TASK2

Yuli Jin

2021/12/7

Task 1 Pick a book

I choose `The Burning Secret` as the my text analysis. This book was written by Zweug, Stefan.

TASK 2 bag of word analysis

First, I use three types of sentiment analysis methods AFINN, Bing and NRC to plot barplot to compare these methods. From the graph below, the AFINN and Bing method fits better. Most of the polt in `The Burning Secret` is in negative. In this book, While being treated for asthma at a country spa, an American diplomat's lonely 12-year-old son is befriended and infatuated by a suave, mysterious baron. But soon his adored friend heartlessly brushes him aside and turns his seductive attentions to his mother. The boy's jealousy and feelings of betrayal become uncontrollable. The story is set in Austria in the 1920s. That is to say, at the beginning of the book, the sentiment of the book is positive, but soon it converts into negative sentiment. However, it is difficult to identify which of the two methods is better. In the following task, I use Bing method to conduct further analysis.
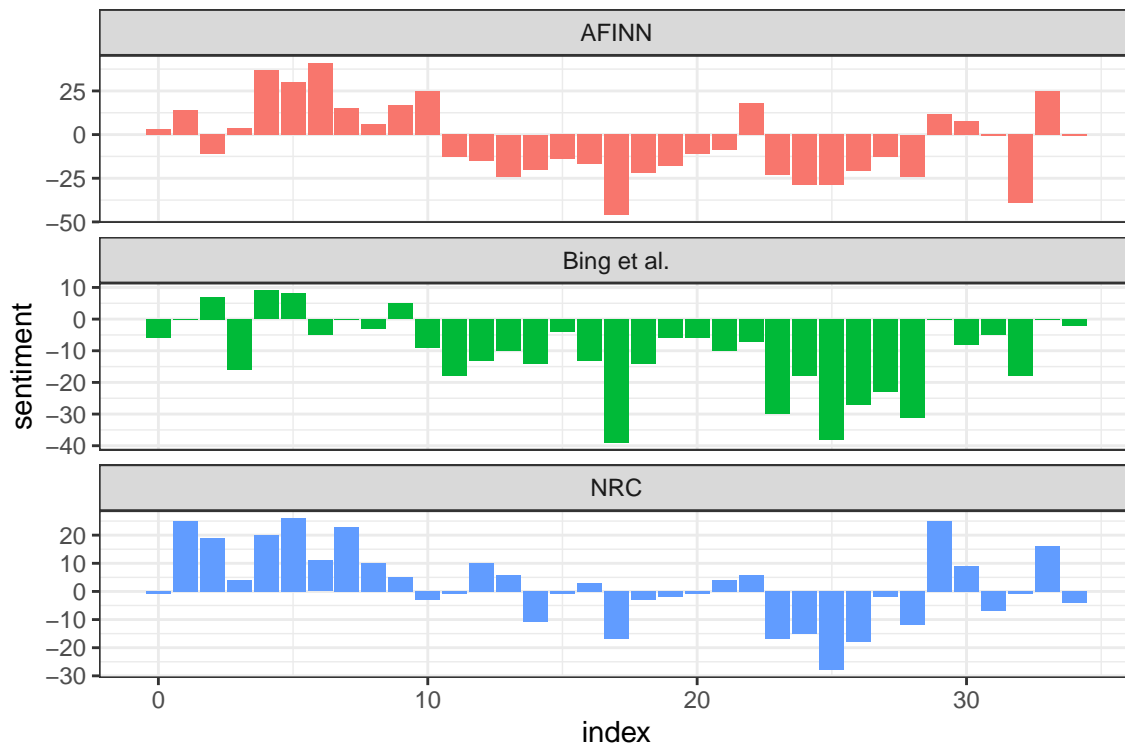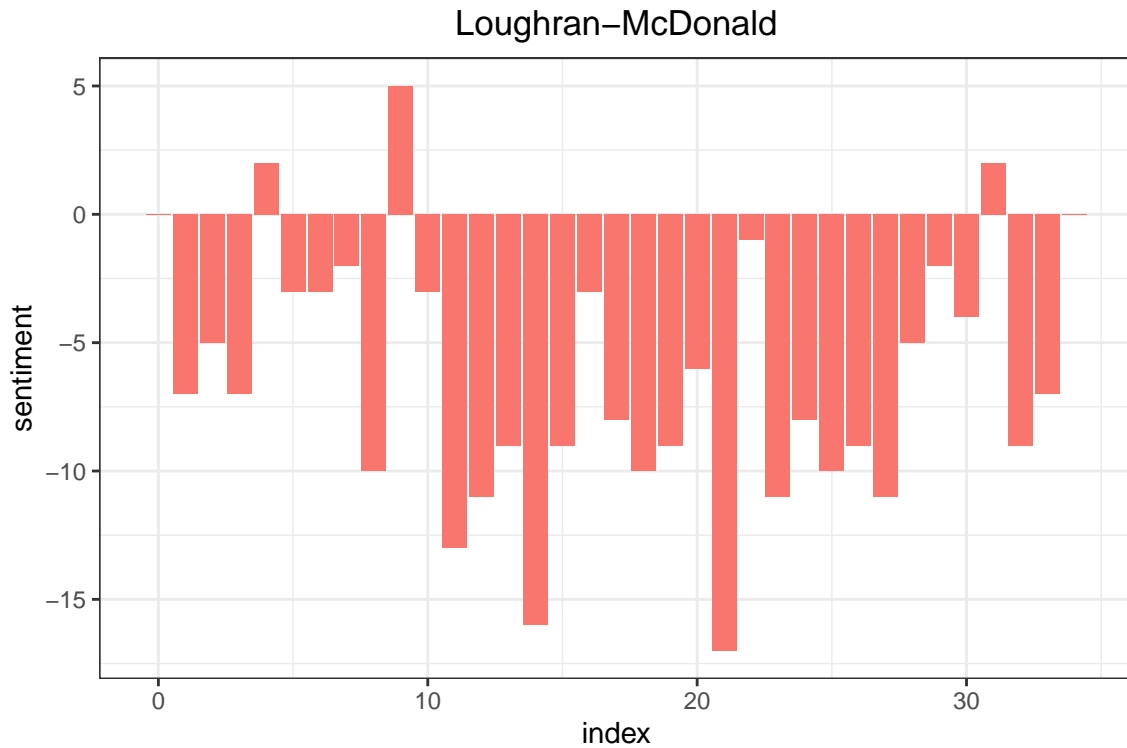

Figure 1: sentiment plot

1

Figure 2: negative positive words count

Figure 2 shows negative and positive word count of each word. For the negative chart, dark is the most common words throughout the whole book. Hate and darkness rank the second and third place respectively. For the positive chart, like is the most common words throughout the whole book. Great and good rank the second and third place respectively.

Figure 3: word cloud

Figure 3 displays word cloud which shows the frequency. As we can see, baron, mother, edgar are the most frequency words among all the words. It is reasonable because they are the main characters in that fiction book. In task 3, I will use two of three characters to conduct further analysis.

Figure 4 generally converts Figure 2's information into word cloud

Figure 4: sentiment word cloud

Task 2 extra credit

In `textdata` package, there are one extra lexicons available to use. This lexicons is called `Loughran-McDonald`. Here I use this new method and plot the similar graph to show the progression from start to finish of the book.

Loughran–McDonald

Then I provide some description. According to the introduction of https://emilhvitfeldt.github.io/textdata/reference/lexicon_loughran.html, this lexicon is created for use with financial documents. Therefore, the graph is complete different from previous lexicon graphs. Therefore, this lexicon cannot correctly reflect the exact sentiment of the book. After all, financial sentiment lexicon isn't necessarily suitable for fiction book.