How to find a Disney movie that everyone likes when everyone get together.

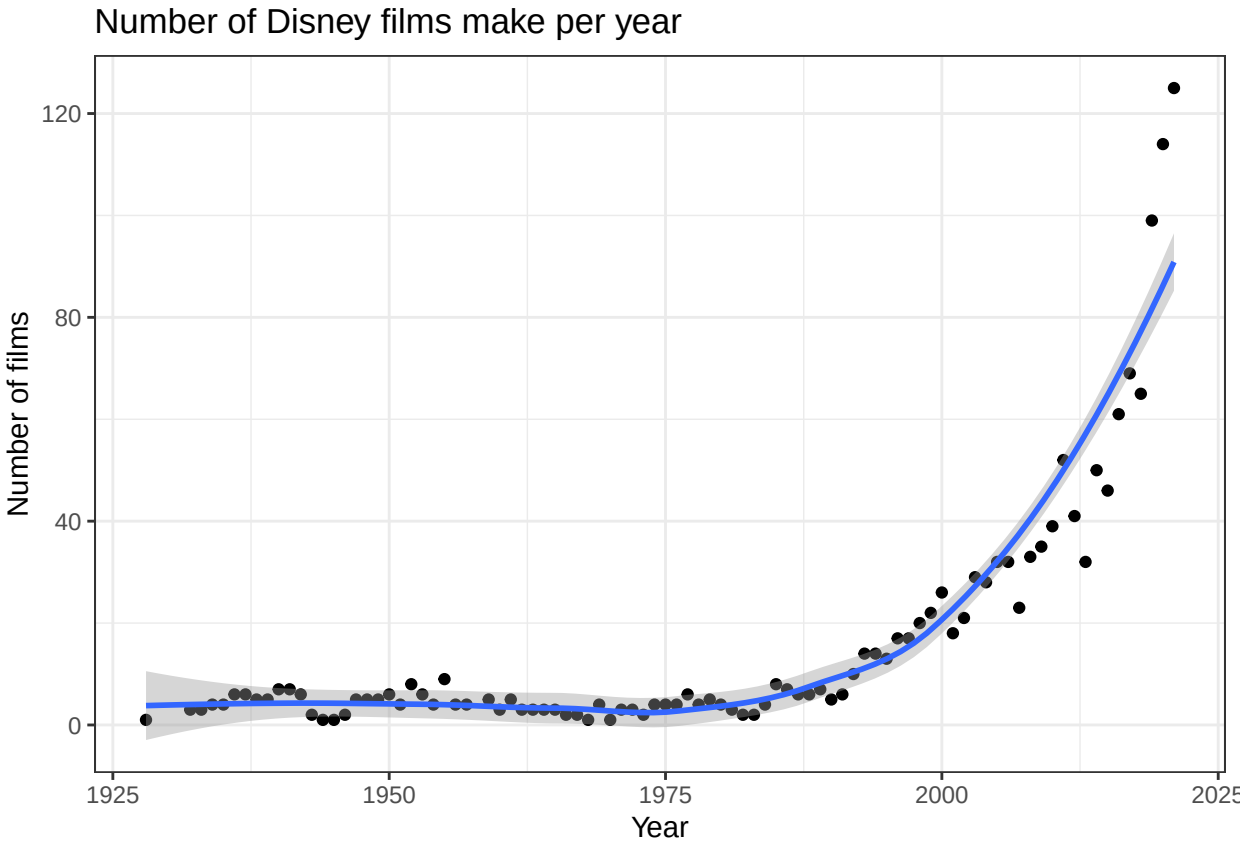
Chuyan YU/Du Rui

2022-12-05

Introduction

These two years Disney movies are getting bigger and bigger, and the number of movies released each year is becoming more and more. Before, we only knew that Disney made animation, but now Disney not only makes animation movies, but also action movies.

There are more and more movies to choose from on a daily basis, as shown in the picture below. The entire Disney film industry has been growing since the year 2000.



Probelm We Facing

A very important question arises at this time. Disney makes so many movies every year. If we can't filter out some of the junk and pick only the quality movies, the extra choices will waste a lot of our time. So it is very important to choose a good quality movie.

In addition, when we watch a movie together, we should look for a movie that meets the tastes of the public, we do not need to enjoy it alone, we need to think about the public and the circulation of the film. Some movies may be very popular for a small group of people, but for the majority of people, they may not be so fond. Therefore, when we get together to choose a movie, we should choose the one that meets the taste of the public, not the one that meets our own taste.

Here I will analyze how we should choose a movie from two perspectives.

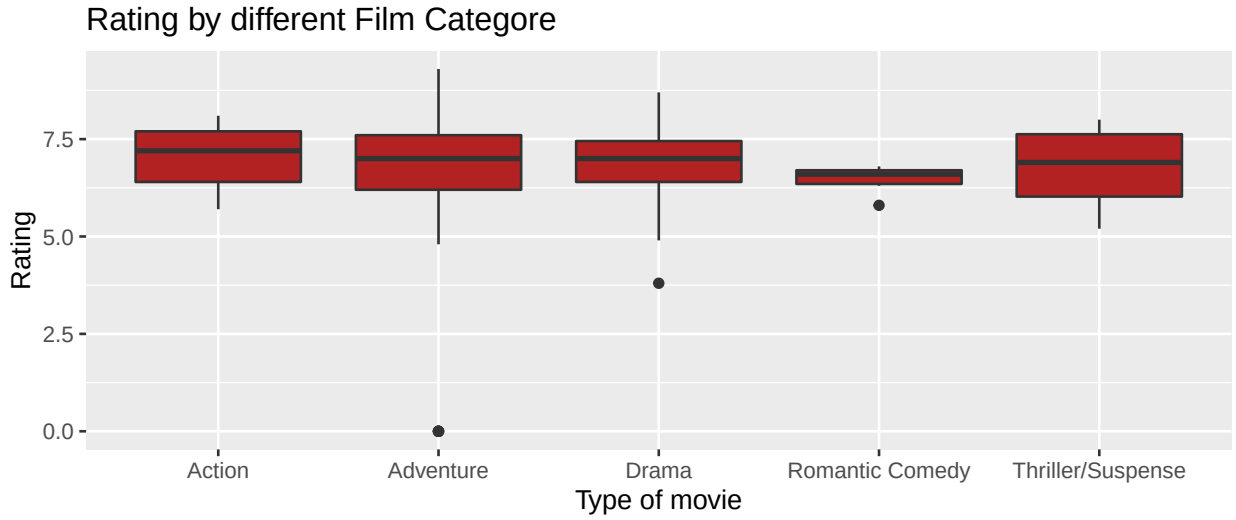
Data Recource

For this research project, I used two different data sources, one of which was found through Github and the other from Kaggle. Both of these sources basically contain information about the box office of the movie, the name of the movie, the type of movie, the release period of the movie, and the rating of the movie. But to make the whole data more useful, I have organized my data. For example, in this two data bases the movie release dates and movie ratings are in two tables, so I need to select the same movie titles and put their release dates and ratings in one table.

First we can look Different film subjects

Disney releases thousands of movies each year, but each store has its own theme, such as a comedy movie, a horror movie or an anime movie.

We all know that Disney started out as an animated movie artist, and he created many familiar characters, such as Mickey Mouse and what else? But does this mean that all Disney comedy movies are very successful. Let's analyze the data from Github[1]. What is the average rating of each movie. So let's look at what Disney is good at.



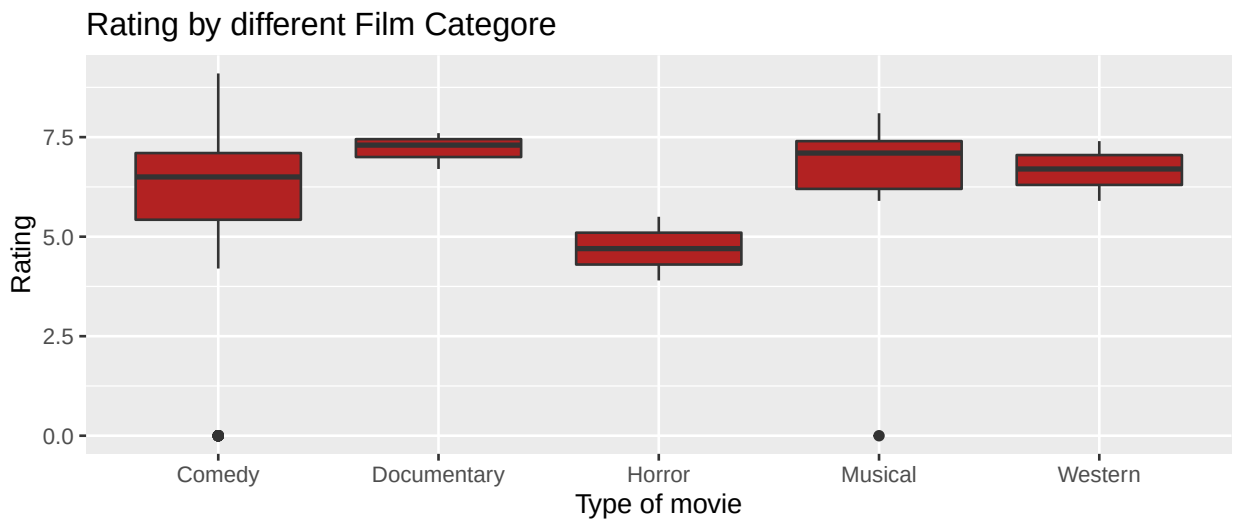


Table 1: Film Categore and Rating separate relation for Movie

	Subject matter	Rating	Min	First quartile	Meaidan	Third quartile	Mean	Stander Divination	Max
11	Action	Rating	5.7	6.400	7.2	7.700	7.007692	0.7410146	8.1
12	Adventure	Rating	0.0	6.200	7.0	7.600	6.670833	1.6802830	9.3
13	Black Comedy	Rating	5.6	5.600	5.6	5.600	5.600000	NA	5.6
14	Comedy	Rating	0.0	5.425	6.5	7.100	5.730000	2.4662873	9.1
15	Documentary	Rating	6.7	7.000	7.3	7.450	7.200000	0.4582576	7.6
16	Drama	Rating	3.8	6.400	7.0	7.450	6.934286	1.0140028	8.7
17	Horror	Rating	3.9	4.300	4.7	5.100	4.700000	1.1313708	5.5
18	Musical	Rating	0.0	6.200	7.1	7.400	6.322222	2.4828299	8.1
19	Romantic Comedy	Rating	5.8	6.350	6.6	6.700	6.510000	0.3107339	6.8
110	Thriller/Suspense	Rating	5.2	6.025	6.9	7.625	6.750000	1.2556539	8.0
111	Western	Rating	5.9	6.300	6.7	7.050	6.666667	0.7505553	7.4

The Box plot above shows that the comedies we thought we'd see didn't lead the ratings, but rather the documentaries and action films were rated higher.

In addition, through my analysis of the annual release figures for these two types of films below, it appears that documentaries and action films are becoming more and more important to Disney. Before 2010, Disney released very few action movies and documentaries, but after 2010, Disney released more action movies and documentaries, so we can assume by this change that the public's taste is now more inclined to action movies and documentaries. Therefore, it is a good choice to choose action and documentaries when choosing these movies

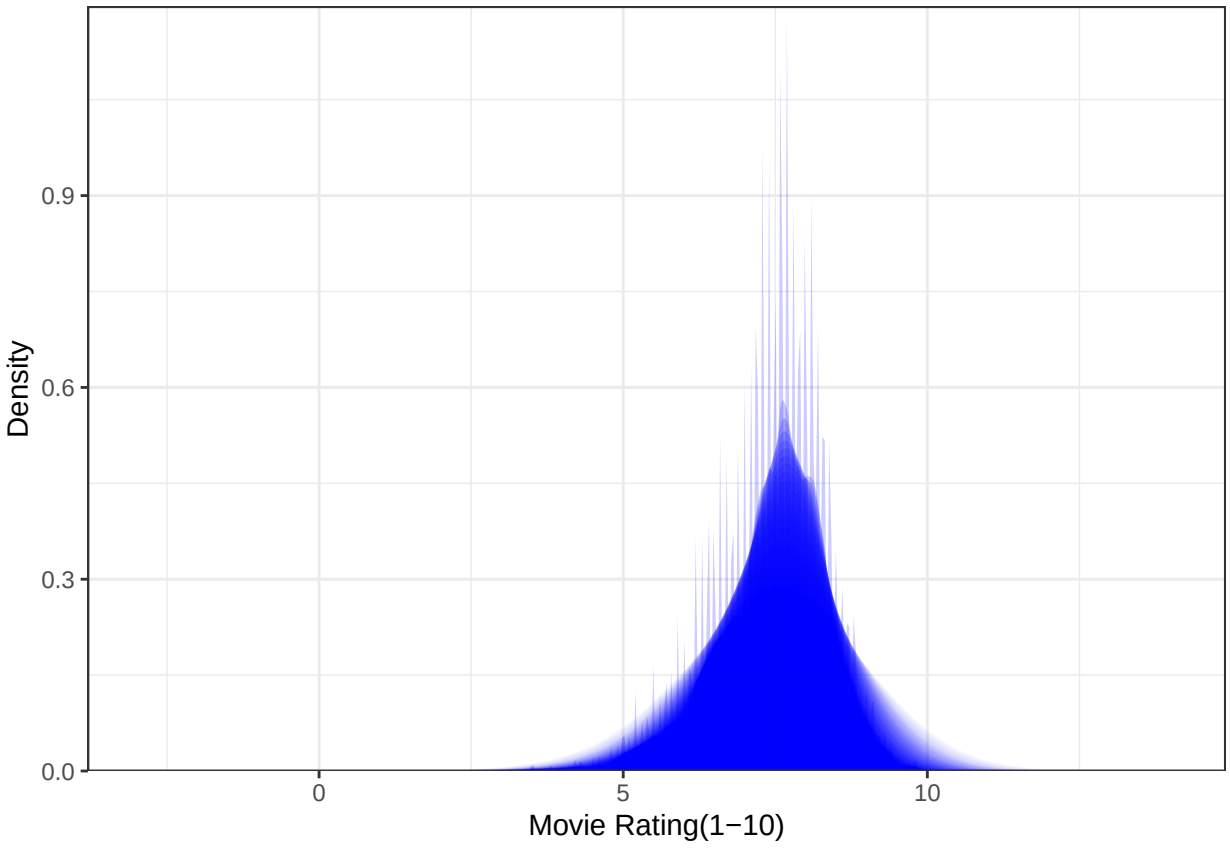
Analysis of popularity from ratings

Now we have a better variety of movies to watch. But this one choice still doesn't allow us to make a good choice. Because for a certain category, Disney can also make bad movies.

In addition, we can see that the ratings of individual movies are too extreme, and we should think about whether the ratings are true and reliable, or whether the movies with high ratings are popular.

So, in the chart below, I have analyzed the frequency of all Disney ratings. The number of ratings largely determines the popularity of the movie. A higher number of scores and reviews means that the movie is more popular.

We can see that in the following chart, I analyzed the frequency of all Disney ratings, and we can see that the highest number of ratings is around 7.5. It also accounts for the highest percentage. So we can focus on the movies that have a rating around 7.5. This is to ensure that most people are interested in the movie.



Summarize

With my analysis above we have basically determined how to find a movie that we all like when we all watch movies together.

Now for the movie categories, we can focus on action movies and documentaries, because these two categories are the ones with higher ratings compared to other Disney movies.

When we choose these two movies, we can also focus on those with a rating around 7.5. This will ensure that the content of the movie will be in line with the public taste and that there will not be a situation where only a small number of people will like the movie and give it a high rating, while others will give it a very low rating.

To sum up, we should choose action movies and documentaries around 7.5 as the source for watching movies together. This will not only ensure that the quality of the movie is not bad, but also that everyone will love the movie and no one will be bored.

Reference

[1] Majumder, P . (2019), Disney Movies Data . [Kaggle Dataset] . Available from https://www.kaggle.com/datasets/prateekmaj21/disney-movies

[2] duncandam86 (2019), Disney Movies and Box Office Success . [Github Dataset] . Available from https://github.com/duncandam86/DataCamp-Projects/tree/master/Disney%20Movies%20and%20Box%20Office%20Success/datasets

```
knitr::opts_chunk$set(echo = TRUE)

library(tidyr)  #load package
library(ggplot2)
library(psych)
library(dplyr)
library(kableExtra)
library(readxl)

DisneyBox <- read_excel("/Users/yuchuyan/Desktop/DisneyBox.xlsx")  #input/read  data


disney <- read.csv(#input/read  data
  file="/Users/yuchuyan/Desktop/disney_plus_projects.csv",
  header=TRUE,
  sep=","
)

disney2 <- read.csv(#input/read  data
  file="/Users/yuchuyan/Desktop/disney_plus_titles.csv",
  header=TRUE,
  sep=","
)



Disney6 <- inner_join( #combine the data
  x = disney,
  y = DisneyBox,
  by = c("title" = "movie_title")
)

disney2 %>%  #first data
  group_by(release_year) %>% #clean data
  summarize(n = n()) %>%
  ggplot(mapping = aes(x = release_year, y = n),) +
  geom_point(fill = "#8B4513") +
  geom_smooth() +
  labs(
    x = "Year",  #graph label
    y = "Number of films",
    title = "Number of Disney films make per year"
  )+
  theme_bw()
Disney6 %>%
 filter(genre.y %in% c("Adventure", "Thriller/Suspense", "Action", "Romantic Comedy", "Drama")) %>%  #c
 ggplot() +
  aes(x = genre.y, y = rating) +
  geom_boxplot(fill = "#B22222") +
  theme_gray()+
```

```
  labs(
    x = "Type of movie",  #graph label
    y = "Rating",
    title = "Rating by different Film Categore ",
  )



Disney6 %>%
 filter(genre.y %in% c("Comedy", "Musical", "Western", "Horror", "Documentary")) %>% #clean data
 ggplot() +
  aes(x = genre.y, y = rating) +
  geom_boxplot(fill = "#B22222") +
  theme_gray()+
  labs(
    x = "Type of movie", #data label
    y = "Rating",
    title = "Rating by different Film Categore ",
  )



data= Disney6 #Film Categore and Rating separate relation for Movie tabel
disneystat <- describeBy(
  x = Disney6$rating, #data resource
  group = Disney6$genre.y,
  data = Disney6,
  mat = TRUE,
  quant = c(0.25, 0.75)
) %>%
  mutate(
    vars = case_when(
      vars == 1 ~ "Rating",
    )
  ) %>%
  select("Subject matter" = group1, Rating = vars,  #determine want we want to show
         Min=min, "First quartile"
         = Q0.25 ,Meaidan=median,"Third quartile" = Q0.75, Mean=mean,"Stander Divination"=sd,Max=max) #
disneystat %>%
  kable(
    caption = "Film Categore and Rating separate relation for Movie " , #table name
    booktabs =TRUE,
    align = c("l", rep("c",8))

  ) %>%      #basic table set
  kableExtra::kable_styling( #graph style and position
    bootstrap_options = c("striped", "condensed"),
    font_size = 16,
    latex_options = c("scale_down","HOLD")


  )
```

```r
DisneyY <-disney %>%
  filter(rating>0)   #clean data

source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")  #use shadow graph f
shadowgram(  #shadow graph function
  dataVec =(DisneyY$rating),
  label = "Movie Rating(1-10)",
  layers = 30,
  aStep = 6,
  color = "blue"
)


# show all the code at the end of pdf
```