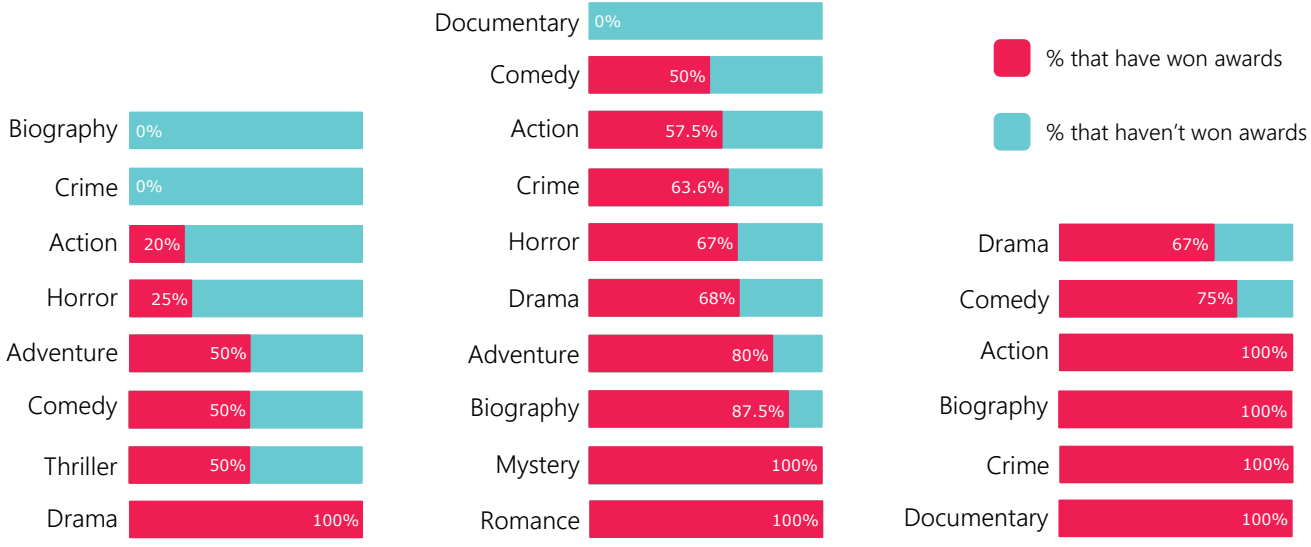
IMDb Ratings vs. Percentage of Movies That Have Won Awards by Genres



Firstly, rather than covering all observations in the dataset, I shrank the study range by filtering out some less intriguing observations. To keep the data's versatility and keep the topic fascinating, I decided to only focus on movies filmed after 2000. I wanted to explore whether the films shot in Toronto since 2000 received similar ratings by IMDb users and the agencies. Thus, I picked out the three most significant elements to depict this topic: the number of times the movie has won an award, IMDb rating, and genre. I came up with a graph that allows viewers to get a sense of what genres are preferred by the IMDb viewers but may not be appreciated by the experts, and vice versa. For example, one may infer that the regular audiences and the professional media have the same taste in action movies filmed in Toronto since 2000, as the percentage of movies that have won awards increases from the low IMDb rating category to the high IMDb rating category. Moreover, the graph also illustrates which genre has the most awards in different rating ranges. In other words, one may find that drama movies shot in Toronto since 2000 have won more awards than other film genres that are rated below five on IMDb, suggesting those users have higher standards for drama films. Last but not least, viewers can quickly identify the range of IMDb ratings of a specific genre. For instance, thrillers' ratings are generally less than 5, indicating that thrillers are usually not welcomed by the public.

Analyzing this dataset using genre is difficult since each movie can have multiple genres. By scanning over the genre column, I noticed that some genres are indeed a minor category. For example, “sport” is a genre given to only a few movies as a subcategory. These subcategories often do not represent the film accurately. Therefore, instead of including all genres and repeating movie counts in all the genres they belong to, I decided only to categorize the movies by their primary genre, being the first one that is listed. However, even with reduced genres, there is still too much information to display if I use colors to symbolize them. As a result, I replaced the colored genre legend with lists of names. Furthermore, I found it was easier to have the genre list on the left side for all three IMDb rating columns instead of having a single axis because I figured it was easier for the interpreters to identify their target. In addition, it does not affect them to compare the data that belongs to the same genre; people can quickly memorize three items due to humans' visual short-term memory capacity.

To help audiences further analyze the data, I added the percentage of movies that have received awards in the bar. Besides looking at the colored bars, these derived data allow viewers to make the comparison more efficiently. To illustrate the percentage bars, I utilized a warm tone color and a cool tone color to contrast each other. Choosing colors for my graph was a straightforward decision because I do not rely heavily on them to depict my thoughts. Furthermore, those bars are ordered attributes; hence, positioning them on a common scale and ordering them by length enhances communication through the magnitude channel.