

Trump's Twitter: Shifts in Content, Sentiment and Engagement Through Time and Context

A Text Analysis and Modeling Project

Aidan Ohlson

01. Project Data

Overview

- Data is taken from Donald Trump's official Twitter/X account
- Tweets range from 2009-2019
- Analysis will be done on an overall basis as well as categorical based on time period
 - Non-presidency (2009-2015, 2021-2023), Pre-presidency (2016, 2024), Active presidency (2017-2021)

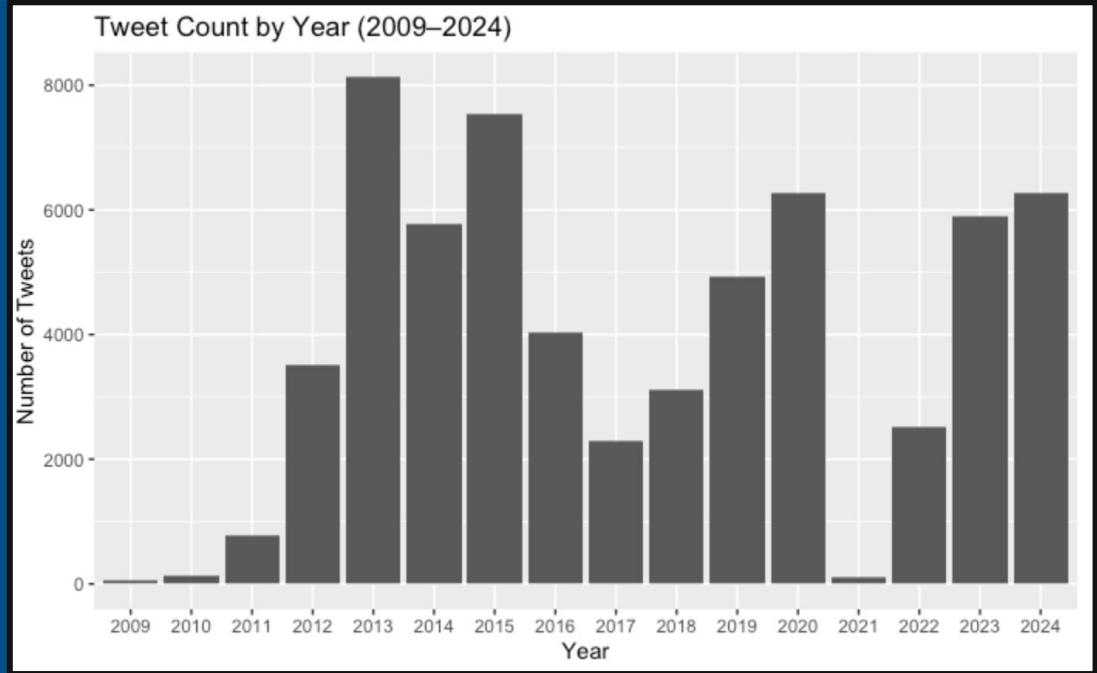
Cleaning

- Set to lowercase
- Removed punctuation
- Removed numbers
- Removed english stopwords
- Stripped whitespace
- Stemmed tokens

Exploratory Data Analysis

Variables of Interest:

- Date (date of tweet)
- Text (cleaned tweet content)
- Favorites (number of likes)
- Retweets (number of retweets)

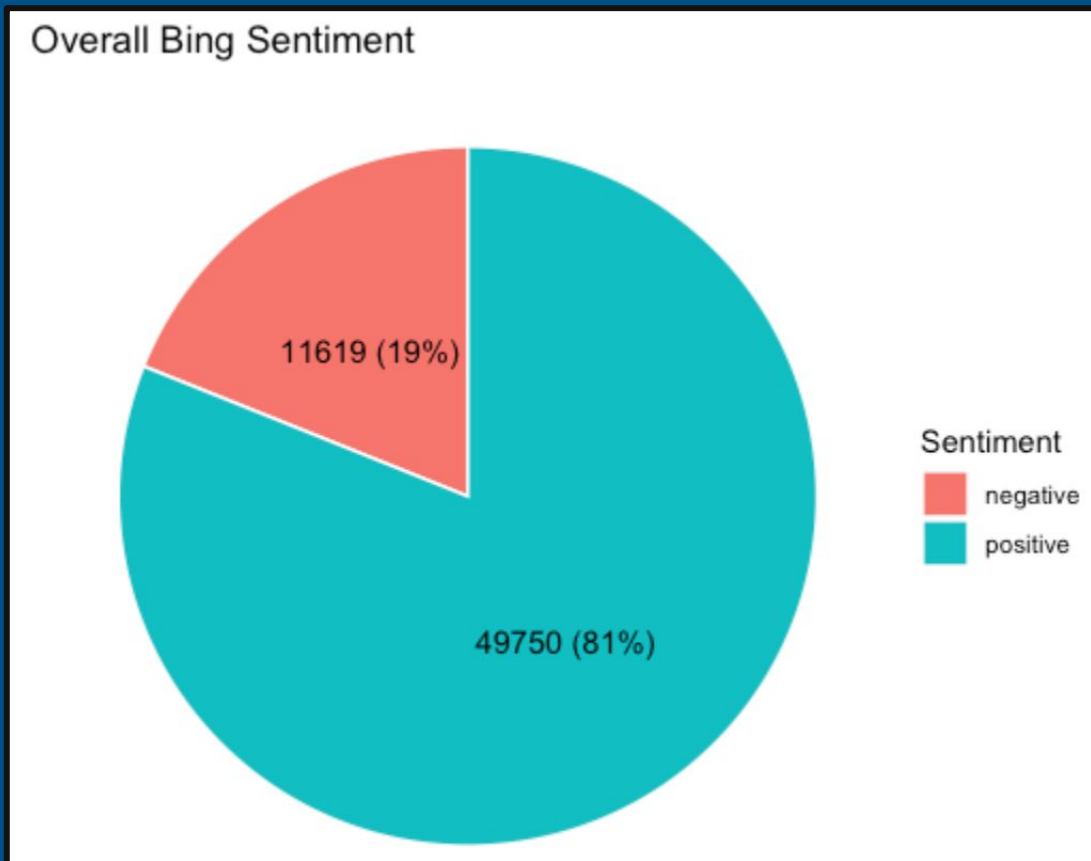


Average likes: 30,104

Average Retweets: 7,098

02. Sentiment Analysis

Positive/Negative Terms Overall

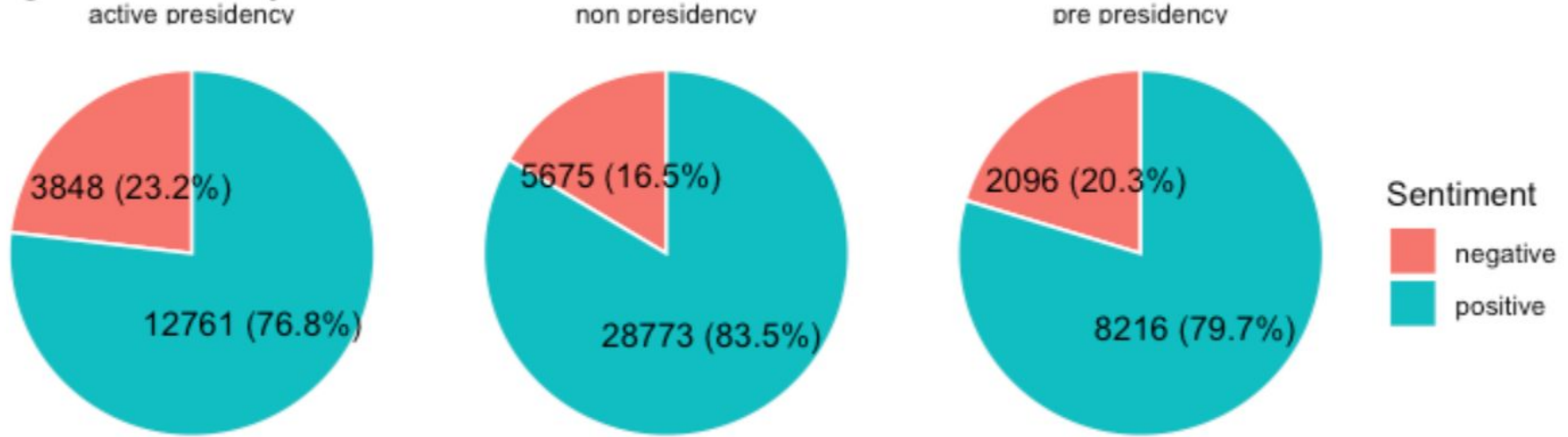


Bing Lexicon assigns specific words a numeric value and the sum of these values determines the sentiment of a given tweet

Roughly 4 of every 5 tweets is positive

Positive/Negative Terms by Context

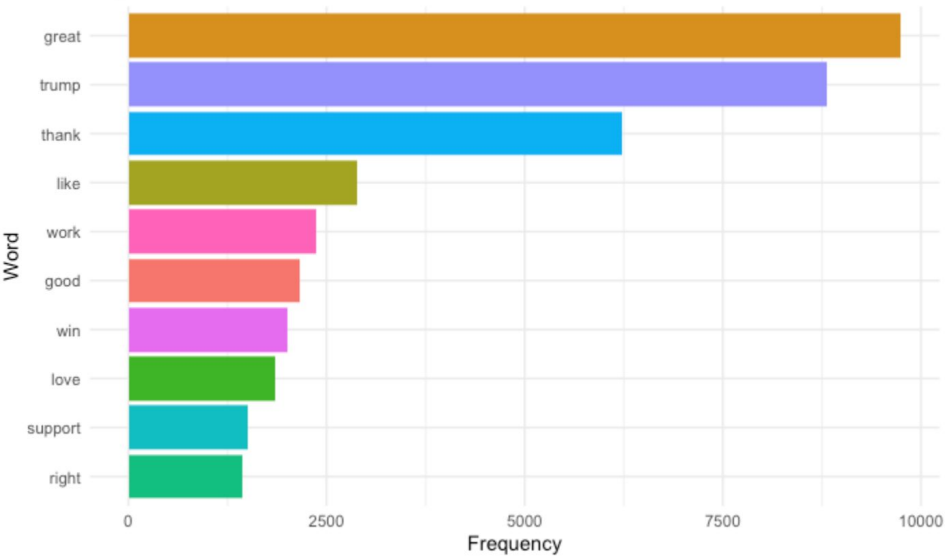
Bing Sentiment by Period



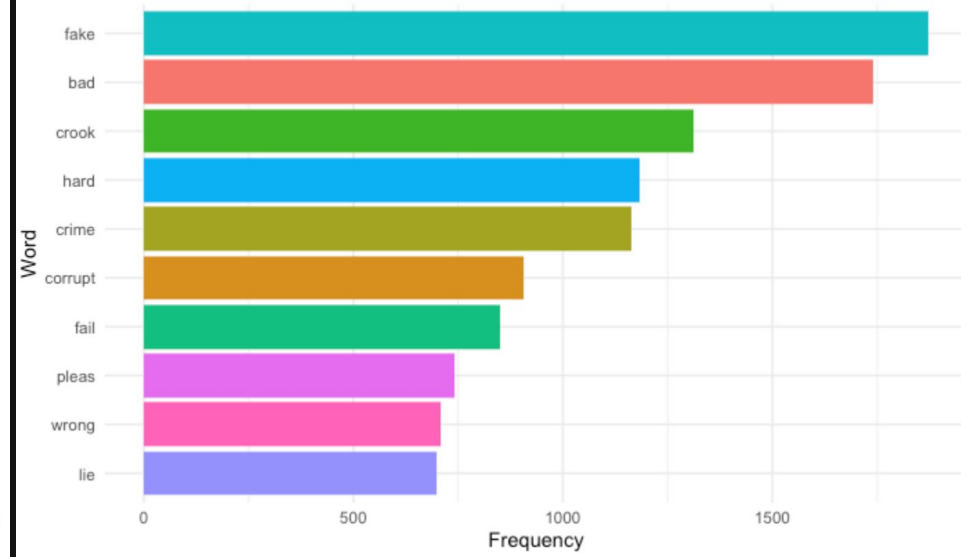
Notice the jump in negative tweets during active presidency and decline in negative tweets when in a non presidential context

Most Frequent Positive/Negative Terms

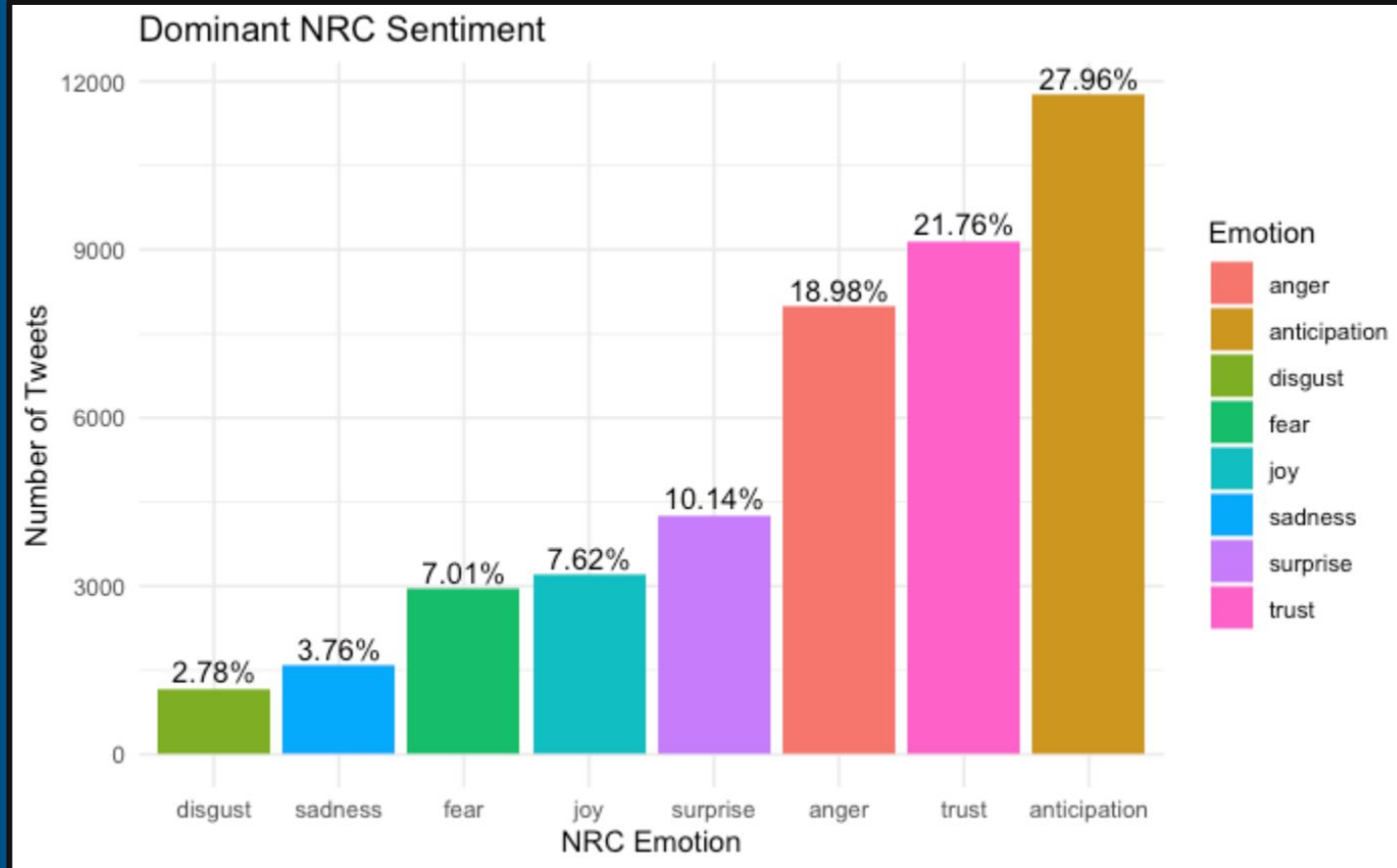
Top 10 Positive Words



Top 10 Negative Words



Overall NRC Sentiment

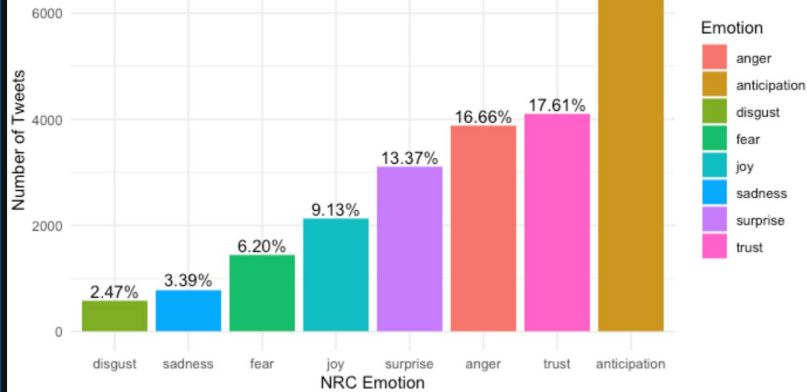


NRC lexicon
assigns
emotions to
specific words

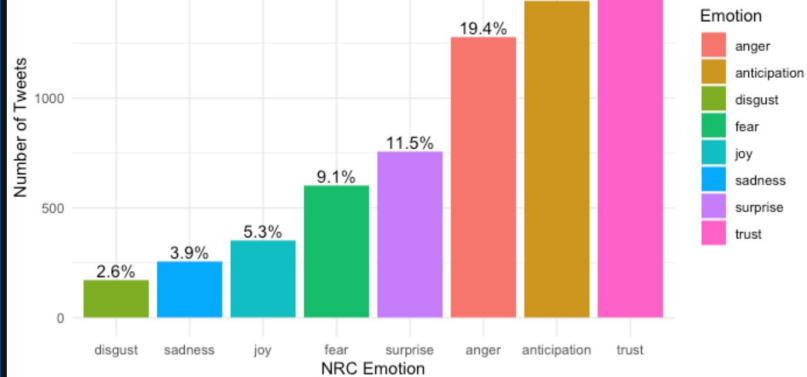
Classified tweets
by taking most
prevalent
assignment per
tweet

NRC Sentiment by Context

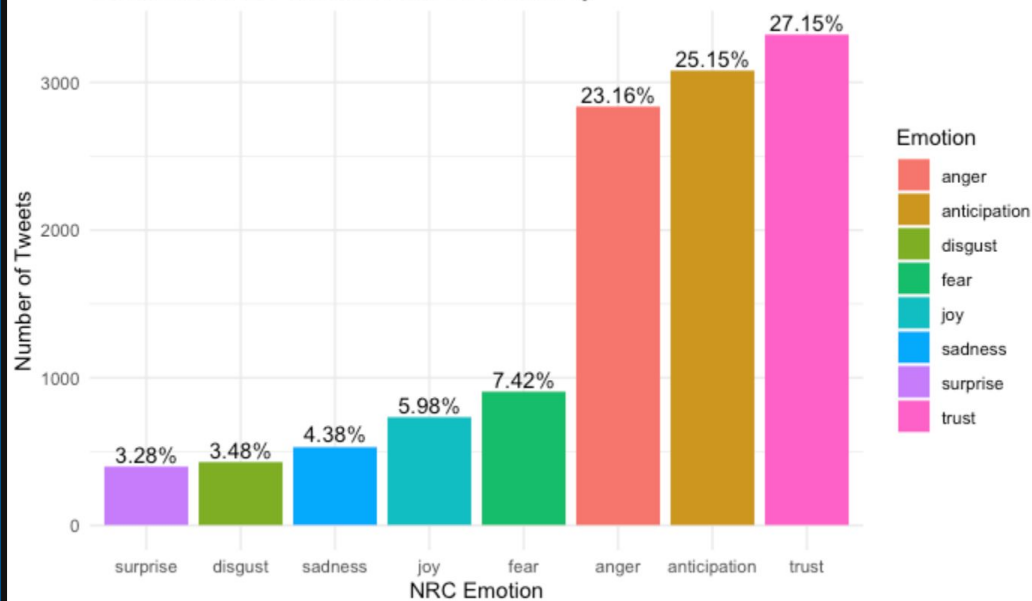
Dominant NRC Emotion: Non Presidency



Dominant NRC Emotion: Pre Presidency

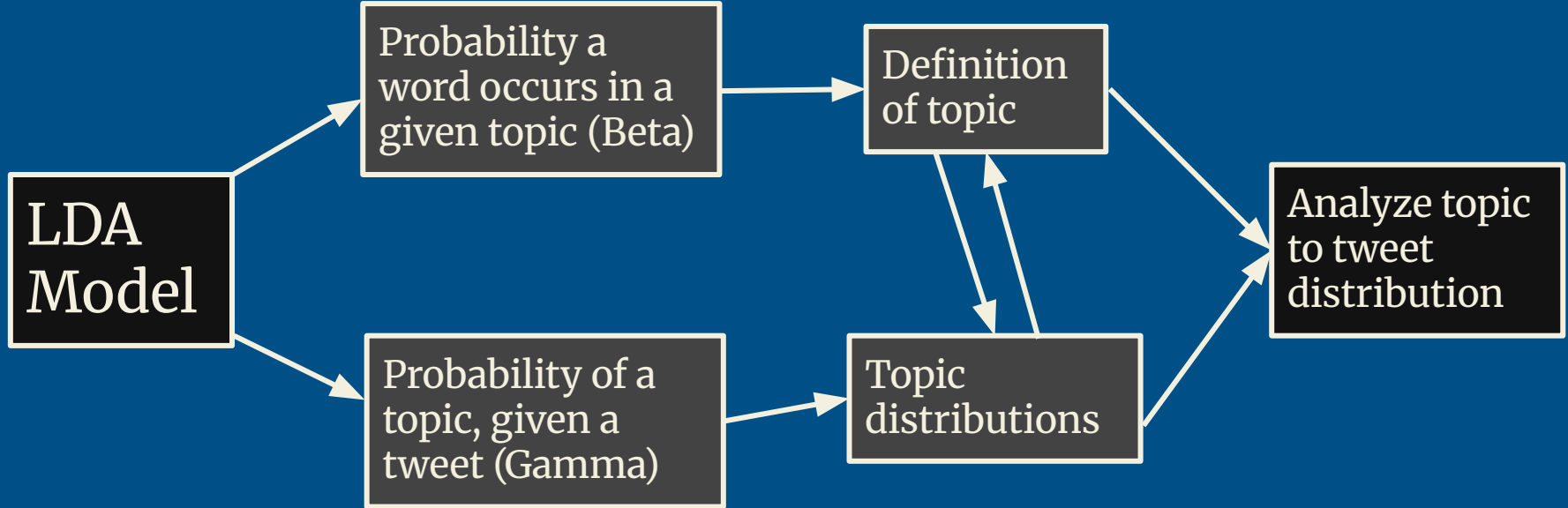


Dominant NRC Emotion: Active Presidency

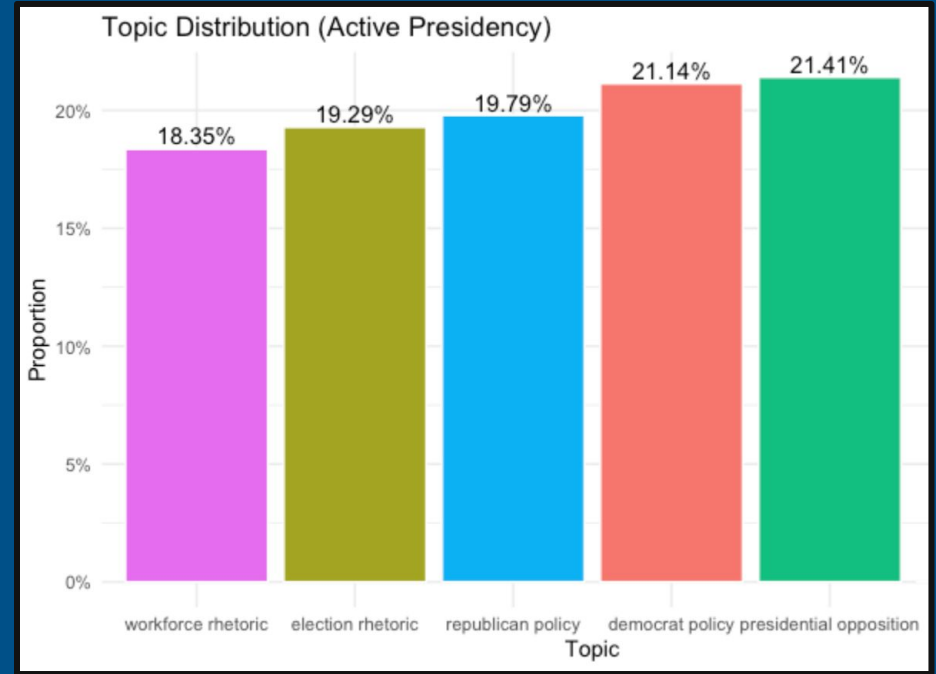
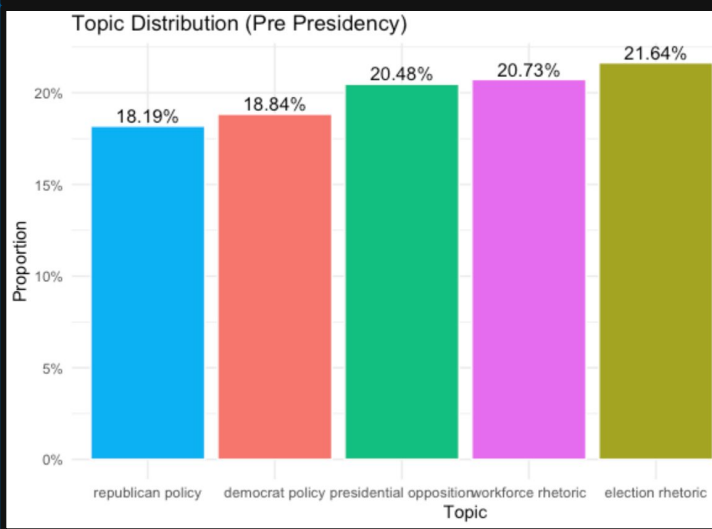
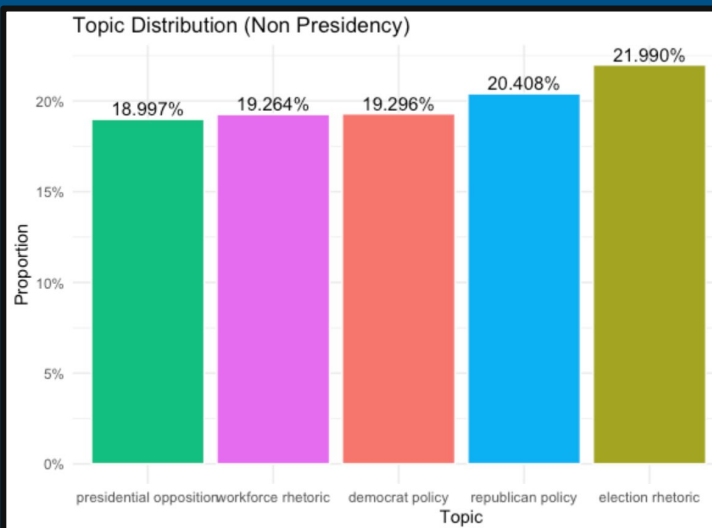


03. Analyzing Topics

Latent Dirichlet Allocation (LDA)

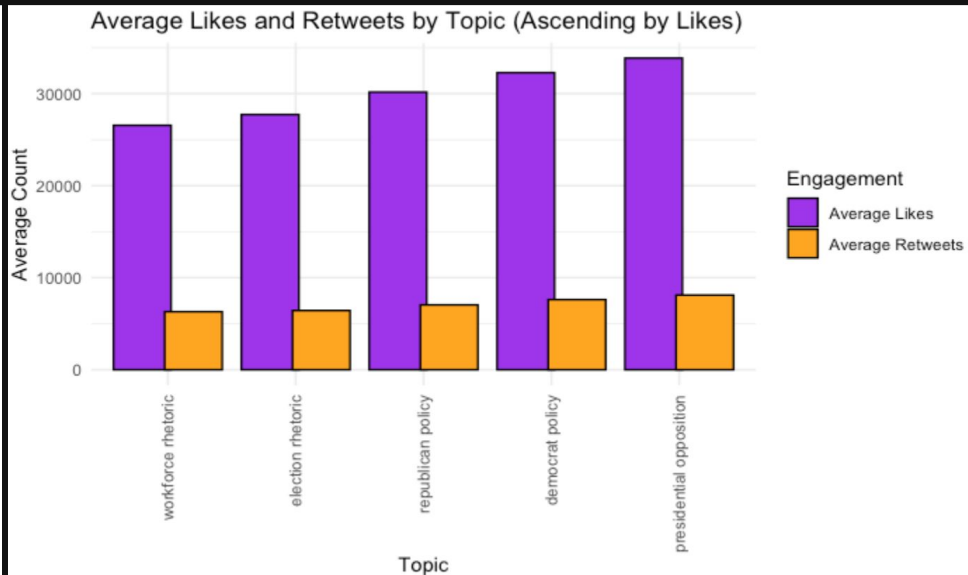
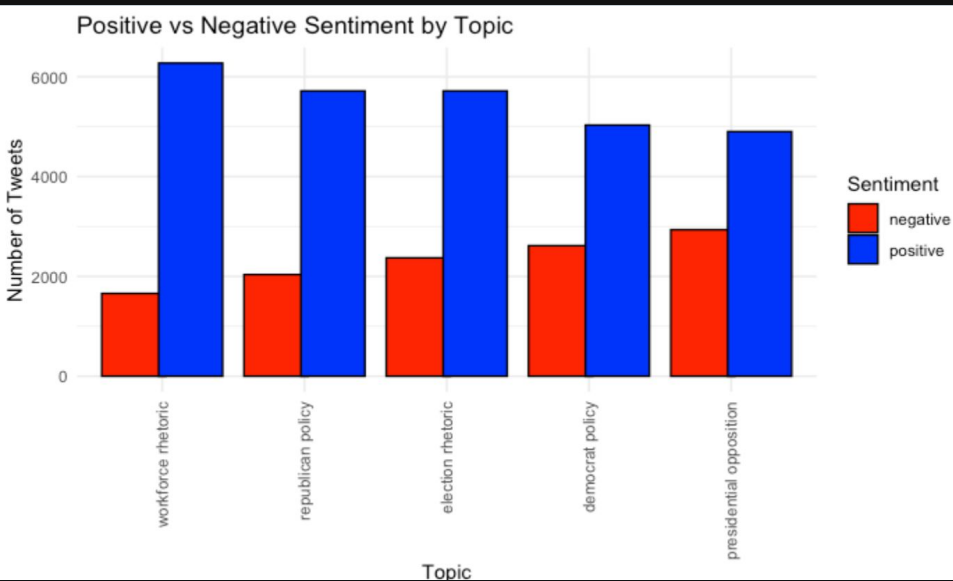


LDA Results



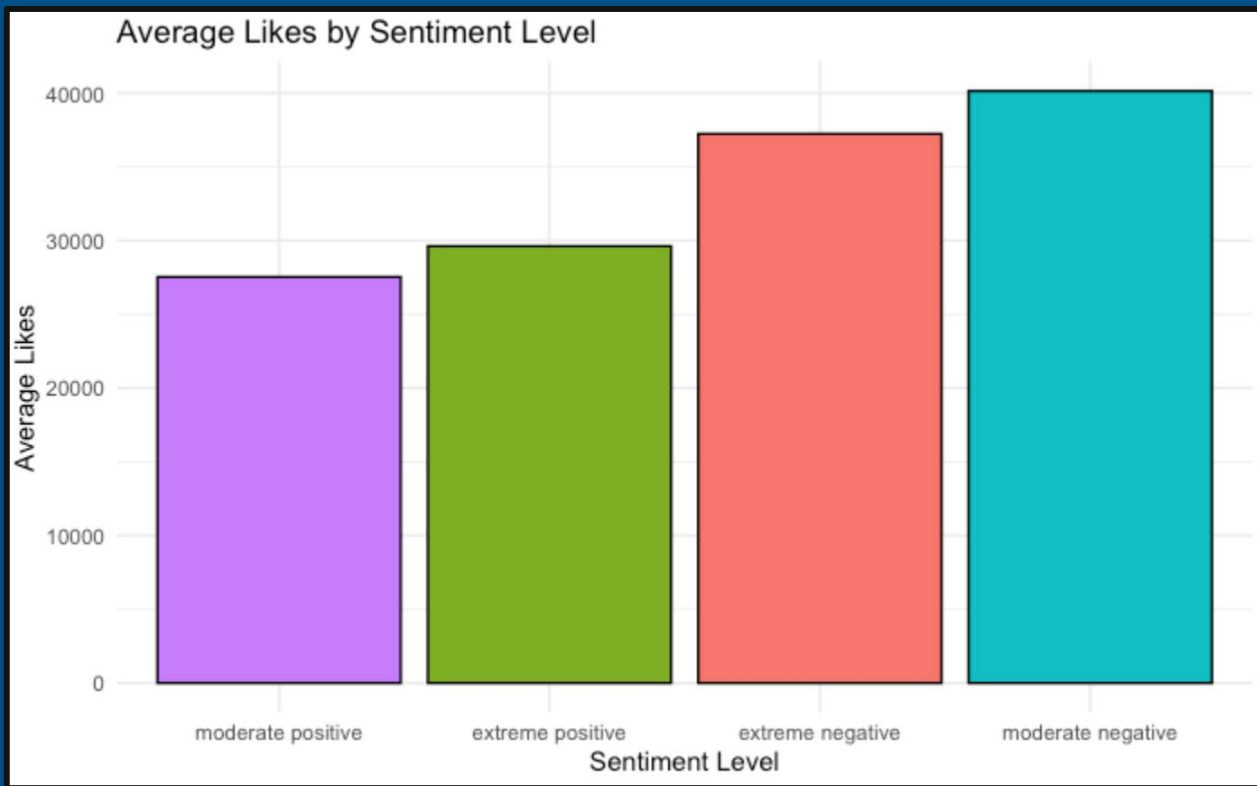
Connecting Sentiment and Topics to Engagement

Engagement Distributions



The first plot ascends by negative tweet count and the second ascends by likes

Looking at Extremes



Moderate was classified as any tweet within -5 and 5 using the Bing Lexicon. Extreme is anything outside of that range

Modeling

Modeling engagement via multiple linear regression. Likes is the outcome, with Bing score and Topic of the tweet as the predictors (only significant predictors shown)

| | | | | |
|-------------------------|---------|-------|---------|----------|
| Intercept | 34199.1 | 626.5 | 54.583 | <2e-16 |
| Bing | -2217.7 | 140.0 | -15.846 | <2e-16 |
| Election Rhetoric | -3749.8 | 870.7 | -4.306 | 1.66e-05 |
| Presidential Opposition | 2330.5 | 876.6 | 2.659 | 0.00785 |
| Workforce Rhetoric | -4354.1 | 881.1 | -4.942 | 7.76e-07 |

Key takeaways: Bing and Presidential Opposition values

Thank you!