2010 and 2017 CBP Apprehensions Analysis

Janie Briglio, Ben Gaudiosi, Kayla Ippongi

2/19/2018

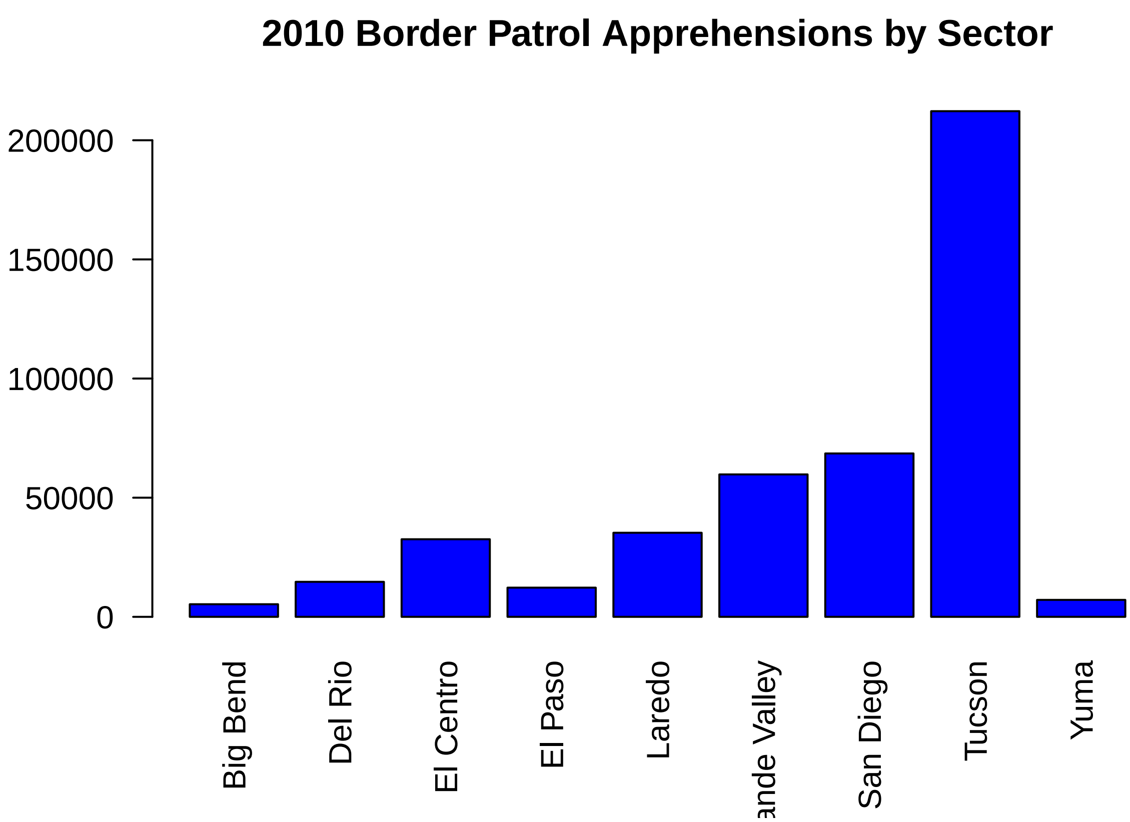
Bar Graph Comparison of Border Partol Apprehensions

entering the data for 2010 and making a column and row for the total for the locations and the months

```
data1 <- read.csv("BP apprehensions 2010.csv", header = TRUE, stringsAsFactors = FALSE)
rownames(data1) <- data1[,1]
x <- subset(data1, select= -c(Sector))
x <- rbind(x, colSums(x))
rownames(x) <- c(rownames(x)[-length(rownames(x))], "Total")
x <- cbind(x,rowSums(x))
colnames(x) <- c(colnames(x)[-length(colnames(x))], "Total")
```

bar graph of the totals for for each location

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

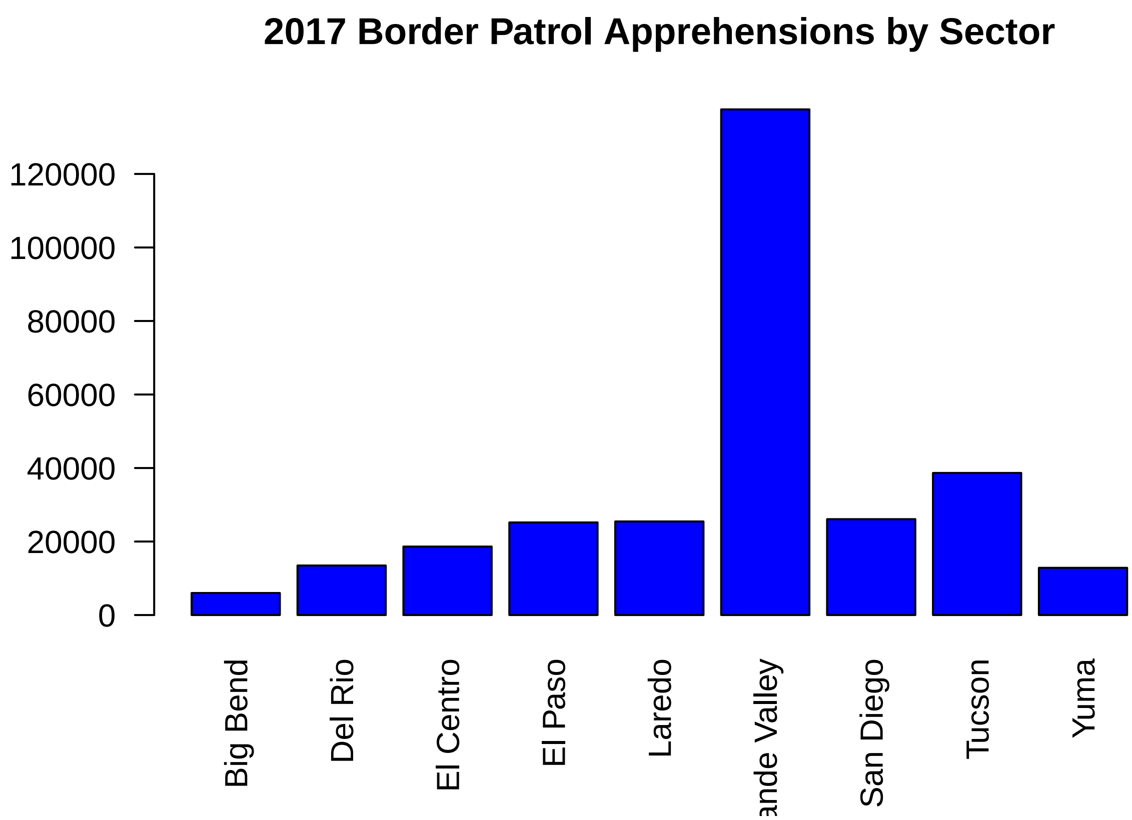


entering the data for 2017 and making a column and row for the total for the locations and the months

```r
data2 <- read.csv("PB Apprehensions 2017.csv", header = TRUE, stringsAsFactors = FALSE)
data2 = head(data2[1:13],-1)
rownames(data2) <- data2[,1]
y <- subset(data2, select= -c(Sector))
y <- rbind(y, colSums(y))
rownames(y) <- c(rownames(y)[-length(rownames(y))], "Total")
y <- cbind(y, rowSums(y))
colnames(y) <- c(colnames(y)[-length(colnames(y))], "Total")
```

bar graph for the totals of each location

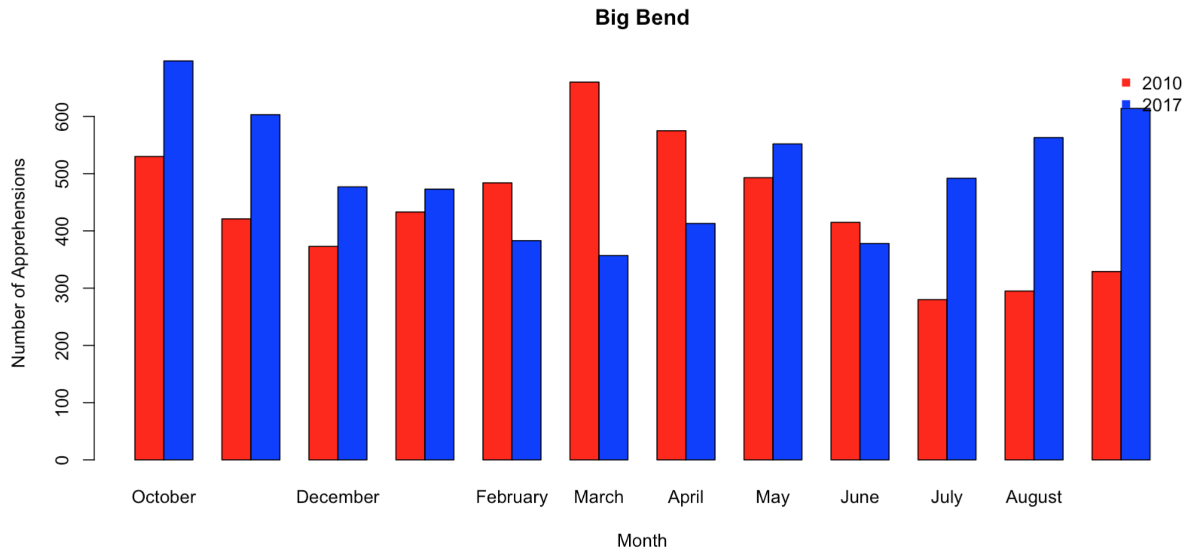
```r
barplot(y[1:9, 13], names.arg = rownames(y)[1:9],
        las=2,
        axisnames=TRUE,
        main="2017 Border Patrol Apprehensions by Sector",
        border="black",
        col="blue")
```



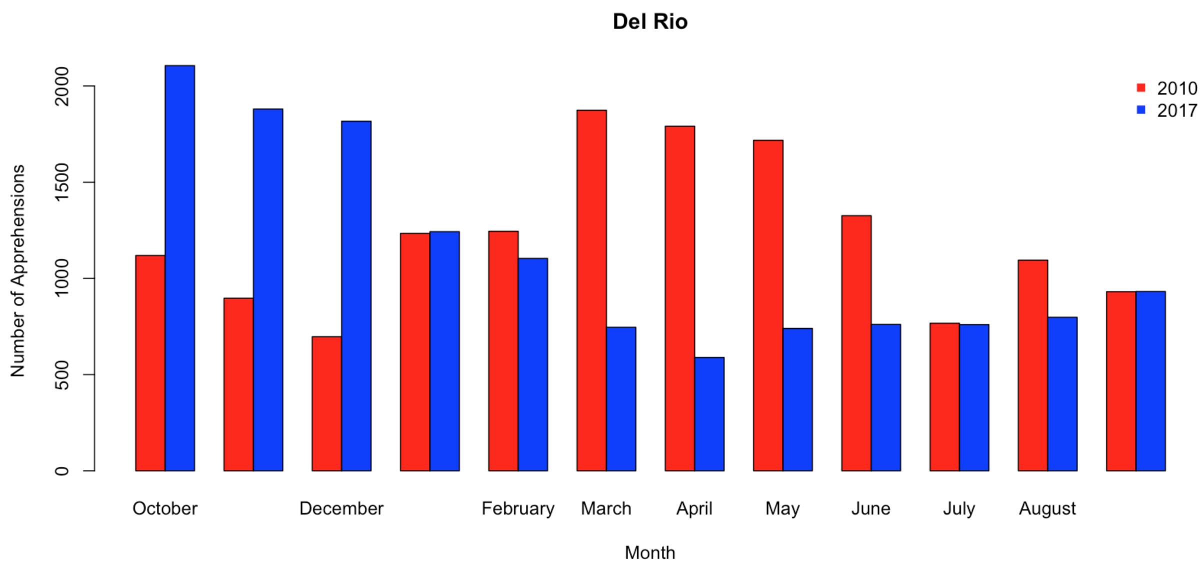
2010 vs 2017 Comparison by Sector

The following bargraphs below compares the 2010 and 2017 CBP data by sector and month. It is apparent that like the CNN article identified, there has been a significant drop in apprehensions compared to 2010, with the exception of the El Paso sector.

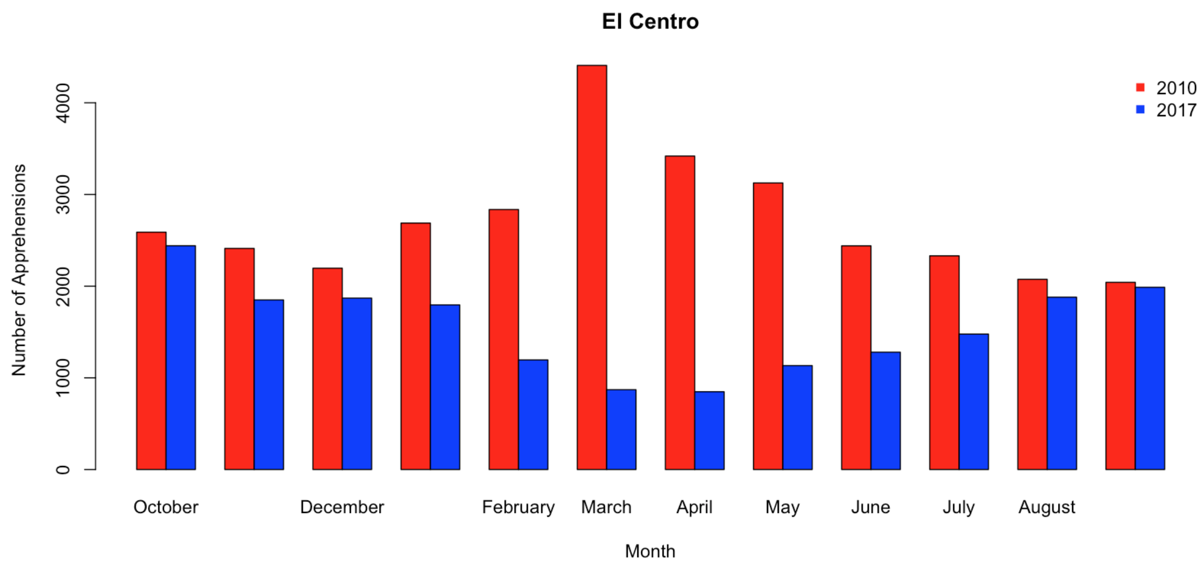
Big Bend Comparison



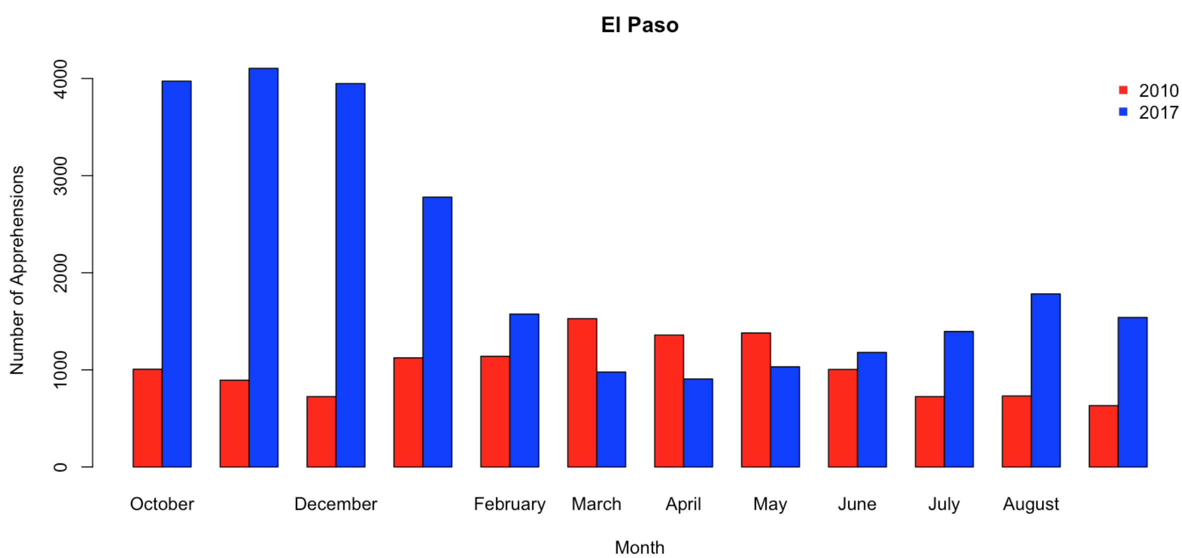
Del Rio Comparison



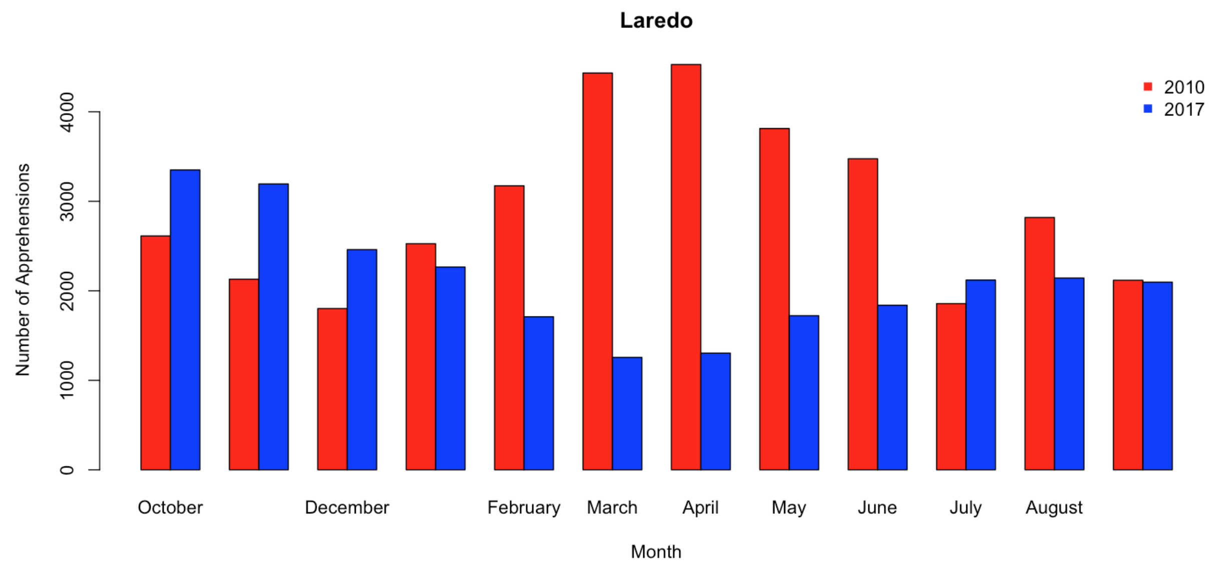
El Centrio Comparison



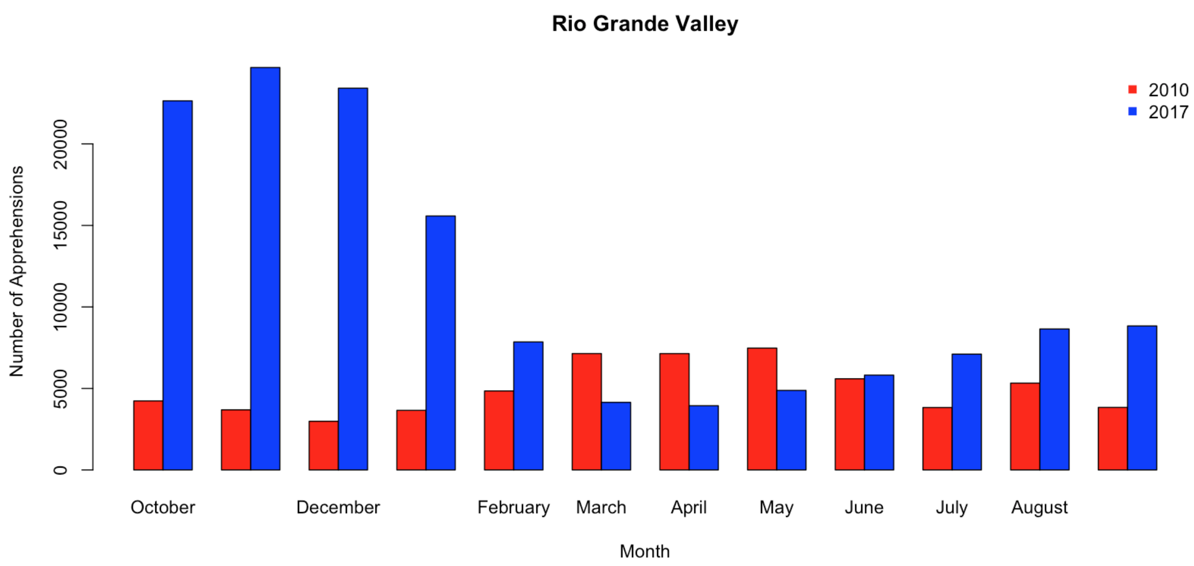
El Paso Comparison



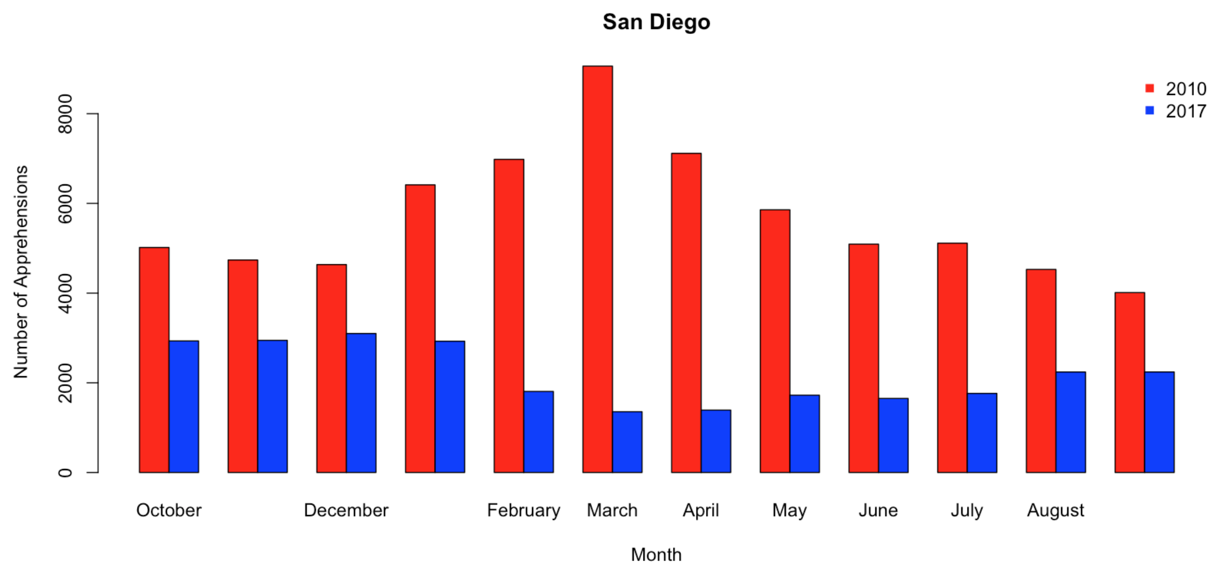
Laredo Comparison



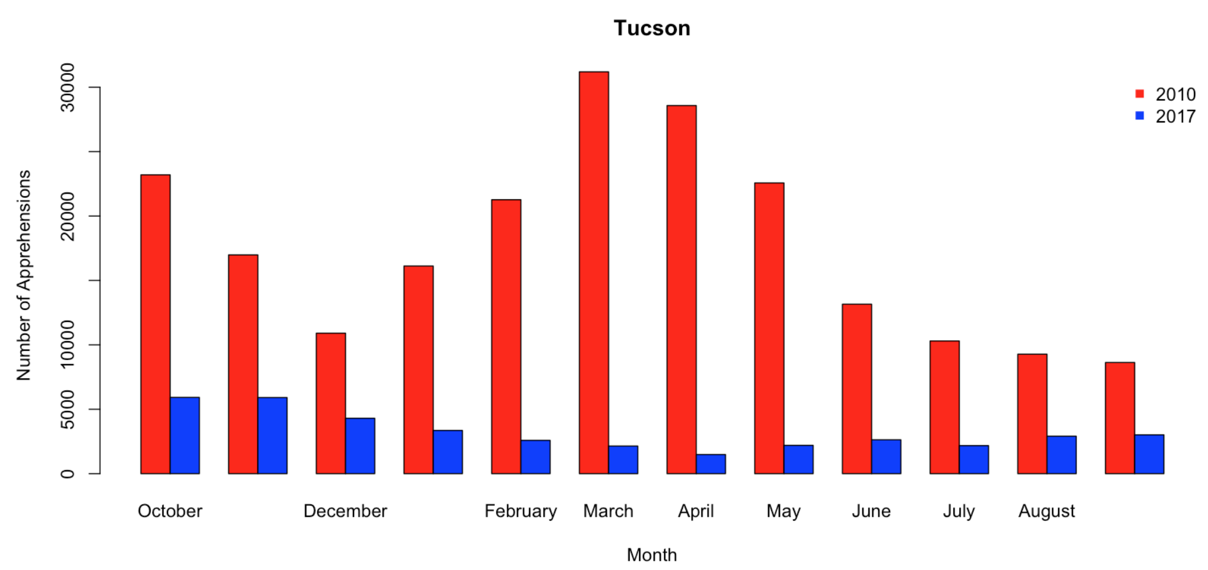
Rio Grande Valley Comparison



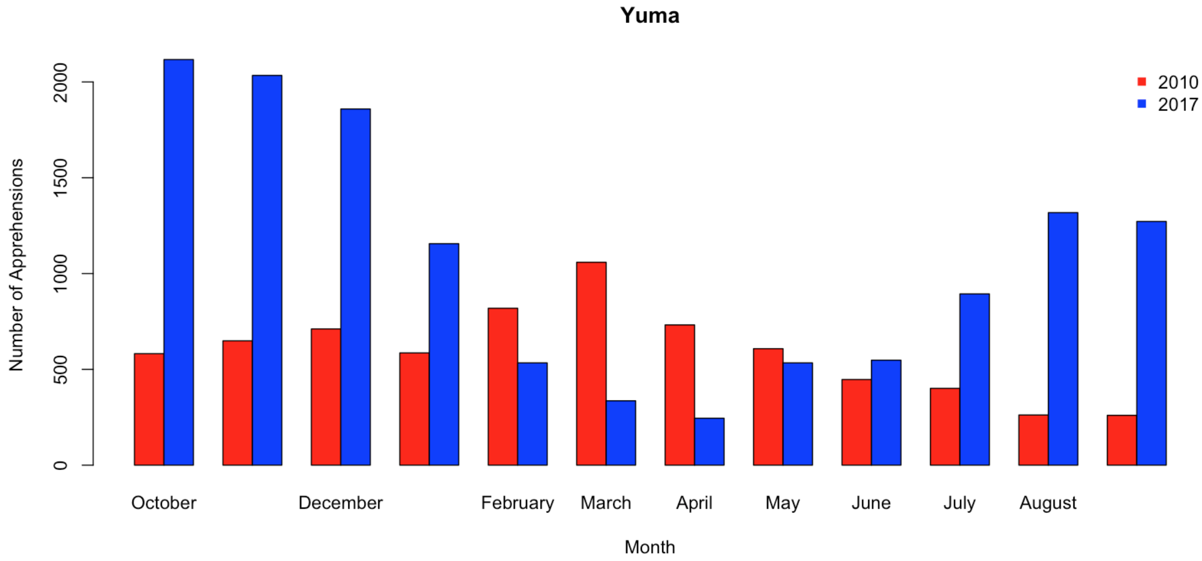
San Diego Comparison



Tuscon Comparison



Yuma Comparison

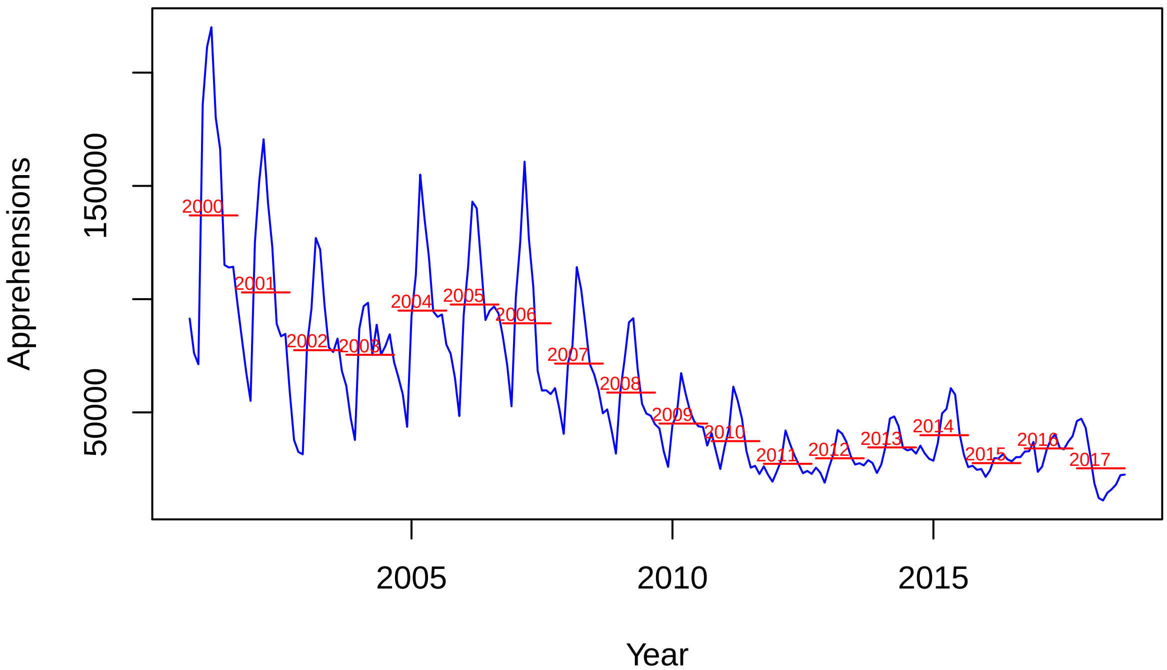


Time Series

This time series graph represents the total yearly apprehensions from 2000-2017. The trend shows the total apprhensions slowly decreasing as the years move towards 2017.

```
PBSum <- read.csv("PB monthly summaries.csv")
ts.plot(ts(as.vector(rev(t(rev(PBSum[-1])))), start = c(2000,10), frequency=12),
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[18,-1])),each=12)  )  )), start = c(2000,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[17,-1])),each=12)  )  )), start = c(2001,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[16,-1])),each=12)  )  )), start = c(2002,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[15,-1])),each=12)  )  )), start = c(2003,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[14,-1])),each=12)  )  )), start = c(2004,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[13,-1])),each=12)  )  )), start = c(2005,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[12,-1])),each=12)  )  )), start = c(2006,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[11,-1])),each=12)  )  )), start = c(2007,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[10,-1])),each=12)  )  )), start = c(2008,10), frequen
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[9,-1])),each=12)  )  )), start = c(2009,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[8,-1])),each=12)  )  )), start = c(2010,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[7,-1])),each=12)  )  )), start = c(2011,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[6,-1])),each=12)  )  )), start = c(2012,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[5,-1])),each=12)  )  )), start = c(2013,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[4,-1])),each=12)  )  )), start = c(2014,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[3,-1])),each=12)  )  )), start = c(2015,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[2,-1])),each=12)  )  )), start = c(2016,10), frequency
      ts(as.vector(  t(  rev(rep(c(rowMeans(PBSum[1,-1])),each=12)  )  )), start = c(2017,10), frequency

      gpars=list(xlab="Year", ylab="Apprehensions", lty=1, col=c('blue',rep(c('red'),times=18), ltw=2))
                        );text(rev(PBSum[,1])+1,rev(c(rowMeans(PBSum[-1])))+4000,labels=paste(rev(PBSu
```



TTests

By looking at the resulting T-test values, March April and May have the highest T-test values which matches the 2010 CBP total apprehensions data - as March, April and May have the highest total apprensions in 2010 but the lowest in 2017. Comparatively, October November and December have the lowest T-test values, going into the negative range. This 3 month period has the highest total apprensions in 2017 but is not too different from it's 2010 data, which is why the t-test values are closer to zero.

entering data as a matrix and removing non numerical data

```
data1 <- read.csv("BP apprehensions 2010.csv", header = TRUE, stringsAsFactors = FALSE)
data2 <- read.csv("PB Apprehensions 2017.csv", header = TRUE, stringsAsFactors = FALSE)
data2 <- head(data2[1:13],-1)

x <- subset(data1, select= -c(Sector))
y <- subset(data2, select= -c(Sector))
```

running comparisons between the months for each location ##October Comparison:

```
t.test(data1$October, data2$October, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$October and data2$October
## t = -0.19561, df = 8, p-value = 0.8498
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7522.671  6346.226
## sample estimates:
## mean of the differences
##               -588.2222
```

November Comparison:

```r
t.test(data1$November, data2$November, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$November and data2$November
## t = -0.57437, df = 8, p-value = 0.5815
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8021.535  4822.424
## sample estimates:
## mean of the differences
##               -1599.556
```

December Comparison:

```r
t.test(data1$December, data2$December, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$December and data2$December
## t = -0.81876, df = 8, p-value = 0.4366
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7724.951  3676.729
## sample estimates:
## mean of the differences
##               -2024.111
```

Janurary Comparison:

```r
t.test(data1$January, data2$January, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$January and data2$January
## t = 0.16907, df = 8, p-value = 0.8699
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4505.171  5218.060
## sample estimates:
## mean of the differences
##                356.4444
```

Feburary Comparison:

```r
t.test(data1$February, data2$February, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$February and data2$February
## t = 1.2559, df = 8, p-value = 0.2446
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2233.163  7574.497
## sample estimates:
## mean of the differences
##                2670.667
```

March Comparison:

```r
t.test(data1$March, data2$March, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$March and data2$March
## t = 1.793, df = 8, p-value = 0.1107
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1563.07 12488.85
## sample estimates:
## mean of the differences
##                5462.889
```

April Comparison:

```r
t.test(data1$April, data2$April, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$April and data2$April
## t = 1.7273, df = 8, p-value = 0.1224
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1642.025 11444.247
## sample estimates:
## mean of the differences
##                4901.111
```

May Comparison:

```r
t.test(data1$May, data2$May, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$May and data2$May
## t = 1.6862, df = 8, p-value = 0.1303
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1328.549  8556.549
## sample estimates:
## mean of the differences
##                    3614
```

June Comparison:

```r
t.test(data1$June, data2$June, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$June and data2$June
## t = 1.6274, df = 8, p-value = 0.1423
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -781.4508 4529.8953
## sample estimates:
## mean of the differences
##                1874.222
```

July Comparison:

```r
t.test(data1$July, data2$July, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$July and data2$July
## t = 0.7663, df = 8, p-value = 0.4655
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1656.988  3306.321
## sample estimates:
## mean of the differences
##                824.6667
```

August Comparison:

```r
t.test(data1$August, data2$August, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$August and data2$August
## t = 0.51189, df = 8, p-value = 0.6226
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1607.187  2524.298
## sample estimates:
## mean of the differences
##                458.5556
```

September Comparison:

```r
t.test(data1$September, data2$September, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  data1$September and data2$September
## t = 0.031118, df = 8, p-value = 0.9759
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2103.832  2161.387
## sample estimates:
## mean of the differences
##                28.77778
```

running a comparison between the totals of each location in both years

```r
total2010<- rowSums(x[1:12])
total2017 <- rowSums(y[1:12])
t.test(total2010, total2017, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  total2010 and total2017
## t = 0.71295, df = 8, p-value = 0.4961
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -35705.54  67664.43
## sample estimates:
## mean of the differences
##                15979.44
```