Tiffany Wu

UEP 237

Wind Impacts on Mobile Air Quality Measurements in Somerville, MA

Abstract

Air pollution in today's globalized, industrialized society poses a major environmental health risk, and millions of premature deaths from respiratory disease or heart attacks can be attributed to ambient pollution (World Health Organization, 2021). This problem is particularly acute in urbanized areas. As part of the effort to identify and combat air pollution, air quality monitors have been deployed by governments to measure pollutants with high levels of accuracy. Advances in technology suggest that low-cost mobile monitors can help augment existing fixed monitors in understanding air quality levels on a more granular level, such as in a neighborhood or street by street.

Wind speed can also impact the spread of pollutants, with suspended particulates responsible for smog conditions observed in London and Los Angeles (Cichowicz, et al. 2020). It is important to understand and control for how wind influences the spread of pollutants in order to understand pollution in a given area. Using existing data from a 2020 Tufts University study on urban traffic and air quality, this project will analyze how atmospheric conditions have impacted levels of ultrafine particles (UFP), black carbon (BC) and nitric oxide (NO) through the binning wind speeds and wind directions through the pandas cut function and comparing the mean concentration of the three pollutant types under each condition. Statistical data was further analyzed through the Seaborn and Plotly libraries for visualization, and k-means clustering was performed to compare unsupervised learning with domain expert categorization from the Beaufort Scale for wind speed classification. Spatial data was saved as a GeoDataframe and uploaded onto the kepler.gl web application through GeoPandas.

(Keywords: air quality, spatiotemporal, mobile, sensor network, wind speed)

Introduction

The City of Somerville is located in Greater Boston and sits at the crisscross of several highways, including Interstate 93 and State Highways 28 and 38. There is also a dense network of streets throughout residential and commercial portions of the city. Given that over 200,000 vehicles travel these roads on a daily basis (Jiang, Hudda, Durant, 2020) there is a significant amount of traffic-related pollution that can have adverse health impacts on nearby residents. In order to measure air quality more closely, a specially equipped vehicle with sensors has been set up to capture particulate data during drives. The Tufts Air Pollution Lab (TAPL) contains instruments to measure black carbon, nitrogen dioxide, and ultrafine particles.

With the construction of a large casino across the Mystic River in Everett, there is concern that additional traffic to and from the site could worsen air pollution levels in Somerville. To gauge potential impacts of the casino being built, a 2020 Tufts University study (Jiang, Hudda, Durant, 2020) analyzes the baseline air quality in the region and looks at spatiotemporal trends to better understand future air quality from changes in traffic volume. This study aims to build off the original study by analyzing how wind speed – collected at the same time as the rest of the data – may have influenced air pollutant concentrations and subsequent analysis.

Literature Review

In terms of using mobile sensors on vehicles to gauge air pollution, there are a few studies that come to mind. One is from a partnership between the Environmental Defense Fund and Google. By attaching specialized monitors to Google Street View vehicles, which travel through all local roads as part of their digital coverage, the researchers were able to gain a detailed understanding of how particulate matter concentration varied on a block-by-block basis in the city of Oakland, CA. This study involved approximately 3 million data points that were analyzed through various algorithms to produce a high-resolution map of pollution hotspots (Apte, et al. 2017). The researchers commented that this vehicle sensing approach was widely scalable and could be used to provide insight into the effectiveness of public management.

An interesting study that combined analyzing atmospheric conditions with air pollution was a 2020 paper from the Chinese Academy of Sciences, which used spatial interpolation to establish datasets of pollutants and meteorological elements that were then spatially matched from a study period of five years (Liu, et al. 2020) across China. This was a fixed sensor study that involved 896 air quality stations throughout the country, but shows insightful trends in pollutant concentration levels, including that particulate concentrations were generally decreasing, with the exception of ozone. The researchers noted that there was a strong negative correlation with wind speed, humidity, and precipitation. The results helped me formulate a hypothesis that higher wind speeds contribute to lower particulate concentrations and that different wind directions may play a role based on their origin.

Most importantly for this project, the 2020 Tufts paper conducted air quality monitoring using the aforementioned TAPL system on 1-second intervals was the source used for this project. Although travel modes, traffic speeds, and traffic volumes were extensively analyzed, wind speed and direction were included in the data collection but not yet incorporated – I hope to continue this study by researching how wind may have played a role in the study results for my project.

Research Questions

- What impact does wind speed have on particulate matter concentration?
- What impact does wind direction have on particulate matter concentration?

Data

I used data from the 2020 Jiang, Hudda, Durant paper for this project. The data was obtained from the Tufts Air Pollution Lab (TAPL). The data collection took place over several weeks in June and July of 2018, September to November of 2018, and January to February of 2019. Overall, there were 40 study days and 191 vehicle loops were driven around a route intended to capture pollutant concentrations on Interstate 93 and in surrounding neighborhoods in East Somerville.

In order to clean the data, I used pandas to read in my desired variables, *WS (Wind Speed), WD (Wind Direction), Latitude, Longitude, BC (black carbon), NO (nitric oxide), PNC (ultrafine particles)* and also included *Season, Day of Week, Day and Hour*. The variables I was most interested in were *WS, WD, BC, NO, and PNC*, but I included latitude and longitude for the spatial aspect of the study and looked into how seasonality and diurnal trends affected individual variables alone during the preliminary phase. I used pandas to find wind speed / wind direction and mean particulate counts to identify quantifiable trends. As this data has already been collected and is not raw data from TAPL, I focused on processing the data under the Methods section for analysis and visualization.

Methods

I read in the CSV file using pandas for the above variables and dropped duplicate rows. With my dataframe *wind_info*, I used the pandas cut function to bin wind speeds from Low (0-3 knots), Medium (3-6 knots) and High (6-10 knots). I also used the pandas cut function to bin wind directions from 0 to 22.5 degrees, and then in 45 degree increments to account for directionality, starting from N and then moving to NE around the cardinal and intercardinal directions. To account for the cyclical nature, I named the last label N2 and then replaced it with N for the full compass rose. My statistical visualizations were developed using the groupby function on these wind speed and wind direction bins and aggregating the data with mean NO, BC and PNC concentrations. This new dataframe I named *particle_by_bin*. I created a Seaborn facet plot to look at how wind speed and wind direction influenced the three particulates. I also created a wind rose plot using Plotly Express, which was a useful way of visualizing the data in an intuitive manner.

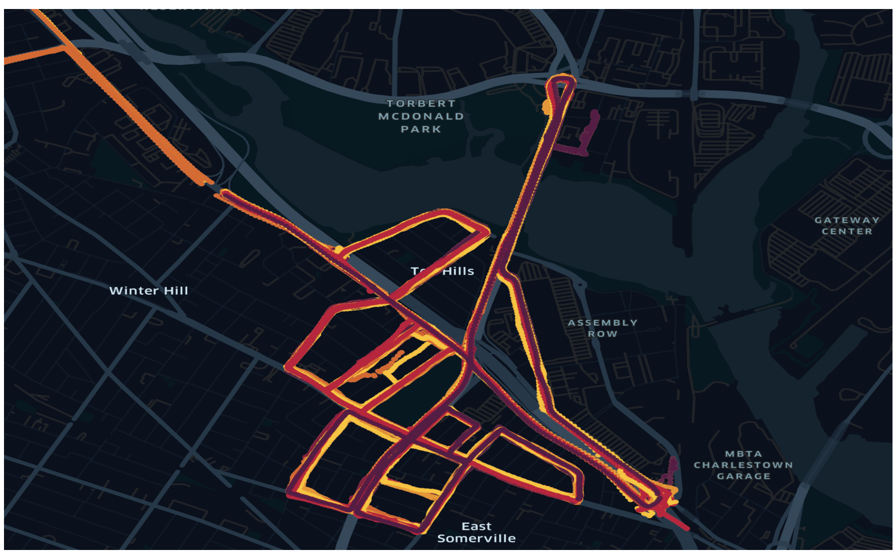
In order to determine what qualified as high, medium or low wind speeds, I utilized the Beaufort Scale (Beaufort, 1805), which is commonly used in navigational settings. On the scale, anything under 1 knot is classified as 0 (“Calm”), 1-3 knots is 1 (“Light Air”), 4-6 knots is 2 (“Light Breeze”), 7-10 knots is 3 (“Gentle Breeze”). For the sake of simplicity in this analysis, I had 0 to 3 knots represent the low wind speed altogether. To see if my decision to use the Beaufort scale with three different bins made sense, I normalized the wind speed and wind direction data and used K-means clustering to perform a silhouette analysis and tested the silhouette scores of clusters sized between 3 and 6.

Spatially, I visualized this data in kepler.gl using a CSV file that I exported from Jupyter notebook with the latitude and longitudes of the data points and joined wind speed information. I created a dataframe called *wind_glimpse*, which was essentially the same as *wind_info* except it included the latitude and longitude information I needed for this part of the study. For each particulate type, I then created a specific dataframe just for that particulate (e.g. nitric oxide). I converted the data to GeoDataframes by setting the geometry from lat and lon and then exported it by setting the driver as GeoJSON.

Results

Spatial Analysis

Concentrations for black carbon under high wind speeds, between 0 and 180 degrees.

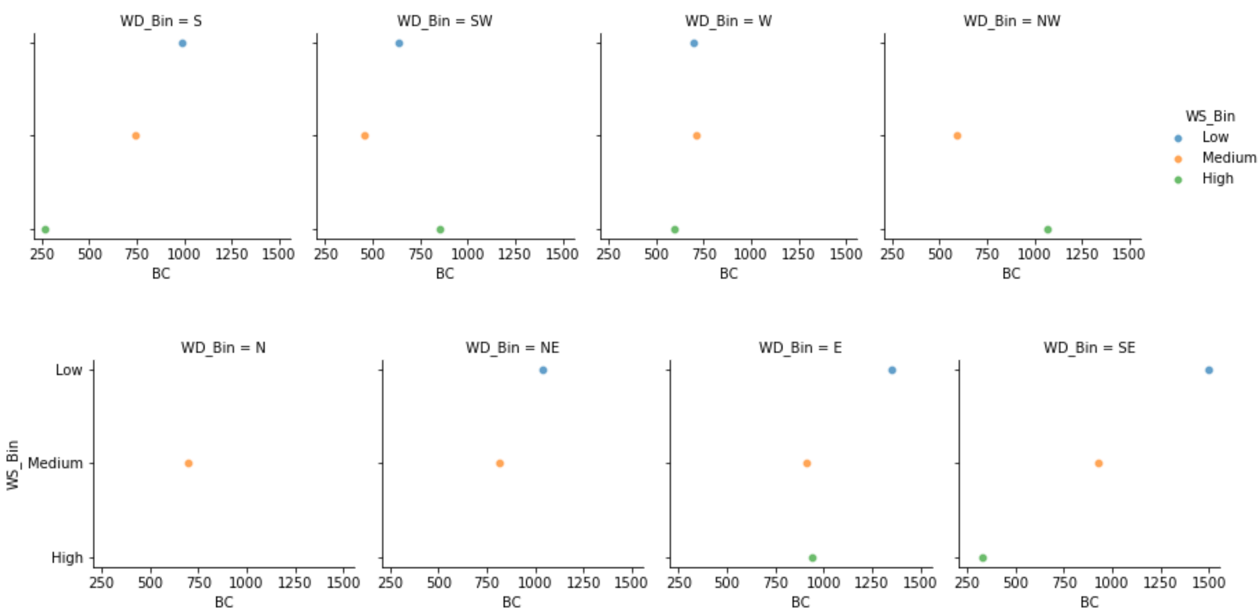


Concentrations for black carbon under high wind speeds, between 180 and 359 degrees.

The maps I produced showed the mean particulate concentrations for different wind speeds and wind directions over the study route taken by TAPL in East Somerville. I used the filter feature to select wind speeds in the High category (6-10 knots) and between 0 to 180 degrees for black carbon. Using the filter feature was helpful in comparing the impacts of both wind and direction on a spatial scale – because there were so many possibly combinations, I chose to include the two above for my screenshots in this report. The raw spatial data for BC, NO and PNC is included with the rest of my project submission. From these two visualizations, it is apparent that wind direction has an impact on particulate concentrations, as the second visualization shows much more range in concentrations overall. This is especially noticeable in the McGrath Highway area, which has lower black carbon levels compared to side streets in the first but the opposite in the latter. Here are the links to the kepler maps for black carbon (BC), nitric oxide (NO), and ultrafine particles (PNC) with filters applied. One can also toggle by Day of Week or Season to see spatiotemporal trends.

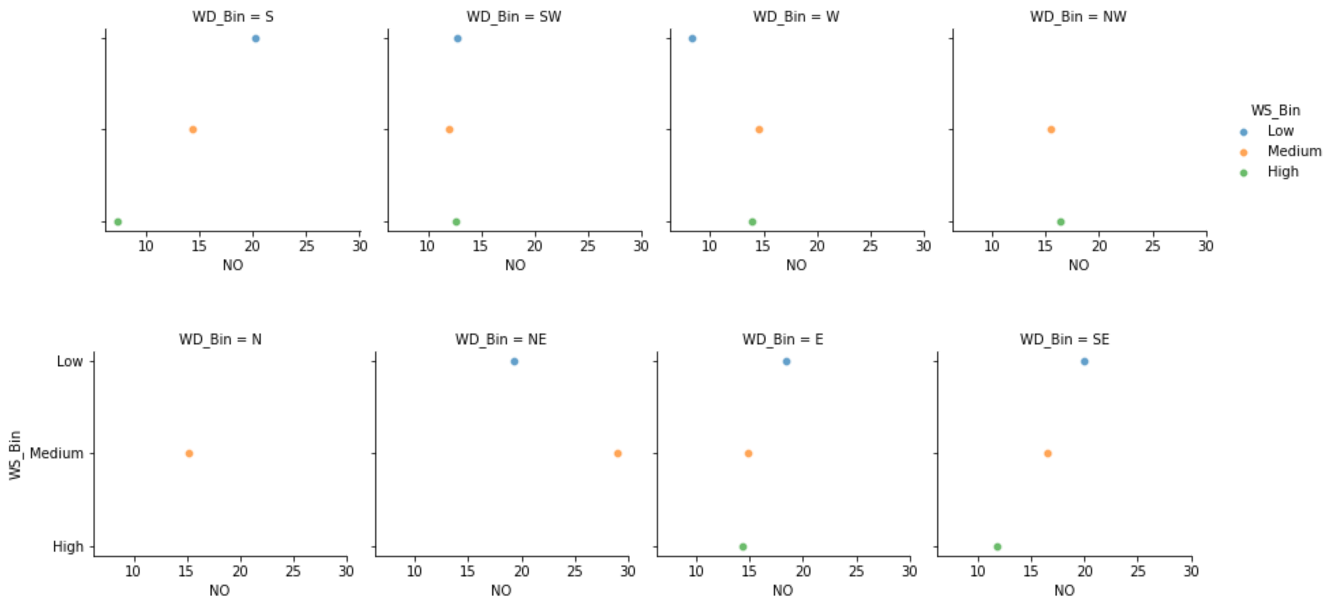
Statistical Analysis

I used Seaborn to create facet plots that showed the impacts of wind speed and wind direction on the mean concentration of each pollutant with the *particle_by_bin* dataframe.



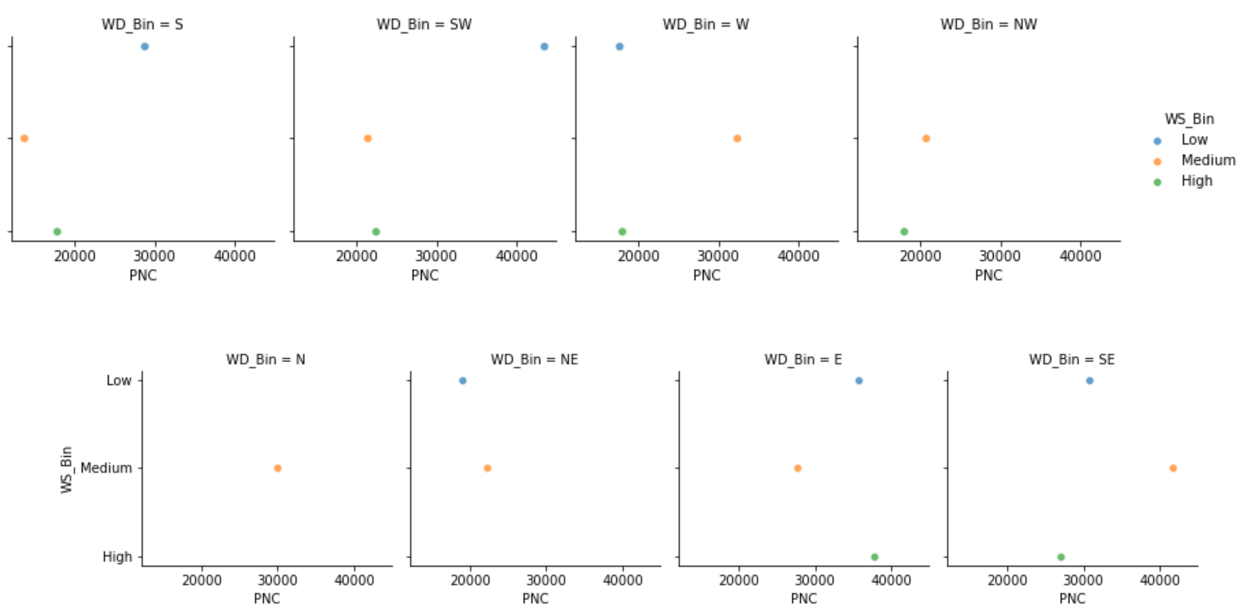
Mean concentrations of black carbon in ng/m^3 *in eight wind directions at three wind speeds.*

With black carbon, it seems that southeasterly and southern winds had the greatest impact on mean black carbon concentration levels. The mean concentration of black carbon with southeasterly winds in particular was around 250 ng/m^3 with high wind speeds, around 1000 ng/m^3 for medium, and over 1500 ng/m^3 for low wind speeds, so the effect of both speed and direction is quite noticeable.



Mean concentrations of nitric oxide in ppb in eight wind directions at three wind speeds.

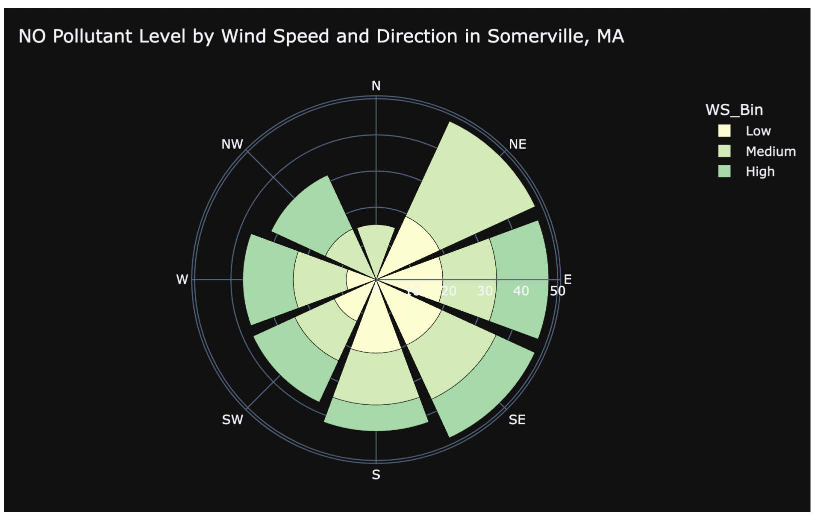
Nitric oxide appeared to be similarly influenced by wind speed and direction. Southeasterly and eastern winds again had the greatest impact when it came to different wind speed bins. Winds from the southwest appeared to have less of an impact on mean nitric oxide particulate levels.

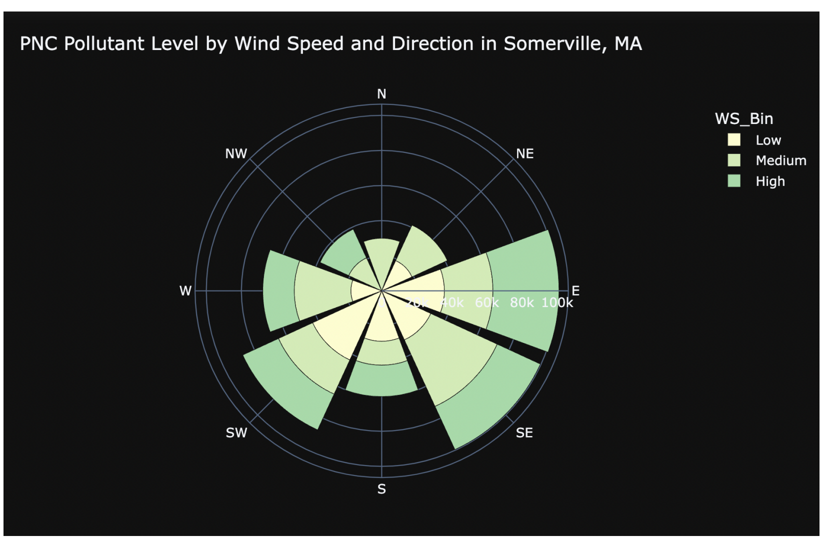


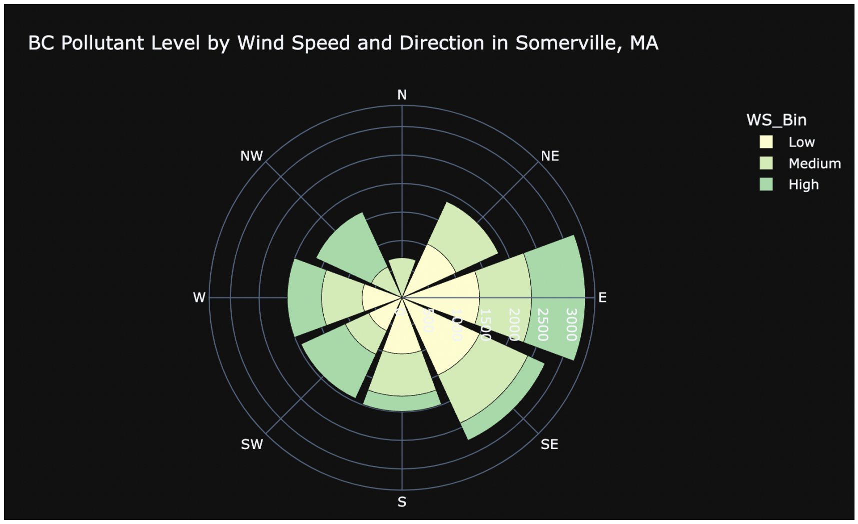
Mean concentrations of ultrafine particles in eight wind directions at three wind speeds.

Ultrafine particulates, measured in particles/cm^3, responded to varying wind directions and speeds in a manner that was not expected. This could be due to the concentration of particulates already present in an area or being generated by ambient vehicle traffic, which would mean that the concentrations are not as influenced by the two factors.

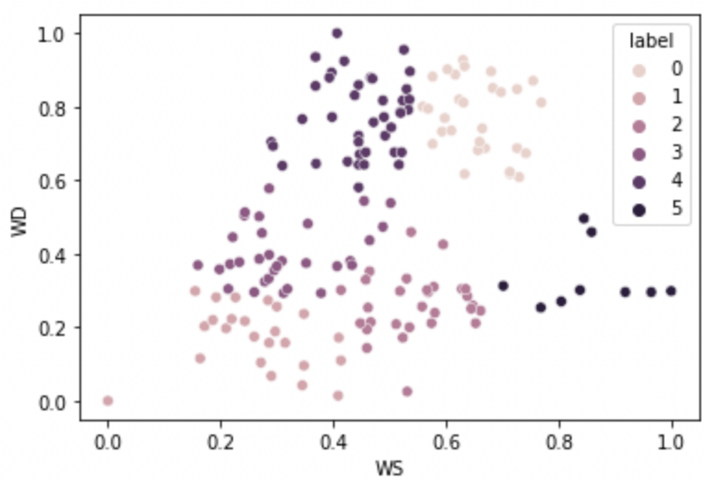
I used Plotly Express to create the wind rose plots shown below, using the *particle_by_bin* dataframe. I created separate plots for NO, BC and PNC. The size of the petals refers to the concentration of particulates, the directions are based on typical cardinal and intercardinal directions, and the color is designated by the wind speed. It was interesting to note that the mean concentrations were highest when the wind was blowing from the east or southeast for all three pollutant types. This type of plot illustrates the impact of wind speed and direction more clearly than the facet plot above. Interestingly, the NO mean pollutant level could be interpreted as being more concentrated with northeasterly winds when the mean pollutant level for BC and PNCs were much lower in this category and resembled each other more closely.



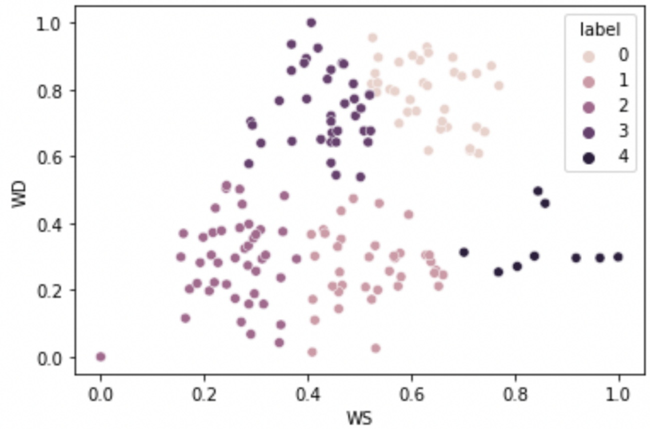




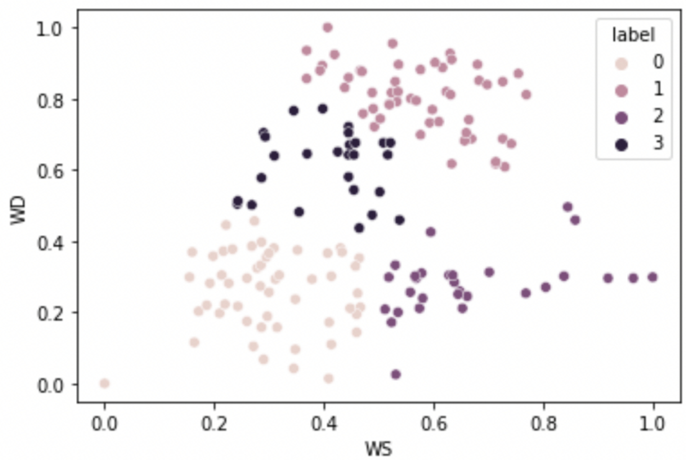
The last part of my statistical analysis was to run a K-means clustering analysis on the data to see if unsupervised learning from a silhouette analysis would agree with the domain expert Beaufort wind speed analysis, which used three bins. I had tested whether there were clustering patterns using K-means in the initial phase of my project, but had not normalized the data, so the large difference between degrees and knots was skewing the data towards the former. This time, I did normalize the data beforehand. I used scikit-learn for this portion and imported Kmeans and silhouette_score. The range of clusters I chose was 3, 4, 5, and 6 because I felt that two clusters was too few and more than double the domain expert classification was likely excessive. Having normalized the *wind_info* dataframe by importing MinMaxScaler, I proceeded to predict the cluster index. I then created a scatterplot of the results, using hue to differentiate between clusters in Seaborn.



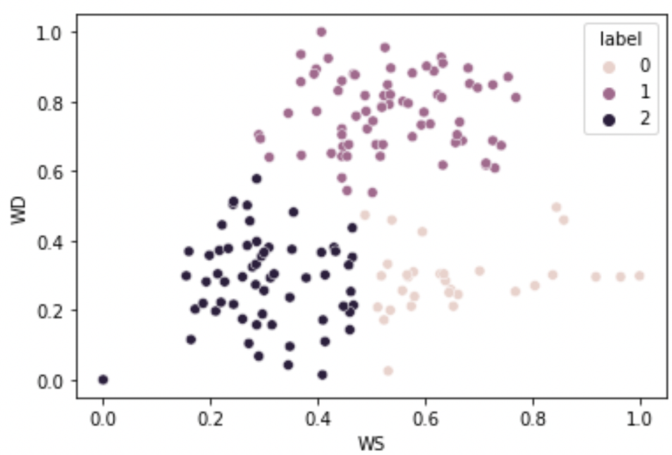
Visualization of clustered data with six clusters.



Visualization of clustered data with five clusters.



Visualization of clustered data with four clusters.



Visualization of clustered data with three clusters.

```
In [62]: k_means_info
Out[62]: {3: 0.4969151451779979,
          4: 0.39844532834427565,
          5: 0.37299444568164697,
          6: 0.3598092166904892}
```

Here were the silhouette scores – from the scores themselves, it appears that having three clusters is ideal for this analysis. But since we are looking at wind speed and wind direction together, it actually seems like two clusters for wind speed is sufficient based on the clustering. For roughly half the wind directions (from SW to N), wind speeds were just one cluster. For the other half, from N to SE or S, wind speeds formed two clusters. This again suggests that wind direction plays a role – more than I originally expected – in the mean concentrations of pollutants. One possible direction for further analysis would be to label the data based on wind speed and wind direction clusters, and analyze the results of the clusters separately.

Policy Implications

The potential impacts of my proposed research include building upon existing literature analyzing air pollution patterns and hotspots. Air quality monitoring, analysis and visualization also has implications for policymakers who may be interested in gauging whether a proposed regulation has been successful at curbing emissions. Since pollutant concentrations were most affected by wind speeds (causing their dispersion elsewhere) coming from the south and southeast, I looked at potential causes.

Geographically, Boston Logan Airport is east / southeast of Somerville, and aviation pollution can have a major impact on environmental quality. When winds blow from this direction towards the study area, they are likely picking up pollutants and taking them with them. Downtown Boston is south / southeast of Somerville, and the high density of traffic and emissions from engine combustion in the urban core of the city may also lead to particulates being tracked into Somerville when the wind direction brings it there. Because East Somerville and surrounding areas are considered an environmental justice community, it would be pragmatic for Massport and the City of Boston to look into ways to decarbonize the transportation and aviation sector. Some examples of how this can be achieved include using alternative fuels for aircraft (Massport, n.d.) Convincing policymakers to offer incentives to either the airport or airline companies that are making quantifiable efforts to decrease their emissions may be promising as well.

Future Work

I would like to continue exploring different facets of this dataset, such as potentially studying the impacts of green infrastructure on air quality by interpolating air quality data in nearby parks around East Somerville and comparing the results to data from roads with high levels of traffic. I could also incorporate demographic or socioeconomic data for an environmental justice and equity focused analysis that looks at which groups might benefit the most from parks and green infrastructure, and where new forms of greenery could be installed to help clean the air. Taking traffic levels into account could help with the question of ambient air pollution observed with the facet plots for PNC. I also want to look integrating diurnal or temporal distributions of mean pollutant levels, which is something I did not get to in this study but definitely hope to include in a subsequent project – looking into air quality on different weekdays or times of year (taking into account seasonality) using the same wind speed and wind direction parameters would likely reveal insightful results.

Appendix

Team member contribution statement: I worked on processing, visualizing, and analyzing the spatial and statistical data for this project on how wind speed and wind direction impacted particulate level concentrations in East Somerville and presented my findings to the class on Monday, December 11th.

Sources

1. Apte, J. et al., 2017. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environmental Science & Technology.* https://pubs.acs.org/doi/10.1021/acs.est.7b00891
2. Beaufort, F. 1805. Beaufort Wind Scale. *UK Royal Navy.* https://www.spc.noaa.gov/faq/tornado/beaufort.html
3. Chichowicz, R. et al. 2020. Effect of wind speed on the level of particulate matter PM10 concentration in atmospheric air during winter season in vicinity of large combustion plant. *Journal of Atmospheric Chemistry.* https://link.springer.com/article/10.1007/s10874-020-09401-w.
4. deSouza P. et al., 2020. Air quality monitoring using mobile low-cost sensors mounted on trashtrucks: Methods development and lessons learned. *Sustainable Cities and Society.* http://senseable.mit.edu/papers/pdf/20200601_deSouza-etal_AirQuality_SCS.pdf
5. Ferguson, L. 2021. Taking to the Road to Measure Air Quality. *Tufts Now.* https://now.tufts.edu/2021/05/11/taking-road-measure-air-quality
6. Liu, Y, et al. 2020. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Scientific Reports.* https://www.nature.com/articles/s41598-020-71338-7
7. Lim, C. et al. 2019. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environment International.* https://www.sciencedirect.com/science/article/pii/S0160412019304854
8. Jiang, S., Hudda, N., and Durant, J. 2021. Estimating the Relationship between Urban Traffic and Air Quality with Smartphone Data. *Proceedings of the 17th International Conference on Computational Urban Planning and Urban Management (CUPUM)*.
9. Massport. n.d. Air Quality at Boston Logan Airport. https://www.massport.com/massport/business/capital-improvements/sustainability/air-quality/.
10. U.S. Environmental Protection Agency. (2008). Reducing urban heat islands: compendium of strategies. https://www.epa.gov/sites/default/files/2017-05/documents/reducing_urban_heat_islands_ch_3.pdf
11. World Health Organization (2021). Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.