wonderful_task2

Peng Liu

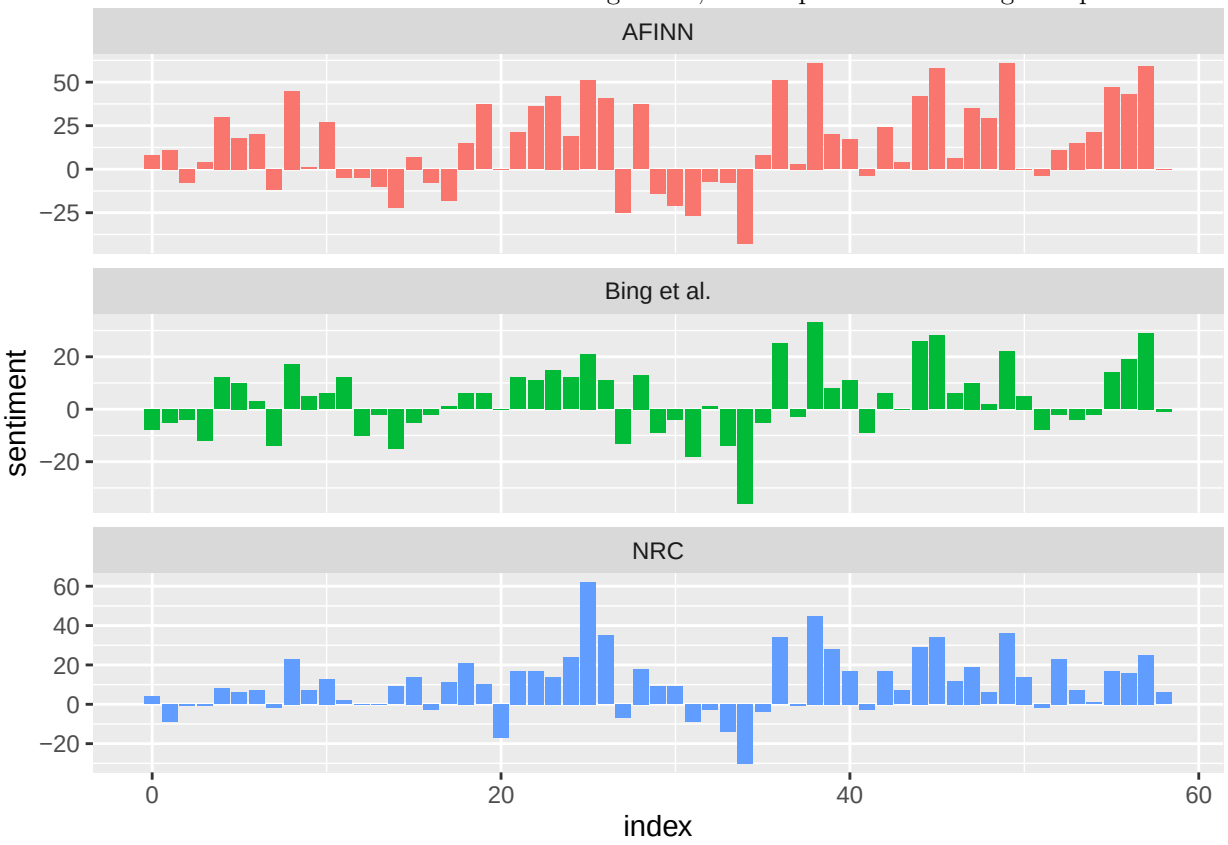
12/9/2021

Task One

I chose Frank Baum's the Wonderful Wizard of Oz as the analysis object of my assignment

Task Two

First, I use inner_join() to analyze the Wizard of Oz using three different sentiment analysis methods: AFINN, Bing, and NRC. Through Figure 1, we can find that the novel's trajectories of these three methods are the same. But among them, NRC produced the highest positive value.




```
## # A tibble: 2 x 2
##   sentiment     n
##   <chr>     <int>
## 1 negative   3318
## 2 positive   2308
```

```
## # A tibble: 2 x 2
##   sentiment     n
##   <chr>     <int>
## 1 negative   4781
## 2 positive   2005
```

Because the ratio of negative words to positive words in the Bing dictionary is higher than that in the NRC dictionary, this will help the effect we see in the image above and any systematic differences in word matching. Therefore, I will use the Bing method to analyze the novel.

From Figure 2, we can find that the word "great" is the most frequently occurring in positive and the most frequently occurring word among all positive and negative words. The second is "good," which appears more often than the first word, "wicked," negatively. This further shows that this novel is more optimistic.

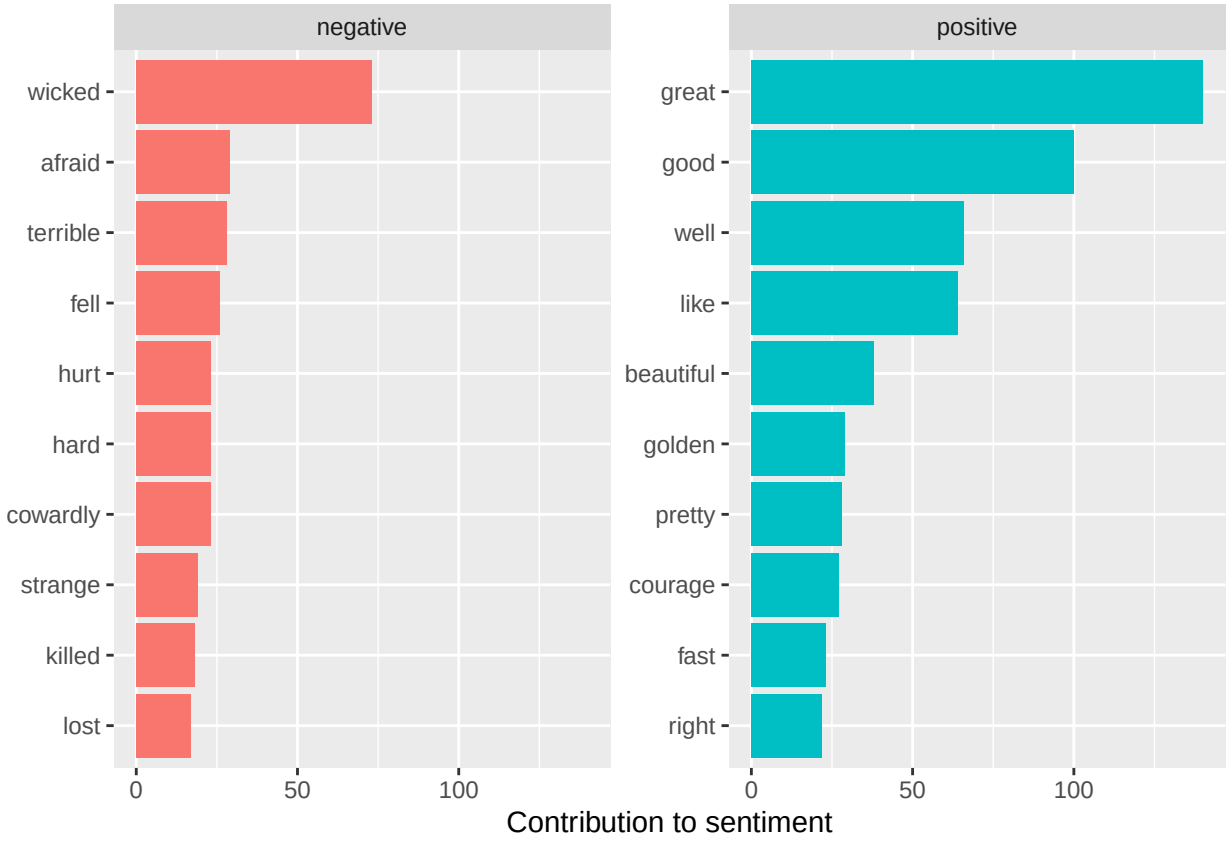


Figure 1: Figure 2: Words that contribute to positive and negative sentiment

Figure 3 and Figure 4 respectively show the word cloud of the frequency of words in the novel and the word cloud of positive and negative words. In Figure 3, the top three in frequency are scarecrow, lion, and oz. Scarecrow and lion are the names of the protagonists in the novel, and oz is the name of the destination they are going to.

Extra Credit

Regarding the extra credit, I chose the Loughran method for analysis. In addition to the visualization of the Loughran dictionary, I also drew the frequency diagrams of six emotions about Loughran: "positive," "negative," "litigation," "uncertain," "constraint" and "redundant."

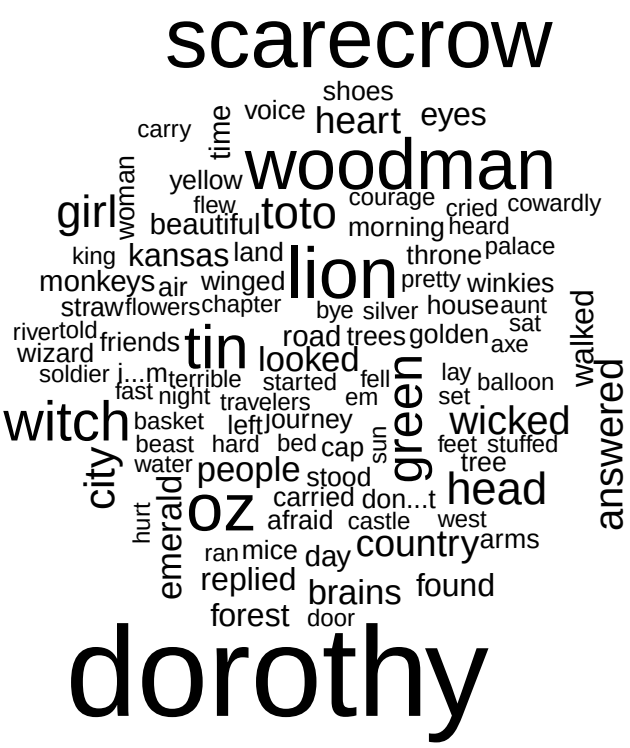
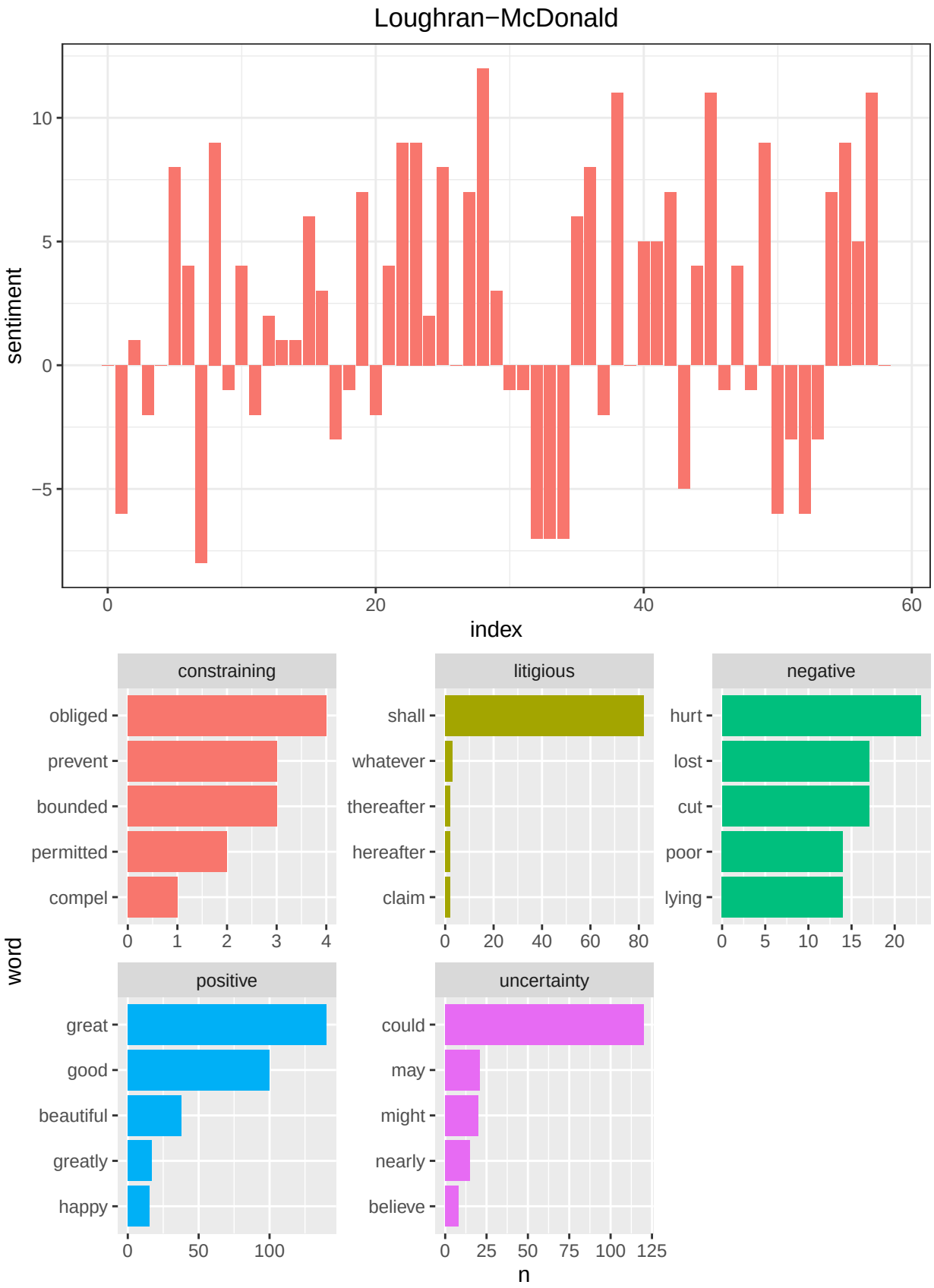


Figure 2: Figure 3: The most common words in the Wonderful Wizard of Oz

Figure 3: Figure 4: Most common positive and negative words in the Wonderful Wizard of Oz


Loughran−McDonald
sentiment
index
10
5
0
−5
0
20
40
60
word
n
constraining
obliged
prevent
bounded
permitted
compel
0
1
2
3
4
litigious
shall
whatever
thereafter
hereafter
claim
0
20
40
60
80
negative
hurt
lost
cut
poor
lying
0
5
10
15
20
positive
great
good
beautiful
greatly
happy
0
50
100
uncertainty
could
may
might
nearly
believe
0
25
50
75
100
125

Reference

Julia Silge and David Robinson, Text Mining with R: A Tidy Approach, (June 8, 2017), or from https://www.tidytextmining.com/

The Comprehensive R Archive Network, textdata:https://cran.r-project.org/web/packages/textdata/textdata.pdf