

goodreads

judging a book by its ~~cover~~ DATA

Emma Gillen IST 719

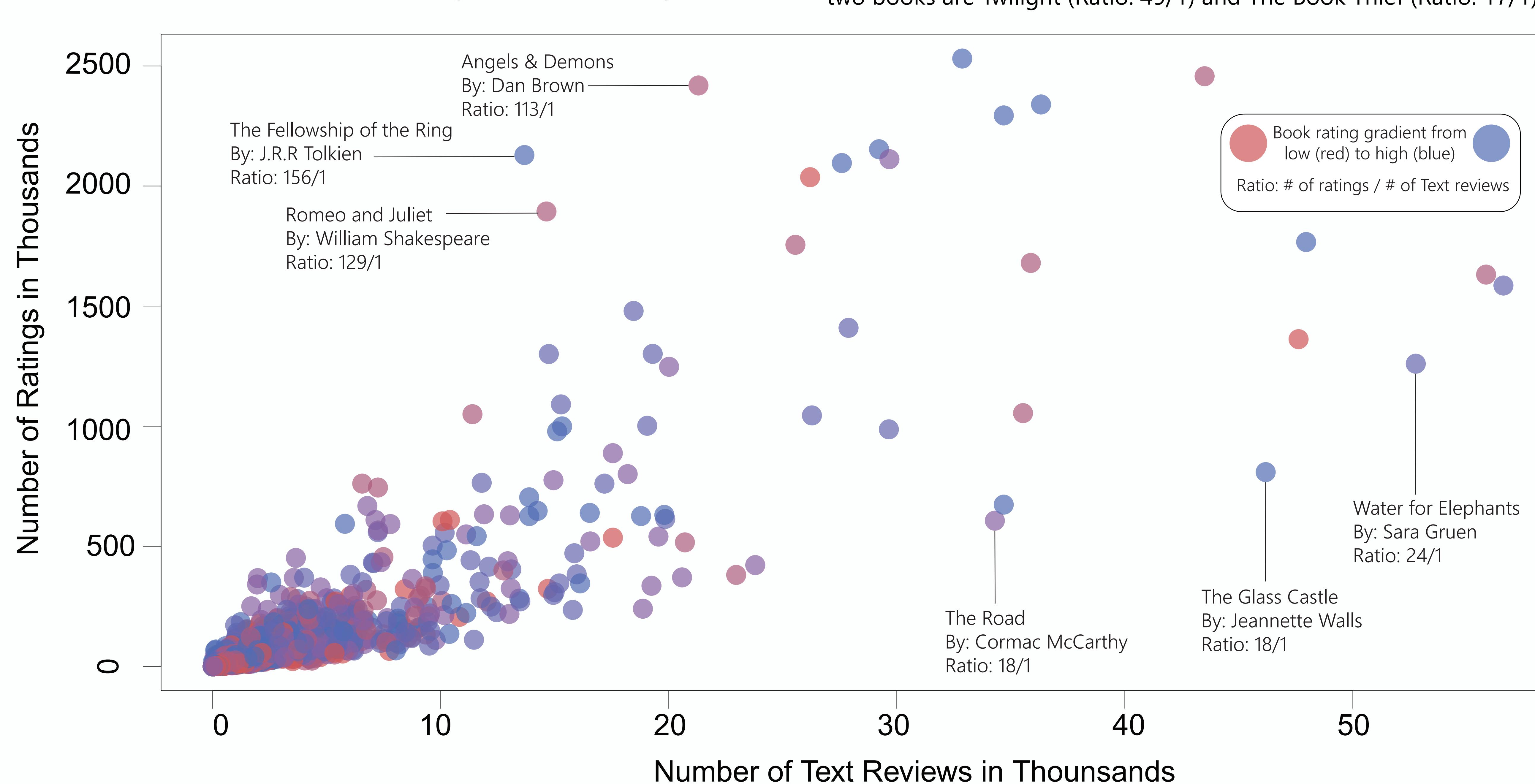
Story: The goodreads website contains entries on thousands of books as well as user reviews and ratings. Based on this user feedback, goodreads provides recommendations.

Motivation: Popular or trending media is often recommended. However, there is often a conflict between popularity and quality. This study focuses on what combination of factors can be considered to find a "good read".

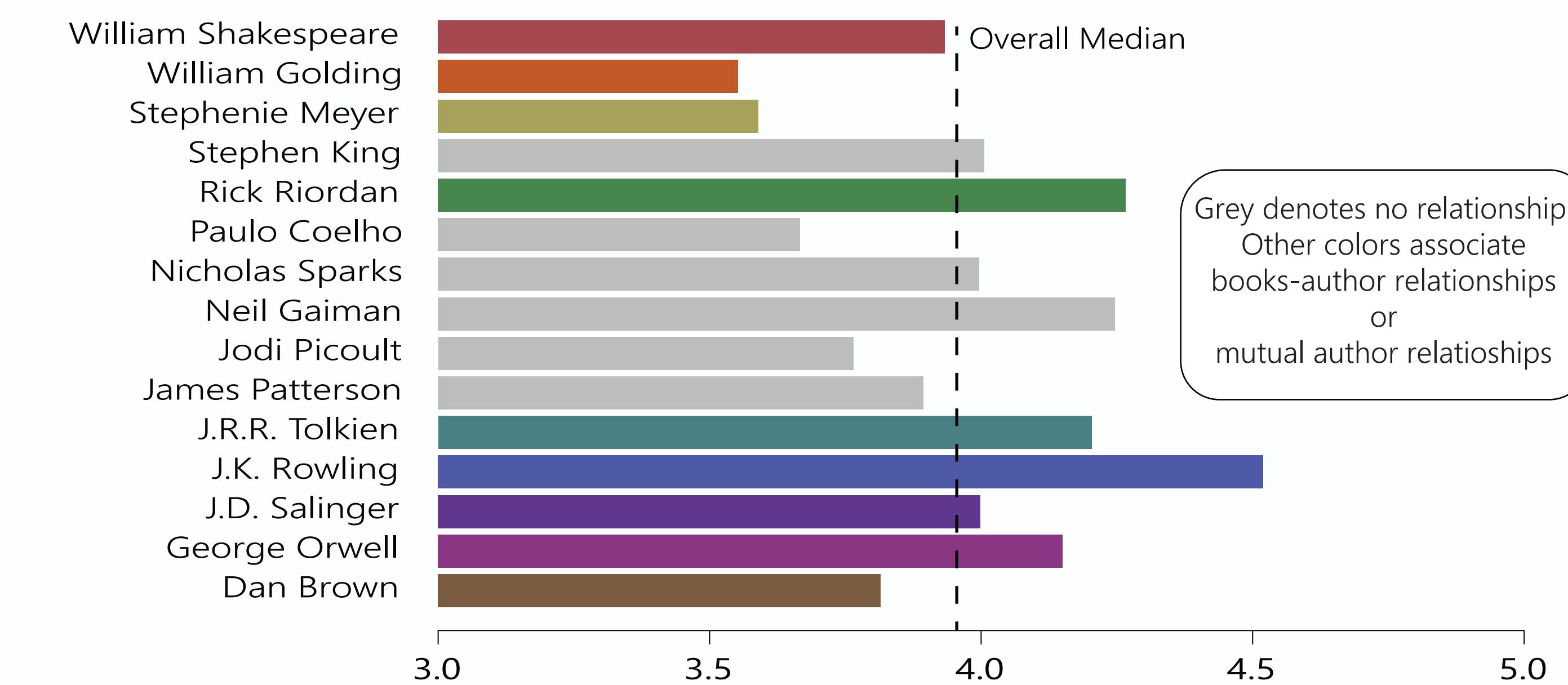
The Data: The dataset contains 11,131 rows and 12 columns. For analysis, books with less than 10 reviews (too obscure) or less than 100 pages in length (largely children's books and incorrect entries) were removed, leaving 9,584 rows. The key columns used for analysis include the title, author, average rating, number of ratings, number of written reviews, and page length.

Does the average rating have a relationship with the amount of feedback a book receives? An author?

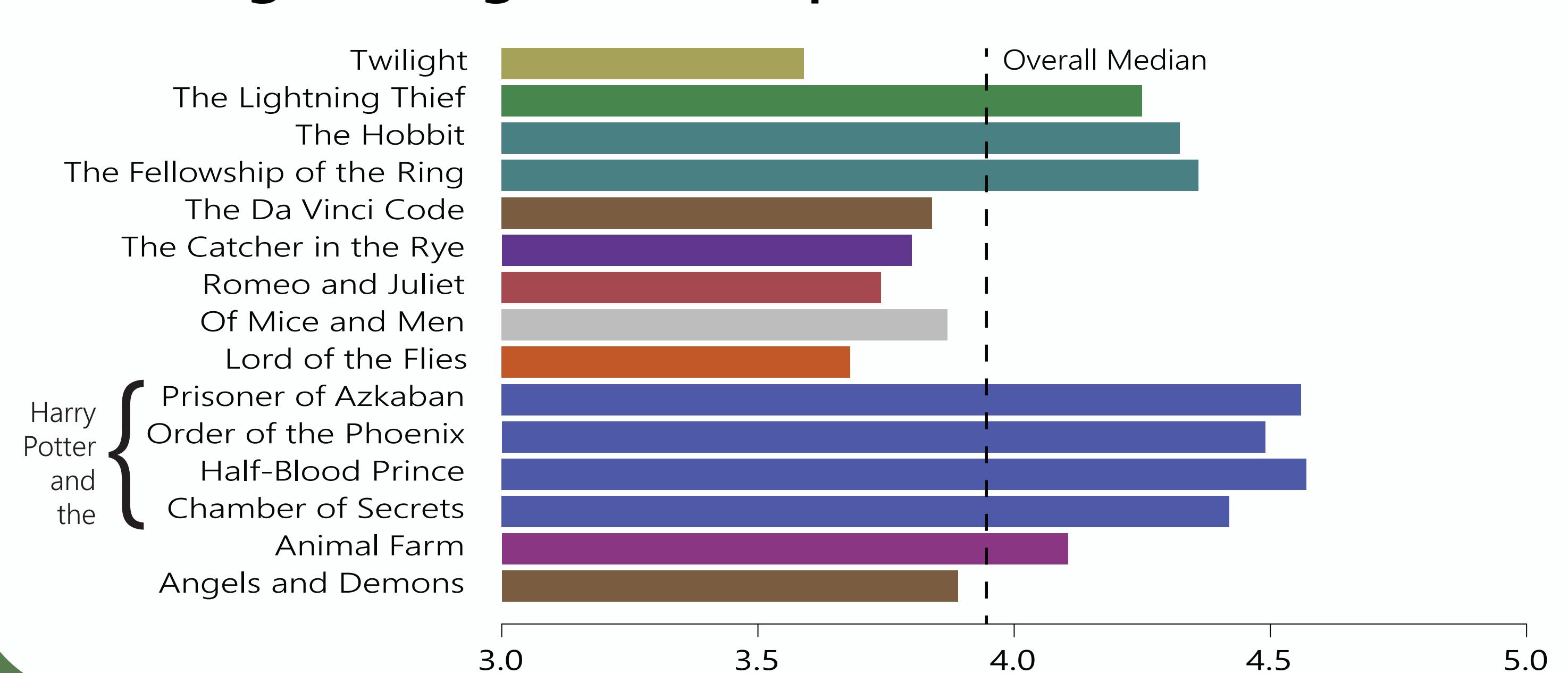
Review & Rating Counts by Score



Average Rating for the Top Fifteen Most Rated Authors



Average Rating for the Top Fifteen Most Rated Titles

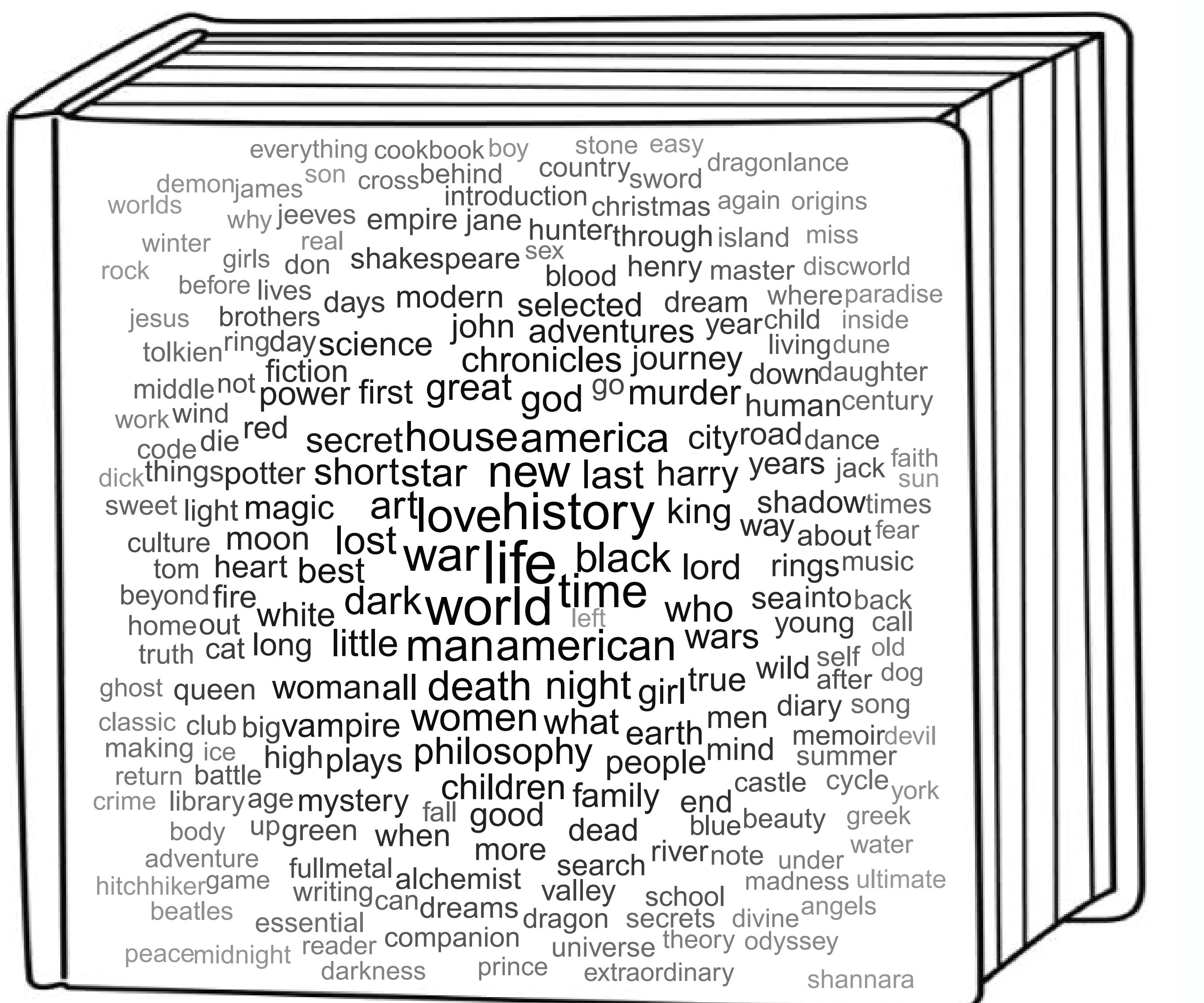


What are these books about?

Most Common Words in Book Titles

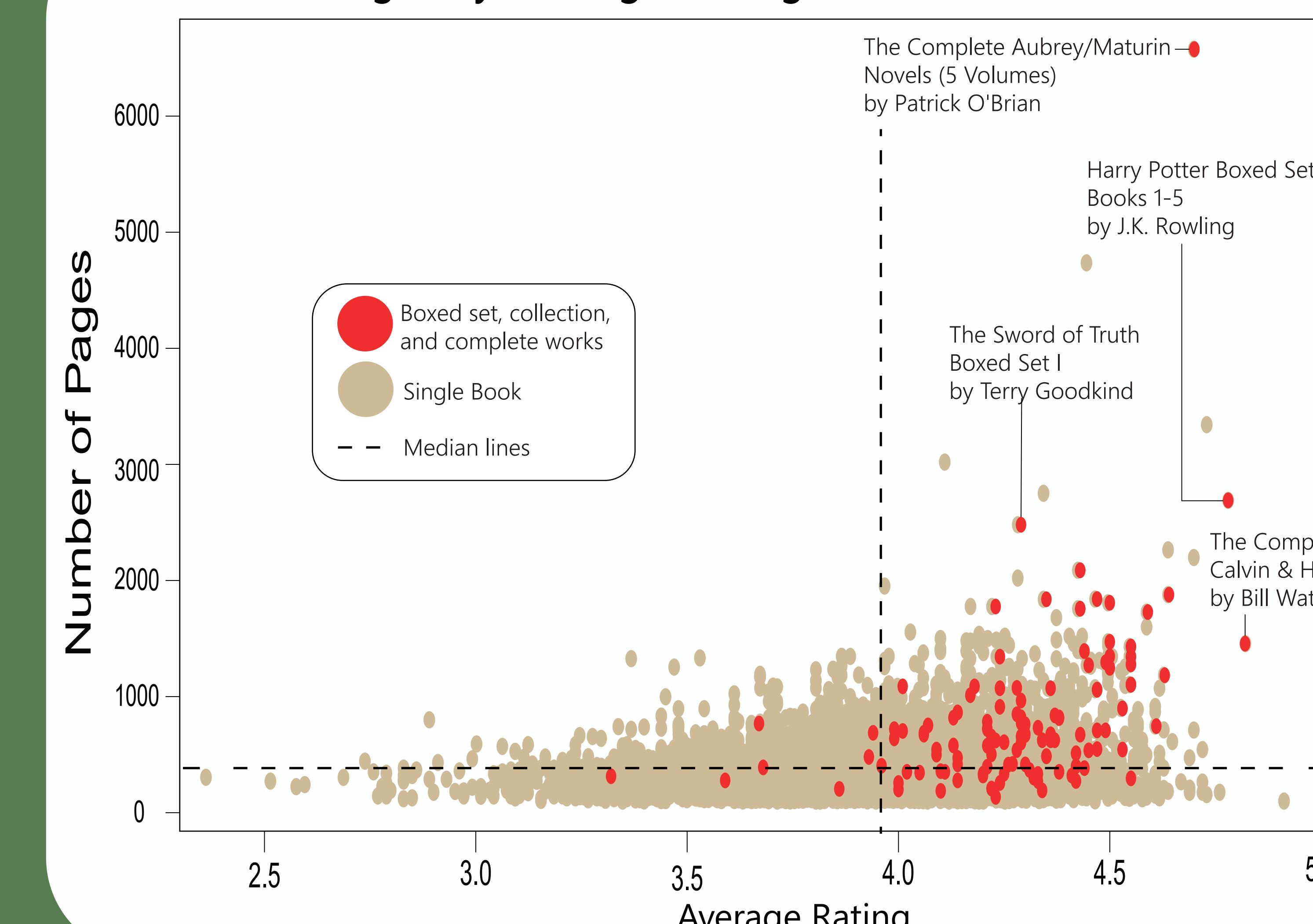
Function words, or words with little meaning were filtered from the word cloud.

Words that occurred less than twenty times were filtered.

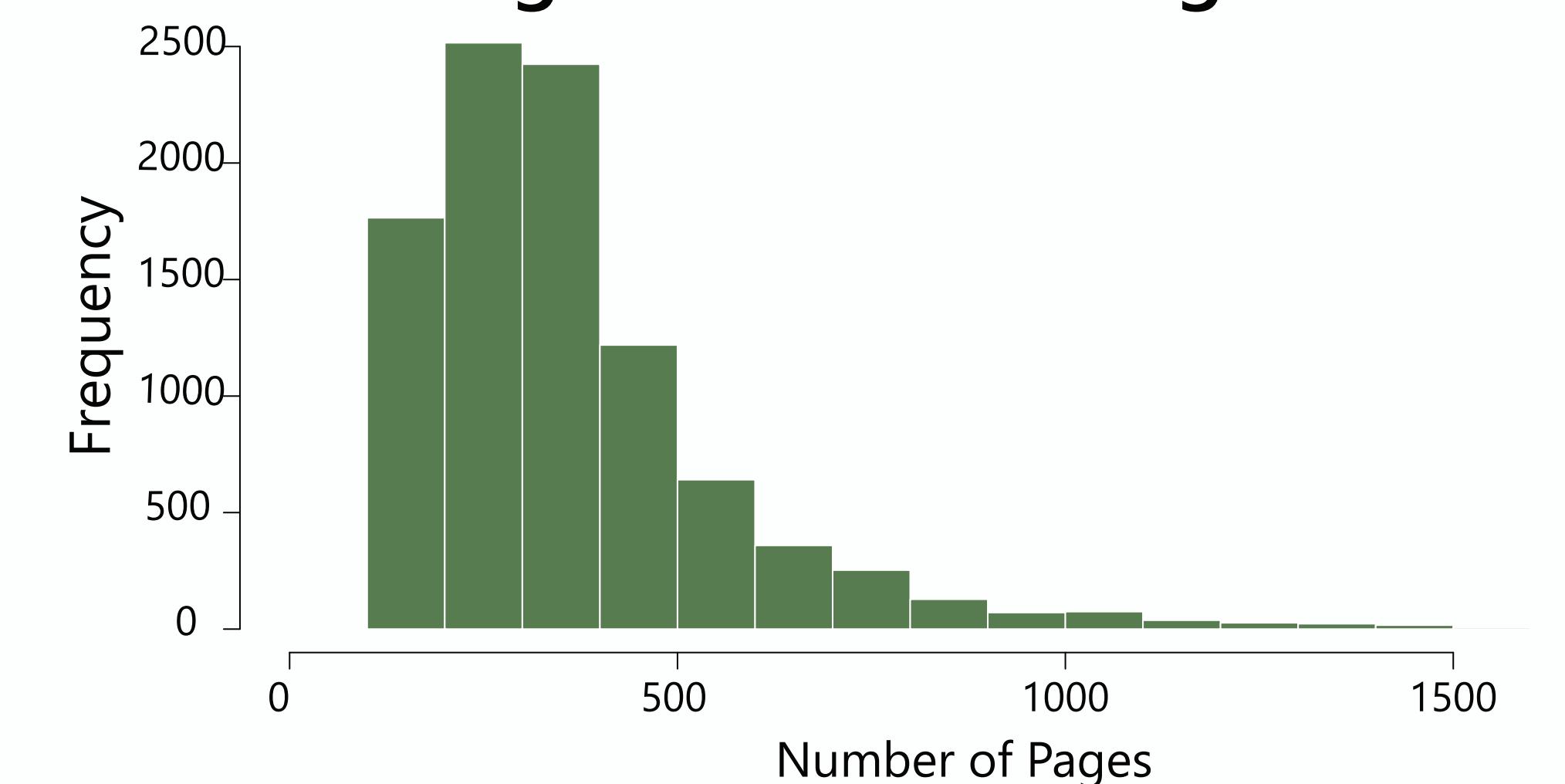


Does the page length have a relationship with rating?

Book Length by Average Rating



Histogram of Book Lengths



Density of Average Ratings

