Project Name:
NYC Yellow Cab Data
and tipping percentage patterns

Team Member: Yingjun Mou & Xiaoyan Wang

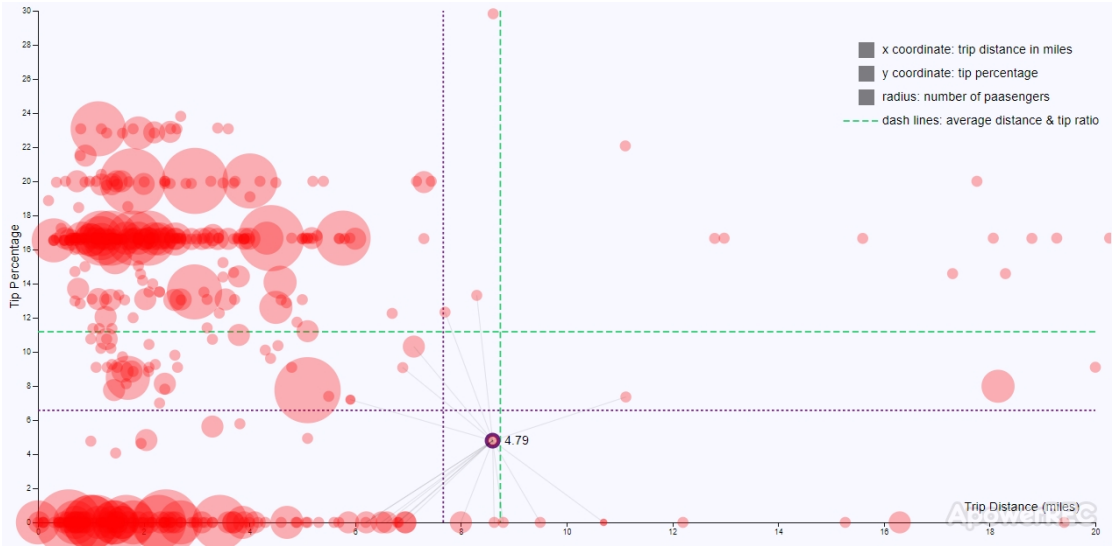
The Yellow Cab is one of the emblematic symbols of NYC. While nowadays more and more people using Ubers with a prescribed amount of tip percentages, we are curious about whether there is a behavior patterns of conventional tipping the yellow cab drivers, and whether there is any correlation between tip percentages and the trip distance or passenger number.

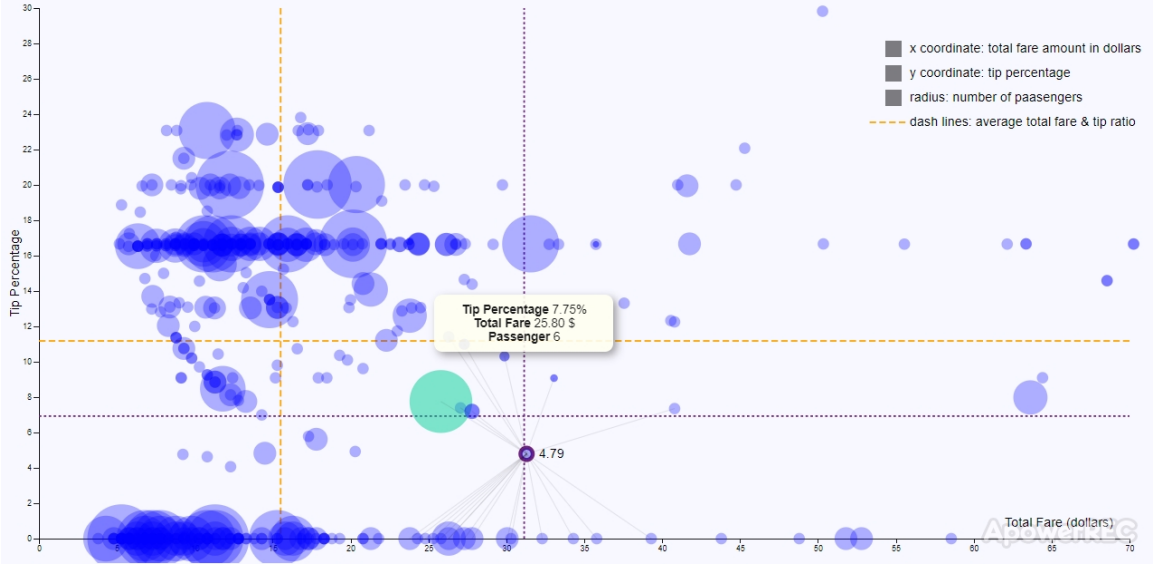
We would like to figure out the statistical correlations between 4 different factors:

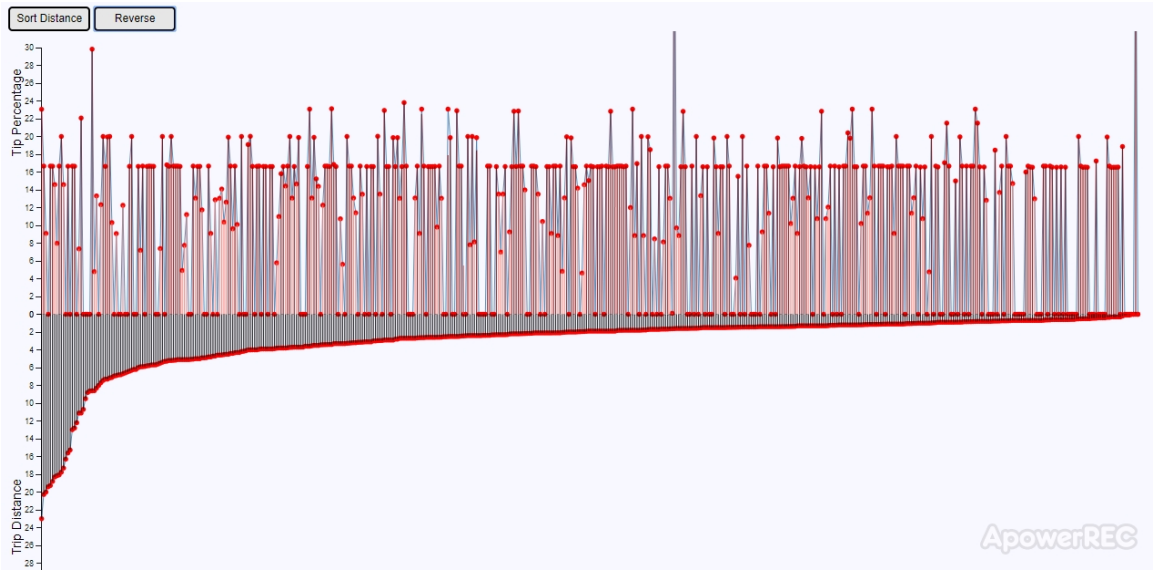
1. Tip percentage
2. Trip total distance
3. Trip total base fare
4. Number of passengers

Besides the correlations, we would also like to know the ranges and averages of these statistics. In order to do this, we got our data from NYC 311 dataset:
https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq

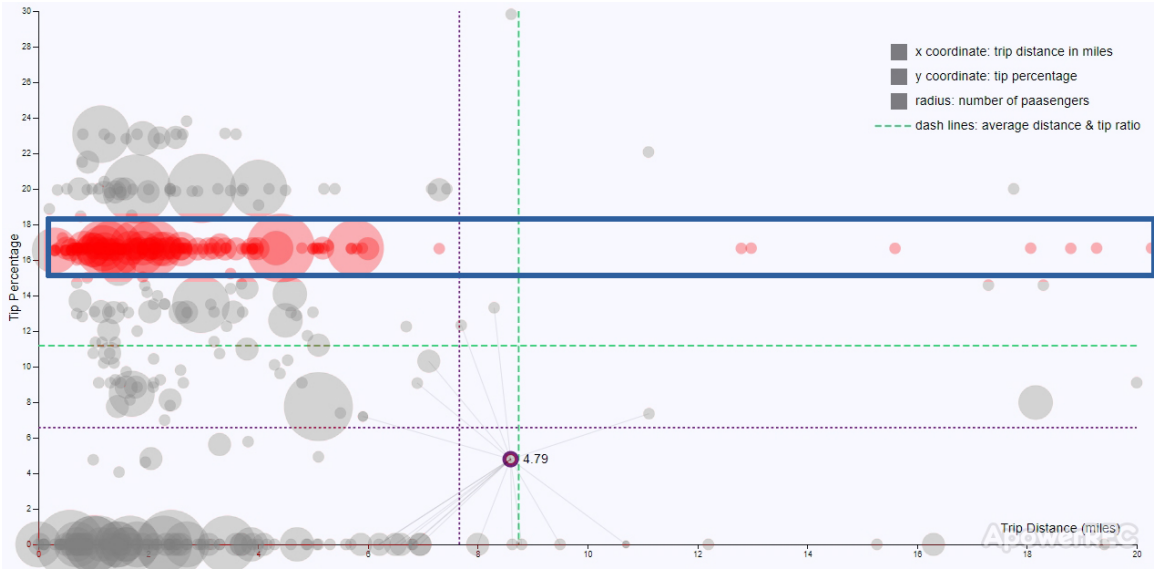
We decide to use two scatter plots to show the relationship between the tip ratio versus trip distance and tip ratio versus total fare amount. Also, we created another diverging bar chart to show the relationship between trip distance and the trip ratio to better illustrate the correlation.

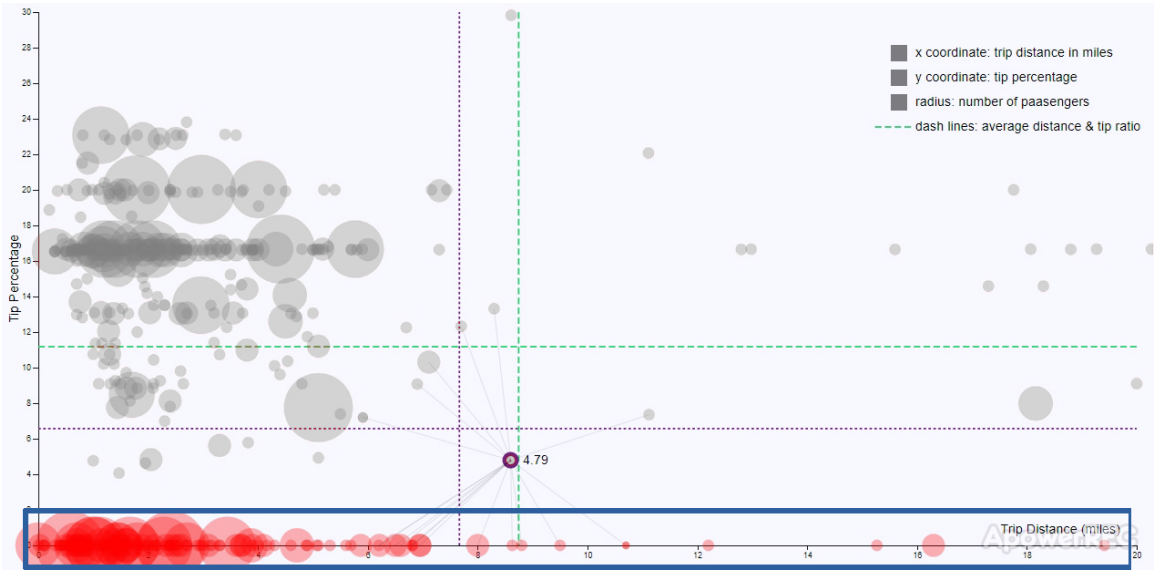


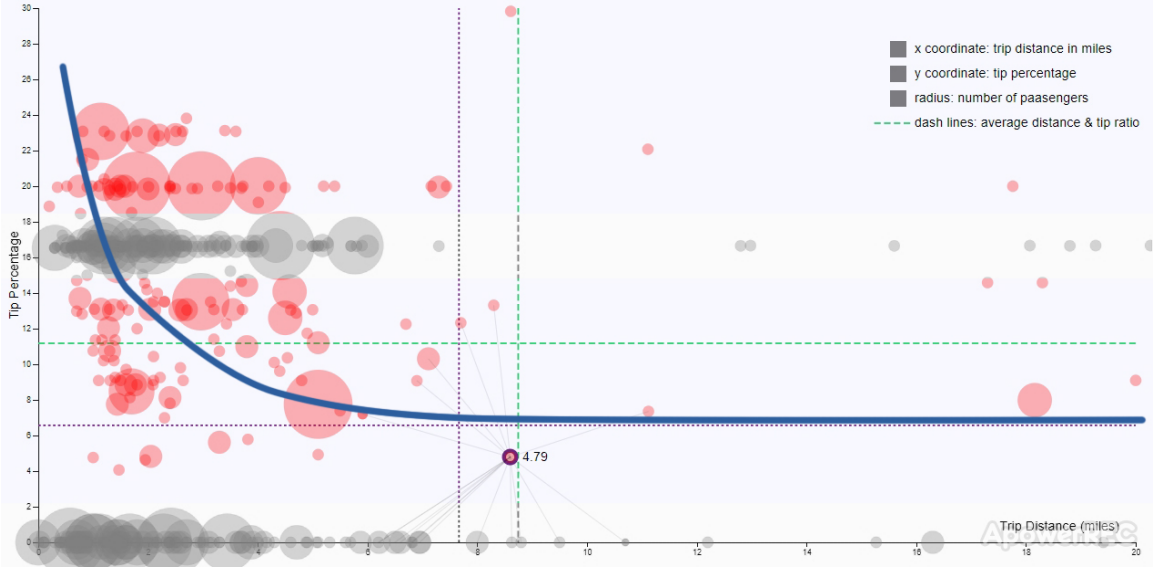




And then we did an analysis of the data based on the plots and few conclusions have been drawn:


x coordinate: trip distance in miles
y coordinate: tip percentage
radius: number of paasengers
dash lines: average distance & tip ratio
Tip Percentage
4.79
Trip Distance (miles)


x coordinate: trip distance in miles
y coordinate: tip percentage
radius: number of paasengers
dash lines: average distance & tip ratio
Tip Percentage
4.79
Trip Distance (miles)



Among the 1000 samples:

(1) the tip percentage range from 0% to 29.82%, with its average equals to 11.2%
(2) the trip distance range from 0 to 20.27 miles, with its average equals to 8.9 miles
(3) the total fare amount range from 0 to $70.27, with its average equals to $15.6
(4) from scatter plots, there are 2 different types of tipping patterns:

(a) tipping in a constant rate (most of them 16.5% or 0%) and
(b) tipping in inverse proportion to the total amount paid and the trip distance.

Regarding the findings, we gave few explanations:

Why there are different types
of tipping patterns:

It may because there are two different groups of riders. For example, people who always tip with a certain percentage (those fall into 16.5%, 20%, and 23% lines) versus some other people who cultural against the ideas of tipping (those fall into 0% line). Some foreign tourists from countries which doesn't have tipping cultures may not be aware of the customs here.

Why there is
an inverse proportional relationship:
This may be resulted from the same group of riders but in different scenarios.
When the total fare is low, for example, $10, even giving some small changes like $2 would be 20%.
On the other hand, as the fare amount increases, the absolute dollar amount of the tip will seem to meet or exceed the fair value for a gratuity. And riders will be more sensitive to the tip percentages.