CSC110 Final Project: The Spread of Hate: How the Pandemic Circulated Anti-Asian Discrimination

Jay Lee, Andy Feng, Jamie Yi

Monday, November 5, 2021

Problem Description and Research Question

It is a commonly accepted theory that the COVID-19 virus originated in Wuhan, China. As this became common knowledge alongside the fact that the first human case of COVID-19 was also in Wuhan, some members of the western public began to unjustly blame people of Asian descent (AAPI) for the pandemic. In the US, the decline of public opinion was propagated by then-president Trump using the term 'Chinese Virus' on Twitter and other legislators using terms such as 'Wuhan Virus' (Brian, 2021, p. 6). These statements essentially endorsed public hate speech. They encouraged feelings of xenophobia and (false) liability towards AAPIs. The number of threats, racist abuse, and discrimination aimed at AAPIs saw an increase as well. Thus, anti-Asian sentiment grew and there was an increase in the number of hate crimes motivated against AAPIs (p. 6). Canada also saw an increase in hate crimes (p. 3); however, it was not motivated by any prominent Canadians. Canadian public officials did not endorse any hate speech and other such actions on social media or on news coverage. One of our goals is to draw a correlation between the rate of hate crimes occurring and prominent public officials encouraging the sentiments behind them.

All three of us are Asian-Canadians, so this problem domain is particularly relevant and interesting to us. When Canada and the US pride themselves on being multicultural, it matters to us that our demographic is being treated as equally as the others. We want to ask: **How has the frequency of anti-Asian hate crimes increased in response to the pandemic (and were there any secondary factors affecting the increase)? Does the increase of hate crimes in an area correlate to the proportion of the population that is AAPI? Does the increase of hate crimes in an area correlate to the political leaning of the state that the city is in?** Our dataset contains data from Canadian and American cities. However, we only plan on aggregating data for the American cities. We also look for trends by comparing the red (Republican) and blue (Democratic) states. Then if we find any patterns or correlations, we will see if they can accurately predict the increase in hate crimes in Canadian cities. This is keeping in mind that was less Canadian vocal hate directed at AAPIs.

Dataset Description

The first dataset we use is from on a report released by the Center for the Study of Hate and Extremism at California State University, San Bernardino in May 2021 (Levin, 2021); found on page 4 of the attached pdf on MarkUs. The report includes relevant data on Anti-Asian hate crimes we plan to extract and use for our project and analysis. The title of this data frame is "Anti-AAPI Hate Crime Data for Select U.S. Cities/U.S. Counties and Major Cities in Canada (2020-2019)". It includes 8 columns/variables: (US City Population / US County Population / Canada City Population, Total Hate Crimes 2019, Total Hate Crimes 2020, % Change for Total Hate Crimes 2019-2020, % of Population – AAPI, Change Anti-Asian Hate Crimes, 2019 Anti-Asian, 2020 Anti-Asian) and each row/observation in the data frame represents a major city or county in the US or a major city in Canada. The 24 US cities shown in the table are the 24 most populous cities in the country. To convert the table, in the pdf to a csv, we just copied and pasted the text into a csv file, and then used the find/replace tool to replace every whitespace with a comma. This file is named `hate_crime_data.csv`. In the csv file we have also omitted all the Canadian cities. The original pdf can be retieved from `https://www.csusb.edu/sites/default/files/Report%20to%20the%20Nation%20-%20Anti-Asian%20Hate%202020%20Final%20Draft%20-%20As%20of%20Apr%2030%202021%206%20PM%20corrected.pdf`

The second dataset we use is from `https://www.politico.com/2020-election/results/president/`. It's a website that shows the state results for the 2020 US presidential election. We used it so we could figure out the colour

(majority political leaning) of each state. We were unable to find any premade .csv files that contained this data so we had to make one of our own, using the data on the website. Check `state_colour_data.csv`. The colours of the states are also represented as 'royalblue' and 'crimson' as opposed to 'blue' and 'red'. This is becuase 'royalblue' and 'crimson' are preset colours for the graphing module we use, so it's easier to represent the data as those preset colours so when reading the data we can just pass it to the graphing module without change. We will be only be using the rows with states that are present in `hate_crime_data.csv`.

The third dataset we use is from `https://simplemaps.com/data/us-cities`. The free .csv file downloaded from the site contains a lot of information and columns for each US city on it (over 28000 of them). We are interested in only the latitude and longitude values for the cities that are listed in `hate_crime_data.csv`. So we actually don't end up using a large portion of the downloaded .csv file. It is named `uscities.csv`.

Computational Overview

Our bubble map marks each of the 24 US cities present in `hate_crime_data.csv`. When creating it, we used the pandas.Dataframe object. It's two-dimensional, size-mutable, tabular data. It's like a 2D array in that it has columns and rows; and both entire columns OR entire rows can be accessed with one statement. Individual items can be accessed with the syntax `sample_dataframe[column_name][row_number]`, which is useful for mutation (entire columns can be read with `sample_dataframe[column_name]` and entire rows can be read with `sample_dataframe.iloc[row_number]`). When a csv is read and the data is moved to a Dataframe (using `pandas.read_csv()`), the columns become matching to the csv headers and the rows are numeric. From there we may add new columns to the Dataframe (which we do). We add the new columns `lat` (latitude), `lon` (longitude), (colour), and `percentage`. For `lat`, `lon`, and `colour`, we read other csv files to locate and extract data from them (check the in line comments in `process_hate_crime_csv.py` for more detail). `percentage` represents the relative size of the bubble we want. The size of the bubble is proportionate to the percent increase in anti AAPI hate crimes from 2019-2020. An increase of 0% or a decrease of any percent (i.e., a -% increase) is represented as a very small bubble size so the city remains visible on the map (they are very small so you might have to squint). Hovering over a bubble will also show you the exact percentage increase in hate crimes so there should be no confusion. One thing to note about size: in cases where there were 0 anti AAPI hate crimes in 2019 and some other positive number of crimes in 2021, there was no best way to represent the increase. Since we did want to differentiate between say, a change from 0 to 1 and a change from 0 to 6, our strategy was to multiply the number of crimes in 2020 by 100 and call that the percentage increase. So a change from 0 to 1 would be represented as 100% and a change from 0 to 6 would be represented as 600%. We recognize that this isn't ideal but making the bubble map like this turned out to be more informative than basing the circle sizes on only the number increase in cases (delta-change in cases vs. percentage change in cases). We have included the latter for reference, the statement to call it should be commented in `main.py`.

To graph this Dataframe, first it was split into two separate Dataframes, each Dataframe corresponding to red (republican), or blue (democrat) with regards to the city's state's political leaning. Then we create a `plotly.graph_objects.Figure()` object to display our data to. We plot the data contained within the two Dataframes using `Figure.add_trace()` (by the Plotly website: "A trace is just the name we give a collection of data and the specifications of which we want that data plotted."). We call `Figure.add_trace()` twice, once for each Dataframe to plot (the code that calls `Figure.add_trace()` does have in-line comments so feel free to check those out in `bubble_map.py` for additional details).

The final bubble map can be interacted with in two ways:
1. You can hover over the bubbles to see the name of the city, the percent increase in anti AAPI hate crimes between 2019 and 2020, and the political leaning of its state.
2. You can click on the legend on the right side that differentiates between Republican cities and Democratic cities to toggle them.

We implemented a tree map to visualize a part-to-whole relationship amongst a large number of categories, in a hierarchical structure by splitting up a large rectangular area into smaller rectangles for subcategories. For our purposes, we wanted to visualize the relationship between the US cities, the states they belong to, and the political party each state represents. To create the tree map, we used the plotly library, more specifically the plotly.express module. The data frame is created using the helper function `process_hate_crime_csv` which is imported from `process_hate_crime_csv.py`. The helper function uses the pandas library to create the data frame and returns it after some manipulations. We then added a new column to the data frame called `party` which represents the political

leaning of each city in the data frame (Republican if the colour is red and Democratic if the colour is blue). Once the data frame has all the information we need, all we had to do was use the `treemap()` function in the plotly express module to generate the tree map. Five arguments were passed into this function. The first argument is the data frame we are creating the tree map from, the second argument is the path of the tree map, which takes in a list of column names in our data frame, the order of these column names is based on the hierarchy we want to display (i.e, party before the state before the city). The third argument is the values we want to display in the label, in this case we chose the total percentage increase of anti AAPI hate crimes from 2019-2020 as the value to display in the label. The fourth argument is the colour, and we base it on the colour column in our data frame (either red for republican or blue for democratic). The last argument is the title for the tree map. Finally, `update_layout()` function sets the margin for the tree map when it is displayed and the `show()` function displays the resulting tree map in the browser. The tree map displayed in the browser is interactive and a user can hover above and click different areas to zoom in and out of rectangles and read the labels. The larger the rectangles, the higher the total percentage of anti AAPI hate crimes from 2019-2020 in that area. As a final note, the interactive components were already coded into the plotly express module so we did not have to include additional code.

We have also visualized another sets of data using a bar graph, and a scatter plot. As the other visualizations we have done above, these two also read the csv file data using Panda library's function and converts the read data into Panda dataframe. In the `draw_anti_asian_comparison()`, the plotly.express's bar figure model takes the Panda dataframe and extracts the columns "US State", "2019 Anti-Asian", "2020 Anti-Asian". Plotly then uses "US State" column to show a bar graph comparison of the number of Anti Asian hate crimes in 2019 vs the number of Anti-Asian hate crimes in 2020. The scatter plot takes a similar procedure. `draw_aapi_vs_anti_asian_hatecrime()` also reads from the csv file using the Panda library and parses the read data to a dataframe. This time, plotly's `scatter` model comes in handy. However, before forming the scatter plot figure, our csv data file's "% of Population-AAPI" column contains "%" character at the end. That prevents plotly.express from automatically sorting the data since the columns are parsed as strings, not floats. So I had to strip out "%" characters and cast them into float datatype. We added a new column to the dataframe with those fixed data. Then using the % of the Asian population as our x-axis, we graphed two scatter points graph using the numbers of Asian hate crimes in 2019 and 2020 in the y-axis.

Running the Program

All the datasets you need should be submitted under MarkUs. If you wish to look at/download the originals, check the links in the above section. In order to fun our program properly, please create a subfolder called **data** in the same directory as the other files like **main.py**. It should look something like this:

- ⌄ 📁 final_copy
  - ⌄ 📁 data
    - 📄 hate_crime_data.csv
    - 📄 state_colour_data.csv
    - 📄 uscities.csv
  - 🐍 bar_graph.py
  - 🐍 bubble_map.py
  - 🐍 bubble_map_delta_increase.py
  - 🐍 main.py
  - 🐍 process_hate_crime_csv.py
  - 🐍 scatter_plot.py
  - 🐍 tree_map.py

Then make sure to download all the packages required under 'requirements.txt'. After, you can directly run main.py to display all four types of graphs. If you uncomment the interactive terminal GUI launcher in main.py, you can expect a GUI screen welcoming you, just like the screenshot below.

```
*-----------------------------------------------------*
| |      Welcome to our Final CSC110 Project!        |
|*=================================================*
| |      Here, you can use the terminal I/O to       |
| |      interactively display our visual data       |
|*=================================================*
| 1) Display a bar graph comparing Asian hate crimes  |
|  in 2019 and 2020                                   |
| 2) Display a scatter plot showing the relationship  |
|   between the % of AAPI and hate crime              |
| 3) Display a treemap that shows the number of  Asian |
|  hate crimes to the political pref of the cities    |
| 4) Display a bubble map of Asian hate crimes using a |
|  geographical map of US as the board (based on      |
|  a percentage increase)                             |
| 5) Display a bubble map of Asian hate crimes using a |
|  geographical map of US as the board (based on the  |
|  net increase in hate crimes)                       |
| 6) Quit the program                                 |
*-----------------------------------------------------*

YOUR INPUT(1-6):
>?
```

Just as the screenshot suggests, the user can enter (1-5) to view the according visual representation of data. Enter 6 to exit out of the program.

Running 1:

The graph displayed is quite interactive. You can drag an area to zoom into a specific section of the bar graph. On the top right corner, you can also click either 2019 or 2020 to toggle the view of each bars.

# Changes between Plan and Final Submission

For the data visualization, we have done as we planned. We have organized our datas into a csv format and aggregated them into program-readable Panda dataframe. We have also used plotly.express effectively to make 4 different types of data visualization as we planned. From the data visualization, we figured that for the most parts, the Asian hate crime numbers in most states have increased from 2019 to 2020. However, since we could not find any correlation between the relationship between political preference and the number of Asian hate crimes, and the percentage of Asian population and the number of Asian hate crimes, we could not form a prediction model as we originally planned. If you look at the scatter plot for example, both the dots from 2019 and 2020 are all over the place. There is no strong correlation between the x-axis and the y-axis. As the % of the Asian population increases, the number of Asian hate crimes sometimes increased, and sometimes decreased. This will be touched more in depth in the discussion section.

Expected Result of the Program
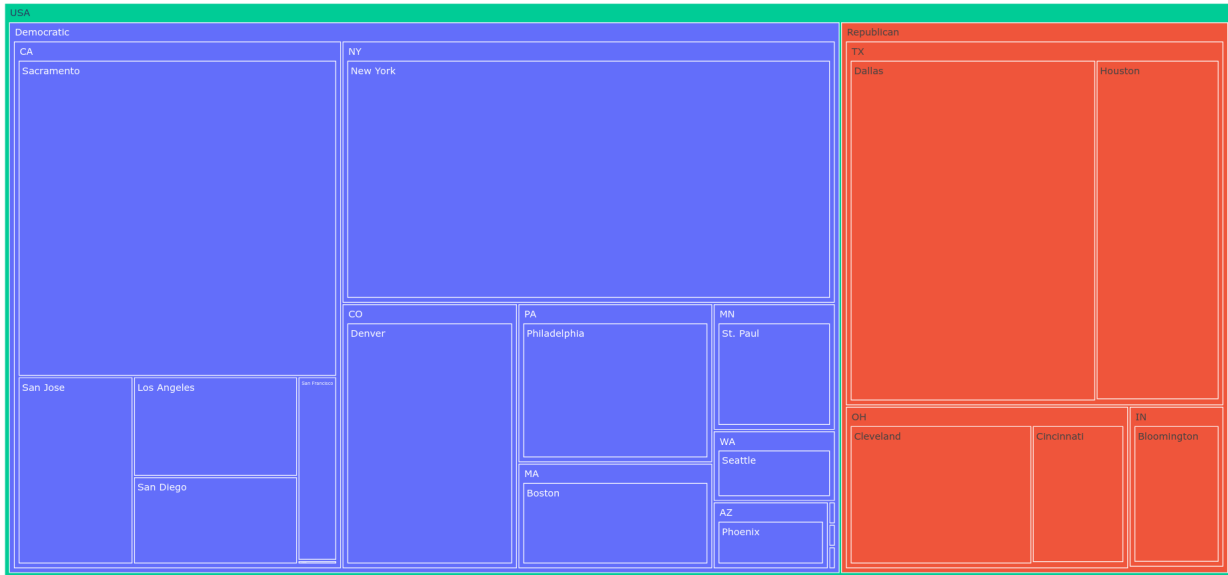
When you enter 1 to display the bar graph:

2019 Anti-Asian Hate Crimes vs 2020 Anti-Asian Hate Crimes
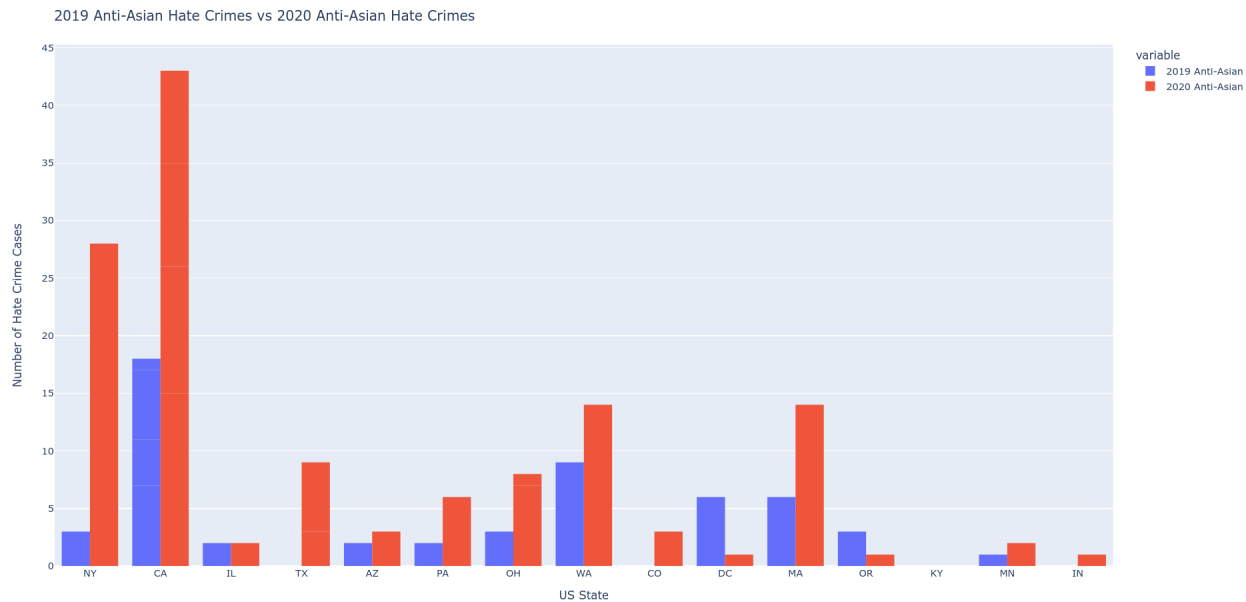
When you enter 2 to display the scatter plot:

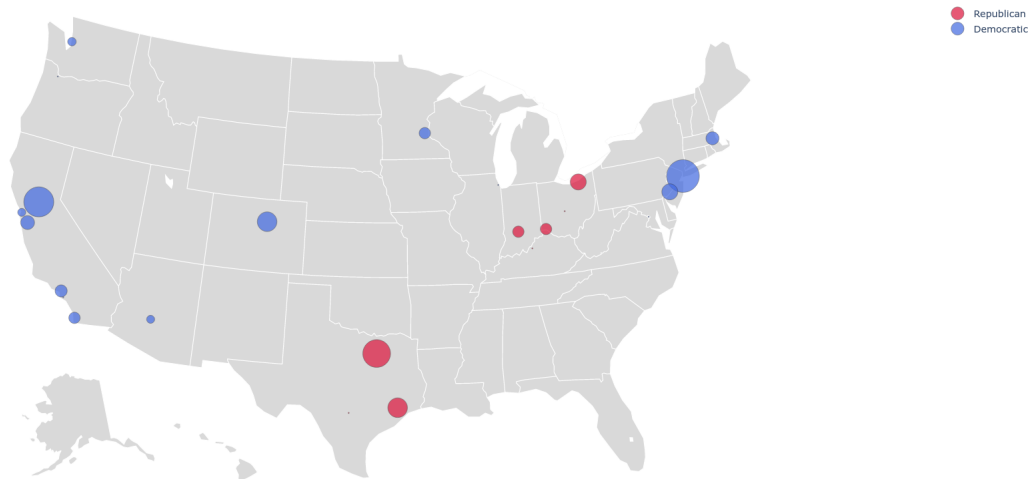Asian Population % to Anti-Asian Hate Crime Cases in 2019
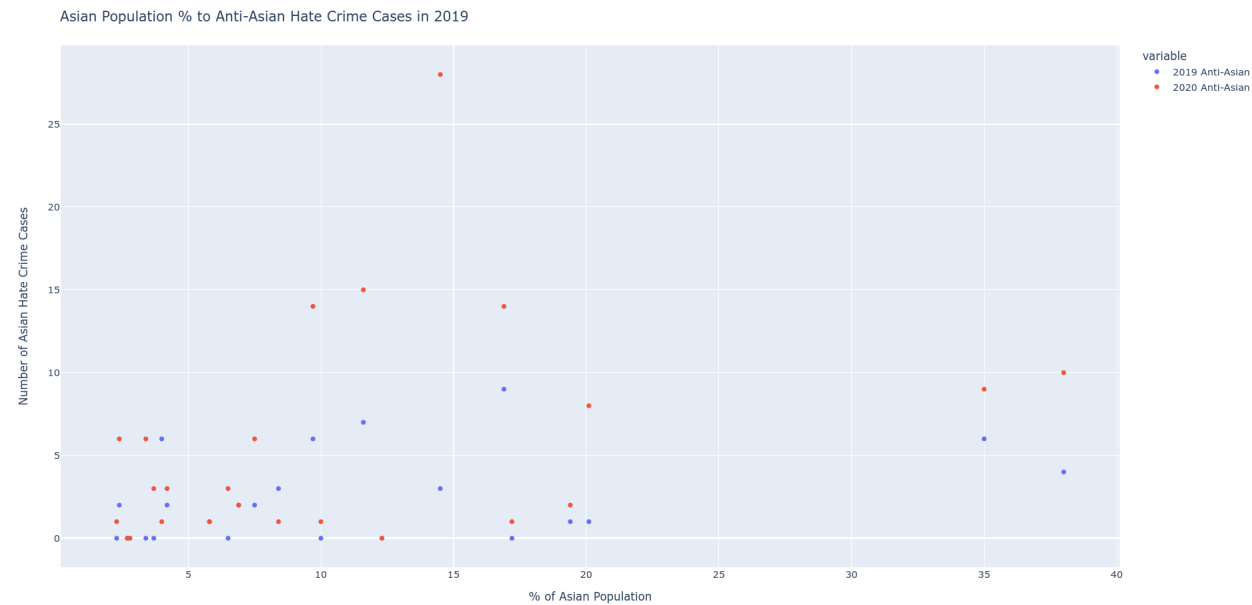
When you enter 3 to display the tree map:

Total Percentage Increase of AAPI Hate Crimes in Major US Cities, 2019-2020
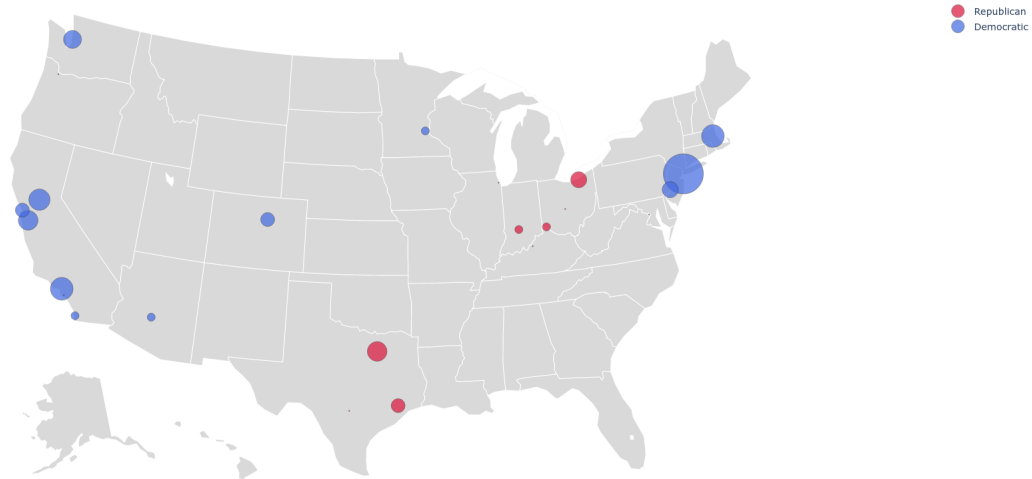

When you enter 4 to display the bubble map:

Percentage Increase in anti-AAPI Hate Crimes in American Cities
(Click legend to toggle traces)

When you enter 5 to display the bubble map that uses flat increase in the numbers:

Total Increase in anti-AAPI Hate Crimes in American Cities
(Click legend to toggle traces)

Republican
Democratic

Discussion

There is no correlation between % AAPI population to the % increase in anti AAPI hate crimes. This is not what we expected at the beginning of the project. We expected that upto a certain % of Asian population, there would be an increase of Asian hate crimes as there are more Asian population, but a decrease after that point as the majority of the population becomes Asian. Our scatter plots show that that was not the case. There is hate crime regardless of the % of the Asian population. One thing to note is that if there was a sufficient data between 20% to 35% Asian populated cities that corresponded with the trend, then we might had been able to conclude that there exists such trend that we predicted at the beginning of our project. Sadly, due to the data having no correlation, we cannot form a mathematical model to predict the situation in Canada.

The average increase in hate crimes between republican cities and Democratic cities is about the same, debunking a common idea that Democratic cities are safer. This breaks the modern stereotype of how so-called 'Trump supporters' are more racist than Democrat supporters. Regardless of political preferences, if a person does evil deeds, they will do evil deeds. This can be checked in the bubble map as the red bubbles(Republicans) and the blue bubbles(Democratic) have roughly the same size. Originally, we intended to use the flat increase in the number of Asian hate crimes to map the bubble map, but we soon realized that doing so produces inaccurate representation. Therefore we ended up using percentage change that is proportional to the change, which is more meaningful.

There has been an evident overall increase of Asian hate crimes from 2019 to 2020. As we have expected, the bar graph shows that compared to Asian hate crimes in 2019, there has been an increase in the number in 2020. For example in NY, there was an increase of 25 more Asian hate crime cases. Most of the cities have increased in the numbers of Asian hate crimes from 2019 to 2020. Bar graph is an excellent way to show the differences as we can compare the data visually side by side. Note that this bar graph is just to show you the flat increase in the data. The increase relative to the population will be dealt with in other graphs. The stats show that the increase in the number of Asian hate crime cases post-COVID outbreak is true.

Based on the tree map, it would appear that the two major cities with the largest increase in anti AAPI hate crimes from 2019 to 2020 are both Democratic cities. The two cities are Sacramento at 700% and New York at 833%, since they take up the largest rectangular areas. Although we do not have data on all the cities in the US, this result is quite surprising to see as we expected the Republican cities to have the largest percentage increase because they are known to share far greater Anti-Asian sentiments than their Democratic counter parts, especially under the influence of then president Donald Trump. Trump was known to spread Anti-Asian hate through social media platforms such as twitter, by using terms such as "Chinese Virus" and "Kung Flu" to describe the coronavirus. This result further demonstrates that the stereotype of Republican cities being more hostile to Asian people is not exactly

accurate and may need to be reassessed.

References

Report to the Nation: Anti-Asian Prejudice Hate Crime - Data Tables Levin, B. (2021). (rep.). Report to the Nation: Anti-Asian Prejudice Hate Crime (pp. 1–34). San Bernardino, California: Center for the Study of Hate and Extremism.

Live election results: The 2020 presidential race. POLITICO. (2021, January 6). Retrieved December 11, 2021, from https://www.politico.com/2020-election/results/president/.

United States Cities Database. simplemaps. (n.d.). Retrieved December 11, 2021, from https://simplemaps.com/data/us-cities.