

P2 Write-Up  
apr74, drl222, jwn43  
4/24/2015

### **Data:**

Our data comes from two primary sources, IMDB (for title and genre), and <https://www.aggdata.com/awards/oscar> (for academy nomination and winner data). For the academy data, we began by cleaning out several awards that were not relevant to our inquiry. These included honorary awards for retired actors and awards that were not directed at a single movie in particular. We also determined that an award like “Best Actor” should be considered an award of the movie in which that actor played a part.

The IMDB data was an enormous list of movies, years, and genres. This list was truly enormous (about two million entries), so it would be impractical to join it with the academy awards data each time the page is loaded. Thus, we used a Python script to extract the data from the genres list (since it was not very conveniently formatted) and turn it into a CSV. Then, we used a second script to join the academy data to the genre data (due to the nature of the data, a sort-merge join was used, matching the title and year of one data set to the other). We ended up with a single CSV containing the Title, Genre, Year, and Award information, with a manageable size of just over twenty thousand entries. This is the file that is imported by the browser and used as data to plot.

### **Mapping:**

The basic idea was to have three sections of information: a timeline, a bar graph, and a list of movie titles. The timeline allows the user to filter data by year. We use the built in d3 brush to select any particular range of years the user chooses. We manually edit the extent after the user selection is made to ensure that we snap to whole years, with a minimum span of a single year. After the extent is decided we re-filter the data so that the bar graph only displays the desired range.

The bar graph is the main part of the data. Given the selected timeframe, it shows the distribution of the awards across genres, with each genre represented as a horizontal bar whose size matches the number of awards. Note that each award-movie-genre combination is counted separately, so that, if a film received multiple nominations and has multiple genres, it is counted several times. Also note that the x-axis changes scale when the timeframe is updated, so that the distribution can be clearly distinguished, whether there are a few entries or a few thousand entries.

The list of movie titles is an interactive feature that shows the names of the movies that were factored into each bar. When a bar in the bar graph is clicked, the list of movies (and the heading of the list) is updated to show the movies that factor into the relevant bar. Any given movie is only listed once. The current selection persists until another bar is clicked, even if the timescale is changed, but this is noted with the year range in the title.

Finally, there is a checkbox that allows the user to toggle between showing data for all nominations, or for only winners. This selection affects both the bar graph and the movie title list.

Our color choices were intended to draw the user towards the “click” functionality of the infographic. The red directions on page-load indicate the user to click one of the red bars for a list of all the movie titles in that genre.

### The Story:

We wanted to demonstrate the changing trends in public reception to various movie genres. Academy Awards aren't always a perfect representation of the public's attitude towards a film, but it allows users to spot trends that may underlay deeper societal beliefs. The most obvious note of interest is the fact that drama always overshadows the other categories. It may say something about America's unquenchable thirst for the dramatic flair that these movies are without fail nominated for Academy Awards. More so however, it probably reflects one of our design decisions: to increment the bar size of **each** genre to which a nominated movie was categorized. Since nearly all other genres could be considered drama, this category becomes a functional superset of many others. The fact that horror never really steals the show might also reveal our fragile stomachs, or that such films aren't artistically respected from members of the industry. Other more subtle observations include Film Noir disappearing after the 1950s, Adventure films taking home plenty of nominations but relatively little wins, and the rise of Romance films during times of war. Whatever you may discover, the central concern of this data is not the behavior of the Academy but rather the behavior and mindset of America throughout the century.