

# Border Apprehension Analysis

*Maria Ren, Yuehan Xiao, Xu Xu*

*February 19, 2018*

## A short Summary of the Background

Apprehension statistics record the number of foreigners who are caught while illegally entering the United States. These people are charged with violation of the Immigration and Nationality Act, and are subject to removal by the US Border Patrol. According to CNN politics, US Border Patrol reported a historic low of apprehensions under the Trump administration. Our project uses the 2010 and 2017 Illegal Alien Apprehension statistical report and the Border Patrol Monthly Summary dataset to prove the decreasing number of apprehension across 9 main sectors of the US. With simple statistical tests and data analytic methods in R, our graphical analysis also shows the decreasing trend in apprehensions from 2000 to 2017.

## Citations

Kopan, Tal. "US-Mexico Border Apprehensions Hit 17-Year Lows." CNN, Cable News Network, 9 May 2017, [www.cnn.com/2017/05/09/politics/border-crossings-apprehensions-down-trump/index.html](http://www.cnn.com/2017/05/09/politics/border-crossings-apprehensions-down-trump/index.html).

```
A2010<- read.csv('PB Apprehensions 2010.csv', header=TRUE, stringsAsFactors = FALSE)
A2017<- read.csv('PB Apprehensions 2017.csv', header=TRUE, stringsAsFactors = FALSE)
Monthly<- read.csv('PB monthly summaries.csv',header=TRUE, stringsAsFactors = FALSE)
```

find the sum of each sector without the first column in 2010

```
sector_sum_2010<-rowSums(A2010[,-1])
sector_sum_2010
```

```
## [1] 5288 14694 32562 12251 35287 59766 68565 212202 7116
```

find the sector with the most apprehensions

```
max(sector_sum_2010)
```

```
## [1] 212202
```

Tucson is the sector with the most apprehensions.

find the sum of each sector in 2017

```
sector_sum_2017<-rowSums(A2017[,-1])
```

find the sector with the most apprehensions in 2010

```
max(sector_sum_2017)
```

```
## [1] 137562
```

Rio Grande Valley is the sector with the most apprehensions.

Two samples t-test

Extracting tucson's monthly apprehensions from the data set and set as vector.

```
row.names(A2010) <- A2010[,1]
A2010<-A2010[,2:13]
A2010_tucson <- as.vector(A2010[8,],mode='numeric')
```

Extracting Rio Grande Valley monthly apprehensions from the data set and set as vector.

```

row.names(A2017) <- A2017[,1]
A2017 <- A2017[,2:13]
A2017_RGV <- as.vector(A2017[6,],mode='numeric')

```

Two Sample T-test

```

intake.A2010_sector <- A2010_tucson
intake.A2017_sector <- A2017_RGV
t.test(intake.A2010_sector, intake.A2017_sector)

```

```

##
##  Welch Two Sample t-test
##
## data:  intake.A2010_sector and intake.A2017_sector
## 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

```

Since the p-value is 0.06346 which is greater than 0.05, so we fail to reject null hypothesis. It indicates that there are no significant difference in the mean apprehensions of 2010 Tucson and 2017 Rio Grand Valley. So there is no change in the sector's maximum. Through bar plots, we observed the 3 month periods with the most apprehensions in 2010 is March, April and May And the 3 month periods with the most apprehensions in 2017 are October, November and December.

extract March, April and May's apprehensions from 2010.

```

A2010_3 <- as.vector(A2010[,6], mode='numeric' )
A2010_4 <- as.vector(A2010[,7], mode='numeric' )
A2010_5 <- as.vector(A2010[,8], mode='numeric' )

```

combine them into a matrix and then into a vector

```

A2010_345 <- as.vector(cbind(A2010_3,A2010_4,A2010_5))

```

extract October, November and December's apprehensions from 2017

```

A2017_10 <- as.vector(A2017[,1], mode='numeric' )
A2017_11 <- as.vector(A2017[,2], mode='numeric' )
A2017_12 <- as.vector(A2017[,3], mode='numeric' )

```

combine them into a matrix and then into a vector

```

A2017_101112 <- as.vector(cbind(A2017_10,A2017_11,A2017_12))

```

two sample t-test

```

intake.A2010_month <- A2010_345
intake.A2017_month <- A2017_101112
t.test(intake.A2010_month, intake.A2017_month)

```

```

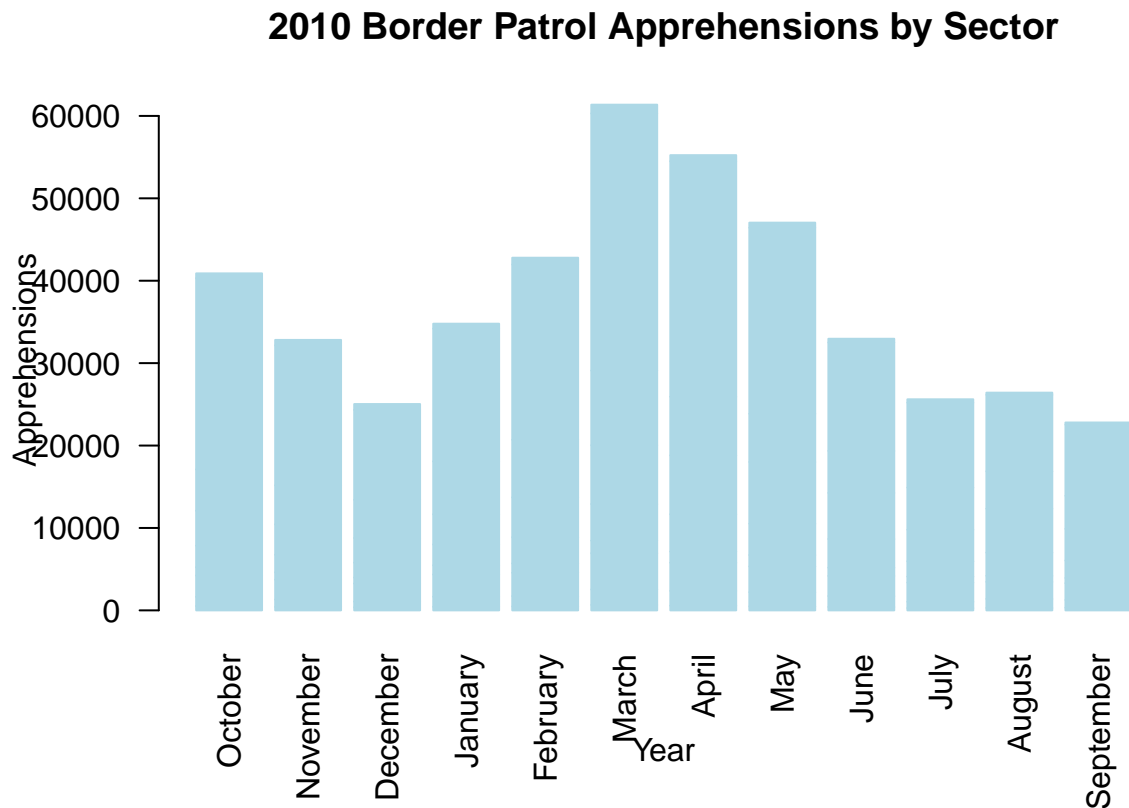
##
##  Welch Two Sample t-test
##
## data:  intake.A2010_month and intake.A2017_month
## t = 0.48741, df = 50.321, p-value = 0.6281
## alternative hypothesis: true difference in means is not equal to 0

```

```
## 95 percent confidence interval:
## -3119.921  5119.699
## sample estimates:
## mean of x mean of y
## 6060.852  5060.963
```

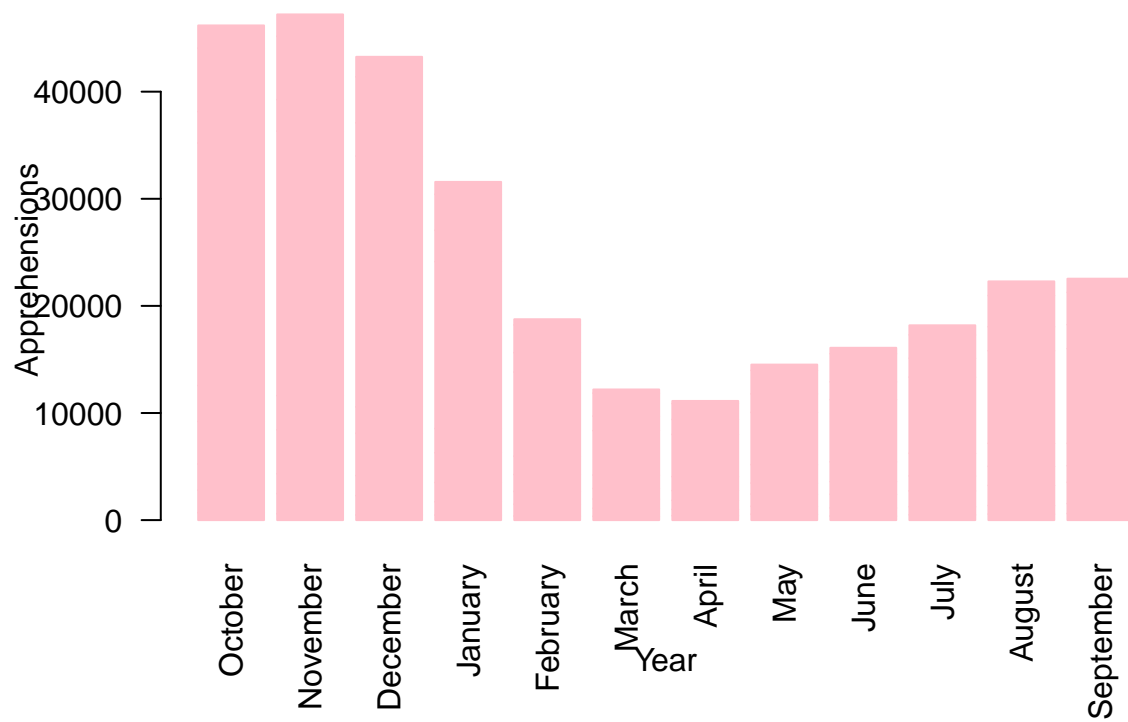
Since the p-value is 0.2075 which is greater than 0.05, so we fail to reject null hypothesis. It indicates that there are no significant difference in the mean apprehensions of 2010 March, April and May, and 2017 October, November and December. So there is no change in the month's maximum.

2010 Monthly Bar plot

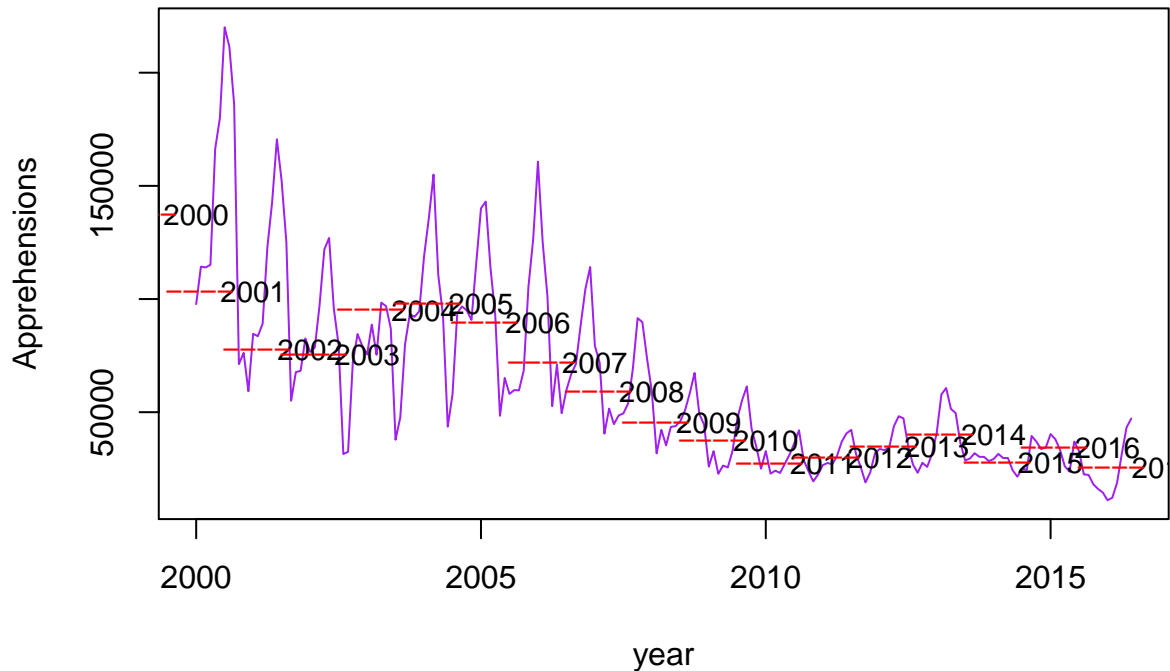


2017 Monthly Bar plot

## 2017 Border Patrol Apprehensions by Sector



Time series plot



From the time series plot, we can see that the average monthly BP Apprehensions decreases across the years between 2000 and 2017.