

# Task 1 & Task 2

Yujia Wang

12/8/2021

## Task 1: Pick a book

Alice's Adventures in Wonderland is a science fiction novel that I loved to read when I was young. So I chose this book to commemorate my lost childhood.

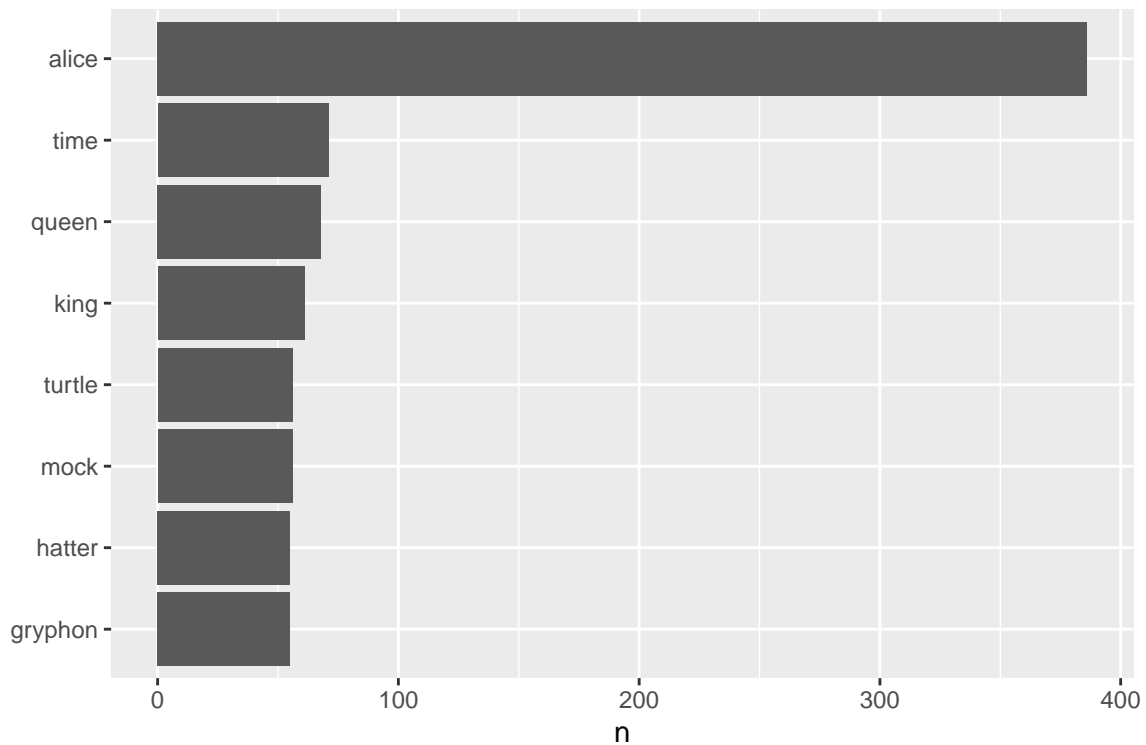


Figure 1: xxx

I calculated the eight most frequent words in the book, and the results are as follows. It can be found that because Alice is the protagonist, the frequency of using her name is much higher than other words.

At the same time, there are 5 words that are the names of characters, such as Queen, King, Turtle, Hatter and Gryphon, which shows that they are the main supporting roles of the book.

Time is the second most frequent word in this book, reflecting the urgency of Alice's adventure in her dream.

## TASK 2: bag of word analysis

## # A tibble: 2,477 x 2

```
##      word      value
##      <chr>      <dbl>
## 1 abandon      -2
## 2 abandoned     -2
## 3 abandons      -2
## 4 abducted      -2
## 5 abduction     -2
## 6 abductions    -2
## 7 abhor         -3
## 8 abhorred      -3
## 9 abhorrent     -3
## 10 abhors       -3
## # ... with 2,467 more rows
```

```
## # A tibble: 6,786 x 2
##      word      sentiment
##      <chr>      <chr>
## 1 2-faces      negative
## 2 abnormal     negative
## 3 abolish      negative
## 4 abominable   negative
## 5 abominably   negative
## 6 abominate    negative
## 7 abomination  negative
## 8 abort        negative
## 9 aborted      negative
## 10 aborts      negative
## # ... with 6,776 more rows
```

```
## # A tibble: 13,875 x 2
##      word      sentiment
##      <chr>      <chr>
## 1 abacus      trust
## 2 abandon     fear
## 3 abandon     negative
## 4 abandon     sadness
## 5 abandoned   anger
## 6 abandoned   fear
## 7 abandoned   negative
## 8 abandoned   sadness
## 9 abandonment anger
## 10 abandonment fear
## # ... with 13,865 more rows
```

```
## # A tibble: 88 x 2
##      word      n
##      <chr>    <int>
## 1 found      32
## 2 garden     15
## 3 baby       14
## 4 beautiful  13
## 5 dance      13
## 6 grow       13
```

```
## 7 deal      12
## 8 child     11
## 9 glad      11
## 10 tree      9
## # ... with 78 more rows
```

Based on the plot trajectory of Alice's Adventures in Wonderland, I calculated the emotional score. The x-axis tracks the narrative time in the text part, and the y-axis is the difference between positive and negative emotions.

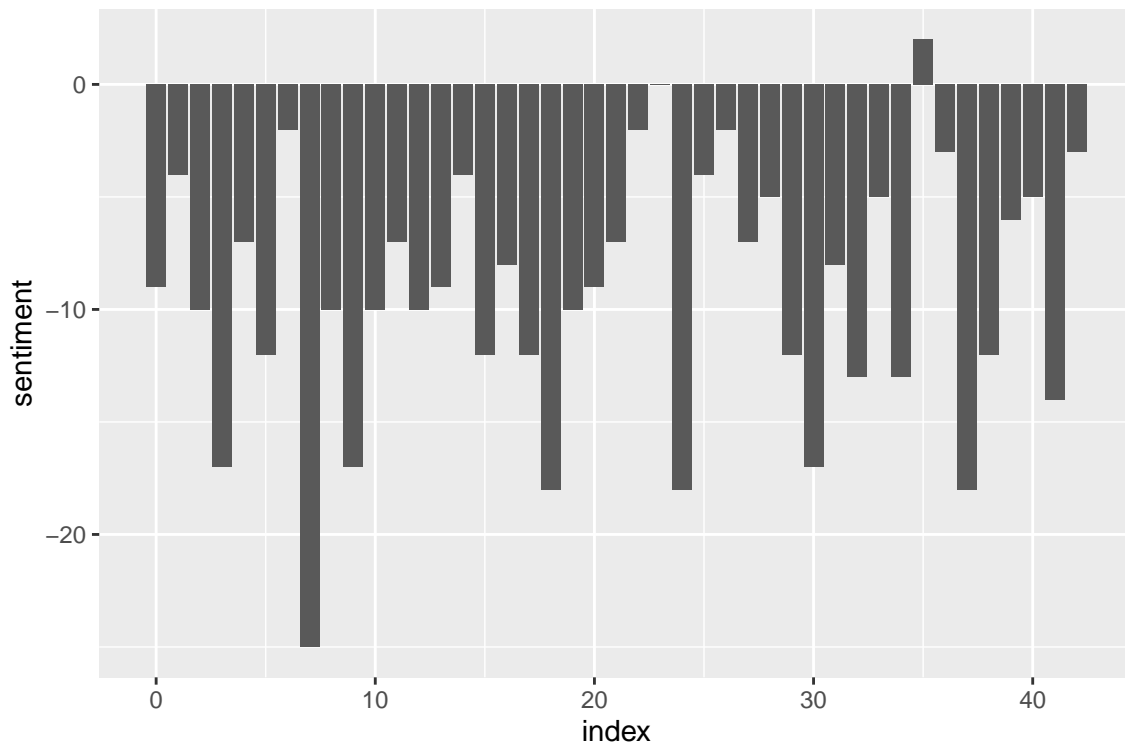


Figure 2: xxx

It can be seen that almost all the plots in this book are negative emotions. This result coincides with the book's absurd, bizarre fictional scenes, weird characters, and thrilling adventure stories.

### comparing the three sentiment dictionaries

According to text mining with R, I imported AFINN dictionary, Bing dictionary and NRC dictionary. It can be seen that the results of AFINN and Bing tend to be consistent, but the results of NRC are more positive, and the result of AFINN has more variance.

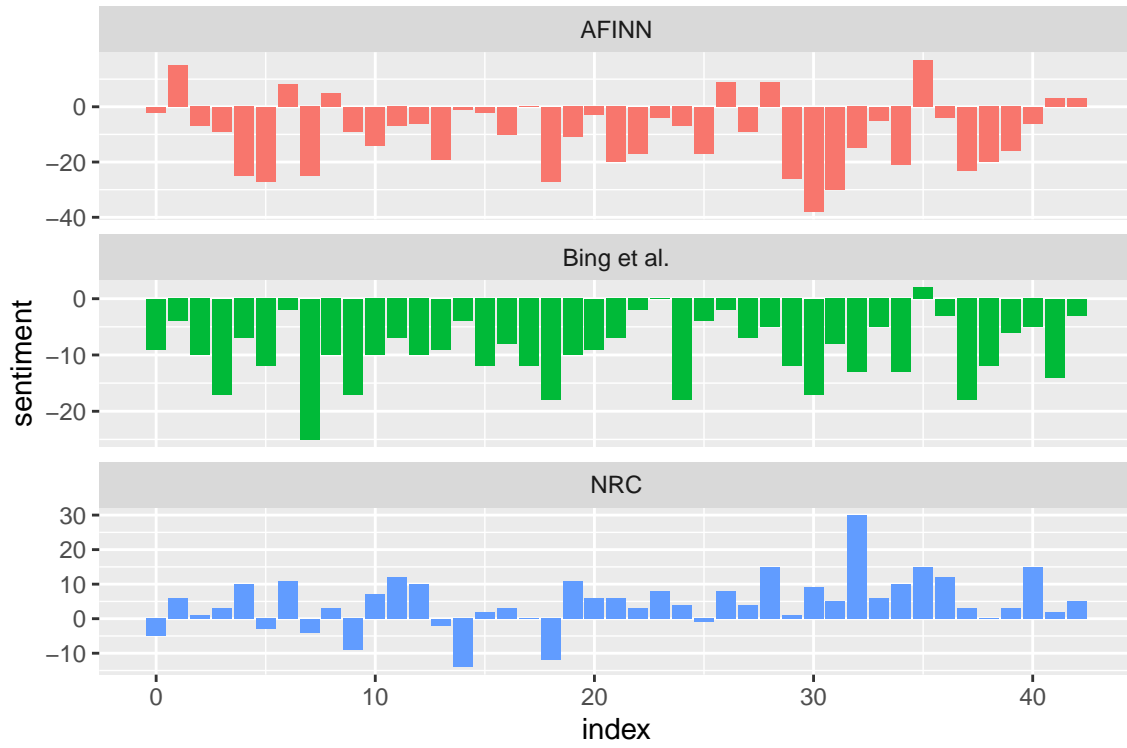


Figure 3: xxx

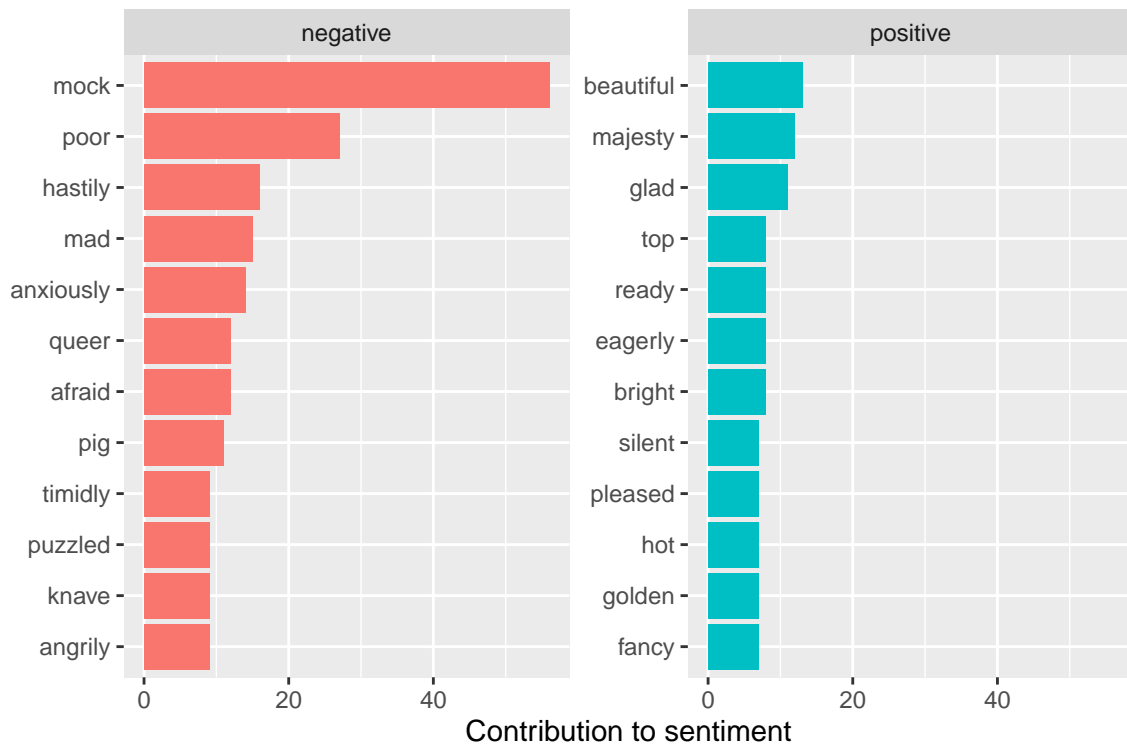


Figure 4: xxx

## word cloud

The plot above shows the 12 most frequently positive or negative words.



Figure 5: sentiment word cloud

Subsequently, I drew two word cloud plots of frequency. This is the visualization of the previous plot.

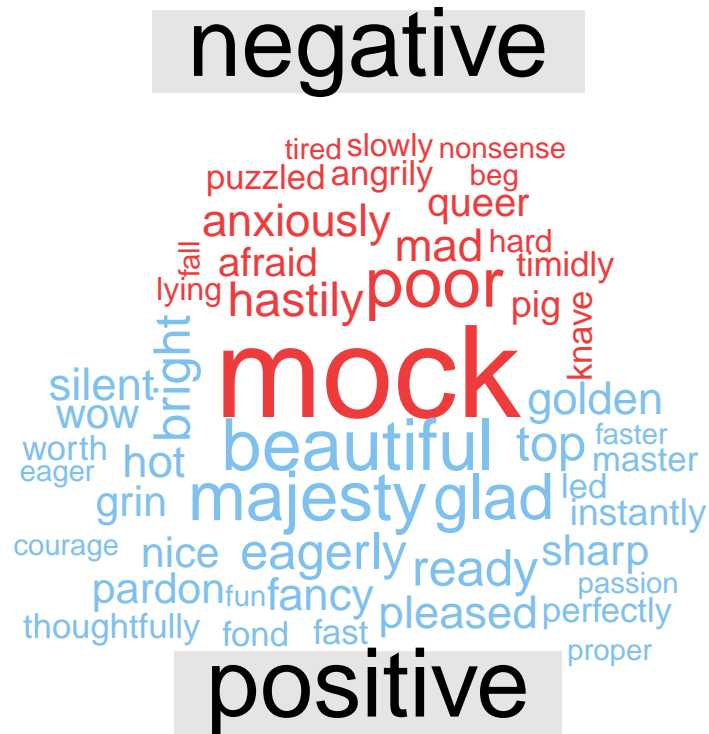


Figure 6: sentiment word cloud

### Extra credit: another lexicon

Coincidentally, I found a new method when the system reported an error, which said “Error in match.arg(lexicon) : ‘arg’ should be one of “bing”, “afinn”, “loughran”, “nrc””.

### comparing the four sentiment dictionaries

Finally, I drew a comparison chart of the four methods. I found that loughran is like a neutralized version of AFINN and Bing. Maybe this is a more reliable sentiment analysis package.

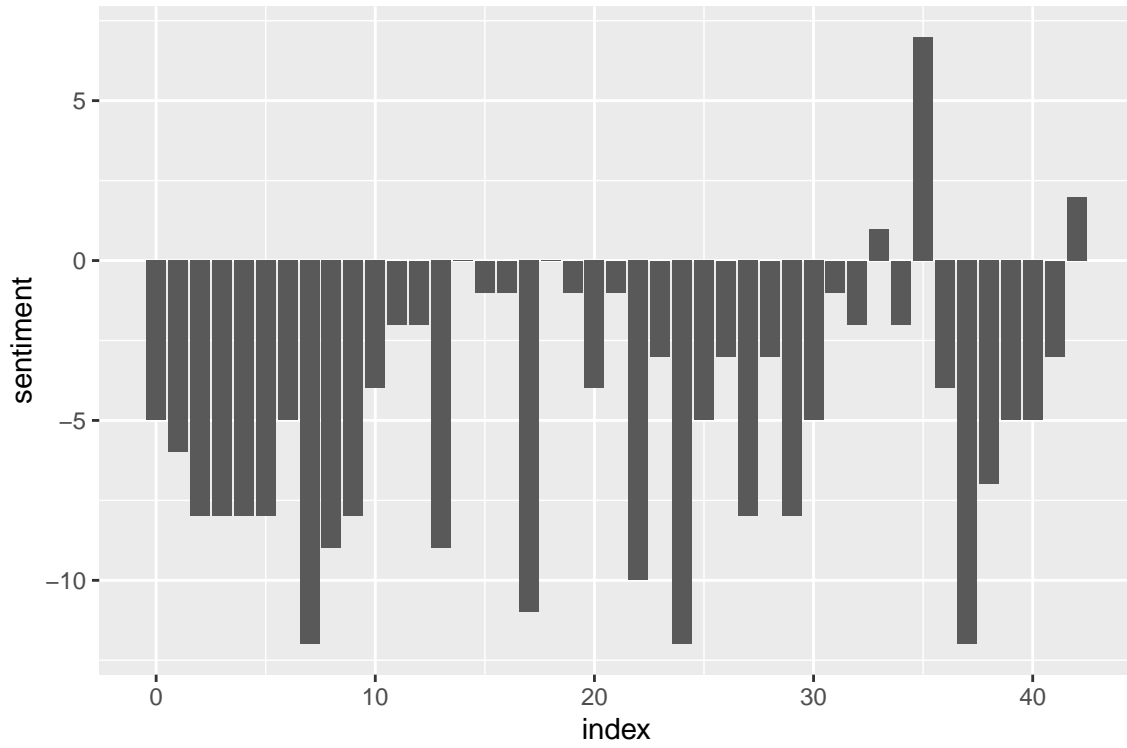


Figure 7: xxx

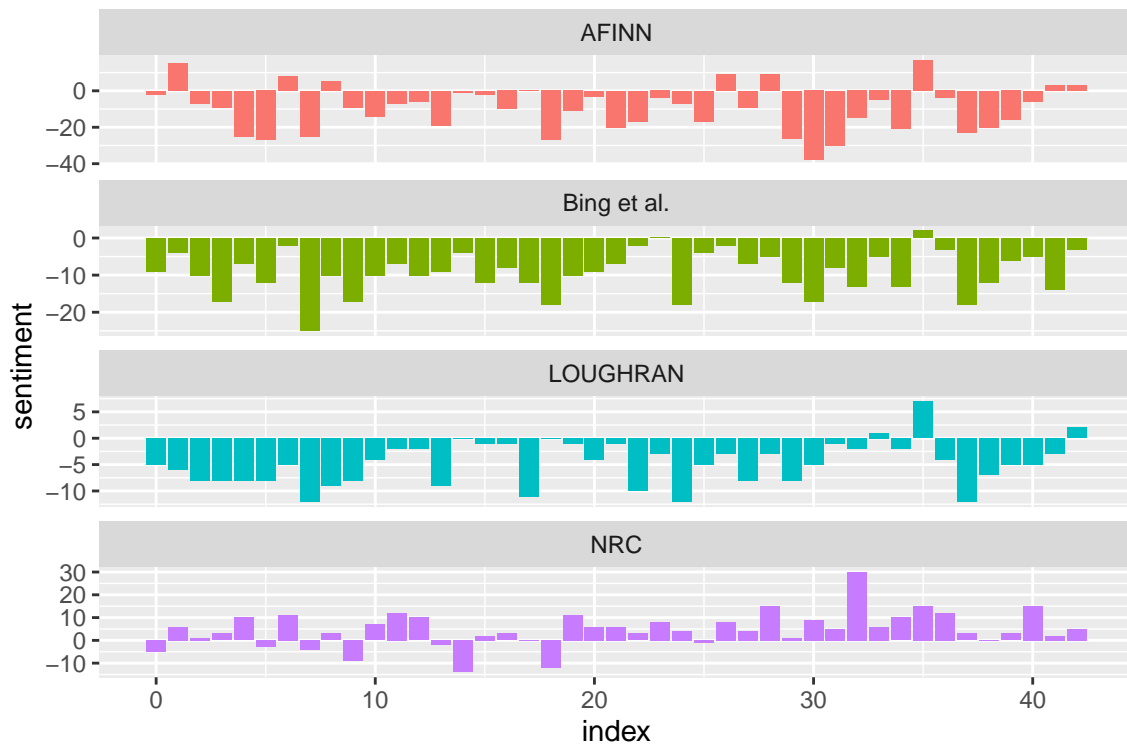


Figure 8: xxx

## Reference

<https://www.gutenberg.org/ebooks/11> <https://www.tidytextmining.com/sentiment.html> <https://www.rdocumentation.org/>  
[https://www.rdocumentation.org/packages/textdata/versions/0.4.1/topics/lexicon\\_loughran](https://www.rdocumentation.org/packages/textdata/versions/0.4.1/topics/lexicon_loughran)