Bhargavi Poyekar
CH33454

Assignment 1

Part 1: Tool Review- <u>Tableau Public and Matplotlib</u>

Tableau Public is a free version of Tableau Desktop. It is a data visualization tool, with capabilities like- creating a hierarchy of qualitative data, hiding fields that are not used, customizable tooltips, joining multiple tables, creating new fields with numerical aggregations, and many more. It supports 6 data types: Text (String), Date and Time, Numerical which can be decimal or whole numbers, boolean, and geographical which are used with maps. Different types of basic charts supported include tables, heatmaps, maps, pie charts, bar graphs, tree maps, circle views, line charts, area charts, scatterplots, box and whiskers, Gantt charts, and bubbles. But this is not the limit for charts in tableau, we can use these basic charts, and create a combination to make detailed and complex charts. Tableau has a very user-friendly interface. It has a data source tab, where we can view the data, and combine multiple data. Its' sheets' tab allows users to make visualization, by selecting the rows and columns for the chart and lets you customize the colors, size, and tooltips. And the most interesting and fun part is that in this tool, we don't have to code, it is drag and drop, which makes it very easy to use. Except for this version of the tableau, tableau desktop, prep, online, and server is paid licensed tools.

Matplotlib is an open-source library for data visualization in python. It supports all platforms and any type of OS. It is interactive to some extent, where users can zoom or pan the figure. The image can be exported to many file formats. It supports all data types present in python. The basic visualization types supported are line plots, bar plots, scatter plots, and many statistical plots like histograms, box plots, and many more. The seaborn library, built using matplotlib is a higher version of matplotlib, making data visualization more customizable with respect to the sizes, shapes, and colors to display. Many complex charts can be created using matplotlib. As matplotlib is a library, it doesn't have any interface, it can be imported into any python environment.

The most important difference between Matplotlib and Tableau is that Tableau doesn't require coding which makes it easier for any user who can be a beginner or advanced. The interactivity is better in tableau, as the user can filter out and highlight the required information as needed, while matplotlib is better for exploratory data analysis. Tableau can be used by end users, who don't know to code. At the same time, matplotlib is better for expert coders, who want to analyze the data in various different formats, and complex structures, without the need to pay, in an open source environment.

Part 2: Visualization

Dataset: The data I am using for viz is the auto mpg dataset.

## Tableau image 1: Scatter Plot matrix for Numerical features.
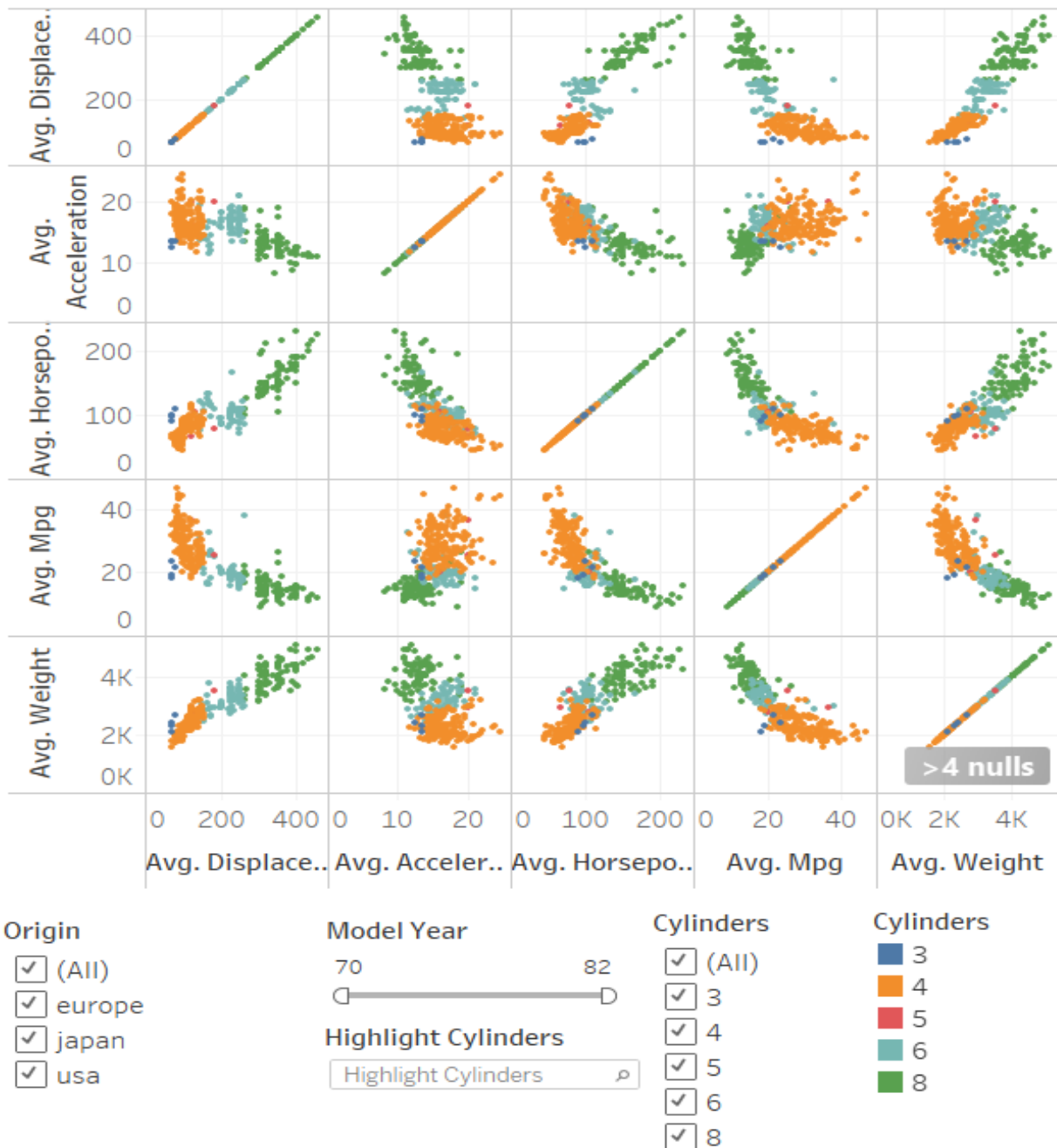
The above figure shows a scatter pair plot matrix for all the quantitative features. The color shows details about Cylinders. The scatter plot is based on Car-Name. The data is filtered on Model Year and Origin. The Model Year filter ranges from 70 to 82. The Origin filter keeps Europe, Japan, and the USA. The view is filtered on Cylinders, which keep 3, 4, 5, 6, and 8. I can find a few insights after analyzing the visualization:

1. As the weight of the car increases the mileage per gallon decreases.
2. We can also see that the more horsepower of the car, the less mileage.
3. The average displacement also follows an inverse relation with the mileage.

Tableau image 2: Distribution Plot of MPG:
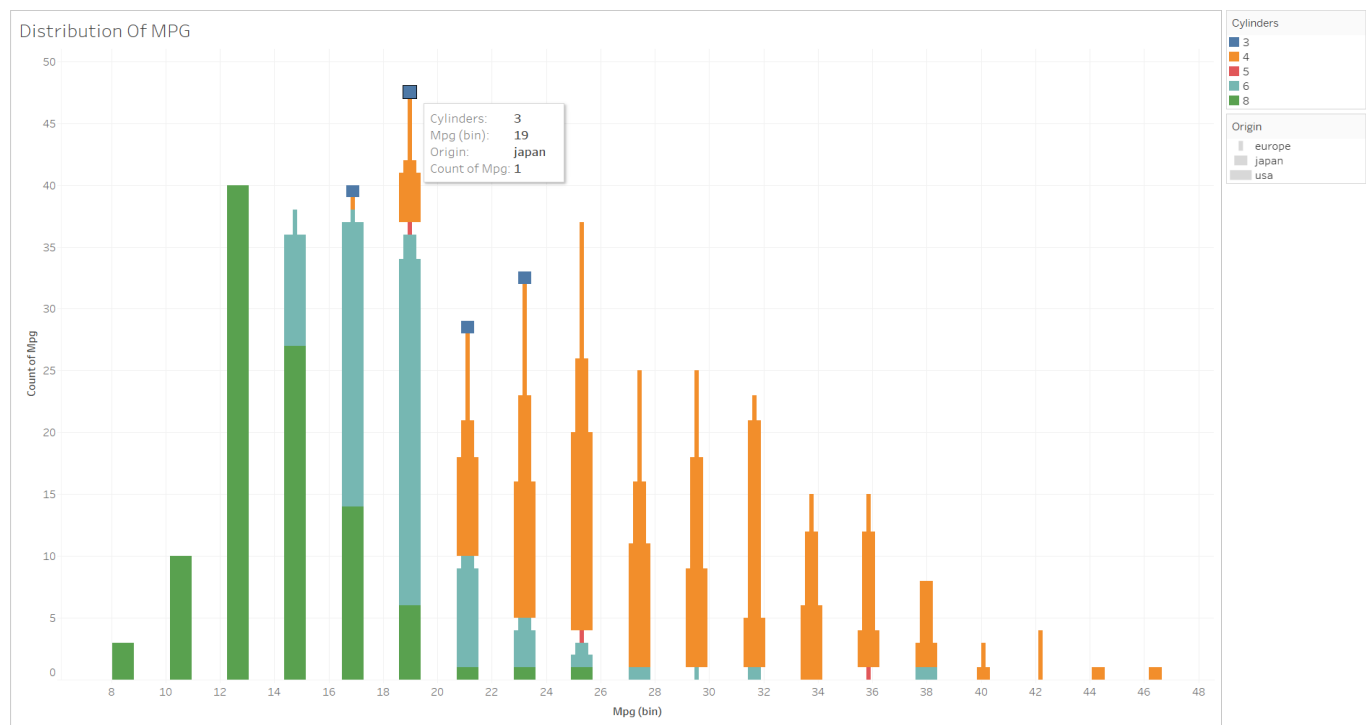
The above figure shows the distribution for the MPG using a histogram. It is a right-skewed normalized distribution that shows a trend of the count of Mpg for Mpg (bin). The color shows details about Cylinders. Size shows details about Origin. We can see from the figure that as the cylinders decrease, the mileage increases. I cannot see any particular trend with respect to the origin and mileage.

Matplotlib Image 1: Relationship between Horsepower and Acceleration
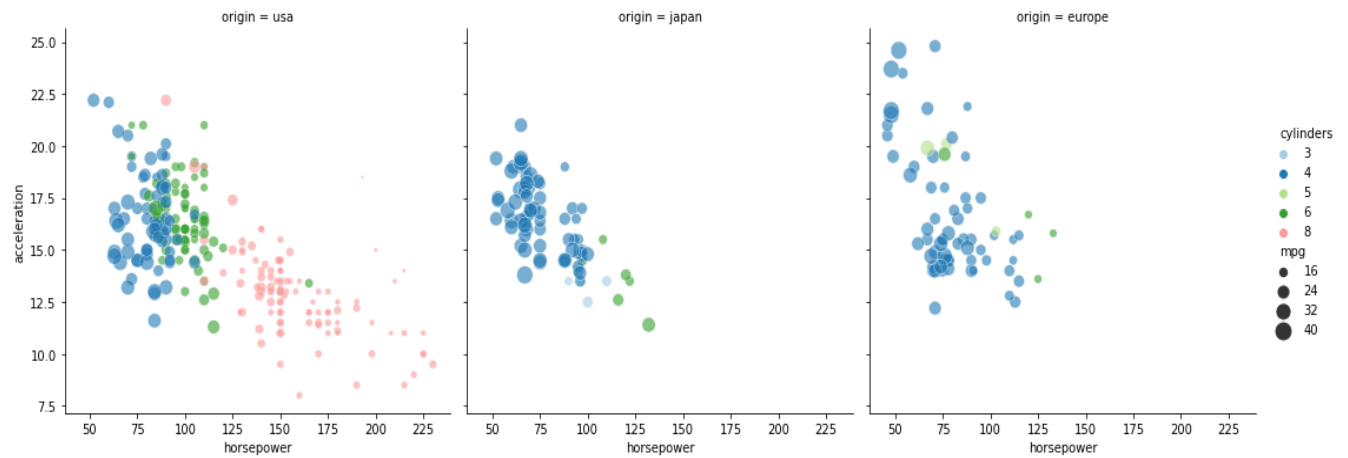
The above figure shows the relationship between Horsepower and Acceleration bifurcated by the origin, cylinders, and mpg level. The size of bubbles increases as the mpg increases and the colors are based according to the no of cylinders. From the figure I can tell:

1. In every region, I can see a negative correlation between horsepower and acceleration.
2. Vehicles having high horsepower and low acceleration have low mileage.
3. Vehicles with high horsepower and low acceleration have less number of cylinders..

Matplotlib Image 2: Change in quantitative features with model_year and origin.
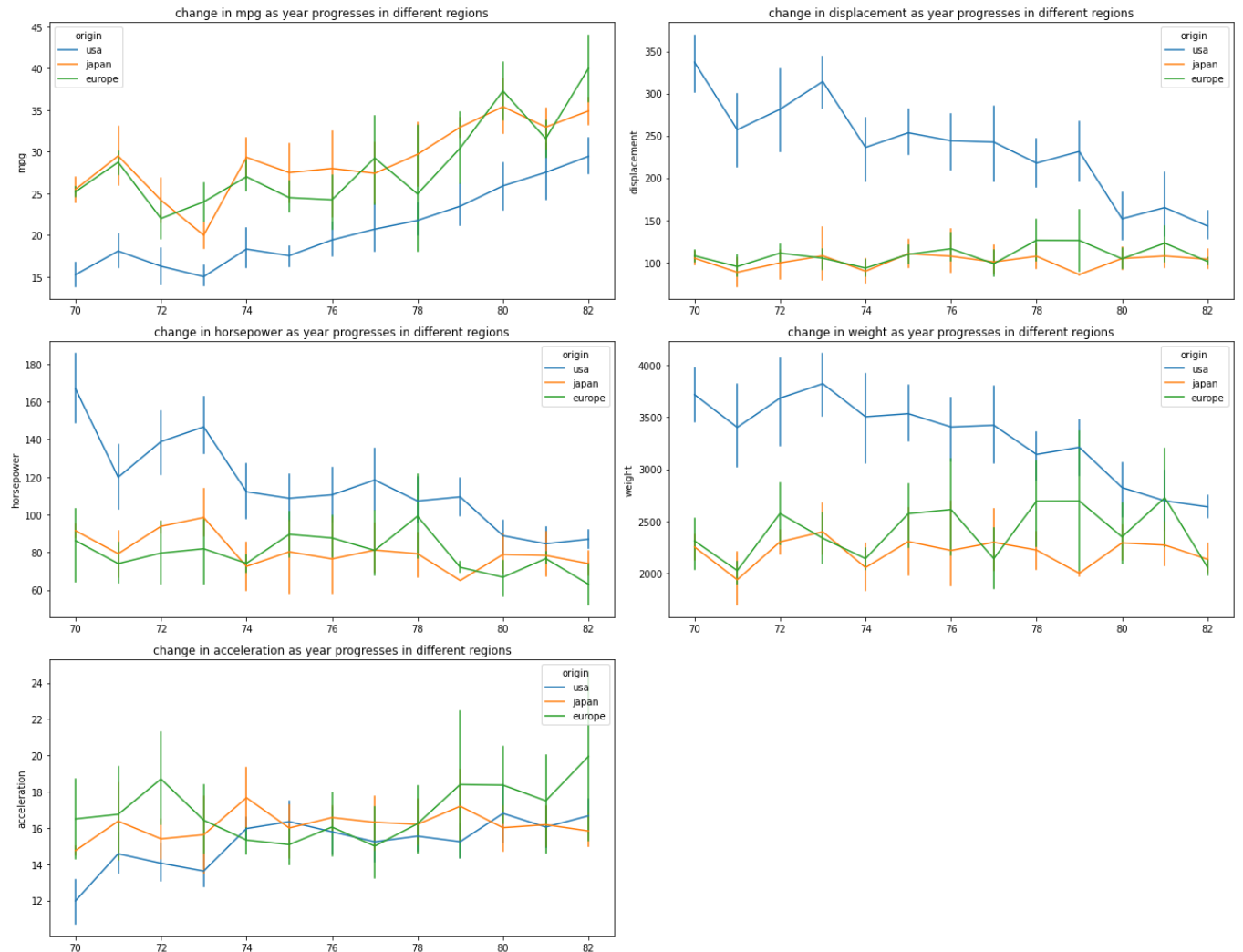
The above figure shows the change in the quantitative features with respect to the model year for all regions. The colors represent the regions. From the figure, I can analyze:

1. The mpg increases in all regions, as the year progresses.
2. Displacement, horsepower, and weight decrease in the USA, as the year increase, and there is no significant trend for Japan and Europe for these 3 features.
3. The acceleration increased in the USA, as the year progresses.

References:

1. https://help.tableau.com/current/pro/desktop/en-us/gettingstarted_overview.htm
2. https://help.tableau.com/current/pro/desktop/en-us/buildexamples_scatter.htm
3. https://help.tableau.com/current/pro/desktop/en-us/buildexamples_histogram.htm
4. https://www.quora.com/Which-one-is-the-best-between-matplotlib-and-Tableau
5. https://matplotlib.org/stable/plot_types/index.html
6. https://www.oreilly.com/library/view/python-data-science/9781491912126/ch04.html