

Reading in texts using the `gutenbergr` package

```
titles <- c("Twenty Thousand Leagues under the Sea", "The War of the Worlds")
books <- gutenberg_works(title %in% titles) %>%
  gutenberg_download(meta_fields = "title")

> books
# A tibble: 18,609 x 3
  gutenberg_id text title
      <int> <chr> <chr>
1         36 The War of the Worlds The War of the Worlds
2         36 "" The War of the Worlds
3         36 by H. G. Wells [1898] The War of the Worlds
4         36 "" The War of the Worlds
5         36 "" The War of the Worlds
6         36 " But who shall dwell in these worlds if they be" The War of the Worlds
7         36 " inhabited? . . . Are we or they Lords of the" The War of the Worlds
8         36 " World? . . . And how are all things made for man?--" The War of the Worlds
9         36 " KEPLER (quoted in The Anatomy of Melancholy)" The War of the Worlds
10        36 "" The War of the Worlds
# ... with 18,599 more rows
```

- In the above code I read in two books using the `gutenbergr` package which reads them from the Project Gutenberg library (containing over 58,000 free books). The object `books` contains the text of both books contained in the “text” column

```
text_waroftheworlds <- books %>%  
  filter(title == "The War of the Worlds") %>%  
  unnest_tokens(word, text) %>%  
  anti_join(stop_words)  
  
text_underthesea <- books %>%  
  filter(title == "Twenty Thousand Leagues under the Sea") %>%  
  unnest_tokens(word, text) %>%  
  anti_join(stop_words)
```

- I then filter by book title and ‘unnest’ the text of each book so that we have one column in each of two dataframes corresponding to the words in each book...