Stephanie Cruz

Scripting For Data Analysis

Table of Contents

[Project Background and Description 2](#_Toc42439025)

[Project Scope and Context of the Analysis 2](#_Toc42439026)

[Data Description 2](#_Toc42439027)

[Research Questions 2](#_Toc42439028)

[Initial Assessment and Unknowns 3](#_Toc42439029)

[Cleaning the Data 3](#_Toc42439030)

[Data Analysis Methods 3](#_Toc42439031)

[Analysis 3](#_Toc42439032)

[Breakdown of Types of Gun Violence by Congressional District 4](#_Toc42439033)

[Congressional Districts with Worst Mass Shooting Numbers 4](#_Toc42439034)

[Congressional Districts with Worst Officer Involved Shooting Numbers 7](#_Toc42439035)

[Congressional Districts with Worst School Shooting Numbers 10](#_Toc42439036)

[Comparison of 2014-2015 Gun Violence Numbers to 2016-2017 Gun Violence Numbers 12](#_Toc42439037)

[Relationship Between Officer Involved Shootings and Overall Gun Violence Numbers 14](#_Toc42439038)

[Relationship Between Number of State Laws and Overall Gun Violence Numbers 17](#_Toc42439039)

[Summary of Results 18](#_Toc42439040)

[Next Steps and Outstanding Questions 19](#_Toc42439041)

[Appendix A: Python Code 19](#_Toc42439042)

# Project Background and Description

An analysis will be performed on three datasets that relate to firearm provisions in the United States, gun violence incidents in the United States and election info as it pertains to congressional districts. This data will be used to analyze and determine whether relationships exist between the three.

# Project Scope and Context of the Analysis

The scope of this analysis includes looking at firearm provisions and gun violence incidents between the years 2014-2017 in the United States. Election data will be analyzed for the 2014, 2016 and 2018 elections and encompasses the timeframe captured in the gun violence and firearm provisions datasets. Analytical techniques such as descriptive statistics, visualizations, and linear modeling will be used to address the research questions outlined below. The analysis will identify datapoints that are correlated with each other as well as determine if the number of officer involved shootings and the number of state gun laws are predictive indicators of gun violence.

# Data Description

The following three datasets were used for this data analysis project:

* **Firearms Provisions in US States**
  + This data is from the State Firearm Laws project which has a goal of providing data to determine the effectiveness of firearm laws. The data obtained cover firearm laws for all 50 states from 1991-2017 and includes information about firearm laws as they pertain to: firearm sales, background checks, firearm purchase age restrictions, handgun restrictions, vendor licensing requirements, firearm sales record retention, and ammunition sales.
* **Aggregate US Congressional District Data**
  + This data is an aggregation of US house data from the 2012 election through the 2018 election. The data set includes information about spending, election data, demographic data and geographic data.
* **Gun Violence Data**
  + This data was scraped from the Gun Violence Archive and includes incidents from 1/1/2013-3/31/2018. It includes data about gun violence incidents, including state and congressional district, incident characteristics, number of people involved, etc.

All three datasets were obtained from Kaggle.

# Research Questions

The data analysis will answer the following four research questions:

1. What is the breakdown of types of gun violence incidents by congressional districts?
2. Compare 2014-2015 gun violence numbers to 2016-2017 gun violence numbers. What happened to gun violence rates after the nominal of Donald Trump?
3. How many gun violence incidents involved a police officer and is this predictive of the overall number of gun violence incidents?
4. How do the number of state gun laws relate to the amount of gun violence incidents?

# Initial Assessment and Unknowns

After an initial assessment of the data, it was determined that the data is reasonably complete, with the following caveats:

* The 2013 and 2018 years in the ‘Gun Violence Data’ dataset appear to be incomplete as there are significantly less incidents for these two years than the 2014-2017 subset. Therefore, these years will be omitted from our analysis.
* The ‘Gun Violence Data’ dataset is missing congressional district information for about 5% of the rows. As this is below the 10% missingness threshold that would indicate the results could be biased, these rows will be dropped from the dataset.
* The first year of election information (2012) is after redistricting, and therefore difficult to compare to previous election cycles. For this reason, we will omit it from our analysis.

It should also be noted that while all congressional districts have approximately the same number of people in them (711,000), they vary widely with regards to population density. Population density information was not included in our data analysis, so that should be kept in mind when reviewing analysis between congressional districts.

# Cleaning the Data

The data obtained was fairly complete and clean. Only a small amount of data cleaning was needed in order to conduct the data analysis, as outlined below:

* **Firearms Provision in US States**
  + Remove columns we do not want to analyze
  + Remove rows for years 1991-2012
* **Aggregate US Congressional District Data**
  + Remove columns we do not want to analyze
* **Gun Violence Data**
  + Remove columns we do not want to analyze
  + Remove rows that do not include congressional district information
  + Remove rows for years 2013 and 2018

# Data Analysis Methods

The following methods were used in the data analysis:

* Descriptive statistics to describe the breakdown of the data
* Visualizations of the data including histograms and tables
* Correlation calculation
* Regression analysis

# Analysis

An in-depth analysis was done of types of gun violence by congressional districts as well as the relationship between officer involved shootings and state laws and the total number of gun violence incidents in a state per year. An analysis was also done to determine the change between gun violence numbers before the republican nomination of Donald Trump and after. See in depth results below.

## Breakdown of Types of Gun Violence by Congressional District

For this analysis, the worst ten districts for each category of gun violence incident are included in a table along with a histogram breakdown of the number of incidents of that category of gun violence that occurred in each congressional district

### Congressional Districts with Worst Mass Shooting Numbers

For the category of mass shootings, the worst ten districts for each of the years 2014-2017 are shown below along with the histogram breakdowns. It should be noted that a ‘mass shooting’ is described as a shooting involving four or more victims.

#### 2014 Overview

In the below table, we can see that Illinois’ seventh district had the highest number of mass shootings with ten shootings in 2014. Also notable is that two districts in Michigan are in the bottom ten.



We can see below that most districts had very few mass shootings, but a small number had significantly more.

A screenshot of a cell phone

Description automatically generated

#### 2015 Overview

In the below table we can see that Illinois’ seventh district again had the highest number of mass shootings – this time with fifteen mass shootings in 2015. Louisiana’s second district and Florida’s twenty-fourth district were in the bottom ten congressional districts for the second year in a row.



We can see again that the majority of districts had no mass shootings and only a select few had a high number of them.

A screenshot of a cell phone

Description automatically generated

#### 2016 Overview

We can see once again that Illinois’ seventh district had the worst mass shooting numbers, recording nineteen in 2016. Louisiana’s second was in the bottom ten for the third year in a row.



The histogram below shows again that the vast majority of districts do not have mass shooting incidents.

A close up of a piece of paper

Description automatically generated

#### 2017 Overview

In 2017, Illinois’ seventh once again had the most instances of mass shootings, and Louisiana’s second is in the bottom ten for the fourth year in a row. Also notable is that Mississippi has two districts in the bottom ten for 2017.



As expected, the 2017 histogram data shows that the majority of districts do not experience mass shooting incidents, while some districts experience a far larger number.

A close up of a piece of paper

Description automatically generated

### Congressional Districts with Worst Officer Involved Shooting Numbers

For the category of officer involved shootings, the worst ten districts for each of the years 2014-2017 are shown below along with their histogram breakdowns.

#### 2014 Overview

In 2014, Missiouri’s first district had the most officer involved shootings with thirty-six of them. Also notable is that both Illinois’ seventh and Louisiana’s second district are both in the bottom ten for officer involved shootings



In the histogram breakdown below, we can see that the majority of districts had between zero and ten officer involved shootings, and a small number of districts had over twenty.

A screenshot of a cell phone

Description automatically generated

#### 2015 Overview

In 2015, Ohio’s eleventh district had the most officer involved shootings with forty-nine incidents. Illinois’ seventh and Louisiana’s second are again in the bottom ten districts.



Below we can see that the majority of districts experienced between five and ten officer involved shootings, which is an increase from the previous year.

A screenshot of a cell phone

Description automatically generated

#### 2016 Overview

We can see that in 2016, Indiana’s seventh had the most officer involved shootings with sixty. Louisiana’s second is the third worst. We can also see that the highest number of officer involved shootings in a single district has increased three years in a row.



We can see below that once again, most districts had between five and ten officer involved shootings.

A screenshot of a cell phone

Description automatically generated

#### 2017 Overview

In 2017, Missouri’s first district had the most officer involved shootings with forty-three. Illinois’ seventh was in the bottom ten again.



The average number of officer involved incidents declined in 2017 to between below ten for the first time in two years.

A screenshot of a cell phone

Description automatically generated

### Congressional Districts with Worst School Shooting Numbers

For the category of school shootings, the worst ten districts for each of the years 2014-2017 are shown below along with their histogram breakdowns.

#### 2014 Overview

In 2014, Tennessee’s ninth district and Alabama’s first district both had eight gun violence incidents involving a school. Florida had two district’s in the bottom ten.



In 2014, the vast majority of districts did not have any incidents involving gun violence at a school.

A close up of a piece of paper

Description automatically generated

#### 2015 Overview

Tennessee’s ninth district once again had the most gun violence incidents at a school, with ten incidents. Both Tennessee and Georgia had two districts in the bottom ten.



In 2015 only 160 districts had no gun violence cases at a school. Almost two thirds of districts had at least one incident.

A screenshot of a cell phone

Description automatically generated

#### 2016 Overview

Florida’s fourth district and Tennessee’s ninth had the most gun violence incidents at a school, with eleven instances. Tennessee once again had both their ninth and fifth districts in the bottom ten.



Once again, the vast majority of districts did not have a gun violence incident at a school.

A screenshot of a cell phone

Description automatically generated

#### 2017 Overview

Tennessee’s ninth and Florida’s fourth once again had the worst numbers, with twelve gun violence incidents at a school. Tennessee again had two districts in the bottom ten, as did Kentucky.



Once again, in 2017, the majority of districts did not report a gun violence incident involving a school.

A screenshot of a cell phone

Description automatically generated

## Comparison of 2014-2015 Gun Violence Numbers to 2016-2017 Gun Violence Numbers

An analysis was completed to determine how gun violence numbers have changed around and after the election in 2016. The average number of gun violence incidents was obtained for both 2014-2015 and 2016-2017. We can see from the table below that some districts decreased and some increased, but overall, the number of gun violence incidents has risen 6.67 percent since Donald Trump’s nomination as the republican presidential candidate.



## Relationship Between Officer Involved Shootings and Overall Gun Violence Numbers

An analysis was done to determine the relationship between officer involved shootings and the occurrence of other types of gun violence for each year from 2014-2017.

#### 2014 Analysis:

We can see from the following correlation matrix that the number officer involved shootings are positively correlated with every other kind of gun violence, with the highest correlation existing between officer involved shootings and mass shootings. This means there is a relationship between officer involved shootings and every other form of gun violence we are analyzing, and as one increases so will the other.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2014 Correlation Matrix** | | | | | | |
|  | **mass** | **officer** | **domestic** | **ms** | **spree** | **school** |
| **mass** | 1 | 0.449013 | 0.201428 | 0.098405 | 0.157934 | 0.226097 |
| **officer** | 0.449013 | 1 | 0.369286 | 0.220239 | 0.155151 | 0.230092 |
| **domestic** | 0.201428 | 0.369286 | 1 | 0.287385 | 0.179672 | 0.331708 |
| **ms** | 0.098405 | 0.220239 | 0.287385 | 1 | 0.034635 | 0.069603 |
| **spree** | 0.157934 | 0.155151 | 0.179672 | 0.034635 | 1 | 0.151189 |
| **school** | 0.226097 | 0.230092 | 0.331708 | 0.069603 | 0.151189 | 1 |

A regression analysis was performed to determine if we can predict the number of total gun violence incidents based on the number of officer involved shootings. The results are shown below and the R-squared value is .668, which means that we can predict the total number of incidents based on the number of officer involved shootings, and that about 67% of the variance in the total number of gun violence incidents can be explained by the number officer involved shootings. Keep in mind that due to the high correlation numbers we cannot say that officer involved shootings are causes other forms of gun violence, just that we can predict one from the other.

**2014 Regression Results:**

OLS Regression Results

=======================================================================================

Dep. Variable: officer R-squared (uncentered): 0.669

Model: OLS Adj. R-squared (uncentered): 0.668

Method: Least Squares F-statistic: 886.7

Date: Sun, 07 Jun 2020 Prob (F-statistic): 2.04e-107

Time: 12:42:47 Log-Likelihood: -1339.6

No. Observations: 440 AIC: 2681.

Df Residuals: 439 BIC: 2685.

Df Model: 1

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

numIncidents 0.0422 0.001 29.778 0.000 0.039 0.045

==============================================================================

Omnibus: 74.773 Durbin-Watson: 1.309

Prob(Omnibus): 0.000 Jarque-Bera (JB): 435.611

Skew: -0.560 Prob(JB): 2.56e-95

Kurtosis: 7.744 Cond. No. 1.00

#### 2015 Analysis

After generating a correlation matrix for the 2015 data, we can see that, once again, the number of officer involved shootings is positively correlated with every other form of gun violence. This time, both domestic violence incidents and mass shooting incidents are strongly positively correlated with officer involved shootings.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2015 Correlation Matrix** | | | | | | |
|  | **mass** | **officer** | **domestic** | **ms** | **spree** | **school** |
| **mass** | 1 | 0.457475 | 0.287985 | 0.085886 | 0.076735 | 0.278353 |
| **officer** | 0.457475 | 1 | 0.510509 | 0.185117 | 0.086627 | 0.29266 |
| **domestic** | 0.287985 | 0.510509 | 1 | 0.383565 | 0.070429 | 0.469685 |
| **ms** | 0.085886 | 0.185117 | 0.383565 | 1 | -0.02948 | 0.201645 |
| **spree** | 0.076735 | 0.086627 | 0.070429 | -0.02948 | 1 | 0.008035 |
| **school** | 0.278353 | 0.29266 | 0.469685 | 0.201645 | 0.008035 | 1 |

Once again, we can see based on our 2015 data that once again, the number of officer involved shootings are predictive of the overall number of gun violence incidents, with 68% of the variance explained by the number of officer involved shootings.

**2015 Regression Results:**

OLS Regression Results

=======================================================================================

Dep. Variable: officer R-squared (uncentered): 0.680

Model: OLS Adj. R-squared (uncentered): 0.679

Method: Least Squares F-statistic: 927.1

Date: Sun, 07 Jun 2020 Prob (F-statistic): 5.55e-110

Time: 12:46:57 Log-Likelihood: -1462.4

No. Observations: 437 AIC: 2927.

Df Residuals: 436 BIC: 2931.

Df Model: 1

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

numIncidents 0.0579 0.002 30.449 0.000 0.054 0.062

==============================================================================

Omnibus: 57.213 Durbin-Watson: 1.129

Prob(Omnibus): 0.000 Jarque-Bera (JB): 344.514

Skew: -0.329 Prob(JB): 1.55e-75

Kurtosis: 7.300 Cond. No. 1.00

#### 2016 Analysis

The 2016 data is consistent with previous years, showing a positive correlation between officer involved shootings are all other forms of gun violence.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2016 Correlation Matrix** | | | | | | |
|  | **mass** | **officer** | **domestic** | **ms** | **spree** | **school** |
| **mass** | 1 | 0.376956 | 0.259967 | 0.078221 | 0.132042 | 0.223421 |
| **officer** | 0.376956 | 1 | 0.558471 | 0.178382 | 0.096348 | 0.322798 |
| **domestic** | 0.259967 | 0.558471 | 1 | 0.439361 | 0.196599 | 0.455534 |
| **ms** | 0.078221 | 0.178382 | 0.439361 | 1 | 0.044619 | 0.21155 |
| **spree** | 0.132042 | 0.096348 | 0.196599 | 0.044619 | 1 | 0.087144 |
| **school** | 0.223421 | 0.322798 | 0.455534 | 0.21155 | 0.087144 | 1 |

The 2016 regression results show similar results as well, however this year only 62% of the variance can be explained by the number of officer involved shootings.

**2016 Regression Results:**

OLS Regression Results

=======================================================================================

Dep. Variable: officer R-squared (uncentered): 0.620

Model: OLS Adj. R-squared (uncentered): 0.619

Method: Least Squares F-statistic: 713.2

Date: Sun, 07 Jun 2020 Prob (F-statistic): 7.09e-94

Time: 12:48:24 Log-Likelihood: -1528.7

No. Observations: 438 AIC: 3059.

Df Residuals: 437 BIC: 3064.

Df Model: 1

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

numIncidents 0.0549 0.002 26.706 0.000 0.051 0.059

==============================================================================

Omnibus: 91.631 Durbin-Watson: 1.037

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1978.497

Skew: -0.105 Prob(JB): 0.00

Kurtosis: 13.410 Cond. No. 1.00

#### 2017 Analysis:

As in previous years, officer involved shootings are positively correlated with all other forms of gun violence.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2017 Correlation Matrix** | | | | | | |
|  | **mass** | **officer** | **domestic** | **ms** | **spree** | **school** |
| **mass** | 1 | 0.391482 | 0.249649 | 0.118732 | 0.171975 | 0.258102 |
| **officer** | 0.391482 | 1 | 0.588241 | 0.336775 | 0.3006 | 0.32688 |
| **domestic** | 0.249649 | 0.588241 | 1 | 0.483758 | 0.298708 | 0.470029 |
| **ms** | 0.118732 | 0.336775 | 0.483758 | 1 | 0.196409 | 0.277296 |
| **spree** | 0.171975 | 0.3006 | 0.298708 | 0.196409 | 1 | 0.229932 |
| **school** | 0.258102 | 0.32688 | 0.470029 | 0.277296 | 0.229932 | 1 |

As expected, officer involved shootings once again predict the number of overall incidents, with 63% of the variance in the total number of shootings being explained by the number of officer involved shootings.

**2017 Regression:**

OLS Regression Results

=======================================================================================

Dep. Variable: officer R-squared (uncentered): 0.631

Model: OLS Adj. R-squared (uncentered): 0.631

Method: Least Squares F-statistic: 750.5

Date: Sun, 07 Jun 2020 Prob (F-statistic): 5.46e-97

Time: 12:49:25 Log-Likelihood: -1455.7

No. Observations: 439 AIC: 2913.

Df Residuals: 438 BIC: 2917.

Df Model: 1

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

numIncidents 0.0455 0.002 27.395 0.000 0.042 0.049

==============================================================================

Omnibus: 145.712 Durbin-Watson: 1.268

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1032.069

Skew: -1.233 Prob(JB): 7.75e-225

Kurtosis: 10.095 Cond. No. 1.00

==============================================================================

## Relationship Between Number of State Laws and Overall Gun Violence Numbers

An analysis was completed to determine if the number of firearm laws in a state has an affect on the number of gun violence incidents.

We can see below that the state laws are consistently positively correlated with the number of gun violence incidents. This is the opposite of what was expected – a negative correlation would be expected given that one would expect that the more laws regulating guns would lead to a decline in the number of gun violence incidents.



A regression analysis was completed to determine if the number of laws is a predictive factor for the number of gun violence incidents. We can see below that it is, and the R-squared value is .468, which means that roughly 47% of the variance in gun violence incidents can be explained by the number of state laws that pertain to gun regulation. Once again, because of the high correlation numbers, we cannot say that they are causing gun violence incidents, just that we can predict one from the other.

**Regression Results:**

OLS Regression Results

=======================================================================================

Dep. Variable: lawtotal2017 R-squared (uncentered): 0.479

Model: OLS Adj. R-squared (uncentered): 0.468

Method: Least Squares F-statistic: 44.99

Date: Sun, 07 Jun 2020 Prob (F-statistic): 1.88e-08

Time: 13:07:29 Log-Likelihood: -236.64

No. Observations: 50 AIC: 475.3

Df Residuals: 49 BIC: 477.2

Df Model: 1

Covariance Type: nonrobust

==================================================================================

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

2017 Incidents 0.0167 0.002 6.707 0.000 0.012 0.022

==============================================================================

Omnibus: 7.441 Durbin-Watson: 1.721

Prob(Omnibus): 0.024 Jarque-Bera (JB): 6.462

Skew: 0.822 Prob(JB): 0.0395

Kurtosis: 3.631 Cond. No. 1.00

==============================================================================

# Summary of Results

Our overall analysis found the following with regards to how gun violence, congressional districts, and the number of state laws pertaining to gun violence interact.

* Certain congressional districts (for example: Illinois’s seventh and Tennessee’s ninth) have significant gun violence problems in nearly all categories that were analyzed
* More than half of all congressional districts have had between five and ten officer involved gun violence incidents in every year from 2014-2017
* More than half of all congressional districts had at least one gun violence incident at a school in 2016
* Congressional districts that started with high levels of gun violence saw large increases after the nomination of Donald Trump as the republican presidential nominee
* Since the nomination of Donald Trump as the republic presidential nominee, there has been a 6.67% increase in total gun violence incidents nation wide
* Officer involved shootings are positively correlated with all other kinds of gun violence for all four years that were analyzed
* Officer involved shootings are predictive of the total number of gun violence incidents in a congressional district
* An average of 64% of the variance in the total number of gun violence incidents can be explained by the number of officer involved shootings
* The number of state laws are positively correlated with the total number of gun violence incidents
* The number of state laws is predictive of the total number of gun violence incidents
* An average of 47% of the variance in the total number of gun violence incidents for an entire state in a year can be explained by the number of state laws regulating gun violence

# Next Steps and Outstanding Questions

* If would be useful to analyze state representatives as well as federal representatives
* Population density should be analyzed for each congressional district to determine how that affects the number of gun violence incidents; it would also allow for more helpful comparisons between similar congressional districts
* This analysis focused on officer involved shootings, but multivariable regression could be completed to determine how all factors affect the overall number of gun violence incidents
* More examination needs to be done to understand why there is a positive correlation between the total number of gun violence incidents and the total number of state laws regulating gun violence. Questions that could be asked:
  1. Do states pass more gun regulations because there are already high rates of gun violence in the state
  2. Do states that have more gun regulations have more high-density cities
  3. Are there grandfather clauses incorporated into the gun regulations which may decrease the effectiveness of the regulation

# Appendix A: Python Code

# -\*- coding: utf-8 -\*-

"""

Created on Sat Jun 6 12:13:12 2020

@author: thecr

"""

import pandas

import numpy

import csv

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator

import matplotlib.pyplot as plt

from sklearn import linear\_model

import statsmodels.api as sm

#Read our data from csv

electionInfoData = pandas.read\_csv('election\_info.csv', header='infer', nrows=436)

gunViolenceData = pandas.read\_csv('gun\_violence\_data.csv', header='infer', nrows=239678)

stateFirearmLawData = pandas.read\_csv('state\_firearms.csv', header='infer', nrows=1351)

#Trim our data to only include the columns we want

electionInfoData = electionInfoData[['District', 'rep\_party\_2012', 'winning\_party\_2012', 'rep\_party\_2014',

'winning\_party\_2014', 'rep\_party\_2016', 'winning\_party\_2016', 'rep\_party\_2018',

'winning\_party\_2018','white','black','latino','asian and pacific island',

'native','other','bach degree or higher among 25up','whiteBA','median\_income',

'noncollege\_white', 'CITYLAB\_CDI']]

gunViolenceData = gunViolenceData[['date','state','n\_killed','n\_injured','congressional\_district',

'gun\_type','incident\_characteristics','participant\_age\_group','participant\_gender',

'participant\_status','participant\_type']]

#Drop data that doesn't include the congressional district

gunViolenceData = gunViolenceData[gunViolenceData['congressional\_district'] != 0]

gunViolenceData = gunViolenceData.dropna(subset=['congressional\_district'])

#split gunViolenceData into 4 data frames based on the year

gunViolenceData2014 = gunViolenceData[gunViolenceData['date'].str.contains('2014')]

gunViolenceData2015 = gunViolenceData[gunViolenceData['date'].str.contains('2015')]

gunViolenceData2016 = gunViolenceData[gunViolenceData['date'].str.contains('2016')]

gunViolenceData2017 = gunViolenceData[gunViolenceData['date'].str.contains('2017')]

#Trim state firearm law data to only include the columns we want

stateFirearmLawData = stateFirearmLawData[['state', 'year','lawtotal']]

#Remove years from state firearm law data that we aren't looking at for this project

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1991].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1992].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1993].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1994].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1995].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1996].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1997].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1998].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==1999].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2000].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2001].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2002].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2003].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2004].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2005].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2006].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2007].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2008].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2009].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2010].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2011].index

stateFirearmLawData.drop(indexNames, inplace=True)

indexNames = stateFirearmLawData[stateFirearmLawData['year']==2012].index

stateFirearmLawData.drop(indexNames, inplace=True)

#Define column names for aggregated gun violence data

columnNames = ['state','district','numIncidents','numKilled','numInjured']

#Create aggregated gun violence dataframes for each year that we are looking at

aggregated2014gunViolenceData = pandas.DataFrame(columns = columnNames)

aggregated2015gunViolenceData = pandas.DataFrame(columns = columnNames)

aggregated2016gunViolenceData = pandas.DataFrame(columns = columnNames)

aggregated2017gunViolenceData = pandas.DataFrame(columns = columnNames)

#Group our gun violence data by state and congressional district

grouped2014gunViolenceData = gunViolenceData2014.groupby(['state', 'congressional\_district'])

grouped2015gunViolenceData = gunViolenceData2015.groupby(['state', 'congressional\_district'])

grouped2016gunViolenceData = gunViolenceData2016.groupby(['state', 'congressional\_district'])

grouped2017gunViolenceData = gunViolenceData2017.groupby(['state', 'congressional\_district'])

#TODO Wordcloud -- see below

#for name, group in grouped2014gunViolenceData:

# for i in group['incident\_characteristcs']:

# wordcloud = WordCloud().generate(i)

# plt.imshaw(wordcloud,interpolation='bilinear')

#Loop through our grouped gun violence data for 2014

incidentCharString2014 = ""

for name, group in grouped2014gunViolenceData:

state = name[0]

district = name[1]

numIncidents = group['n\_killed'].count()

numKilled = group['n\_killed'].sum()

numInjured = group['n\_injured'].sum()

massShooting = 0

officerInvolvedIncident = 0

domesticViolence = 0

murderSuicide = 0

spreeShooting = 0

schoolIncident = 0

#Look for key words in our incident characteristics column

for i in group['incident\_characteristics']:

try:

incidentCharString2014 = incidentCharString2014 + i

if(i.find('Mass Shooting') > -1):

massShooting +=1

if(i.find('Officer Involved Incident') > -1):

officerInvolvedIncident +=1

if(i.find('Domestic Violence') > -1):

domesticViolence +=1

if(i.find('Murder/Suicide') > -1):

murderSuicide +=1

if(i.find('Spree Shooting') > -1):

spreeShooting +=1

if(i.find('School Incident') > -1):

schoolIncident +=1

except AttributeError:

print('error')

except TypeError:

print('error')

#Do this outside, concatenate and then create the wordcloud

#wordcloud = WordCloud().generate(i)

#plt.imshow(wordcloud,interpolation='bilinear')

#create an array of our data

data = [[state, district, numIncidents, numKilled, numInjured, massShooting, officerInvolvedIncident, domesticViolence, murderSuicide, spreeShooting, schoolIncident]]

#Create a temp dataframe

tempDf = pandas.DataFrame(data, columns = ['state','district','numIncidents','numKilled','numInjured', 'mass', 'officer', 'domestic','ms','spree','school'])

#Append it to our existing dataframe

aggregated2014gunViolenceData = aggregated2014gunViolenceData.append(tempDf, True)

#print(incidentCharString2014)

#wordcloud = WordCloud().generate(incidentCharString2014)

#plt.imshow(wordcloud, interpolation='bilinear')

#plt.axis('off')

#plt.show()

#Loop through our grouped gun violence data for 2015

for name, group in grouped2015gunViolenceData:

state = name[0]

district = name[1]

numIncidents = group['n\_killed'].count()

numKilled = group['n\_killed'].sum()

numInjured = group['n\_injured'].sum()

massShooting = 0

officerInvolvedIncident = 0

domesticViolence = 0

murderSuicide = 0

spreeShooting = 0

schoolIncident = 0

#Look for key words in our incident characteristics column

for i in group['incident\_characteristics']:

try:

if(i.find('Mass Shooting') > -1):

massShooting +=1

if(i.find('Officer Involved Incident') > -1):

officerInvolvedIncident +=1

if(i.find('Domestic Violence') > -1):

domesticViolence +=1

if(i.find('Murder/Suicide') > -1):

murderSuicide +=1

if(i.find('Spree Shooting') > -1):

spreeShooting +=1

if(i.find('School Incident') > -1):

schoolIncident +=1

except AttributeError:

print('error')

#create an array of our data

data = [[state, district, numIncidents, numKilled, numInjured, massShooting, officerInvolvedIncident, domesticViolence, murderSuicide, spreeShooting, schoolIncident]]

#Create a temp dataframe

tempDf = pandas.DataFrame(data, columns = ['state','district','numIncidents','numKilled','numInjured', 'mass', 'officer', 'domestic','ms','spree','school'])

#Append it to our existing dataframe

aggregated2015gunViolenceData = aggregated2015gunViolenceData.append(tempDf, True)

#Loop through our grouped gun violence data for 2016

for name, group in grouped2016gunViolenceData:

state = name[0]

district = name[1]

numIncidents = group['n\_killed'].count()

numKilled = group['n\_killed'].sum()

numInjured = group['n\_injured'].sum()

massShooting = 0

officerInvolvedIncident = 0

domesticViolence = 0

murderSuicide = 0

spreeShooting = 0

schoolIncident = 0

#Look for key words in our incident characteristics column

for i in group['incident\_characteristics']:

try:

if(i.find('Mass Shooting') > -1):

massShooting +=1

if(i.find('Officer Involved Incident') > -1):

officerInvolvedIncident +=1

if(i.find('Domestic Violence') > -1):

domesticViolence +=1

if(i.find('Murder/Suicide') > -1):

murderSuicide +=1

if(i.find('Spree Shooting') > -1):

spreeShooting +=1

if(i.find('School Incident') > -1):

schoolIncident +=1

except AttributeError:

print('error')

#create an array of our data

data = [[state, district, numIncidents, numKilled, numInjured, massShooting, officerInvolvedIncident, domesticViolence, murderSuicide, spreeShooting, schoolIncident]]

#Create a temp dataframe

tempDf = pandas.DataFrame(data, columns = ['state','district','numIncidents','numKilled','numInjured', 'mass', 'officer', 'domestic','ms','spree','school'])

#Append it to our existing dataframe

aggregated2016gunViolenceData = aggregated2016gunViolenceData.append(tempDf, True)

#Loop through our grouped gun violence data for 2017

for name, group in grouped2017gunViolenceData:

state = name[0]

district = name[1]

numIncidents = group['n\_killed'].count()

numKilled = group['n\_killed'].sum()

numInjured = group['n\_injured'].sum()

massShooting = 0

officerInvolvedIncident = 0

domesticViolence = 0

murderSuicide = 0

spreeShooting = 0

schoolIncident = 0

#Look for key words in our incident characteristics column

for i in group['incident\_characteristics']:

try:

if(i.find('Mass Shooting') > -1):

massShooting +=1

if(i.find('Officer Involved Incident') > -1):

officerInvolvedIncident +=1

if(i.find('Domestic Violence') > -1):

domesticViolence +=1

if(i.find('Murder/Suicide') > -1):

murderSuicide +=1

if(i.find('Spree Shooting') > -1):

spreeShooting +=1

if(i.find('School Incident') > -1):

schoolIncident +=1

except AttributeError:

print('error')

#create an array of our data

data = [[state, district, numIncidents, numKilled, numInjured, massShooting, officerInvolvedIncident, domesticViolence, murderSuicide, spreeShooting, schoolIncident]]

#Create a temp dataframe

tempDf = pandas.DataFrame(data, columns = ['state','district','numIncidents','numKilled','numInjured', 'mass', 'officer', 'domestic','ms','spree','school'])

#Append it to our existing dataframe

aggregated2017gunViolenceData = aggregated2017gunViolenceData.append(tempDf, True)

#Print our reports of aggregated gun violence data for each year from 2014-2017

aggregated2014gunViolenceData.to\_csv("2014AggregatedGunViolenceData.csv")

aggregated2015gunViolenceData.to\_csv("2015AggregatedGunViolenceData.csv")

aggregated2016gunViolenceData.to\_csv("2016AggregatedGunViolenceData.csv")

aggregated2017gunViolenceData.to\_csv("2017AggregatedGunViolenceData.csv")

#Convert column numIncidents to be an integer

aggregated2014gunViolenceData['mass'] = aggregated2014gunViolenceData['mass'].astype(int)

#Yearly Breakdown Plots

aggregated2014gunViolenceData.plot.box(figsize=(15,12))

plt.title("2014 Gun Violence Overview")

aggregated2015gunViolenceData.plot.box(figsize=(15,12))

plt.title("2015 Gun Violence Overview")

aggregated2016gunViolenceData.plot.box(figsize=(15,12))

plt.title("2016 Gun Violence Overview")

aggregated2017gunViolenceData.plot.box(figsize=(15,12))

plt.title("2017 Gun Violence Overview")

#Histogram Plots

#aggregated2014gunViolenceData.hist(column='mass')

#plt.title("2014 Mass Shooting Breakdown")

#aggregated2015gunViolenceData.hist(column='mass')

#plt.title("2015 Mass Shooting Breakdown")

#aggregated2016gunViolenceData.hist(column='mass')

#plt.title("2016 Mass Shooting Breakdown")

#aggregated2017gunViolenceData.hist(column='mass')

#plt.title("2017 Mass Shooting Breakdown")

#aggregated2014gunViolenceData.hist(column='officer')

#plt.title("2014 Officer Involved Shooting Breakdown")

#aggregated2015gunViolenceData.hist(column='officer')

#plt.title("2015 Officer Involved Shooting Breakdown")

#aggregated2016gunViolenceData.hist(column='officer')

#plt.title("2016 Officer Involved Shooting Breakdown")

#aggregated2017gunViolenceData.hist(column='officer')

#plt.title("2017 Officer Involved Shooting Breakdown")

#aggregated2014gunViolenceData.hist(column='domestic')

#plt.title("2014 Domestic Violence Shooting Breakdown")

#aggregated2015gunViolenceData.hist(column='domestic')

#plt.title("2015 Domestic Violence Shooting Breakdown")

#aggregated2016gunViolenceData.hist(column='domestic')

#plt.title("2016 Domestic Violence Shooting Breakdown")

#aggregated2017gunViolenceData.hist(column='domestic')

#plt.title("2017 Domestic Violence Shooting Breakdown")

#aggregated2014gunViolenceData.hist(column='school')

#plt.title("2014 School Shooting Breakdown")

#aggregated2015gunViolenceData.hist(column='school')

#plt.title("2015 School Shooting Breakdown")

#aggregated2016gunViolenceData.hist(column='school')

#plt.title("2016 School Shooting Breakdown")

#aggregated2017gunViolenceData.hist(column='school')

#plt.title("2017 School Shooting Breakdown")

#Function to map state codes to state name so that we can merge our data sets

def extractStateFromCode(code):

switcher = {

"AL":"Alabama",

"AK":"Alaska",

"AZ":"Arizona",

"AR":"Arkansas",

"CA":"California",

"CO":"Colorado",

"CT":"Connecticut",

"DE":"Deleware",

"FL":"Florida",

"GA":"Georgia",

"HI":"Hawaii",

"ID":"Idaho",

"IL":"Illinois",

"IN":"Indiana",

"IA":"Iowa",

"KS":"Kansas",

"KY":"Kentucky",

"LA":"Loisiana",

"ME":"Maine",

"MD":"Maryland",

"MA":"Massachusetts",

"MI":"Michigan",

"MN":"Minnesota",

"MS":"Mississippi",

"MO":"Missouri",

"MT":"Montana",

"NE":"Nebraska",

"NV":"Nevada",

"NH":"New Hampshire",

"NJ":"New Jersey",

"NM":"New Mexico",

"NY":"New York",

"NC":"North Carolina",

"ND":"North Dakota",

"OH":"Ohio",

"OK":"Oklahoma",

"OR":"Oregon",

"PA":"Pennsylvania",

"RI":"Rhode Island",

"SC":"South Carolina",

"SD":"South Dakota",

"TN":"Tennessee",

"TX":"Texas",

"UT":"Utah",

"VT":"Vermont",

"VA":"Virginia",

"WA":"Washington",

"WV":"West Virginia",

"WI":"Wisconsin",

"WY":"Wyoming"

}

return switcher.get(code, "Unk")

#Function to extract the district fomr our code

def extractDistrictFromCode(code):

if(code == 'AL'):

return "1"

else:

return code

#electionInfoData.info()

#Function to determine if a congressional seat changed political parties

def determineIfSeatChangedParties(partyA, partyB):

if(partyA == 'Open Post-Redistrict'):

return "false"

elif(partyB == 'NA'):

return "false"

elif(partyA == 'Open-Used to be Dem' and partyB == 'D'):

return "false"

elif(partyA == 'Open-Used to be GOP' and partyB == 'R'):

return "false"

elif(partyA == 'GOP' and partyB == 'R'):

return "false"

elif(partyA == 'DEM' and partyB == 'D'):

return "false"

elif(partyA == partyB):

return "false"

else:

return "true"

#Funciton to determine the party if it's not a "swing" party

def calculatePartyIfNotSwing(repParty2014, winParty2014):

if(repParty2014 == "Open-Used to be Dem" or repParty2014 == 'Open-Used to be GOP'):

return winParty2014

else:

return repParty2014

#Define column names for our data frame to map the districts to their political parties

columnNames = ['state','district','party']

#Create our new dataframe

congressionalDistrictPartyData = pandas.DataFrame(columns = columnNames)

#Loop through our election info data

for index, row in electionInfoData.iterrows():

#Extract our state and district

state = extractStateFromCode(row['District'][0:2])

district = extractDistrictFromCode(row['District'][2:])

#Determine if it's a "Swing" district

swing = determineIfSeatChangedParties(row['rep\_party\_2014'], row['winning\_party\_2014'])

if(swing == 'false'):

swing = determineIfSeatChangedParties(row['rep\_party\_2016'], row['winning\_party\_2016'])

if(swing == 'false'):

swing = determineIfSeatChangedParties(row['rep\_party\_2018'], row['winning\_party\_2018'])

#Calculate the political party affiliated with this district

party = ""

if(swing == "true"):

party = "Swing"

else:

party = calculatePartyIfNotSwing(row['rep\_party\_2014'], row['winning\_party\_2014'])

#Create an array of our data

electionData = [[state, district, party]]

#Create a temporary dataframe

tempElectionDf = pandas.DataFrame(electionData, columns=['state','district','party'])

#Append it to our dataframe

congressionalDistrictPartyData = congressionalDistrictPartyData.append(tempElectionDf, True)

#Write our our congressional district party data report

congressionalDistrictPartyData.to\_csv("congressionalDistrictPartyData.csv")

#Convet numIncidents to numeric

aggregated2014gunViolenceData['numIncidents'] = pandas.to\_numeric(aggregated2014gunViolenceData['numIncidents'])

#Calculate the total number of incidents by state for 2014

total2014IncidentsByState = aggregated2014gunViolenceData.groupby('state').sum()

total2014IncidentsByState = total2014IncidentsByState.reset\_index()

total2014IncidentsByState = total2014IncidentsByState[['state','numIncidents']]

total2014IncidentsByState.columns = ['state','2014 Incidents']

#Calculate the total number of incidents by state for 2015

aggregated2015gunViolenceData['numIncidents'] = pandas.to\_numeric(aggregated2015gunViolenceData['numIncidents'])

total2015IncidentsByState = aggregated2015gunViolenceData.groupby('state').sum()

total2015IncidentsByState = total2015IncidentsByState.reset\_index()

total2015IncidentsByState = total2015IncidentsByState[['state','numIncidents']]

total2015IncidentsByState.columns = ['state','2015 Incidents']

#Calculate the total number of incidents by state for 2016

aggregated2016gunViolenceData['numIncidents'] = pandas.to\_numeric(aggregated2016gunViolenceData['numIncidents'])

total2016IncidentsByState = aggregated2016gunViolenceData.groupby('state').sum()

total2016IncidentsByState = total2016IncidentsByState.reset\_index()

total2016IncidentsByState = total2016IncidentsByState[['state','numIncidents']]

total2016IncidentsByState.columns = ['state','2016 Incidents']

#Calculate the total number of incidents by state for 2017

aggregated2017gunViolenceData['numIncidents'] = pandas.to\_numeric(aggregated2017gunViolenceData['numIncidents'])

total2017IncidentsByState = aggregated2017gunViolenceData.groupby('state').sum()

total2017IncidentsByState = total2017IncidentsByState.reset\_index()

total2017IncidentsByState = total2017IncidentsByState[['state','numIncidents']]

total2017IncidentsByState.columns = ['state','2017 Incidents']

#Create four dataframes to hold the state firearms law data for each year from 2014-2017

stateLawData2014 = stateFirearmLawData[stateFirearmLawData['year'] == 2014]

stateLawData2015 = stateFirearmLawData[stateFirearmLawData['year'] == 2015]

stateLawData2016 = stateFirearmLawData[stateFirearmLawData['year'] == 2016]

stateLawData2017 = stateFirearmLawData[stateFirearmLawData['year'] == 2017]

#Rename column names to be more descriptive

stateLawData2014 = stateLawData2014[['state','lawtotal']]

stateLawData2014.columns = ['state','lawtotal2014']

stateLawData2015 = stateLawData2015[['state','lawtotal']]

stateLawData2015.columns = ['state','lawtotal2015']

stateLawData2016 = stateLawData2016[['state','lawtotal']]

stateLawData2016.columns = ['state','lawtotal2016']

stateLawData2017 = stateLawData2017[['state','lawtotal']]

stateLawData2017.columns = ['state','lawtotal2017']

#Merge our incident dataframes and statelaw dataframes

mergedIncidentDataByState = pandas.merge(total2014IncidentsByState, stateLawData2014, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, total2015IncidentsByState, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, stateLawData2015, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, total2016IncidentsByState, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, stateLawData2016, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, total2017IncidentsByState, on='state',how='inner')

mergedIncidentDataByState = pandas.merge(mergedIncidentDataByState, stateLawData2017, on='state',how='inner')

mergedIncidentWithoutLawInformation = mergedIncidentDataByState[['state','2014 Incidents','2015 Incidents','2016 Incidents', '2017 Incidents']]

#Write out our mergedincidents by state report

mergedIncidentDataByState.to\_csv("violenceCountsByStateAndYear.csv")

mergedIncidentWithoutLawInformation.to\_csv("violenceCountsByStateAndYearWithoutLawInfo.csv")

#Convert district column to float type

congressionalDistrictPartyData['district'] = congressionalDistrictPartyData['district'].astype(float)

#Merge our gun violence data and congressional party data for 2014-2017

mergedIncidentsByCongressionalDistricts2014 = pandas.merge(aggregated2014gunViolenceData, congressionalDistrictPartyData, on=['state','district'], how='inner')

mergedIncidentsByCongressionalDistricts2015 = pandas.merge(aggregated2015gunViolenceData, congressionalDistrictPartyData, on=['state','district'], how='inner')

mergedIncidentsByCongressionalDistricts2016 = pandas.merge(aggregated2016gunViolenceData, congressionalDistrictPartyData, on=['state','district'], how='inner')

mergedIncidentsByCongressionalDistricts2017 = pandas.merge(aggregated2017gunViolenceData, congressionalDistrictPartyData, on=['state','district'], how='inner')

#Write out our reports of merged incidents by congressional districts

mergedIncidentsByCongressionalDistricts2014.to\_csv("violenceIncidentsByCongressionalDistrictType2014.csv")

mergedIncidentsByCongressionalDistricts2015.to\_csv("violenceIncidentsByCongressionalDistrictType2015.csv")

mergedIncidentsByCongressionalDistricts2016.to\_csv("violenceIncidentsByCongressionalDistrictType2016.csv")

mergedIncidentsByCongressionalDistricts2017.to\_csv("violenceIncidentsByCongressionalDistrictType2017.csv")

groupedAndMergedIncidentsByCongressionalDistricts2014 = mergedIncidentsByCongressionalDistricts2014.groupby('party')

#Calculate the average number of firearm laws between 2014-2017 per state

averageStateFirearmLawData = stateFirearmLawData.groupby('state').mean()

averageStateFirearmLawData = averageStateFirearmLawData.reset\_index()

averageStateFirearmLawData = averageStateFirearmLawData[['state','lawtotal']]

#Count the number of rep, dem and swing seats per state

republicanRepsInState=mergedIncidentsByCongressionalDistricts2014.groupby('state')['party'].apply(lambda x: (x=='R').sum()).reset\_index(name='count')

democraticRepsInState=mergedIncidentsByCongressionalDistricts2014.groupby('state')['party'].apply(lambda x: (x=='D').sum()).reset\_index(name='count')

swingRepsInState=mergedIncidentsByCongressionalDistricts2014.groupby('state')['party'].apply(lambda x: (x=='Swing').sum()).reset\_index(name='count')

#Merge our counts of dems, reps and swings with the average amount of firearm laws between 2014-2017

mergedFirearmLawsWithCongressionInformation = pandas.merge(averageStateFirearmLawData, republicanRepsInState, on=['state'], how='outer')

mergedFirearmLawsWithCongressionInformation = pandas.merge(mergedFirearmLawsWithCongressionInformation, democraticRepsInState, on=['state'], how='outer')

mergedFirearmLawsWithCongressionInformation = pandas.merge(mergedFirearmLawsWithCongressionInformation, swingRepsInState, on=['state'], how='outer')

mergedFirearmLawsWithCongressionInformation.columns = ['state','lawtotal','rep count','dem count','swing count']

#Print our report of firearm law relationship with number of congressional districts by party

mergedFirearmLawsWithCongressionInformation.to\_csv("firearmsWithCongressionalInformation.csv")

#Correlation of officer involved shootings

aggregated2014Correlation = aggregated2014gunViolenceData.corr(method='pearson')

aggregated2014Correlation.to\_csv('aggregated2014CorrelationInfo.csv')

aggregated2015Correlation = aggregated2015gunViolenceData.corr(method='pearson')

aggregated2015Correlation.to\_csv('aggregated2015CorrelationInfo.csv')

aggregated2016Correlation = aggregated2016gunViolenceData.corr(method='pearson')

aggregated2016Correlation.to\_csv('aggregated2016CorrelationInfo.csv')

aggregated2017Correlation = aggregated2017gunViolenceData.corr(method='pearson')

aggregated2017Correlation.to\_csv('aggregated2017CorrelationInfo.csv')

OIX2014 = aggregated2017gunViolenceData['officer']

OIY2014 = aggregated2017gunViolenceData['numIncidents']

OI2014Reg = linear\_model.LinearRegression()

OI2014Model = sm.OLS(OIX2014,OIY2014).fit()

#print(OI2014Model.summary())

mergedIncidentDataByStateCorrelation = mergedIncidentDataByState.corr(method='pearson')

mergedIncidentDataByStateCorrelation.to\_csv('mergedIncidentDataByStateCorrelation.csv')

#TODO - LINEAR REGRESSION

#print(mergedIncidentDataByState)

X = mergedIncidentDataByState['lawtotal2017']

Y = mergedIncidentDataByState['2017 Incidents']

regr = linear\_model.LinearRegression()

model = sm.OLS(X,Y).fit()

print(model.summary())