

STA313 Assignment 1 Rationale

Do Movies with Longer Runtimes Have Higher IMDB Ratings?

Do longer movies tell better stories? Or are shorter movies enough to please an audience? For this visualization, I explored whether the length of a movie impacted its IMDB rating and whether highly rated movies tend to win more awards while communicating the decades in which they were released.

Design Decisions

I chose a scatter plot because comparing two pieces of quantitative data reveals potential correlations between them. In this case, my independent variable (x-axis) is runtime, and my dependent variable (y-axis) is IMDB rating.

As for the visual encodings, I used colour hue to represent each decade, area (2D size) to illustrate the number of awards won, and stars to represent each point. Since decades are ordinal data, I used the Viridis colour scale to represent the categorical differences between each decade and their ordered nature. The early decades are depicted using lighter colours (yellow and lime green) and the more recent decades are depicted using darker colours (dark blue and dark purple). Viridis is also colourblind-friendly, making my visualization more accessible. For the number of awards won, I chose to represent that using 2D area because it is an ordered attribute. Therefore, bigger stars indicate more awards won. As for the shape, I chose stars because a high number of stars is often used to signify success in the media. Therefore, a bigger star indicates more wins, showing higher success in a movie.

For the number of awards won, I combined the number of general wins and the number of Oscar wins. I did this to ensure a holistic measure of the success of each movie as movies may have received numerous awards from different organizations but no Oscars.

Key Insights

The scatter plot reveals a positive correlation between movie runtime and IMDB rating, suggesting longer movies tend to receive higher IMDB ratings. Also, movies with larger stars, indicating more wins, are more likely to be rated highly. Additionally, the visualization shows movies from more recent decades cluster towards the top, indicating higher ratings for more recent movies.

Limitations

A limitation I faced was overlapping points that obscure other data points, especially when larger stars cover the smaller ones. This makes it difficult to see outliers such as movies with higher ratings, but fewer awards won. Additionally, area (2D size) lacks precision when it comes to the positioning of each data point, making it hard to read the exact IMDB rating of each movie.

Alternatives

When testing different visual encodings, I first struggled with the encoding for decades. I noticed it was categorical and ordered so I did not know whether to use colour hue or luminance. By combining both using the Viridis scale, I was able to effectively communicate each decade. When choosing a visualization type, I also considered using a line graph but it would not be effective when showing the relationship between runtime, IMDB rating, decade, and awards won simultaneously.