

States of The Unions: Presidential speech analysis

Metis Project 4:



Unsupervised Learning

By: Gabriel Vieira Equitz



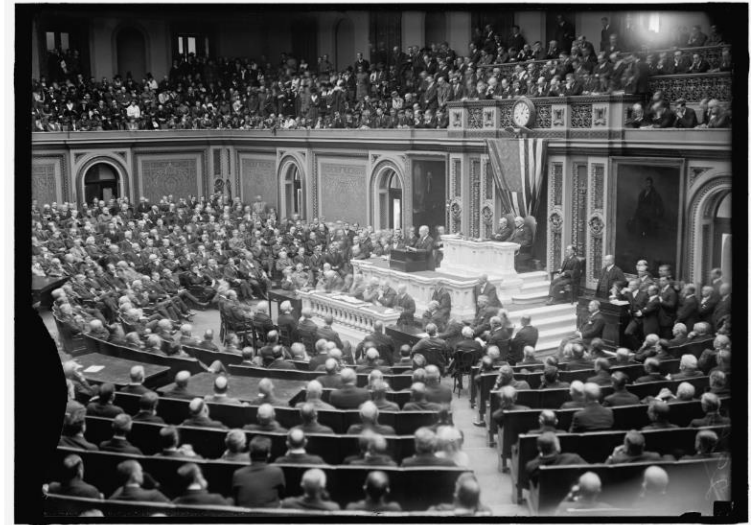
Presented: February 25, 2021

State of the Union addresses 1790-2018

Every year, since 1790, the **President** has given an **address to Congress**, talking about the **condition of America** and its people

All speeches until 2018: **228 total**

from Washington (1790) to Trump (2018)



What we are trying to learn?

NLP can give us insights into the content of speeches, that humans may not otherwise get

Business case: **See how topics change over time.** People who are interested in American politics and want a deeper understanding of the first modern Democracy.

- Includes politicians, economists, historians, and you (hopefully)



How do we Process Natural Language (NLP)

- Topic Modeling can be used for Unstructured Learning
- Exclude common words (e.g. “Shall”, “America”, “People”) and outlier “words” (e.g. “aaa”, “(applause)”)

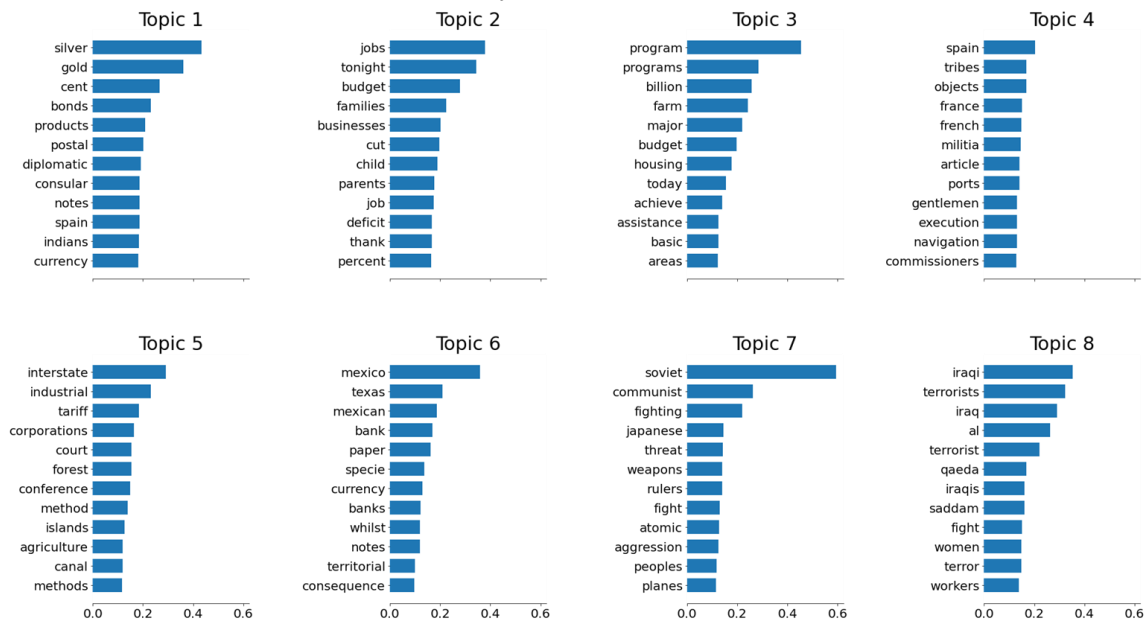
Max_df = 0.42

Min_df = 0.01

Topics

Using NMF, I narrowed down to **8 topics**, and focused on the **top 12 words** of each topic

Topics in NMF model



Groups

The 8 topics are further divided into **3 groups**:

- 1) Topics 1 & 2: Economics

- 1) Topics 3 & 5: National Development (Industry & Gov't Institutions)

- 1) Topics 4, 6, 7, 8: War and Foreign Policy

Topics 1 & 2: Economics Then & Now

Topic 1

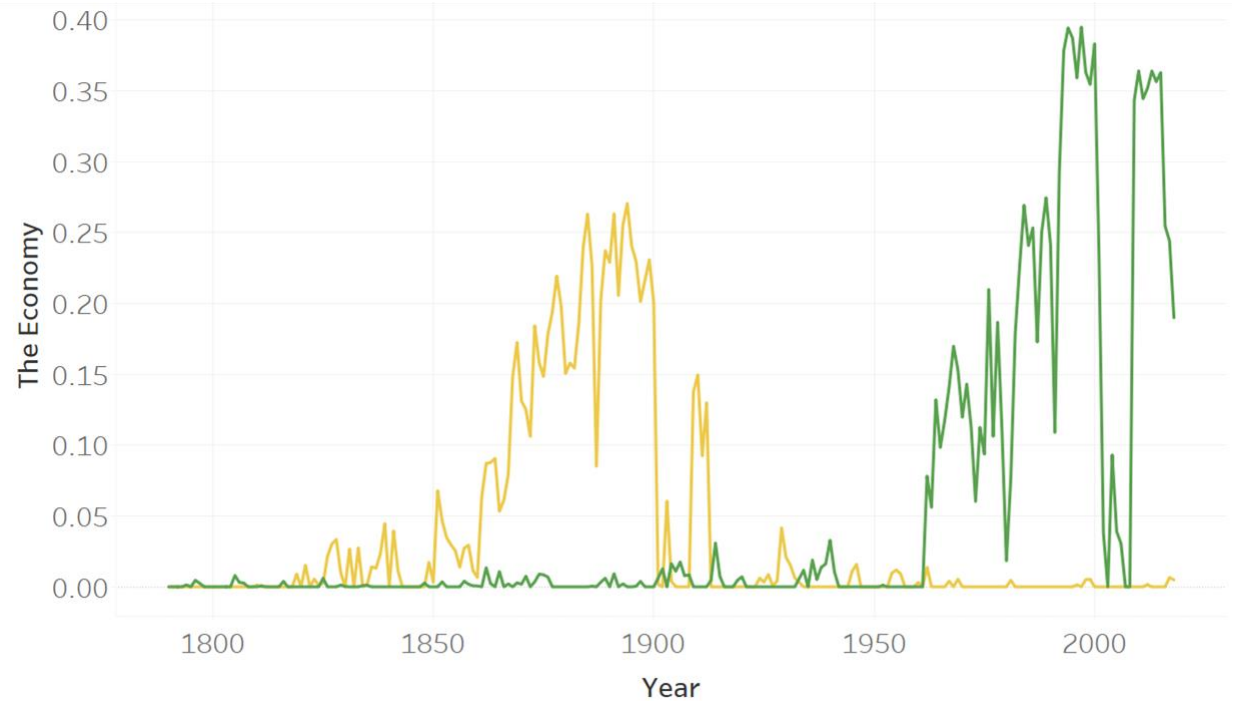
19th Century Econ

(In Gold)

Topic 2

Modern Economics

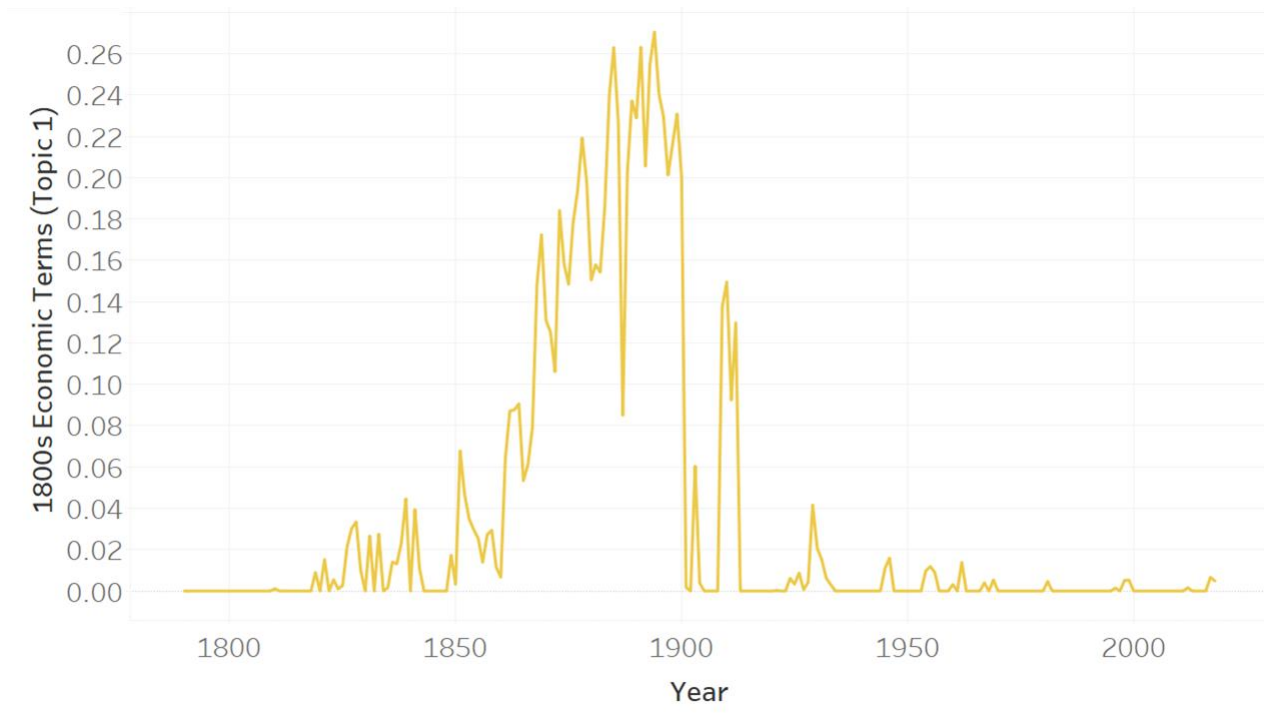
(In Green)



Topic 1: 19th Century Economic Language

Relevant words

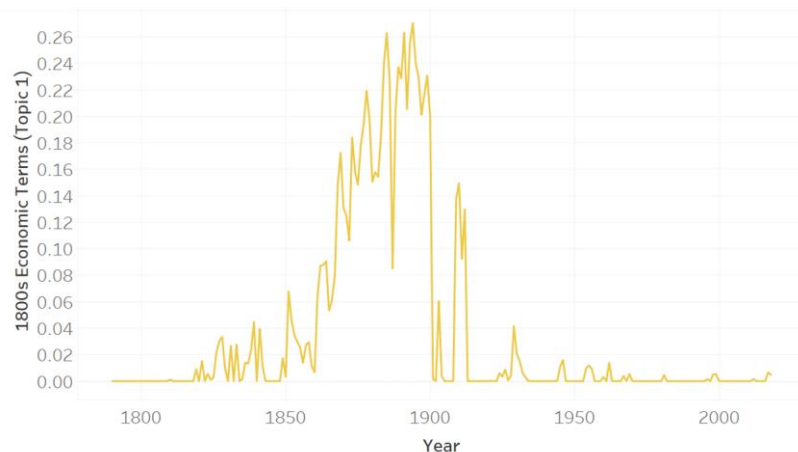
Silver
Gold
Cent
Bonds
Notes
Currency



Topic 1: Economics Then

Topic 1 reflects economic language from the 19th century (peak is from 1869-1900, w/ resurgence under Taft 1909-1912)

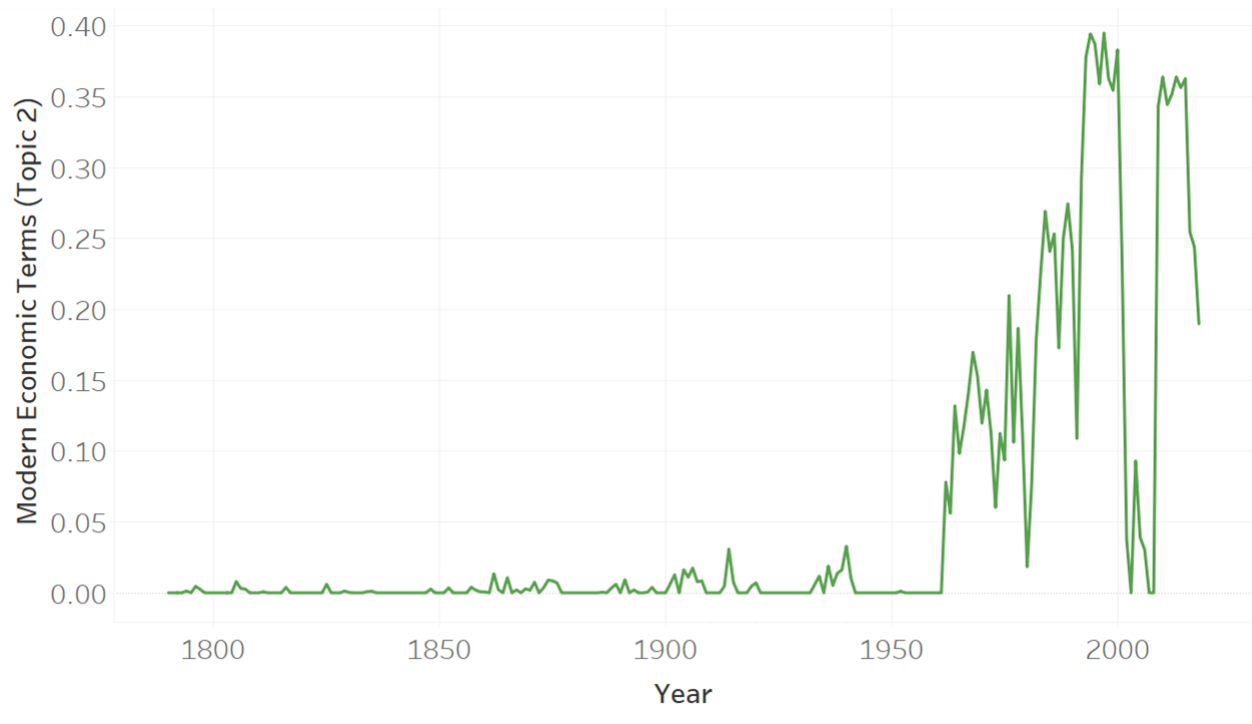
- Apparent Gold & Silver dominated late 1800s
(Bimetallism debate was big)
- Terms like Cents, Bonds, and Notes are less used today, showing how the Presidential language used has changed



Topic 2: Modern Economic Language

Relevant Words

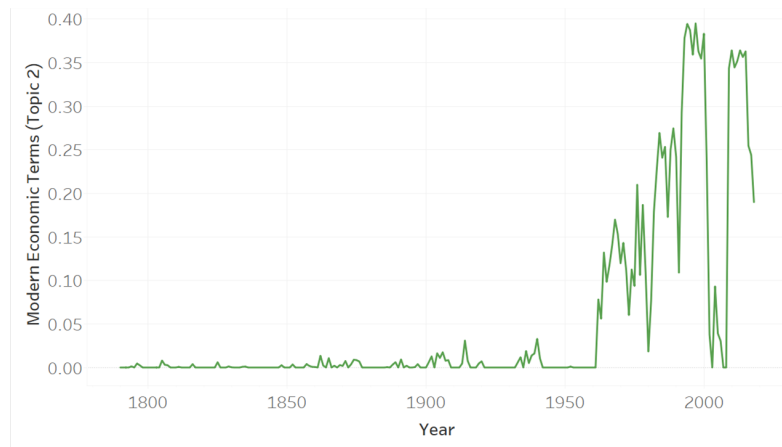
Jobs
Budget
Businesses
Cut
Job
Deficit



Topic 2 Economics Now

Topic 2 meanwhile shows how more recent presidents talked about the economy. In addition, presidents used words “Families”, “Child”, and “Parents” more

- Topic 2 started rising under JFK in 1961.
Growth until peak w/ Clinton (1993-2000).
- Sharp decrease in 2002 following 9/11.
Doesn't rise again until Obama in 2009



Topic 3 & 5: Industry Institution Development

Topic 5

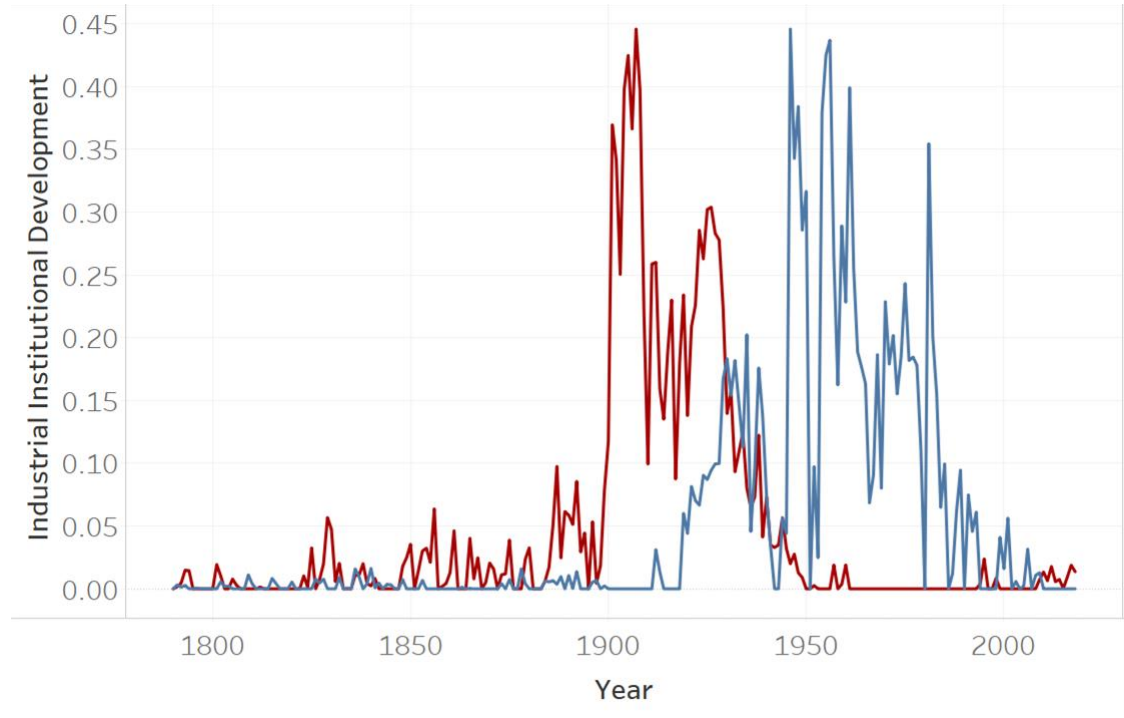
19th Century Econ

(In Blue)

Topic 3

Modern Economics

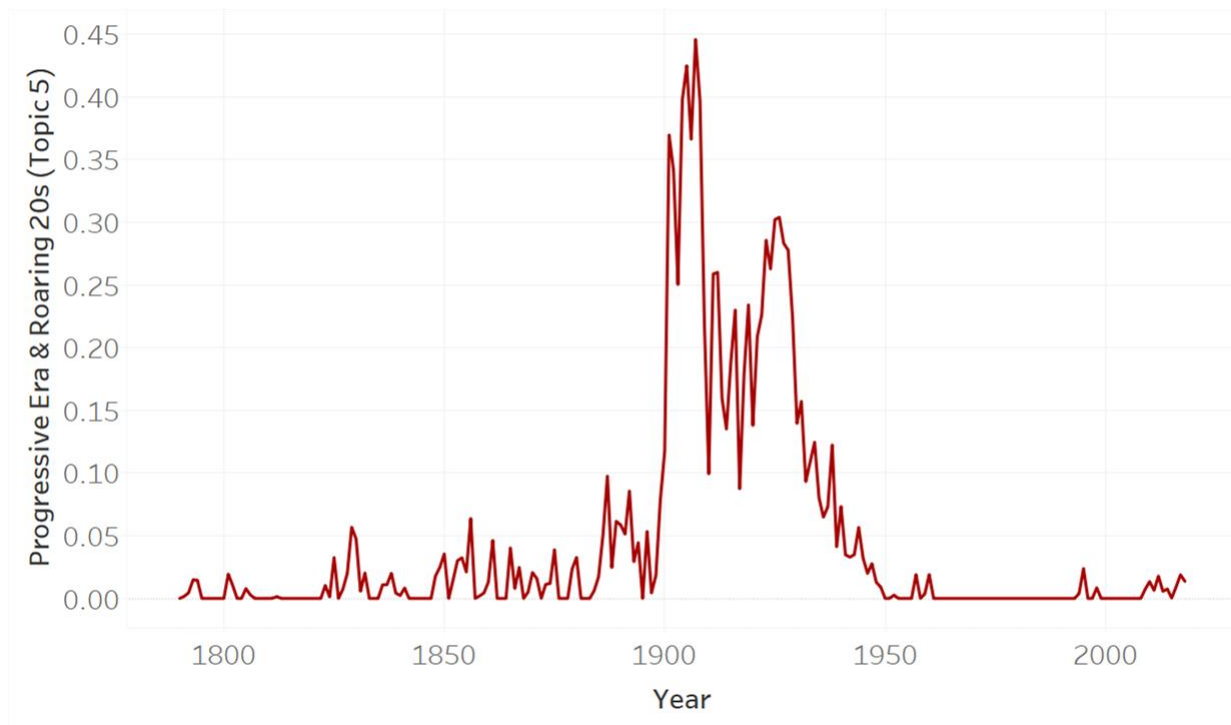
(In Red)



Topic 5: Progressive Era & Roaring 20s

Relevant Words:

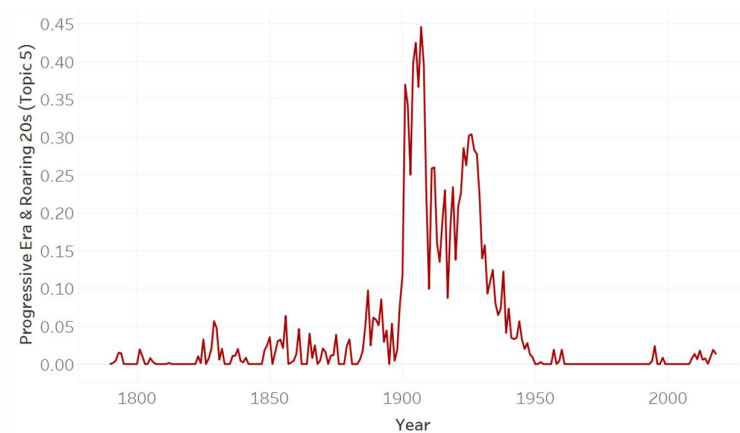
Interstate
Industrial
Tariff
Corporations
Court
Canal



Topic 5: Industrial Revolution

Peaks in 1900-1930, Progressive Era & Roaring 20s

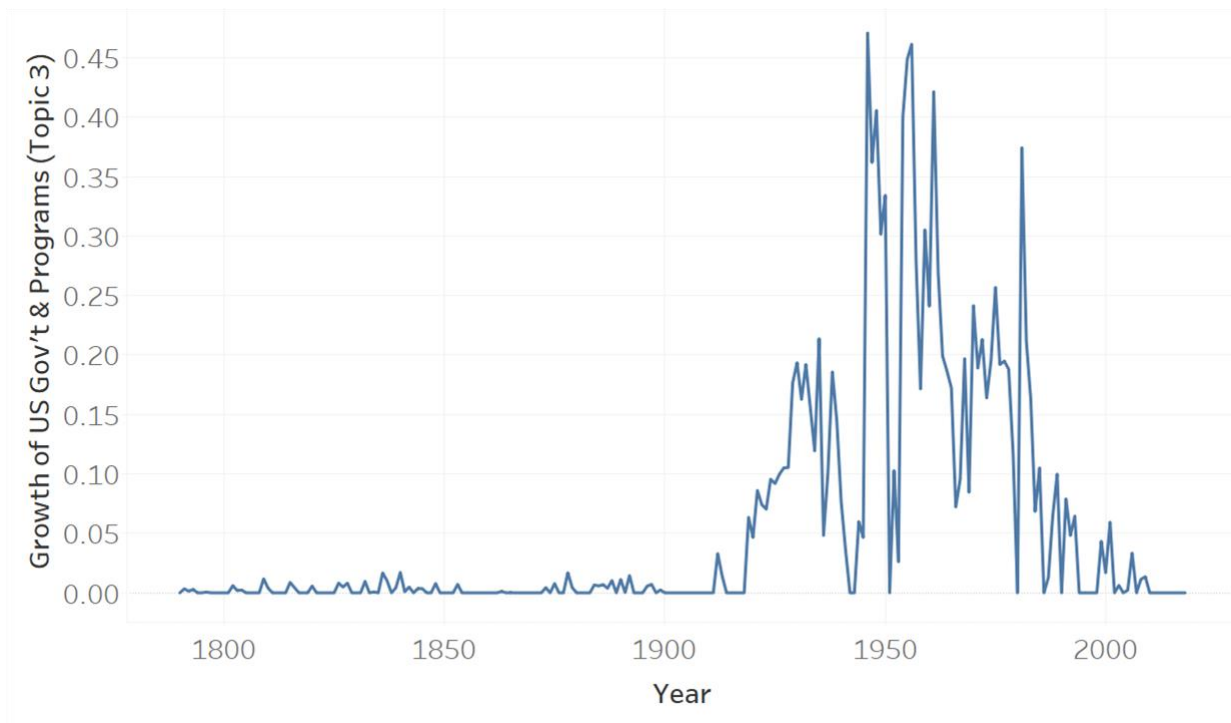
- Represents the industrial revolution and The rise of capitalism and increased trade
- Moderate pause during WW1
- Fizzles out in 1930, Great Depression
- Correlation with relatively high peace



Topic 3: Growth of US Gov't (+Departments)

Relevant Words:

Program
Programs
Farm
Budget
Housing
Assistance



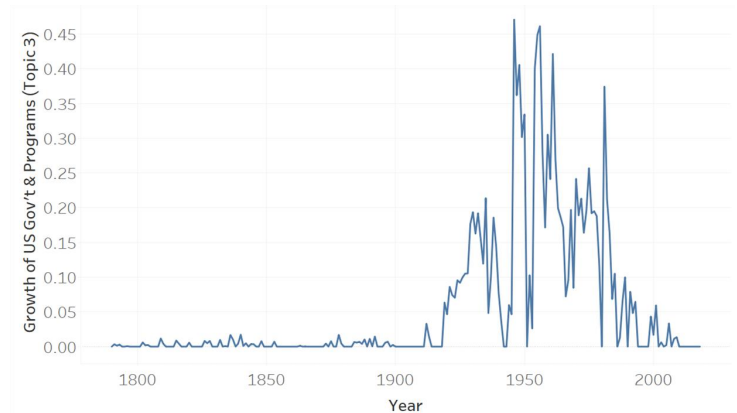
Topic 3: More Public Programs

This represents the creation and expansion of federal welfare programs

- Early peak correlates w/ New Deal (1930s)
- Apex under Truman, Eisenhower (1946-1961)

After WW2

- Lowers until Reagan, in which this topic is
More or less gone by his 2nd term (1984)



Topics 4/6/7/8: WAR! (and foreign policy)

Topic 4

1790-1830

(In Blue)

Topic 6

1830-1880

(In Gold)

Topic 7

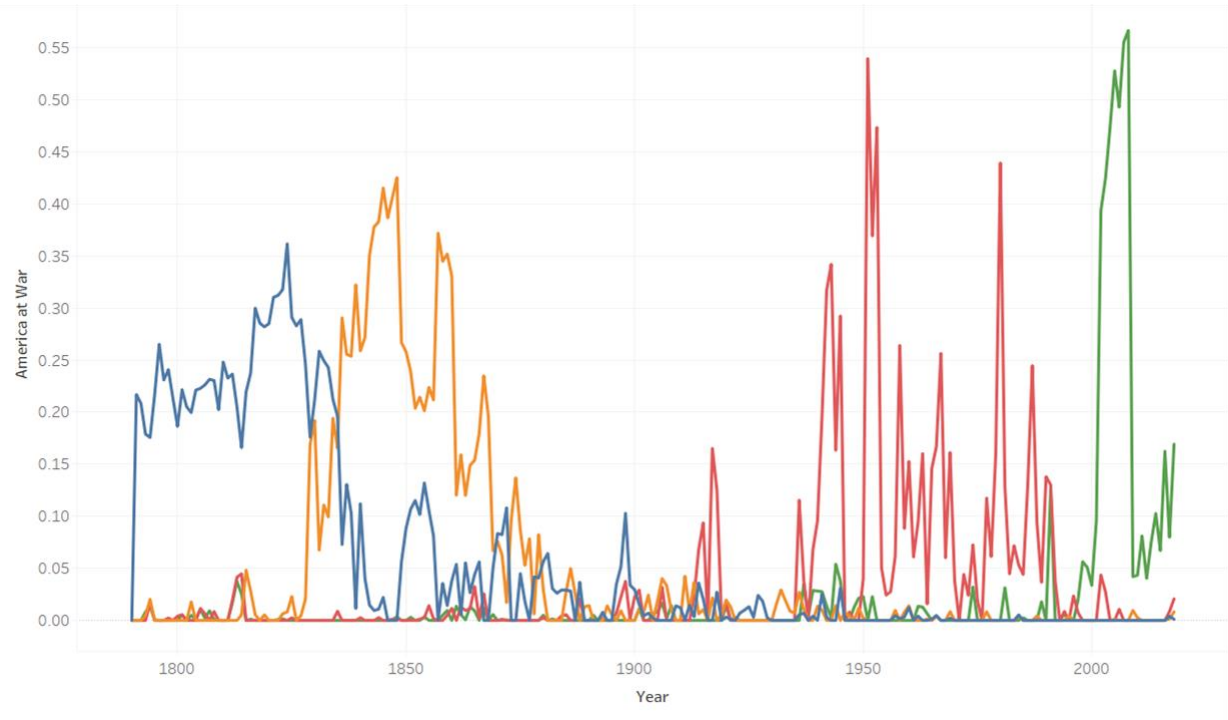
1940-1990

(In Red)

Topic 8

2002-now

(In Green)



Topic 4: Early Foreign Policy & War

Relevant Words:

Spain

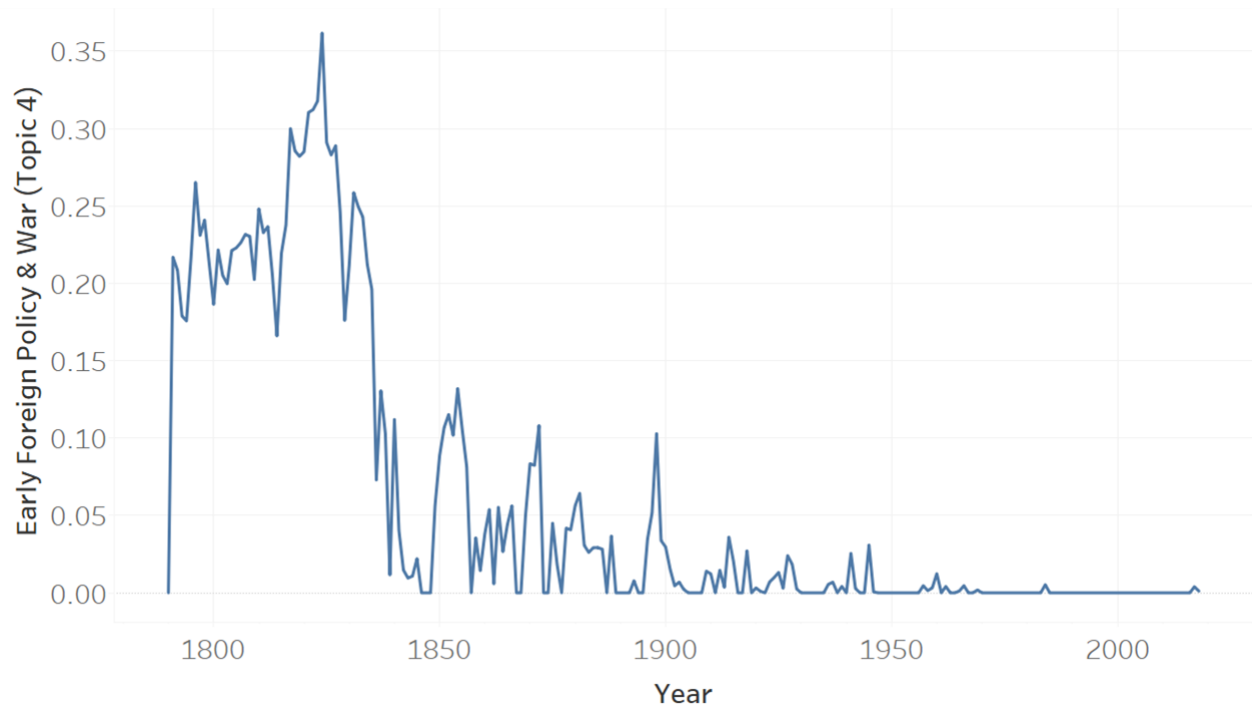
Tribes

France

Militia

Execution

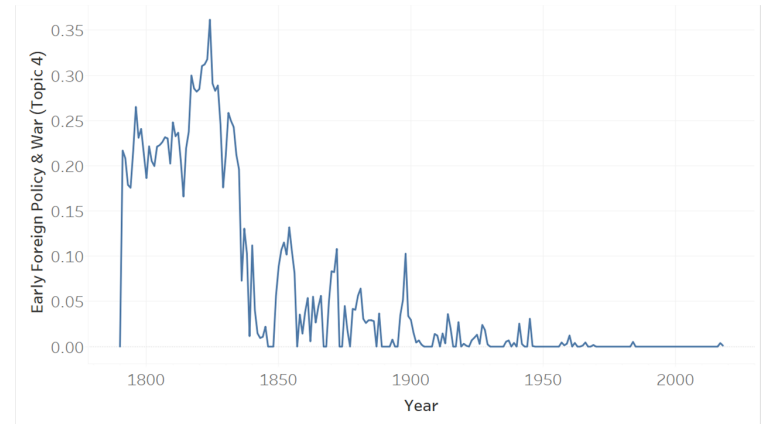
Commissioners



Topic 4: Olde War (1790-1830)

These words reflect the nature of war and foreign policy in the times of the founding fathers

- The US bought land from the French and Spanish. Took land from Native Americans
- Militia, Commissioners and other such Terms are almost never in SoTUs now (refers to lack of standing army)



Topic 6: 1830-1880 in War

Relevant Words:

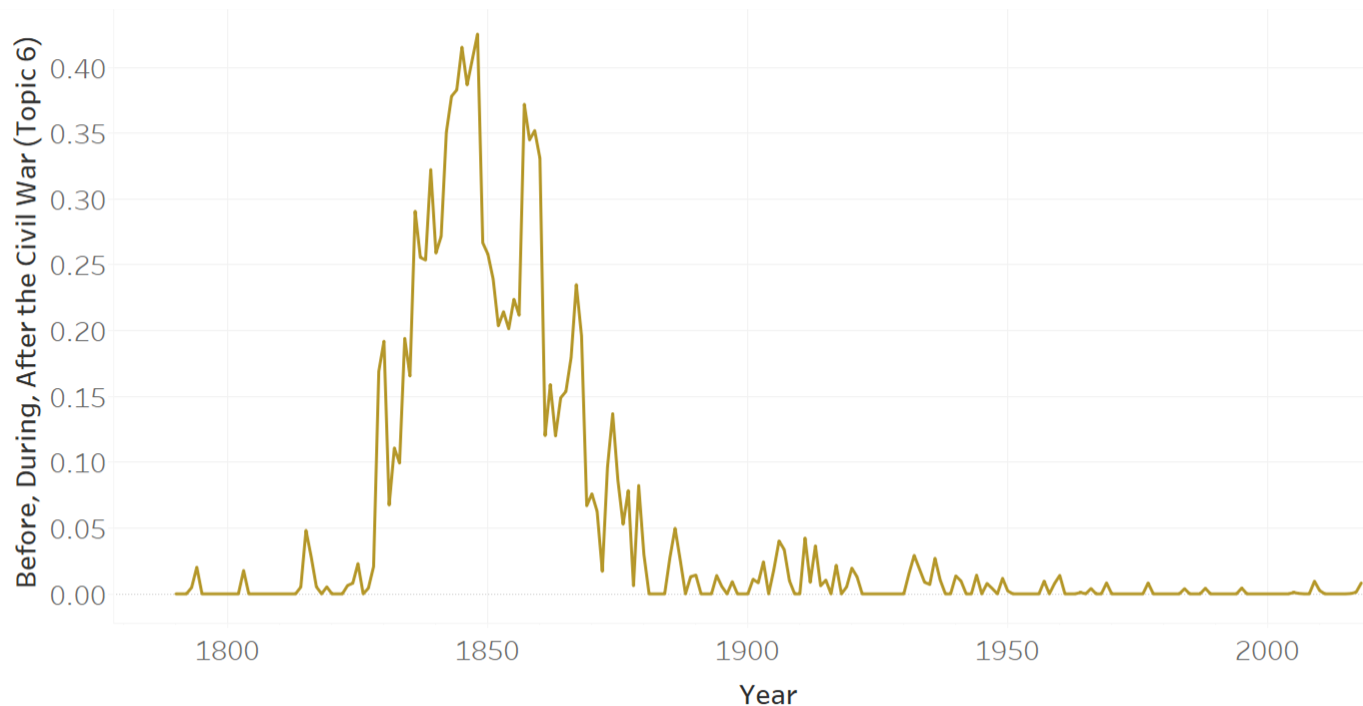
Mexico

Texas

Mexican

Territorial

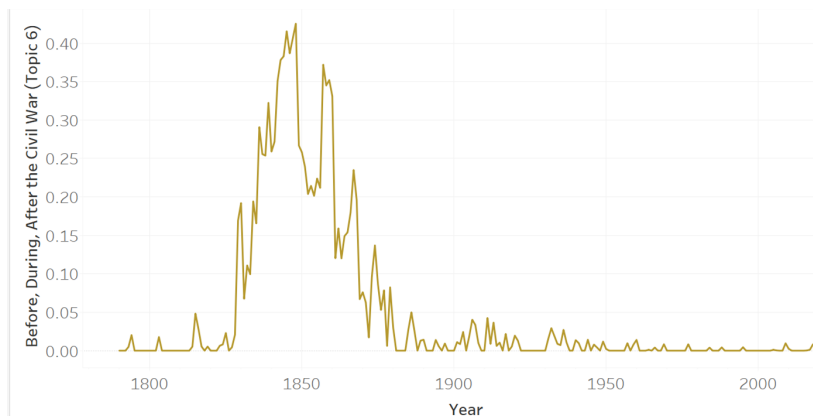
Consequence



Topic 6: Before, During, After the Civil War

This topic highlights the period from 30 years before the US Civil War to 15 years after

- Starts with Jackson (1830), but peaks
in Mexican-American War (1846-1848)
- Remains high through Civil War, and
Basically ends with Reconstruction
(end of US military occupation of South)



Topic 7: World Wars & Cold War

Relevant Words:

Soviet
Communist
Fighting
Japanese
Atomic
Aggression



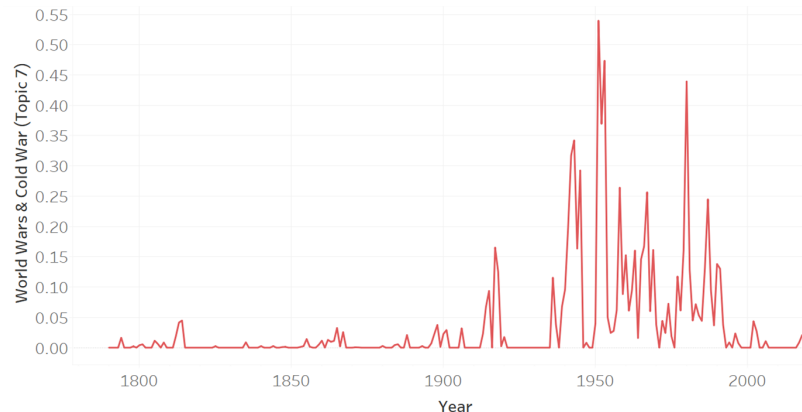
Topic 7: War has changed, Again

We see references to a new world order and new technologies

- Massive spike a response to escalation of Cold War (Russia develops Nuclear bombs

And escalation of Korean War (1951)

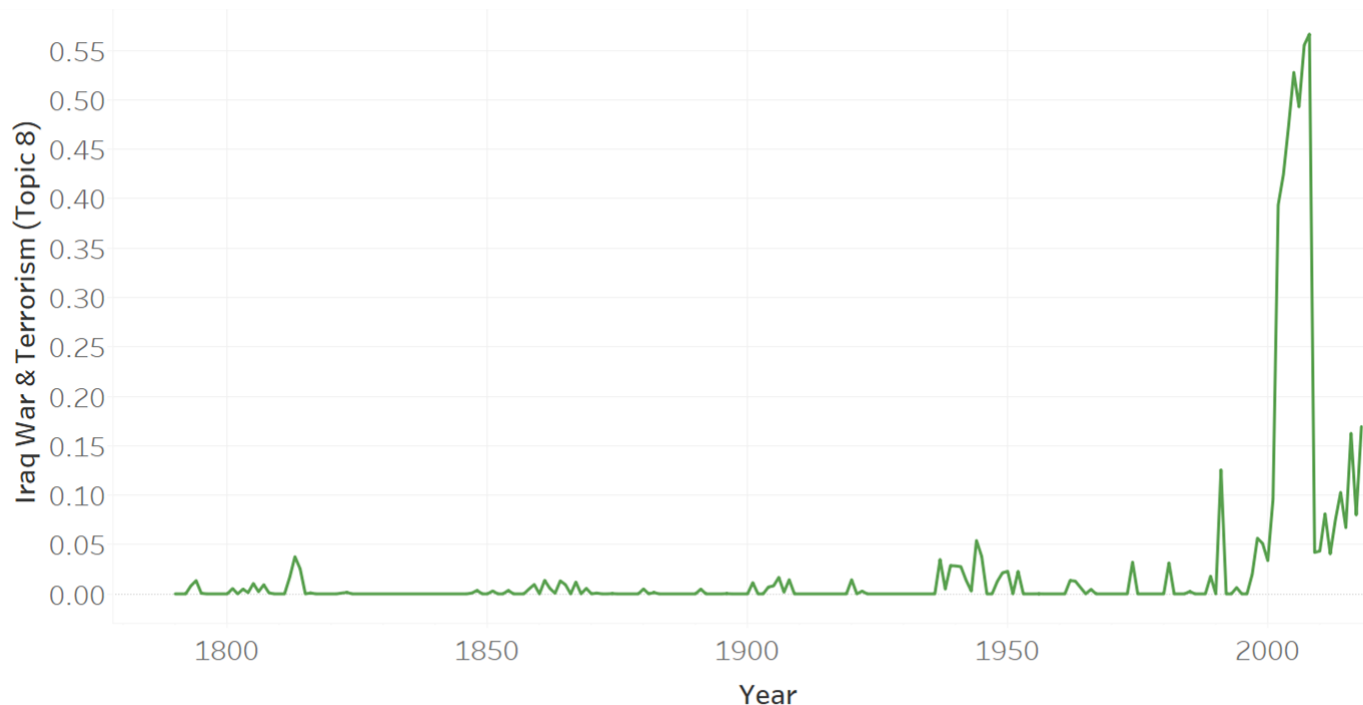
- End corresponds w/ fall of the USSR (1990)



Topic 8: Iraq War & Terrorism

Relevant Words:

Iraqi
Terrorists
Qaeda
Saddam
Fight
Terror



Topic 8: This is Modern War

The aftermath of 9/11 and the Iraq War defined the presidency of George W. Bush

- 2001 speech is Topic 2, but Topic 8

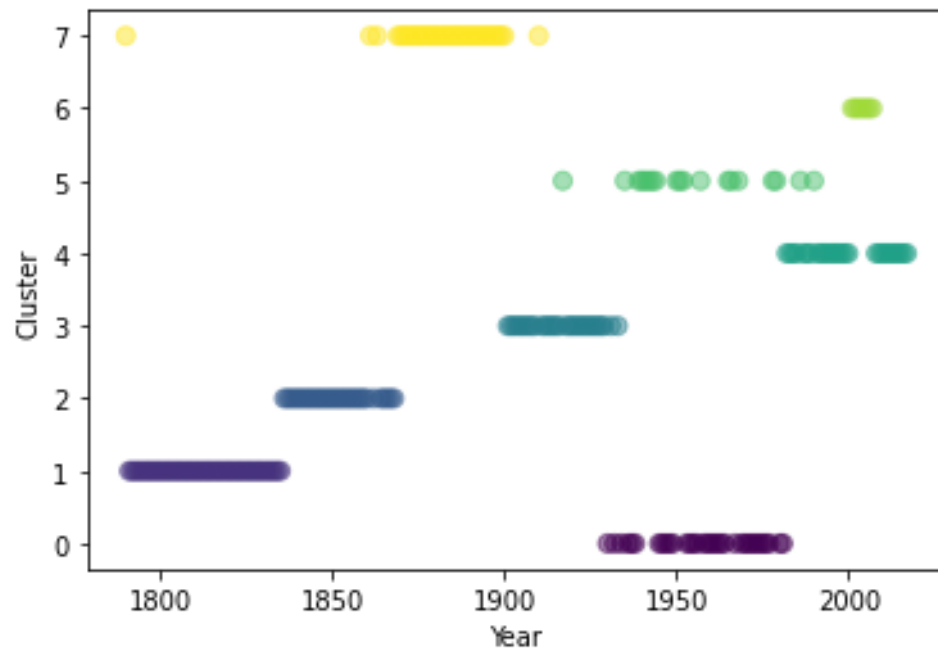
Dominates 2002-2008

- Rise w/ Bush Sr, Gulf War (1991)
- Somewhat increases w/ ISIS (2014-)



Scatter Plot

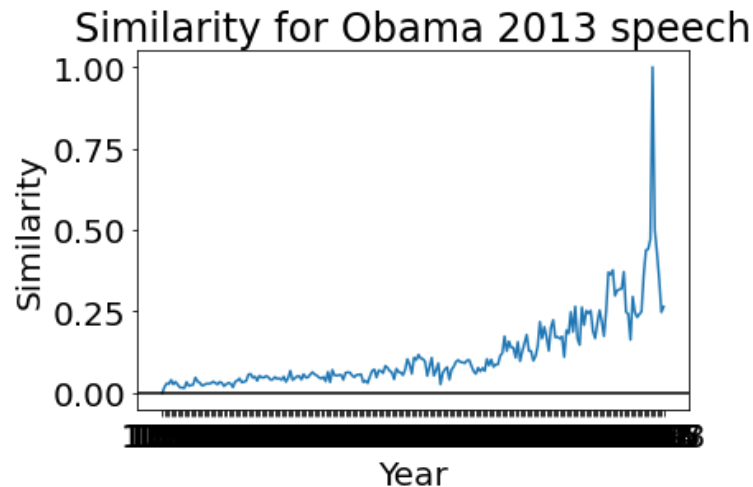
Correspond to largest topics for each
speech/year



Most Similar Speeches

The most similar speeches, when not by the same president, are typically those of their immediate predecessors and successors.

- Obama is an exception: His speeches are Most similar to Clinton



Conclusions

- Topics 1 & 2 show how presidents State of the Union speeches have changed in regards to economics (**less Silver & Gold, more Jobs & Budgets now**)
- Topics 3 & 5 show how language on industry and government institutions have evolved with the industrial revolution and increased size of the US gov't
- Topics 4, 6, 7, 8 show how war has changed: Who is fighting, what are the technologies, and tactics

Also: 200 years ago, 'gentlemen' was a top word, now 'women' is a top word
(Topic 8 Iraq)

Potential Future Work

Add and compare economic data (GDP/GNP)

Analyze by party, see if Democrats and Republicans switch in 1960s

Does American politics have 40 year cycles? (Schlesinger)

- 1780-1820, 1820-1860, ...1940-1980, 1980-2020

Addendum

Word Cloud!

WordCloud of Lincoln_1864.txt



The End

