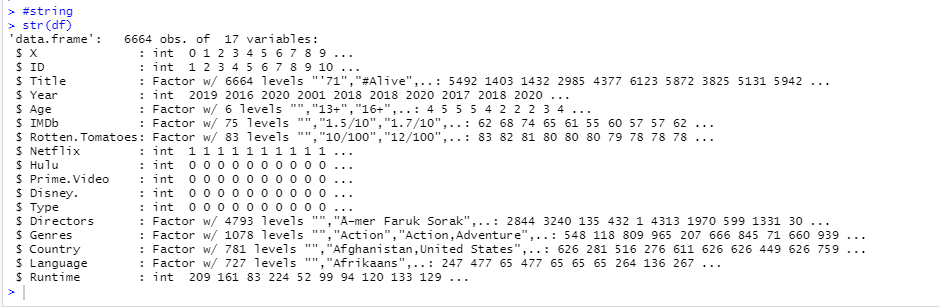
**Yijun Wang**

**ALY 6010**

**Introduction**

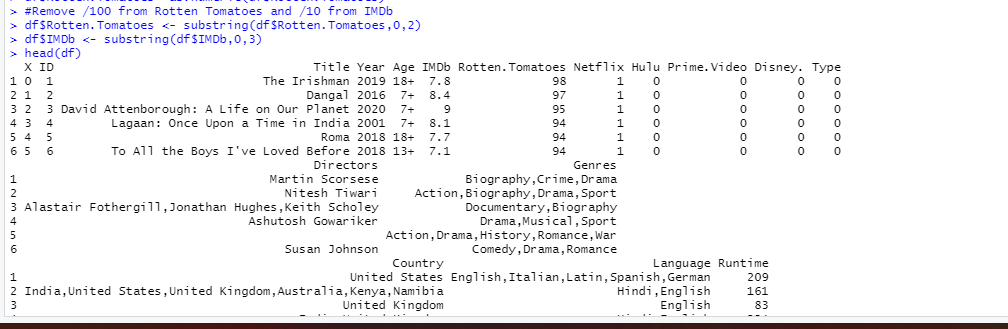
I got this dataset from the Kaggle website, and I picked the data set that included movies from Netflix, Prime Video, Hulu, and Disney+. This topic was chosen because of covid-19. The long-term decline of movie theaters has created an opportunity for enormous expansion for different streaming media providers. The reason I'd like to study this data is because I'm curious about which streaming service individuals should use.



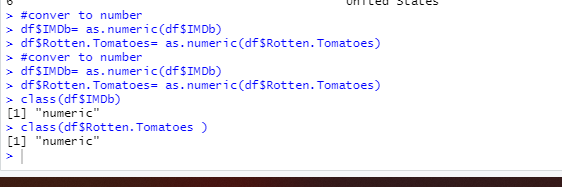


this dataset has 6664 observations and 17 columns. All of dataset are talking around four different platforms: Netflix, Hulu, Disney, and Prime. Video. The goal of this study is to examine the dataset statistically and provide relevant, in-depth findings about the dataset, including data levels, descriptive statistics, and data distribution. Rest of dataset are Movies’ title, publish year, Age’s level, Rotten.Tomatoes’s scores, IMDb’s score, Directors, Genres, Country, Language and run time. These data are composed of int and factors.

From the column of “IMDb”, the number is followed by the symbol "/10." Although we recognize that this is a scoring system, it is quite cumbersome when undertaking data processing. As a result, I eliminated the "/10" and changed these as number.



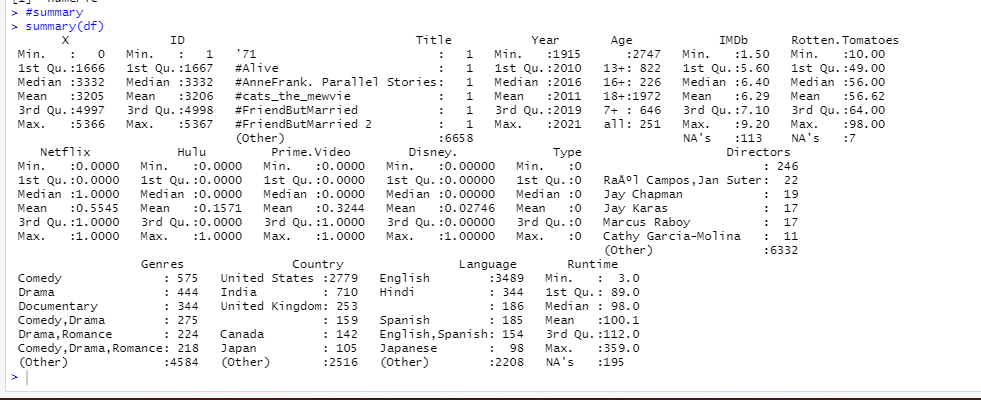
And I did same thing to “Rotten.Tomatoes”.



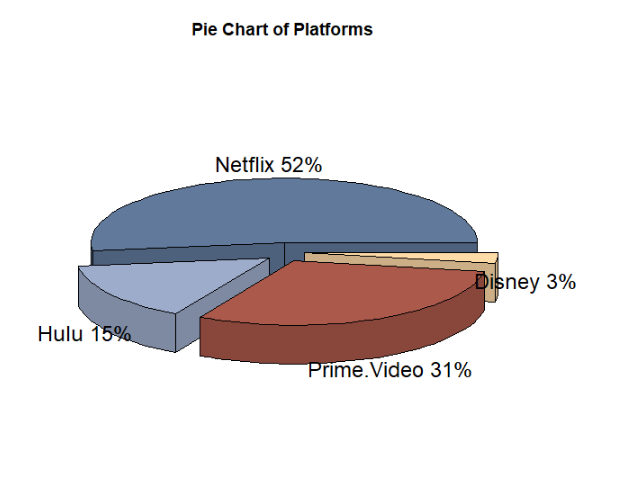
Besides that, I remove blank value from “Age”.



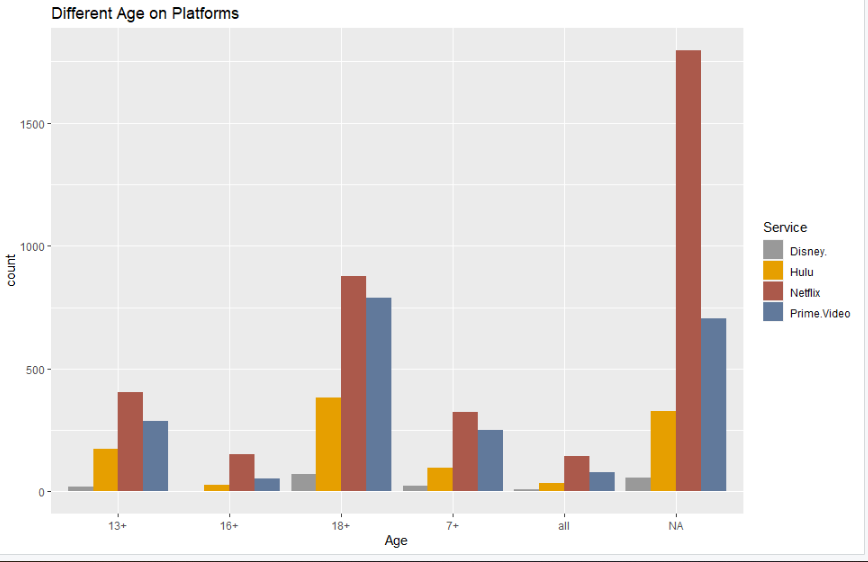
Here is the summary of this dataset.



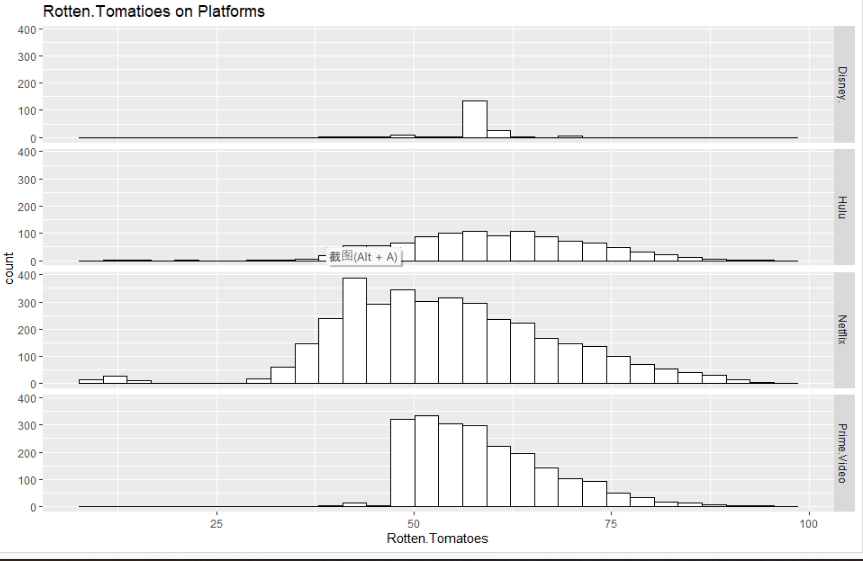
**Data Analysis**



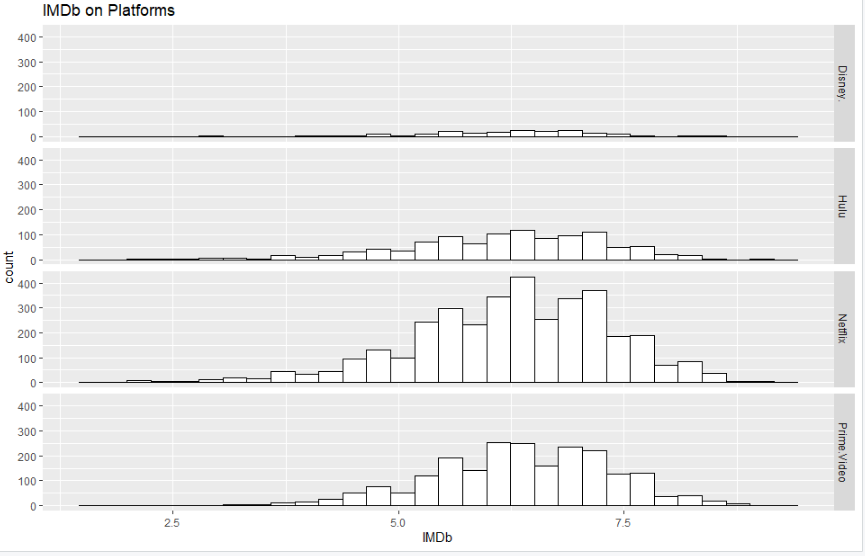
According to this pie chart, Netflix has the most total number of movies played, followed by Prime. The last but not least is Disney.



This bar graph depicts the total number of movies watched by each age group on various platforms. According to the graph, Netflix has the most broadcasts, while Disney has the fewest.



As shown on the following graph, the average score of Netflix platform on Rotten. Tomatoes is around 50. Prime. Video has a substantially higher average score than the other three formats.



This is the number of times count has received a rating on IMDb. In contrast to Rotten. Tomatoes, Netflix have considerably better IMDb ratings, especially when compared to Disney.

**Summary**

Although Prime Video has the highest average rating on Rotten. Tomatoes, Netflix is the best option if you value quality over IMDb. The platform with the most films published in the last ten years and over 90% IMDb approval. On the other hand, if you want to meet the needs of your children, Disney+ is a fantastic subscription because it is specifically designed for them.

Kaggle's movie data contains information about movies that have been played on four different platforms, as well as other relevant information. This data set comprises 17 attributes associated with 6664 observations, as well as the title, year, and director of the film. The process of data cleansing is critical. We can remove extraneous information, such as blank values in Age. Furthermore, the use of vertical charts, pie charts, or box plots helps us to examine the direction of the data and the relevant correlations in a more natural manner.

**Question**: Based on the years in the data, can it be concluded about which year the movies are more popular?

