A rationale for your design decisions. How did you choose your particular visual encodings and interaction techniques? What alternatives did you consider and how did you arrive at your ultimate choices?

Our visualization takes an IMDb movie dataset to answer the following question: Does a movie's certificate affect its performance? A movie's certificate is a rating that reflects its suitability for certain audiences, for example R for Restricted. Since a movie's performance can be evaluated in a variety of ways, we decided to focus on financial success (i.e. a movie's gross) and public success (i.e. a movie's rating). Therefore, we added the option to toggle between graphs of two performance metrics: one plot of the distribution of movie ratings and one plot of the distribution of movie grosses.

In terms of visualization type, we were inspired by Matt Daniel's project, The Largest Vocabulary in Hip Hop, a beeswarm plot that sorted rappers based on the number of unique words used in their lyrics. We chose a beeswarm plot over other visualizations that also display distribution because we wanted to focus audience attention on exploring individual movies rather than looking at overall trends or averages. Here are some alternatives we considered and why we decided against them:

- Dot plot - While similar to a beeswarm plot in that it visualizes each individual movie, our rating data was not suitable for this type of plot due to the large number of movies with the same rating. Choosing a dot plot would have meant having a very vertical graph, making it inaccessible to explore and hard to see the overall picture. On the other hand, a beeswarm plot visualizes the data in a way that the data points don't overlap, yet "swarm" together based on their x-value, making it a better choice for visualizing our distribution.
- Histogram - A histogram could display the distribution of movie ratings and gross very effectively, but the type of plot would only count the number of movies that fall under each bin, without visualizing the movies themselves. This type of plot would not be effective for identifying specific movies' ratings and would also not allow us to see what type of certification each movie has.
- Box plot - Using a box plot would have led to similar problems as described with a histogram: more focus on the overall descriptive statistics instead of each individual movie. A box plot would be a better choice than a histogram due to plotting multiple plots of the distribution of ratings/gross for each certification category on the same set of axes being more legible and space-saving compared to plotting multiple histograms, but separating the data by certification category in this manner makes it more difficult to compare individual movies across categories.
- Bar chart - Lastly, we considered using a bar chart because it would allow us to plot a categorical variable (certification) and a continuous category (movie rating/gross). However, we cared less about the number of our average rating/gross in each category and more about the distribution.

Another alternative we considered was examining the relationship between a movie's genre and its performance. We decided against focusing on a movie's genre because we felt like it has been studied many times before. It was also much more interesting to learn about certificates, how many there are, and what it takes to assign certain certificates to movies. Lastly, we decided to visualize the top 200 movies instead of the 1,000 in the entire dataset due to the limitation of space. Our beeswarm plot looked congested when we had all 1,000 data points on the visualization, so the best decision was to minimize our dataset to only 200 movies. We chose quality and effective visualization over quantity.

Our final solution? A beeswarm plot where each individual point is a movie, visualized by its movie poster, and its location on the x-axis corresponding with its rating or gross, depending on the graph. The movie posters as data points add a layer of clarity to the graph compared to using solid-colored circles because it allows for user recognition of some movie immediately. On hover, a tooltip appears with the movie's name and its rating or gross, further clarifying the data. The inclusion of the movie rating or gross in the tooltip is essential because a disadvantage of the beeswarm plot is that it sacrifices a level of accuracy (placing points *exactly* where they should be on a graph) to receive a higher level of clarity and legibility. The labeling on the tooltip makes up for the data accuracy that the beeswarm plot lacks. Lastly, in order to give users the tools to explore our visualization to answer our opening question, we added a dropdown menu that allows users to highlight movies with a specific certification, while other movies are faded out, to minimize the information overload of viewing 200 movies at the same time. Dropdown menus also save time for users by allowing them to quickly find all the data points they are looking for by simply clicking on the certificate they want to examine. For example, if a user hypothesizes that movies with U (Unrestricted) certification have a high grosses due to it being more accessible to audiences of varying ages as opposed to A (Adults only) certification, they can use our menu dropdown feature to explore those two specific groups of movies and their gross distributions.

An overview of your development process. Describe how the work was split among the team members. Include a commentary on the development process, including answers to the following questions: Roughly how much time did you spend developing your application (in people-hours)? What aspects took the most time?

Our team practiced pair programming throughout the entire duration of this project. This allowed us to collaborate on every aspect of the visualization, from deciding on and cleaning the dataset, to choosing specific visualization encodings and interaction techniques, to coding and debugging our website and graph. We spent roughly 15 hours working on this application together from start to finish, with the most time being spent on the beeswarm plot. In the beginning we weren't even sure how to describe the visualization we wanted to create, calling it a scatterplot with no y-axis, a dotted histogram, and a dot plot before learning that it was called a beeswarm plot thanks to Sam. From there, looking for resources and help online became easier with the right keywords, but we still found many aspects of D3 and Svelte to be difficult and time-consuming, such as generating the graph with proper point placement, adding images to points, and incorporating interactivity.