P5 Description

Team Members: Kevin Byrne, Brian Glowniak

Dataset Chosen: Movies

Design Overview:

When exploring the dataset initially, the most pressing question we had was what attributes are correlated with movies that are profitable and which are correlated with movies that perform poorly at the box office. These questions were also extended to discover which attributes are typical of movies with low or high budgets. By using this as our main paradigm for design, our first attempt was to use a parallel coordinates graph with some color encoding to answer these questions. We quickly concluded that there were far too many data cases to allow for a visualization like that and occlusion was a huge problem. Our next attempt was a scatter plot which worked out far better. While there was some occlusion of the data, there was far less and the distribution of data cases worked out far better. This also allowed us to put in a line above which were movies that turned a profit and below which were movies that lost money.

We then extended this scatter plot by linking it to a bar chart that represented either the genres of the movie, continent the movie was filmed on, or the rating of the movie. It was linked through a brush that the user creates and dynamically updates as the brush moves. Everything that updates through user interaction was designed with animated transitions to allow the user to easily see what is changing and how the values are increasing or decreasing. Both the scatter plot and bar chart support details on demand when a user clicks on them so that the user can explore specifically what they want to. Past our main communicative objective of the gross vs. budget of various movies, we did not try to limit the user. We do not assume what the user is interested in, so hopefully the vis answers as many questions as they can develop. By not restricting where the user can brush or how large the brush can be and by giving them three different attributes of the data to link in the bar chart, we tried to give the users as much freedom as possible.

Supported User Tasks:

Our visualization supports the following user interaction tasks: select, encode, abstract/elaborate, filter, and connect. For select, we allow users to select a certain bar from the

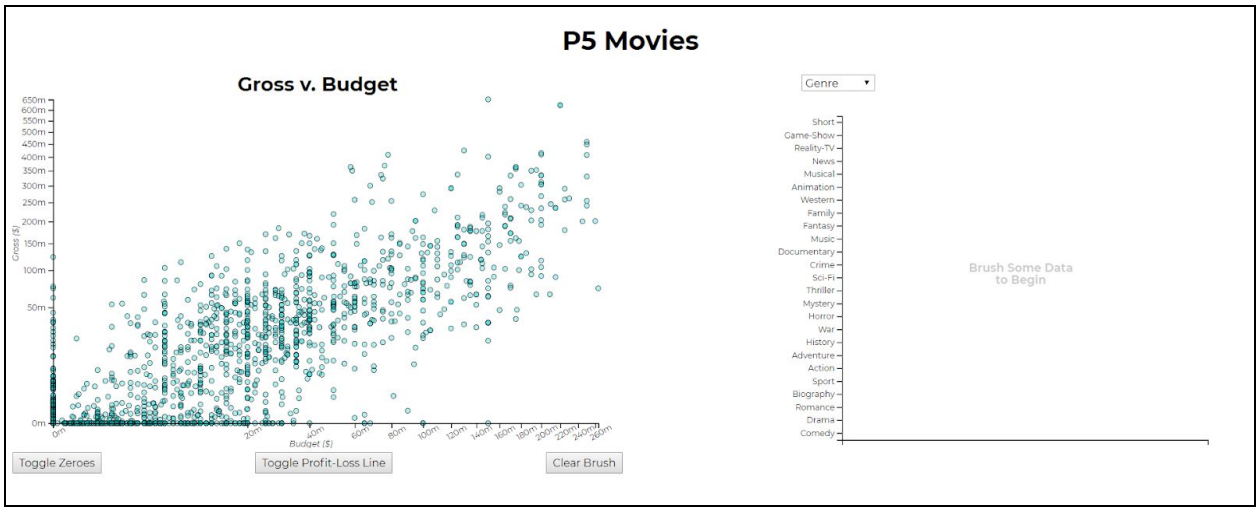
bar chart on the right and it will remain selected as they move their brush around. This allows the user to keep track of a certain genre, continent, or rating they are particularly interested in and see how different budgets or grosses affect it. Encode is simply implemented through allowing the user to change the color encoding of the circles in the scatter plot. Opacity is used to indicate where circles overlap to try and avoid issues related to occlusion, but we allow the user to toggle a profit-loss line where the circle's colors are now encoded to show if the movie had a higher gross than budget or not.

Details on demand is the method in which we chose to allow the user to accomplish the abstract/elaborate task. On both the scatter plot and the bar chart, the user can click on a circle or bar to reveal more information. For the circles, a details box appears with all of the other attributes of the data not shown in the visualization. The user can click on up to 3 movies at a time to compare the values between them. For the bars, the user can also click on them to reveal further details in a box below the chart. For genre and rating, it just shows the user the exact count, and for continent it shows the component countries and how many movies were filmed in each. All three of these also dynamically update as the brush is moved on the scatter plot.

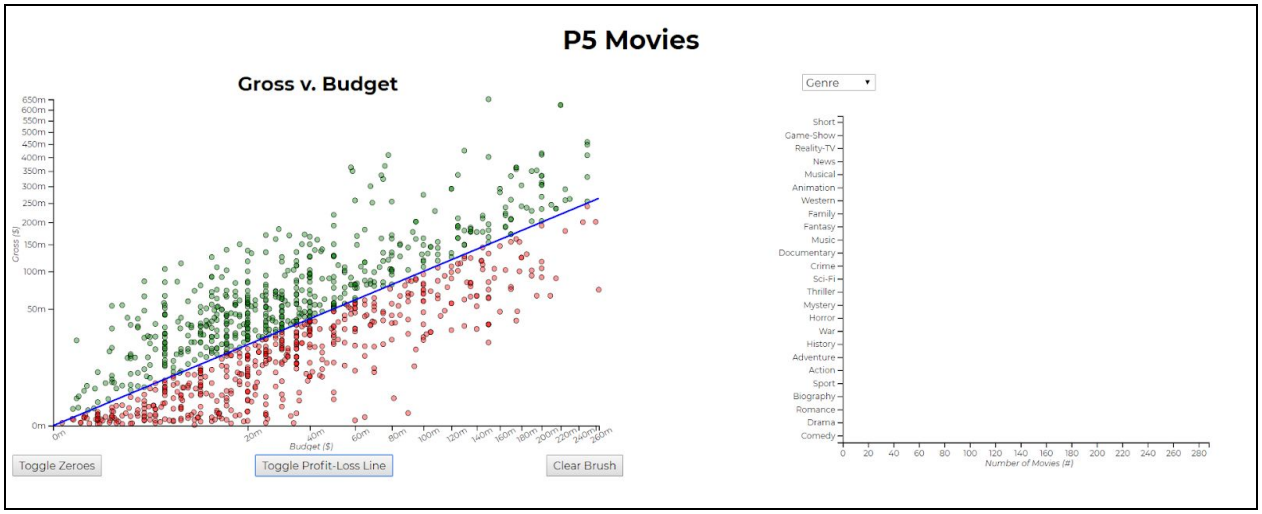
The user is given the ability to filter through two different mechanisms. The first was created out of necessity due to the data having a lot of missing values. For any movie that did not have a value for its budget or gross, we give the user the ability to filter out those circles from the scatter plot. The second method is with the brush that the user can draw over the scatter plot. This updates the bars on the right and allows the user to create the bars however they please.

The last user task is connect. This is intimately related to all of the previous tasks, but it is explicit in the brushing and linking. The chart on the right is empty until the user creates a brush to link some data to it. Since our main narrative of the visualization is the profit of a movie, allowing the user to explore correlations through various other attributes, we have the linked bar chart. The user can choose from three different attributes to explore the data through and to answer any potential questions they may have about movies with various values for their budgets and grosses.

Explanation of Specific Aspects:



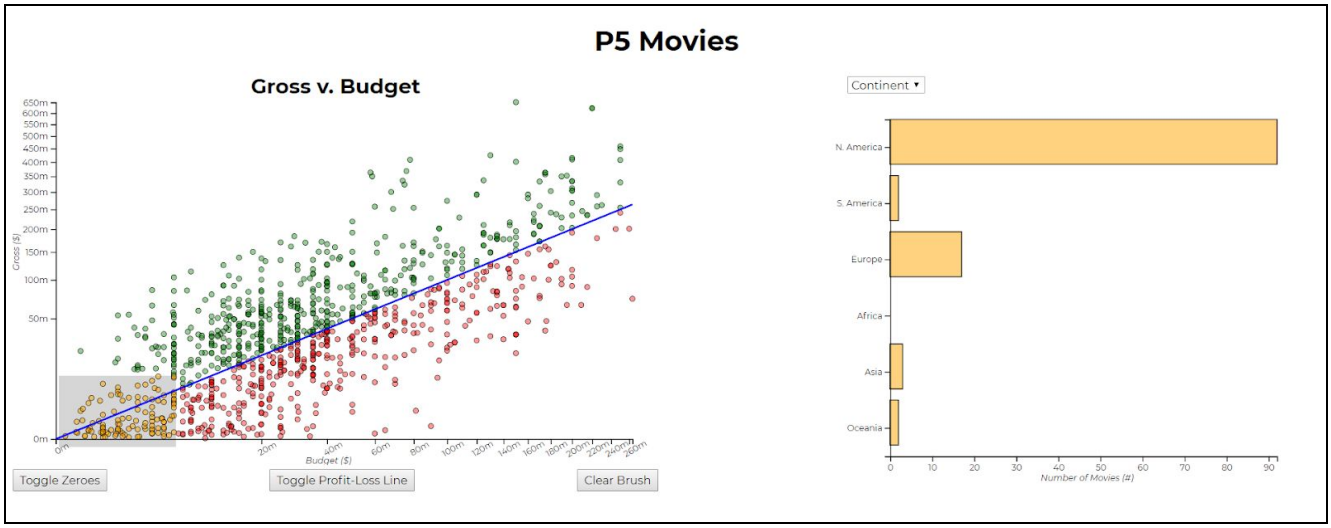
This is the screen that appears when a user first loads up the project. It depicts the two main components of the visualization: on the left, a scatter plot charting budget vs. gross, and on the right, a bar chart depicting the number of movies with certain attributes. Since there was overlap between a decent amount of the data points, we scaled the axes exponentially and added some translucency to each individual encoding. The bar chart starts off blank, as it is populated once the user brushes data on the scatter plot. This is explained to the user by means of a small message in the bar chart that says “Brush Some Data To Begin.” A drop down menu allows the user to select from Continent, Rating, and Genre for the attribute to focus on.



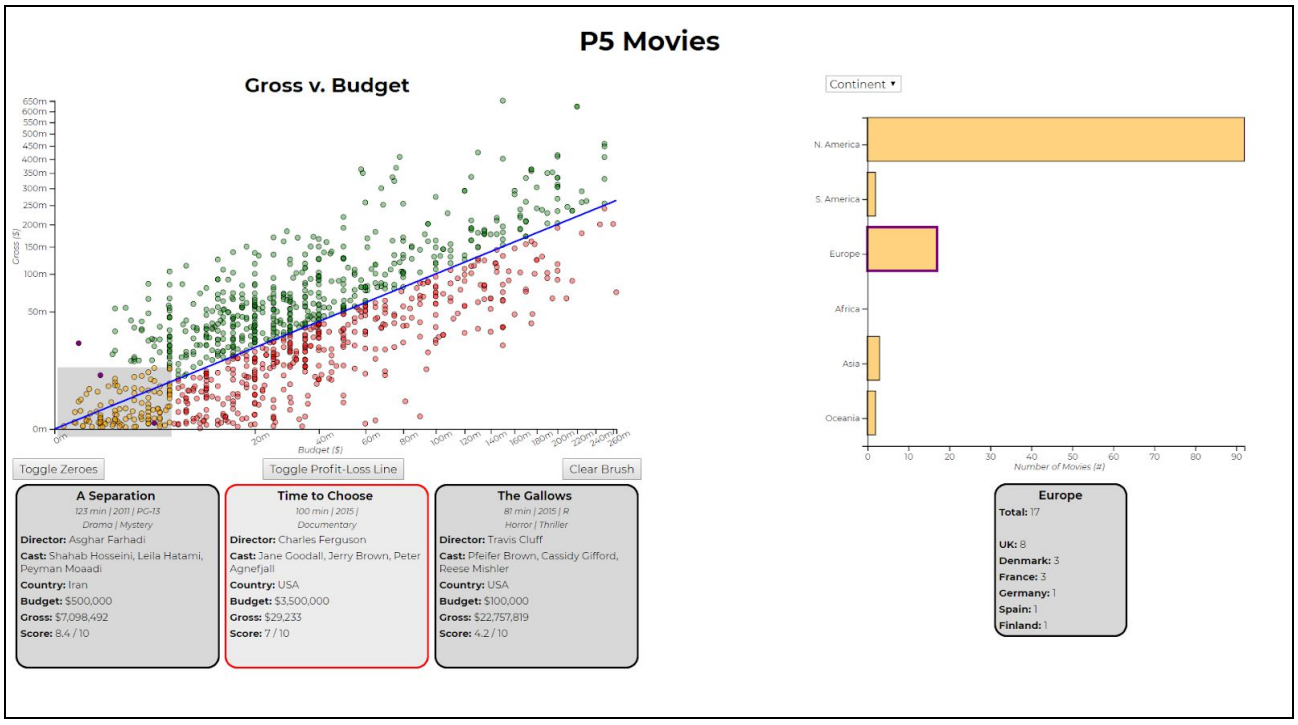
This second screenshot depicts the functionality of the first two buttons located underneath the scatter plot. The first, “Toggle Zeroes,” allows the user to hide or show the data cases on the plot

that either have a zero budget or a zero gross. This was done because we wanted to include all the data provided, but these specific elements were reported erroneously and we wanted to give the option to hide them when exploring the data.

The second, "Toggle Profit-Loss Line," places a blue line on the plot that shows where budget equals gross. Any points above this line are colored green (as they made a profit), and any points below the line are colored red (as they lost money), and this allows a user to more easily see which movies were successful and which were not as they explore the vis. These color changes are done over a small duration to clearly show the split.



This screenshot depicts the brushing capability of the vis. When a user brushes a set of points on the scatter plot, they are highlighted yellow, and the number of points selected are split up and depicted in the bar chart by the focused attribute from the drop down menu. The bars animate and change dynamically as the user moves and adjusts the brush. An example use case of this is a user wanting to see the most common continent for movies with lower budgets. Using the "Clear Brush" button removes the brush and fully resets the bar chart.



This image depicts the details on demand functionality of the vis. For the scatter plot, a user can select up to three movies to see more attributes and to compare them. These are highlighted purple on the plot (independent of whether the selected points are brushed or not). In the same vein, on the bar chart, a user can click on a specific bar to the number of movies that bar represents (i.e. how many movies brushed were made in Europe?). If the focused attribute is Continent, the details box also includes breaks it down by countries within that continent.

A user can deselect a movie by either clicking the details box (which changes color and border when a user hovers over it) or by clicking on the same encoding used to select it. Clicking the "Clear Brush" button will also remove the details box below the bar chart.