

# Lab 01

Anonymous

2023-08-23

## Text Preliminaries

```
little_women <- readLines("./data/little_women.txt")
lines <- length(little_women)
lw_clean <- little_women[little_women != ""]
lw_text <- paste(lw_clean, collapse = " ")
```

## Lab Answers

1. There are 47 chapters and 21,022 lines in Little Women.
2. Done in the above.
3. Chapter splitting done below:

```
lw_chaps <- strsplit(lw_text, "CHAPTER ")
lw_chap <- lw_chaps[[1]] #Why does this thing count the blurb at the start with the Project Gutenberg
#as a chapter?
lw_chapters <- lw_chap[2:length(lw_chap)]

check_length <- length(lw_chapters)
```

Note that the length of `lw_chapters` reflects the correct number of chapters. The only thing I don't understand is why I have to remove the first element from my list, as the phrase "CHAPTER" does not appear in the Project Gutenberg blurb, but seems to have been included in `lw_chap`.

## Counting and Storing Appearances

```
library(stringr) #loading the stringr library
#counting appearances in each chapter
count_Amy <- str_count(lw_chapters, "Amy")
count_Beth <- str_count(lw_chapters, "Beth")
count_Jo <- str_count(lw_chapters, "Jo")
count_Laurie <- str_count(lw_chapters, "Laurie")
count_Meg <- str_count(lw_chapters, "Meg")

#placing cumulative counts in a data frame for easier plotting
counting <- data.frame(
  cumul_count_Amy = cumsum(count_Amy),
  cumul_count_Beth = cumsum(count_Beth),
  cumul_count_Jo = cumsum(count_Jo),
  cumul_count_Laurie = cumsum(count_Laurie),
  cumul_count_Meg = cumsum(count_Meg)
)
```

Required Plot

### cumulative counts of character appearances

