Sooah Moon, Sally Park

Professor Endert

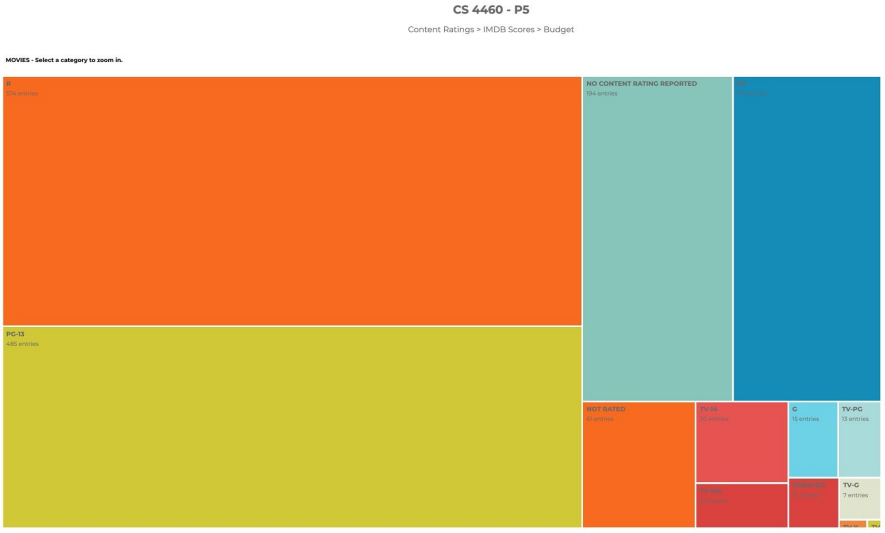
CS 4460

2 December 2019

Program Overview

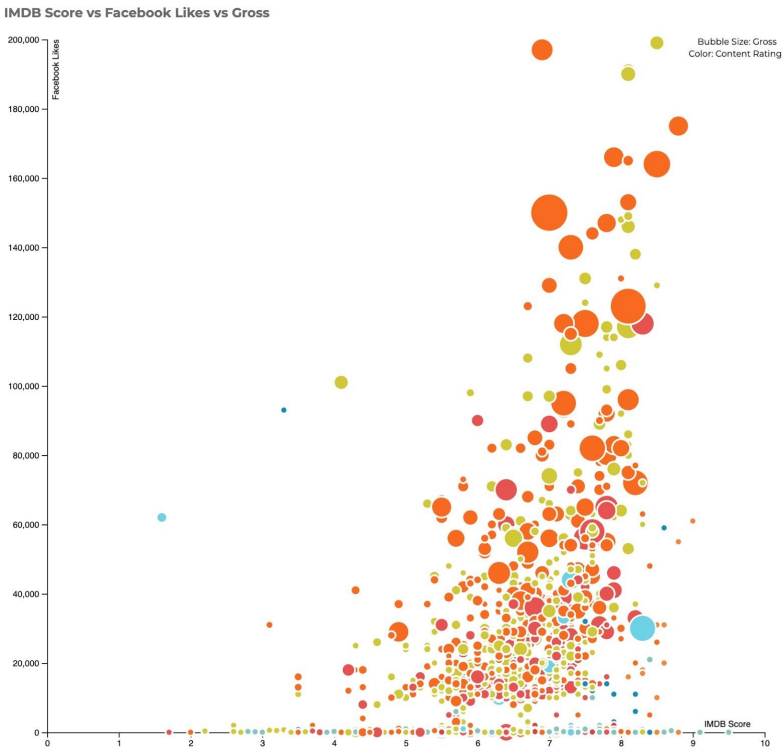
For our final programming assignment, we created a visualization using the Movie dataset provided using two layouts - a treemap and a bubble plot. Both layouts support a large number of analytical tasks, such as looking over the dataset to find interesting information, analyzing to find patterns and outliers by displaying all relevant categories, filtering by selecting a category, adjusting the level of abstraction by allowing the viewer to unfold categories, finding anomalies by displaying all categories so the viewer can clearly see outliers, and highlighting associations and relationships by implementing a zoomable treemap. Additionally, the viewer can retrieve a value by hovering over a data point for more detail in the bubble plot and the clustering by showing points allows the user to see groups of similar items.

For the visualization, we explored parts of Shneiderman's visualization mantra for visual analysis (overview first, zoom and filter, then details-on-demand) to guide the viewer through



visual analysis by allowing them to gain an understanding for the overall pattern, then focus on areas of interest using interactive features, and finally progress deeper to locate details of interest. The treemap begins by showing an overview of all ratings possible for movies, laid out in a tree format to successfully show hierarchical information. Each block's size is encoded based on the number of items in the data set pertaining to that content rating. The user is able to select a block and have it unfold into another visualization that is filtered by IMDB ratings. Eventually, they can filter even further to view another treemap with rectangle's size encoded to the budget movies with that score .

When exploring the bubble plot layout, the viewer can first see an overview of a movie's IMDB score, the number of Facebook likes it received, and its gross revenue. While the size of



the bubble is encoded to the how much the movie earned, its color is corresponds to the content rating. When hovering over a data point, the viewer can see the movie title and its gross revenue. By interacting with the data, we hope that the viewer is able to answer any questions they might have while exploring the visualization. For example, a viewer might wonder whether kid-friendly movies always have higher satisfaction scores, why might that be, and whether it might be related to the amount they spend on a budget. They could go on to explore the bubble

plot, where conclusions can be drawn on why Twilight was poorly received by the public but still made a significant amount of revenue, and what movies seem to be well-received by the public on all platforms despite making very little revenue.

At any point in the process of exploring the treemap, the viewer can select the header bar to zoom out, which is noted in the navigation bar in addition to a message for the zoom in option. It was important for us that our visualization be intuitive and comprehensive as well as visually pleasing. As a result, not only were we able to clearly and accurately represent information, but the viewer was also able to participate in user interaction to figure out what the information means. While unreported data could possibly influence making conclusions taken from our visualizations, we felt it still effectively communicated the objectives we had.

Something to note about both visualizations is that there are advantages and disadvantages to each representation. When using a treemap, while it does make the exploration of large data sets fairly easy, sometimes it can be difficult for the viewer to keep track of hierarchy when exploring the dataset. On the other hand, bubble plots are useful when visualization situations and patterns, but provide little insight on the why and how. We felt that pairing both layouts together displayed different ways to view the same dataset.