

DS4200: Information Presentation and Visualization

Professor Yang

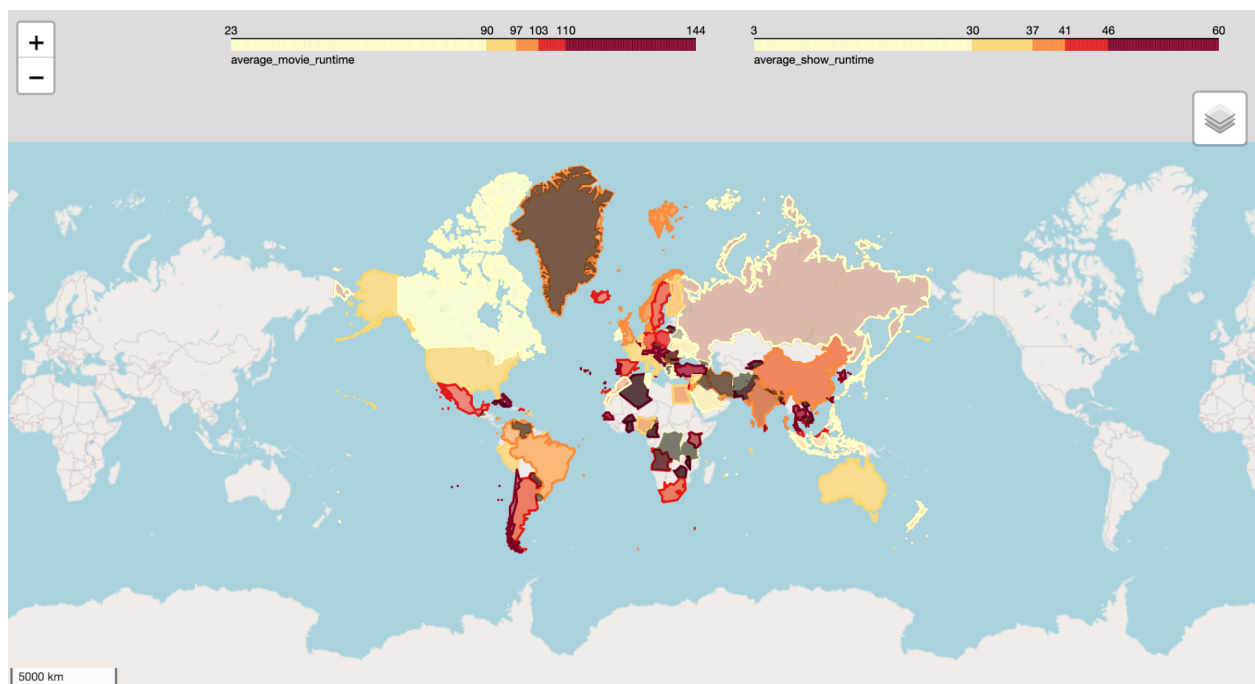
I Man Liu, Joseph Trisnandi, Khanh Nguyen

April 19th, 2024

Deployed on <https://alice0217.github.io/DS4200-Project/>

Final Project Design

Overall Map

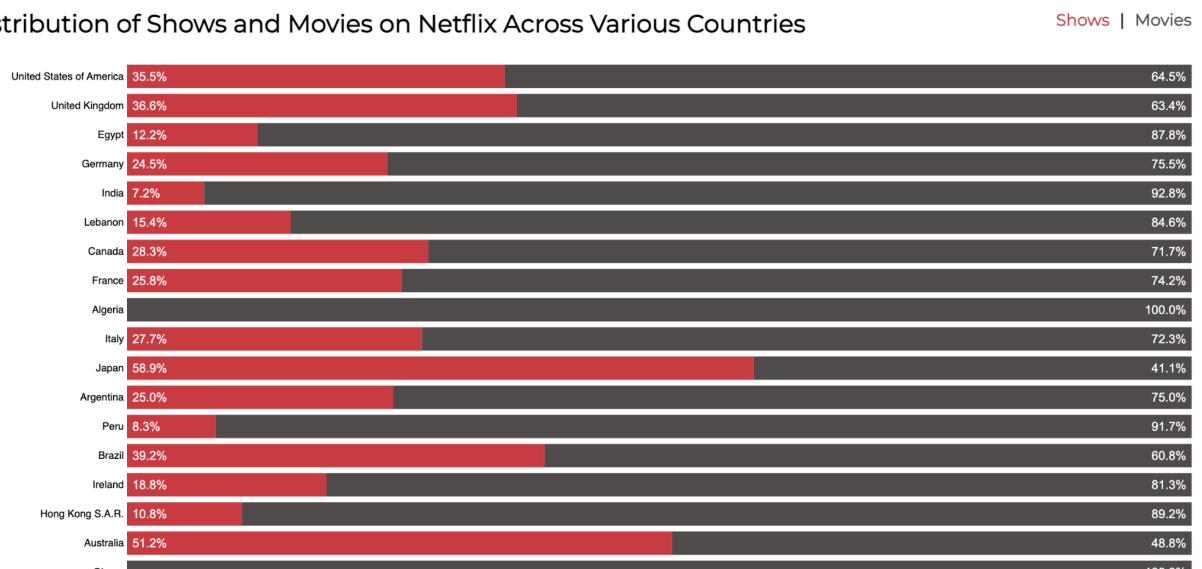


Our group created an overall choropleth geographical map to illustrate the average runtime of shows and movies across countries and continents. Our color theme was chosen to match Netflix's branding to help users easily discern the length preferences of each country in their productions. In this graph, each country is represented by an area (mark), with color (one of the channels) indicating the average runtime: darker colors represent longer average runtimes.

To enhance the viewer's experience, we also include tooltip and popup functions for viewers that are interested to learn more about all the available information that relate to the movies and shows on Netflix of that country. Hovering over a country displays its name in the tooltip, making it easy for viewers to locate specific countries. The popup provides additional information such as the earliest show release year, latest movie release year, average imdb scores, and more. Therefore, this graph aims to provide viewers a grasp of all general information and attributes that we are going to explore of how the shows and movies on Netflix produced by different countries differ.

Shows and Movies Distribution Stacked Bar Plot

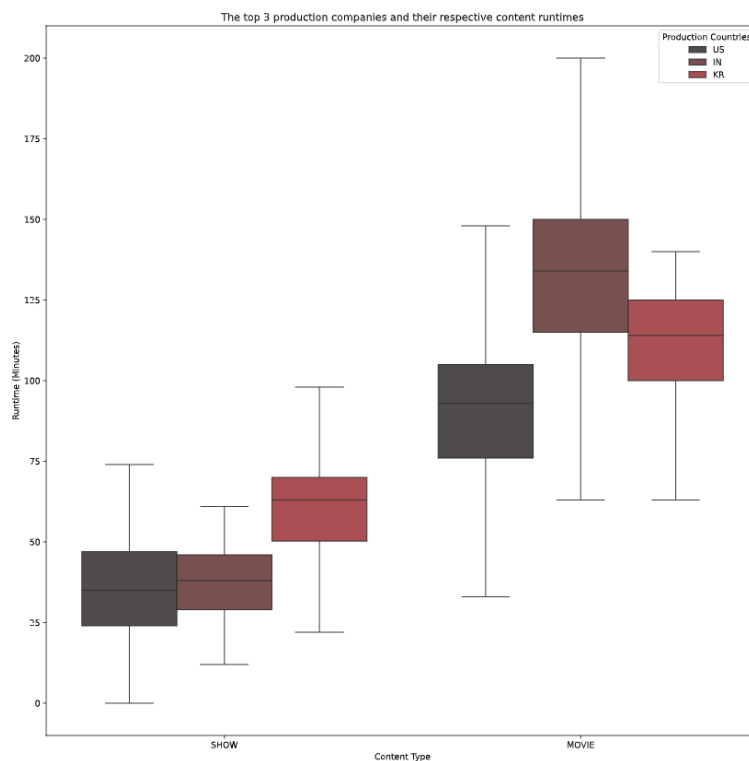
Distribution of Shows and Movies on Netflix Across Various Countries



Our group wanted to create a stacked bar plot because we are interested in seeing the movies and shows distribution across all the countries in the dataset. Therefore, we choose red and black as our channel (corresponding to the labels 'Shows | Movies'), resembling Netflix's branding color and also making it straightforward for users to identify which proportion is larger. The red and black colors have a high contrast, which is good for finding patterns. Moreover, we

added texts to display the percentages, to better quantify the differences. For the position channel, we displayed the bars horizontally, also for the purpose to make it easier to observe. Since there are 109 countries in our dataset, it would be more pleasing to scroll up and down instead of left and right. Moreover, it would be more straightforward to observe the split from left and right.

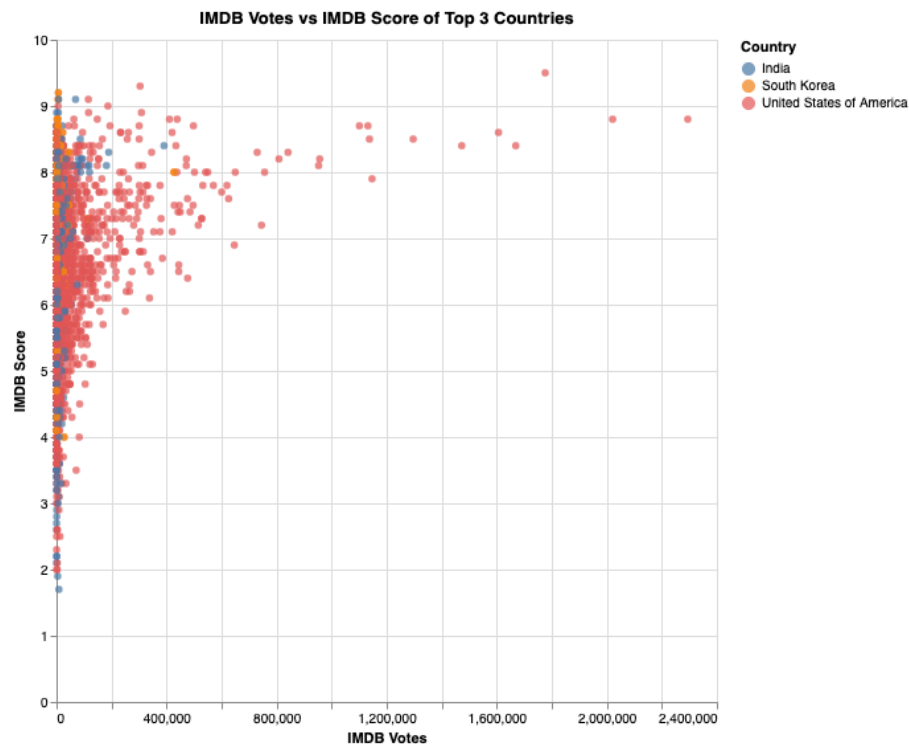
Runtimes of Content Types Side-by-Side Boxplot



IMDB Votes vs. IMDB Score of Top 3 Countries Scatter Plot

Next, our group wanted to find the distribution runtimes of each of the content types. We decided since we wanted as much data as possible, we obtained the top 3 production countries in the dataset and made a side by side boxplot of the country and their distribution. This allowed us to

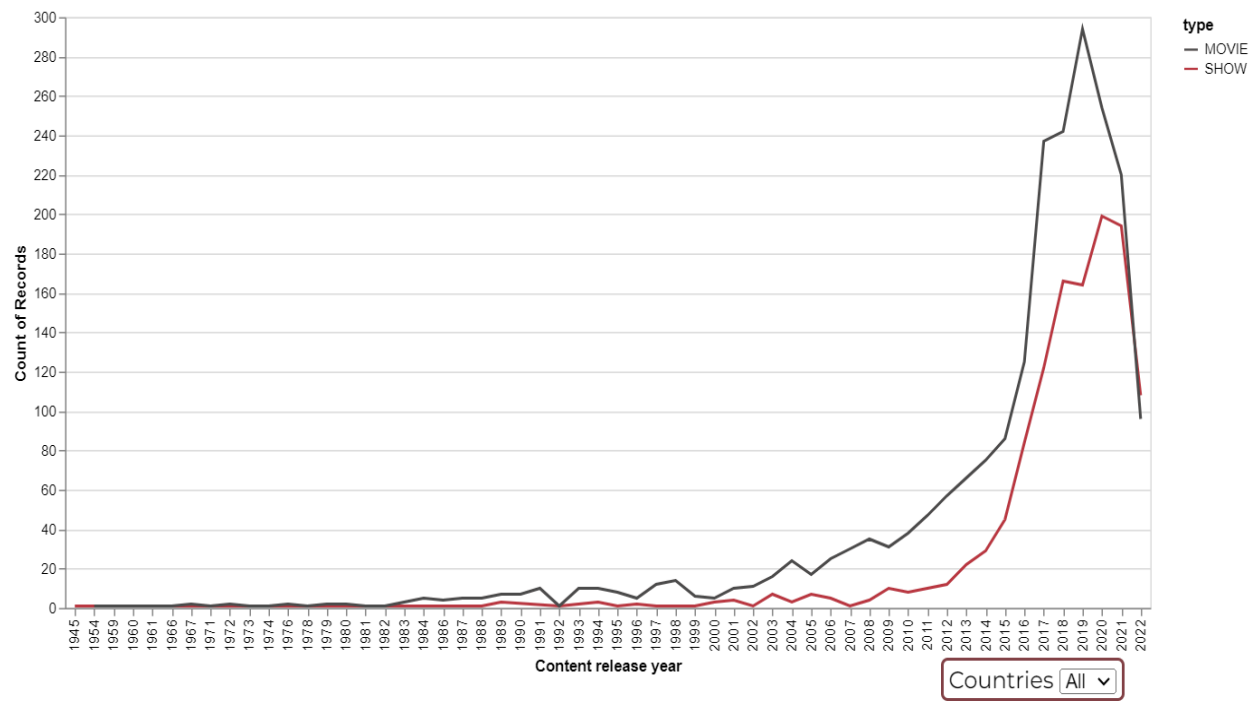
easily compare the difference of runtime across countries and also across content type. For example, we could easily see that movies tended to have longer runtime than shows and we could easily identify how Korean shows were typically longer than US or Indian shows. The marks of a boxplot are the lines which help represent the boxes and their quartiles, while the channels are size (length of line) and color (to differentiate countries).

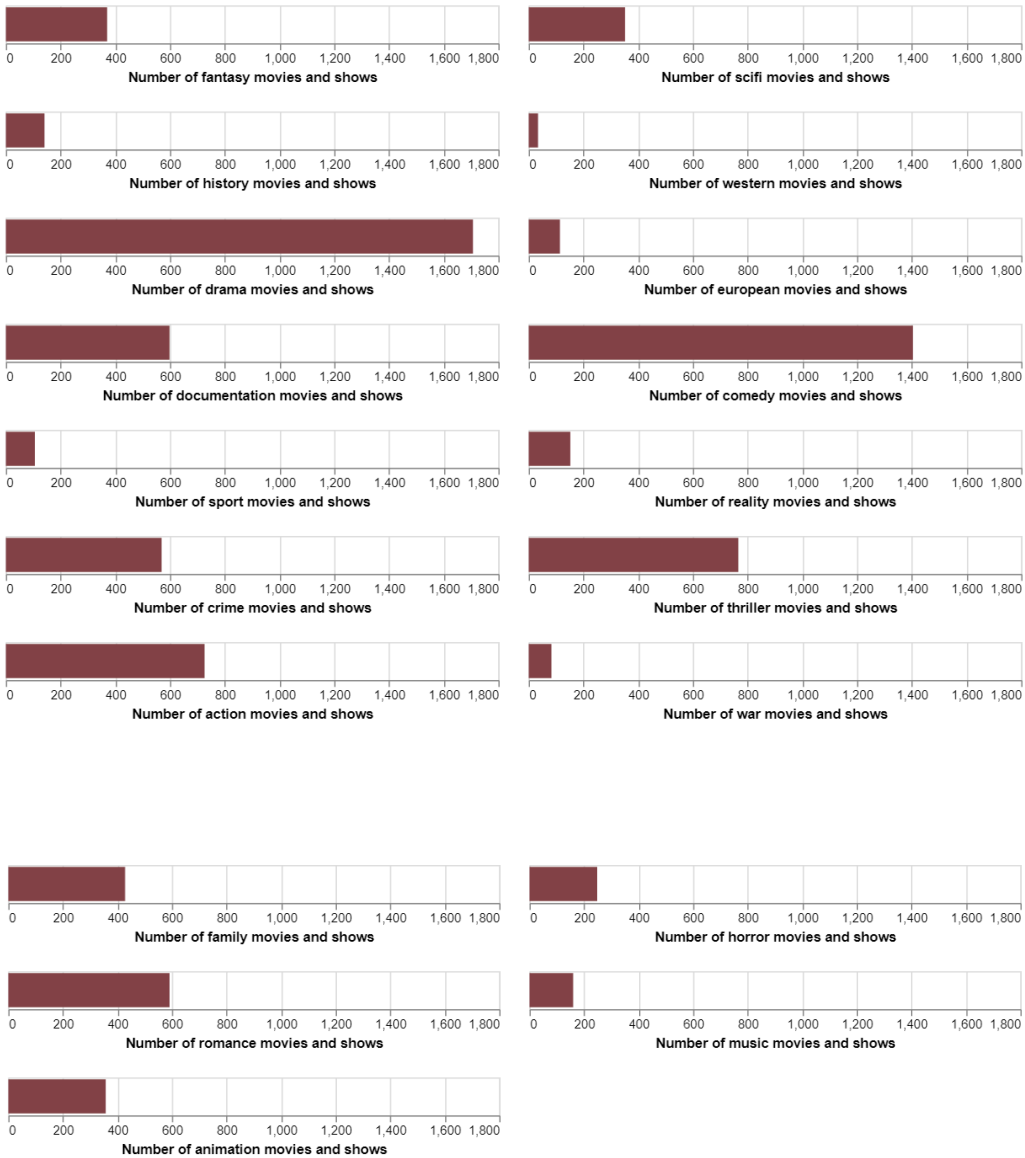


We created a scatterplot to explore the distributions, spread and correlation between imdb votes and imdb scores of the top 3 countries which are the United States of America, South Korea, India. We perform this graph to understand the amount of imdb votes and the imdb scores of each vote coming from the viewers of each country to understand more of the viewer's taste and how they correspond to the movies and shows. The mark for this graph is points where each point represents a vote for that specific country. The channels for this graph are position (horizontal), and color. We decide to have it vertical based on how the imdb scores/ratings are based on the imdb votes, and the color is to differentiate between the top 3 countries. For

instance, we notice that movies and shows with more votes tend to have higher ratings. Another thing that we observe is how the US covers the widest range of IMDB Votes and Scores compared to India and Korea.

Content Genres and Release Years Interactive Line + Bar Plot

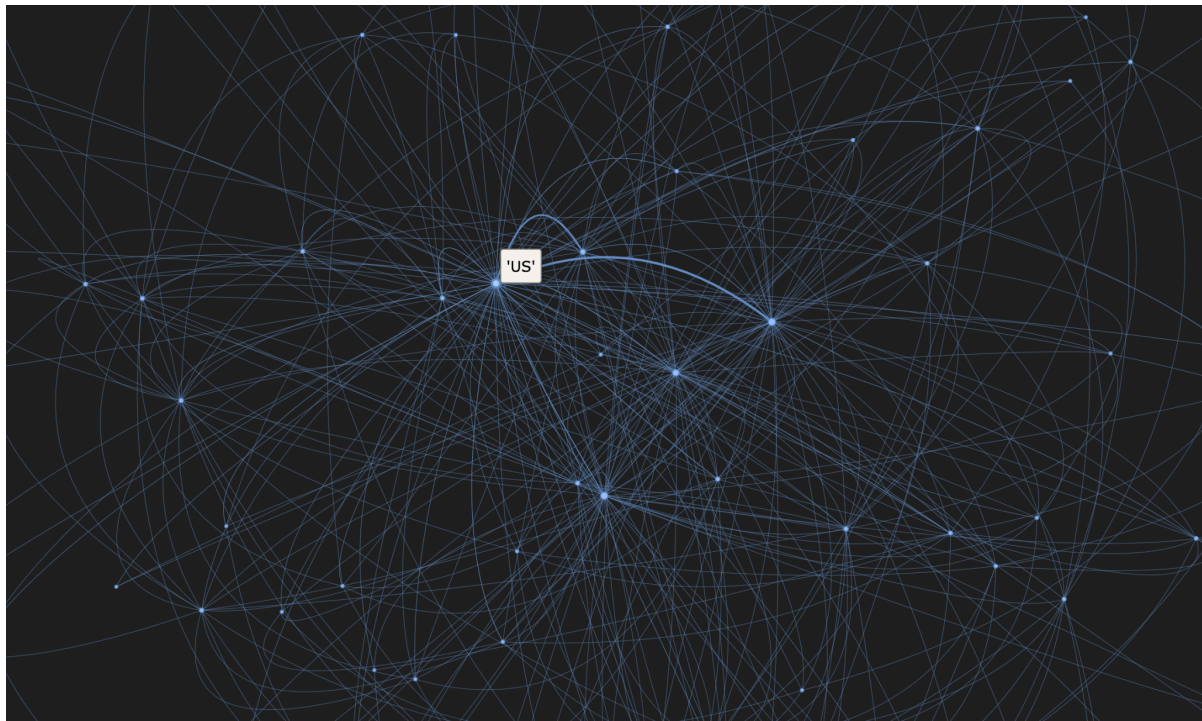




We created this graph to identify the number and content type of releases made by the top 3 countries over a period of time. The marks of the line chart included line, while the channel would be tilt and color (to differentiate between content type). The dropdown bar allows you to switch between countries, allowing the user to see trends in each country. In addition, we wanted to answer what type of genres were released during a block of time. We allowed the user to selectively choose a period of time, where the bar charts would reflect the count of content genres released at that period. The mark for the bar chart is a line, while the channel for the chart

is the size. To optimize screen space, we organized the bars into two rows, allowing comparison of multiple genres at a glance on a single screen, compared to a single row format.

Countries Partnerships Network



We want to create a country network because we want to find out the collaborations of countries in making movies and shows that are on Netflix. The marks are points (countries) and lines (edges between countries). The channel is size, the thickness of the edges that represents their weight, for example, we find the edge between US and Great Britain is particularly thick, representing that they've had a lot of partnerships in producing movies and shows on Netflix. Moreover, the color channel helps us identify countries like the US and France that have a large degree, because they are brighter than other countries in the network. Lastly, the position channel helps us identify outliers such as Sri Lanka that have a much smaller degree than the countries located more in the center.