Lab 6

Enter the names of the group members here: Austine Do, Graceanne Becker, Catherine Zhong

This assignment is due by the end of the lab. Only one student in the group submits a pdf file on Gradescope.

For all questions, include the R commands/functions that you used to find your answer (show R chunk). Answers without supporting code will not receive credit. Write full sentences to describe your findings.

In this lab, you will explore one of the published books of Jane Austen, accessible through the `janeaustenr` package. Let's first install it:

```r
# Install package (Note, eval=FALSE means this code chunk is not submitted when knitting)
install.packages("janeaustenr")
```

Then load that package and other necessary packages for today:

```r
# If you haven't installed these packages remember to do so from the Text Mining Worksheet
library(janeaustenr)
library(tidyverse)
library(tidytext)
library(ggwordcloud)
```

Let's take a quick look at the data available for the books:

```r
# Save data in your environment
austen_books <- austen_books()

# Take a quick look
austen_books |>
  tail()
```

```
## # A tibble: 6 x 2
##   text                                                      book
##   <chr>                                                     <fct>
## 1 "possible, more distinguished in its domestic virtues than in its" Persuasion
## 2 "national importance."                                     Persuasion
## 3 ""                                                         Persuasion
## 4 ""                                                         Persuasion
## 5 ""                                                         Persuasion
## 6 "Finis"                                                    Persuasion
```

The text is split by line for each book. The goal of the lab is to conduct some sentiment analysis for one of these books.

Question 1: (3 pts)

Find the number of lines for each book in `austen_books`. Which book has the greatest number of lines?
Note: it's ok to count empty lines, titles, etc. Just find the total number of lines.

```
# Finds the total number of lines by novel
austen_books |>
    group_by(book) |>
    summarize(length = n()) |>
    arrange(length)
```

```
## # A tibble: 6 x 2
##   book                length
##   <fct>                <int>
## 1 Northanger Abbey      7856
## 2 Persuasion            8328
## 3 Sense & Sensibility  12624
## 4 Pride & Prejudice    13030
## 5 Mansfield Park       15349
## 6 Emma                 16235
```

The book that has the greatest number of lines is the novel 'Emma'

Question 2: (6 pts)

What does the code below do?

The code below creates a variable called `chapter` that keeps track of the chapter of the line in each book. Note: `ungroup()` is considering text as a whole again, not just per book.

```
# Select one book and code the number of chapters
austen_books |>
  group_by(book) |>
  mutate(chapter = cumsum(str_detect(text, "^Chapter|^CHAPTER"))) |>
  ungroup()
```

```
## # A tibble: 73,422 x 3
##    text                   book                chapter
##    <chr>                  <fct>                 <int>
##  1 "SENSE AND SENSIBILITY" Sense & Sensibility      0
##  2 ""                     Sense & Sensibility       0
##  3 "by Jane Austen"       Sense & Sensibility       0
##  4 ""                     Sense & Sensibility       0
##  5 "(1811)"               Sense & Sensibility       0
##  6 ""                     Sense & Sensibility       0
##  7 ""                     Sense & Sensibility       0
##  8 ""                     Sense & Sensibility       0
##  9 ""                     Sense & Sensibility       0
## 10 "CHAPTER 1"            Sense & Sensibility       1
## # i 73,412 more rows
```

```
# Continue the code
```

Continue the code above to:

1. Choose to keep **only one of the books** (the book of your choice!),

2. Get rid of the lines for chapter 0,

3. Get rid of the empty lines,

4. Get rid of the lines showing the chapter sections. *Hint: str_detect() with Regex would be useful again!* Note: some novels also have "volumes". Get rid of those lines as well.

Finally, call the resulting dataset `book`.

How many chapters were contained in the `book` you chose?

```r
# Creates a data frame of lines from Pride & Prejudice excluding chapter 0 and section titles

book <- austen_books |>
            group_by(book) |>
            mutate(chapter = cumsum(str_detect(text, "^Chapter|^CHAPTER"))) |>
            ungroup() |>
            filter(book == 'Pride & Prejudice' & chapter != 0 & text != '' & !str_detect(text, '^CH

length(unique(book$chapter))
```

```
## [1] 61
```

There are 61 chapters in the novel 'Pride & Prejudice'

Question 3: (3 pts)

Next, split each line into words with `unnest_tokens()` and save the resulting dataset as `words_book`:

```r
# This creates a data frame of only words from Pride & Prejudice
book |>
    unnest_tokens(input = text, output = word) -> words_book
```

What are the 10 most common words in the book you chose? Do they reveal any pattern?

```r
# Finds the top 10 most common words in Pride & Prejudice
words_book |>
    group_by(word) |>
    summarize(frequency = n()) |>
    slice_max(n = 10, frequency)
```

```
## # A tibble: 10 x 2
##     word   frequency
##     <chr>      <int>
##  1 the         4331
##  2 to          4162
##  3 of          3610
##  4 and         3584
##  5 her         2203
##  6 i           2065
##  7 a           1954
##  8 in          1880
##  9 was         1843
## 10 she         1695
```

The most common words in the book we chose were 'the', 'to', 'of', 'and', 'her', 'i', 'a', 'in', 'was', and 'she'. This does not show any pattern as these words don't mean anything as stand alone words.

Question 4: (5 pts)

After getting the words by themselves, let's get rid of the stop words with the `SMART` lexicon:

```
# Recall the SMART lexicon
SMARTstops <- stop_words |> filter(lexicon == "SMART")
```

Use a joining function to get rid of stop words in `words_book` then find the 10 most common words and display them in a word cloud (most frequent words should appear bigger and in a different color). Do you notice any pattern in these words?

```
# Finds the top 10 most common words excluding words from the SMART lexicon and displays a word cloud
words_book |>
    group_by(word) |>
    summarize(frequency = n()) |>
    anti_join(SMARTstops,  by = 'word') |>
    slice_max(n = 10, frequency) |>
    ggplot(aes(label = word, size = frequency, color = frequency)) +
        geom_text_wordcloud() +
        scale_size_area(max_size = 20) +
        scale_color_gradient(low = 'blue', high = 'red')
```

```
theme_minimal()
```

```
## List of 97
##  $ line                    :List of 6
##   ..$ colour      : chr "black"
##   ..$ linewidth   : num 0.5
##   ..$ linetype    : num 1
##   ..$ lineend     : chr "butt"
##   ..$ arrow       : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ rect                    :List of 5
##   ..$ fill        : chr "white"
##   ..$ colour      : chr "black"
##   ..$ linewidth   : num 0.5
##   ..$ linetype    : num 1
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##  $ text                    :List of 11
##   ..$ family      : chr ""
##   ..$ face        : chr "plain"
##   ..$ colour      : chr "black"
##   ..$ size        : num 11
##   ..$ hjust       : num 0.5
##   ..$ vjust       : num 0.5
```

```
##   ..$ angle       : num 0
##   ..$ lineheight  : num 0.9
##   ..$ margin      : 'margin' num [1:4] 0points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ title                  : NULL
## $ aspect.ratio           : NULL
## $ axis.title             : NULL
## $ axis.title.x           :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 1
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 2.75points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.x.top       :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 0
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 0points 0points 2.75points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.x.bottom    : NULL
## $ axis.title.y           :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 1
##   ..$ angle       : num 90
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 0points 2.75points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.y.left      : NULL
```

```
##  $ axis.title.y.right       :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 0
##   ..$ angle       : num -90
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 0points 0points 0points 2.75points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text                 :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : chr "grey30"
##   ..$ size        : 'rel' num 0.8
##   ..$ hjust       : NULL
##   ..$ vjust       : NULL
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : NULL
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x               :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 1
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 2.2points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x.top           :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : NULL
##   ..$ vjust       : num 0
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 0points 0points 2.2points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
```

```
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.x.bottom       : NULL
## $ axis.text.y              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.2points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.y.left         : NULL
## $ axis.text.y.right        :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.2points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.ticks               : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
## $ axis.ticks.x             : NULL
## $ axis.ticks.x.top         : NULL
## $ axis.ticks.x.bottom      : NULL
## $ axis.ticks.y             : NULL
## $ axis.ticks.y.left        : NULL
## $ axis.ticks.y.right       : NULL
## $ axis.ticks.length        : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
## $ axis.ticks.length.x      : NULL
## $ axis.ticks.length.x.top  : NULL
## $ axis.ticks.length.x.bottom: NULL
## $ axis.ticks.length.y      : NULL
## $ axis.ticks.length.y.left : NULL
## $ axis.ticks.length.y.right : NULL
## $ axis.line                : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
## $ axis.line.x              : NULL
## $ axis.line.x.top          : NULL
## $ axis.line.x.bottom       : NULL
## $ axis.line.y              : NULL
## $ axis.line.y.left         : NULL
```

```
##   $ axis.line.y.right        : NULL
##   $ legend.background         : list()
##    ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##   $ legend.margin             : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##    ..- attr(*, "unit")= int 8
##   $ legend.spacing            : 'simpleUnit' num 11points
##    ..- attr(*, "unit")= int 8
##   $ legend.spacing.x          : NULL
##   $ legend.spacing.y          : NULL
##   $ legend.key                : list()
##    ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##   $ legend.key.size           : 'simpleUnit' num 1.2lines
##    ..- attr(*, "unit")= int 3
##   $ legend.key.height         : NULL
##   $ legend.key.width          : NULL
##   $ legend.text               :List of 11
##    ..$ family    : NULL
##    ..$ face      : NULL
##    ..$ colour    : NULL
##    ..$ size      : 'rel' num 0.8
##    ..$ hjust     : NULL
##    ..$ vjust     : NULL
##    ..$ angle     : NULL
##    ..$ lineheight : NULL
##    ..$ margin    : NULL
##    ..$ debug     : NULL
##    ..$ inherit.blank: logi TRUE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##   $ legend.text.align         : NULL
##   $ legend.title              :List of 11
##    ..$ family    : NULL
##    ..$ face      : NULL
##    ..$ colour    : NULL
##    ..$ size      : NULL
##    ..$ hjust     : num 0
##    ..$ vjust     : NULL
##    ..$ angle     : NULL
##    ..$ lineheight : NULL
##    ..$ margin    : NULL
##    ..$ debug     : NULL
##    ..$ inherit.blank: logi TRUE
##    ..- attr(*, "class")= chr [1:2] "element_text" "element"
##   $ legend.title.align        : NULL
##   $ legend.position           : chr "right"
##   $ legend.direction          : NULL
##   $ legend.justification      : chr "center"
##   $ legend.box                : NULL
##   $ legend.box.just           : NULL
##   $ legend.box.margin         : 'margin' num [1:4] 0cm 0cm 0cm 0cm
##    ..- attr(*, "unit")= int 1
##   $ legend.box.background      : list()
##    ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##   $ legend.box.spacing        : 'simpleUnit' num 11points
##    ..- attr(*, "unit")= int 8
```

```
##  $ panel.background        : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ panel.border            : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ panel.spacing           : 'simpleUnit' num 5.5points
##   ..- attr(*, "unit")= int 8
##  $ panel.spacing.x         : NULL
##  $ panel.spacing.y         : NULL
##  $ panel.grid              :List of 6
##   ..$ colour      : chr "grey92"
##   ..$ linewidth   : NULL
##   ..$ linetype    : NULL
##   ..$ lineend     : NULL
##   ..$ arrow       : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ panel.grid.major        : NULL
##  $ panel.grid.minor        :List of 6
##   ..$ colour      : NULL
##   ..$ linewidth   : 'rel' num 0.5
##   ..$ linetype    : NULL
##   ..$ lineend     : NULL
##   ..$ arrow       : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ panel.grid.major.x      : NULL
##  $ panel.grid.major.y      : NULL
##  $ panel.grid.minor.x      : NULL
##  $ panel.grid.minor.y      : NULL
##  $ panel.ontop             : logi FALSE
##  $ plot.background         : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ plot.title              :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : 'rel' num 1.2
##   ..$ hjust       : num 0
##   ..$ vjust       : num 1
##   ..$ angle       : NULL
##   ..$ lineheight  : NULL
##   ..$ margin      : 'margin' num [1:4] 0points 0points 5.5points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.title.position     : chr "panel"
##  $ plot.subtitle           :List of 11
##   ..$ family      : NULL
##   ..$ face        : NULL
##   ..$ colour      : NULL
##   ..$ size        : NULL
##   ..$ hjust       : num 0
##   ..$ vjust       : num 1
```

```
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 5.5points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ plot.caption             :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : num 1
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 5.5points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ plot.caption.position    : chr "panel"
## $ plot.tag                 :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 1.2
##   ..$ hjust        : num 0.5
##   ..$ vjust        : num 0.5
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ plot.tag.position        : chr "topleft"
## $ plot.margin              : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##   ..- attr(*, "unit")= int 8
## $ strip.background         : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
## $ strip.background.x       : NULL
## $ strip.background.y       : NULL
## $ strip.clip               : chr "inherit"
## $ strip.placement          : chr "inside"
## $ strip.text               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : chr "grey10"
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 4.4points 4.4points 4.4points 4.4points
```

```
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug       : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ strip.text.x             : NULL
## $ strip.text.x.bottom      : NULL
## $ strip.text.x.top         : NULL
## $ strip.text.y             :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : num -90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ strip.text.y.left        :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ strip.text.y.right       : NULL
## $ strip.switch.pad.grid    : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
## $ strip.switch.pad.wrap    : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
## - attr(*, "class")= chr [1:2] "theme" "gg"
## - attr(*, "complete")= logi TRUE
## - attr(*, "validate")= logi TRUE
```

Most of these top 10 most common words are just titles or names.

Question 5: (6 pts)

Let's take a look at the sentiments associated with words in the book and how these sentiments change as the story goes. Consider the sentiment value associated with each word from the `afinn` lexicon:

```r
# Sentiments value
get_sentiments("afinn") |> head()
```

```
## # A tibble: 6 x 2
##   word        value
##   <chr>       <dbl>
## 1 abandon        -2
## 2 abandoned      -2
## 3 abandons       -2
## 4 abducted       -2
## 5 abduction      -2
## 6 abductions     -2
```

Follow these steps to keep track of the sentiments as the story goes:

1. Use a joining function to only keep the words in `words_book` that are associated with a sentiment value.

2. Find the average sentiment value per chapter.

3. Create a `ggplot` with `geom_line()` to represent the average sentiment value across the chapters.

How do the sentiments evolve as the story goes?

```r
# Joins the sentiment data frame with the words_book data frame and summarizes the average sentiment pe
words_book |>
    inner_join(get_sentiments('afinn')) |>
    group_by(chapter) |>
    summarize(average_sentiment = mean(value)) |>
    ggplot(aes(x = chapter, y = average_sentiment)) +
        geom_line() +
        labs(title = 'Average sentiment as the novel \'Pride & Prejudice\' progresses',
             x = 'chapter number',
             y = 'average sentiment')
```

Average sentiment as the novel 'Pride & Prejudice' progresses
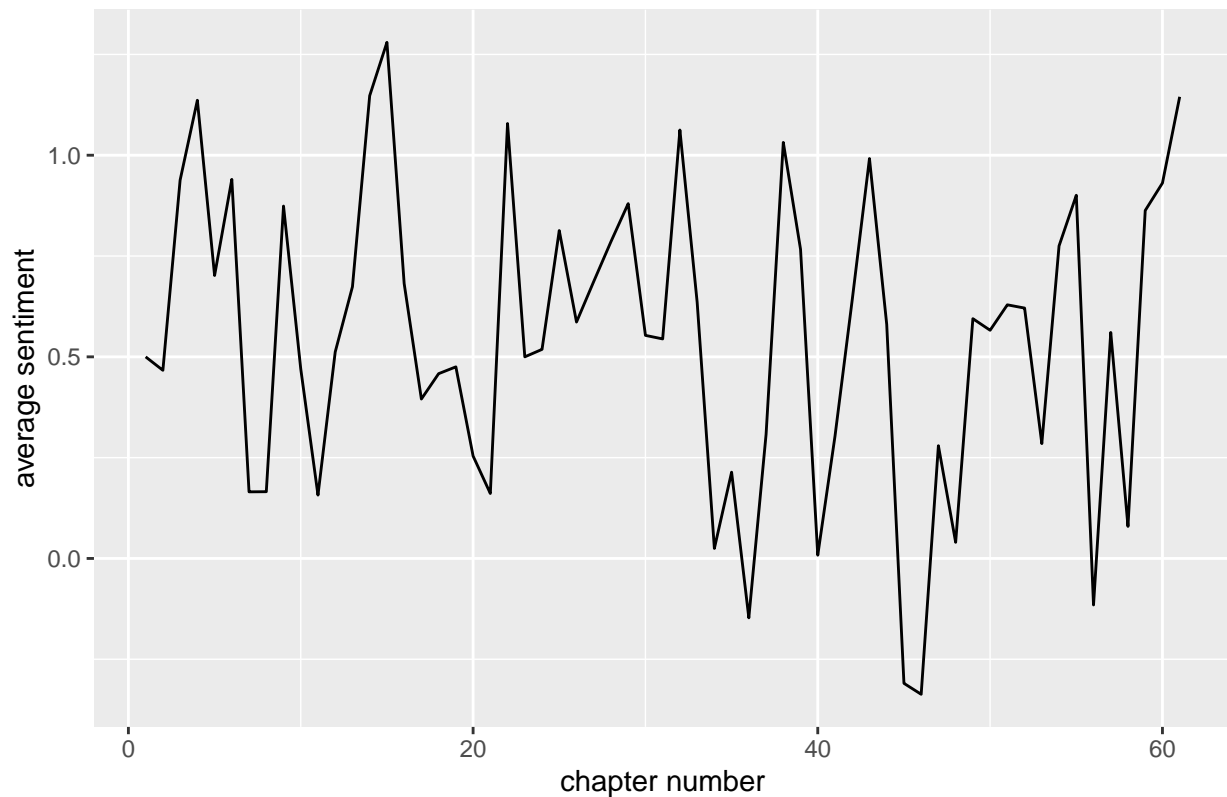
The sentiment varies a lot throughout the novel Pride & Prejudice so there is not an obvious sentiment uptrend or downtrend as the novel progresses.

Question 6: (1 pt)

After investigating how the sentiments change over the chapters, did the data match your expectations or not? If the data differed from your expectation, provide a possible explanation for why the data differed from what you expected.

The data does not meet our expectation because we thought there would be a clear trend in sentiment whether it was an uptrend or downtrend but the graph shows us that there was a lot of variance in the average sentiment as the novel progressed. One possible reason for this might be because the author might not want to use exclusively negative or positive words as the novel/novel series is a critique so she may want to address both the good and bad aspects of the topics which requires a mixture of both negative and positive sentiment words.

Formatting: (1 pt)

Make sure the names of all group members are included at the beginning of the document.

Knit your file! You can knit into pdf directly or into html. Once it knits in html, click on `Open in Browser` at the top left of the window pops out. Print your html file into pdf from your browser.

Any issue? Ask other classmates or TA!

Finally, remember to select pages for each question when submitting your pdf to Gradescope and to identify your group members.