

# Top Issues in America

*According to Party Manifestos & PollingReport.com*



David Wang

Stern Program for Undergraduate Research

Professor Vishal Singh

05/17/16

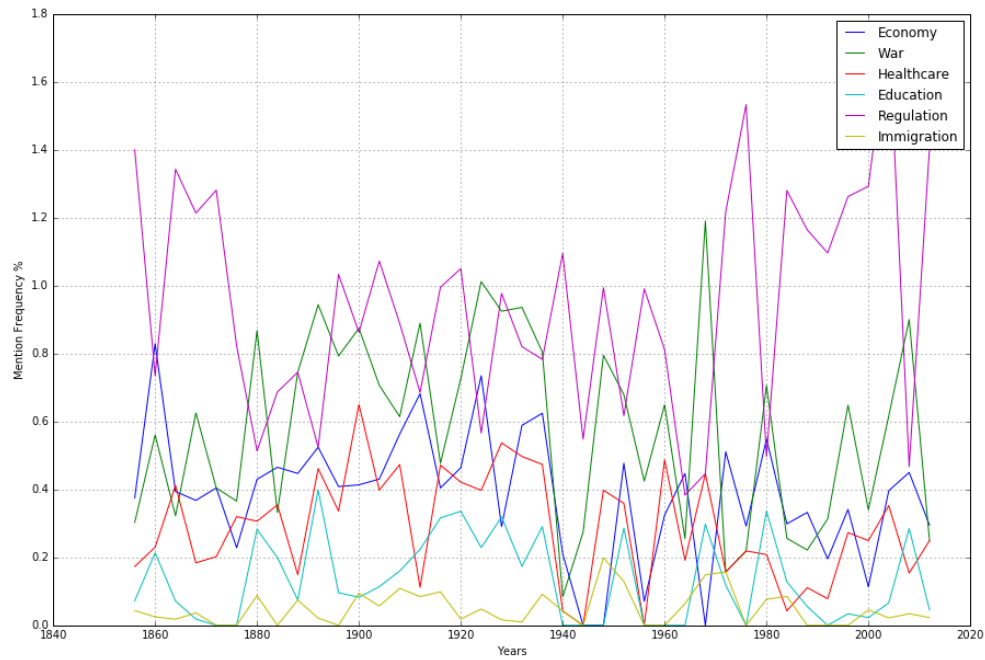
## Overview

The goal of this analysis was to analyze how the importance of certain issues have changed in America, over time. I used two data sources and various techniques to analyze this:

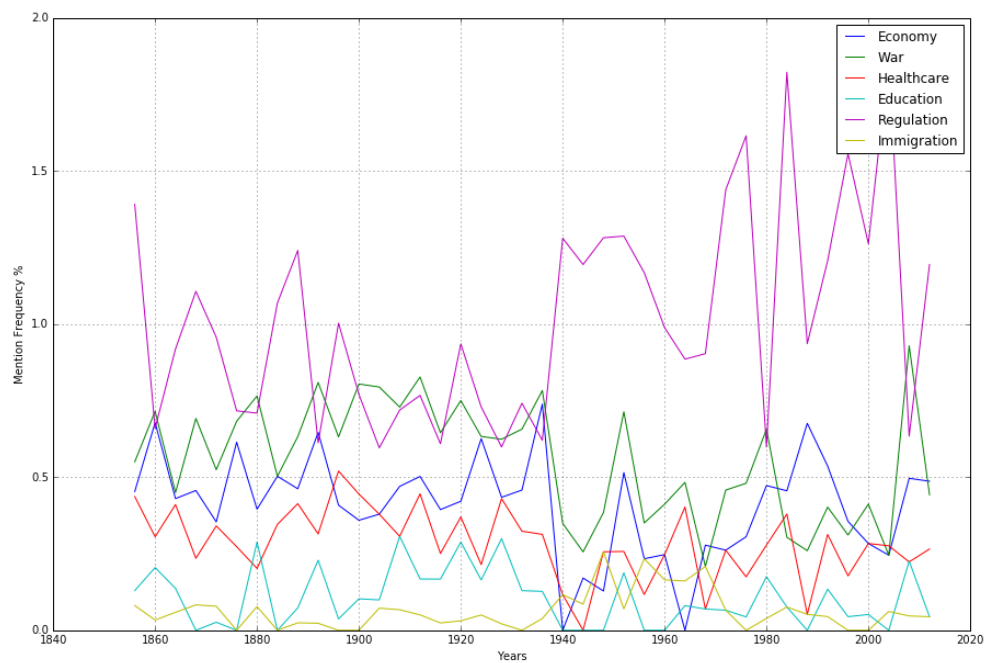
1. Party manifesto data
  - a. Every four years, Democrats and Republicans release a party manifesto that outlines the party's goals and accomplishments, as well as the key problems to address both in America and within their party.
  - b. The files were classified by year and by party, from 1856 to 2012.
  - c. Process:
    - i. Data Wrangling:
      1. Imported TXT files for each speech and converted them to a string. Each string was then stored in a dictionary as a value, with the key being the year
      2. Used NLTK to clean the data through tokenization and lemmatization
      3. Sifted through each speech and counted the number of times certain topics were mentioned (matched against a list of words frequently used to describe a topic), and divided this count by the total number of words in the speech to get the Mention Frequency % (or how often this topic was talked about that year by each party). These values were then stored in dictionaries, with year as the key, and used for various visualizations.
    - ii. Visualization & Insights
      1. Used Pandas, matplotlib, and ggplot2 to create frequency distributions, multi-series line plots, a correlation matrix, and a stacked bar plot
2. PollingReport.com data
  - a. PollingReport.com collects poll results conducted by major news websites, such as CNN, Fox, and CBS, and organizes them by theme – the theme this analysis focuses on is top issues in America, from 2002 to present day.
  - b. Process:
    - i. Web Scraping
      1. Built a Python web scraper to analyze all the topics in each poll, however the HTML trees were different for many of the polls. This inconsistency invalidated the web scraper and made it hard to gather the accurate polling result percentage. Therefore, the rest of the data was manually inputted.

## Party Manifesto Data – Results

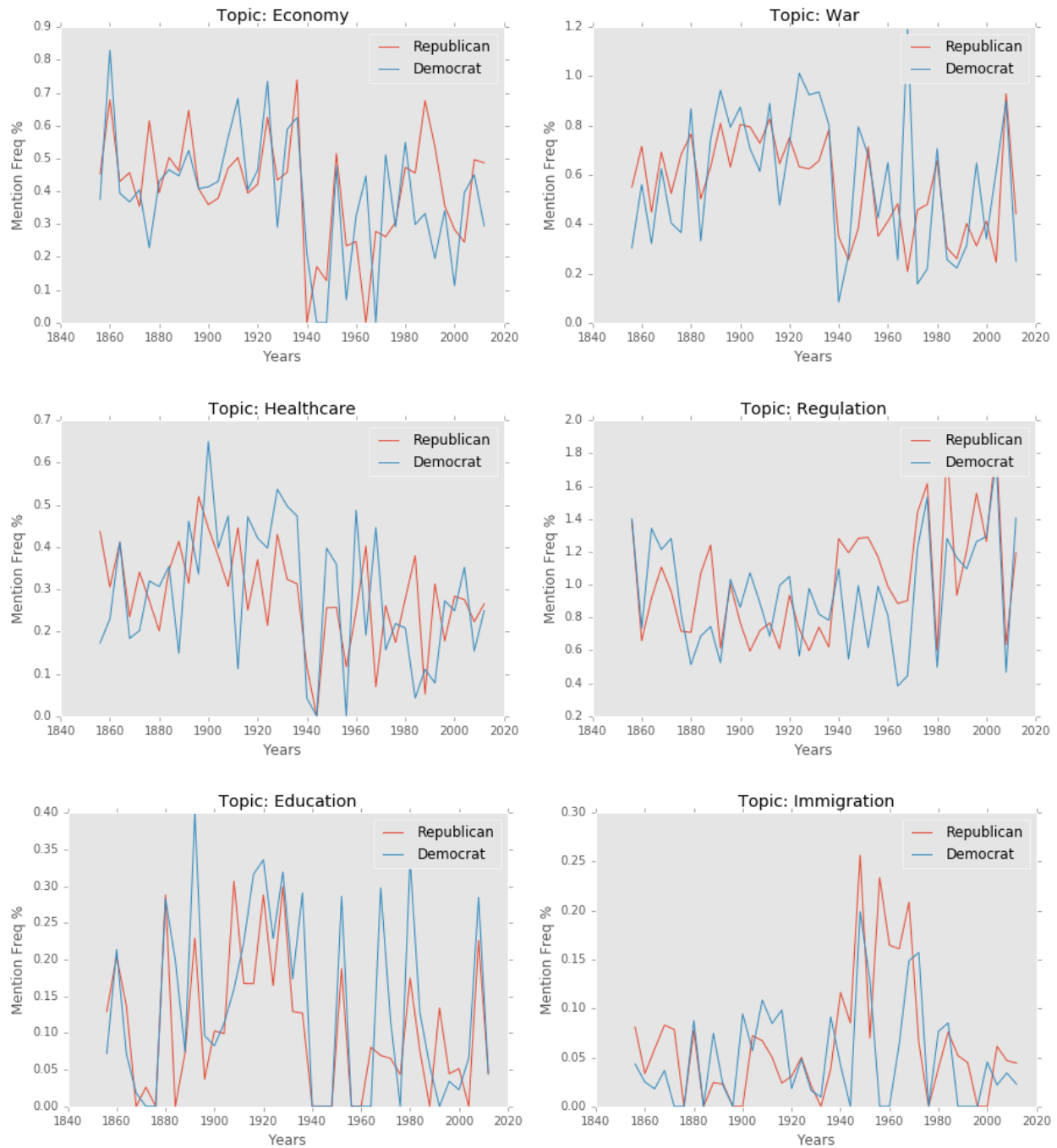
*Democrat Mention Frequency per Issue 1856 - 2012*



*Republican Mention Frequency per Issue 1856 - 2012*



## Individual Issue Comparison between Parties



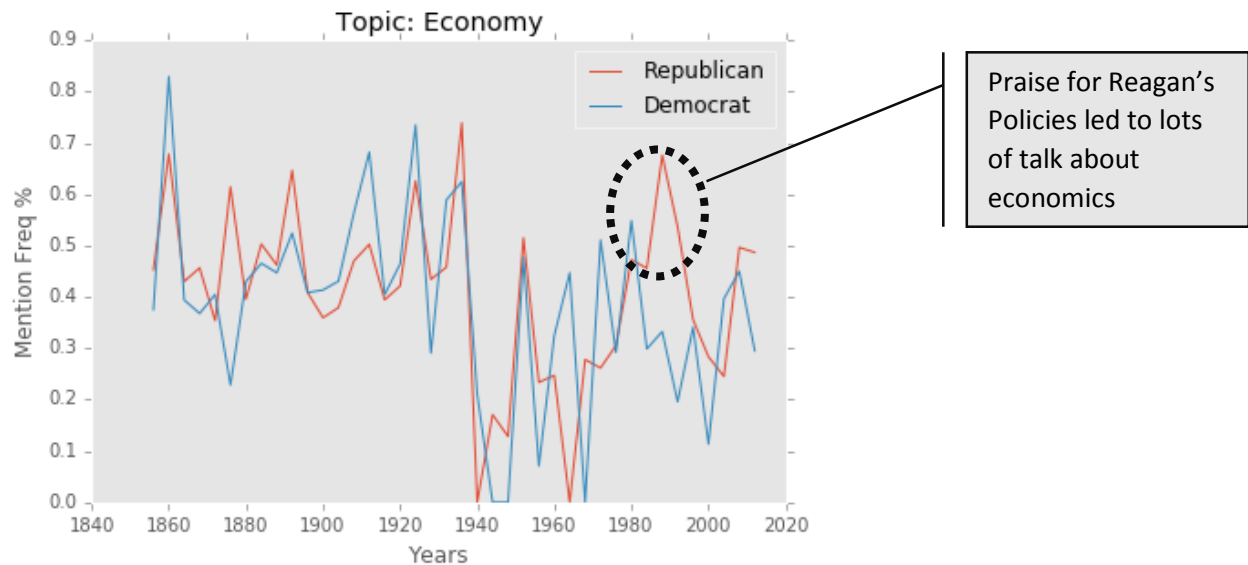
Here, we can see that Republicans tend to talk more about socioeconomic issues, such as the economy and regulation. They also tend to talk more about immigration.

For Democrats, we can see that they have a heavier emphasis on social issues, such as healthcare and education.

## Interesting Takeaways

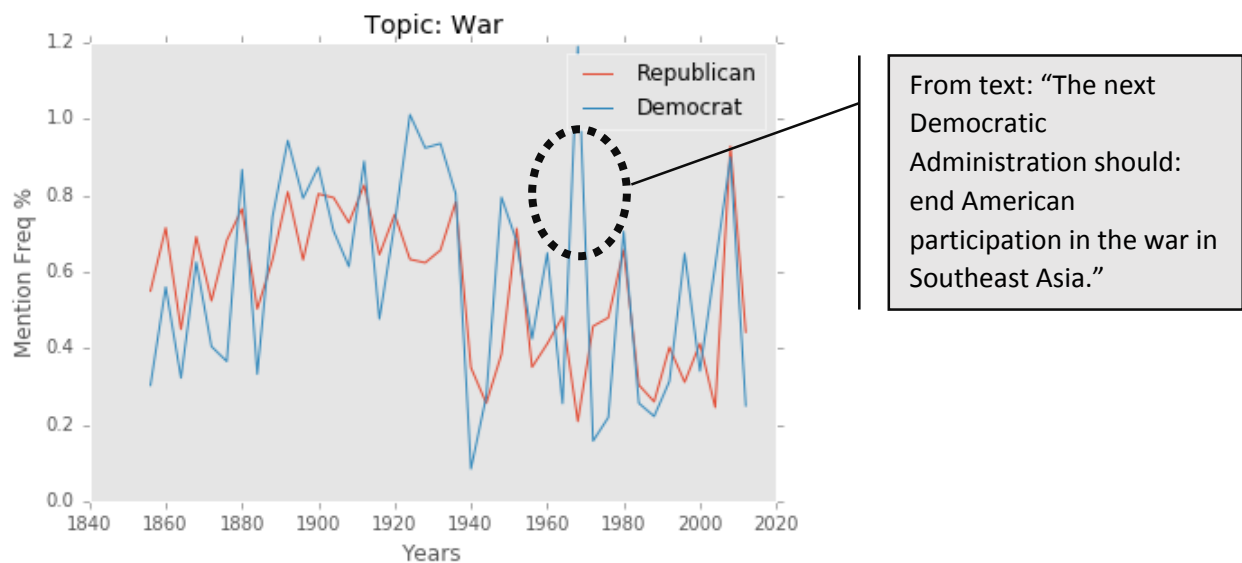
### Reagonomics

Taking a closer look at the issue of Economy, Republicans talked a lot more about this issue around the end of the 1980s. Going back into the text, I saw that the Republicans talked a lot about Reagonomics and were praising about how his economic policies benefited the country. From the text: "The Reagan revolution has become a republican renaissance. Our country's back in business and back on top again."



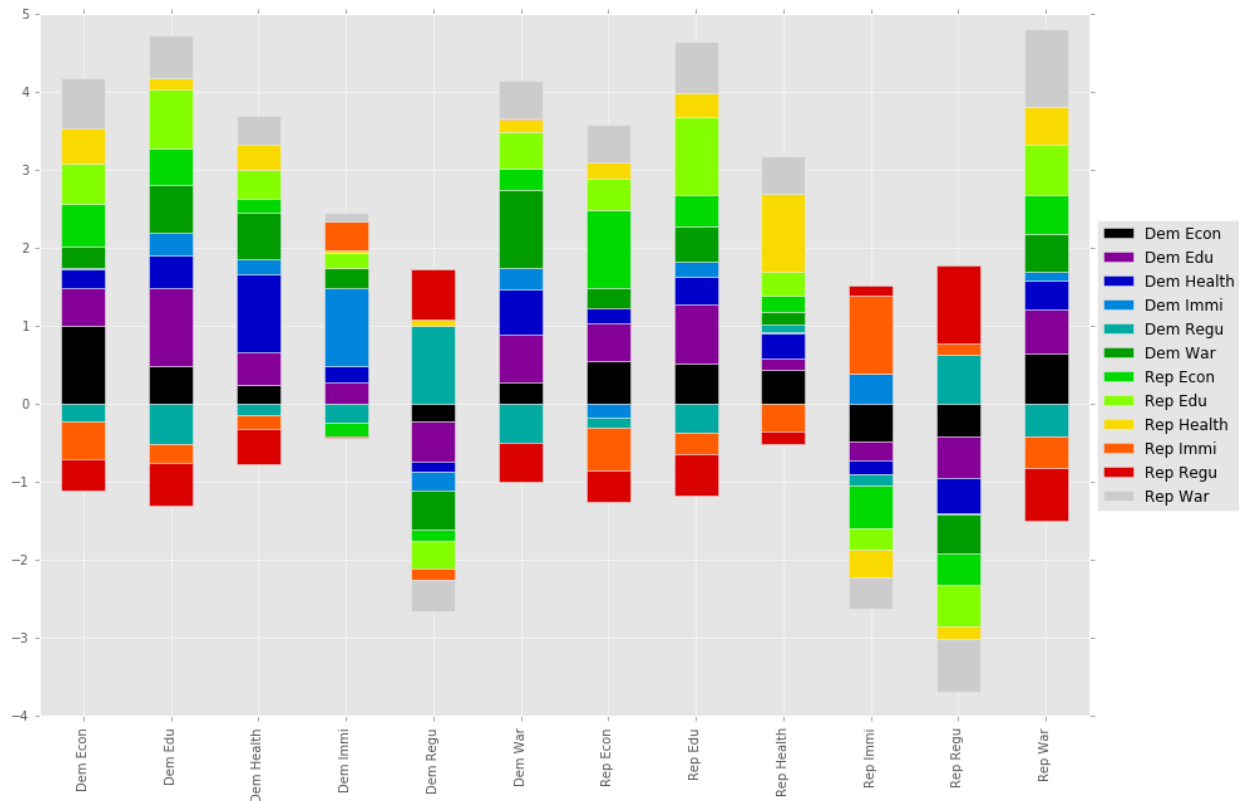
### Vietnam War

Looking at the issue of War, Democrats had a huge spike in Mention Frequency % for this topic around the late 1960s. Checking into the text for those years, I found that Democrats heavily opposed the Vietnam War.



## Issue Correlation

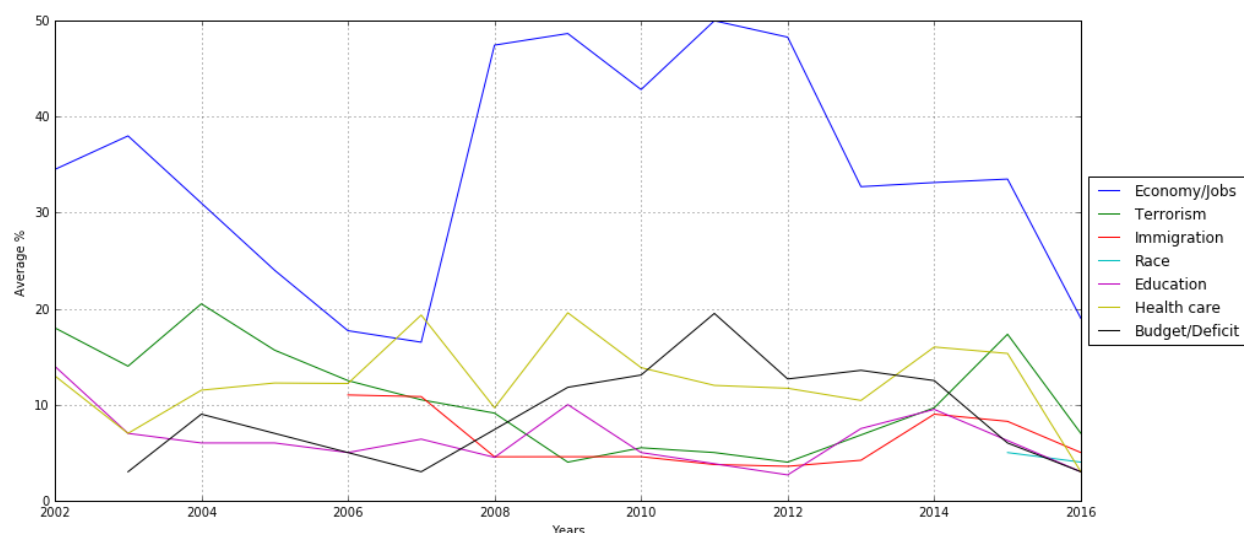
The goal of this visualization was to see which topics are frequently talked about together. This was done by producing a correlation matrix between all the topics, and then creating a stacked bar plot. For each category, the bigger the box within each bar represents more correlation.



From this, we see that the topics of healthcare and education are frequently talked about together within the same year, between parties and across parties. In addition, the topics of economy and immigration are talked about less together, within the same year.

## PollingReport.com Data – Results

Using the Jupyter Python development environment, I was able to issue SQL queries directly into Pandas dataframes, and then manipulate these dataframes with matplotlib to create interesting visualizations.



This graph shows the aggregated polling results across all polls conducted within each year, regardless of source (this normalized for differences in readership demographics for each poll for each news source). These polls asked Americans what issue was the most pressing in America at that time – therefore by aggregating these polling results, we have an interesting snapshot of what Americans thought were the most important issues facing America and how it has changed given the issues that come and go. This is an interesting perspective because the previous data source was from the viewpoint of politicians and part leaders. However, the data here was limited to 2002, which was when PollingReport.com first started collecting the data.

### Interesting Takeaways

- The issue of economy took a huge spike during the financial crisis
- The issue of terrorism has taken a more recent spike in recent years given the rise of ISIS in the Middle East
- In 2011, the budget crisis sparked public outrage, and thus received higher polling percentages.

## Conclusion

In conclusion, the analysis shows interesting trends throughout history and recent years in terms of how politicians and citizens viewed the issues that affected our country. This is important because we see that the issues the public deems important are often quite near-sighted, and affected by events that recently happened. The rhythmic rise and fall of certain issues in the polls gives us insight into the attention span of people when key events happen, such as the financial crisis.

This analysis could be improved and added upon in numerous ways, most notably with sentiment analysis in the party manifesto data to understand whether the party talked negatively or positively about certain topics in a particular year. The main objective of sentiment analysis would be to understand party rhetoric, as well as see if politicians actually met the goals and attitudes they set out in their manifesto. For the PollingReport.com data, further analysis could be done in terms of comparing polling responses by news source to understand how certain reader groups view issues.

For Python and SQL code, as well as data sources, please see Dropbox folder.