Kimberly Brown, Hannah Ajmani, Kevin Ma

Kcb65, hna8, kym5

CS 3300- Project 2

Gender’s Role in Film

Description of the Data

1. Overview

For our project, we focused on movies in in the last 15 years and how they split the dialogue between genders. We compared the percentage of male dialogue to percentage of female dialogue and used information such as the gender of producer, director, and producer and the year of the film as filters on the data. We also needed the title of every film and the gross revenue for our visual.

1. Description of Database
2. Data Selection
3. Multiple Data Sets

Description of Mapping to Visual Elements

1. Plotting the Data

We have three separate plots in our visual: a Pi chart, histogram, and scatterplot. The first one we created was the Pi chart. The idea behind showing our data this way first was to give the viewer a clear visual of the split between male dominated dialogue in movies and woman dominated dialogue. In order to create this graph, we wrote a function that plotted each data point as a circle in the pi chart, first for all the woman-dominated dialogue films in one area and then for all the man-dominated dialogue films in the other area. There was an issue with the sizing of the pi chart circle, which was correlated with the number of data points being graphed. We did our best to minimize the size of this circle while fitting all the data points, but the results weren’t exact. The color of each data point was designated as either red or blue depending on which gender dominated the dialogue in the film. This way both color and placement in the Pi chart differentiated male dominated films from female dominated ones.

The histogram was plotted in order to show a further breakup in the percentage of dialogue. We kept the color of each data point the same as in the Pi chart to maintain consistency across our visual. The x-axis is the percentage of male dialogue in the film and was scaled linearly from 0 to 105 percent while the y-axis is the number of films and was scaled linearly from zero to the number of films in the largest bin plus five.

The scatterplot was our last visual and was created in order to show why most dialogue in films appears to be dominated by men. Once again we kept the color of each data point the same. The x-axis is the percentage of male dialogue in the film and was scaled linearly from 0 to 100 percent, while the y-axis was scaled linearly from 0 to the highest revenue a film made in the dataset. The size of the circles is another representation of the gross revenue of each film in comparison to the other data points. We calculated the radius as a multiple of the square root of the film’s revenue divided by the highest revenue in the dataset. By changing the size of the circles, we give another visualization of how much revenue a film made aside from the location along the y-axis.

1. Filtering Data

In order to give the viewer more detail about the dataset, we allow the data to be filtered by the gender of the producer, director, and writer as well as the year the film was released. These filters can be applied across all three visuals and give a clearer view of the data especially in the last graph. All four filters can be applied together or separately. Everytime When applied in the Pi chart, the size of the graph is reduced and a grey circle is created to show to size of the graph when no filters are applied. In the histogram and scatterplot, the filters can affect the scale of the y-axis depending on which data points are removed.

1. Adding the Popups

The Story

Our visualization shows how male dominated the film industry is in Hollywood. Not only are the majority of speaking roles in film dominated by men, but also so are the creative roles behind the camera. The absence of women behind the camera can be seen in the lack of data points when male directors, producers, or writers are filtered out. However, the visualization shows that when women are a part of the creative process behind the camera, there is either a more even distribution of speaking roles or women dominate the dialogue more often. In this way the lack of women behind the camera explains the lack of dialogue spoken by women on camera.

Our third graph shows the correlation between revenue and the extent that male dialogue dominates a film.