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

Aidan Ohlson - 405815712

2025-03-21

Introduction

Over the past decade, social media—Twitter (now X) in particular—has played a key role in political communication. President Donald Trump's Twitter account became a major source of political news, controversy, and engagement, providing us with a large volume of tweets to analyze. In this report, we investigate several core questions about the content, sentiment, and engagement of Trump's tweets, and how they vary over time and context.

Key Research Questions
- How does sentiment in Trump's tweets change over time and context?
- Are there key topics he speaks on, and how do sentiments shift within those topics?
- How does the Twitter public respond (in terms of likes and retweets) to different sentiments and topics?
- Do more extreme sentiments or certain topics generate higher or lower engagement?
- Can we predict engagement (favorites/likes) from sentiment and topic features?

We divide time into three eras:

1. **Non Presidency**: 2009–2015 and 2021–2023
2. **Pre Presidency**: 2016 and 2024
3. **Active Presidency**: 2017–2020

These categories reflect distinct contexts: Before holding office or running to do so, the year before an active presidency, and during a presidency.

Data Overview

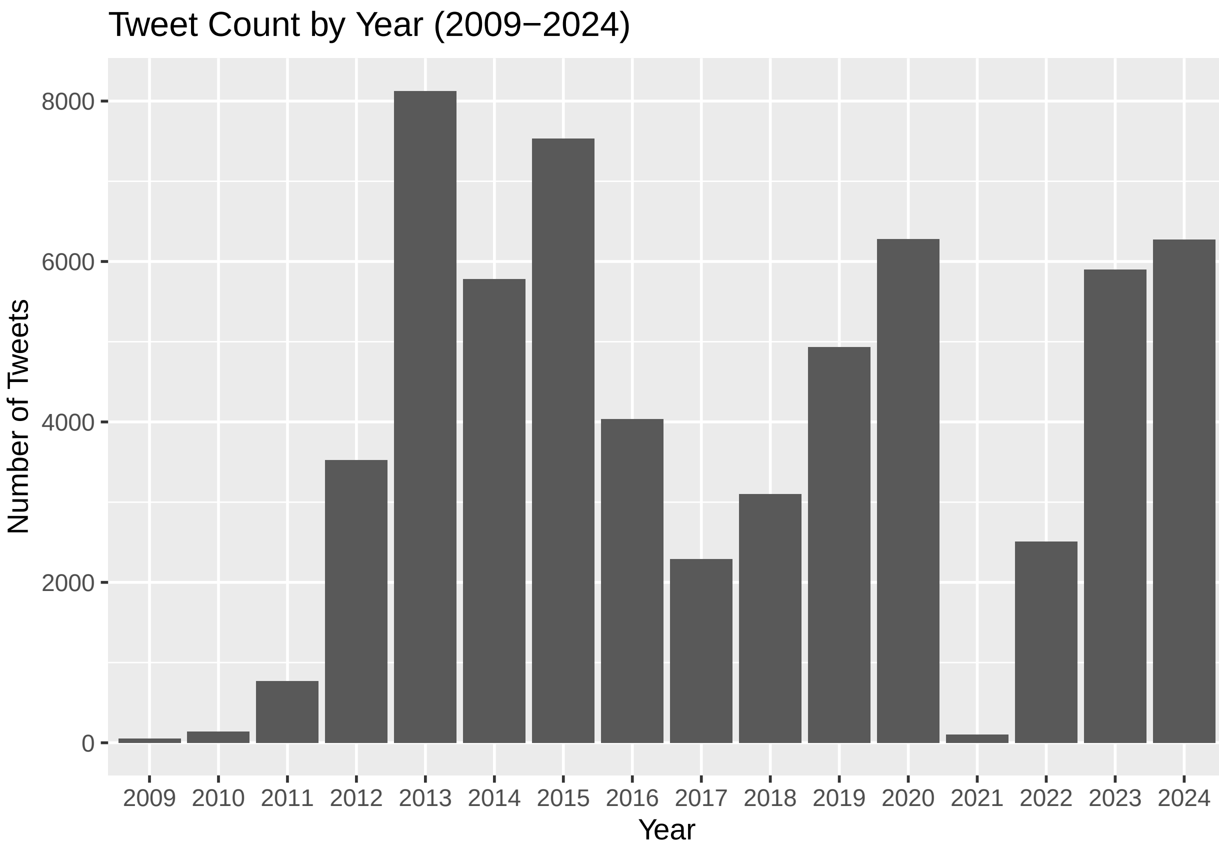
The dataset consists of Trump's tweets from 2009–2024, compiled from publicly available Twitter data (provided by the professor). Each row represents a single tweet with these main columns:

- **Date**: When the tweet was posted
- **Text**: The content of the tweet (cleaned)
- **Favorites**: Number of likes
- **Retweets**: Number of retweets
- **is_president** or **period** indicators for the three time contexts
- **Other** (e.g., a numeric `bing` sentiment score, assigned topics, etc.)

Data Cleaning & Preprocessing
- Converted text to lowercase
- Removed punctuation and numbers
- Removed English stopwords (`tm` and `tidytext`)
- Stripped excess whitespace
- Stemmed words (using `SnowballC`)

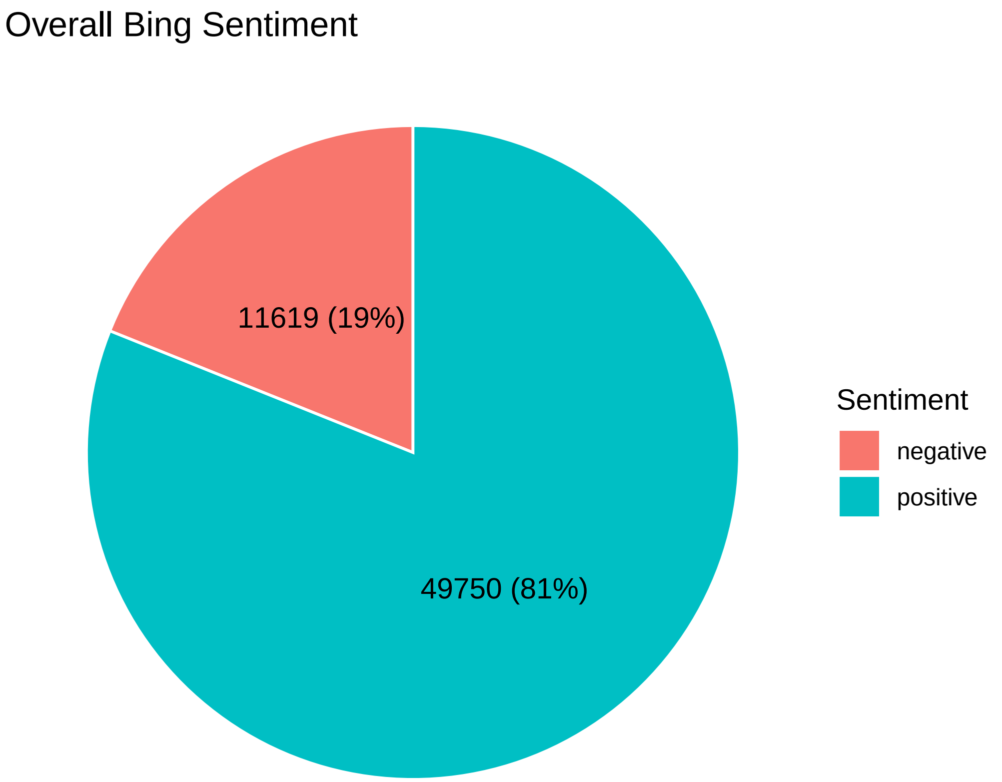
This ensures a cleaner corpus for our text-mining, sentiment analysis, and topic modeling.

After cleaning, we note that average likes is around 30,000 and average retweets is around 7,000. The tweets are then partitioned into Non Presidency, Pre Presidency, and Active Presidency based on the year. Additionally, we then provide a plot showing number of tweets per year, for additional contextual understanding.

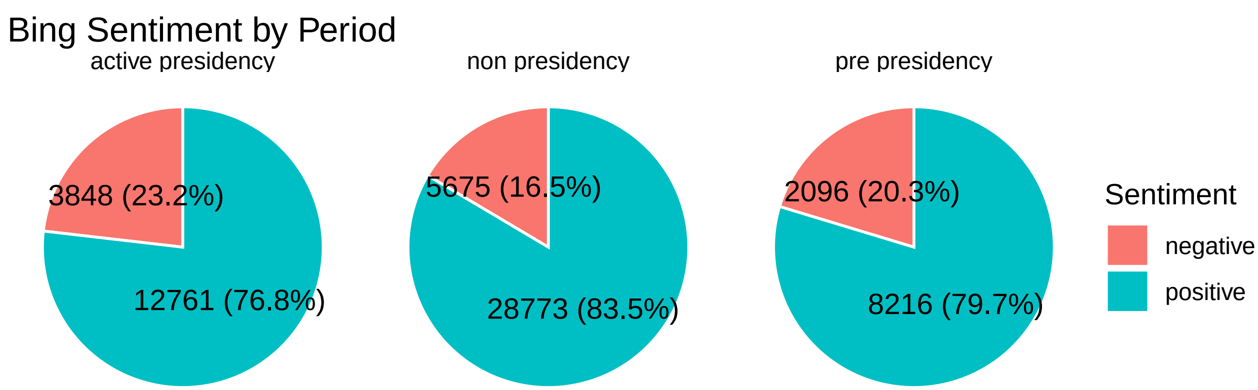


Sentiment Analysis

We applied **Bing** lexicon (positive/negative) and **NRC** lexicon (emotions) to each tweet's text.



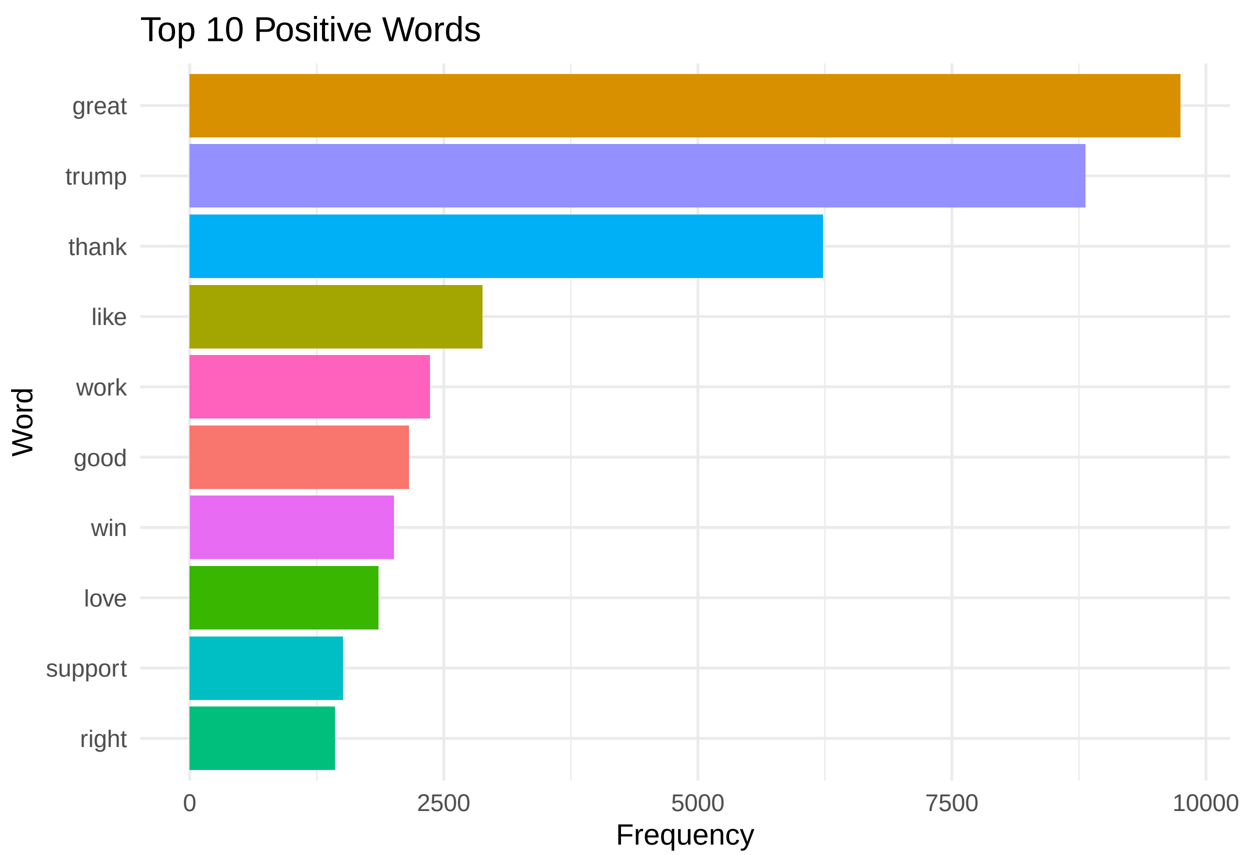
Overall Bing Distribution
Roughly 80% of tweets scored as "positive," suggesting that single words like "great," "winning," or "love" are frequently used. Negative tweets formed about 20%.

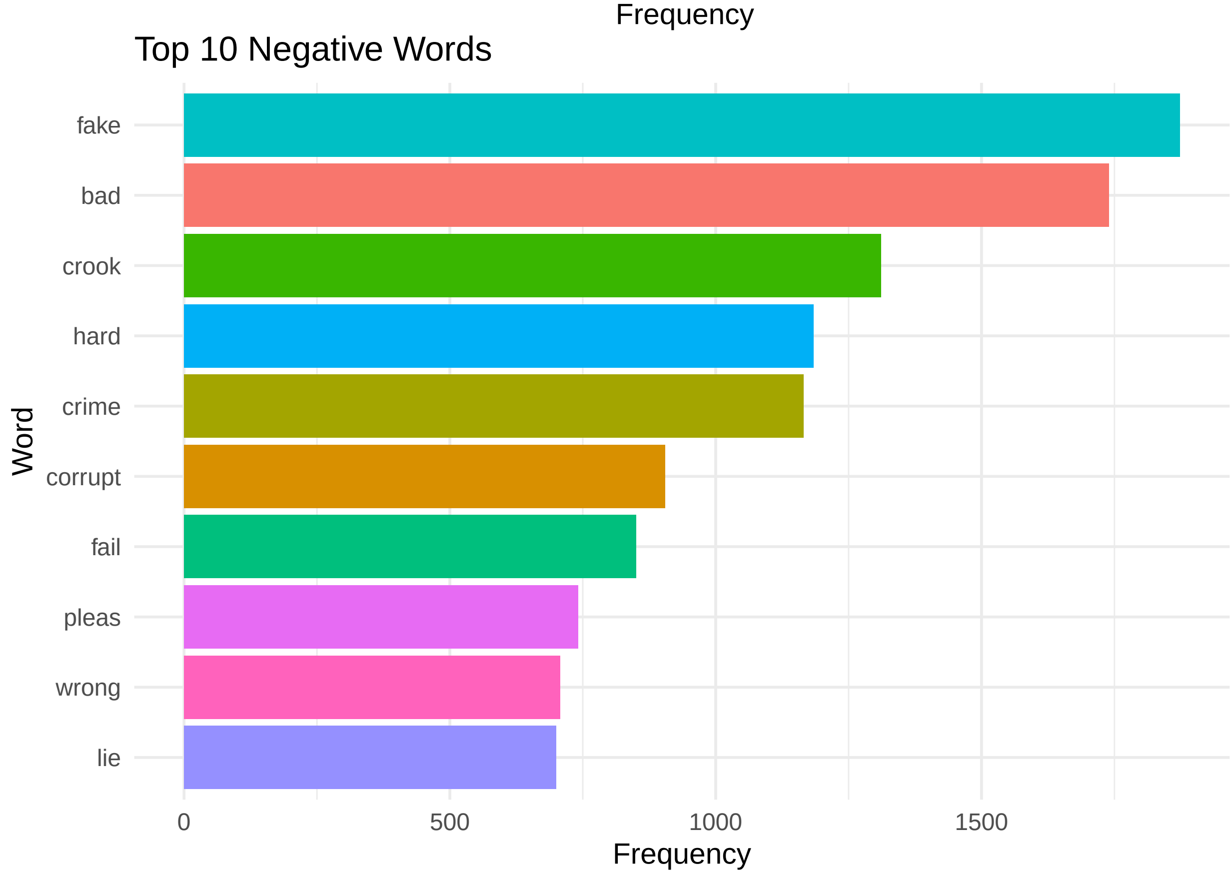


Sentiment by Context
- **Non Presidency**: Noticeably fewer negative tweets.
- **Pre Presidency**: Slight increase in negativity.
- **Active Presidency**: Negative tweets spike further, possibly reflecting controversies or more combative language.

Most Frequent Positive/Negative Words

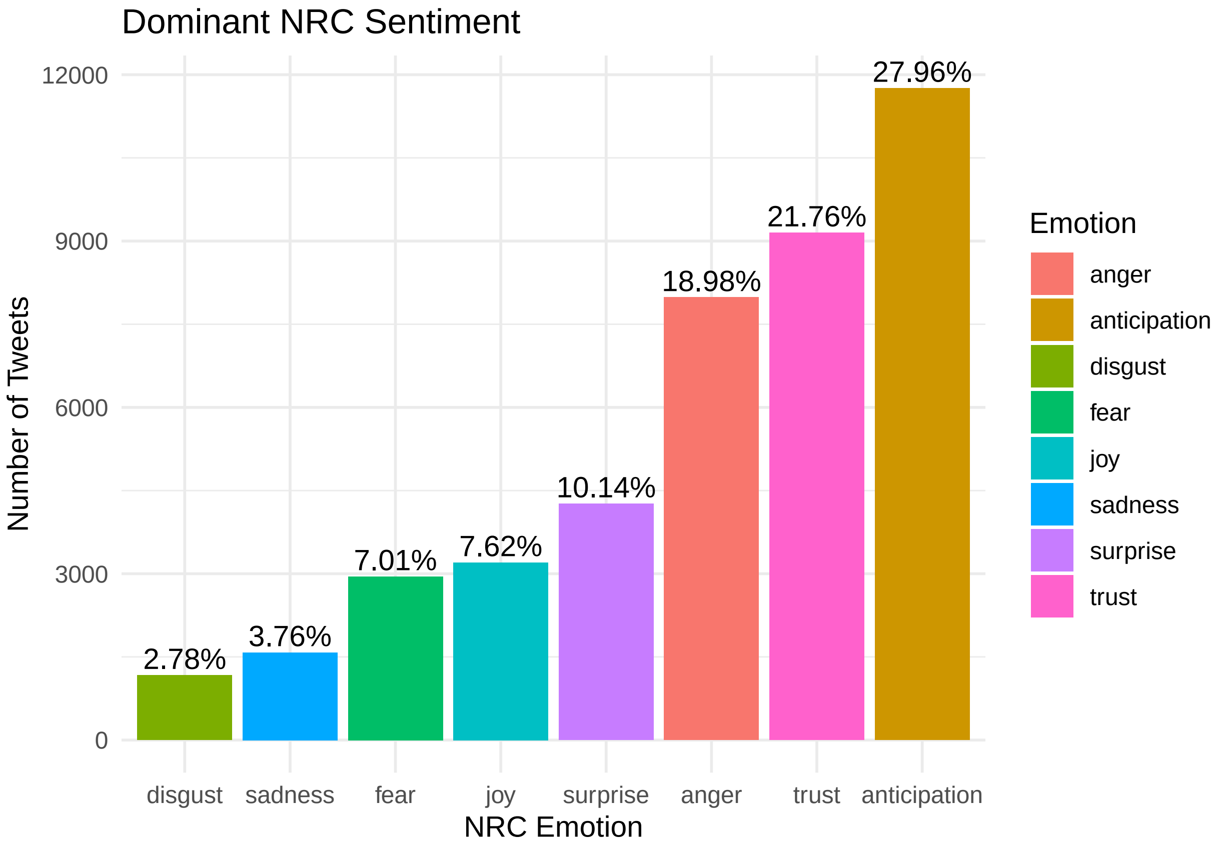
To see which terms contributed most to positivity or negativity, we examined the top 10 tokens each:


Top 10 Positive Words
Word
great
trump
thank
like
work
good
win
love
support
right
0
2500
5000
7500
10000
Frequency

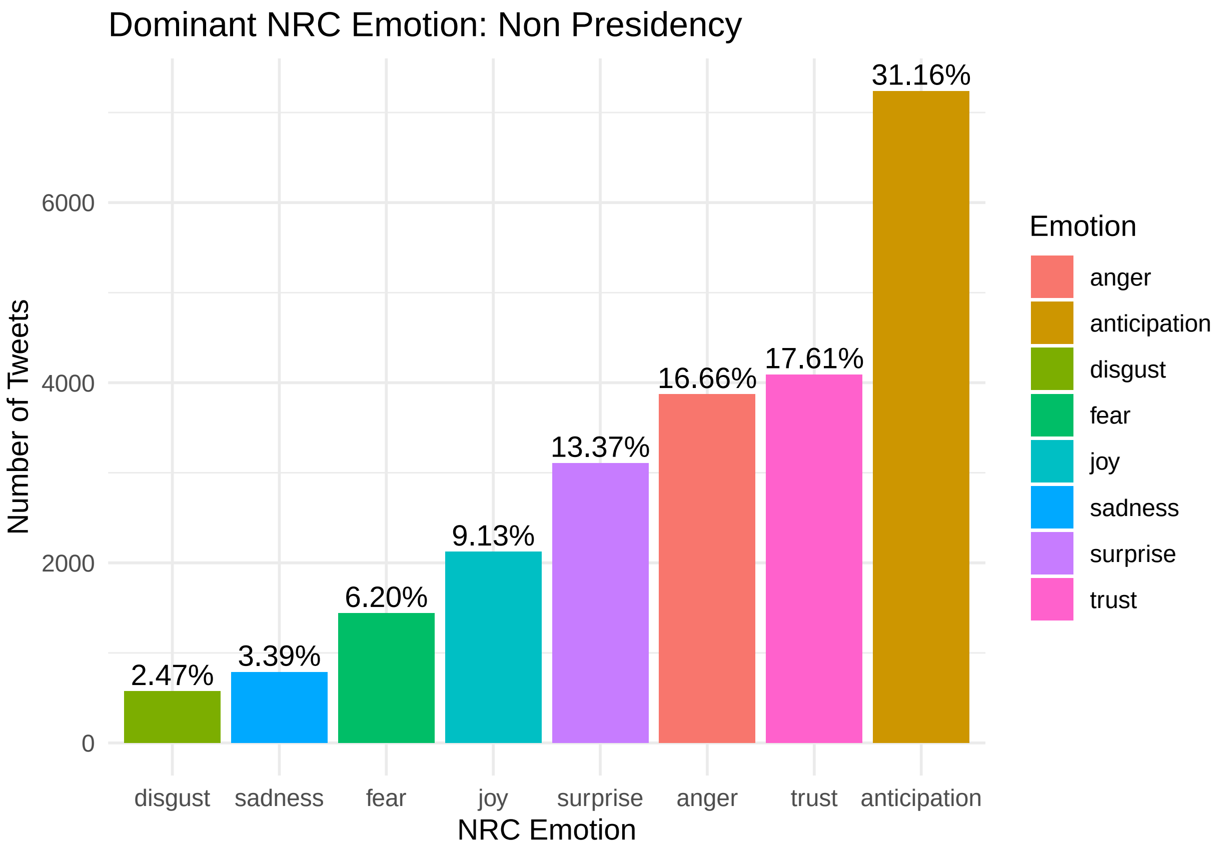

Top 10 Negative Words
Word
fake
bad
crook
hard
crime
corrupt
fail
pleas
wrong
lie
0
500
1000
1500
Frequency

NRC (Emotion) Sentiment

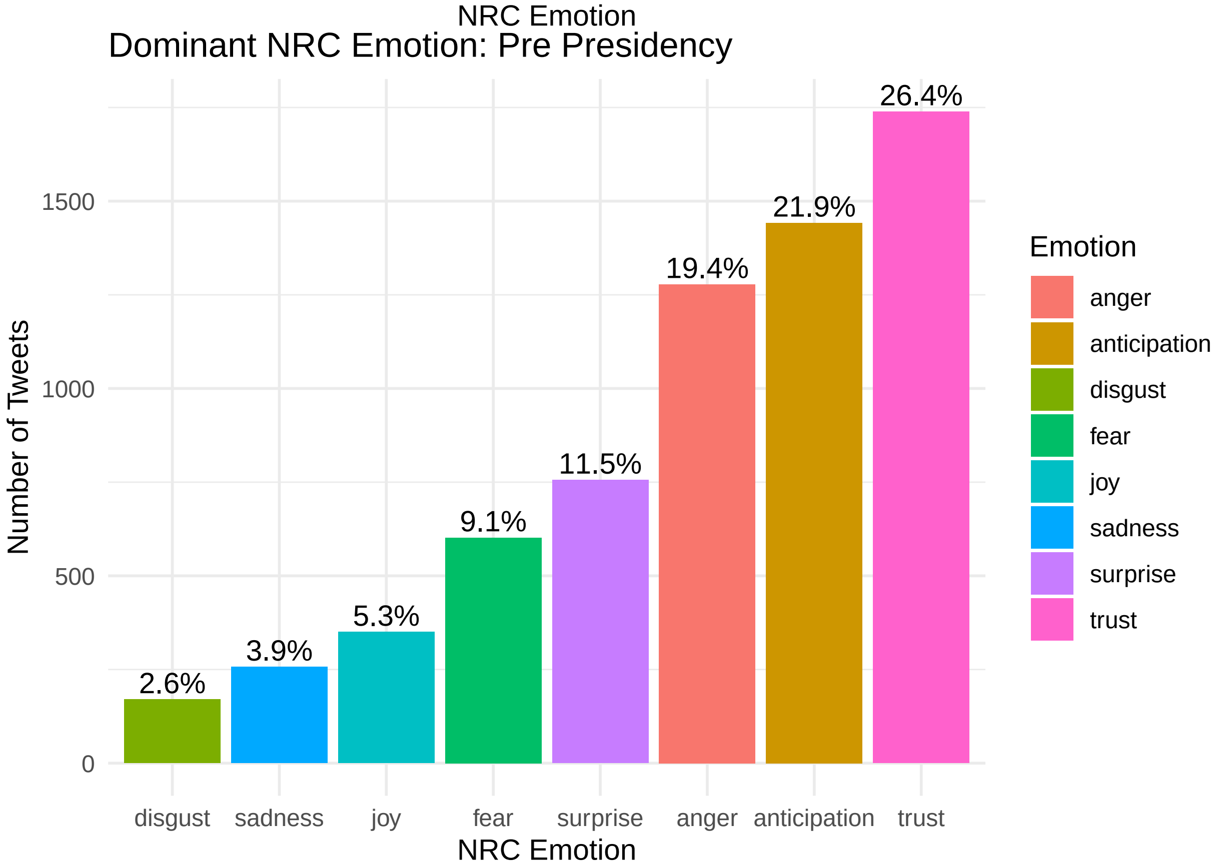
Using the NRC lexicon, each word is mapped to emotional categories like anger, joy, fear, sadness, trust, and so on. Each tweet is labeled by the dominant emotion (the emotion with the highest word count in that tweet).



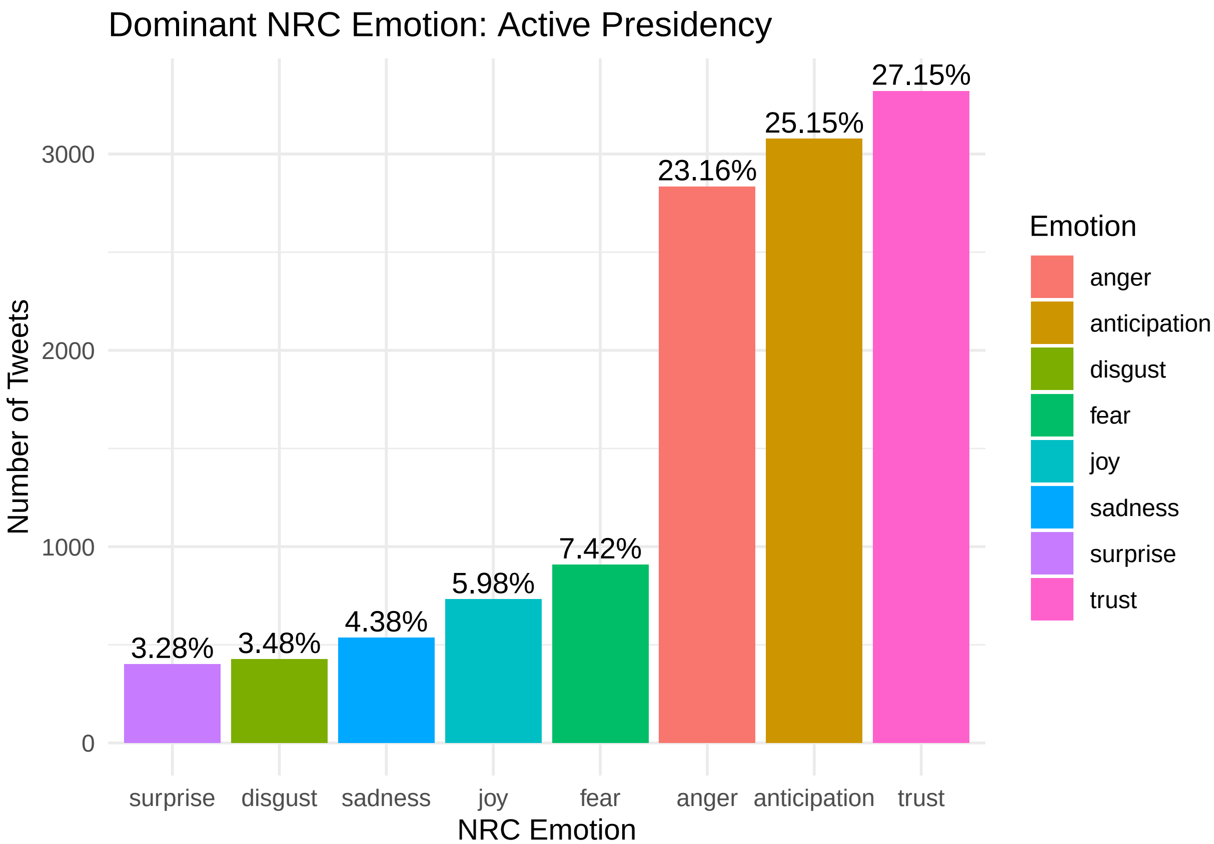
- **Overall**: Anticipation, trust, and anger were dominant categories, with moderate counts in surprise, joy, and fear.



Dominant NRC Emotion: Non Presidency
Number of Tweets
0
2000
4000
6000
2.47%
3.39%
6.20%
9.13%
13.37%
16.66%
17.61%
31.16%
disgust
sadness
fear
joy
surprise
anger
trust
anticipation
NRC Emotion
Emotion
anger
anticipation
disgust
fear
joy
sadness
surprise
trust



Dominant NRC Emotion: Pre Presidency
Number of Tweets
0
500
1000
1500
2.6%
3.9%
5.3%
9.1%
11.5%
19.4%
21.9%
26.4%
disgust
sadness
joy
fear
surprise
anger
anticipation
trust
NRC Emotion
Emotion
anger
anticipation
disgust
fear
joy
sadness
surprise
trust

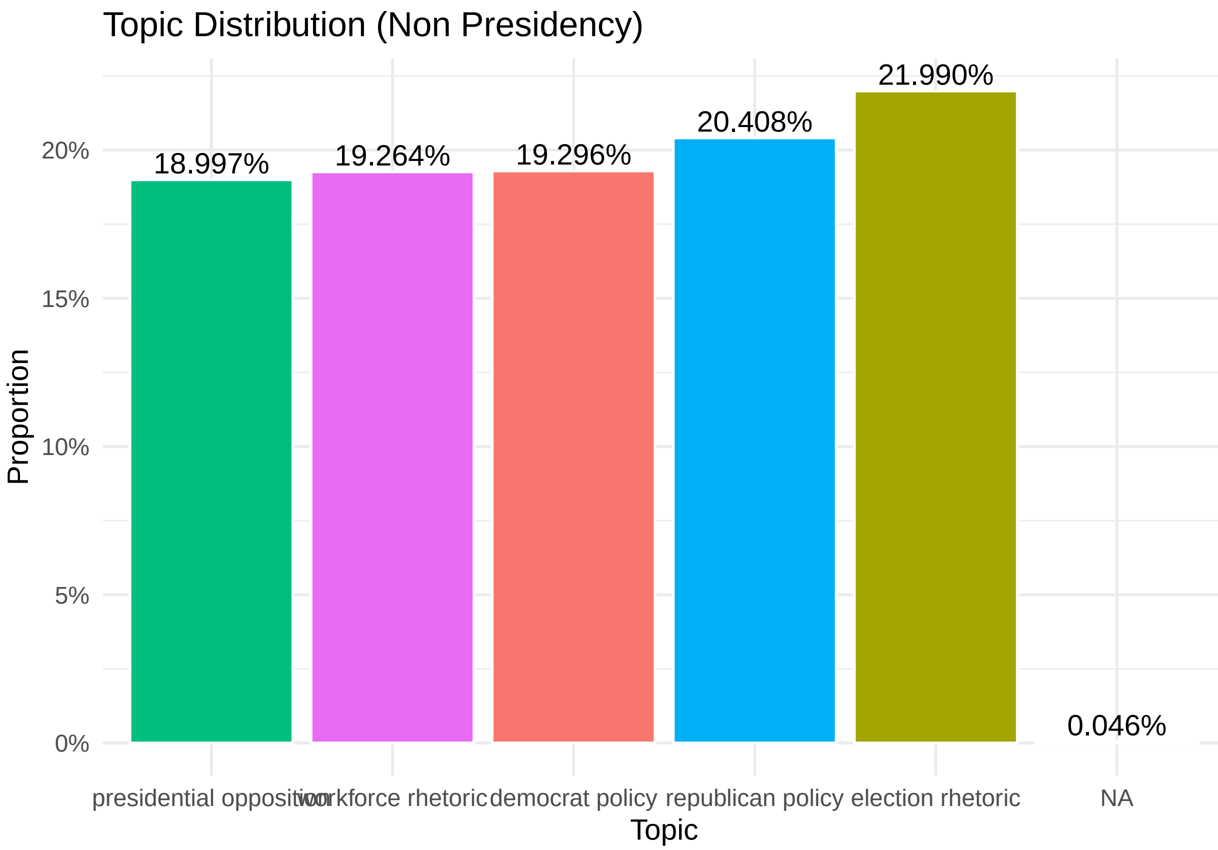


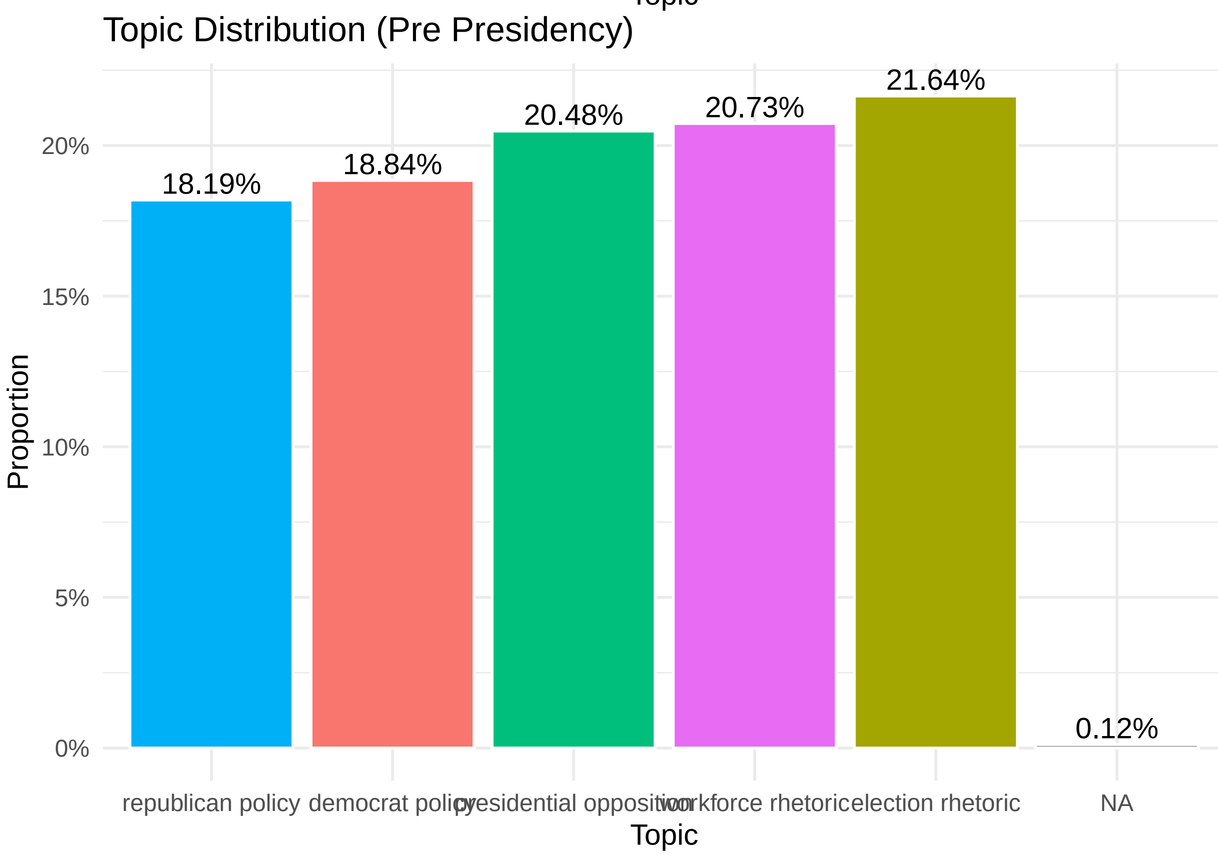
By Context: During the active presidency era, anger- and fear-related words increased, while non-presidency periods show more joy/trust.

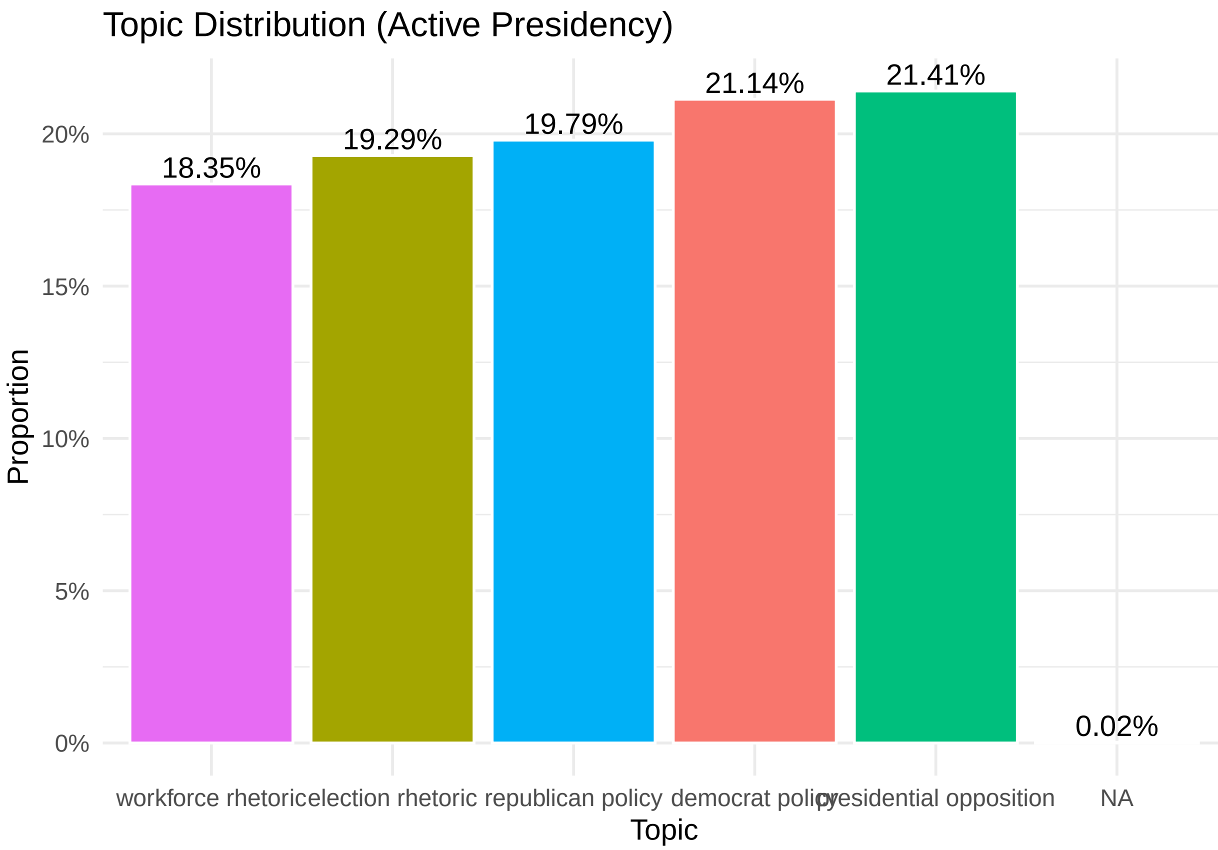
Analyzing Topics with LDA

We used Latent Dirichlet Allocation (LDA) to identify key themes (topics) in Trump's tweets. After removing very common words (e.g., "amp," "will," "trump") and building a document-term matrix, we ran an LDA with $k = 5$ topics.

1. **democrat policy**
2. **presidential opposition**
3. **republican policy**
4. **workforce rhetoric**
5. **election rhetoric**


Topic Distribution (Non Presidency)
Proportion
Topic
0%
5%
10%
15%
20%
18.997%
19.264%
19.296%
20.408%
21.990%
0.046%
presidential opposition
workforce rhetoric
democrat policy
republican policy
election rhetoric
NA


Topic Distribution (Pre Presidency)
Proportion
Topic
0%
5%
10%
15%
20%
18.19%
18.84%
20.48%
20.73%
21.64%
0.12%
republican policy
democrat policy
presidential opposition
workforce rhetoric
election rhetoric
NA



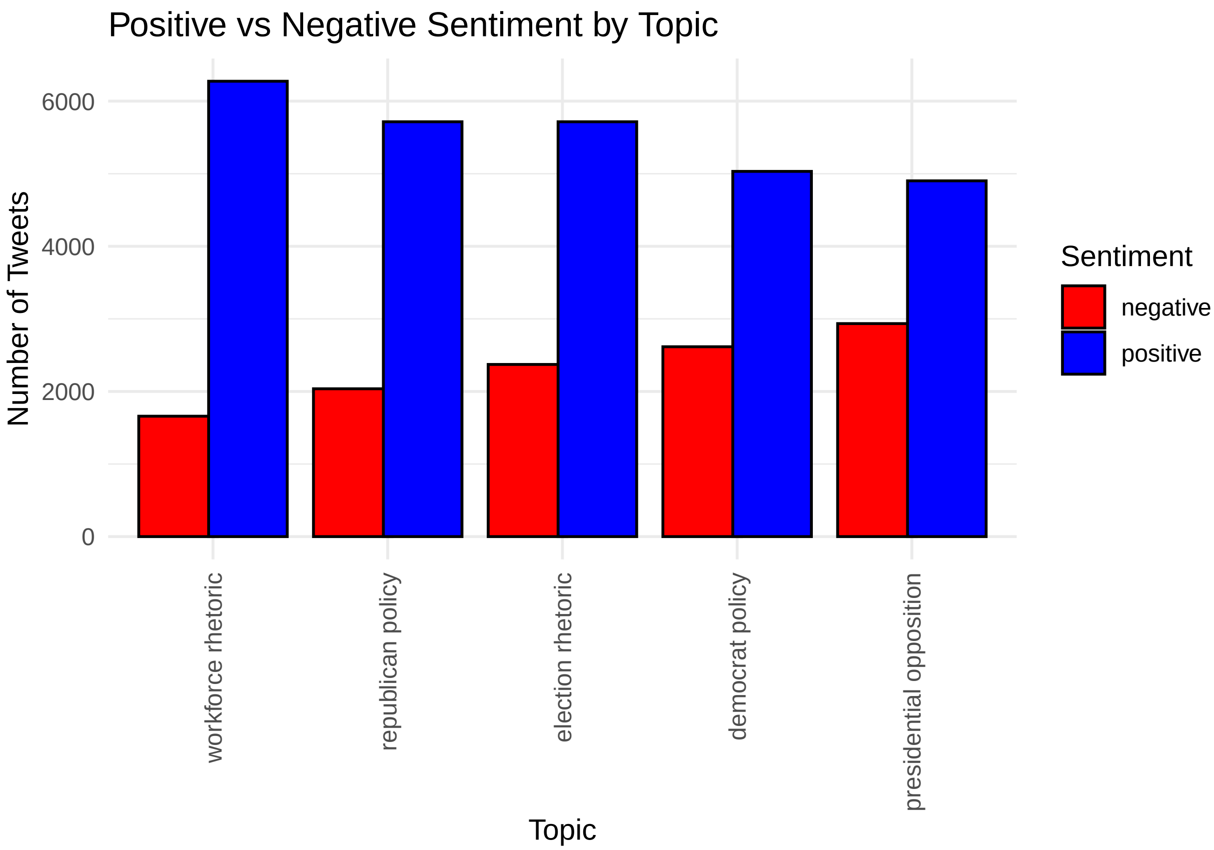
Topic Distribution by Eras

We observed how these five topics vary in proportion across the three time periods: Trump speaks more about presidential opposition and democrat policy than anything else when he is in active presidency. This shifts to election rhetoric and workforce rhetoric in the year leading up to his presidential terms. Interestingly, republican policy is only dominant in years which he was not politically relevant.

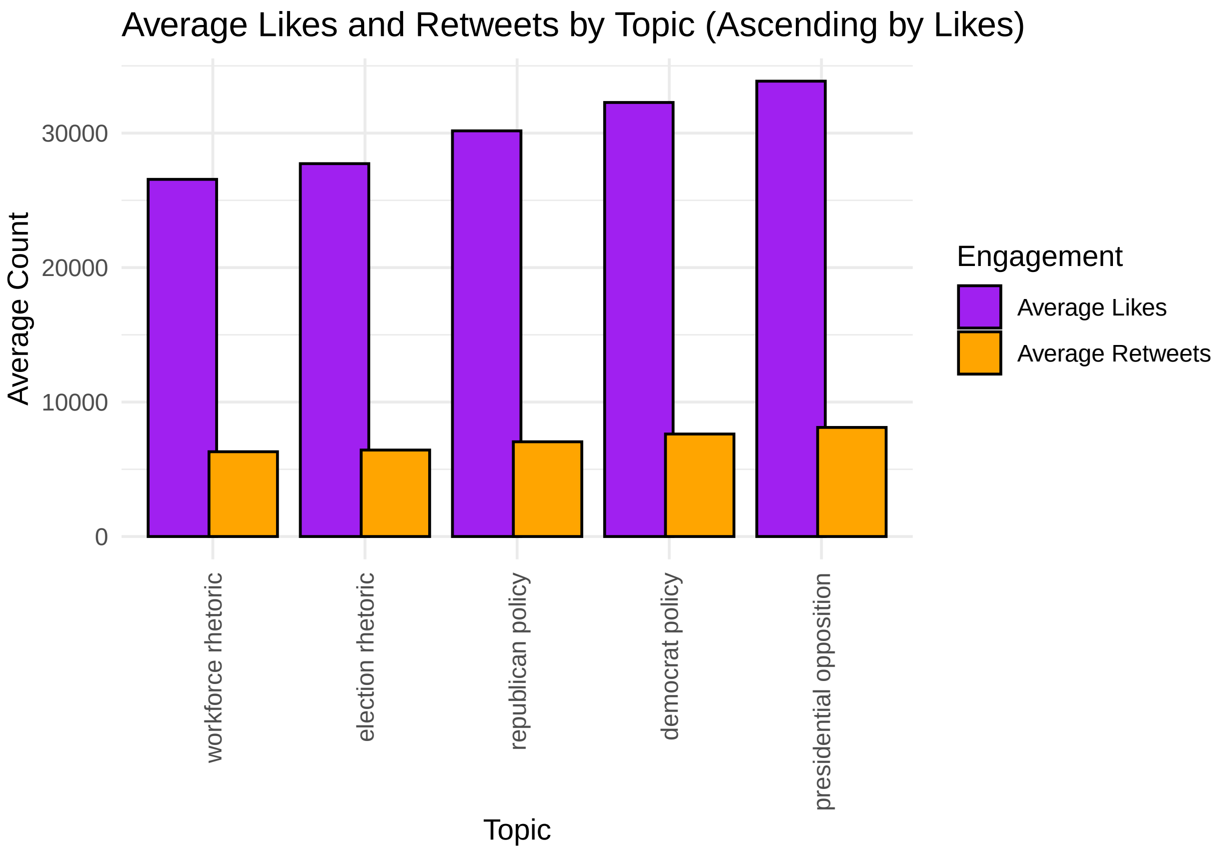
Connecting Sentiment and Topics to Engagement

We next examined how favorites (likes) and retweets vary by sentiment and topic.

Positive vs. Negative



Interestingly, topics referring to Trump's opponents, or differing viewpoints tended to amass many more negative tweets than topics referring to his own policy or the american workforce.

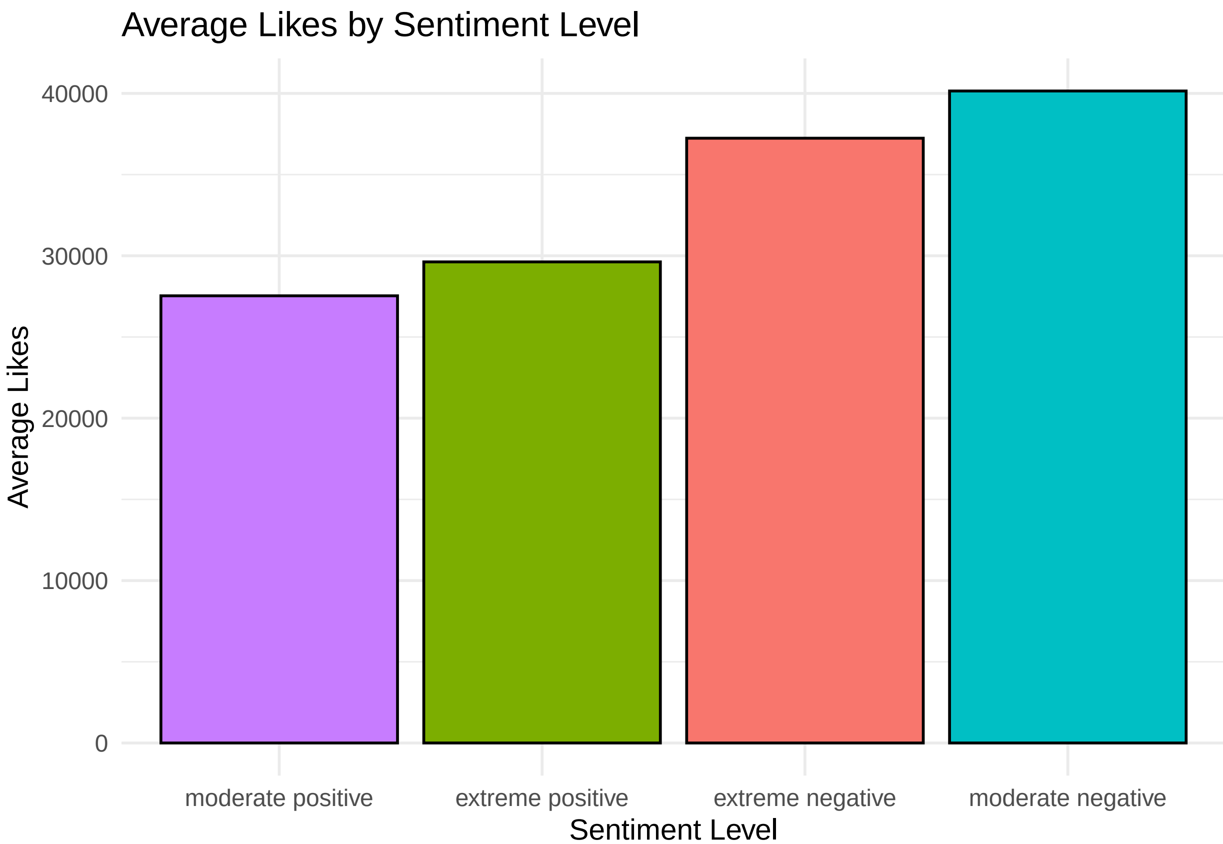


Connecting these results to the last plot, we can see that there is a positive correlation between the number of negative tweets in a topic and the average likes a certain topic gets. The more negative tweets, the more likes a topic is expected to get. This exemplifies the issue of social media tending to inflate hurtful and negative rhetoric.

Extreme vs. Moderate Sentiment

We further classified tweets based on the **Bing** score magnitude:

- **Moderate Negative**: bing between -5 and -1
- **Moderate Positive**: bing between 1 and 5
- **Extreme Negative**: bing $<$ -5
- **Extreme Positive**: bing $>$ 5



We find that negative tweets in general drive higher engagement. Trump is expected to get around ten thousand more likes on average if he posts something negative versus posting something positive. This further shows how negative sentiment and touchy topics drive engagement.

Predictive Modeling

Finally, we attempted to predict **favorites (likes)** using the numeric Bing score and the assigned topic as predictors.

Example Linear Regression

A multiple linear regression of:

$$\text{Likes} \sim \text{BingScore} + \text{TopicName}$$

yielded the following significant predictors:

- **Bing**: Coefficient is -2218, suggesting strongly negative correlation with likes.
- **Election Rhetoric**: Coefficient is -3750 (tweets about elections often have fewer likes).
- **Presidential Opposition**: +2330 (opposition tweets may rally supporters, boosting likes).
- **Workforce Rhetoric**: -4354 (topics focusing on jobs or workforce appear to get fewer likes).

These results further show the trend of this analysis, increased negativity and focus on opposition tend to yield Trump a higher like count.

Conclusions and Future Directions

Our analysis of Trump's tweets reveals several core insights:

1. **Sentiment**: The majority of tweets are positive by word count, though negative or combative tweets increased markedly during the presidency era.

2. **Topics**: Five main topics emerged. "Election rhetoric" and "presidential opposition" spiked in certain periods, aligning with real-life events.

3. **Engagement**: Tweets with more negative sentiment receive higher likes, suggesting that negative content fuels engagement.

4. **Predictive Modeling**: Simple regressions show both topic and sentiment significantly shape favorites, highlighting the interplay between content and engagement.

Our findings underscore that tweets with negative sentiment and opposition-related topics consistently attract higher engagement. By analyzing average likes across a range of sentiment scores and topic categories, we observed that tweets framed around criticism or contentious issues—particularly those targeting political opponents—tend to generate more interaction from followers and detractors alike. This pattern aligns with the broader tendency on social media for emotionally charged or provocative content to garner disproportionate attention, reflecting how polarized or "touchy" subjects can fuel heightened user response on the Twitter/X platform.

In future work, we might incorporate more refined sentiment scoring (e.g., transformer-based classifiers), compare retweets vs. likes for a fuller picture, or investigate how Twitter policies and timeline algorithms influenced the visibility of certain tweets.
