Final Project Report (CS&SS 321)

Emma Favier, Georgia Pertsch, Aakash Krishna, Ethan Side

Introduction, research question, and hypotheses

According to the United States Bureau of Economic Analysis, the Gross Domestic Product, or GDP, is "a comprehensive measure of economic activity... [and] the most popular indicator of the nation's overall economic health" (USBEA 2023). GDP measures the total monetary value of goods and services produced in the US in any given year, so it can be used to compare between years or countries to reveal economic productivity. Countries are given a label of high-income, upper-middle income, lower-middle income, or low income depending on what range the GDP of the given country falls into. Economic prosperity of a given country and the successful industries within that country are related, as industry directly influences the economy and vice versa.

The global film industry "brings jobs, revenue, and related infrastructure development" created by the numerous departments within film, such as construction, production, and even tourism, "providing an immediate boost to local economies" (Motion Pictures Association, Inc. 2023). The potential for film industries to thrive relies on the accessibility to basic infrastructure. While the film industry provides essential work for thousands of people, it is not part of the basic and essential industries that fuel the economy's growth like healthcare, construction, technology, and manufacturing. Thus, the film industry can be viewed as one indicator of a nation's economic prosperity based on its size and scale. When considering the countries included in the top 10 GDPs globally, the top five film industries in the world–the United States, China, United Kingdom, Japan, and India–are within the top six GDP (World Atlas 2018). Ratings may be another indicator to inform how successful a given country's film industry is among the citizens, those who actually watch the film that their country produces. It is important to note that, in this context, film includes both movies and television shows.

Research question and hypotheses

In this research project, we are investigating the relationship between GDP and ratings of film in our given countries: United States, China, United Kingdom, Japan, and India. Our research questions are: are film ratings higher (as in more restricted) in countries with higher GDPs and lower (as in less restricted) in countries with lower GDPs, and, do countries with higher GDPs produce more movies than countries with lower GDPs?

We suspect that countries in higher GDP ranges will have higher film ratings, and countries in lower GDP ranges will have lower film ratings. Additionally, we speculate that countries in a higher GDP range on average produce more movies than countries in a lower GDP range because of accessibility of resources needed for film to thrive.

Research design (Part 1)

The main data set we use consists of listings of all the movies and TV shows available on Netflix, along with details such as genre, cast, rating, and release year. We decided to look at four of the higher producing entertainment countries in the world, two being high income and two being lower middle income, choosing

these countries as those in the two income groups that produced the most movies. We wanted to see if countries that are not high income have a different movie genre proportion than lower income countries. In order to do so, we first merged our GDP and movie data sets, and created data subsets for the 3 primary entertainment genres we chose to focus on: action, drama, and comedy.

```
gdp_data$country = gdp_data$TableName
merged_data <- data %>% left_join(gdp_data)
```

```
## Joining with `by = join_by(country)`
```

```
merged_data <- merged_data %>% rowwise() %>% mutate(Drama = ifelse(("Dramas" %in% listed_in) | ("TV Dra
merged_data <- merged_data %>% rowwise() %>% mutate(Comedy = ifelse(("Comedies" %in% listed_in) | ("TV 
merged_data <- merged_data %>% rowwise() %>% mutate(Action = ifelse(("Action & Adventure" %in% listed_i
merged_data <- merged_data %>% filter(!is.na(IncomeGroup) & IncomeGroup != "")

drama_data_set <- merged_data %>% mutate(Drama1 = grepl("Dramas", listed_in))
comedy_data_set <- merged_data %>% mutate(Comedy1 = grepl("Comedies", listed_in))
action_data_set <- merged_data %>% mutate(Action1 = grepl("Action & Adventure", listed_in))

drama_plot <- ggplot(drama_data_set, mapping = aes(x = IncomeGroup, fill = Drama1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Drama Movies and TV shows by Income Group", x = "Income Group", y = "Total
ggplotly(drama_plot)
```

```
comedy_plot <- ggplot(comedy_data_set, mapping = aes(x = IncomeGroup, fill = Comedy1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Comedy Movies and TV shows by Income Group",x = "Income Group", y = "Total
ggplotly(comedy_plot)
```

```
action_plot <- ggplot(action_data_set, mapping = aes(x = IncomeGroup, fill = Action1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Action Movies and TV shows by Income Group", x = "Income Group", y = "Tota
ggplotly(action_plot)
```

After looking at this data, we then wanted to test each country's proportion of Drama movies and Tv shows produced to see if lower middle income countries were more inclined to produce Dramas as it may relate to their population more. We then did the same with comedies to see if higher income countries had a similar trend in the other direction.

We filtered these four producers of film and then individually compared their drama and comedy-producing proportions.

```
US_data_set <- drama_data_set %>% filter(TableName == "United States")
Nigeria_data_set <- drama_data_set %>% filter(TableName == "Nigeria")
India_data_set <- drama_data_set %>% filter(TableName == "India")
UK_data_set <- drama_data_set %>% filter(TableName == "United Kingdom")
US_drama_proportion <- table(US_data_set$Drama1)
Nigeria_drama_proportion <- table(Nigeria_data_set$Drama1)
India_drama_proportion <- table(India_data_set$Drama1)
```

```r
UK_drama_proportion <- table(UK_data_set$Drama1)

(US_drama_proportion[2] / (US_drama_proportion[1] + US_drama_proportion[2]))
```

```
##      TRUE
## 0.2767921
```

```r
(Nigeria_drama_proportion[2] / (Nigeria_drama_proportion[1] + Nigeria_drama_proportion[2]))
```

```
##      TRUE
## 0.6736842
```

```r
(India_drama_proportion[2] / (India_drama_proportion[1] + India_drama_proportion[2]))
```

```
##      TRUE
## 0.6656379
```

```r
(UK_drama_proportion[2] / (UK_drama_proportion[1] + UK_drama_proportion[2]))
```

```
##      TRUE
## 0.1622912
```

```r
US_data_set1 <- comedy_data_set %>% filter(TableName == "United States")
Nigeria_data_set1 <- comedy_data_set %>% filter(TableName == "Nigeria")
India_data_set1 <- comedy_data_set %>% filter(TableName == "India")
UK_data_set1 <- comedy_data_set %>% filter(TableName == "United Kingdom")
US_comedy_proportion <- table(US_data_set1$Comedy1)
Nigeria_comedy_proportion <- table(Nigeria_data_set1$Comedy1)
India_comedy_proportion <- table(India_data_set1$Comedy1)
UK_comedy_proportion <- table(UK_data_set1$Comedy1)

(US_comedy_proportion[2] / (US_comedy_proportion[1] + US_comedy_proportion[2]))
```

```
##      TRUE
## 0.2689851
```

```r
(Nigeria_comedy_proportion[2] / (Nigeria_comedy_proportion[1] + Nigeria_comedy_proportion[2]))
```

```
##      TRUE
## 0.4315789
```

```r
(India_comedy_proportion[2] / (India_comedy_proportion[1] + India_comedy_proportion[2]))
```

```
##      TRUE
## 0.3436214
```

```
(UK_comedy_proportion[2] / (UK_comedy_proportion[1] + UK_comedy_proportion[2]))
```

```
##      TRUE
## 0.1622912
```

Research Design (Part 2)

The research design for this portion of our analysis is breaking down how ratings change in one country as GDP fluctuates in order to to analyze without the different impacts studying countries different countries can have. In order to do this, we created a merged data file that was merged with both country and year/ year of release as the common factor.

```
datax <- read_csv("netflix_titles.csv") %>%
  mutate(year = release_year) %>%
  filter(country %in% c("United States","India","United Kingdom"))
```

```
## Rows: 8807 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): show_id, type, title, director, cast, country, date_added, rating,...
## dbl  (1): release_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
gdpchange <- read_csv("gdpchange.csv") %>%
  filter(country %in% c("United States","India","United Kingdom"))  %>%
  pivot_longer(GDP_1960:GDP_2021, names_to = "year", values_to = "gdp") %>%
  mutate(year = as.numeric(str_sub(year, 5L, -1L))) %>%
  select(-`2022`)
```

```
## Rows: 266 Columns: 67
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): country, Country Code, Indicator Name, Indicator Code
## dbl (61): GDP_1961, GDP_1962, GDP_1963, GDP_1964, GDP_1965, GDP_1966, GDP_19...
## lgl  (2): GDP_1960, 2022
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
mergeddata <- datax %>%
  left_join(gdpchange)
```

```
## Joining with `by = join_by(country, year)`
```

Then, we created a new column labeled ' GDPcategory ' which was labelled 1, 2, 3, or 4 based on which quartile the country's GDP was in that year.

```
usa <- mergeddata %>% filter(country == "United States")
summary(usa$gdp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  -2.768   1.667   2.288   1.900   2.945   7.237      21
```

```
usa <- usa %>%
  mutate(GDPcategory = case_when(gdp <1.667 ~ 1,
                                  between(gdp, 1.667, 2.288)~ 2,
                                 between(gdp, 2.288, 2.945)~ 3,
                                 TRUE ~ 4 ))
```

Then we made 4 different graphs showing the percent break down of rating dependent on for the years in which the US was either in quartile 1, 2, 3, or 4.

We will be looking at the differences in these graphs to view change.

Data: the source of the data, measurement, and uses plots to summarize the dependent variable

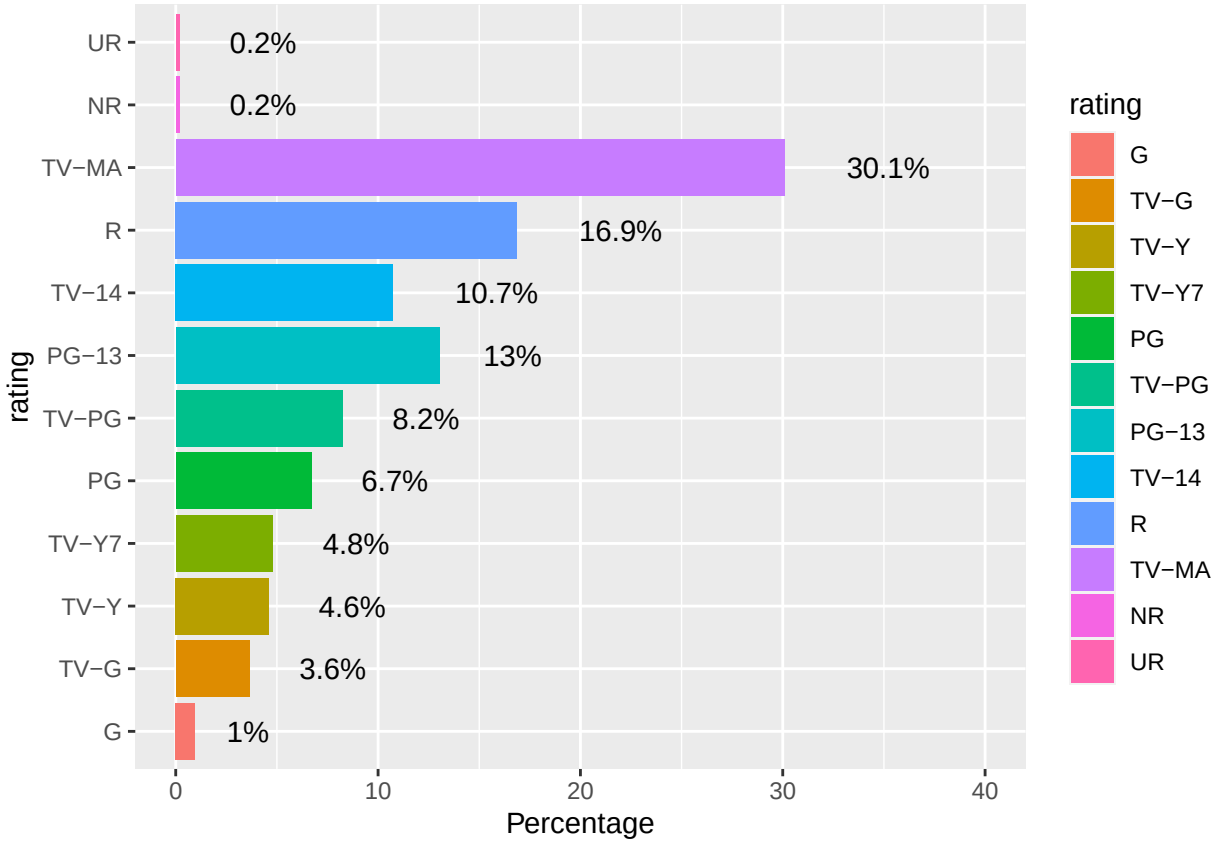
We acquired the film data from Kaggle.com, a website that offers a free and online repository of community-uploaded datasets. In the dataset, GDP was measured as high-income, upper-middle income, lower-middle income, or low income depending on what range a given country falls into. Ratings were measured as TV-7 (suitable for 7 yrs and older) to TV-MA (Mature) for television and G (General Audiences) to R (Restricted) for movies. We acquired the GDP data from the Bureau of Economic Analysis of the United States government.

```
lowestquartileUSA <- usa %>%
  filter(GDPcategory == 1)

rating_counts <- lowestquartileUSA %>%
  count(rating) %>%
  mutate(percentage = n / sum(n)*100)
print(rating_counts)
```

```
## # A tibble: 12 x 3
##    rating     n percentage
##    <chr>  <int>      <dbl>
##  1 G          5      0.958
##  2 NR         1      0.192
##  3 PG        35      6.70
##  4 PG-13     68     13.0
##  5 R         88     16.9
##  6 TV-14     56     10.7
##  7 TV-G      19      3.64
##  8 TV-MA    157     30.1
##  9 TV-PG     43      8.24
## 10 TV-Y      24      4.60
## 11 TV-Y7     25      4.79
## 12 UR         1      0.192
```

```
rating_counts %>%
  rename(Percentage = percentage) %>%
  mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
mutate(rating = factor(rating,
                       levels = c("G","TV-G","TV-Y",
                                  "TV-Y7","PG","TV-PG",
                                     "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
ggplot(aes(x = Percentage, y= rating,
           fill = rating, label = pct_label)) +
geom_col() +
  geom_text(hjust = -0.75) +
  scale_x_continuous(limits = c(0, 40))
```



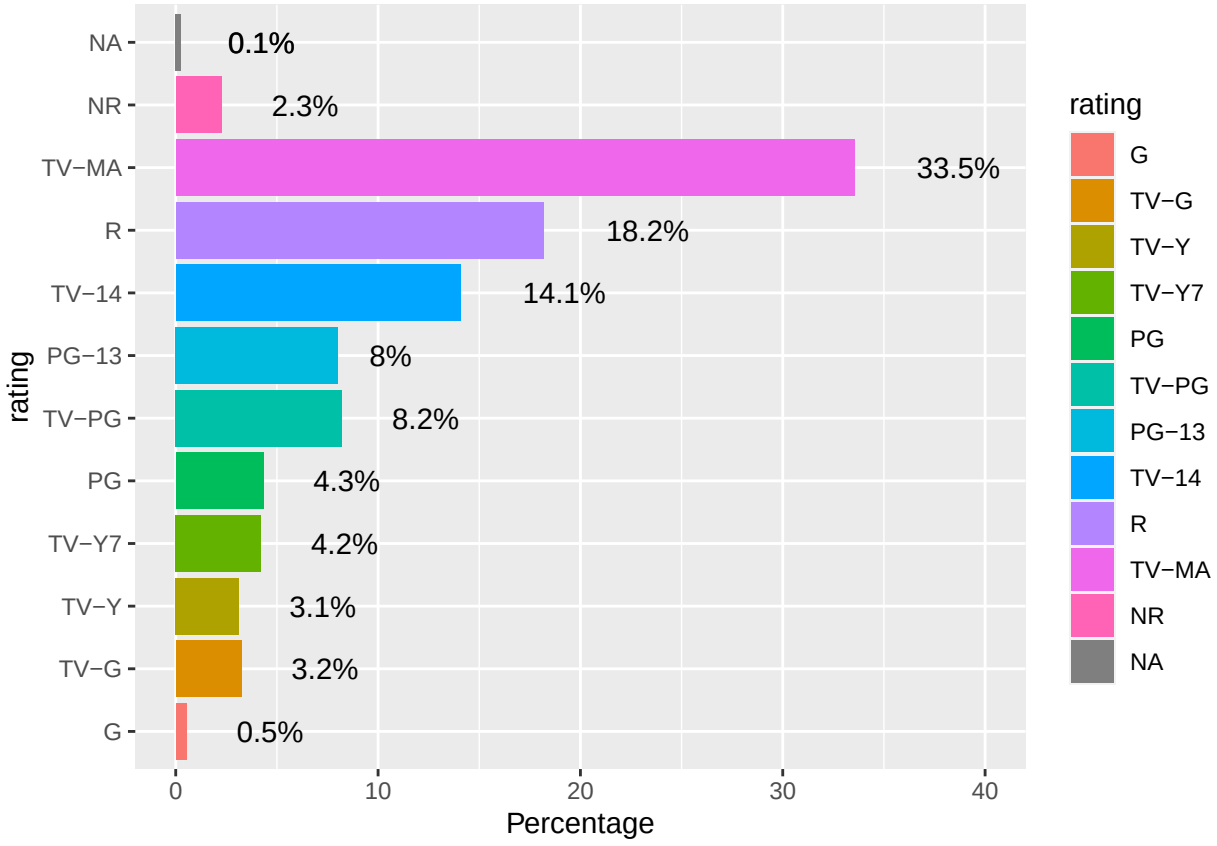

```
secondquartileUSA <- usa %>%
  filter(GDPcategory == 2)

rating_counts2 <- secondquartileUSA %>%
  count(rating) %>%
  mutate(percentage = n / sum(n)*100)
print(rating_counts2)
```

```
## # A tibble: 13 x 3
##    rating       n percentage
##    <chr>    <int>      <dbl>
##  1 74 min       1      0.108
```

```
##  2 G             5      0.541
##  3 NR           21      2.27
##  4 PG           40      4.33
##  5 PG-13        74      8.01
##  6 R           168     18.2
##  7 TV-14       130     14.1
##  8 TV-G         30      3.25
##  9 TV-MA       310     33.5
## 10 TV-PG        76      8.23
## 11 TV-Y         29      3.14
## 12 TV-Y7        39      4.22
## 13 TV-Y7-FV      1      0.108
```

```
rating_counts2 %>%
  rename(Percentage = percentage) %>%
  mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
  mutate(rating = factor(rating,
                         levels = c("G","TV-G","TV-Y",
                                    "TV-Y7","PG","TV-PG",
                                    "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
  ggplot(aes(x = Percentage, y= rating,
             fill = rating, label = pct_label)) +
  geom_col() +
  geom_text(hjust = -0.75) +
  scale_x_continuous(limits = c(0, 40))
```

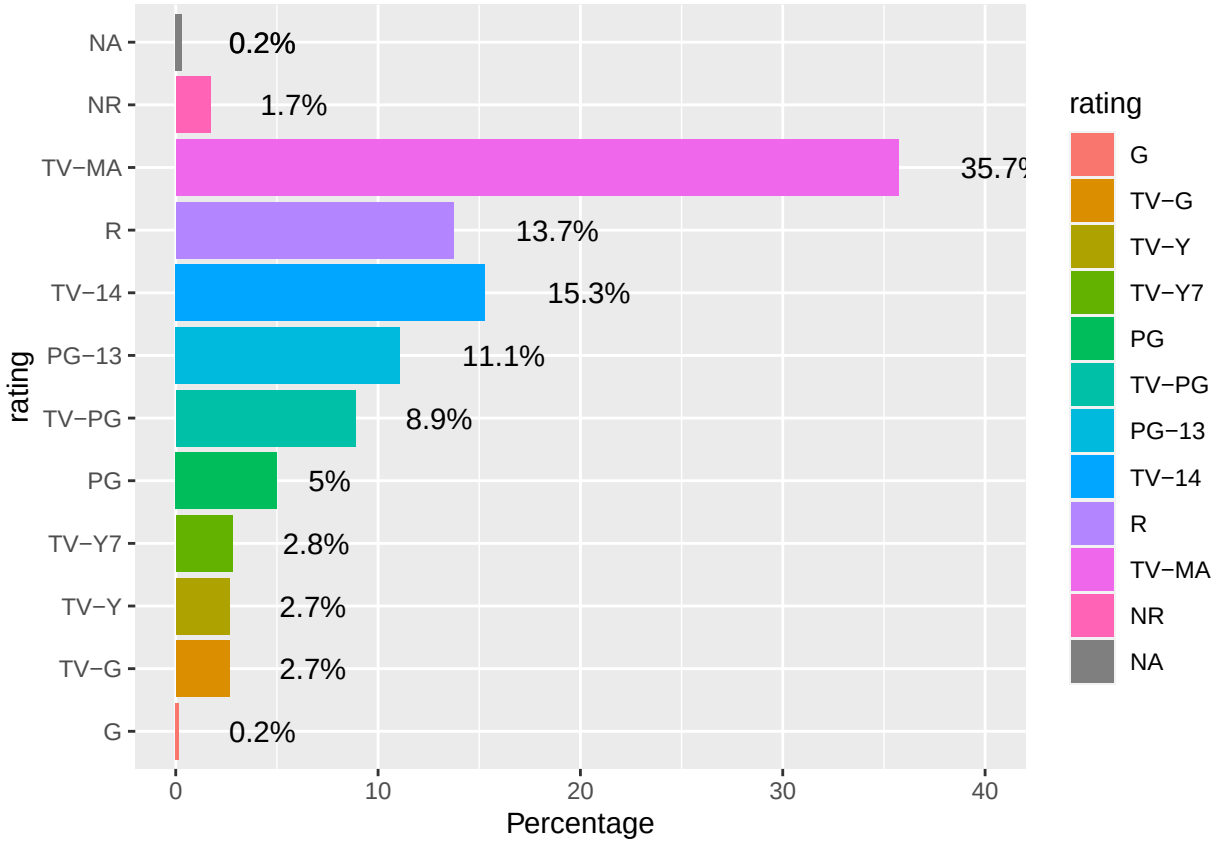


```r
thirdquartileUSA <- usa %>%
  filter(GDPcategory == 3)

rating_counts3 <- thirdquartileUSA %>%
  count(rating) %>%
  mutate(percentage = n / sum(n)*100)
print(rating_counts3)
```

```
## # A tibble: 13 x 3
##    rating     n percentage
##    <chr>  <int>      <dbl>
##  1 66 min     1      0.156
##  2 84 min     1      0.156
##  3 G          1      0.156
##  4 NR        11      1.72
##  5 PG        32      4.99
##  6 PG-13     71     11.1
##  7 R         88     13.7
##  8 TV-14     98     15.3
##  9 TV-G      17      2.65
## 10 TV-MA    229     35.7
## 11 TV-PG     57      8.89
## 12 TV-Y      17      2.65
## 13 TV-Y7     18      2.81
```

```r
rating_counts3 %>%
  rename(Percentage = percentage) %>%
  mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
  mutate(rating = factor(rating,
                         levels = c("G","TV-G","TV-Y",
                                    "TV-Y7","PG","TV-PG",
                                    "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
  ggplot(aes(x = Percentage, y= rating,
             fill = rating, label = pct_label)) +
  geom_col() +
  geom_text(hjust = -0.75) +
  scale_x_continuous(limits = c(0, 40))
```

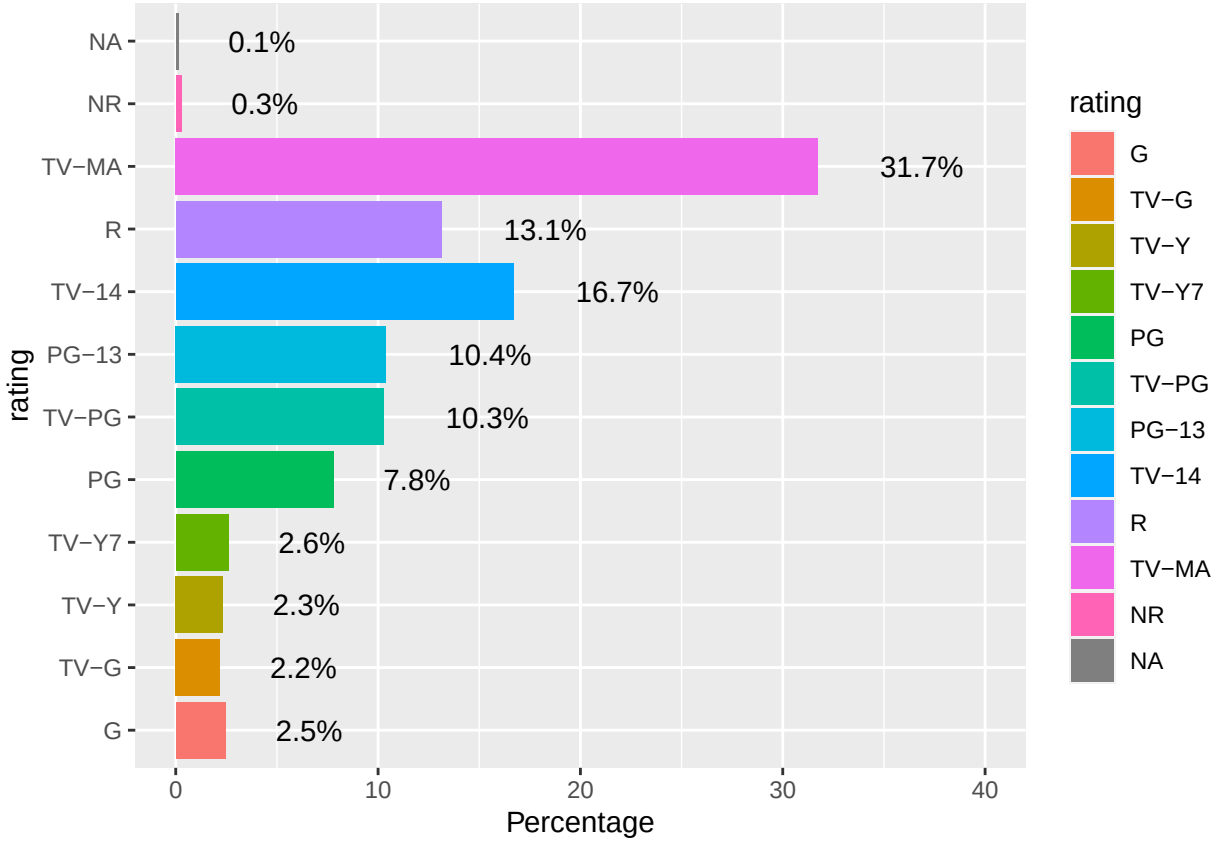



```
fourthquartileUSA <- usa %>%
  filter(GDPcategory == 4)

rating_counts4 <- fourthquartileUSA %>%
  count(rating) %>%
  mutate(percentage = n / sum(n)*100)
print(rating_counts4)
```

```
## # A tibble: 12 x 3
##    rating     n percentage
##    <chr>  <int>      <dbl>
##  1 G         18      2.46
##  2 NC-17      1      0.137
##  3 NR         2      0.274
##  4 PG        57      7.80
##  5 PG-13     76     10.4
##  6 R         96     13.1
##  7 TV-14    122     16.7
##  8 TV-G      16      2.19
##  9 TV-MA    232     31.7
## 10 TV-PG     75     10.3
## 11 TV-Y      17      2.33
## 12 TV-Y7     19      2.60
```

```
 rating_counts4 %>%
   rename(Percentage = percentage) %>%
   mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
   mutate(rating = factor(rating,
                          levels = c("G","TV-G","TV-Y",
                                     "TV-Y7","PG","TV-PG",
                                     "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
   ggplot(aes(x = Percentage, y= rating,
              fill = rating, label = pct_label)) +
   geom_col() +
   geom_text(hjust = -0.75) +
   scale_x_continuous(limits = c(0, 40))
```

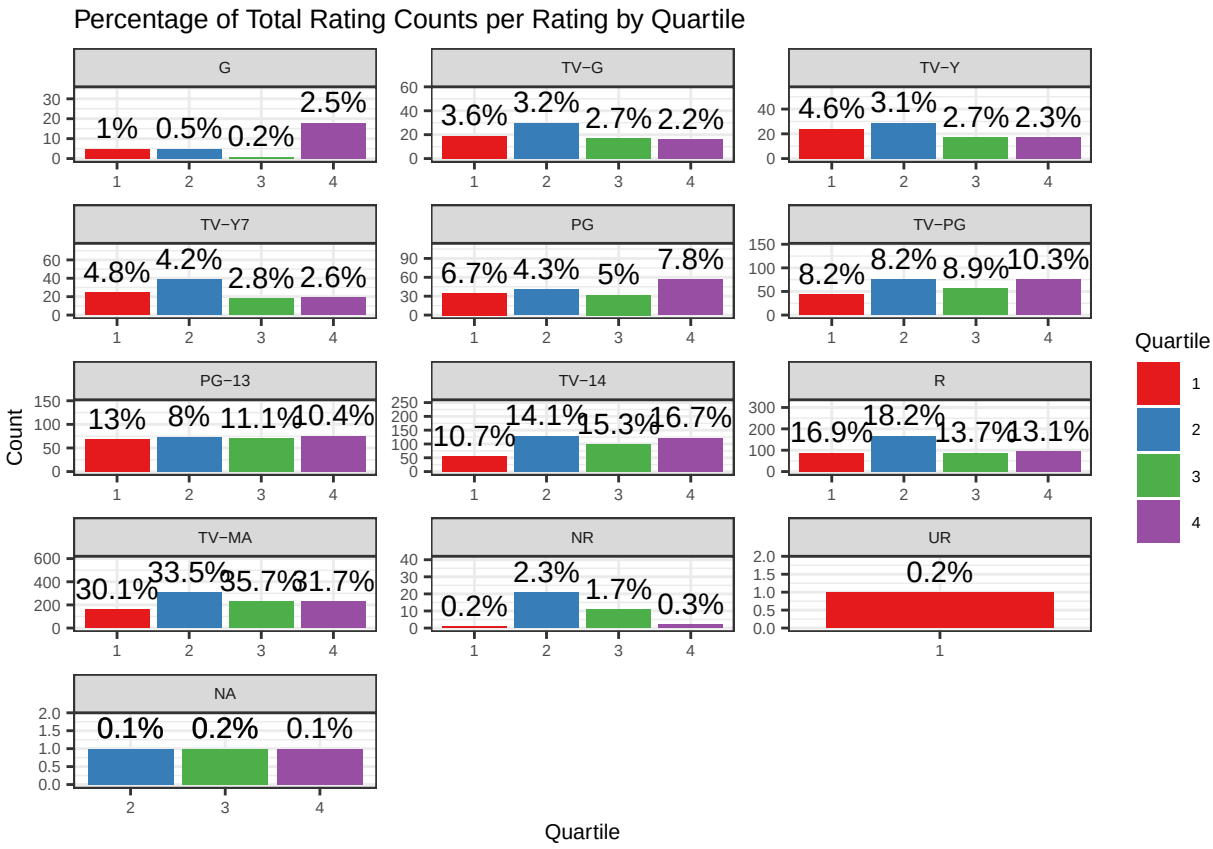



```
 combined_counts <- bind_rows(
    data.frame(Quartile = "1", rating_counts),
    data.frame(Quartile = "2", rating_counts2),
    data.frame(Quartile = "3", rating_counts3),
    data.frame(Quartile = "4", rating_counts4)
 )

combined_plot <- combined_counts %>%
    mutate(rating = factor(rating,
                           levels = c("G","TV-G","TV-Y",
                                      "TV-Y7","PG","TV-PG",
                                      "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
```

```
  ggplot(aes(x = Quartile, y = n, fill = Quartile)) +
  geom_col(position = "dodge") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
            position = position_dodge(width = 0.9),
            vjust = -0.5) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Percentage of Total Rating Counts per Rating by Quartile",
       x = "Quartile", y = "Count") +
  facet_wrap(~ rating, scales = "free", ncol = 3) +
  theme_bw() +
  scale_y_continuous(expand = expansion(mult = c(0.1, 1.0))) +
  theme(text = element_text(size = 8))

combined_plot
```



Results: scatterplot of the main relationship of interest and output for the main regression

To conclude our exploratory analysis we split all the countries in our dataset into gdp quartiles to see the distribution of television and movie parental ratings. Ranging from TV-7 to TV-MA for television ratings and General Audiences to Restricted for movies. In our exploration of these distributions there were significant takeaways that could be seen by the distribution. Fourth quartile countries had a significantly higher percentage of movies rated G and PG than any other quartile country while their R rated movie percentage was 13.1% which was the lowest percentage of any quartile. This trend proposes that wealthy

countries, whose quality of life is higher, make more family friendly films as their audiences prefer happy go-lucky films as it may be more relatable content for their citizens. This higher proportion of lighter cinema may also be because families in the wealthiest nations in the world can afford to buy movie tickets for their entire family, a purchase that could be significantly less probable for a family less fortunate. This key distinction may be a reason that production companies in lower GDP countries make fewer G and PG rated films as they believe that there won't be the same box office dollars compared to more adult themed films. As countries in the first and second quartile of GDP produce more R rated films due to this disparity in potential box office dollars.

In contrast, the ratings for television shows of TV-G, TV-Y, and TV-Y7 in fourth quartile countries is the smallest percentage of all the GDP quartiles. This contrast in kid friendly content when it comes to television is possibly due to TV network executives theorizing that wealthier countries have a higher education rate meaning children are less likely to be watching television on a day to day basis and countries that have a lower GDP may have a larger amount of kids at home during the day watching TV. While there is a contrast from the movie ratings and television ratings aimed at children once the TV ratings become aimed at teenagers the fourth quartile countries again have the higher percentage as TV-PG and TV-14 proportions increase with each increasing quartile. This is due to the fact that unike young children teenagers are much more capable of controlling the television meaning that shows directed at teens are more likely to get watched during weekends and non-school hours as the older a person gets the more they are able to flip through channels and find their show whereas young children would need some kind of assistance meaning younger audiences are less likely to be consuming media via a TV.

Conclusion: a brief conclusion summarizing the results, assesses the extent to which you found support for your hypothesis, and describes limitations of your analysis and threats to inference

Based off our findings it is safe to say that while mature films and television shows dominate the overall industry, for in both industries the most mature rating is the highest proportion of content produced, the proportion of production on content for more broad audiences is based off the lifestyles of people in their countries respective GDP quartile. Another insight that we gained was how differences in a countries wealth affect not only the ratings of a movie but what types of movies they produce. From our analysis we found that less fortunate countries like Nigeria and India have a significantly higher proportion of movies that fall in the drama category, of all the movies that Netflix has in their data set Nigeria and India have over 66% of their movies categorized as dramas compared to the United States and the United Kingdom which only have 27.6% and 16.2% of their film project categorized as dramas respectively. This massive difference is possibly due to both the UK and US falling under high income countries which have a significantly higher amount of movies produced on Netflix but could also be that countries with lower incomes produce dramas as their citizens prefer the escapism that happens when watching dramas. This trend also happens when looking at the proportion of comedies produced as again Nigeria and India have much higher proportions than their high income counterparts. While these proportions are useful to get a grasp as to what certain countries produce the data we are working with only views movies and TV shows that are on the Netflix app so to come to conclusions about a countries cultural association with their cinema based off just what Netflix has would not paint the full picture. As their is a good chance that there may be different proportions if we were to look at all streaming service apps or an entire file of each countries movies produced all time. Yet even though our research isn't conclusive enough to make umbrella statements about a countries relationship to their cinema it is conclusive enough to say that each countries variation in genre and ratings of films produced has much to do with the income of the citizens who reside there.

Works Cited

https://www.worldatlas.com/articles/largest-film-industries-in-the-world.html https://www.kaggle.com/datasets/shivamb/netflix-shows https://www.motionpictures.org/what-we-do/driving-economic-growth/ https://www.bea.gov/data/gdp/gross-domestic-product

Appendix: R Code

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

setwd("D:/university/SOC321/proj/321MovieProject")

# load libraries:
library(tidyverse)
library(ggplot2)
library(dplyr)
library(plotly)
# load data:
data <- read.csv("netflix_titles.csv")
gdp_data <- read.csv("CountryIncomeBrakcets.csv")
gdp_data$country = gdp_data$TableName
merged_data <- data %>% left_join(gdp_data)
merged_data <- merged_data %>% rowwise() %>% mutate(Drama = ifelse(("Dramas" %in% listed_in) | ("TV Dra
merged_data <- merged_data %>% rowwise() %>% mutate(Comedy = ifelse(("Comedies" %in% listed_in) | ("TV
merged_data <- merged_data %>% rowwise() %>% mutate(Action = ifelse(("Action & Adventure" %in% listed_i
merged_data <- merged_data %>% filter(!is.na(IncomeGroup) & IncomeGroup != "")

drama_data_set <- merged_data %>% mutate(Drama1 = grepl("Dramas", listed_in))
comedy_data_set <- merged_data %>% mutate(Comedy1 = grepl("Comedies", listed_in))
action_data_set <- merged_data %>% mutate(Action1 = grepl("Action & Adventure", listed_in))

drama_plot <- ggplot(drama_data_set, mapping = aes(x = IncomeGroup, fill = Drama1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Drama Movies and TV shows by Income Group", x = "Income Group", y = "Total
ggplotly(drama_plot)

comedy_plot <- ggplot(comedy_data_set, mapping = aes(x = IncomeGroup, fill = Comedy1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Comedy Movies and TV shows by Income Group",x = "Income Group", y = "Total
ggplotly(comedy_plot)

action_plot <- ggplot(action_data_set, mapping = aes(x = IncomeGroup, fill = Action1)) +
  geom_bar() +
  theme_classic() +
  labs(title = "Proportion of Action Movies and TV shows by Income Group", x = "Income Group", y = "Tota
ggplotly(action_plot)
US_data_set <- drama_data_set %>% filter(TableName == "United States")
Nigeria_data_set <- drama_data_set %>% filter(TableName == "Nigeria")
India_data_set <- drama_data_set %>% filter(TableName == "India")
UK_data_set <- drama_data_set %>% filter(TableName == "United Kingdom")
```

```
US_drama_proportion <- table(US_data_set$Drama1)
Nigeria_drama_proportion <- table(Nigeria_data_set$Drama1)
India_drama_proportion <- table(India_data_set$Drama1)
UK_drama_proportion <- table(UK_data_set$Drama1)

(US_drama_proportion[2] / (US_drama_proportion[1] + US_drama_proportion[2]))

(Nigeria_drama_proportion[2] / (Nigeria_drama_proportion[1] + Nigeria_drama_proportion[2]))

(India_drama_proportion[2] / (India_drama_proportion[1] + India_drama_proportion[2]))

(UK_drama_proportion[2] / (UK_drama_proportion[1] + UK_drama_proportion[2]))


US_data_set1 <- comedy_data_set %>% filter(TableName == "United States")
Nigeria_data_set1 <- comedy_data_set %>% filter(TableName == "Nigeria")
India_data_set1 <- comedy_data_set %>% filter(TableName == "India")
UK_data_set1 <- comedy_data_set %>% filter(TableName == "United Kingdom")
US_comedy_proportion <- table(US_data_set1$Comedy1)
Nigeria_comedy_proportion <- table(Nigeria_data_set1$Comedy1)
India_comedy_proportion <- table(India_data_set1$Comedy1)
UK_comedy_proportion <- table(UK_data_set1$Comedy1)

(US_comedy_proportion[2] / (US_comedy_proportion[1] + US_comedy_proportion[2]))

(Nigeria_comedy_proportion[2] / (Nigeria_comedy_proportion[1] + Nigeria_comedy_proportion[2]))

(India_comedy_proportion[2] / (India_comedy_proportion[1] + India_comedy_proportion[2]))

(UK_comedy_proportion[2] / (UK_comedy_proportion[1] + UK_comedy_proportion[2]))
datax <- read_csv("netflix_titles.csv") %>%
  mutate(year = release_year) %>%
  filter(country %in% c("United States","India","United Kingdom"))

gdpchange <- read_csv("gdpchange.csv") %>%
  filter(country %in% c("United States","India","United Kingdom"))  %>%
  pivot_longer(GDP_1960:GDP_2021, names_to = "year", values_to = "gdp") %>%
  mutate(year = as.numeric(str_sub(year, 5L, -1L))) %>%
  select(-`2022`)


mergeddata <- datax %>%
  left_join(gdpchange)

usa <- mergeddata %>% filter(country == "United States")
summary(usa$gdp)

usa <- usa %>%
  mutate(GDPcategory = case_when(gdp <1.667 ~ 1,
                                  between(gdp, 1.667, 2.288)~ 2,
                                 between(gdp, 2.288, 2.945)~ 3,
                                 TRUE ~ 4 ))
lowestquartileUSA <- usa %>%
```

```
  filter(GDPcategory == 1)

rating_counts <- lowestquartileUSA %>%
  count(rating) %>%
  mutate(percentage = n / sum(n)*100)
print(rating_counts)


  rating_counts %>%
    rename(Percentage = percentage) %>%
    mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
  mutate(rating = factor(rating,
                         levels = c("G","TV-G","TV-Y",
                                    "TV-Y7","PG","TV-PG",
                                       "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
  ggplot(aes(x = Percentage, y= rating,
             fill = rating, label = pct_label)) +
  geom_col() +
    geom_text(hjust = -0.75) +
    scale_x_continuous(limits = c(0, 40))


  secondquartileUSA <- usa %>%
    filter(GDPcategory == 2)

  rating_counts2 <- secondquartileUSA %>%
    count(rating) %>%
    mutate(percentage = n / sum(n)*100)
  print(rating_counts2)


  rating_counts2 %>%
    rename(Percentage = percentage) %>%
    mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
    mutate(rating = factor(rating,
                           levels = c("G","TV-G","TV-Y",
                                      "TV-Y7","PG","TV-PG",
                                      "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
    ggplot(aes(x = Percentage, y= rating,
               fill = rating, label = pct_label)) +
    geom_col() +
    geom_text(hjust = -0.75) +
    scale_x_continuous(limits = c(0, 40))


  thirdquartileUSA <- usa %>%
    filter(GDPcategory == 3)

  rating_counts3 <- thirdquartileUSA %>%
    count(rating) %>%
    mutate(percentage = n / sum(n)*100)
  print(rating_counts3)
```

```
  rating_counts3 %>%
    rename(Percentage = percentage) %>%
    mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
    mutate(rating = factor(rating,
                           levels = c("G","TV-G","TV-Y",
                                      "TV-Y7","PG","TV-PG",
                                      "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
    ggplot(aes(x = Percentage, y= rating,
               fill = rating, label = pct_label)) +
    geom_col() +
    geom_text(hjust = -0.75) +
    scale_x_continuous(limits = c(0, 40))


  fourthquartileUSA <- usa %>%
    filter(GDPcategory == 4)

  rating_counts4 <- fourthquartileUSA %>%
    count(rating) %>%
    mutate(percentage = n / sum(n)*100)
  print(rating_counts4)


  rating_counts4 %>%
    rename(Percentage = percentage) %>%
    mutate(pct_label = paste0(round(Percentage, 1), "%")) %>%
    mutate(rating = factor(rating,
                           levels = c("G","TV-G","TV-Y",
                                      "TV-Y7","PG","TV-PG",
                                      "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
    ggplot(aes(x = Percentage, y= rating,
               fill = rating, label = pct_label)) +
    geom_col() +
    geom_text(hjust = -0.75) +
    scale_x_continuous(limits = c(0, 40))

 combined_counts <- bind_rows(
    data.frame(Quartile = "1", rating_counts),
    data.frame(Quartile = "2", rating_counts2),
    data.frame(Quartile = "3", rating_counts3),
    data.frame(Quartile = "4", rating_counts4)
  )

combined_plot <- combined_counts %>%
    mutate(rating = factor(rating,
                           levels = c("G","TV-G","TV-Y",
                                      "TV-Y7","PG","TV-PG",
                                      "PG-13","TV-14", "R","TV-MA", "NR","UR"))) %>%
    ggplot(aes(x = Quartile, y = n, fill = Quartile)) +
    geom_col(position = "dodge") +
    geom_text(aes(label = paste0(round(percentage, 1), "%")),
              position = position_dodge(width = 0.9),
              vjust = -0.5) +
```

```
    scale_fill_brewer(palette = "Set1") +
    labs(title = "Percentage of Total Rating Counts per Rating by Quartile",
         x = "Quartile", y = "Count") +
    facet_wrap(~ rating, scales = "free", ncol = 3) +
    theme_bw() +
    scale_y_continuous(expand = expansion(mult = c(0.1, 1.0))) +
    theme(text = element_text(size = 8))

  combined_plot
# Notice, eco=TRUE and eval=FALSE
```