Asian Hate Crime Analysis: A Visualization Design Study on the Change of Asian Hate Crime in the Past 20 Years

Shuoxin Liu
The City College of New York

Yi Lin
The City College of New York

Huihong Yu
The City College of New York

Sambeg Raj Subedi
The City College of New York

Ruixiang Huang
The City College of New York

Dor Ulman
The City College of New York

Abstract

We have chosen the topic Asian Hate Crimes for the data visualization project. The main reason behind choosing this topic is that in recent years, we have seen or heard about the hate crimes on Asian people more frequently through television or social media. This made us more familiar and curious to know more about this hate crime in detail. This visualization paper will be analysis the hate crime from 1991 to 2019 and will focus on the anti-Asian hate then compare the number of hate crime between 1991 until 2019 and we also find out what cause this situation.

1 Introduction

Visualization is a useful method to show the relation of the data and if you choose a right visualization of the data and also see how is the trend of the data. The past two years are the restless years because of the COVID-19 virus. There were many people dead and the economy also decreased for all the country. Although, there are already 2 years about the epidemic, the situation is not getting very better for the whole world. However, the situation is getting better than before in some countries which have the vaccine, such as US, China and so on. During the hard time of the epidemic, there appeared many of the crime and the hate crime increased very fast and most of the hate crimes were anti-Asian because of many people thought that the virus came from China and there hate Chinese, thus there were some people attacked Asian no matter if there were Chinese. Because of this situation, we decide to make a visualization to see how's the trend of hate crime from this few 10 years or 20 years. We collected the data from the CDE (The Crime Data Explorer) which was a program from FBI's Uniform Crime Reporting Program. The data we got was a csv file which included all kind of the hate crime data from 1991 to 2019. We separated it, analyzed it and use the analyzed data to make a visualization and saw how difference of the hate crime change.

2 Data set that has been used

2.1 FBI DATA SET

The Hate Crime Statistics dataset provides annual statistics on the number of incidents, offenses, victims, and offenders in reported crimes that are motivated in whole, or in part, by an offender's bias against the victim's perceived race, gender, gender identity, religion, disability, sexual orientation, or ethnicity. Hate Crime data are captured by indicating the element of bias present in offenses already being reported to the UCR Program.

All law enforcement agencies, whether they submit Summary Reporting System (SRS) or National Incident-Based Reporting System (NIBRS) reports, can contribute Hate Crime Data to the UCR Program using forms specified to collect such information.

The Hate Crime Statistics Program of the FBI's Uniform Crime Reporting (UCR) Program collects data regarding criminal offenses that were motivated, in whole or in part, by the offender's bias against a race, gender, gender identity, religion, disability, sexual orientation, or ethnicity, and were committed against persons, property, or society. Because motivation is subjective, it is sometimes difficult to know with certainty whether a crime resulted from the offender's bias. Moreover, the presence of bias alone does not necessarily mean that a crime can be considered a hate crime. Only when a law enforcement investigation reveals sufficient evidence to lead a reasonable and prudent person to conclude that the offender's actions were motivated, in whole or in part, by his or her bias, should an agency report an incident as a hate crime.

- `Incident types`: The UCR Program collects data about both single-bias and multiple-bias hate crimes. A single-bias incident is an incident in which one or more offense types are motivated by the same bias. Beginning in 2013, law enforcement agencies could report up to five bias motivations per offense type. Therefore, the definition of a multiple-bias incident has been revised to an incident in which one or more offense types are motivated by two or more biases.

- `Offense types`: The law enforcement agencies that voluntarily participate in the Hate Crime Statistics Program collect details about offenders' bias motivations associated with 13 offense types already being reported to the UCR Program: murder and nonnegligent manslaughter, rape (revised and legacy definitions), aggravated assault, simple assault, intimidation, human trafficking commercial sex acts, and human trafficking involuntary servitude (crimes against persons); and robbery, burglary, larceny-theft, motor vehicle theft, arson, and destruction/damage/vandalism (crimes against property). The law enforcement agencies that participate in the UCR Program via NIBRS collect data about additional offenses for crimes against persons and crimes against property. These data appear in Hate Crime Statistics in the category of other. These agencies also collect hate crime data for the category called crimes against society, which includes drug or narcotic offenses, gambling offenses, prostitution offenses, weapon law violations, and animal cruelty offenses. Together, the offense classification other and the crime category crimes against society include 39 Group A offenses that are captured in NIBRS, which also collects the previously mentioned 13 offense types. (The NIBRS User Manual provides an explanation of the 52 Group A offenses.) Beginning in 2015, all law enforcement agencies could report human trafficking offenses. However, no human trafficking offenses with a bias motivation were reported during 2016.

- `Crimes against persons, property, or society`: The UCR Program's data collection guidelines stipulate that a hate crime may involve multiple offenses, victims, and offenders within one incident; therefore, the Hate Crime

Statistics Program is incident-based. According to UCR counting guidelines:

  - One offense is counted for each victim in crimes against persons.
  - One offense is counted for each offense type in crimes against property.
  - One offense is counted for each offense type in crimes against society.

- `Victims`: In the UCR Program, the victim of a hate crime can be an individual, a business/financial institution, a government entity, a religious organization, or society/public as a whole. Law enforcement can indicate the number of individual victims, the number of victims 18 years of age and older, and the number of victims under the age of 18.

- `Offenders`: According to the UCR Program, the term known offender does not imply that the suspect's identity is known; rather, the term indicates that some aspect of the suspect was identified, thus distinguishing the suspect from an unknown offender. Law enforcement agencies specify the number of offenders (adults and juveniles) and, when possible, the race and ethnicity of the offender or offenders as a group.

- `Race/ethnicity`: The following five racial designations in the Hate Crime Statistics Program are: White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander. In addition, the UCR Program used the ethnic designations of Hispanic or Latino and Not Hispanic or Latino.

- `Data reporting` : Law enforcement agencies report hate crimes brought to their attention monthly or quarterly to the FBI either through their state UCR Programs or directly. These agencies submit hate crime data electronically in a NIBRS submission, the hate crime record layout, or a Microsoft Excel Workbook Tool.

- `Reporting via NIBRS`: Agencies that report offense data to the FBI via NIBRS use a data element within their reporting software to indicate whether an incident was motivated by bias. Because NIBRS is an incident-based, comprehensive data collection system, these agencies report considerably more information about a hate crime than is captured in the other electronic record or the Excel workbook. For example, the data element that indicates bias motivation applies to 52 Group A offenses, and agencies can report information such as the age, sex, race, and ethnicity of victims, offenders, and arrestees. Although the additional data collected via NIBRS are not maintained in the hate crime database, they are available in NIBRS flat files. When agencies submit a Group A Incident Report with a bias indicator of ìNone,î a Group B Arrest Report (because no offenses [bias-motivated or otherwise] occurred in their respective jurisdictions), or a Zero Report (because no offenses [bias-motivated or otherwise] or arrests occurred), the FBI records zero hate crime incidents for that agency for the reporting period.

- `Reporting via the electronic hate crime record layout`: Law enforcement agencies that do not report via NIBRS may use the hate crime record layout specified in the publication Hate Crime Technical Specification.

- `Reporting via Microsoft Excel Workbook Tool`: Agencies that use the Excel Workbook Tool capture the following information about each hate crime incident:

  - Offense type and the respective bias motivation
  - Number and type of victims
  - Location of the incident
  - Number of known offenders
  - Race and ethnicity of known offenders
  - Number of adult or juvenile victims and offenders

  For each calendar quarter, law enforcement agencies submit a Hate Crime Incident Report for each bias-motivated incident or designate a Zero Report for the month. Each month, law enforcement agencies submit a Hate Crime Incident Report for each bias-motivated incident as part of their regular SRS submissions. When reporting zero incidents of hate crime, the agency should select the zero report box within the Microsoft Excel Workbook Tool. For updating purposes, the agency should retain a copy of the report. Agencies should select the adjustment box within the Microsoft Excel Workbook Tool to make corrections/updates and provide the appropriate changes. Agencies should select the delete box within the Microsoft Excel Workbook Tool to delete incidents.

- `Population estimation` : For the 2016 population estimates, the FBI computed individual rates of growth from one year to the next for every city/town and county using 2010 decennial population counts and 2011 through 2015 population estimates from the U.S. Census Bureau. Each agency's rates of growth were averaged; that average was then applied and added to its 2015 Census population estimate to derive the agency's 2016 population estimate.

- `Universities and colleges`: The figures listed for universities and colleges are student enrollments that were provided by the United States Department of Education for the 2015 school year, the most recent available. The enrollment figures include full-time and part-time students.

- `County designations`: Based on the Office of Management and Budget's standards for defining Metropolitan Statistical Areas, the UCR Program refers to suburban counties as metropolitan counties and to rural counties as non metropolitan counties.

- `Publication Annotation`: Narrative portions of this publication present percentage breakdowns for various facets of tabular data. Where percentage breakdowns are used, percentages may not add to 100.0 percent due to rounding.

Link to the data in our github: `https://raw.githubusercontent.com/visualization-project-group/surprise-map-assignment/master/hate_crime.csv`

2.2 NYPD DATA SET

NYC Hate Crime Report Data set is a arrest statistics involving hate crime incidents. The following data are collected: Arrestee Gender, Race, Age and Bias Motivation.

Link to the data in our github: `https://raw.githubusercontent.com/Lsx961029/DATA-SET/main/hate-crime-arrests-by-motivation-annual-2018.csv`

2.3 Twitter DATA SET

A random 1% sample of publicly available tweets were collected from June 2015 to July 2018 using Twitter's Streaming Application Programming Interface (API). We restricted our analyses to English language tweets from the United States with latitude and longitude coordinates or other place attributes that permitted identification of the state or county location where the tweet was associated. All tweets included in the analysis used one or more of 518 race-related keywords. These keywords were compiled from racial and ethnic categories used by prior studies examining race-related online conversations (Bartlett, Reffin, Rumball, & Williamson, 2014; Pew Research Center, 2016) and an online database of racial slurs ("The Racial Slur Database," 2018). Tweets were classified into five main racial/ethnic categories: Asians, Arabs, Blacks, Latinos, and Whites according to the keywords used. Details of the data collection process including the full keyword list have been previously published (Nguyen et al., 2019).

We performed a sentiment analysis on the Twitter data set. This procedure has been previously described (Nguyen et al., 2020). Briefly, we utilized Support Vector Machines (SVM), a supervised machine learning model to label the tweets. We obtained training data from manually labeled Sentiment140 (n=498) (Sentiment140, 2011), Kaggle (n=7,086) (Kaggle in Class, 2011), Sanders (n=5,113) (Sanders Analytics, 2011) and 6,481 tweets labeled by our research group. Sentiment140, Sanders, and Kaggle datasets are all publicly available training datasets specifically labeled for sentiment analysis. We compared negative tweets (assigned a value of 1) to all other tweets (assigned value of 0). We used 5-fold cross validation to assess model performance and reached a high level of accuracy for the negative classification of tweets (91%) and a high F1 score (84%). We then labeled all the collected tweets using SVM model by assigning a dichotomized sentiment value (1 versus 0) to each tweet. State and year specific sentiment variables were created by averaging the dichotomous sentiment value of tweets referencing various racial/ethnic groups. State-level sentiment scores are a continuous measure of the proportion of tweets that are negative and scaled so that a one-unit increase represents a 10% increase in proportion of tweets that have a negative sentiment. State and year specific racial sentiment data are then merged with state and year specific data on hate crimes, racial attitudes from the General Social Survey, and explicit and implicit bias from Project Implicit.

Link to the data in github: `https://raw.githubusercontent.com/visualization-project-group/surprise-map-assignment/master/twitter_data.csv`

3 Background and related work

3.1 Related work

We build the work through the classes of visualization techniques that we apply, which in many cases depends on the analysis tasks to be supported. In summary, first of all, we review the visualized data methods and highlight the changes in Asian Hate Rate over the past 10 and 20 years, including in the United States. Second, we will discuss systems that provide an overview of multiple sequences,typically represented as visual colors. We summarize multivariate data visualization techniques, focusing on Asian hate crime visualization.

3.2 Definition of Hate Crime

Hate crimes are crimes motivated by race, religion, sexual orientation, or racial prejudice.In some states, it includes bias against gender, age, and gender identity.Hate crime laws have been passed in 47 states and the federal government since the 1980s.Currently, only Arkansas, South Carolina, and Wyoming do not have hate crime laws.To be prosecuted as a hate crime, the crime – whether battery, murder, or vandalism – must be an act of direct assault on a person motivated by prohibited prejudice.In other words, hate crimes punish motivation; Prosecutors must convince a judge or jury that victims were targeted because of their race, faith, sexual orientation, or other legally protected characteristics.

3.3 Visualization 1: PCA Visualization of Asian Hate Crime

3.3.1 Data Set: FBI

3.3.2 Method: PCA

Principal component analysis (PCA) is a statistical method of dimension reduction, and it is by using an orthogonal transformation, the original random vector that is relevant to the component into its component is not related to the new random vector. This appears to be the initial random vector on the algebra of co-variance matrix transformation into a diagonal matrix, on the geometry of the initial coordinate transformation into a new orthogonal coordinate system. Then, the multi-dimensional variable system can be converted into a low-dimensional variable system with higher precision. Then, the low-dimensional system can be further transformed into a one-dimensional system by constructing an appropriate value function.

The principle of principal component analysis is to try to recombine the original variables into a new set of several comprehensive variables which are unrelated to each other. At the same time, according to actual needs, a statistical method that can take out several fewer sum variables to reflect the information of the original variables as much as possible is called principal component analysis or principal component analysis, and it is also a method to deal with dimensionality reduction in mathematics. Principal component analysis (PCA) is an attempt to replace the original index by recombining the numerous initial indicators with specific correlation (such as P indicators) into a new set of comprehensive indicators that are unrelated to each other. The usual mathematical treatment takes the original P indices as a linear combination as a new composite index. The most classic approach is to use the variance of F_1 (the first linear combination selected, i.e., the first comprehensive index) to express, that is, the larger$VA(rF1)$ is, the more information F1 contains. Therefore, F1 selected from all linear combinations should be the one with the largest variance, so it is called the first principal component. If the first principal component is not enough to represent information of original P indicators, then consider to choose F2 is to choose the second linear combination, in order to effectively reflect the original information, F1's existing information is not needed to appear again in the F2, is expressed in mathematical language to Cov (F1, F2) = 0, says F2 for the second principal component, and so on can be constructed out of the third, fourth, ..., the P principal component.

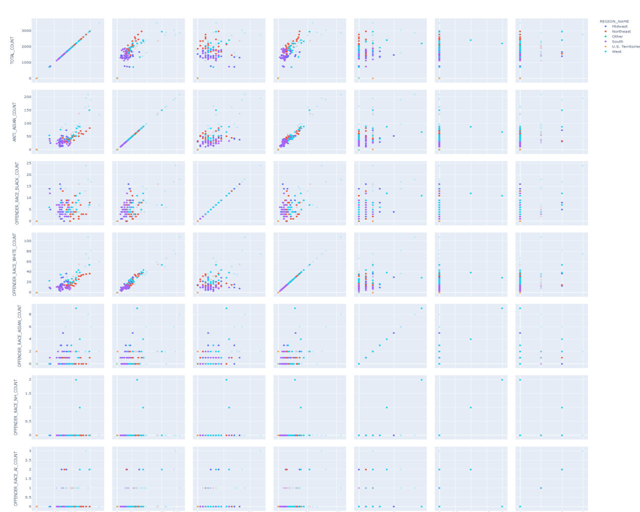

Figure 1: Scatter Plot Matrix

We build two PCA visualization based on the count of Asian hate crimes during 1991-2019 in different regions of America. We choose the each region in each year within the range as an instance and use different offender race count of Asian hate crimes as attributes. Initially, in the figure 1 we have 7×7 scatter plot matrix which contains 7 variables for each instance. Then we use dash transformed it into a PCA which contains 2-6 variables accordingly. Therefore, a 7 dimensional system can be reduced to 2 dimensional system which has a explained variance 69.74%. The total variance is the sum of variances of all individual principal components. The fraction of variance explained by a principal component is the ratio between the variance of that principal component and the total variance. The second PCA we have is just a example of 3 Components in a 3 dimensional system. We may find the points with same color as a cluster since the population and race difference in different regions of America.

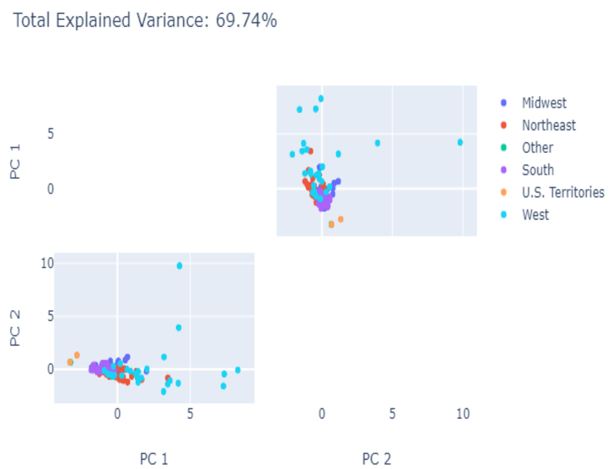

Figure 2: 2 Components PCA Scatter Plot Matrix

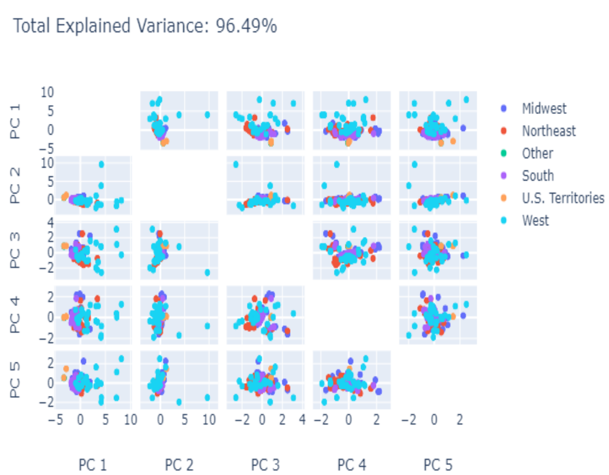

Figure 3: 5 Components PCA Scatter Plot Matrix

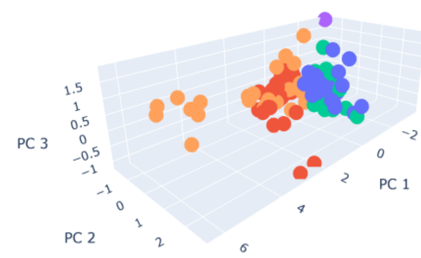

Figure 4: 3D PCA

3.3.3 Analysis

3.4 Question 1: What's the level of Asian Hate in 2019 before COVID-19

3.4.1 Data set: FBI

3.4.2 Method: Bayesian Surprise Map

In 1977, the astronomer Jerry Ehman worked with SETI to study extraterrestrial life when he stumbled upon an exciting radio signal stuck like a needle in the haystack of all SETI's electromagnetic signal monitoring devices.It's a potent radio signal that meets many of the parameters we'd expect to see to determine whether aliens are trying to communicate with us.Circle the signal in red ink and write, "Wow!"When analyzing data, we are usually not interested in day-to-day business: we care about rules and outliers.But when we visually display the data, anomalies and outliers can disappear into the usual sea of change.For geographic data, our proposed solution is called a "surprise map": a heat map that gives more weight to unexpected data.The idea behind the surprise map is that when we look at data, we often have all sorts of expectation models: what we expect to see in the data and what we don't see.If we have these models, we can also measure the deviations or differences in these models.The deviation was unexpected, and the data surprised us.These surprising numbers are sometimes necessary, at least for subsequent analysis.The surprise chart is helpful when the raw data on its own doesn't tell us much: visual patterns may look complex but only convey statistical noise, or designs may look simple but hide exciting features.

3.4.3 Analysis

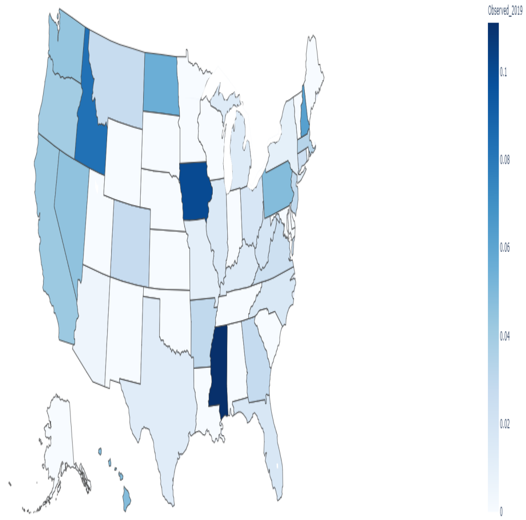

Figure 5: USA 2019 Asian Hate Crimes Rate Observed Heat Map

There are 3 important graphs in this visualization. First we made this USA 2019 Asian Hate Crimes Rate Observed Heat Map to show a the proportion of Asian hate crimes over all kinds of hate crimes and the proportion is color encoded. We can see that in 2019, Iowa, Idaho and Mississippi have a high Asian hate crime rate and some states like South Dakota, Minnesota, etc...They have a Asian hate crime rate which is 0. As we know simple heat map has no accounting for confounds and also has no accounting for variance. Then we process the data and calculate the surprise. The following are the formulas:

$P(A|B) \sim P(B|A)P(A)$
$P(O|M_{seasonal}) \approx 1–|O-E|$
$P(M_s|O) \sim P(M_s)P(O|M_s)$
$Surprise \approx D_{KL}(P(M|O)||P(M))$

The following 2 surprise maps are based on last 10 years and last 20 years respectively:

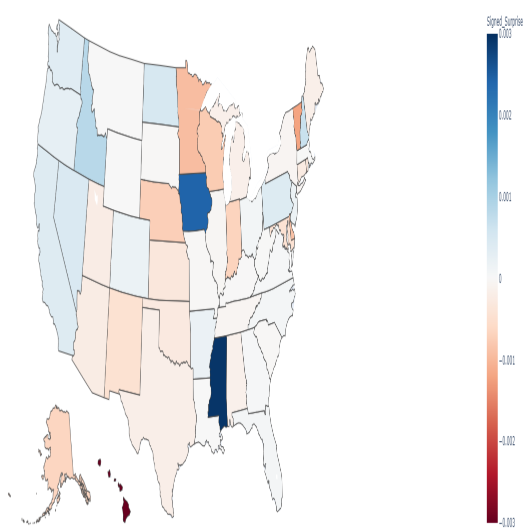

Figure 6: USA 2019 Asian Hate Crimes Bayesian Surprise Map based on last 10 years

Comparing 2019 with last 10 years, Iowa and Mississippi do have a unmoral Asian Hate rate appears in the surprise map. However a lot of the states also the decreasing trends for the proportion of Asian hate crimes over all kinds of hate crimes.

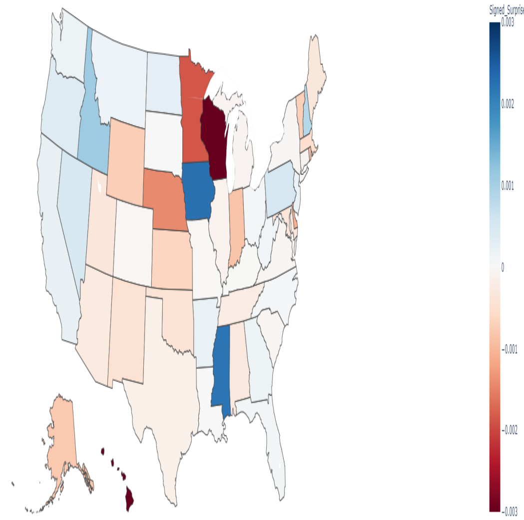

Figure 7: USA 2019 Asian Hate Crimes Bayesian Surprise Map based on last 20 years

If we look back even farther. Comparing 2019 with last 20 years, we can find a larger variance between each state. The state like South Dakota, Minnesota and Wisconsin pop up showing a large decreasing on the proportion of Asian hate crimes.

3.5 Question 2: What are the time in history when hate crimes had a higher occurrence than other times?

3.5.1 Data Set: FBI

3.5.2 Method: Scatter Plot

The visualization function is that the Y-axis is the number of people and the X-axis is year. The attribute also includes Anti Asian, Anti-White, Anti-Black or African American, Anti-Hispanic or Latino, Anti-American Indian, Alaska Native, and all hate Crimes.In addition, each color represents each attribute, which enables us to understand better the total hate crime of each year and the cases of each race.In addition, we went for three years, five years, ten years, and all of them.In this way, we can find the required number of years and quickly compare the data of these three years.At the same time, when we want to know the number of hate crimes in each year in detail, we can use the mouse to click each year to be prepared to know how many cases occur each year and each race.

3.5.3 Analysis

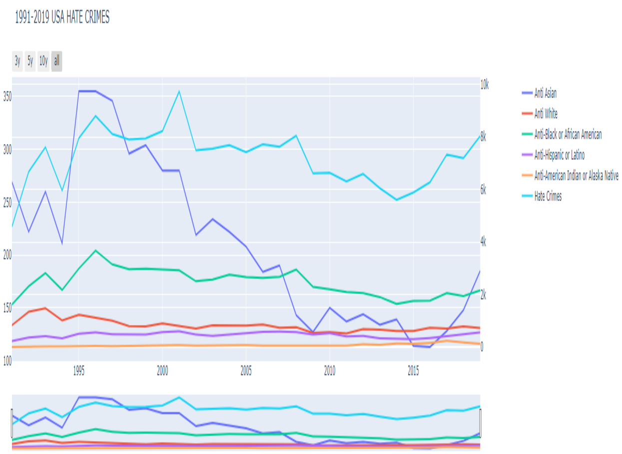

Figure 8: 1991-2019 USA hate crime

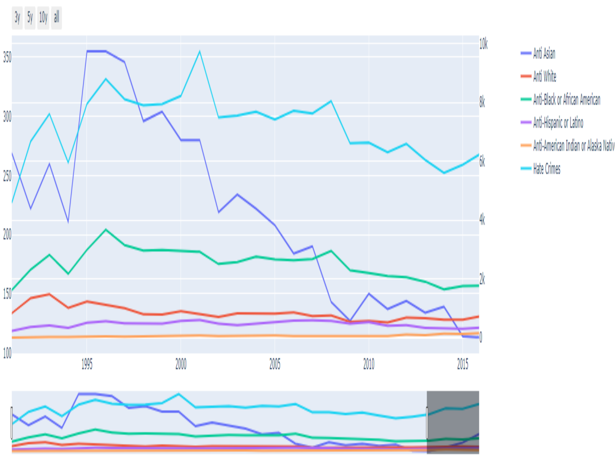

Figure 9: 1995-2015 USA hate crime

The number of routes in 2019 is growing significantly faster than in the past decade.Moreover, the numbers are growing faster in some states, where Asian-Americans commit significantly more hate crimes than their race.In this visualization, we can see that anti-Asian populations are increasing again after 15 years of decline.The FBI has released Hate Crime Statistics (2019), an updated compilation of the Uniform Crime Reporting (UCR) project covering bias-based motivated incidents across the country.In 2019, data submitted by 15,588 law enforcement agencies provided information on crime locations, victims, offenders, and hate crimes. Law enforcement agencies filed 7,314 criminal cases, including 8,559 involving race, ethnicity, descent, religion, sexual orientation, disability, gender, and gender identity.Please note that UCR procedures do not assess crimes committed by institutions that have not submitted reports.Here are the highlights of the 2019 hate crime statistics.Law enforcement agencies can designate one of the 46 places where hate crimes occur. In 2019, most hate crime incidents 24.6% occurred in or near residential/premises.More than 18%(18.2) on highways/highways/alleys/streets/sidewalks;9.6%Studying at school/university;4.7%Study in Parking/Parking/Garage;The location was reported to be another/unknown location, accounting for 11.2% of hate crime incidents.The remaining 27.3 percent of hate crimes occurred elsewhere or in multiple locations

The United States has about 1.6 million Chinese immigrants (including those born in Hong Kong), making them the fourth-largest immigrant group after Mexicans, Filipinos, and Indians.Although Chinese immigration to the United States dates back to the 19th century, the Chinese immigrant population multiplied in the 1990s and 2000s.Today, there are almost as many Native American citizens of Chinese ancestry as there are Chinese immigrants.The American economy is undergoing profound changes.Compared with the 1980s, the American economy in the 1990s has the characteristics of "three low, three high and three increase" and the corresponding sustained and moderate growth.There are several reasons why the U.S. economy is doing well.The fact that the United States is leading the way in the new round of scientific and technological revolution and the widespread adoption of recent scientific and technical achievements are prominent reasons, but not the only ones.This relatively good momentum may continue for some time, but this does not rule out the possibility of a cyclical downturn hitting the U.S. economy around the turn of the century.The U.S. economy is far from solving all its economic and social problems.On the contrary, chronic diseases such as polarization have become more prominent and become hidden dangers that cannot be ignored in American society. Economic decline has been a significant problem in the postwar American economy. Faced with job opportunities being snatched up by new immigrants, native residents have a bad mood and motivation to emigrate.In addition, many Asians have deficient status in the United States, with jobs like laundry and dish-washing.Therefore, the Asian hate rate was exceptionally high in 1998. When the American economy realized the "zero point plan" of the budget balance in 2002, the American economy began to increase, and the Asian hate crime and all hate crime began to decline.Some Asian people began to enter the high-level people of all walks of life, and their status became higher and higher.

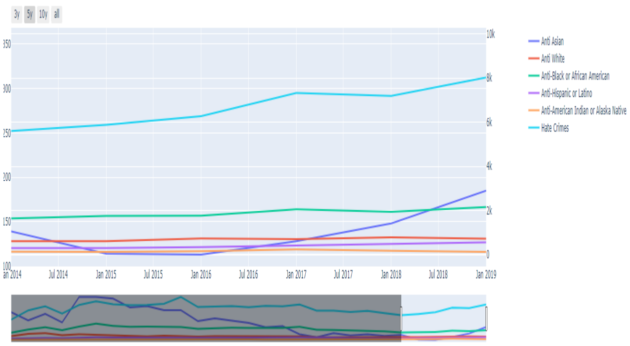

Figure 10: 2014-2018 USA hate crime

The US National Security Strategy Report released by the Trump administration on December 18, 2017 regards China as its primary strategic competitor. The trade war between China and the US has affected the relationship between China and the US. Through news reports and some untrue news, other races feel more hate towards Asian people, believing that Asian people are preventing the development of the US.Rising suspicions and worst-case scenarios have taken on a grim mood, raising the prospect of future political or military crises in Asia between the US and China that could end up propelling the two sides into a cold war or worse.As the balance of power in Asia shifts in the First Island Chain, the relationship between the two countries has become increasingly suspicious and hostile.The negative turn also reflects the region's general failure to resolve several contentious issues, including the Korean Peninsula, Taiwan, maritime disputes and military-related intelligence, surveillance and reconnaissance activities.These factors make life easier for Asians under the pressure of American news.Therefore, after 2017 years, Asian hate crime began to increase rapidly again

3.6 Question 3: The influence of COVID-19 to Asian hate between 2018-2020

3.6.1 Data Set: NYPD

3.6.2 Method: Bar Chart

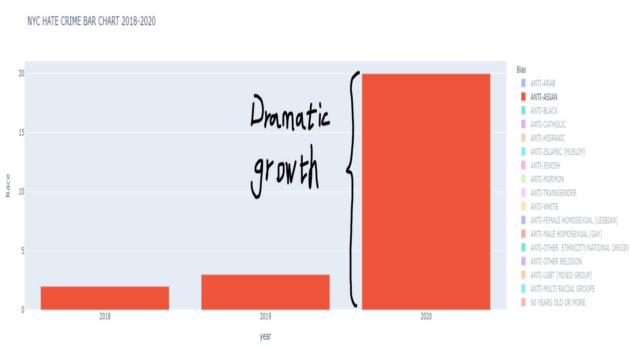

Figure 11: NYC hate Crime bar chart 2018-2020

3.6.3 Analysis

This is a bar chart that analysis the anti-Asian before COVID-19 and during COVID-19. We have the select bar on the right side which can change the highlight on the bar chart. We only select the anti-Asian to see how difference between 3 years. We can see that the red bar show the number of anti-Asian hate crime during that year. Number of anti-Asian hate crime almost 10 times of 2019 and 2018. The growth trend is dramatic. Because of the COVID-19, many of the Asian became the hate crime attack target by racist. There appear a website site which called Stop AAPI Hate to help Asian people to report crime. Most of the racist attack the old man/women and most of them do not speak English, thus they are hard to get help from other.

Eight people, six Asian women, were killed in three separate shootings at massage parlors in Atlanta, Georgia, on March 16.The suspect, Robert Aaron Long, a 21-year-old white man, said he did not kill him in a racial hatred but to eliminate a hidden temptation he found at a massage parlor.After the shooting, Asian-Americans, who had been building anxiety and panic over the past year, took to online platforms to complain bitterly."This mass shooting targeted Asian women and their businesses," social media expert Mark King wrote on Twitter.

This is not an isolated incident."However, a police spokesman said the suspect had been in a bad mood that day.This news will make Asian Americans feel very scared because "you are a virus," "you don't belong here," "you *** immigrant"...Amid insults and name-calling, Asian-Americans have been pushed, spit on, beaten, and shot, regardless of gender, age, or occupation.Racism and hatred have developed into a "spiritual plague" that has made the living environment of the Asian American community dangerous. Tragedy continues to unfold, leaving Asian Americans living under the pressure of insecurity and fear that they will be killed."If the status quo does not change, who can guarantee that the next victim will not be my relatives or me?""Wrote an Asian citizen on the Internet.And it's not just about generic Asian hate crime. Anna Mok, a Chinese-American executive at Deloitte, a consulting firm, who lives in San Francisco, acknowledged her privilege but said that because of hate crimes against Asians over the past year, she has had friends advise her not to even go for a walk.

"I don't think I've ever been so worried about the threat of physical violence," Ms. Mo said. She added that many Asians who work for large corporations have described similar pressures to her: "There's no buffer, there's no segregation.No matter how much money you make, no matter how successful you are, that's the reality of being Asian in the United States."When yellow skin becomes a reason for violent crime, Asians can only live in fear, keeping a high alert to the dangers that may occur around them at any moment[5].

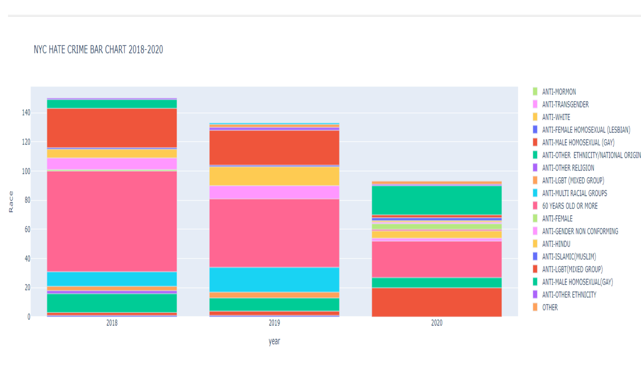

Figure 12: NYC hate Crime bar chart 2018-2020

As of early May 2020, more than 1.8 million people have tested positive for COVID-19 in the United States alone, and there have been more than 105,000 deaths, with the number increasing every day. Although researchers have traced the virus to European travelers in the United States and travelers within the United States, some public members believe Asians are Asians. Americans suspect and are carriers of the disease. On April 28, 2020, NBC News reported that30% of Americans had witnessed someone blaming Asians for the coronavirus.

The COVID-19 pandemic has exposed the negative perceptions of Asian-Americans that have long been prevalent in American society. In the United States, many people consider the virus to be alien, phenotypically blaming Asians as transmitters of the virus. Not only are they at risk of being exposed to COVID-19, but they also have to deal with the additional risk of being victimized, which may increase their anxiety.

Historically, from the late 19th century to the mid-20th century, popular culture and the news media have described Asians in the United States as "yellow danger," a symbol of Western fears of uncivilized, non-white Asian invasion and domination. During the COVID-19 period, the perceived threat of the yellow peril is likely to re-emerge. The spread of the coronavirus and the intensification of the pandemic have increased the fear and panic of most Americans because of the physical limitations and financial hardship that COVID-19 poses.

To date, 42 states have issued "Order at Home," resulting in 95% of the US population facing restrictions that affect their daily lives. Innovative efforts to end the pandemic across the state have led to business closures. As a result, more than 30 million Americans have applied for unemployment since the outbreak of the coronavirus crisis.

Because the virus has been identified as foreign, for some people, their feelings have manifested themselves in xenophobia, prejudice, and violence against Asian-Americans. Due to the unprecedented impact of COVID-19 on people's lives, these negative perceptions and actions have received attention, and institutions such as the University of California, Berkeley, have even normalized these reactions. But racism and xenophobia are not a "natural" response to the threat of the virus; Instead, we speculate that the historical legacy of whiteness and citizenship generates these responses, with many likely interpreting Asian-Americans as foreigners and at higher risk of disease transmission. The FBI has warned of a potential increase in hate crimes against Asian Americans due to COVID-19, as "a segment of the American public will associate COVID-19 with China and Asian Americans".

3.7 Question 4: How does social media affect the rate of Asian hate crimes?

3.7.1 Data Set: Twitter and FBI

3.7.2 Method: Scatter Plot

The twitter data collected 30,977,757 tweets from June 2015–July 2018 containing at least one keyword pertaining to specific groups (Asians, Arabs, Blacks, Latinos, Whites). The data set characterized sentiment of each tweet (negative vs all other) and averaged at the state-level. These racial sentiment measures were merged with other measures based on: hate crime data from the FBI Uniform Crime Reporting Program. We want to show how the social media (based on twitter data) affect the rate of Asian-hate crimes.

The methods that we used is using scatter plot to show the relationship between anti-Asian tweets and anti-Asian crime rate at state level. Each points in the graph encodes each states, and the x-coordinate is the anti-Asian crime rate which is given by: the number of total Anti-Asian crime occurs in the state from June 2015-July 2018 over the number of total hate crime occurs in that state in the same period. And the y-coordinate is given by the percentage of sentiment tweets about Asian of all sentiment tweets.

3.7.3 Analysis

Figure 5. shows the result of the visualization. As we can see from the result, Tweets referencing Asians had a low proportion of negative sentiment (around 6% to 9%), respectively. The Delaware state has the highest proportion of negative sentiment tweets referencing to Asians, but the proportion of Asian crime in all hate crimes is still pretty low (around 2%). Actually, if we exclude some outliers in the data like the Delaware state, Hawaii state and Lowa state. Most state's Asian Hate crime rate lies between the interval from 0% to 6%, and the rate of anti-Asian tweets is in between 6% to 8%. they form a cluster, not a linear relationship. This suggests that there is no positive correlation between negative social media sentiment against Asians and racial crimes against Asians.

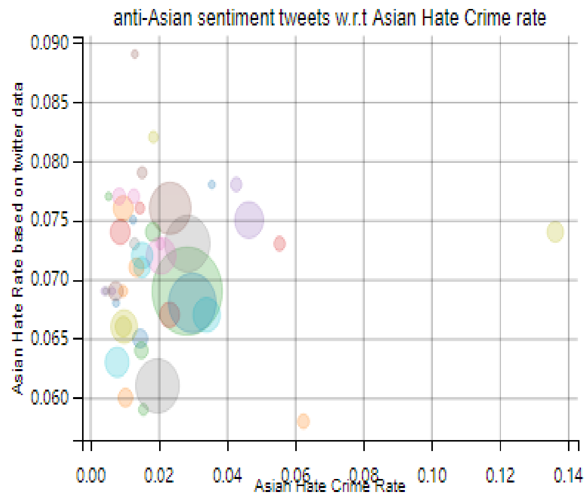

Figure 13: Tweets vs Anti-Asian crime 2015-2018

In fact, this result can be verified by an paper Called "Evaluating associations between area-level Twitter-expressed negative racial sentiment, hate crimes, and residents' racial prejudice in the United States" [12], the paper uses the same data sets that we have. And the result states that Negative sentiment referencing other subgroups (Black, Asians, and Arabs) were also not significantly associated with hate crimes against the respective subgroup. Negative sentiment tweets referencing racial minorities as a group was not related to any hate crimes, which also confirms that our conclusion is correct from the side.

3.8 Question 5: What's the relationship between the anti-Asian crime and the hate crime towards the other races in specific state/year? And is the anti-Asian crimes correlated any offender's race?

3.8.1 Data Set: FBI

3.8.2 Method

To solve this problem, we can implement a visualization that is a dynamic scatter plot. It can dynamically show the user selected variables of our FBI Hate crime data set as a axis of our dynamic scatter plot, the variables could either be the hate crime towards any specific race (in our case, Anti-Asian, Anti-Black or African American, Anti-White, Anti-American Indian or Alaska Native, Anti-Native Hawaiian or Other Pacific Islander) or the offender's race recorded in anti-Asian crime case.
The scatter plot is interactive, there is a slider bar provided for user to select a specific year, and each point of the scatter plot, encoding a state's data corresponding to user's selected variable. The scale of

the scatter plot could either be linear based of logarithm based. There will be two line charts next to the scatter plot, one of them is a line charts of how the x-variable changes for a specific state corresponding to the time changes, and the other one is also a line charts show that how the y-variable changes for a specific state corresponding to the time changes. When a user hover the mouse to any point of the scatter plot, the 2 line charts will dynamically change to the corresponding line charts of user's selected variable of the user's hovered state. Then we can see the result of the relationship of any of two variables of a certain state/year that in the data sets and see how those variables changes with respect to the time change.

3.8.3 Analysis

The visualization of this dynamic scatter plot allow us to show any of two relationship between hate crimes of any two race. for example. here is a figure that shows the state level anti-Asian hate crime vs. anti-Black hate crime of year 2012. As we can see from the result, it seems there is a positive correlation between the two variable since it seems the more anti-Asian crimes for a state has, the more anti-black crime cases for the state has.

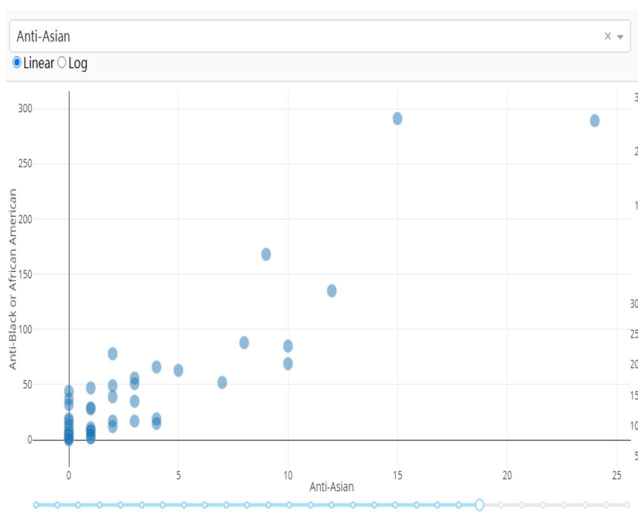

Figure 14: State level Anti-Asian crime vs Anti-Black crime of year 2012

And as we hover our mouse to any points in the scatter plot, as in this case, I hover the point of New Jersey, it shows the figures of line charts of the anti-Asian crime and anti-Black crime's data of New Jersey with respect to the time changes.

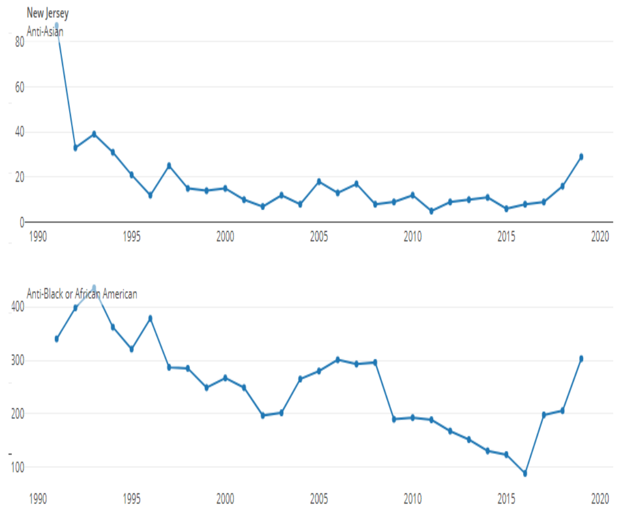

Figure 15: State level Anti-Asian crime vs Anti-Black crime of year 2012

And this visualization can also shows how the anti-Asian crimes correlated any offender's race for any time and any state. For example, here is a figure shows that the state level anti-Asian hate crime vs. the anti-Asian offender race to be black of year 2018. As we can see from the result, we can also see some linearly relationship between the two variable since it seems numbers of the black offenders are correlated to the number of anti-Asian crimes.

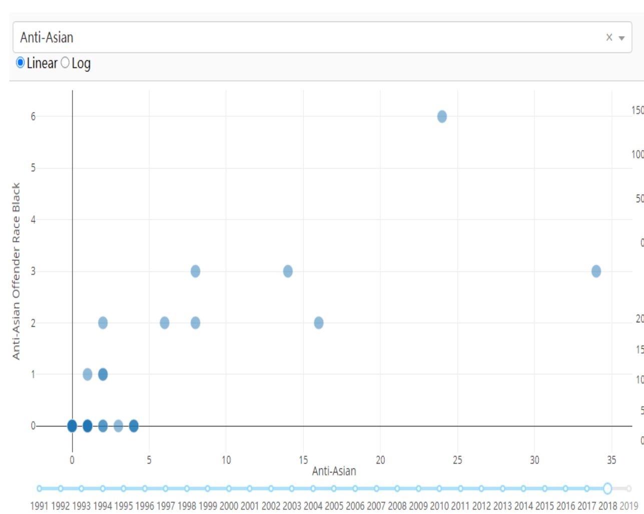

Figure 16: Anti-Asian crime vs Anti-Black crime of New Jersey with time changes

And if we select a specific state, for instance, The state California in this case, the lines charts further provides a evidence since the two variables are more likely have the same trends.

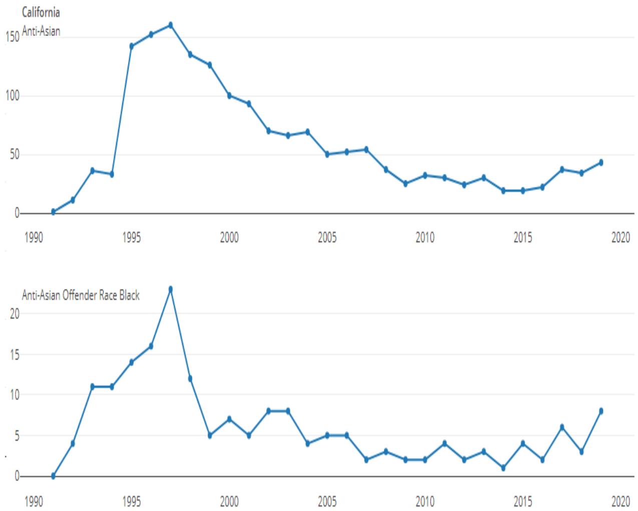

Figure 17: Anti-Asian crime vs Anti-Black crime of New Jersey with time changes

3.9 Question 6: Is the number of reported hate crimes throughout the year consistent? Does it spike during certain month?

- Here, we have used the heatmap to detect which month or the season has the highest hate crimes. Using the Vega-lite, we have created the visualization, where the more concentrated color(blue) represents the higher number of victims whereas lighter color represents the fewer victims meaning, less crime.X axis represents the Month and Y axis represents the year. Through the code we determined the hate crime incident occurred in each month of the year. As per the visualization, we can analyze that during the interval from March to October, the hate crime rates are usually higher. However, we could not see any trends, spikes, or clusters over certain seasons such as summer, spring, or winter. From this, we can conclude that several factors and incidents lead to hate crimes, not just a single factor. If just a single factor was responsible then there would be some cluster or concentrated color just over certain months or for some interval.

- Original dataset :`https://raw.githubusercontent.com/visualization-project-group/surprise-map-assignment/master/hate_crime.csv`

- This dataset provides data about the hate crimes from the date 1991 to 2019. It has broad data, meaning, it is not specific to Asian hate crimes. Thus, we extracted just the Anti-Asian incidents reported.

- Dataset after extraction:`https://raw.githubusercontent.com/Sambeg-Subedi/Visualization/main/Asianhatecrimes.csv`

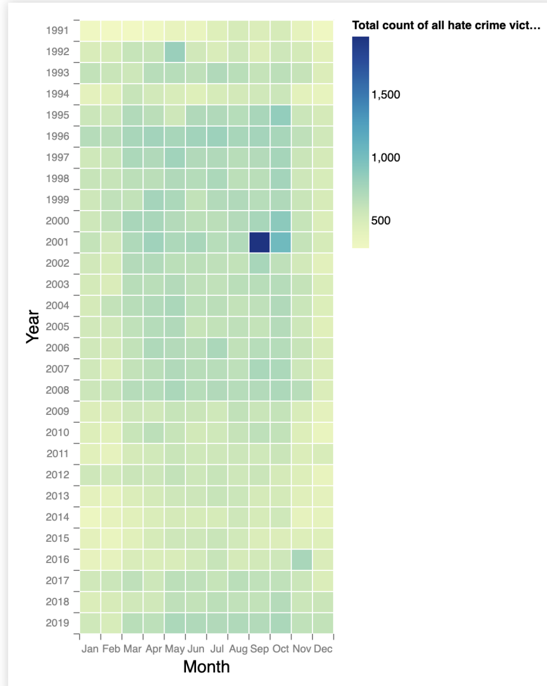

Figure 18: Fig: Heatmap visualization for all hate crimes

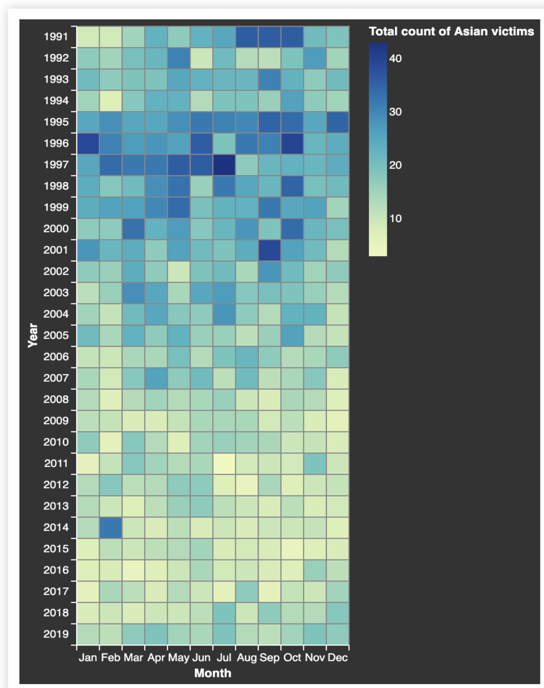

Figure 19: Fig: Heatmap visualization for Asian hate crimes

Similarly, for Asian hate crimes too, we couldn't see any spikes or any clusters on certain season or months, indicating several factors are responsible for Asian hate crimes. However we do see that from year 1991 to 2007 the number of Asian hate crime was higher and from 2007 to 2017 the crimes decrease but from 2017 the number of such incident again start increasing.

3.9.1 Number of Asian hate crimes reported till 2019 on each States

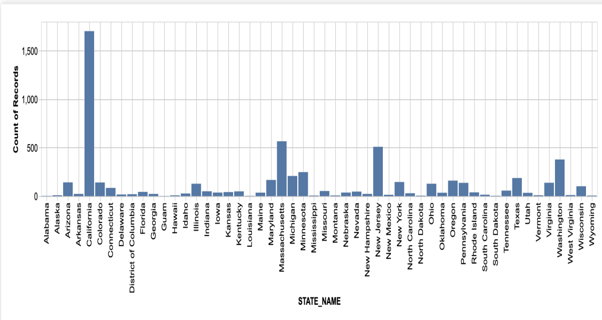

Figure 20: Fig: Visualization representing State name vs Total victim count

Here, in the visualization, we can see that California has recorded the highest number of Asian hate crimes during the time period 1991 to 2019 which is above 1500. Along with this, Massachusetts, New Jersey, and Washington have also recorded the Asian hate crime on above 400 Asian people.

3.9.2 Displaying the frequency of Asian hate crimes on Different location

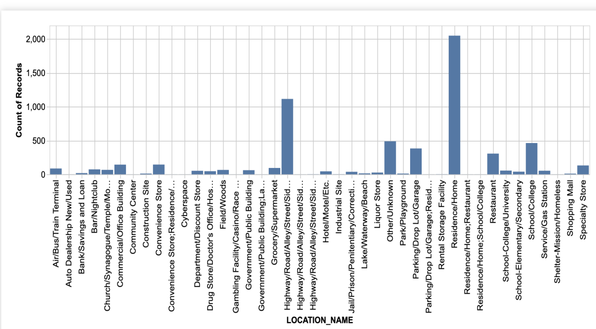

Figure 21: Fig: Visualization representing Location name vs Victim count

Asian hate crimes occur in different locations such as Grocery stores, schools, parks, food services, and so on. Here, in this visualization, we see that the maximum number of the incident had happened at the Resident/Home. After the resident or home, such incidents have more likely happened at Highway/Roads/Streets, Parking lot, Restaurant and Schools/ University.

3.9.3 Analyzing the types of offense during Asian hate crimes

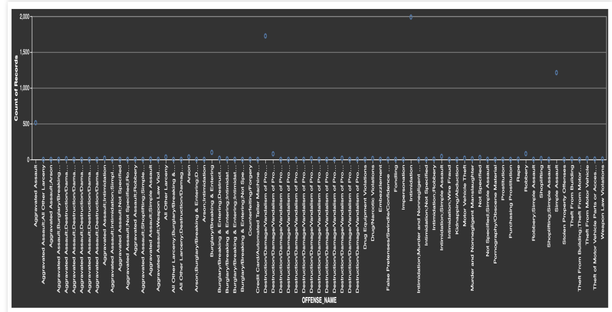

Figure 22: Fig: Visualization representing Offense name vs Sum of Victim count

There are different types of an offense such as Aggravated Assult, Burglary, Impersonation, Rape, Robbery and so on reported to the police. Visualizing that data, we can see that the most number of the victims reported Intimidation along with Destruction/ Damage/Vandalism and Simple assault.

4 Limitation and Scope

The data was old and we did not know whether the people who collected the data were biased or not. And because many people had received racial crimes they dared not report to the police. There was a frenzy of immigration between 1990 and 2000 and many of undocumented immigrants during that time. All of these are difficult to collect in the data and these factors make the data inaccurate. We also hard to find the accurate data of 2020 COVID-19 anti-Asian hate crime. So now the data is limited and the data we got from Twitter also have those questions.

5 Conclusion

Through this project, we were able to implement different forms of visualization for different data types. We used Vega-lite and python code in Jupyter Notebook to create the visualization. Through the visualization, we are able to determine regions and states with higher hate crimes, analyzed if there are any spikes on certain months or seasons, found out the most probable location and offense type for such hate crimes in the United States. We found that PCA is not that useful to describe Anti-Asian hate because there are not too many related attributes between each case of Anti-Asian crime. However Bayesian Surprise Map is a very good way to visualize the related change of Asian hate crime in a certain time. It uses color saturation to remind people which state has a higher ”surprise” value. Also the geographical information is encoded as a intuitive map to let viewers directly relate the ”surprise” value of each state.

We might have many guess that why anti-Asian rate change from high to low and from low to high from 1991 to 2019. However, there are only one guess for the anti-Asian hate crime increase from 2018 to 2020, that’s because the COVID-19.

through the Scatter Plot,I know the Asian hate crime rate over the last 20 years, why 1998 was the highest hate crime rate in the previous 20 years, and then it went down. In addition, because of the spread of epidemic rumors, Asian hate crimes have peaked in 10 years.

Many people may think that social media hate comments against Asians affects the rate of hate crimes against Asians, but in reality they are not necessarily related. The result states that Negative sentiment referencing subgroups of Asian were also not significantly associated with hate crimes against the respective subgroup. Negative sentiment tweets referencing racial minorities as a group was not related to any hate crimes.

References

[1] Hate Crime Statistics, Nov. 2020.

[2] Principal component analysis, May 2021. Page Version ID: 1023655415.

[3] S. M. . August 2016. 'Wow!' Again? SETI Mystery Signal Could Long Puzzle Astronomers.

[4] R. Fausset, N. Bogel-Burroughs, and M. Fazio. 8 Dead in Atlanta Spa Shootings, With Fears of Anti-Asian Bias. *The New York Times*, Mar. 2021.

[5] G. Grinstein, D. Keim, and M. Ward. Information visualization, visual data mining, and its application to drug design. IEEE Visualization Course #1 Notes, October 2002.

[6] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23, 2017. To appear. doi: 10.1109/TVCG.2016.2615308

[7] G. Kindlmann. Semi-automatic generation of transfer functions for direct volume rendering. Master's thesis, Cornell University, USA, 1999.

[8] Kitware, Inc. *The Visualization Toolkit User's Guide*, January 2003.

[9] M. Levoy. *Display of Surfaces from Volume Data*. PhD thesis, University of North Carolina at Chapel Hill, USA, 1989.

[10] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Computer Graphics*, 21(4):163–169, Aug. 1987. doi: 10.1145/37402.37422

[11] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, June 1995. doi: 10.1109/2945.468400

[12] T. T. Nguyen, D. Huang, E. K. Michaels, M. M. Glymour, A. M. Allen, and Q. C. Nguyen. Evaluating associations between area-level twitter-expressed negative racial sentiment, hate crimes, and residents' racial prejudice in the united states. *SSM - Population Health*, 13:100750, 2021. doi: 10.1016/j.ssmph.2021.100750

[13] G. M. Nielson and B. Hamann. The asymptotic decider: Removing the ambiguity in marching cubes. In *Proc. Visualization*, pp. 83–91. IEEE Computer Society, Los Alamitos, 1991. doi: 10.1109/VISUAL.1991.175782

[14] S. W. Pan, G. C. Shen, C. Liu, and J. H. Hsi. Coronavirus stigmatization and psychological distress among Asians in the United States. *Ethnicity & Health*, 26(1):110–125, Jan. 2021. doi: 10.1080/13557858.2020.1849570

[15] M. D. Swaine and M. D. Swaine. The Deepening U.S.-China Crisis: Origins and Solutions.

[16] H. Tessler, M. Choi, and G. Kao. The Anxiety of Being Asian American: Hate Crimes and Negative Biases During the COVID-19 Pandemic. *American Journal of Criminal Justice*, 45(4):636–646, Aug. 2020. doi: 10.1007/s12103-020-09541-5

[17] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, 2nd ed., 2004. doi: 10.1016/B978-155860819-1/50001-7

[18] G. Wyvill, C. McPheeters, and B. Wyvill. Data structure for *soft* objects. *The Visual Computer*, 2(4):227–234, Aug. 1986. doi: 10.1007/BF01900346