

Analysis of Best Original Screenplay Nominees from 2024 and 2025

Anna Ryan, Elaine Zou, Kelsey Caldwell; John Rachlin

Background, Motivation, and Goals

Box office accomplishment does not always equate to award show success, and vice versa. In fact, award shows often seem to play by their own set of unwritten criteria. This project aims to explore what textually defines a critically-acclaimed screenplay. We were able to utilize natural language processing to investigate word frequency, overlap, and complexity of the ten most recently nominated films for Best Original Screenplay at the Academy Awards.

Process and Methods

Our process has an emphasis on functional programming and reusability of code. We aimed to craft a coding framework that allowed for the processing of large documents (e.g. screenplays) efficiently. We chose visualizations that we hoped would give us insight into both how screenwriters choose words and structure for their scripts and what choices are rewarded by the Academy.

Findings and Products

Fig.1: Sankey Diagram. The sankey shows us the most frequently used words in each screenplay and which words overlap between scripts. As to be expected, the top words were mostly character names, with some exceptions such as “day,” “like,” and “yeah.”

Fig.2: Screenplay Complexity Heatmap. The heatmap measures the complexity of the language used by each script through the normalized values of textual grade level, overall word count, and the number of words used with more than three syllables. Generally, we are able to see that a single script tends to score similarly across all three complexity measures.

Fig.3: Word Frequency Bar Graphs. The bar graph subplots show the top k-word frequencies per screenplay. Though not all ten are displayed, the subplots visualize differences in word usage for each screenplay, highlighting differences in language and emphasis across scripts.

Conclusion and Next Steps

Our research helps to identify why and how screenplays are being written, and what styles and techniques award shows tend to reward. To broaden our research and achieve more answers, we would like to incorporate multiple years of data and compare the differences between scripts that ultimately won the category versus scripts that were simply nominated. Expanding even further to comparing screenplays that were box office successes versus those that were critically acclaimed would also provide more context and findings into what textual choices lead to successful critical results.

Author Contributions | Anna: Debugging; sankey visualization; pdf parser updates | Elaine: Debugging; third visualization | Kelsey: code framework; pdf parser; second visualization

