

Border Patrol Apprehensions: 2010 vs 2017

Han Nguyen, Carlos Echeverri, Nathan Mokhtarzadeh

February 19, 2018

Since Donald Trump became President, apprehensions at the US-Mexico border have been very low compared to previous years. This is because people have stopped illegally crossing the Southern border, which may be due to the President's "rhetoric and aggressive push" to enforce immigration laws. "Would-be" immigrants may be waiting and watching instead of crossing the border.

In this project, we compare the number of border patrol apprehensions in the years 2010 and 2017, as well as monthly summaries to see if there has been a significant decrease in apprehensions. We compare different months in each sector: Big Bend, Del Rio, El Centro, El Paso, Laredo, Rio Grande Valley, San Diego, Tucson, and Yuma.

Ahead we call in the two data tables, which now include totals for each sector and month.

```
# read bp tables for each year
A2010 <- read.csv("bp appre 2010.csv", header = TRUE, stringsAsFactors = FALSE)
A2017 <- read.csv("bp appre 2017.csv", header = TRUE, stringsAsFactors = FALSE)
```

```
# get name of each sector
rownames(A2010) <- A2010[,1]
A2010[,1]
```

```
## [1] "Big Bend"      "Del Rio"       "El Centro"
## [4] "El Paso"       "Laredo"        "Rio Grande Valley"
## [7] "San Diego"     "Tucson"        "Yuma"
```

```
rownames(A2017) <- A2017[,1]
```

```
# function to get totals of each sector, month, and data as a whole
getTotals <- function(data) {
  data <- subset(data, select= -c(Sector))
  rownames(data)

  data <- rbind(data, colSums(data))

  rownames(data)

  -length(rownames(data))

  rownames(data) <- c(rownames(data)[-length(rownames(data))], "Total")

  data <- cbind(data, rowSums(data))

  colnames(data) <- c(colnames(data)[-length(colnames(data))], "Total")
  return(data)
}
```

```
# Append totals for data
A2010 <- getTotals(A2010)
A2017 <- getTotals(A2017)
```

Visualization of the Data

We compare the number of apprehensions in 2010 and 2017 by months and sector in a side by side barplot. Below are 9 barplots that represent each sector.

```
# a function to create side-by-side barplots comparing the 2010 and 2017 statistics by
# month and sector
compare <- function(){

  for (x in 1:9) {

    title <- rownames(A2010)[x]

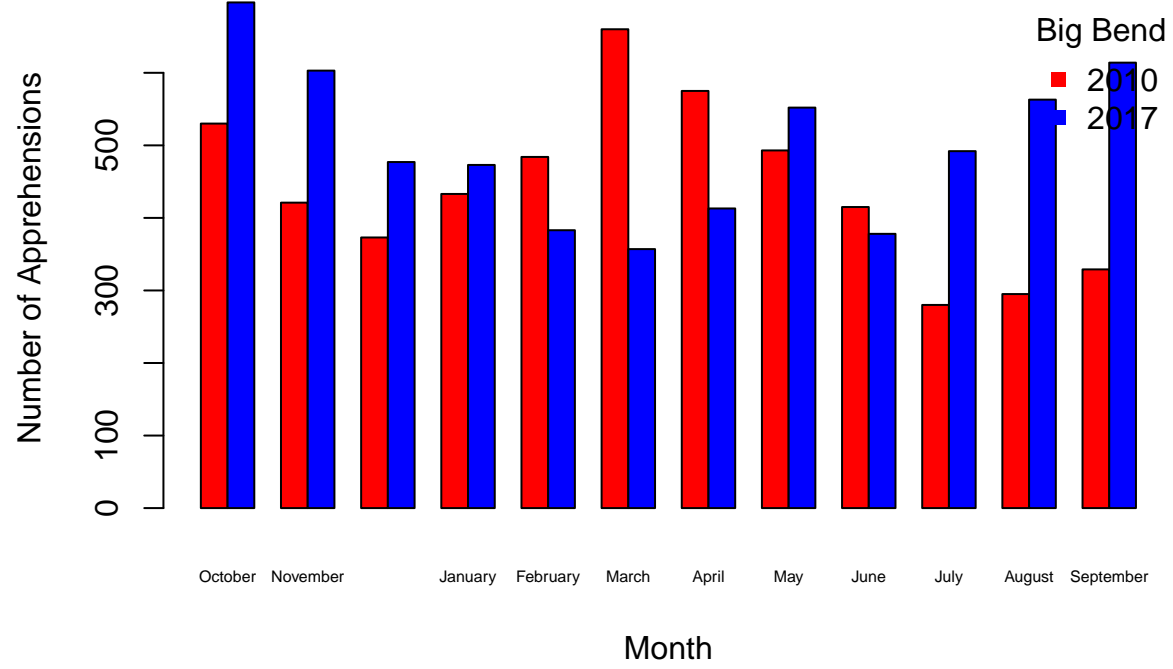
    both <- rbind(A2010[x,1:12], A2017[x,1:12])

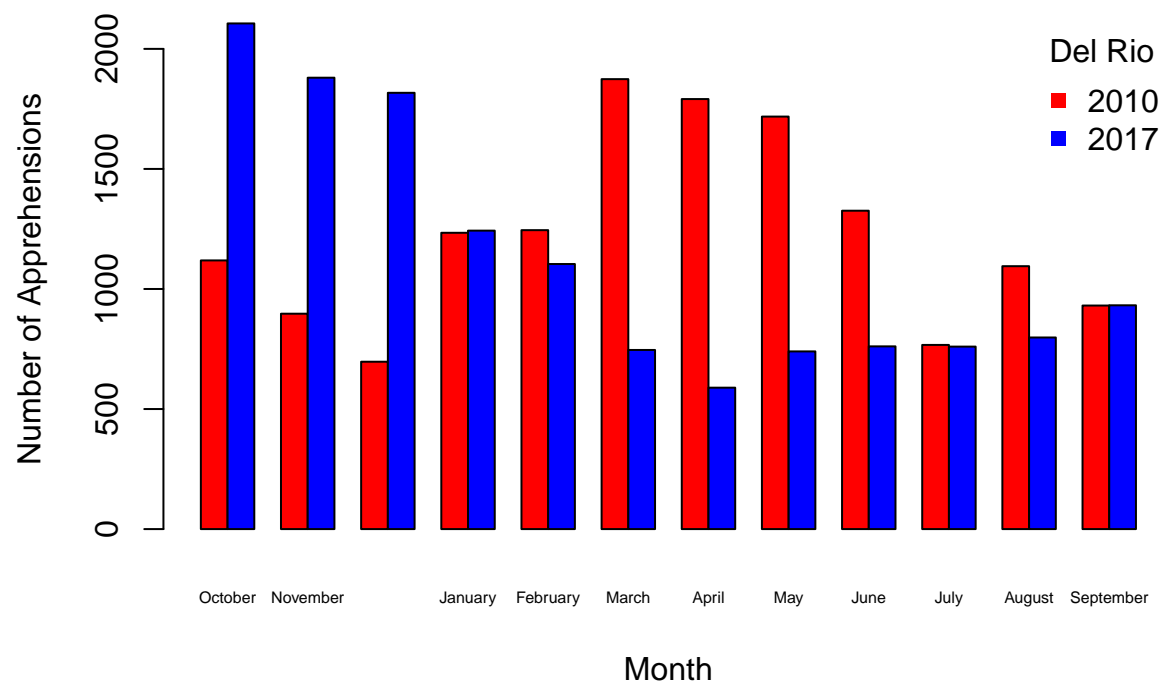
    row.names(both) <- c("2010", "2017")

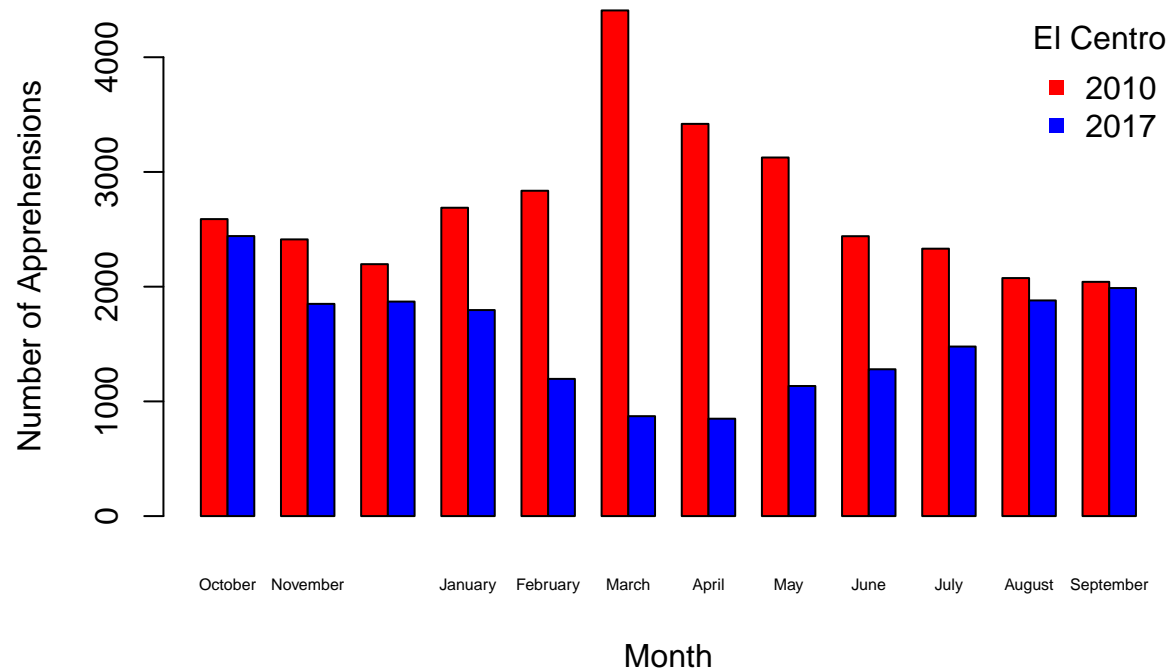
    barplot(as.matrix(both), beside = TRUE, col = c("red", "blue"), bty="n", xlab = 'Month',
            ylab = 'Number of Apprehensions', cex.names = 0.5 )

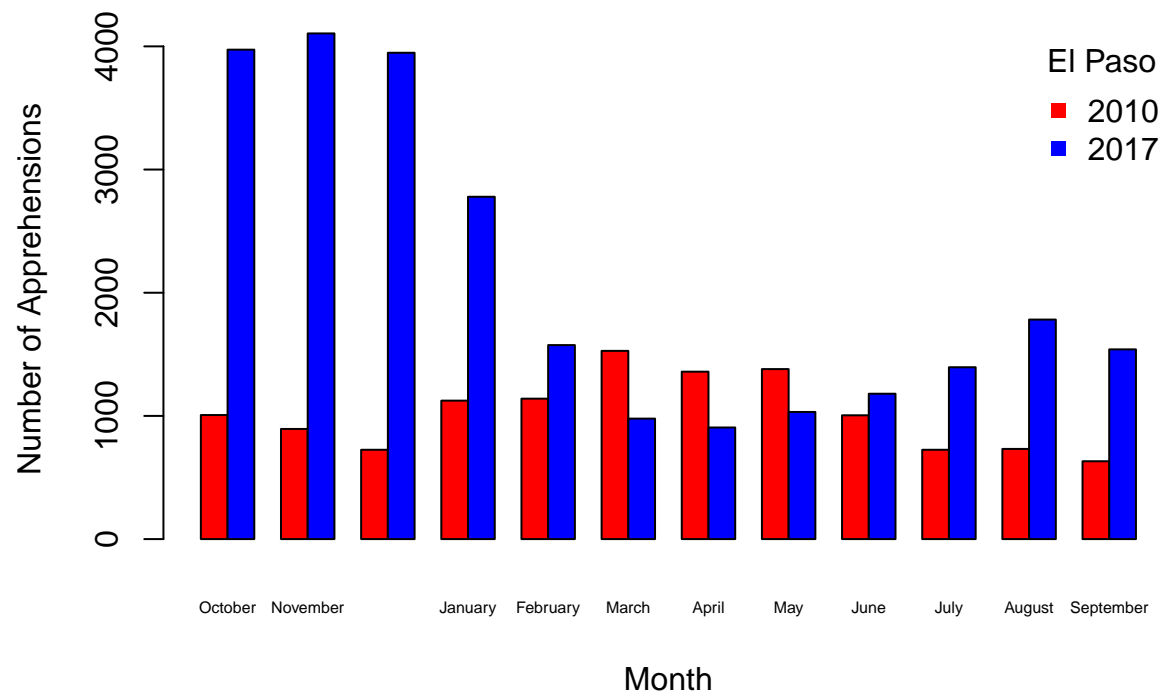
    legend("topright", c("2010","2017"), pch=15, col=c("red","blue"), bty="n",
           title = title)

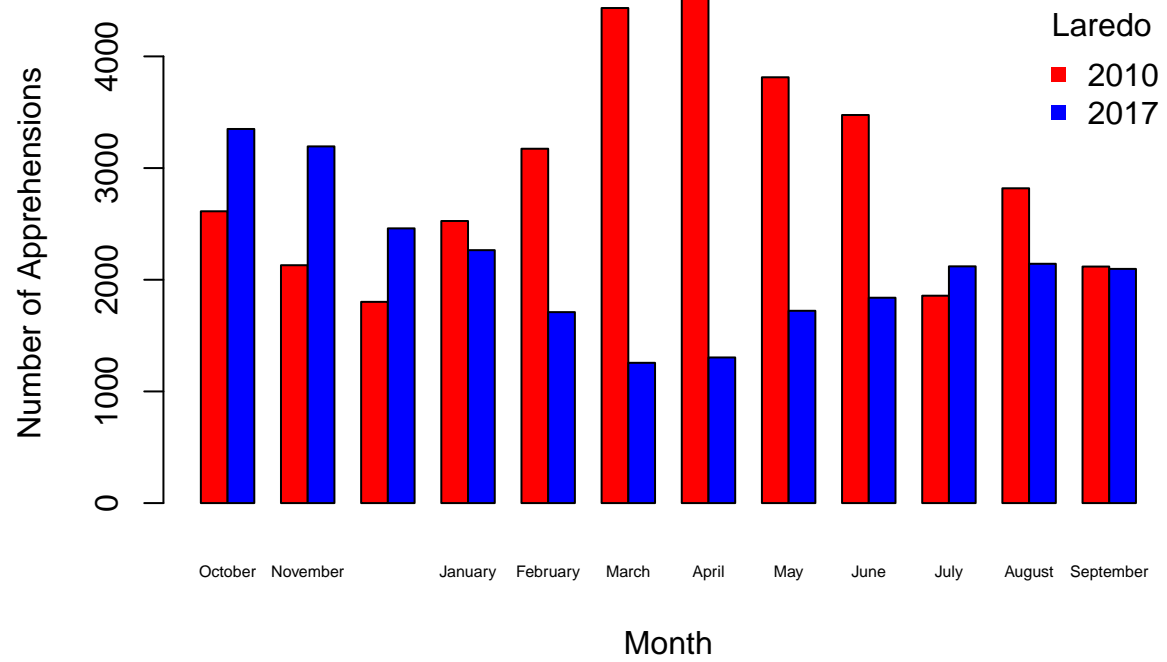
  }
}
compare()
```

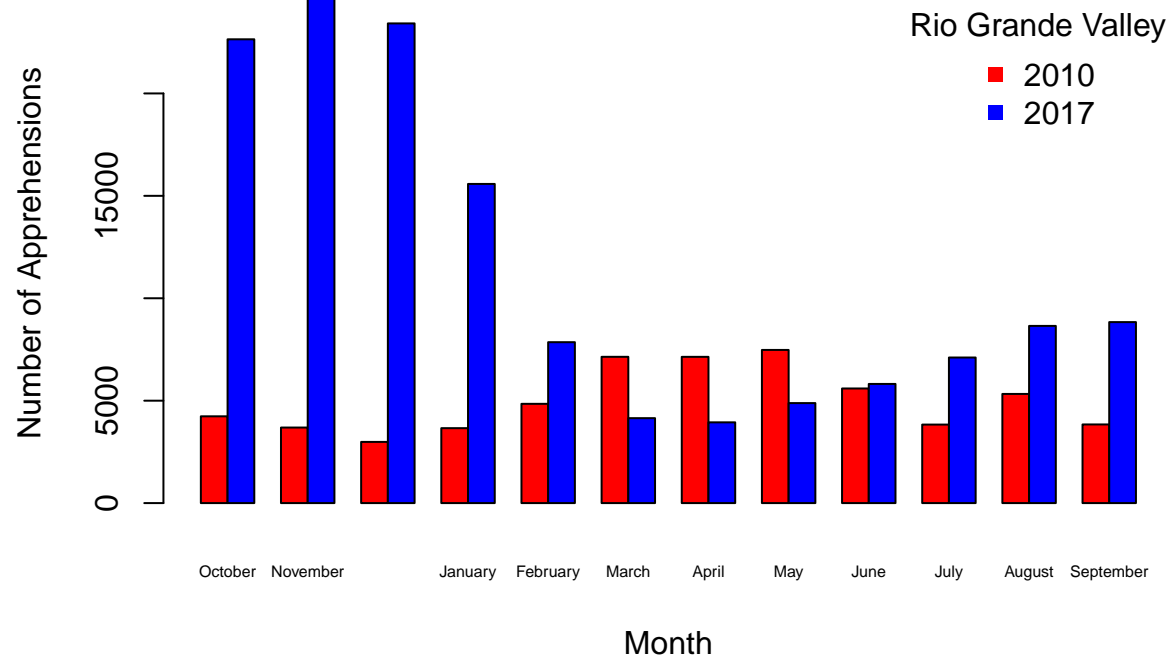


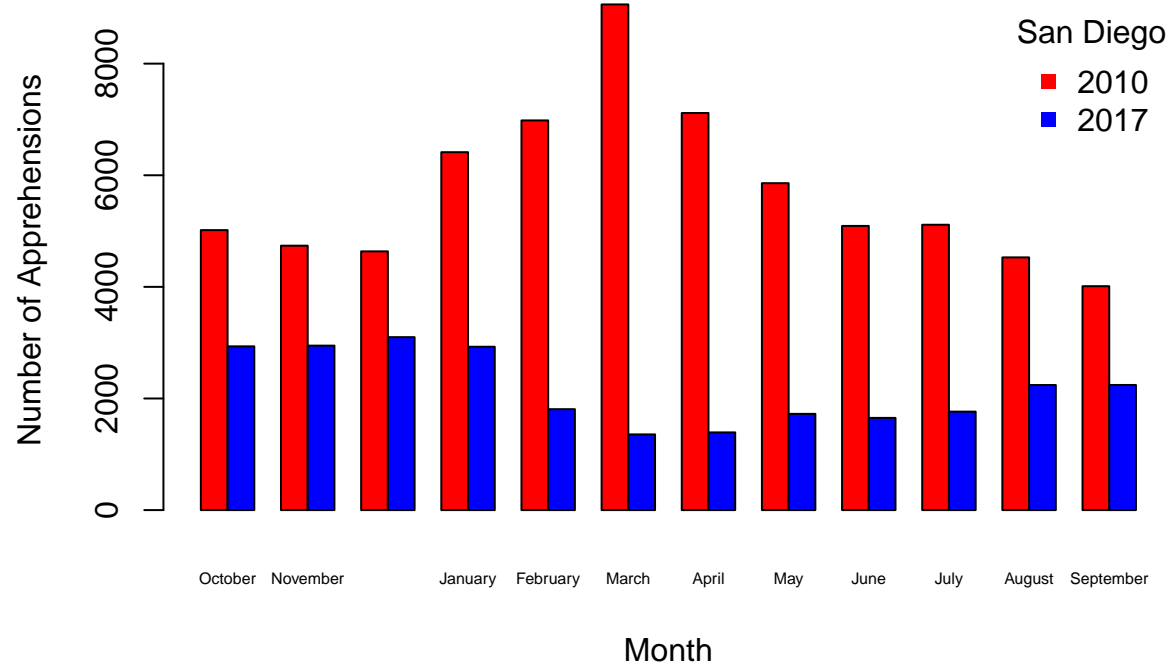


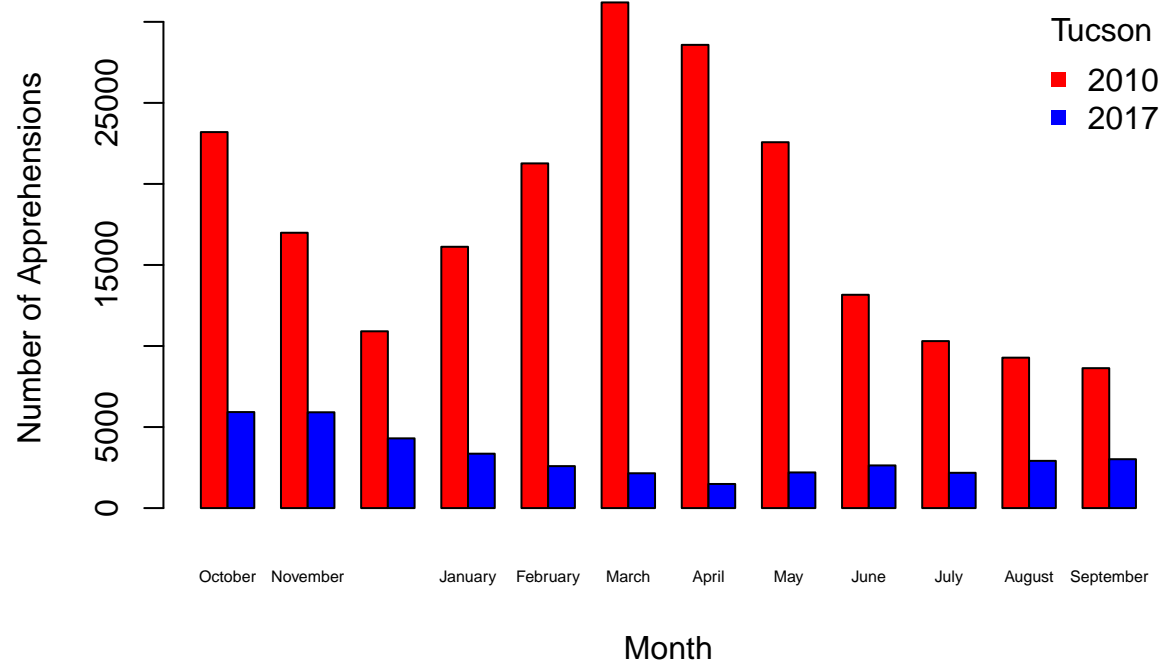


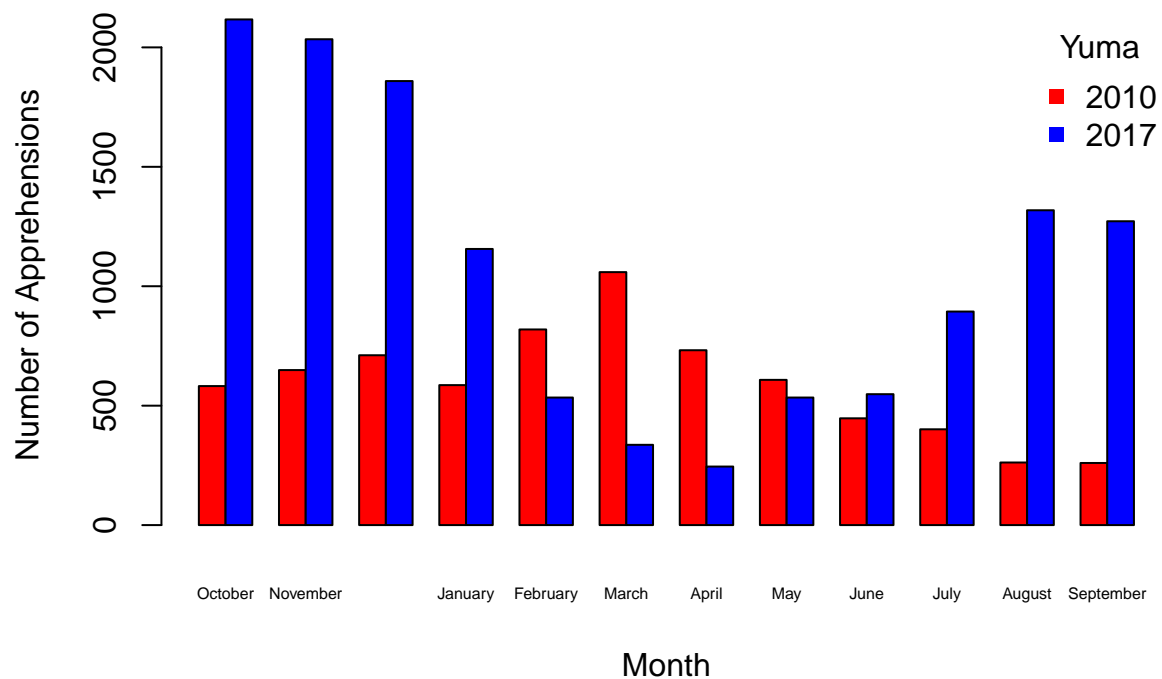












In general, there seemed to be a much higher number of apprehensions in 2010 between the months of February and May, and a higher number of apprehensions in 2017 between September and December. Tucson and San Diego had the most change when looking at the barplots. There was a large decrease in number of apprehensions from 2010 to 2017.

Statistical Testing

Most Apprehensions by Sector

We first compared the sector with the most apprehensions in 2010 and 2017. In 2010, Tucson had the most apprehensions, and in 2017, Rio Grande Valley had the most apprehensions. With this data, we took a t-test to see if there was a significant difference in these maximums. The null hypothesis is that the means between these apprehensions are equal to 0, and the alternative hypothesis is that they are not equal to 0.

```
# T Test for most apprehensions by sector
customTest <- function(earlyData, newData) {
  #Assumes getTotals function has been run on data
  #Data for old data input
  maximumOld <- max(earlyData[1:9,13])
  maxRowIndexOld <- which.max(earlyData[1:9,13])
  earlyDataMean <- rowMeans(earlyData[maxRowIndexOld,1:12])

  #Data for new data input
  maximumNew <- max(newData[1:9,13])
  maxRowIndexNew <- which.max(newData[1:9,13])
  newDataMean <- rowMeans(newData[maxRowIndexNew,1:12])

  t.test(earlyData[maxRowIndexOld,1:12], newData[maxRowIndexNew,1:12])
}
customTest(A2010,A2017)

##
## Welch Two Sample t-test
##
## data: earlyData[maxRowIndexOld, 1:12] and newData[maxRowIndexNew, 1:12]
## 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
```

Although the mean of Tucson in 2010 was 17684 apprehensions, and the mean of Rio Grande Valley in 2017 was 11464, there seems that there would be a difference in the mean. Surprisingly, the t-test gave a t-score of 2, and a p-value of 0.06. Using an alpha of 0.05, we fail to reject the null hypothesis since the p-value is greater than alpha. This p-value is still low, but not low enough. This means that the difference in the means of apprehensions in 2010 and 2017 are not significantly different. This is also seen in the confidence interval, which is (-380, 12820), and includes 0, meaning there is not a significant difference.

Then, we compared the 3 month periods with the most apprehensions in 2010 and 2017. In 2010, March, April, and May had the most apprehensions, and in 2017, October, November, and December had the most apprehensions. Again, with this data, we took a t-test to see if there was a significant difference in these maximums.

Most Apprehensions in 3 Month Period

```
#Get top three months
topMonths <- function(data){
  copy <- data
  copy <- sort(copy[10,1:12], decreasing = TRUE)[1:3]
  copy
}

# T test for top three months in apprehensions
compareThreeMonthPeriod <- function(earlyData, newData) {
  earlyDataThree <- topMonths(earlyData)
  newDataThree <- topMonths(newData)
  # get indices of these months
  earlyColName1 <- grep(colnames(earlyDataThree[1]), colnames(earlyData))
  earlyColName2 <- grep(colnames(earlyDataThree[2]), colnames(earlyData))
  earlyColName3 <- grep(colnames(earlyDataThree[3]), colnames(earlyData))
  earlyIndex <- min(earlyColName1,earlyColName2,earlyColName3)

  newColName1 <- grep(colnames(newDataThree[1]), colnames(newData))
  newColName2 <- grep(colnames(newDataThree[2]), colnames(newData))
  newColName3 <- grep(colnames(newDataThree[3]), colnames(newData))
  newIndex <- min(newColName1,newColName2,newColName3)

  earlyMean <- mean(c(earlyData[1:9,earlyIndex],earlyData[1:9,earlyIndex+1],
                     earlyData[1:9,earlyIndex+2]))
  newMean <- mean(c(newData[1:9,newIndex],newData[1:9,newIndex+1],newData[1:9,newIndex+2]))

  t.test(c(earlyData[1:9,earlyIndex],earlyData[1:9,earlyIndex+1],
           earlyData[1:9,earlyIndex+2]),c(newData[1:9,newIndex],newData[1:9,newIndex+1],
                                           newData[1:9,newIndex+2]))
}

compareThreeMonthPeriod(A2010,A2017)

##
## Welch Two Sample t-test
##
## data: c(earlyData[1:9, earlyIndex], earlyData[1:9, earlyIndex + 1], and c(newData[1:9, newIndex], r
## 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
```

Although the mean for 2010 was 6061, and the mean for 2017 was 5061, which is a decrease in apprehensions, the t-test said that this difference was insignificant. The t score was 0.5 and the p-value was 0.6, which is way above the alpha of 0.05. Since the p-value was above alpha, we fail to reject the null hypothesis, so the

true difference in means is equal to 0, which means there is no significant difference in the means comparing most apprehensions in 3 month periods. The confidence interval says the same since it is (-3120,5120), which includes 0, meaning the differences in means are insignificant.

Most Apprehensions Overall

Because we believed there still may be a significant difference, we took a t-test of each data as a whole in 2010 and 2017, to see if there may be a difference there. The total number of apprehensions in 2010 was 447731, and in 2017 was 303916. This looks like a huge difference, since it nearly decreased by 15000.

```
# T test for whole data  
t.test(A2010, A2017)
```

```
##  
## Welch Two Sample t-test  
##  
## data: A2010 and A2017  
## t = 0.93944, df = 226.33, p-value = 0.3485  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -4856.601 13706.755  
## sample estimates:  
## mean of x mean of y  
## 13776.338 9351.262
```

Surprisingly, although there seemed to be a big change, there was not a significant difference in the means of apprehensions. The t test gave a t score of 0.9 and the p-value was 0.3, which is above the alpha of 0.05, meaning we fail to reject the null hypothesis. So, there is no true difference between the means in 2010 and 2017. The confidence interval says the same, being (-4857,13707), which includes 0, saying that there is no true difference between the means.

Time Series Chart

```
# time series chart for monthly summaries
```

```
ts <- readClipboard(raw = FALSE)
```

```
ts1 <- read.table("clipboard", sep = "\t")
```

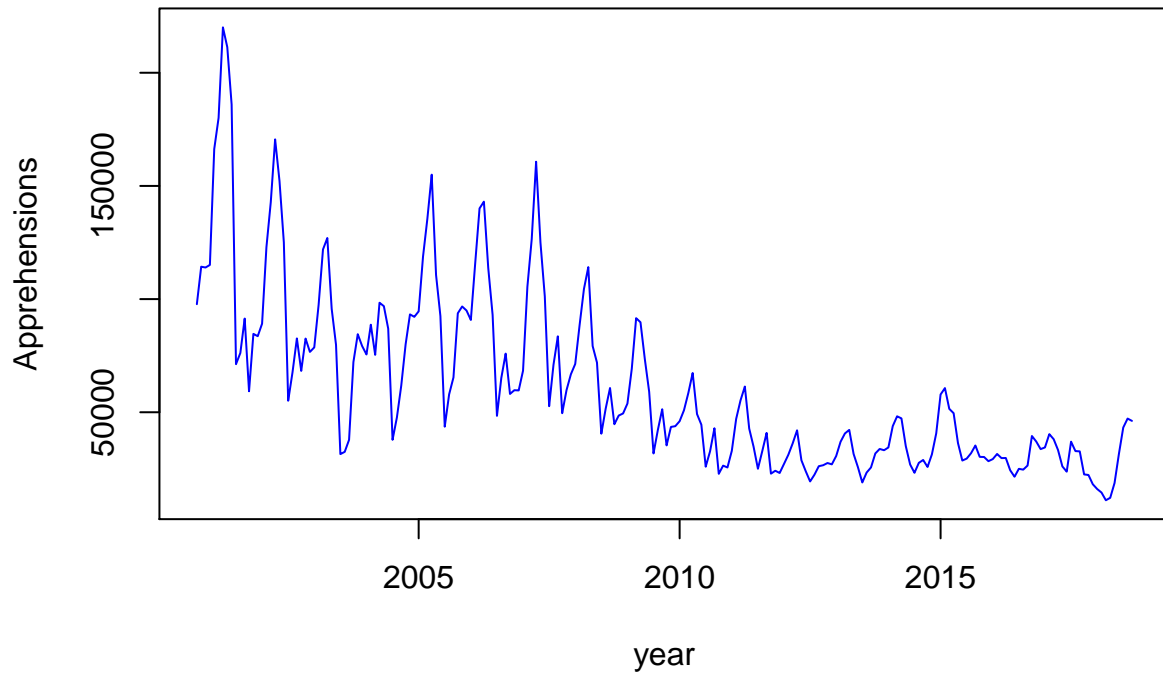
```
ts1
```

##	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
## 1	46184	47211	43251	31576	18754	12195	11127	14519	16087	18187
## 2	32724	32838	37014	23758	26072	33316	38089	40337	34450	33723
## 3	26450	24641	25019	21514	24376	29791	29750	31576	29303	28388
## 4	35312	31896	29528	28668	36403	49596	51502	60683	57862	40708
## 5	28929	27636	23243	26921	35042	47293	48212	43856	34436	33230
## 6	25612	23368	18983	25714	31579	42218	40628	36966	30669	26978
## 7	26165	22405	19429	23926	28786	42014	36251	31236	27166	23170
## 8	40890	32815	25034	34784	42790	61361	55237	47045	32955	25609
## 9	42938	32780	25947	44502	49211	67342	58493	50884	46044	43843
## 10	51339	42209	31802	59028	73483	89770	91566	69233	53854	49472
## 11	60713	51594	40527	71934	79268	114137	104465	88504	71338	66782
## 12	83557	70975	52673	101195	125046	160696	126538	105450	68366	59641
## 13	75913	65135	48406	93020	113775	143048	140062	115823	90786	94954
## 14	65391	57894	43614	92521	110669	154981	135468	118726	94590	92165
## 15	61792	47731	37824	86925	96869	98399	75359	88690	75530	79284
## 16	37812	32506	31501	79793	95724	126992	121921	97424	78655	76661
## 17	82632	67709	55081	125090	152229	170580	142813	122927	89131	83602
## 18	91410	76196	71252	185979	211328	220063	180050	166296	115093	113956
##	V11	V12								
## 1	22288	22537								
## 2	37048	39501								
## 3	30239	30286								
## 4	31388	25825								
## 5	33797	31802								
## 6	27567	26591								
## 7	24166	22863								
## 8	26415	22796								
## 9	43522	35359								
## 10	48541	44708								
## 11	59795	49581								
## 12	59751	58084								
## 13	96733	93741								
## 14	93246	80017								
## 15	84486	72176								
## 16	82557	68263								
## 17	84648	59276								
## 18	114312	97744								

```
ts2 <- rev(as.vector(t(ts1)))
```

```
ts3 <- ts(ts2, start = c(2000,10), frequency=12)
```

```
ts.plot(ts3, gpars=list(xlab="year", ylab="Apprehensions", lty=c(1:3)), col = 'blue')
```



The time series chart shows a decrease in apprehensions throughout the years from 2000 to 2017. There seems to be a large decrease, although our statistical tests say otherwise for 2010 and 2017.

Overall

In conclusion, there seems to be a decrease in apprehensions throughout the years, probably due to the fact that less “would-be” immigrants are crossing the border since the president has been changing immigration laws. The t-tests, though, don’t show a significant difference between mean of apprehensions in 2010 and 2017. But the flaw in this may be that we are only comparing the maximum number of apprehensions by sectors or months in each year, disregarding what month and sector they are actually in. If there was a comparison between each sector and each month, there may be a bigger significant difference.