NLP Analysis of Inaugural and State of the Union Discourse

Kevin Huang, Shuyang (Simon) Bian, Roberto Franzosi[1]

[1] Emory College of Arts and Sciences, Emory University, Atlanta, GA;

Introduction

Both US Presidential Inaugural addresses and State of the Union speeches have been heavily analyzed over the years, using both qualitative and computational methods. Generally, these analyses follow a common framework by which researchers attempt to draw some kind of insight into the political atmosphere, presidential language and goals of politicians, or underlying historical trends by using various tools in NLP or other methodologies. We apply tools from the NLP Suite to examine similar questions surrounding the two corpora.

Methods and Materials

I. We collected both the Inaugural and State of the Union corpora from the UCSB American Presidency Project (The American Presidency Project, n.d.) and divided them into **three sub-corpora: war/peacetime, economic growth/decline, and party affiliation.** The inaugural corpus consists of speeches from 1789 to 2020, totaling 63 speeches. The State of the Union corpus contains 242 total speeches from 1790 to 2023.
   I. The data for wartime and peacetime is taken from a US Congressional Research Service analysis (Daggett, Stephen. (2010). Costs of Major U.S. Wars. 9).
   II. The division based on economic status is founded on the observation of three variables: inflation, unemployment, and recession dates, and taken from Federal Reserve FRED NBER data.
   III. Finally, we separate presidents into different sub-corpora based on their political party. For most analyses, we focus on the Democratic and Republican parties.
II. All analyses are run through the NLP Suite ((https://github.com/NLP-Suite/NLP-Suite/wiki). The purpose of the NLP Suite is to provide humanists and social scientists a wide range of computational tools for the analysis and visualization of smaller datasets, the more typical datasets humanists and social scientists use.

Results

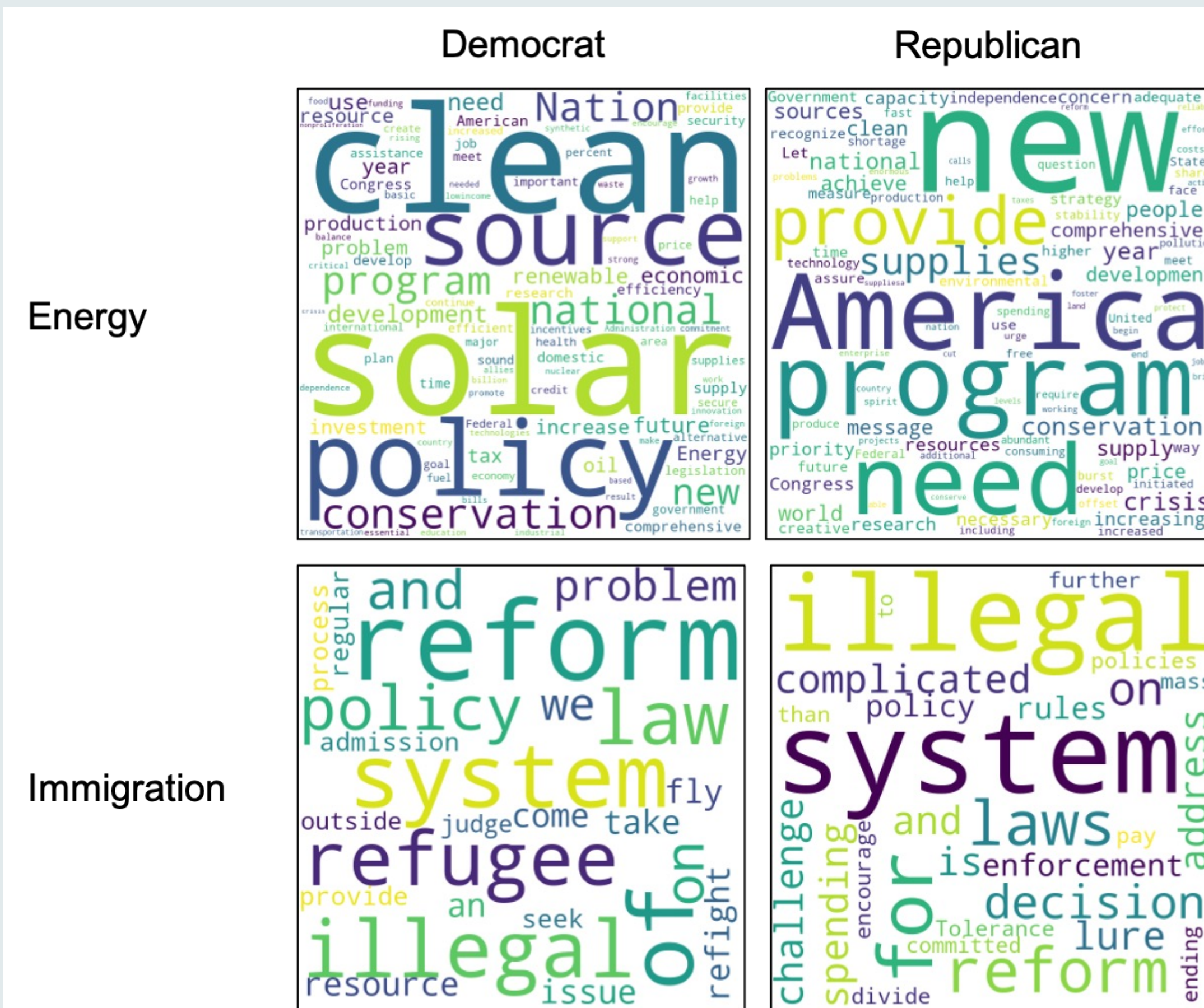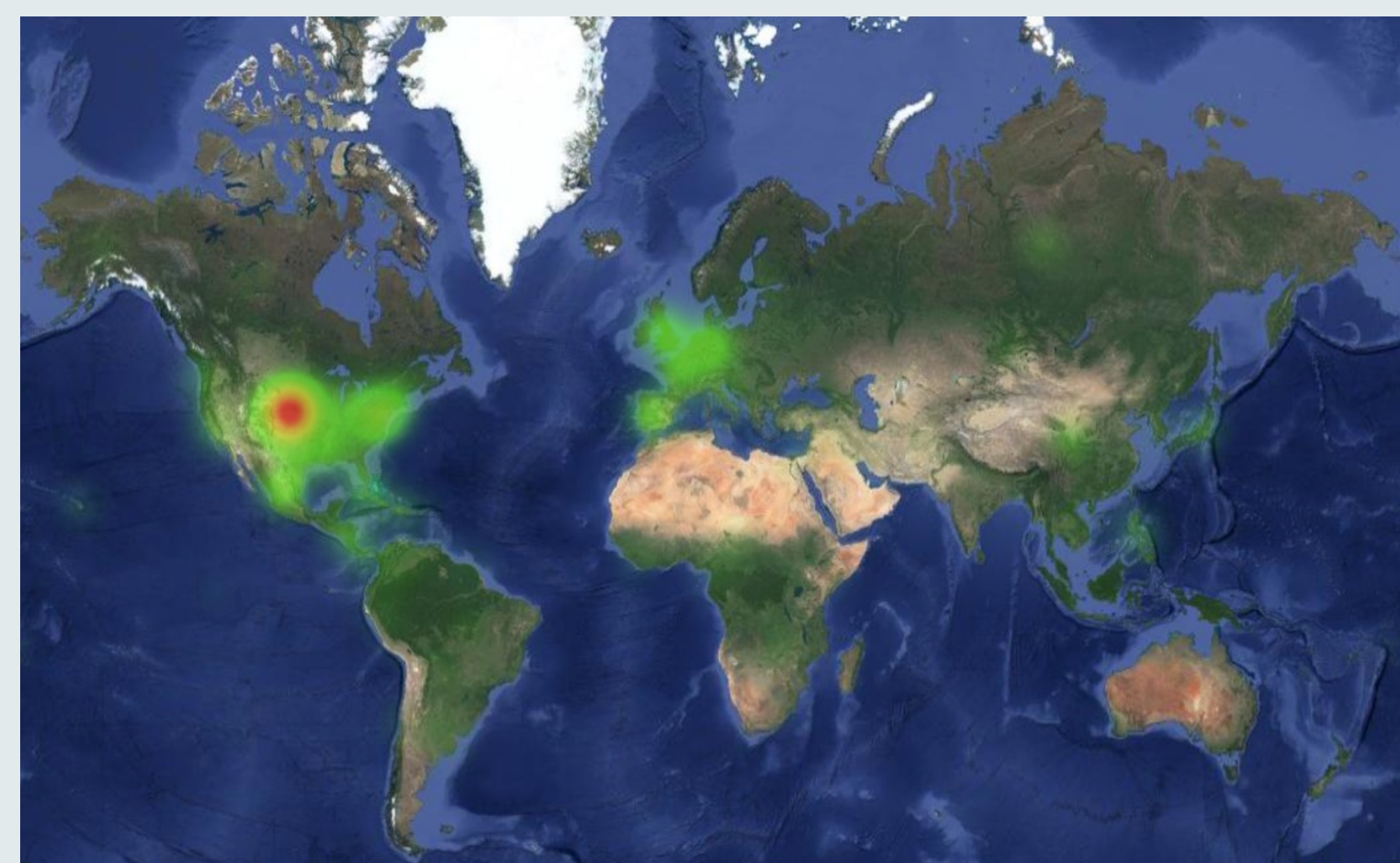**Figure 1: Selected CoNLL table word clouds based on political party from 1964 onward**

Figure 2: Stanford CoreNLP GIS Heatmap of Inaugural Adresses

Figure 3: Stanford CoreNLP GIS Heatmap of SOTU speeches

Figure 4: Frequency of "God" usage in SOTU speeches by President
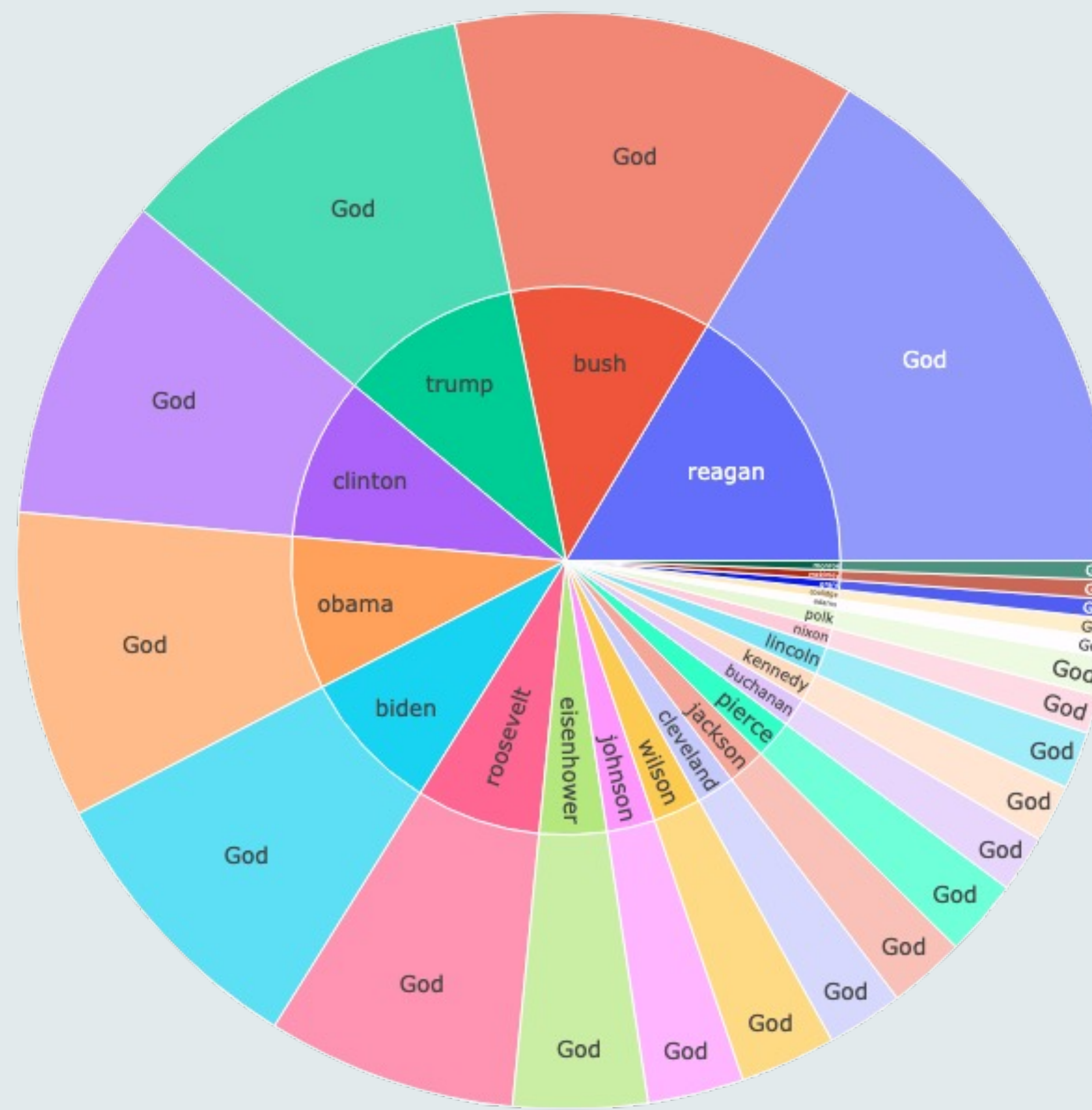
Figure 5: Sentiment scores of SOTU speeches vs. U-3 unemployment rate and inflation rate at time of speech
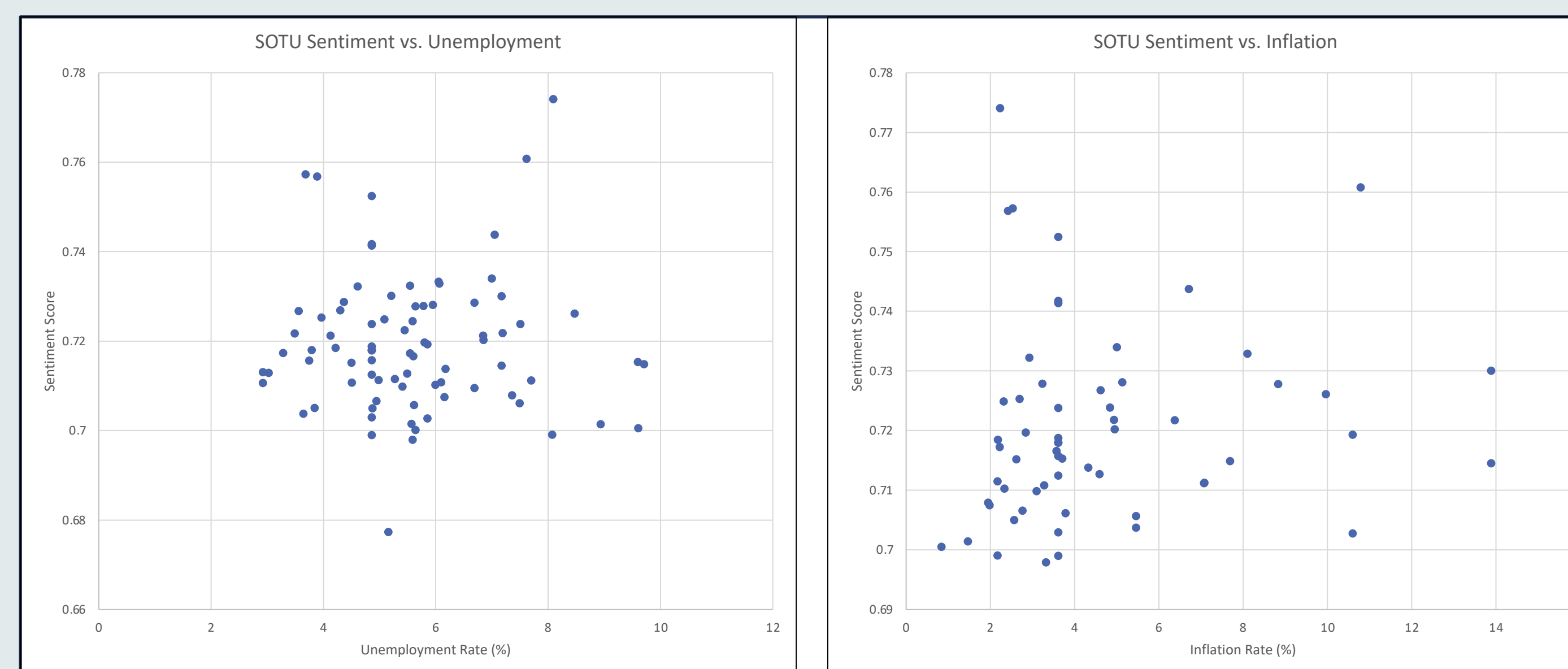
Figure 6: Average Yngve complexty score by SOTU speech sorted chronologically
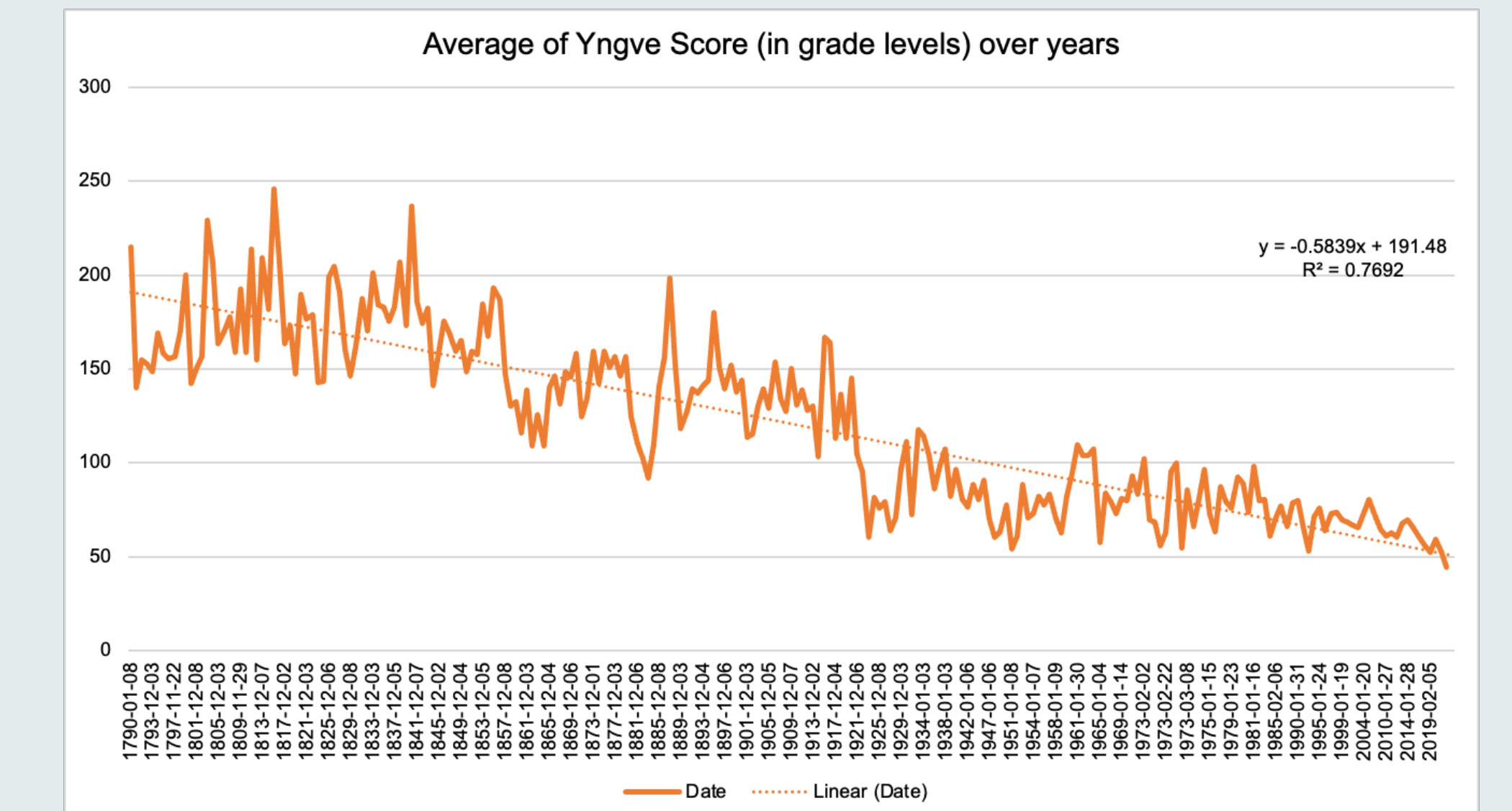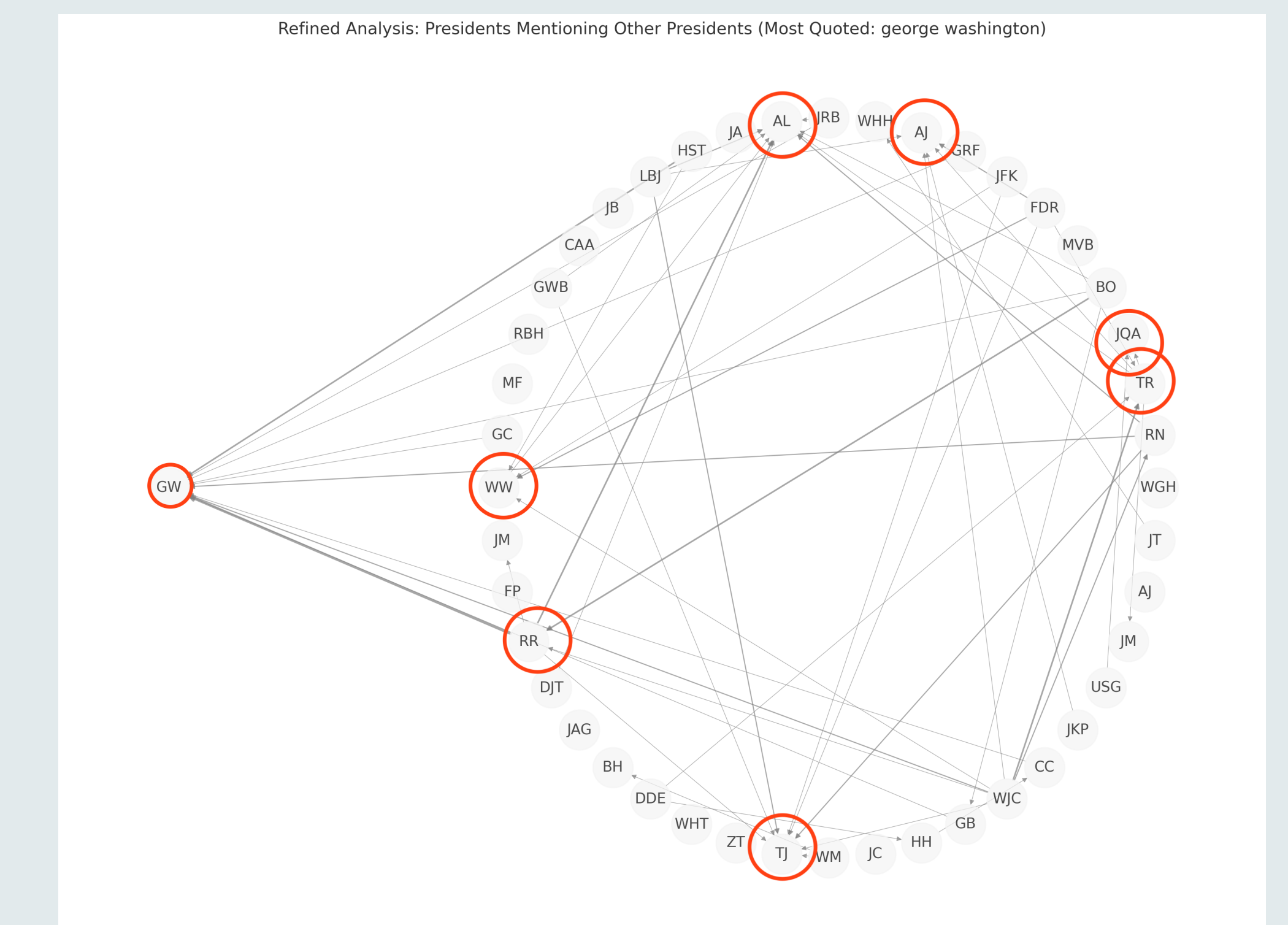

Figure 7: Network Analysis of US Presidents through name references in both SOTU and Inaugural speech corpora

"JRB": "joseph r biden", "AL": "abraham lincoln", "JA": "john adams", "HST": "harry s truman", "LBJ": "lyndon b johnson", "JB": "james buchanan", "CAA": "chester a arthur", "GWB": "george w bush", "RBH": "rutherford b hayes", "MF": "millard fillmore", "GC": "grover cleveland", "WW": "woodrow wilson", "JM": "james monroe", "FP": "franklin pierce", "RR": "ronald reagan", "DJT": "donald j trump", "JAG": "james a garfield", "BH": "benjamin harrison", "DDE": "dwight d eisenhower", "WHT": "william howard taft", "ZT": "zachary taylor", "TJ": "thomas jefferson", "WM": "william mckinley", "JC": "jimmy carter", "HH": "herbert hoover", "GB": "george bush", "WJC": "william j clinton", "GW": "george washington", "CC": "calvin coolidge", "JKP": "james k polk", "USG": "ulysses s grant", "AJ": "andrew jackson", "JT": "john tyler", "WGH": "warren g harding", "RN": "richard nixon", "TR": "theodore roosevelt", "JQA": "john quincy adams", "BO": "barack obama", "MVB": "martin van buren", "FDR": "franklin d roosevelt", "JFK": "john f kennedy", "GRF": "gerald r ford", "WHH": "william henry harrison"

Conclusion

Usage of the NLP Suite yielded novel findings in CoNLL table analysis, sentiment analysis, GIS mapping, sunburst graphical representation, and readability. For example, State of the Union speeches have decreased in complexity score (Yngve) over time. On the topic of energy, Democrats were more focused on "clean" and "solar," whereas Republicans mentioned fewer renewables, speaking more about "need," "America," and "new." On the issue of immigration, Democrats had higher mentions of "reform" than did Republicans. Reagan, Trump, and Bush were the presidents with the highest occurrence of "God" in their State of the Union speeches. George Washington, Abraham Lincoln, Thomas Jefferson, Andrew Jackson, Woodrow Wilson, Theodore Roosevelt, John Quincy Adams, and Ronald Reagan were the most referenced presidents throughout both corpora, in that order. Sentiment did not appear to correlate with U-3 unemployment or inflation rates. GIS mapping showed that the State of the Union corpus was much more international than the Inaugural, meaning that these speeches were more likely to mention foreign nations/locations.

Future Plans

Further analysis is planned into various avenues of the graphs above. For example, GIS heatmaps may be graphed on a time series to identify trends based on different periods of time, wars, etc. Especially of interest is the analysis of sentiment based on real world factors. Preliminary results above show that there does not appear to be a correlation between sentiment and traditional economic indicators. However, other variables may prove to be correlated as they are tested. Overall, these and future results will allow for a more complete picture of the corpora using NLP or computational methods.

Contact
Dr. Roberto Franzosi
College of Arts and Sciences, Emory University
rfranzo@emory.edu