Homer Ayuste

CP321

March 31, 2023

Student ID: 190658320

Student Email: [ayus8320@mylaurier.ca](mailto:ayus8320@mylaurier.ca)

CP321 Project Report: What Conditions Affected the Amount of Air Pollution Cyclists Encountered in Ottawa During the Summer of 2010?

Introduction and motivation:

As a bike commuter and environmentalist, I was curious what conditions affect the amount of air pollution cyclists encounter in an urban setting. So, I set out to research pollution and find any potential datasets that contained information about cyclists and pollution. In general, pollutants that cyclists often encounter when cycling include: PM 2.5, nitrogen oxides, ultrafine particles, black carbon and more. In general, these emissions can come from many sources, but cyclists will often find that the most common source for exposure comes from riding near motor vehicles including cars, buses, and trucks. Thus, I will be looking at some of these pollutants and how exposure to motor vehicles can affect the amount of pollutants that cyclists are exposed to.

Approach:

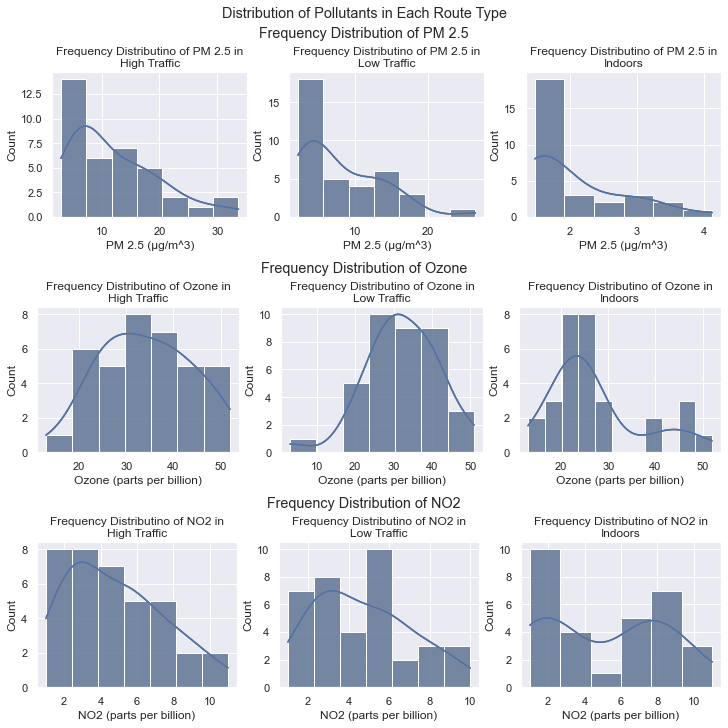
The dataset that I analyzed was published by Health Canada and contains measurements of pollutants recorded by cyclists. The data was gathered from a study of 42 healthy adults during the summer of 2010 in Ottawa. Participants cycled for one hour along high and low traffic routes, as well as indoors, and gathered measurements for ultrafine particles (UFPs) 0.1um in aerodynamic diameter, particulate matter (PM 2.5) 2.5 µm in aerodynamic diameter, and black carbon were measured along each cycling route.

Another dataset that was used is a dataset from the OECD that recorded each nation’s annual PM 2.5 average. This dataset will supplement the first dataset by providing the Canadian national average PM 2.5 level in 2010, thus allowing comparisons to be made to a baseline. Thus, working with these two datasets, I will research what conditions affected the amount of air pollution cyclists encountered in Ottawa during the summer of 2010.

For data processing, the function “read\_csv” from the Pandas library is used to convert the csv obtained from the government of Canada open data website into a Pandas dataframe. This creates a table with columns and rows of data. The columns include: participant ID, occasion, date, age, sex, wind speed, wind direction, wind direction in degrees, atmospheric pressure, ozone measured in parts per billion (ppb), NO2 measured in ppb, SO2 measured in ppb, average ultra fine particles (UFP) measured in number per cubic centimetre, black carbon measured in nanograms per cubic metre, bike temperature in degrees Celsius, relative humidity in percentage, particulate matter less than 2.5 microns in diameter (PM 2.5) measured in microgram per cubic metre, and route type or cycling location which include low traffic, high traffic and indoors.

There were some changes required to make the data understandable. First, the values in the sex column were originally represented by binary values, 0 representing male and 1 representing female. So, the values were converted to a string, “male” or “female”, thus making these values more understandable than 0 or 1. Next, the values in the ‘date’ column were transformed from a string to datetime format. This helps with plotting values in a time series visualization. And lastly, all rows containing any missing values were dropped from the table.

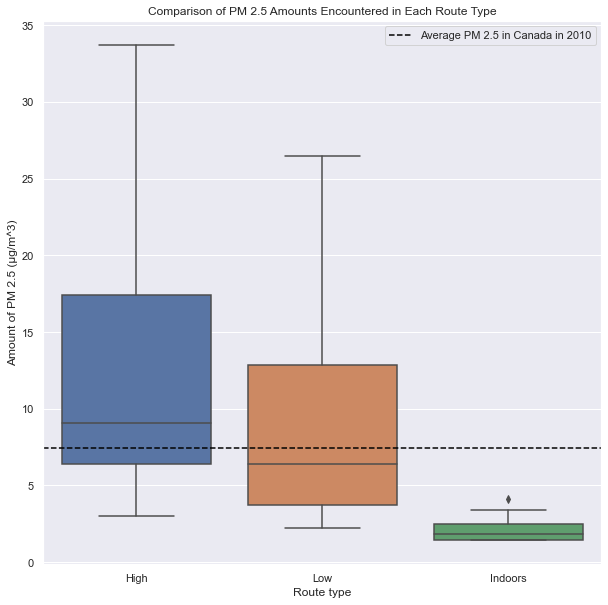
For the second dataframe containing nations’ annual PM 2.5 average, unnecessary columns were dropped and rows containing Canada were kept while every other row were dropped as well. Then, values in the ‘Year’ column were converted from a string to datetime format.



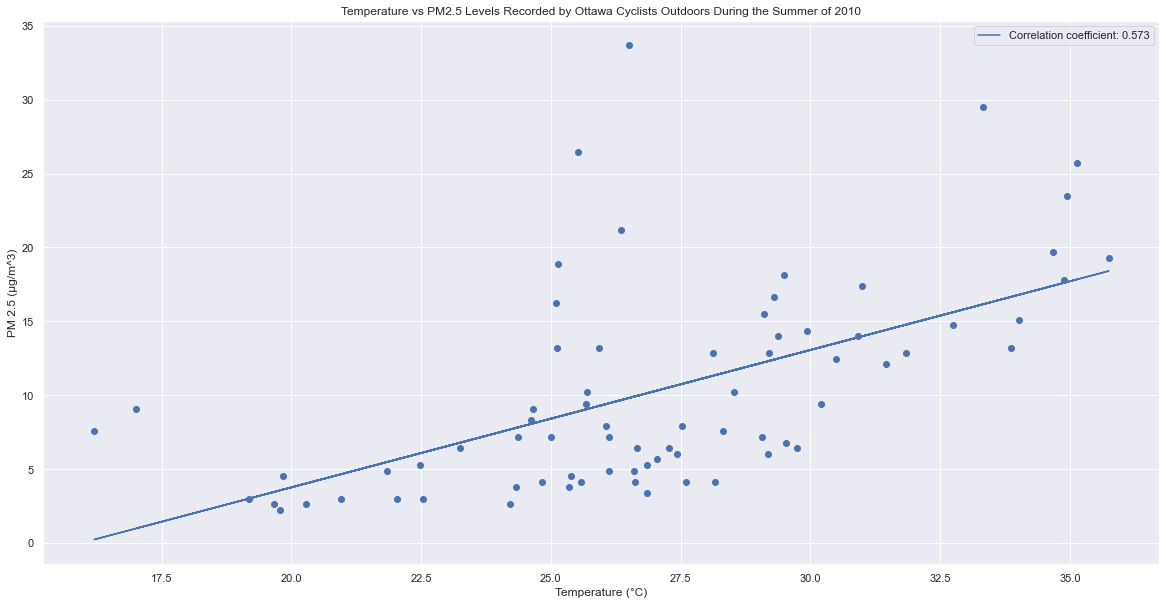
The above graph shows multiple histograms with each row showing the distribution for a different pollutant and each column showing a route. There is also a kernel density plot overlayed with each histogram. Looking at PM 2.5 levels, each graph is right skewed showing that the median is to the left of the mean. This means that most values lie to the left side of the graph and most riders in high traffic encounter lower levels of PM 2.5 than they encounter higher levels. However, most extreme PM 2.5 values are higher than lower (i.e., there are relatively few outlying low values).

When looking at ozone levels, there was no consistent skew type in the histograms/density plots amongst the different routes and each density plot showed a different shape. For example, the density plots for the ozone levels in indoors compared to low traffic show two different skews, right skewed versus left skewed. There is insufficient data collected to draw any conclusion on ozone levels and route types. But we can see that indoor rides encounter smaller amounts of ozone than outdoor rides. Yet, indoor riders can still encounter high ozone levels that are seen in high and low traffic rides. However, we can see that both high and low traffic routes generally have more occurences of high ozone levels compared to indoor rides.

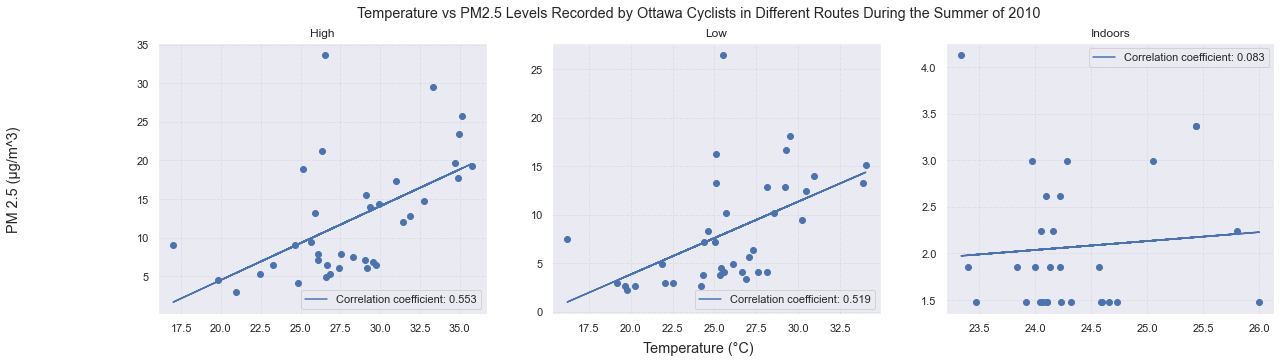
For NO2 levels, both high and low traffic routes show a right skew while indoor rides show a bimodal distribution. There may be an underlying phenomenon causing the bimodal distribution. For example, there could be days where less NO2 gets indoors and days where more NO2 gets indoors, but perhaps less days where a moderate amount of NO2 gets indoors because of other external factors. Interestingly, there appears to be more occurrences of higher levels of PM 2.5 indoors than in either traffic routes. However, since there are few data entries, it is possible that this density plot is not an accurate representation of NO2 levels seen indoors or outdoors.



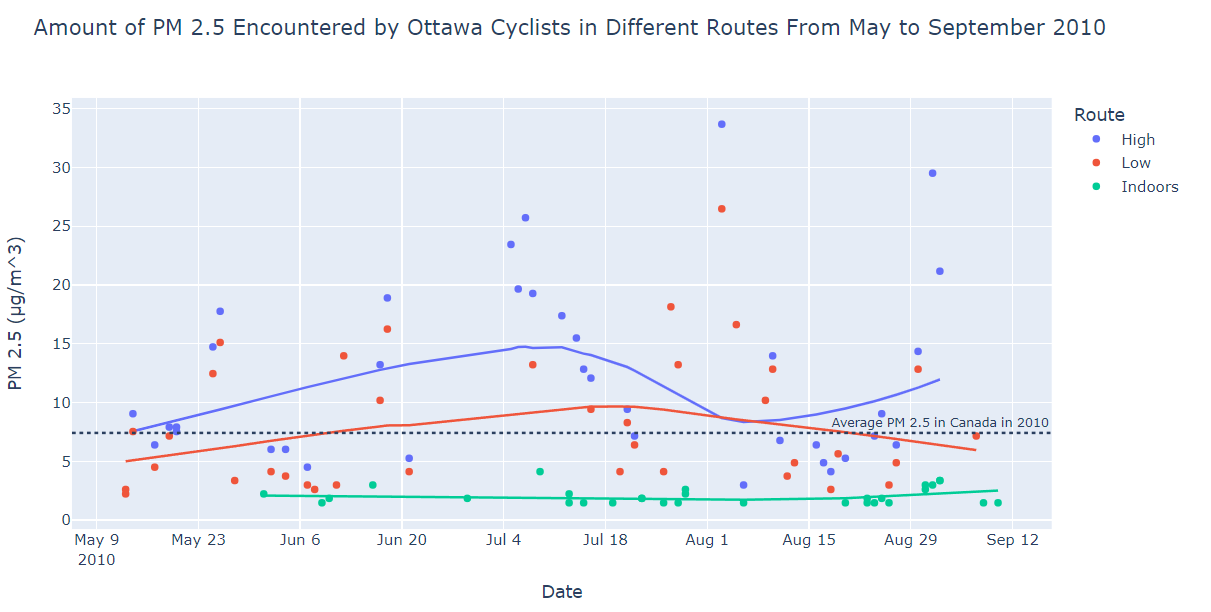
In this visualization, boxplots of each route type are shown to compare the amount of PM 2.5 encountered in each. The 2010 national average PM 2.5 amount of 7.5 µg/m3 is also shown to compare each route type to this baseline. We can see that cycling indoors never reaches that level and only reaches a maximum of approximately 4 µg/m3. For riding in low traffic areas, the median and minimum PM 2.5 level is below the national average. However, we can see that since the national average lies past the median, the percentage of rides through low traffic that encountered PM 2.5 values above the national average were less than 50%. In comparison to riding in high traffic, more than 50% of all rides encountered more than the national average PM 2.5 levels. An interesting observation is the minimums for both high and low traffic routes can reach levels as low as what is seen when riding indoors. Overall, riding through areas with high or low traffic appears to affect the amount of PM 2.5 that riders encounter. But riding through high traffic is most likely to expose cyclists to higher levels of PM 2.5. Meanwhile, riding indoors will expose cyclists to tiny amounts of PM 2.5.



This visualization shows a scatterplot with a line of best fit comparing temperature and PM 2.5 levels for rides recorded outdoors. The graph shows a moderate positive correlation between temperature and PM 2.5 levels. This indicates that temperature and PM 2.5 levels increase and decrease together. However, this relationship does not imply that an increase in one variable causes an increase in the other. It is possible that pollution and particulate matter coming from motor vehicles increase temperature. And it is also possible that an increase in temperature could cause more pollution to be created. For example, on hotter days, more people may drive thus causing more pollution. But this data cannot give any information on causality between variables, and it can only indicate whether there is a linear correlation between variables and the strength and direction of that correlation.



This visualization is like the previous one except a subplot is used for each route type. We can see that both high and low routes both show a moderate positive correlation between temperature and PM 2.5, as seen in similarly in the previous visualization. However, indoor rides show no linear correlation at all between temperature and PM 2.5. The lack of a correlation can be explained by common sense. Indoor cyclists are not near traffic, where cars emit pollution, and there are fewer heavy emitters of pollution indoors.



In this visualization, a scatter plot with dates as the x-axis. When riding indoors, the amount of PM 2.5 stays consistently around 1.7 to 2 µg/m3. Thus, there is no noticeable positive or negative trend noticed during the summer months for PM 2.5 encountered indoors. When riding through low traffic areas, the trend line shows that it increases from May before peaking in mid-July and declining from late-July to September to similar levels found in June. According to trends, PM 2.5 levels for low traffic routes go above the 2010 national average PM 2.5 amount of 7.5 µg/m3 starting in June before going below this level in late August. Meanwhile, trends for high traffic routes show that trends were consistently above the national average. However, this does not indicate that all rides on high traffic routes encountered levels above the national average as shown in the graph above. Given that there were less than 200 recorded rides in total, this may not be enough of a sample size to say these trends are accurate. Overall, it is possible that PM 2.5 levels start low in May and peak around July when riding through high or low traffic. We can surmise that the hottest month of the summer, July, is possibly correlated to the higher amount of PM 2.5 seen.

Conclusion:

Looking at some of the conditions recorded in the first dataset and observing whether they affected or correlated with the amount of pollution encountered by cyclists led to some interesting results. Regarding cycling location, cycling outdoors, especially in high traffic, leads to greater exposure to pollutants compared to riding indoors as seen in both the boxplot of PM 2.5 levels and histograms of PM 2.5, ozone, and nitrogen dioxide levels for each route type. We also looked at temperature which saw that PM 2.5 levels there is a positive and moderate correlation between these two variables when riding through traffic. As well as looking at a scatter plot by date and PM 2.5, we created trendlines for PM 2.5 levels based on the recorded route and found that PM 2.5 levels peaked in July, which is also the hottest month in the summer, for outdoor rides while indoor rides found essentially no changes in PM 2.5 levels throughout the summer. Overall, cyclists will find higher levels of pollutants in high traffic areas and during months with hotter temperatures in the summer.

Sources:

<https://open.canada.ca/data/en/dataset/b4673681-d738-4c35-acdb-9af63b00c59f>

<https://stats.oecd.org/Index.aspx?DataSetCode=EXP_PM2_5>