State of the Union Address Sentiment & Topic Analysis

Text Mining for Public Policy, Spring 2018

Background & Literature Review

The Constitutionally Mandated Address

- President of the United States (POTUS) presents a message to a joint-session of Congress *every year*
- Historically, addresses about policy, including budgetary and economic outlooks along with other legislative priorities
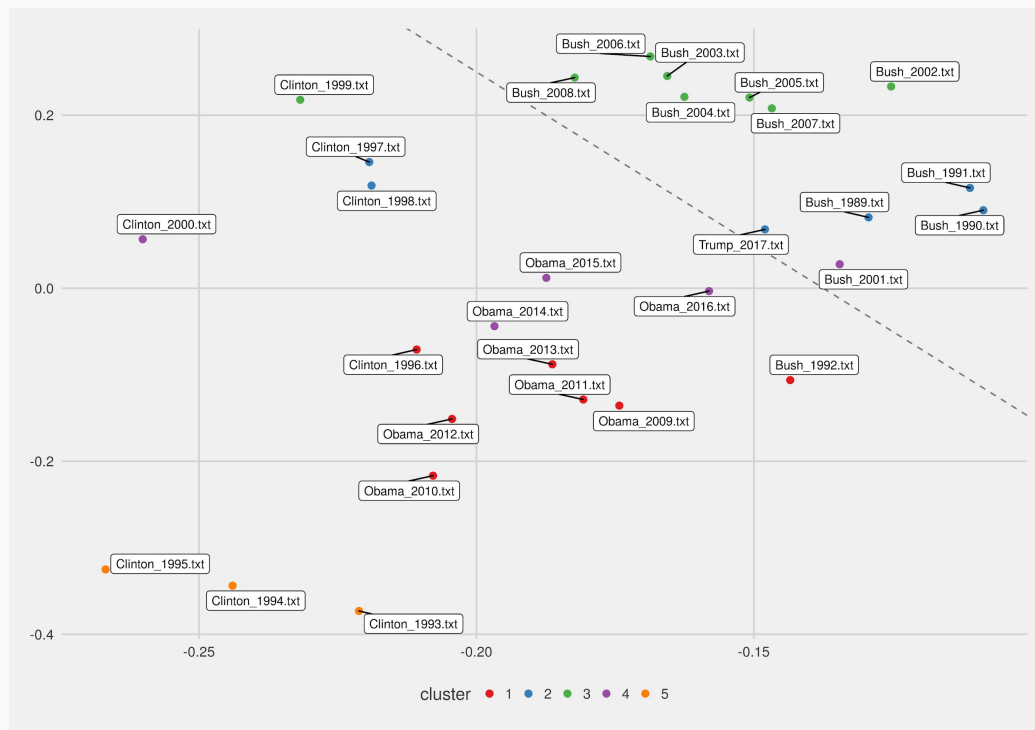- Have become more culturally motivated with rise of television

A happier time

Literature Review

Analyzing State of the Union is not new

Existing Body of Work

- Cluster analysis using SVD & Kmeans

Jonathan Bouchet's _NLP analysis on the SOTU addresses_ shows Clinton's addresses differ greatly during his second term, while W's all fall within one cluster
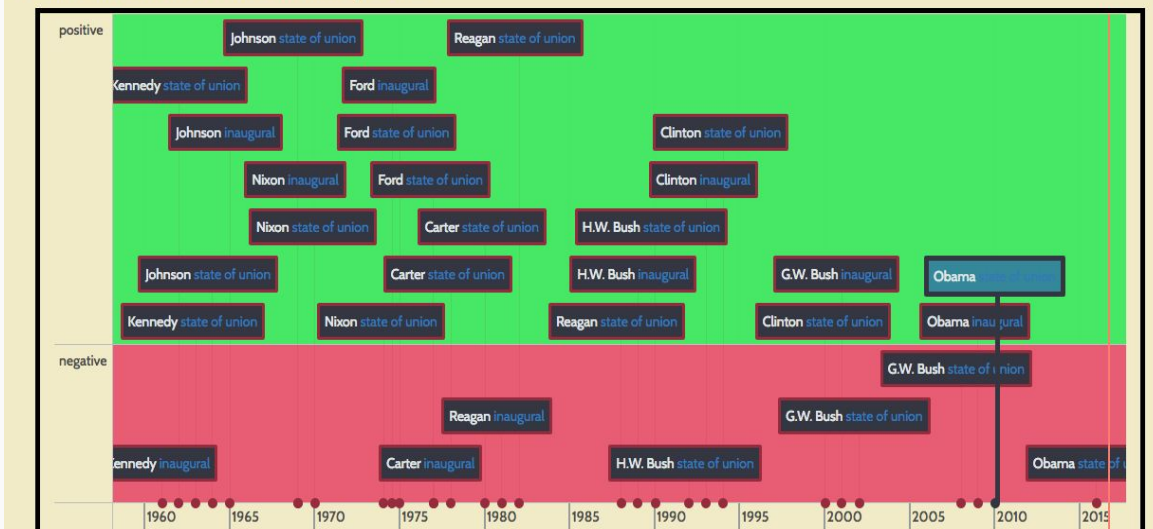
Literature Review

Analyzing State of the Union is not new

Existing Body of Work

- Cluster analysis using SVD & Kmeans

- Sentiment analysis using nltk
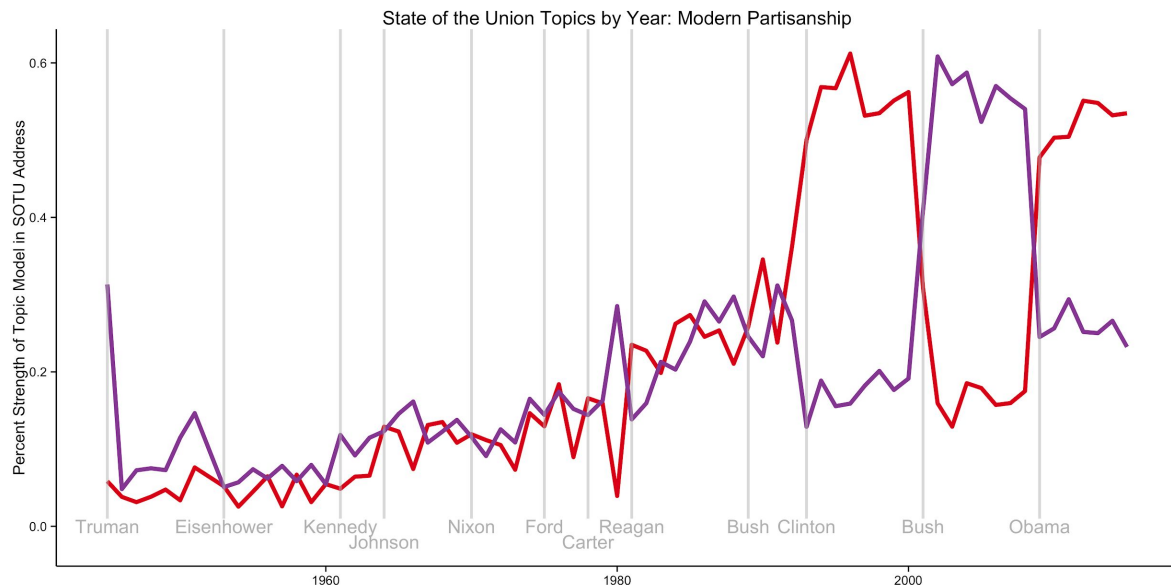
Overall Sentiment of Presidential Speeches

positive

Johnson state of union
Reagan state of union
Kennedy state of union
Ford inaugural
Johnson inaugural
Ford state of union
Clinton state of union
Nixon inaugural
Ford state of union
Clinton inaugural
Nixon state of union
Carter state of union
H.W. Bush state of union
Johnson state of union
Carter state of union
H.W. Bush inaugural
G.W. Bush inaugural
Obama
Kennedy state of union
Nixon state of union
Reagan state of union
Clinton state of union
Obama inaugural

negative

G.W. Bush state of union
Reagan inaugural
G.W. Bush state of union
Kennedy inaugural
Carter inaugural
H.W. Bush state of union
Obama state of

1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015

Jennifer Dixon's *Presidential Speech Analysis* trains a Naive Bayes classifier using political speech texts to classify 30 State of the Union and Inaugural Addresses as positive or negative

Literature Review

Analyzing State of the Union is not new

Existing Body of Work

- Cluster analysis using SVD & Kmeans

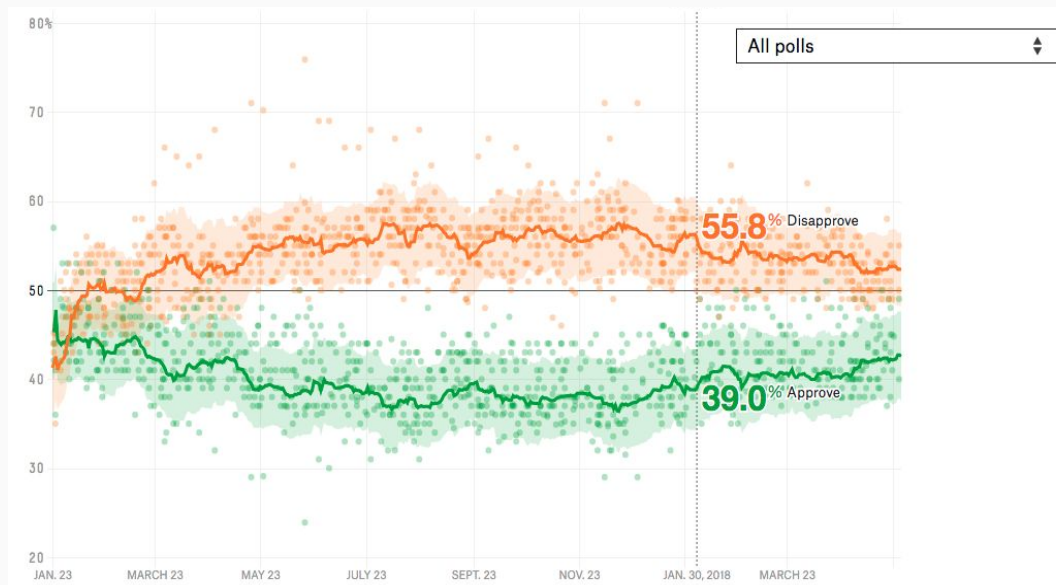- Sentiment analysis using nltk

- Topic modeling using LDA

State of the Union Topics by Year: Modern Partisanship

Frank Evan's *Topic Modeling of the State of the Union Address* characterizes speeches since 1913 into four main topics and examines how these topics have evolved over time, specifically with the rise of televisions and more recently the 24-hour news cycle

Research Question Development

Scrolling through [FiveThirtyEight's presidential approval poll aggregator](#) gave me the idea to tie together State of the Union addresses and approval ratings

Can we predict presidential approval using State of the Union sentiment?

Trump's Approval Rating Following 2018 SOTU

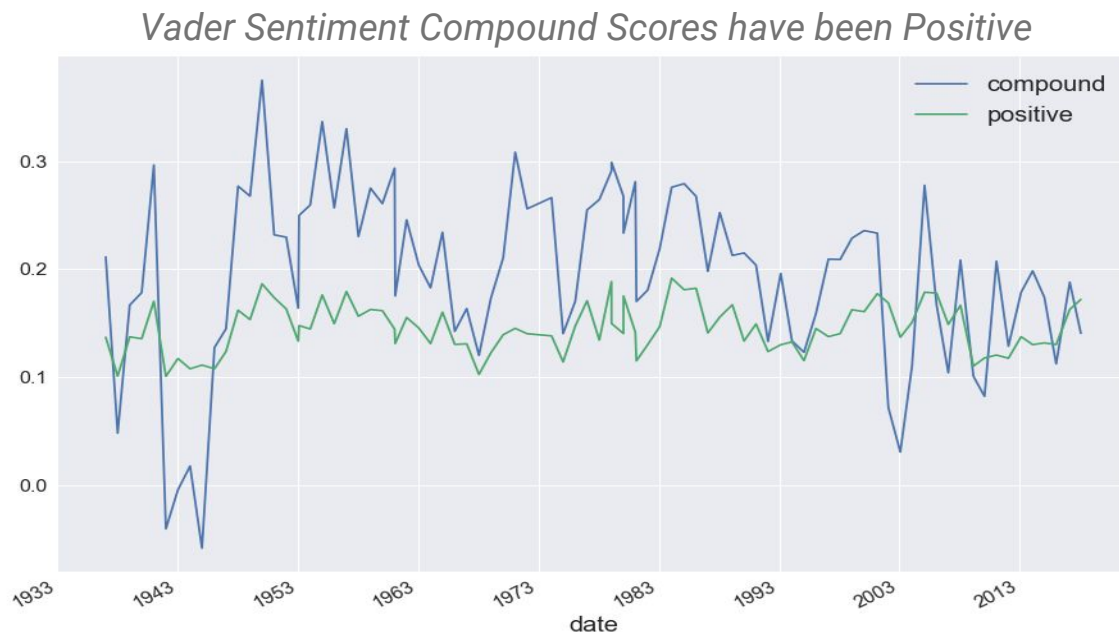# Dataset Description

UCSB's [The American Presidency Project](#)

- Corpus built by scraping State of the Union transcripts
  - Output stored as .txt file for each address, including both written and spoken addresses

Roper Center for Public Opinion Research [Presidential Approval Project](#)

- Database of public presidential polls starting from 1945
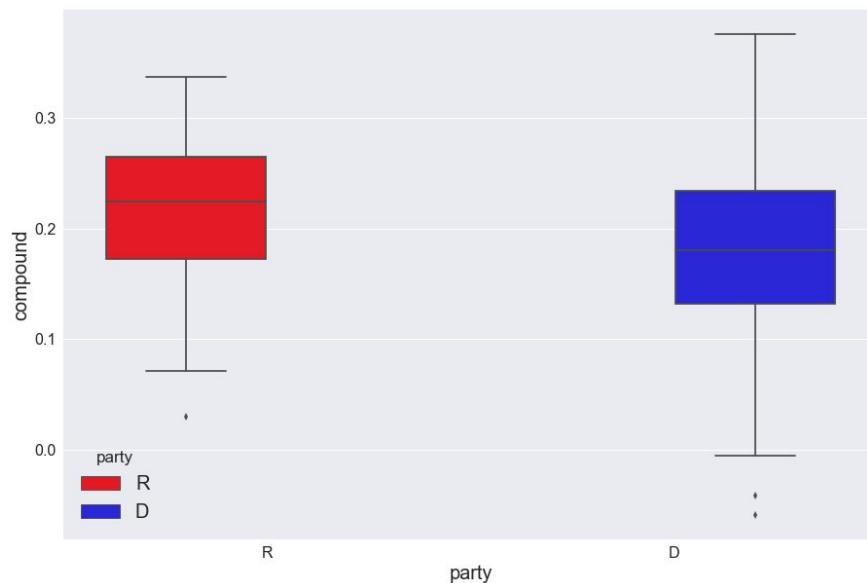
Sentiment Analysis Over Time

- Speeches have been fairly positive overall
  - Only negative scores from FDR in 1945 and 1942

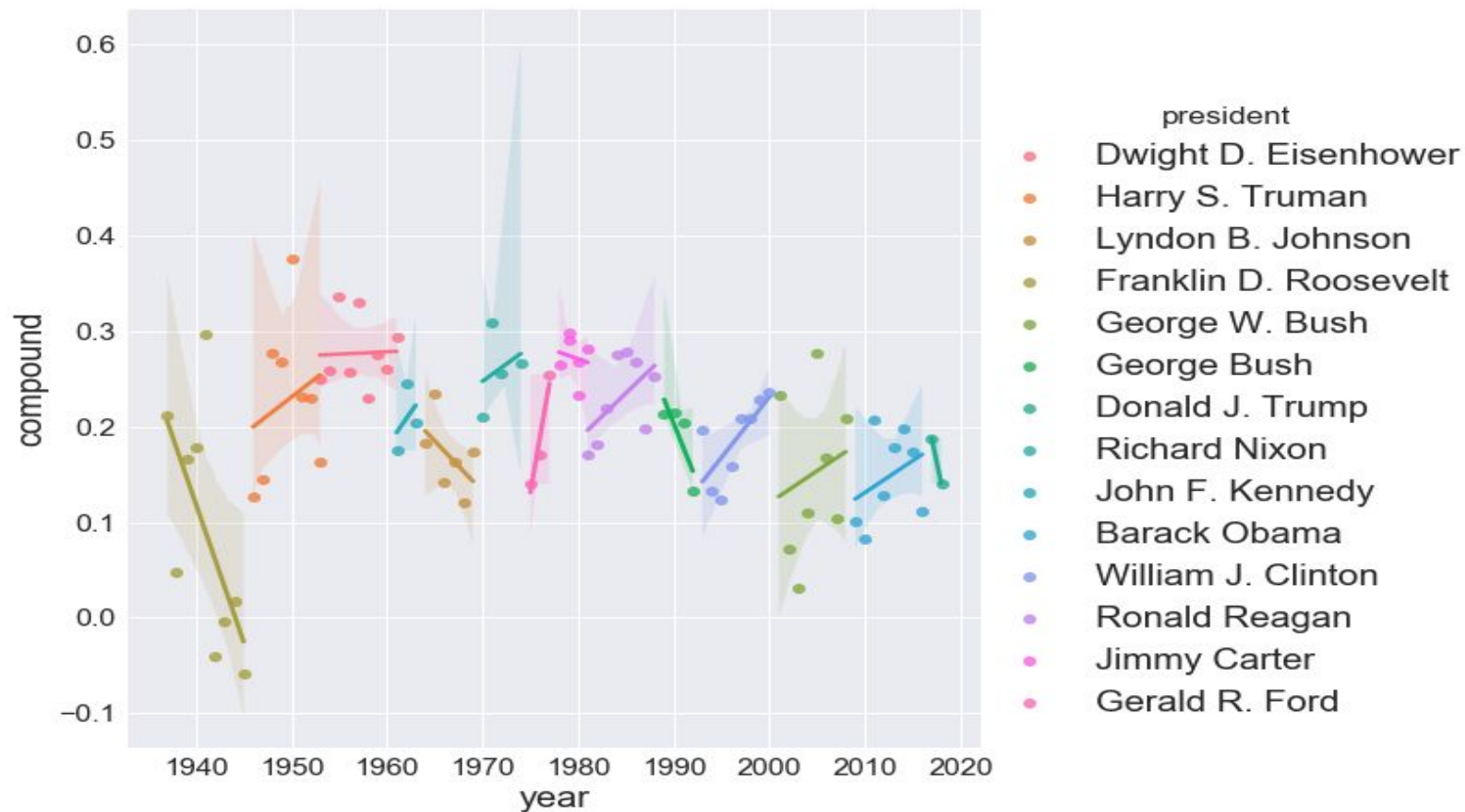*Vader Sentiment Compound Scores have been Positive*

Sentiment Analysis by Party

- Republicans have a higher median sentiment score and smaller variance since FDR
- FDR's negative scores are outliers

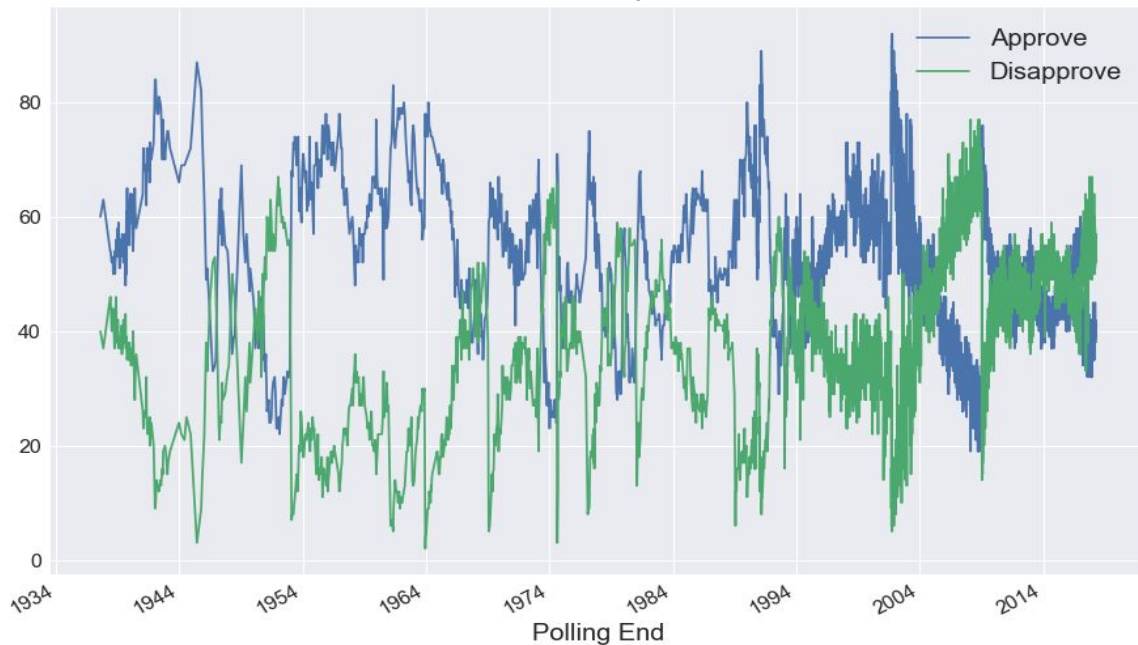*Democrats have had a much wider spread of sentiment*

Sentiment Analysis by President

Approval Ratings

- Approval ratings trends more difficult to decipher recently with emergence of a diverse polling industry
  - Higher approval percentages at beginning of first term, following national disasters

Number of Polls has Exploded since 1994

Predicting Approval

- Trained logistic regression models using various hyperparameters
- Accuracy around 62%
- Poor precision and recall

```
get_precision(cnf_matrix)
```
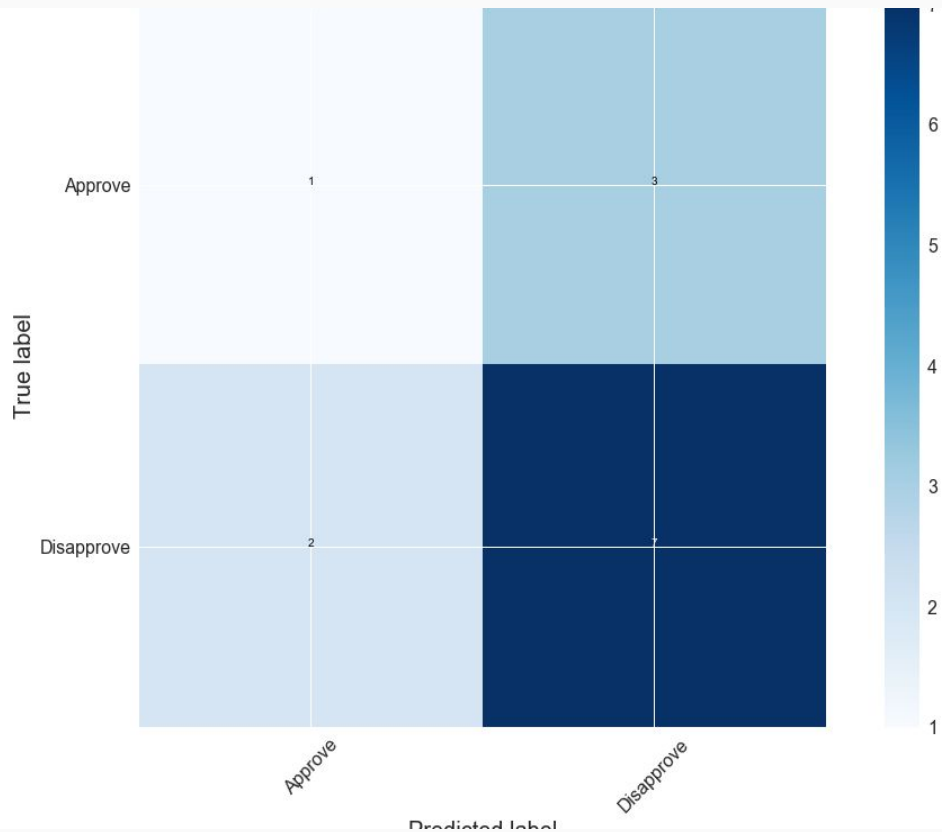```
0.3333333333333333
```

```
get_recall(cnf_matrix)
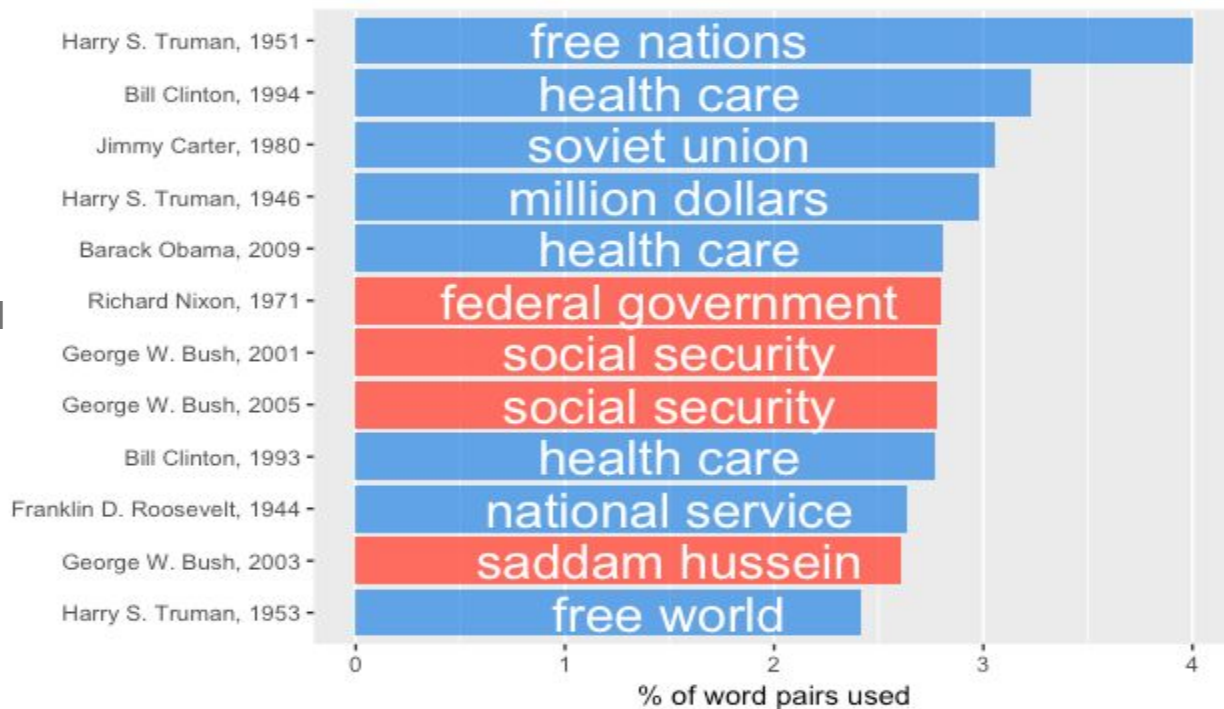```
```
0.25
```

```
get_accuracy(cnf_matrix)
```
```
0.6153846153846154
```


Confusion Matrix for Top AUC Model

Top Bigrams

- Bigrams typically reflect top policy goals and the biggest "villains" of the time
  - Clinton and Obama wanted to pass healthcare bills
  - Nixon, Carter, and W. really disliked the Federal Government, Soviet Union and Saddam Hussein respectively

Topic Modeling

Topic 1: Policy/Legislative Words

legislation public
policy labor
defense federal billion
military administration
programs strength
development

Topic 2: Middle Eastern Conflicts

afghanistan
fight weapons iraq
september enemy qaida
enemies terror
al
iraqi terrorist

Topic 3: How to end Communism?

conflict commitments south think
billion try
recommend vietnam
tonight north men
crime believe

Topic 4: Cold War & WWII

europe soviet
strength
military union defense
threat communist oil
korea atomic
countries aggression

Limitations & Potential Applications

- Sentiment analysis of State of the Union addresses is limited in policy scope

- Code pipeline is proof of concept

- Pipeline has potential applications in other areas:

    - Political campaigns: using stump speech sentiment to predict polling figures

    - Advertising/Marketing: using product sentiment to predict future interactions

Thank you!

Questions?

References

- Pablo Martinez Monsivais/Getty Images, *State of the Union Photo Gallery*
  - http://www.wbur.org/hereandnow/2016/01/11/obama-first-congressional-address
- Jonathan Bouchet, *NLP analysis on the SOTU addresses*
  - https://www.kaggle.com/jonathanbouchet/nlp-analysis-on-the-sotu-addresses
- Jennifer Dixon, *Presidential Speech Analysis*
  - https://github.com/jennifro/Feeling-Presidential
- Frank Evan, *Topic Modeling of the State of the Union Address*
  - https://dzone.com/articles/topic-modeling-the-state-of-the-union-address-with
- FiveThirtyEight, *Presidential approval poll aggregator*
  - https://projects.fivethirtyeight.com/trump-approval-ratings/?ex_cid=rrpromo
- UCSB, The American Presidency Project
  - http://www.presidency.ucsb.edu/sou.php
- Roper Center for Public Opinion Research Presidential Approval Project
  - https://presidential.roper.center/