Text analysis task3

Daniel(Chen) Xu BUID:U49903384

12/07/2021

Task 1 Pick a book

I choose `Twenty Thousand Leagues under the Sea` written by Jules Verne as the book for my text analysis assignment.

I use three types of sentiment analysis methods AFINN, Bing and NRC to plot barchart plots to compare these methods. From the graph below, the AFINN and Bing method fits better. From the sentiment analysis plot, we can see that scores for positive/negative sentiment in `Twenty Thousand Leagues under the Sea` seems balanced throughout the whole book.

Twenty Thousand Leagues Under the Sea tells the story of marine biologist Pierre Aronnax, his manservant Conseil and harpoonist Ned Land, who – after joining the hunt for a mysterious sea monster – are thrown overboard when the monster attacks and find themselves prisoners of Captain Nemo, probably one of Verne's most memorable yet elusive characters. On board the Nautilus, a technologically advanced submarine that everyone has mistaken for a sea monster, the three companions get to experience the vast and endlessly fascinating world under the sea. In 1866, a monster suspected of being a narwhal was discovered on the sea, and Professor Aronnax and his servant Conseil were invited to join the hunt. In pursuit, they and Their harpooner Ned Land fell overboard and landed on the monster's back. They discovered that the monster was not a narwhal, but a strangely constructed submarine. Nemo secretly built the submarine on a deserted island in the middle of the ocean. It was sturdy and powered by seawater. Captain Nemo invited Aronnax on a voyage under the sea. They set out from the Pacific, passed coral islands, the Indian Ocean, the Red Sea, the Mediterranean Sea, the Atlantic Ocean, and saw many rare plants and animals and strange sights in the sea. On the way, I also experienced a lot of dangerous situations, such as stranding, aboriginal siege, shark fighting, iceberg blocking, octopus attack and so on. Finally, when the submarine reached the Norwegian coast, the three men left without saying goodbye and returned to his hometown.

That is to say, at the beginning of the book, the sentiment of the book is negative, but soon it converts into positive sentiment. And converts to be negative when goes to the end of the book. 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.

task 3 sentence-level analysis

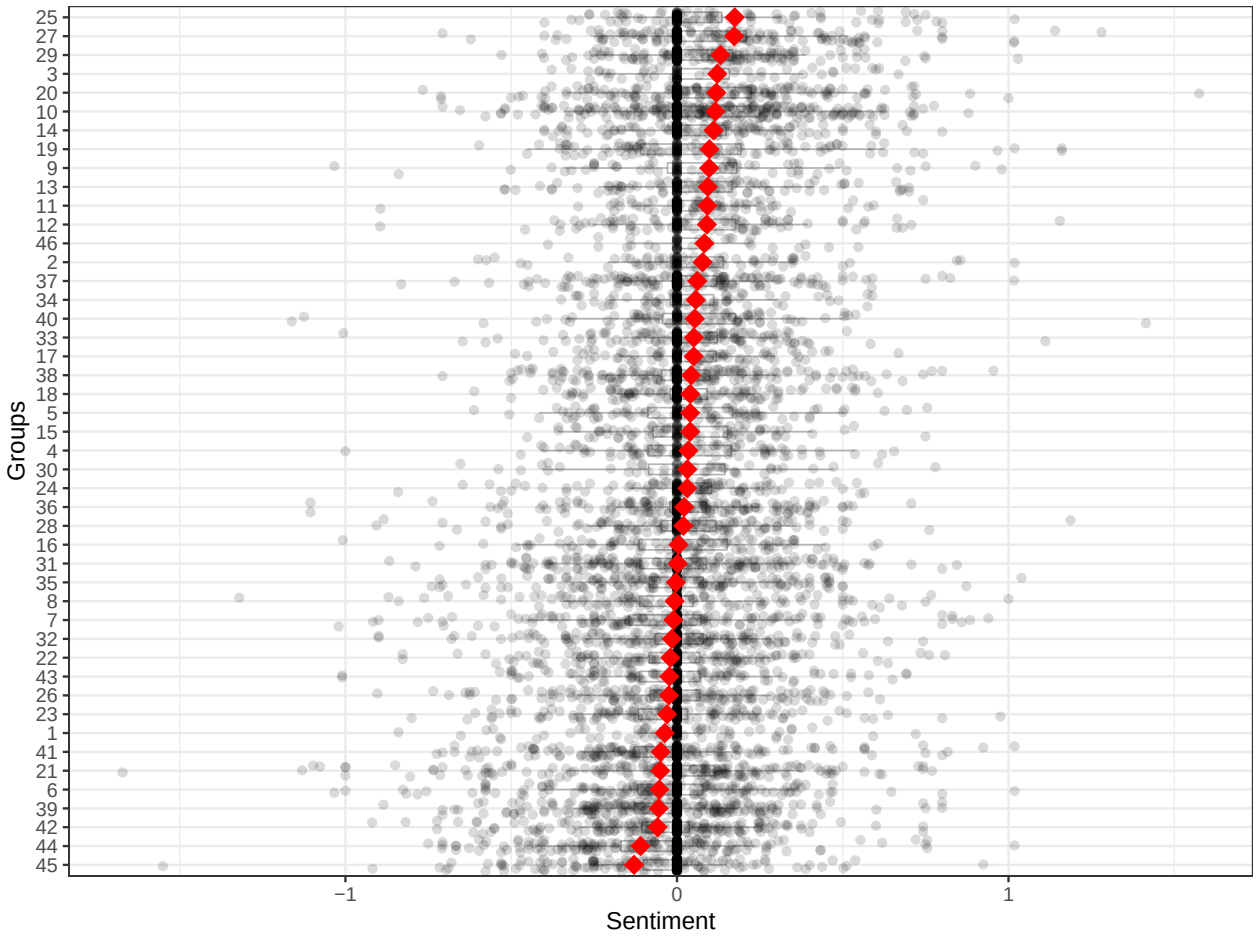
tnum

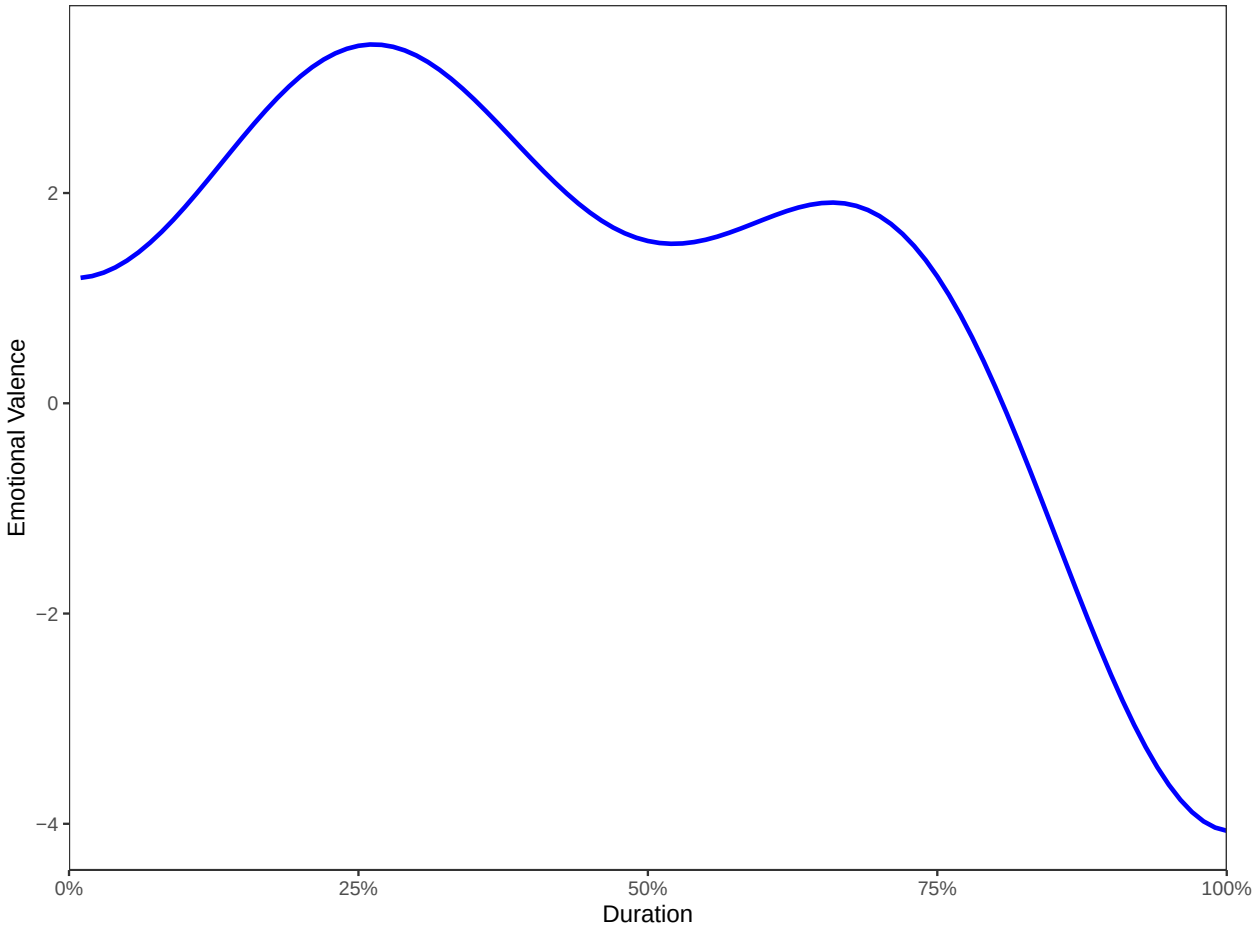
First, I put the book into tnum, the following table shows evidence of my tnum database.

```
##   [1] ""
##   [2] "wells/hw_time"
##   [3] "wells/hw_time_1"
##   [4] "Twain/Tom_Sawyer"
##   [5] "twain/tom_sawyer"
##   [6] "cervantes/don_quixote"
##   [7] "Dickens/david_copperfield"
##   [8] "dickens/david_copperfield"
```

```
##   [9] "well/hw_time_1"
##  [10] "hw_time_1/heading"
##  [11] "hw_time_1/section"
##  [12] "hw_time_2/heading"
##  [13] "hw_time_2/section"
##  [14] "Alcott/Little_women"
##  [15] "alcott/little_women"
##  [16] "Kafka/Metamorphosis"
##  [17] "kafka/metamorphosis"
##  [18] "bronte/wuthering_heights"
##  [19] "twain/wz_huck_1"
##  [20] "Thomas"
##  [21] "thomas"
##  [22] "Thomas/Tess"
##  [23] "thomas/tess"
##  [24] "wells/hw5"
##  [25] "wells5/hw5"
##  [26] "Kipling/junglebook"
##  [27] "kipling/junglebook"
##  [28] "cervantes/don_quixote_v2"
##  [29] "jane_eyre/heading"
##  [30] "jane_eyre/section"
##  [31] "wells6/hw6"
##  [32] "wells7/hw7"
##  [33] "wells8/hw8"
##  [34] "wells9/hw9"
##  [35] "Twain/Tom_Sawyer_TEST"
##  [36] "twain/tom_sawyer_test"
##  [37] "barrie/Assignment4"
##  [38] "barrie/assignment4"
##  [39] "wilde/section"
##  [40] "gatsby/hw_1"
##  [41] "zweig/test1"
##  [42] "tnum_pp/assignment4"
##  [43] "randb/heading"
##  [44] "randb/section"
##  [45] "cervantes/don_quixote_v3"
##  [46] "cervantes/don_quixote_v4"
##  [47] "Zweig/test1"
##  [48] "Twain/Tom_Sawyer2"
##  [49] "twain/tom_sawyer2"
##  [50] "Zweig/test2"
##  [51] "zweig/test2"
##  [52] "glasser/the_martian"
##  [53] "wells10/hw10"
##  [54] "wells11/hw11"
##  [55] "wells12/hw12"
##  [56] "hssb/book"
##  [57] "hssb/book1"
##  [58] "glasser/the_martian_1"
##  [59] "Glasser/the_martian_1"
##  [60] "jv_ttluts/textanalysis"
##  [61] "Austen/persuation"
##  [62] "austen/persuation"
```

```
##  [63] "hvsb/book3"
##  [64] "hvdsb/book4"
##  [65] "MachiavelliPrince/heading"
##  [66] "machiavelliprince/section"
##  [67] "Thomas/Tess_new"
##  [68] "thomas/tess_new"
##  [69] "Machiavelli_Prince/heading"
##  [70] "machiavelli_prince/section"
##  [71] "jv_ttluts/textanalysis1"
##  [72] "Jane_Austen/persuation"
##  [73] "jane_austen/persuation"
##  [74] "anton/hw_russian"
##  [75] "william/test2"
##  [76] "howard/test2"
##  [77] "homer/the_odyssey"
##  [78] "Homer/The_Odyssey"
##  [79] "gdm/test2"
##  [80] "xihao"
##  [81] "je/heading"
##  [82] "je/section"
##  [83] "Montgomery/Anne"
##  [84] "montgomery/anne"
##  [85] "ballantyne/doggie"
##  [86] "yuanming"
##  [87] "Thomas/Tess_new2"
##  [88] "thomas/tess_new2"
##  [89] "johanna/heidi"
##  [90] "Johanna/Heidi"
##  [91] "xihao/hw4_v2"
##  [92] "ballantyner/doggieandi"
##  [93] "BallantyneR/DoggieandI"
##  [94] "yuanming/happyprince"
##  [95] "holmes/hw4"
##  [96] "charlse_dickens/a_christmas_carol"
##  [97] "Charlse_Dickens/A_Christmas_Carol_2"
##  [98] "charlse_dickens/a_christmas_carol_2"
##  [99] "Charlse_Dickens/A_Christmas_Carol_3"
## [100] "charlse_dickens/a_christmas_carol_3"
## [101] "holmes/hw4_2"
## [102] "love_paddington/heading"
## [103] "love_paddington/section"
## [104] "twain/huck"
## [105] "holmes/holmes"
## [106] "holmes/hound"
## [107] "mobydick/zby"
## [108] "Maupassant/Pierre_Jeans"
## [109] "maupassant/pierre_jeans"
## [110] "peter_pan/assignment4"
## [111] "Charlse_Dickens/A_Christmas_Carol_4"
## [112] "charlse_dickens/a_christmas_carol_4"
## [113] "Alexandre_Dumas/Camille_Book92"
## [114] "alexandre_dumas/camille_book92"
```


Groups
25
27
29
3
20
10
14
19
9
13
11
12
46
2
37
34
40
33
17
38
18
5
15
4
30
24
36
28
16
31
35
8
7
32
22
43
26
23
1
41
21
6
39
42
44
45
−1
0
1
Sentiment



Compare this analysis with the analysis you did in Task TWO

It is difficult to directly compare Sentimentr and Bing's score. Therefore, I apply `scale` function to standardize two sentiment values generated from two tasks into the same criteria. Then I use ggplot to plot barchart plot. From the Figure below, we can see that the trends that positive and negetive directions are approximately similar. But the exact number differs from two methods. Generally, as we can see from the figure blow, Bing lexicons text analysis method is more optimistic than sentimentr package.

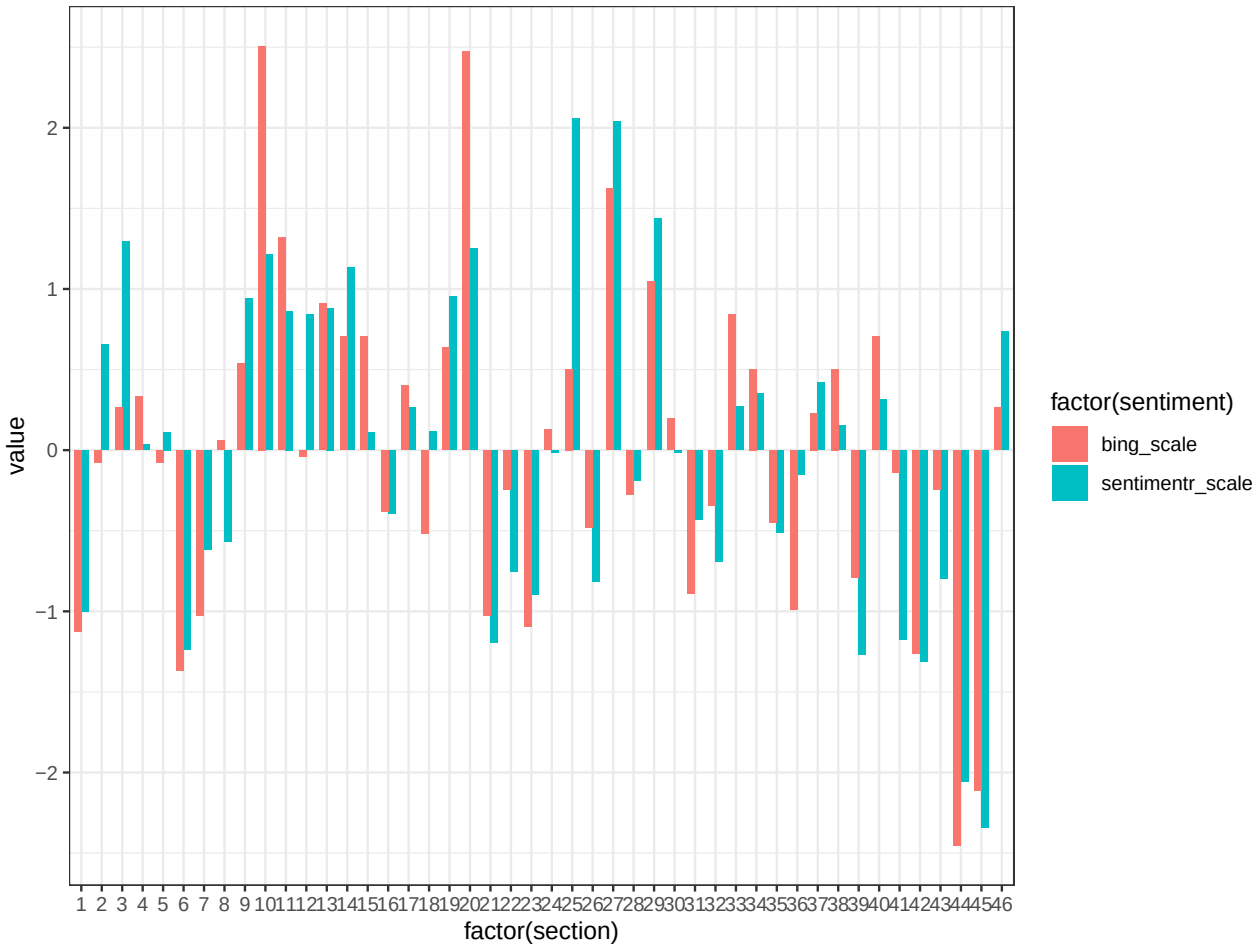


Figure 1: sentiment comparison