Assignment 3 - Section 1 project

Nanfang Hong (U92430309), Jiajian Guo (U43469879)

February 17, 2018

Background

US-Mexico border apprehensions hit 17-year lows

(Washington, CNN) Apprehensions at the US-Mexico border reached historic lows in April, continuing a downward slide in the first few months of the Trump administration. Out of curiosity, let's see what data can really tell us.

Tables

We collected the data of US-Mexico border apprehensions at 2010 and 2017.

	Big Bend	Del Rio	El Centro	El Paso	Laredo	RG Valley	San Diego	Tucson	Yuma
Oct	530	1119	2589	1007	2613	4236	5017	23197	582
Nov	421	897	2412	894	2130	3688	4738	16986	649
Dec	373	697	2196	725	1802	2987	4636	10907	711
Jan	433	1234	2688	1124	2526	3658	6413	16122	586
Feb	484	1245	2836	1140	3173	4845	6982	21266	819
Mar	660	1874	4408	1528	4433	7141	9061	31197	1059
Apr	575	1791	3419	1359	4528	7139	7115	28579	732
May	493	1718	3126	1380	3813	7477	5858	22572	608
Jun	415	1326	2440	1005	3475	5595	5092	13160	447
Jul	280	767	2331	725	1857	3832	5113	10303	401
Aug	295	1095	2075	732	2819	5329	4528	9280	262
Sep	329	931	2042	632	2118	3839	4012	8633	260

Table 1: Data Table of Border Apprehesions 2010

	Big Bend	Del Rio	El Centro	El Paso	Laredo	RG Valley	San Diego	Tucson	Yuma
Oct	697	2106	2441	3973	3350	22642	2934	5924	2117
Nov	603	1880	1850	4105	3194	24686	2947	5912	2034
Dec	477	1817	1870	3948	2460	23418	3099	4303	1859
Jan	473	1243	1796	2779	2265	15580	2927	3357	1156
Feb	383	1104	1196	1575	1710	7855	1808	2589	534
Mar	357	746	871	978	1256	4147	1356	2148	336
Apr	413	589	849	906	1304	3942	1392	1487	245
May	552	740	1134	1032	1722	4882	1724	2199	534
Jun	378	761	1280	1180	1839	5817	1652	2632	548
Jul	492	760	1478	1395	2120	7107	1764	2177	894
Aug	563	798	1880	1782	2143	8650	2241	2913	1318
Sep	614	932	1988	1540	2097	8836	2242	3016	1272

Table 2: Data Table of Border Apprehesions 2017

Data Visualization

Comparing 2010 and 2017 statistics by month and sector:

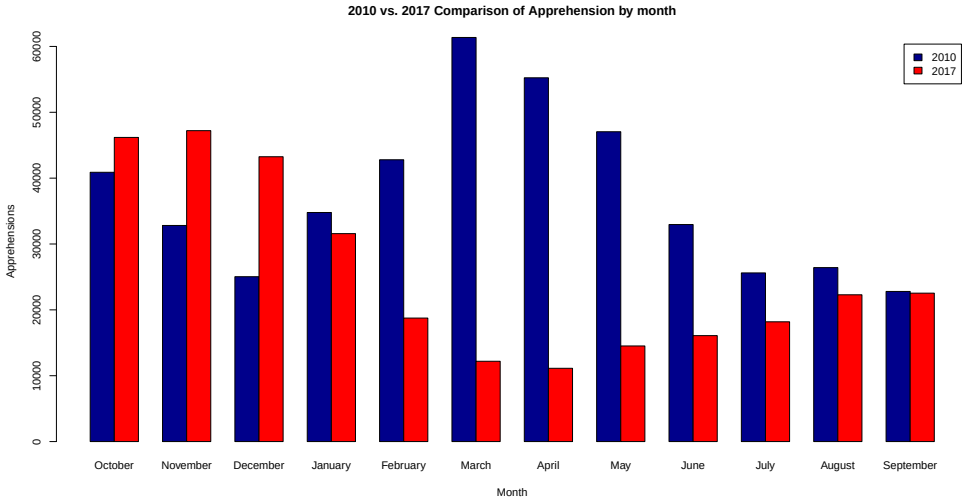

Figure 1: 2010 vs. 2017 Comparison of Apprehension by month

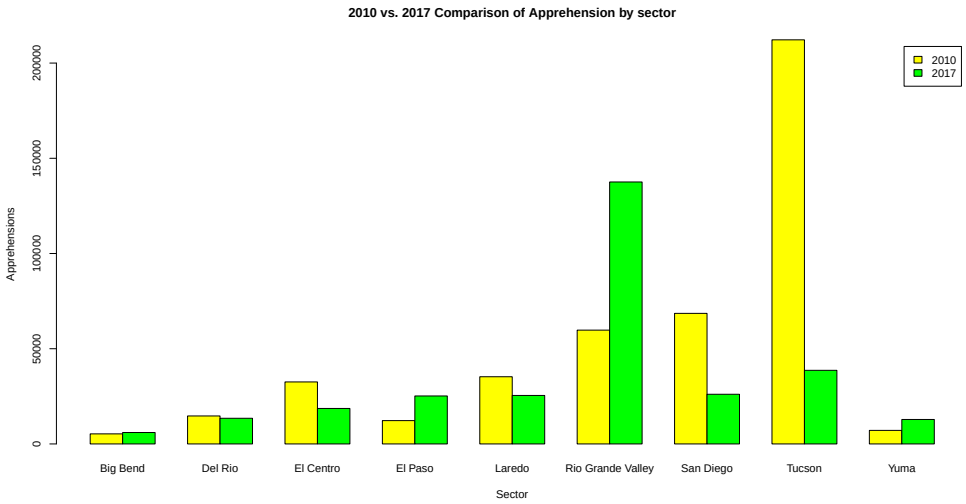

Figure 2: 2010 vs. 2017 Comparison of Apprehension by sector

Statistical Analysis

Then we ask: has there been a change in the maximum of apprehensions of sectors in 2010 and 2017? has there been a change in the maximum of apprehensions of 3 consecutive months in 2010 and 2017?

The sector with the most apprehensions for 2010 is Tucson. The sector with the most apprehensions for 2017 is Rio Grande Valley. Then we do a two-sample Welch t-test. As a result, the p-value is **0.06346**. Fail to reject the null hypothesis of no change, here is **no** significant change in the maximum of 2010 and 2017

Then we do a two-sample Welch t-test again on 3 month periods with the most apprehensions in 2010 and 2017. As a result, the p-value is **0.4042**. Fail to reject the null hypothesis of no change, here is **no** significant change in the maximum 3 consecutive monthly apprehensions.

```
> #3 month period for 2010 and the most is March-April-May:
> a2010 <- which.max(sapply(1:10, function(i) sum(dMat2010[, i]
+         + dMat2010[, i+1] + dMat2010[, i+2])))
> #The data of 9 sectors during March-April-May 2010 (27 entries):
> dMat2010[a2010:(a2010 + 26)]

 [1]  4236  5017 23197   582   421   897  2412   894  2130  3688  4738 16986
[13]   649   373   697  2196   725  1802  2987  4636 10907   711   433  1234
[25]  2688  1124  2526

> #3 month period for 2017 and the most is October-November-December:
> a2017 <- which.max(sapply(1:10, function(i) sum(dMat2017[, i]
+         + dMat2017[, i+1] + dMat2017[, i+2])))
> #The data of 9 sectors during October-November-December 2017 (27 entries):
> dMat2017[a2017:(a2017 + 26)]

 [1]   697  2106  2441  3973  3350 22642  2934  5924  2117   603  1880  1850
[13]  4105  3194 24686  2947  5912  2034   477  1817  1870  3948  2460 23418
[25]  3099  4303  1859

> t.test(dMat2010[a2010:(a2010 + 26)],
+        dMat2017[a2017:(a2017 + 26)],  var.equal = FALSE)

        Welch Two Sample t-test

data:  dMat2010[a2010:(a2010 + 26)] and dMat2017[a2017:(a2017 + 26)]
t = -0.84137, df = 49.056, p-value = 0.4042
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4738.737  1941.700
sample estimates:
mean of x mean of y
 3662.444  5060.963

> #Comparison of 2010 and 2017 maximum sector.
> #the sector with the most apprehensions for 2010:
> df2010$Sector[which.max(rowSums(dMat2010))]

[1] Tucson
9 Levels: Big Bend Del Rio El Centro El Paso Laredo ... Yuma

> #the monthly data of Tucson
> dMat2010[which.max(rowSums(dMat2010)), ]

  October  November  December   January  February     March     April       May
    23197     16986     10907     16122     21266     31197     28579     22572
     June      July    August September
    13160     10303      9280      8633

> #the sector with the most apprehensions for 2017:
> df2017$Sector[which.max(rowSums(dMat2017))]
```

```
[1] Rio Grande Valley
10 Levels:  Big Bend Del Rio El Centro El Paso Laredo ... Yuma

> #the monthly data of Rio Grande Valley
> dMat2017[which.max(rowSums(dMat2017)), ]

  October  November  December   January  February     March     April       May
    22642     24686     23418     15580      7855      4147      3942      4882
     June      July    August September
     5817      7107      8650      8836

> t.test(dMat2010[which.max(rowSums(dMat2010)),],
+        dMat2017[which.max(rowSums(dMat2017)),],  var.equal = FALSE)

        Welch Two Sample t-test

data:  dMat2010[which.max(rowSums(dMat2010)), ] and dMat2017[which.max(rowSums(dMat2017)), ]
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
```

Historical Data Visualization

Finally, we collected all apprehensions data from 2000 to 2017, and plot in time series. Let's see the big picture. The trend is definitely going down.

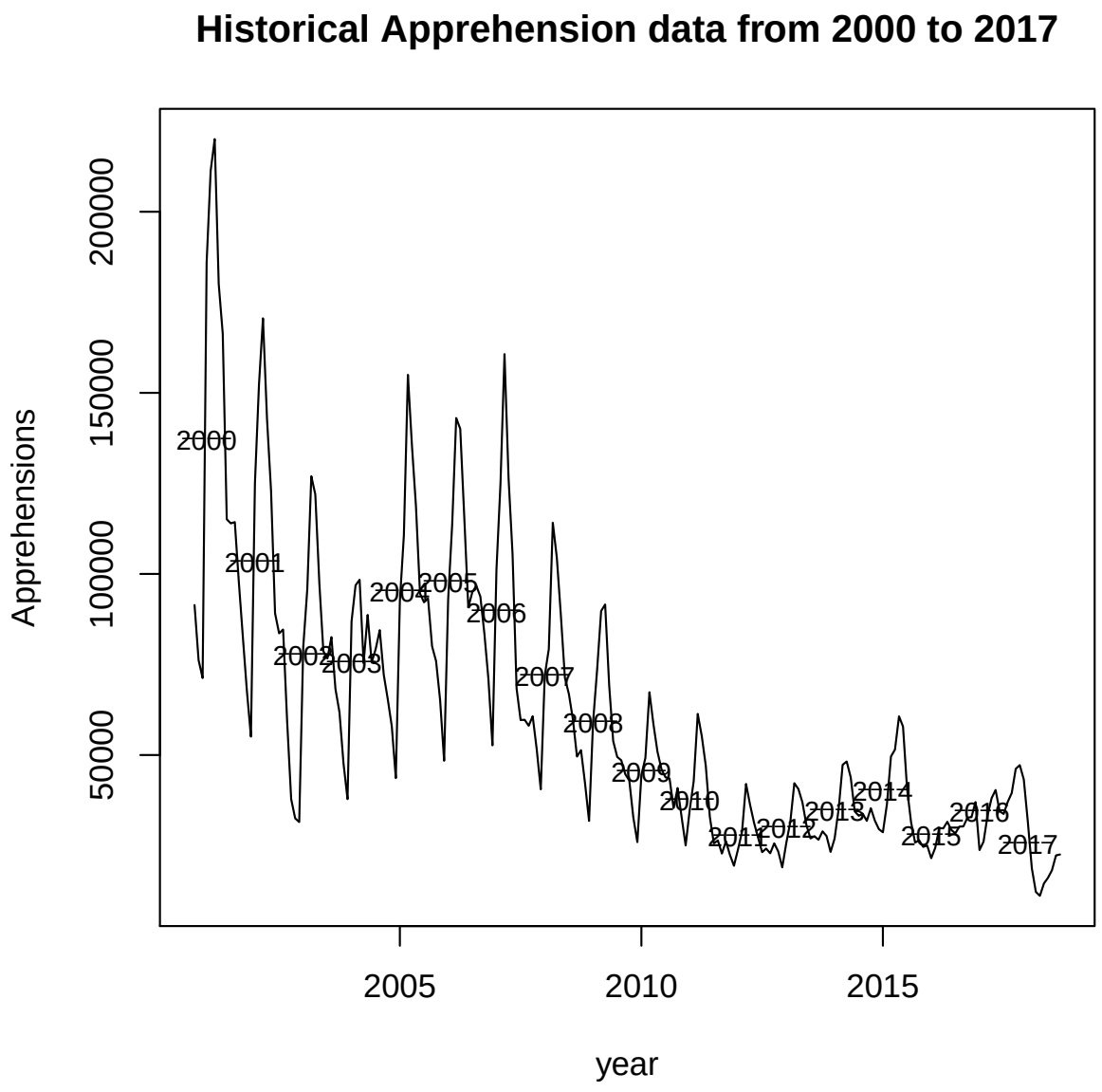

Figure 3: Historical Apprehensions data from 2000 to 2017