MA415 Project One

Lin Li, Lingjia Zhang, Zirui Liu

Background

According to the CNN news, apprehensions at the US-Mexico border reached the historic low rate in April 2017, and have a continuing downward trend during the first few months of the Trump administration. The drop of the apprehension rate was interpreted as the sign of "President Donald Trump's rhetoric and aggressive push to enforce immigration laws is having an effect." Through this project, we want to analyze by using past data to estimate whether the declining rate of apprehension relates to the President Donald's new law.

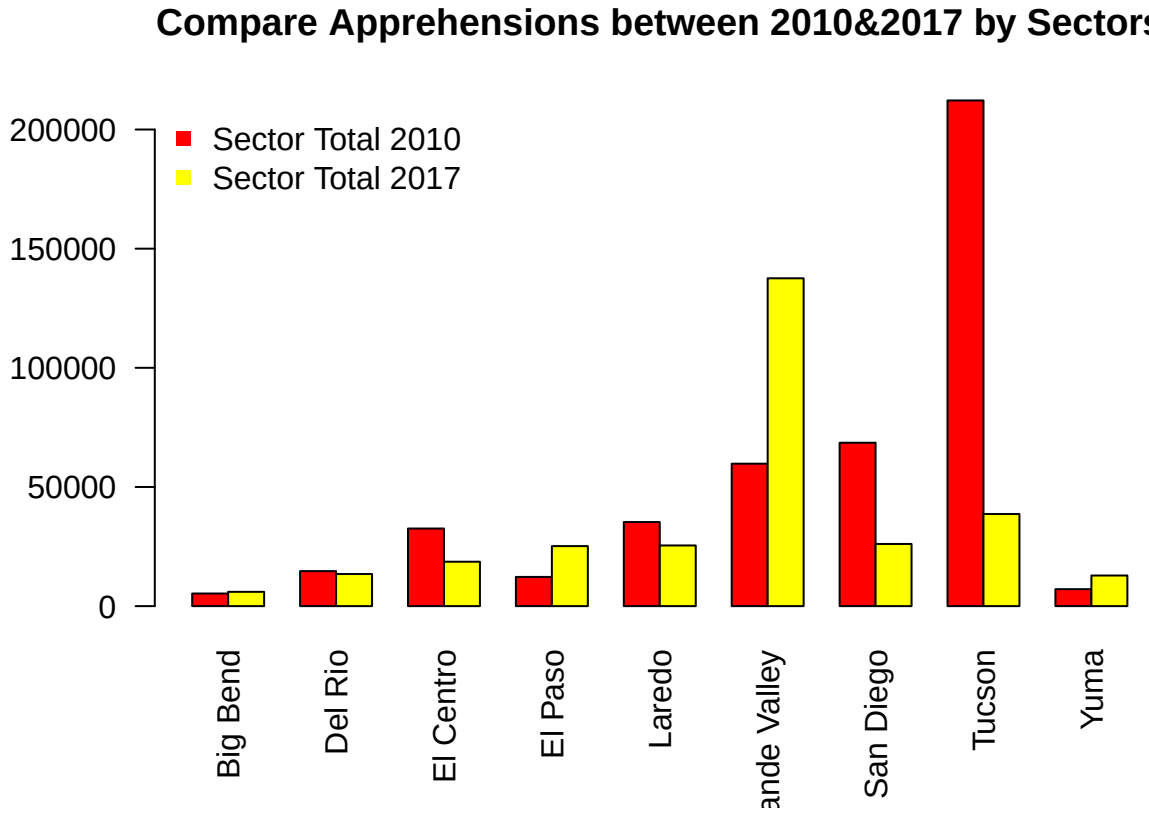
Data Visualization

```
T2010 <- read.csv("bp appre 2010.csv")
T2017 <- read.csv("bp appre 2017.csv")

rownames(T2010) <- T2010[,1]
rownames(T2017) <- T2017[,1]

# sum apprehensions of sector and month in 2010
T2010 <- subset(T2010, select = -c(Sector))
# sort month in order
T2010 <- cbind(subset.data.frame(T2010)[,4:12],subset.data.frame(T2010)[,1:3])
sum_m2010 <- colSums((T2010))
T2010 <- rbind(T2010,sum_m2010)
rownames(T2010)[10] <- "Sector Total"

sum_s2010 <- rowSums(T2010)
T2010 <- cbind(T2010,sum_s2010)
colnames(T2010)[13] <- "Month Total"

# sum apprehensions of sector and month in 2017
T2017 <- subset(T2017, select = -c(Sector))
# sort month in order
T2017 <- cbind(subset.data.frame(T2017)[,4:12],subset.data.frame(T2017)[,1:3])
sum_m2017 <- colSums((T2017))
T2017 <- rbind(T2017,sum_m2017)
rownames(T2017)[10] <- "Sector Total"

sum_s2017 <- rowSums(T2017)
T2017 <- cbind(T2017,sum_s2017)
colnames(T2017)[13] <- "Month Total"

#compare apprehensions by months
sector <- rownames(T2010)
month <- colnames(T2010)

#The function takes in the index of the sector as input
```

```r
plot_by_months <- function(i){
  year <- rbind(T2010[i,1:12],T2017[i,1:12])
  barplot(as.matrix(year), beside = T, col = c("red", "yellow"), bty="n",las=2,
          main = "Compare Apprehensions Between 2010&2017 by Months")
  legend("topright",
         c("Month Total 2010","Month Total 2017"),
         pch=15,
         col=c("red","yellow"),
         bty="n"
         )
}

#The function takes in month as input
plot_by_sectors <- function(i){
  T2010<-t(T2010)
  T2017<-t(T2017)
  year <- rbind(T2010[i,1:9],T2017[i,1:9])
  getOption("scipen")
  opt <- options("scipen" = 20)
  barplot(year, beside = TRUE, col = c("red", "yellow"), bty="n",las=2,
          main = "Compare Apprehensions between 2010&2017 by Sectors"
          )
  legend("topleft",
         c("Sector Total 2010","Sector Total 2017"),
         pch=15,
         col=c("red", "yellow"),
         bty="n")
}
```

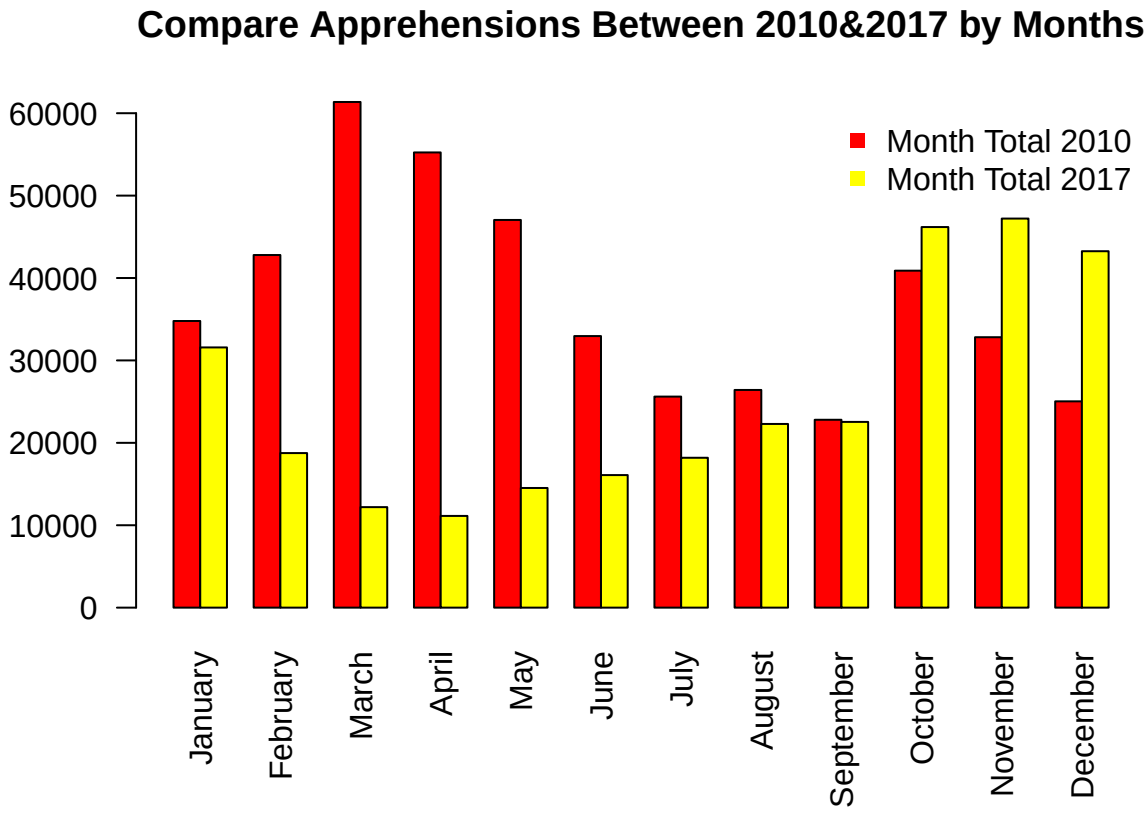
Based on csv files, we compare the apprehensions by sectors and months between 2010 and 2017. First of all, we sum up apprehensions by sectors and create the following plot.

```r
plot_by_sectors(13)
```



Second of all, we sum up apprehensions by months and create the following plot.

```
plot_by_months(10)
```



Based on these two graphs, it is easy to conclude that in 2010, Tucson has the most apprehensions. In 2017, Rio Grande Valley has the most apprehensions. In 2010, from March to May, there are most apprehensions. In 2017, from October to December, there are most apprehensions.

To illustrate the problem and see the problem in a more dynamic way, we also build a shiny app and the website is: https://eileenli.shinyapps.io/Assignment3/.

Statistic Analysis

First problem

Use simple statistical tests linke those demonstrated in class to compare the sector with the most apprehensions for 2010 and with the sector with the most apprehensions in 2017. Has there been a change in the maximum?

```
#In 2010, the sector that has most apprehensions is Tucson
most_2010 <- match(max(T2010[1:9,13]), T2010[1:9,13])
rownames(T2010)[most_2010]
```

```
## [1] "Tucson"
```

```
#In 2017, the sector that has most apprehensions is Rio Grande Valley
most_2017 <- match(max(T2017[1:9,13]), T2017[1:9,13])
rownames(T2017)[most_2017]
```

```
## [1] "Rio Grande Valley"
```

To evaluate wheter the most apprehensions between 2010(Tucson) and 2017(Rio Grande Valley),we have following null hypothesis and alternative hypothesis:

$$H_0 : \mu_{T,2010} - \mu_{R,2017} = 0$$
$$H_a : \mu_{T,2010} - \mu_{R,2017} \neq 0$$

```
##
##  Welch Two Sample t-test
##
## data:  T2010_T and T2017_R
## t = 1.9547, df = 21.973, p-value = 0.06346
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -379.5935 12819.5935
## sample estimates:
## mean of x mean of y
##   17683.5   11463.5
```

From t-test, we get p-value $= 0.06346 > 0.05$, and we fail to reject the null hypothesis. At significance level of 5%, the mean of apprehension of Tucson in 2010 is equal to the mean of apprehension of Rio Grande Valley in 2017. Thus, there is no significant difference between the maximum of apprehensions by secoter in 2010 and 2017.

Second problem

From the grpah, we know that in 2010, March, April, and May are ranking top 3. In 2017, November, October, and December are ranking top 3. The statistic analysis we want to evaluate is whether the mean of 3-month highest apprehensions has changed.

```r
# In 2010, March, April,and May have the most apprehensions
top3_2010 <- order(T2010[10,1:12], decreasing = TRUE)[1:3]
colnames(T2010)[top3_2010]
```

```
## [1] "March" "April" "May"
```

```r
# In 2017, November, October,and December have the most apprehensions
top3_2017 <- order(T2017[10,1:12], decreasing = TRUE)[1:3]
colnames(T2017)[top3_2017]
```

```
## [1] "November" "October"  "December"
```

Null hypothesis and alternative hypothesis:

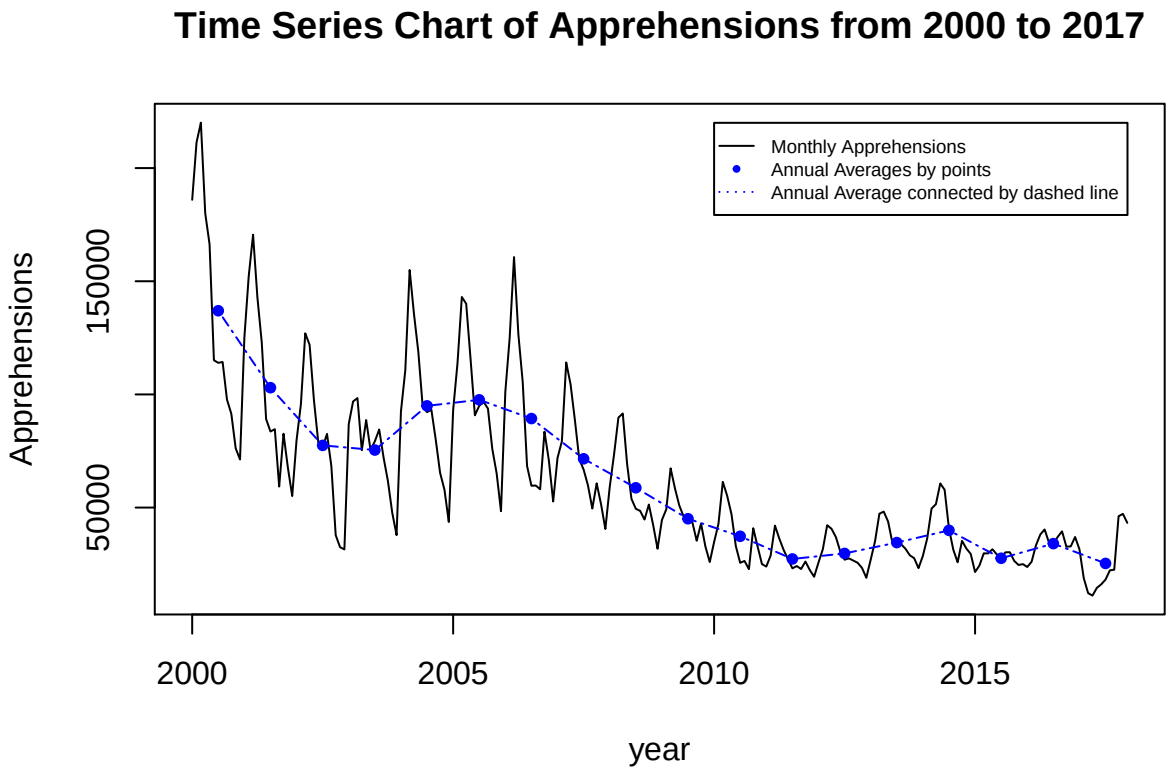
$$H_0 : \mu_{3months,2010} - \mu_{3months,2017} = 0$$

$$H_a : \mu_{3months,2010} - \mu_{3months,2017} \neq 0$$

```
##
##  Welch Two Sample t-test
##
## data:  T2010_three_months and T2017_three_months
## t = -5.9857, df = 11.05, p-value = 0.00008944
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -31591.20 -14611.46
## sample estimates:
## mean of x mean of y
##   2225.00  25326.33
```

From t-test, we get p-value $= 0.3454 > 0.05$, and we fail to reject the null hypothesis. At significance level of 5%, the mean of apprehensions from March to May in 2010 is equal to the mean of apprehensions from October to December in 2017. Thus, there is no significance difference betweeen maximum of apprehensions from March to May in 2010 and from October to December in 2017.

Time series



Conclusion

Through Time Series Chart of Apprehension from 2000 to 2017 above, we can observe that the decreasing trend happens long before the president Donald Trump was elected. Although we cannot deny that the new law might have some restrained effect on the drop of apprehension, the effect is limited. Then, the t-tests shows that the data through years are not significantly different. To get a better understanding of the US-Mexico border apprehensions, we need to obtain more data about the reasons that the apprehensions in each sector changes through months and years.