Background

I obtained data of *2019 Netflix Movies and TV Shows* at the website Kaggle (https://www.kaggle.com/shivamb/netflix-shows). This dataset consists of tv shows and movies available on Netflix as of 2019. The dataset is collected from Flixable which is a third-party Netflix search engine. These data can perform some possibly social questions such as: understanding what content is available in different countries, identifying similar content by matching text-based features, finding interesting insights, and so on.

Summary of Data

There are 12 variables and more than 6,000 observations in the dataset. The variables include the type of the show (Movie or TV show), title, genre, countries where they are serviced, etc. there are some missing values on the director and cast name sections. Figure 1 shows that how many movies and TV shows are in each MPAA and TV Parental Guideline Rating System. Most movies and TV shows got in TV-14 and TV-MA ratings. Figure 2 shows that how many movies and TV shows are serviced in each country (not including all countries in the earth). Interestingly, India is the second most with Netflix videos, and United Arab Emirates is the third most. Further, I also want to see if there is any relationship between the audience-age rating and number of movies and TV shows in countries.

Potential Problem

Most of the data are categorical and text data, so I may have to figure out how to analyze the texts. Related to the Netflix data, I may need other datasets such as rotten tomatoes so that I can do further related analysis.
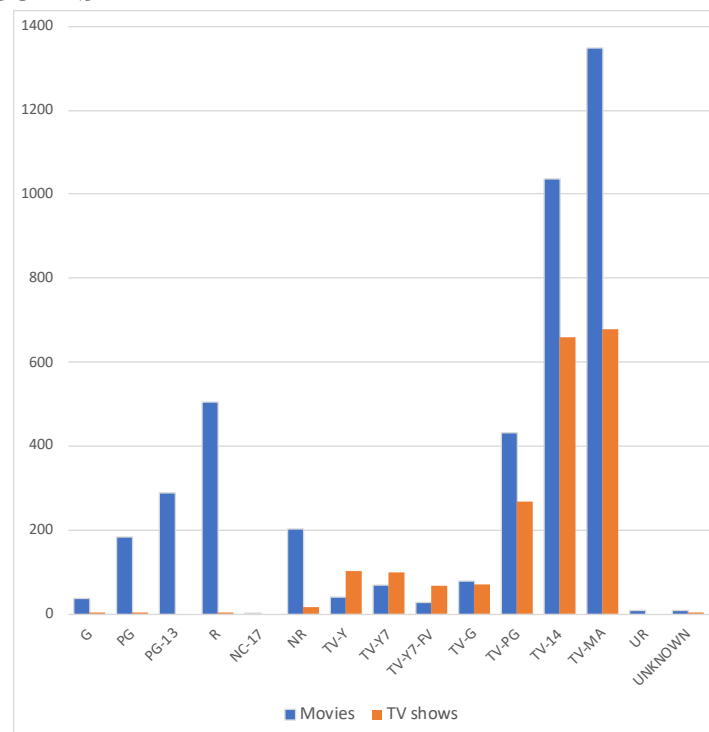
TABLES AND FIGURES

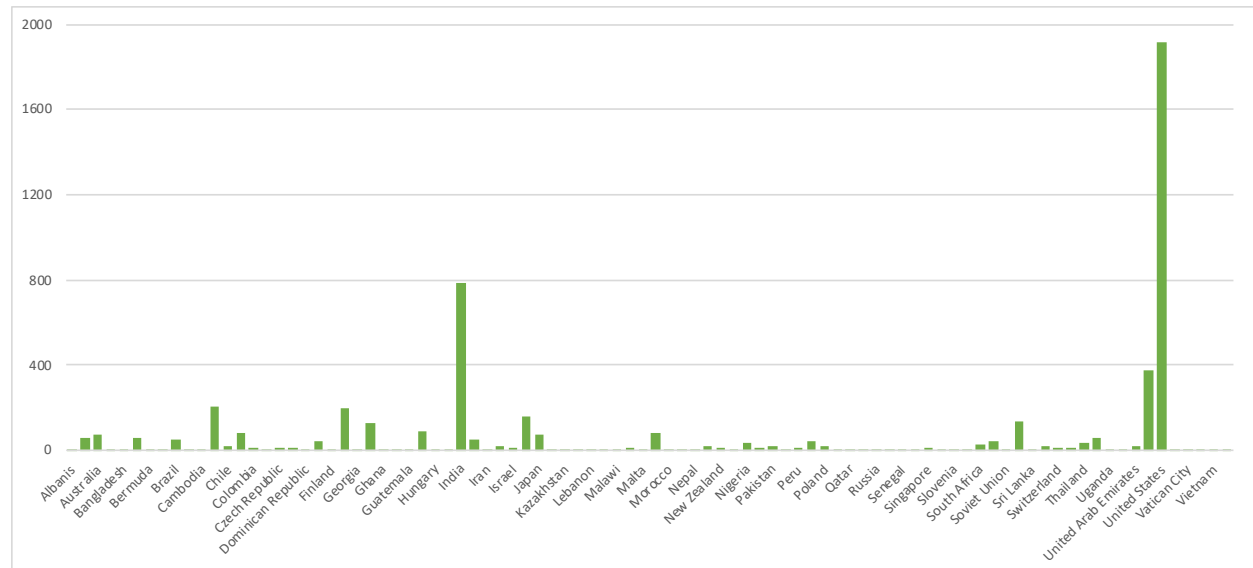Figure 1: MPAA and TV Parental Guideline Rating System of Movies and TV shows

Figure 2: Number of Movies and TV Shows per Countries