***Determinants to Hate Crimes***

**Marlene Aviles**

**DSC680: Applied Data Science**

<https://maviles2.github.io/Hate_Crimes/>.

**Which Domain?**

GitHub will be the main domain utilized for this project, and the dataset on the subject of hate crime will be analyzed. The dataset contains factors from several websites and organizations such as the FBI, Southern Poverty Law Center, United States Census, and the Kaiser Family Foundation. This data will look at the determinants of hate crimes in the United States.

1. Dataset was retrieved from GitHub on hate crimes. Retrieved from <https://github.com/fivethirtyeight/data/blob/master/hate-crimes/README.md>

2. Southern Poverty Law Center website data on hate crimes. Retrieved from <https://www.splcenter.org/20161129/ten-days-after-harassment-and-intimidation-aftermath-election>

3. FBI Hate Crime data. Retrieved from <https://ucr.fbi.gov/hate-crime>

4. Kaiser Family Foundation data on Median Annual Income in the U.S. Retrieved from <https://www.kff.org/other/state-indicator/median-annual-income/>

5. FiveThirtyEight published story on this hate crime dataset. Retrieved from <https://fivethirtyeight.com/features/higher-rates-of-hate-crimes-are-tied-to-income-inequality/>

6. United States Census data on educational attainments. Retrieved from <https://www.census.gov/prod/2012pubs/p20-566.pdf>

7. United States Election Projects data on 2016 election turnout. Retrieved from <http://www.electproject.org/2016g>a

8. Kaiser Family Foundation data on race and ethnicity distribution in the U.S. Retrieved from <https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>

9. Kaiser Family Foundation data on unemployment rate (seasonally adjusted) in the U.S. Retrieved from <https://www.kff.org/other/state-indicator/unemployment-rate/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>

10. BBC news article on the rise of hate crimes. Retrieved from <https://www.bbc.com/news/world-us-canada-41975573>

**Which Data?**

<https://github.com/fivethirtyeight/data/blob/master/hate-crimes/README.md>

FiveThirtyEight is a website that is focused on statistical analysis to tell compelling stories about politics, economics, science, and more through the use of data. FiveThirtyEight has created the following dataset to analyze hate crimes in the United States after the 2016 election. The dataset was created with the use of different reference sources such as the Southern Poverty Law Center, the Kaiser Family Foundation, the FBI, and the United States Census. Each source listed managed to provide FiveThirtyEight data regarding several socio-economic factors that affected hate crimes in the United States, such as the unemployment rate, median household income, distribution on race, and more.

The dataset contains a total of 12 variables listed below and 52 rows that represent the states in the United States.

*Variable overview:*

State: State name

Median\_household\_income: Median household income, 2016

Share\_unemployed\_seasonal: Share of the population that is unemployed, 2016

Share\_population\_in\_metro\_areas: Share of the population that lives in metropolitan areas, 2015

Share\_population\_with\_high\_school\_degree: Share of adults 25 and older with a high-school degree, 2009

Share\_non\_citizen: Share of the population that are not U.S. citizens, 2015

Share\_white\_poverty: Share of white residents who are living in poverty, 2015

Gini\_index: Gini index, 2015

Share\_non\_white: Share of the population that is not white, 2015

Share\_voters\_voted\_trump: Share of 2016 U.S. presidential voters who voted for Donald Trump

Hate\_crimes\_per\_100k\_splc: Hate crimes per 100,000 population by the Southern Poverty Law Center (Nov 9-18, 2016)

Avg\_hatecrimes\_per\_100k\_fbi: Average annual hate crimes per 100,000 population, FBI 2010- 2015

**Research Questions? Benefits? Why analyze these data?**

A hate crime is described as a criminal offense against a person or property which is motivated in whole or part by an offender’s bias against race, religion, disability, sexual orientation, ethnicity, gender, or gender identity (FBI, n.d.). These types of crimes, usually involve violence and are considered a top priority for organizations such as the Federal Bureau of Investigation (FBI), this priority is due to the devastating impact that they cause on individuals, families, and communities (FBI, n.d.). This following project will analyze several factors that impacted the increase of hate crimes after the 2016 election. The results and analysis of this particular project will assist as a way of identifying the determinants of hate crimes in the United States during this period. Being able to identify which factors contribute to hate crimes will help in reducing future hate crimes and in maintaining safer communities. In 2016, the presidential election took place and during this particular period, an increase in hate crimes was noted. This analysis will help in determining which factors affected the increase of these heinous crimes after this particular election and why. The variables in the dataset will be analyzed to find significant trends or patterns in the data via exploratory data analysis, correlations, and multivariate linear regression. The following research questions will be analyzed for this particular project:

* What is the distribution of hate crimes in the United States?
* Does your political affiliation (Democrat or Republican) influence your involvement in a hate crime?
* Is your economic status a predictor of committing a hate crime or being a victim of a hate crime?
* What factors are the best predictors to hate crimes?
* Do hate crimes vary by geographical region?

**What Method?**

The methods utilized for this particular project will include exploratory data analysis (EDA), regression analysis, and predictive analytics. EDA will assist in identifying the most important variables in the dataset. EDA will aid in executing visuals to look at distributions, outliers, and meaningful trends in the dataset. By utilizing exploratory data analysis methods, the most appropriate models and algorithms will be selected. The use of correlations and regression analysis will be applied to identify the best predicting variables of hate crimes. A predictive model will be applied based on the results of EDA and regression analysis.

**Potential Issues?**

This project has limited resources given that hate crimes are under-reported and not accurately documented. The federal government tracks hate crimes systematically voluntarily only. This means that not all hate crimes may have been included in the data, causing a misrepresentation of the actual data on these crimes. The trajectory of this project will be tracked via a weekly schedule to make sure this project does not go off schedule. The data has been collected and the dataset has been created, given this, the project should not be off schedule.

**Concluding Remarks**

The dataset created by FiveThirtyEight based on hate crimes will serve as a beneficial analysis to help predict which variables are highly correlated with hate crimes. Hate crimes are usually violent crimes that affect individuals, families, and communities all around the United States. Significant events that occur during the trajectory of this data must be noted, such as the 2016 election, given that they may have a significant impact on the data recorded. Hate crimes and being able to identify the variables that cause them to increase will only help in ensuring the safety and well-being of all of the individuals residing in the United States.

**Sources Cited:**

FBI. ( n.d.) Hate crimes. Retrieved from <https://www.fbi.gov/investigate/civil-rights/hate-crimes>