

Intro:

The dataset I used for this analysis was from Kaggle (<https://www.kaggle.com/brllrb/uber-and-lyft-dataset-boston-ma>), which was collected from 11-26-18 to 12-18-18 in Boston, MA. The data was gathered from various entities including Uber and Lyft.

Analysis:

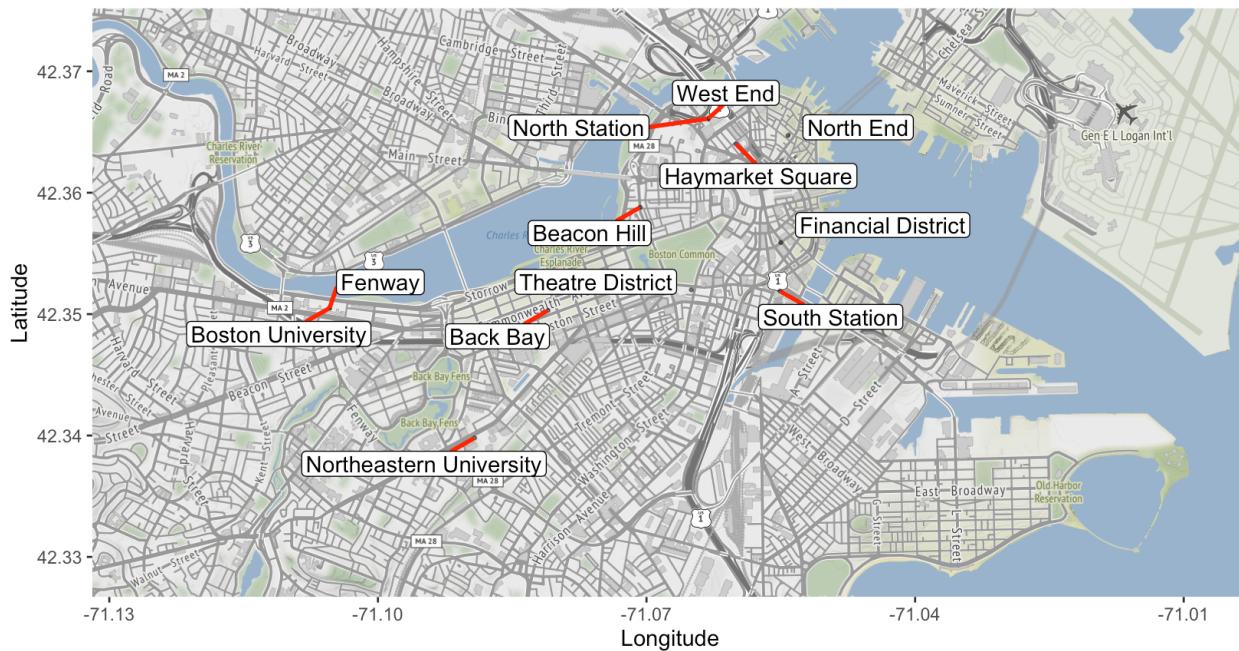
Upon studying this dataset, I found that of the 637,976 rows of data, the split between Uber and Lyft rides was roughly 50/50. When coding, I searched for complete cases in order to avoid any NA values in the dataframe. I registered my own API with google in order to assist with making any desired maps.

I. Overview of Boston Neighborhoods:

There are twelve neighborhoods included in this analysis. Below, I've attached a labeled map of the neighborhoods, a density diagram, and a graph examining the average cab price per neighborhood route.

A.

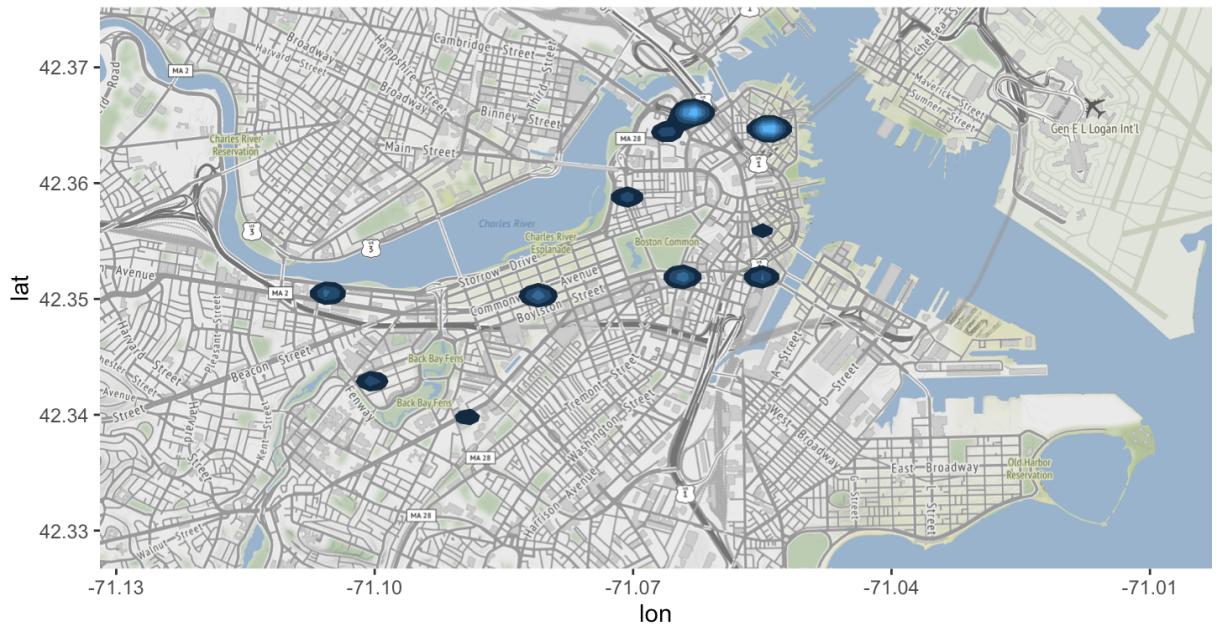
Labeled Map of Neighborhoods in Boston



I used both ggmap and get_stamenmap to create the labeled map featured above. I used latitude and longitude to clearly indicate where each neighborhood is located.

B.

Density Map of Cab Rides in Boston Neighborhoods

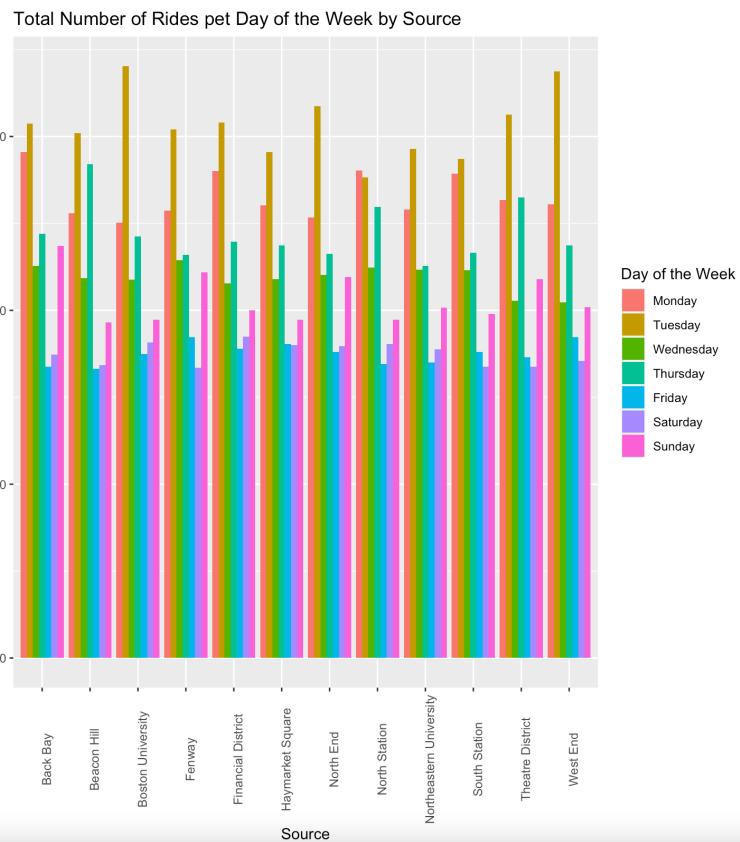


This map outlines the density of total rides in each of the neighborhoods. The lighter blue indicates a larger number of rides.

II. Rides per Day of the Week and per Hour:

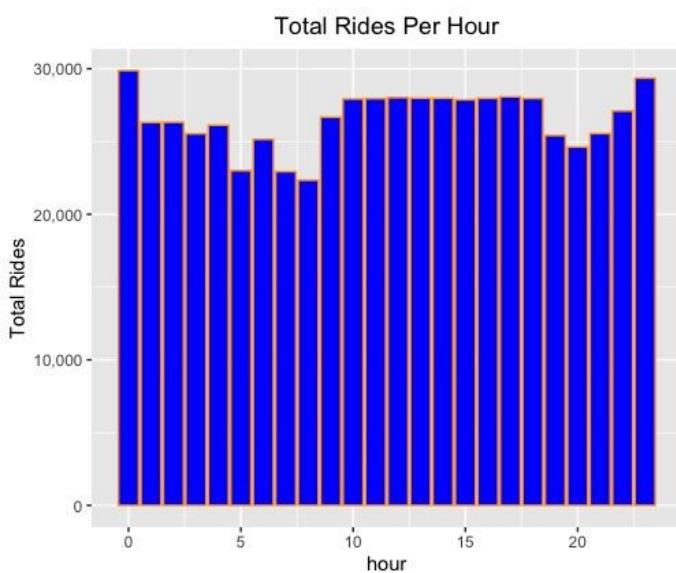
This section first examines the distribution of total number of cab rides per source and then per hour.

A.



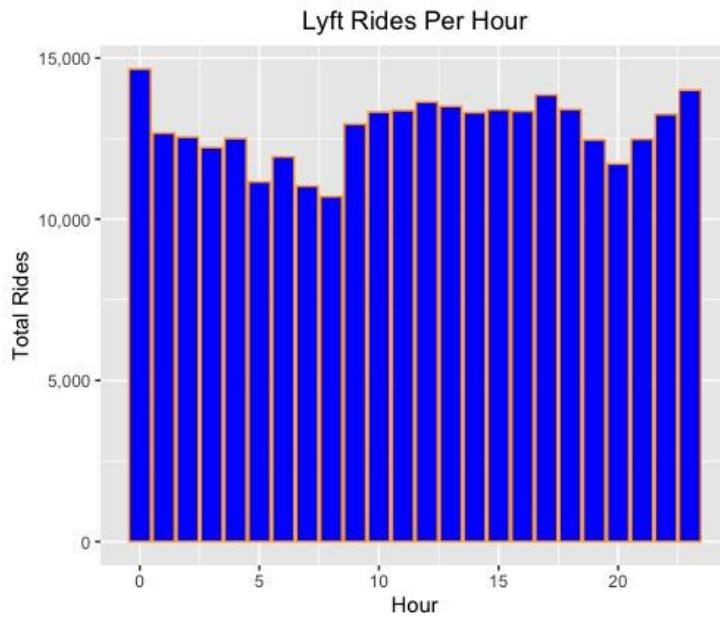
This graph clearly shows that regardless of source, the largest number of cab rides took place on Monday and Tuesday. Specifically, Tuesday tended to have even more cab rides than Monday.

B.



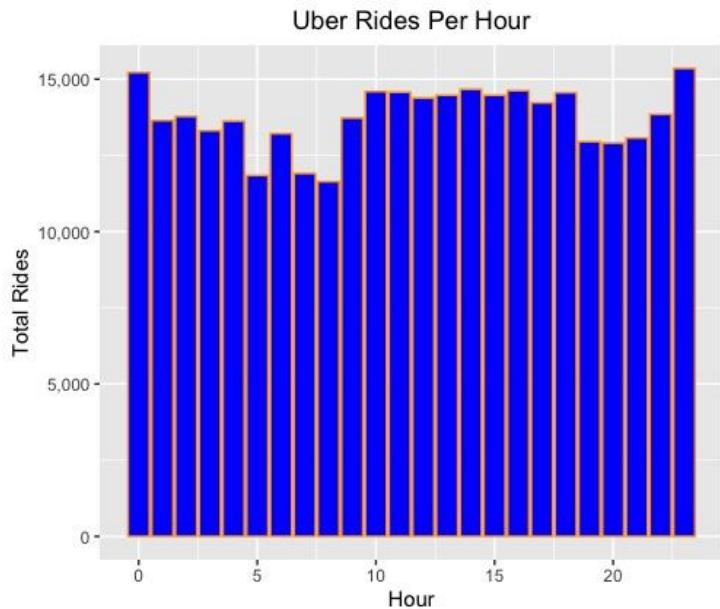
This graph shows that the majority of rides occurred between 11PM and 12AM.
The second highest number of rides occurred between 5PM and 6PM.

C.



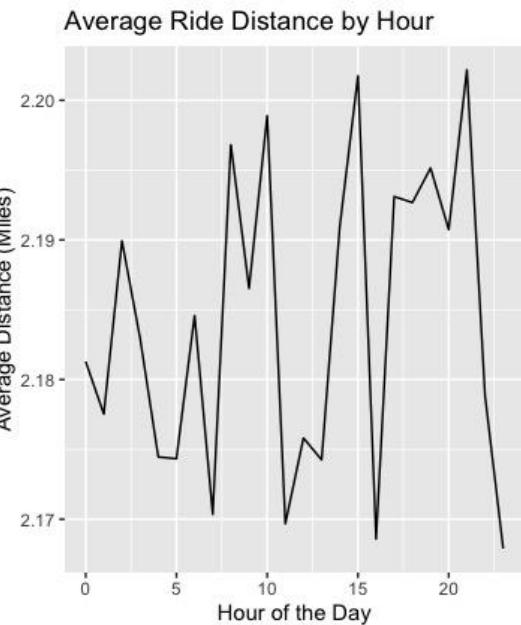
Similarly, this graph shows that the majority of rides occurred between 11PM and 12AM. The second highest number of rides occurred at 5PM.

D.

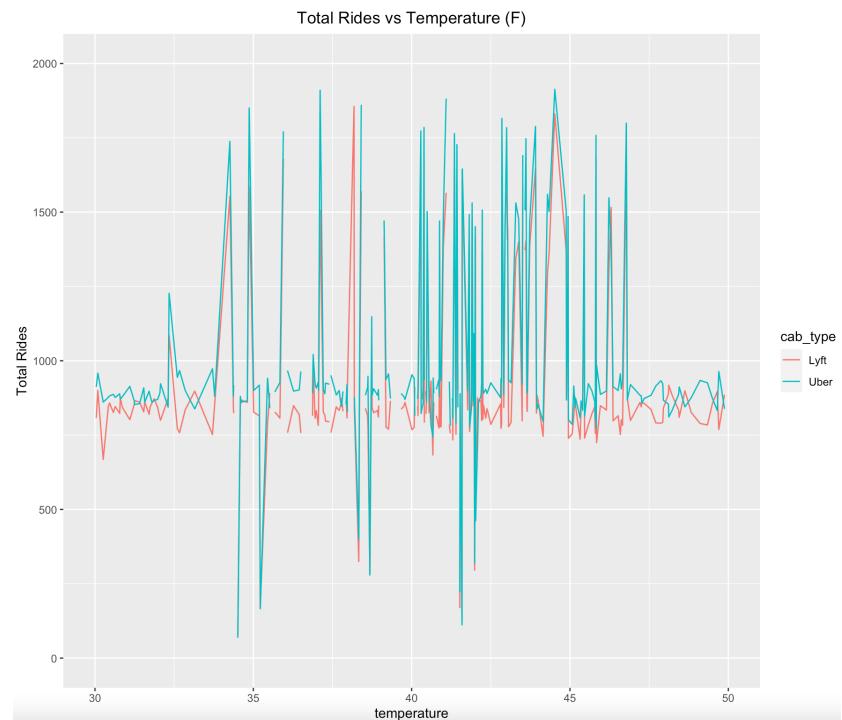


This graph shows the majority of rides occurred between 11PM and 12AM.

E.



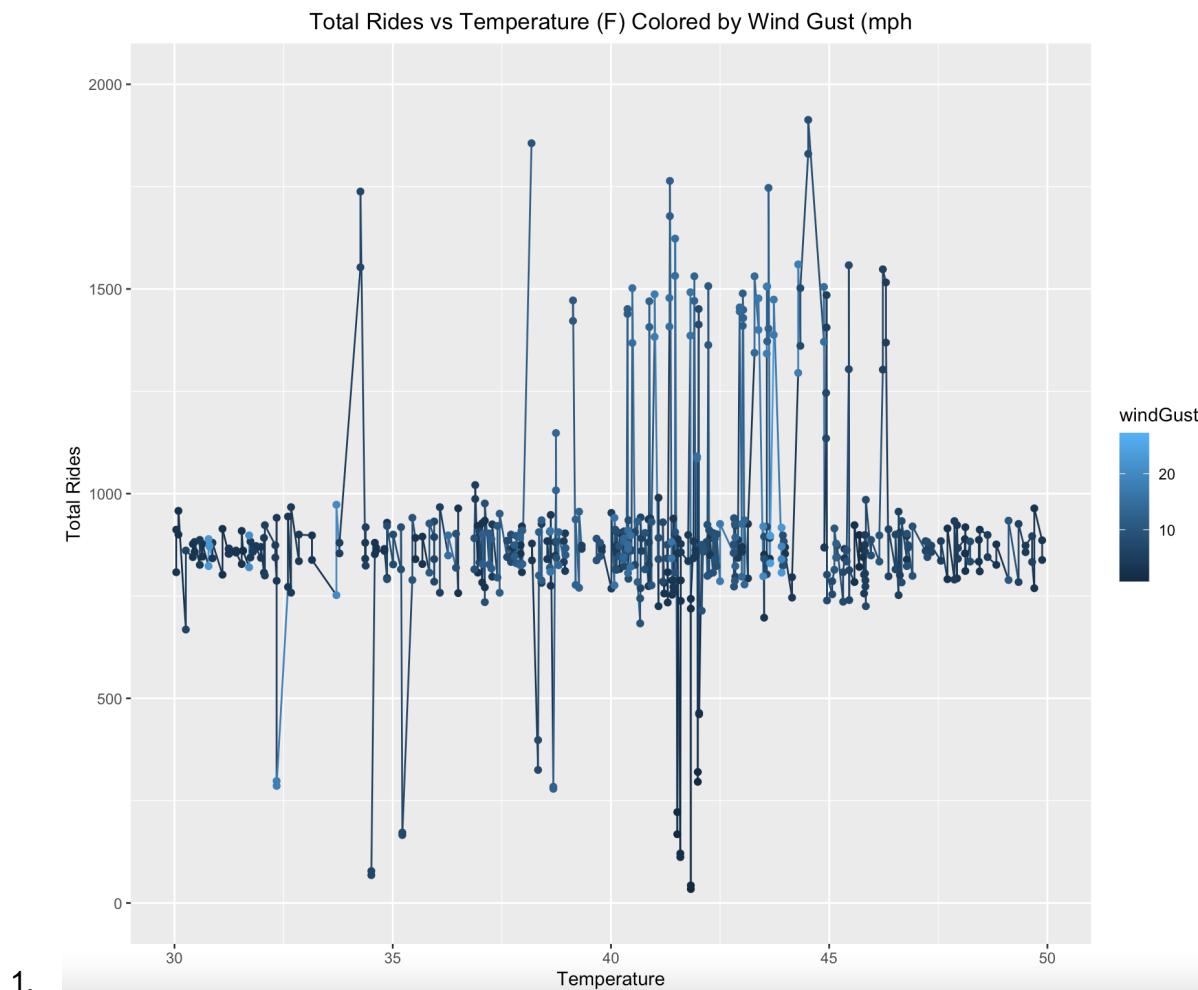
This graph shows that on average, the longest ride distance took place at roughly 3PM and 9PM. The next longest ride distance intervals were at 7:30AM and 10AM. The points all make sense because the morning data correlates to the early commute and the evening data correlates to the latter commute.



F.

I narrowed the temperature range of this graph in order to more accurately show the spikes in the number of cab rides. Regardless of cab type (Uber or Lyft), there were similar major surges and drops in the number of cab rides right

around 35, 38, and 42 degrees. There were surges in the number of both Uber and Lyft rides ordered between 43 and 47 degrees Fahrenheit.

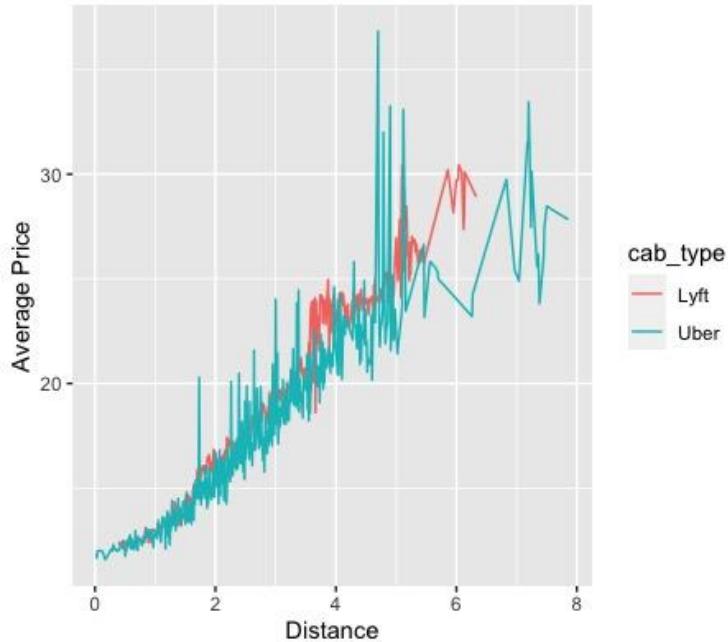


1.

To explore the temperature and total rides relationship further, I made another plot looking again at temperature vs rides, but colored by wind gust speed (mph). The higher speeds correlate to a lighter blue color and these surges in wind gust speeds fall between 42 and 45 degrees Fahrenheit, which was when there was an increase in both Uber and Lyft cab rides in the previous plot. In other words, this means that there is a positive correlation between an increase in rides and higher wind gusts although it is small (0.246).

G.

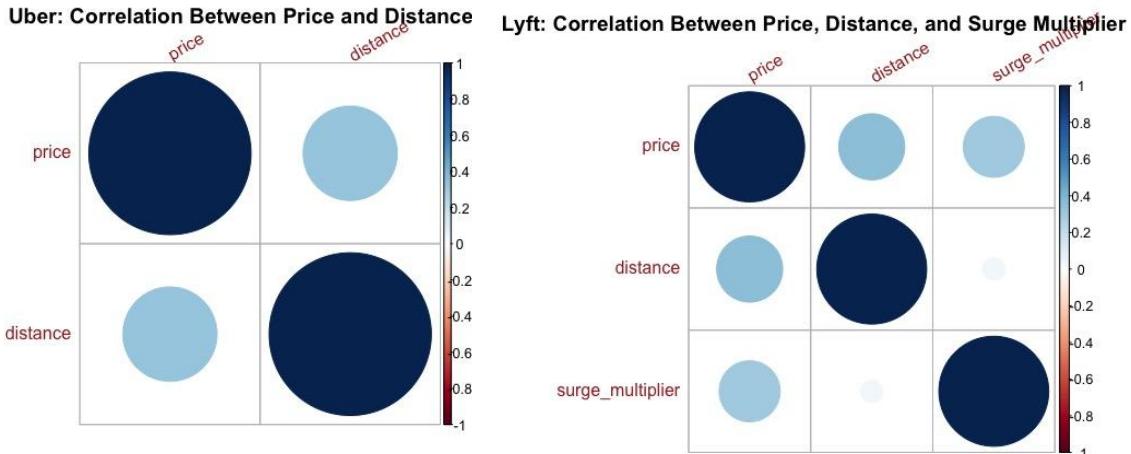
Average Price by Distance for both Lyft and Uber



This graph shows that regardless of distance traveled, riders prefer to use Uber over Lyft.

III. Lyft/Uber Prices (Correlation):

Before diving into the analysis, I wanted to determine if there was an obvious correlation between distance and price for both types of cabs. I noticed that the surge multiplier only applied to Lyfts, so I made two graphs. The first (Uber) was a correlation plot between distance and price and the second (Lyft) was a correlation plot between distance, price, and surge multiplier.



In addition to creating these correlation plots, I also found the direct correlation data for each cab type. Looking at the Uber graph, price and distance are weakly correlated with

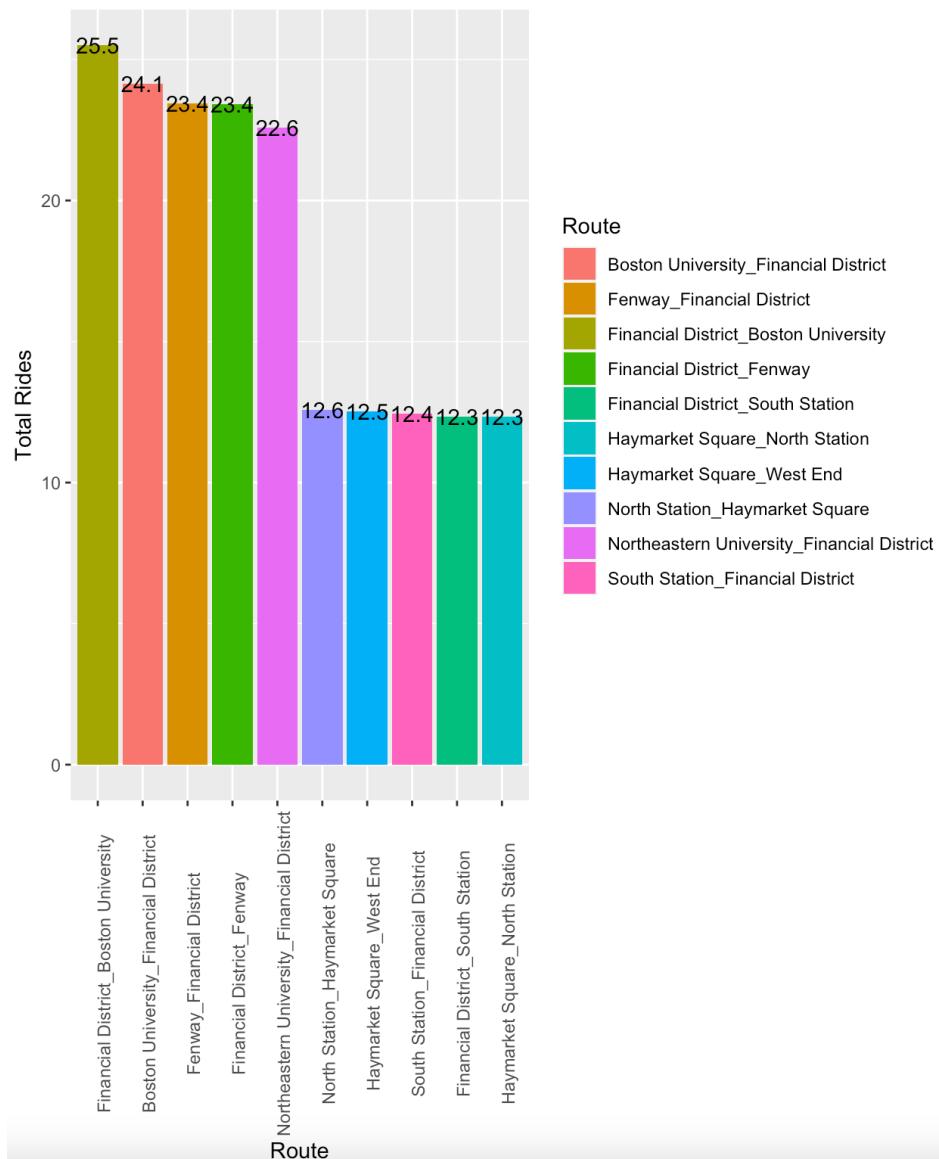
an exact correlation value of 0.3362. The Lyft graph shows that price and distance too aren't strongly correlated (correlation value of 0.3616) and similarly surge_multiplier and price are weakly correlated (correlation value of 0.3082) and surge_multiplier and distance are even more weakly correlated (correlation value of 0.04008).

IV. Relationship Between Route and Number of Cab Rides:

I was interested in seeing if there was a correlation between a cab route and the total number of rides. This section includes all available ride data for both Uber and Lyft. With respect to Lyft, this means that I've included both surge and non-surge data.

A.

Five Priciest and Five Least Priciest Ride Price vs Cab Route

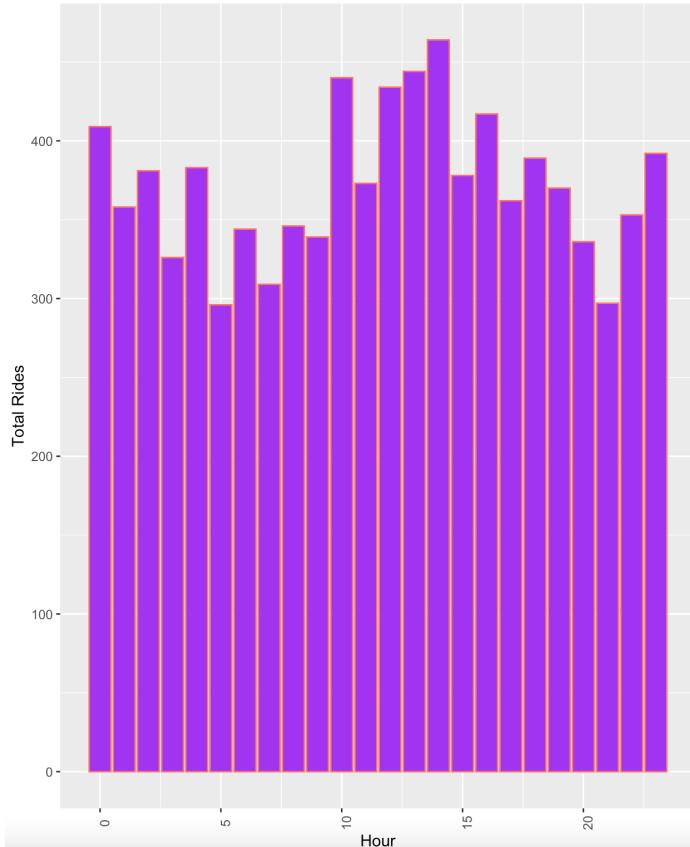


Above is a graph of the top ten mean cab rides in the aforementioned Boston neighborhoods. This plot shows that the Financial District to Boston University route has the highest average cab ride price. Inversely, the Haymarket Square to North Station route has the lowest average cab ride price. This is reasonable as

the distance between the Financial District and Boston University is 4 miles whereas the distance between Haymarket Square and North Station is 0.5 miles. However, this translates to the most expensive ride equating to \$6.38 per mile and the least expensive ride being a whopping \$24.60 per mile.

B. Priciest Route - Financial District to Boston University:

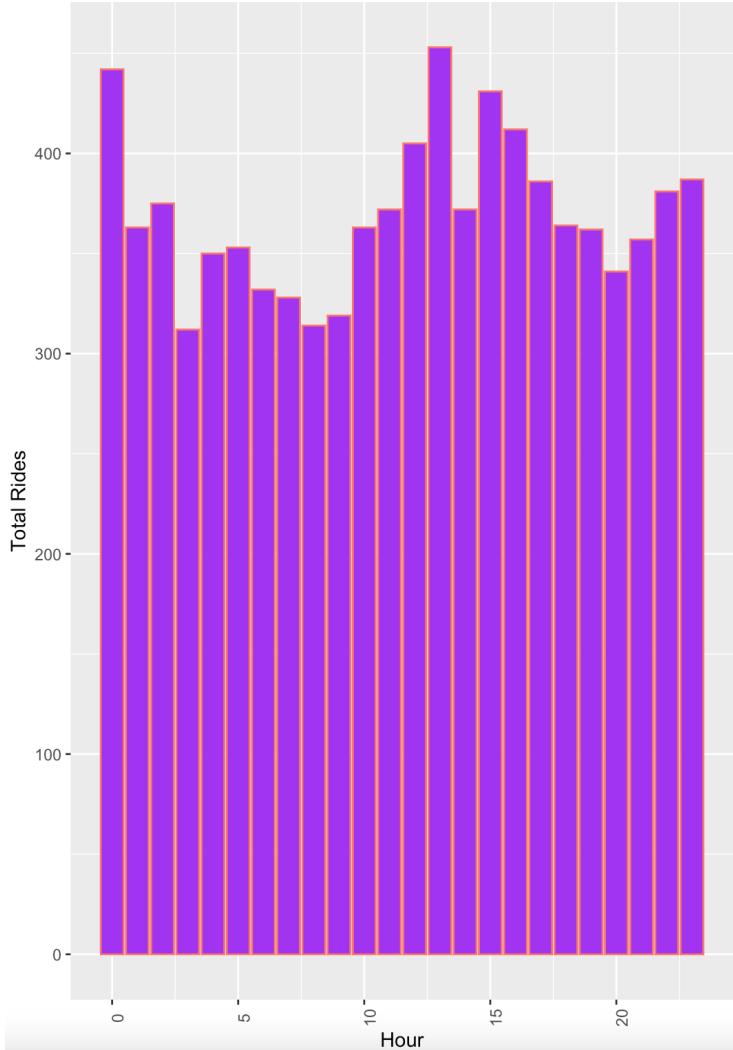
Financial District to Boston University - Number of Rides Per Hour



The previous plot (which included the five most expensive and least expensive routes on average) showed that the priciest route was from the Financial District to Boston University. Here, I filtered for this route looking at the total number of rides per hour. The plot shows that the highest number of rides (464) occurred during the 2 o'clock hour (PM).

C. Cheapest Route - Haymarket Square to North Station:

Haymarket Sqare to North Station - Number of Rides Per Hour

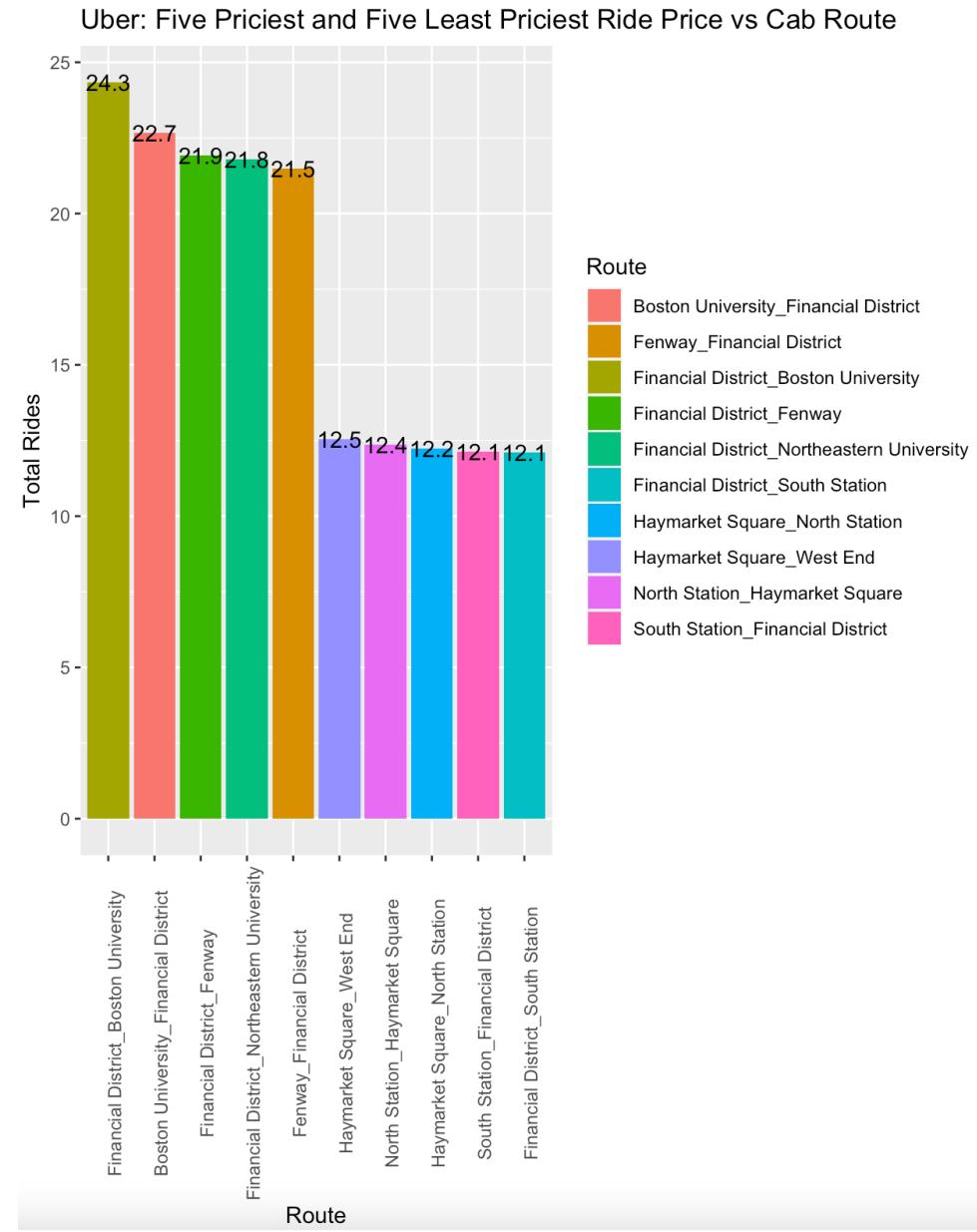


The aforementioned price plot (which included the five most expensive and least expensive routes on average) showed that the least expensive route was from the Haymarket Square to North Shore. Here, I filtered for this route looking at the total number of rides per hour. The plot shows that the highest number of rides (453) occurred during the 1 o'clock hour (PM). Interestingly enough, the highest number of rides took place during the same time of the day for both the most expensive and least expensive route.

V. Relationship Between Route and Number of Cab Rides:

In this section, I separated the data in order to look at the different cab types (Uber, Lyft) individually. Specifically, I focused on all Uber data and the Lyft data where no surge (`surge_multiplier = 1.00`) was present.

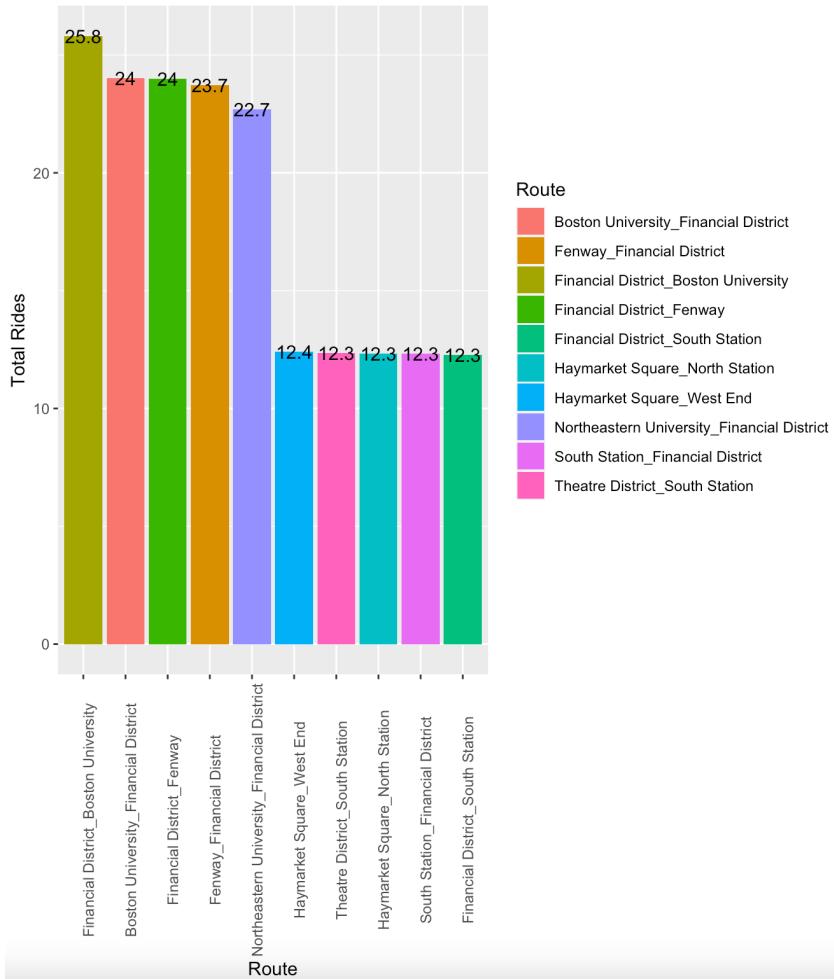
A.



This graph shows that the most expensive Uber average cab route was from the Financial District to Boston University (4 miles) and the least expensive route was from the Financial District to South Station (0.6 miles). This translates to the most expensive route being \$6.08 per mile and the least expensive route being \$20.17 per mile.

B.

Lyft (No Surge): Five Priciest and Five Least Priciest Ride Price vs Cab Route



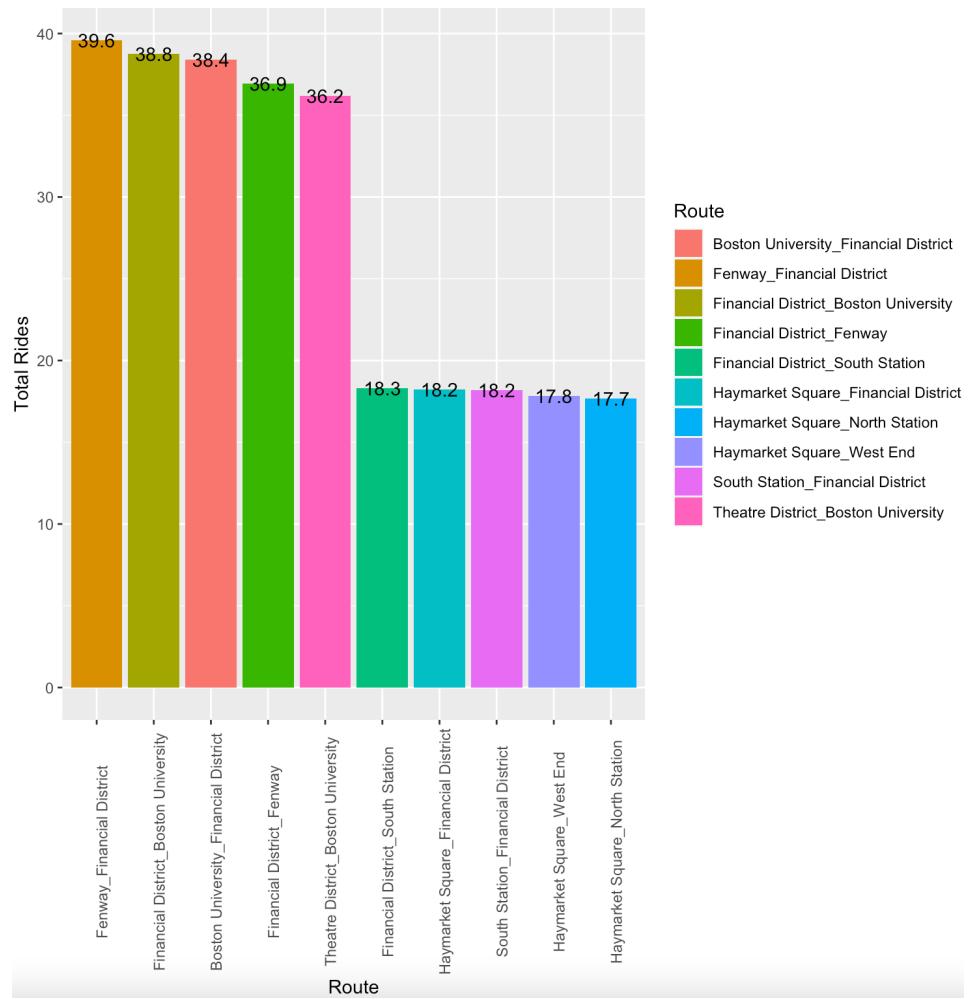
This graph shows that the most expensive Lyft average cab route was from the Financial District to Boston University (4 miles) and the least expensive route was from the Financial District to South Station (0.6 miles). This translates to the most expensive route being \$6.45 per mile and the least expensive route being \$20.50 per mile.

VI. Lyft: Relationship Between Surge Multiplier, Price, and Total Rides

The surge multiplier only applied to Lyft rides, so Uber data wasn't considered in this analysis. In this section, I separated the surge multiplier into two groups (1.25 - 1.75) and (2.00 - 3.00) in order to examine the most expensive cab route on average with respect to the two surge groups.

A.

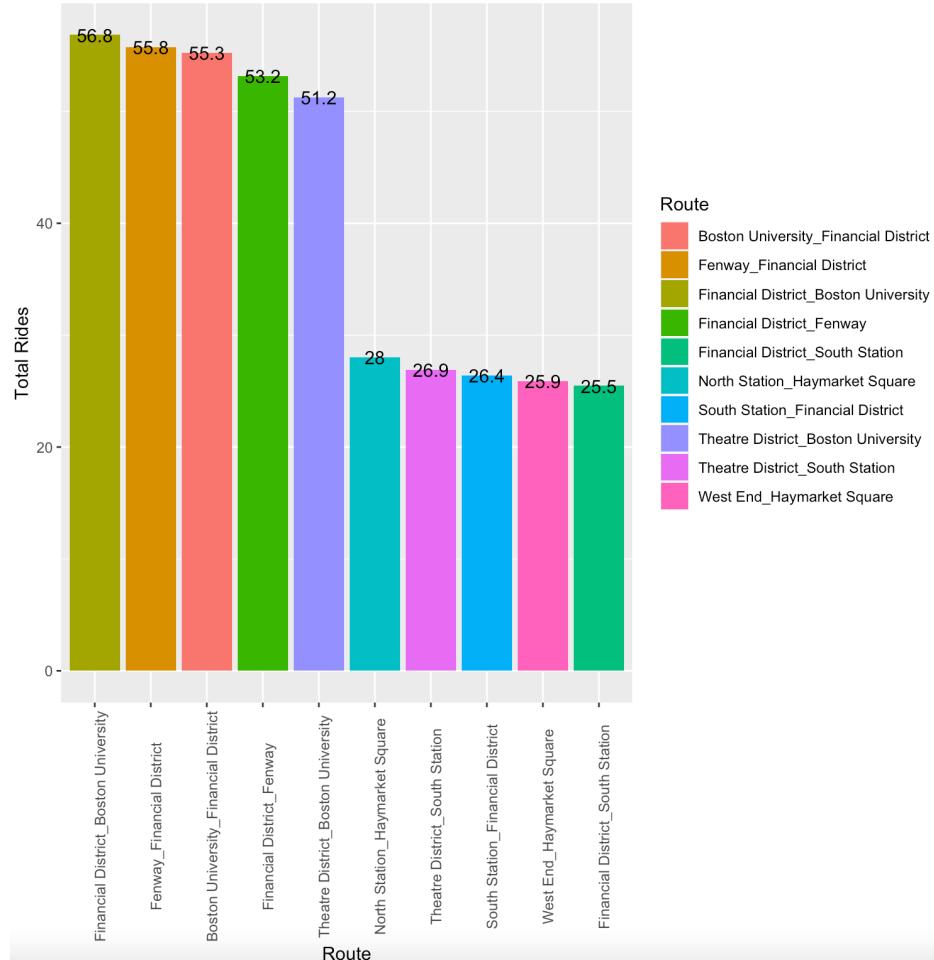
1.25-1.75 Surge Data: Five Priciest and Five Least Priciest Ride Price vs Cab Route



This graph shows that the most expensive route on average was from Fenway Park to the Financial District (4.4 miles) and the least expensive route was from Haymarket Square to North Station (0.5 miles). This translates to the most expensive route being \$9.00 per hour and the least expensive being \$35.40 per hour. The average surge multiplier in this group is 1.5, which makes sense as the most expensive route is roughly 1.5 times the ride price from the no surge data. However, the least expensive route is actually 1.75 times the ride price from the no surge data.

B.

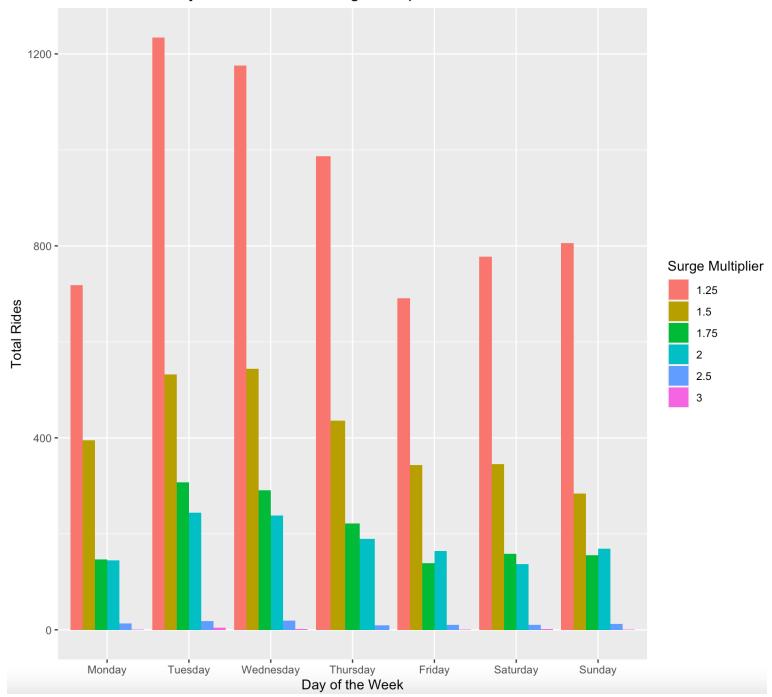
2.00-3.00 Surge: Five Priciest and Five Least Priciest Ride Price vs Cab Route



This graph shows that the most expensive route on average was from the Financial District to Boston University (4 miles) and the least expensive route was from the Financial District to South Station (0.6 miles). This translates to the most expensive route being \$14.20 per hour and the least expensive being \$42.50 per hour. The average surge multiplier in this group is 2.5, which means that the most expensive route is actually almost \$2.00 cheaper than anticipated. Similarly, the least expensive route is \$10 cheaper than anticipated. This can likely be explained by the difference in number of cab_rides. The first surge multiplier group contained 18,570 samples whereas this group only contained 2,405 samples. This difference well explains the variance in cab route price data.

C.

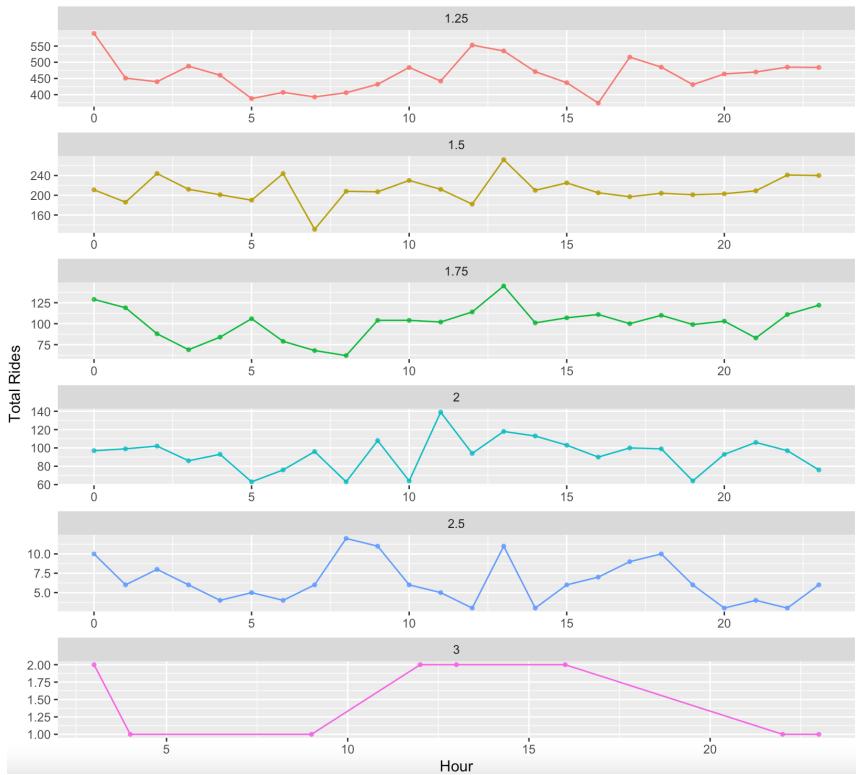
Total Rides vs Day of the Week Per Surge Multiplier



This plot examined the total number of Lyft rides per day of the week and their respective surge multiplier. Regardless of surge multiplier, the highest number of rides occurred on Tuesday and Wednesday.

D.

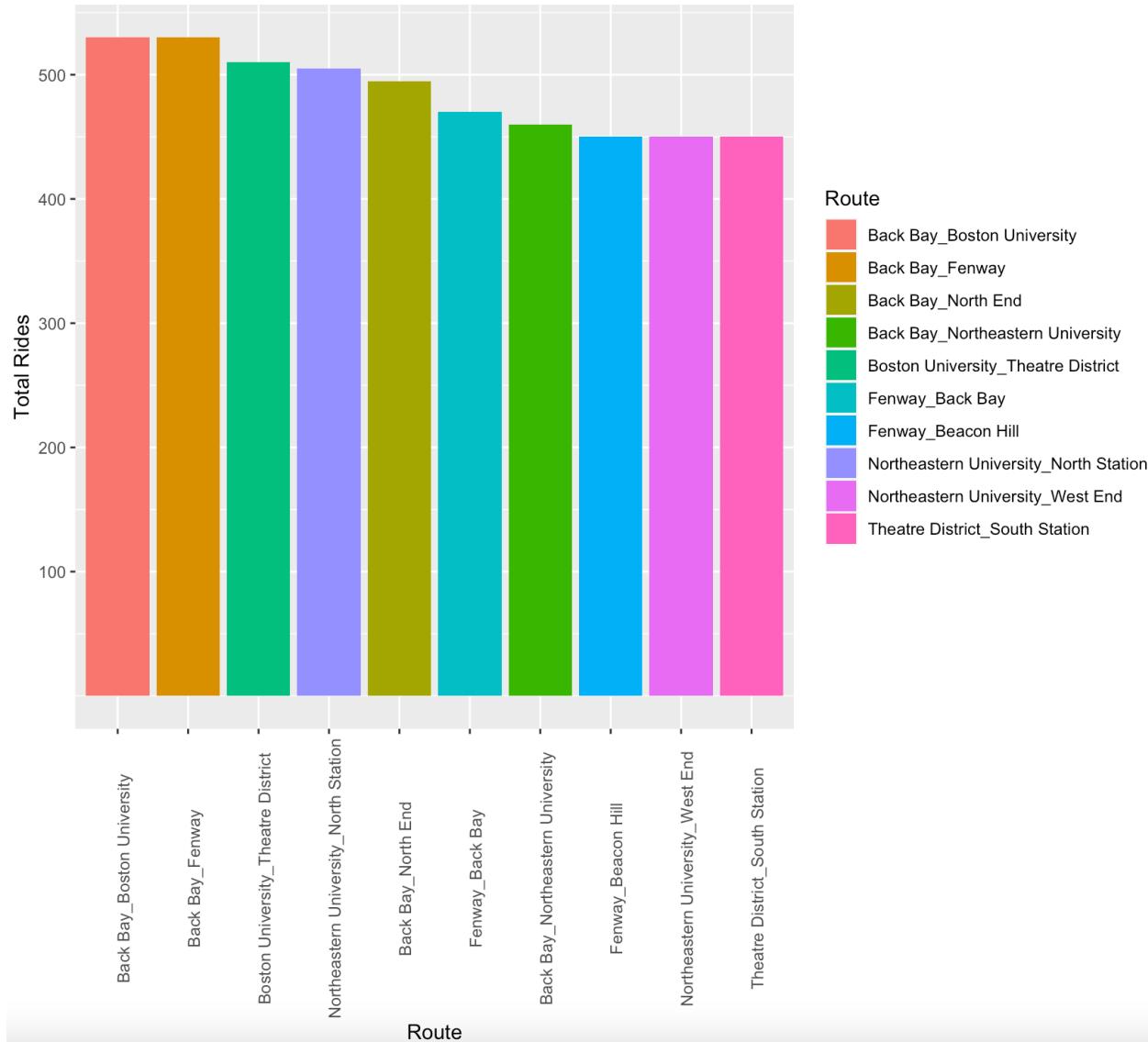
Lyft: Total Rides vs Hour of the Day Per Surge Multiplier



This plot examined the total number of Lyft rides per hour of the day and their respective surge multiplier. Regardless of surge multiplier, the highest number of rides took place between 12 and 1 o'clock in the afternoon.

E.

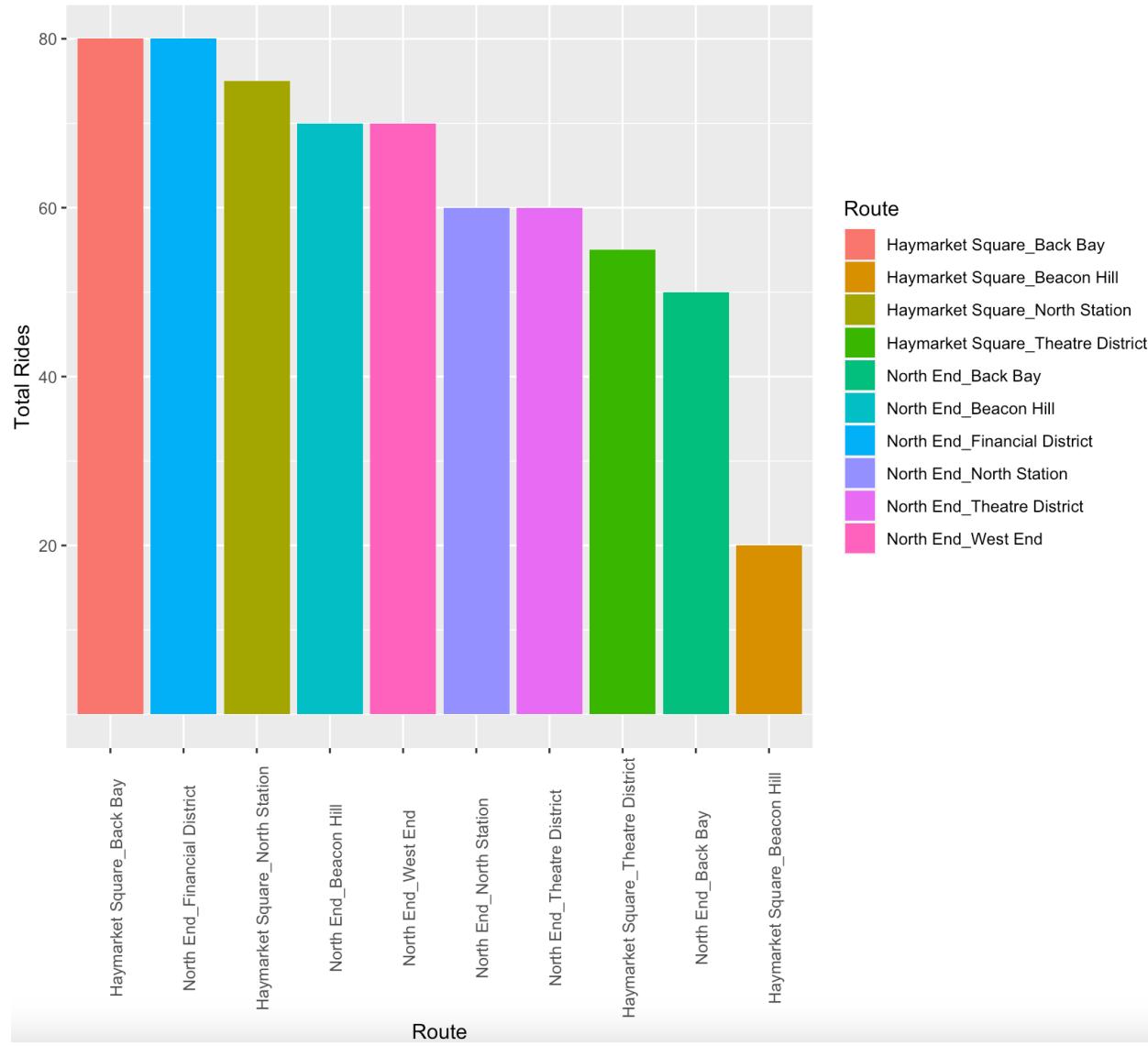
Surge: Relationship Between Lyft Rides and Routes - Upper Bound



This plot shows that the routes from Back Bay to Boston University and Fenway more often correlated with a surge in Lyft prices. In addition, rides originating in Back Bay constituted forty percent of this graph.

F.

Surge: Relationship Between Lyft Rides and Routes - Lower Bound



This plot shows that routes from Haymarket Square to Beacon Hill less likely experienced a surge in Lyft prices. Furthermore, routes originating in Haymarket Square and North End constituted this entire graph with respect to less common surge prices.