

Goal

Our goal is to analyze various aspects of former President Donald Trump's tweets given his particular influence on American politics. We aim to examine:

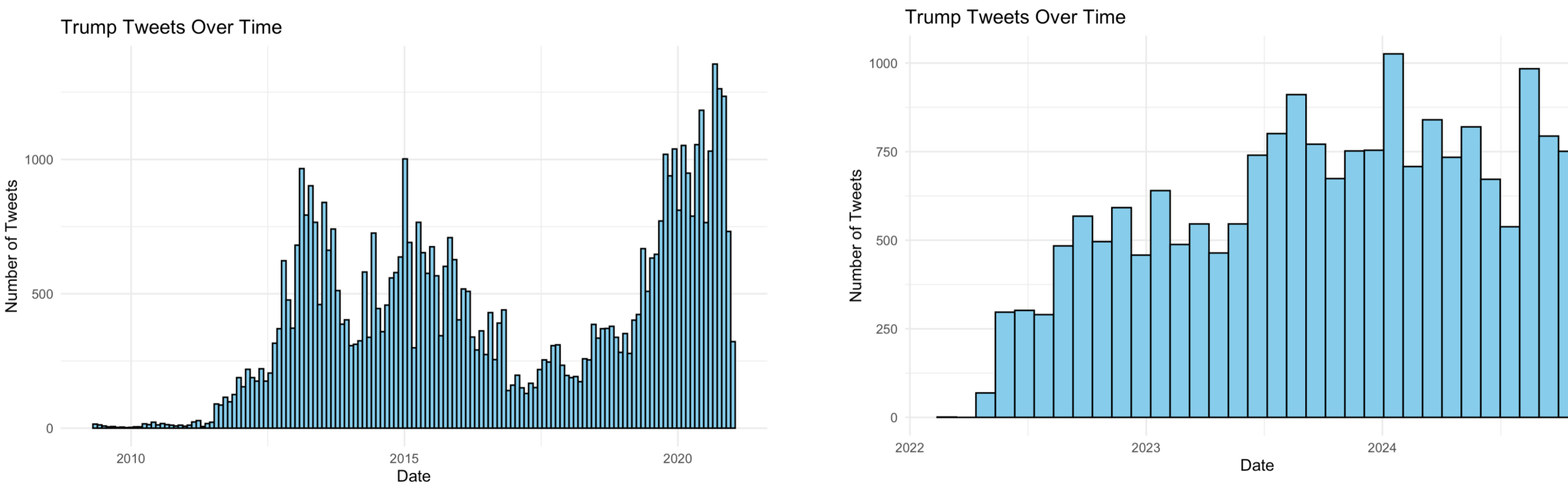
- His frequency of tweets over different time periods.
- His flagged versus non-flagged tweet favorite counts.
- His average tweet length between flagged and non-flagged tweets.

Additionally, we will apply text-based analysis to develop a model that estimates the probability of Trump's post-2021 tweets being flagged, based on patterns observed in his pre-2021 tweets. This analysis is conducted within the context of Elon Musk's acquisition of Twitter and Trump's subsequent unbanning.

Exploratory Data Analysis & Hypothesis Testing

Using our visualizations from Exploratory Data Analysis, we are able to draw some hypotheses and test them:

- **Hypothesis 1:**
 - H_0 : Number of tweet frequency doesn't go up during election season.
 - H_a : Number of tweet frequency goes up during election season.



(a) Number of tweets over time BEFORE BAN

(b) Number of tweets over time AFTER BAN

The two histograms show that Trump tends to tweet more frequently around election season. A Welch's Two-Sample t-test confirmed this with a significant p-value, supporting Trump tweets more during election season.

- **Hypothesis 2:**
 - H_0 : The distribution of favorites is the same for flagged and non-flagged tweets.
 - H_a : Trump's flagged tweets have more favorites than non-flagged tweets.

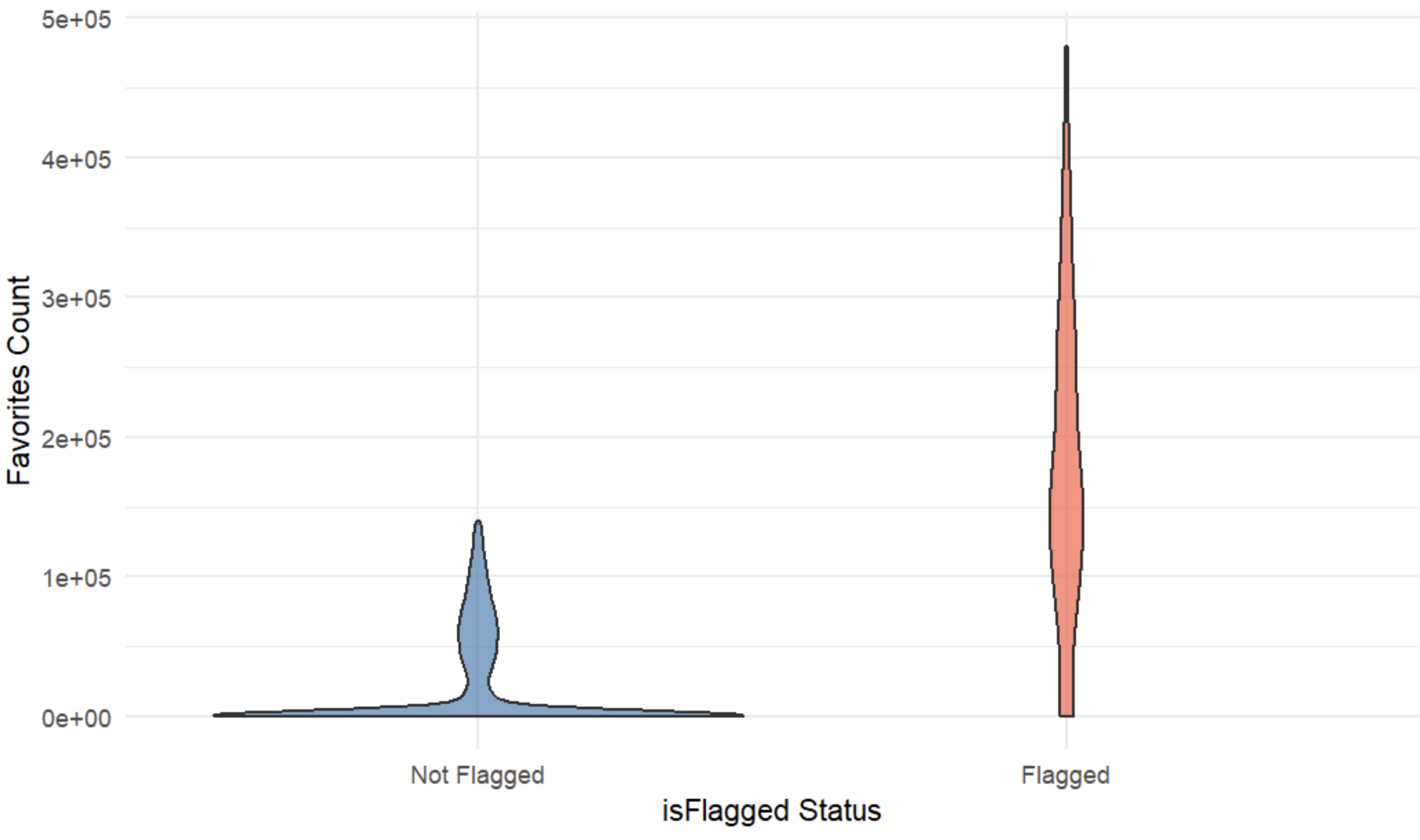


Figure 1. Comparison of Favorites Count between Flagged & Non-Flagged Tweets

The violin plot shows that flagged tweets have a wider range and higher favorite counts compared to non-flagged tweets. A Wilcoxon signed-rank test confirmed this with a significant p-value, supporting the conclusion that flagged tweets generally receive more favorites.

Hypothesis Testing (Continued)

- **Hypothesis 3:**
 - H_0 : Trump's pre-2021 flagged tweets are not longer than his normal tweets.
 - H_a : Trump's pre-2021 flagged tweets are longer than his normal tweets.

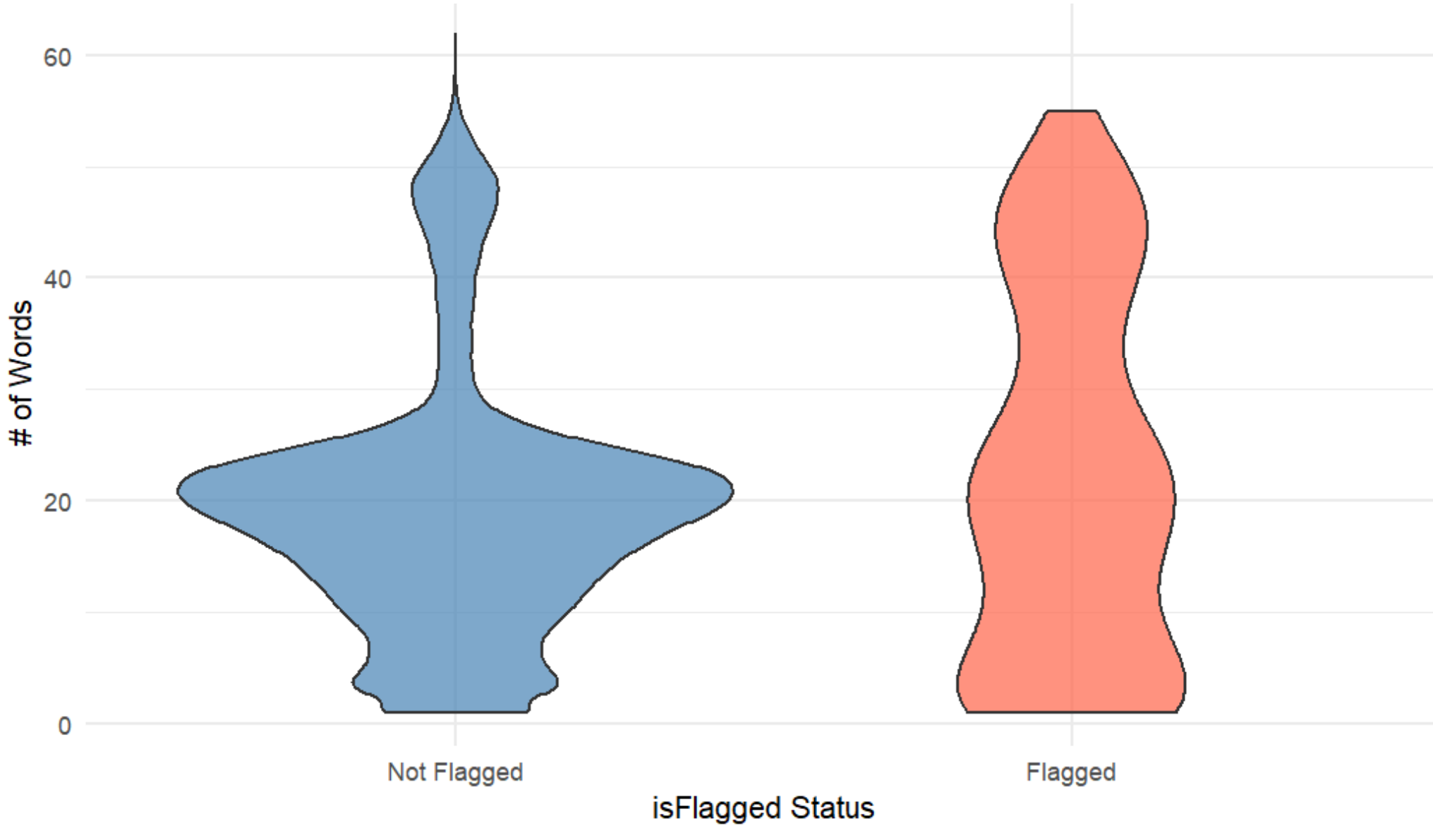


Figure 2. Comparison of Word Count between Flagged & Non-Flagged Tweets

The violin plot shows that flagged tweets tend to have more words compared to non-flagged tweets. A Welch's Two-Sample t-test confirmed this with a significant p-value, supporting the conclusion that flagged tweets are generally longer.

Classification Model

In our modeling, we had to account for the textual nature of the data and the imbalanced Flagged to non-Flagged observations.

To tackle the imbalance between Flagged and non-Flagged observations in our dataset, we applied sampling techniques such as oversampling, undersampling, and SMOTE.

We then tested several models, including Logistic Regression, XGBoost, and Random Forest. However, all models struggled with high accuracy due to overtraining on non-Flagged data, while precision and recall for Flagged observations remained low.

Adding sentiment analysis as a feature had minimal impact on performance. Ultimately, we opted for a simpler approach: TF-IDF (Term Frequency-Inverse Document Frequency). This method captures how often terms appear in tweets while accounting for their overall frequency across the dataset, delivering similar results to the more complex models.

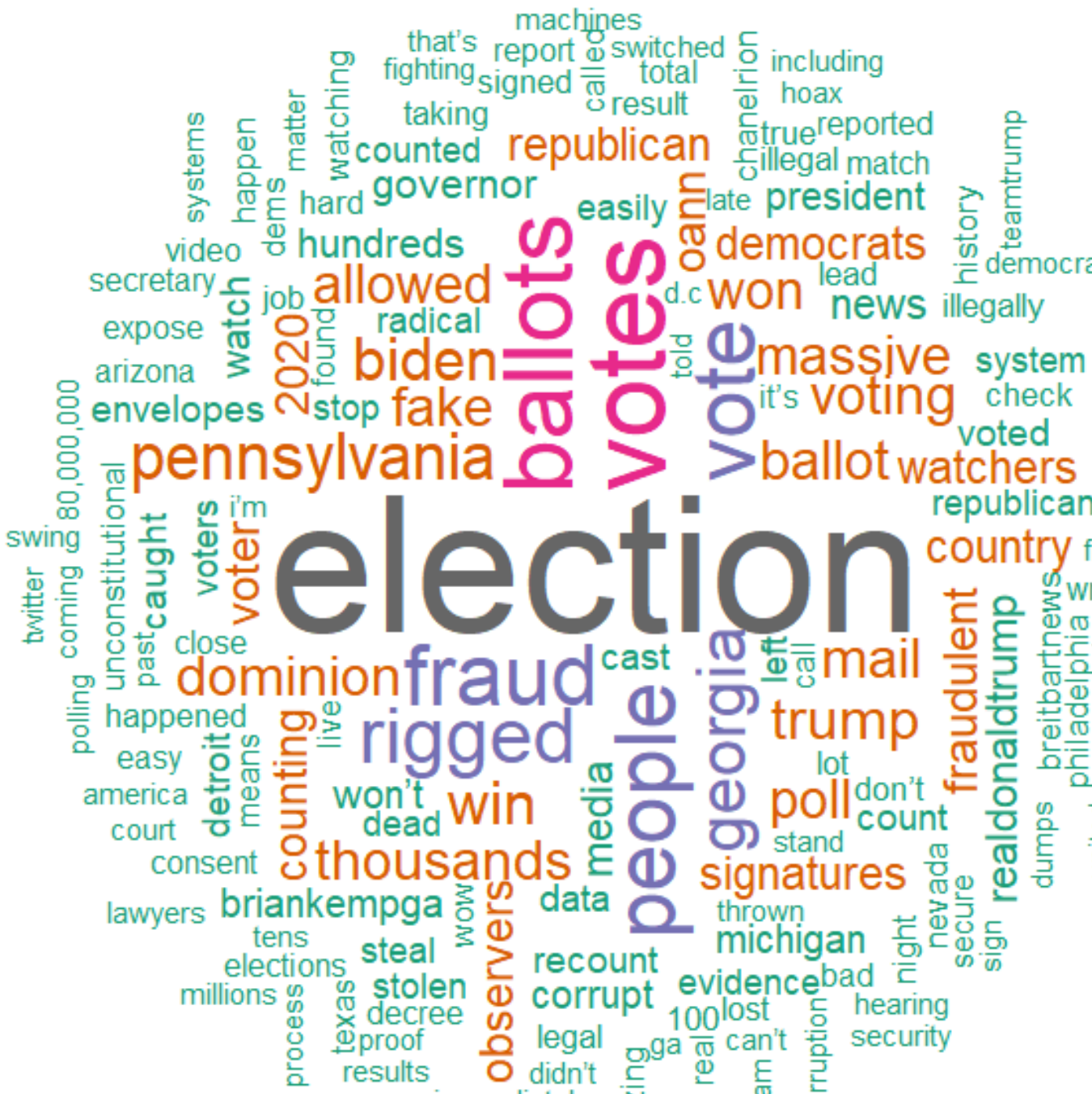


Figure 3. Trump's Favorite Flagged Words

Algorithm

We separated the dataset based on the isFlagged status, applying TF-IDF to each subset. The parameters were set to extract a maximum of 3,000 features, with each feature being a unigrams, bigrams, or trigrams (1–3 words in length). For instance, phrases like "rigged election" were identified as significant features. This process generated weighted representations of words and phrases specific to both flagged and non-flagged data. By comparing the two sets, we identified the terms present in the flagged set but absent from the non-flagged set.

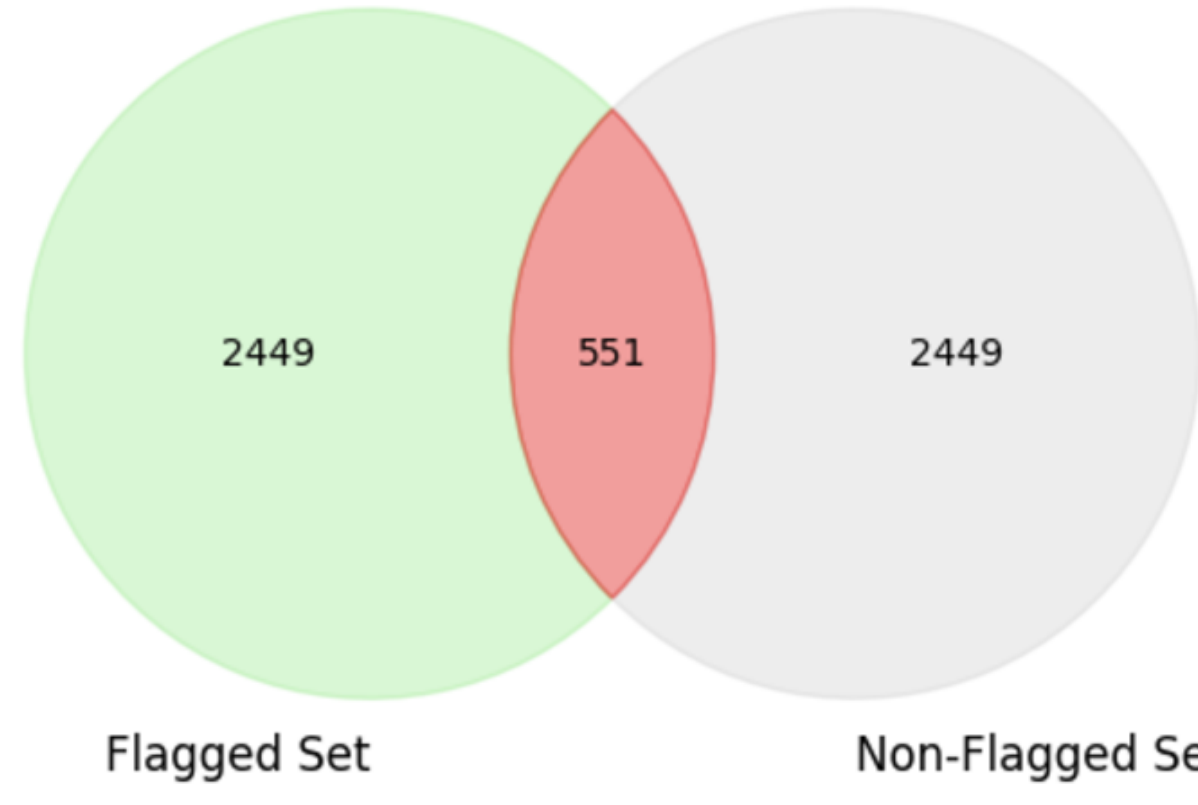


Figure 4. We keep the green and remove the red overlap.

These distinctive words and phrases were then used to classify tweets as flagged or non-flagged.

Model Analysis

After running our model on the original dataset, we created a frequency table to show actual vs. prediction.

	Predicted: No Flag	Predicted: Flag
Actual: No Flag	53,677	2,590
Actual: Flag	120	184

Table 1. Confusion Matrix

Model Performance Summary:

- **Accuracy:** 95.48% of total predictions were correct, indicating a high overall prediction rate.
- **True Positives (Flagged Tweets):** Only 6.63% of the flagged tweets were correctly identified as positive by the model.
- **Recall:** The model successfully identifies 60.53% of the actual flagged tweets.
- **F1 Score:** The low F1 score of 11.91% suggests that while the model detects some flagged tweets, it also produces many incorrect predictions.

Though these are not the best results, they are comparable to other complicated models applied to the training data.

Conclusions

When applied to post-2021 data, our model predicted **17,305 non-flagged tweets and 2,206 flagged tweets**. While these results align with our current efforts, the model is not the optimal solution given the constraints of time and resources.

In future iterations, we would prioritize addressing sample imbalance through improved sampling methods and explore more computationally intensive models to better capture nuanced patterns in the tweets.

It would also be nice to implement the Topic Models that we did come up with but did not have the time to use.