

MA615 Assignment4 Task Two

Yuyang Li

2021/12/8

Task one : Pick one book

The book I selected called Old Granny Fox written by Thornton W. Burgess from Gutenberg. It's a fairy tale.

Task two : Words analysis

Firstly I cleaned the data into tidy format and tokenized it, and imported three methods to analysis the sentiment of text.

Figure 1 shows box plot of sentiment using BING lexicon, and figure 2 shows analysis with three sentiment lexicons. I found that using BING lexicon fits best among these figures. And the main sentiment of whole book tend to be positive.

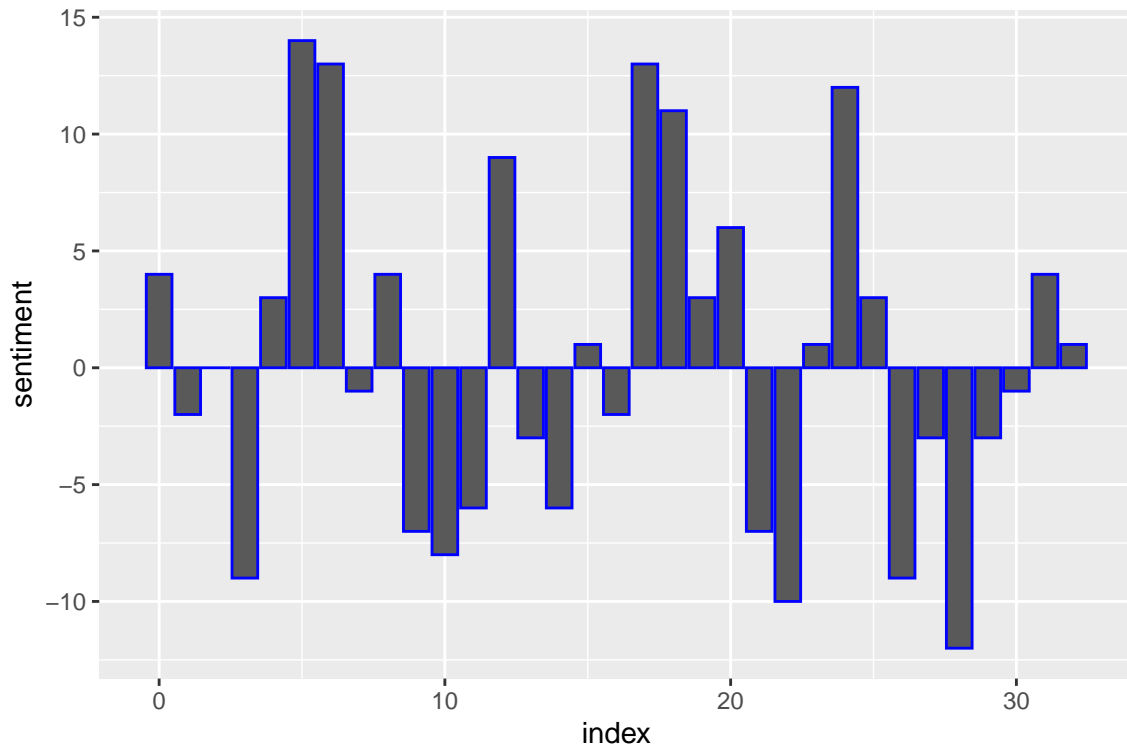


Figure 1: sentiment plot using bing

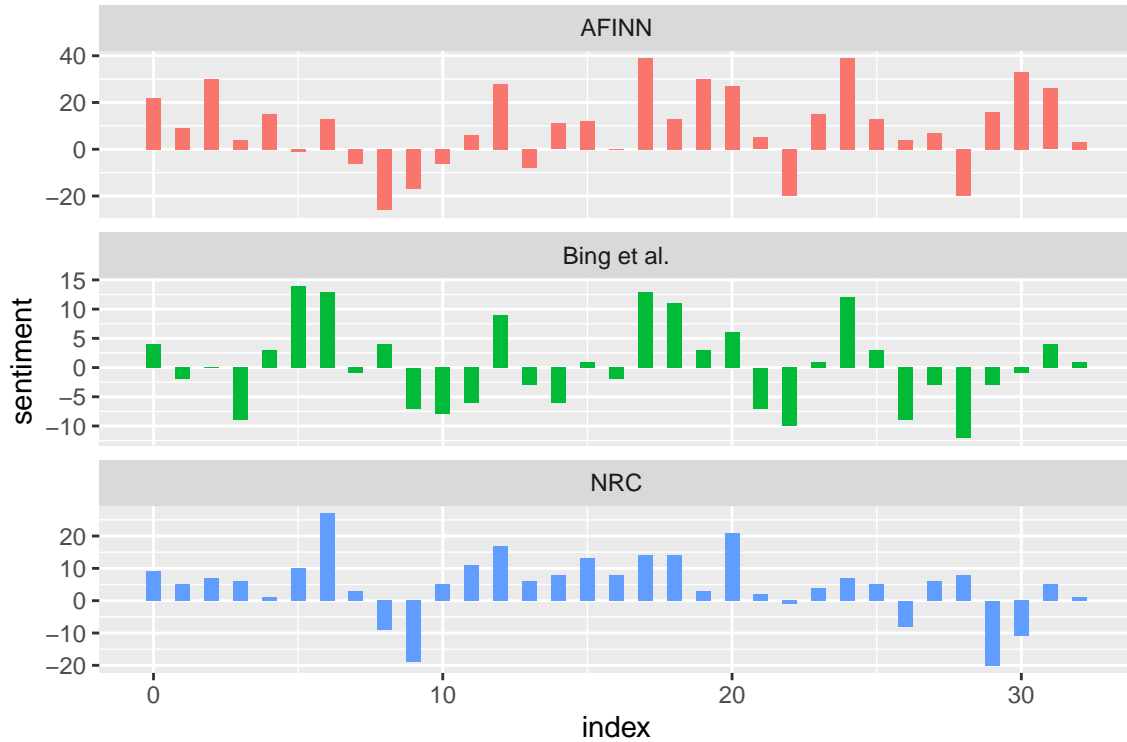


Figure 2: sentiment plot

Figure 3 shows the number of positive and negative words appeared in book. It's clear that the frequency of positive words is higher than negative. The highest contribution word in positive group is “right” and it's “foolish” in negative group. As we know that this book is a fairy tale, and the main content of the book is when a deep winter snow carpets the Green Forest and nearby meadow, Granny Fox and Reddy have some disagreements on how best to find some food. But Granny – with her years of experience – wins out over Reddy and teaches him quite a bit about patience, common sense, and resourcefulness. So the the whole sentiment of book should be positive and relaxing.

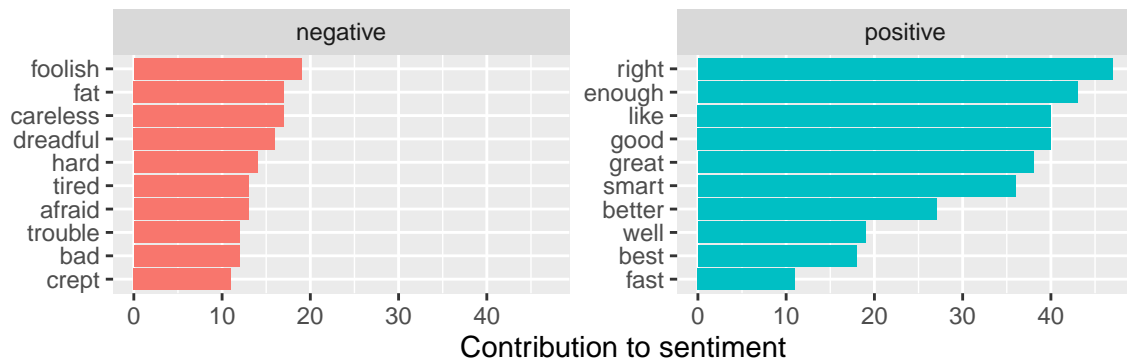


Figure 3: negative and positive words count

I also got word cloud figures, figure 4 shows the frequency of words appeared in book, the two main characters' name is the biggest part of figure. And figure 5 is about sentimental words, which are separate into positive side and negative side.

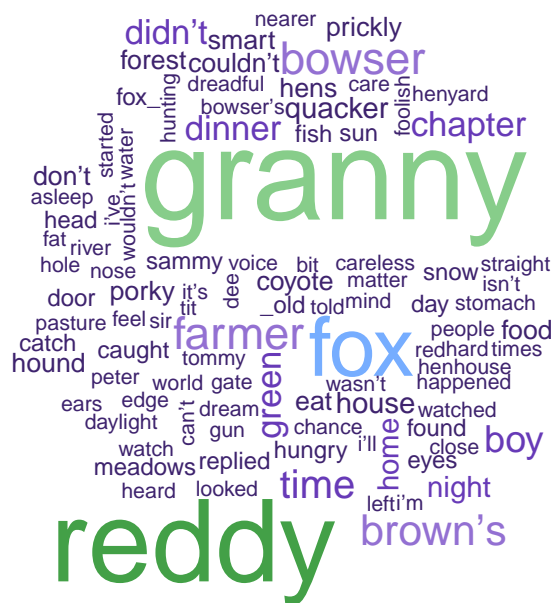


Figure 4: word cloud



Figure 5: sentiment word cloud

Extra credit of Task two

I found another lexicon provided in package ‘textdata’ called ‘Loughran-McDonald’, when I used this new lexicon to get similar figure and result as before.

