

Ranjan Sikand, Barath Palanisamy, Viswanath Chadalapaka
ECS 163
March 16, 2020
Final Project Write Up

Overview - Dataset and Description

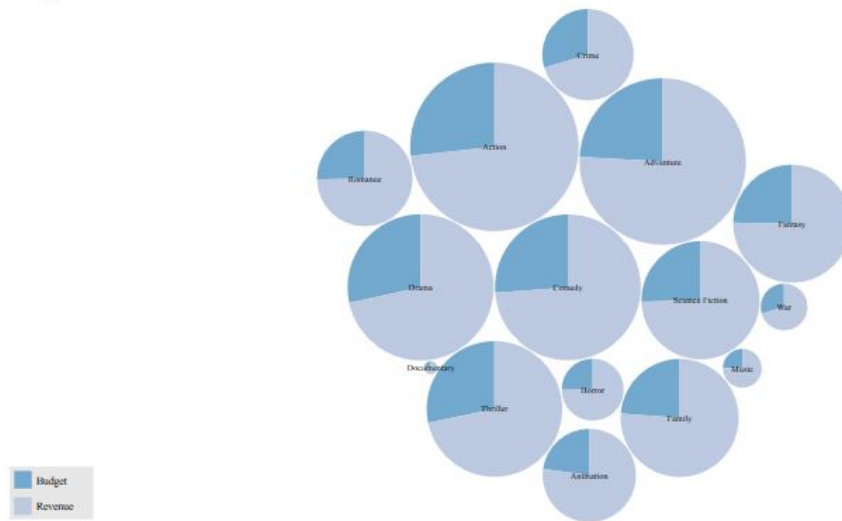
Our system is defined primarily by the visuals and interactions between three key components: A Bubble Chart, a Force-Directed Graph, and a Parallel Coordinates Plot. The dataset we have decided to use consists of 5k movies, and can be found at the following URL: <https://www.kaggle.com/tmdb/tmdb-movie-metadata>.

From the conception of our system, we wanted users to be able to intuitively build connections between the success of a movie, its various attributes, including its “correlation” to other movies with respect to the similarity between various facets which go into movie development, including cast, crew, etc. Thus, the visualization could potentially be geared towards those in the movie industry, for example, who would like to examine what exactly goes into the success of a film. We implemented the requirements of data filtering, through interactions between the parallel coordinate plot, interactive remapping of marks and channels in interactions with the graph, and data selection interaction by filtering the data passed to the parallel coordinate plot through interaction with the force-directed graph. What follows are design-choice level descriptions for each of the visualizations employed.

Bubble Chart

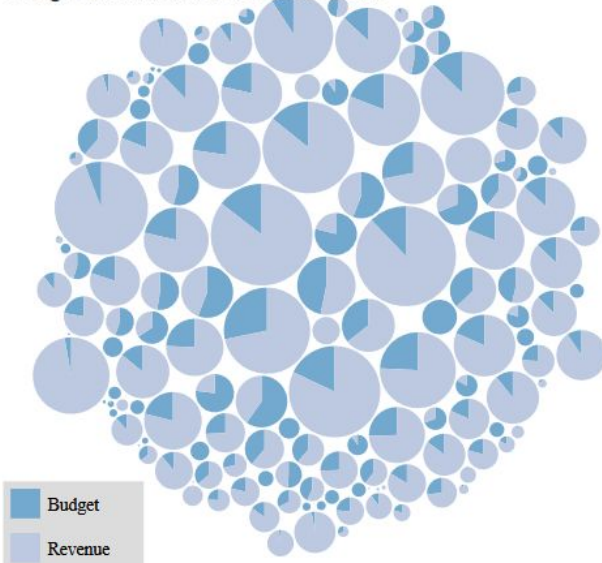
Our ultimate result was a bubble-chart of pie charts that depicted the budget and revenue of movies across genres, a radial graph showing the five most profitable movies in a given genre, a force-directed layout showing the links in genre, keywords, or cast and crew among movies, and a parallel coordinate plot showing trends in runtime, release date, budget, and audience score.

Budget and Revenue Across Genres

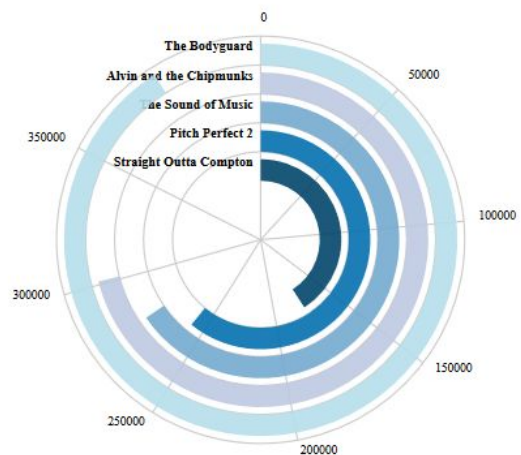


The bubble chart above depicts the relative sizes of each genre, based upon the sum of the budget and revenue of all films in that genre. This helps paint trends among film categories better; while drama is the largest category by a significant margin, it is only slightly bigger than action and adventure, and even thriller approaches it in size, despite having far fewer movies. Clicking on any of the genres expands the bubble to give more detail on the films in that category, and hovering over any mark generates a tooltip that indicates the highlighted data value. For example, clicking on the music category renders the following view:

Budget and Revenue for Music Movies



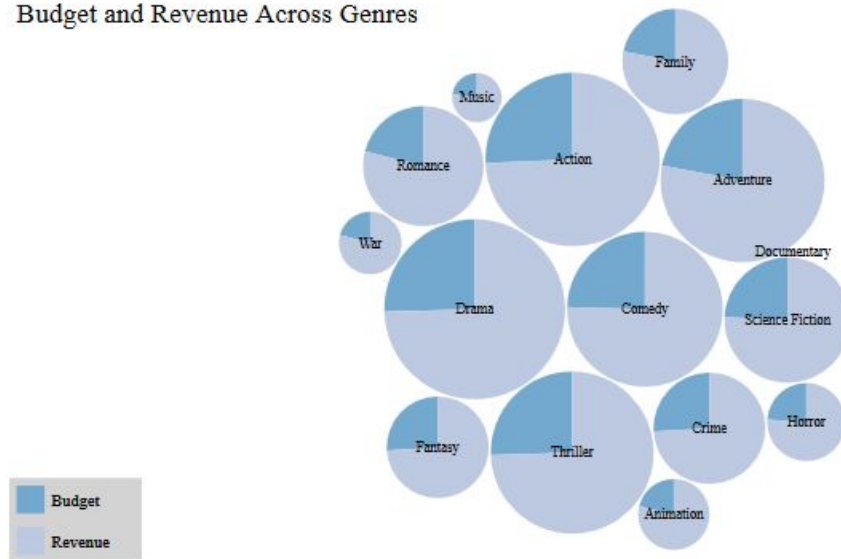
Most Profitable Movies in Genre (in thousands)



This view depicts the budget and revenue of all films in the Music genre. This ranges from movies like “The Bodyguard” starring Whitney Houston, which follows the life of a popstar, to “The Sound of Music,” a more traditional musical.

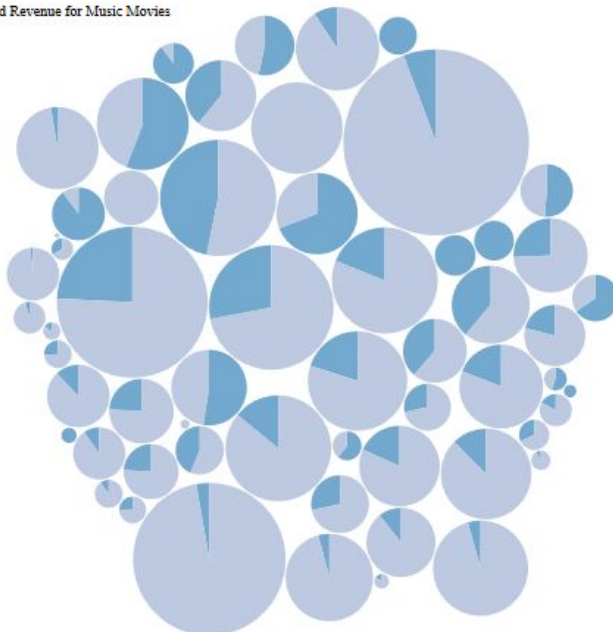
The bubble chart can be filtered through interaction with the parallel coordinates chart. For example, if I wanted to only look at films that released before the year 2000, I could simply brush along the Year axis, which would generate this view:

Budget and Revenue Across Genres

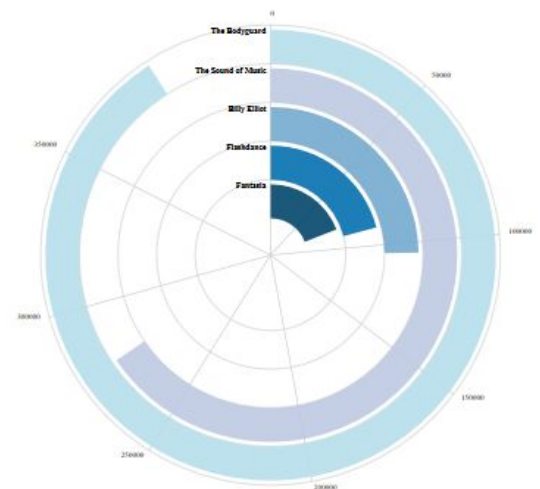


With this filter, the action and adventure genres are less dominant over the chart. The fantasy genre also takes a sizable hit, while the Drama category maintains its dominance over the rest of the chart. Clicking on the Music genre again brings us to this view:

Budget and Revenue for Music Movies



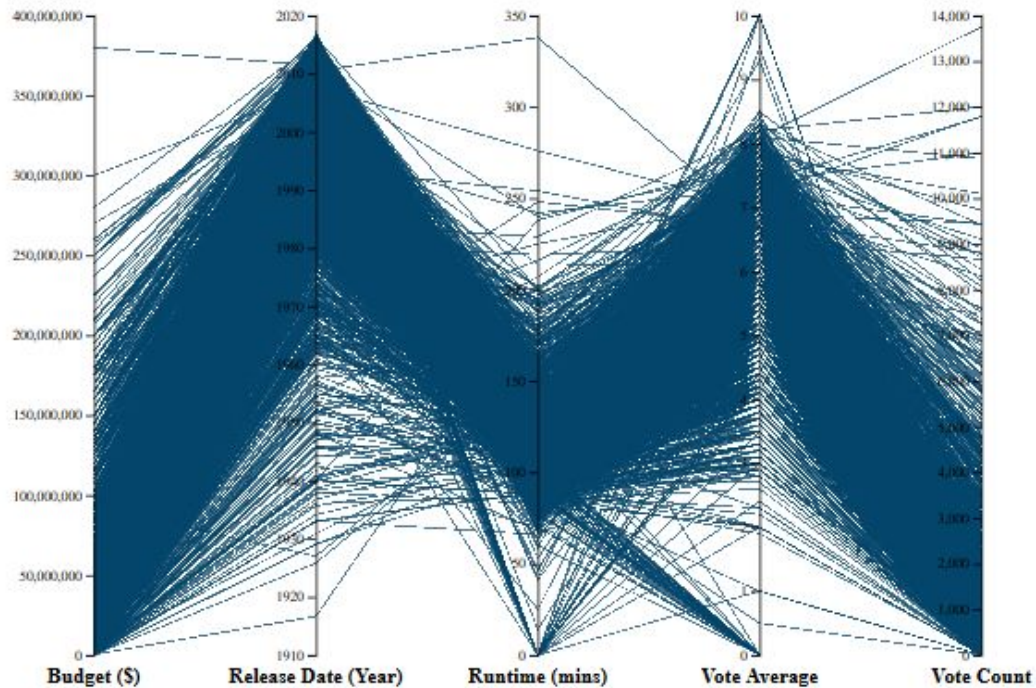
Most Profitable Movies in Genre (in thousands)



“The Bodyguard,” released in 1992, remains at the top of our rankings, but our entries at 2, 4 and 5 have fallen off the chart. This allows for the classic Disney film “Fantasia,” released in 1940, to sneak into fifth place with a profit just over \$81 million.

Parallel Coordinates Chart

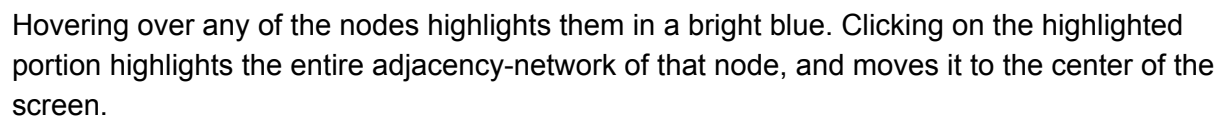
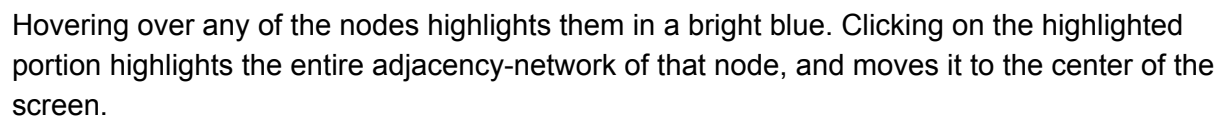
The parallel coordinate chart plots the attributes of over 4000 films, showing trends in 5 separate categories. For example, the vast majority of films in this dataset were released after 1960, and have a runtime between 60 and 180 minutes. However, one film can be seen shooting into the 340-minute range: 2006's "Carlos," boasting a runtime of 5 hours 34 minutes. On the other of the spectrum are movies like "A Charlie Brown Christmas," totaling only 25 minutes from title sequence to end credits.



A brush tool allows users to filter the data, highlighting only the selected portions and regenerating the other two visualizations. Also, as the user filters, paths that line in the brushed portion remain highlighted blue while the rest of the lines are greyed out.

Force-Directed Graph

On loading, the force-directed graph looks as follows:



You can also filter from this graph to the parallel chart. Under the settings, we have a filter slide, which tells how the threshold to draw lines between dots. Filtering via sliders updates the parallel chart. This is multiple filtering via multiple views

Together, these visualizations give you an idea of what movies in any genre share that binds the genre together, and how distinct genres are defined. When selecting a movie on Netflix, for example, this system will make it easier to pick which movie to watch based on previous films you had liked.