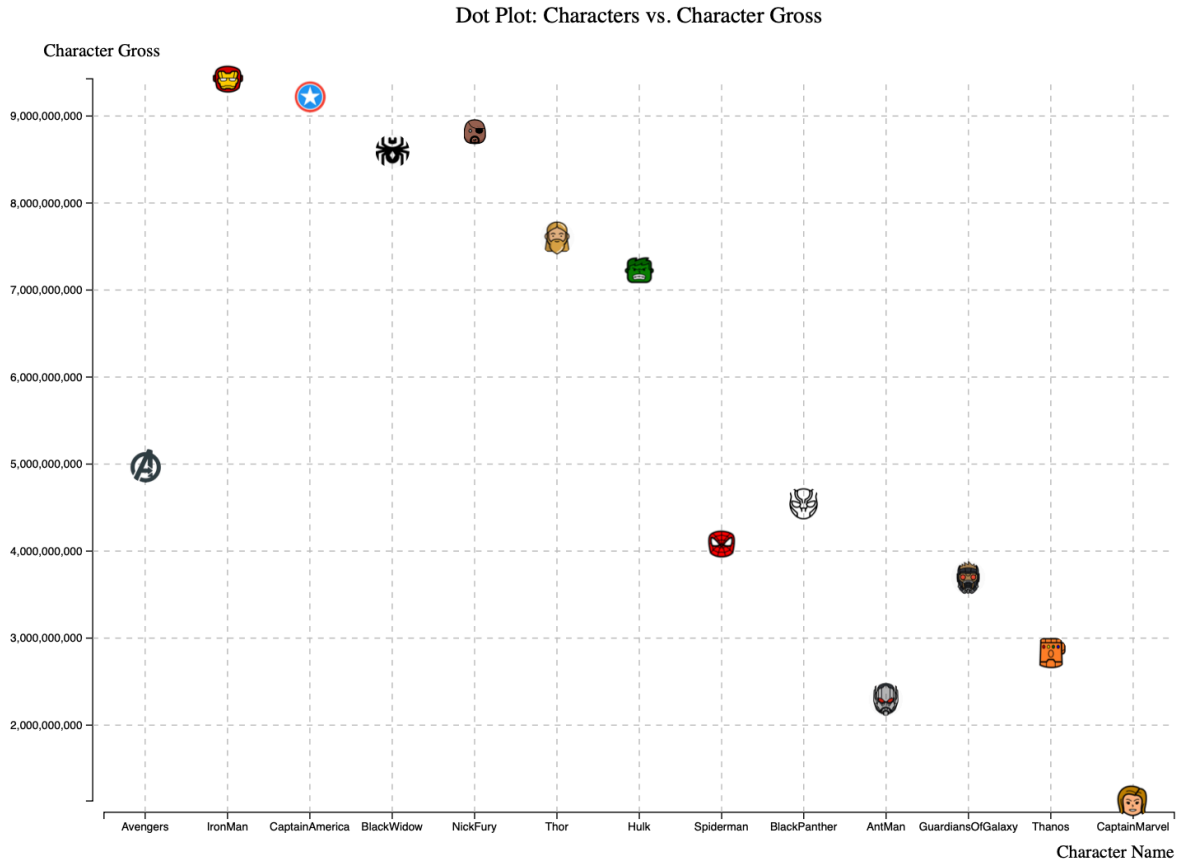
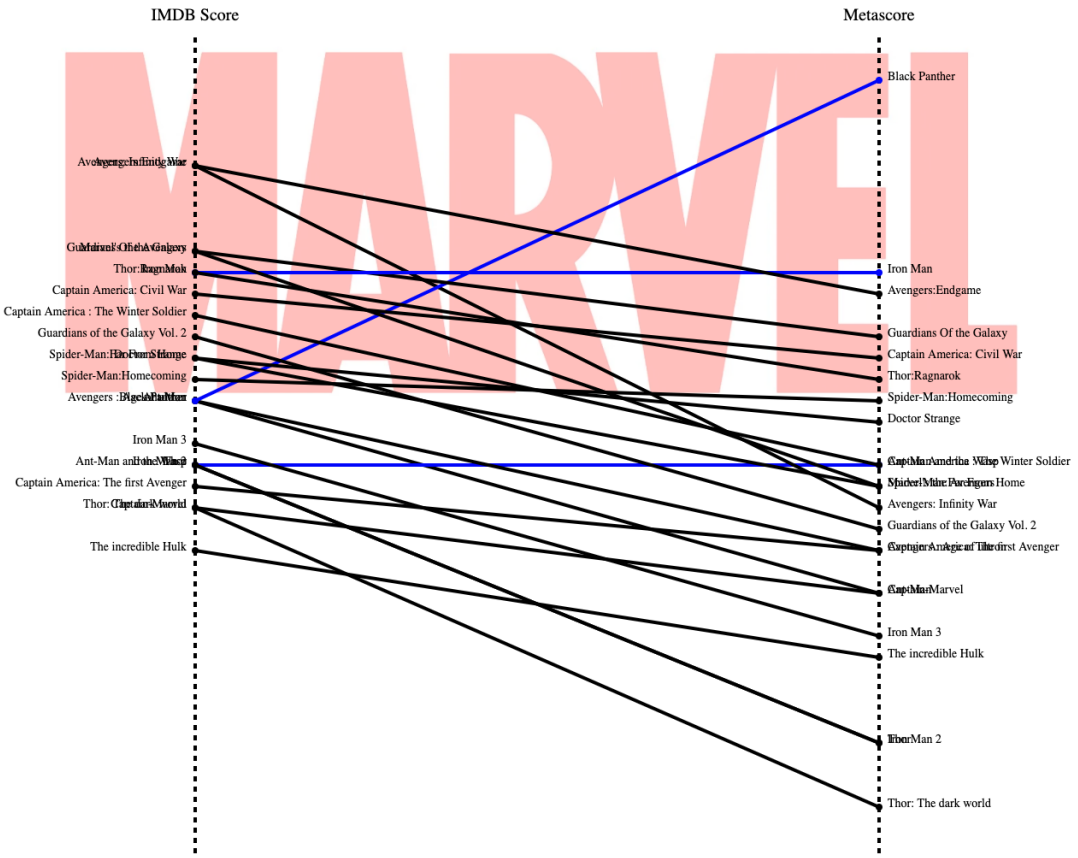
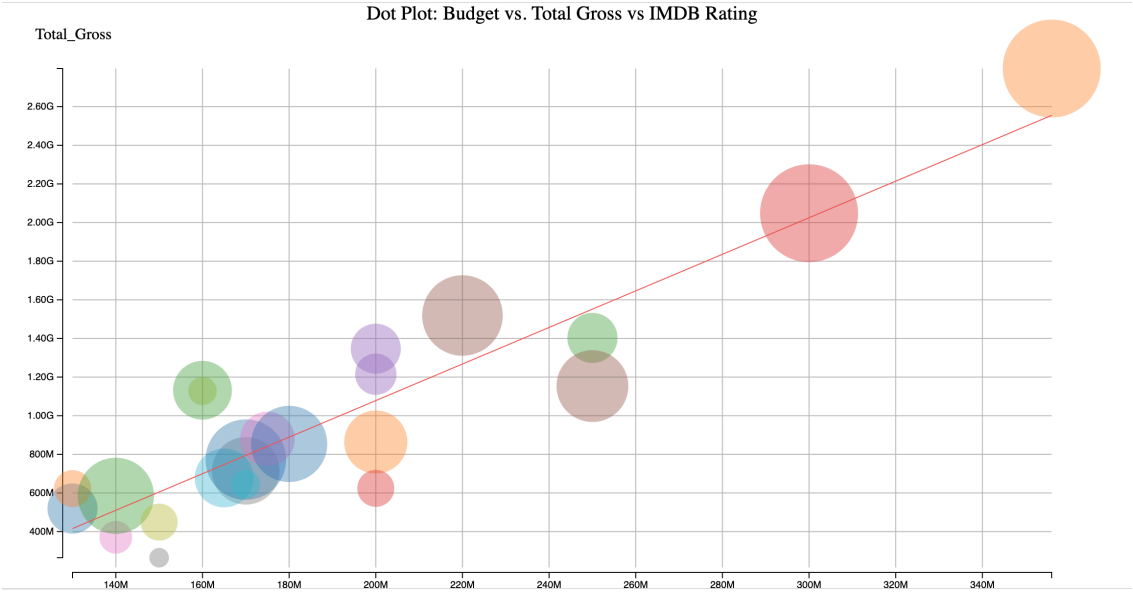
a) Screenshot of final, static visualization.



Slope Graph: IMDB Score VS. Metascore





b) A description of the data.

This is a notebook to explore the MCU dataset (https://www.kaggle.com/bcruise/mcu-analysis/data). The dataset contains 16 columns holding information about all the movies in Marvel Cinematic Universe from 2008 up to 2019, including their budget, domestic gross, oversea gross and ratings (IMDB and Metascore). The original dataset was scraped from the IMDB website. To gain a better insight into the gross that each Marvel Superhero brought, we integrated the MCU dataset with the Marvel Universe Dataset (https://github.com/pramodkumavat/MCU/blob/master/MarvelUniverse.csv). The Marvel Universe Dataset counts up the number of appearances for each superhero in each movie.

All data processing was completed using Python.

Preprocessing

We dropped those overlapped columns (e.g. domestic gross, overseas gross …) and concatenated two datasets. One thing to notice is that the Marvel Universe Dataset only contains information on 22 movies (excl. Spider-Man: Far From Home & Avengers: Endgame), so we filled missing columns with NaN.

One-hot-encoding

Before going into ML, we borrowed some knowledge from machine learning and used one-hot encoding to transform categorical variables in the Marvel Universe Dataset to binary values of 1 or 0.

Adding Variable

To visualize how each superhero contributes to the gross of the Marvel Cinematic Universe, we added a new variable called “Character Gross”. “Character Gross” is calculated by adding the total gross of movies that each character appeared in. For example, Ant-Man appeared in two Ant-Man movies as well as the Captain American: Civil War; therefore we added the total gross of these three movies and assigned that value to be Ant-Man’s character gross.

c) An overview of your design rationale.

1. **Slope Graph: IMDB VS. Metascore**

Marks: lines, color hue

Channel: slope of lines

We chose a slope graph for this visualization because slope graphs can show the relationship between two relevant variables, especially when IMDB and Metascore share a roughly similar scale. Different line colors are used to make further distinctions. Lines with positive slopes are in blue, while negative slopes are in black. A “Marvel” background logo was added to improve the visual effect. Not only because the red color in the logo provides a high contrast with the blue and red line, but also the logo clearly displays the theme of our design.

2. **Dot Plot: Box-Office Gross of Characters**

Marks: icons

Channel: horizontal and vertical aligned positions

We were looking for an ideal graph to display the connection between each Marvel character and the gross box-office return each of them created. In order to make the graph more vivid, we decided to combine the icons of each character into our graph, and it came naturally to use the dot plot and have the dots replaced by the character icons. Horizontal and vertical grid lines were added to help viewers to easily locate the name of and box-office gross created by each Marvel character.

3. **Dot Plot: Total gross and budget**

Marks: circle

Channel: horizontal and vertical grid line

We choose a visual graph to show the relationship between budget, total gross and IMDB score. Different colors of circles refer to different movies. In this way, it can help us to distinguish different movies since they are overlapping in the graph. The radius represents the IMDB score of each movie. Horizontal and vertical grid lines help me to intuitively realize the relation between budget and total gross. Finally we interpolate a red fitting line to help us get the conclusion.

d) The story.

Nowadays, there are many different movie rating systems including IMDB, Rotten Tomatoes, MetaCritic. Although the rating system institutions don't disclose their rating formula, from the result of the same movies in different rating systems, we can conclude that the rating formula is quite different. As for most of the movie viewers, it's hard for them to choose which rating system to trust and know the difference between them. Our project is to compare the rating of many MCU movies from two different rating systems, MetaCritic and IMDB.

MCU movies are known for a variety of true to life characters. Nevertheless, different characters' levels of popularity are varied. Our team aims to visualize different box office gross of different characters in MCU movies, in order to compare their popularity visually.

1. **Slope Graph: IMDB VS. Metascore**

IMDB score is voted by IMDb registered users (from 1 to 10) on every released title in the database, and these individual votes are then aggregated and summarized as a single IMDb rating. In contrast, metascore is the weighted average of many reviews coming from reputed critics from mainstream media such as The Guardian, New York Times, Rolling Stone, etc. Actually, Metascore owns a specific column called UserScore for normal movie viewers to rate. But it has little influence on the Metascore result. As usual, the Metacritic team reads the reviews and assigns each a 0–100 score, which is then given a weight, mainly based on the review's quality and source.

Our visualization tells viewers the comparison of Marvel movie scores under the two different rating systems. The comparison allows us to seek out what Marvel movie is more supported by fans (blue lines), and what Marvel movie is higher valued by the critics (black lines).

Both of the two movie rating systems have their downsides: IMDB score can be inflated by multiple fake accounts created by users, while metascore's weighting coefficients are confidential, so we don't get to see the extent to which each review counted in the metascore.

From our result, there is no absolute relation between these two ratings because the aesthetics of professional critics is different from the ordinary audience's. However, for most of

the movies, the difference between two rating systems is little and their scores are in a certain range. So we conclude that both of the two rating systems are trusty because their results are consistent. As for which one to trust, it depends on the viewers.

2. **Dot Plot: Box-Office Gross of Characters**

The visualization reveals the box-office gross created by each Marvel Character. Iron man is the character who has created the most revenue, followed by Captain America. Capital Marvel contributed the least box-office gross among all the characters. The median box-office gross was created by the team of Avengers, which is about $5,000,000,000. The box-office gross also shows us the popularity of Marvel heroes. The most popular heroes are IronMan, CaptainAmerica, and BlackWidow; the least popular heroes are Captain Marvel, AntMan, and Thanos.

What's more, from our collected data, the more popular the characters are, the more times they appear in the movies. Also, many popular characters often appear in the movies at the same time. For example, IronMan, CaptainAmerica and BlackWidow always appear in the movies Avengers: Age of Ultron, Avengers: Infinity War and Marvel's The Avengers, all of which are the most successful Marvel movies.

3. **Dot Plot: Budget and total gross**

The dot plot shows the relation between the movie budget and total gross. The x axis refers to the budget of each movie, the y axis refers to the total gross, and the radius of the circle refers to the IMDB score. From the visualization, we can conclude that totally, the total gross is positively related to the budget. However, there is no absolute relationship between the IMDB score with budget or total gross. Therefore, if the producer wants the movie to be more beneficial, they should spend more money on it.

Outline of Team Contribution

- Chuhan Yu: data cleaning, created the slope graph, revised the dot plot
- Hang Jiang: revised the dot plot, collected the character icons, wrote the story

- Hedy Yang: created the dot graph, wrote the story and rationale
- Huaxiang Zeng: created an dot graph, wrote the story