# Analysis of Trends in Apprehensions Along the United States Southern Border

*Tiffany Cheng and Xiru Lyu*

*2/19/2018*

## Background

Illegal crossing of the United States southern border has been a central topic of debate in the current political climate. Data from US Customs and Border Protection show that apprehensions in federal fiscal year 2017 are at an all-time low when compared to data from fiscal year 2000. In the following analysis, we aim to study whether apprehensions have actually decreased from 2000 to 2017 through a time series plot. In particular, we would analyze data from fiscal year 2010 and 2017 through the use of both bar plots and statistical tests.

## Data Comparison

### Comparison of 2010 and 2017 Data by Sector

Figure 3 and 4 show the comparison of total number of illegal alien apprehensions in fiscal year 2010 and 2017 by sector. In some sectors such as El Centro, San Diego, and Tucson, there was a significant drop in total number of apprehensions for each month in 2017 compared with data in 2010. In contrast, the total number of apprehensions increased sharply in every month in sectors such as El Paso, Rio Grande Valley, and Yuma in 2017 compared with data in 2010.

### Comparison of 2010 and 2017 Data by Month

Figure 1 and 2 show the comparison of total number of illegal alien apprehensions in fiscal year 2010 and 2017 by month. The total number of apprehensions is generally lower in 2017 compared with data in 2010 for each month. There was a sharp decrease in the number of total apprehensions in Tucson from 2010 to 2017. However, it's noteworthy to point out that there was a large increase in the total number of apprehension in Rio Grande Valley from 2010 to 2017.

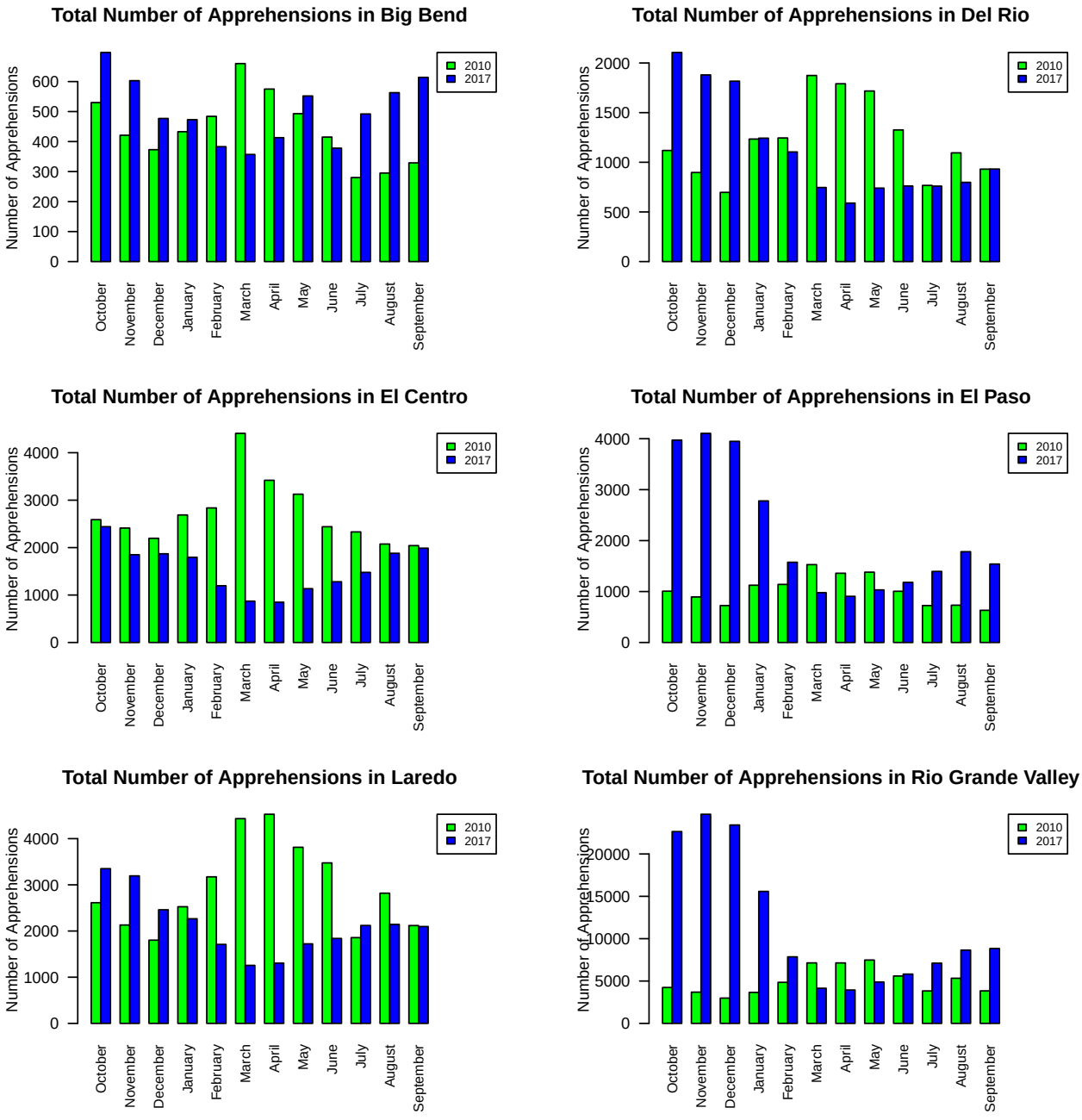


Figure 1: Comparison of Data by Sector

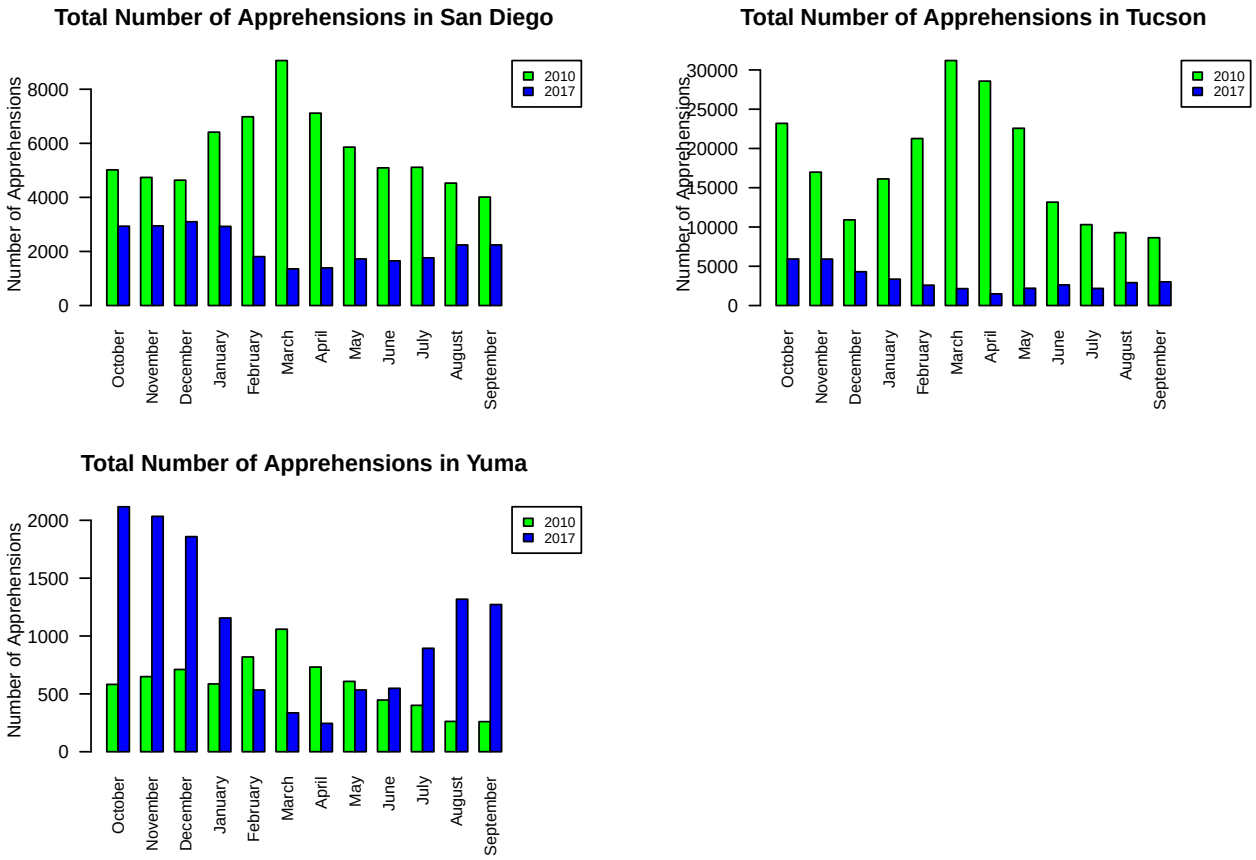


Figure 2: Comparison of Data by Sector

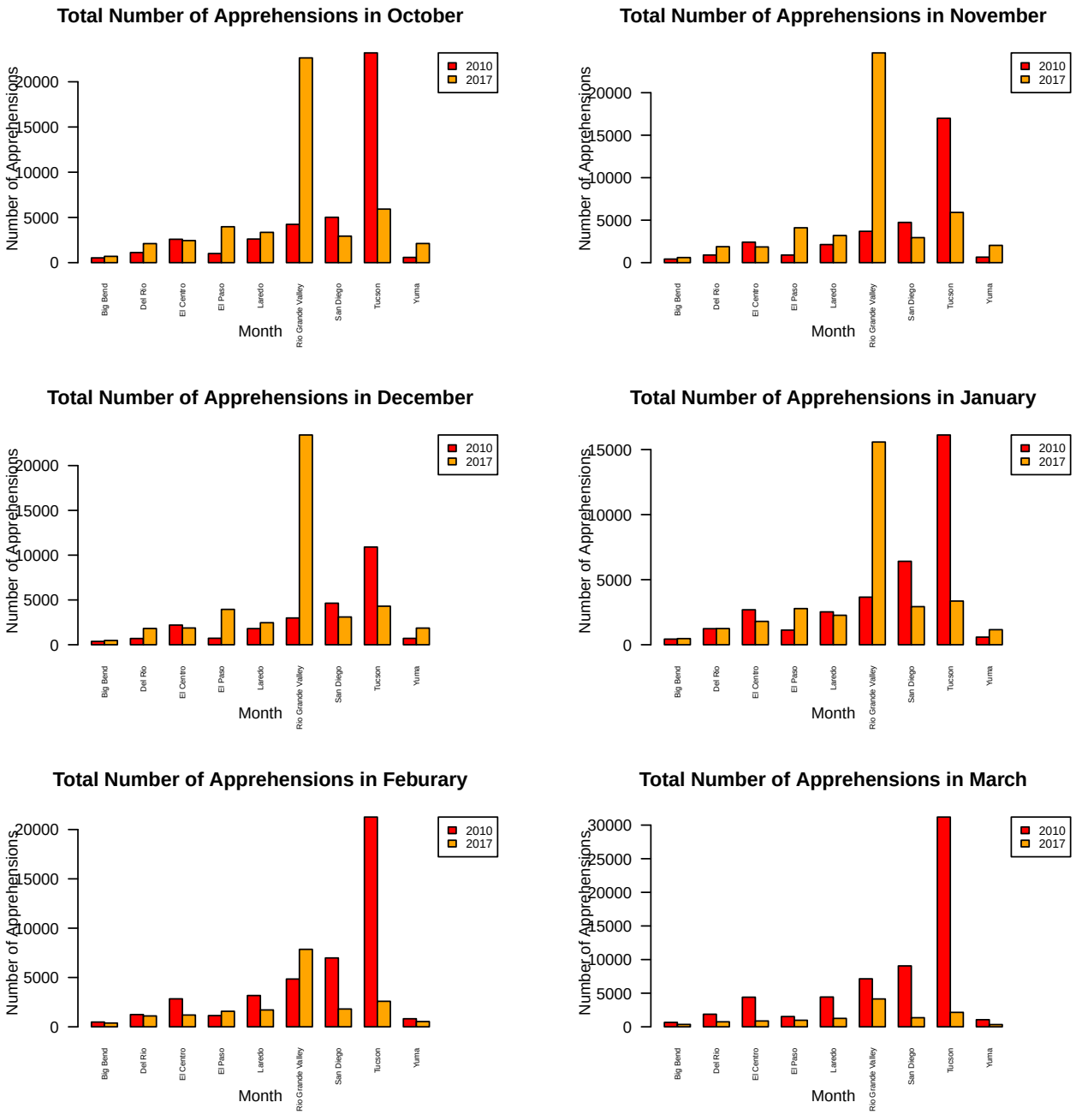


Figure 3: Comparison of Data by Month

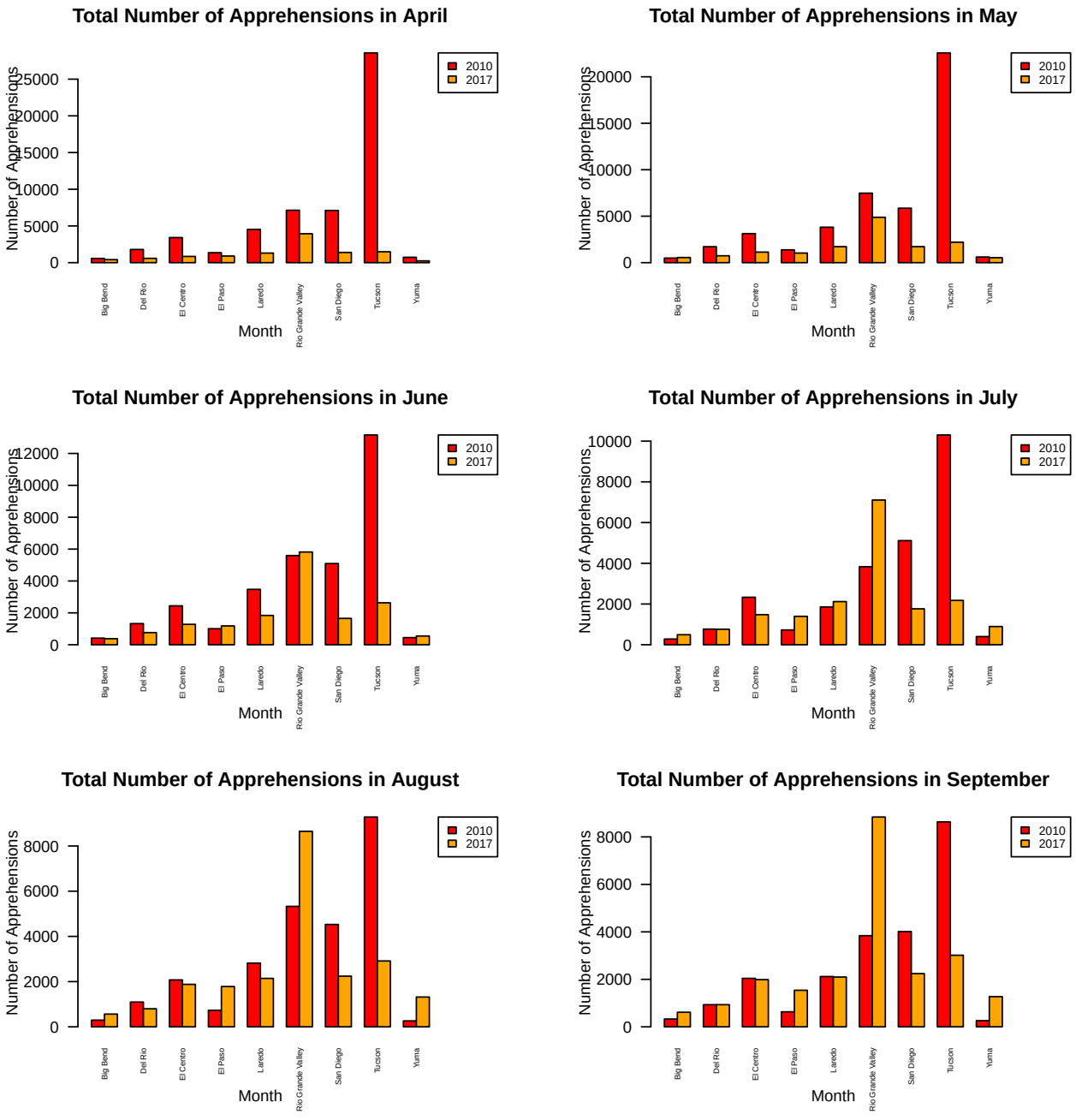


Figure 4: Comparison of Data by Month

# Statistical Tests

## Difference in Mean Maximum Apprehensions for One-month Period

We would like to find out whether there has been a change in the maximum level of apprehensions in 2017 compared with 2010. This is a comparison between the sector with the most apprehensions in 2010 (Tucson) and the sector with the most apprehensions in 2017 (Rio Grande Valley). Before conducting the t-test, we performed an F-test to compare the variances of the two sectors with the most apprehensions in 2010 and 2017, respectively. The hypothesis is as follows:

$$H_0 : \sigma^2_{Tucson} = \sigma^2_{RioGrandeValley}$$

$$H_A : \sigma^2_{Tucson} \neq \sigma^2_{RioGrandeValley}$$

The result of the test, as shown below, points out that variances of the two samples are equal.

```
##
##  F test to compare two variances
##
## data:  most[, 1] and most[, 2]
## F = 0.93205, num df = 11, denom df = 11, p-value = 0.9092
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2683166 3.2376648
## sample estimates:
## ratio of variances
##           0.932051
```

We then performed a two independent sample t-test to test whether the mean of maximum apprehension level in 2010 differs from that in 2017, which has the following hypothesis:

$$H_0 : \mu_{Tucson} = \mu_{RioGrandeValley}$$

$$H_A : \mu_{Tucson} > \mu_{RioGrandeValley}$$

As seen by the summary table shown below, with a p-value of 0.03172 and a significance level of $\alpha = 0.05$, we would reject the null hypothesis. This leads to the conclusion that the mean of maximum apprehension level in 2010 and is higher than that in 2017.

```
##
##  Two Sample t-test
##
## data:  most[, 1] and most[, 2]
## t = 1.9547, df = 22, p-value = 0.03172
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  756.0001      Inf
## sample estimates:
## mean of x mean of y
##   17683.5   11463.5
```

## Difference in Mean Maximum Apprehensions by Three-Month Period

We then would like to know whether there is a difference in the maximum level of apprehension by three-month period in 2010 and in 2017. This is a comparison between the time period with the largest total number of

apprehensions in 2010 (Mar-May) and the time period with largest total number of apprehensions in 2017 (Oct-Dec). Before conducting the actual t-test, we performed a variance test to see whether the variances of the two samples are the same. The hypothesis is as follows:

$$H_0 : \sigma^2_{2010(Mar-May)} = \sigma^2_{2017(Oct-Dec)}$$

$$H_A : \sigma^2_{2010(Mar-May)} \neq \sigma^2_{2017(Oct-Dec)}$$

The result of the test, which is shown below, suggests that there is no difference between two variances.

```
##
##  F test to compare two variances
##
## data:  max_three$max_2010 and max_three$max_2017
## F = 1.43, num df = 9, denom df = 9, p-value = 0.6027
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3551856 5.7570716
## sample estimates:
## ratio of variances
##           1.429975
```

After checking the variances, we performed a two independent sample t-test to test whether the mean of the maximum level of apprehensions in the three-month period in 2010 is different from that in 2017. The hypothesis of the test is as follows:

$$H_0 : \mu_{2010(Mar-May)} = \mu_{2017(Oct-Dec)}$$

$$H_A : \mu_{2010(Mar-May)} \neq \mu_{2017(Oct-Dec)}$$

The result of this t-test is shown below. With a significance level of $\alpha = 0.05$ and a p-value of 0.8032, we can say that there is no difference in the means of the two samples. Thus, we can conclude that there is no change in the maximum three-month period apprehension level in 2010 compared with that in 2017.

```
##
##  Two Sample t-test
##
## data:  max_three$max_2010 and max_three$max_2017
## t = 0.25287, df = 18, p-value = 0.8032
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -39459.94  50258.74
## sample estimates:
## mean of x mean of y
##   32728.6   27329.2
```

# Time Series Plot

Figure 5 is a time series plot for the monthly apprehension level across all southern US sectors from fiscal year 2010 to 2017. Note that the federal fiscal year starts in October and for the convenience of labeling, we converted fiscal year into normal calendar year. We set the starting point of the time series object, October 2000, to be $(2000, 1)$, which signifies the first month of the fiscal year 2000. Therefore, values on the x-axis of the time series plot are denoted in fiscal years. The red segmented lines in the plot reflect the average number of apprehensions for each fiscal year and each red line covers the whole period of each fiscal year, with a label indicating the year positioned to the right of the line. The plot shows that the level of apprehensions

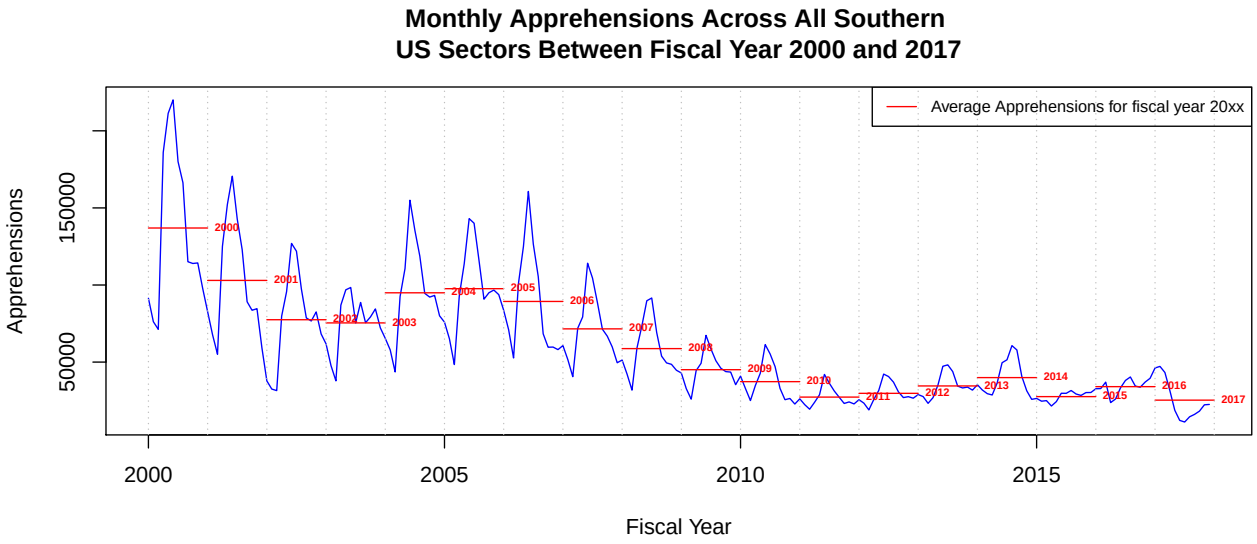


Figure 5: Time Series Plot for Apprehensions Across All Sectors

decreases over time from fiscal year 2000 to 2017, with 2000 having the highest level of apprehensions and 2017 having the lowest level of apprehensions.

# Conclusion

Through the use of bar plots and statistical tests, we see that the trend has not changed significantly from 2010 to 2017. However, after looking at the trend over 17 years in the time series plot, it is evident that there has been a downward decline in border apprehensions from 2000 to 2017 and that the number of apprehensions in 2017 is at indeed an all-time low.