

# HATE CRIMES IN THE UNITED STATES (1991 – 2018)

Mandy Yeung

## OVERVIEW

### 1. Executive Summary

- ✓ *The FBI's Uniform Crime Reporting (UCR) Program's Hate Crime Statistics Program gathers information on criminal acts that were influenced, either wholly or partially, by the perpetrator's prejudice against a race, gender, gender identity, religion, disability, sexual orientation, ethnicity, and that targeted individuals, property, or society.*

### 2. Data Source

- ✓ *Data Source: This data is publicly available via Kaggle. It is an open-source data downloaded from [United States Hate Crimes \(1991-2018\) via Kaggle](#).*

*Data Collection: This data is administrative data. This dataset was retrieved from the [FBI: Crime Data](#). The data is collected every year.*

*Data Contents: The data contains the date, state, offender race, victim count, type of offense, etc. The dataset ranges from 1991 to 2018.*

*Data Relevance: This data is relevant and significant to answering the questions of this project.*

### 3. Limitations and Ethics (not done yet)

- ✓ *The data aims to provide information and is typically reliable, with instances of bias being uncommon. However, because motives can be subjective, determining definitively whether a crime stems from bias can be challenging. Additionally, the mere presence of bias does not automatically classify a crime as a hate crime. Law enforcement must gather adequate evidence during their investigation to reasonably conclude that the offender's actions were influenced, at least in part, by bias before categorizing an incident as a hate crime.*

*Furthermore, the data may not reflect current conditions since it is updated annually, and the most recent recorded information dates to 2018.*

## 4. Data Cleaning and Data Consistency Checks

- ✔ ➤ *Adjusted data type.*
- *Checked for missing data.*
- *Checked for duplicate data.*
- *Checked for mixed data type.*

## 5. Data Profile

- ✔ *Original data consists of 201,403 rows and 28 columns.*  
*Prepared data consists of 199,797 rows and 22 columns.*

## 6. Column Details

Column	Column Description	Data Type	Time Variant (Y/N?)
INCIDENT_ID	identification number assigned to the incident	Qualitative, nominal	No
DATA YEAR	The data is from 1991-2018	Qualitative, ordinal	No
ORI	State abbreviations followed by numerical values assigned to each incident ID.	Qualitative, ordinal	No
PUB_AGENCY_NAME	Name of agency.	Qualitative, nominal	No
AGENCY_TYPE_NAME	Type assigned to agency	Qualitative, ordinal	No
STATE_ABBR	State abbreviations	Qualitative, nominal	No
STATE_NAME	Name of the state.	Qualitative, nominal	No
DIVISION_NAME	Name of the division	Qualitative, nominal	No
REGION_NAME	Name of the region	Qualitative, nominal	No
POPULATION_GROUP_CODE	Range of the population codes	Qualitative, ordinal	No

POPULATION_GROUP_DESC	Population assigned to each population group code.	Qualitative, ordinal	No
INCIDENT_DATE	Date when incident happened.	Qualitative, ordinal	No
TOTAL_OFFENDER_COUNT	How many offenders are there for the incident.	Quantitative, discrete	Yes
OFFENDER_RACE	Race of the offender	Qualitative, nominal	No
VICTIM_COUNT	Number of victims.	Quantitative, discrete	Yes
OFFENSE_NAME	Name of the offense.	Qualitative, nominal	No
TOTAL_INDIVIDUAL_VICTIMS	Total victims of the incident.	Quantitative, discrete	Yes
LOCATION_NAME	Location of where incident happened.	Qualitative, ordinal	No
BIAS_DESC	Description of bias.	Qualitative, nominal	No
VICTIM_TYPES	Type of victims. (individual, organization, etc.)	Qualitative, ordinal	No
MULTIPLE_OFFENSE	S – assigned if it's single offense. M – assigned if it's multiple offenses.	Quantitative, ordinal	No
MULTIPLE_BIAS	S – assigned if there were single bias. M – assigned if it's multiple bias.	Qualitative, ordinal	No

## 7. Questions to Explore



1. What type of hate crimes are more common throughout 1991-2018?
2. Which ethnic group was most and least affected?
3. Which states had the most hate crimes reported throughout 1991-2018?