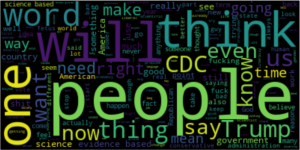
Recently, the White House has enacted a ban inside the Center for Disease Control (CDC), stating that they cannot use the words 'fetus', 'science-based', 'evidence-based', 'vulnerable', 'diversity', 'transgender', and 'entitlement'. In a scientific sense, these are very important words because they give us both a deeper and broader picture of what the state of the population is at: have any diseases, such as the flu, come to diverse environments and how should we run a protocol to deal with it? What medical approaches are there for Zika as it affects fetuses in pregnant women? Transgender peoples have a higher rate of depression among them - why is that, and how can we combat that? These words shape policy. Banning them strips us of the ability to react well to novel diseases, leaving us a needless step behind a disease that may evolve into something more sinister than it is now.

Nonetheless, I decided to look at public opinion on how this ban was received. For this, I built a Python program ([link to code](https://github.com/ajitkoduri/Code-for-Blog-Posts/blob/master/CDC%2B7%2BWords%2BBan%2B-%2Bfor%2BReddit.py)) that went through the Reddit comments of the initial post about this ban in Reddit on the subreddit r/Politics. If you don't know about Reddit, it's an internet forum where you can browse a variety of topics, known as subreddits, ranging from politics to sports to your favorite video games. These subreddits contain within them 'threads', wherein a discussion of some subtopic occurs. You can also vote on what other people have said about these subtopics, and, while the voting system is meant for establishing relevancy to the topic, it usually becomes a stamp of agree or disagreement.

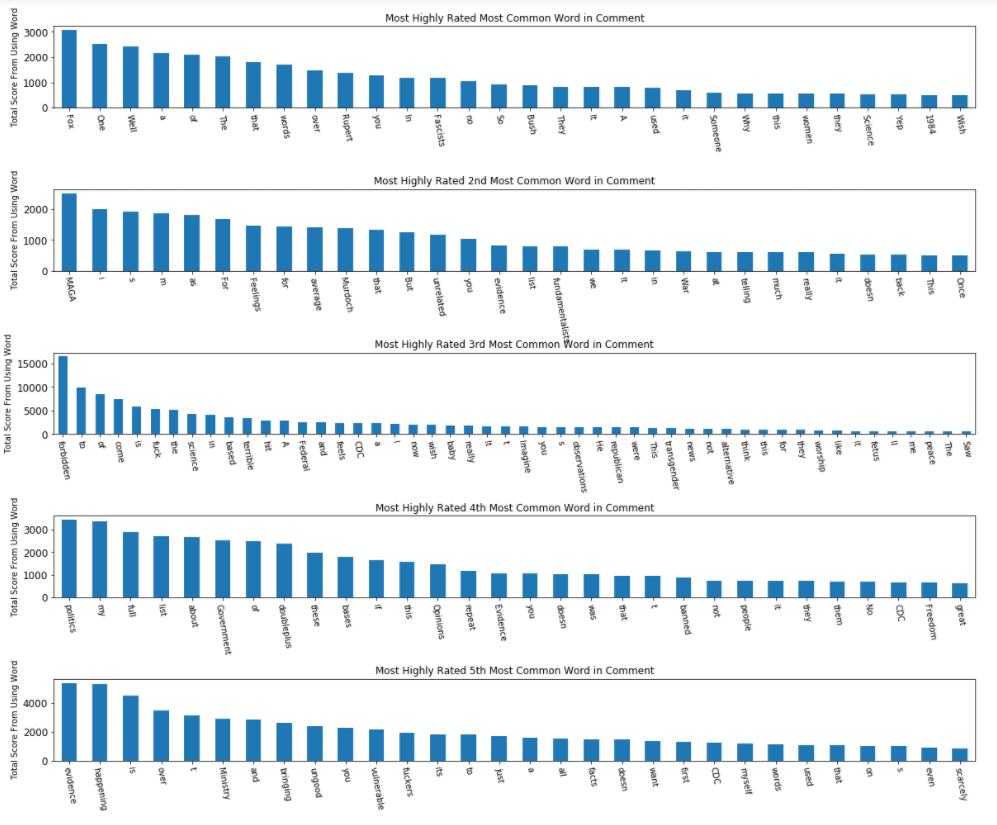
The r/Politics subreddit is a politically liberal subreddit, and most of the articles on there are from either liberal or left-leaning media. So how did r/Politics receive this ban? The data I looked at consisted of close to 6000 comments from over 3000 unique users.

**What words did they use?**

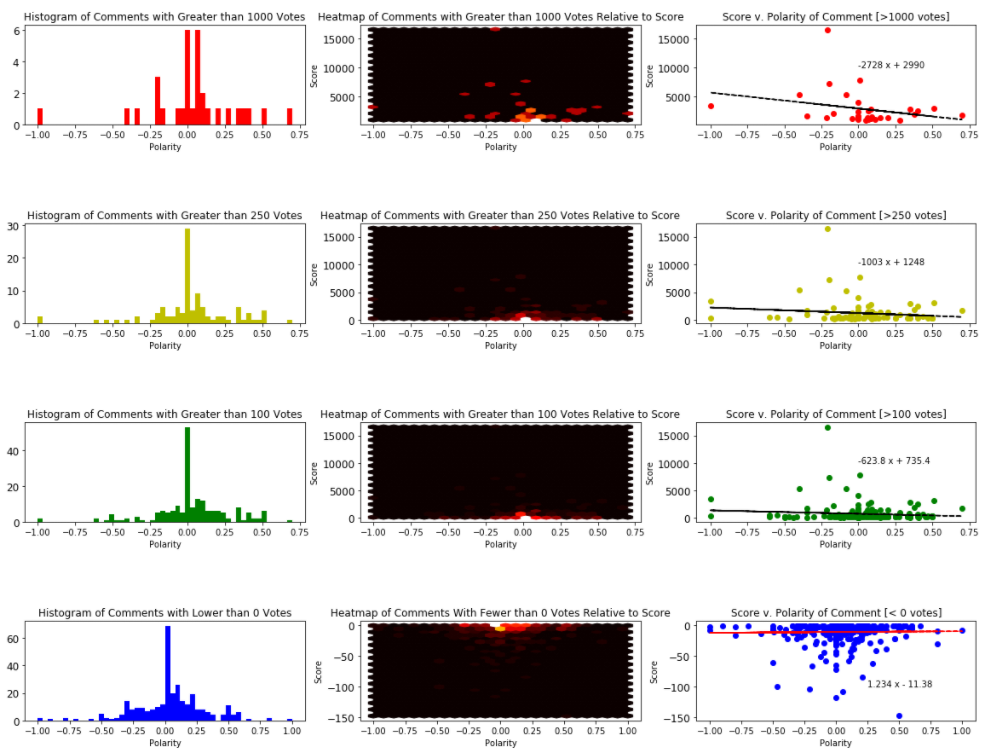
[](https://ajitdoes.files.wordpress.com/2017/12/wordcloud.png)

From the word cloud to the left, the general population is talking mostly about the ban on the words evidence and science-based specifically. This is pretty logical, given that the CDC is a medical branch of our government, focusing primarily on getting our people to safety from diseases. Of course, there's more to it - there's discussions about the Trump administration, some people are attacking the legislation, some people are discussing the American government's outlook and the international perception of the American government, but the other specific words, 'fetus', 'diversity', etc. aren't being talked about nearly as much.

Below is a series of graphs of the most common words associated with each post, and how well each of these words did according to the score of the posts they were inside.

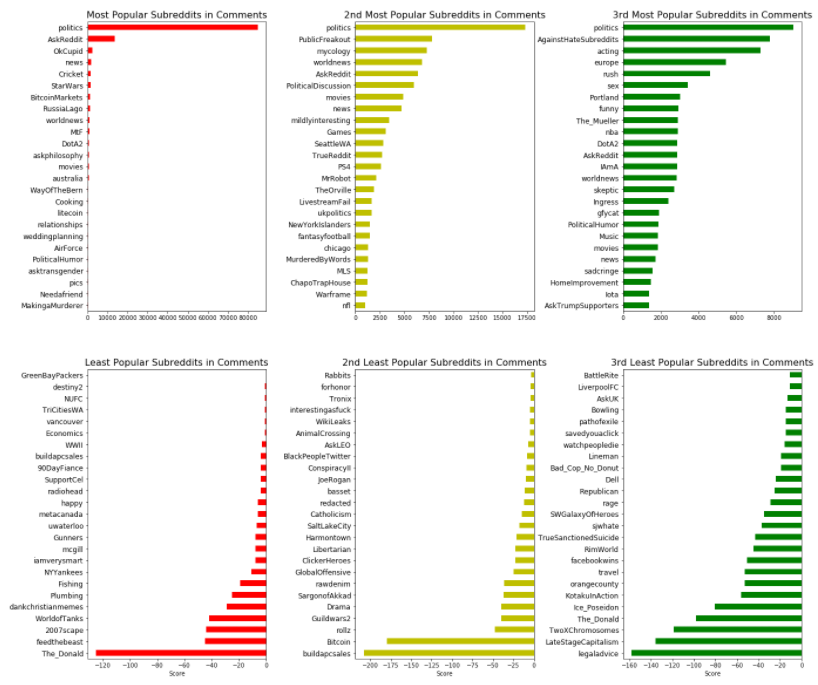
[](https://ajitdoes.files.wordpress.com/2017/12/wordspopularity.png)

What you'll instantly see is that most people are talking about FOX (most probably FOX news), fascists, Rupert Murdoch, fundamentalists, republicans, and probably slamming the right-wing for this new policy's enaction. Comparisons to 1984, talks about the CDC, sarcastic uses of 'MAGA' and 'alternative news', these all make up the words associated with the most popular comments, a precursor to how the sentiment of the comments affected their popularity.

**Sentiment Analysis[](https://ajitdoes.files.wordpress.com/2017/12/comments_histogram.png)**

The above graphs are depictions of how the comments are distributed. The plots were histograms, heat maps, and scatter plots of the polarity, how powerfully people were feeling either for or against in each comment, versus the score the comment received. What you'll notice is that comments that were fairly popular, the more the were polarized negatively. The line of best fit, with the equation given in the 3rd graph of each of them, tells the story about the popularity of each sentiment. The histograms for each of the splits of the comments is around 0, however the most popular comments were on the negative side. On the other hand, the least popular comments tended to be read as positive. Following the very popular comments on the positive polarity found them to be mostly replies to a negative comments, which means they were probably internet memes.

The lines of best fit for each subsection of the comments strongly show the quick drop off for having a positive view of this ban, with each batch of more popular comments having a heavier and heavier drop off, culminating in an almost 3000 vote difference between comments that are negative and comments that are positive for the ban. On the other end, most of the comments are approximately normally distributed around having a 0 polarity, with a slight leaning to the positive end, probably indicating that those comments are highly irrelevant or Trump spam.

[](https://ajitdoes.files.wordpress.com/2017/12/subredditpopularity2.png)A look into the above graph indicates what subreddits each of the commenters come from, and how popular each of the subreddit's posters were on this thread. What you'll notice is that a significant portion of the subreddits that had the highest points were left-leaning, but also diverse, meaning that most of the users had popular comments tended to frequent Reddit and the various subreddits it has.

On the other side of the table, many of the users went to subreddit's that are highly right-wing, including the infamous 'The\_Donald', 'KotakuInAction', 'Republican', and 'Libertarian'. Alongside those, it also includes many subreddits dedicated to video games that the other side doesn't. They also tend to visit subreddits related to America more, while popular comments also came from the 'ukpolitics' and 'Australia' subreddits. So the split in popularity is clearly between liberal commenters and their conservative counterparts, as expected.

I would have looked at the subreddit's based on their average Sentiment value, however it would probably not contain any useful data through just a simple computer program that guesses their sentiment based on the text of any given comment alone. It would require analyzing it with the context of the comment chain it is in, but that would become computationally more difficult the deeper in the comment chain you go! Additionally, memes and comments that end with '/s' (a term in reddit meaning that the comment was sarcastic and the meaning was the opposite of what you understood) would stymie those efforts as well.

This analysis certainly wasn't finding ground-breaking things in the data; I mostly did it to get my feet wet by looking at how Natural Language Processing could be used to understand data, and I thought that a politically polarizing piece like this would have intuitive results, which it luckily did: the majority of the unpopular commenters were from right-wing subreddits and tended to be 'All-American', while the popular commenters tended to be from left-wing subreddits and were more diverse. Additionally, there was a clear separation in video game subreddits frequented, but that is explained by the fact that most people in the alt-right are video-gamers (though the inverse is not true). The popular comments on the left-wing subreddit are those that align against the current government and its beneficiaries (in this case, the anti-science collaboration in the Republican party).

Overall, people were not polarized against any direction, though being highly against the ban was very popular. The one thing that was cute in this data was that while the overall comments, without regard for their score or poster's name, was slightly positive of the ban, taking into account the score reversed that - more of the people commenting were for the ban, but more of the people voting were against.

I hoped you enjoyed reading this one! I know the findings weren't anything novel or exhilarating, but doing it really helped me get a feel for some of the more exciting technologies. Also, I could replicate it to a thread on Reddit (or even a whole subreddit) and get the general perception and word choice there - something that could be very beneficial for a future post!