

MA615 Assignment4

Text Analysis of Pierre and Jean

Yanbing Chen

2021/12/6

Task One: Pick a book

According to the task one requirement, I pick a fiction book named Pierre and Jean written by Guy de Maupassant on The Gutenberg Project to do text analysis. The book size is 284kB.

Task Two: Bag of words analysis

In this part, I tidy information in the book to gain a clean data file called 'tidy_pierre'. Because there are many options for sentiment lexicons, I try three methods that are AFINN, BING, and NRC to make box plot to find which measure is the best. As figure1 shown, the BING fits better, therefor, finally I chose to apply NRC method in deeper text analysis.

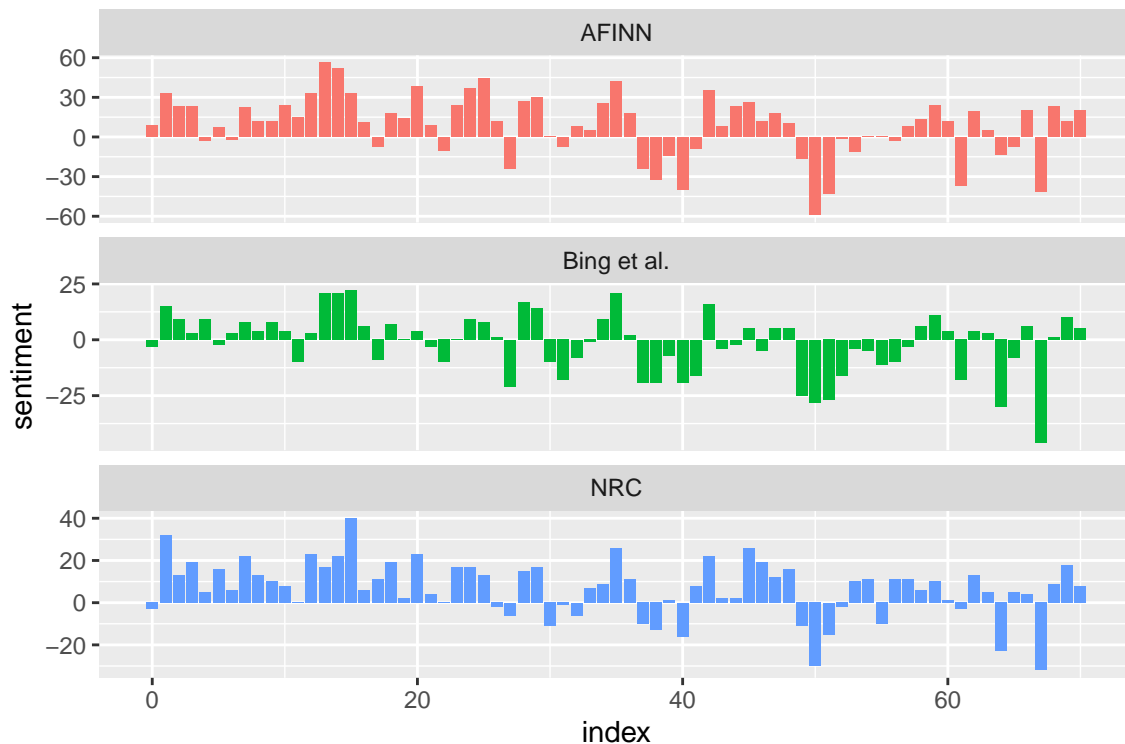


Figure 1: sentiment plot

