

# Data Visualization Final Delivery

Contributor:  
Jenna Adams

Date:  
12/15/2022

Group Number: 23

[Github Repository](#)

[Data Visualization Page](#)

[Video Presentation](#)

## Overview and Motivation

Movies give viewers more than just entertainment, they call for the viewer to look from a different perspective and sometimes explore worlds unknown to them. Movies have the ability to reflect culture, influence culture, and promote economic progress makes them crucial to society.

Even though movies can have such a large impact on society, this doesn't mean that every movie that is released is a good one. So who decides when a movie is good? Who's to say that they don't have bias? Rotten Tomatoes is an American review-aggregation website for film and television. They have made a name for themselves and are notorious for their movie reviews leaving a movie with a final score and a verdict of being Fresh or Rotten.

We chose to investigate this because we are curious to see if this system of critics has any trends. Each person in this group grew up watching movies of all sorts and found different genres to be our favorite. We want to understand what genres of movies score better than others or if movies with certain actors typically received higher ratings. We grew up with this entertainment and are curious to find out how movies and ratings may have changed over time.

## Project Objectives

One of our first questions we wanted to answer is "Does a certain genre of movies tend to score better than other genres and how has the quality of those genres changed over time?" We wanted to understand what genres tend to have better ratings and which genres tend to have lower ratings. The benefits to understanding this is understanding what to expect from certain genres, if a new movie comes out, depending on its genre we have a general idea of where on the scale it may fall in the ratings. This would also help us determine if we want to see a movie at all. And out of curiosity, we also wanted to see how the quality of genres have changed over time. Interpreting this would help us understand what movie genres are in their prime now and what genres have aged with the times. This information would help us make better decisions when it comes to a movie night.

Our second question asks to find how an audience's review differs from a critics, "Is there a correlation between the audience score and the critics score of a movie?" Our data set provides both review ratings, and obviously every movie created isn't a cinematic masterpiece, but that movie may still be worth watching. This question aims to answer what differences there are between these two ratings and will help us understand if the critic reviews are as accurate when it comes to all movies.

Our final question is "Does content rating have an effect on the critic's score?" Content ratings limit audience viewing in most theaters and households, but would a movie rated M still have the same quality as one that is rated G? We hope to find that answer and learn if we should be taking our dates to see mature movies or if PG-13 movies are a better option

## Data

We found this [Rotten Tomatoes Dataset](#) on Kaggle. Thankfully this data is complete and did not need to be collected firsthand. It contains over seventeen thousand movies and their related critic reviews that were scraped from Rotten Tomatoes on 2020-10-31.

## Data Processing

The data that we originally planned to use was the genre/s of the movie, the release date, the critic score and the audience score. We also needed to filter the data so that any fields we wanted to use did not contain null values. This removed about four thousand movies from the final delivery.

The data we needed to derive was the average audience score per genre per year, and the average critic score per genre per year. We also needed to get the individual movie critic and audience ratings. We originally thought we would need other calculations to be done, and we were correct. We were able to make all of it work in D3.

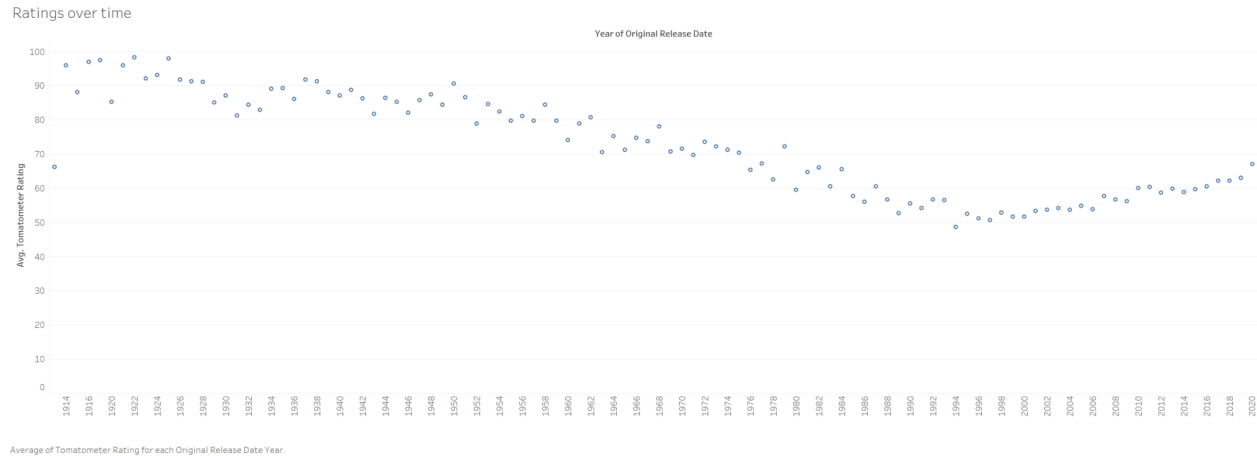
## Exploratory Data Analysis

We initially looked at our data within Kaggle and Excel to get a better understanding of the different columns and the size of the dataset. We then took to tableau to find what visualizations may make sense for this type of data set. I started looking at bubble charts, as those seem to be a personal favorite. Although I found them aesthetically pleasing they made it easy to misconstrue the data. One bubble may be larger than the other at a glance, when in fact it's not, it varies by the hundreds with a dataset of this size. So that idea was thrown to the wolves.

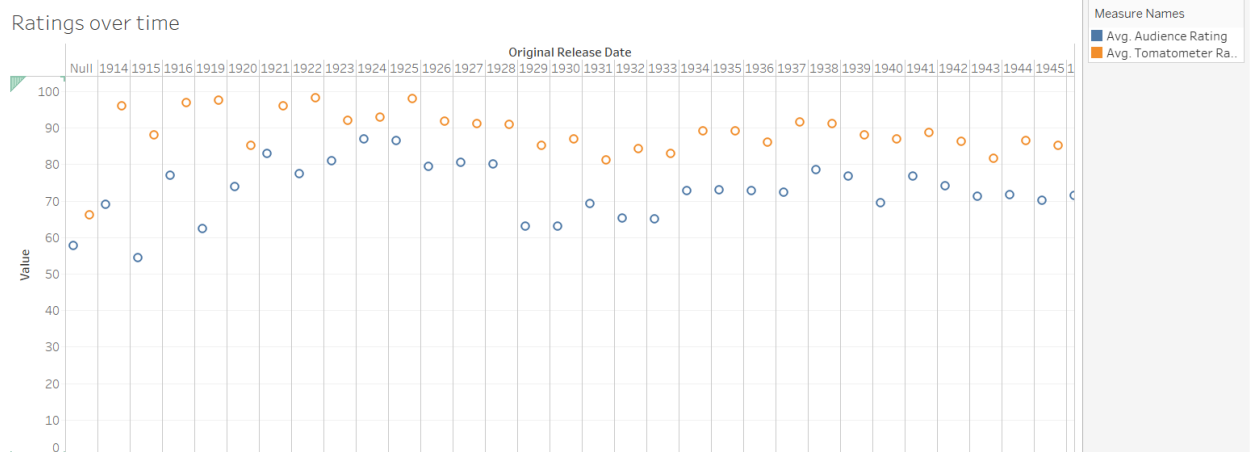
Exploratory Data Analysis did help with understanding how many genres a movie could have attached to it, and in turn how many combinations that would make for a bar chart, heatmap, etc. This also brought to light how each genre's total movies varied greatly in this dataset. The genre "Cult Movies" had a total of 7 movies in it, while "Drama" had thousands. This also helped me realize that doing anything with the individual movies would have to be done on a larger scale, as there were over 17000 movies in this dataset.

## Visualization Design

### Alternative Prototype designs:

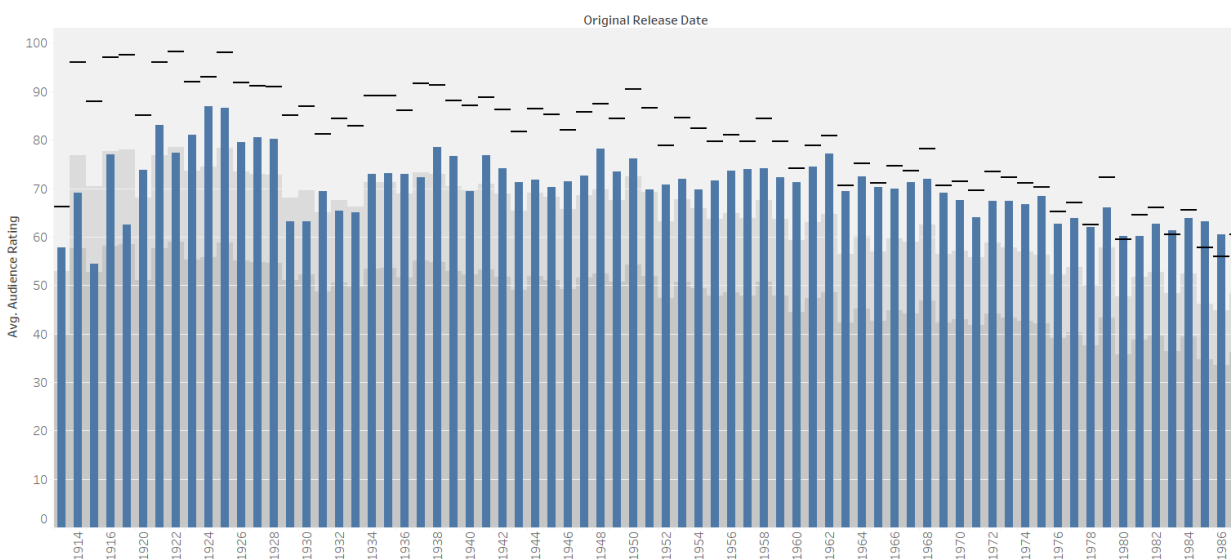


This design specifically highlights Average Critic Ratings of movies based on their release date. Something like this would be great to show two attributes at a time. This would also be a great way to show data with a shorter window of release date ranges. This graph is easy to read and shows an overall trend, but we want our viewers to be able to see more than a two dimensional graph.

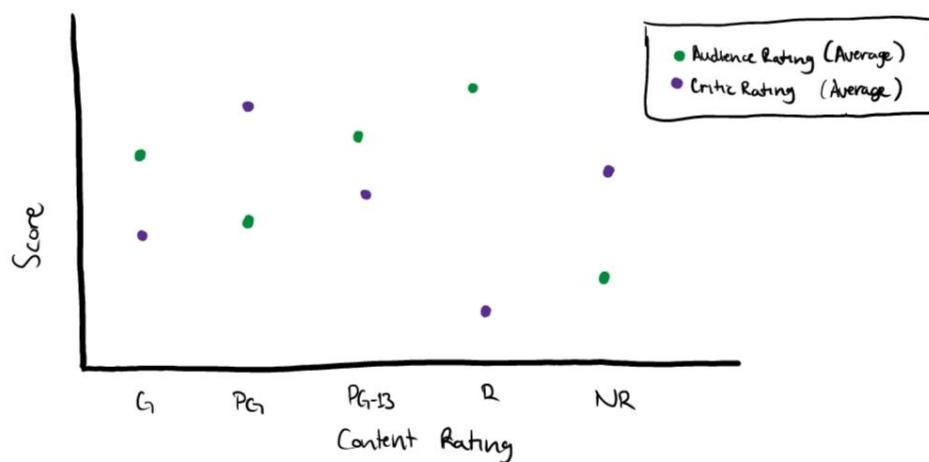


This graph shows Average Audience Ratings and Average Critic Ratings of movies based on their release date. This graph highlights how different a critics rating can be compared to the audiences rating. This graph shows more information but still doesn't allow users to draw their own conclusions and is limited to 3 attributes.

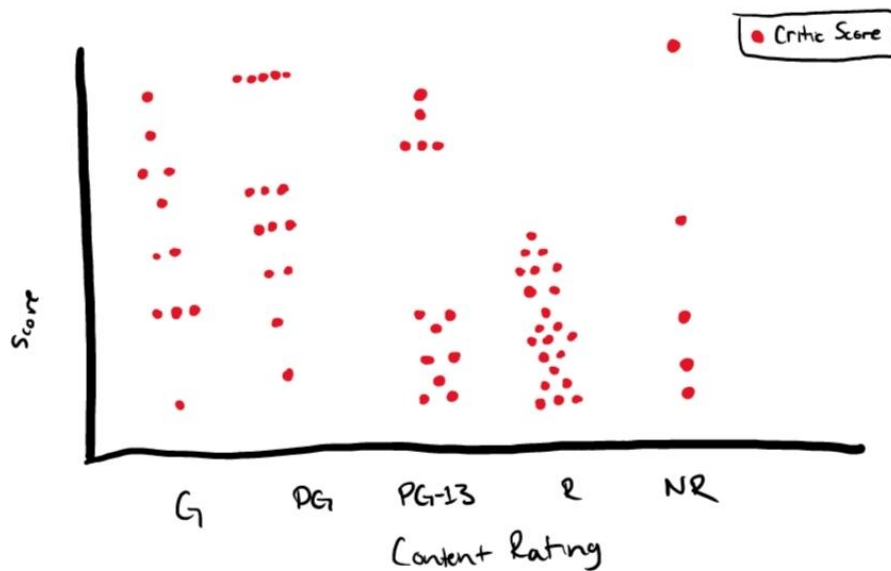
Critic vs Audience reviews



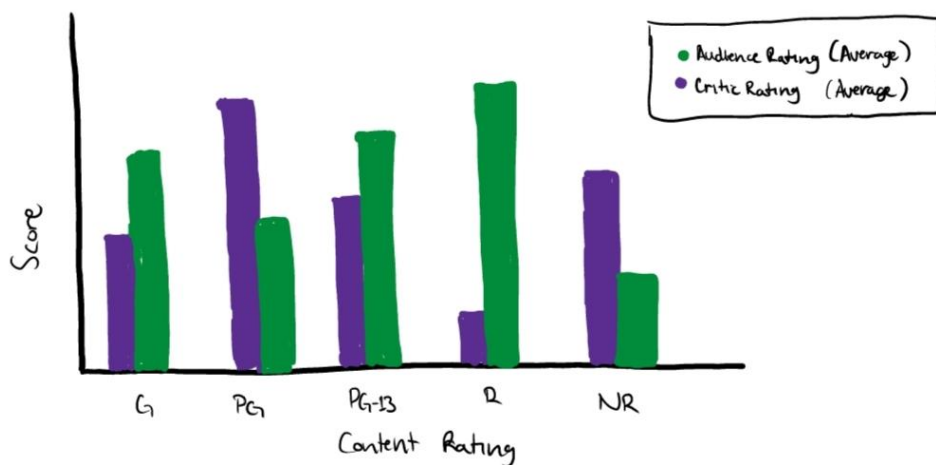
This graph shows a few different attributes, it shows the year a movie was released, the average critic rating (black line), average audience rating (blue bar), a 60% average of the average critic rating (dark gray bar), and a 80% average of the average critic rating (light gray bar). The gray bars really do not give any information about the ratings as they've been computed as the average of an average which has skewed the data so much it would only lead viewers to misunderstand, and it's honestly confusing to look at. This graph could have potentially been used without the gray bars, but for now it will stay an alternative design.



This graph shows the relationship between content rating and the average Rotten Tomatoes score from both critics and the audience for each given content rating. This design does a good job at allowing the viewer to see the difference between audience and critic scores as well as differences between scores between different content ratings.

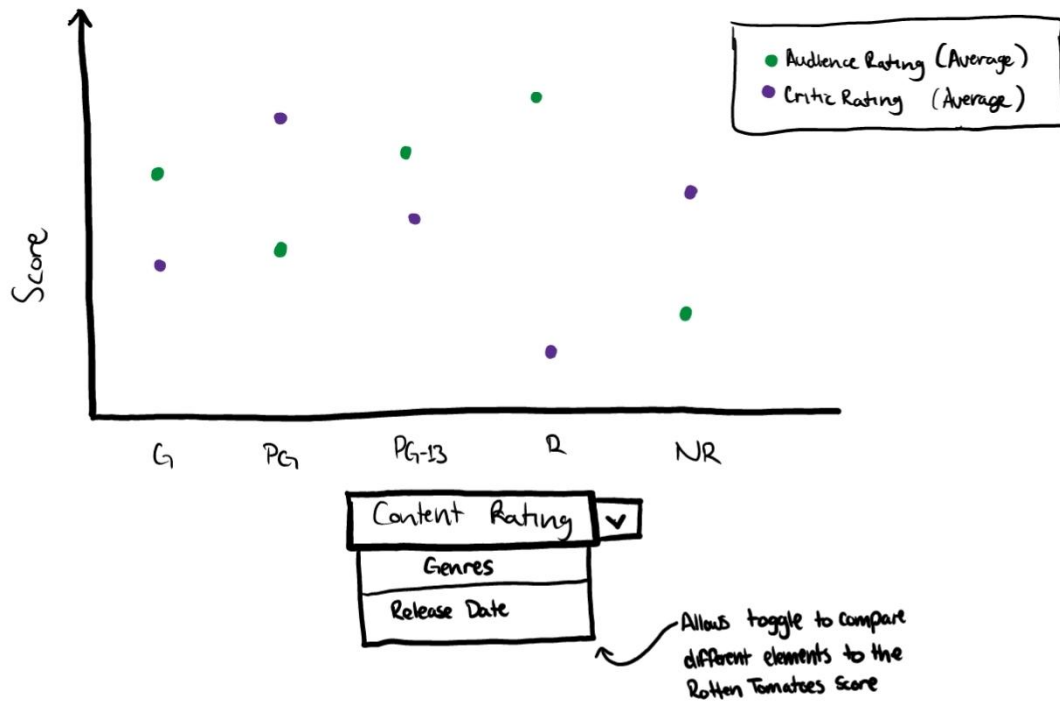


This visualization allows the user to better see the critic's score of individual movies given their content rating than the previous, since each score is individually plotted. Although, the visualization can become easily clustered with a large amount of movies. An implementation of this visualization can also allow the user to toggle between the critic and audience scores of individual movies under a given content rating.



This visualization is an alternate version of the dotplot which plots the average critic and audience scores under a given content rating, but uses bars instead of dots at the visualization technique. This allows a user to better see differences in scores across different content ratings since it is easier for the eye to decode than a point on a graph.

Final Design:



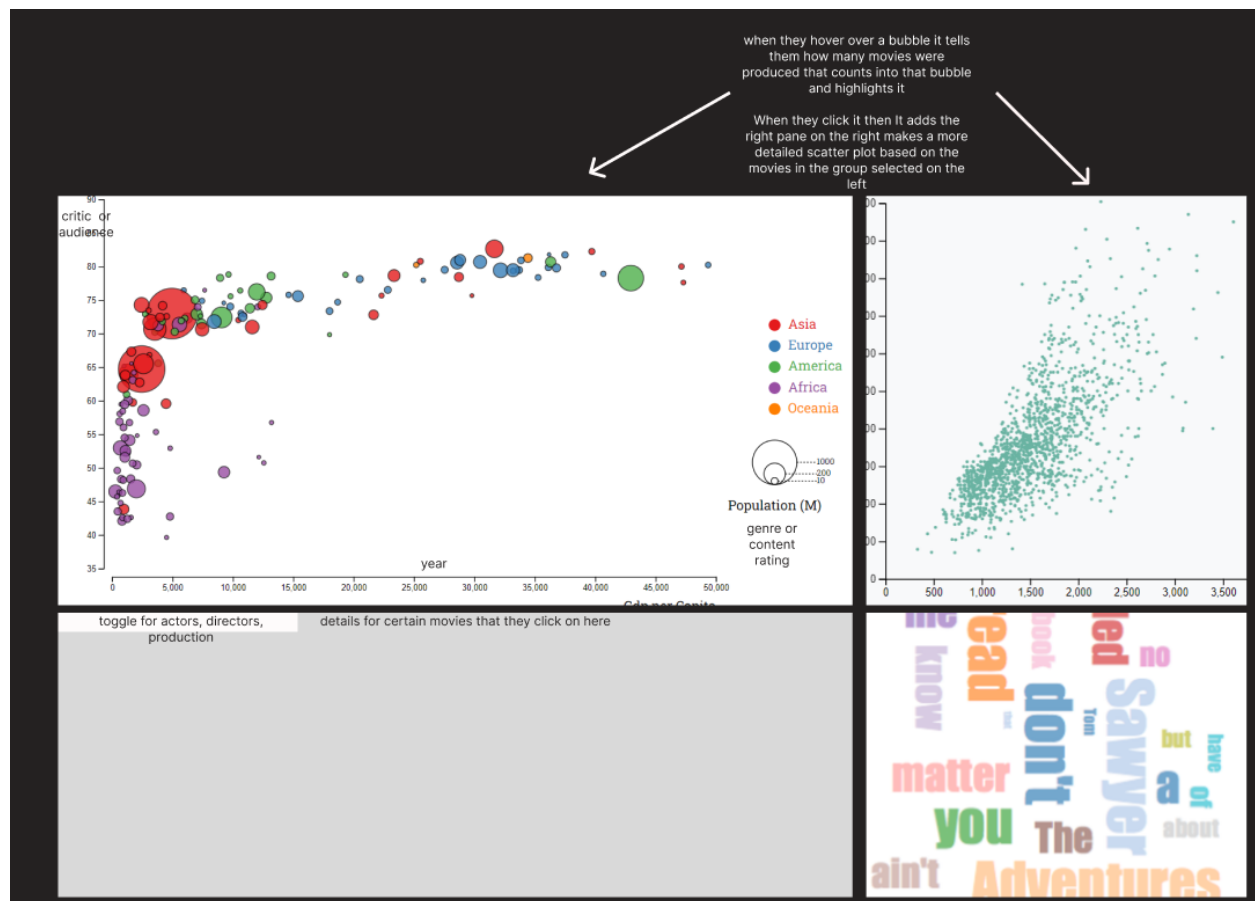
For our final design, we wanted to give viewers the chance to interact with the data and see different views. In order to create a visualization that isn't too cluttered or complicated, we will give the user the ability to change the content on the X-axis. We hold rating/score constant on the Y-axis. Some of the things the user will be able to toggle through is Content Rating, genres, and release date.

The channels will also be different depending on the data being shown. When we want to show aggregated data, like averages, we will use bar charts. When we want to show more granular data, or change over time, we will use scatter plots.

## Design Evolution

The section above displayed what was originally designed. I was able to build parts of this design, but found that it didn't convey enough information in one screen. I decided to look at this visualization system in a different way. I wanted the user to have the same type of layout with multiple pieces of information instead of showing one at a time.

As a result I recreated my design in figma with similar graphs, see below.



I knew that I still wanted to answer the key questions that started this project. I wanted to know how the critic scores may change over time. I wanted more detail about each movie. I wanted a word cloud to help describe the genres I, or other users were not familiar with. The bottom left corner was a mystery, there was so much information in this dataset I didn't know what to display. Keeping the same questions and key ideas in mind, began to build this visualization system.

In the end I decided to make a scatter plot in the top left corner that showed the individual movies over time, highlighting each genre with a different color. The top right graph became a lollipop graph which compared the average critic score per genre with the average audience score per genre, still following the color scheme the scatter plot followed. The bottom right word cloud derived the most common words in the critic consensus from each movie within a genre. This word cloud also follows the color schemes of each genre. Finally it was time to design the bottom right corner. I remembered that one of the original questions this project was centered around was how certain movies perform when looking at the content rating. I decided to create a bar graph that would allow the user to view one genre at a time, but would allow for them to see how the genre performed in relation to content rating, per the audience and critic score.

Overall this was almost a complete rendition of the original designs, but they were more compact and helpfully contained more information at a glance.

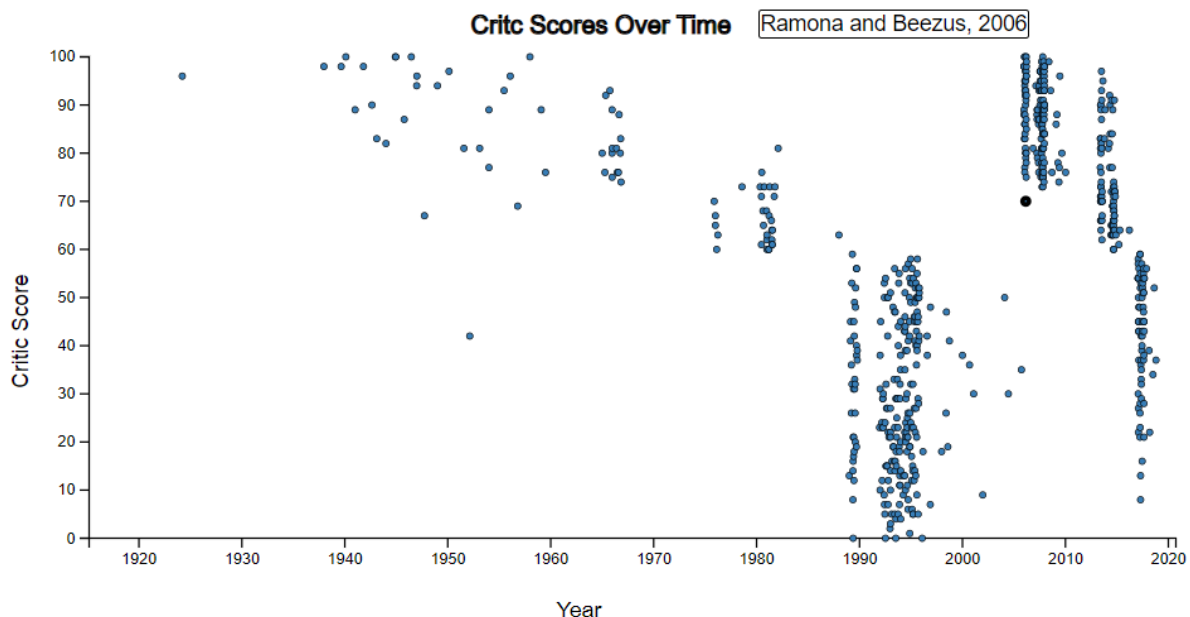


## Implementation

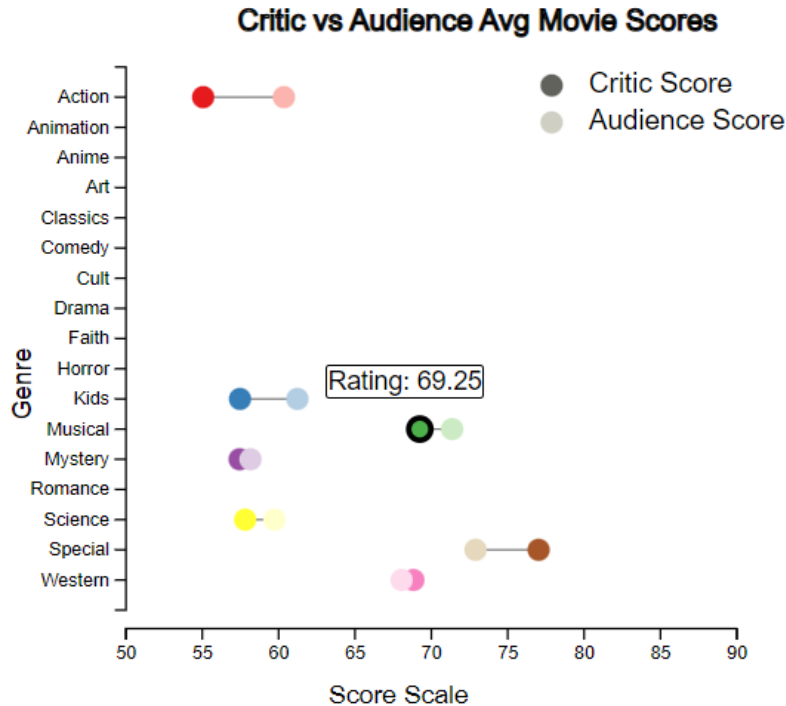
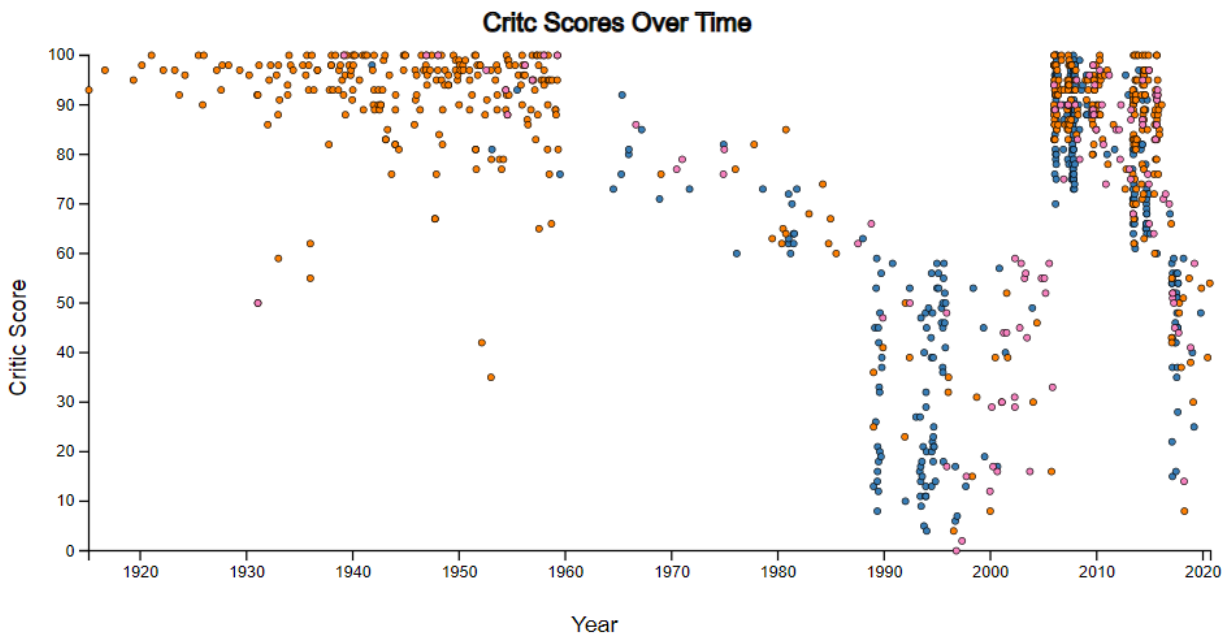
The overall design goal was for the user to be able to easily navigate this data visualization per genre. Each genre follows a specific color code and is implemented by selecting a checkbox on the left of the screen. A goal of mine was for this visualization to have a flow to it, considering english is read left to right, I considered giving the user their overall choice in that area, with a predetermined genre already selected so they would be able to infer how this visualization functioned. In the image below you can see that the genre “Classics” has been checked, and everything in this visualization is colored orange to show that.



Next, because this dataset is so large, I wanted them to be able to interact with the data and see exactly what data is being displayed. I included a tooltip for each data visualization in the system so that the user can move over the specific data point to get that information. See examples below.



In this example, the mouse is hovering over a point that is highlighted black, which causes the tooltip above to showcase the movie and its release year. This graph is also able to show multiple genres at a time, so that users can view different genre trends at the same time. See below.

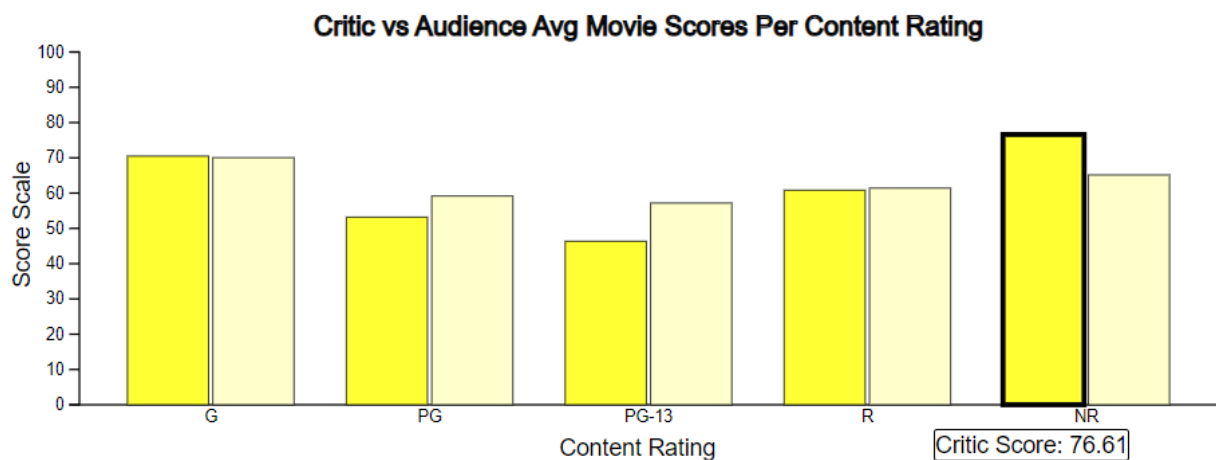


In the graph above, you can see that this visualization allows the user to select multiple different genres, which enables the users ability to compare the average rating given by both the critic and the audience.

Each data point can be hovered over to show the exact score and will highlight the point so that the user understands which point is being shown.



This graph is updated every time the user selects a new genre. It only shows one genre at a time so that there isn't any confusion and will help the user understand what a genre is about. This Genre is in relation to "Western" and displays that this genre is known for being a classic western genre with memorable performances, with an honorable mention to Clint Eastwood.



The final visualization in this system features another genre based graph. This graph allows for the user to compare the average critic score per genre with the average audience score per genre. This also allows for user interaction by hovering over a bar to showcase its average score.

With all of these visualizations combined the user will be able to get an understanding of how different aspects of a movie can alter the movie rating. For example, through moving through different genres the user will be able to see different trends, such as what type of movies perform well with different content ratings, or even what genres perform better at an audience level than a critics.

## Evaluation

Through evaluating this dataset with my data visualizations I was able to answer my questions. Using the scatter plot I found that over time, movie ratings have varied, but within the past ten years it looks like they're consistently better than they ever have been before. Then using the content rating bar chart I could see that it also seems like there may be some correlation between content rating and the critics score, from looking at the bar chart in most genres the NR and G tend to score better than any other. Finally, I was able to use the lollipop chart to see that the average critic score per genre is normally within five points of the audience's average score. This leads me to believe that the critics rating these movies aren't as far off as I once thought. A few more interesting conclusions I found were that, by using the bar chart I could see that on average critics score NR movies better than the audience did, which I found interesting since the R rated comedy movies were almost the same between the audience and critics. Then I found with the scatter plot that on average all movies released between 1989 and 2005 were rated much lower than any other time frame in the past hundred years. Last, with the lollipop graph, I found that on average, the audience tends to agree with the critics ratings on animation movies, but greatly disagree on anime and manga movies. I find this greatly interesting, since they're both animated in a sense.

Overall I feel like my visualization system answers all of the questions I wanted to find. There are many improvements to be made. I would love to add a brush to the scatter plot to be able to see the data in more detail. I would want the word cloud to show a percentage of top words from the critics consensus of each genre selected instead of one at a time. Finally I would like to implement the bar graph in a way that it would show two to three genres in one graph instead of just one, so that the user is easily able to compare the information.

## Extension

This project has been insightful and an exciting experience to be a part of. I look forward to continuing my journey in data visualization and honing my skills so that next time, my chart will have a zoom and pan function.