Laura Fonseca Pauthz, Josette Vignos, Peter Huang, Sung Woo Min
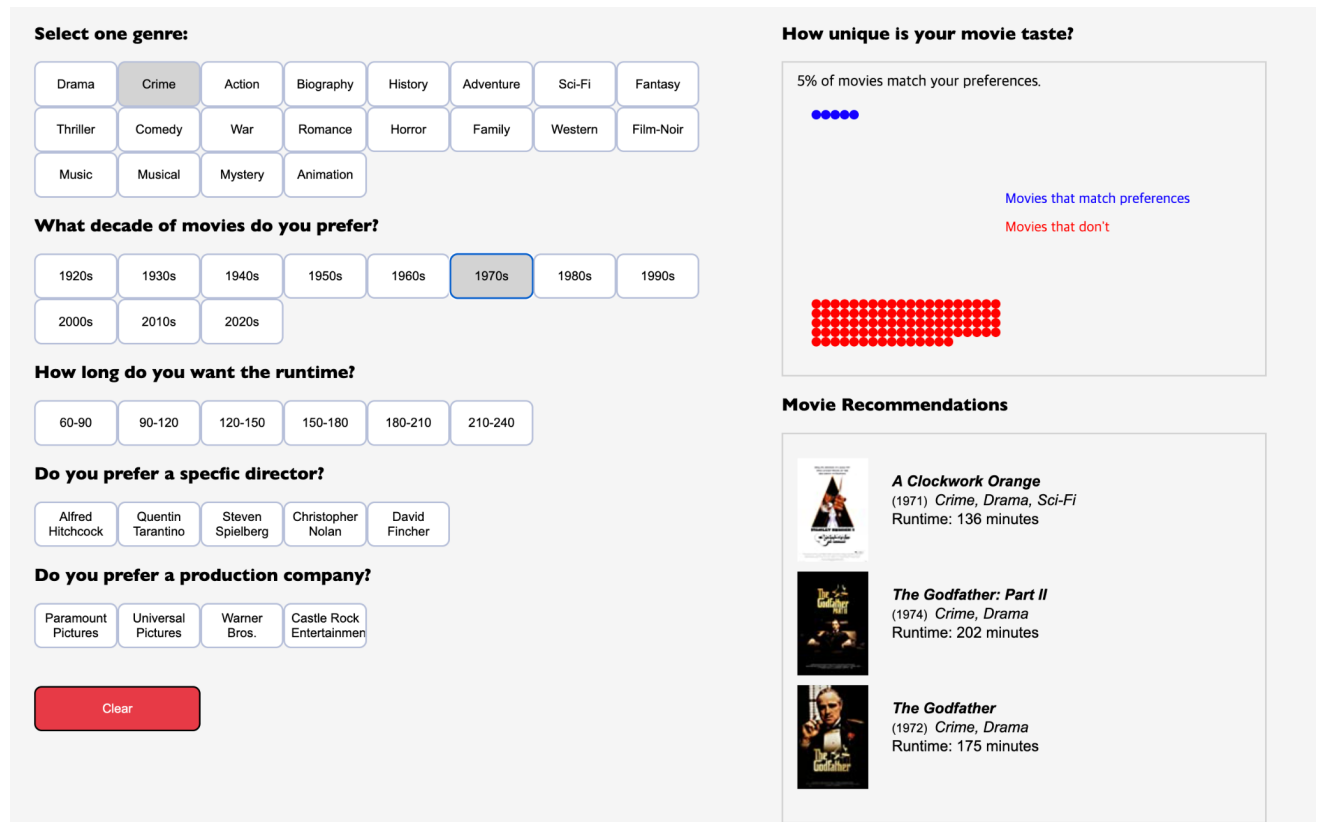
INFO 3300 Project 2

15 November 2022

Data Description

      The data gathered to work on this project were sourced from Kaggle and GitHub. These comprehensive datasets had information on the highest rated movies on IMDb and Rotten Tomatoes. The variables we used to build our visualization were genres, eras, runtimes, directors, and production companies. These were important for our filtering process so that users can easily find and choose the specific features they preferred, and a way for us to calculate and find the percent pool of movies that match a user's taste. We integrated the different datasets by using the movie's information and ratings from the smaller IMDb data, and the poster links from the larger IMDb data. We filter down possible movies by using the variable `filterdata`, which tracks a user's preferences and choices, and returns a list of movies that fit that criteria by relying on the .filter() function in JavaScript. This key variable helps the visualization by showing what

percentage of movies match a user's taste, and also helps in presenting what movies would best suit the user based on their choices.

Design Rationale

One of the main inspirations for our project was the McDonald's and police violence visualization projects presented in the Project 2 handout. In order to follow conventional design principles, we kept the filtering options to the right, and the visualization to the left. This allows for users to read the questions, interact with the project, and see the visualization in action seamlessly. One tradeoff about this way of presenting movies is that users lose a lot of information about the movies that they don't choose, and only get recommendations for the top 3 choices that match their answers. The marks in our visualization are the dots in our chart, each representing a movie. The channels for the visualization are horizontal positioning and color, with horizontal movement to the left indicating that it's a movie that matches, and color indicating whether a movie fits a filter. Blue indicates that it fits the filtering criteria, and red means that it does not. This helps users visualize how their choices match up with well known movies, and how unique their movie tastes are. The movie recommendations section includes the title, genre, year released, runtime, and the poster. Presenting these pieces of information lets users see potential movies they would be interested in, and gives the users basic information about the movie. These movies change depending on the user's choices.

As for our interactive elements, we decided on using labeled buttons as our primary form of interaction. Since this is an interactive way to filter down movies and connect a user's tastes to the movies in our dataset, it's important that users see all the available options given to them. It also encourages users to experiment with many different possible options, whether with genre, runtimes, and even directors. We decided to format all the questions the same, with all options being presented as buttons. Considering how there is a wide array of movies, all with different combinations of options available, we thought it would be more user-friendly and similar to other online quizzes like Buzzfeed compared to other options like dropdowns, radio buttons, and checkboxes. The buttons also darken when an option is chosen, letting a user see the options that they chose. This way, they can also see how their different options yield different results.

The Story

This visualization has a variety of different stories for different people. Perhaps some people's tastes aren't as popular as others. Maybe some users find that their tastes and preferences are very common, and they can find movies to watch that they're interested in. In terms of the suggestions given and the percentages shown, it will always differ among different people. However, that's the goal and insight we wanted to provide - that people's tastes are different and unique. What's interesting is the distribution of genres within the top 10, and how tastes may not be as popular as one would think. For example, drama is the most common genre, with drama movies 71%, and musicals being the least common genre, with 1% of movies being musicals. It's interesting that musicals is even a genre considered for the top rated movies in IMDb.

Team Contributions

As a group, we planned the project together, and tried to find ways to equally split the work. In our first meeting, we decided that we all pitch 1-2 ideas each, and tasked ourselves to look for 2-3 datasets relating to our suggestions. In our second meeting, we each presented full-fledged ideas using the datasets we found, and the ideas suggested in the last meeting. At the end of that meeting, we all voted on which we liked best, and planned what our visualizations and interactions would be like, and what takeaways we want our users to have. We split the work in such a way that we could fit our parts together for the final part, but continually adjusted as we worked together. Peter helped initialize the filtering, Josette worked on filtering functionality and data visualization, Laura worked on the report and movie recommendations implementation, and Sung Woo created the GitHub repository, implemented helper functions for filtering and movie recommendations, and styled the project.

Contributions Time Breakdown:

Josette: hard-coded visualization and layout of page for in-class demo [2 hours], implemented filtering for all quiz questions [5 hours], implemented dynamic data visualization [4 hours]

Laura: Movie recommendations visualization [2 hours], most of the report [2 hours].

Peter: initialized the filtering functions [2 hours], styling, report, and miscellaneous [1 hour]

Sung Woo: implemented helper functions and features for filtering and movie recommendations [6 hours], majority of styling [1 hour]