```r
# Load the necessary packages
library(dplyr)
library(tidyr)
library(kableExtra)
library(stringr)
library(ggplot2)
library(scales)
```

```r
# Load the Disney+ & Netflix data from the provided URL
disney_url <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vT2CVS1o_R5Dq-ATuNlVRInKOWHG9(

disney_data <- read.csv(disney_url)

netflix_url <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vRhYpV_lEupTdc9VgHeH2814Qtwa(

netflix_data <- read.csv(netflix_url)
```

```r
# Add a platform column to the Disney+ dataset
disney_plus <- disney_data %>%
  mutate(platform = "Disney Plus")
# Add a platform column to the Netflix dataset
netflix <- netflix_data %>%
  mutate(platform = "Netflix")
# Join the two datasets
moviesandtv <- bind_rows(disney_plus, netflix)
```

```r
# Add N/A values to replace missing data
moviesandtv <- moviesandtv %>%
  mutate(across(where(is.character), ~na_if(., "")))
```

```r
#Separate the moviesandtv table into two distinct tables

movies <- moviesandtv %>%
  filter(type == 'Movie') %>% # Filter the type column
  mutate(duration = as.integer(str_remove(duration, " min"))) # Remove the string 'min' and c

tvshows <- moviesandtv %>%
  filter(type == 'TV Show') %>% # Filter the type column
  mutate(duration = str_replace(duration, "Seasons", "Season")) %>% # Replace the Seasons to
  mutate(duration = as.integer(str_remove(duration, " Season"))) # Remove the string 'Season
```

```r
### How many TV shows and movies from a specific year are available on Disney+ and Netflix ?

# For TV Shows
# Step 1: Wrangle the TV shows dataset
# Filter the dataset to include only TV shows released between 1950 and 2025
tv_trend <- tvshows %>%
  filter(release_year >= 1990 & release_year <= 2025) %>%
  group_by(platform, release_year) %>% # Group by platform and release year
  summarise(count = n()) # Count the number of TV shows

# Step 2: Create the trend line plot
ggplot(tv_trend, aes(x = release_year, y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + # Create bar plot with grouped bars
  labs( title = "Number of TV Shows on Netflix vs. Disney+ From a Particular Year",
    x = "Release Year",
    y = "Number of TV Shows",
    fill = "Streaming Platforms")+ # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5, size = 12), #center the title and change the fron
    legend.position = "bottom") # Positions the legend at the bottom of the plot
```

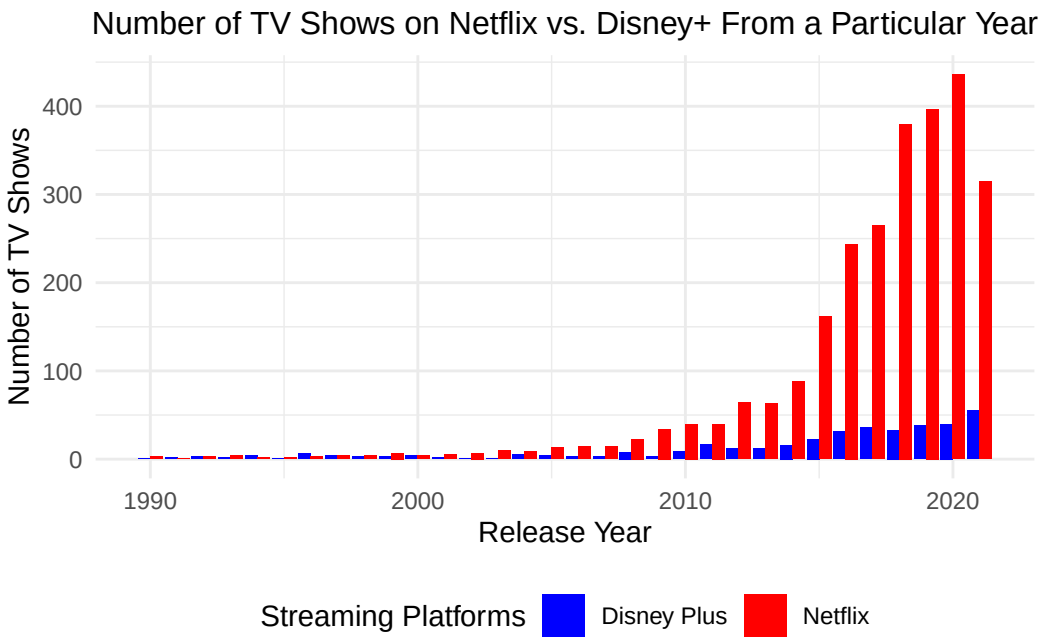


Figure 1: Comparison of TV Shows available on Disney+ and Netflix

```
# For Movies
# Step 1: Wrangle the Movies dataset
# Filter the dataset to include only Movies released between 1950 and 2025
movie_trend <- movies %>%
  filter(release_year >= 1990 & release_year <= 2025) %>%
  group_by(platform, release_year) %>% # Group by platform and release year
  summarise(count = n()) # Count the number of movies

# Step 2: Create the bar graph plot
ggplot(movie_trend, aes(x = release_year, y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + # Create bar plot with grouped bars
  labs( title = "Number of Movies on Netflix vs. Disney+ From a Particular Year",
    x = "Release Year",
    y = "Number of Movies",
    fill = "Streaming Platforms") + # Add titles, labels, and key for the plot
  scale_fill_manual( values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal()+ # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5, size = 12), #center the title and change the front
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
```

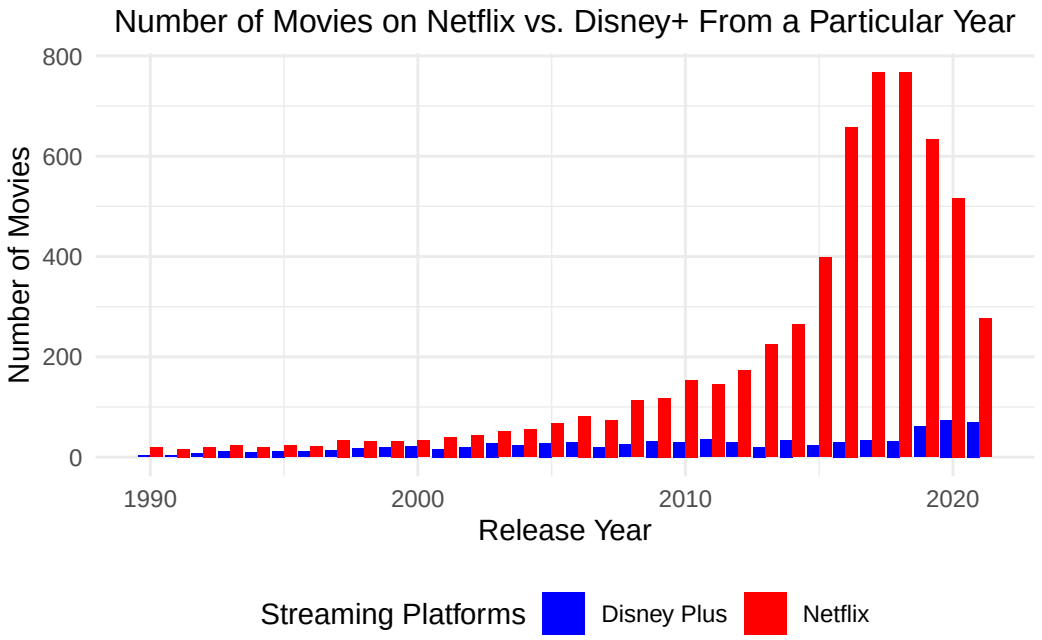


Figure 2: Comparison of Movies available on Disney+ and Netflix

```
### What is the average duration of movies and TV shows on Disney+ and Netflix?
# For TV Shows
# Step 1: Wrangle the TV Shows dataset
average_duration_tv <- tvshows %>%
  group_by(platform) %>% # Group by platform
  summarise(avg_seasons = mean(duration, na.rm = TRUE)) # Calculate mean number of seasons

# Step 2: Plot the bar graph plot
ggplot(average_duration_tv, aes(x = platform, y = avg_seasons, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # Create a bar plot
  labs(title = "Average Seasons of TV Shows on Netflix vs. Disney+",
    x = "Platform",
    y = "Average Number of Seasons" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
```

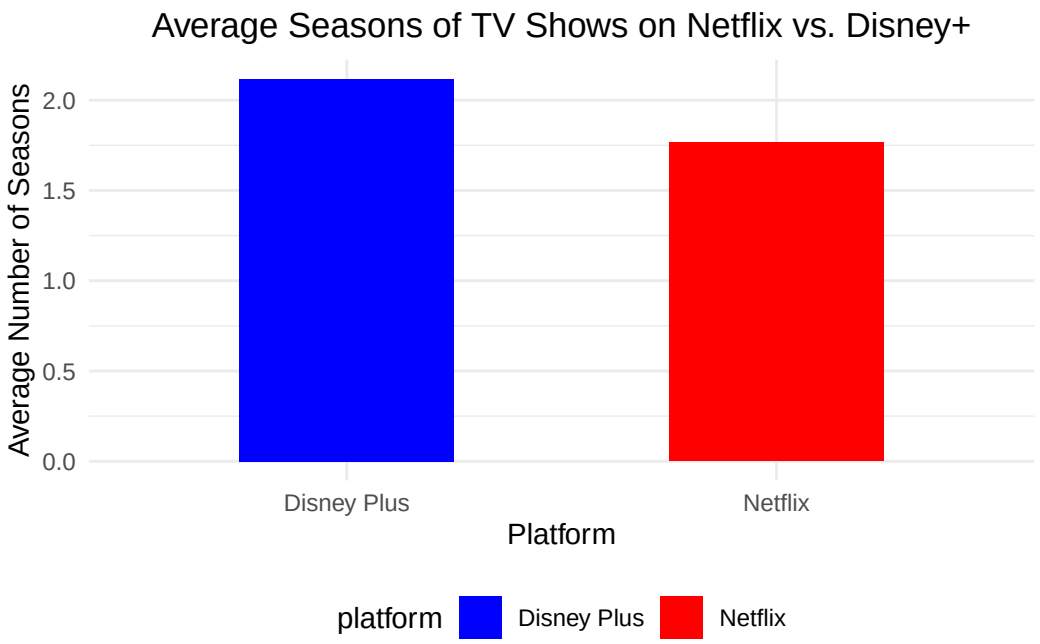


Figure 3: Comparison of average duration of TV Shows on Disney+ and Netflix

```r
# For Movies
# Step 1: Wrangle the Movies dataset
average_duration_movies <- movies %>%
  group_by(platform) %>% # Group by platform
  summarise(avg_time = mean(duration, na.rm = TRUE)) #Calculate mean number of time

# Step 2: Plot the bar plot
ggplot(average_duration_movies, aes(x = platform, y = avg_time, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # Create a bar plot
  labs( title = "Average Time of Movies on Netflix vs. Disney+",
    x = "Platform",
    y = "Average Time (mins) of Movies"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
```

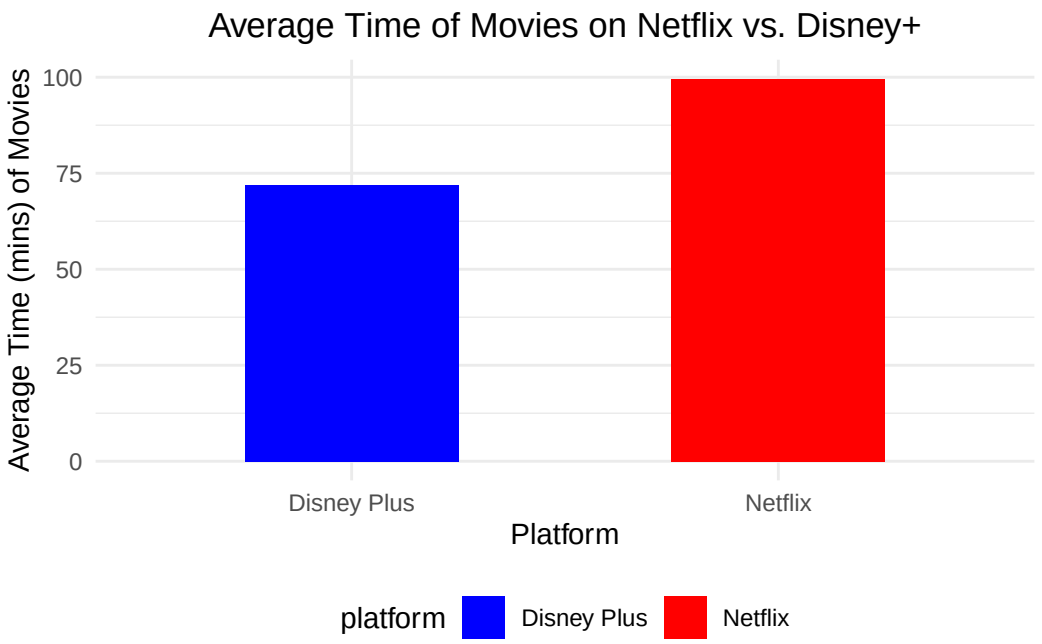


Figure 4: Comparison of average duration of Movies available on Disney+ and Netflix

```
### Which directors are most frequently featured on Disney+ and Netflix?
# For TV shows
# Step 1 : Wrangle the TV shows dataset
director_data_tv <- tvshows %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  separate_rows(director, sep = ",") %>% # Separate rows with a separator
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of rows

# Step 2: Get the top directors for each platform
top_directors_tv <- director_data_tv %>%
  group_by(platform) %>% # Group by platform
  top_n(10, count) %>% # Select top 10 directors by count
  ungroup() # Ungroup the data

# Step 3: Plot the bar data
ggplot(top_directors_tv, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity") + # Create a bar plot
  coord_flip() + # Flipping coordinates for better readability
  labs(title = "Top TV Shows Directors on Disney+ and Netflix",
    x = "Director",
```

```
    y = "Number of TV Shows",
    fill = "Streaming Platform" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal()+ # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
```

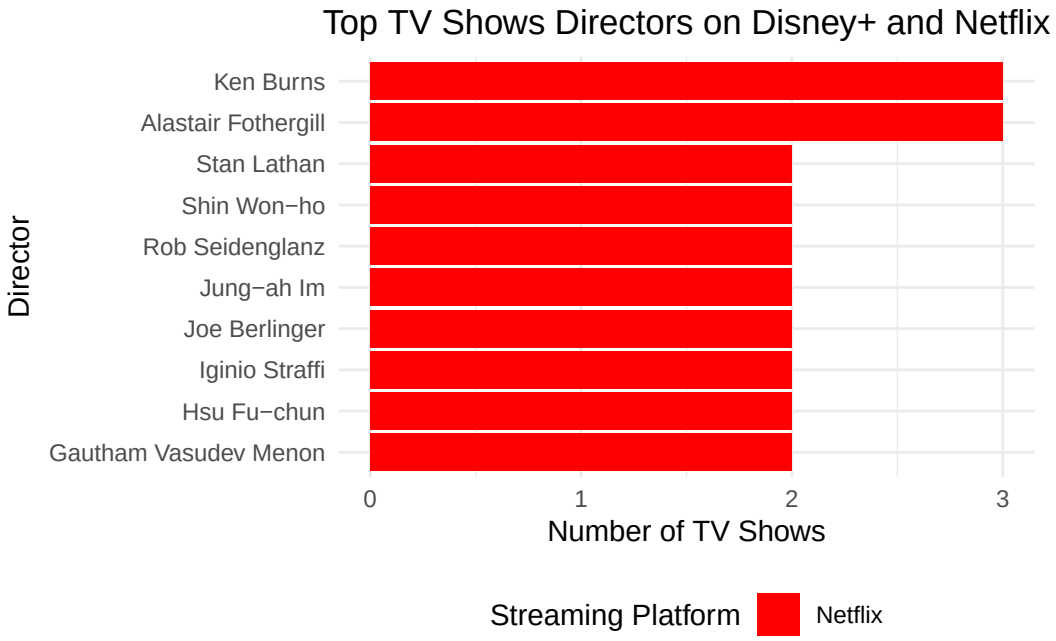


Figure 5: Comparison of top 10 TV show directors available on Disney+ and Netflix

```
# For movies
# Step 1 : Wrangle the Movies dataset
director_data_movies <- movies %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of rows

# Step 2: Get the top directors for each platform
top_directors <- director_data_movies %>%
  group_by(platform) %>% # Group by platform
  top_n(10, count) %>% # Select top 10 directors by count
  ungroup() # Ungroup the data
```

```r
# Step 3: Plot the data
ggplot(top_directors, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity", show.legend = FALSE) + # Bar plot without legend
  coord_flip() + # Flip coordinates for better readability
  labs(title = "Top Movie Directors on Disney+ vs. Netflix",
    x = "Director",
    y = "Number of Movies",
    fill = " Streaming Platform" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red") ) + # Set custom co
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom") # Positions the legend at the bottom of the plot
```

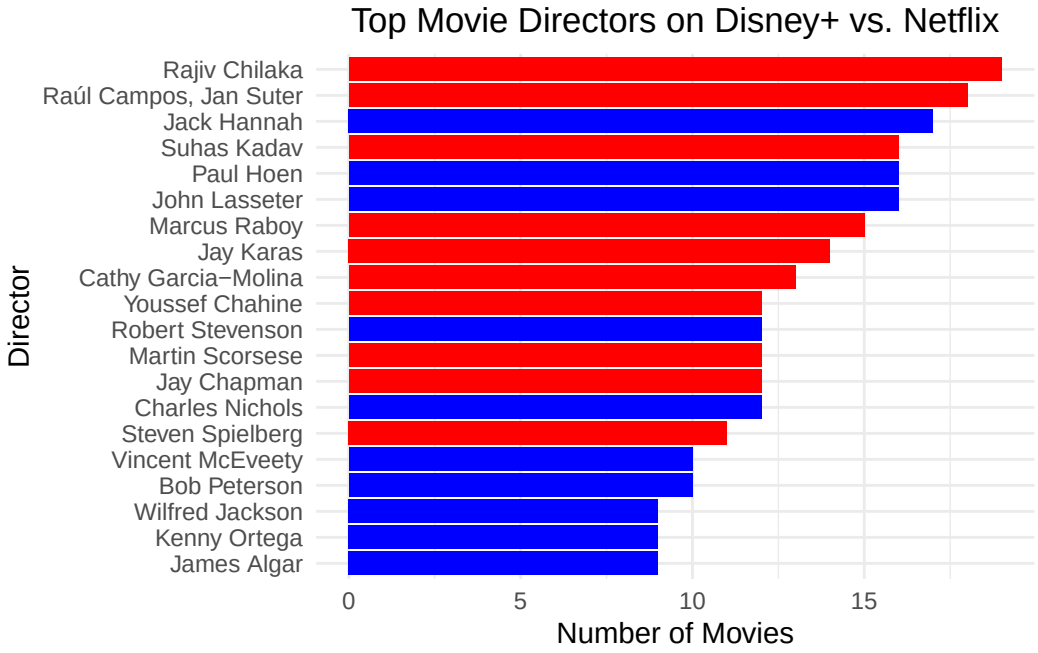


Figure 6: Comparison of top 10 Movie directors available on Disney+ and Netflix

```r
# What is the regional distribution of Disney+ and Netflix  content (based on country/region

#Step 1: Wrangle the data to select Disney+ and sort by country of production
disney_distribution <- moviesandtv %>%
  filter(platform == "Disney Plus") %>% #filter only Disney Plus entries
```

```r
  separate_rows(country, sep = ", ") %>% #Separate the list of country entries by the comma a
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country) & country != "United States") %>% #remove NA values and not include
  slice_head(n = 15) #Select only the top 15 production countries

#Step 2: Wrangle the data to select Netflix and sort by country of production
netflix_distribution <- moviesandtv %>%
  filter(platform == "Netflix") %>% #filter only Netflix entries
  separate_rows(country, sep = ", ") %>% #Separate the list of country entries by the comma a
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country) & country != "United States") %>% #remove NA values and not include
  slice_head(n = 15) #Select only the top 15 production countries

#Step 3: Create a bar graph displaying the regional distribution for Disney+
ggplot(disney_distribution, aes(x = reorder(country, count), y = count, fill = "Disney Plus")
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE) + #creates a bar graph
  labs( #provides title and axis labels
    title = "Regional Distribution of Content on Disney+",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  scale_fill_manual(values = c("blue")) + #display the data in blue for Disney+
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1) #adjust the y axis content t
  )
```

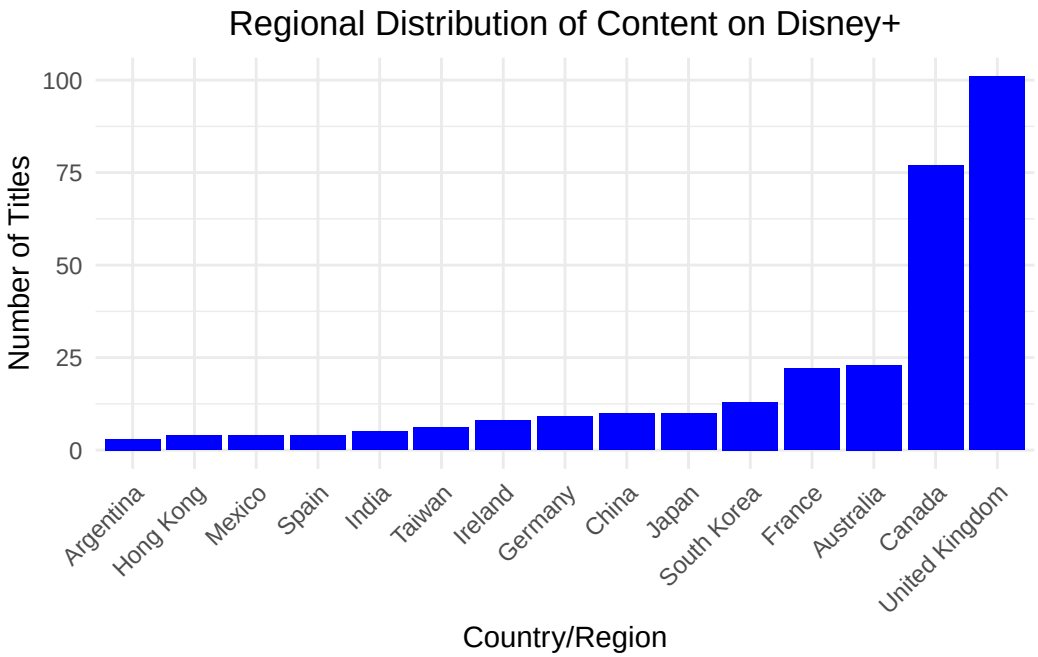


Figure 7: Regional Distribution of Disney+ and Netflix Content

```
#Step 3: Create a bar graph displaying the regional distribution for Netflix
ggplot(netflix_distribution, aes(x = reorder(country, count), y = count, fill = "Netflix")) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE) + #creates a bar graph
  labs( #provides title and axis labels
    title = "Regional Distribution of Content on Netflix",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  scale_fill_manual(values = c("red")) + #display the data in red for Netflix
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1) #adjust the y axis content t
  )
```

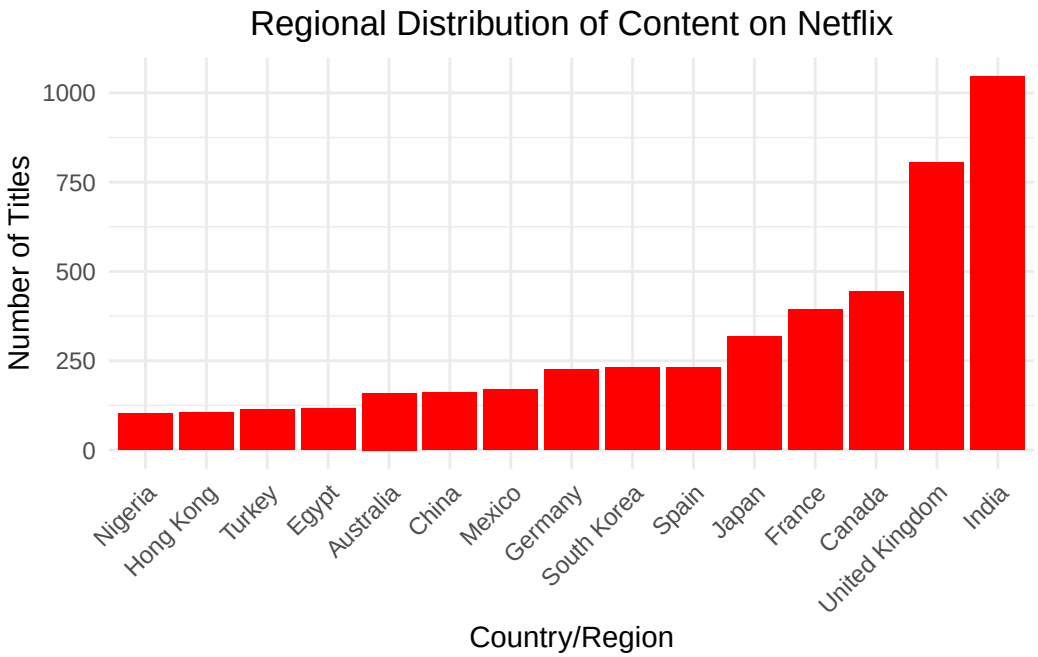


Figure 8: Regional Distribution of Disney+ and Netflix Content

```
# Is there a trend in content ratings? For instance, does Disney Plus predominantly feature r

#Step 1: Wrangle the moviesandtv dataset to analyze the ratings with their count
ratings <- moviesandtv %>%
  filter(!grepl("min", rating))%>% #removed random values in ratings containing "min" entries
  group_by(platform, rating) %>% #group the data to display both Netflix and Disney+ and ther
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count))

#Step 2: Create a bar plot displaying ratings and their count for both Netflix and Disney+
ggplot(ratings, aes(x = reorder(rating, count), y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + #creates a bar graph
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + #fills the bars
  labs( #provides title, axis labels, and legend
    title = "Content Ratings on Disney Plus and Netflix",
    x = "Rating",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  theme_minimal() +
  theme(
```

```
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1), #adjust the y axis content
    legend.position = "bottom" #position the legend on the bottom of the plot
  )
```

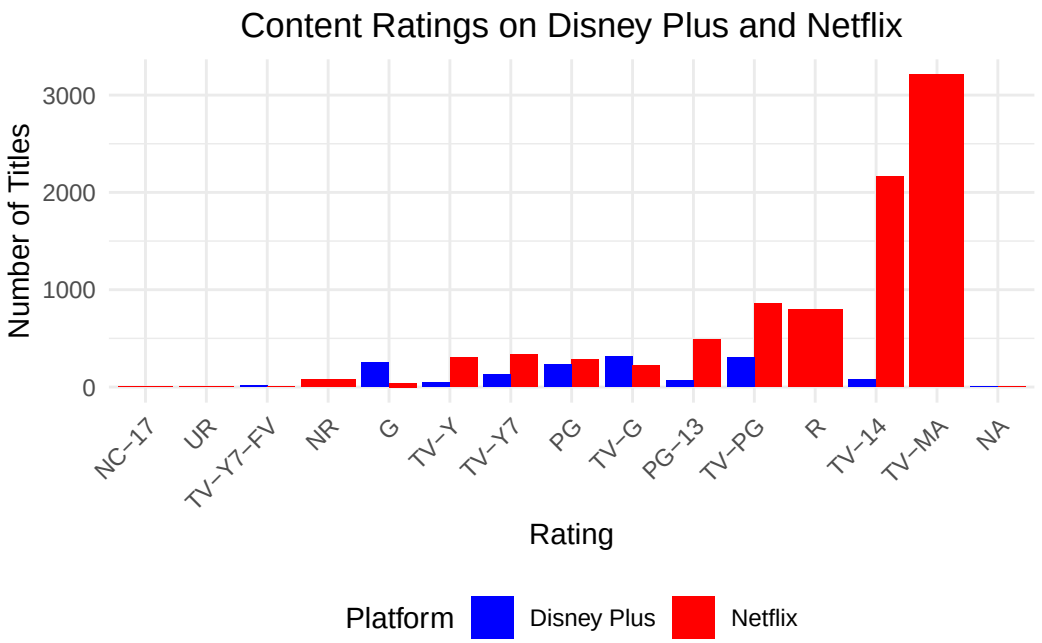


Figure 9: Comparison of Content Ratings of Disney+ and Netflix

```
# How does the time between a movie's release and its addition to Disney+ and Netflix vary?

#Step 1: Wrangle the dataset so that the date_added is a integer of the year and not a string
moviesandtv_datechange <- moviesandtv %>%
  filter(!is.na(date_added), !is.na(release_year), release_year > 2000) %>% #filters out any
  mutate(added_year = as.integer(str_trim(str_sub(date_added, -4)))) #mutates the table to di

#Step 2: Create a plot displaying the correlation between the year releasted and year added t
ggplot(moviesandtv_datechange, aes(x = release_year, y = added_year, color = platform, size =
  geom_point() + #Creates a scatter plot with both Disney+ and Netflix data
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "grey") + #This is refer
  scale_color_manual(values = c("Netflix" = "red", "Disney Plus" = "blue")) + #setting Netfli
  scale_size_manual(values = c("Netflix" = 1, "Disney Plus" = 2)) +  # Custom size for each p
  labs( #provides title, axis labels, and legend
    title = "Release Year vs. Year Added for Movies and TV Shows",
```

```
  x = "Release Year",
  y = "Year Added",
  color = "Platform"
) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5),  # Center the title
  legend.position = "bottom" #position the legend on the bottom of the plot
)+
guides(size = "none") #removes the second legend displaying size as a key
```

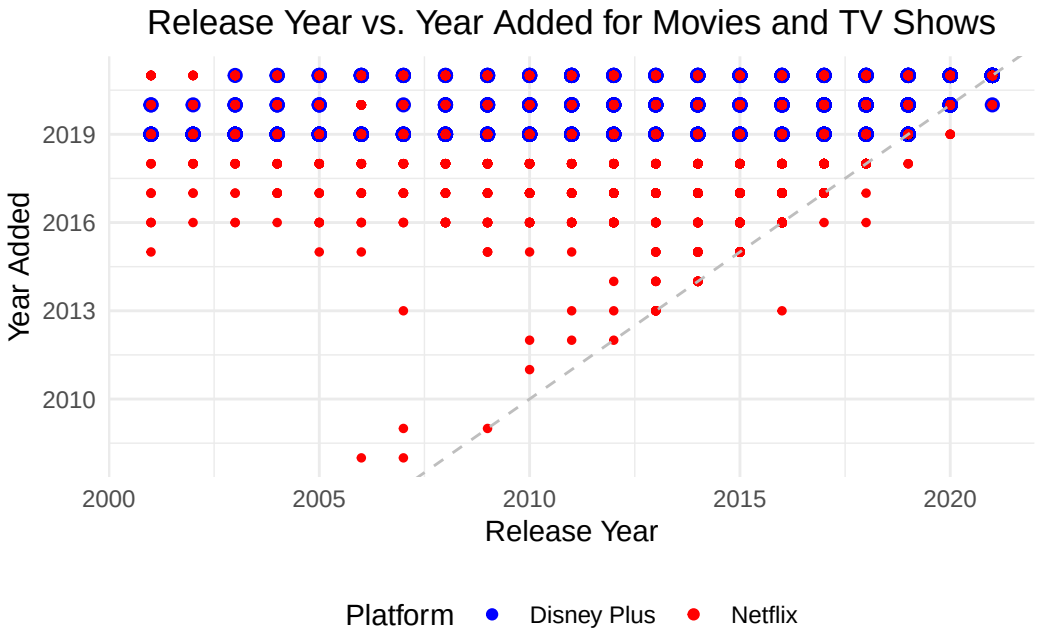


Figure 10: Comparison of Release Year to Year Added to Netflix and Disney+