# Text Analysis: A Timeline of the Federal Reserve Board's Interests

Atul Lanka
Northwestern University
atul.lanka@northwestern.edu
11/24/2020

## INTRODUCTION

The Federal Reserve – or the "Fed" - is the central bank of the United States solely responsible for enacting monetary policy as well as bank regulation and supervision.

When commenting on the economy, the Fed tends to communicate in "Fedspeak," a term given to the “turgid dialect of English” used by the Federal Reserve Board chairmen. That is, the Fed is known to make verbose and more ambiguous statements when commenting on the current state of the economy. The intention behind this is to reduce market instability

This paper will highlight the Natural Language Processing and Machine Learning methods, in combination with domain knowledge to:

- Categorize the Federal Reserve Chairman’s speeches into categories or topics of interest
- Identify trends in topics that each Chairman has uniquely discussed in their term
- Identify the numerous agencies they've created or supported in order to implement their policies
- Compare quantitative measures and clusters against key economic indicators such as Fed Funds Rate and Unemployment as well as historic events that have impacted our economy in the last two decades (2000 dot com bubble, 9/11, 2008 Housing Crisis, 2020 Covid-19 pandemic) to determine whether there is any palpable correlation.
- Use sentiment analysis to compare tone and impartiality for each Chair’s addresses over their respective terms.

## BACKGROUND

The Federal Reserve ideally comprises of a Chair, Vice Chair and 2 to 4 Governors. This Board of Governors are appointed by the President, which means this corpus will span over 4 presidencies from Clinton to Trump. The Federal Reserve website includes transcripts of all speeches delivered since 1996. Therefore, this paper will focus on the last 24 years, which means it will range from the Clinton presidency to the Trump presidency. Over this period of time, 4 people have served as Chairman – Alan Greenspan, Ben Bernanke, Janet Yellen and Jerome Powell.

This inception of this paper wouldn't be possible without the existing research papers that have been published. Papers and articles about assessing Fed Chair hopefuls [1] and whether an speech will impact the fed funds rate in the future [2] have influenced me to learn more about Federal Reserve.

## DATA

### Summary

As aforementioned, all speeches available on the Federal Reserve website date back to 1996. These transcripts have been further filtered to only include the speeches made by the Chairman. This brought the dataset to 430 speeches. Each speech roughly spans 1500-4000 words depending on the nature of the address.

### Data Scraping

Each speech's transcript is a htm file on the website and so I used the BeautifulSoup package in Python to scrape the text of each speech into a data frame as well as include each speech's metadata for analysis. The metadata includes:

- Title of the speech
- Link to the HTML version in the Fed's official website
- Speaker name
- The year and full date of the speech.
- Raw speech text
- Number of words in the speech.

### Data Cleaning

Once the dataset was created, the raw text itself had to be cleaned in order to make it suitable for analysis. Firstly, all speeches under 250 words were immediately removed, as they would've simply been introductory remarks made with no substance. Subsequently, all references and

hyperlinks in the transcript were removed. All punctuations and contractions – for which a function was created to replace those words with its expanded form – numbers and stop words were removed. The remaining text was then made lower case and lemmatized. Finally, a dictionary of all the tokenized words as well as an array of each speech 'cleaned up' is generated for analysis.

## EDA – Exploratory Data Analysis

Before diving into the textual analysis, provided is visual exploration of the dataset to add more context:

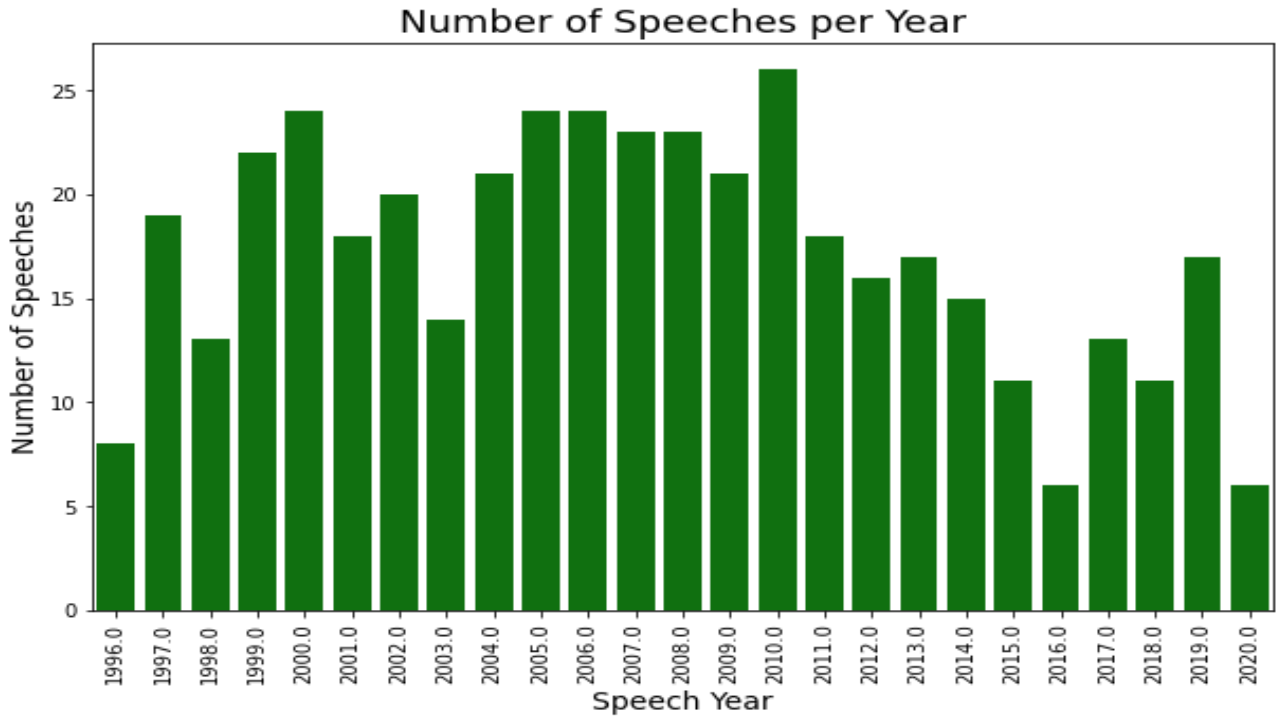
Years as Chairman

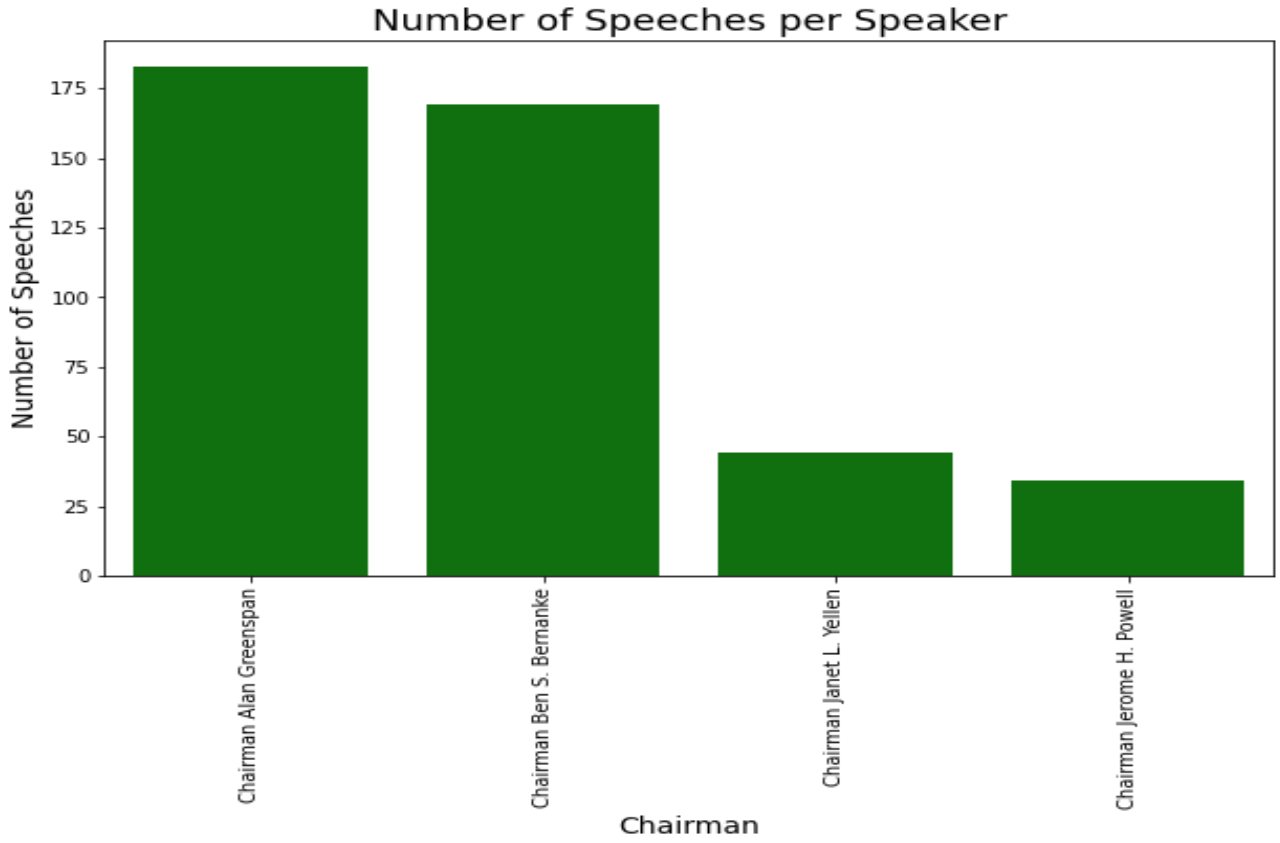
Alan Greenspan: 1996 – 2006 [10 years]

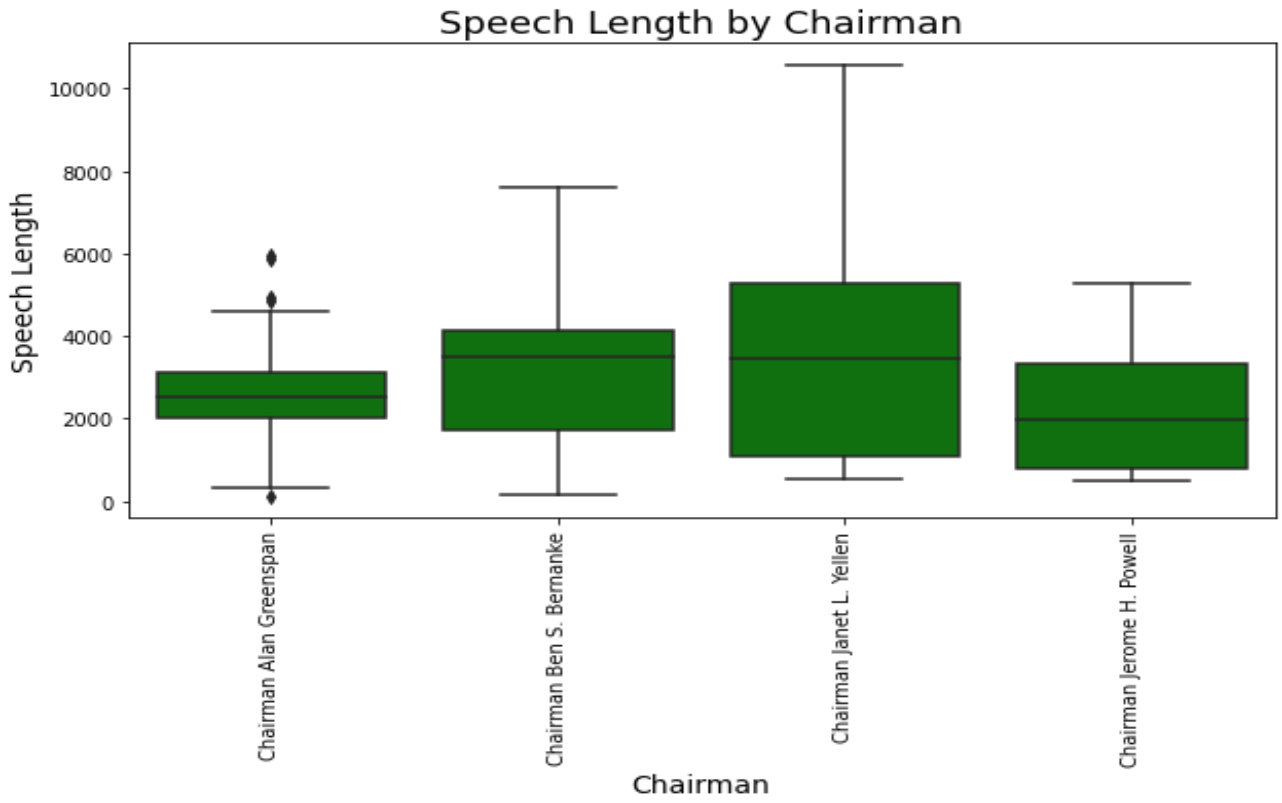
Ben Bernanke: 2006 – 2014 [8 years]

Janet Yellen: 2014 – 2018 [4 years]

Jerome Powell: 2018 – 2020 [2 years; Incumbent]


Number of Speeches per Year
Number of Speeches
0
5
10
15
20
25
1996.0
1997.0
1998.0
1999.0
2000.0
2001.0
2002.0
2003.0
2004.0
2005.0
2006.0
2007.0
2008.0
2009.0
2010.0
2011.0
2012.0
2013.0
2014.0
2015.0
2016.0
2017.0
2018.0
2019.0
2020.0
Speech Year


Number of Speeches per Speaker
Number of Speeches
0
25
50
75
100
125
150
175
Chairman Alan Greenspan
Chairman Ben S. Bernanke
Chairman Janet L. Yellen
Chairman Jerome H. Powell
Chairman


Speech Length by Chairman
Speech Length
10000
8000
6000
4000
2000
0
Chairman Alan Greenspan
Chairman Ben S. Bernanke
Chairman Janet L. Yellen
Chairman Jerome H. Powell
Chairman

## Research Design

To identify topics overall and unique to each speaker, the dataset is vectorized using TFIDF, Doc2Vec and Word2Vec as well. K-means clustering is applied on both the TFIDF and Doc2Vec matrices to get an initial idea at how the topics are shaping out. LDA topic modeling will also refine our results of topics and term similarity. This was performed using the TFIDF matrix as well as a bag of words.

An initial ontology of topics will be created to reflect the overall scope of the Federal Reserve. In order to clean up the vocabulary and accentuate the topic clusters, terms from the TFIDF clusters, output vectors *similar_by_word* function for Word2Vec and domain knowledge aided in creating equivalence classes.

I used Doc2Vec to visualize speech similarity in a 2D space. Using Doc2Vec to reduce the dimensionality of the speech to a 20-dimensional vector, I then used Multi-Dimensional Scaling to plot each speech in two dimensions.

LDA (Latent Dirichlet Allocation) was used to group speeches together based on probability of a word's frequency in a certain topic vs the frequency in all topics. A web-based interactive visualization of the topics created from the LDA is also generated to corroborate our initial results and understand which topics are most closely related.

Sentiment analysis is run on the original text dataset to test the idea that the Federal Reserve meticulously chooses its verbiage in order to abate market fluctuations and overreactions from analysts. These metrics will also be compared to the fluctuations of the Federal interest, unemployment and inflation rates to observe if there is any correlation.

## RESULTS

### WordCloud:


### Primary Clusters:

From the first iteration of clusters and topics generated, the overarching themes covered by each speaker were largely:

- Monetary Policy: Inflation rates, Unemployment in the country, interest rates
- Bank Regulation, supervision and relations with other central banks of the world such as the ECB (European Central Bank)
- Rise in technological advancements in the financial world
- International Trade and Globalization
- Commodities such as Oil and Natural Gas; Energy

- Housing and Mortgages
- Community Investment and Entrepreneurship – Rural and Minority Communities

While these were the central topics covered that spanned 24 years of speeches, it is worthy to note the unique ones for each speaker, influenced either by certain events or preference.

**Alan Greenspan**

From K-means clusters and LDA outputs, these were prominent topics and terms:

| **Topic** | **TFIDF K-means / LDA Terms** | | | | |
|---|---|---|---|---|---|
| Intellectual Property | China | Protection | Property Right | Reagan | Technology |
| Currency | New Note | Counterfeit | Legal Tender | Recirculated | Older Note |
| Century date change | CDC | breakdown | readiness | | |

To add more context to these topics, Greenspan widely lauded Former President Reagan's effort bolster our industrial competitiveness against nations such as Japan. A blue-ribbon commission appointed by President Reagan found that "inadequate protection of intellectual property rights" was "among the reasons for a decline in the U.S. comparative advantage in high technology." In an effort to improve the current technological space, Greenspan brought up Reagan's efforts as inspiration – how his support of the patent system and his policies that offered pro-growth tax cuts and reduced needless regulation accelerated this.

The $20 note is the most frequently counterfeited note in the United States. So in 1998 and subsequently in 2003, the $20 note primarily went through significant redesigns critical to weeding out the problem of counterfeiting US currency.

As the nation was embracing the a new millennium, it was feared that such a misreading would lead to software and hardware failures in computers used in such important areas as banking, utilities systems, government records, and so on, with the potential for widespread chaos on and following January 1, 2000 – the Y2K phenomenon. Market participants resolved to limit their exposure to Y2K-related risks by cutting back normal trading activities. The Federal Reserve foresaw that this strategy could lead to serious liquidity problems in key financing markets, and had to launch programs for backup funding.

**Ben Bernanke**

| Topic | TFIDF K-means / LDA Terms | | | | |
|---|---|---|---|---|---|
| Minority | TALF | procurement | depository | diverse | CDFI |
| Housing | Foreclosure | Delinquency | mortgage | recession | subprime |
| Regulation | Basel III | Hedge fund | bank | Clearinghouses | |
| Happiness | Wellbeing | Satisfaction | | | |

Bernanke oversaw one of the most devastating financial crises in recent memory – the 2008 Housing Crisis. In order to jumpstart the economy and restore faith with the people, the Federal Reserve created a program in November 2008 called TALF or Term Asset-Backed Securities Loan Facility to boost consumer spending. It primarily benefited students and small businesses

owned by minority communities and women with credit revival for loans. Furthermore, it also promoted to general welfare of MDIs (Minority depository institutions which are essentially banks and credit unions that are owned or operated by minorities.

Along with this came swift changes to bank regulation and supervision policies to prevent a similar disaster. The Basel Accords were a framework for financial regulation and the 3rd installment was ratified as a result of the housing crisis which made changes to the behavior of banks and hedge funds.

He was also one of the few speakers to consistently bring up the economics of happiness and human wellbeing.

**Janet Yellen**

| Topic | TFIDF K-means/ LDA Terms | | | | |
|---|---|---|---|---|---|
| Women | participation | Woman history | diversity | diverse | CDFI |
| Family | household | SCF | asset | inequality | |

In a significant portion of her speeches, Janet Yellen, the first female to Chair the Federal Reserve, also preached an increased in female participation and diversity in the financial world as well as the importance of Asset Building for Low and Middle Income households as the country continued to recover from the housing crisis.

**Jerome Powell**

| Topic | TFIDF K-means / LDA Terms | | | | |
|---|---|---|---|---|---|
| Financial Stability | Stress Test | Sustainability | Recovery | diverse | CDFI |
| Coronavirus | covid | virus | pandemic | | |
| Middle Class | Inequality | Working age | | | |

Jerome Powell, the current Chairman, highlights the necessity of economic stress tests to continuously diagnose how resilient our economy can be in the face of a crisis, notably with the onset of a pandemic that has halted practically the entire world.

**The entire breakdown of the topics and their relation to each other can be seen in the ontology attached in the Appendix section at the end.**

**Programs**

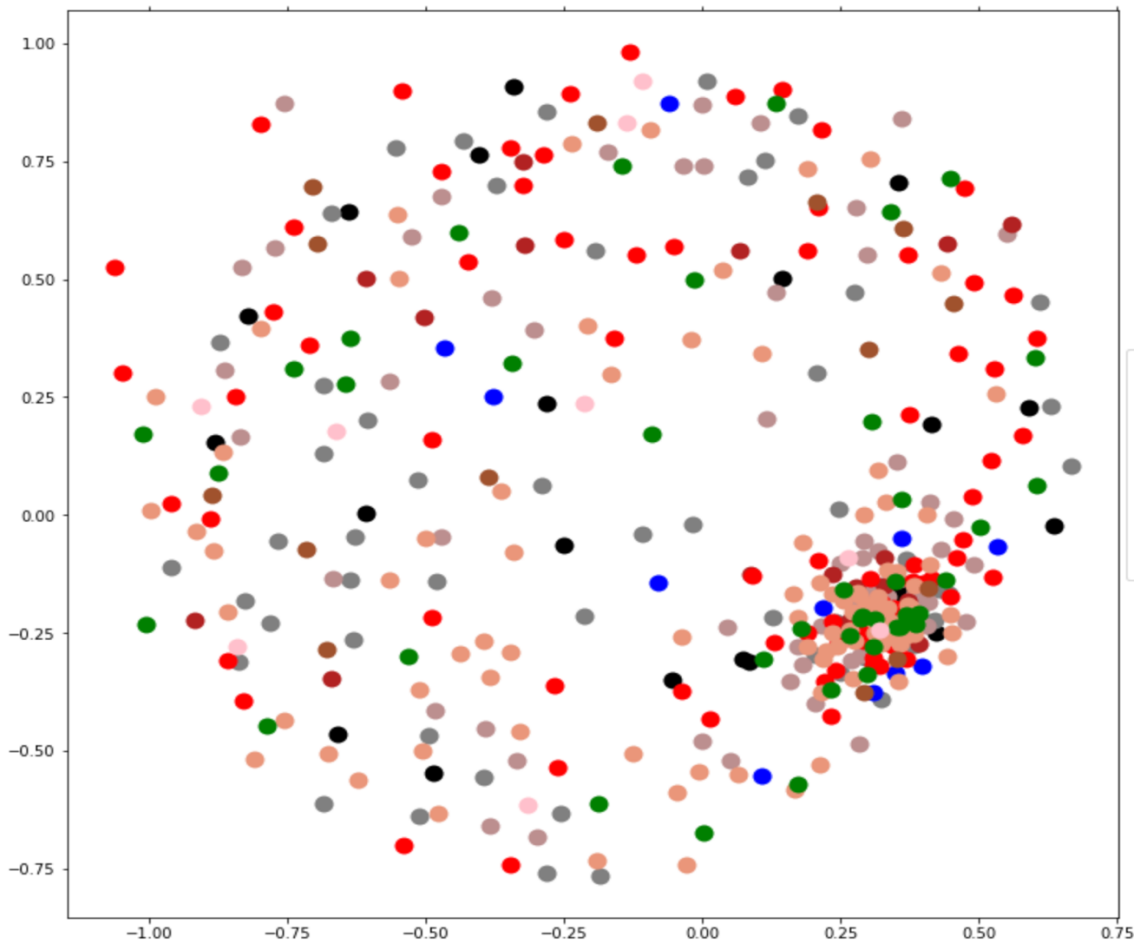
These are the programs and committees that the Federal Reserve oversaw or assisted with:

- FOMC - Federal Open Market Committee: a committee within the Federal Reserve System responsible for overseeing the nation's interest rates and growth
- CRA - Community Reinvestment Act: A federal law designed to encourage commercial banks to help meet the needs of borrowers in low-income and rural communities
- TALF - Term Asset-Backed Securities Loan Facility : Credit Revival for loan seekers in minority communities
- CDFI - Community development financial institutions: Promotes economic revitalization in distressed communities throughout the United States by providing financial assistance

- GSE - Government Sponsored Enterprise: Intended function is to enhance the flow of credit to targeted sectors of the economy: agriculture, home finance and education
- FSOC – Financial Stability Oversight Council: Identify risks and respond to incoming threats to the economy, while also promoting market discipline

**Doc2Vec Modeling:**

Using Multidimensional Scaling to plot the K-means clusters of the document vectors based on the distance matrix:

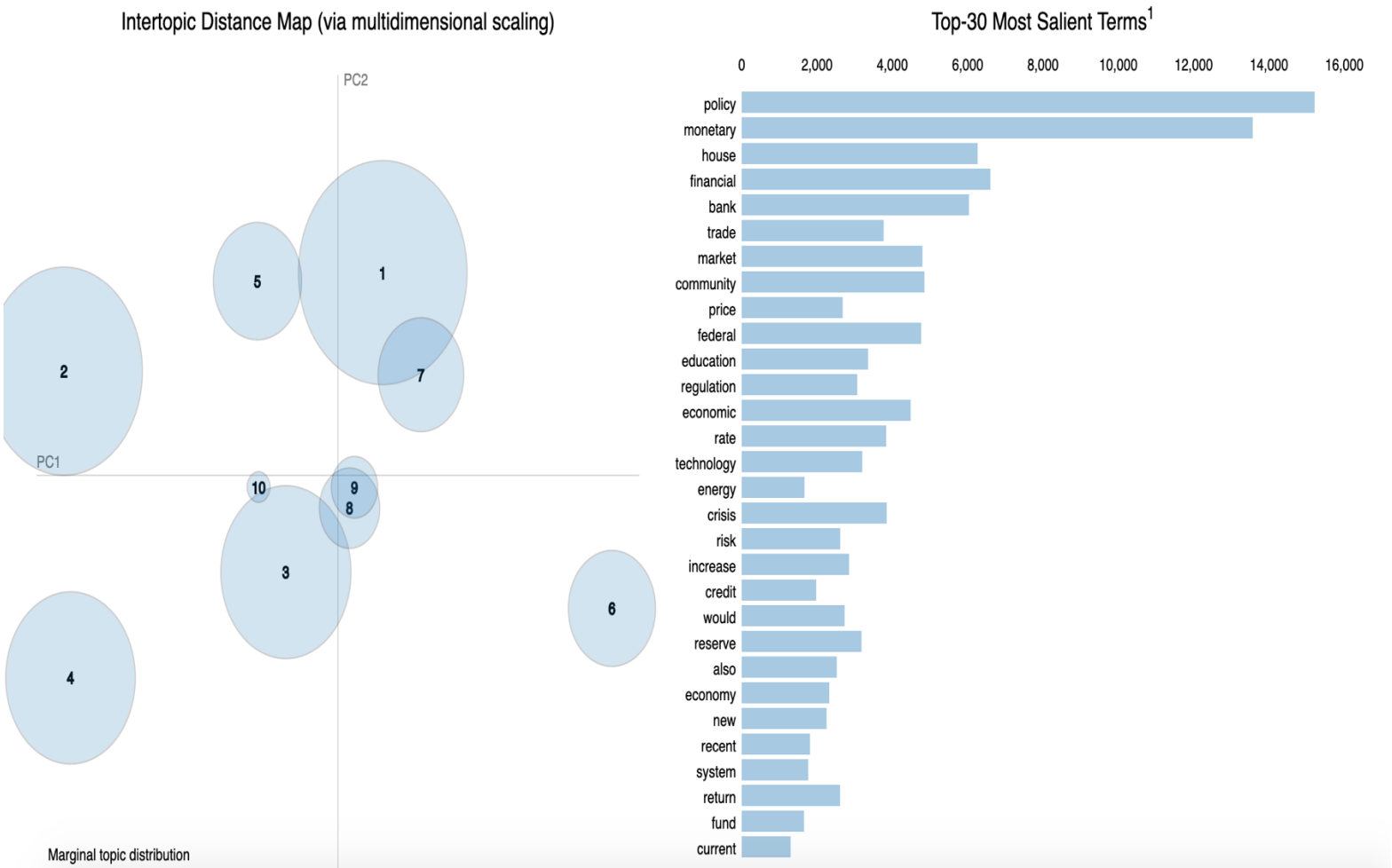



**Cross-validation Score on Random Forest Classifier: 0.907**

This shows that the K-means and RF models aren't that off in classifying the documents in such a manner which is slightly concerning. I anticipated there to be more vivid clusters based on topic. Two explanations can justify this: either the data and model need to be further improved in its the equivalence classes and hyperparameters or perhaps the verbiage of the Fed is too common regardless of the topic. Basic economic vocabulary may be so prevalent in each speech, regardless of topic, that it perhaps is difficult to distinguish them. Furthermore, some of these speeches were titled 'Introductory Remarks' or 'Brief Remarks' which had very generic material so that probably impacted the model's performance as well.

## LDA Modeling

The process was repeated using an LDA Model of bag of words corpus:



The size of the bubble represents how prevalent the top 30 words are for that topic

**Perplexity Score: -7.32**

**Coherence Score: 0.41**

This visualization and the subsequent metrics are more promising than the previous experiment. There is a decent balance of distance and overlapping between the clusters. However, this size of each cluster are fairly and the top 30 salient terms are fairly generic which shows that there may not be enough to distinguish between topics and use the model for further classification tasks.

## SENTIMENT ANALYSIS

Here are the results of sentiment analysis using both *VADER's SentimentIntensityAnalyzer* and TextBlob:

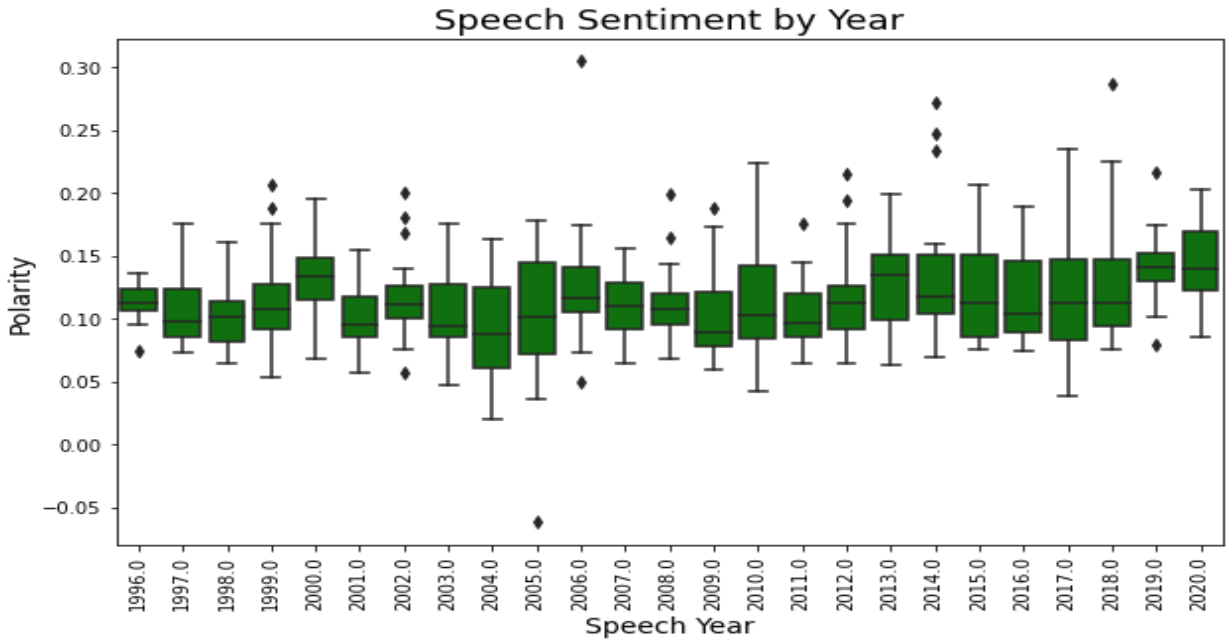
VADER:

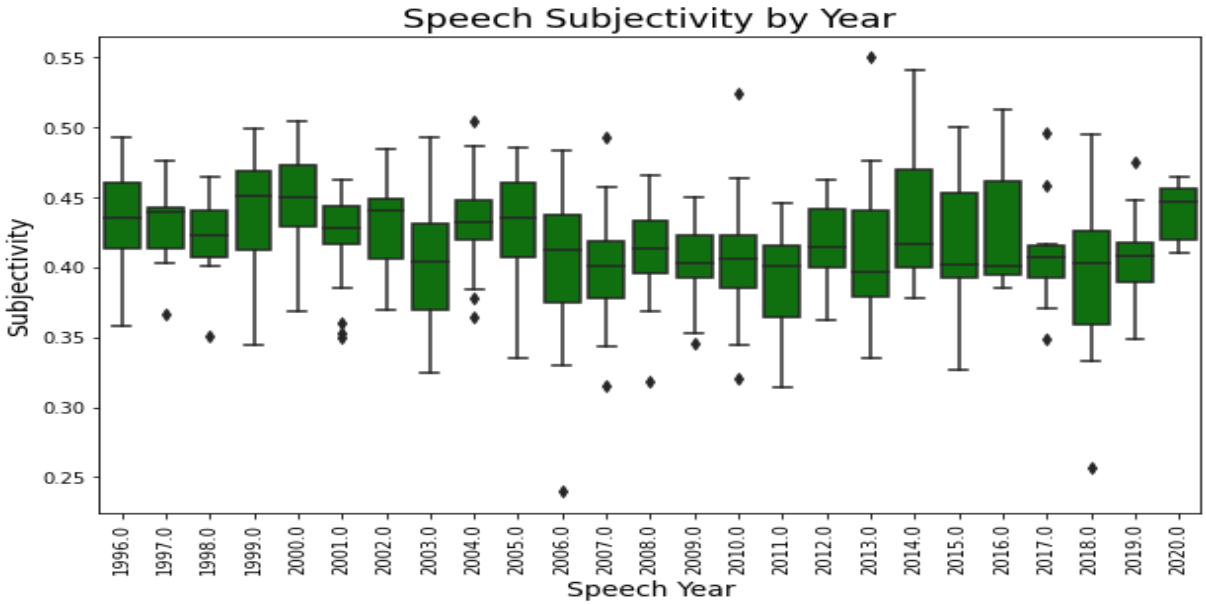
| Speaker | Positive | Neutral | Negative | Compound |
|---|---|---|---|---|
| Greenspan | 0.147 | 0.781 | 0.072 | 1.0 |
| Bernanke | 0.145 | 0.784 | 0.07 | 1.0 |
| Yellen | 0.129 | 0.81 | 0.061 | 1.0 |
| Powell | 0.134 | 0.79 | 0.076 | 1.0 |

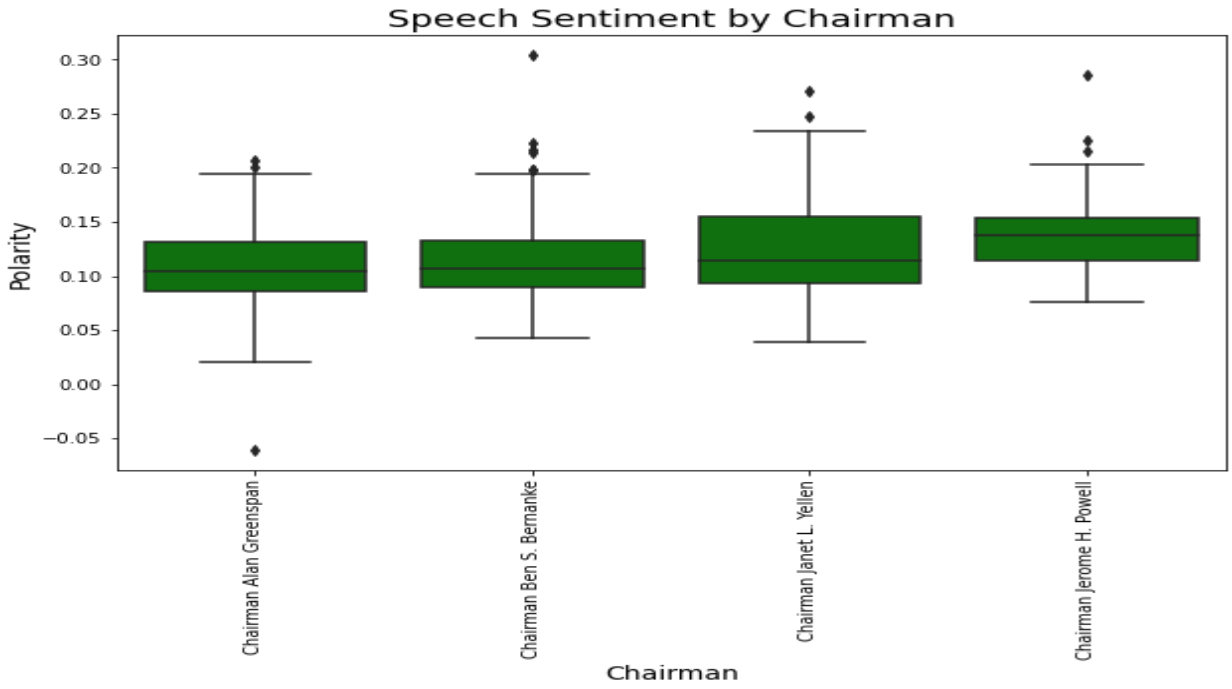
Examples of positive words: growth, increase, improve, sustainability

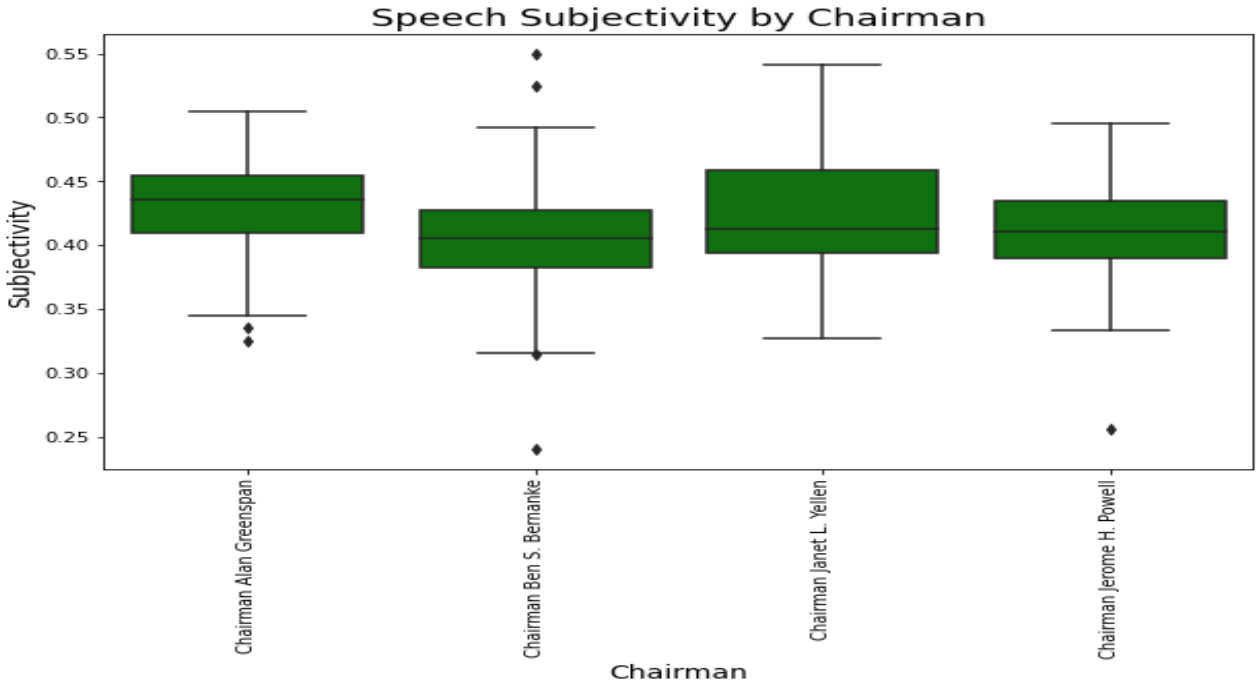
Examples of negative words: decline, crisis, turmoil

TextBlob:


Speech Sentiment by Year
Polarity
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
1996.0
1997.0
1998.0
1999.0
2000.0
2001.0
2002.0
2003.0
2004.0
2005.0
2006.0
2007.0
2008.0
2009.0
2010.0
2011.0
2012.0
2013.0
2014.0
2015.0
2016.0
2017.0
2018.0
2019.0
2020.0
Speech Year


Speech Subjectivity by Year
Subjectivity
0.55
0.50
0.45
0.40
0.35
0.30
0.25
1996.0
1997.0
1998.0
1999.0
2000.0
2001.0
2002.0
2003.0
2004.0
2005.0
2006.0
2007.0
2008.0
2009.0
2010.0
2011.0
2012.0
2013.0
2014.0
2015.0
2016.0
2017.0
2018.0
2019.0
2020.0
Speech Year


Speech Sentiment by Chairman
Polarity
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
Chairman Alan Greenspan
Chairman Ben S. Bernanke
Chairman Janet L. Yellen
Chairman Jerome H. Powell
Chairman


Speech Subjectivity by Chairman
Subjectivity
0.55
0.50
0.45
0.40
0.35
0.30
0.25
Chairman Alan Greenspan
Chairman Ben S. Bernanke
Chairman Janet L. Yellen
Chairman Jerome H. Powell
Chairman

The polarity, which ranges from -1 to 1, is right around 0.1 - 0.2 for each speaker, save for a few exceptions. Overall, this fits the general tone that the Federal Reserve wants to achieve and they tend to be more positive than negative in verbiage as well. In terms of subjectivity scores, which range from 0 to 1, I was a little surprised the speeches were averaging around 0.4 – 0.45, walking a tight line of somewhat opiniated and somewhat factual. I figured given their whole idea of not trying to impact market trends, the scores would be much lower. However, a case can be made that simply rattling out facts is not an effective way to address economists and the rest of the country when painting a picture of our economy. Leadership requires a certain level of subjectivity.

Comparing it to specific years of crisis – 2001, 2008, 2020 – there isn't any trend that can be discerned. Even comparing it to constantly fluctuating GDP growth, unemployment, interest, and fed funds rates year by year would be moot since there is hardly any fluctuations in speech polarity. Conclusively, the Federal Reserve has done a good job in maintaining its tone.

## CONCLUSION

This paper was a fairly successful result in analyzing the speeches made by Chairman. While it was effective in identifying topics of interest over the years that matched well with the overall timeline and narrative in our economy and nation, the modeling results weren't as promising. This dissuaded me from using it more extensively as a classifier in topic prediction, which would have been the next step. Using domain knowledge and research was an incredible resource in refining the topics. Sentiment analysis was reaffirming in understanding 'Fedspeak'.

This paper is only the beginning of a continuous effort to study the Federal Reserve Board and classify its interests.

## POTENTIAL IMPROVEMENTS AND FUTURE WORK

- Consider removing more generic economic terms from the dataset to improve model performance
- Further consolidating similar words (supervisory, supervision, supervisor) using either stemming and/or refining part-of-speech (POS) tagging
- Using pre-trained word embeddings like GloVe (Global Vectors for Word Representation) to possibly improve doc2vec model
- Increasing dataset size by including more Fed speeches and even published communications from the Fed, such as press releases and testimonies.
- Create a recommendation system for speeches that are similar to a speech of interest. This may need to utilize neural networks (RNN) as well. This would also require more time in labeling the speeches manually in order to improve classification.
- Using Neural Networks to learn how to 'FedSpeak'

## REFERENCES

[1] - November 1, 2017 Author: Sentieo. "Assessing Fed Chair Hopefuls With NLP Analysis Of Past Speeches." Sentieo, October 15, 2020. https://sentieo.com/assessing-fed-chair-hopefuls-with-nlp-analysis-of-past-speeches/.

[2] - Ahuja, Sameer. "Can Neural Networks Learn to Fed-Speak?" Medium. Towards Data Science, July 22, 2019. https://towardsdatascience.com/can-neural-networks-learn-to-fed-speak-27cd8c546963.

- Doh, Taeyoung, Dongho Song, and Shu-Kuei Yang. "Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements." *The Federal Reserve Bank of Kansas City Research Working Papers*, 2020. https://doi.org/10.18651/rwp2020-14.

- "Board of Governors of the Federal Reserve System." Federal Reserve Board . Accessed November 20, 2020. https://www.federalreserve.gov/.

- Amadeo, Kimberly. "Compare Today's Unemployment with the Past." The Balance, April 3, 2020. https://www.thebalance.com/unemployment-rate-by-year-3305506.

**APPENDIX:**

**Ontology:**