Movie Facts Visualization

Team Members

- Zhiqing Li zl3064@columbia.edu
- Yunfan Ling yl4811@columbia.edu
- Jianing Wang jw3995@columbia.edu
- Stacey Zuo mz2875@columbia.edu

Abstract

We are interested in exploring cinema with Data Visualization. As a visual medium, film itself is already aesthetically rich in content, but it leaves too few clues to interpret itself, too little information for the audience to see the whole picture, to understand the stories or statistics behind and beyond the scenes. We believe data visualization techniques can be used to extract industry-wise meaningful patterns, and compare countless pictures in a new light. We aspire to transcend graphic visuals for cinematic insights with visualizations of movie data.

To begin with, we will explore some of the biggest data sources for movies, such as IMDb and Rotten Tomatoes (as they contain user ratings for millions of films and all the relevant detailed movie metadata) to understand the relationships among various descriptive characteristics, for instance, the correlation between audience rating and critical reception. We might focus on a particular aspect or topic of cinema to visualize as we dive deeper, such as the chromatic choices of films of different genres and death count in famous films.

Data Overview

We may take advantage of the following datasets:

1) IMDB Dataset

IMDb is considered the most comprehensive and authoritative source of information on movies. Subsets of IMDb data are available for access to customers for personal and non-commercial use.

2) On-Screen Movie Kill Counts of Films

This file contains the on-screen death counts for roughly 550 films and updated at the end of 2013. Other information like actors, movie genres, year of release, MPAA rating, director, movie length, and IMDb ratings. (The data in this file were scraped from www.MovieBodyCounts.com, so it will save us a lot of time in terms of conducting analysis)

3) TheMovieDB

This dataset is organized from TMDB API and mainly includes information about movie awards, casting crews, and 500 great directors.

4) Douban Movie Dataset

Douban Movie is a Chinese website that allows Internet users to share their comments and viewpoints about movies. Users are able to post short or long comments on movies and give them marks. This dataset contains more than 2 million short comments of 28 movies in Douban Movie website. Essential movie characters such as name and genres are also included.

5) Movie Dialog Corpus

This dataset is collected by Cornell University. It contains 220,579 conversational exchanges between 10,292 pairs of movie characters, involving 9,035 characters from 617 movies. Other essential features are also included.

6) Daily Earnings of Movies

This dataset contains daily data for 3400 movies, along with number of active theaters and daily ranings rank. It may be utilize to explore the relationship between marketing strategies or social media campaigns and their effectiveness on the box office.

7) TMDB

The Movie Database (TMDB) is a popular community-built movie and TV database. Its API is a resource for any developers to integrate movie, TV show, and cast data along with posters or movie fan art.

8) Rotten Tomatoes movies and critic reviews datasets

This dataset is scraped from the rotten tomatoes website as of 2020-10-31. It contains information such as movie title, description, genres, duration, director, actors, users' ratings, and critics' ratings; the critics dataset further includes critic name, review publication, date, score, and content.

Visualizations:

The visualizations included in this project will be mainly focusing on seven parts, each consisting of one specific sub-topic from the main topic. The project aims to study and visually represent the subtle details from the greatest movies in history by showing the audiences statistics of which they often easily ignore during their appreciation of the motion picture itself through the content and the plot.

This section provides a general introduction to the methods and the subtopics that will be examined in the project. A more detailed description and representation will be provided in the appendix.

1. **User Review vs. Critic Review (Scatter Plot and Bubble Plot)**

The first topic that we are interested in is the difference between the user reviews and the critic reviews on a movie. It is not rare that a critically acclaimed movie can receive very contrasting responses from the general public due to the differences in standards and academic backgrounds. We are interested in how the two groups differ in terms of their review on a basket of movies. The preferred method would be a scatter plot or a bubble plot.

2. **Death Count in famous movies (Barplot)**

Deaths of film characters are more and more common in contemporary movies, especially in action or assassin movies. However, in some cases, the audience gets so used to the deaths in movies that they often mistakenly underestimate the importance of a character's death in a movie. Therefore, by visualizing the death counts in famous movies, people can visually understand the seriousness of deaths. (In some cases, the death count is surprisingly large)

3. **Profanity in movies (Barplot or Other representations)**

Movies are often defined as a carrier for culture, history and literature. However, through the history of all kinds of artistic developments, profanity is inevitable as it is a part of people's daily lives. The Motion Picture Association film rating system plays an important role to prevent teenage audiences from getting access to too much profanity and nudity. We are interested in seeing how profanity progresses through time and how profanity becomes more and more acceptable to the general audience.

4. **Movie themes timeline (Timeline)**

The history of movies is also the history of the world. People tend to record the things that happened through artistic means such as literature, music, drawings, and also, movies. We think it would be interesting to visualize the progression of movie themes through a timeline with respect to the real-life events that happened during each time period in history.

5. **Filming Locations mapping (GGmap)**

Collaboration among movie creators in different countries all around the world has become a very common phenomenon in the contemporary era. American movies don't necessarily have to take place in America. The collaboration among countries can work as a catalyst for improving international relationships in some sense. It would be inspirational to visualize how filming locations change through time.

6. **Oscar Data Representation (Various)**

The Academy Awards is often regarded as the most prestigious award in the movie industry. Winning an Academy Awards is considered as the greatest honor for all movie creators and producers. Therefore, it would be insightful to examine the characteristics of Academy Award winners and see how the whole movie industry develops over time, and answer the ultimate question: How to produce a movie that can win an Oscar?

Appendix: Brainstorming and Ideas

"Movie Data Visualization."

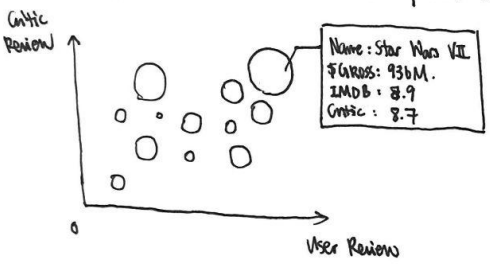
I. Introduction.

The visualization aims to provide the most interesting facts of motion pictures, one of the best mediums for carrying messages & aesthetic pleasure. While most of the movie watchers pay attention to the plots & cast, we pay more attention to the facts & details that most of the people ignore & might be stunned or dazzled to see some of the statistics we provide in this data visualization project.

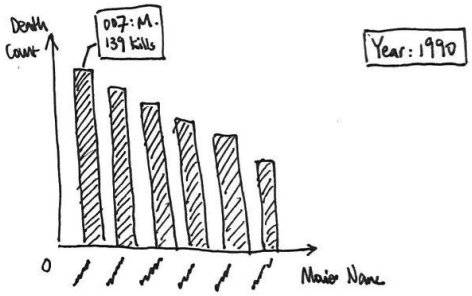
1 - User Review vs. Critic Review
2 - Death Count in famous movies.
3 - Profanity Count.
4 - Movie themes timeline visualization.
5 - Filming locations visualization
6 - Oscar movie data.

II. User Reviews vs. critic Review.

(Bubble plot) * The bubble size can represent the gross profit ($)

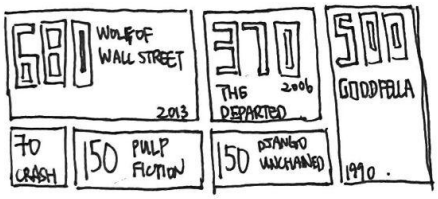


III. Death Count.



IV. Profanity Count.

~~Option 1~~:

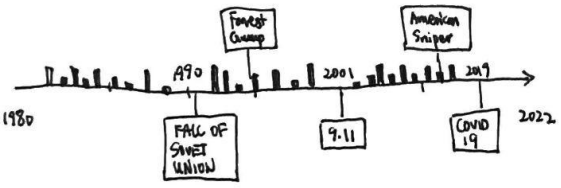


...

V. Filming locations

* Heatmap

VI. Movie Theme Timeline (Work in progress)



VII. Oscar Data Visualization.

- Movies with most Academy Award nominations & wins
- Movies that are popular but no oscar nomination.
- Actors & Actresses with most wins & nominations
- Themes of Academy Award winners.