

Sta 141a Homework #3

2.



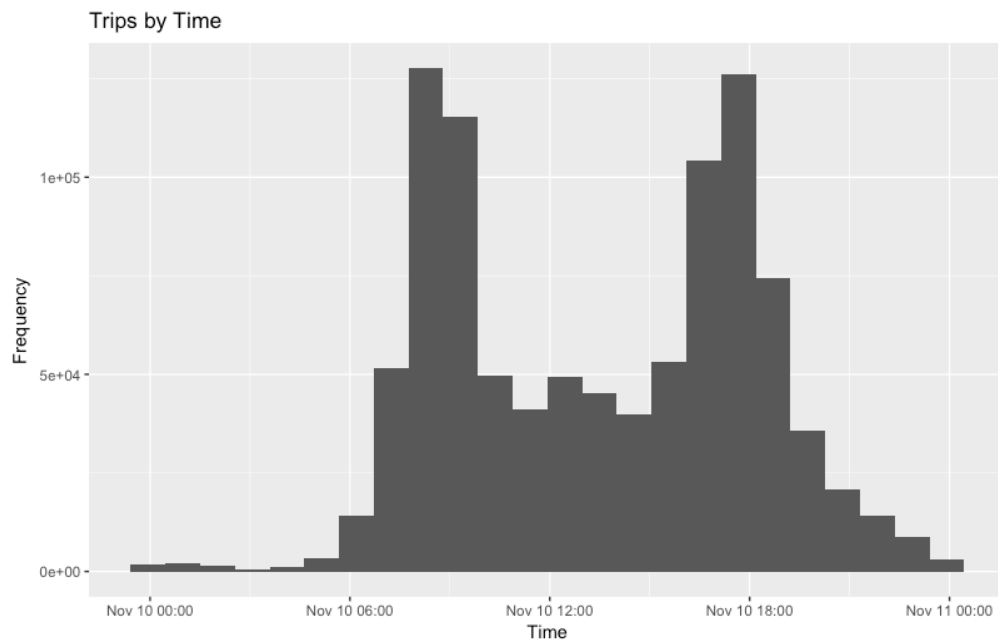
This map shows us that most of the locations of the bike share stations in San Francisco are located along the coast and along Market Street. Furthermore, the most frequent stations also appear to be along Market Street, although there are a couple stations near Mission Bay that appear to be frequent starting points. This tells us that the most densely traveled points tend to be in the middle of Downtown San Francisco, and that people living in the Mission Bay may frequently use the bikes to travel to work.

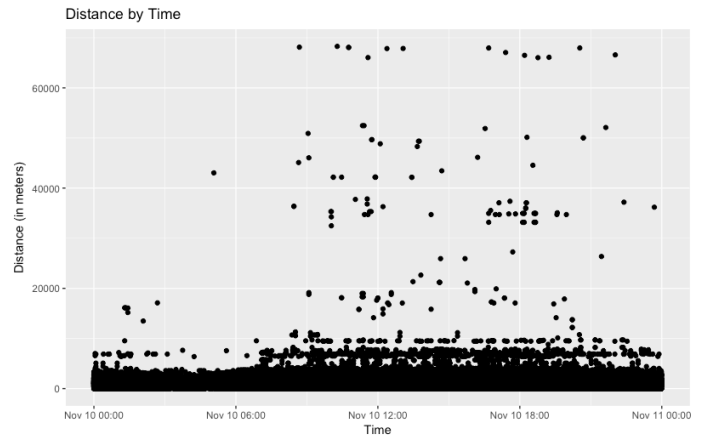
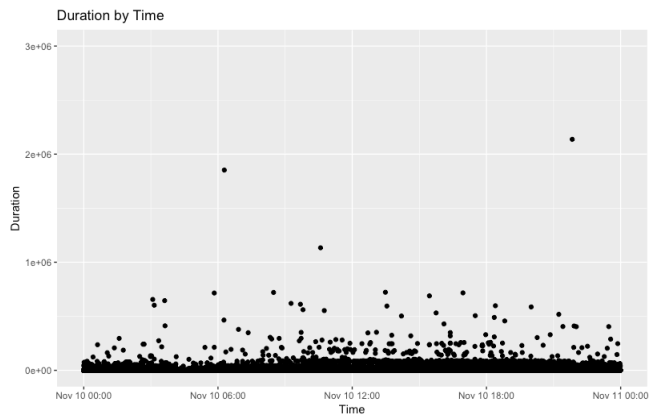
4.



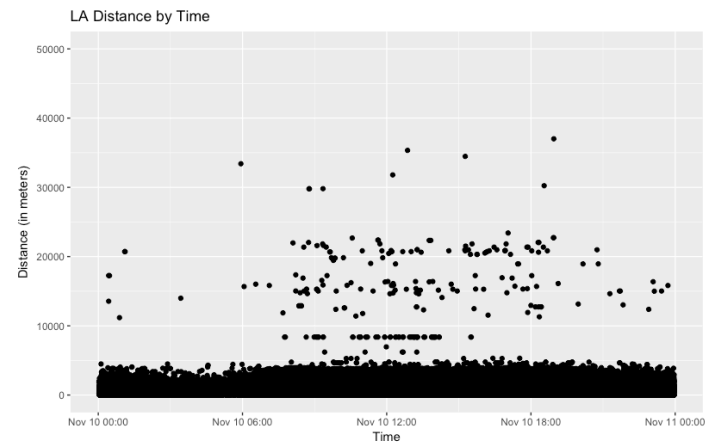
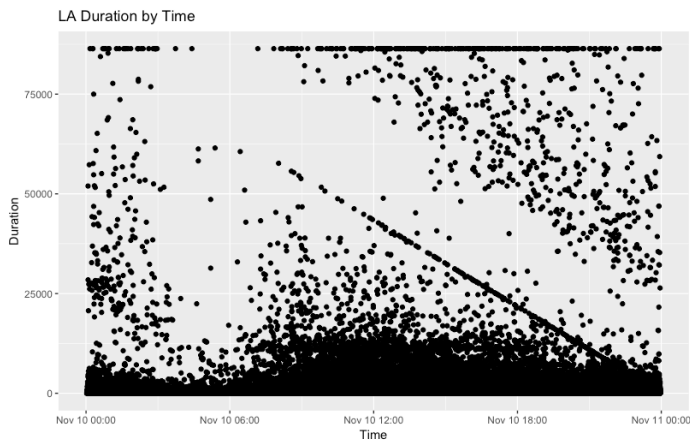
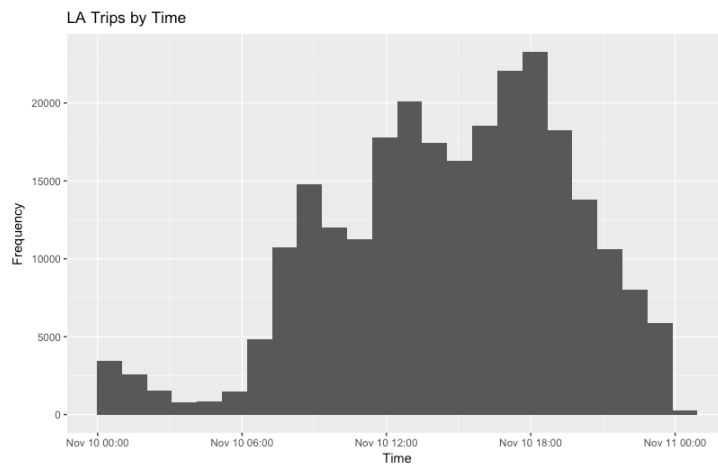
This map shows us that the most popular stations are clustered along Main Street and Broadway, in the center of Downtown Los Angeles. The stations do not appear to be located in any specific pattern, as they are spread throughout downtown Los Angeles. However, as the distance from the center increases, the popularity of the stations tends to decrease. (1), (2)

5.



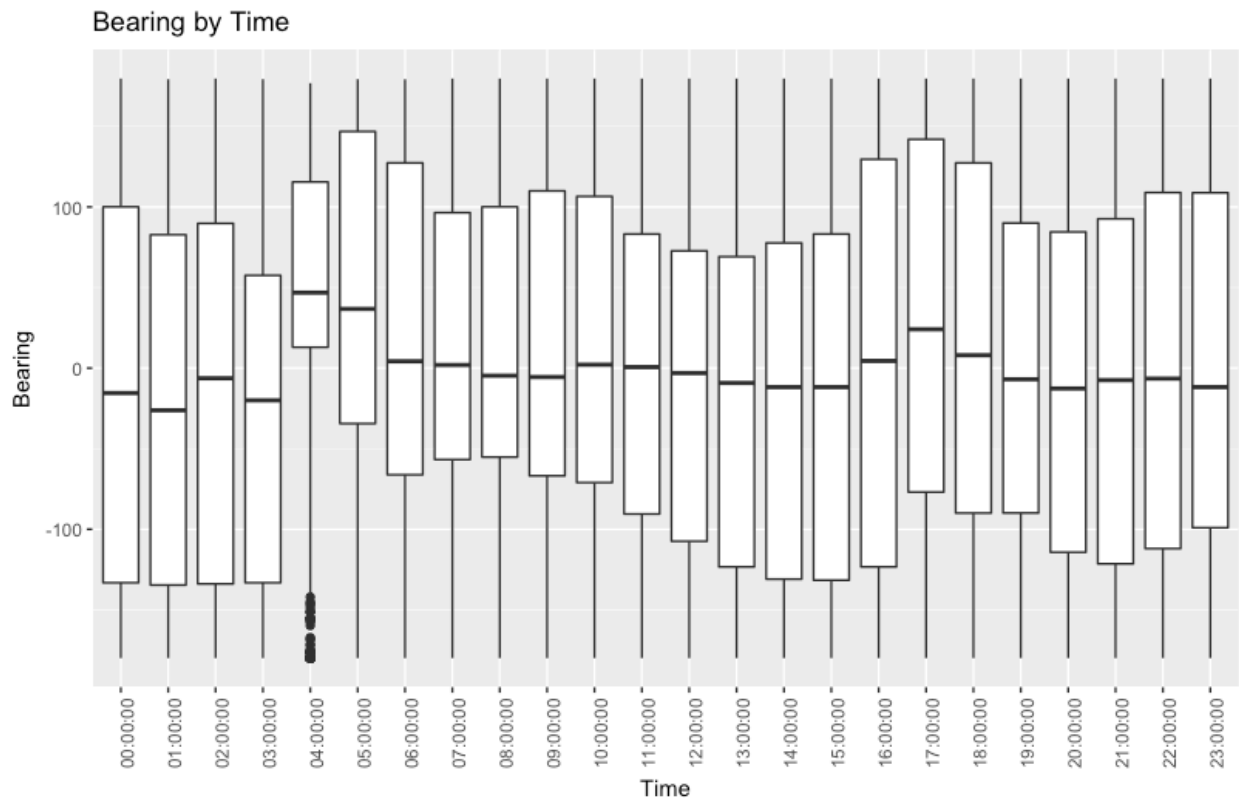


In the Bay Area, the bike stations are most frequently used between 7am and 8pm, with major spikes around 8 am and 4pm. This makes sense, as these spikes follow the expected starting and ending times for most office workers. These spikes are likely caused by people going to and from work via bicycle. There is no apparent pattern between duration and time, and there is a weak pattern in which the distance is slightly larger during the middle of the day.



In Los Angeles, much of the same trends that were seen in the San Francisco data were also followed. For example, the rush hours for bike usage followed the same pattern as the expected work schedule for most people, although the spikes were less dramatic in Los Angeles than they were in the Bay Area. Furthermore, there was no apparent pattern regarding duration, while the middle of the day had the longest trips. I noticed that the data for Los Angeles had some flaws. For example, one of the stations was a “virtual station,” so no latitude or longitude coordinates were provided. Therefore, any trips that involved the virtual station were removed from the plot. Also, one station had coordinates of 0 latitude, 0 longitude, which is obviously an error. Thus, when calculating the distance, our data would be heavily skewed and if desired, calculating an accurate regression equation would not be possible if we include this data. Also, regarding both the Bay Area and Los Angeles data, there were numerous cases of round trips, in which the distance would appear to be zero. In order to truly gauge the distance traveled, one would have to find the halfway mark, or where the person initially traveled with the bike.

6.



Although the median bearing throughout most of the day is about zero, we can still conclude a clear pattern by looking at the box plot for each hour. In the morning, the bearing tends to be positive, while as the afternoon approaches, the bearing becomes more negative. This shows us that traffic into the city is very strong in the morning, while traffic going out of the city is strongest in the afternoon, when people may be either leaving work or going to get lunch.

- 1) <https://stackoverflow.com/questions/34659910/remove-zeros-from-data-frame>
- 2) <https://stackoverflow.com/questions/32772539/in-r-merging-rows-where-a-column-has-same-value-but-different-case>
- 3) <https://stackoverflow.com/questions/36682451/rotating-x-label-text-in-ggplot>