COMPSCI 571

Process Book @ Milestone Phase

Soniya Gaikwad
32927398
sgaikwad@umass.edu

Manan Parikh
34968148
mananrajeshb@umass.edu

Shiva Sravani Mudiyanur
34965596
shivasravani@umass.edu

Project Title: Montage - Exploring Movie Trends Over Time

GitHub Repository: https://github.com/soniyagaikwad/movie-trends-data-vis

1   Overview & Motivation

Our team decided to pursue "Montage: Exploring Movie Trends Over Time" because of our love and interest in entertainment, especially movies. Also, some of us have projects revolving around media, so we wanted to continue looking into this industry and its crossover with data visualization.

2   Related Work

Some related works that inspired us to work on "Montage: Exploring Movie Trends Over Time" include the visualization from Rotten Tomatoes from Homework 1 and the PokeData Final Project from the University of Utah shown in class. We wanted to combine aspects we liked from both visualizations to create a data visualization website for movie-related data.

3   Project Objectives and Questions

The primary questions we are aiming to answer with our visualization include:

- While looking at a specific (movie/genre/MPAA rating/budget), how much revenue was generated?
- What are the relationships between the MPAA ratings, budget, gross revenue, release dates, genres, runtimes, and ratings the movies have received?
- What are the highest and lowest revenue-generating movies?
- What are the highest and lowest-budget movies?
- What are the ratings for the movies depending on their genres?
- Which genre has the most interest over the years?

We would like to learn how these pieces of data related to movies correlate with one another and theorize how they potentially play out in the entertainment industry. Some benefits of learning and accomplishing these aspects include helping us and others understand what factors could be critical to the movie industry for potential box office success and how the public interacts with these movies based on these factors.

4   Data

We are using a dataset from Kaggle by Yashwanth Sharaff called "Movies Performance and Feature Statistics: Analyzing Box Office Performance, Rating and Audience Reactions" [1] to visualize our project.

https://www.kaggle.com/datasets/thedevastator/movies-performance-and-feature-statistics

5   Data Processing

In the dataset "Movies Performance and Feature Statistics: Analyzing Box Office Performance, Rating and Audience Reactions," we gain access to a large amount of data related to movies, such as their titles, MPAA ratings, budget, gross revenue, release dates, genres, runtimes, ratings they have received, actor id, and many more. Based on our initial designs, we focus on each movie's title, MPAA ratings, budget, gross revenue, genres, and summaries.

In terms of data processing and cleaning, we process it using Javascript. We start importing our csv file, which is kept inside the data folder. We use d3's csv function to read the csv file, and only consider the first 510 rows as only they are required, then we consider these columns only to proceed further: title, MPAA rating, budget, gross revenue, genre, runtime, rating, and release date.

Moreover, there is a column named 'release date' in our dataset, which we don't intend to use in its entirety. Based on our brainstorming, we will only extract the release year from it. In addition to the release year as a derived attribute, another derived attribute would be the profit/loss, in which we take the revenue - budget.

To support our visualizations, we created specific utility functions in JavaScript. One such function, groupGrossByRating, groups movies based on their MPAA rating and collects their gross revenues into sorted sets. This helps in preparing data for boxplots, where revenue distribution by rating can be visualized more effectively.

Another function, getMoviesByRatingAndYear, filters the dataset based on a given MPAA rating and release year. It returns a simplified list of objects containing only the movie title and its gross revenue, making it useful for focused comparisons or detailed listings in visual outputs.

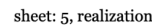## 6   Exploratory Data Analysis

To begin exploring our data and planning our visualization tool, we created a basic table to display a list of movies along with their key details like title, MPAA rating, genre, release year, budget, revenue, profit, and rating. With these details, we decided to use a combination of simple bar charts, pie charts, and scatter plots.

For example, to visualize the budget and revenue of each film, we decided to use bar charts and pie charts. To observe the different movie details and how each aspect potentially relates to another aspect, we decided to work with different graphs that fit the attributes appropriately. For example, while digging deeper into the data, we noticed there could be a potential pattern between MPAA ratings and gross revenue, which we haven't explored yet in other data visualization tools. So, to explore and communicate this relationship clearly, we decided to use a box plot, which effectively shows how revenue varies across different rating categories.

However, we noticed that if we placed all these charts on a single webpage, it would look cluttered and overwhelming. To solve this, we decided to create an interactive interface where users can

select their own X and Y axes, allowing them to generate the specific visualizations they're interested in. This makes the experience more user-friendly and customizable.

These insights helped shape our final design.

7 Design Evolution

At our Project Proposal phase, we considered the following design that integrated all the aspects we wanted to use from the alternate designs each member created.

Figure 1: Sheet 5, Realization, of the 5 Design Sheet Methodology (/project-proposal)

The design decisions we made using the perceptual and design principles we've learned in class include the following important features:

- User-Friendly Interface
  - A clean, intuitive design with easy navigation.
  - As described in Figure 1 and seen in Figure 2, we are using an orange and blue palette for the UI to ensure the visualization tool is accessible to all users.
- Movie Search Functionality
  - Users can search for movies by name to retrieve relevant details.
- Movie Details Display
  - Essential information such as title, MPAA rating, budget, revenue, release date, genre, runtime, number of ratings, and summary should be presented.
- Sorting & Filtering Options
  - Allow users to sort movies by popularity, revenue, release date, or rating.
- Clickable Links
  - Users can click on a movie to get the information summary accordingly.
- Trends over Time
  - Graphs or visualizations showing revenue trends over time.

In addition to these important features, we had some other ideas we thought would be nice to have, but were not critical to our project such as:

- User Ratings & Reviews
  - Users can rate and review movies.
- Social Sharing
  - Allow users to share movie details on social media.
- Streaming Availability
  - Indicate where the movie is available for streaming (Netflix, Disney+, etc.).
- Movie Recommendations
  - Suggest similar movies based on selected titles or genres.

However, we may have to deviate from these features due to the lack of quick and accessible data. Another aspect we may have to consider deviating from is the bar graph that comes from the box plot as seen in Figure 1. While we still aim to work towards this feature, it may be difficult to fit meaningfully within the UI.

For the majority of our visualization tool design, we continued to stick to the design decisions we originally made; however, to improve the design, we decided to enlarge the "Compare All Movies" aspect, so that the data plots are easier to interact with and analyze, as we continue to allow the user to play around with the x-axis and y-axis with different attributes and plots that fit these attributes, as seen in Figure 2.

8   Implementation

In Figure 2, we have our key design of our website, "Montage: Exploring Movie Trends Over Time."

On the left side of the website in Figure 2, we have a table that will list all the movies in the dataset with their appropriate details, such as their title, MPAA rating, genre, release year, budget, gross revenue, profit, and rating. In this table, the user will have the ability to search for a specific movie or filter the movie based on their attributes.

If the user selects a specific movie in the table, the selected movie's data summary will show on the window to the right side of the table on the website. In addition to the movie's summary, there will be two plots shown; one of which provides the rating information using a circle percentage graph and another providing information about the budget and revenue using a bar chart.

Below the table and movie summary section, the user has the opportunity to compare all the movies in the dataset. With this graph, the user can interact with it by changing the x-axis and y-axis labels by using the data attributes from the table, such as their MPAA ratings, genres,

release years, budgets, gross revenues, profits, and ratings. This part of the graph works dynamically and changes according to the different types of data plots that make sense with the provided axes.
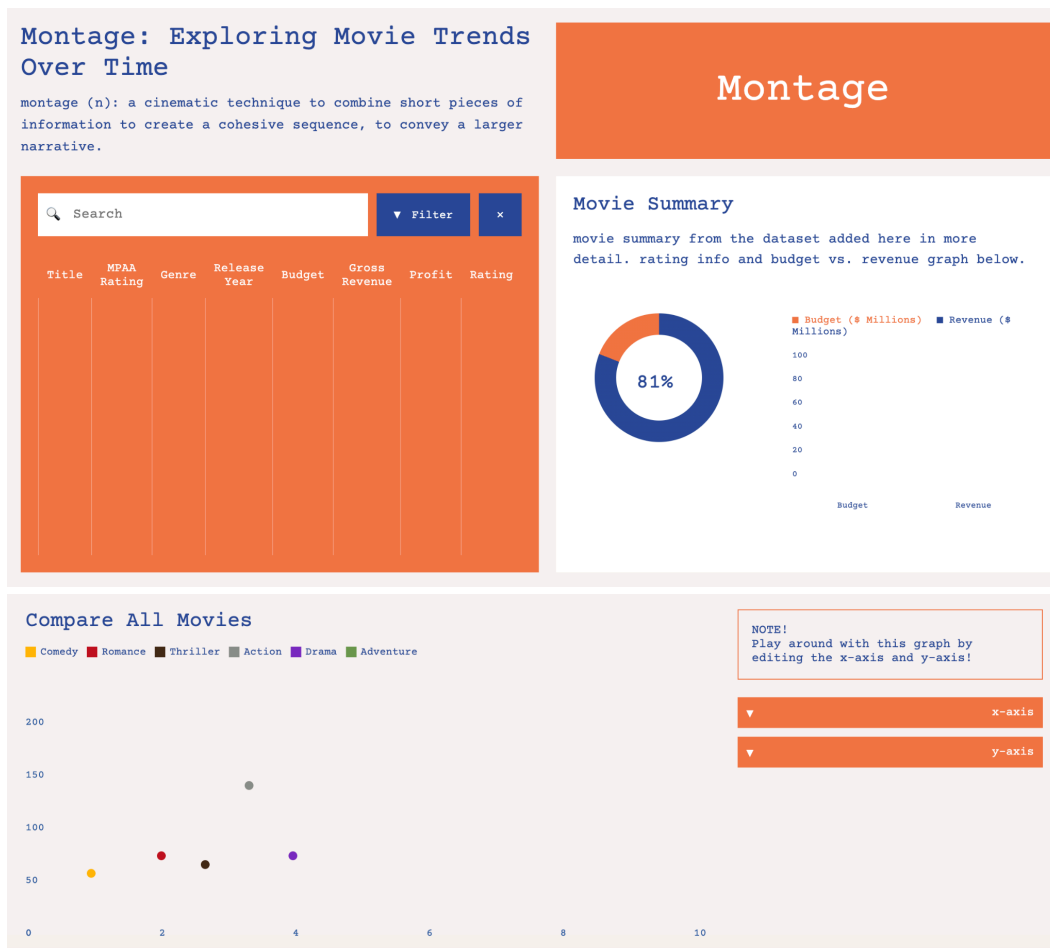

Figure 2: UI Implementation of the Visualization (montage.html and montage-design.css)

9   Evaluation and Progress

As we're making progress on implementing our web-based interactive visualization tool, we are continuing to answer our research questions and how we can improve the ways we represent the data plots appropriately for the different variables in the dataset.

We continue to work together and delegate responsibilities based on our interests, experiences, and strengths each step of the way.

Tasks	Expected Deadline	Progress Phase
Data Processing + *Project Review	*3/7	Completed
Create the UI + *Project Review	*3/14	Completed

Work on the Backend Functionality	3/28	Completed
Visualization Prototype for 2 Datasets + *Project Milestone	*4/11	Completed
*Peer Feedback	*4/16 + 4/18	Pending
Fully Interactive Website w/ Few Bugs	4/23	In Progress
Final Touches	4/25	In Progress
*Project Screencast	*5/2	Pending
*Final Project Submission + Group Feedback	5/9 + *5/12	Pending

* Course Deadlines, Subject to Change

References

[1] TheDevastator. (2023). Movies performance and feature statistics: Analyzing Box Office Performance, Rating and Audience Reactions. Kaggle. https://www.kaggle.com/datasets/thedevastator/movies-performance-and-feature-statistics