

# Assignment 4 Task1 & 2

Boyu Chen

12/8/2021

## Task 1

I choose a book "Persuasion" written by Jane Austen. Here is the introduction:

"Persuasion" is the story of a young man and woman who go through a difficult time and end up in love. Anne's youth is fading, and she is in the prime of her life when she finally parted ways with her intended, Colonel Wentworth because she accepted the "persuasion" of others. Although her youth is gone and she suffers from the indifference of her vain father and sister, her noble kindness makes her the most popular person among her family and friends. When the two meet again eight years after their breakup, the Colonel cannot freeze his grudge and pursues others against his will; and Anne comes close to accepting her hypocrite cousin's proposal of marriage. In a series of interactions over family matters, the Colonel discovers that Anne's selflessness and steadfastness are unmatched. Anne finds the courage to love again and again through self-convincing. They find that reuniting is happier than falling in love for the first time, and so, having been tested, they cease to be indifferent and begin to remember and confess their love to the fullest.

```
book<-gutenberg_download(105)  # Download the book through gutenberg

## Determining mirror for Project Gutenberg from http://www.gutenberg.org/robot/harvest

## Using mirror http://aleph.gutenberg.org

newbook <- book %>%              # Identify the linenumber, chapter, and get the each word for each sentence
  mutate(
    linenumber = row_number(),
    chapter = cumsum(str_detect(text,
                                regex("^chapter [\\divxlc]",
                                       ignore_case = TRUE)))) %>%
  unnest_tokens(word, text)

get_sentiments("bing")          # Get the sentiment word

## # 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
```

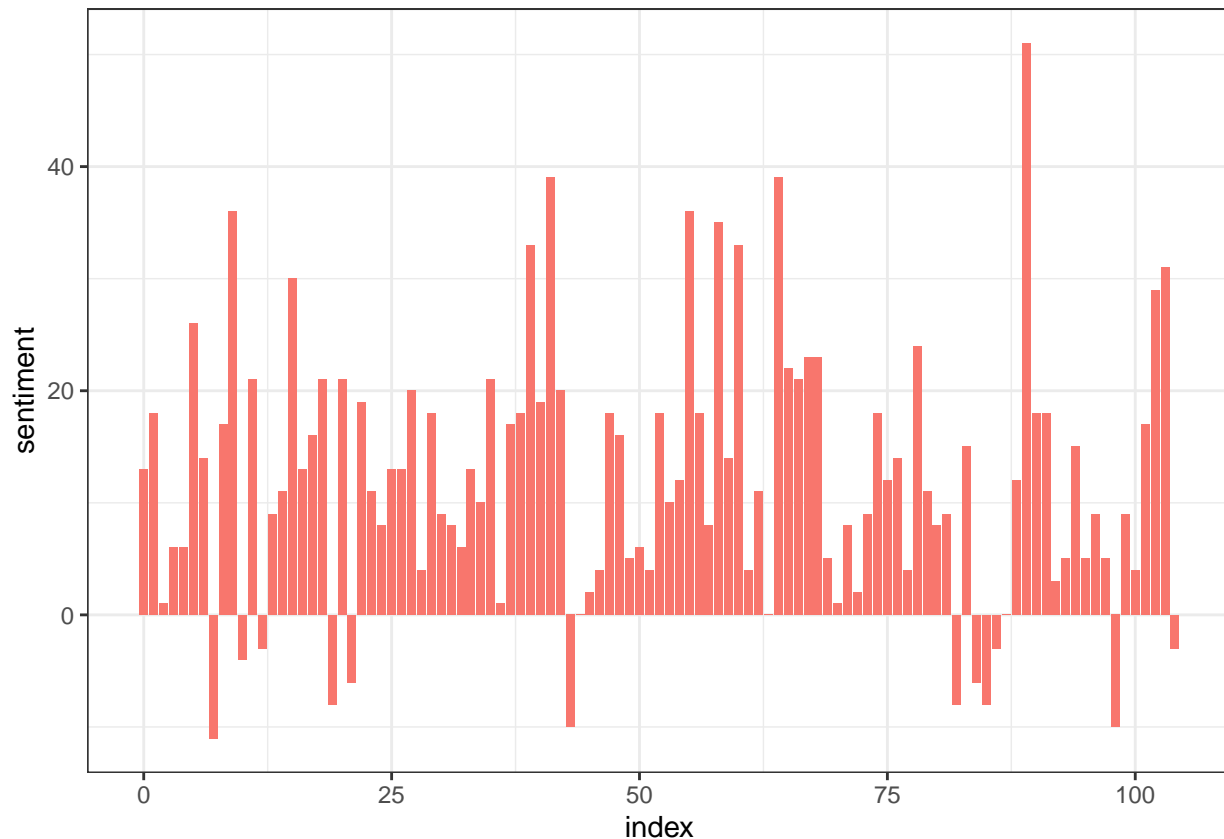
```
bing_neg <- get_sentiments("bing") %>% # Choose the negative sentiment word from bing list
  filter(sentiment == "negative")
```

```
# Do tidy book
tidy_book <- book %>%
  mutate(
    linenumber = row_number(),
    chapter = cumsum(str_detect(text,
                                regex("^chapter [\\divxlc]",
                                       ignore_case = TRUE)))) %>%
  unnest_tokens(word, text)
```

```
# Get the book sentiment word which in bing list
newbook_sentiment <- newbook %>%
  inner_join(get_sentiments("bing")) %>%
  count(index = linenumber %/% 80, sentiment) %>%
  pivot_wider(names_from = sentiment, values_from = n, values_fill = 0) %>%
  mutate(sentiment = positive - negative)
```

```
## Joining, by = "word"
```

```
# Make the plot
ggplot(newbook_sentiment, aes(index, sentiment, fill='red')) +
  geom_col(show.legend = FALSE) + theme_bw()
```



From the plot we can find that at first Anne fall in love with navy captain wintersworth. Which is match the first stage of the plot. However, when they plan to marriage, because the other party has no money and no power, family members are strongly opposed to the marriage, especially Anne's godmother, Mrs. Russell, she urged Anne to give up the idea, Anne out of "prudence" and "responsibility" to accept the godmother's advice, with the beloved to break off the marriage, Colonel Wentworth was hurt, and decided to leave. Which match the second stage. More positive with less negative. After war, Wentworth go back to hometown and meet Anne again. This story is watch the middle stage of plot, more positive number. At last, Wentworth share his love agin with Anne shows the large positive sensitive.

```
# Make the comparison between three different kind of dictionary
```

```
afinn <- newbook %>%
  inner_join(get_sentiments("afinn")) %>%
  group_by(index = linenumber %/% 80) %>%
  summarise(sentiment = sum(value)) %>%
  mutate(method = "AFINN")
```

```
## Joining, by = "word"
```

```
bing_and_nrc <- bind_rows(
  newbook %>%
    inner_join(get_sentiments("bing")) %>%
    mutate(method = "Bing et al."),
  newbook %>%
    inner_join(get_sentiments("nrc")) %>%
    filter(sentiment %in% c("positive",
                          "negative"))
```

```

) %>%
  mutate(method = "NRC") %>%
  count(method, index = linenumber %/% 80, sentiment) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)

```

```

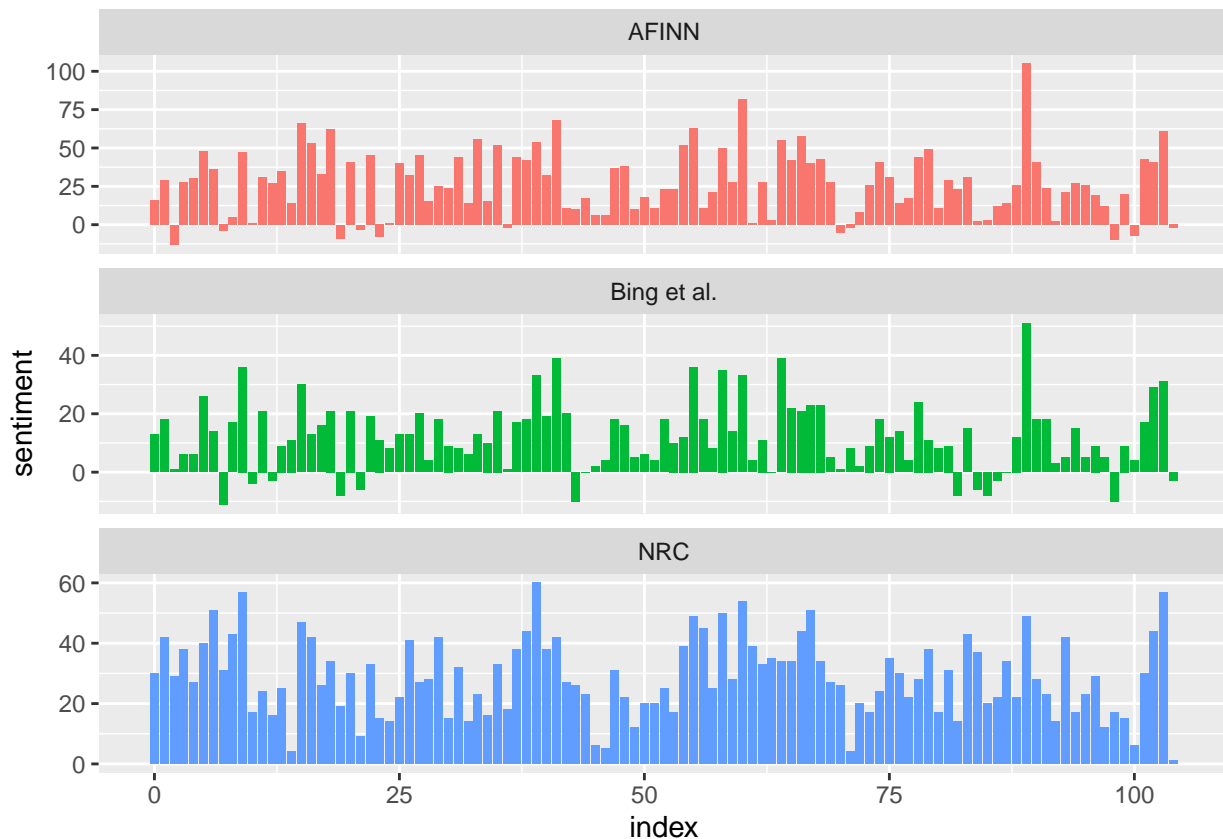
## Joining, by = "word"
## Joining, by = "word"

```

```

bind_rows(afinn,
          bing_and_nrc) %>%
  ggplot(aes(index, sentiment, fill = method)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~method, ncol = 1, scales = "free_y")

```



We can see that NRC doesn't show any negative emotion. and Bing and AFINN show the similar trend of emotion.

```

get_sentiments("nrc") %>%
  filter(sentiment %in% c("positive", "negative")) %>%
  count(sentiment)

```

```

## # A tibble: 2 x 2

```

```
## sentiment      n
## <chr>          <int>
## 1 negative     3318
## 2 positive     2308
```

*# According to NRC, we can find 3318 negative words and 2308 positive words*

```
get_sentiments("bing") %>%
  count(sentiment)
```

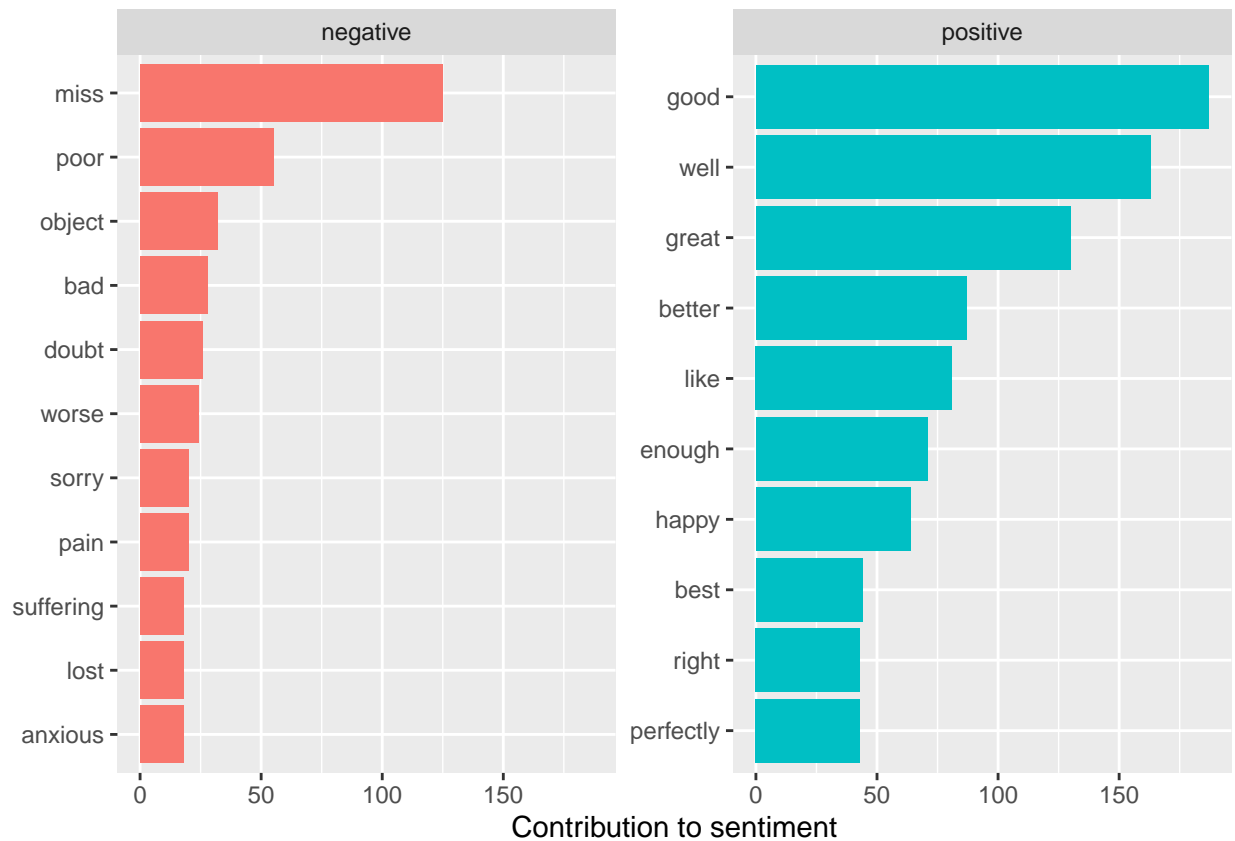
```
## # A tibble: 2 x 2
## sentiment      n
## <chr>          <int>
## 1 negative     4781
## 2 positive     2005
```

*# According to bing, we can find 4781 negative words and 2005 positive words*

```
bing_word_counts <- newbook %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
```

```
## Joining, by = "word"
```

```
bing_word_counts %>%
  group_by(sentiment) %>%
  slice_max(n, n = 10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(x = "Contribution to sentiment",
       y = NULL)
```



*# From the chart, we can see which words contribute more on negative or positive words most in book  
 # Obviously, "good, well, great" contribute most to positive emotio. and "miss" contribute most to negat  
 # emotion.*

```
# Make words cloud without stop word
newbook %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))
```

```
## Joining, by = "word"
```

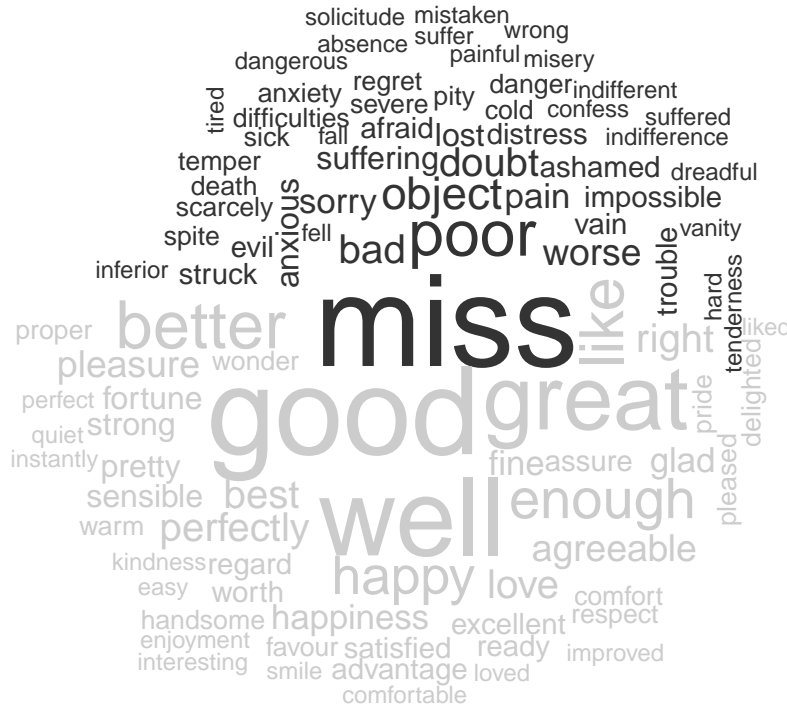


*# From the word cloud, we can find Anne and Captain appear the most,  
# which are the name of main role in book*

```
newbook %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("gray20", "gray80"),
    max.words = 100)
```

```
## Joining, by = "word"
```

# negative



# positive

*#This word cloud shows the negative and positive most words show in book.*

*#For negative word, "miss" show the most. For postive word, "good" show the most.*

Reference: <https://www.gutenberg.org/ebooks/105> Code are reference from Yuli Jin. Github:[https://github.com/MA615-Yuli/MA615\\_assignment4\\_new](https://github.com/MA615-Yuli/MA615_assignment4_new) and from <https://www.tidytextmining.com/sentiment.html>

Difference between my code and Jin's code is that due to my book have much more positive words than negative words, I use zscore instead of scale to compare. According to zscore, I can find the similar trend. Without zscore, I found the different trend for emotion for each section.