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

Jessie Xu

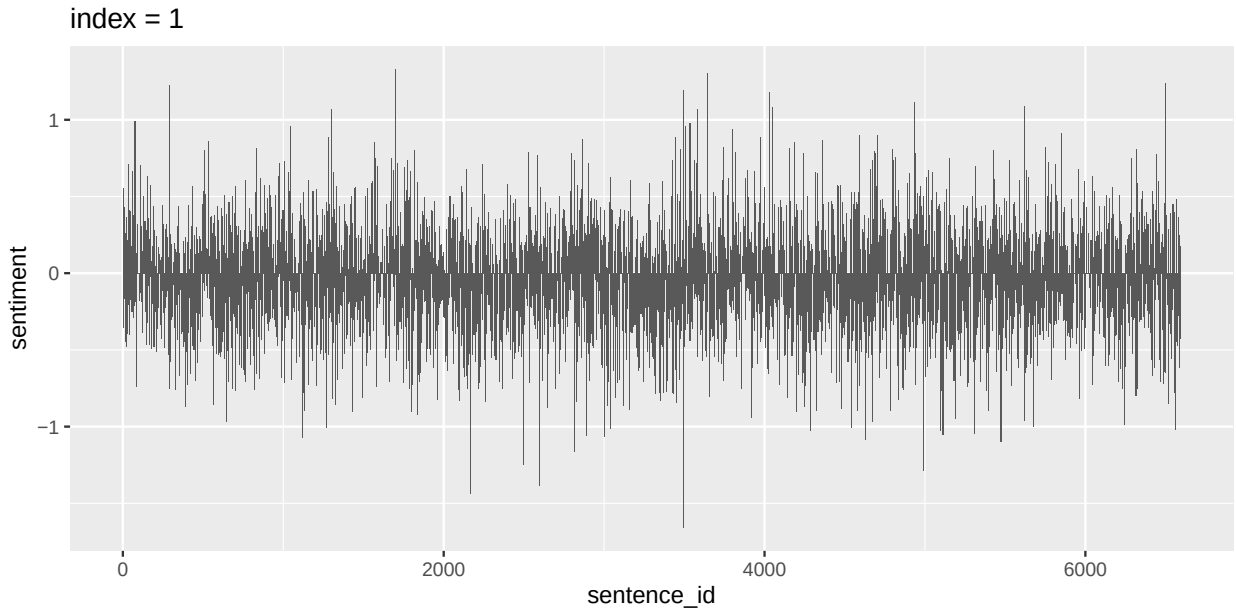
2021/12/7

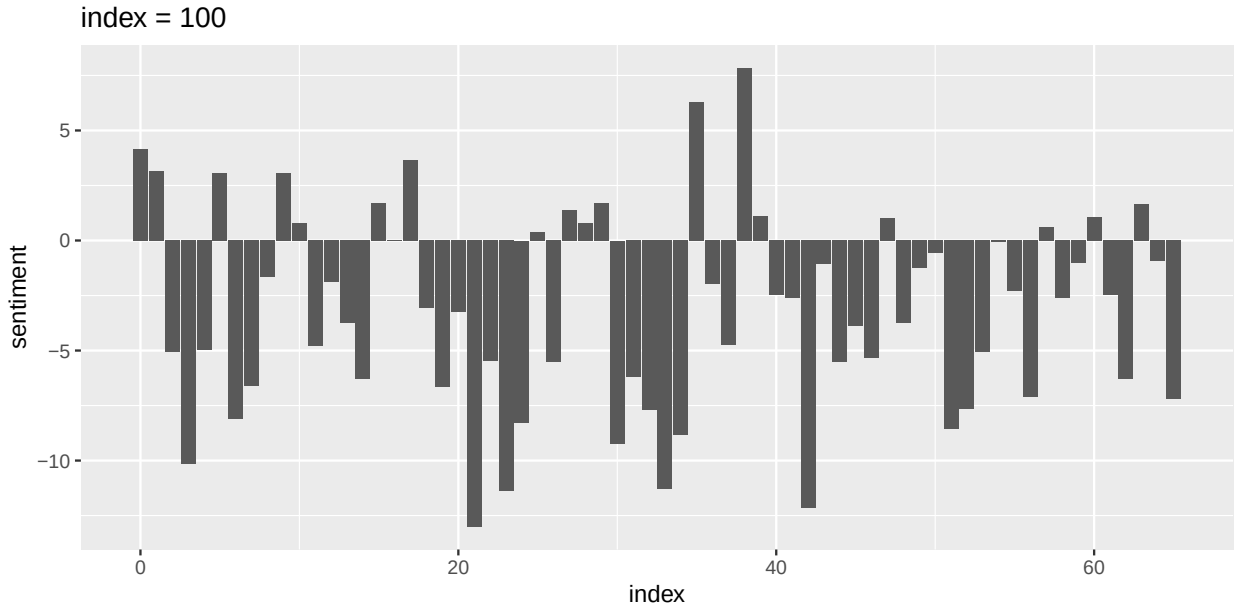
Split Sentences Using Truenumber

To verify my assessment of the text, the table below shows the first 3 rows and 2-3 columns 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





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)

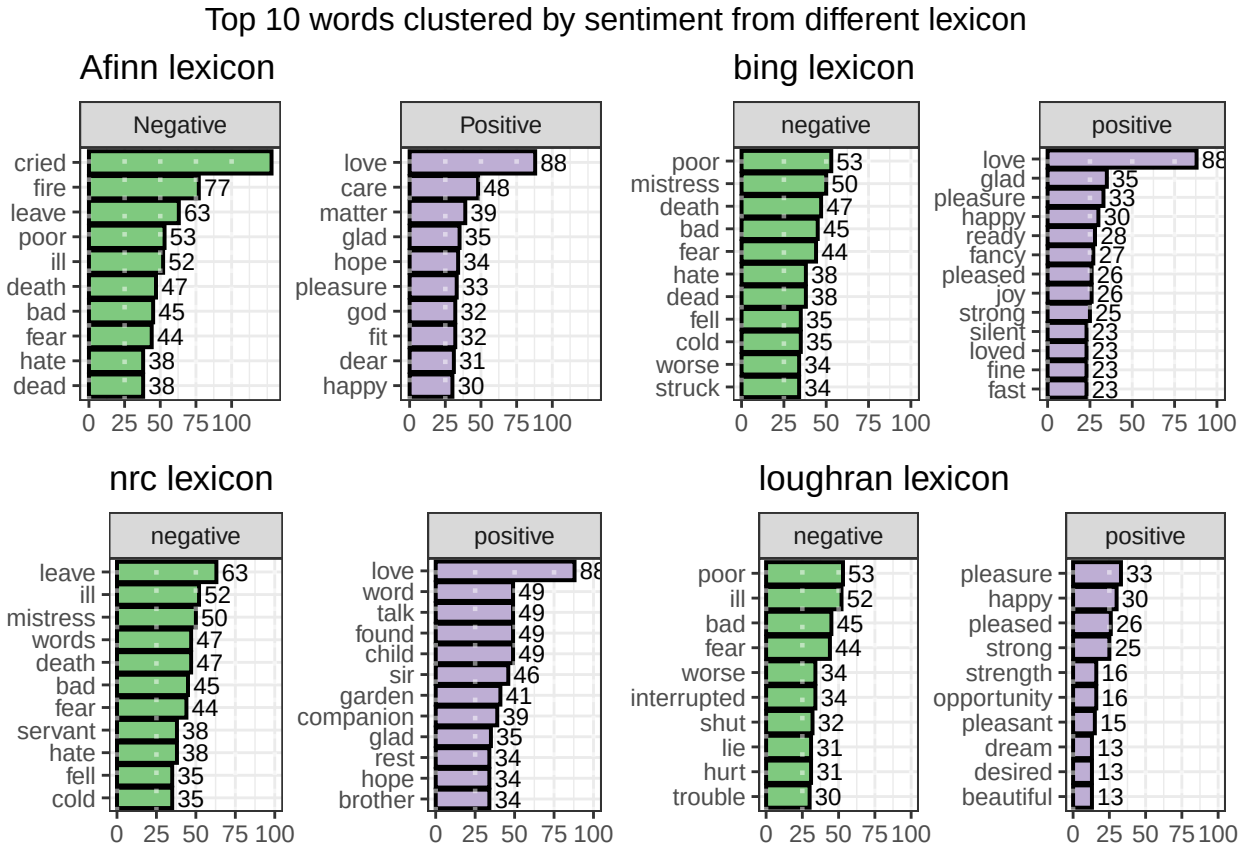
1. an example using "nrc" lexicon

I select out the most frequent words filter by sentiment "joy" from "nrc" lexicon. Comparing to the Task2, the results are almost the same, but I can't figure out why the frequency of some 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:



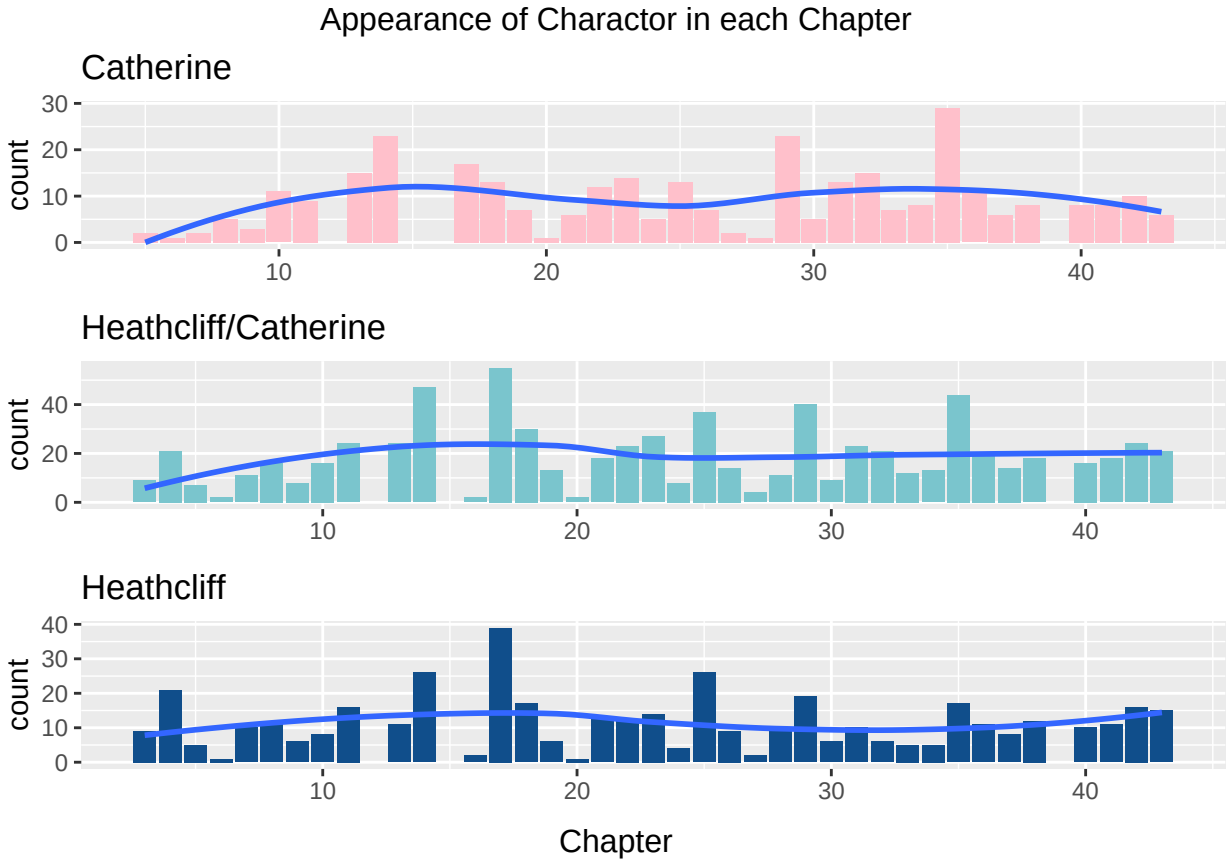

```
## NULL
```

4. Word Cloud

Comparing to the Task2, the results are almost the same, only the frequency of some words changes a little.

EXTRA CREDIT: character analysis

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



Inference:

- All the Haviland and Allen's code;
- The sentimentr video:
  - https://www.r-bloggers.com/2020/04/sentiment-analysis-in-r-with-sentimentr-that-handles-negation-valence-shifters/
- The EDA in the appendix:
  - https://github.com/BruceMallory/Truenumbers/blob/e017672768b13c1fd98b9dc54c19c1aedbcac582/JaneAusten.pdf

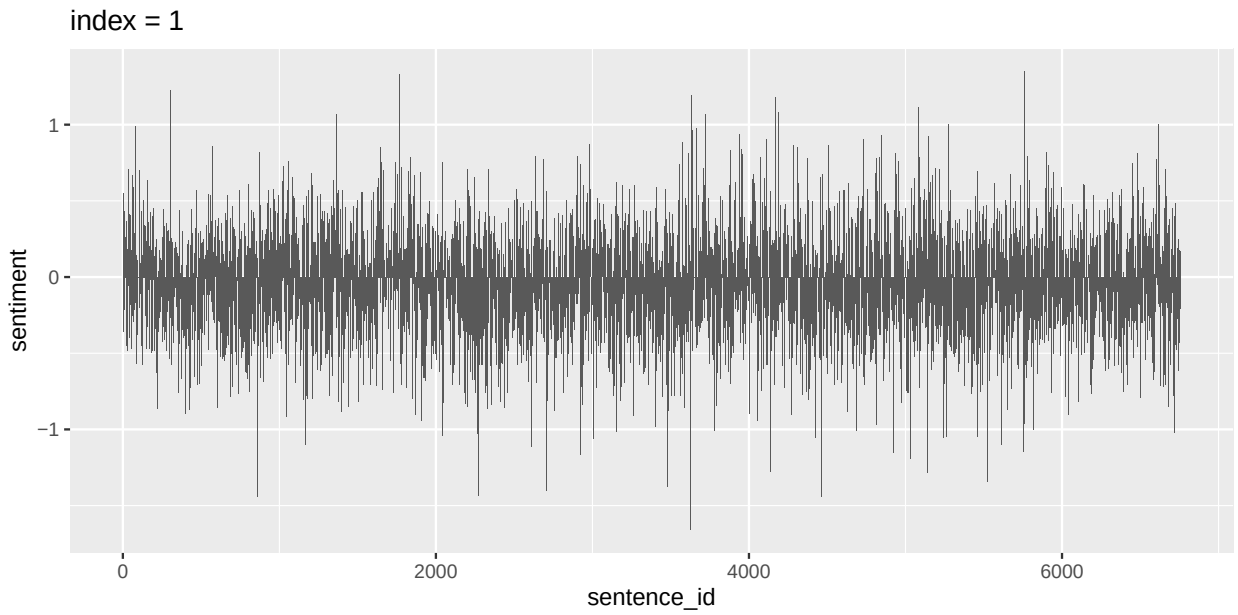
Appendix:

Split Sentences Using the Way I Learned Myself

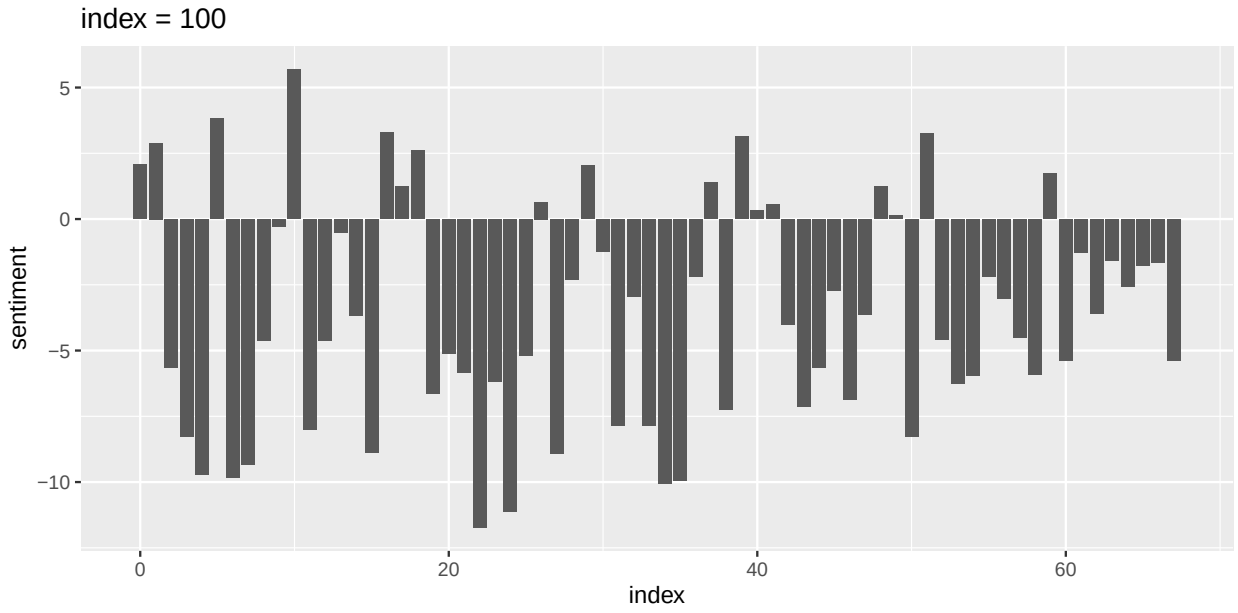
```
wh_s = readLines("https://raw.githubusercontent.com/MA615-Jessie-Xu/Assignment-4/main/WutheringHeights.t
wh_s = paste0(wh_s, collapse = ' ')
wh_s = gsub('\\t|\\r|\\n', ' ', wh_s)
wh_s = gsub('\\s+', ' ', wh_s)
wh_s = trimws(wh_s)
wh_s <- as.data.frame(wh_s) # 1x1

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


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




```
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")
```




```
# 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 colors refer to different paragraphs.
2. Word Cloud: Plot the most frequent words which have real meaning in the whole passage, instead of stop and custom words or personal pronouns.

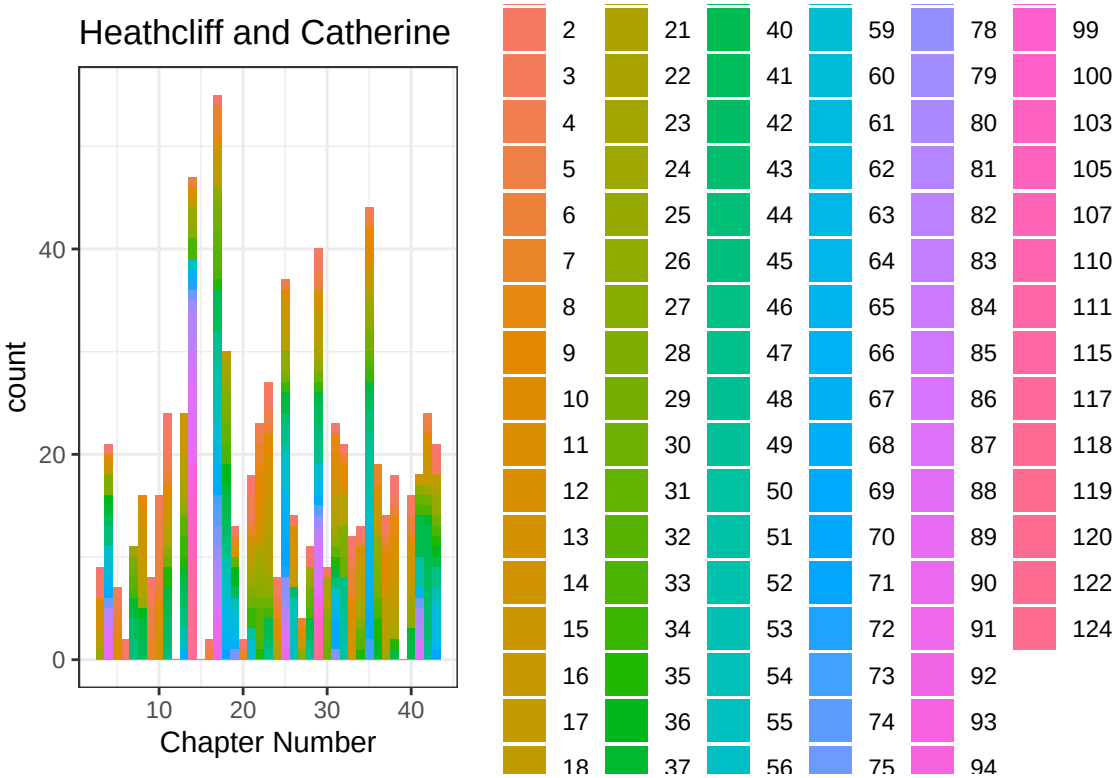
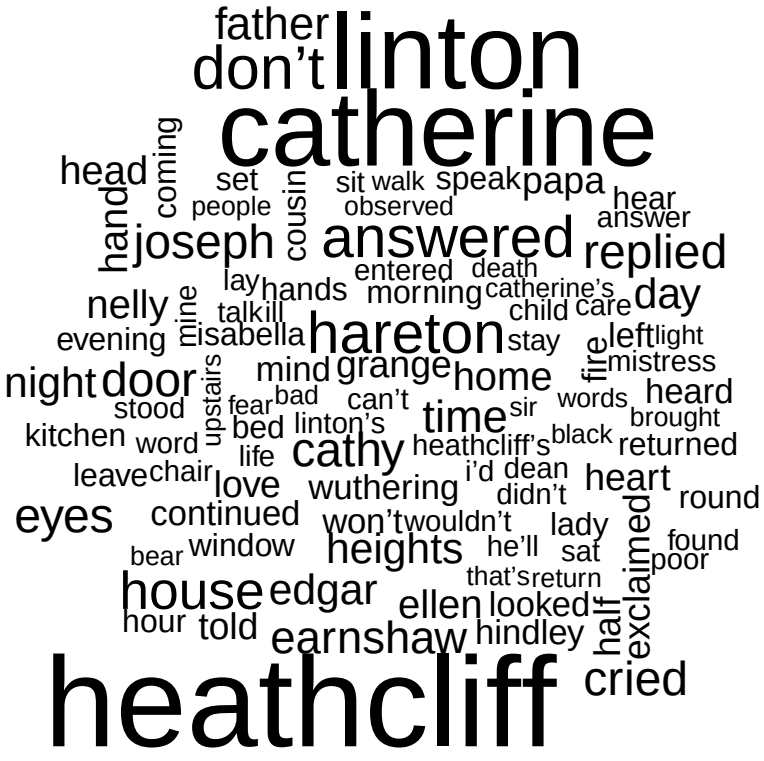


Figure 1: Distribution of the Appearance of Heathcliff and Catherine in each Chapter



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 (characters). 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 installed the "tnum" package, the error said I can't use the function. I would like to figure it out later.