P5 - Putting It All Together
Team Members: John Ho, Yi Li
Data Set: filtered_movies.csv (original was movies.csv)

Design Overview

        When looking through all the P5 data sets provided to us, we noticed that we had to deal with over 1000+ data cases. We found the movie data set to have some interesting attributes and proceeded in designing a multiview graph. Our goal was to create multiple, linked visualizations that included interactivity and coordinated views. The first challenge we faced was we had to decide on which attributes to use and how to clean the "movies.csv". We noticed many attributes had blank spaces, and the movie_title column had weird characters. So we decided to clean the data set by inserting a "0" for any blank spaces and got rid of all those weird characters, and then copied over clean columns in a new csv file titled "filtered_movies.csv".

        Next, we had to decide what types of visualizations in order to represent our selected attributes. We enjoyed P4 and thought using a dual-linked scatterplot with brushing and linking capabilities with other charts would be an awesome extension. Additionally, due to the amount of data cases we had to include, we thought it'd be the best way to be able to click on a single data point. After many scatterplot views, we decided that having a Number of Critic Reviews and User Reviews vs IMDb score would be the best viewing distribution. Finally, for ordinal attributes such as the year, we decided to go with a bar chart. For nominal attributes such country and language, we went with a pie chart.

        Examples of the analytical questions we thought of were: "What year were there movies with the highest IMDb score?", "How many top rated movies were from the USA vs other countries?", and "For a single movie, how many user reviews were there vs user reviews and how does that impact our frame of a good movie?". For our communication goals, we wanted to explore aspects of analysis and presentation with a multiview graph. We tried to design a nice interface that looks visually fantastic and could give great details through interactivity, but also give useful functions for analysis such as the ability to filter, sort, brush and link. We emphasized a sound, functional system implementation that had multiple coordinated views. We believe we've effectively combined all the ideas we learned this semester (bar charts, pie charts, scatterplots, brushing + linking, multiple analysis tasks, and details on demand). Additionally, we exhibited some creativity by integrating a **Movie API** that could convert our movie_title attribute into the main movie poster image to give our details on demand section the extra flare it deserves!

List of Analysis Tasks:

1) Retrieve Value - when **double clicking** a point on the scatterplot, retrieve values for movie attributes of: title, year, rating, duration, director, genres, budget, and gross if provided.
2) Filter - given concrete conditions of year, country, or language, a user is able to _hover_ over and click on a bar on the barchart or slice on pie chart, and then have those movies that match that filter highlighted on the two linked scatterplots below.
3) Compute Derived Value - when **hovering** over a bar on bar chart or a slice on pie chart, compute value for total count of movies in that category and have it show up on graph.
4) Sort - for the bar chart we have sorted all the data cases according to the year they were released, with an extra 'N/A' section for movies with an unknown release year. For the countries and language pie charts, we sorted the movies based on their each respective attribute and made separate slices, with a 'Other' category for movies which there aren't over 10 movies from that country, or over 5 movies in that language respectively.
5) Determine Range - given all the movie cases, we found the range of movies from a near 0 IMDb score to a perfect 10 IMDb score.
6) Find Anomalies - on the scatterplot you can visually see the outliers that are spread further apart from other data points, with some points being at extreme ranges on both ends of the X and Y axis spectrums.
7) Cluster - on the scatterplot you can visually see clusters of movies with a similar rating and number of critic reviews/user reviews. A user can **brush** over an area and view a similar cluster on the other linked scatterplot.
8) Correlate - the biggest thing with our multiview graph is that all graphs are correlated with each other! If a user **clicks** on a bar in the year chart, all movies released that year will be highlighted on both scatterplots. Similarly, if a user **clicks** on a country or language slice on our pie charts, those respective movies will also be highlighted. This can be extremely useful for analysis questions.

Individual Interface descriptions

Title: Wrath of the Titans
Year: Unknown
Rating: PG-13
Duration: Unknown
Director: Jonathan Liebesman
Genres: Action|Adventure|Fantasy
Budget: Unknown
Gross: Unknown

1. Information Section (Details on Demand)

This is the information chart for the visualization. When you double-click a data point in either of the two scatterplots, the detail information of the movie will be shown in this chart. The information includes the tile, year, rating, duration, director, genres, budget and gross. If any of the information is not available in the original dataset, it will be classified as unknown. Additionally, with the help of a Movie API script, we were able to give it our movie title and output the main movie poster if the user is connected to the internet!
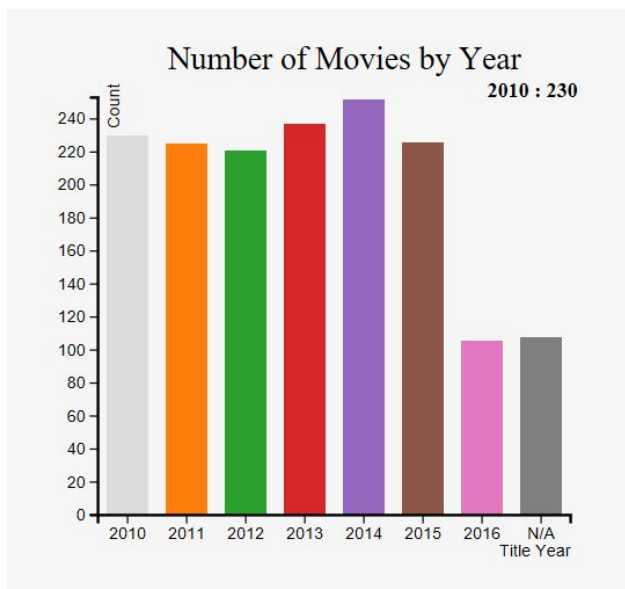

2. Number of Movies by Year

The Number of Movies by Year chart is an interactable bar chart in the visualization. It shows how many movies were published during each year. If a movie doesn't have a specific title year, it will be counted into N/A category. When you move the mouse over the bar, the bar will turn into a silver-gray color and the data will be shown on the top right corner of the graph. When you move your mouse away from the bars, the information will disappear and the color will be back to its original color. When you click on a bar of a certain year, the two scatterplots will highlight the movies that were published during that year.

Movies by Countries

UK : 136
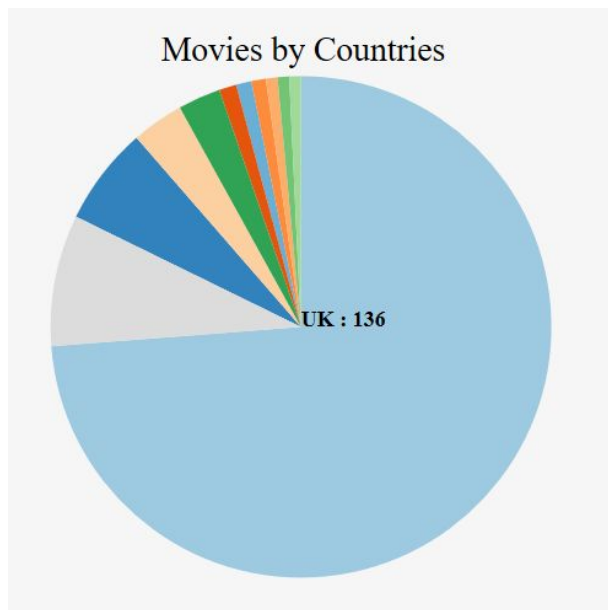
3. Movies by Countries Chart

The Movies by Countries chart is an interactable pie chart in the visualization. It shows how many movies were produced in different countries. The other category holds the movies was produced by a country that has less than ten movies listed in the dataset. When you move the mouse over the bar, the bar will turn into a silver-gray color and the data will be shown in the center of the pie chart. When you move your mouse away from the bars, the information will disappear and the color will be back to its original color. When you click on a certain country, the two scatterplots will highlight the movies that were produced by the country.

Movies by Language

French : 24
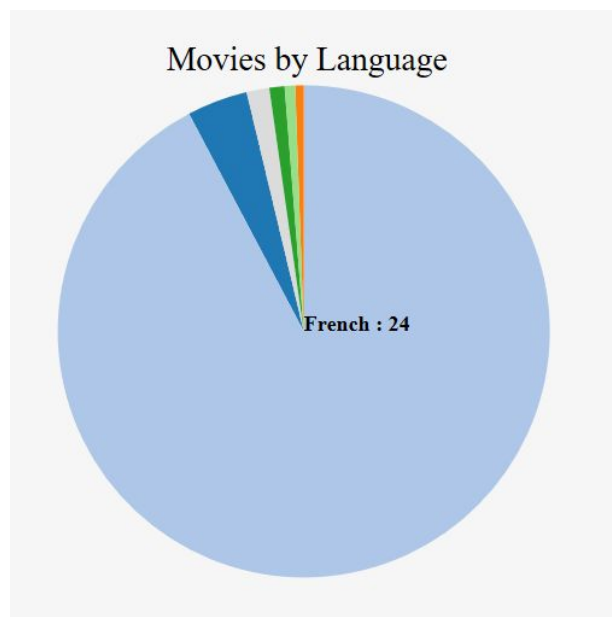
4. Movies by Language Chart

The Movies by Language chart is an interactable pie chart in the visualization. It shows how many movies were in different languages. The other category holds the movies was in a language that has less than five movies listed in the dataset. When you move the mouse over the bar, the bar will turn into a silver-gray color and the data will be shown in the center of the pie chart. When you move your mouse away from the bars, the information will disappear and the color will be back to its original color. When you click on a certain language, the two scatterplots will highlight the movies that were in that language.

5. Critic Reviews vs IMDb Score data plots
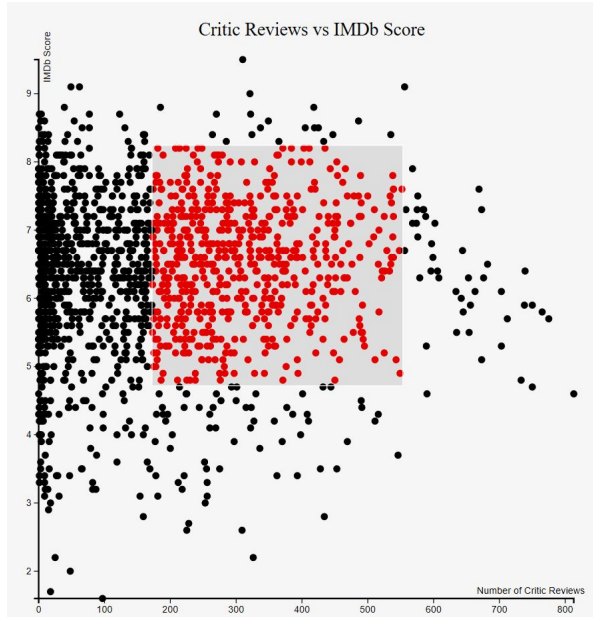
The Critic Reviews vs IMDb Score is an interactable scatterplot in the visualization. It shows the relationship between critic reviews and the IMDb score. When you double-click a certain data point at the graph, the current data point and data point corresponded to the current data point in the other scatter plot will be highlighted, and the information chart will display the details of the movie. When you highlight a certain region of the scatterplot, the data points in the other scatterplot, which correspond to the data points that are currently highlighted will also be highlighted. When you move the highlighted region, the highlighted points in the other scatterplot will also be shifted to the current points. When you single click on the graph, the brush will clear.

6. User Reviews vs IMDb Score data plot
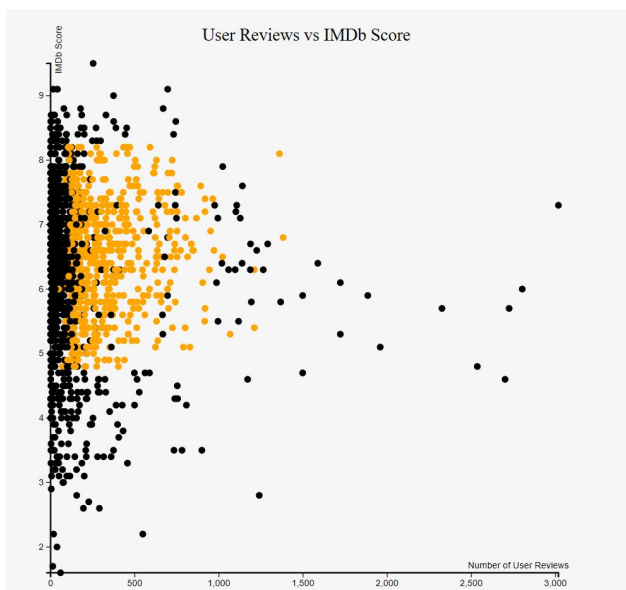
The Critic Reviews vs IMDb Score is an interactable scatterplot in the visualization. It shows the relationship between user reviews and IMDb score. When you double-click a certain data point at the graph, the current data point and data point corresponded to the current data point in the other scatter plot will be highlighted, and the information chart will display the details of the movie. When you highlight a certain region of the scatterplot, the data points in the other scatterplot, which correspond to the data points that are currently highlighted will also be highlighted. When you move the highlighted region, the highlighted points in the other scatterplot will also be shifted to the current points. When you single click on the graph, the brush will clear.

An entire Movie MultiView screenshot:

Movie MultiView

Title: After Earth
Year: 2014
Rating: PG-13
Duration: 100 min
Director: M. Night Shyamalan
Genres: Action|Adventure|Sci-Fi
Budget: $165000000
Gross: $187991439

Number of Movies by Year

Movies by Countries

Movies by Language

Critic Reviews vs IMDb Score

User Reviews vs IMDb Score