

Project 1 Writeup

The data used for this project's visualization concerned the Bechdel Test, a test that checks for an instance of two women talking about something other than men during the course of a movie. The data set itself was found on the FiveThirtyEight Github Repository, a repository that contains various datasets on a wide range of topics used on the FiveThirtyEight website. Through our project, we looked to break down the data and represent it in ways that highlight how gender inequality in America has been reflected in American media.

The data itself was a CSV file containing a set of 1,795 movies, each containing values for various measures, such as the movie's budget, domestic and international revenue, and how the movie fared under the Bechdel Test. The latter points indicated whether the movie passed or failed. In the case where it failed, the data set grouped the failure into one of many categories, such as having no women in the movie, having women present but do not talk, and having women that talk but only about men. We also gathered more information to look at the data at other angles, including looking at individual IMDB ratings. To gather the ratings for each of the various movies, the IMDB API was used along with the provided movie IDs from the data set in order to add the values to the CSV files.

In designing the layout and the graphs, all decisions were made to make the data visualization as clear as possible, and allow viewers to quickly identify trends and understand the overall claim of the project. In the first graph, each of the various categories of failures was given a distinct color in order to allow for easy differentiation of each, along with a subsequent more translucent version of the same color to indicate where experts disagreed about if said movie fell under such a failure. This allows the viewer to quickly distinguish the various groupings, as well as easily understand which translucent colors go with which solid color. The categories were then made into bars and stacked on top of each other for each grouping on the x-axis, with each bar indicating the percent of movies that fell into that individual category, and the entire bar indicating how many movies failed the Bechdel Test in that group. The stacking of

bars that represent each of the various categories also serve to make the downward trend of failures through the years more clear. For the second graph, lines were used to represent the evolution of the average IMDB scores of movies that failed and passed from year to year, with red representing those who failed and blue representing those who passed. One of the main design decisions was to shrink the y-axis to only include values between 5 and 8, instead of having the full 0 to 10 scale. Because the average IMDB ratings fell between this range, switching the y-axis allowed smaller trends within the data to be more emphasized and easily distinguished. Finally, for the last graph, each of the movies was made into a data point and graphed in terms of their year and the budget they received. Circles were colored red if they failed the test and blue if they passed, mirroring the color decisions of the second graph.

Each of the three graphs in our data visualization aims to make obvious a trend that we observed in the data. The first graph aims to highlight the downward trend of movies failing the test, an indication of American media being more willing to have females take a larger role in movies involving plot outside of romance for a male central character. The second graph looks to reflect how the average IMDB ratings of movies that passed the test were substantially lower than those who failed the test for much of the 1980s and 1990s, before beginning to converge during the 2000s. However, the average IMDB ratings for movies that pass the Bechdel Test still stay reliably under the average of those who do not, possibly indicating some sort of bias towards male-dominated movies that have continued to carry over from past decades. Finally, the last graph helps illuminate the fact that movies that fail the Bechdel Test tend to have higher budgets than those who pass, also indicating some bias towards male-dominated media. Overall, the trends in the graph serve to convey the idea that during the 1980s and 1990s, women saw extreme biases within the film industry, in the way of diminished roles, smaller budgets, and worse critical response. The visualizations also reflect the idea that while these trends have begun to correct themselves, these inherent biases in favor of males still exist in our society today.

Project Work BreakDown:

Zach Brody (ztb5)

- Worked on the Stacked Bar Graph (First Graph)
- Worked on Formatting Data/Fetching Data from API
- Worked on Styling
- Worked on Finding Data Sets

Eldor Bekpulatov (eb654)

- Worked on the Scatter Plot (Third Graph)
- Worked on Styling
- Worked on Finding Data Sets
- Brainstormed ideas on how to visualize the data in an effective way

Nathan Bala (nb378)

- Worked on the Line Graph (Second Graph)
- Worked on Description labels for the graphs and overall project
- Worked on Styling
- Worked on Finding Data Sets

Finding the data set took up a large portion of our time, due to the fact that our initial idea ended up being not too feasible. Once the data set was found, each individual member spent the time formatting and parsing the data in their separate javascript files and creating their graph for the data visualization. This aspect of the project took the longest for each individual. Finally, the graphs were put together and descriptions and styling updates were made to create the final project.