**Homework Assignment 3**

**STA 141A A02**

**Submitted By:**

**Que1.** Write a function that loads the Bay Area bike share trip data from a CSV file, converts the columns to appropriate data types, and then saves the tidied data frame to an RDS file. Your function should have arguments to set the path for the input CSV file and the output RDS file. Write a second function that does the same thing for the Bay Area bike share station data.

**Solution 1:**The functions takes two arguments i.e. first is the path of the input CSV file and next the path of RDS file where the data is to be stored. This data is to be stored into the RDS file after proper conversion of the predefined columns. Like conversion of some of the columns include converting "start date" , "end date" to the corresponding date format. As well as converting "subscriber type" to factor.

**Que2.** Create a map that shows the locations of the Bay Area bike share stations in San Francisco (only). Label each station with its name. Make the size of each point correspond to the number of trips started from that station. Discuss what you can conclude from the map.

**Solution2:** Based on the map and calculation Ferry Building Station, Caltrain Station are few among the most popular stations in San Francisco. All of these except Embarcadero at Sansome are next to major public transit hubs for Caltrain, BART or buses. It shortest leg of their journey. Popularity of Embarcadero is due to tourist, as Embarcadero is a tourist destination.

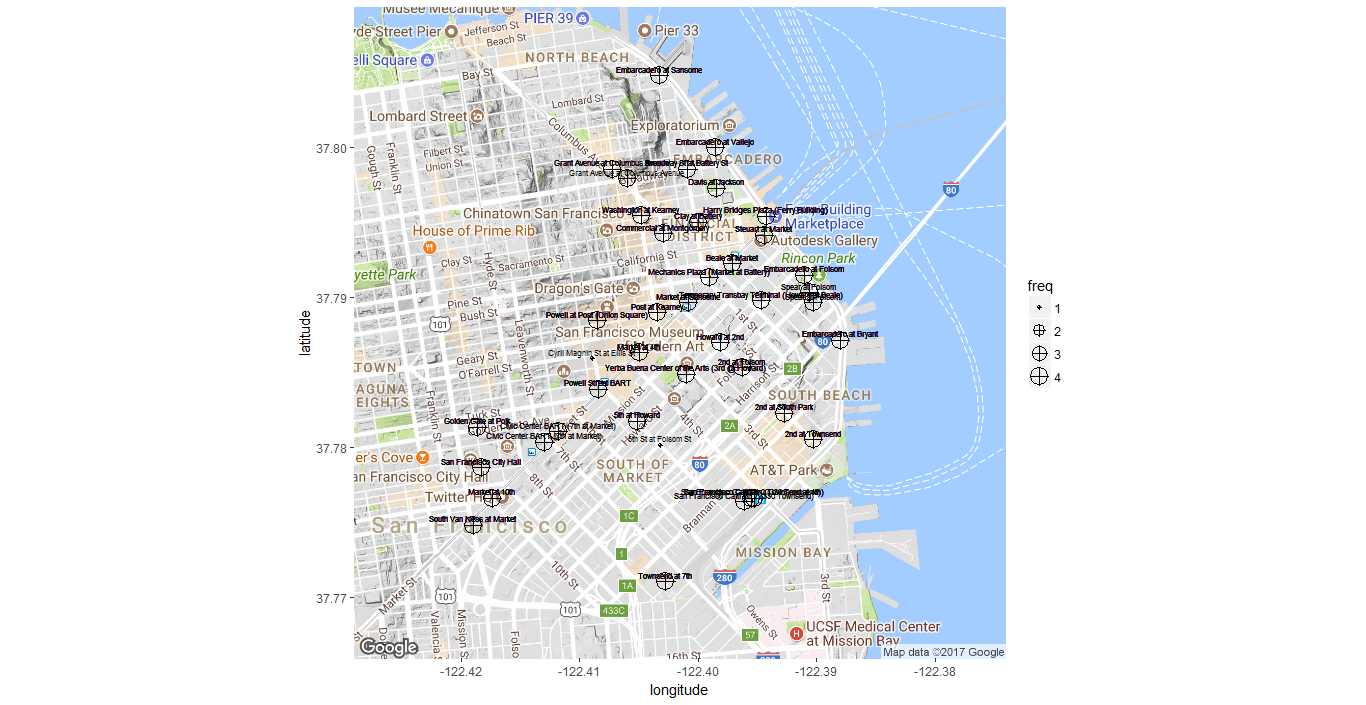


Figure 1:Stations name with their location in San Francisco

**Que3.** Write a function that loads the Los Angeles bike share trip data from the 5 provided CSV files, binds them into one data frame, converts the columns to appropriate data types, and saves the tidied data frame to an RDS file. Your function should have arguments to set the path for the input directory and the output RDS file. Keep your function short and simple by using an apply function rather than repeating code. Write a second function that loads, tidies, and saves the Los Angeles bike share station data.

**Solution3:**This function for Los Angeles bike share trip takes six arguments i.e. first five are is the path of the input CSV file and next the path of RDS file where the data is to be stored. This data is to be stored into the RDS file after proper conversion of the predefined columns. Like conversion of some of the columns include converting "start date" , "end date" to the corresponding date format. As well as converting "passholder\_type" and "trip\_route\_category" to factor.

This function for Los Angeles bike share station takes two arguments i.e. first is the path of the input CSV file and next the path of RDS file where the data is to be stored. This data is to be stored into the RDS file after proper conversion of the predefined columns. Like conversion of some of the columns include converting "start date" , "end date" to the corresponding date format. As well as converting "region " and "status" to factor.

**Que4.** Create a map that shows the locations of the Los Angeles bike share stations near downtown Los Angeles (only). Label each station with its name. Make the size of each point correspond to the number of trips started from that station. Discuss what you can conclude from the map.

**Solution4.** Based on the map and calculation 7th and Flower Station, Grand and 7th Station are few among the most popular stations in Los Angeles. There is a huge density of stations near the locality of Preshing Square, Gammy Museum and China town where as less stations on the outskirts of the city.

The region for the popularity for Gammy Museum is due to the presence of tourists as is the most popular tourist place.

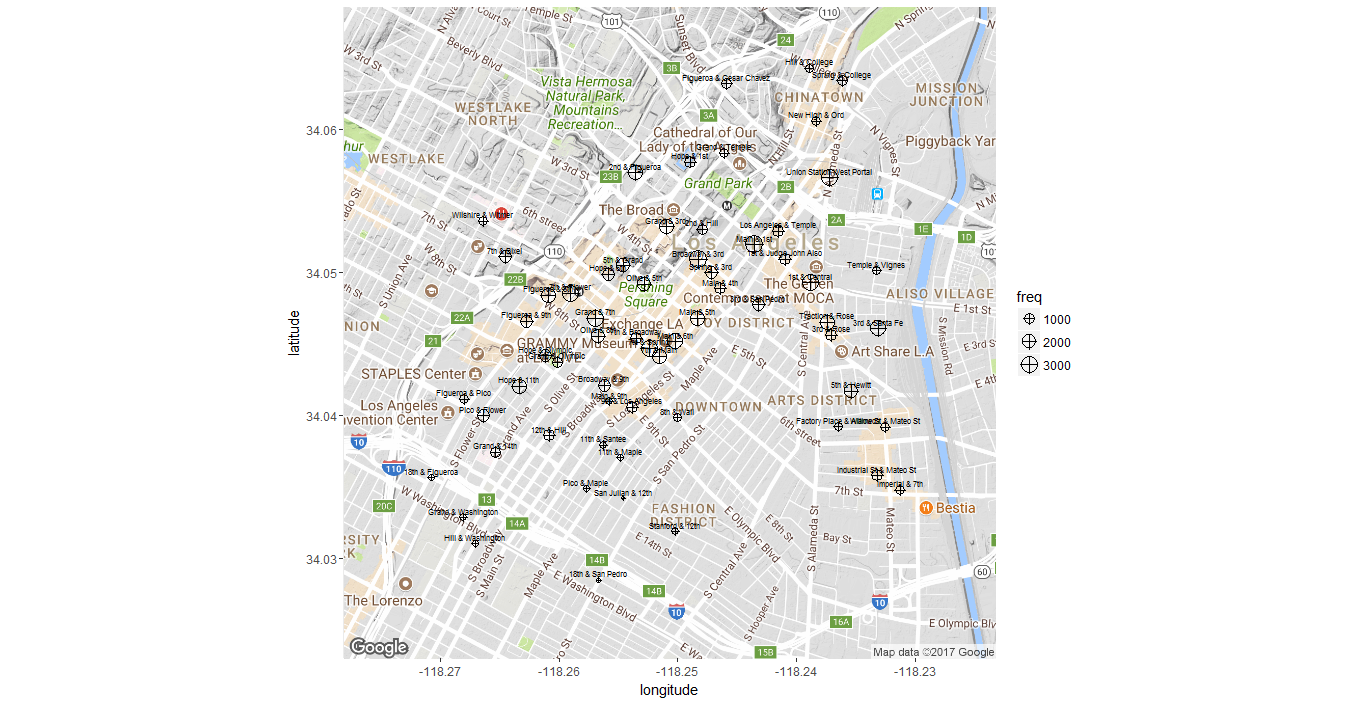


Figure 2: Station names and their location in Los Angeles

**Que5.** How do trip frequency, distance, and duration change at different times of day? Investigate for both the Bay Area bike share and the Los Angeles bike share. Compare your findings. The geosphere::distGeo()1 function can compute distances for longitude and latitude coordinates.

**Solution5.**

Based on the graph we can conclude on comparing the medians for the box plot in "frequency" vs. "Time of day" for Los Angles so more trip frequency can be observed at night then in morning etc. This is due to the reason people in that area goes to their office and comes from their during that particular time of the day.

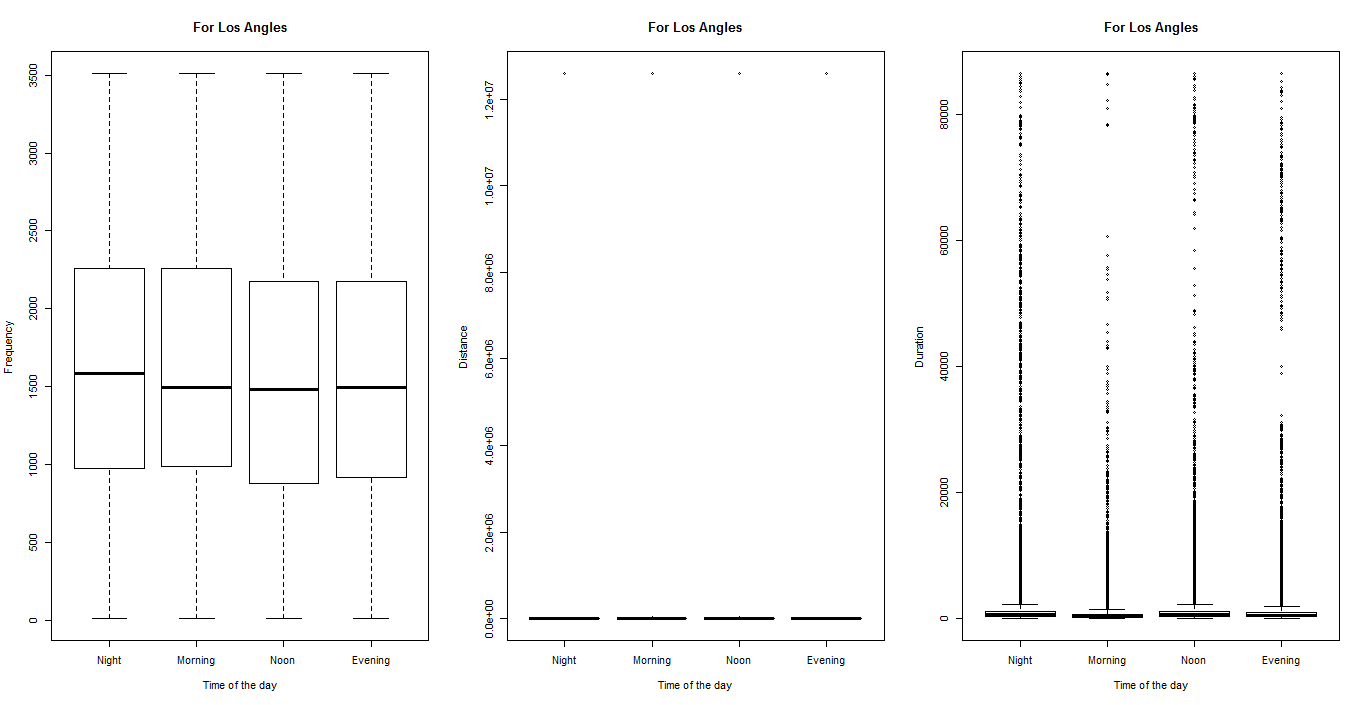
****

Figure 3: Distance,Duration,Frequency variation with Time of the day for Los Angeles

The another box plot tells about the "distance covered" during different "Time of the Day" as seen from the graph it remains almost constant.

The 3rd box plot tells about the duration between which trip happens. For night time as it is spread throughout the duration which states people like doing trip through night as most of them are free.

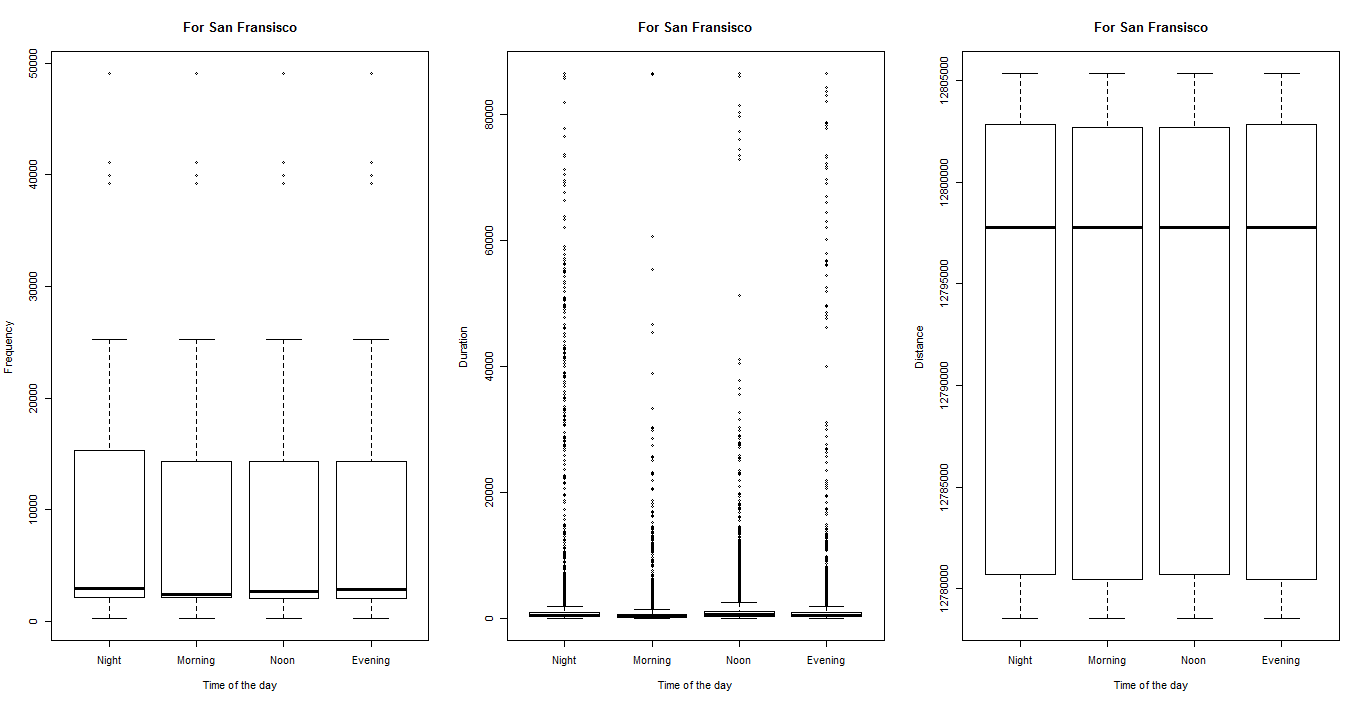
****

Figure 4:Distance,Duration,Frequency variation with Time of the day for San Fransisco

Based on the graphs above for "Frequency" and "Time of the day" the San Francisco region same as for Los-Angeles people prefer doing trips during night and evening times.

For the another graph i.e. "Duration" v/s "Time of the day" this is almost same for different time of the day. More outliers at night.

As well as same distance covered during the trip at different times of the day.

But trip frequency in San Francisco is less as compared to Los Angeles.

**Que6.** For Bay Area bike share trips in San Francisco, how does bearing (angle) change at different times of day? What can you conclude about traffic patterns in the city? The geosphere::bearing() function can compute bearings for longitude and latitude coordinates.

**Solution6:**As there is a very high density of data present in for bearing angle 0 that means if north is considered as the direction for bearing angle zero. A huge amount of traffic is directed towards north during the night time. That means most of the people doing trip during night go towards north side. Reason for this might be most of the people homes and offices are situated in northern part of the city.

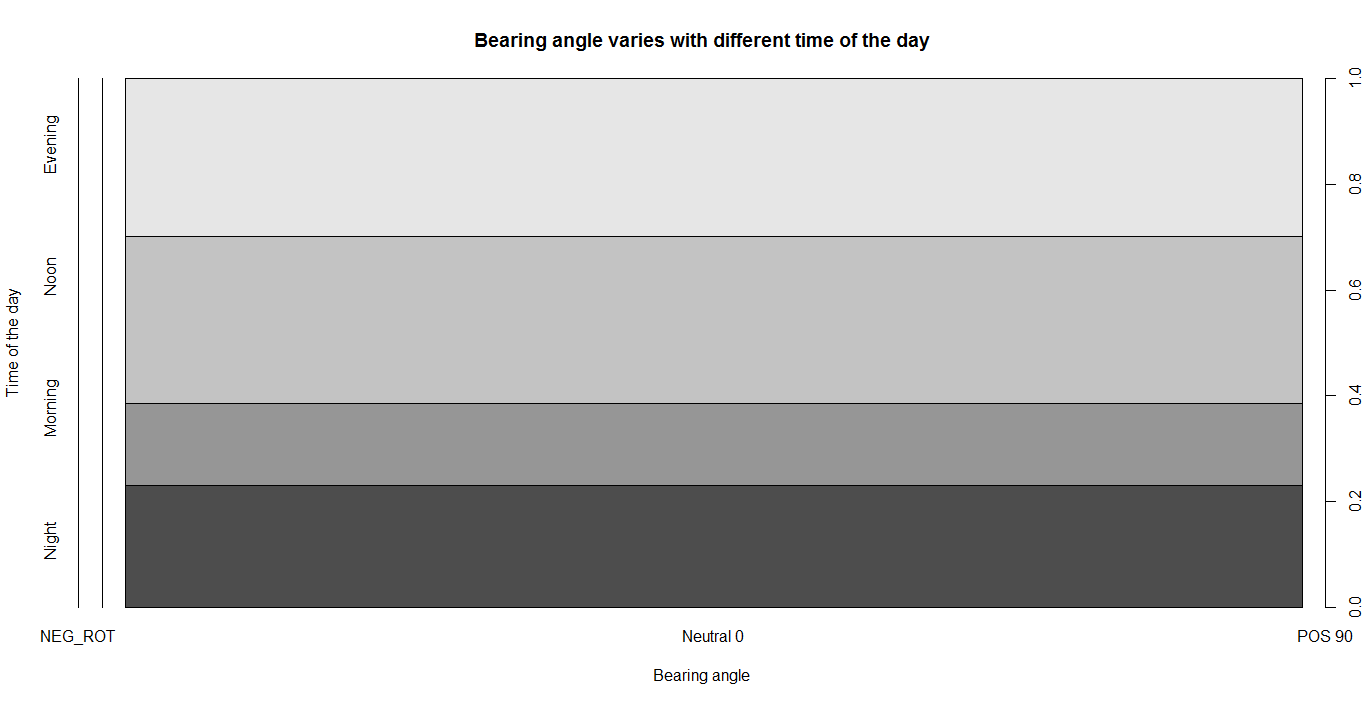
****

Figure 5: Bearing angle variation with different time of the day for San Francisco.

**References: 1.)Wikipedia:** http://ggplot2.org/

**2.)Piazza:** Que2 how to resolve data loss due to zoom

**3.)Date and time formatting:** https://www.stat.berkeley.edu/~s133/dates.html