COVID-19 Global Mobility

Preliminary Data Cleaning

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
data <- read.csv(file = "Global_Mobility_Report.csv", as.is = TRUE)
data <- select(data, -3:-7)

colnames(data)[1]  <- "Country Code"
colnames(data)[2]  <- "Country"
colnames(data)[4]  <- "Retail_and_Recreation_Percent_Change"
colnames(data)[5]  <- "Groccery_and_Pharmacy_Percent_Change"
colnames(data)[6]  <- "Parks_Percent_Change"
colnames(data)[7]  <- "Transit_Percent_Change"
colnames(data)[8]  <- "Work_Percent_Change"
colnames(data)[9]  <- "Residential_Percent_Change"

data$date <- gsub("2020-", "", data$date)
data$month <- gsub("02-.*", "2", data$date)
data$month <- gsub("03-.*", "3", data$month)
data$month <- gsub("04-.*", "4", data$month)
data$month <- gsub("05-.*", "5", data$month)
data$month <- gsub("06-.*", "6", data$month)
data$month <- gsub("07-.*", "7", data$month)
data$month <- as.numeric(data$month)

data$day <- gsub("02-|03-|04-|05-|06-|07-", "", data$date)
data$day <- as.integer(data$day)

data$daysSince <- data$day + car::recode(data$month, "2 = 31; 3 = 60; 4 = 91; 5 = 121; 6 = 152; 7 = 182")

head(data)
```

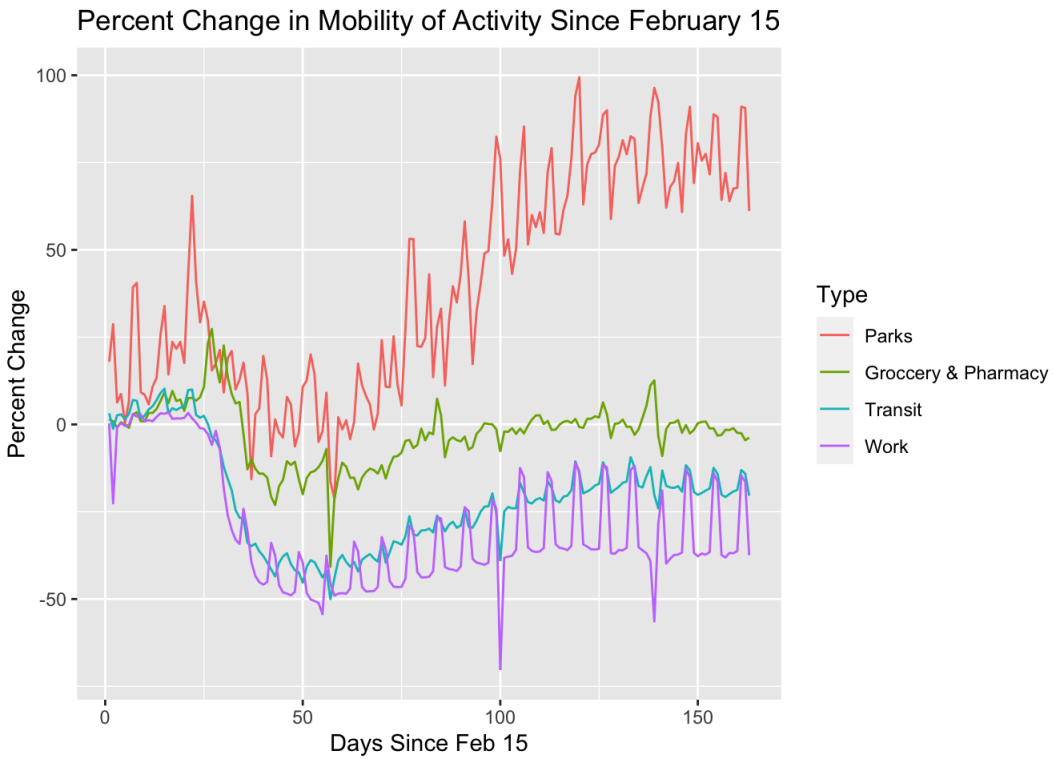
```
##   Country Code              Country  date Retail_and_Recreation_Percent_Change
## 1           AE United Arab Emirates 02-15                                    0
## 2           AE United Arab Emirates 02-16                                    1
## 3           AE United Arab Emirates 02-17                                   -1
## 4           AE United Arab Emirates 02-18                                   -2
## 5           AE United Arab Emirates 02-19                                   -2
## 6           AE United Arab Emirates 02-20                                   -2
##   Groccery_and_Pharmacy_Percent_Change Parks_Percent_Change
## 1                                   4                    5
## 2                                   4                    4
## 3                                   1                    5
## 4                                   1                    5
## 5                                   0                    4
## 6                                   1                    6
##   Transit_Percent_Change Work_Percent_Change Residential_Percent_Change month
## 1                      0                   2                           1     2
## 2                      1                   2                           1     2
## 3                      1                   2                           1     2
## 4                      0                   2                           1     2
## 5                     -1                   2                           1     2
## 6                      1                   1                           1     2
##   day daysSince
## 1  15        46
## 2  16        47
## 3  17        48
## 4  18        49
## 5  19        50
## 6  20        51
```

#More Data Cleaning

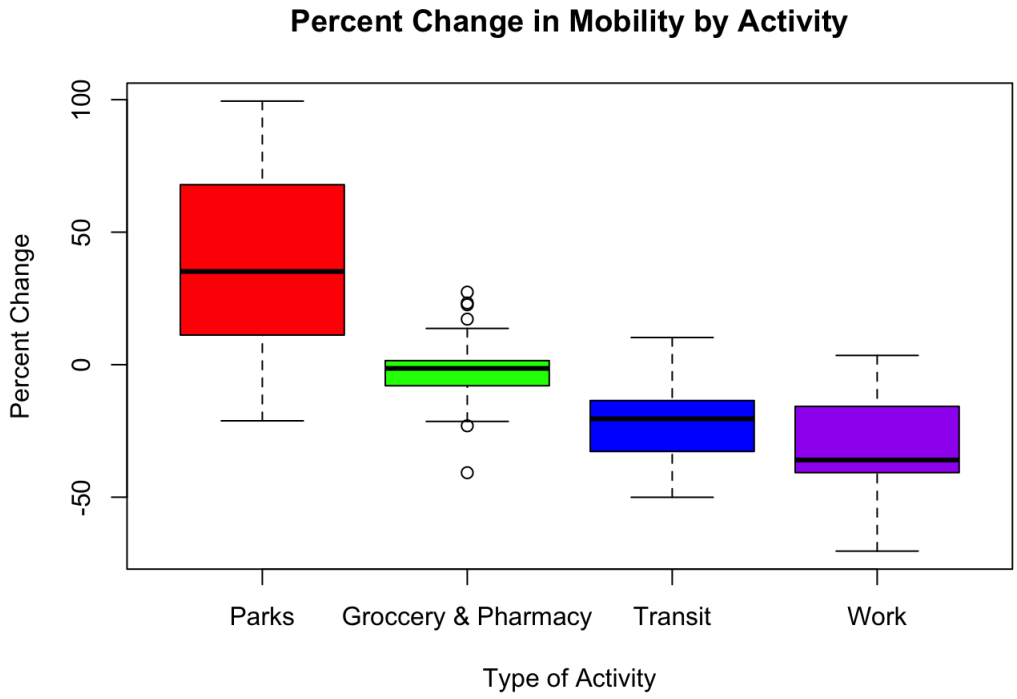
```
#Mean Percent Change in Work per Day


data <- na.omit(data)
attach(data)

work_place <- c()
for (day in 46:209) {
  work_place[day-46] <- mean(data[which(Country == "United States" & daysSince == day), "Work_Percent_Change
"])
}


transit <- c()
for (day in 46:209) {
  transit[day-46] <- mean(data[which(Country == "United States" & daysSince == day), "Transit_Percent_Change
"])
}

parks <- c()
for (day in 46:209) {
  parks[day-46] <- mean(data[which(Country == "United States" & daysSince == day), "Parks_Percent_Change"])
}

g_p <- c()
for (day in 46:209) {
  g_p[day-46] <- mean(data[which(Country == "United States" & daysSince == day), "Groccery_and_Pharmacy_Perc
ent_Change"])
}


r_r <- c()
for (day in 46:209) {
  r_r[day-46] <- mean(data[which(Country == "United States" & daysSince == day), "Retail_and_Recreation_Perc
ent_Change"])
}
```

#Data Visualization

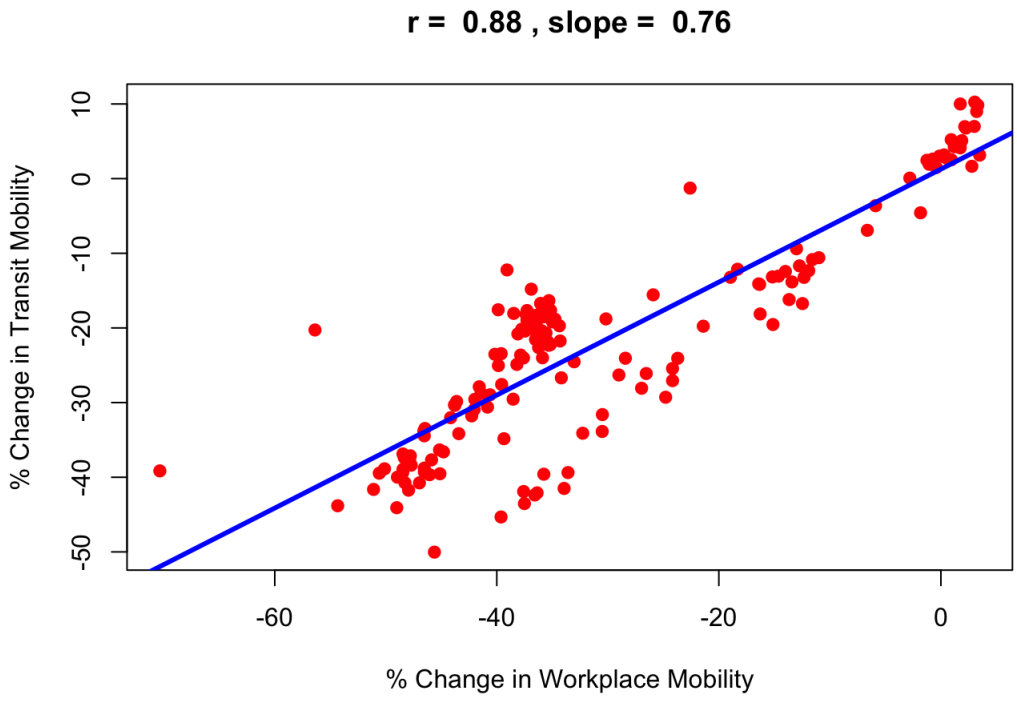
```
days_since <- 1:163

change_df <- data.frame(days_since, work_place, transit, parks, g_p)

Type <- rep("Work", 163)
Change <- work_place
work_df <- data.frame(days_since, Change, Type)

Type <- rep("Transit", 163)
Change <- transit
transit_df <- data.frame(days_since, Change, Type)

Type <- rep("Parks", 163)
Change <- parks
parks_df <- data.frame(days_since, Change, Type)

Type <- rep("Groccery & Pharmacy", 163)
Change <- g_p
gp_df <- data.frame(days_since, Change, Type)


combined <- rbind(parks_df, gp_df, transit_df, work_df)

ggplot(data = combined, aes(x=days_since, y=Change)) + geom_line(aes(colour=Type)) + xlab('Days Since Feb 15
') + ylab('Percent Change') +
ggtitle("Percent Change in Mobility of Activity Since February 15")
```



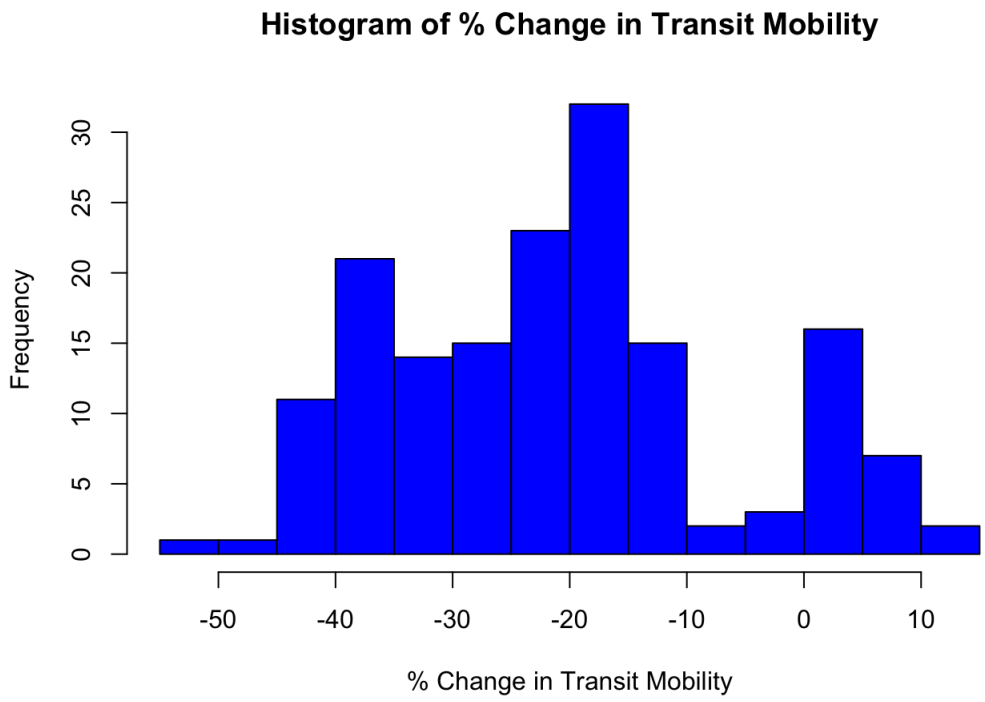

```
boxplot(Change~Type, data=combined, main="Percent Change in Mobility by Activity",
   xlab="Type of Activity", ylab="Percent Change", col=(c("red","green", "blue", "purple")))
```




```
myCor <- function(x, y){
  plot(x, y, pch = 19, col = "red")
  mtext(paste("Sample Correlation =", round(cor(x, y), 3)), cex = 1.2)
}

cor1 <- cor(work_place, transit)
lm1 <-  lm(transit ~ work_place)

plot(data.frame(work_place, transit), main = paste("r = ", round(cor1,2) , ", slope = ", round(lm1$coef[2],
2)),
     pch = 19, col = "red", xlab = "% Change in Workplace Mobility", ylab = "% Change in Transit Mobility")

abline(lm1$coef, col = "blue", lwd = 3)
```




```
hist(transit, col = "blue", breaks = 10, main = "Histogram of % Change in Transit Mobility", xlab = "% Chang
e in Transit Mobility")
```




```
hist(work_place, col = "blue", breaks = 10, main = "Histogram of % Change in Work Mobility", xlab = "% Chan
ge in Work Mobility")
```