Proposal- Interactive Storytelling using the Goodreads book reviews dataset

Team:
Madhura Sridhar (ms10146)
YiChun Yeh (ycy274)

Project goal:
We hope to explore some of the common questions that arise when going through the Goodbooks review dataset, concerning the popularity of books based on the user ratings they garnered, visualize more information about the authors using the existing dataset and visualize the review of the most popular author in the form of a word cloud, all using D3, HTML, CSS, JAVASCRIPT, using the Bootstrap front-end framework.

Dataset:
Goodreads book reviews dataset
Source: https://www.kaggle.com/gnanesh/goodreads-book-reviews
Attributes inside the dataset:
bookID
title : Title of the book
Author : author of the book
Rating : average rating of the book on Goodreads
ratingsCount : Sum total of the rating for the book
reviewsCount: Sum total of the reviews for the book
reviewerName: name of the person who reviewed the book
reviewerRatings: rating that this reviewer gave the book.
Review: textual review for the book given by the reviewer.

Analytical Questions and Proxy Tasks

- Analytical question 1: What is the most popular book?
  Proxy task:
  Most popular book can be considered as the one with highest rating and reviewerRatings.
  Proxy value:
  bookID, rating, ratingsCount

- Analytical question 2: Who is the most popular author in the dataset?
  Proxy task:
  Most popular author can be considered as the author who has written the top 5 liked books, which in turn is related to the top 5 popular books that we derived from question 1.
  Proxy Value:
  rating, ratingsCount, author

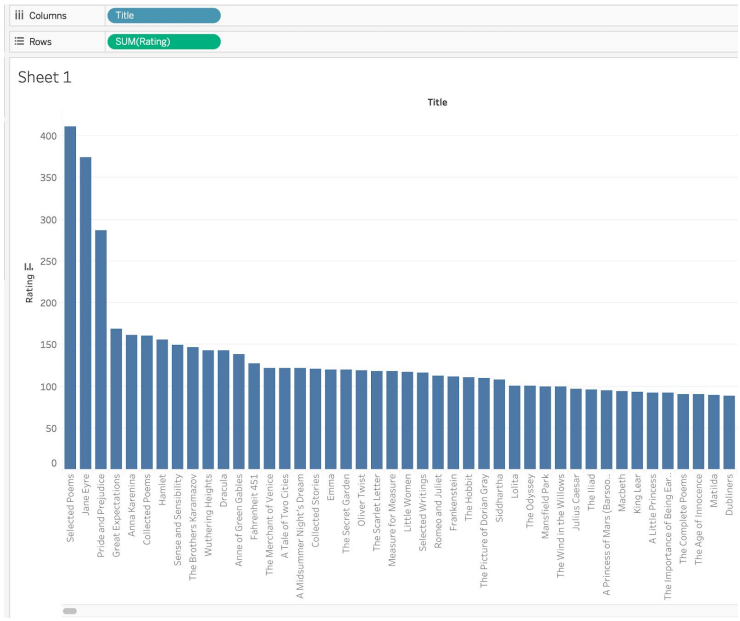
- Analytical question 3: Which author is known to have written books that have high rating, while the author himself/herself has relatively below average rating?
  Proxy task: Which author has written a book in the top 10 most popular books while he/she has below average ratings for all the books he/she has written.
  Proxy value: author, rating, ratingsCount.

- Analytical Question 4: Which popular author has been reviewed with a 'positive' word set?
  Proxy task: Based on who is the most popular author in question 2, we can determine the wordcloud of the review that particular author got.
  Proxy value: author, reviews.

Story Design

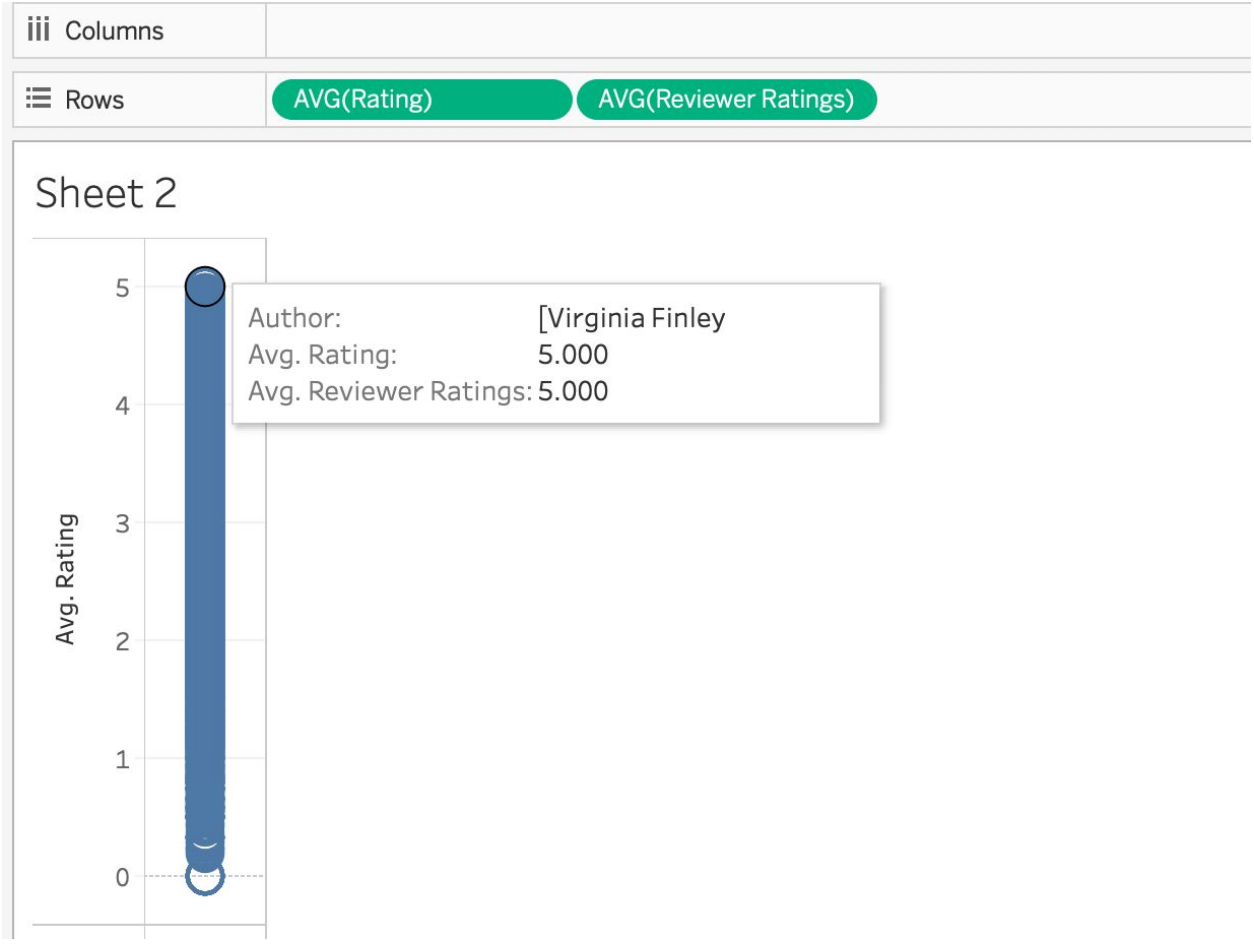
Data Analysis and sketches

- Analytical question 1: What is the most popular book?



Description: We tried to determine the most popular book by considering the sum of ratings they got all-over.

Analytical question 2: Who is the most popular author in the dataset?



Description: Here, we tried to analyze the popularity of authors based on the combination of average rating and total ratings (ratingsCount).

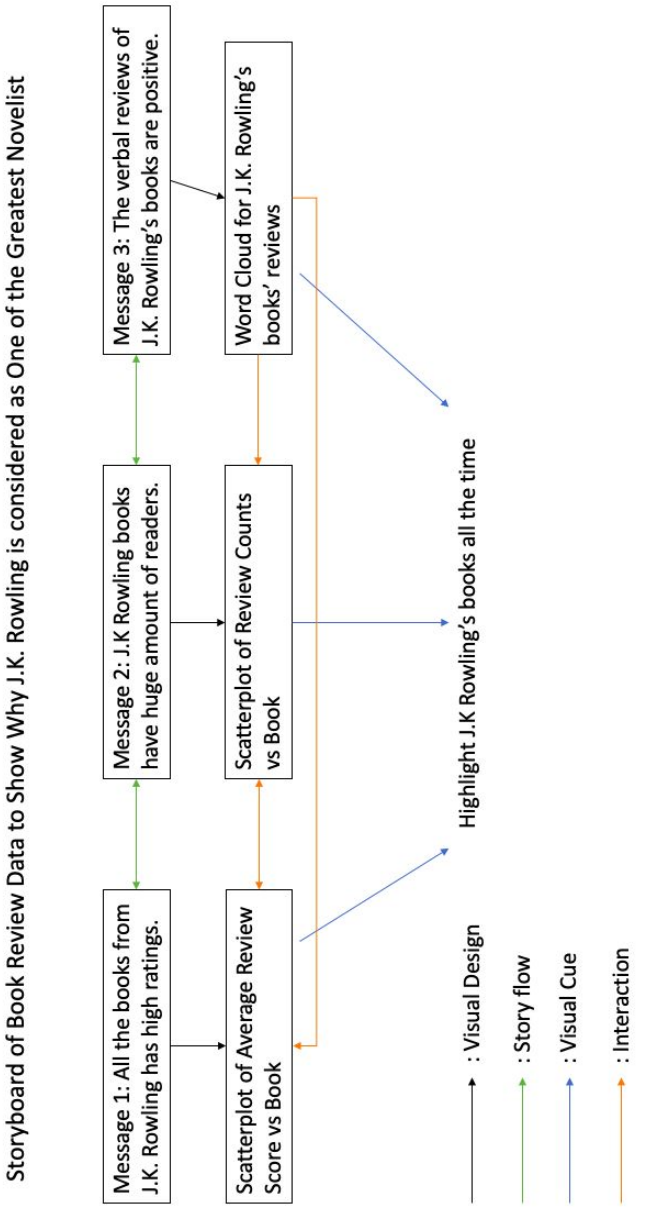
- Analytical question 3: Which author is known to have written books that have high rating, while the author himself/herself has relatively below average rating?
  We can use the same graph we generated for Analytical question 2.

- Analytical Question 4: It is visualized by D3 using the Word cloud feature as shown below.

Storyboard

We apply the posit storyboard technique to illustrate the data story driven from our analysis from the data. The flowchart structure is inspired from http://www.storytellingwithdata.com/blog/2015/8/24/how-i-storyboard.
The flow chart shows the interactions between different visual encodings which correspond to convey respective key messages.

- Based on the graphical visualizations created in d3, we started off building the story for the visualizations.
- They include visualization and story for popular books, popular authors, and word cloud of the reviews for the most popular author, J.K. Rowling.



Changelog

- We added more analytical questions trying to find the popularity of books and authors.
- Based on the feedback given last week by the TA, we tried to make the story more cohesive and well-woven with the visualizations we had.

Implementation

- Github repository: https://github.com/NYU-VIS-FALL2018/storytelling-group-19
- Demo link: https://nyu-vis-fall2018.github.io/storytelling-group-19/