Firework

International Movie Ratings/Popularity by Country

Fan Guo, Sylvia Lai, Ian Lin, Varoot Phasuthadol, Hao Zeng

{guofan|szuhsuan|ianlin|varoot|haozeng}@umich.edu

http://infovis.varoot.com/firework/

Abstract

Movie data visualization is becoming increasingly popular for both commercial and research data analysis. Most of the existing visualizations only present discrete data of individual movies. We explored different visualization techniques to see how higher-level groupings help analysts better understand the overall statistics. We present Firework, a novel information visualization system that groups movies by year, country, and genre, and enables the exploratory data analysis on average rating and popularity. Firework consists of three interactive parts: a scatter plot, a time slider with trend view, and optional side filters. We believe that our system is especially helpful for movie distributors/importers to research on movie genres from foreign countries that would have higher potentials to make a hit in US market. We recruited 10 participants for the usability test of six tasks, identified three major findings, and made corresponding improvements afterwards. Finally, we propose a plan of future work to further extend the possibilities of our system.

1 Introduction

The movie preference regarding country and genre has been changing over time. Starting from 1960s, the "Golden Age" for world cinema, film industries outside of the United States has thrived and become popular among U.S. audience. While some countries produced a lot of good movies across various genres, others might choose to focus only on a specific genre. This results in changing movie preference over time, genre, and country of origin.

We analyzed movie preference with two factors: popularity and rating. We believed that these two variables are independent from each other. A movie can be popular but low-rated (e.g. bad movies with lots of advertising) or high-rated but less well-known (e.g. good documentary movies that attract only specific audiences). Our goal is to provide a system to visualize movies' rating and popularity, grouped by country of origin, and also allows users to see trends over the years. These are some questions we tried to answer with our visualization:

- How do movie popularity and ratings differ from one country to another?
- For a specific country, how does each genre differ from one another in terms of popularity and rating?
- How has the movie preference of a specific country changed from 1960 to 2011?

2 Related Work

Movie data visualization has become increasingly popular recently for both commercial and research use. Different aspects of movie data have been drawn from different field of interest, ranging from social network to movie pattern development. The New York Times published "The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008" (Bloch, M. & Byron, L. et al., 2008), showing how movies fared at box office using Stream Graph. Netflix came up with "Movie Similarity" visualization, linking 5,000 movies using color line representing strength of the similarities. Liveplasma (http://www.liveplasma.com/) offers similar visualization using Amazon.com's recommendation engine. However, most of the existing visualizations only show data from individual movie. None of them look at the movie data at a higher level. Our system addresses this issue by grouping movies by country and genre.

2.1 Design Principles

There are six variables to be represented in the visualization: country of origin, total number of movies, average rating, popularity (votes/movie), genre, and time. We found it very difficult to encode all these six variables into a single chart. Therefore, we utilized the following techniques for our system:

A. Scatter Plot

A scatter plot shows a relationship between two quantitative variables: one on x-axis and the other on y-axis (Cleveland & McGill, 1984). Since popularity and rating are the most important variables we wanted to display. We encode each country's popularity and average rating on x and y axes respectively. To maximize the use of scatter plot, we encode one more variable on the chart: total number of movies of each country, as the size (area) of a dot. The size of a dot also makes it easier for user to spot important data (countries that produce a lot of movies). To help finding a specific country, we also wanted to color each dot. However, we have about 50 countries and it is not possible to find 50 distinguishable colors, so we decided to group countries according to their culture and region. We explored different grouping options to make sure each group has comparable number of countries. These are the eight groups we use: English-speaking, West Europe, North Europe, South Europe, Russian/Slavic, Latin America, Middle East, and Asia.

B. Interaction

Since we could not encode all the variables on the same plot, we decided to use interaction to help users explore the data. There are seven categories of interaction based on the user intents (Yi & Kang et al., 2007). We implemented three of these interaction techniques to facilitate information finding: select, filter, and explore. Search function allows users to highlight specific countries. Filters allow users to see the data from a specific genre. Users can also explore the data by clicking each dot on the plot. We use two plots to show the detailed level of data: Firework, which we will discuss in the next section, and Time Slider.

C. Time Slider

Time is another important dimension of our visualization, and we need to allow the users to select a specific year as well as seeing the trend of the whole time span. Since our temporal data is linear, we believed it is best to use time as x-axis to illustrate the trend over the 50-year interval (Aigner & Miksch et al., 2008). We use y-axis to represent average ratings and popularity simultaneously, by using line chart and bar chart respectively. This *Trend Chart* allows users to see the peak or any outstanding data points of both variables at the same time.

3 System Description

3.1 Firework

According to the nature of our data, the genres can be seen as the child nodes of each country. The variables we need to display for each genre are the same as the ones we have for each country (total number of movies, average rating, and popularity). Since these variables have already been covered in the scatter plot, we think it is best that we also plot them on the same chart. To show that each genre is a child of a country (parent node), we connect them together with a line. We also use animation to show that a child node comes from the parent node. We call this plot *Firework*, since the child nodes are exploded from the parent like a firework. In order to prevent confusion or information

overload, we only show a firework when users click on each country dot and remove the previous firework from the plot so that only one firework can be seen at a time. Using Firework also makes it possible to compare different genres within one country. Moreover, by using the same coordination system, we save users from re-learning how to interpret the visualization.

3.2 Data Collection

We use the data from IMDB, which can be acquired from http://www.imdb.com/interfaces. The data is offered in plain text, and we re-constructed a database using a Python script. We also pre-processed the data by removing irrelevant entries, such as TV shows, adult films, and movies outside the year range (1960-2011). To keep the number of countries low, we only selected top 50 countries that produce the most movies for the past 10 years. We later dropped the United States from our data since it has produced far greater number of movies than other countries, thus skewing our scale.

3.3 Implementation

We implemented the system using HTML5, CSS3, and three JavaScript libraries: D3, jQuery, and jQuery UI. D3 is mainly used for the data visualization, jQuery for DOM (Document Object Model) manipulation, and jQuery UI for creating widgets and user interface.

3.4 Use Case

To better illustrate the use of Firework, we came up with a scenario in which film distributors are trying to figure out what movies he should invest next by using this visualization.

While reading the box office on a regular Monday morning, John, a movie distributor, tries to figure out which movie genres are more popular among his target audience so that he could import the right titles. John also finds himself facing an increasingly globalized movie pool with a large number of Asian, European and South American movies that have potential in US market. He opens up Firework and tries to find out the prospect country.

First, John wants to figure out which country group has higher popularity and ratings. By default the country groups are all selected, and each of the culture groups has a color and is associated with the dots which belong to this group. Since John has no specific country in mind, he deselects all country group first, selects each of them individually, and moves the time slider to see which group has a trend that both popularity and ratings grow overtime. He thinks his prospect country would be one of the Asian countries.

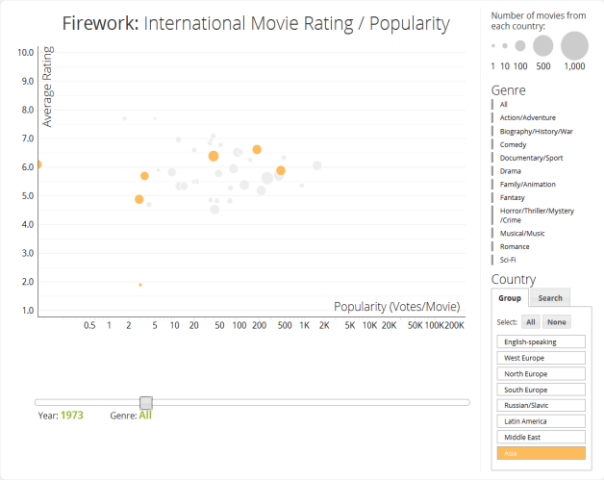

Figure 1: The Time Slider points to the current year and can be dragged to any specific year between 1960 to 2011.

By hovering over the dot, the label shows up and he realizes that "Hong Kong" has an increasing popularity over 50 years, "Japan" has an increasing rating and "India" has produced more movies over time (because the size of the dot is bigger). He clicks on these dots to understand the audience preference toward different genres. For Japanese movies, action and horror movies are more popular, but drama has the highest rating. For Indian movies, sci-fi has the highest popularity but biography and music have highest ratings. The Trend Chart under the main visualization also shows that Japanese movies have increasing rating but decreasing popularity, while Indian movies have increasing popularity but decreasing rating, especially for the past ten years.

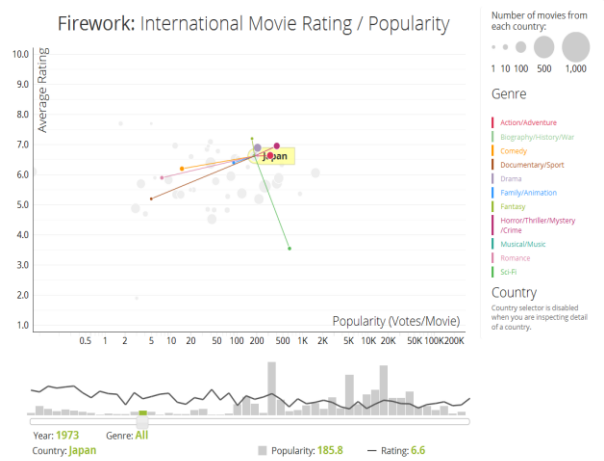

Figure 2: Firework is shown for Japan. The trend chart is also shown below the main visualization.

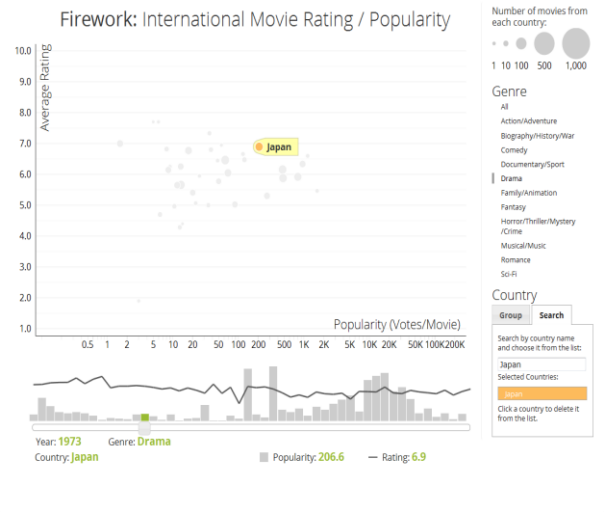

Figure 3: Genre filter and Country filter are used to pinpoint a specific country and genre.

After getting better understanding toward these two countries, John is curious how Japanese drama movies are doing. John clicks on "Drama" on Genre filter and searches for "Japan" on Country filter's search tab.

John thinks it is interesting that Japanese dramas have higher rating and popularity around 2000-2005. In the recent years, the rating remains high but the popularity has dropped. He thinks that Japanese dramas are good for him to invest. He is satisfied with the result and closes the system.

4 Evaluation

We conducted a usability test with ten college and graduate students from the University of Michigan. Each participant assumed a role of a film distributor who wants to import movies from outside the US and followed the same instructions to complete the following tasks:

1. Name at least one country with low popularity but at the same time has a relatively high ratings in 2011.
2. Identify the trend (in term of ratings and popularity) from 1960 to 2011 of this country.
3. Locate India in the visualization.
4. Identify the most popular genre of Indian movies in 2006.
5. Identify the trend of Indian music/musical movies from 1960 to 2011.
6. Name at least one country which got both high popularity and high ratings for fantasy movies in 2010.

We then had follow-up questions based on our observation of the tester's performance. From the test results, we have obtained three key findings as follows.

Finding 1: The current implementation of Firework needs improvement.

When testers were performing Task 4 to find genre details, the majority of them were confused after expanding the firework. Because of the lack of labels for

genres, the testers needed to look back and forth between Firework and Genre legend to match the colors with genre names. Moreover, the genre colors were close to one another. They also assumed that the genre dots should be clickable because that was how country dots work. Finally, it was difficult for them to close the firework, since there was no indication anywhere. We fixed some of these issues in our final implementation. However, the system can still be largely improved by hinting how to close the firework, choosing more distinctive colors for genres, and enabling Genre filter in Firework mode.

Finding 2: Better labeling is needed for the main plot and Trend Chart.

Many testers reported that some parts of the visualization cannot be interpreted easily, and more labels of explanation might help. The meaning of dot size was not clear, and some of the users thought it represented popularity. The label for current year and the legend for Trend Chart were too small, resulting in some confusion about the bar and the line graph. We added the legend for dot size after the testing.

Finding 3: Locating desired countries was not intuitive as we expected.

The country filter on the right-hand side was not used frequently by users. Some users would rather hover over each country than using the group filter or search function. Many of them reported they never noticed search on the second tab. A possible solution for this is putting search and country group in the same view instead of using tabs.

5 Discussion

From the user testing, the visualization is effective in the following aspects:

- Facilitate browsing movie ratings & popularity for countries/genres
- Provide intuitive transition between the overview perspective (the scatter plot) and the more detailed perspective (Firework)
- Provide easy to understand visualization for ratings, popularity and trend
- Allow intuitive navigation through the timeline
- Provide engaging aesthetic

Few alternatives were considered and prototyped for the visualization, but we finally realized they might not be as effective as Firework. Star glyph and pie chart are options we considered to display details about each country. Given the nature of the information to be visualized as being categorized two-dimensional quantitative variables, using star glyph alone is not sufficient since it best handles categorized one-dimensional quantitative variables. Pie chart can be used here, however, people are not very good at comparing quantitative data visualized as angles or lengths. Instead, positions are most effective in encoding such information. Therefore Firework is chosen as the popularity and rating of each genre are represented with horizontal and vertical positions respectively.

We also considered representing the temporal data differently. One technique we came up with is "Firefly", where the data from previous years are animated and shown on the same plot as a fading trail.

The advantage of this technique is that all the information is shown at once in a shared metric, thus showing traces of multiple countries is possible. However, our data is very noisy and therefore looks very confusing even when showing one trail, so we abandoned the idea.

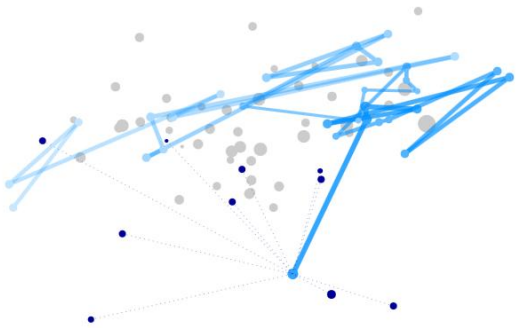

Figure 4: Firefly shows data from previous years as a fading trail.

6 Future Work

Our design currently does not allow users to compare the genres of one country with other countries since we only show one firework at a time. Allowing multiple fireworks of different countries to be displayed simultaneously would be very helpful for comparison. However, Trend Chart is currently only capable of displaying data for a single country, and thus has to be adjusted to allow viewing multiple countries.

We also have a dot-overlapping issue, making it difficult to click on a specific dot. A zoom-in/zoom-out function would effectively solve this problem. We can also show or hide some information according to the zoom level, which will make the visualization cleaner.

We have tried to fix those problems addressed by our testers, but the implementations of some of the suggestions may essentially affect our design concept and software architecture. However, some of these suggestions can be applied for the future work. For example, some of the users reported that the change of the meanings of the axes costs too much mental effort to interpret the visualization: The x-axis represents "popularity" in the scatter plot, but "time" in Trend Chart. Other users desired to have one more level of exploration about movie lists of each genre in each year in our visualization, which would make more sense about the reason why the rating/popularity was so high/low.

We also want to allow the users to choose a specific time span they are interested in. In this way, each country dot will represent the average popularity and ratings of that country in that time span, and can be treated as a miniature of the trend in that time span. This idea will thus further facilitate the users to compare the trend between different countries with much less mental efforts since all the miniature trends can be shown in a single chart at the same time.

7 References

Aigner, W., Miksch, S., Muller, W., Schumann, H. & Tominski, C.(2008) Visual Methods for Analyzing Time-Oriented Data. *IEEE Transactions on Visualization and Computer Graphics, 14(1)* pp 47 - 60.

Bloch, M., Byron, L., Carter, S. & Cox, A. (2008) The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008. Retrieved December 6, 2011 from http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

Cleveland, W. & McGill, R.(1984) Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association, Vol. 79, No. 387.*, pp. 531-554.

Hefele, C. (2009) Netflex: movie similarity. Retrieved December 6, 2011 from

http://www.the-ensemble.com/content/netflix-prize-movie-similarity-visualization

Ji Soo Yi, Youn ah Kang, Stasko, J.T.& Jacko, J.A. (2007) Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics, 13(6)* pp. 1224-1231.

Liveplasma. Retrieved December 6, 2011 from http://www.webdesignerdepot.com/2009/06/50-great-examples-of-data-visualization/