Citibike trip histories

Data Visualization

Ravi Shankar – 01647513

Data Visualization (DSC 530/CIS 602-02)

Final Project Designs

Table of Contents

Data Set:

The dataset was about the New York Citibike histories dataset, has been taken from the Citibike New York webpage, which contain the data from July – 2013 to Dec – 2013 and Jan 2014 – Feb 2014, total of 8 months' data.

Following are the detail of Data sets each column name shown below

1. Station Id: The unique id for all the stations (Quantitative)
2. Start Time and Date: at what time the ride starts, also calculated the separated the hour field from the time. (Date and time field)
3. Station Latitude and longitude: each station has the longitude and latitude (Quantitative)
4. Year of Birth: the age was calculated with the birth year (Year field)
5. Frequency: frequency is used to show the number of total rides by stationed and hours

The second related database I found is the Weather data. Which is related like in the sunny / cold weather do people prefer to ride the bikes or in snow how much population rides the bike,

Following are the detail of Data sets each column name shown below

1. Trips: trip is the quantitative field, used to show the number of total rides.
2. Min Temperature and max temperature field: is used to show the maximum and minimum temperature of the month.
3. Stations_in_service: contains the station Id (unique id, Quantitative)
4. Frequency: frequency is used to show the number of total trips.

The dataset has been taken from the following website, this dataset contains the one main CSV file.
URL: https://www.citibikenyc.com/system-data

The dataset has been taken from the following website, this dataset contains the one main CSV file.
URL: https://github.com/toddwschneider/nyc-citibike-data/blob/master/data/daily_citi_bike_trip_counts_and_weather.csv

Processing of the Dataset:

I have processed my dataset through the R language using the R-Studio. The R-queries are updated here

https://raw.githubusercontent.com/raviladhar/visualization/master/R%20quries.txt
https://raw.githubusercontent.com/raviladhar/visualization/master/R%20quries1.txt

Questions:

Though this is the interesting dataset, I am going to answer the few questions below to know some interesting information about the dataset.

1- Showing the Number of Rides by Stations and showing what parts of New York have higher or lower usage of rides. Showing the Number of Rides for different stations on various hours and determining the busiest hours?

2- Showing the trend between weather and number of rides and if it has any impact.

3- To check what age groups, ride the most and least. Also does certain hours see more of specific age group

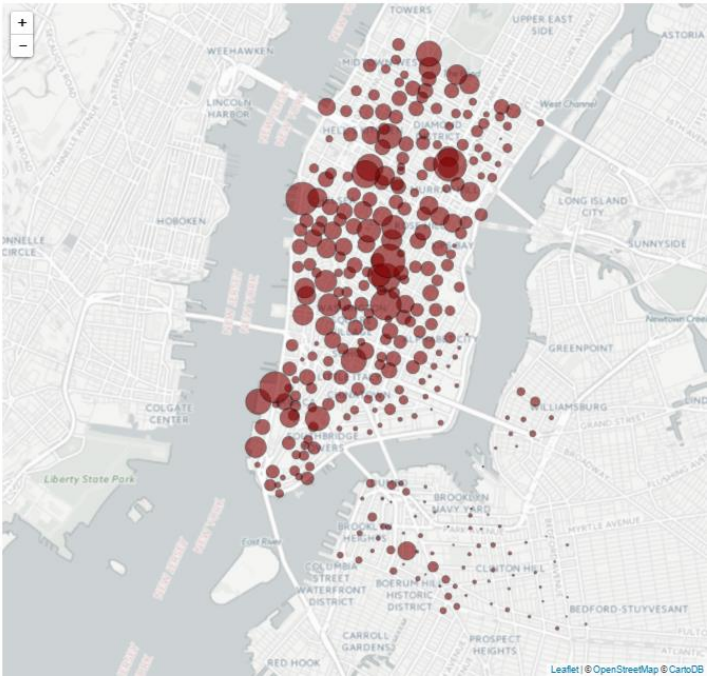
I am visualizing all these all three questions by map plot, line chart, bar chart and the scatter plot.

Visualization:

To answer the different question, we used multiple different idioms, so for Question 1:

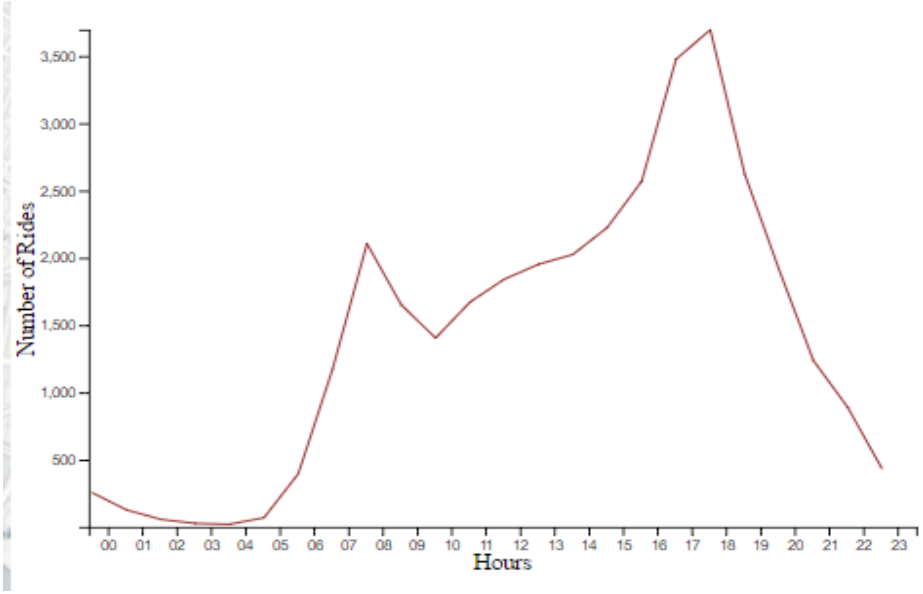
1- Showing the Number of Rides by Stations and showing what parts of New York have higher or lower usage of rides. Showing the Number of Rides for different stations on various hours and determining the busiest hours?

   To answer this question, I used the Leaflet library to show the map of the specific location for New York city, the map contains the bubbles, the bubble on the map shows the specific station and the size of the bubble shows the max number of rides for that stations.

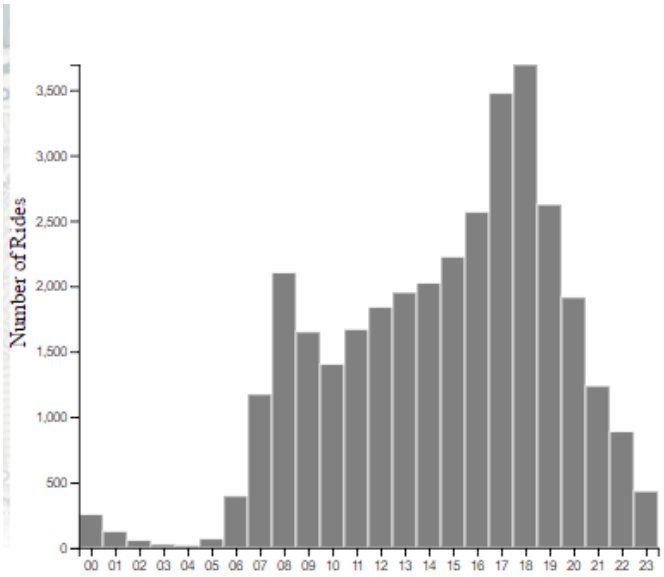


So, the picture shows the bubble chart and it shows the multiple circles or bubbles, which is representing the specific stations and the big and small size of the circles shows the maximum number of rides for that specific stations.

In detail if we will click on the specific station (circle), it will show us the other trends for the precise number of rides per hour through the line chart and the bar chart,



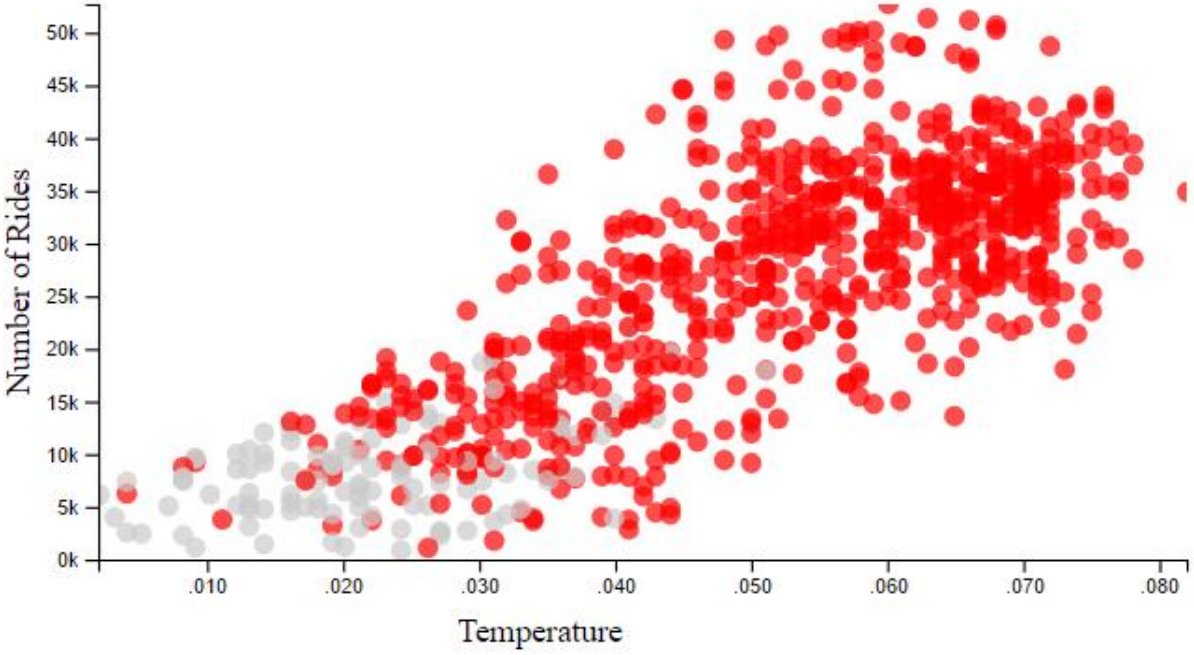
The line chart and the bar chart both are similar, but the line chart is useful here to show the trends such as there are more number of rides at the evening as compare to the morning hours, and there are very few number of rides at the early morning around 3 – 4 AM in the morning.



The bar chart is also viewing the same details as the line chart, bar chart is also helping us to show the trends, such as the number of rides are more in evening than the morning, and the other reason to show the bar chart here, it is viewing the exact number of rides per hour, which is more specific to know, at what time, there are more number of rides, and when there will be no bike rides.

2- Showing the trend between weather and number of rides and if it has any impact.

I answered this question through the scatterplot, where its showing the trend for the number of rides according to the weather condition and temperature

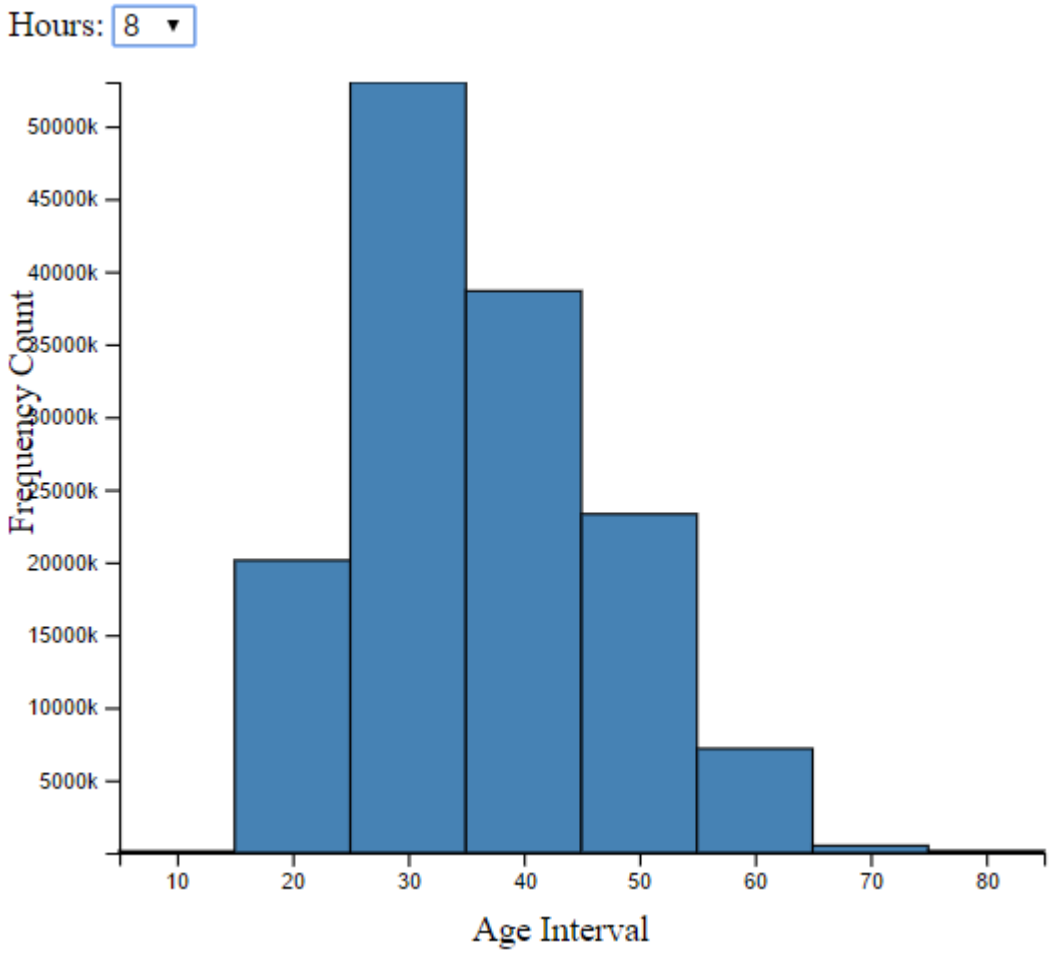


This visualization above for scatterplot is showing the trend, that when the temperature is low around 1 – 10 degree Fahrenheit the number of rides are decreasing, but as the temperature is increasing to 70 and 80, like its hotter weather, the number of rides are also increasing.

Here this visualization is indicating the two categorical colors of dots, the red dots are showing the normal temperature, and the gray color dots are showing the snowy weather, like when there is too cold and its snowy weather, the number of rides are very less.

3- Checking what age groups ride the most and least. Also does certain hours see more of a specific age group

I answered this question through the bar chart, by plotting bars based on hours. Here I have plotted the bars by creating age groups of 10, and the bar chart shows the trend of what age group of people use the bike service on particular hours of the day.



It is observed that ages from 20 -40 use the bike service the most, the bar chart is updated based on the hours which are selected from the drop-down list.

My processed datafiles and code links are available here:
https://github.com/raviladhar/visualization
http://bl.ocks.org/raviladhar/ea22e3a4b00138aafb6204c3666d31db for my GitHub code

References:

http://bl.ocks.org/weiglemc/6185069
http://stackoverflow.com/questions/23697854/line-graph-dynamic-data-updates-in-d3
http://stackoverflow.com/questions/10784018/how-can-i-remove-or-replace-svg-content
http://www.d3noob.org/2016/08/create-simple-line-graph-using-d3js-v4.html
http://bl.ocks.org/d3noob/9267535
http://stackoverflow.com/questions/40199108/d3-v4-scaleband-ticks