

Check in beyond  
about component

## 3Lattes Final Project Proposal

Group name: 3Lattes

Title: Civil Rights in States: Exploring U.S. Hate Crime Data in 2019-2020

### Purpose

“Civil rights investigations are at the heart of what we do at the FBI for the simple reason that civil liberties and civil rights are the very heart of who we are as Americans.” FBI Associate Deputy Director Jeffrey Sallet said during an FBI conference in Denver on June 2021<sup>1</sup>.

On In 2020, it seemed like there were sensational headlines in the news everyday. From the question of the origins of Covid-19 to the death of George Floyd followed by the Black Lives Matter movement, tensions between individuals and communities arose within the U.S. For our group, the onset of Hate Crimes reports on the news caught our attention. In social media, there are many claims that there have been more anti-Asian and anti-African-American hate crimes over the past years. In fact, Sallet said that “the FBI’s Criminal Investigative Division has elevated civil rights violations to its highest-level national threat priority—a measure of how the FBI allocates money and resources”<sup>2</sup>. From this, we want to see what the FBI data says about Hate Crimes to determine where the money and resources should be allocated to. If Civil Rights is at the heart of who Americans are, then the FBI will want to address the following questions to figure out which states and racial groups need the most attention:

- Which states have seen the most significant change in the count of race-based hate crimes between 2019 and 2020?
- On a state-by-state basis, which specific race-based hate crimes have significantly changed during 2019-2020?
- Is there a relationship between race-based hate crime with population or income? How can we group states up by these relationships and are there significant differences between each group?

### Data

We will have 2 main datasets detailed below.

#### *Hate Crime Dataset*

We are primarily exploring the FBI Hate Crime data in 2019-2020 ( <https://crime-data-explorer.fr.cloud.gov/pages/explorer/crime/hate-crime>). The downloaded file from the website is in a CSV format. However, we anticipate having challenges in filtering for the specific years and selecting variables. For simplicity, we will focus on single offenses.

Selected variables:

- Year: The year of the hate crime should filter for 2019-2020

---

<sup>1</sup><https://www.fbi.gov/news/stories/hate-crimes-and-civil-rights-elevated-to-top-national-threat-priority-063021>


<sup>2</sup>Ibid.

- State\_Abbr: State ID (e.g: AK, MA, etc)
- Bias\_Desc: FBI's category for the type of bias involved in the hate crime

Our final dataset will have the following variables:

- State
- Percent Change in total race-based hate crimes
- Percent Change in total hate crimes by specific biases including:
  - Anti-Black or African American
  - Anti-White
  - Anti-Hispanic or Latino
  - Anti-Asian
  - Anti-Arab
  - Anti-Native Hawaiian or Other Pacific Islander
  - Anti-American Indian or Alaska Native
  - Anti-Multiple Races
  - Anti-Other Race/Ethnicity/Ancestry

### *Population & Income Dataset*

We will take data from KFF.org to find the following information. The 2020 and 2019 data is based on an analysis of the U.S Census Bureau's March Supplement to the Current Population Survey (the CPS Annual Social and Economic Supplement or ASEC) from 2016, 2018, and 2020 (due to the challenges of the pandemic). 

Selected variables:

- Population per state (2020) (<https://www.kff.org/other/state-indicator/total-number-of-residents-cps/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>)
- Race-specific population per state (2019) (<https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?dataView=1&currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>)
- Median household income per state (2019) (<https://www.kff.org/other/state-indicator/median-annual-income/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>)

Our final dataset will have the following variables:

- State
- Total race-based hate crimes (2019)
- Population per state (2019)
- Race-specific population per state (2019)
  - White
  - Black
  - Hispanic
  - Asian
  - American Indian or Alaska Native
  - Native Hawaiian or Other Pacific Islander
  - Multiple Races
- Median household income per state (2019)

Limitations: As our data for each state is an aggregate of every city and county, the results of our analysis is very generalized for each state. For further, more accurate research, finding city or county-based statistics would be better, however, we were limited in the resources we could find.

*Yes ok*

## Final Product

To address our first question, we will make a map to visually represent the change in hate crime from 2019 to 2020. For the second question, we plan to create a bar graph to show the top 10 states with the highest percent change in hate crime based on the specific bias.

For the last question, we will use clustering analysis to see if we can characterize states together using the variables of the second dataset. We will initially make a clustering matrix, then select which variables to look into specifically. We will then run an ANOVA analysis to determine whether there are significant differences between each cluster.

What aspect is "new"? Beyond component

## Schedule

April 28 - Finish wrangling data (e.g create all usable datasets), and start to make our map and graph.

May 5 - Finish making data visualizations, figure out which clustering analysis to use, and work on the report & presentation.

May 10 - Presentation

May 13 - Finish report

May 20 - Finish reflection

## Questions

- Which clustering analysis would be the most appropriate?

↳ It's more you need something new to go "beyond".

You could run k-means and compare it to a new method. → hclust, agnes, diana, clara, etc.

Or, just use k-means

but explore cluster validation  
→ there's an R pkg for this and it would give you some statistics to explore (and can compare methods w/o learning them all).