Bicycling and the built environment: route choice and road safety

Peng Chen

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Qing Shen, Chair

Anne Vernez Moudon

Cynthia Chen

Linda Ng Boyle

Program Authorized to Offer Degree:

Interdisciplinary Ph.D. Program in Urban Design and Planning

University of Washington

Abstract

Bicycling and the built environment: route choice and road safety

Peng Chen

Chair of the Supervisory Committee:

Professor Qing Shen

Department of Urban Design and Planning

Bicycling is an environment-friendly, healthy, and low-cost transportation mode that is especially suitable for short distance travels. There is a bicycling renaissance in the North America, and Seattle takes a leading role in this change. Increasing bicycle use in the US will amplify the exposure to crashes and injuries, and the outcomes will likely depend on the bicycle route choice and the risk factors in the built environment.

Both bicyclists' route preferences and bicycling safety performance are connected to the built environment features. To understand the effect of the built environment on bicycling, this dissertation addresses three interrelated research questions: (1) what built environment features are correlated with bicycle route choice, (2) how built environment features are correlated with bicycle crash frequency and bicycle crash risk, and (3) how built environment features are correlated with bicyclist injury severity. Using Seattle's data, the research methodologies include advanced discrete choice models and count data models as key components.

The results of this dissertation research show that changing the factors of land use, demographics, road network and design, contribute to a convenient, safe, and attractive bicycling environment, which encourages more bicycle use. The most significant built environment features impacting bicyclists' route preferences and safety outcomes are: land use mixture, household density, employment density, bicycle facility types, waters and parks, commercial land use, street lights, street trees, slopes, and posted speed limit.

Several policy implications can be drawn from the aforementioned results. First, in light of bicyclists' route preferences, transportation planners should add cycle tracks and bike lanes on shortcuts in flat areas. In addition, local authorities should lower posted speed limits, improve street lighting conditions, and plant more street trees. Second, to reduce bicycle crash frequency and bicycle crash risk, local authorities should influence bicycling and driving behaviors with higher degrees of mixed land use, and place a greater percentage of commercial lands along popular bike routes. Transportation planners should encourage dense development and separate bike lanes from road traffic. Third, to mitigate bicyclist injury severity, local authorities should lower posted speed limits to reduce the risk of severe bicyclist injuries or separate bicycle lanes from road traffic. And transportation planners should, once again, advocate for the dense development and mixed land use, while improving street lighting and avoiding placing bike lanes on steep slopes in planning practice. Overall, encouraging compact development and implementing considerate roadway designs would promote safety and create a favorable bicycling environment.

Table of contents

List of figures

List of tables

Acknowledgements

I wish to express sincere appreciation to my mentor and advisor, Prof. Qing Shen, for his invaluable assistance and guidance throughout my Ph.D. life. His academic passion, insightful advice, and supportive help greatly inspired my deep interest in the field of land use and transportation planning. He acts as a role model to enlighten me of how to become a respectful scholar with charming personalities. Equally as indispensable, I thank Professors Anne V. Moudon, Cynthia Chen, and Linda Ng. Boyle for assisting me to accumulate analytical skills in research design and statistical modeling. A special 'thank you' has to be sent to Prof. Robert Plotnick for serving in the thankless role of graduate school representative.

I would also like to acknowledge the help from the faculty and staff of the Interdisciplinary Ph.D. Program under the Department of Urban Design and Planning at the University of Washington, especially Prof. Branden Born, Ms. Jean Rogers, Jan Brooks, and Larissa Maziak. Without their assistance, I would not have been able to receive continual generous financial support. In addition, other financial support was provided by the Department of Civil and Environmental Engineering, the Pacific Northwest Transportation Consortium, the Harborview Injury Prevention and Research Center, and the International Association for China Planning. Specially, gratitude goes to Professors Yinhai Wang and Charles Mock. Finally, a huge thank to my dear friends for providing me confidence, hugs, joys, encouragement, and optimisms, and reminding me that there is an appealing new life after graduation.

Dedication

To my mother, with love, heart and soul.

Chapter 1. Introduction

1.1 Problem statement

Bicycling is a green transportation mode widely advocated for by the public, authorities, and scholars. The United States (US) faces many environmental and public health problems as a consequence of longtime car dependence (Younger *et al.* 2008). Regular bicycling activities generate physical and mental benefits, such as improving fitness and reducing stress (Blair *et al.* 2001). Furthermore, an increase in the amount of bicycling reduces car use, parking demand, energy consumption, road congestion, and traffic related air-pollution (Litman and Burwell 2006, Woodcock *et al.* 2009).

In order to promote bicycle use, state and local authorities advocate building urban areas that are convenient, safe, and have attractive bicycling environments. As a result, many US cities are creating or implementing bicycle master plans to encourage bicycling, such as Seattle (Seattle Department of Transportation 2014). However, these plans focus on adding bicycle facilities. Connections between other built environment[1] features, such as factors of land use and road network, and bicycling behaviors are not adequately analyzed and discussed.

The problems related to bicycle planning, specifically for route choice and bicycle safety, are not adequately answered. Therefore, this dissertation aims to inform bicycle planning practices by addressing the following questions.

- What built environment features are correlated with bicycle route choice?

[1] The built environment, short for BE, is defined as the man-made settings for human activities. In this dissertation, the built environment includes land use and road network features.

- How are built environment features correlated with bicycle crash frequency and bicycle crash risk?
- How are built environment features correlated with bicyclist injury severity?

How bicyclists interact with the built environment is essential in deciding where to place new bicycle facilities. In order to understand bicycle route choice, it is important to investigate areas in which bicyclists are traveling to. Before the development of GPS trackers, the built environment features of bicycle routes remained ambiguous to researchers, because there was no effective way to record the bicycling environment features associated with bicyclists' preferences. Previous studies largely relied on attitudinal surveys or retrospective interviews, creating a perception of less reliable results. Emerging data collection technology helps track the observed bicycle routes, where the connections between bicycle routes and built environments can be modeled with revealed-preference data and produce more accurate estimates of factors impacting bicyclists' route preferences.

The importance of the bicycling environment in mitigating bicycle crash risk is somehow underestimated in US safety studies. Among existing studies focusing on built environment factors, bicycle route types and intersection design are given more considerations. Land use is somehow ignored. Compact development, such as higher densities and greater mixed land use, shorter trip distances, and more difficulties in owning, driving, and parking a car, are shown to be correlated with greater bicycling popularity (Pucher & Buehler, 2006, 2008). In addition, bicycle mode choice and route choice are also dependent on road safety. In fact, safety acts as the most important deterrent preventing people from bicycling because bicyclists are more vulnerable than motorists in sharing roads (Wegman *et al.* 2012, Wei and Lovegrove 2012). Mitigating the conflicts between bicyclists and motorists through improving bicycle routes may result in mode switch from driving to bicycling. Integrating the consideration of risk factors[2] into bicycle planning contributes to improved bicycle use. Improved road safety is essential to the success of bicycle plans and programs. Finally, due to the low mode share of bicycling, the requirements and techniques of bicycle planning in the US are different from countries in Europe and Asia, and their successful experiences of bicycle planning may not be directly transferable. To plan convenient, safe, and attractive US bicycling environments, the efforts to understand safety issues are essential.

1.2 Background: bicycle mode share and bicycle safety worldwide

Bicycling is an affordable transportation mode. Bicycling uses no fossil energy, relieves mental pressures, and promotes the livability of cities (Pucher and Buehler 2008). Compared with walking, bicycling covers greater travel distances at higher speeds. Compared with riding transits, bicycling provides a seamless door-to-door mobility option. Compared with driving, bicycling is more environment-friendly and physically healthy (OECD 2013). However, to support a bicycling renaissance in the US, it is important to review the trends in bicycle use around the world and the issues that have manifested with increasing bicycle use.

1.2.1 Bicycle mode share worldwide

Bicycle mode share[3] varies greatly among Asia, Europe, Australia and Oceania, South America, and North America (see *Figure 1-1)*. Megacities in countries such as the Netherlands, Germany, India, and China, have bicycling as the primary transportation mode, where they are

---

[2] Risk factor is any attribute, characteristic or exposure of an individual that increases the probability of involving an injury.

[3] Bicycle modal share, also called bicycle mode split, is the percentage of trips made using bicycles among all transportation modes. Table 1-1 shows modal shares for the bicycle, which consists of the conventional bicycle and the electric bicycle.

used for utilitarian trip purposes (Pucher and Buehler 2008, Tiwari *et al.* 2008). In comparison, the bicycle mode share remains low in countries such as the US, United Kingdom, Canada, and Australia (Moudon *et al.* 2005, Pucher *et al.* 2011). The variations in bicycle popularity can be due to economic development, transportation policies, the built environment, bicycling culture, topography, and weather (Pucher and Buehler 2008, Buehler *et al.* 2011, Gallop *et al.* 2012, Hankey *et al.* 2012, El Esawey *et al.* 2015).

The bicycle was invented in Europe, where the bicycling culture runs deep. Many European countries have built convenient, safe, and attractive bicycling environments. The Netherlands, Denmark, and Germany have taken advantage of planning and policy tools to maintain bicycling as a primary transportation mode (Pucher and Buehler 2008). However, as shown in *Table 1-1*, bicycling is no longer popular in other European countries (e.g., UK) because of automobile-oriented polices (Pucher and Buehler 2008).

Bicycling has been an important transportation mode in Asian countries since the early 20th century, and the bicycle continually serves as a primary mobility tool. However, as automobiles are becoming more affordable with continual economic growth, bicycle use is steadily decreasing in China and India (Tiwari *et al.* 2008, Mittal *et al.* 2015). For instance, bicycle mode share in Beijing dropped from 62.70% in 1986 to 14.00% in 2010 (Yang *et al.* 2014).

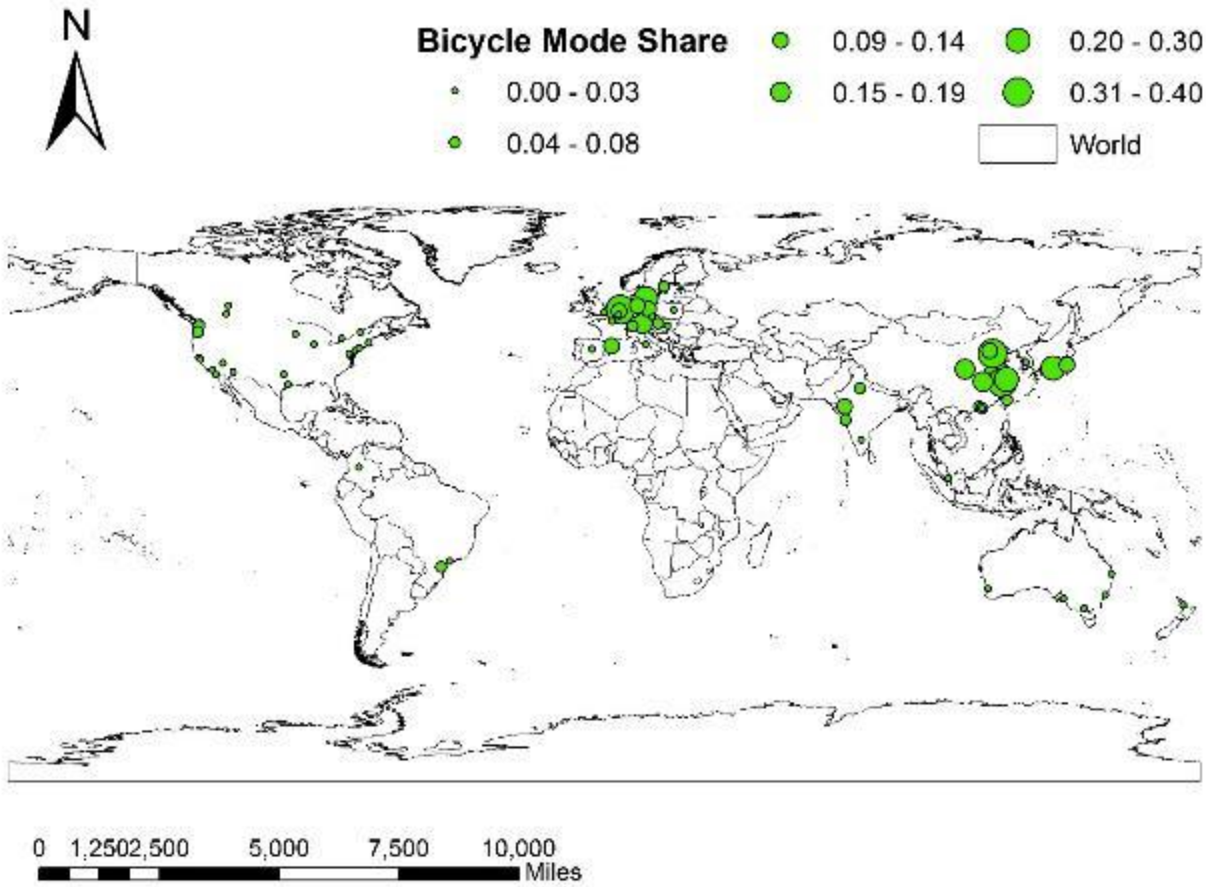


Figure 1-1: Bicycle mode share in world megacities

Source: (Singapore land transport and authority 2014, City lock magazine 2014, Charting transport 2014, European cyclists' federation 2015)

In China, the bicycle mode share is negatively associated with per-capita income but it still remains popular in many Chinese cities. Based on dense population, relatively well-planned bicycle facilities are built in some Chinese megacities. It is worth noting that electric bicycles are increasingly popular in China in recent decades because bicyclists desire faster speeds and greater mobility (Cherry 2007).

Table 1-1: Bicycle mode share in selected large cities Worldwide

City	Cycling	Year	Region
Beijing	14.00%	2013	Asia
Daejeon	2.00%	2012	Asia
Delhi	6.00%	2012	Asia
Mumbai	6.00%	2008	Asia
Osaka	24.00%	2012	Asia
Shanghai	28.00%	2014	Asia
Singapore	1.00%	2011	Asia
Taipei	5.00%	2013	Asia
Tokyo	14.00%	2008	Asia
Ahmedabad	14.00%	2008	Asia
Bangalore	3.00%	2010	Asia
Guangzhou	8.00%	2015	Asia
Hangzhou	34.00%	2007	Asia
Hong Kong	1.00%	2011	Asia
Nanjing	38.00%	2015	Asia
Shenzhen	6.00%	2010	Asia
Tianjin	37.00%	2010	Asia
Wuhan	19.00%	2008	Asia
Xi'an	18.00%	2011	Asia
Adelaide	1.00%	2011	Australia
Brisbane	1.00%	2011	Australia
Melbourne	2.00%	2011	Australia
Perth	1.00%	2011	Australia
Sydney	1.00%	2011	Australia
Auckland	1.00%	2009	Oceania
Bogota	2.00%	2008	Brazil
San Paulo	1.00%	2012	Brazil
Toronto	2.00%	2011	Canada
Calgary	3.00%	2011	Canada
Edmonton	2.00%	2010	Canada
Montreal	2.00%	2010	Canada
Vancouver	3.70%	2011	Canada
Curitiba	5.00%	2011	Columbia
Amsterdam	40.00%	2014	Europe
Barcelona	12.00%	2012	Europe
Berlin	13.00%	2012	Europe
Brussels	3.00%	2010	Europe
Budapest	2.00%	2004	Europe
Copenhagen	30.00%	2012	Europe
Hamburg	12.00%	2008	Europe
London	3.00%	2011	Europe
Madrid	0.00%	2006	Europe
Munich	17.00%	2011	Europe
Paris	3.00%	2010	Europe
Prague	1.00%	2009	Europe
Rome	0.00%	2001	Europe
Rotterdam	14.00%	2004	Europe
Stockholm	4.00%	2012	Europe
Utrecht	34.00%	2015	Europe
Vienna	6.00%	2013	Europe
Warsaw	1.00%	2009	Europe
Zurich	4.00%	2005	Europe
Boston	2.00%	2009	US
Chicago	1.00%	2009	US
Dallas	0.00%	2009	US
Houston	0.00%	2009	US
Las Vegas	0.00%	2009	US
Los Angeles	1.00%	2009	US
New York	1.00%	2009	US
Philadelphia	2.00%	2009	US
Phoenix	1.00%	2009	US
Portland	6.00%	2009	US
San Diego	1.00%	2009	US
San Francisco	3.00%	2009	US
San Jose	1.00%	2009	US
Seattle	4.00%	2013	US
Washington DC	2.00%	2009	US
Minneapolis	1.00%	2009	US

Sources:
(Singapore land transport and authority 2014, City lock magazine 2014, Charting transport 2014, European cyclists' federation 2015)

In India, the bicycle mode share varied from 7.00% to 15.00% in large cities, and from 13.00% to 21.00% in medium and small cities (Tiwari *et al.* 2008). The bicycle mode share is also negatively correlated with per capita income in India. Bicycling is the most affordable and convenient mobility mode for short distance travels for low-income people.

In the US, only 0.50% of commuters use bicycles (American Association of State Highway and Transportation Officials and US Departmemt of Transportation 2013). The majority of bicycle trips in the US are for exercise and recreational purposes (Moudon *et al.* 2005, Pucher and Buehler 2008). The past decade has observed an increase in bicycling cultures across large US cities, the scope is still small when compared to the international level. As shown in *Figure 1-2*, the percentage of bicycle commuters in the 50 states varies from 0.00% to 1.90%.

To summarize, bicycle mode share varies substantially by nations. In general, bicycle mode share decreases sharply in Asian developing countries, keeps stable in European developed countries, and observes a bicycling renaissance in North America. The popularity of bicycling depends on many factors, such as weather, season, topography, city size, travel distance (Martens 2004, Rietveld and Daniel 2004), bicycle facility, bicycling culture, transportation policy, impedance, safety concern, and relative costs of using alternative modes (Martens 2004, Rietveld and Daniel 2004, Dill 2009, Winters *et al.* 2011, Pucher *et al.* 2012).

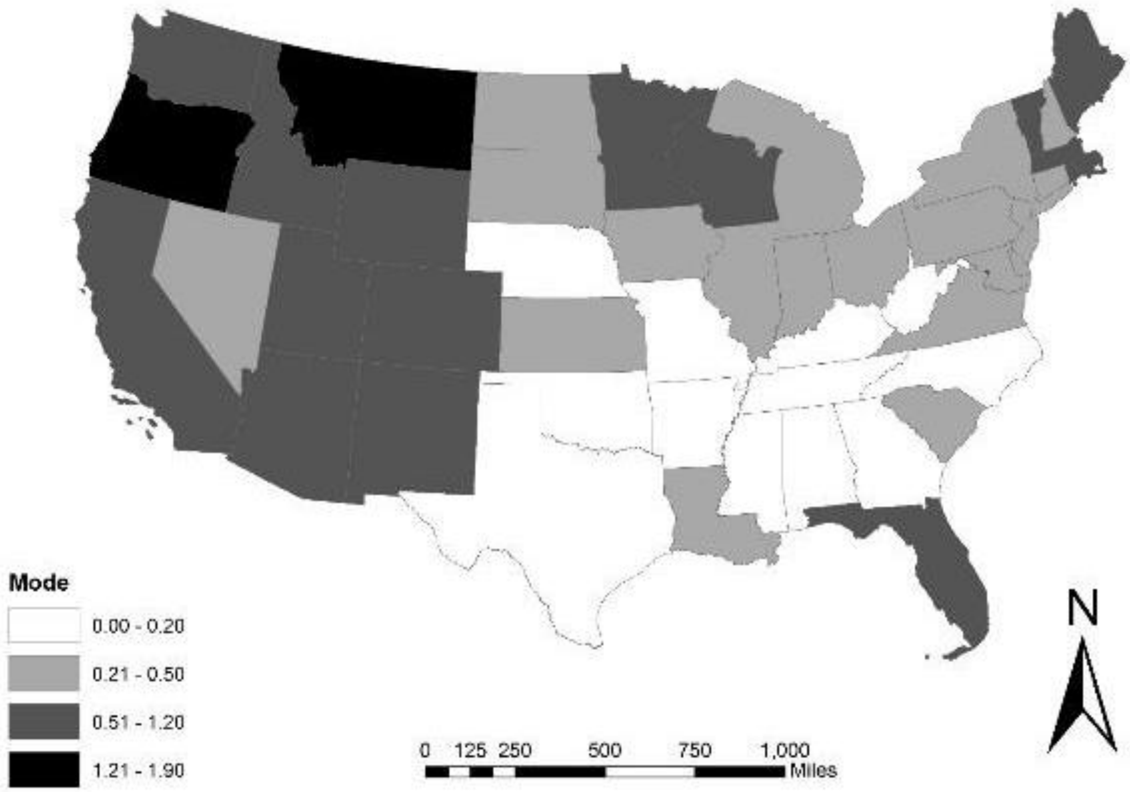


Figure 1-2: Mode share of bicycle commuters in the United States

Source: (United States Department of Transportation 2010)

1.2.2 Bicycle safety worldwide

Bicycle safety issues also vary greatly by country. In China, for instance, motorists treat bicyclists with insufficient respect. Chinese bicyclists enjoy the great mobility that is provided by electrical bicycles. However, the use of electrical bicycles induces major safety issues. Electrical bicyclists are likely to violate traffic laws, such as speeding and red-light running, which poses threats to motorists and other bicyclists (Wu *et al.* 2012, Zhang and Wu 2013).

In India, lack of dedicated bike routes and sufficient surveillance, the contributing factors of bicycle crashes are more complicated but are largely explained by ineffective traffic management. Motorcycles and vehicles easily hit bicyclists and leave the scene ("hit and run") (Mohan *et al.* 2009). For example, nearly 37.00% of reported bicycle crashes occurred on national highways in India (Tiwari *et al.* 2008). This percentage indicates that local authorities fail to prevent bicyclists from entering highways, and fail to monitor bicycle crashes on arterial routes and local streets. The number of bicyclist fatalities in Mumbai and Delhi is fewer than that of pedestrians and motorcyclists, but much more than that of the other transportation modes (Mohan *et al.* 2009). The understanding of India's bicyclist safety is inaccurate and limited in the existing literature.

Comparatively, bicycling is safer in Europe than in the US, India, and China. Existing studies have compared the relative risk of bicycle crashes to car crashes in European countries. In general, a lower degree of crash risk is associated with a larger bicycle mode share. The relative risk[4] of fatal injuries per distance of bicycle crashes to car crashes is about 6 times in Norway and the Netherlands. This relative risk is approximately 11 times in Switzerland and Denmark, and roughly 15 times in the United Kingdom and New Zealand (OECD 2013).

In the US, building separated bike lanes is not realistic or affordable in most cities because of insufficient bicycling demand. Bicyclists have to share roads with motorists and are required to wear helmets. Most bicycle crashes result from conflicts between motorists and bicyclists. As shown in *Figure 1-3*, Florida and Arizona have greater numbers of bicyclist fatalities per million people than the other states.

[4] Relative risk is the ratio of the probability of an injury happening in an exposed group to the probability of that injury happening in a comparison, non-exposed group, or the probability of one type of injury to the probability of another type of injury.

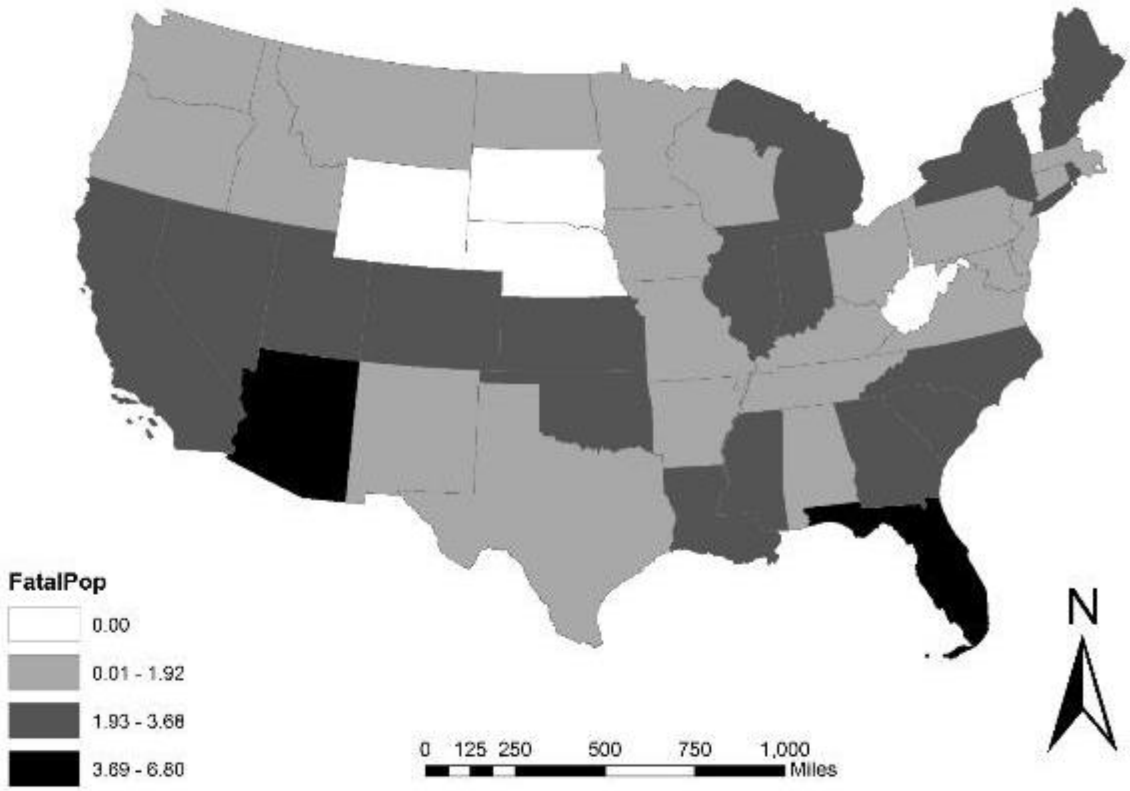


Figure 1-3: Bicyclist fatalities per million people in the United States

Source: (The National Highway Traffic Safety Administration 2013)

To summarize, bicycle mode share and safety concerns vary greatly across the world. In the US, there exists a low level of bicycle mode share and a high level of bicycle crash risk. The bicycle planning experiences of European and Asian countries may not be directly transferable to the US. Subsequently, specific research focusing on the US urban context for bicycling should be conducted. This dissertation studies how to increase bicycle use and reduce bicycle crash risk in order to produce convenient, safe, and attractive bicycling environments in the US.

1.3 Conceptual framework and research objectives

This dissertation originated from a research project to comprehend the factors impacting bicycle route choice, but the dissertation expanded to understanding the relationship between individual bicyclists and the built environment. The public's perception of the lack of safety is an obstacle to the popularity of bicycling (Reynolds *et al.* 2009). In order to make more insightful policy recommendations promoting bicycle use in the US, safety has been included as another component that impacts bicycling.

This dissertation places bicycle route choice and safety issues in the loop between individual bicyclists and the built environment, as shown in *Figure 1-4*. This loop considers the three elements linking individual bicyclists and the built environment: (1) use[5], (2) access[6], and (3) exposure[7]. The three elements are interrelated. Better planned bicycle facilities encourage more bicycle use, promote bicycling accessibility, and hence bring safety benefits to bicyclists. Many issues are related to bicycle use, such as mode choice, bicycle volume, bicycle miles traveled, and trip frequency. Bicycle route choice is the major concern of accessibility. The typical bicycle exposures in the built environment are the risks of collision, injury, and air pollution. For instance, the placement of the barriers installed between bike lanes and drive lanes

[5] "Use" describes an action or state of affairs that was done repeatedly or existed for a period of time. The use of bicycle is related to the mode choice of bicycling, bicycle volume, bicycle frequency, bicycle miles traveled, and bicycle hours traveled.

[6] "Access" refers to the ease of an individual to reach a destination. Accessibility, the measurement of access, is frequently quantified by proximity, which is a location-based distance measurement to different land uses, locations, or activities. Composite accessibility measurement considers all types of land use nearby, which reflects the accurate land use composition, but is hard to be interpreted.

[7] "Exposure" in this study refers to the behavioral or environmental factor that may cause or associate with the outcome of a bicycle crash or a bicyclist injury.

reduces the risk of collision that bicyclists are exposed to motorists. This dissertation only focuses on access and exposure.

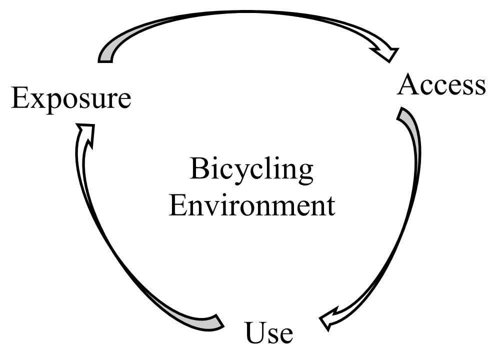


Figure 1-4: Individuals and the bicycling environment

Source: University of Washington URBDP 576 Pedestrian Travel, Land Use, and Urban Form

This dissertation contributes to the literature through the following three aspects. First, traditional route choice cost functions, such as the shortest path, do not adequately capture bicyclists' joint consideration of multiple factors on route preferences. There are many additional factors to be considered in bicycle route choice, such as leisure and safety. Second, due to the lack of appropriate denominators, existing bicycle safety studies have not sufficiently investigated bicycle crash risk. This dissertation project consequently includes a zonal-counted number of bicycle trips to measure the bicycle crash risk. Third, spatial dependence is a common statistical concern in urban studies. This dissertation addresses the concern on spatial autocorrelation in modeling.

This dissertation is aimed at understanding how built environment features are associated with bicycle route choice, bicycle crash frequency, bicycle crash risk, and bicyclist injury severity. This dissertation project employs quantitative methods to investigate the associations between the built environment and bicycling factors in the US context. It is targeted at facilitating designs for bicycling through building convenient, safe, and attractive urban environments, reducing the number of collisions and the severities of injuries, and providing resources for evidence-based policy design. The specific research objectives are listed below:

- To apply the theory of *utility maximization* to the study of bicycling, to connect compact development elements with bicycle route choice, and to use GPS-tracked bicycle routes to understand how bicyclists interact with the built environment.
- To examine the theory of *safety in numbers*, to link the built environment features to bicycle crash data, and to determine what elements influence bicycle crash frequency, bicycle crash risk, and bicyclist injury severity.
- To draw lessons from the research outcomes that can be applied to bicycle facility planning, and promote bicycle accessibility and safety.

1.4 Dissertation structure

The remainder of this dissertation is organized into three interrelated studies and a synthesis. The first study focuses on bicycle route choice in an attempt to understand how bicyclists consider convenience, safety, and leisure in choosing bicycle routes. The second study examines the *safety in numbers* theory through comparing a bicycle crash frequency model with a bicycle crash risk model, and highlights policy implications drawn from density-related measurements. The third study investigates the relationship between built environment risk factors and different types of bicyclist injuries. Finally, a summary of policy implications and future research is presented in the last chapter.

1.4.1 Analyzing bicycle route choice

The first study employs a GPS dataset gathered by a smartphone application called "CycleTracks". The bicycle traces are recorded with the GPS points. The cost functions of the recorded bicycle traces are developed by multiple principal component analyses. Two safety-related elements, including posted speed limit and bicycle facility type, are selected as the primary factors in defining those principal components. Through comparing the traces with the generated routes, the defined route alternatives are used to identify the significant built environment features that impact bicycle route choice using a path size logit model. The assumption is that bicycle route choice is jointly impacted by convenience, safety, and leisure. The objective of this paper is to provide a reference for bicycle planning through evaluating bicyclists' route preferences. This study is under review for possible publication in the journal of "Transportation Research Part A, Policy and Practice".

1.4.2 Analyzing bicycle crash frequency and bicycle crash risk

In the second study, the associations between automobile-involved bicycle crash frequency, bicycle crash risk, and built environment features are investigated by two Poisson-Lognormal random effects models. The hypothesis is that areas with higher density and more mixed land use would have higher bicycle crash frequencies but lower bicycle crash risks. Traffic analysis zone[8] is selected as the spatial unit to match the available exposure variables. This study considers the spatial autocorrelations among analytical units by employing two random effects. The objective of this chapter is to examine the theory of *"safety in numbers"* in the US context, and to provide insights to compare the variations in offering protections for road

[8] TAZ short for Traffic Analysis Zone, is the commonly employed unit for transportation planning modeling.

users on different bicycle facilities. This study has been published as an article, entitled "Built environment features in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach," in the journal *"Safety Science"*.

1.4.3 Analyzing bicyclist injury severity

The third study examines the correlation between bicyclist injury severity and the built environment using a generalized ordered logit model and a general additive model. The hypothesis is that compact developed areas have relatively lower driving speeds, hence bicycle injury severities are less serious. This study is conducted at a disaggregate level, therefore motorist and bicyclist-related factors are better explained. Specifically, injury type is assumed to follow an ordered categorical distribution, and the possible spatial dependence across crash sites is examined. The findings provide useful information for modifying the road network and land use for safety improvements. This study has been published in the journal *"Accident Analysis and Prevention"* as an article, which is titled "Built environment effects on cyclist injury severity in automobile-involved bicycle crashes."

References

American Association of State Highway and Transportation Officials and US Departmemt of Transportation, 2013. Commuting in america 2013: the national report on commuting patterns and trends. 7-8.

Blair, S.N., Cheng, Y., Holder, J.S., 2001. Is physical activity or physical fitness more important in defining health benefits? Medicine and science in sports and exercise 33 (6; SUPP), S379-S399.

Buehler, R., Pucher, J., Merom, D., Bauman, A., 2011. Active travel in Germany and the U.S.: Contributions of daily walking and cycling to physical activity. American Journal of Preventive Medicine 41 (3), 241-250.

Charting transport, 2014. What does the census tell us about cycling to work?

Cherry, C., 2007. Electric bike use in china and their impacts on the environment, safety, mobility and accessibility. UC Berkeley center for future urban transport: a volvo center of excellence.

City lock magazine, 2014. Cycling mode share data for 700 cities.

Dill, J., 2009. Bicycling for transportation and health: The role of infrastructure. Journal of Public Health Policy, S95-S110.

El Esawey, M., Mosa, A.I., Nasr, K., 2015. Estimation of daily bicycle traffic volumes using sparse data. Computers, Environment and Urban Systems 54, 195-203.

European cyclists' federation, 2015. Cycling facts and figures.

Gallop, C., Tse, C., Zhao, J., Year. A seasonal autoregressive model of vancouver bicycle traffic using weather variables. In: Proceedings of the Transportation Research Board 91st Annual Meeting.

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in minneapolis, mn. Landscape and Urban Planning 107 (3), 307-316.

Litman, T., Burwell, D., 2006. Issues in sustainable transportation. International Journal of Global Environmental Issues 6 (4), 331-347.

Martens, K., 2004. The bicycle as a feedering mode: Experiences from three european countries. Transportation Research Part D: Transport and Environment 9 (4), 281-294.

Mittal, S., Dai, H., Shukla, P., 2015. Low carbon urban transport scenarios for China and India: A comparative assessment. Transportation Research Part D: Transport and Environment.

Mohan, D., Tsimhoni, O., Sivak, M., Flannagan, M.J., 2009. Road safety in india: Challenges and opportunities.

Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L., Weather, R.D., 2005. Cycling and the built environment, a US perspective. Transportation Research Part D: Transport and Environment 10 (3), 245-261.

OECD, 2013. Cycling, health and safety.

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.

Pucher, J., Buehler, R., 2008. Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. Transport Reviews 28 (4), 495-528.

Pucher, J., Buehler, R., Seinen, M., 2011. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. Transportation Research Part A: Policy and Practice 45 (6), 451-475.

Pucher, J., Dill, J., Handy, S., 2012. Infrastructure, programs, and policies to increase bicycling: An international review. Preventive Medicine 50, Supplement (0), S106-S125.

Reynolds, C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. Environmental Health 8 (1), 47.

Rietveld, P., Daniel, V., 2004. Determinants of bicycle use: Do municipal policies matter? Transportation Research Part A: Policy and Practice 38 (7), 531-550.

Seattle Department of Transportation, 2014. Seattle master bicycle plan.

Singapore land transport and authority, 2014. Passenger transport mode shares in world.

The National Highway Traffic Safety Administration, 2013. Traffic safety facts: Bicyclists and other cyclists. http://www-nrd.nhtsa.dot.gov/Pubs/812018.pdf.

Tiwari, G., Arora, A., Jain, H., 2008. Bicycling in Asia. Transport Research & Injury Prevention Programme (TRIPP), IITDelhi.

United States Department of Transportation, 2010. FHWA’S bicycle and pedestrian program.

Wegman, F., Zhang, F., Dijkstra, A., 2012. How to make more cycling good for road safety? Accident Analysis & Prevention 44 (1), 19-29.

Wei, F., Lovegrove, G., 2012. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention (0).

Winters, M., Davidson, G., Kao, D., Teschke, K., 2011. Motivators and deterrents of bicycling: Comparing influences on decisions to ride. Transportation 38 (1), 153-168.

Woodcock, J., Edwards, P., Tonne, C., Armstrong, B.G., Ashiru, O., Banister, D., Beevers, S., Chalabi, Z., Chowdhury, Z., Cohen, A., 2009. Public health benefits of strategies to reduce greenhouse-gas emissions: Urban land transport. The Lancet 374 (9705), 1930-1943.

Wu, C., Yao, L., Zhang, K., 2012. The red-light running behavior of electric bike riders and cyclists at urban intersections in China: An observational study. Accident Analysis & Prevention 49 (0), 186-192.

Yang, M., Wang, Q., Zhao, J., Zacharias, J., Year. Rise and decline of the bicycle in Beijing. In: Proceedings of the Transportation Research Board 93rd Annual Meeting.

Younger, M., Morrow-Almeida, H.R., Vindigni, S.M., Dannenberg, A.L., 2008. The built environment, climate change, and health: Opportunities for co-benefits. American journal of preventive medicine 35 (5), 517-526.

Zhang, Y., Wu, C., 2013. The effects of sunshields on red light running behavior of cyclists and electric bike riders. Accident Analysis & Prevention 52 (0), 210-218.

Chapter 2. Built environment effects on bicyclists' route preferences: a GPS data analysis

Abstract

The goal of this study presented in this chapter was to examine the effects of the built environment on bicycle route choice using data from Seattle. This goal was achieved using a smartphone application called "CycleTracks" that records GPS data and associated bicycle route features. Features of land use and road network favored by bicyclists are then identified and analyzed. A route choice set is modeled by the labeling route approach, and a path size logit regression is then estimated. There were six major findings: (1) bicycle route choice involves joint consideration of convenience, safety, and leisure; (2) most bicyclists prefer to cycle on short, flat, and well-planned bicycle facilities; (3) some bicyclists prefer routes with low posted speed limits; (4) a substantial percentage of bicyclists prefer routes surrounded by low floor area ratio; (5) some bicyclists prefer routes surrounded by mixed land use, or near waters and parks; and (6) some bicyclists favor routes planted with street trees and installed with street lights. These conclusions offer valuable insights for planning strategies that facilitate efficient, safe, and comfortable bicycling.

2.1 Background

Bicycling creates economic, environmental, and social benefits by reducing travel costs, decreasing traffic-related pollution, and promoting public health (Pucher and Dijkstra 2003, Sælensminde 2004, Weichenthal *et al.* 2011). According to the American Community Survey, 55 major US metropolitan areas witnessed an increase in the amount of bicycle commuters in the past decade (ACS 2013). Transportation planners face many challenges in attempting to meet this undeniable bicycling growth. For instance, transportation planners must consider where and what types of bicycle routes should be built and how different built environment features encourage bicycling.

Many studies have been conducted to address the above concerns. The effects of bicycle facility provisions on bicycling behaviors have been examined (Tilahun *et al.* 2007, Akar and Clifton 2009, Dill 2009, Chen *et al.* 2012, Teschke *et al.* 2012). As agreed, if new bicycle lanes are built, bicyclists will use them (Dill and Carr 2003). The correlations between bicycle facility changes and bicycling risk factors have been investigated (Parkin *et al.* 2007, Cheng and Liu 2012, Washington *et al.* 2012). For instance, the placement of safety countermeasures is correlated with increased number of bicyclists (Chen *et al.* 2013). A number of studies explored how policy interventions could encourage more bicycling (Dill and Carr 2003, Pucher *et al.* 2012), such as the effect of public bicycle sharing programs (Shaheen *et al.* 2010, Faghih-Imani *et al.* 2014). Additionally, several researchers used stated-preference data to analyze bicycle route choice (Stinson and Bhat 2003, Hunt and Abraham 2007, Sener *et al.* 2009). Due to the limitation in tracking bicycle routes, the interaction between the built environment and bicycle route choice has not been sufficiently investigated in the previous literature. This study leverages new data collection technology to track bicycle routes and extract relationships between the built environment and bicycle route choice.

Many local agencies adopt plans and programs to promote bicycling due to the health and environmental benefits. For example, the City of Seattle has added bicycle lanes throughout Seattle. However, traditional four-step travel demand forecasting models are incapable of effectively supporting bicycle planning. Attaining detailed data related to bicycle route choice is key to advancing travel demand forecasting models. In particular, promoting the modeling capacity is important for the fast-paced propagation of bicycle programs. The success of programs such as the “safe routes to school” program, which has already been adopted in many US states to endorse travel safety for children and teenagers, and to promote active transportation (Boarnet *et al.* 2005, Buckley *et al.* 2013), depends on bicycle planning. Advancements in travel data collection technology, such as GPS applications, generate new opportunities for enhanced understanding of bicycle route choice and create opportunities for more accurate models.

2.2 Literature review

A bicycle route choice model supports bicycling behavior analysis and acts as an indispensable element of trip assignment. Many previous studies used stated-preference data to examine the associations between road network and bicycle route choice. GPS data-based route choice model is anticipated to offer a more precise understanding of the relationship between bicycling behavior and the built environment by employing revealed-preference data. The following sections present the frequently discussed topics of bicycle route choice, including factor selection, choice set generation, and discrete choice modeling.

2.2.1 *Factors influencing bicycle route choice*

Bicycle facility type[9], slope, trip distance, posted speed limit, street lighting, and road signal density are essential elements impacting bicycle route choice (Hunt and Abraham 2007, Sener *et al.* 2009, Menghini *et al.* 2010, Broach *et al.* 2012). Studies focused on stated-preference data emphasize bicyclists' characteristics, street parking, crosswalks, stop signs, pavement quality, and the presence of separated bicycle lanes on bridges (Stinson and Bhat 2003, Sener *et al.* 2009). In general, bicyclists favor shortcut routes detached from vehicle traffic.

2.2.2 *Choice set generation*

Path search methods have explicit procedures for the route choice set generation, which are divided into deterministic and stochastic approaches (Bovy 2009). K-shortest paths and labeling routes are deterministic approaches to generate route alternatives. The K-shortest path[10] approach is restricted to minimize trip distance, which is inadequate for capturing other factors impacting bicycle route choice. The labeling route[11] approach is endorsed for producing a better bicycle choice set because this approach exploits various segment[12] features and creates more realistic route alternatives. In recent bicycle route

---

[9] Bicycle facility type is commonly classified as cycle track, bike lane, and bike boulevard. More detailed descriptions about bike facilities are stated in WSDOT's instruction (WSDOT, 2012. WSDOT road design manual: Bicycle facilities chapter 1520). The order of bicyclists' preference is cycle track, bike lane, bike boulevard, and arterial route.

[10] The K-shortest path routing algorithm is a generalization of the shortest path problem. "Path" is a similar term as "route".

[11] The labeling route approach minimizes costs by creating functions through a linear combination of factors, and labels a route with the prominent factor.

[12] A road segment is defined by the link between two consecutive nodes.

choice studies, a calibration process was applied to improve the existing labeling route methods (Broach *et al.* 2010). Link elimination[13] and link penalty[14] are simulation (stochastic) methods for enumeration (Bekhor *et al.* 2006, Bovy 2009, Prato 2009). Link elimination is not an ideal method for enumeration because it may result in an unsolvable road network. For example, if the bridges connecting north and south Seattle are disregarded in the link elimination process, no additional route alternatives can be created based on the remaining road network. In terms of the number of generated comparable routes, the link penalty approach is better. However, when bicycling distance is long and road network is made of densely connected streets, large numbers of generated route alternatives can be unmanageable.

Assuming that segment length was the primary factor impacting bicycle routing, Broach et al. (2010) conducted a stepwise simulation to acquire a rich bicycle route choice set, and then allocated labels to the added factors, such as maximizing all bike facilities and minimizing upslope. This approach generated realistic routes for great overlap with observed traces[15]. Similarly, Ehrgott (2012) advanced a generic term of suitability by merging several factors and made a bicycle route cost function based on travel time and a suitability score.

---

[13] Link elimination approach is described as continually eliminating the shortest path segments from the road network to find the next best route until converged.

[14] Link penalty approach is labeled as repeatedly increasing the impedances of the shortest path segments to search the next best route until converged.

[15] Trace: a trace is a type of visible mark left by a passenger, such as a footprint. The difference between a trace and a route lies in whether the footprints have been snapped to the road network or not.

2.2.3 Discrete choice modeling: path size logit model

The correlations[16] among created routes remain a primary focus in route choice research. More sophisticated models are required to eliminate the problem of identical and irrelevant alternatives. Current GPS-based bicycle route choice studies select the path size[17] logit model for analysis (Menghini *et al.* 2010, Broach *et al.* 2012). The path size logit model is a variant of the multinomial logit model, which includes a spatially measured route similarity index. Therefore, the path size logit model requires less computation as compared to the other advanced discrete choice models. Also, the path size logit model works efficiently in identifying similarities among candidate routes by accounting for the issue of independence of irrelevant alternatives. Besides the path size logit model, there are other methods for route choice modeling, such as the mixed logit model (Prato 2009).

2.2.4 GPS-tracked bicycle route choice research

Only three GPS data-based bicycle route choice studies are available in the current literature. Those studies were carried out in Zurich (Menghini *et al.* 2010), Portland (Broach *et al.* 2010, Broach *et al.* 2012) and San Francisco (Hood *et al.* 2011). Menghini's study (2010) used the simulated K-shortest path approach with link elimination to create route alternatives. Yet, this study only took segment length, bicycle facility type, slope, and street lighting into account for the cost functions[18], which may

---

[16] Correlation refers to the degree of independence of each alternative route. If two routes have overlapped road segments, they are correlated.

[17] Path size is a spatial index that describes the portion of a route overlapping with other route alternatives.

[18] Transportation cost refers to the sum of input that a passenger pay to reach a destination, such as monetary cost, time cost, and physical effort cost. A cost function is a curve expressing the sum of input.

not be adequate to understand the observed route features. Broach et al.'s study (2010) advanced the calibrated labeling route approach to create route alternatives, which improved the quality of the choice set considerably. However, the independent variables were restricted to traffic volume, the number of turns, slope, and bicycle facility types.

2.3 Research objective and data sources

The objective of this research is to comprehend the effects of built environment features, such as the characteristics of land use and road network, on bicycle route choice. This study employs GPS data from Seattle bicyclists to empirically determine bicycle route choice characteristics that are important to travel demand forecasting.

Seattle is making efforts to become bicycle friendly city. Currently, 4.10% of commuters use bicycles as a primary transportation mode in Seattle, which is higher than that of the US national average (Seattle Department of Transportation 2013). The city is determined to support the mode switch from driving to bicycling.

The GPS data were acquired from a smartphone application called "*CycleTracks*," and the GIS information of the built environment was collected from related agencies. As observed bicyclists' routes are obtained, the origins and destinations of the GPS traces are identified. Then the built environment features are quantified in ArcGIS. The unit of analysis for quantifying the built environment is road segment, and the unit of analysis for discrete choice modeling is route. The approach of labeling route is employed to generate bicycle route alternatives, which make up the sensible and feasible bicycle route choice set. Additionally, a path size logit model is created to estimate the effects of the built environment on bicyclists' route preferences. This study contributes to the field by: (1) measuring built environment features along with the GPS-tracked bicycle routes; (2)

offering a practical method for processing smartphone GPS data; (3) assimilating safety-related indicators as primary factors to generate cost functions using principal component analyses; and (4) taking convenience, safety, and leisure into account for bicycle route choice.

2.4 Research design

2.4.1 Theoretical assumptions

The theory of modeling bicycle route choice is the *utility maximization* theory. Bicycle route choice involves the joint consideration of multiple objectives. The assumed objectives of bicycle route choice in this study are simplified to (1) convenience, (2) safety, and (3) leisure. The first objective, convenience, emphasizes the efficiency concerning bicycle route choices, such as the shortest path or shortest travel time. The second objective, safety, focuses on bicyclists' concerns about minimizing risks in bicycling. Some safety-related elements are intrinsically associated with crash frequency and injury severity, such as posted speed limits and bicycle facility types. The third objective, leisure, highlights the pleasure gaining from bicycling, such as the attractiveness of a route (Hunt and Abraham 2007).

Under the guidance of the *utility maximization* theory, a GPS-data based bicycle route choice model involves the procedures of (1) processing raw GPS data to traces, and traces snapping to the road network, (2) generating cost functions using principal component analyses, (3) creating alternative routes using selected cost functions, (4) quantifying built environment features and selecting variables, and (5) modeling bicycle route choice using the path size logit regression. The detailed research design is expressed by *Figure 2-1*.

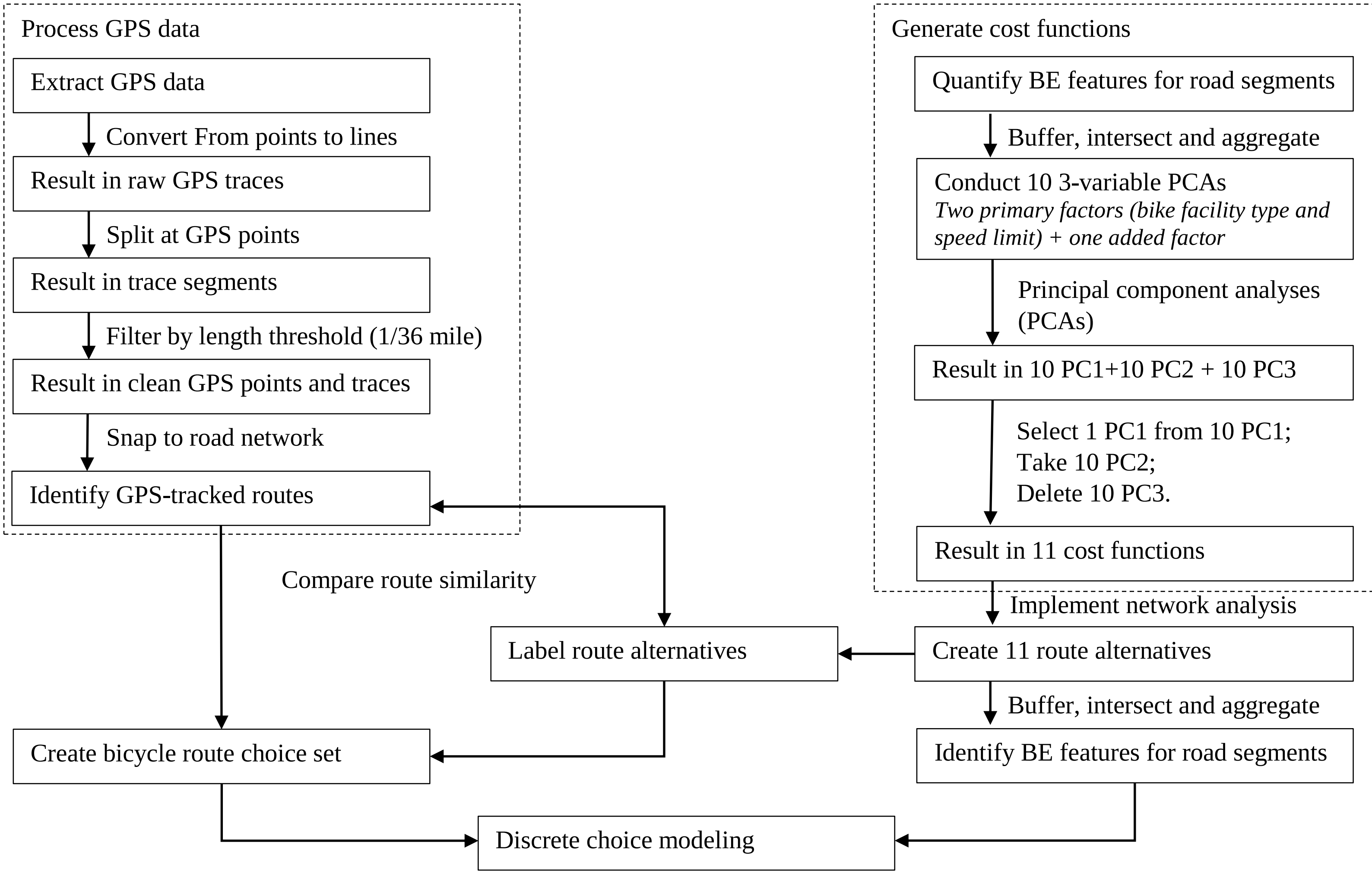


Figure 2-1: Research design

2.4.2 GPS data and processing

The GPS data used in this study were gathered by a smartphone application *"CycleTracks",* which was developed by the San Francisco County Transportation Authority. This application demonstrates that smartphone apps can be operative tools in gathering bicycle route data. The GPS data were collected for the city of Seattle between 11/18/2009 and 03/24/2014 (3.5 years). A total of 4.9 million GPS points were converted to 3,310 routes for 197 bicyclists, as shown in *Figure 2-2*. All app-users are recruited as bicyclists. The application archives one GPS point every two seconds. The reported variables are age, gender, and bicycling frequency. In addition, travel time is simply calculated as the difference between the starting time and the ending time. Employing this GPS data helps better comprehend the bicyclists' route preferences by clearly tracing the route chosen and then associating built environment features along the entire trips.

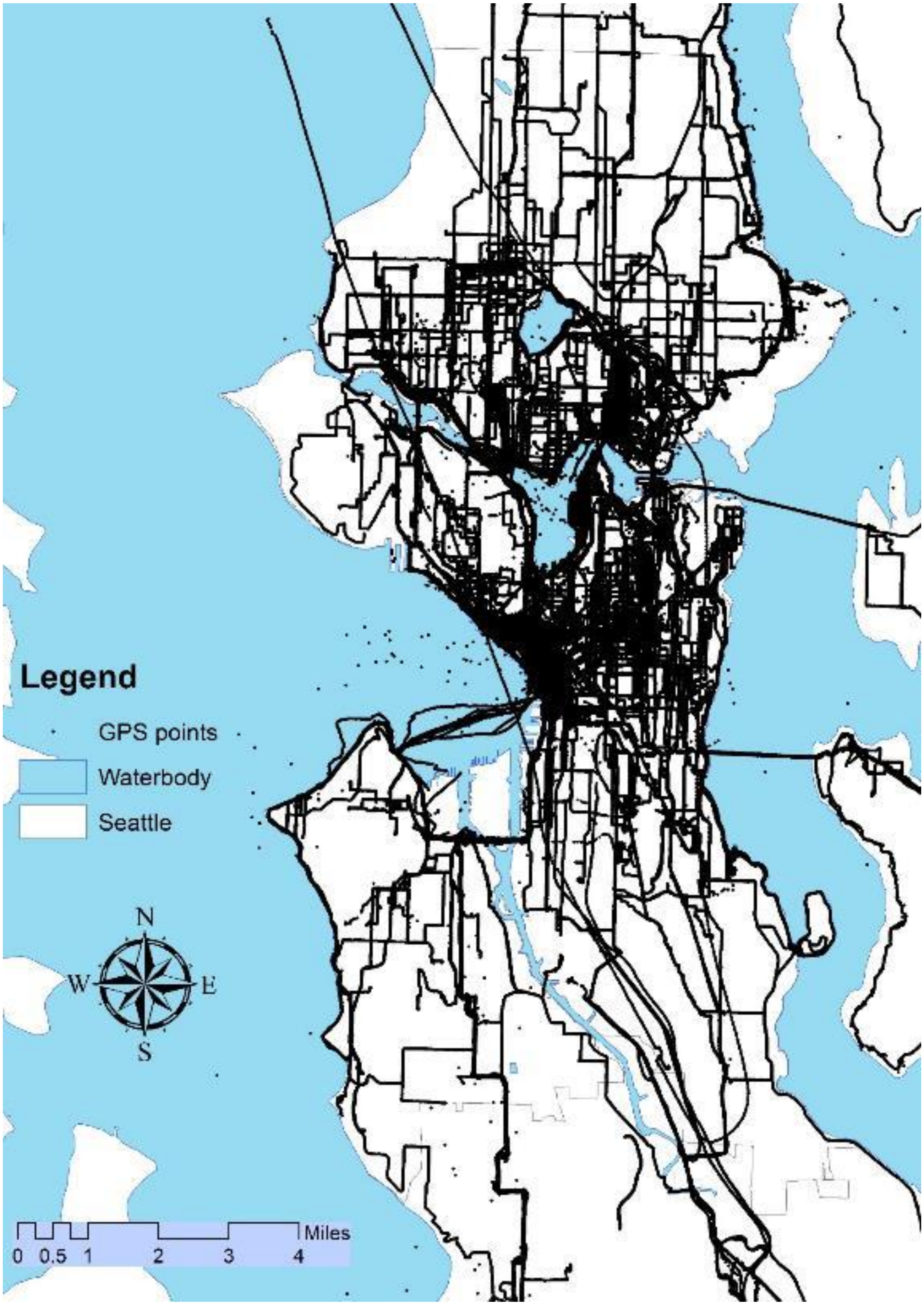


Figure 2-2: Raw data imported from “CycleTracks”

A limitation of the smartphone GPS data was the number of errors observed. The urban canyon[19] effect was evident in the traces and the GPS-tracked routes in areas where there was a cluster of tall buildings. As noted in *Figure 2-2*, reflected signals result in the GPS points jumping away from the traces. Therefore, in this study, several algorithms are used to clean, choose, and snap GPS points to the road network in ArcGIS. First, GPS points are converted to lines where many irregular traces are observed. Second, the converted lines are split at the GPS points, and the distance of each trace segment is thereby measured. The distance of each trace segment between two GPS points higher than a certain threshold is defined to filter problematic points, resulting a sample of 2,922 clean routes. This study assumes that a temporal bicycling speed is impossible to be greater than 50 mph, and the trace segment between two consecutive GPS points cannot be longer than 1/36 mile.

Issues regarding sampling validity require a selection of the responded bicyclists' traces. First, over-representing bias is involved in the sample. Similar routes are identified from bicyclists who reported multiple times, whereas some bicyclists reported only one or two times. Second, some short routes seem to be a portion of a trip. Bicyclists may terminate recording due to an insufficient battery, weak signals, or other factors. Third, rounding tours[20] are not appropriate to be selected for modeling bicycle route choice. It is impossible or ineffective to create realistic routes for such traces.

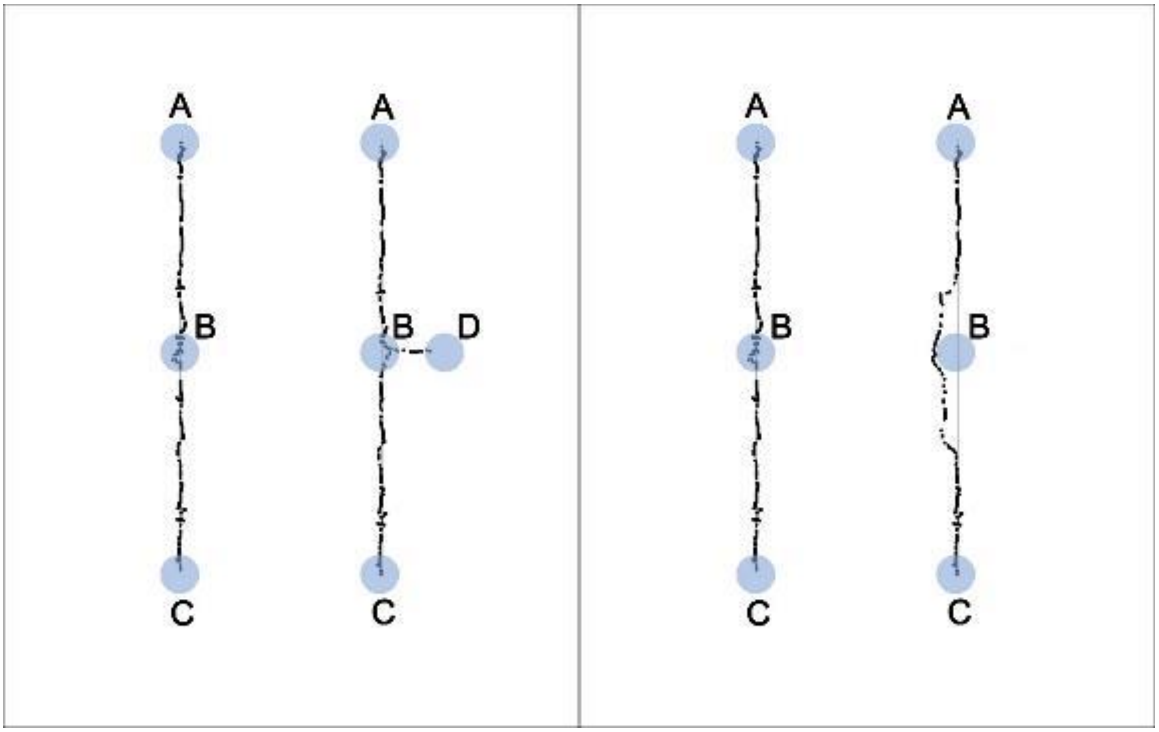
---

[19] Urban canyon effect could be described as smartphone signals being blocked in places where the streets are flanked by buildings.

[20] A rounding tour is defined as a trip starting and ending at a same location or geographically approximating locations.

This selection process creates several possible biases. First, in the sampling procedures, short distance travels are underrepresented. Trip distance less than one mile is disregarded. There is no effective way to discern whether the trips are of real short distances, or are partially recorded due to weak signals or smartphone running out of batteries. Second, long distance bicycle trips (more than five miles) are not so frequent, but are overrepresented in the final sample. The sampling process drops out many short distance routes contributing to the average longer trips in the final sample. For instance, the longest trip distance in the GPS-tracked routes is 22.67 miles, which is reported by an experienced bicyclist. Therefore, a significant assumption made in this study is that experienced bicyclists share similar bicycling environment preferences as regular bicyclists.

To snap the traces to the road network, the GPS points are overlaid with road intersection buffers. Two consecutive intersection IDs are then used to identify the corresponding road segment IDs. Errors occur when GPS points are snapped to inappropriate neighboring intersections or missing correct intersections. For instance, A, B and C are three points on segments AB and BC. If GPS points are snapped to B’s neighbor D, there would be four segments in that route, including AB, BD, DB and BC, as shown in *Figure 2-3 (a)*. Furthermore, assuming B as the breaking/turning point of segments AB and BC, the missing point B result in an incorrect segment AC, which may not exist in the road network, as shown in *Figure 2-3 (b)*. ArcGIS editing tools are employed to correct these small errors.



(a) (b)

Figure 2-3: Two types of GPS data snapping errors

Due to limitations in the data, explained later, traces were filtered based on: each responded bicyclist must have a valid user ID, and each selected trace must be unique for a user. Moreover, "dirty" traces, rounding tours, and multimodal trips were eliminated from the sample. For traces crossing the city boundaries, only the portion inside Seattle was kept. The filtering procedures result in a sample consisting of 544 traces, as presented in *Figure 2-4*.

Figure 2-4: Snapped bicycle routes

2.4.3 Principal component analysis for cost function generation

Principal component analysis is a commonly used multivariate method to identify the underlying structure of a set of variables by considering the correlations among variables. In this study, the intention of using principal component analysis lies in its capability of combining multiple variables by extracting the data (Jolliffe 2002). In other words, a principal component analysis replaces original variables by creating a set of new orthogonal variables through linear combination (Abdi and Williams 2010), and provides a procedure for generating cost functions for bicycle route choice. The principal component analyses are employed to explore the factor compositions within the built environment features and develop composite cost functions in this study.

A commonly observed challenge in developing cost function for route choice is that many qualitative factors are measured as ordered categorical variables. Ordered categorical variables do not have variations for a given category, such as a particular type of bike facility. If two alternative segments have the same type of bike facility, their costs cannot be differentiated using this categorical variable alone. As a solution, principal component analysis can integrate the ordinal variables with continuous variables through linear combination to differentiate the costs[21] of such road segments.

A composite cost function can be defined by combining multiple variables, and a label can be assigned to this cost function. The principal component analysis involves three steps in this study: (1) identifying the primary factors, (2) choosing the added factors, and (3) defining the cost functions and labeling the routes.

---

[21] Transportation cost refers to the sum of costs that a passenger pays to reach a destination, such as monetary cost, time cost, and physical effort cost. A cost function is a curve expressing this sum.

In the first step, two safety-related factors are selected as primary factors for the principal component analyses based on previous literature. Safety is a key concern in bicycle route choice (Hunt and Abraham 2007, Winters *et al.* 2011a, Winters *et al.* 2011b, Ehrgott *et al.* 2012, Teschke *et al.* 2012), which is measured by posted speed limits and bicycle facility types in this study. Bicycle facility types have four levels and are recorded on an ordered categorical scale, which include cycle tracks[22], bicycle lanes[23], bicycle boulevards[24], and arterial routes. The posted speed limits were divided into six levels, including 0, 20, 25, 30, 35 and 40 mph. Traffic volume and the number of lanes are the two other possible factors to quantity road safety. They are excluded from this study due to the lack of required data on local streets. The built environment features are measured in the unit of road segment and standardized[25] for the principal component analyses.

Second, following the idea of labeling route approach, the two safety-related primary factors are integrated with the ten added factors accordingly using principal component analyses, as shown in *Tables 2-1 and 2-2.* Among the composite principal

---

[22] Cycle track is a type of separated route dedicated to bicycling and walking. A cycle track is commonly placed next to a major street, but separated by a curb, a hedge or other physical barriers.

[23] Bike lane is an element of the paved arterial marked with painted lines. Bike lanes are designated exclusively for cyclists, but parallel with drive lane and street parking. In Washington State, the bike lane has three types, including protected bike lanes, buffered bike lanes, and conventional bike lanes.

[24] Bike boulevard is signed as being a bike route in low volume local streets, and may have traffic circles or speed bumps at intersections. Bicyclists use the same lanes with motorists when bicycling in bike boulevard.

[25] Standardization is to rescale the factors into a range of 0 to 1 by

$$X' = (X - \min(X)) / \text{range}(X)$$

Where X' is the standardize value, X is the reported value, min(X) is the minimum X, and range(X) is the difference between the maximum X and minimum X.

component functions, the route label is assigned to the factor with the greater loading. More factors have been considered in this study, but excluded for multicollinearities, such as road signals, crosswalks, and stop signs. Traffic circles[26], parking signs, and bus stops are excluded for generating cost functions of unrealistic alternatives. *Table 2-1* presents the ten added factors that are selected.

[26] The traffic circle contains raised concrete circles centered at local street intersections, and is designed to calm down vehicles as they enter a neighborhood.

Table 2-1: Two primary factors and ten added factors for principal component analyses

Primary factors	Unit
1. Bicycle facility types	Cycle track (0), separated/buffered/conventional bike lane (1), bike boulevard (2), and arterial route (3).
2. Speed limit	Mph
Added factors	
1. Segment length	Miles;
2. Segment slope	Degrees;
3. Floor area ratio[27]	Average floor area ratio per mile in 500-feet buffers;
4. City features	Number of city features per mile in 500-feet buffers, including churches, community centers, libraries, art centers, play grounds, schools, and theatres;
5. Water and parks	Percentage of water and parks per mile in 500-feet buffers, including rivers, lakes, seas, wetlands, and parks;
6. Street lights	Number of street lights per mile in 50 feet buffers;
7. Land use mixture[28]	Land use mix entropy along each segment, measured in 500-feet buffers;
8. Street trees	Number of street trees per mile in 50 feet buffers;
9. Road intersections	Number of intersections per mile;
10. Green space and residential lands	Percentage of residential lands and green spaces in 500-feet buffers.

[27] FAR short for Floor Area Ratio is calculated by the total area of a building divided by the area of the lot.

[28] LUM short for Land Use Mixture, reflects the level of integration within a given area of different types of land uses, which may include residential, office, commercial, water and parks, and public space. Land use mixture is measured by an entropy index, expressed as:

$$\textit{Land use mixture} = -\Sigma(P_i * \ln P_i)/\ln n$$

Where n is the number of different land use types, and P_i is the proportion of land in type i. The resulting variable land use mixture is an entropy index, which varies from 0 (homogeneous land use) to 1 (most mixed) land use.

The third step is to implement the ten three-variable principal component analyses[29], which eventually result in the eleven cost functions, displayed in *Table 2-2*. The fifth column in *Table 2-2* presents the variance proportion of the selected principal components.

The first principal component in each of the ten principal component analyses is explained primarily by the safety-related primary factors, i.e. bike facility types and posted speed limits, and their weights are similar in each case. Therefore, this study only selects one out of the ten first principal components (PC1) to represent the cost function determined mainly by bicycle facility types and posted speed limits, plus a third factor, for example segment length (as shown in the first row listed in *Table 2-2).* The ten second principal components (PC2) are mostly explained by the added factors. Therefore, ten additional cost functions are established, each consists of an added factor and one or two contributing primary factors. These ten additional cost functions are shown from *the 2nd row* to *the 11th row* in *Table 2-2*).

The coefficients of each function are generated by the ten principal component analyses. For example, the quantification of the first cost function listed in *Table 2-2* is expressed by

$$PC1_{segment} = 0.697*Speed\ Limits + 0.695*Bike\ Facility\ Types - 0.178*Segment\ Length$$

Similarly, the quantification of the second cost function listed in *Table 2-2* is expressed by

[29] These analyses produce similar weights in the 10 3-variable principal component analyses, of 53%, of 33%, and of 14%, in the three principal components. Any of the first two principal components in the 10 principal component analyses explain 86% of the variances.

$$PC2_{segment} = 0.111*Speed\ Limits + 0.140*Bike\ Facility\ Types + 0.984*Segment\ Length$$

Table 2-2: Eleven cost functions generated from principal component analyses (PCAs)

Route alternative labels	Weight of primary factors		Weight of added factor	Variance explained by PC1 or PC2 in ten PCAs
	Speed limit	Bike facility type		
1. Max. prioritized bicycle facilities (minimize speed limits simultaneously with segment length as added factor)	0.697	0.695	-0.178	53.72% (PC1)
2. Min. trip distance	0.111	0.140	0.984	32.70% (PC2)
3. Min. trip slopes	-0.140		0.990	33.30% (PC2)
4. Min. floor area ratio			0.998	33.39% (PC2)
5. Max. city features			0.999	33.35% (PC2)
6. Max. water and parks	0.165	0.188	0.968	31.98% (PC2)
7. Max. street lights		-0.243	0.970	33.27% (PC2)
8. Max. land use mixture		-0.155	0.988	33.32% (PC2)
9. Max. street trees			0.993	33.60% (PC2)
10. Min. road intersections			0.998	33.25% (PC2)
11. Max. residential lands and green spaces		0.121	0.993	33.34% (PC2)
Note: PC1, the first principal components; PC2, the second principal components.				

2.4.4 Route choice set generation

A choice set consists of selected routes and a set of sensible route alternatives. The number of route alternatives for an OD pair can be unmanageable given a dense urban street network. The labeling route (deterministic) generation is preferred to obtain a route choice set of a reasonable size.

Based on the spatial overlay, the selected alternatives are defined by the alternatives overlapped the most with the GPS-tracked routes, while the remained candidate routes are included for comparison in the final model. Yet, challenges occur when overlapped road segments in two or more route alternatives are identical. In total, 15 out of 544 OD pairs have two or more identical alternative routes. These 15 OD pairs are therefore removed for offering limited or no insights to differentiate the features correlated with bicycle routes. Route alternatives with no overlap of the GPS-tracked routes are not observed in the sample.

In the process of generating route alternatives, the methods to minimize and maximize costs are slightly different. Some functions are directly employed to minimize the costs of road segments, such as the route labeled by minimizing intersection density. However, some functions are reversed when maximizing their costs.

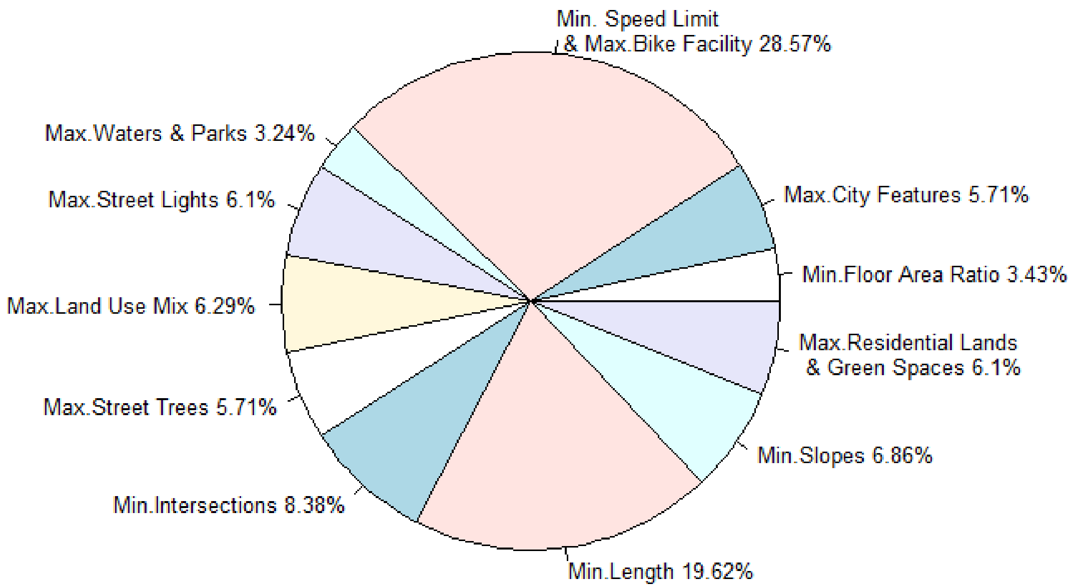


Figure 2-5: Percentages of labeled route alternatives in the defined choice set (n=525)

Figure 2-5 shows the percent of each selected route alternative in the choice set. The alternative maximizing prioritized bicycle facilities takes the greatest share (28.57%) regarding the degrees of overlapping with the GPS-tracked routes. The alternative minimizing trip distance denotes 19.62% in the selected routes. These two labeled route alternatives nearly take a half of the defined choices.

2.4.5 Variable selection

The spatial overlay function in ArcGIS is used to identify the road segment-based built environment features. The built environment features were measured at the segment level, but are aggregated to the trip level in the final model. Floor area ratio, land use mixture, city feature density, and percentage of waters and parks in 500-feet buffers surrounding the road segments are included in the final model. As specified in *Table 2-3*, the road network factors included are the densities of street trees, intersections, street lights, and average posted speed limit.

Several individual-specific factors were considered in the model such as age, gender, bicycling frequency, bicycling speed, and distance to the closest bicycle facilities from origins and destinations. These factors were eliminated from the model because they did not show a statistically significant impact. Hence, the individual-specific information cannot be used to conclusively explain bicyclists' route preferences. The trip purpose reported from the *"CycleTracks"* application was not considered reliable given that the data were self-reported with a high likelihood of reporting the wrong trip purpose. The app-users can slide the phone screen, and select one out of eight trip purposes. However, app-users could easily report wrong trip purposes, and there is no incentive to correct any wrong inputs. Many app-users actually reported similar routes for different trip purposes.

2.4.6 Path size logit model

The path size logit model is chosen to examine the correlation between the built environment and bicyclists' route preferences. The path size logit model has become increasingly popular for route choice studies for its advantage in accounting for spatial similarity. The probability function of a path size logit model is expressed by *Equation 2-1:*

$$P_k = \frac{exp(V_K + \beta_{PS} * lnPS_k)}{\sum exp(V_l + \beta_{PS} * lnPS_l)}$$ *Equation 2-1*

Where PS_k and PS_l are the path size of routes *k* and *l*, β_{PS} is the estimated parameter. P_k is the probability of alternative route *k* being chosen. V_k and V_l are the utilities of routes *k* and *l* respectively. The total utility of choosing alternative route *k* is equal to the utility of *k's* fixed covariates V_k, and the utility resulting from the correlations among alternatives, which is explained by *log (path size)*.

Equation 2-2 shows how to calculate the path size of route *k*. Path size describes the proportion of a route overlapping with other route alternatives, which is depending on the overlapping frequency across route alternatives. The path size of a unique route is equivalent to *1*, while the path size of *n* replicated routes is *1/n* (Ben-Akiva and Bierlaire 1999, Bovy 2009, Prato 2009).

$$PS_k = \sum \frac{L_\alpha}{L_k} \frac{1}{\sum \delta_{\alpha l}}$$ *Equation 2-2*

Where L_α is the length of overlapped segments; L_k is the length of the whole route; $\sum \delta_{\alpha l}$ refers to the occurrences that one road segment of a route is overlapped with the other candidate routes. Path size ranges from *0* to *1*. This path size logit model is simulated with R program using "mlogit" package (Croissant 2012). The elasticities of this path size logit (multinomial logit) model are computed by *Eq. 2-3*:

$$E = \beta(1 - P) * \overline{X}$$ *Equation 2-3*

2.5 Descriptive analysis

Table 2-3 displays the descriptive statistics of selected variables of the eleven route alternatives. The age of responding bicyclists ranges from 20 to 61, of which

72.00% are males. The mean bicycling frequency[30] is 3.23, suggesting that most respondents bicycled daily or several times each week. The average bicycling time is 31. 89 minutes (more than a half hour). Therefore, the respondents are more experienced bicyclists.

The mean length of generated alternatives is 5.33 miles, 0.57 miles longer than the average distance of GPS-tracked routes, which was 4.76 miles. The longest GPS-tracked route is 22.67 miles, while its longest generated alternative route is 22.65 miles.

The average floor area ratio and land use mixture are, respectively, 1.46 and 0.28 in the sample, greater than the city mean of 1.29 and 0.22 correspondingly. Therefore, bicycling is more popular in densely developed areas. The average posted speed limit on these routes is 18.30 mph. The path size ranges from 9.00% to 100.00% in the sample.

[30] Bicycling frequency in this study are coded as "daily"(4), "several times per week"(3), "several times per month"(2), and "less than once a month"(1).

Table 2-3: Variable dictionary and data summary (n=525)

Variable	Mean	S.D.	Min.	Max.	Description
Trip distance	5.33	3.57	0.51	22.65	Distance of generated alternatives, miles;
Speed limit	18.30	7.75	0.00	36.40	Average driving speed limit, mph;
Slope	2.57	1.26	0.05	7.55	Average slope, in percent;
Land use mixture	0.28	0.09	0.03	0.58	Land use mixt entropy, ranging from 0 to 1;
Floor area ratio	1.46	0.71	0.17	7.09	Average floor area ratio;
City feature density	11.89	8.20	0.00	62.45	Number of city features per mile, including churches, community centers, environment education places, libraries, art centers, play grounds, schools, and theatres;
Proportion of waters & parks	0.09	0.07	0.00	0.53	Percentage of water or public parks in 500-feet buffers, including rivers, lakes, seas, wetlands and parks;
Bus stop density	4.09	3.12	0.00	27.74	Number of bus stops per mile in 50 feet buffers;
Prioritized bike lane	0.51	0.27	0.00	1.00	Proportion of cycle tracks and bike lanes of each route alternative
Bike boulevard	0.23	0.22	0.00	1.00	Proportion of bike boulevard of each route alternative
Light density	60.99	23.66	10.10	340.62	Number of street lights per mile in 50 feet buffers;
Tree density	87.47	71.07	0.00	1047.02	Number of street trees per mile in 50 feet buffers;
Path size	0.45	0.25	0.09	1.00	*A ratio of spatial similarity;*
Log (path size)	-0.97	0.59	-2.40	0.00	The log of path size;
Age	39.01	9.18	20.00	61.00	*Age;*
Gender	0.72	0.45	0.00	1.00	*Gender: female (0), male (1);*
Bicycling frequency	3.23	0.81	1.00	4.00	*Frequency: "daily"(4), "several times per week"(3), "several times per month"(2), and "less than once a month"(1);*
Travel time	31.89	19.95	4.35	124.97	*Bicycling time per trip, minutes.*
Variables marked in *"Italic"* are excluded from the final model.					

Table 2-4: Modeling outcome of path size logit for eleven route alternatives (n=525)

	Convenience			Safety			Leisure				
Variable	Min. trip distance	Min. intersection	Min. slope	Max. prioritized bicycle facility	Max. street lights	Max. LUM	Max. waters & parks	Max. trees	Min. FAR	Max. residential lands & green spaces	Max. city features
Trip distance	-0.31***	-0.48***	-0.47***	-0.45***	-0.84***	-0.37***	-0.57***	-0.62***	-0.29**	-0.33***	-0.39*
Speed limit	0.03	-0.07	-0.04	0.00	0.07	0.04	0.04	-0.10***	0.00	0.00	-0.16*
Slope	-0.77**	-0.29	-0.46	-0.96***	-0.91**	-1.05***	-1.19***	-0.66***	0.18	-0.72***	-0.46
LUM	-3.36	0.95	2.04	6.38*	5.18	-4.71	1.67	-1.23	1.20	-0.78	2.52
FAR	-1.68**	-1.13*	-0.19	-1.40*	-0.73*	-0.37	-1.37*	0.07	-1.16	-0.49	-1.00
City feature density	0.06	-0.02	-0.03	-0.02	0.00	0.00	-0.01	-0.06	-0.04	-0.05	-0.02
Proportion of waters & parks	-12.6*	-0.80	2.62	4.39	8.45	-6.19	-0.02	9.07**	-5.78	3.34	-8.00
Prioritized bike lane	4.00**	1.13	2.13	3.98**	3.73*	5.44***	3.99***	1.02	0.45	1.52	7.11**
Bike boulevard	2.22	0.85	2.63	-0.53	2.98	2.16	-0.68	0.59	-0.21	-1.35	1.07
Light density	0.00	0.00	0.01	0.02	-0.02	-0.01	0.01	0.06***	0.00	0.00	-0.01
Tree density	0.00	0.01*	-0.01	0.01	0.00	0.01	0.01	0.00	-0.01	0.02**	0.04***
Log(path size)	0.08	-0.11	0.96	2.56***	3.28*	1.16**	0.53	1.19**	0.42	0.01	2.74**
Indicated by estimates with level of significance (*<0.05, **<0.01, ***<0.001)											
Model fit	Log-likelihood: -882.12										

Table 2-5: Calculated elasticities

	Convenience			Safety			Leisure				
Variable	Min. trip distance	Min. intersection	Min. slope	Max. prioritized bicycle facility	Max. street lights	Max. LUM	Max. waters & parks	Max. trees	Min. FAR	Max. residential lands & green spaces	Max. city features
Trip distance	-1.18	-2.34	-2.33	-1.93	-4.20	-1.85	-2.94	-3.12	-1.49	-1.65	-1.96
Speed limit											-2.76
Slope	-1.41			-1.98	-2.20	-2.53	-2.96	-1.60		-1.74	
LUM				1.44							
FAR	-1.75	-1.51		-1.64	-1.00		-1.94				
City feature density											
Proportion of waters & parks	-0.81							0.77			
Bus stop density											
Prioritized bike lane	1.46			1.63	1.79	2.60	1.97				3.42
Bike boulevard											
Light density								3.45			
Tree density		0.80								1.64	3.30

2.6 Inferential analysis

To summarize, this research suggests the importance of maximizing convenience, safety, and leisure in bicycle routing. Firstly, alternatives labeled as maximizing riding in prioritized bike facilities and minimizing trip distance account for 28.52% and 19.70% of all the selected routes, which aligns with the general expectation. Additionally, the most important features affecting bicyclists' route choices are trip distance, slope, floor area ratio, the proportion of riding in prioritized bicycle facilities, and street tree density.

In particular, most bicyclists prefer short, flat, and prioritized bicycle facilities (defined as cycle tracks and bike lanes in this study). Some bicyclists favor routes with low speed limits, and routes with street lights and street trees. In terms of land use features, some bicyclists prefer routes surrounded by low floor area ratio, near waters and parks, or mixed land use. Other factors, including the densities of bus stops and city features, and the proportion of riding in bike boulevards, do not explain bicyclists' route preferences. The above results of the path size logit model are presented in *Table 2-4.*

To account for the statistical concern of IIA (independence of irrelevant alternatives) issue among route alternatives, the *log (path size)* is included in the final model. As shown in the results, the estimates of the *log (path size)* are significant for the route alternatives labeled by maximizing riding in prioritized bicycle facilities, maximizing city features, maximizing street lights, maximizing street trees, and maximizing land use mixture. These results suggest that these five created route alternatives are spatially overlapped. In other words, a great proportion of the five route alternatives is made of the same road segments.

To further interpret the results, the following sections split the eleven route alternatives by the assumed threefold-objective, namely convenience, safety, and leisure. The categories are as follows. (1) Route alternatives mainly considering convenience include minimizing trip distance, minimizing intersections, and minimizing slopes. (2) Route alternatives focusing more on safety are maximizing riding in prioritized bicycle facilities, maximizing street lights, and maximizing land use mixture. (3) Route alternatives primarily interested in accommodating bicyclists' leisure are minimizing floor area ratio, maximizing residential lands and green spaces, maximizing waters and parks nearby, maximizing street trees, and maximizing city features. It is worth noting that some alternatives could be assigned to either objective, for example, bicyclists spend fewer efforts biking in flat routes, but the alternative may also be attributed to flat routes being safer to biking. However, another classification will not alter the essential interpretation.

Table 2-5 presents the elasticities calculated for the significant variables. The elasticity indicates the probability change of one route alternative resulting from one percent change of an independent variable. Because all the independent variables are continuous measures, the elasticities in this study are calculated using the equation of direct elasticity for the multinomial logit model, as shown in *Eq. 2-3.*

2.6.1 Route alternatives considering convenience

Among the three route alternatives considering convenience, the alternative labeled by minimizing trip distance is the most important one because it has more significant fixed effects and riders most frequently select this route among the three. The probability of selecting this alternative is negatively correlated with trip distance, slope, floor area ratio, and the proportion of waters and parks nearby, but is positively associated with the proportion of riding in prioritized bicycle facilities. For this alternative, trip distance, slope, and prioritized bicycle facilities have elasticities of the smallest magnitudes. A 1.00% increase in these measures is correlated with -1.18%, -1.41%, and 1.46% changes in the probabilities of choosing this alternative respectively. It suggests that choosing this alternative is most inelastic to changes of the significant fixed effects. In addition, choosing this alternative is negatively correlated with the proportion of waters and parks nearby. A possible explanation is that shortcuts are less likely to be close to waters and parks with Seattle's topography.

Compared to the dominant convenience alternative, the alternative labeled by minimizing slopes does not an add explanation to bicyclists' route preferences. On the other hand, the alternative labeled by minimizing intersections has consistent trip distance and floor area ratio effects with the dominant convenience alternative. It is worth noticing that street tree density is positively correlated with choosing this alternative.

2.6.2 Route alternatives considering safety

The three safety-focused route alternatives show largely consistent fixed effects, and spatially overlap regarding the results of the *log (path size).* As a result, any one of these alternatives is representative for the other two.

For the alternative labeled by maximizing riding in prioritized bicycle facilities, trip distance, slope, and prioritized bicycle facilities have greater elasticities than those of the alternative labeled by minimizing trip distance. A 1.00% increase in these measures is correlated with -1.93%, -1.98%, and 1.63% changes in the probabilities of choosing this alternative accordingly. These findings verify that if you build good bicycle facilities (short, flat, and separated), people will use them. In addition, a 1.00% increase in land use mixture is correlated with 1.44% increase of choosing this route.

The alternative labeled by maximizing street lights is the most elastic (-4.20) to the change of trip distance, indicating that long distance trips are much less likely to be on routes defined by maximizing street lights. Also, this alternative is the most inelastic (-1.00) to the change of floor area ratio. Perhaps buildings provide lighting in densely developed areas rather than street lights.

2.6.3 Route alternatives considering leisure

While the fixed effects are similar to the previous alternatives considering safety, the alternative labeled by maximizing waters and parks nearby has shown greater elasticities. For example, a 1.00% increase of trip distance is associated with a 2.94% decrease of choosing this alternative, which is the second largest across the elasticities of the significant fixed effects. Also, a 1.00% increase of floor area ratio is correlated with a 1.97% decrease of choosing this alternative.

The alternative labeled by minimizing floor area ratio does not provide additional explanation to bicyclists' route preferences. The other three route alternatives considering leisure, including maximizing residential lands and green spaces, maximizing city features, and maximizing street trees, all indicate consistent relationships with the core measurements of bicycle route choice (trip distance, slope, and proportion of riding in prioritized bicycle facilities). It is worth noting that street tree density is positively correlated with probabilities of choosing the routes labeled by maximizing residential lands and green spaces, and maximizing city features. In particular, a 1.00% increase of street tree density is associated with a 1.64% and 3.30% increment in the probabilities of choosing these two routes, respectively. Other findings include that a 1.00% increase in the proportion of waters and parks nearby is correlated with a

0.77% increment and a 1.00% increase of street light density is associated with a 3.45% increment of choosing the route labeled by maximizing street trees accordingly.

2.7 Conclusions and limitations

The study presented in this chapter provides insights on the relationship between bicyclists' route preferences and the built environment, which fills the gap in the traditional four-step models that has commonly ignored bicycle route choice. This smartphone-recorded GPS data offer evidence of how bicyclists choose preferred routes in urban settings. The results suggest that a GPS data-based bicycle route choice model can be integrated into a travel demand forecasting model, and can be used to assist bicycle planning.

In a nutshell, this study identifies five factors as the core contributing elements of a convenient, safe, and comfortable bicycling environment: trip length, slope, floor area ratio, prioritized bicycle facilities, and trees. It is worth noting that the significance of posted speed limit is largely undermined by its multicollinearity with prioritized bicycle facilities. Additionally, some bicyclists prefer routes surrounded by mixed land use, or near waters and parks.

Based on the above findings, local authorities should build more separated bicycle facilities with shortcuts and flat routes to promote bicycling. This conclusion is consistent with Menghini et al.'s research (2010). Another policy implication is that local authorities should isolate bicycle routes from vehicle traffic, place new routes in areas with relatively low floor area ratio, and plant street trees. Moreover, local authorities should encourage mixed land use and improve street lighting conditions of the existing bicycle facilities. In addition, low posted speed limits should be encouraged when mixing with bicycle traffic, which is again largely consistent with previous research (Broach *et al.* 2012). However, there are also tradeoffs between bicyclist and motorists as the latter may be displeased with speed reduction because of their decreased utility for driving. Hence, adding cycle tracks and separated or buffered bike lanes could be a better solution than simply lowering posted speed limits.

Several limitations are noted in this research. First, traffic volume and the number of lanes are only available on Seattle's arterials and highways. Otherwise, more safety-related indicators could be accounted for cost functions and possibly creating a greater route choice set that captures the observed bicycle route features. Second, the smartphone GPS data over-represent experienced bicyclists' behaviors, which may not be completely representative for regular bicyclists. Third, smartphone GPS data have many small errors. Thus, further quality enhancements of smartphone gathered GPS data are required for future research. Fourth, the aggregation from the segment level to the trip level raises the issue of "regression towards the mean". More disaggregate analyses for route choice are expected.

To enrich future research, the joint use of different transportation modes should be considered, such as bicycling integrated with riding public transit. Such integration is under-investigated in the existing research, possibly due to restrictions in gathering valid data. Building more attractive bicycling environments needs integration of bicycles and transits.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2 (4), 433-459.

ACS, 2013. American community survey. using American FactFinder.

Akar, G., Clifton, K.J., 2009. Influence of individual perceptions and bicycle infrastructure on decision to bike. Transportation Research Record: Journal of the Transportation Research Board 2140 (1), 165-172.

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice set generation algorithms for route choice models. Annals of Operations Research 144 (1), 235-247.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. Handbook of transportation science. Springer, pp. 5-33.

Boarnet, M.G., Day, K., Anderson, C., Mcmillan, T., Alfonzo, M., 2005. California's safe routes to school program: Impacts on walking, bicycling, and pedestrian safety. Journal of the American Planning Association 71 (3), 301-317.

Bovy, P.H., 2009. On modelling route choice sets in transportation networks: A synthesis. Transport Reviews 29 (1), 43-68.

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 (0).

Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. Transportation Research Record: Journal of the Transportation Research Board 2197 (1), 89-97.

Buckley, A., Lowry, M.B., Brown, H., Barton, B., 2013. Evaluating safe routes to school events that designate days for walking and bicycling. Transport Policy 30 (0), 294-300.

Chen, L., Chen, C., Ewing, R., Mcknight, C.E., Srinivasan, R., Roe, M., 2013. Safety countermeasures and crash reduction in New York city experience and lessons learned. Accident Analysis & Prevention 50 (0), 312-322.

Chen, L., Chen, C., Srinivasan, R., Mcknight, C.E., Ewing, R., Roe, M., 2012. Evaluating the safety effects of bicycle lanes in New York city. American Journal of Public Health 102 (6).

Cheng, Y.-H., Liu, K.-C., 2012. Evaluating bicycle-transit users’ perceptions of intermodal inconvenience. Transportation Research Part A: Policy and Practice 46 (10), 1690-1706.

Croissant, Y., 2012. Estimation of multinomial logit models in r: The mlogit packages. R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf.

Dill, J., 2009. Bicycling for transportation and health: The role of infrastructure. Journal of Public Health Policy, S95-S110.

Dill, J., Carr, T., 2003. Bicycle commuting and facilities in major U.S. Cities: If you build them, commuters will use them. Transportation Research Record: Journal of the Transportation Research Board 1828 (-1), 116-123.

Ehrgott, M., Wang, J.Y., Raith, A., Van Houtte, C., 2012. A bi-objective cyclist route choice model. Transportation research part A: policy and practice 46 (4), 652-663.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (bixi) in Montreal. Journal of Transport Geography 41, 306-314.

Hood, J., Sall, E., Charlton, B., 2011. A gps-based bicycle route choice model for San Francisco, California. Transportation letters 3 (1), 63-75.

Hunt, J.D., Abraham, J., 2007. Influences on bicycle use. Transportation 34 (4), 453-470.

Jolliffe, I., 2002. Principal component analysis Wiley Online Library.

Menghini, G., Carrasco, N., Schussler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. Transportation Research Part A: Policy and Practice 44 (9), 754-765.

Parkin, J., Wardman, M., Page, M., 2007. Models of perceived cycling risk and route acceptability. Accident Analysis & Prevention 39 (2), 364-371.

Prato, C.G., 2009. Route choice modeling: Past, present and future research directions. Journal of Choice Modelling 2 (1), 65-100.

Pucher, J., Dijkstra, L., 2003. Promoting safe walking and cycling to improve public health: Lessons from the Netherlands and Germany. American journal of public health 93 (9), 1509-1516.

Pucher, J., Dill, J., Handy, S., 2012. Infrastructure, programs, and policies to increase bicycling: An international review. Preventive Medicine 50, Supplement (0), S106-S125.

Sælensminde, K., 2004. Cost–benefit analyses of walking and cycling track networks taking into account insecurity, health effects and external costs of motorized traffic. Transportation Research Part A: Policy and Practice 38 (8), 593-606.

Seattle Department of Transportation, 2013. 2013 Seattle bicycle master plan.

Sener, I., Eluru, N., Bhat, C., 2009. An analysis of bicycle route choice preferences in Texas, US. Transportation 36 (5), 511-539.

Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia: Past, present, and future. Transportation Research Record: Journal of the Transportation Research Board (2143), 159-167.

Stinson, M.A., Bhat, C.R., 2003. Commuter bicyclist route choice: Analysis using a stated preference survey. Transportation Research Record: Journal of the Transportation Research Board 1828 (1), 107-115.

Teschke, K., Harris, M.A., Reynolds, C.C., Winters, M., Babul, S., Chipman, M., Cusimano, M.D., Brubacher, J.R., Hunte, G., Friedman, S.M., 2012. Route infrastructure and the risk of injuries to bicyclists: A case-crossover study. American Journal of Public Health 102 (12), 2336-2343.

Tilahun, N.Y., Levinson, D.M., Krizek, K.J., 2007. Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. Transportation Research Part A: Policy and Practice 41 (4), 287-301.

Washington State Department of Transportation, 2012. WSDOT design manual: Bicycle facilities chapter 1520.

Washington, S., Haworth, N., Schramm, A., 2012. Relationships between self-reported bicycling injuries and perceived risk of cyclists in Queensland, Australia. Transportation Research Record: Journal of the Transportation Research Board 2314 (-1), 57-65.

Weichenthal, S., Kulka, R., Dubeau, A., Martin, C., Wang, D., Dales, R., 2011. Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. Environ Health Perspect 119 (10), 1373-8.

Winters, M., Davidson, G., Kao, D., Teschke, K., 2011a. Motivators and deterrents of bicycling: Comparing influences on decisions to ride. Transportation 38 (1), 153-168.

Winters, M., Teschke, K., Grant, M., Setton, E.M., Brauer, M., 2011b. How far out of the way will we travel? Transportation Research Record: Journal of the Transportation Research Board 2190 (1), 1-10.

Chapter 3. Built environment features in explaining automobile-involved bicycle crash frequencies and risks

Abstract

The objective of this study is to examine the relationship between built environment features and automobile-involved bicycle crash frequencies versus risks. The method employed is the Poisson-Lognormal random effects model using hierarchal Bayesian estimation. The City of Seattle is selected for empirical analysis. The traffic analysis zone, short for TAZ, is selected as the unit of analysis to quantify the built environment features. The assembled dataset provides a large set of variables, including factors of road network and design, land use, and travel demand. The research questions are twofold: how are the built environment features associated with bicycle crash frequencies and risks, and are the TAZ-based bicycle crash frequencies versus risks spatially correlated? The findings of this study are: (1) safety improvements should focus on places with more mixed land use, and greater proportions of commercial lands; (2) off-arterial bicycle routes are safer than on-arterial bicycle routes; (3) TAZs with more road signals are likely to have more bicycle crashes and greater risks; (4) TAZs with more total trips have more bicycle crash frequencies, but less bicycle crash risks; and (5) compact urban environment has lower bicycle crash risks. For policy implications, the results suggest that the local authorities should lower the posted speed limits, encourage compact development, and separate bicycle lanes from road traffic.

3.1 Introduction

Regular cycling activity generates physical and mental benefits, such as losing weight, reducing stress, and improving fitness (Clark *et al.* 1998). An increase in the number of bicyclists can reduce car dependence, demand for parking spaces, energy consumption, road congestion, and traffic-related air-pollution.

Regarding the bicycling status in the US, increased popularity of cycling was observed as the number of bicycle trips doubled from 1.7 billion in 2001 to 4.0 billion in 2009 (bikeleague, U.S. Department of Transportation 2009). However, the percentage of bicyclist fatalities steadily increased from 1.50% in 2003 to 2.20% in 2012 (The National Highway Traffic Safety Administration 2013). Bicyclists are more vulnerable road users than motorists (Wegman *et al.* 2012, Wei and Lovegrove 2012). Most of US bicyclists inevitably rode close to automobiles on the roads, and plenty of bicyclists were killed by cars though they were wearing helmets. Though the number of bicyclists is increasing, overall the bicycle volume remains low in the US. Only 0.50% commuters rode bicycles in 2013 (American Association of State Highway and Transportation Officials and US Departmemt of Transportation 2013).

In the US, more research on the relationship between cycling safety and the built environment is needed. Firstly, in explaining the causes of bicycle crashes, prior studies weighted motorist and bicyclist-related factors as important elements, such as helmet use (Attewell *et al.* 2001, Walker 2007). Most of those studies were conducted at the micro-level focusing on individual bicyclists. However, bicyclists in Europe are not required to wear helmets and the bicycle crash risk is lower than that in the US (Teschke *et al.* 2012). Certain built environment features can explain the causes of bicycle collisions beyond motorist and bicyclist-related factors. The macro area-based studies, which highlighted the effects of the built environment features, were greatly underestimated and insufficiently investigated (Siddiqui *et al.*

2012). Secondly, among the bicycle safety studies having connections with the built environment, many of them worked on bicycle facility types (Harris *et al.* 2011, Chen *et al.* 2012, Teschke *et al.* 2012), while relatively less effort has been placed on area-wide land use features. Thirdly, research findings from other countries with compact development may not be applicable to the US. US cities are characterized as low density land use pattern, high degrees of motorization, and low bicycle volume. Therefore, the cycling environment features vary significantly between the European and US cities. Fourthly, studies on the motorist and bicyclist-related factors provide insights for education programs and policy enforcement to reshape driving and cycling behaviors. Research on built environment features contributes to lowering bicycle collision risks through engineering modifications in the road environment. Findings from these two types of research are mutually supportive.

The research objectives of this study are twofold: (1) to explore the effects of the built environment features on bicycle crash frequencies and bicycle crash risks at the TAZ level; (2) to account for the unobserved heterogeneity and spatial dependence among TAZs by modeling two random effects employing the Poisson-Lognormal models. The remainder of this chapter is organized into four sections, starting with a review of the literature, followed by the description of data sources and geo-spatial unit selection, the descriptive and inferential analyses, and ending with a discussion, limitations, and conclusions.

3.2 Literature review

3.2.1 Key definitions

In this study, a bicycle crash is defined as a collision between a bicycle and an automobile. The crash frequency, also known as the incidence, is the number of collisions at a certain location or area per unit time. The incidence rate, also known as the risk, is commonly calculated by the number of crashes reported per 1,000 trips, 1,000 hours or 1 kilometer of exposures (de Geus *et al.* 2012). In this study, the bicycle crash risk is calculated as the number of crashes divided by the corresponding TAZ-counted number of bicycle trips. However, many risk factors are not quantifiable, but are explained by surrogate measures. In this study, cycling risk is explained using factors associated with the built environment features, motorist, and bicyclist. Risks can result in travel resistances such as perceptions in non-safe environments (Schepers *et al.* 2013).

3.2.2 The theory of safety in numbers

"Safety in numbers" is a commonly cited theory as a policy reference to improve bicycle mode share and to support a bicycle plan. According to this theory, there is a non-linear relationship between the number of bicycle crashes and bicycle volume. This theory was developed based on an international comparison study between the European and US cities (Jacobsen 2003, Elvik 2009). Under the framework of this theory, as more bicyclists use the roads, motorists become more aware of their existence and slow down to avoid potential conflicts; and the expectations are that the bicycling environment would be even safer. Bhatia and Wier questioned the applicability of this theory in the US (Bhatia and Wier 2011). Blindly encouraging bicycling by simply inferring the *"safety in numbers"* theory can mislead and diminish the attention on potential environmental hazards. In addition, local authorities must evaluate the adaptability and capability of the built environment features in supporting bicycle programs. Till the relationships between the built environment and bicycle crash frequencies versus risks are identified, it is premature to justify the causal inference of the *"safety in numbers"* theory.

3.2.3 Relationships between built environment features and bicycle crash frequency

A large number of studies have investigated the relationships between built environment features and bicycle crash frequencies. For the lack of appropriate denominators, bicycle crash risk has barely been examined.

The unit of analysis in prior bicycle crash frequency research varies extensively, such as traffic analysis zones (Siddiqui *et al.* 2012, Wei and Lovegrove 2012), census tracts (Narayanamoorthy *et al.* 2013), grid-based structures (Gladhill and Monsere 2012), and locations (Wang and Nihan 2004, Schepers *et al.* 2011, Zahabi *et al.* 2011, Strauss *et al.* 2013, Vandenbulcke *et al.* 2014). Besides, prior research has considered a large set of explanatory variables to investigate the bicycle crash frequency in connection with the built environment features, which include the factors of road network and land use. In addition, factors of travel demand are also included.

Regarding road network features, the densities of intersections, roadways, and bicycle lanes have been included for modeling. Among different types of intersections, positive associations between intersection density and bicycle crash frequency are confirmed (Siddiqui *et al.* 2012, Wei and Lovegrove 2012, Strauss *et al.* 2013). In addition, complex intersections increase the probability of involving bicycle collisions (Vandenbulcke *et al.* 2014). As for the effects of roadway density, more drive lanes and bicycle lanes are positively associated with the number of bicycle crashes (Wei and Lovegrove 2012). Of different types of bicycle facilities, off-road bicycle lanes are safer than on-road bicycle lanes (Reynolds *et al.* 2009, Teschke *et al.* 2012, Hamann and Peek-Asa 2013), and the installation of bicycle lanes does not lead to additional crashes, but a possible increase in the number of bicyclists (Chen *et al.* 2012). Sakshaug et al. (2010) found that adding roundabouts produced more bicycle conflicts as the yielding rules were ambiguous in roundabout areas, contributing to a lower yielding rate and less trust among road users. By differentiating roundabouts at different locations, Daniels et al. found that roundabouts with cycle lanes performed worse than roundabouts in mixed traffic and separated cycle paths (Daniels *et al.* 2009).

In terms of street elements, bus stop density is positively associated with bicycle crash frequencies (Miranda-Moreno *et al.* 2011b, Wei and Lovegrove 2012, Strauss *et al.* 2013). Among current studies, street lighting has only been included for modeling bicycle injury severity (Klop and Khattak 1999, Kim *et al.* 2007), but is rarely considered for bicycle crash frequency. Most cities do not have accurately geo-coded street light data. Also, aggregating the lighting condition of individual crashes to areas is not reasonable.

Parking entrances do not appear to have a significant relationship with bicycle crash frequency (Miranda-Moreno et al. 2011b), but parked automobiles near separated bicycle facilities are associated with an increased crash risk because bike lanes and parking areas are placed together (Vandenbulcke et al. 2014). These findings are hypothesized to be related to three things: (1) when passing through entrances to parking lots, drivers focus on pedestrians and bicyclists passing across the entrance to avoid a collision. (2) Urban roadway mileage has a greater proportion of street parking relative to parking entrances, therefore the likelihood of bicycle crashes occurring at parking entrances when compared to lots is relatively low. (3) Bicyclists are also more likely to be in drivers' blind spots. Because drivers are less likely to see the bicyclists, they are therefore more likely to hit them when backing, starting, or opening doors. However, there are no observed studies on this and further research will be needed to validate these hypotheses.

Among travel demand variables, vehicle volume (Schepers *et al.* 2011, Hamann and Peek-Asa 2013) and bicycle volume (Miranda-Moreno *et al.* 2011b, Schepers *et al.* 2011, Hamann and Peek-Asa 2013, Strauss *et al.* 2013) have been considered and all the models suggest positive associations with the frequency of bicycle crashes. As for road design variables, a higher density of low-speed streets (< 15 mph) is negatively associated with the number of bicycle crashes (Siddiqui *et al.* 2012), while more roads with high-speed limits (> 35 mph) correlate with more bicycle crashes (Siddiqui *et al.* 2012, Chen and Fuller 2014). Additionally, the traffic signal density is positively correlated with the bicycle crash frequencies (Wei and Lovegrove 2012).

In relation to land use factors, the percentage of commercial land use and proximity to it were positively associated with bicycle crash frequency and bicyclist evident injuries (Narayanamoorthy *et al.* 2013, Vandenbulcke *et al.* 2014), but percentage of commercial land use was not a significant predictor of bicycle crashes in Strauss et al.'s study (2012). Inconsistencies remain in the effects of land use factors. Siddiqui et al.'s study showed that the densities of population and employment were positively related to bicycle crash frequency (2012).

3.2.4 Modeling techniques

The concerns of crash frequency modeling include over-dispersion or under-dispersion of count data, unobserved heterogeneity, spatial dependence, and the excess of zeros (Lord and Mannering 2010, Mannering and Bhat 2014). Data over dispersion is the presence of great variability, expressed by that the variance is largely greater than the mean of a variable. Data over dispersion is commonly observed in count data. The basic model used in bicycle crash frequency research is the negative binomial regression, which can handle data over-dispersion. Zero-inflated models can account for the excess of zeros by jointly working with the Poisson or negative binomial model. The generalized additive model and random effects model can calculate spatial dependence. The above modeling advantages can be jointly considered, such as the Poisson-lognormal conditional-autoregressive model (Wang and Kockelman 2013) and the

Bayesian multivariate Poisson-Lognormal model (Park and Lord 2007, Aguero-Valverde and Jovanis 2009). The spatial statistical approach provides a chance to capture the spatial autocorrelation with accurately estimated parameters. It contributes to the generalizability that same treatments can be applied to areas with similar features.

3.3 Modeling approach

This study employs two area-based Poisson-Lognormal random effects models. The models have two attractive features: handling over-dispersion of count data, and accounting for unobserved heterogeneity and spatial dependence. They provide the subject-specific estimates based on conditional probability, as compared to the aggregated population parameters in the other fixed effects models.

The Poisson-Lognormal random effects model is becoming popular in crash frequency and crash risk research. Fixing the random effects can improve the model fit and the precision of posterior estimates (Aguero-Valverde 2013). The marginal distribution of this Poisson-Lognormal model does not have a closed form; hence, it is implemented with the hierarchal Bayesian estimation.

An important statistical concern is to differentiate frequency and risk. According to the theory of *"safety in numbers"*, areas with more bicycle collisions may be less risky for bicycling. With this assumption, the risk is measured by the TAZ-counted number of bicycle crashes divided by the TAZ-estimated number of bicycle trips in this study, as specified in *Equation 3-1*.

$$Risk_i = \frac{Y_i}{E_i}$$ *Equation 3-1*

Y_i refers to the number of bicycle crashes in each TAZ, and E_i is the expected number of bicycle trips. The frequency model and the risk model are expressed in *Equations 3-2 and 3-3*:

$$Y_i \mid \alpha, \beta_i, \mu_i, \nu_i \sim Poisson(e^{\alpha+\beta_i X_i} e^{\mu_i+\nu_i})$$ *Equation 3-2*

$$Y_i \mid \alpha, \beta_i, \mu_i, v_i \sim Poisson(E_i e^{\alpha+\beta_i X_i} e^{\mu_i + v_i})$$ *Equation 3-3*

The difference of the models specified in *Equations 3-2 and 3-3* lies in whether the models have included the E_i as a denominator. Each model has an intercept *α,* and a vector of estimated parameters β_i for the fixed effects. X_i is a vector of the independent variables. The unknown quantities are the coefficients of the vectors of u_i and v_i, which are two latent random effects to compose the posterior distributions of spatial variance (u_i) and unobserved heterogeneity (v_i).

$$u_i \mid u_j, j \in ne(i) \sim N(u_i, \frac{\sigma_u^2}{m_i})$$ *Equation 3-4*

$$\mu_i = \frac{1}{m_i} \sum_{j \in ne(i)} \mu_j$$ *Equation 3-5*

Equations 3-4 and *3-5* describe the distribution of the random effect that is employed to estimate the spatial dependence. u_i is the local spatial random effects assumed to follow a lognormal distribution. To define a neighbor, $ne_{(i)}$ is the set of adjacent polygons of area *i* , and m_i is the number of neighbors of area *i*. The neighbors are defined by at least sharing one border. Specifically, for area *i*, the variance of u_i is conditional on the variance of u_j, and $j \in ne_{(i)}$. These models assign the spatial random effects an intrinsic conditional on the autoregressive prior. The spatial random effect's mean is the mean of the neighboring TAZs, and the variance is proportional to one over the number of its neighbors. This approach was developed by Besag et. al. (Besag *et al.* 1991).

$$v_i \sim N\left(0, \sigma_v^2\right)$$ *Equation 3-6*

Where v_i represents the random effect of unobserved heterogeneity, assuming it follows independent and identical distribution, as expressed by *Equation 3-6.* v_i captures the residual or unexplained log frequency/risk of collisions in area *i*, where the variance σ_v^2 controls the extra-Poisson variation (Aguero-Valverde 2013).

Both σ_u^2 and σ_v^2 are defined with respect to the log scale. σ_u^2 is a conditional variance and its magnitude determines the amount of spatial variation, whereas σ_v^2 has a marginal interpretation. Bayesian inference is carried out through the R INLA package. INLA is short for integrated nested Laplace approximation.

3.4 Data sources and geo-unit selection

The empirical setting of this study is the city of Seattle. As a bicycle friendly city, Seattle Department of Transportation has been working steadily toward developing an urban bicycle trail system to accommodate bicyclists. To better understand bicycle safety, the data employed in this study include two components, the bicycle crash records and the built environment features.

3.4.1 Automobile-involved bicycle crash data

The bicycle crash data were collected by Seattle Department of Transportation from 2010 to 2013. The bicycle crash data had 1,389 records. In that 4-year period, the total number of automobile-involved crashes in Seattle was 46,797, and the bicycle crashes were accounted for 2.97%. The data had five types of injuries, including fatality (dead at scene/ on arrival/ in hospitals), serious injury, evident injury, possible injury and property damage only.

These crash data have limitations because a large number of minor incidents are unreported to authorities (de Geus *et al.* 2012, Wegman *et al.* 2012). The possible biases included are: (1) less severe crashes are widely underestimated, including property damages, possible injuries, and evident injuries; (2) collisions that occurred at local streets and suburban areas are relatively under surveillance; and (3) collisions between bicyclists, conflicts between bicyclists and pedestrians, and single falls are not covered in the sample.

3.4.2 Risk factors

Bicycle crash frequency or risk results from the interaction of three traffic safety pillars: road user(s), bicycle(s) or automobile(s), and the built environment (Schepers *et al.* 2013). Risk factors are defined as any built environment features or risk-taking travel behaviors that increase the probability of a bicycle crash. In this study, the built environment data are supported by Puget Sound Regional Council, Seattle Department of Transportation, and King County. The sources, explanations, and data summary of selected variables are listed in *Table 3-1*. The variable selection based on multicollinearity, which indicates that two or more predictor variables in a multiple regression model are highly correlated. The selection criterion is based on the variance inflation factor (VIF), which measures the severity of multicollinearity in a regression using the ordinary least squares as the estimating method. If the VIF of a variable is greater than 5.0, that variable is excluded in this study.

Being consistent with existing research, the built environment features included for modeling are classified as factors of road network and land use. In addition, variables of road design and travel demand are also included for modeling. The built environment features are quantified with the analysis unit of the TAZ. The road network features includes 3-way intersections, 4-way intersections, and complicated intersections (more than 4-way), the densities of on-arterial versus off-arterial bicycle routes, and zonal mean slope. The land use factors are land use mixture, the proportion of commercial and mixed lands, the proportion of office and government lands, the proportion of industrial lands, household density, and employment density.

In addition, the road design variables include the densities of road signals and stop signs, and zonal mean of posted speed limits. The zonal mean of posted speed limits is equal to the length of road segment multiplied by the corresponding posted speed limit, and divided by the sum length of roads. The street elements are the densities of bus stops, street lights, street trees, and traffic circles. The density of crosswalks and the densities of local streets versus arterial routes are excluded due to multicollinearity.

Also, three travel demand variables are included for modeling. The numbers of bicycle trips and total trips are estimated by Puget Sound Regional Council (PSRC), which are a part of the output of an activity-based travel forecasting model, called SoundCast. The origin based the TAZ-counted number of bicycle trips and the total number of trips are the major outputs of this model[31]. The original data were surveyed and gathered by Puget Sound Regional Council. Traffic volume (annual average daily traffic, AADT) and the number of lanes are not included in this research, because Seattle only has those data for arterial routes. The bicycle volume and bicycle miles traveled are also not available. In this study, the origin-based TAZ-counted number of bicycle trips acts as a substitute for the bicycle volume, while the total number of trips is a substitute for the traffic volume. The bicycle mode share is calculated by the number of bicycle trips divided by the total number of trips in TAZs. The use of these travel demand variables is novel in measuring bicycle exposure.

3.4.3 Geo-spatial analytical unit selection

The geographic unit in quantifying the built environment and aggregating crash frequencies varied in prior studies, such as county (Aguero-Valverde and Jovanis 2006, Huang *et al.* 2010), census tract (Ukkusuri *et al.* 2012), TAZ (Siddiqui *et al.* 2012, Wei and Lovegrove 2012), grid-cell (Gladhill and Monsere 2012) and intersection (Miranda-Moreno *et al.* 2011a, Castro *et al.* 2012, Vandenbulcke *et al.* 2014). There is a trade-off when a geographical scale is

[31] This study uses the origin-counted trips (bicycle versus all transportation modes). The risk of using these data for research include: (1) bicycle crashes usually not happen at origin or destination, but on the way to destinations; (2) Surveyed number of trips is likely to underestimate the non-motorized trips, which are of recreational purposes, and is likely to underestimate intra-zonal travels, which are not of primary transportation modes.

chosen. Larger areas provide more stable rates, but the accuracy of measurements may be reduced due to aggregation. Aggregating at a large geospatial scale may produce the threat of regression toward the mean. Some localized effects can only be detected on a small scale.

This study takes TAZ as the analytical unit, because it matches the census information on important demographic profiles and travel demand characteristics, including population density, employment density, bicycle mode share, the number of bicycle trips, and the total number of trips.

Another challenge is the possible error of counting the bicycle crashes. A collision may just fall into the cross-boundary between two neighboring TAZs. TAZs are usually split by the arterials, and many bicycle crashes occurred on arterial routes. An inaccurate geo-coding error may result in the changes of counted crash occurrences among TAZs. This study uses “spatial join” function in the ArcGIS to count bicycle crashes, assuming that the data were correctly geo-coded by Seattle Department of Transportation.

Table 3-1: Variable definitions and data summary (n = 707) of potential predictors for bicycle crashes in Seattle TAZs.

Variable	Mean	S.D.	Min.	Max.	Unit	Source
Bicycle crash						
Number of bicycle crashes *(E_i)*	1.97	2.58	0.00	28.00	numb/TAZ	SDOT
Road network						
Number of 3-way intersections per ha	0.31	0.26	0.00	1.75	numb/ha	PSRC
Number of 4-way intersections per ha	0.46	0.34	0.00	2.28	numb/ha	PSRC
Number of complicated intersections (5 or more ways) per ha	0.04	0.14	0.00	1.33	numb/ha	PSRC
Number of traffic circles per ha	0.06	0.09	0.00	0.51	numb/ha	SDOT
Length of on-arterial bicycle lanes per ha	0.02	0.02	0.00	0.19	km/ha	SDOT
Length of off-arterial bicycle lanes per ha	0.04	0.03	0.00	0.20	km/ha	SDOT
Zonal mean slope (average gradients, absolute value)	0.21	0.33	0.00	4.11	ratio	SDOT
Number of bus stops per ha	0.26	0.31	0.00	2.74	numb/ha	King County
Number of street lights per ha	5.12	2.22	0.00	15.48	numb/ha	SDOT
Number of street trees per ha	8.53	6.26	0.00	31.46	numb/ha	SDOT
Number of parking signs per ha	4.36	6.97	0.00	43.4	numb/ha	SDOT
Road design						
Number of stop signs per ha	0.50	0.43	0.00	2.87	numb/ha	SDOT
Zonal mean of posted speed limits	25.25	4.53	20.00	48.08	mph	SDOT
Number of traffic signals per ha	0.19	0.34	0.00	2.13	numb/ha	SDOT
Travel demand						
Bicycle mode share	0.02	0.02	0.00	0.18	ratio	PSRC
Number of bicycle trips in TAZs *(Y_i)*	0.07	0.17	0.00	3.68	10^3	PSRC
Number of trips in TAZs	3.55	2.98	0.03	40.86	10^3	PSRC
Land use						
Land-use mixture, ranging from 0 to 1	0.51	0.15	0.00	0.87	ratio	PSRC

Proportion of industrial lands in TAZs	0.07	0.14	0.00	0.91	ratio	PSRC
Proportion of commercial and mixed lands in TAZs	0.10	0.15	0.00	0.81	ratio	PSRC
Proportion of office and government lands in TAZs	0.10	0.15	0.00	0.83	ratio	PSRC
Area of each TAZ	*0.31*	*0.33*	*0.01*	*3.37*	*km^2*	*SDOT*
Demographics						
Household density	0.02	0.03	0.00	0.18	10^3/ha	PSRC
Employment density	0.09	0.23	0.00	2.09	10^3/ha	PSRC
Number of households in the TAZs	*402.21*	*352.24*	*0.00*	*2,181*	*numb/TAZ*	*PSRC*
Number of jobs in the TAZs	*680.36*	*954.75*	*0.00*	*17,304*	*numb/TAZ*	*PSRC*
The variables in *italics* are excluded in modeling due to multicollinearity.						

3.5 Descriptive analysis

The number of bicycle crashes ranges from 0.00 to 28.00, with a mean of 1.97 collisions and a standard deviation of 2.58 collisions, suggesting that the distribution of the bicycle crash frequency is dispersed. The number of bicycle trips ranges from 1 to 3,680. As indicated in *Figure* 3-1, more collisions are clustered in downtown Seattle. A supportive figure to show the pattern of bicycle crash risk is displayed in *Figure 3-2*. As displayed, the bicycle crash risks are greater in North and South Seattle. The clusters of bicycle crash frequency and risk are not matching spatially.

Table 3-2 lists the outcome of the Poisson-Lognormal random effects model using hierarchal Bayesian estimation. The estimates at 2.50% and 97.50% credential intervals with no possible zero parameters are significant factors. The significant factors in the bicycle crash frequency and risk models: the 3-way intersection density, the density of on-arterial bike lanes versus off-arterial bike lanes, the densities of road signals, the zonal mean of posted speed limits, and the total number of trips. The land use mixture is only significant in the bicycle crash frequency model. Bicycle mode share, the density of street trees, the proportion of commercial and mixed lands, household density, and employment density are only significant in the bicycle crash risk model.

The spatial dependence is calculated based on the variances of the two random effects, expressed by $\frac{\sigma_u}{\sigma_u+\sigma_v}$, accounting for 54.49% and 56.82% of the random effects variances in the bicycle crash frequency and bicycle crash risk models accordingly. This model indicates that more than half of the errors can be explained by the spatial autocorrelation or spatial spillover effects. It also suggests 45.50% and 43.18% of unobserved heterogeneities remain in the two models. The same modifications can be applied to areas with similar built environment features to reduce bicycle crash frequency and bicycle crash risk.

In the outcomes of the Poisson-Lognormal random effects models, the 3-way intersection density and the density of off-arterial bicycle lanes suggest negative relationships with bicycle crash frequencies and risks. The density of road signals, the zonal mean of posted speed limits, the total number of trips, and the density of on-arterial bicycle lanes are positively correlated with bicycle crash frequencies and risks.

The elasticity is computed as the regression parameter times the mean value of the variable ($E = \beta_i * \bar{X}_i$) (Ewing and Cervero 2010). By fixing the other included fixed covariates, the calculated elasticity reflects the relationship between 1.00% change in the independent variable and the corresponding percentage change in the dependent variable.

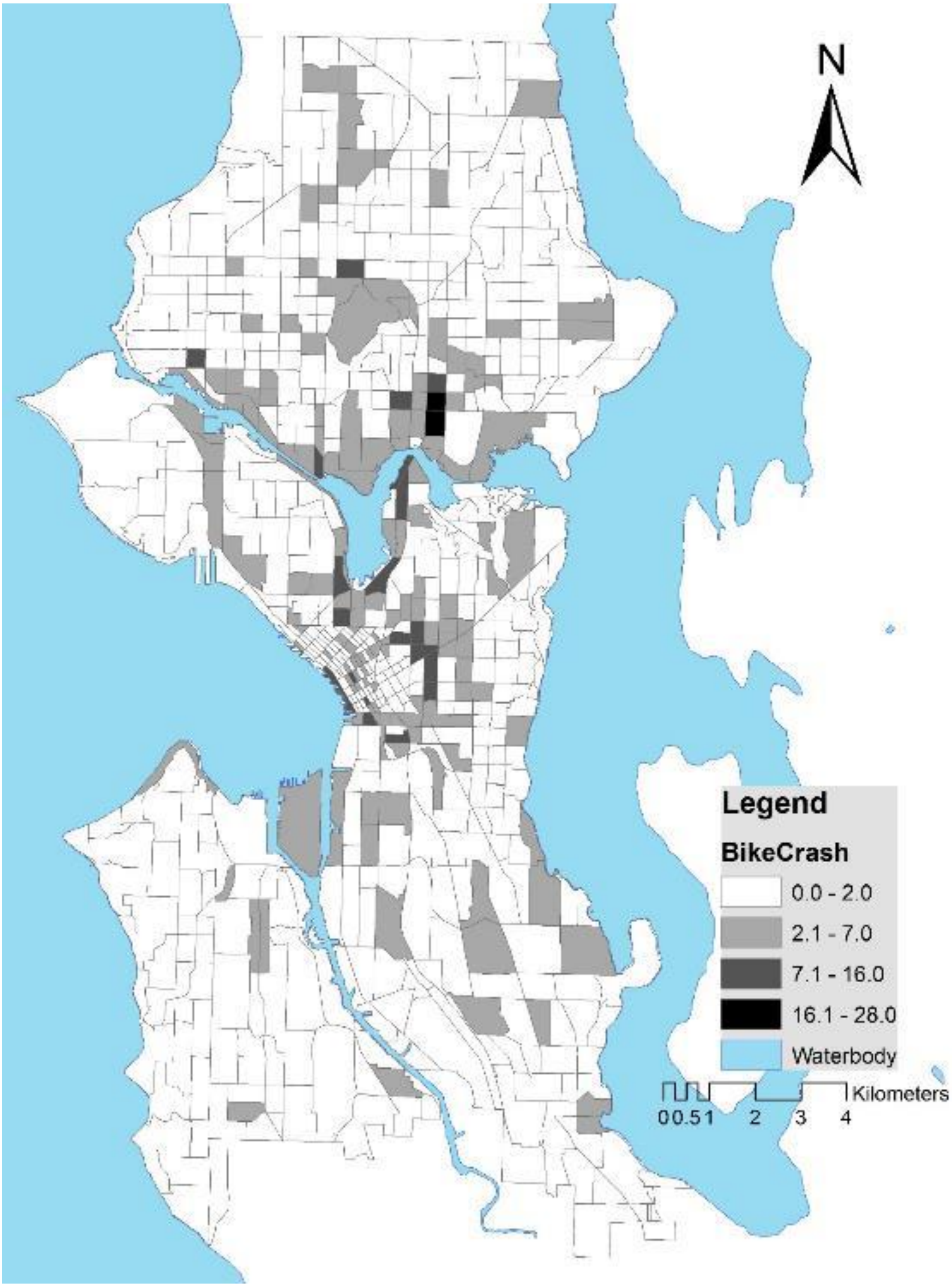


Figure 3-1: Bicycle crash frequencies in Seattle traffic analysis zones, 2010-2013 *(Y_i)*

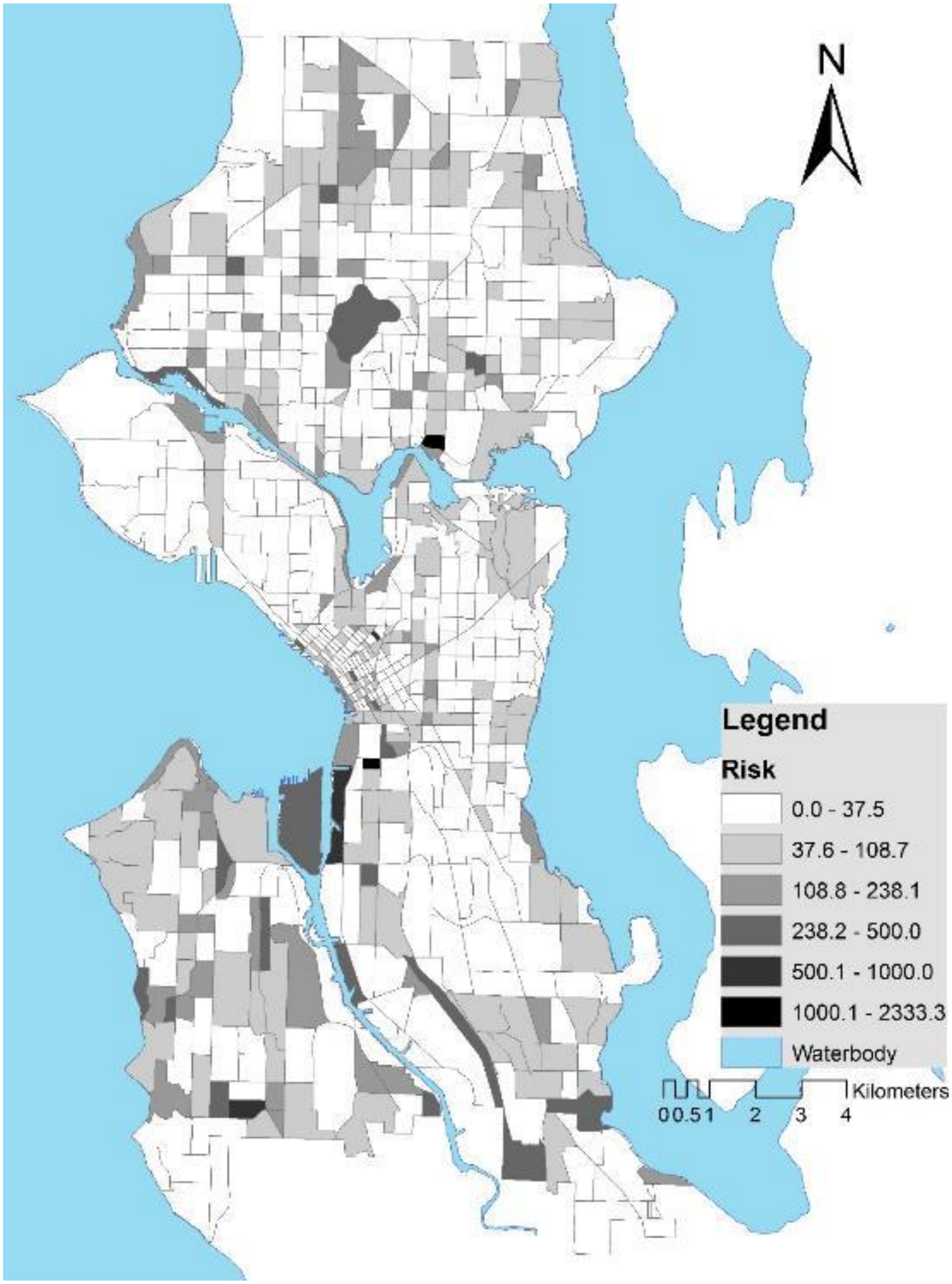


Figure 3-2: Bicycle crash risks in Seattle traffic analysis zones, 2010-2013 *(Y_i/E_i)*

3.6 Inferential analysis

Table 3-2: The estimates of the Poisson-Lognormal random effects models and the elasticities for significant variables

	Bicycle crash frequency model					Bicycle crash risk model				
Fixed effects:	Mean	S.D.	2.5% CI	97.5% CI	Elasti.	Mean	S.D.	2.5% CI	97.5% CI	Elasti.
Intercept	-1.31	0.43	-2.16	-0.47		-2.93	0.46	-3.83	-2.03	
Road network										
Number of 3-way intersections per ha	-0.37	0.17	-0.70	-0.04	-0.11	-0.44	0.18	-0.79	-0.09	-0.14
Number of complicated intersections per ha	-0.57	0.32	-1.21	0.06		-0.58	0.34	-1.25	0.08	
Number of 4-way intersections per ha	-0.20	0.19	-0.57	0.17		-0.36	0.20	-0.76	0.04	
Number of traffic circles per ha	-0.15	0.55	-1.23	0.93		-0.29	0.58	-1.43	0.84	
Length of on-arterial bicycle lanes per ha	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00
Length of off-arterial bicycle lanes per ha	-0.01	0.00	-0.01	-0.00	-0.00	-0.01	0.00	-0.02	-0.01	0.00
Zonal mean slope	-0.30	0.18	-0.67	0.05		-0.09	0.19	-0.47	0.28	
Number of bus stops per ha	0.27	0.18	-0.07	0.62		0.21	0.18	-0.15	0.57	
Number of street lights per ha	0.03	0.02	-0.02	0.07		0.00	0.03	-0.05	0.06	
Number of street trees per ha	-0.01	0.01	-0.02	0.01		-0.02	0.01	-0.03	-0.00	-0.14
Road design										
Number of parking signs per ha	0.01	0.01	-0.01	0.03		0.01	0.01	-0.01	0.03	
Number of stop signs per ha	0.13	0.13	-0.12	0.38		0.21	0.13	-0.05	0.48	
Zonal mean of posted speed limits	0.03	0.01	0.01	0.05	0.80	0.04	0.01	0.01	0.06	0.91
Number of traffic signals per ha	0.62	0.22	0.19	1.05	0.12	0.95	0.23	0.50	1.40	0.18
Travel demand										
Bicycle mode share	-2.76	1.67	-6.05	0.50		-6.41	1.76	-9.89	-2.96	-0.13
Number of bicycle trips	-0.60	0.37	-1.33	0.13		---	---	---	---	
Number of trips	0.06	0.02	0.02	0.11	0.23	-0.11	0.02	-0.14	-0.07	-0.37
Land use										

Land use mixture	1.21	0.35	0.53	1.89	0.62	-0.27	0.36	-0.98	0.43	
Proportion of industrial lands	-0.03	0.43	-0.88	0.81		0.35	0.45	-0.55	1.23	
Proportion of commercial and mixed lands	0.29	0.37	-0.45	1.02		0.80	0.39	0.04	1.57	0.08
Proportion of office and government lands	0.13	0.37	-0.60	0.85		0.38	0.39	-0.38	1.13	
Demographics										
Household density	-0.02	0.26	-0.53	0.47		-0.81	0.27	-1.34	-0.27	-0.02
Employment density	4.06	4.59	-5.29	12.75		-25.54	5.29	-35.95	-15.15	-2.30
Random effects:										
	Mean	S.D.	2.5% CI	97.5% CI		Mean	S.D.	2.5% CI	97.5% CI	
σ_v	8.31	3.43	3.98	17.13		7.13	3.05	3.29	14.97	
σ_u	0.93	0.23	0.53	1.43		0.83	0.22	0.44	1.32	
Marginal likelihood					-1963.59					-1988.99
Spatial dependence $\frac{\sigma_u}{\sigma_u+\sigma_v}$					54.49%					56.82%

3.7 Discussion

The purpose of this study is to examine how the built environment features are related to bicycle crash frequency and risk, particularly to identify the modifiable factors that contribute to better cycling safety. To differentiate bicycle crash frequency and risk, and highlight the importance of quantifying exposure to comprehend the built environment risk factors, this research design is unique through investigating the same set of fixed covariates with two different safety measures.

Both the random effects of the spatial variations in bicycle crash frequency and for bicycle crash risk take more than half of the total variations accordingly. These results confirm the existence of significant spatial autocorrelations among TAZs, indicating that the estimates of nearby TAZs are more related than those of distant TAZs. These results indicate that engineering modifications and planning strategies could be applied to TAZs with similar built environment features.

In relation to the road network features, the modeling outcomes suggest that the 3-way intersection density is negatively associated with bicycle crash frequencies and risks. The effects of the 4-way and complicated intersections remain unclear. However, some prior research found positive relationships between bicycle crashes and intersection density/counts (Siddiqui *et al.* 2012, Wei and Lovegrove 2012). There is a possible explanation for this inconsistency. The exposure to encounters between intersecting automobiles increases as a number of intersections increase; but road networks with more intersections may contribute to lower driving speeds and thereby lessen severe bicycle crashes. Hence, the results are mixed. Cycling on on-arterial bike lanes is more dangerous than cycling on off-arterial bike lanes, which is consistent with the past research on the effects of different bicycle routes (Teschke *et al.* 2012, Wei and Lovegrove 2012, Lusk *et al.* 2013).

The outcome shows that a higher zonal mean of posted speed limits is associated with more bicycle crash frequencies and risks. The elasticities are 0.80% of bicycle crash frequencies and 0.91% of bicycle crash risks for a 1.00% change in the zonal mean of posted speed limits. Prior area-based research did not include zonal mean of posted speed limits as risk factors. However, other micro-level bicycle injury severity studies suggested that higher posted speed limits are associated with more severe bicyclist injuries (Kim *et al.* 2007, Eluru *et al.* 2008, Zahabi *et al.* 2011, Chen and Fuller 2014). In short, all research indicates that lowering posted speed limits is an effective approach to reduce cycling risks.

This study shows that bicycle crash risk is lower when cycling on streets with more trees. Also, the results on traffic signals are consistent with past research (Wei and Lovegrove 2012); a higher signalized intersection density is associated with more bicycle crash frequencies and risks.

In relation to land uses, a study showed that industrial and commercial land uses were positively related to bicyclist injuries (Narayanamoorthy *et al.* 2013), but those results remain unclear with Seattle's data regarding bicycle crash frequency. However, the result of the risk model shows that bicycle crash risk is higher in zones with a greater proportion of commercial and mixed lands. The land use mixture has barely been investigated in prior bicycle crash frequency studies. This study suggests that a 1.00% increase in the land use mixture is associated with a 0.62% increase in the number of bicycle crashes. This positive relationship may result from the conflicts between concentrated human activities in places with different land use purposes.

This study used the total number of trips as a substitute for the traffic volume. Even though the variables are different, the findings are somewhat consistent (Miranda-Moreno et al., 2011; Chen and Fuller, 2014, Vandenbulcke et al., 2014). A positive relationship between the total number of trips and bicycle crash frequencies is identified in this study. However, in the risk model, the corresponding relationship between bicycle crash risks and the total number of trips is negative. These opposite coefficients in the two models indicate that more bicycle crashes could be observed in high traffic area, but the risk of involving a bicycle collision is lower.

To examine the theory of *"safety in numbers"*, the remaining interpretations focus on several density-related measures. As noted in the risk model, the outcome shows that bicycle crash risk is negatively correlated with the bicycle mode share, the total number of trips, household density, and employment density. It is worth noting that a 1.00% change in the employment density is associated with a 2.30% change in the bicycle crash risk. Promoting employment density could bring safety benefits to the bicyclists. These outcomes provide strong evidence to support the conclusion that bicycling is safer in densely developed areas.

3.8 Limitations

Some limitations of this study should be noted. A major challenge comes from the trend confounding effect across a 4-year period. Road design factors are time-varying explanatory variables so that the temporal effects of traffic signals are hard to control. For instance, the signals could have been installed after the crashes had occurred, which could be a plausible alternative explanation for the positive association between bicycle crashes and the number of signals. A similar case can be found in the interpretation of the relationship between bicycle crashes and on-arterial bicycle lanes. The local authority may have given higher priority to improving the bicycle facilities where crashes had occurred. Because the order of the events cannot be confirmed, uncertainty remains in the interpretation. A possible way to control this confounding effect is to add random effects for space-time modeling. However, the space-time model assumes that the crash risks in the temporal trends are linear (Knorr-Held and Besag 1998). This assumption does not apply to Seattle Department of Transportation bicycle crash data. The sample is too small to be further split, and bicycle crash counts in TAZs are likely to have the problem of the excess of zeros.

The second challenge comes from the underreporting/under-surveillance and unavailability of data used in this study. Firstly, there are many unreported minor bicycle collisions, especially in suburban areas and local streets. This underreporting of bicycle crash counts poses a threat to the reliability of the data. Similarly, the traffic volume data are only gathered at arterials and freeways, but missing on local streets. Secondly, data unavailability could have a negative effect on the accuracy of modeling. For instance, this study suggests that the density of on-arterial bicycle lanes is positively correlated with bicycle crash frequency. This conclusion could be impacted by missing bicycle volume data. Bicycle miles traveled and bicycle volume are more accurate in describing the distance bicyclists traveled and the number of on-road bicyclists. Better data should be gathered to enrich future bicycle safety research.

The denominator used to measure bicycle crash risk is the zonal-counted number of bicycle trips. The original number of bicycle trips was collected from regional household travel behavior survey. The survey data are mostly collected from bicycle tours with a primary trip of commuting purpose (Puget Sound Regional Council 2014). Yet the majority of bicycling trips serves the purposes of recreation and exercise. For example, two-thirds of the bicycle trips are of general recreation (Seattle Department of Transportation 2014). And in many cases, bicycle trips are ignored as intra-zonal travels. Therefore, the bicycle crash risk measured in this study is just an approximation.

It remains questionable whether findings generalized from Seattle are applicable to other US cities. This is because, firstly, Seattle is a city of relatively high density, posing a threat to the generalizability in the other low-density cities. It is likely that these results apply the best to other central downtown areas. Secondly, Seattle continually implements the bicycle master plan and cycle share program to promote bicycling as a primary transportation mode. It may not be an appropriate inference for cities having much lower bicyclist volume.

3.9 Conclusions

The popularity of cycling is greatly related to actual and perceived safety. Lowering bicycle crash risk is a key step in increasing the likelihood of bicycle use and in promoting cycling as an active mode of transportation. The safety ramification of increased bicycle use in North America was evaluated in this chapter within an urban environment.

Regression techniques were employed in this study to highlight the importance of measuring spatial dependence and unobserved heterogeneity. The results indicate that zonal bicycle crash frequencies are spatially correlated. This study has included some bicycle exposure variables that had not been investigated in prior studies. Unique variables include the number of bicycle trips versus the total number of trips, the zonal mean of posted speed limits, the zonal mean slopes, and the densities of street trees and parking signs. The significance of these new variables helps to better understand the exposure and risks of cycling behaviors. There were limitations in the bicycle exposure data and future research will need to capture improved data to verify some of the results.

For transportation engineers, planners and policy makers, this study provides several statistically founded recommendations to improve bicycle safety through engineering modifications. For instance, local authorities should lower the posted speed limits, encourage compact development, and separate bicycle lanes from road traffic. Additionally, the incentive of speed limit reduction for bicycle safety is to decrease actual driving speeds. For good roadway design practices, transport engineers should apply the principles of *Vision Zero* (Tingvall and Haworth 2000), and *Functionality, Homogeneity and Predictability* (Wegman *et al.* 2005) to operate a sustainable and safe traffic system. In view of the widely ongoing planning and construction to promote increased cycling in the North America, continual research on bicycle crash risks is urgently needed. Safety cannot be traded for the sake of mobility (Tingvall and Haworth 2000).

References

Aguero-Valverde, J., 2013. Full bayes poisson gamma, Poisson-Lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis & Prevention 50 (0), 289-297.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson-Lognormal models for crash severity modeling and site ranking. Transportation Research Record: Journal of the Transportation Research Board 2136 (1), 82-91.

American Association of State Highway and Transportation Officials and US Departmemt of Transportation, 2013. Commuting in america 2013: the national report on commuting patterns and trends. 7-8.

Attewell, R.G., Glase, K., Mcfadden, M., 2001. Bicycle helmet efficacy: A meta-analysis. Accident Analysis & Prevention 33 (3), 345-352.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43 (1), 1-20.

Bhatia, R., Wier, M., 2011. "Safety in numbers" re-examined: Can we make valid or practical inferences from available evidence? Accident Analysis & Prevention 43 (1), 235-240.

Bikeleague, Bicycle commuting data.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transportation Research Part B: Methodological 46 (1), 253-272.

Chen, D., Fuller, D., 2014. Analyzing road surface conditions, collision time, and road structural factors associated with bicycle collisions from 2000 to 2010 in Saskatoon, Saskatchewan. Journal of Transport & Health 1 (1), 40-44.

Chen, L., Chen, C., Srinivasan, R., Mcknight, C.E., Ewing, R., Roe, M., 2012. Evaluating the safety effects of bicycle lanes in New York city. American Journal of Public Health 102 (6).

Clark, A., Thornley, B., Tomlinson, L., Galletley, C., Norman, R.J., 1998. Weight loss in obese infertile women results in improvement in reproductive outcome for all forms of fertility treatment. Human Reproduction 13 (6), 1502-1505.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2009. Injury crashes with bicyclists at roundabouts: Influence of some location characteristics and the design of cycle facilities. Journal of safety research 40 (2), 141-148.

De Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2012. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accident Analysis & Prevention 45, 683-693.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40 (3), 1033-1054.

Elvik, R., 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. Accident Analysis & Prevention 41 (4), 849-855.

Ewing, R., Cervero, R., 2010. Travel and the built environment: A meta-analysis. Journal of the American Planning Association 76 (3), 265-294.

Gladhill, K., Monsere, C.M., 2012. Exploring traffic safety and urban form in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board 2318 (1), 63-74.

Hamann, C., Peek-Asa, C., 2013. On-road bicycle facilities and bicycle crashes in iowa, 2007–2010. Accident Analysis & Prevention 56 (0), 103-109.

Harris, M.A., Reynolds, C.C., Winters, M., Chipman, M., Cripton, P.A., Cusimano, M.D., Teschke, K., 2011. The bicyclists' injuries and the cycling environment study: A protocol to tackle methodological issues facing studies of bicycling safety. Injury Prevention 17 (5), e6-e6.

Huang, H., Abdel-Aty, M.A., Darwiche, A.L., 2010. County-level crash risk analysis in Florida. Transportation Research Record: Journal of the Transportation Research Board 2148 (1), 27-37.

Jacobsen, P.L., 2003. Safety in numbers: More walkers and bicyclists, safer walking and bicycling. Injury prevention 9 (3), 205-209.

Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis & Prevention 39 (2), 238-251.

Klop, J.R., Khattak, A.J., 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. Transportation Research Record: Journal of the Transportation Research Board 1674 (1), 78-85.

Knorr-Held, L., Besag, J., 1998. Modelling risk from a disease in time and space. Statistics in medicine 17 (18), 2045-2060.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44 (5), 291-305.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2013. Risk of injury for bicycling on cycle tracks versus in the street. Injury prevention 17 (2), 131-135.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1 (0), 1-22.

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

Miranda-Moreno, L.F., Strauss, J., Morency, P., 2011b. Disaggregate exposure measures and injury frequency models of cyclist safety at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board 2236 (1), 74-82.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B: Methodological 55 (0), 245-264.

Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record: Journal of the Transportation Research Board 2019 (1), 1-6.

Puget Sound Regional Council, 2014. Activity-based travel model: Soundcast. http://www.psrc.org/data/models/abmodel/.

Reynolds, C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. Environmental Health 8 (1), 47.

Schepers, J., Kroeze, P., Sweers, W., Wüst, J., 2011. Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. Accident Analysis & Prevention 43 (3), 853-861.

Schepers, P., Hagenzieker, M., Methorst, R., Van Wee, B., Wegman, F., 2013. A conceptual framework for road safety and mobility applied to cycling safety. Accident Analysis & Prevention (0).

Seattle Department of Transportation, 2014. Seattle master bicycle plan.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention 45 (0), 382-391.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: A bayesian modelling approach. Accident Analysis & Prevention 59 (0), 9-17.

Teschke, K., Harris, M.A., Reynolds, C.C., Winters, M., Babul, S., Chipman, M., Cusimano, M.D., Brubacher, J.R., Hunte, G., Friedman, S.M., 2012. Route infrastructure and the risk of injuries to bicyclists: A case-crossover study. American Journal of Public Health 102 (12), 2336-2343.

The National Highway Traffic Safety Administration, 2013. Traffic safety facts: Bicyclists and other cyclists. http://www-nrd.nhtsa.dot.gov/Pubs/812018.pdf.

Tingvall, C., Haworth, N., Year. Vision zero: An ethical approach to safety and mobility. In: Proceedings of the 6th ITE International Conference Road Safety & Traffic Enforcement: Beyond.

United States Department of Transportation, 2009. 2009 national household travel survey.

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

Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in brussels: A spatial case–control approach. Accident Analysis & Prevention (0).

Walker, I., 2007. Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. Accident Analysis & Prevention 39 (2), 417-425.

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

Wang, Y., Nihan, N.L., 2004. Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. Accident Analysis & Prevention 36 (3), 313-321.

Wegman, F., Dijkstra, A., Schermers, G., Van Vliet, P., Year. Sustainable safety in the Netherlands: The vision, the implementation and the safety effects. In: Proceedings of the Proceedings of the 3rd International Symposium on Highway Geometric Design. Chicago.

Wegman, F., Zhang, F., Dijkstra, A., 2012. How to make more cycling good for road safety? Accident Analysis & Prevention 44 (1), 19-29.

Wei, F., Lovegrove, G., 2012. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention (0).

Zahabi, S.A.H., Strauss, J., Manaugh, K., Miranda-Moreno, L.F., 2011. Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. Transportation Research Record: Journal of the Transportation Research Board 2247 (1), 81-90.

Chapter 4. Built environment features in explaining bicyclist injury severity

Abstract

This analysis employs a generalized ordered logit model and a generalized additive model to estimate the built environment effects on bicyclist injury severity in automobile-involved bicycle crashes, and to accommodate the possible spatial autocorrelation among bicycle crash sites. The research data are obtained from the Seattle Department of Transportation bicycle collision profiles. This study categorizes the bicyclist injury types as property damage only, possible injury, evident injury, and severe injury or fatality. The modeling results indicate that: (1) employment density is negatively associated with bicyclist injury severity; (2) severe injury or fatality is negatively associated with the degrees of mixed land use; (3) lower probability of injuries is observed for bicyclists dressing reflective clothes; (4) better street lighting is associated with less probabilities of bicyclist injuries; (5) a higher posted speed limit increases the probabilities of evident injury and severe injury or fatality; (6) older bicyclists are more vulnerable to severe injury or fatality; and (7) bicyclists are more apt to be severely injured when large vehicles are involved in crashes. One recommendation drawn from this research is that cities should encourage mixed land use and promote compact urban development, optimally lower posted speed limits on streets with both bicycles and automobiles, avoid place bicycle lanes in steep streets, and improve lighting conditions to provide a safe bicycling environment. In addition, bicyclists are encouraged to wear reflective clothes.

4.1 Introduction

Bicycling is an active transportation mode, which offers health, environmental, and social benefits, such as decreased rate of obesity, lowered greenhouse gas emissions, reduced congestion, and improved livability. Bicycling activities are increasingly popular in the US due to better awareness of eco-friendly lifestyles. However, bicyclist injuries remain a serious public health problem. Safety issue acts as an important deterrent preventing people from bicycling. According to a recent report, only 0.50% of commuters in the US take bicycles as their primary transportation mode (American Association of State Highway and Transportation Officials and US Departmemt of Transportation 2013). While the number of deaths in traffic crashes steadily decreased in the past four decades (The National Highway Traffic Safety Administration 2012), the number of reported injured bicyclists increased from 45,000 in 2001 to 49,000 in 2012. The percentage of bicyclist fatalities among total traffic deaths increased from 1.70% to 2.20% in the same period (The National Highway Traffic Safety Administration 2013). Therefore, it is vital to investigate what the leading causes are correlated with bicyclist injury severity.

Human factors that are directly associated with bicyclist injury severities include helmet use by the bicyclist, and intoxication and distraction by the drivers (Attewell *et al.* 2001, Cummings *et al.* 2006, Walker 2007, Goldenbeld *et al.* 2012). The exploration of the correlations between the built environment and bicyclist injury severity has not been sufficiently examined for two reasons. First, exploring this link requires qualified bicyclist injury records and capability in conducting interdisciplinary research. Second, in contradiction of posted speed limit, vehicle type, and bicyclist age, built environment features are perceived as indirectly correlated with bicyclist injury severity. Hence, built environment features were mostly adjusted as confounders, or disregarded, in previous research.

This chapter uses a generalized ordered logit model and a generalized additive model to identify the foremost land use and road network features factors correlated with bicyclist injury severities for Seattle, Washington. This study highlights the effects of employment density and land use mixture in mitigating bicyclist injury severity.

4.2 Literature review

The majority of transportation safety studies concentrates on two issues: crash frequency and injury severity. The objective is to identify risk factors and draw policy recommendations and road network features guidelines for safety improvements. Previous bicyclist injury severity studies considered a large set of influential factors, reclassified as: (1) individual socio-demographics, such as age and gender of motorists and bicyclists; (2) behavioral factors, such as drinking alcohol using drugs, being distracted and inattentive, violating traffic regulations such as misusing helmets; (3) vehicle types; (4) road network features, such as bicycle facility types, slopes, and other features associated with intersections or mid-blocks; (5) road design variables, such as signals, stop signs, and posted speed limits; (6) environmental factors, such as the time of day and weather conditions; (7) land use variables, such as density and land use mixture; and (8) crash characteristics, such as the directions and movements of driving and bicycling.

Posted speed limit, vehicle type, and age of injured bicyclists are significant factors that directly contributed to severe bicyclist injuries (Kim *et al.* 2007, Walker 2007, Eluru *et al.* 2008, B Í *et al.* 2010, Chong *et al.* 2010, Yan *et al.* 2011). As for motorist and bicyclist-related factors in the existing safety research, the momentary activities of road users are regarded as more important factors in explaining bicyclist injury outcomes. Furthermore, the effects of using protective equipment and making improper driving behaviors have been evaluated frequently (Kim *et al.* 2007, B Í *et al.* 2010, Chong *et al.* 2010, Boufous *et al.* 2011, Moore *et al.* 2011). For instance, a research article showed that helmet use mitigated the severity of bicyclists' brain injury by more than 85.00% (Moore *et al.* 2011).

According to the existing research, environmental settings are also related to the severity of bicyclist injuries. Darkness, measured by the time of day, is an important factor associated with fatality (Klop and Khattak 1999, Eluru *et al.* 2008, B Ì *et al.* 2010, Boufous *et al.* 2011). Additionally, adverse environmental circumstances, such as wet surfaces, ice, and fog, increase the probability of serious bicyclist injuries (Moore *et al.* 2011).

In relation to road network features, it is safer to ride bicycles on signalized intersections with good lighting conditions (Eluru *et al.* 2008, B Ì *et al.* 2010, Zahabi *et al.* 2011). The factors correlated with intersection and mid-block bicyclist injury severity are slightly different (Klassen *et al.* 2014). Roadway classifications and on-street parking are the factors affecting the injury severity of the mid-block bicycle crashes (Klassen *et al.* 2014). Zahabi et al. found that compared to the collisions occurring at mid-blocks, bicyclist injuries occurred more frequently but less severely at intersections due to driving speed reduction at intersection areas (2011). Another study showed that the proportions of industrial and commercial land use were positively associated with evident bicyclist injuries (Narayanamoorthy *et al.* 2013). Other land use variables, such as land use mixture, population density, and road connectivity, showed no significant correlations with bicyclist injury severity (Zahabi *et al.* 2011).

Some challenging methodological issues are debated for the ordered categorical attribute of bicyclist injury severity (Savolainen *et al.* 2011, Yasmin and Eluru 2013, Mannering and Bhat 2014). Injury severity is usually classified into ordered categories of fatality, severe injury, evident injury, possible injury, and property damage only. How injury severities are correlated with the independent variables are frequently examined by ordered and unordered response models. However, injury severity types are interrelated in nature. Ordered logit model, also called proportional odds model, is capable of capturing the ordinal attribute across different levels of injuries (Mooradian *et al.* 2013). The underlying assumption of an ordered logit model forces the coefficients for covariates to remain constant for all injury types. In other words, the coefficients of one factor on all injury types are assumed to be in the same direction. However, some variables may decrease the probability of one injury type while increasing the probability of another. The impacts of some covariates can be biasedly reported under the ordered logit modeling framework.

On the other hand, injury severity is considered as a categorical variable which allows the independent variables to influence response levels differently in the unordered response models (Yasmin and Eluru 2013). In addition, minor bicycle collisions are widely underreported (de Geus *et al.* 2012, Wegman *et al.* 2012), particularly those occurred on local streets and in suburban areas. In this context, the more effective approaches of modeling are the unordered response models (Yasmin and Eluru 2013), such as multinomial logit and mixed logit models. Even though the nested logit model can capture the ordinal nature inherent in injury levels within nests, some studies found that the added complexity from the nested structure could not be justified by its limited improvement in the prediction accuracy (Abdel-Aty 2003, Mooradian *et al.* 2013).

Another appealing choice is the partial proportional odds model. It loosens the assumption of an ordinal data attribute. As a substitute to the ordered logit model, it assumes that a subset of explanatory variables affects an injury category independently (Mooradian *et al.* 2013). The generalized ordered logit model offers an even more flexible modeling framework, which relaxes the constant cutoff point across injury cases (Eluru 2013, Yasmin and Eluru 2013).

The spatial dependence among collision sites is another methodological issue debated in prior research (Savolainen *et al.* 2011, Castro *et al.* 2013). Two types of spatial dependencies, called "spatial spillover" and "spatial correlation," have been discussed (Castro *et al.* 2013). "Spatial spillover" causes the injury risk tendency at one place to affect the probability of injury at its nearby places. "Spatial correlation" results in the same type of places sharing similarities in injury risks. The spatial dependence will cause the cases to no longer be independent. Several ways of calculating spatial dependence have been applied to injury severity (Castro *et al.* 2013, Klassen *et al.* 2014).

4.3 Modeling approach

A generalized ordered logit model and a generalized additive model are employed in this chapter to discern how built environment features are associated with different bicyclist injury categories, where bicyclist socio-demographics and motorist momentary behaviors are adjusted as confounders. The generalized ordered logit model relaxes the ordered logit modeling framework by allowing covariates to affect bicyclist injury severities differently. In addition, a generalized additive model is selected to investigate the possible spatial dependence.

4.3.1 *Generalized ordered logit model*

The generalized ordered logit model is shown in *Equation 4-1,* which assumes the vector of unobserved utility having a cumulative distribution (Williams 2006, Agresti and Kateri 2011). More specifically, supposing an ordered categorical dependent variable Y_i has M values, the generalized ordered logit model produces a set of estimates, including $M-1$ cutoff points, at which Y_i can be dichotomized.

$$\ln\left(\frac{P(Y_i > j)}{1-P(Y_i > j)}\right) = \ln\left(\frac{g(\beta_j X_i)}{1-g(\beta_j X_i)}\right) = \alpha_j + \beta_j X_i,\ j = 1,2,...,M-1 \quad \textit{Equation 4-1}$$

The probability functions when Y_i is equivalent to each of the values *1, 2, ..., M* are presented in *Equations 4-2, 4-3* and *4-4* (Williams 2005, Eluru *et al.* 2008, Kaplan and Prato 2012).

$$P(Y_i = 1) = 1 - g(\beta_1 X_i) \quad \textit{Equation 4-2}$$

$$P(Y_i = j) = g(\beta_{(j-1)} X_i) - g(\beta_j X_i),\ j = 2,...,M-1 \quad \textit{Equation 4-3}$$

$$P(Y_i = M) = g(\beta_{(M-1)} X_i) \quad \textit{Equation 4-4}$$

Where *M* is the number of ordinal categories of bicyclist injury severities, $P(Y_i)$ is the probability of any given injury type for case *i*, *j* represents the cutoff points between different injury categories, X_i is the vector of fixed covariates that explain bicyclist injury severity, α_j is the intercept, and β_j is the vector of corresponding estimates. In the generalized ordered logit model, three cutoff points are specified, including the cutoff points between generalized ordered logit model and possible injury, possible injury and evident injury, evident injury and severe injury or fatality. *Equation 4-2* represents the probability of the first category (property damage only), *Equation 4-3* represents the probabilities for the middle categories (evident injury and possible injury), whereas *Equation 4-4* represents the probability of the last category (severe injury or fatality).

4.3.2 Generalized additive model

A generalized additive model is specified to capture the possible spatial dependence. The generalized additive model is built on a standard ordered logit model by adding a smooth function to account for the spatial dependence among crash locations (Wood 2006, Agresti and

Kateri 2011). An ordered logit model assumes the estimated coefficients for fixed covariates keep the same for all ordinal categories. The additional term, $\delta(S_i)$, is a smooth spline function made by the vector of latitude and longitude for spatial covariates, as listed in *Equation 4-5.*

$$\delta(\mathrm{S_i}) = \mathrm{S_i}(\mathrm{x_{coori}}, \mathrm{y_{coori}}) \qquad \textit{Equation 4-5}$$

The generalized additive model is shown in *Equation 4-6,* where the $i, j, \alpha, \beta, P(Y_i > j)$ *and* X_i are the similar items as specified in the generalized ordered logit model (Wang and Kockelman 2005).

$$\ln\left(\frac{P(Y_i > j)}{1 - P(Y_i > j)}\right) = \alpha + \beta X_i + \delta(\mathrm{S_i}) \qquad \textit{Equation 4-6}$$

The generalized ordered logit model is implemented with R “VGAM” package (Yee 2010), and generalized additive model is estimated using R “mgcv” package (Wood 2001).

4.3.3 Elasticity formulas for generalized ordered logit model

Elasticity calculating formulas of the generalized ordered logit model is derived as follows.

$$\frac{\partial P}{\partial X} = P(1-P)\frac{\partial V}{\partial X} \qquad \textit{Equation 4-7}$$

For all models, elasticity is written as $E = \frac{\partial P}{\partial X}\frac{X}{P}$. If *X* is a dummy variable, elasticity is computed as:

$$E = \frac{P_{x=1} - P_{x=0}}{P_{x=1} + P_{x=0}} \qquad \textit{Equation 4-8}$$

However, if *X* is a continuous variable, for this generalized ordered logit model, the elasticities for the four injury categories are computed as *Equations 4-9 to 4-12*:

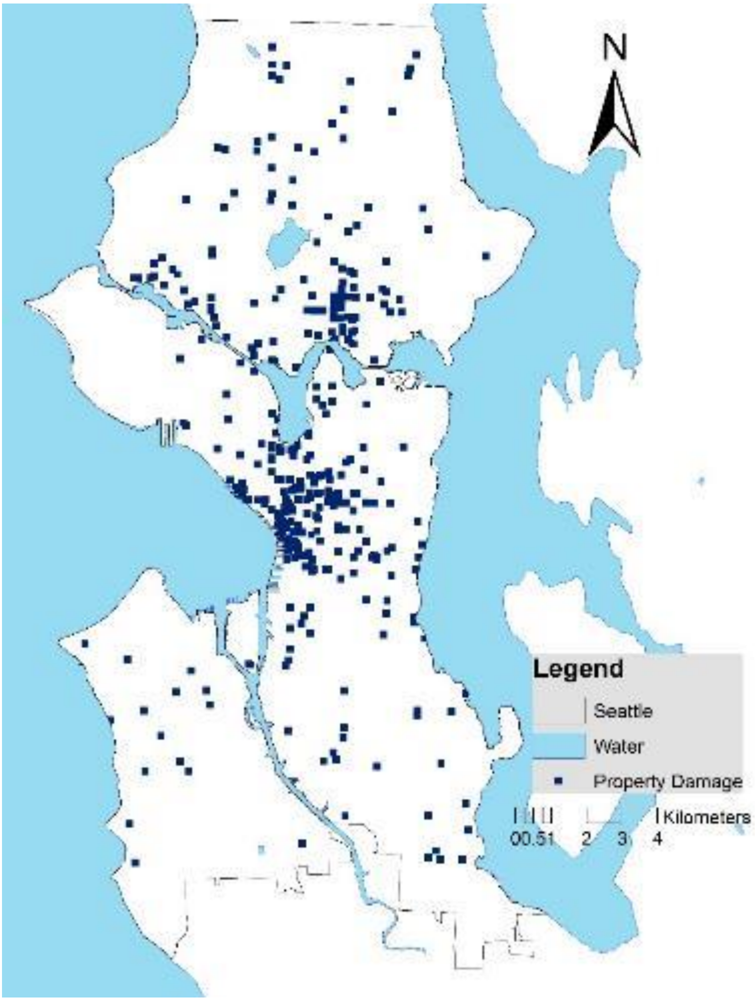
$$E_{PDO} = -\beta_1(1 - P_1)X_i$$ *Equation 4-9*

$$E_{PI} = \beta_1 * (P_2 + P_3 + P_4) * P_1 \frac{X_i}{P_2} - \beta_2 * (P_3 + P_4) * (P_1 + P_2)\frac{X_i}{P_2}$$ *Equation 4-10*

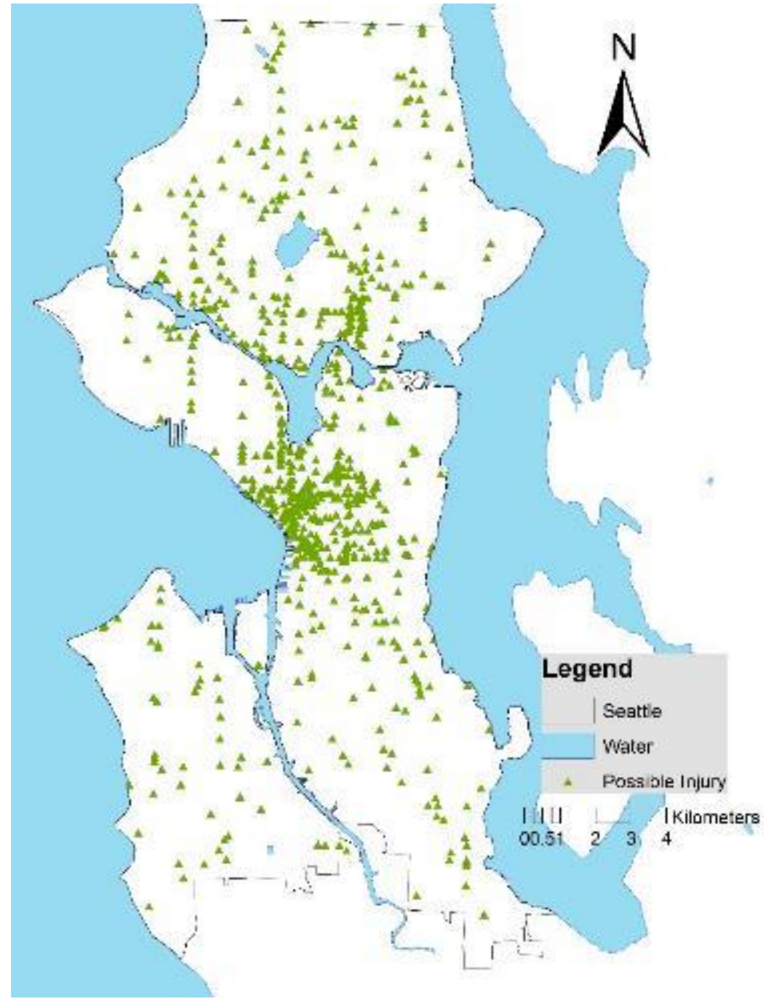
$$E_{EI} = \beta_2 * (P_3 + P_4) * (P_1 + P_2)\frac{X_i}{P_3} - \beta_3 * P_4 * (P_1 + P_2 + P_3)\frac{X_i}{P_3}$$ *Equation 4-11*

$$E_{SIF} = \beta_3(1 - P_4)X_i$$ *Equation 4-12*

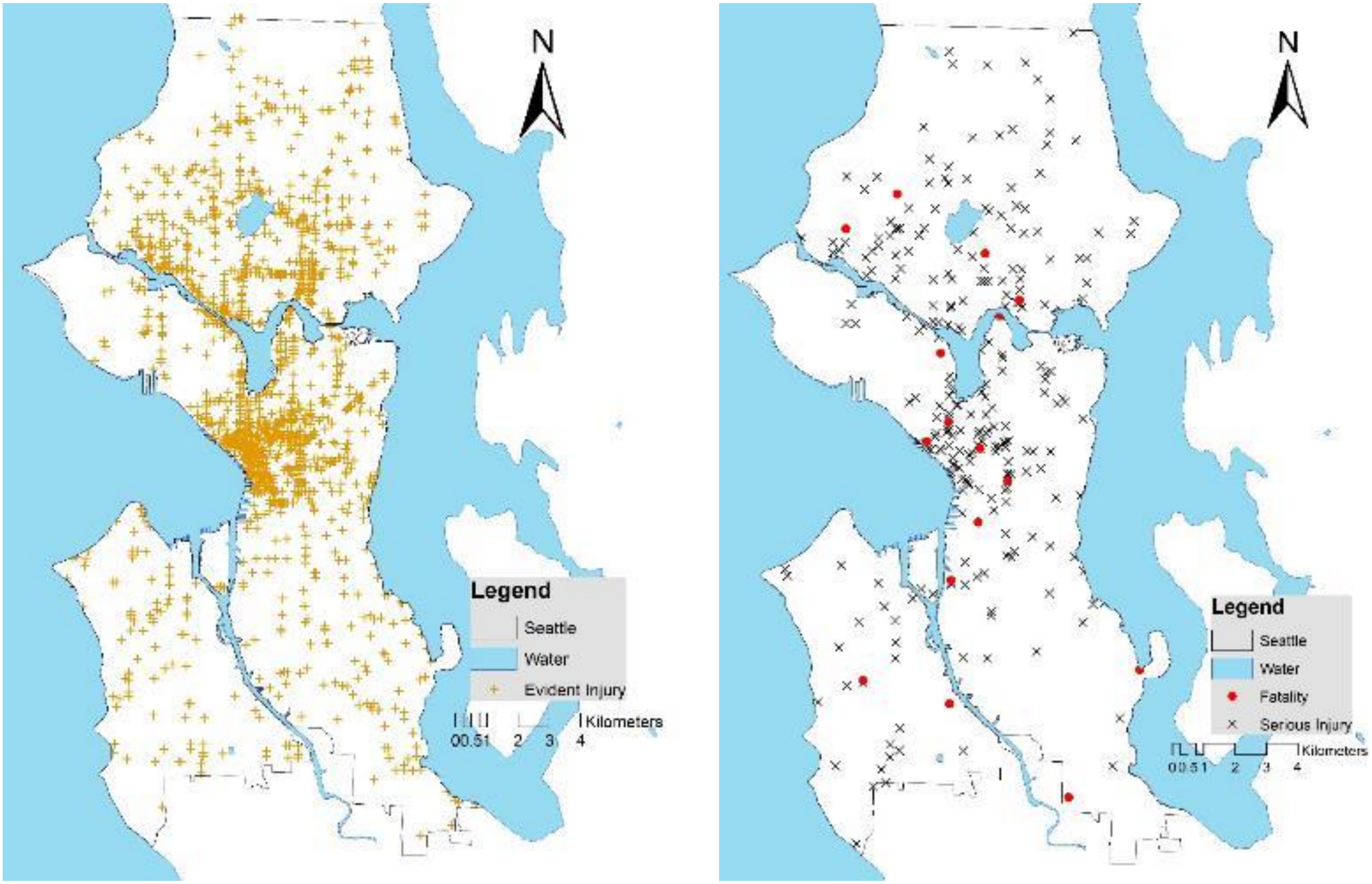
4.4 Data



(a) Property damage only



(b) Possible injury



(c) Evident injury

(d) Severe injury or fatality

Figure 4-1: Spatial distribution of bicyclist injury severity in Seattle, 2004-2013

Seattle is ranked third among large US cities in regards to the percentage of commuters using bicycles (Seattle Department of Transportation 2014). In addition, Seattle is implementing a bicycle master plan, which continually provides bicycle routes and contributes to the steady growth in the number of bicyclists. The total number of bicyclist injuries increased in the past decade due to increased use of bicycles, while the injury rate per bicyclist decreased in that period (Seattle Department of Transportation 2013).

The employed data for this empirical analysis includes two components: the bicycle crash records and the built environment features. The bicycle crash records were collected by Seattle Department of Transportation from January 2004 to March 2013. The geo-coded data, displayed in *Figure 4-1 (a) to (d),* helps locate the bicycle crashes and identify the built environment characteristics. Seven types of bicyclist injury severity categories were recorded in the original crash data, including three detailed categories for fatal crashes (dead at scene/on arrival/at hospitals). To ensure each injury type has a sufficient sample size to obtain valid statistical inference, severe injury and fatality are merged into one category. Hence, injury severities are aggregated into four ordered categories: property damage only, possible injury, evident injury, and severe injury or fatality.

However, missing values for various fields are observed in many records, such as bicyclist age, bicyclist gender, and posted speed limit. In total the sample had 3,310 crashes; the number of observed cases with injury severity reported was 2,911. After excluding records with missing values, a sample of 1,502 cases was acquired for the final model. The number and the percentage of each bicyclist injury category is presented in *Table 4-1*. The full sample with complete injury severity information and the final sample yield very similar descriptive statistics. Therefore, the estimation based on the final sample does not misrepresent the results greatly.

Table 4-1: The number and percentage of each bicyclist injury type in the crash data

Injury types	Property damage only	Possible injury	Evident injury	Severe injury	Fatality
Sample with injury severity information (2,911 cases)	267 (9.17%)	834 (28.65%)	1,551 (53.28%)	243 (8.35%)	16 (0.55%)
Final sample (1,502 cases)	113 (7.52%)	416 (27.70%)	826 (54.99%)	137 (9.12%)	10 (0.67%)

In the full sample, 16 bicyclists were fatally injured, and 10 of them remained in the final sample for modeling. Property damage only, possible injury, evident injury, and severe injury or fatality accounted for, respectively, 8%, 28%, 55% and 9% of the total. The other data acquired for modeling contains built environment features quantified through ArcGIS spatial overlay analysis. Seattle Department of Transportation and Puget Sound Regional Council provide GIS maps of the built environment. The data summary is shown in *Table 4-2*.

4.5 Conceptual framework and variable selection

Figure 4-2 shows the modeling framework that presents the variables assumed to be associated with bicyclist injury severity. Eleven categories of variables are considered, including features of land use, road network, road design, individual profiles of bicyclists and motorists, behavioral factors, vehicle information, motorist movement, and other conditions.

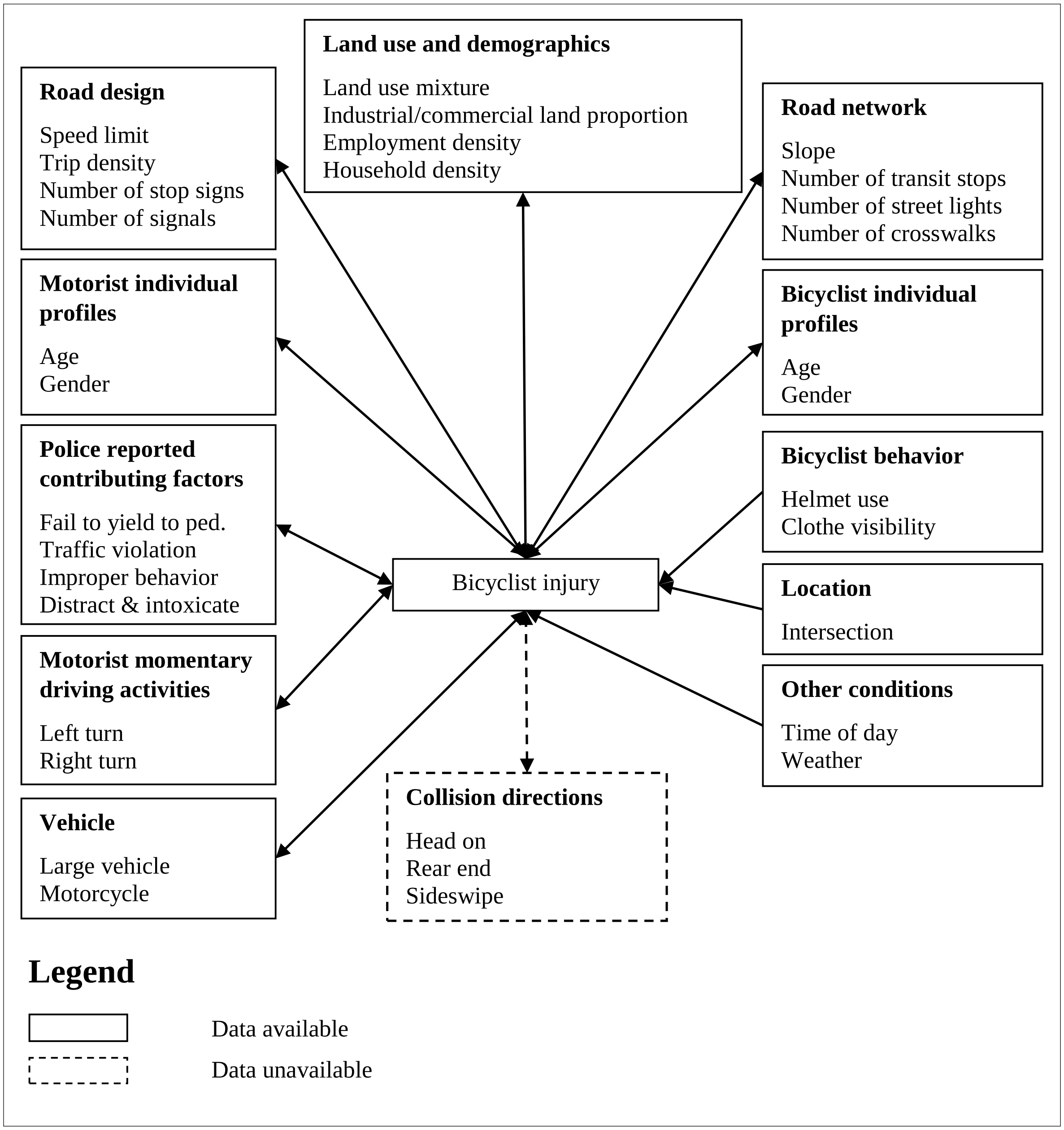


Figure 4-2: Modeling framework

In the final models, the factors of collision direction and motorist demographics are not included. Collision direction for bicycle crashes is unavailable in the Seattle Department of Transportation data. Motorist demographics are excluded due to a large percentage of missing values. Other conditions, time of day and weather, are included. As for bicyclist demographics, bicyclist age is included, but bicyclist gender is excluded due to insignificance.

Vehicle type is an important element frequently investigated in previous studies. Yet, only 62 large vehicles (more than 10,000 lbs, including trucks, farm tractors and buses) and 7 motorcycles were involved in the 1,502 bicycle crashes. Large vehicle is selected for modeling, while the motorcycle is excluded for limited observations.

The built environment features comprise the factors of land use and demographics, road network and design. Three land use and demographic variables— household density, employment density, land use mixture—are included for modeling. The slope and the number of street lights nearby collision sites are included. As for road design factors, because momentary driving speed when a collision occurred is not available, police-reported posted speed limit is included as a substitute. Other land use, road network and design features, such as the number of crosswalks, the proportion of commercial land use, the proportion of industrial land use, the number of stop signs, and the number of signals, are excluded due to insignificance or collinearity issues.

Wearing reflective clothing and equipping with helmets are included. The motorists' momentary driving activities, such as going straight, turning left, turning right, backing, stopping, starting, parking, changing lanes and merging traffic, are considered. Motorist turning left and right are the momentary driving activity included in the final models. Police-reported contributing factors, such as motorist performing improper behaviors, motorist being distracted or intoxicated, motorist violating traffic regulations, and motorist failing to yield to pedestrians, are considered but excluded from the final models for showing no statistical significance.

4.6 Descriptive analysis

Among the 1,502 observations, 74.30% of injured bicyclists were males, which is close to the national average of 77% (AARP and APTA 2014). The average age of injured bicyclists was 32.89 years old. In the sample, only 7.66% of bicyclists dressed reflective clothing, while 71.30% of them wore helmets. As for motorists' momentary activities, 28.10% of bicyclists were turning left, 19.44% were turning right, 34.75% were going straight, and less than 1.00% were backing. Regarding police-reported contributing factors, dozens of contributors were reported in the crash data. Failing to yield to pedestrians accounted for 48.14%, whereas another 15.84% were comprised of small proportions. Of this 15.84%, motorists performing improper behaviors, such as improper turns and following too close, accounted for 4.46%; traffic violations accounted for 4.46%; and being intoxicated, fatigued or distracted accounted for 6.92%.

The built environment features, such as land use mixture and street light density, are captured using ArcGIS overlay functions. The distance (the radius of buffers) used to quantify built environment factors is 50 meters in this study. The rationale to determine such distance is relative to the maximum width of a road surface and the minimum distance between the central lines of two neighboring streets. For example, if a road has six lanes (3.7 meters each) with barriers, collectors, and bike lanes on each side, the road surface could be 30 meters wide. In that case, 50 meter is a reasonable distance to capture the built environment features along the roads, such as land use nearby.

Table 4-2: Variable dictionary and summary of selected variables quantified at crash sites (n=1,502)

Variable	Mean	S.D.	Min.	Max.	Source	Description
Land use and demographics						
Household density	2.97	3.40	0.00	17.69	PSRC	Household density in corresponded TAZ, in 1k/km^2;
Employment density	9.97	22.47	0.00	208.56	PSRC	Employment density in corresponded TAZ, in 1k/km^2;
Land use mixture	0.39	0.09	0.13	0.64	PSRC	Measured by entropy of five types of land use, including residential, green spaces, offices, commercial and industrial, within 50-meter buffers of collusion sites;
Road network and design						
Number of street lights	2.88	2.00	0.00	10.00	SDOT	Number of street lights within 50-meter buffers of collusion sites;
Slope	0.03	0.02	0.00	0.13	SDOT	The slope of the crash site;
Speed limit	29.06	3.38	10.00	50.00	SDOT	Police-reported posted speed limit at crash locations, in mph;
Bicyclist socio-demographics						
Bicyclist age	32.89	12.76	3.00	79.00	SDOT	Age of injured bicyclist;
Other conditions						
Weather	0.12	---	0.00	1.00	SDOT	If the crash occurred in rainy and snowy days, 1, else 0;
Day time	0.20	---	0.00	1.00	SDOT	If the crash occurred in darkness, 1, else 0;
Vehicle type						
Large vehicle	0.04	---	0.00	1.00	SDOT	If a large vehicle is involved in a bicycle crash (>10,000 lbs, including trucks, farm tractors and buses) 1, else 0;
Bicyclist-related behavioral factors						
Bicyclist reflective suit	0.08	---	0.00	1.00	SDOT	Bicyclist dressed in reflective clothing 1, else 0;
Bicyclist helmet	0.71	---	0.00	1.00	SDOT	Bicyclist wore helmet, 1, else 0;
Motorist momentary driving activities						
Motorist left turn	0.28	---	0.00	1.00	SDOT	Crashes occurred when motorists turning left 1, else 0;
Motorist right turn	0.20	---	0.00	1.00	SDOT	Crashes occurred when motorists turning right 1, else 0;

4.7 Inferential analysis

4.7.1 Estimated results for generalized ordered logit model

Insignificant variables (*p*-value is greater than 0.1) on the three cutoff points are removed from the final generalized ordered logit model. *Table 4-3* shows the results of the estimated generalized ordered logit model. The estimated parameters (β_j) and *p*-values are shown in the 2-paralleled columns under the three cutoff points.

4.7.1.1 Land use, demographics, road network and design features

This research measures many built environment features to explore how they are correlated with bicyclist injury severity. Among the land use and demographic factors, employment density is negatively associated with bicyclist injury severity, whereas the land use mixture is negatively correlated with severe injury or fatality. Of road network features, the parameter estimated for the number of street lights is marginally significant at the first cutoff point, indicating that better lighting condition is associated with a lower probability of bicyclist injuries as compared to property damage only. Furthermore, the estimate for the slope is marginally significant at the third cutoff point, indicating that the probabilities of severe injury or fatality in steep streets are greater than other types of injury. As for road design, posted speed limit is a positive predictor of evident injury and severe injury or fatality.

Table 4-3: Generalized ordered logit modeling results (n=1,502)

Variable	Cutoff point Between property damage only and possible injury		Cutoff point Between possible injury and evident injury		Cutoff point between evident injury and severe injury or fatality	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Intercept	1.61	0.10	-0.70	0.21	-2.66**	0.00
Land use and demographics						
Household density	0.00	0.99	-0.01	0.43	-0.03	0.35
Employment density	-0.01***	0.00	-0.00*	0.05	-0.01	0.34
Land use mixture	0.96	0.43	0.31	0.65	-2.93**	0.01
Road network and design						
Number of street lights	-0.09	0.08	-0.03	0.24	-0.05	0.27
Slope	5.03	0.29	3.14	0.21	6.91	0.06
Speed limit	0.02	0.54	0.05**	0.00	0.04	0.15
Bicyclist individual profiles						
Bicyclist age	0.01	0.48	-0.01	0.20	0.01*	0.04
Other conditions						
Weather	0.02	0.95	-0.06	0.71	0.12	0.66
Day time	0.04	0.88	-0.04	0.78	-0.07	0.74
Vehicle type						
Large vehicle	0.16	0.76	0.02	0.95	1.11***	0.00
Bicyclist-related behavioral factors						
Bicyclist reflective suit	-0.82**	0.01	-0.41*	0.05	-1.13*	0.02
Bicyclist helmet	0.16	0.48	0.16	0.21	-0.10	0.60
Motorist momentary driving activities						
Motorist left turn	0.22	0.39	0.09	0.49	0.55**	0.01
Motorist right turn	-0.17	0.50	-0.19	0.20	-0.22	0.41
Indicated by estimate with level of significance (*<0.05, **<0.01, ***<0.001)						
Degrees of freedom: 4,461						
Model fit	Residual deviance: 3,246.80			Log-likelihood: -1,623.40		

Table 4-4: Generalized additive modeling results (n=1,502)

	Estimate	*p*-value
Intercept	0.43	0.40
Land use		
Household density	-0.01	0.53
Employment density	-0.01*	0.04
Land use mixture	-0.34	0.60
Road network and design		
Number of street lights	-0.04	0.11
Slope	4.77*	0.04
Speed limit	0.04**	0.00
Bicyclist socio-demographics		
Bicyclist age	0.00	0.98
Other conditions		
Weather	0.00	0.99
Day time	-0.02	0.84
Vehicle type		
Large vehicle	0.38	0.16
Bicyclist-related behavioral factors		
Bicyclist reflective suit	-0.56**	0.00
Bicyclist helmet	0.08	0.48
Motorist momentary driving activities		
Motorist left turn	0.20	0.10
Motorist right turn	-0.18	0.18
Indicated by estimate with level of significance (*<0.05, **<0.01, ***<0.001)		
Significance of smooth terms	Degrees of freedom	*p*-value
S(X coordinate, Y coordinate)	5.76	0.40
Deviance explained = 22.40%	- Restricted maximum likelihood = 1,670.800	

4.7.1.2 Motorist and bicyclist-related factors

For bicyclists, age is positively correlated with severe injury or fatality, while wearing reflective clothing is negatively correlated with injury severity. Motorist turning left is positively correlated with the probability of bicyclist injuries.

4.7.1.3 Vehicle type

The involvement of a large vehicle is positively associated with severe injury or fatality.

4.7.2 Estimated results for generalized additive model

To capture the ordered categorical nature of injury severity, and to measure the possible spatial dependence among bicycle collision sites, a generalized additive model is further estimated. This generalized additive model is specified as an ordered logit model plus a spline smoothing function. *Table 4-4* presents the modeling results. Among significant predictors of bicyclist injury severity in the generalized ordered logit model, five of them show highly consistent effects in the generalized additive model: slope, employment density, posted speed limit and bicyclist dressing reflective clothing are significant factors, and motorist turning left is only marginally significant.

The *p*-value estimated for the smoothing terms is not significant (*p*-value = 0.40), which suggests that the "spatial spillover" or the "spatial autocorrelation" effect is not observed among the collision sites. Therefore, bicyclist injury severity is not spatially auto-correlated in Seattle.

4.7.3 Model comparison

The generalized ordered logit model offers richer insights on specific injury cutoff points and injury types, and considers the ordered categorical nature of the injury severity variable. The generalized additive model has the advantage in accounting for the spatial dependence. For this study, the generalized ordered logit model provides greater details into the relationship between bicyclist injury severity and the built environment.

4.7.4 Elasticity effects from the generalized ordered logit model

The elasticities calculated for the preferred generalized ordered logit model are shown in *Table 4-5*, measured for the four types of bicyclist injury severities accordingly. The numbers are interpreted as the percentage change in the probability of an individual bicyclist injury severity category due to 1.00% change in the explanatory variables.

The outcomes suggest that severe injury or fatality is generally more sensitive to the explanatory variables than other injury types. The largest elasticity in absolute value is associated with land use mixture, indicating that a 1.00% increase in this measure is associated with a 1.13% reduction in the probability of a bicyclist experiencing severe injury or fatality. Also quite strongly relating to decreases in the probabilities of severe injury or fatality are bicyclists dressing reflective suit and the number of street lights within the buffer area, with elasticities of -0.49 and -0.14 accordingly. It is worth noting that a 1.00% increase in slope is associated with a 0.20% increment in the probability of severe injury and fatality.

On the other hand, increasing posted speed limit is most strongly associated with increased probability of a bicyclist experiencing severe injury or fatality; specifically, a 1.00% increase in posted speed limit is related to a 0.95% higher probability of severe injury or fatality. Involvements of a large vehicle and an older bicyclist in bicycle crash are also quite strongly associated with higher probabilities of severe injury or fatality, with calculated elasticities of 0.46 and 0.30 accordingly.

Table 4-5: Elasticities calculated for the generalized ordered logit model

	Property damage only		Possible injury		Evident injury		Severe injury or fatality	
	Elasticity	S.D.	Elasticity	S.D.	Elasticity	S.D.	Elasticity	S.D.
Household density	0.00	0.00	0.01	0.01	-0.02	0.03	-0.08	0.10
Employment density	0.07	0.13	-0.08	0.62	0.01	0.02	-0.10	0.22
Land use mixture	-0.33	0.08	0.02	0.00	0.19	0.02	-1.13	0.27
Number of street lights	0.20	0.13	-0.06	0.06	-0.08	0.06	-0.14	0.10
slope	-0.14	0.10	-0.01	0.01	0.05	0.02	0.20	0.14
Speed limit	-0.52	0.06	-1.21	0.24	-0.15	0.09	0.95	0.09
Bicyclist age	-0.29	0.12	0.16	0.04	-1.23	1.09	0.30	0.11
Weather	-0.01	---	0.02	---	-0.04	---	0.06	---
Day time	-0.02	---	0.01	---	-0.01	---	-0.03	---
Large vehicle	-0.07	---	0.02	---	-0.25	---	0.46	---
Bicyclist reflective suit	0.30	---	-0.07	---	-0.08	---	-0.49	---
Bicyclist helmet	-0.07	---	-0.02	---	0.07	---	-0.05	---
Motorist left turn	-0.09	---	0.01	---	-0.04	---	0.25	---
Motorist right turn	0.07	---	0.02	---	-0.06	---	-0.10	---

4.8 Conclusions and future research

Employing Seattle's data, this research has investigated the associations between bicyclist injury severity and several categories of factors, including bicyclist socio-demographics, vehicle type, road design, motorist momentary behavior, road network, and land use. The modeling results are largely consistent with the significant effects of a number of important variables identified in the existing literature. First, older bicyclists are more likely to be severely injured in bicycle crashes. Second, large vehicle-involved bicyclist injuries are more likely to be severe. Third, street lighting condition is negatively associated with severe injury or fatality. Finally, posted speed limit is positively correlated with evident injury and severe injury or fatality.

This research has gained additional insights in understanding the importance of particular behavioral factors and the built environment features. Regarding behavioral factors, bicyclists dressing reflective clothing suggest a lower probability of injuries, and crashes happened while motorists are turning left are more likely to result in severe injuries. As for the built environment features, employment density is negatively associated with bicyclist injury severity, the slope is positively associated with severe injury or fatality, and land use mixture is negatively associated with severe injury or fatality.

These findings suggest that certain environmental treatments and road improvements can improve bicycle safety. Several important policy recommendations can be obtained from this analysis. First, the safety effects of diversifying land use and densifying employment should be considered in zoning. Second, improving street lighting conditions to bicycle routes, placing bicycle lanes in flat areas, and optimally lowering posted speed limit for streets should be considered where motorists and bicyclists share road spaces. Third, bicyclists are encouraged to dress reflective clothing. These principles should be jointly considered in bicycle master plans.

While the outcome indicates the benefit of lowering posted speed limits for the sake of safety. Safety concerns should be balanced with other road users' mobility requirements. Perhaps a more feasible approach to mitigate the conflict is to add more separated bike lanes to isolate bicyclists from road traffic. However, since bicycling accounts for only a small proportion of the transportation mode share, lowering road speed limits of the whole transportation system may not be cost-effective.

Future research may investigate continuous measurements of the built environment. An assumption made in this research is that the built environment is stable in the ten-year period as quantified in the GIS maps. This assumption may not be realistic for many locations. . For instance, employment densities in some areas may have increased over years, which no longer represent the true employment densities correlated with crashes happened earlier. Even though the built environment is relatively stable in most US cities, measuring the built environment more frequently is desirable.

References

AARP, APTA, 2014. The alliance benchmarking report: Bicycling and walking in the United States. http://www.aarp.org/content/dam/aarp/livable-communities/documents-2014/2014-Bike-Walk-Benchmarking-Report.pdf.

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of safety research 34 (5), 597-603.

Agresti, A., Kateri, M., 2011. Categorical data analysis Springer.

American Association of State Highway and Transportation Officials and US Departmemt of Transportation, 2013. Commuting in America 2013: the national report on commuting patterns and trends. 7-8.

Attewell, R.G., Glase, K., Mcfadden, M., 2001. Bicycle helmet efficacy: A meta-analysis. Accident Analysis & Prevention 33 (3), 345-352.

B ĺ, M., B ĺová, M., Müller, I., 2010. Critical factors in fatal collisions of adult cyclists with automobiles. Accident Analysis & Prevention 42 (6), 1632-1636.

Boufous, S., Rome, L.D., Senserrick, T., Ivers, R., 2011. Cycling crashes in children, adolescents, and adults a comparative analysis. Traffic Injury Prevention 12 (3), 244-250.

Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis & Prevention 52, 188-203.

Chong, S., Poulos, R., Olivier, J., Watson, W.L., Grzebieta, R., 2010. Relative injury severity among vulnerable non-motorised road users: Comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions. Accident Analysis & Prevention 42 (1), 290-296.

Cummings, P., Rivara, F.P., Olson, C.M., Smith, K., 2006. Changes in traffic crash mortality rates attributed to use of alcohol, or lack of a seat belt, air bag, motorcycle helmet, or bicycle helmet, united states, 1982–2001. Injury Prevention 12 (3), 148-154.

De Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2012. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accident Analysis & Prevention 45, 683-693.

Eluru, N., 2013. Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. Accident Analysis & Prevention 55, 1-11.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40 (3), 1033-1054.

Goldenbeld, C., Houtenbos, M., Ehlers, E., De Waard, D., 2012. The use and risk of portable electronic devices while cycling among different age groups. Journal of Safety Research 43 (1), 1-8.

Kaplan, S., Prato, C.G., 2012. Risk factors associated with bus accident severity in the united states: A generalized ordered logit model. Journal of Safety Research 43 (3), 171-180.

Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis & Prevention 39 (2), 238-251.

Klassen, J., El-Basyouny, K., Islam, M.T., 2014. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A city of edmonton case study. Safety science 62, 295-304.

Klop, J.R., Khattak, A.J., 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. Transportation Research Record: Journal of the Transportation Research Board 1674 (1), 78-85.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1 (0), 1-22.

Mooradian, J., Ivan, J.N., Ravishanker, N., Hu, S., 2013. Analysis of driver and passenger crash injury severity using partial proportional odds models. Accident Analysis & Prevention 58 (0), 53-8.

Moore, D.N., Schneider Iv, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accident Analysis & Prevention 43 (3), 621-630.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B: Methodological 55 (0), 245-264.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis & Prevention 43 (5), 1666-1676.

Seattle Department of Transportation, 2013. 2012 traffic report.

Seattle Department of Transportation, 2014. Seattle master bicycle plan.

The National Highway Traffic Safety Administration, 2012. A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system. http://www-nrd.nhtsa.dot.gov/Pubs/812032.pdf.

The National Highway Traffic Safety Administration, 2013. Traffic safety facts: Bicyclists and other cyclists. http://www-nrd.nhtsa.dot.gov/Pubs/812018.pdf.

Walker, I., 2007. Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. Accident Analysis & Prevention 39 (2), 417-425.

Wang, X., Kockelman, K.M., 2005. Occupant injury severity using a heteroscedastic ordered logit model: Distinguishing the effects of vehicle weight and type. Transportation Research Record 1908, 195-204.

Wegman, F., Zhang, F., Dijkstra, A., 2012. How to make more cycling good for road safety? Accident Analysis & Prevention 44 (1), 19-29.

Williams, R., 2005. Gologit 2: Generalised logistic regression. Partial Proportional.

Williams, R., 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. Stata Journal 6 (1), 58-82.

Wood, S., 2006. Generalized additive models: An introduction with R CRC press.

Wood, S.N., 2001. Mgcv: Gams and generalized ridge regression for R. R news 1 (2), 20-25.

Yan, X., Ma, M., Huang, H., Abdel-Aty, M., Wu, C., 2011. Motor vehicle-bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. Accident Analysis & Prevention 43 (5), 1751-1758.

Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis & Prevention 59, 506-521.

Yee, T.W., 2010. The vgam package for categorical data analysis. Journal of Statistical Software 32 (10), 1-34.

Zahabi, S.a.H., Strauss, J., Manaugh, K., Miranda-Moreno, L.F., 2011. Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. Transportation Research Record: Journal of the Transportation Research Board 2247 (1), 81-90.

Chapter 5. Synthesis: contributions, implications, and future research

This concluding chapter: (1) synthesizes the findings; (2) highlights the scholarly contributions; (3) draws implications that can inform researchers, transportation planners, and policy makers; and (4) discusses the directions for future research.

5.1 Summary of findings

This dissertation investigates the determinants of bicycling in the US context for route preference and road safety. Specifically, this dissertation has employed the theory of *"utility maximization"* to understand bicycle route choice and its connection with built environment features. Also, in referencing to the theory of *"safety in numbers"*, this dissertation has compared the differences between the frequency of and the risk of bicycle crashes, and has highlighted the bicycling risk factors in the built environment, as presented in *Table 5-1*.

5.1.1 Land use features

Improving density and land use mixture to build a safe bicycling environment is beneficial for encouraging bicycling. As shown in *Table 5-1*, for bicycle safety, household density is negatively associated with bicycle crash risk, and employment density is negatively correlated with bicyclist injury severity and bicycle crash risk. In addition, for bicycle route choice, many bicyclists prefer bicycle routes surrounded by lands of low floor area ratio. Some bicyclists prefer bicycle routes surrounded by mixed land use and near waters and parks. Though bicycle routes with higher land use mixture may have higher rates of collisions, the land use mixture is negatively associated with the injury severity incurred in bicycle crashes. On the other hand, the percentage of commercial land use is positively correlated with the risk of involving a bicycle crash.

Table 5-1: Significant built environment variables in the three studies

Variable	Route choice	Crash frequency	Crash risk	Injury severity
Household density	---		Negative	
Employment density	---		Negative	Negative
Floor area ratio	Negative	---	---	---
Proportion of commercial and mixed land	---		Positive	---
Proportion of parks and waters	Positive/Negative	---	---	---
Land use mixture	Positive	Positive		Negative
Slope	Negative			Positive
Street lighting	Positive			Negative
Posted speed limit	Negative	Positive	Positive	Positive
Intersection density	---	Negative	Negative	---
Signal density	---	Positive	Positive	---
On-arterial bike lanes (bike lanes)	---	Positive	Positive	---
Off-arterial bike lanes (cycle tracks + bike boulevards)	---	Negative	Negative	---
Prioritized bike lanes (cycle tracks + bike lanes)	Positive	---	---	---
Street tree density	Positive		Negative	---
Note: (1) Positive: positive and significant; (2) Negative: negative and significant; (3) ---: considered but excluded for being insignificant or multicollinearity; (4) Blank: included but not significant.				

5.1.2 Road network variables

As listed in *Table 5-1*, bike lanes in hilly areas are associated with more severe bicyclist injuries. In addition, bicycle facility types affect route choice, crash frequencies, and crash risks. A denser street network is negatively correlated with bicycle crash frequencies and risks. Of roadway design elements, street tree density is positively associated with bicyclists' route preferences, and negatively associated with bicycle crash risks. Improving street lighting is likely to attract bicyclists and reduce the risk of severe injury or fatality. Furthermore, both bicycle crash frequency and bicycle crash risk are positively correlated with signal density. Of utmost importance, higher posted speed limit detrimentally impacts bicycle route choice and safety outcomes.

5.2 Contributions

5.2.1 Conducting interdisciplinary research

Interdisciplinary research is motivated by drawing findings from different fields, and by employing diverse methodologies and concepts. With the idea of applying interdisciplinary research, this dissertation transcends the conventional boundary of urban planning by extending to transportation engineering, public health, and statistics. Consequently, the three studies of this dissertation are published in interdisciplinary journals or presented at different conferences. Both the practical and academic values of this dissertation interest a wide body of groups, such as government officers and researchers.

5.2.2 Applying advanced statistical methods

This dissertation applies multiple advanced statistical models to understand bicycle planning research questions. Building on quantitative reasoning, three different statistical approaches were employed to examine spatial relationships. Description, estimation, comparison, interpretation, and limitation of analytical techniques for planning practice are also covered in this dissertation.

5.2.3 Examining theories

The dominant theories explaining the relationship between individual bicycling behavior and the built environment in this dissertation are *utility maximization* and *safety in numbers*. Convincing empirical evidence validating these theories is identified and discussed in this dissertation.

5.2.4 Drawing practical policy implications

Beyond the academic contribution, this dissertation identifies and delivers explicit policy recommendations to policy makers, transportation planners and engineers, and advocacy groups. The substantial findings related to public policy and planning practice are separately stated in previous chapters and collectively summarized in this chapter.

5.3 Policy implications

Through multiple statistical analyses, this dissertation gains rich insights towards optimized promotion of safe and efficient bicycling. Utilizing the advancement in GPS data collection and well-documented bicycle crashes, the results of this project could assist in increasing bicycle route attractiveness and minimizing bicycling risks.

Concerning land use features, policy makers should (1) encourage mixed land use, (2) utilize green spaces and water resources in planning bicycle facilities, and (3) minimize the potential conflicts between motorists and bicyclists near commercial areas.

To provide a safe road network, policy makers and planners should focus on the slopes, bicycle facility types, intersection density, and street trees. Transportation planners should avoid placing bike lanes in steep areas and should separate road traffic and bicycle traffic as much as possible. Separation is a desirable solution to balance the velocity discrepancies between motorists and bicyclists. To improve street connectivity for better bicycling environments, transportation planners should design dense blocks, where the traffic operating speeds are likely to be slower. Authorities should plant more street trees, because more street trees contribute to improved attractiveness of bicycling environments and reduced risks of bicycle crashes.

Considering road design variables, policy makers and planners should make priorities for managing road signals and posted speed limits. The conflicts between motorists and bicyclists are likely to happen in areas with greater signal densities. Authorities should promote surveillance in signalized areas to monitor possible traffic violations. In addition, wherever conflicts may occur between bicyclists and motorists, posted speed limits should be adjusted to an acceptable level to ensure safe bicycling. As suggested, greater bicycle mode share contributes to reduced bicycle crash risks, but no significant increase in either injury severity or crash frequency. Though more trips are associated with reduced bicycle crash risks, more trips are also associated with a greater number of bicycle crashes. Regardless the overall effect, it is not sensible for the authorities to encourage more driving.

To summarize, as supported by the theories of *utility maximization* and *safety in numbers*, planning strategies advocated for the compact development provide appealing solutions in building an attractive bicycling environment. Neighborhood changes to the built environment should focus on but not limit to increasing commercial land use, green spaces and waters, mixed land use, household density, employment density, street lighting, prioritized bike facilities, and intersection density, and decreasing slopes and posted speed limits.

5.4 Future research

To promote evidence-based policy making, assist route planning, improve road safety, guide municipal investments, and establish maintenance priorities, future bicycle research beyond this dissertation may focus on the following issues. The first future research direction highlights the possible ways to advance data collection and integration, and the next three future research directions contribute to a more in-depth understanding of bicycling.

5.4.1 Utilizing emerging data

More robust data would lead to better-informed bicycle policy making, cost-effective bicycle infrastructure investments, and overall improved bicycling outcomes. For instance, current bicycle crash data include police-reported collisions, online self-reported crashes, hospital trauma reviews, and insurance company records. The integration of such data is not well-developed. With the increased complexity of analysis activities, there is an urgent need to utilize all available data resources effectively.

Collecting and utilizing camera data for bicycling research is an unexplored field. Camera data can assist in the evaluation of infrastructure quality and in the measurement of bicyclists' perceived risks. In general, bicycle use is closely attached to the perceived risks of bicycling. The joint utilization of GPS trackers and cameras can build rich data to investigate bicyclists' instant reactions in response to different road environments.

Utilizing smartphone applications to assist research is increasingly popular for researchers. The development of an app to promote bicycle safety through the collaboration of police, health professionals, and injured bicyclists can lead to more efficient reporting. App-recorded data are easily standardized, and are accessed across cities, states, or even nations. Police and health professionals can use the same patient ID to fill out different sections requested in the app under the same response menu. This would link police-reported data and trauma interviews. Individuals can participate in reporting. In that way, the widely underestimated minor collisions can be better captured by authorities and documented for researchers. This assists authorities to notice how many minor crashes are occurring. In addition, the design of such an app could minimize the possible errors in transcribing from paper forms to computers. Finally, on-site pictures could be taken by police for health professionals to reevaluate the crash outcomes.

5.4.2 Analyzing bicycle volume

This dissertation project has done research on the interactions between individuals and the built environment on *exposure* and *access*; however, what is not yet clear is the impact of bicycle *use*. Barely any bicycle use studies have been done due to the lack of well-recorded bicycle count data. From 2011 January, the City of Seattle began to count bicycles at 50 selected locations. These data could be used to investigate two questions: (1) Considering weather conditions, and adjusting for effects of seasons, weekends, and peak hours, which built environment features are associated with bicycle counts? (2) Does the implementation of the bicycle master plan and of the cycle share program predict the bicycle volume? The final output of such research would be used to predict bicycle volume for all road intersections.

5.4.3 Exploring perceived risk and route acceptability

Perceived risk and route acceptability are obstacles to promote bicycling. Many people fear cycling in the US. A better understanding of these issues assists planners to improve bicycle facilities from a psychological perspective, as well as provide policy makers evidence for policy design and decision making. Current research on perceived risk relies greatly on data collected through interviews on bicyclists' experiences and perceptions of riding in different built environments. However, retrospective information may not be consistent with real behaviors. The joint use of camera data, GPS data, interviews, and surveys provides an opportunity to better understand why people are reluctant to ride bicycles. The findings may provide informative insights for bicycle planning.

5.4.4 Understanding cycling risks in different built environments

Research on bicyclist injury severity tends to overlook cities that have relatively few bicyclists, and consequently, cities that have different levels of bicycling and different built environments for bicycling have not been analyzed. As cities make bicycling a key component of their urban transportation plans in response to the growing demand, it is time to initiate serious research efforts that improve our understanding of the connections between different built environments and bicycling collisions. The new knowledge can effectively inform planning and policy making, and thus help transform our cities into safer places for bicycling. Future research advancing the understanding of the effects of built environment features on bicyclist injury severity and crash frequencies should be conducted with the following question: are bicycling risks significantly different among cities with different levels of bicycling and different built environments?

Appendix A. Generated alternative bicycle routes

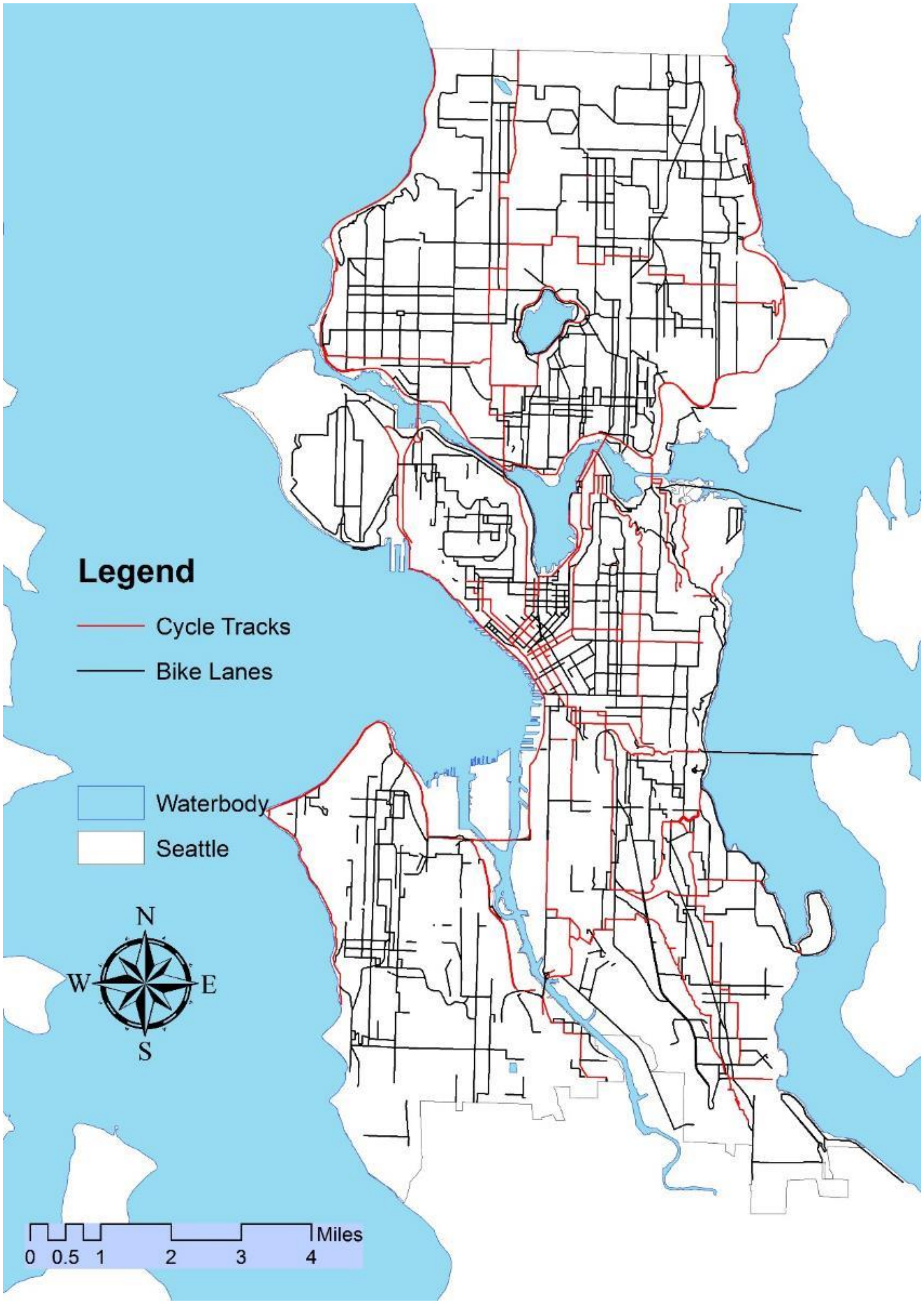


Figure* A-1*: Bicycle facilities in Seattle

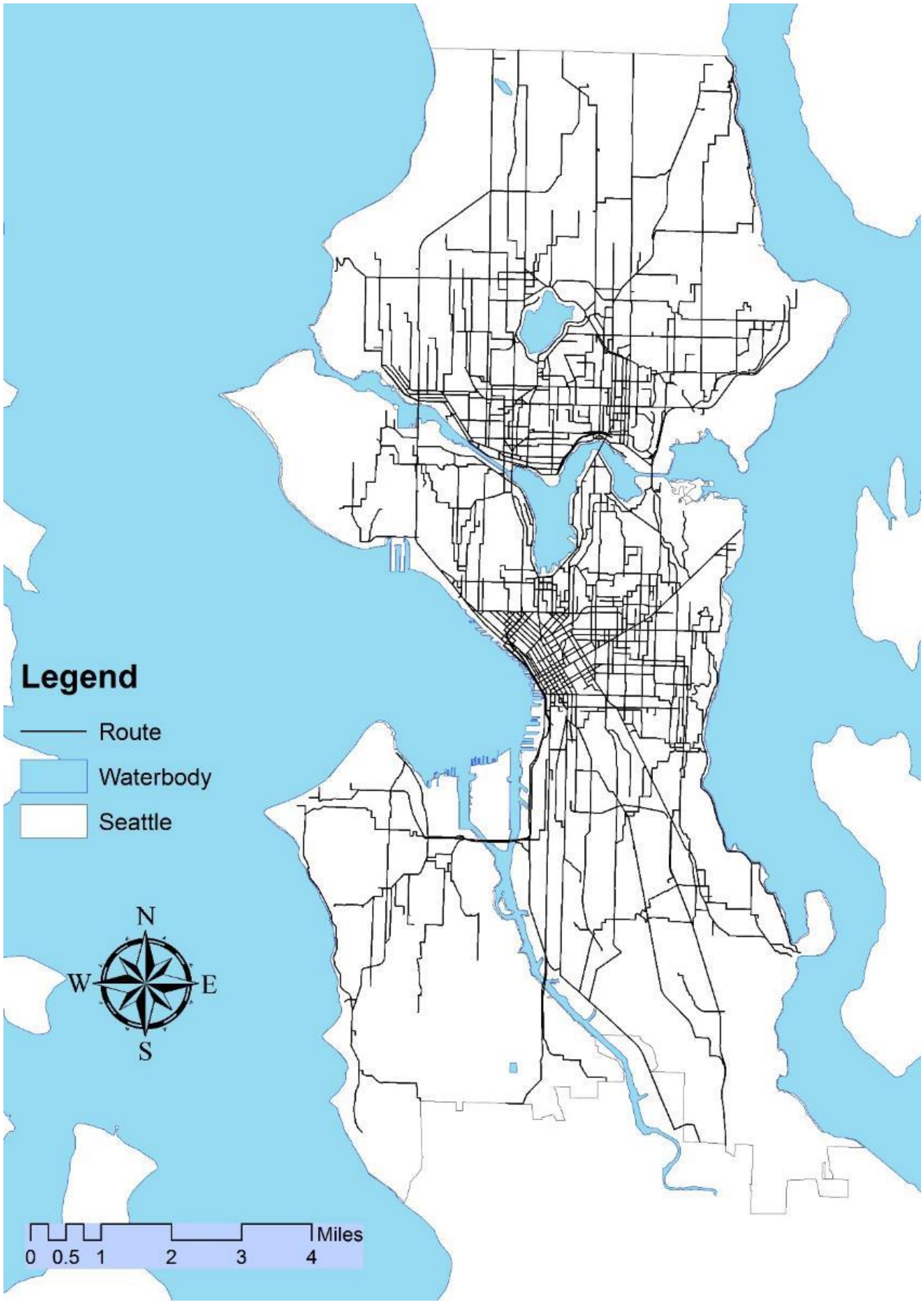


Figure A-2: Alternative routes labeled by minimizing trip distance

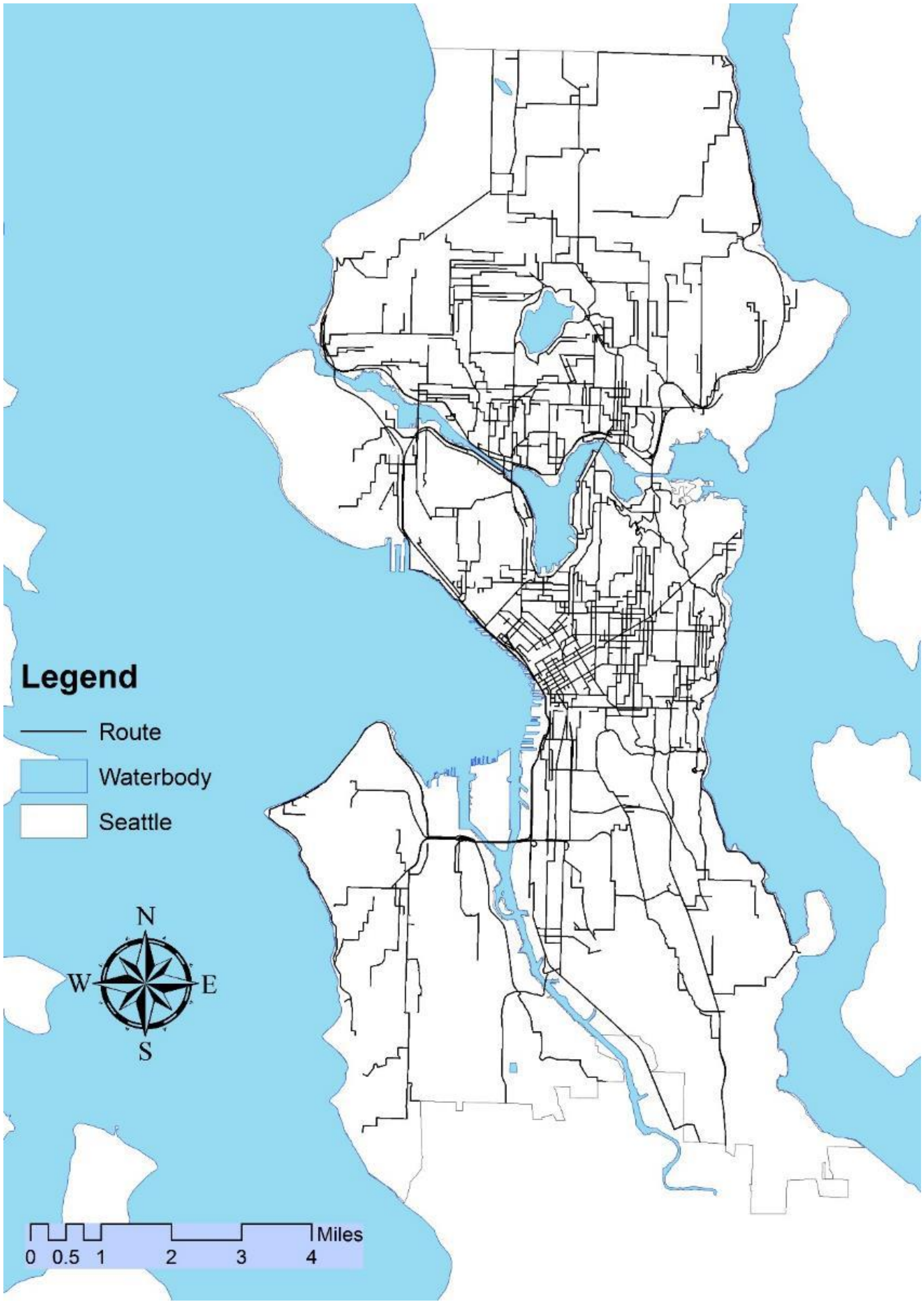


Figure A-3: Alternative routes labeled by minimizing slope

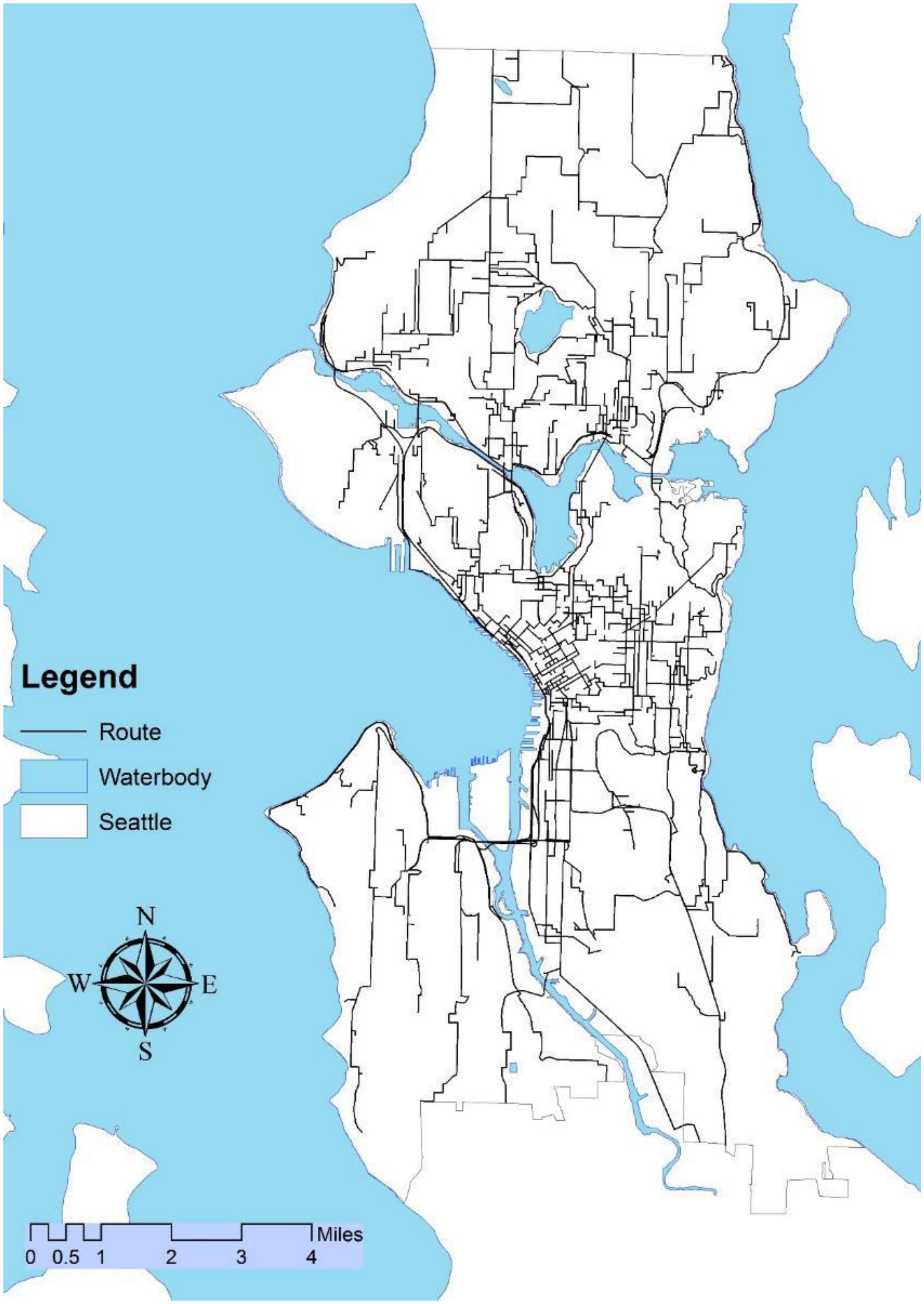


Figure A-4: Alternative routes labeled by minimizing floor area ratio

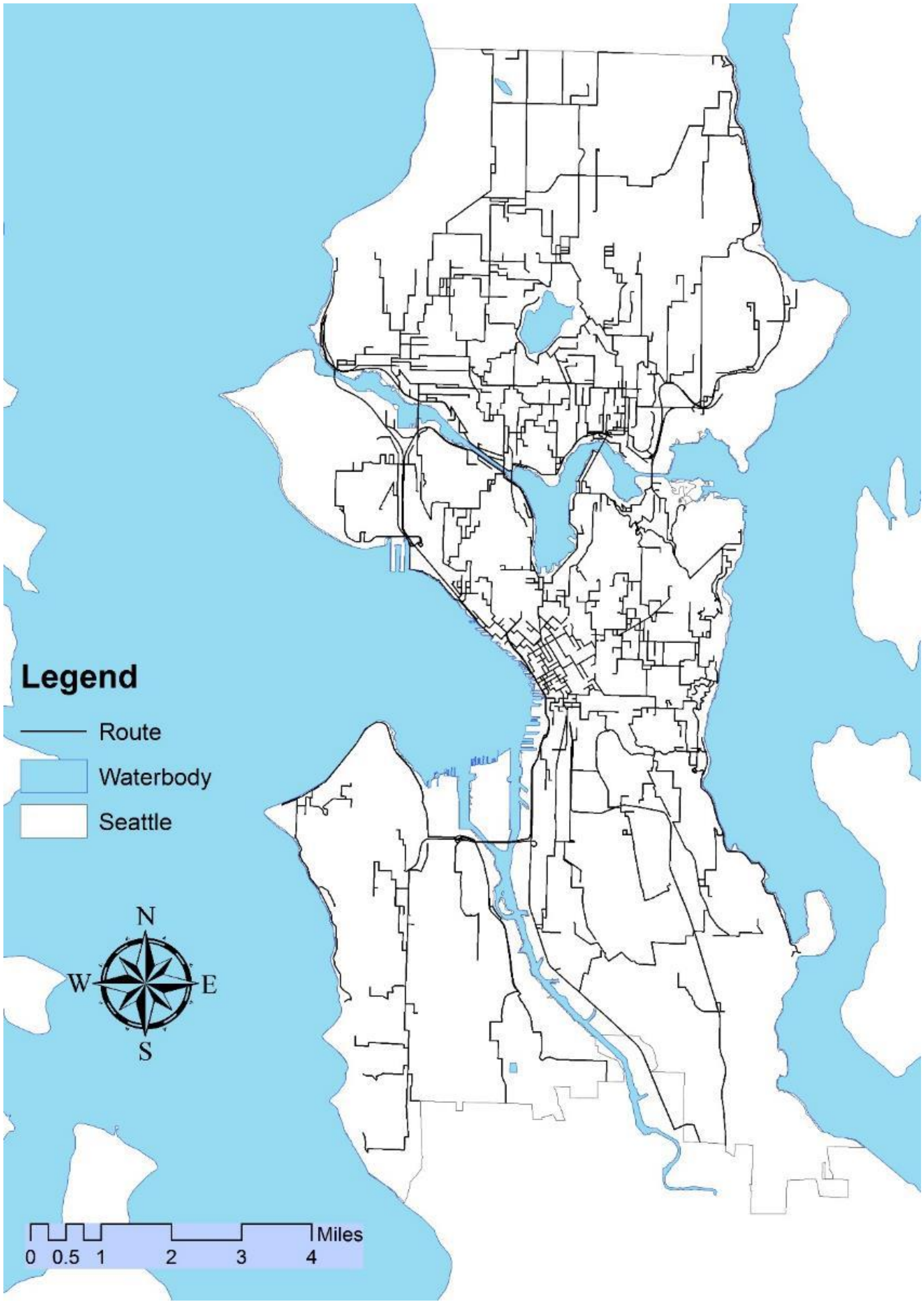


Figure A-5: Alternative routes labeled by maximizing city features

Figure A-6: Alternative routes labeled by maximizing waters and parks

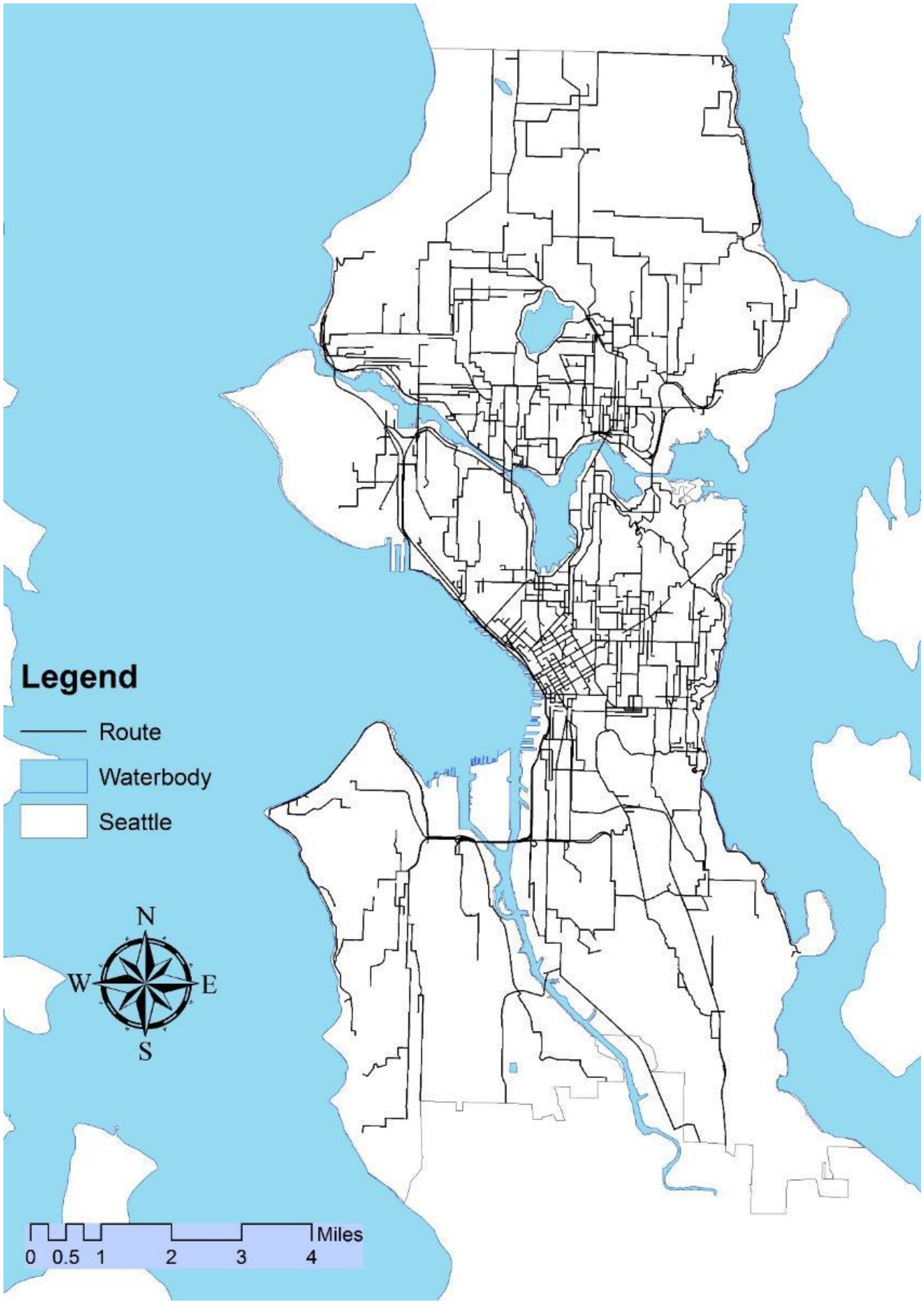


Figure A-7: Alternative routes labeled by maximizing street lights

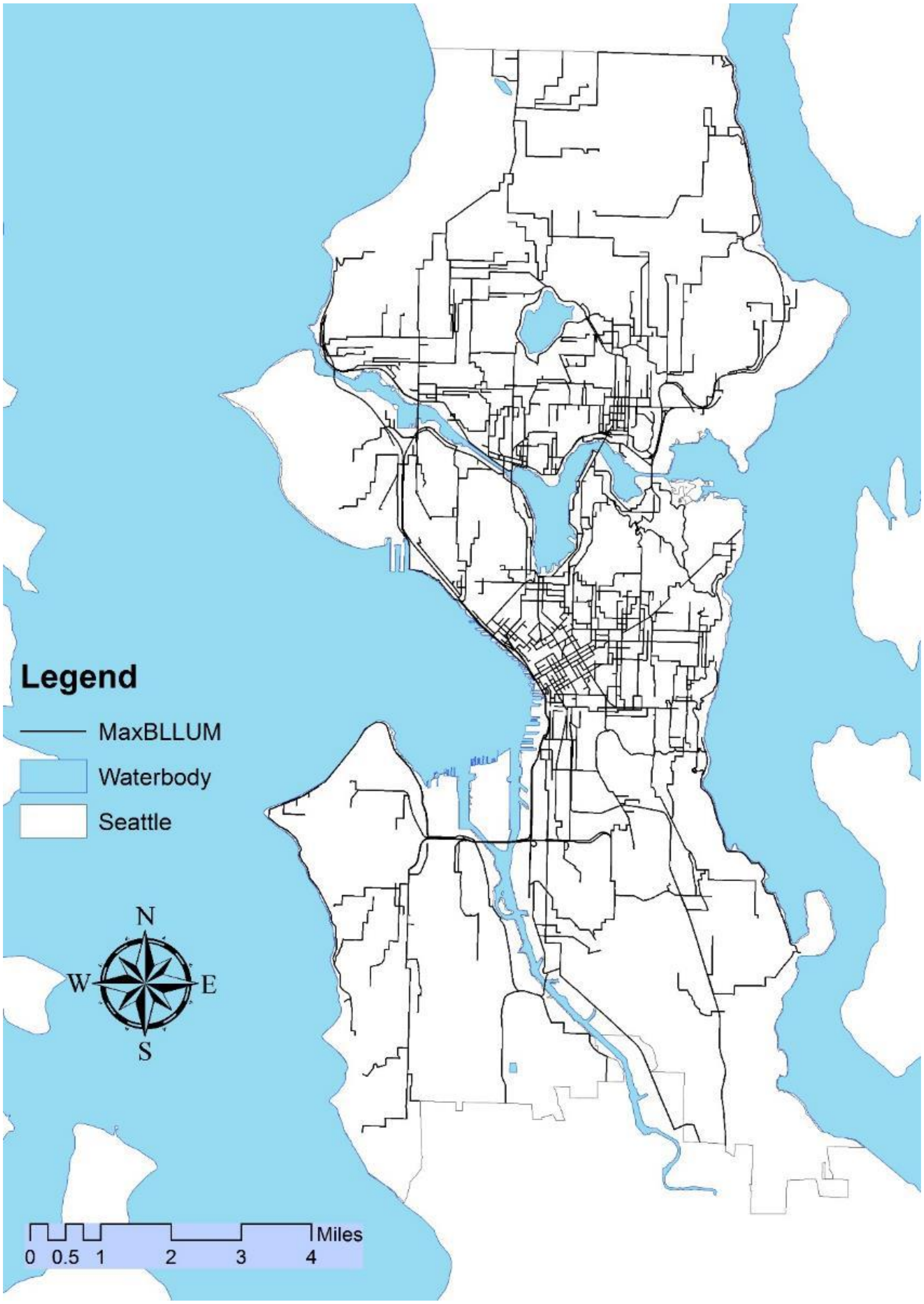


Figure A-8: Alternative routes labeled by minimizing land use mixture

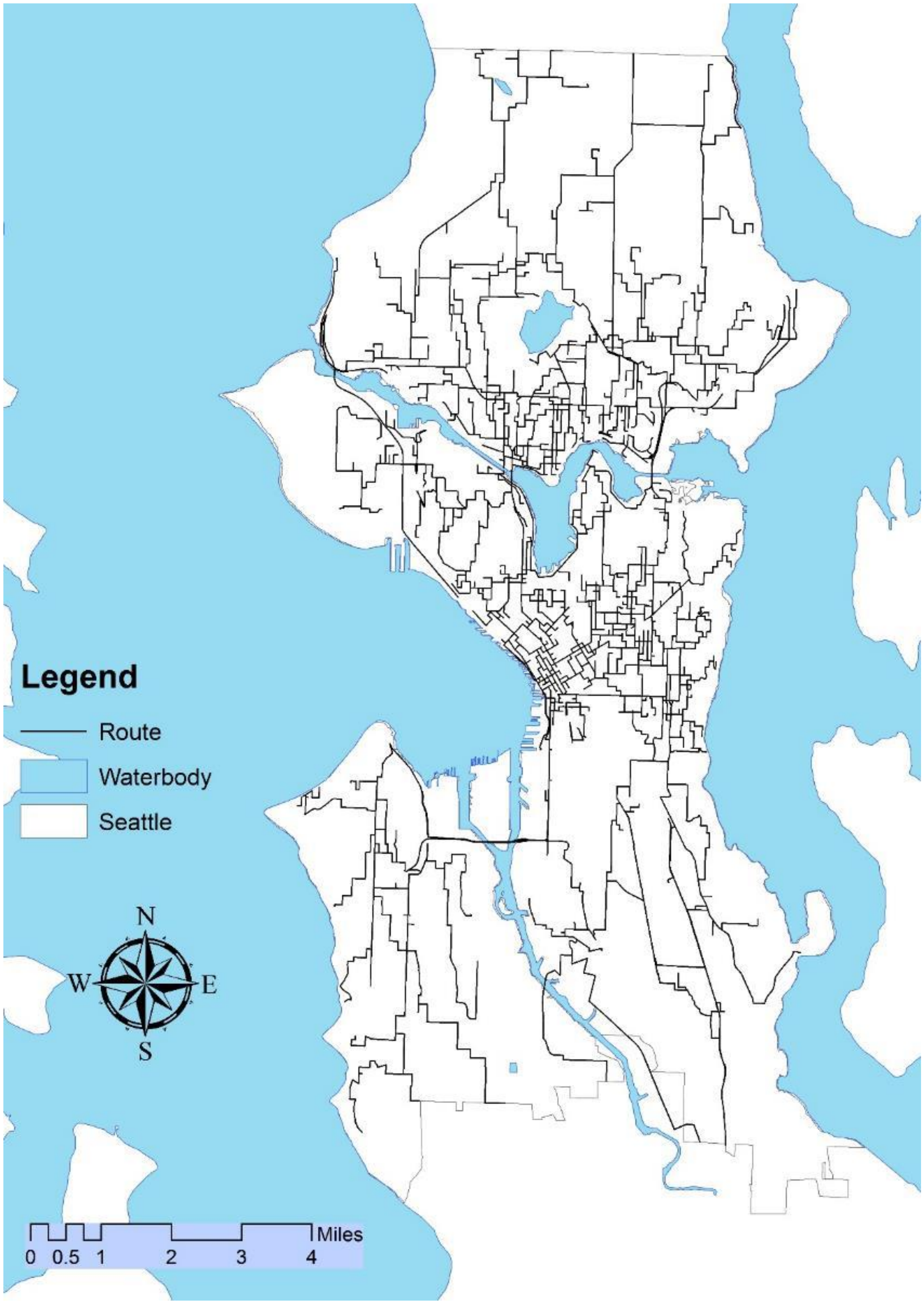


Figure A-9: Alternative routes labeled by maximizing street trees

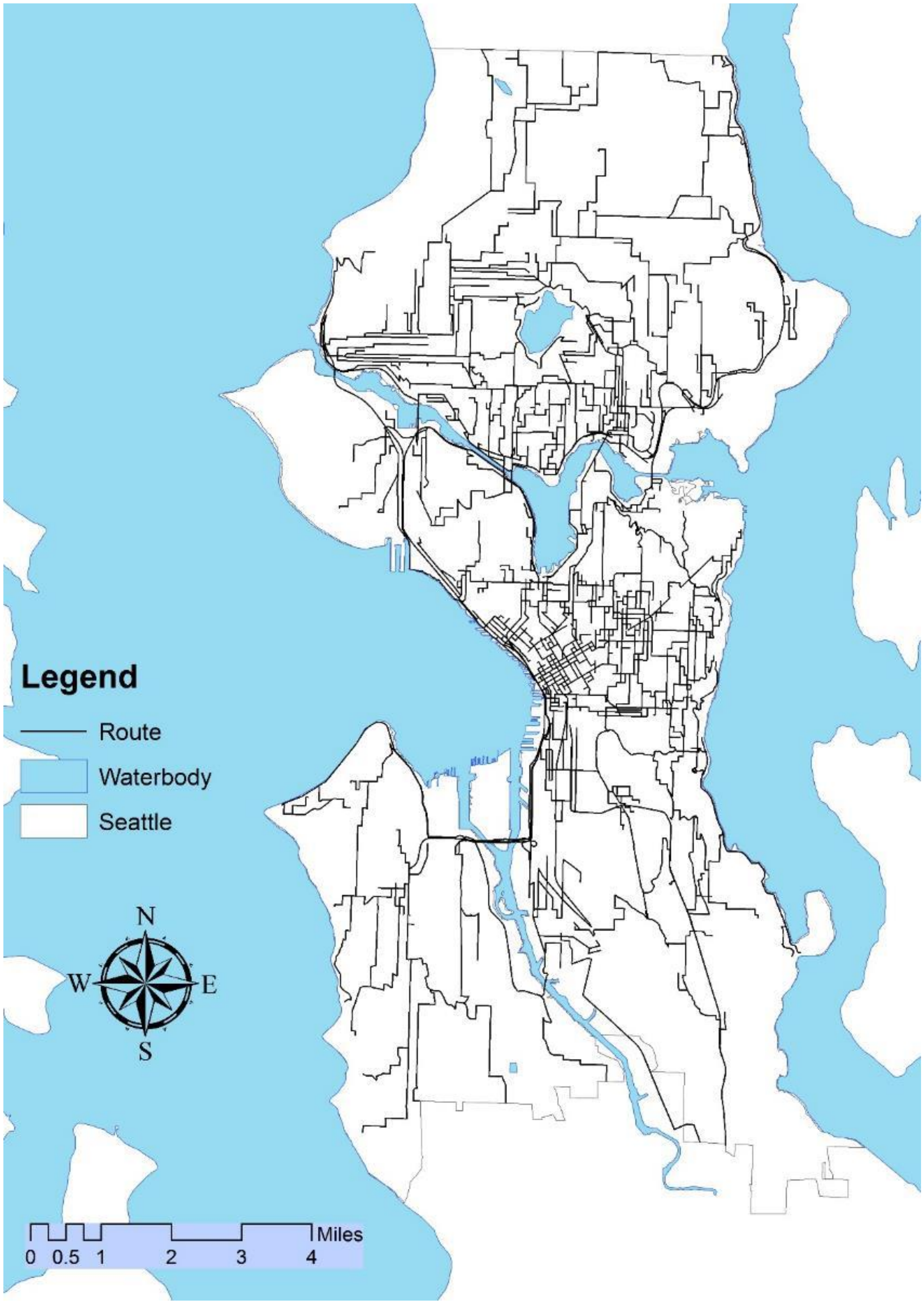


Figure A-10: Alternative routes labeled by minimizing the number of intersections

Figure A-11: Alternative routes labeled by maximizing prioritized bicycle facilities

Appendix B. Road network and design, and land use data

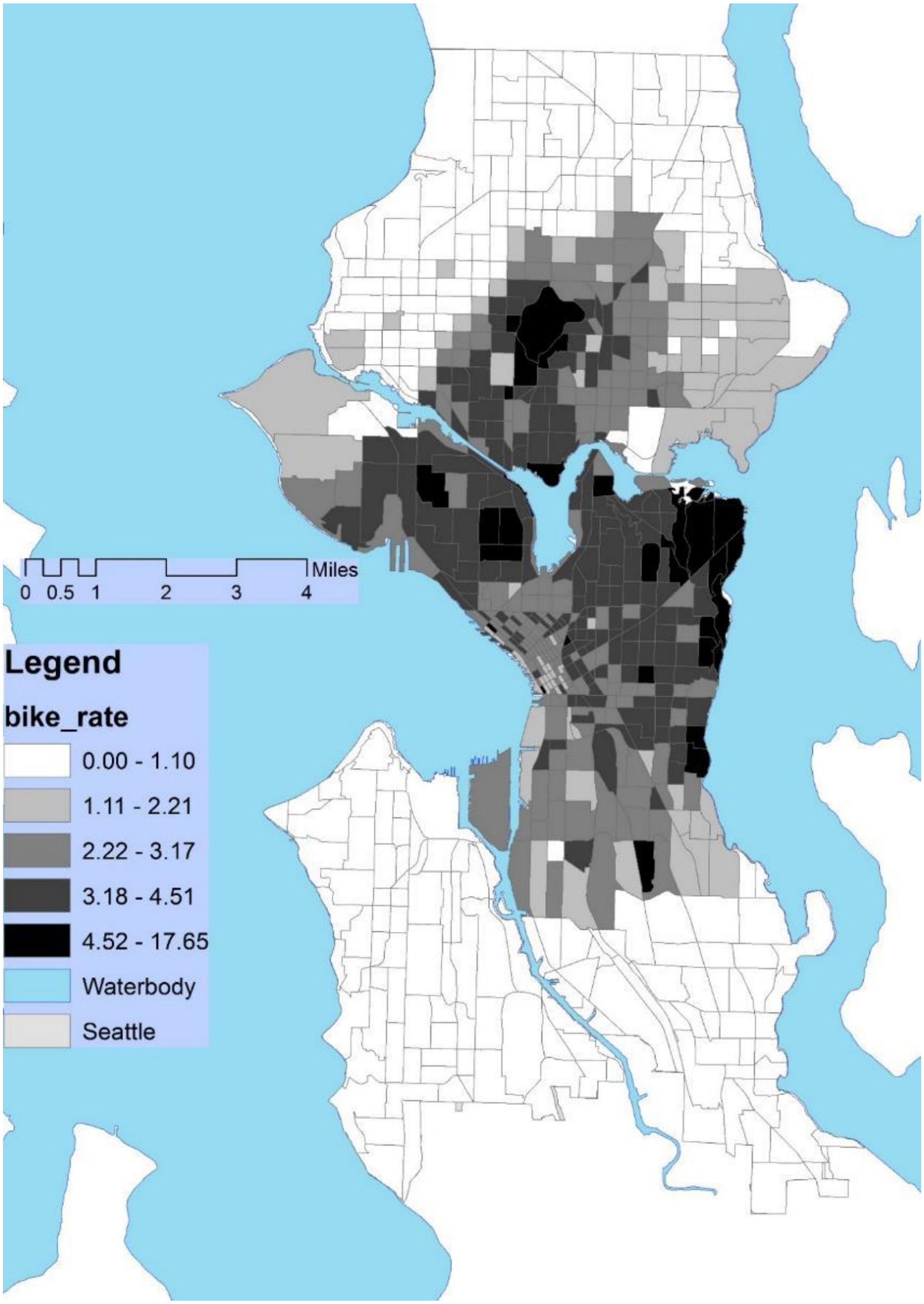


Figure B-1: Bicycling rate of each traffic analysis zone in Seattle

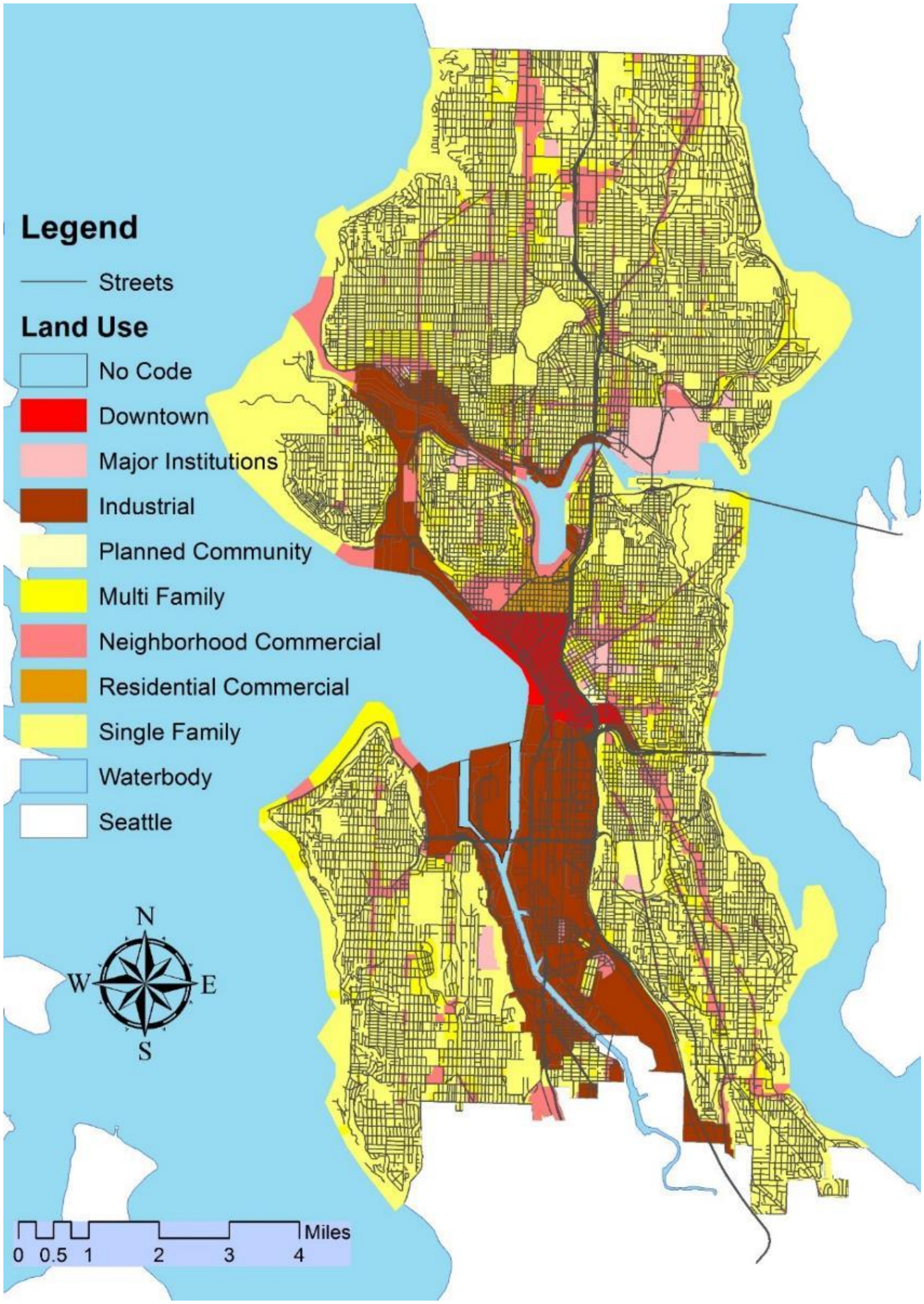


Figure B-2: Land use in Seattle *(Zoning)*

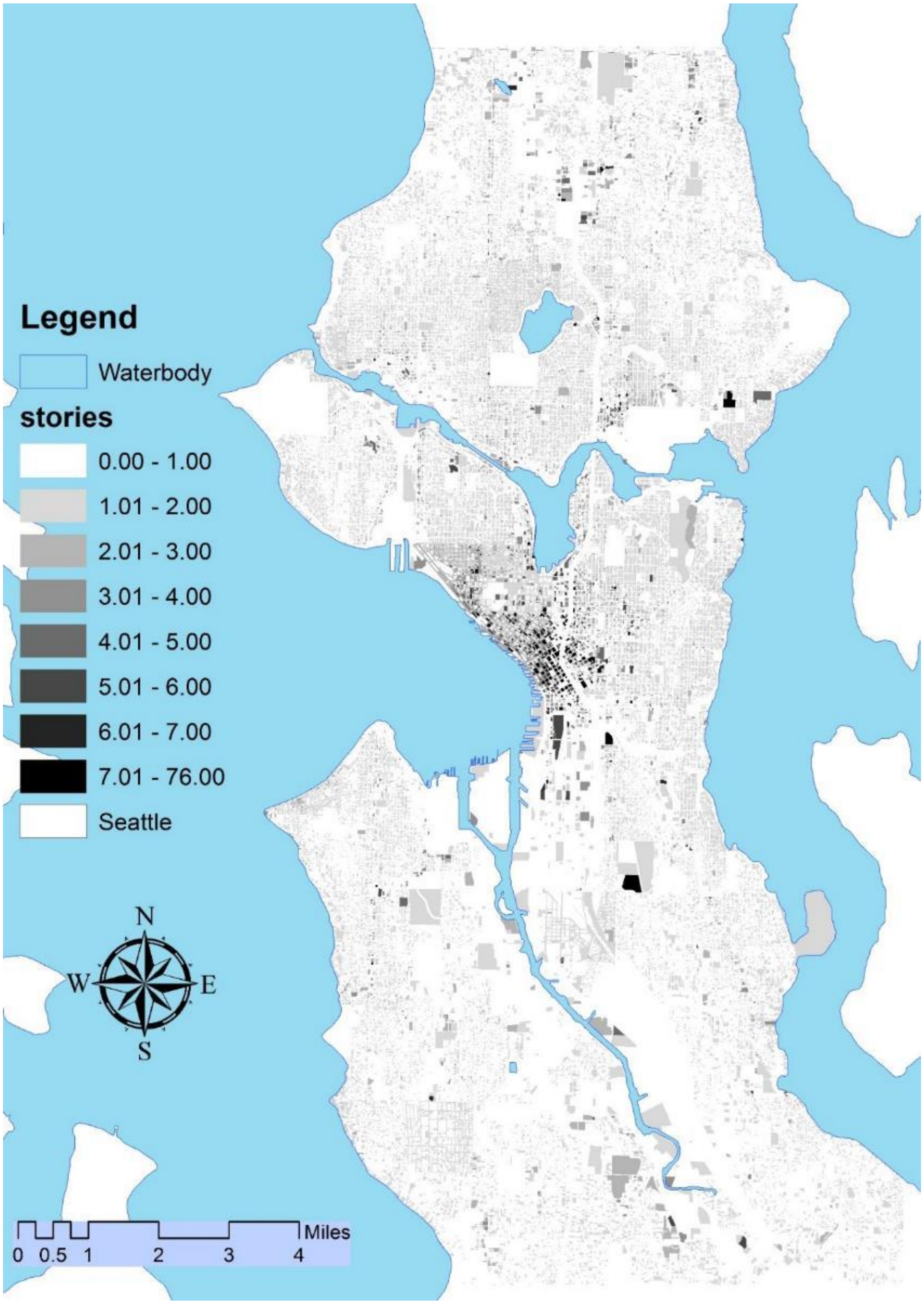


Figure B-3: Floor area ratio in Seattle *(number of stories)*

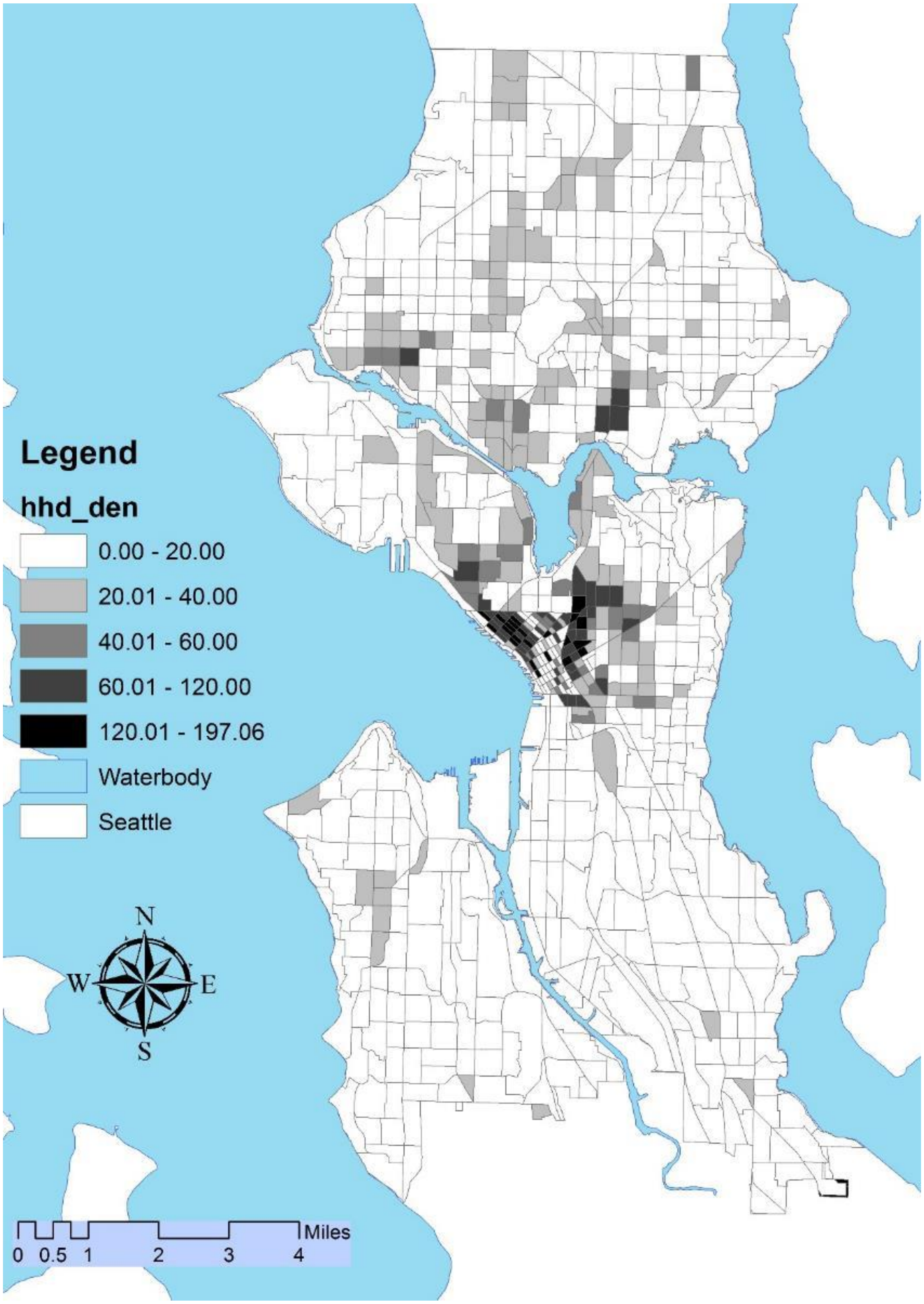


Figure B-4: Household density in Seattle

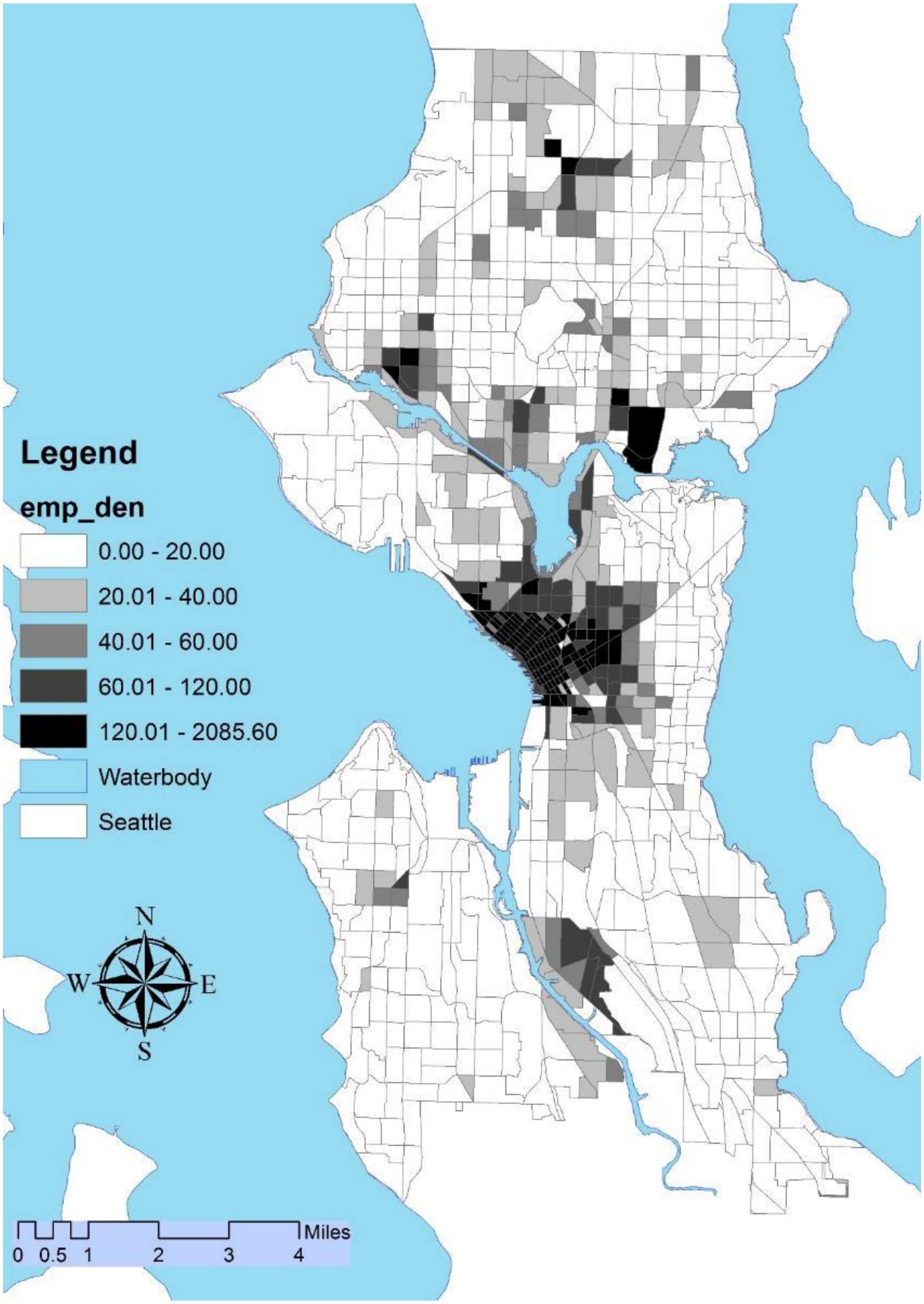


Figure B-5: Employment density in Seattle

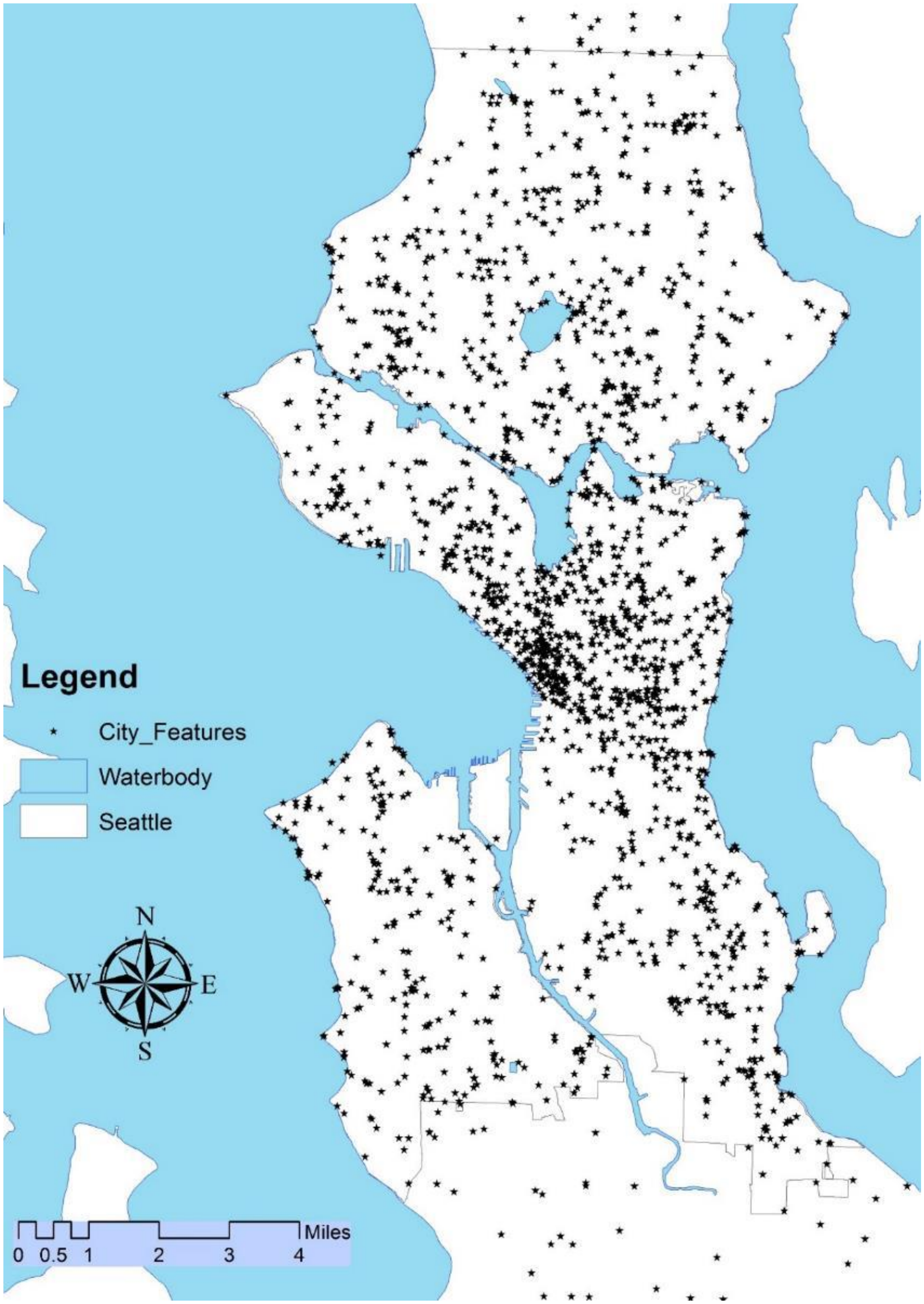


Figure B-6: Public city features in Seattle
(Churches, parks, libraries, art centers, play grounds, schools, theatres, etc.)

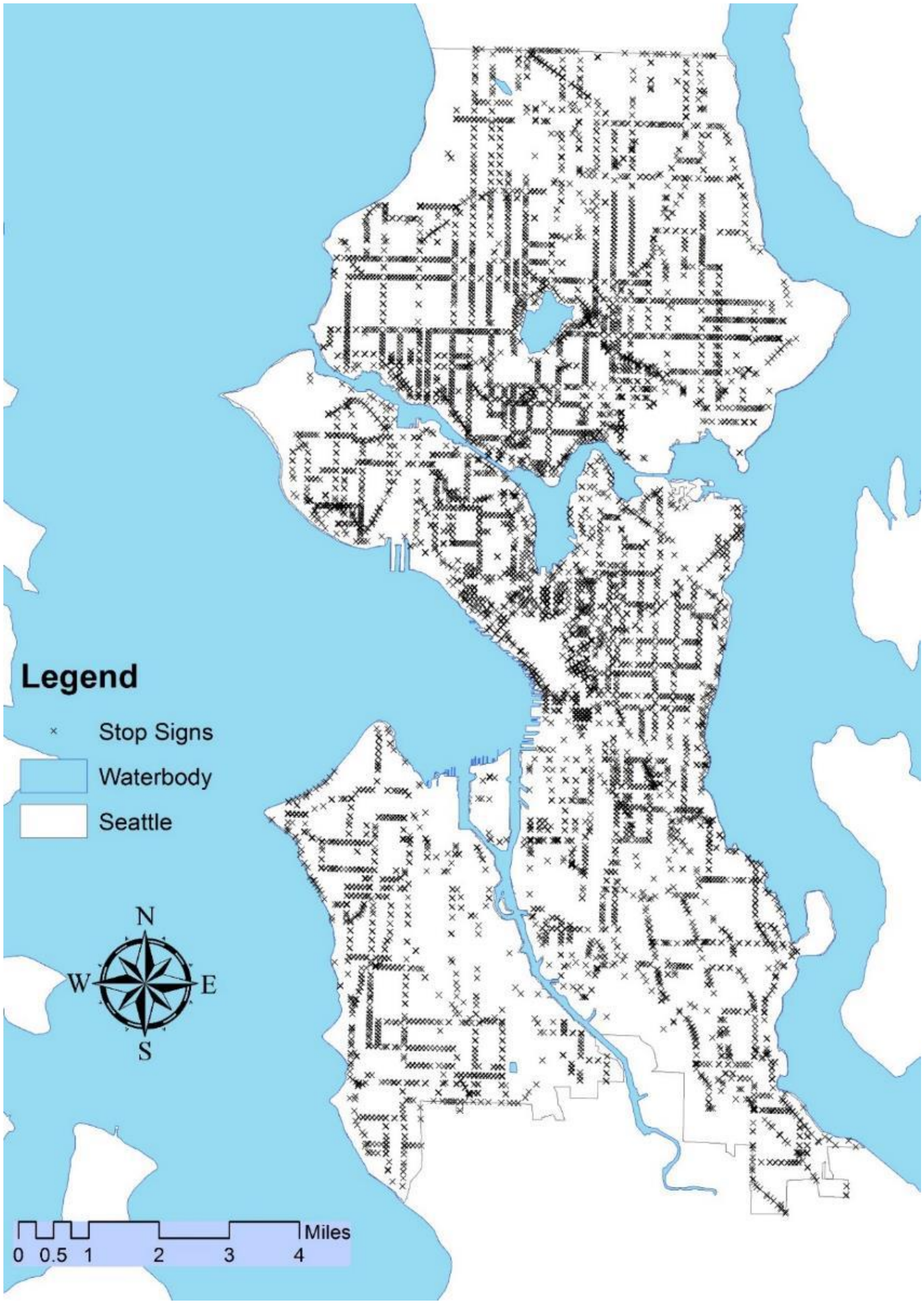


Figure B-7: Stop signs in Seattle

Figure B-8: Traffic signals in Seattle

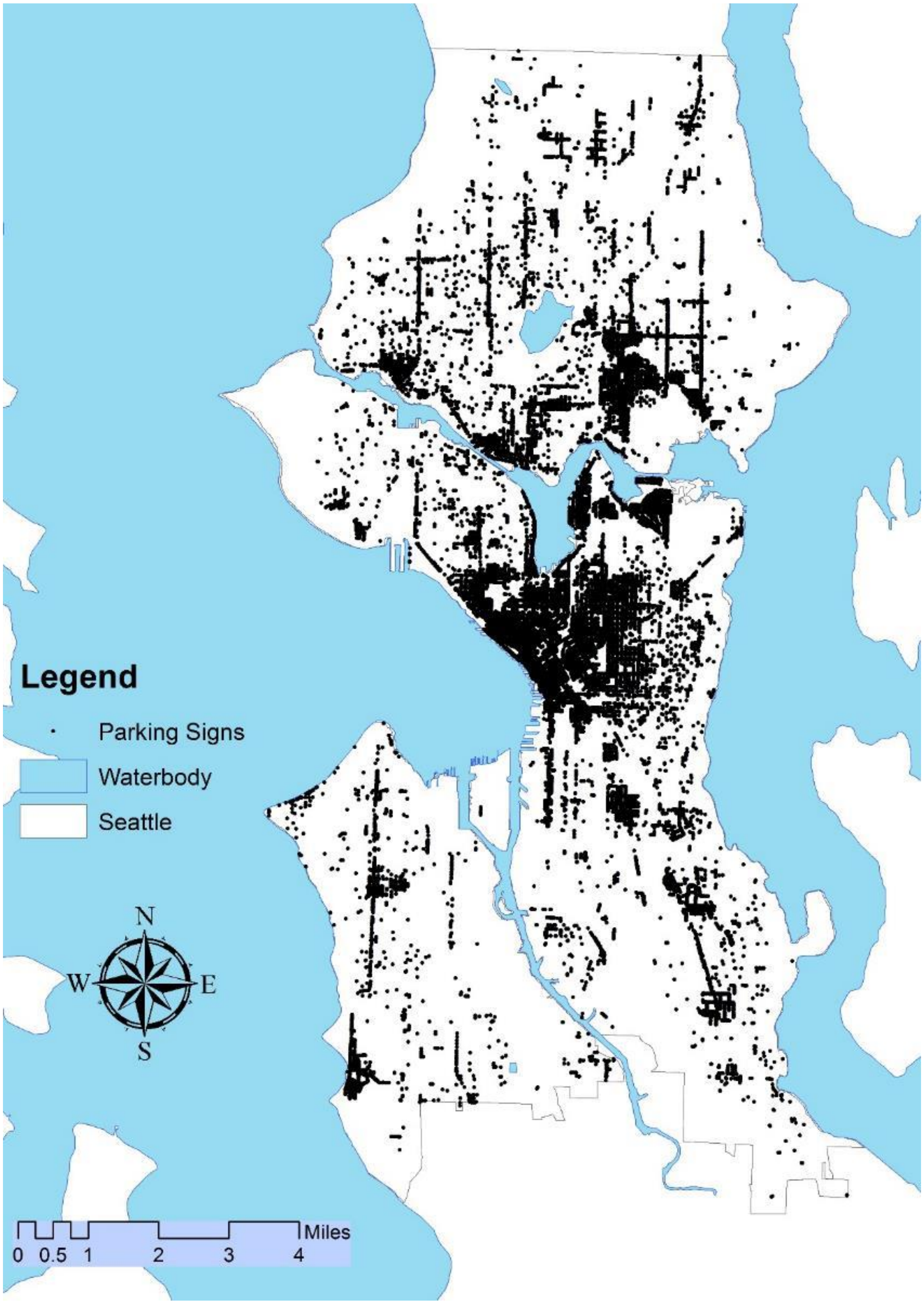
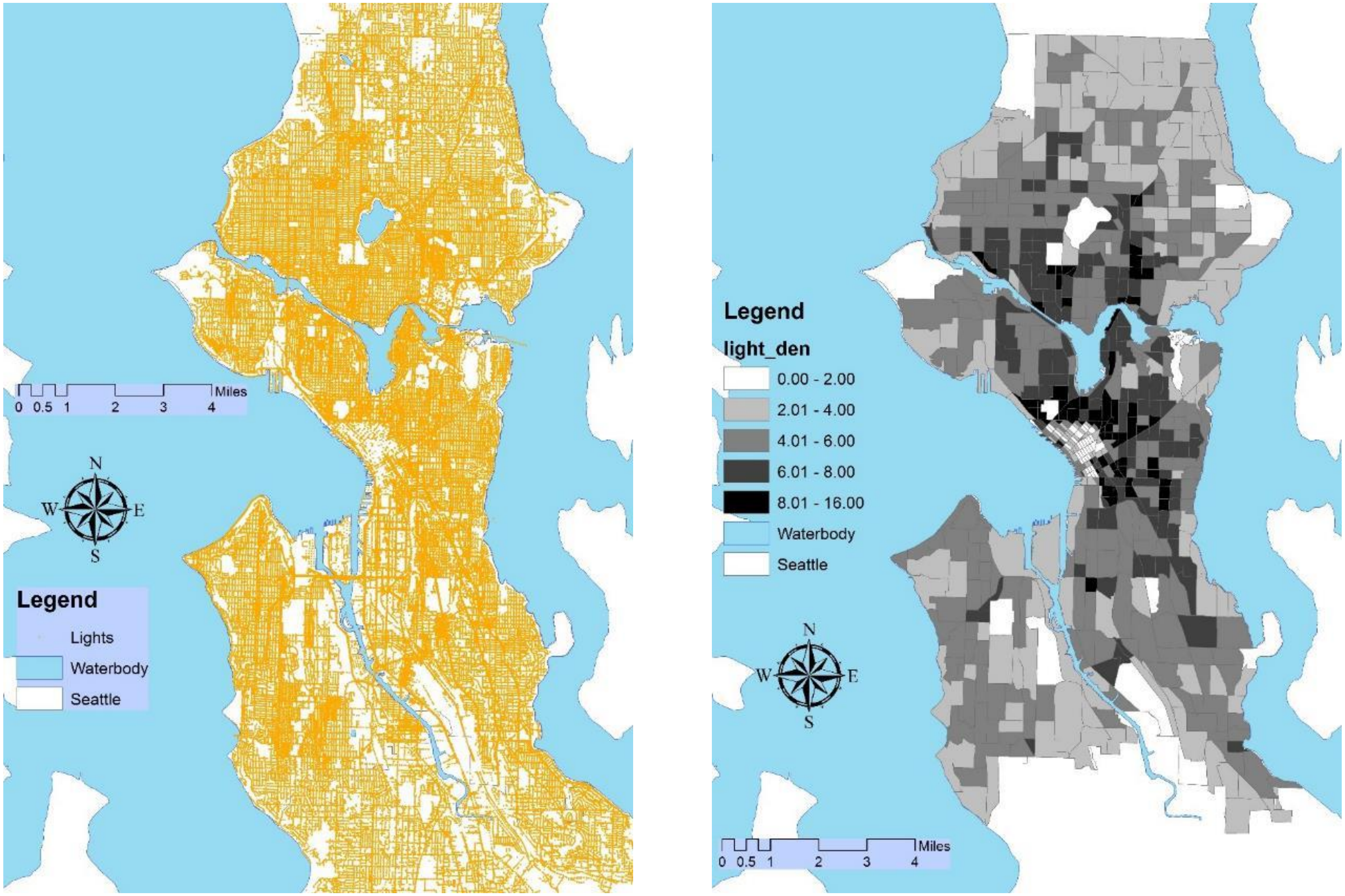


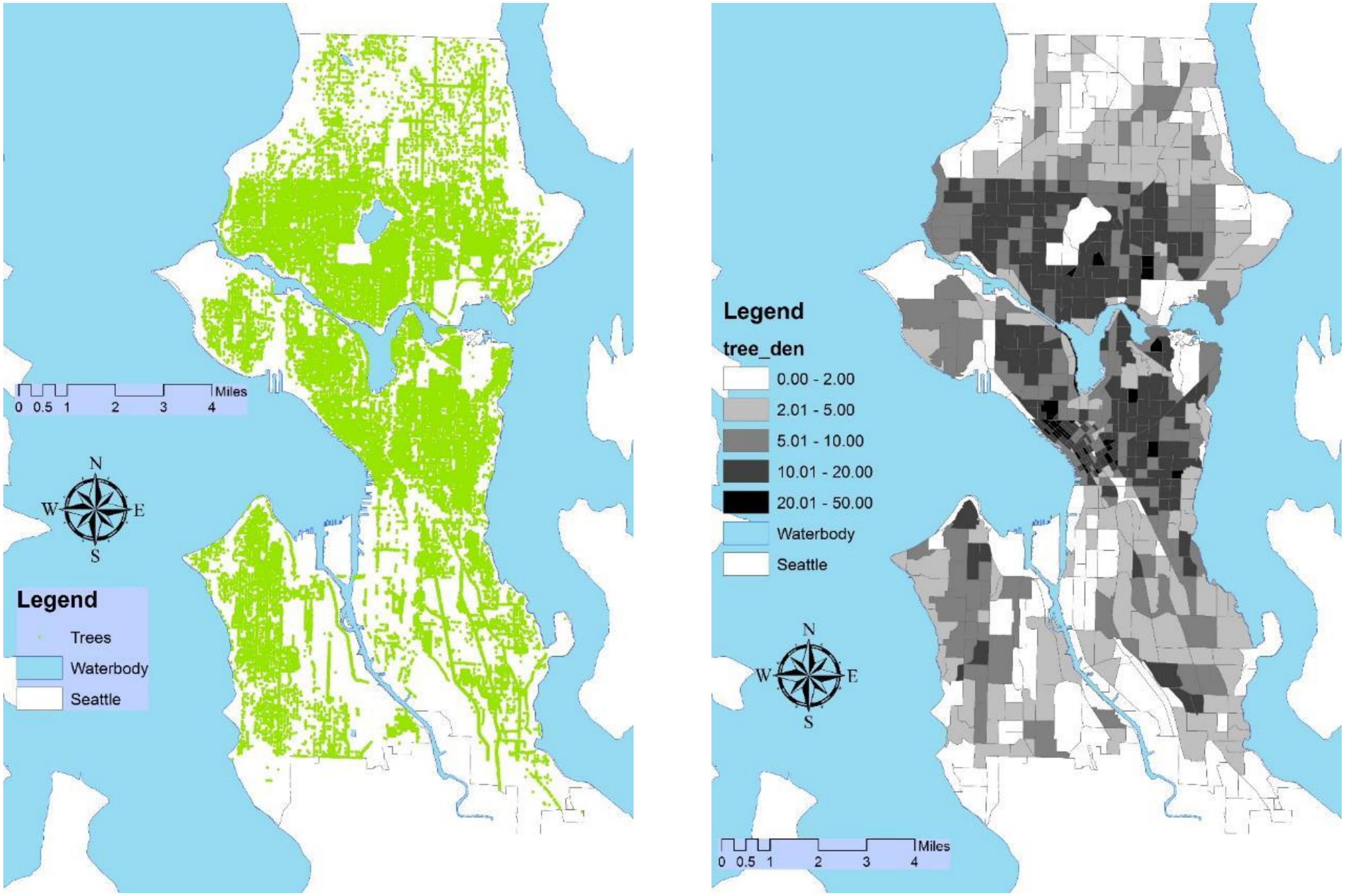
Figure B-9: Street parking signs in Seattle

Figure B-10: Transit routes and bus stops in Seattle



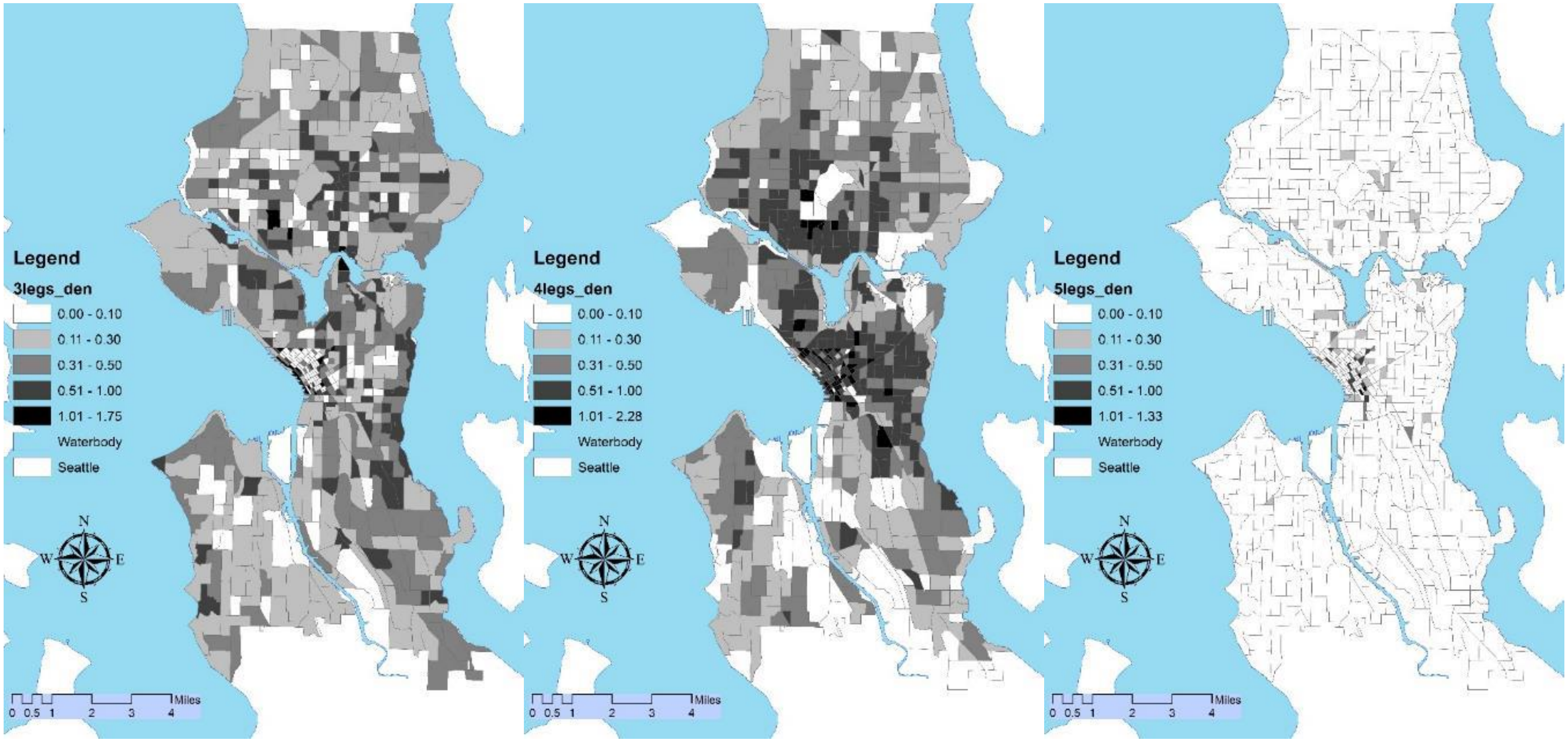
(a) (b)

Figure B-11: Street lights in Seattle (a) and zonal street light density *(numb/ha)* (b)



(a) (b)

Figure B-12: Street trees in Seattle (a) and zonal street tree density *(numb/ha)* (b)



(a) (b) (c)

Figure B-13: 3-way, 4-way, and more than 4-way intersection density in Seattle *(numb/ha)*