Hate Crimes in the United States 2013 - 2018

By: Brittany Fowle and Julia Chong

What is a hate crime?

- Criminal offense such as murder, arson, or vandalism with an added element of **bias**.
- Bias includes being motivated by *race, religion, disability, sexual orientation, ethnicity, gender, or gender identity.*
- Hate itself is not a crime - freedom of speech is protected.

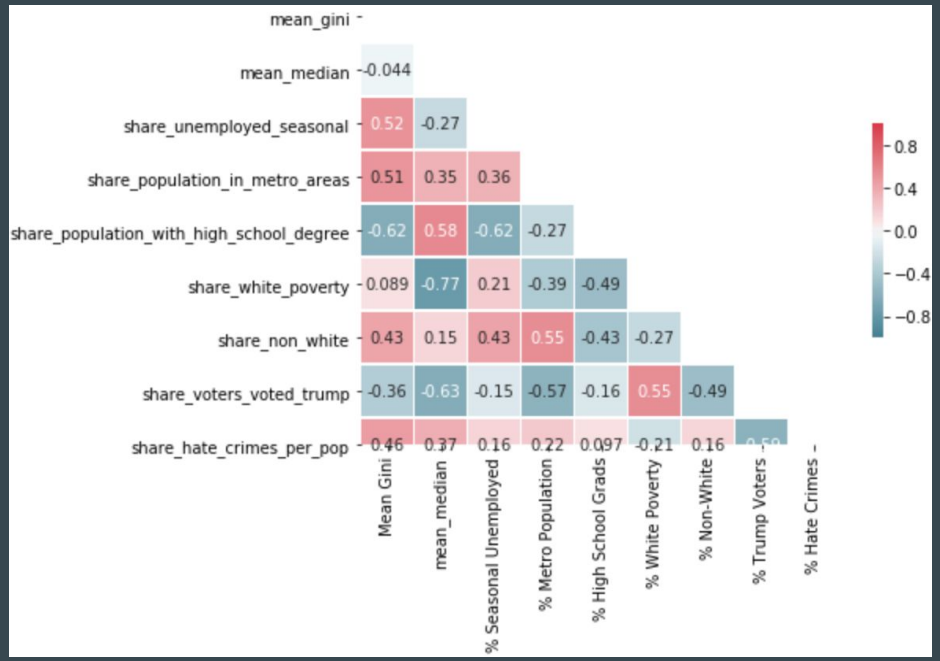
- Defining a Hate Crime, '*https://www.fbi.gov/investigate/civil-rights/hate-crimes*'

Datasets

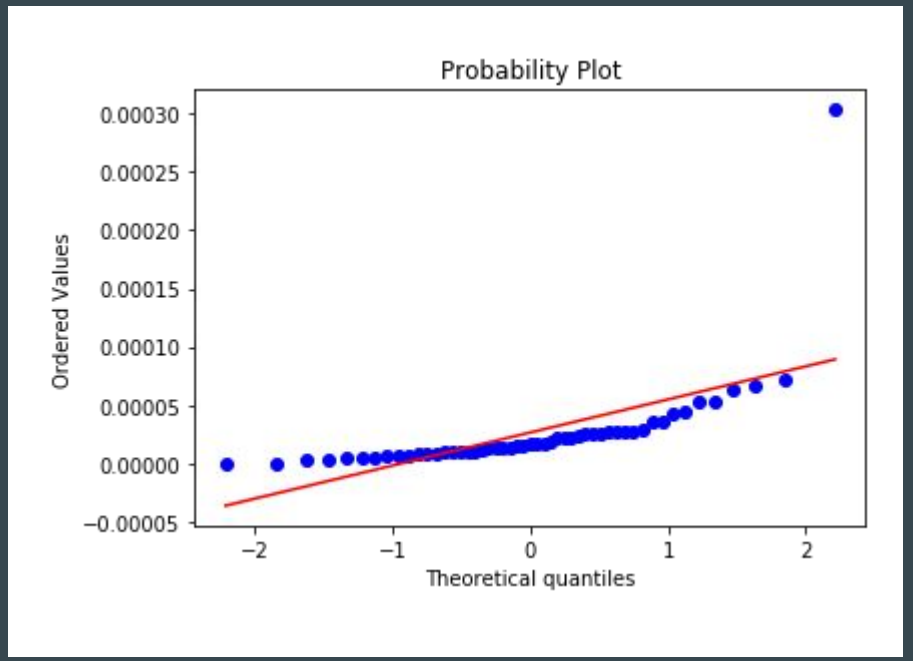
Each of our datasets included all 50 states, for each year between 2013 - 2018.

- Reported hate crimes as a proportion of the total population
- Median household income
- The Gini coefficient - an economic measure of income inequality
- Percentage of the population that voted for Trump in 2016 (static variable, data only from 2016 election)

Initial Testing



Pearson coefficient heat map



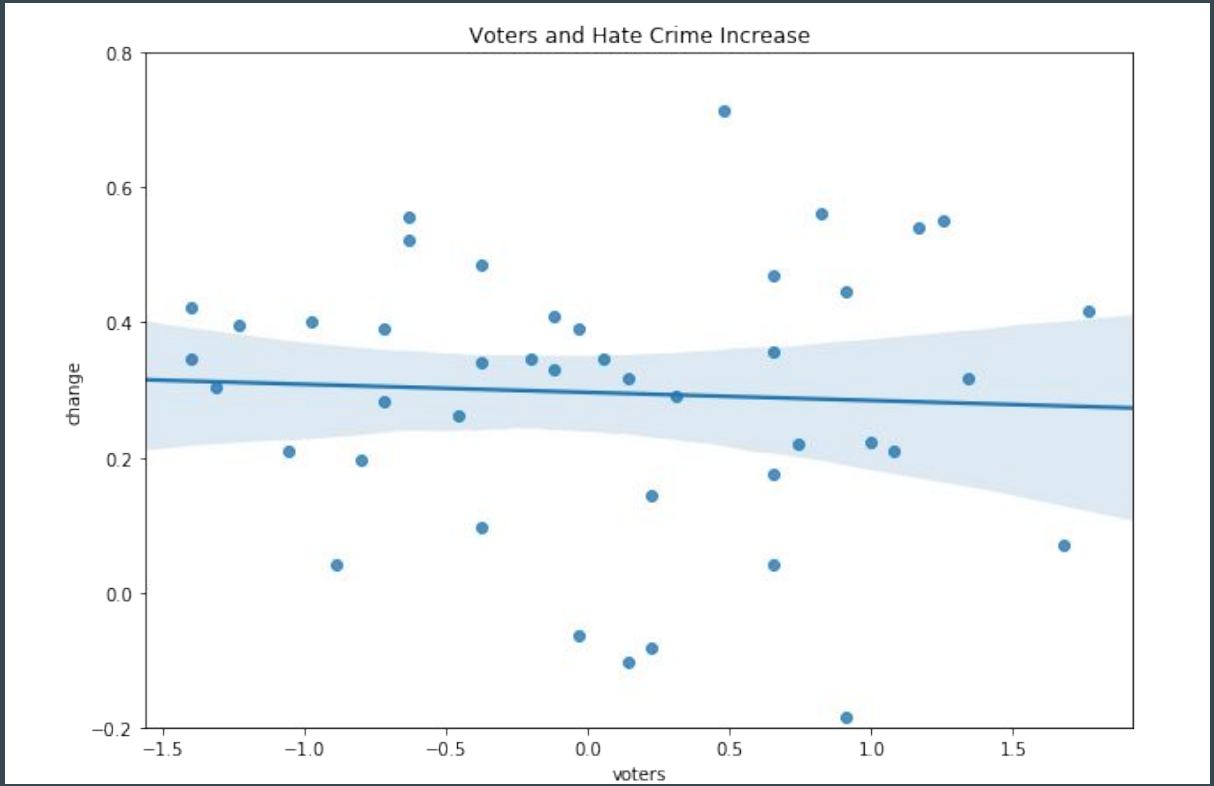
QQ Plot for Hate Crimes

Questions for Exploratory Data Analysis

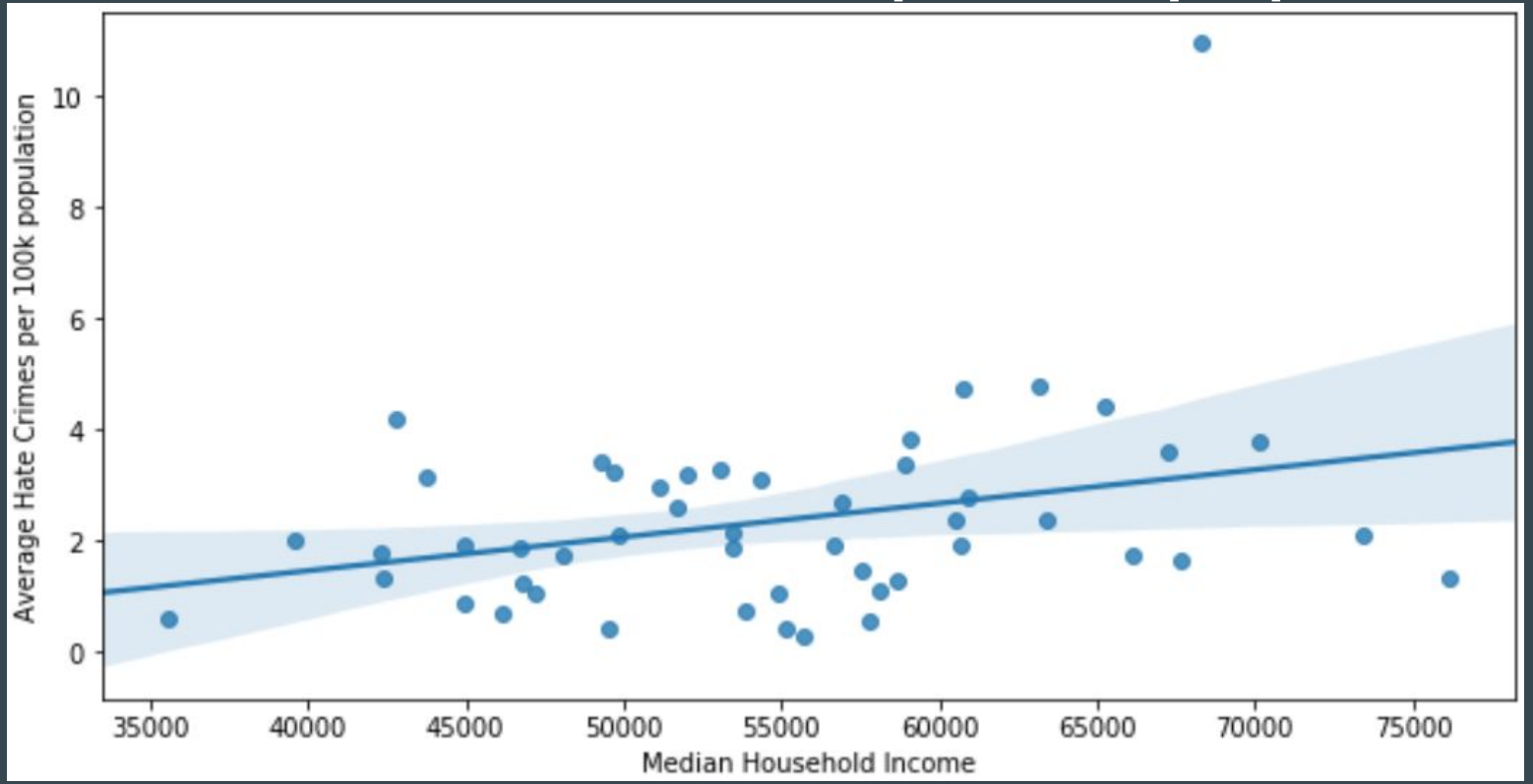
1) Did hate crimes increase after Trump was elected in states with higher than average rates of Trump supporters?
    a) No correlation
2) Does a wider gap between the rich and poor instigate more hate crime?
    a) Weak correlation
3) Is household income correlated with hate crime?
    a) Weak correlation

Hate Crime Change and Percent of Trump Voters

- 2016 to 2017
- Percent change in hate crimes
- Percent of population who voted for Trump
- Both variables normalized using Z Score
- R^2 = .003

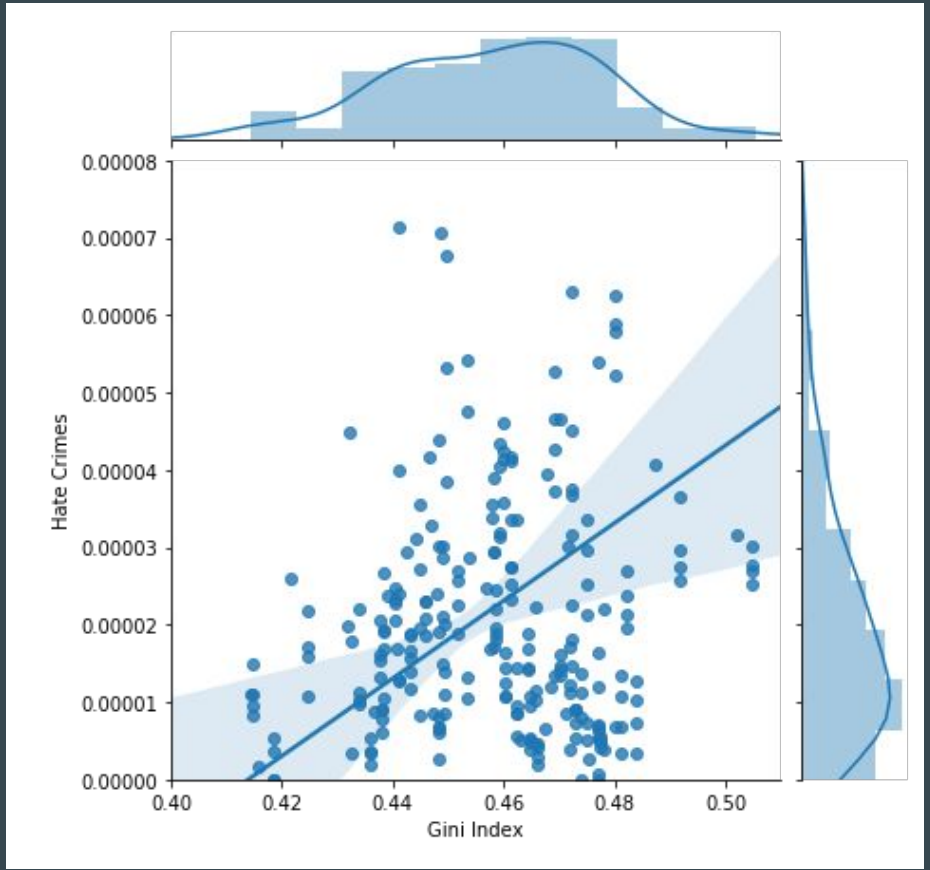


Household Income vs Hate Crimes per 100k people



Hate Crime and Income Inequality

- R^2 = .159
- Gini Coefficient by state
- Hate Crime Population Proportion by State
- 2013 - 2017



OLS + VIF

-R2 = 0.52

-Prob (F-statistic) = 2.47e-5

-Significant variables:

- Mean Gini (p-value = 0.002)

- Share Voters Voted Trump (p-value = 0.048)

-Variance inflation factors < 3.4

Retest with significant variables

-R2 = 0.406

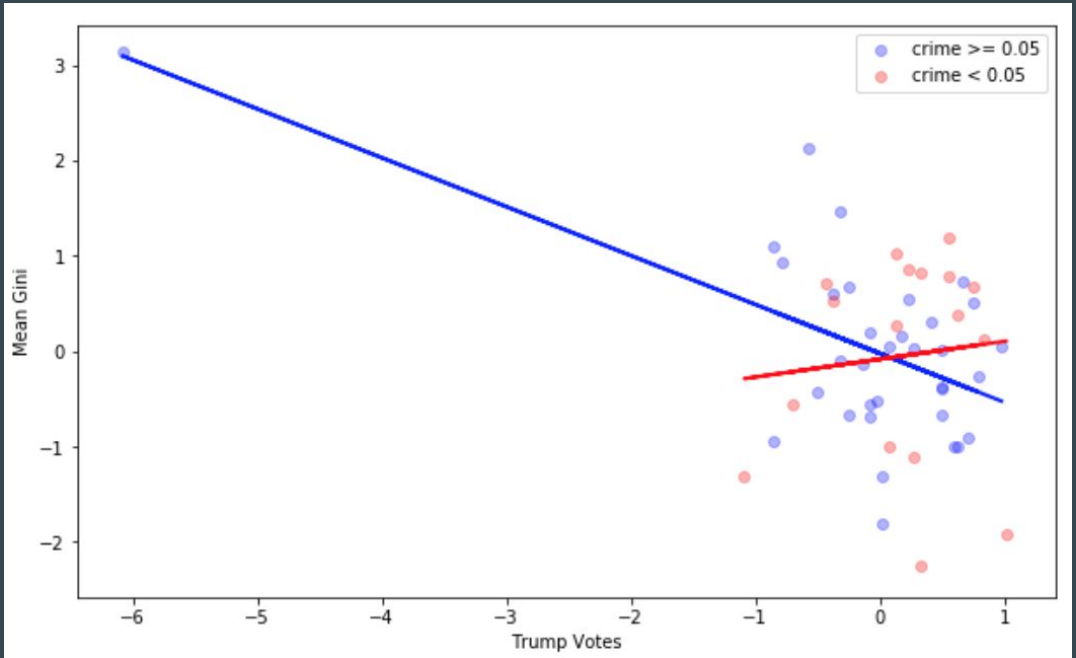
- Prob (F-statistic) = 4.78e-06

Model:

Y = 2.354e-05 + 1.547e-05(X1) - 1.804e-10(X2) + 3.806 e-06(X3) + 8.086e-05(X4) + 4.212e-05(X5) - 3.868e-06 (X6) - 9.679e-06(X7)

X1 = Mean Gini Coefficient
X2 = Mean Median Income
X3 = Percent Seasonal Unemployed
X4 = Percent With High School Degree
X5 = Percent White Poverty
X6 = Percent Non-White
X7 = Percent Trump Voters

Interaction



- Test for interaction between Trump Votes and Gini Coefficient
- Take logs of raw data to standardize
- Significant interaction between Trump Votes and Gini Coefficient

Interaction term = -0.7441326352870261

Model Training: Ridge/Lasso

```
Train Error Ridge Model share_hate_crimes_per_pop    2.880014e-09
dtype: float64
Test Error Ridge Model share_hate_crimes_per_pop    1.325702e-09
dtype: float64


Train Error Lasso Model share_hate_crimes_per_pop    3.031821e-08
dtype: float64
Test Error Lasso Model share_hate_crimes_per_pop    3.952532e-09
dtype: float64
```

```
Train Error Unpenalized Linear Model share_hate_crimes_per_pop
3.380282e-38
dtype: float64
Test Error Unpenalized Linear Model share_hate_crimes_per_pop
1.237645e-38
dtype: float64
```

Final Model

Y = 2.354e-05 -1.1142205e-10(X1X7) - 1.804e-10(X2) + 3.806 e-06(X3) + 8.086e-05(X4) + 4.212e-05(X5) - 3.868e-06 (X6) - 9.679e-06(X7)

X1 = Mean Gini Coefficient X2 = Mean Median Income X3 = Percent Seasonal Unemployed X4 = Percent With High School Degree X5 = Percent White Poverty X6 = Percent Non-White X7 = Percent Trump Voters

Improvements

- Polynomial regression to test for a better fit to reduce bias
-Remove outliers
- Treat underlying dataset differently (use time series data as instances instead of aggregating to increase sample size)
-Explore other demographic variables that might have stronger correlation