Abstract:

Introduction:

Political data analysis has been a hot topic in recent years. Many analyses focus on polling data and the trends over time as well as geographically located. Leading into the last election, the majority of polls had Hillary Clinton winning by a fairly large margin. As we all know, the polls were wrong. We chose to instead look for our data from the discourse on online forums, rather than using polls and political questionnaire data. Where better to glean insight than in the intellectual battleground of social media, where the majority of the conversation has moved in recent years. The basis of our analysis stems from the deep current divide between Liberals and Conservatives and the hate that we see online between the two groups. Both sides accuse the other of being the more hateful group, so our analysis uses a Sentiment Analysis package on data gathered by Web Scraping popular political forums on Reddit. The 2 main forums that we used for contrast are r/Conservative and r/Liberal. The goal in this analysis is to attempt to quantify overall tone of discourse and compare the groups using a more numerical system of judgement, rather than a subjective exploration of the forums. Sentiment analysis, in short, attaches a positive or negative numerical value to words based on their connotation. Our resulting Dashboard of visualizations shows the Sentiment analysis of both r/Conservative and r/Liberal with descriptive statistics of both groups as well as a visualization of the most popular words used in those forums, which gives the viewer a good sense of what the conversation was focused on in that given forum.

Research was also done on the nature of online forums and political discourse. Much of the research has looked at sites like Twitter, but there are similarities between the political discourse on almost all mainstream social media sites, such as Facebook, Twitter, Reddit, etc. One aspect of social media and politics is the development of what are called Echo Chambers. An Echo Chamber is when a group is able to insulate themselves from opinions of an opposing view and conversation further strengthens their already formed opinions. Research on this idea may shed some light into the overall negative tones of both groups and how forums may lead individuals to stray further from a moderate viewpoint.

Methods:

Instead of searching for previously compiled data, our group gathered its own data. Web Scraping using the Reddit API Praw gathered Reddit comments from specified subreddits. Praw outputs a data frame with the posts title, score, url, the number of comments, the date that it was posted, and the body of the post. Once the data was gathered and compiled into a data frame, the Sentiment Analysis of the body of the post was analyzed using the Sentiment Intensity Analyzer from the Vader Sentiment package within python. The Sentiment Analyzer gives a numerical score to words and phrases based on how positive or negative the connotation is. This process was used on 2 separate subreddits, r/Conservative and r/Liberal.

A word frequency operation was also used on the body of these posts, in order to gather the most popular topics on the forums. The body data from the data frame was cleaned for stop words using the Natural Language Tool Kit, or NLTK, from python. Stop words and punctuation were removed in order to find the relevant words most commonly used. The resulting list of word frequency was used in the Dashboard for a word cloud, illustrating words most commonly used in both subreddits.

Tableau was utilized to create the dashboard. Attributes included in the dashboard are as follows: word clouds for both groups, descriptive statistics of sentiment analysis for both sides, a butterfly chart comparing sentiments of both groups, and a list of most common words color coded by party and frequency.

Analysis:

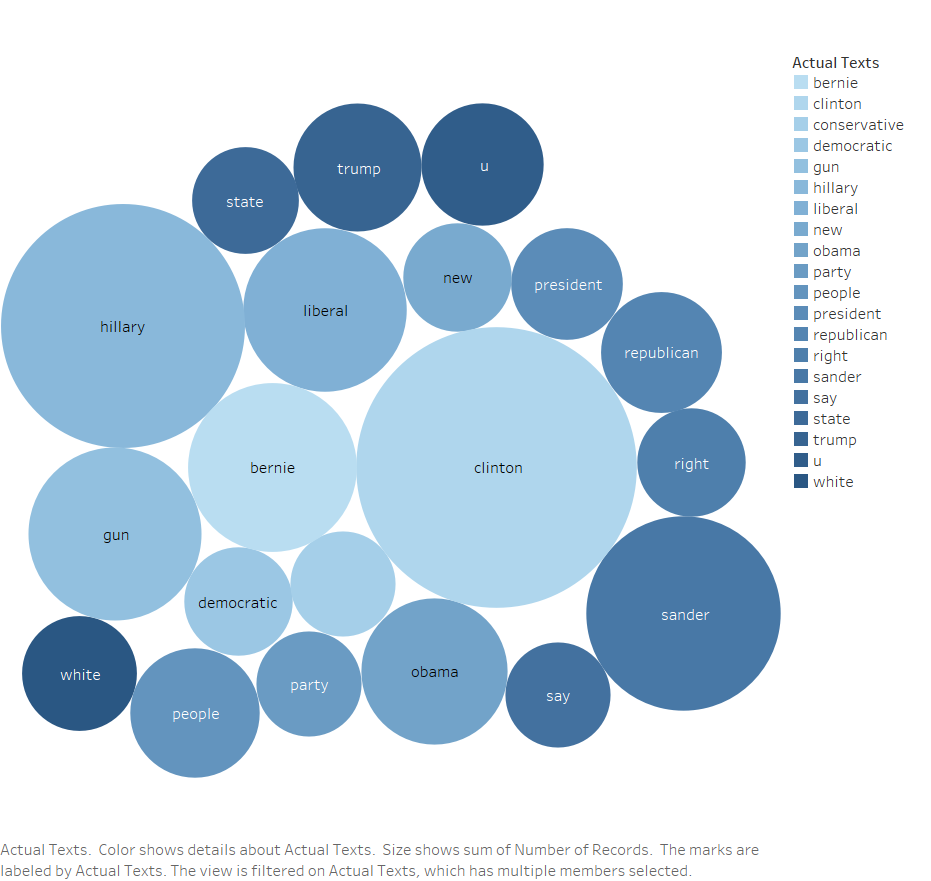
The rise of social media has led to an increasing amount of political discourse moving to sites such as Facebook, Twitter, and Reddit. This increased connectivity has allowed more individuals to share ideas, debate each other, and potentially come to new, more informed conclusions based on their experiences online. With every new technology, however, there are emergent properties that could not have been fully predicted prior to the introduction of said technology. One such property is the emergence of the echo chamber on social media platforms. There are many definitions of an echo chamber, but all converge to the same idea.

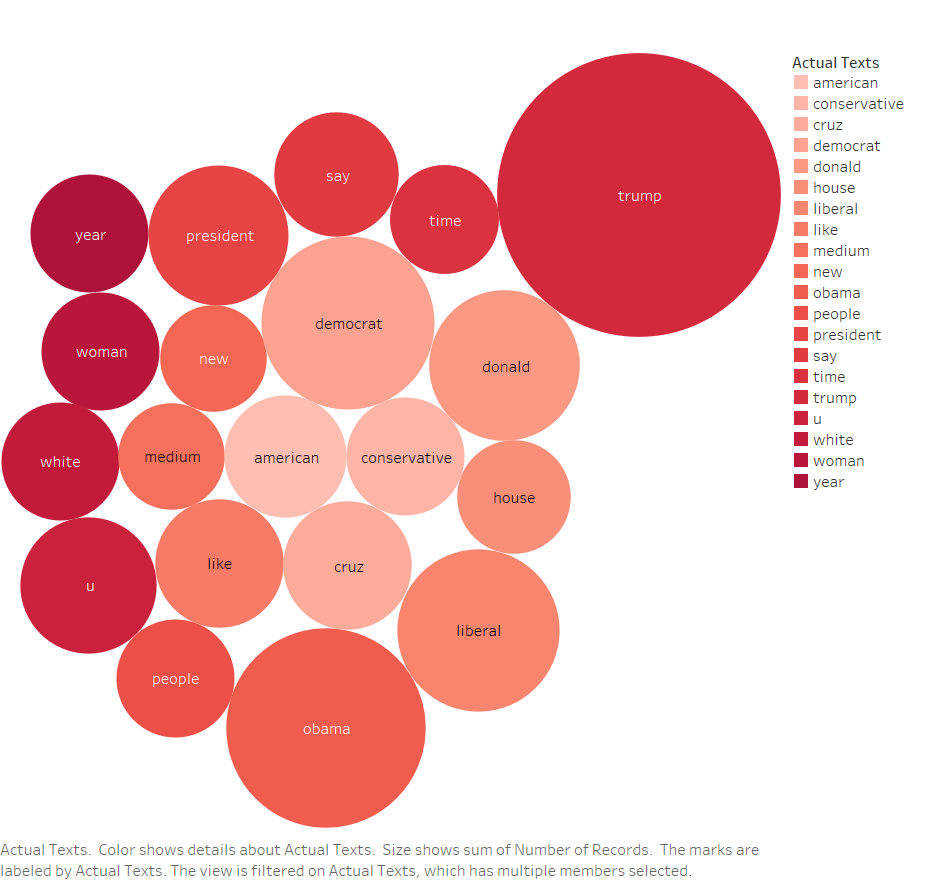
“In news media, echo chamber is a metaphorical description of a situation in which beliefs are amplified or reinforced by communication and repetition inside a closed system. By visiting an "echo chamber", people are able to seek out information which reinforces their existing views, potentially as an unconscious exercise of confirmation bias. This may increase political and social polarization and extremism.” (Wikipedia, 2019)

For the purposes of a general definition of an echo chamber, Wikipedia had one of the most specific and usable definitions. While not specifically directed at social media, one can safely argue that political discourse online has surpassed conventional news media in terms of volume and influence. Prior to the rise of social media, one may have assumed that given an online public square and the constraints of the medium, individuals would have interacted freely with everyone else. The emergence of the echo chamber is an interesting phenomenon. The phrase makes reference to being in a cave, or an area where sounds echo. The louder the sound, or the more sounds collectively produced bounce off of walls and amplify each other. In the case of online political forums, you replace sounds for ideas and sentiments in this theoretical situation. In a paper from 2018, this idea is echoed:

“The opinion corresponds to content items shared by users, while the underlying social network is what allows their propagation. We say that an echo chamber exists if the political leaning of the content that users receive from the network agrees with that of the content they share.” (Garimella, De Francisi Morales, Gionis, Mathioudakis, 2018).

Given the agreement on the term of echo chambers, we move forward. Our analysis of the 2 subreddits showed that among Liberal forums, the most commonly used words over the past 5 years were Clinton, Hillary, Bernie, Conservative, Democratic, Sanders, and Gun. From the Conservative side, the most common words were Trump, Donald, Obama, Democrat, and Liberal. Given these word frequencies, we can safely assume that both sides had a lot to say about candidates from their party in the previous election. Conservatives also had a lot to say about Barrack Obama, while Liberals had a lot of posts concerning the word gun, which can be tied to the gun control debates.

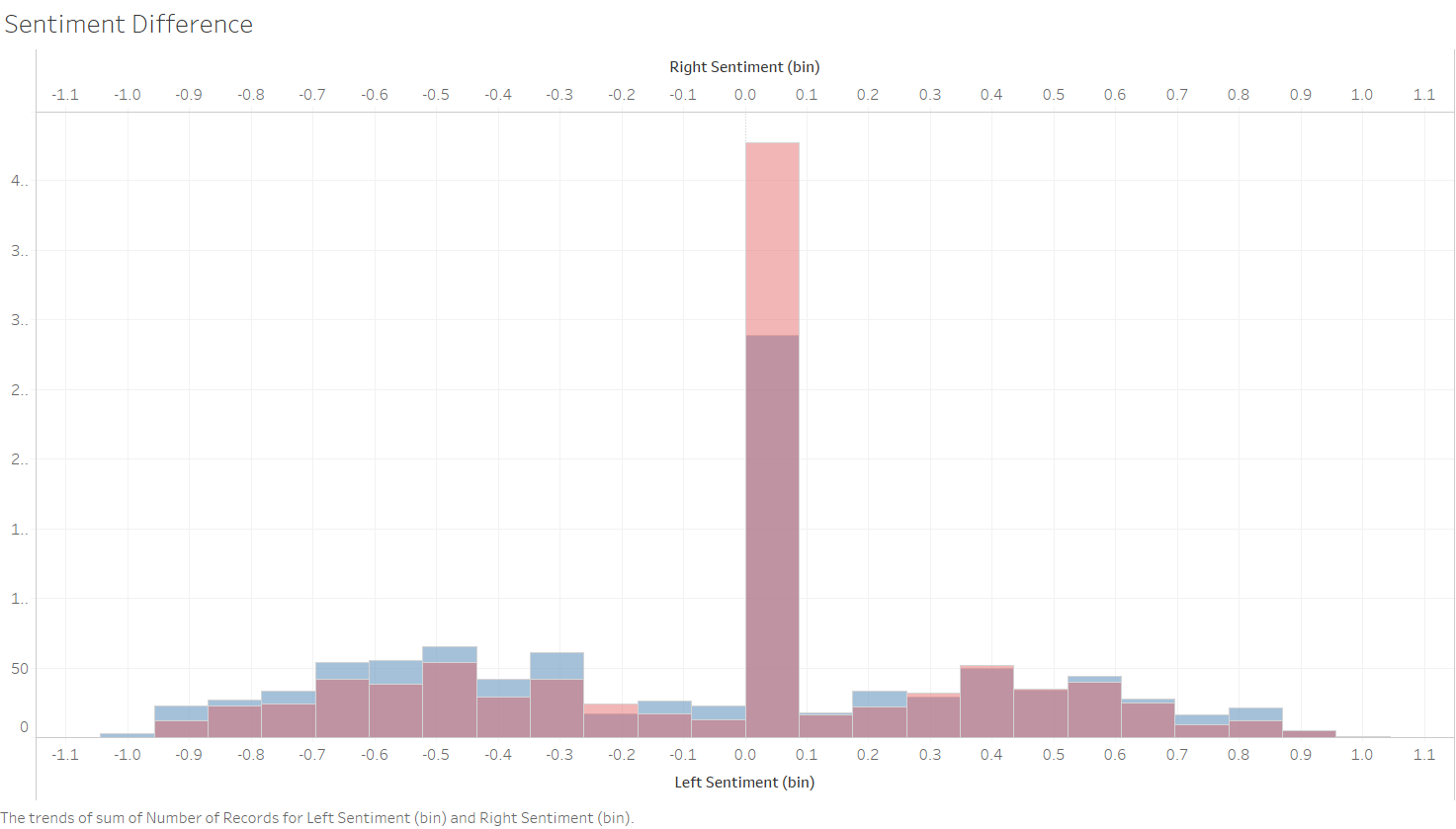




The sentiment analysis of both groups yielded average sentiments of -0.07870 for Liberals, and -0.04110 for Conservatives. A negative number with regard to the sentiments refers to the generally negative language used in posts. From this we can draw the conclusion that both groups have generally negative things to say, although Liberals were slightly more negative by our metrics. The standard deviation for Liberals was 0.4428 and 0.3882 for Conservatives, showing slightly more variability among the sentiments of Liberals than conservatives. The main conclusion that we draw from this analysis is that sentiments in these forums is generally negative in tone, and that due to a Conservative President being in office, the Liberal groups have more negative things to say, due to the controversial nature of the current administration.

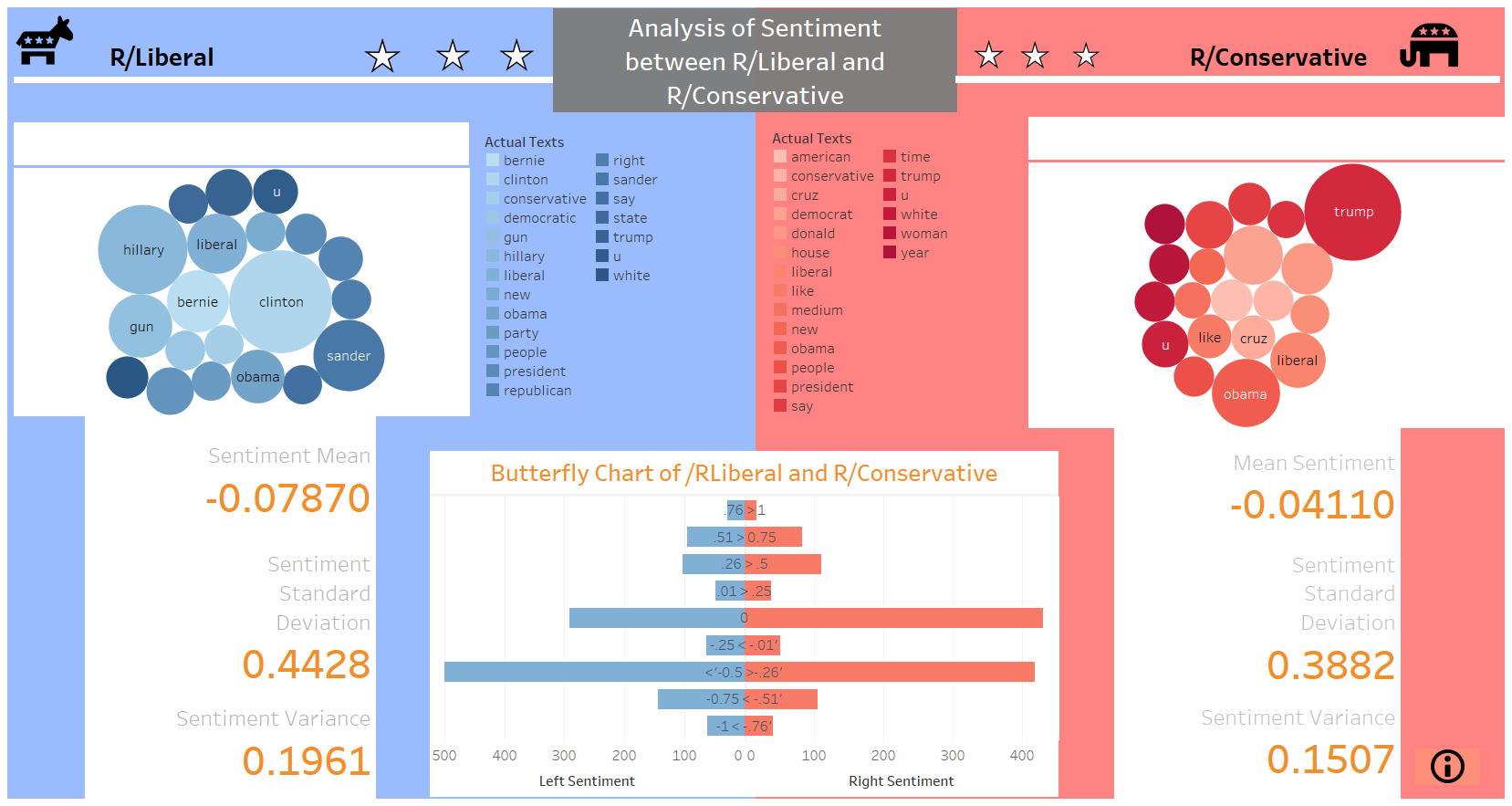
There is some evidence of these Subreddits being echo chambers to some degree. The format of Reddit is more conducive to seeking out only opinions shared by the user. Joining specific discussion boards with titles that you identify is clearer cut than on Twitter or Facebook. Twitter’s Hashtag does allow for searching and interacting with content based on keywords, which can also contribute to echo chambers. Facebook groups also allow for this activity. We can attribute some of the negativity shown in our analysis to the fact that the reinforcement of ones own views on these discussion boards. There is the possibility that this reinforcement can foment hate and negativity on forums such as this.

The Sentiment Analysis tool used is well suited for our purposes, but potentially even better for the purpose of hate speech detection. “Hate speech is a particular form of offensive language where the person using it is basing his opinion either on segregative, racist or extremist background or on stereotypes.” (Watanabe, Bouazizi, Ohtsuki, 2018). The average scores of these forums are generally negative but the interesting statistics are those at the fringes of our visualizations, shown below.



The red is coded for Conservative, while the blue is Liberal. As you can see, conservative has a much higher concentration centered around zero than liberal, with the liberal having higher concentrations in the lower ranges of the histogram. Hate speech is generally classified on these low fringes.

In conclusion, our analysis of the data collected as well as the literature is that online political forums do tend towards average behavior that exhibits echo chamber characteristics. The overall negativity of both sides as well as the distributions of sentiments shows this, and due to the format of Reddit, there doesn’t appear to be much interaction between the two sides, at least on these two major subreddits.



Bibliography:

* Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW ’18*, 913–922. <https://doi.org/10.1145/3178876.3186139>
* Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. *IEEE Access*, *6*, 13825–13835. <https://doi.org/10.1109/ACCESS.2018.2806394>
* Echo chamber (media). (2019). In *Wikipedia*. Retrieved from <https://en.wikipedia.org/w/index.php?title=Echo_chamber_(media)&oldid=925469265>

**Data Sources:**

Subreddit r/Conservative

<https://www.reddit.com/r/Conservative/>

Subreddit r/Liberal

<https://www.reddit.com/r/Liberal/>

Praw Documentation

<https://praw.readthedocs.io/en/latest/>

VaderSentiment Documentation

<https://pypi.org/project/vaderSentiment/>