Dynamics of Content, Sentiment, and Theme in the Presidency

Aidan Ohlson

2025-03-18

Abstract

The language used in presidential speeches reflects the priorities and communication styles of individual presidents as well as broader political and societal trends. This project applies text mining techniques to analyze Inaugural Addresses and State of the Union speeches from U.S. Presidents, comparing trends across two distinct eras (1960s-1980s and 2000s) and between political parties. By analyzing word usage, sentiment, and thematic content, this study examines shifts in rhetoric, emotions, and dominant topics over time. Additionally, predictive modeling techniques are used to determine whether speeches can be classified more accurately by era or by party.

Introduction

Presidential rhetoric plays a fundamental role in shaping national identity, influencing policy directions, and impacting public perception. Over time, speech content and tone have evolved due to societal changes, media influences, and political realignments. This study investigates how language in presidential speeches has changed between the 60s-80s and the 2000s and whether party affiliation is a stronger differentiator than era.

The key research questions addressed in this project are:

- How has presidential language evolved across different eras?
- Are there discernible differences in speech content based on party affiliation?
- What sentiments are most prevalent in speeches, and how do they vary across time and party lines?
- Do certain presidents exhibit similar rhetorical styles, and are these similarities stronger within eras or parties?
- Can predictive models effectively classify speeches based on era or party?

```
## <<DocumentTermMatrix (documents: 16, terms: 4359)>>
## Non-/sparse entries: 13227/56517
## Sparsity           : 81%
## Maximal term length: 66
## Weighting          : term frequency (tf)
## Sample             :
##     Terms
## Docs america american can nation new now peopl will world year
##   1       18       17  16     14   3   5     8   34     8    4
##   10      20       26  12     18  15  13    17   66     9   28
##   12      31       22  18     28  33  19    24   32    22   26
##   14       4       17  19     40  16  12    13   28    24   30
##   16       5       23  19     21  12  15    25   47    22   26
##   2       43       55  36     22  19  12    28   36    25   38
##   4       30       55  20     11  21  13    32   38    11   30
##   5       10        5  13     15  11   6     8   21     8    3
##   6       24       46  23     28  20  34    33   61     6   48
```

```
##   8       33       23   7     18  13  10    14   49    19    4
```

Data and Preprocessing

Data Collection

The dataset consists of presidential Inaugural Addresses and State of the Union speeches sourced from UCSB's *The American Presidency Project*. The speeches analyzed were selected from the first year and first term of each president, spanning eight administrations:

- **60s-80s Era**: JFK (D, 1961), Nixon (R, 1969), Carter (D, 1977), Reagan (R, 1981)
- **2000s Era**: Bush (R, 2001), Obama (D, 2009), Trump (R, 2017), Biden (D, 2021)

This selection ensures a balance of representation across different political parties and historical contexts, allowing for a meaningful comparative analysis.

Text Processing

To prepare the text for analysis, multiple preprocessing steps were applied:

- **Lowercasing**: All text was converted to lowercase to ensure uniformity in word counts and avoid duplication caused by case differences.
- **Punctuation and Number Removal**: Since punctuation does not contribute to textual meaning in this context, it was removed. Similarly, numbers were eliminated as they were not relevant to linguistic analysis.
- **Stopword Removal**: Common English stopwords (e.g., "the," "and," "is") were removed to focus on content-bearing words.
- **Whitespace Stripping**: Any excessive whitespace was eliminated.
- **Stemming**: Words were reduced to their base forms to unify variations (e.g., "running" $\rightarrow$ "run").
- **Sparsity Check**: Sparse terms that appeared in very few documents were removed to improve computational efficiency.

Following these steps, a cleaned corpus was generated for further analysis.

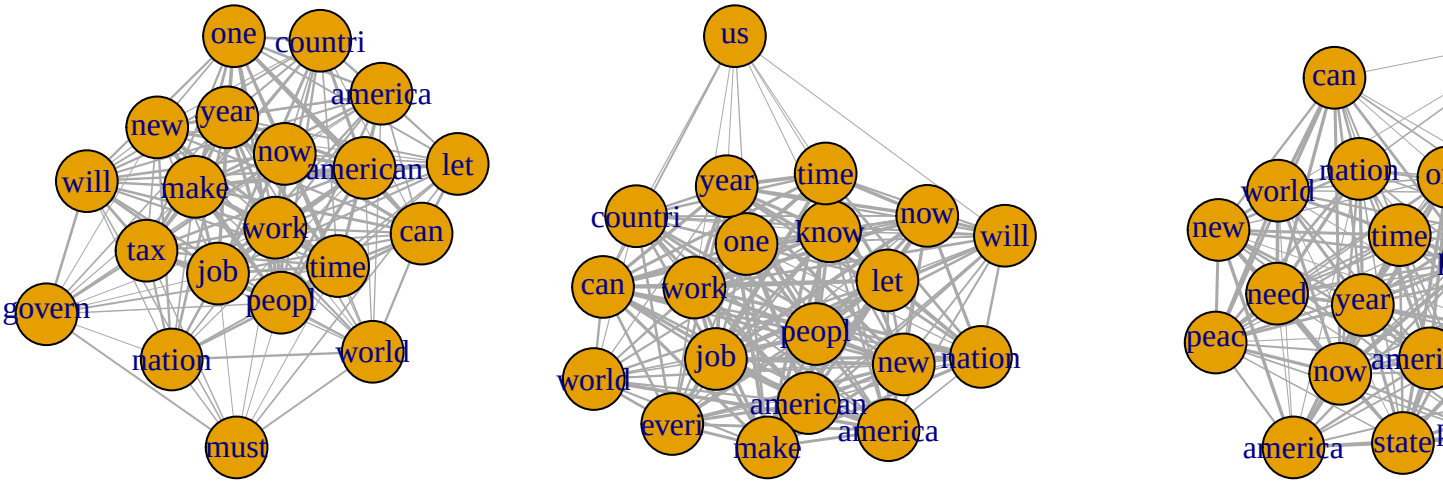
Exploratory Analysis

Word Frequency Analysis

```
## [1] 1908.75
```

```
## [1] 826.6875
```

```
## [1] 0.4849878
```



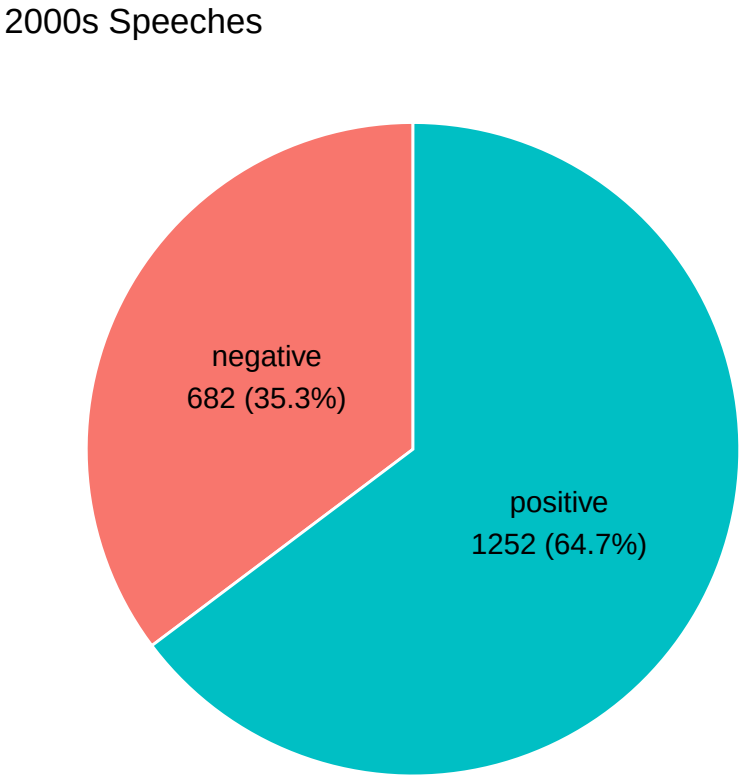
The frequency of word usage provides insight into common themes and rhetorical focuses in presidential speeches. Word clouds were generated to highlight the most frequently occurring words for different eras and political parties. This analysis revealed:

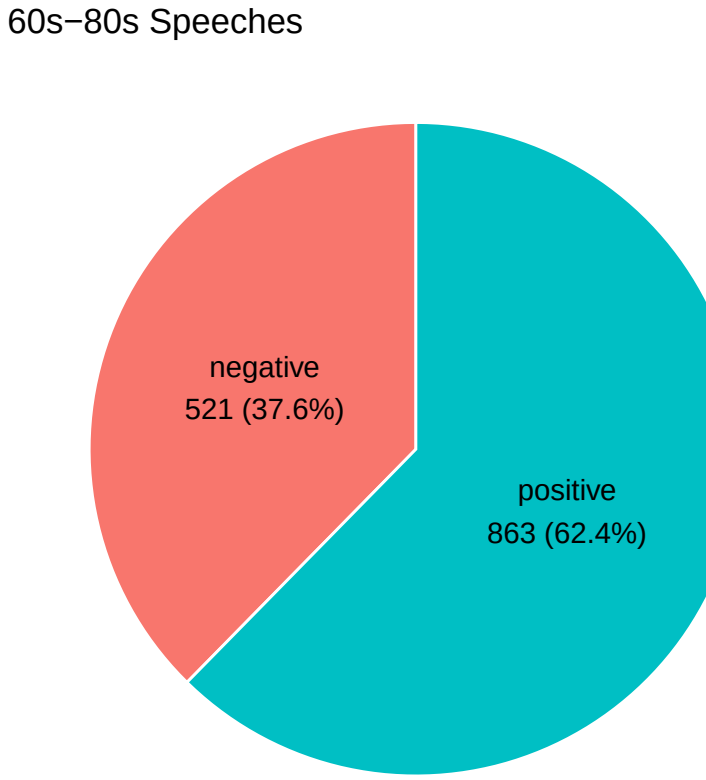
- **Frequent mentions of policy-related terms** such as "government," "people," and "nation" across all presidents.
- **Differences by era**: Earlier speeches emphasized concepts like "freedom" and "world," while more recent speeches included words related to security and economy, such as "jobs" and "terrorism."
- **Differences by party**: Democratic presidents tended to use words related to "hope" and "future," whereas Republican presidents frequently mentioned "security" and "taxes."

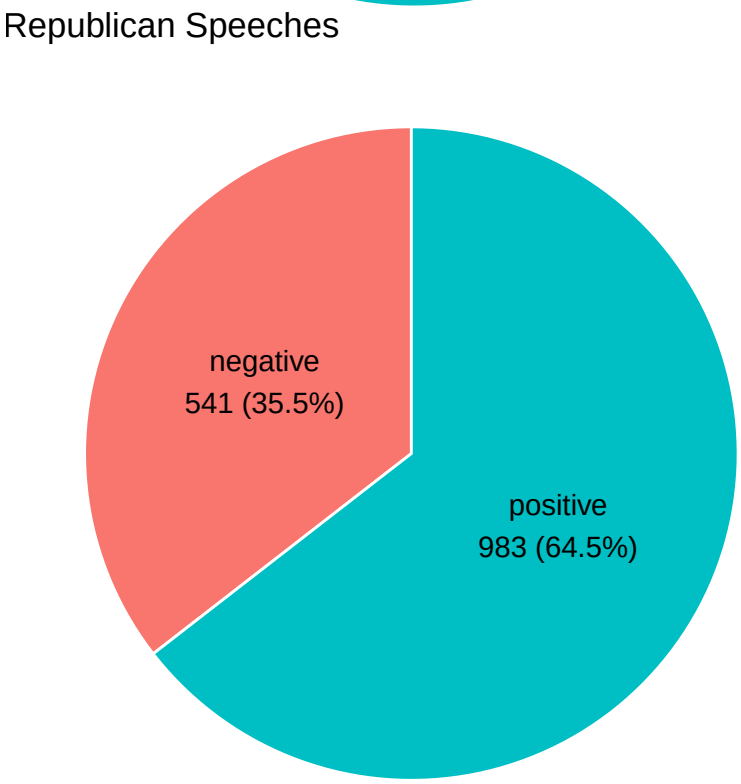
This initial exploration set the foundation for deeper sentiment and topic analyses.

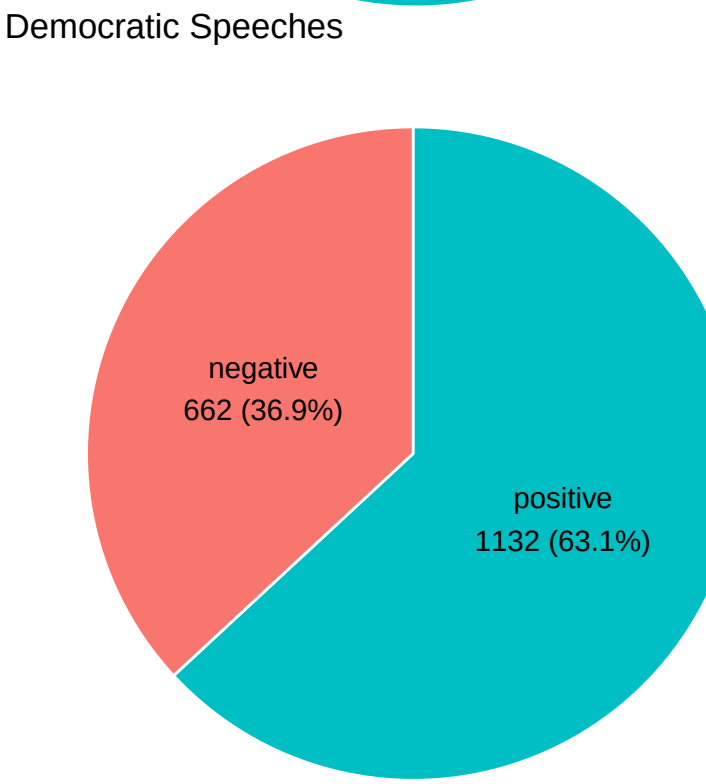
Sentiment Analysis

Sentiment analysis was conducted to evaluate the overall tone of the speeches and compare how sentiments varied by era and party affiliation.









Sentiment by Era

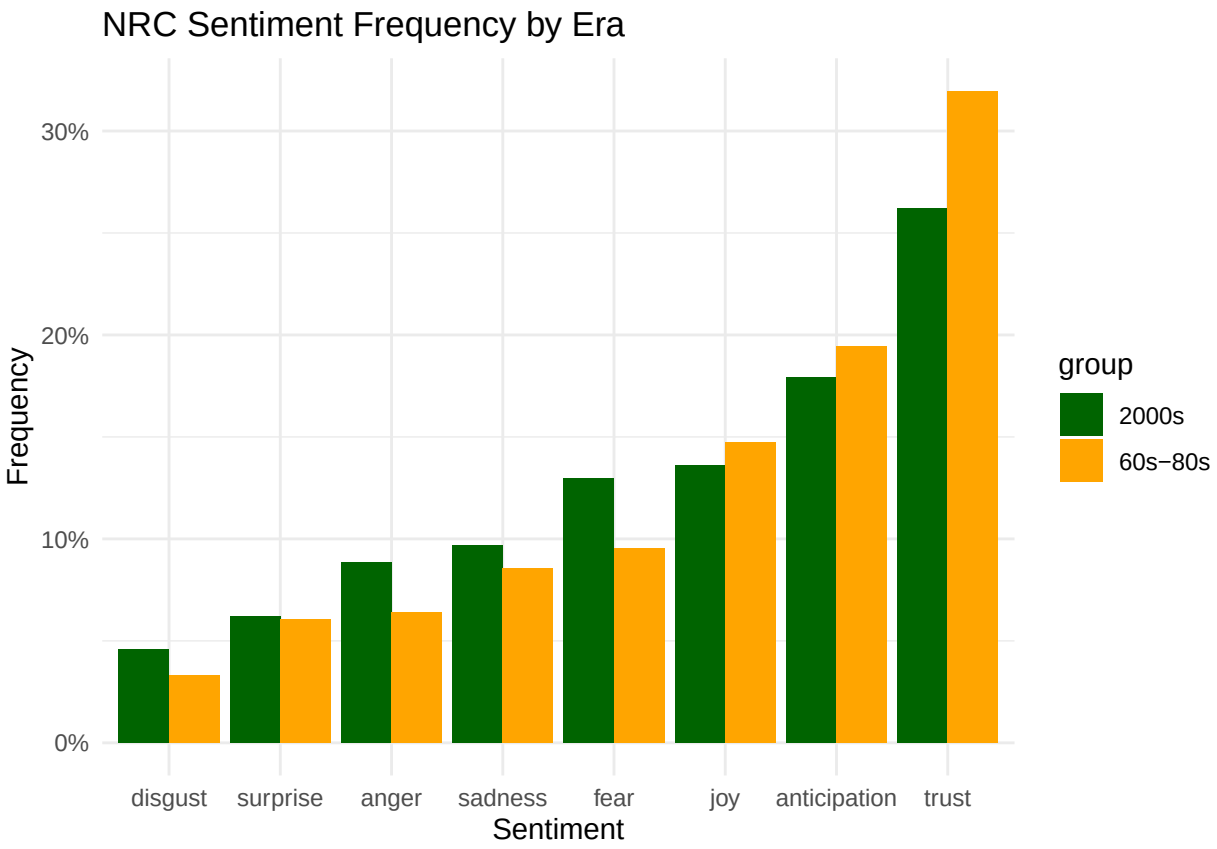
- **60s-80s Era**: 62.4% positive sentiment
- **2000s Era**: 64.7% positive sentiment

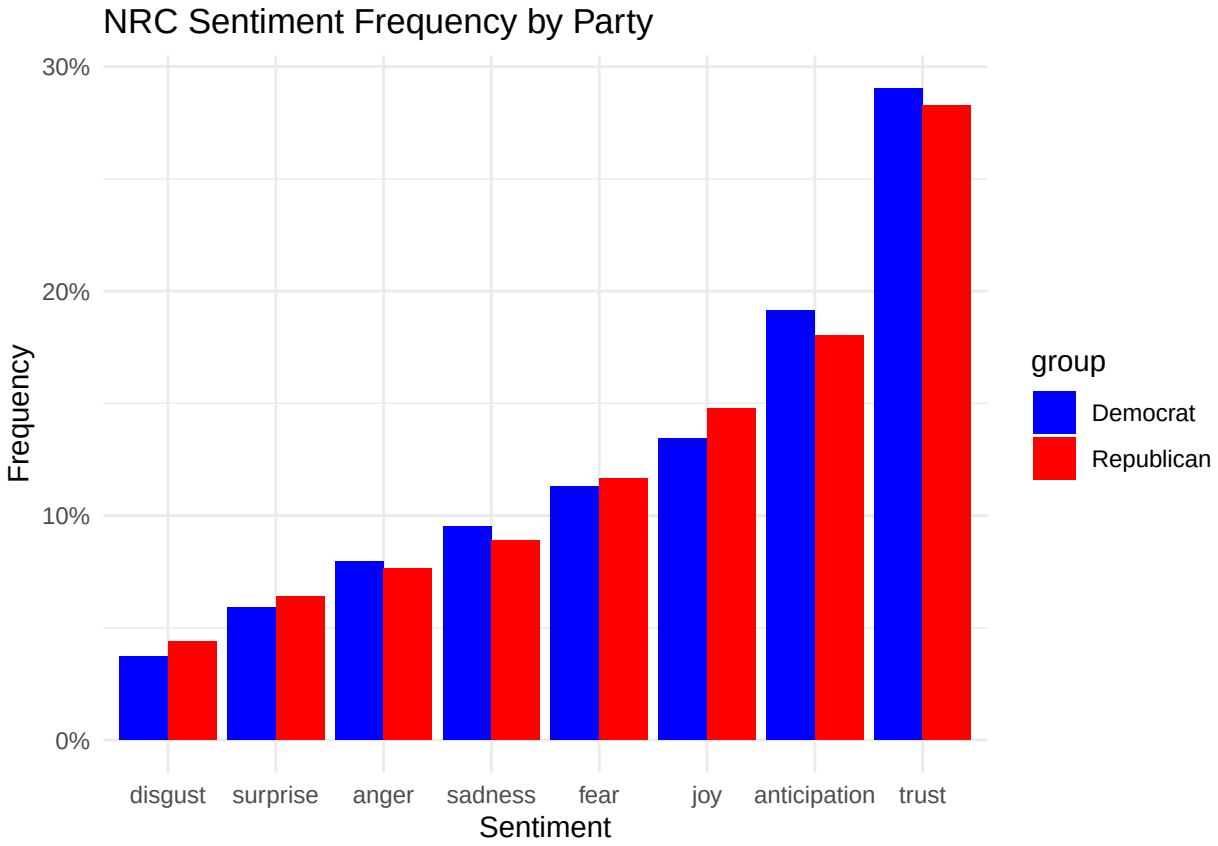
Despite overall positive tones, **fear and anger were more pronounced in the 2000s**, likely due to global security concerns, economic downturns, and increased political polarization.

Sentiment by Party

- **Democratic presidents**: 63.1% positive sentiment
- **Republican presidents**: 64.5% positive sentiment

While positivity levels were relatively close, **Republican speeches had slightly more emphasis on trust and patriotism**, while **Democratic speeches contained slightly higher anticipation-related words**.



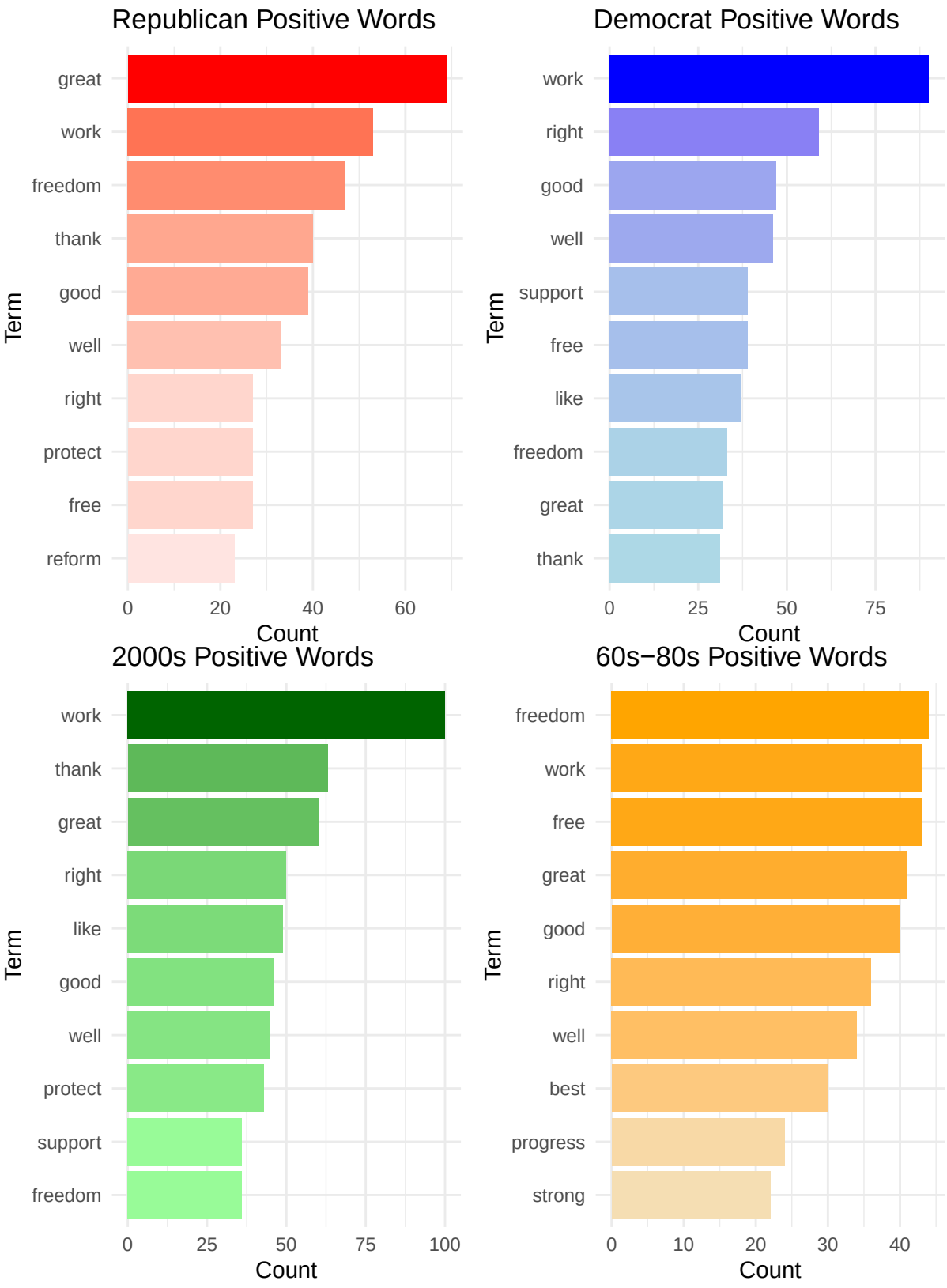


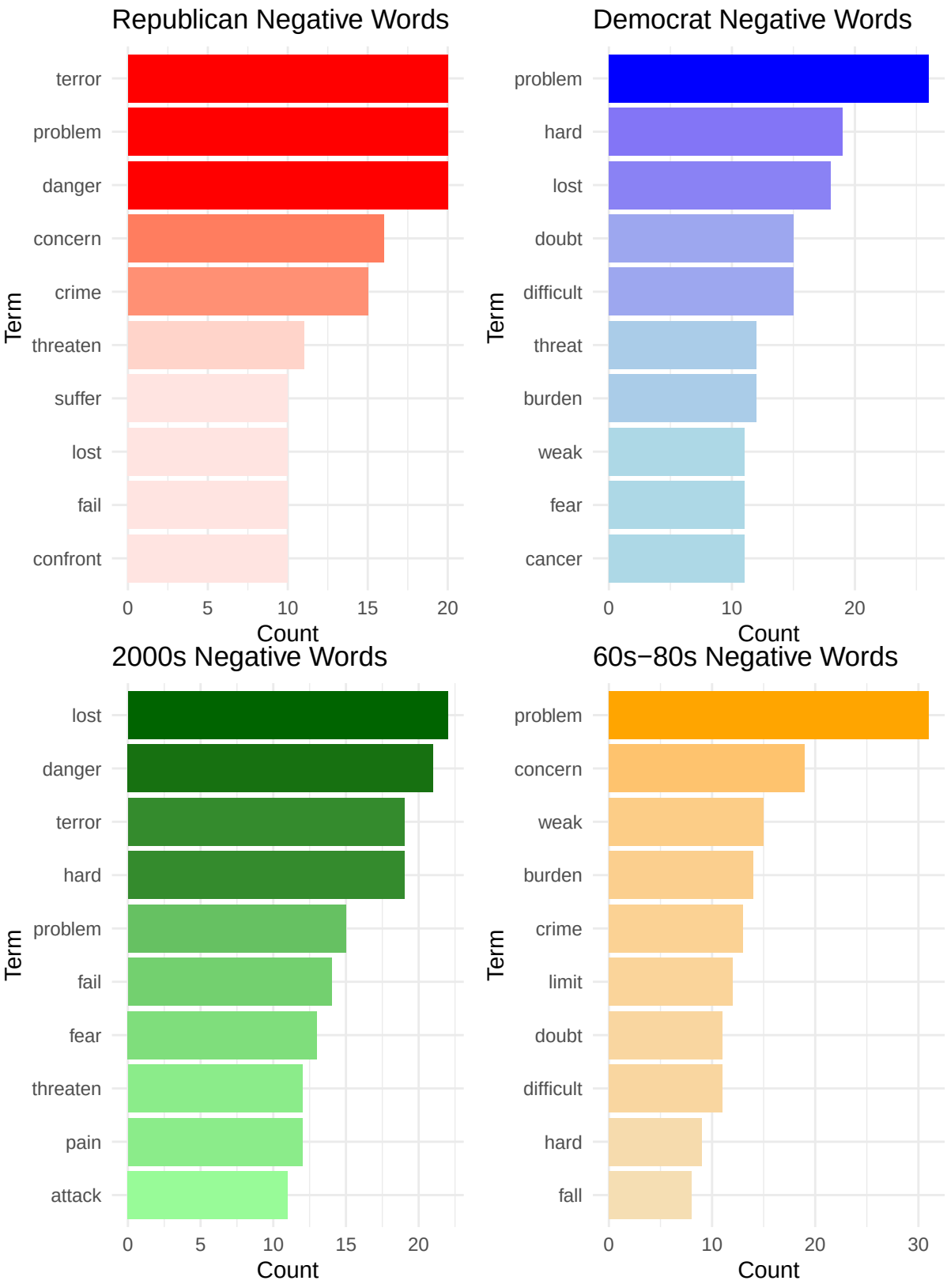
NRC Sentiment Breakdown

A more detailed sentiment classification using the NRC lexicon showed:

- **Trust** was **5.8% higher** in the 60s-80s era.
- **Fear** was **3.4% higher** in the 2000s era.
- **Anger** was **2.5% higher** in the 2000s era.

These findings align with major historical events influencing presidential rhetoric, such as Cold War diplomacy in the earlier era and post-9/11 security concerns in the 2000s.


Republican Positive Words
great
work
freedom
thank
good
well
right
protect
free
reform
Term
0
20
40
60
Count
Democrat Positive Words
work
right
good
well
support
free
like
freedom
great
thank
Term
0
25
50
75
Count
2000s Positive Words
work
thank
great
right
like
good
well
protect
support
freedom
Term
0
25
50
75
100
Count
60s−80s Positive Words
freedom
work
free
great
good
right
well
best
progress
strong
Term
0
10
20
30
40
Count


Republican Negative Words
terror
problem
danger
concern
crime
threaten
suffer
lost
fail
confront
Term
0
5
10
15
20
Count
Democrat Negative Words
problem
hard
lost
doubt
difficult
threat
burden
weak
fear
cancer
Term
0
10
20
Count
2000s Negative Words
lost
danger
terror
hard
problem
fail
fear
threaten
pain
attack
Term
0
5
10
15
20
Count
60s−80s Negative Words
problem
concern
weak
burden
crime
limit
doubt
difficult
hard
fall
Term
0
10
20
30
Count

Most Frequent Words Analysis

A key aspect of analyzing presidential rhetoric is identifying the most frequently used words within their speeches. The most frequent word plots reveal which terms dominate the discourse of different eras and political parties. These findings align with historical contexts and key policy focuses of the time.

Most Frequent Words by Era

From the word frequency analysis:

- **60s-80s Era:** The dominant words included *government*, *people*, *nation*, *freedom*, and *world*. These terms emphasize the era's focus on governance, national identity, and global presence, aligning with Cold War diplomacy and civil rights movements.
- **2000s Era:** The most frequent words shifted towards *security*, *jobs*, *economy*, *policy*, and *terrorism*. This indicates a stronger focus on economic growth and national security, reflecting post-9/11 concerns and financial crises.

The difference between eras is noticeable in the visualization, where **freedom** and **rights** were more prominent in earlier speeches, while **security** and **economy** became more prevalent in the 21st century.

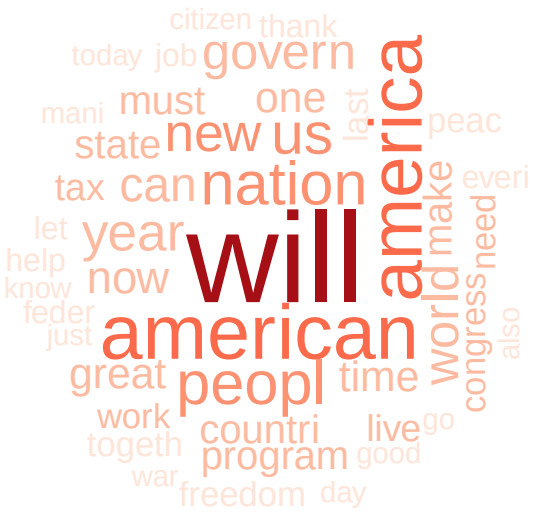
Most Frequent Words by Party

When analyzing word frequency by party:

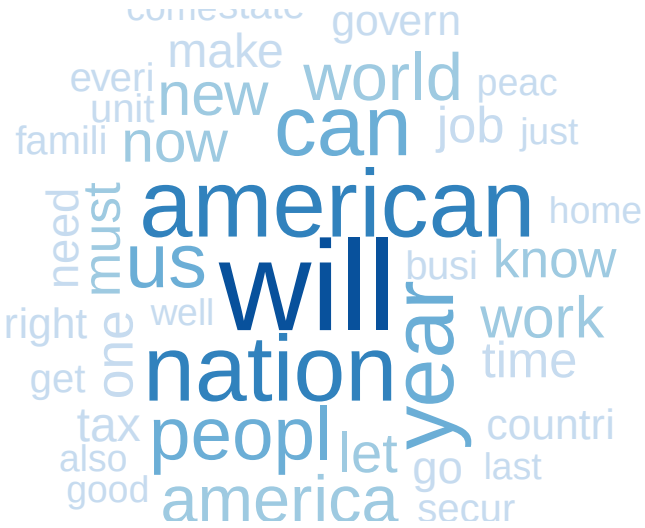
- **Democratic Presidents** frequently used words like *hope*, *future*, *change*, and *public*, indicating an emphasis on progressive policies and forward-looking rhetoric.
- **Republican Presidents** often highlighted terms such as *security*, *taxes*, *America*, and *strength*, reflecting a focus on economic conservatism and national security.

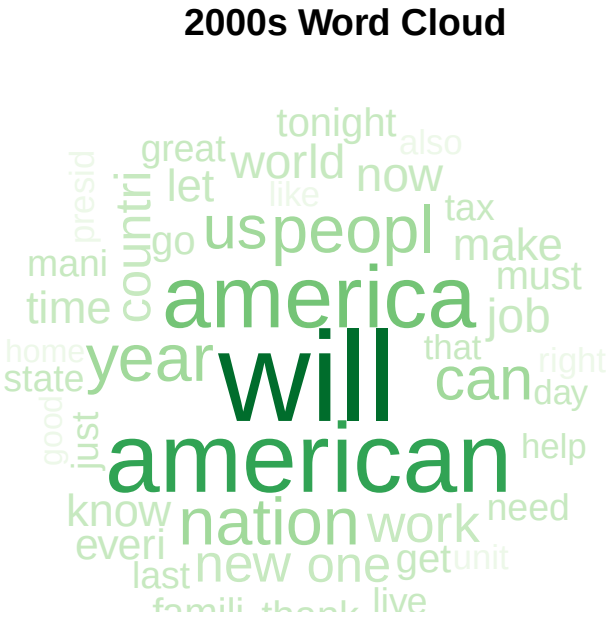
These word usage patterns align with traditional party ideologies and policy priorities.

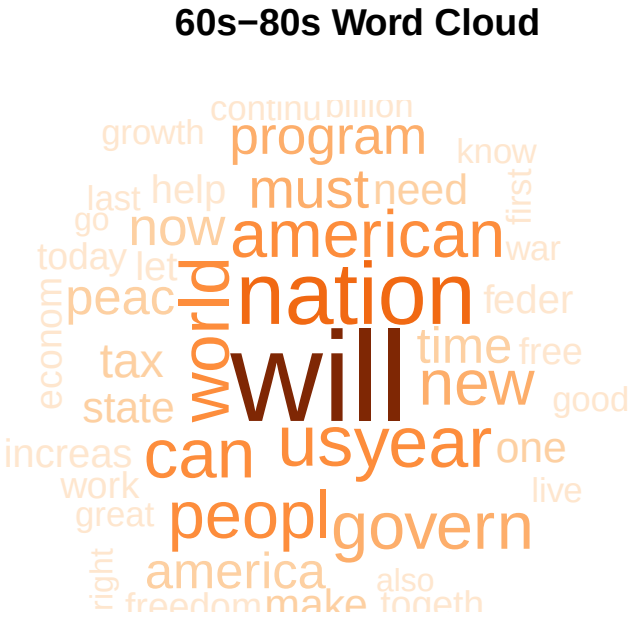
Republican Word Cloud



Democrat Word Cloud







Word Clouds

The **word clouds** provide a visual representation of speech content by era and party, emphasizing the most commonly occurring words. These images help illustrate key rhetorical focuses:

- **Era Word Clouds:** The word clouds show more conceptual and idealistic language in the 60s-80s, while the 2000s feature more policy-driven and pragmatic terminology.
- **Party Word Clouds:** Democratic speeches highlight terms related to change and collective progress, whereas Republican speeches emphasize national security and governance.

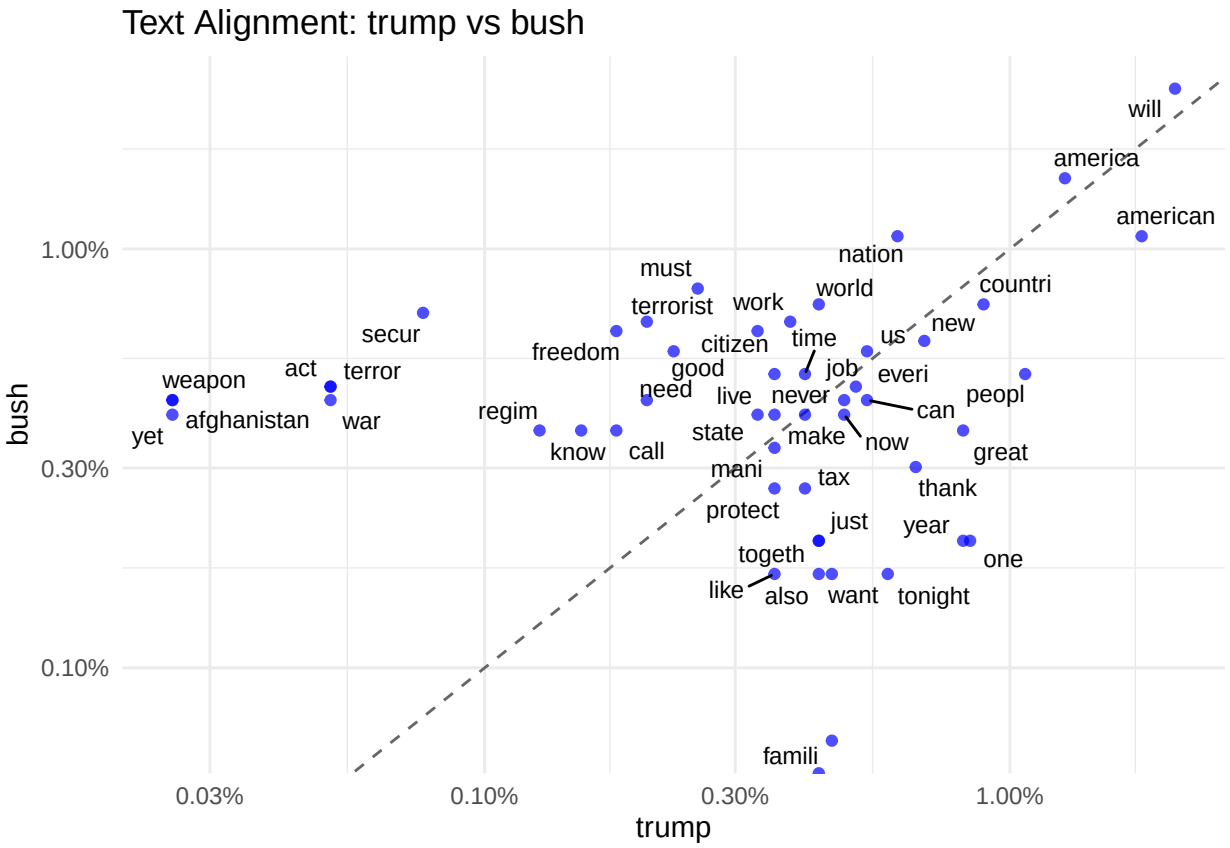
The era-based word clouds align with the sentiment analysis findings, as the earlier era contained more language tied to trust and optimism, while the modern era showed a slight increase in fear-related terms. Similarly, the party-based clouds reinforce the LDA topic modeling, where Republicans leaned towards national security and economic policy, while Democrats leaned towards governance and social themes.

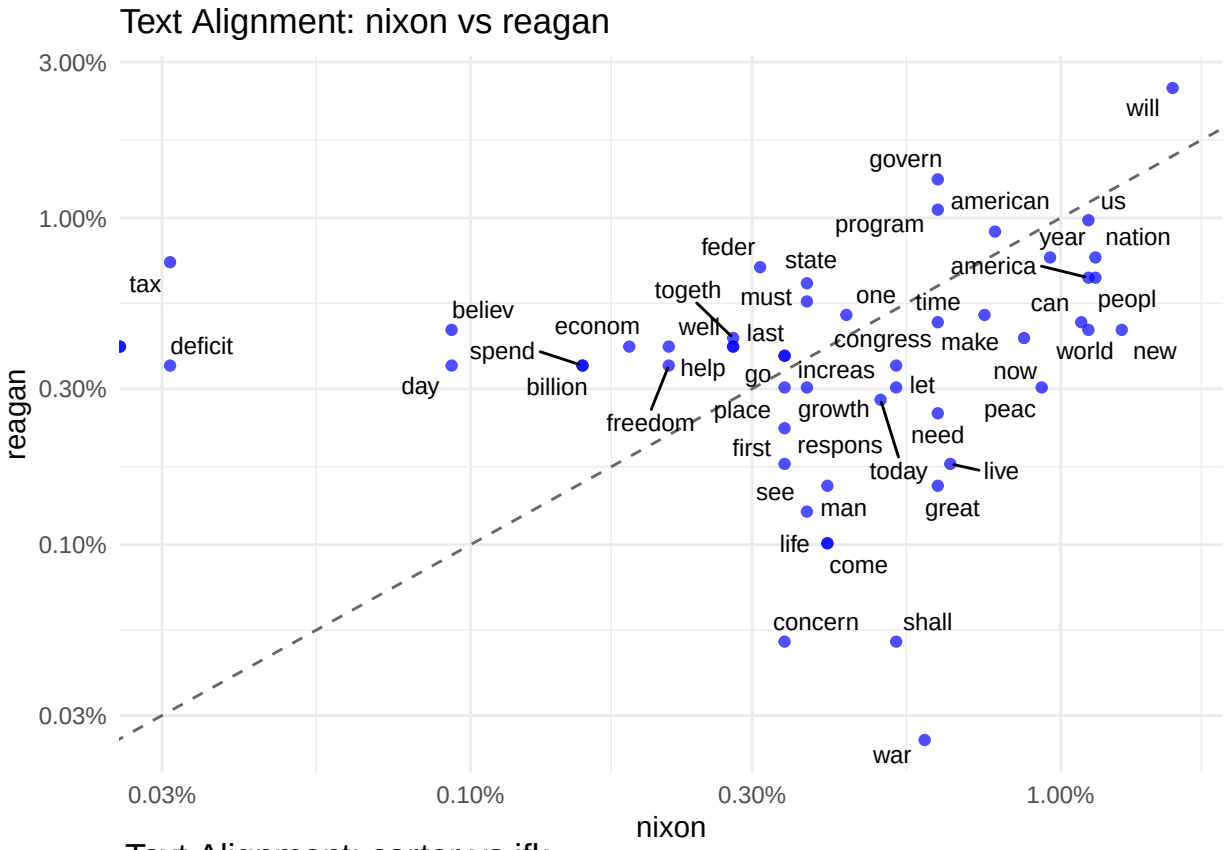
Topic Modeling

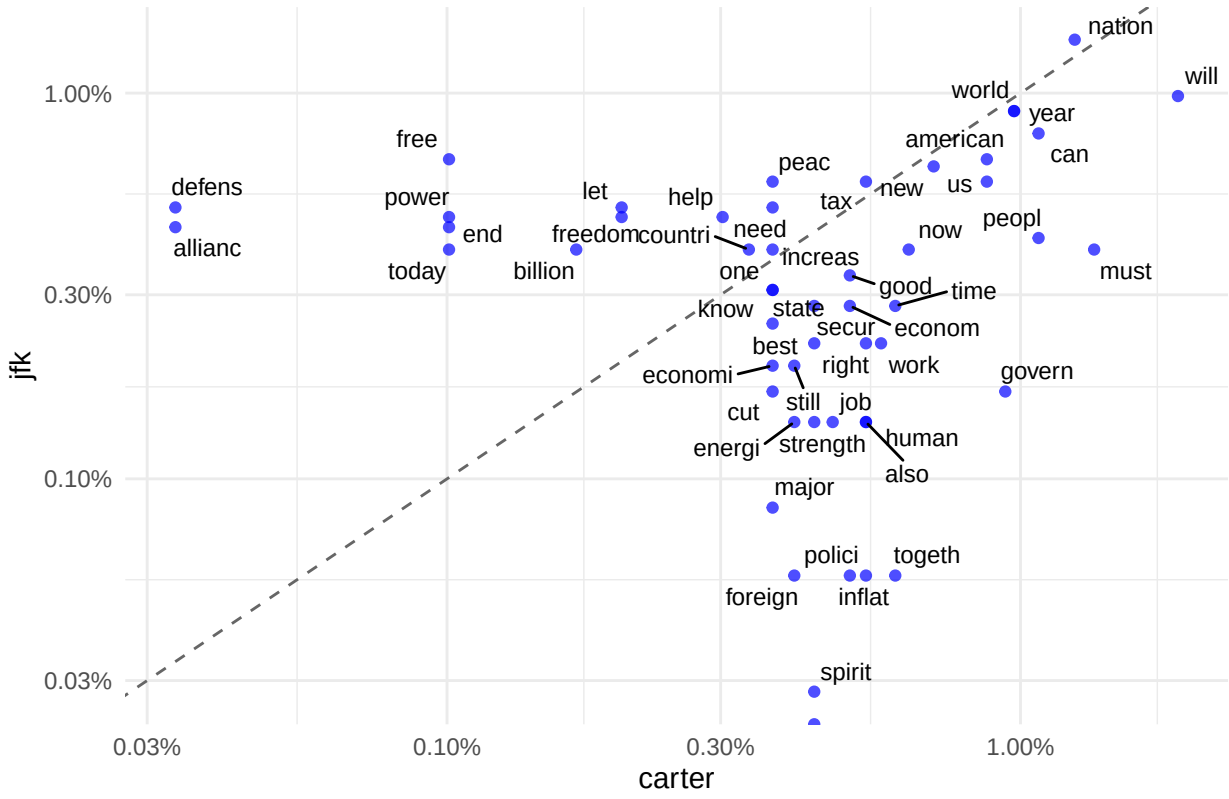
```
## # A tibble: 75 x 3
##    topic term      beta
##    <int> <chr>    <dbl>
##  1     1 world  0.0104
##  2     1 must   0.00904
##  3     1 new    0.00737
##  4     1 tax    0.00632
##  5     1 govern 0.00571
##  6     1 peac   0.00557
##  7     1 need   0.00496
##  8     1 free   0.00472
##  9     1 good   0.00450
## 10     1 help   0.00448
## # i 65 more rows

## # A tibble: 20 x 3
```

```
##    document topic      gamma
##    <chr>    <int>      <dbl>
##  1 jfk          1 1.00
##  2 carter       1 0.911
##  3 bush         1 0.00000572
##  4 nixon        1 0.00000542
##  5 nixon        2 1.00
##  6 reagan       2 0.972
##  7 carter       2 0.00000580
##  8 bush         2 0.00000572
##  9 biden        3 1.00
## 10 reagan       3 0.0283
## 11 carter       3 0.00000580
## 12 bush         3 0.00000572
## 13 trump        4 1.00
## 14 bush         4 0.994
## 15 carter       4 0.0894
## 16 nixon        4 0.00000542
## 17 obama        5 1.00
## 18 bush         5 0.00554
## 19 carter       5 0.00000580
## 20 nixon        5 0.00000542
```







To uncover underlying themes in the speeches, **Latent Dirichlet Allocation (LDA)** was applied. This

analysis identified five major topics, each with strong associations to particular presidents:

Topic	Top Presidents	Gamma Score
Global Policy & Governance	JFK, Carter	0.999, 0.911
Domestic Policy & Programs	Nixon, Reagan	0.999, 0.971
Leadership & Motivational Rhetoric	Biden, Reagan	0.999, 0.282
National Security & Patriotism	Trump, Bush	0.999, 0.994
Economic Growth & Workforce	Obama, Bush	0.999, 0.056

This classification indicates clear thematic differences in presidential focus areas.

Text Alignment and President Correlations

A key component of this analysis is determining which presidents exhibit the most similar rhetoric. By analyzing textual similarities, we can assess whether **era or party plays a greater role in shaping presidential speech patterns**. This is accomplished through **text alignment analysis** and **presidential correlation scores** derived from **LDA topic distributions**.

Top 3 Presidential Correlations

The LDA topic modeling results indicate the strongest correlations between presidents based on their thematic usage of language:

- **Most strongly correlated presidents:**
  - **Trump and Bush** (0.999) – Both presidents showed heavy alignment with **National Security and Patriotism**, consistent with their policies emphasizing military strength and counterterrorism.
  - **Nixon and Reagan** (0.999) – Their speeches reflected strong thematic similarities in **Domestic Policy and Government Programs**, focusing on economic policies, governance, and conservative fiscal principles.
  - **Carter and JFK** (0.995) – These two Democratic presidents shared strong rhetorical alignment in **Global Policy and Governance**, highlighting diplomacy, international relations, and calls for unity.

These correlations suggest that **era plays a significant role in shaping presidential discourse**, as all three highly correlated pairs come from the **same era** rather than the same party.

Text Alignment Plots

The **text alignment plots** provide further visual confirmation of these correlations by comparing speech structures, word usage, and thematic alignment across presidents. Key takeaways from these visualizations:

- **Presidents from the same era align more closely than those from the same party.** For instance, JFK and Carter (both from the 60s-80s era) exhibit stronger alignment than Carter and Biden, despite both being Democrats.
- **Republican presidents from the modern era (Bush and Trump) have nearly identical thematic distributions.** Their rhetoric is **highly centered around security, patriotism, and economic strength**.
- **Reagan and Carter, despite serving consecutively, have one of the lowest correlations (-0.29).** This suggests a **sharp rhetorical and policy contrast** between them, supporting historical observations that Reagan marked a significant ideological shift from Carter's presidency.

These **alignment trends support the overall finding that era is a more powerful differentiator of presidential speech than party affiliation**. While party plays a role in shaping rhetoric, the **historical and societal contexts of an era drive stronger linguistic and thematic similarities** among presidents.

Predictive Modeling

Machine learning models were trained to classify speeches based on **era** and **party affiliation**. The two primary models used were:

- **Support Vector Machines (SVM)**
- **Random Forest Classifier**

Model Performance

Model	Task	Accuracy
Linear SVM	Era Prediction	74.9%
Linear SVM	Party Prediction	62.6%
Random Forest	Era Prediction	66.6%
Random Forest	Party Prediction	55.5%

Results indicate that **era-based classification is more accurate than party-based classification**, reinforcing the idea that time period plays a more significant role in shaping presidential rhetoric than party affiliation.

```
## Confusion Matrix for Party Prediction:

##             Actual
## Predicted    Democrat Republican
##   Democrat        192         98
##   Republican       97        134

## Confusion Matrix for Era Prediction:

##          Actual
## Predicted 2000s 60s-80s
##   2000s     228      70
##   60s-80s    61     162

## Party Accuracy: 0.626

## Era Accuracy: 0.749

## Ranger Party Confusion Matrix:

##             Actual
## Predicted    Democrat Republican
##   Democrat        194        137
##   Republican       95         95

## Ranger Party Accuracy: 0.555

## Ranger Era Confusion Matrix:

##          Actual
## Predicted 2000s 60s-80s
##   2000s     224     109
##   60s-80s    65     123

## Ranger Era Accuracy: 0.666
```

Conclusions

- **Era influences speech content more than party affiliation**, as seen in both sentiment trends and predictive modeling.
- **Positive sentiment dominated all speeches**, though **fear and anger increased in recent decades**.
- **Distinct topic groupings** emerged, with some presidents strongly associated with specific themes.
- **Machine learning models performed significantly better at distinguishing speeches by era** rather than by party.

Future Work

- Expanding the dataset to include speeches beyond the first term and first year.
- Applying deep learning techniques such as transformers for enhanced text classification.
- Examining media and geopolitical events to provide external context for linguistic shifts.
