Investigation of Hate Crimes in the United States

Authors: Ryan Singh, Luka Marceta, Jordano Mark

Executive Summary

1. How has the number of hate crimes in a year changed from 1991-2018?

   Hate crimes have sporadically increased from 4,589 since 1991, reaching a peak in 2001 with 9,730 incidences of hate crimes. Incidences declined in 2002 and have stayed mostly around that number until experiencing another decrease in 2009. This decrease continued slowly until 2014 when incidences began to rise again from 5,599 to 7,194 in 2018.

2. What groups are the most targeted based on different demographic factors?

   When considering hate crimes from 1991-2018, Anti-Black or African American hate crimes have occurred 69,056 times, far more than those biased against any other racial or ethnic group. With regards to gender-biased hate crimes, Anti-Female hate crimes are the most prevalent at 159 incidences. Anti-gay (male) hate crimes are the most frequent among sexuality-biased hate crimes, and anti-Jewish hate crimes far exceed hate crimes with any other religious bias.

3. Where do most hate crimes occur? Does this change depending on the bias of the crime?

   California has the most reported hate crimes from 1991-2018, followed by New York and New Jersey. These states, along with Michigan and Massachusetts, consistently make up the highest percentage of hate crimes in the United States regardless of the crime's bias.

Motivation

With the recent increase in Asian-American hate crimes and anti-semetic rhetoric, we've found it important to consider the history of hate crimes in America. We hope to identify any geographic or temporal trends relating to hate crimes so that we can ask further questions surrounding the "why?" of these occurrences. The first step is identifying what crimes are happening, where, and against whom.

Dataset

This dataset provides information about differing hate crime incidents from 1991 to 2018 as provided by the FBI: Crime Data Explorer. It includes the date, state, offender race, victim count, type of offense, etc. To visualize our data using GeoPandas, we needed to use a shapefile from the United States Census Bureau that contained the geometry for every state in the United States.

Method

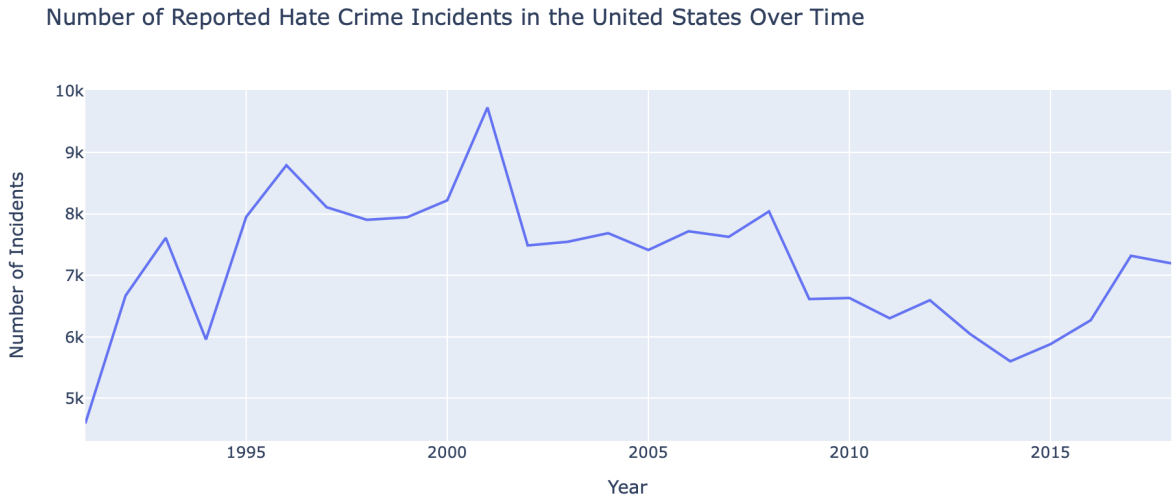
Because the hate crime dataset is over 50 MB, we decided to run our data analysis in a Jupyter Notebook on the collaborative environment JetBrains Datalore. To begin, we imported the Pandas, GeoPandas, Plotly Express, and Plotly Graph Objects libraries that will be necessary to complete our analysis. We also imported assert_equals from cse163_utils for testing. Then we loaded in the hate crime data as a Pandas dataframe to easily perform data analysis functions.

To answer the first research question, we calculated the total number of hate crimes for each year in the dataset by grouping the data by year, and counting the number of incident ID's for each year. From there, we achieved our external library challenge goal and used Plotly Express to plot a line plot of the number of hate crime incidents in a particular year. By using a line plot with year on the x-axis and number of incidents on the y-axis to answer this first question, we can easily see the directional trend in hate crimes incidents from year to year. Plotly Express makes it easy to see the whole range of our data and observe incident counts for a year by hovering over the graph. We used assert statements to compare the figures in the visualization to the data from our calculations.

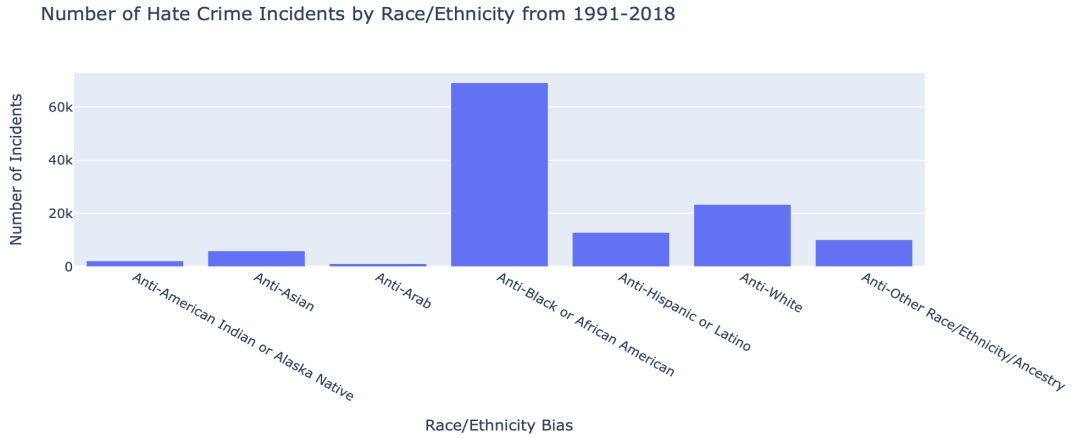
For our second question, we grouped the original dataframe by the bias of the crime and calculated the total crimes for each bias. Since there were a large number of bias descriptions including incidents with multiple biases, we decided to create four separate dataframes to analyze hate crimes on the demographic factors of race/ethnicity, gender, sexuality, and religion. For each dataframe, we selected relevant bias descriptions to the demographic factor, and we plotted the number of incidences from 1991-2018 on a bar chart with the bias description on the x-axis using Plotly Express. These bar charts allow us to clearly see the counts of hate crimes for different biases within a demographic characteristic. This information allows us to dive deeper into which groups of people experience a majority of hate crimes, and we can see the counts of hate crimes relative to populations within a demographic characteristic. We used assert statements for each visualization to compare their figures to the data from our calculations.

Our last question requires us to import the state shapefile data as a GeoPandas file. We then grouped the original data by state and got a count of the number of incident IDs in each state. We then fulfilled the purpose of our second challenge goal by merging the state hate crime occurrences with the state geometry data. We used an assert statement to test the accuracy of our merge by comparing the rows in the original dataframe and the dataframe resulting from the merge. We used Plotly Graph Objects to make a choropleth map of the number of hate crime incidents in each state. We replicated this process to visualize the percentage distribution of anti-gay (male) and anti-Black hate crimes among states, but instead of an occurance count for each state, we calculated a percentage by dividing the hate crime incident count for the state by the total number of incidences in the dataframe. We used assert statements to compare the figures in the maps to the data from our calculations. Displaying this analysis as a map allows the states with the highest counts and percentages of hate crimes to stand out, and there is potential to observe regional similarities and differences in these numbers if they exist.

Results

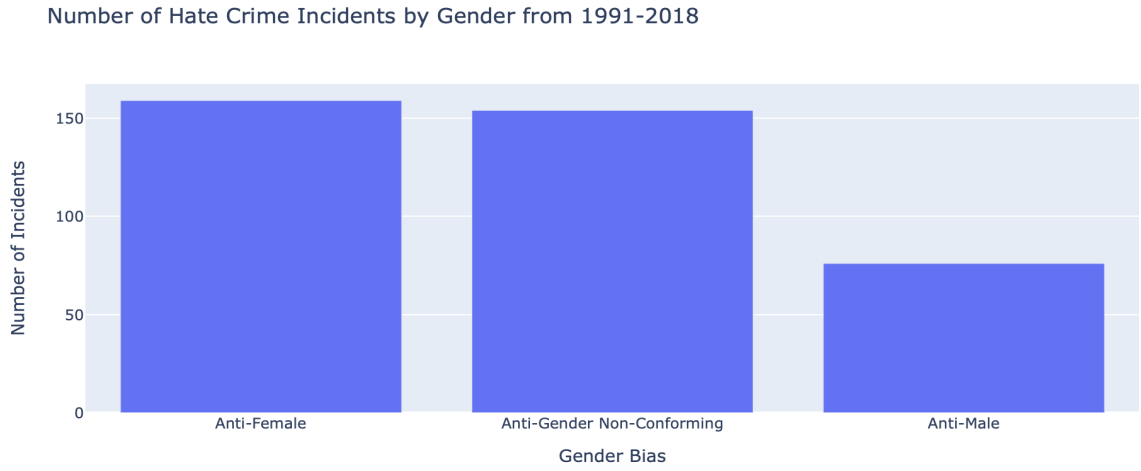


1. Our first research question gave us some interesting results when considering changes over time. Every year since the beginning of our data in 1991, the number of hate crimes has been higher than it was in 1991. After sharp increases, hate crimes started to level off around 7,900 incidents a year until there was another sharp spike in 2001. 2001 had the most hate crime incidents in our dataset at 9,730 incidents. This high level of hate crimes in 2001 could come from the rise in islamophobia stemming from the September 11 attack on the twin towers. Many of the spikes in our data come in the years surrounding a presidential election. This may happen because election periods in the United State typically bring a lot of political tension between many Americans, and a lot of political issues surround immigration, social programs, and social equality. The rise of hate crimes from 2014 to 2017 likely surrounded the rhetoric of President Trump's election which is often associated with white supremacist and anti-immigrant sentiments. Knowing when hate crimes tend to spike could allow people to anticipate when hate crimes in general will occur, and warnings and safety measures could be taken during times of anticipated tension in the United States.

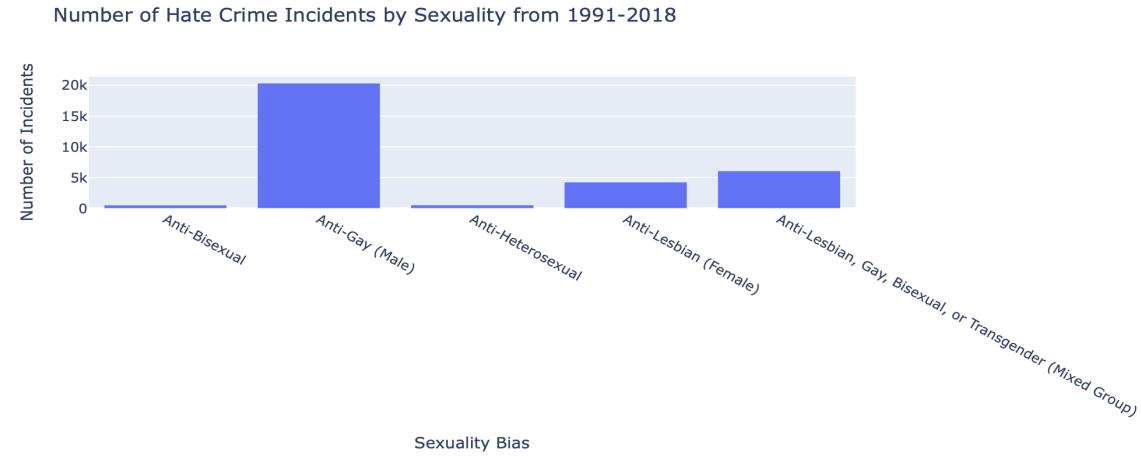


2. When looking at race and ethnicity, anti-Black or African American hate crimes far exceed others with 69,056 incidents reported. This large statistic is likely an outcome of long standing racism and prejudice against Black people and African Americans in American that began with slavery and was further perpetuated in society and culture

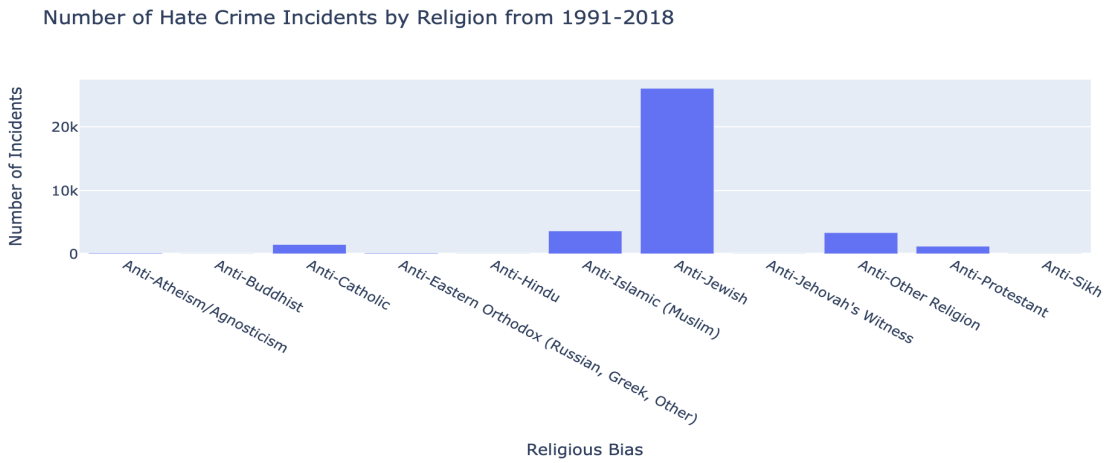
through systemic injustice and mass incarceration. With the increased publicity of anti-asian hate crimes recently, we expected the number of anti-asian hate crimes to be among the highest; however, anti-asian hate crimes from 1991-2018 only totaled 5,913 incidents. This discovery further emphasizes the disproportionality with which Black and African American people experience hate crimes relative to other minorities.



Gender biased hate crimes have happened much less frequently than race/ethnicity based hate crimes, but we found it interesting to see that the 76 anti-male hate crimes were less than half of the anti-female and anti-gender non-conforming hate crimes that totalled 159 and 154 incidents respectively. We believe this reflects the patriarchal and misogynistic nature of American society.

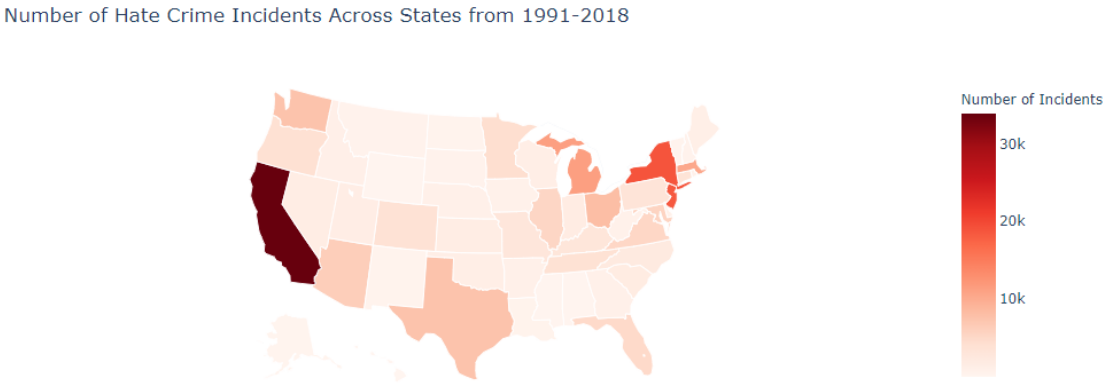


Looking at sexuality biased hate crimes reveals that anti-gay (male) hate crimes far exceed others, totalling 20,316 reported incidents. We believe this disparity is because the mass recognition of sexualities other than homo- and hetero-sexual has begun relatively recently, so people likely targeted gay males as a way to attack anyone whose sexuality wasn't heterosexual.

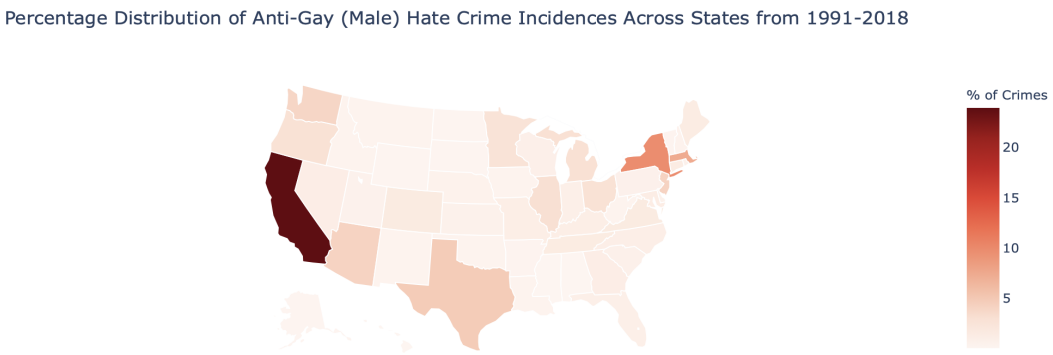


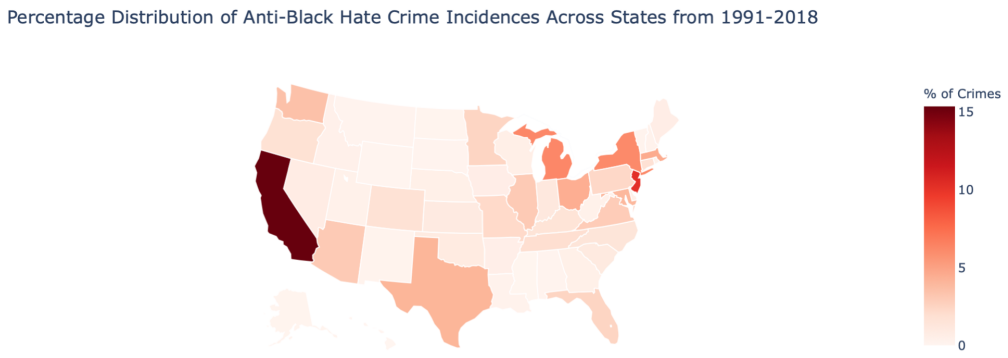
Anti-Jewish hate crimes are extremely high among religion-biased hate crimes, totalling 26,109 incidents from 1991-2018. Acknowledging the long and complex history of the Jewish people and their relationship with a predominantly Protestant United States, there have often been instances of religious conflict in America with Jewish people. Anti-Jewish hate crimes have been persistant throughout American history, and they continue into present day.

Knowing these hate crime disparities, we can begin to see which populations experience the most hate in America. This leads us to ask the larger question of “why” these groups are targeted, and people can start to make efforts to remedy historic contributors to prejudice against particular groups.



3. This question allows us to see where hate crimes are occuring. Our analysis shows that most hate crimes have occured in California with 33,891 incidents in that state, followed by New York and New Jersey. What's interesting is while California and New York are among the top in population size and hate crime incidents, states like New Jersey, Michigan, and Massachusetts have many more hate crimes than larger states like Texas, Florida, and Pennsylvania. This result suggests that hate crime occurrence might not be directly correlated to population size, and there may be other factors that contribute to hate crime occurrence.





Comparing the percentage distribution of anti-Black crimes among states to the percentage distribution of anti-gay (male) hate crimes, California constitutes the highest percentage of hate crimes for both visualizations. An interesting state to look at is New Jersey, who had the third most instances of hate crimes among states and has the second highest percentage of anti-Black hate crimes (10.22%) but only holds only 4.18% of anti-gay (male) hate crimes. This comparison reinforces the idea that hate crime occurrences can vary based on factors other than population, and it makes our team interested to see if any other states constitute high percentages of hate crimes for other biases that we did not analyze. These results could be used by people determining where they want to live, and they can also paint narratives about the culture or people in certain states that may or may not be true. These results could also be used to implement initiatives to prevent hate crimes on a more locationally intentional level.

Challenge goals

1. External Library: Plotly Express and Plotly Graph Objects were both implemented to create effective visualizations for our data analysis. We were originally going to just use Plotly, but we found that the data we wanted to visualize could be handled better with Plotly Express and Plotly Graph Objects.
2. Multiple Datasets: In addition to our hate crime dataset from the FBI, we used a dataset with geometry for every state in the United States. Combining these two datasets we were able to uncover insights surrounding the location of hate crimes. We had proposed to meet the Machine Learning challenge goal, but we realized that we would need to join

a dataset with state geometry to our hate crime data, and we weren't able to find an ethical and practical use for machine learning in our data analysis.

Work plan evaluation

Our work plan estimates were slightly overestimated, with the exception of step 3 of our plan. The first step of completing statistical analysis ended up taking around an hour instead of two, and the following step of creating meaningful visualizations took around an hour for the visualizations associated with our first two research questions. We think these estimates were overestimated because we didn't anticipate the dataset we used to be formatted so well, especially when we utilized Pandas to run our analysis. The Plotly visualizations were also relatively easy to learn because they provided examples with code that matched the visualizations we aimed to create. The last step of our plan wasn't applicable because we didn't end up creating a machine learning algorithm. Instead, we spent around two hours trying to make geospatial visualizations. This took a lot of time because we had to figure out how to work with the shapefiles from the Census Bureau to create choropleth graphs with Plotly.

Testing

Much of our testing came from our decision to use a Jupyter Notebook in conducting our analysis. Using a Jupyter Notebook allowed us to quickly display our dataframes and visualizations after each code block to make sure we were doing the analysis we intended to do and to make sure we were creating the proper visualizations. When we merged two datasets, we used assert_equals to ensure the dataframe we used for the geospatial visualizations had the same number of rows as our original data. Additionally, we tested our visual graphs with assert statements to ensure that the number of hate crimes in a state or for a given group matches what we calculated in our data analysis.

Collaboration

Aside from TA's and ourselves, we consulted the Plotly documentation to better understand how to create visualization with their libraries. We also explored the Census Bureau's website to understand the formatting of the shapefiles and how to use them. Kaggle was very helpful in providing documentation to explain the columns for our hate crime dataset, and from there we were able to figure out which columns were most relevant for our analysis.