DATA VISUALIZATION PROJECT

New York City Yellow Taxi Rides 2013

SARAT KARASALA

30098831

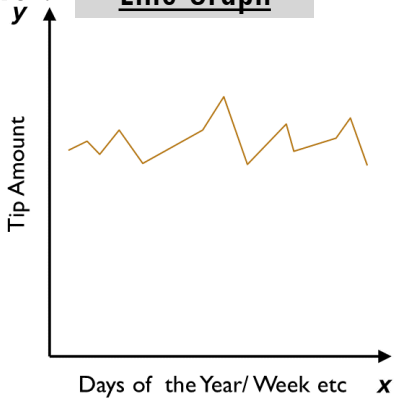
MOTIVATION

- New York City is a very densely populated city with Manhattan boasting a population density of 27,000 people per square kilometre
- This means a LOT of taxis
- Just the familiar Yellow Taxis alone cross about 165 Million rides in a year
- We can visualize the data and find out how New Yorkers tip per cab ride
- We can answer questions like 'What day of the week New Yorkers are most generous?' and a few others…

Sheet 1: Brainstorming

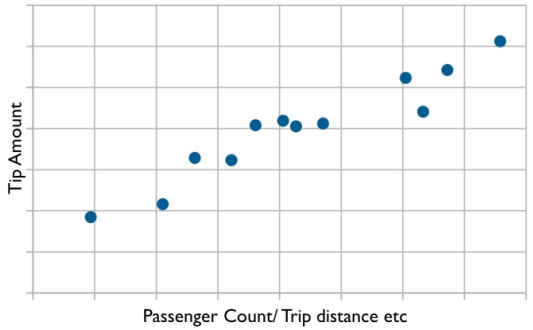
Ideas :

Line Graph



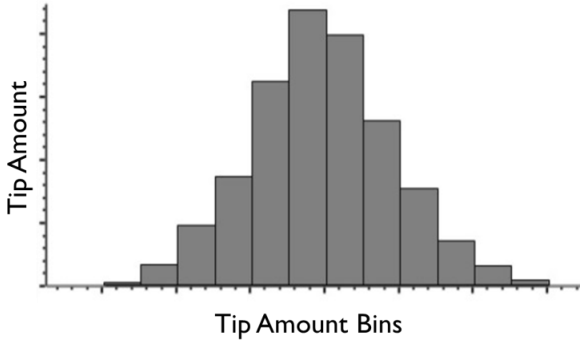
Daily average thorough the year

Scatter Plot



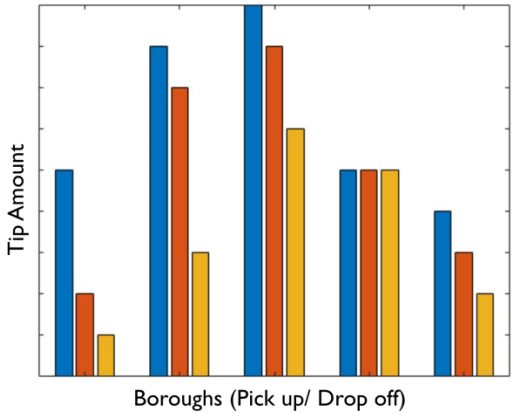
Tips vs Other parameters like trip time, passenger count etc

Histogram



Can show the distribution of tip amounts for different days, months etc

Bar Plot

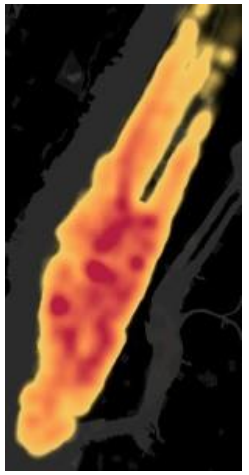


Stacked or Grouped bar charts can add another layer

Map

Plot the 5 boroughs and display statistics

Density Map



Plot areas with high cab density on a map

Filtering Ideas

A Line Graph is redundant compared to the other graphs

Density maps work best when we have the details of all suburbs. We only have the boroughs

Categorize

- Scatter plot, Bar plot and Histogram
- Map of the Boroughs

Combine and Refine

Scatter Plot and Bar plot can show tipping behaviour for different factors

A Histogram plot can be used to observe the distribution of tips in the given data

A Map of the 5 Boroughs can be used to show the basic statistics for each borough

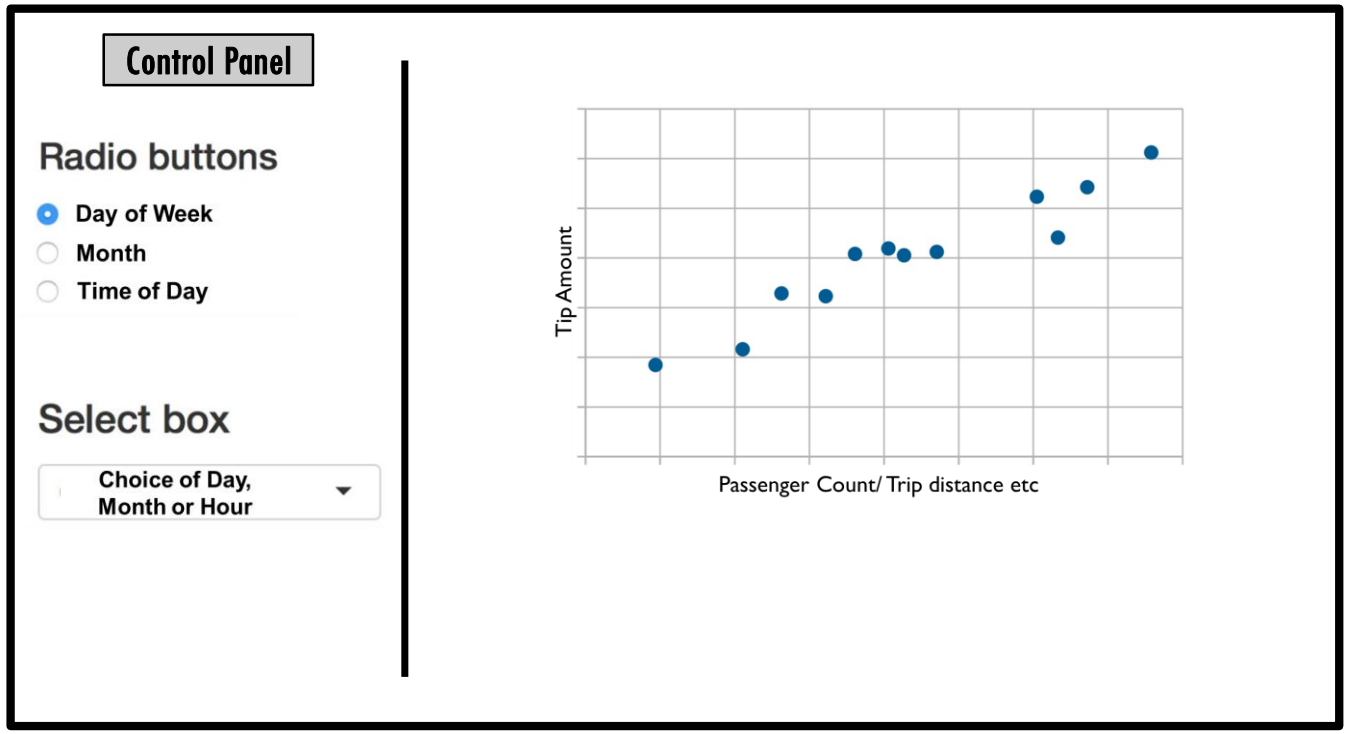
TITLE: BRAINSTORMING

AUTHOR: SARAT KARASALA

DATE: 07/06/2020

Sheet 2: Design

Layout : Scatter Plot



Operations : We can filter the data by day of week, month or time of day.

Pick one and choose the specific value in the selection box. For example I want to see the tips offered on Sundays or at 12 mid night etc

Focus

The Scatter plot will show the correlation between the Tip offered for a ride and any parameter (passenger count/ trip distance)

Discussion

Pros : Scatter plots are are very effective when there is a strong correlation between variables

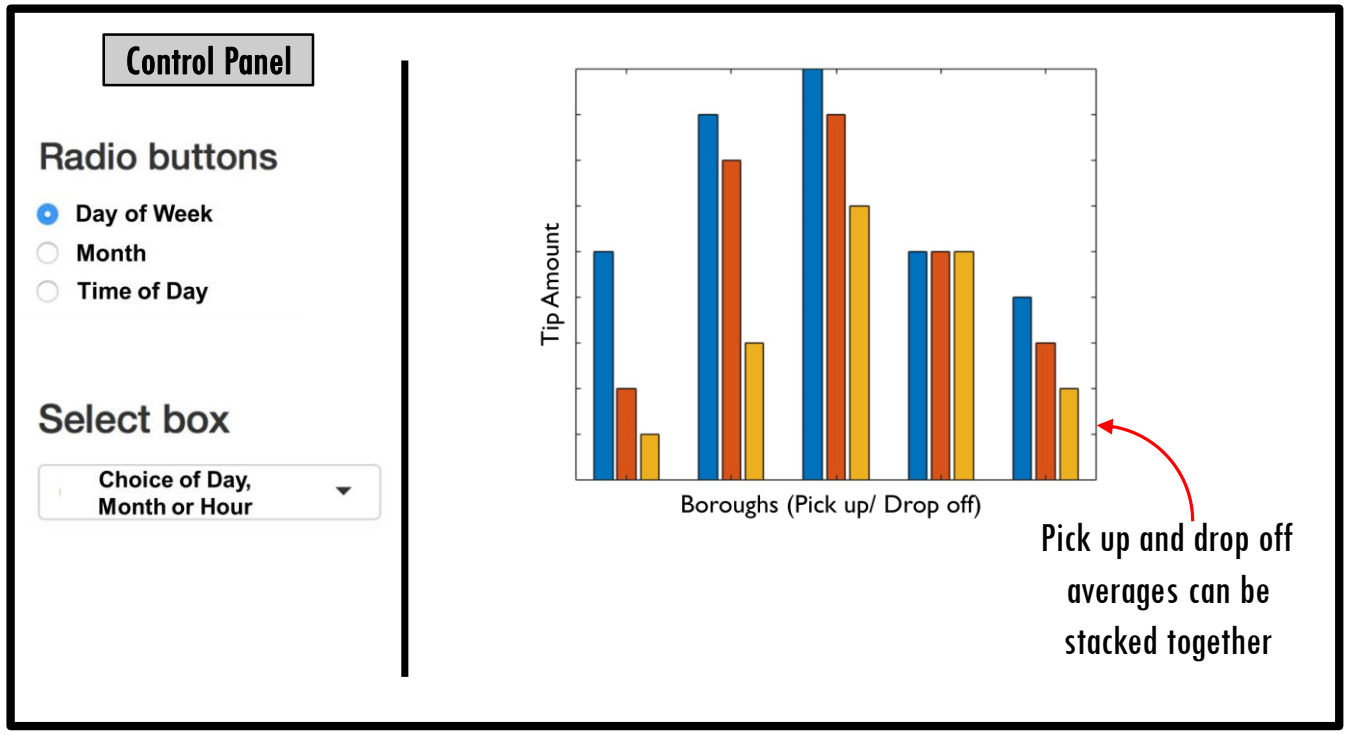
Cons : Time taking – 450,000 lines of data

Due to the large amount of data the graph is completely filled

TITLE: DESIGN SHEET 2
AUTHOR: SARAT KARASALA
DATE: 07/06/2020

Sheet 3: Design

Layout : Bar Plot



Operations : Filter the data to see the average tip offered in each borough for a given day of week, month or time of day

Pick one from the radio button list and choose the specific value in the selection box below. For example I want to see the tips offered on in December.

Focus

Aggregate methods: The Bar plot will show the "average" tip offered in each borough for the given selection of filters

Discussion

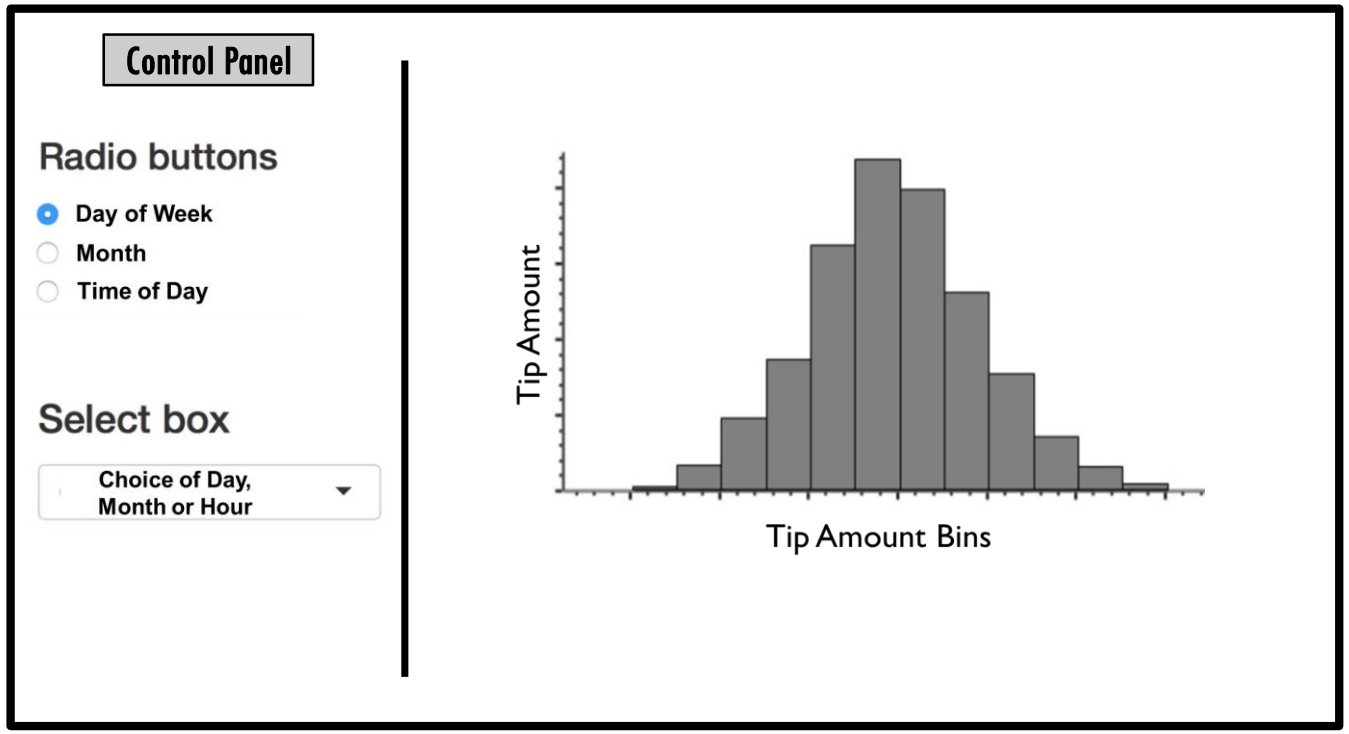
Pros : Using aggregation functions makes the visualization more meaningful and doesn't feel cluttered due to too much data

Cons : Bar Plot can't handle continuous data so we cannot use trip time and trip distance without splitting them into bins

TITLE: DESIGN SHEET 3
AUTHOR: SARAT KARASALA
DATE: 07/06/2020

Sheet 4: Design

Layout : Histogram



Operations : Filter the data to see the distribution of the tip amounts offered for a given day of week, month or time of day

Pick one from the radio button list and choose the specific value in the selection box below. For example I want to see how the tips are distributed in December.

Focus

When the bar plot and histogram are used together we get a lot of information

The bar chart shows us how people from different boroughs tip and the histogram shows us the distribution of different tip amounts

Discussion

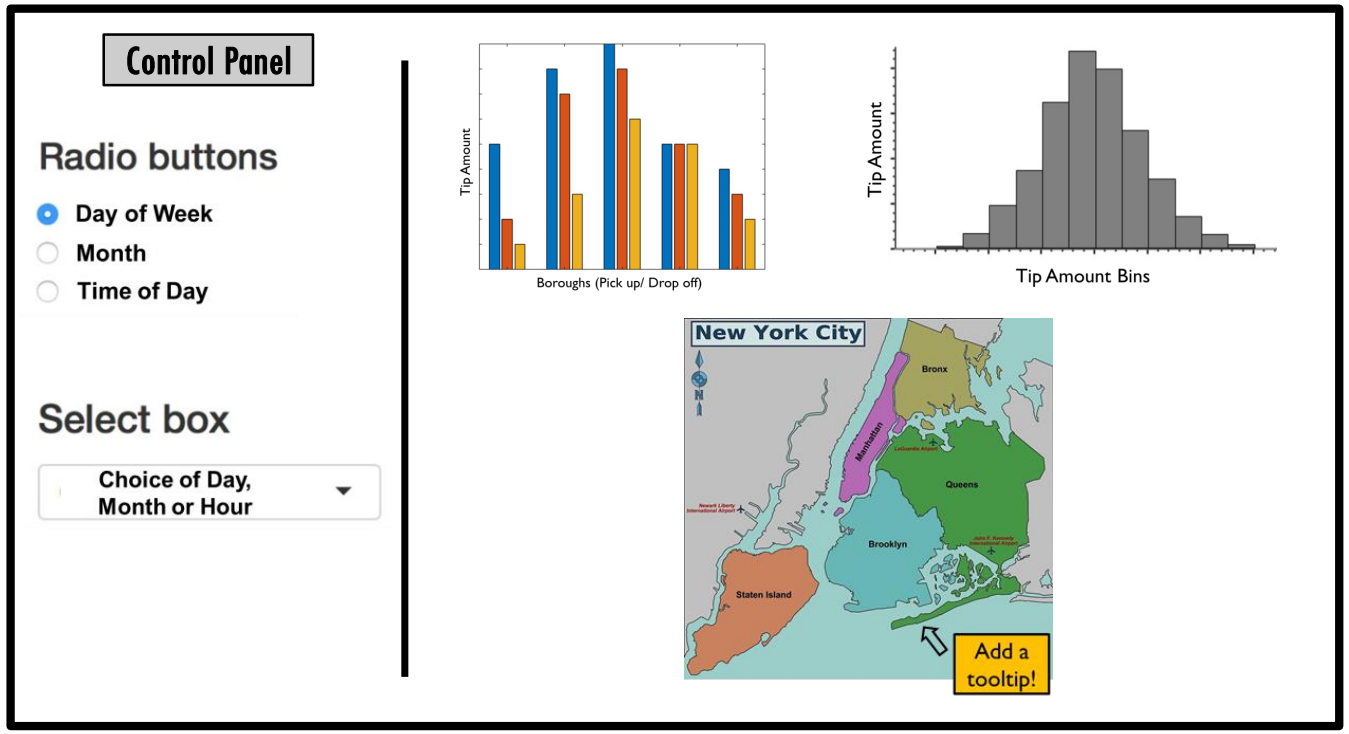
Pros : Shows the distribution of tips. Adds another layer of information when used with a bar plot

Cons : By itself there isn't a lot of information here.

TITLE: DESIGN SHEET 4
AUTHOR: SARAT KARASALA
DATE: 07/06/2020

Sheet 5: Realization

Layout :



Operations :

We can filter the data to change the visualizations based on a given day of week, month or time of day

The selection box lets us pick a specific value of the parameter chosen from the radio button list.

Hovering over different boroughs in the map will display a tooltip with statistics for each borough

Focus

The bar plot and histogram give a good idea of how New Yorkers tip.

The map has a tool tip that can give us some statistics for each borough
Example : Average passenger count, trip distance, trip time, average tip etc

Details

Platform & Algorithm :
R Shiny, Widgets, ggplot2, leaflet
Tracking mouse hover on graphs
Aggregation functions on R Dataframes

Data:
New York City Yellow Taxi Trips 2013
NYC Borough Boundary Shape Files

TITLE: REALIZTION
AUTHOR: SARAT KARASALA
DATE: 07/06/2020

THANK YOU