or as vehicle-miles of travel along a segment in that period. The sum-of-the-ranks (SR) method combines the selected individual methods in the calculation of a single rank value for each zone. The Crash Score (CS) method is based on normalizing the values to the same scale to obtain a score for each method. In the sum-of-the-ranks method, crash density (CD_A), crash rate based on vehicular volume(CR_{VV}), and crash rate based on population (CR_{PP}) are used to estimate the overall crash score using the CS method. The individual scores for each method are normalized to a 0 –100 scale using the following equations:

$$\text{Score } CD_A = (CD_A / \text{maximum } CD_A) \times 100$$

$$\text{Score } CR_{VV} = (CR_{VV} / \text{maximum } CR_{VV}) \times 100$$

$$\text{Score } CR_{PP} = (CR_{PP} / \text{maximum } CR_{PP}) \times 100$$ (Butchart et al., 2000)

The highest score for a method is equal to 100. The individual scores for each method are then summed to estimate the crash score for the zone:

$$CS = \text{Score } CD_A + \text{Score } CR_{VV} + \text{Score } CR_{PP}$$ (Butchart et al., 2000)

2.2.2. Statistical Approach. Mapping the risk of collision between bicyclists and motor vehicles in urban areas is important to provide safe bicycle route opportunities. It might be helpful to make an informed transportation planning decisions and explore patterns of injury epidemiology. Yiannakoulias, Bennet, and Scott (2012) stated that bicycle crash risk can be measured by four parameters e.g. Crash frequencies, crash per capita, crash rates per bicyclist, and crash rates per distance traveled. The relative risk of bicyclist collisions can be expressed as

$$\theta_i = (y_i / e_i)$$ (Yiannakoulias et al., 2012)

where y_i is the observed and e_i is the expected number of collisions in census tract i. The expected number of collisions is equal to the total bicyclist kilometers travelled in tract i multiplied by the study area rate of collisions (total collisions divided by total kilometers travelled). Given i, the number of collisions y_i in the census tract can be treated as an independent Poisson process (Yiannakoulias et al., 2012). Using the shortest distance route choice algorithm, the study estimated the total distance traveled by all commuter bicyclists in Hamilton, Canada. This measure can be used to represent the underlying geography of biking risk more realistically, and provide more geographically and empirically meaningful information to those interested in understanding how bike safety varies over space. Moreover, the understanding the geography of biking collision risk could also be important for improving the understandings of spatial inequities of biking risk, and linking these inequities to features of the social environment (Yiannakoulias et al., 2012).

Bike crash data can be investigated to understand the bike route safety using regression models. The type and distribution of data define which regression model to apply. Because each model has its underlying assumptions and data need to verify those assumption criteria for the model to apply. For instance, the data need to be normally distributed to for a linear regression model to apply. If the data contains lots of zero and has over-dispersion problems, then generalized linear models or zero inflated models would be a better fit.

2.2.2.1. Generalized Linear Model (GLM). Over-dispersion (variance is much larger than mean) and excess zeros are two problems that typically occur in count data sets in economics and the social sciences. Poisson regression model (quasi-Poisson model) and negative binomial regression model are very useful tool to deal with overly dispersed data. Both models belong to generalized linear models (GLM). On the other hand, zero-inflated Poisson regression is conducive to model count data that have an excess of zero counts. (Zeileis, Kleiber, & Jackman, 2008). Three generalized linear models are commonly used for count data analysis, i.e. Poisson, quasi-Poisson and negative binomial model.

Poisson distribution is the simplest method and commonly used for modeling count data with probability density function f(y;μ) = exp(−μ)·μy y!. It is a special case of the generalized linear model framework. The canonical link is g(μ) = log(μ) resulting in a log-linear relationship between linear predictor and mean. In the Poisson model, the variance is identical to the mean, thus the dispersion is fixed at φ = 1 and the variance function is V (μ) = μ (Zeileis et al., 2008). In R, it is specified in the *glm()* call just by setting family = Poisson. While describing the mean using Poisson model it is important not to underestimate the variance of the data. (Zeileis et al., 2008).

Another way of dealing with over-dispersion is to use the mean regression function and the variance function from the Poisson Generalized Linear Models (GLM). In this case the dispersion parameter φ unrestricted and φ is not assumed to be fixed at 1 rather it is estimated from the data. This strategy leads to the same coefficient estimates as the standard Poisson model but inference is adjusted for over-dispersion. Consequently, both models (quasi-Poisson and sandwich-adjusted Poisson) adopt the estimating function view of the Poisson model and do not correspond to models with fully specified likelihoods. In R, the quasi-Poisson model with estimated dispersion parameter can also be fitted with the *glm() function,* simply setting family = quasi-Poisson. (Zeileis et al., 2008)

The third way of modeling over-dispersed count data is to assume a negative binomial (NB) distribution. Package MASS provides the family function *negative.binomial()* that can directly be plugged into *glm()* provided the argument theta is specified (θ = 1).

2.2.2.2. Hurdle Model. Some data sets contain more zero observations than that would be allowed for the Poisson model. Hurdle Model can capture both over-dispersion and zero inflation phenomenon. Hurdle model has two components: A truncated count component, such as Poisson, geometric or negative binomial, is employed for positive counts, and a hurdle component models zero vs. larger counts. In statistical package R, the *hurdle() function* from the *pscl* package is used to fit count data. The arguments of *hurdle()* are given as follows (Zeileis et al., 2008).

hurdle(formula, data, subset, na.action, weights, offset, dist = "poisson", zero.dist = "binomial", link = "logit", control = hurdle.control(...), model = TRUE, y = TRUE, x = FALSE, ...

2.2.2.3. Zero-Inflated Model. Zero-inflated models can deal with excess zero counts in the data. These models are two-component mixture combining a point mass at zero with a count distribution such as Poisson or negative binomial. Thus, zeros might come from two different sources i.e. from both the point mass and the count component. For modeling the unobserved state (zero vs. count), a binary model is used. In statistical package R, *zeroinfl() function* from the *pscl* package is used to fit zero inflated count data model. The arguments of *zeroinfl()* are given as follows.

zeroinfl(formula, data, subset, na.action, weights, offset, dist = "poisson", link = "logit", control = zeroinfl.control(...), model = TRUE, y = TRUE, x = FALSE, ...)

Where all arguments have almost the same meaning as for *hurdle()*. The main difference with hurdle model is the absence of *zero.dist argument.* In the *zero.dist function* of hurdle model, a binomial model is used for distribution in the zero-inflation component. (Zeileis et al., 2008).

In short, generalized linear model, hurdle model, and zero inflated model are very useful to deal with wide range of data. Poisson regression model is very useful to model the count data when the variance of the data is equal to the mean. If data are characterized with over-dispersion, then zero inflated model are useful i.e. zero inflated Poisson, zero inflated negative binomial. This section also summarizes the R code to conduct these regression model in statistical package - Rstudio. Table 2 summarizes statistical model, its context of application with variable data type.

Table 2: Summary of statistical models, and their applications

Model	**Application**	**Dependent Variable**	**Independent Variable**
Linear regression	Description of a linear relationship	Continuous	Continuous and/or categorical
Logistic regression	Prediction of the probability of belonging to groups (outcome: yes/no)	Dichotomous (success of treatment: yes/no)	
Poisson regression	Modeling of counting processes	Count data	
Zero Inflated Poisson (ZIP)	Modeling of Count data with lots of zeros	Count data	
Zero Inflated Negative Binomial (ZINB)	Modeling of count data where data are over dispersed (Variance > Mean)	Count Data	

2.2.3. Survey Based (Participatory) Approach. Primary data analysis is a powerful tool to analyze any real world problems and conducting surveys is a common way to collect primary data. The direct input from the stakeholders (e.g. bicyclist) makes the plan more applicable and user-friendly. Survey can be conducted at KU campus, downtown, or bicyclists on the bus to get the untold stories about the challenges of riding bikes in Lawrence. These inputs might address many issues that the city of Lawrence is currently unaware of.

2.3. Strategies to Providing Bicycle Facility

This section has three components i.e. factor to consider in providing bike route facilities, provision of public bike system, and geological and cultural context for providing bike route facilities.

2.3.1. Factors to consider in the provision of bike route facility The selection of a bicycle facility may depend on many factors, including vehicular and bicycle traffic characteristics, adjacent land use and expected growth patterns. Table 3 lists some of these factors to consider. Safe, comfortable and well-designed facilities are essential to encourage bicycle use (AASHTO Executive Committee, 1999). Limitations of each type of facility must also be considered. Obstructions and impediments to bicycle travel include incompatible grates, debris, shoulder rumble strips, narrow lanes, driveways, rough pavements, curbside auto parking, bridge expansion joints, railroad tracks, poor sight distance and traffic signals that are not responsive to bicycles (AASHTO Executive Committee, 1999).

Table 3: Factors to consider in selection of bicycle facilities

Skill level of users	Barriers	Directness
Motor vehicle parking	Crash reduction	Accessibility
Aesthetics	Conflicts(Between motor vehicle and cyclist)	Pavement surface quality
Personal Safety/Security		Truck and bus traffic
Stops	Maintenance	Traffic volumes and speed
Bridges	Intersection corridors	Costs/ funding

Source: (AASHTO Executive Committee, 1999)

In an effort to become ‘Bike Friendly’, it is best to find the required and desired features and amenities that need to be included in the development of bicycle facilities. Table 4 shows results from a bicycle tourism study from Taiwan that illustrated numerous factors bicyclists found important in selecting routes (Chang & Chang, 2005). Factors like bike path, low traffic volume, flat terrain, pavement quality, signage and interpretation are not only important for recreational purpose but also affect the safety of the bicyclists. Some of these factors have been considered in this study as well.

Table 4: Order of importance of factors in the bicycling facility for recreational cyclists.

Order of importance	Factors	Order of importance	Factors
1	Safety	12	Weather and climate
2	Low flow of traffic	13	Bike rental provided
3	Bike path	14	Scenery and greenery
4	Restroom	15	Bike route length long enough
5	Tourist attraction	16	Café and restaurant
6	Bicycle map	17	Touring activity of cyclists
7	Rest place	18	Flat terrain preferred
8	Signage and interpretation	19	Friendly residents
9	Pavement quality	20	Convenient store
10	Racks and locker provided	21	Accommodation
11	Challenging terrain preferred	**Source**: (Chang & Chang, 2005)	

2.3.2. Provision of Public Bicycle System. Bike sharing is becoming very popular in the North American cities. The City of Lawrence, KS is no exception of this trend. The city has been conducting bike sharing feasibility study currently and attempting to connect with the public transit system. Historically, the bicycle is very convenient and efficient for last mile travel. Yang, Lin, and Chang (2010) proposed public bicycle system to connect major origins and destinations with some fixed pickup and drop off points (see Figure 3). For shorter trips, it could be very pragmatic solution and the city of Lawrence, as a college town, could possibly adapt this idea to promote bicycle in the town among the high school, college, and even young professionals. Undoubtedly, this would create a more engaged and lively community in the town.

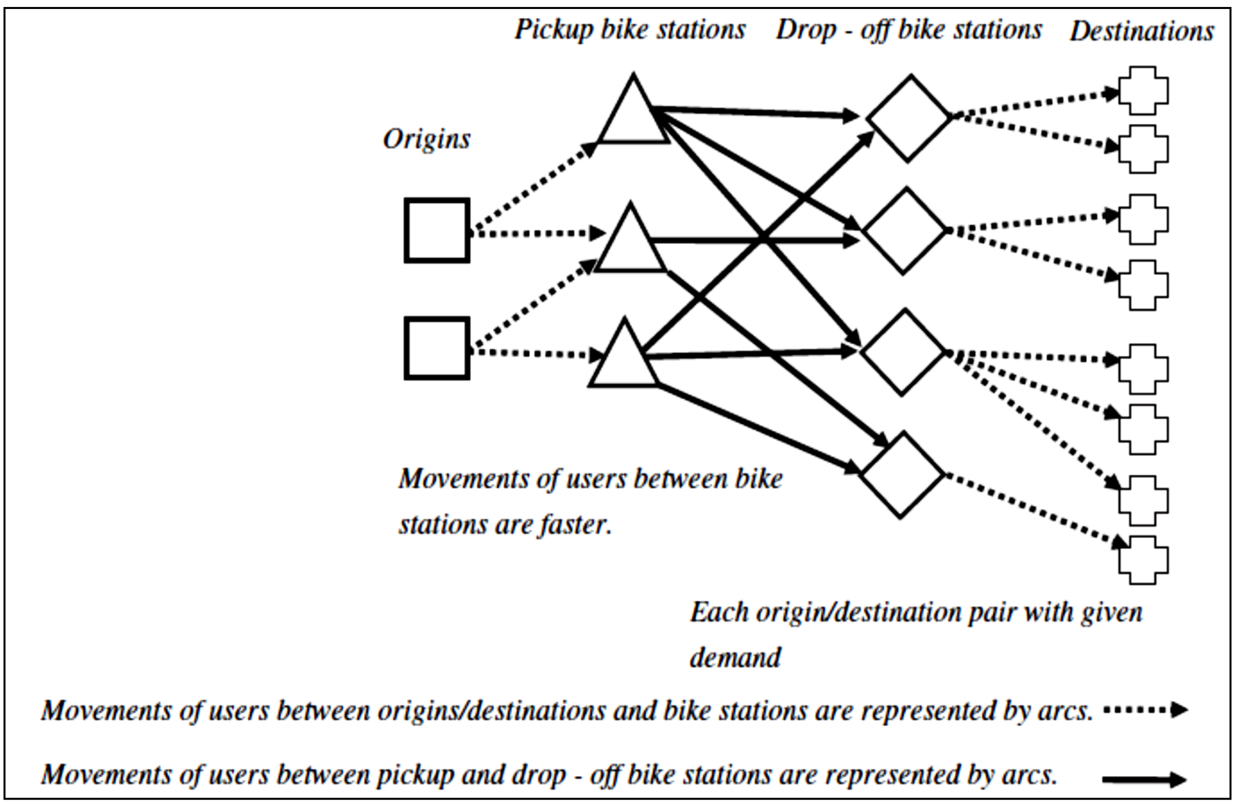


Figure 3: Network structure of public bicycle system. Source: Yang et al. (2010)

2.3.3. Transit Oriented Development and Bike Sharing. Historically, western European countries e.g. Denmark, Germany have better transit and bicycle friendly infrastructure. The higher population density, mixed land use pattern (which generates shorter trips within city limits), conducive policies have promoted bicycling in those countries. In addition to good bike infrastructure, the people in those geographical areas i.e. the Netherlands, Denmark love bicycling, which is partly cultural phenomenon. However, the land use and density of North American cities are quite different from European cities. Even though USA and Canada have similar climatic and cultural context, Canada has a higher bike share of urban trips than the United States. There exists a misconception that cold weather might restrict the bicycling, but Pucher and Buehler (2006) concluded that cool climate does not prevent biking and a warm climate does not necessarily assure it. For instance, Yukon Territory- with roughly the same latitude as Alaska—has a bike share of work trips more than twice as high as California's (2.0 vs. 0.8%) and more than three times as high as Florida's (0.6%) (Pucher & Buehler, 2006). The study also examined factors like biking safety, land use patterns, car ownership rates, and costs of car use, per capita income, climate, and cultural differences to find bicycling trends in USA and Canada. The study also urged transit oriented development policies and the integration of bicycle with transit network to promote bicycling and enhance its safety (Pucher & Buehler, 2006). The city of Lawrence is conducting a feasibility study for adopting bike sharing strategy and it would probably be an efficient last mile solution. Better connectivity with bus stop would help bicyclists to avoid very steep and risky routes as well.

2.4. Conceptual Framework

A conceptual framework illustrates which variables contribute to the bike crash incidence. As shown in Figure 4, traffic characteristics, road geometry, infrastructure, environmental characteristics, and characteristics of drivers and bicyclists have an impact on bike crash incidence. Neither the city of Lawrence nor the Kansas Department of Transportation have all these data available. So, the lane number, traffic volume, speed, and bike route facility have been selected as independent variable for the bike route safety analysis in Lawrence. Street-wise slope data are not available, but it has been calculated from Digital Elevation Model (DEM) data using ArcMap.

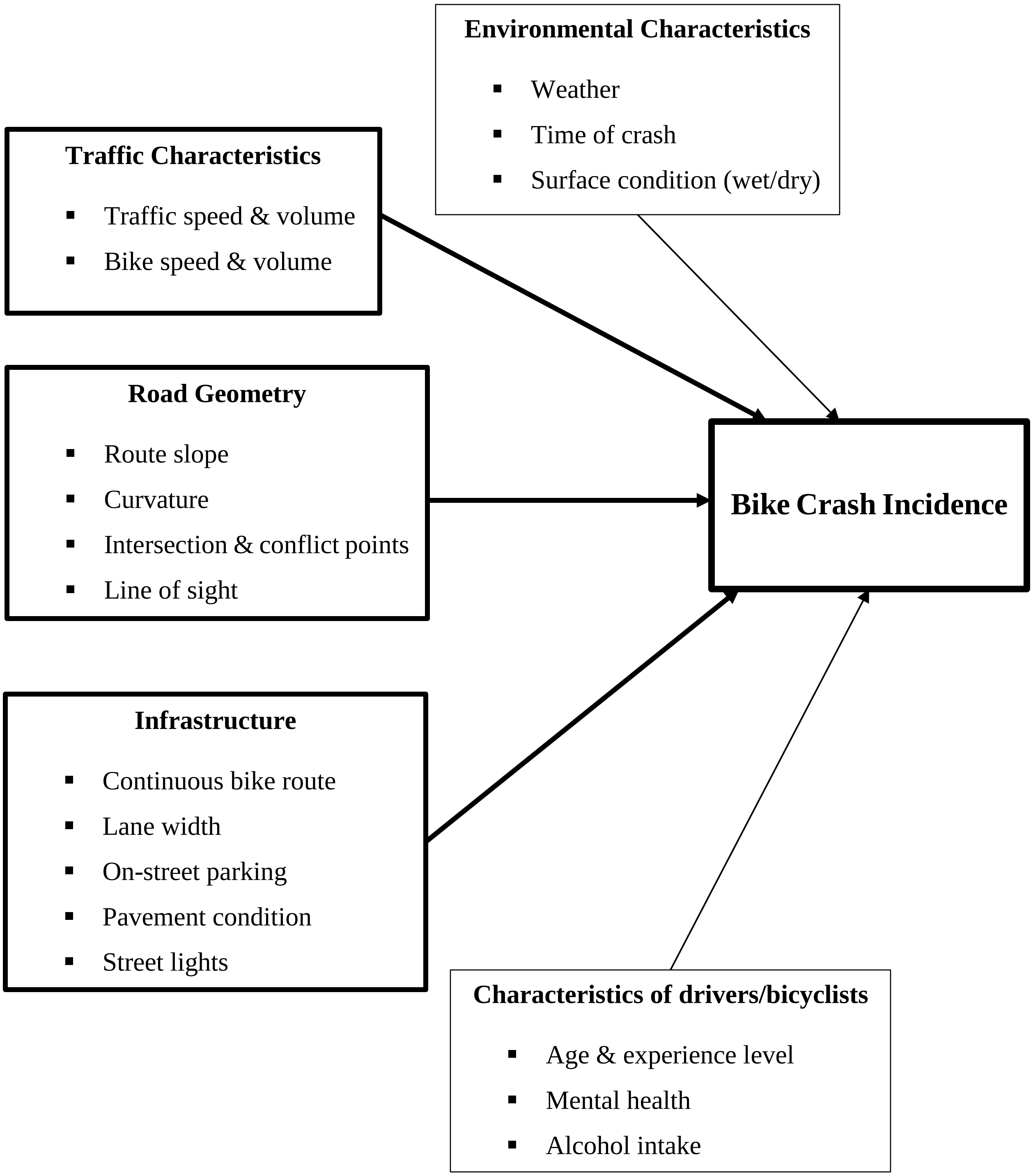


Figure 4: Conceptual framework of the study

Addition of some primary data i.e. actual bike count and speed data, on-street parking etc. in the model could probably increase the predictability. As a large number of bike incidence occurred close to intersections, proximity to intersections could be considered as an independent variable. Variables that are very significant but not available are actual bike count and speed, characteristics of bicyclists/drivers, on-street parking, line of sight etc.

2.5. Summary of Literature Review

The main road geometry, infrastructure, and traffic characteristics that influence bike safety are roadway width, number of lanes, pavement type, and condition, slope of the route, the presence of bike routes, traffic volume and speed. Actual bicyclists speed is important to understand maneuvering related crash incidence, and bicyclists movement directions in steep slope helps to measure the crash severity. The presence of trucks on the road and left turn at the intersection also affect the safety of bicyclists. The City of Lawrence does not have all these data and the collection of certain data for each street are very expensive for the city e.g. bicyclists volume and speed in city streets etc. Therefore, it is important to find patterns from the available data using advanced statistical methods. Previous study on bike safety analysis using crash data is non-existent. Some studies used bicycle user ratings as a measure of safety and applied statistical tools to find correlation between variables. But none of those were dealing with zero inflated data. This study explores how to deal with zero inflated data in bike safety analysis. Bike crash data have been used as dependent variable and the independent variables are route slope, lane number, traffic volume, speed, the availability of bike route facilities.

Chapter 3: Methodology

Literature review in Chapter 2 helped to narrow down the independent and dependent variables for the study. This chapter will describe the steps that have been followed to investigate the impacts of road geometry, infrastructure, and traffic characteristics on bike route safety. Contents of this chapter include the discussion of underlying assumptions, model structure and its components for all the regression models that have been applied in this study. It is important to validate the methodology, check the applicability of these models for the collected data set. Exploration of the distribution of independent and dependent variable data is a prerequisite for data analysis. The details of the study area and existing bike route facilities has been discussed in the chapter 1 (subsection 1.4). At this point, data type, statistics of the dependent and independent variables have been discussed in the following section.

3.1 Exploring Data, Type, and Collection Method

The literature review shows that the presence of bike route facilities, route slope, speed, traffic volume, number of lanes play a vital role in describing the safety issues for bicyclist. It is generally believed that the presence of bike routes helps to reduce the bike crash incidence and lower speed with a smaller number of lanes are less prone to bike crash. The steepness of slope is very subjective and less experienced bicyclists tend to ride on relatively flat route. These are all predicted variables for bike safety analysis. On the other hand, the bike crash incidence is a measure of safety, because it reflects how safe a bicyclist is at any route segment. Therefore, the bike crash incidence has been considered as response variable for this study. These data are available in the City of Lawrence and others are extracted from collected data using ArcMap software.

Table 5: Exploring the type of data and its applications

Data	Applications (Remarks)
Street Centerline	Speed limit, traffic flow direction, and route surface characteristics
Digital Elevation Model (DEM) data for Lawrence	To calculate slope of each route segment
Bikeway System T2040 Plan	Existing and future bike infrastructure plan, facility type and width of bike infrastructure i.e. width of bike lane, route, shared use path etc.
Traffic Volume	24 hours' traffic volume data, AM peak, and PM peak traffic volume data.
Number of lane	An independent variable to develop the compatibility index
Road Hierarchy	Three major road classes' i.e. primary, secondary and local roads. It is a proxy for traffic volume data and the perception of safety that influence the route choice by bicyclists.

Data source: City of Lawrence, KS.

3.2.1 Street Centerline. The street shapefile is downloaded from City of Lawrence website. The shapefile has some important information i.e. posted speed limit for each route, road surface characteristics, traffic flow direction (one way or both way) and segments length etc. It is good to know the definition of route segments in the city of Lawrence. Section 2.3.2 of the Kansas 911 GIS Data model (http://kansasgis.org/initiatives/NG911/index.cfm) discusses road center line creation in relation to address ranges. A segment will be split if it crosses a city limit or has a street name change. At Douglas County & the City of Lawrence, road segments were split where address ranges change. For example, in Lawrence on W 22nd Terrace between *Ousdahl* Road and Naismith Drive, the street is made up of 4 different segments. Each segment represents the address range within a "100" block. From Naismith Drive, the address range begins as the 1300 block and moving west towards *Ousdahl* the next segments represent the 1400, 1500 and 1600 block respectively. This split of route segment, in fact, provides more details for route level of bike analysis.

3.2.2 Streetwise slope calculation. The Digital Elevation Model (DEM) data with cell size of 3 feet has been collected from the City of Lawrence. The slope has been calculated using surface tool under spatial analyst toolbox of ArcGIS. In order to calculate the level of service for each road segment, streetwise slope is measured, which is considered as one of the major determining factors of bicyclists' route choice. Streetwise slope has been calculated using 'Add surface information' tool under '3D Analyst' toolbox of ArcMap. The average slope of each route segment has been calculated using the bilinear method. Bilinear is an interpolation method for raster data to determine a value from the values of the four nearest cells.

3.2.3 Traffic volume. The traffic volume of a route segment plays an important role in determining its compatibility for the bicyclists. The greater the traffic volume, the higher the risk of potential bicycle crash incidence. It is a more concerning issue for the inexperienced bicyclists than experienced bikers. However, it is an important variable, in general, for bicycle compatibility of any route. The City of Lawrence and Kansas Department of Kansas (KDOT) collect traffic volume data at some designated points in the city streets. It is very challenging to get traffic volume for each route segment in Lawrence. So, the road hierarchy (functional class) has been used as proxy of traffic volume for each route segments. For example, a principal arterial would have more traffic volume than a collector, a collector would have more traffic than a city street etc. The functional class of roads have been converted to factor variable and considered as a proxy for volume data. In this study, freeways, ramp, turnaround, and private road segments have been ignored in data analysis. So, there are three categories left under functional road class of the data and they are street, collector (Major/just collector) and arterial (Major & minor).

3.2.4 Number of Lanes and Traffic Flow Direction. The number of traffic lanes in a route segment has a strong influence over bicyclists' route choice. The bicyclists generally prefer fewer (two) lane routes and the motorists tend to pay more attention to other larger vehicles than to bicyclists in the wider routes (more lanes). So, the number of traffic lanes and flow direction (one way or two way) are very important variables to analyze the bicyclists comfort level and safe features. These data have been collected from City of Lawrence, but the number of lanes data had some accuracy problem and lots of missing data. Therefore, the data have been cross-checked using google maps as background in the ArcMap.

3.2.5 Existing Bike Infrastructure. It is generally believed that the presence of bike infrastructure (bike lane, bike route, shared use path, shared lane marking) reduces the bike crash incidence which, in turns, increase the safety features for bicyclists. Bike lanes are on-road lanes marked with paint, dedicated to cycling and typically excluding all motorized traffic. A shared use path supports multiple modes, such as walking, bicycling, inline skating and people in wheelchairs. Bicyclists share route with motorists in bike route and routes with shared lane marking. The bike infrastructure data for city of Lawrence have been collected from 'Bikeway System T2040' which was prepared by City of Lawrence. For the regression model, bike infrastructure variable has been converted as dichotomous (binary) variable which consists of 0 (no bike infrastructure) and 1 (have bike infrastructure).

3.2.6 Bike Crash Data. Bike crash incidence reflects the safety and comfort level of bicyclists in any route segments. It can be used as response variable to understand the route safety for bicyclists. The city stores each crash incidence data with latitude and longitude that helps to project the crash location on the map using ArcMap. The crash locations have been joined to the city street shapefile using 'spatial join' tool of ArcMap. The tool helps to add total number of crash incidence for each route segment in Lawrence. It sums up all crashes that intersect or closest to the street centerline.

3.2.7 Sample Size. The sample size is an important indicator for conducting statistical analysis. In any empirical study, the sample size affects the precision of inferences made about the population. In practice, it is very expensive and time-consuming to work with whole population; therefore, the selection of an appropriate sample size determines the result of the analysis. The smaller sample size might result in a biased answer and some regression model e.g. Poisson cannot produce meaningful results with small sample sizes. Due to inappropriate sample sizes, there might be two types of errors during hypothesis testing. Type I error (false-positive) occurs when the null hypothesis is rejected even though it is true in the population. Type II error (false-negative) arises if null hypothesis is accepted but it is not true in the population.

The street centerline shapefile which was collected from City of Lawrence has 5773 route segments within the city boundary. The functional class of the street shapefiles has 13 identical categories and 5 of these route classes have been ignored in this study. It is inappropriate to provide bike facilities in freeways, freeways ramp, frontage, other ramp, and turnaround. Thus, total sample street segments have declined to 5081 which is a large enough sample for conducting any statistical analysis.

3.2.8 Statistics of the Model Variables

Table 6 summarizes and explains the dependent and independent variables considered in the study. The data type and summary statistics are also included in Table 6 to understand the pattern of the data. Road hierarchy has been used as proxy of traffic volume. It is assumed that arterial roads would have a higher traffic volume than collector roads. Volume is used as categorical variable in the model.

Table 6: Summary statistics of dependent and independent variables

Variables	**Data Type**	**Summary Statistics**	**Remarks**
Crash Incidence	Count	Mean: 0.054 Range: 0 to 8	Dependent Variable
Number of Lane	Count	Mean: 2.29 Range: 1 to 7	Independent variable
Volume (Dummy)	Categorical	Arterial:808 Collector: 896 Street: 3377	Independent variable
Presence of bike route	Dichotomous (Binary)	No: 4041 Yes: 1040	Independent variable
Traffic speed	Count	<30 mph: 472 30-45mph: 4592 >45 mph: 17	Independent variable
Slope	Continuous	$< 1.5^0$: 1776 $1.5 - 4^0$: 2606 $>4^0$: 699 Average slope: 2.45^0 Below average route segments: 3332	Independent variable

3.2 Data Analysis

The criteria and underlying assumptions to choose a statistical model have been discussed in this section. Depending on the type and distribution of the data, multiple regression methods have been tested to find a model that describes the data best. Before initiating data analysis, it is important to discuss the selection of appropriate regression model. The model selection depends on the data type of dependent and independent variable. For example, linear regression model requires continuous dependent variable, logistic regression model is applicable for dichotomous dependent variable, and the count data requires Poisson regression, zero-inflated Poisson regression, zero-inflated negative binomial regression model. The coefficients of variables, R squared value, p-value play important role in the selection of best model.

The better the model the lower the error between predicted and actual values. To find the best model a sequential statistical data analysis has been conducted. The variables have both continuous and categorical data. The sample size is 5081 and the data are characterized with lots of zeros. Next, we investigate if there is any linear relationship between dependent variable (Crash incidence) and independent variables (speed, slope, presence of bike routes, traffic volume and number of lanes).

3.2.1. Phase I: Linear Regression Model. In a linear regression model, Ordinary Least Square (OLS) method is used to find linear relationships between independent and dependent variables. The underlying assumptions of linear regression include data linearity, normality, independence of error and equality of variance.

According to Figure 5, bike crash data are not normally distributed and are positively skewed and do not follow any pattern. Over 4000 route segments do not have bike route facility (it is dichotomous data) and slope data are positively skewed as well. The majority of the route segments have a speed limit of less than 30 mph and two lanes. It is evident from Figure 5 that data have violated the underlying assumptions of linear regression model. Thus, the ordinary least square method would result in significant error in the estimation. Next, a logistic regression model is tested as it can handle nonlinear relationship in the data.

3.2.2 Phase II: Logistic Regression Model. Instead of linear relationships between the dependent and independent variable, logistic regression model works better for nonlinear relationships among variables. In this case, the dependent variable should be binary data (yes =1,

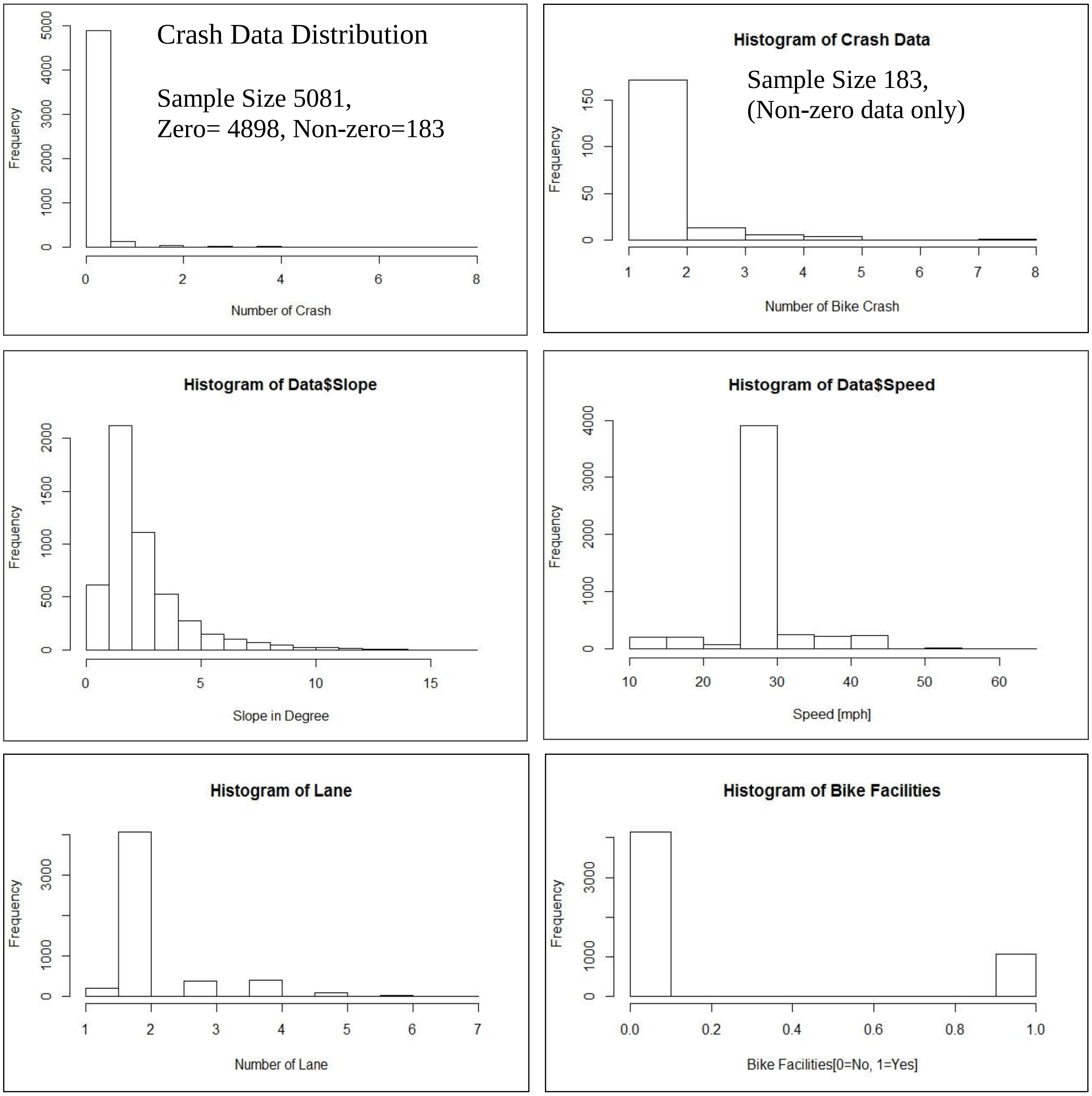


Figure 5: Histogram distribution of dependent and independent variables

no=0) and the independent variables would help to predict whether there will be crash incidence in a route segment or not. The logistic regression makes no assumption about the distribution of the independent variables. Data do not have to be normally distributed, linearly related or of equal variance within each group. Also, the logistic model can handle any large or small values of independent variable with its link function (log odds transformation or logit) and the logit is the natural parameter of binomial distribution. Noteworthy to mention is that the dependent variable has been converted into binary data (1= Route segments have one or more crash incidence and 0= No crash incidence). The outcome of the logistic regression model did not produce any meaningful coefficients of the independent variables. Additionally, the logistic model outcome is binary in nature and the bike crash data are not necessarily binary. The crash incidence data can be considered as count data because crash data are either zero or any positive number. The Poisson regression model is very useful to analyze the count data.

3.2.3 Phase III: Poisson Regression Model. The Poisson model is the best known generalized linear model for count data and the model typically uses log of the mean. Count data represent all nonnegative integer values i.e. 0,1,2,3.. etc. The Poisson distribution assumes that the variance of the dependent variable would be same as the mean. Poisson regression model has been applied for the variables in this study and the coefficients of the independent variables i.e. speed, slope, lane, and volume, are statistically significant at 99.9% confidence interval (Appendix A). But the interpretation of the coefficients is misleading because the bike crash data show greater variability than the Poisson model allows. For bike crash data, the variance (0.1236) is much larger than mean (0.0562) which implies data have an over-dispersion issue. This is a clear violation of the underlying assumptions of Poisson regression model. The possible reasons for over-dispersion could be heterogeneity of independent variables or lots of zero in the data. To overcome this issue a Zero-inflated Poisson model (ZIP) or a negative binomial model can be applied.

3.2.4 Phase IV: Zero-inflated Poisson Model (ZIP). ZIP is a useful statistical tool for modeling data with excess zeros. Zero-Inflated Poisson (ZIP) model has two components i.e. count model and logit model for predicting excess zeros. The first process is governed by a Poisson distribution that generates counts, some of which may be zero and logistic part generates structural zeros. ZIP model has been run for bike safety dataset and it shows a significant improvement from conventional Poisson model (model results are attached in Appendix A).

3.2.5 Phase V: Negative Binomial. The negative binomial model can deal with overly dispersed count data and it is very useful for discrete data over an unbounded positive range whose sample variance exceeds the sample mean. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean. Having lots of zero in the data does not necessarily mean that zero inflated model is the sole solution. Depending on the distribution and type of data, negative binomial could be a better solution than zero inflated negative binomial model. The output of the negative binomial model shows that the independent variables are strongly related with the dependent variable (the results are attached in the appendix A).

3.2.6 Phase VI: Zero-inflated Negative Binomial (ZINB). The ZINB is considered as the upgraded version of negative binomial. Sometime but not always ZINB shows better fit to the data than that of conventional negative binomial model. Thus, ZINB model has been run to see if there is any significant improvement from the negative binomial model in terms of meaningful regression coefficients and fit to the data. The ZINB model result and R code have been attached in the appendix A.

3.3 Best Regression Model Selection

Data predictability, p-value, and significance level play key role in the selection of best regression model. In this study, linear regression model, logistic model, and Poisson model are not the best fitted model because bike crash data have violated the assumptions of these regression models. So, possible candidates for the best fitted model are negative binomial, zero inflated Poisson and zero inflated negative binomial model.

Desmarais and Harden (2013) stated that *Vuong* test is useful to determine whether the zero-inflated model fits the data significantly better than count regression with a single equation. For instance, comparisons of zero-inflated count models with their non-zero-inflated analogs (e.g., zero-inflated negative-binomial versus negative-binomial, or zero-inflated Poisson versus conventional Poisson). Vuong test outcomes have three test statistics i.e. Raw, AIC corrected and BIC corrected. Each test statistic (Raw, AIC corrected and BIC corrected) performs similarly well for zero inflated data. The raw difference between actual and predicted counts for each outcome can be used as a model fit indicator. Akaike's information criterion (AIC) balances model's goodness-of-fit to the data and the model with a smaller AIC value is considered better fit to the data. Lastly, the Bayesian information criterion (BIC) is very useful to select the correct model when the sample size is large and the smaller the number better the model. Considering all these factors, ZINB model has been selected as best fitted model and the coefficients of ZINB model has been used to explore the impacts of road geometry and traffic characteristics on bike route safety. The coefficients of independent variable in the negative binomial (NB) model are also statistically significant with the dependent variable. A comparative analysis and interpretations of the ZINB and NB regression coefficients are discussed in section 4.4.

3.4 Impacts of Road Geometry Factors to Bike Route Safety

The coefficients of the zero-inflated negative binomial model are log coefficients, so these are converted into a number using the exponentials of those log coefficients. If the values are less than 1 it means the independent variables would negatively impact the dependent variable, and a greater than 1 value means positive impact on dependent variable. All independent variables have been interpreted to find the correlation of these variables with bike safety in the city of Lawrence.

In order to identify the bike crash zone in Lawrence, the kernel density tool of ArcMap has been used to generate bike crash density map showing low to high bike crash prone areas in the city (Figure 8). To visualize the relationship between the existing bike route infrastructure and crash incidence, the bike route shapefile is superimposed on top of bike crash zone map layer. This map (Figure 10) would help to understand whether the presence of bike route infrastructure help to reduce the crash incidence or not. Moreover, city planner would get an idea where to build new bike route infrastructure in the city. Similarly, Figure 9 shows the spatial relationship between route slope and crash incidence in Lawrence.

Finally, the bike compatibility map has been developed using the (ZINB) regression coefficients of independent variables i.e. slope, lane numbers, speed, existing bike facilities etc. The regression coefficients were multiplied by the corresponding variables and then added with intercept to get the index value for each route segment in Lawrence. This index value has been classified by Jenks natural break method in ArcMap. Natural break method identifies the break points for each class in a way that minimizes variance within the class and maximizes the difference between classes. These classes are based on the natural groupings that is inherent in the data. The bike compatibility map for the City of Lawrence (Figure 11) has been prepared and this outcome will be compared with existing city's bike rideability map (Figure 12) which was prepared by user experience ratings. The similarities and differences of these two maps would be analyzed with justifications.

3.5 Expert Interview

Two expert interviews have been conducted to get insight on bike route safety. City engineer of Lawrence and senior transportation planner from Mid-America Regional Council (MARC) have gone through the bike compatibility map and regression outcomes. Both have made comments and suggested some measures to improve the model results.

3.6 Summary of the Methodology

After exploring the type of dependent and independent variable data, it is revealed that the data are characterized by over-dispersion and zero inflation. The independent variables are not linearly related with dependent variable. Therefore, several models have been applied and analyzed to find the best representation of the data. The logistic regression model improved the explanatory power of the model from linear regression as it is not bounded by assumptions of equal variance or normality. But the dependent variable is binary in logistic model and the crash incidence could be better represented as count data because the number of crash incidence for a street segment could be more than one and crash number is not necessarily always binary data. Therefore, generalized linear models have been applied to improve the model output. Firstly, the Poisson model has been applied to explore the model improvement from logistic regression and the result shows that R^2 value increased a bit. But the underlying assumption about the equality of mean and variance has made the outcome questionable as the data have an over-dispersion problem. In order to resolve this issue, zero inflated Poisson model and negative binomial model have been applied. Both these models have showed improvement from the Poisson model. If the data have lots of zero then zero inflated model usually improves the model result. To test that hypothesis, zero inflated negative binomial model has been applied to compare the result with negative binomial model. It is found that zero inflated negative binomial model produces significantly better result than negative binomial. The independent variables are strongly correlated with dependent variable and the coefficients are significant as well. To validate this result, *Vuong* test has been conducted and found that zero inflated negative binomial model (ZINB) shows significant improvement from negative binomial model. So, the coefficients of each independent variable in the ZINB model have been used to explain the relationship with the dependent variable. As the coefficients of ZINB model are log based, the exponentials of the coefficients have been used to quantify how one-unit change of independent variable would affect the dependent variable by keeping all other variables constant. Finally, the bike compatibility map has been generated using the ZINB regression coefficients and natural break (Jenks) classification method was used to create the class intervals for the map. The outcomes of this compatibility map have been compared with the city's bike rideability map and the analysis includes both similarities and differences. Chapter 4 (Experimental Results) would analyze and summarize negative binomial and zero inflated negative binomial model results. It will also synthesise each model result and explain the contribution of each variables to the bike route safety in Lawrence, KS.

Chapter 4: Experimental Results

The goal of this chapter is to explore and synthesize the regression model outputs. The results would explain how road geometry and traffic characteristics affect the bike route safety in the City of Lawrence, KS. Meanwhile, chapter 3 illustrates the process to conduct the regression models and this chapter will explain the outcomes. The null hypothesis is that road geometry, infrastructure, and traffic characteristics have no impact on the bike route safety. The alternative hypothesis is that bike route safety increases with the increase of the bike routes facilities, and it is negatively related with traffic volume, number of lanes, slope, and speed. To draw a conclusion, statistical analysis has been conducted to find the correlation between independent and dependent variable. It is noteworthy to mention that chapter 3 laid the foundation for this analysis by discussing the underlying assumptions and structural components of each regression model. Prior to beginning with the statistical analysis, the regression model components and model evaluation parameters are discussed in the following subsections.

4.1 Outline of Regression Model Components

The model outline includes variables name and the regression models to be applied in this study. The regression models are linear regression, logistics regression, Poisson regression, Zero-inflated Poisson regression, negative binomial regression, and zero-inflated negative binomial regression model. For all these models, predictor, response variables, and sample size are same.

Dependent variable: Crash incidence

Independent variable:

- Number of lanes
- Route slopes (Average Slope)
- Existing bike routes

- Traffic volume
- Posted speed limits

Total route segments: 5081

4.2 Key Model Evaluation Parameters

4.2.1 P-value. The p-value for each variable test the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that the null hypothesis cannot be accepted. Alternatively, an independent variable with low p-value is likely to have a significant impact on the response variable and it would be a good addition to the model. On the other hand, a larger (insignificant) p-value suggests that changes in the predictor are not associated with changes in the response. Typically, p-value helps to determine which independent variable is strongly related with the dependent variable and which one should keep in the model.

4.2.2 R-Squared Value. In the linear regression model, ordinary least squares (OLS) regression method minimizes the sum of the squared residuals. Generally, a model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased. R-squared is a statistical measure of how close the data are to the fitted regression line. R squared value represents the percentage of the response variable variation that is explained by a linear model. However, R^2 value has two basic limitations. First, R-squared cannot determine whether the coefficient estimates and predictions are biased, which is why residual plot analysis is required. Second, R-squared does not indicate whether a regression model is adequate. It is likely to have a low R-squared value for a good model, or a high R-squared value for a model that does not fit the data. Regardless of the R-squared, the statistically significant coefficients still represent the mean change in the response for one unit of change in the independent variable while holding other predictors in the model constant. Therefore, conclusions drawn from the statistically significant predictors are still very important and extremely valuable.

4.2.3 Regression Coefficients. Regression coefficients represent the mean change in the response variable for one unit of change in the independent variable while holding other independent variables in the model constant. This statistical control that regression provides is important because it isolates the role of one variable from all of the others in the model.

4.3 Statistical Analysis to Fit the Data

In Poisson model, the independent variable i.e. speed, slope, lane, bike route and volume are significantly related with dependent variable (Appendix A). But the variance (0.1236) of the dependent variable is greater than mean (0.0562) which is the violation of the Poisson model assumptions. This implies that the data have over-dispersion problem and the Poisson model output cannot be accepted to illustrate the impact of road geometry, infrastructure, and traffic characteristics on bike route safety in Lawrence, KS.

To deal with over-dispersion, Zero Inflated Poisson (ZIP) and Negative Binomial (NB) model have been applied to fit the data better. The outcome of ZIP model indicates statistically significant improvement from Poisson model (Appendix A). The independent variables e.g. speed, lane number, bike route facility, slope, and traffic volume are negatively correlated with the crash incidence. The coefficient of slope and low traffic volume street are not statistically significant, and the application of ZIP model coefficients have produced some erroneous outcomes in the bike safety map for Lawrence i.e. local and low volume streets are unsafe for bicyclists in few neighborhoods. Thus, ZIP model coefficients would not be the best choice to analyze bike safety in Lawrence. The next step is to explore the negative binomial and zero inflated negative binomial model as the data are heavily zero inflated and overly dispersed.

There is no consensus in literature to choose zero inflated negative binomial model over negative binomial without verifying the data distribution and model result. Having lots of zeros in the data do not necessarily always mean that zero inflated negative binomial model will produce better result than that of conventional negative binomial model. Depending on the distribution and type of the data, negative binomial model sometimes generates better result than zero inflated negative binomial model. Certainly, there are some cases when ZINB model would generate more meaningful outcomes than NB model. For instance, if the response variable is the number of children ever born to a sample of 52 years old women, then it is very likely that many women at that age are biologically sterile. In this case, variation in the independent variables might not change (very little chance) the response variable outcome. Surely this type of restriction does not apply for bike crash incidence. There is always a possibility of crash incidence at any section of roads anytime.

As discussed in the previous chapter (subsection 3.3) that *Vuong* test has been applied in this study to measure the model improvement between zero inflated model and their corresponding non-zero equivalent model. Rstudio generated results are summarized below to compare negative binomial, conventional Poisson model with their zero inflated counterparts.

```
> vuong(Poisson,ZIP: Comparison between ZIP and conventional Poisson
-------------------------------------------------------------
              Vuong z-statistic             H_A    p-value
Raw                   -4.747265 model2 > model1 1.0309e-06
AIC-corrected         -4.563891 model2 > model1 2.5107e-06
BIC-corrected         -3.964876 model2 > model1 3.6717e-05
> vuong(ZINB,NB): Comparison between ZINB and conventional Negative Binomial
-------------------------------------------------------------
              Vuong z-statistic             H_A   p-value
Raw                   3.4384711 model1 > model2 0.0002925
AIC-corrected         2.5839216 model1 > model2 0.0048842
BIC-corrected        -0.2075769 model2 > model1 0.4177797
> vuong(ZIP,ZINB): Comparison between ZIP and ZINB
-------------------------------------------------------------
              Vuong z-statistic             H_A  p-value
Raw                   -1.715317 model2 > model1 0.043144
AIC-corrected         -1.715317 model2 > model1 0.043144
BIC-corrected         -1.715317 model2 > model1 0.043144
```

According to *Vuong* test result, ZINB would be a better model than that of NB or ZIP. Zero inflated Poisson model is certainly an improvement from the conventional Poisson model and all three test statistics i.e. Raw, AIC and BIC corrected are statistically significant with very low p value. On the other hand, Raw and AIC value suggest that ZINB model is better choice than conventional NB model to analyze bike crash data for city of Lawrence. Although BIC value indicates other way around but p value is not statistically significant. Finally, all three *Vuong* test statistics suggest that ZINB would be better option than that of zero inflated Poisson model.

By considering the pattern and type of bike crash data, and Vuong test result, it has been decided to consider the output of both ZINB and NB model in investigating the bike safety factors. Negative binomial model generates meaningful coefficients for all variable but bike route facilities, and independent variables are statistically significant as well. The model coefficients are log based, so these numbers are exponentiated to get the odds ratio. The exponentiated value of each variable explains the impact (positive or negative) of road geometry and traffic characteristics on bike route safety. The detail explanations of predictor variable coefficients and a comparison with ZINB model coefficients have been summarized in the following section (section 4.5) of chapter 4.

Negative Binomial Model Coefficients:

```
                Estimate Std. Error z value Pr(>|z|)    
(Intercept)     -1.71193    0.68010  -2.517 0.011830 *  
Speed           -0.04359    0.01648  -2.644 0.008191 ** 
Slope           -0.18934    0.06230  -3.039 0.002374 ** 
Lane             0.41030    0.12054   3.404 0.000665 ***
Bike_Route       0.32529    0.21076   1.543 0.122723    
VolumeCOLLECTOR -0.18197    0.28576  -0.637 0.524270    
VolumeSTREET    -1.07524    0.32719  -3.286 0.001015 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.0793) family taken to be 1)

    Null deviance: 822.52  on 5080  degrees of freedom
Residual deviance: 705.02  on 5074  degrees of freedom
AIC: 1861.5
```

Zero inflated negative binomial model has two components e.g. count model and zero inflated model (logistic model). The log link function of the count model is negative binomial and interpretation of coefficients is similar to the negative binomial model. On the other hand, the coefficients of the zero-inflated model are on the logit scale, and the odds ratio has been calculated by exponentiating coefficients value. It is, however, noteworthy to mention that zero inflated model predicts non-occurrence of the outcome.

```
Count model coefficients (negbin with log link):
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.279871   1.019487   2.236  0.02533 *
Speed           -0.057911   0.022458  -2.579  0.00992 **
Slope           -0.005619   0.105647  -0.053  0.95758
Lane            -0.106037   0.114798  -0.924  0.35565
VolumeCOLLECTOR -0.666485   0.432336  -1.542  0.12317
VolumeSTREET    -1.262345   0.459714  -2.746  0.00603 **
Bike_Route      -0.899831   0.287944  -3.125  0.00178 **
Log(theta)      -0.381392   0.495144  -0.770  0.44114

Zero-inflation model coefficients (binomial with logit link):
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      4.12116    1.12503   3.663 0.000249 ***
Speed            0.03267    0.02477   1.319 0.187157
Slope            0.17844    0.10898   1.637 0.101564
Lane            -1.08748    0.19868  -5.474 4.41e-08 ***
VolumeCOLLECTOR -0.58098    0.49097  -1.183 0.236681
VolumeSTREET    -0.59902    0.56534  -1.060 0.289332
Bike_Route      -1.76222    0.42783  -4.119 3.81e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 0.6829
Number of iterations in BFGS optimization: 78
Log-likelihood: -894.6 on 15 Df
```

The exponentiated coefficients and the range of coefficients with 95% confidence interval for both ZINB and NB model have been attached to the appendix A. All regression results have been generated using Rstudio software.

4.5 Interpretation of Model Coefficients

The coefficient interpretation of predictor variables would enable how road geometry, infrastructure, and traffic characteristics affect the bike route safety in Lawrence. The exponentiated coefficients have been used to interpret the sensitivity of the independent variables. In other words, the exponentiated coefficients value would help to quantify how much one-unit change of independent variable will affect the dependent variable.

Table 7 lists the exponentiated model coefficients and significance level of each independent variable. The coefficients without any asterisk means that there is one in the 95% confidence interval range which implies that coefficient is not significant. The odds ratio of greater than one means positive correlation and less than one means the negative correlation between independent and dependent variable.

Table 7: ZINB and NB model results with exponentiated regression coefficients

Independent Variable	**Exponentiated Coefficients: NB**	**Exponentiated Coefficients: ZINB**	
		Count Model	Logistic Model
Speed	0.96 **	0.94 **	1.03
Slope	0.83 **	0.99	1.19
Lane	1.51 ***	0.89	0.34 ***
Bike Route	1.38	0.41 **	0.17 ***
Volume_Collector	0.83	0.51	0.56
Volume_Street	0.34 **	0.28 **	0.55

Significance Code (α): 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’

4.5.1 Speed. In this study, posted speed limit for each route segment has been used rather than the actual bicyclist's riding speed. As citywide bicycle speed data are not available, vehicular speed limit has been considered as a proxy measurement of bicyclist's safety for city of Lawrence. Streetwise bike speed and volume data collection is expensive for a small city like Lawrence, and the use of vehicular speed as a proxy is also common practice in bike safety studies. The posted speed limit for a route segment provides a relative idea about the bicyclists' speed. Bike crash severity is positively related with the speed limit of a street. The severity of crash incidence increases with the increase of speed. The higher the speed the higher the possibility of skidding and imbalance.

Both ZINB and NB model conclude that speed has a strong correlation with bike route safety and the correlation with dependent variable is statistically significant (α is 0.01). The absence of one in the 95% confidence interval also implies strong correlation with the dependent variable. The odds ratio is less than one which means bike crash frequency is negatively related to speed. Both negative binomial model and count model coefficients support the above conclusion. According to NB model, a one-unit increase of traffic speed would affect bike crash count by a factor of 0.96 with remaining all other variables constant. ZINB model also suggests identical impact of speed on bike route safety. The lower speed local neighborhood streets are more prone to bike crash than high-speed arterial roads. According to City of Lawrence bike crash data, two-thirds of the bike crash incidents occurred on the city streets with a speed of no more than 30 mph. It might be because of the higher volume of bicyclists ride on the neighborhood level city streets. In other words, bike crash incidence reduces in the high-speed road (arterials) because there are not so many bicyclists on arterial roads. The addition of predictor variable like actual bicycle volume in the model would probably produce more logical outcome (which means the crash incidence increases with the increase of bicyclist speed).

4.5.2 Slope (Average). The city of Lawrence has a unique topographic feature i.e. a flat downtown and hilly university campus adjacent to each other. Both locations are the major attractions (either origin or destination) for the bicyclists. So, the presence of steep slope streets between major origin and destination might have (or maybe not) a potential impact on bicycle route choice, especially for the inexperienced bicyclists. On the other hand, low to moderate route slope may not affect route choice for experienced bicyclists but the inexperienced bicyclists usually have preference to bike on relatively flat terrain. Pre-assumption before testing the data, the probability of bike crash incidence would increase with the increase of slope.

Negative binomial model strongly concludes that there is negative relationship between crash incidence and slope. The coefficient of the slope is statistically significant at 0.01 α level with a p-value of 0.0023. The odds ratio of 0.83 implies that one-unit increase of slope would decrease the bike crash incidence by a factor of 0.83 with all other variables constant. Alternatively, the probability of bike crash incidence in the steep slope route is relatively lower. On the other hand, ZINB model coefficients are not statistically significant and the odds of count model concludes like negative binomial model. The odds ratio of 1.019 for zero inflated logistic model part indicates that the slope has a positive proportional relationship with the crash incidence variable. Remaining all other variables constant in the model, a one-unit increase in slope variable would increase the possibility of crash incidence by a factor of 1.019. But slope coefficients in the ZINB model is not significant and it can be concluded that slope does not have strong influence on bike safety for Lawrence.

4.5.3 Lanes. Bicyclist's comfort level usually decreases with the increase of lane number on the street. A higher lane number indicates a higher traffic volume and high-speed vehicles on the road. This creates a stress and discomfort to the bicyclists on the road, but this impact might be very limited in the shared use path (off road bike facilities). In the negative binomial model, the correlation between lane number and crash incidence is statistically significant with significance level (α) of 0.001. The odds ratio of 1.51 indicates that lane number is positively related with crash incidence. Holding all other variables constant, a one-unit increase of lane number would increase the probability of crash incidence by a factor of 1.51. The lower the number of lanes on the road the safer the ride for bicyclists. In contrary, the zero-inflation part of ZINB model is statistically significant at 95% confidence level, but the count model coefficient is not significant at all. The odds ratio of zero inflated part indicates negative relationship between crash incidence and lane number.

4.5.4 Bike Route. Usually, the presence of bike route facility has a positive impact on the comfort and safety of bicyclists. Riders tend to prefer riding on facility bike routes than that of routes without bike route infrastructure. It is important to investigate whether the presence or absence of bike route facilities have any influence on bike route safety in Lawrence.

Negative binomial model coefficient for bike route is not statistically significant, but ZINB model coefficients are statistically significant at significance (α level) of 0.01 and the p-value is very small (0.002). This implies that bike route has significant impact on the safety of bicyclists. The odds ratio of 0.41 indicates that bike route facility would have a negative impact on bike crash incidence which is consistent with other peer reviewed studies. ZINB model concludes that presence of bike routes would decrease crash incidence by a factor of 0.41 with holding all other factors constant in the model. In other words, a one-unit increase of bike route facilities would decrease the probability of crash incidence by the margin of 59%.

4.5.5 Volume. The volume is a factor variable of functional class of road and data were collected from City of Lawrence. Three major road types e.g. street, collector and arterial have been categorized as a proxy of the low, medium and high volume of traffic. In this regression model, the coefficients of collector and street are relative to the arterial road and reference level is arterial road.

The volume is significant in both NB and ZINB model. Bike route safety decreases with the increase of traffic volume. According to ZINB model result, the odds ratio for medium traffic volume route is 0.51 relative to high traffic volume route which means medium traffic volume roads would have a lower possibility of crash incidence than that of high volume traffic roads. In the case of low volume traffic roads, the probability of crash incidence would be decreased by a factor of 0.28 relative to the high-volume traffic roads. The possibility of crash incidence increases with the increase of traffic volume on the street.

To recapitulate, major collector and arterial roads are not the safest option for bicyclists because of higher lane numbers, traffic volume, and speed. The presence of bike route facilities helps to reduce bike crash incidence and route slope is not a strong determinant of bicyclists safety in Lawrence. Both NB and ZINB model coefficients are helpful to get better insights on bike route safety for bicyclists. Only bike route variable in negative binomial model and lane number variable in ZINB model are not statistically significant. Interestingly, both model coefficients have produced almost identical bike compatibility map for Lawrence. Before interpreting the bike compatibility map of Lawrence, it is important to explore bike crash location and its spatial relationship with route slope and bike route facility.

4.6 Identification of Bike Crash Zone in Lawrence

Identification of bike crash zone is important in developing efficient and effective strategies to enhance bike safety. It would be a useful tool for the city planner to decide in which locations safety improvement measure is required. The bike crash data from 2009 to 2013 were collected from the City of Lawrence for crash zone identification. The collected crash data were geocoded by the City of Lawrence which means that the first step of creating crash density map is already done. Each crash incidence represents one dot on the GIS map, but the difficulty of representing the crash data with a dot is that several crashes may have occurred at the same point. To resolve this issue, a density map has been created using spatial analyst tool of ArcMap. The density can be measured using two methods in ArcMap: simple method (point density and line density) and kernel density. In a simple method, the entire study is divided into a preset number of cells (Figure 6). The large circular dots in Figure 6 represent crashes near the cell. The individual cell density values are then calculated as the ratio of number of crashes that fall within the search area to the size of the search area (Pulugurtha et al., 2007)

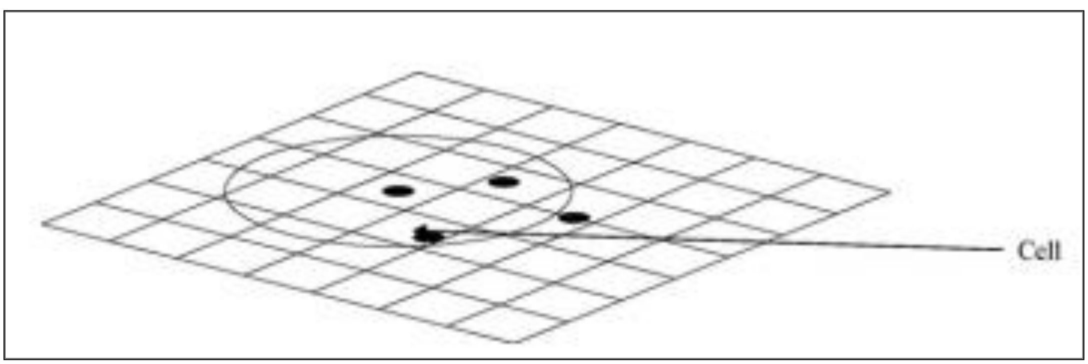


Figure 6: Circular search area around each cell in simple method, Source: (Pulugurtha, Krishnakumar, & Nambisan, 2007)

In a simple method, the outcome of the map is biased by the radius of the search area. For instance, the larger the radius the smoother the density surface. In the kernel method, a circular search area is drawn around each crash (large circular dots in Figure 7) instead of the cell as in the case of the simple method. A kernel function is then applied to each crash to calculate the kernel values. The surface value is highest at the location of the crash. It decreases with increasing distance from the crash, reaching 0 at the radius distance from the crash. The ArcMap uses a quadratic kernel function to estimate kernel density. The individual cell density values are then calculated as the sum of the overlapping kernel values over that cell. (Pulugurtha et al., 2007).

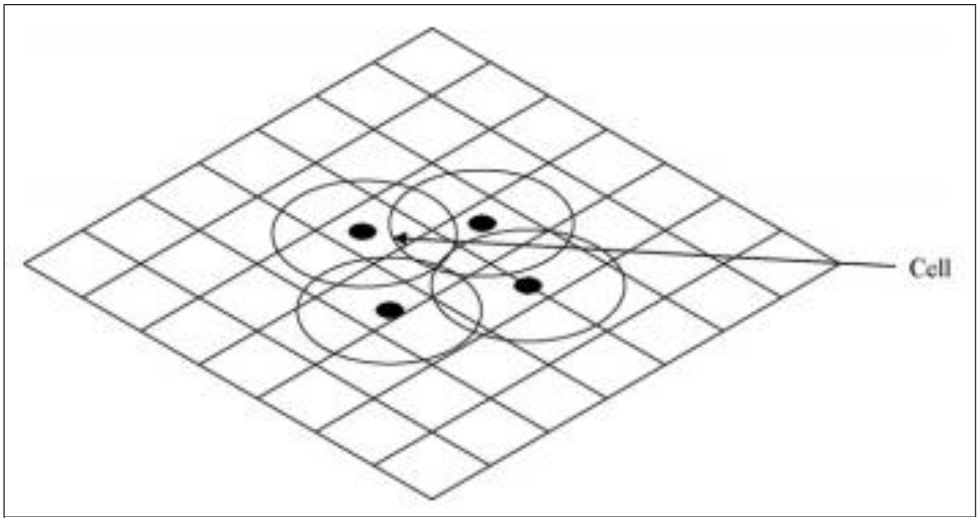


Figure 7:Circular search area around each crash- Kernel Method, Source:(Pulugurtha et al., 2007)

Usually, the kernel method generates smoother looking density surface than simple method. But kernel density map is also affected by the radius of search neighborhood (the greater the radius the flatter the resulting kernel).

To overcome the issue of subjectivity, the concentration areas in the crash density map are categorized into very low, low, medium, high, and very high-risk locations. These ranges are identified for these five classes based on the quintiles for the corresponding variable. Thus, the “very low” class consists of values from 0 to the 20th percentile in the range of values of the 80th percentile.

.

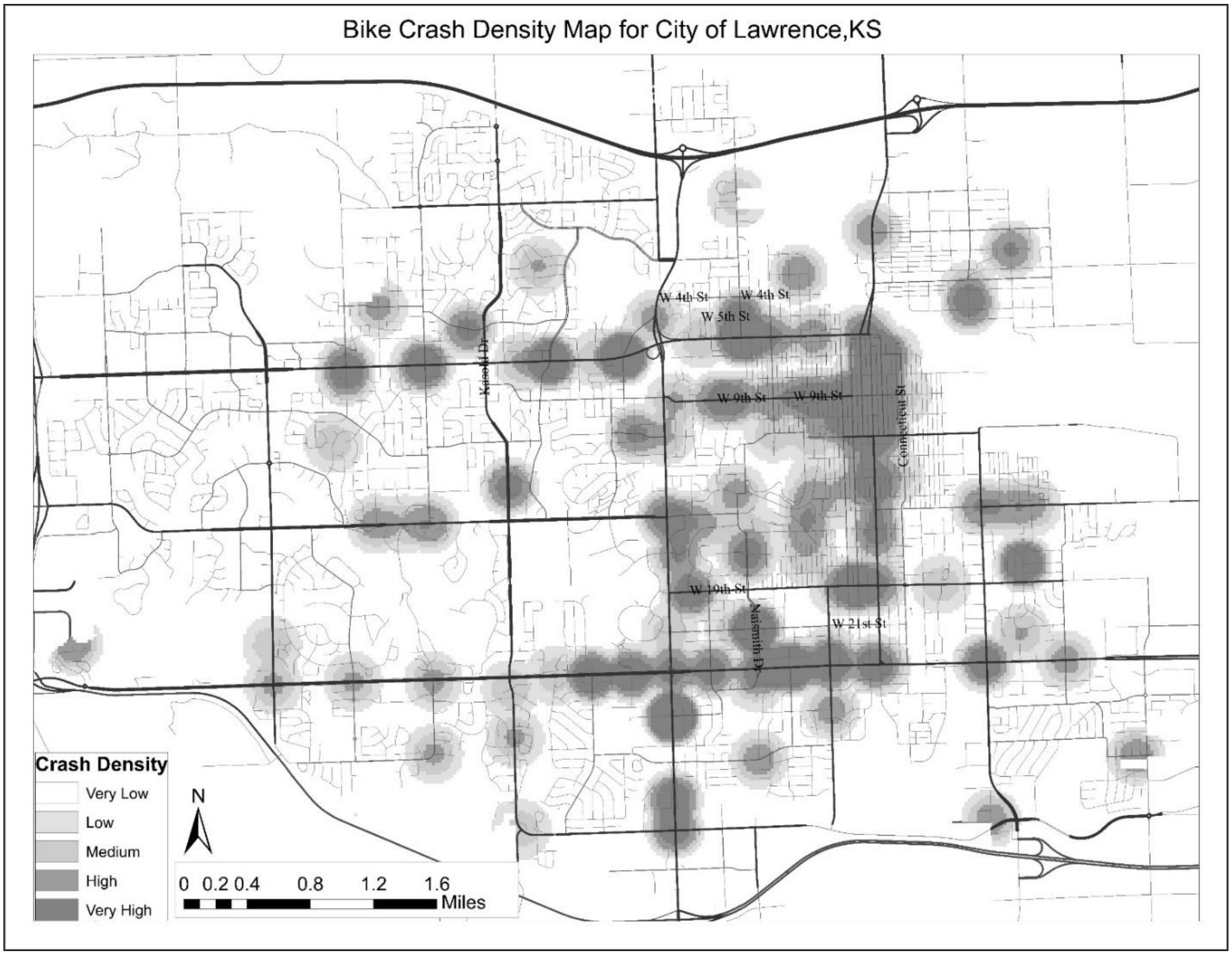


Figure 8: Bike crash density map for the City of Lawrence

Figure 8 shows the crash density in the city of Lawrence and the clusters represent the crash zone in the city. In the preparation of Kernel density map, the search radius and cell size were 1000 feet and 50 feet respectively. The crash density map shows that the largest bike crash zone is located in and around the downtown area. The extent of the crash zone includes 6th street to W 19th street along the Massachusetts Street. The bike crash zone classification is summarized in Table 8.

Table 8: Summary finding of crash zones in Lawrence

Crash Severity	**Streets/Intersections**
Very high crash zone	Massachusetts Street (6th street to 15th street) W 9th street (Mississippi street to Cincinnati street) W 19th street (Louisiana street to Massachusetts street) 23rd street (Naismith to Louisiana street, Crestline drive to Atchison Avenue) 15th and Iowa intersection 6th street and Lawrence Avenue intersection 6th and Rockledge road intersection 6th and Monterey way intersection 6th an S Folks road intersection
High Crash zone	*Kasold* drive and Trail road intersection Monterey way and Bob Billings Parkway intersection Wakarusa drive and Clinton Parkway intersection Clinton Parkway and Inverness drive intersection Crossgate drive and Clinton Parkway intersection
Medium crash zone	23rd and *Kasold* drive intersection W 9th street and Iowa street intersection

4.7 Spatial Relationship between Crash Location and Route Slope

Route slope is an interesting road geometric feature in Lawrence where route slope changes within a short distance e.g. flat downtown and hilly college campus very close to each other. Route slope is a deciding factor for inexperienced bicyclists and hilly landscape of the City of Lawrence has made biking more interesting. The severity of crash incidence in upward hill and downward hill would be different. Presumably, the crash incidence in the downward hill would be more severe and might cause fatal injuries.

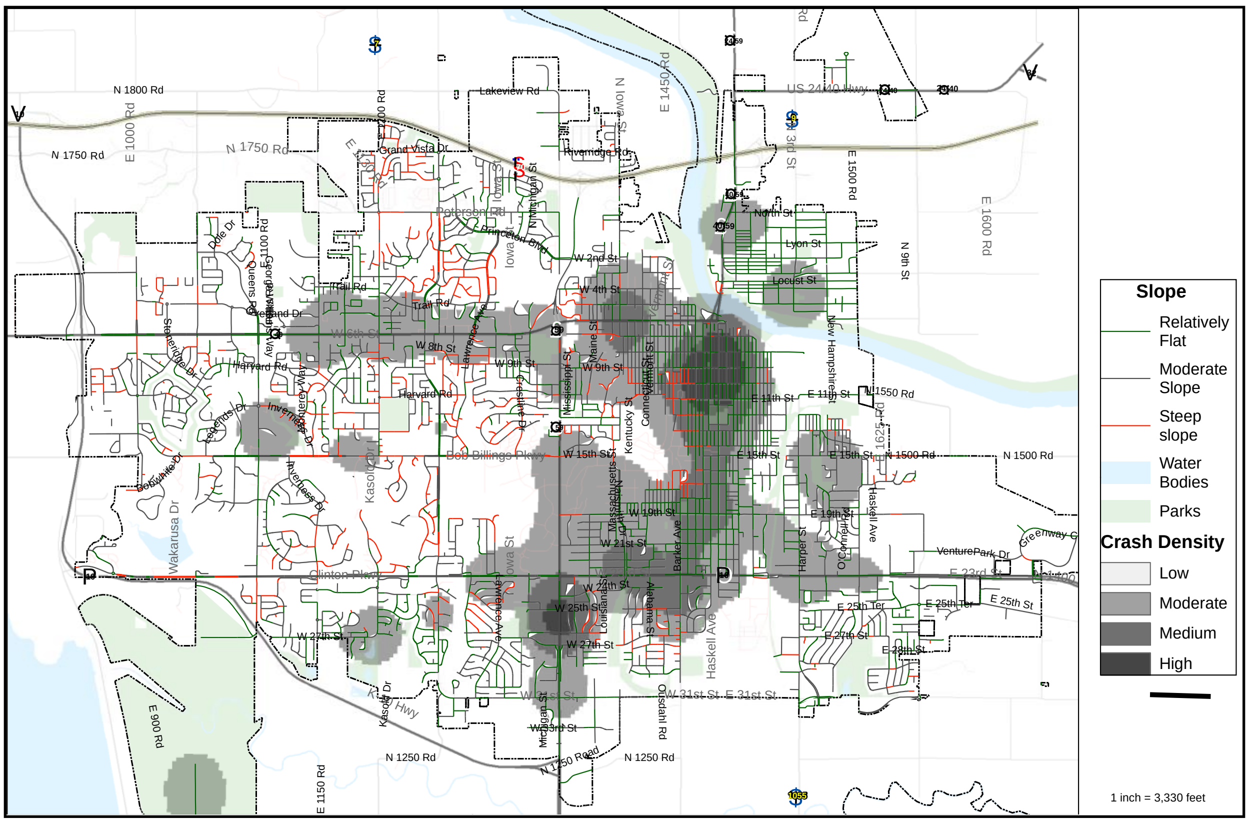


Figure 9: Spatial relationship between bike crash and route slope in Lawrence

Figure 9 shows the spatial relationship between crash incidence and slope distribution. The route slope data have been classified into three categories using natural break method in ArcMap and the classes are relatively flat moderate slope, and steep slope. This layer is then superimposed on top of crash zone map to explore spatial relationship between slope and crash incidence. The average route slope for Lawrence is 2.456^0 and two-thirds of routes have lower than average slope. The summary of the slope and crash density map is listed in Table 9.

Table 9: Summary of route slope and bike crash density map

Route Slope	Range	Remarks
Relatively flat	$<1.5^0$	35.3% route segments fall into this category
Moderate slope	$1.5 – 4.0^0$	51% route segments fall into this category
Steep slope	$4.0 – 16.08^0$	13.7% routes have slope of more than 4^0

Figure 9 shows that most of the bike crash incidence occurred in flat to moderate slope routes. But there are steep slope areas with medium to high crash density in Lawrence e.g. east side of Tennessee street between 14^{th} street and 9^{th} street. This study has concluded that the correlation between bike crash and route slope is not statistically significant. Thus, slope has very little impact in the bike route safety for Lawrence. Slope of major roads and neighborhoods in Lawrence have been extracted from Figure 9.

The slope distribution of major streets in Lawrence are following:

- Iowa Street (Flat to moderate Slope)
- 23^{rd} Street (Flat to moderate Slope)
- Clinton Parkway: Flat to moderate slope (Between Iowa Street and Wakarusa Drive)
  Moderate to steep slope (Further west of Wakarusa Drive)
- W 6^{th} Street (Flat to moderate slope)
- Bob Billings Parkway (combination of steep and moderate slope)
- W 19^{th} Street (Flat to moderate slope)
- E 15^{th} Street (Moderate slope)
- Massachusetts Street (Flat)

- Tennessee Street (Flat)
- Kentucky Street (Flat)
- Wakarusa Drive (Moderate slope, steep slope south of Clinton Parkway)
- Kasold Drive (Mostly flat to moderate slope, steep slope in the Deerfield area)
- Haskell Avenue (Flat to moderate slope)

Synthesis of neighborhood street slope is following:

- Flat to moderate Route Slope
  - Downtown Lawrence
  - Lawrence cultural district
  - Old West Lawrence
  - North Lawrence
  - East Lawrence
  - Lawrence High School neighborhood
  - Prairie Park
  - Free State High School neighborhood
- Moderate to Steep Slope Route
  - Deerfield
  - Sunset Hills
  - West Hills
  - Quail Run and Perry Park
  - The Jayhawk club block

4.8 Relationship Between Crash Location and Bike Route

The presence of bike route facility has a significant influence on bike route safety, and this study has concluded that the probability of bike crash incidence would decrease with the increase of bike route infrastructure. The provision of bike routes not only create safe and convenient places for bicyclists but also give confidence to the people of all ages and abilities to ride safely. With the increase of bike route infrastructure over the years, number of bicyclists in Lawrence has increased 65.7% from 2009 to 2015 (City of Lawrence, 2015). This implies that people tend to bike more if there are bike route facilities available on the street. In addition to the provision of bike route facility, creating a bike culture is also important to reduce bike crash incidence. Higher bicyclist volume in any area does not necessarily mean that bike crash rate would be higher. For instance, the Netherlands has the lowest bike crash rate even though bicyclists volume is highest in the world. So, it is important to encourage people to bike on the street which would help to create a strong bike culture in Lawrence.

Out of 5081 route segments, only 1041 segments have bike route facilities. Figure10 shows the location of bike crash incidence, and existing bike route facilities in Lawrence. In the downtown Lawrence, there are bike routes on Massachusetts street, Vermont street, and New Hampshire street but this area is still very prone to bike crash. Higher traffic and bicyclists volume and frequent conflict points among themselves might be a reason.

The 23rd street is another important route in the city of Lawrence and lots of student housing are located nearby the 23rd street. A bicycle could be the primary mode to commute to the school. Figure 10 shows that 23rd street has so many high crash prone intersections and it would give city planners a heads-up to emphasize on these intersections.

The North Lawrence has some bike crash zone with no existing bike route facilities. The city planners and decision makers can think about some intersections and surrounding neighborhoods i.e. Locust street and N 7th street, Maple street and N 8th street etc., to provide bike route infrastructure.

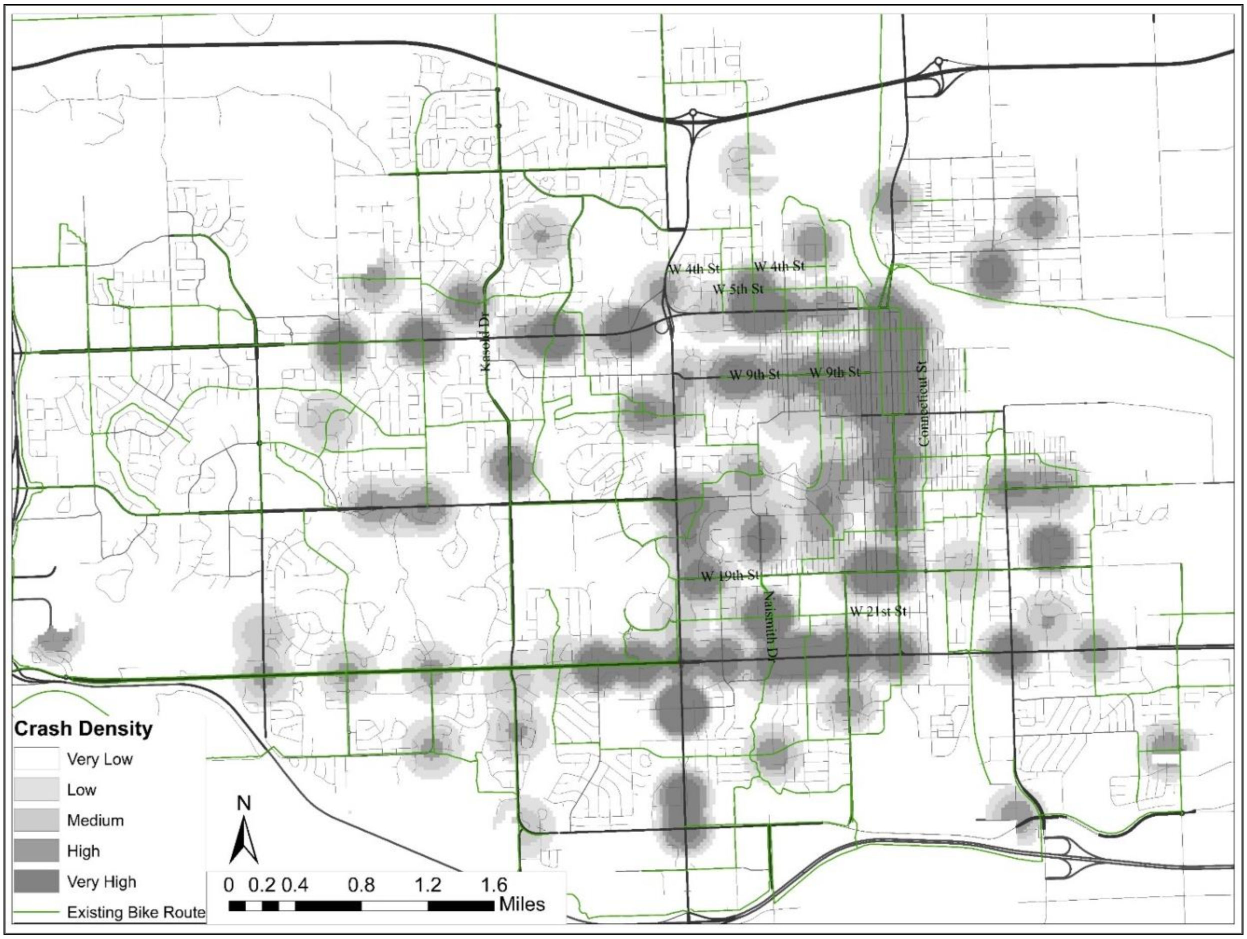


Figure 10: Spatial relationship between bike crash location and bike route facilities

With the understanding of all independent variables and their regression coefficients, bike compatibility map for Lawrence has been prepared using both NB and ZINB model coefficients. Bike compatibility map based on negative binomial model is attached in the appendix B. The next section will explain bike route safety and comfort level using bike compatibility map which is prepared based on ZINB regression coefficients. It will also compare the outcome of bike compatibility map and city's bike rideability map to explore similarities and differences.

4.9 Case Study for Bike Compatibility in Major Corridors

The bike compatibility map provides a very clear idea about which routes in Lawrence are good for all bicyclists and the routes that need some improvements. The map has five class interval ranges from very safe to unsafe for bicyclists. For a better understanding of the map, the major corridors in the city are discussed below.

4.9.1 Clinton Parkway and 23rd Street. According to Figure 11, Clinton Parkway is an unsafe bike route option in the City of Lawrence. The number of lanes, high traffic volume, and high-speed traffic have made the Clinton Parkway as unsafe option for bicyclists. This study also concludes that crash incidence probability increases with higher traffic volume and higher lane numbers on the street. Regardless of flat route slope, high traffic volume and traffic speed have turned Clinton Parkway unsafe for bicyclists. This outcome does not match with the city's *Bike Rideability* Map (Figure 12).

On the other hand, the 23rd Street (right side of Iowa Street) is one of the busiest streets in Lawrence and there are lots of commercial, shopping, restaurants, retail and office buildings on both side of the streets. High volume of traffic, major intersections, and lots of conflict point interrupts mobility of bicyclists flow, frequent entrance and exit of commercial establishments to the 23rd street make the route very difficult for the bicyclists. There are no designated bike routes, but some segments have a wider sidewalk along the road. According to the ZINB regression coefficient value, the presence of bike route has significant impact on bike route safety and high traffic volume negatively affects the safety of bicyclists. Thus, 23rd Street is designated as unsafe option for bicyclist as well (Figure 11).

4.9.2 W 6th Street. The W 6th Street is another high traffic volume and high-speed street in Lawrence. In addition to that, the lack of bike facilities and frequent interruptions with motor vehicles at intersections makes W 6th street unattractive to the bicyclists. As concluded in the ZINB coefficients interpretation, the lack of bike route facilities and high traffic volume decrease the bicyclist's comfort and safety on the street. Major intersections along the 6th street i.e. 6th and Michigan, 6th and Rockledge road, 6th and Lawrence Avenue, 6th and Monterey Way, are very crash prone (Figure 8). Therefore, the bike compatibility map shows that W 6th street belongs to moderate to unsafe option for bicyclists (Figure 11). This outcome does not match with the *Bike Rideability* Map (Figure 12).

4.9.3 Iowa Street. The Iowa Street is a major arterial road in Lawrence and it connects with all other major streets and interstate I-70 as well. According to Kansas Department of Transportation traffic count data, the volume of the traffic at Iowa Street varies between 23000 and 34500 per day. N Iowa Street (even further north of 6th street) has relatively less traffic which is around 9000 per day (City of Lawrence, 2017). Apart from high traffic volume, the Iowa Street has some difficult intersections for bicyclists i.e. 15th street and Iowa, 19th and Iowa street, 21st, and Iowa, 23rd, and Iowa etc. Referring to the ZINB regression coefficients interpretation, the lack of bike facilities on top of above issues makes the Iowa street unattractive to the bicyclists. Both bike compatibility map and City's *Bike Rideability* map have drawn similar conclusion for Iowa Street.

4.9.4 19th Street. The 19th street has medium to high traffic volume ranges from 6000 to 21000 per day based on KDOT and City of Lawrence traffic count (City of Lawrence, 2017). The E 19th street (Barker Avenue to O'Connell road) has comparatively less traffic than West side of the 19th street (Iowa to Massachusetts street). In addition to low traffic, there is also bike facilities after Barker and 19th street intersection which makes the street more friendly and comfortable for the bicyclists. As per ZINB model result, bicyclist's comfort increases with the decrease of traffic volume. Therefore, the compatibility map and rideability map draw the same conclusion. On the other hand, biking on the W 19th street is a bit challenging and bicyclists need to be careful about some major intersections i.e. Louisiana Street and W 19th street, Tennessee and W 19th street etc. Bicyclists need to be very watchful at the W 19th street and Iowa intersections as many vehicles right turn quickly to get off the major arterial even in the red signal. Also, Figure 8 indicates that the route segment between Iowa and *Ousdahl* road has high crash density. Overall, 19th street is a safe option for bicyclists in Lawrence.

4.9.5 Bob Billings Parkway. This street is characterized by medium to high traffic volume (11000 to 15500 per day) and route slope changes frequently (City of Lawrence, 2017). Bob Billings Parkway is very difficult for the inexperienced bicyclists as riders need to handle both uphill and downhill route slope very often. The route segment between Wakarusa Drive and *Kasold Drive* are very challenging for the bicyclists. Figure 8 indicates that there is two high bike zones i.e. Monterey Way and Bob Billings intersection, in this route segment. The bike compatibility map (Figure 11) indicates that this segment is moderately unsafe option for bicyclists. Along the Bob Billings Parkway, there are three apartment housings i.e. Meadowbrook, Orchard Corners, and Aspen West apartment between Iowa Street and *Kasold Drive*. Lots of KU students who live in these apartment use bicycle for their daily commute. As concluded in the regression model interpretation that lane numbers, high speed, and traffic volume affect bike route safety significantly. The Bob Billings and Iowa street intersection is one of the most difficult intersection for the bicyclists as it is very high crash prone area (Figure 8). Thus, Bob Billings Parkway is designated as moderately unsafe to unsafe option for bicyclists in Lawrence. The Inverness Drive is one of the safest connector routes for bicyclists between Bob Billings Parkway and Clinton Parkway.

4.9.6 Downtown Lawrence. The downtown is a major activity hub in Lawrence and it attracts a lot of people from all over the city for various purposes. The city of Lawrence has facilitated adequate bike route facilities to ensure bike connectivity with the Lawrence downtown. Streets with bike route facilities in downtown are Massachusetts street (between 11th street and 23rd street), Vermont street (between 11th and 6th street), New Hampshire street (between 11th and 6th street), Connecticut street (between 7th and 15th street), and 7th street (between Michigan street and Delaware street). Downtown Lawrence is also connected with north Lawrence through Massachusetts and Vermont street with bike route which is separated from motor vehicles.

The bike compatibility map (Figure 11) shows that biking on the downtown Lawrence is a safer option for the bicyclists. The underlying road geometry and traffic factors that contribute to safer bike route in downtown Lawrence are lower number of lanes, existing bike route facilities, flat terrain, low vehicular speed, and relatively low traffic volume. For instance, routes are two lane, route slope is less than 1.5 degree, traffic volume on Vermont street is just over 4000 per day (City of Lawrence, 2017). The 7th street has bike route facilities with relative lower traffic volume (around 1000-1200 vehicles per day) which makes it a good choice for bicyclists to move east-west direction from downtown (City of Lawrence, 2017).

4.9.7 9th Street. The 9th street is another major east-west connector street in Lawrence. It is characterized by medium to high volume traffic (around 16,700 per day) and lots of major intersections i.e. Kentucky and 9th, Tennessee and 9th, Maine and 9th, Mississippi and 9th street etc. The street has bike route facility and it is relatively flat terrain. According to the bike compatibility map, the 9th street is mostly a safe option for bicyclist.

4.9.8 Deerfield (North of 6th street). The neighborhood around Lawrence Country Club is very safe and comfortable for the bicyclists according to the bike compatibility map. Low volume of traffic, fewer interruptions with vehicles, and the presence of bike facilities on Lawrence Avenue, Princeton Boulevard, Trail road which connects Lawrence Avenue and Monterey way make these routes better and safer choice for biking. Interestingly, city's *Bike Rideability* map also recommends these routes safe for all riders.

4.9.9 North and East Lawrence. According to bike compatibility map (Figure 11), both north and east Lawrence are very bike friendly and the underlying factors might include low traffic volume, presence of bike route facilities, and fewer lane numbers.

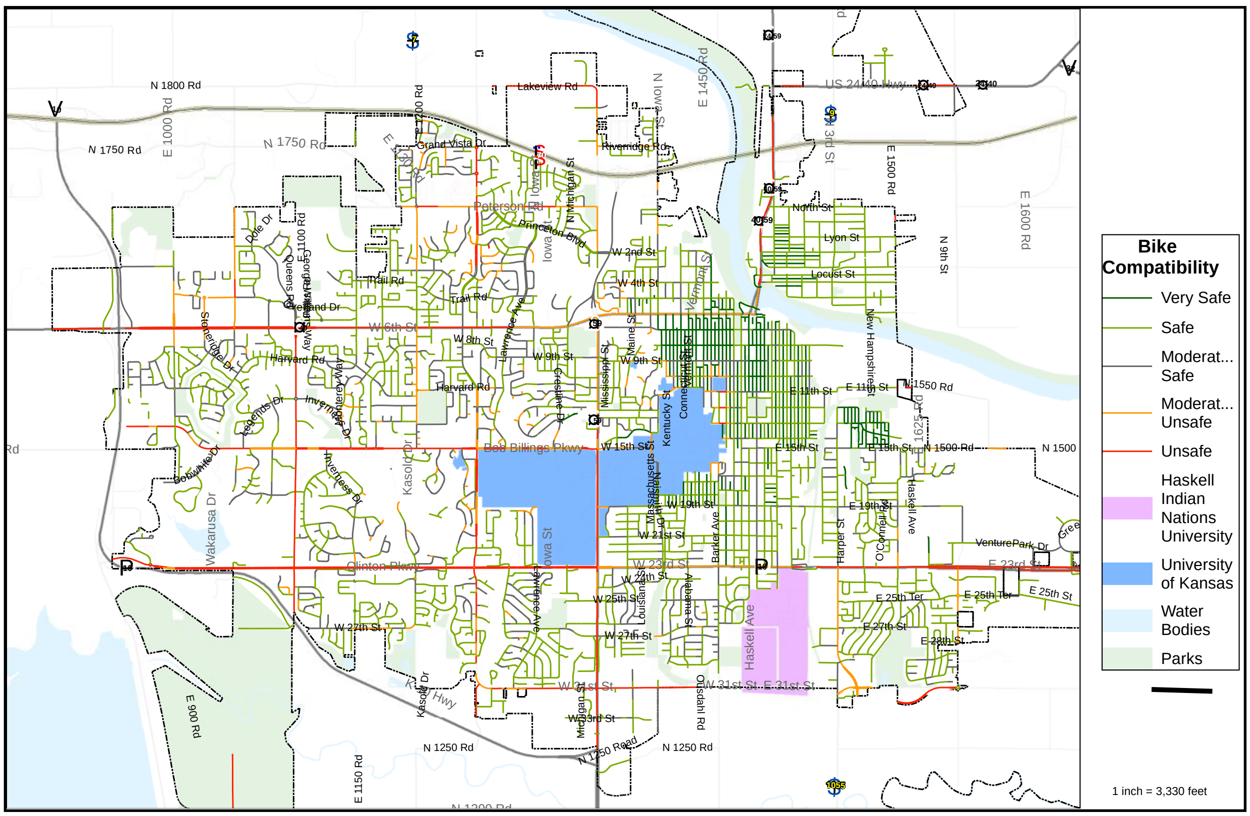


Figure 11: Bike Compatibility Map for City of Lawrence

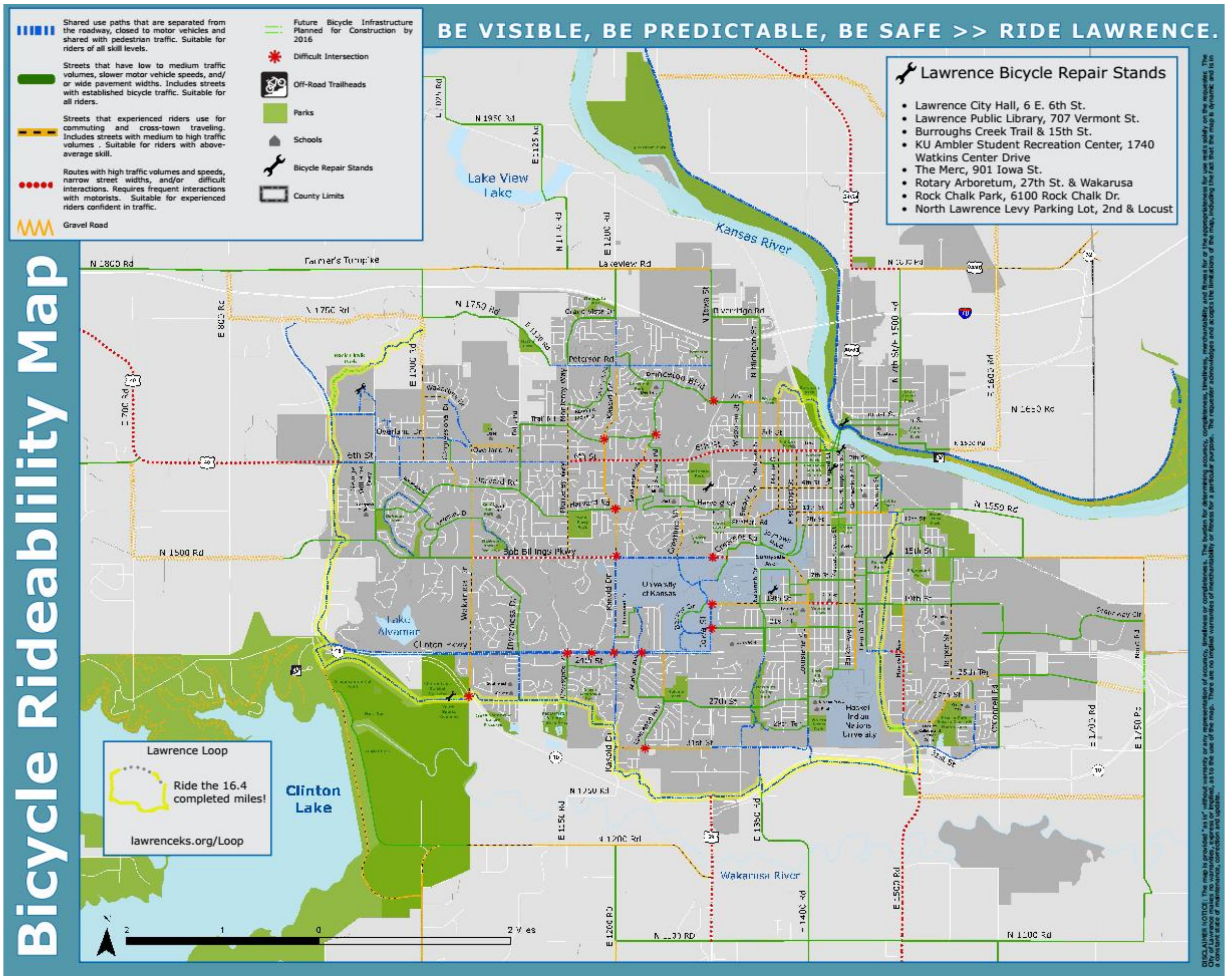


Figure 12: Bike Rideability Map for City of Lawrence

Source: https://assets.lawrenceks.org/assets/mpo/bicycle/BikeMap.pdf

4.9.9 Criticism and limitations of the Compatibility Map. Despite the advantages of the bike compatibility map, it has some downsides as well and the map could produce more detailed and precise results on bike route safety. In this study, road functional class has been used as a proxy of traffic volume, because actual bike count data are not available for Lawrence. The addition of actual bike count data and bicyclist speed in the regression model would probably make a significant difference in the model output. Only installation of bicycle route facility may not always improve the bicyclist's safety, considering pavement conditions of bike routes and sidewalks, on-street parking, and land use data might add some values in the explanatory power of the model. The conceptual framework (section 2.4) described all possible factors that might have an impact on bike route safety, but data unavailability is a problem for this study. The city of Lawrence has started collecting bike count data voluntarily but an emphasize can be added on bike count and speed data collection to reduce bike crash incidence in Lawrence.

Route curvature might have an impact on bike route safety especially bike crashes due to sharp edge maneuvering. The addition of horizontal route curvature as a predictor variable in the model might improve some predictability. The '*Curve Detective'* tool which is available in '*HighwaySafety.pyt'* (a customized Python toolbox) can be used to calculate the bike route curvature for Lawrence. The last but not the least, the study area is very small, and the bigger city with more bike crash incidents and relevant road geometry data would result in better regression model output.

4.10 Interview with Professional Expertise

Apart from literature review, the insight of professional expertise is a useful tool to improve knowledge base. Therefore, two interviews have been conducted to get their insight and comments on the outcome of the bike compatibility map. These professionals are David Cronin, Lawrence City Engineer, and Aaron Bartlett, Senior Transportation Planner at Mid-America Regional Council (MARC).

Aaron Bartlett has agreed with the variables considered to prepare the bike compatibility map for Lawrence. He commented that these are all good variable to include in bike safety analysis. He mentioned that "the study area, Lawrence, is very small city and it does not have enough bicycling or network to produce reliable results. Look at Portland or a larger city. My guess is that your R squared is relatively low due to this factor."

City Engineer of Lawrence have gone through the bike compatibility map and he agreed with most of the outcomes. He recommended conducting the bike route safety analysis using actual traffic volume data instead of proxy data to see if that makes any difference. City of Lawrence and Kansas Department of Transportation jointly collect traffic volume data for some designated routes in Lawrence. Streetwise traffic volume data have been inserted to the ArcMap database and then regression model has been run again. In this case, the sample size was 1698 route segments and routes have actual traffic volume have been considered. The model also concludes that volume has negative correlation with bike route safety. This implies that routes with high traffic volume are more prone to bike crash incidence and are less safe for bicyclists. Addition of actual traffic volume in the model does not change or produce any meaningful coefficients for the other independent variables. Yet it is good to know that information, and it would be helpful for future researchers. The next possible step would be adding actual bicycle volume to get better model results. The city of Lawrence has started collecting bike count data and it is now conducted by volunteers in different locations of the city. Volunteers are using National Bike and Pedestrian Documentation (NBPD) project method for collecting bike count data. When city of Lawrence would have enough bike count data with decent sample size, then this regression model would probably produce more pragmatic results. The next chapter will summarize the major findings of the study and recommendations to improve the bike safety in the City of Lawrence.

Chapter 5: Recommendation and Conclusion

This chapter will summarize the major findings by synthesizing the experimental results and explain how the objectives of the research are met. The main findings of the data analysis and statistical model will be recapitulated in this chapter. Finally, some recommendations will be provided to improve the bike safety in the city of Lawrence.

5.1 Major Findings

The main goal of the study was to find what road geometry and traffic characteristics affect the bike route safety. In order to achieve that goal, some supplementary questions were set in the beginning of the study. Table 10 summarizes how these objectives are met in the study and the corresponding results.

Table 10: Summary of major findings of the study

Queries	Results	Remarks
Does the presence of bike routes affect the safety for the bicyclists?	Yes, the bike route variable is strongly related with response variable with a significance level of 0.01 and p value of 0.002. The variable would have negative impact on the response variable (crash incidence) The odd ratio is 0.41 which implies that the presence of bike routes would decrease crash incidence by a factor of 0.41 (given all other factors are constant).	Details can be found in Chapter 4, Section 4.5.4
	The slope is not statistically significant in ZINB model, but it is negatively correlated with bike crash incidence according to NB model at 0.01	Details: Chapter 4,

Is there a relationship between the route slope and the number of crashes?	α level where the p-value is 0.0023 which is significant as well. NB model concludes that bike crash incidence does not necessarily increase with the increase of route slope.	Section 4.5.2
Where are the existing bicycle crash zones located in Lawrence?	Very high crash zone: Mass St. (between 6th and 15th St.) W 19th St. (between Louisiana & Mass) 15th & Iowa intersection etc. High crash zone: Monetary way and Bob billings intersection Wakarusa Dive. & Clinton Parkway Intersection etc. Medium Crash Zone: 23rd & *Kasold Drive* intersection. W 9th St. & Iowa intersection etc.	Details at Chapter 4 Section 4.6 Figure 6 Table 8
Does high traffic volume affect bicyclists' safety?	Yes, traffic volume is negatively correlated with bike safety. This implies that local neighborhood streets are much safer for bicyclists than collector and arterials roads. The probability of crash incidence increases with the increase of traffic volume. In this study, functional road class has been used as a proxy of traffic volume. The whole model has been re-run with actual traffic	Details at Chapter 4 Section 4.5.5

	volume in selected routes (sample size was 1698). This new model result also concludes that traffic volume has negative correlation with bike route safety.	
Relationship between crash incidence and traffic speed	Statistically significant (α level is 0.001) relationship exists between traffic speed and crash incidence. Speed is negatively correlated with crash incidence. It indicates that collector and arterial roads are less prone to bike crash incidence. It might be because there are fewer number of bicyclists on these high-speed roads. Given that two-thirds of crash incidents occurred on city streets with a speed of no more than 30 mph. In this study posted speed limit has been used instead of actual bicyclist speed. Availability of bicycle speed data would probably improve the model result.	Details at Chapter 4 Section 4.5.1
Does the bike crash incidence increase with the increase of lane numbers?	The probability of bike crash incidence increases with the increase of lane numbers. Local streets are safer than collector or arterial streets.	Details at Chapter 4 Section 4.5.3
Spatial relationship between bike crash location and bike route facilities.	Figure 10 shows which routes and intersections are crash prone. It implies where the city of Lawrence needs to extend bike route facilities.	Details at Chapter 4 Section 4.8

The rideability map shows bike suitability for collector and arterial routes based on the bicyclist's user rating. The compatibility map shows more details of bike route comfort and safety based on road geometry, infrastructure, and traffic characteristics of all types of streets. There are both similarities and differences between the bike rideability map and the bike compatibility map. The bike compatibility map indicates that neighborhood level streets that are characterized by low traffic volume, fewer lane numbers etc. are safer option for bicyclists. Moreover, the presence of bike route facility makes a difference on bicyclists safety. A synthesis of bike compatibility map is listed below.

Recommended routes for bicyclists are:

- Local neighborhood streets
- W 19th street (watch out at Louisiana and W 19th, Naismith and W 19th intersections)
- Inverness Drive (it connects Bob Billings and Clinton Parkway)
- Harvard road
- Trail Road (Monterey Way to Lawrence Avenue)
- W Princeton Boulevard (Peterson to S Iowa street)
- Lawrence Avenue (W Princeton Boulevard to W 6th street)
- Peterson road (Monterey way to N Iowa street)
- W 27th street
- Harper Street
- Haskell Avenue (North part of 23rd street)
- Connecticut street
- E 13th street
- Downtown Lawrence

- North Lawrence neighborhood
- East Lawrence neighborhood
- W 4th street (McDonald Drive to Indiana street)

Route segments where inexperienced bicyclists are not recommended include major collector and arterial roads. The city of Lawrence needs to take some actions to improving bike safety on those roads. The actions might include providing off road bike facilities (shared use paths), improvement in the sidewalk pavement, or providing more bicycle friendly signage at the major intersections etc. List of streets that require city's attention are following.

- W 6th street
- Iowa Street (W 6th street to W 31st street)
- 23rd street
- Bob Billings Parkway
- Clinton Parkway
- Kasold Drive
- Wakarusa Drive
- Kasold Drive and Petersen road intersection in Deerfield area
- W 15th street (Iowa street to Naismith Drive)
- Haskell Avenue (23rd street to E 31st street)

The high bike crash density intersections that require special attentions are:

- Iowa and W 15th street
- Iowa and 23rd street
- W 19th street and Iowa street
- *Kasold Drive* and W 6th street

- Massachusetts street and 23rd street
- Rockledge road and W 6th street
- Lawrence Avenue and W 6th street
- Monterey way and W 6th street
- W 6th street and S Folks road
- Louisiana street and W 19th street

Both negative binomial (NB) and zero inflated negative binomial (ZINB) model generate identical bike compatibility maps. Bayesian Information Criterion (BIC) test statistics predicts model improvement better when sample size is large. In this study, BIC indicates negative binomial model coefficients are better than zero inflated negative binomial model coefficients, but p value was very high (0.42). On the other hand, Raw and Akaike Information Criterion (AIC) test statistics suggest that ZINB model is better than conventional NB where p value is statistically significant for both AIC and Raw test statistics. In a nutshell, there is no consensus that zero inflated negative binomial model would generate better result than conventional negative binomial model. It depends on the type and distribution of data, independent and dependent variable, and research queries.

5.2 Recommendations to Improving Bike Safety

According to the League of American Bicyclists, the number of bicycle commuters nearly doubled in the 70 largest cities of US between 1990 and 2012. In 2014, Americans make more than four billion trips by bike each year (Petersen, 2014). In the US, bike crashes are increasing day by day with the increase in the number of bicyclists. But the situation is the other way around in the Netherlands, because of the excellent bike route design, the provision of adequate route facilities, and the establishment of biking as a part of their culture. In this study, the term “crash” is used instead of “accident” to emphasize that design and planning of the transportation system contributes to most of the bike fatalities and injuries. A set of recommendations have been proposed based on the findings of this study.

- The presence of bike route facility helps to reduce bike crash incidence, and it would increase bike route safety considerably. Most of the collector and arterial roads in Lawrence are not safe for bicyclists. As these roads have high traffic volume with high speed. Off street bike route facility e.g. shared use path or barrier protected bike lane could be an effective way to improve bike route safety. Between 2009 and 2013, two-thirds of bike crash incidence occurred in neighborhood level streets. To improve bike safety in neighborhood streets, dedicated right of way for bicyclists could be a better choice. In fact, encouraging biking in the neighborhood streets would help to create a bike culture in a community. This might be helpful to reduce bike crash incidence in the local streets of Lawrence.
- Citywide bike count data is a prerequisite to better understand the spatial distribution of bicyclists and their trip pattern. The city of Lawrence is currently collecting bike count data on a voluntary basis. Prioritizing the bike count data collection project would help to minimize bike crash incidence in Lawrence. It is noteworthy to mention that availability of streetwise bike count data could result in better regression coefficients which might be helpful insight to improve bike safety in Lawrence.
- While analyzing bike crash data it was found that the city of Lawrence stores bike crash information in ArcGIS platform with a single line geometry. Instead of one single line, bike crash data can be stored in two lines with much more detailed information for each

direction of the street. This would help to extract the directional impact on bike route safety especially for a city like Lawrence where route slope changes drastically.

5.3 Conclusions

The bike route safety study based on road geometry, infrastructure, and traffic characteristics conclude that high traffic volume streets with larger lane numbers e.g. collector and arterial roads are unsafe for bicyclists. Local neighborhoods level streets are comparatively safer routes for bicyclists in Lawrence. Route slope does not have significant impact on bike route safety for Lawrence even though slope changes drastically very often. Surprisingly, speed is negatively related with crash incidence and underlying reason is that this study used posted speed limit as speed variable. Availability of actual bicycle speed for each street would probably change the model outcome. Instead of using traffic volume, actual bike count data for the whole city would surely bring some meaningful insights in the bike safety analysis. It can be concluded that complex statistical analysis on bike safety using available data adds some value in the understanding of bike safety for the city of Lawrence. The bike compatibility map indicates which routes are safer option for bicyclists and which streets are unsafe for bicyclists. This compatibility map has some flaws i.e. details are missing which can be improved by incorporating variables like actual bike count data, bicyclists speed etc. These limitations of bike compatibility map can be overcome with better data (as listed on the conceptual framework), and advanced statistical methods would not make any difference unless better data available.

Bicycle crash incidence is very much under reported, and the city of Lawrence does not have enough crash data to develop regression model with high predictability. Conducting this type of research for big cities like Portland would produce more meaningful outcomes. An addition of variable in the model like on-street parking, actual bicycle volume and speed for different street segments, pavement condition of streets and sidewalks, land use type- residential or commercial, the curvature of streets etc. would probably improve the model predictability. In short, there is probably no better way than empirical analysis to explore inherent causal relationship among variables with a comprehensive data available.

References

AASHTO Executive Committee. (1999). Guide for the Development of Bicycle Facilities. *American Association of State Highway and Transportation Officials, Washington, DC*.

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice, 46*(10), 1730-1740.

Butchart, A., Kruger, J., & Lekoba, R. (2000). Perceptions of injury causes and solutions in a Johannesburg township: implications for prevention. *Social science & medicine, 50*(3), 331-344.

Casello, J., Nour, A., Rewa, K., & Hill, J. (2011). *An analysis of stated preference and GPS data for bicycle travel forecasting*. Paper presented at the 90th Annual Meeting of the Transportation Research Board.

Center for Sustainability. (2016). KU Bike Plan. Retrieved from http://sustain.ku.edu/campus-bike-plan

Chang, H.-l., & Chang, H.-W. (2005). Comparison between the differences of recreational cyclists in national scenic bikeway and local bike lane. *Journal of the Eastern Asia Society for Transportation Studies, 6*, 2178-2193.

Chu, Y., Azer, M., Catalanotto, F., Ungar, H., & Goodman, L. (1999). *Safety/GIS models reviewed and related to long island arterial needs study*. Paper presented at the Enhancing Transportation Safety in the 21st Century ITE International Conference.

City of Lawrence. (2015). Lawrence Kansas Bicycle and Pedestrian Documentation Data Report 2009-2015. Retrieved from http://worldonline.media.clients.ellingtoncms.com/news/documents/2016/02/03/Results_of_2015_count.pdf

City of Lawrence. (2017). Traffic Counts. Retrieved from http://lawrenceks.maps.arcgis.com/apps/webappviewer/index.html?id=576ac2f2c4214bf6ad0f67c185b459f3

Desmarais, B. A., & Harden, J. J. (2013). Testing for zero inflation in count models: Bias correction for the Vuong test. *The Stata Journal, 13*(4), 810-835.

Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of public health policy, 30*(1), S95-S110.

Hu, J., Zhong, G., Cheng, Z., & Wang, D. (2012). *GIS-based road safety evaluation model for cyclist in campus of Higher Education Mega Center*. Paper presented at the Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on.

Hyodo, T., Suzuki, N., & Takahashi, K. (2000). Modeling of bicycle route and destination choice behavior for bicycle road network plan. *Transportation Research Record: Journal of the Transportation Research Board*(1705), 70-76.

Larsen, J., & El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation research part D: transport and environment, 16*(2), 172-177.

Mortinger, J. (2015). *Pedestrian safety crash analysis*. Retrieved from https://assets.lawrenceks.org/assets/mpo/pedplan/memoCrashAnalysisDRAFT.pdf

Noland, R. B., & Kunreuther, H. (1995). Short-run and long-run policies for increasing bicycle transportation for daily commuter trips. *Transport Policy, 2*(1), 67-79.

Petersen, M. (2014). Gadgets to Boost Bike Safety. *The New York Times*. Retrieved from https://well.blogs.nytimes.com/2014/03/10/gadgets-to-boost-bike-safety/?_r=0

Petritsch, T., Landis, B., Huang, H., & Challa, S. (2006). Sidepath safety model: bicycle sidepath design factors affecting crash rates. *Transportation Research Record: Journal of the Transportation Research Board*(1982), 194-201.

Pucher, J., & Buehler, R. (2006). Why Canadians cycle more than Americans: a comparative analysis of bicycling trends and policies. *Transport Policy, 13*(3), 265-279.

Pulugurtha, S. S., Krishnakumar, V. K., & Nambisan, S. S. (2007). New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis & Prevention, 39*(4), 800-811.

Sener, I., Eluru, N., & Bhat, C. (2008). An analysis of bicyclistts and bicycling characteristics: who, why and how much are they bicycling. *Transportation Research Record, 2*, 63-72.

Singleton, A. D., & Lewis, D. J. (2012). Including accident information in automatic bicycle route planning for urban areas. *Urban Studies Research, 2011*.

Snizek, B., Nielsen, T. A. S., & Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography, 30*, 227-233.

Stinson, M., & Bhat, C. (2004). Frequency of bicycle commuting: internet-based survey analysis. *Transportation Research Record: Journal of the Transportation Research Board*(1878), 122-130.

Stinson, M. A., & Bhat, C. R. (2005). *A comparison of the route preferences of experienced and inexperienced bicycle commuters.* Paper presented at the TRB 84th Annual Meeting Compendium of Papers.

United States Census Bureau. (2014). Quick Facts,Lawrence,Kansas. Retrieved from https://www.census.gov/quickfacts/table/PST045215/2038900

Winters, M., Teschke, K., Grant, M., Setton, E., & Brauer, M. (2010). How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. *Transportation Research Record: Journal of the Transportation Research Board*(2190), 1-10.

Yang, T.-H., Lin, J.-R., & Chang, Y.-C. (2010). *Strategic design of public bicycle sharing systems incorporating with bicycle stocks considerations.* Paper presented at the Computers and Industrial Engineering (CIE), 2010 40th International Conference on.

Yiannakoulias, N., Bennet, S. A., & Scott, D. M. (2012). Mapping commuter cycling risk in urban areas. *Accident Analysis & Prevention, 45*, 164-172.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of statistical software, 27*(8), 1-25.

Appendix

Appendix A: Regression Model Results

Poisson Regression Model:

R Code:

Step 1: Install necessary packages and Data in R

```
> install.packages("pscl")
>install.packages("MASS","lattice","ggplot2")
> install.packages("boot")
> Data <- read_csv("C:/Users/SHOFI/Desktop/Thesis Analysis/Data.csv")
```

Step 2: Proxy of volume data (factor variable of Functional road class)

```
> Data$FUNCTCLASS<-factor(Data$FUNCTCLASS)
> table(Data$FUNCTCLASS)
> Data$Volume<-Data$FUNCTCLASS
```

Step 3: Run Poisson model

```
Poisson<-glm(Crash~Speed+Slope+Lane+Bike_Route+Volume,
family=poisson,data=Data)
```

```
> summary(Poisson)

Call:
glm(formula = Crash ~ Speed + Slope + Lane + Bike_Route + Volume,
    family = poisson, data = Data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0675  -0.3648  -0.2348  -0.2106   7.6028

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.30932    0.47307  -2.768 0.005645 **
Speed           -0.07337    0.01090  -6.733 1.66e-11 ***
Slope           -0.16767    0.04780  -3.508 0.000452 ***
Lane             0.52537    0.07230   7.267 3.68e-13 ***
Bike_Route       0.23063    0.13469   1.712 0.086829 .
VolumeCOLLECTOR -0.14802    0.18466  -0.802 0.422805
VolumeSTREET    -1.16391    0.22754  -5.115 3.13e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1995.5  on 5080  degrees of freedom
Residual deviance: 1754.4  on 5074  degrees of freedom
AIC: 2184.8

Number of Fisher Scoring iterations: 6
```

Zero-Inflated Poisson (ZIP) Model:

R Code:

```
ZIP<-zeroinfl(Crash~Speed+Slope+Lane+Bike_Route+Volume,data=Data)
```

Model Summary:

```
> summary(ZIP)

Call:
zeroinfl(formula = Crash ~ Speed + Slope + Lane + Bike_Route + Volume, data =
Data)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-0.6121 -0.1786 -0.1253 -0.1126 74.2310

Count model coefficients (poisson with log link):
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.28830    0.73418   3.117  0.00183 **
Speed           -0.03715    0.01717  -2.164  0.03048 *
Slope           -0.00392    0.07809  -0.050  0.95996
Lane            -0.16057    0.09388  -1.710  0.08719 .
Bike_Route      -0.79038    0.22907  -3.450  0.00056 ***
VolumeCOLLECTOR -0.43037    0.29778  -1.445  0.14838
VolumeSTREET    -1.04124    0.33428  -3.115  0.00184 **

Zero-inflation model coefficients (binomial with logit link):
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.64774    0.95438   3.822 0.000132 ***
Speed            0.05017    0.01903   2.637 0.008376 **
Slope            0.16177    0.08173   1.979 0.047764 *
Lane            -0.92775    0.17742  -5.229 1.70e-07 ***
Bike_Route      -1.30038    0.31461  -4.133 3.58e-05 ***
VolumeCOLLECTOR -0.28920    0.35650  -0.811 0.417237
VolumeSTREET    -0.15271    0.46739  -0.327 0.743871
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 26
Log-likelihood: -904.2 on 14 Df
```

Negative Binomial Model:

R Code:

```
> NB<-glm.nb(Crash~ Speed+Slope+Lane+Bike_Route+Volume,data=Data)
```

Model Output Summary:

```
> summary(NB)

Call:
glm.nb(formula = Crash ~ Speed + Slope + Lane + Bike_Route +
    Volume, data = Data, init.theta = 0.0792526012, link = log)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-0.5656  -0.3075  -0.2261  -0.2032   4.1660

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.71193    0.68010  -2.517 0.011830 *
Speed           -0.04359    0.01648  -2.644 0.008191 **
Slope           -0.18934    0.06230  -3.039 0.002374 **
Lane             0.41030    0.12054   3.404 0.000665 ***
Bike_Route       0.32529    0.21076   1.543 0.122723
VolumeCOLLECTOR -0.18197    0.28576  -0.637 0.524270
VolumeSTREET    -1.07524    0.32719  -3.286 0.001015 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.0793) family taken to be 1)

    Null deviance: 822.52  on 5080  degrees of freedom
Residual deviance: 705.02  on 5074  degrees of freedom
AIC: 1861.5

Number of Fisher Scoring iterations: 1


              Theta:  0.0793
          Std. Err.:  0.0120

 2 x log-likelihood:  -1845.4770
```

As the coefficients are log based, it is important to exponentiate the values to get the odds ratio.

```
> exp(cbind(CO=coef(NB),confint(NB)))
                       CO     2.5 %    97.5 %
(Intercept)     0.1805173 0.0433694 0.7374630
Speed           0.9573494 0.9281210 0.9872143
Slope           0.8275090 0.7292815 0.9303408
Lane            1.5072632 1.2150088 1.9023711
Bike_Route      1.3844321 0.9184350 2.0869742
VolumeCOLLECTOR 0.8336274 0.4646651 1.5010804
VolumeSTREET    0.3412142 0.1767108 0.6626395
```

Zero Inflated Negative Binomial (ZINB) Model:

R Code:

```
> ZINB<-zeroinfl(Crash~ Speed+Slope+Lane+Volume+Bike_Route, data=Data,dist=
"negbin")
```

Model Output:

```
> summary(ZINB)

Call:
zeroinfl(formula = Crash ~ Speed + Slope + Lane + Volume + Bike_Route, data =
Data,
    dist = "negbin")

Pearson residuals:
     Min       1Q   Median       3Q      Max
 -0.4807  -0.1809  -0.1201  -0.1075 110.7639

Count model coefficients (negbin with log link):
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.279871   1.019487   2.236  0.02533 *
Speed           -0.057911   0.022458  -2.579  0.00992 **
Slope           -0.005619   0.105647  -0.053  0.95758
Lane            -0.106037   0.114798  -0.924  0.35565
VolumeCOLLECTOR -0.666485   0.432336  -1.542  0.12317
VolumeSTREET    -1.262345   0.459714  -2.746  0.00603 **
Bike_Route      -0.899831   0.287944  -3.125  0.00178 **
Log(theta)      -0.381392   0.495144  -0.770  0.44114

Zero-inflation model coefficients (binomial with logit link):
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      4.12116    1.12503   3.663 0.000249 ***
Speed            0.03267    0.02477   1.319 0.187157
Slope            0.17844    0.10898   1.637 0.101564
Lane            -1.08748    0.19868  -5.474 4.41e-08 ***
VolumeCOLLECTOR -0.58098    0.49097  -1.183 0.236681
VolumeSTREET    -0.59902    0.56534  -1.060 0.289332
```

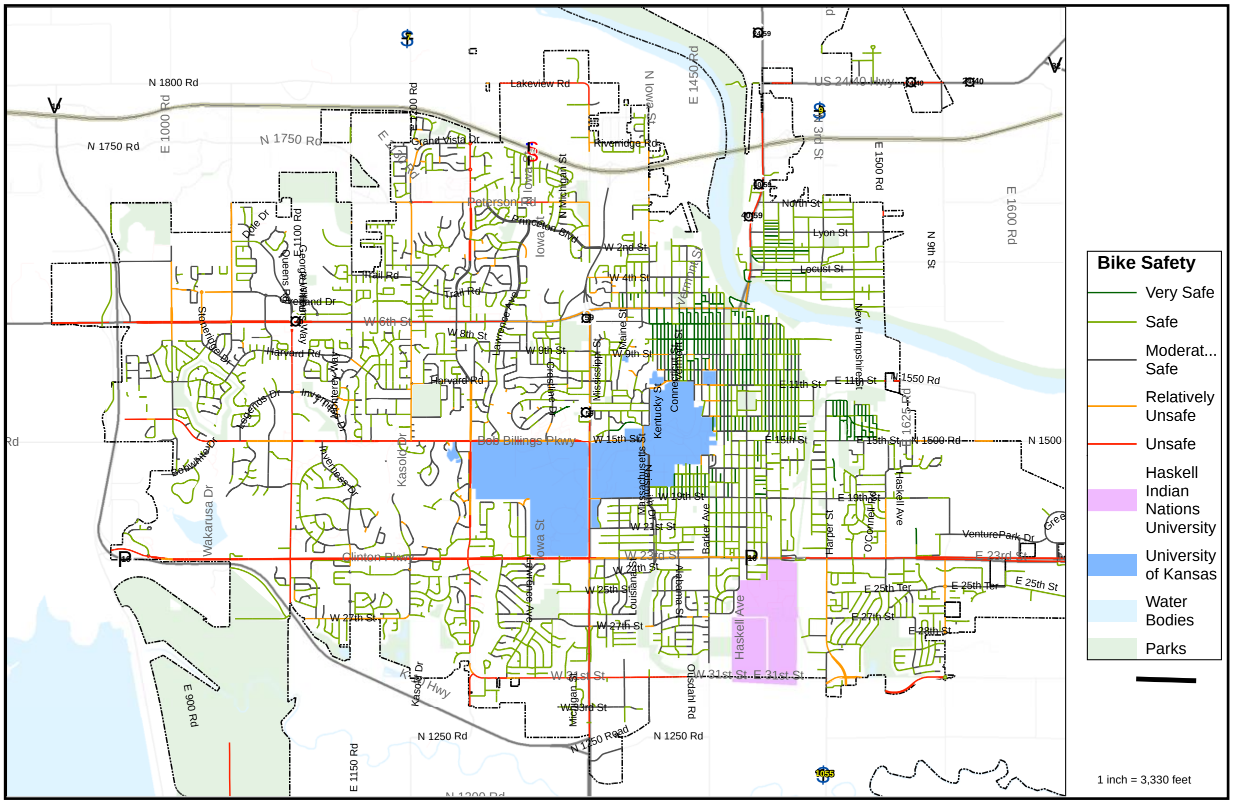
```
Bike_Route      -1.76222    0.42783  -4.119 3.81e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 0.6829
Number of iterations in BFGS optimization: 78
Log-likelihood: -894.6 on 15 Df
> exp(cbind(CO=coef(ZINB),confint(ZINB)))
                              CO       2.5 %       97.5 %
count_(Intercept)      9.7754240 1.32539856   72.0982487
count_Speed            0.9437344 0.90309456    0.9862030
count_Slope            0.9943966 0.80841120    1.2231703
count_Lane             0.8993910 0.71817866    1.1263273
count_VolumeCOLLECTOR  0.5135105 0.22006159    1.1982692
count_VolumeSTREET     0.2829896 0.11493743    0.6967542
count_Bike_Route       0.4066386 0.23126498    0.7150021
zero_(Intercept)      61.6308812 6.79471284  559.0178131
zero_Speed             1.0332077 0.98425162    1.0845989
zero_Slope             1.1953530 0.96544771    1.4800063
zero_Lane              0.3370640 0.22834682    0.4975420
zero_VolumeCOLLECTOR   0.5593500 0.21368131    1.4642012
zero_VolumeSTREET      0.5493478 0.18139800    1.6636511
zero_Bike_Route        0.1716636 0.07421819    0.3970507
```

Model Comparison:

```
> vuong(Poisson,ZIP)
-------------------------------------------------------------
              Vuong z-statistic             H_A    p-value
Raw                   -4.747265 model2 > model1 1.0309e-06
AIC-corrected         -4.563891 model2 > model1 2.5107e-06
BIC-corrected         -3.964876 model2 > model1 3.6717e-05
> vuong(ZINB,NB)
-------------------------------------------------------------
              Vuong z-statistic             H_A   p-value
Raw                   3.4384711 model1 > model2 0.0002925
AIC-corrected         2.5839216 model1 > model2 0.0048842
BIC-corrected        -0.2075769 model2 > model1 0.4177797
> vuong(ZIP,ZINB)
-------------------------------------------------------------
              Vuong z-statistic             H_A  p-value
Raw                   -1.715317 model2 > model1 0.043144
AIC-corrected         -1.715317 model2 > model1 0.043144
BIC-corrected         -1.715317 model2 > model1 0.043144
```

Appendix B: Map

Bike Compatibility Map for City of Lawrence



* Negative binomial model coefficients have been used to prepare this map.