Project Part 1

Description of Data

Background Information

This dataset is a list of all the Disney movies released before the year 2017 and contains their genre, MPAA rating, box office earnings in the dollar amount of the year released, the dollar amount adjusted for inflation, the month the movie was released, and the year the movie was released. Disney is one of the most successful film companies of all time, making movies since the 1930s and continuing to have an impact on the industry into the 21st century (https://thewaltdisneycompany.com/about/#our-businesses). They have produced some of the most popular films of all time, and this dataset presents those as well as their least successful attempts at the box-office.

Explanation of the rows and variables

Each row contains the name of the film, genre, rating, year released, month released, nominal box office earnings, and inflation-adjusted box office earnings. There are 579 movies in the list spanning from December of 1937 (Snow White and the Seven Dwarves) to March of 2016 (Zootopia). The ratings, although not constant over time due to changes in the industry, are G, PG, PG-13, and R. The genres encapsulate every popular genre in the movie industry. The real earnings are necessary because it gets rid of the temptation to compare movie earnings without taking inflation into account, so including this allows us to compare them on a level playing field. We can use the following breakdowns for the data:

Movie name, genre, rating, year: factors

Nominal earnings, real earnings: numeric

Month: character

Population

The data represents a population. This dataset includes every Disney movie that has been released since 1937. The data was collected from various websites that track these particular statistics for each movie. The most important website was (https://www.the-numbers.com/

movies/distributor/Walt-Disney#tab=year), which includes the name of the movies and information regarding the earnings per movie. I accessed this data from data.world, which allows access to data publically.

Potential issues

Some of the films lack a rating or have a rating that was determined in a different time period, which could make time an underlying factor that affects rating. The MPAA was established in 1968, so the ratings of movies before then are not necessarily indicative of their potential rating at the time of release (https://www.motionpictures.org/film-ratings/). Also, genre, rating, and both variables concerning earnings have some datapoints that lack entries. Therefore, we have to omit some of those points when comparing those variables despite our desire to take all datapoints into account. One of the glaring problems with a dataset like this is the lack of randomness when talking about the earnings of each film. Obviously, some stories are far better than others and deserve to perform better at the theatre. For example, the story of Snow White and the Seven Dwarves has been one of the most popular movies throughout its entire existence and its popularity is shown in its earnings. In short, the storyline and narrative are the factors that are determining its success, not necessarily any of the factors we are presented with in this dataset.

Numerical Summary

```
library(ggplot2)

# NECESSARY CODING TO FIND AVERAGE EARNING FOR EACH GENRE

# Average_Earnings1 <- as.data.frame(tapply(disney1$Real_Earnings,
# disney1$Genre, mean))

# Average_Earnings2 <- as.numeric(Average_Earnings1[-9,])

# Genres1 <- c("Unassigned Genre", "Action", "Adventure",
# "Black Comedy", "Comedy", "Concert/Performance", "Documentary",
# "Drama", "Horror", "Musical", "Romantic Comedy",
# "Thriller/Suspense", "Western")
```

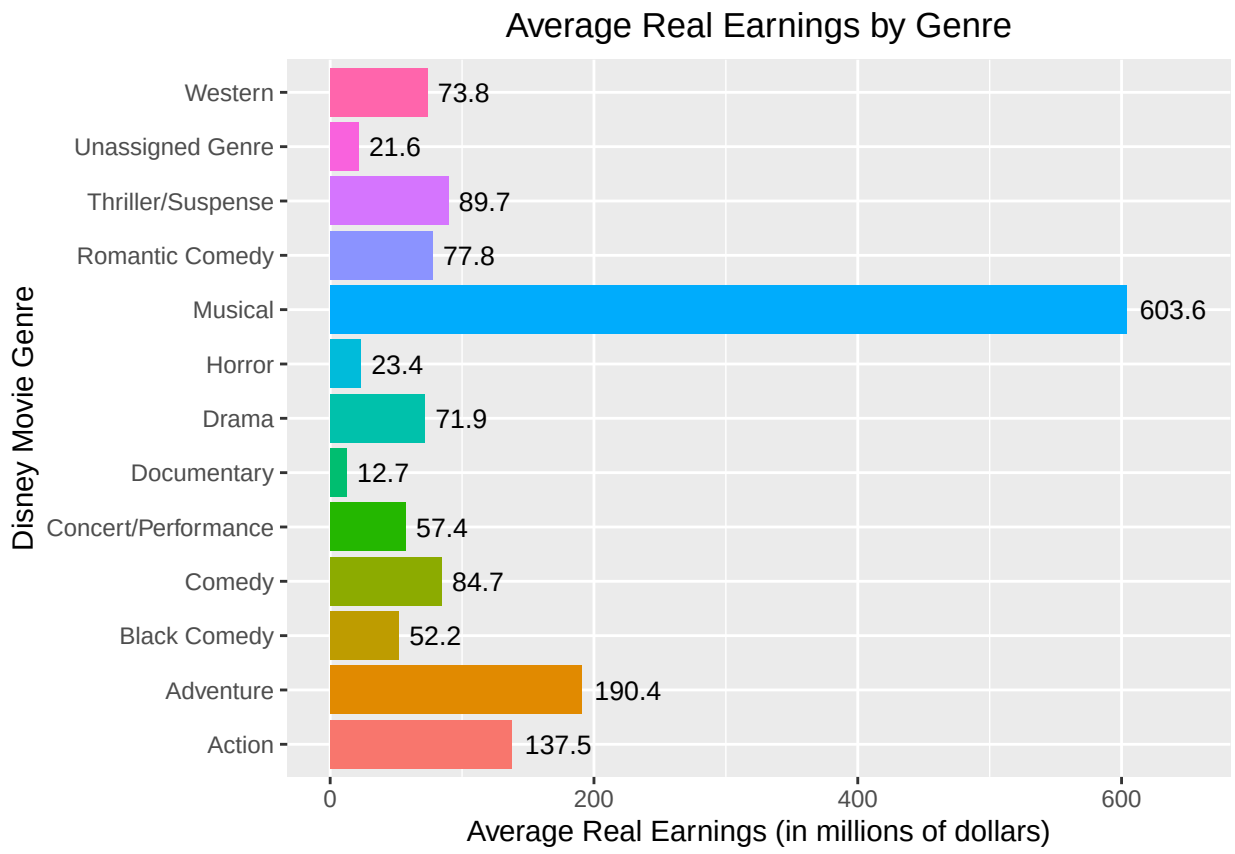
```r
# Average_Earnings3 <- data.frame(Genres1, Average_Earnings2)

# WHEN TRANSFERRING THE DATA FROM ANOTHER R MARKDOWN FILE, THE CODE ABOVE
# DID NOT KNIT CORRECTLY, SO I WROTE THE FILE OUT AND READ IT BACK IN
# TO MAKE THE GRAPH THE SAME AS THE OTHER FILE.

Average_Earnings4 <- read.csv("/Users/joey/Downloads/Average_Earnings3")

ggplot(Average_Earnings4, aes(x=Genres1,
                              y=(round(Average_Earnings2/1000000,1)),
                              fill=Genres1)) + geom_bar(stat="identity") +
  theme(legend.position="none", plot.title = element_text(hjust = 0.5)) +
  labs(title="Average Real Earnings by Genre", x="Disney Movie Genre",
       y="Average Real Earnings (in millions of dollars)") +
  stat_summary(geom="text", aes(label=..y..), color="black",
               size = 3.5, hjust=-.2) + coord_flip(ylim = c(0, 650))
```

```
## No summary function supplied, defaulting to `mean_se()
```

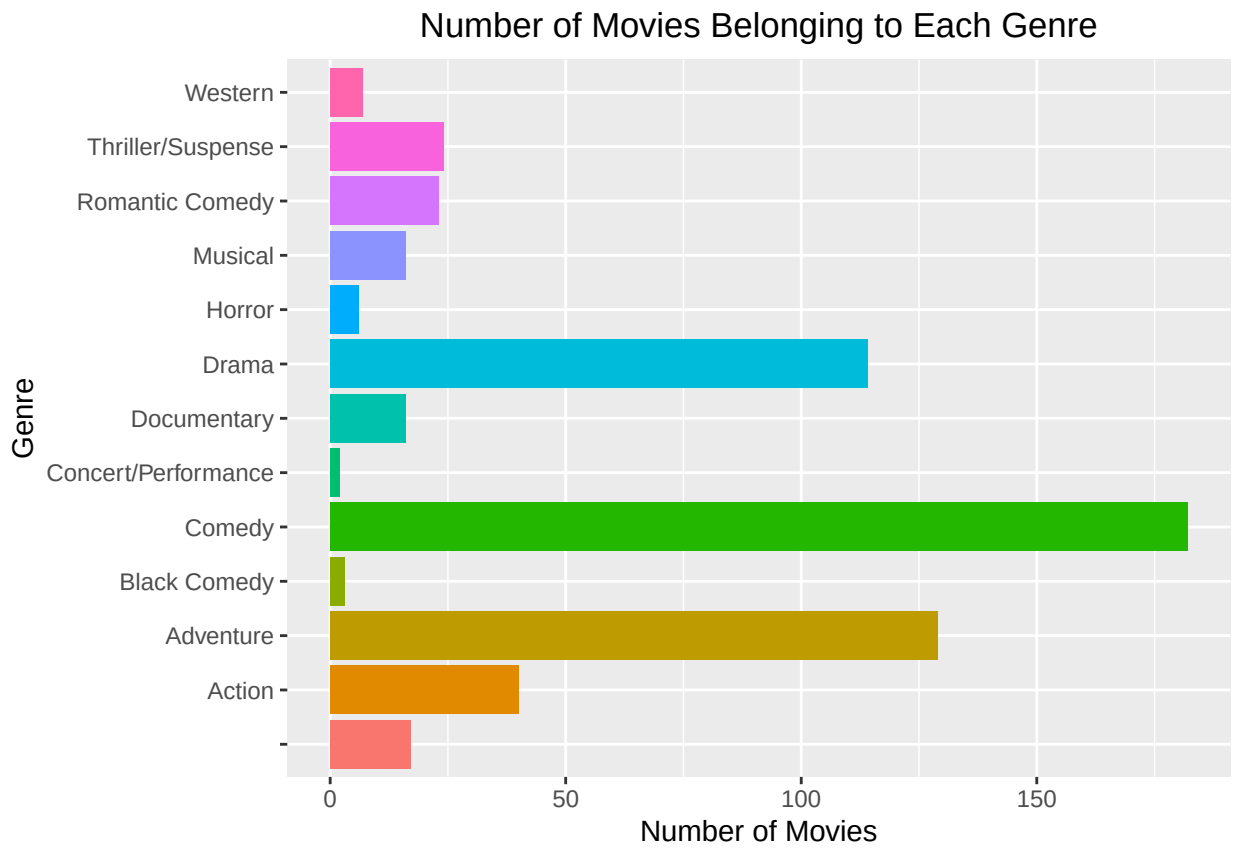


Conclusion

This graphic shows the average earnings for films belonging to each genre. The genre of each Disney movie is clearly related to the box-office success. Musicals are by far the most successful at the box office and documentaries bring in the lowest earnings. Action and adventure movies are also some of the most successful in the box office. If Disney was determining what genre of movie to release their next movie under, I would suggest that they work on a musical. However, adventure and action movies are also worth pursuing.

Graphical Summary

```
ggplot(disney, aes(x=Genre, fill=Genre)) + geom_bar() +
  theme(legend.position="none", plot.title = element_text(hjust = 0.5)) +
  coord_flip() +
  labs(title="Number of Movies Belonging to Each Genre", x="Genre",
       y="Number of Movies")
```



Conclusion

This graph summarizes the number of films that belong to each genre and puts the last graphic into perspective. We can see that because Disney has released a relatively low number of musicals, the potential earnings have less weight than a movie genre such as comedy or adventure. Adventure films have the second highest average earnings and second highest number of films released under that genre, so a safe bet would be to work on those films in light of both graphics. The salmon colored bar represents the movies that had an unassigned genre in the dataset. Ideally, they would be assigned a genre, but for some reason, the dataset lacked a label for them.

References

1. https://data.world/kgarrett/disney-character-success-00-16/workspace/file?filename=disney_movies_total_gross.csv
2. https://www.sugarcane.com/data/walt-disney-animation-studios-films-1
3. http://www.the-numbers.com/movies/distributor/Walt-Disney

4. https://en.wikipedia.org/wiki/List_of_Disney_animated_universe_characters
5. https://thewaltdisneycompany.com/about/#our-businesses
6. https://www.motionpictures.org/film-ratings/