Twitter Malicious Speech and Racially Motivated Hate Crimes in California

Directions: Submit your proposal in your project repository. Make sure it is labeled clearly (it helps if you put a link in your README.md). The proposal should be .pdf or Jupyter Notebook (.ipynb). 1-2 pages. Only one person needs to submit. Due Sunday night.

Topic

1. What questions does your topic address or what problems does your topic solve? Why and to whom are these meaningful?

With the rise of social media, the translation of malicious speech from internet discourse into tangible action has been an increasingly tangible threat. This can be seen in the 2017 Charlottesville Unite the Right Rally and January 6th 2021 attack on the whitehouse, both organized through social media.

We are interested in the relationship between social media and smaller scale hate crimes in the state of California. Identifying trends in hateful rhetoric online and crime could be meaningful in promoting national security by determining risk of violence against certain racial and ethnic minorities.

We hope to answer a few key questions through our project. What minority groups are receiving the most racially motivated offensive terms/slurs? What racial minority groups are receiving malicious threats that indicate "intent"? These questions will be answered -with some additional linguistic research- using Twitter data.

We will also answer some additional questions using the Hate Crime and California Racial Demographics data sets. What California minority groups are most likely to be the victim of violent crime relative to their population size?

Our final questions will be answered using all three data sets. Is there a relationship between which groups are receiving offensive tweets and which groups are being targeted by hate crimes? Does this relationship to hate crime rate change when a tweet's "intent" is considered?

2. What data source(s) will your team use? Briefly describe each data source and explain how you think you will use it. Provide a link for each data source
    a. Hate crimes in California
    Variables of interest:
    "MostSeriousUcrType" Type of attack, ie property crime, violent crime, etc.
    "MostSeriousBias" Specific biases, ie anti-black, anti-gay male, anti-jewish etc.
    "MostSeriousBiasType" Groupings of biases, ie racial/ethnic, religious, sexual orientation, gender, disability, etc.

b. Racial/ethnic Demographics in California
   Variables of interest:
   "population of one race"
   "population of two race"
c. Twitter offensive speech (Nico will find)
   Variables of interest:

The California Department of Justice Open Access data sets, from which we extracted the hate crime data, are publicly available and credible. The US census data, from which we extracted the demographics data, is also a very credible source.

3. What's challenging about your topic? Is a 6-week project long enough to explore the topic reasonably well?

We will have to do some background research on linguistics in order to best analyze the text content of twitter data, which will be a unique challenge. Working with the twitter data will also require quite a bit of natural language processing.

It could be challenging for us to draw conclusions while acknowledging the limitations of working with government data. Some of these limitations include the fact that (1) Demographic data is not collected for all ethnic subgroups,(2) Hate crimes aren't always reported, and (3) hate crimes data does not provide particularly comprehensive information about the victims of hate crimes, only the perceived quality that is being targeted by the perpetrator.

Six weeks should be sufficient time to do background research on linguistics, to clean and analyze these datasets, produce a written report, and practice a presentation of preliminary results. We have started in a strong place, with clearly defined questions and with group members holding a strong contextual understanding of linguistics and politics.

4. What skills from STA 141A-B or other statistics classes do you expect to use on the project?

- Natural language processing (lecture 6-1)
- Confidence intervals (introductory statistics classes)
- Visualizations (lecure 3-1)
- Indexing, Data frames (lecture 2-2)
- Reading, inspecting, aggregating, and grouping data (lecture 2-2)

NOTES:
Hate Speech data sets:
https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset
https://www.kaggle.com/vkrahul/twitter-hate-speech

Articles:
https://www.researchgate.net/publication/351431715_The_Lexicogrammar_of_Hate_Speech_The_Case_of_Comments_Responding_to_New_Zealand_Mass_Shooting_Online_News_Video

1. What questions does your topic address or what problems does your topic solve? Why and to whom are these meaningful?

Hate crimes, of "crime[s] motivated by bias against race, color, religion, national origin, sexual orietnation, gender, gender identity, or disability" (The United States Department of Justice) have been rising in the United States (Kaplan, 2006; Levin & Reitzel, 2018; Tessler et al., 2020). The goal of this study is to better understand how the nature and target of hate crimes have changed over time by conducting an exploratory data analysis on hate crime data and by employing natural language processing techniques on hateful rhetoric on social media.

To narrow the scope of our analysis of hate crimes, we will be investigating hate crimes in California from 2001 to 2020 (California Department of Justice). Our questions are as follows: (1) How has the proportion of hate crimes relative to the population changed over time? (2) How has the geographical distribution of hate crimes shifted over time in California?

We will also conduct a linguistic analysis of hateful and non-hateful social media posts. Racist and prejudiced rhetoric proliferates on social media platforms (Awan, 2014; Relia et al. 2019), and some scholars have even argued that a higher usage of Facebook is causally linked to hate crimes against refugees (Müller & Schwarz, 2021). Therefore, an analysis of hateful speech online is a fitting complement to our study of hate crime data. In order to gain greater insight into the psychology of the Twitter users, we will conduct a semantic and lexical analysis of the posts that tests two hypotheses.

The first hypothesis is that the definite article "the" should appear before group names, such as "blacks", more often in the hateful posts than in the non-hateful posts (e.g., if a non-hateful poster writes "Blacks should be treated with respect", a hateful poster is predicted to write "The blacks should not be treated with respect"). Acton (2019) argues that when speakers employ the definite article in grammatically unnecessary ways (e.g., "Blacks should not be treated with respect" is still grammatical, so "the" is not necessary in the example above), the speakers signal that they are not part of the group they are referring to, and this is the behavior we expect from posters of tweets that denigrate these groups.

The second hypothesis is that among the pronouns that are employed, hateful posts should have a higher proportion of third person pronouns, such as "they", compared to non-hateful posts. This follows from research suggesting that hate speech tends to be characterized by "othering", or the positioning of oneself or one's own social group against another individual or group (Meddaugh & Kay, 2009). If the writers of the hateful posts employ a higher proportion of third person pronouns, this may indicate that they are engaging in more "othering" behavior, as others who have studied hate speech have suggested (ElSherief et al., 2018).

To tie the hate crime data and the social media data together, our final question will be as follows: Has the proportion of hate crimes relative to the population in California increased as social media usage has increased? To simplify this question, we will compare the proportion of hate crimes from before 2011 to that after 2011 as 2011 was when half of American adults stated that they used at least one social media site (Pew Research Center). This analysis will help to test whether or not there is a relationship between hate crimes and social media usage.

By analyzing hate crimes in California temporally and conducting a semantic and lexical analysis of hateful speech on the Internet, this study has the potential to inform policymakers as to trends in crime over time and to benefit linguists interested in learning more about hate speech.

2. What data source(s) will your team use? Briefly describe each data source and explain how you think you will use it. Provide a link for each data source.

For the analysis of hate crimes in California, we will be employing the "Hate Crime" data set from the California Department of Justice and merging this with a data set from the United States Census Bureau on the demographics of California. Each row in the hate crime data set corresponds to every reported hate crime from 2001 to 2020 in California. The year and the month of each crime are included, which will allow for a temporal analysis. Moreover, the county where each crime occurred is provided, meaning that a geographical analysis can be conducted. Joining the hate crimes data set to the data set from the census, which contains information about the population of United States counties, will allow us to assess the number of hate crimes that occurred in each county relative to the population.

Two data sets with social media posts will be employed for the semantic and lexical analysis. The first data set consists of posts from the white supremacist forum Stormfront (de Gibert et al., 2018), and the second data set was created by searching for Twitter posts (Mollas et al., 2021). Data from different social media platforms was selected to make the sample of posts more diverse in our study. For both data sets, each row corresponds to a single post and they have been tagged by the authors as either hateful (1) or non-hateful (0).

3. What's challenging about your topic? Is a 6-week project long enough to explore the topic reasonably well?

One challenge will be learning how to work with geospatial data in Python, which we have not covered in this class. Rather than writing a paragraph detailing how many crimes have occurred in each county, a more effective way to visualize how trends in hate crime have changed in California counties over time is to create a colored map. To do this, however, we will have to learn how to employ packages like GeoPandas to manipulate geospatial data.

Another challenge will be for us to draw conclusions while acknowledging the limitations of working with government data. Some of these limitations include the fact that (1) demographic data is not collected for all ethnic subgroups, (2) hate crimes are not always reported, and (3) hate crime data sets do not provide particularly comprehensive information about the victims of hate crimes.

Six weeks should be enough time to clean and analyze these datasets, produce a written report, and practice a presentation of preliminary results. We have started in a strong place, with clearly defined questions and with group members holding a strong contextual understanding of linguistics and politics.

4. What skills from STA 141A-B or other statistics classes do you expect to use on the project?

We are planning on referring to the lectures during Weeks 2 and 3 to help us with cleaning and visualizing the hate crimes data set. Moreover, we will employ the natural language processing techniques that we learned in Week 6 for the social media data sets.