

Text Analytics exploratory problem I

```
review.words <- reviews %>%
  unnest_tokens(word, review_text) %>%
  mutate(word = str_extract(word, "[a-z']+"),
         word.stem = wordStem(word))

review.words <- review.words %>%
  anti_join(stop_words, by = "word")

word.count <- review.words %>%
  group_by(word) %>%
  summarize(word.count = n()) %>%
  arrange(desc(word.count))

new_stop_words <- tibble(word = c("book", "read", "story", "books", "reading",
                                   "kid", "children", "kids", "children's",
                                   "lots", "little", "written", "writting",
                                   "writes", "adults", "lots", "http", "found",
                                   "words", "word", "find", "person", "pages"))

review.words <- review.words %>%
  anti_join(new_stop_words, by = "word")

word.count2 <- review.words %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count))

review.words$word <- gsub("loved", "love", review.words$word)
review.words$word <- gsub("characters", "character", review.words$word)
review.words$word <- gsub("friends", "friend", review.words$word)
review.words$word <- gsub("stories", "story", review.words$word)
review.words$word <- gsub("pictures", "picture", review.words$word)
review.words$word <- gsub("stars", "star", review.words$word)
review.words$word <- gsub("students", "student", review.words$word)
```

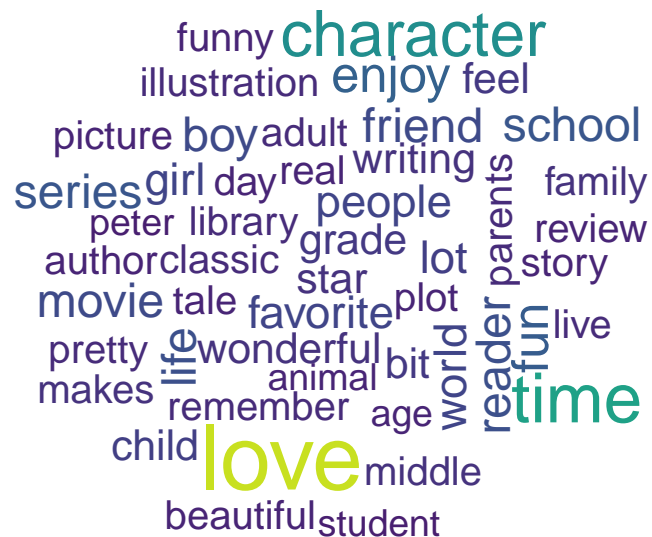
```

review.words$word <- gsub("animals","animal", review.words$word)
review.words$word <- gsub("enjoyed","enjoy", review.words$word)
review.words$word <- gsub("boys","boy", review.words$word)
review.words$word <- gsub("girls","girl", review.words$word)
review.words$word <- gsub("times","time", review.words$word)
review.words$word <- gsub("readers","reader", review.words$word)
review.words$word <- gsub("lives","live", review.words$word)
review.words$word <- gsub("illustrations","illustration", review.words$word)

new_word_count <- review.words %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count))

wordcloud(words = new_word_count$word, freq = new_word_count$count,
          max.words = 50, colors = viridis(50), scale=c(3,1))

```



References

1. Resources are listed here. Links should be surrounded by `<>`.