# Comparing the Content Strategy of Leading Streaming Platforms: Netflix and Disney+ BUS 314 Final Project Alexis Park

### Summary of Research Questions and Results

#### RQ1: On both Netflix and Disney+ platforms, in which countries are most movies/TV shows produced in?

Understanding this information can help us visualize and examine the platforms’ content strategy and preferences towards particular regions. Additionally, it can provide insight into the diversity of the platforms’ content and global reach.

The analysis showed that outside the U.S., top producers of content on Disney+ and Netflix include India, the United Kingdom, South Korea, and Japan.

#### RQ2: How many movies/TV shows were added throughout the years to Netflix and Disney+ platforms?

Asking this question can help us better understand the platforms’ content acquisition and production strategy, the growth of the platforms, and its global expansion into non-U.S. markets.

The analysis showed that over time, more movies were added that were produced from different countries, from more genres, and from a variety of ratings.

#### RQ3: Is there a relationship between duration and release year?

The visualizations produced in response to this question will give us insight into duration trends throughout the time-period of our datasets to help with understanding whether content has shortened, lengthened, or remained constant over the lifespan of the dataset. Potential reasons why it may be important to know this relationship, from a business perspective, is that duration can have a direct impact on production and other associated costs, as longer TV shows and films typically require more resources and time to produce. Marketing is another factor to consider when analyzing the relationship between duration and media content, as the duration of TV shows and movies may be able to help determine target demographics and campaign strategies.

The analysis found that there appears to be a relationship between duration and release year that varies by country and by type (movie vs TV show).

#### RQ4: How do movies/TV shows that are available for streaming on both Disney+ and Netflix differ, if they do?

The answer to this question can provide insights into how each platform rates their content to better understand their audience preferences and types of content that appeals to their consumers.

The analysis found that there are indeed differences on the same movie/TV show depending on the streaming platform, such as how it is rated, the duration, and when it was added to the streaming service.

**Motivation and Background**

With the significant rise in popularity of online streaming platforms like Netflix and Disney+, there is a growing need to understand their content strategy and audience preferences. In fact, Netflix and Disney+ make up 2 out of the top 3 U.S. video streaming app market share in 2021 and 2022, according to JustWatch data. This comparative analysis between Netflix and Disney+ can help us from a business perspective better understand the differences between two competitors’ marketing strategies, leading to improved market positioning and informed decision-making when it comes to content creation and distribution.

RQ1 can shed light on the platforms' global reach and diversity of their content, as the number of shows/movies produced in different regions can reveal each platform's preferences and interests. This information is crucial for content creators, as they can use it to tailor their content to a specific region, thereby increasing its chances of being accepted by these platforms.

RQ2 is crucial in understanding the growth of each platform and how it has expanded globally. This information can be used to predict future growth patterns and to identify areas where each platform may need to improve. Knowing how many movies and shows were added each year can also help movie creators understand the growing demand for TV and film.

RQ3 aims to understand if there is a relationship between the duration of movies/TV shows and their release year. This information can be useful in predicting future trends in movie and TV show durations, which can help media companies tailor their productions to fit the expected demand. It can also reveal how the content industry has evolved over time and how changes in technology and consumer preferences have affected movie and TV show durations.

Finally, RQ4 provides insights into each platform's audience preferences and the types of content that appeal to their viewers by seeing how they differ depending on platform. Media companies can use this information to create content that has a higher chance of being accepted by these platforms and to cater to the specific preferences of their target audience.

**Datasets**

The datasets we used are from kaggle.com, one consisting of listings of all movies and TV shows on Netflix ([available here](https://www.kaggle.com/datasets/shivamb/netflix-shows)) and the other consisting of listings of all movies and TV shows on Disney+ ([available here](https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows)), two popular media and video streaming platforms with millions of subscribers globally. The datasets were published in mid-2021, both with columns including a unique ID, type (movie or show), title, director, cast, country, date added, release year, rating, and duration.

### Ethical Considerations of Data

* Bias: As Disney+ and Netflix are American companies, there may be a bias of adding shows from countries that the United States have a positive relationship with.
* Historical considerations: Disney in the past was known for incorporating American propaganda into its movies. There could be a valid concern of whether the movies/TV shows added by Disney+ or Netflix incorporates American propaganda to its audience members.
* Data access: The content on Disney+ and Netflix changes depending on where the user is located. The data therefore may not incorporate every single movie/TV show available to a user; or, the data may include movies/TV shows that not every single user has access to without a VPN.
* Respect for cultural differences: When asking RQ1, RQ2 and RQ3, it is important to recognize that each country has its unique culture and ways of expressing themselves through media. The questions should not be asked in a way that promotes any bias towards one region over another.
* Privacy and data protection: It is crucial to ensure that the data being collected and analyzed is done so ethically and in compliance with data protection laws. Personal information should not be collected or shared without the user's consent.
* Transparency and fairness: When asking RQ4, it is important to be transparent about how content is rated and what criteria is being used. The rating system should be fair and unbiased, and not influenced by any external factors.
* Inclusivity and representation: In both RQ1 and RQ4, it is important to ensure that the analysis takes into account the diverse representation of people and cultures. The content should reflect the diversity of the audience it serves, and the analysis should not exclude or marginalize underrepresented groups.

### Methodology

This analysis aims to explore the content strategy and preferences of leading streaming platforms, Netflix and Disney+. Our research questions seek to understand the countries with the highest rate of production for movies and TV shows, the number of movies and shows added to each platform over recent years, and how Netflix and Disney+ rate the same movie/TV show. This information can help streaming platforms make better informed decisions regarding content acquisition, production, and creation.

We used two data sets for this analysis, one for each online streaming platform, and were analyzed to understand content strategy and audience preferences. The methodology used involved analyzing the datasets and creating visualizations to understand the relationships between production rate, content acquisition, and ratings. These visualizations include a world map to visualize in which countries, and how much, content was created, bar and line graphs to visualize how many movies and TV shows were produced in the top producing countries of these platforms, line graphs that depict movie and TV show trends, pie charts that display the proportion of titles with the same rating across both platforms, and bar graphs that that show the differences in when titles across both platforms were added and their duration.

**Results**

RQ1:

##### Map of Countries Most Movies/TV Shows were Produced In

Background pattern

Description automatically generated with low confidence

Given that both Netflix and Disney+ are United States based video streaming platforms, it makes sense that there is a significant portion of movies/TV shows added from the United States. Therefore, we filtered out the United States to see what other countries have a significant presence on these streaming platforms. Noticeably from this map, there are significant movies/TV shows produced in India, the United Kingdom, and East Asia (Japan and South Korea). We deduce that media from these regions are popularly consumed by video streaming platforms' audiences, leading to more media to be added to Disney+ and Netflix from these regions.

##### % Difference from Mean Movies/TV Shows from Top 10 Countries

Chart, histogram

Description automatically generated

These visualizations also help answer RQ1. From these graphs, we are more explicitly able to see the diversity of international movies and TV shows on Netflix and Disney+. In particular, there are more than the average number of movies produced in India and the United Kingdom among the top 10 countries; on the other hand, there are more than the average number of TV shows produced in the United Kingdom, Japan, and South Korea among the top 10 countries. This tells us that certain countries produce media tailored to movies or more tailored towards TV shows.

##### RQ2:

##### Line Graphs of the Number of Movies/TV Shows Added to Platforms per Country

Chart, line chart

Description automatically generated

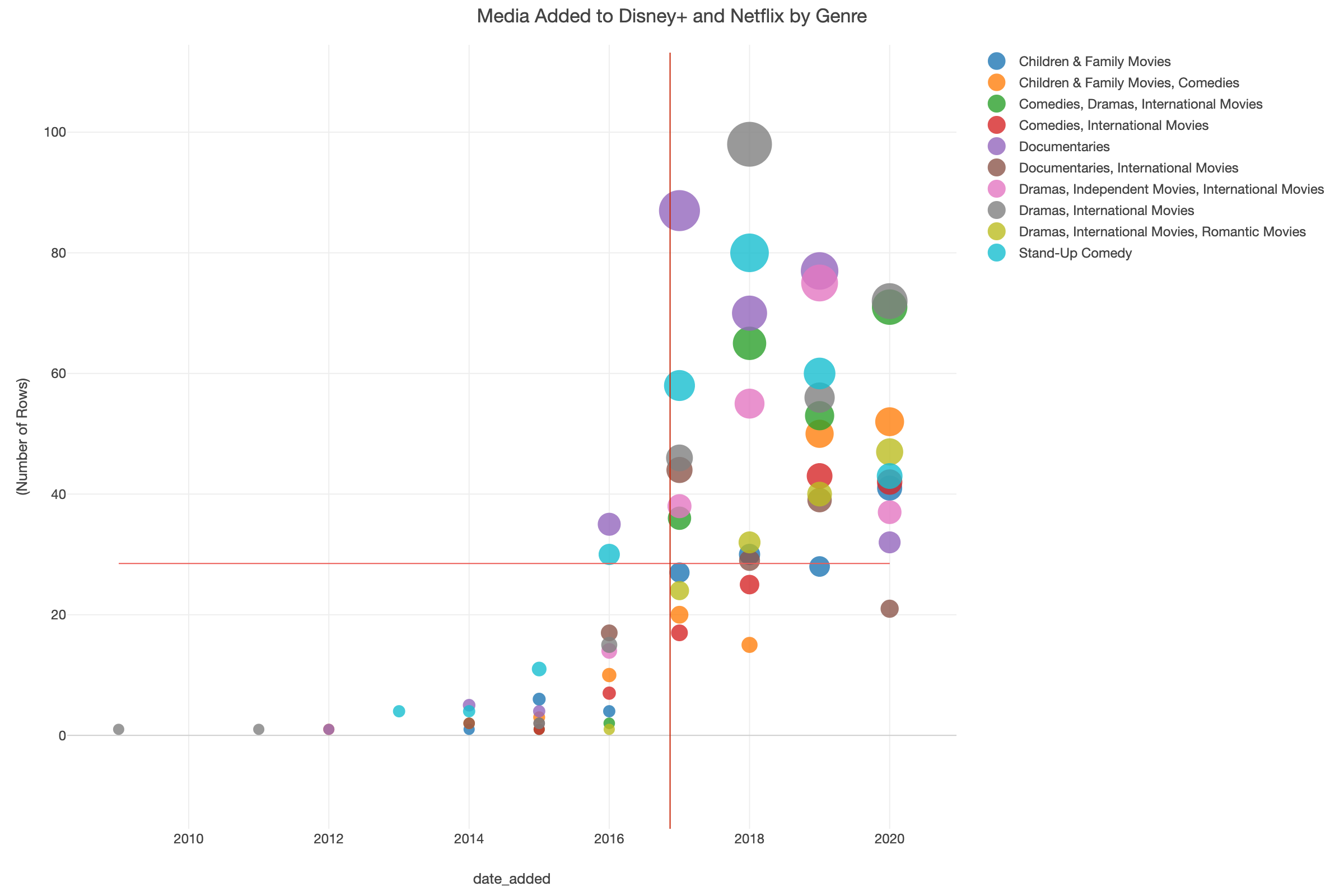
This visualization helps us answer RQ2. Throughout the years that movies/TV shows were added to Netflix and Disney+ by country, we found that in general, more movies were added, as seen by the larger range in Y Axis for the movie graph. We decided to not sync the axes so that we could better visualize the trends in TV show acquisition. We also have to keep in mind that this dataset does not contain all of the movies/TV shows that were added for 2021, leading to us deciding to filter out the data from 2021 as it would lead to most of these graphs to show a downward trend in 2021.

One finding from this data that we wanted to point out was with Japan and South Korea, two of the top 3 countries that produced TV shows that were added to Disney+ and Netflix the most. Given the audience of Japanese and South Korean movies/TV shows, these results make sense: Japanese anime series and South Korean drama series (K-dramas) are more popularly consumed over Japanese and South Korean movies, according to one Nielsen study. Therefore, it makes sense for Netflix and Disney+ to acquire TV shows from these countries over movies to better cater to their audiences.

We also see that among the top 10 countries that media was added from, not every country had an upward trend in terms of media produced from it being added to Netflix and Disney+. For example, TV shows produced in Taiwan and China decreased in 2020, as well as movies produced in India.

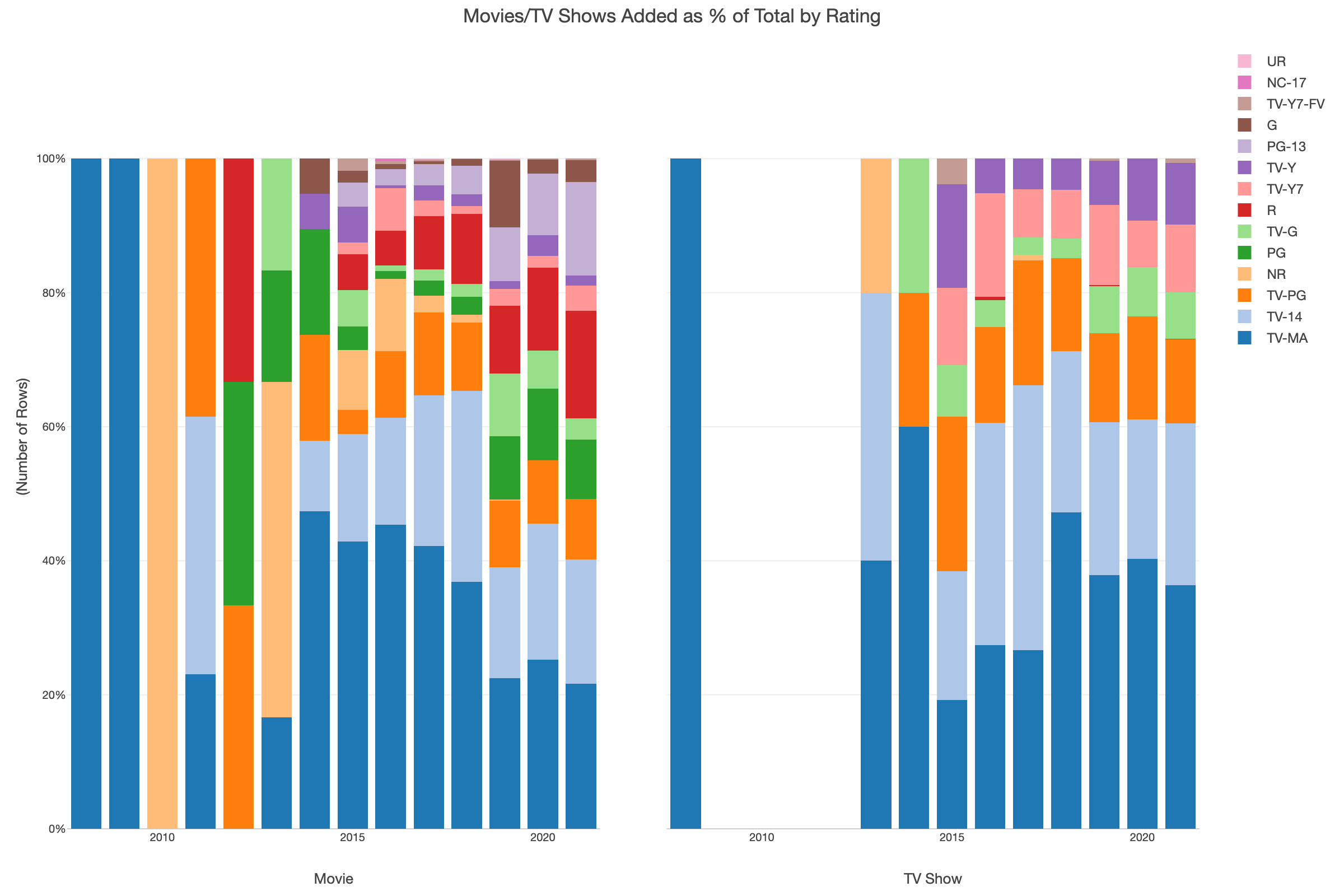
Generally, we see that there is an upward trend in movies added that are produced in foreign countries, and a mix in positive and negative trends in TV shows added that are produced in foreign countries. We deduce from these visualizations that Netflix and Disney+ are likely expanding their platforms to audiences from countries where they are acquiring more movies/TV shows.

##### Scatterplot of Media Added to Disney+ and Netflix by Genre



This visualization shows how many movies/TV shows were added to Disney+ and Netflix by genre rather than by country, and we included the U.S. in this visualization. Notice how even with data on movies/TV shows produced in the U.S. was included, there is still a significant number of content of the "Dramas, International Movies" category added to Netflix and Disney+. Media from this category was added more recently, as seen by more gray circles being seen after the vertical red reference line. Other visually noticeable genres include "Stand-Up Comedy," "Documentaries," "Comedies, Dramas, International Movies," and "Dramas, Independent Movies, International Movies." The general trend we see is that media from most genres were added to a greater extent each year as seen by how most circles “shift” from the bottom left quadrant to the top right quadrant.

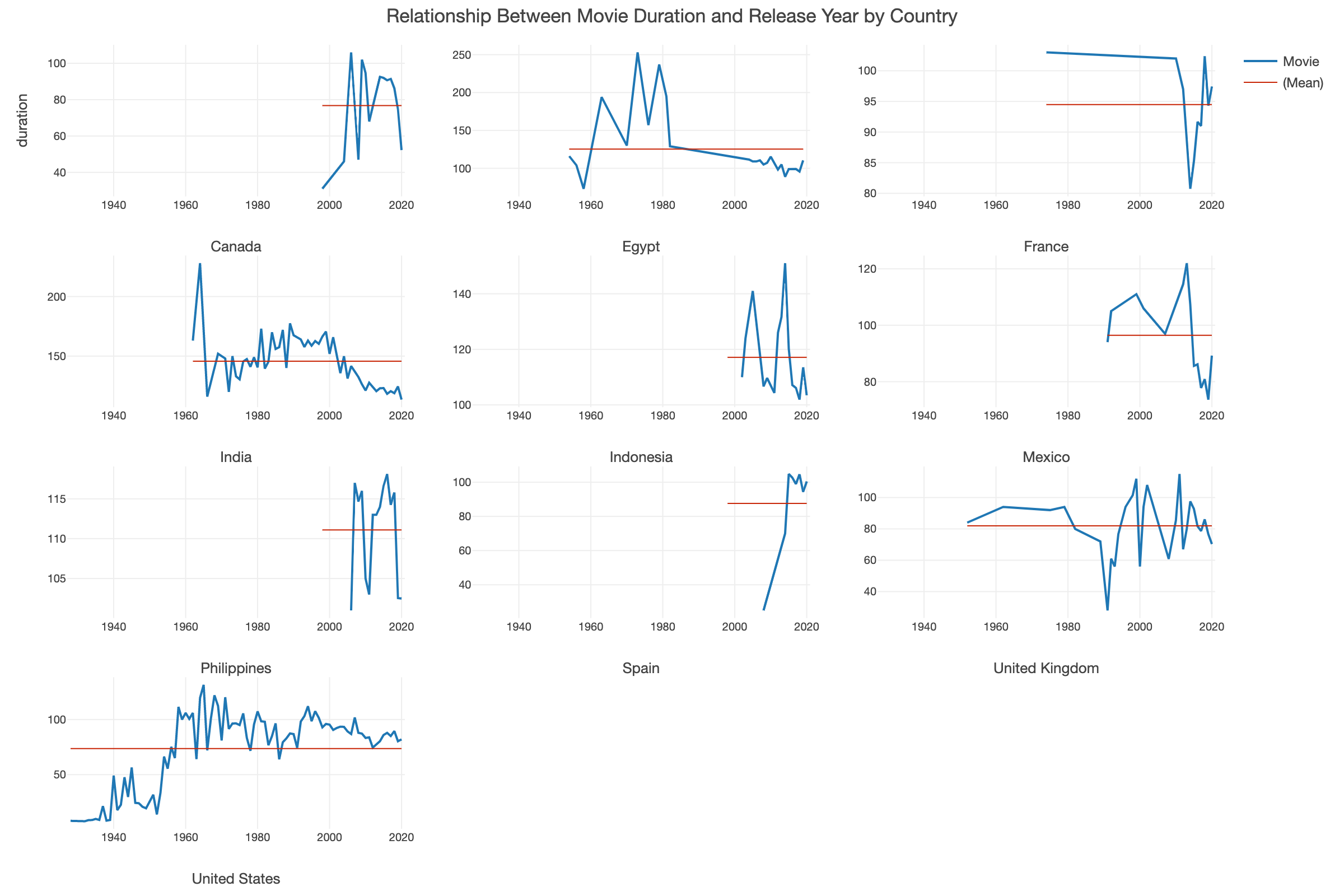
##### Bar Graph of Media Added to Disney+ and Netflix by Rating



This visualization helps us see the distribution of ratings of all movies/TV shows added to Disney+ and Netflix until 2020, and how this distribution changes over time. In the Movies graph, notice how only movies rated TV-MA was added until 2010. Subsequent years show a decline in the percentage of TV-MA movies added as a total of all movies added to Disney+ and Netflix. More PG and TV-14 rated movies were added as a percentage of all movies added, in addition to many other rated movies, showing how Disney+ and Netflix added a larger variety of movie content to their platforms. On the other hand, TV shows added to Disney+ and Netflix shows a different trend. All TV shows that were added in 2008 were rated TV-MA, and then no TV shows were added until 2013. At this point, Disney+ and Netflix were adding a more divese range of rated TV shows. However, notice how there are less colors in the TV shows graph compared to the movies graph. This shows that there is less of a variety in ratings in added TV shows compared to added movies, potentially because TV shows are more recently popular compared to movies.

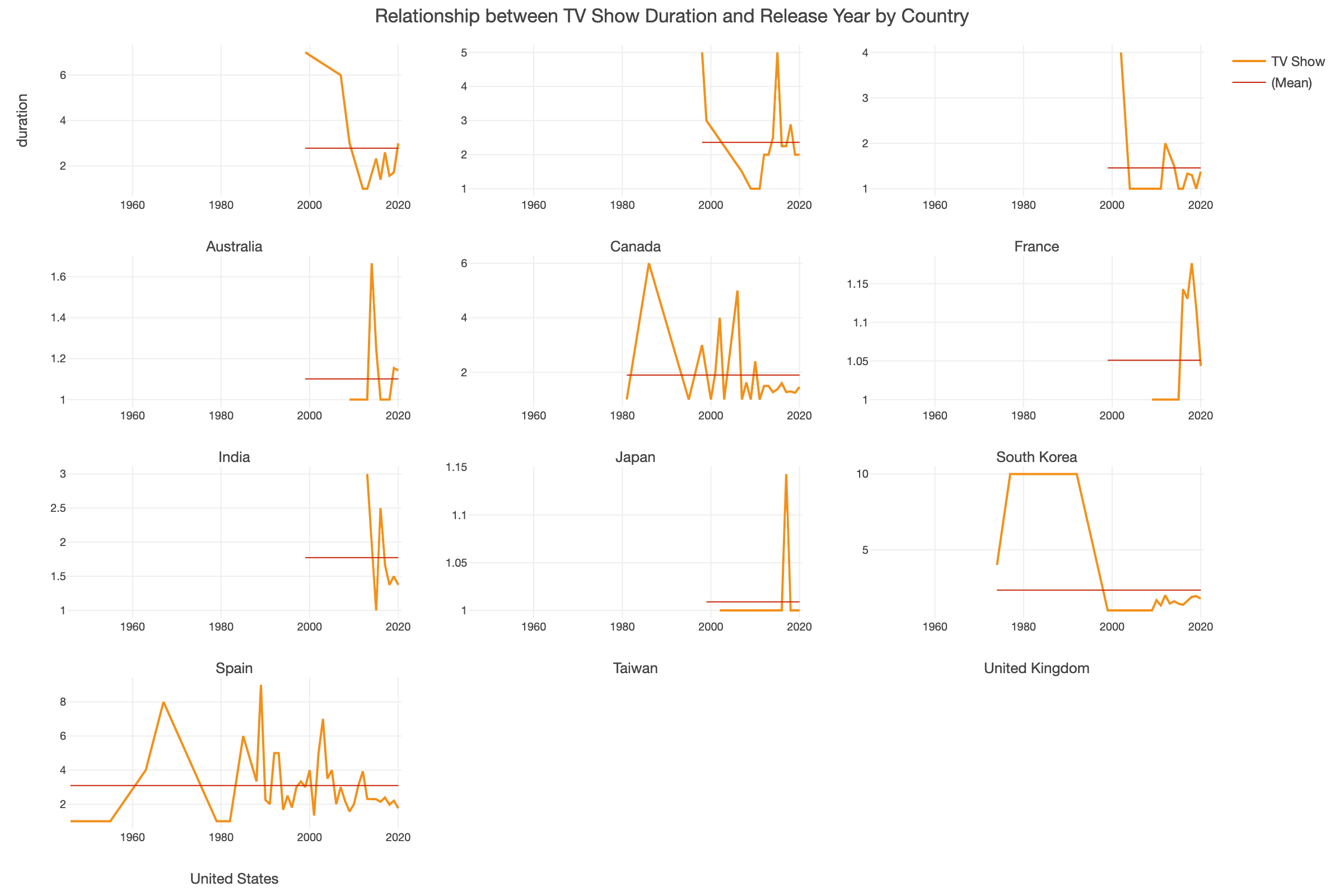
RQ3:

##### Line Graphs of Movie Duration vs Release Year by Country



Five out of the top 10 countries have movie duration trending down in the last year of data (United Kingdom, Philippines, Indonesia, Canada, and India) while the other five are trending up in the last year of data (Egypt, France, Mexico, United States, and Spain). 6 of the top 10 have average movie durations under 100 minutes (United States, United Kingdom, Spain, Mexico, France, and Canada) while the other four have average movie durations over 100 minutes (Egypt, India, Indonesia, and Philippines). United States has lowest average movie duration of its releases with 75.69 minutes, whereas India has the highest average movie duration of its releases with 144.82 minutes. For some of these countries that have averages of under 100 minutes, such as United States which have long histories of releasing movies, this could be due to the large amounts of short films that used to be released when movies first began to be produced which would depress the all-time average. 7 of the 10 countries have been releasing movies since before the year 2000 in this dataset (United States, United Kingdom, Canada, France, India, Mexico, and Egypt) while the rest have only been releasing movies since after the year 2000 in this dataset (Indonesia, Philippines, Spain). Over a longer period of time though, most of these countries have seen a long-term increase in the average movie duration time in their releases with the exception of Canada, Egypt, India, Indonesia, and Mexico.

##### Line Graphs of TV Show Duration vs Release Year by Country



5 of the top 10 countries have TV Show duration trending down in the last year of data (South Korea, Spain, United Kingdom, India and United States), 3 are trending upwards in the last year of data (Australia, France, and Japan), and 2 remained constant in the last year (Canada and Taiwan). Out of all of the countries, only the United States has an average TV Show duration of over 3 seasons (3.49) while the rest sit between 1.01 (Taiwan) and 2.66 (Australia). Another important note is that only 3 countries (United States, United Kingdom, and Japan) have data for TV Show duration from before the year 2000. The rest only have data from the year 2000 or later. This shows that some countries have been releasing TV Shows for much longer periods of time than others.

RQ4:

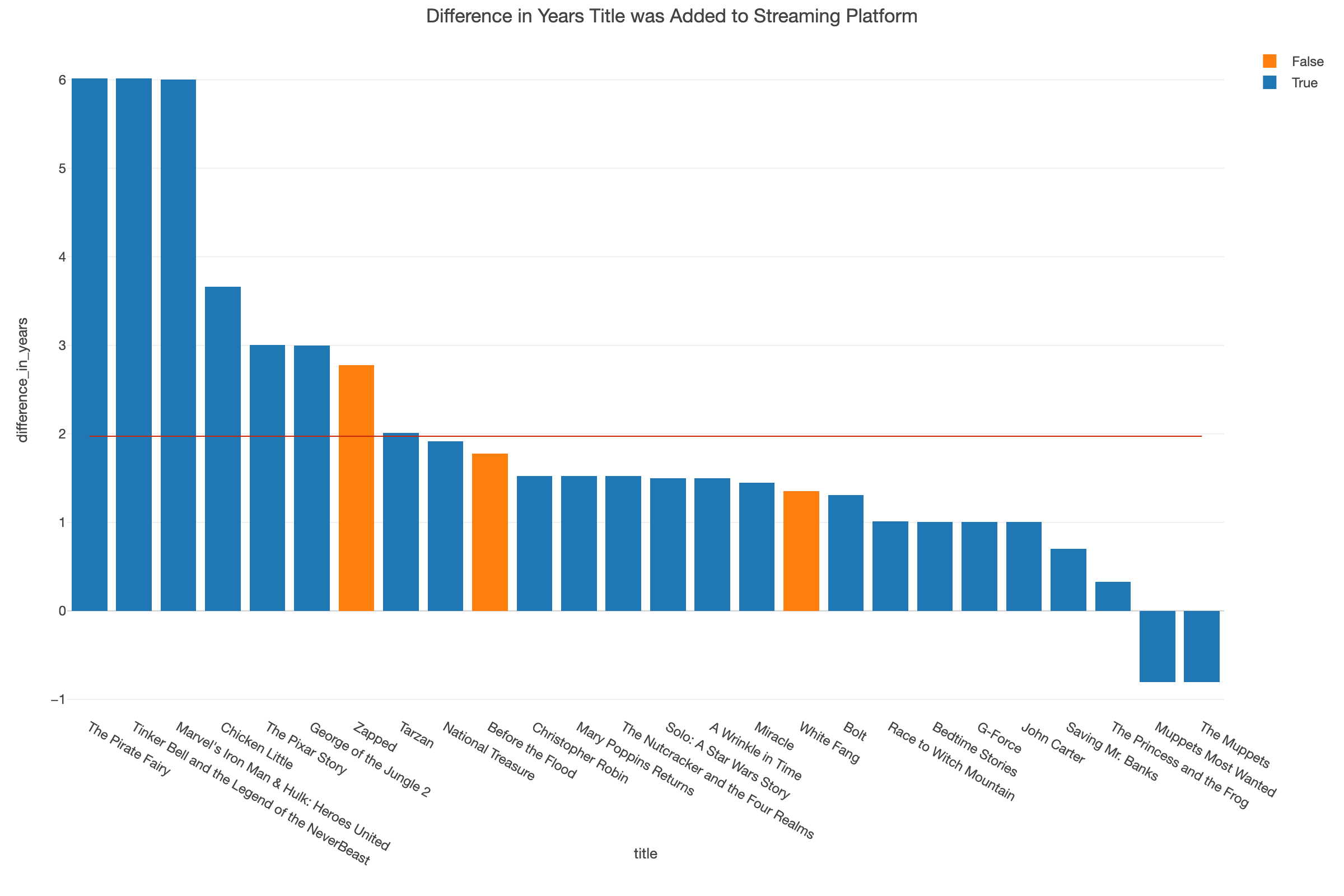
##### Pie Chart of Percentage of Same Ratings

Shape, circle

Description automatically generated

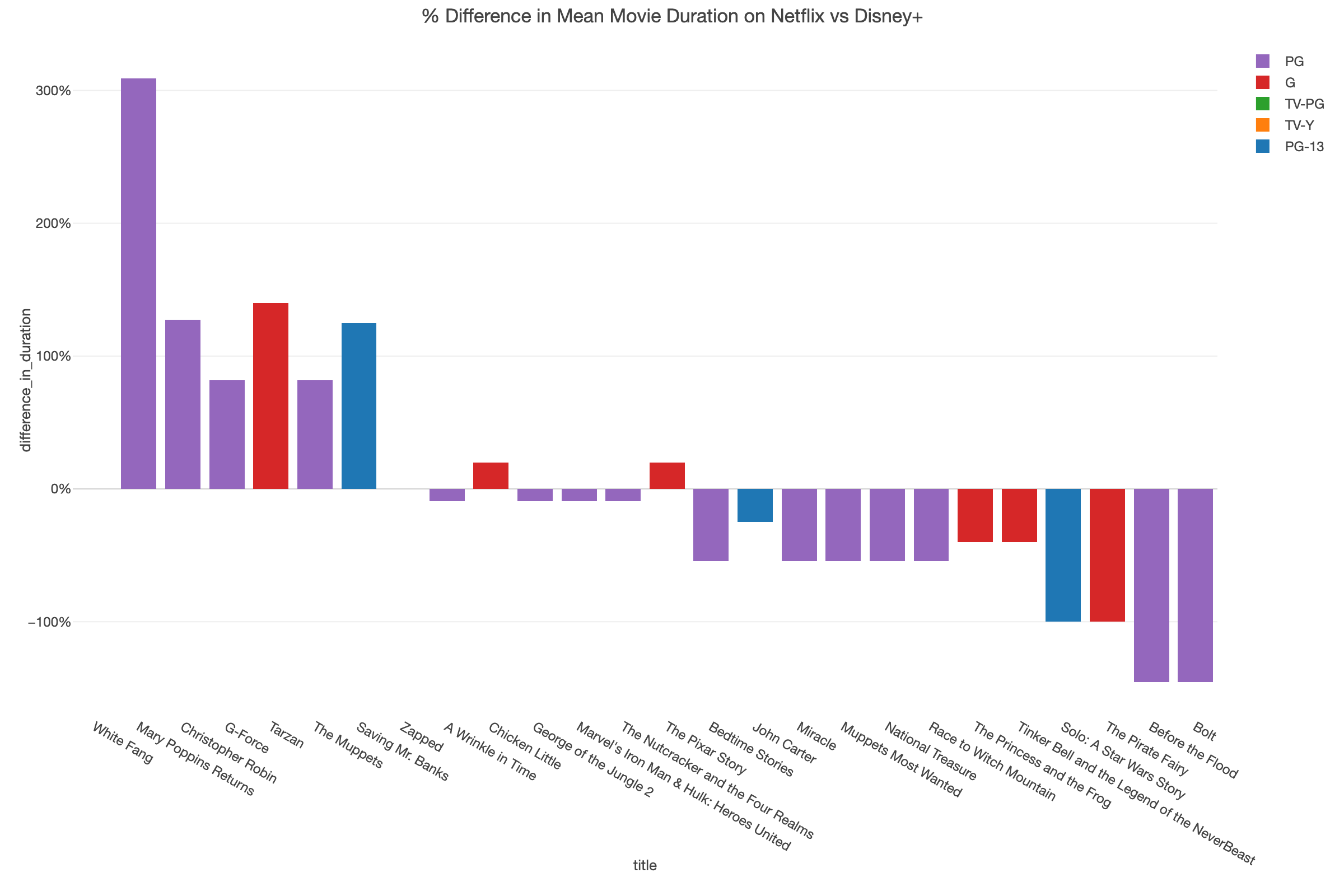
This visualization aids in answering RQ4 by revealing the ratio of ratings that are the same or not the same among the titles that are available on both Disney+ and Netflix, grouped by each rating. We can see that the movies/TV shows that were not rated the same were those rated as PG, TV-14, TV-G, TV-PG, and TV-Y7 by Disney+. Movies/TV shows that were rated the same were those rated as G, PG-13, and TV-Y by Disney+ and Netflix. This visualization reveals that not all movies/TV shows are rated the same on Disney+ and Netflix. It also shows that the movies/TV shows that are available for streaming on both Disney+ and Netflix are generally catered to a younger audience, as most of these ratings are less strict (G, PG, TV-Y, etc.).

##### Bar Chart of Difference in When Title was Added



From this visualization, we can see that there is mostly a positive difference in when the title was added. This means that most titles were added to Netflix later than Disney+. The only two titles that were added to Netflix earlier than Disney+, as depicted by a negative difference in years, are *The Muppets* and *Muppets Most Wanted.* The average difference in years of when a title was added between Netflix and Disney+ was 2 years, in which most titles have less than this average difference. Additionally, all titles that don't have the same rating were added to Netflix later than Disney+.

##### % Difference in Mean Movie Duration



Lastly, this visualization shows us that even with the same movie, there is a difference in movie duration. There doesn't appear to be a specific relationship between rating and difference in mean duration. % Difference in mean movie duration is larger for Disney+ (as seen by the lengths of the upward bars), but there are more movies that are longer on Netflix (as seen by more quantity of downward bars).

### Reproducing Results

We will have a total of 4 datasets: the original Netflix dataset, the original Disney+ dataset, the merged dataset, and the inner joined dataset. After clicking the black “Download” button on the upper right hand corner of each dataset link, import the csv files (remember to unzip the downloaded files!). To do so,

* Click the + sign next to Data Frames.
* Import the data as file data by clicking “File Data” then “Text File (CSV, Deliminited)."
* Check both “disney\_plus\_titles.csv” and “netflix\_titles.csv” and click “Import” then “OK for All.”
* Now, do this again, but this time click “Import & Merge,” then “Save,” and rename this dataframe as “merged\_titles.”
  + Now you should have 3 datasets under Data Frames. We can merge these files in this manner because all of the columns are the same.

To create the inner join dataframe,

* Create a copy of the disney\_plus\_titles.csv dataframe by clicking the hamburger menu then “Duplicate.”
* Rename this dataframe as “inner\_join\_titles” then click “OK.”
* On the summary view of this duplicated dataframe, click the down arrow next to “title,” then “Join (Add Columns).”
* Under Join Type, click “Inner Join (Keep only matching rows)” and select “netflix\_titles” as the Target Data Frame.
* Then under Current Column, select “title” for both dropdown menus. Then hit Run.
  + You should now have 4 different dataframes.

We will be working with the merged\_titles dataframe and the inner\_join\_tites dataframe for our visualizations.

Notice how in the Summary view under the date\_added column (or date\_added.x, date\_added.y in the inner join dataframe), the date is in character format. To change this to Date/Time,

* Click the down arrow next to date\_added
* Hover over “Change Data Type” then “Convert to Date / Time”
* Select “Month Day Year.”
* We will be overwriting this column, so keep that option selected, and hit Run.

Now, we will create visualizations to answer RQ1. The first visualizaiton we will create is the world map.

* Select merged\_titles under Data Frames. Go to Chart view and under Type, select “Map - Standard.”
* Under Map, select “World Countries (Atlantic Ocean Centered)” and under country select “Country.”
* Then under both Color By and Size, select “(Number of Rows).”
* Notice how the blue color scheme makes it difficult to see the differences in the number of films/shows produced in each country.
  + To change this, select the hamburger menu next to Color By and select “Color, Group, & Sort.”
  + Then under Color Palette, select “Yellow-Orange-Red,” drag the Opacity slider to 0.6, and hit Apply.

Films/shows were produced the most in the United States, which makes sense given that both of these platforms are U.S. based. To clearly see what other countries are top producers of films/shows on these streaming platforms, let’s filter out the U.S.

* Click “+ Add a new step” under Steps and then “Filter.”
* Under Column, select “country,” use the Operator “Not Equal To,” and then type “United States.”
* If prompted, select the option with just “United States” and no other listed countries. Then hit Run.
* Drag the green pin to this last step and see how the map changes. We can now see large and dark circles in India, East Asia, and the United Kingdom.
* To see these circles even more distinctly, change the map to Dark mode by clicking the gear icon and selecting "Dark" under Type in the Style section.

Secondly, to clearly see how many movies/TV shows were produced in the top producing countries of these streaming platforms, let’s create a bar chart.

* First, click the + sign under Chart view and select “Bar” under Type. Select “country” under X Axis and “(Number of Rows)” under Y Axis.
* Second, check the box next to Sort By.
* Thirdly, click the hamburger menu next to X Axis and select Limit Values.
* Under Type, select Top and type in 10 under Number of Results. Hit Apply.
* Then, under Color By and Repeat By, select "Type."
  + This will show separate graphs for movies and TV shows.
* To add a title, select the gear icon and scroll down to the Title section. Under Text, type in the title (for example, "% Difference from Mean of Movies/TV Shows Produced in Top 10 Countries").
* To show a % Difference from Mean, select the hamburger icon next to Y Axis and select "Window Calculation."
* Then select "% Difference From" under Calculation Type, select "Mean (Average)" under % Difference From, and select "Color By" under Group By.
  + Remember to check the box next to "Remove NA"
  + Then hit Apply.

Now we can clearly see the top 10 countries that movies were produced in on Disney+ and Netflix, excluding the U.S, are India, United Kingdom, Canada, Egypt, Spain, Philippines, France, Indonesia, Mexico, and Nigeria. The top 10 countries that TV shows were produced in on Disney+ and Netflix, excluding the U.S., are the United Kingdom, Japan, South Korea, India, Taiwan, Canada, France, Australia, Spain, and China. We can also see that some countries produce more movies/TV shows than average among the top 10 producing countries. For example, India and the United Kingdom are significant producers of movies, and the United Kingdom, Japan, and South Korea are significant producers of TV shows from the Disney+ and Netflix datasets.

To answer RQ2, we will create a line chart.

* Select + under Chart view and select Line under the Type dropdown menu.
* Select “date\_added” under X Axis and “(Number of Rows)” under Y Axis.
* Then select “type” under Color (Group By) and “country” under Repeat By.
* Next to Repeat By, click the hamburger menu and select Limit Values, then click “Top” under Type and enter 10 under Number of Results to show the top 10 countries.
* To see these charts without data from movies/TV shows added in 2021 (because it does not include all movies/TV shows added in 2021),
* Click + Add new step and select Filter,
* Then under Column select "date\_added," select the Operator "Is Not Equal," select "Year (e.g. 2020)" under How to Set Value, and type 2021 under Value. Then hit Run.
* Lastly, deselect Sync Y Axis located at the top of the charts.
* To add a title to the chart, select the gear icon and scroll down to the Title section. Then, under Text, type in the title (for example, “Movies/TV Shows Added by Country of Production”).

Notice how in general, of movies/TV shows that are added to Netflix and Disney+ from the top 10 countries that these movies/TV shows are produced in, most countries have more movies added. The only two countries where this isn’t the case are Japan and South Korea, which makes sense given the audience of Japanese and South Korean media (anime/K-dramas).

Next we will look at another way to answer RQ2: visualizing the trends in movies/TV Shows that were added over the years by genre. We will be visualizing this as a scatterplot.

* Select + under Chart view and select Scatterplot (With Aggregation) under the Type dropdown menu.
* Select "type" under Group by, "date\_added" under X Axis, and "# (Number of Rows)" under Y Axis.
* Next to the hamburger menu under X Axis, select Reference Line.
  + Under Reference Line Type, select "Mean (Average)," select "All" under Group By, and "Red" under Color. Then hit Apply.
  + Do the same with the hamburger menu under Y Axis.
* Under Color By, select "listed\_in" and limit the values to Top 10.
* Under Size, select "# (Number of Rows)."

From this visualization, we can see that there was a general upward trend. Generally, most genres were added to a greater extent over time.

Lastly, we will visualize the distribution in ratings of movies/TV Shows that were added over the years as a response to RQ2. We will be visualizing this as a % of Total Bar Graph.

* Select + under Chart view and select Bar under the Type dropdown menu.
* Select "date\_added" under X Axis, "# (Number of Rows)" under Y Axis, "rating" under Group By, and "type" under Repeat By.
  + To create the % of Total graph, select the hamburger menu next to Y Axis.
  + Select "Window Calculation."
  + Select "% of" under Calculation Type, "Sum (total)" under Summarize Values Using, and "X Axis" under Group By.
  + Check the box next to "Remove NA."
  + Then hit Apply.
* Make sure to check the boxes next to "Sync X Axis" and "Sync Y Axis."
* Notice how some of the genres are not actually genres. We will need to filter these out.
  + Click "+ Add a new step" and select the Filter option.
  + Select "rating" under Column and "Is Not In (Multiple Values)" under Operator.
  + Now select "66 min," "74 min," and "84 min" to filter out.
  + Hit Run.
* Add a title by clicking the gear icon and typing in your title under Text.

This visualization shows the decrease in TV-MA rated movies and TV Shows in favor for adding more diverse and more audience-inclusive ratings, such movies and TV Shows for younger audiences.

To answer RQ3, we will create the following visualizations using the merged dataset that we created in the beginning. First, we will create the visualization on the relationship between movie duration and release year by country.

* After navigating to the “merged\_titles” dataset, using the table view, click on the dropdown menu on any column and click on “Summarize (Aggregate).”
* Next, under the “Group By” section:
  + You will choose “title” first.
  + Then “release\_year” second, but after choosing this ensure that just to the right of it says “INT” for integer, and if it doesn’t click on that, showing a dropdown menu where you will choose “integer.”
  + Then, “type” goes third.
  + Then, “country” goes fourth.
  + Now, move on to the “Value section.”
  + You will choose “duration.”
  + At first, the value next to “duration” will say “UNQ,” however, we don’t want that.
    - So, click on the “UNQ” and a dropdown menu will pop up and you should choose “Max Value.”
  + Then, hit “Run” giving you a simplified table with only values you want.
* After that, select the dropdown menu on the “country” column.
  + Hover over “Filter” and then select “Remove NA.”
  + Then hit “Run” in the pop up window.
* Now, select the dropdown menu on the “duration” column.
  + Hover over “Change Data Type” and then select the “Convert to Numeric” option so that you can utilize averages as an option in this chart.
  + Next, you will then hit “Run” in the pop up window that appears.
* Next, click the dropdown menu on the “release\_year” column.
  + Hover over “Change Data Type” and then hover over “Convert to Date/Time” then select “Year” giving you a pop up window.
  + Then, hit “Run” in that window.
* To finish up the table, select the dropdown menu on the “type” column.
  + Hover over “Filter” and then select “Is Equal To…”
  + Then in the pop up window select the “Movie” option in the dropdown menu there.
  + Then hit “Run.”
* Now, to create the chart click on “Chart” at the top:
  + Drag the green pin to the last step that you did (Filtering only Movies).
  + In the left side of your screen under “Type” select “Line.”
  + Under “X-Axis” select “release\_year.”
  + Under “Y-Axis” select “duration” but then click on the “UNQ” that is just to the right of it giving you a dropdown menu where you will select “Mean (Average).”
  + Then click on the hamburger menu just to the right of where it now says “MEAN.”
  + Select “Reference Line.”
  + Under “Reference Line Type” select “Mean (Average).”
  + Under “Group By” select “Repeat By.”
  + Then, click “Apply.”
* Now in the “Color (Group By)” Section select “type.”
* In the “Repeat By” Section select “country.”
* In the pop up window select “Limit the values to show - (Top 20).”
* Now, underneath the “Repeat By” you should see small green wording saying “Top 20 by (current Y).”
  + Click on that green wording and a pop up window will appear.
  + In the “Number of Results” section type in “10.”
  + In the “Based On” section click on the dropdown menu and select “Number of Rows.”
  + Then click “Apply.”
* Finally, above your two graphs ensure that all options are unchecked. Uncheck the “Fit,” “Sync X Axis,” and “Sync Y Axis” options by clicking on them.
* Before exiting, make sure that the Reference Line selection under the Y Axis remains grouped by “Repeat By” - if it has changed to show “All,” return to the Y-Axis chart selection step to fix that.

We can interpret these results by comparing these line graphs to each other, keeping in mind that the Y Axis is different for each graph. We can also look at the general trend line to see how it compares to the mean for each graph.

Now, for the second visualization to answer RQ3:

* After navigating to the “merged\_titles” dataset, using the table view, click on the dropdown menu on any column and click on “Summarize (Aggregate).”
* Next, under the “Group By” section:
  + You will choose “title” first.
  + Then “release\_year” second, but after choosing this ensure that just to the right of it says “INT” for integer, and if it doesn’t click on that, showing a dropdown menu where you will choose “integer.”
  + Then, “type” goes third.
  + Then, “country” goes fourth.
  + Now, move on to the “Value section.”
  + You will choose “duration.”
  + At first, the value next to “duration” will say “UNQ,” however, we don’t want that.
    - So, click on the “UNQ” and a dropdown menu will pop up and you should choose “Max Value.”
  + Then, hit “Run” giving you a simplified table with only values you want.
* After that, select the dropdown menu on the “country” column.
  + Hover over “Filter” and then select “Remove NA.”
  + Then hit “Run” in the pop up window.
* Now, select the dropdown menu on the “duration” column.
  + Hover over “Change Data Type” and then select the “Convert to Numeric” option so that you can utilize averages as an option in this chart.
  + Next, you will then hit “Run” in the pop up window that appears.
* Next, click the dropdown menu on the “release\_year” column.
  + Hover over “Change Data Type” and then hover over “Convert to Date/Time” then select “Year” giving you a pop up window.
  + Then, hit “Run” in that window.
* To finish up the table, select the dropdown menu on the “type” column.
  + Hover over “Filter” and then select “Is Equal To…”
  + Then in the pop up window select the “Movie” option in the dropdown menu there.
  + Then hit “Run.”
* Now, to create the chart click on “Chart” at the top:
  + Drag the green pin to the last step that you did (Filtering only Movies).
  + In the left side of your screen under “Type” select “Line.”
  + Under “X-Axis” select “release\_year.”
  + Under “Y-Axis” select “duration” but then click on the “UNQ” that is just to the right of it giving you a dropdown menu where you will select “Mean (Average).”
  + Then click on the hamburger menu just to the right of where it now says “MEAN.”
  + Select “Reference Line.”
  + Under “Reference Line Type” select “Mean (Average).”
  + Under “Group By” select “Repeat By.”
  + Then, click “Apply.”
* Now in the “Color (Group By)” Section select “type.”
  + Then, click on the hamburger menu just to the right then select “Color, Group, & Sort.”
  + Then in the pop up window towards the bottom where it says “Color” and change from the default color to “Light Orange.”
  + Then click “Apply” and then “Close.”
* In the “Repeat By” Section select “country.”
  + In the pop up window select “Limit the values to show - (Top 20).”
  + Now, underneath the “Repeat By” you should see small green wording saying “Top 20 by (current Y).”
  + Click on that green wording and a pop up window will appear.
  + In the “Number of Results” section type in “10.”
  + In the “Based On” section click on the dropdown menu and select “Number of Rows.”
  + Then click “Apply.”
* Finally, above your two graphs ensure that all options are unchecked. Uncheck the “Fit,” “Sync X Axis,” and “Sync Y Axis” options by clicking on them.
* Before exiting, make sure that the Reference Line selection under the Y Axis remains grouped by “Repeat By” - if it has changed to show “All,” return to the Y-Axis chart selection step to fix that.

Likewise for this graph, we can interpret these graphs by comparing these line graphs to each other. Also note that the unit of duration is the number of season, as these are TV shows, rather than minutes. Also note that not all TV shows were released in the same time frame from each country.

To answer RQ4, we will transition and look at the inner join dataframe and aggregate titles that are available for streaming on both platforms.

* Select merged\_titles under Data Frames and select Chart view.
* We will first filter out movies/TV shows that although have the same title, are actually different movies/TV shows.
* To do so, we will create a Summary Table.
  + Select “Summary Table” under Type, “title” under Group By, and “release\_year.x,” “release\_year.y,” “director.x,” and “director.y” under Value.
  + From here, we shall determine which movies/TV shows are actually not the same.
  + If the release years and directors are different, then clearly the movie/TV show is not the same.
  + We have determined that these movies/TV shows are *Amy*, *Brain* *Games*, *For the Birds, Genius, Inspector Gadget, Once Upon a Time, The Bad Batch, The Blue Umbrella, The Jungle Book,* and *The Little Mermaid.*
* We will now be filtering these out of our comparison.
  + Click “+ Add a new step” under Steps and select Filter.
  + Then, under Column select “Title,” under Operator select “Is Not In (Multiple Values),” and then select all of the aforementioned movies/TV shows.
  + Then hit Run.

We will now compare the ratings of these remaining movies/TV shows that are indeed the same movie/TV show.

* To do so, we will first create a boolean calculation which returns “True” if the ratings are the same and “False” if ratings aren’t the same.
* Under Table view, select the down arrow next to “rating.x,” hover over “Create Calculation,” and then select “Standard.”
* Here, we will type our if-else condition. Under Calculation Editor, type the following:
  + if\_else(rating.x == rating.y, “True”, “False”)
* Rename this column as “is\_same\_rating.” Then hit Run.
* Now, create a new chart by going to Chart view and clicking the + sign, then “Summarize Table” under Type.
* Under Group By, select “title” and under Value, select “rating.x,” “rating,y,” and “is\_same\_rating.”
* Check and make sure that those ratings that are the same are indeed listed as “True,” and different ratings are “False.”
* The movies/TV shows that have different ratings are Becoming, Before the Flood, PJ Masks, White Fang, and Zapped.
* To create a title, select the gear icon and scroll down to Title.
* Then under Text, type in your title (for example, "Comparing Ratings Between Netflix and Disney+").

To create a visual of this data, let’s create a Pie chart.

* Click + under Chart view and select “Pie” under Type.
* Under Color (Group By), select “is\_same\_rating,” under Value select “(Number of Rows),” and under Style select “Pie.”
* Add a title by clicking the gear icon and scrolling down to Title, and under Text type in your title (for example, "Percentage of Titles with Same Rating").
* Under Repeat By, select "rating.x"

These charts show how many movies/TV Shows from Disney+ are the same on Netflix. We can see that all movies/TV Shows that are rated TV-14 and TV-Y7 on Disney+ are rated differently on Netflix, and there are a few movies/TV Shows rated as PG, TV-G, and TV-PG on Disney+ that are rated differently on Netflix.

Now let's see the difference between when the titles were added to Netflix and Disney+.

* Click + under Chart view and select “Bar” under Type.
* Select "+ Add a new step," then hover over "Create Calculation (Mutate)" and select "For Single Column."
  + Create your new column name as "difference\_in\_years."
  + Under Calculation Editor, type the following:
    - (date\_added.x - date\_added.y)/365
  + Then hit Run.
  + Now create a new "Create Calculation (Mutate)" step and name this column "rounded\_difference\_in\_years"
  + Under Calculation Editor, type the following to round the difference to 2 decimal places:
    - round(difference\_in\_years, digits = 2)
    - Hit Run.
* Make sure the Pin is selected to this new step.
* Now select "title" under X Axis, "difference\_in\_years" under Y Axis, and check the box next to Sort By.
* We can see that in general, titles that are available for streaming on both platforms were added to Disney+ before Netflix.
* To see if there is a trend between these titles and our new "is\_same\_rating" column, select "is\_same\_rating" under Color (Group By).
  + This shows us that interestingly, all titles that do not have the same rating were added to Disney+ before Netflix.

This visualization shows that in general, most titles that are available for streaming on both Disney+ and Netflix were added to Disney+ before Netflix, as shown by most of the bars being above 0.

Lastly, let's see the difference between the same movie's duration on Netflix compared to Disney+.

* First, we need to change the data type under "duration."
  + Under Table View, select the down arrow next to the column "duration.x," hover over "Change Data Type" and select "Convert to Numeric."
  + Repeat this step for "duration.y."
* Next, we want to filter out all TV shows.
  + Click "+ Add new step" and select "Filter."
  + Select "type.x" under Column, "Equal To" under Operator, and "Movie" under Value.
  + Hit Run.
* For our last step before creating the graph, we will create another "Create Calculation (Mutate)" step.
  + Name this new column "difference\_in\_duration."
  + Type in the following:
    - duration.x - duration.y
  + Hit Run.
* Now, click + under Chart view and select “Bar” under Type.
* Select "title" under X Axis, "difference\_in\_duration" under Y Axis, and check the box next to Sort By.
* Click the hamburger menu next to Y Axis and select "Window Calculation."
  + Under Calcuation Type select "% Difference From," under % Difference From select "Mean (Average)," and under Group By select "Color By."
  + Hit Apply.
* To see if there is a relationship with the % difference in mean movie duration and ratings, select "rating.y" under Color By.
  + From here, we can see that there is no immediate relationships between rating and % difference in mean movie duration.

This visualization shows a wide distribution of difference in mean movie duration of the same title. Some titles on Disney+ are longer than the same title on Netflix and vice versa.

### Collaboration/Bibliography

Curry, David. “Video Streaming App Revenue and Usage Statistics (2023).” Business of Apps, Business of Apps, 9 Jan. 2023, <https://www.businessofapps.com/data/video-streaming-app-market/>.

Chase, Marilyn. “How Disney Propaganda Shaped Life on the Home Front during WWII.” Smithsonian.com, Smithsonian Institution, 11 July 2022, <https://www.smithsonianmag.com/history/how-disney-propaganda-shaped-life-on-the-home-front-during-wwii-180979057/>.

Nielson. “Nielsen Study Reveals Significant Growth of Asian Representation on TV in 2021.” Cision, 19 May 2022, <https://www.prnewswire.com/news-releases/nielsen-study-reveals-significant-growth-of-asian-representation-on-tv-in-2021-301550193.html>.

### Reflection

Completing this assignment has provided us with valuable insights into the content strategies and audience preferences of Disney+ and Netflix. We have learned that these platforms are continually expanding their content libraries and expanding their global audience. Additionally, we discovered that many movies and TV shows are being filmed and produced in multiple countries, which can provide a unique perspective on different cultures and lifestyles that are depicted on screen.

Looking back at our analysis, we realized that there were some areas where we could have expanded our research. One of the most significant opportunities for improvement was to include other streaming platforms in our analysis, such as HBO, Hulu, and Apple TV. By doing so, we could have gained a more comprehensive understanding of the streaming industry's landscape and identified any patterns or trends across different platforms.

Furthermore, we acknowledge that our analysis was limited by the data available to us. While we were able to gain valuable insights from the datasets we used, we would have liked to have more comprehensive data, including total viewership, revenue, and other metrics that would have provided a more in-depth understanding of the industry's performance. With this additional data, we could have explored more nuanced questions, such as the impact of individual titles on viewership and revenue, and identified areas of potential growth or concern for each platform.

For future students embarking on this project, we would advise that they find datasets that not only have many columns to aggregate through, but also have data that they can create research questions to. We found it difficult to create visualizations for some of our questions because of the nature of the question. We would also advise to think of what kind of visualizations to create before looking for datasets, so that one would know what to look for in the dataset to create the visualizations. Lastly, when looking for datasets, students should ideally be interested or familiar with the data beforehand so that it is easier to think of research questions.

In conclusion, while our analysis provided us with valuable insights into the content strategies and audience preferences of Netflix and Disney+, there were significant opportunities for improvement. We recognize the importance of including additional platforms in our analysis and utilizing more comprehensive datasets to gain a more accurate and complete understanding of the online streaming industry.