TV Shows and Movies listed on Netflix

Sathish Kumar Rajendiran

IST 719 - HW3

Description

This dataset is sourced from https://www.kaggle.com/shivamb/netflix-shows.

It consists of movies and tv shows in Netflix available as of 2020. Over 2000, titles released in various countries between 1925 and 2020 are available in this dataset. In addition, it includes title, show ratings, listed category, description, running duration, cast, director etc.

From this dataset, we can explore interesting facts like, contents available in different countries, increase in trend over years, famous categories, etc. Data visualization task includes (but not limited to),

1. Dataset Preparation
    a. Data extraction
    b. Preprocessing
    c. Dataset Size analysis

2. Build Single Dimension Plots
    a. Titles by Year released – Bar Plot
    b. Titles by Rating – Line chart
    c. Total number of Titles by Type – Pie Chart
    d. Top 10 Genre – Horizontal bar plot

3. Build Multi Dimension Plots
    a. Available shows by year released by Type - Stacked bar chart
    b. Available shows by year released by Rating - Stacked bar chart - beside

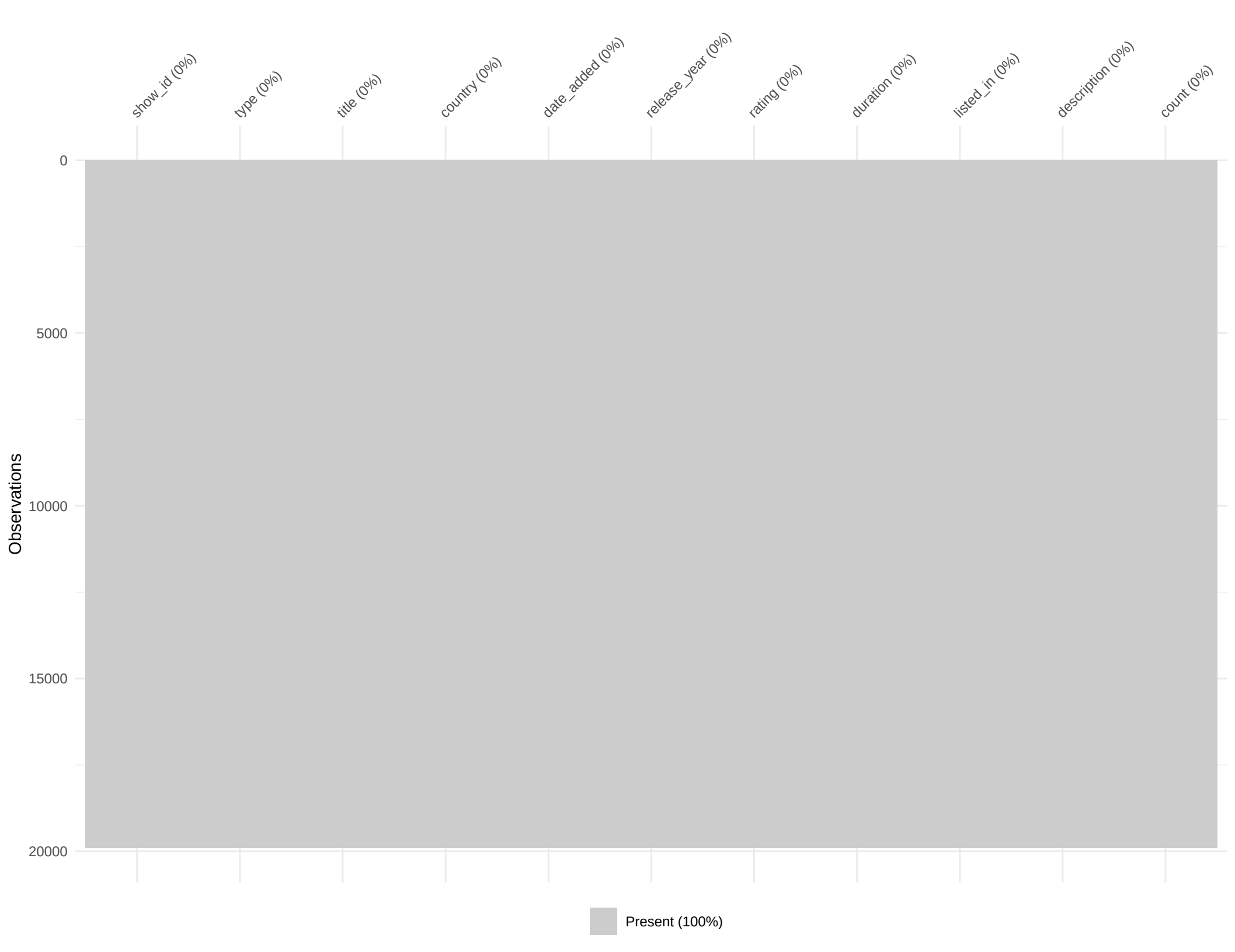
4. Other Analysis
    a. Wordcloud

Data Size

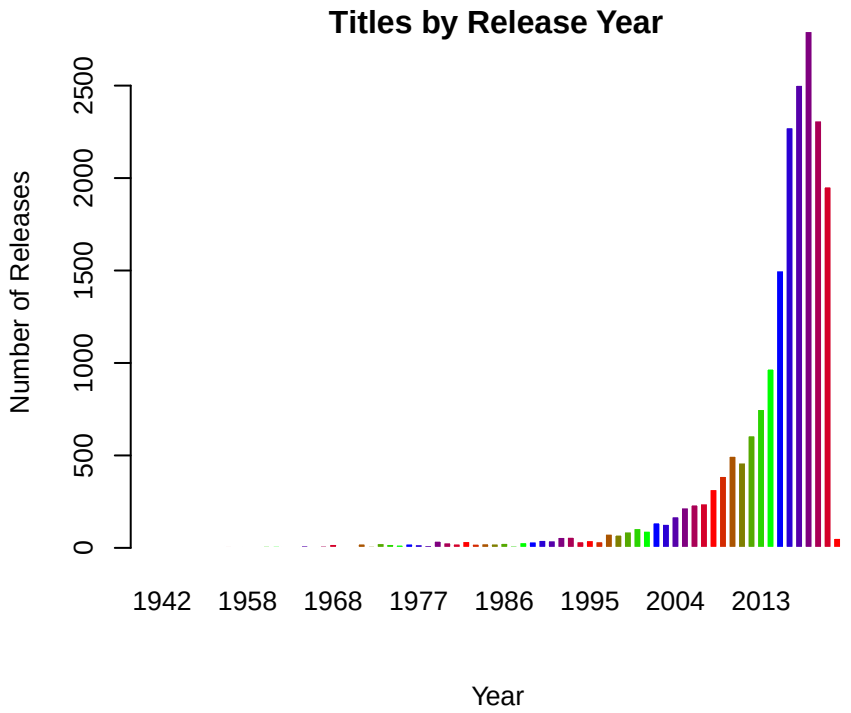
```
> # Print data size
> dataVolume(netflix)

Total Number of Observations:  7787
Total Number of Variables:  12
Data Size is  3737.76
```

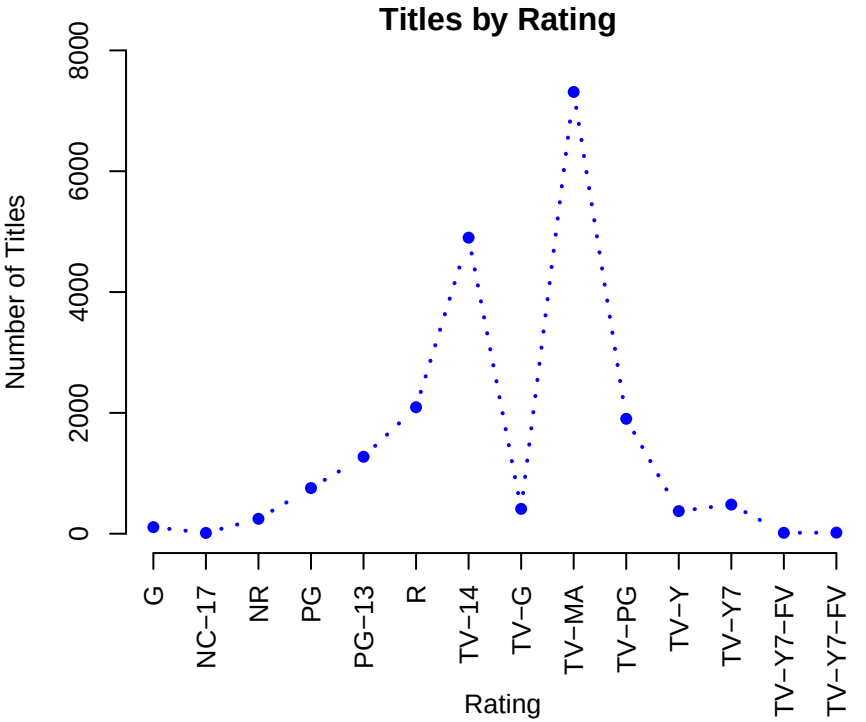
Structure

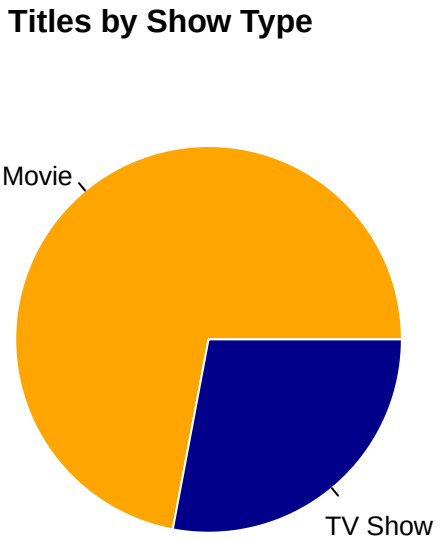
```
 $ show_id     : chr [1:19904] "s1" "s1" "s1" "s2" ...
 $ type        : chr [1:19904] "TV Show" "TV Show" "TV Show" "Movie" ...
 $ title       : chr [1:19904] "3%" "3%" "3%" "7:19" ...
 $ country     : chr [1:19904] "Brazil" "Brazil" "Brazil" "Mexico" ...
 $ date_added  : chr [1:19904] "August 14, 2020" "August 14, 2020" "August
 $ release_year: int [1:19904] 2020 2020 2020 2016 2016 2011 2011 2009 200
 $ rating      : chr [1:19904] "TV-MA" "TV-MA" "TV-MA" "TV-MA" ...
 $ duration    : chr [1:19904] "4 Seasons" "4 Seasons" "4 Seasons" "93 mir
 $ listed_in   : chr [1:19904] "International TV Shows" "TV Dramas" "TV Sc
 $ description : chr [1:19904] "In a future where the elite inhabit an isl
d paradise far from the crowded slums, you get one chance to join t"| __tr
cated__ "After a devastating earthquake hits Mexico City, trapped survivor
 $ count       : num [1:19904] 1 1 1 1 1 1 1 1 1 1 ...
 - attr(*, "na.action")= 'omit' Named int [1:1045] 43 44 94 95 96 167 237
  ..- attr(*, "names")= chr [1:1045] "43" "44" "94" "95" ...
```

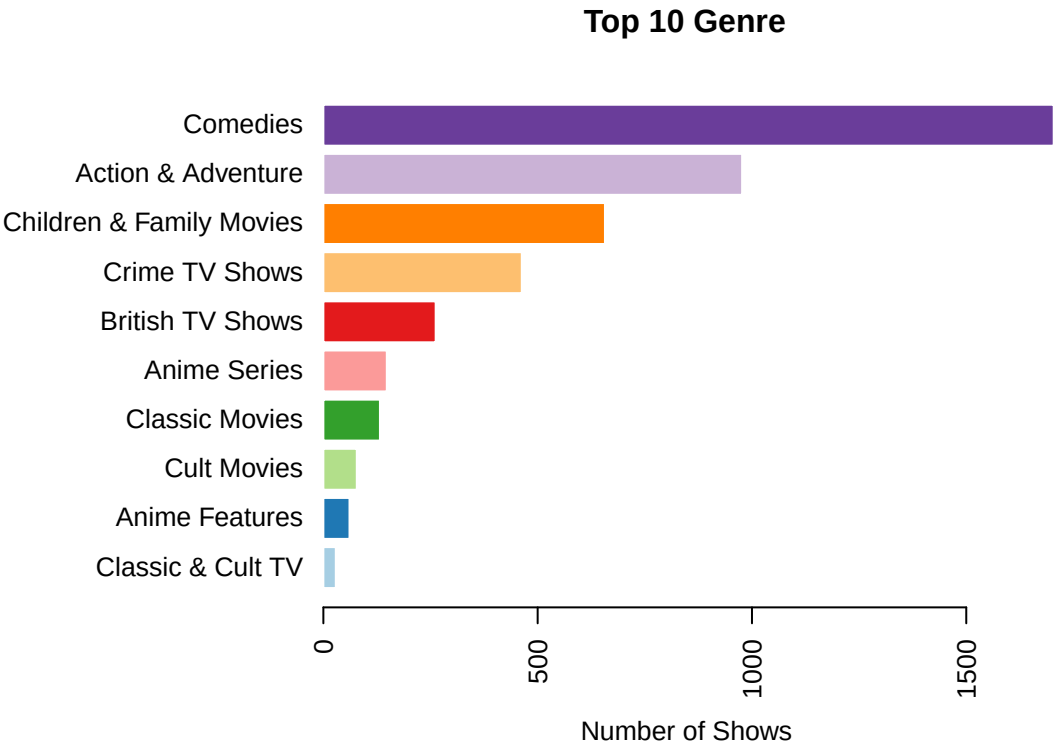

show_id (0%)
type (0%)
title (0%)
country (0%)
date_added (0%)
release_year (0%)
rating (0%)
duration (0%)
listed_in (0%)
description (0%)
count (0%)
0
5000
10000
15000
20000
Observations
Present (100%)

Single Dimension Plots



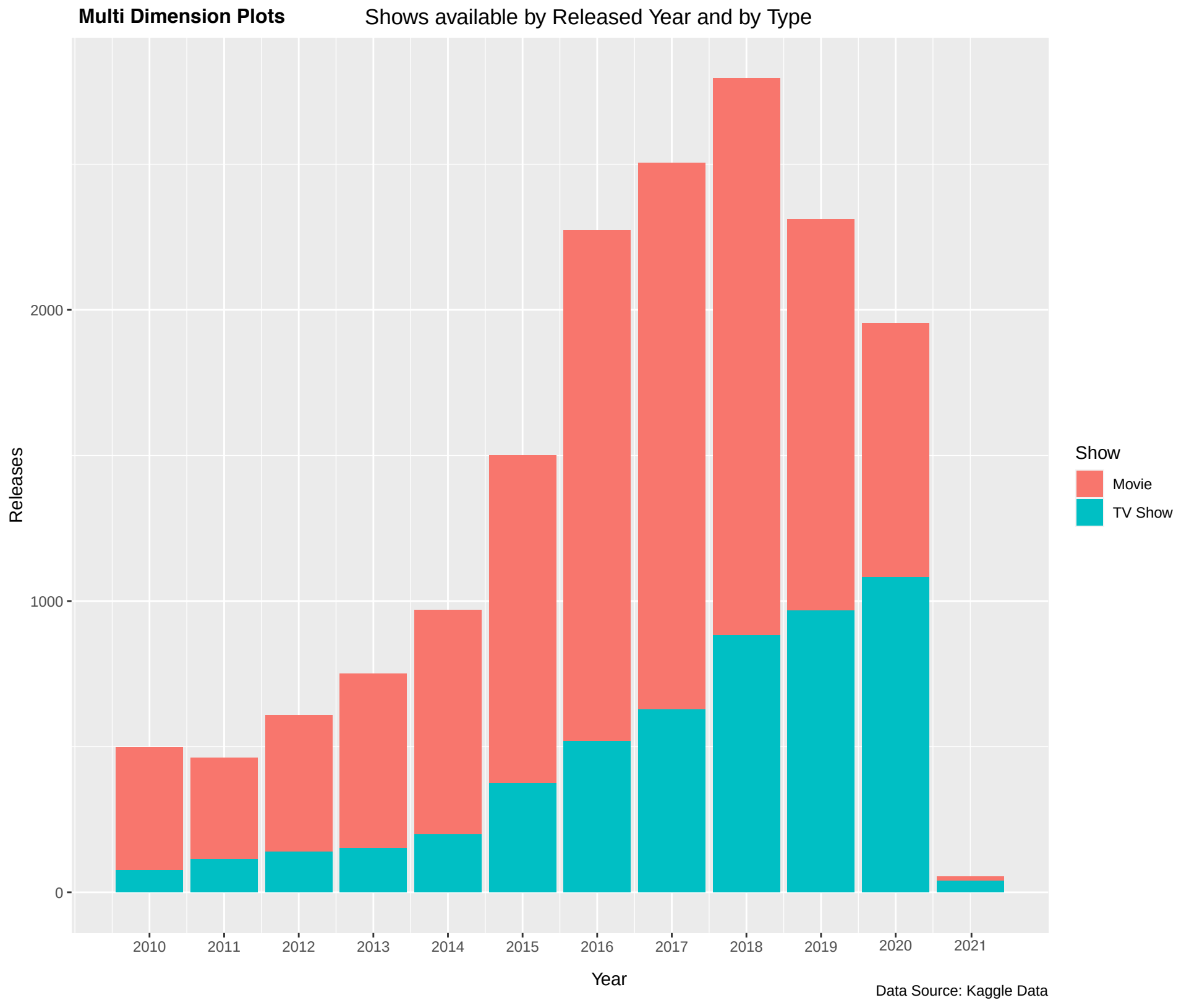


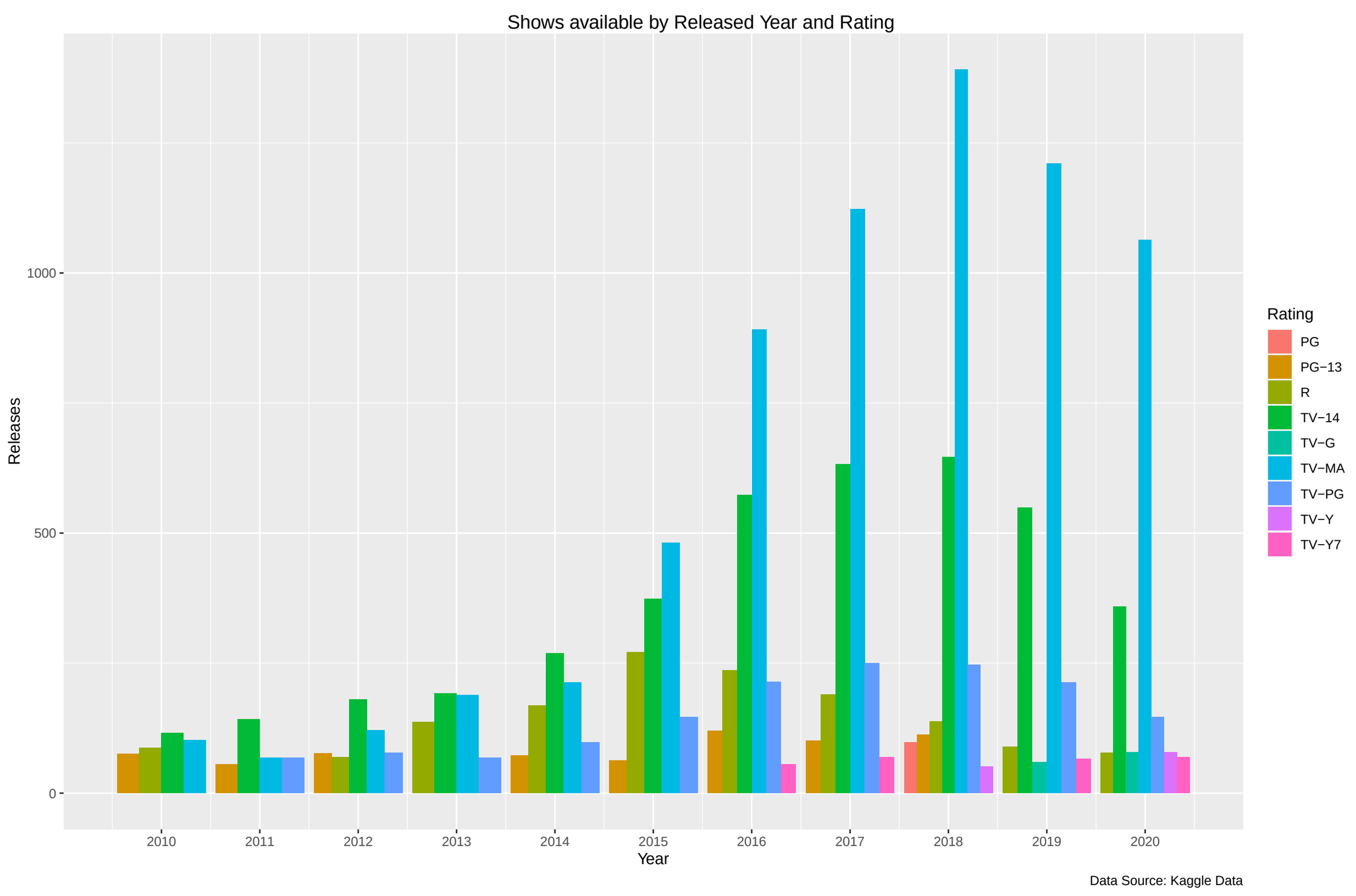


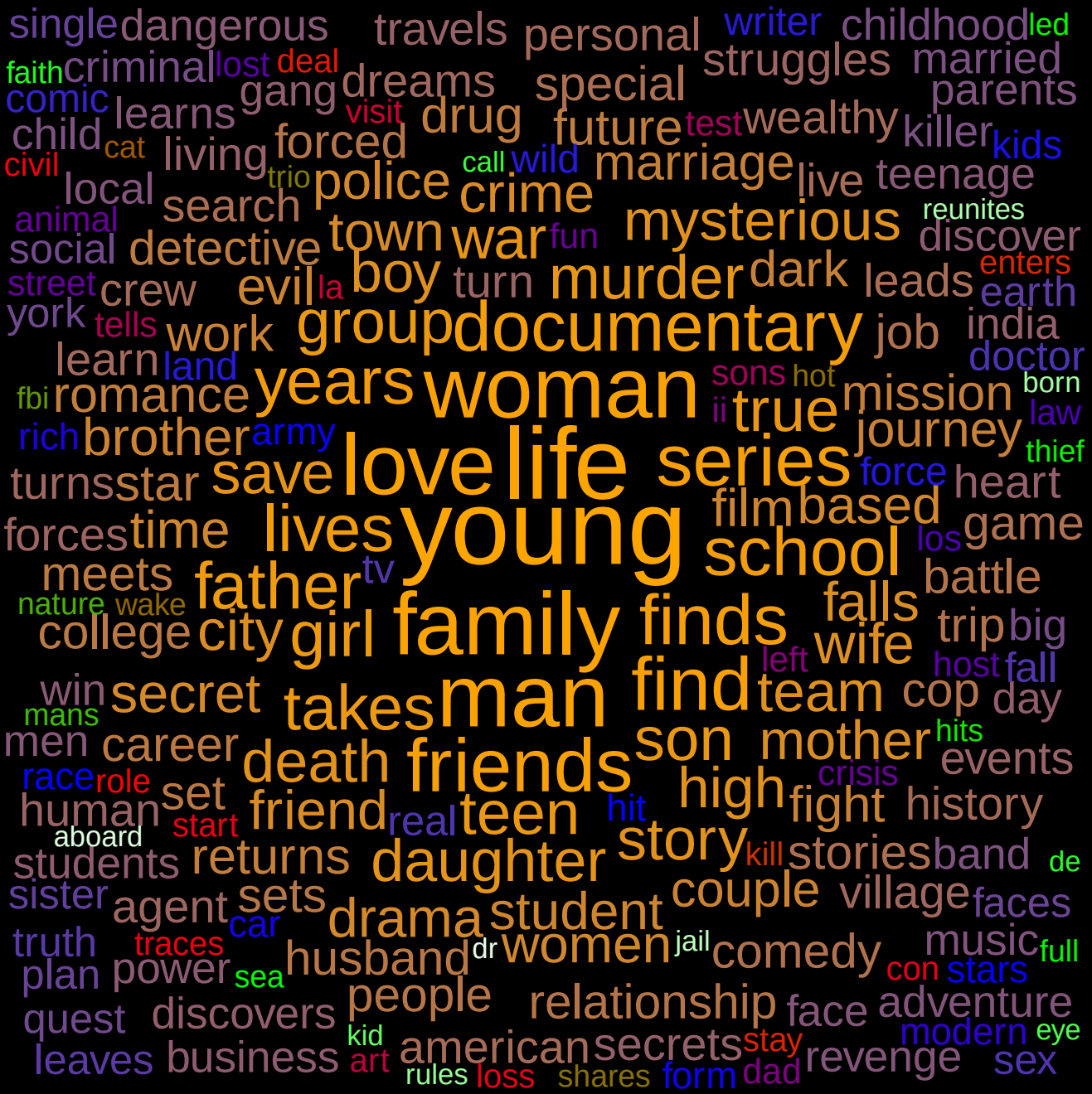


Data Source: Kaggle Data

Multi Dimension Plots




Shows available by Released Year and Rating
Releases
1000
500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year
Rating
PG
PG−13
R
TV−14
TV−G
TV−MA
TV−PG
TV−Y
TV−Y7
Data Source: Kaggle Data



Source: Kaggle Data