Text analysis can also be used to discover patterns in writing, and for a writer, may be helpful in discovering when we depend too much on certain words and phrases. For today’s demonstration, I read in my (still in-progress) novel – a murder mystery called *Killing Mr. Johnson* – and did the same type of text analysis I’ve been demonstrating in recent posts.

To make things easier, I copied the document into a text file, and used the read\_lines and tibble functions to prepare data for my analysis.

setwd("~/Dropbox/Writing/Killing Mr. Johnson")  
  
library(tidyverse)

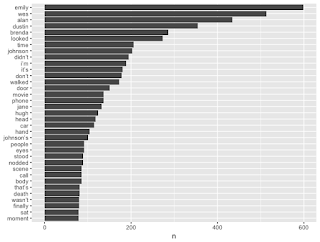
KMJ\_text <- read\_lines('KMJ\_full.txt')  
  
KMJ <- tibble(KMJ\_text) %>%  
 mutate(linenumber = row\_number())

I kept my line numbers, which I could use in some future analysis. For now, I’m going to tokenize my data, drop stop words, and examine my most frequently used words.

library(tidytext)  
KMJ\_words <- KMJ %>%  
 unnest\_tokens(word, KMJ\_text) %>%  
 anti\_join(stop\_words)

## Joining, by = "word"

KMJ\_words %>%  
 count(word, sort = TRUE) %>%  
 filter(n > 75) %>%  
 mutate(word = reorder(word, n)) %>%  
 ggplot(aes(word, n)) +  
 geom\_col() + xlab(NULL) + coord\_flip()

[](https://i2.wp.com/4.bp.blogspot.com/-GjsxpFsisa0/W3nyypJA0vI/AAAAAAAANco/2z0pPhkqfeszs8EwOeHFrlwPoHg6cCfzQCLcBGAs/s1600/words.png?ssl=1)

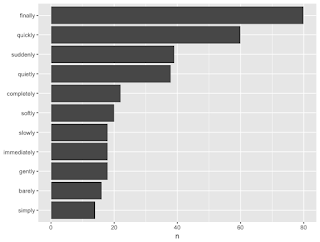
Fortunately, my top 5 words are the names of the 5 main characters, with the star character at number 1: Emily is named almost 600 times in the book. It’s a murder mystery, so I’m not too surprised that words like “body” and “death” are also common. But I know that, in my fiction writing, I often depend on a word type that draws a lot of disdain from authors I admire: adverbs. Not all adverbs, mind you, but specifically (pun intended) the “-ly adverbs.”

ly\_words <- KMJ\_words %>%  
 filter(str\_detect(word, ".ly")) %>%  
 count(word, sort = TRUE)  
  
head(ly\_words)

## # A tibble: 6 x 2  
## word n  
##   
## 1 emily 599  
## 2 finally 80  
## 3 quickly 60  
## 4 emilyâ€s 53  
## 5 suddenly 39  
## 6 quietly 38

Since my main character is named Emily, she was accidentally picked up by my string detect function. A few other top words also pop up in the list that aren’t actually -ly adverbs. I’ll filter those out then take a look at what I have left.

filter\_out <- c("emily", "emily's", "emilyâ€s","family", "reply", "holy")  
  
ly\_words <- ly\_words %>%  
 filter(!word %in% filter\_out)  
  
ly\_words %>%  
 filter(n > 10) %>%  
 mutate(word = reorder(word, n)) %>%  
 ggplot(aes(word, n)) +  
 geom\_col() + xlab(NULL) + coord\_flip()

[](https://i1.wp.com/4.bp.blogspot.com/-piiSDtlVYuw/W3ny-syG7LI/AAAAAAAANcs/WpYXho_Ao_Ykt4v9nLSx8CfF4-W8iiXQgCLcBGAs/s1600/ly_words.png?ssl=1)

I use “finally”, “quickly”, and “suddenly” far too often. “Quietly” is also up there. I think the reason so many writers hate on adverbs is because it can encourage lazy writing. You might write that someone said something quietly or softly, but is there a better word? Did they whisper? Mutter? Murmur? Hiss? Did someone “move quickly” or did they do something else – run, sprint, dash?

At the same time, sometimes adverbs are necessary. I mean, can I think of a complete sentence that only includes an adverb? Definitely. Still, it might become tedious if I keep depending on the same words multiple times, and when a fiction book (or really any kind of writing) is tedious, we often give up. These results give me some things to think about as I edit.

Still have some big plans on the horizon, including some new statistics videos, a redesigned blog, and more surprises later! Thanks for reading!