Democrats vs Republicans:

Text Analysis of US Presidents' Inauguration Speeches

Done by: Si Han Ang

This project sets out to answer the research question: What are the similarities and differences between democratic and republican party presidents' inauguration speeches?

For the analysis, I would be using the dataset of the US Presidential Inaugural Addresses within the NLTK toolkit which comes in text files ranging from the first president's speech till Trump's speech in 2017. I have also included an additional self-compiled dataset which provides the mapping of each president to their respective parties.

I deep-dived into the speeches to explore the similarities/differences in topics and people mentioned by the republicans and democrats.
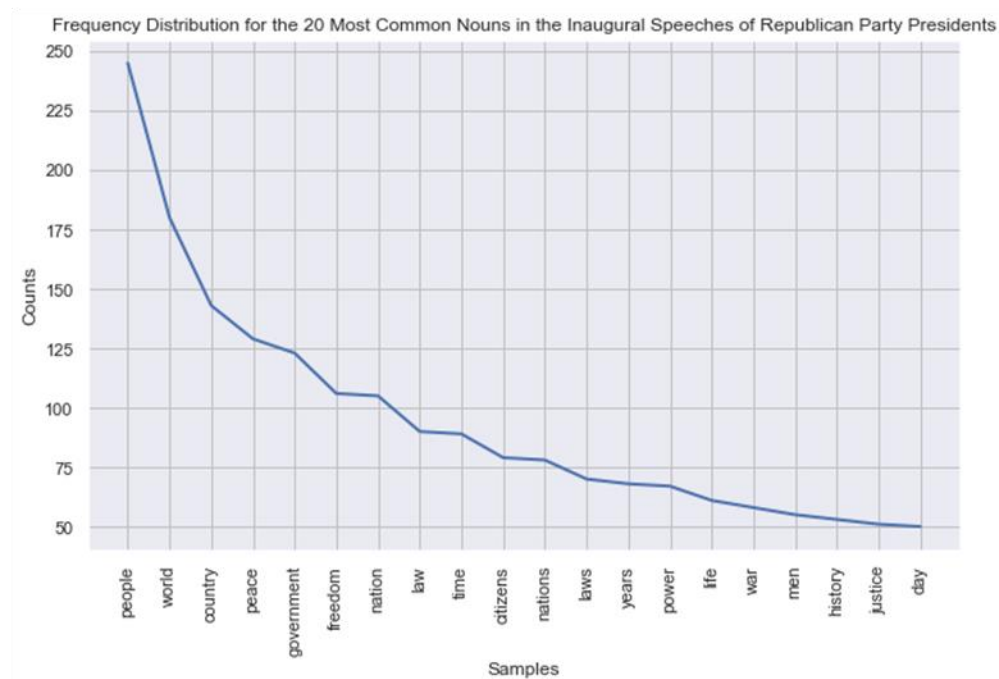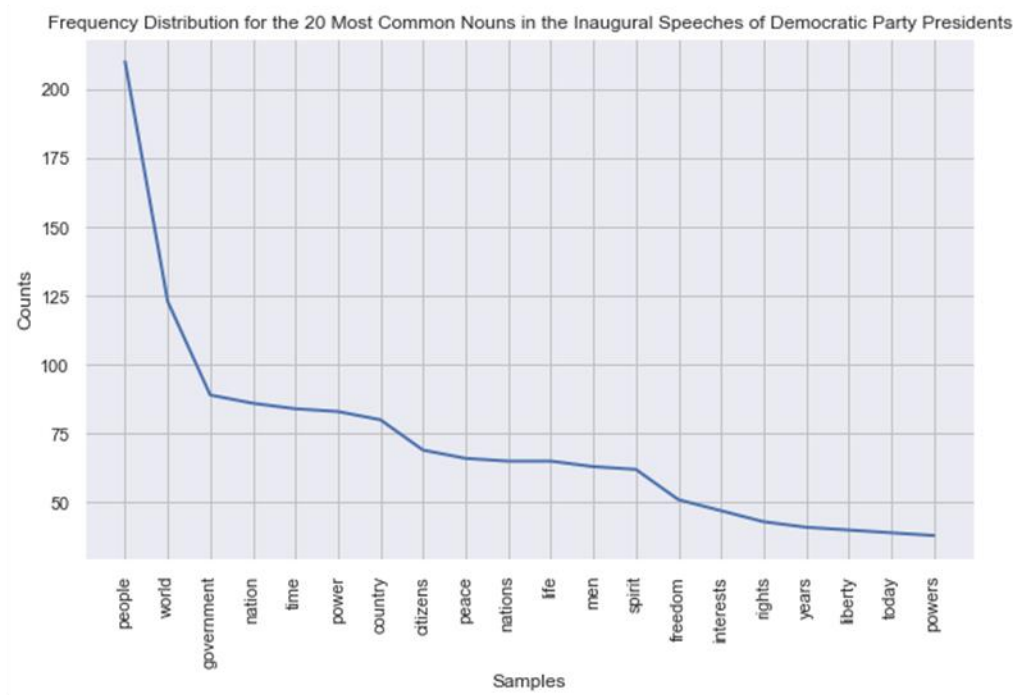
I used Named Entity Recognition (NER) to derive people mentioned by the presidents of both parties in word clouds below. Republicans and Democrats both tend to mention God and the past US presidents.

Visualisation 1: Word Clouds of People Mentioned by Presidents from Democratic and Republican Party

To explore on topics of speeches, I used the part-of-speech (POS) tagger in NTLK and obtained words tagged as nouns. I then ran the frequency distribution plots of the top 20 most common mentioned by republicans and democrats as shown below:
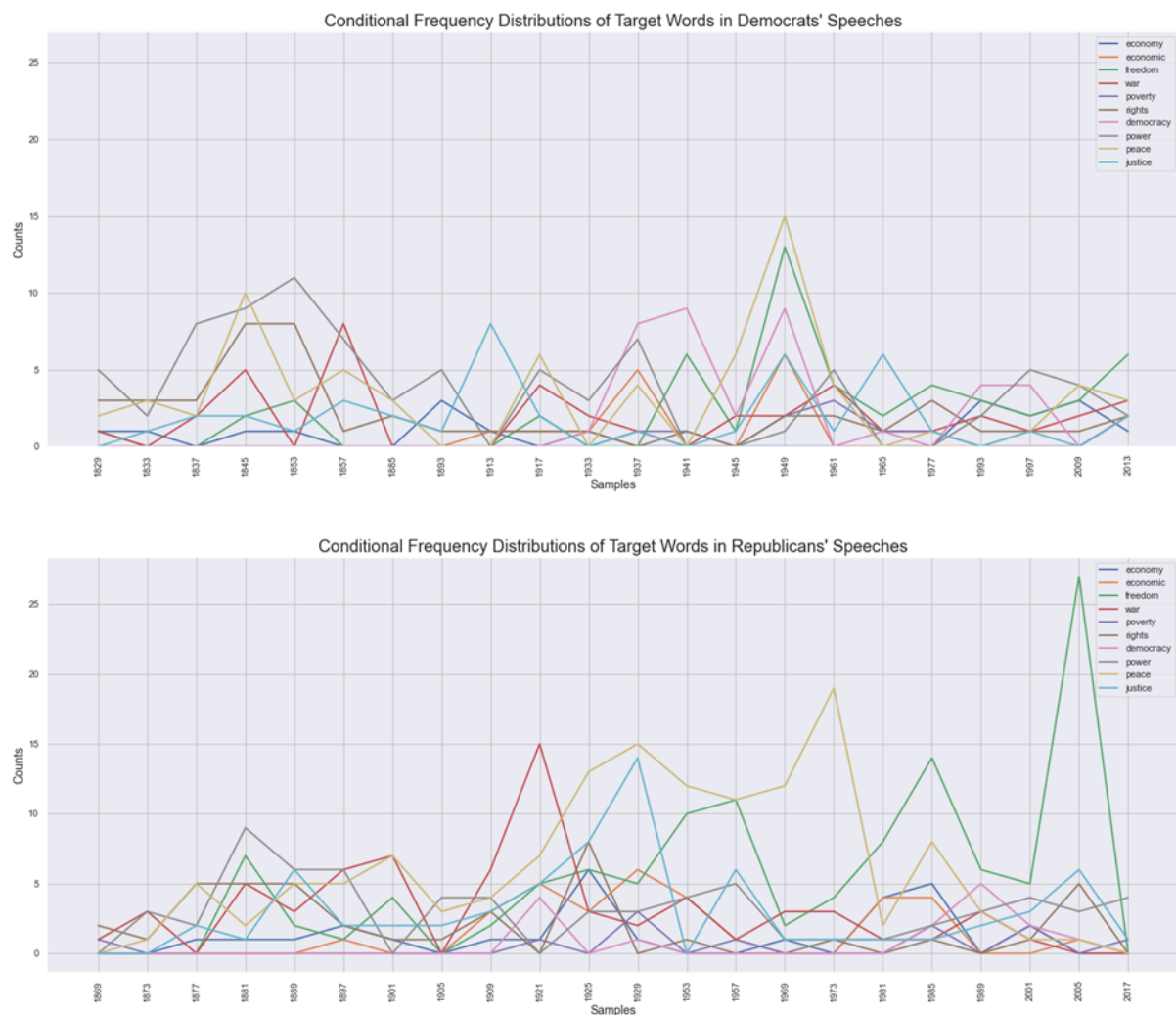
Visualisation 2: Frequency Distribution of 20 Most Common Nouns Mentioned by Presidents from Democratic and Republican Party

Frequency Distribution for the 20 Most Common Nouns in the Inaugural Speeches of Democratic Party Presidents

Frequency Distribution for the 20 Most Common Nouns in the Inaugural Speeches of Republican Party Presidents

From these distribution plots, it shows that both parties commonly talk about people, world, country/nation, peace, government etc. However, the apparent differences in ranking of frequencies of the words such as justice, power etc. points to potential differences in political stances by the parties worth investigating.

In light of this, I looked into the conditional frequency distribution of target words which could provide a sense of what each party's emphasis is over the years. First, I compiled a list of target words related to economy ('economy', 'economic'), social issues ('freedom', 'justice', 'rights', 'poverty', 'democracy') and military ('war', 'power'). Next, I removed stop words (using those obtained from NLTK, digits, punctuations and common stop words) and made sure that words are alphanumeric. Finally, I plotted these plots of both parties separately using the function from NLTK as follows.

Visualisation 3: Conditional Frequency Distribution of Target Words Mentioned by Presidents from Democratic and Republican Party over the Years

In terms of social issues, democrats mentioned more about "rights" than republicans which is in line with democratic party's focus in human rights and equal opportunities. Both parties have emphasized more on freedom in recent few decades with the republicans being more focussed on freedom than the democrats. As for the word "peace", from 1900 onwards, peace has been mentioned a lot more times in periods of uncertainty in the presidential speeches from both parties, for instance during the speech of 1917 during World War 1 (1914 - 1918), 1920s where turf wars among rival gangs were prevalent in the USA and 1929 stock market crash, 1945 through 1973, when second world war happened (1942-1945), Korean War (1950-1953) and Vietnam War (1955-1975). In periods of relative peace from 1980s onwards, presidents from both parties made fewer mention of this word and

interestingly, the republicans made a lot fewer mentions of peace compared to their predecessors while Obama's speeches (2009, 2013) led to an uptick in mention of peace by the democrats.

In terms of economic issues, presidential inauguration speeches from both parties tended not to mention much about economy-related matters.

In terms of military related words, there were more mentions of war by democrats than republicans in the last 3 decades which could be deliberate since the recent wars US engaged in happened during the presidential terms of republicans and this controversial topic implied by the word "war" could have led republicans to stay clear of this.

To wrap up, I also used topic modelling using Latent Dirichlet Allocation (LDA) from gensim library after lemmatization. The results from both parties are in line with the previous findings on social issues and military issues.

Visualisation 4: LDA Model Output (Topics) from Inauguration Speeches of Presidents from Democratic and Republican Party

```
[(0,
 '0.016*"nation" + 0.009*"people" + 0.009*"life" + 0.009*"great" + '
 '0.009*"man" + 0.008*"spirit" + 0.006*"government" + 0.006*"thing" + '
 '0.005*"year" + 0.005*"purpose"'),
(1,
 '0.017*"people" + 0.013*"government" + 0.008*"public" + 0.008*"citizen" + '
 '0.005*"american" + 0.005*"interest" + 0.005*"service" + '
 '0.005*"constitution" + 0.004*"law" + 0.004*"national"'),
(2,
 '0.014*"government" + 0.013*"state" + 0.010*"people" + 0.009*"power" + '
 '0.008*"country" + 0.007*"nation" + 0.007*"union" + 0.006*"great" + '
 '0.005*"constitution" + 0.005*"duty"'),
(3,
 '0.001*"people" + 0.001*"government" + 0.001*"nation" + 0.001*"country" + '
 '0.001*"state" + 0.000*"great" + 0.000*"power" + 0.000*"time" + 0.000*"man" '
 '+ 0.000*"citizen"'),
(4,
 '0.014*"nation" + 0.012*"people" + 0.010*"america" + 0.008*"time" + '
 '0.007*"man" + 0.007*"today" + 0.007*"work" + 0.007*"government" + '
 '0.006*"great" + 0.006*"change"')]
```

LDA Results for Presidential Speeches from Democratic Party

```
[(0,
 '0.001*"government" + 0.001*"people" + 0.000*"great" + 0.000*"nation" + '
 '0.000*"country" + 0.000*"law" + 0.000*"state" + 0.000*"america" + '
 '0.000*"citizen" + 0.000*"work"'),
(1,
 '0.012*"country" + 0.011*"government" + 0.010*"great" + 0.009*"nation" + '
 '0.009*"people" + 0.006*"party" + 0.005*"peace" + 0.005*"state" + '
 '0.005*"law" + 0.004*"political"'),
(2,
 '0.012*"people" + 0.011*"nation" + 0.011*"america" + 0.009*"country" + '
 '0.009*"freedom" + 0.008*"great" + 0.007*"government" + 0.006*"time" + '
 '0.006*"citizen" + 0.006*"law"'),
(3,
 '0.011*"government" + 0.010*"nation" + 0.008*"great" + '
 '0.006*"america" + 0.006*"peace" + 0.006*"law" + 0.006*"man" + 0.005*"time" '
 '+ 0.005*"state"'),
(4,
 '0.013*"government" + 0.013*"people" + 0.009*"law" + 0.008*"state" + '
 '0.007*"nation" + 0.006*"great" + 0.006*"peace" + 0.006*"public" + '
 '0.004*"american" + 0.004*"progress"')]
```

LDA Results for Presidential Speeches from Republican Party

(600 Words)

References

1. 262588213843476. "NLTK's List of English Stopwords." *Gist*, https://gist.github.com/sebleier/554280.
2. B., Sowmya V., et al. *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly, 2020.
3. Bird, Steven, et al. "Natural Language Processing with Python." *NLTK Book*, https://www.nltk.org/book/.
4. CR, Aravind. "Topic Modeling Using Gensim-LDA in Python." *Medium*, Analytics Vidhya, 31 July 2020, https://medium.com/analytics-vidhya/topic-modeling-using-gensim-lda-in-python-48eaa2344920.
5. "Democrat vs Republican." *Diffen*, https://www.diffen.com/difference/Democrat_vs_Republican.
6. "Doc · Spacy API Documentation." *Doc*, https://spacy.io/api/doc.
7. Fontanella, Clint. "75 Stop Words That Are Common in SEO & When You Should Use Them." *HubSpot Blog*, 25 Nov. 2020, https://blog.hubspot.com/marketing/stop-words-seo.
8. Li, Susan. "Topic Modeling and Latent Dirichlet Allocation (LDA) in Python." *Medium*, Towards Data Science, 1 June 2018, https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24.

9. "Presidents." *The White House*, The United States Government, 16 Feb. 2021, https://www.whitehouse.gov/about-the-white-house/presidents/.

10. "Trivia - Major Events / 1920s Major Events." */ 1920s Major Events*, https://www.humbleisd.net/domain/10764.