Nov. 13:

Group Members:

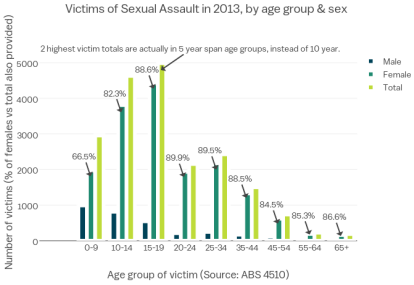
- ➢ Riyana Basu, Michelle Chan, Valerie Hernandez, Alicia Kim

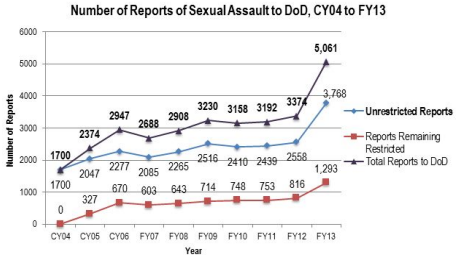
Project Title: Society and Sexual Assault

Project Topic:

For our project, we will be analyzing sexual assault through several different lenses, using both pre-gathered data as well as data gathered by scraping Twitter. We will be investigating factors such as the locations, industries/occupations and ethnicities where the magnitude of cases being reported are coming from, as well as where the #metoo Twitter tag is most prominently coming from.

Potential Design:





Task Division and Progress:

Riyana:

Riyana will look at datasets of various locations across the US to compare population or socio-economic standing of the locations in relation to the density of #metoo mentions. She will also use Wordle or Wordcloud to visualize the most common words used in the #metoo hashtag.

Current Status & Future Work: My initial exploration consisted of searching for tweets with the topic query “sexual assault.” I then wanted to connect my exploration to socioeconomic status, and thus used the geolocation parameter to look at tweets from places of varying socioeconomic diversity. The ranking was created by WalletHub. Wordcloud came in handy to better understand how the tweets from varying regions differed. So far, I have only looked at popular tweets because I thought these tweets would provide insight into how people’s opinions

differ based on their socioeconomic status. Because the crux of the Twitter-related parts of this project is the #metoo tag, the remainder of my work will be centric on repeating the described work with the tag instead of just with the query “sexual assault.” I would like to generate visualizations that paint an easy to understand picture, and thus will likely make a heat map of sorts that is related to the #metoo tag. Furthermore, I would like to see if I can explore not just correlation, but also causation. This will likely be much harder to prove, but some use of statistics should allow me to reach some conclusions.

Michelle:
Michelle will be responsible for scraping data from Twitter related to the #metoo tag in order to further explore the density of the responses across different areas in the US. She will also use Tableau or some other visualization tool to visualize the scraped data.

Current Status & Future Work: Scraped Twitter data with the query #meToo as well as the location corresponding to each tweet, generating a map to visualize where the majority of the responses fall. Would like to clean the data more and account for more responses so that the map is more filled out. Compare the self-generated map with other maps displaying a certain variable such as political standing effectively. Look into other details such as number of retweets or number of followers that can be turned into visualizations.

Valerie:
Valerie will look into how household income factors into reports of sexual assault. She will also create visualizations for the overall trend of sexual assault over time, including data about different ethnicities affecting sexual assault report rate.

Current Status & Future Work: I currently have information on household income from the Crime Victimization Survey, and the data is organized in Jupyter. I still need to arrange and clean some of the data for visualizations.

Alicia:
Alicia will be responsible for gathering the data from the Crime Victimization Survey to visualize what industries are reporting the most sexual assault cases. She will also use that data set to understand whether marital status affects likelihood of reporting sexual violence. This analysis will be done in Jupyter.

Current Status & Future Work: I have gathered, cleaned, and visualized the data from the Crime Victimization Survey. I specifically have visualized married versus unmarried reported sexual assault cases. The data I have lacks information about specific industries reporting sexual assault cases. Instead, I have visualized the overall trend related to the amount of sexual assault cases reported, and I will be looking at the spikes and falls in this trend to see if I can find associations between certain events and the times that these changes in sexual assault cases occurred. Overall, the trend from the data I collected shows a decrease in sexual assault cases reported since 1993. In case I am not able to find events/research that helps explain the

spikes and falls, I will turn to the final piece of data I was able to gather which concerns the amount of cases of sexual assault reported in urban/suburban/rural areas .

Challenges Faced:
It was hard to scrape Twitter data because of the limit they imposed, but this was fixed by pickling data so that it could be loaded anytime. In addition, many users do not have their geocode enabled so it was hard to find the most specific and accurate location for their tweet.

Remaining Timeline:

11/20: have individual components done

11/27: have everything uploaded to website