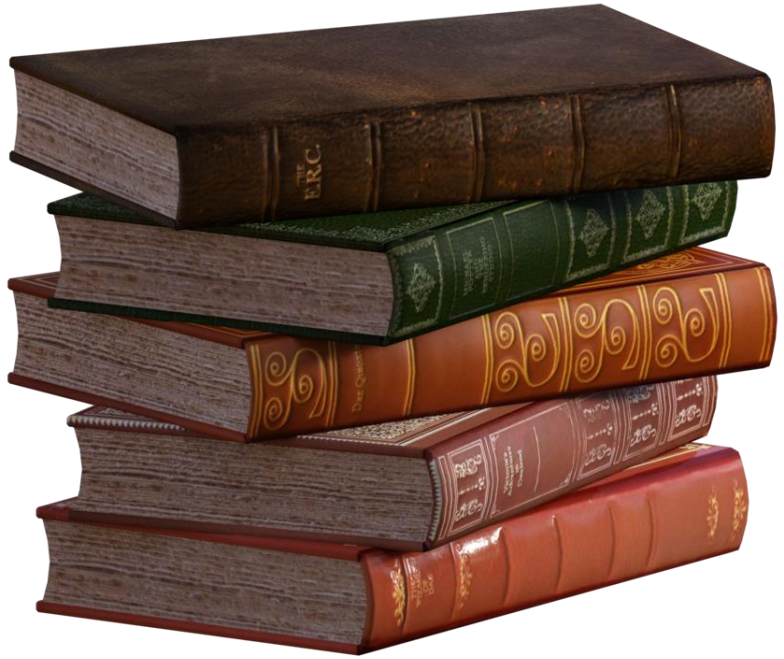
Goodreads Data Analysis

SHIVANI BHAKTA, PAYAL SINGH, YICHEN ZHANG, WENDA CHEN

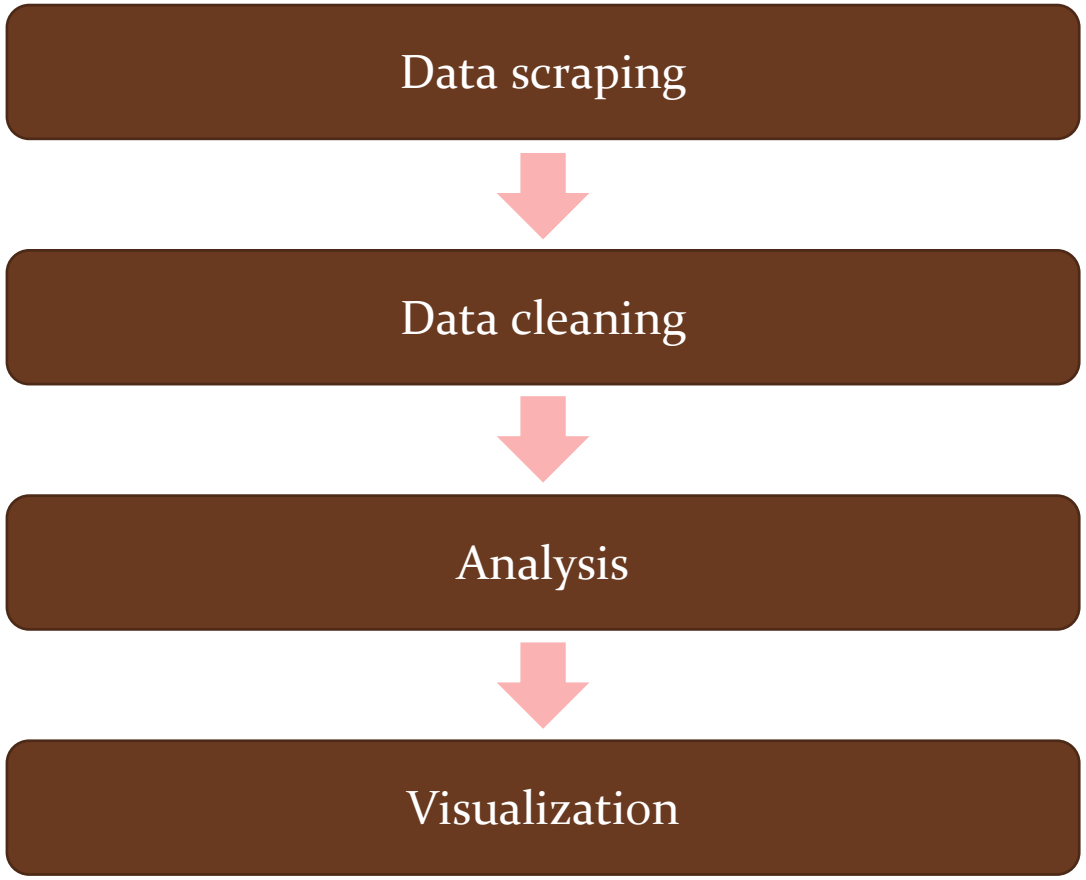
Data Overview

- Total books: 2.6 billion
- Our dataset: 24k books
- Goodreads API restriction: 1 call/sec

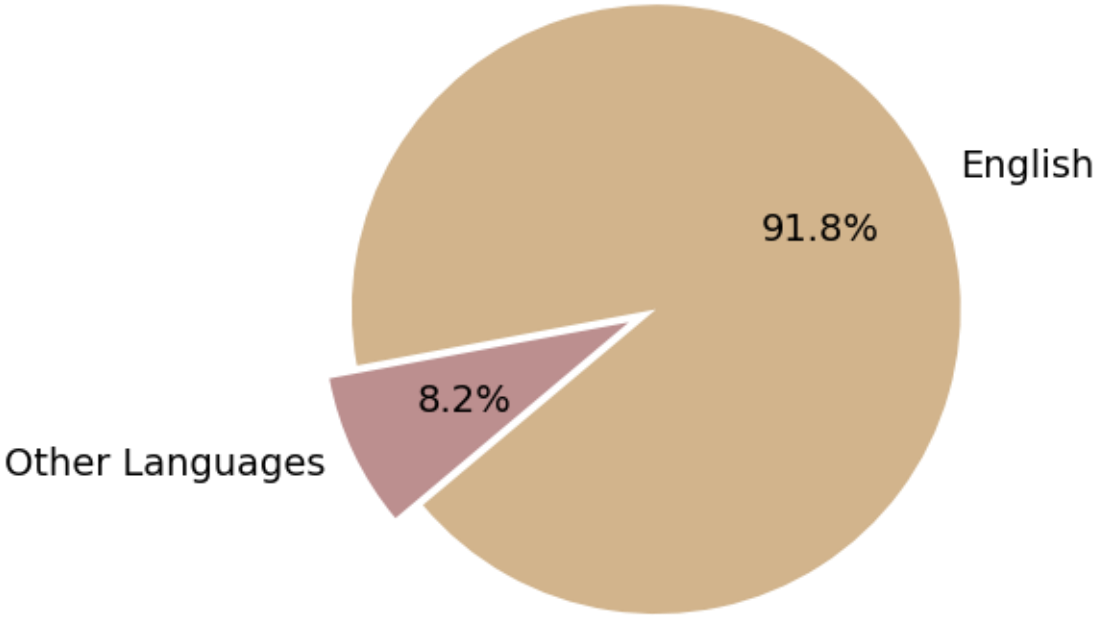


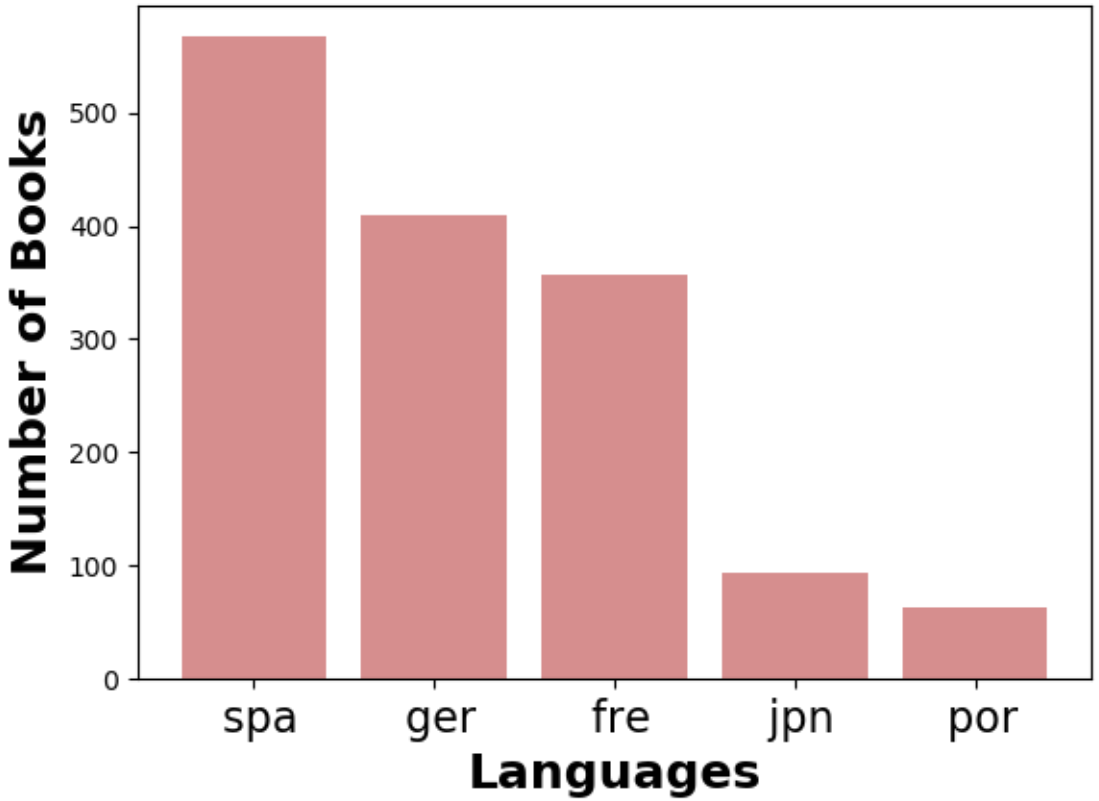
*it would take ~**83 years** to collect all of the data currently on Goodreads using the API (assuming no new books were added in that time)

Methodology

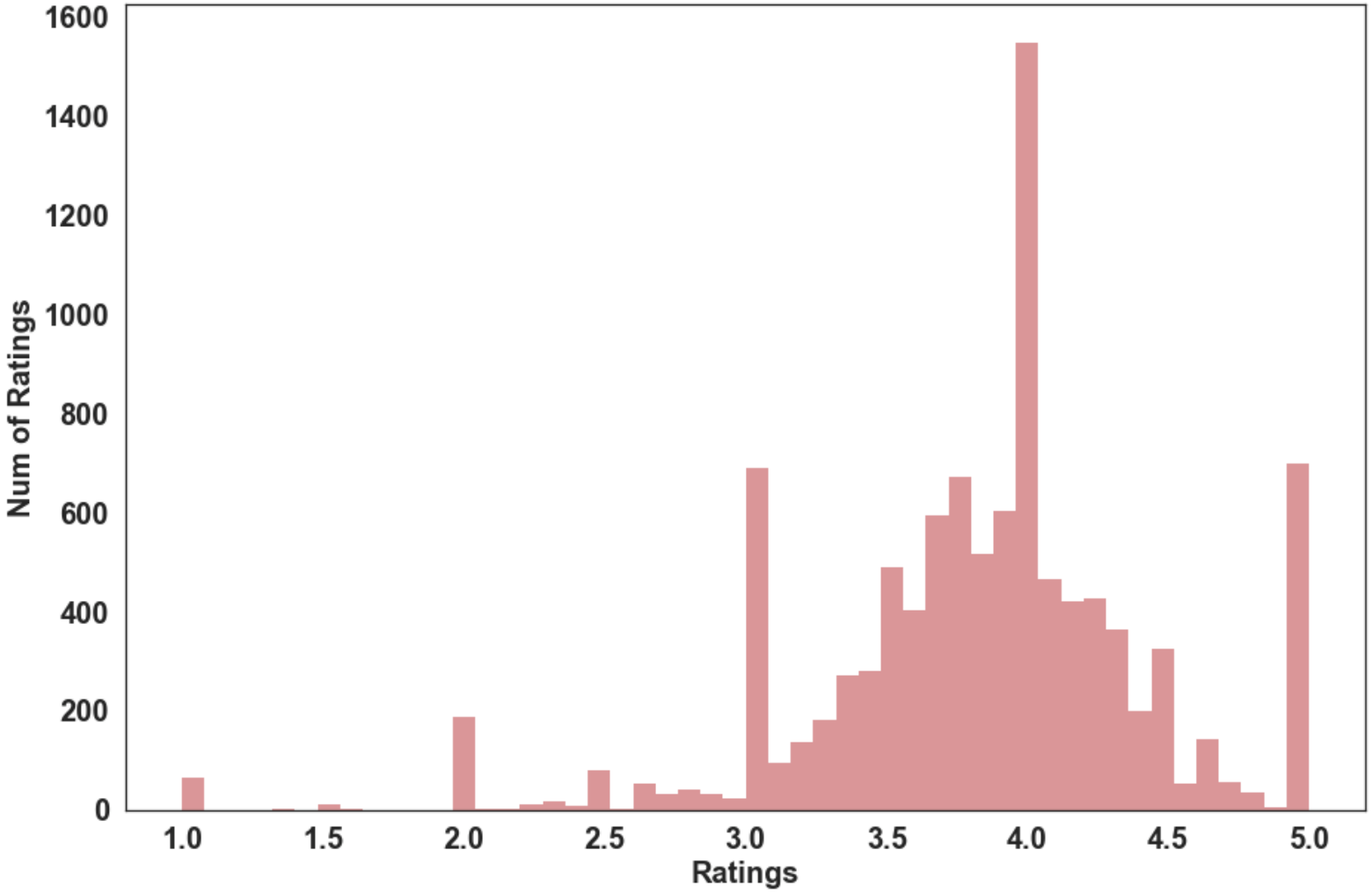


What languages are most common?

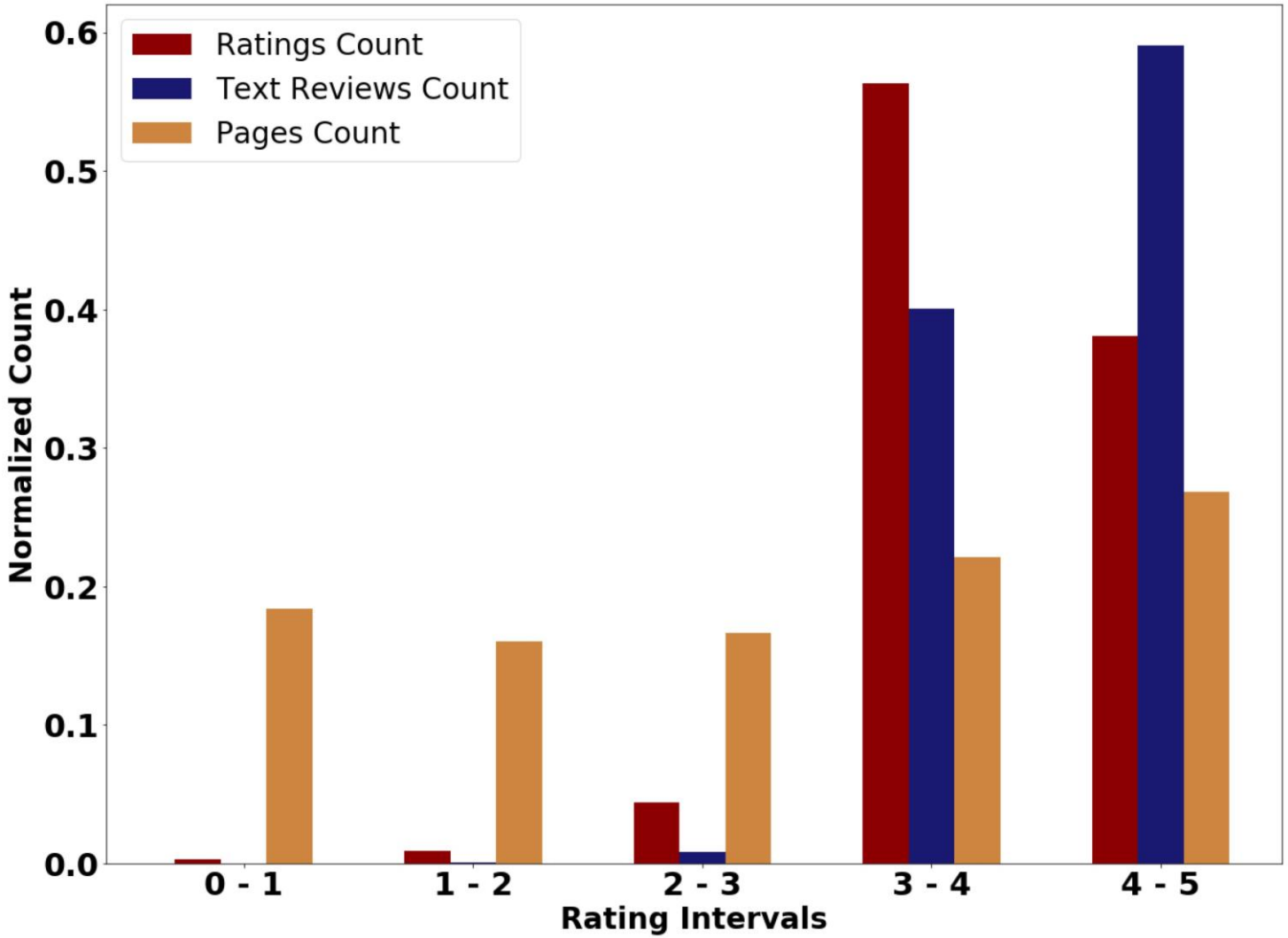




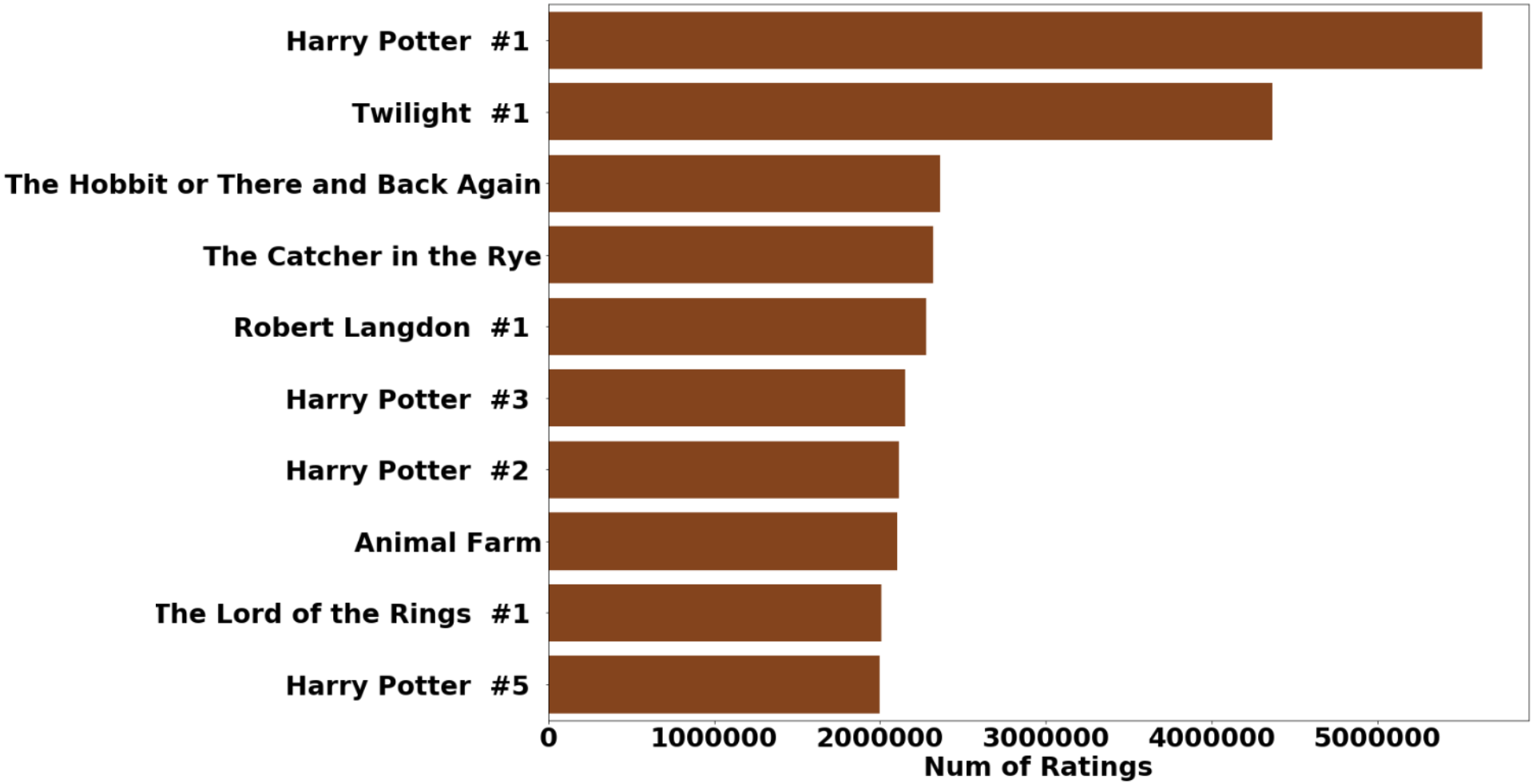
Rating distribution for all books in dataset



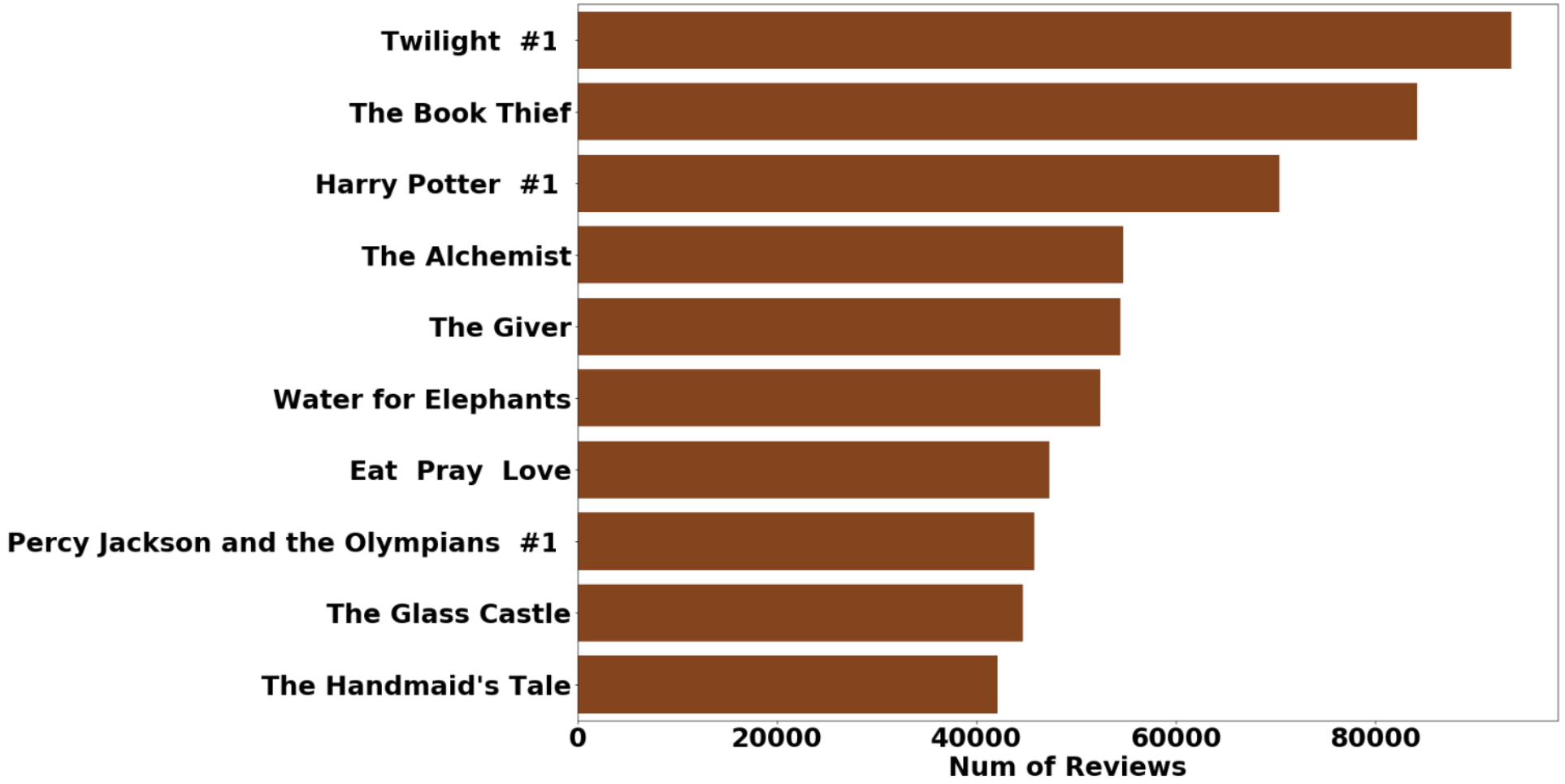
Rating vs. Rating count, Text Reviews count, Page count



What are the books with the highest number of ratings?

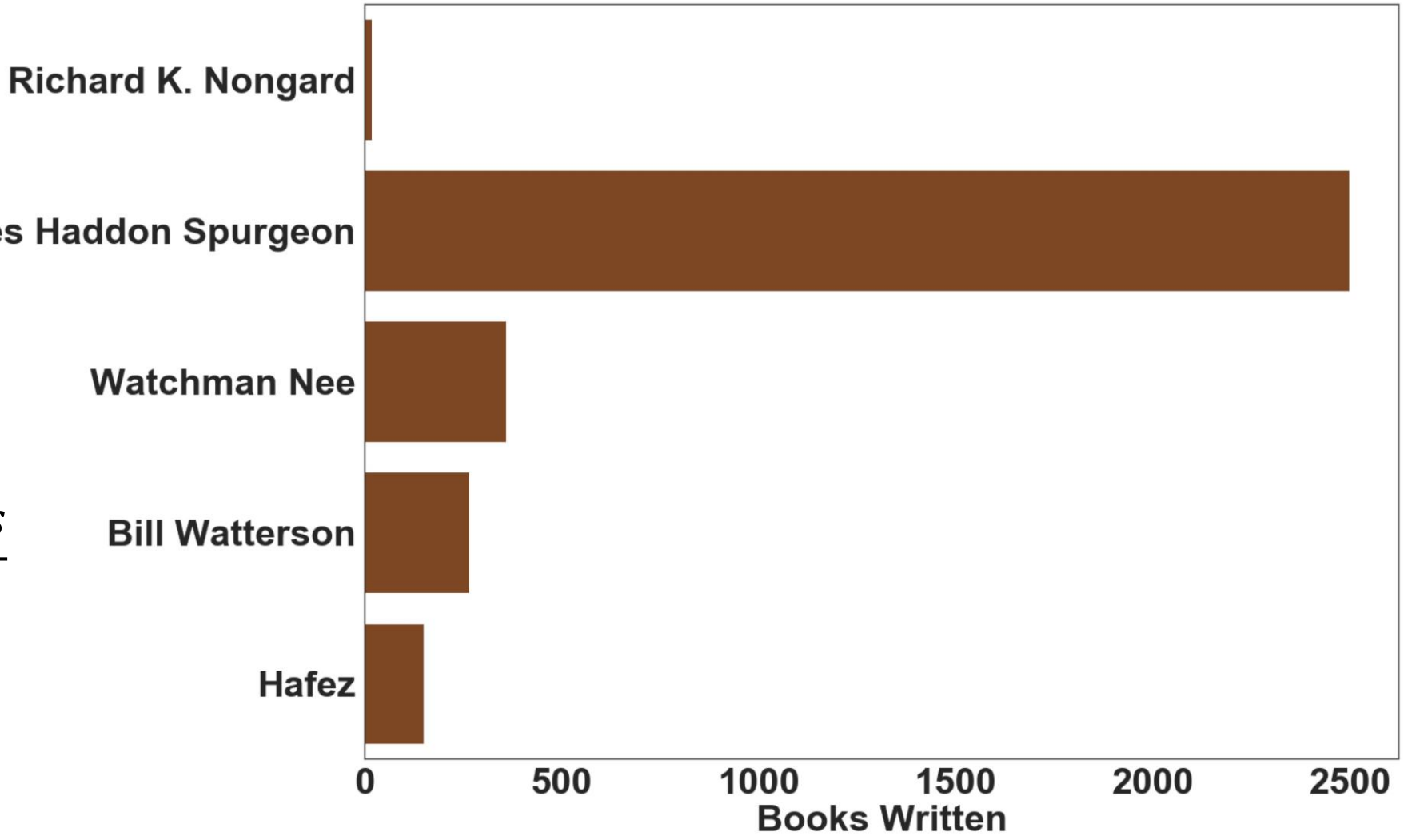


What are the books with the highest number of reviews?

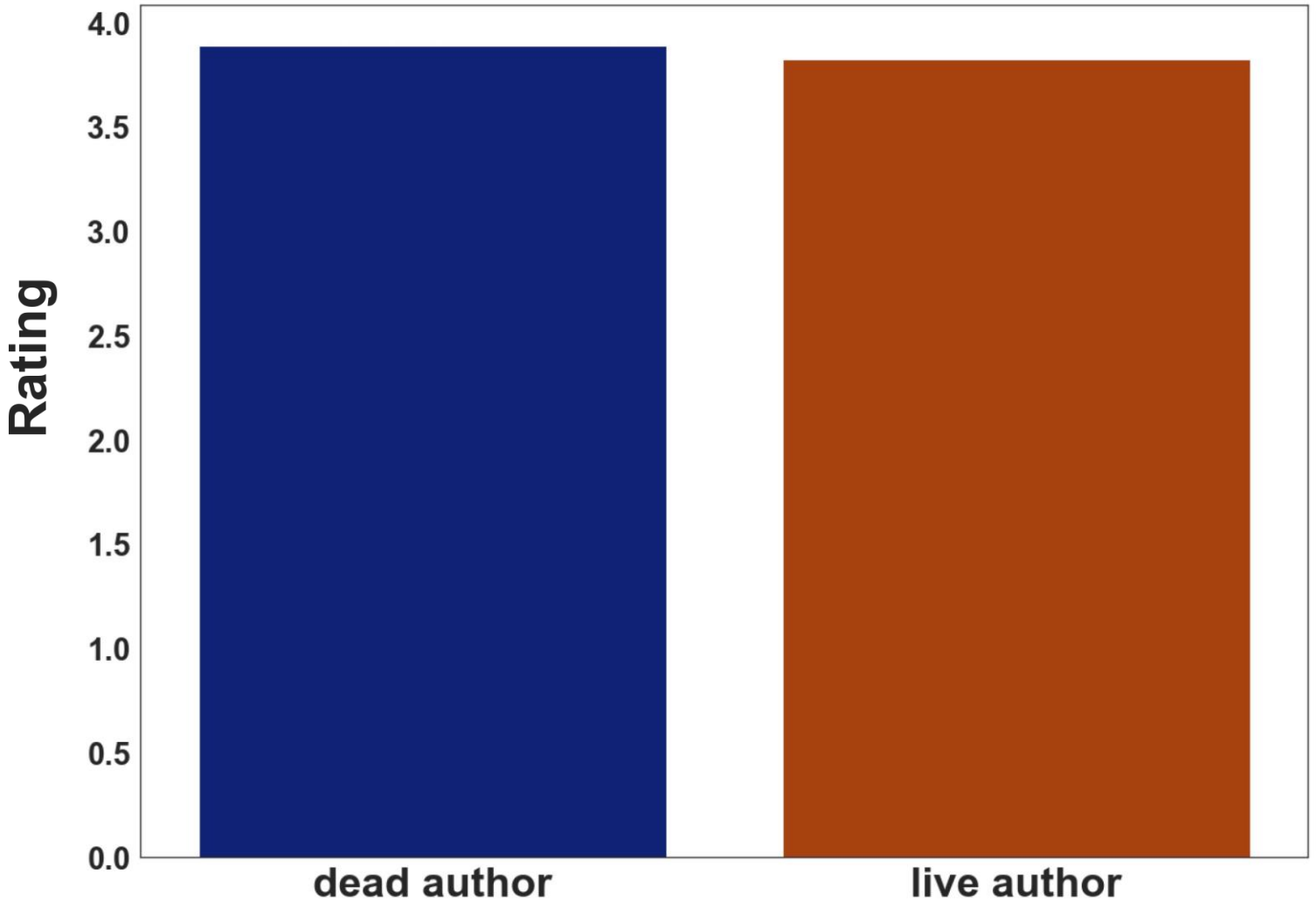


How many books did the top 5 rated authors write?

$$Author\ Rating = \frac{\sum Book's\ Ratings}{\#Books}$$



Best authors – alive or dead?



Conclusions

- English > other languages
- Young adult novels!
- Dead or alive, we like all authors

Thank you!

HAPPY THANKSGIVING