Text Analysis of Wuthering Heights (Task3 - Sentence Level Sentiment Analysis)

Jessie Xu

2021/12/7

Split Sentences Using Truenumber

- verify my assessment of the text

Show the first 3 rows and 2-3 colunms of the data frame.

property	string.value
text	"1801—I have just returned from a visit to my landlord—the solitary neighbour that I shall be trou
text	"This is certainly a beautiful country! In all England, I do not believe that I could have fixed on a s
text	"A perfect misanthropist's Heaven—and Mr Heathcliff and I are such a suitable pair to divide the d

Sentiment Analysis by Sentimentr

index = 1

index = 100
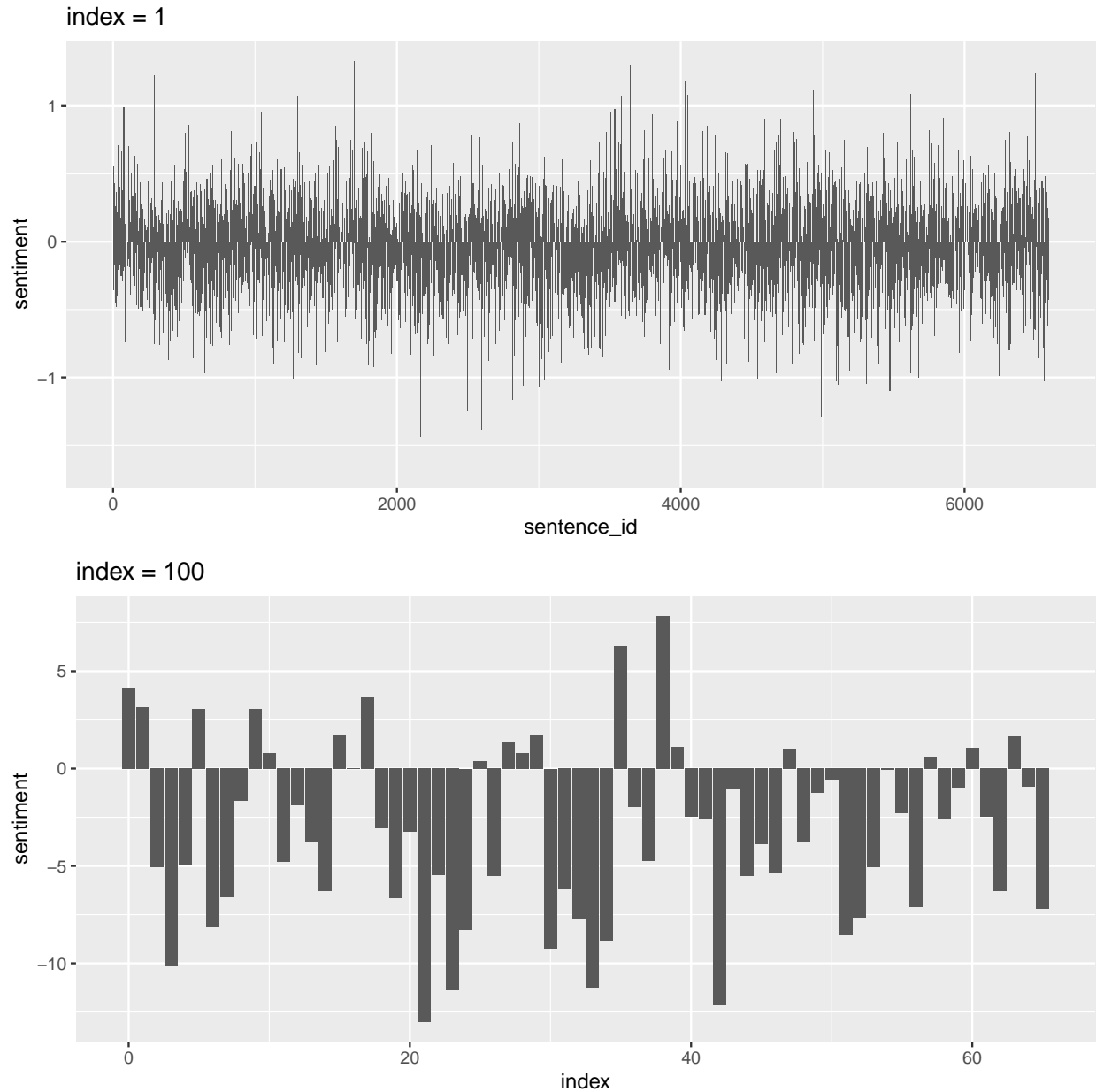
According to the plots above, we noticed that they have similar relative trajectories through the novel as the previous four lexicons plots (in task 2). We see similar dips and peaks in sentiment at about the same places in the novel. And since we use TNs to split the passage into sentence instead of lines, so the scale has changed.

Words Analysis (compared with Task 2)

1. an example using "nrc" lexicon

I select out the most frequent words filter by sentiment "joy" from "nrc" lexicon, almost the same, but I can't figure out why the frequency of a word changed.
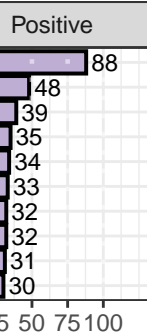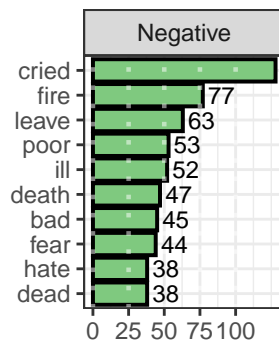
word	n
love	88
child	49
found	49
garden	41
companion	39
glad	35
hope	34
god	32
friend	31
happy	30

2. Comparing the three sentiment dictionaries: visualize sentiments using 4 lexicons (still working)
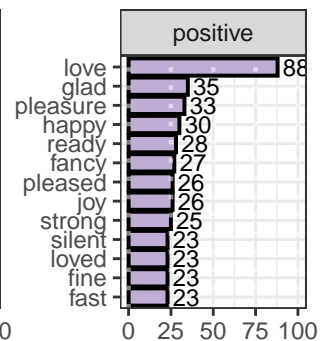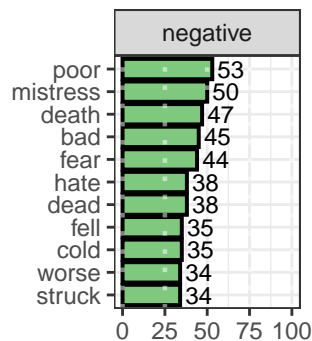
3. Top 10 words clustered by sentiment from different lexicon:

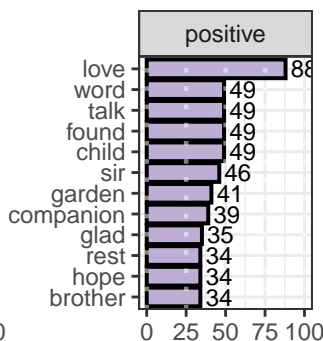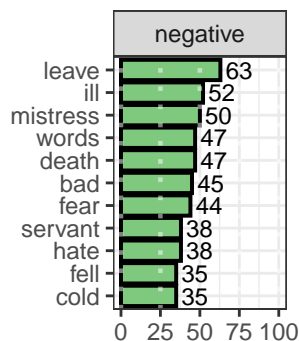## Top 10 words clustered by sentiment from different lexicon
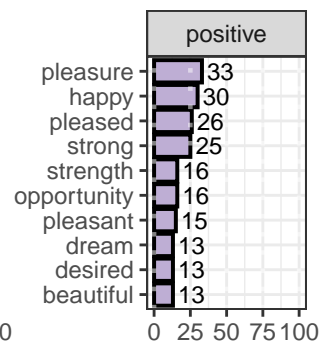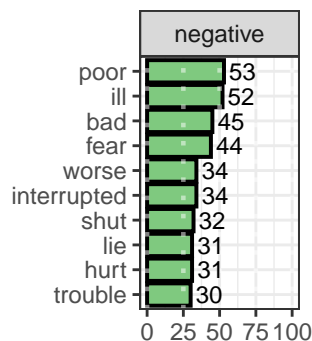
Afinn lexicon

Negative	Positive
cried	love 88
fire 77	care 48
leave 63	matter 39
poor 53	glad 35
ill 52	hope 34
death 47	pleasure 33
bad 45	god 32
fear 44	fit 32
hate 38	dear 31
dead 38	happy 30

bing lexicon

negative	positive
poor 53	love 88
mistress 50	glad 35
death 47	pleasure 33
bad 45	happy 30
fear 44	ready 28
hate 38	fancy 27
dead 38	pleased 26
fell 35	joy 26
cold 35	strong 25
worse 34	silent 23
struck 34	loved 23
	fine 23
	fast 23

nrc lexicon

negative	positive
leave 63	love 88
ill 52	word 49
mistress 50	talk 49
words 47	found 49
death 47	child 49
bad 45	sir 46
fear 44	garden 41
servant 38	companion 39
hate 38	glad 35
fell 35	rest 34
cold 35	hope 34
	brother 34

loughran lexicon

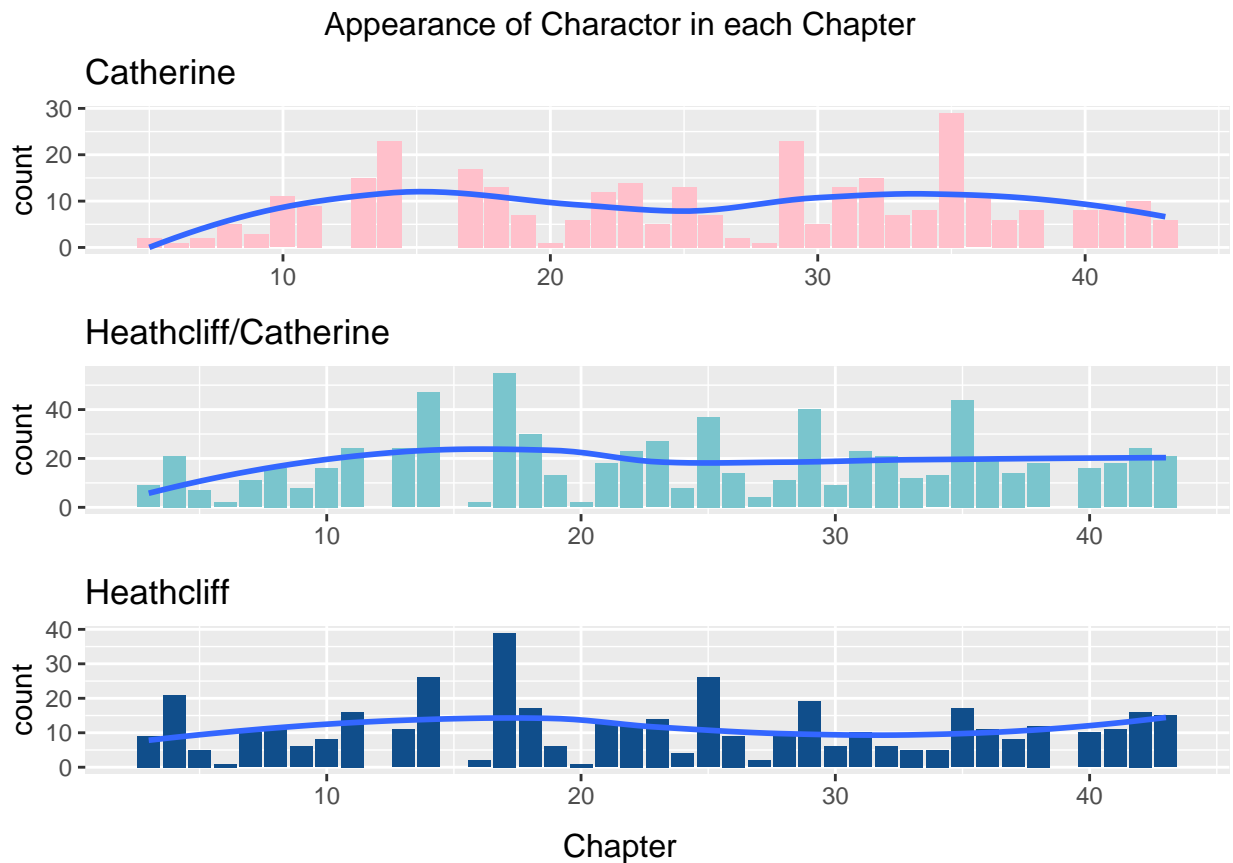| negative	positive
poor 53	pleasure 33
ill 52	happy 30
bad 45	pleased 26
fear 44	strong 25
worse 34	strength 16
interrupted 34	opportunity 16
shut 32	pleasant 15
lie 31	dream 13
hurt 31	desired 13
trouble 30	beautiful 13

```
## NULL
```

4. Word Cloud

Almost the same, only the frequency of some words change a little.

EXTRA CREDIT: character analysis

I pick Heathcliff and Catherine as my character, and plot the number of times each one appears in each chapter and the number of times both characters appear in the same paragraphs.

Appearance of Charactor in each Chapter

Inference:

All the Haviland and Allen's code;

https://www.r-bloggers.com/2020/04/sentiment-analysis-in-r-with-sentimentr-that-handles-negation-valence-shifters/;

https://towardsdatascience.com/doing-your-first-sentiment-analysis-in-r-with-sentimentr-167855445132;

https://github.com/BruceMallory/Truenumbers/blob/e017672768b13c1fd98b9dc54c19c1aedbcac582/Jane Austen.pdf;

# Appendix:

- Split Sentences Using the Way I Learned by Myself

```
wh_s = readLines("https://raw.githubusercontent.com/MA615-Jessie-Xu/Assignment-4/main/WutheringHeights.t
wh_s = paste0(wh_s, collapse = ' ')
wh_s = gsub('\\t|\\r|\\n', ' ', wh_s)
wh_s = gsub('\\s+', ' ', wh_s)
wh_s = trimws(wh_s)
wh_s <- as.data.frame(wh_s) # 1x1

wh_with_senti <-
  wh_s%>%
  get_sentences()%>%
  sentiment()


wh_with_senti %>%
  ggplot() + geom_col(aes(x = sentence_id, y = sentiment))+ labs(title="index = 1")
```
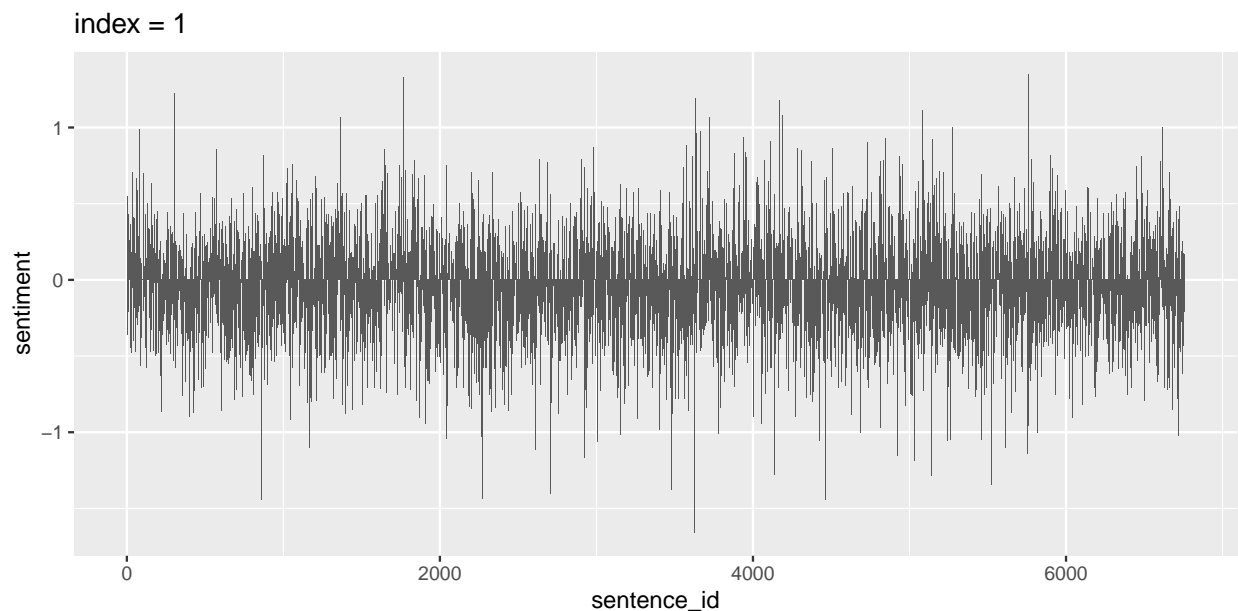
index = 1


```
wh_with_senti %>%
  group_by(index = sentence_id %/% 100) %>%
  summarize(sentiment = sum(sentiment)) %>%
    ggplot() + geom_col(aes(x = index, y = sentiment))+ labs(title="index = 100")
```

index = 100
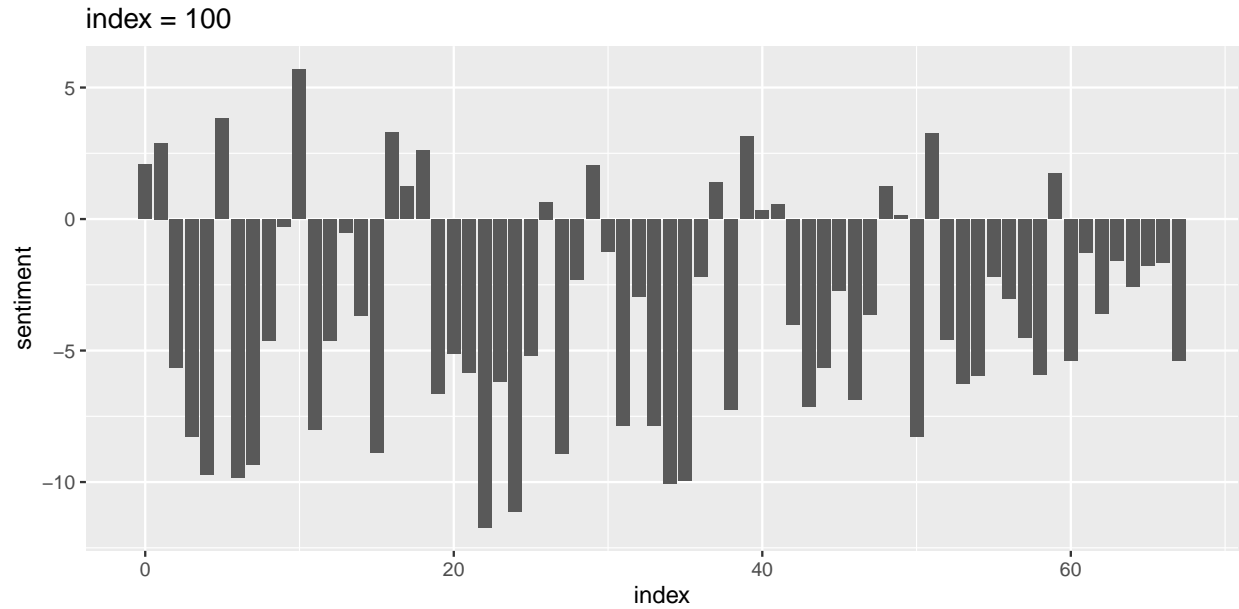
```r
# text_s$wh_s %>%
#   #get_sentences() %>%
#   sentiment_by() %>% #View()
#   highlight()
```

- Other EDA

1. This EDA shows the distribution of the occurrence of Heathcliff and Catherine in each Chapter. Difference color refer different paragraphs.

Distribution of the Appearance of Heathcliff and Catherine in each Ch
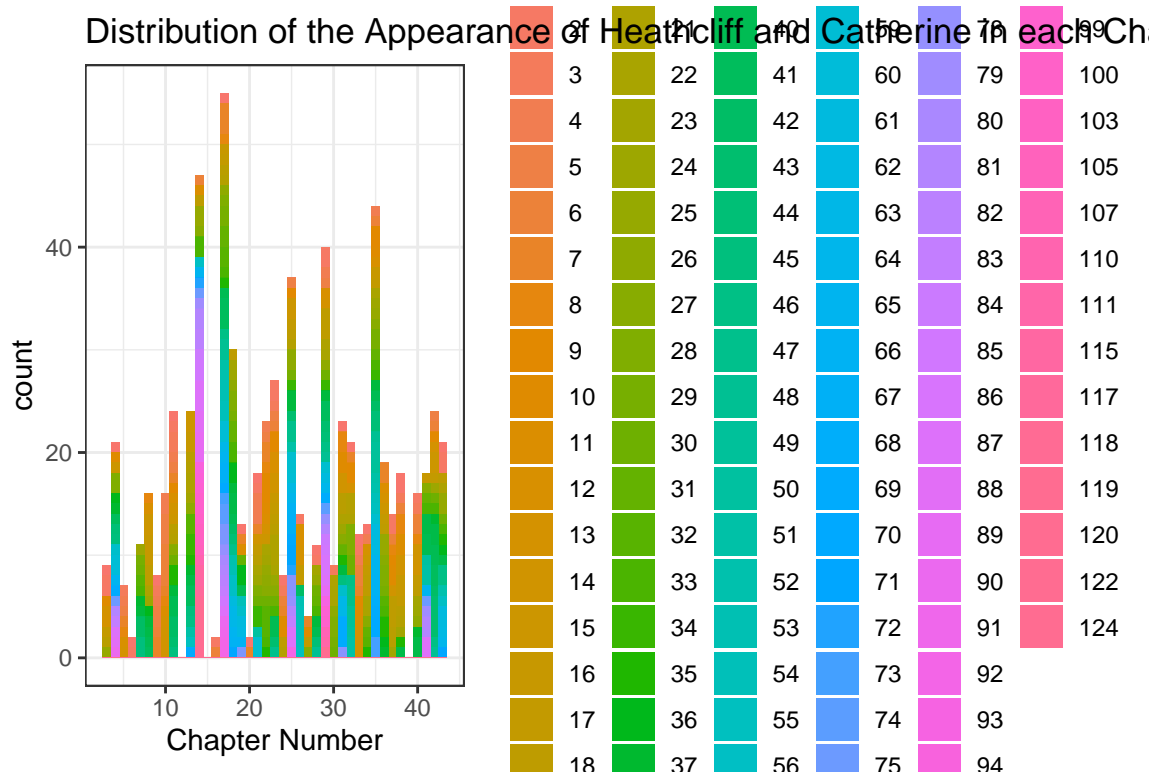
2. Word Cloud:Plot the most frequent words which has real meaning in the whole passage, instead of stop and

custum words or personal pronouns.
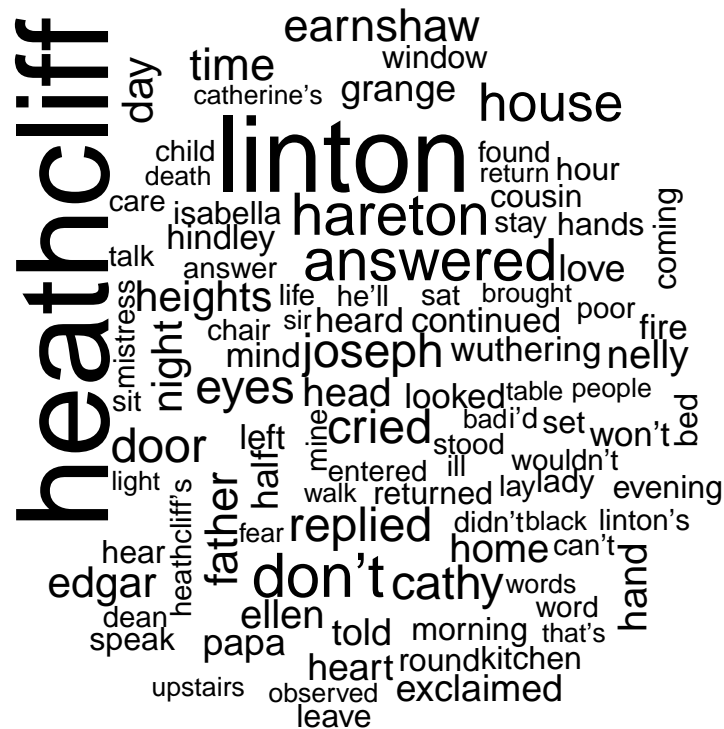
- Further Steps using tnum

```
CHgraph<-tnum.graphTnumList(df_CH$subject)
tnum.plotGraph(CHgraph)
```

As you see, I try to make a tree diagram to plot the logical connection between chapters. But I can't use the function like <tnum.makeTnumPhraseGraph> or<tnum.plotGraph> offered by the "tnum—instructions-and-examples-v1.0.pdf" and other tutorial about tnum. Even I already install the "tnum" package, the error said I can't use the function. I would like to figure it out later.