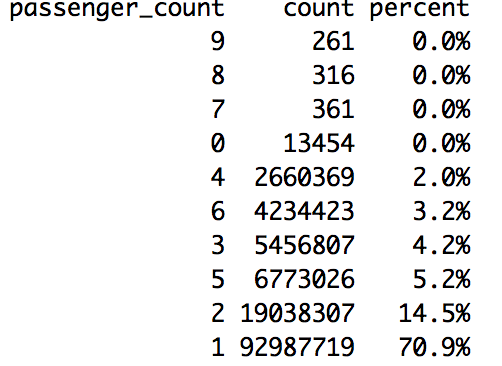
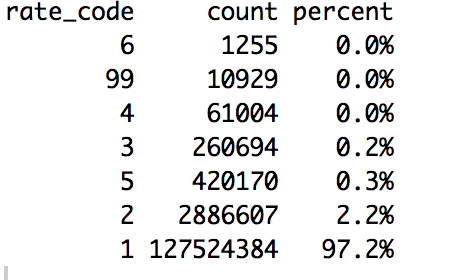
**New York City Yellow Taxi Trip 2016 Analysis**

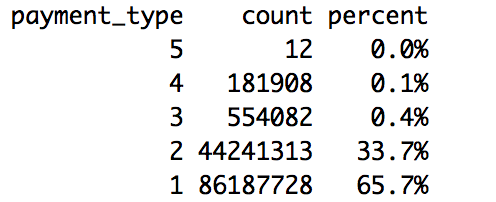
* Distribution of categorical variables
  + Passenger count: Most of the passengers are taking taxi alone (70.9%) or in pairs (14.5%). The proportion of trips with more than 2 passengers is around 15%. Also, there are some records with 0 passenger count. This could be due to taxi driver never key in number of passengers.

****

* + Rate code: Majority of the trips use standard rate (97.2%). There are 2.2% trips with JFK rate and 0.2% with Newark rate. This could be due to JFK is the largest international airport in NYC leading to more taxi trips from/to there. There are also some outlier data with rate code as 99.



* + Payment type: There are more trips paid in credit card (65.7%) as compared to cash (33.7%).

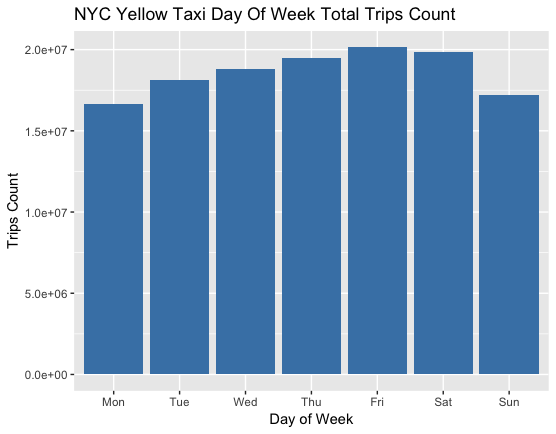
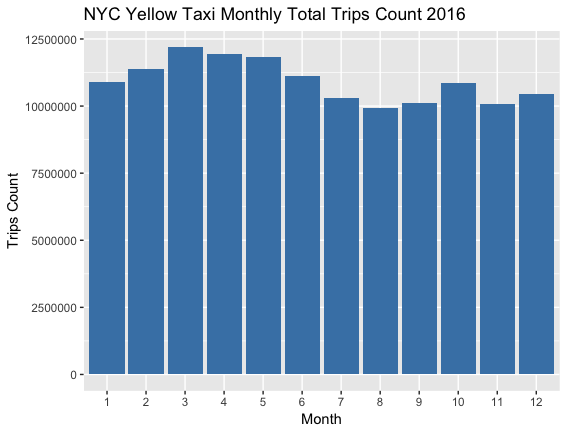


Data Selection Note: New York City Yellow Taxi Trips in 2016 was used since 2016 is the most recent year. Outliers data with fare amount <0 and trip distance >=5000 are excluded from the following analysis.

**Total trips count by month and day of week**

Monthly total trip count is around 10 million. There are slightly more trips in the first half of the year compared to the second half of the year.

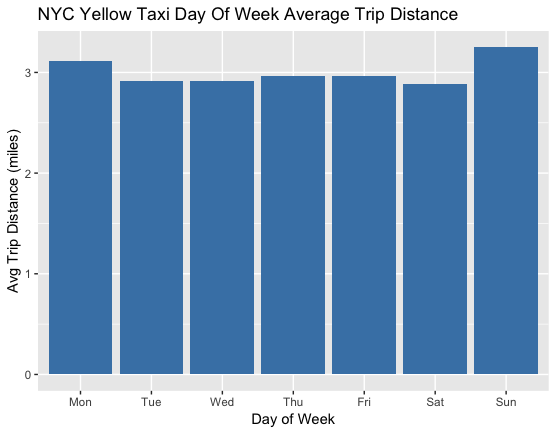
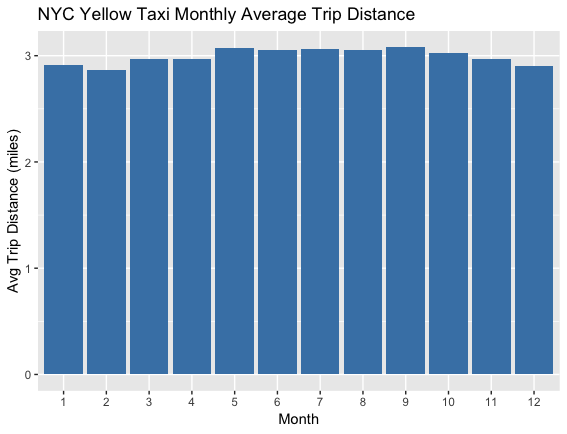
Generally, Friday and Saturday have the most number of trips. Monday and Sunday has the least number of trips.



**Average monthly trip distance & Average trip distance by day of week**

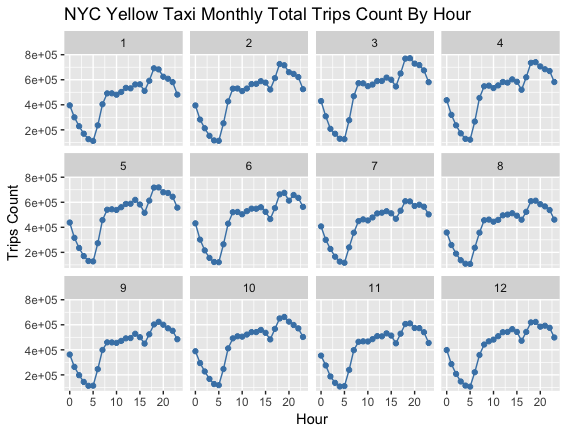
Monthly average trip distance is stable throughout the year and is around 3 miles.

Average trip distance on Monday and Sunday is slightly longer than other days in a week.

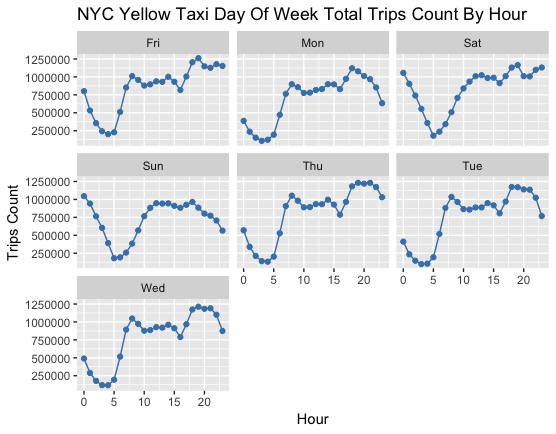


**Peak Hour Analysis: Total trip counts by hour and month & Total trip counts by hour and day of week**

There is no significant difference in peak/off-peak hour pattern across different months. Peak hour is 6-7pm and off-peak hour is 4-5am.

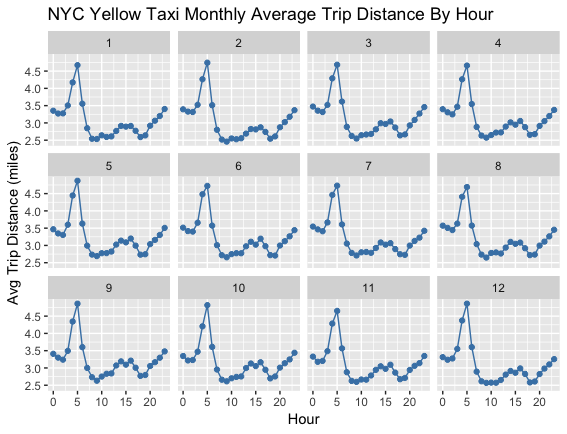


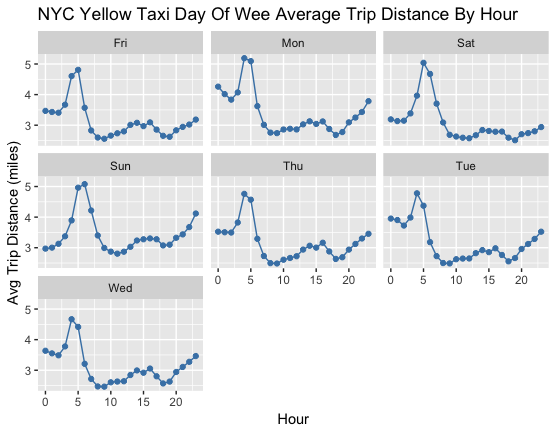
There is a different peak/off-peak hour pattern between weekdays and weekends. Weekday off-peak hour is from 2am to 5am. Weekday peak hour is from 6pm-9pm except Friday and 6pm to midnight on Friday. Saturday peak hour is from 6pm to 1am and off-peak hour is around 5am. Sunday peak hour is from midnight to 1am and off-peak hour is still around 5am. The pattern on Friday and Saturday can be explained by more people having dinner outside/hanging out overnight. Sunday’s night is not peak hour since people usually go back home earlier to prepare for work on the following Monday.



**Average trip distance by hour and month**

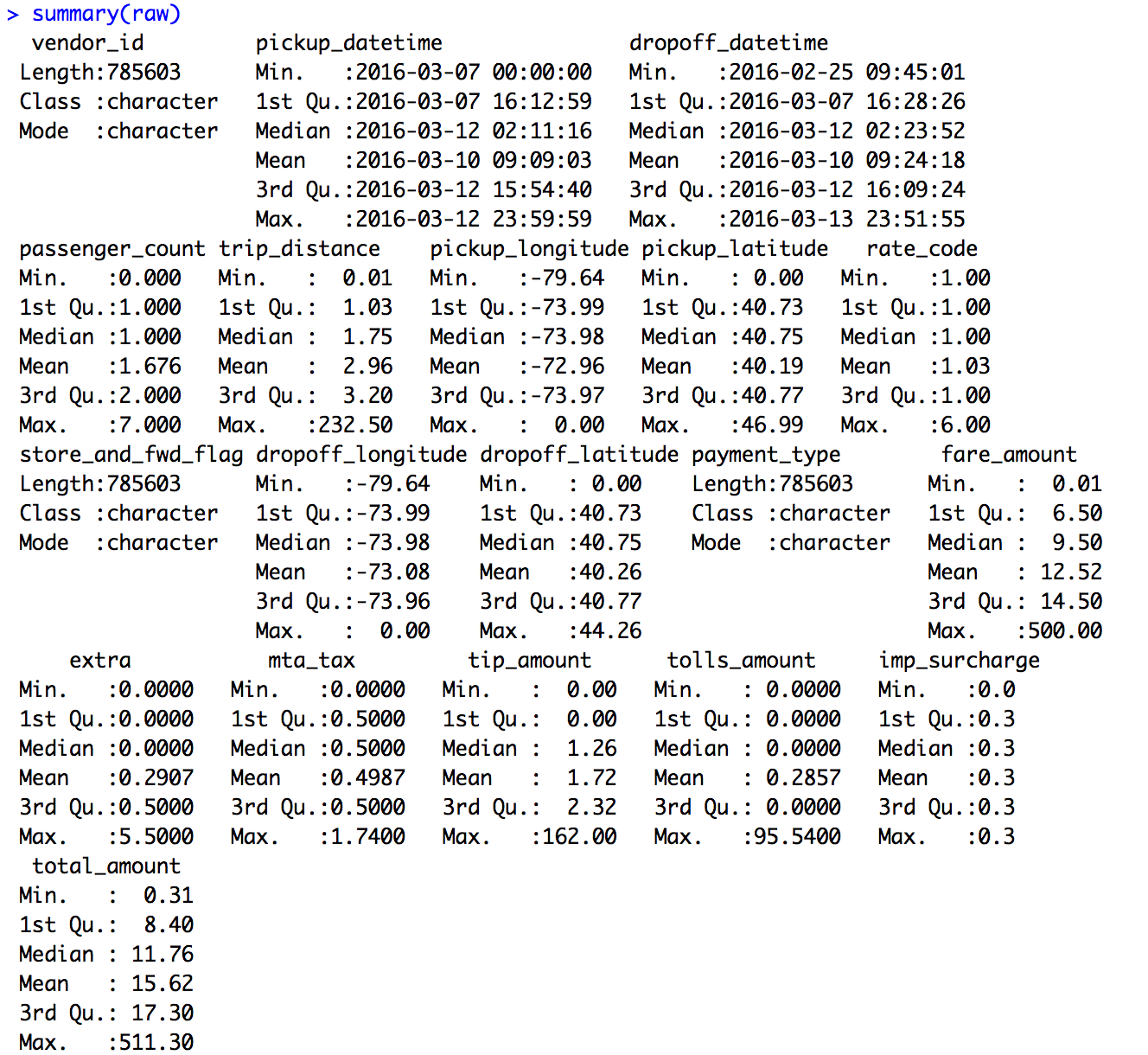
Same pattern is observed for 12 months. Trips with pick-up time earlier than 6am have longer distance, especially trips with pick-up time from 4am to 5am. This is possibly contributed by trips from and to the airports as shown in the latter part of the report. Even though there are more trips from 6-7pm, the trip distance is relatively short. Similar trend is observed by splitting data by day of week and hour.





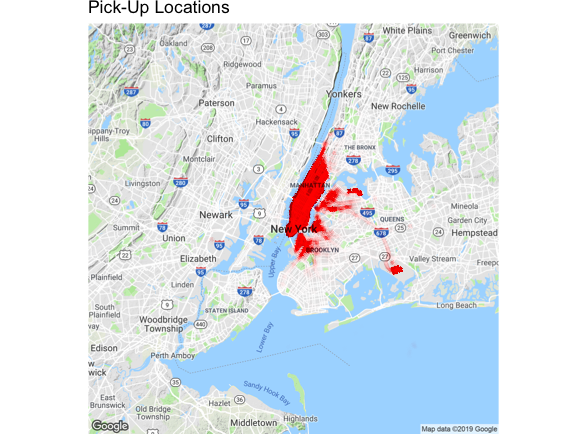
Next, we choose only 2 days’ data to perform geolocation analysis due to the huge data volume. Monday and Saturday are chosen to represent weekday and weekend. The same outlier filtering rule is still applied.

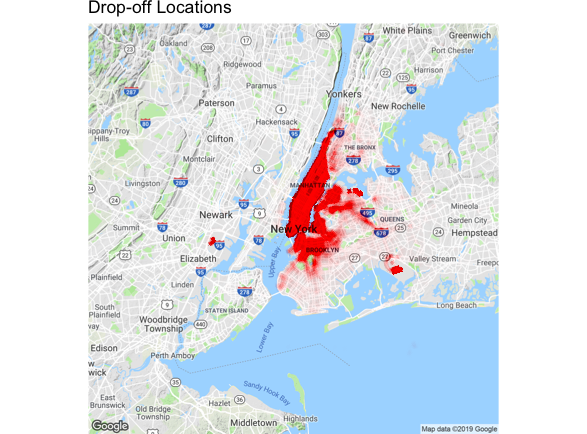
First, we take a look at the data summary for each column. We can see the min. drop-off date time is smaller than the min. pick-up date time which means there are invalid records. There are also data with longitude/latitude of 0 which should be excluded from geolocation analysis as well.



**General popular pick-up and drop-off locations**

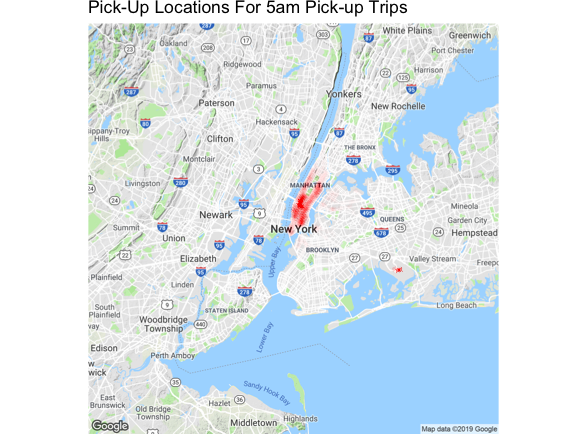
The most popular pick-up locations are Manhattan, Brooklyn Northwest, JFK airport (bottom right on the map) and LaGuardia airport (top center on the map). There is a larger spread of popular drop-off locations, including all popular pick-up locations and Newark airport (bottom left on the map).

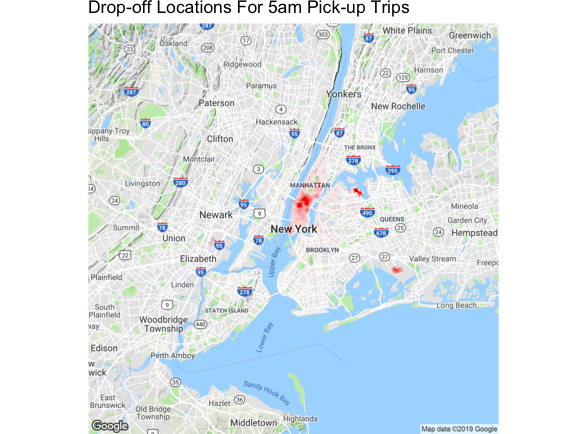


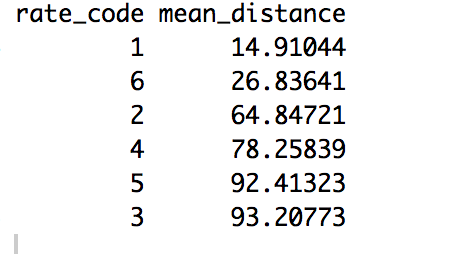


**Pick-up and drop-off locations for trips with pick-up time at 5am (longest distance)**

Areas with the most number of pick-ups are Manhattan (near New York Time Square) and JFK airport. Similarly, areas with the most number of drop-offs are Manhattan (near New York Time Square), JFK airport, LaGuardia Airport and Newark airport. Trips from and to the airports usually have longer trip distance, as seen by the table below that average trip distance is longer for rate code 2(JFK) and 3(Newark). This explains the trend we have seen earlier that average trip distance is the longest for trips with pick-up time around 5am.

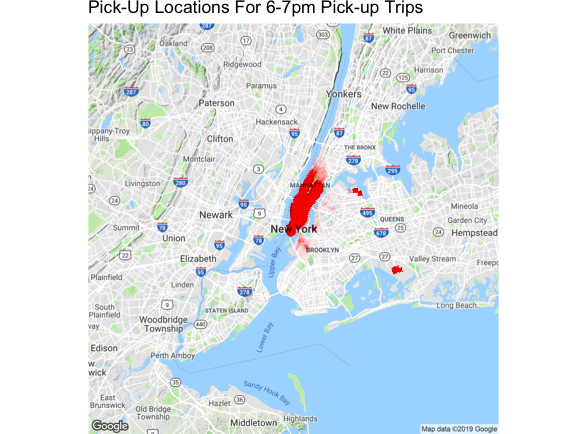


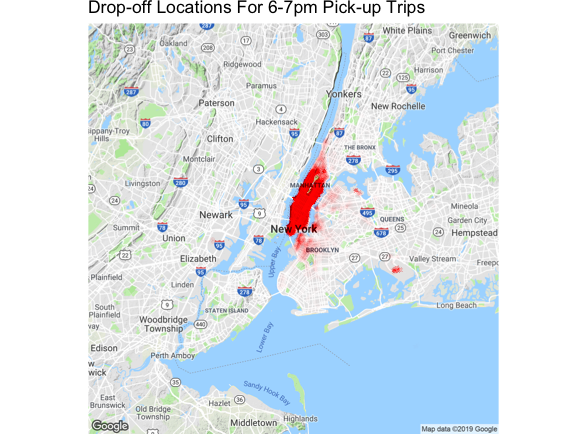




**Pick-up and drop-off locations for trips with pick-up time at 6-7pm (shortest distance but largest number of trips)**

Although Manhattan and JFK airports remain as popular locations, there are much more trips within Manhattan area as compared to 5am pick-up trips. Hence, the average trip distance is shorter.

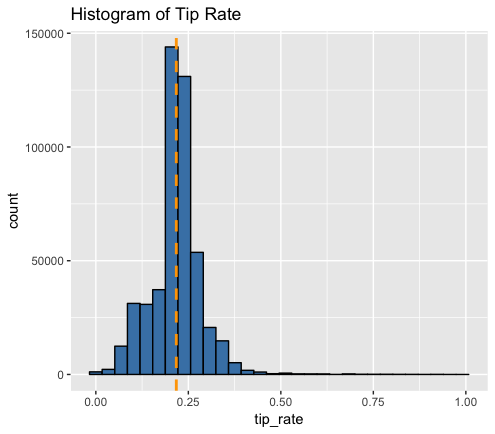


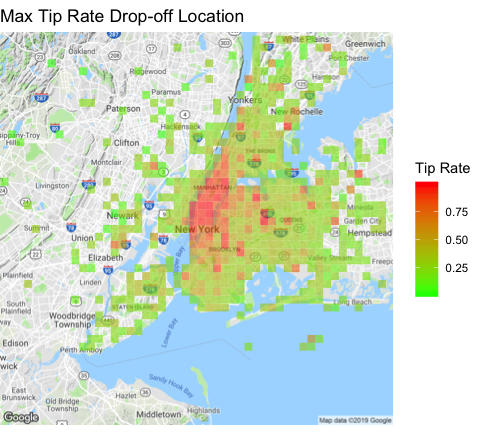
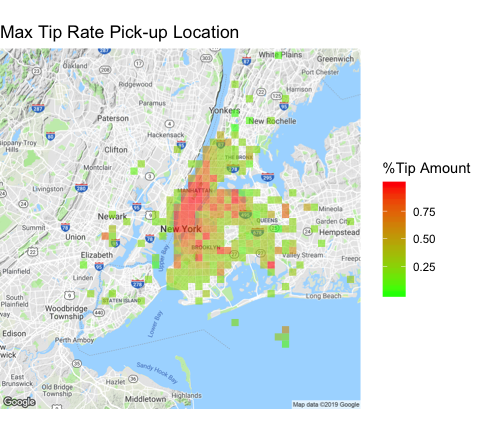


**Tip Rate Analysis**

After filtering out extreme outliers (tip rate >100%), the mean and median tip rate is about 20%. Majority of the trips have tip rate from 5% to 35%.

From the max tip rate heat map shown below, we can see trips with higher than usual tip rate (>50%) are mainly from Manhattan and airports area.

****

****