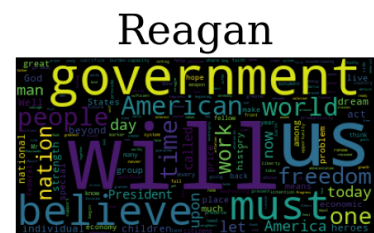
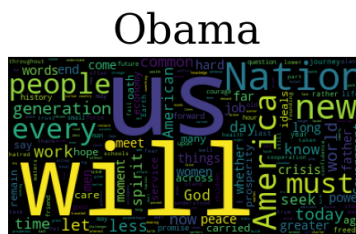
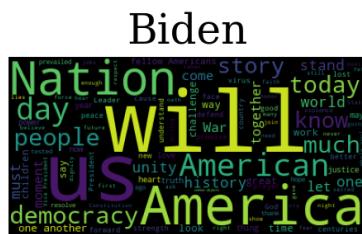


Presidential Speech Analysis

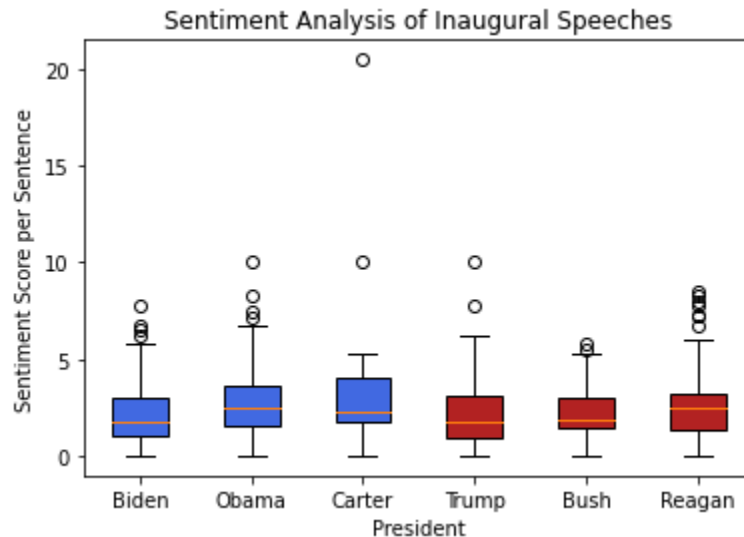
Hi, it's Daniel. For homework #3 of our DS2500 class, we conducted an in-depth analysis of 6 previous presidents' inaugural speeches to try and find any recognizable patterns or trends in their expression. It was a tough homework assignment and overall, my code feels super messy and unrefined, but I tried my best to keep it organized and to thoroughly answer some of the questions presented in the homework document. I could have made ample use of classes in this homework, but so far, we've only touched the surface of classes, so I held off on it. Without further ado, here are the results of my analysis.

WordClouds:



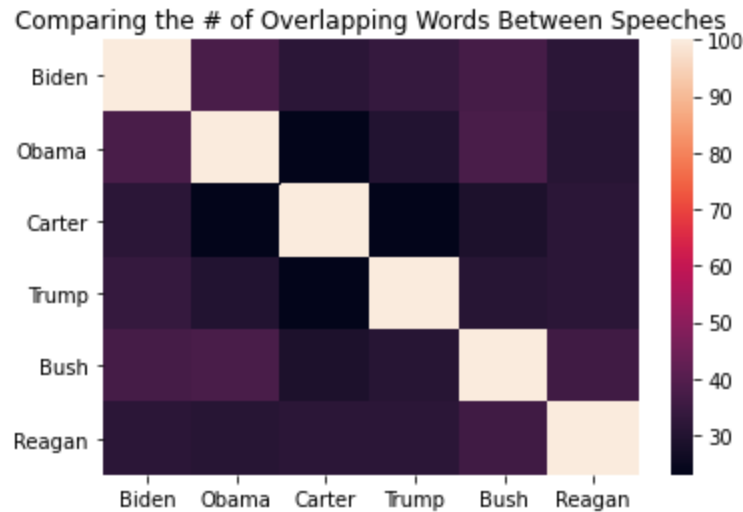
My first task for this homework was to create a figure visualizing the word clouds between each of the 6 speeches I chose. Implementation was extremely easy with the wordcloud package doing all of the heavy lifting. Something that sticks out from this picture is the prevalence of the word: “will”. It makes sense why this would be such a common word after accounting for all the stop words, but interestingly, President Carter seems to have avoided the word altogether in his speech. He must have gotten bored of hearing it from past Presidents. Another interesting observation is that Trump tends to favor “America(n)” over “US”, an occurrence that actually makes a lot of sense when thought about. It's possible that Trump uses intense American themes in his speeches and expressions to galvanize and rally his supporters.

Sentiment Analysis:



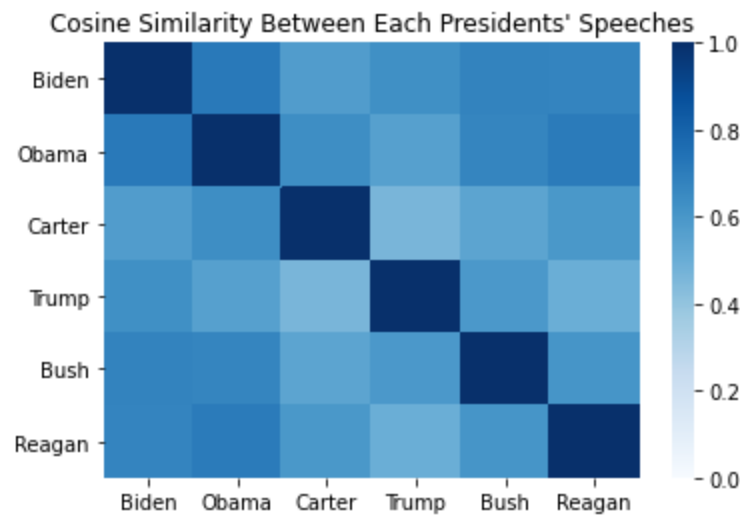
This was a fun graph to make, sort of. I used matplotlib to make a bar graph after realizing that using seaborn may not be the best idea, and I used this website (<https://www.pythonprogramming.in/change-box-color-in-boxplot.html>) to change the colors of my bar graph to match the political party colors of each President. As you can see, the democrats - Carter absolutely carries - seem to slightly outscore the Republicans in the most positive speech department. Carter's outlier occurs in his third to last sentence which is mind boggling long and contains six bullets on strengthening American ideals. The sentiment score of this sentence is nearly twice as high as any other sentence from any other speech, although, if you look at the tail end of Carter, he's actually the lowest. The rest of the field is all pretty tight other than George Bush who probably made his speech a little too real.

Overlapping Words:



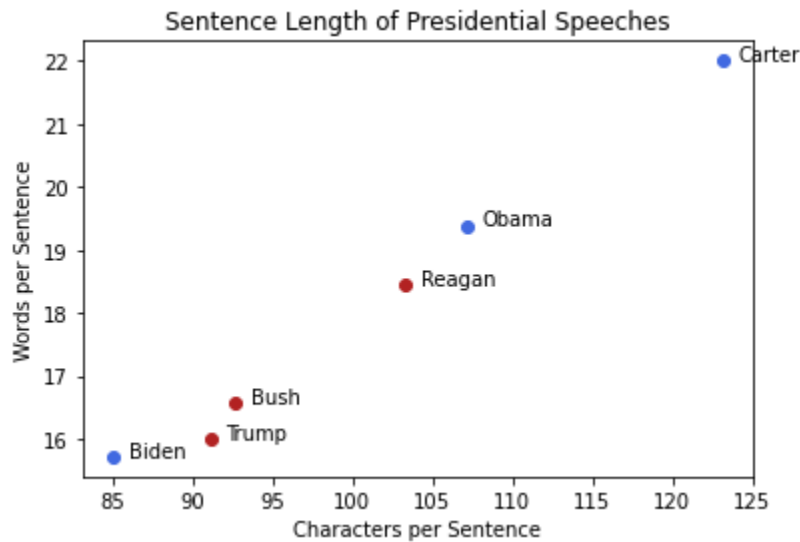
This graph shows the amount of overlapping words between each combination of speeches. I first found the k most common words - using the Counter module - in every speech and then compared every speech's common words with a different speech's common words until every combination was met, its number of overlapping words counted up, and its value inputted into an empty array. From this heatmap, we see that, obviously, Biden's speech is identical to his own and so on. We also see that the two people with the most different speeches are surprisingly Obama and Carter and not as surprisingly, Trump and Carter. In general actually, Carter seems to be the biggest maverick of them all when it comes to speech writing. His speech is the most dissimilar to everyone else's and doesn't seem to have 40 overlapping words with any other President's speech. It seems that the most similar speeches would be President Obama's and President Bush's.

Cosine Similarity:



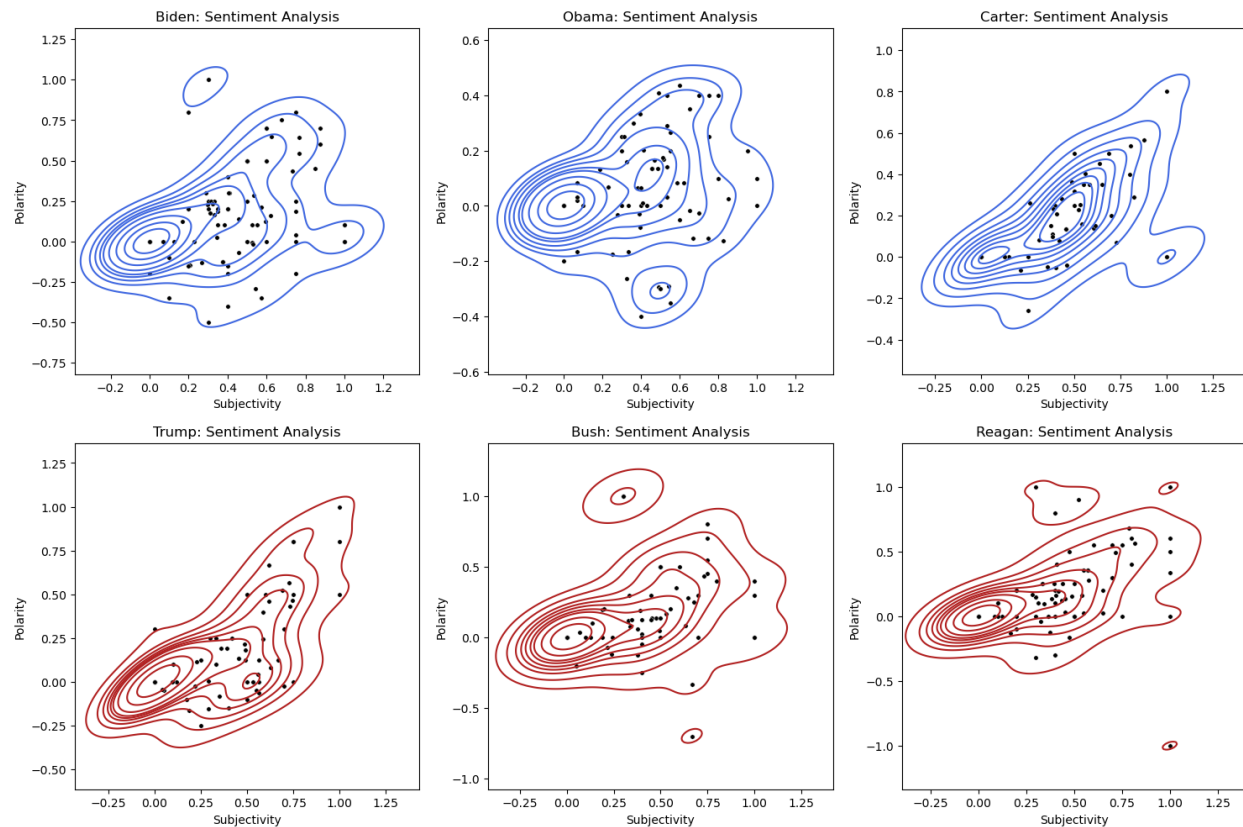
Moving onto the final part of the homework, we have another heatmap, this time depicting the cosine similarity between each Presidents' speech. The calculation and implementation of cosine similarity is similar to that of the overlapping words functions, but with cosine similarity, we go a bit more in depth than just whether or not two words overlap. Instead of simply counting the number of words that overlapped between two speeches, we create a huge vector for each speech, every element representing a unique common word with its value representing the number of times it appears in the speech. We then compute the cosine similarity between every two presidents and plot them out in the heatmap above. We see a similar result from above where Trump and Carter seem very dissimilar, but we also see new results where now, Obama and Reagan seem quite similar and so do Biden and Obama. What trends can we make out from this? Well, taking from this and the last heatmap maybe Obama was having a little fun copying from everyone's speech. A better observation though, is that Joe was Obama's vice president and a close succeeding President to Obama which could both potentially explain some of their similarity. Obama and Reagan were also both known to be incredibly charismatic presidents, perhaps a trait that contributed to the similarity of their speeches.

Sentence Length:



Next up is the simplest and most straightforward graph I created. It finds the average amount of words and characters within each speech and just plots a point on a graph. The findings here; however, are still incredibly telling. For one, they offer an explanation as to why Carter had such a high sentiment score for some sentences: that Carter packed tons of words into his sentences, increasing the opportunity for those words to be hopeful and sanguine. We can also corroborate some other sentiments spread around by people and news outlets online. *Not saying that any of these are true* For example, Sleepy Joe isn't the most articulate or well-versed speaker and neither is Trump for that matter. A word length function would be a nice follow-up to this figure which could potentially illustrate how well-spoken each president is.

Polarity and Subjectivity:



Finally, we have our polarity vs subjectivity figure. The figure contains 6 subplots for each of the 6 speeches, visualizing the polarity and subjectivity scores of every sentence within each speech. The code for this figure was actually relatively straightforward. It involved using Textblob's built in sentiment analysis features and seaborn's built in Kernel Density Estimate plot function which were both thankfully demonstrated in class. The polarity of these graphs represents the positiveness and negativeness associated with each sentence while the subjectivity represents how subjective (opinion-based) and objective (fact-based) each sentence is. From what I can tell, although the graph looks a little misleading, President Carter still seems to have the most positively associated speech, an observation that supports the findings from my own sentiment analysis. Such an extensive graph like this allows for so many thought-provoking ideas to explain each visualization and so many intriguing conclusions to be drawn from each result. Take Obama's speech for example. Coming off of the Great Recession, Obama had to address the 10% of Americans who were jobless and the more than 5 million Americans who had lost their home. Being positive during that era would have been hard for any President. There are so many thoughts to be had, so many connections to be made, yet I've only touched the tip of the iceberg in this analysis.