Project 1 Report

Yingdi Liu

My system is designed based on the movie dataset. It consists of three charts.

The first one is a simple histogram with a dropdown that can narrow down data to a dataset of a specific production country. This histogram can either tell users how many movies are produced each year worldwide or nationwide. When the cursor hovers above the bars, a tooltip will show up telling the exact number of movies and their production year this bar represents.

The second chart uses parallel coordinates to display movie information in terms of genre. There are three axes, representing runtime, budget and revenue of movies respectively. Each line represents a movie and connects the corresponding values on the three axes with the color of the line represents of a specific genre. On each axis, we can fix the range at will, the corresponding lines will stay on the plot while others will turn grey and hide beneath. Through this plot, we can discover interesting patterns of a specific genre of movie. For example, documentaries and horror movies tend to have shorter time and least budget and revenue. In contrast, action and adventure movies are most likely to have the highest budget and considerable revenue.

The third chart is a scatterplot. The x axis is the profit calculated from budget and revenue, and the y axis is average rating of movies on a one to ten basis. Each circle represents a movie, with the color indicates its genre and the size indicates its popularity. When the cursor hovers above a circle, the detailed information of the movie this circle represents will show up. The color legend of this plot is the same as the second one. This chart can also respond to the dropdown of first chart, movies of a certain production country can be selected. From this chart, I notice that most of the movies are American made (obviously), and almost all the profitable movies are from United States (Hollywood more accurately), including Avatar, which has the insanely highest profit among those movies. However, the majority of movies that lost money comes from Europe, which is understandable since a lot of their movies are anti-mainstream art movies and they are not made to be blockbusters.

These three charts may not reflect the reality of the movie market worldwide since there are only limited data and some of the items are incomplete, and this is not a bugless system, but still I had fun making this system and exploring the dataset. I think this system can be improved from several perspectives: first, I think the first chart can be more complex, the chart including the dropdown can be replaced by a zoomable world map; The styling can be improved as well; The colorscale of the second chart sometimes does not correspond to the third chart; There is also a problem with screen fitting, users may need to zoom in or zoom out the have the perfect display.

Peer Evaluation by Jiayu Tang
Merits:

- Good looking colors and fonts.
- Visualized most data in the dataset.
- Includes great transition animations for all three views.

Shortcomings:
- In the visualization of average rating and profit, when there're too many movies presenting, some dots may coincide, which makes it hard to know details on these movies under such view.
- There're some zooming issues.