

A Story of Uber in New York City

Uber, as a new worldwide traffic phenomenon founded in 2009, had a strong market entry advantage in the for-hire-vehicle (FHV) industry. We are interested in the footsteps of Uber in NYC and would like to provide some business and consumer insights about Uber. In this website, we compared the usage of Uber with other for-hire-vehicle companies like Lyft, regarding pick-up locations and time. We also interpreted Uber usage with respect to weather conditions, weekdays/weekends and holidays. Moreover, we analyzed some customer reviews of Uber.

Authors

Xiaoyu Dai @lonadai0121 xd2236@columbia.edu
Wendy Li @Wendylwy wl2710@columbia.edu
Yixin Lin @rebeccalyx yl4293@columbia.edu
Nikky Xiong @nikkyxiong wx2225@columbia.edu

More about the project

[Click here to see our process book](#)
[Click here check out our datasets](#)
[Click here to see the code](#)



Data Source

The original data source that contains data on over 4.5 million Uber pickups in New York City from April to September 2014 and 14.3 million more Uber pickups from January to June 2015, as well as trip-level data on 10 other for-hire vehicle (FHV) companies. (<https://github.com/fivethirtyeight/uber-tlc-foil-response/>)

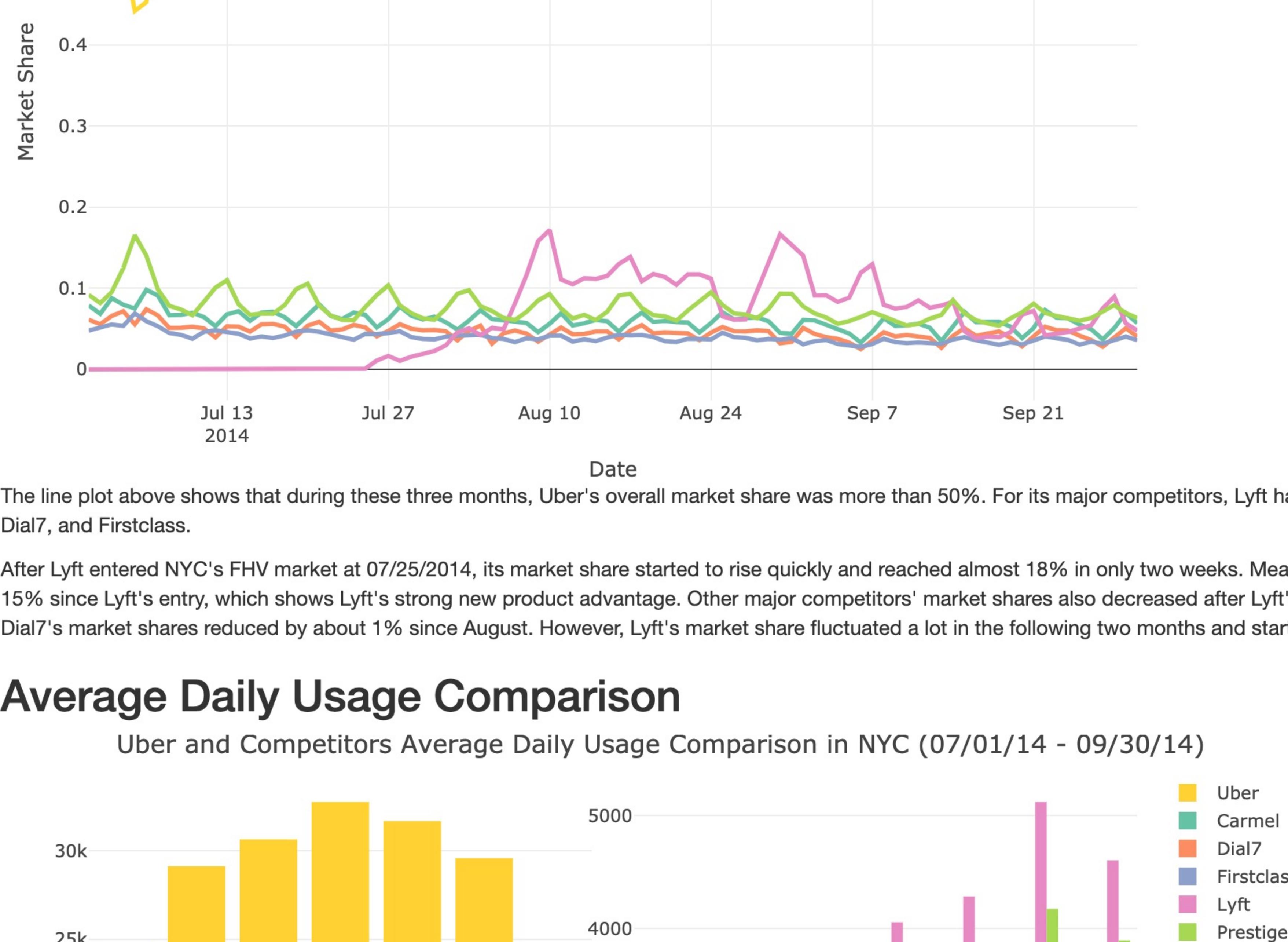
The data is obtained by FiveThirtyEight from the NYC Taxi & Limousine Commission (TLC) by submitting a Freedom of Information Law request on July 20, 2015.

In order to compare the usage of Uber and its major competitors in New York City, we chose the time period from July to September in 2014 during which these FHV companies all have available trip data in this dataset.

Data Visualization

Market Share Comparison

Market Share Comparison in NYC (07/01/14 - 09/30/14)

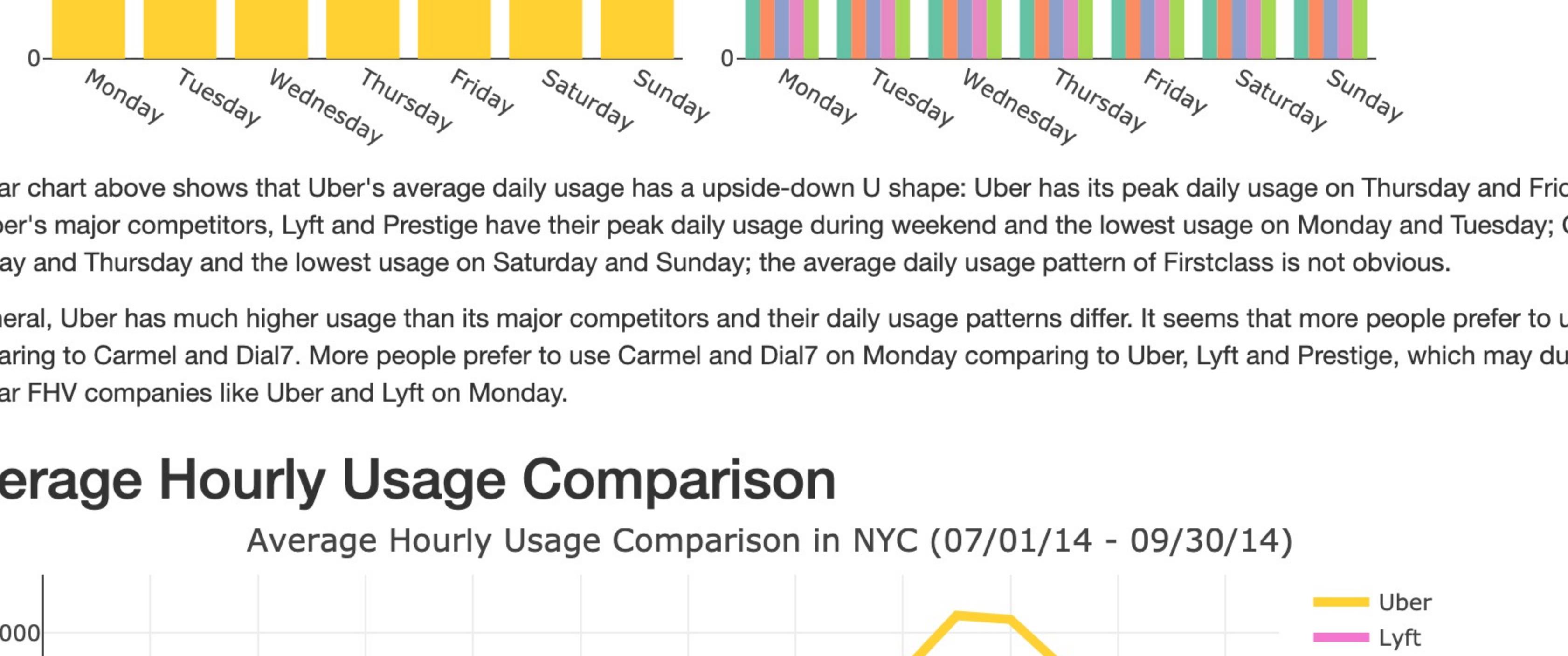


The line plot above shows that during these three months, Uber's overall market share was more than 50%. For its major competitors, Lyft has the highest share, followed by Prestige, Carmel, Dial7, and Firstclass.

After Lyft entered NYC's FHV market at 07/25/2014, its market share started to rise quickly and reached almost 18% in only two weeks. Meanwhile, Uber's market share dropped by more than 15% since Lyft's entry, which shows Lyft's strong new product advantage. Other major competitors' market shares also decreased after Lyft's entry but not much. For example, Carmel and Dial7's market shares reduced by about 1% since August. However, Lyft's market share fluctuated a lot in the following two months and started to tumble later.

Average Daily Usage Comparison

Uber and Competitors Average Daily Usage Comparison in NYC (07/01/14 - 09/30/14)

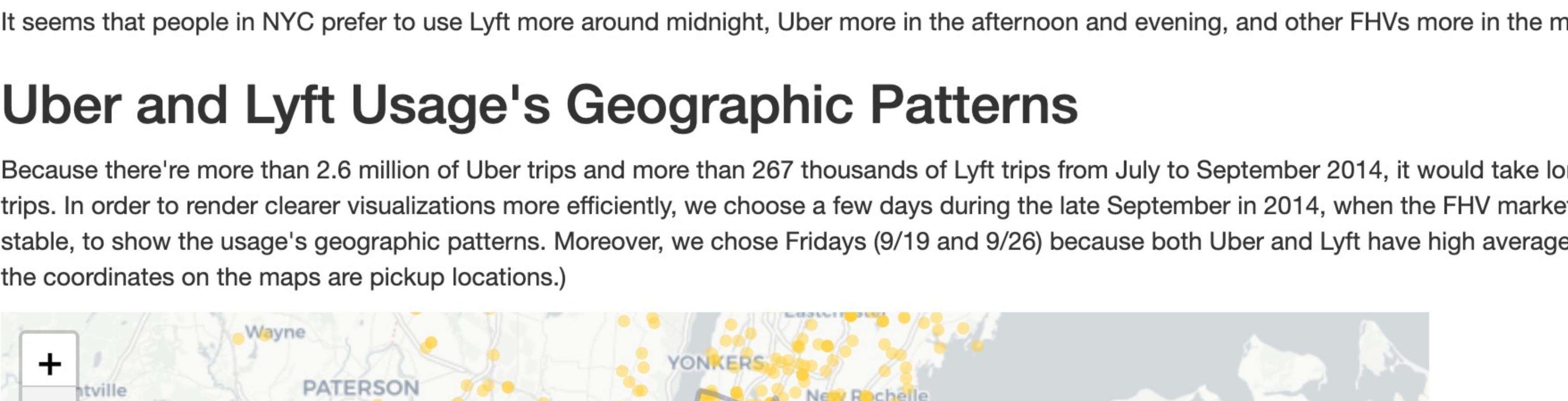


The bar chart above shows that Uber's average daily usage has an upside-down U shape: Uber has its peak daily usage on Thursday and Friday and the lowest usage on Monday and Sunday. As for Uber's major competitors, Lyft and Prestige have their peak daily usage during weekend and the lowest usage on Monday and Tuesday; Carmel and Dial7 have their peak daily usage on Monday and Thursday and the lowest usage on Saturday and Sunday; the average daily usage pattern of Firstclass is not obvious.

In general, Uber has much higher usage than its major competitors and their daily usage patterns differ. It seems that more people prefer to use Uber, Lyft and Prestige on Friday and weekend comparing to Carmel and Dial7. More people prefer to use Carmel and Dial7 on Monday comparing to Uber, Lyft and Prestige, which may due to people want to avoid peak usage periods of some popular FHV companies like Uber and Lyft on Monday.

Average Hourly Usage Comparison

Average Hourly Usage Comparison in NYC (07/01/14 - 09/30/14)



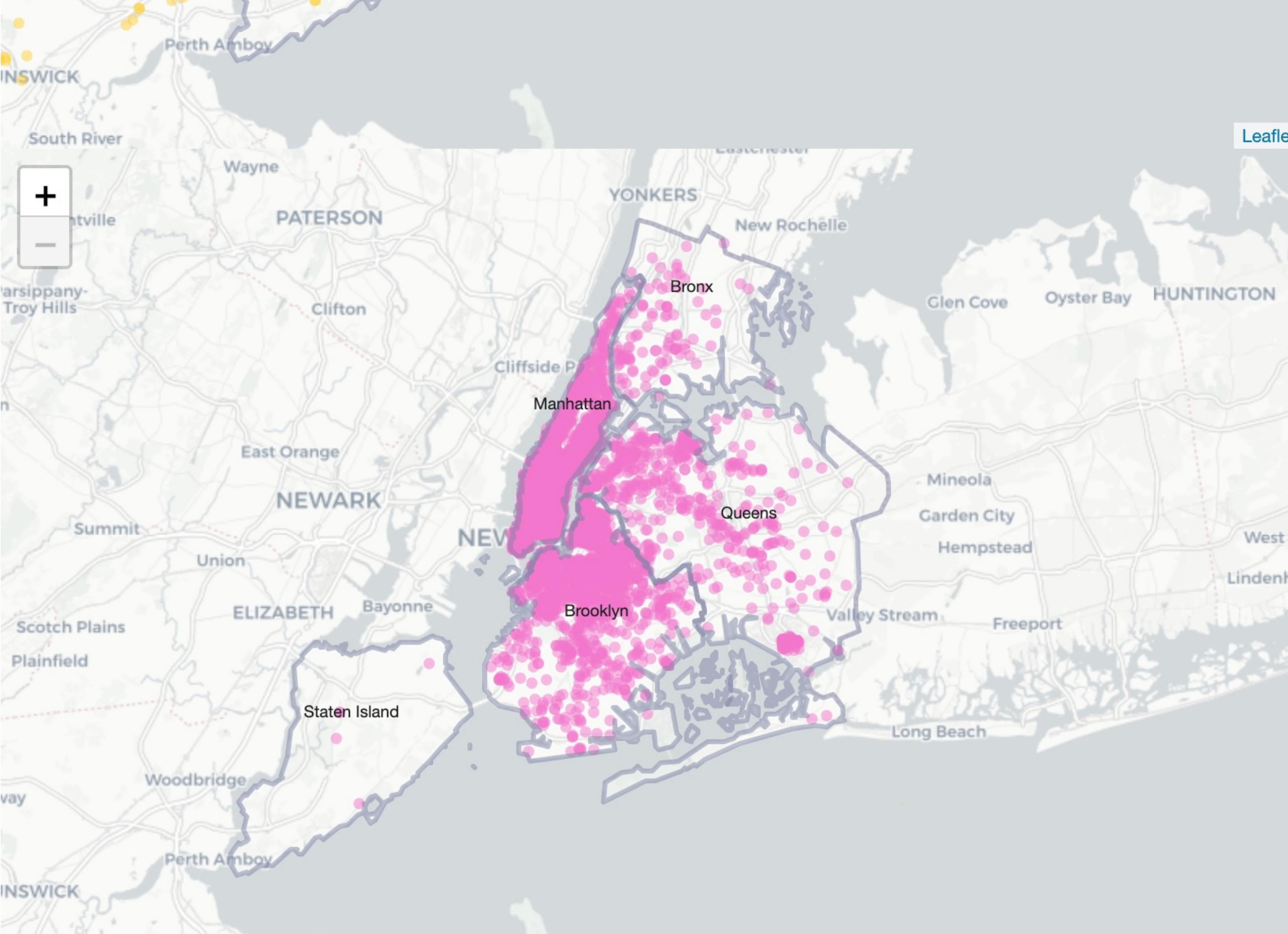
The plot above shows that Uber has its average peak hours from 3PM to 9PM during which it has more than 1700 hourly usage in NYC. Uber has its off-peak hours from 12AM to 6AM during which it has less than 900 hourly usage.

Lyft has much different hourly usage patterns with its average peak hours from 10PM to 2AM during which it has more than 280 hourly usage in NYC. Lyft has its off-peak hours from 7AM to 11AM during which it has less than 100 hourly usage. For other FHVs combined, they have similar off-peak hours with Uber but has its highest usage from 7AM to 10AM.

It seems that people in NYC prefer to use Lyft more around midnight, Uber more in the afternoon and evening, and other FHVs more in the morning.

Uber and Lyft Usage's Geographic Patterns

Because there're more than 2.6 million of Uber trips and more than 267 thousands of Lyft trips from July to September 2014, it would take long time to load the crowded maps if we plot all the trips. In order to render clearer visualizations more efficiently, we choose a few days during the late September in 2014, when the FHV market in NYC is less affected by Lyft's entry and more stable, to show the usage's geographic patterns. Moreover, we chose Fridays (9/19 and 9/26) because both Uber and Lyft have high average usage on Friday based on the daily usage plot. (Note: the coordinates on the maps are pickup locations.)



Uber (top) and Lyft (bottom) pickup locations in NYC on 9/19/14 and 9/26/14 combined

The map shows that Uber and Lyft have similar pickup density difference in NYC's five different boroughs: both Uber and Lyft have their most pickups at Manhattan, followed by Brooklyn, Queens, Bronx and Staten Island. Moreover, they both have higher pickup density within and around Manhattan.

Uber has much higher pickup density and covers more distant and border areas in NYC compared to Lyft, which shows Uber drivers' broader coverage in NYC.

Description and Plots

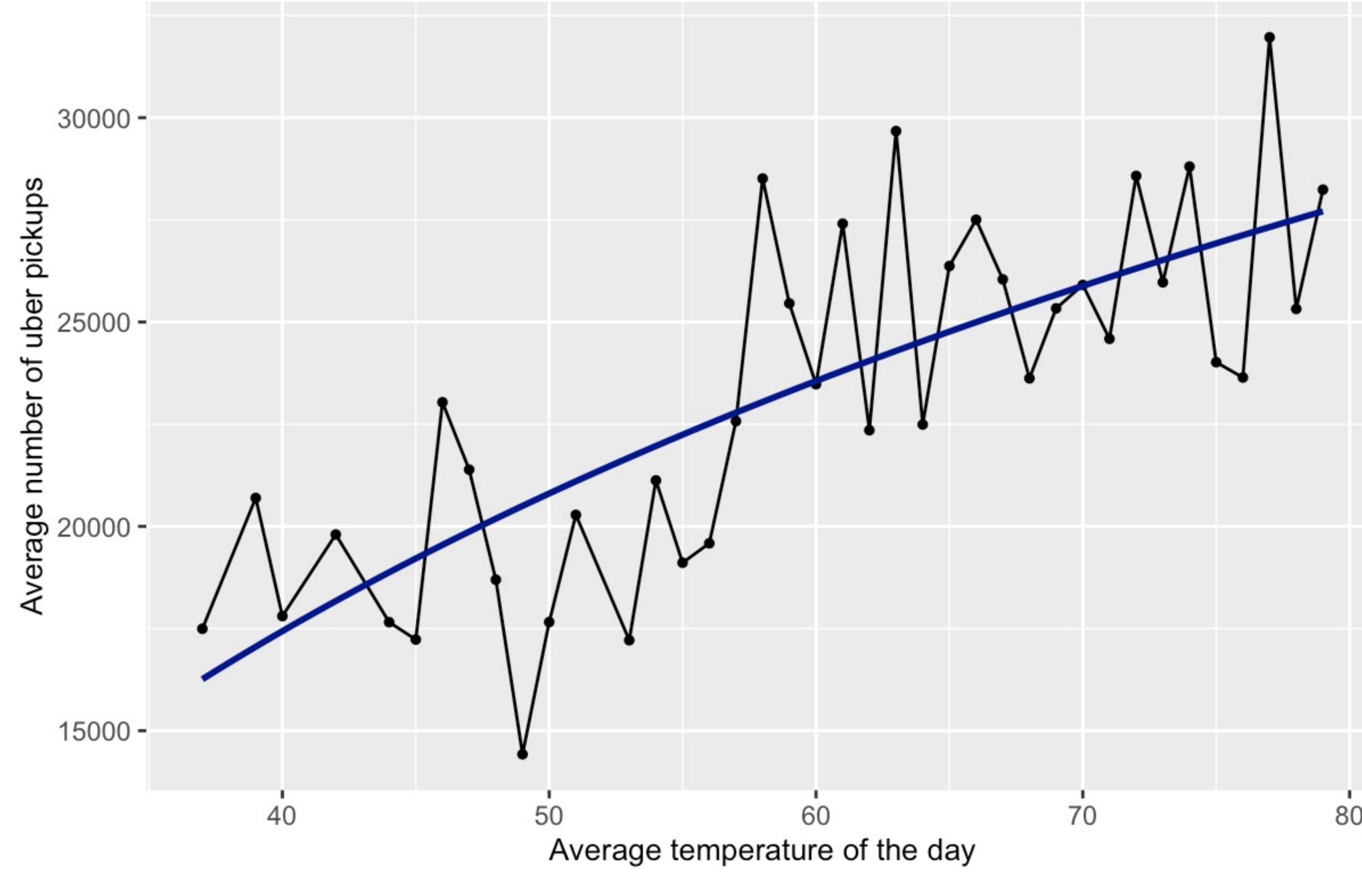
Map of NYC pickups

Weather API

We're using the weather API from [WorldWeatherOnline](#) to acquire useful weather information based on location, date and time stamp in the cleaned existing dataset of uberpickups.

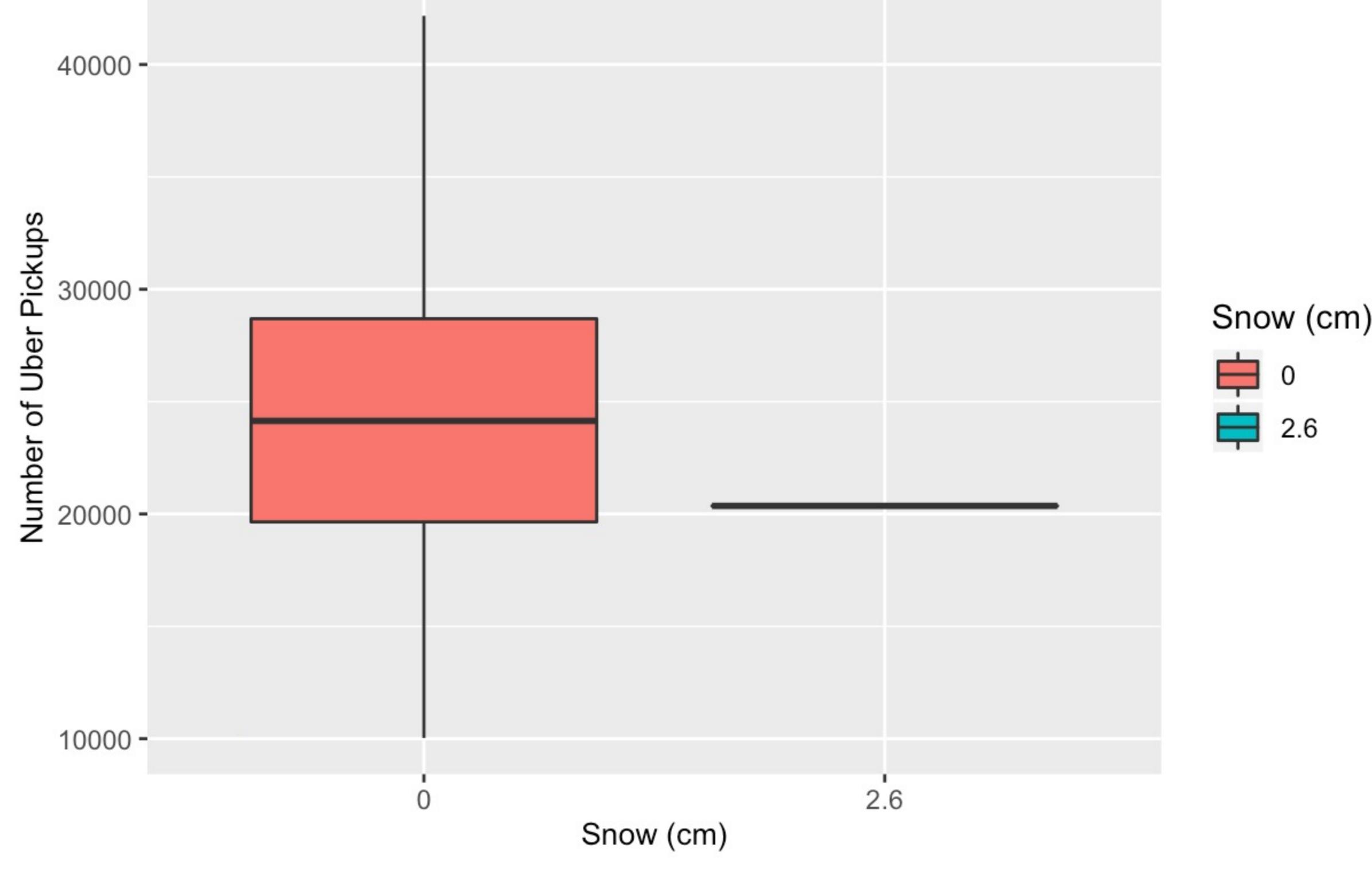
Visualization

Number of Uber Pickups Vs. Average Temperature



From the line plot with smoothing line, we can see in general, as the average temperature of the day increased, the average number of uber pickups increased. In addition, the line plot reached a lowest average number of uber pickups approximately at 49F; and from approximately 45F to 56F, the line plot was below the smoothing line. It is reasonable that when the average temperature was very low or very high, the uber pickups tended to be more than those formed in a relatively comfortable temperature.

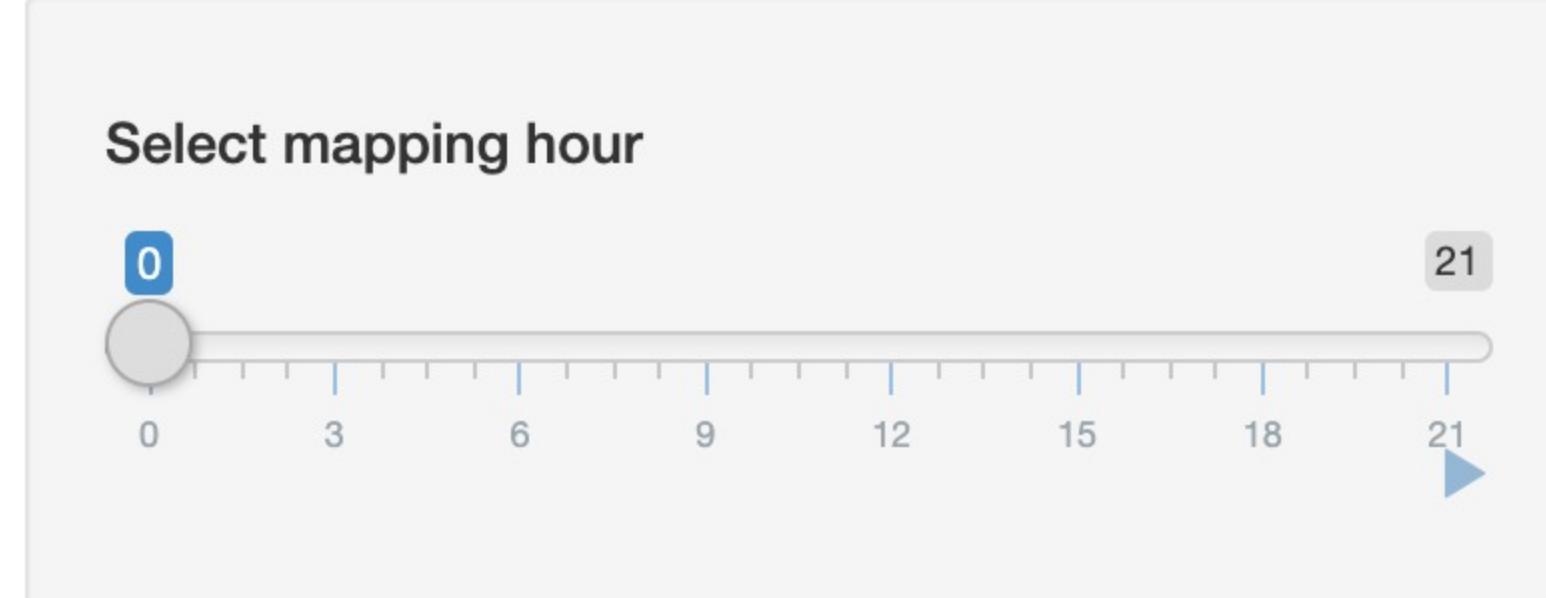
Number of Uber Pickups Vs. Snow (cm)



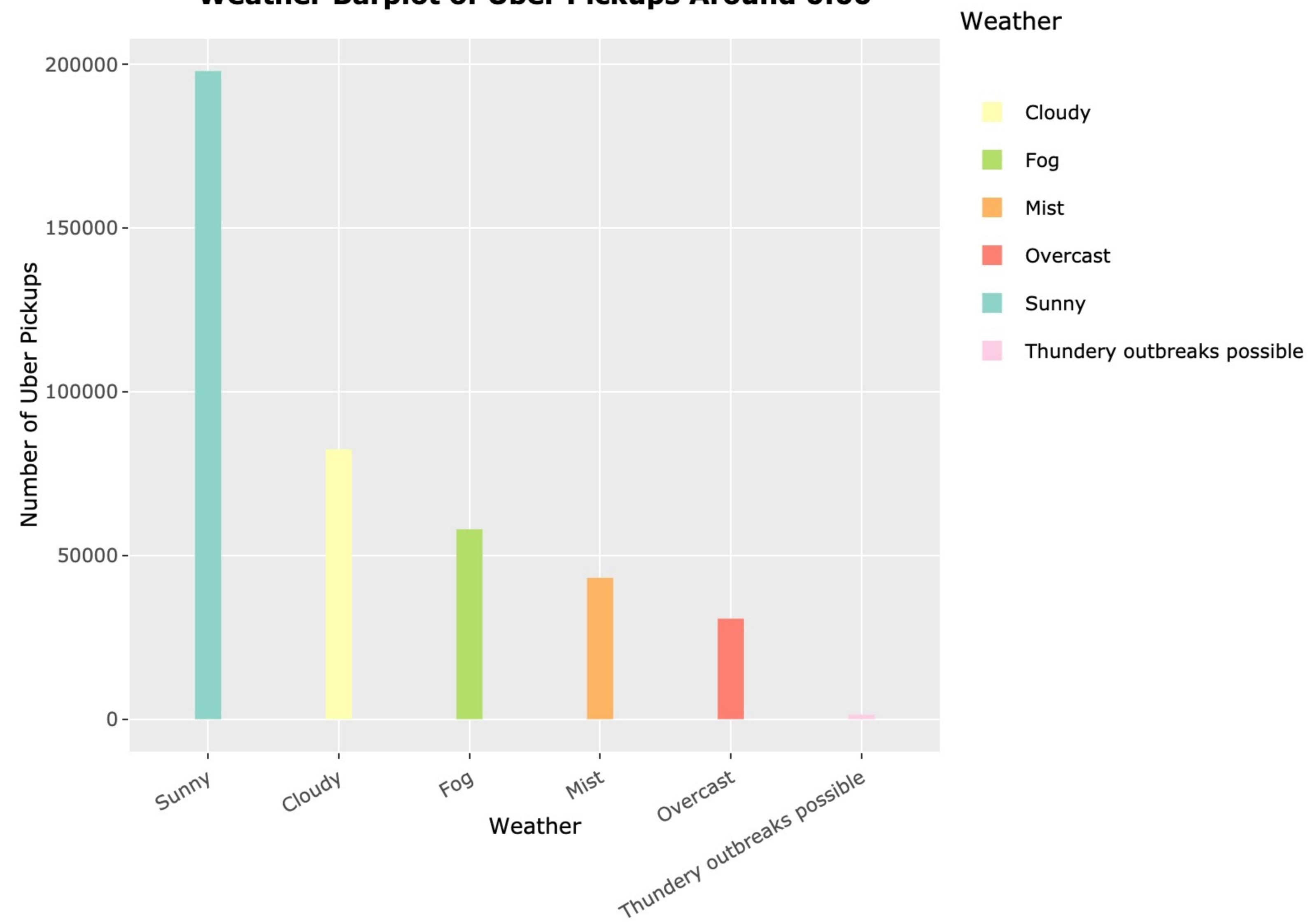
Since the data of uber only covered from April to September, there is only one snow day. Even though there might be bias towards the incompleteness of the data, we can still see approximately 75% of the average number of uber pickups in a day without snow is higher than the number of uber pickups in a day with snow.

Since there are a lot of descriptions of weather and some of them are very similar, we decided to combine some weather descriptions and the left ones are "Sunny", "Light rain", "Moderate rain", "Heavy rain", "Cloudy", "Overcast", "Mist", "Fog", "Thundery outbreaks possible", "Patchy snow possible" and "Heavy snow".

To minimize the effect of time of the day on the number of uber pickups and in order to have a better look at the relationship between weather and the number of uber pickups, I will show the barplots of weather in different time stamps.



Weather Barplot of Uber Pickups Around 0:00



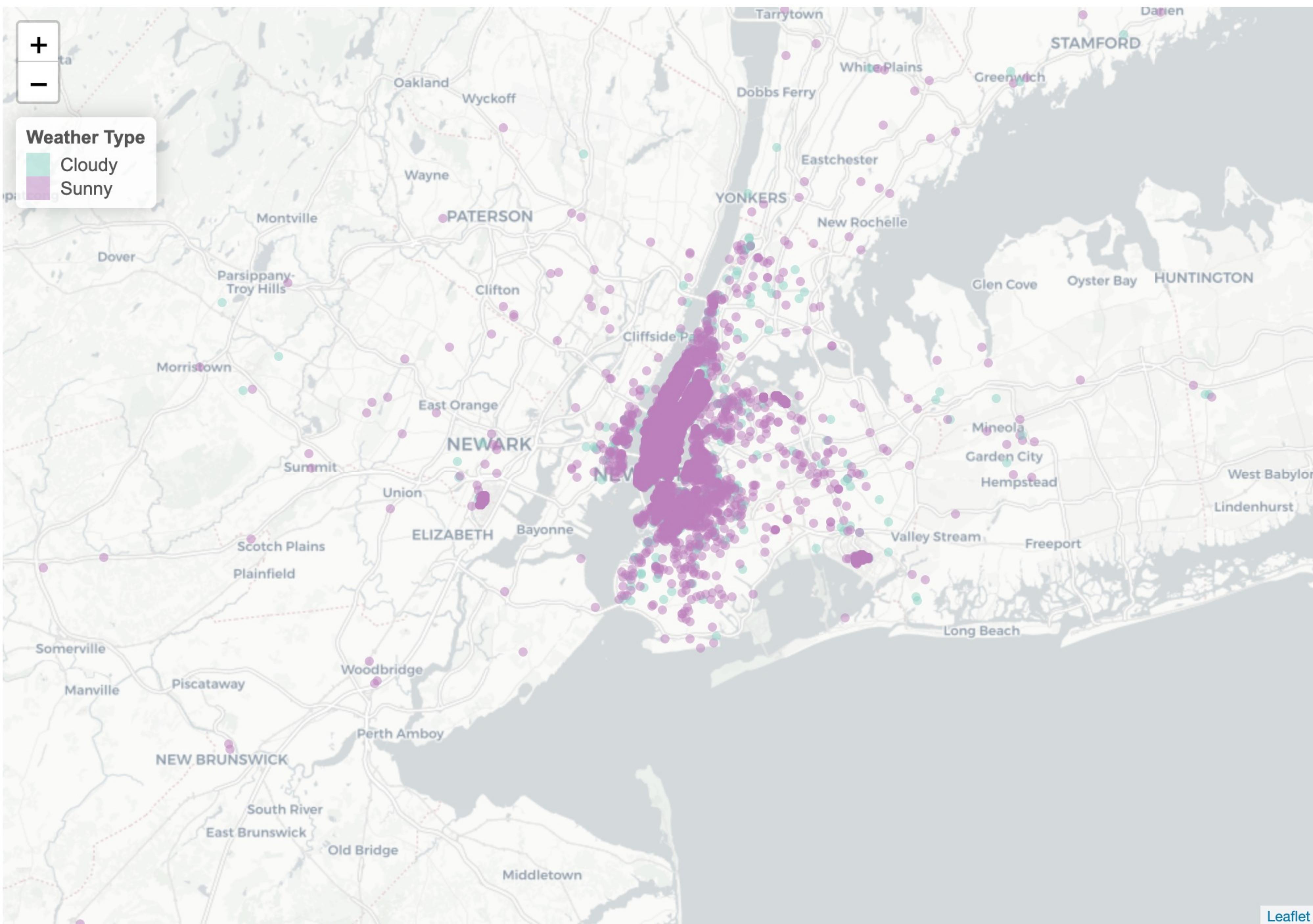
I also used different colors to represent different types of weather for an easier comparison. We can see in all the time stamp plots, "Sunny" weather received the most number of Uber pickups. "Cloudy" or "Light rain" reached the second place. As observed from the plots, in general, if the weather was very bad, such as "Heavy snow", "Thundery outbreaks", "Heavy rain", "Overcast", "Mist" and "Fog", the number of uber pickups tended to be very small.

[Description and Plots](#)[Map of NYC pickups](#)**Select date:**

2014-04-01

Choose Weather:

NULL



Weekdays vs. Weekends

Holidays vs. Nonholidays

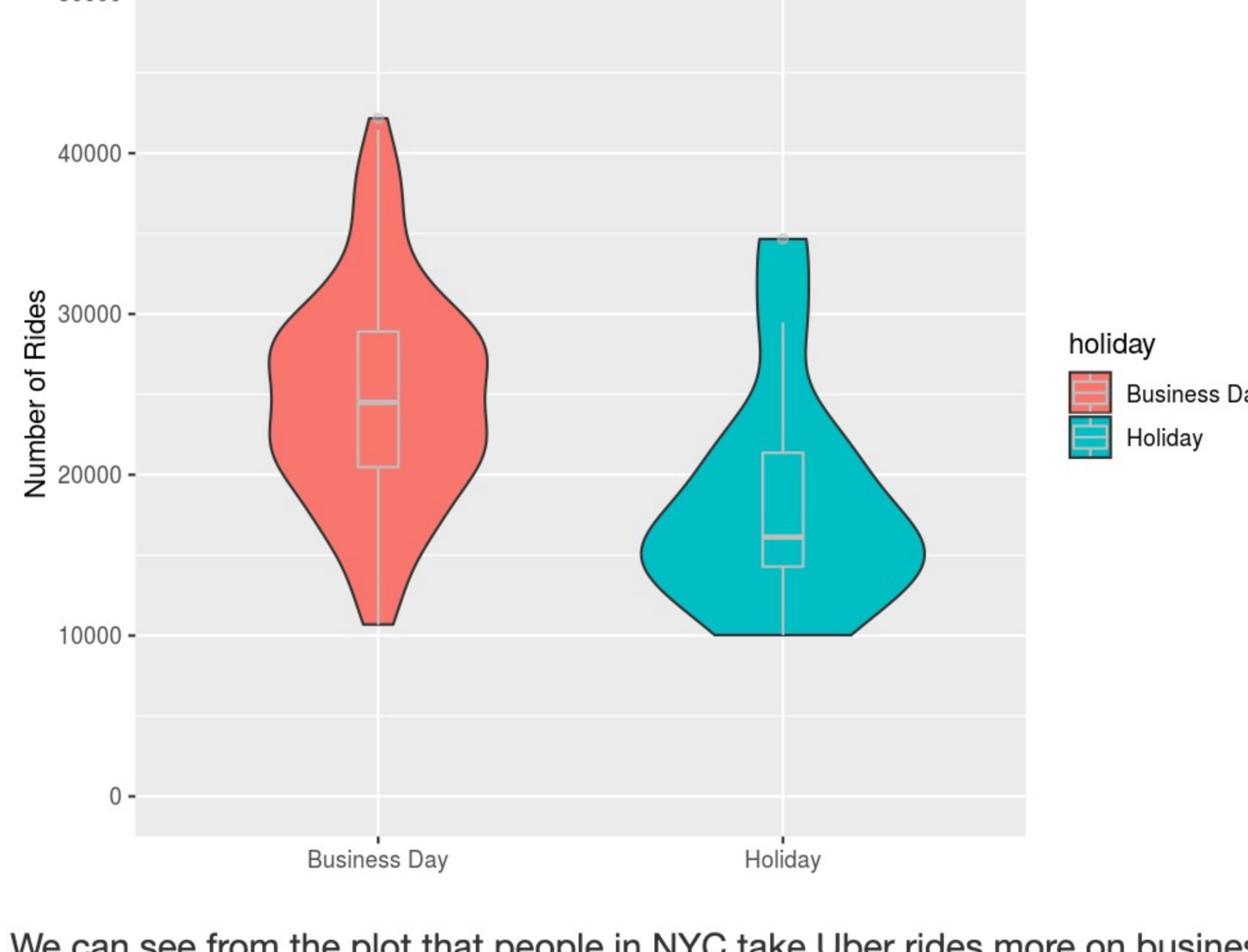
Holidays Included

We include both federal holidays and non-federal holidays from April to September, 2014.

| date | name |
|-----------|-------------------|
| 4/1/2014 | April Fool's Day |
| 4/18/2014 | Good Friday |
| 4/20/2014 | Easter |
| 4/22/2014 | Earth day |
| 5/1/2014 | Mother's Day |
| 5/17/2014 | Armed Forces Day |
| 5/26/2014 | Memorial Day |
| 6/8/2014 | Pentecost |
| 6/15/2014 | Father's Day |
| 7/4/2014 | Independence Day |
| 7/27/2014 | Parents' Day |
| 9/1/2014 | Labor Day |
| 9/7/2014 | Grandparents' Day |
| 9/17/2014 | Citizenship Day |

Distribution of Uber Rides on Holidays and Nonholidays

Distribution of Uber Rides: Nonholiday vs. Holiday

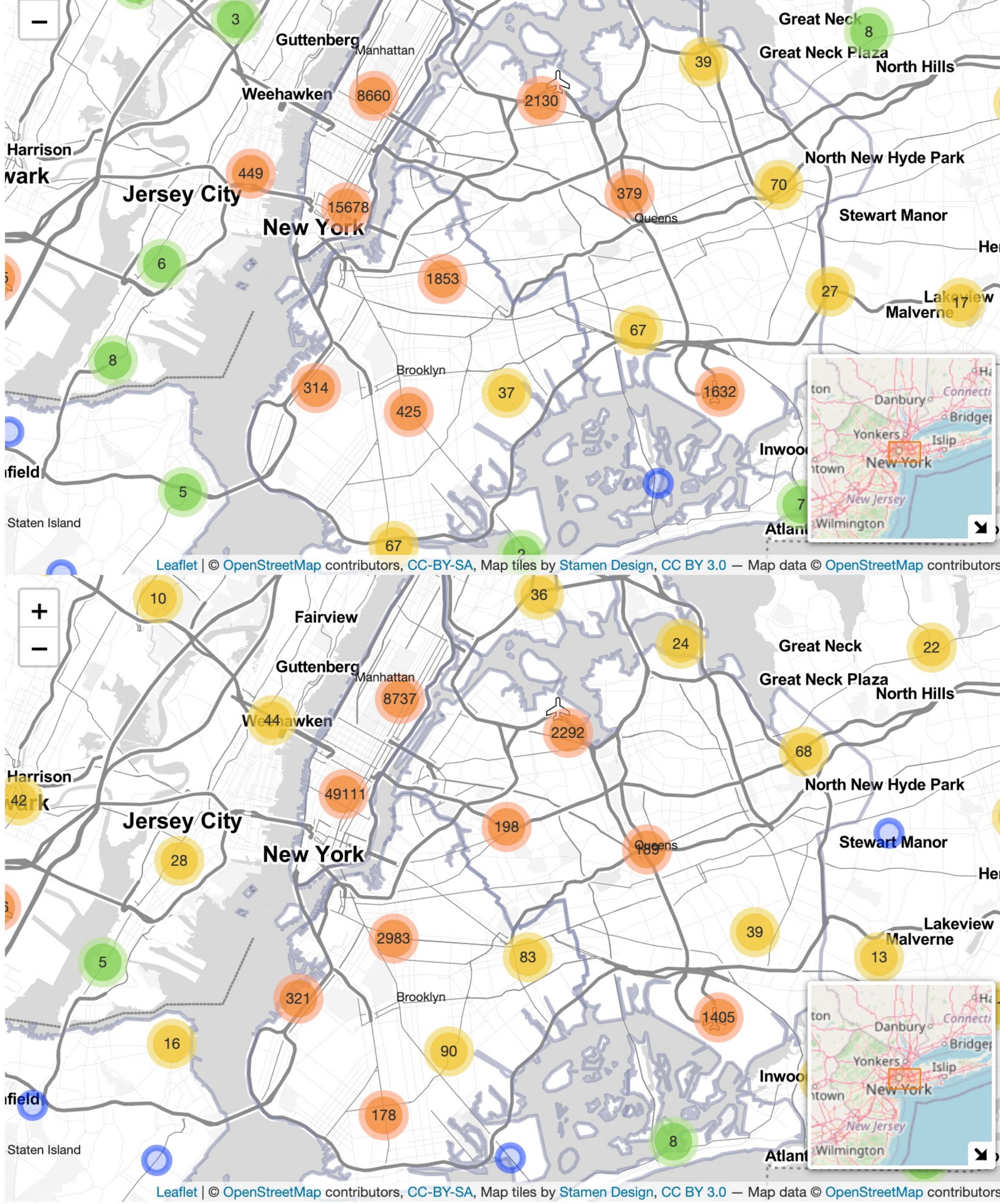


We can see from the plot that people in NYC take Uber rides more on business day than holidays. The average number of rides on business day is around 25000 per day, and the average number of rides on holidays is around 15000 per day. Also, from the plot we can see that Uber rides on holidays are clustered in the bottom.

Map of pick-up places in Holidays vs. Nonholidays

To visualize the pickup locations on business day versus holiday, I picked two federal holidays in 2014, 7/4/2014, the Independence Day and 9/1/2014, the Labor Day, to represent holidays. And picked two business day close to these two holidays, 7/10/2014, Wednesday, and 9/4/2014, Thursday.

Cluster Maps



From the first map we can see that, in New York City, uber rides are more intensive in Manhattan, while Queens and Brooklyn have relatively lower number of Uber rides. And from the second map we can notice, in business days, Manhattan Midtown and downtown show a great increase in Uber rides.

Data Source

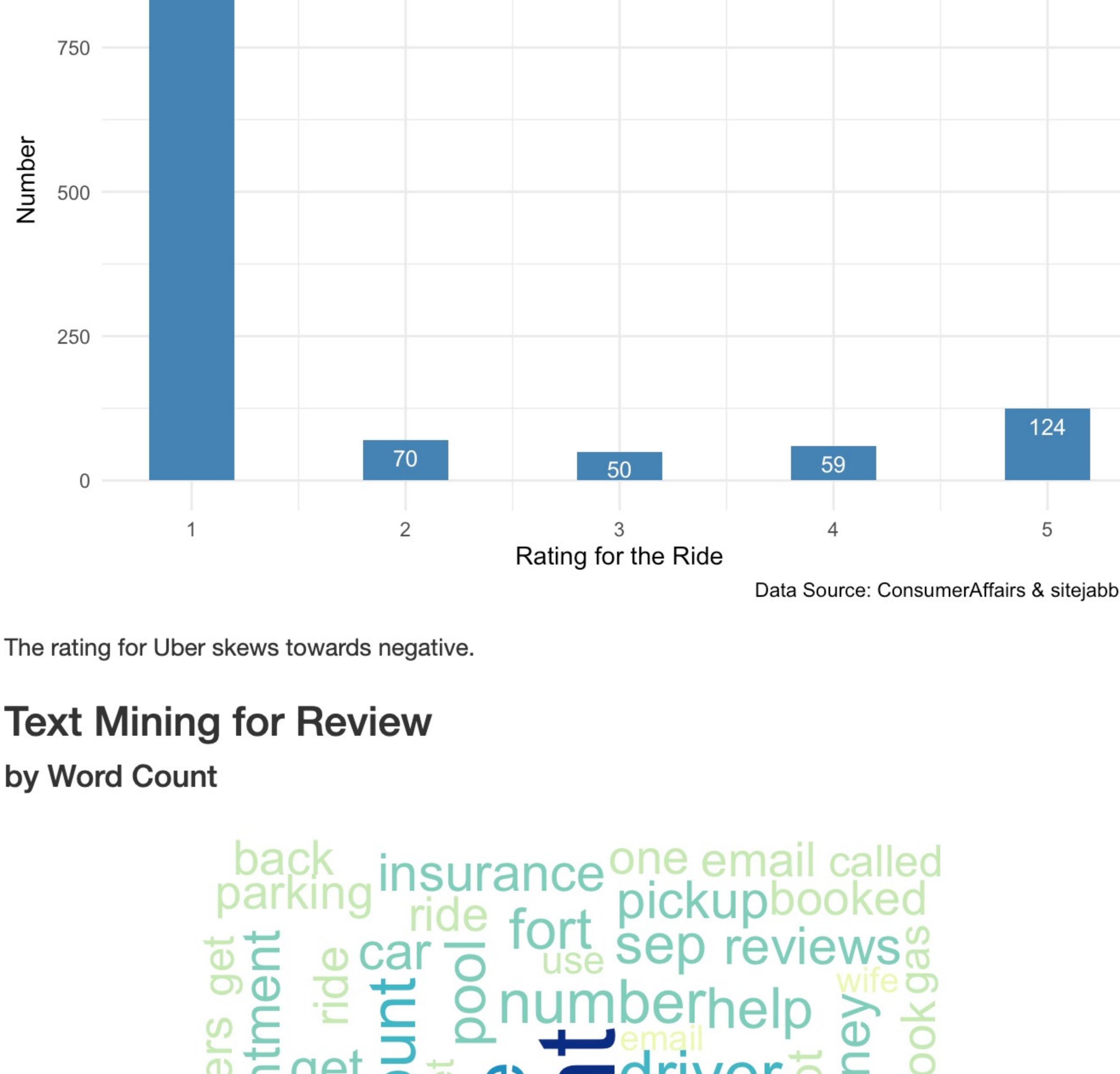
The user scrapped the uber ride reviews from two websites, <https://www.consumeraffairs.com/travel/uber.html> and <https://www.sitejabber.com/reviews/www.uber.co>. The user also created the column “Ride Sentiment” using the ride’s rating. If the rating is above 2, the review is treated it as a positive sentiment(1). Otherwise, it’s a negative sentiment(0).

The user also created the column "Ride Sentiment" using the ride's rating. If the rating is above 3, the review is treated it as a positive sentiment(1). Otherwise, it's a negative sentiment(0).

Data Visualization

Rating

1000



cal ap di s a O

driver gift
ride gift



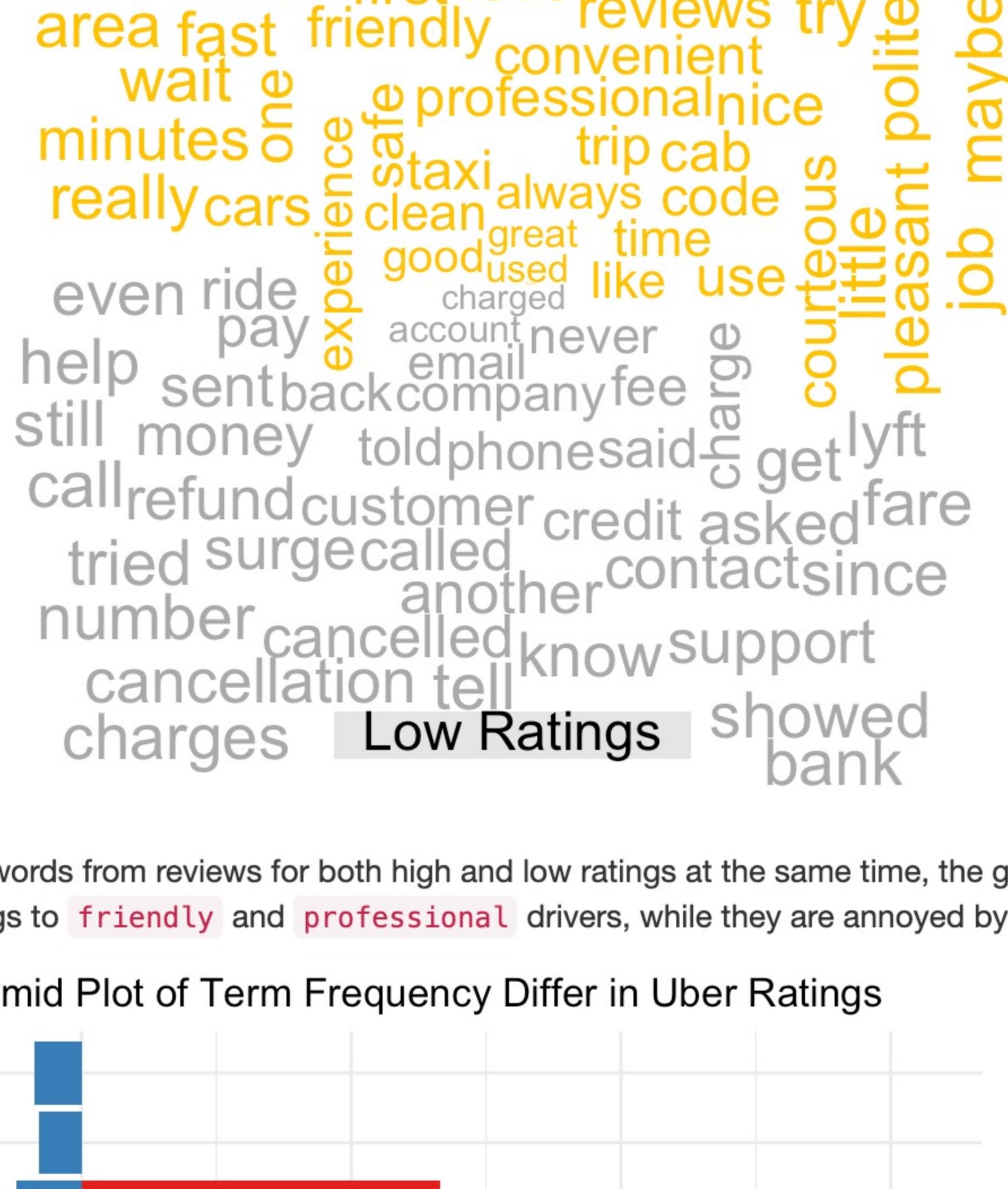
Key unigrams like `crooked`, `horrible`, and `shut` are the most significant in reviews for I like.

Key diagrams
words do not



The bigram word cloud provides more information about the review for Uber. For instance, phone number, surge pricing, customer care, driver cancelled, cancellation fee are top issues that mentioned in reviews as shown in the graph.

by Ratings



A horizontal stacked bar chart comparing the breakdown of costs for 'service' and 'ride'. The y-axis lists the categories: 'service' and 'ride'. Each category has a total bar divided into two segments: a blue segment on the left and a red segment on the right. The blue segments represent a smaller portion of the total cost, while the red segments represent a larger portion.

| Category | Blue Segment (Estimated) | Red Segment (Estimated) |
|----------|--------------------------|-------------------------|
| service | \$100 | \$200 |
| ride | \$100 | \$300 |

A horizontal bar chart with two bars. The first bar is blue and positioned near the left edge. The second bar is red and extends further to the right, indicating a higher value than the blue bar.

