Capstone Project

Raghav Raheja

February 2, 2019

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
library(igraph)
```

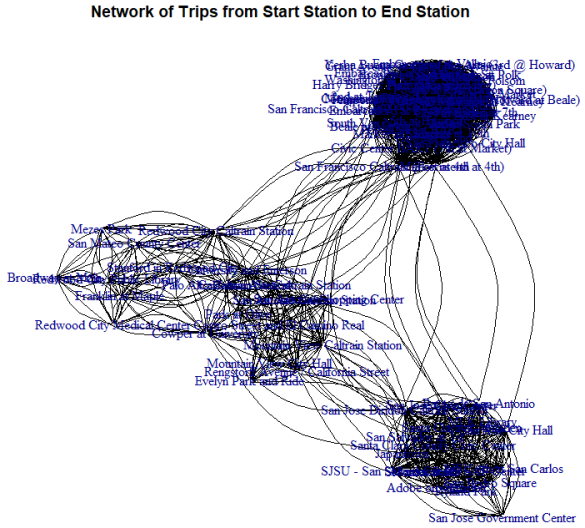
```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

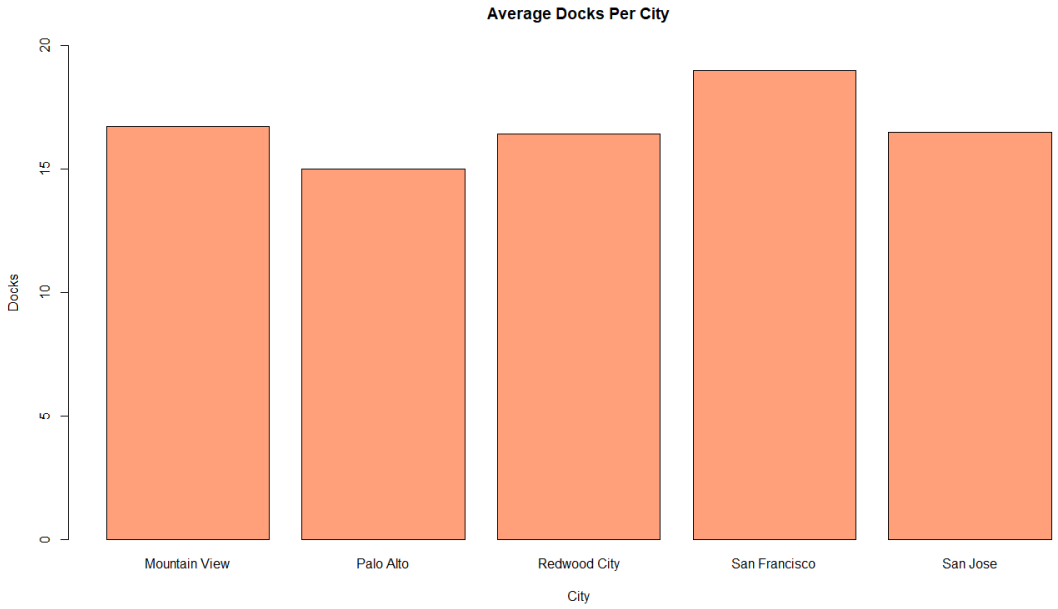
```
#View(Station)

file2 <- "C:\\Users\\Raghav\\Downloads\\Bike Data\\trip.csv"
Trips <- read.csv(file2)
#View(Trips)

Trips <- Trips[!is.na(Trips$start_station_name),]
Trips <- Trips[!is.na(Trips$end_station_name),]

# Network Between Start to End Station

edgelist <- as.matrix(Trips[c("start_station_name", "end_station_name")])
g <- graph_from_edgelist(edgelist, directed = TRUE)
g <- simplify(g)
plot.igraph(g,
            edge.arrow.size=0,
            edge.color="black",
            edge.curved=TRUE,
            edge.width=1,
            vertex.size=2,
            vertex.color=NA,
            vertex.frame.color=NA,
            vertex.label=V(g)$name,
```

```
            vertex.label.cex=1,
            layout=layout.fruchterman.reingold,
            main = "Network of Trips from Start Station to End Station"
)


# The Network below shows that there are two clear and two closely
interlinked clusters


# Docks Per City
library(ggplot2)
```

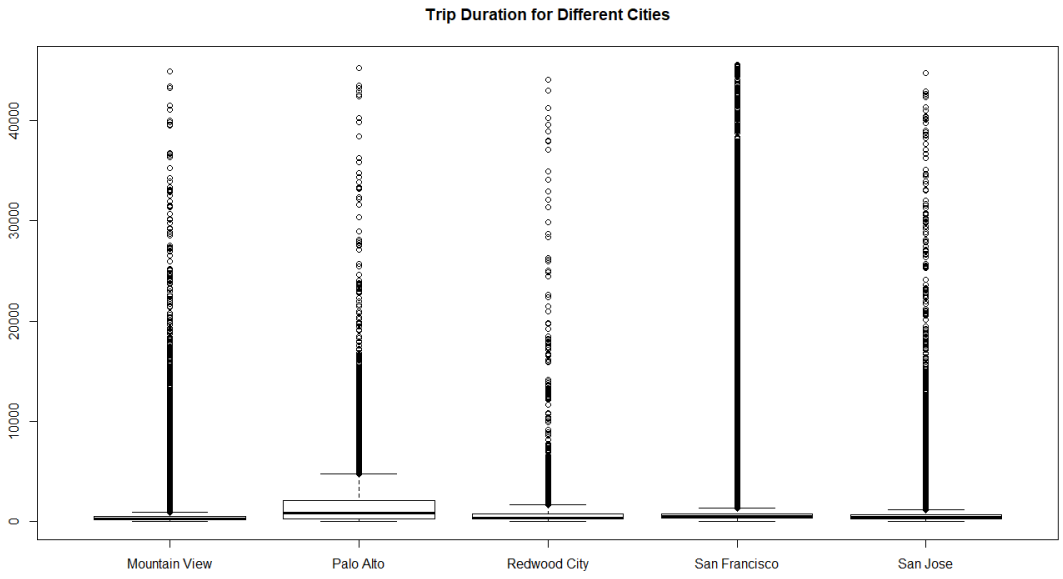



```
#library(maptools)
library(plyr)
library(knitr) # for the kable() function, which prints data frames as
tables:

## Warning: package 'knitr' was built under R version 3.5.2

# Plot 2

#file <-
Station <- read.csv('Station.csv')

#no_of_docks <- data.frame(Average_Dock =
tapply(Station$dock_count,Station$city, mean))

no_of_docks = ddply(Station, "city", summarize, Average_Dock =
mean(dock_count))
```

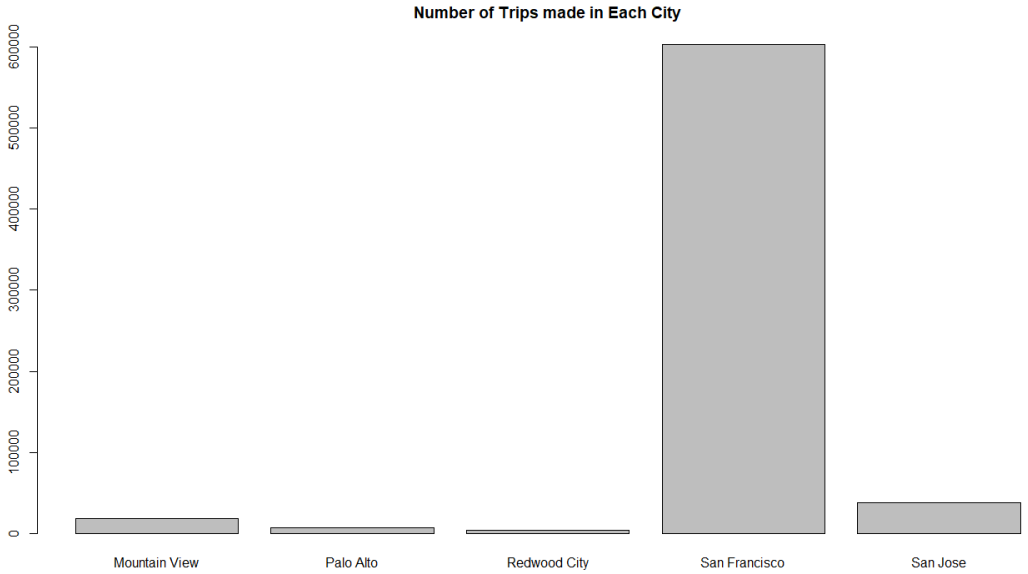
```r
#par(mar = c(2,2,2,2))
barplot(no_of_docks$Average_Dock, names.arg = no_of_docks$city,
                                                main = "Average Docks Per City",
xlab = "City", ylab = "Docks", col = "lightsalmon"
        , ylim = c(0,20))
```



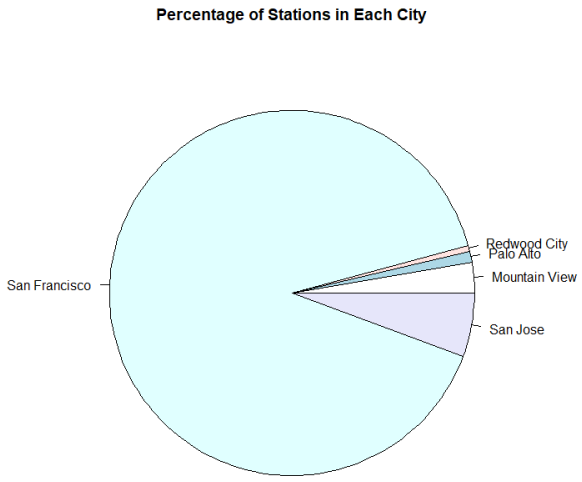

```r
# Plot 3
colnames(Trips) <-
c("Trip_id","duration","start_date","start_station_name","id","end_date","end
_station_name","end_station_id","bike_id",
"subscription_type","zip_code")
Station_trip <- merge(Trips,Station, By = "id")

# Removing the Outliers by removing the points from 2 deviations from the
mean
md <- mean(Station_trip$duration)
std <- sd(Station_trip$duration)
lower.l <- md - 2*std
higher.l <- md + 2*std
Station_trip <- Station_trip[-(which(Station_trip$duration<lower.l |
Station_trip$duration >higher.l)),]

# Plotting Boxplot for Trip Duration in Each City
options(scipen = 999)
boxplot(Station_trip$duration~Station_trip$city
        , main = "Trip Duration for Different Cities")
```




```r
# Plot 4
tpc <- ddply(Station_trip, "city", summarize, trips_per_city = length(id))
barplot(tpc$trips_per_city
        , names.arg = tpc$city
        , main = "Number of Trips made in Each City")
```




```r
# Plot 5
no_of_stations <- ddply(Station_trip, "city", summarize, number = length(id))
no_of_stations$percentage <- no_of_stations$number/sum(no_of_stations$number)
pie(no_of_stations$percentage, labels = no_of_stations$city
    , main = "Percentage of Stations in Each City")
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.