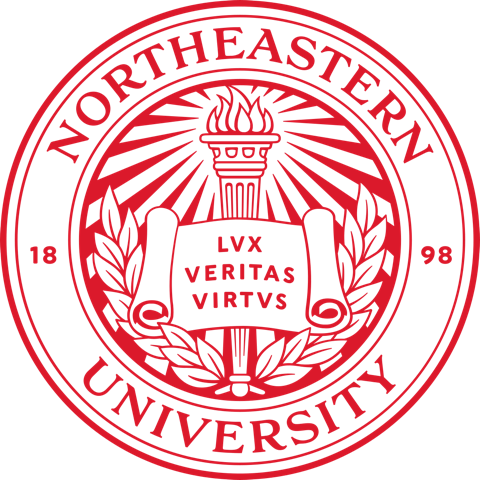
**Course:** ALY 6040

**Instructor’s Name:** Prof. Dr. Justin Grosz



**Submitted By:**

Sunil Raj Thota

**Date:** 10/21/2021

**Introduction:**

The March on Washington and King's address are largely seen as pivotal milestones in the Civil Rights Movement, bringing racial equality demands and confrontations from the South to a national platform. These are the thoughts that sparked a revolution. The speech's main theme is that all individuals are created equal, and that, while this was not the case in America at the moment, it must be in the future. He argued strongly and aggressively. This speech is intended to educate the American public about the injustice of racial inequity and encourage them to stop discriminating based on race.

Using this speech as an exercise, we can assess the pragmatic contents and sentence patterns of the speech. We can also look at how MLK used the speech to effectively advocate for equality and overcome racism. Text mining allows researchers to swiftly evaluate large amounts of data. Mining can reveal key connections between individuals that could go undiscovered. We can uncover hidden facts and gain sufficient insights. It also aids in the discovery and development of novel hypotheses. This speech has the potential to yield more detailed findings that have suggested important trends and facts. I wanted to use this chance to analyze the words he used in his speech to better comprehend the feelings he was expressing.

**Analysis:**

**Step 1: Getting the information and loading the packages**

For this assignment, I used many packages for natural language processing, data cleaning, and text mining. Text mining is made simpler by using tools like tm. Text stemming is done by using SnowballC. Word cloud is used to generate the word clouds. RColorBrewer provides a range of color palettes. Rcpp package provides integration of R and C++ library integrations. Here, the result is a structure of type Corpus that is loaded into the memory. It is a list of 46 items. The main goal of this presentation is to analyze text data.

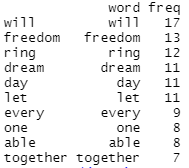
**Step 2: Clean up the textual data**

The following are some of the most common data cleaning activities connected with text mining: Changing all of the text to lower case Getting rid of punctuation marks (periods, commas, hyphens, etc.). Stop words (very common words like "and," "or," "not," "in," "is," and so on) are removed. Stop words are words that are so prevalent that they have a low information value. Numbers are being removed. Filtering off terms that we don't want to see like punctuations. Extra whitespace has been removed. The tm package includes numerous functions for performing the tasks, which are then applied to the data set as modifications using the tm map() method. Create content transformers, i.e., functions which modify the content of an R object. tm\_map is used for an interface to apply transformations to corpora.

Text stemming, which lowers terms to their root form, is another crucial preprocessing step. To put it another way, this procedure removes suffixes from words to simplify them and determine their common origin. A stemming process, for example, lowers the terms "moving," "moved," and "movement" to the root word "move."

**Step 3: Generate a TermDocumentMatrix**

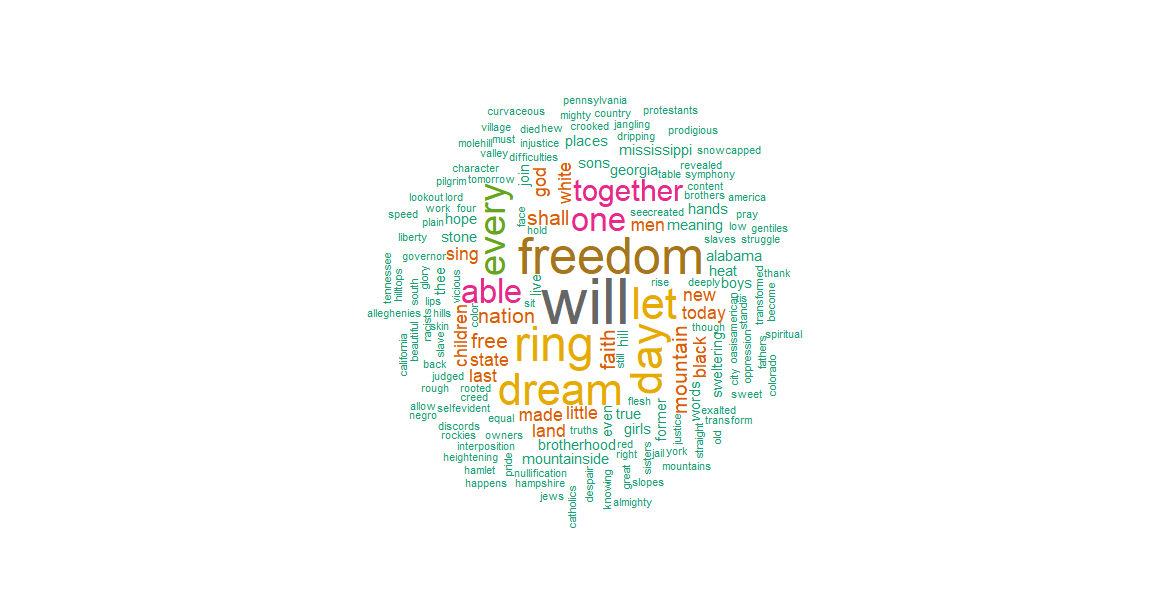
TermDocumentMatrix in R is used. It is the development of a term-document matrix from a corpus is a typical text-mining approach: components in this matrix represent the presence of a term (a word, or an n-gram) in a corpus document. As a next step, create a data frame with each term listed in the first column and their frequency listed in the second column. The TermDocumentMatrix function from the tm package can be used to create a document term matrix. This is how the data frame looks like.



**Step 4: Generate a Word Cloud**

To make word clouds, you'll need the R package word cloud as well as the RcolorBrewer package for the colors. They provide text data in a simple cloud structure, in which the size of the words is proportional to their frequency. As a result, they are pleasing to the eye as well as simple and quick to comprehend. Word clouds are illuminating.

Their visual structure encourages us to think and reach the best conclusions possible, based on what we're looking at.



The five most essential terms of Martin Luther King's "I have a dream speech" are clearly shown in the following word cloud: "Will," "freedom," "dream," "day," and "together."

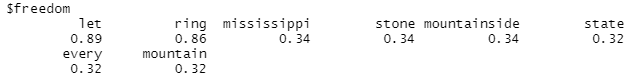
**Step 5: Frequent Terms**

In a document-term or term-document matrix, look for common terms.

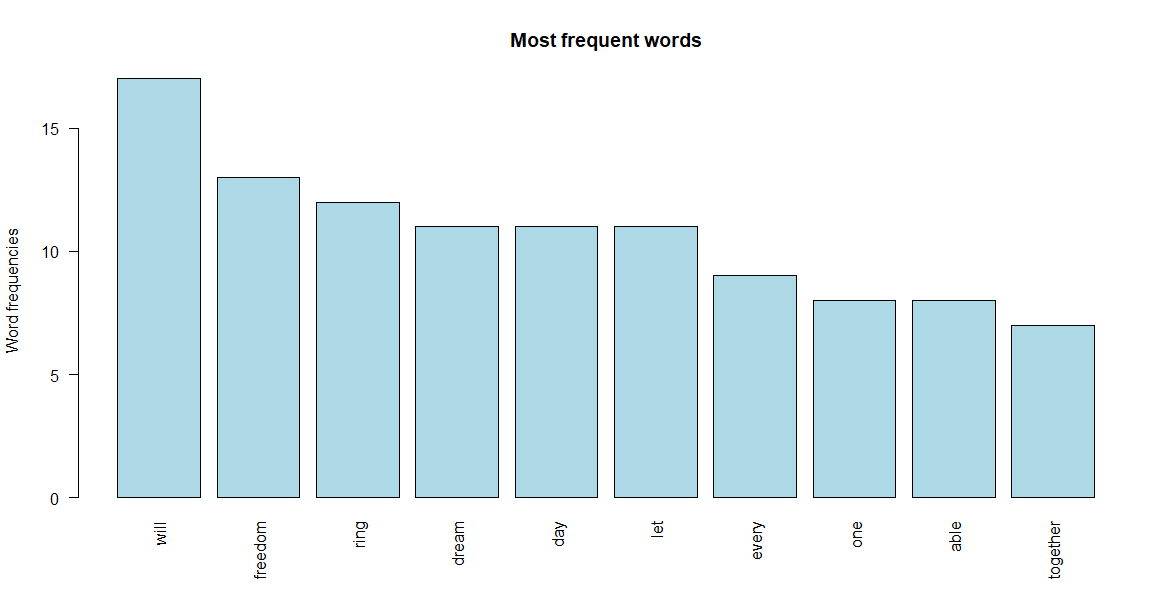


**Step 6: Find Associations**

Find associations in a document-term or term-document matrix. We may also use the findAssocs() function to investigate the relationship between common phrases. This lists the words in the I Have a Dream speech that is related to the word freedom.



**Step 7: Most frequent words – Visual Representation**



In a document-term or term-document matrix, let’s look for common terms from the above picture. We can use this data to show the most common words in the I Have a Dream speech once we've completed all of the text analysis. To display our findings, we've created a bar graph. The terms will, freedom, and ring are obviously at the top of the list of his most commonly used words in the speech.

**Interpretation and Recommendations:**

The bar plot tells an interesting message where it shows the good sentiments and positive impact. Words with a positive meaning occur more frequently in the discourse. This is a wonderful example of why it's critical to examine data from several angles. Another thing I'd want to point out is that if you look at the Step 5 results, do you see any of the same words here in Step 7? Some of them are missing, as you will observe. His speech includes basic emotions like trust, sadness, joy, anger, fear, surprise, disgust, and expectations. I identified the speech is in the English language which is a process in lexalytics. Tokenization is the process of separating text documents into their constituent parts. I would recommend this approach to use for other speeches. For Data owners, these results give a good understanding of the emotions which are necessary to boost the individual and also to build a better future together. They can also focus on the driving forces for the text mining processes and business impacts.

The next step that I would love to perform is sentiment analysis on the speech using the tidytext package's get sentiments() function. This allows us to organize sentiment lexicons into orderly lines, one for each word. This will classify words as either good or negative sentiments and sort them by frequency. The most prevalent positive and negative terms in this speech can now be easily plotted. We can examine word counts that correlate to each emotion when we have a data frame containing both sentiment and term.

**Conclusion:**

In his speech, MLK Jr. used more good words than negative phrases, as shown in this picture. While the accessibility and complexity of each little segment change greatly, the overall level remains pretty similar all through the speech. The most important factors in memorability were clarity and placement. The speech portrays a picture of a better, more equitable future characterized by racial peace and inclusion. The most renowned sentence is the theme "I have a dream," which is repeated over and over to emphasize King's motivational ideas.

Perhaps this is one of the factors why this speech is more remembered and had a significant impact on the civil rights movement in the 1960s. Even after more than 50 years, his optimistic word choice made the "I Have a Dream" speech so compelling. The speech begins and finishes on a positive note, with a positive middle portion punctuated by two troughs to break up the monotony.