Joanna Riascos

Dr. Sylvain Jaume

Draft Manuscript

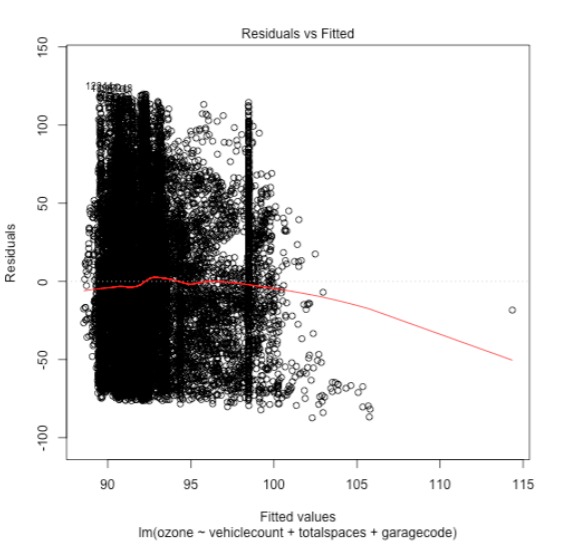
March 31st, 2017

**Draft Manuscript (Results)**

Our main proposal was to investigate if smart parking can help save the environment. There has always been an issue with parking worldwide. Finding a parking spot sometimes is nearly impossible. Since there are more cars then there exists a greater amount of car emissions that affect our cities’ environments. Therefore, we took the parking and the pollution dataset and merged these two datasets together to explore and analyze the different approaches and methods that we could use to contribute. Furthermore, we used our most significant variables from the dataset to implement and perform our analysis.

The variables that we used were the “total spaces,” “vehicle count,” “garage code,” and the “ozone layer.” Then we went ahead and used the stepwise model selection by AIC and we acknowledged that we needed to use the ozone layer as the dependent variable and the vehicle count, total spaces, and garage code as the independent variables. Next, we fitted our model and ran a multiple regression analysis. We used multiple regression analysis to better understand the relationship between the independent and the dependent variables. Also, multiple regression is a very flexible and commonly used statistical method in comparison to other statistical techniques.

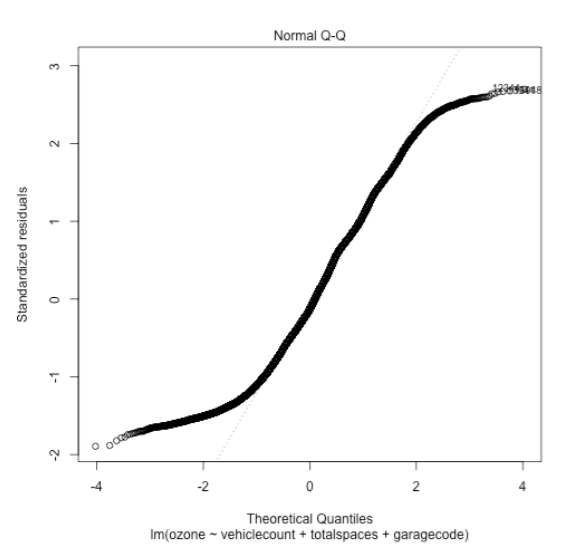
**Residuals vs Fitted**



First, we plotted the residuals plot to identify any potential outliers and to help us better understand if we could improve our regression model based on the residuals. This plot is also useful for checking the assumptions of linearity and homoscedasticity.

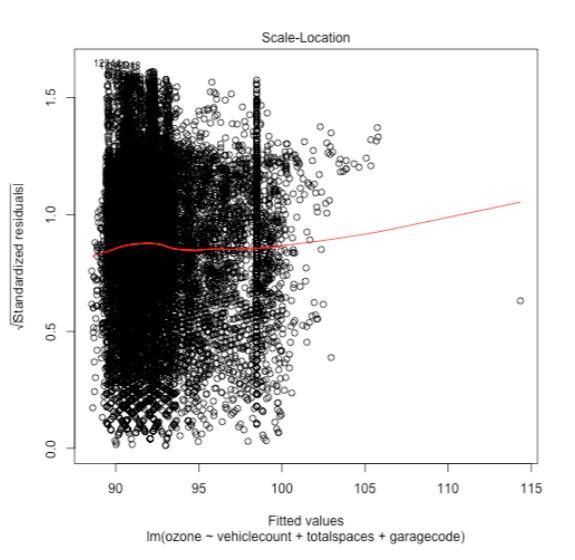
**Normal QQ plot**

 The quantile-quantile plot shows whether residuals are normal distributed or not. The normality assumption is evaluated based on the residuals. The QQ plot compares the residuals to “normal observations.”



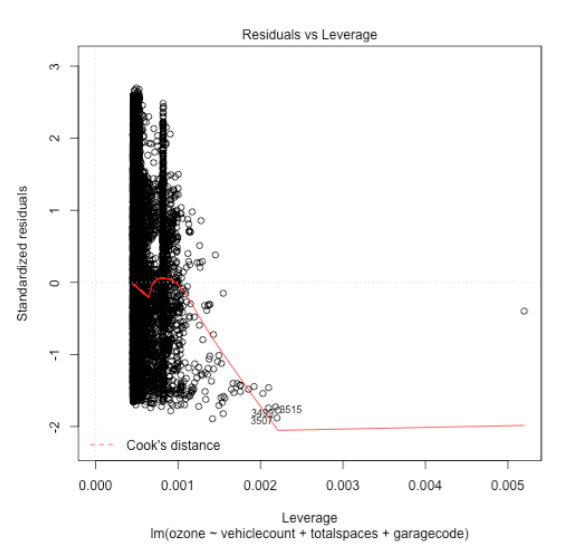
**Spread-Location plot**

The scale location plot depicts whether residuals are spread in equal portions along the ranges of predictors. We used this plot to check the assumptions of equal variance (homoscedasticity)



**Residuals vs Leverage**

The residuals vs leverage plot helps us identify any “influential cases.” Normally, not all outliers are influential in a multiple linear regression analysis.



After running the multiple regression analysis and plotting our models we then analyzed the structure of our model and created a couple of visual representations.

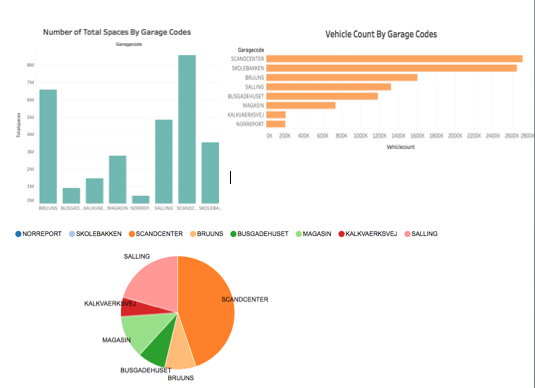
**System Data Center**

This system data center is a flowchart that depicts the structure of the “smart parking” system. The smart parking system checks whenever a car is parked or not. It checks when a car leaves the garage and when it enters the garage. When there is an available parking spot the system immediately notifies a user that was indenting to make a parking reservation. The driver can also check on the smart parking application if a parking spot is available. It also sends the user information of the parking spot details. All of this data gets stored in the data center. Similarly, the “QR Code” will be generated when a car goes in the parking garage. Through the “QR Code” the driver should be able to pay their parking reservation fee.

Untitled Diagram (10).png

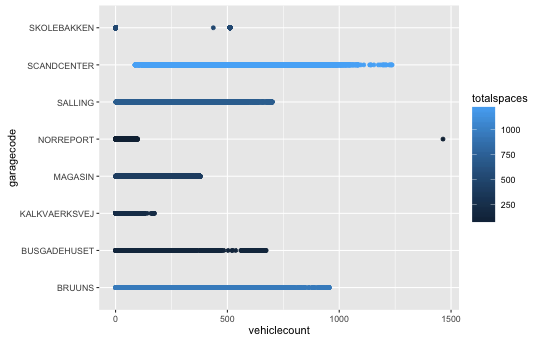
**Data Visualizations**

The following data visualizations depict the vehicle count by the garage codes. As we can see below in the charts, the most used garage code in reference to the vehicle count are the “scandcenter,” “skolebakken,” and “brunns."

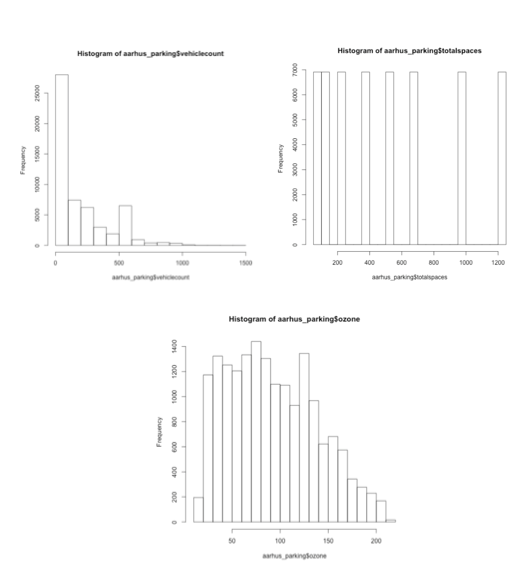


**R Scatterplot**

The r scatterplot depicts the garage codes, vehicle count, and the total spaces. It shows a visual representation of the amount of cars in accordance to the total spaces. The plot shows a line but in reality, they are dots. The dots represent the amount of cars that have the same garage code.

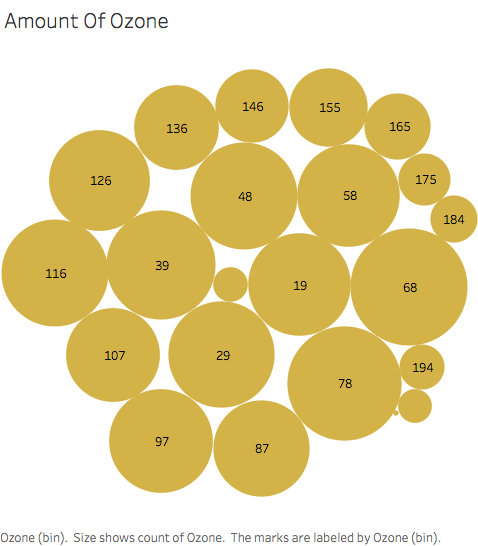


Furthermore, after running our multiple regression analysis we called the “summary” function and we plotted the outputted values using the “hist” function. The first histogram represents the frequency by the vehicle count. The second histogram depicts the frequency by the total spaces. Lastly, the third histogram represents the frequency by the ozone layer.

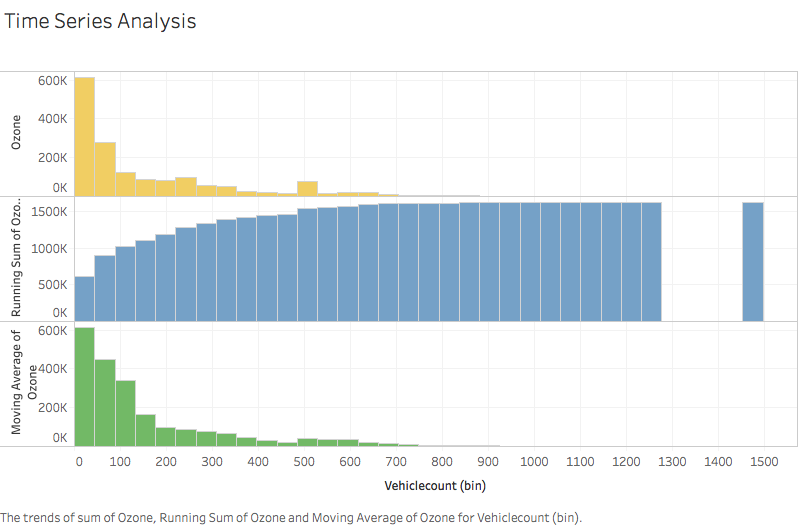


**Ozone**

Afterwards, the bubble chart shows the amount of ozone layer that we took from our dataset. As you can see in the chart below, the bubbles have different sizes. The bigger the size of the bubble is, the greater amount of ozone there is.



**Time Series Analysis**

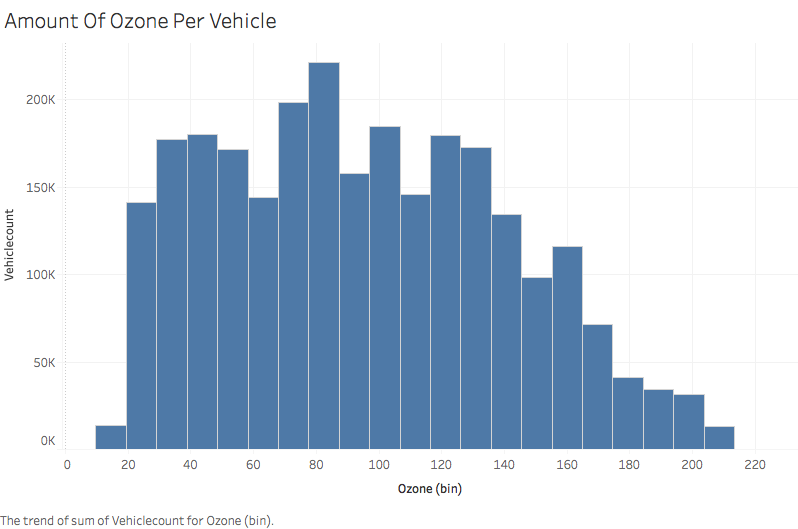


Moreover, we then performed the time series analysis to get a better understanding of the impact that our variables had on the ozone layer. We took the count of the vehicle count, the sum of the vehicle count, and the rank of the vehicle count. The first part of the chart shows the amount of the ozone layer, the second part of the chart shows a running sum of the ozone layer, and the third part shows a moving average of ozone.

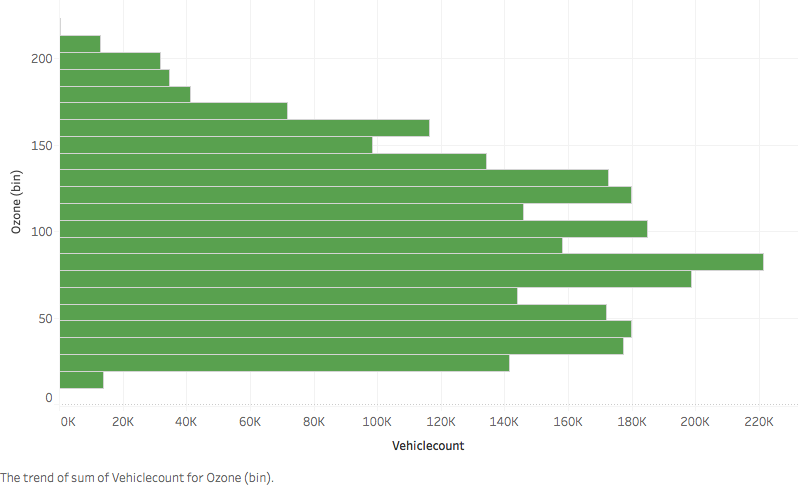
Time series analysis is composed of a series of methods for analyzing time series data that extracts meaningful statistics and other characteristics and parts of data. The time series analysis will help us identify numerous patterns in the correlated data. It also provides a compact description of the data. Similarly, it will enable us to understand and better model our dataset.

**Ozone Layer Per Vehicle**

The following chart depicts the impact of the vehicle count in regards to the ozone layer. As we can see in the screenshot below, when the vehicle count increases the ozone layer count decreases. The ozone layer is what helps the earth from being exposed directly to the sun’s ultraviolet rays. If the ozone layer decreases then there is a greater chance of environmental disasters.



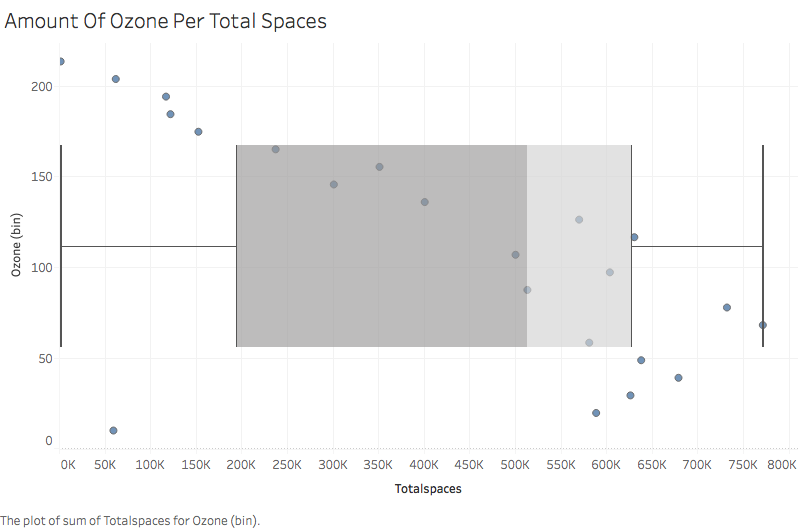
The following chart is very similar to the previous chart. The only thing that really changes is the field that we chose for the x-axis. The ozone layer field was chosen as the y-axis. There is an increase in the vehicle count and a decrease on the ozone layer.



**Boxplot**

We have also created a box plot for the amount of ozone and the total parking spaces.

As indicated in the box plot, the amount of total spaces is quite large. As we can see, the box plot is skewed left and most of the observations are shown on the high end of the scale.



In conclusion, we were able to identify which variables had the most impact on the ozone layer. We learned that the multiple regression analysis and the time series analysis helped us immensely in discovering any potential outliers. Similarly, these statistical techniques helped us in improving our model.