# Disney+ and Netflix Movies and TV Shows Analysis

Jacqueline Maloney & Aruushi Naik

## Trends and Patterns in DisneyPlus and Netflix Movies and TV Shows: A Comparative Analysis

**Netflix and Disney+**

Image Credits - https://static0.gamerantimages.com/wordpress/wp-content/uploads/2022/08/Netflix-vs-Disney-Plus.jpg

### Introduction

Disney+ and Netflix are two of the most well-known streaming services in the world, providing millions of users with access to a vast variety of TV series and films. We conduct an exploratory data analysis (EDA) of the films and television series that are available on Netflix and Disney+ in this analysis. Given the dynamic nature of digital entertainment, knowing

how content trends and patterns have changed over time offers important insights regarding platform expansion, production tactics, and consumer preferences.

Key patterns in Disney+ and Netflix content will be examined in this analysis, such as variations in the number of movies and TV shows over time, average running times, regional production distribution, content ratings, etc.Disney+ and Netflix's future content choices and marketing plans will be influenced by the insights this analysis provides into consumer preferences, strategic planning for content, and platform evolution.

**The broader research question for our analysis is :**

**What trends and patterns can be observed in Disney Plus and Netflix TV shows and movies?**

## Dataset Overview

To address the research question, we will conduct an in-depth analysis of the Disney+ and Netflix data sets, examining various aspects of the content offered on both platforms.

Disney Data Preview

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|
| s1 | Movie | Duck the Hal... | Alonso Ramir... | Chris Dia-man... | | November 26,... | 2016 | TV-G | 23 min | Animation, F... | Join Mickey ... |
| s2 | Movie | Ernest Saves... | John Cherry | Jim Varney, ... | | November 26,... | 1988 | PG | 91 min | Comedy | Santa Claus ... |
| s3 | Movie | Ice Age: A M... | Karen Disher | Raymond Albe... | United States | November 26,... | 2011 | TV-G | 23 min | Animation, C... | Sid the Slot... |
| s4 | Movie | The Queen Fa... | Hamish Hamil-ton | Darren Criss... | | November 26,... | 2021 | TV-PG | 41 min | Musical | This is real... |
| s5 | TV Show | The Beat-les:... | | John Lennon,... | | November 25,... | 2021 | | 1 Season | Docuseries, ... | A three-part... |
| s6 | Movie | Becoming Cou... | Liz Garbus | Jacques Yves... | United States | November 24,... | 2021 | PG-13 | 94 min | Biographical | An inside lo... |
| s7 | TV Show | Hawkeye | | Jeremy Renne... | | November 24,... | 2021 | TV-14 | 1 Season | Action-Adven... | Clint Bar-ton... |
| s8 | TV Show | Port Pro-tect... | | Gary Muehlbe... | United States | November 24,... | 2015 | TV-14 | 2 Seasons | Docuseries, ... | Residents of... |
| s9 | TV Show | Secrets of t... | | Dr. Ray Ball... | United States | November 24,... | 2019 | TV-PG | 2 Seasons | Animals & Na... | A day in the... |
| s10 | Movie | A Muppets Ch... | Kirk R. That... | Steve Whitmi... | United States | November 19,... | 2008 | G | 45 min | Comedy, Fami... | Celebrate th... |

Netflix Data Preview

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| s1 | Movie | Dick Johnson... | Kirsten Johnson | | United States | September 25... | 2020 | PG-13 | 90 min | Documentaries | As her fathe... |
| s2 | TV Show | Blood & Water | | Ama Qamata, ... | South Africa | September 24... | 2021 | TV-MA | 2 Seasons | Internationa... | After crossi... |
| s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajil... | | September 24... | 2021 | TV-MA | 1 Season | Crime TV Sho... | To protect h... |
| s4 | TV Show | Jailbirds Ne... | | | | September 24... | 2021 | TV-MA | 1 Season | Docuseries, ... | Feuds, flirt... |
| s5 | TV Show | Kota Factory | | Mayur More, ... | India | September 24... | 2021 | TV-MA | 2 Seasons | Internationa... | In a city of... |
| s6 | TV Show | Midnight Mass | Mike Flanagan | Kate Siegel,... | | September 24... | 2021 | TV-MA | 1 Season | TV Dramas, T... | The arrival ... |
| s7 | Movie | My Little Po... | Robert Culle... | Vanessa Hudg... | | September 24... | 2021 | PG | 91 min | Children & F... | Equestria's ... |
| s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba... | United State... | September 24... | 1993 | TV-MA | 125 min | Dramas, Inde... | On a photo s... |
| s9 | TV Show | The Great Br... | Andy Devonshire | Mel Giedroyc... | United Kingdom | September 24... | 2021 | TV-14 | 9 Seasons | British TV S... | A talented b... |
| s10 | Movie | The Starling | Theodore Melfi | Melissa McCa... | United States | September 24... | 2021 | PG-13 | 104 min | Comedies, Dr... | A woman adju... |

The tables above were built using the head() method and display the first few values from the data sets. It provides a quick preview of the data set's structure, column names, and value types, which is important for understanding the data's overall format and content.

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Length:1450 | Length:1450 | Length:1450 | Length:1450 | Length:1450 | Length:1450 | Length:1450 | Min. :1928 | Length:1450 | Length:1450 | Length:1450 | Length:1450 |
| Class :character | Class :character | Class :character | Class :character | Class :character | Class :character | Class :character | 1st Qu.:1999 | Class :character | Class :character | Class :character | Class :character |
| Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Median :2011 | Mode :character | Mode :character | Mode :character | Mode :character |
| NA | NA | NA | NA | NA | NA | NA | Mean :2003 | NA | NA | NA | NA |
| NA | NA | NA | NA | NA | NA | NA | 3rd Qu.:2018 | NA | NA | NA | NA |
| NA | NA | NA | NA | NA | NA | NA | Max. :2021 | NA | NA | NA | NA |

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Length:8807 | Length:8807 | Length:8807 | Length:8807 | Length:8807 | Length:8807 | Length:8807 | Min. :1925 | Length:8807 | Length:8807 | Length:8807 | Length:8807 |
| Class :character | Class :character | Class :character | Class :character | Class :character | Class :character | Class :character | 1st Qu.:2013 | Class :character | Class :character | Class :character | Class :character |
| Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Mode :character | Median :2017 | Mode :character | Mode :character | Mode :character | Mode :character |
| NA | NA | NA | NA | NA | NA | NA | Mean :2014 | NA | NA | NA | NA |
| NA | NA | NA | NA | NA | NA | NA | 3rd Qu.:2019 | NA | NA | NA | NA |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NA | NA | NA | NA | NA | NA | NA | Max. :2021 | NA | NA | NA | NA | |

Additionally, we used the summary() function to create a summary table. Each column in the data data frame is summarized in the table that is shown above.

### Data Cleaning and Wrangling

To streamline the analysis, the datasets for Disney+ and Netflix were merged into a single dataset. This joining process ensures a consistent structure for comparative analysis and enables efficient exploration of trends across both the platforms.

```
# 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)
```

Upon looking at the *moviesandtv* dataset, we came across empty cells and decided to replace it with N/A values to account for missing data. This would maintain uniformity throughout the dataset and allow for more accurate analysis by clearly identifying missing information across the Disney+ and Netflix datasets.

To facilitate targeted analysis for each type of content, we separated the *moviesandtv* dataset into two distinct tables: one for movies and one for TV shows. For movies, we changed the format for duration by removing the "min" label, while for TV shows, we converted the number of seasons into integers to ensure clear and specific analysis of content.

TV Shows Data

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | platform |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| s5 | TV Show | The Beatles:... | NA | John Lennon,... | NA | November 25,... | 2021 | NA | 1 | Docuseries, ... | A three-part... | Disney Plus |
| s7 | TV Show | Hawkeye | NA | Jeremy Renne... | NA | November 24,... | 2021 | TV-14 | 1 | Action-Adven... | Clint Barton... | Disney Plus |
| s8 | TV Show | Port Protect... | NA | Gary Muehlbe... | United States | November 24,... | 2015 | TV-14 | 2 | Docuseries, ... | Residents of... | Disney Plus |
| s9 | TV Show | Secrets of t... | NA | Dr. Ray Ball... | United States | November 24,... | 2019 | TV-PG | 2 | Animals & Na... | A day in the... | Disney Plus |
| s14 | TV Show | Dr. Oakley, ... | NA | Dr. Michelle... | United States | November 17,... | 2013 | TV-PG | 10 | Action-Adven... | Meet Dr. Mic... | Disney Plus |
| s18 | TV Show | Disney Fancy... | NA | Mia Jenness,... | United State... | November 12,... | 2018 | TV-PG | 3 | Animation, Kids | Nancy makes ... | Disney Plus |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| s19 | TV Show | Disney Inter... | NA | Carolina Dom... | NA | November 12,... | 2021 | TV-PG | 1 | Comedy, Comi... | Allegra is r... | Disney Plus |
| s29 | TV Show | Olaf Presents | NA | Josh Gad | NA | November 12,... | 2021 | TV-PG | 1 | Animation, F... | Olaf goes fr... | Disney Plus |
| s52 | TV Show | Disney Amphibia | NA | Justin Felbi... | United State... | November 3, ... | 2018 | TV-Y7 | 3 | Animation, C... | Anne Boonchu... | Disney Plus |
| s53 | TV Show | Photo Ark | NA | Joel Sartore | United States | November 3, ... | 2017 | TV-PG | 1 | Animals & Na... | National Geo... | Disney Plus |

Movies Data

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | platform |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| s1 | Movie | Duck the Hal... | Alonso Ramir... | Chris Diaman... | NA | November 26,... | 2016 | TV-G | 23 | Animation, F... | Join Mickey ... | Disney Plus |
| s2 | Movie | Ernest Saves... | John Cherry | Jim Varney, ... | NA | November 26,... | 1988 | PG | 91 | Comedy | Santa Claus ... | Disney Plus |
| s3 | Movie | Ice Age: A M... | Karen Disher | Raymond Albe... | United States | November 26,... | 2011 | TV-G | 23 | Animation, C... | Sid the Slot... | Disney Plus |
| s4 | Movie | The Queen Fa... | Hamish Hamilton | Darren Criss... | NA | November 26,... | 2021 | TV-PG | 41 | Musical | This is real... | Disney Plus |
| s6 | Movie | Becoming Cou... | Liz Garbus | Jacques Yves... | United States | November 24,... | 2021 | PG-13 | 94 | Biographical | An inside lo... | Disney Plus |
| s10 | Movie | A Muppets Ch... | Kirk R. That... | Steve Whitmi... | United States | November 19,... | 2008 | G | 45 | Comedy, Fami... | Celebrate th... | Disney Plus |
| s11 | Movie | Adventure Th... | John Gleim | Don Hahn, Ka... | NA | November 19,... | 2020 | TV-PG | 59 | Documentary | Explore the ... | Disney Plus |
| s12 | Movie | Puppy for Ha... | NA | NA | NA | November 19,... | 2020 | TV-G | 4 | Comedy, Fami... | Check out Da... | Disney Plus |
| s13 | Movie | The Pixar Story | Leslie Iwerks | Stacy Keach,... | United States | November 19,... | 2007 | G | 91 | Documentary | A groundbrea... | Disney Plus |
| s15 | Movie | America the ... | NA | Michael B. J... | NA | November 12,... | 2021 | TV-PG | 2 | Animals & Na... | Epic, grand ... | Disney Plus |

## Exploratory Data Analysis(EDA)

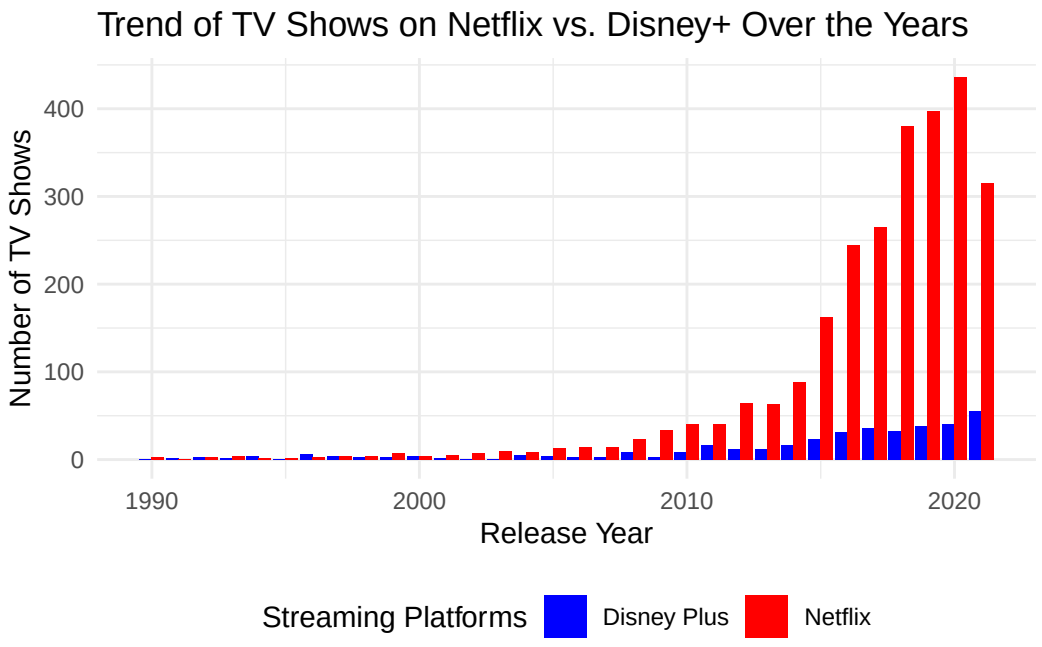
### How many TV shows and movies from a specific year are available on Netflix and Disney+ ?

### TV Shows

```
# 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
```

```
`summarise()` has grouped output by 'platform'. You can override using the
`.groups` argument.
```

```r
# 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 = "Trend of TV Shows on Netflix vs. Disney+ Over the Years",
    x = "Release Year",
    y = "Number of TV Shows",
    fill = "Streaming Platforms"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) +
  theme_minimal() + # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
```



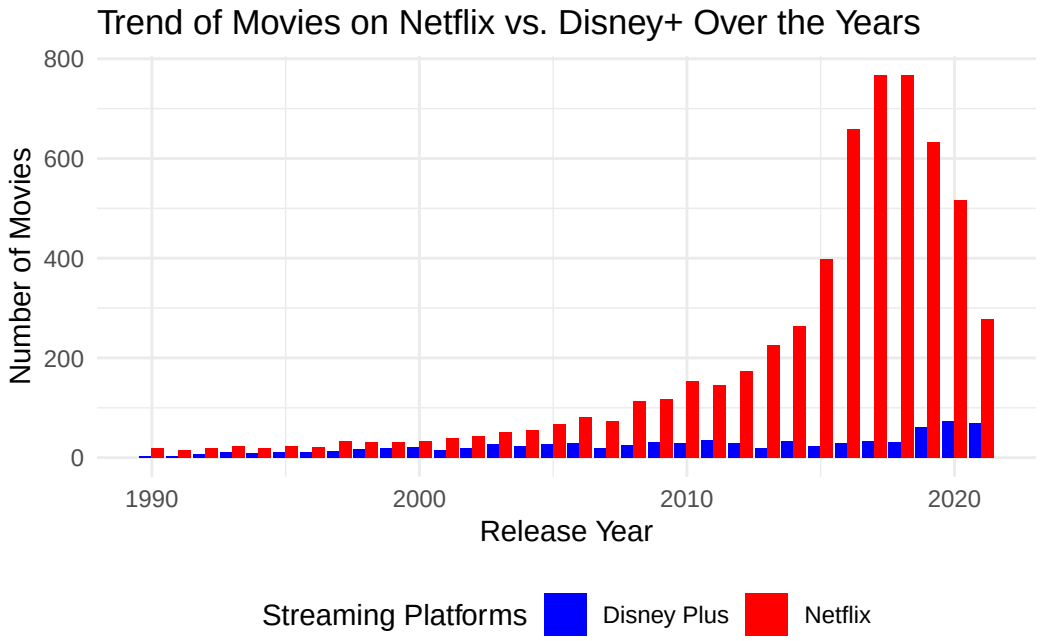

**Movies**

```r
# 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
```

```
`summarise()` has grouped output by 'platform'. You can override using the
`.groups` argument.
```

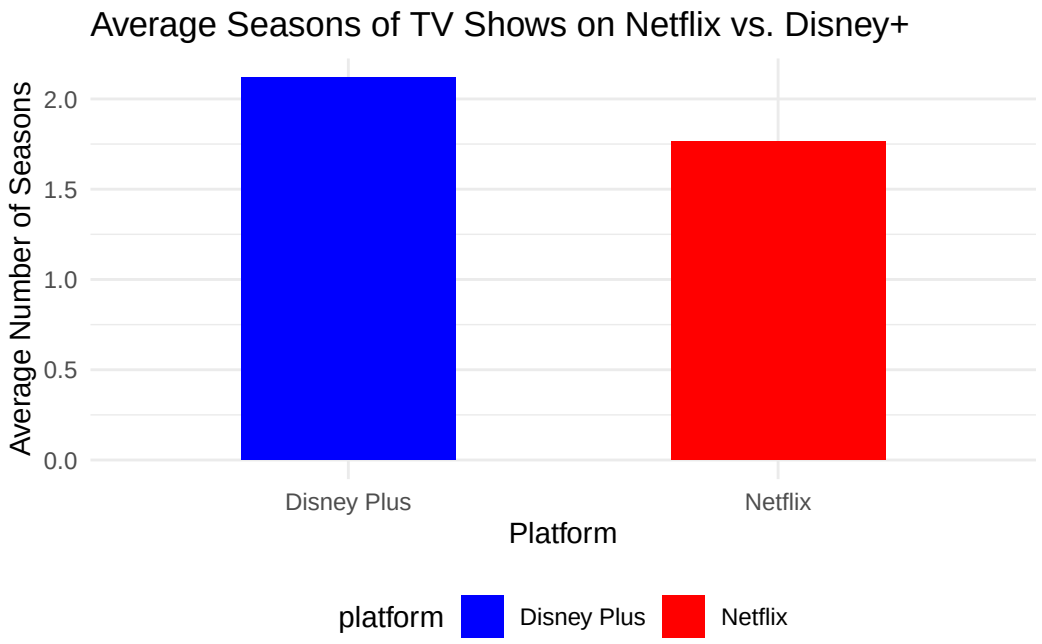
```
# Step 2: Create the trend line plot
ggplot(movie_trend, aes(x = release_year, y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + # Create bar plot with grouped bars
  labs(
    title = "Trend of Movies on Netflix vs. Disney+ Over the Years",
    x = "Release Year",
    y = "Number of Movies",
    fill = "Streaming Platforms"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) +
  theme_minimal()+ # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
```



**What is the average duration of movies and TV shows?**

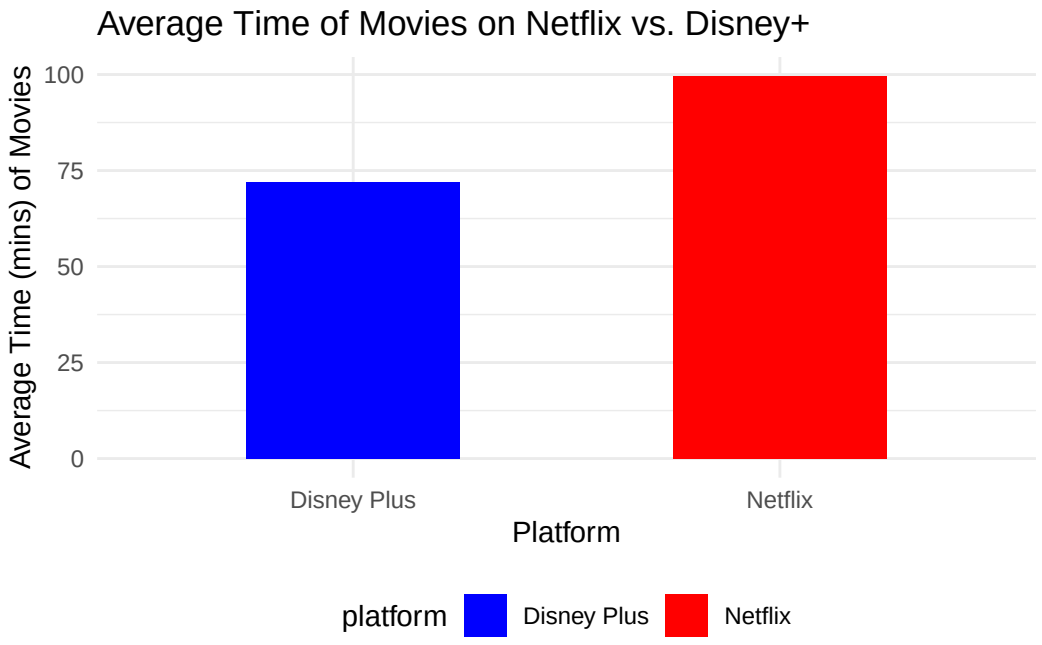
**TV Shows**

```
# For TV Shows
# Step 1: Wrangle the Movies 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 plot
ggplot(average_duration_tv, aes(x = platform, y = avg_seasons, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # 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(legend.position = "bottom")  # Positions the legend at the bottom of the plot
```



**Movies**

```
# For Movies
average_duration_movies <- movies %>%
  group_by(platform) %>% # Group by platform
  summarise(avg_time = mean(duration, na.rm = TRUE)) #Calculate mean number of seasons

# Step 4: Plot the data
ggplot(average_duration_movies, aes(x = platform, y = avg_time, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # 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 col
  theme_minimal() + # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
```



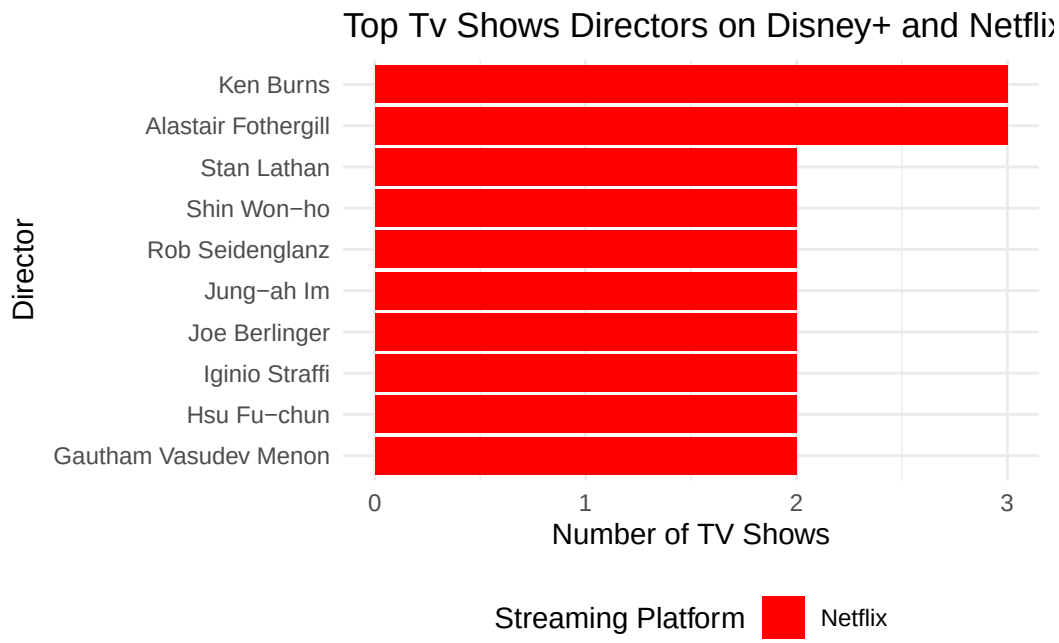

**Which directors are most frequently featured on Disney+?**

**TV Shows**

```r
# for tv shows
director_data_tv <- tvshows %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  separate_rows(director, sep = ",") %>%
  group_by(platform, director) %>%
  summarise(count = n()) # Count the number of appearances
```

```
`summarise()` has grouped output by 'platform'. You can override using the
`.groups` argument.
```

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

# Step 3: Plot the data
ggplot(top_directors_tv, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity") + # Bar plot without legend
  coord_flip() + # Flip 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( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) + # Set custom colors
  theme_minimal()+ # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
```



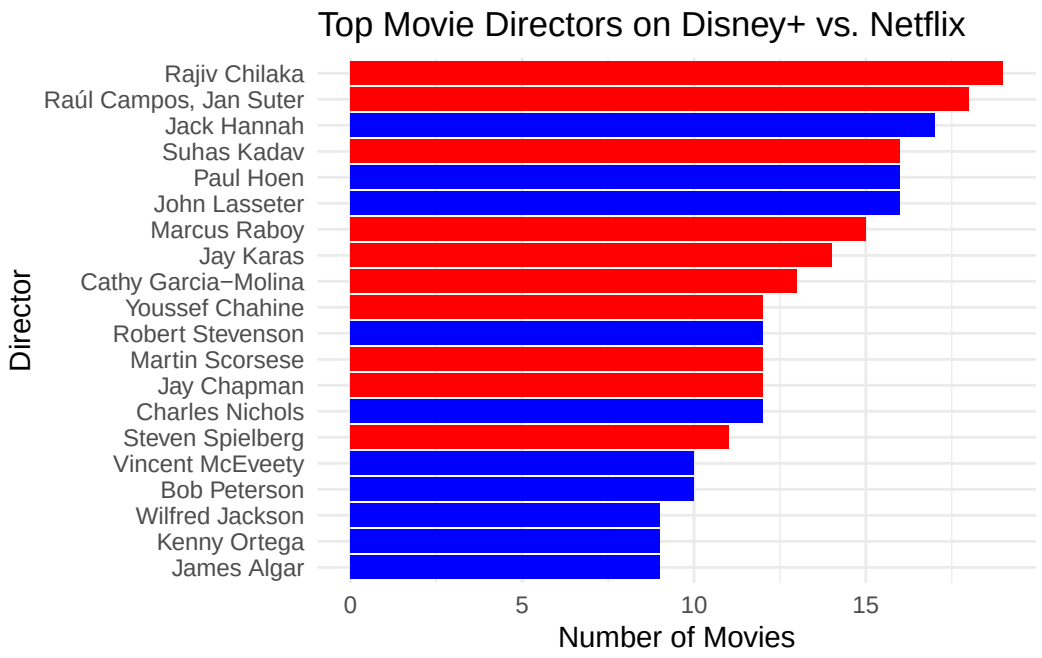

**Movies**

```
# For movies
director_data_movies <- movies %>%
  filter(!is.na(director)) %>%
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of appearances
```

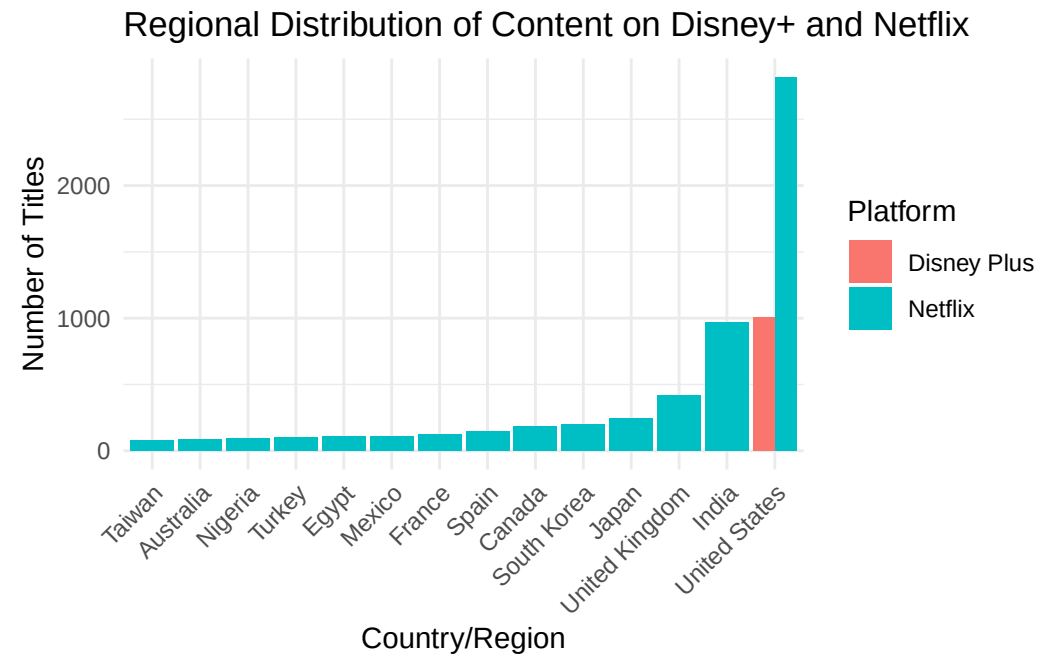
```
`summarise()` has grouped output by 'platform'. You can override using the
`.groups` argument.
```

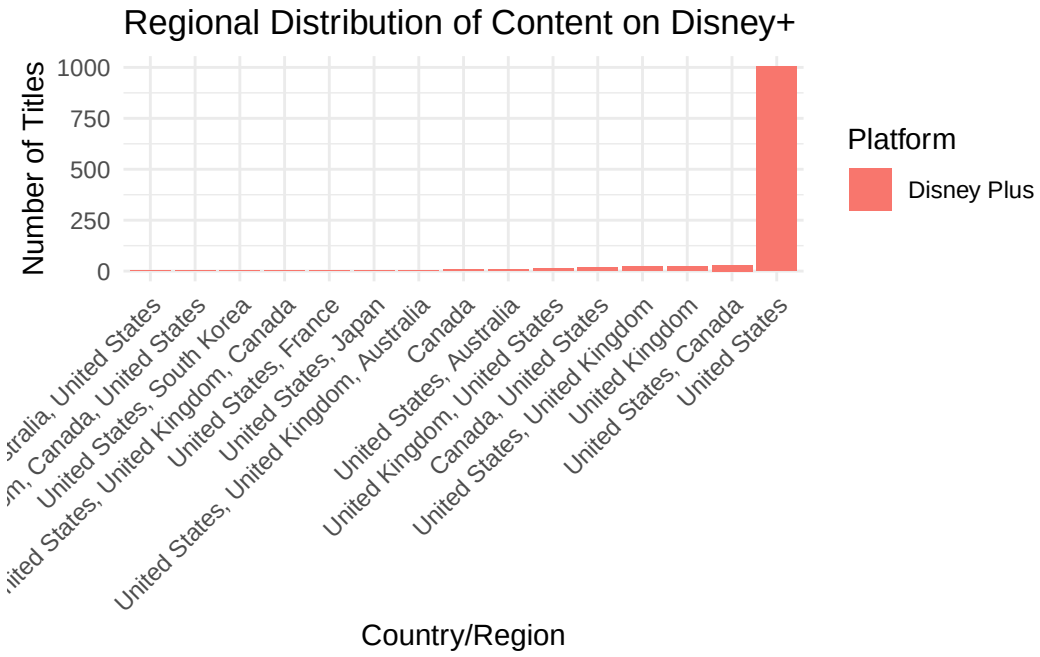
```
# Step 2: Get the top directors for each platform
top_directors <- director_data_movies %>%
  group_by(platform) %>%
  top_n(10, count) %>% # Select top 10 directors by count
  ungroup() # Ungroup the data

# 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
```

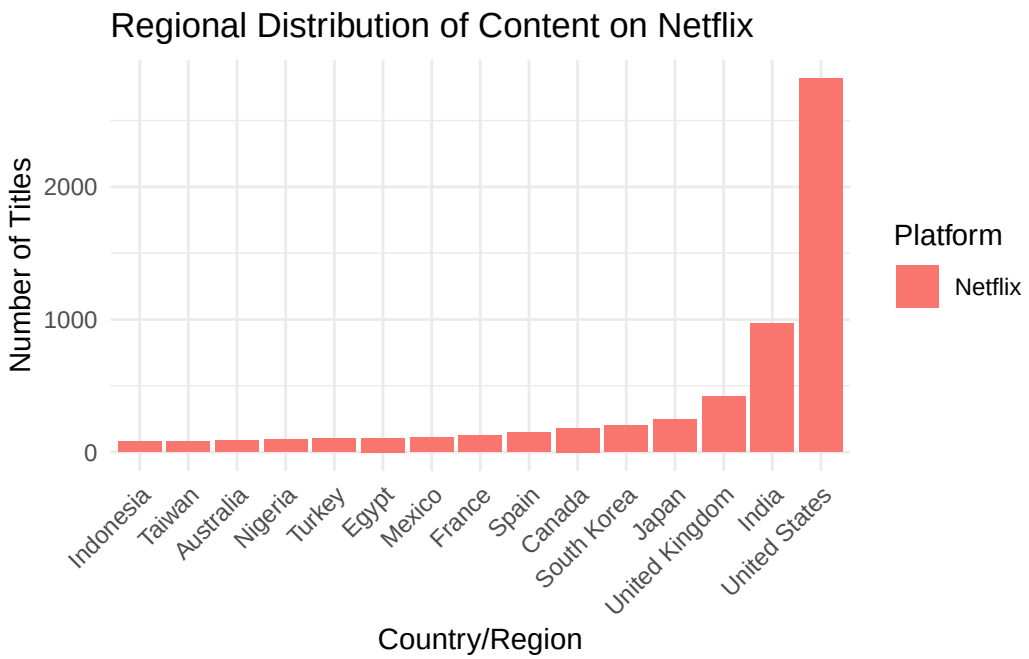
```r
  labs(
    title = "Top Movie Directors on Disney+ vs. Netflix",
    x = "Director",
    y = "Number of Movies",
    fill = " Streaming Platform"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) + # Set custom colors
  theme_minimal() + # Add a clean theme
  theme(legend.position = "bottom") # Positions the legend at the bottom of the plot
```



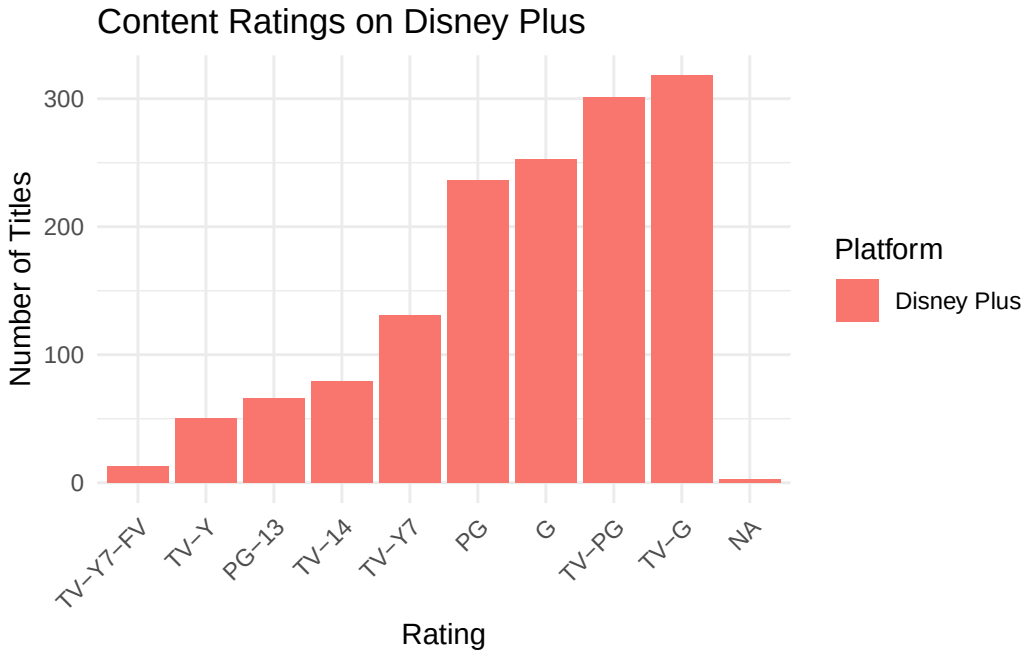
**What is the regional distribution of Disney+ content (based on country/region of production)?**

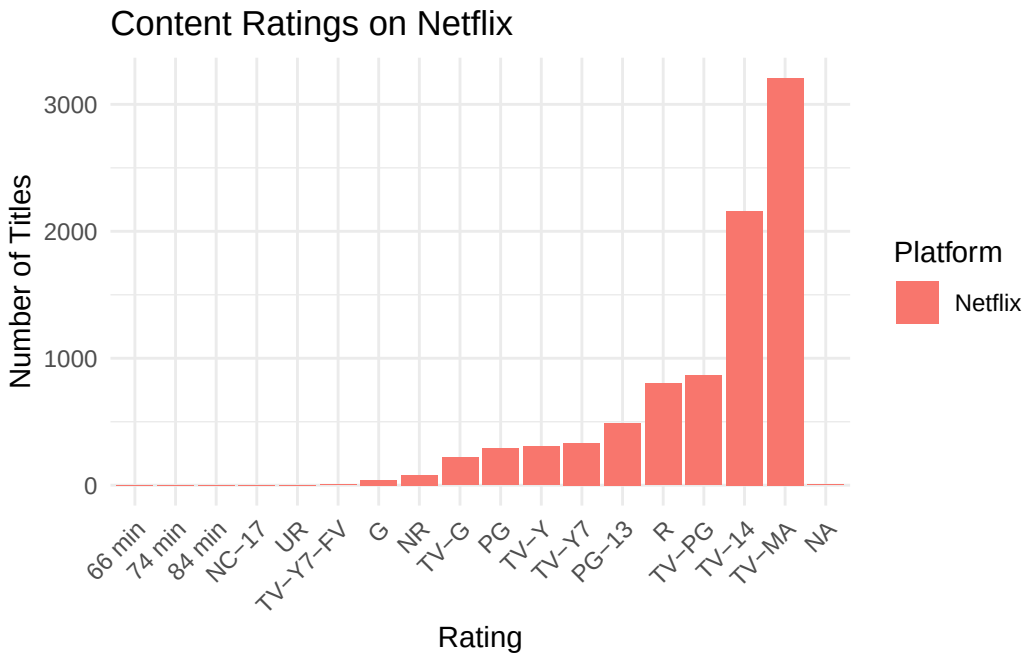





**Is there a trend in content ratings? For instance, does Disney Plus predominantly feature family-friendly content, or is there a growing inclusion of mature-rated titles?**






Disney Plus tailors to a younger audience whereas Netflix tailors to teens and young adults.

**Outside the US, where are most of Disney+ movies and TV Shows produced?**

**Date added to Disney Plus - Are the movies newer to Disney or have been on the platform for a long time?**

## Code Appendix

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

```r
netflix_data <- read.csv(netflix_url)
# Preview Disney+ (first 10 entries) with truncated text and adjusted font size for the table
head(disney_data, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 6) #Modify cell font size
# Preview Netflix the data
# Preview Disney+ (first 10 entries) with truncated text and adjusted font size for the table
head(netflix_data, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 6) #Modify cell font size

# Summarize Disney+ the data
summary(disney_data) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%
  row_spec(0, bold = TRUE, font_size = 4.5) %>%
  kable_styling(font_size = 6)%>%
  column_spec(1, extra_css = "text-align: left;")
# Summarize Netflix the data
summary(netflix_data) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%
  row_spec(0, bold = TRUE, font_size = 4.5) %>%
  kable_styling(font_size = 5)%>%
  column_spec(1, extra_css = "text-align: left;")
# 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)
# Add N/A values to replace missing data
moviesandtv <- moviesandtv %>%
  mutate(across(where(is.character), ~na_if(., "")))
#Separate the moviesandtv table into two distinct tables

movies <- moviesandtv %>%
  filter(type == 'Movie') %>%
  mutate(duration = as.integer(str_remove(duration, " min")))

tvshows <- moviesandtv %>%
  filter(type == 'TV Show') %>%
  mutate(duration = str_replace(duration, "Seasons", "Season")) %>%
  mutate(duration = as.integer(str_remove(duration, " Season")))
head(tvshows, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "0.95cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 5) #Modify cell font size
head(movies, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "0.95cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 5) #Modify cell font size
# 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 = "Trend of TV Shows on Netflix vs. Disney+ Over the Years",
    x = "Release Year",
    y = "Number of TV Shows",
    fill = "Streaming Platforms"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) +
  theme_minimal() + # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
# 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 trend line 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 = "Trend of Movies on Netflix vs. Disney+ Over the Years",
    x = "Release Year",
    y = "Number of Movies",
    fill = "Streaming Platforms"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) +
  theme_minimal()+ # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
# For TV Shows
# Step 1: Wrangle the Movies 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 plot
ggplot(average_duration_tv, aes(x = platform, y = avg_seasons, fill = platform)) +
```

```
  geom_bar(stat = "identity", width = 0.5) +  # 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(legend.position = "bottom")  # Positions the legend at the bottom of the plot
# For Movies
average_duration_movies <- movies %>%
  group_by(platform) %>% # Group by platform
  summarise(avg_time = mean(duration, na.rm = TRUE)) #Calculate mean number of seasons

# Step 4: Plot the data
ggplot(average_duration_movies, aes(x = platform, y = avg_time, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # 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(legend.position = "bottom")  # Positions the legend at the bottom of the plot
# for tv shows
director_data_tv <- tvshows %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  separate_rows(director, sep = ",") %>%
  group_by(platform, director) %>%
  summarise(count = n()) # Count the number of appearances

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

# Step 3: Plot the data
ggplot(top_directors_tv, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity") + # Bar plot without legend
  coord_flip() + # Flip 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( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) + # Set custom colors
  theme_minimal()+ # Add a clean theme
  theme(legend.position = "bottom")  # Positions the legend at the bottom of the plot
# For movies
director_data_movies <- movies %>%
  filter(!is.na(director)) %>%
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of appearances

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

# 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( # Set custom colors
    values = c("Disney Plus" = "blue", "Netflix" = "red")
  ) + # Set custom colors
  theme_minimal() + # Add a clean theme
  theme(legend.position = "bottom") # Positions the legend at the bottom of the plot

distribution <- moviesandtv%>%
  group_by(platform,country)%>%
  summarise(count = n(), .groups="drop")%>%
```

```
  arrange(desc(count))%>%
  filter(!is.na(country))%>%
  slice_head(n=15)

ggplot(distribution, aes(x = reorder(country, count), y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Regional Distribution of Content on Disney+ and Netflix",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  )+
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1) ) # Adjust label position
library(dplyr)
library(ggplot2)

disney_distribution <- moviesandtv %>%
  filter(platform == "Disney Plus") %>%
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country)) %>%
  slice_head(n = 15)

netflix_distribution <- moviesandtv %>%
  filter(platform == "Netflix") %>%
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country)) %>%
  slice_head(n = 15)


ggplot(disney_distribution, aes(x = reorder(country, count), y = count, fill = "Disney Plus"
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Regional Distribution of Content on Disney+",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
```

```
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )



ggplot(netflix_distribution, aes(x = reorder(country, count), y = count, fill = "Netflix")) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Regional Distribution of Content on Netflix",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )

disney_ratings <- moviesandtv %>%
  filter(platform == "Disney Plus") %>%
  group_by(rating) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count))

netflix_ratings <- moviesandtv %>%
  filter(platform == "Netflix") %>%
  group_by(rating) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count))

ggplot(disney_ratings, aes(x = reorder(rating, count), y = count, fill = "Disney Plus")) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Content Ratings on Disney Plus",
    x = "Rating",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  theme_minimal() +
```

```
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )

ggplot(netflix_ratings, aes(x = reorder(rating, count), y = count, fill = "Netflix")) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Content Ratings on Netflix",
    x = "Rating",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  )
```