Sentiment analysis on Roe v. Wade using Social Media data

Team 15: Mansi Bhandari, Varad Ghodake, Marcos Grillo, Vaani Radhakrishnan, Sindhu Raghuram, Suraj Shourie

CSE 6242: Data and Visual Analytics

Introduction

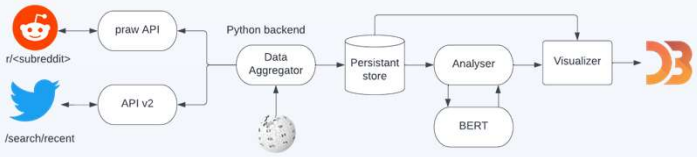
The intent was to study a significant event in the US: The Supreme Court overturned the historical Roe v. Wade case, opening the possibility of individual states across the country enforcing state-wide bans on abortion. This, in turn, has brought the topic of abortion into the center of public interest nation-wide. We maintain the goal of analyzing sentiment around a politically charged topic, focused on abortion stances within the US during the past year. Our objective with this project is to examine the varying responses to this issue and the spatial distribution of sentiment in the United States. Health systems, social scientists, clinicians, and charitable organizations geared towards supporting women are especially interested in understanding the sentiment around this historic overturning. Understanding the trend and consensus would help them in developing better response plans, guiding women through complex pregnancies' and identifying underrepresented populations to provide better support to them.

1. What is the spatial distribution of pro-abortion and anti-abortion sentiments in the country following the overturning of Roe v. Wade?
2. Can we observe any alignment between the disposition of the leading political party in the state and the general sentiment of the public?
3. As Reddit is completely anonymized, does anonymity lead to viewing more unfiltered and potentially extreme opinions online as compared to Twitter?

Data

We collected Twitter data Pythonically from ***Twitter API V2***. The raw collected dataset size is about 100k posts on Twitter. Leveraged ***praw*** interface to collect Reddit data, it's around 50k Reddit posts on selected subreddits. Government related data was collected from Wikipedia and entered manually to a persistent store in order to gauge how the opinions sway compared to the overall political stance in the state.

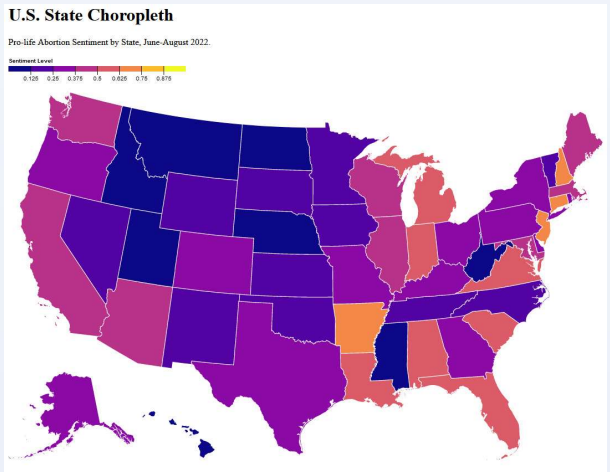
We used D3 choropleths for visualizing the main sentiment for each US state. We are using the ***sentiment score*** from our ***BERT*** model. Sentiment score factors in the severity of the emotion in the said content. This score is calculated for each entry in our data set which is then paired with the location data of each tweet and Reddit post. We grouped the tweets based on the state from which the tweet was created, and an average sentiment score was calculated on a state level.



1. Severity of opinions on abortion stance by state

The color scale shows that states with a deeper shade have a lower average sentiment score while states with a lighter shade have a higher average score. There are two visualizations: for prolife and prochoice. Both of them show distribution of sentiment scores on the US map.

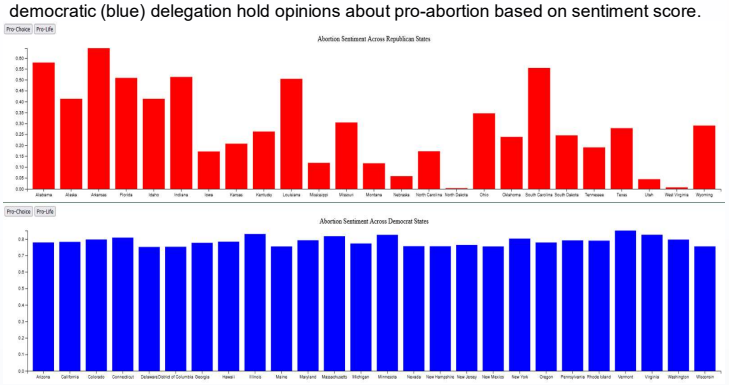
For example, this is how *prolife* related sentiments are expressed in different states. Darker shade corresponds to a stronger anti-abortion stance.



2. Political affiliations and abortion opinion comparison

We also made interactive charts using D3 to compare average level sentiment for pro and anti abortion tweets for each state and for both major political parties. There are two visualizations, one for states with democratic delegation in 2022 and one for republican. We are providing an onclick filter at the top to show *pro-life* and *pro-choice* sentiments for all these states categorized into democrat and republic. We used this visualization as it is intuitive and interactive for the end user.

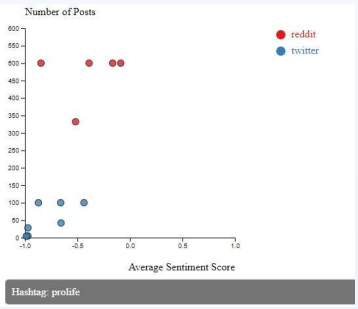
For example, this is a view with ***Pro-Choice*** filter to show states with republican (red) and democratic (blue) delegation hold opinions about pro-abortion based on sentiment score.



3. Cross-platform inclination: Reddit and Twitter

We are leveraging a bubble chart to compare cross-platform sentiments. We categorized Twitter hashtags in two groups: prolife and prochoice and juxtaposed their score with corresponding subreddits' aggregate sentiment scores. In our graph the x-axis corresponds to the sentiment score and the y-axis corresponds to the number of tweets made (Twitter) or posts and comments (Reddit) made in the corresponding subreddit. With this visualization, we were able to immediately identify how sentiments differ for a topic of discussion between Reddit and Twitter.

For example, group of 5 *pro-life* categorized hashtags and their corresponding subreddits. Hashtag/subreddit names will be shown on bubble hover.



Experiments and results

We successfully analyzed the Twitter dataset to visualize how states express their opinions about this issue. Our visualizations confirmed the political partisanship and consensus in almost all the states on this issue have a strong correlation hence confirming the fidelity of data as well as visualization.

The cross-platform sentiment analysis delineated how different social media mediums are given the depth of anonymity offered on these platforms. Reddit opinions inclined a lot toward the extreme ends of the opinion spectrum compared to Twitter.

User Study: In terms of data clustering suggestion, users preferred having different bubbles for different hashtags instead of one giant blob for prolife and prochoice for Reddit and Twitter.

The state political stances and abortion related opinions received the most positive feedback on its ability to distinguish state's sentiment scores. The social media juxtaposition chart between Reddit and Twitter also received a good response. a lot of users also echoed the sentiment that the results align with what they have observed on these sites, personally. We also heard some clamor for the lack of descriptions for what 'sentiment scores' meant. The state choropleth and hashtags scatterplot received a very positive feedback on its ability to explain each state's sentiment score through hashtags.