

COMP 5070 – Statistical Programming for Data Science

Exam - What is Good to Read?

Anisha Mariam Abraham

110416080

LMDS

Introduction

Goodreads is an American social cataloguing website that allows individuals to search its books, annotations, quotes, and reviews database. Users can register books to generate library catalogues and reading lists. This is a report to scrap data regarding two popular genres -Science Fiction and Romance and investigate reading recommendations from GoodReads.com.

To study the book recommendations from GoodReads.com, we will use some competitive aspects of R. Data has been collected from <https://www.goodreads.com/shelf/show/science-fiction> and <https://www.goodreads.com/shelf/show/romance> and we will be focusing specifically on the book details collected on **20th November 6 pm**.

Science Fiction books' details were scraped into *booksSciFi* and Romance books' details were scraped into *booksRomance*. There were a total of 50 books in each genre with various information available for each book as mentioned below.

	Title	Author	Genre	PublishingYear	AverageRating	NumberOfRating	Pages	Decade
1	Dune (Dune, #1)	Frank Herbert	Science Fiction	1965	4.26	1272597	658	1960
2	Ender's Game (Ender's Saga, #1)	Orson Scott Card	Science Fiction	1985	4.31	1342565	324	1980
3	The Hitchhiker's Guide to the Galaxy (Hitchhiker's Gui...	Douglas Adams	Science Fiction	1979	4.23	1789946	216	1970
4	The Martian (Hardcover)	Andy Weir	Science Fiction	2011	4.41	1079803	384	2010
5	Ready Player One (Ready Player One, #1)	Ernest Cline	Science Fiction	2011	4.23	1126891	374	2010
6	Fahrenheit 451 (Kindle Edition)	Ray Bradbury	Science Fiction	1953	3.97	2309672	194	1950
7	1984 (Paperback)	George Orwell	Science Fiction	1949	4.19	4384557	368	1940
8	The Hunger Games (The Hunger Games, #1)	Suzanne Collins	Science Fiction	2008	4.33	8243228	374	2000
9	Brave New World (Paperback)	Aldous Huxley	Science Fiction	1932	3.99	1796187	268	1930

Showing 1 to 9 of 50 entries, 8 total columns

Table 1: *booksSciFi* containing details of Science Fiction books

Table 1 shows *booksSciFi* has **50 rows** and **8 columns**. The columns are as below:

1. **Title:** Contains the title of the book.
2. **Author:** Contains the author of the book.
3. **Genre:** Contains the genre of the book.
4. **PublishingYear:** Contains the year of first publication of the book.
5. **AverageRating:** Contains average rating received by readers online for the book.
6. **NumberOfRating:** Contains the number of people who rated the book.
7. **Pages:** Contains the total number of pages of the book.
8. **Decade:** Contains the decade year in which it was first published.

	Title	Author	Genre	PublishingYear	AverageRating	NumberOfRating	Pages	Decade
1	Pride and Prejudice (Paperback)	Jane Austen	Romance	1813	4.28	4093474	279	1810
2	Beach Read (Paperback)	Emily Henry	Romance	2020	4.02	977697	400	2020
3	Red, White & Royal Blue (Paperback)	Casey McQuiston	Romance	2019	4.11	886656	423	2010
4	It Ends with Us (It Ends with Us, #1)	Colleen Hoover	Romance	2016	4.21	2925056	386	2010
5	Book Lovers (Hardcover)	Emily Henry	Romance	2022	4.16	909709	377	2020
6	People We Meet on Vacation (Paperback)	Emily Henry	Romance	2021	3.90	1017024	364	2020
7	Twilight (The Twilight Saga, #1)	Stephenie Meyer	Romance	2005	3.64	6329913	498	2000
8	The Hating Game (Paperback)	Sally Thorne	Romance	2016	3.91	670186	365	2010
9	The Kiss Quotient (The Kiss Quotient, #1)	Helen Hoang	Romance	2018	3.91	413397	323	2010

Showing 1 to 9 of 50 entries, 8 total columns

Table 2: booksRomance containing details of Romance books

Table 2 shows *booksRomance* has **50 rows** and **8 columns**. The columns are as below:

1. **Title:** Contains the title of the book.
2. **Author:** Contains the author of the book.
3. **Genre:** Contains the genre of the book.
4. **PublishingYear:** Contains the year of first publication of the book.
5. **AverageRating:** Contains average rating received by readers online for the book.
6. **NumberOfRating:** Contains the number of people who rated the book.
7. **Pages:** Contains the total number of pages of the book.
8. **Decade:** Contains the decade year in which it was first published.

The most popular Science Fiction book is *The Hunger Games (The Hunger Games, #1)* by Suzanne Collins with 8243228 ratings. The least popular Science Fiction book is *The Long Way to a Small, Angry Planet (Wayfarers, #1)* by Becky Chambers with 137849 ratings. The most popular Romance book is *Twilight (The Twilight Saga, #1)* by Stephenie Meyer with 6329913 ratings. The least popular Romance book is *The Bride Test (The Kiss Quotient, #2)* by Helen Hoang with 157452 ratings. Average number of ratings for Science Fiction books is 1078736. The average number of ratings for Romance books is 1209266.

The longest Science Fiction book is *Dune (Dune, #1)* by Frank Herbert with 658 pages. The shortest Science Fiction book is *The Time Machine (Paperback)* by H.G. Wells with 118 pages. The longest Romance book is *Outlander (Outlander, #1)* by Diana Gabaldon with 850 pages. The shortest Romance book is *A Walk to Remember (Kindle Edition)* by Nicholas Sparks with 215 pages. Average number of pages for Science Fiction books is 332.5. The average number of pages for Romance books is 424.08.

The latest Science Fiction book is Project Hail Mary (Hardcover) by Andy Weir published in 2021. The oldest Science Fiction book is Frankenstein: The 1818 Text (Paperback) by Mary Wollstonecraft Shelley published in 1818. The latest Romance book is Happy Place (Hardcover) by Emily Henry published in 2023. The oldest Romance book is Pride and Prejudice (Paperback) by Jane Austen published in 1813.

Distribution of the length of a book

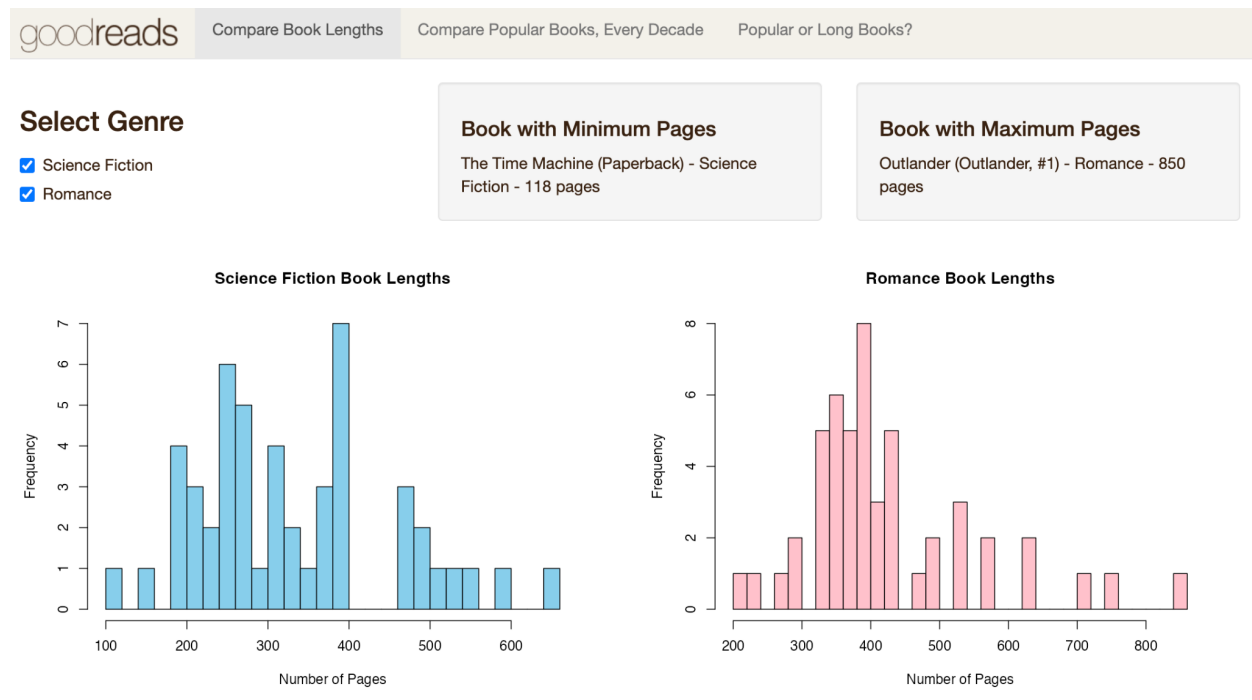


Figure 1: Histograms to compare the number of pages in Science Fiction and Romance books

The length of Science Fiction books typically ranges from 100 to 700 pages. The majority of these books cluster around the 250 to 300-page mark, with a significant number comprising 400 pages. However, there's a scarcity of books with page counts below 180 or above 500, indicating a concentration within this particular range.

In contrast, Romance novels span a wider page count range, usually falling between 200 to 900 pages. The prevalent lengths for Romance books cluster around 310 to 440 pages. Similar to Science Fiction, there's a scarcity of books below 300 pages or above 600 pages in the Romance genre.

Show entries

Search:

	Title	Genre	Pages
14	The Time Machine (Paperback)	Science Fiction	118
20	All Systems Red (The Murderbot Diaries, #1)	Science Fiction	144
34	The Martian Chronicles (Mass Market Paperback)	Science Fiction	182
23	The War of the Worlds (Paperback)	Science Fiction	192
6	Fahrenheit 451 (Kindle Edition)	Science Fiction	194
46	Annihilation (Southern Reach, #1)	Science Fiction	195
38	The Giver (The Giver, #1)	Science Fiction	208
100	A Walk to Remember (Kindle Edition)	Romance	215
3	The Hitchhiker's Guide to the Galaxy (Hitchhiker's Guide to the Galaxy, #1)	Science Fiction	216
43	A Wrinkle in Time (Time Quintet, #1)	Science Fiction	218

Showing 1 to 10 of 100 entries

Previous 2 3 4 5 ... 10 Next

Table 3: Titles of each book of selected genre arranged in increasing order of pages

Comparing the lengths of Science Fiction and Romance books reveals notable differences. While Science Fiction tends to have a more concentrated distribution, with a majority of books gravitating toward the 250-300-page range, Romance novels exhibit a more diverse range of lengths. Romance books commonly span the 310-440-page range, showcasing a broader distribution compared to Science Fiction.

Overall, Romance books generally have a wider variety of page counts, including a significant number of longer books, whereas Science Fiction books tend to concentrate more around a specific page count range, with fewer outliers at extreme lengths.

Compare Popular Books by every decade of first publication

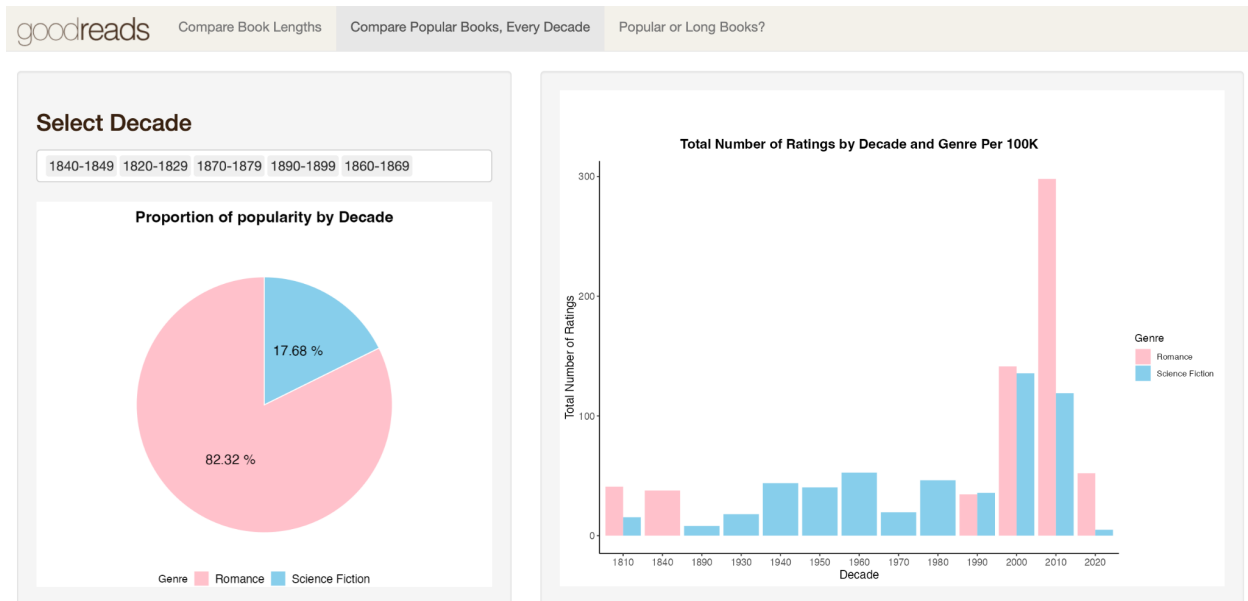


Figure 2: Pie chart and bar plot to compare the popularity of Science Fiction and Romance books through the years

The pie chart provides a visual comparison of the number of ratings received by different genres across selected decades, presenting their respective proportions. It allows us to discern which years experienced a higher proportion of popular books within a particular genre.

In the 2010s, Romance books garnered significant attention, evident from the substantial number of ratings, reaching up to 300000K from readers. This decade marked a notable surge in the popularity of Romance literature compared to Science Fiction. However, if we backtrack from 1840 to 1990, there was a relative absence of Romance books receiving substantial ratings. The genre only began gaining traction from the 1990s onwards, indicating a significant shift in readership and interest.

Contrarily, Science Fiction books have enjoyed sustained popularity across several decades. Particularly in the 2000s, they experienced a peak in popularity, surpassing other periods in terms of the number of ratings received. The following decade, the 2010s, remained consistently strong for Science Fiction, showcasing a continuous trend of popularity. As for the most recent decade, the 2020s, due to the incomplete data at present, it's challenging to make definitive observations or comment on the ongoing trend.

In summary, while Romance books experienced a resurgence in popularity from the 1990s onwards, the Science Fiction genre maintained a strong and consistent presence across various

decades, especially peaking in the 2000s and remaining robust through the 2010s. The evolving trends in the readership and popularity of these genres reflect changing preferences and interests among book enthusiasts over time.

Show 10 entries

Search:

	Title	Author	Genre	PublishingYear	AverageRating	NumberOfRating	Pages	Decade
23	The War of the Worlds (Paperback)	H.G. Wells	Science Fiction	1897	3.83	304979	192	1890
14	The Time Machine (Paperback)	H.G. Wells	Science Fiction	1895	3.9	506037	118	1890
94	Wuthering Heights (Paperback)	Emily Brontë	Romance	1847	3.89	1740351	464	1840
76	Jane Eyre (Paperback)	Charlotte Brontë	Romance	1847	4.15	2034941	532	1840

Showing 1 to 4 of 4 entries

Previous 1 Next

Table 4: Information about books and their popularity according to selected decades

Popular or Long Books?



Figure 3: Scatterplots to compare the relationship between the length of books and their popularity

The first scatterplot in Figure 3, which represents Science Fiction books, exhibits an intriguing trend. As the number of pages in these books increases, there's a discernible positive correlation with the ratings they receive. This indicates that longer Science Fiction books tend to attract

higher ratings, showcasing a pattern of positive reception from readers. However, it's notable that among these longer books, there's a cluster that receives relatively low ratings, falling below 2000K from the audience. The most highly rated Science Fiction book in the dataset contains roughly 400 pages, marking a peak in both length and popularity among the Science Fiction genre.

In contrast, the second scatterplot within Figure 3, depicting Romance books, portrays a different relationship between book length and reader ratings. Here, as the number of pages increases, there's a visible decrease in the ratings garnered by Romance books. This implies that longer Romance books tend to receive comparatively lower ratings, indicating a potential preference among readers for shorter works within this genre. Interestingly, a significant number of books in the range of 300 to 400 pages have received ratings from a smaller audience, less than 2000K. The highest-rated Romance book in the dataset is approximately 500 pages long, representing a notable outlier with both substantial length and higher popularity among Romance readers.

These observations suggest varying reader preferences between genres—Science Fiction books appear to gain more positive attention with increased length, while Romance books exhibit a trend where longer works may not necessarily attract higher ratings. The nuances in reader behaviour and genre-specific preferences revealed through these scatterplots offer valuable insights into how book-length might impact reader engagement and reception within distinct literary genres.

Show
10
entries

Search:

	Title	Genre	NumberOfRating	Pages
42	The Long Way to a Small, Angry Planet (Wayfarers, #1)	Science Fiction	137849	518
89	The Bride Test (The Kiss Quotient, #2)	Romance	157452	296
49	The Forever War (The Forever War, #1)	Science Fiction	163461	278
74	Get a Life, Chloe Brown (The Brown Sisters, #1)	Romance	170958	384
26	The Left Hand of Darkness (Paperback)	Science Fiction	172744	304
44	Second Foundation (Foundation, #3)	Science Fiction	179703	256
40	Old Man's War (Old Man's War, #1)	Science Fiction	191334	318
41	Foundation and Empire (Foundation, #2)	Science Fiction	198796	256
20	All Systems Red (The Murderbot Diaries, #1)	Science Fiction	203361	144
84	One Last Stop (Paperback)	Romance	216037	418

Showing 1 to 10 of 100 entries

Previous
1
2
3
4
5
...
10
Next

Table 4: Titles of each book of selected genre arranged in increasing order of NumberOfRating

Conclusion

In conclusion, the analysis conducted on Science Fiction and Romance genres sourced from Goodreads provides valuable insights into the readership trends and preferences. The examination of book details, including ratings, page counts, publication years, and reader engagement, illuminated distinctive characteristics between these popular genres.

Science Fiction books showcase a tendency for increased popularity with longer page counts, as depicted in the positive correlation between book length and reader ratings. Notably, while the majority of Science Fiction books cluster around the 250-300-page range, outliers at the 400-page mark garner higher ratings, reflecting positive reception among readers.

Contrarily, Romance novels exhibit a different trend, where longer works tend to receive comparatively lower ratings. The genre encompasses a wider range of page counts, with a preference towards books in the 310-440-page range, yet longer Romance novels may not necessarily resonate as strongly with readers in terms of ratings.

Moreover, the analysis of popular books by decade highlights the evolving trends in readership over time. Romance literature experienced a resurgence from the 1990s, while Science Fiction maintained consistent popularity, notably peaking in the 2000s. These observations underscore the dynamic nature of readers' preferences and the genre-specific nuances that shape their engagement with literary works.