Is a new Mexican border wall really necessary?

(Lavandera, 2017).

Erin Tsai

MSBA 320

Spring 2018

Class ID: 17049260

Table of Contents

I. Introduction

After President Trump was elected, immigration has become a hot topic throughout the news media on an almost daily basis. President Trump's most popular promise during his election campaign was that he would build a great wall, and make Mexico pay for it, in order to prevent illegal entry to the United States through the Mexican/American border. (Ballasteros, 2018).

The actual border stretches approximately 1,900 miles. (Beaver, 2006). Approximately 650 miles of the 1,900 miles is already protected by the wall. (Ballasteros, 2018). Families from other countries cross the border annually, most often in seek of asylum, new opportunities, a better life here in the United States, at the risk of death by starvation, dehydration, and other dangers. (Jenkins, 2015).

Based on President Trump's claims, the goal is to keep out Mexicans because they are "drug dealers, criminals and rapists." (Lee, 2015). The question becomes, based on the trends of illegal immigration, are we actually keeping out Mexicans, or other people? Moreover, what kind of people is America deporting? Are we deporting mostly criminals? Or mostly young vulnerable children, with no knowledge of any other country in their lives?

We will use this opportunity to analyze the data to determine what the data tends to show as America's priorities.

II. Data Collection

Our data comes from mostly the U.S. Customs and Border Protection, which operates under U.S. Department of Homeland Security. There are annual data tables offered each year. There is an extremely large number of tables. However, we will focus primarily on border crossing and deportations (specifically table 41D through U.S. Department of Homeland Security, and apprehensions at the border by U.S. Customs and Border Protection).

We have pulled data from Kaggle regarding illegal immigrants. (U.S. Customs and Border Protection, 2017). Additionally, we are utilizing Table 41D from the 2016 Immigration Statistics from the U.S. Department of Homeland Security. (U.S. Department of Homeland Security, 2018). We cleaned up the data files to make it easier to work with, by transposing and eliminating rows and columns that we were not looking into. We also utilized the data to create columns to determine percentages as a comparison.

III. Data Analysis Conducted

We use R as our main tool for conducting data analysis. With both data sets involved, we initially looked at the data as a whole. We tried to determine if there were any trends that may be of use to us to determine how to further approach analysis.

We looked at the data visually to get an easy idea of whether or not the data tends to show decrease or increase. Because we are told certain trends by the media, it is important for us to truly determine if these trends hold true. After a brief visual inspection of the data, we conduct a few quick descriptive statistics determinations to see if it agrees with what we have seen. Then, we move into finding linear regression and making predictions.

With respect to our second data set, we still do a quick visual inspection of the data on Mexicans to find trends, to see if it agrees with what we expect it to. Then we utilize hypothesis testing to determine if there is any difference between how Mexicans are being treated and how everyone else is being treated as far as deportations go.

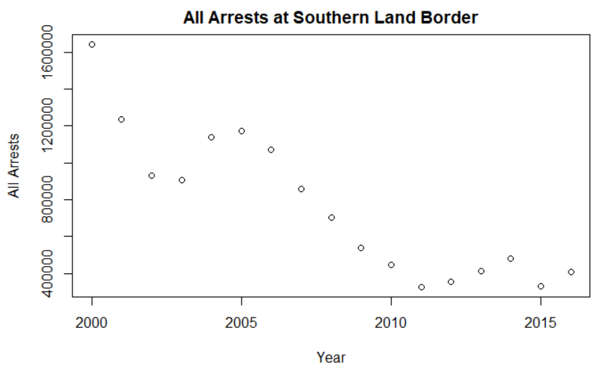
IV. Analysis and Interpretation

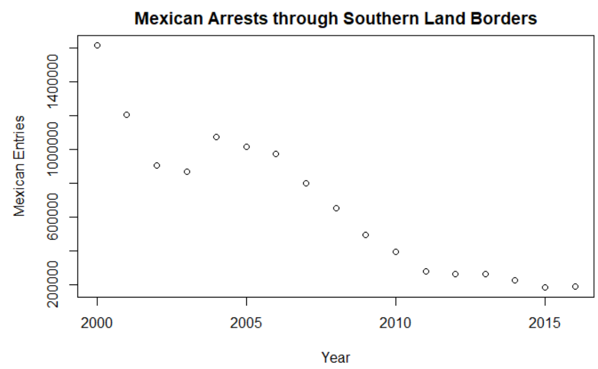
A. Trends of Arrests by Customs and Border Protection at the Southern Border

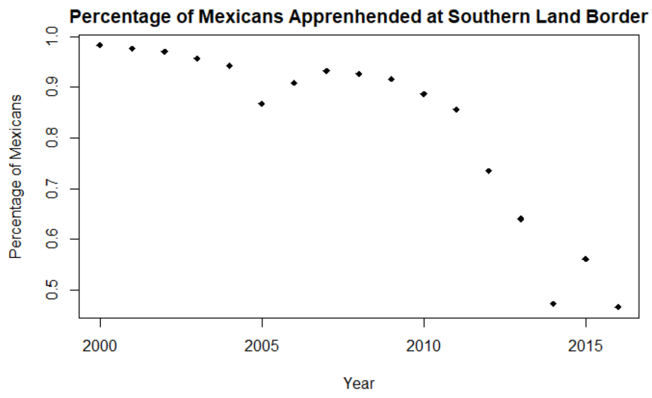
a. Exploratory Data Analysis

Initially, we reviewed the data based on descriptive statistics and visually to determine if there are any trends of relevance we may want to look into.

The plots below each indicate a downward trend of arrests at the Southwest border, which is the Southern Land border of the United States, regardless of whether we are looking at all arrests, just Mexicans, or the percentage of Mexicans. All of the plots indicate a similar downward trend.







The charts agree with our correlation data, which tends to indicate a negative value for almost all sectors, as compared to year.

```
                   Big.Bend    Del.Rio  El.Centro    El.Paso     Laredo Rio.Grande.Valley
Big.Bend          1.0000000  0.8773931  0.8736508  0.8979049  0.9711289        0.65061687
Del.Rio           0.8773931  1.0000000  0.9753969  0.8057345  0.8912691        0.56074915
El.Centro         0.8736508  0.9753969  1.0000000  0.8075173  0.8791536        0.46495716
El.Paso           0.8979049  0.8057345  0.8075173  1.0000000  0.8872123        0.47450877
Laredo            0.9711289  0.8912691  0.8791536  0.8872123  1.0000000        0.71352183
Rio.Grande.Valley 0.6506169  0.5607492  0.4649572  0.4745088  0.7135218        1.00000000
San.Diego         0.8366344  0.6497540  0.7034958  0.8264843  0.8276306        0.46108886
Tucson            0.9341044  0.8370121  0.8621053  0.9267025  0.9264147        0.51036261
Yuma              0.8067900  0.7449764  0.7101617  0.9496094  0.7907743        0.42137145
All               0.9626309  0.8901869  0.8929246  0.9403178  0.9667969        0.61753546
USA               0.9636297  0.8909983  0.8939606  0.9394382  0.9676778        0.61809473
Year             -0.4737740 -0.5429531 -0.5901676 -0.5789442 -0.4292502        0.03916075
                  San.Diego     Tucson       Yuma        All        USA        Year
Big.Bend          0.8366344  0.9341044  0.8067900  0.9626309  0.9636297 -0.47377405
Del.Rio           0.6497540  0.8370121  0.7449764  0.8901869  0.8909983 -0.54295314
El.Centro         0.7034958  0.8621053  0.7101617  0.8929246  0.8939606 -0.59016755
El.Paso           0.8264843  0.9267025  0.9496094  0.9403178  0.9394382 -0.57894418
Laredo            0.8276306  0.9264147  0.7907743  0.9667969  0.9676778 -0.42925018
Rio.Grande.Valley 0.4610889  0.5103626  0.4213714  0.6175355  0.6180947  0.03916075
San.Diego         1.0000000  0.9396167  0.7376907  0.9008657  0.9008129 -0.44240435
Tucson            0.9396167  1.0000000  0.8626158  0.9851174  0.9849571 -0.54774458
Yuma              0.7376907  0.8626158  1.0000000  0.8731618  0.8709253 -0.52787789
All               0.9008657  0.9851174  0.8731618  1.0000000  0.9999632 -0.51966294
USA               0.9008129  0.9849571  0.8709253  0.9999632  1.0000000 -0.52032236
Year             -0.4424043 -0.5477446 -0.5278779 -0.5196629 -0.5203224  1.00000000
```

As we can see in the above correlation chart, the correlation values under Year are almost all negative, except for Rio Grande Valley, which is very near 0, indicating little or no correlation to year at all. We will take that into account when we conduct linear regression.

b. Linear Regression and Predictions

Because our exploratory data analysis indicates a downward trend, we attempt to utilize linear regression to estimate the numbers for 2017. Our first linear model takes into account All Arrests by Year. We obtain the following linear model:

```
Call:
lm(formula = AllArrests[, "All"] ~ AllArrests[, "Year"])

Coefficients:
         (Intercept)  AllArrests[, "Year"]
           141482666                -70080
```

We utilize ANOVA to determine significance:

```
                    Df    Sum Sq   Mean Sq F value   Pr(>F)    
AllArrests[, "Year"]  1 2.004e+12 2.004e+12   63.06 9.43e-07 ***
Residuals           15 4.766e+11 3.178e+10                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

With a negative slope for Year, it agrees with our expectations. Moreover, based on the F value and the extremely small probability, this indicates that the Year would be a good predictor in our regression equation, that is to say, that there is a strong trend between Year and All Arrests.

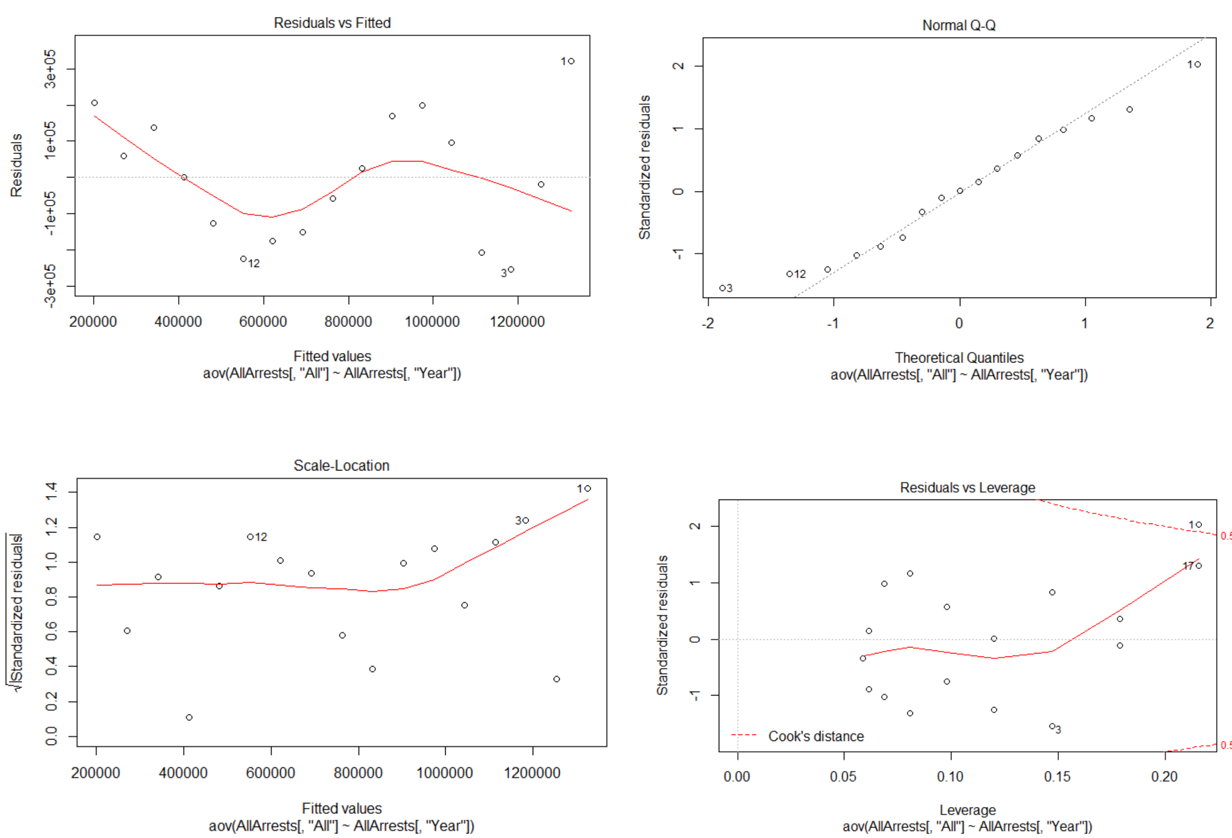
We compare this model with a few others, one for the Percent data, to see if the data results in similar data. Although the probability and F value still indicate significance, we cannot really compare the two models appropriately as one is based on percentages of Mexicans out of all arrests, and the other one is merely all arrests. However, it is important to note for the purposes of our understanding, that, not only are the number of arrests in total decreasing, but in particular ***the percentage of Mexicans being arrested at the Southern border are decreasing too.*** The question remains, why is America pushing so hard to build a wall to “keep out the Mexicans” if the number of Mexicans entering is decreasing so quickly already, ***without*** the wall?

We further make another model to include Rio Grande Valley as a factor, given that it was the only sector providing a positive correlation value. The resulting model tends to indicate that inclusion of Rio Grande Valley may significantly affect the linear regression model, as indicated by the high probability value.

Because we have created a linear regression model, with a fairly good fit of an adjusted R squared value of 0.795, based solely on the All Arrests value, we can attempt to make a prediction for 2017.

The estimate based on our linear regression equation of Model 1, based on All Arrests, indicates an estimate of 131,306. The actual value of apprehensions at the Southwest border for 2017 was 104,997. (U.S. Customs and Border Protection, Dec. 2017) Although the numbers physically are far apart, the trend was maintained. Based on the graphs in the above section, we are inclined to believe that the trend may possibly not be linear, rather it may be logarithmic. If so, then logarithmic regression may produce more accurate values. Logarithmic regression would be something to look into for further research.

A review of our fit based on the plots indicates that the fit is mostly okay.

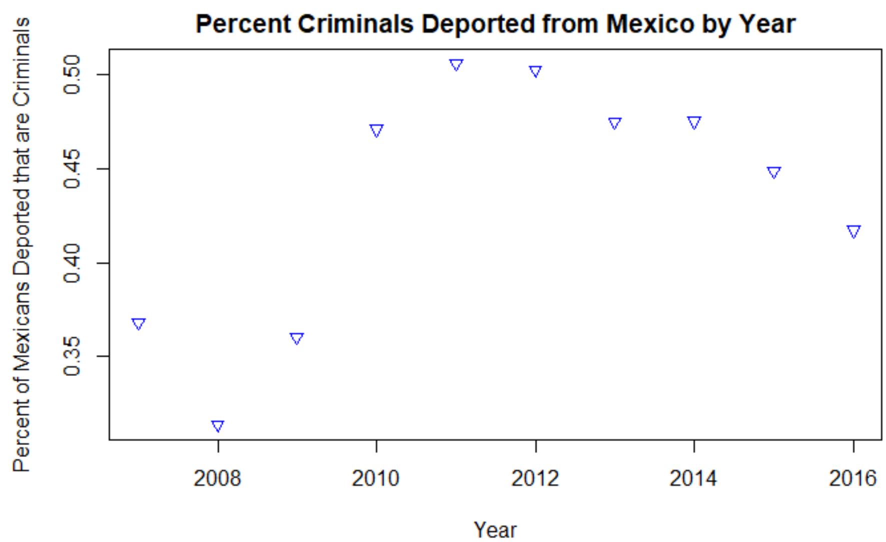


Here, the residuals do not tend to show that they are getting much larger as values get higher. They are generally along the same gap. The linear graph of the Normal QQ indicates that our errors are mostly normal. The Scale Location plot also does not give us any red flags to be wary of. In our Cooks Distance plot, there does not seem to be any major influence except for the 3 and 17. This is something to possibly to look into in thefuture to find a better fit.

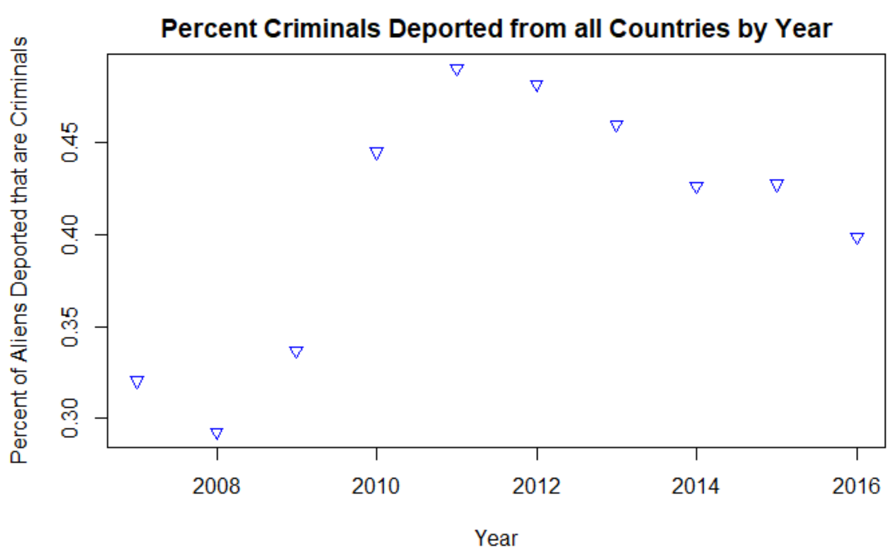
B. Deportations of Criminals from Mexico

a. Exploratory Data Analysis

Based on our initial review of the data of criminals vs. non criminals being deported that are from Mexico, we note that there was an increase in number of criminals being deported between 2008 and approximately 2011. However, after 2011, there is a continuing downward trend of criminals being deported that originated from Mexico. Initially, based on President Obama's administration, Immigration and Customs Enforcement were supposed to prioritize criminals for deportation starting 2014. (Zamora, 2017). This graph below seems to indicate the opposite when it comes to focusing on Mexicans. Therefore, the remaining question we will present is whether the percentage of criminals being deported is the same for Mexicans, as it is for all of the countries combined. We will utilize hypothesis testing for this task.



To begin, we will graphically view the data regarding criminals deported from all countries as a whole, as an initial comparison, before delving into hypothesis testing.



From the two graphs, we can see graphically, the trend is very similar. That is to say, that the percent of criminals being deported is ***shrinking***. This is contrary to the decision under the Obama administration in 2014 to prioritize deportations of criminals. (Zamora, 2017). What this also means, is that the media has been portraying the data in the wrong light, causing us to believe that more of the deportations occurring are criminal deportations, as opposed to non-criminal deportations.

b. Hypothesis Testing

We further explore the data regarding criminal deportations by comparing the means through hypothesis testing. We compare the percent of deportations that are criminals between Mexico and the all countries as a whole to determine if there is a difference in the means. Are Mexicans being treated worse? Better? Neither?

The t-test is set up as follows:

Level of significance selected (α): 5% or 0.05

Null Hypothesis: The difference of means of percent of deportations that are criminals is 0 between Mexico and all Countries.

Alternative Hypothesis: The difference of means is ***not*** 0, or in other words, there ***is*** a significant difference in means of percent of deportations that are criminals.

The results of the Welch's t-test are as follows:

```
Welch Two Sample t-test

data:  DepMexicans[, "Percent.Crim"] and DepUSA[1:10, "Percent.Crim"]
t = 0.86577, df = 17.97, p-value = 0.398
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03736340  0.08973211
sample estimates:
mean of x mean of y
0.4338738 0.4076894
```

Based on our t-test results, we ***cannot*** reject the null hypothesis because our p-value is 0.398, which is much greater than 0.05. Therefore, we cannot find that there is a significant difference in means. We acknowledge that the 95% confidence interval ***does*** contain the value of 0.00, so as to confirm that we cannot find a significant difference in means. What this tends to indicate is that, deportations by Immigration and Customs Enforcement, do not indicate that there is a primary focus on criminals, and that Mexicans are not being treated differently with respect to deporting criminals over non-criminals.

V. Conclusion

Based on our preliminary statistical analysis based on the data of border apprehensions in the Southern land border of the United States, and deportations by Immigration and Customs Enforcement, it seems there are a few pieces missing.

President Trump often cites the need to deport Mexicans, as a priority, because they are "drug dealers, criminals and rapists." However, the deported Mexicans are more often than not non-criminals. So, what is the purpose of having this border? Moreover, why would we need to spend billions of dollars to add more miles to a border wall, if the number of arrests have been declining over the years so quickly? What is the point of spending all this money to build a border, when clearly the implemented rules by Immigration and Customs Enforcement and the Department of Homeland Security, are already preventing people from coming into the United States illegally?

We have been told in the past that criminals are priorities for deportation, but the data does not reflect that. Somehow, news articles have been able to present "data" that indicate as such, even though our data, that comes directly from the U.S. Department of Homeland Security page does not replicate the statements made. So, how does this kind of information come out? At this point, what can we read and believe? Are there other factors involved that needed to be taken into consideration?

The data analysis conducted here only leaves us with more questions. But at the very least, we can easily see that spending billions of dollars on a border wall is clearly a waste of money and unnecessary, when illegal immigration is already under a lot more control than before years.

VI. References

Ballasteros, Carlos. (2018, January 18). Trump's Border Wall by the Numbers: 10 Facts about the President's Most Popular Campaign Promise. *Newsweek.* Retrieved from: http://www.newsweek.com/trump-border-wall-facts-immigration-daca-dreamers-784693.

Beaver, Janice Cheryl. (2006, November 9). *U.S. International Borders: Brief Facts*. Retrieved from https://fas.org/sgp/crs/misc/RS21729.pdf.

Jenkins, Jack. (2015, February 10). What It's Really Like to Cross the U.S.-Mexico Border. *ThinkProgress.* Retrieved from: https://thinkprogress.org/what-its-really-like-to-cross-the-u-s-mexico-border-a375d80f3cc/.

Lavandera, Ed. (2017, February 9). Tensions and tunnels in an ever-changing border. *CNN.* Retrieved from: https://www.cnn.com/2017/01/27/politics/border-series-california/index.html

Lee. Michelle Ye Hee. (2015, July 8). Donald Trump's false comments connecting Mexican immigrants and crime. *Washington Post.* Retrieved from: https://www.washingtonpost.com/news/fact-checker/wp/2015/07/08/donald-trumps-false-comments-connecting-mexican-immigrants-and-crime/?noredirect=on&utm_term=.b8876837bbc9.

U.S. Customs and Border Protection. (2017, January 27). Illegal Immigrants Arrested by US Border Patrol. Available from Kaggle website: https://www.kaggle.com/cbp/illegal-immigrants/data.

U.S. Customs and Border Protection. (2017, December 5). CBP Border Security Report. Retrieved from: https://www.cbp.gov/sites/default/files/assets/documents/2017-Dec/cbp-border-security-report-fy2017.pdf.

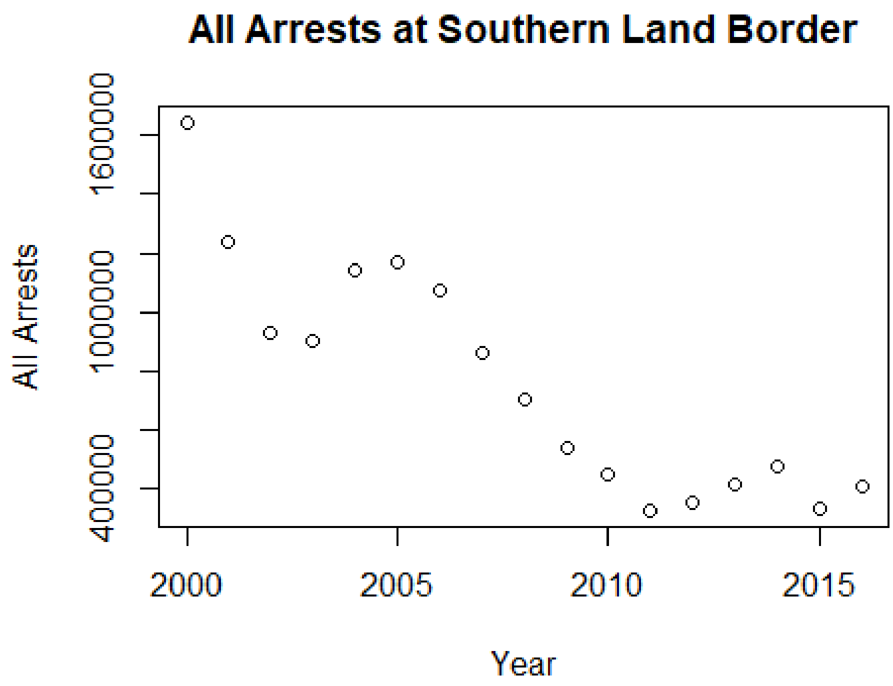
U.S. Department of Homeland Security. (2018, April 6). Yearbook of Immigration Statistics 2016. Available from U.S. Department of Homeland Security website: https://www.dhs.gov/immigration-statistics/yearbook/2016.

Zamora, Lazaro (2017, February 27). Comparing Trump and Obama's Deportation Priorities. *Bipartisan Policy Center.* Retrieved from: https://bipartisanpolicy.org/blog/comparing-trump-and-obamas-deportation-priorities/.

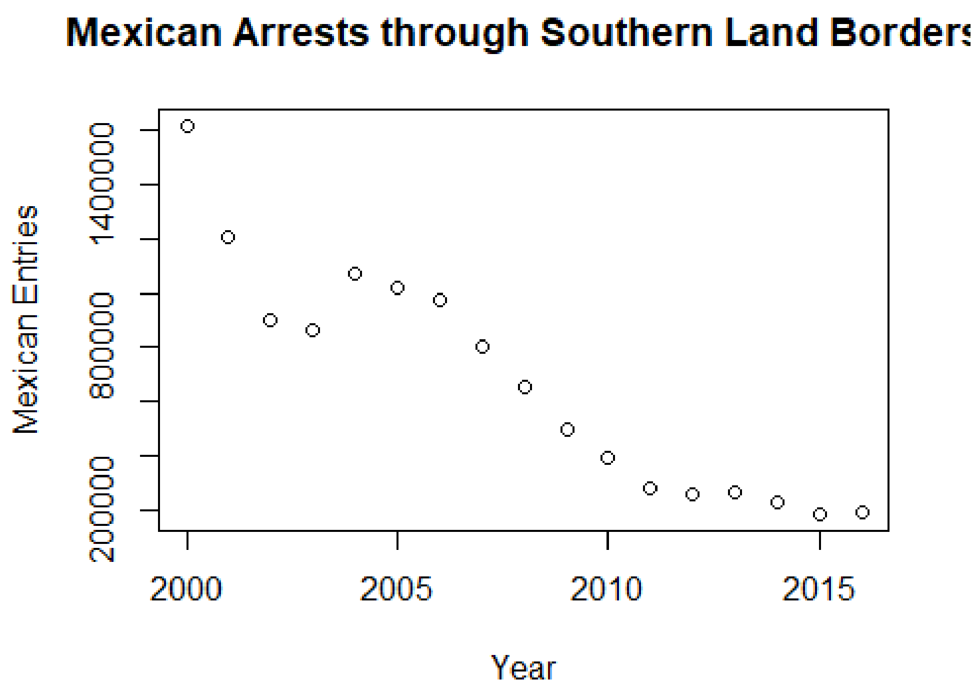
VII. Appendix

```
Southwest<-read.csv("C:\\Users\\etlaw\\OneDrive\\Documents\\MSBA 320\\Final P
roject\\SouthwestOnly.csv")

attach(Southwest)
names(Southwest)

##  [1] "Big.Bend"          "Del.Rio"           "El.Centro"
##  [4] "El.Paso"           "Laredo"            "Rio.Grande.Valley"
##  [7] "San.Diego"         "Tucson"            "Yuma"
## [10] "All"               "USA"               "Year"
## [13] "Type"

Mexican<-subset(Southwest,Type=="Mexican")

attach(Mexican)

## The following objects are masked from Southwest:
##
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

names(Mexican)

##  [1] "Big.Bend"          "Del.Rio"           "El.Centro"
##  [4] "El.Paso"           "Laredo"            "Rio.Grande.Valley"
##  [7] "San.Diego"         "Tucson"            "Yuma"
## [10] "All"               "USA"               "Year"
## [13] "Type"

Percent<-subset(Southwest,Type=="Percent")

attach(Percent)

## The following objects are masked from Mexican:
##
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

## The following objects are masked from Southwest:
##
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

names(Percent)

##  [1] "Big.Bend"          "Del.Rio"           "El.Centro"
##  [4] "El.Paso"           "Laredo"            "Rio.Grande.Valley"
##  [7] "San.Diego"         "Tucson"            "Yuma"
## [10] "All"               "USA"               "Year"
## [13] "Type"
```

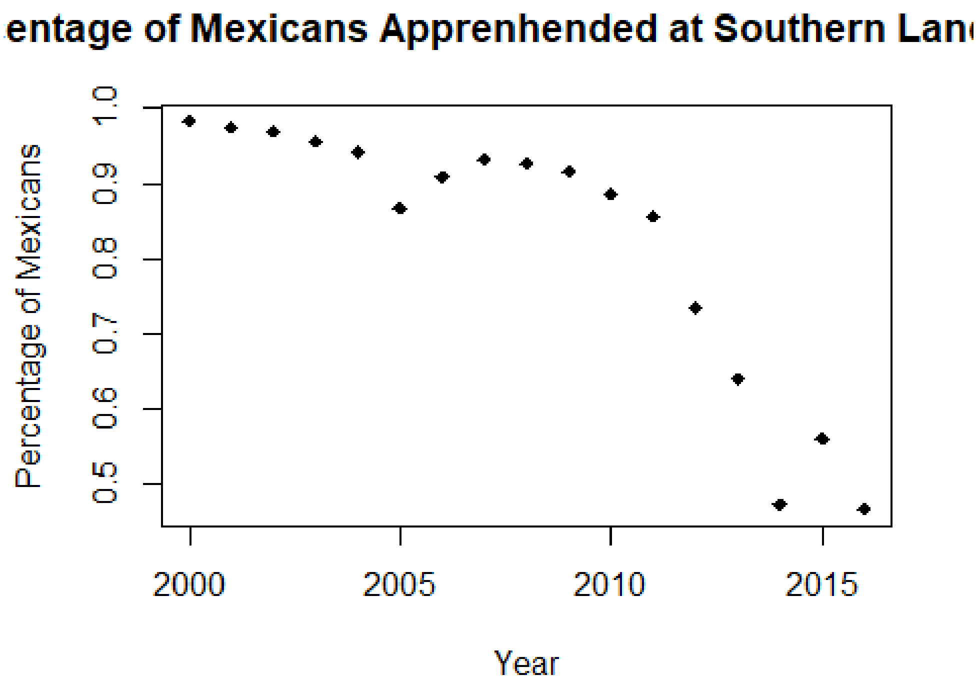
```r
AllArrests<-subset(Southwest,Type=="All")

attach(AllArrests)

## The following objects are masked from Percent:
## 
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

## The following objects are masked from Mexican:
## 
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

## The following objects are masked from Southwest:
## 
##     All, Big.Bend, Del.Rio, El.Centro, El.Paso, Laredo,
##     Rio.Grande.Valley, San.Diego, Tucson, Type, USA, Year, Yuma

names(AllArrests)

##  [1] "Big.Bend"          "Del.Rio"           "El.Centro"        
##  [4] "El.Paso"           "Laredo"            "Rio.Grande.Valley"
##  [7] "San.Diego"         "Tucson"            "Yuma"             
## [10] "All"               "USA"               "Year"             
## [13] "Type"

plot(AllArrests[,'Year'],AllArrests[,'All'],xlab="Year",ylab="All Arrests",ma
in="All Arrests at Southern Land Border")
```




```r
plot(Mexican[,'Year'],Mexican[,'All'],xlab="Year",ylab="Mexican Entries",main
="Mexican Arrests through Southern Land Borders")
```



```r
plot(Percent[,'Year'],Percent[,'All'],main="Percentage of Mexicans Apprenhended at Southern Land Border",xlab="Year",ylab="Percentage of Mexicans",pch=18)
```




```r
summary(AllArrests)
```

```
##     Big.Bend         Del.Rio          El.Centro          El.Paso
##  Min.   : 3684   Min.   : 14694   Min.   : 12820   Min.   :  9678
##  1st Qu.: 5031   1st Qu.: 20761   1st Qu.: 23916   1st Qu.: 12339
##  Median : 6360   Median : 23510   Median : 40961   Median : 30312
##  Mean   : 7401   Mean   : 43959   Mean   : 63713   Mean   : 57502
##  3rd Qu.:10530   3rd Qu.: 53794   3rd Qu.: 74467   3rd Qu.:104399
##  Max.   :13689   Max.   :157178   Max.   :238126   Max.   :122679
##      Laredo        Rio.Grande.Valley   San.Diego          Tucson
##  Min.   : 35287   Min.   : 59243    Min.   : 26290   Min.   : 63397
##  1st Qu.: 40569   1st Qu.: 75473    1st Qu.: 31891   1st Qu.:120939
##  Median : 50749   Median : 97762    Median :110075   Median :317696
##  Mean   : 58703   Mean   :112825    Mean   : 92365   Mean   :282358
##  3rd Qu.: 74840   3rd Qu.:134186    3rd Qu.:138608   3rd Qu.:392074
##  Max.   :108973   Max.   :256393    Max.   :162390   Max.   :616346
##       Yuma              All                USA               Year
##  Min.   :  5833   Min.   : 327577   Min.   : 337117   Min.   :2000
##  1st Qu.:  6951   1st Qu.: 414397   1st Qu.: 420789   1st Qu.:2004
##  Median : 14170   Median : 705005   Median : 723825   Median :2008
##  Mean   : 43973   Mean   : 762799   Mean   : 779613   Mean   :2008
##  3rd Qu.: 78385   3rd Qu.:1071972   3rd Qu.:1089092   3rd Qu.:2012
##  Max.   :138438   Max.   :1643679   Max.   :1676438   Max.   :2016
##       Type
```

```
##  All    :17
##  Mexican: 0
##  Percent: 0
##
##
##
```

```r
summary(Mexican)
```

```
##     Big.Bend        Del.Rio          El.Centro          El.Paso
##  Min.   : 2177   Min.   : 10196   Min.   : 11320   Min.   :  8915
##  1st Qu.: 3417   1st Qu.: 12404   1st Qu.: 22511   1st Qu.: 10677
##  Median : 5002   Median : 14916   Median : 40159   Median : 29137
##  Mean   : 6366   Mean   : 35235   Mean   : 62413   Mean   : 54604
##  3rd Qu.: 9568   3rd Qu.: 43931   3rd Qu.: 73741   3rd Qu.:100842
##  Max.   :12851   Max.   :150467   Max.   :236346   Max.   :117780
##      Laredo       Rio.Grande.Valley   San.Diego          Tucson
##  Min.   : 25337   Min.   : 38353    Min.   : 24269   Min.   : 46494
##  1st Qu.: 29131   1st Qu.: 47823    1st Qu.: 27871   1st Qu.:102303
##  Median : 34403   Median : 55401    Median :109281   Median :305429
##  Mean   : 47934   Mean   : 60781    Mean   : 90632   Mean   :270711
##  3rd Qu.: 62220   3rd Qu.: 63468    3rd Qu.:136847   3rd Qu.:382610
##  Max.   :105637   Max.   :122501    Max.   :160818   Max.   :614145
##       Yuma             All               USA               Year
##  Min.   :  3513   Min.   : 186017   Min.   : 188122   Min.   :2000
##  1st Qu.:  5416   1st Qu.: 265409   1st Qu.: 267734   1st Qu.:2004
##  Median :  8016   Median : 653035   Median : 661766   Median :2008
##  Mean   : 42484   Mean   : 671160   Mean   : 679943   Mean   :2008
##  3rd Qu.: 77974   3rd Qu.: 973819   3rd Qu.: 981066   3rd Qu.:2012
##  Max.   :136767   Max.   :1615081   Max.   :1636883   Max.   :2016
##       Type
##  All    : 0
##  Mexican:17
##  Percent: 0
##
##
##
```

```r
summary(Percent)
```

```
##     Big.Bend          Del.Rio           El.Centro          El.Paso
##  Min.   :0.4285   Min.   :0.4204   Min.   :0.7384   Min.   :0.5329
##  1st Qu.:0.8620   1st Qu.:0.5719   1st Qu.:0.9413   1st Qu.:0.9137
##  Median :0.8990   Median :0.7185   Median :0.9804   Median :0.9580
##  Mean   :0.8373   Mean   :0.7178   Mean   :0.9514   Mean   :0.9051
##  3rd Qu.:0.9081   3rd Qu.:0.8167   3rd Qu.:0.9893   3rd Qu.:0.9659
##  Max.   :0.9388   Max.   :0.9580   Max.   :0.9943   Max.   :0.9879
##      Laredo       Rio.Grande.Valley   San.Diego          Tucson
##  Min.   :0.6025   Min.   :0.2475    Min.   :0.7958   Min.   :0.7165
##  1st Qu.:0.7180   1st Qu.:0.4049    1st Qu.:0.9595   1st Qu.:0.8525
##  Median :0.7903   Median :0.6815    Median :0.9857   Median :0.9614
```

```
##  Mean   :0.7901   Mean   :0.6075    Mean   :0.9653   Mean   :0.9156
##  3rd Qu.:0.8329   3rd Qu.:0.7387    3rd Qu.:0.9903   3rd Qu.:0.9821
##  Max.   :0.9694   Max.   :0.9194    Max.   :0.9929   Max.   :0.9964
##       Yuma              All              USA               Year
##  Min.   :0.2479   Min.   :0.4666   Min.   :0.4641   Min.   :2000
##  1st Qu.:0.9106   1st Qu.:0.7351   1st Qu.:0.7286   1st Qu.:2004
##  Median :0.9585   Median :0.9084   Median :0.9008   Median :2008
##  Mean   :0.8806   Mean   :0.8234   Mean   :0.8154   Mean   :2008
##  3rd Qu.:0.9879   3rd Qu.:0.9422   3rd Qu.:0.9350   3rd Qu.:2012
##  Max.   :0.9953   Max.   :0.9826   Max.   :0.9764   Max.   :2016
##       Type
##  All    : 0
##  Mexican: 0
##  Percent:17
##
##
##
```

```
sapply(Southwest,mode)
```

```
##          Big.Bend           Del.Rio         El.Centro           El.Paso
##         "numeric"         "numeric"         "numeric"         "numeric"
##            Laredo Rio.Grande.Valley         San.Diego            Tucson
##         "numeric"         "numeric"         "numeric"         "numeric"
##              Yuma               All               USA              Year
##         "numeric"         "numeric"         "numeric"         "numeric"
##              Type
##         "numeric"
```

```
cor(Southwest[,1:12], use = "complete.obs",method="pearson")
```

```
##                     Big.Bend    Del.Rio  El.Centro    El.Paso     Laredo
## Big.Bend           1.0000000  0.8773931  0.8736508  0.8979049  0.9711289
## Del.Rio            0.8773931  1.0000000  0.9753969  0.8057345  0.8912691
## El.Centro          0.8736508  0.9753969  1.0000000  0.8075173  0.8791536
## El.Paso            0.8979049  0.8057345  0.8075173  1.0000000  0.8872123
## Laredo             0.9711289  0.8912691  0.8791536  0.8872123  1.0000000
## Rio.Grande.Valley  0.6506169  0.5607492  0.4649572  0.4745088  0.7135218
## San.Diego          0.8366344  0.6497540  0.7034958  0.8264843  0.8276306
## Tucson             0.9341044  0.8370121  0.8621053  0.9267025  0.9264147
## Yuma               0.8067900  0.7449764  0.7101617  0.9496094  0.7907743
## All                0.9626309  0.8901869  0.8929246  0.9403178  0.9667969
## USA                0.9636297  0.8909983  0.8939606  0.9394382  0.9676778
## Year              -0.4737740 -0.5429531 -0.5901676 -0.5789442 -0.4292502
##                   Rio.Grande.Valley  San.Diego     Tucson       Yuma
## Big.Bend                 0.65061687  0.8366344  0.9341044  0.8067900
## Del.Rio                  0.56074915  0.6497540  0.8370121  0.7449764
## El.Centro                0.46495716  0.7034958  0.8621053  0.7101617
## El.Paso                  0.47450877  0.8264843  0.9267025  0.9496094
## Laredo                   0.71352183  0.8276306  0.9264147  0.7907743
## Rio.Grande.Valley        1.00000000  0.4610889  0.5103626  0.4213714
```

```
## San.Diego           0.46108886  1.0000000  0.9396167  0.7376907
## Tucson              0.51036261  0.9396167  1.0000000  0.8626158
## Yuma                0.42137145  0.7376907  0.8626158  1.0000000
## All                 0.61753546  0.9008657  0.9851174  0.8731618
## USA                 0.61809473  0.9008129  0.9849571  0.8709253
## Year                0.03916075 -0.4424043 -0.5477446 -0.5278779
##                     All        USA        Year
## Big.Bend          0.9626309  0.9636297 -0.47377405
## Del.Rio           0.8901869  0.8909983 -0.54295314
## El.Centro         0.8929246  0.8939606 -0.59016755
## El.Paso           0.9403178  0.9394382 -0.57894418
## Laredo            0.9667969  0.9676778 -0.42925018
## Rio.Grande.Valley  0.6175355  0.6180947  0.03916075
## San.Diego         0.9008657  0.9008129 -0.44240435
## Tucson            0.9851174  0.9849571 -0.54774458
## Yuma              0.8731618  0.8709253 -0.52787789
## All               1.0000000  0.9999632 -0.51966294
## USA               0.9999632  1.0000000 -0.52032236
## Year             -0.5196629 -0.5203224  1.00000000
```

```
lm(AllArrests[,'All']~AllArrests[,'Year'])
```

```
##
## Call:
## lm(formula = AllArrests[, "All"] ~ AllArrests[, "Year"])
##
## Coefficients:
##          (Intercept)  AllArrests[, "Year"]
##            141482666                -70080
```

```
model1<-aov(AllArrests[,'All']~AllArrests[,'Year'])
```

```
summary(model1)
```

```
##                      Df    Sum Sq   Mean Sq F value   Pr(>F)
## AllArrests[, "Year"]  1 2.004e+12 2.004e+12   63.06 9.43e-07 ***
## Residuals            15 4.766e+11 3.178e+10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm(Percent[,'All']~Percent[,'Year'])
```

```
##
## Call:
## lm(formula = Percent[, "All"] ~ Percent[, "Year"])
##
## Coefficients:
##       (Intercept)  Percent[, "Year"]
##          62.39576           -0.03066
```

```
model2<-aov(Percent[,'All']~Percent[,'Year'])
```

```
summary(model2)

##                     Df Sum Sq Mean Sq F value   Pr(>F)
## Percent[, "Year"]  1 0.3836  0.3836    46.6 5.73e-06 ***
## Residuals         15 0.1235  0.0082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm(AllArrests[,'All']*AllArrests[,'Rio.Grande.Valley']~AllArrests[,'Year'])

##
## Call:
## lm(formula = AllArrests[, "All"] * AllArrests[, "Rio.Grande.Valley"] ~
##     AllArrests[, "Year"])
##
## Coefficients:
##          (Intercept)  AllArrests[, "Year"]
##            1.209e+13            -5.978e+09

model3<-aov(AllArrests[,'All']*AllArrests[,'Rio.Grande.Valley']~AllArrests[,'
Year'])

summary(model3)

##                        Df    Sum Sq   Mean Sq F value Pr(>F)
## AllArrests[, "Year"]  1 1.458e+22 1.458e+22   7.209  0.017 *
## Residuals            15 3.034e+22 2.023e+21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model1,model2)

## Warning in anova.lmlist(object, ...): models with response '"Percent[,
## \"All\"]"' removed because response differs from model 1

## Analysis of Variance Table
##
## Response: AllArrests[, "All"]
##                        Df     Sum Sq    Mean Sq F value    Pr(>F)
## AllArrests[, "Year"]  1 2.0038e+12 2.0038e+12  63.057 9.434e-07 ***
## Residuals            15 4.7665e+11 3.1777e+10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model2,model3)

## Warning in anova.lmlist(object, ...): models with response '"AllArrests[,
## \"All\"] * AllArrests[, \"Rio.Grande.Valley\"]"' removed because response
## differs from model 1

## Analysis of Variance Table
##
```

```
## Response: Percent[, "All"]
##                   Df  Sum Sq Mean Sq F value    Pr(>F)    
## Percent[, "Year"]  1 0.38362 0.38362  46.599 5.731e-06 ***
## Residuals         15 0.12349 0.00823                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model1,model3)
```

```
## Warning in anova.lmlist(object, ...): models with response '"AllArrests[,
## \"All\"] * AllArrests[, \"Rio.Grande.Valley\"]"' removed because response
## differs from model 1
```

```
## Analysis of Variance Table
## 
## Response: AllArrests[, "All"]
##                      Df     Sum Sq    Mean Sq F value    Pr(>F)    
## AllArrests[, "Year"]  1 2.0038e+12 2.0038e+12  63.057 9.434e-07 ***
## Residuals            15 4.7665e+11 3.1777e+10                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

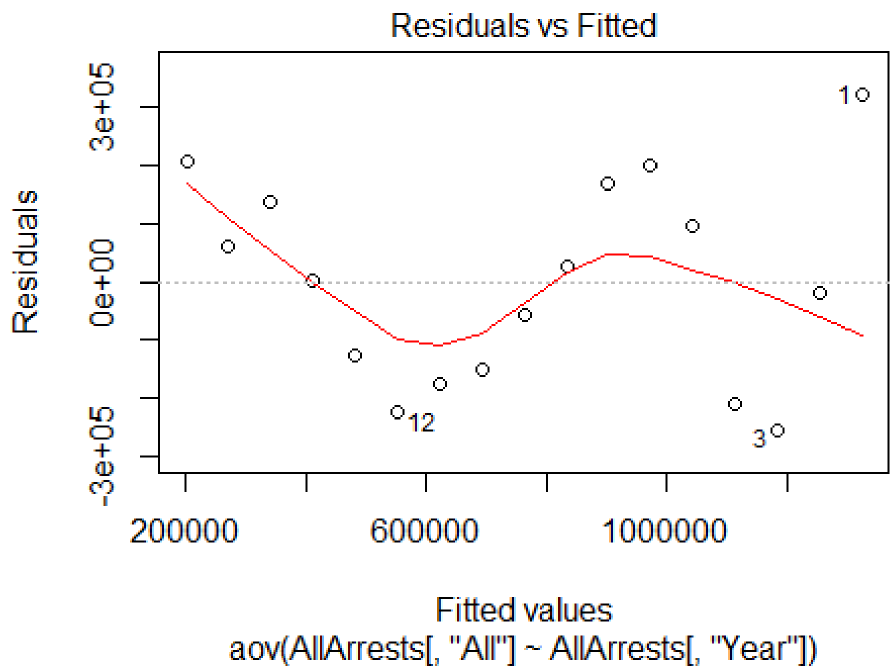
```
#Based on the above anova runs, there is not a huge difference between each o
f the tests.  Based on the p value of the first model, it should provide suff
icient information to estimate for 2017.
lmMod <-lm(AllArrests[,'All']~AllArrests[,'Year'])
summary(lmMod)
```

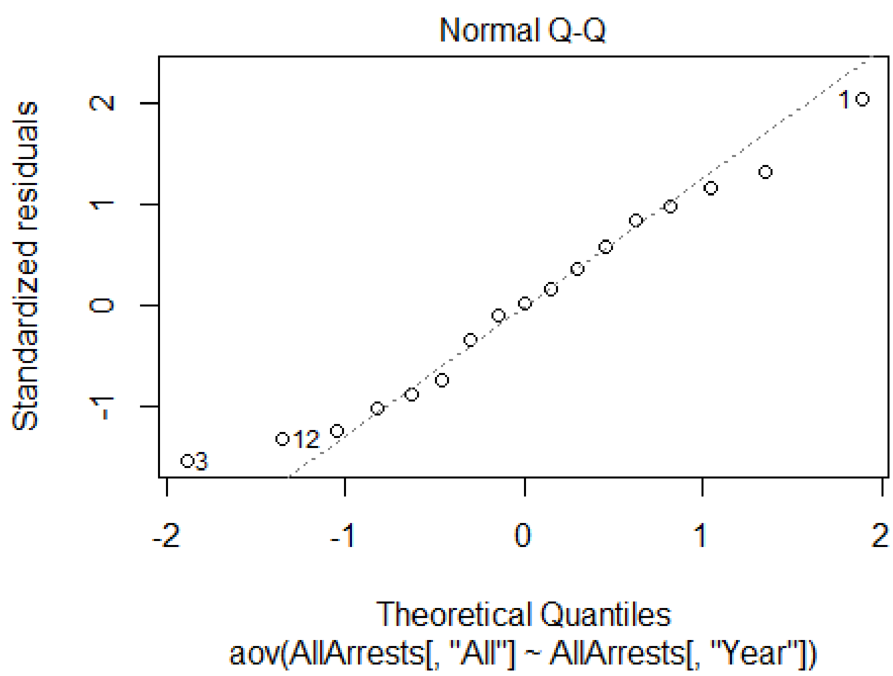
```
## 
## Call:
## lm(formula = AllArrests[, "All"] ~ AllArrests[, "Year"])
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -253468 -151854    1996  137050  320243 
## 
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          141482666   17721008   7.984 8.83e-07 ***
## AllArrests[, "Year"]    -70080       8825  -7.941 9.43e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 178300 on 15 degrees of freedom
## Multiple R-squared:  0.8078, Adjusted R-squared:  0.795 
## F-statistic: 63.06 on 1 and 15 DF,  p-value: 9.434e-07
```

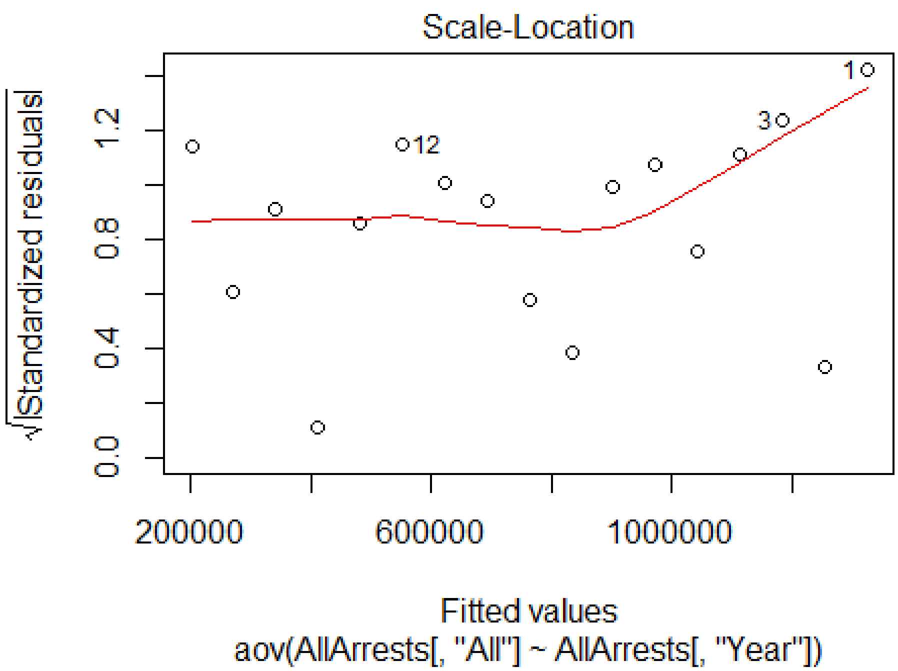
```
#Estimate 2017 value
141482666-(70080*2017)
```

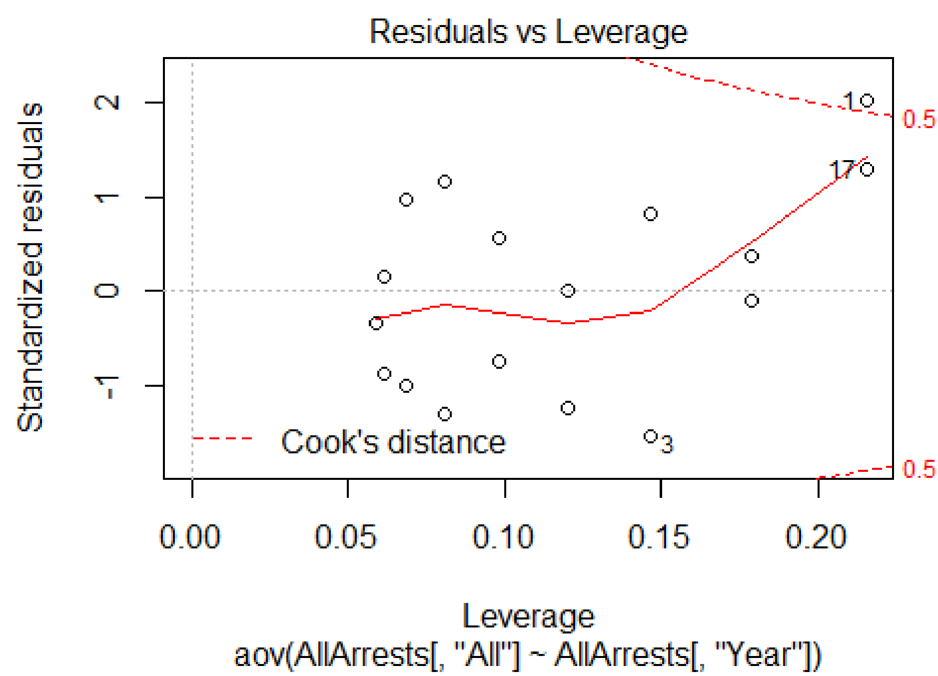
```
## [1] 131306
```

```r
plot(model1)
```



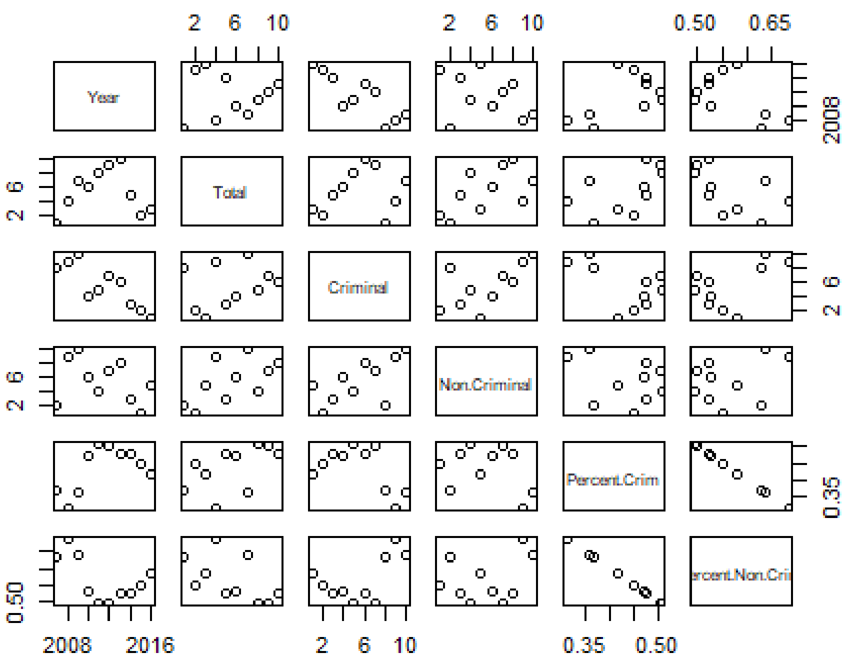




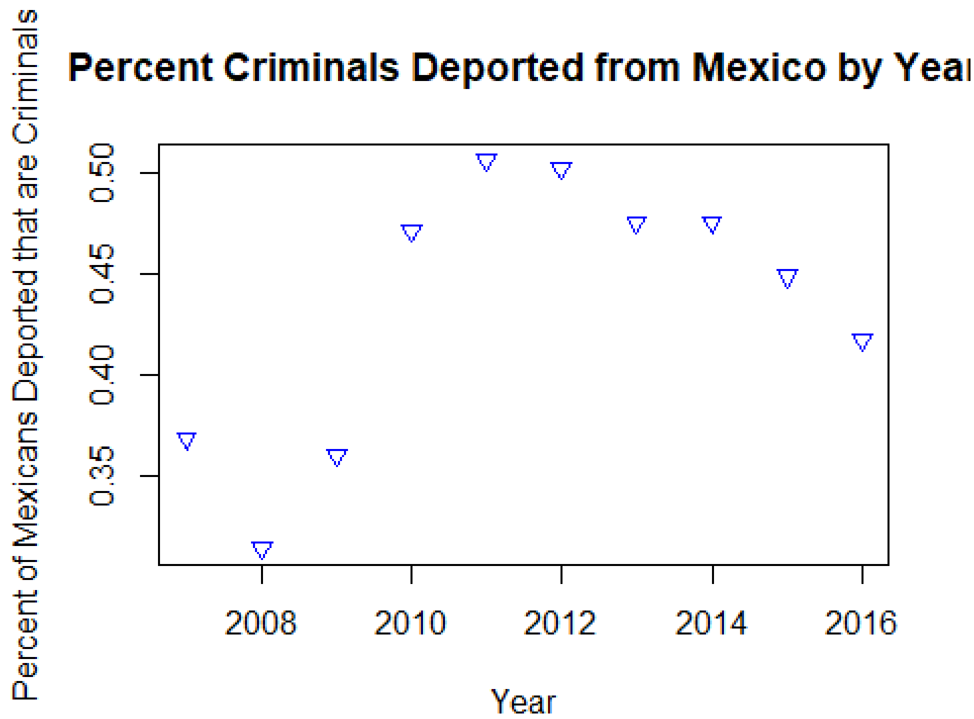



```
DepMexicans<-read.csv("C:\\Users\\etlaw\\OneDrive\\Documents\\MSBA 320\\Final
 Project\\Mexico41D.csv")

attach(DepMexicans)
```

```
## The following object is masked from AllArrests:
##
##     Year

## The following object is masked from Percent:
##
##     Year

## The following object is masked from Mexican:
##
##     Year

## The following object is masked from Southwest:
##
##     Year

names(DepMexicans)

## [1] "Year"              "Total"             "Criminal"
## [4] "Non.Criminal"      "Percent.Crim"      "Percent.Non.Crim"

plot(DepMexicans)
```



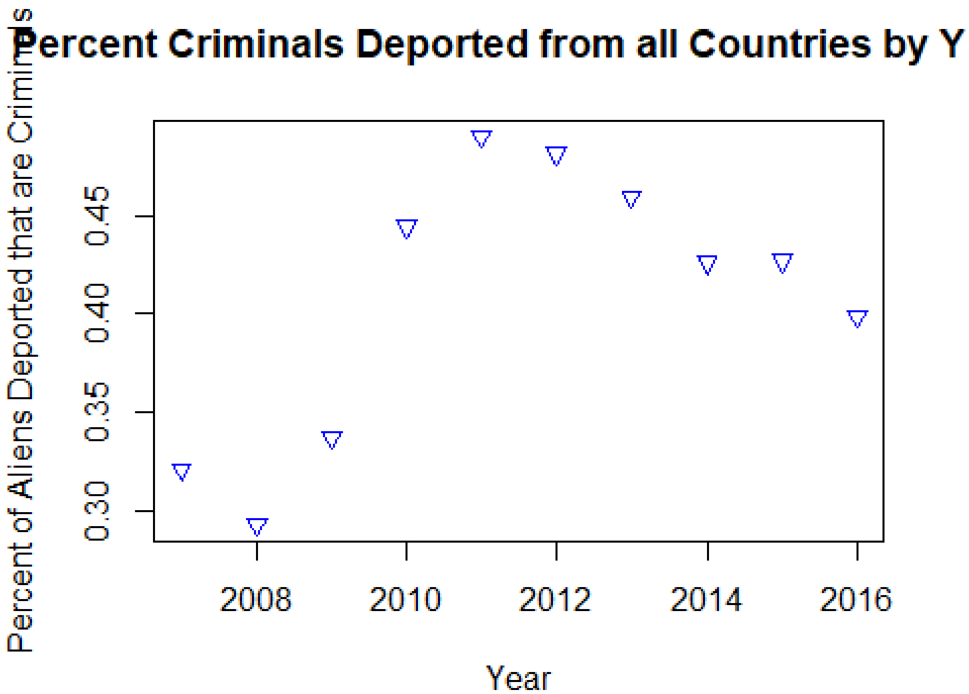

```
plot(DepMexicans[,'Year'],DepMexicans[,'Percent.Crim'],pch=6,col="blue",xlab=
"Year",ylab="Percent of Mexicans Deported that are Criminals",main="Percent C
riminals Deported from Mexico by Year")
```




```
DepUSA<-read.csv("C:\\Users\\etlaw\\OneDrive\\Documents\\MSBA 320\\Final Proj
ect\\All41D.csv")

attach(DepUSA)

## The following objects are masked from DepMexicans:
##
##     Criminal, Percent.Crim, Total, Year

## The following object is masked from AllArrests:
##
##     Year

## The following object is masked from Percent:
##
##     Year

## The following object is masked from Mexican:
##
##     Year

## The following object is masked from Southwest:
##
##     Year

names(DepUSA)
```

```
## [1] "Year"              "Total"             "Criminal"          "NonCrim"

## [5] "Percent.Crim"    "Percent.NonCrim"
```

```
plot(DepUSA[,'Year'],DepUSA[,'Percent.Crim'],pch=6,col="blue",xlab="Year",yla
b="Percent of Aliens Deported that are Criminals",main="Percent Criminals Dep
orted from all Countries by Year")
```




```
t.test(DepMexicans[,'Percent.Crim'],DepUSA[1:10,'Percent.Crim'],alt="two.side
d")
```

```
##
##  Welch Two Sample t-test
##
## data:  DepMexicans[, "Percent.Crim"] and DepUSA[1:10, "Percent.Crim"]
## t = 0.86577, df = 17.97, p-value = 0.398
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03736340  0.08973211
## sample estimates:
## mean of x mean of y
## 0.4338738 0.4076894
```