Task 1&2

Zening Ye

12/6/2021

Book Selection

The book I used for text analysis is The Martian by Allen Glasser. The book talked about a series of stories that occur when a person from a different planet comes to a different place. A stranger from a different world, bewildered by his new surroundings, unable to express his desires, falls into the hands of others, and his fate may not be happy.

Word Analysis

At the beginning of the analysis, I used inner_join to join three different lexicons with the original data to identify the sentiment word.By inner join three different lexicons we will have following plots.
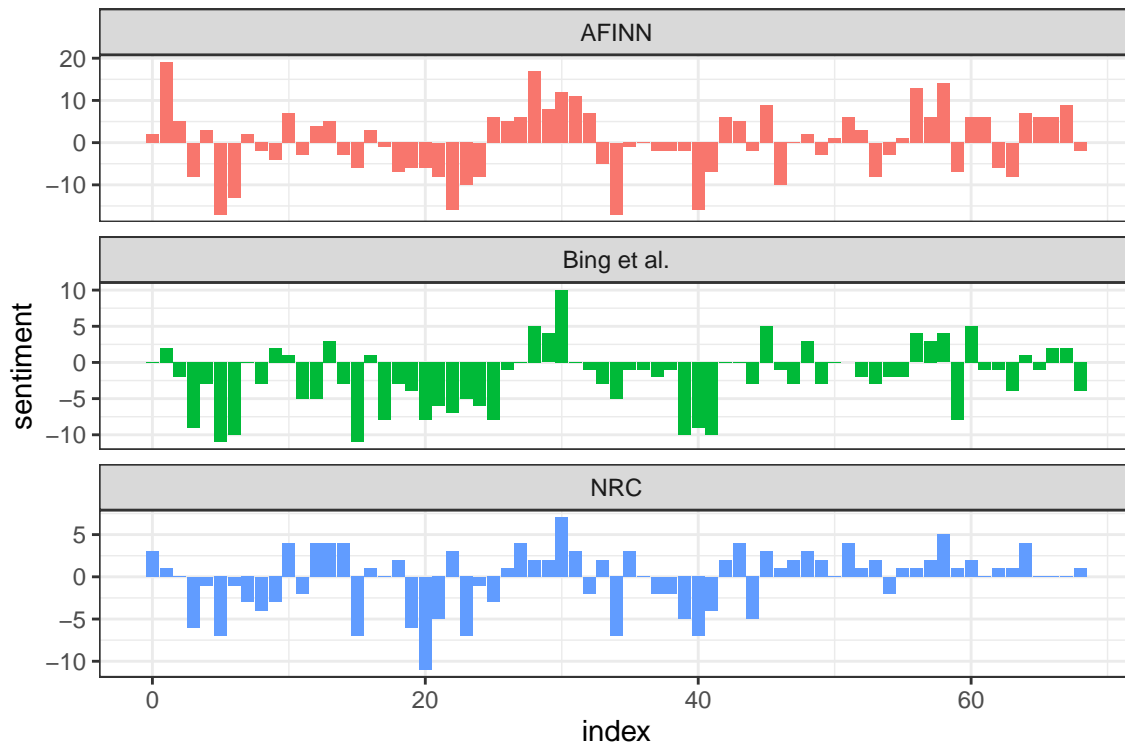
Figure 1: Visualization of Three Lexicons

After these plots, I realized there might be more negative emotion than positive emotion in this book. As you can see in "bing" lexicons, for instance, it indicates there are more negatives from the beginning of the book. By looking for other plots, we can also realize the negative emotion starting from the beginning. The three lexicons use different scales to measure the sentiment level, it is obvious that the definition of the data joined by these individual lexicons are different. Therefore, it is reasonable to have different interpretations under different lexicons.

Next, I compared positive and negative sentiment under the original database. Even though we can see the word "great" used a lot in the book, it is still hard to change the entire analysis for the book.
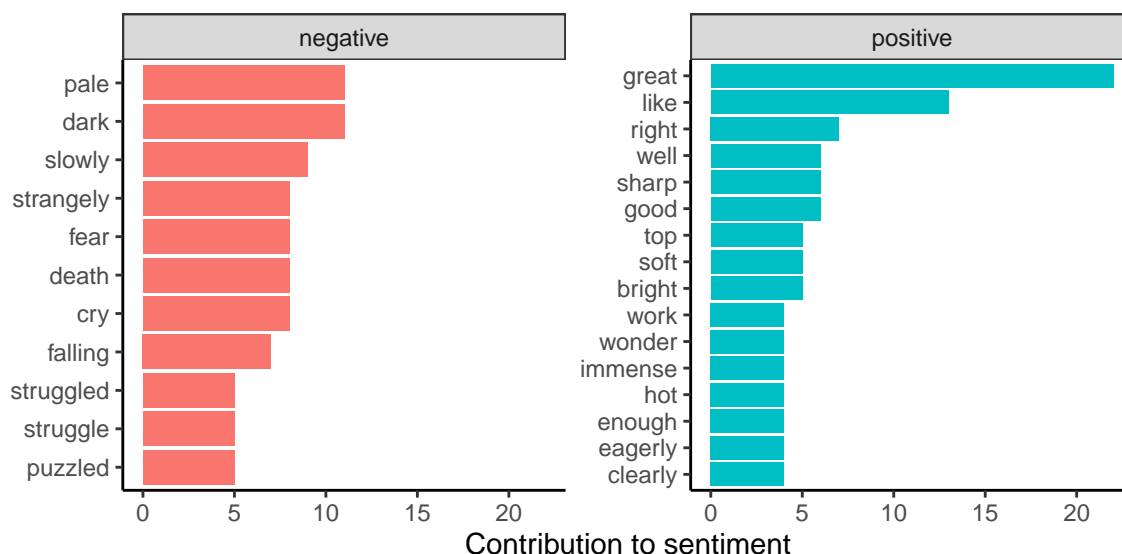

Figure 2: Negative and Positive Word Count

Wordclouds

A word cloud is a way to show how often words are used throughout the book, which allows us to understand better how the words are used in the book, with larger ones indicating more frequent use and vice versa. The word cloud is equivalent to presenting the most common words in the book as a whole, in a better visual format. On Figure 3, the word "eye" and "light" under specific conditions that show it came up a lot.

Positive and negative emotional word clouds (Figure 4) can more visually reflect the distribution of emotions throughout the book, and we can present different emotional word clouds by changing the number of words that are used. The positive and negative word cloud might be similar with the positive and negative word count that shown above, but this is more on the entire data frame.

Conclusion

Under a series text analysis, I have a basic view for the book I chose. Even though the Martian is science fiction, it came up with a different perspective I never thought about before. Furthermore, I realized that although the number of positive words is higher than the number of negative words, the total ratio of negative words is higher than that of positive words.

Extra Credit

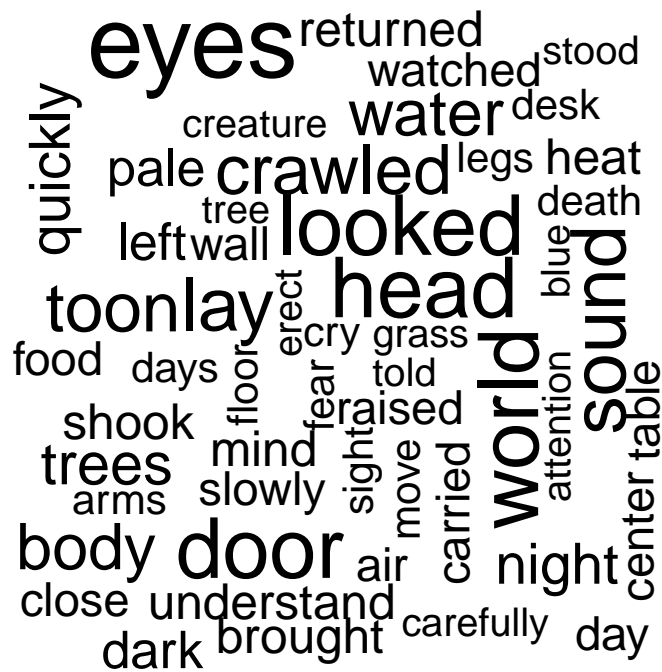Using different lexicon for sentiment analysis: *loughran*

Figure 3: WordCloud For the Book

Figure 4: Positive and Negative Word Cloud

As you can see follow, the result by using this package is similar to the result I got from other three packages. In addition, this graph is very similar to the second graph I have on Figure 1.
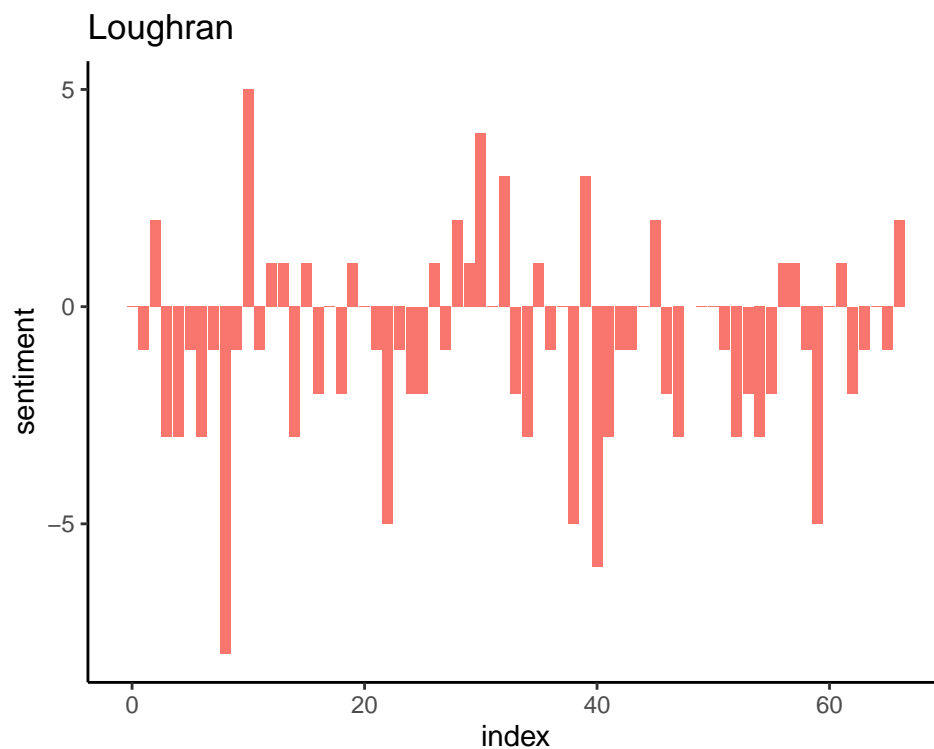

Figure 5: Loughran

Citation:

Text Mining with R: https://www.tidytextmining.com/