Rishi Raj and Yashveer Singh

CS4460 Final Project Report

Dataset: Movies

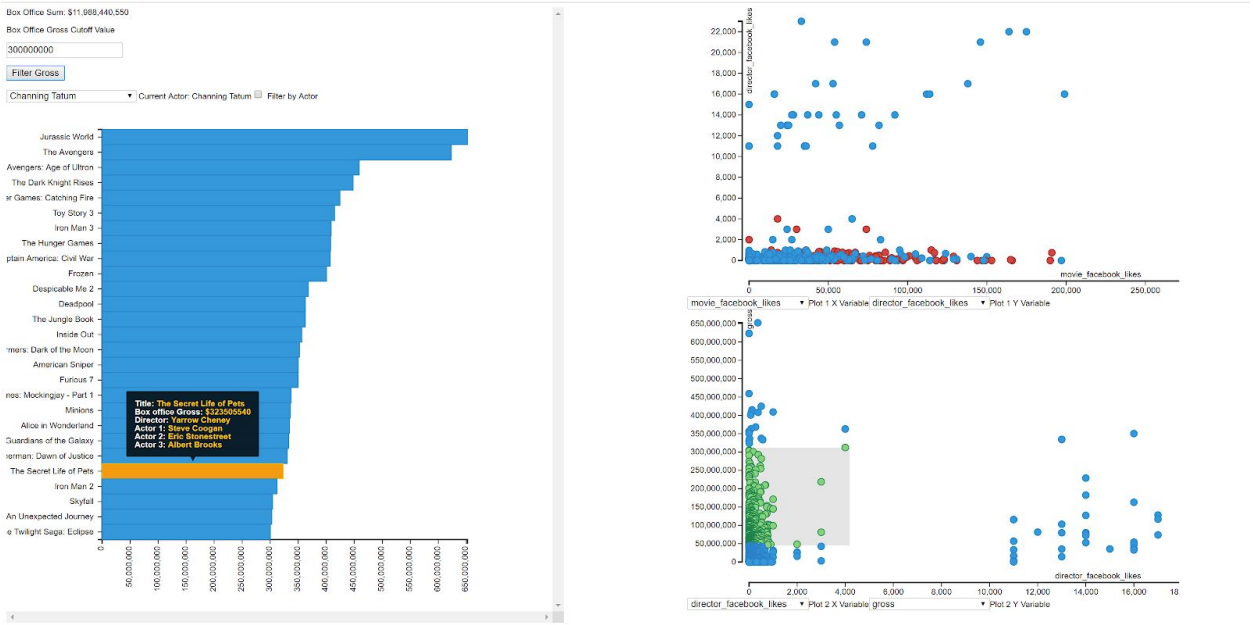
Analytic Tasks: Finding Extremum, Correlation, Retrieving Values, Filtering, Sorting, Clustering

Design Overview

We elected to use the movie dataset. Before we began, we asked ourselves which questions we wanted to answer about the dataset. The following were several of the questions we asked or felt that someone else might ask:

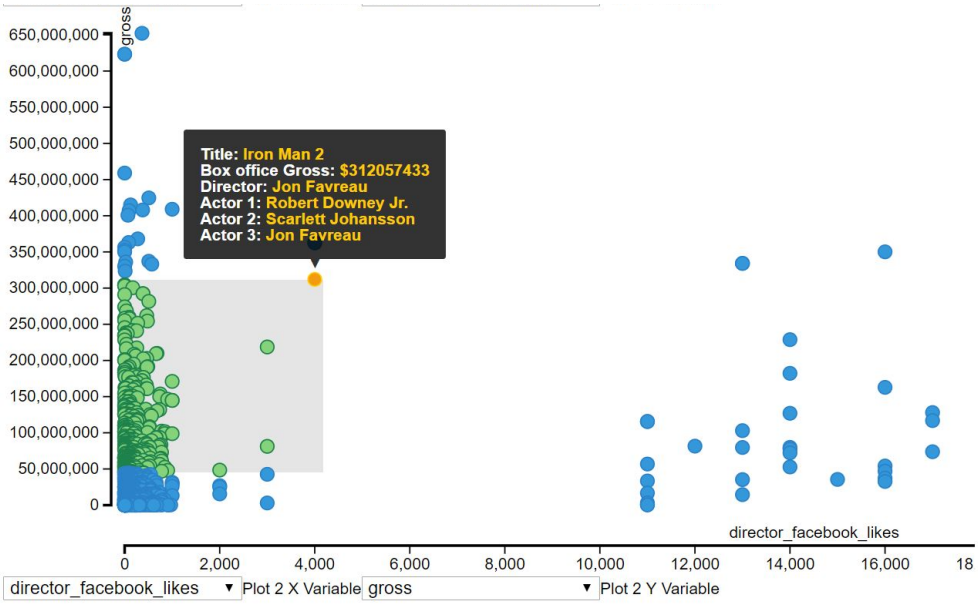
1. How does the number of facebook likes for the various actors, cast, and director, relate to the gross income for the movie?
2. How does does the budget for the movie affect the gross income for the movie?
3. Which movies did x actor star in?
4. Who was involved in the highest grossing and lowest grossing movies?
5. Which movie has the greatest gross income?
6. Which movie has the highest budget?
7. Which factors influence the movie's gross the most? For example, does the director's facebook likes have more of a relation with the gross than the top billed actor?
8. What is the most money a movie has made?
9. What is the largest budget spent on a movie?
10. Which actors consistently star in high grossing films?
11. How many films have grossed over $__ million?

We felt that we could answer the questions we had with an interactive bar chart and two side-by-side scatterplots, and some well-formatted tooltips. When you first open the page, you see a bar graph and two scatter plots.



The initial bar graph plots movie titles versus their gross revenue. There are a large number of movies, so you have to scroll down to see them all. However, this is just meant as a light overview of the data. We offer a number of filtering and sorting operations to narrow down the bar graph. In particular, we allow you to only look at bars above a certain gross revenue. Also, you can choose to look at movies corresponding to a particular actor. We have a drop down menu that lists various actors, and when you select one, it filters the bar graph to only include the movies that have that actor in it. Additionally, the bars are always sorted in decreasing order according to gross revenue. Our bar graph scatterplots have tooltips, where you can hover your mouse over a data point and see relevant information. Our scatterplots allow you to change the variables plotted on the x and y axes, so you can make many different comparisons. We chose to have two scatterplots next to each other to make it easier to see if one variable has a greater correlation with a fixed variable then some other one. For example, we could plot the director's facebook likes against box office gross, and the top actor's facebook likes against box office gross, and see if one has a stronger correlation.

Using the many elements and options available in the visualization that were described above, users can answer many questions about the dataset effectively. For questions 1 and 7, a user can assign the y axis of one scatterplot to gross income. Then they can cycle through the facebook like counts for the director, lead actor, and cast and see how the values correlate. They can observe if the points cluster together in a certain part of the graph or if the points are evenly distributed suggesting no correlation at all. Because we have side by side scatterplots, we could see if "director Facebook likes" has a stronger correlation with gross revenue than "actor_1_facebook_likes".

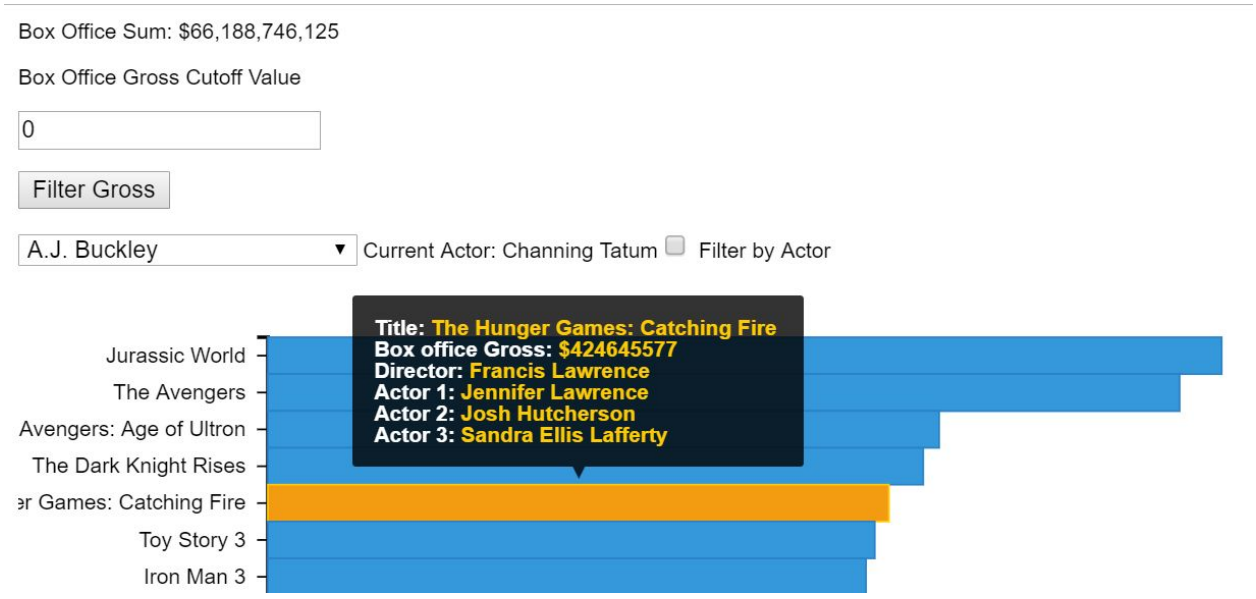


To answer question 3, all you would have to do is check the box that says “filter by actor”, select the actor from the drop down menu, and then look at the movies on the y-axis. To answer question 4, you can just hover the mouse over the largest or smallest rectangle in the bar graph and the tooltip will show the director and top 3 actors.

Analytic Tasks: Finding Extremum, Correlation, Retrieving Values, Filtering, Sorting, Clustering, Computing Derived Values, Find Anomalies

Finding Extremum - Our sorted bar graph in combination with tooltips make it easy to find extreme values in our visualization. We can easily find the highest and lowest grossing movies, and the exact amounts associated with them. Our bar graph is naturally sorted, so if we uncheck the “filter by actor” box and set our gross cutoff value to 0, it will list all the movies in descending order by gross revenue. We can easily see the titles of the movies on the y-axis, and then hover the mouse over the bar to get information about the movie.

Filtering- We wanted to give many filtering options to the user to maximize insights into the data. In the beginning, our bar chart has no filters and contains a list of every single movie in the database. The resulting list to far too long to draw insights. Users can filter that chart with a box office gross cutoff value and filter by a certain actor’s filmography. The graph will redraw itself whenever either value or both are changed. We also embedded the bar chart in a scrollable div and made it dynamically resizable so it can handle any bar chart size as a result of a query.
In our scatter plots, users can take advantage of our brushing features to filter out linked points between two scatter plots. Users can highlight some points in one scatterplot and discover where these points lie in another attribute comparison on the other scatterplot. The linked points may cluster in another part of the plot or may be widely distributed in the other scatterplot.



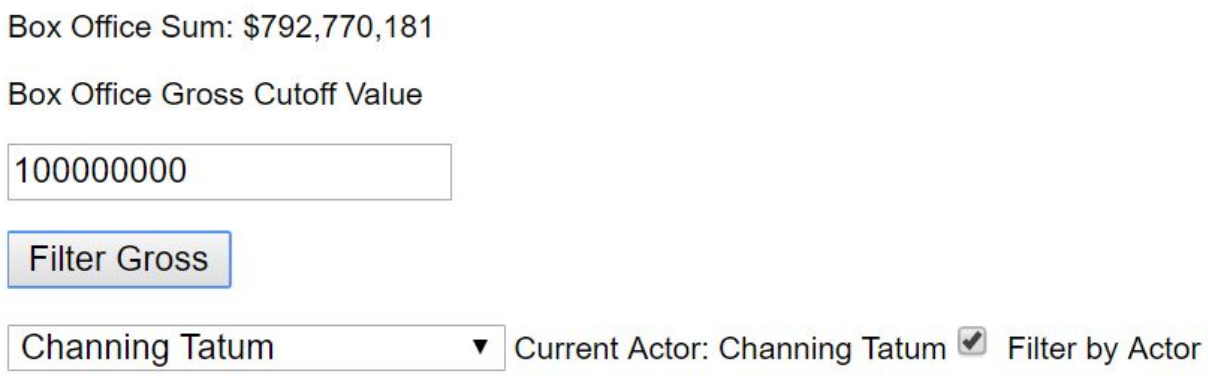
Correlation- Our two scatterplots make it easy to determine correlation between two data attributes.

Sorting- Our bar chart is sorted in descending order of box office gross. This is typically one of the most sought after data points and it made sense to presort the bar chart this way.

Clustering- Our scatterplots are effective for spotting data clusters. We can easily observe where most movies lie in relation to two data attributes. Conversely, this can also let us find outliers from data clusters very easily.

Retrieving Values - Our tooltips make it simple to retrieve values about data cases. We simply hover the mouse over a rectangle or circle, and we can immediately see the title, gross, actors, and director. So, if you wanted to know who directed the highest grossing movie, you would just hover your mouse over it.

Compute Derived Value - When you filter the bar graph by actor, we show the total box office gross of all of that actor's listed movies.



Find Anomalies - Our scatterplots make it easy to see outliers if there are any.