

Dynamics of Content, Sentiment, and Theme in the Presidency

A Text Analysis and Modeling Project

Aidan Ohlson

01. Project Data

Background

- The texts being analyzed are from the Inaugural Addresses and State of the Union Speeches of the former US Presidents listed below
 - Each speech was from the first year and first term of the presidency
- There are four presidents from the 60s-80s era and four from the 2000s era
 - Split evenly by party in each respective era

<u>60s-80s</u>	JFK (D) 1961	Nixon (R) 1969	Carter (D) 1977	Reagan (R) 1981
<u>2000s</u>	Bush (R) 2001	Obama (D) 2009	Trump (R) 2017	Biden (D) 2021

General Research Goal: Comparing the content, sentiments, and trends in US presidential speeches between era and party

Cleaning

1. Read speeches in from PDF format
2. Converted to lowercase
3. Removed punctuation
4. Removed numbers
5. Removed English stopwords
6. Stripped whitespace
7. Stemmed documents

*Sparsity was low as is without `removeSparseTerms`

Inspection

For the combined speeches:

Sparsity: 81%

Maximal Term Length: 24

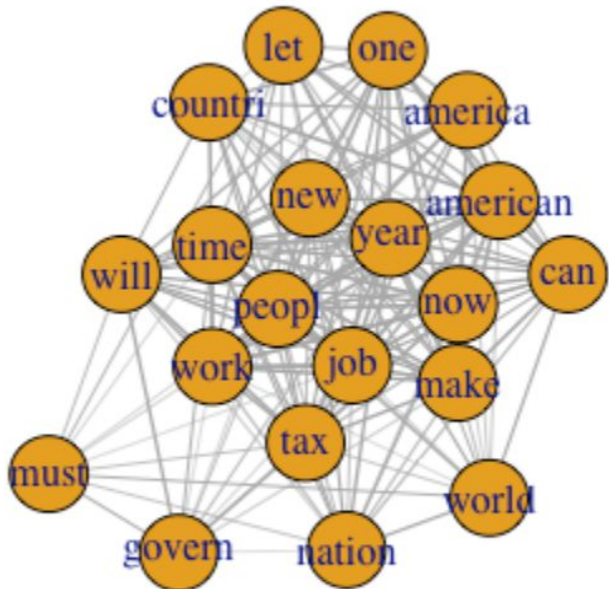
Average Words per Document: 1908.75

Average Unique Words per Document:
826.69

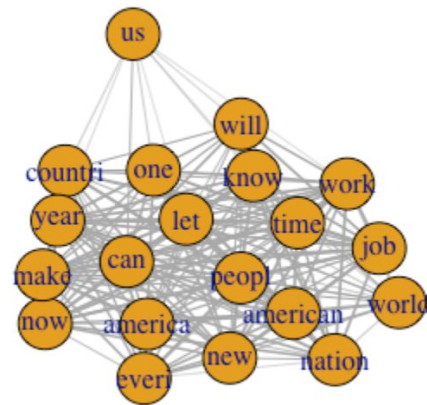
Average Vocab Density: 0.485

Term Association

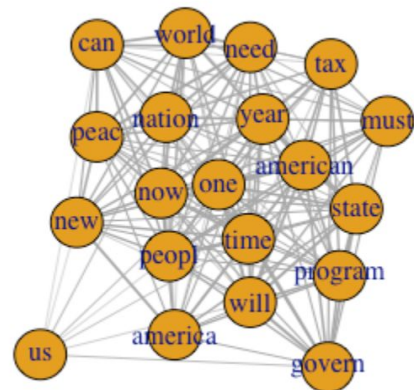
Top Term Associations (All)



Top Term Associations 2000s Era



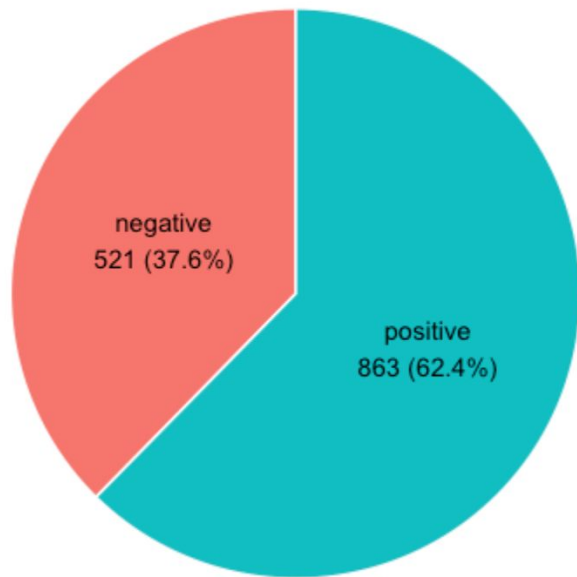
Top Term Associations 60s-80s Era



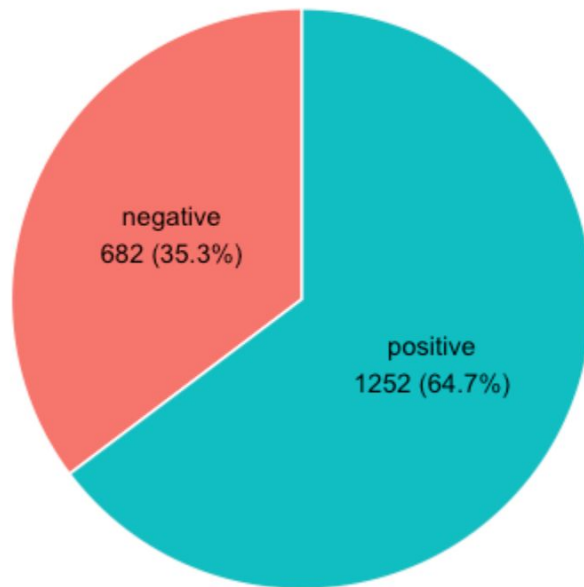
02. Sentiment Analysis

Positive/Negative Terms by Era

60s-80s Speeches



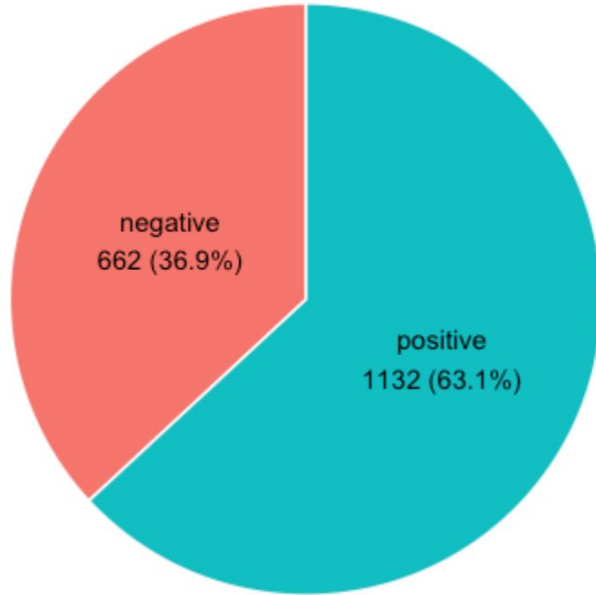
2000s Speeches



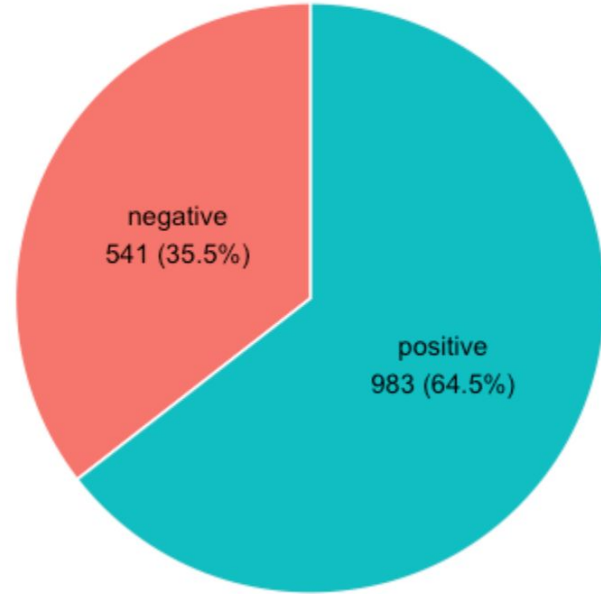
62.4% vs. 64.7% positive sentiment

Positive/Negative Terms by Party

Democratic Speeches



Republican Speeches



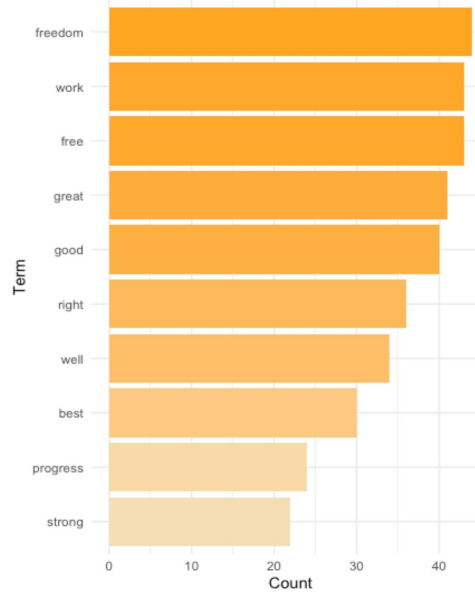
63.1% vs. 64.5% positive sentiment

Most Frequent Positive/Negative Terms by Era

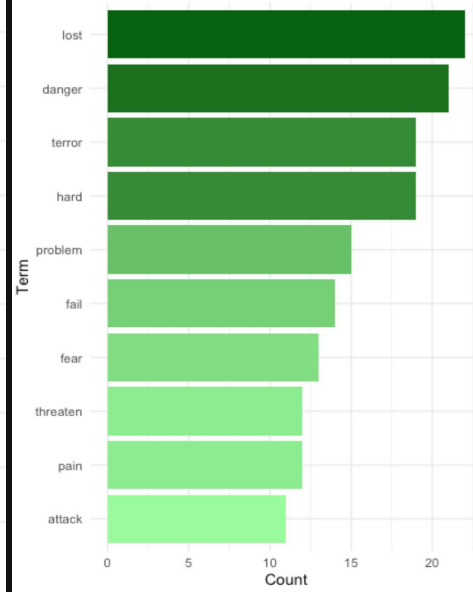
2000s Positive Words



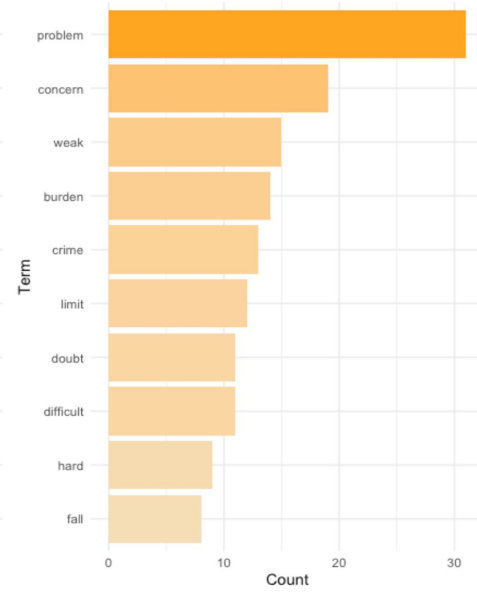
60s-80s Positive Words



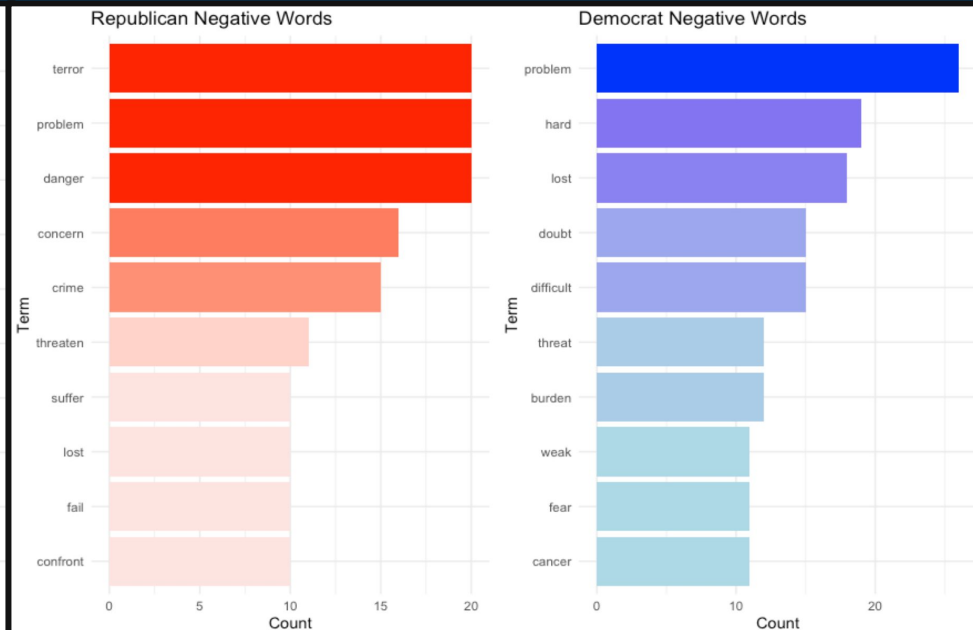
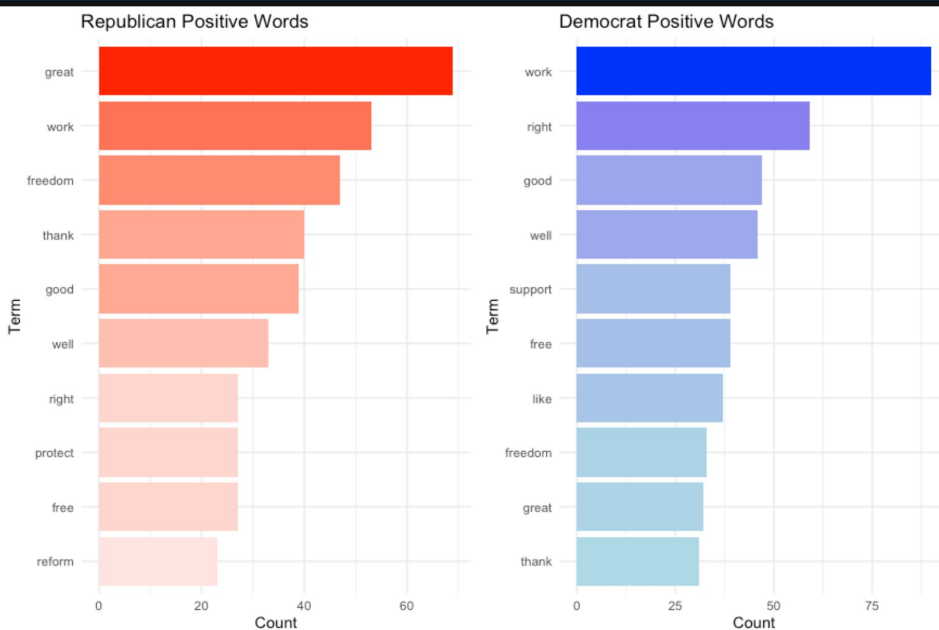
2000s Negative Words



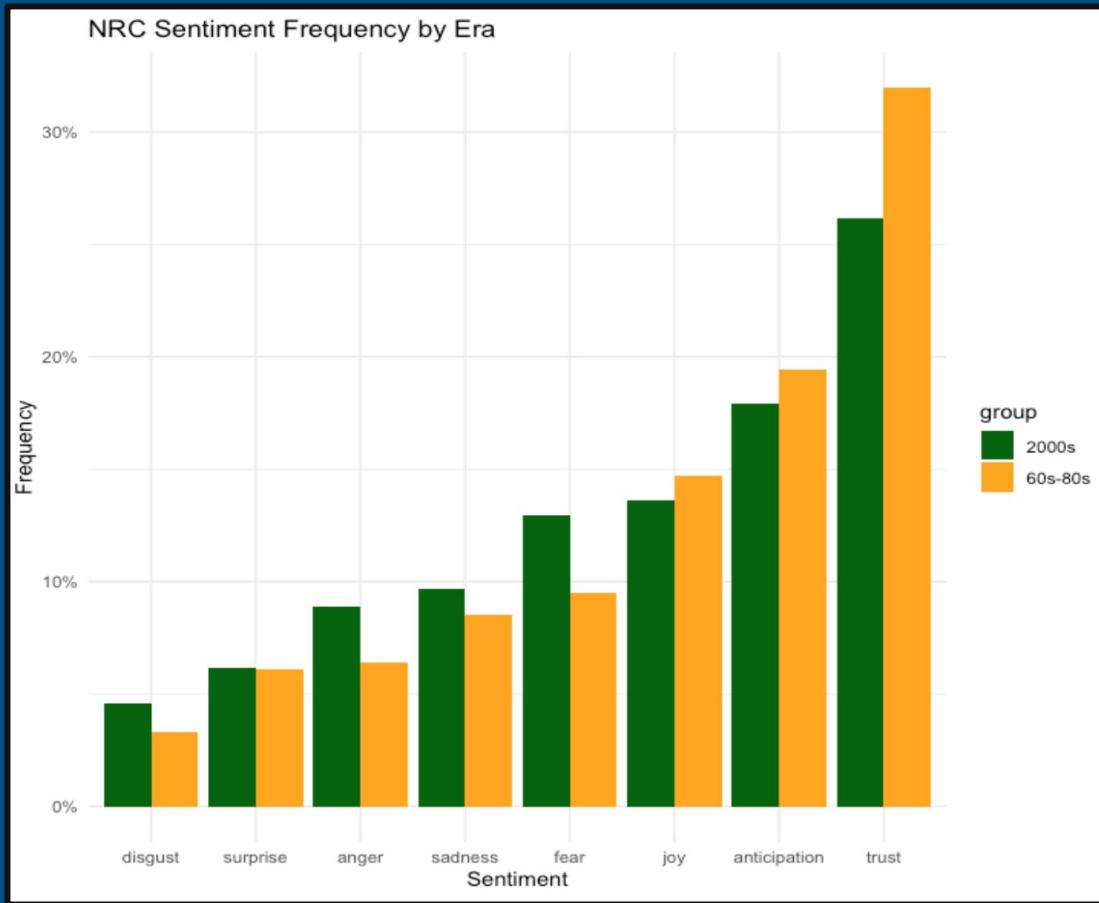
60s-80s Negative Words



Most Frequent Positive/Negative Terms by Party



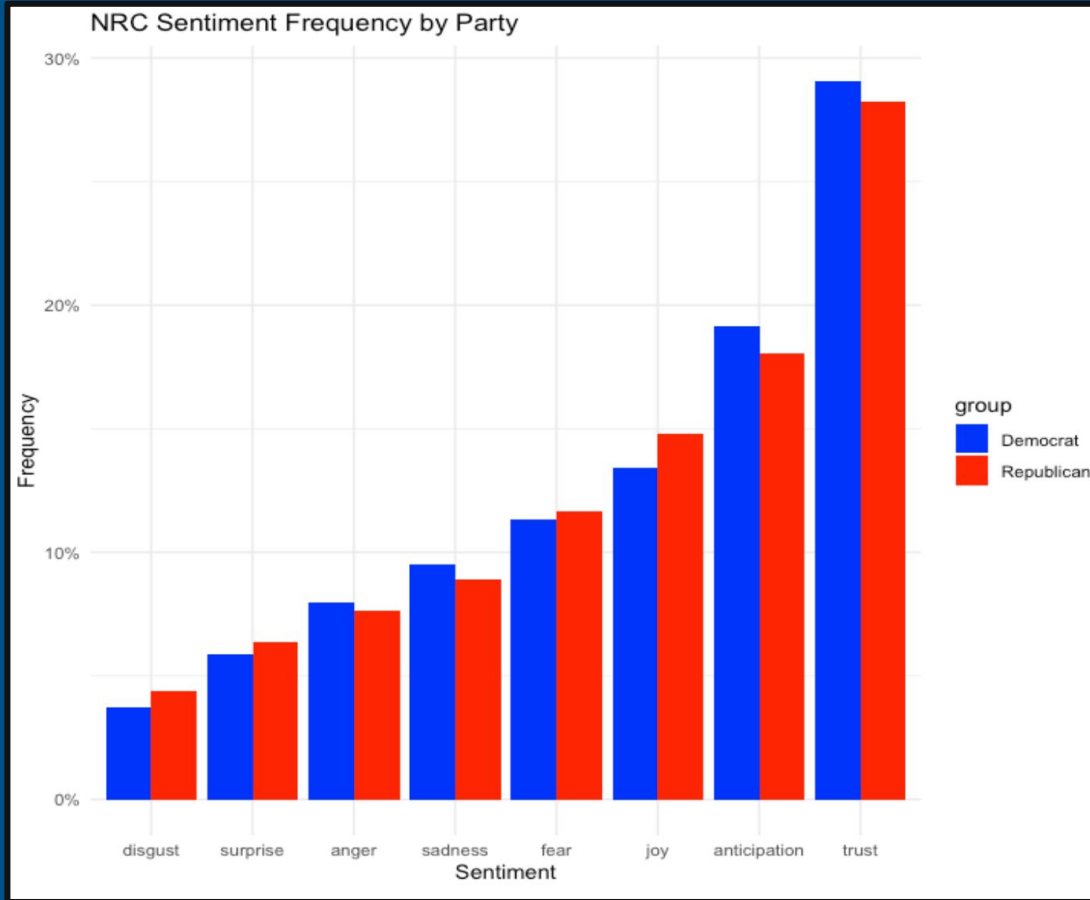
NRC Sentiments by Era



Top 3 Differences (by frequency %):

- Trust: 5.8% greater for 60s-80s
- Fear: 3.4% greater for 2000s
- Anger: 2.5% greater for 2000s

NRC Sentiments by Party

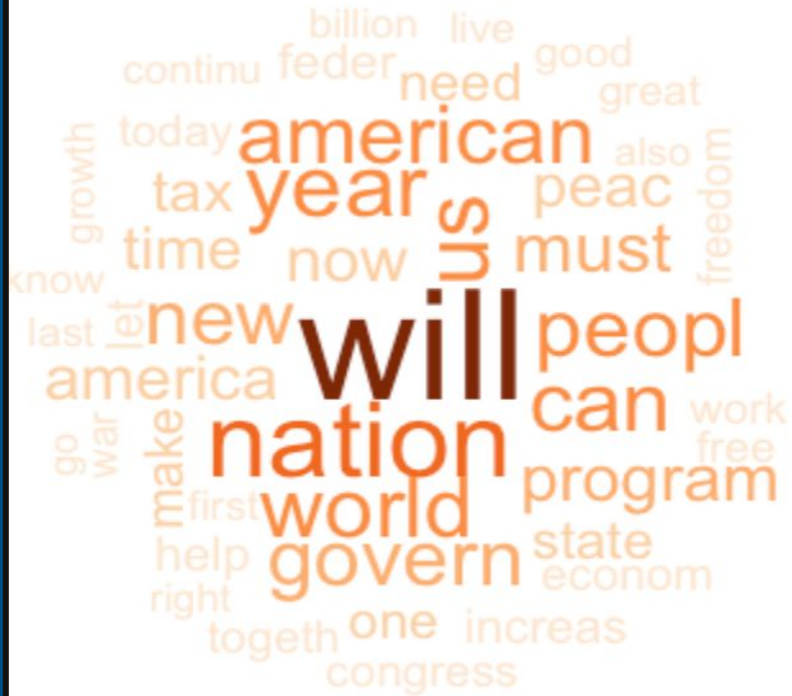


Top 3 Differences (by frequency %):

- Joy: 1.4% greater for Republican
- Anticipation: 1.1% greater for Democrat
- Trust: 0.7% greater for Democrat

Era Word Clouds

60s-80s Word Cloud



2000s Word Cloud



Party Word Clouds

Republican Word Cloud

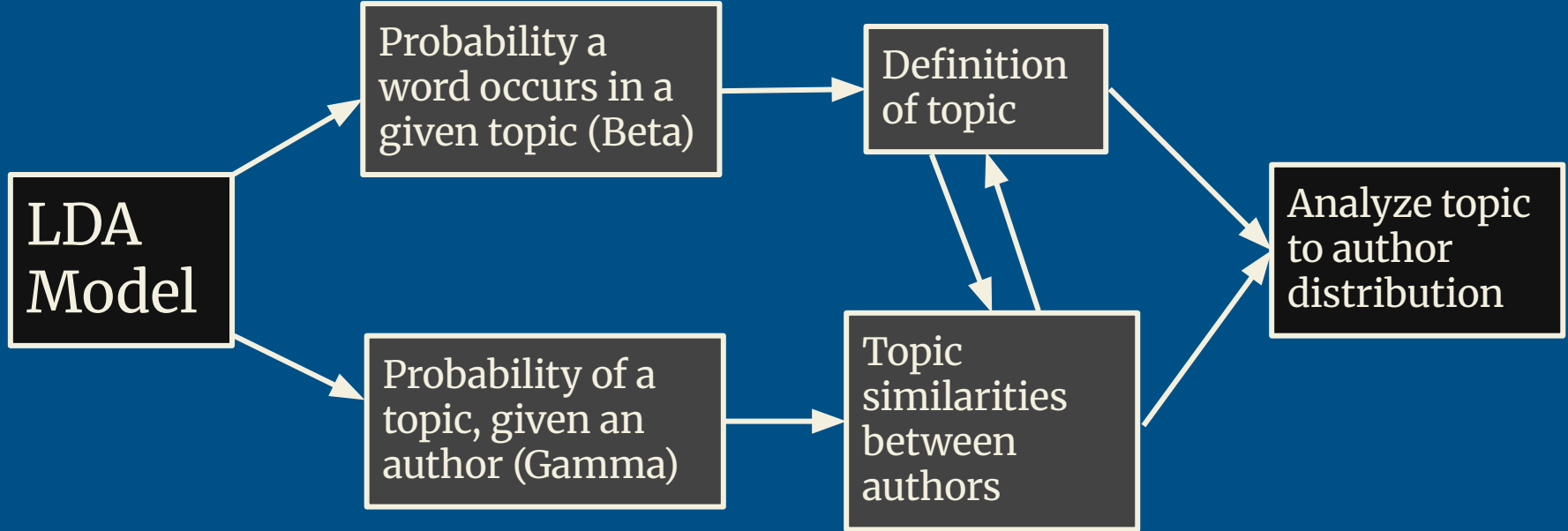


Democrat Word Cloud



03. Comparing Authors and Topics

Latent Dirichlet Allocation (LDA)



LDA Results

Topic	Top Presidents	Gamma
Global Policy and Governance	JFK Carter	0.999 0.911
Domestic Policy and Government Programs	Nixon Reagan	0.999 0.971
Leadership/Motivational Rhetoric	Biden Reagan	0.999 0.0282
National Security and Patriotism	Trump Bush	0.999 0.994
Economic Growth and Workforce Policy	Obama Bush	0.999 0.0056

Top Author Correlations:

Positive:

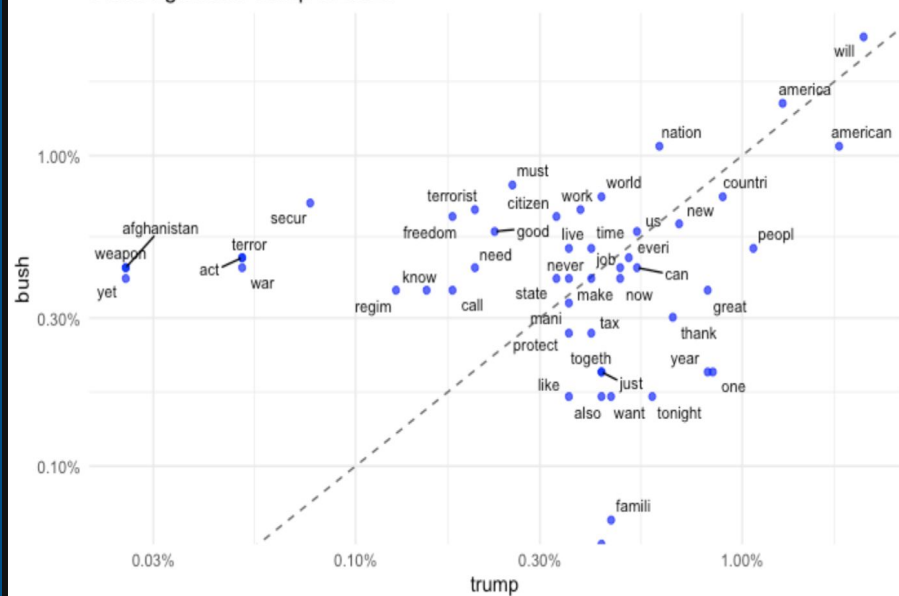
- Trump and Bush (0.999)
- Nixon and Reagan (0.999)
- Carter and JFK (0.995)

Negative:

- Reagan and Carter (-0.29)
- Biden and Carter (-0.28)
- Nixon and Carter (-0.28)

Text Alignment (Highest Correlators)

Text Alignment: trump vs bush

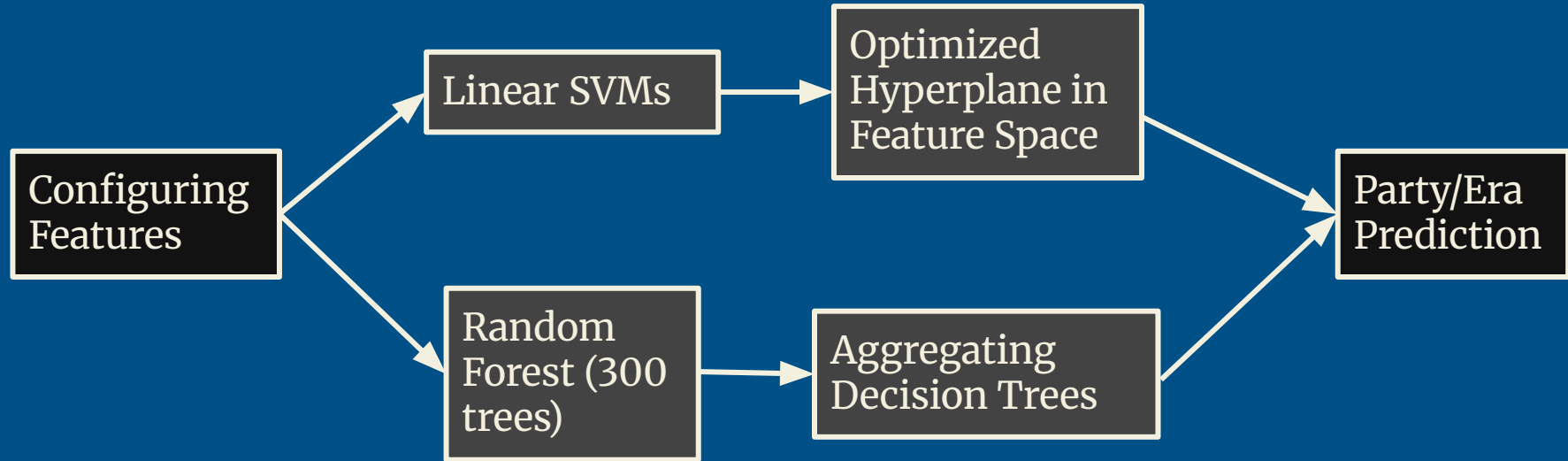


Text Alignment: nixon vs reagan



04. Predictive Modeling

Linear SVM and Random Forest



Goal: Evaluate differences in both model's prediction of Era versus Party

Model Results

Linear SVM

P\A	2000s	60-80
2000s	228	70
60-80	61	162

P\A	Dem	Rep
Dem	192	98
Rep	97	134

Testing Accuracies:
74.9% Era versus 62.6% Party

Random Forest

P\A	2000s	60-80
2000s	224	109
60-80	65	123

P\A	Dem	Rep
Dem	194	137
Rep	95	95

Testing Accuracies:
66.6% Era versus 55.5% Party

>10% greater accuracy in both models
when predicting Era

Conclusions

- All presidents shared certain overarching themes/features
- More variation in specific sentiments/terms than binary (pos/neg) sentiment
- Presidents aligned with similar Era and Party when identifying key topics
- Era dynamics were overall more differentiable than Party, both models more accurately predicting Era
- Ideal relation in Presidents likely combines important features from both Era and Party