IMDb Movie Data Visualization
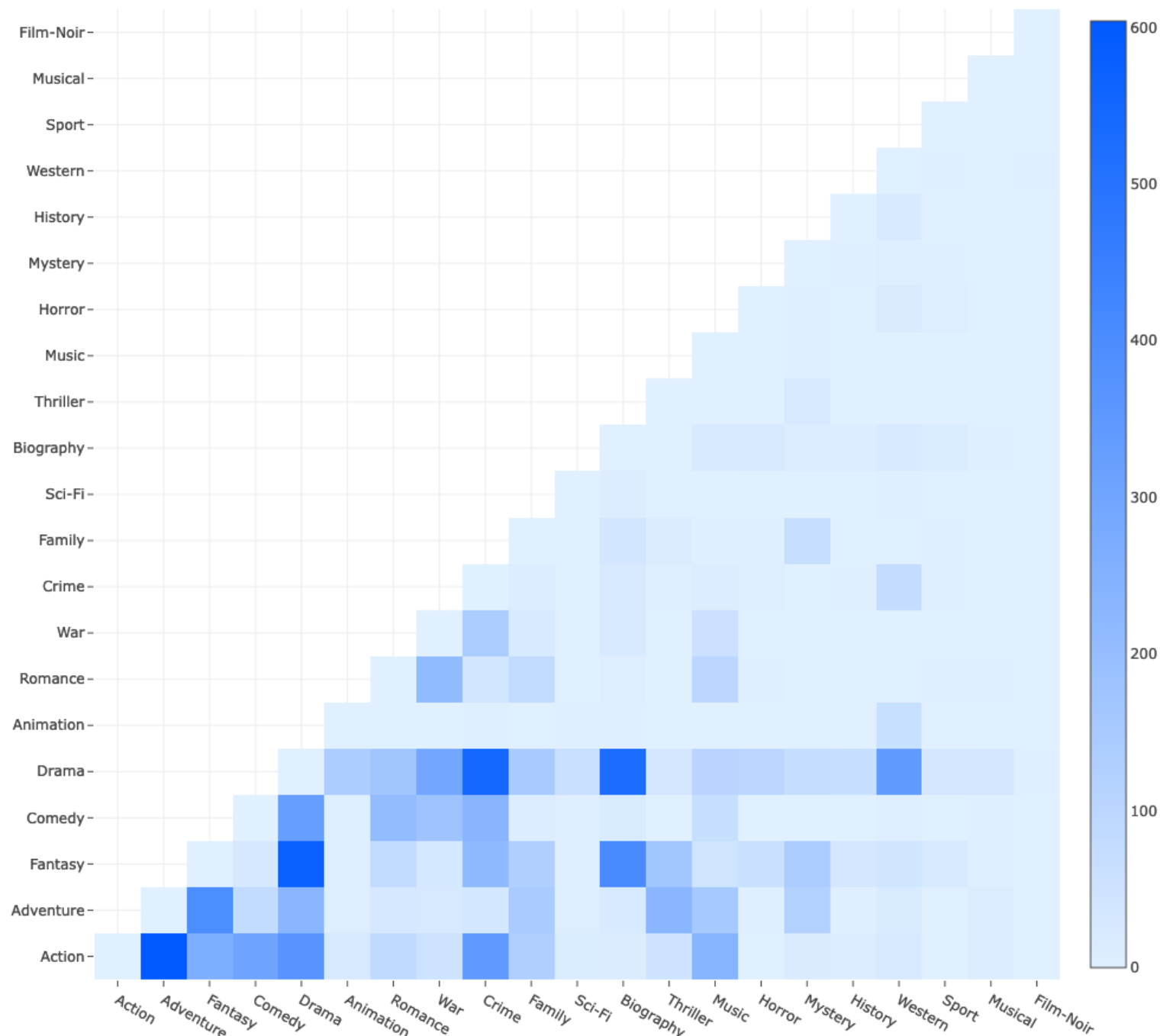
Objective

Introduction

As time progresses, people have an increasing number of choices for movie-watching. As a result, various video rating sites have emerged. The IMDb movie website is one such site, which offers detailed information about movies. This information is incredibly valuable for movie enthusiasts, researchers, and industry professionals seeking to gain a better understanding of movie culture and audience preferences. Our data visualization tool will provide an intuitive and interactive representation of the IMDb movie dataset. By referring to our visualization, users can make informed decisions about which movies to watch, analyze market trends, and gain a deeper understanding of movie culture. We hope to explore the relationship between different variables of IMDb films and the feature of IMDb users

1. What is the correlation between different film categories?
2. What is the number of films in different genres or countries along the time?
3. What is the distribution of different film ratings by country, year, and film length?
4. What are average IMDb scores of films by different countries?
5. What are frequent words in film titles?

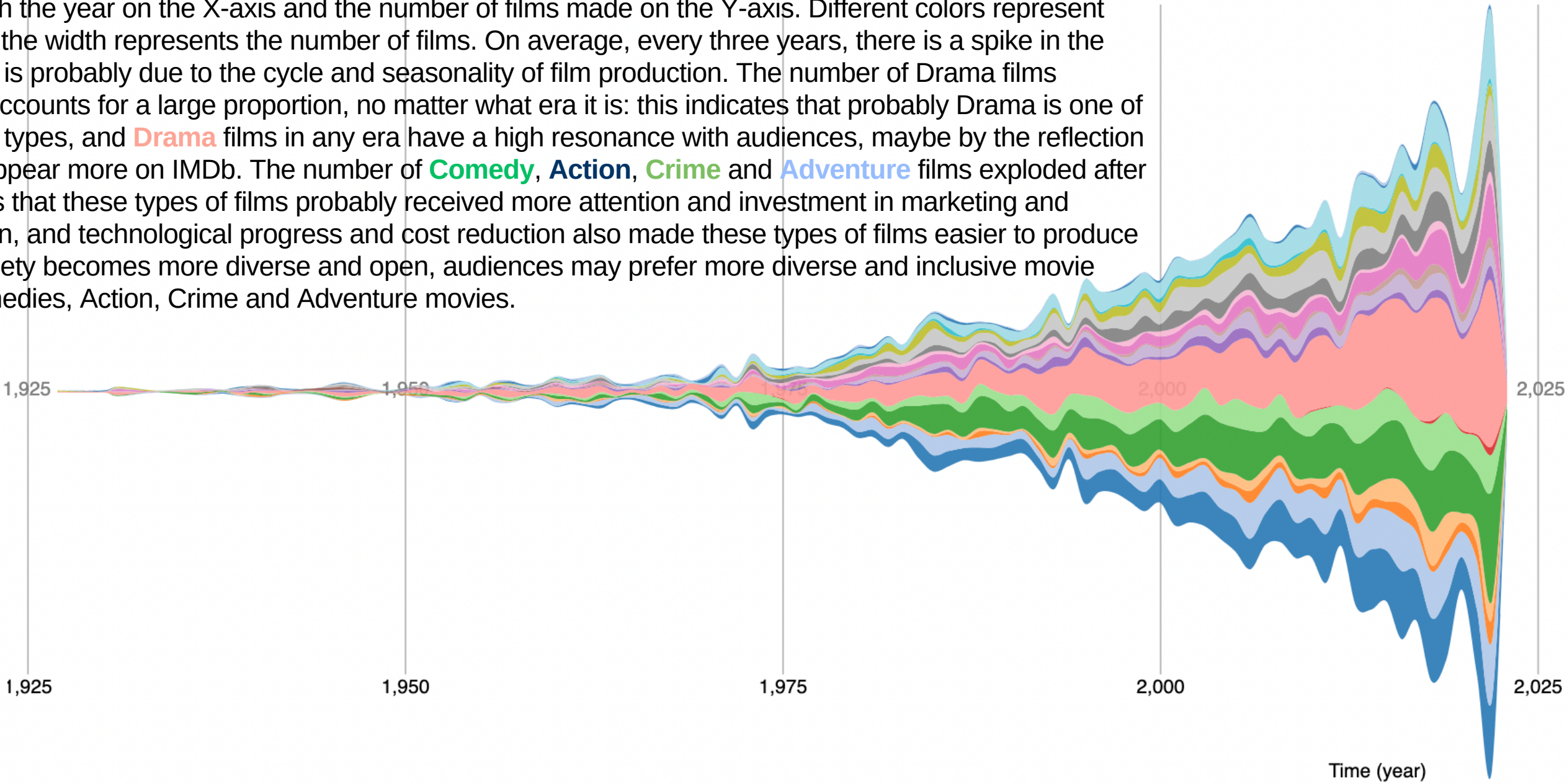## Heatmap of the Correlation between Genres

vis.1 What is the correlation between different film categories?

This is a heatmap. The two coordinate systems list the categories of movies, and the color depth of the squares represents the frequency (correlation) of the two categories appearing in the same film. It can be observed that the top four categories with a high correlation are Adventure and Action, Drama and Fantasy, Drama and Crime, and Drama and Biography, which shows that creators prefer to create movies in these categories and these combinations have a favored status. In terms of individual categories, Action and Drama are related to more categories than others, which indicates that the films in these two categories are highly integrated and diversified. This versatility and flexibility in storytelling can make movies in these categories more appealing to different types of viewers, and help them achieve broader commercial success.
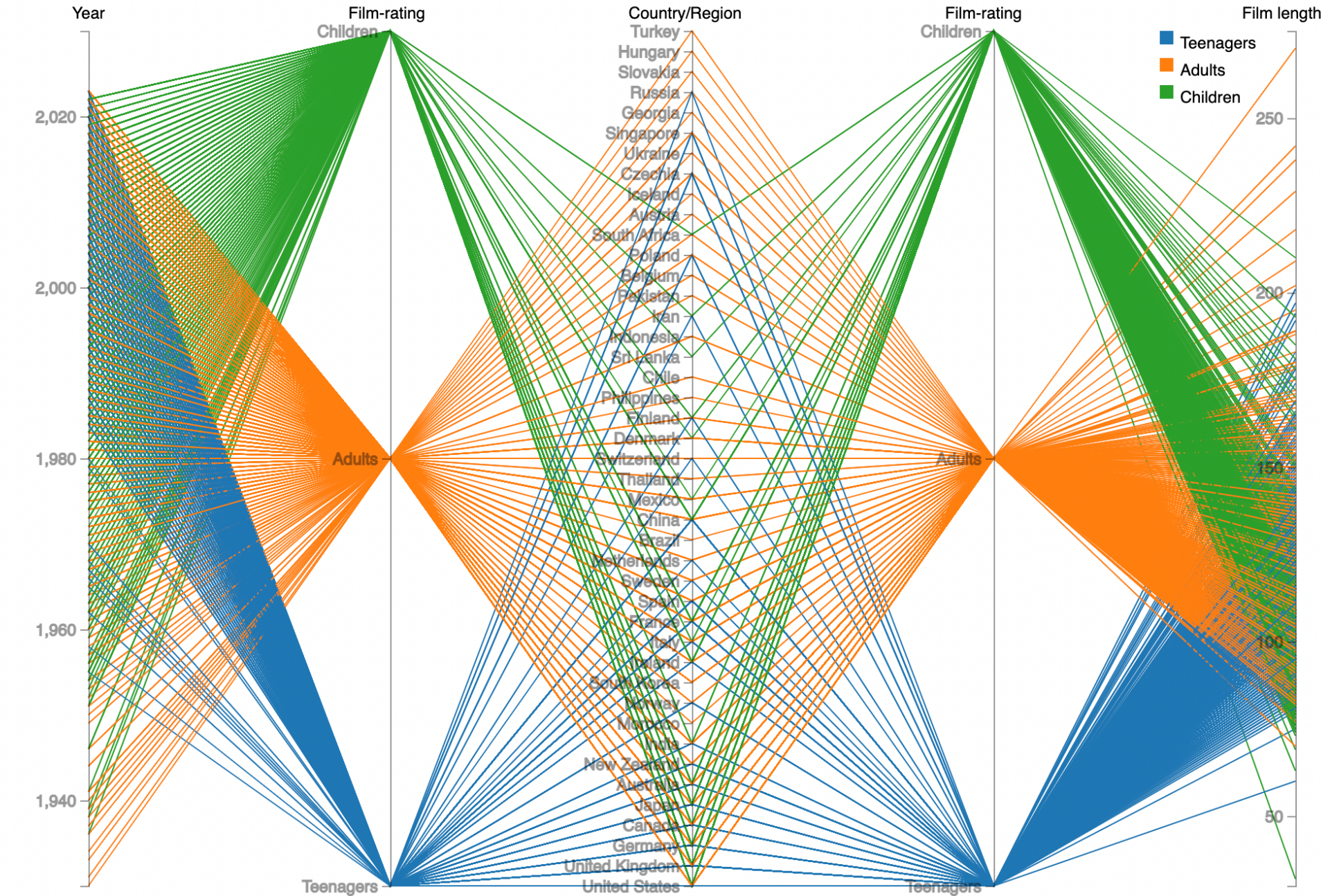
Timeriver of Number of Movies of Genres along the Time

vis.2 What is the number of films in different genres or countries along the time? (Poster only shows genres)

This is a timeriver with the year on the X-axis and the number of films made on the Y-axis. Different colors represent different genres, and the width represents the number of films. On average, every three years, there is a spike in the number of films. This is probably due to the cycle and seasonality of film production. The number of Drama films appearing on IMDb accounts for a large proportion, no matter what era it is: this indicates that probably Drama is one of the most popular film types, and Drama films in any era have a high resonance with audiences, maybe by the reflection on society, so they appear more on IMDb. The number of Comedy, Action, Crime and Adventure films exploded after 1975, which indicates that these types of films probably received more attention and investment in marketing and commercial promotion, and technological progress and cost reduction also made these types of films easier to produce and promote. As society becomes more diverse and open, audiences may prefer more diverse and inclusive movie themes, such as comedies, Action, Crime and Adventure movies.

Parallel Plot of the Distribution of Film-Ratings by Years, Countries/Regions and Lengths
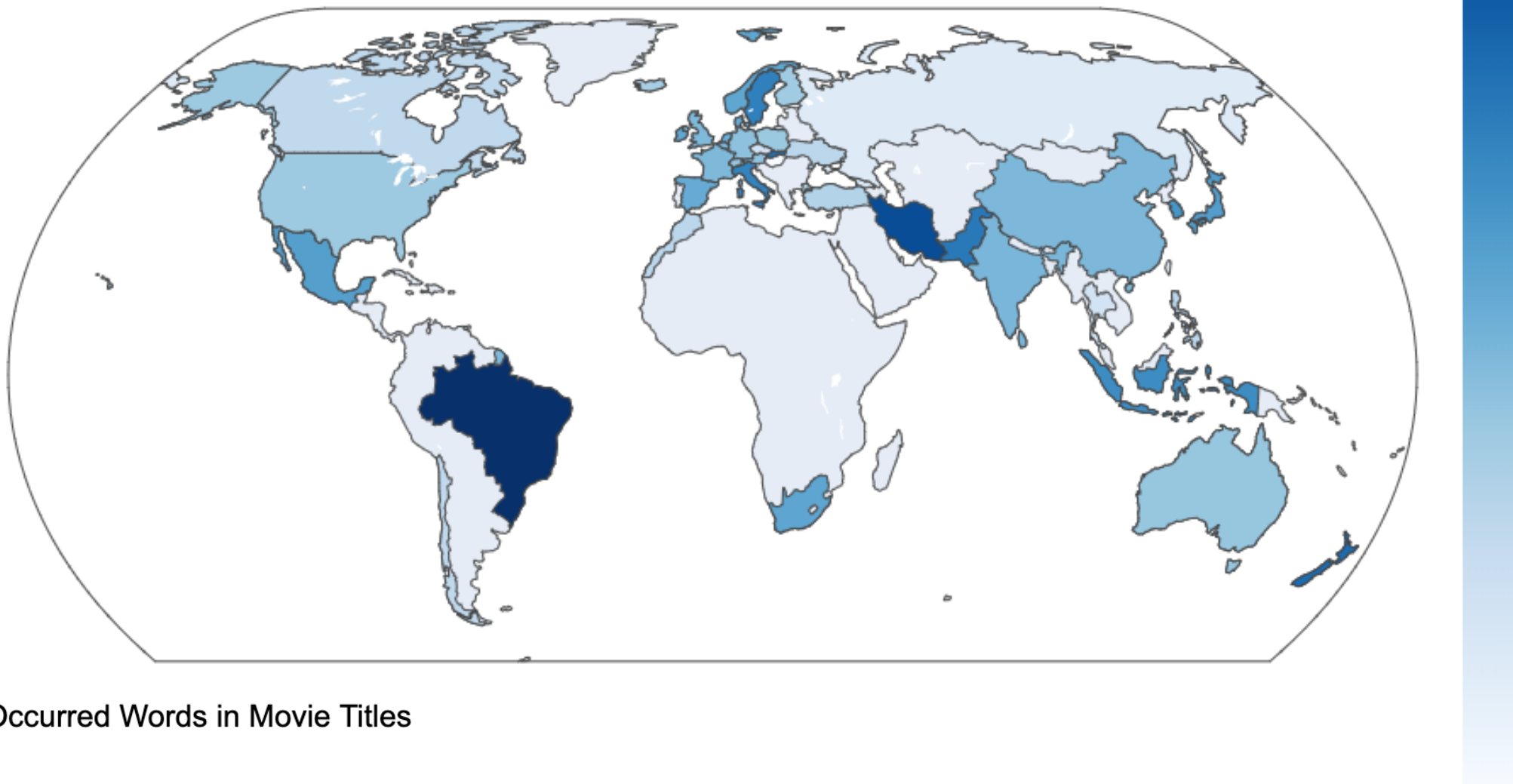

vis.3 What is the distribution of different film ratings by country, year and film length?

This is a parallel plot with the Y-axis showing the year, rating, country, and time of the film. The three colors represent three different ratings (Children, Adults, and Teenagers), with each line representing a movie and the intersection with the Y-axis representing its corresponding attribute. Films for Teenagers tend to concentrate after the 1970s, while the distribution of others is relatively even. This may reflect the cultural and social changes of this period and shows that films for teenagers get more attention and production in this period, which may be related to the development of the rating system. In addition, IMDb users are likely to pay more attention to movies from Adults and Children. Movies for Adults have appeared in all countries, while films for Children and Teenagers distribute sparsely. This may reflect that adults are the main group of moviegoers, and there are the different demands and market scales of movies for teenagers and children in these countries. On the other hand, IMDb users can pay more attention to adult movies from all regions. Films rated to Adults have a large span of time, while those of Children and Teenagers are relatively shorter. This may be because the interests and demands of adult audiences are more diverse, thus the film lengths are more diverse, while those of young audiences have limited tolerance to the length of films.

Average Score in Different Countires/Regions

Word Cloud of the Top 200 Occurred Words in Titles

vis.4 What are the average IMDb scores of films by different countries?

This is a map showing all the countries or regions in the world, with the color of the region representing the average film score (audience rating) for that region. It can be observed that the average score of the United States, which has the largest number of creations, is not outstanding, which indicates that the audience's evaluation of the film varies from high to low quality. South America, Europe and Asia, Brazil, Switzerland, Italy, Iran, and Japan all have high average scores, albeit small numbers, indicating that these films are popular with their audiences, and the films are also competitive. The competitive nature of the film industry means that even these local popular films are not necessarily guaranteed to have a high level of popularity globally.

vis.5 What are frequent words in film titles?

This is a word cloud, which shows the keywords that appear in the title of a movie, text. It can be observed that the word "Man" appears most frequently in the movie, followed by "World" may reflect filmmakers' exploration and expression of human experience, society, and movie themes like love, death, loneliness, and so on. This fits with Drama and Action being more grab viewers' attention and get them interested in what the film is about. and the frequency of the words reflected in the shade and size of the "World", "Love", "Last", "Dead", "Night" and other words. Words like "Man" and relationships. Words like "Love," "Last", "Dead" and "Night" refer to classic emotional diverse and popular. This suggests that filmmakers may be using these common words to

Authors

Boqing Zheng, Eric Zhang, Ruiqi Chen