MA615 Assignment4

Text Analysis of Pierre and Jean Task Two

Yuyang Li

2021/12/6

Task one : Pick one book

The book I selected called Old Fox Granny written by Thornton W. Burgess from Gutenburg. 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 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.

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

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.

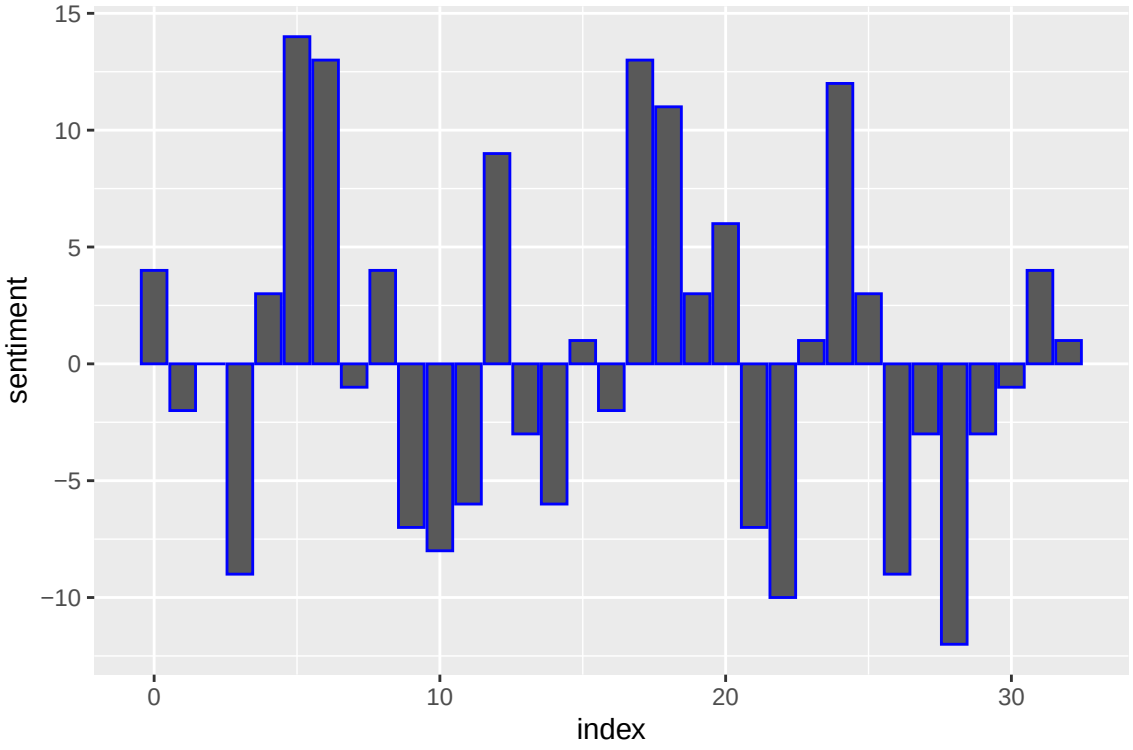


Figure 1: sentiment plot using bing

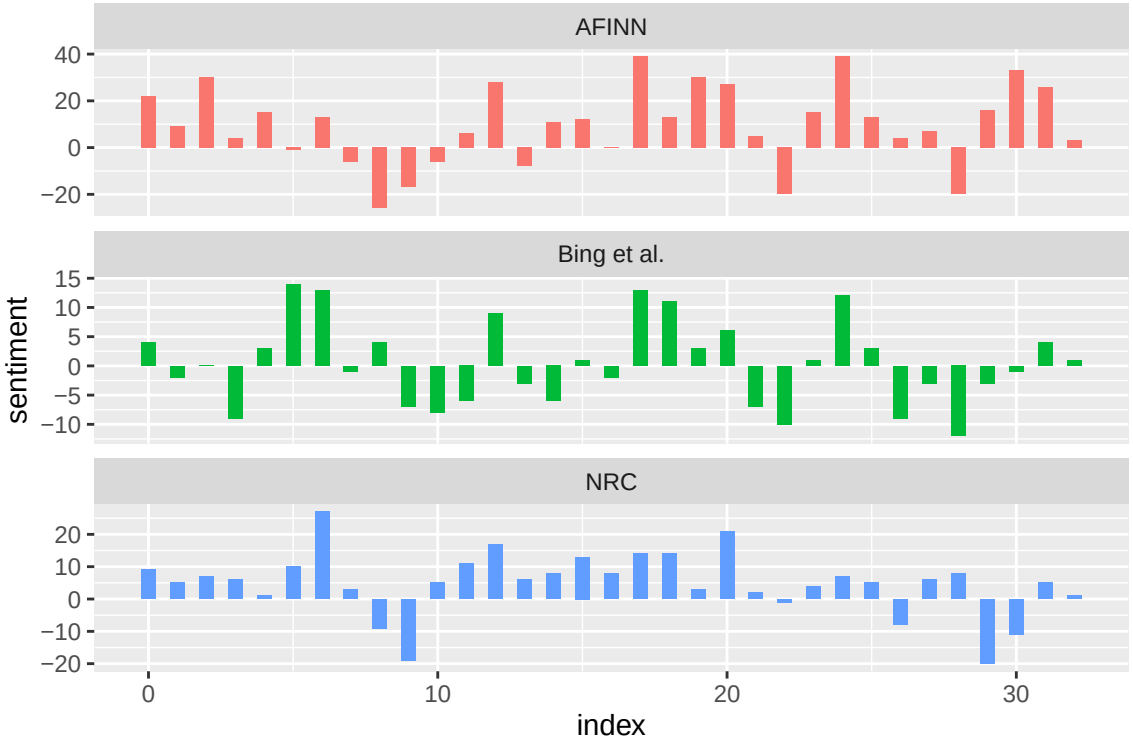
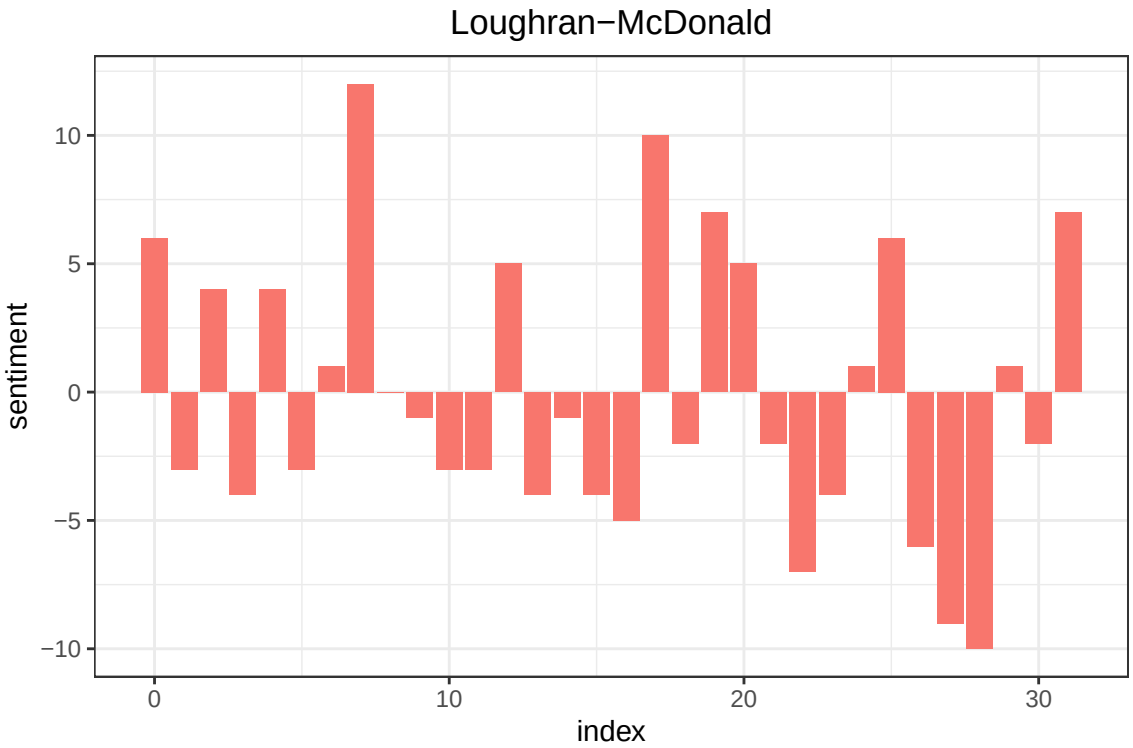


Figure 2: sentiment plot

Figure 3: negative and positive words count

Figure 4: word cloud

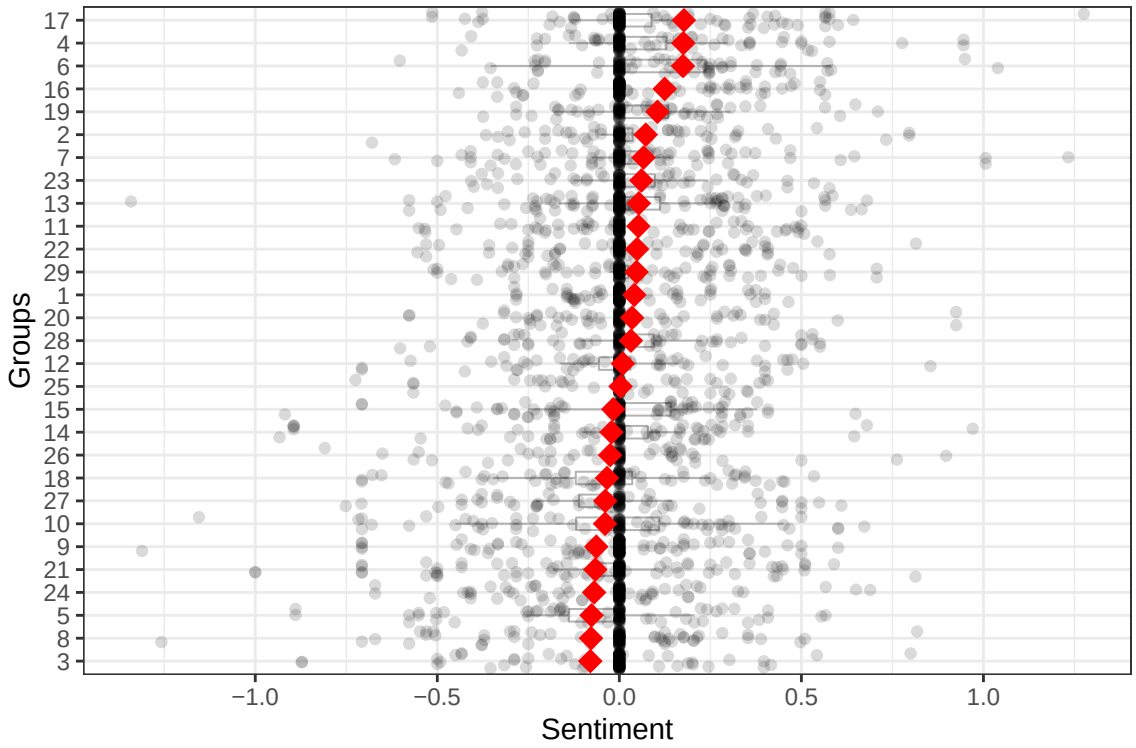
Figure 5: sentiment word cloud



Task three

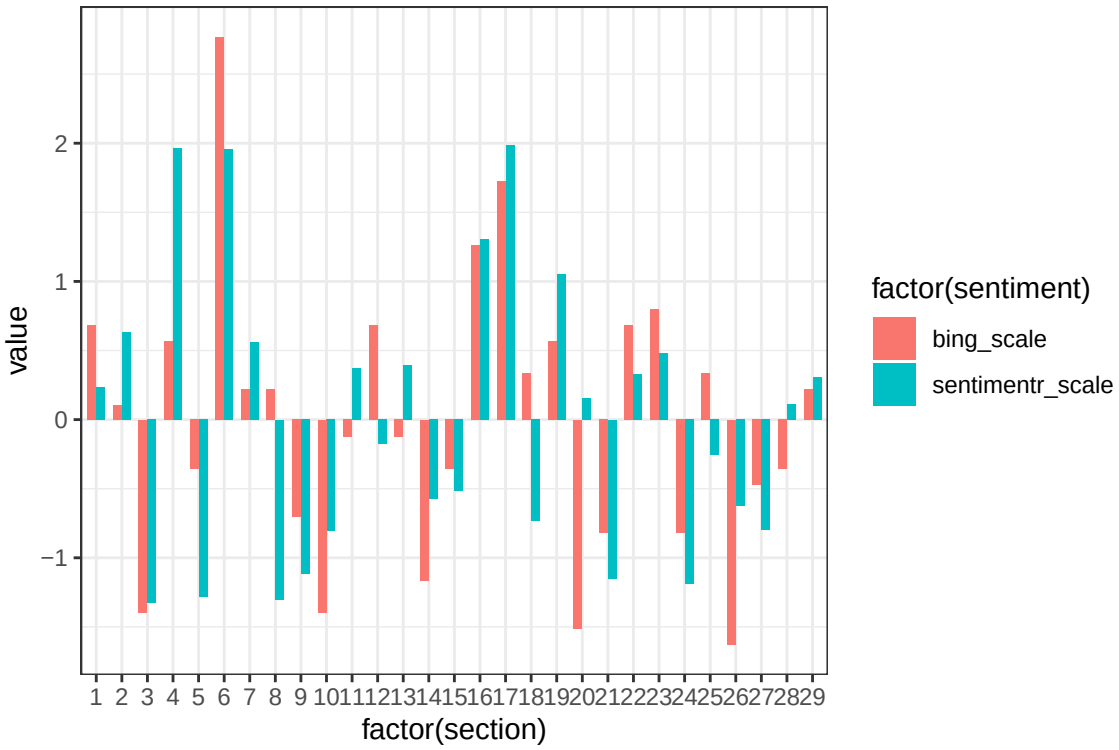
First of all, I change my book into tnum form, and the content of my tnum database is shown as follow:

Secondly I use sentimentr to get sentiment score group by these scores with section to get the average result. The plot sort the average sentiment score from high to low.



Compare the analysis result with Task two

Because the scale of scores of two methods are different, so I need to change them into one consistent scale using function, then I get the bar plot for two scores.



Extra credit of Task three: character analysis

As I said in Task two, there are two main characters in this book, one is Granny Fox and other is Reddy. So I select them from my book and count their appeared frequency in each chapter. seeing the table I found in most of the chapter Granny appears more than Reddy.

section	granny	reddy
1	9	8
2	14	13
3	16	15
4	9	1
5	14	11
6	19	3
7	14	0
8	12	0
9	14	18
10	16	14
11	1	0
12	7	10
13	18	11
14	3	17
15	5	16
16	16	15
17	17	11
18	10	12
19	8	9
20	18	8

section	granny	reddy
21	14	13
22	12	13
23	7	6
24	14	13
25	14	11
26	6	10
27	8	8
28	8	7
29	12	17

section	paragraph	both_appear
1	3	1
1	5	1
2	3	1
2	5	1
2	6	2
2	8	1
2	11	1
2	13	1
3	3	1
3	8	1
3	16	1
3	17	1
3	21	1
5	7	1
6	8	2
9	3	1
9	6	1
9	9	1
9	21	2
10	6	2
10	8	1
12	3	1
12	7	2
12	8	1
13	6	1
13	13	1
13	14	1
13	16	1
13	18	2
14	6	1
15	14	1
16	8	1
16	10	1
16	13	1
16	14	1
17	3	1
17	5	2
17	10	1
17	15	1

section	paragraph	both_appear
18	5	1
18	7	1
18	9	1
18	14	1
19	5	1
19	6	2
19	7	2
19	9	1
19	11	1
20	5	1
20	6	2
20	7	1
20	8	1
21	1	1
21	5	1
21	6	1
21	8	1
21	9	1
22	18	1
22	20	1
22	21	1
23	6	1
23	8	1
23	12	1
24	5	1
24	12	1
24	16	1
25	5	2
25	6	1
25	7	1
25	8	1
25	14	2
25	17	1
26	5	1
26	7	1
26	10	1
26	12	1
27	10	1
27	11	1
29	7	1
29	13	1
29	16	1
29	17	1
29	20	1