DS 2500: Intermediate Programming with Data (Fa21)
Prof. Rachlin and Prof. Park
Northeastern University

The Presidential Inaugural

Problem Description:

In the United States, when a president is elected or re-elected, the newly sworn-in president gives an inaugural address, presenting their goals and priorities to the nation. At six of the more recent inaugural ceremonies, there has been a poetry reading by poets such as Robert Frost, Maya Angelou, and most recently Amanda Gorman. In this assignment we'll use natural language processing and sentiment analysis to analyze political speeches and political thought. You may focus *either* on six historical presidential inaugural addresses of your own choosing or the six poems that have been recited at these events.

What can Data Science tell us about these texts? What do they suggest about the social, economic, and political events of their day? What do they reveal about the values and hopes of the American people? Is the tone universally positive and uplifting or do events such as war and political division cast these texts into a different light?

This assignment lays out some minimal goals and guidelines for your analysis, but you should not feel constrained to carry out this work exactly as suggested below. Be creative. Try to come up with innovative and original approaches to visualizing your insights. Over the next couple of weeks our collective assignment is to discover what we can about this data.

Some Recommended Data Sources:

The American Presidency Project: https://www.presidency.ucsb.edu/documents

Poems at Presidential Inaugurals: https://poets.org/inaugural-poems-history

Twitter sentiment analysis tutorial: https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/tree/master/data/opinion-lexicon-English

AFINN: http://www2.imm.dtu.dk/pubdb/pubs/6010-full.html

Another source of words with sentiment score:
https://github.com/hitesh915/sentimentstrength/blob/master/wordwithStrength.txt

Broad Goals and Possibilities:

1. If you are choosing to focus on presidential inaugural addresses, you should focus on at least six presidents of your choosing. We recommend choosing a variety of both Democrats and Republicans and you may want to avoid some of the very early speeches which tended to be very short. For example, the following list might make for a good selection.

    a. Donald Trump (R - 2017)
    b. Barack Obama (D - 2013)
    c. George W. Bush (R - 2001)
    d. Ronald Reagan (R - 1981)
    e. John F. Kennedy (D - 1961)
    f. Franklin D. Roosevelt (D - 1941)

    For those of you wishing to look at the poetry reading, there are only six such readings and they were all given at the inaugural ceremonies of Democrats (Kennedy, Clinton (2x), Obama (2x), and Biden).

2. Generate an array of word clouds, one for each speech or poem.

3. Conduct a sentiment analysis of the different texts and explore ways of presenting a visual comparison. Is sentiment correlated with political party affiliation?

4. Identify the most common non-trivial words for each text and determine whether the most commonly used words evolved over time. For example, it would be interesting to identify, for each speech, the top 100 words of length > n (for some n) and count the number of overlapping words for each pair of presidents, plotting the results as a heat map using Seaborn.

5. Develop other comparative measures of speech analysis, such as:

    a. Average sentence length
    b. Average number of words of length > n per sentence (or per 100 words)
    c. Number of unique words per 1000 words
    d. Polarity and Subjectivity
    e. Cosine similarity

    and generate some comparative visualizations. *Interpret your results!*

What to submit:

You may submit your code as one or more python files (.py) or as one or more Jupyter notebooks. If you use Jupyter notebooks, then all of your commentary, analysis, and visualization can be integrated into the notebook. Be sure that the cells have all been executed so that graders can see your outputs without having to run your code. If you submit standard python code, write up a separate report describing your methodology, analysis, and visualization as a single .pdf file.