Ayo Akinnawonu

EMP 532

Professor Woodson

April 12, 2019

Text Analysis Project

Part 1: Alice’s Adventures in Wonderland by Lewis Carroll

The dataset that I will first analyze will be the novel Alice’s Adventures in Wonderland, more commonly known as Alice in Wonderland by Lewis Carroll. The dataset was acquired from Project Guttenberg at <http://www.gutenberg.org/cache/epub/28885/pg28885.txt>. It should be noted that I consulted with the following sources:

1. Tidy Text Mining - <https://www.tidytextmining.com/tidytext.html>
2. R- Bloggers - <https://www.r-bloggers.com/intro-to-text-analysis-with-r/>

These two sites helped to provide references in which to write R code and helped in consultation of which functions to use in order to write such code.

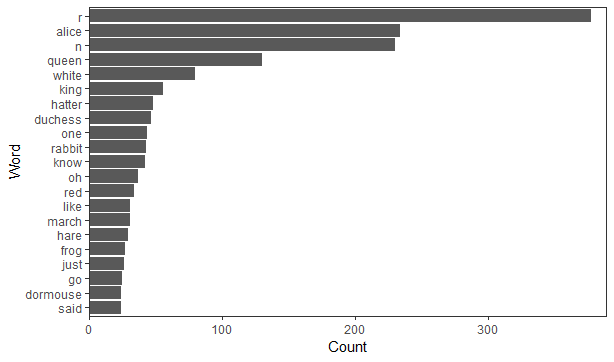
About the data:

Alice’s Adventures in Wonderland is a fictional novel by the author Charles Dodgson writing under the pseudonym of Lewis Carroll. The novel was published in 1865 and it tells the tale of a girl named Alice as she travels down the rabbit hole and meets all sorts of strange and interesting characters. It should be noted that Lewis Carroll was also a mathematician, a logician specifically, and the novel Alice’s Adventures in Wonderland has been interpreted as a game of chess. This interpretation come from a view that Wonderland is a chess board and Alice is a pawn. She has no power and can only move forward however the story takes but as she makes to the end, she can become whatever piece that she wants to be and the most powerful piece on the board is the queen and Alice’s biggest obstacle is the opposing queen.

Preparing the data:

The book in its full text was available for download from Project Guttenberg. However, the formats that the book was available in was not compatible with RStudio, so I copied a plain text version of the book and pasted to a blank document in Microsoft Word. I then saved this as a text file and imported it into RStudio. In order to do analysis on this text I found that I had to convert the value of the words. The plain text file on Project Guttenberg encoded each unique word as an integer and so I converted the text to a character string in order to continue.

Key Words in Context

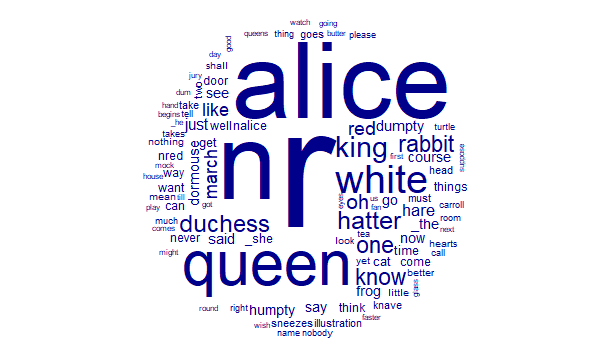


The plot above represents the result of the 20 most frequently occurring words that are found in the text of Alice’s Adventures in Wonderland. There is a question to be raised by the presence of hundreds of singular “r” and “n” in the frequency count. The best explanation I can reason for that is that some characters were corrupted and there was a production of these letters within the character string of the text. They do not appear in the regular text and so they will be ignored. Some of the words that occur they most within the text are Alice, Queen and King. These of the main characters of Alice in Wonderland. Alice occurs the most, but the rather interesting observations occur between the Queen and the King.

When you think about these terms, you think of an authority that rules over their dominions. Perhaps, the king is more important than the queen, In Alice in Wonderland, that does not appear to be the case. Queen appears more than twice then the king and more authoritative actions are closely associated to the appearance of the Queen. The court of Wonderland appears to sentence many people to be beheaded rather frequently. When using the kwic() function, it was discovered that the word head appears often in conjunction with the Queen. The most quoted line from Alice’s Adventures in Wonderland is “Off with their heads.” The Queen shouts this over to the court more than a few notable characters including Alice.

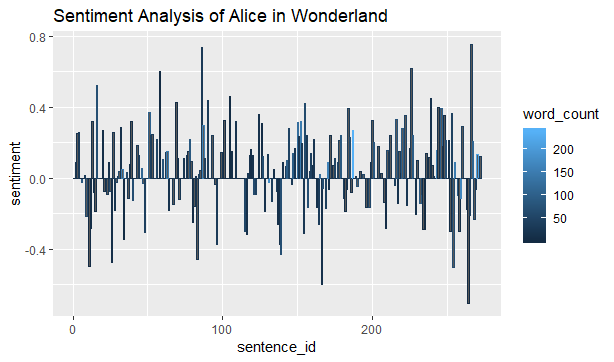
The King, for his part goes along with the Queen but the analysis shows that he does not make any of these decisions himself and is much more lenient than the Queen. The analysis of the text surrounding the King is significantly more passive than that of his wife. The role of the these two is very similar to their roles in a game of chess. The king is at the center of everything because the objective of the game is to corner and take out the king. The Queen on the other hand is the most powerful piece and she often decides the course of the game. The Queen of Hearts makes all the decisions and the entire kingdom fears her.

The world cloud below is another representation of the key words within the novel. It serves as a helpful illustration of how much the Queen is pivotal in the text which is fitting as the antagonist to the heroine. The names of the characters are among the most frequently occurring words in the text. Alice’s Adventures in Wonderland is written like a play but even when considering that the characters names would precede their respective lines. There is a greater count of the mentions of the names then there are lines attributed to the characters which means that they are mentioned within the lines spoken by other characters. This suggests that in the interaction of the characters, there is a lot of discussion which character discuss one another, particularly Alice and the Queen of Hearts



Sentiment Analysis

Alice’s Adventures in Wonderland is a rollercoaster of a tale for our protagonist Alice. During her time down the rabbit hole, Alice is unsure of where here adventure will take her. In that time, she meets all sort of weird and colorful personalities. Some of which do not mean to be her friend. To this end, it would help to know how positive Alice’s experiences were during her time in Wonderland. The plot below demonstrates the ups and down of Alice’s adventure.



The text of the novel was parsed into sentences and they were referenced for how positive and how negative they appear in RStudio. The x- axis uses the sentence order to plot the data wo the graph represents the book in order from start to finish. The y – axis represents the value that the sentiment function placed on each sentence by referencing their content to a predefined list of positive and negative words. The graph shows that over the course of the novel the overall sentimentality is positive but there were many sentences in which the sentimentality of was negative. It is important to note that longer sentences tended to be more, or at least, closer to neutral in their score when compared to short sentences. Perhaps longer sentences are more nuanced in their descriptions and so this leads to sentiment scores that are closer to zero. A theory could be that the shorter the sentence, the more weight that each word that appears has on the score of the sentence score. Shorter sentences, therefore, are more binary in their sentiment and that greater dichotomy can signal a lack of expression within the sentence.

Part two: The Farewell Address of Presidential Dwight D. Eisenhower (1961)

The text is the final speech of the presidency of Dwight D. Eisenhower (1953-1961). The text from Presidential Eisenhower’s farewell address was retrieved from ourdocuments.gov at the URL <https://www.ourdocuments.gov/print_friendly.php?flash=false&page=transcript&doc=90&title=Transcript+of+President+Dwight+D.+Eisenhowers+Farewell+Address+%281961%29>

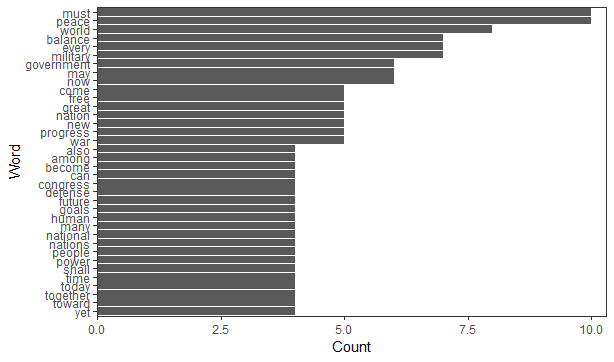
About the data

It is customary for a departing president to give one final message to the American People before they depart from office. Commonly know as the farewell address, each president who has had the opportunity to leave the office on their own terms has taken this moment leave some parting wisdom to the American populace. In 1961, President Dwight D. Eisenhower was departing office after serving two terms. Before that, he led Allied forces through the bloodiest conflict in human history and help set up the North Atlantic Treaty Organization or NATO and serve as its first Supreme Allied Commander Europe. This speech would serve as a reflection of over fifty years of service which saw some of the worst atrocities committed in history. In addition, Eisenhower would be leaving a United States that would be entering a new sort of conflict. An era that would be defined by opposing superpowers who have the power to end life on Earth. The Cold War would shape geopolitics for half of the twentieth century and its legacy has had lasting impacts on our modern word. Eisenhower channeled his experiences in war and diplomacy to give a stark warning about the challenges that lay ahead for the country.

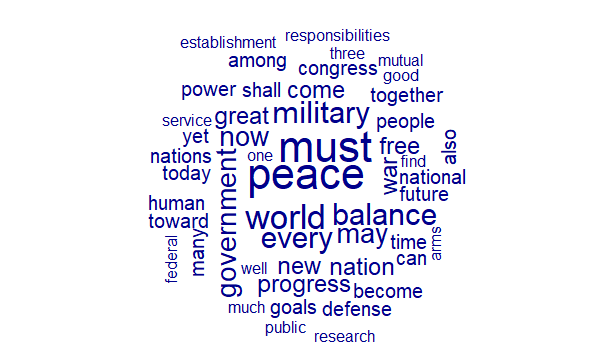
Preparing the Data:

At just over 1800 words, I copied the text of the speech and pasted it into a character string and defined it as a variable. From there, the text was set up as a corpus and tokenized. Finally, the stop words were removed so that analysis could be done.

Key Words in Context



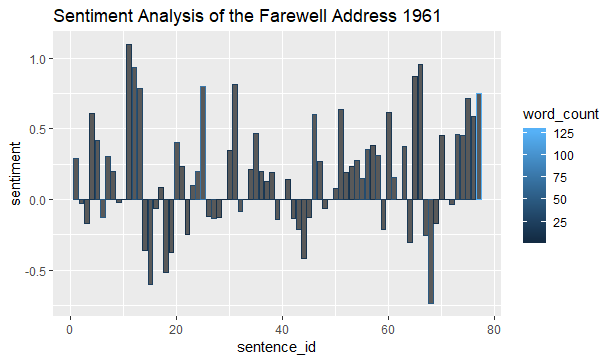
The above plot visualizes the most common non stopwords that appear in Eisenhower’s farewell address. The word “must” is tied for being the most frequently occurring word in the text. Found in 10 out of 77 sentences in the speech. Its repeated use signifies that the text is declarative in nature. Out of 10 sentences, must is preceded by the word “we” in five of them. These sentences are warnings about what the American people must be, must become and what they mustn’t become. The recurrent use of such declarations using “must” supports a narrative that Eisenhower can speak with authority. Having had command over millions for a period of over 20 years can instill such confidence in an individual to command such authority and likely with it, respect. The word “must” is tied with “peace” as the most frequently repeated word. For career military officer turned president who prosecuted a war during his administration, peace is an interesting word to focus on. Perhaps, the person who has seen the most war is one who strives for the most peace. The use of the word is often in conjunction with other ideological terms. “Peace” is near the use of the words progress, betterment, prosperity and justice. Form this use, it appears the President Eisenhower wishes to invoke a feeling of what can America, perhaps what it deserves from a hard-fought peace.



The word cloud above represents of the rhetoric used within the farewell address. The terms that made their way into the cloud through the frequency of their use reveal an overarching story about the message of the speech. Peace is paramount but the needs of the nation and the world is balanced by the use of the military. Now, President Eisenhower is not advocating the use of armed force for its own sake. Likely, he advocates that a measured hand is necessary in order to wield such awesome power. Military, power and war is tempered by peace, service and progress. Eisenhower is trying to convey a future, a word which is featured on the word cloud, in which the conventional means of power in previous generations. The use of the word “balance” is important in this respect. Mentioned seven times in the text, five instances of the word occur near each other. “World” is also reoccurring. During the Eisenhower administration, America became a world superpower. Eisenhower has to frame America as a leader in the world and so the world is brought up often as a means of comparison and perspective.

Sentiment Analysis

A farewell address to a nation can be an emotion goodbye to a country that the president has had the privilege of serving for years. Such a speech can reveal how a president they have left the country and what final warnings or advice they feel they would like to leave the country. A president’s farewell address can ultimately help to define the legacy of that president. The following plot shows the sentimental analysis of Eisenhower’s speech. The text is separated by sentences using the sentences\_by function. Then, using the sentimentr function, Rstudio can calculate a sentiment score for each sentence. Using ggplot, I was then able to graph the data.



Despite the recurring mention of war and the military, the speech overall had a net positive sentimentality. And the graph, which the columns represent the individual sentences and their sentiment scores were mostly positive. Alice’s Adventures in Wonderland, by comparison, had an even split between positive and negative sentences. The time in which the speech was written and the subject matter that it wishes to discuss could have been a reason in which it was decided to go for a hopeful and positive measure even in the face of potential nuclear holocaust. Eisenhower’s sentences are short and succinct with many of his sentences being 20 words or less. He conveys a point and moves on. There is little room for interpretation in Eisenhower’s sentiment, he knows what it is he wishes to say and just says it. As for the end of the speech, it is overwhelmingly positive. A fitting end to an overall positive speech. Eisenhower tackles some broad and usually negative topics and ultimately turns it into an uplifting final message to the American people.