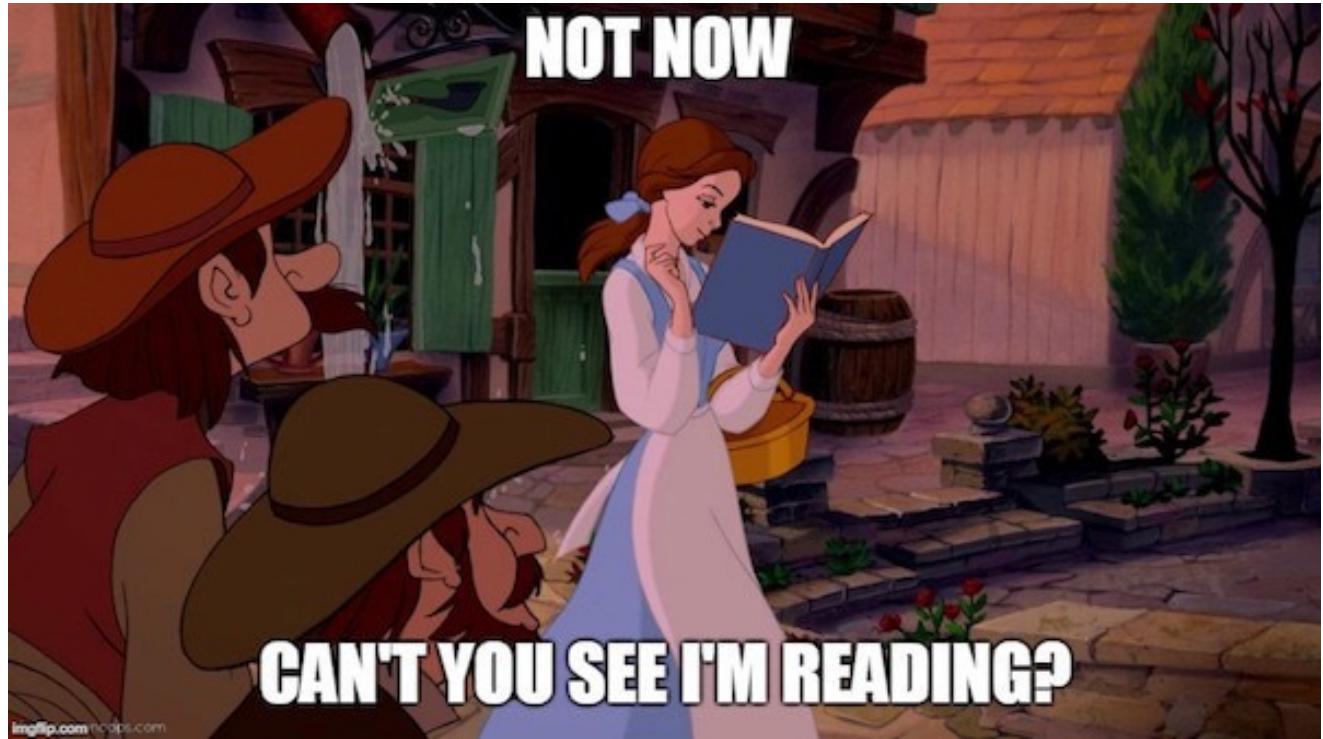


Good Reads on Goodreads

Melanie Ackerman



An Opportunity for Penguin Random House

- Expansive catalog of books spanning many genres and authors
- Wide net of readers, whose interests likely overlap in genre
- Harness data to help publishers with content decisions
- Translate reader preferences into a personalized book recommendation system



Proposed Solution

- Segment reader audiences using clustering algorithms
 - Could genres with overlapping audiences borrow content/styles from each other?
 - Next suggested read?
- Measures of success
 - Survey authors/publishers about content decisions based on data
 - Survey the users of recommendation system



Taking Advantage of Goodreads

- Public data from goodreads.com (2018):
 - 10K popular books
 - Book metadata (e.g. author, year, ratings, etc.)
 - Goodreads tags/shelves/genres
 - Books marked ‘to read’ by users
- Preliminary stats to investigate:
 - Top rated books vs. most popular books
 - Breakdown of genres



Risks and Assumptions

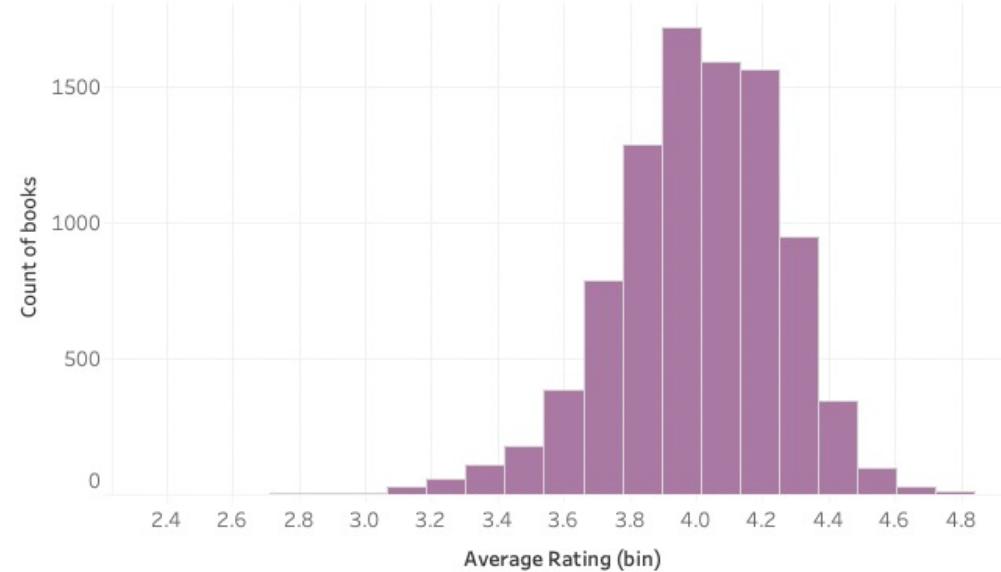
- Readers may not want genres to overlap more than they already do.
- Authors may not appreciate content guidance.
- Goodreads users may not represent general reader audience.
- Goodreads tags are user-created → no structure/rules for genres



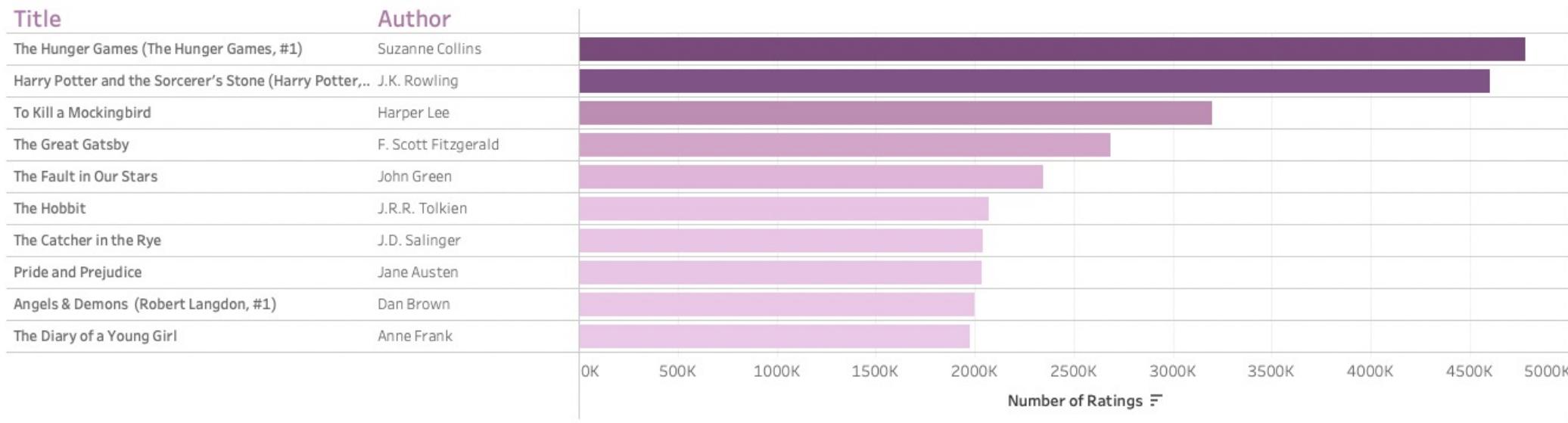
Top 10 Books by Rating

Title	Author	Avg. Rating
The Complete Calvin and Hobbes	Bill Watterson	4.82
Words of Radiance (The Stormlight Archive, #2)	Brandon Sanderson	4.77
Mark of the Lion Trilogy	Francine Rivers	4.76
It's a Magical World: A Calvin and Hobbes Collection	Bill Watterson	4.75
Harry Potter Boxset (Harry Potter, #1-7)	J.K. Rowling	4.74
There's Treasure Everywhere: A Calvin and Hobbes Collection	Bill Watterson	4.74
Harry Potter Collection (Harry Potter, #1-6)	J.K. Rowling	4.73
The Authoritative Calvin and Hobbes: A Calvin and Hobbes Tr...	Bill Watterson	4.73
The Indispensable Calvin and Hobbes	Bill Watterson	4.73
A Court of Mist and Fury (A Court of Thorns and Roses, #2)	Sarah J. Maas	4.72
Attack of the Deranged Mutant Killer Monster Snow Goons	Bill Watterson	4.72

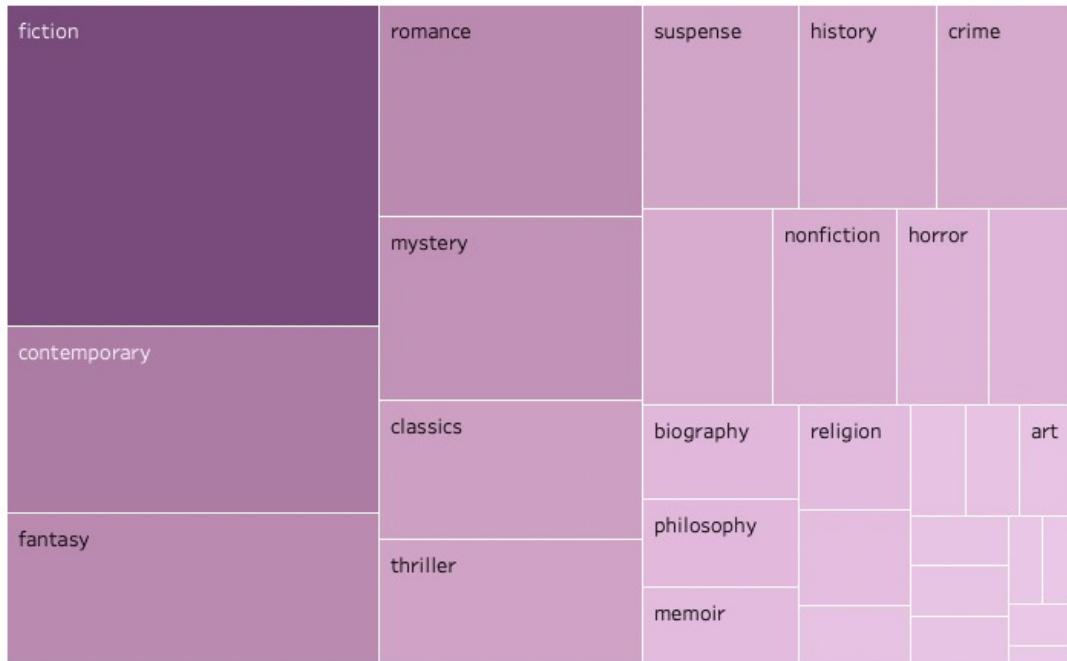
Avg. Ratings Distribution



Top 10 Most Popular Books



All Genres

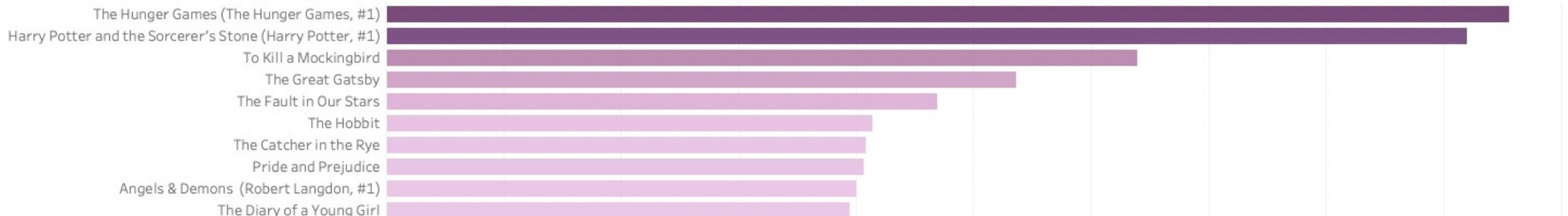


Fiction Sub-Genres



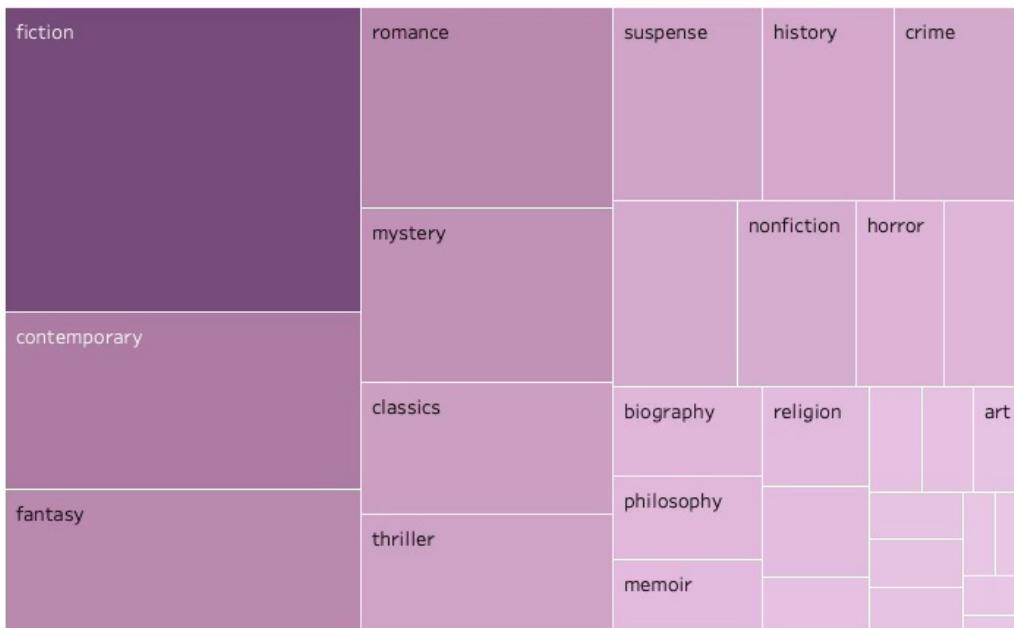
Top 10 Most Popular Books by Genre

Title



Ratings Count

All Genres

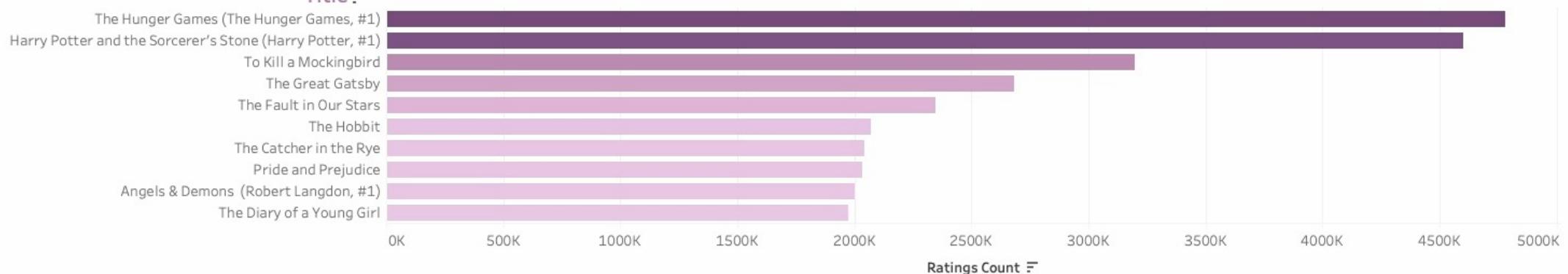


Fiction Sub-Genres



Top 10 Most Popular Books by Genre

Title



Moving Forward

- Devise rules for categorizing books
- Explore data options beyond goodreads
 - Potential propriety data from Penguin Random House
 - New York Times Best Sellers
 - Book sales data
- NLP to detect style/tone/structure in books
 - Would add information to reader preferences findings



"I declare after all
there is no enjoyment
like reading!"

- Jane Austen