Final report HSH40 (GitHub: https://github.com/rickkenobi0513/HSH40_Visualization_2023.git)
Title:
"Exploring Film Dynamics: Budget, Revenue, Ratings, and Runtimes"

Main Figure:
The main figure is a combination of three subplots:

Budget vs. Revenue with Rating Differentiation:
Scatter plot displaying the relationship between budget and revenue.
Color-coded by movie ratings.
Trendline overlayed for better trend visualization.
Each point represents a movie, with hover information for the original title.

Violin Plot of Ratings Across Eras:

Violin plot showing the distribution of movie ratings across different eras (years).
Box plots within the violins provide additional statistical information.
Hover information displays the original title of each movie.

Runtime vs. Profit with Rating Size Differentiation:
Scatter plot illustrating the relationship between movie runtime and profit.
Size and color of points represent the movie's average rating.
Logarithmic scale on the x-axis for better visibility.
Hover information for the original title of each movie.

Legend:
Legend for Budget vs. Revenue Plot:
Points: Individual Movies
Trendline: Ordinary Least Squares (OLS) Regression Line
Colors: Rating Categories

Legend for Violin Plot:
Violin Plots: Distribution of Ratings
Box Plots: Statistical Summary within Each Violin

Legend for Runtime vs. Profit Plot:
Points: Individual Movies
Size and Color of Points: Movie Ratings
Logarithmic Scale on X-axis
Findings:
Strong positive correlation between budget and revenue, with variations based on movie ratings.
Movie ratings have evolved across eras, with the distribution of ratings showcased in the violin plot. This might cause by the annual numbers of movies is higher compared to the old time.

No clear correlation between runtime and profit, but for shorter runtime film (about an hour), it does seems like the rating is worse than those within one and two hour.

Data and Method:
Data:
Two datasets, 'new' and 'whole,' were used, with missing values dropped in 'whole.'
Data includes columns such as budget, revenue, rating, era (year), vote average, runtime, profit, and original title.

Method:
Visualization created using Plotly Express in Python.
Scatter plots, violin plots, and trendlines used to explore relationships.
OLS regression applied for trendlines.

Significance Statement:
The presented figure is crucial for understanding the multifaceted dynamics of the film industry. It combines financial aspects (budget, revenue, profit), audience response (ratings), and temporal trends (eras). This holistic view can aid filmmakers, producers, and analysts in making informed decisions about budgeting, content creation, and understanding the evolving landscape of audience preferences over time.
In addition to the three plots above, there are also several plots regarding the genre and the profit. They are in the GitHub repo as well.

Budget and Revenue

Rating of the age
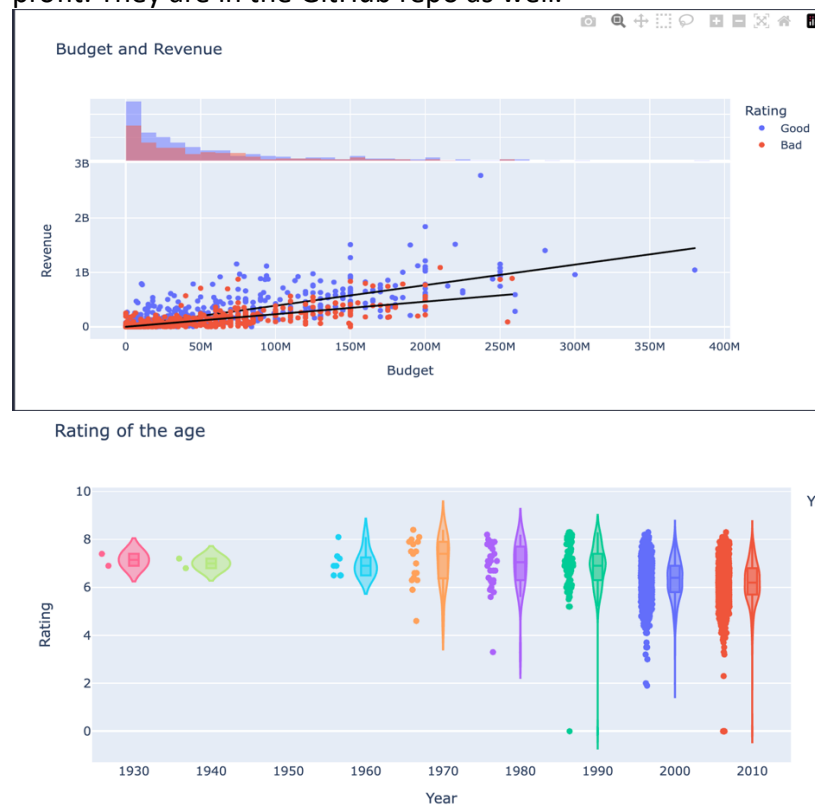
Does movie runtime matter to the profit?
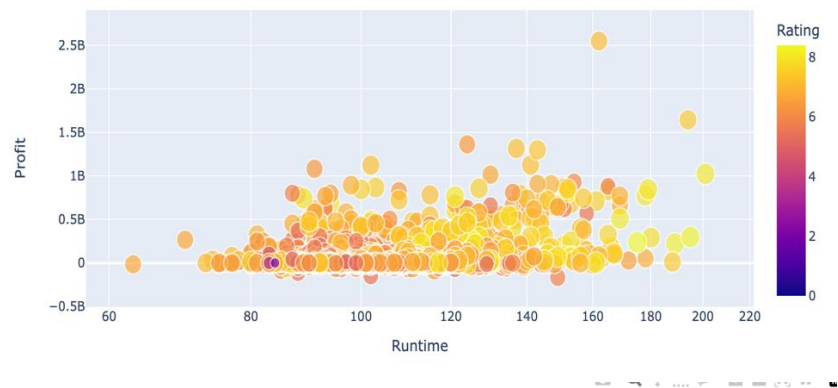
Total Revenue and Rating through the years