Proposal: Was the Movie Worth it?

Goal:

Our visualization tries to investigate multiple relationships between a movie's investment and its outcome under a given release year. The visualization will show the following 3 relationships of a movie released in a given year: Budget vs. Rating, Budget vs Revenue, and Revenue vs Rating. Only one relationship is displayed at a time under a user's choice.

Data Pre-processing:

Our data comes from tmdb_5000_movies.csv in the kaggle dataset "TMDB 5000 Movie Dataset" at https://www.kaggle.com/tmdb/tmdb-movie-metadata/data. The dataset offers interesting film statistics that we plan to visualize as a scatterplot. Specifically, we will use the columns "budget", "revenue" and "vote_average" together with the film title and the year of release. In our data pre-processing, we will use Excel to extract out the relevant columns and use filtering to only look at non-zero "budget" and non-zero "revenue". The 0 value entries are left out because they indicate missing data. We then format the release date and sort the data by release year, and only look at the data during 1990-2016, saved to a new CSV file.

Visualization Plan:

Our visualization would be a combination of scatterplots and a scrollbar, using mainly d3. We would employ a drop-down selector to allow users to choose among the three relationships presented in the goal section of the proposal. Each choice shows its corresponding scatterplot that visualizes the relationship it represents. A scrollbar is placed under the plot to let the user select which release year they want to focus on, or see the general trend of how a selected relationship changes with time. Each dot on the scatterplot would show the movie it represents if the user puts the mouse over the dot, allowing the user to identify the movies.
Our visualization looks like a combination of the following picture:
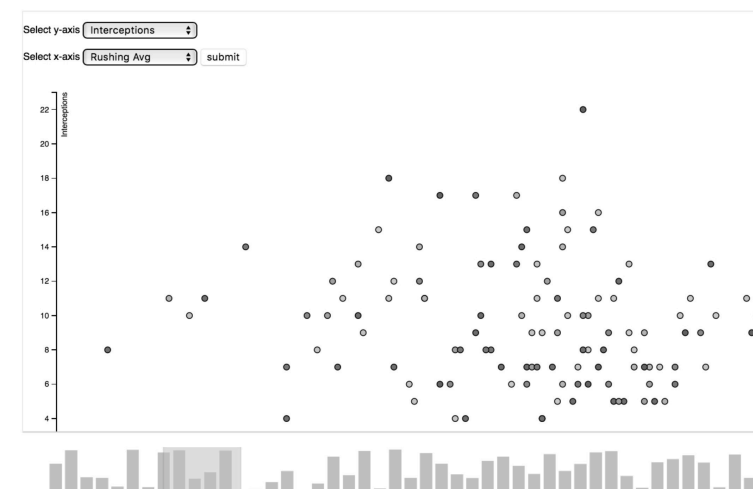
Note that we do not use grayscale encoding in our visualization, and the drop-down selector/scrollbar designs are placeholders. The plot shown here is just a close approximation of what we are trying to do.