# Hate Crimes in the Unites States (1991-2018)

**Data sourcing and profiling**

1. Project Overview
   1. **Motivation**:
      1. America is a country with a long history of prejudice that continues to this day. Prejudice takes place in many forms within America and is based on many components of individuality like race, gender, or sexuality. Prejudice can manifest in many ways, within America, prejudice manifests in one of its most violent and dangerous forms: hate crime.
   2. **Objective**:
      1. Determine ways in which hate crimes have behaved within the United States over the past few decades.
   3. **Scope**:
      1. The dataset come from the FBI’s Hate Crime Statistics Program, Uniform Crime Reporting (UCR) Program and includes data for all of America. Hate crime can be described as criminal offenses that are motivated, in part, or wholly by an offenders bias against another individual’s race, gender, gender identity, religion, disability, sexual orientation, or ethnicity. Hate crimes include crimes that are not only committed against one person but also include crimes against persons, property, or society.
2. Research Hypothesis
   1. TBD
3. Data Source and Overview
   1. [**United States Hate Crimes (1991-2018) | Kaggle**](https://www.kaggle.com/datasets/louissebye/united-states-hate-crimes-19912017?resource=download)
      1. Data Source: Data is publicly available and is open-source data. The dataset is stored on Kaggle and was made available by Louisse Bye. As stated in the metadata section, the data is uploaded under the “U.S. Government Works” license and therefore falls within the public domain and can be used without obtaining explicit permission or paying copyright fee.
      2. Data Collection: The data is administrative data and was retrieved from the [FBI: Crime Data Explorer](https://cde.ucr.cjis.gov/). This data is collected annually.
      3. Data Contents: The data contains the date, state, location of incident, offender race, victim count, type of offense, motivation, etc. The data encompassed the period of 1991-2018. The data was collected by law enforcement agencies that chose to participate with the Hate Crimes Statistic Program.
      4. Data Structure and Organization: Data is contained within a CSV files with over 200,000 rows of data. Dataset contains both qualitative data and quantitative data.
      5. Data Relevance: Though this data is old, it is still relevant to the significance of the project and can provide insight and answers to the project questions.
4. Data Limitations and Ethical considerations
   1. The dataset only contains data from the years 1991-2018, meaning that it is outdated by approximately 6 years. Additionally, the data contains a significant amount of null values, while these values will be looked at during analysis, this may be a potential limiting factor to the quality of the results of the analysis.
   2. Data within this analysis is also rather subjective, bias may have entered into the data through the individuals who were collecting and submitting it. While the data is meant to be informative and the likelihood of bias is rare, motivation of an individual is subjective and therefore a difficult factor for an those collecting data to determine. Additionally, bias alone is not the definitive factor in determining if a crime is considered a hate crime. A law enforcement investigation must reveal sufficient evidence to lead a reasonable and prudent person to conclude that the individuals actions were indeed motivated by their bias, and only then should the agency report the incident as a hate crime.
5. Data Cleaning and Consistency Checks
   1. Original Data: 201,403 rows and 28 columns. New data: 199,797 rows and 21 columns.
   2. Adjusted data types
   3. Checked for missing data
   4. Checked for duplicate data
   5. Check for mixed data types
6. Column Details

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Column Description | Data Type | Time Variant (Y/N) |
| INCIDENT\_ID | IDENTIFICATION NUMBER ASSIGNED TO AN INCIDENT | QUALITATIVE, NOMINAL | NO |
| DATA\_YEAR | YEAR IN WHICH THE INCIDENT occurred | QUALITATIVE, ORDINAL | NO |
| ORI | 9 digit identifier for the agency requesting a criminal record check | qualitative, nominal | no |
| PUB\_AGENCY\_NAME | Name of the public agency | qualitative, nominal | no |
| AGENCY\_TYPE\_NAME | Assigned agency type | qualitative, nominal | no |
| STATE\_ABBR | Abbreviation of the state | qualitative, nominal | no |
| STATE\_NAME | name of the state | qualitative, nominal | no |
| DIVISION\_NAME | name of the devision | qualitative, nominal | no |
| REGION\_NAME | region of the usa | qualitative, nominal | no |
| POPULATION\_GROUP\_CODE | group code assigned to the population | Qualitative, Ordinal | no |
| POPULATION\_GROUP\_DESC | description of the population group code | qualitative, ordinal | no |
| INCIDENT\_DATE | the date of the incident in M/D/Y format | qualitative, ordinal | no |
| TOTAL\_OFFENDER\_COUNT | quantity of offenders involved in the incident | quantitative, discrete | yes |
| OFFENDER\_RACE | race of the offending party | qualitiative, nominal | no |
| VICTIM\_COUNT | quantity of victims involved in the incident | quantitative, discrete | yes |
| OFFENSE\_NAME | type of the offense commited | qualitative, nominal | no |
| LOCATION\_NAME | location where incident occurred | qualitative, nominal | no |
| BIAS\_DESC | description of bias | qualitative, nominal | no |
| VICTIM\_TYPE | type of victim, individual, organization, etc. | qualitative, nominal | no |
| MULTIPLE\_OFFENSE | s or m assigned to the incident if singal (s) or multiple (m) offenses occurred | qualitative, ordinal | no |
| MULTIPLE\_BIAS | s or m assigned to the incident if singal (s) or multiple (m) biases occurred | qualitative, ordinal | no |

1. Questions to exploore
   * 1. **What type of hate crimes are more prevalent throughout 1991-2018.** 
        1. Additionally, consider comparison over 3 decades. 1991-1999, 2000-2009, and 2010-2018.
        2. How has hate crime behaved over the years, an increase or decrease of incidence?
     2. **Which ethnic group was most affected by hate crimes, which was least affected?**
     3. **Which states have the largest incidence of hate crime?**
     4. **What types of hate crime was most common?**