

# Visualizing Political Subreddit Similarity

Jourdan DeVies, Tess Leggio, Rachel Shah, Alex Sitaras, Lauren Skorb, Darsh Thakkar

## How do subreddits differ in their discussion of political topics?

The current climate of political discussion, especially on social media, is extremely polarized. We analyzed which subreddits discuss selected political topics similarly. We leverage Reddit data to expose individuals to all sides of a given topic and make inferences about groups of online communities.

**Google BigQuery** Using SQL queries within Google BigQuery, all submissions and comments for 2018 were extracted for the 30 largest news and politics subreddits and exported into Google's storage bucket.

### Sentiment Analysis

Text sentiment was determined using *SentimentAnalyzer's* rules-based approach to assign positive, negative, neutral, and compound sentiment scores.

### hLDA

hierarchical Latent Dirichlet Allocation conducted hierarchical topic clustering by identifying words shared among subreddits for a political topic.

## Conclusion

Distinct language is used across subreddits to discuss various polarizing issues, resulting in clusters of similar language that change by topic. Further, users are successfully able to interact with our visualization to learn new information.

### Reddit

>1.7 billion Reddit submissions and comments are available as public datasets.

### Python

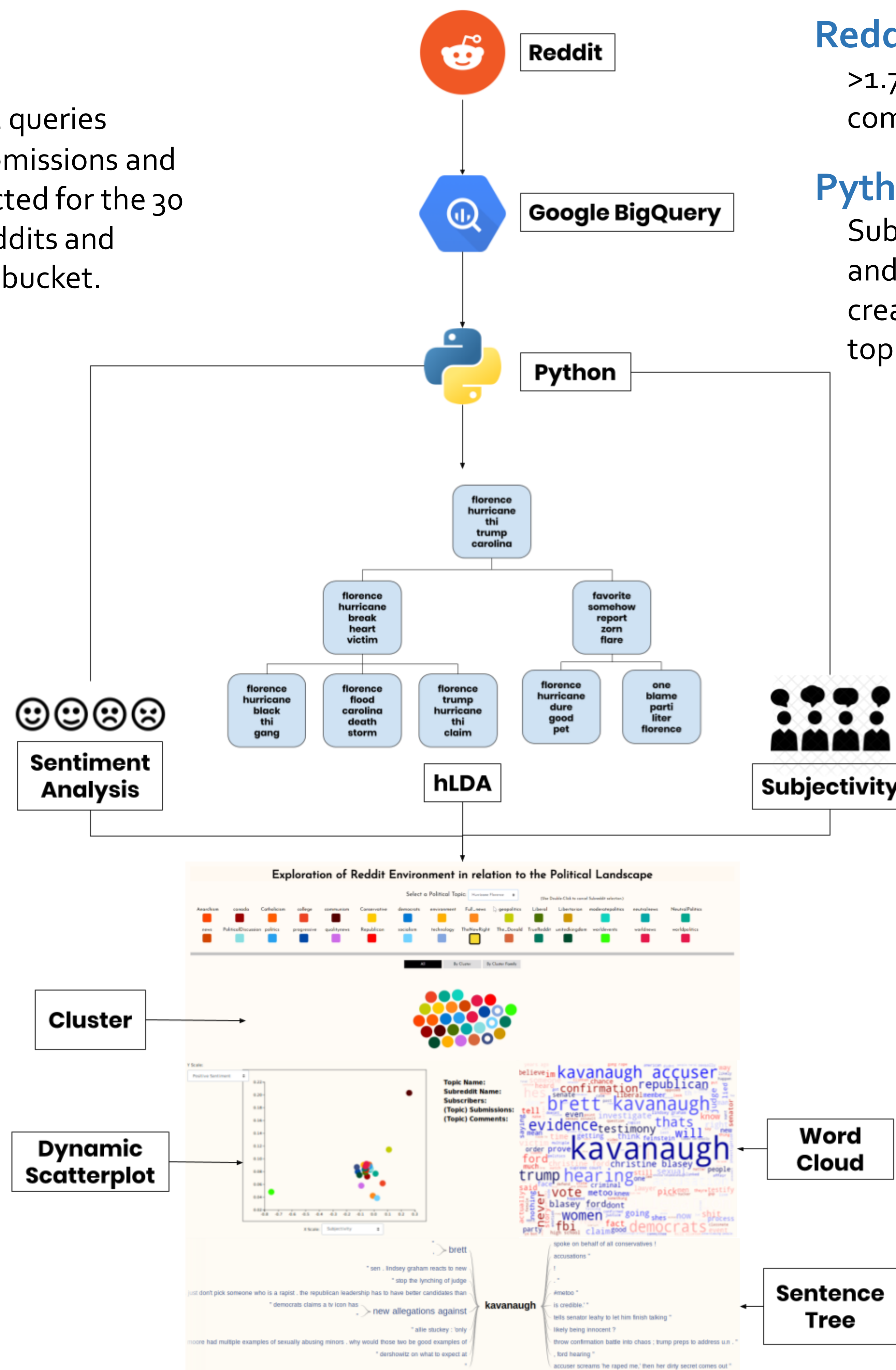
Submissions and comments were cleaned and combined, and bag-of-words files were created for each combination of political topic and subreddit using Python scripts.

### Subjectivity

The extent to which subreddits use emotive and subjective language rather than facts was determined by averaging *TextBlob's* predetermined subjectivity scores and intensity words across each subreddit's submission and comment corpus.

### Interactive Visualization

Created using D3, the visualization shows how each subreddit's **cluster** changes for a political topic. The **dynamic scatterplot** allows users to select metrics of interest to visualize. Interacting with the cluster and dynamic scatterplot reveal a **word cloud** and **sentence tree** for the subreddit of interest.

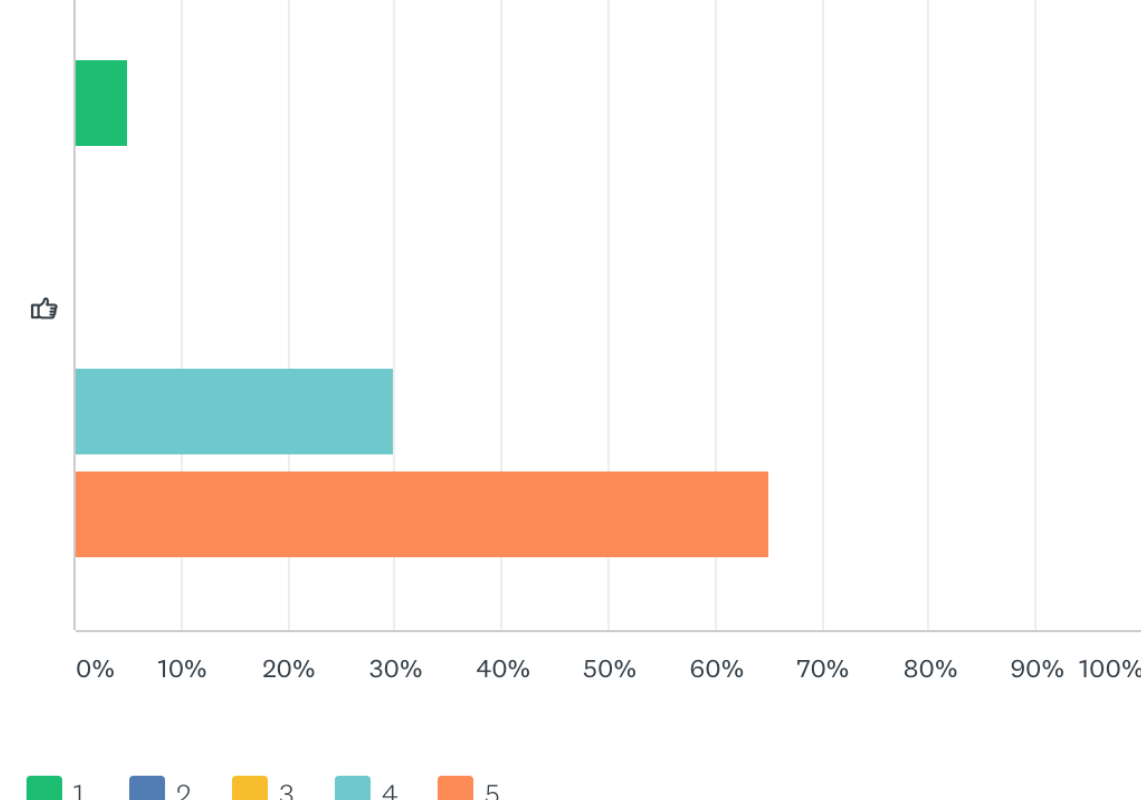


## User Survey Results

Provide evidence that our tool allows users to interact with information in novel ways

This website presents new information than what I have seen before.

Answered: 20 Skipped: 0



It was easy to interact with this website

Answered: 20 Skipped: 0

