

Text Mining

Yuanming LENG

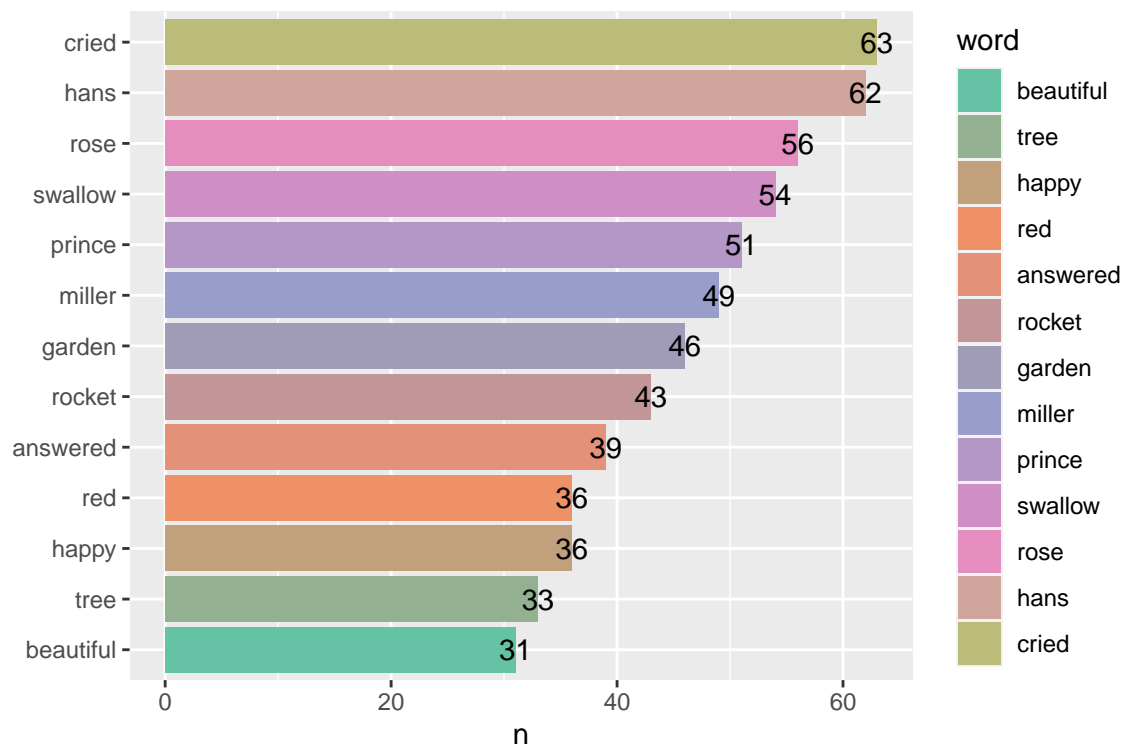
12/6/2021

Task one

The Happy Prince and Other Tales (or Stories) is a collection of stories for children by Oscar Wilde first published in May 1888. It contains five stories: “The Happy Prince”, “The Nightingale and the Rose”, “The Selfish Giant”, “The Devoted Friend”, and “The Remarkable Rocket”. And the five stories in this book are unrelated. This five stories are very meaningful for children.

Task two

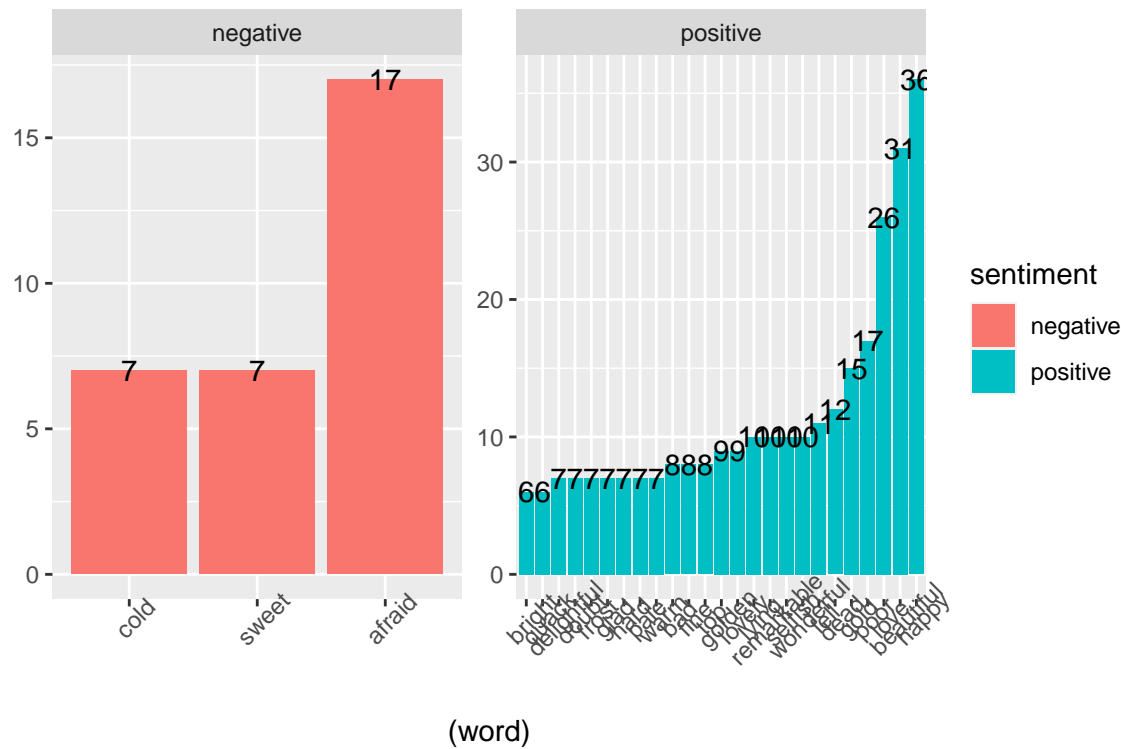
Cuting words



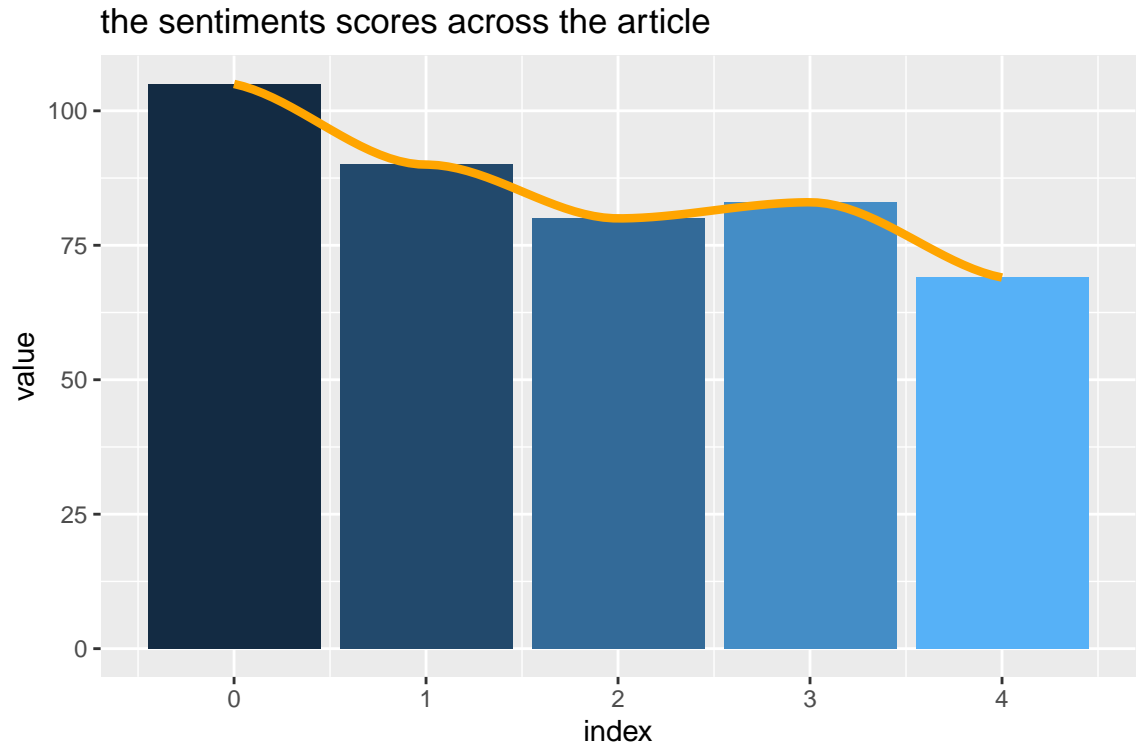
According to the plot, we choose words count over 30 and we found those words covered for all five stories. For example, *tree* for *The Selfish Giant*, *prince* and *swallow* for *The Happy Prince*, *rocket* for *The Remarkable Rocket*, *rose* for *The Nightingale and the Rose*.

Using NRC lexicon to analysis the book, then according to the plot, this book contains more positive words.

Lexicon : **bing**

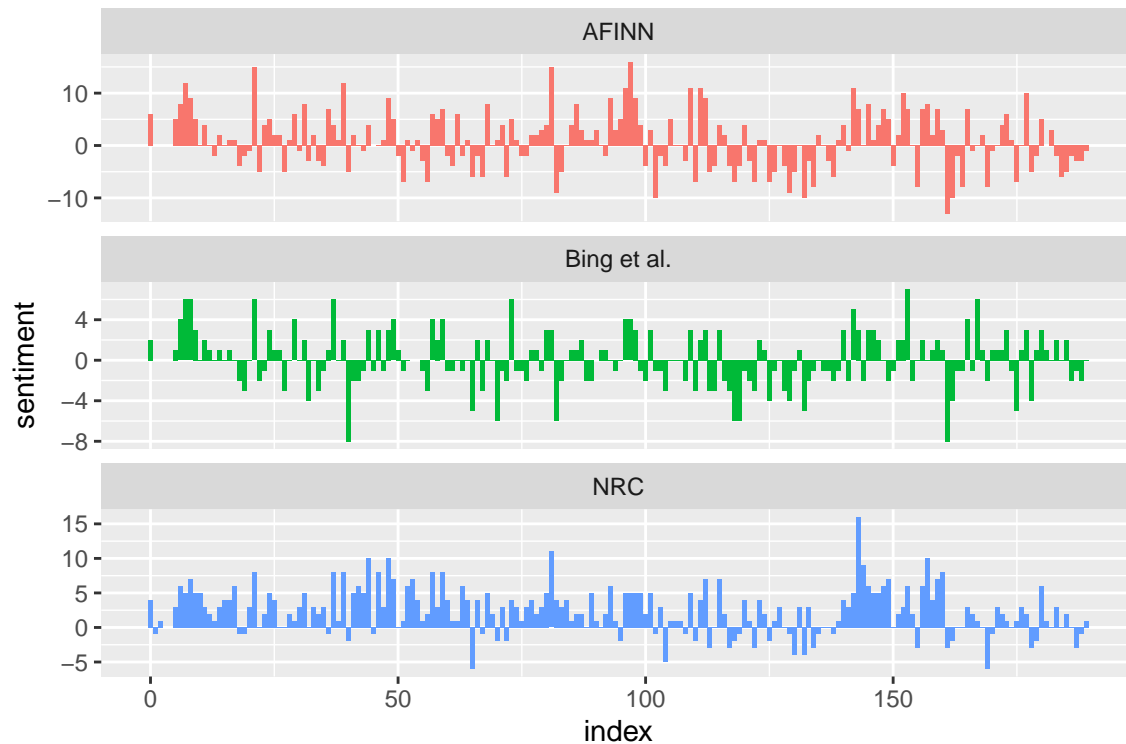


According to the plot, the book talks more positive words when analysis it in the NRC lexicon.



Lexicon: AFINN

Comparison for three lexicons



The plot shows that in different lexicons, the book has different proportions for the positive and negative parts. Since there are different summarize methods for different lexicons. In the ‘afinn’ lexicon, the value calculates by sum each word’s value for every ten lines, and it may conclude a broader range. In the ‘NRC’ and ‘bing’ lexicon, they label each word. ‘bing’ marks each word as positive or negative, while ‘NRC’ will label of brutal ways and we only choose words which are marked as positive or negative, which means we will drop many values. In the NRC lexicon, we have more positive values. But with the ‘bing’ lexicon, we have more negative values.

Also, we should consider that this book talks about five fairy tales, and five tales are independent. Five different stories have different characters, plots, and sentiments, so it is hard to say whether the figure matches the plot. More reliable analysis should plot the sentiment for five parts of this book, respectively.

Reference

- [1] <https://www.r-bloggers.com/2020/04/sentiment-analysis-in-r-with-sentimentr-that-handles-negation-valence-shifters/>
- [2] <https://github.com/MA615-XihaoCao/HW4>
- [3] <https://www.tidytextmining.com/sentiment.html>
- [4] https://en.wikipedia.org/wiki/The_Happy_Prince_and_Other_Tales#The_Selfish_Giant