Group Members:
- ➢ Riyana Basu
- ➢ Michelle Chan
- ➢ Valerie Hernandez
- ➢ Alicia Kim

Project Proposal:

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.

Dataset:

1) National Crime Victimization Survey: a dataset gathered by the Bureau of Justice Statistics going on from 1973 to 2015 regarding criminal victimization in the workforce and the different components of it.
2) Twitter API: self-scraped data from Twitter
3) US Census Data: data on various cities across the US

Motivation:

We are motivated by the recent Harvey Weinstein case responses and trending Twitter tag, #metoo, which Twitter claims will help exemplify the "magnitude of sexual assault". We believe that we can utilize the #metoo Twitter tag to even better exemplify the state of sexual assault cases in the country. What specific industries are these cases most prevalent? Is there an association between who has reported sexual assault cases and ethnicity, occupation, socioeconomic status? Specifically with the #metoo tag, where are the tags coming from? What are the demographics of the tag uses?

Expected visualization:

In the end, we will be presenting all the visualizations and data in a self-generated website. Various visualizations for data will include but are not limited to infographics, maps, line charts,
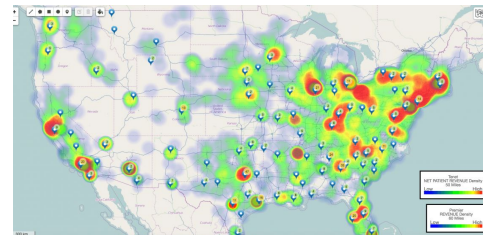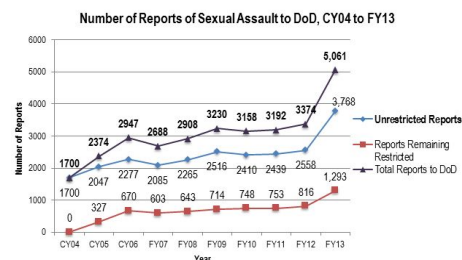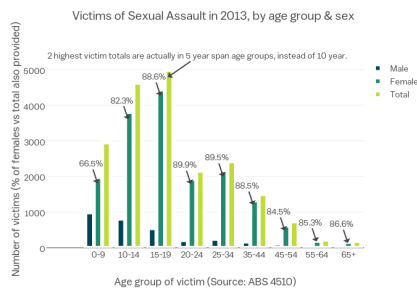
and bar graphs. For example, the density of #metoo Twitter mentions can be visualized in a heat map.

Potential Design:

Tools:

We will be using a mixture of R, Tableau, World, Wordcloud, and Python to perform the visualizations. For the website, we will be using HTML, CSS, and JS.

Task Division:

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.

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.

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.

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. She will also be responsible for gathering all the final presentation data into a self-generated website.

Related Work:

Twitter analysis on emojis on election day; uses Twitter API & visualizes scraped data

Text analysis of tweets; uses Twitter API & analyzes word frequencies

Visualizations of sexual assault in campuses; looks at our topic but in a different setting

Proposed Timeline:

11/13: be at least halfway done with individual components

11/20: have individual components done

11/27: have everything uploaded to website