Assignment 3

yaru peng

2018/2/12

Introduction

US-Mexico boarder apprehensions hits 17-year lows in April, continuing a downward slide in the first few months of the Trump administration. Using data of BP Apprehensions 2010, BP Apprehensions 2017, and BP monthly summary, we want to discover if there has been a change in the maximum by sectors and 3-month periods.

Data cleanning

```r
data2010<-read.csv("bp appre 2010.csv")
data2017<-read.csv("bp appre 2017.csv")
datas<-read.csv("BP monthly summaries.csv")
datas2 = apply(t(datas[,2:13]),1,rev)
#datas[,2:13]
#datas = apply(t(datas),1,rev)
datas$mean<-apply(datas[,2:13],1,mean)
ndata<-nrow(datas[,2:13])
rownames(datas)<-datas[,1]
datas<-cbind(datas,rowSums(datas))
colnames(datas)<-c(colnames(datas)[-length(colnames(datas))],"Total")
datas<-subset(datas,select=-c(year))
#View(datas)
rownames(data2010)<-data2010[,1]
data2010<-subset(data2010,select=-c(Sector))
rownames(data2017)<-data2017[,1]
data2017<-subset(data2017,select=-c(Sector))
data2010 <- rbind(data2010, colSums(data2010))
#rownames(data2010)
#-length(rownames(data2010))
rownames(data2010) <- c(rownames(data2010)[-length(rownames(data2010))], "Total")
data2010 <- cbind(data2010,rowSums(data2010))
colnames(data2010) <- c(colnames(data2010)[-length(colnames(data2010))], "Total")
#View(data2010)
data2017 <- rbind(data2017, colSums(data2010))
#-length(rownames(data2010))
rownames(data2017) <- c(rownames(data2017)[-length(rownames(data2017))], "Total")
data2017 <- cbind(data2017,rowSums(data2017))
colnames(data2017) <- c(colnames(data2017)[-length(colnames(data2017))], "Total")
#View(data2017)
```

Statistic data

```r
knitr::opts_chunk$set(echo = TRUE)
mean2010s<-apply(data2010[,1:12],1,mean)
```

```r
mean2010m<-apply(data2010[,1:12],2,mean)
mean2017s<-apply(data2017[,1:12],1,mean)
mean2017m<-apply(data2017[,1:12],2,mean)
bpsector<-cbind(mean2010s,mean2017s)
bpmonth<-cbind(mean2010m,mean2017m)
sec2010<-as.data.frame(t(data2010[,1:12]))
as.character(colnames(sec2010))
```

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

```r
sec2017<-as.data.frame(t(data2017[,1:12]))
as.character(colnames(sec2017))
```

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

```r
t.test(sec2010$Tucson,sec2017$`Rio Grande Valley`)
```

```
##
##  Welch Two Sample t-test
##
## data:  sec2010$Tucson and sec2017$`Rio Grande Valley`
## 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
```

```r
month2010<-as.data.frame((data2010[,6:8]))
as.character(rownames(month2010))
```

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

```r
month2010$sum<-apply(month2010,1,sum)
maxmonth2010<-max(month2010$sum)
month2017<-as.data.frame((data2017[,1:3]))
as.character(rownames(month2017))
```

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

```r
month2017$sum<-apply(month2017,1,sum)
maxmonth2017<-max(month2017$sum)
t.test(month2010$sum,month2017$sum)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  month2010$sum and month2017$sum
## t = -0.027015, df = 17.537, p-value = 0.9788
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -53963.24  52595.64
## sample estimates:
## mean of x mean of y
##   32728.6   33412.4
```
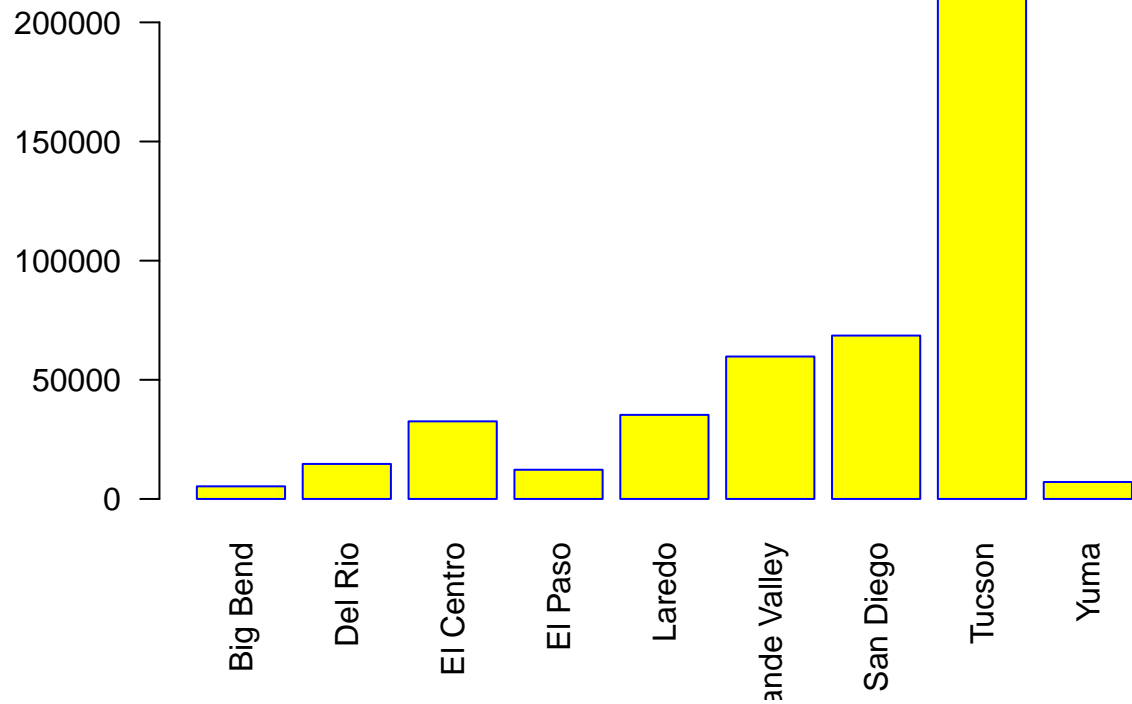
Barplots

```r
knitr::opts_chunk$set(echo = TRUE)
barplot(data2017[1:9,13], names.arg = rownames(data2017)[1:9],
        las=2,
        axisnames=TRUE,
        main="2017 Border Patrol Apprehensions by Sector",
        border="blue",
        col="red")
```


2017 Border Patrol Apprehensions by Sector
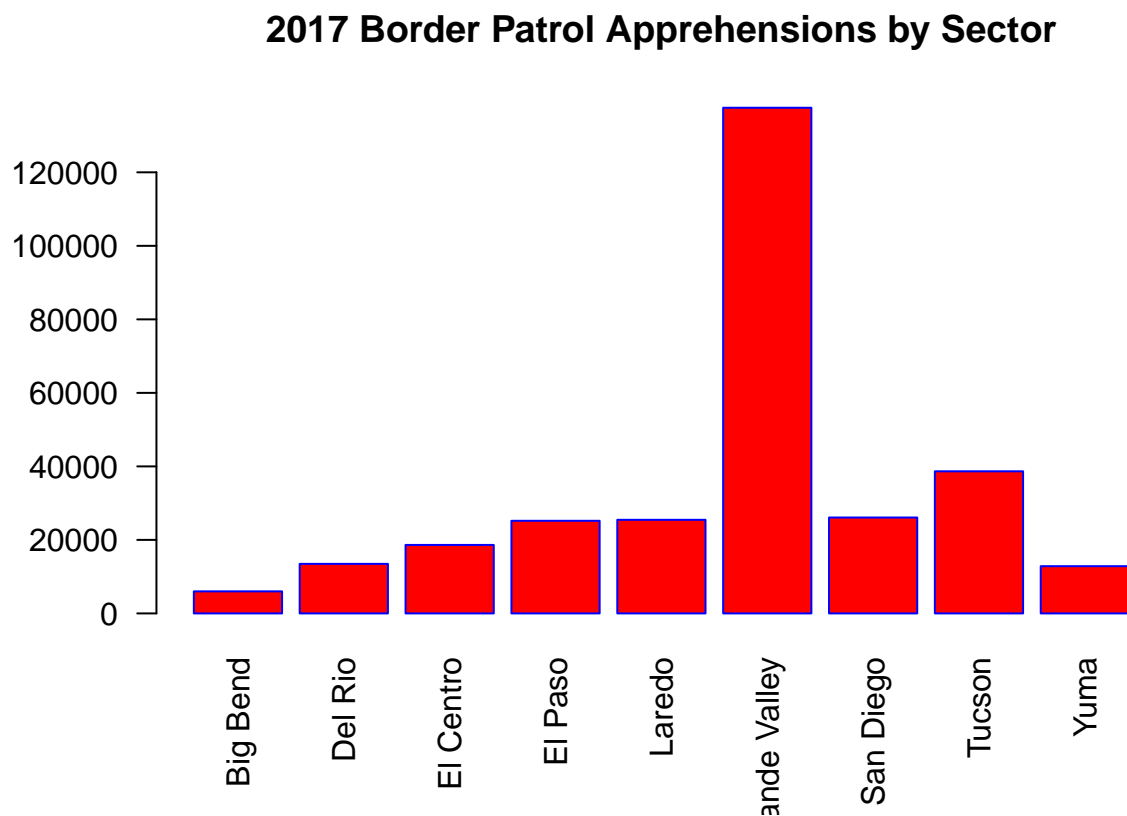
```r
barplot(data2010[1:9,13], names.arg = rownames(data2010)[1:9],
        las=2,
        axisnames=TRUE,
        main="2010 Border Patrol Apprehensions by Sector",
```
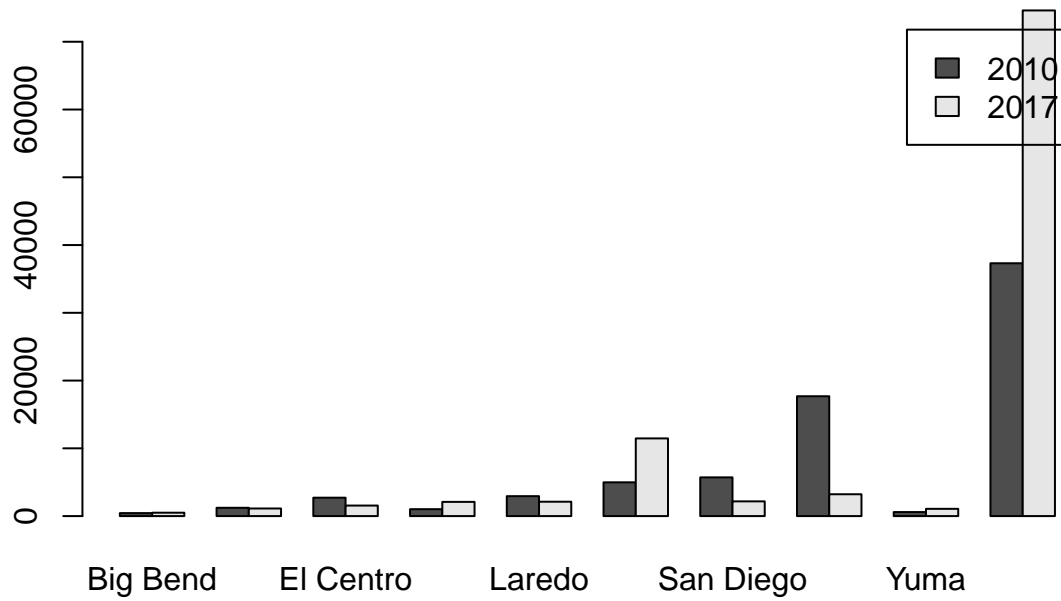
```
                border="blue",
                col="yellow")
```
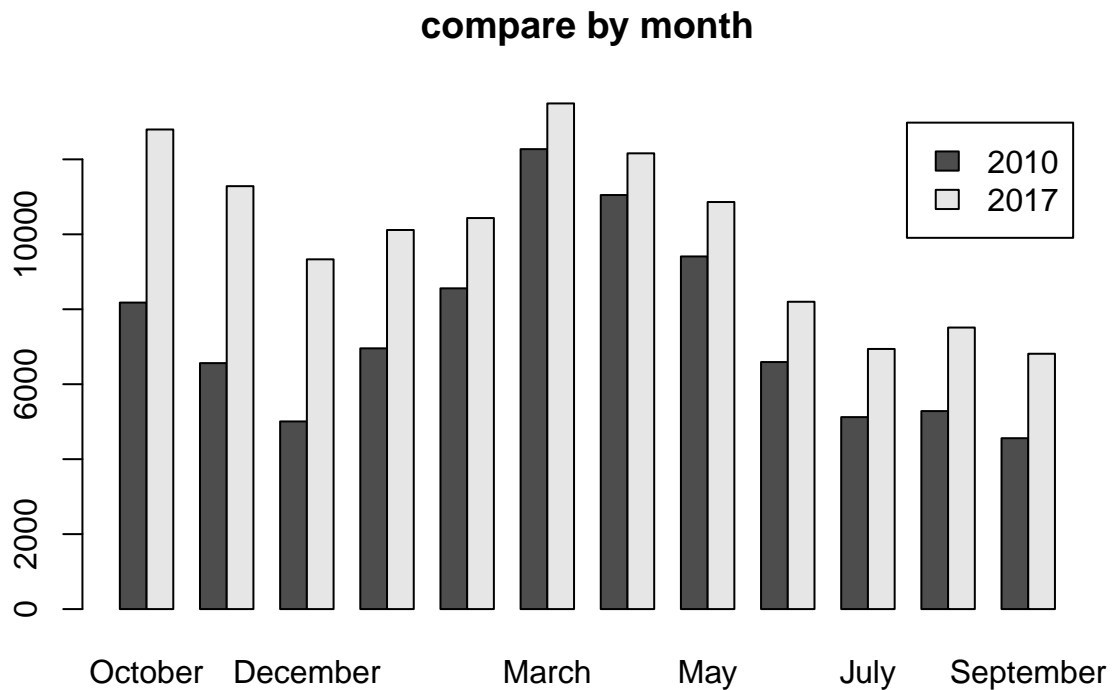
2010 Border Patrol Apprehensions by Sector


```
barplot(t(bpsector),beside = TRUE,legend.text = c("2010","2017"),main = "compare by sector")
```

compare by sector


```r
barplot(t(bpmonth),beside = TRUE,legend.text = c("2010","2017"),main = "compare by month")
```

compare by month

Analysis

From the bar plots, we can see that the sector with most apprehensions for 2010 is Tucson with total number of 212202 apprehensions. The sector with the most apprehensions for 2017 is Rio Grand Valley with total number of 137562 apprehensions. So, we use Waltch 2 Sample t test using these data. The p-value we got for this t-test is 0.06346. Since the p-value is larger than 5% significance level, we fail to reject the null hypothesis and conclude that there is no significant change in the maximum. We also did a Waltch 2 Sample t test on the 3 month periods with the most apprehensions in 2010 and 2017. The 3 month periods with the most apprehensions in 2010 is March, April, and May. The 3 month periods with the most apprehensions in 2017 is October, November, and December. The p-value we got for this t-test is 0.9788. Since the p-value is way larger than 5% significance level, we fail to reject the null hypothesis and conclude that there is no significant change in the maximum of the 3 consecutive monthly apprehensions.

Summary

```
ts2 <- as.vector(t(datas2))
#ts2
ts3 <- ts(ts2,frequency = 13,start = c(2000,10))
ts.plot(ts3,gpars = list(xlab="year",ylab="appre"))
datasmean <- apply(datas2,1,mean)
points(c(2001:2018), datasmean, pch = 19, col = 'red')
text(c(2001:2018), datasmean, c(2000:2017))
```