US-Mexico Border Apprehensions

Kevin Chan, Brianna Kincaid, Lauren Vanvalkenburg

Over the past 17 years, apprehensions at the US-Mexico border have reached a historic low. Data on U.S. Border Patrol Southwest Border Apprehensions released by the US Customs and Border Protrection (CBP) in 2017 revealed that apprehensions of individuals attempting to illegally cross the Southern border has decreased from 447,731 in 2010 to 303,196 in 2017. Some credit the decrease in apprehensions to President Donald Trump's enforcement of immigration laws. In April 2017, there were 11,129 apprehensions at the Southwest border, a 62% drop from the previous April. February and March also saw a decrease in apprehensions, which has not been the case since 2000.

1 Data Import and Cleaning

1.1 Import Data

```
> A2010 <- read.csv( "BP Apprehensions 2010.csv" , header = TRUE, stringsAsFactors = FALSE)
> A2017 <- read.csv("PB Apprehensions 2017.csv", header = TRUE, stringsAsFactors = FALSE)
```

	October	November	December	January	February	March	April	May	June	July	August	September
Big Bend	530.00	421.00	373.00	433.00	484.00	660.00	575.00	493.00	415.00	280.00	295.00	329.00
Del Rio	1119.00	897.00	697.00	1234.00	1245.00	1874.00	1791.00	1718.00	1326.00	767.00	1095.00	931.00
El Centro	2589.00	2412.00	2196.00	2688.00	2836.00	4408.00	3419.00	3126.00	2440.00	2331.00	2075.00	2042.00
El Paso	1007.00	894.00	725.00	1124.00	1140.00	1528.00	1359.00	1380.00	1005.00	725.00	732.00	632.00
Laredo	2613.00	2130.00	1802.00	2526.00	3173.00	4433.00	4528.00	3813.00	3475.00	1857.00	2819.00	2118.00
Rio Grande	4236.00	3688.00	2987.00	3658.00	4845.00	7141.00	7139.00	7477.00	5595.00	3832.00	5329.00	3839.00
San Diego	5017.00	4738.00	4636.00	6413.00	6982.00	9061.00	7115.00	5858.00	5092.00	5113.00	4528.00	4012.00
Tucson	23197.00	16986.00	10907.00	16122.00	21266.00	31197.00	28579.00	22572.00	13160.00	10303.00	9280.00	8633.00
Yuma	582.00	649.00	711.00	586.00	819.00	1059.00	732.00	608.00	447.00	401.00	262.00	260.00

Table 1: CBP Data for 2010

	October	November	December	January	February	March	April	May	June	July	August	September
Big Bend	697.00	603.00	477.00	473.00	383.00	357.00	413.00	552.00	378.00	492.00	563.00	614.00
Del Rio	2106.00	1880.00	1817.00	1243.00	1104.00	746.00	589.00	740.00	761.00	760.00	798.00	932.00
El Centro	2441.00	1850.00	1870.00	1796.00	1196.00	871.00	849.00	1134.00	1280.00	1478.00	1880.00	1988.00
El Paso	3973.00	4105.00	3948.00	2779.00	1575.00	978.00	906.00	1032.00	1180.00	1395.00	1782.00	1540.00
Laredo	3350.00	3194.00	2460.00	2265.00	1710.00	1256.00	1304.00	1722.00	1839.00	2120.00	2143.00	2097.00
Rio Grande	22642.00	24686.00	23418.00	15580.00	7855.00	4147.00	3942.00	4882.00	5817.00	7107.00	8650.00	8836.00
San Diego	2934.00	2947.00	3099.00	2927.00	1808.00	1356.00	1392.00	1724.00	1652.00	1764.00	2241.00	2242.00
Tucson	5924.00	5912.00	4303.00	3357.00	2589.00	2148.00	1487.00	2199.00	2632.00	2177.00	2913.00	3016.00
Yuma	2117.00	2034.00	1859.00	1156.00	534.00	336.00	245.00	534.00	548.00	894.00	1318.00	1272.00

Table 2: CBP Data for 2017

The CBP data we analyze here includes total apprehensions broken down by month and sector for both 2010 and 2017. The data includes statistics for Big Bend, Del Rio, El Centro, El Paso, Laredo, Rio Grande Valley, San Diego, Tucson, and Yuma.

1.2 Clean Data

Here we are organizing the data with column and row totals with the appropriate row and column names.

1.2.1 2010 Data

Use the strings in column 1 as row names.

```
> rownames(A2010) <- A2010[,1]
```

Drop column 1

```
> A2010 <-  subset(A2010, select= -c(Sector))
```

Use rbind to add the sum across columns to the dataframe.

```
> A2010 <- rbind(A2010, colSums(A2010))
```

Drop that name that rbind assigns and rename the row that contains the column totals as ”Total”

```
> rownames(A2010) <- c(rownames(A2010)[-length(rownames(A2010))], "Total")
```

Use cbind to add the sum across rows to the dataframe.

```
> A2010 <- cbind(A2010,rowSums(A2010))
```

Drop that name that cbind assigns and rename the column that contains the row totals as ”Total”

```
> colnames(A2010) <- c(colnames(A2010)[-length(colnames(A2010))], "Total")
```

1.2.2 2017 Data

Use the strings in column 1 as row names.

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

Drop column 1

```
> A2017 <-  subset(A2017, select= -c(Sector))
```

Use rbind to add the sum across columns to the dataframe.

```
> A2017 <- rbind(A2017, colSums(A2017))
```

Drop that name that rbind assigns and rename the row that contains the column totals as ”Total”

```
> rownames(A2017) <- c(rownames(A2017)[-length(rownames(A2017))], "Total")
```

Use cbind to add the sum across rows to the dataframe.

```
> A2017 <- cbind(A2017,rowSums(A2017))
```

Drop that name that cbind assigns and rename the column that contains the row totals as ”Total”

```
> colnames(A2017) <- c(colnames(A2017)[-length(colnames(A2017))], "Total")
```

2 2010 and 2017 Compared

2.1 By Sector

Here we compare the total apprehensions in each sector for 2010 and 2017. Extracting sector data for 2010:

```
> year2010s <- t(as.data.frame(matrix(A2010[1:9,13])))
> colnames(year2010s) <- rownames(A2010[1:9,])
> colnames(year2010s)[6] <- "Rio Grande"
```

Extracting sector data for 2017:

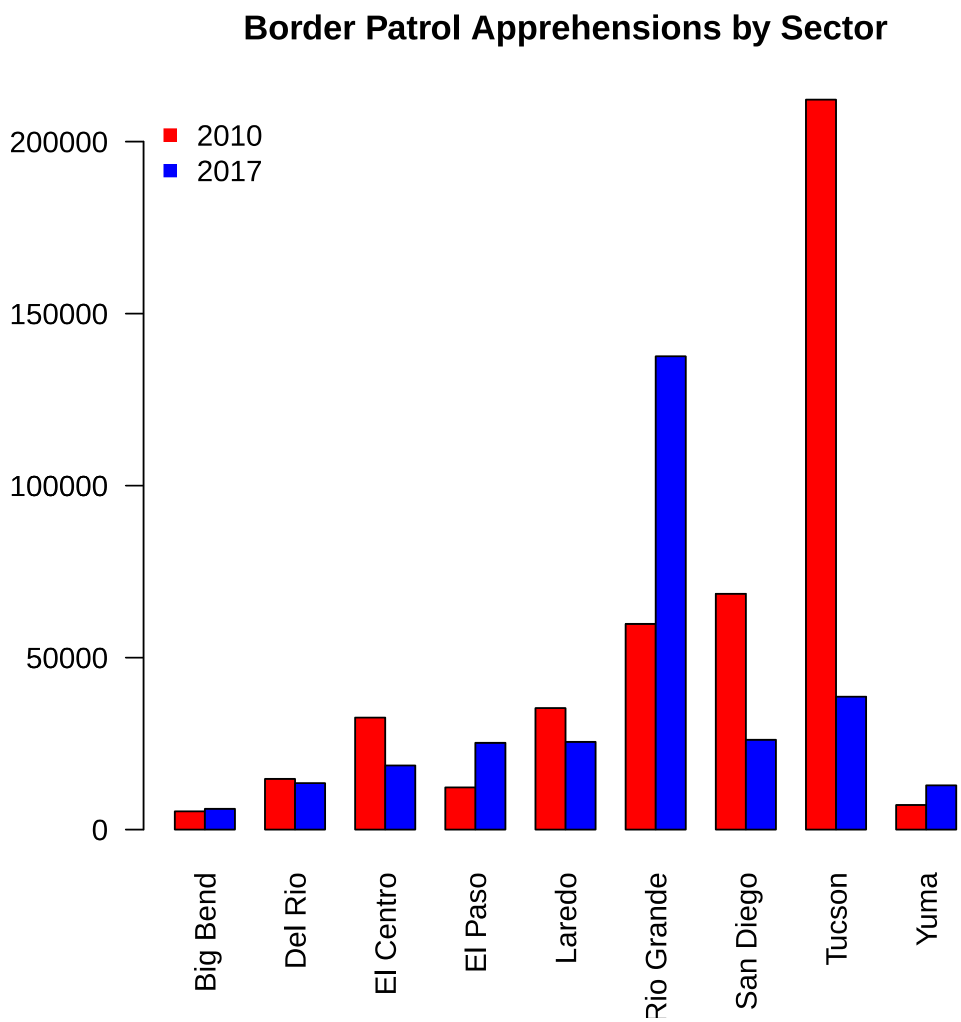
```
> year2017s <- t(as.data.frame(matrix(A2017[1:9,13])))
> colnames(year2017s) <- rownames(A2017[1:9,])
> colnames(year2017s)[6] <- "Rio Grande"
```

Combining sector data for 2010 and 2017:

```
> year2010_17s <- rbind(year2010s, year2017s)
> row.names(year2010_17s) <- c("2010", "2017")
```

Creating a bar plot:

```
> barplot(as.matrix(year2010_17s), beside = TRUE, col = c("red", "blue"), bty="n",las=2)
> legend("topleft", c("2010","2017"), pch=15,  col=c("red","blue"),  bty="n")
> title("Border Patrol Apprehensions by Sector")
```



The sector with the largest number of apprehensions for the year changed from Tucson in 2010 to Rio Grande Valley in 2017.

2.2 By Month

Here we compare the total apprehensions for each month in 2010 and 2017.

Extracting monthly data for 2010:

```
> year2010m <- t(as.data.frame(matrix(unname(t(A2010)[1:12,10]))))
> colnames(year2010m) <- colnames(A2010[1:12])
```

Extracting monthly data for 2017:

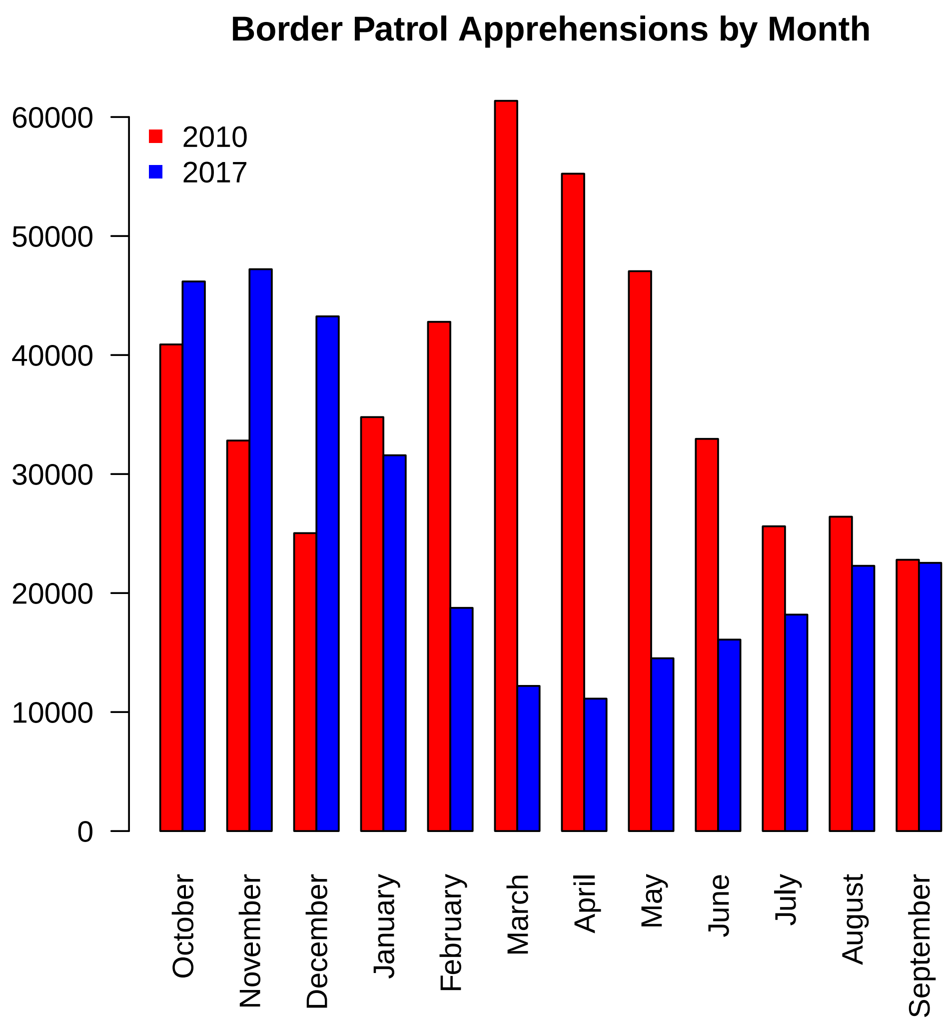
```
> year2017m <- t(as.data.frame(matrix(unname(t(A2017)[1:12,10]))))
> colnames(year2017m) <- colnames(A2017[1:12])
```

Combining monthly data for 2010 and 2017:

```
> year2010_17m <- rbind(year2010m, year2017m)
> row.names(year2010_17m) <- c("2010", "2017")
```

Creating a bar plot:

```
> barplot(as.matrix(year2010_17m), beside = TRUE, col = c("red", "blue"), bty="n",las=2)
> legend("topleft", c("2010","2017"), pch=15,  col=c("red","blue"),  bty="n")
> title("Border Patrol Apprehensions by Month")
```



As can be seen in the data, the trend throughout the year regarding total apprehensions across sectors changed drastically from 2010 to 2017. The two months with the highest apprehensions in 2010 (March and April) had the least amount of apprehensions in 2017. Furthermore, the month with the highest number of apprehensions in 2010 had more than 10,000 more apprehensions that the month ith the greatest umber of apprehensions in 2017.

2.3 t-test

2.3.1 A comparison between the sector with most apprehensions for 2010 and the sector with most apprehensions in 2017

First we must extract the total apprehensions by sector for 2010.

```
> Sector_Totals_2010 <- A2010[1:9,13]
> names(Sector_Totals_2010) <- rownames(A2010[1:9,])
```

Then we do the same thing for 2017.

```
> Sector_Totals_2017 <- A2017[1:9,13]
> names(Sector_Totals_2017) <- rownames(A2017[1:9,])
```

Then we use these totals to determine the indeces of the sector with the greatest number of apprehensions.

```
> MA_2017_index <- which(Sector_Totals_2017 == max(Sector_Totals_2017))
> MA_2010_index <- which(Sector_Totals_2010 == max(Sector_Totals_2010))
```

We use these indeces to extract the monthly data for the sectors with the greatest number of apprehensions in 2010 and 2017.

```
> MA_2017 <- A2017[MA_2017_index,1:12]
> MA_2010 <- A2010[MA_2010_index,1:12]
```

We compared the sector with the most apprehensions for 2010 with the sector with most apprehensions in 2017 at a 5-percent significance level. The sectors with the highest apprehensions in 2010 and 2017 are Tuscan and Rio Grande Valley, respectively.

```
> t.test(MA_2010,MA_2017)
```

Welch Two Sample t-test
data: MA2010 and MA2017
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: 17683.5
mean of y: 11463.5

The t test resulted in a p-value of .06346, which at 5-percent significance means the difference between the two sectors is not statistically significant.

We also decided to compare 2010 with 2017 Tuscan and 2010 with 2017 Rio Grande Valley.

```
> t.test(MA_2010,A2017[MA_2010_index,1:12])
```

Welch Two Sample t-test
data: Tuscan 2010 and Tuscon 2017
t = 6.4303
df = 11.781
p-value = 3.545e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: 9551.716 19372.450

sample estimates:
mean of x: 17683.500
mean of y: 3221.417

The t test between 2010 and 2017 Tuscan resulted in a p-value of .00003545, which is statistically significant. Which means there is a considerable difference between the amount of apprehensions in Tuscan from 2010 to 2017.

```
> t.test(MA_2017,A2010[MA_2017_index,1:12])
```

Welch Two Sample t-test
data: Rio Grande Valley 2010 and Rio Grande Valley 2017
t = 2.7789
df = 11.846
p-value = 0.01686
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: 1392.65 11573.35

sample estimates:
mean of x: 11463.5
mean of y: 4980.5

The t test between 2010 and 2017 Rio Grande Valley resulted in a p-value of .01686, which is also statistically significant. There is a noticeable difference between the number of apprehensions in Rio Grande Valley from 2010 to 2017.

2.3.2 A comparison between the 3 month periods with the most apprehensions in 2010 and 2017

```
> col <- c("Oct-Dec", "Jan-Mar", "Apr-Jun","Jul-Sep")
> Monthly_Totals_2010 <- (t(A2010)[1:12,10])
> A2010_3 <- rbind(sum(Monthly_Totals_2010[1:3]),sum(Monthly_Totals_2010[4:6]),sum(Monthly_Totals_2010[7:9
> Monthly_Totals_2017 <- (t(A2017)[1:12,10])
> A2017_3 <- rbind(sum(Monthly_Totals_2017[1:3]),sum(Monthly_Totals_2017[4:6]),sum(Monthly_Totals_2017[7:9
```

In 2010, the three month period with the most apprehensions is January, February and March. In 2017, the three month period with the most apprehensions is October, November and December. We ran a t test comparing these three month periods with a 5-percent significance level.

```
> t.test(Monthly_Totals_2010[4:6],Monthly_Totals_2017[1:3])
```

Welch Two Sample t-test
data: Jan/Feb/March 2010 and Oct/Nov/Dec 2017
t = 0.095848
df = 2.0908
p-value = 0.932
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: -32101.2 33627.2

sample estimates:
mean of x: 46311.67
mean of y: 45548.67

The t test resulted in a p-value of .932, which is not statistically significant. There is not a significant difference in the amount of apprehensions in January, February and March of 2010 and October, November and December of 2017.

3 Overall Trends

3.1 Data Cleaning

Here we import monthly summaries of apprehensions across all sectors for years 2000 to 2017.

```
> A2000.2017 <- read.csv("PB monthly summaries.csv", header = TRUE, stringsAsFactors = FALSE)
```

Use the strings in column 1 as row names.

```
> rownames(A2000.2017) <- A2000.2017[,1]
```

Drop column 1

```
> A2000.2017 <-  subset(A2000.2017, select= -c(year))
```

Reorder the rows so that the years list from 2000 to 2017.

```
> A2000.2017 <- A2000.2017[18:1,]
```

	October	November	December	January	February	March	April	May	June	July	August	September
2000	91410	76196	71252	185979	211328	220063	180050	166296	115093	113956	114312	97744
2001	82632	67709	55081	125090	152229	170580	142813	122927	89131	83602	84648	59276
2002	37812	32506	31501	79793	95724	126992	121921	97424	78655	76661	82557	68263
2003	61792	47731	37824	86925	96869	98399	75359	88690	75530	79284	84486	72176
2004	65391	57894	43614	92521	110669	154981	135468	118726	94590	92165	93246	80017
2005	75913	65135	48406	93020	113775	143048	140062	115823	90786	94954	96733	93741
2006	83557	70975	52673	101195	125046	160696	126538	105450	68366	59641	59751	58084
2007	60713	51594	40527	71934	79268	114137	104465	88504	71338	66782	59795	49581
2008	51339	42209	31802	59028	73483	89770	91566	69233	53854	49472	48541	44708
2009	42938	32780	25947	44502	49211	67342	58493	50884	46044	43843	43522	35359
2010	40890	32815	25034	34784	42790	61361	55237	47045	32955	25609	26415	22796
2011	26165	22405	19429	23926	28786	42014	36251	31236	27166	23170	24166	22863
2012	25612	23368	18983	25714	31579	42218	40628	36966	30669	26978	27567	26591
2013	28929	27636	23243	26921	35042	47293	48212	43856	34436	33230	33797	31802
2014	35312	31896	29528	28668	36403	49596	51502	60683	57862	40708	31388	25825
2015	26450	24641	25019	21514	24376	29791	29750	31576	29303	28388	30239	30286
2016	32724	32838	37014	23758	26072	33316	38089	40337	34450	33723	37048	39501
2017	46184	47211	43251	31576	18754	12195	11127	14519	16087	18187	22288	22537

Table 3: Total apprehensions for all sectors for years 2000 to 2017

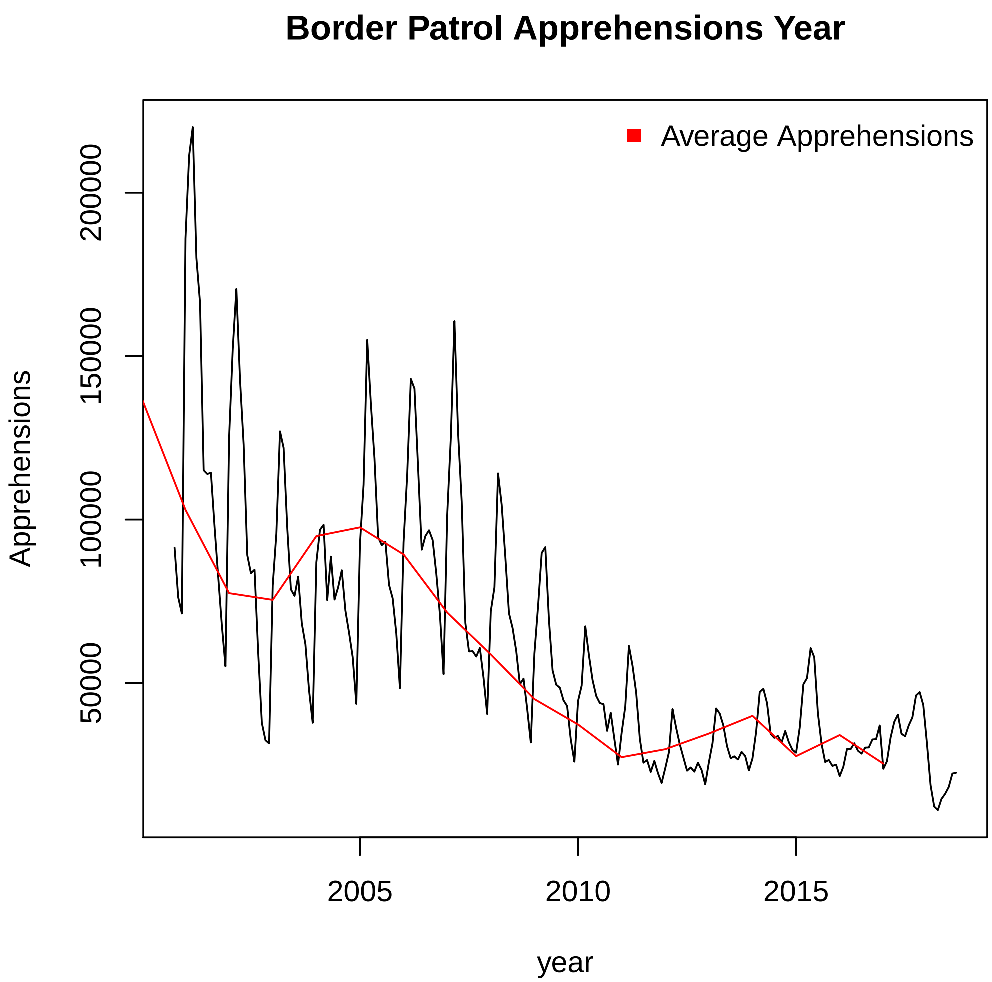
Remove the rownames and column names from the data frame in order to prepare to make it a time series.

```
> rownames(A2000.2017) <- c()
> A2000.2017 <- unname(A2000.2017)
```

3.2 Time Series

Here we use the monthly summaries data in order to create a time series.

```
> ts2 <- as.vector(t(A2000.2017))
> time_series <- ts(ts2, start = c(2000,10), frequency=12)
> ts.plot(time_series, gpars=list(xlab="year", ylab="Apprehensions", lty=c(1:3)))
> meanbyyear <- rowMeans(A2000.2017)
> years <- c(2000:2017)
> lines(years,meanbyyear,col="red")
> title("Border Patrol Apprehensions Year")
> legend("topright", c("Average Apprehensions"), pch=15,  col="red",  bty="n")
```


Border Patrol Apprehensions Year
Average Apprehensions
Apprehensions
200000
150000
100000
50000
2005
2010
2015
year