The Relationship of Built Environment and Weather with Bike Share – Evidence from the Pronto Bike Share System in Seattle

Wei Ding

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Urban Planning

University of Washington

2016

Committee:
Chang-Hee Christine Bae
Branden Born
Philip M. Hurvitz

Program Authorized to Offer Degree:
Urban Design and Planning

University of Washington

Abstract

The Relationship of Built Environment and Weather with Bike Share –Evidence from the Pronto Bike Share System in Seattle

Wei Ding

Chair of the Supervisory Committee:
Professor C.-H. Christine Bae
Urban Design and Planning

The purpose of this research is to identify correlations with bike station ridership for Pronto, a bike share program in Seattle. The daily number of trips and station-level trips from October 13th, 2014 to October 12th, 2015 was obtained from Pronto. Data for independent variables were from various sources. Polynomial regression models with cubic and quadratic terms are used to evaluate the relationship between weather and temporal factors on daily system-wide ridership and the effects on different types of users. Multiple linear regression models are used to investigate the effects of built environment variables on station-level ridership in 0.25-mile, 0.5-mile, and 0.75-mile scales. The models have high goodness of fit and identify a number of variables having statistically significant correlations with ridership. Temperature and wind speed are not linearly associated with daily ridership. Rain, weekends, and holidays decrease daily ridership. Different effects of weather and temporal factors on annual members and short-term pass holders are also captured in this research. Annual members, who are less affected by rainfalls than short-term pass holders, are more likely to use Pronto on weekdays while short-term pass holders tend to use Pronto on weekends and holidays. In addition, the station-level ridership is negatively associated with job density, proximity to parks,

and proximity to waterfront in all three buffers. The findings will help planners and managers to predict daily ridership, optimize bike locations, and improve bike rebalance efficiency.

Table of Contents

List of Figures

List of Tables

CHAPTER 1 INTRODUCTION

1.1 Background

With economic development and urbanization, consumption of transportation-related energy, especially non-renewable energy, keeps increasing and it will continuously increase in the future. The increasing consumption can lead to serious environmental problems, such as air pollution and climate change. In return, sustainable transportation largely based on non-motorized transportation is gaining more and more attention in transport planning field. Among different types of sustainable transportation programs, bike share, a shared use of bike approach, stands out. It is becoming a part of global urban mobility idea. There are more than 600 cities around the world that have bike share programs (Wikipedia). As a public bicycle program, it is initiated with the idea of increasing cycling, bridging the gap between first mile/last mile and a transportation network, and alleviating environmental issues (DeMaio, 2009).

Various advantages of bike share help explain the reasons why this type of travel modes is becoming popular. First, bike share helps save travel time when people want to avoid getting stuck in traffic congestion and bike share also saves travel time when people do not want to walk for a long distance. Second, when combined with transit, bike share can easily connect the gaps between transportation networks or travel destinations (Shaheen, Guzman, & Zhang, 2010). Bike-and-ride, integrating bicycling with transit, provides people with longer travel distance at a lower cost and physical difficulty. Third, bike share plays an important role in increasing health benefits. Studies show that bicycling can be a healthful exercise in daily life (Pucher & Dijkstra, 2003). People who have higher levels of cycling tend to have a lower possibility of obesity (Pucher, Buehler, Bassett, & Dannenberg, 2010). Lastly, bike share is a low-carbon travel option. Compared with automobile and public transit, bike share generates almost zero carbon emission even when the carbon emission generated by program operation are taken into consideration.

These advantages provide reasons for the government to alleviate transportation and environment issues. More and more cities have begun bike share programs. The first bike share program in the world, “White Bikes,” was released in Amsterdam in 1968 (Shaheen, Guzman, & Zhang, 2010). The prototype of this program was proposed in *The White Bicycle Plan (White Fietsenplan)*, which was announced in 1965 by the group called “Provo” which was a rebellious, radical anarchist organization pursuing the restoration of a sustainable environment by banning personal automobiles, and allowing cycling, walking and public transit only in Amsterdam ("Provo", 2016). They painted the bicycle in white for simplicity and healthy living. It was designed as free, unlocked bicycles that were always ready to ride in the city center of Amsterdam ("The whimsical anarchism of the White Bicycle revolution", 2015). As soon as Provo provided 50 White Bicycles, Dutch police confiscated all of them. People could borrow a bike for free with unlimited trip duration. However, this unrestricted rule led to serious problems, such as stealing, and the program eventually failed. Having learned from this experience, other European cities started to propose their own bike share programs with different rules and technologies to trace bike usage (DeMaio, 2009). The first bike share program in the U.S “Yellow Bikes” launched in Portland, Oregon, in 1994. Yellow Bikes was a community free bicycle as a herald of bicycle socialism, ("Bicycle Master Plan, City of Portland", 2016), but it was not well received by Portlanders.

As people could use this program for free, the program suffered from theft and damage. It eventually stopped operating in 2001. Other cities, such as Boulder, Colorado in 1995, Minneapolis and St. Paul in 1996, and Washington in 1996, joined Portland with their own bike share program (Shaheen, Guzman, & Zhang, 2010; "Encyclopedia of transportation: social science and policy", 2015). However, the bike share programs that have been revived in the 2010s are different from those in the 1990s. The bicycles are locked in a station, and users pay a fee. Nowadays, more than 50 cities in the USA have bike share programs.

Figure 1 The first white bike in Amsterdam

Source: Dangerousminds.net (2015)

Figure 2 Portland's Yellow Bike Project

Source: Oregononlive (2016)

In Seattle, Pronto is the first bike share program, launched on October 13th, 2014. It has more than 500 bikes with the initial 50 stations spread throughout central areas of Seattle. In order to maximize the ridership and ensure financial stability, the sites of stations were designed to be located based on population density, popular sightseeing points, existing transportation facilities, and technical feasibility ("Pronto", 2016). Many of the stations are located in the University District, Downtown Seattle, Capitol Hill, and South Lake Union. People can purchase an annual membership for $85 per year, or a 1-Day or 3-Day pass for $8 and $16, and can then borrow and return bikes to any open station. Pronto staff constantly redistributes bikes to ensure users will not have trouble finding or returning them. (This is called "rebalance"). Users can make unlimited trips while trip duration is under 30 minutes in order to avoid an extra charge. Because Washington has a law requiring the use of bike helmets, helmets are available for rental in any open station.

Figure 3 A Pronto bike station

Source: (Soper T., GeekWire 2014)

Figure 4 People riding Pronto

Source: City of Seattle (2014)

After one-year of operation, Pronto opened its first-year data to the public on October 14th, 2015.The dataset includes trip data with bike numbers, trip start day & time, trip end day & time, trip start station, trip end stations, and rider types (annual or short-term pass holder). Annual members' trip data also includes the member's gender and year of birth.

The available data provides researchers opportunities to study Pronto users' travel behavior. To study people's travel behavior not only helps to optimize the operation system and bike rebalance, but also creates a better plan for future expansion, such as choosing a location and determining the size of bike stations. This study investigates the effects of weather and built environmental characteristics near bike share stations on daily ridership and station-level ridership.

1.2 Research objectives and hypothesis

For the purpose of this thesis, ridership is defined as the sum of trips. For instance, daily ridership means the daily number of people using Pronto. Station-level ridership means the number of people starting or ending their trips per each bike share station. When a trip starting time is after 12:00 am on a certain day, this trip is considered as a trip of that day. In order to meet the purpose and potential application of this research, this paper focuses on investigating the following questions and objectives:

1.2.1 Research questions:

1. Are weather and temporal factors associated with the daily number of people using Pronto?

If so, how are they related with the ridership? And is there a difference between the relationship of weather and temporal factors with ridership of different user types?

2. Are nearby built environment factors related to station-level ridership?

If so, how are they related to the ridership? And is there a difference between the relationships of built environment factors and station-level ridership generated by different user type? Do different spatial scales of these elements have a different impact on nearby station ridership? If so, how does it affect station-level ridership?

1.2.2 Research hypotheses:

1. Weather and temporal factors both are related to total daily ridership, annual members' ridership, and short-term pass holders' ridership

- Precipitation and wind speed are negatively associated with ridership, with less impact on annual members than short-term pass holders;

- Temperature is positively associated with ridership, while less impact on annual members than short-term pass holders;
- More trips are generated during weekdays than weekends and holidays.

2. Nearby built environment elements are associated with station-level ridership:
    - Trips tend to be generated in high-job-density and high-housing-density areas. Job density is positively associated with station-level ridership.
    - The stations that are closer to parks and waterfront have more trips. Trips generated by short-term pass holders are more negatively related to distance to parks and waterfront;
    - Nearby bus ridership has an impact on station-level ridership; more frequent transit services, more people use nearby bike share station;
    - Shorter distance to light rail stations and other bike share stations, more people use nearby bike share station.

1.3 Paper structure

This paper consists 5 sections:

1. Introduction - Introduces the history of bike share program and the background of Seattle's bike share program, as well as the needs to study Seattle's bike share program.
2. Literature review - Identifies the relationship between weather, temporal factors and built environment, and bike share ridership, and then summarizes the limitations of existing literature.
3. Data and Methodology - Outlines the dataset and data exploration and introduces the

regression model used for this study.

4. Findings - Presents the analysis results.

5. Conclusion and Limitations - Discusses the results, recommendations and limitations.

CHAPTER 2 LITERATURE REVIEW

Research on bikes and bike share programs is increasing. Some of it has an emphasis on descriptive analysis, such as the benefit of bike share, policy study, and user impact (Shaheen, Guzman, & Zhang, 2010; Shaheen, Cohen, & Martin, 2013). However, increasing attention has been paid to empirical studies, focusing on how various factors, such as weather, built environment, and transportation infrastructure impact bicycling and bike share ridership. Based on the analysis results of how weather, built environment, socio-demography, and transportation infrastructure affect bike usage or bike share ridership in different cities and in different scales, research can provide planning suggestions and policy to encourage cycling and improve the performance of bike share programs in urban areas. This chapter consists of three parts. First, it introduces the existing research on how weather and temporal elements associate with bike and bike share usage, as well as trip duration. Second, it states the existing studies on investigating the relationship between built environment, transportation infrastructure and socio-demographic elements, and bike usage and station-level bike share ridership. The last part summarizes the limitations of existing research.

2.1 Weather and temporal factors affecting bicycling

Weather is a theme that has been taken into consideration in much cycling-related research. It has been found that short-term weather conditions and long-term (seasonal) climate change are associated with cycling (Nankervis, 1999). Miranda-Moreno and Nosal look at the impact of weather on cycling in Montreal, Canada by using log-linear model and Poisson regression, and notice that temperature has non-linear impact on ridership-- when temperature is higher than 28˚C and humidity is greater than 60%, they have a negative impact on ridership. Otherwise, they are positively associated with ridership. Precipitation is another important factor. Bicycle volume is affected by not only the presence of rain in the same hour, but also in the previous 3 hours or in the morning (Miranda-Moreno & Nosal, 2011). Nosal and Miranda-Moreno also investigate the impact of weather on bicycling in Montreal, Ottawa, Vancouver, and Portland. Using regression model, they identify that temperature and humidity have non-linear associations with cycling, while precipitation has a dramatic negative impact on cycling flows (Nosal & Miranda-Moreno, 2014). Sears et al. study how seasonal factors and weather affect bicycle commuting in Vermont, USA by using a generalized linear regression model. Temperature, precipitation, and wind speed are all significantly associated with the likelihood of bicycling – temperature and wind speed have positive and negative impacts on the likelihood, and the rain absence in the morning doubles the likelihood of biking (Sears, Flynn, Aultman-Hall, & Dana, 2012). Saneinejad et al use multinomial logit modeling to investigate the impact of weather on not only bicycling, but also four other active transport modes, including auto driver, auto passenger, transit, and walking. The interaction has also been conducted to compare demography and weather. The results support the finding from other previous research (Saneinejad, Roorda, & Kennedy, 2012). More than focusing on investigating the relationship between bike share ridership and weather, Gallop et al. create a seasonal autoregressive model to predict bike traffic. Temperature, humidity, wind speed, precipitation, and fog are found to be significantly related to bicycling, and are used for predicting bike usage levels (Gallop, Tse, & Zhao, 2011).

From the previous studies, temperature and humidity show a significant positive or non-linear impact on bicycling, while wind speed has a negative impact on ridership. Precipitation usually has a negative impact on ridership, while the presence of rain before bicycling also matters. These results are not only widely found in cycling-related research, but also in bike share related programs. Gebhart and Noland find that cold temperature, rain, and excessive heat can reduce the likelihood of using bike share programs as well as trip duration by using negative binomial regression for the bike share program in Washington, DC (Gebhart & Noland, 2014). Corcoran et al. explore the impact of site-specific weather conditions on the public bicycle sharing program in Brisbane, Australia. Temperature, wind speed, and precipitation, as well as calendar events, are included as independent variables in this research. By using Poisson regression model, the results show that stronger winds and precipitation have a significant negative influence on total number of trips. No significant relationship is found between temperature and total number of trips, though after control of other variables, a small positive impact of temperature has been shown (Corcoran, Li, Rohde, Charles-Edwards, & Mateo-Babiano, 2014).

2.2 Built environment attributes affecting bicycling

Research on built environment and travel behavior are widely conducted in the planning field. Sufficient studies show that built environment is an important factor that affects physical activity and human behavior. For instance, in a neighborhood scale, Cervero identifies that in dense and mixed-use neighborhoods, the percentage of walk or bike travel to work is higher than other types of neighborhoods (Cervero, 1996). Demographics, socio-economic attributes, and neighborhood characteristics are significantly associated with trip generation (Kockelman, 1997). Handy et al. investigate the relationship between neighborhood characteristics and travel behavior while also taking travel preference and neighborhood preference into account by using quasi-longitudinal study design. The results support the assumption that the built environment has a causal effect on travel behavior (Handy, Cao, & Mokhtarian, 2005).

Built environment features are measured at different scales of geography. Research focuses either on a predefined geographic unit, such as a neighborhood, city, or regional scale, or a custom unit defined by the researcher, such as a buffer around a location (Handy, Boarnet, Ewing, & Killingsworth, 2002). For instance, Buehler and Pucher focus on bike-related transport facilities in a city unit and analyze the influence of bike paths and lanes on bike commuting in large American cities. The results showed that cities with more bike paths and lanes had higher bike commute rates (Buehler & Pucher, 2011).

Other studies focus on built environment within a specific buffer size. For instance, Moudon et al. use 240-meter buffers to summarize the built environment characteristics near respondents' household locations. They find that the likelihood of cycling tends to increase in locations with presence of offices, clinics/hospitals, fast food restaurants, and regional trails. However, in general, the impact of the neighborhood environment on cycling is limited (Moudon et al., 2005). Similar research has also been conducted to investigate the relationship between built environment and bike share. Imani et al. focus on the bicycle flows at the station level of the bike share program BIXI in Montreal, Canada. They use a 250-meter buffer to compute the number of commercial enterprises, the presence of other bike share stations, and the presence of transit near bike share stations, as well as the distance to the central business district as a part of built environment variables. The results show that the presence of bike paths and lanes near a station increases bike usage, and a shorter distance to CBD and nearby transit stops encourages more bike trips (Faghih-Imani, Eluru, El-Geneidy, Rabbat, & Haq, 2014). Focusing on proximity of nearby businesses and jobs to stations, Wang et al. create 400-meter buffers for each bike share stations for the Nice Ride Minnesota program, then measure built environment and transportation infrastructure features within the buffers. The results indicate that the distance to water, CBDs, and parks have negative impact on daily trips, while the presence of trails has positive effect on trips (Wang, Lindsey, Schoner, & Harrison, 2016). By aggregating the data to the 400-meter buffer surrounding each station, Rixey investigates the impacts of built environment and socio-demographic characteristics near bike share stations in three systems in Washington D.C, Denver, and Minneapolis by using multivariate linear regression. He finds that socio-demographic variables, including population density, median income, education, non-while population, and retail job density are statistically significant. Moreover, bike, walk, and transit commuters, presence of bikeways, and proximity to a bike sharing stations are also found significant (Rixey, 2013). El-Assi et al. calculate the number of neighboring stations within a 200-meter buffer as a variable to analyze the effects of built environment on Toronto's bike share program. Bicycle infrastructure and number of intersections with a major road, as well as distance to other bike stations are associated with bike share trips. (El-Assi, Salah Mahmoud, & Nurul Habib, 2015).

2.3 Limitations of prior research

The following section identifies some limitations of existing literature from the perspectives of methodology and case study.

- Ignorance of different user types' travel behavior. Most existing research does not distinguish the number of trips made by annual members from trips of short-term pass holders. It ignores the possibility that different types of users can have different travel behavior. As an improvement, this paper investigates the different impact of weather, built environment, temporal, and other variables on the number of trips generated by different types of users.
- Different scales to measure built environment. Prior research has measured built environment in various ways, but lacks investigation on the relationship between built environment and bike share ridership across different spatial scales. This paper measures built environment and transportation elements in different scales and compares the different results.
- Special features of Seattle. From previous literature, the relationship between weather and bike share ridership is clear. However, no studies have looked at the impact of weather on Seattle's ridership. Considering the special weather and climate conditions in Seattle, the relationship between weather and travel behavior may be different from other locations.
- Spatial relationship. Most researchers use linear regression models to examine the significant level of built environment elements, which means that each observation is considered independently and there is no difference between them spatially. Therefore, it is assumed that if there is no difference between built environment

elements in different locations, then there is no difference in the magnitude of the effect of built environment on nearby bike share stations.

CHAPTER 3 METHODOLOGY

To investigate the factors affecting bike share ridership, weather information, built environment data, and bike ridership are collected to create a complete dataset. This section first describes the study area and data collection, then summarizes hourly, daily, and monthly ridership, as well as weather information and built environment attributes. This section also includes model development and discusses model choice.

3.1 Study area and data collection

This paper focuses on Seattle's bike share program with 54 stations and more than 500 bikes. Most of them are located in the downtown area, South Lake Union, Capitol Hill, and University District. Figure 5 demonstrates bike station location in these main Seattle neighborhoods. The neighborhood boundaries data, including Capitol Hill & First Hill Urban Center, Downtown Urban Center, University District, and South Lake Union Planning Organization Focus Area, are from the City of Seattle ("City of Seattle GIS Data", 2016).

Bike stations are mostly scattered evenly within the four neighborhood boundaries. Ten of the 54 bike stations are located in University District as well as University of Washington campus; 7 stations are located near Lake Union, with 5 being located within the boundary of the South Lake Union neighborhood. Eighteen stations, over 33% of all stations, are located in downtown area, and 12 of the 54 stations are in the Capitol Hill and First Hill neighborhoods. There are 4 stations which do not fall into any neighborhood boundaries, with one located in Lower Queen Anne, two located in East Lake Union, and one near Seattle Children's Hospital.

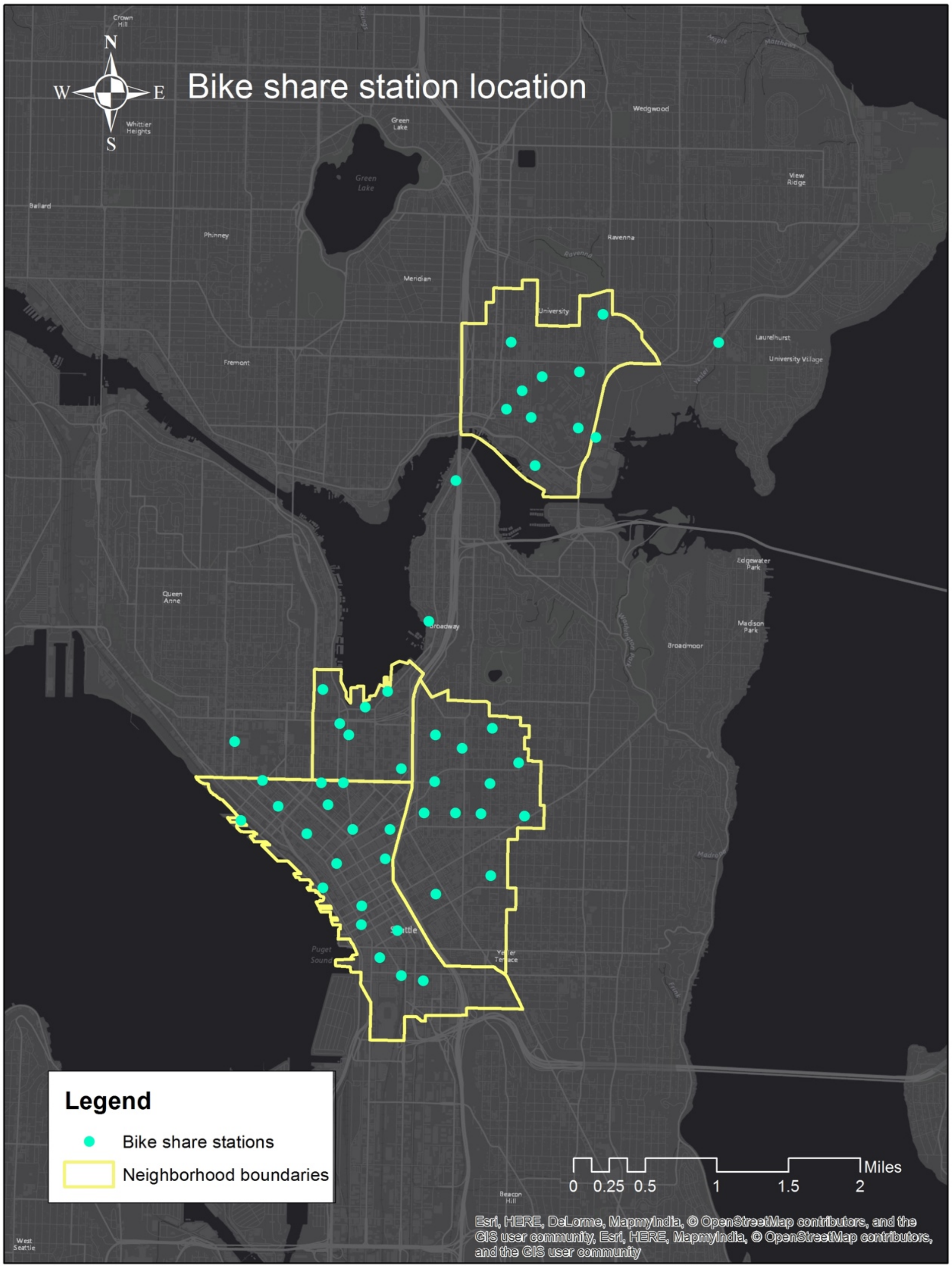


Figure 5 Pronto bike share location

The data used in this research comes from various sources, including Pronto, King County, the City of Seattle, and the University of Washington Urban Form Lab. The data descriptions are listed below:

- Pronto operation data from October 13th 2014 to October 12th 2015. The entire data set consists of 4 elements, including station status data, trip data, weather data, and station data. 1) Station status data include the information about the available bikes and docks at 50 of all the stations. Most status data were collected every 1 second from October 15th 2014 to October 30th 2015, and a small part of the data were collected every 10 seconds. 2) Trip data, one of the most important data elements in this research, contains a unique trip ID column, trip starting time, ending time, bike ID, trip duration, and the station information about where user checked out and returned the bike. If a user is an annual member, the trip data includes his/her gender and birth year information as well. If a user is a short-term pass holder, the data of gender and age are left as blank. 3) Weather data is another important data set in this research. It includes the daily maximum, average, and minimum information about temperature, dew point, humidity, sea level, visibility, wind speed, and precipitation. It contains the summary of daily weather conditions as well, such as rain, fog, snow, and thunderstorm. 4) Station data contain longitude, latitude, the number of docks, station name, station ID, and the opening date.
- King county block group census data. 1) Socio-demographic data include population, gender, race, and median household income from the American Community Surveys 2010-2014 (5-year estimates).

- King county parcel, employment density, and transit service shapefiles are provided by the Urban Form Lab of the Urban Planning Department, University of Washington. These layers are used to calculate the job density, the number of bus boardings and alightings, and housing units in proximity to bike share stations.
- The City of Seattle provides the park, waterfront, and light rail station shapefile. ArcGIS was used to measure distances to parks, waterfront, and light rail stations near the bike share station which are calculated and utilized in this research.

3.2 Descriptive Statistics

3.2.1 Bike trip summary

Daily ridership for the first year ranges from 34 to 941 times per day, with an average of 391 and a standard deviation of 164.1. Table 1 shows the annual ridership and ridership by user type and gender. The total number of trips among the 54 stations from October 13th, 2014 to October 12th, 2015 is 142,846. Annual members generated 61% of all trips, with 21% by females, 77% by males, and 2% by unspecified. On the other hand, only 39% of the total count was generated by short-term pass holders.

Table 1 Summary of the trip by user type and gender

	Total trip (%)	Female (%)	Male (%)	Unspecified (%)
Total trips	142,846 (100)	N/A	N/A	N/A
Trips by annual members	87,360 (61)	18,245 (21)	67,608 (77)	1,507 (2)
Trips by short-term pass holders	55,486 (39)	N/A	N/A	N/A

3.2.2 Trips by location

Figure 6 plots the station-level ridership by origin and destination by using the circle size to present the magnitude of ridership. Many trips start from the Capitol Hill area. Downtown and South Lake Union have high counts of both trip origins and destinations. Compared with the station-level ridership in other areas, the overall ridership in University District is relatively low for both origins and destinations.

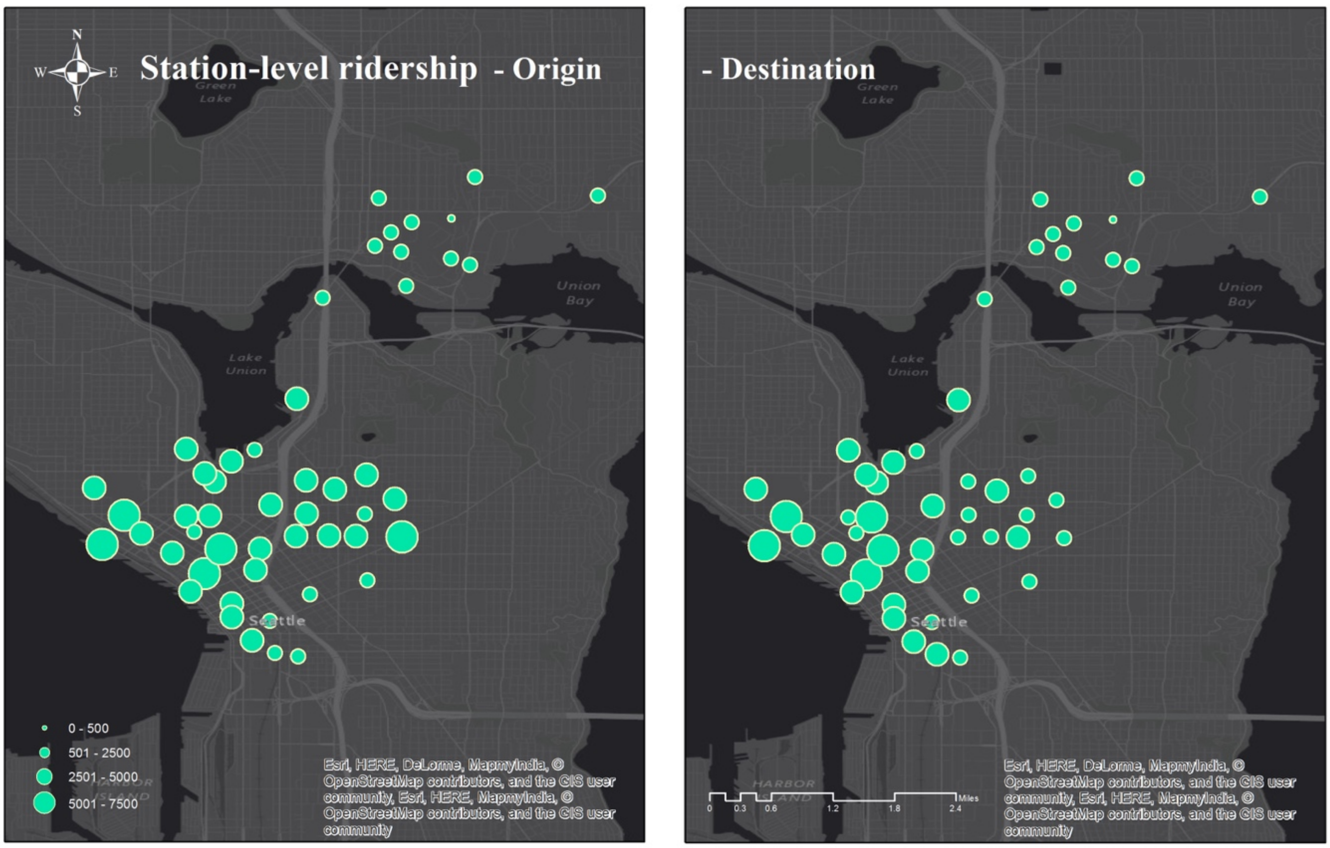


Figure 6 Station-level ridership by origin and destination

Table 2 presents the top 10 stations with greatest counts of origins and destinations. Four are in the downtown area, three are in Capitol Hill, and another three are in South Lake Union.

Among the ten destinations with greatest the counts, downtown includes 6, with the remaining 4 being in Capitol Hill. None of the stations in the University District is on the list.

Table 2 Top 10 popular station by origin and destination

ORIGIN			DESTINATION		
Station Id	Ridership	Station Location	Station Id	Ridership	Station Location
WF-01	3973	Downtown	WF-01	4166	Downtown
BT-01	3543	Downtown	CBD-13	3798	Downtown
CBD-13	2744	Downtown	BT-01	3438	Downtown
CH-07	2570	Capitol Hill	SLU-15	2659	South Lake Union
SLU-15	2478	South Lake Union	WF-04	2501	Downtown
CH-08	2471	Capitol Hill	SLU-07	2401	South Lake Union
CH-02	2168	Capitol Hill	PS-04	2330	Downtown
SLU-19	2154	South Lake Union	SLU-19	2308	South Lake Union
BT-03	1990	Downtown	CBD-06	2301	Downtown
SLU-01	1833	South Lake Union	SLU-16	2280	South Lake Union

Figure 7 shows the top 30 most popular origin and destination pairs for annual members and short-term pass holders. The patterns are very different. For short-term pass holders, the most popular routes concentrate in the downtown area, especially the stations near the waterfront. Moreover, the route from University to South Lake Union is the only route that connects the University District. Stations in the Capitol Hill area seem to be unpopular to short-term pass holders as none of the 30 trips happen in that area. For annual members, none of the popular pairs include the University District. As with short-term pass holders, downtown, especially near the waterfront, contains popular routes for annual members. However, many trips generated by annual members start or end in the Capitol Hill and South Lake Union areas.

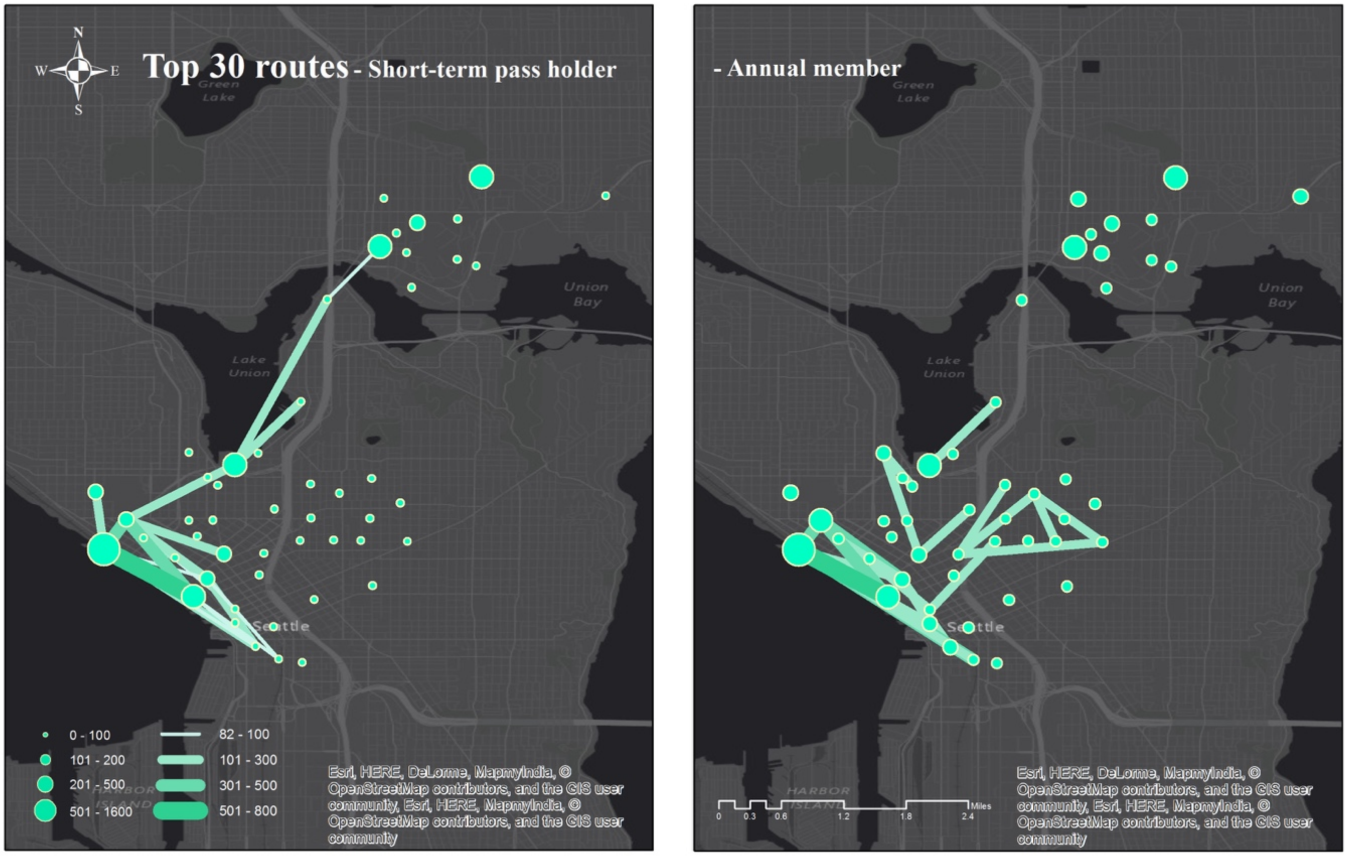


Figure 7 Popular bike stations by user type

3.3.3 Trips by month

Figure 8 plots monthly bike counts for the period of October 13th, 2014 to October 12th, 2015. It should be noted that October of both 2014 and October 2015 are not fully measured months. Ridership peaks in July 2015 and hits a low point in December 2014 at the annual scale.

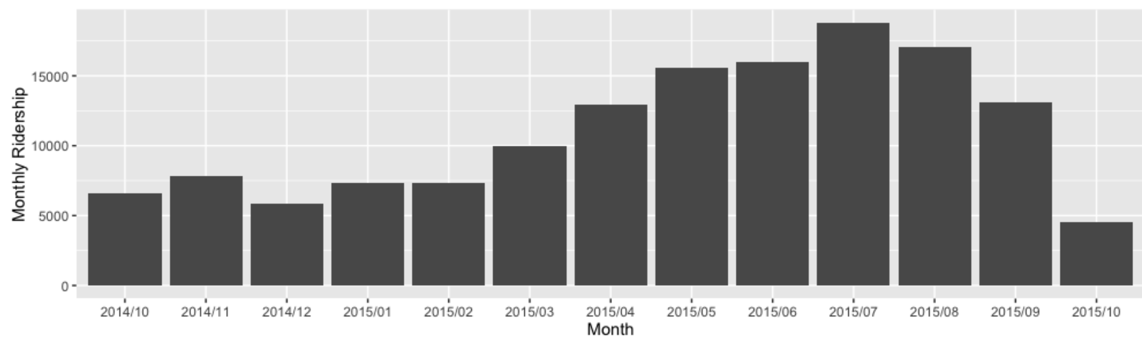


Figure 8 Monthly ridership

Figure 9 shows the monthly ridership by user type. The overall trends of both annual members and short-term pass holders fit the trend of total monthly ridership. It is obvious that the overall monthly trips generated by annual members is greater than by short-term pass holders. With an average gap of 2,452, the maximum gap of monthly ridership between annual members and short-term pass holders is 4,260 in January 2015, while the minimum is 1,266 in July 2015. The difference between ridership of annual members and short-term pass holders tends to shrink from winter to summer and it starts to increase from summer to autumn.

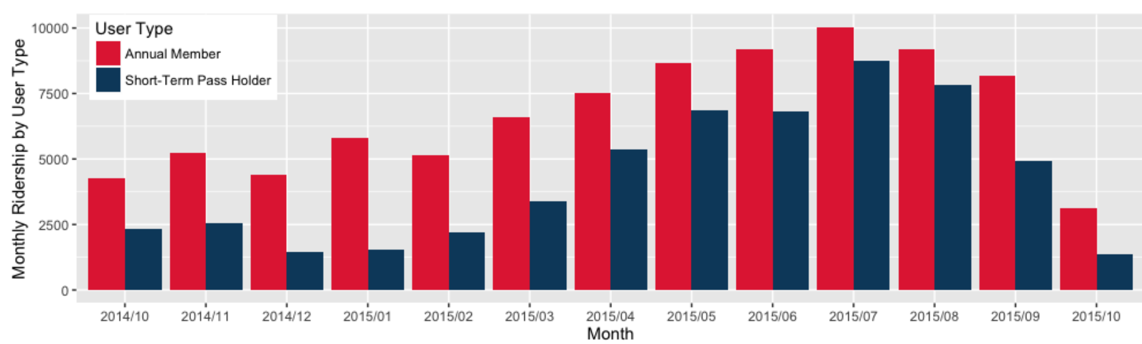


Figure 9 Monthly ridership by user type

3.2.4 Trips by day of the week

As shown in Table 3, the total number of trips made by annual members on weekdays, weekends, and national holidays is larger than short-term pass holders. Also, the average daily trip on weekdays for annual members is about 290 while short-term pass holders generated an average daily trip count of about 117. However, this situation changed on weekends and national holidays with daily trip averages of about 236 by short-term pass holders versus about 128 by annual members on weekends and 164 versus 115 on national holidays.

Table 3 Trips by day of the week and user type

	Annual member		Short-term pass holder	
Day of week	Total	Average	Total	Average
Weekday	72,944	290.6	29,352	116.9
Weekend	13,268	127.6	24,497	235.5
National holiday	1,148	114.8	1,637	163.7

3.2.5 Trips by hour

In this study, peak period is defined as being between 6:00 and 9:00 and between 15:00 and 18:00, based on King County Metro Fare ("Fares & ORCA", 2016). Figure 10 shows the total hourly ridership. During peak period, ridership is much higher than other hours. A dramatics ridership increase starts at 6:00, and hits a peaking point at 8:00. Moderate changes happen between 9:00 and 15:00. After 15:00, another increase starts and peaks at 17:00, then a noticeable decrease begins right after 17:00.

Approximately 52% of total trips are generated during peak hours while about 48% is made during off-peak hours. Trips generated during peak hours contribute to the highest two peaking points of hourly ridership.

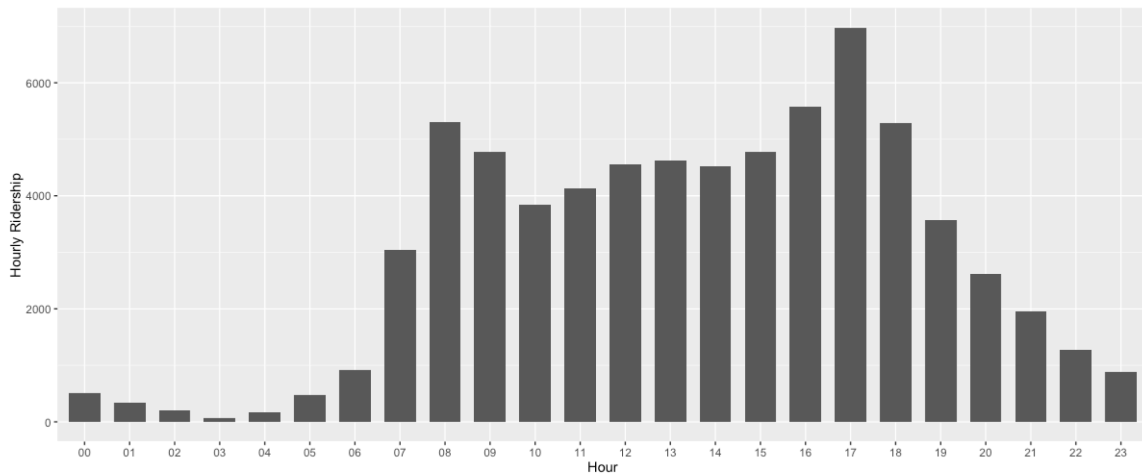


Figure 10 Hourly ridership

Figure 11 shows the accumulatively hourly ridership by user type on weekdays. It is a strong pattern that the trips generated by annual members are concentrated during peak hour -- about 60% are during peak period. Two peaks happen at 8:00 and 17:00. It is reasonable to assume that bikes are used for commuting. Unlike annual members, short-term pass holders have a relatively moderate bike usage. Only about 40% trips are made during peak hour while about 70% trips are made during 11:00 to 18:00. Ridership of short-term pass holders starts to increase gradually from 6:00 and hits the highest point at 14:00, after which it starts to gradually decrease. Figure 12 shows the hourly ridership by user type on weekends. The pattern is very different from the previous one – short-term pass holders generated two to three times more trips, especially during the day time, than annual members.

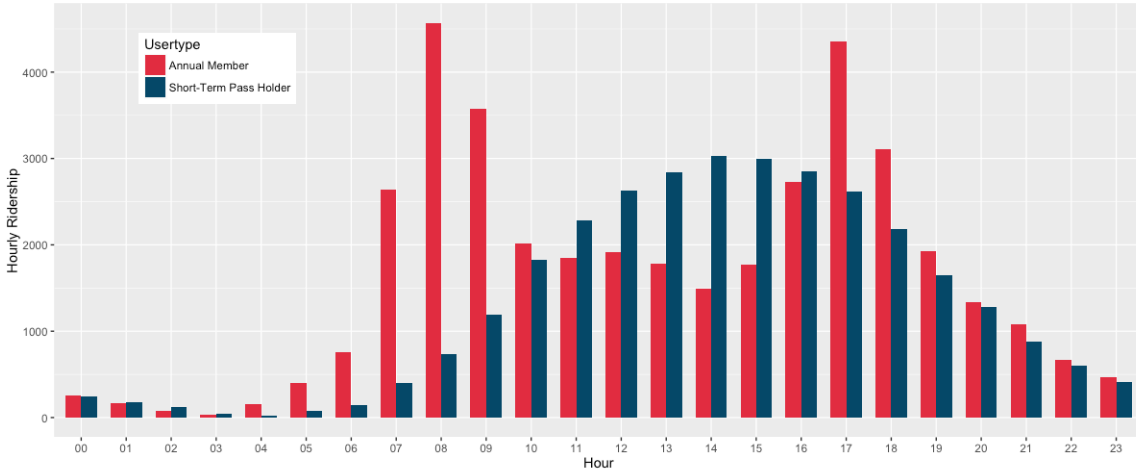


Figure 11 Hourly ridership on weekdays by user type

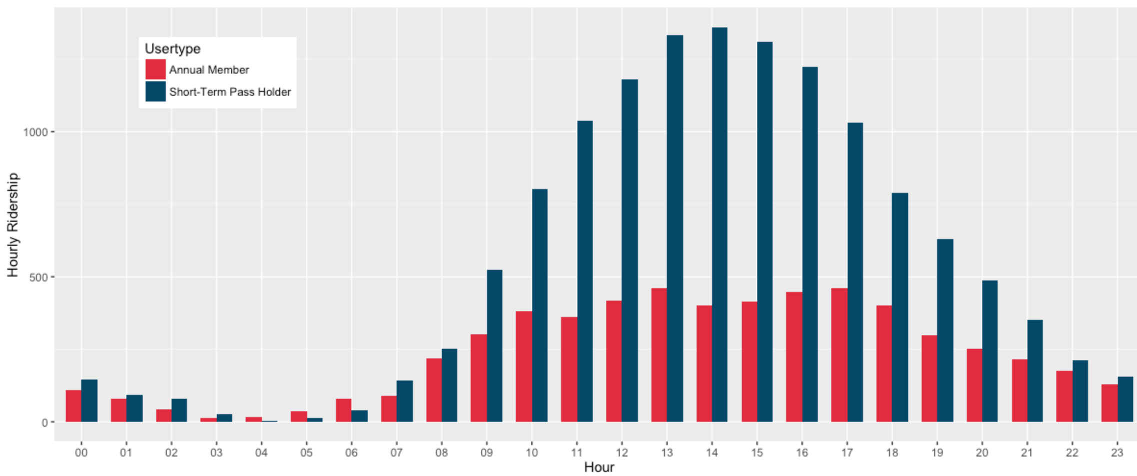


Figure 12 Hourly ridership on weekends by user type

3.2.6 Trip duration

According the bike sharing rules, both annual members and short-term pass holders can use the bike for 30 minutes without additional fees. Figure 13 shows the trip duration distribution aggregated to of 1-minute intervals by user type. Most bikes were returned within 30 minutes after trip started. More than 98% trips generated by annual members were under 30 minutes while about 70% trips made by short-term pass holders were under 30 minutes. For annual members, the average trip duration is about 9 minutes with a 5-minute peak and median of 8 minutes. However, the trip duration pattern of short-term pass holders is significantly different from pattern of annual members, with an average of 36 minutes and median of 20 minutes.

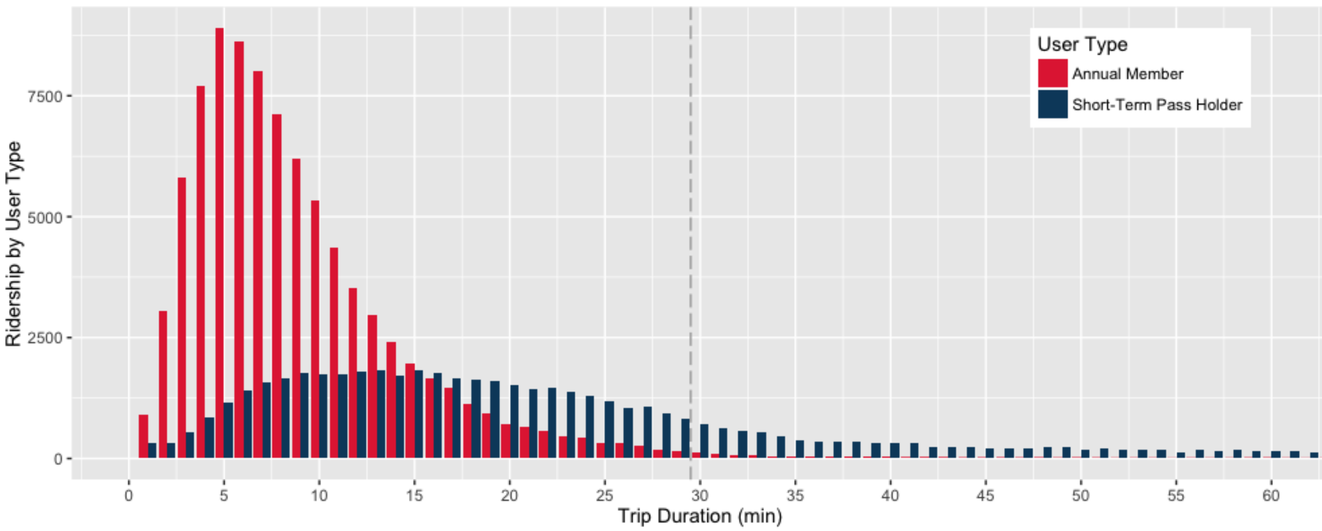


Figure 13 Trip duration by user type

3.3 Preliminary analysis

In order to better understand the relationship between weather and ridership, this section describes the preliminary analysis. For this purpose, the daily bike flows and daily weather information are considered. Moreover, this section contains the description of built environment variables.

3.3.1 Temporal variables

The large difference between the ridership on weekdays and weekends and holidays points to the importance of considering temporal factors on ridership. Two studies have identified temporal factors being significantly associated with bike ridership (Corcoran, Li, Rohde, Charles-Edwards, & Mateo-Babiano, 2014) (Gebhart & Noland, 2014). However, the findings are not consistent; compared to weekdays, the effect of the weekend is negative in one study, while another finds that it is positive. It would be interesting to examine how the effects of temporal factors on Pronto's ridership. Moreover, it would be meaningful to investigate the pattern of average ridership generated by different types of users on weekdays versus weekends and holidays. In addition, this study focuses on the effects of various factors on daily ridership; the temporal variables could play an important role.

3.3.2 Weather variables

The literature agrees that weather is an important determinant for bicycling. Temperature, wind speed, and precipitation are identified by previous research as variables that can affect bicycling ridership and bike share ridership (Miranda-Moreno & Nosal, 2011; Gebhart & Noland, 2014). In general, temperature is identified to be positively associated with bike count and bike share ridership, while wind speed and precipitation are negatively related to bike ridership. Precipitation is widely regarded to have negative effects on bike ridership (Gebhart & Noland, 2014). The following part discusses these three independent variables with more detail. Based on these findings, temperature, precipitation, and wind speed are chosen to be independent variables in this study.

3.3.2.1 Temperature

Figure 14 shows an increase in ridership that increases with temperature. Although generally monotonous, this ridership is linear. When temperature is higher than 70˚F, the increasing ratio of ridership starts to decrease. The same happens when temperature is lower than 50˚F. Figure 15 plots the average daily temperature and trip count generated by annual members and short-term pass holders. The ratio of the two fitted lines are similar. When temperature reaches approximately 72 ˚F, the ridership of annual members tends to decline while the increasing rate of short-term pass holders seems to be less impacted by high temperature.

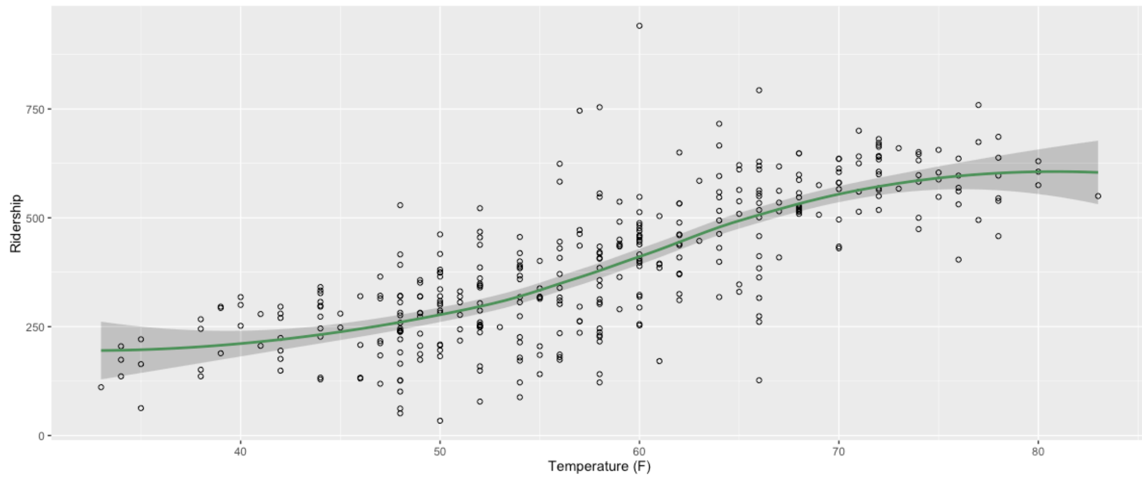


Figure 14 Average daily temperature and daily ridership

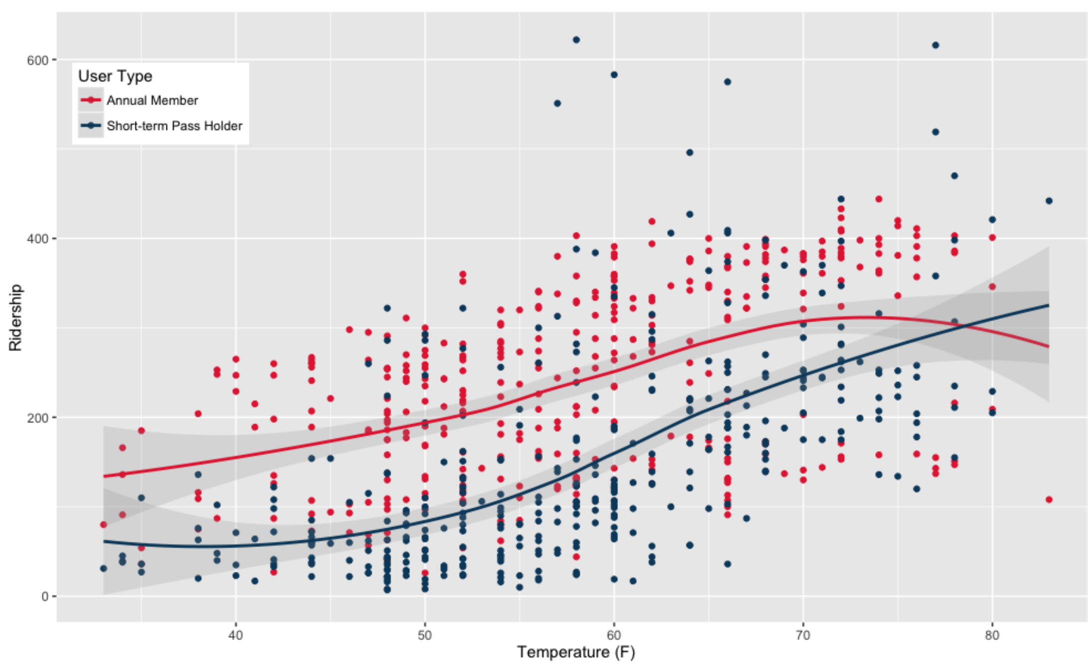


Figure 15 Average daily temperature and daily ridership by user type

3.3.2.2 Precipitation

Based on the standards created by the American Meteorological Society and National Meteorological Library, contiguous precipitation data has been transformed into categorical data for the purpose of this study. As shown in Table 4 Precipitation level standards221 days from October 13th 2014 to October 12th 2015, have no precipitation; 67 days have light precipitation; 50 days have moderate precipitation; 26 days have heavy precipitation while only 1 day has violent precipitation. That means about 60% of the days have no precipitation while 40% have less or more precipitation.

Table 4 Precipitation level standards

Precipitation Category	None	Light	Moderate	Heavy	Violent
Standard (inch)	0	0 - 0.1	0.1 - 0.3	0.3 - 2.0	Over 2.0
Frequency (days)	221	67	50	26	1

Figure 16 Boxplot of precipitation and total ridershipshows precipitation and total ridership and ridership by annual members and short-term pass holders. Both figures demonstrate similar patterns. When precipitation level gets higher – from none precipitation to violent – the ridership tends to decrease significantly.

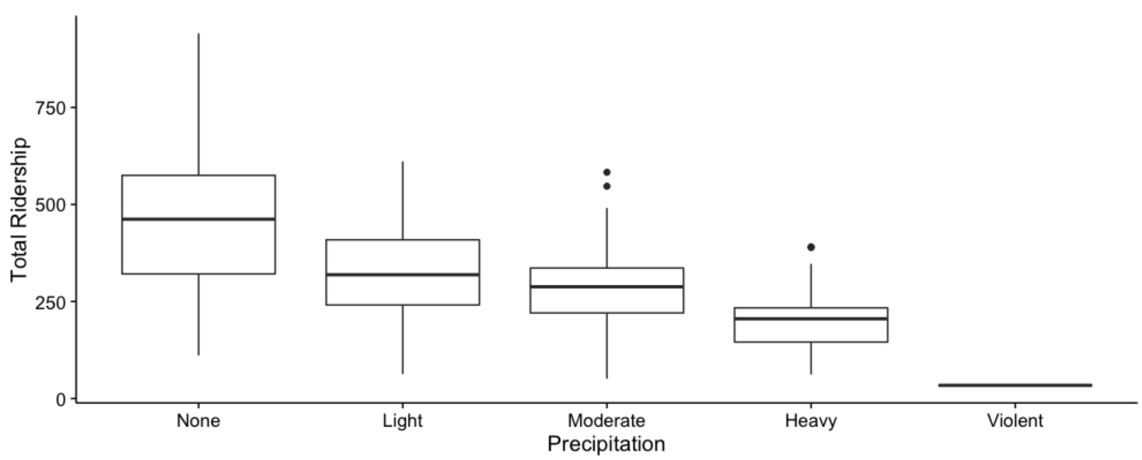


Figure 16 Boxplot of precipitation and total ridership

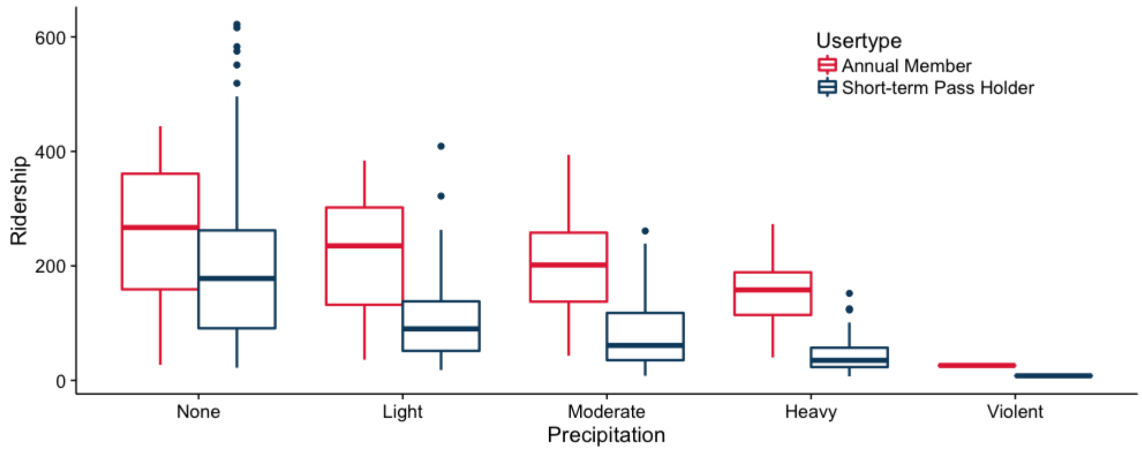


Figure 17 Boxplot of precipitation and ridership by user type

Figure 17 plots the relationship between trip duration and precipitation by user type. The average trip duration of short-term pass holders tends to decrease when precipitation increases, while annual members seem to be affected less.

3.3.2.3 Wind speed

Figure 18 shows the relationship between ridership and wind speed; at low levels, when wind speed increases, ridership tends to increase as well; when wind speed is higher than 4 mph, ridership starts to decrease substantially.

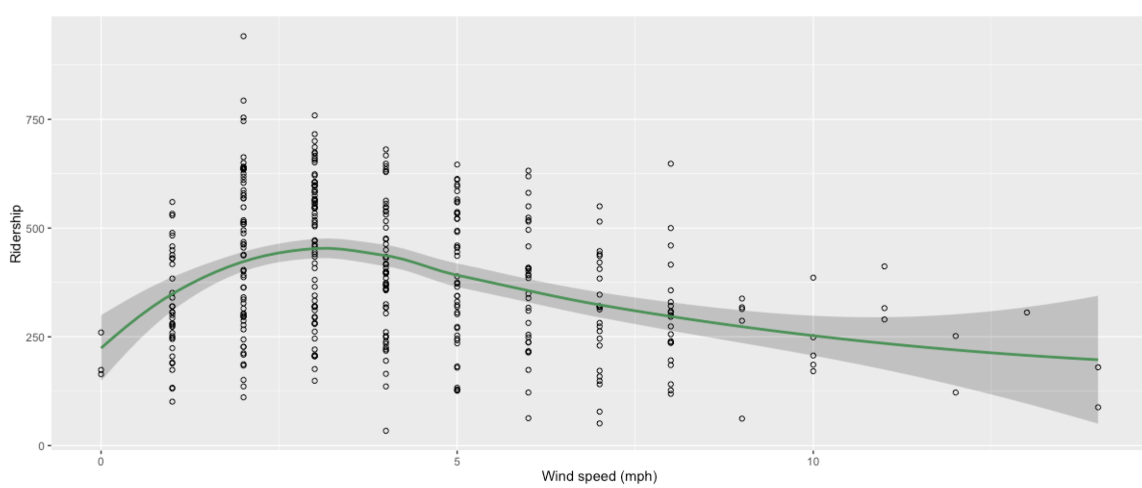


Figure 18 Wind speed and total daily ridership

Figure 19 plots wind speed and ridership by annual members and short-term pass holders. The patterns for different user types are similar –when wind speed is between 0 to 3 mph, ridership tends to increase, while when wind speed is higher than 3 mph, ridership of both annual members and short-term pass holders starts to decrease, and the ratio of short-term pass holders is slightly larger than the ratio of annual members. When wind speed is greater than 10 mph, there is a slight increase in the number of trips generated by short-term pass holders.

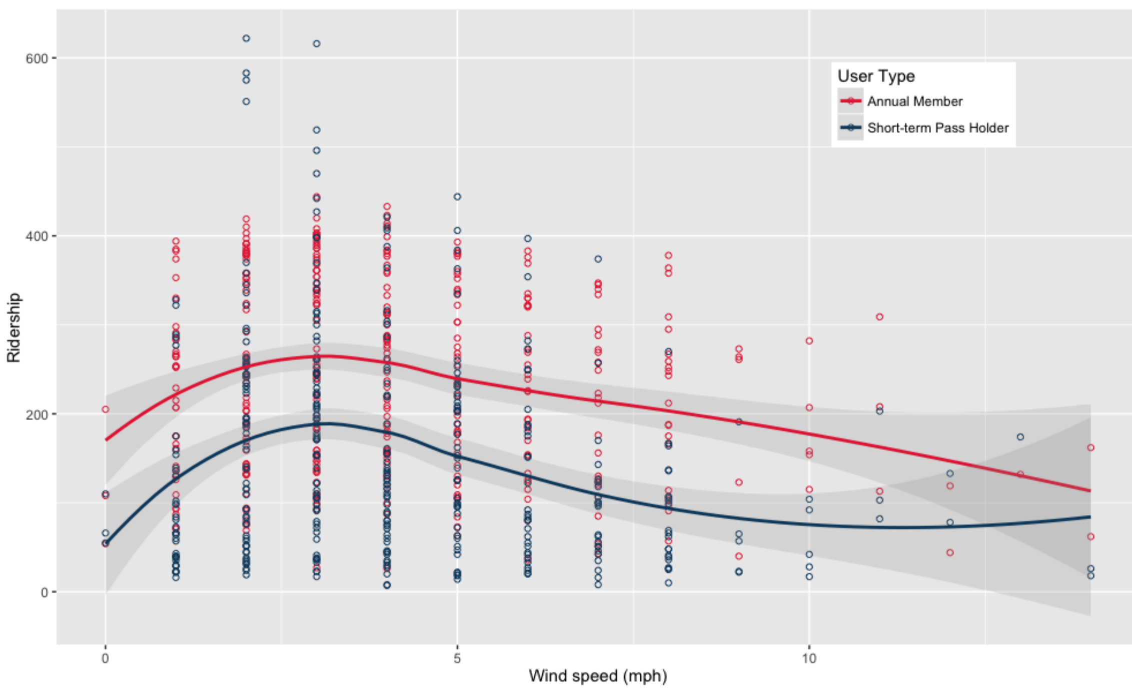


Figure 19 Wind speed and daily ridership by user type

3.3.3 Built Environment Variables

Mentioned in previous chapter, one of the limitations of the existing literature is that no consistent spatial scale is used to measure built environment, transportation, or sociodemographic variables. It is reasonable to assume the different measures could lead to different analysis results. To investigate the effects of scale, this paper uses 0.25-mile, 0.5-mile, and 0.75-mile buffers -- estimated 5-min, 10-min, and 15-min walking distances, to measure the built environment, transportation infrastructure, and socio-demographic features of each station. Figure 20 displays these redial buffers around bike stations. A separate regression model was developed for each buffer distance.

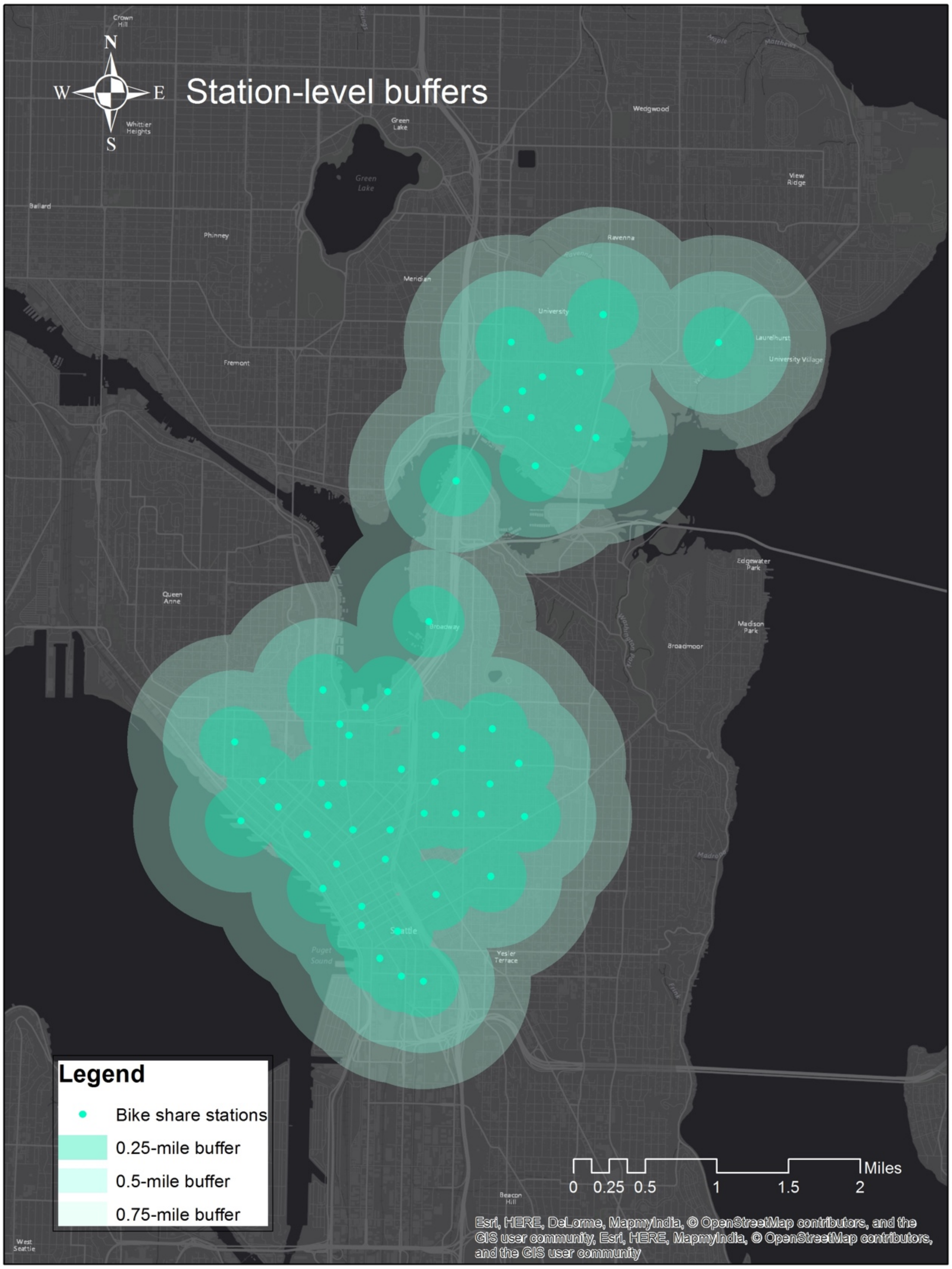


Figure 20 Station-level buffer, including 0.25-mile, 0.5-mile, and 0.75-mile

In order to ensure the consistency of the dataset, 50 of the total 54 stations are used because 4 stations launched after October 13th 2014. To measure the effects of the built environment around each station on station-level ridership, variables in proximity to various destinations are considered to be points of interest or trip generation points. The variables are the Euclidean distance of the station, in miles, to the nearest park boundary, waterfront, other bike stations, and light rail stations. It is expected that station-level ridership is positively associated with being closer to parks, waterfront, bike share stations, and light rail stations.

This paper hypothesizes that transportation infrastructure, such as bike lanes and public transit, supports station-level ridership because it increases the accessibility of a bike share station. The variables bus boarding and alighting density indicate the density of the year-round transit records of each bus stop. This variable indicates the popularity of a transit stop, and therefore shows the potential points of interest. It is expected that station-level ridership is positively associated with bus boarding and alighting. The variables length of bike lane and length of principal transit route measure the density of bike lanes and transit routes. It is expected that the station-level ridership is positively associated with these two variables.

After identifying possible independent variables, Pearson's Correlation Coefficient is conducted to measure the strength of the linear correlation between independent variables. The value of Pearson's Correlation Coefficient ranges from +1 to -1, where +1 indicates an absolute positive correlation, 0 means no correlation at all, and -1 shows an absolute negative correlation.

The standard for determining the correlation between variables varies. In this paper, 0.8 ~ 1 and -0.8 ~ -1 are used to identify strong correlation between variables, 0.8~0.5 and -0.8~ -0.5 is used as a sign of a weak correlation, and the range between 0.5 ~ -0.5 is used to identify a non-significant correlation. Table 5, Table 6, Table 7, Table 8, and Table 9 present Pearson's correlation coefficient matrix for all three buffers and the selected independent variables with more details. Population, multifamily unit density, and residential building density variables are eliminated after Pearson correlation coefficient analysis because of its strong correlation with other variables. As a measure of distribution of population, housing unit density, which is the sum of multifamily unit density and residential building density, is used as an alternative. In addition, bus boarding and alighting records are combined as one variable to present bus ridership. In order to ensure a valid model, Variation Inflation Factor (VIF) is calculated to determine whether the correlation between variables can be safely ignored in next section.

Table 5 Built environment, transportation, and socio-demographic variable summary

Variables	Description	Unit
Weekend	The number of trips generated on weekends	-
Holiday	The number of trips generated on holidays	-
Short-term pass holder	The number of trips generated by short-term pass holders	-
Origin	The number of trips starting from a certain station	-
Distance to park	The distance from a bike station to nearest park	Mile
Distance to waterfront	The distance from a bike station to nearest waterfront	Mile
Distance to nearby bike station	The distance from a bike station to nearest bike station	Mile
Distance to nearby LINK station	The distance from a bike station to nearest light rail station	Mile
Bus ridership density	The average of bus boarding and alighting records within a certain buffer	1000 time/ buffer
Bike lane length density	The sum of bike lane length within a certain buffer	Mile/buffer
Job density	Employment density in a certain buffer	1000/buffer
Housing unit density	The number of housing unit within a certain buffer	1000 units/buffer

Table 6 Pearson's correlation for 0.25-mile buffer

	Job density	Distance to light rail	Distance to park	Distance to bike station	Bike lane density	Residential building density	Multifamily housing density	Distance to waterfront	Bus ridership density
Job density	1.00								
Distance to light rail	0.45	1.00							
Distance to park	0.16	-0.04	1.00						
Distance to bike station	-0.05	0.43	0.02	1.00					
Bike lane density	-0.02	-0.05	-0.16	-0.11	1.00				
Residential building density	-0.22	0.21	0.08	0.50	-0.14	1.00			
Multifamily housing density	-0.27	-0.23	0.14	0.10	-0.18	0.05	1.00		
Distance to waterfront	-0.22	-0.01	0.01	0.08	-0.14	0.51	0.43	1.00	
Bus ridership density	-0.10	-0.37	-0.07	-0.27	0.38	-0.34	-0.08	-0.18	1.00

Table 7 Pearson's correlation for 0.5-mile buffer

	Multifamily housing density	Residential building density	job density	Distance to light rail	Distance to park	Distance to bike station	Bike lane density	Distance to waterfront	Bus ridership density
Multifamily housing density	1.00								
Residential building density	-0.19	1.00							
job density	-0.36	-0.17	1.00						
Distance to light rail	-0.35	0.09	0.33	1.00					
Distance to park	-0.20	0.12	0.20	-0.04	1.00				
Distance to bike station	-0.27	0.24	-0.16	0.43	0.02	1.00			
Bike lane density	0.15	-0.06	0.25	0.02	-0.19	-0.08	1.00		
Distance to waterfront	0.24	0.62	-0.08	-0.01	0.01	0.08	-0.04	1.00	
Bus ridership density	0.39	-0.45	-0.08	-0.40	-0.14	-0.32	0.34	-0.13	1.00

Table 8 Pearson's correlation for 0.75-mile buffer

	Job density	*Distance to light rail*	*Distance to park*	*Distance to bike station*	*Bike lane density*	*Residential building density*	*Multifamily housing density*	*Bus ridership density*	*Distance to waterfront*
Job density	1.00								
Distance to light rail	0.43	1.00							
Distance to park	0.18	-0.04	1.00						
Distance to bike station	-0.05	0.43	0.02	1.00					
Bike lane density	0.22	-0.10	-0.01	-0.23	1.00				
Residential building density	0.06	0.29	0.17	0.51	-0.18	1.00			
Multifamily housing density	-0.34	-0.23	-0.20	-0.19	-0.13	0.05	1.00		
Bus ridership density	-0.19	-0.43	-0.16	-0.31	0.21	-0.64	0.20	1.00	
Distance to waterfront	-0.09	-0.01	0.01	0.08	-0.09	0.59	0.52	-0.06	1.00

Table 9 Independent variable comparison summary

	0.25-mile buffer			*0.5-mile buffer*			*0.75-mile buffer*		
	Max	Min	Median	Max	Min	Median	Max	Min	Median
Multifamily housing density	161.94	0	14.40	672.54	0	193.00	76.78	0.64	11.82
Residential building density	2.90	0	0.01	5.66	0	0.38	2.47	0	0.51
job density	5.40	0.02	0.45	24.05	0.65	1.92	24.69	1.46	4.50
Distance to light rail	1.72	0.04	0.29	1.72	0.04	0.29	1.72	0.04	0.29
Distance to park	0.36	0	0.10	0.36	0	0.10	0.36	0	0.10
Distance to bike station	0.83	0.03	0.21	0.83	0.03	0.21	0.83	0.03	0.21
Distance to waterfront	1.30	0	0.43	1.30	0	0.43	1.30	0	0.43
Bike lane density	11.61	0	6.38	8.08	2.44	5.67	6.97	3.75	6.97
Bus ridership density	480.84	3.31	42.338	195.91	2.49	35.34	120.34	1.72	27.99

3.4 Model development

In order to answer this study's research questions regarding the relationship between weather and temporal factors on total daily ridership and the relationship between built environment and station-level ridership, linear regression modeling is employed. However, standard Ordinary Least Square linear regression models are not appropriate to this study area based on the following reasons. Because the relationships between temperature and ridership and wind and ridership are not linear, a polynomial regression model was used to account for these non-linear relationship. Specifically, I employ a polynomial regression model that builds on linear regression while incorporating quadratic terms and a logarithmic transformation applied to the dependent variables. Moreover, to investigate the effects of built environment on station-level ridership, I begin with a simple linear regression model, which was compared with logarithmic and root square transformations. The choice of best-fit model is based on the purpose of analysis and residual comparison.

3.4.1 Choice of models

In this part, I expect to compare different models and find the best fit one. To begin with, polynomial regression models with different degrees are examined and compared based on AIC. Then, collinearity test of independent variables is investigated to improve prediction accuracy as well.

Linear regression

Linear regression is an approach to model the relationship between a dependent variable and one or multiple independent variables. With more than one independent variable, the linear regression is called multiple linear regression. As the name indicates, a linear regression model expects a linear relationship between independent variables and dependent variable.

$$\boldsymbol{y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_mx_m + \varepsilon_i}$$

The models are estimated using the Ordinary Least Square (OLS) method. OLS is a widely used method for estimating variables in linear regression models. As we can learn from the words "Least square", OLS determines a best-fit line of regression model by minimizing the sum of squares between a real data point and regression line. All models in this study are developed using the statistics software R.

Polynomial regression

Polynomial regression is a special linear regression model. Unlike simple linear regression to explain the relationship between independent variables and dependent variable, polynomial regression attempts to explain independent and dependent variables as an Nth degree polynomial.

A typical polynomial regression model is as follows:

$$\boldsymbol{y = a_0 + a_1x_i + a_2x_i^2 + \cdots + a_mx_i^m + \varepsilon_i}$$

where i = 1, 2, … n

Akaike Information Criterion

Akaike information criterion (AIC) is widely used to measure the relative quality of a collection of models for one given dataset. RSS is the residual sum of squared errors, $\boldsymbol{\sigma}^2$

indicates the variance of the error in how we observe our observations, and $\boldsymbol{p}$ denotes the number of predictors. $\boldsymbol{\sigma}^{\wedge}\mathbf{2}$ equals to $\frac{\boldsymbol{RSS}}{\boldsymbol{n-p}}$[1]. A large value of AIC indicates a low quality of a model, therefore, a larger AIC is considered better when it's positive, while a smaller AIC is considered better when it's negative.

Variance Inflation Factor

Multicollinearity exists when two or more independent variables in a regression model are strongly or moderately correlated. Multicollinearity can decrease the precision of regression coefficients, and therefore limit the reliability of research findings and conclusions. In order to avoid this problem, this paper uses Variance Inflation Factors to detect multicollinearity. Variance Inflation Factor is a measure of how much a variance is inflated. The value of VIP starts at 1, which indicates that there is no correlation between variables. A small number of VIF indicates a less possibility of multicollinearity existence. In general, a VIF exceeding 10 indicates a serious multicollinearity ("Detecting Multicollinearity Using Variance Inflation Factors", 2016).

3.4.1.1 Polynomial regression model

To compare model fit, this paper tests different degrees of polynomial regression for the temperature and wind speed variables using Akaike Information Criterion or AIC. Table 10, Table 11, and Table 12 present the AIC and adjusted R-squared values for polynomial regression models with different degree and common transformations, including standard,

[1] Statistics consulting report

log, and square-root. We can see that the standard polynomial regression with the degree of temperature is 3 and the degree of wind speed is 2, the AIC is the lowest value, indicating it has a relatively good fit. While the log transformed polynomial regression with quadratic term of temperature and quadratic term of wind speed has relatively high fit. With square root transformation, polynomial regression with cubic term of temperature and quadratic term of wind speed has relatively high quality. Therefore, these three models are selected for next step.

Table 10 AIC and adjusted R-square for different degree of temperature and wind speed with none transformation

Degree of temperature	Degree of wind speed	AIC	Adjusted R^2
1	1	8515.928	0.7505
2	1	8492.708	0.7586
3	1	8479.107	0.7634
4	1	8480.624	0.7632
3	2	8470.508	0.7665

Table 11 AIC and adjusted R-square for different degree of temperature and wind speed with log transformation

Degree of temperature	Degree of wind speed	AIC	Adjusted R^2
1	1	113.0743	0.732
2	1	32.806	0.7603
3	1	31.72392	0.7609
4	1	33.52649	0.7607
3	2	16.67887	0.7661
3	3	18.52629	0.7658

Table 12 AIC and adjusted R-square for different degree of temperature and wind speed with square root transformation

Degree of temperature	Degree of wind speed	AIC	Adjusted R^2
1	1	3202.185	0.7554
2	1	3150.775	0.7723
3	1	3143.115	0.775
4	1	3145.103	0.7747
3	2	3129.835	0.7794
3	3	3130.163	0.7796

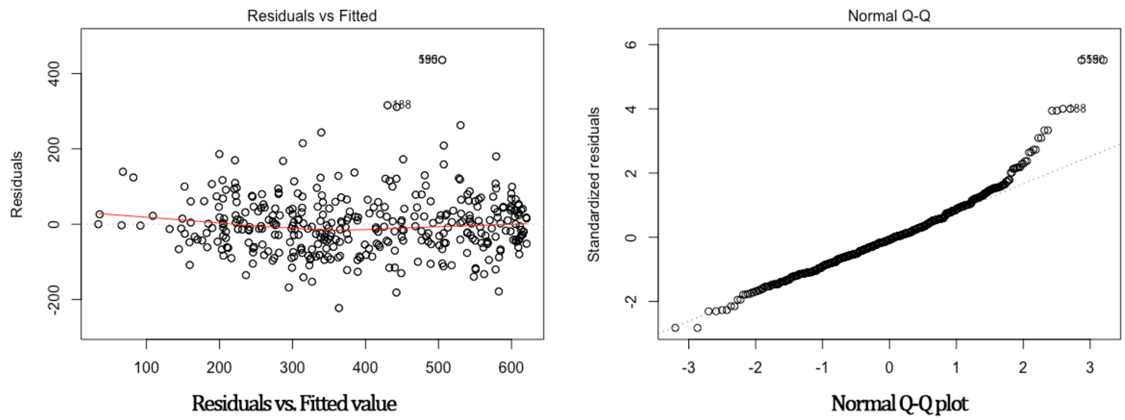


Figure 21 Residual versus fitted value and normal Q-Q plots of polynomial regression with none transformation

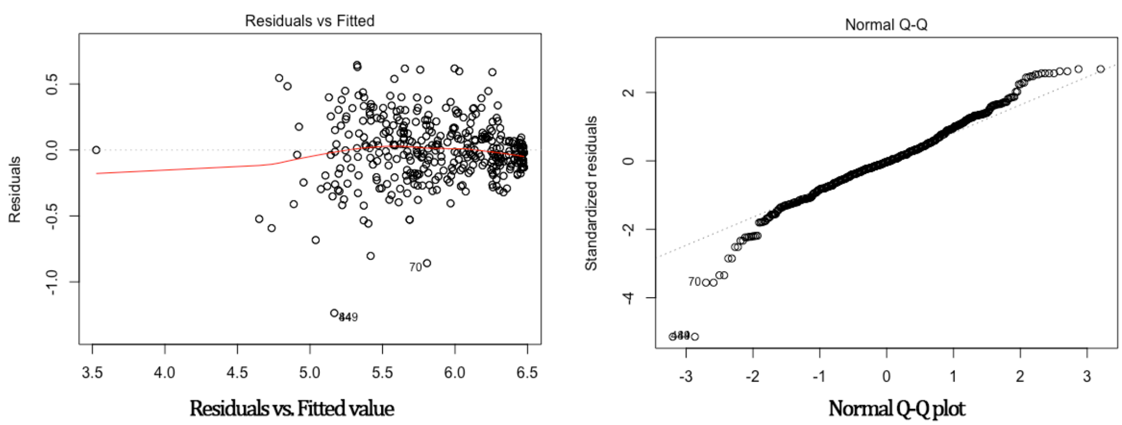


Figure 22 Residual versus fitted value and normal Q-Q plots of polynomial regression with log transformation

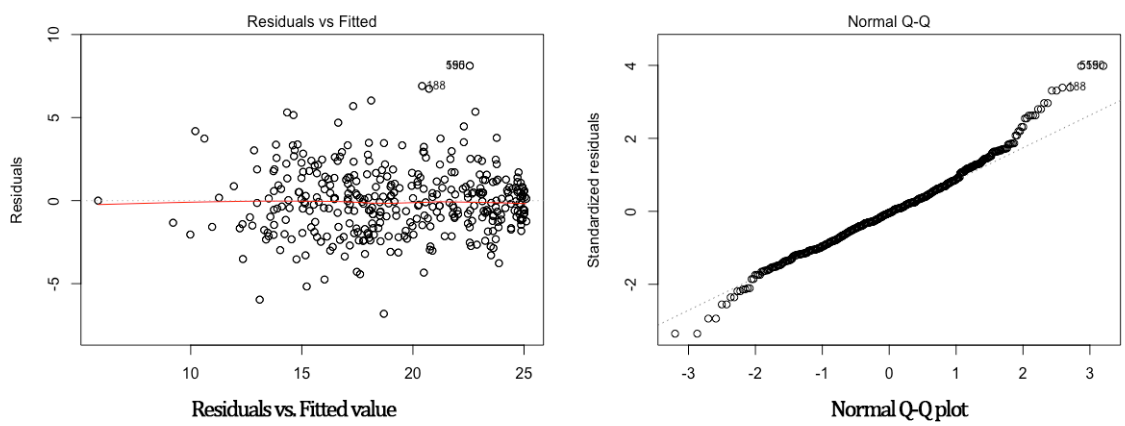


Figure 23 Residual versus fitted value and normal Q-Q plots of polynomial regression with root square transformation

Residual versus fitted value and normal Q-Q plots are used to compare the good fit of the selected models. Residual versus fitted value plot is one of the most common plot that is used to detect non-linearity. A good regression model is expected to have its residuals randomly distributed around the zero line. It indicates a linear relationship is reasonable. In addition, the residuals of an ideal model should roughly form a band around the zero line to ensure the variances of the error terms are equal. From the plots above, we can observe that the first (in Figure 21) and third (in Figure 23) polynomial regression models have a clear band, concentrating around the zero line, suggesting good fit of models.

The normal Q-Q plot is a tool to assess if a set of data is normally distributed and satisfies the regression assumptions. A normal Q-Q plot is expected to have observed values close to the line. While as shown in plots, the second normal Q-Q plot has very heavy tails on both sides, indicating a poor normal distribution of dependent variable. The normal Q-Q plots of the standard regression model has a heavy tail on the right side, while the third one has a slight tail on the right side, indicating a good normal distribution. Even though the third model does not have a perfectly normal Q-Q plot, it is still relatively good. Based on the analysis above, polynomial regression model with root square transformation, cubic term of temperature, and quadratic term of wind speed is selected to detect multicollinearity.

Based on AIC, residual distribution, and Q-Q plot, the polynomial regression with cubic term of temperature and quadratic term of wind speed is selected in this research. Temperature, wind speed, and precipitation are chosen as weather variables and day of the week is used as temporal variable.

It should be noted that for the purpose of interpretation, this study does not use orthogonal polynomial regression. Polynomial regression provides an easy way to interpret results because its independent polynomial terms can be treated as a series of linear regression terms. In this thesis, temperature, quadratic term of temperature, and cubic term of temperature are correlated. However, based on the definition of polynomial regression, the correlations between temperature and its quadratic and cubic terms and the correlation between wind speed and its quadratic term are expected. Multicollinearity should not be of much concern in this case.

The model is as follows:

$$
\begin{aligned}
\sqrt{\boldsymbol{Daily\ trip\ count}} \\
&= \boldsymbol{a_0} + \boldsymbol{a_1 Temperature} + \boldsymbol{a_2 Temperature^2} + \boldsymbol{a_3 Temperature^3} \\
&+ \boldsymbol{a_5}\textbf{Wind speed} + \boldsymbol{a_5}\textbf{Windspeed}^{\boldsymbol{2}} + \boldsymbol{a_5 Precipitation_{None}} \\
&+ \boldsymbol{a_6 Precipitation_{Light}} + \boldsymbol{a_7 Precipitation_{Moderate}} \\
&+ \boldsymbol{a_8 Precipitation_{Heavy}} + \boldsymbol{a_9 Precipitation_{Violent}} \\
&+ \boldsymbol{a_{14} Day\ of\ the\ week_{Weekday}} + \boldsymbol{a_{15} Day\ of\ the\ week_{Weekend}} \\
&+ \boldsymbol{a_{16} Day\ of\ the\ week_{Holiday}} + \boldsymbol{\varepsilon}
\end{aligned}
$$

3.4.1.2 Multiple linear regression

In this study, a multiple linear model is proposed for investigating the relationship between built environment and station-level ridership. Because the dependent variable does not follow a normal distribution, log transformation is conducted in order to meet the linear regression assumption. Variance inflation factor is calculated to detect multicollinearity. As shown in

Table 13, the VIF value of multifamily housing increases with the increase of spatial scale. The largest VIF value is over 5, indicating a possible multicollinearity. These results confirm the correlation between the multifamily housing density variable and other variables based on Pearson's Correlation Coefficient results. Therefore, the housing density variable, the sum of residential building density and multifamily unit density, is selected to reduce multicollinearity. After the transformation, the new largest VIF value is 2.01, which is much smaller than 10 in Table 14. This confirms that correlation between independent variables can be ignored here.

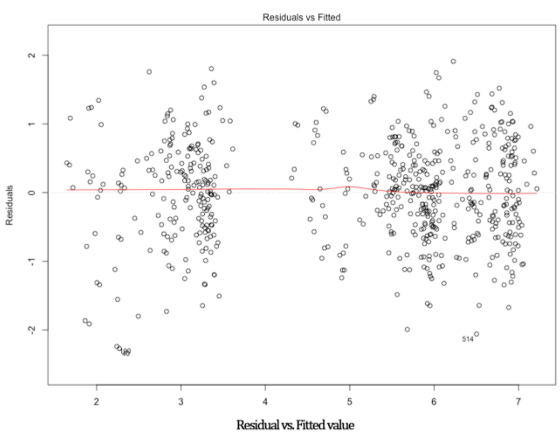
Figure 24, Figure 25, and Figure 26 depict residual versus fitted value plots for 0.25-mile, 0.5-mile, and 0.75-mile buffers. Most residuals are scattered randomly around the zero line, indicating the data follows normal distribution. Moreover, three clear bands are depicted, which tells us the regression models have no heterogeneity issue. Lastly, there are no significant standout points, indicating outliers do not serious affect the models. The normal Q-Q plots support the previous findings. Although we can observe the slight tails on both sides, most observed values are close to the line.

Table 13 Variance inflation factor of different scales with residential housing density and multifamily unit density variables

VARIABLES	0.25-MILE VIF	0.5-MILE VIF	0.75-MILE VIF
ORIGIN DESTINATION	1	1	1
DAY OF THE WEEK	1	1	1
DISTANCE TO PARK	1.147728	1.195936	1.152228
DISTANCE TO WATERFRONT	1.971887	2.391881	3.244188
DISTANCE TO BIKE STATION	1.884197	1.594188	2.123291
DISTANCE TO LIGHT RAIL	2.014145	1.754429	1.923261
JOB DENSITY	1.624931	2.166026	1.62922
RESIDENTIAL BLDG DENSITY	1.644028	2.123882	1.856235
MULTIFAMILY UNIT DENSITY	2.378605	3.016091	5.416394
BIKE LANE	1.227648	1.536759	1.194253
BUS SERVICE	1.496596	1.941218	2.855703

Table 14 Variance inflation factor of different scales without multifamily unit density and residential housing density variables

VARIABLES	0.25-MILE VIF	0.5-MILE VIF	0.75-MILE VIF
ORIGIN DESTINATION	1	1	1
DAY OF THE WEEK	1	1	1
DISTANCE TO PARK	1.109867	1.161971	1.134944
DISTANCE TO WATERFRONT	1.322218	1.154067	1.568267
DISTANCE TO BIKE STATION	1.429302	1.587712	1.489226
DISTANCE TO LIGHT RAIL	2.01311	1.749905	1.869316
JOB DENSITY	1.541556	1.776201	1.619857
HOUSING DENSITY	1.453441	1.697391	1.857889
BIKE LANE	1.227255	1.348679	1.187544
BUS SERVICE	1.422881	1.57315	1.362751



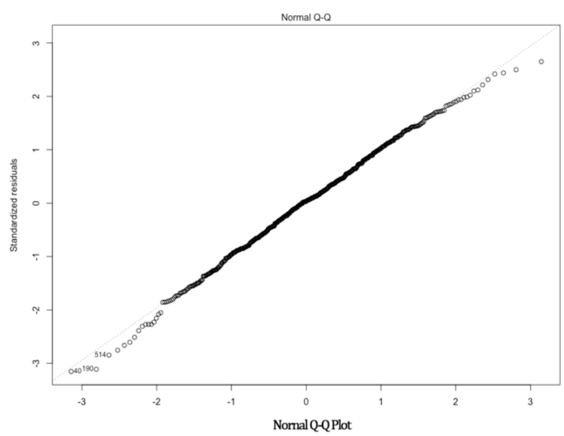


Figure 24 Residual versus fitted value and normal Q-Q plots of multiple linear regression with log transformation in 0.25-mile scale

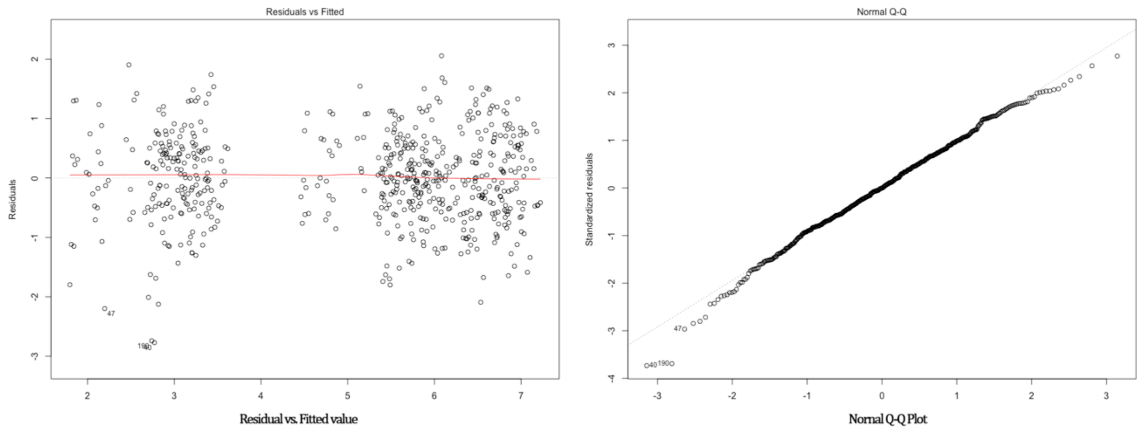


Figure 25 Residual versus fitted value and normal Q-Q plots of multiple linear regression with log transformation in 0.5-mile scale

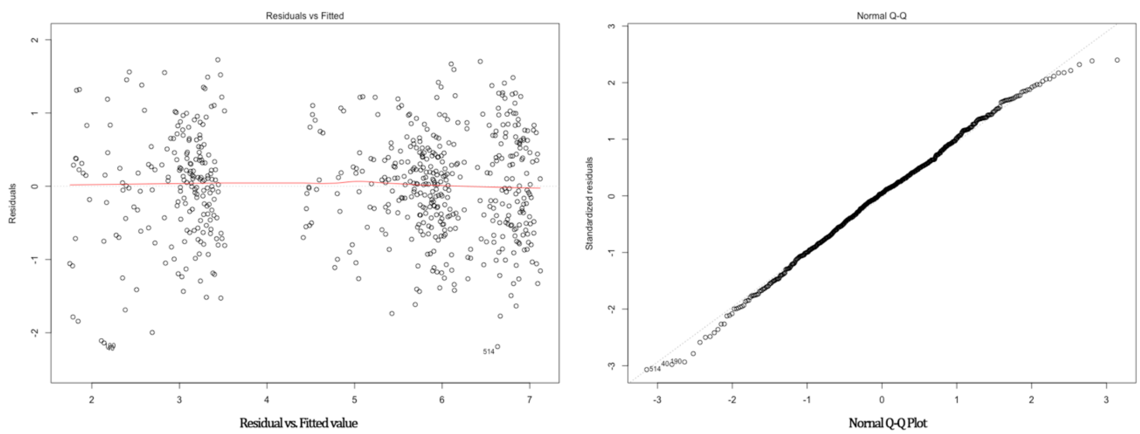


Figure 26 Residual versus fitted value and normal Q-Q plots of multiple linear regression with log transformation in 0.75-mile scale

Based on the data distribution, AIC, residual, and other tests, log transformed linear regression model is selected for investigating the effects of built environment on station-level ridership. The model function is as follows:

$$
\begin{aligned}
\boldsymbol{log\ Trip} \ = \boldsymbol{a_0} + \ & \boldsymbol{a_1 Origin} + \ \boldsymbol{a_2 Destination} + \boldsymbol{a_3 Distance\ to\ park} \\
& + \boldsymbol{a_4 Distance\ to\ waterfront} + \ \boldsymbol{a_5 Distance\ to\ bike\ share\ station} \\
& + \ \boldsymbol{a_6 Distance\ to\ light\ rail\ station} + \boldsymbol{a_7 Bus\ density} \\
& + \boldsymbol{a_8 Shortterm\ pass\ holder} + \ \boldsymbol{a_9 Weekend} + \ \boldsymbol{a_{10} Holiday} \\
& + \ \boldsymbol{a_{11} Bike\ lane\ density} + \ \boldsymbol{a_{12} housing\ unit\ density} + \boldsymbol{\varepsilon}
\end{aligned}
$$

CHAPTER 4 RESULTS

In this section, I present the results of polynomial regression and multiple linear regression estimations to understand the different effects of weather, temporal factors, and built environment on the bike share usage in Pronto. It should be noted that for the purpose of ease, the same model is used for investigating the effects of dependent variables on bike usage of different user type. The results are included in this section.

4.1 Polynomial regression model – weather and temporal variables

Data for this model is used to develop weather and ridership on a daily basis for a full year. The results of polynomial regression model are shown in Table 15. This model has a very high fit: the adjusted R-square is 0.78. In other words, more than 77% of the variation of daily ridership can be explained by weather and temporal variables. Overall, this is a relatively high fit model.

Table 15 Polynomial regression modeling results for weather and temporal variables

Variables	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	15.65	8.145	1.921	0.0551 .
Temperature	-0.5754	4.42e-01	-1.302	0.19342
Temperature^2	0.0201	7.82e-03	2.573	0.010291 *
Temperature^3	-0.0001	4.52e-05	-3.178	0.001549 **
Wind	0.0831	1.00e-01	0.831	0.406424
Wind^2	-0.0330	8.47e-03	-3.897	0.000106 ***
Precipitation-Light	-2.0900	2.15e-01	-9.722	<2.00e-16 ***
Precipitation-Moderate	-3.1210	2.51e-01	-12.424	<2.00e-16 ***
Precipitation-Heavy	-5.30900	3.21e-01	-16.535	<2.00e-16 ***
Precipitation-Violent	-11.960	1.46e+00	-8.188	1.21e-15 ***
Day of the week-Weekend	-1.2330	1.70e-01	-7.236	1.19e-12 ***
Day of the week-Holiday	-2.0530	4.74e-01	-4.328	1.72e-05 ***

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'
*Precipitation-None and Day of the week -Weekday are reference groups.
*Adjusted R-square is 0.7794

Temperature

The proposed hypothesis is that temperature is positively associated with ridership. The results first confirm there is a significant and strong relationship between temperature and ridership. However, this relationship is non-linear. The estimate of quadratic term is positive while the coefficient for linear and cubic terms are negative. These numbers indicate that the relationship between temperature and ridership is curvilinear. The basic hypothesis of a simple linear relationship between temperature and ridership is rejected based on the regression results. However, this does not completely dismiss the relationship.

The following few paragraphs presents two examples to better understand its relationship. Since the polynomial regression model includes more than one variables, we need to consider other variables as constant if we want to know the unit change of ridership caused by temperature. For example, the following functions show the derivative of temperature when assuming a summer weekday with 7 mile-per-hour wind speed and light precipitation and a winter weekday with 8 mile-per-hour and heavy precipitation:

$$\frac{\delta y}{\delta x} = 2 * (12.5247 - 0.5754x + 0.0201x^2 - 0.0001x^3)(-0.5754 + 0.0402x - 0.0003x^2)$$

$$\frac{\delta y}{\delta x} = 2 * (8.8938 - 0.5754x + 0.0201x^2 - 0.0001x^3)(-0.5754 + 0.0402x - 0.0003x^2)$$

In order to better understand the marginal change of ridership caused by unit change of temperature, we can look at:

$$\Delta y = \frac{\delta y}{\delta x}\Delta x$$

Δy presents the unit change of daily ridership, $\frac{\delta y}{\delta x}$ *presents the derivative function*, and Δx presents the unit change of temperature. Therefore, when assuming temperature is 69 ˚F, then one-degree increase in temperature is associated with approximately 56 increase in daily ridership. When assuming temperature is 44 ˚F, then one-degree increase in temperature is associated with approximately 17 increase in daily ridership.

Wind speed

Based on the model results, we can find that the linear wind speed and quadratic term are both significant. One coefficient is positive and the other is negative. This result shows that wind speed has a non-linear impact on ridership. Controlling other variables as constant, assuming a weekday with temperature is 69 ˚F, with light precipitation and assuming daily ridership is y, and wind speed is x, we can have the derivative function of wind speed:

$$\frac{\delta y}{\delta x} = 2 * (0.0831 - 0.0664x)(24.63 + 0.0831x - 0.0332x^2)$$

Moreover, the following function presents the unit change of daily ridership caused by the unit change of wind speed.

$$\Delta y = 2 * (0.0831 - 0.0664x)(24.63 + 0.0831x - 0.0332x^2)\Delta x$$

When assuming wind speed is 7 mile-per-hour, then one-unit increase in wind speed is associated with approximately 27 decrease in ridership. When assuming wind speed is 8 mile-per-hour, then one-unit increase in wind speed is associated with approximately 21 decrease in ridership. Overall, the impact of wind speed on daily ridership is positive, then negative.

Precipitation

All precipitation levels are significant in this study. No precipitation is used as the reference group here. Because the dependent variable is transformed by root square, the unit change of ridership caused by the change of precipitation is based on origin value of the dependent variable. In order to better understand the effects of precipitation on ridership, we can assume ridership is 225 when precipitation is none and other variables are constants. The ridership with light, moderate, heavy, and violent precipitation would be approximately 167, 141, 94, and 9. Overall, the greater the precipitation level, the lower ridership when other variables are held constants. This clearly shows the negative impact of precipitation on ridership and it confirms the hypothesis.

Day of the week

Day of the week and holiday are found significant in this study. To better understand the effects of temporal variables on ridership, we can assume ridership is 225 on a weekday and hold other variables as constants, the ridership on weekend and holiday would be approximately 190 and 168. The example clearly shows that Pronto users are less likely to ride Pronto on weekends, and lesser on holidays than weekdays. The results support the hypothesis.

4.2 Polynomial regression – annual member and short-term pass holder

The previous model focuses on how weather and temporal factor impact the overall trip count, while the two models in this section focus on the different impact of weather and temporal factor on annual members and short-term pass holders. As shown in Table 16 Polynomial regression models for annual member and short-term pass holder, both polynomial regression models for annual members and short-term pass holders have very high fits: the adjusted R-square values are 0.8539 and 0.7416, respectively, indicating that over 85% of the variation in annual members' trips can be explained by these variables and over 74% of the variation in short-term pass holders' trips can be explained. Most of the independent variables are significant, and the majority of them are significant at p = 0.0001 level, which also supports the goodness of fit.

Similar with the results of the model in previous section, non-linear relationships between temperature and wind speed and ridership of annual members and short-term pass holders are shown. Even temperature variables are not found significant in short-term pass holders' model. However, it does not necessarily mean the effects of temperature and wind speed on annual members' ridership does not exist, as the first model confirms temperature has a non-linear relationship with ridership.

According to the polynomial regression results, we can have the following temperature functions when assuming it is a weekday with light precipitation and 7 mile-per-hour wind speed:

$$\Delta y = 2 * (7.9061 - 0.0667x + 0.0074x^2 - 0.0001x^3)(-0.0667 + 0.0148x - 0.0003x^2)\,\Delta x$$

$where\ \Delta y$ presents the marginal change of daily ridership generated by annual members.

$$\Delta y = 2 * (0.0898 - 1.6220x + 0.0368x^2 - 0.0002x^3)(-1.6220 + 0.0736x - 0.0006x^2)\Delta x$$

$where\ \Delta y$ presents the marginal change of daily ridership generated by short $-$ term pass holder.

Assuming temperature is 69 ˚F, then one-degree increase in temperature is associated with 5 decrease in daily ridership of annual members and approximately 3 decrease in short-term pass holders' ridership. These numbers demonstrate the non-linear relationship between temperature and ridership of annual members and short-term pass holders.

Negative and positive association between wind speed and ridership are observed from models of annual members and short-term pass holders.

Precipitation is negatively associated with ridership of both annual members and short-term pass holders. However, the impact on annual member is less than short-term pass holders. Assume the origin ridership is 225 on a sunny day, then with other variable constant, the ridership of annual members on light, moderate, heavy, and violent day would be about 199, 181, 141, and 85; while the ridership of short-term pass holders would be 165, 144, 109, and 17. Figure 27 demonstrates the different changes of annual members and short-term pass holders.

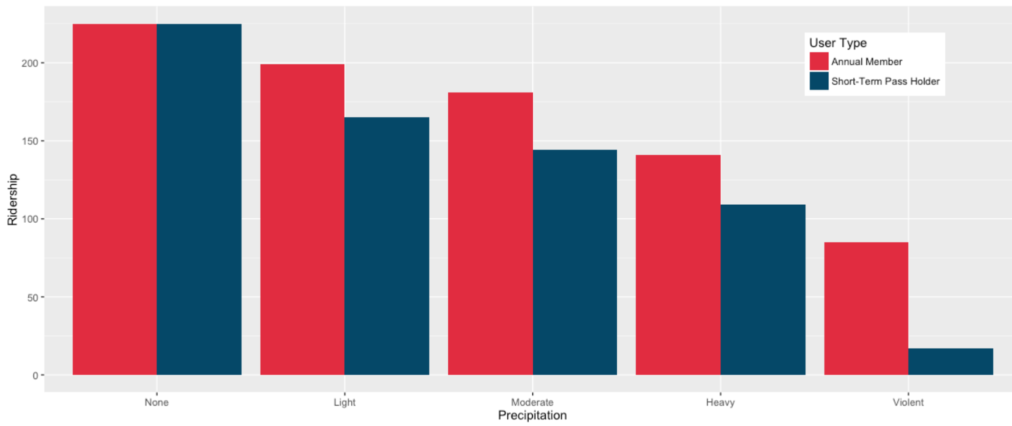


Figure 27 Effects of precipitation on ridership of annual member and short-term pass holder (assuming origin ridership is 225)

The coefficients of day of the week are very different for annual members and short-term pass holders—annual members use less bike share program on weekends and holidays than weekdays, while short-term pass holders use more on weekends and holidays. If assumed that the origin ridership on a weekday is 225, then for annual members the ridership on a weekend and holiday would be about 88 and 87; while for short-term pass holders the ridership on a weekend and holiday would be about 382 and 342. Figure 28 depicts the estimated numbers.

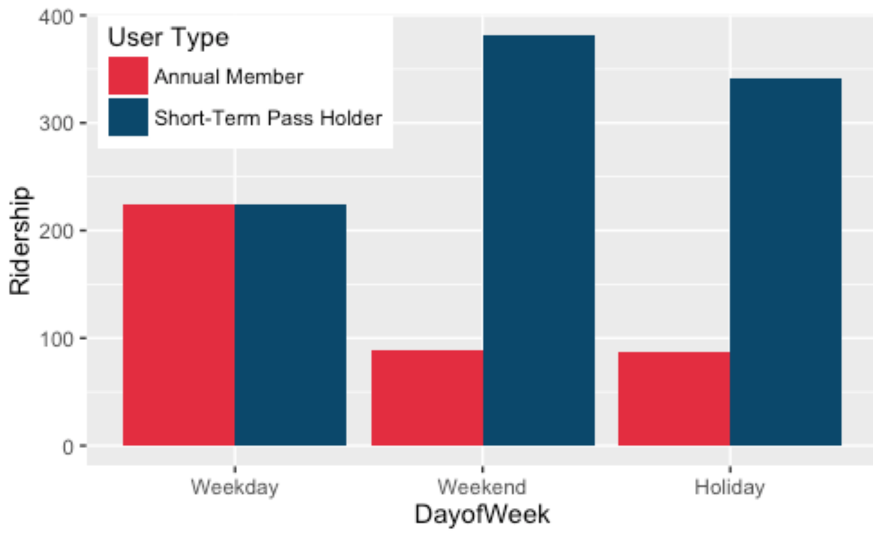


Figure 28 Effects of temporal variables on daily ridership of annual member and short-term pass holder (assuming origin ridership is 22)

Table 16 Polynomial regression models for annual member and short-term pass holder

	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
	Model 2 - Annual Member				Model 3 - Short-term Pass Holder			
Intercept	9.448	7.706	1.226	0.221	2.684	1.365	1.966	0.0501 .
Temperature	-0.0667	0.4182	-0.16	0.873	-1.6220	0.7409	-2.1890	2.92e-02 *
Temperature^2	0.0074	0.0074	1.001	0.318	0.0368	0.0131	2.8110	5.22e-03 **
Temperature^3	-0.0001	0.0000	-1.4780	0.1400	-0.0002	0.0001	-3.0420	2.52e-03 **
Wind	-0.0121	0.0947	-0.1280	0.8980	0.1864	0.1677	1.1120	2.67e-01
Wind^2	-0.0112	0.0080	-1.3980	0.1630	-0.0394	0.0142	-2.7780	5.77e-03 **
Precipitation-Light	-0.9084	0.2034	-4.4660	0.0000 ***	-2.1480	0.3603	-5.9610	6.07e-09 ***
Precipitation-Moderate	-1.5490	0.2377	-6.5160	0.0000 ***	-3.0190	0.4210	-7.1700	4.42e-12 ***
Precipitation-Heavy	-3.1130	0.3038	-10.246	<0.000 ***	-4.5560	0.5382	-8.4660	6.91e-16 ***
Precipitation-Violent	-5.7580	1.3810	-4.1680	0.0000 ***	-10.8400	2.4470	-4.429	1.27e-05 ***
Weekend	-5.6350	0.1612	-34.954	<0.000 ***	4.5530	0.2856	15.942	<2.00e-16 ***
Holiday	-5.6820	0.4488	-12.660	<2.00e-16 ***	3.5040	0.7950	4.407	1.39e-05 ***
Adjusted R-square	0.8539				0.7416			

*Significant level: 0 ‘***’, 0.001 ‘**’, 0.01’*’, 0.05’.’
*Precipitation-None, Season-Spring, and Day of the week -Weekday are reference groups.

4.3 Multiple linear regression model

In this section, I discuss the results of the multiple linear regression models to estimate the relationship of built environment characteristics on the station-level bicycle usage. It must be noted that for the purpose of ease, same multiple linear regression models are used to investigate the effects of built environment on bicycle usage of annual members and short-term pass holders. Three models were developed, for buffers 0.25, 0.5, and 0.75 mile buffers, and each has a good fit with over 80% of the variation in station-level ridership can be explained by these variables. However, we should note that not all variables are statistically significant for each buffer distance.

4.3.1 In 0.25-mile buffer

Table 17 shows the results of multiple linear regression for station-level analysis for a 0.25-mile buffer. The adjusted R-squares are 0.8267. The high adjusted R-square indicates a good fit of the model.

People tend to use Pronto more on weekdays than weekends and holidays as highlighted by the negative coefficient of the temporal variables. However, the length of bike lane in a 0.25-mile buffer of each station, user type, and housing density variables are not statistically significant. The results provide an indication that 0.25-mile scale is not as productive for investigating the effects of bike infrastructure and housing variables on station-level ridership.

Table 17 Multiple linear regression results of 0.25-mile buffer

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.7852	0.1473	25.691	<2e-16 ***
Origin	0.0308	0.05954	0.517	0.60505
Short-term holder	0.0417	0.05954	0.701	0.48356
Weekend	-0.9398	0.0729	49.43	<2.00e-16 ***
Holiday	-3.6045	0.0729	36.541	<2.00e-16 ***
Distance to park	-1.2697	0.4092	-3.103	0.00201 **
Distance to waterfront	-0.5523	0.1071	-5.157	3.44E-07 ***
Distance to bike station	-0.7103	0.2713	-2.619	0.00905 **
Distance to light rail	0.0181	0.1487	0.121	0.90349
Job density	-0.3107	0.02846	-10.917	<2.00e-16 ***
Housing density	0.0017	0.0012	1.352	0.17696
Bike lane density	-0.0173	0.0131	-1.319	0.18785
Bus density	0.0008	0.0003	2.33	0.02017 *
Adjusted R-square	0.8267			
Degree of freedom	0.7292 on 587 degrees of freedom			

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'

It is expected that the station-level ridership increases when the distance to parks, waterfront, and bike stations decrease. The negative coefficients confirm the hypothesis. To be more specific, being 0.1 mile closer to a park, waterfront, or other bike share station is associated with about 12.7%, 5.5%, and 7.1% increase in the station-level ridership. Pronto users more often combined their trip with recreational activity. The positive impact of bus service density near bike stations on station-level ridership is recognized. The results indicate that higher level of Pronto usage is likely to occur in area with greater bus ridership compared to an area with lower ridership. There was no statistically significant relationship between bike ridership and the distance to light rail variable. One possible explanation for this situation is because the service area of light rail is relatively small when we compare it with public bus services. The relationship between distance to light rail and station-level ridership may not be extracted in this model.

Some similar patterns were observed when dichotomizing by membership type (see Table 18). Both annual members and short-term pass holders tend to use more on weekdays than weekends and holidays. This finding follows the previous results. The housing density variable is significant at the $p = 0.05$ level in annual members' model. Even though the significance is not high, it still demonstrates a positive impact of housing density on station-level ridership. This finding provides an indication that short-term pass holders tend to be tourists who do not live in Seattle. Regarding job density, Pronto stations with higher local job density tend to have lower bike usage. One of the possible explanations is that in high job density areas, traffic tends to be heavy and Pronto users may consider it unsafe to ride a bike in that area.

Distance of stations to parks, waterfront, and other bike station have a negative relationship with bike usage in the previous model. However, the effects of proximity on bike usage of annual members and short-term pass holders are different from each other. Possible explanations are that Pronto annual members more often combine their trip with other bike station and activity in a park, while short-term pass holders combine their trip with a waterfront visit rather than a park and other bike station. As the waterfront is one of the most popular points for tourists, the highlighted correlation between proximity to waterfront and short-term pass holders' bike usage indicates that short-term pass holders tend to be tourists.

This result also provides an indication that annual members and short-term pass holders have different travel purposes.

The bike usage of short-term pass holders increases when there are more bike lanes nearby a Pronto station. Even though the significant level is not high, we can still expect to see that short-term pass holders tend to use bike lane for safety concerns. However, the negative impact of bike lanes on the bike usage of annual members indicates that annual members do not care as much as short-term pass holders.

Table 18 Multiple linear regression model results for 0.25-mile buffer

	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
	Annual Member				Short-term Pass Holder			
igin	-0.0071338	0.0716066	-0.1	0.920712	0.0687502	0.0623436	1.103	0.27105
eekend	-1.6903076	0.0876998	48.711	<2.00e-16 ***	-0.1893686	0.076355	38.465	<2.00e-16 ***
liday	-4.2719429	0.0876998	29.437	<2.00e-16 ***	-2.9369899	0.076355	35.985	<2.00e-16 ***
stance to park	-1.6714896	0.4921664	-3.396	0.000779 ***	-0.8679127	0.4285002	-2.025	0.04374 *
stance to waterfront	0.0753009	0.128801	0.585	0.559254	-1.1798401	0.1121394	-10.521	<2.00e-16 ***
stance to bike station	-1.1239331	0.3262061	-3.445	6.55e-04 ***	-0.2967017	0.2840084	-1.045	0.29704
stance to light rail	-0.1528966	0.1789539	-0.854	3.94e-01	0.1889971	0.1558046	1.213	0.22611
o density	-0.3484393	0.0342224	-10.182	<2.00e-16 ***	-0.2728694	0.0297954	-9.158	<2.00e-16 ***
ousing density	0.002846	0.0014816	1.921	5.57e-02 .	0.0004847	0.0012899	0.376	7.07e-01
ke lane density	-0.0612208	0.0157886	-3.878	1.31e-04 ***	0.026602	0.0137462	1.935	5.39e-02 .
s density	0.0006328	0.0004196	1.508	0.132671	0.0009928	0.0003653	2.718	0.00697 **
ljusted R-square	0.8988				0.8757			
gree of freedom	0.6201 on 288 degrees of freedom				0.5399 on 288 degrees of freedom			

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'

4.3.2 In 0.5-mile buffer

A multiple linear regression model is estimated for data measured within a 0.5-mile buffer. The adjusted R-square of this model is 0.82, indicating that over 80% variance can be explained by the independent variables in Table 19.

Table 19 Multiple linear regression results of 0.5-mile buffer

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.1063	0.1879	16.534	<2e-16 ***
Origin	0.0308082	0.0613932	0.502	0.61598
Short-term holder	0.0417391	0.0613932	0.68	0.49686
Weekend	-0.9398381	0.075191	47.937	<2.00e-16 ***
Holiday	-3.6044664	0.075191	35.438	<2.00e-16 ***
Distance to park	-0.9444141	0.4317591	-2.187	0.02911 *
Distance to waterfront	-0.459855	0.1031693	-4.457	9.94E-06 ***
Distance to bike station	-0.2892204	0.29477	-0.981	3.27E-01
Distance to light rail	-0.4101642	0.143048	-2.867	0.00429 **
Job density	-0.0379882	0.0063801	-5.954	4.50E-09 ***
Housing density	0.0008419	0.0001946	4.325	1.79E-05 ***
Bike lane density	0.0559443	0.0271405	2.061	0.03972 *
Bus density	-0.0008235	0.0006558	-1.256	0.20972
Adjusted R-square	0.8158			
Degree of freedom	0.7519 on 587 degrees of freedom			

*Significant level: 0 ‘***’, 0.001 ‘**’, 0.01’*’, 0.05’.’

Similar to the results in the 0.25-mile buffer, people tend to bicycle more on weekdays than weekends and holidays as highlighted by the negative coefficient of the weekend and holiday variables. The negative coefficient of job density variable mirrors the same results in the 0.25-mile buffer. In the 0.5-mile buffer model, bike lane density is positive associated with bike usage. It indicates that Pronto usage is likely to occur in dense bike lane area compared to less dense area.

The negative coefficients of distance to park, waterfront, and light rail variables support that stations closer to parks, waterfront, or light rail stations, have greater bike usage, similar to the results in previous mode. However, bus ridership density and distance to bike station are not statistically significant in the 0.5-mile buffer model. The results indicate that Pronto users more often combine their trip mode with light rail then bus transit in the 0.5-mile buffer. The combination of bus service and bike share mode is preferred by Pronto users in 0.25-mile buffer. It highlights that adding more transit services in a 0.25-mile buffer of each station is likely to attract more bike usage.

The models were also used to estimate the relationship between built environment and bike usages of annual members and short-term pass holders. The adjusted R-squares of these two models are 0.89 and 0.86, indicating that over 85% variance can be explained by the independent variables in Table 20. The high adjusted R-square values support good fits of the models.

Similar to previous findings, distance to parks is negatively associated with the bike usage of annual members, while distance to waterfront is negatively related to the bike usage of short-term pass holders. The result provides an indication that parks are more attractive to annual members while waterfront is more popular to short-term pass holders. It also highlights the possibility that adding more recreational area is likely to increase Pronto usage. A strong correlation between proximity to light rail station and bike usage of annual members is also found. Pronto annual members more often combine their trip mode with light rail in the 0.5-mile buffer.

The results suggest that housing unit density exerts a significant and positive effect on bike usage of annual members. It indicates that Pronto stations near denser neighborhoods are more likely to experience a higher volume of bike usage. This finding offers an indication that annual members start or end their trip at the Pronto station near their home. It also indicates that Pronto bike share probably is used as a tool to bridge the "last mile" to people's destination.

Table 20 Multiple linear regression model results for 0.5-mile buffer size

	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
	Annual Member				Short-term Pass Holder			
igin	-0.0071338	0.0746774	-0.09.60	0.923962	0.0687502	0.0656059	1.048	0.2956
eekend	-1.6903076	0.0914607	-18.481	<2.00e-16 ***	-0.1893686	0.0803505	-2.357	1.91e-02 *
liday	-4.2719429	0.0914607	-46.708	<2.00e-16 ***	-2.9369899	0.0803505	-36.552	<2.00e-16 ***
stance to park	-1.2463279	0.5251826	-2.37	1.83E-02 *	-0.6425003	0.4613857	-1.393	1.65e-01
stance to waterfront	0.1794087	0.1254929	1.43	1.54E-01	-1.0991186	0.1102486	-9.97e+00	<2.00e-16 ***
stance to bike station	-0.3476318	0.3585519	-0.97	3.33E-01	-0.230809	0.3149966	-0.733	0.4643
stance to light rail	-0.6364258	0.1740005	-3.658	3.03E-04 ***	-0.1839026	0.1528637	-1.203	0.2299
density	-0.0336858	0.0077606	-4.341	1.97E-05 ***	-0.0422906	0.0068178	-6.20e+00	1.92e-09 ***
using unit density	0.0014428	0.0002368	6.094	3.52E-09 ***	0.0002409	0.000208	1.158	2.48e-01
ke lane density	-0.0248032	0.0330131	-0.751	0.453078	0.1366919	0.0290028	4.713	3.80e-06 ***
s density	-0.0014918	0.0007977	-1.87	0.062467 .	-0.0001551	0.0007008	-0.221	8.25e-01
ljusted R-square	0.89				0.8624			
gree of freedom	0.6467 on 288 degrees of freedom				0.5682 on 288 degrees of freedom			

*Significant level: 0 ‘***’, 0.001 ‘**’, 0.01’*’, 0.05’.’

4.3.3 In 0.75-mile buffer

This section covers the results of multiple linear regression estimation for the 0.75-mile buffer. From Table 21, the adjusted R-square of this model is 0.83, indicating that over 82% variance can be explained by the independent variables.

Table 21 Multiple linear regression results of 0.75-mile buffer

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.1705	0.2822	14.778	<2e-16 ***
Origin	0.0308082	0.059254	0.52	0.60331
Short-term holder	0.0417391	0.059254	0.704	0.48146
Weekend	-0.9398381	0.072571	49.668	<2.00e-16 ***
Holiday	-3.6044664	0.072571	36.718	<2.00e-16 ***
Distance to park	-1.1064048	0.4118398	-2.686	0.00743 **
Distance to waterfront	-0.3498049	0.1160759	-3.014	0.00269 **
Distance to bike station	-0.7233145	0.2755339	-2.625	8.89e-03 **
Distance to light rail	-0.144832	0.1426965	-1.015	3.11e-01
Job density	-0.0654685	0.0062083	-10.545	<2.00e-16 ***
Housing density	-0.0005276	0.0028991	-0.182	8.56e-01
Bike lane density	-0.0619299	0.0474251	-1.306	1.92e-01
Bus density	0.0004649	0.0009125	0.509	0.61065
Adjusted R-square	0.8284			
Degree of freedom	0.7257 on 587 degrees of freedom			

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'

From Table 22, we can find results similar to the results of previously used temporal variables -- people tend to use Pronto more on weekdays in all three models. Job density is identified to have negative impacts on bike usage in three buffers. It rejects the previous assumption that Pronto usage is more likely to occur with higher job density. As mentioned, a possible explanation is because the heavy traffic in high job density area increases people's safety concern. Increased proximity to parks and waterfront are all negatively associated with bike usage. These results are repeatedly found in all three buffers, indicating the strong attractiveness of parks and waterfront to Pronto users. Distance to bike stations is negatively related to bike usage in 0.75-mile scale.

The results of increased usage with lower distance to parks and waterfront may suggest recreational activity being the trip purpose for annual members and short-term pass holders. In order to attract more bike users, it would be reasonable to increase the access to parks or waterfront from a bike station.

Table 22Mmultiple linear regression model results for 0.75-mile buffer

	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
	Annual Member				Short-term Pass Holder			
rigin	-0.0071338	0.0686122	-0.104	0.917264	0.0687502	0.0660494	1.041	2.99E-01
eekend	-1.6903076	0.0840325	-20.115	<2.00e-16 ***	-0.1893686	0.0808936	-2.341	1.99E-02 *
oliday	-4.2719429	0.0840325	-50.837	<2.00e-16 ***	-2.9369899	0.0808936	-36.307	<2.00e-16 ***
istance to park	-1.0106083	0.4768837	-2.119	3.49E-02 *	-1.2022013	0.4590706	-2.62E	0.00929 **
istance to waterfront	0.3253	0.1344084	2.42	0.016129 *	-1.0249097	0.1293878	-7.921	5.16E-14 ***
istance to bike station	-1.101741	0.3190503	-3.453	0.000637 ***	-0.3448881	0.3071328	-1.123	0.2624
istance to light rail	-0.2318837	0.1652332	-1.403	1.62E-01	-0.0577803	0.1590613	-3.63E-01	7.17E-01
ob density	-0.0807212	0.0071888	-11.229	<2.00e-16 ***	-0.0502158	0.0069203	-7.256	3.71E-12 ***
ousing unit density	0.0017501	0.003357	0.521	0.602534	-0.0028054	0.0032316	-0.868	3.86E-01
ike lane density	-0.0955943	0.0549152	-1.741	0.082793 .	-0.0282656	0.052864	-0.535	5.93E-01
us density	0.0003086	0.0010566	0.292	0.77042	0.0006211	0.0010172	0.611	5.42E-01
djusted R-square	0.9071				0.8605			
egree of freedom	0.5942 on 288 degrees of freedom				0.572 on 288 degrees of freedom			

*Significant level: 0 ‘***’, 0.001 ‘**’, 0.01’*’, 0.05’.’

The following tables summarize the coefficients of built environment and significance of three models. As mentioned in Chapter 2, one of the limitations in existing literature is that researchers use different scales to measure built environment and it may lead to different results. As shown in Table 23, we can observe that only distance to parks, distance to waterfront, and job density are significant in three spatial scales, while other variables are either significant in one or two scales. These findings not only provide an indication that the subjective measurement of built environment could lead to biased results, but also highlight the important effects of parks and waterfront on station-level ridership.

Table 23 Compare correlation significance by buffer size

	0.25-MILE	0.5-MILE	0.75-MILE
DISTANCE TO PARK	-1.270 **	-0.944 *	-1.106 **
DISTANCE TO WATERFRONT	-0.552 ***	-0.460 ***	-0.350 **
DISTANCE TO BIKE STATION	-0.710 **	-0.289	-0.723 **
DISTANCE TO LIGHT RAIL	0.0180	-0.410 **	-0.14483
JOB DENSITY	-0.311 ***	-0.038 ***	-0.0654 ***
HOUSING UNIT DENSITY	0.00167	0.0008 ***	-0.00053
BIKE LANE DENSITY	-0.0173	0.0559 *	-0.06193
BUS DENSITY	0.0008 *	-0.0008	0.00047

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'

Table 24 and Table 25 demonstrate the effects of built environment on the station-level trips generated by annual members and short-term pass holders. The following observations can be made from the tables. First, distance to parks is the only built environment variable associated with station-level ridership of annual members. Second, distance to waterfront is the only built environment variable associated with short-term pass holders' ridership. Third, it is interesting to see that the job density variable is negatively related to station-level ridership in all models.

The results clearly highlight that increasing distance to parks and waterfront results in a decrease of bike usage. These findings can be helpful when planners try to understand the best location to allocate new stations to ensure high usage. Moreover, the results show that Pronto users more often avoid biking in dense employment areas. Possible explanations for this findings are as follows: First, a high job density area is usually associated with heavy traffic, which can arise bicyclists' safety concerns. Second, in a high job density area, there are more travel alternatives. For instance, people can take bus, light rail, or taxi to get to their destination. Compared to other travel modes, bike share can be less competitive in terms of travel time and speed. Third, as can be discerned from the hourly count ridership figure, the strong commute pattern of annual members suggests the assumption that annual members use Pronto for daily work commuting. The quality of job density data might be the cause of the opposite results.

Table 24 Compare correlation significance of annual member by buffer size

ANNUAL MEMBER	0.25-MILE	0.5-MILE	0.75-MILE
DISTANCE TO PARK	-1.671 ***	-1.246 *	-1.011 *
DISTANCE TO WATERFRONT	0.0753	0.1794	0.325 *
DISTANCE TO BIKE STATION	-1.123 ***	-0.3476	-1.10 ***
DISTANCE TO LIGHT RAIL	-0.1529	-0.636 ***	-0.23188
JOB DENSITY	-0.348 ***	-0.0337 ***	-0.0807 ***
HOUSING UNIT DENSITY	0.003 .	0.001 ***	0.00175
BIKE LANE DENSITY	-0.0612 ***	-0.0248	-0.0955 .
BUS DENSITY	0.0006	0.001 .	0.00031

*Significant level: 0 '***', 0.001 '**', 0.01'*', 0.05'.'

Table 25 Compare correlation significance of short-term pass holder by buffer size

SHORT-TERM PASS HOLDER	0.25-MILE	0.5-MILE	0.75-MILE
DISTANCE TO PARK	-0.868 *	-0.6425	1.2022 **
DISTANCE TO WATERFRONT	-1.180 ***	-1.099 ***	-1.025 ***
DISTANCE TO BIKE STATION	-0.2967	-0.2308	-0.3449
DISTANCE TO LIGHT RAIL	0.1890	-0.1839	-0.0578
JOB DENSITY	-0.273 ***	-0.042 ***	-0.050 ***
HOUSING UNIT DENSITY	0.0005	0.0002	-0.0028
BIKE LANE DENSITY	0.0266 .	0.136 ***	-0.02827
BUS DENSITY	0.001 **	-0.0002	0.00062

*Significant level: 0 ‘***’, 0.001 ‘**’, 0.01’*’, 0.05’.’

CHAPTER 5 CONCLUSION AND LIMITATIONS

This thesis provides new perspectives on bike share usage, with particular focus on weather and built environment characteristics. It is hoped that it will be used by planners and Pronto managers to enhance bike station location decisions, aid in development of expansion plans, and improve efficiency of bike rebalance.

5.1 Weather and times

Seattle is a city with marked weather variation. Seattle has substantial amounts of rain, has cold and humid days in the winter and hot and dry days in the summer. This study provides planners with analysis of the relative relationships of various weather conditions and temporal factors with bike share ridership. The results show that fewer trips are generated on days with heavy rain, high wind speed, and cold temperatures for both annual members and short-term pass holders.

Many of these results are not surprising. The differential relationship of weather and temporal factor with the ridership of annual members and short-term pass holders can be used to optimize bike rebalancing and can contribute to the Pronto expansion plan. For instance, at popular stations for short-term pass holders, information kiosks with guidance and information on how to ride in the rain, or even provision of rain gear could increase ridership, given that short-term pass holders seem to be more sensitive to weather change than annual members. With respect to varying daily ridership, annual members usually make fewer trips on weekends than on weekdays; however, short-term pass holders make more trips on weekends. Therefore, Pronto managers can expect a higher demand for bike rebalance at bike stations popular with short-term pass holders on weekends and holidays.

5.2 Station location

The study results provide new insights into built environment factors that correlated with station-level bike usage and can be used by planners to locate new stations and optimize current system. Bike share program is proposed to solve the “last mile” problem therefore bike locations are normally located in high population density and points of interest area. The unanticipated negative coefficient on the measure of job density in three scales and negative coefficients of proximity to waterfront and parks indicate Pronto mainly serves recreational purposes rather than utilitarian purposes.

Because Pronto has real-time information on the number of available bikes and docks, the results provide insights of why station use varies and why stations attract different users from a built environment perspective. Moreover, planners can use the multiple linear regression model to predict station usage at particular locations, based on built environment information, to assist with the locating of new bike stations. Further, the results from separate regression models for annual members and short-term pass holders may also be used for potential user outreach.

In sum, this study fills a gap in research on Seattle’s bike share program. It develops regression models that can be applied to predict daily ridership, and moreover, looks into the impact of built environment on station-level ridership across different spatial scales. The models in the study can provide guidance for bike share program planners.

5.3 Limitations

One limitation of this research is that the total number of membership is not taken into account as control variable. Steadily increasing annual membership has tracked with Pronto's development. From October 13th, 2014 to October 12th, 2015, the total number of annual members increased, given that the membership is valid for one year. The increase in membership may have led to an increase in total ridership, but a decrease in average trips per person. This missing information threatens the accuracy of analysis results. It would be meaningful to consider this problem in future research when the data are available. Moreover, topography, socio-demography, and many other variables are not taken into consideration in this study.

Although this study provides new insights into station-level ridership across different scales, it also raises new questions. For instance, this study does not consider any spatial relationships that may exist between different variables. Built environment variables can be geographically associated (i.e. stations that are close to each other are likely to have similar characteristics, and therefore are not completely independent with respect to statistical modelling). Future research could use spatial statistical analytics, such as Moran's I and Geographically Weighted Regression models to test whether built environment variables are clustered and whether they are spatially auto-correlated.

Additional research into specific times that affect use could provide important information for bike rebalances. For instance, studying station-level ridership and origin-to-destination trips during peak hour and off-peak period could be useful for projecting level of service.

Second, a comparative study of a variety of different bike share programs can be another approach for achieving successful bike share management. Lastly, using street network service areas instead of buffers based on Euclidean distance may improve the quality of data used in the models.

References

Bicycle Master Plan, City of Portland. (2016). *Portlandoregon.gov*. Retrieved 28 May 2016, from https://www.portlandoregon.gov/shared/cfm/image.cfm?id=40414

Buehler, R. & Pucher, J. (2011). Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*, *39*(2), 409-432. http://dx.doi.org/10.1007/s11116-011-9355-8

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American Housing Survey.*Transportation Research Part A: Policy And Practice*, *30*(5), 361-377. http://dx.doi.org/10.1016/0965-8564(95)00033-x

City of Seattle GIS Data. (2016). *Washington State Geospatial Data Archive*. Retrieved 1 January 2016, from https://wagda.lib.washington.edu/data/geography/wa_cities/seattle/index.html

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. *Journal Of Transport Geography*, *41*, 292-305. http://dx.doi.org/10.1016/j.jtrangeo.2014.09.003

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. *Journal Of Transport Geography*, *41*, 292-305. http://dx.doi.org/10.1016/j.jtrangeo.2014.09.003

DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal Of Public Transportation*, *12*(4), 41-56. http://dx.doi.org/10.5038/2375-0901.12.4.3

The whimsical anarchism of the White Bicycle revolution. (2015). *DangerousMinds*. Retrieved 9 June 2016, from http://dangerousminds.net/comments/the_white_bicycle_revolution

Detecting Multicollinearity Using Variance Inflation Factors. (2016). *Regression Methods*. Retrieved 28 May 2016, from https://onlinecourses.science.psu.edu/stat501/node/347

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2015). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto.*Transportation*. http://dx.doi.org/10.1007/s11116-015-9669-z

Encyclopedia of transportation: social science and policy. (2015). *Choice Reviews Online*, *52*(09), 52-4548-52-4548. http://dx.doi.org/10.5860/choice.189065

Faghih-Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal Of Transport Geography*, *41*, 306-314. http://dx.doi.org/10.1016/j.jtrangeo.2014.01.013

Fares & ORCA. (2016). *King County Metro*. Retrieved 28 May 2016, from King County Metro: http://metro.kingcounty.gov/fares/

Gallop, C., Tse, C., & Zhao, J. (2011). A seasonal autoregressive model of Vancouver bicycle traffic using weather variables. *I-Manager's Journal On Civil Engineering*, *1*(4), 9.

Soper, T. (2014). *Seattle's bike-share program launches today: Here are the docking station locations*. *GeekWire*. Retrieved 9 June 2016, from http://www.geekwire.com/2014/seattles-bike-share-program-launches-today-heres-docking-stations/

Gebhart, K. & Noland, R. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, *41*(6), 1205-1225. http://dx.doi.org/10.1007/s11116-014-9540-7

Handy, S., Boarnet, M., Ewing, R., & Killingsworth, R. (2002). How the built environment affects physical activity. *American Journal Of Preventive Medicine*, *23*(2), 64-73. http://dx.doi.org/10.1016/s0749-3797(02)00475-0

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport And Environment*, *10*(6), 427-444. http://dx.doi.org/10.1016/j.trd.2005.05.002

Kockelman, K. (1997). Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area. *Transportation Research Record: Journal Of The Transportation Research Board*, *1607*, 116-125. http://dx.doi.org/10.3141/1607-16

Miranda-Moreno, L. & Nosal, T. (2011). Weather or Not to Cycle. *Transportation Research Record: Journal Of The Transportation Research Board*, *2247*, 42-52. http://dx.doi.org/10.3141/2247-06

Moudon, A., Lee, C., Cheadle, A., Collier, C., Johnson, D., Schmid, T., & Weather, R. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport And Environment*, *10*(3), 245-261. http://dx.doi.org/10.1016/j.trd.2005.04.001

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A: Policy And Practice*, *33*(6), 417-431. http://dx.doi.org/10.1016/s0965-8564(98)00022-6

Nosal, T. & Miranda-Moreno, L. (2014). The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts. *Transportation Research Part A: Policy And Practice*, *66*, 213-225. http://dx.doi.org/10.1016/j.tra.2014.04.012

Joseph Rose: Remembering Portland's disastrous Yellow Bike Project (photos). (2016). *OregonLive.com*. Retrieved 9 June 2016, from http://www.oregonlive.com/history/2016/01/portlands_disastrous_yellow_bi.html

Pronto. (2016). *Prontocycleshare.com*. Retrieved 28 May 2016, from http://www.prontocycleshare.com/faq

Provo. (2016). *Bl.uk*. Retrieved 28 May 2016, from http://www.bl.uk/learning/histcitizen/21cc/counterculture/assaultonculture/provo/provo.html

Pucher, J. & Dijkstra, L. (2003). Promoting Safe Walking and Cycling to Improve Public Health: Lessons From The Netherlands and Germany. *Am J Public Health*, *93*(9), 1509-1516. http://dx.doi.org/10.2105/ajph.93.9.1509

Pucher, J., Buehler, R., Bassett, D., & Dannenberg, A. (2010). Walking and Cycling to Health: A Comparative Analysis of City, State, and International Data. *Am J Public Health*, *100*(10), 1986-1992. http://dx.doi.org/10.2105/ajph.2009.189324

Rixey, R. (2013). Station-Level Forecasting of Bikesharing Ridership. *Transportation Research Record: Journal Of The Transportation Research Board*, *2387*, 46-55. http://dx.doi.org/10.3141/2387-06

Saneinejad, S., Roorda, M., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport And Environment*, *17*(2), 129-137. http://dx.doi.org/10.1016/j.trd.2011.09.005

Sears, J., Flynn, B., Aultman-Hall, L., & Dana, G. (2012). To Bike or Not to Bike. *Transportation Research Record: Journal Of The Transportation Research Board*, *2314*, 105-111. http://dx.doi.org/10.3141/2314-14

Shaheen, S., Cohen, A., & Martin, E. (2013). Public Bikesharing in North America. *Transportation Research Record: Journal Of The Transportation Research Board*, *2387*, 83-92. http://dx.doi.org/10.3141/2387-10

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia.*Transportation Research Record: Journal Of The Transportation Research Board*, *2143*, 159-167. http://dx.doi.org/10.3141/2143-20

Wang, X., Lindsey, G., Schoner, J., & Harrison, A. (2016). Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations. *J. Urban Plann. Dev.*, *142*(1), 04015001. http://dx.doi.org/10.1061/(asce)up.1943-5444.0000273

Wikipeida, List of Bicycle Sharing Programs, https://en.wikipedia.org/wiki/List_of_bicycle-sharing_systems Accessed in June 2016