Pedestrian Exposure to PM2.5 in Commercial Core, Seattle

Qi Chen

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

Master of Urban Planning

University of Washington
2020

Committee:
Chang-Hee Christine Bae
Edmund Seto

Program Authorized to Offer Degree:
Urban Design and Planning

University of Washington

Abstract

Pedestrian Exposure to PM2.5 in Commercial Core, Seattle

Qi Chen

Chair of the Supervisory Committee

Chang-Hee Christine Bae

Department of Urban Design and Planning

The thesis explores how environmental factors influence pedestrian exposure to PM2.5 by collecting and analyzing air quality data in Downtown Seattle, Washington. After reviewing related research, the influences from outdoor smoking, bus loading activities, and restaurants are chosen to be analyzed in this research. The study collected measurements along a fixed route surrounded by 1st Ave, 3rd Ave, Pike Street, and Yesler Way. After data cleaning, measurements of PM2.5 intensity, temperature, and other indicators a linear mixed-effects model. This model is applied because of clustered data by day. Data were collected for five workdays with a similar pattern, satisfying the assumptions of the linear mixed effect model. Analysis results indicate that outdoor smoking ranks top among all factors that have a significant influence on pedestrian exposure to PM2.5. In contrast, restaurants and bus loading are not considered as important. Therefore, this research motivates further policy on how to regulate outdoor smoking behavior through city regulations.

Table of Contents

List of Figures & Tables

Acknowledgements

I feel fully supported by professors, school resources, and my family and my friends. I would like to express my sincerest thanks to Chang-Hee Christine Bae and Edmund Seto. They reviewed my thesis and proposed many constructive suggestions for my project. I am also grateful to all the school resources provided by the University of Washington library. Staff in Odegaard Writing and Research Center provides are very enthusiastic in helping me with writing. Jackson Koch helps with my writing and my presentation. I would like to say thanks to him. There are not enough thanks to my family who give me all the financial support and love. Additionally, I appreciate all the accompanying from my friends. Without you, I could not finish this thesis.

Chapter1: Introduction

1.1 Health Damage from PM2.5

1.1.1 Current Air pollution

No single human being could breathe and live without air. Clean air is a public good which relates closely to the quality of our life. However, increasing human activities reduce the quality of our atmosphere. Air pollution is a big concern for human society especially in some developing countries with more heavy industries. Figure 1 is the global ambient air pollution map from the World Health Organization official website. The red area expresses the high concentration of PM2.5. In contrast, the area covered in green has better air quality. As the map shows, the air polluted countries are mainly from Asia, Africa, and South America where the economy is still under development. The United States Environmental Protection Agency (EPA) sets National Ambien Air Quality Standards (NAAQS) for “criteria air pollutants” under the requirement of the Clean Air Act. The six main air pollutants include Ground-level Ozone(O3), Particulate Matter (PM), Carbon Monoxide (CO), Lead (Pb), Sulfur Dioxide (SO2), and Nitrogen Dioxide (NO2). Among the six criteria pollutants, the Ozone and Particular Matter are two main threats in the USA.(United States Environmental Protection Agency, 2019) From the data provided by the WHO, 4.2 million people died from exposure to outdoor air pollution every year, and 91% the world’s population live in places where the air quality do not satisfy the WHO standards.

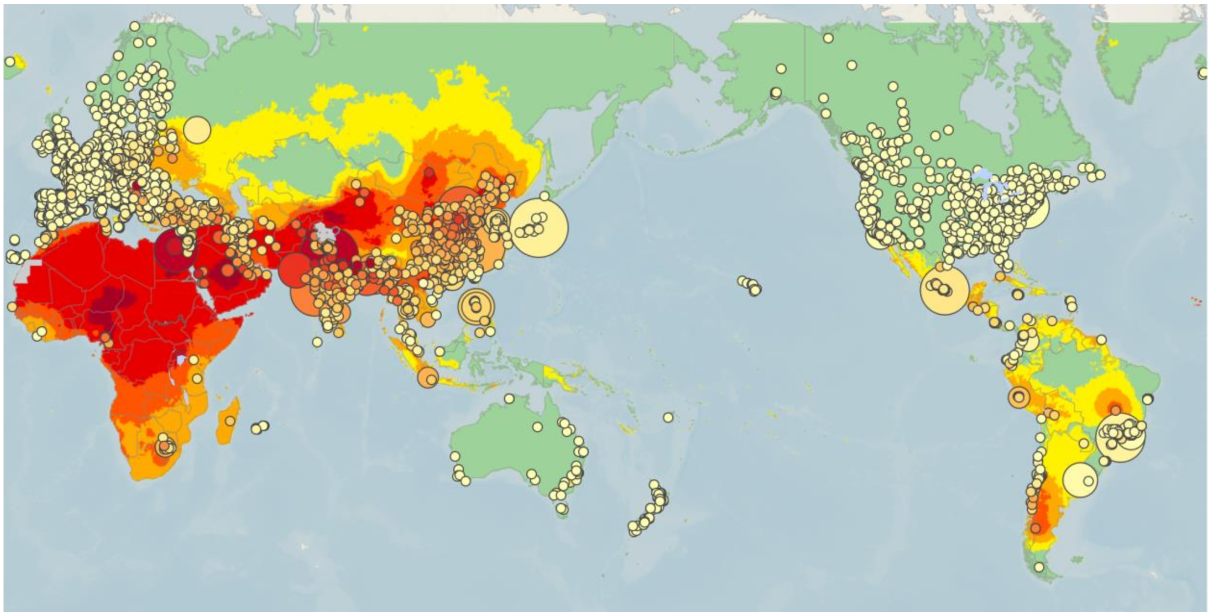


Figure 1. WHO Global Ambient Air Pollution Map

1.1.2 What is PM2.5?

The full name of PM is particulate matter which is one of the six criteria pollutants in the US. Particulate matters are tiny mixture of solid particulars and droplets. They are too tiny to be detected by naked eyes. PM2.5 is a kind of atmospheric particulate with a diameter < 2.5 micrometers, which is also called fine particulate matters. The size of PM2.5 is so small that its diameter is 33 times smaller than that of human hairs. (Bliss Air, 2014)

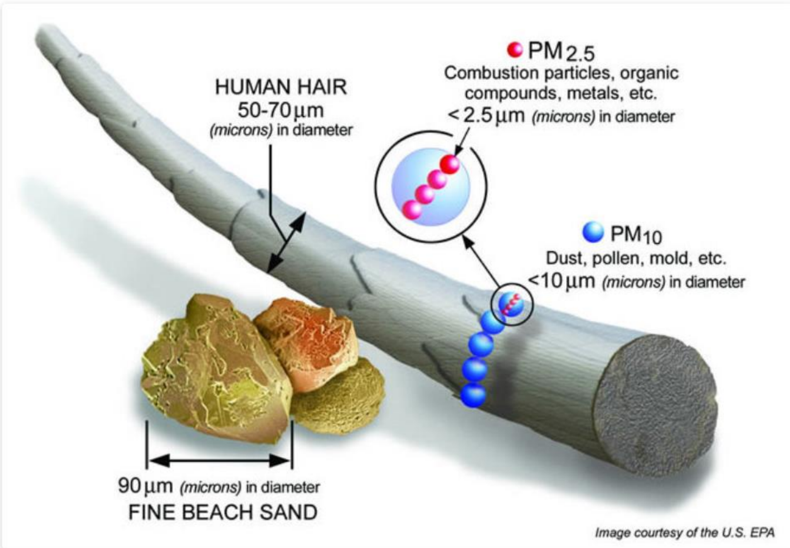


Figure 2. Comparisons of Particular Matters

(Bliss Air, 2014)

The sources of PM2.5, divided into stationary sources and mobile sources, vary by location. Construction, fuel burning and other human activities at stationary sites that produce PM2.5 are stationary sources of PM2.5. Traffic related air pollution is an example of mobile sources. As one of the mobile sources, traffic related pollution attracts lots of attention from researchers. PM2.5 could come from cars, trucks, and even the dust splashed by vehicles close to the road surface. The air pollution near roadside usually is worse than off-road locations. Besides air pollutions from traffic, PM2.5 could also come from burning solid fuel for producing electricity, cooking and heating. Moreover, the small sized fine particles are able to stay for a long time in the air and have some chemical combination with other particles in nature. Environmental tobacco smoke is another important source of air pollution.

1.1.2.1 How does PM2.5 affect human health?

The small size allows PM2.5 to stay longer in the air and makes it easier to concentrate together and influence human health in many ways. Fine particulates that are sized at 3% diameter of human hairs bypass the barriers set by the nose and throat and go deep into lungs and the circulatory system of the human being. (Bliss Air, 2014) The functions of arteries are limited by the concentration of the fine particulates. Therefore, the ratio of lung cancer and heart disease has a positive correlation with the PM2.5 index.(Department of Health, 2020) Other human diseases caused by particulate matter include but are not limited to irregular heartbeat, aggravated asthma, and increased respiratory symptoms. Besides influencing human beings, particulate matter can damage the soil and our water system in some degree. (United States Environmental Protection Agency, 2019)

PM2.5 (μg/m3)	Air Quality Index (μg/m^3)	PM2.5 Health Effects	Precautionary Actions
0 to 12.0	Good 0 to 50	Little to no risk.	None.
12.1 to 35.4	Moderate 51 to 100	Unusually sensitive individuals may experience respiratory symptoms.	Unusually sensitive people should consider reducing prolonged or heavy exertion.
35.5 to 55.4	Unhealthy for Sensitive Groups 101 to 150	Increasing likelihood of respiratory symptoms in sensitive individuals, aggravation of heart or lung disease and premature mortality in persons with cardiopulmonary disease and the elderly.	People with respiratory or heart disease, the elderly and children should limit prolonged exertion.
55.5 to 150.4	Unhealthy 151 to 200	Increased aggravation of heart or lung disease and premature mortality in persons with cardiopulmonary disease and the elderly; increased respiratory effects in general population.	People with respiratory or heart disease, the elderly and children should avoid prolonged exertion; everyone else should limit prolonged exertion.
150.5 to 250.4	Very Unhealthy 201 to 300	Significant aggravation of heart or lung disease and premature mortality in persons with cardiopulmonary disease and the elderly; significant increase in respiratory effects in general population.	People with respiratory or heart disease, the elderly and children should avoid any outdoor activity; everyone else should avoid prolonged exertion.

Figure 3. 24-Hour PM2.5 Standard

(United States Environmental Protection Agency, 2012)

Figure 3 shows how fine particles with different concentration levels influence the physical health of the human being and what people should do to protect themselves in different situations. Figure 3 indicates that the elderly and children are more vulnerable to terrible air condition and that people should refrain from outdoor activities when the PM2.5 is at an undesirable level. When PM2.5 concentration is less than 12.0 $\mu g/m^3$, the PM2.5 health effects are deemed little to no risk. When PM2.5 concentration is in-between 12.1 $\mu g/m^3$ to 35.4 $\mu g/m^3$, sensitive folks may feel uncomfortable. When PM2.5 concentration reaches the range of 35.5 $\mu g/m^3$ to 55.4 $\mu g/m^3$, more diseases including heart or lung disease and premature mortality have increasing likelihood of influencing sensitive groups. The air quality is unhealthy when PM2.5 concentration is within the range of 55.5 $\mu g/m^3$ to 150.4 $\mu g/m^3$ with the increased aggravation of diseases mentioned above. Sensitive groups should avoid any outdoor activities and everyone should avoid long exposure to the outdoor environment when the PM2.5 concentration is in the range of 150.5 $\mu g/m^3$ and 250.4 $\mu g/m^3$. These ranges of PM2.5 are for 24-hour average concentrations, and generally standards to not exist for shorter time averages.

1.2 Background of the study area

The study area is located at downtown Seattle. The rectangular study route in the right image of Figure 4 takes the researcher 40 mins to walk. In each data collection shift, the researcher started the data collection from the intersection of Pike St & 3rd Ave, walked along the 3rd Ave from the Pike St to Yesler St and then turned into the 1st Ave from Yesler back to Pike St, and finally went east to the starting point.

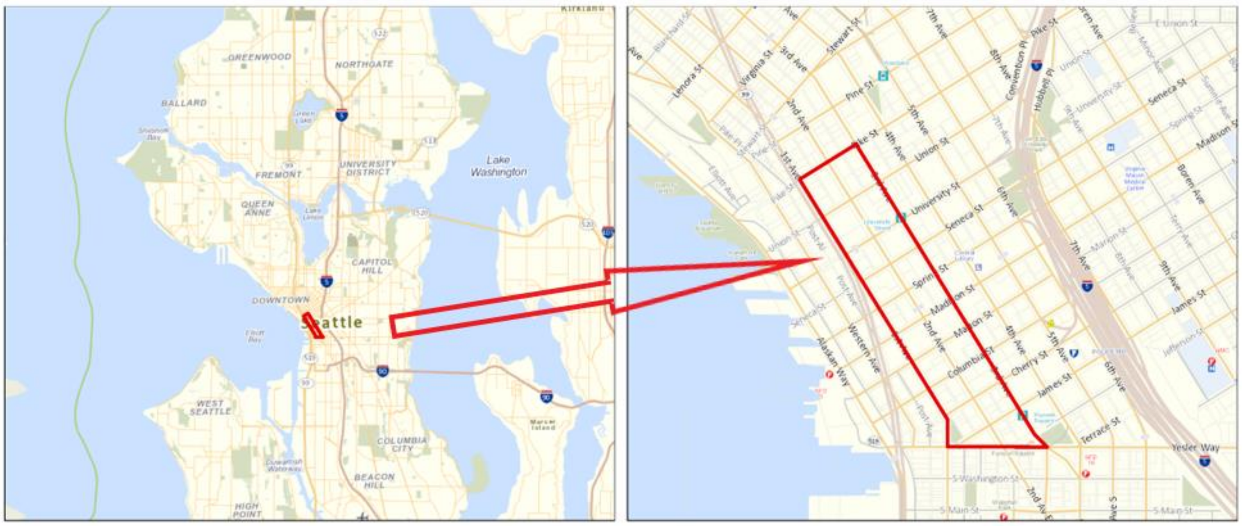


Figure 4. The Map of Study Route

The Study route connects the Pike Place Market and the Pioneer Square. Pike Place Market, established in 1907, embraced local seafood, flowers, and handcrafts from many small businesses. As one of most famous symbols of Seattle, Pike Place attracts many tourists from the whole world. Pioneer Square is the original settlement of Seattle with the longest history among neighborhoods of Seattle. (Wikipedia, 2020) Pioneer Square station of the central link light rail is located at the corner of Yesler Street and Third Avenue, which is a transit pivot for people who work in and visit downtown. Third Ave has been changed into a bus prioritized road since 2018. No private vehicles are allowed to enter Third Avenue from 9AM to 3PM seven days a week. The whole dataset used in the thesis was collected at midday (1-2pm) when all the private vehicles were banned on the 3rd Ave. As part of the heart of Seattle, the area surrounded by the study route is occupied by lots of mixed-use buildings shared by restaurants, shops, and office places. Pedestrians are commuters working in the neighborhood, visitors shopping at the local merchant stores, and tourists visiting various parts of Seattle Downtown like Pike Place Market and Seattle Art Museum.

Although the nearest fixed monitoring station is located at the 10th and Weller site, it is mainly impacted by the I-5 freeway traffic. The Seattle Beacon Hill Station is the next closest fixed station monitoring that records ambient background PM2.5 data. Seattle Beacon Hill Station locates at Jefferson Park of Seattle, which is 3.8 miles away from the study area. The graph below is published by the Puget Sound Clean Air Agency reflecting the fluctuations of PM2.5 each data collection days. The fixed station which is located at 4103 Beacon Ave S is 4.2 miles away from the study route, and the land use type is very different from downtown Seattle. Therefore, the PM2.5 data shown in the graph only serves as a reference estimate of the background air quality in the Seattle region. Note these are 1-hour averages in Figure 5, which are not the same as the 24-hour health-related categories in Figure 5.

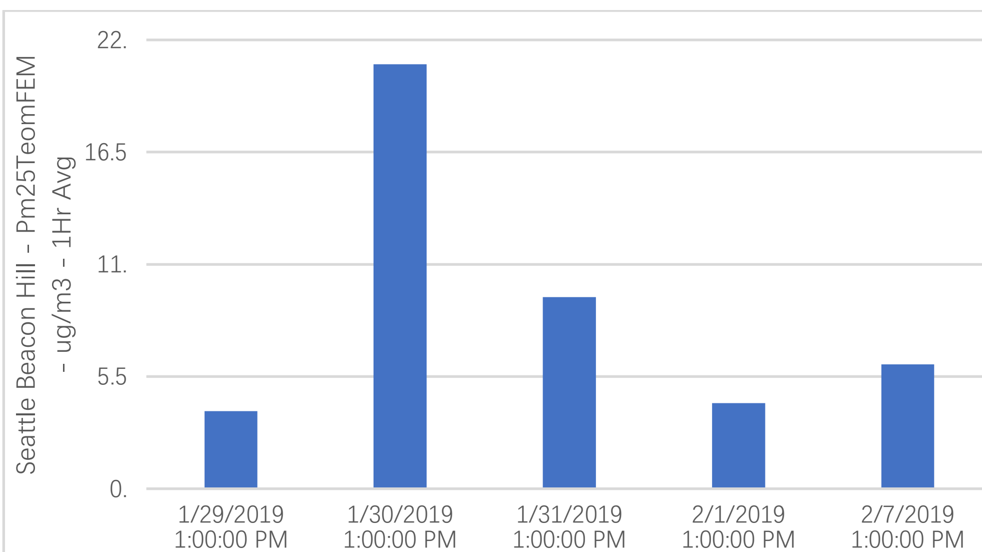


Figure 5. PM2.5 One Hour Average Concentration of PM2.5, Seattle Beacon Hill Station (Puget Sound Clean Air Agency, 2019)

1.3 Research purpose and thesis structure

Research on particulate matter has an important social significance, especially for developing countries with dense populations. As clarified before, exposure to high concentrations of PM2.5 increases the likelihood of heart and lung related diseases and are especially harmful to the elderly, children and other sensitive groups. Exposure to air pollution is also related to nonaccidental mortality. (Vanos, Cakmak, Kalkstein, & Yagouti, 2015) There are some researchers working on how different transportation modes people use affect their exposure to PM2.5. There is a finding that the people in the low speed of pedestrian travel are exposed to 9 times higher PM 2.5 than people taking an air-conditioned car, which is the mode with the lowest exposure in the research. (Goel, Gani, Guttikunda, Wilson, & Tiwari, 2015)

Bae and her students built a foundational work on analyzing pedestrian exposure to PM2.5 in different neighborhoods of Seattle. They research air pollution questions using qualitative methods and general linear regression models. (Bae, Sa, & Sinha, 2018) My research aims to learn how bus stops, smokers, and restaurants influence pedestrian exposure to PM2.5 in linear mixed model based on their findings. Linear mixed model considers the fixed effects from the three factors considered and also the random effects from other unrelated factors. With the more developed model and proper variable transformation, the thesis finds that smokers contribute a lot to pedestrian exposure to PM2.5 and surprisingly the bus stops have no evident influence on pedestrian exposure.

The thesis is written in the order of research from the background of topics and the study site to the data collection and analysis. Chapter 2 is the literature review of the three independent variables in this research and considerations on the performance of portable sensors. Chapter 3 covers methodology for the process of statistically analyzing the PM2.5 data, involving the data collection, data cleaning and model selection. After using linear mixed model, the variable interpretation and the model performance are in the chapter 4 results. Further improvements and potential policy implications are discussed in the chapter 5.

Chapter2: Literature Review

2.1 PM2.5 Effect on traffic mode/ Microenvironment

Paul Lioy & Clifford Weisel defined a microenvironment as a place where people doing different activities have the same exposure. (Lioy & Weisel, 2014) In other words, all the environmental elements influencing exposure could be treated as homogenous within one microenvironment. Under the definition of Lioy and Weisel, commuters taking different transportation modes are considered to be exposed to different transport microenvironments. Two Chinese researchers Weng and Jin in 2015 studied on air pollution in different transportation microenvironments, taking the concentration of formaldehyde as the research object. (Weng & Jin, 2015) Not surprisingly researchers found that the density of air pollution sources varied from one microenvironment to another. In this case the microenvironments are the inside of cars, buses, and bus stations. In order to understand pedestrian exposure to PM2.5, learning how a traffic microenvironment influences Particulate Matter 2.5 is the first step.

In 2001, three scholars conducted the first comprehensive study of personal exposure to PM2.5 in multi-traffic modes. (Adams, Nieuwenhuijsen, & Colvile, 2001) The study, taking Central London as the study area, was split into two separate parts. Firstly, the field study was accomplished by two teams in a 3-week duration when one team conducted the study in summer and the other focused on the winter season. Each team measured PM2.5 along fixed routes, and volunteers collected PM 2.5 data from four microenvironments, including bicycles, cars, buses, and the underground subway on weekdays. The second part of the field study was also a 3 week-long study on weekdays. The researchers recruited 24 volunteer cyclists recording their real commute when they took bicycles as their commute tool. The different emphasis shaped different research design. The first part conducted in fixed route highlighted the comparison of individual exposure in different seasons. The second field study only selected volunteers biking from offices to home and had more focus on personal exposure in real commuting routes in real life. The result of the first multi-seasonal study shows that in general summer has higher personal exposure to PM2.5 than winter among all three transportation modes (bicycle, underground subway, cars) excluding buses. Commuters in the second filed study served as a control group. From the result of the second study, there is no significant statistical difference between the “real commuters” and “data collectors in the first study". The second study proved the reliability of designing the first field study. Adams, Nieuwenhuijsen and Colvile's scholarship painstakingly built a framework of researching on personal exposure to PM 2.5. In the paper, they brought many researching angles like a comparison between routes, seasons, setting up a control group which inspired lots of researchers deep exploring in this field.

Air quality is a big concern in India as well as other developing countries. Four Indian researchers analyzed on-road PM2.5 data collected from 11 microenvironments along the 8.3-km fixed arterial route in Delhi. (Goel et al., 2015) The research took four months to conduct covering 8 different traffic modes which could be divided into traveling within open environments and within a closed environment. For example, walking, cycling, motorized two-wheelers, open-windowed car, and auto rickshaw are considered as open transportation environments. Meanwhile, air-conditioned cars and the underground metro are within closed environments. Metro bus on the street is a little special because the opening and closing of bus doors as well as some drivers tending to open the window influence the analysis results. Researchers estimated personal exposure based on the distance, and also took the travel speed into account. Not surprisingly, walking ranked the top mode getting most pollution, followed by cycle and buses. Air-conditioned cars with the help of ventilation had the lowest exposure to pollution. The study innovatively grouping microenvironments into open and closed environments raised concerns of pedestrian exposure to PM2.5 which ranked very high as an open microenvironment.

Taking Sacramento, California as a background city, four scholars researched the commuter exposure. (Ham, Vijayan, Schulte, & Herner, 2017) With the sampling of 161 commuters, they focus on multi pollution sources including PM2.5, BC, and UFP in six transport microenvironments. In the research, the ventilation settings of vehicles could effectively reduce the in-vehicle air pollution by up to 75%. Similar to the previous findings in Delhi, ventilation settings serve as a key factor in vehicle passenger exposure. But both of these sources didn't give further analysis on pedestrian exposure on the road and any policy implication from a planning perspective.

2.2 Main factors affecting PM2.5

2.2.1 Seasons

Many air pollution scholars have found that the concentration pattern of PM2.5 varies from seasons. In Goel’s study on Delhi’s personal exposure to PM2.5, the air quality in winter is worse than in summer.(Goel et al., 2015) The finding matches with what other researchers found in different cities. Through two weeks of PM2.5 sample collection in International District, Seattle, Sinha and Bae clarifies that the concentration of PM2.5 has a seasonal trend. From their dataset, Seattle has more polluted air in winter compared to air in spring. (Bae & Sinha, 2016) Similarly in January and June 2015, five researchers monitored PM2.5 and CO index at Hong Kong with TSI DustTrak and Q-Trak portable monitors. (Li, Che, Frey, Lau, & Lin, 2017) From Figure 6 below, the mean concentration of PM2.5 in winter obviously is higher than summer where the former ranges from 31ug/m^3 to 47ug/m^3 and the latter only ranges from 10 ug/m^3 to 23 ug/m^3. However, researchers from Britain get a opposite conclusion from their dataset. (Adams et al., 2001) In their study, all transportation modes have lower PM2.5 exposure level in winter than summer. The logic of how seasons influence on PM 2.5 exposure is out of scope of this research. But it is reasonable to think meteorological factors could play a role in the PM 2.5 levels such as the wind and rain, etc.

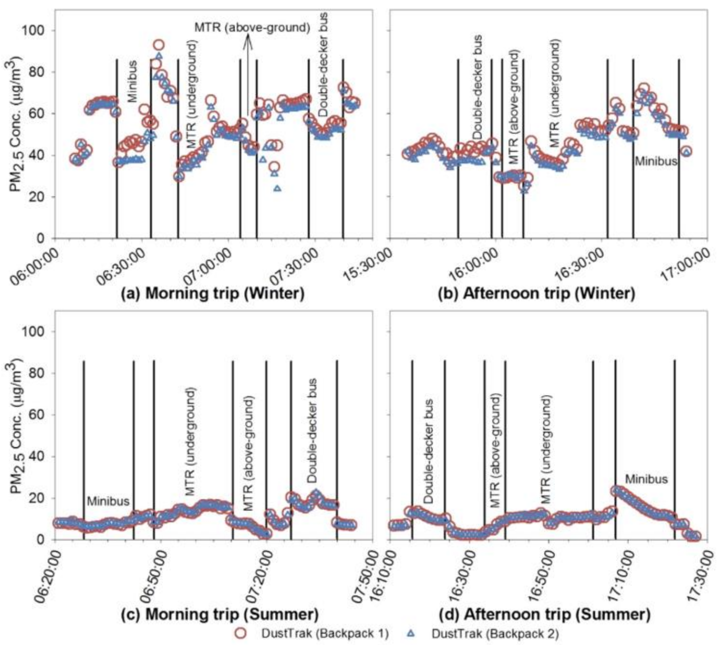


Figure 6. The Apparent Difference of PM 2.5 Concentration between Winter and Summer in Hong Kong

(Li et al., 2017)

2.2.2 Cooking

In 2018, Bae and Sa, at the University of Washington, collected pedestrian exposure to PM2.5 in South Lake Union and Pike Place, Seattle where her study area geographically covers the research area of this thesis.(Bae et al., 2018) They found that restaurant-related land use has a positive relationship with PM2.5 exposure. PM2.5 was high when Bae carried a portable sensor close to food trucks. It makes sense because it has been proved that using solid fuel to cook has a negative impact on air quality. (Carter et al., 2016) Although using solid fuel as energy is more common in the developing countries (Hartin, Seto, & Yost, 2015), restaurants in Seattle mostly use natural gas to cook. In the study of Particulate Matter from Cao and Thompson, they found that cooking activities contribute most in the personal exposure to PM2.5. (Cao & Thompson, 2017) But pedestrians don't partake in cooking activities directly, and air filtering facilities of buildings may induce the higher concentration of PM2.5. It is not clear how restaurants influence pedestrian exposure to PM2.5, but it is an important factor that needs to be researched. As a result, restaurants become one of main independent variables in this research. Aside from burning of solid fuels, cooking also aerosolizes fats and oils, which contribute to PM2.5.

2.2.3 Outdoor Smoking

It is surprising to find that outdoor smoking is an inevitable factor influencing pedestrian exposure to PM2.5 from Bae's study. (Bae et al., 2018) In her dataset collected in downtown Seattle, the PM2.5 peaks often related to smokers near the bus stops or groups of smokers gathering at the corner park of the south end of Third Avenue. Bereitschaft conducted a study of pedestrian exposure to PM2.5 in 2015. He monitored PM2.5 levels in six mixed-use sites in Ohama in a 2 Km fixed route. He found that 56% pollutions are from background concentration which is monitored by nearby fixed monitoring station, and short PM2.5 peak concentrations are from vehicle emissions as well as tobacco smoke. (Bereitschaft, 2015) However, there are few other scholars working on pedestrian pollution exposure caused by outdoor smoking. Smoking is only treated as a source of indoor air pollution in most research.

2.2.4 Bus Stops

Bus lines are located along the arterials of cities to provide a low-cost and convenient public transportation choice. Some researchers make efforts to understand the relationship between commuter exposure to PM2.5 with waiting at bus stops. After collecting data from 28 bus shelters in the urban center of Buffalo, New York, researchers found four key factors determining the concentration of PM2.5. (Baldwin, David, Stinson, & Park, 2010) They are the time of the day, locations where passengers wait, land use types near the bus shelter, and the presence of smokers near the bus stops. Not surprisingly the more confined the bus shelters are, the higher air pollution commuters are exposed to. (Narotzki, Reznick, Mitki, Aizenbud, & Levy, 2014) However, people waiting close to bus shelters are not exactly pedestrians. Therefore, how bus stops influence pedestrian exposure to PM2.5 was not well defined in the Narotzki's research.

2.3 Portable Sensors

Portable sensors are broadly applied in environmental research with the advantages of low cost and portability. But the reliability and prerequisites of applying portable sensors need to be discussed. There are several studies finding that air pollution detected by portable sensors is usually higher than fixed monitoring stations. (Bae & Sinha, 2016) Part of the reason may be that portable sensors tend to be closer to pollution sources while some fixed stations are located far from pollution sources. Another factor may be that the concentration of humidity interferes the detection of PM2.5. (Jayaratne, Liu, Thai, Dunbabin, & Morawska, 2018) Fixed monitoring stations have dryers or heaters inside to remove liquid in the air, which is not harmful to human health. However, low-cost sensors lacking dryers may detect concentration of humidity as particulate matter. In consequence, portable detectors may report a denser concentration of PM2.5 than exists in the real situation. Nevertheless, if the humidity density is under 75%, the influence of humidity on measuring PM2.5 is not considered to be significant. (Jayaratne et al., 2018) Moreover, some researchers compared portable monitors with fixed monitory stations using three different portable monitors. (micro-aethalometer AE51, DiscMini, Dusttrak DRX) in 2015. They found relative differences between portable and fixed stations of less than 20%, and argued that the reliability of portable sensors should be considered high. (Viana et al., 2015)

Many researchers have applied portable sensors in their environmental studies. Cao and Thompson conducted portable sensor study in 2016. They developed a field-portable device to detect PM2.5 which combines an Arduino microprocessor, SD card reader, and an optical dust monitor. They had several individuals carrying the device and collecting data in Lubbock, TX and Atlanta, GA. (Cao & Thompson, 2017) In another study, three researchers used seven Portable University of Washington Particle (PUWP) monitors, low-cost particulate sensors using Shinyei PPD42NS sensors. These sensors are able to identify potential PM2.5 hotspots in high PM2.5 concentration environments. (Gao, Cao, & Seto, 2015)

Chapter3: Methodology

This thesis uses a similar approach as Bae and Sa to answer how environmental factors influence pedestrian exposure to PM2.5 by collecting and analyzing air quality data in Downtown Seattle. (Bae et al., 2018) The methodology chapter, as described in Figure 7, is split into three parts: Data Collection, Data Cleaning, and Data analysis. AirBeam2, Google Maps, and Go Pro were used to collect data. After data collection, Excel, RStudio, and ArcGIS was applied as tools to clean data. RStudio was used for data analysis in this research.

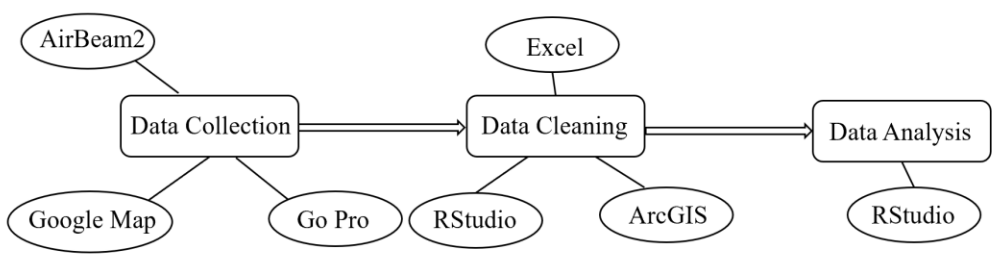


Figure 7. Data Processing Flow Chart

3.1 Data collection

The data collection has two parts: physical field collection and online data searching. The whole dataset used for this research was collected in winter when Seattle's weather was not consistent. Some days were windy while some days saw rain. In order to avoid unnecessary interference from factors like seasons and weather, the whole dataset was collected during similar weather within the 2-week period (Figure 8). The researcher gathered real-time PM2.5 walking along the study route carrying three devices, smartphone, Go Pro camera, and AirBeam2 (Figure 9,Figure 10). PM 2.5 indicators were collected on five weekdays, Jan29, Jan30, Jan31, Feb01, and Feb07, 2019. The interferences of meteorological factors need to be considered when weather conditions and temperatures are different every day. However, in this case the differences in both weather and temperatures are not significant concerns. All the PM2.5 data was detected around 1 PM on workdays. Based on the meteorological data for collecting days, the temperature at 1 PM ranges from 54°F to 37°F. Regarding the weather, four of five days were sunny or cloudy, which were rare in Seattle's winter, but ideal for air pollution study. The geographical information about bus stops and restaurants in the study area are manually exported from Google Map by the researcher.

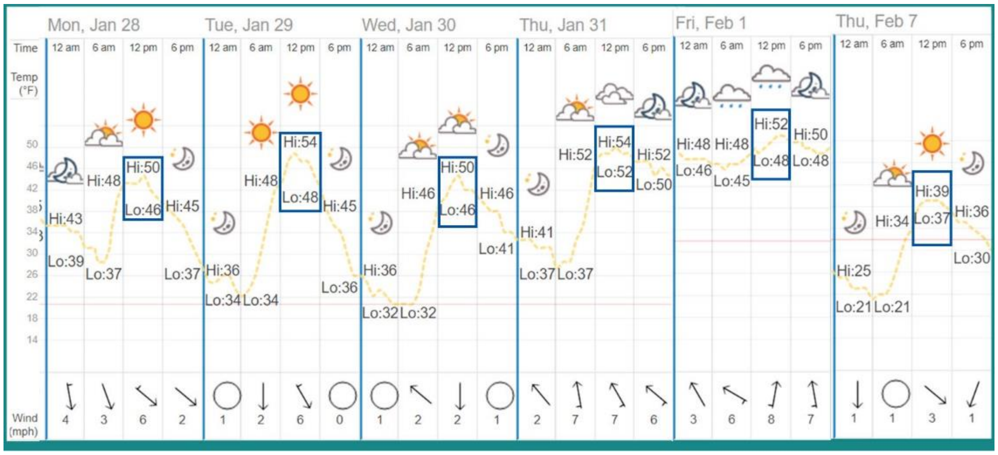


Figure 8. Meteorological Records of Data Collection Days

Figure 9. The image of Go Pro 6

(“GoPro Official Website - Capture + share your world - hero 6,”)

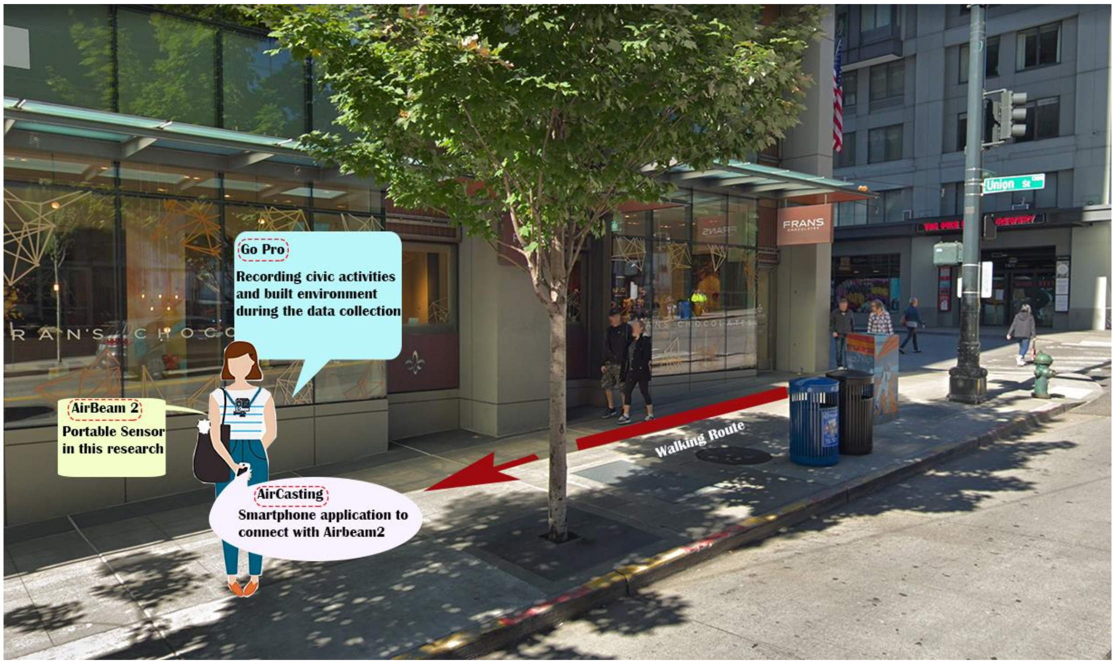


Figure 10. The Image of Collecting Data on 1st Ave

3.1.1 PM 2.5

As a numeric dependent variable, high quality PM2.5 data is a prerequisite to get accurate results. The AirBeam Network consisting of a portable sensor, a sharable online platform, and an Android application well supports measuring the pollution index. In the field, the researcher can read the real-time data in line charts and maps from the Android Application connecting with the portable sensor via Bluetooth. In the meantime, the AirCasting Platform automatically updates the dataset online. The dataset is open to the public, which everyone could download as a CSV file. None of the humidity data detected from AirBeam2 in this research hit the bar of 75%. Therefore, AirBeam2 could be considered as one kind of reliable detector. However, the sensor data were not formally compared or calibrated to other reference PM2.5 instruments. So for the purposes of this study, the PM2.5 measurements should be considered only a semi-quantitative indicator of PM2.5 exposure variation, rather than true PM2.5 concentrations.

3.1.1.1 AirBeam2

AirBeam2 is a low-cost, handy portable sensor used in this research to measure real-time pollution exposures. The market price of AirBeam 2 is $249. It is able to measure PM1, PM2.5, PM10, relative humidity, and temperature two times per second on average. The handy portable size which is only $5.21'' \times 3.87'' \times 1.10''$ makes the sensor easy to be put into a pocket.

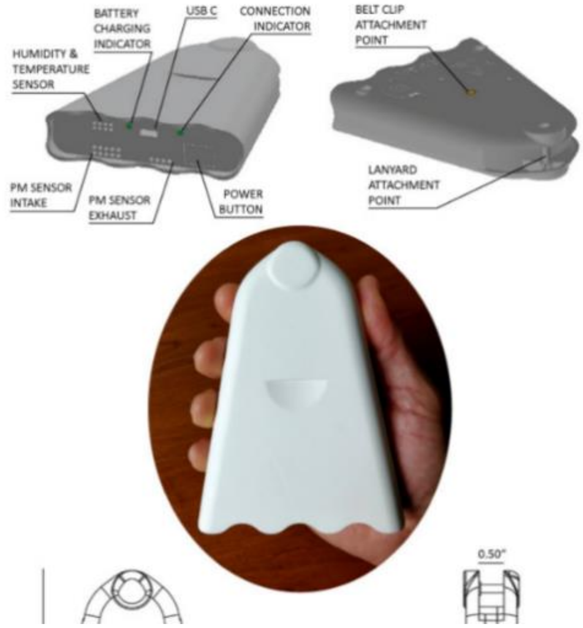


Figure 11. AirBeam 2 Model

3.1.1.2 AirCasting Mobile Application

The application showing in Figure 12 is developed to support AirCasting Platform with powerful data visualization function. However, the application requires an android phone to install, and is not available in Apple Store.

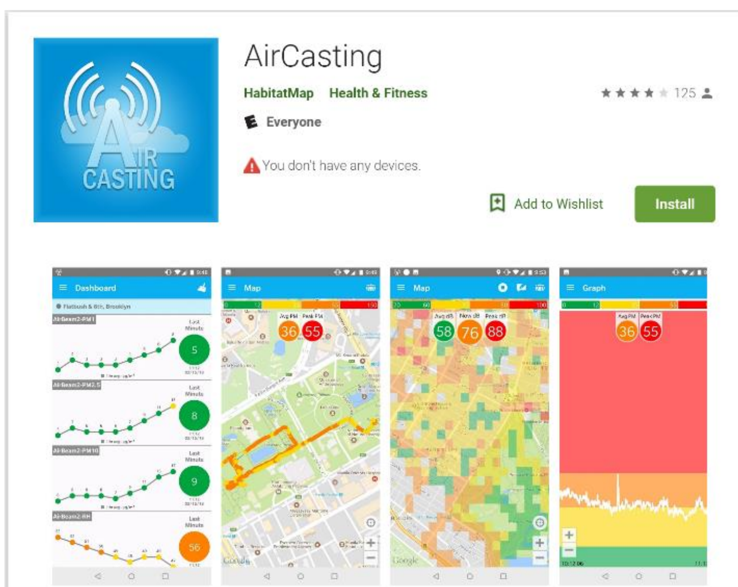


Figure 12. AirCasting Application

3.1.1.3 AirCasting Platform

The open source website provides recording, mapping, sharing pollution measurement services. Researchers can download data uploaded by any other portable sensors for free. Moreover, customers can personalize the data visualization by adjusting the legend on the left side of the webpage.

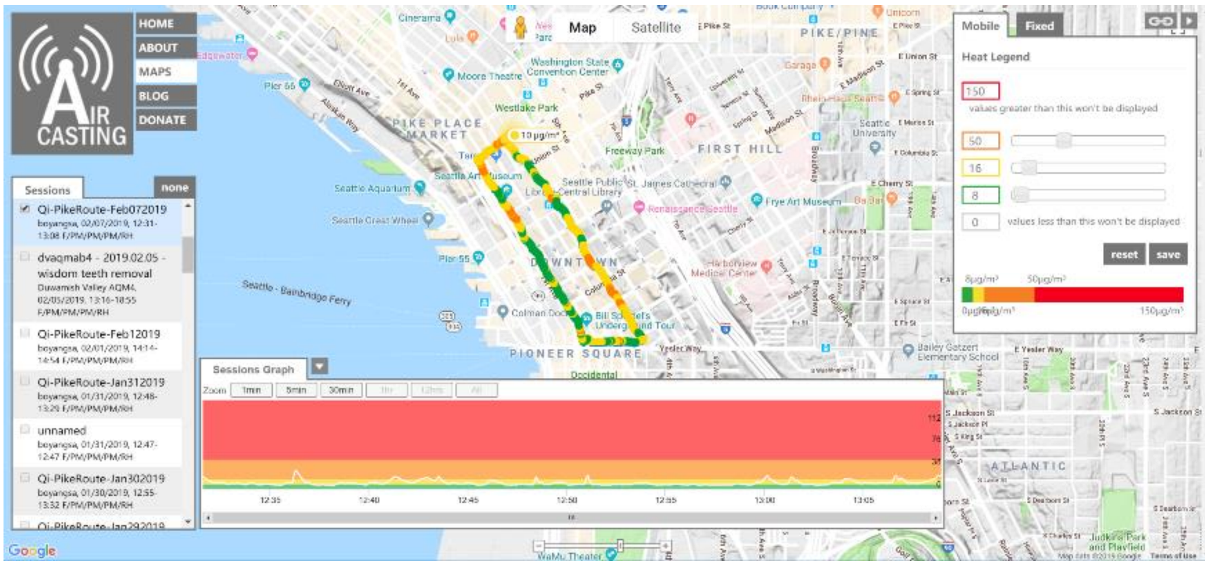


Figure 13. AirCasting Platform

3.1.2 Bus Stops and Restaurants

The counts of bus stops and restaurants near the studying route are straightforward to collect without measuring error. There is a function called “My Map” in the Google Map, where users can identify places manually and export maps in KMZ file. Because the researcher didn’t find any online GIS data of buses and restaurants within the study area, the thesis uses the layers of buses and restaurants downloaded from “My Map”. Bus Stops are places where passengers board or alight from a bus. However, the definition define restaurants in environmental studies is less clear. The intention of the research is to understand the relationships between restaurants and air pollution. From the literature review part, it was seen that outdoor cooking can have a positive influence on PM2.5. (Carter et al., 2016) In this case, cafes, like Starbucks and Peet’s Coffee, were excluded from the restaurant category, and we assume their cooking with oils and fats is minimal to none.

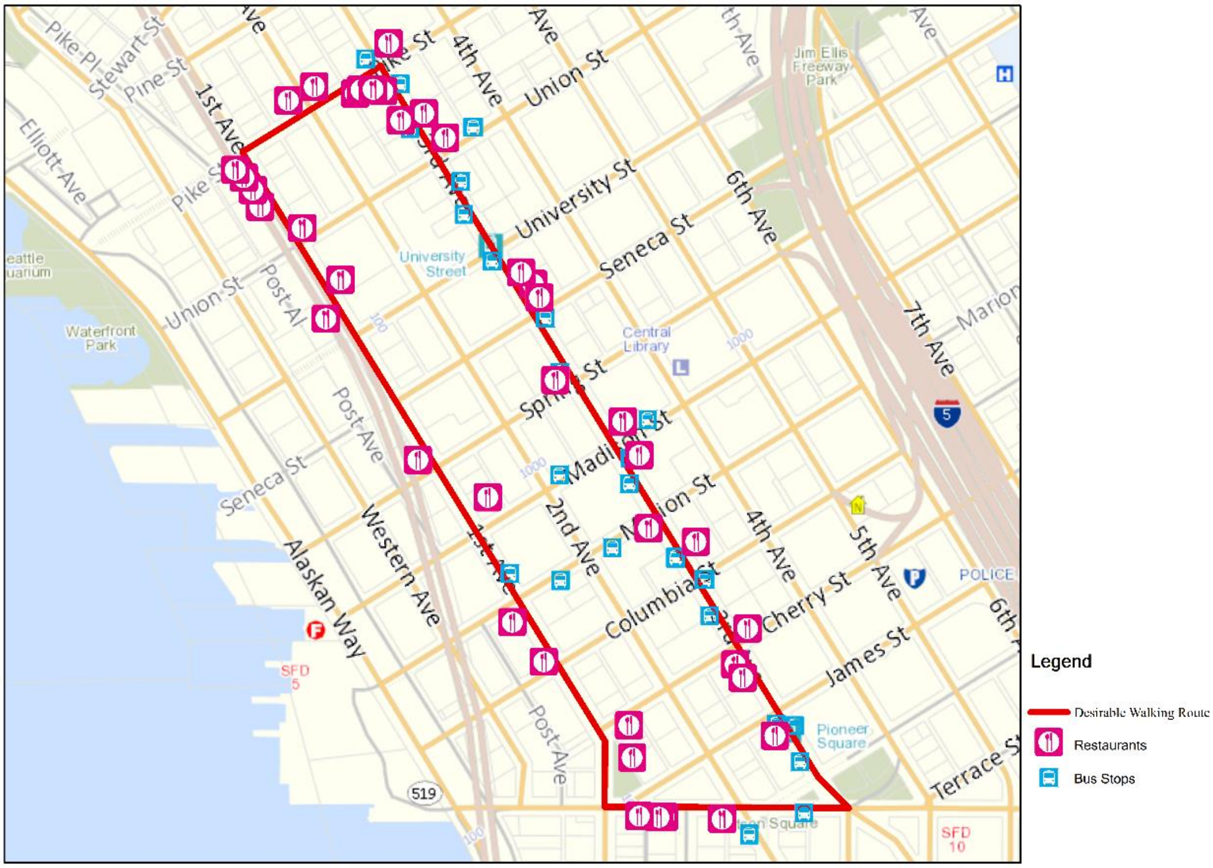


Figure 14. The Distribution of Buses and Restaurants in the Study Area

3.1.3 Smokers

As one of the three independent variables in this research, outdoor smoking plays an inevitable role in the increasing PM2.5 index. (Bae et al., 2018) However, there is no clear interpretation of how important the role is. It is not reliable to count smokers based solely on personal field observation, as the researcher's vision can easily be blocked or otherwise imperfect during a 40-min walk in a dense urban environment. The combination of Go Pro and Airbeam2 can do a better job of counting smokers than a human being. Go Pro is a type of portable video camera with a long-lasting battery. The reason for choosing Go Pro is threefold. First, the Go Pro's viewing angle at 140° is wider than a human's, at 114°. In other words, Go Pro can catch more information than humans without turning the body around. Second, all the civic activities and built environment characteristics will be recorded in the video. Last but not least, records could be saved for a long time and sharable. It's easy to pause the video at any moment needed for further research.

In this study, the Aircasting Platform provides possibilities to find smokers in a short time. Firstly, the researcher looked for all the PM 2.5 peaks of the charts, which were automatically created by the Aircasting Platform, and then recorded the time when PM2.5 peaks appeared. Going back to Go Pro records, the researcher captured screenshots at the specific moments which matched the PM2.5 peaks. Smokers were counted from the screenshots of the Go Pro videos. The final step was updating the origin dataset downloaded from the AirCasting Platform with new smoker data. We recognize that this creates a biased data set by only looking at the presence of smoking during peaks, and not during non-peak PM2.5 periods, which we acknowledge in our discussion below.

3.2 Data cleaning

AirBeam2 monitors multiple pollution sources, including PM1, PM2.5 as well as PM10 once or twice times per second. In the meantime, the sensor records geographic locations through longitude and latitude. When it comes to a 40-minute data collection procedure per day, the dataset downloaded from the AirCasting Platform is a CSV file containing over 2000 records each day. In total, the whole PM2.5 dataset has 10,000 geographic points in 5 weekdays. Each GPS point is considered as one sample in the research. Besides filtering the PM2.5 index and GPS information, combining buses, restaurants, and smoker counts with each GPS point is the main task of data cleaning.

3.2.1 PM2.5 & Relative Humidity

The original dataset was massive and disordered. Because AirBeam2 monitors multiple pollution sources including PM1, PM2.5 as well as PM10 once or twice times per second. The daily dataset downloaded from the AirCasting Platform has more than 2000 samples blending with other useless information. As with particulate matters, the longitude and latitude of geo-points were recorded more than one time each second. Because the researcher as a pedestrian was unlikely to walk several steps within one second, when the geolocation data is visualized, there are many geolocation points overlapping together since they share latitudes and longitudes. The cleaning of PM2.5 and relative humidity has two steps: separating the dataset by second and deleting rows with NULL value. The first step is to group PM 2.5 data by every second in RStudio and then take the mean value of every second as the final PM2.5 value. After deleting redundant columns and sorting rows in reasonable order, the original csv file was converted into a well-organized xls file. Figure 15 shows PM2.5 data collected on Jan 29; the red line shows the study routes.

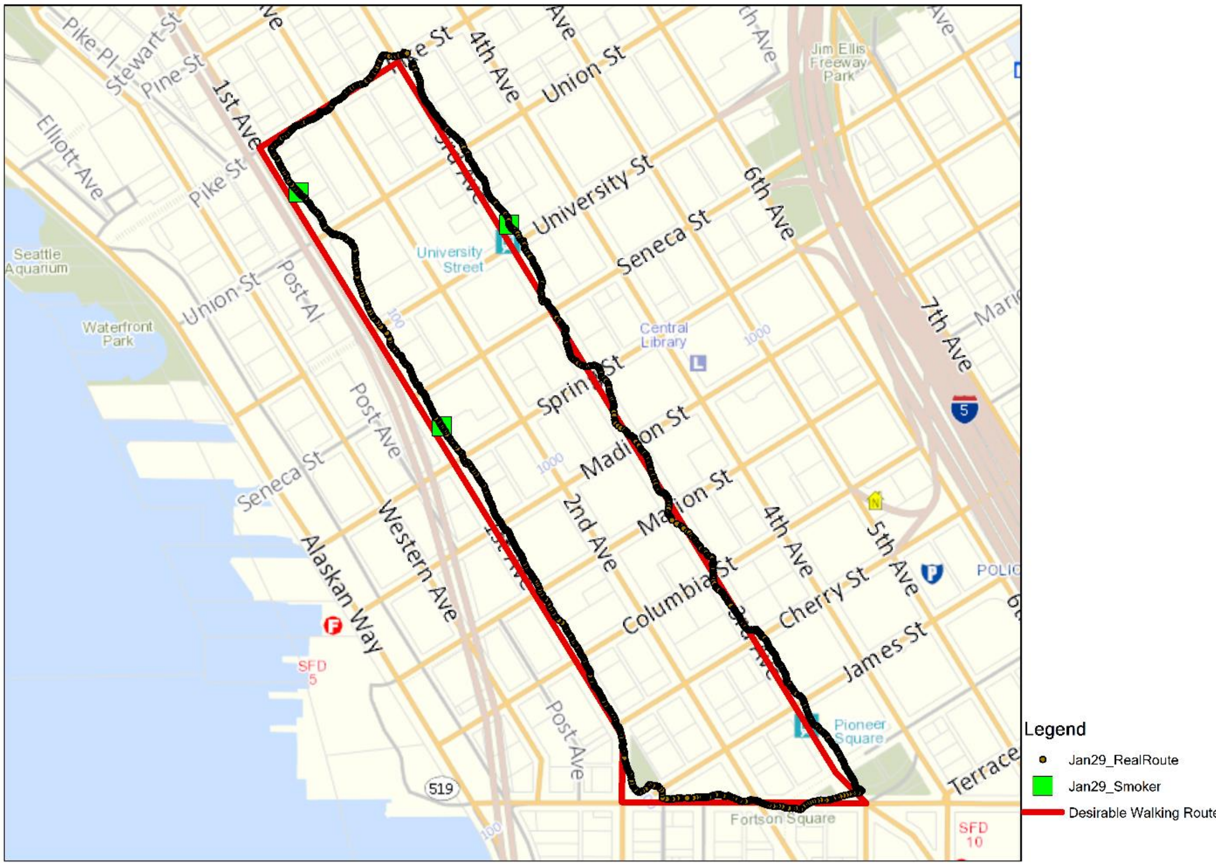


Figure 15. The Studying of PM2.5 Data Collected in Jan29,2019 Showing as Separated Dots

3.2.2 Bus Stops and Restaurants

After separating the original PM2.5 dataset second by second, the whole table downloaded from the AirCasting Platform consists of 10,000 geographic separated points collected from 5 weekdays. How to assign each geographic point with certain values associating with bus stops is a pivotal question.

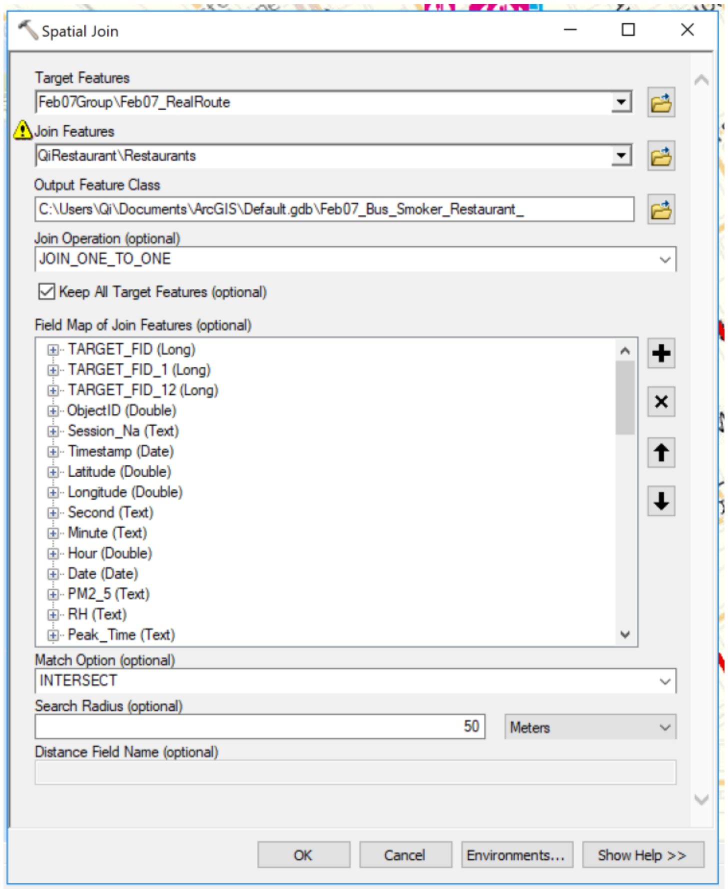


Figure 16. Assigning Restaurant Counts As a Value to The Study Units via ArcGIS

Commuters are exposed to PM2.5 while they are waiting at bus stops. (Ngoc, Kim, Bui, Park, & Lee, 2018) As mentioned in the literature review, bus shelters, land use near the bus stops, and other factors work together to influence commuter exposure to air pollution. In this research, the distance from bus stops is assumed as the only measurement for the influence from bus stops. In general, the longer the distance from bus stops is, the less PM2.5 bus loading caused. In the research, the area influenced by bus stops is set as a circle with a radius of 20m. The common block size at Pioneer Square is $120m \times 170m$ (as measured from Google Maps). Each sample receives a value of bus counts based on how many bus stops influence circles it falls within. Similarly, every geographic point has a value associated with restaurant counts. In this case, the radius of the circle influenced of restaurants is 50m (and again coffee and bakeries are not included).

3.2.3 Smokers

Ott and five other researchers measured non-smokers ’exposure to PM2.5 near a smoker. (Ott et al., 2014) They found that there is a linear relationship between non-smoker exposure and their distance from smokers. From this research, Ott provided three sets of PM2.5 related data. The mean PM2.5 individual exposures are 59, 40, 28 $\mu g/m^3$ respectively at distances as 0.5m, 1.0m, 1.5m from smokers, while the PM2.5 concentration in the background is 1.7$\mu g/m^3$. After running a simple linear regression model based on the three sets of data, 2.3m is an ideal boundary of the smokers’ area of influence. However, in real life, smokers are not stagnant. Some pedestrians smoke while they are walking. Smoking also is a kind of social activity. It’s not rare to see two or three smokers gathering together to smoke and chat. Thus, the real influence area can be larger than the circle with a radius of 2.3m. For the convenience of researching, taking influencing radius as 10m is reasonable.

```
Smoker Regression Results
==============================================
                         Dependent variable:
                     ---------------------------
                                  y
----------------------------------------------
x                             -31.000*
                               (4.041)

Constant                      73.333**
                               (4.365)

----------------------------------------------
Observations                      3
R2                              0.983
Adjusted R2                     0.967
Residual Std. Error        2.858 (df = 1)
F Statistic              58.837* (df = 1; 1)
==============================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Figure 17. Simple Linear Regression on Influence of Smoking

3.3 Model Selection

3.3.1 Multiple Linear Regression Model

Multiple linear regression model is a fundamental regression analysis to analyze the relationships between several independent variables and the response variable. It is also a good start to understand the statistical characteristics of the whole dataset. Bae chose the generalized linear regression model to analyze pedestrian exposure in 2018. (Bae & Sinha, 2016) In her model, land use type related to restaurants, bus stops, and smokers were independent variables. Applying the multiple linear regression model, the output from R default package is showed below. Among all three variables involved, only smokers has a positive influence on pedestrian exposure to PM2.5 while other two variables have p-values larger than 0.05. The estimated coefficient of Smoker Count is 9.006. The interpretation of the multiple linear regression model is that one unit increase of smokers will increase 9.006 $\mu g/m^3$ PM2.5 when other variables keep constant. However, the adjusted R Square is only 0.05367 which means the current model only could explain 5.367% of the variance in PM2.5. The low adjusted R Square indicates that the multiple linear regression doesn't fit the dataset well.

	Dependent variable:
	PM2_5
SmokerCount	9.006***
	(0.384)
BusCount	0.033
	(0.147)
RestaurantCount	0.068
	(0.048)
Constant	13.539***
	(0.077)
Observations	9,702
R2	0.054
Adjusted R2	0.054
Residual Std. Error	5.059 (df = 9698)
F Statistic	184.409*** (df = 3; 9698)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 18. Multiple Linear Regression Output

There are a few assumptions in the multiple linear regression model, which are linearity, homoscedasticity, independence, and normality. From the four diagnostic plots, the four assumptions could be checked.

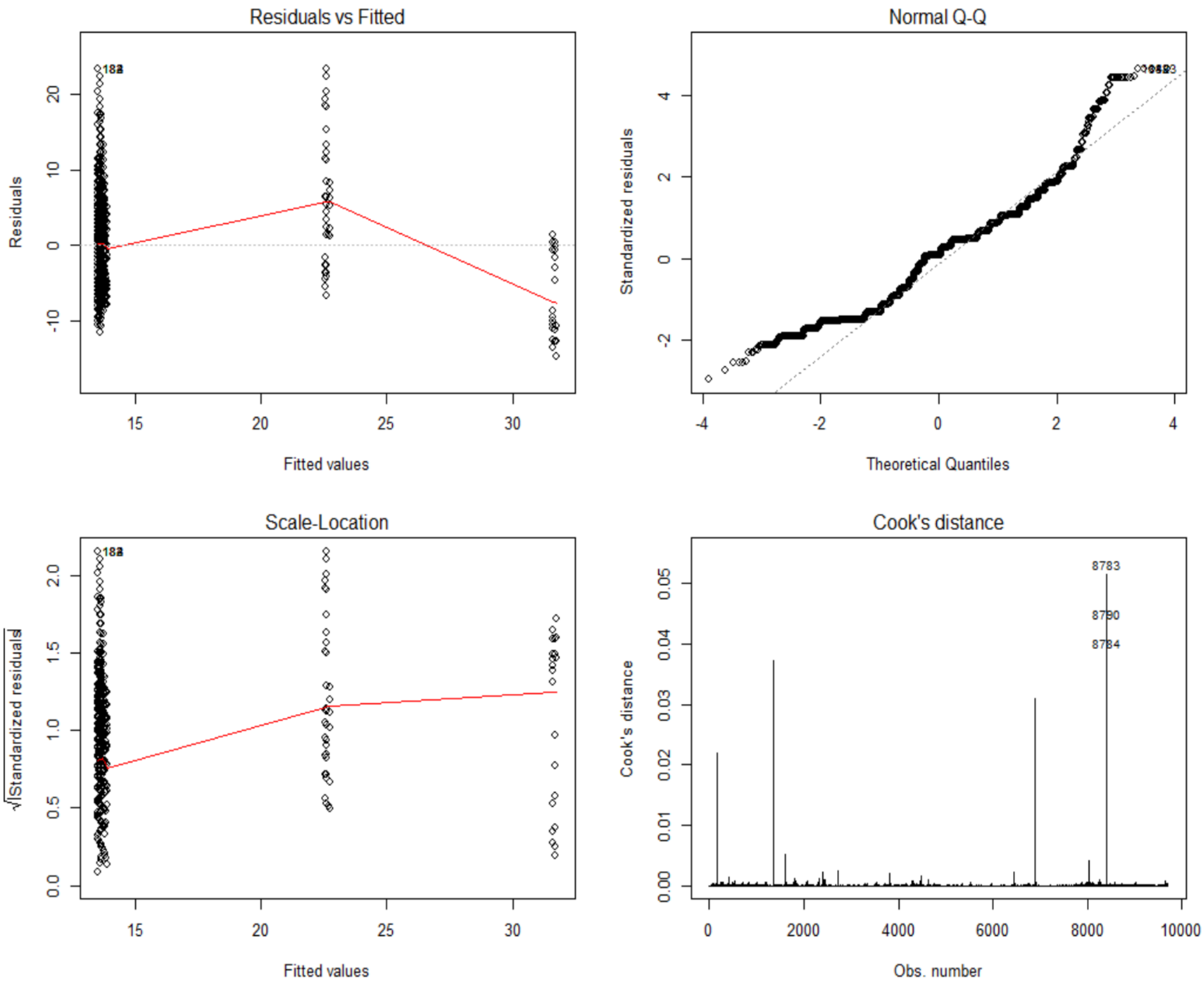


Figure 19. Diagnostic Plots of the Multiple Linear Regression Model

The Residuals vs Fitted plot indicates the linearity assumption is basically satisfied. However, the fitted value is cluttered on the value around 12, 23 and 33 in the Residuals vs Fitted plot. It's not sure why they are clustered in this way. One reasonable guess could be that 23 and 33 are most common value of PM2.5 peaks, and meanwhile 12 is the value close to the background PM2.5 value. It is hard to say that the response variable is normally distributed from the Q-Q plot. The homoscedasticity is not fully satisfied from the Scale-Location. With the low Cook's distance, it is fair to determine that all the observations are independent. The AIC of the model is 58995.98. More advanced and potential variable transformation are worthy of discussion.

3.3.2 Linear Mixed Effects Model

A linear mixed effects model is an extension of linear regression models. Incorporating random effects and fixed effect, the linear mixed model is a better choice than multiple linear regression model for panel data. (Consulting, 2019) Frees systematically introduced the longitudinal data and how to analyze the longitudinal data with different models. (Frees, 2004) Among the models mentioned before, the linear mixed effects model is an effective approach for data with inherently correlated/clustered measurements.

Before the application of the linear mixed model, the concepts of fixed and random effects, together with why PM2.5 data is longitudinal data, need to be discussed. A linear mixed effects model is a hybrid of a regression model and a time series analysis model — the regression model indicates the relationships between different subjects and the time analysis model focus on explaining and predicting fewer subjects over time. In general, fixed effects, similar to the regular regression models, represent the relationships between independent variables the researcher self-determined. However, if the whole data collection repeats over time, there will be some random effects interfering differently in each unit. Most times, those intervening variables are invisible and not factored into the research design. To compensate, random effects in the linear mixed model are used to represent the undefined interferences showing over time.

In this research, the whole data collection process lasted five days. As mentioned in the literature review, there are other factors influencing pedestrian exposure like temperature, wind level. Those factors differed every day and were not considered as independent variables. It is natural to group the whole dataset containing 9702 samples into 5 different days.

$$\overbrace{Y}^{9702\times1} = \overbrace{\underbrace{X}_{9702\times4}\;\underbrace{\beta}_{4\times1}}^{9702\times1} + \overbrace{\underbrace{Z}_{9702\times5}\;\underbrace{\mu}_{5\times1}}^{9702\times1} + \overbrace{\epsilon}^{9702\times1}$$

Outcome Variable Fixed Effects Random Effects Residuals

Figure 20. Equation of Linear Mixed Model in the Research

Figure 20 shows the Linear mixed model in this research. Y represents the outcome variable which is PM2.5 levels in this case having 9702 samples. X and Z respectively are coefficients of fixed effects and random effects. β represents all three independent variables and one intercept influencing on the fixed effects, and μ represents all the random effects grouped by 5 days while € is a 9702×1 column vector of the residuals.

Similar to the multiple linear regression model, there are a few assumptions in the linear mixed model. We assume that the random effects are normally distributed with mean zero. The residual error € also follows the normal distribution. However, the linear mixed model is proved still robust even if the residual error and the random effects don’t follow the normal distribution. The estimated β keeps unbiased. (Bartlett, 2014)

Chapter4: Results

4.1 Variable Description

4.1.1 PM 2.5

The remaining dataset contains 9702 samples after cleaning PM2.5, removing values of zero. The mean PM2.5 over five days is 13.73, while the median with a little higher value is 14.00. The maximum of PM2.5 is 42, and the minimum value is 2. Figure 21 shows the descriptive statistic characteristics of PM 2.5. The variation of median and mean value in each day reflects different PM2.5 concentration.

* **2019/1/29**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2	6	7	9.398	10	42

* **2019/1/30**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8	14	15.25	15.8	18	31

* **2019/1/31**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12	16	16	17.21	19	33

* **2019/2/1**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10	14	16	15.99	17	46

* **2019/2/7**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5	7	9	9.635	10	31

Figure 21. The Summary of PM2.5

The mean PM2.5 varies on different days. On Jan 29 and Feb 7, PM 2.5 were relatively low, which were respectively 9.398μg/m^3 and 9.635 μg/m^3. The other three days had higher values which were 15.8 μg/m^3 on Jan 30, 17.21 μg/m^3 on Jan 31, and 15.99 μg/m^3 on Feb 1. Compared with the PM2.5 concentration detected by the Beacon Hill Station, the value of PM2.5 on 1 PM from Jan29 to the last study day respectively are 3.8 μg/m^3, 20.8 μg/m^3, 9.4 μg/m^3, 4.2 μg/m^3, and 6.1 μg/m^3 on average. PM2.5 concentration monitored by the fixed station is lower than the PM 2.5 value from AirBeam2 on average except Jan 30. Figure 23 shows the comparison of PM2.5 concentration between Beacon Hill data and AirBeam data. The red line shows the value of Beacon Hill Data, which are 3.8 μg/m^3, 20.8 μg/m^3, 9.4 μg/m^3, 4.2 μg/m^3, and 6.1 μg/m^3. The PM2.5 concentration monitored by AirBeam fluctuates with the time, while the data collected by Beacon Hill Station is a red horizontal line in the figure. From the box plot below, it is easy to find that the first day and the last day had lower PM2.5 than the other three days. Outliers with PM 2.5 higher than 30μg/m^3 exist in each group. The outliers could be shown as the separated dots on the Figure 22.

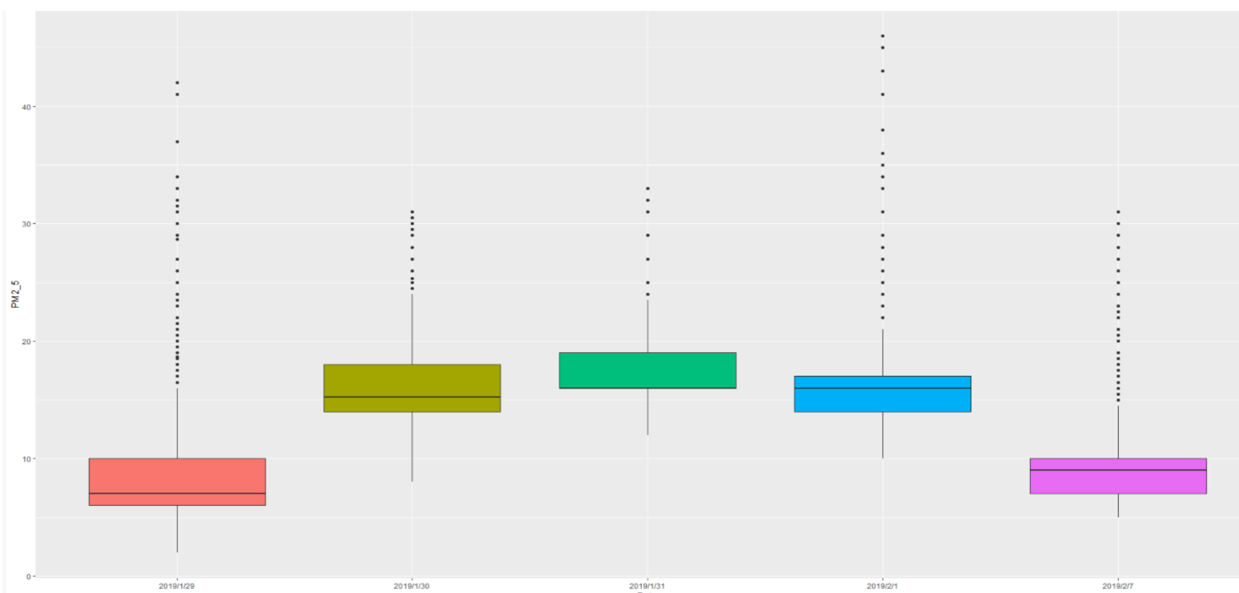
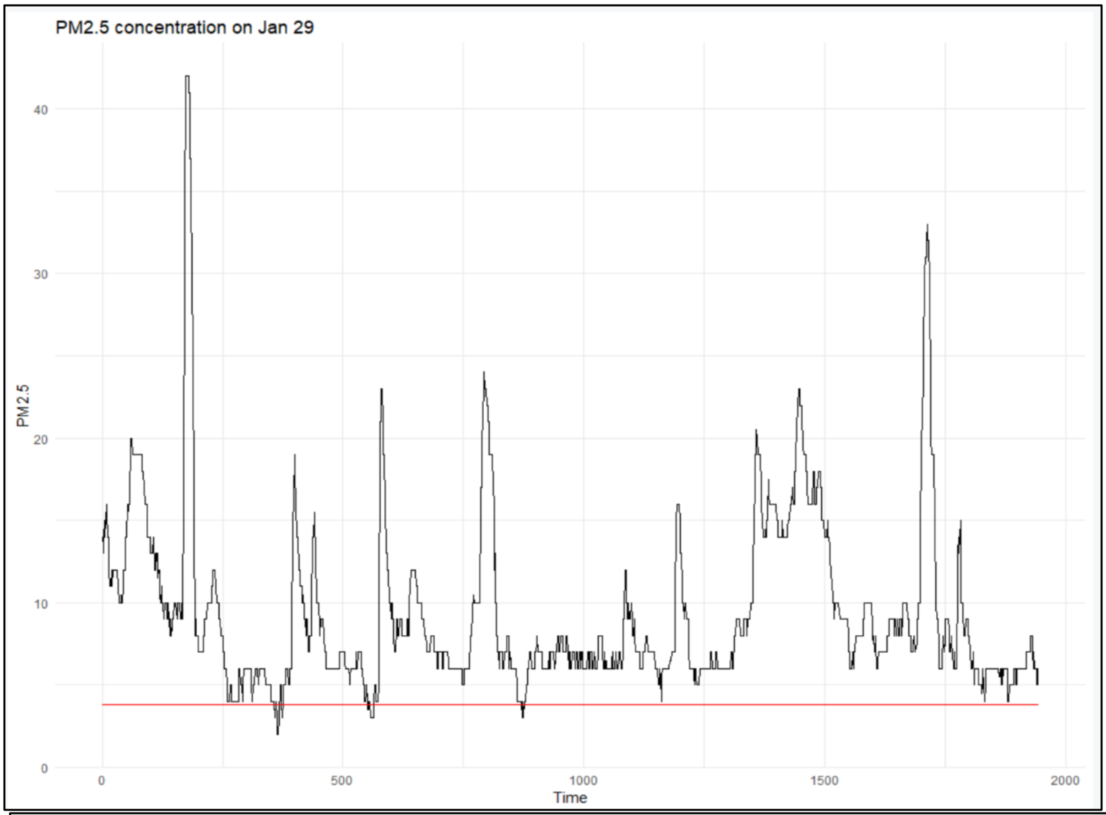
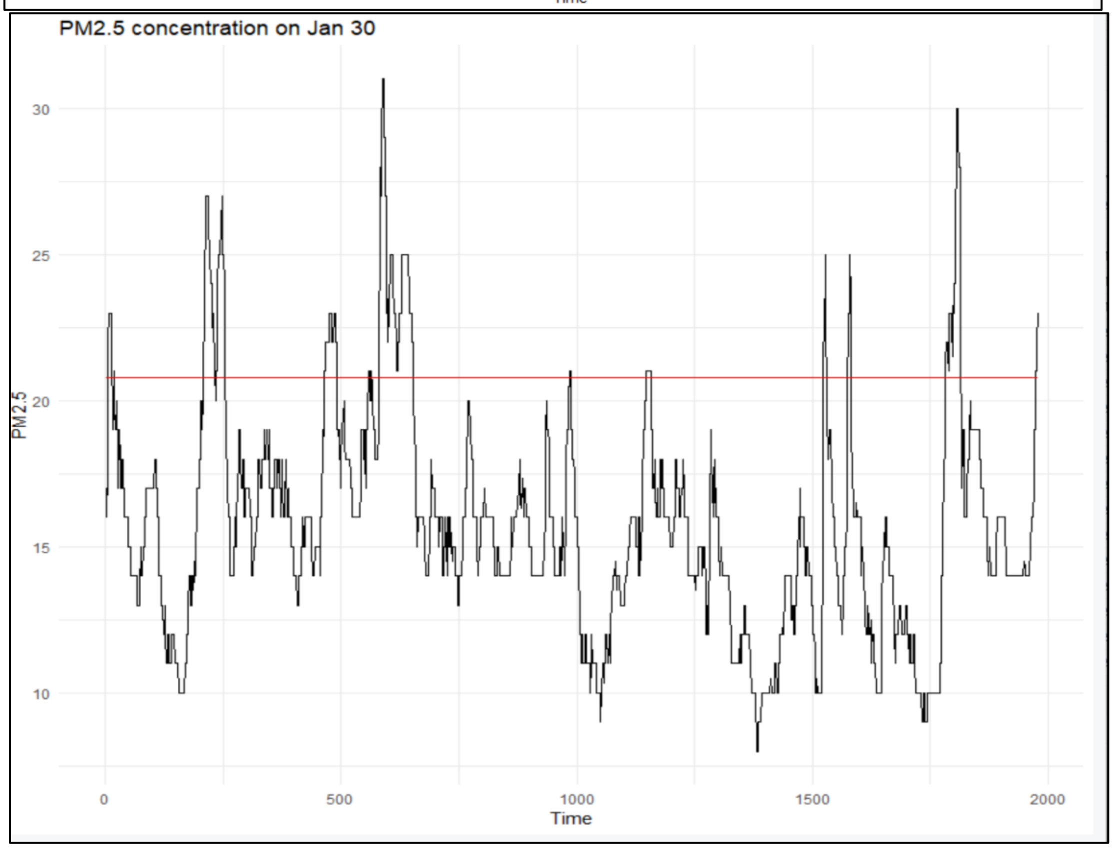
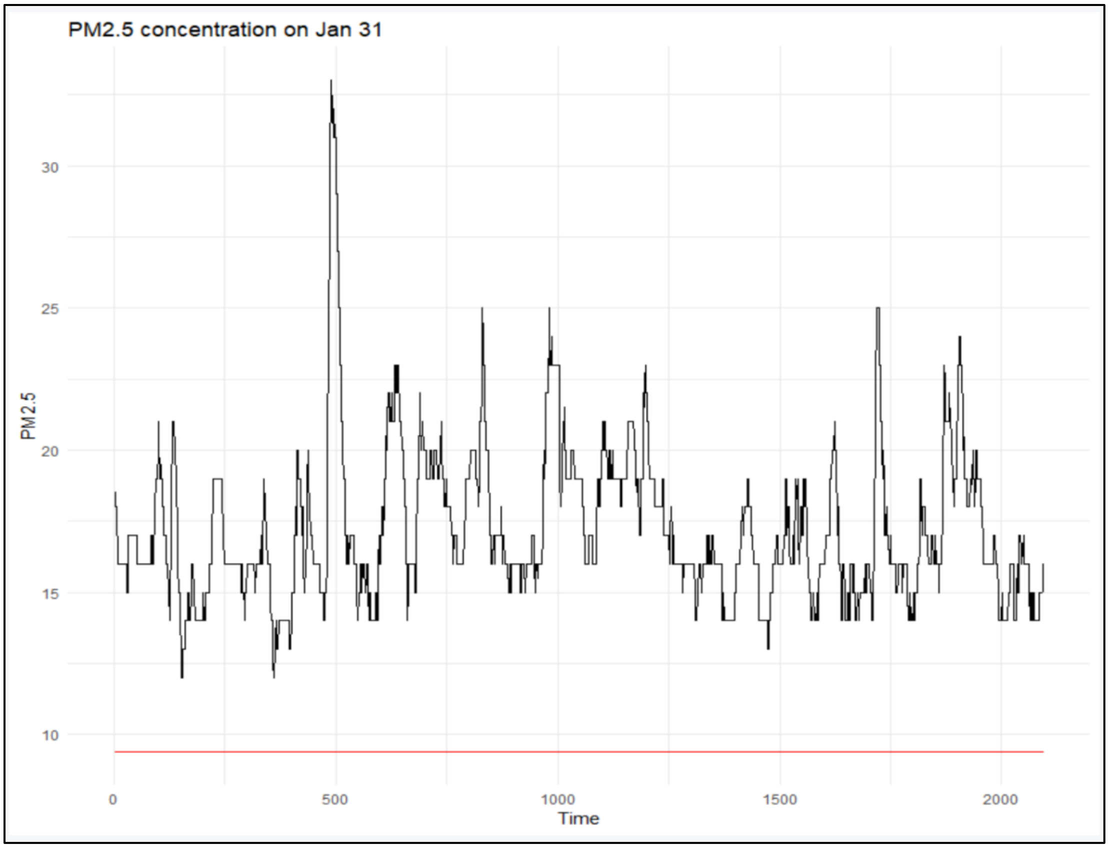
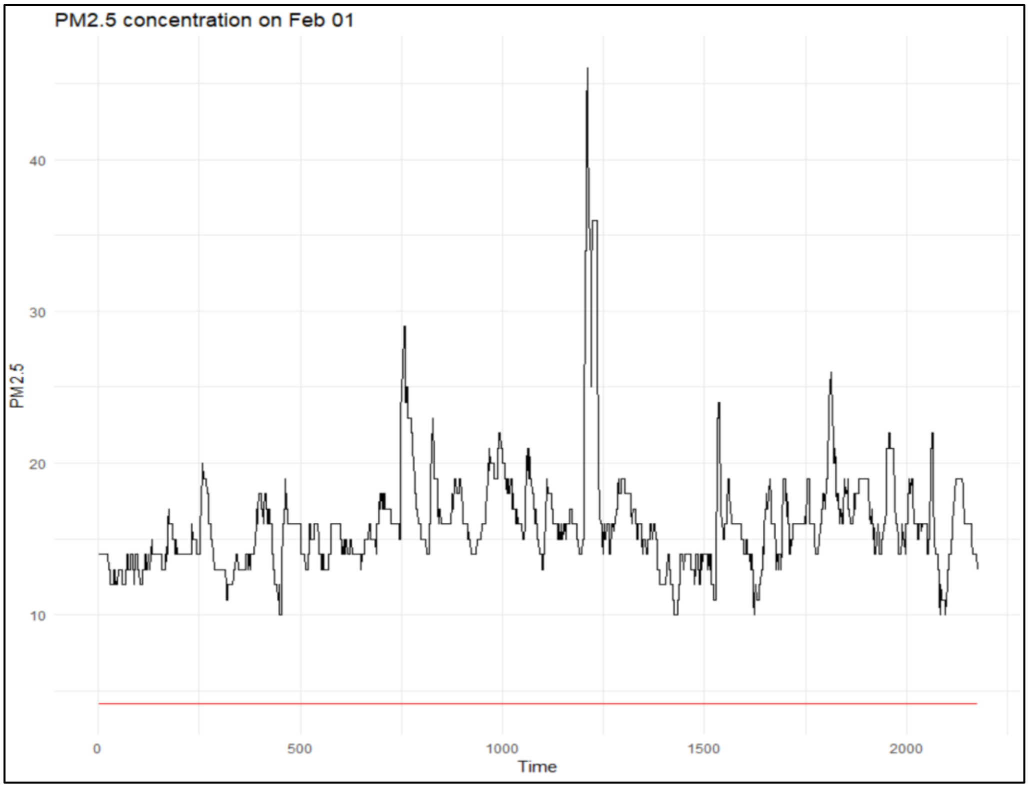


Figure 22. PM2.5 Boxplots in Five Days


PM2.5 concentration on Jan 29
PM2.5
0
10
20
30
40
0
500
1000
1500
2000
Time


PM2.5 concentration on Jan 30
PM2.5
10
15
20
25
30
0
500
1000
1500
2000
Time


PM2.5 concentration on Jan 31
PM2.5
30
25
20
15
10
0
500
1000
1500
2000
Time



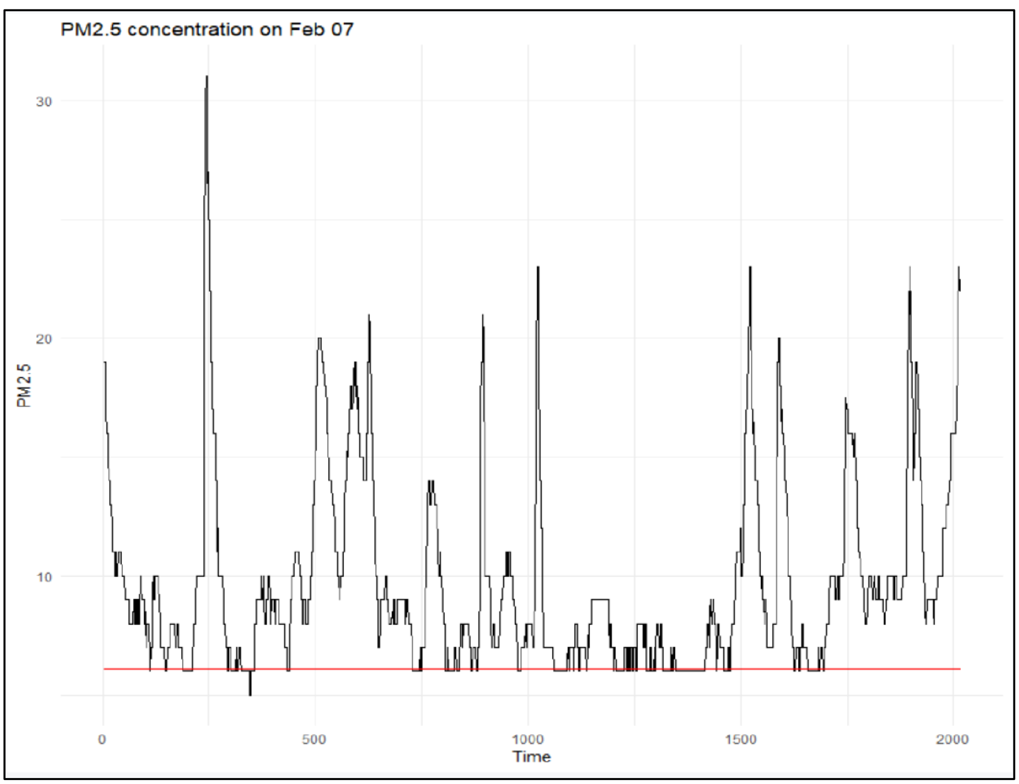


Figure 23. Comparing PM2.5 Concentration between Fixed Station and Portable Sensor Among Five Days

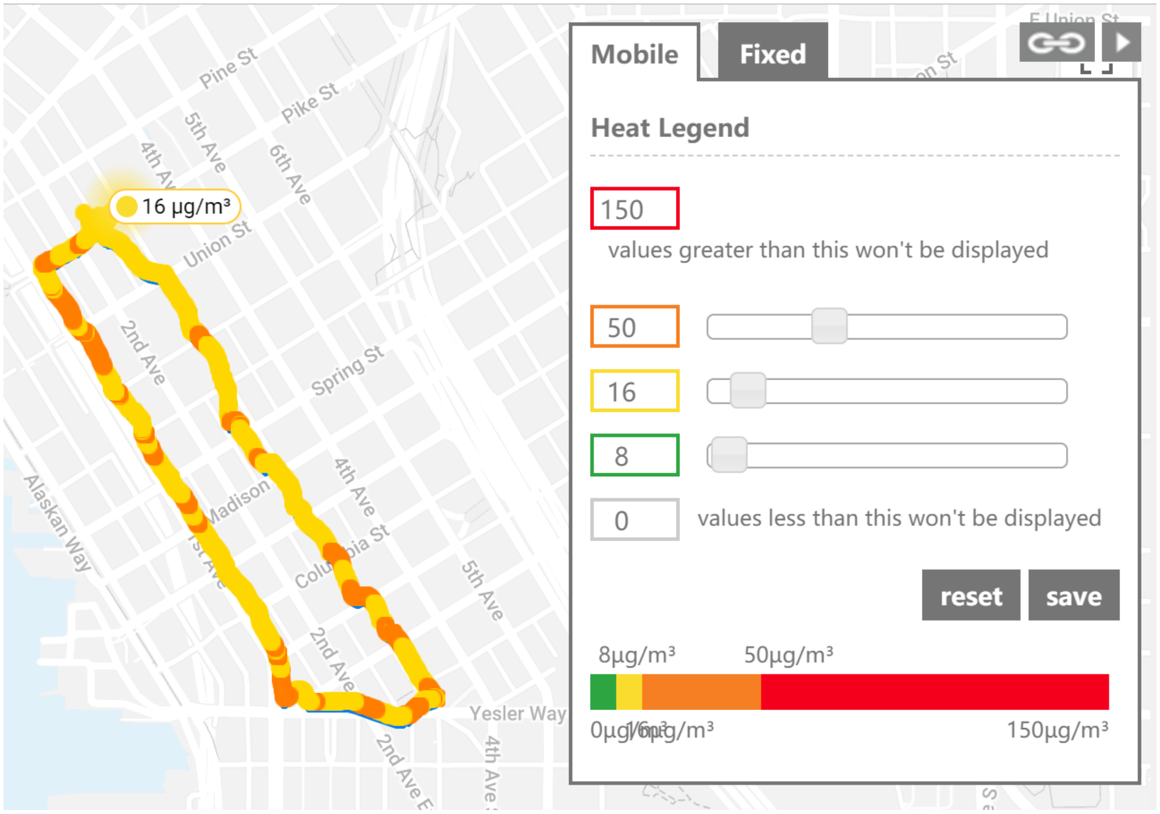
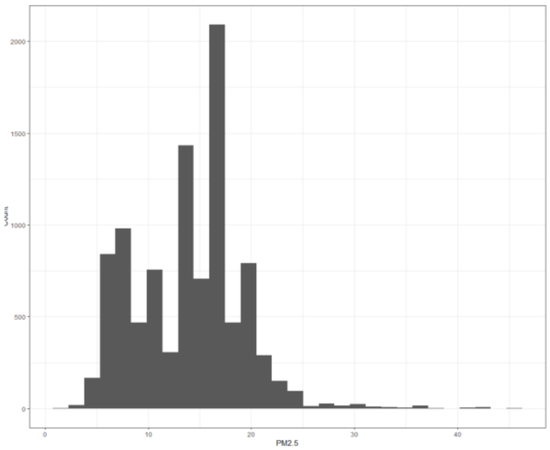


Figure 24. PM2.5 of Feb 01 Showing on AirCasting Platform

The histogram of PM2.5 shows the distribution of the response variable. From the plot, the distribution is slightly right-skewed. However, the mean of PM2.5 which is 13.73 is slightly lower than the median, which indicates the left-skewed distribution. After the log transformation, the histogram looks closer to normal distribution. The comparison of model fitness will be discussed later.



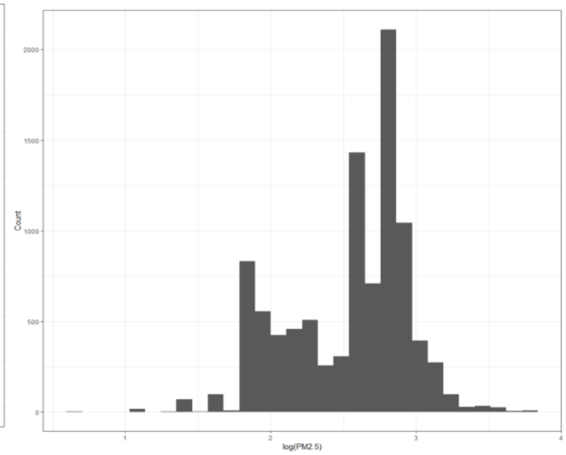


Figure 25. Histogram of PM2.5

Figure 26. Histogram of log(PM2.5)

4.1.2 Smoker

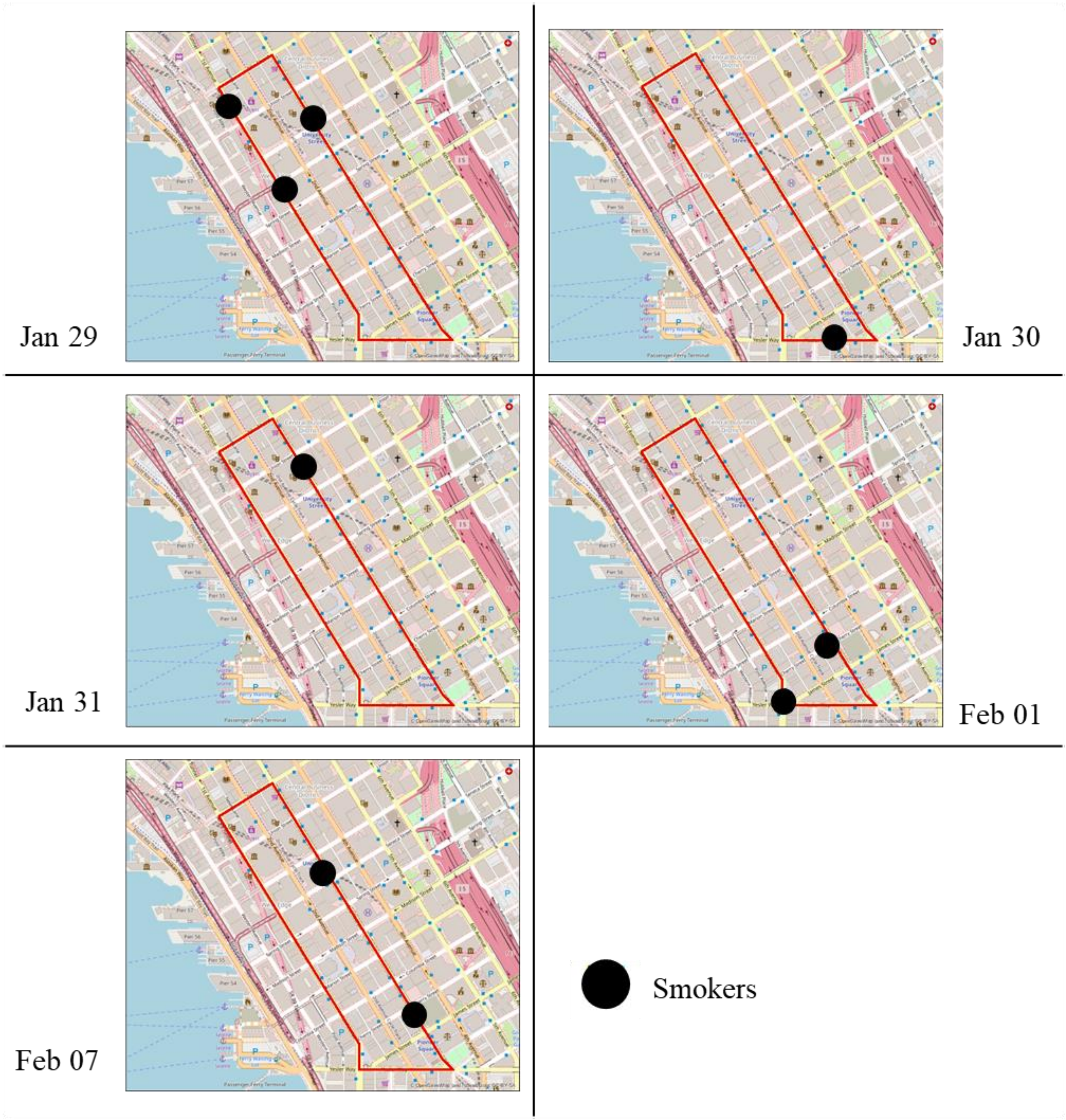


Figure 27. Distribution of Smokers in 5 Days

Figure 27 presents distributions of smokers in 5 days. Based on the five maps, there are not apparent popular smoking places along the studying routes. In general, only two or three smokers show on each map. The number of smokers is not as high as expected. But the low volume of smokers does not mean that they have no significant influence on pedestrian exposure. Figure 28 shows the fluctuation of the number of pedestrians in each intersection over five research days. Overall, the intersections locate at Madison St & 3th Ave, University St & 1st Ave, and Pike St & 2nd Ave are the intersections with the highest pedestrian volume.

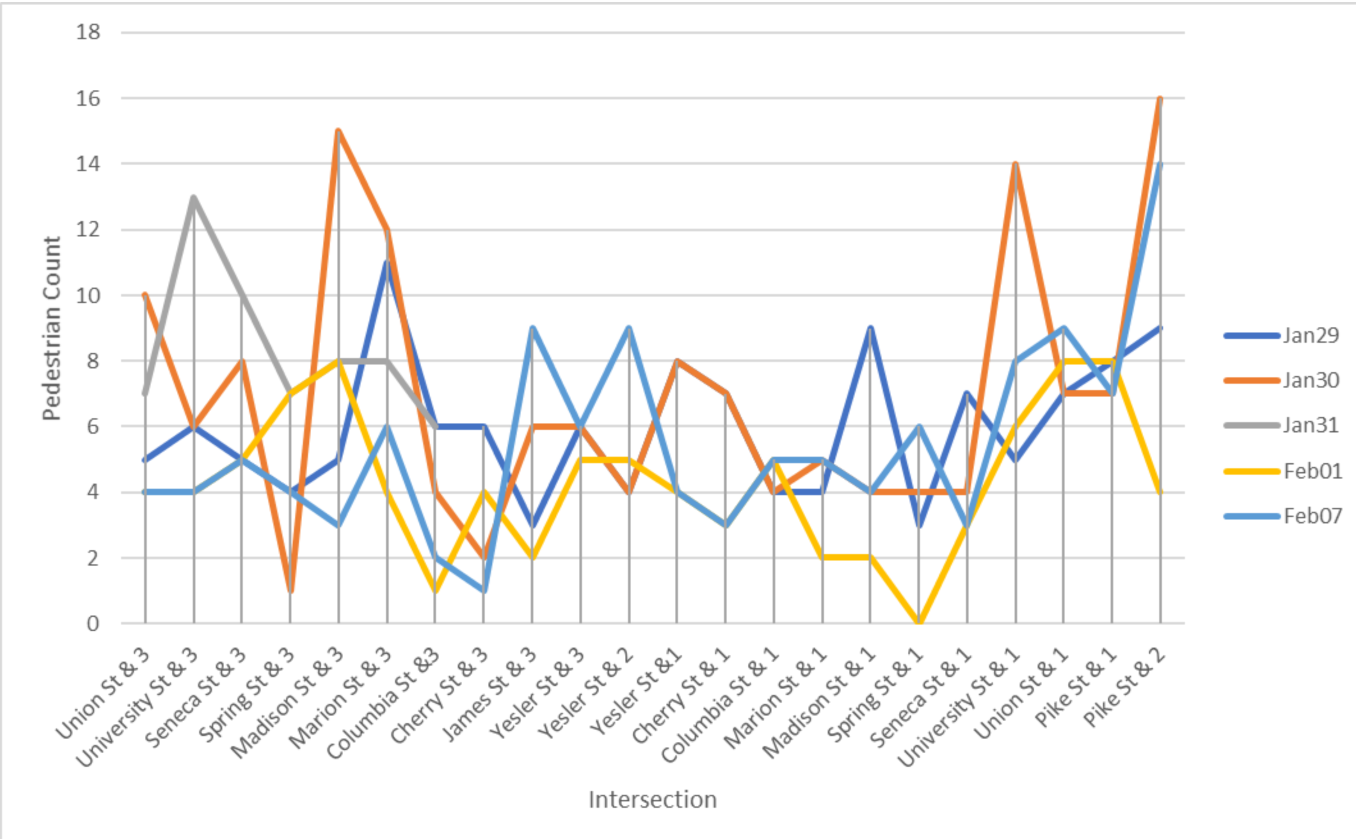


Figure 28. Pedestrian Counts in Each Intersections

4.1.3 Bus Stops

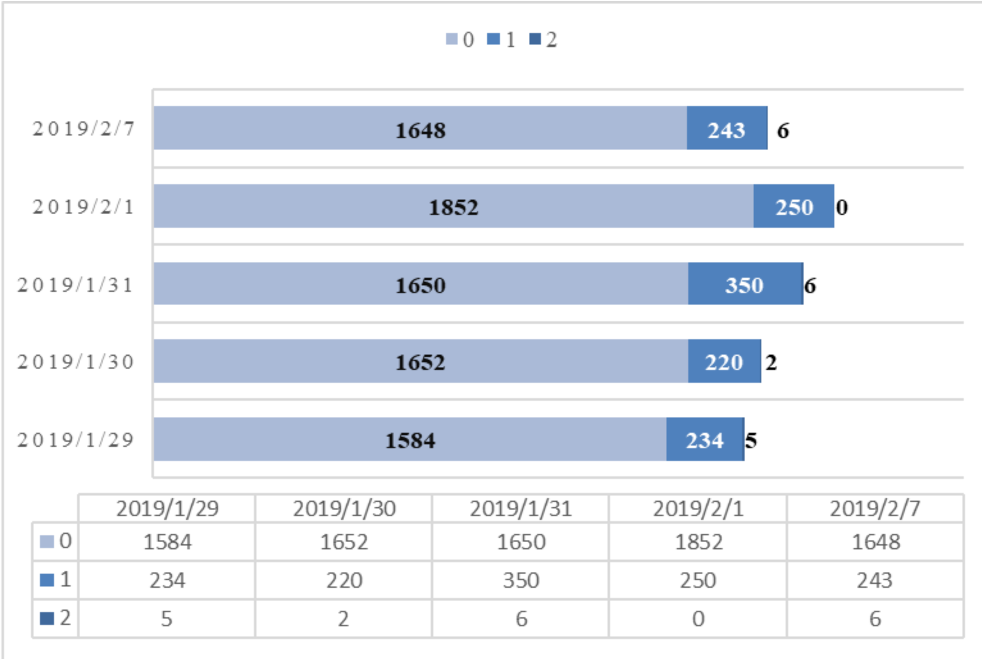


Figure 29. The Frequency Table of Buses in 5 days

The size of area influenced by the bus stops was set up as circles within 20m radius from the data cleaning process. Figure 29 shows the frequencies of samples influenced by bus stops each day. There are only 3 values which are 0, 1, and 2 within the frequency table. Most samples are not influenced by any bus stops where the value of bus count is 0. If a sample is influenced by one bus stop, the value will be 1. Similarly, the value will be 2 if a sample is influenced by two bus stops. From Figure 29, a few samples are influenced by bus stops. In detail, 5 samples fall into the influenced area of two bus stops on Jan29, while the values on Jan 30, Jan 31, Feb 01, and Feb 07 are 2, 6, 0, 6 respectively. Overall, there are some distinctions of bus counts among 5 days. Whether the differences are big enough to affect the PM2.5 needs to be explored in the further step.

4.1.4 Restaurants

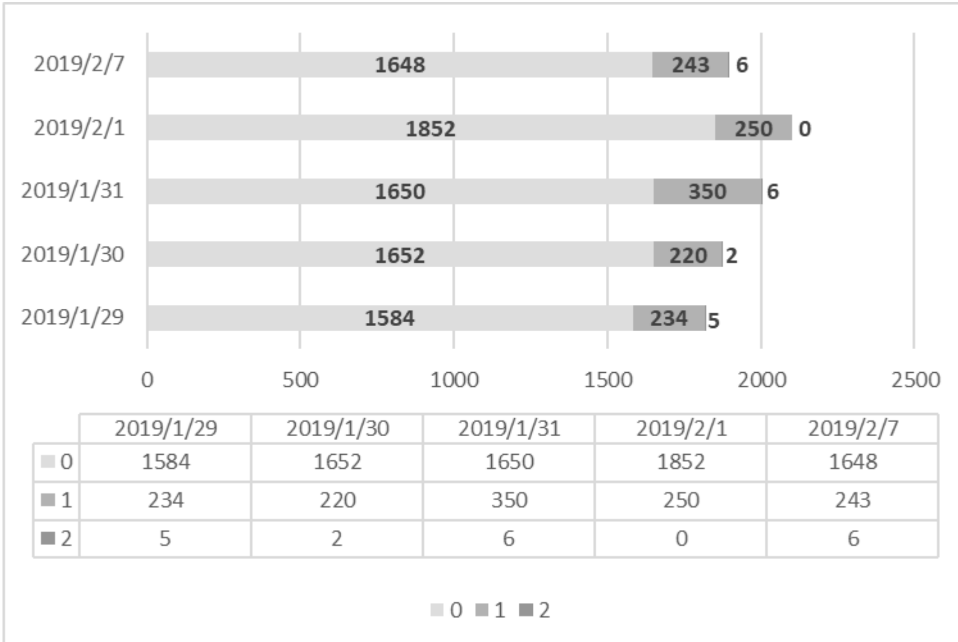


Figure 30. The Frequency Table of Restaurant Count

Same with the bus counts, there are only 3 distinct value in the frequency table of restaurant count which are 0, 1, and 2. (Figure 30) Most of samples are out of the influencing zone of any restaurants while only a few points are affected by more than one restaurant. On Jan 31, there are 350 samples influenced by one restaurant where the number is higher than other days. Other four days have around 230 sample points influenced by one restaurant on average.

4.2 Model Interpretation

The research takes bus stops, restaurants, and smokers as independent variables. PM2.5 serving as the dependent variable has 9702 samples collected from 5 separated workdays. All the variables are transferred as numeric variables. Among them, bus stops and restaurants originally are geographic points showing geolocations. With the help of ArcGIS, numerical value associated with bus stops and restaurants was assigned to samples. Because PM2.5 collected in the same day may also be influenced by other unrevealed factors and shared common features. The linear mixed model is applied into this research since it specially works better for clustered data.

$$\overbrace{Y}^{9702\times 1} = \overbrace{\underbrace{X}_{9702\times 4} \quad \underbrace{\beta}_{4\times 1}}^{9702\times 1} + \overbrace{\underbrace{Z}_{9702\times 5} \quad \underbrace{\mu}_{5\times 1}}^{9702\times 1} + \overbrace{\epsilon}^{9702\times 1}$$

Figure 31. Linear Mixed Model in the Research

R Package nlme was developed by Jose Pinheiro and other three researchers for the linear and non-linear mixed effect model. (Jose Pinheiro,2013) Through the function *lme()* in the nlme package, the outcome of the model is displayed below.

	Dependent variable:
	PM2_5
SmokerCount	10.326***
	(0.283)
BusCount	-0.128
	(0.108)
RestaurantCount	0.220***
	(0.035)
Constant	13.257***
	(1.550)
Observations	9,702
Log Likelihood	-26,524.900
Akaike Inf. Crit.	53,061.790
Bayesian Inf. Crit.	53,104.870
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 32. Summary of Linear Mixed Model

Samples collected in the same day was treated as one group where the model contains five groups in total. Counted Smokers, Bus Stops, and Restaurants as the fixed effects, the model is: $PM2.5 \sim SmokerCount + BusCount + Restaurant$

$$\overbrace{Y}^{9702\times 1} = \overbrace{\underbrace{X}_{9702\times 4} \quad \underbrace{\beta}_{4\times 1}}^{9702\times 1} + \overbrace{\underbrace{Z}_{9702\times 5} \quad \underbrace{\mu}_{5\times 1}}^{9702\times 1} + \overbrace{\epsilon}^{9702\times 1}$$

$$y = \begin{bmatrix} PM2.5 \\ 14 \\ 13.5 \\ \cdots \\ 22 \end{bmatrix} \begin{matrix} n_{ij} \\ 1 \\ 2 \\ \cdots \\ 10089 \end{matrix} \qquad X = \begin{bmatrix} Intercept & Smoker & BusStops & Restaurants \\ 1 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 3 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 13.26 \\ 10.33 \\ -0.1281 \\ 0.2199 \end{bmatrix}$$

Figure 33. Linear Mixed Model

As showed in the equation, the coefficients of smokers, bus stops and restaurants are respectively 10.33, -0.1281, 0.2199 while the intercept is 13.26. Only the relationships between smokers and pedestrian PM2.5 and restaurants and pedestrian PM2.5 are statistically significant ($p < 0.05$). The coefficient for smoking is large, indicating a strong effect on pedestrian PM2.5 exposure. Notice that the coefficient of restaurants is 0.2199, which is low compared to the coefficient of the model intercept. It means the impact of restaurant is limited although it does have a statistically significant impact on pedestrian exposure. Figure 33 shows the coefficients of random effects from 5 different days. Similar with the mean value of PM2.5 each day, Jan 29 and Feb 07 have comparatively lower PM2.5 than other three days.

4.2.1 Variable Transformation

Since the distribution of PM2.5 is little skewed, the log-transformation is considered. The output of model showing below:

$log(PM2.5) \sim SmokerCount + BusCount + Restaurant$

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2.497	0.1381	9694	18.08	6.785e-72
SmokerCount	0.6075	0.02146	9694	28.31	1.691e-169
BusCount	-0.01129	0.008195	9694	-1.378	0.1683
RestaurantCount	0.02244	0.002683	9694	8.365	6.862e-17

Table: Fixed effects: log(PM2_5) ~ SmokerCount + BusCount + RestaurantCount

Min	Q1	Med	Q3	Max
-5.006	-0.5635	-0.07441	0.4126	5.422

Table: Standardized Within-Group Residuals

	Observations	Groups	Log-restricted-likelihood
Date	9702	5	-1502

Figure 34. Output of transformed linear mixed model

The result from transformed linear mixed model is similar to the previous linear mixed model. Smokers and restaurant variables are statistically significantly associated with pedestrian exposure to PM2.5 ($p < 0.05$). In contrast, there is no association between bus stops and pedestrian exposure. The intercepts of random effects show in the Figure 35.

	(Intercept)	SmokerCount	BusCount	RestaurantCount
2019/1/29	2.082	0.6075	-0.01129	0.02244
2019/1/30	2.707	0.6075	-0.01129	0.02244
2019/1/31	2.809	0.6075	-0.01129	0.02244
2019/2/1	2.724	0.6075	-0.01129	0.02244
2019/2/7	2.164	0.6075	-0.01129	0.02244

Figure 35. Coefficients of random effects in the transformed linear mixed model

4.2.2 Model Selection

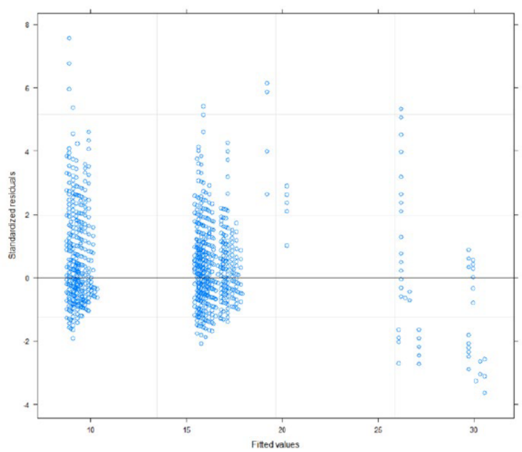
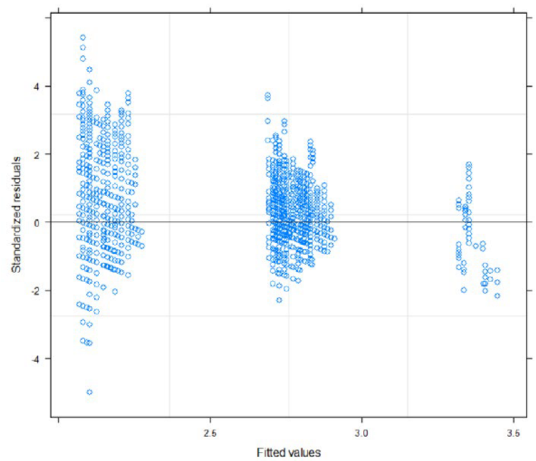
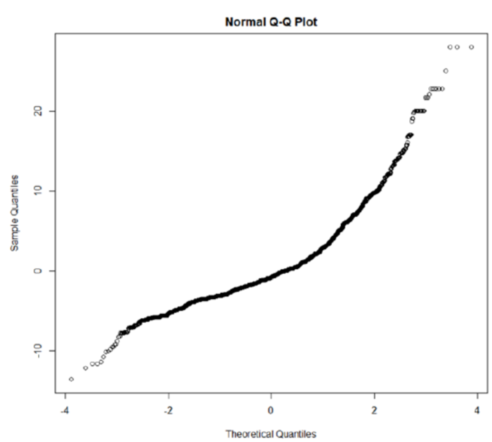
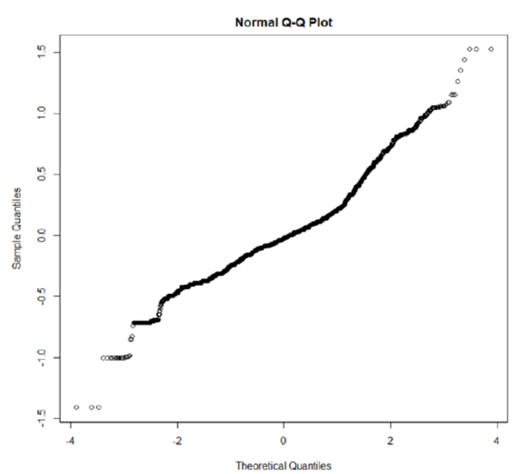
<table>
<tr><th>Regular Linear Mixed Model</th><th>Transformed Linear Mixed Model</th></tr>
<tr><td colspan="2">Standard Residuals Vs. Fitted Values Plot</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Normal Q-Q Plot</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">RMSE</td></tr>
<tr><td>4.442587</td><td>0.3555291</td></tr>
</table>

Figure 36. Model Comparison Chart

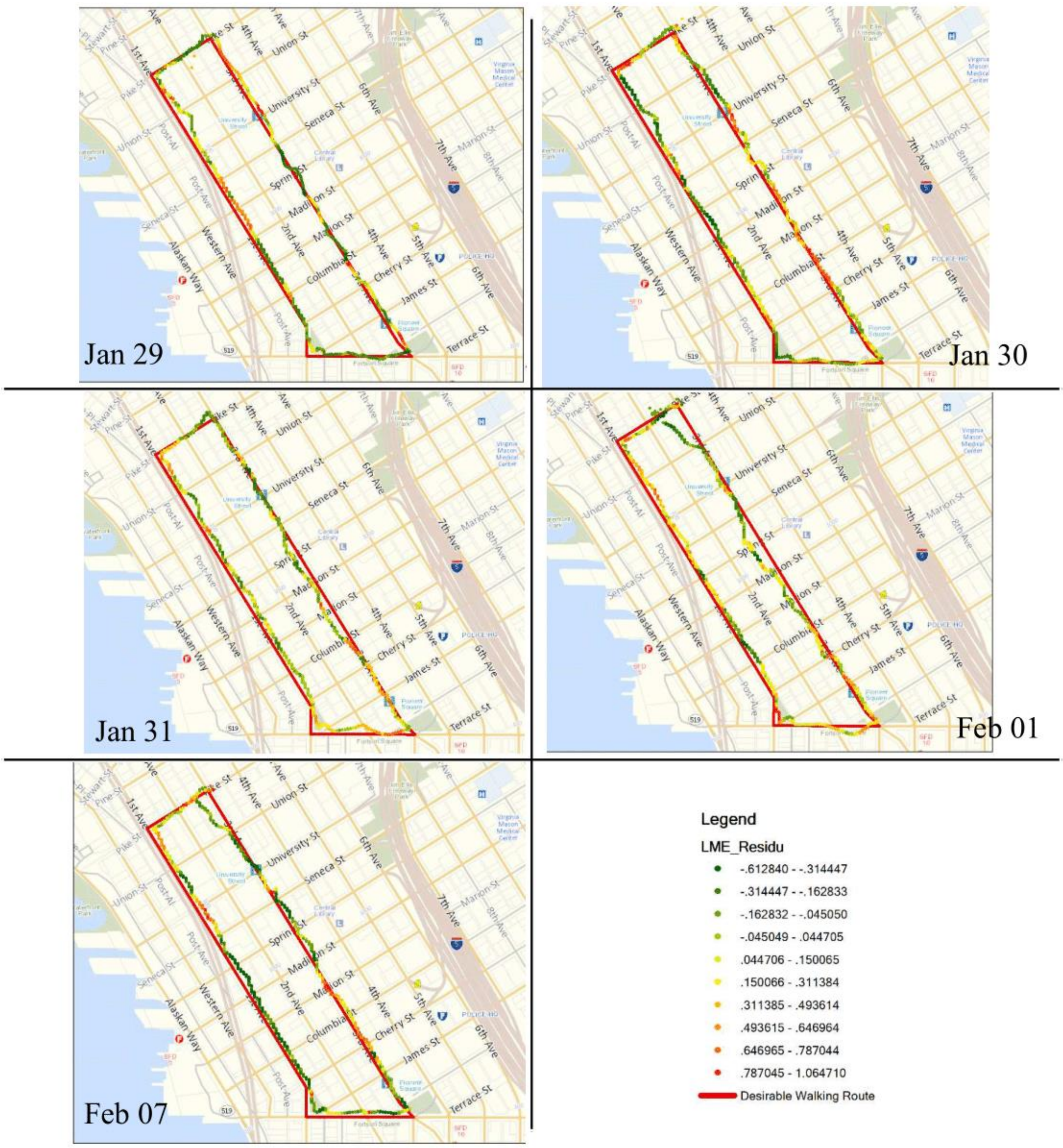


Figure 37 Residuals of Linear Mixed Effects Model in 5 study days

Figure 36 above represents the four criteria to select the model fitting better in the research. Through verifying assumptions of linear mixed model and comparing RSME of two models, the transformed linear mixed model performs better. The standard residuals vs. fitted value plot verifies whether the independent variables have a linearity relationship with dependent variable which is one of assumption of linear regression model. In the plot of Standard Residuals versus Fitted Values, the regular model has more outliers and meantime the transformed model is more symmetric about 0. Thus, the transformed model satisfies the linearity assumption better. But the plots show that residuals cluster into three groups. If the residuals are plotted separately in 5 days, a similar geographic pattern is easy to be observed in Figure 37. Positive residuals of the model mean that the model is underestimated, which are shown as red color in Figure 37. Negative residuals indicate that the model is over estimated, which are shown as green color. It is not sure the reason why the model is underestimated in some certain areas. Dense red dots are close to larger bus stops on the Third Avenue, a park between Yesler & 2nd Ave and Yesler & 3rd Ave. In Jan 29, Jan 31, Feb 01, Feb 07, they all have clustered green dots from Seneca St to Madison St on 3rd Avenue. There might be some built environment characteristics related to the pattern. Whether the construction site or gathered human influence the residuals are worthy of being discussed in future research. If the slop of samples line in Q-Q plot is closer to 1, the residuals of the model follow the normal distribution better. Among the two models, the transformed model performs better. The last criteria is the value of RSME. The smaller RSME is, the closer the observed data points are to the predicted data. The RSME of transformed linear mixed model is 0.355 while the RSME of regular model is 4.442. Based on the results of model diagnostics, the model with transformed response variable performs better.

Chapter 5: Discussion

5.1 Research Conclusion

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2.497	0.1381	9694	18.08	6.785e-72
SmokerCount	0.6075	0.02146	9694	28.31	1.691e-169
BusCount	-0.01129	0.008195	9694	-1.378	0.1683
RestaurantCount	0.02244	0.002683	9694	8.365	6.862e-17

Figure 38. Coefficients Table of Transformed Linear Mixed Model

From the linear mixed model conducted in this research, there is enough confidence to confirm that smokers and restaurants can increase the exposure of nearby pedestrians to PM2.5, but the impact of bus stops is surprisingly not clear. One possible explanation could be that I use fixed bus stations as my variable. But when I collected data, the bus stations could have no loading buses. Then the influence brought by bus is not statistically significant in this study. Based on Figure 38, the unconditional expected mean of log (PM2.5) is 2.497. In other words, the default value of PM2.5, when the influences of smokers and restaurant are ignored, is exp (2.497) =12.146. Although restaurants have positive influences on pedestrian exposure to PM 2.5, the value is exp (0.0224) = 0.0073, which is comparatively low. The value of coefficient of smokers is exp (0.6075) = 1.835. It means that one unit of increasement of smokers could increase 1.835 units of PM2.5. A reasonable interpretation of the findings reported for the variable restaurants is that there is no food truck in the study route. As discussed in the literature review, Bae found the PM2.5 indicator went high when she was close to food trucks. (Bae et al., 2018)

5.2 Limitations

The research inevitably has some limitations from the perspectives of research assumptions, research design framework, and the data collection work in real world. First, the method to quantify influences of bus stops and restaurants on air pollution is not error-proof. In this research, all the samples are thousands of geographic points assigned with scores related to each factor. In other words, every sample has values associated with buses and restaurants. The values represent the number of restaurants and bus stops influencing the sample. The influencing radius of restaurants is 50m while the influencing radius of bus stops is 20m. However, the buffer of the influencing area is set up without a systematical research guideline. The inexact buffer size may bring some bias or error into the research.

Second, one of the assumptions in this research is that smokers are static during the data collection process. It is not always true in real world. Some smokers were walking while they were smoking. The researcher could be exposed to smoking a longer time when someone was smoking and moving in the same direction with the researcher. But the model doesn't take this situation into account, which may bring some deviations into the research. Moreover, the model is built on the assumption that smokers have a net positive influence on levels of pedestrian exposure to air pollution. Thus, in the data clean process, the researcher counted smokers by reviewing the video record when PM2.5 detection shows peaks of PM2.5. Although the way of counting smokers could be biased, it is the most time-efficient way to count smokers.

5.3 Policy Implication

The research indicates that pedestrians suffer the risk of breathing secondhand smoking while they are walking on the sidewalks. In other words, smokers have ineligible impacts on pedestrian exposure to PM2.5. There is no nation-wide law in the USA to ban smoking in public spaces. Some states like California and Washington have state-wide regulations for tobacco, while other states don't have similar regulations. The Center for Tobacco Policy & Organizing published a report on researching Secondhand Smoke Ordinances in California. (America Lung Association, 2019) Table 1 shows the list of cities having specific regulations on sidewalks in CA. Inspired by the research result of this project, I suggest legislations to set a boundary in the urban core area to ban smoking on sidewalks.

Comprehensive Outdoor Secondhand Smoke Ordiances:Somking is Prohibited					
City/County	Time	City/County	Time	City	Time
Solvang	November 2018	Cloverdale	January 2017	Coronado	October 2013
Citrus Heights	November 2018	Laguna Beach	January 2017	Walnut Creek	October 2013
Half Moon Bay	October 2018	Bell	December 2017	Arcata	July 2013
Tiburon	July 2018	Sonoma City	June 2016	Fremont	November 2012
Los Altos	March 2018	Los Gatos	May 2016	Alameda	November 2011
Dana Point	January 2018	San Rafael	April 2016	Hermosa Beach	November 2011
Alturas	February 2017	Pleasant Hill	November 2015	Concord	September 2011
Capitola	October 2015	Palo Alto	May 2015	Carpinteria	February 2011
Davis	September 2015	Livermore	April 2015	Eureka	July 2010
Daly City	August 2015	Hemet	March 2015	Camarillo	April 2010
Oceanside	March 2015	Piedmont	December 2014	San Luis Obispo	April 2010
Agoura Hills	January 2015	Foster City	October 2014	Del Mar	December 2009
El Cerrito	January 2015	Manhattan Beach	July 2014	Palm Desert	November 2009
Moorpark	September 2009	Loma Linda	June 2008	Hayward	May 2008
Santa Cruz	September 2009	Albany	May 2008	Berkeley	December 2007
El Cajon	August 2007	Burbank	April 2007	Santa Monica	October 2006
Santa Rosa	June 2006				

Table 1. List of Cities Regulating Smoking on Sidewalks in CA

5.4 Recommendations for future research

I have a few recommendations for future research on pedestrian exposure to air pollutions based on lessons from this research. Most suggestions are for the data collection of studies. Firstly, the sample data could be collected in a more extended period with consistent weather conditions. Secondly, Electric cigarette samples are worthy of being separated from regular cigarette samples in terms of second-hand smoking studies. Last but not least, a counter could be applied through the data collection process, while counting smokers via videos brings bias into the research.

Chapter 6: References

Adams, H. S., Nieuwenhuijsen, M. J., & Colvile, R. N. (2001). Determinants of fine particle (PM 2.5) personal exposure levels in transport microenvironments, London, UK. *Atmospheric Environment*, *35,* 4557–4566. https://doi.org/10.1016/S1352-2310(01)00194-7

America Lung Association. (2019). *Comprehensive Outdoor Secondhand Smoke Ordinances*. Retrieved from www.Center4TobaccoPolicy.org/tobacco-policy/

Bae, C.-H. C., Sa, B., & Sinha, D. (2018). Real-time Environmental Exposure To Pedstrains: The Case of Downtown Seattle, 1–10.

Bae, C.-H. C., & Sinha, D. (2016). Measuring Pedestrian Exposure to PM-2.5. *Transportation Research Record: Journal of the Transportation Research Board*, *2570*, 139–147. https://doi.org/10.3141/2570-15

Baldwin, D., David, P., Stinson, A. E., & Park, J. (2010). Determinants of exposure to fi ne particulate matter (PM 2 . 5) for waiting passengers at bus stops. *Atmospheric Environment*, *44*(39), 5174–5182. https://doi.org/10.1016/j.atmosenv.2010.08.046

Bartlett, J. (2014). Robustness of linear mixed models. Retrieved April 5, 2020, from https://thestatsgeek.com/2014/08/17/robustness-of-linear-mixed-models/

Bereitschaft, B. (2015). Pedestrian exposure to near-roadway PM2.5 in mixed-use urban corridors: A case study of Omaha, Nebraska. *Sustainable Cities and Society*, *15*, 64–74. https://doi.org/10.1016/j.scs.2014.12.001

Bliss Air. (2014). What is PM2.5 and Why You Should Care. Retrieved April 3, 2020, from https://blissair.com/what-is-pm-2-5.htm

Cao, T., & Thompson, J. E. (2017). Portable, Ambient PM2.5 Sensor for Human and/or Animal

Exposure Studies. *Analytical Letters*, *50*(4), 712–723. https://doi.org/10.1080/00032719.2016.1190736

Carter, E., Archer-nicholls, S., Ni, K., Lai, A. M., Niu, H., Secrest, M. H., … Baumgartner, J. (2016). Seasonal and Diurnal Air Pollution from Residential Cooking and Space Heating in the Eastern Tibetan Plateau. https://doi.org/10.1021/acs.est.6b00082

Consulting, U. S. (2019). Introduction to Linear Mixed Models. Retrieved February 5, 2020, from https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/

Department of Health, N. Y. S. (2020). Fine Particles (PM 2.5) Questions and Answers. Retrieved April 3, 2020, from https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm

Frees, E. W. (2004). Longitudinal and Panel Data:Analysis and Applications for the Social Sciences. *Cambridge University Press*, 1–364.

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environmental Pollution*, *199*, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013

Goel, R., Gani, S., Guttikunda, S. K., Wilson, D., & Tiwari, G. (2015). On-road PM2.5 pollution exposure in multiple transport microenvironments in Delhi. *Atmospheric Environment*, *123*, 129–138. https://doi.org/10.1016/j.atmosenv.2015.10.037

Ham, W., Vijayan, A., Schulte, N., & Herner, J. D. (2017). Commuter exposure to PM2.5, BC, and UFP in six common transport microenvironments in Sacramento, California. *Atmospheric Environment*, *167*, 335–345. https://doi.org/10.1016/j.atmosenv.2017.08.024

Hartin, K., Seto, E., & Yost, M. (2015). Fine Particle(PM2.5) Composition of Inddor and Outdoor Air Samples Collected in Xuanwei County, China.

Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, *11*(8), 4883–4890. https://doi.org/10.5194/amt-11-4883-2018

Li, Z., Che, W., Frey, H. C., Lau, A. K. H., & Lin, C. (2017). Characterization of PM 2 . 5 exposure concentration in transport microenvironments using portable monitors *. *Environmental Pollution*, *228*, 433–442. https://doi.org/10.1016/j.envpol.2017.05.039

Lioy, P., & Weisel, C. (2014). *Exposure Science: Basic Principles and Applications*. *Exposure Science: Basic Principles and Applications*. https://doi.org/10.1016/C2013-0-13335-1

Narotzki, B., Reznick, A. Z., Mitki, T., Aizenbud, D., & Levy, Y. (2014). Bus Stop Air Quality: An Empirical Analysis of Exposure to Particulate Matter at Bus Stop Shelters. *Rejuvenation Research*. https://doi.org/10.1089/rej.2013.1540

Ngoc, L. T. N., Kim, M., Bui, V. K. H., Park, D., & Lee, Y. C. (2018). Particulate matter exposure of passengers at bus stations: A review. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph15122886

Ott, W. R., Acevedo-Bolton, V., Cheng, K. C., Jiang, R. T., Klepeis, N. E., & Hildemann, L. M. (2014). Outdoor fine and ultrafine particle measurements at six bus stops with smoking on two California arterial highways-Results of a pilot study. *Journal of the Air and Waste Management Association*, *64*(1), 47–60. https://doi.org/10.1080/10962247.2013.832712

Puget Sound Clean Air Agency. (2019). Seattle Beacon Hill - Air Graphing. Retrieved April 3, 2020, from https://secure.pscleanair.org/airgraphing

United States Environmental Protection Agency. (2012). 2012 National Ambient Air Quality Standards (NAAQS) for Particulate Matter (PM).

United States Environmental Protection Agency. (2019). Your Health. Retrieved April 3, 2020, from https://www.airnow.gov/index.cfm?action=topics.health

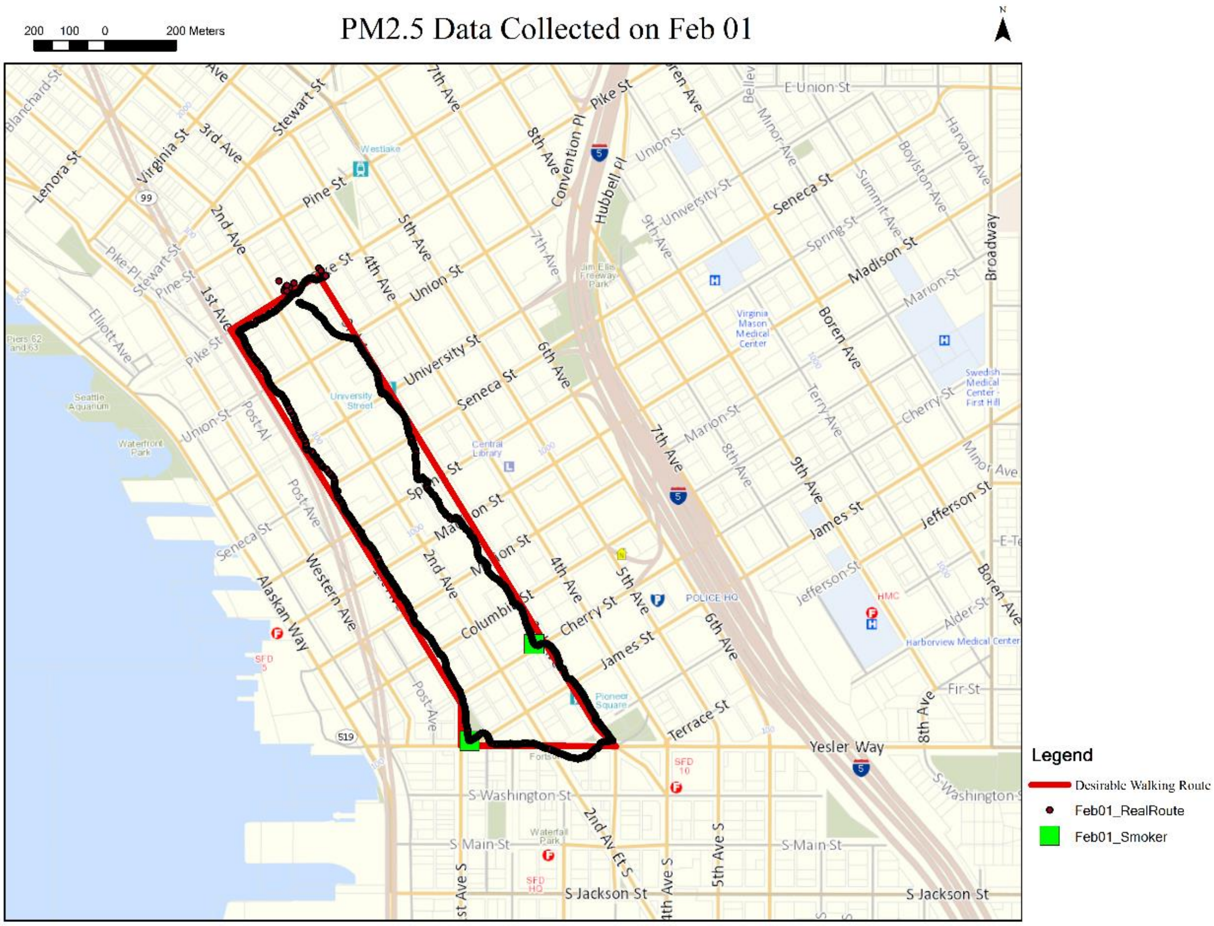
Vanos, J. K., Cakmak, S., Kalkstein, L. S., & Yagouti, A. (2015). Association of weather and air pollution interactions on daily mortality in 12 Canadian cities. *Air Quality, Atmosphere and Health*, *8*(3), 307–320. https://doi.org/10.1007/s11869-014-0266-7

Viana, M., Rivas, I., Reche, C., Fonseca, A. S., Pérez, N., Querol, X., … Sunyer, J. (2015). Field comparison of portable and stationary instruments for outdoor urban air exposure assessments. *Atmospheric Environment*, *123*, 220–228. https://doi.org/10.1016/j.atmosenv.2015.10.076

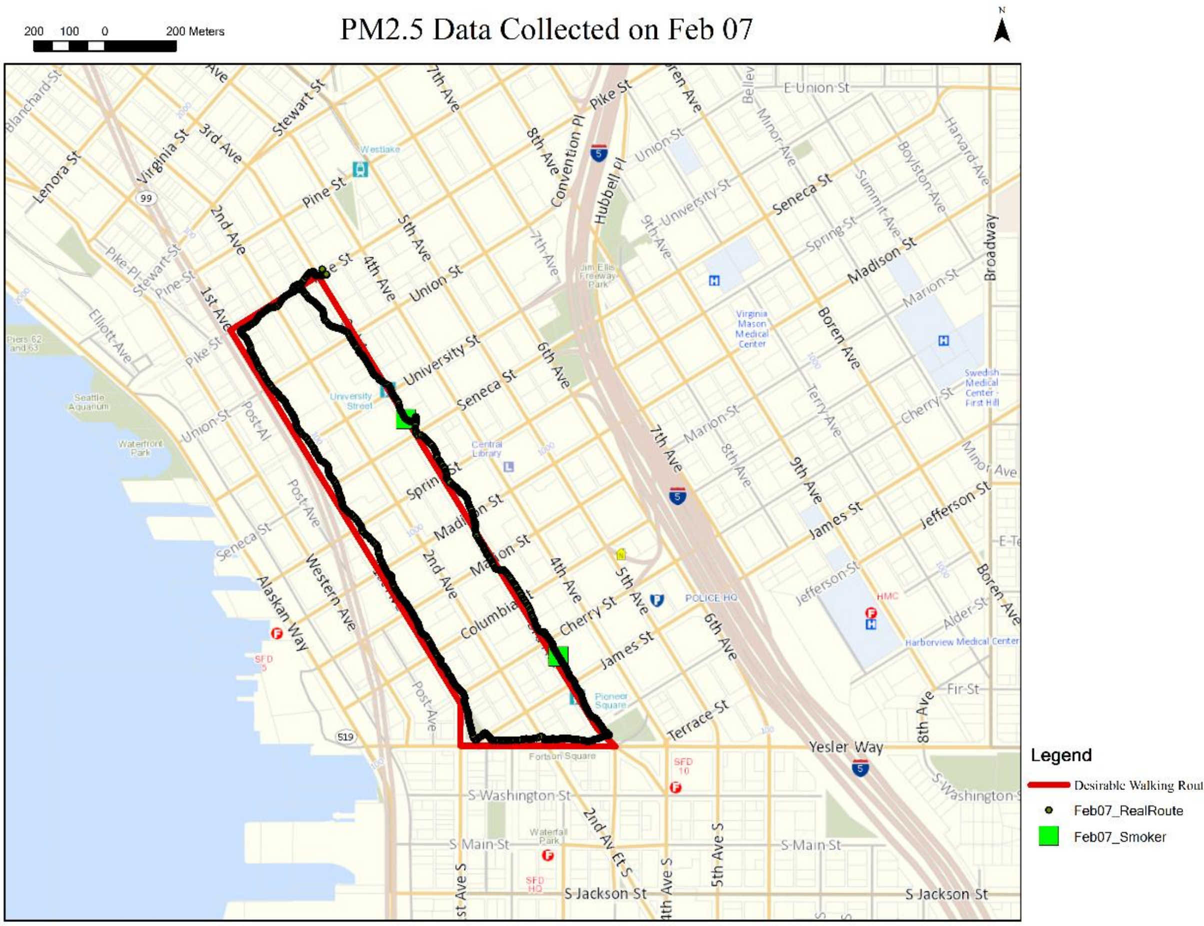
Weng, M., & Jin, X. (2015). Study on the air pollution in typical transportation microenvironment: Characteristics and health risks. *Journal of the Air and Waste Management Association*, *65*(1), 59–63. https://doi.org/10.1080/10962247.2014.962648

Wikipedia. (2020). Pioneer Square, Seattle - Wikipedia. Retrieved April 3, 2020, from https://en.wikipedia.org/wiki/Pioneer_Square,_Seattle

Appendix

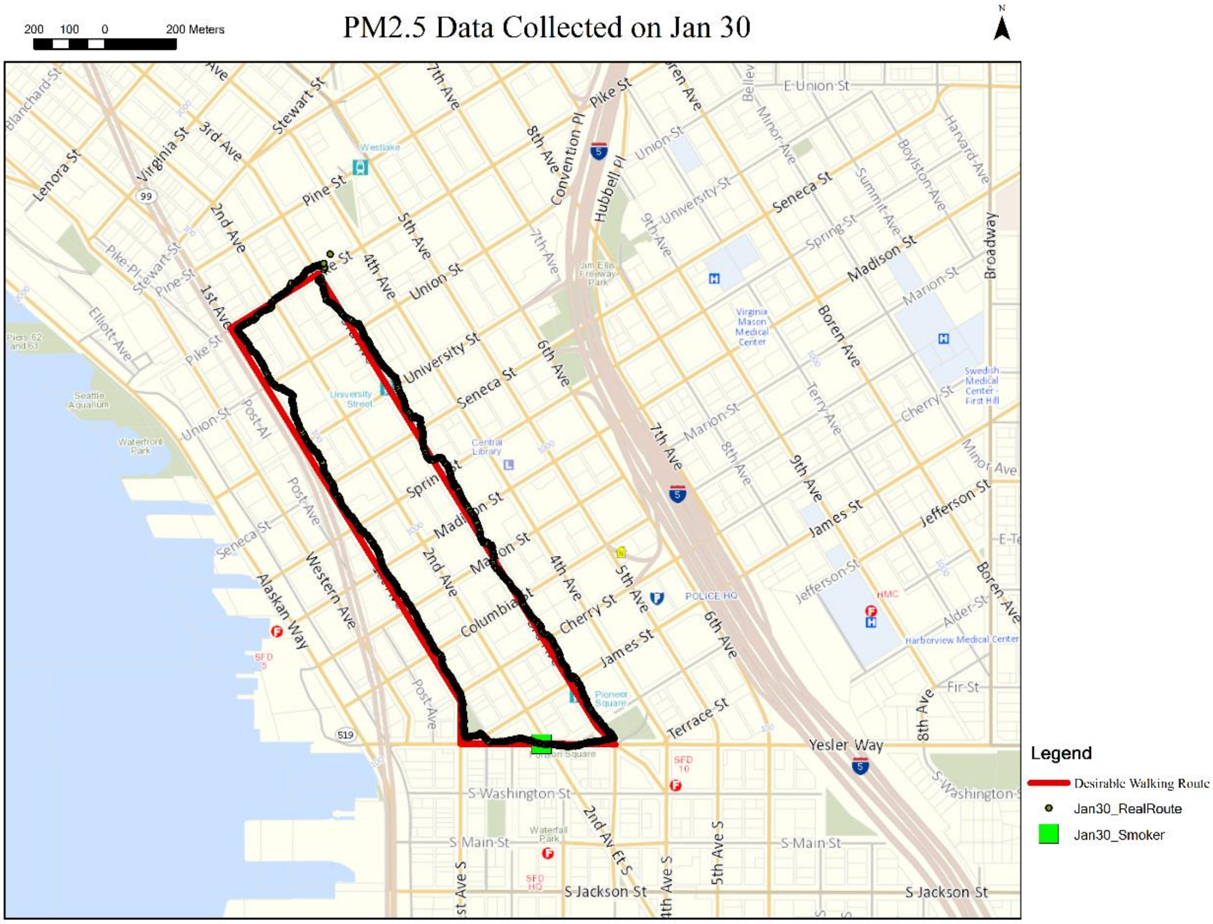

PM2.5 Data Collected on Feb 01
200 100 0 200 Meters
N
Legend
Desirable Walking Route
Feb01_RealRoute
Feb01_Smoker

PM2.5 Data Collected on Feb 07



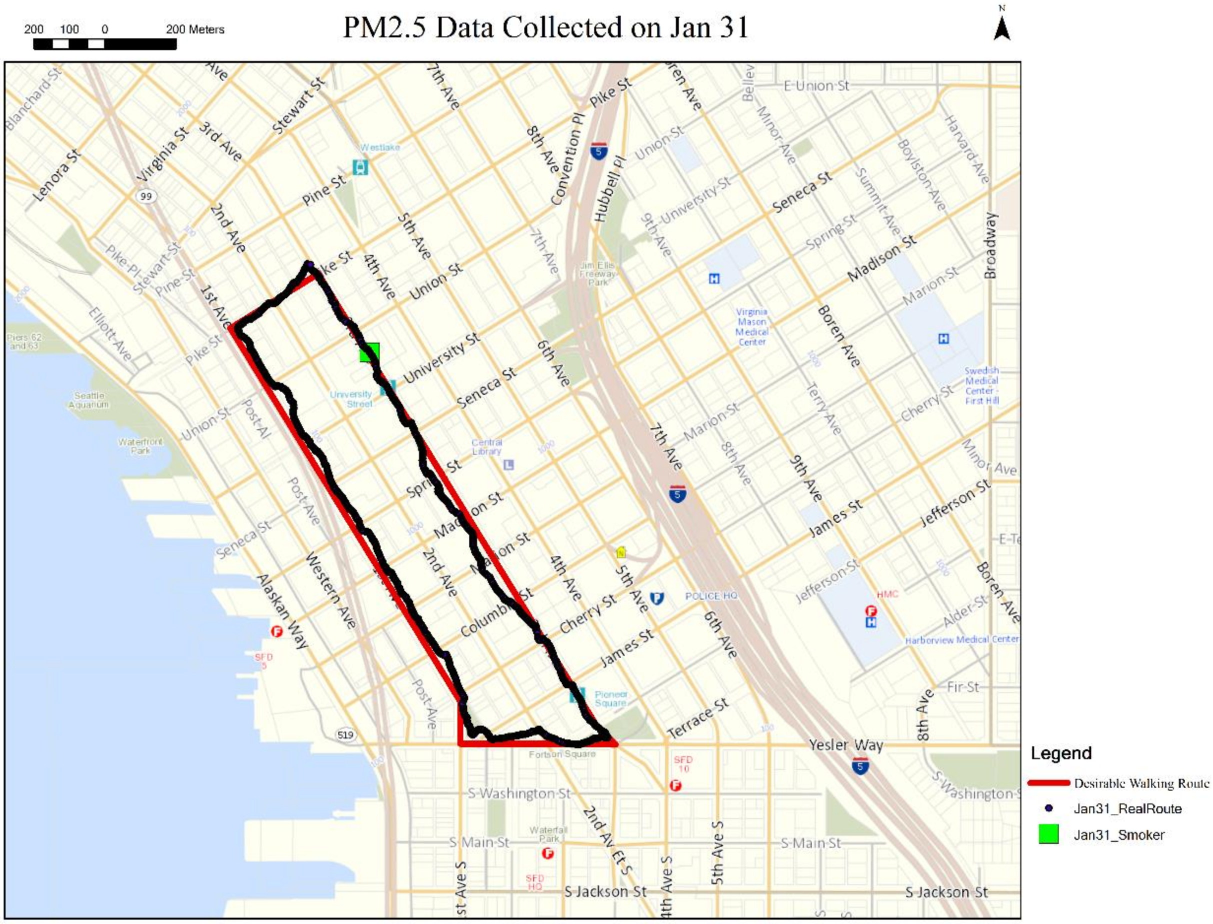

PM 2.5 Collected on Jan 29
125 62.5 0 125 Meters
Legend
Jan29_RealRoute
Jan29_Smoker
Desirable Walking Route
Pike Pl
Stewart St
Pine St
1st Ave
4th Ave
Union St
Elliott Ave
Pike St
University St
Seneca St
6th Ave
Seattle Aquarium
University Street
Union St
Post Al
Waterfront Park
Central Library
7th Ave
Spring St
Madison St
Post Ave
Seneca St
Marion St
Western Ave
2nd Ave
4th Ave
5th Ave
Alaskan Way
Columbia St
Cherry St
James St
SFD 5
Pioneer Square
Jim Ellis Freeway Park
Post Ave
519
Terrace
Fortson Square
SFD 10

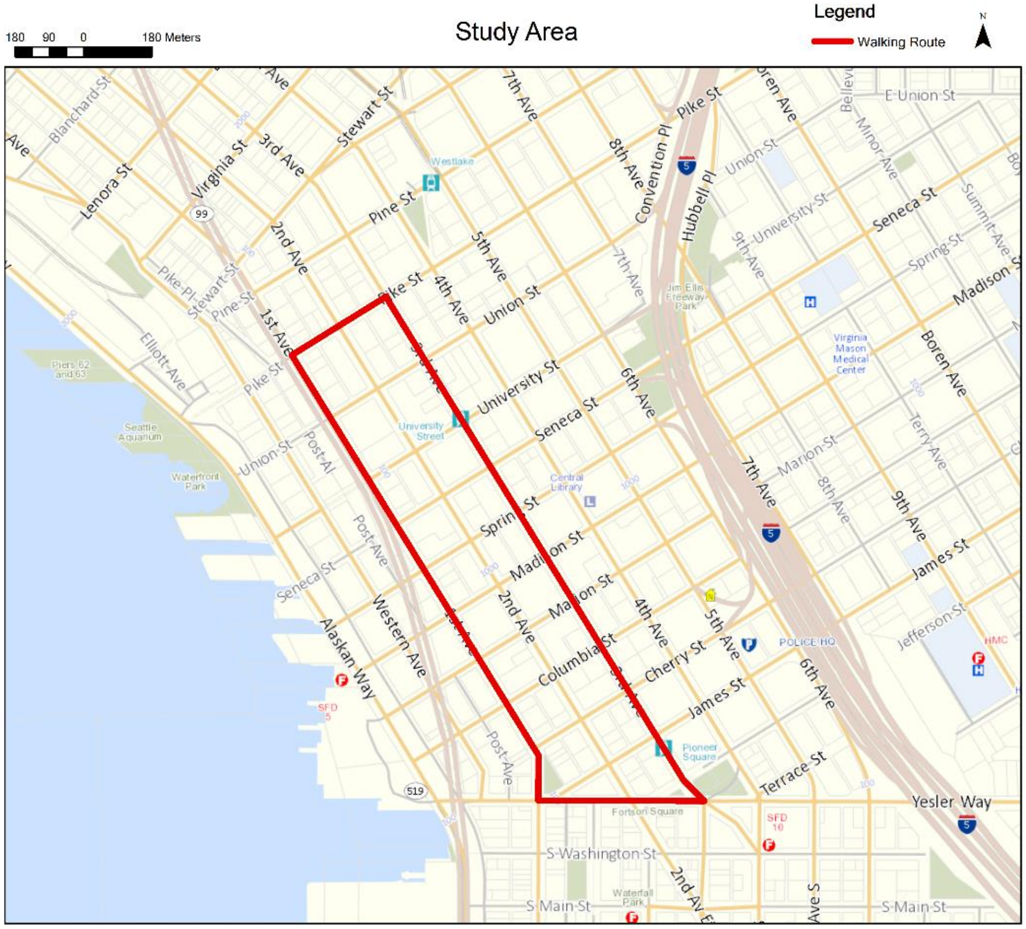

125 62.5 0 125 Meters
Bus Stops and Restaurants Along the Studying Route
N
Legend
Desirable Walking Route
Restaurants
Bus Stops



PM2.5 Data Collected on Jan 30
200 100 0 200 Meters
N
Legend
Desirable Walking Route
Jan30_RealRoute
Jan30_Smoker

PM2.5 Data Collected on Jan 31




Study Area
Legend
Walking Route
N
180 90 0 180 Meters
Stewart St
7th Ave
8th Ave
Convention Pl
Pike St
Hubbell Pl
E Union St
Lenora St
Virginia St
3rd Ave
Westlake
Pine St
2nd Ave
5th Ave
Seneca St
Madison St
1st Ave
4th Ave
Union St
Jim Ellis Freeway Park
Virginia Mason Medical Center
Boren Ave
Elliott Ave
Pike St
University St
Piers 62 and 63
Seattle Aquarium
University Street
Seneca St
6th Ave
Waterfront Park
Union St
Post Al
Central Library
7th Ave
Marion St
8th Ave
9th Ave
Spring St
Madison St
James St
Post Ave
Seneca St
Marion St
4th Ave
5th Ave
Alaskan Way
Western Ave
2nd Ave
Columbia St
Cherry St
POLICE HQ
Jefferson St
6th Ave
James St
HMC
SFD 5
Pioneer Square
Post Ave
Terrace St
519
Yesler Way
Fortson Square
SFD 10
S Washington St
S Main St
Waterfall Park
S Main St