

24 MAY 2022

Goodreads Data Analysis Project

An Analysis in Tableau

goodreads

Written by

Hailey Nguyen

GOOD READS

WHAT IS IT?

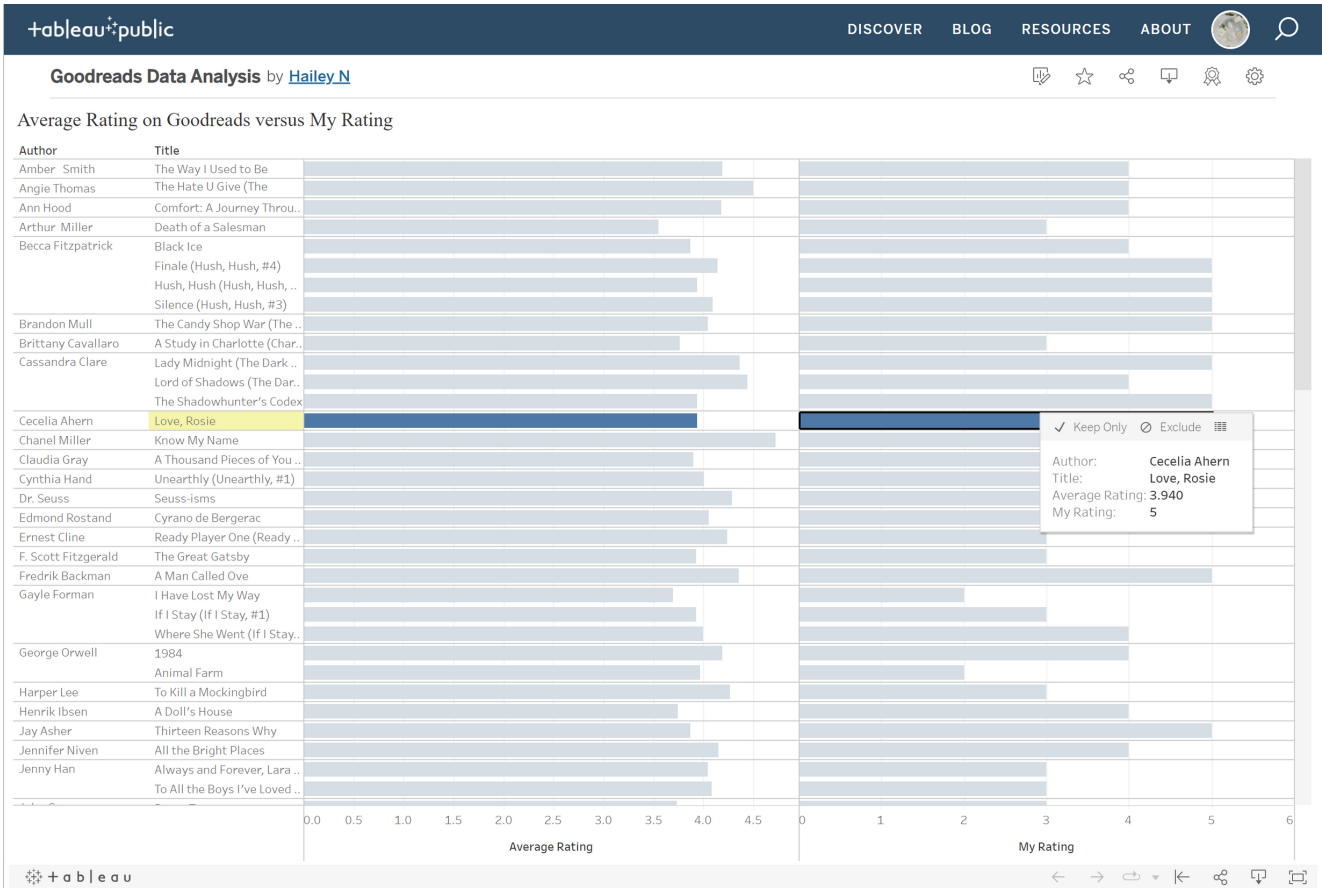
Goodreads is an online book logging platform that allows users to perform activities such as cataloging books they read or want to read, setting reading goals, writing book reviews, and even tracking statistics about reading books. Users can befriend others on the site and participate in polls and discussions about books. In addition to its website presence, the platform is available in the Google Play Store and Apple App Store.

PURPOSE

I created this project because I love reading books and I use Goodreads quite frequently to track books I read. After using Tableau and other statistical applications to work with data in my university courses, I wanted to apply the concepts that I learned to my own personal projects. Thus, the idea of using data from my own Goodreads profile came to mind. I hope this project brings some interest in the realm of books and data analysis. Let's begin!

I began this project by exporting the Goodreads data from my Goodreads profile and uploading it into Tableau. From the Goodreads data, I wanted to determine my average rating of books I read compared to Goodreads users' average ratings, identify the average rating of books in the data set on Goodreads, visualize books I have read versus books I want to read along with their read counts (the number of times I read a book), identify the best publishers on Goodreads based on average ratings, and recommend books based on book page length. Before analyzing the data in Tableau, I removed all 'null' values in my analyses to keep the data cohesive and reduce errors in my analysis.

In this [bar graph](#) comparing my average rating of books to that of Goodreads users' average ratings, one can identify that I really disliked reading *As I Lay Dying* as I gave the book one star. Since I had to read this book for an English class and did not enjoy any part of it, I graced it with a one-star rating . . . I assume any student reading this now could say the same (unless they actually did enjoy reading it). Most of my low rated books (one or two-star ratings) were required readings for my high school English class with the exception of Gayle Forman's *I Have Lost my Way*, Lance Rubin's *Denton Little's Deathdate*, Robert Cormier's *The Chocolate War*, and Stephen Chbosky's *The Perks of Being a Wallflower*.



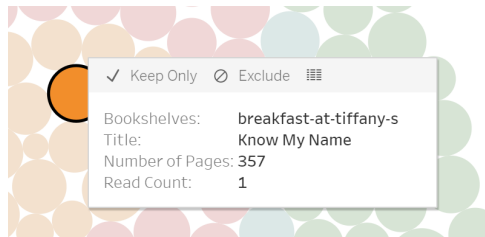
In the photo above, I highlighted one of my favorite books I read (Love, Rosie). I gave the book a five-star rating. The average Goodreads rating for this book is roughly 3.9. Other books I enjoyed reading were Fredrik Backman's *A Man Called Ove*, Chanel Miller's *Know My Name*, and Leigh Bardugo's *Six of Crows*.

Moving on to this [chart](#), one can see the highest average rated books and the lowest average rated books. I am a bit surprised to see Earnest Cline's *Ready Player Two* at the bottom of the list. I assumed the book would possess a higher average rating in this data set, though it appears that readers did not enjoy the sequel to *Ready Player One*.

I am also surprised to see the highest average rated book in this data set is *Vodka Soup for the Widowed Soul: Stories of Grief, Alcohol, Infidelity, Cursing, and Hope* with an average of 4.7. Judging from the large number of positive reviews on Goodreads, I may likely read this soon!

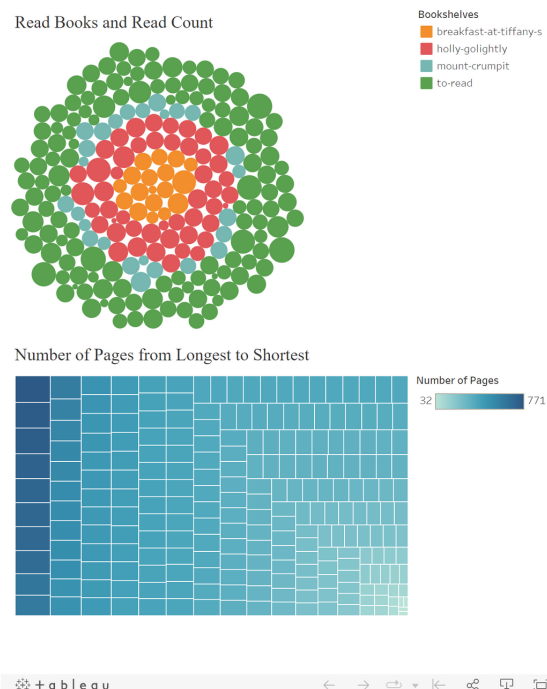
According to this [chart](#). The highest rated publisher on Goodreads from this data set is Atria Books with an average rating of 36.79 and 3,155 pages followed closely by HarperTeen with an average rating of 35.99 and 3,444 pages total. The lowest rated publisher is Knopf Publishing Group with an average rating of 3.44 with 336 pages.

Looking at the [bubble chart](#) below, one can see that I have many books I want to read based on the green bubbles. Hovering over each bubble will display something like this:



Hovering over the majority of the bubbles will show that I typically read books only once.

Lastly, the [tree map](#) below the bubble chart depicts the sum of all pages for all books in this data set. If one wanted to read a book with a substantial number of pages, then this map suggests *The Goldfinch* may be of interest with a sum of 771 pages. In contrast, someone who would prefer to read a short book may find *Seuss-isms* ideal with a total page length of 32 pages.



CLOSING THOUGHTS

I hope you enjoyed reading this analysis and found some new books to add to your 'to-read' list. This was an enjoyable experience for me, and I had a ton of fun dissecting my Goodreads data!

LINKS

- [Goodreads data set](#)
- My [Public Tableau](#)