Capstone Project

Raghav Raheja

February 2, 2019

This is an [R Markdown](http://rmarkdown.rstudio.com) 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*.

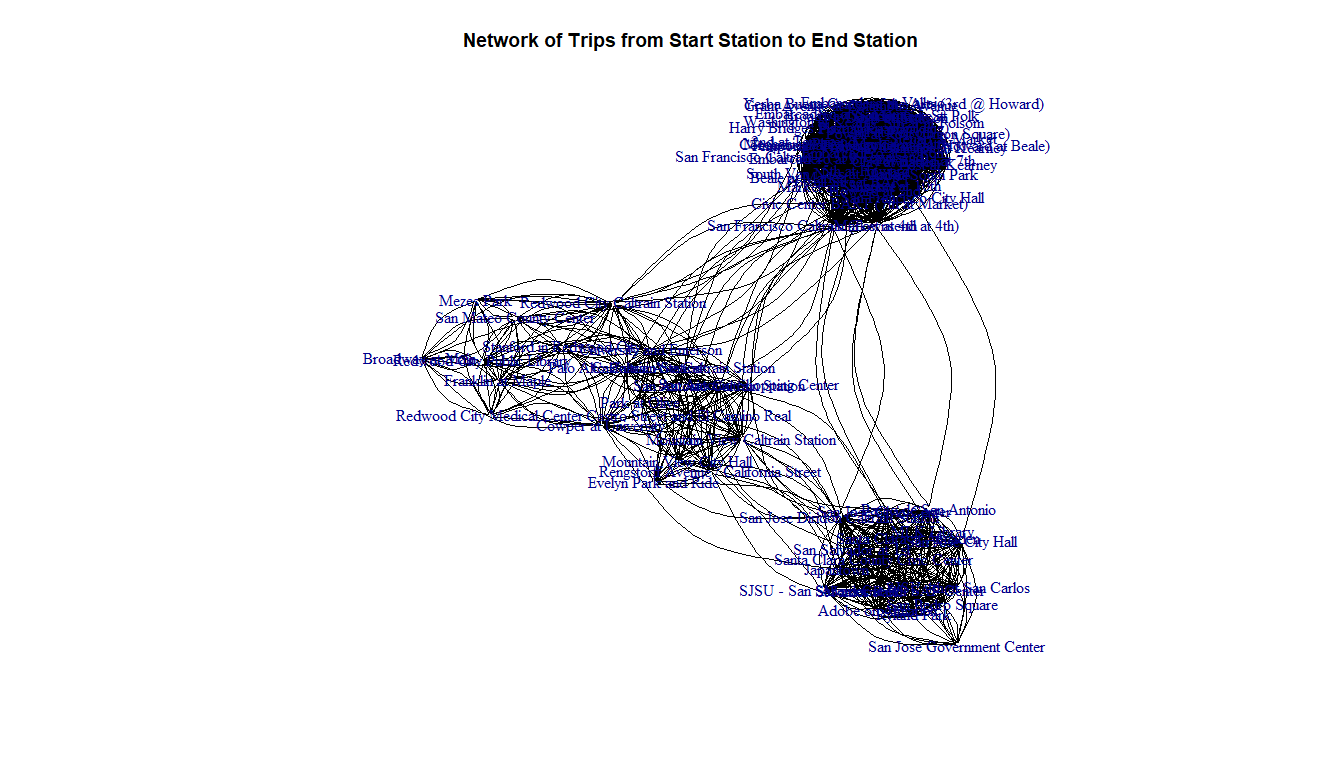
library(igraph)

##   
## Attaching package: 'igraph'

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

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

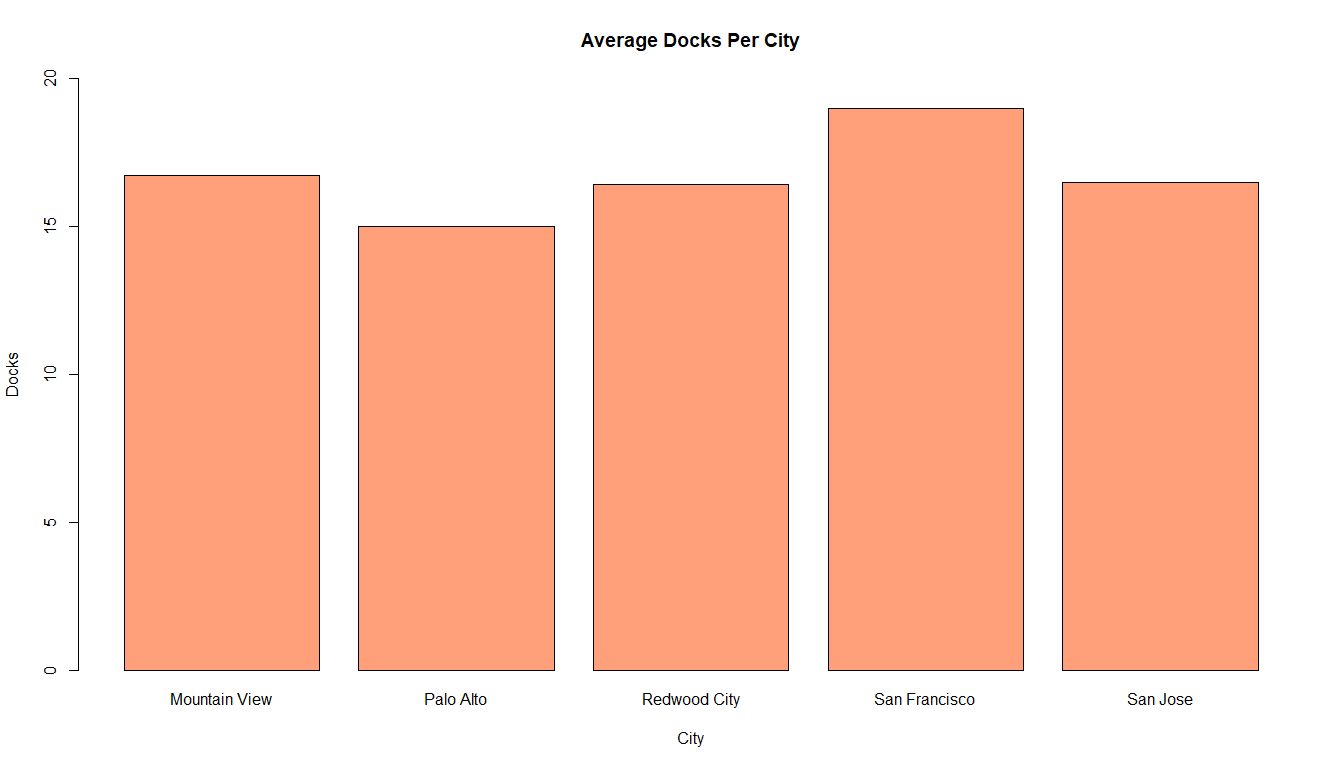
#View(Station)  
  
file2 <- "C:\\Users\\Raghav\\Downloads\\Bike Data\\trip.csv"  
Trips <- read.csv(file2)  
#View(Trips)  
  
Trips <- Trips[!is.na(Trips$start\_station\_name),]  
Trips <- Trips[!is.na(Trips$end\_station\_name),]  
  
# Network Between Start to End Station  
  
edgelist <- as.matrix(Trips[c("start\_station\_name", "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)



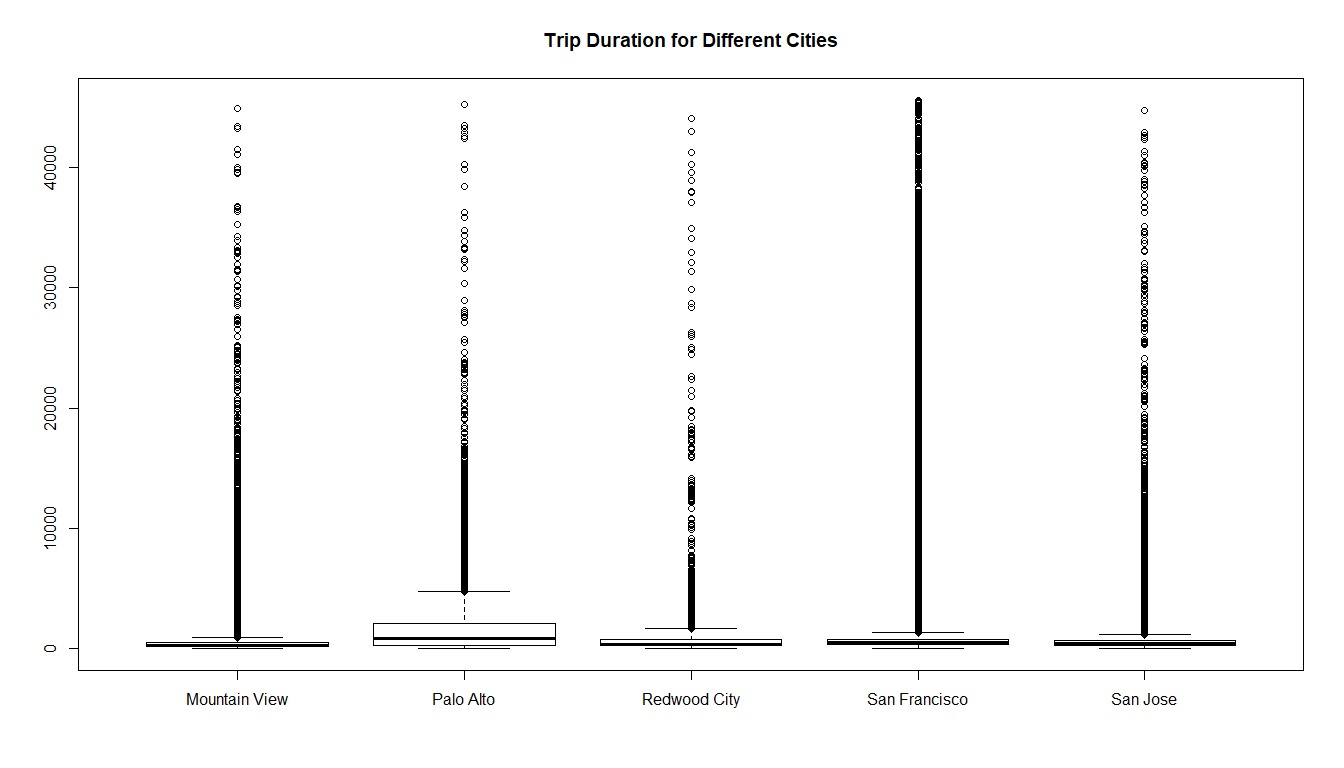
#library(maptools)  
library(plyr)  
library(knitr) # for the kable() function, which prints data frames as tables:

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

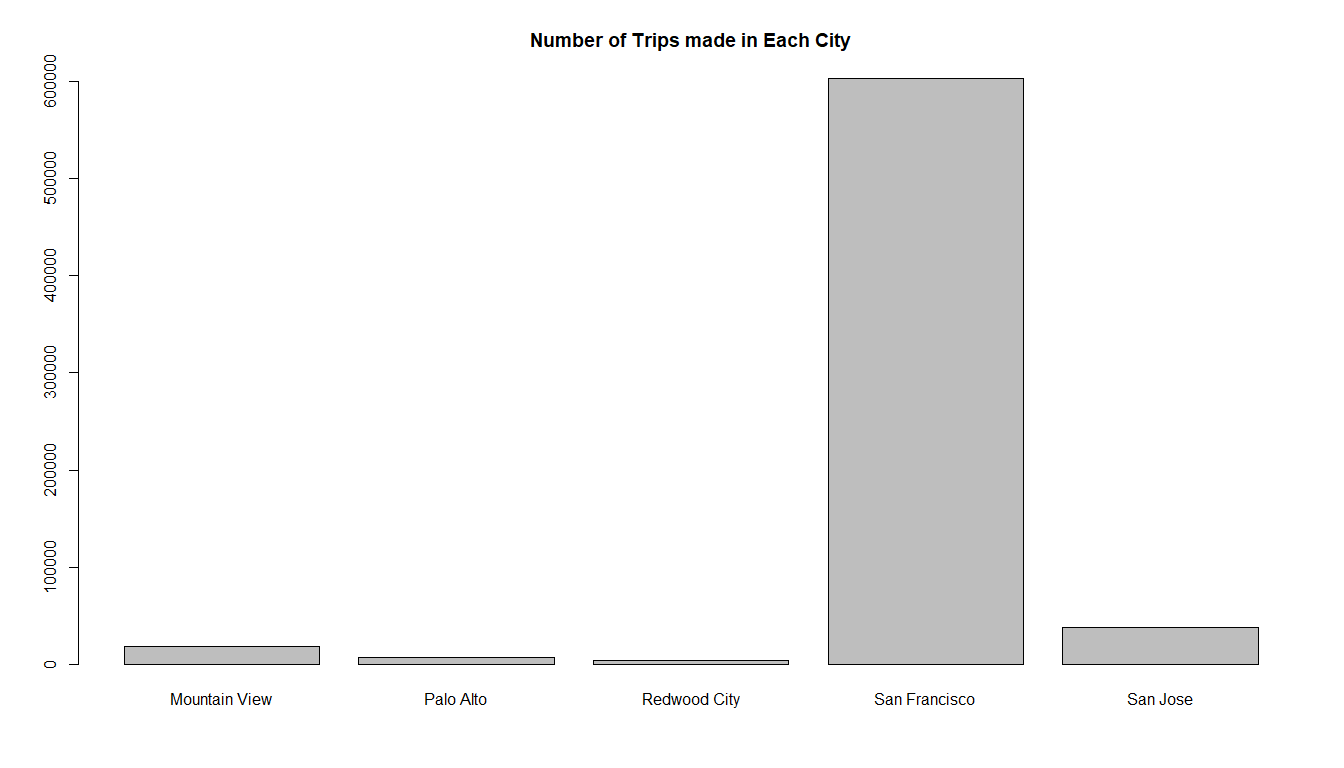
# Plot 2  
  
#file <-   
Station <- read.csv('Station.csv')  
  
#no\_of\_docks <- data.frame(Average\_Dock = tapply(Station$dock\_count,Station$city, mean))  
  
no\_of\_docks = ddply(Station, "city", summarize, Average\_Dock = mean(dock\_count))  
  
#par(mar = c(2,2,2,2))  
barplot(no\_of\_docks$Average\_Dock, names.arg = no\_of\_docks$city,  
 main = "Average Docks Per City", xlab = "City", ylab = "Docks", col = "lightsalmon"  
 , ylim = c(0,20))



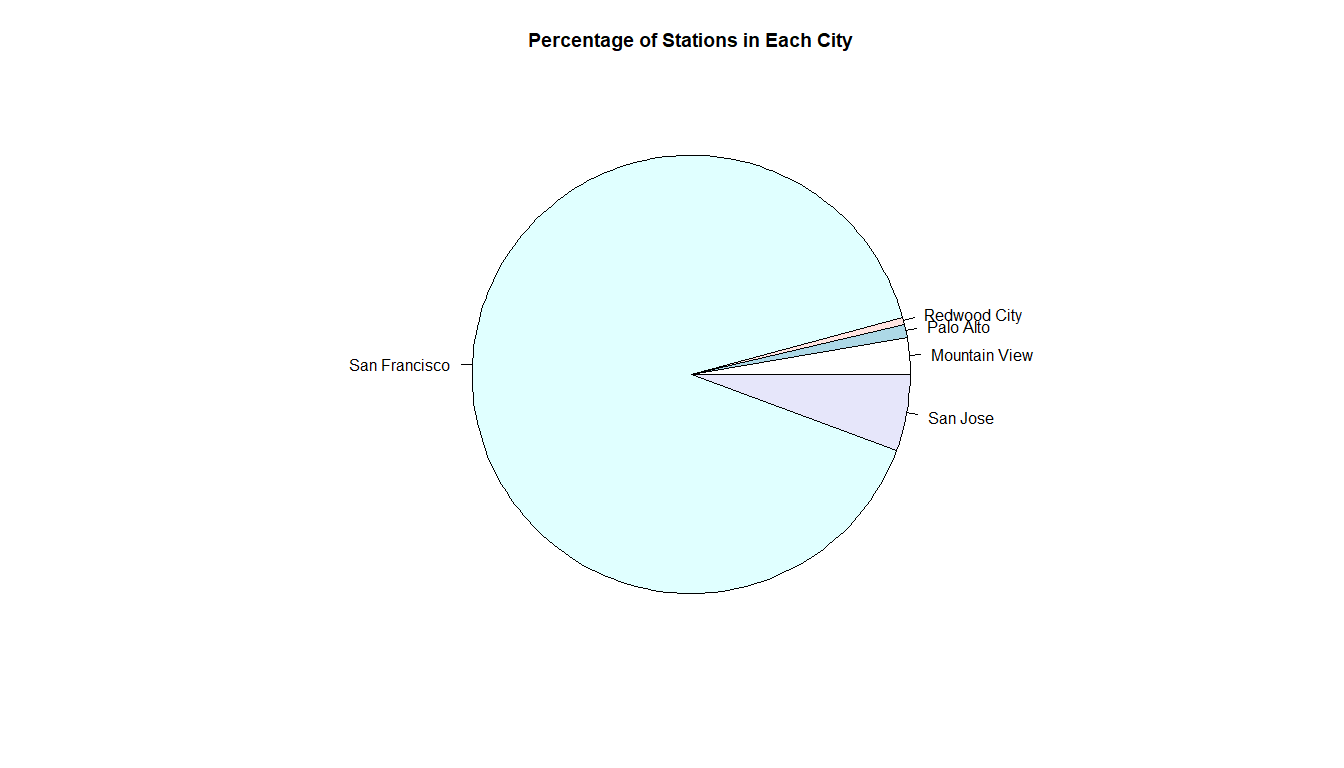
# Plot 3  
colnames(Trips) <- c("Trip\_id","duration","start\_date","start\_station\_name","id","end\_date","end\_station\_name","end\_station\_id","bike\_id", "subscription\_type","zip\_code")  
Station\_trip <- merge(Trips,Station, By = "id")  
  
# Removing the Outliers by removing the points from 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")



# 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")



# 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.