

Assignment 3

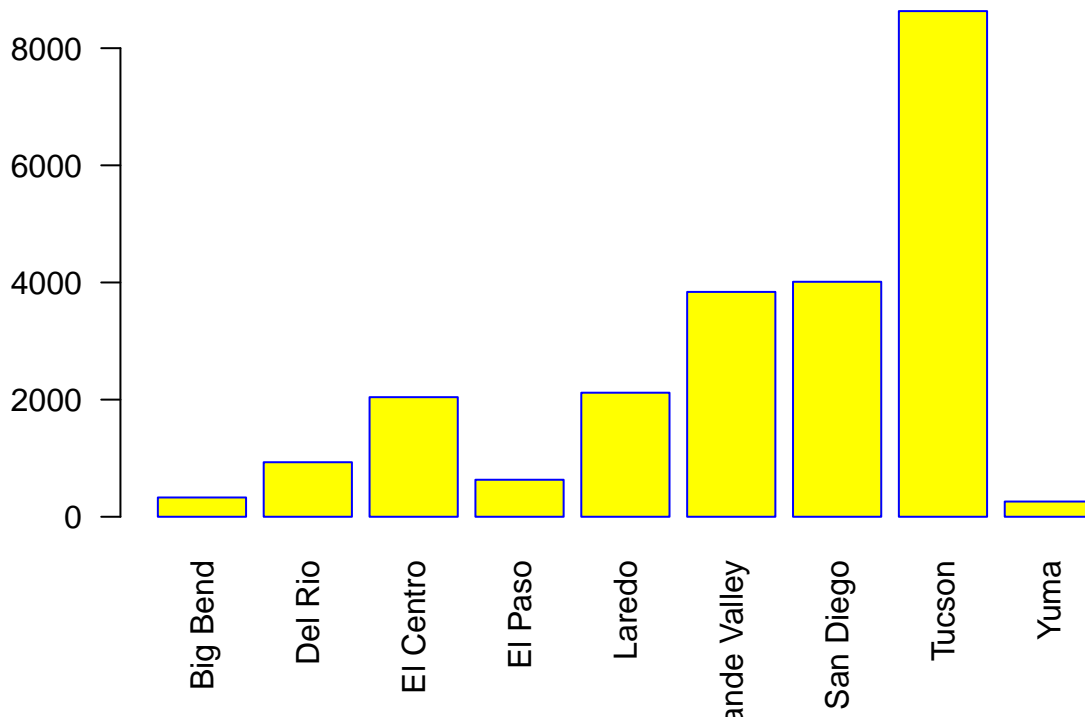
Steven Tran

February 7, 2018

```
## Reading in Data
BP2010 <- read.csv("BP Apprehensions 2010.csv", header = TRUE, stringsAsFactors = FALSE)
PB2017 <- read.csv("PB Apprehensions 2017.csv", header = TRUE, stringsAsFactors = TRUE)
PBmonthly <- read.csv("monthly_sum.csv", header = TRUE, stringsAsFactors = TRUE)
rownames(PBmonthly) <- PBmonthly[,1]

## Displaying Data By Sector
rownames(BP2010) <- BP2010[,1]
x <- barplot(BP2010[1:9,13], names.arg = rownames(BP2010)[1:9],
             las=2,
             axisnames=TRUE,
             main="2010 Border Patrol Apprehensions by Sector",
             border="blue",
             col="yellow")
```

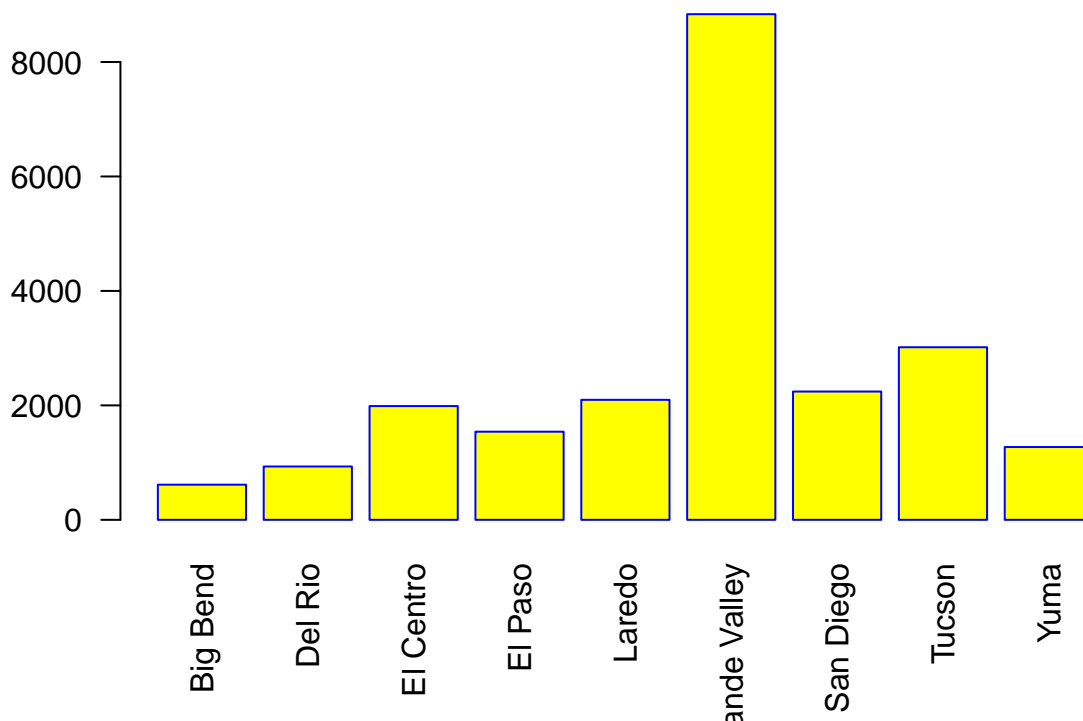
2010 Border Patrol Apprehensions by Sector



```
rownames(PB2017) <- PB2017[,1]
barplot(PB2017[1:9,13], names.arg = rownames(PB2017)[1:9],
        las=2,
        axisnames=TRUE,
```

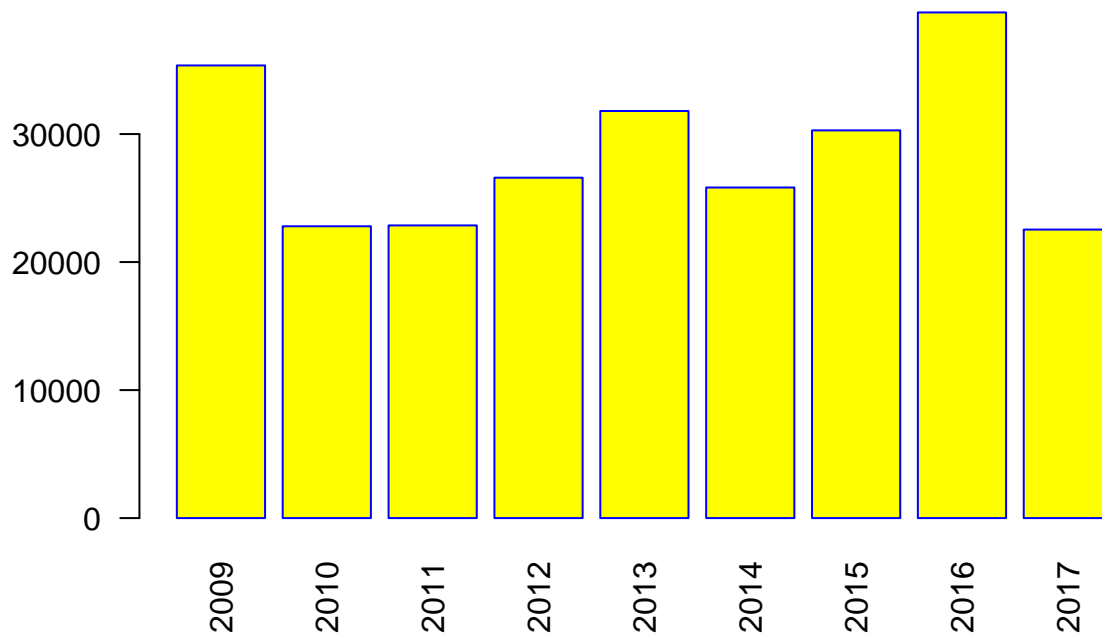
```
main="2017 Border Patrol Apprehensions by Sector",
border="blue",
col="yellow")
```

2017 Border Patrol Apprehensions by Sector



```
##Display Data By Year
barplot(rev(PBmonthly[1:9,13]), names.arg = rev(rownames(PBmonthly)[1:9]),
        las=2,
        axisnames=TRUE,
        main="Border Patrol Apprehensions by Year",
        border="blue",
        col="yellow")
```

Border Patrol Apprehensions by Year



From this t-test, we are sure that the two means are not statistically different with 95% confidence.

T-test for Tucson in 2010 and Rio Grande Valley in 2017, the sectors with the most apprehensions in

```
t2010 <- t(BP2010)
t2017 <- t(PB2017)
t.test(as.numeric(t2010[2:13, 8]), as.numeric(t2017[2:13, 6]))
```

```
##
## Welch Two Sample t-test
##
## data: as.numeric(t2010[2:13, 8]) and as.numeric(t2017[2:13, 6])
## 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
```

```
## Altering table to see sum of three month periods
m2010 <- subset(BP2010, select = -c(Sector))
m2010 <- rbind(m2010, colSums(m2010))
m2010 <- cbind(m2010, rowSums(m2010[1:3]))
m2010 <- cbind(m2010, rowSums(m2010[4:6]))
m2010 <- cbind(m2010, rowSums(m2010[7:9]))
m2010 <- cbind(m2010, rowSums(m2010[10:12]))
colnames(m2010) <- c(colnames(m2010)[1:12], "Oct-Dec", "Jan-Mar", "Apr-Jun", "Jul-Aug")
```

```

m2017 <- subset(PB2017, select = -c(Sector))
m2017 <- rbind(m2017, colSums(m2017))
m2017 <- cbind(m2017, rowSums(m2017[1:3]))
m2017 <- cbind(m2017, rowSums(m2017[4:6]))
m2017 <- cbind(m2017, rowSums(m2017[7:9]))
m2017 <- cbind(m2017, rowSums(m2017[10:12]))
colnames(m2017) <- c(colnames(m2017)[1:12], "Oct-Dec", "Jan-Mar", "Apr-Jun", "Jul-Aug")

```

We see here that both three period months of both years are not statistically different from one another.

T-test for Jan-Mar period for 2010 and Oct-Dec period for 2017

```

t.test(m2010[1:9, 4:6], m2017[1:9, 1:3])

```

```

##
## Welch Two Sample t-test
##
## data: m2010[1:9, 4:6] and m2017[1:9, 1:3]
## t = 0.044719, df = 51.904, p-value = 0.9645
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3719.530 3889.085
## sample estimates:
## mean of x mean of y
## 5145.741 5060.963

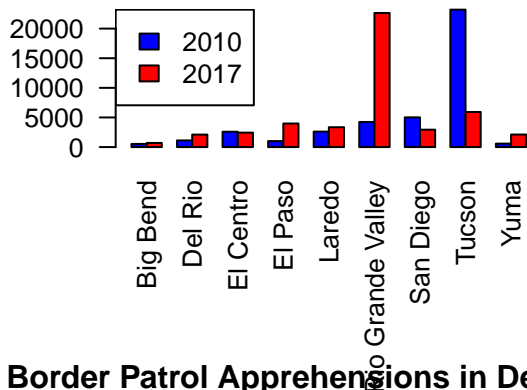
## Creates a 2 x 9 matrix from the 2010 data and 2017 for specified month
sideBySideMatrix <- function(month){
  matrix(c(BP2010[1:9,month], PB2017[1:9,month]), nrow = 2, byrow = TRUE)
}

## Creates the barplot for a given month
sideBySideBarPlot <- function(month, monthString){
  barplot(sideBySideMatrix(month), names.arg = rownames(BP2010),
    las=2,
    axisnames=TRUE,
    beside=TRUE,
    col=c("blue", "red"),
    main = paste("Border Patrol Apprehensions in", monthString, sep=" "))
  legend("topleft",
    c("2010", "2017"),
    fill = c("blue", "red"))
}

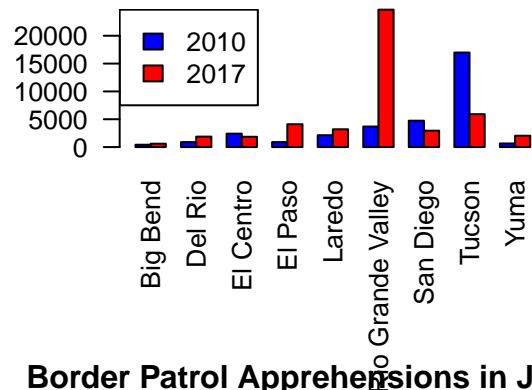
## Creates the side by side bar plots for each month
par(mfrow=c(2,2))
sideBySideBarPlot(2, "October")
sideBySideBarPlot(3, "November")
sideBySideBarPlot(4, "December")
sideBySideBarPlot(5, "January")

```

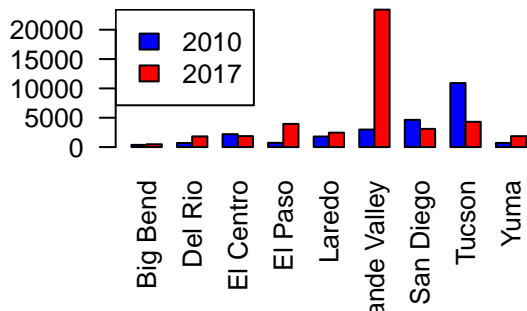
Border Patrol Apprehensions in October



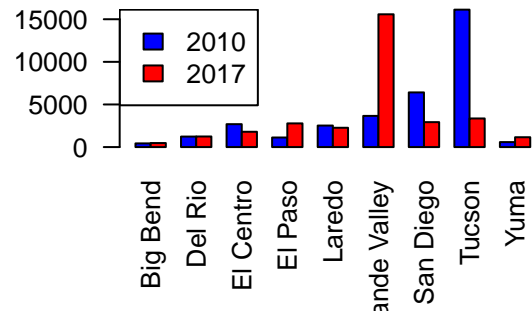
Border Patrol Apprehensions in November



Border Patrol Apprehensions in December

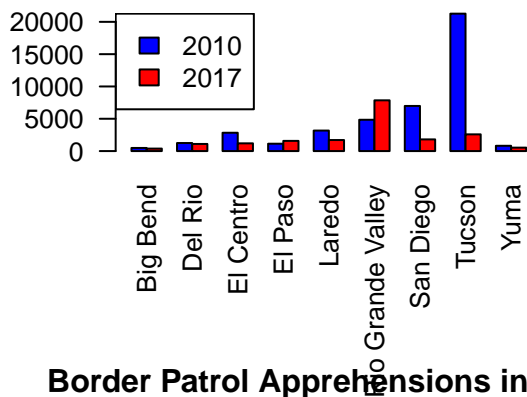


Border Patrol Apprehensions in January

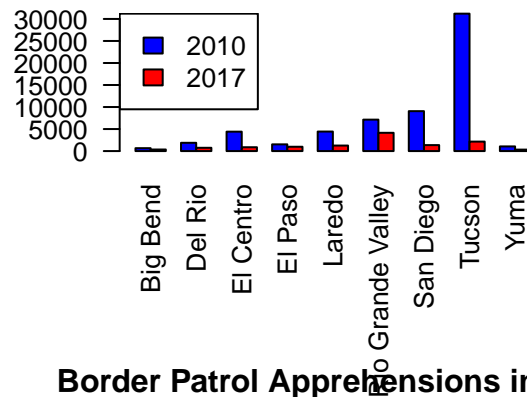


```
sideBySideBarPlot(6, "February")
sideBySideBarPlot(7, "March")
sideBySideBarPlot(8, "April")
sideBySideBarPlot(9, "May")
```

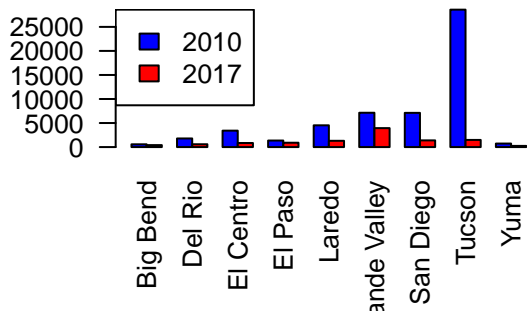
Border Patrol Apprehensions in February



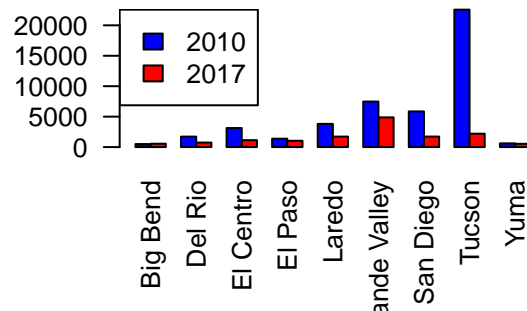
Border Patrol Apprehensions in March



Border Patrol Apprehensions in April

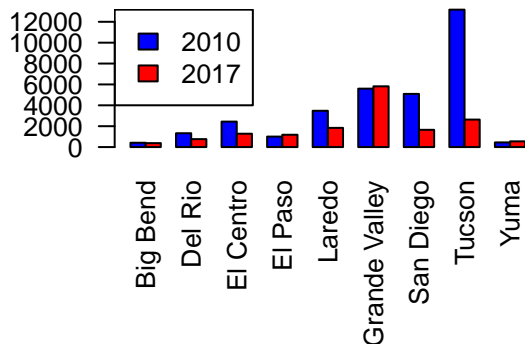


Border Patrol Apprehensions in May

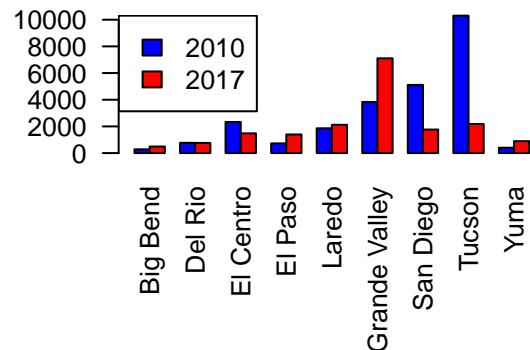


```
sideBySideBarPlot(10, "June")
sideBySideBarPlot(11, "July")
sideBySideBarPlot(12, "August")
sideBySideBarPlot(13, "September")
```

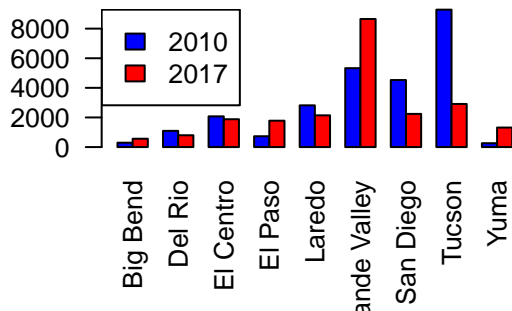
Border Patrol Apprehensions in June



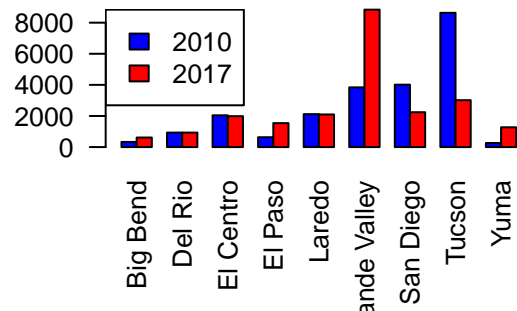
Border Patrol Apprehensions in July



Border Patrol Apprehensions in August



Border Patrol Apprehensions in September



```
par(mfrow=c(1,1))
```

```
#overlays the plots of 2 lines 1 blue and 1 red; for use with the apprehensions by month of 2010 and 2017
twolineplot <- function(){
  x <- factor(2:13, labels = c("October", "November", "December", "January", "February", "March", "April", "May", "June", "July", "August", "September", "October"))
  t2010 <- as.numeric(PBmonthly[1,2:13])
  t2017 <- as.numeric(PBmonthly[8, 2:13])
  plot(t2010 ~ x, type="p", xlab = "Month", ylab = "Apprehensions", main = "Apprehensions By Month", ylim=c(0, 12000))
  lines(t2010 ~ x, col="blue")
  lines(t2017 ~ x, col="red")
  legend("topleft",
        c("2010", "2017"),
        fill = c("blue", "red"))
}
```

Here we can see the plots of 2010 and 2017 monthly overlaid. As you can tell, the number of apprehensions rose steeply from the beginning to the April 2017, decreasing thereafter, whilst there was a fairly steady decline in apprehensions in 2010 until April 2010, followed by a slight incline.

```
twolineplot()
```

Apprehensions By Month

