Eduardo Mestanza
Daniel Armstrong
Date: 11/30/2017
CS: 4460

Data set

After analysing the data sets we determined the that movies.csv would be the best choice for project 5.

List of analytical tasks

- What are the highest and lowest grossing films
- Which films had the highest and lowest budgets.
- Was the film a financial success
- What are the most popular film
- What are the most critically acclaimed film
- Who are the members of the cast of the film
- In general which type of cast member tends to be the most popular
- Which are the most popular genres of films
- Which genre of films are the most critically acclaimed
- Correlation between popularity and critical acclaim
- Correlation between popularity and successes.
- What are the elements that make a movie popular
- How do the visualization variables say about the success of a film
- Do genres have certain trends in regards to their budgets and gross
- If so those that have effect over their overall popularity and critical acclaim.
- Can an overall guide line be found to determine the success of a film.

Design Overview

The following visualization uses the movie.csv data set to demonstrate the level of quality of various films. Before we look at how this overall message is conveyed let's look at the individual components of our visualization and then put it all together. Let's start with the two main graphical components of the visualizer. The first graphical component is a scatter plot. The x-axis represents the budget of the film and the y-axis represents the grossing of the film. Each film is represented by circle and is colored in by which member of the cast is the most popular according to the number of facebook likes that individual has. The second graph demonstrates a histogram for the various films. The x-axis represents the imbd score of the film. The y-axis is the number of films in that histogram bin. Each film is represented by a circle and is colored coded by the

number of likes that film has in facebook. At the top is a box for selecting the genre of films that will be displayed by the visualizer. It is set to all genres by default but can be changed so that it displays specific genres. Notice that films can appear again in various genres. There are two main interactive tools in the design. The first one is the brush which can be used in either graphs. One can click and drag which will show a box being made. Any film that is inside that box is shown while the rest are hidden. Notice that this will cause the other graph to highlight the same films as the one found in the box. You can drag the box around the graph so you can highlight different films. The last element is the mouse over tool tip. If one hovers the mouse over a specific film a table will appear either on top or on the bottom. This table has details about that film, specifically it gives the actual values for each of the elements the graphs already represent. Also contained in the table are the movie's title, content rating (G, PG, PG-13, R), duration in minutes, year produced, as well as the country of origin.
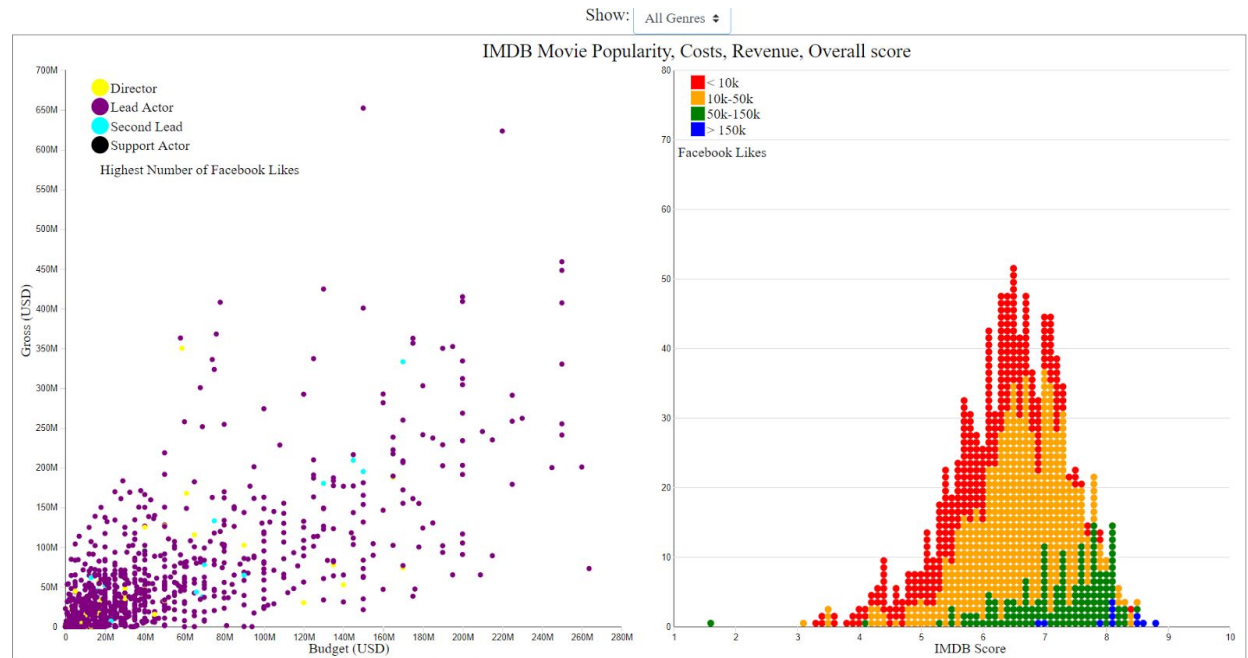
With these elements in mind let's look at how these elements come together for as an overall design. The visualization was designed according to the following steps: Overview -> Filtering -> Zooming -> Specificity. Initially one can see all of the data in both graphs which allows for simple questions about the data set to be answered such as the initial tasks presented in the list of analytical tasks. From here we use the selector to sort by genre thus reducing the clustering of data points. Here we answer simple questions about genre popularity and scores. Then we move into zooming. While the data points don't actually change in size the brush allows us to look at films at specific domain ranges which allows for comparisons between the different variables. Finally we reach specificity where our hover over table comes into play. From here we can take steps back as ask more complicated questions like looking for trends found in genres. Comparing various elements among various genres. Now we can look back at the overview and look for and determine how these variables work together and which ones are the more predominant when determining the quality of a film or what is expected of a film given the values given to the variables explored. For example is a high budget needed for producing successful action films?

Some clarifications

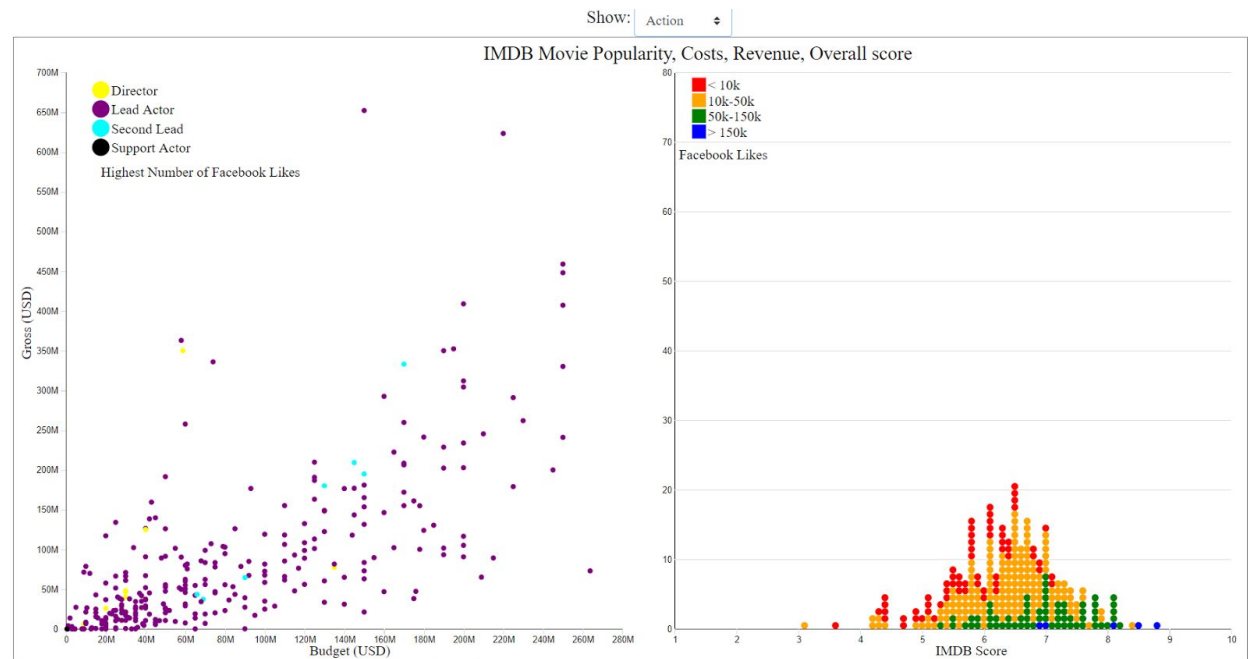Given the number of variables found in the movie.csv file we decided to filter out many of them such as the films color, keyword descriptions, aspect ratio etc. These elements didn't seem to add much towards determining the quality or success of a film, Also they would be too difficult or lack information for us to represent them whether using visuals or even text.
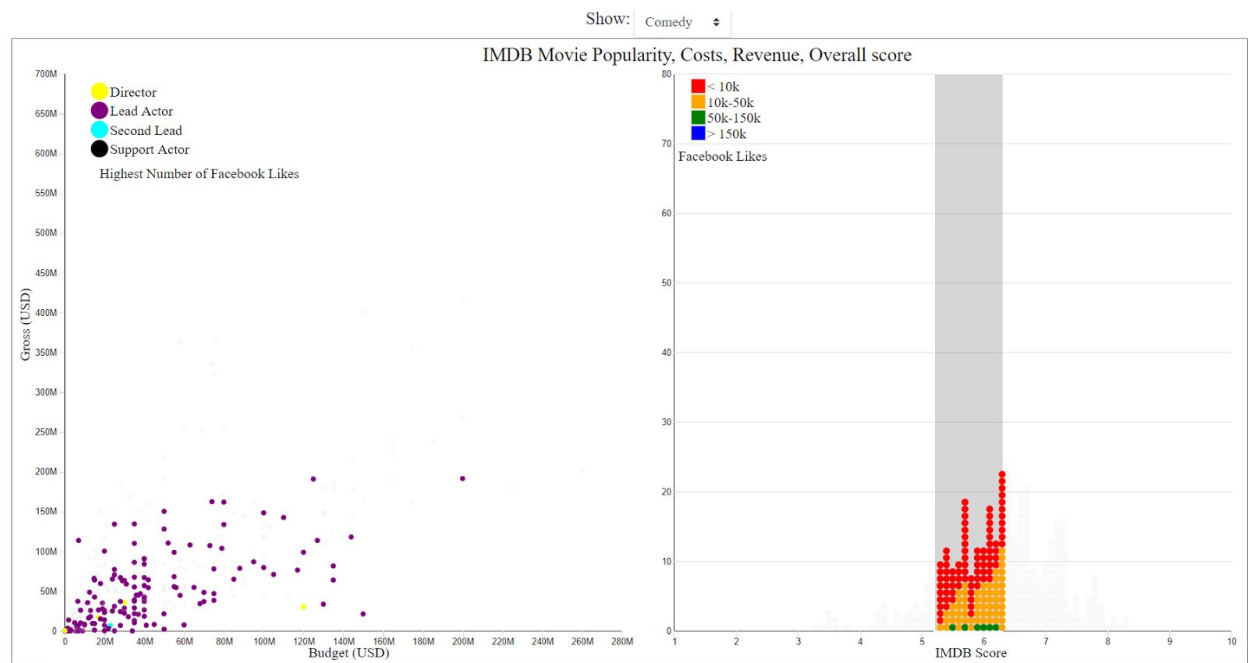
UI Snippets

Overview:

Filtered

Zooming:

IMDB Movie Popularity, Costs, Revenue, Overall score

- Director
- Lead Actor
- Second Lead
- Support Actor

Highest Number of Facebook Likes

- < 10k
- 10k-50k
- 50k-150k
- > 150k

Facebook Likes

Gross (USD) vs Budget (USD)

IMDB Score

Specific details:

IMDB Movie Popularity, Costs, Revenue, Overall score

- Director
- Lead Actor
- Second Lead
- Support Actor

Highest Number of Facebook Likes

- < 10k
- 10k-50k
- 50k-150k
- > 150k

Facebook Likes

Alvin and the Chipmunks: Chipwrecked

Name of director	Year
Mike Mitchell	2011
Duration (min)	Content Rating
87	G
Gross (USD)	Budget (USD)
133103929	75000000
country	Movie Facebook likes
USA	0
Lead Actor	Content Number of likes
Amy Poehler	1000
Second Lead Actor (USD)	Number of Likes
Jesse McCartney	1000
Supporting Actor	Number of Likes
Lauren Gottlieb	733

Gross (USD) vs Budget (USD)

IMDB Score