Final Project Report

Gabriel Casillas and Peter Melling

University of Denver

LIS4210 Data Visualization

Dr. Shimelis Assefa

June 1, 2021

**Final Project Report**

**Documentation**

The data used in the analysis was a dataset compiled and scraped by Shivam Bansal from flixable.com and hosted on Kaggle. The dataset focused on Netflix’s current holdings—as of 2020—in the data fields of title, release date, country, genre the production was listed in, type of production, runtime, maturity rating, cast, producer, and description. Some of the data fields contained multiple elements in a single row, e.g., country—United States, Poland. It was not specified as to if the dataset was all of Netflix’s current holdings, or if it was content only accessible in a single region. It was also not specified within the data field ‘country’ as to whether or not the multiple countries actively produced the content or if those were the countries where the content was filmed. The last data field that was not specified was runtime for TV shows; the number of seasons currently on Netflix may not accurately reflect the number of seasons that have been produced or have been greenlit for production.

The original dataset did not contain viewer ratings. In order to further explore the dataset, two smaller datasets were created tabularly and introduced—IMDB’s Top 10 Movies, and IMDB’s Top 10 TV Shows. The two datasets were of current—2021—rankings and did not specifiy the collection date or how many viewers had ranked the items in the lists.

**Questions**

1. What is the correlation between runtime and country? What is the relationship between runtime and type of production?
2. Which countries are the most represented on Netflix? Of the most represented countries, what is the split between TV shows and movies?
3. Which countries are represented in co-productions? Which type of production is co-produced the most?
4. How long do TV shows run on Netflix? What is the correlation between number of seasons, genre, and country?
5. What is the correlation between IMDB popularity ratings with Netflix release year? How often are Netflix productions reflected in IMDB’s Top Rated?

**Hypotheses**

1. Certain runtimes would be more culturally acceptable in certain countries.
2. English-language media would have the largest share of the platform, with the United States being the most-prolific producer of media.
3. Major Westernized nations will be represented the most in co-produced titles. Movies will represent the most co-produced type of production hosted on Netflix.
4. The TV shows hosted on Netflix will have five seasons or under. The average number of TV shows produced will come from Westernized nations. The most popular TV show genres will have more seasons of runtime and will be produced by Westernized nations.
5. Multiple titles hosted on Netflix will appear in IMDB’s Top Rated TV shows and movies.

\*Note: No hypothesis was generated for correlation between IMDB popularity ratings with Netflix release year.

**R Packages**

RStudio and RMarkdown were used to analyze, create visualizations, and display the data sets. Multiple packages and tools were used throughout the data analysis, including the publishing of the analysis and visualizations.

The packages used were: tidyverse, readr, [corrplot], [RColorBrewer], dplyr, ggplot2, htmlwidgets, maps, hrbrthemes, flexdashboard, and shiny. tidyverse has a wide range of functionality due the large library of tools house within it. readr was used to read the .csv files that housed the data. ggplot2 was used to plot and produce many of the visualizations. flexdashboard was used to produce an interactive dashboard in which the data, code, visualizations, and text chunks could be published and viewed on the web. shiny was used alongside flexdashboard in order to create an interactive visualization on the web.

Some of the packages were not used due to the nature of the package falling outside of the scope of analysis. Packages such as lubridate, tmap, leaflet, plotly, maps, htmlwidgets, and hrbrthemes were not used in the final RMarkdown script or the final shiny dashboard. While these packages were not used, they would be helpful in reexamining the data for further exploration and visualizations.

**Data Explorations**

***Cleaning***

The original dataset taken from Kaggle was relatively clean. The original file was not UTF-8 encoded and needed to be changed to correctly display diacritics; it was brought into Excel and reformatted. When looking for unique values to work with, a row was identified that had many of the columns shifted left into incorrect field columns. To make sure the row was realigned to the correct columns, the dataset was brought into Excel again and the individual fields shifted by hand; this resulted in a few columns having a null placement that was not previously there beforehand and filled fields that had previously been null.

***Transforming***

Two of the columns—country and listed\_in—possessed multiple factors deliminated by a comma and an ampersand. Director, cast, and description also had multiple factors in each column, but due to the analysis not focusing on these data points they were left alone. In order to make sure the data split and was attributed to the correct id, the data was brought into OpenRefine. The columns country and listed\_in were split on commas and on ampersands, which created new rows for each separate factor. The new rows were null minus the new splits. To make sure each of the new rows was attributed to the correct id, each column was filled down so that id’s with multiple factors in the country and listed\_in columns could be identified. This transformation of the dataset was saved as a separate file so that there would be two distinct files that could be worked with; one with the cleaned data, and one with the clean and transformed data for country and listed\_in.

**Data Analysis**

***Runtime***

The initial hypothesis was that certain runtimes were more culturally acceptable in certain countries, which a platform like Netflix was bound to reflect. However, the data proved to be a problem.

Due to the nature of how Bansal coded runtime data, movie runtimes were character strings. It can be postulated that Bansal was attempting to account for both TV shows and movies in the same dataset. In order to narrow this part of the data, a filter was used to select only “Movie”-type items. After that, a combination of gsub and as.numeric removed the “min” suffix and made all the runtimes numeric. Despite these efforts, the data would not cooperate well with plotting. Instead of a 0-100, 101-200, etc. pattern, the data adopted a 10-100, 20-200, etc. pattern. This made for murky plotting, revealing little in the way of useful information. On top of that, a lack of grouping runtimes and the 110 countries made the chart barely-readable from any distance.

***International Representation***

**Top Representation**

Given that there were 110 countries individually represented on Netflix, it was imperative to determine which nations were the most-represented on the service. The transformed dataset was used to determine which countries have the most media on Netflix due to the nature of the splitting of countries on a delimiter. There was a postulation that English-language media would have the largest share of the platform, with the United States being the most-prolific producer of media.

The first inquiry was into the ten most-represented countries on the platform. Through a simple count and sort code, the countries that were identified with the most titles on Netflix were:

1. United States (6802)
2. India (3045)
3. United Kingdom (1459)
4. Canada (750)
5. Japan (709)
6. France (627)
7. South Korea (626)
8. Spain (502)
9. Germany (359)
10. Mexico (343)

These results made sense with the theorizing on English-language media and the US’ production rates. India’s second-place spot also made sense when considering the firmly-established South Asian communities within the United States and the overall population size/film industry output of India. India’s film industry produced double the amount of movies that the US did and Indian movies have no dedicated streaming service (e.g., HBO Max, Disney+, Paramount+, etc.)(Keelery, 2020). Based on this data, a conclusion can be drawn about Netflix’s deal with Indian film distributors. With regards to English-language media, the United Kingdom and Canada being at numbers three and four respectively followed a similar trend in that they have had the largest film and television industries in English. It can be extrapolated that the amount of United Kingdom titles could be larger, but that number has likely gone down due to the two largest UK TV producers (BBC and ITV) forming the streaming platform Britbox in 2017 (Szalai, 2017). Countries like Japan, Spain, and Mexico being on the list again followed the trend of population demographics and the size of their film/TV industries.

**Top Production Type**

Following top production countries, the data could be further narrower to determine whether the ten countries produced or hosted more TV shows or movies on the platform. The hypothesis for this line of inquiry was that movies would have more representation on the platform than TV shows, due to both the smaller cost overhead from producing/licensing movies than TV and the existence of many platforms designed for serialized programing (e.g., Hulu+, Britbox, etc.). Through both another count argument and a merge with the list established for the previous part, it was determined the split between each of the ten countries’ film and TV representation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | country | n.x | type | n.y |
| 1 | Canada | 750 | TV Show | 240 |
| 2 | Canada | 750 | Movie | 510 |
| 3 | France | 627 | TV Show | 166 |
| 4 | France | 627 | Movie | 461 |
| 5 | Germany | 359 | Movie | 267 |
| 6 | Germany | 359 | TV Show | 92 |
| 7 | India | 3045 | Movie | 2790 |
| 8 | India | 3045 | TV Show | 255 |
| 9 | Japan | 709 | Movie | 294 |
| 10 | Japan | 709 | TV Show | 415 |
| 11 | Mexico | 343 | TV Show | 138 |
| 12 | Mexico | 343 | Movie | 205 |
| 13 | South Korea | 626 | TV Show | 485 |
| 14 | South Korea | 626 | Movie | 141 |
| 15 | Spain | 502 | TV Show | 170 |
| 16 | Spain | 502 | Movie | 332 |
| 17 | United Kingdom | 1459 | TV Show | 694 |
| 18 | United Kingdom | 1459 | Movie | 765 |
| 19 | United States | 6802 | Movie | 4873 |
| 20 | United States | 6802 | TV Show | 1929 |

Table 1. *Ten countries’ film and TV representation*.

For most countries, the hypothesis about the discrepancy between movie and TV show representation was correct. Most notable was the drop in India’s representation, with only 255 shows compared to 2790 movies. It can be theorized that there were likely different licensing agreements between movie studios and TV production houses, as well as the added cost of translating many episodes of TV shows versus the cost of subtitling movies. It would also explain the relatively smaller numbers of all non-English TV shows on the platform.

Japan and South Korea went opposite of the trend with more TV shows than movies on the platform. This could be the result of which media companies Netflix made distribution deals with or an overall reflection of a discrepancy in industry size/export demand. Japan’s prolific animation industry could also be driving its TV numbers higher than movies since animated programing has long been its most notable entertainment export worldwide (Pineda, 2020). This could change in the future with the rise in Japanese animation-focused streaming services (e.g., Crunchyroll), but it looks to be a stable snapshot given the overall export numbers of media and Netflix’s likely pre-existing deals with distributors and production companies.

***International Co-Production***

The next progression in questioning the data was in regards to number of co-productions internationally. It should be noted that the dataset and the Kaggle repository did not define if the column country was the countries that produced the content, locations in which the content was filmed, or where the content was available to be viewed. Looking at individual instances in the data, it could be co-production or filming locations depending on the metadata for each individual instance and ISAN (International Standard Audiovisual Number).

Since the data was split into two separate sets—cleaned and transformed—it was imperative to look at just the cleaned set to calculate how many listings were co-productions so as to not count each instance multiple times through the transformed set. The function count was used to locate all the commas in the country column for 682 values; 777 can be found through looking for country, which returns co-productions and individual productions; using the grepl function, 659 can be found. With such a large dataset with two separate delimiters being used, locating the number of co-productions can vary depending on how the data is being filtered.

Out of the co-productions, there was a definitive difference between the movies and TV shows hosted on Netflix. Movies were most often co-produced as opposed to TV shows (Fig.1).

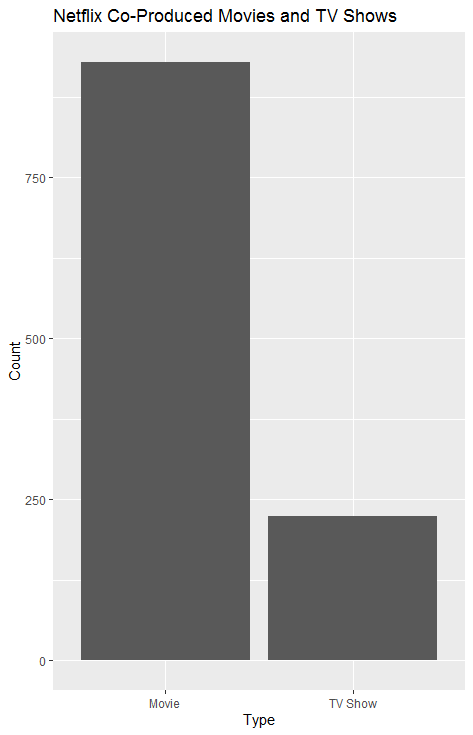


Figure 1. *Netflix co-produced movies and TV shows.*

In order to determine why, it could further be explored against date of production. It can be theorized that there are more movie co-productions than TV shows due to the fact that a single film is much easier to collaborate and produce than an extended TV show that may be affected by numerous factors such as transportation, casting, language, budget, seasonal weather, contracts, viewership ratings, etc.

**Netflix Shows**

There was much that could be looked at in terms of Netflix and their TV show offerings, such as number of seasons and genre, or number of seasons and country or production, or even number of seasons and production date.

***Seasons***

Overall, the average number of seasons for a TV show hosted on Netflix is 1. While a single season is the most common, there are even shows that have up to 16 seasons being hosted on Netflix. Looking at seasons alone, there is much that could be hypothesized about; one such instance being: Netflix has begun to produce content in more recent years, which has led to an influx of new TV shows that were all currently in their first season.

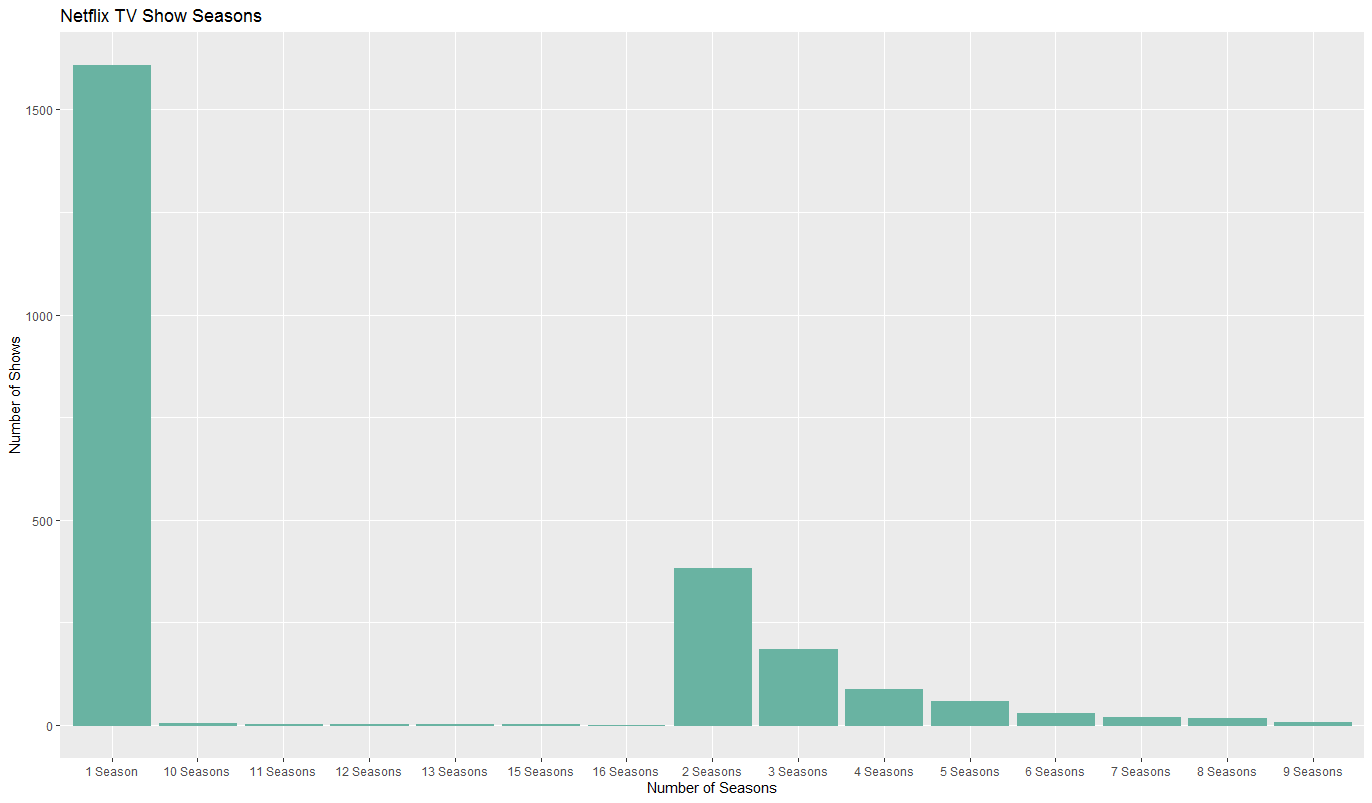


Figure 2. *Netflix TV show seasons*.

Top of Form

Most shows that Netflix offers were found to only have one or two seasons at most. The data did not take into account whether these shows were created for Netflix, were being hosted on Netflix, if the shows have been cancelled, or if the shows were currently in production or have been greenlit for further production. The graph also does not reflect the number of seasons and year of production.

***Genre***

Looking to how many seasons TV shows have on Netflix as well as what genres those TV shows may be in, the question would need to be further broken up and explored by individual season or by a single genre at a time. An interactive graph would also allow for further filtering and narrowing of results. The genres of shows that Netflix hosts and produces covered a wide range. They also differ widely across the number of seasons for each show.

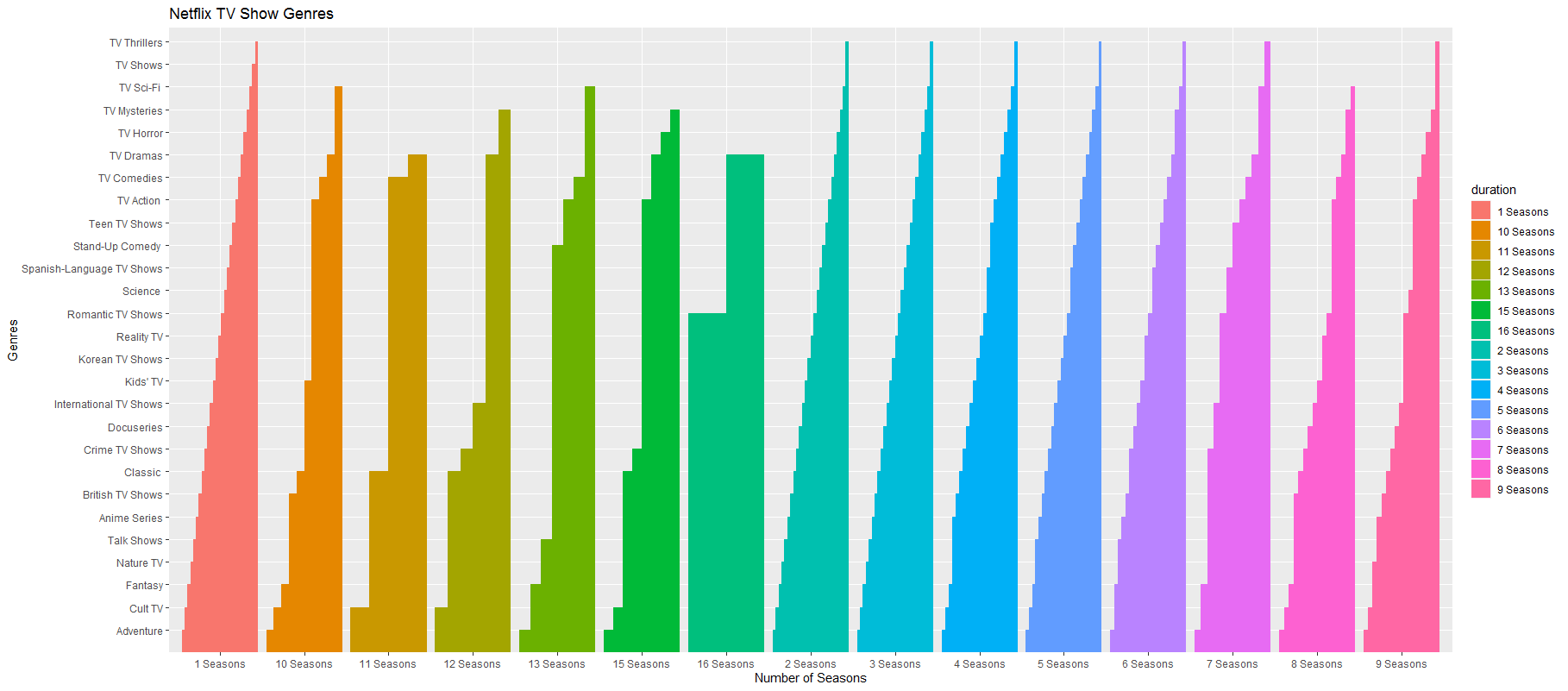


Figure 3. *Netflix TV show genres.*

Figure 3 displays the more popular genres at the bottom. The thickness of the bar graph shows the amount of shows within that genre. Genres that do not have many seasons can be viewed at the top of the graph and the thickness of the individual bars. Lower seasons tend to have more content being produced across the various genres. Adventure, Cult TV, Fantasy, Nature TV, and Talk Shows are amongst the top five that tend to have the most seasons as well as the most number of shows. Thrillers, Shows in general, and Sci-Fi seem to have far fewer numbers in any number of seasons.

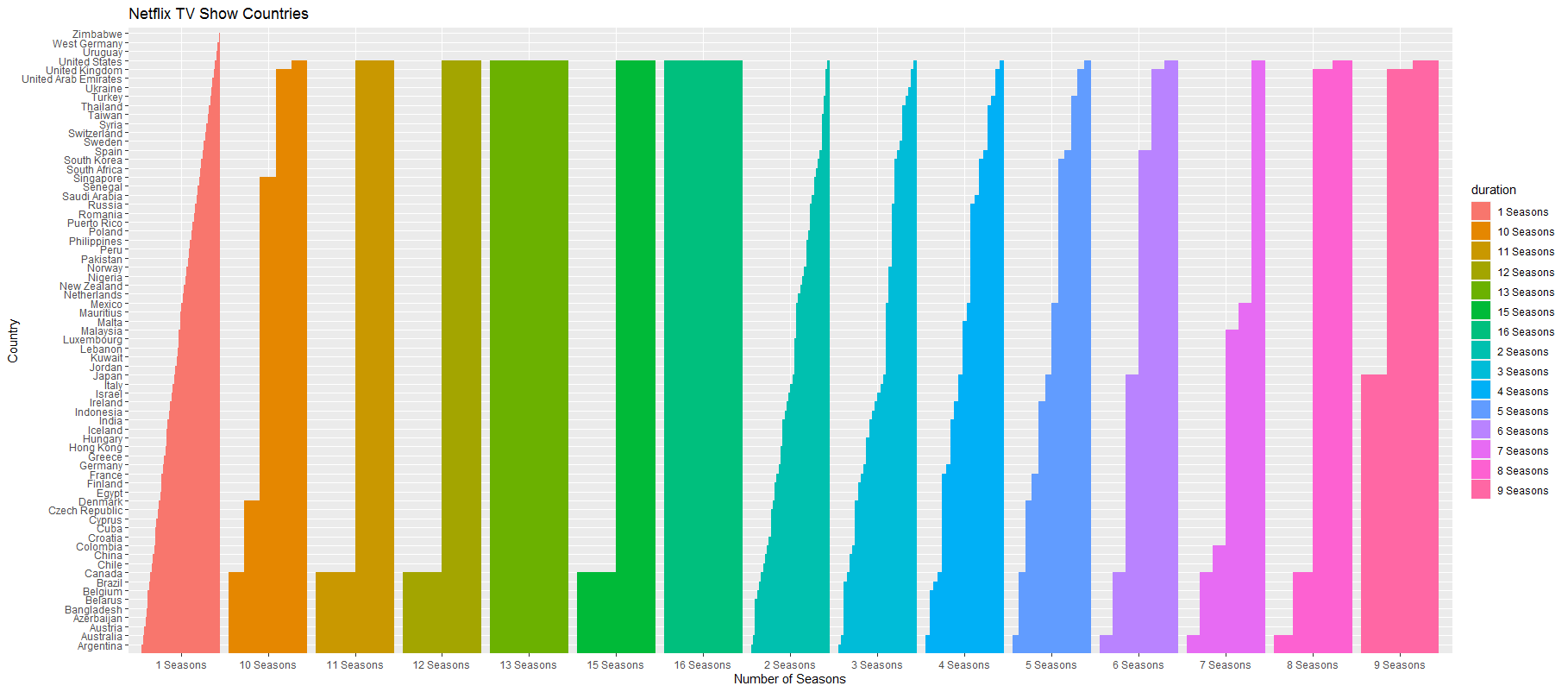
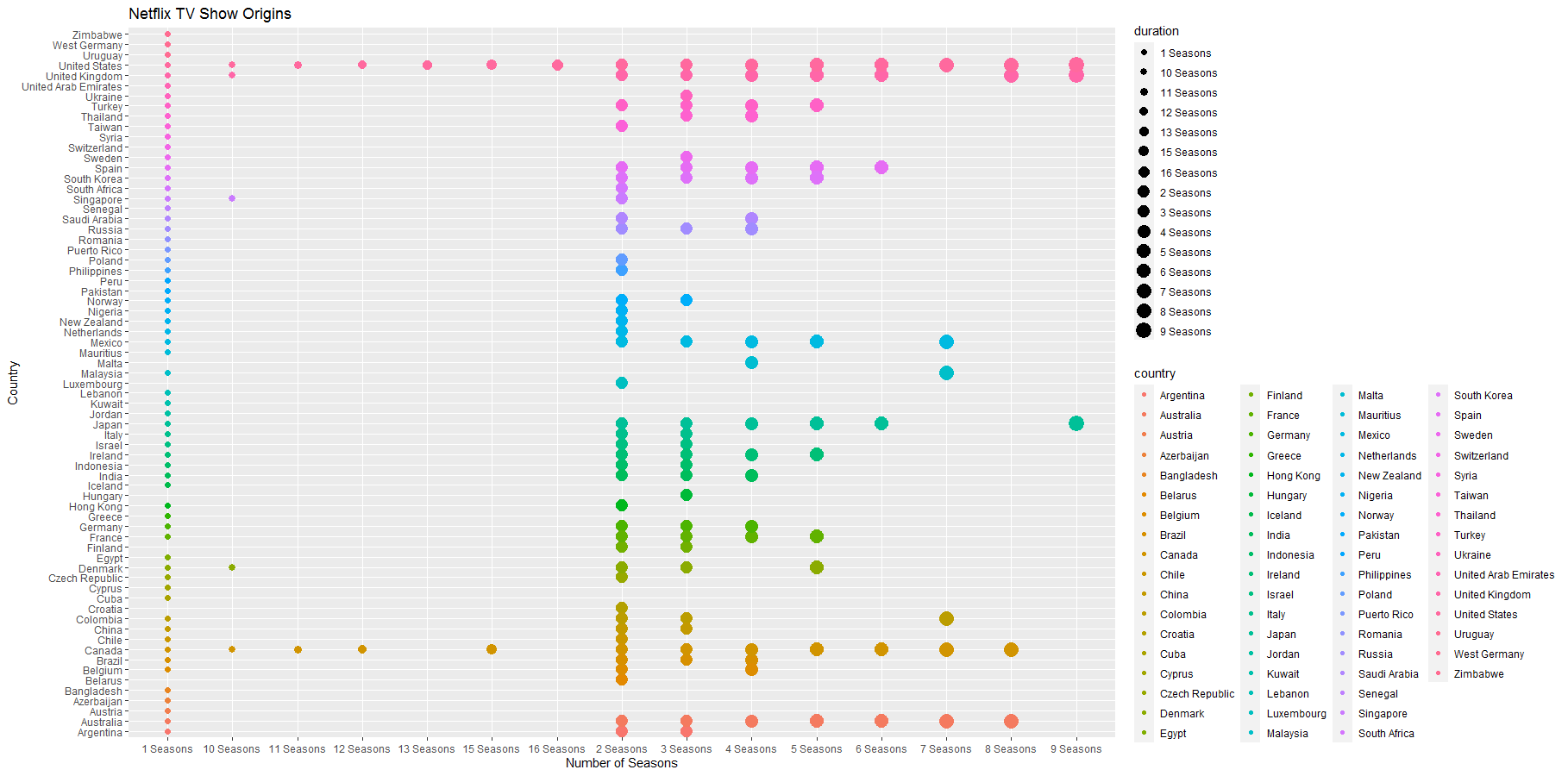
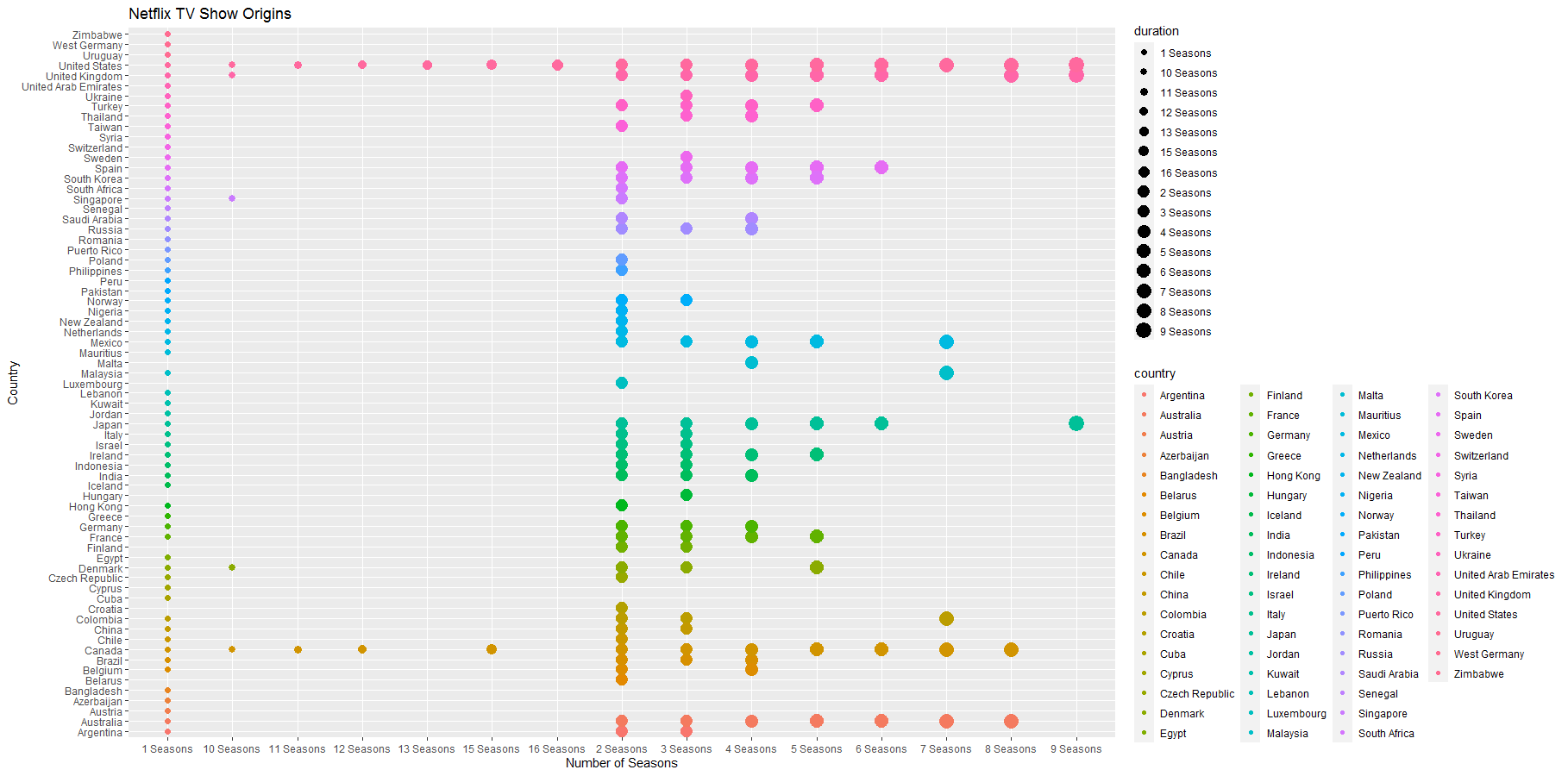


Figure 4. *Netflix TV show countries.*

The second graph that addressed seasons and another factor were individual countries that host or produce content for Netflix. Many countries only produce a small number of seasons for a show, which can be viewed as the thinner portions of the colored bars in the graph. The countries that produce the most content are found along the bottom of the graph. As the number of seasons went up, the number of countries producing content went down. Since there were a lot of individual countries that produced content, a further exploration of the question and the data with an interactive graph centered on region of countries, it may lend itself to being easier to work with, compute, and visualize.

When looking at multiple factors, such as number of seasons produced, the country that produced the content, and the amount of content they produced, the data can quickly become just as cluttered (fig. 5).

Bottom of Form

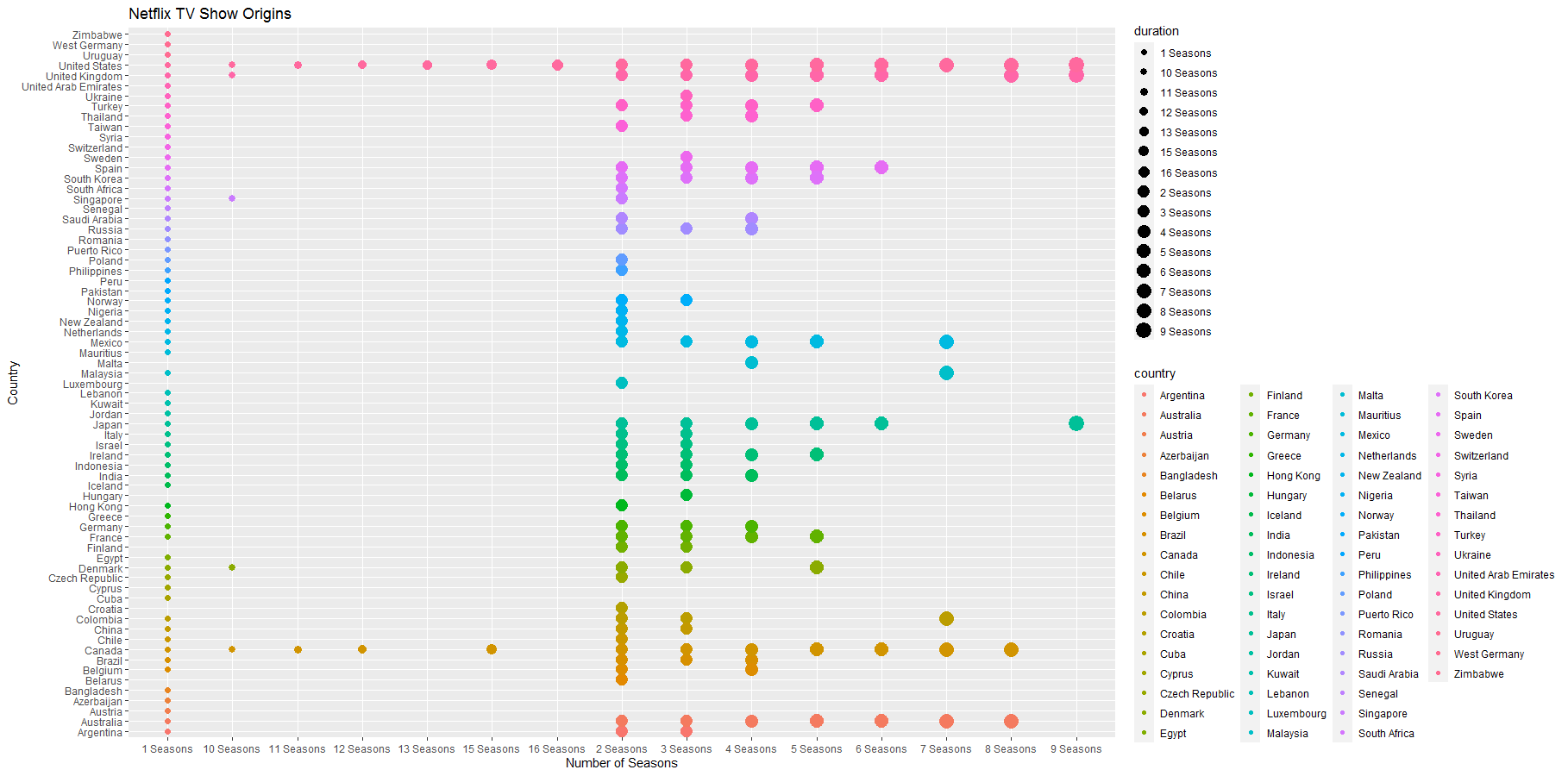


Figure 5. *Netflix TV show origins.*

Despite the difficulty in seeing more of the data graphed, Many countries have produced TV shows in season 1 and season 2 where it then further narrows. The top producers of content can be seen in the higher seasons, such as the United States, Canada, Singapore, and the United Kingdom.

***IMDB***

**Ratings and Release Year**

When first assessing the data, one of the questions the analysis looked to investigate was the correlation between release year and IMDB ratings on either TV shows, movies, or both. This exploration was not explored, but the question can still be readdressed in future explorations of the data.

**Netflix and IMDB**

Two separate sets of data were introduced in order to draw further conclusions about the Netflix data and what it had to offer. The top ten results from both the Top 250 movies and TV shows was tabulated and charted against what Netflix had in its listings. Out of both TV shows and movies, Netflix only had three TV shows and three movies (table 2).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| title | type | release\_year | rated | director | country | rating | duration | listed\_in |
| Blue Planet II | TV Show | 2017 | 9.3 |  | United Kingdom | TV-G | 1 Season | British TV Shows, Docuseries, Science & Nature TV |
| Our Planet | TV Show | 2019 | 9.3 |  | United States, United Kingdom | TV-PG | 1 Season | Docuseries, Science & Nature TV |
| Planet Earth II | TV Show | 2016 | 9.5 |  | United Kingdom | TV-G | 1 Season | British TV Shows, Docuseries, Science & Nature TV |
| Pulp Fiction | Movie | 1994 | 8.8 | Quentin Tarantino | United States | R | 154 min | Classic Movies, Cult Movies, Dramas |
| Schindler’s List | Movie | 1993 | 8.9 | Steven Spielberg | United States | R | 195 min | Classic Movies, Dramas |
| The Lord of the Rings: The Return of the King | Movie | 2003 | 8.9 | Peter Jackson | New Zealand, United States | PG-13 | 201 min | Action & Adventure, Sci-Fi & Fantasy |

Table 2. *Netflix movies and TV shows in IMDB’s top 10.*

The Netflix dataset was joined with the two IMDB datasets in order to produce a series of tables to compare some of the information between IMDB and Netflix, as well as to help build further on the Netflix set by including a popularity rating.

Based on the joined tables alone, Netflix hosts short nature or short docuseries for TV shows, as well as classic movies in order to draw in viewership. Comparing release year and IMDB rating, viewers may prefer newer TV shows for factors such as content, current, length, and cinematography; viewers may also prefer older movies that are longer and have well known directors.

**Visualizations**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | country | n.x | type | n.y |
| 1 | Canada | 750 | TV Show | 240 |
| 2 | Canada | 750 | Movie | 510 |
| 3 | France | 627 | TV Show | 166 |
| 4 | France | 627 | Movie | 461 |
| 5 | Germany | 359 | Movie | 267 |
| 6 | Germany | 359 | TV Show | 92 |
| 7 | India | 3045 | Movie | 2790 |
| 8 | India | 3045 | TV Show | 255 |
| 9 | Japan | 709 | Movie | 294 |
| 10 | Japan | 709 | TV Show | 415 |
| 11 | Mexico | 343 | TV Show | 138 |
| 12 | Mexico | 343 | Movie | 205 |
| 13 | South Korea | 626 | TV Show | 485 |
| 14 | South Korea | 626 | Movie | 141 |
| 15 | Spain | 502 | TV Show | 170 |
| 16 | Spain | 502 | Movie | 332 |
| 17 | United Kingdom | 1459 | TV Show | 694 |
| 18 | United Kingdom | 1459 | Movie | 765 |
| 19 | United States | 6802 | Movie | 4873 |
| 20 | United States | 6802 | TV Show | 1929 |

Table 1. *Ten countries’ film and TV representation*.

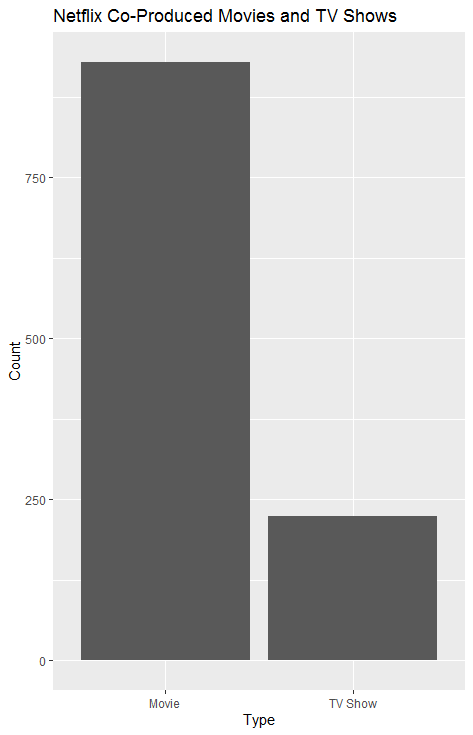


Figure 1. *Netflix co-produced movies and TV shows.*

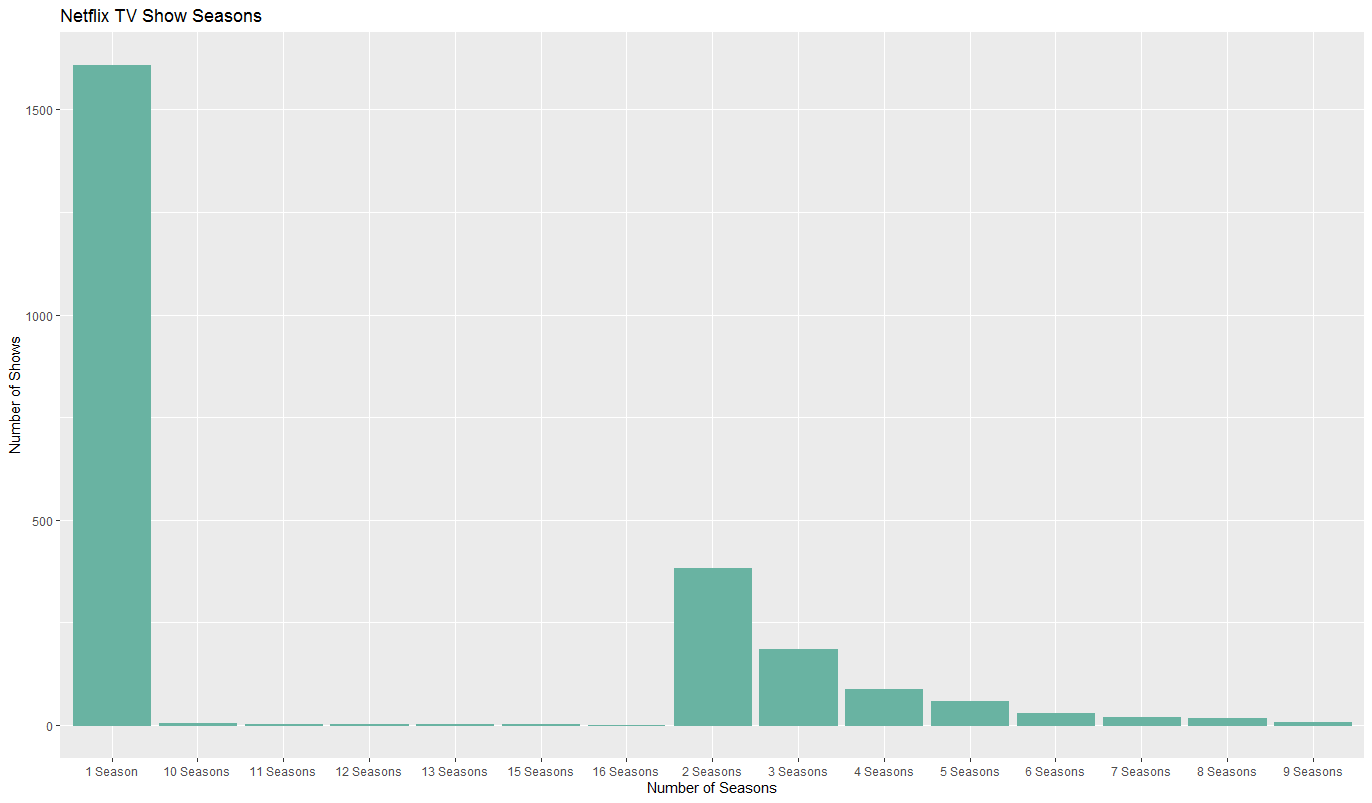


Figure 2. *Netflix TV show seasons*.

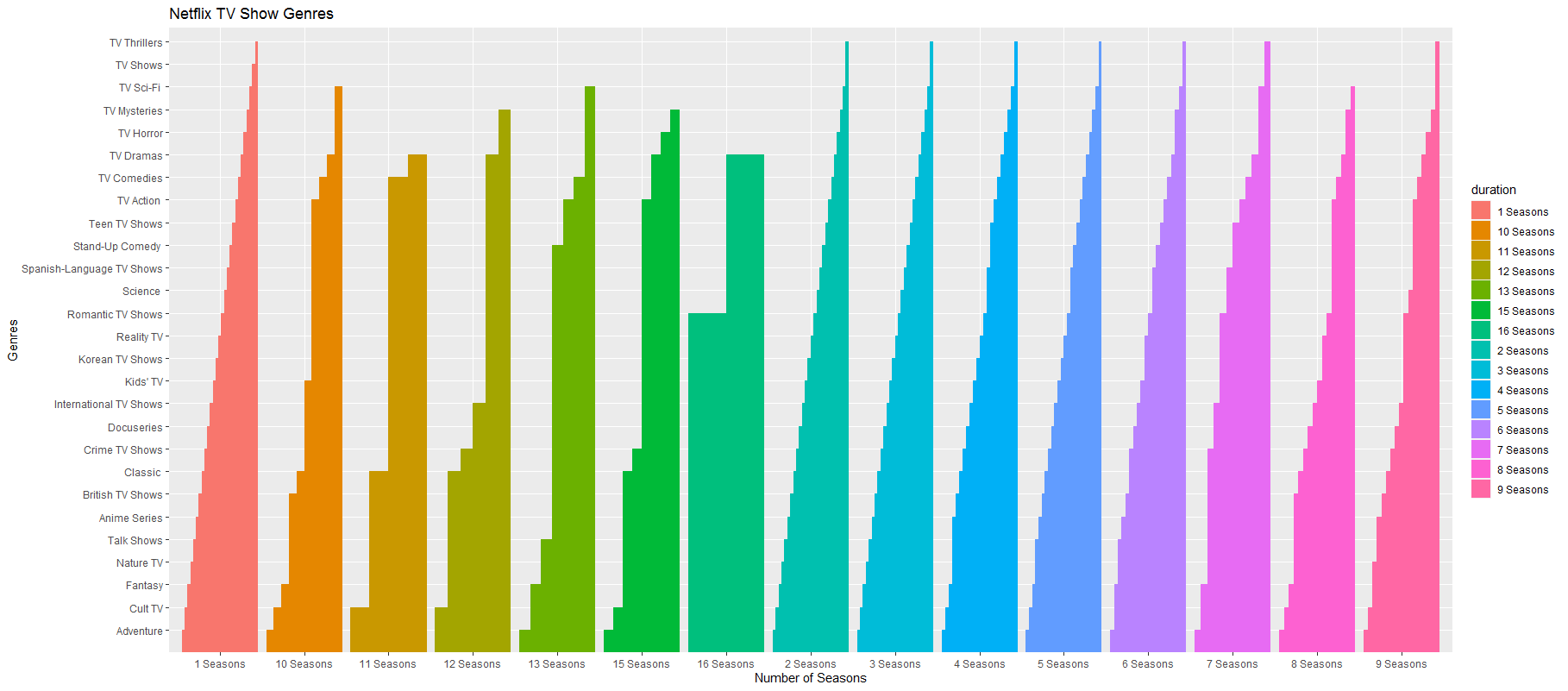


Figure 3. *Netflix TV show genres.*

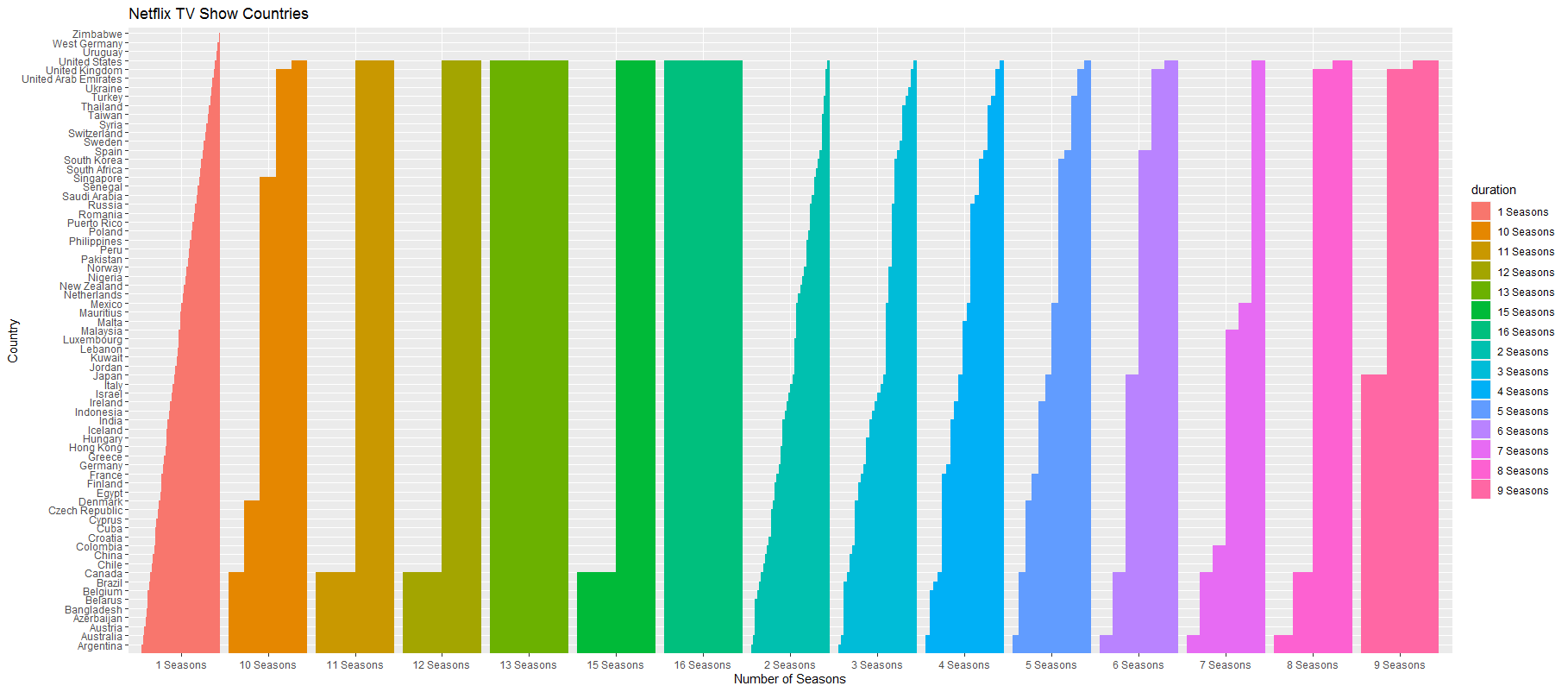


Figure 4. *Netflix TV show countries.*

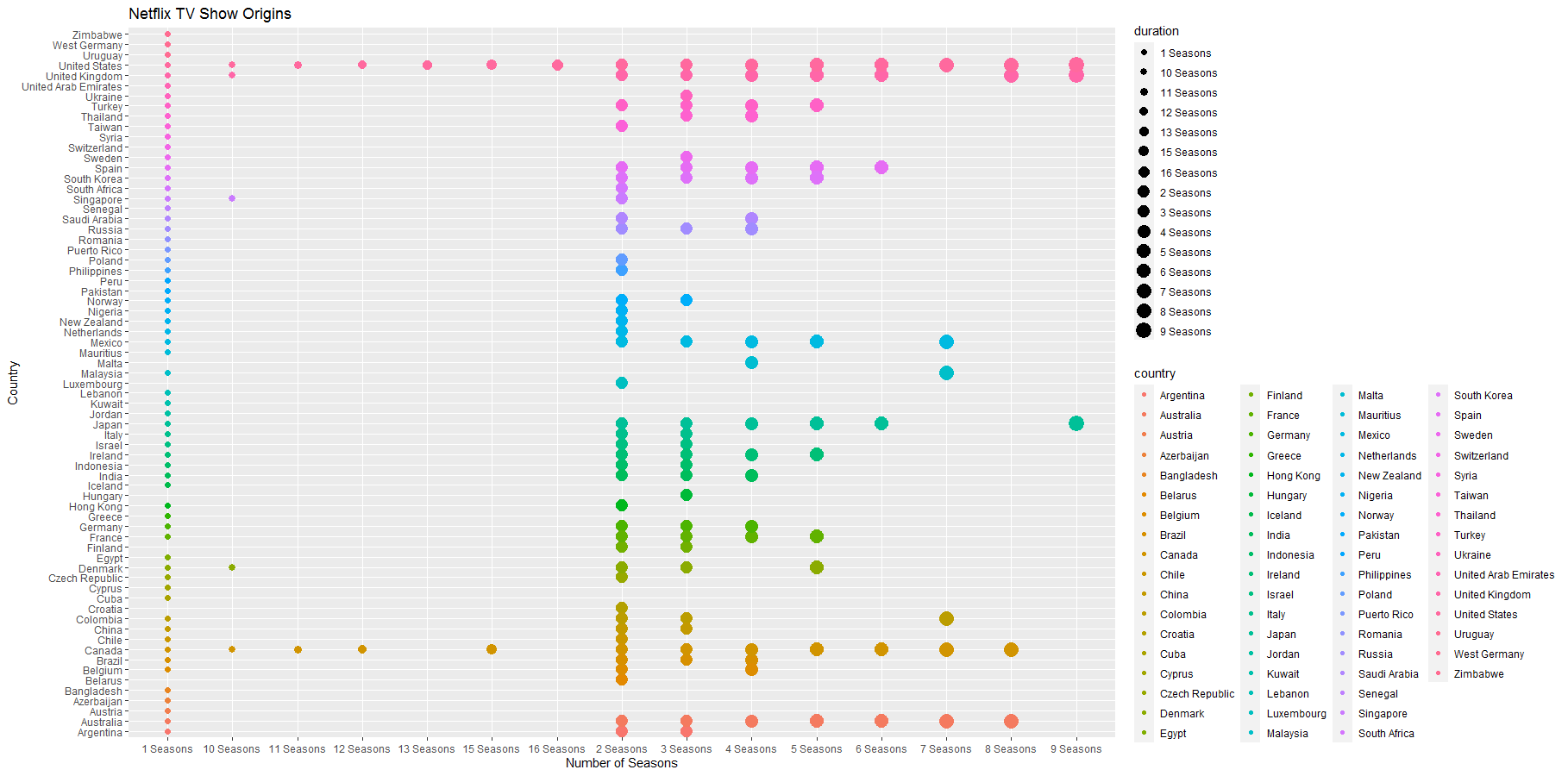


Figure 5. *Netflix TV show origins.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| title | type | release\_year | rated | director | country | rating | duration | listed\_in |
| Blue Planet II | TV Show | 2017 | 9.3 |  | United Kingdom | TV-G | 1 Season | British TV Shows, Docuseries, Science & Nature TV |
| Our Planet | TV Show | 2019 | 9.3 |  | United States, United Kingdom | TV-PG | 1 Season | Docuseries, Science & Nature TV |
| Planet Earth II | TV Show | 2016 | 9.5 |  | United Kingdom | TV-G | 1 Season | British TV Shows, Docuseries, Science & Nature TV |
| Pulp Fiction | Movie | 1994 | 8.8 | Quentin Tarantino | United States | R | 154 min | Classic Movies, Cult Movies, Dramas |
| Schindler’s List | Movie | 1993 | 8.9 | Steven Spielberg | United States | R | 195 min | Classic Movies, Dramas |
| The Lord of the Rings: The Return of the King | Movie | 2003 | 8.9 | Peter Jackson | New Zealand, United States | PG-13 | 201 min | Action & Adventure, Sci-Fi & Fantasy |

Table 2. *Netflix movies and TV shows in IMDB’s top 10.*

**R Script**

---

title: "Netflix Productions and IMDB"

author: "Gabriel Casillas and Peter Melling"

date: "5/31/2021"

output: flexdashboard::flex\_dashboard

orientation: rows

runtime: shiny

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

```{r include=FALSE, echo=FALSE}

library(tidyverse)

library(readr)

library(corrplot)

library(RColorBrewer)

library(dplyr)

library(ggplot2)

library(htmlwidgets)

library(maps)

library(hrbrthemes)

```

```{r, echo=FALSE}

netflix <- read.csv('netflix.csv')

net <- read.csv('net.csv')

```

# Netflix

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

The dataset focused on Netflix’s current holdings—as of 2020—in the data fields of title, release date, country, genre the production was listed in, type of production, runtime, maturity rating, cast, producer, and description.

The original dataset did not contain viewer ratings. In order to further explore the dataset, two smaller datasets were created tabularly and introduced—IMDB’s Top 10 Movies, and IMDB’s Top 10 TV Shows.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### Tools and Packages

The tools that were used were RStudio, RMarkdown, Shiny, and OpenRefine. OpenRefine was used to clean the data; countries and genres were split on separators and filled down. RStudio was used to create, analyze, and run the script and generate visualizations. RMarkdown was used to create an HTML document. Shiny was used to create a series of stories within a dashboard to display the different questions and visualizations from the analysis.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### Questions

1. What is the correlation between runtime and country? What is the relationship between runtime and type of production?

2. Which countries are the most represented on Netflix? Of the most represented countries, what is the split between TV shows and movies?

3. Which countries are represented in co-productions? Which type of production is co-produced the most?

4. How long do TV shows run on Netflix? What is the correlation between number of seasons, genre, and country?

5. What is the correlation between IMDB popularity ratings with Netflix release year? How often are Netflix productions reflected in IMDB’s Top Rated?

# 1. Country of Origin and Runtime

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

There are 110 countries represented within Netflix's listings, with a wide variety of movies from each nation. There is an outlier in the dataset that had a null classification of country. Despite assumptions to the contrary, there is no apparent correlation between the runtime of movies and their country of origin.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 1

```{r include=FALSE, echo=FALSE}

movies <- filter(net, type == "Movie")

movies$runtime <- gsub("min","", movies$duration)

as.numeric(movies$runtime)

movies\_cleaned <- movies[order(movies$runtime),]

```

```{r echo=FALSE}

# 1

# scatter

ggplot(movies\_cleaned,aes(x=country, y=runtime, color=country, size=duration)) +

geom\_point() +

labs(title="Netflix Movie Origins",

x="Country", y="Runtime") +

theme(legend.position="none",

axis.text.x = element\_text(

size=6, angle=45),

axis.text.y = element\_text(

size=4.5))

```

# 2. International Representation

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

There are 110 countries represented within Netflix's listings, with a wide variety of both TV shows and movies. There is an outlier in the dataset that had a null classification of country. For the ten most represented countries on the platform, more movies are present on the platform than TV shows.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 2a. What are the top-ten most represented countries on Netflix?

```{r include=FALSE, echo=FALSE}

country\_count <- count(net, country, sort = TRUE)

country\_count\_sort <- country\_count [1:10,]

country\_count\_sort

```

```{r echo=FALSE, results='asis'}

library(knitr)

kable(country\_count\_sort)

```

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 2a

```{r echo=FALSE}

# 2a

ggplot(country\_count\_sort, aes(fill=country, y=n, x=country))+

geom\_bar(postion="dodge", stat="identity")+

labs(title="Top Ten Most-Represented Countries on Netflix", x="Country", y="Amount of Media")+

scale\_fill\_manual("country", values = c("Canada" = "red", "France" = "blue", "Germany" = "green", "India" = "orange1", "Japan" = "purple", "Mexico" = "darkgreen", "South Korea" = "brown4", "Spain" = "yellow", "United Kingdom" = "grey50", "United States" = "navy"))

```

# International Representation

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

There are 110 countries represented within Netflix's listings, with a wide variety of both TV shows and movies. There is an outlier in the dataset that had a null classification of country. For the ten most represented countries on the platform, more movies are present on the platform than TV shows.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 2b. Of these countries, what is the split between TV shows and movies??

```{r include=FALSE, echo=FALSE}

country\_count\_type <- count(net, country, type, sort = TRUE)

country\_count\_split <- merge(country\_count\_sort, country\_count\_type, by = c("country"))

country\_count\_split

```

```{r echo=FALSE, results='asis'}

library(knitr)

kable(country\_count\_split)

```

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 2b

```{r echo=FALSE}

# 2b

ggplot(country\_count\_split, aes(fill=country, y=n.y, x=type))+

geom\_bar(postion="dodge", stat="identity")+

labs(title="Movies and TV Shows of the Top Ten", x="Type", y="Amount of Media")+

scale\_fill\_manual("country", values = c("Canada" = "red", "France" = "blue", "Germany" = "green", "India" = "orange1", "Japan" = "purple", "Mexico" = "darkgreen", "South Korea" = "brown4", "Spain" = "yellow", "United Kingdom" = "grey50", "United States" = "navy"))

```

# 3. International Co-Production

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

There are 659 unique productions that include multiple countries having produced a TV Show or Movie together. There is an outlier in the dataset that had a null type of TV Show or Movie. Out of the previous productions that are co-produced, Netflix contains more movies that are co-produced among multiple nations.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 3a. How many listings are international co-productions?

```{r include=FALSE, echo=FALSE}

class\_count <- dplyr::count(netflix, type, country)

class\_count

```

```{r echo=FALSE}

ggplot(class\_count, aes(x=type, y=n)) +

geom\_bar(stat = "identity") +

labs(title="Netflix Movies and TV Shows",

x="Type", y="Count")

```

```{r echo=FALSE}

# 3a

ggplot(net, aes(fill=type, y=country, x=type)) +

geom\_bar(position="dodge", stat="identity") +

labs(title="Netflix Movies and TV Shows",

x="Type of Production", y="Country of Origin")

```

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 3b. Do movies or TV shows get co-produced the most?

```{r include=FALSE, echo=FALSE}

class\_count <- dplyr::count(netflix, type, country)

class\_count

tr <- class\_count %>%

filter\_all(any\_vars(grepl(',', .)))

```

```{r echo=FALSE}

ggplot(tr, aes(x=type, y=n)) +

geom\_bar(stat = "identity") +

labs(title="Netflix Co-Produced Movies and TV Shows",

x="Type", y="Count")

```

# 4. Netflix Shows

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

Most shows that Netflix offers are found to only have one or two seasons at most. The data did not take into account whether these shows were created for Netflix, are being hosted on Netflix, if the shows have been cancelled, or if the shows are currently in production or have been greenlit for further production.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 4a. How long do most shows run on Netflix?

```{r include=FALSE, echo=FALSE}

runn <- filter(netflix, type == "TV Show")

```

```{r echo=FALSE}

ggplot(runn, aes(x=duration)) +

geom\_bar(fill="#69b3a2", color="#69b3a2") +

labs(title="Netflix TV Show Seasons",

x="Number of Seasons", y="Number of Shows") +

theme(axis.text.x = element\_text(

size=10, angle=45),

axis.text.y = element\_text(

size=8))

```

# Netflix Show Seasons

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

The genres of shows that Netflix hosts and produces are a wide range.They also differ widely across the number of seasons for each show. The first graph displays the more popular genres at the bottom. The thickness of the bar graph shows the amount of shows within that genre. Genres that do not have many seasons can be viewed at the top of the graph and the thickness of the individual bars.

The second graph shows the individual countries that host or produce content for Netflix. Many countries only produce a small number of seasons for a show, which can be viewed as the thinner portions of the colored bars in the graph. The countries that produce the most content are found along the bottom of the graph.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 4b. Is there a correlation between show seasons and genre?

```{r include=FALSE, echo=FALSE}

genre <- filter(net, type == "TV Show")

```

```{r echo=FALSE}

ggplot(genre, aes(fill=duration, y=listed\_in, x=duration)) +

geom\_bar(position="dodge", stat="identity") +

labs(title="Netflix TV Show Genres",

x="Number of Seasons", y="Genres") +

theme(axis.text.x = element\_text(

size=10, angle=45),

axis.text.y = element\_text(

size=8))

```

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 4c. Is there a correlation between show seasons and country of origin?

```{r include=FALSE, echo=FALSE}

origin <- filter(net, type == "TV Show")

```

```{r echo=FALSE}

# bar graph

ggplot(origin, aes(fill=duration, y=country, x=duration)) +

geom\_bar(position="dodge", stat="identity") +

labs(title="Netflix TV Show Countries",

x="Number of Seasons", y="Country") +

theme(axis.text.x = element\_text(

size=10, angle=45),

axis.text.y = element\_text(

size=6))

```

# Netflix Countries

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

Netflix hosts and produces many TV shows across the world. A large portion of the shows produced only have a few seasons. The graph depicts how many seasons of a show a country produces. The more size of the points corresponds to the number of shows that country may have--the smaller the point, the greater the number of shows; the larger the point, the fewer the number of shows. Very few countries produce content after the first five seasons of a show.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### 4c.

```{r echo=FALSE}

# scatter

ggplot(origin,

aes(x=duration,

y=country,

color=country,

size=duration)) +

geom\_point() +

labs(title="Netflix TV Show Origins",

x="Number of Seasons", y="Country") +

theme(legend.position="none",

axis.text.x = element\_text(

size=6, angle=45),

axis.text.y = element\_text(

size=4.5))

```

# 5. Netflix and IMDB

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

Out of both TV Shows and Movies that Netflix hosts on their service, these are the only titles in the right table that have been rated to be within the Top 10 on IMDB.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### Top 10 TV Shows on IMDB

```{r include=FALSE, echo=FALSE}

title <- c("Planet Earth II", "Planet Earth", "Band of Brothers", "Breaking Bad", "Chernobyl", "The Wire", "Blue Planet II", "Our Planet", "Cosmos: A Spacetime Odyssey", "Cosmos")

type <- c("TV Show", "TV Show", "TV Show", "TV Show", "TV Show", "TV Show", "TV Show", "TV Show", "TV Show", "TV Show")

release\_year <- c(2016, 2006, 2001, 2008, 2019, 2002, 2017, 2019, 2014, 1980)

rated <- c(9.5, 9.4, 9.4, 9.4, 9.3, 9.3, 9.3, 9.3, 9.2, 9.2)

from <- c("IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB")

imdb\_show <- data.frame(title, type, release\_year, rated, from)

```

```{r echo=FALSE, results='asis'}

library(knitr)

kable(imdb\_show)

```

<br>

<br>

<br>

<br>

### Top 10 Movies on IMDB

```{r include=FALSE, echo=FALSE}

title <- c("The Shawshank Redemption", "The Godfather", "The Godfather: Part II", "The Dark Knight", "12 Angry Men", "Schindler's List", "The Lord of the Rings: The Return of the King", "Pulp Fiction", "The Good, the Bad and the Ugly", "The Lord of the Rings: The Fellowship of the Ring")

type <- c("Movie", "Movie", "Movie", "Movie", "Movie", "Movie", "Movie", "Movie", "Movie", "Movie")

release\_year <- c(1994, 1972, 1974, 2008, 1957, 1993, 2003, 1994, 1966, 2001)

rated <- c(9.2, 9.1, 9.0, 9.0, 8.9, 8.9, 8.9, 8.8, 8.8, 8.8)

from <- c("IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB", "IMDB")

imdb\_movie <- data.frame(title, type, release\_year, rated, from)

```

```{r echo=FALSE, results='asis'}

library(knitr)

kable(imdb\_movie)

```

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### Netflix Movies and TV Shows in IMDB's Top 10

```{r include=FALSE, echo=FALSE}

finaldt <- rbind(imdb\_show, imdb\_movie)

```

```{r echo=FALSE, results='asis'}

new <- merge(finaldt, netflix, by = c("title", "type", "release\_year"))

newer <- select(new, -c("from", "show\_id", "cast", "date\_added", "description"))

```

```{r echo=FALSE, results='asis'}

library(knitr)

kable(newer)

```

# References

Inputs {.sidebar}

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

The dataset focused on Netflix’s current holdings—as of 2020—in the data fields of title, release date, country, genre the production was listed in, type of production, runtime, maturity rating, cast, producer, and description.

The original dataset did not contain viewer ratings. In order to further explore the dataset, two smaller datasets were created tabularly and introduced—IMDB’s Top 10 Movies, and IMDB’s Top 10 TV Shows.

Row

-----------------------------------------------------------------------

<br>

<br>

<br>

<br>

### References

Bansal, S. (2020). Netflix. Kaggle. https://www.kaggle.com/shivamb/netflix-shows

IMDB. (2021). IMDb top 250 movies. IMDB. https://www.imdb.com/chart/top?pf\_rd\_m=A2FGELUUNOQJNL&pf\_rd\_p=4da9d9a5-d299-43f2-9c53-f0efa18182cd&pf\_rd\_r=T4YW4W6DRTV4R7PN4SMA&pf\_rd\_s=right-4&pf\_rd\_t=15506&pf\_rd\_i=toptv&ref\_=chttvtp\_ql\_3

IMDB. (2021). IMDb top 250 tv shows. IMDB. https://www.imdb.com/chart/toptv?pf\_rd\_m=A2FGELUUNOQJNL&pf\_rd\_p=4da9d9a5-d299-43f2-9c53-f0efa18182cd&pf\_rd\_r=6BVQC60H3T168VTH020K&pf\_rd\_s=right-4&pf\_rd\_t=15506&pf\_rd\_i=top&ref\_=chttp\_ql\_6

**References**

Bansal, S. (2020). Netflix. *Kaggle.* https://www.kaggle.com/shivamb/netflix-shows

IMDB. (2021). IMDb top 250 movies. *IMDB*.

https://www.imdb.com/chart/top?pf\_rd\_m=A2FGELUUNOQJNL&pf\_rd\_p=4da9d9a5-d299-43f2-9c53-f0efa18182cd&pf\_rd\_r=T4YW4W6DRTV4R7PN4SMA&pf\_rd\_s=right-4&pf\_rd\_t=15506&pf\_rd\_i=toptv&ref\_=chttvtp\_ql\_3

IMDB. (2021). IMDb top 250 tv shows. *IMDB*.

https://www.imdb.com/chart/toptv?pf\_rd\_m=A2FGELUUNOQJNL&pf\_rd\_p=4da9d9a5-d299-43f2-9c53-f0efa18182cd&pf\_rd\_r=6BVQC60H3T168VTH020K&pf\_rd\_s=right-4&pf\_rd\_t=15506&pf\_rd\_i=top&ref\_=chttp\_ql\_6

Keelery, S. (2020, September 30). *Topic: Film Industry in India*. Statista.

https://www.statista.com/topics/2140/film-industry-in-india/

Pineda, R. A. (2020, June 8). *Japan’s Internal Affairs Ministry: Anime Comprises 80% of*

*Broadcast Exports—News—Anime News Network*. Anime News Network.

https://www.animenewsnetwork.com/news/2020-06-08/japan-internal-affairs-ministry-anime-comprises-80-percent-of-broadcast-exports/.160362

Szalai, G. (2017, March 7). BBC Worldwide, ITV SVOD Venture BritBox Launches With $6.99

Monthly Price. *The Hollywood Reporter*.https://www.hollywoodreporter.com/tv/tv-news/bbc-worldwide-itv-svod-venture-britbox-launches-699-monthly-price-983887/