

# An Investigation of People's Attitudes about Bicycling in Urban Areas in the US: An Exploratory Case Study in Washington DC

A thesis submitted by

YAQI SHEN

In partial fulfillment of the requirements of the degree of Master of Arts

in

Urban and Environmental policy and planning

Tufts University

May 2016

Copyright 2016, YAQI SHEN

Advisor: Justin Hollander, Reader: Sumeeta Srinivasan

## **Abstract**

Biking as the travel mode has become more and more common and popular recently. However, some problems occurred in development of cycling. This thesis explores the use of micro-blog data in the form of sentiment analysis and statistical analysis to determine if there are relationships existed between people's attitude, bicycling index, and locations where people talking about bicycling. Furthermore, this thesis lays the groundwork for a deeper understanding of bikeability by making quantitative analysis. My results demonstrate that there is relationship existed between peoples' attitude and few bicycling facilities and physical environment factors. Furthermore, some correlation results between each independent variables and sentiment scores indicated. I also provide suggestions about some good strategies of developing cycling for bicycling planners and policy makers by using the results indicated in this study.

## **Acknowledgements**

I would first like to thank my Thesis Advisor, Justin Hollander for patience, support and always being willing to give me suggestions of my research framework and questions throughout the whole process. I would also like to thank my Reader, Sumeeta Srinivasan, for teaching me how to use ArcMap, giving me suggestion about spacial analysis, and provide comments into my research. I would especially like to thank my parents, Hui Shen and Yan Zhang, for funding me through graduate school, for always being supportive and positive, for encouraging me anytime.

## TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
ABSTRACT.....	i
ACKNOWLEDGEMENTS .....	ii
TABLE OF CONTENTS .....	iii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
CHAPTERS	
CHAPTER 1 – Introduction.....	1
CHAPTER 2 – Literature Review .....	4
CHAPTER 3 – Overview of Current Bikeability in Washington DC.....	14
CHAPTER 4 – Methodology.....	28
CHAPTER 5 – Results and Analyses.....	38
CHAPTER 6 –Conclusions, Limitations, and Recommendations for Future Research .....	59
REFERENCES .....	66
APPENDICES	
Appendix I .....	71
Appendix II .....	73
Appendix III.....	83
Appendix IV .....	89

## List of Tables

Table 1: Steps in producing score maps.....	17
Table 2: Methods of creating distance/value score variables.....	35
Table 3: Sentiment analysis results: Sentiment analysis results: Random sample Tweets vs. bike-focus sample Tweets.....	39
Table 4: Sentiment analysis results: Sentiment scores and distance from bike trails.....	41
Table 5: Correlation between distance to bike trail and sentiment scores .....	42
Table 6: Sentiment analysis: Sentiment scores and distance from Capital Bikeshare locations.....	44
Table 7: Correlation between distance to Capital Bikeshare locations and sentiment scores.....	45
Table 8: Sentiment analysis: Sentiment scores and distance to signed bike routes...47	
Table 9 Correlation between distances to signed bike routes and sentiment score...48	
Table 10: Sentiment analysis results: Sentiment and intersection density scores.....	50
Table 11: Correlation between Intersection density scores and sentiment scores.....	51
Table 12: Sentiment analysis: Sentiment and hilliness scores.....	53
Table 13: Correlation between hilliness scores and sentiment scores.....	54
Table 14: Sentiment analysis: Sentiment and destination density scores.....	56
Table 15: Correlation between destination density scores and sentiment scores.....	57
Table 16: Results of linear regression between independent and dependent variables.....	54

## List of Figures

Figure 1: Conceptual model.....	10
Figure 2: Destination density scores.....	18
Figure 3: Intersection density scores.....	19
Figure 4: Hilliness scores.....	20
Figure 5: Signed bike routes scores.....	21
Figure 6: Capital bike share location scores.....	22
Figure 7: Bike trail scores.....	23
Figure 8: Biking scores.....	24
Figure 9: Biking Scores with Census Tract Overlay.....	26
Figure 10: Street Map with Census Tract Overlay.....	27
Figure 11: Sample tweets from Washington DC.....	29
Figure 12: Geolocation of Tweets in Washington DC (Random sample and Bike-focus sample of Tweets).....	30
Figure 13: Example of bike-focus sample Tweets from Washington DC.....	31
Figure 14: Example of random sample Tweets from Washington DC.....	31
Figure 15: Example of the sentiment analysis for a single Tweet.....	33
Figure 16: Example of sentiment analysis for Tweets after modification.....	33
Figure 17: Example of Emoji.....	34
Figure 18: Example of creating buffers with line shapefile (signed bike routes).....	36
Figure 19: Example of creating buffers with points shapefile (Capital Bikeshare locations).....	36
Figure 20: Distribution of geolocated Tweets by bike trails.....	40
Figure 21: Distribution of geolocated Tweets by capital bike share locations.....	43
Figure 22: Distribution of geolocated Tweets by signed bike routes.....	46
Figure 23: Distribution of geolocated tweets by intersection density.....	49
Figure 24: Distribution of geolocated Tweets by hilliness.....	52
Figure 25: Distribution of geolocated tweets by destination density.....	55
Figure 26: Example of Bike-focus sample tweets from one user.....	62

## **Chapter 1: Introduction**

Recently, non-motorized modes of travel, particularly bicycling, have become more common and popular. Bicycle use increased 39 percent nationally from 2001 to 2011, because riding a bike saves on the cost of living in urban areas, and also, bikes are available readily, and are convenient to ride anywhere and anytime compared to public transportation (Ben, 2011). Cities such as San Francisco, Chicago, and Washington DC have seen huge increases in cycling (Ben, 2011); however, cycling plans in urban areas have some problems and the development of such plans has not kept pace with the massive rise in cyclists.

Through a case study in Washington, DC, this thesis evaluated the ways in which bicycling facilities and city cycling plans influence people's attitudes and behaviors. The study includes a discussion of the Bicycling Score in Washington, an analysis of people's attitudes from Twitter data collected from March 3 to April 10, 2015, and an analysis of the relationship between cyclists' attitudes, cycling facilities, and city bike plans. Finally, the thesis provides a detailed discussion and evaluation of the current research on cyclists' attitudes about bicycling.

The following research questions were explored:

1. How is bicycling ability represented in Washington DC?
  - What is the Bicycling Score in Washington, DC, and how is the index

used?

- What factors enter into the index, such as cycling infrastructures, hilliness, desirable amenities, and road connectivity?
2. What are people's attitudes (positive or negative) about bicycling in Washington?
  3. Is there a relationship between people's attitudes, the bicycling index (cycling infrastructures, hilliness, desirable amenities, and road connectivity), and locations where people talk about bicycling?

The thesis also discusses the implications of these research questions with respect to bicyclists, bicycling facilities, and bicycling planning within the Washington DC area. I will also draw conclusions from the Washington data for application to the rest of the US, and provide some useful guidelines for future planning about bicycling based on human and social considerations.

Twitter data collected from March 3 to April 10, 2015 provide the most recent data, and thus represent people's attitudes well. Current attitudes about bicycling contributed to the analysis of the research questions and identified problems better.

To answer the research questions above, the following software was used: ArcMap, SentiStrength, and STATA. ArcMap was used to produce maps of current

bikability in Washington DC, and helped calculate the bicycling score there as well. In addition, the data created by ArcMap contributed to other statistical analyses. SentiStrength was used to estimate the sentiment score for each Twitter post, while STATA provided the ability to conduct statistical analyses of the sentiment scores and other biking data calculated by ArcMap. The methods with which these software programs were used is described in detail in the next chapters.

This thesis is presented as follows. Chapter 2 includes a literature review that provides a description of the benefits of bicycling and its status in the US at present. The chapter also discusses several case studies of people's attitudes about bicycling in urban areas and the ways in which bikeability affects those attitudes. Chapter 3 offers an overview of current bikeability in Washington DC, and includes maps to make the results more straightforward. The methods of determining factors, collecting data, and producing maps also are detailed and described specifically in Chapter 3. Chapter 4 presents the methodologies employed in the thesis for data collection, determination of sentiment scores, and the relationship between sentiment scores and factors associated with biking. Chapter 6 describes the conclusions and limitation of the study, and offers ideas for future relevant studies.

## **Chapter 2: Literature Review**

### **Background**

Bicycling has become a popular topic of discussion on social media today. Many transportation institutes and government agencies have realized that non-motorized travel can benefit communities in many ways, and have determined that current cycling plans are problematic and provide considerable scope for improvement (VTPI, 2015; Grabow, 2013). More than 90% of commuting trips in the US are made in private motor vehicles, and the balance of commuting trips are made on foot, and by bicycle and public transit (NHTS, 2009).

A strong movement has developed over the past decade to increase bicycling in order to improve health through greater activity, reduce vehicle miles travelled, and improve air quality, among other benefits (Gotschi, 2011; Krizek 2007; Dill & Carr, 2003). Although bicycling can make significant contributions to social goals related to public health, energy independence, climate change, air quality, traffic congestion, mobility, and economy, only mass cycling can reach those goals (Furth, 2012).

This literature review includes a brief description of the benefits of bicycling, the status of bicycling in the US today, and findings about the relationship between people's attitudes, bicycling, and other urban issues.

## **Benefits of Bicycling: Health, Social, Environmental**

The bicycle plays an important role in the transportation system today. First, it is a low cost, non-polluting transportation option that makes efficient use of limited roadway capacity. Further, for those individuals who do not have a vehicle, the bicycle is an effective means of transport, particularly for trips that are too long for walking or are not served by quality public transit (Murphy& Knoblauch, 2004).

Another major benefit is that bicycling contributes to public health. Physical activities reduce obesity, which has reached epidemic proportions in the nation, and bicycling is an excellent form of physical exercise (Killingsworth, 2003). Many case studies have shown that the health benefits of bicycling far exceed the risk of traffic injuries; moreover, as levels of bicycling increase, the rate of injuries decreases (Elvik, 2009; Jacobsen, 2003; Robinson, 2005).

Bicycling also is an efficient way to reduce vehicular pollution, which is a major contributor to greenhouse gas emissions; based on a 2013 EPA report on total US greenhouse gas emissions, 27% of such emissions result from vehicles. Thus, non-motorized transportation, including bicycling, can help alleviate this problem.

In conclusion, bicycling offers numerous benefits. It can ameliorate certain environmental issues, and as a means of transport and recreation, is beneficial to people's health. Bicycling also is an efficient transportation alternative to driving or

walking.

### **Status of Bicycling in the US**

At present, the bicycle is not a common mode of daily transportation in the US. Only one percent of all one-way trips in 2009 were made by bicycle (NHTS, 2009). However, increasing evidence has confirmed that bicycling should be considered a viable transportation option. First, it is easy to engage in bicycling, as it requires less exertion than does other physical modes of transportation (Frank, Engelke, & Schmid, 2003). In addition, bicycling can be a relaxing and enjoyable activity that helps people relieve stress and increase work productivity. However, based on the current situation, there are several barriers to bicycling in the US, primarily the fact that bicycle facilities remain rudimentary, and the implementation of bicycling plans requires an investment of financial resources (Grabow, 2013).

The greatest obstacle to bicycle planning is the lack of “separation criteria” for cyclists in the US. Currently, there are no criteria related to when cyclists should be separated from fast or heavy traffic (American Association of State Highway Transportation Officials, 1999). Further, there are no limits to the speed or number of lanes on roads that incorporate bike lanes, or even those designated as “shared lanes” (Furth, 2012). A primary goal of any bicycle infrastructure program must be to provide sufficient separation from traffic, such that it attracts the mainstream

population (Furth, 2012). Some shortcomings and strengths of planning are obvious in urban areas in the US; in the next few chapters, I will discuss using social media data to make future suggestions for bicycle planning.

Currently, because of the increasing confirmed benefits associated with bicycling, more US municipalities are beginning to promote and develop policies and plans to make bicycling a safer and easier form of transport. Transportation planners have begun to realize that public participation in planning bicycle projects is an important part of the process, specifically with respect to the demand for bicycle facilities, network planning, suitability modeling, and other related issues (Molina, 2014). In the next few chapters, I will argue that data collected from Twitter can be an effective method of public participation. Using social networks is helpful in tracking people's attitudes on many objects, and acquiring feedback and suggestions. An analysis of Twitter data specifically could contribute to future bicycle planning.

### **What are People's Attitudes about Bicycling in Urban Areas? Does the Ability to Bicycle Affect their Attitudes?**

#### **Empirical study of people's travel attitudes and urban design features**

There is a consensus within planning and urban design policy that the design of a sustainable urban environment would encourage people to reduce the use of automobiles and choose more sustainable modes of travel. However, based on one

case study in the UK, sustainable features of urban design did not cause people to change their behavior, although such features may change their attitudes about walking and bicycling in the long-term (Susilo, Williams, Morag, & Dair, 2012). Some sustainable features of urban design relevant to people's attitudes and behaviors, including secured bike storage, high connectivity of neighborhoods to nearby areas, natural surveillance, high quality public realm, and issues of heavy traffic will be discussed.

### **Facts from Survey of People's Attitudes and Behaviors about Bicycling**

Data from the 2012 National Survey of Bicyclist and Pedestrian Attitudes and Behavior are worth mentioning: 46% of 7,509 respondents have bicycles paths available within a quarter mile of where they live, and 39% have bicycle lanes available on roads within a quarter mile of where they live (NHTSA, 2012). Among the 1,350 respondents who do not have bicycle paths nearby, 40% never ride on bike paths. Further, among the 1,176 people who do have bicycle paths nearby, only 12% of them always ride on bicycle paths, and another 11% never use these paths (NHTSA, 2012). Moreover, the same report indicated that 38% of 1,551 respondents ride a bike more often by comparison to a year ago, while 20% ride a bike less often than in 2011 (NHTSA, 2012). These facts and case studies show that, although people's bicycling behaviors are changing, there still is no clear pattern in their

attitudes about bicycling as it relates to the ability to bike in urban areas.

### **Conceptual Model from Previous Case Studies in the US**

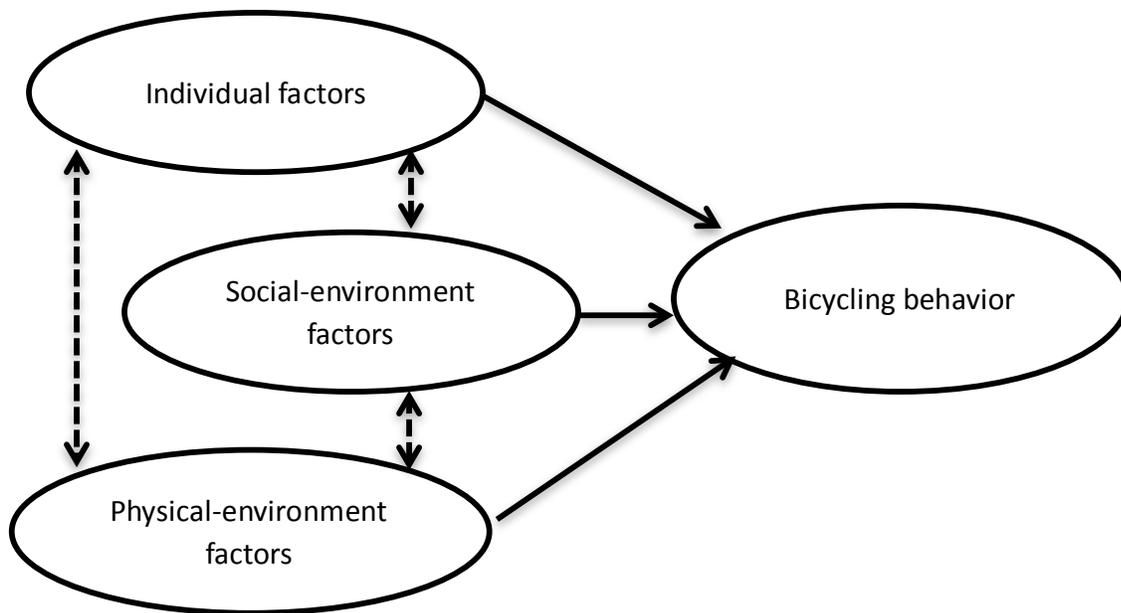
There is a conceptual framework that is used widely in research on physical activity, specifically in the field of public health (Sallis & Owen, 2008). This framework is composed of three factors—individual, social-environmental, and physical-environmental—that are used to explain people’s behavior and attitudes (Handy, Xing, & Buehler, 2010). Individual factors (e.g., attitudes) about the physical-environmental factors in transportation infrastructures and land use patterns both offer strong support for the arguments in the next sections in this thesis.

Figure 1 shows that these three factors are hypothesized to affect bicycling behaviors directly (Handy et al., 2010). Individual factors contribute to the motivation to bicycle, and social and physical environments are key factors in determining the quality of bicycling conditions that may enable and encourage people to bicycle (Handy 1996, 2009). These factors are interrelated, and all influence bicycling behaviors. For example, bicycle infrastructure, one of the physical-environmental factors, influences an individual’s attitudes and behaviors with respect to bicycling; because bicycle infrastructure affects travel time, safety, enjoyment, and other related bicycling experiences, it has an important influence on whether or not people decide to bicycle. In turn, increasing individual bicycling

would help generate a supportive physical environment.

While individual factors are important in explaining bicycling behaviors (Handy et al., 2010), aspects of the physical environment also are important, including the distance to destinations as determined by land use patterns in transportation bicycling, and the presence of a network of off-street bicycle paths for both transportation and recreational purposes (Handy et al., 2010).

All of these are useful pieces of evidence that directed the research and analysis in the next few chapters.



*Figure 1. Conceptual model (Handy, Xing, & Buehler, 2010)*

## **Approaches to Measure People's Attitudes about Bicycling and Other Urban Issues**

Because people's attitudes about bicycling and other urban issues are complex, an analysis of those attitudes is necessary and helpful in understanding people's perceptions. Moreover, social networking provides good tools for exploring people's attitudes, because of how much they share. I will use data from Twitter, because Twitter has 320 million monthly active users and supports more than 35 languages ([Twitter.com](https://twitter.com), 2015). Thus, this large number of users will provide a sufficient sample size for exploring and analyzing attitudes. In the article "You are what your friends eat," the authors use social media data to track attitudes and explore changes in eating behavior and body weight (Adam, Chris, & Annik, 2011). This is a good example and provides evidence to show that social media are a powerful tool to identify people's attitudes and for research.

People's attitudes may be measured by either qualitative or quantitative methods. Qualitative methods are powerful tools to explore such complex issues because they allow investigation of an individual's own explanations of their behavior and attitudes (Beirao & Cabral, 2007; Clifton & Handy, 2001). Qualitative methods include participant observation, direct observation, unstructured interviews, and case studies.

Quantitative approaches have the advantage of measuring the responses of many subjects to a limited set of questions, and thus allow comparison and statistical aggregation of the data (Patton, 1990).

### **Qualitative Case Study**

One qualitative study addressed travelers' attitudes about transport (Beirao & Cabral, 2007). They used in-depth interviews of 24 individuals who ranged from 18 to 70 years of age and conducted a comparative analysis of the respondents' demographics (e.g., work status, car status, and income level).

### **Quantitative Case Study**

In contrast, a quantitative study by Grabow (2013) presented predictors from urban design and used them to model commuting behavior. The author developed the "active transportation index" (ATI) by using data from surveys and geographic information system (GIS) sources.

A brief description of these studies will be helpful in demonstrating the methodology employed in this thesis. The thesis combined both of these methods and included data collection, quantitative analysis by SPSS, and spatial analysis by GIS.

### **Findings and Conclusions**

Thus far, the literature has shown that the benefits of bicycling cannot be

ignored, but that there are significant shortcomings in urban bicycle planning at present. Data from Twitter will be useful in obtaining feedback from current bicyclists, and suggesting options that will facilitate bicycle planning.

Further, individual, social-environmental, and physical-environmental factors all affect people's attitudes, which provided strong support for the purpose of this thesis, to evaluate how bicycling facilities and city cycling planning influence people's attitudes and behaviors through a case study in Washington, DC.

Moreover, the literature related to the various methodologies used to assess attitudes and urban issues was helpful in the selection of the methods used in this thesis. Ultimately, I concluded that a mixed quantitative and qualitative design would be an ideal way to address the hypotheses and perform the analyses.

## **Chapter 3: Overview of Current Bikeability in Washington DC**

### **Description**

The primary focus of this study was an exploration of the relationship between people's attitudes about biking and bicycling facilities and city cycling planning in Washington DC. In preparation for the remainder of the analyses, I first conducted a bikeability analysis in Washington DC using GIS data.

The data for this analysis were derived from four main resources that contributed to six individual maps. The main goals of this analysis were to obtain an overview of bikeability in Washington DC in a map format that makes the results easy to interpret. Further, this analysis identified the best aspects of six indices separately and represented the highest biking score area as maps as well. Moreover, parts of these maps were used in data collection and the statistical analyses.

The physical environment is an important variable in explaining bicycling behaviors, particularly the distances to destinations as determined by land use patterns for transportation bicycling, and the network of off-street bicycle paths for bicycling for both transportation and recreational purposes (Handy et al., 2010). According to the empirical evidence from the opinion survey, and travel behavior analyses, four main domains combined to identify bikeability: bicycle facilities, street connectivity, topography, and neighborhood land use (Megan, Michael,

Eleanor, & Kay, 2012).

Thus, based on this evidence and the references, I used six physical environment factors in this analysis to evaluate bikeability in Washington DC: bike trails, signed bike routes, Capital Bikeshare locations, hilliness, destination density, and intersection density. This resulted in individual maps for the six factors, after which I calculated an overall score (scale of 1 to 5, with 5 the highest) to identify the best biking areas in Washington.

The reasons for using the Capital Bikeshare locations should be mentioned here. The Capital Bikeshare location is organized by the Capital Bikeshare (CaBi) system, which has operated since 2010 as a public-private partnership with Alta Bicycle Share, and is a bicycle sharing system that provides service in Washington DC, and in some cities in Virginia and Maryland. This system has more than 300 stations and 2,500 bikes, all of which are owned by local governments (Capital Bikeshare, 2016). This system improved the development of bicycle transportation and enhanced people's attitudes about biking as well. Therefore, I used the Capital Bikeshare locations as one of the factors to evaluate bikeability in Washington DC.

### **Data Resources**

In this analysis, four main data resources were used to construct six individual maps: DCGIS open Data ([opendata.dc.gov](http://opendata.dc.gov)), ReferenceUSA, Social

Explorer, and [census.gov](https://www.census.gov). The DC Open Data is a program that belongs to the Office of Open Government (OOG), an independent office under the Board of Ethics and Government Accountability (BEGA) charged with advancing open governance in the District of Columbia. All of the data from GOS Open Data are updated and added to at regular periods, so that the data used in this analysis were all recent. The original data on the distribution of bike trails and signed bike routes, Capital Bikeshare locations, and spot locations were downloaded as shapefiles (.SHP) directly from the DC Open Data website.

The data used to make destination density maps were drawn from ReferenceUSA. The research focused on two groups (retail trade and services) among major industry groups that encompass over ten categories. In these two groups, I selected reasonable destinations within the retail trade, such as grocery stores, meat and fish markets, and fruit and vegetable markets; within the category of services, colleges and universities, libraries, schools, and educational services were selected (Appendix I). After filtering the results, I downloaded the data with longitude and latitude columns, and then processed them in GIS. In addition, other useful data from Social Explorer and [census.gov](https://www.census.gov) were downloaded directly as .CRV files or Shapefiles in preparation for the next steps in the analyses.

## **Primary Methods**

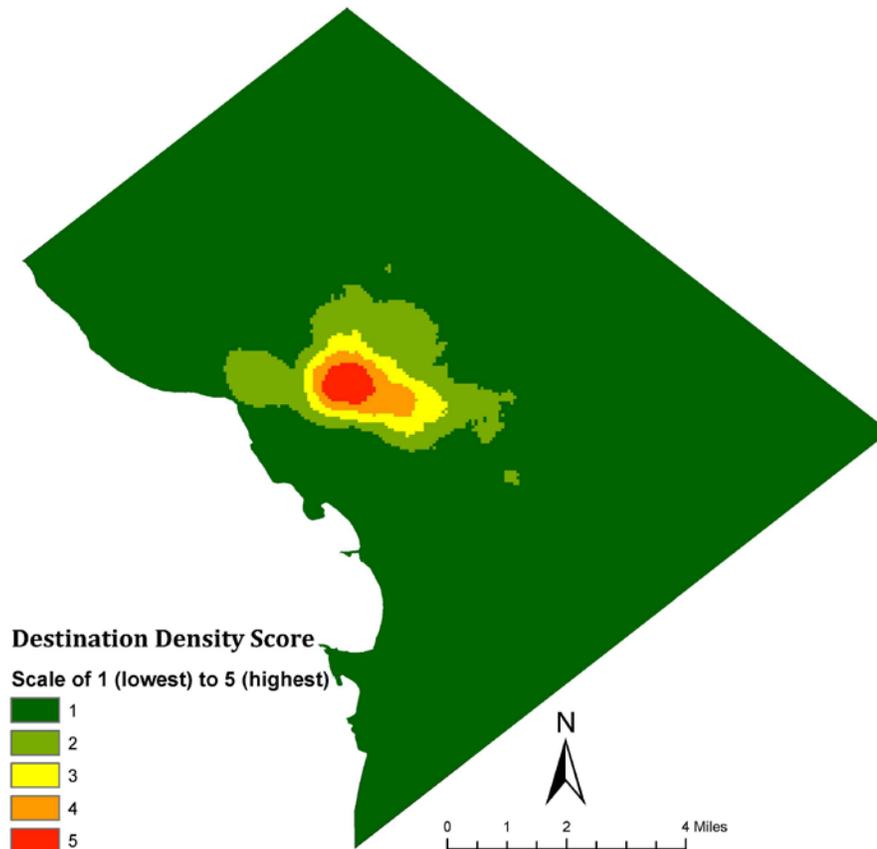
The primary steps used to produce different maps differed slightly, and are summarized in the following table.

Factors	Methods
<b>Destination Density Score</b>	Point data from ReferenceUSA were processed with the “Display XY Data” and “Point Density” tools to rank proximity to existing destinations
<b>Intersection Density Score</b>	Line data from M Drive in GIS Lab operated by Tufts University were calculated with the “intersect” tool, and then the results were filtered with “Selection” when more than two lines intersected (ICOUNT>=3); I then processed the layers created in the previous steps with the “Point Density” tool to rank proximity to existing intersection points.
<b>Hilliness Score</b>	Spot elevation data were processed with the IDW interpolation and Slope tools to rank hilliness.
<b>Signed Bike Routes Score</b>	A line shapefile was processed with the Line Density tool to rank proximity to existing signed bike routes.
<b>Capital Bikeshare Location Score</b>	Point shapefiles of Capital Bikeshare locations were processed with the Point Density tool to rank proximity to existing Capital Bikeshare locations.
<b>Bike Trail Score</b>	A line shapefile of bike trails was processed with the Line Density tool to rank proximity to existing bike trails.
<b>Overall Biking Score</b>	“Raster Calculator” was used to calculate the biking score overall by the six factors described above, which were given equal weights (100/6).

*Table1. Steps in producing score maps*

## Results and Analyses

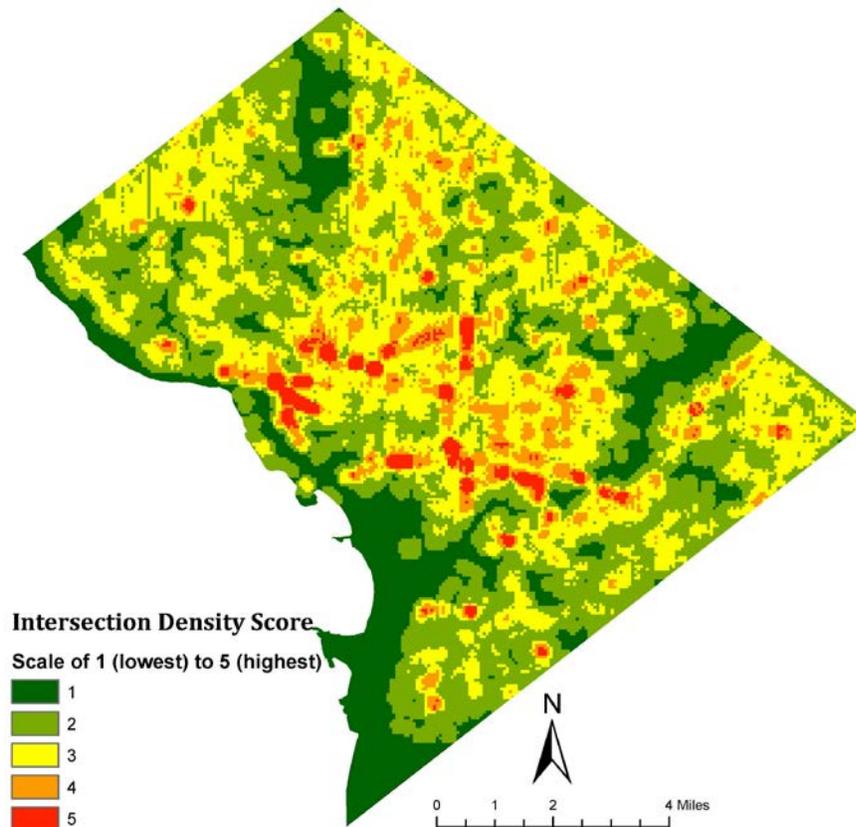
A comparison of Figures 1 to 6 shows that there was no clear pattern among the maps of these six factors. The highest score for destination density was located in central Washington DC, and decreased gradually around that area (see Figure 2). Most of the area in the map is green, which indicates that in Washington DC, most areas have low scores. Thus, from the perspective of destination density, except in the central area, most areas of Washington DC are not suitable for biking.



*Figure 2. Destination density scores*

There also was no clear pattern in the map of intersection density (see Figure

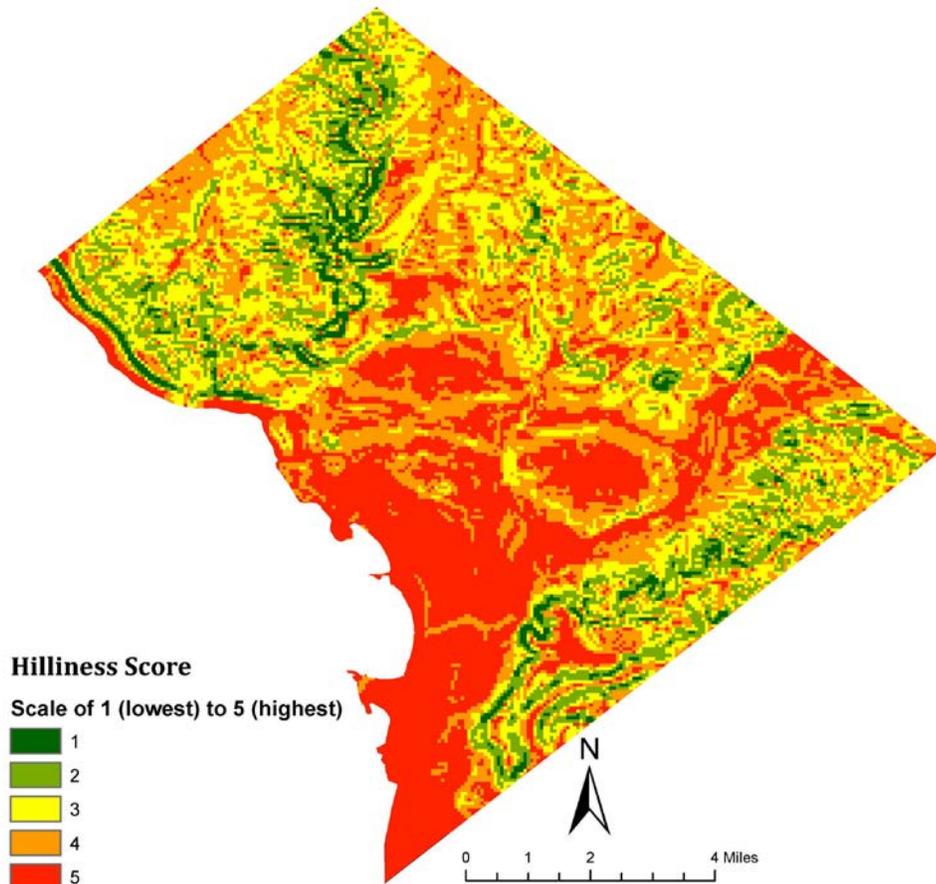
3). The red area on the map had the highest score, and most of the red areas were located in central Washington. In addition to these red areas, one can see some dense orange areas on the map. Most areas on the map are yellow, which indicates that most areas in Washington DC have scores of 3 or higher. From the point of view of this factor, Washington DC is suitable for biking.



*Figure 3. Intersection density scores*

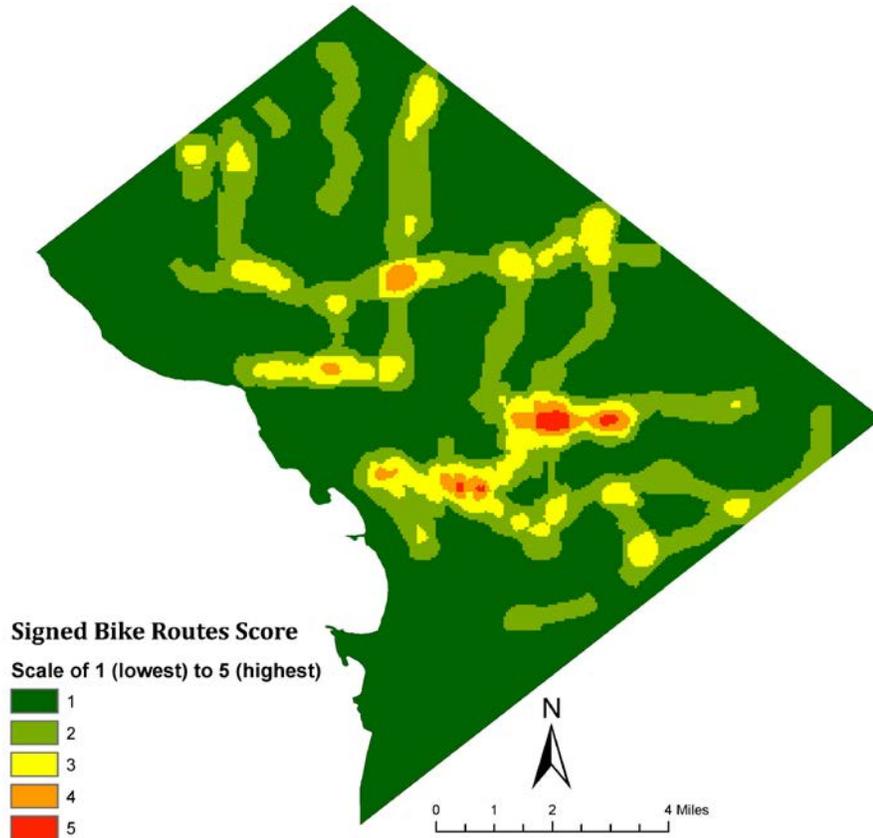
Figure 4 indicates the hilliness score, which was higher than the other factors; nearly half of the areas were red, and most parts of Washington were yellow, orange, and red. Therefore, when we consider only hilliness as the factor in

evaluating bikeability, the bikeability in Washington DC is positive and strong.



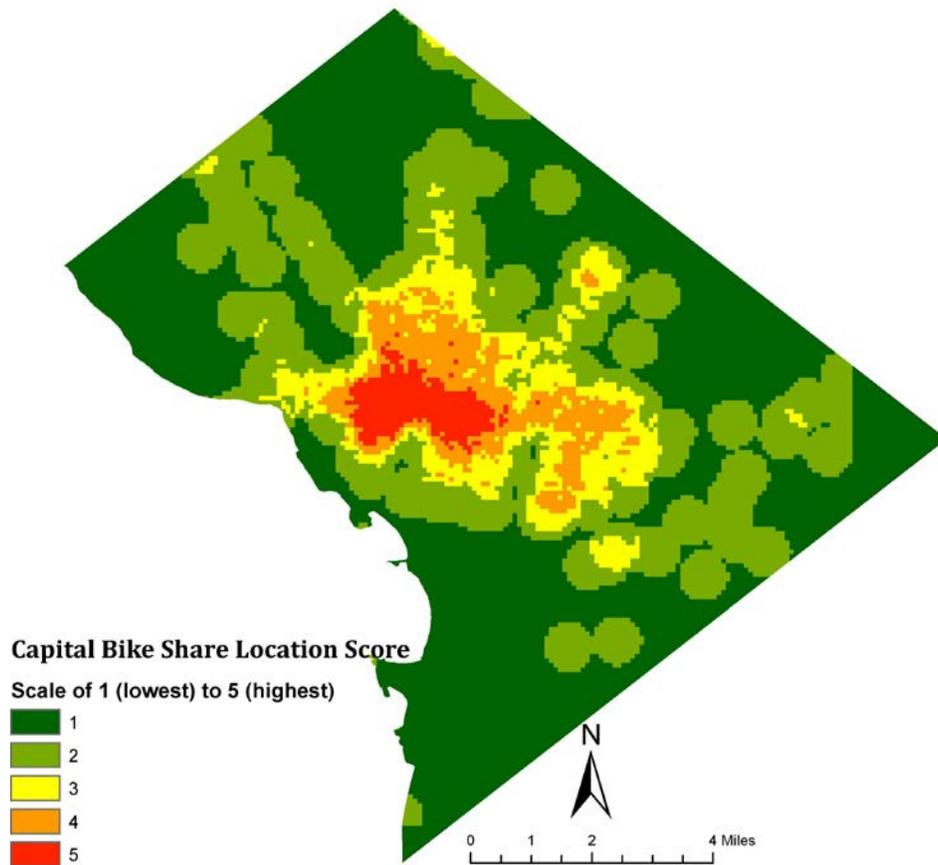
*Figure 4. Hilliness scores*

Compared with other factors described before, the map of signed bike routes was different (see Figure 5), and the trend in the density was consistent with the signed bike routes. Most areas in Washington DC have a low score, indicating that, in general, there are few signed bike routes in Washington.



*Figure 5. Signed bike routes scores*

In Figure 6, the area with the highest score was located in central Washington, and was similar to the map of destination density scores. Generally, the Capital Bikeshare locations are distributed centrally, with fewer locations outside the central area.



*Figure 6. Capital Bikeshare location scores*

Similar to the map of signed bike routes, the trend in density was consistent with the bike trail distribution (see Figure 7), with the high-density area located around the edges of the Washington DC boundary.

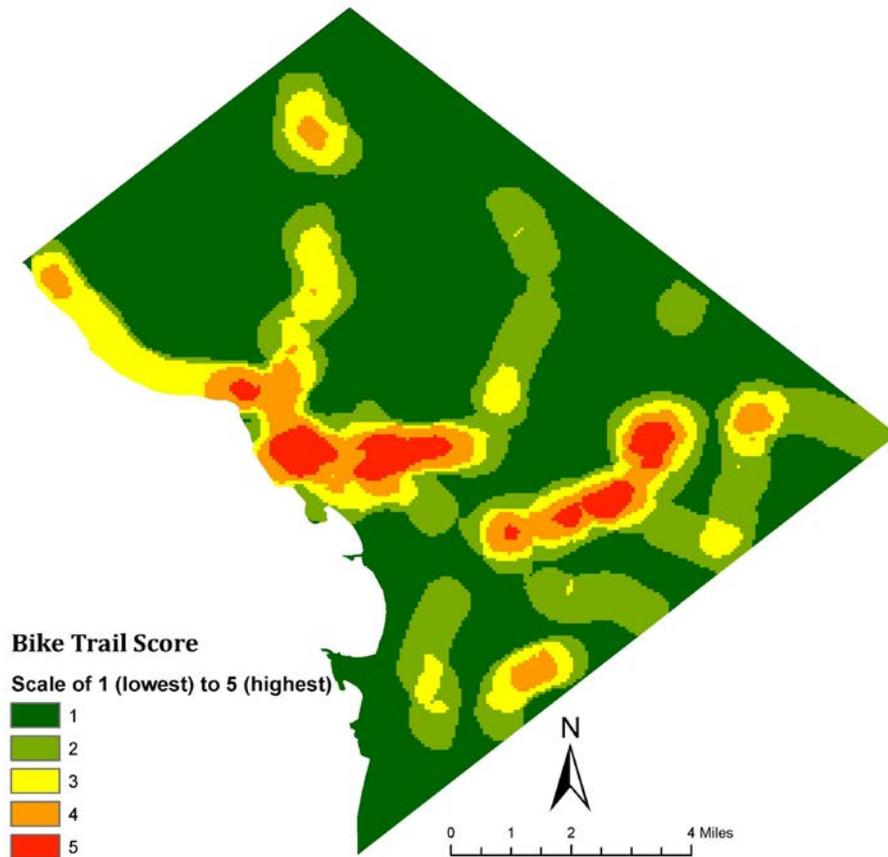
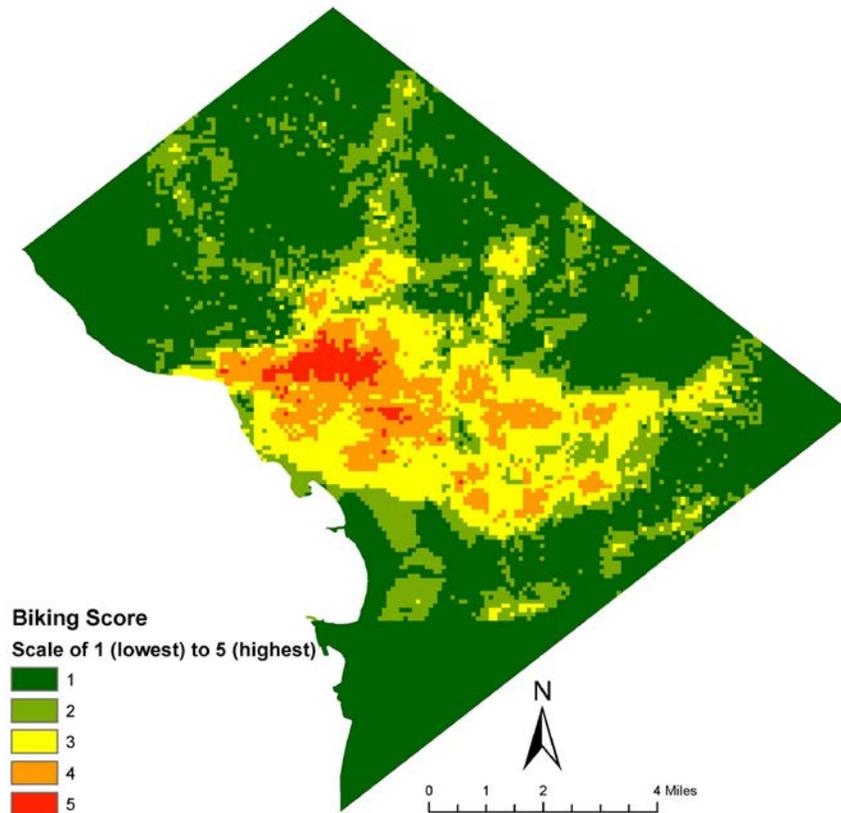


Figure 7. Bike trail scores

In evaluating the biking scores, I used all six factors in the following formula:

*Biking Score = Destination Density Score \* (100/6) + Intersection Density Score \* (100/6) + Hilliness Score \* (100/6) + Signed Bike Routes Score \* (100/6) + Capital Bike Share Location Score \* (100/6) + Bike Trail Score \* (100/6).* The best biking areas were adjacent, and biking scores between 3 and 5 were located primarily in central Washington. Compared to the individual maps of the other six factors, the destination density score and Capital Bikeshare location scores followed a pattern similar to the biking scores (see Figure 8).



*Figure 8. Biking scores*

To explain this more deeply and to give a better understanding, I use population density as the reference in Figure 9. The top left-hand map shows the population density in Washington DC. The red areas have a high population density and the green areas have a low population density. The top right-hand map is overlaid with the census tract number. We can see that the highest population density areas are located in Census Tracts 50.02, 37, 28.02, and 28.01. Furthermore, another seven of the second highest population density areas are located adjacent to the highest population density areas. The map on the bottom of the figure is a biking

score map overlaid with the census tract number. The best biking area is in Census Tract 107 and adjacent areas. Although Census Tract 107 has a low population density, the surrounding areas have a high population density. To conclude, the higher population density areas do not exactly match the best biking areas, but they are closed.

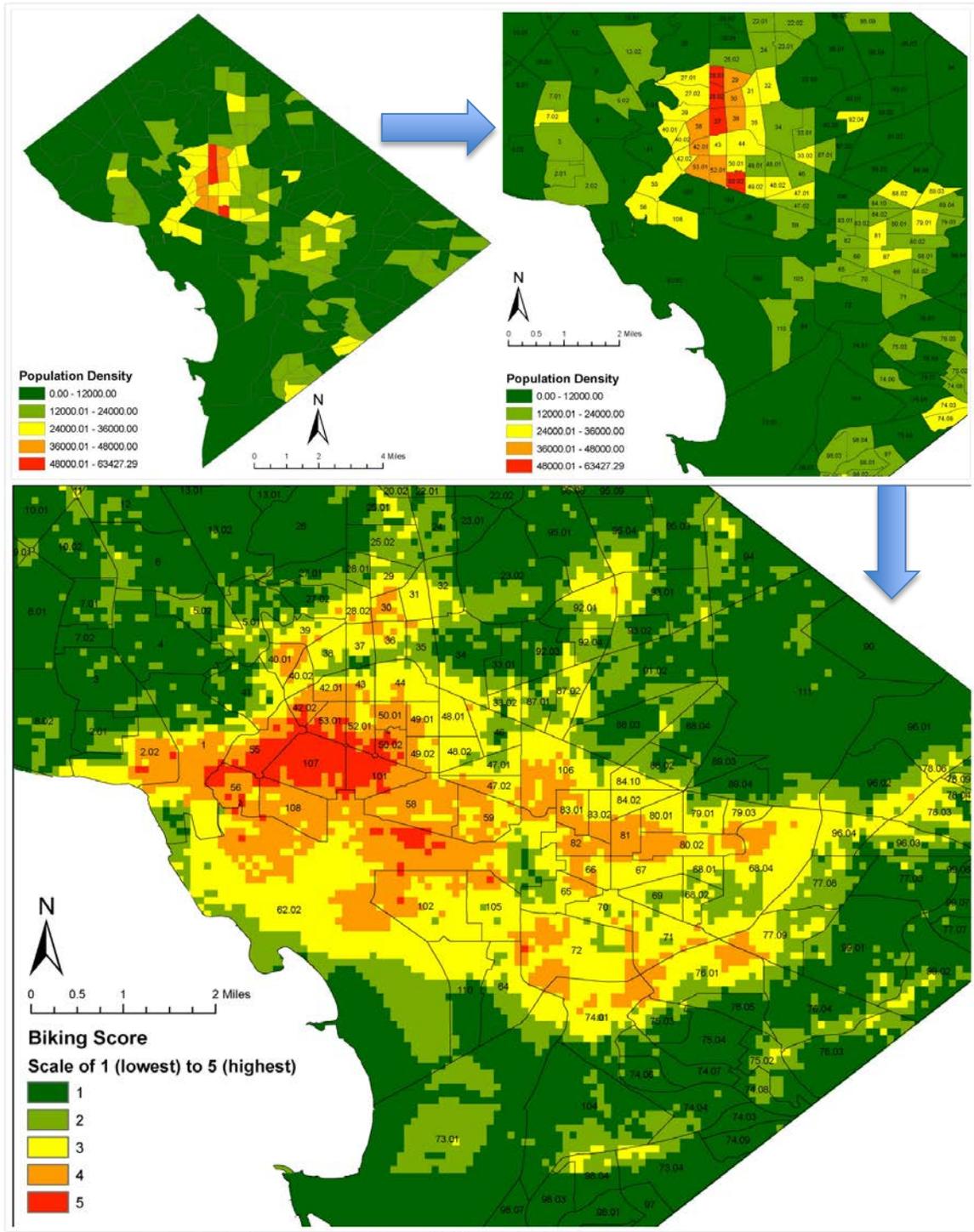


Figure 9. Biking Scores with Census Tract Overlay



## **Chapter 4: Methodology**

As mentioned in the previous chapters, the final primary purpose of this study was to investigate people's attitudes about bicycling in urban areas in Washington DC. In order to collect these data, I used the following method, based on the Tufts Urban Attitudes lab (UAL), a research team affiliated with the department of Urban and Environment Policy and Planning at Tufts University, directed by Dr. Justin Hollander. Those methods described below are correlated closely, and should be considered together.

### **Data Collection**

In order to collect and filter the micro-blog data, I used a computer program developed by the UAL to collect data from the Twitter Streaming API. This program collects and indexes a continuous stream of Tweets available publicly, which total less than 1% of all Tweets generated at a given location (Hollander, Graves, Levanthal, 2014). Geotagged Tweets were collected from March 3 to April 10, 2015 (total of 13 .CSV files). Figure 11 shows a small set of sample Tweets as displayed within UAL that were exported to Excel in the .CSV file format. Column A is the Twitter ID, column B is the user's ID, column C is Tweets, columns D and E display longitude and latitude, and column F is the time the Tweets were posted.

A	B	C	D	E	F	G
5.73E+17	2802200413	@heartoutlander @beulahcrusoe @stenbergmika @Heughliots @heughanize	-77.086826	38.889995	2015-03-03	15:32:36
5.73E+17	996017294	Back	-77.000631	38.925059	2015-03-03	15:32:38
5.73E+17	190787153	Aint going tell u twice I'm tell u once	-76.917868	38.853918	2015-03-03	15:33:05
5.73E+17	349921132	Nancy Pelosi is in the house. #Israel #IranDeal #jpost	-77.009266	38.88853	2015-03-03	15:33:16
5.73E+17	180458590	Este sueño me tiene de muy mal humor y CERO tolerancia y CERO paci	-77.026458	38.99114	2015-03-03	15:33:16
5.73E+17	375383005	@georgetown @RobertMGroves if we want to reach all survivors confi	-77.070697	38.907674	2015-03-03	15:33:19
5.73E+17	1387436876	#AAP shld consider replacing the Jhadu with the scorpion as their s	-77.039532	38.905299	2015-03-03	15:33:22
5.73E+17	190787153	@RafielVistaGWO: What's understood don't gotta be explained _ _	-76.917899	38.85392	2015-03-03	15:33:47
5.73E+17	2802200413	@beulahcrusoe @stenbergmika @Heughliots @heughanizers @Fans0SamHeug	-77.086826	38.889995	2015-03-03	15:33:48
5.73E+17	2641339091	Friendship height we here's today.!!! <a href="http://t.co/XdZk8gH1JK">http://t.co/XdZk8gH1JK</a>	-77.085308	38.959943	2015-03-03	15:33:50
5.73E+17	311267542	Im So Over These Big Ass Prenatal Pill	-76.978076	38.845022	2015-03-03	15:34:04
5.73E+17	380290690	Because cat Sabi swim no mean say E be catfish this Wan sef na pun	-76.99386	38.926592	2015-03-03	15:34:08
5.73E+17	128467838	Never get sloppy drunk but alcohol is problem solving!	-77.011489	38.895257	2015-03-03	15:34:10
5.73E+17	2587789764	#Nursing #Job alert: RN / Registered Nurse / Travel RN job   Supple	-77.052129	38.904237	2015-03-03	15:34:15
5.73E+17	450113186	Cannot stand it when people in college brag about their grades #not	-76.996555	38.9341	2015-03-03	15:34:20
5.73E+17	59785902	_ @qweenpush: <a href="http://t.co/VHUcoitZiK">http://t.co/VHUcoitZiK</a> me	-77.020233	38.924031	2015-03-03	15:34:23

Figure 11. Sample Tweets from Washington DC

All 13 .CSV files included more than 400,000 rows of data, and thus, it was necessary to select only those Tweets related to “biking.” Therefore, all data were filtered by the following key words: bike; bicycling/cycling; biking; bicycle; cyclists/bicyclists; bike facilities; bike infrastructure; bikeability/bikeabilities. Ultimately, 366 data points were obtained. Further, I collected the same number of randomly selected data points (366) from the original data, except the 366 Tweets about biking selected previously. This helped compare people’s attitudes about biking to that about other subjects.

The location of all 366 Bike-focus sample Tweets and 366 random sample Tweets was determined by longitude and latitude; Twitter has a setting that users can turn on or off to post the geolocations of their Tweets. Thus, I used ArcGIS to geolocate all Tweets (see Figure 12). The green dots shows the Bike-focus samples, and the red dots are random samples. From the figure 12, we cannot observe any strong patterns of distribution for both bike-focus samples and random samples, and these tweets

just posted everywhere in Washington DC. The Figure 13 and 14 shows the example of the Bike-focus tweets and random sample tweets in excel file (detailed excel file in Appendix II). In Figure 13, we can see people talked about “bike” in many ways, such as someone complained drivers, and someone expressed their love of biking.

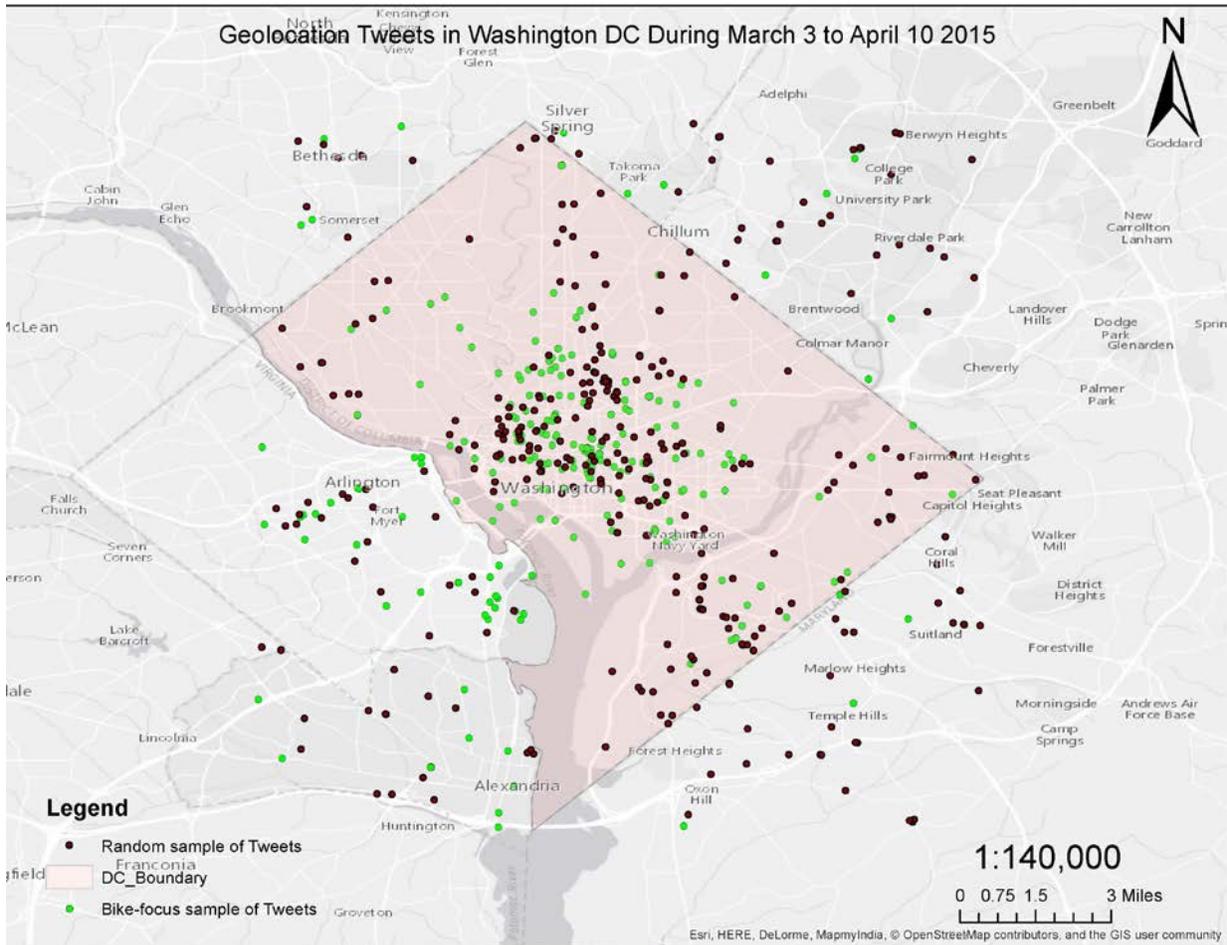


Figure 12. Geolocation of Tweets in Washington DC (Random sample and Bike-focus sample of Tweets)

5.72787E+17	259467435.00	Things I'm going to do in spring: read outside. Picnic. Ride my <b>bike</b> to school like it's the 90's all over again.	38.951828	-77.002712
5.73123E+17	237996608.00	Amazing any if us are alive 2day @oldpicsarchive: Einstein riding a <b>bicycle</b> down range from a nuclear bomb test. <a href="http://t.co/xLq6x9PAJ">http://t.co/xLq6x9PAJ</a>	38.897955	-77.010209
5.7287E+17	259340026.00	@USBicycleRoutes: RT @acaroutes: Time to act! US Reps voting Amtrak bill. Tell ur rep to support roll-on 4 <b>bikes</b> <a href="http://t.co/cqVmusK9E">http://t.co/cqVmusK9E</a>	38.916799	-77.005173
5.72899E+17	282196492.00	Moeen Khan talking about #RediscoveringPakistan-untold tale: <b>Motorbike</b> journey through Pakistan <a href="http://t.co/Tgyz35OhJF">http://t.co/Tgyz35OhJF</a>	38.943397	-77.068232
5.72899E+17	282196492.00	Moeen Khan talking about #RediscoveringPakistan-untold tale.at#PakEmbassy. <b>Motorbike</b> journey through Pak <a href="http://t.co/MECRHOAKTC">http://t.co/MECRHOAKTC</a>	38.943397	-77.068232
5.73107E+17	12609882.00	#bikedc has anyone been on mt vernon trail btwn alexandria and bridges lately? How are conditions?	38.818375	-77.057225
5.7319E+17	16987475.00	DDOTDC a reminder that the paint at 19th and M NW is usually ignored. Drivers make up their own rules in DC #bikedc <a href="http://t.co/iahO7fwoic">http://t.co/iahO7fwoic</a>	38.905925	-77.043672
5.7319E+17	498852873.00	Why is Arlington even considering cuts to walking/ <b>biking</b> improvements? The story coming up at 4:30 on @wamu85news #bikeva	38.882074	-77.076369
5.73199E+17	384657755.00	Few things have made me as happy as riding my <b>bike</b> in Amsterdam did <a href="#">#</a>	38.951653	-76.972017
5.73203E+17	556246838.00	<b>Bike</b> is being transported to dc. #nicee	38.908302	-76.997939
5.73232E+17	498852873.00	Tune to @wamu85news at 4:30 for my story on Arlington County considering cuts to <b>bike</b> /walk improvements. #bikedc #bikeva	38.928067	-77.034041
5.73233E+17	81913437.00	Gear Prudence: Why Don't Cyclists Use Available <b>Bike Lanes</b> ? Gear Prudence: Occasionally when I'm driving along a <a href="http://t.co/hULbgYSW0p">http://t.co/hULbgYSW0p</a>	38.901815	-77.037336
5.73283E+17	58186248.00	#Pakistan <b>biker</b> & director #MoinKhan introducing his doc Rediscovering Pakistan @DifferentAgenda @GirlRisingPak <a href="http://t.co/AZHbgfHrY7">http://t.co/AZHbgfHrY7</a>	38.898846	-77.047954
5.73291E+17	215452671.00	Shattered. # <b>BikeDC</b> #AnacoastiaRiver @ Frederick Douglass National Historic Site <a href="https://t.co/YVgnxihpjp">https://t.co/YVgnxihpjp</a>	38.863704	-76.984425
5.73312E+17	21544197.00	Best <b>bike</b> for a winter unsanctioned underground parking garage criterium? @Van_Dessel #Alcoominator of course.	38.864729	-77.051082
5.73347E+17	308802453.00	Only thing im missin is a car. I refuse ta ride up on a <b>bike</b> . Lmtao.	38.852656	-76.930918
5.73375E+17	133487008.00	MJHershey09 in response to that favorite I did NOT fall off the <b>bike</b> the boys just made that up cuz they're mean <b>dy™...dy™...dy™...</b>	38.89956	-77.04869

Figure 13. Example of bike-focus sample Tweets from Washington DC

5.73E+17	2802200413	@heartoutlander @beulahcruose @stenbergmika @Heughliots @heughanizers @FansOSamHeughan thank u!	38.889995	-77.088626
5.73E+17	1606990741	Deberian exportar Paso de los Toros. Estoy necesitada.	38.986135	-77.088226
5.73E+17	185434420	I like doing what we do when nobody around... <b>s</b>	38.923723	-77.021206
5.73E+17	49530624	@TalkerNewYorker amazing. Same w/ Ted. Those who fought the nazis and saved others displayed an incredible level of bravery and courage.	38.888833	-77.030568
5.73E+17	152819240	When bitches make mistake do yall feel bad about it	38.860555	-76.949044
5.73E+17	47081057	@shawmblanc oh yeal! "I start reading it _omorrow I think.	38.90382	-77.030916
5.73E+17	1890536647	Just woke up	38.925295	-77.004523
5.73E+17	427830477	Thanks for everything @CutonDime25.	38.900051	-77.048335
5.73E+17	66604591	@Chantellesay guilty..."	38.936629	-77.024647
5.73E+17	17931851	Big sappy pinot fruit #delectableapp @ Rasika West End <a href="https://t.co/9b44NkC0u5">https://t.co/9b44NkC0u5</a>	38.905468	-77.047572
5.73E+17	2367317738	B5 Ref's!	38.884446	-77.112643
5.73E+17	13913575	Nini Krever and me at #AIPAC. We met in early1960s at Bay Beach in Ontario Canada where our <a href="https://t.co/IUqwin3JU">https://t.co/IUqwin3JU</a>	38.814138	-77.04044C
5.73E+17	532286836	@jonkarl @jasoninthehouse You are a political hack. Surprised ABC hasn't promoted you after this hit piece.	38.831993	-77.007542
5.73E+17	419810713	Imao luv life	38.995274	-76.992623
5.73E+17	220508211	@imfreddiemac already	38.802371	-77.08365
5.73E+17	232673125	I stayed with the latest BabyPhat on my back	38.922069	-77.018206
5.73E+17	279272025	911 I've heard the word millennials three times before 9 am.	38.939251	-77.084953

Figure 14. Example of random sample Tweets from Washington DC

## Sentiment Analysis

Once I collected all of the “biking” Tweets, I conducted a sentiment analysis using SentiStrength, which uses a set of Tweets and processes the dataset line by line to match words in a sentiment dictionary, after which it assigns either a positive or negative score to each Tweet.

SentiStrength is available in ten languages, among them English, Arabic, French, Greek, Italian, Persian, Polish, Portuguese, and Swedish, and it is possible to apply Tweets in each language to the same scale. This software also can apply certain emoticons, such as “:(, “:)”, “^\_^” and “<3.” This lexicon has been used in a number of peer-reviewed articles and multiple studies and projects, and thus is

respected academically (Thelwall, Buckley, & Paltoglou, 2011; Pfitzner, Garas, & Schweitzer, 2012; Garas, Garcia, Skowron, & Schweitzer, 2012).

For each Tweet, SentiStrength (2016) reports negative emotion scores from -1 (not negative) to -5 (extremely negative), and positive emotion scores from 1 (not positive) to 5 (extremely positive). For the purposes of this study, I used positive, negative, and overall emotion scores, and analyzed them separately; the overall score was equal to the sum of the positive and negative scores. Figure 15 shows an example of the sentiment score for a single Tweet; the overall score for this sample would be 1, because the positive emotion score was 2, and the negative emotion score was -1. When a large number of data are processed, the software creates a .txt file with the same concept as the example. To begin the sentiment analysis, I used SentiStrength to process the data needed, and exported them to a .txt file. Then, I modified it as an Excel file (see Figure 16) to prepare for the next statistical analyses.

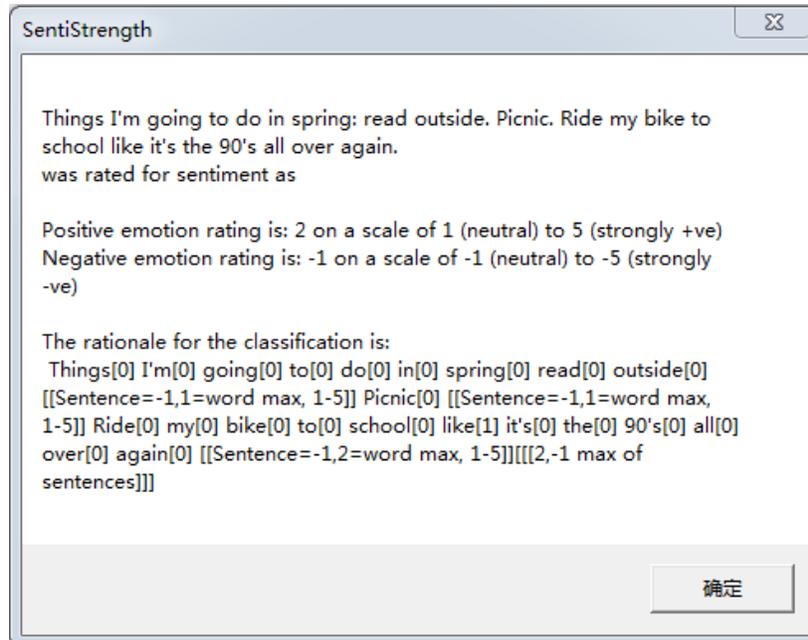


Figure 15. Example of the sentiment analysis for a single Tweet

A	B	C	D	E	F	G
Twitter	TwitterUserID	Tweets	latitude	longitude	positive_e	negative_e
5.73596E+17	275797956	Adult (barely) snow day #dcwx #running @ Rock Creek Park Bike Trail <a href="https://t.co/UUf14mK">https://t.co/UUf14mK</a>	38.9235	-77.0487	2	-1
5.73611E+17	59793427	Trying to enjoy the winter weather. @ Rock Creek Park Bike Trail <a href="https://t.co/yoS71xLEk">https://t.co/yoS71xLEk</a>	38.9235	-77.0487	3	-1
5.74193E+17	3014630120	The ole girl is steady for spring. #bicycle #fitnessgirls #npcbikini #fitnessmotivation <a href="http://t.co">http://t.co</a>	38.90372	-77.05855	1	-1
5.7495E+17	1055491120	Today's bike adventure. Somehow I cut my thumb. SE is not very bike friendly. Lol @ Frederick	38.86875	-77.00523	2	-2
5.7503E+17	10400902	Oh so good #BikeDC #Bliss @ Washington Monument National Monument <a href="https://t.co/psUDw">https://t.co/psUDw</a>	38.88988	-77.03458	2	-1
5.75033E+17	556246838	The bike scene out here is litt	38.90788	-77.00308	1	-1
5.75033E+17	10400902	It's melting... It's melting #BikeDC @ Potamic River <a href="https://t.co/aQbjw02LQ8">https://t.co/aQbjw02LQ8</a>	38.88091	-77.05033	1	-1
5.75048E+17	182447811	I love SE when that weather breaks ðŸ˜† 4 wheelers...bike...ALL the weed men be out....	38.83985	-76.99352	3	-2
5.75051E+17	296404744	I want an uber bike service where I hop on the back of some pegs	38.90254	-77.06257	1	-1
5.75799E+17	279211139	Sunset in DC behind Kennedy Center #BikeDC @ Watergate complex <a href="https://t.co/OZdm6DGXII">https://t.co/OZdm6DGXII</a>	38.89876	-77.05555	1	-1
5.75815E+17	9314472	First late night bike trek of the season around the National Mall. So happy right now #bikeDC @	38.8893	-77.05012	2	-1
5.7583E+17	387881212	Lookin for A Dirtbike Shop or Website No Pit Bikes	38.8992	-76.94164	1	-1
5.75832E+17	387881212	â€œ@_Bigpoppingmils: @CHB_Hugo16 I'm pullin the 4wheels out this summer brah ðŸ˜†ðŸ˜†CE	38.89919	-76.94151	1	-1
5.76109E+17	1042533690	Riding in circles around the Washington Monument with @bikeshare #bikeDC #bikettravel #bi	38.88988	-77.03458	1	-1

Figure 16. Example of sentiment analysis for Tweets after modification

While this method is the best way to process large amounts of data, it does have several weaknesses. Firstly, although this software can be applied in ten languages, the number of words in each lexicon in the software differs. For example, there is a total of 1,964 words in the Portuguese lexicon, of which 1,627 words (83%) are negative (SentiStrength, 2016). By comparison, 2,476 words were found in the English lexicon, and 1,585 (64%) of those were negative (SentiStrength,

2016). Although these biking data all used English, some of the random data selected used other languages. Further, emoji (see Figure 17) have become more and more popular in social media, but SentiStrength does not yet have a lexicon of emoji. In addition, the software can perform a sentiment analysis on only 122 emoticons; therefore, it has limited ability to test the remainder of the emoticons used. Thus, these weaknesses might have influenced the outcome of the sentiment analysis.



*Figure 17. Examples of Emoji*

## **ArcMap**

I used ArcMap to connect the sentiment scores and six physical environment factors: destination density; intersection density; hilliness; signed bike routes; Capital Bikeshare locations, and bike trails. The main purpose of using this method was to be able to input the geolocation of the Tweets into the maps, and to view the number of Tweets located in the areas identified; these results also were combined with the sentiment scores to prepare for the next part of the analysis. The table below presents a simple explanation:

Factors	Methods
<b>Signed bike routes, Capital Bikeshare locations, &amp; Bike trails</b>	<ol style="list-style-type: none"> <li>1) Imported geolocation Tweets into the ArcMap file</li> <li>2) Imported line/point shapefile about signed bike routes/ Capital Bikeshare location/bike trail from DCGIS Open Data</li> <li>3) Created four buffers and a line or point shapefile: 0.05 mile, 0.1 mile, 0.15 mile, and 0.2 mile</li> <li>4) Selected points located within the 0-0.05, 0.05-0.1 mile, 0.1-0.15 mile, and 0.15-0.2 mile buffers (see examples in Figures 18 &amp; 19) within Washington DC, and 0.2 mile was the bikeable distance</li> <li>5) Exported results as .txt file</li> <li>6) Modified .txt file and saved as Excel file:</li> <li>7) Provided the distance score for all points: 1 for points located within 0-0.05 mile, 2 for points within 0.05-0.1 mile, 3 for those within 0.1-0.15 mile, and 4 for those within 0.15-0.2 mile.</li> </ol>
<b>Destination density, Intersection density, &amp; Hilliness</b>	<ol style="list-style-type: none"> <li>1) Imported geolocation Tweets into the ArcMap file</li> <li>2) Imported shapefile of destination density score/intersection density score/hilliness score created from bikeability analysis (Chapter 3)</li> <li>3) Transferred raster to polygon</li> <li>4) Selected points located in score areas 1 to 5 individually within Washington Exported results as .txt file</li> <li>5) Modified .txt file and saved as Excel file:</li> <li>6) Provided the scores for all points: 1 for points located in area 1, 2 for points in area 2, 3 for points in area 3, and 4 for points in area 4.</li> </ol>

*Table 2. Methods of creating distance/value score variables*

A distance/value score was calculated to create an independent variable for quantitative analysis.

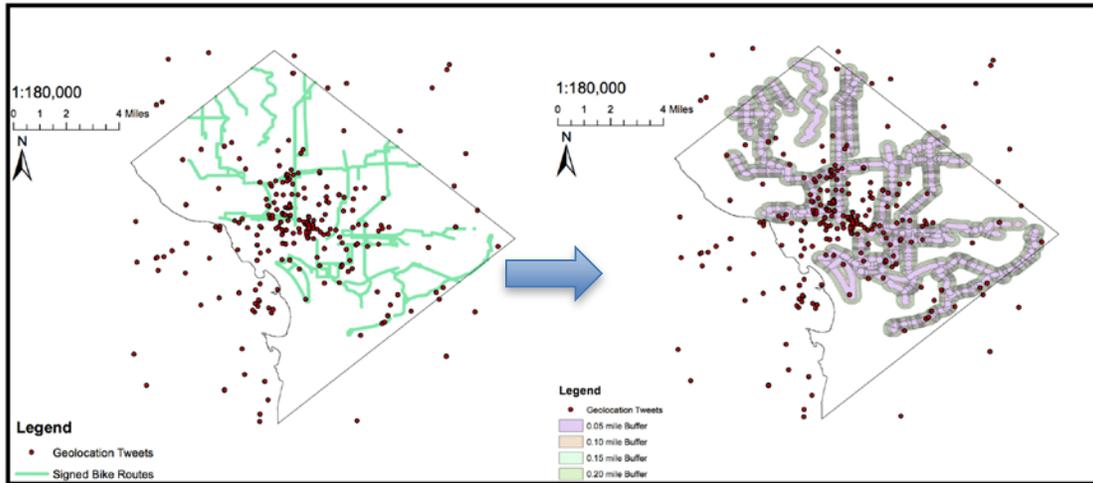


Figure 18. Example of creating buffers with line shapefile (signed bike routes)

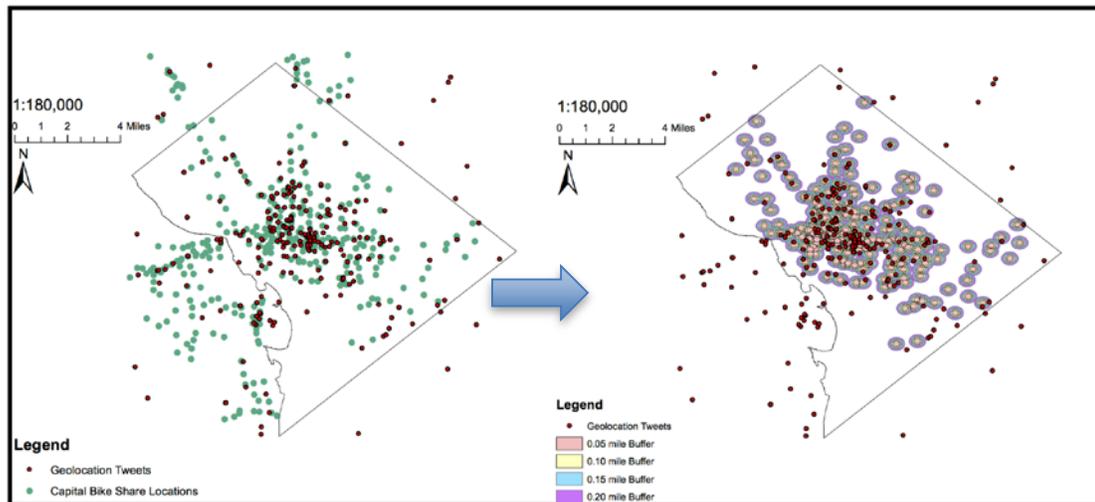


Figure 19. Example of creating buffers with points shapefile (Capital Bikeshare locations)

## Statistical Analyses

Based on all data collected, I performed a sentiment analysis, and used

ArcMap to create the independent variable. Then, I began the analysis using the general-purpose statistical software, STATA.

Firstly, I used the summarizing tool to obtain detailed summary statistics and tables that were constructed to compare positive and negative sentiment scores across the various distances/values to determine whether there were noticeable differences in the scores across these categories. Moreover, I used the graphic tool to create histograms used to evaluate the sentiment scores visually for trends across the distance/value measures. Moreover, a series of Chi-squared and Kruskal Wallis tests was used to explore whether the differences were statistically significant. In addition, Spearman correlations were used to explore the statistical relationship between variables. Finally, a univariate linear regression model was performed to explore the linear relationship between the sentiment scores and the six physical environment variables.

## **Chapter 5: Results and Analyses**

In this chapter, I first present the results of the comparison between all data collected related to biking and the data selected randomly from all original data. Then, I present the analysis of the sentiment scores for the six individual physical environment variables in several ways. Finally, I describe the results of a linear regression used to test the relationship between the sentiment score and those six variables.

### **Section 1: Biking data vs. Random data**

The purpose of comparing bike-focus sample Tweets and random sample Tweets was to determine whether people's attitudes about biking were below or above a standard level. Thus, because the original database was large, I selected another 366 Tweets from over 450,000 Tweets. The results of the data selected randomly from the original database represented the sentiment level of all Tweets.

Table 3 shows the comparison between these two datasets. The mean score of the random sample overall was -0.44, and the mean of the biking sample was 0.26. Thus, people's attitudes about biking were positive and higher than the sentiment level of all Tweets.

	Random sample Tweets			Bike-focus sample Tweets		
	Positive emotional score	Negative emotional score	Overall score	Positive emotional score	Negative emotional score	Overall Score
<b>Observation</b>	366	366	366	366	366	366
<b>Median</b>	1	-1	0	1	-1	0
<b>Mean</b>	1.44	-1.49	-0.44	1.54	-1.28	0.26
<b>Standard deviation</b>	0.69	0.85	1.10	0.74	0.61	0.94

*Table 3. Sentiment analysis results: Random sample Tweets vs. bike-focus sample Tweets*

## Section 2: Six factors analysis

- **Bike Trails**

Figure 20 shows that nearly half of the Tweets posted were within 0-0.05 mile of a bike trail, while only 8 Tweets were located between 0.1-0.15 mile from a bike trail. The figure shows no clear trend in the relationship between the number of Tweets posted and the distance to a bike trail. Further, because the total data sample is 366, and we can calculate that only 67 of the Tweets were posted within the distance identified, 299 Tweets were posted more than 0.2 miles from a bike trail.

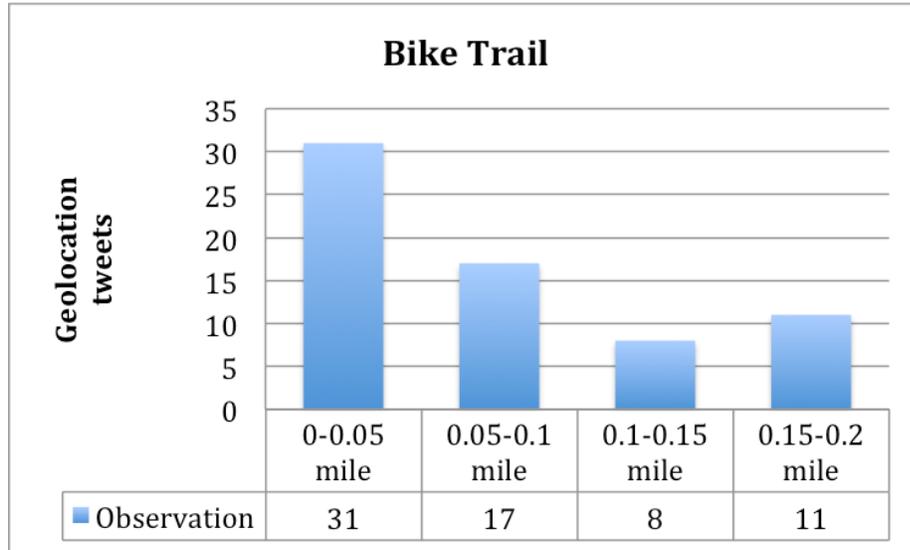


Figure 20. Distribution of geolocated Tweets by bike trails

Table 4 shows the detailed statistical summary of sentiment scores based on the distance to a bike trail. The mean of the sentiment scores overall for all four distance ranges was greater than 0, which indicates that people’s attitudes about biking on bike trails in Washington are positive. However, there was no specific pattern in those numbers.

Distance	Mean			Standard deviation		
	Positive score	Negative score	Overall Score	Positive score	Negative score	Overall score
0-0.05 mile	<b>1.58</b>	<b>-1.16</b>	<b>0.42</b>	<b>0.84</b>	<b>0.37</b>	<b>0.89</b>
0.05-0.1 mile	<b>1.47</b>	<b>-1.41</b>	<b>0.06</b>	<b>0.62</b>	<b>0.71</b>	<b>1.00</b>
0.1-0.15 mile	<b>1.62</b>	<b>-1.25</b>	<b>0.38</b>	<b>0.92</b>	<b>0.46</b>	<b>0.92</b>
0.15-0.2 mile	<b>1.36</b>	<b>-1.36</b>	<b>0</b>	<b>0.67</b>	<b>0.50</b>	<b>1.00</b>

*Table 4. Sentiment analysis results: Sentiment scores and distance from bike trails*

In order to understand the extent of the association between the independent and dependent variables, we must look at the correlations in Table 5. The hypotheses were that there is no association between the dependent and independent variables, respectively, and the null hypothesis was rejected at  $p < 0.05$ .

The  $p$ -values between the following variables were all greater than 0.05: “distance” and “positive score,” “distance” and “negative score,” “distance” and “overall score,” and “negative score” and “positive score,” which indicates that they were not statistically significant associated with each other. In addition, the  $p$ -value was 0.17 between the “negative score” and “distance,” and although it was greater than 0.05, it can be considered only marginally significant. Thus, this suggests that there was a very slight inverse relationship between “distance” and “negative score.”

Several measures had a  $p=0$ : “overall score” and “positive score,” and “overall

score” and “negative score;” further, their r-values all were greater than 0.5, indicating that there was a very high and positive relationship between “distance” and “positive score” and “negative score.”

	1	2	3	4
1.distance	r=1.00			
2. positive score	r=-0.07 p=0.60	r=1.00		
3. negative score	r=-0.17 p=0.17	r=0.00 p=0.10	r=1.00	
4.overall score	r=-0.15 p=0.23	r=0.80 p=0.000	r=0.57 p=0.000	r=1.00

Table 5. Correlation between distance to bike trail and sentiment scores

- **Capital Bikeshare Locations**

Figure 21 shows that a total of 207 Tweets was posted within 0.1 mile of Capital Bikeshare locations, and 111 Tweets were within the range of 0.05 to 0.1 mile. Only 40 Tweets were within 0.1 to 0.15 miles, and 19 Tweets were within 0.15 to 0.2 mile. Moreover, 266 (73%) of the 366 Tweet samples were located within 0.2 mile of a Capital Bikeshare location. Thus, over 70% of Tweets were posted within the range identified, and 111 (30%) were located in the 0.05 to 0.1 mile range.

Figure 16 does not show any trend in the distribution of Tweets, and we can say only that over 50% of Tweets (207) were posted within 0.1 mile of a Capital Bikeshare

location.

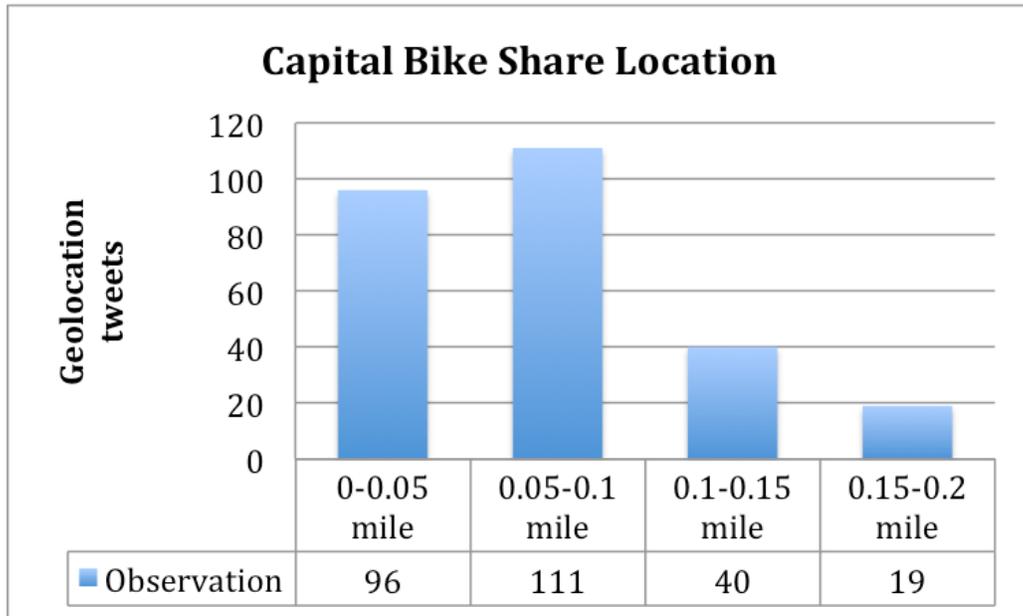


Figure 21. Distribution of geolocated Tweets by Capital Bikeshare locations

Table 6 indicates that the means of the overall scores were positive in the following distance ranges: 0-0.05 mile, 0.05-0.1 mile, and 0.15-0.2 mile, while the mean of the overall score was negative between 0.15 mile and 0.2 mile from a Capital Bikeshare location. This suggests that people’s attitudes were positive in posts within 0.15 mile of these locations; however, beyond a distance of 0.15 mile, people’s attitudes became negative. Because of the limitations of this study, we do not have any data for distances over 0.2 mile, and thus this result has possible to be inaccurate.

Distance	Mean			Standard deviation		
	Positive score	Negative score	Overall score	Positive score	Negative score	Overall score
0-0.05 mile	1.69	-1.34	0.42	0.85	0.37	0.89
0.05-0.1 mile	1.50	-1.18	0.32	0.75	0.49	0.82
0.1-0.15 mile	1.55	-1.20	0.35	0.64	0.52	0.89
0.15-0.2 mile	1.11	-1.32	-0.21	0.32	0.58	0.54

*Table 6. Sentiment analysis: Sentiment scores and distance from Capital Bikeshare locations*

Table 7 indicates that the  $p$ -value exceeded 0.05 for the following: “negative score” and “distance,” “overall score” and “distance,” and “positive score” and “negative score,” indicating that there was statistically insignificant association between them.

However, the  $p$ -value between “positive score” and “distance” was less than 0.05 ( $p=0.03$ ), and they had a low association ( $r=0.14$ ). Further, there was a significant inverse association between the positive sentiment score and distance to a Capital Bikeshare location. Thus, an increase in the distance to a Capital Bikeshare location decreased the positive sentiment score. In addition, there was a strong, positive association between “overall score” and “positive score,” and between “overall score” and “negative score.”

	1	2	3	4
1. Distance	r=1.00			
2. Positive score	r=-0.14 p=0.03	r=1.00		
3. Negative score	r=0.08 p=0.21	r=0.03 p=0.65	r=1.00	
4. Overall score	r=-0.08 p=0.20	r=0.81 p=0.000	r=0.57 p=0.000	r=1.00

Table 7. Correlation between distance to Capital Bikeshare locations and sentiment scores

- **Signed Bike Routes**

Figure 22 indicates that 72 Tweets were located in the 0 to 0.05 mile distance range, 32 were located between 0.05 to 0.1 mile, 16 were located between 0.1 to 0.15 mile, and 65 were located between 0.15 to 0.2 mile from a signed bike route. There was no significant or clear pattern in the relationship between distance and the number of Tweets. Further, 185 (51 %) of all Tweets were located within 0.2 mile of signed bike routes.

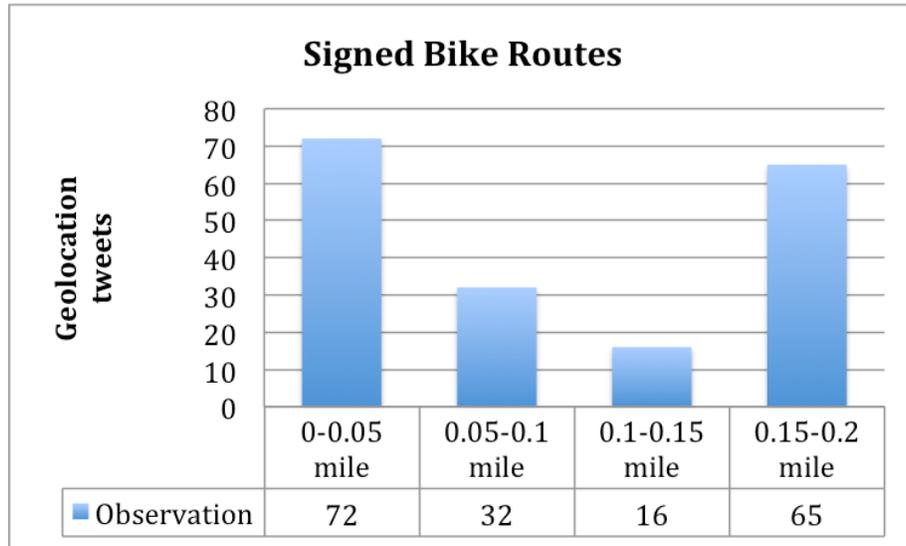


Figure 22. Distribution of geolocated Tweets by signed bike routes

Table 8 shows that the means of the overall scores in all four distance ranges did not follow any specific pattern. Within the distance range of 0 to 0.05 mile and 0.15 to 0.2 mile, the means of the sentiment score overall were positive (0.42 and 0.35). Between 0.05 mile to 0.1 mile, the mean of the score overall was 0, indicating that people had neutral attitudes and no preference in biking when only the factor of signed bike routes was considered. However, for the distance range of 0.1 to 0.15 mile, the mean of the overall sentiment became negative (-0.125), which suggests that people’s attitudes were negative overall with respect to biking in this distance range.

Distance	Mean			Standard deviation		
	Positive score	Negative score	Overall score	Positive score	Negative score	Overall score
0-0.05 mile	1.61	-1.19	0.42	0.85	0.49	0.98
0.05-0.1 mile	1.41	-1.41	0	0.61	0.84	0.98
0.1-0.15 mile	1.13	-1.25	-0.125	0.50	0.45	0.72
0.15-0.2 mile	1.58	-1.23	0.35	0.70	0.49	0.92

*Table 8. Sentiment analysis: Sentiment score and distance to signed bike routes*

Table 9 shows the results of the correlations between distance to signed bike routes and the three dependent variables. From the table, we can see that the  $p$ -value was clearly insignificant between the following variables: “distance” and “positive score,” “distance” and “negative score,” “distance” and “overall score,” and “positive score” and “negative score.” Thus, there was no statistically significant association between the variables listed above. Two sets of data in the table had  $p$ -values less than 0.05: “positive score” and “overall score,” and “negative score” and “overall score.” These had a strong positive association.

	1	2	3	4
1. Distance	r=1.00			
2. Positive score	r=-0.00 p=0.98	r=1.00		
3. Negative score	r=-0.04 p=0.60	r= 0.05 p=0.52	r=1.00	
4. Overall score	r=-0.03 p=0.71	r=0.83 p=0.000	r=0.56 p=0.000	r=1.00

Table 9. Correlation between distances to signed bike routes and sentiment score

In comparing the three variables above, we can see that 67 of the Tweets were posted within 0.2 mile of bike trails, 111 Tweets were within 0.2 mile of Capital Bikeshare locations, and 185 Tweets were located within 0.2 mile of signed bike routes. Thus, more people Tweeted close to signed bike routes, and fewer people Tweeted close to bike trails.

- **Intersection Density**

Figure 23 indicates that 184 (52%) of the total Tweets posted were located in an area where the intersection density score was equal to 3. The second highest number of Tweets (67) was distributed in areas with a score of 2. There were 56 Tweets in score 1 areas, 33 Tweets in score 4 areas, and 13 Tweets in score 5 areas. In addition, 353 Tweets were posted in score areas 1 to 5, and only 13 Tweets were posted outside those signed areas. Thus, almost all Tweets were located in signed score areas 1 to 5. Moreover, there was no clear trend in the distribution of these

Tweets and intersection density scores.

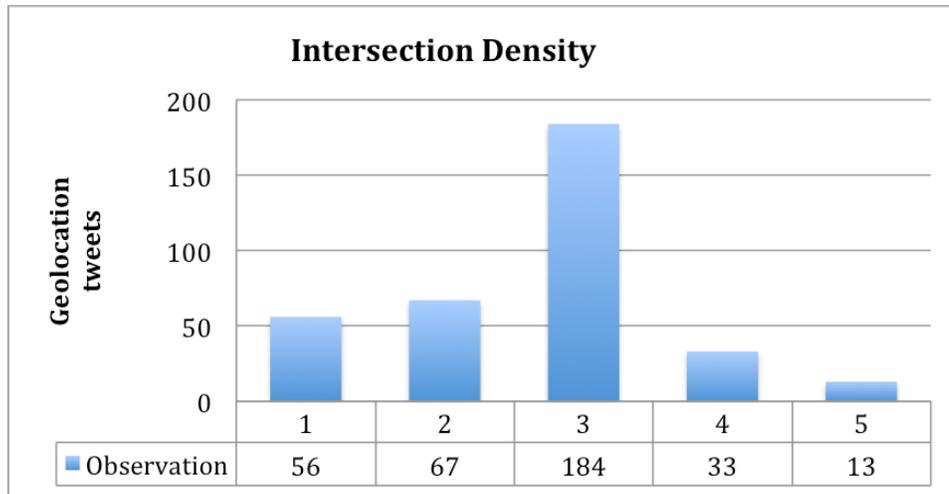


Figure 23. Distribution of geolocated Tweets by intersection density

Table 10 shows that the means of the overall scores in all five identified areas were positive. Thus, considering intersection density alone, people’s attitudes about biking were positive. However, although the mean of the overall score for area 5 was positive, it was significantly lower than the others (mean= 0.08), suggesting that people’s attitudes in this area were close to neutral. Moreover, there was no clear pattern among those numbers in the five areas.

Intersection Score	Mean			Standard deviation		
	Positive score	Negative score	Overall score	Positive score	Negative score	Overall score
1	1.57	-1.23	0.34	0.63	0.50	0.67
2	1.39	-1.24	0.15	0.67	0.55	0.89
3	1.63	-1.26	0.37	0.81	0.61	1.00
4	1.45	-1.27	0.18	0.67	0.57	0.85
5	1.31	-1.23	0.077	0.63	0.44	0.76

Table 10. Sentiment analysis results: Sentiment and intersection density scores

Table 11 shows the correlations between the intersection density scores and the three independent variables: positive score, negative score, and overall score. The *p*-value between the following variables far exceeded 0.05: “intersection density score” and “positive score,” “intersection density score” and “negative score,” “intersection density score” and “overall score,” and “positive score” and “negative score.” Thus, these variables were not statistically significant associated.

However, two sets of data in the table had *p*-values less than 0.05: “positive score” and “overall score,” and “negative score” and “overall score.” Therefore, they were associated strongly and positively.

	1	2	3	4
1. Intersection density score	r=1.00			
2. Positive score	r=-0.06 p=0.92	r=1.00		
3. Negative score	r=-0.01 p=0.82	r= -0.03 p=0.59	r=1.00	
4. Overall score	r=-0.01 p=0.85	r=0.81 p=0.000	r=0.53 p=0.000	r=1.00

Table 11. Correlation between Intersection density scores and sentiment scores

- **Hilliness**

Figure 24 illustrates that 363 (99%) Tweets were posted in all five signed areas. Thus, only 3 Tweets were posted outside these areas. 143 Tweets were posted in score 1 areas, and 178 Tweets in score 2 areas; therefore, nearly 90% (321) of Tweets were posted in these two areas. It is worth mentioning here that only 8 and 6 Tweets, respectively, were posted in score areas 4 and 5; this might affect the sentiment analysis in the next part because of the small sample size.

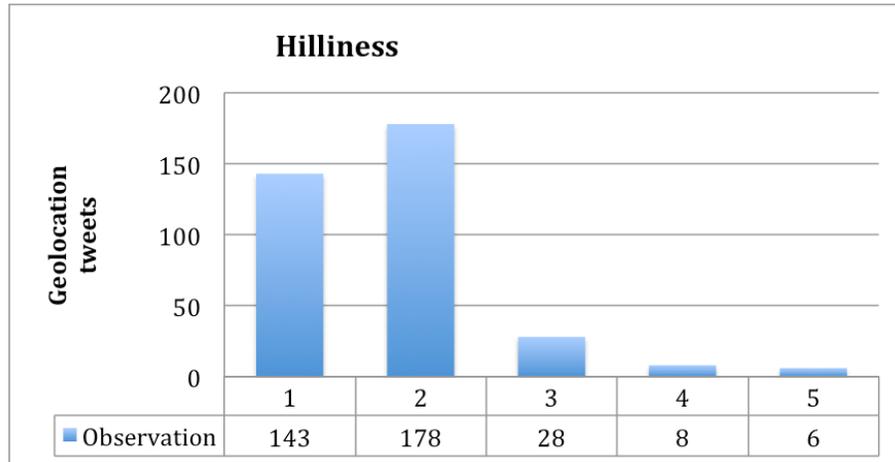


Figure 24. Distribution of geolocated Tweets by hilliness

Table 12 presents a detailed summary of the sentiment and hilliness scores. There was no significant clear trend, although the means of the overall scores were all equal to or greater than 0. The highest mean score overall, 1, occurred in score 5 areas. The means of the overall scores in areas 1, 2, and 3 were positive, indicating that people’s attitudes about biking with respect to hilliness were positive in those areas. In score 4 areas, the means of the overall score were 0, indicating that people’s attitudes were neutral. As mentioned before, the samples for score 4 and 5 areas were very small; thus, the results might be erroneous.

Hilliness Score	Mean			Standard deviation		
	Positive score	Negative score	Overall score	Positive score	Negative score	Overall score
1	1.52	-1.37	0.15	0.73	0.73	1.03
2	1.55	-1.21	0.34	0.75	0.52	0.88
3	1.5	-1.25	0.25	0.69	0.58	0.93
4	1.25	-1.25	0	0.71	0.46	0.53
5	2	-1	1	0.89	0	0.89

Table 12. Sentiment analysis: Sentiment and hilliness scores

Table 13 shows the results of the correlations between hilliness score and the three independent variables. There was statistically insignificant association between “positive score” and “hilliness score,” and “negative score” and “positive score,” because their *p*-values were 0.87 and 0.76, respectively.

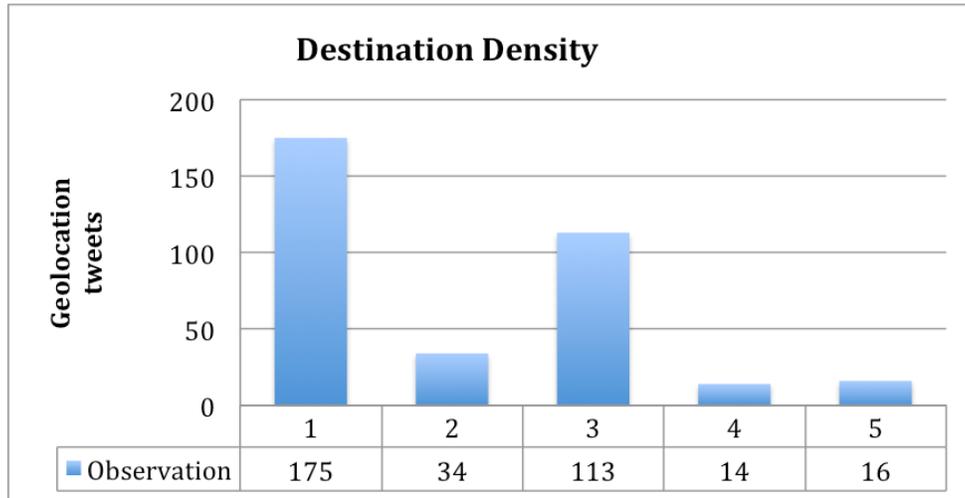
Because the *p*-value for the hilliness and negative sentiment scores was 0.06, it was marginally significant. However, their association was significant, as the *r* value was 0.10. Further, the *p*-value between the “overall score” and “hilliness score” was 0.20, which was marginally significant, indicating a weak relationship between the two. The *p*-value was 0 between the “overall score” and “positive score,” and “overall score” and “negative score,” indicating a strong, positive association between the two.

	1	2	3	4
1. Hilliness score	r=1.00			
2. Positive score	r=0.01 p=0.87	r=1.00		
3. Negative score	r=0.10 p=0.06	r=-0.02 p=0.76	r=1.00	
4. Overall score	r=0.07 p=0.20	r=0.80 p=0.000	r=0.56 p=0.000	r=1.00

Table 13. Correlation between hilliness scores and sentiment scores

- **Destination density**

Figure 25 shows no straightforward pattern in the results, as the Tweets were posted randomly in the five categories. A total of 175 Tweets was posted in score 1 destination density areas, and 113 Tweets were posted in score 3 areas. Only a few Tweets were located in score 2, 4, and 5 areas compared to the other two areas (34, 14, and 16). Thus, nearly all of the Tweets (352) were posted in the areas identified.



*Figure 25. Distribution of geolocated Tweets by destination density*

Table 14 shows that the means of the overall sentiment scores were positive in score areas 1, 2, 3, and 4, and all four were similar. Thus, we can conclude that, with respect to destination density, people’s attitudes about biking were similarly positive in those score areas. The mean of the overall sentiment score was negative in score area 5, indicating that people’s attitudes about biking became negative at that destination density.

Destination density Score	Mean			Standard deviation		
	Positive score	Negative score	Overall score	Positive score	Negative score	Overall score
1	1.54	-1.26	0.28	0.73	0.55	0.88
2	1.5	-1.18	0.32	0.82	0.45	0.88
3	1.6	-1.26	0.34	0.75	0.65	0.96
4	1.57	-1.14	0.43	0.85	0.36	1.02
5	1.25	-1.375	-0.125	0.44	0.62	0.89

*Table 14. Sentiment analysis: Sentiment and destination density scores*

Table 15 presents the correlations between the destination density scores and the three categories of sentiment scores. The *p*-value was extremely high between the following variables: “destination density score” and “positive score,” “destination density score” and “negative score,” “destination density score” and “overall score,” and “positive score” and “negative score,” indicating that there were no significant associations among them statistically.

Similar to all of the variables above, a strong association existed between “overall score” and “positive score,” and “overall score” and “negative score.” The reason for this was that the overall score was calculated by adding the positive and negative scores, as mentioned in Chapter 4.

	1	2	3	4
1. Destination density score	r=1.00			
2. Positive score	r=-0.01 p=0.88	r=1.00		
3. Negative score	r=0.03 p=0.60	r=-0.03 p=0.57	r=1.00	
4. Overall score	r=0.00 p=0.99	r=0.81 p=0.000	r=0.53 p=0.000	r=1.00

Table 15. Correlation between destination density scores and sentiment scores

### Section 3: Linear Regression

Table 16 shows the results of the regression between the sentiment scores and the six individual independent variables. The hypothesis was that there is no significant relationship between the dependent and independent variables. As the table shows, most  $p$ -values were higher than 0.05, and thus, the null hypothesis was accepted.

However, two  $p$ -values less than 0.05 are highlighted in the table. The  $p$ -value between the hilliness and negative sentiment scores was 0.04, and thus, the null hypothesis was rejected in that case. When the hilliness score increased by 1, the negative sentiment score increased by 0.08, indicating that a higher hilliness score reduced people's negative attitudes about biking (detailed tables in Appendix IV). The  $r$ -squared were very low in this case, and they were around 0.01 and 0.03

shown in the table 16. It means the model explains only 1% and 3% of the response data around their means.

There was another significant relationship between the distance to Capital Bikeshare locations and the positive sentiment score. The  $p$ -value between them was 0.01, and thus, the null hypothesis was rejected. When the distance to a Capital Bikeshare location increased by 1 unit, the positive sentiment score fell by 0.14. Thus, the greater the distance to a Capital Bikeshare location, the more negative people’s attitudes about biking (detailed tables in Appendix IV).

Six independent variables	$p$ -value		
	Positive score	Negative score	Overall score
Intersection density score	0.92	0.76	0.78
Destination density score	0.69	0.91	0.69
Hilliness score	0.68	0.04 coeff.=0.82 R-squared=0.0119	0.10
Distance to signed bike routes	0.72	0.87	0.71
Distance to Capital Bikeshare locations	0.01 coeff. =-0.14 R-squared=0.0280	0.31	0.12
Distance to bike trail	0.53	0.26	0.25

Table 16. Results of linear regression between independent and dependent variables

## **Chapter 6: Conclusions, Limitations, and Recommendations for Future Research**

In this concluding section, I will discuss the results of the research in light of the research questions. I will also discuss the limitations of this study, and will make recommendations for future relevant studies.

### **Conclusions**

The first research question asked, “How is bicycling ability presented in Washington DC?” To address the question, I used six indices to calculate bicycling scores in Washington: intersection and destination density, hilliness, bike trails, signed bike routes, and Capital Bikeshare locations. I was unable to provide specific scores for the whole city, but by using these six individual factors and processing ArcMap, I found that the best biking areas were adjacent, and were located primarily in the central areas of the city.

The second research question was, “What are people’s attitudes (positive or negative) about bicycling in Washington?” and the answer overall was “positive.” As Chapter 5 mentioned, by comparing the biking data to random data, the sentiment scores for biking data were positive and higher than were those for random data. Based on the results presented in Chapter 5, I concluded that people’s attitudes towards biking in Washington are positive.

The last research question asked, “Is there a relationship between people’s

attitudes, the bicycling index (cycling infrastructures, hilliness, desirable amenities, and road connectivity), and locations where people talk about bicycling?” Because the answer is complex, I will describe it specifically as follows, based on the six biking indices. Firstly, there was no clear pattern between people’s attitudes and the distance to bike trails and signed bike routes. Further, the results showed no relationship between people’s attitudes and different scores for intersection and destination densities.

However, the analyses did demonstrate some significant relationships. The linear regression showed relationships between the hilliness score and negative sentiment scores, and distance to Capital Bikeshare locations and positive sentiment scores. When the hilliness score increased by 1, the negative sentiment score increased by 0.08, indicating that a higher hilliness score reduced people’s negative attitudes about biking. Another set of data were related to the distance to Capital Bikeshare locations and positive sentiment scores. When the distance to Capital Bikeshare location increased by 1 unit, positive sentiment scores fell by 0.14, suggesting that greater distances to Capital Bikeshare locations reduced positive attitudes.

Thus, the answers to the research questions were relatively conclusive. However, these results could be improved further by optimizing the whole process. I

discuss the limitations of this study in the next section.

## **Limitations**

- *Data collection*

Some limitations in the process overall included the ways in which the data were collected, as well as the statistical analyses, either or both of which may have influenced the outcomes of the study.

The first limitation relates to data collection. Firstly, I collected Twitter data for only 5 weeks. Thus, the sample size was not sufficiently large to be representative. In addition, all of the data collected were from Twitter, and contained words and emoticons only. Therefore, some pictures, videos, and emoji were not considered and applied in the sentiment analysis. Further, within the small sample size (366), some samples came from one Twitter user. Thus, if a user was overrepresented, some elements of his/her personality may have biased the data. For example, if a user was a negative person, s/he might be likely to complain about everything, and thus, his/her Tweets might skew the results negatively, given that the sample size was small already. The Figure 26 is the example of selected Tweets from one users. I pick up user 46416755 as the example here. The user 46416755 posted 27 tweets during 5 weeks about biking, and the mean of positive score from these tweets is 1.52, and the mean of negative score from these tweets is -1.19.

Based on these evidences, the attitudes of user 46416755 is tend to positive, and have possible to influence the results.

TwitterID	TwitterUserID	Tweets	positive emotion rating	negative emotion rating
573195000000000000	498852873	Why is Arlington even considering cuts to walking/biking improvements? The story coming up at 4:30 on @wamu	2	-1
573232000000000000	498852873	Tune to @wamu885news at 4:30 for my story on Arlington County considering cuts to bike/walk improvements. #	2	-1
573472000000000000	498852873	Progressive Arlington may cut funding for bike/pedestrian programs: <a href="http://t.co/GLeFnUoNK4">http://t.co/GLeFnUoNK4</a> @wamu885news #	1	-1
573909646999744000	498852873	@WABADC says @DDOTDC not effectively implementing bike projects at a pace that meets demands of growing	1	-1
573910303253086000	498852873	@WABADC's Greg Billing says DDOT can't get projects done. And now... they're attempting to lower expectations	1	-1
573911440500854000	498852873	@darsal @beyonddc I didn't realize there were bike lanes there.	1	-1
573916116839448000	498852873	Oversight hearing continues. Dornsjo yet to testify. Lots of testimony from bike advocates ANCs re: need for bett	2	-2
573916330719603000	498852873	@darsal @beyonddc The number of miles of protected bike lanes is rather small. Can't remember off top of my h	1	-1
573918034215235000	498852873	@marychey says we still don't have comprehensive bike lane network. Key word: network. Also mentioned lack c	1	-2
576374174110862000	498852873	Look at all that traffic! #bikedc <a href="http://t.co/FRvg7kbvVW">http://t.co/FRvg7kbvVW</a>	1	-1
576374604572332000	498852873	Look who is riding a bike to work today. (Hint: @marychey) <a href="http://t.co/RHwfoYteFJ">http://t.co/RHwfoYteFJ</a>	1	-1
576380878324161000	498852873	Is @DDOTDC moving fast enough to build a protected bike lane network? Tune in Monday to @wamu885news fo	1	-1
576410565482000000	498852873	@jeffreanders19 @maustermuhle Protected bike lanes are the clincher for some riders who are iffy about doing	1	-1
574271447222321000	896556121	It's all good @ewilliams0305! Always great riding mighty #HillsOfAnacostia w/ you & @moehers! @BicycleS	3	-1
574286354374355000	896556121	Frederick Douglas Bridge a #bikedc @DDOTDC disaster! Today's @BicycleSPACE #HillsOfAnacostia ride w/ @ewilli	1	-2
574334957075700000	896556121	My @jamisbicycles I rode on the mighty #HillsOfAnacostia today! @BicycleSPACE @ewilliams0305 @moehers #t	1	-1
574364168654076000	896556121	It was all @jamisbicycles riding the mighty #HillsOfAnacostia today! @BicycleSPACE #JamisBikesDC #bikedc <a href="http://t.co/896556121">http://t.co/896556121</a>	1	-1
574615748624171000	896556121	On @BicycleSPACE #CityExplorers ride! @MOMsOrganicMirt food stop! @CrMoBike @KuglerCycles @Jrkinsella #	1	-1
576697318763914000	896556121	Rain or shine! Its no jive! I am riding the mighty #HillsOfAnacostia today! 8a @BicycleSPACE @moehers #bikedc I	1	-1
576832105784717000	896556121	Stoked! to see @bikeshare carrying @jamisbicycles! Nice! #JamisBikesDC #bikedc <a href="http://t.co/36HxcM0nFM">http://t.co/36HxcM0nFM</a>	2	-1
577238546362269000	896556121	Mighty @PhilKinDC showing off his old school racing strip sweats on @BicycleSPACE #CityExplorers ride today! #b	1	-1
58555070867211000	896556121	Walking enriches your life! Just found 1989 Lincoln Penney on the National Mall! @allwalkscd @BikeWalk #walkk	1	-1
585780020402114000	896556121	My @GiantBicycle riding across the giant Frederick Douglass Bridge! Wrong Mayor on sign! @DDOTDC @MayorB	1	-2
585952071901392000	896556121	Today @DCCPoliceDept #bikedc patrol officer showed me proper way to carry my bike up the stairs in John Marsh	2	-1
586140799060024000	896556121	Wal-Mart on H St in #bikedc no customer parking for bike riders! No bike racks in parking garage! Thanks @Walm	2	-1
575352290124566000	46416755	This is a removable/lockable rear trunk! The design is sweet! #womenbike #NBS15 <a href="http://t.co/UWtwx3N8Lj">http://t.co/UWtwx3N8Lj</a>	2	-1
575360082713571000	46416755	@MCMHandles worked on advocacy in Boyle Heights- all the people selected for leading the project were from B	1	-1
575360379303126000	46416755	If you are asking community leaders to do outreach for your Transpo project. make sure you fund them for their t	1	-1
575362938388619000	46416755	To reach the Boyle Heights community @MCMHandles went to local businesses- not come to me but go to them	1	-1
575367220257161000	46416755	Just had a presenter drop #Kept100 at the National Bike Summit. #nbs15 #womenbike	1	-2
575377683120422000	46416755	The entire room at #womenbike just chanted outdoor Afro #NBS15	1	-1
575378317701873000	46416755	@ambrown @qpxd definitely is- @peopleforbikes study says Women are more likely to bike for rec alone and rec	1	-1
575379289930750000	46416755	All of these discussions at #NBS15 confirm for me: bike fun is the future of coalition building for bike advocacy. #w	2	-1
575379663024881000	46416755	@BikePortland lots of people still don't think that's true! And sadly many of them are the people working on this	2	-2
575381164258561000	46416755	@GoddardTara I'm not quite sure what your question is but what I mean is our biggest potential for growth is cul	2	-1
575381854250884000	46416755	@ambrown I agree we need more family friendly bike fun- but I think the advantage of bike fun it is specifically tu	2	-1
575382065357019000	46416755	@gerikkkransky @GoddardTara should've come to my talk on Bike Fun building advocates last year ;)	2	-1
5753823347121014000	46416755	@GoddardTara I'm not saying replacing car trips- that's the. Broadening the bike community means focusing on b	1	-1
575383468672057000	46416755	@GoddardTara @gerikkkransky if we focus on only commute trips we lose those that can't commute by bike beca	1	-2
575383797174140000	46416755	@GoddardTara @gerikkkransky we find women are more likely to participate in activities/recreation in groups- hei	2	-1
575384514085580000	46416755	@GoddardTara @gerikkkransky Utility=Fun isn't what I was saying. Bike Fun- community based rides specially by&	2	-1
575384948971953000	46416755	@GoddardTara To build women&minority bike leaders we first need to get them on bikes. Bi	2	-1
575385921916264000	46416755	@GoddardTara @gerikkkransky Andando en Bicicletas en Cully and We all Can Ride and Black Women Bike- great t	3	-1
575386316646416000	46416755	@GoddardTara @gerikkkransky The #womenbike keynote last year talked about how in the 1970s no women ran	1	-1
575386564215230000	46416755	@gerikkkransky @GoddardTara I didn't say Bike Fun is the only key to getting people on bikes. But it absolutely car	2	-1
575386915089879000	46416755	@gerikkkransky @GoddardTara It's much easier to invite someone to a neighborhood bike ride than to a city coun	1	-1
575386316646416000	46416755	@GoddardTara @gerikkkransky The #womenbike keynote last year talked about how in the 1970s no women ran	1	-1
575386564215230000	46416755	@gerikkkransky @GoddardTara I didn't say Bike Fun is the only key to getting people on bikes. But it absolutely car	2	-1
575386915089879000	46416755	@gerikkkransky @GoddardTara It's much easier to invite someone to a neighborhood bike ride than to a city coun	1	-1
575387246112587000	46416755	@gerikkkransky @GoddardTara really? How about ABC members who lead Day of the dead rides and also write gri	1	-3
575388389161049000	46416755	@gerikkkransky @GoddardTara it's the first step - fee want to advocate for bikes until they start riding them.	1	-1
575402978133307000	46416755	@ladyfleur @cyclicious here's the bike fun podcast! <a href="http://t.co/6AOss2R6MZ">http://t.co/6AOss2R6MZ</a>	2	-1

Figure 26. Example of Bike-focus sample Tweets from one user

Another limitation resulted from setting the distance variable. Further, when I created buffers on the bike trail map, Capital Bikeshare location map, and signed bike routes map, I created four buffers from 0 to 0.2 mile, and thus, only partial data were included in the statistical analyses. For example, four buffers were created for bike trails, and only 67 Tweets were posted within 0.2 mile, while 299 samples

(81%) were not controlled in this case. Consequently, these data collection and modification methods would have affected the accuracy of the results.

Moreover, as mentioned previously, only 67 Tweets were posted within 0.2 mile from bike trails, 111 Tweets were posted within 0.2 mile from a Capital Bikeshare location, and 185 Tweets were posted within 0.2 mile from signed bike routes. Therefore, different sample sizes were used in the sentiment analysis and statistical analyses, and the sample size was reduced further by filtering. These factors also may have biased the results.

When I made the bicycling score map in Chapter 3, I gave six physical environmental factors equal weights when calculating the score, which may have affected the map's accuracy.

Finally, I used longitude and latitude to geolocate Tweets, but the locations where the Tweets were posted are not necessarily the same as the locations to which the people were referring. For example, people may talk about a bike trail when they are at home. The user's location in this case would have been tracked as his/her home address rather than the location s/he discussed. Although I faced this limitation from the beginning, at present, there are no methods to address the problem.

- *Methodology*

Some other methods used in this thesis have limitations as well. First, for each set of data in the statistical analyses, there was only one independent and dependent variable each. Therefore, I considered only one factor when I conducted the statistical analyses, and fewer independent variables will create some biases in the results.

Further, the software I used to obtain the sentiment score has limitations. This software uses different scales for positive and negative words in different languages. Although my biking data included only Tweets in English, the random data included other languages. In addition, this software cannot be applied to mixed-language Tweets, both of which factors might be reflected in the results.

### **Future recommendations**

Although the study had some limitations, some suggestions and recommendations can be offered. The first suggestion addresses future relevant studies. Based on the limitations described above, future studies may be able to overcome or minimize the limitations in this study. For example, adding more physical environmental factors would address some of the potential biases in the statistical analyses. The sample size also should be expanded considerably. Moreover, identifying more relationships or including more factors in the analysis would help cycling planners and policymakers make decisions in the development of

biking.

In Chapter 5, I indicated that the distance to Capital Bikeshare locations was related negatively to the positive sentiment score. Thus, it is clear that Capital Bikeshare locations are very important to bicyclists in Washington DC. Some urban areas have already implemented Capital Bikeshare location programs or similar programs, and increasing the number of such programs might be a good strategy to meet bicyclists' needs. In cities that do not yet have such a program, planners can begin to develop them as a good start in implementing effective bicycling plans.

Further, higher hilliness scores were found to have a statistically significant relationship with peoples' negative attitudes. While this analysis was limited and many variables were not included, there may be an important relationship between topography and bike user attitudes that deserves further exploration by local government planners. But more than anything, the findings presented here illustrate the remarkable power of social media data to shed light on urban policy issues and help shape the way that planning can happen.

## References

- American Association of State Highway and Transportation Officials. (1999). *Guide for the development of bicycle facilities*. Washington DC: American Association of State Highway and Transportation Officials.
- Adam, F, Chris, B. & Annik, S. (2011). You are what your friends eat: systematic review of social network analyses of young people's eating behaviours and bodyweight. *BMJ Global Health*, 1-8.
- Beirao, G. & Gavral, J. A. S. (2007). Understanding attitudes towards public transport and private cars: A qualitative study. *Transport Policy*, 14, 478-489.
- Ben, A. (2011). The Rise of Urban Biking. *The Nation* Instigating progress Daily. Retrieved from <http://www.thenation.com/article/rise-urban-biking/>
- Clifton, K. L. & Handy, S. L. (2001). *Qualitative methods in travel behavior research*. Paper presented at the International Conference on Transport Survey Quality and Innovation, Kruger National Park, South Africa.
- Dill, J. & Carr, T. (2003). Bicycle commuting and facilities in major US cities: If you build them, commuters will use them. *Journal of the Transportation Research Board*, 1828(1), 116-123.
- Eivik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accid. Anal. Prev.*, 41, 849-855.
- Frank, L. D., Engelke, P. O., & Schmid, T. L. (2003). *Health and community design: The*

- impact of the built environment on physical activity*. Washington, DC: Island Press.
- Furth, P. G. (2012). *City cycling: Chapter 6: Bicycling infrastructure for mass cycling: A transatlantic comparison*. Boston, MA: MIT Press. pp. 105-139.
- Gotschi, T. (2011). Cost and benefits of bicycling investments in Portland, Oregon. *Journal of Physical Activity and Health, 8*(1), 49-58.
- Grabow, M. L. (2013). *Active transportation and health: Predictors from urban design and modeled commuting behavior*. ProQuest LLC. 1-140.
- Garas, A., Garcia, D., Skowron, M., & Schweitzer, F. (2012). Emotional persistence in online chatting communities. *Scientific Report, 2*, article 402.
- Handy, S. L. (1996). Urban form and pedestrian choices: A study of Austin neighborhood. *Transp. Res. Rec., 1552*, 135-144.
- Handy, S. L. (2009). *Walking, bicycling, and health*. Healthy Equitable Transportation Policy. Oakland, CA: PolicyLink.
- Handy, S. L., Xing, Y., & Buehler, T. J. (2010). Factors associated with bicycle ownership and use: A study of six small U.S. cities. *Transportation, 37*, 967-985.
- Hollander, J., Graves, E., & Levanthal, T. (2014). Using big data to study urban sentiments: Twitter data vs. published meeting minutes. Tufts University.

- Jacobsen, P. (2003). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Inj. Prev.*, 9, 205-209.
- Killingsworth, R. E. (2003). Health promoting community design: A new paradigm to promote healthy and active communities. *Am. J. Health Promote.*, 17, 169-170.
- Krizek, K. J., Wl-Geneidy, A., & Thompson, K. (2007). A detailed analysis of how an urban trail system affects cyclist's travel. *Transportation*, 34(5), 611-624.
- Litman, T., Blair, R., Demopoulos, B., Eddy, N., Fritzel, A., Laidlaw, D.,...& Forster, K. (2015). *Pedestrian and bicycle planning: A guide to best practices*. Canada: Victoria Transport Policy Institute.
- Molina, J. (2014). *The case for crowdsourcing in bicycle planning: An exploratory study*. ProQuest LLC. 2-135.
- Murphy, N. B. & Knoblauch, R. (2004). *Hispanic pedestrian and bicycle safety*. Washington DC: The Federal Highway Administration (FHWA) and National Highway Transportation Safety Administration.
- Meghan,W., Michael, B., Eleanor, M.B., & Kay, T. (2012). Mapping bikeability: A spatial tool to support sustainable travel. *Environment and Planning B: Planning and Design 2013*, 40, 865-883.
- National Household Travel Survey (NHTS). (2009). US Department of Transportation and Federal Highway Administration, 2001.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.

Pfitzner, R., Garas, A., & Schweitzer, F. (2012). Emotional divergence influences information spreading in Twitter, ICWSM-12.

Robinson, D. (2005). *Safer in numbers in Australia: More walkers and bicyclists, safer walking and cycling*. *Health Promot. J. Austr.*, 16, 47-51.

Schroeder, P. & Wilbur, M. (2013). *2012 National survey of bicyclist and pedestrian attitudes and behavior, Volume 1: Summary report*. Report No. DOT HS 811 841 A. Washington, DC: National Highway Traffic Safety Administration.

Susilo, Y. O., Williams, K., Morag, L., & Dair, C. (2012). The influence of individuals' environmental attitudes and urban design features on their travel patterns in sustainable neighborhoods in the UK. *Transportation Research Part D*, 17, 190-200.

Sallis, J.F., Owen, N. (2008). Ecological models of health behavior. *Health behavior and health education: theory, reseach, and practice*, 4, 465-486.

Sentistrength. (2016). About Sentistrength. Retrieved from

<http://sentistrength.wlv.ac.uk/#About>

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-

418.

Twitter.com. (2015).About Twitter. Retrieved

from<https://about.twitter.com/company>

United States Environmental Protection Agency. (2013). *Sources of Greenhouse Gas*

*Emissions*. Washington DC: United States Environmental Protection Agency.

# Appendices

## Appendix I

Major Industry Group Search Tips ⓘ Collapse ▾ Remove ✕

- Agriculture, Forestry, & Fishing (01-09) 0 selected
- Mining (10-14) 0 selected
- Construction (15-17) 0 selected
- Manufacturing (20-39) 0 selected
- Transportation (40-49) 0 selected
- Wholesale / Distributors (50-51) 0 selected
- Retail Trade (52-59) 1153 selected
  - 52-BUILDING MATERIALS & HARDWARE 0 selected
  - 53-GENERAL MERCHANDISE STORES 0 selected
  - 54-FOOD STORES 81 selected
    - 5411-GROCERY STORES 9 selected
    - 5421-MEAT & FISH MARKETS 18 selected
    - 5431-FRUIT & VEGETABLE MARKETS 6 selected
    - 5441-CANDY NUT & CONFECTIONERY STORES 0 selected
    - 5451-DAIRY PRODUCTS STORES 0 selected
    - 5461-RETAIL BAKERIES 0 selected
    - 5499-MISCELLANEOUS FOOD STORES 48 selected
  - 55-AUTOMOTIVE DEALERS & SERVICE STATIONS 0 selected
  - 56-APPAREL & ACCESSORY STORES 103 selected
    - 5611-MENS & BOYS CLOTHING STORES 0 selected
    - 5621-WOMENS CLOTHING STORES 0 selected
    - 5632-WOMENS ACCESSORY & SPECIALTY STORES 17 selected
    - 5641-CHILDRENS & INFANTS WEAR STORES 0 selected
    - 5651-FAMILY CLOTHING STORES 0 selected
    - 5661-SHOE STORES 14 selected
    - 5699-MISC APPAREL & ACCESSORY STORES 72 selected
  - 57-HOME FURNITURE & FURNISHINGS STORES 0 selected
  - 58-EATING & DRINKING PLACES 969 selected
  - 59-MISCELLANEOUS RETAIL 0 selected
- Finance, Insurance, & Real Estate (60-67) 0 selected
- Services (70-89) 0 selected
- Public Administration (91-98) 0 selected
- Nonclassified EstablishmentS (99) 0 selected

[Clear Field\(s\)](#)

*Filter results in retail trade*

- Services (70-89) 161 selected
  - 70-HOTELS ROOMING HOUSES & CAMPS 0 selected
  - 72-PERSONAL SERVICES 0 selected
  - 73-BUSINESS SERVICES 0 selected
  - 75-AUTO REPAIR SERVICES & PARKING 0 selected
  - 76-MISCELLANEOUS REPAIR SERVICES 0 selected
  - 78-MOTION PICTURES 0 selected
  - 79-AMUSEMENT & RECREATION SERVICES 0 selected
  - 80-HEALTH SERVICES 0 selected
  - 81-LEGAL SERVICES 0 selected
  - 82-EDUCATIONAL SERVICES 161 selected
    - 8211-ELEMENTARY & SECONDARY SCHOOLS 0 selected
    - 8221-COLLEGES & UNIVERSITIES 12 selected
    - 8222-JUNIOR COLLEGES & TECHNICAL INSTITUTES 3 selected
    - 8231-LIBRARIES 14 selected
    - 8243-DATA PROCESSING SCHOOLS 3 selected
    - 8244-BUSINESS & SECRETARIAL SCHOOLS 8 selected
    - 8249-VOCATIONAL SCHOOLS NEC 31 selected
    - 8299-SCHOOLS & EDUCATIONAL SERVICES NEC 90 selected
  - 83-SOCIAL SERVICES 0 selected
  - 84-MUSEUMS ART GALLERIES & GARDENS 0 selected
  - 86-MEMBERSHIP ORGANIZATIONS 0 selected
    - 8611-BUSINESS ASSOCIATIONS 0 selected
    - 8621-PROFESSIONAL MEMBERSHIP ORGANIZATIONS 0 selected
    - 8631-LABOR UNIONS & SIMILAR ORGANIZATIONS 0 selected
    - 8641-CIVIC SOCIAL & FRATERNAL ASSOCIATIONS 0 selected
    - 8651-POLITICAL ORGANIZATIONS 0 selected
    - 8661-RELIGIOUS ORGANIZATIONS 0 selected
    - 8699-MEMBERSHIP ORGANIZATIONS NEC 0 selected
  - 87-ENGINEERING & ACCOUNTING & MGMT SVCS 0 selected
  - 89-MISCELLANEOUS SERVICES NEC 0 selected
- Public Administration (91-98) 0 selected
- Nonclassified Establishment S (99) 0 selected

[Clear Field\(s\)](#)

*Filter results in services*

















5.80882E+17 3087725546 @CrockgangUAY\_CongratsLPS1\*0Y\*Y 38.855277 -76.99

5.80882E+17 251410333 @That\_Guy\_Mizz: @viteezzyfromVA@\_2Raw I bumped it today... it's tight lol :| 38.848951 -76.946564

5.80882E+17 227336795 @yungwies im too hip 38.930281 -77.106378

5.81294E+17 2845770484 Baby lma hood st ardyey 38.892765 -76.911538

5.81294E+17 81420784 Nutter Butters are life 38.991614 -76.985178

5.81294E+17 1203151244 I regret getting my belly pierced (never again) 38.90174 -76.977336

5.81294E+17 37288434 Rip wvu 38.923659 -77.028032

5.81294E+17 475194300 I spent a little time with Dr. Don Chitty. 10x Purple Heart recipient visited the Vietnam Wall today. #capconn15 <http://t.co/6IALUDz40> 38.802223 -77.079322

5.81294E+17 386353555 The fire alarm never fails. The Vietnam War. You gotta be shittin me 38.837652 -77.01539

5.81294E+17 225577497 @ShelsDretta you wannatake it lol? 38.862235 -77.062396

5.81294E+17 377529064 @PHMantz I don't blame the refs. Our style we foul a lot so it's expected. We needed to make shots we don't usually make to have a chance 38.930595 -77.134247

5.81294E+17 1895036647 Refuse to let my heart keep takin L ydy 38.925234 -77.004656

5.81301E+17 16273359 Eivto told me i was one of the only girls he knows who looks good without makeup and i love him even more 38.888552 -77.093583

5.81301E+17 3048461905 head hurt 0Y\* 38.941518 -77.002414

5.81304E+17 1852296150 DChhttp://t.co/11p3BkjB 38.899888 -77.005952

5.81304E+17 18517521 I should've paid attention in Spanish class. They were going in though @ The Bier Baron Tavern <http://t.co/afvW2H4Gc> 38.91028 -77.048768

5.81304E+17 317122648 @TheOnly\_Yayaa: kkae@\_L\_Vulgarity: @TheOnly\_Yayaa i miss him too 0Y\* @k285Y\* 0Y\* 0Y\* shut up 0Y\* 0Y\* 0Y\* 0Y\* 0Y\* 0Y\* just tryna make you laugh 38.848773 -76.94932

5.81306E+17 246821237 Aww my baby 0Y\* 38.96635 -76.967688

5.81306E+17 76442156 Tbfm kkae@Desire\_Nae: When bae comes over and brings food <http://t.co/uf6r1h4H3p4> 38.994948 -76.912598

5.81306E+17 3048461905 tired 38.941482 -77.002488

5.81309E+17 326945201 kkae@yogajenny: Why is my dad talking shit 0Y\* 38.941194 -76.925044

5.81309E+17 2760944895 @TheOnlyKaypop : I don't even be press for the weekends no more 0Y\* 38.86027 -76.976644

5.81309E+17 430167058 all these riggas the same 0Y\* 38.836393 -76.953489

5.81313E+17 180145104 my bath was so much love 0Y\* 38.877524 -77.014633

5.81313E+17 242528187 stark face w/ it <http://t.co/h03JDzCQI> 38.921619 -77.017313

5.81318E+17 182978572 I love him more then words can express too 0Y\* @SweettAUSugari: Aiden loves his Godparents 38.897399 -76.970731

5.81318E+17 378379832 @TweetThis\_DiCK try updating Twitter 38.944541 -76.947818

5.81318E+17 614598674 @TIAE no she's annoying and hasn't done a thing to deserve such recognition 38.910197 -77.038109

5.81318E+17 133848037 We did. I just can't remember her name RT @eb4pres: I know yk\*\* all had a candy lady 38.83463 -76.926386

5.81318E+17 236119621 Disappear wen im bored. 38.894908 -76.980998

5.81325E+17 95488345 Going to NYC @ Union Station in Washington DC <http://t.co/5CTRPC313r> 38.897563 -77.006167

5.81325E+17 235186357 @\_XiaMonique\_ something confused upset n more. I miss you 0Y\* 38.838888 -76.982124

5.81325E+17 16113342 @HsARELbitches @kykstorm 0Y\* 0Y\* 0Y\* omgthess 38.965239 -76.978789

5.81328E+17 67685933 Dated number 2 0Y\* <http://t.co/BiUuJmAgf> 38.980755 -76.935831

5.81328E+17 816571716 How didi get here??? Lol that's how good the drinks are :p (at @CobaltDc in Washington DC) <http://t.co/SodLoJBE5V> 38.912666 -77.039830

5.81328E+17 228729720 I'll have you like she never had you 0Y\* 38.941512 -76.949596

5.81329E+17 180145184 kkae@\_lmao @Des\_ : Taurus are crazy 0Y\* @Des\_ : @no we are not 0Y\* 38.877475 -77.014623

5.81332E+17 14560696 When pot holes are so deep they hurt your soul 2D Cproblems 38.884847 -77.084785

5.81332E+17 41469816 kkae@SQUID: Bruh ain't even hant <http://t.co/20HvBbmQUk> my nans is straight outta Assasins Creed lmao 38.860382 -77.082579

5.81332E+17 232679125 I feel fat 38.92202 -77.018181

5.81334E+17 180145104 here come lil nae 0Y\* 38.877505 -77.014565

5.81334E+17 16917475 @wmat a @Metrorailinfo @MartiniCaro this says it all: metro employees report metro is NOT safe <http://t.co/T2d26Jexec> 38.838115 -77.078721

5.81334E+17 150404352 @sweden Didn't notice those. Impressive weight. 38.975271 -77.019379

5.86553E+17 170011798 Im 240 0Y\* i need to get back to when i was 220 38.984886 -76.987766

5.81335E+17 35622007 This thunderstorm is making me so happy! It's finally spring 0Y\* @C0000E 38.906444 -77.004905

5.82932E+17 75562204 DC brunch life #R058dc #bestoffee #finally \*Y\* \*Y\* CL @ Ted's Bulletin <http://t.co/ThqinOQFIS> 38.802429 -76.995141

5.83052E+17 18278344 @justaddScott just be glad you aren't an internet journalist. 38.876645 -77.006222

5.82935E+17 284485933 kkae@freshDRB: What do yall think i need to do to take my career to another level? kkae have Talent meh. just do what you do best! I God dey. 38.902362 -76.997534

5.83007E+17 411194625 Solidarity @ The Potter's House <http://t.co/2QZQVdH3P1> 38.92511 -77.089896

5.83013E+17 561039137 @drtprincess @amygeurden @1SACE85 @Caro\_L\_Etc @li79793244 @susiel3 0Y\* 38.825256 -77.081818

5.83018E+17 3087725546 I might let your boy chauffeur me 0Y\*... but he gotta eat the booty like groceries 0Y\* 38.871567 -76.990783

5.83037E+17 222375221 @new\_dynasty Kill My Power Haven't Went Out. When it Storms Though But Just in Case Though 0Y\* 38.854886 -76.968003

5.83069E+17 16423480 We missed the photo of the day yesterday. So here is photo one of two. #DaddysGnt #steepyheads <http://t.co/4PpvF283mz> 38.860239 -77.055002

5.82981E+17 1172892470 I got a job moe!!!!!!! 38.892318 -76.94891

5.83121E+17 348039701 It's already April 1st. Time speeds up every year. 38.945433 -76.980031

5.83048E+17 24093612 #HotOnEvent @ Warner Theatre <http://t.co/h0alSRkTxw> 38.896488 -77.029527

5.82938E+17 1554354618 LOOK AT THAT GLOO ON ME 0Y\* 38.853398 -76.984886

5.85532E+17 1167526610 Just saw nick walking around ballston i am so happy today rocks 38.879423 -77.103963

5.85532E+17 2428861079 My son be like mama i love ya 0Y\* 0Y\* 38.93693 -77.110995

5.85538E+17 373626937 Damn this is so tight salute to @K0sevendence. One of the realist in the game 0Y\* <http://t.co/RJMor7i6k> 38.95272 -76.945442

5.85538E+17 17354416 kkae@thegrance: HBOS @veepb0 should win some kind of a swag award for this. [@OfficialJD @veepb0 @Buzbush](http://t.co/1edD83QkKsgreed) 38.917002 -77.096221

5.85538E+17 92440223 I want to write. However I really don't want to blog help. 38.902336 -77.034809

5.85538E+17 263796484 LittleItalyNewYorkNewYorkCityNewYorkNewYorkEENY claudlavacac @ amayam atos @ Little Italy New York <http://t.co/d3kPQhAdT> 38.901485 -77.026086

5.85543E+17 1493040350 my trap down the street 38.849868 -76.982489

5.85543E+17 182739066 New Event: Magical Ecuador. Boleros @ Gunston Arts Center (Arlington VA) <http://t.co/vANz22hvC2> 38.847742 -77.068625

5.85543E+17 182739066 New Event: Magical Ecuador. La Escoba (The Broom) @ Gunston Arts Center (Arlington VA) <http://t.co/xxmVnVM6mC> 38.847742 -77.068625

5.85548E+17 1431857568 @SolomonMiles Sure I'd be happy to learn more about it. Do tell... 38.902529 -77.03975

5.85548E+17 14397600 Oh DC you are looking gorgeous today! Your spring colors are lovely. #cherryblossom #washingtndc <http://t.co/kv1e95p93> 38.890126 -77.086872

5.85544E+17	72239013	@rll_alysiaC that's my biggest #pitpeeve. What social sorority were you in? P.S. Happy Birthday!	38.090126	-77.086872
5.85544E+17	35949918	It's ok to be completely jealous of my niece's wardrobe. @mathalaign itsaminthehty #babyfadonk! https://t.co/UTUvqDvgat	38.916303	-77.036879
5.85545E+17	582198153	Straight off the rack from the World Bank bookstore's sidewalk sale... For \$31 an intriguing! http://t.co/1T1rGj2CS	38.899016	-77.041791
5.85989E+17	44672726	@JDunnah did someone say...dark knight http://t.co/VNq5ObK9Qp	38.843704	-77.111205
5.85989E+17	20359460	@kex@vintagedoll : @tokyoDiamond bitch stop lying @948567 bitch you saw it @94	38.897335	-76.976439
5.85989E+17	224818894	you catch a nigga in a lie & they dead lie about lying. @948567 bitch you saw it @94	38.920987	-77.017691
5.85991E+17	61699674	Charo @ know she will get her ass whooped that's why she gets security backs down like a damn puppy #RBDivastLA	38.991155	-77.030183
5.85991E+17	174377381	No this Tobin dude didn't just call my Nelson a @kexpat @kex. Watch your mouth @CSiCyber	38.845433	-76.978716
5.85991E+17	15149829	I want to get home! Please driver it can't be this hard...	38.854959	-77.04440
5.85991E+17	2511325591	I wish I would have went to go play ball. but homework was like FUCK NO BABY!	38.892032	-76.952254
5.85992E+17	168328184	MCW with 30 tonight	38.948903	-76.953206
5.85992E+17	169052647	I make ppl laugh	38.92527	-77.004550
5.85993E+17	2531065892	It's over	38.864783	-76.997597
5.85994E+17	331331370	I'm supposed to be studying @ Verizon Center https://t.co/W3Dg80OV6D	38.898113	-77.021114
5.85995E+17	105187267	Church of Scientology paid for Pis to follow leader's father according to police reports: http://t.co/6vmQzB8KEV	38.919534	-77.026587
5.85995E+17	335600139	Lost a real nigga & i can't get em back @aymarc	38.847586	-76.975071
5.85996E+17	129545659	Nunca mais quero ir em museu puta porra chata	38.910042	-77.045761
5.85996E+17	125929372	Howard out here making moves on the low and think nobody is hip...	38.920612	-77.024433
5.85996E+17	561039197	@Card_L_ECore like Season 1 all over again!! @vashville so good!!	38.825258	-77.081119
5.85996E+17	22239013	When the Anz and #capp games get out at the same time... #theStruggle	38.881775	-77.099661
5.86113E+17	985839554	Tay and I are now on our way to our second White House tour in the private west wing. Isn't it great to have a best friends with connections	38.901072	-77.039902
5.86113E+17	2491363370	I bailed on my other plans tonight... time to see @GraveMusic and fam kill it	38.942741	-77.02182
5.86113E+17	180145104	on my way hm from my coin	38.925255	-77.105722
5.86113E+17	472588042	@OnThisJourney71 nigga yes that was another level her nigga was funny looking too and spoke bad English @94	38.898462	-77.028208
5.86117E+17	90440931	Mind your business @94 https://t.co/HfItLWys6x	38.853058	-76.983150
5.86117E+17	1474402608	The songs on the radio are okay but my taste in music is your face	38.892078	-77.049741
5.86118E+17	2580359229	She drivin me crazy wit these braids	38.831764	-77.004332

Detailed excel file of random sample of Tweets from Washington DC

## Appendix III

```
. spearman distancscore positive_e negative_e overall_score
(obs=185)
```

	distancscore	positive_e	negative_e	overall_score
distancscore	1.0000			
positive_e	0.0019	1.0000		
negative_e	-0.0396	0.0478	1.0000	
overall_score	-0.0271	0.8305	0.5629	1.0000

```
. spearman distancscore positive_e
```

```
Number of obs = 185
Spearman's rho = 0.0019
```

```
Test of Ho: distancscore and positive_e are independent
Prob > |t| = 0.9797
```

```
. spearman distancscore negative_e
```

```
Number of obs = 185
Spearman's rho = -0.0396
```

```
Test of Ho: distancscore and negative_e are independent
Prob > |t| = 0.5923
```

```
. spearman distancscore overall_score
```

```
Number of obs = 185
Spearman's rho = -0.0271
```

```
Test of Ho: distancscore and overall_score are independent
Prob > |t| = 0.7141
```

```
. spearman negative_e positive_e
```

```
Number of obs = 185
Spearman's rho = 0.0478
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.5181
```

```
. spearman overall_score positive_e
```

```
Number of obs = 185
Spearman's rho = 0.8305
```

```
Test of Ho: overall_score and positive_e are independent
Prob > |t| = 0.0000
```

```
. spearman overall_score negative_e
```

```
Number of obs = 185
Spearman's rho = 0.5629
```

```
Test of Ho: overall_score and negative_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and hilliness*

```
. spearman distancscore positive_e negative_e overall_score
(obs=67)
```

	distancscore	positive_e	negative_e	overall_score
distancscore	1.0000			
positive_e	-0.0649	1.0000		
negative_e	-0.1716	0.0021	1.0000	
overall_score	-0.1489	0.8041	0.5689	1.0000

```
. spearman distancscore positive_e
```

```
Number of obs = 67
Spearman's rho = -0.0649
```

```
Test of Ho: distancscore and positive_e are independent
Prob > |t| = 0.6020
```

```
. spearman distancscore negative_e
```

```
Number of obs = 67
Spearman's rho = -0.1716
```

```
Test of Ho: distancscore and negative_e are independent
Prob > |t| = 0.1651
```

```
. spearman distancscore overall_score
```

```
Number of obs = 67
Spearman's rho = -0.1489
```

```
Test of Ho: distancscore and overall_score are independent
Prob > |t| = 0.2291
```

```
. spearman negative_e positive_e
```

```
Number of obs = 67
Spearman's rho = 0.0021
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.9868
```

```
. spearman overall_score positive_e
```

```
Number of obs = 67
Spearman's rho = 0.8041
```

```
Test of Ho: overall_score and positive_e are independent
Prob > |t| = 0.0000
```

```
. spearman overall_score negative_e
```

```
Number of obs = 67
Spearman's rho = 0.5689
```

```
Test of Ho: overall_score and negative_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and distance to bike trails*

```
. spearman distancscore positive_e negative_e overall_score
(obs=266)
```

	distancscore	positive_e	negative_e	overall_score
distancscore	1.0000			
positive_e	-0.1363	1.0000		
negative_e	0.0772	0.0279	1.0000	
overall_score	-0.0778	0.8141	0.5728	1.0000

```
. spearman distancscore positive_e
```

```
Number of obs = 266
Spearman's rho = -0.1363
```

```
Test of Ho: distancscore and positive_e are independent
Prob > |t| = 0.0262
```

```
. spearman distancscore negative_e
```

```
Number of obs = 266
Spearman's rho = 0.0772
```

```
Test of Ho: distancscore and negative_e are independent
Prob > |t| = 0.2096
```

```
. spearman distancscore overall_score
```

```
Number of obs = 266
Spearman's rho = -0.0778
```

```
Test of Ho: distancscore and overall_score are independent
Prob > |t| = 0.2060
```

```
. spearman negative_e positive_e
```

```
Number of obs = 266
Spearman's rho = 0.0279
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.6511
```

```
. spearman overall_score positive_e
```

```
Number of obs = 266
Spearman's rho = 0.8141
```

```
Test of Ho: overall_score and positive_e are independent
Prob > |t| = 0.0000
```

```
. spearman overall_score negative_e
```

```
Number of obs = 266
Spearman's rho = 0.5728
```

```
Test of Ho: overall_score and negative_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and distance to Capital Bikeshare locations*

```
. spearman distancscore positive_e negative_e overall_score
(obs=185)
```

	distan~e	positi~e	negati~e	overal~e
distancscore	1.0000			
positive_e	0.0019	1.0000		
negative_e	-0.0396	0.0478	1.0000	
overall_score	-0.0271	0.8305	0.5629	1.0000

```
. spearman distancscore positive_e
```

```
Number of obs = 185
Spearman's rho = 0.0019
```

```
Test of Ho: distancscore and positive_e are independent
Prob > |t| = 0.9797
```

```
. spearman distancscore negative_e
```

```
Number of obs = 185
Spearman's rho = -0.0396
```

```
Test of Ho: distancscore and negative_e are independent
Prob > |t| = 0.5923
```

```
. spearman distancscore overall_score
```

```
Number of obs = 185
Spearman's rho = -0.0271
```

```
Test of Ho: distancscore and overall_score are independent
Prob > |t| = 0.7141
```

```
. spearman negative_e positive_e
```

```
Number of obs = 185
Spearman's rho = 0.0478
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.5181
```

```
. spearman overall_score positive_e
```

```
Number of obs = 185
Spearman's rho = 0.8305
```

```
Test of Ho: overall_score and positive_e are independent
Prob > |t| = 0.0000
```

```
. spearman overall_score negative_e
```

```
Number of obs = 185
Spearman's rho = 0.5629
```

```
Test of Ho: overall_score and negative_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and distance to signed bike routes*

```
. spearman value positive_e negative_e overall_score
(obs=353)
```

	value	positive_e	negative_e	overall_score
value	1.0000			
positive_e	-0.0053	1.0000		
negative_e	-0.0118	-0.0288	1.0000	
overall_score	-0.0100	0.8086	0.5334	1.0000

```
. spearman value positive_e
```

```
Number of obs = 353
Spearman's rho = -0.0053
```

```
Test of Ho: value and positive_e are independent
Prob > |t| = 0.9211
```

```
. spearman value negative_e
```

```
Number of obs = 353
Spearman's rho = -0.0118
```

```
Test of Ho: value and negative_e are independent
Prob > |t| = 0.8246
```

```
. spearman value overall_score
```

```
Number of obs = 353
Spearman's rho = -0.0100
```

```
Test of Ho: value and overall_score are independent
Prob > |t| = 0.8515
```

```
. spearman negative_e positive_e
```

```
Number of obs = 353
Spearman's rho = -0.0288
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.5895
```

```
. spearman overall_score positive_e
```

```
Number of obs = 353
Spearman's rho = 0.8086
```

```
Test of Ho: overall_score and positive_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and intersection density*

```
. spearman value positive_e negative_e overall_score
(obs=352)
```

	value	positive_e	negative_e	overall_score
value	1.0000			
positive_e	-0.0081	1.0000		
negative_e	0.0285	-0.0302	1.0000	
overall_score	0.0009	0.8075	0.5338	1.0000

```
. spearman value positive_e
```

```
Number of obs = 352
Spearman's rho = -0.0081
```

```
Test of Ho: value and positive_e are independent
Prob > |t| = 0.8798
```

```
. spearman value negative_e
```

```
Number of obs = 352
Spearman's rho = 0.0285
```

```
Test of Ho: value and negative_e are independent
Prob > |t| = 0.5940
```

```
. spearman value overall_score
```

```
Number of obs = 352
Spearman's rho = 0.0009
```

```
Test of Ho: value and overall_score are independent
Prob > |t| = 0.9868
```

```
. spearman negative_e positive_e
```

```
Number of obs = 352
Spearman's rho = -0.0302
```

```
Test of Ho: negative_e and positive_e are independent
Prob > |t| = 0.5726
```

```
. spearman positive_e overall_score
```

```
Number of obs = 352
Spearman's rho = 0.8075
```

```
Test of Ho: positive_e and overall_score are independent
Prob > |t| = 0.0000
```

```
. spearman overall_score negative_e
```

```
Number of obs = 352
Spearman's rho = 0.5338
```

```
Test of Ho: overall_score and negative_e are independent
Prob > |t| = 0.0000
```

*Detailed results of correlations between sentiment scores and destination density*

## Appendix IV

**. regress positive\_e distancscore**

Source	SS	df	MS	Number of obs	=	266
Model	<b>4.30412802</b>	<b>1</b>	<b>4.30412802</b>	F(1, 264)	=	<b>7.60</b>
Residual	<b>149.560534</b>	<b>264</b>	<b>.566517173</b>	Prob > F	=	<b>0.0063</b>
				R-squared	=	<b>0.0280</b>
				Adj R-squared	=	<b>0.0243</b>
Total	<b>153.864662</b>	<b>265</b>	<b>.580621365</b>	Root MSE	=	<b>.75267</b>

positive_e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
distancscore	<b>-.1428979</b>	<b>.051843</b>	<b>-2.76</b>	<b>0.006</b>	<b>-.2449762 -0.0408196</b>
_cons	<b>1.824998</b>	<b>.1102966</b>	<b>16.55</b>	<b>0.000</b>	<b>1.607825 2.042171</b>

*Detailed linear regression between sentiment scores and Capital Bikeshare locations*

**. regress negative\_e value**

Source	SS	df	MS	Number of obs	=	363
Model	<b>1.62199702</b>	<b>1</b>	<b>1.62199702</b>	F(1, 361)	=	<b>4.34</b>
Residual	<b>134.829794</b>	<b>361</b>	<b>.373489733</b>	Prob > F	=	<b>0.0379</b>
				R-squared	=	<b>0.0119</b>
				Adj R-squared	=	<b>0.0091</b>
Total	<b>136.451791</b>	<b>362</b>	<b>.376938648</b>	Root MSE	=	<b>.61114</b>

negative_e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
value	<b>.0823936</b>	<b>.0395374</b>	<b>2.08</b>	<b>0.038</b>	<b>.0046411 .1601462</b>
_cons	<b>-1.421884</b>	<b>.0772288</b>	<b>-18.41</b>	<b>0.000</b>	<b>-1.573759 -1.270009</b>

*Detailed linear regression between sentiment scores and hilliness*