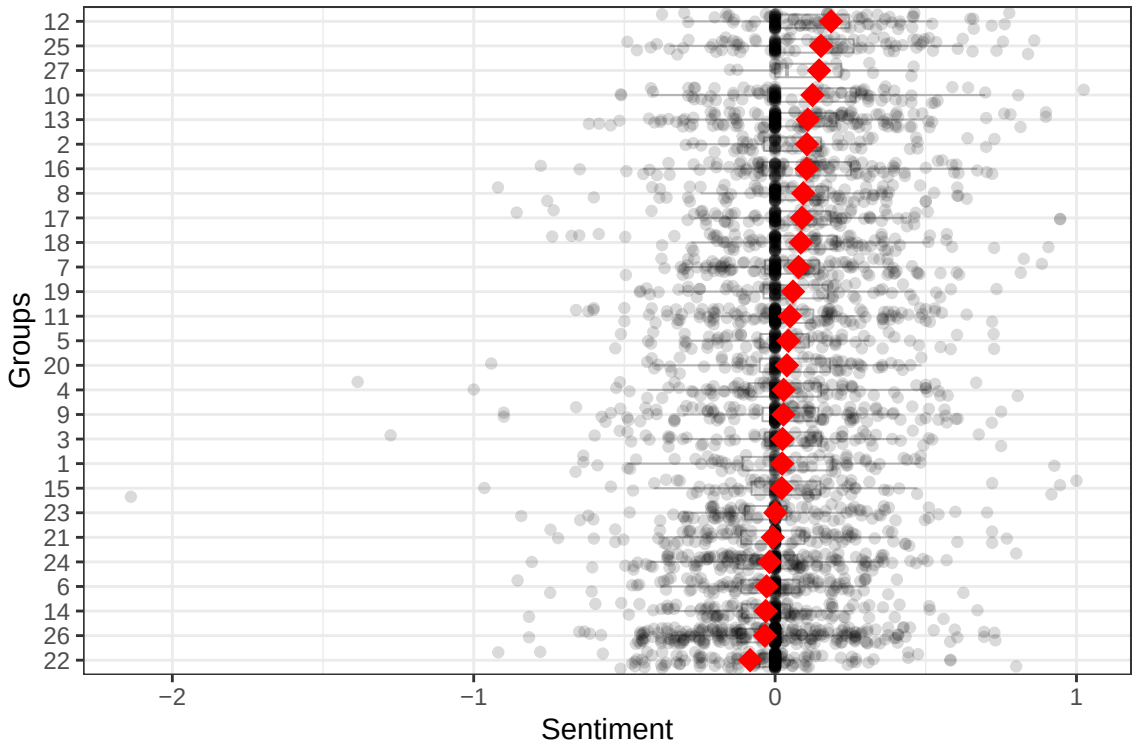
Task Three

Wuji Shan

12/7/2021

Task Three

```
library(tnum)
tnum.authorize("mssp1.bu.edu")
tnum.setSpace("test2")
source("Book2TN-v6A-1.R")


# adjust the book type
Cami_book <- gutenberg_download(1608)
#write.table(Cami_book, "Cami_book.txt")
book_fix <- read.table("Cami_book.txt", header = T)

# tnBooksFromLines(book_fix$text, "Alexandre_Dumas/Camille_Book92")

tidy_Cami_3 <- book_fix %>%
  mutate(
    linenumber = row_number(),
    chapter = cumsum(str_detect(text,
                                regex("chapter",
                                      ignore_case = TRUE)))) %>%
  unnest_tokens(word, text)

df_Cami <- tnum.query('Alexandre_Dumas/Camille_Book92/section# has text', max = 10000) %>% tnum.objects

Cami_sentence <- df_Cami %>% separate(col = subject,
                  into = c("path1", "path2","section","paragraph","sentence"),
                  sep = "/",
                  fill = "right") %>%
  select(section:string.value)

Cami_sentence <- Cami_sentence %>% mutate_at(c('section','paragraph','sentence'),~str_extract_all(.,"\\d
                                          %>% unlist()
                                          %>% as.numeric())


sentence_out <- Cami_sentence %>% dplyr::mutate(sentence_split = get_sentences(string.value)) %$%
    sentiment_by(sentence_split, list(section))

plot(sentence_out)
```



Compare

```
# create a new bing with index=chapter
new_bing <- tidy_Cami_3 %>%
    inner_join(get_sentiments("bing")) %>%
    mutate(method = "Bing et al.") %>%
    count(method, index = chapter, sentiment) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)

# scale sentiment to keep unit same
new_bing2 <- new_bing %>%
  mutate(bing_scale = scale(sentiment)) %>%
  select(method, index, bing_scale)

# change colname in order to join by section
colnames(new_bing2)[2]='section'

# scale sentiment to keep unit same
sentence_out <- sentence_out %>% mutate(sentimentr_scale = scale(ave_sentiment))

# join two df
sentence_out_2method <- left_join(sentence_out, new_bing2, by='section') %>%
  select(section,bing_scale,sentimentr_scale)

# use pivot longer for ggplot
```

```
sentence_out_2method_plot <- sentence_out_2method %>%
  pivot_longer(cols = c('sentimentr_scale','bing_scale'), names_to = 'sentiment')

# create barplot to compare
sentence_out_2method_plot %>% ggplot(aes(y = value,x = factor(section))) +
  geom_bar(aes(fill = factor(sentiment)), stat = 'identity', position = "dodge",width = 0.7) +
  theme_bw()
```