Final Project

Datasets: Ride Austin

What the dataset is about:

- This dataset includes Uber and Lyft trips information have been made in Austin from 6/3/2016 through 2/7/2017
- The columns including start/complete time, distance, Driver rating, Rider rating, car category, color, make, model, year, data, speed, start location and end location

How I got it?

Link:
https://public.opendatasoft.com/explore/dataset/rideaustin/table/?dataChart=eyJxdWVyaWVzIjpbeyJjaGFydHMiOlt7InR5cGUiOiJjb2x1bW4iLCJmdW5jIjoiQ09VTlQiLCJ5QXhpcyI6ImRpc3RhbmNlX3RyYXZlbGxlZCIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMyQzNGNTYifSx7InR5cGUiOiJsaW5lIiwiZnVuYyI6IlNVTSIsInlBeGlzIjoiZGlzdGFuY2VfdHJhdmVsbGVkIiwic2NpZW50aWZpY0Rpc3BsYXkiOnRydWUsImNvbG9yIjoiI0VDNjQzQyJ9XSwieEF4aXMiOiJkYXRlIiwibWF4cG9pbnRzIjoiIiwidGltZXNjYWxlIjoiZGF5Iiwic29ydCI6IiIsImNvbmZpZyI6eyJkYXRhc2V0IjoicmlkZWF1c3RpbiIsIm9wdGlvbnMiOnt9fX1dLCJkaXNwbGF5TGVnZW5kIjp0cnVlLCJhbGlnbk1vbnRoIjp0cnVlfQ%3D%3D

This is an open dataset from OpenDataSoft.

StoryLines:

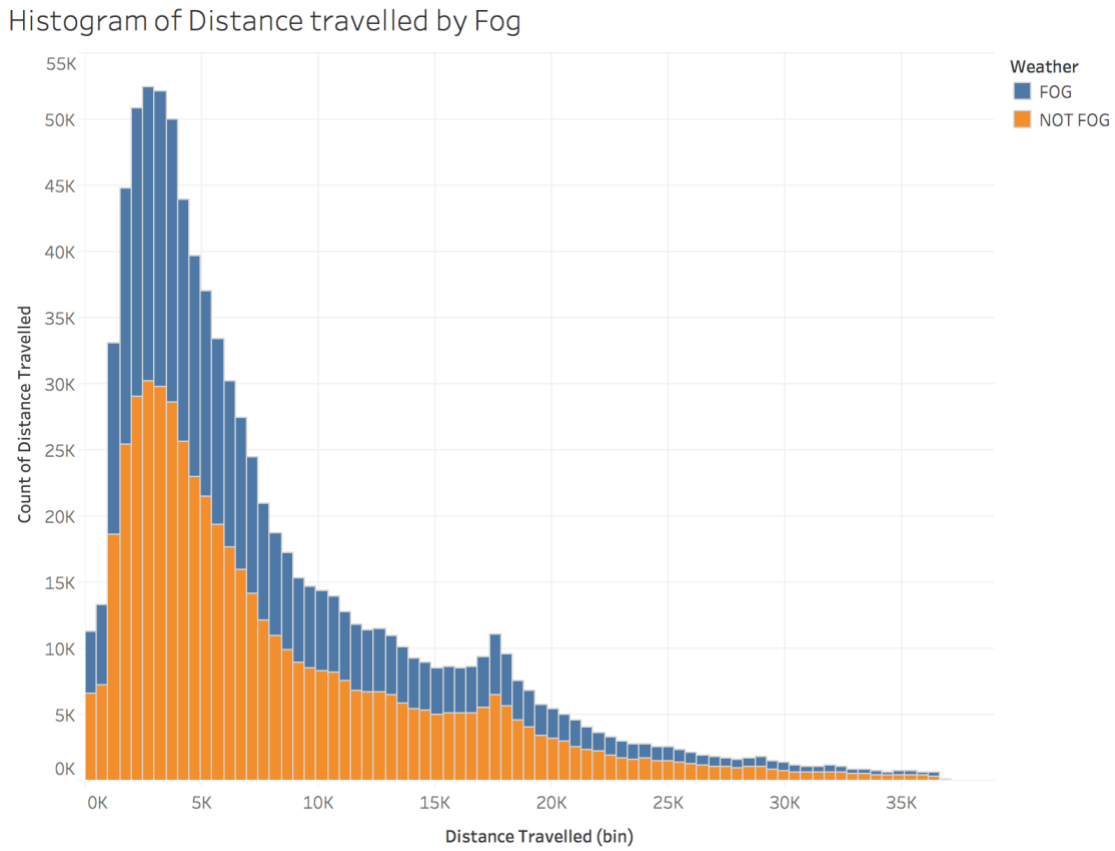
Based on this dataset, I mainly analyzed and visualized in three perspective:

1. The relationship between weather and travel distance
2. The relationship between car make and travel distance & ratings
3. The relationship between Trip and Geographic & Time

Plot

Weather and Travel Distance

- **Histogram**

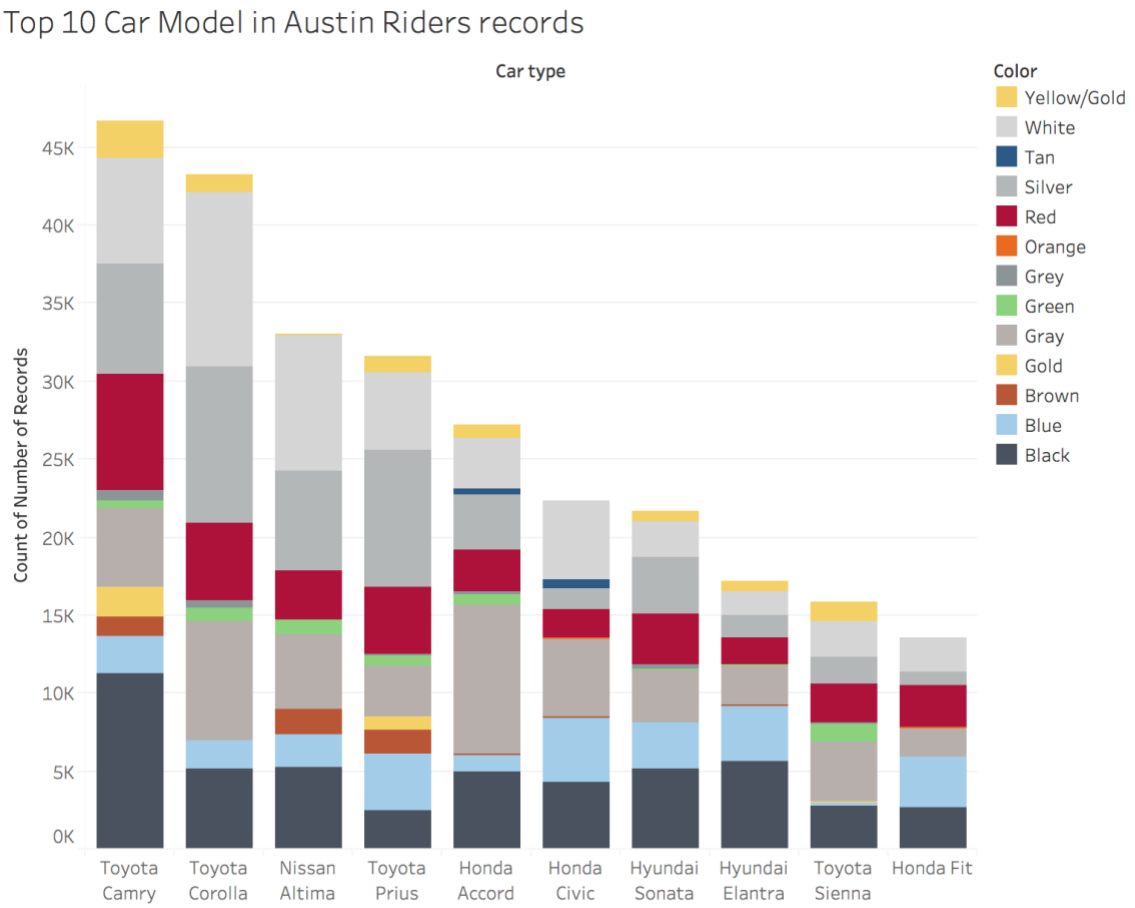


The plot of count of Distance Travelled for Distance Travelled (bin). Color shows details about Calculation2.

From this plot, we can see that with the Fog Weather, there are more prople choose shareriders than usual time. But the overall distribution of Distance travelled is similar.

Car Make & Distance Travelled & Ratings

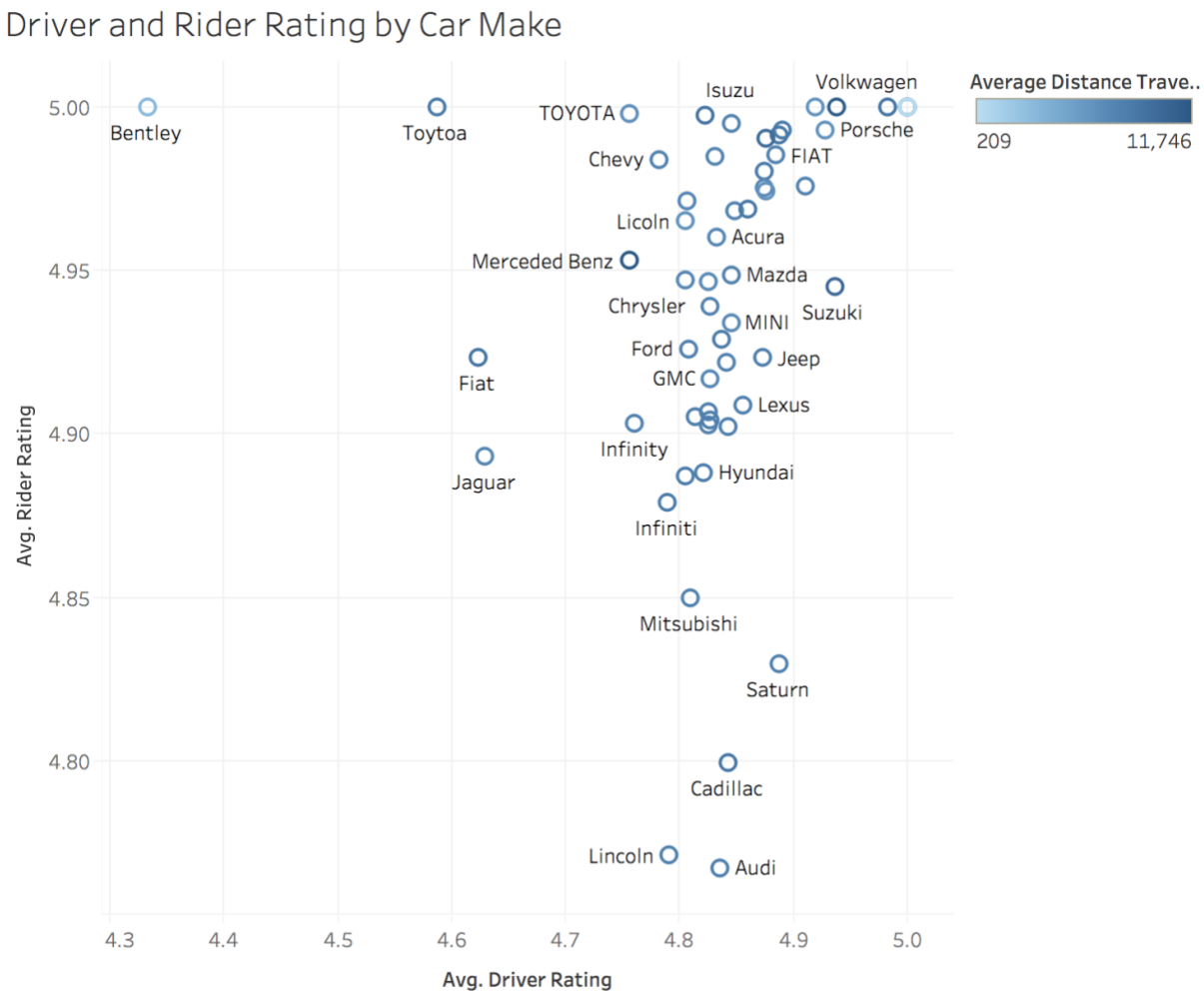
- **Bar plot**



Count of Number of Records for each Car type. Color shows details about Color. The view is filtered on Car type, which keeps 10 members.

From this Plot, we can see that among all share riders' car models in Austin, the most common car models are from Toyota, Nissan, Honda and Hyundai which are all economic car make. We can also see that black, white and grey are the most common colors of share riders' car.

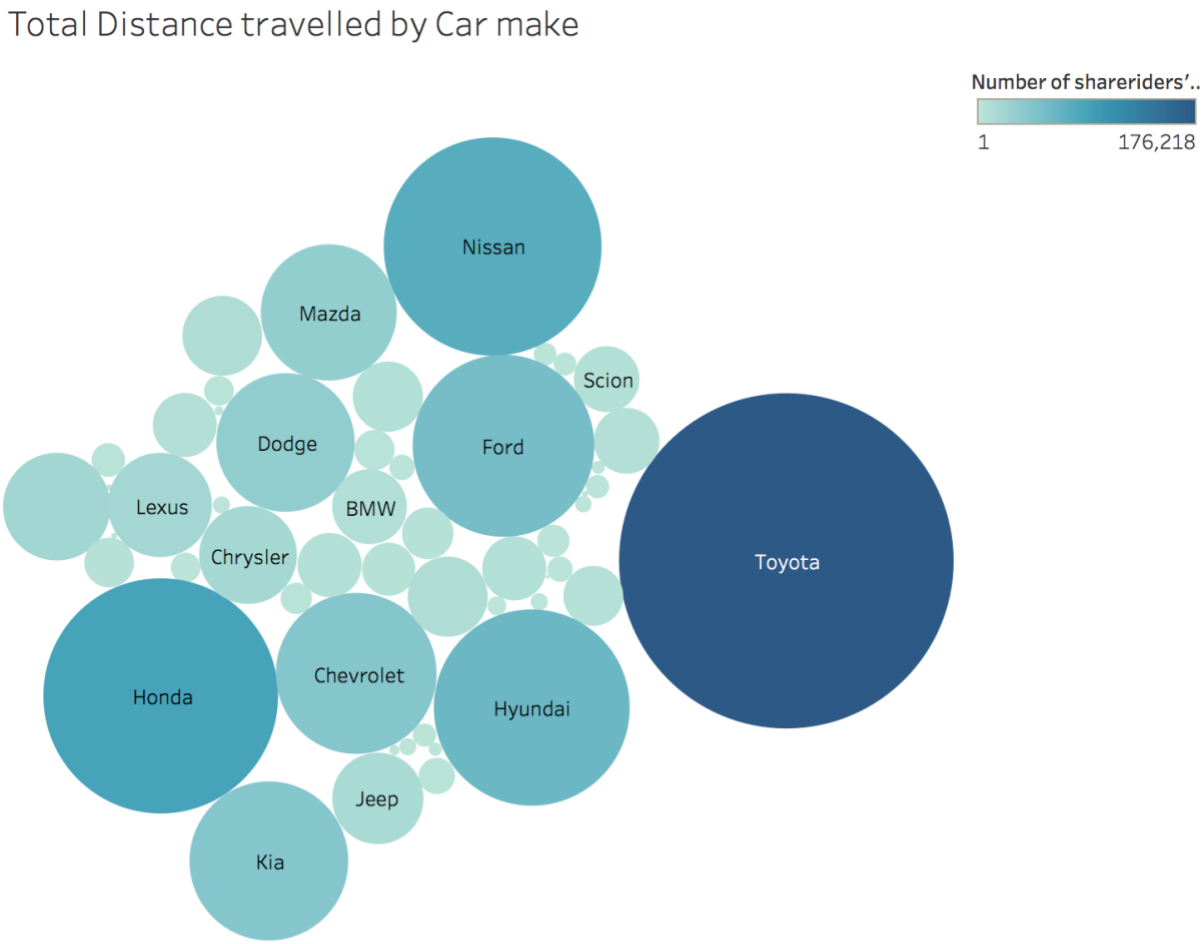
- **Scatter Plot**



Average of Driver Rating vs. average of Rider Rating. Color shows average of Distance Travelled. The marks are labeled by Make. Details are shown for Make.

From the Scatter plot, we can see that the relationship between average Driver and Rider Ratings by car make. On the right upper corner are the car make with higher Driver ratings as well as higher Rider ratings. In our case, Volkswagen is the best car make.

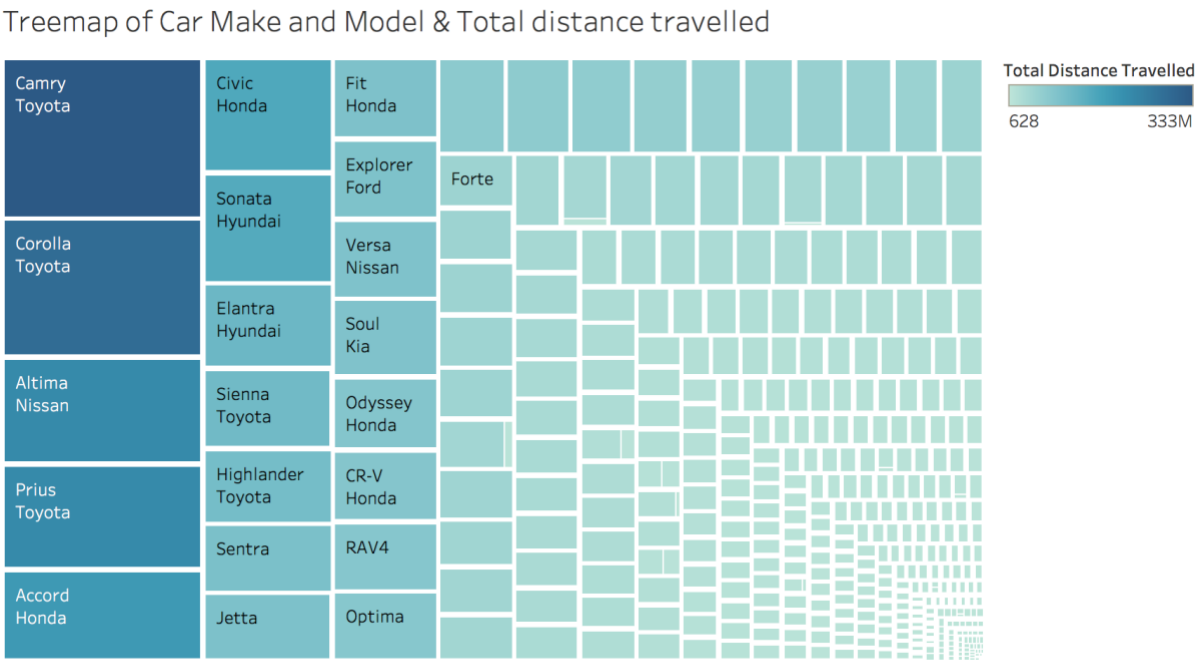
- **Bubble Map**



Make. Color shows count of Make. Size shows sum of Distance Travelled. The marks are labeled by Make.

From this plot, we can see that Total distance travelled by car make, colored by the count of trips. We can see that Toyota, Honda, Nissan, Hyundai, Chevrolet and Ford are the most popular car makes in Austin area.

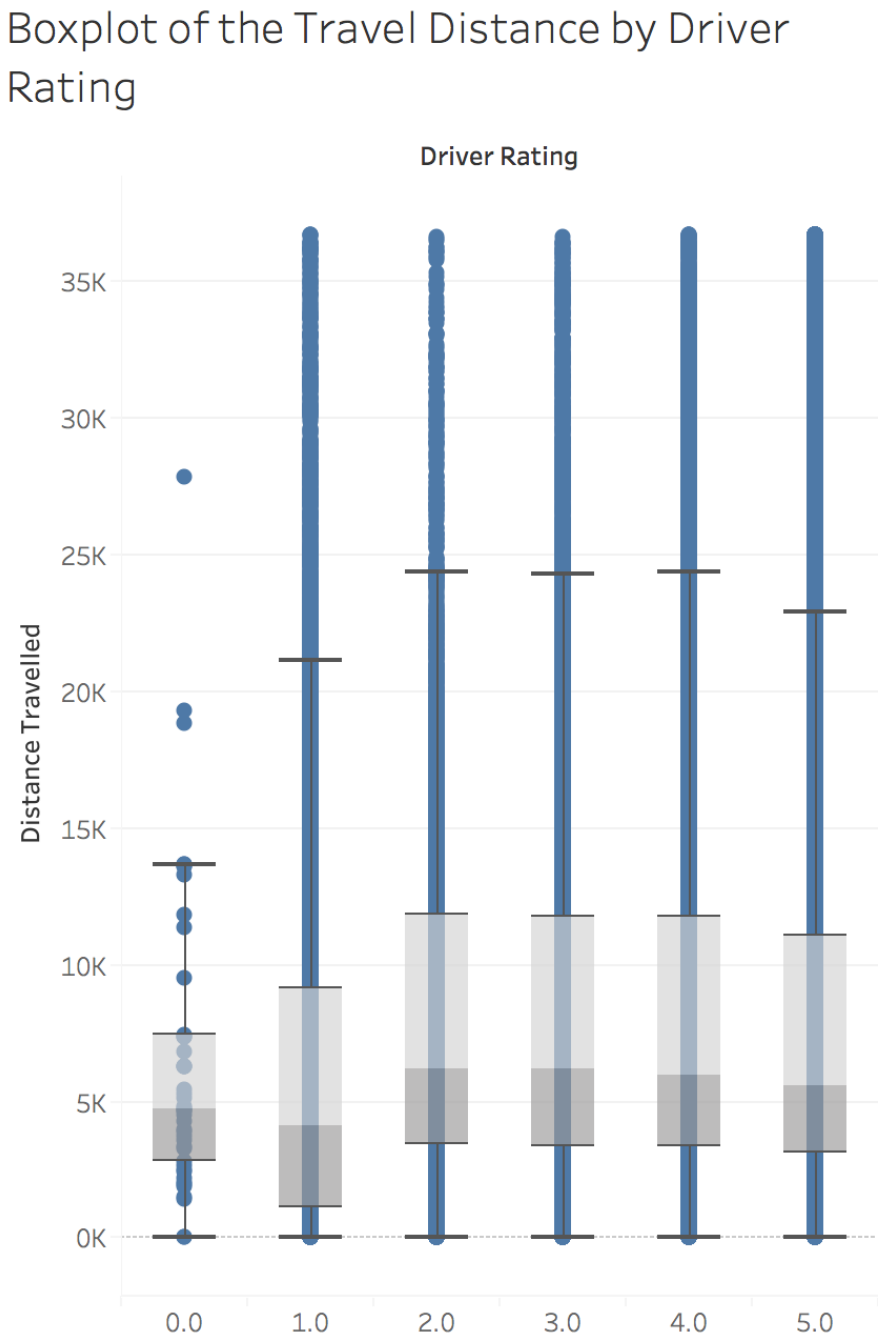
- **Treemapping**



From this Treemapping, we can see that car make segmentations in the share rides market in Austin Area. The Toyota Camry, Corolla Toyota, Altima Nissan, Prius Toyota and Accord Honda take the largest proportions in the market.

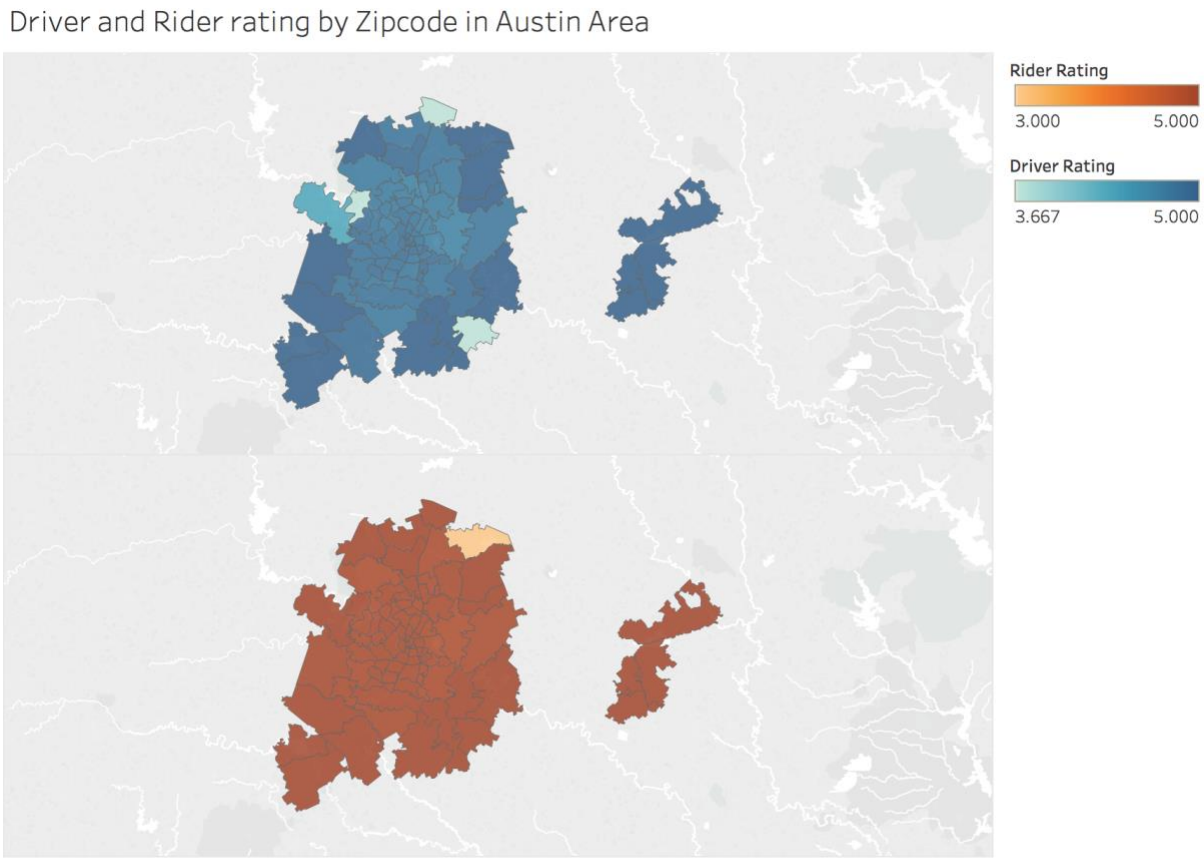
Trip Rating

- **Box Plot**



From this plot, we can see that Different Driver Rating level has slightly different distance travelled amount. For Rating 2,3,4 has the relative higher travelled distance.

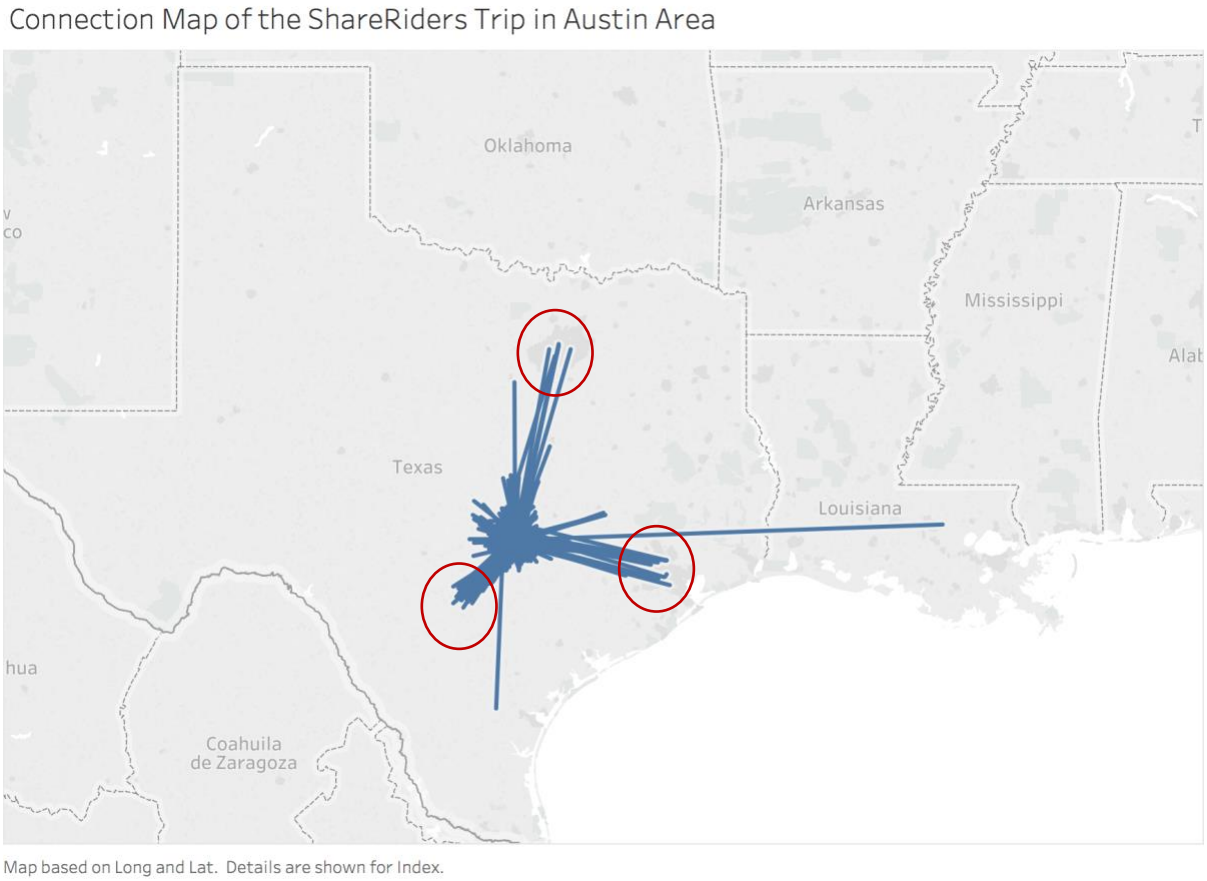
- **Chloropleth Map**



These two maps compared the driver and rider ratings by zip code areas in Austin area. We can see that there is very small change in riders rating, but relative more changes in drivers rating. This indicates that riders generally have different opinions on drivers while drivers less likely to have bad opinions on riders. For drivers' rating, we can see that the middle (downtown) area has lower average drivers' rating, this might be caused by worse traffic conditions in downtown areas.

Trip in Geographic and Time Series

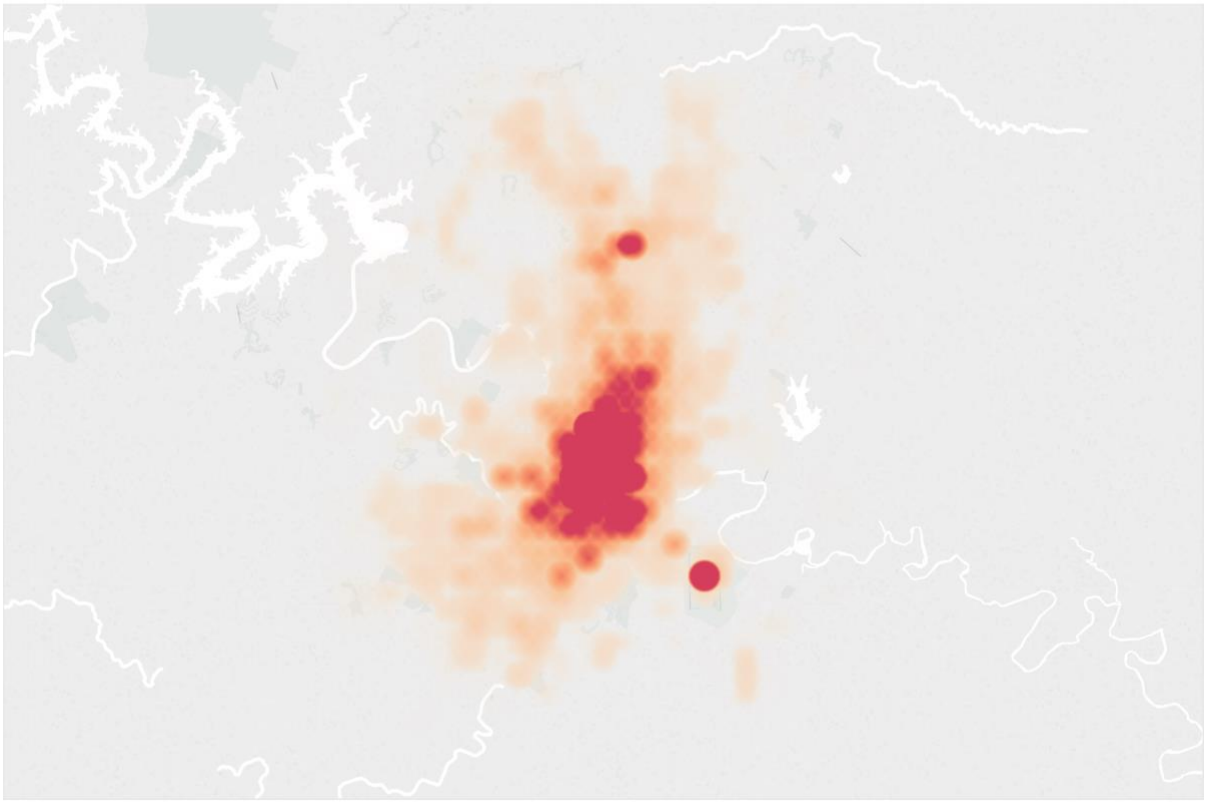
- **Connection Map**



The connection map connects the start point and end point of the trips. We can see that there are both short distance trips as well as some long-distance trips. The long-distance trips mostly concentrated on three nearby cities.

- **Heat Map**

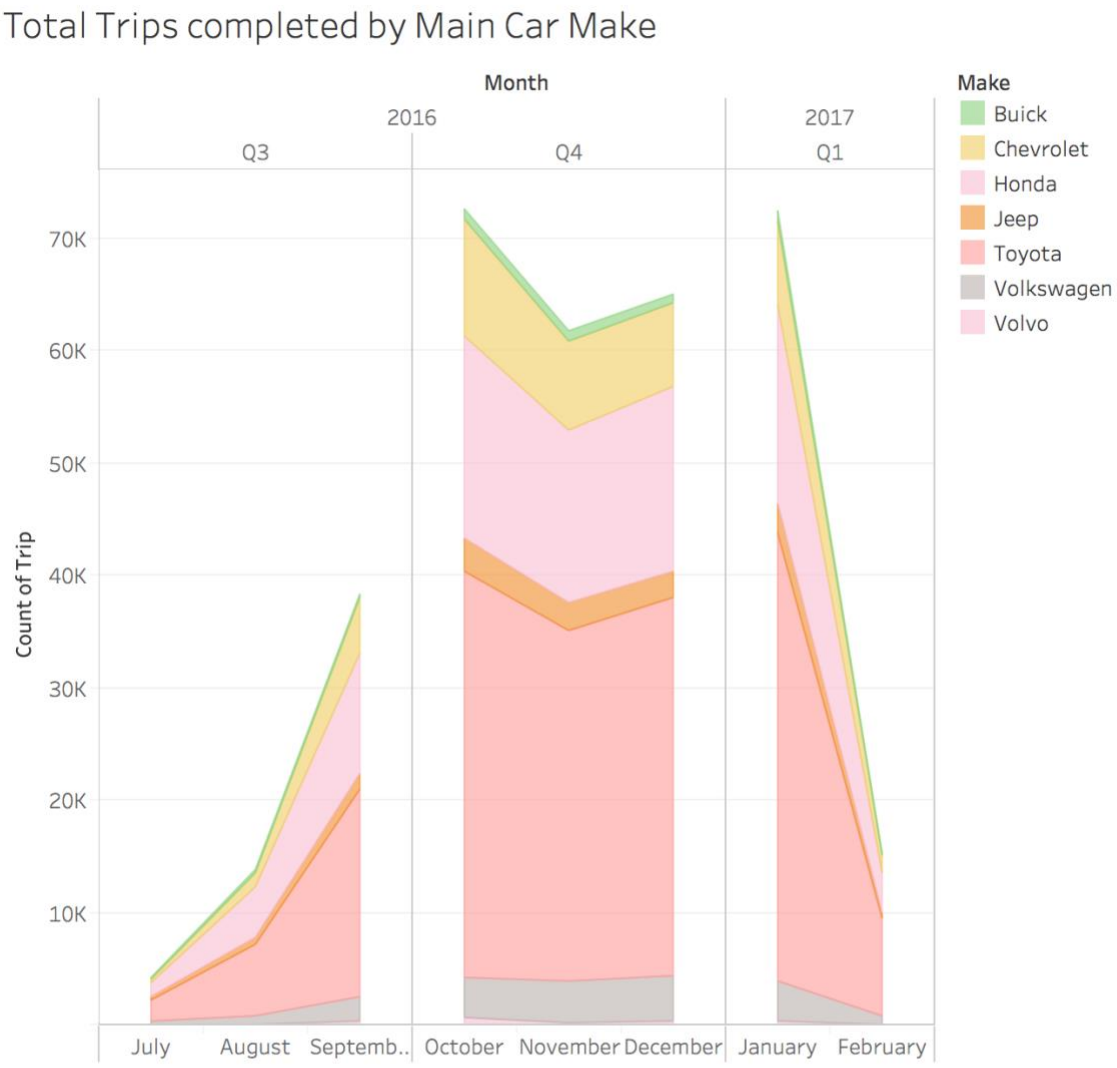
Rides Pickup



Map based on average of Start Long and average of Start Lat. Details are shown for Trip Id. The view is filtered on Trip Id, which keeps 903,316 of 908,344 members.

From this plot, we can see the density of Rides Pickup Area. The red areas mean there is more need of rides need. The orange areas represent less need of share rides in that area.

- **Stacked area graph**



Count of Trip for each Month Month broken down by Month Year and Month Quarter. Color shows details about Make. The view is filtered on Make, Exclusions (Make,MONTH(Date),QUARTER(Date),YEAR(Date)) and Month Month. The Make filter keeps 7 of 56 members. The Exclusions (Make,MONTH(Date),QUARTER(Date),YEAR(Date)) filter keeps 397 members. The Month Month filter excludes June.

From this Stacked area graph of the total trips completed by Main car make, we can see that there are more share ride needs in October, November and December. This is probably due to the holiday seasons. We can also see that Toyota and Honda are the most popular share ride car type in Austin area. The car make in this table is filtered by the total number of cars in the dataset.

Summary

This visualization series is focus on analyze the share rides in Austin Areas. Based on my analysis,

- Weather has some inverse influence on the share rides. In bad weathers, there are more needs of share rides.
- The main car makes in share ride market in the Austin area are Toyota, Honda and Nissan. Majority of the car models are Economic types. Car makes and Models has influence on Driver ratings.
- The Trips districted imbalanced in geographic and times. The downtown area has more ride needs. There are also more ride needs in holiday seasons.

Appendix and Link to the GitHub Page: Final Project Github Page