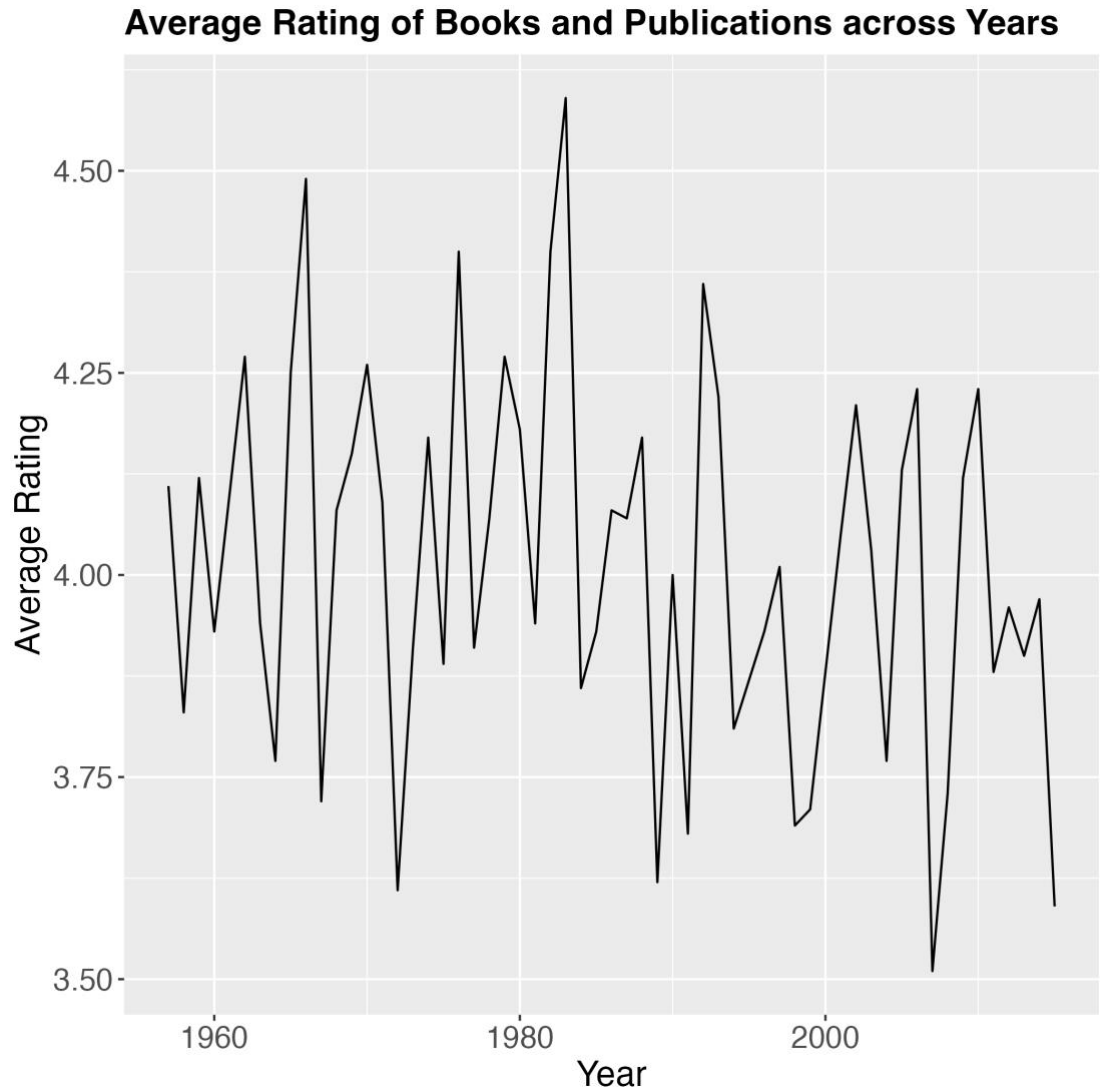**Problem 1**

1. Time-Dependent Graph

Figure 1. Time-series analysis of average ratings of books and publications from 1957 to 2015.

Data story message:

There are 2 variables in this time-dependent graph, namely the year of publication and average rating the book or publication received. As years went by from 1957 to 2015, average rating fluctuated often but went on a slightly decreasing trend. The highest rating occurred in roughly 1982, and the lowest rating happened in the year of about 2007. This time-series analysis plot might indicate that the quality of books and publications has been decreasing in recent years. For users and readers, old books and publications might be more suitable choices for academic references.
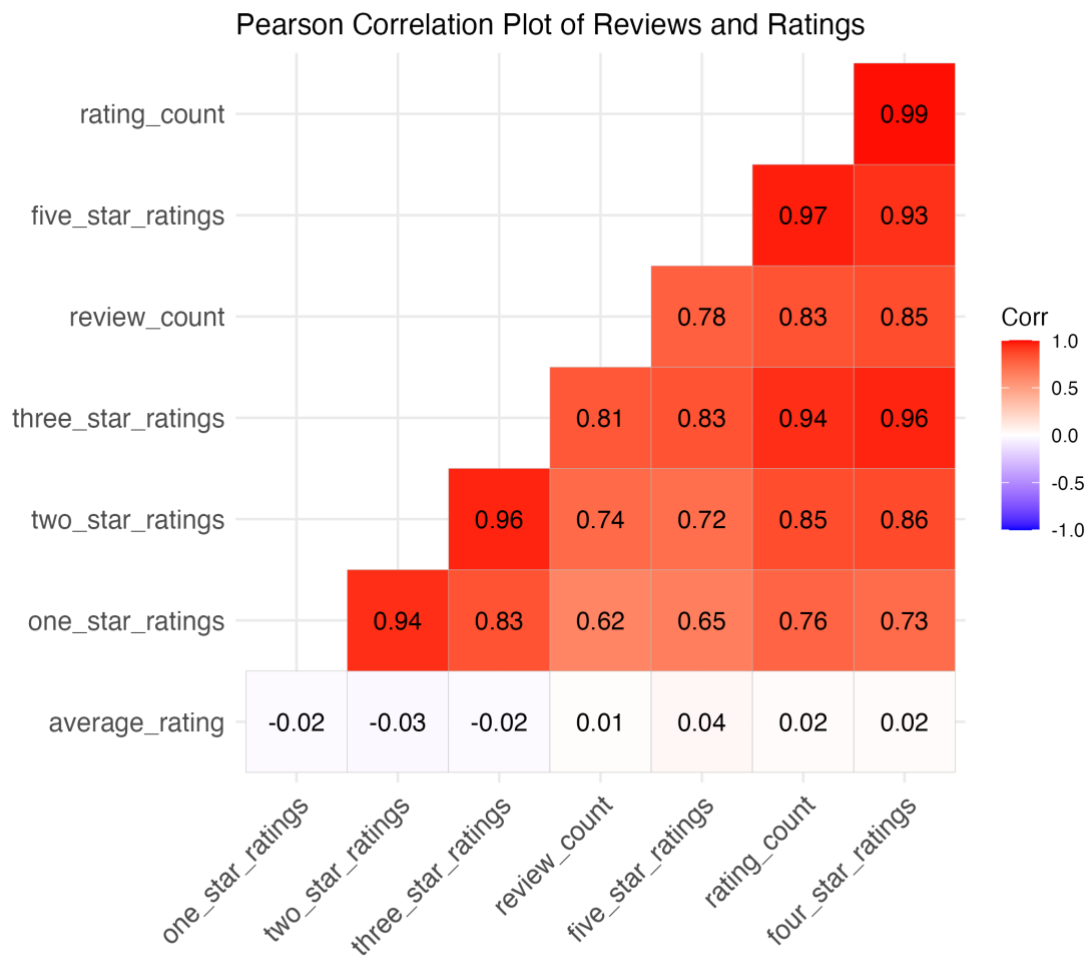
2. Correlation Plot

Figure 2. Pearson correlation plot among properties regarding reviews and levels of ratings.

Data story message:

There are 7 variables in this time-dependent graph, namely the number of reviews, the number of ratings, average rating scores, and the number of different star ratings where stars range from one star to five stars. This is a visualization plot revealing pair-wise internal connections among reviews and ratings using Pearson correlation. -1.0 represents negatively linearly related, while 1.0 represents positively linearly related, and approaching 0 means no linear relationships between a pair of variables. The average rating score is linearly independent of any other review and rating properties, indicating the justice of the rating score since it is not determined by any single variable or property. Except average rating score, all other variables are strongly and positively linearly related to each other, indicating that most reviews and ratings are objective and reliable. This also shows the similar distribution patterns of different levels of ratings. New readers and users are encouraged to take reviews and ratings into consideration when choosing

reference books and publications.

3. Linear Regression

Linear Regression between five-Star-rating Count and one-star-rating Count
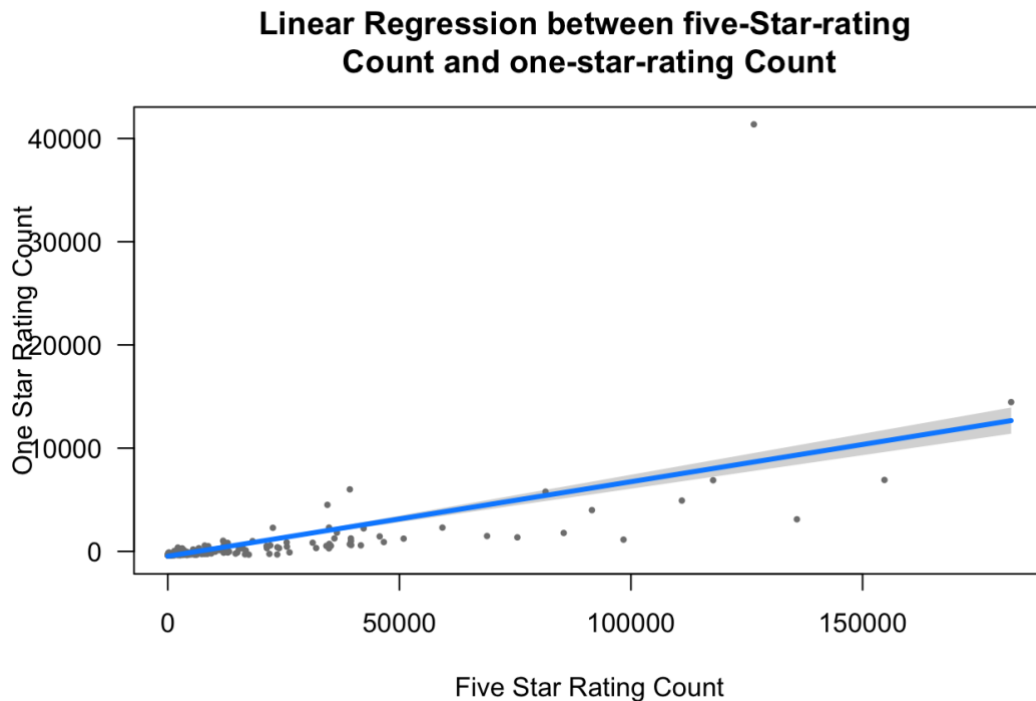
Figure 3. Linear regression analysis between five-star rating number and one-star rating number.

Data story message:

There are 2 variables in this time-dependent graph, namely the five-star rating count and the one-star rating count. Each book or publication sampled received roughly 30x more five-star ratings than one-star ratings, and the five-star rating count is approximately linearly related to the one-star rating count. This linearity between the highest rating count and lowest rating count across all sampled books and publications shows the consistency between high-rating reviewers and low-rating reviewers. To some extent, the ratings of books and publications are reliable and referable when choosing references for new users and readers. Those books and publications are mostly of good quality as well.
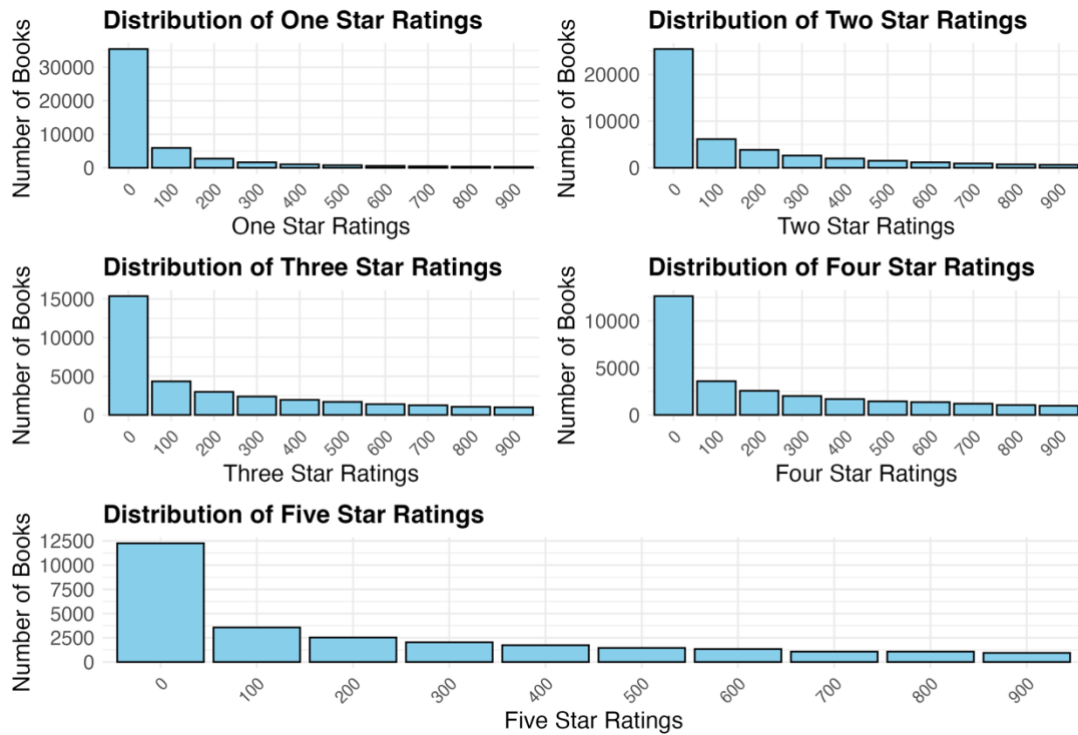
4. Multi-panel graph

Figure 4. Comparisons of different levels of ratings' distribution patterns.

Data story message:

This multi-panel figure has 5 components, each showing a distribution pattern of a rating level to avoid overloaded information in a single panel. Considering the differences in total mounts of each rating level, separating them into different panels also helps us understand their distribution patterns. Most books and publications received less than 100 ratings at each level. Most books and publications received higher levels of ratings than low levels of ratings, but the rating distributions are similar across levels as a rapidly decreasing curve as rating counts increase at each level. This indicates that most books and publications are of good quality. To some extent, the ratings of books and publications are reliable and referable when choosing references for new users and readers.

Problem 2

Nowadays there are more and more users contact with their friends via social media spaces, and Twitter is one of the most popular platforms among them. Twitter friends consist of many complex connections forming diverse social network structures. Himelboim et al [1]. proposes a conceptual and practical model for the classification of topical Twitter networks, based on their network-level structures. Social network analysis helps reveal the density, modularity, centralization, and the fraction of isolated users. Furthermore, it also leads us to suggest six structures of information flow: divided, unified, fragmented, clustered, in and out hub-and-spoke networks.

Himelboim et al [1]. demonstrate the value of diverse social network structures by sampling 60 Twitter topical social media network datasets into these six different patterns of connections. They also illustrate how different topics of conversations exhibit different patterns of information flow. Since each pattern of information flow has unique properties, this classification is quite useful and essential when evaluating social media activity. We can also utilize this classification to construct recommendation strategies for new-comers in a certain social media platform.

[1] Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. *Social media+ society*, *3*(1), 2056305117691545.