Understanding Gun Violence

Final Project for W207

David He
Rory Liu
Kathryn Plath
Andrea Sy

Background and Approach

The subject of gun violence has captured national attention for the past few years. In order to dig deeper and get a better understanding of this issue, we decided to conduct an independent analysis with the dataset provided by Kaggle and Gun Violence Archive.

Our final project aims to explore the relationship between certain features (related to gun violence, demographics, socioeconomic factors, NRA ratings, etc), number of incidents, number of casualties, and likelihood of mass events.

Our Approach:

Data cleaning and processing

Exploratory analysis

Modeling: Predicting incident volume

Modeling: Predicting casualties

Modeling: Predicting mass events

Data Cleaning and Processing

Cleaning

- **parsing dates and removing 2013 & 2018**
  - Initial analysis shows incomplete data from 2013 and 2018, so we focused on years where we have full data
- **parsing incident types and forming binary incident categories**
  - Original data listed all qualifying characteristics in a single variable
  - We created binary incident types (e.g. accidents) and removed non-violent types (e.g. gun buy-back)
- **parsing participants characteristics to find suspect and victim demographics**
  - Original data listed participants demographics info in the same variable
  - We created binary demographics variables for suspect and victim types

Merging in external data sources

- **NRA ratings**
  - Average NRA rating for senators by state
  - NRA rating for house representative by congressional district
- **State political leaning**
  - Binary flag for blue states based on 2016 election
- **Congressional district population and demographics indicators**
  - Total population, population density, education level, percent male population, per capita income by congressional district

Exploratory Data Analysis

Chart 1 - Gun Violence over Time

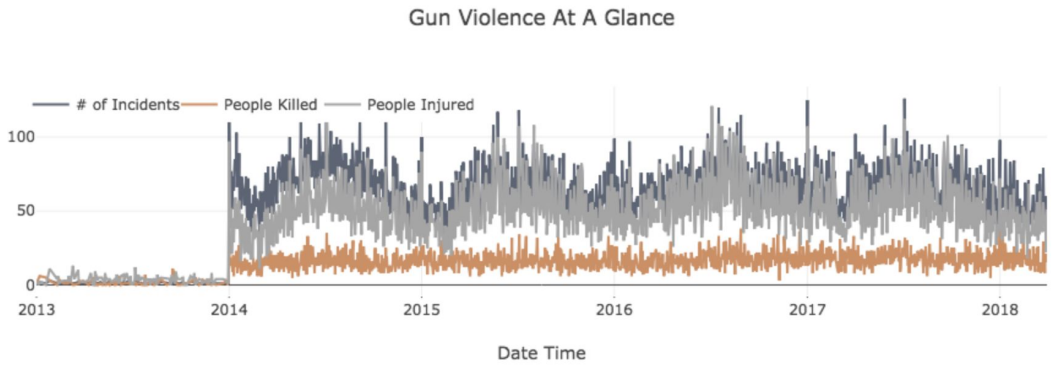
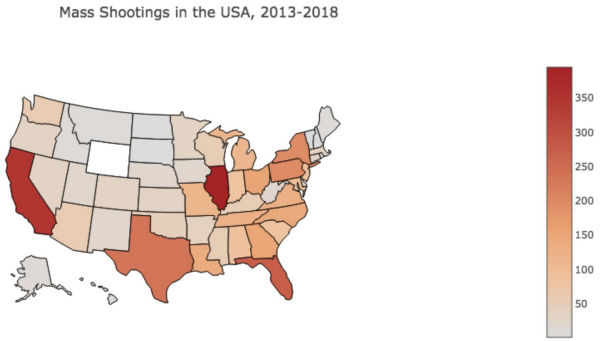
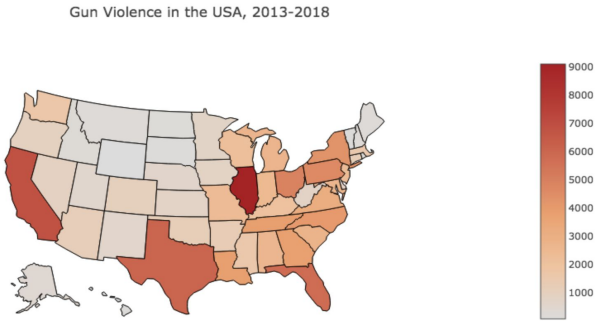


Chart 2 - Mass Shootings & Gun Violence by State

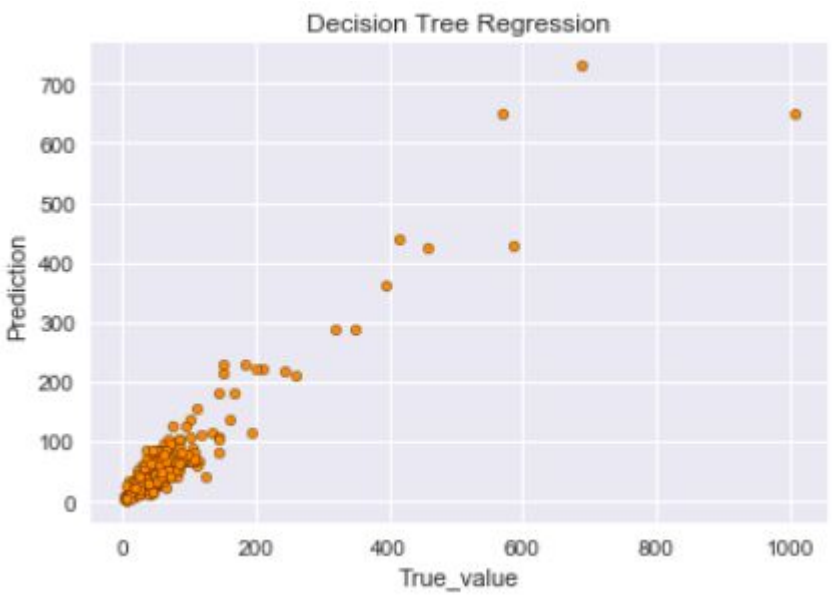


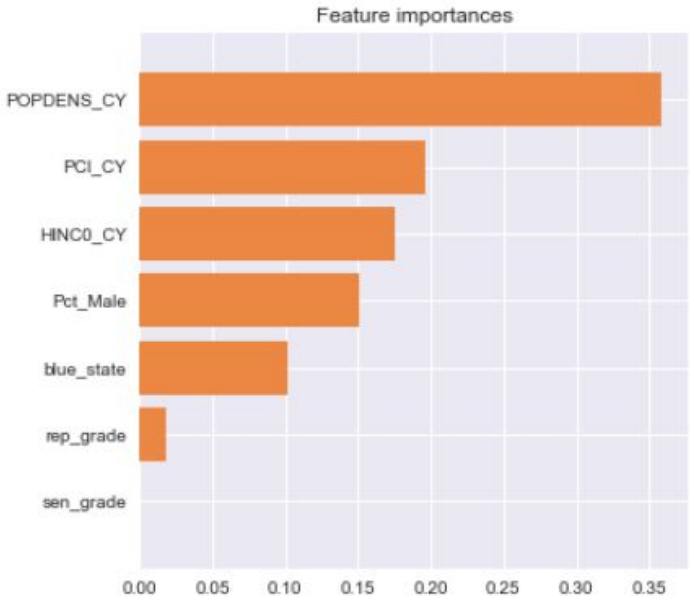


Modeling & Findings

Predicting annual incident volume based on district characteristics

- **Input**: Population density, per-capita income, # of household w/ income <15k, percent of male population, blue state, NRA senator grade, NRA rep grade
- **Outcome variable**: annual incident volume
- **Algorithm used**: Linear regression, Linear regression with regulation, Decision tree regressor, random forest regressor.
- **Winner**: Decision tree regressor: score = 0.9 (no limit on max_depth; score = 0.7 w/ depth of 5)





Modeling & Findings

Predicting casualties/deaths based on incident types and suspect demographics

- **Input**: Incident Types, Suspect Types, NRA_sen_grade, NRA_rep_grade
- **Outcome variable**: n_killed
- **Algorithm used**: Linear Regression, Random Forest
- **Winner**: No clear winner this yielded very low accuracy results of about .16 for a Linear Regression (OLS) and .25 for a Random Forest.

Results:

OLS Regression Results

Dep. Variable:	n_killed	R-squared:	0.157
Model:	OLS	Adj. R-squared:	0.157
Method:	Least Squares	F-statistic:	744.0
Date:	Sun, 05 Aug 2018	Prob (F-statistic):	0.00
Time:	19:59:55	Log-Likelihood:	-75473.
No. Observations:	103560	AIC:	1.510e+05
Df Residuals:	103533	BIC:	1.513e+05
Df Model:	26		
Covariance Type:	nonrobust		

```
Linear Regression MSE of test data predictions: 0.251655461526
Random Forest MSE of test data predictions: 0.250785454565
```

Modeling & Findings

Predicting mass events

- **Input**: 28 Features on categories of incidents
- **Outcome variable**: Binary variable for if a mass shooting occurred
- **Algorithm used**: Logistic Regression
- **Winner**: 96% accurate

Results

Factors that increase the likelihood that a mass shooting occurs include if children are involved, domestic violence, gangs, fights, and hate crimes.

Factors that do not affect the likelihood of a mass shooting include the NRA ratings of the elected officials, and how many incidents per capita occur in that state.

	Features	Coefficients	Odds Ratio
2	Child	1.673477	5.330671
11	Murder	1.300376	3.670676
17	Social_gathering	1.278586	3.591559
9	Hate_crime	1.080098	2.944969
7	Gang	0.695743	2.005198
6	Fight	0.675669	1.965348
4	Domestic_Violence	0.667206	1.948784
15	School	0.585573	1.796020
3	Defense	0.348482	1.416915
18	terrorism	0.346885	1.414655
13	Political_Violence	0.293256	1.340786
21	Teen_Male_SubjectSuspect	0.127471	1.135951
24	Child_Female_SubjectSuspect	0.117975	1.125216
8	Ghost_gun	0.110792	1.117162
10	Hostage	0.100497	1.105720
0	n_guns_involved	0.066875	1.069162
5	Drug_Alcohol	0.023417	1.023693
25	sen_grade	0.008743	1.008782
27	inc_per_capita	-0.006372	0.993648
19	Adult_Male_SubjectSuspect	-0.058525	0.943155
26	rep_grade	-0.077255	0.925654
12	Police	-0.275975	0.758832
16	Sex_crime	-0.292106	0.746690
22	Teen_Female_SubjectSuspect	-0.305971	0.736408
14	Robbery	-0.510188	0.600383
20	Adult_Female_SubjectSuspect	-0.514867	0.597580
23	Child_Male_SubjectSuspect	-0.602948	0.547196
1	Accidents	-2.081078	0.124796