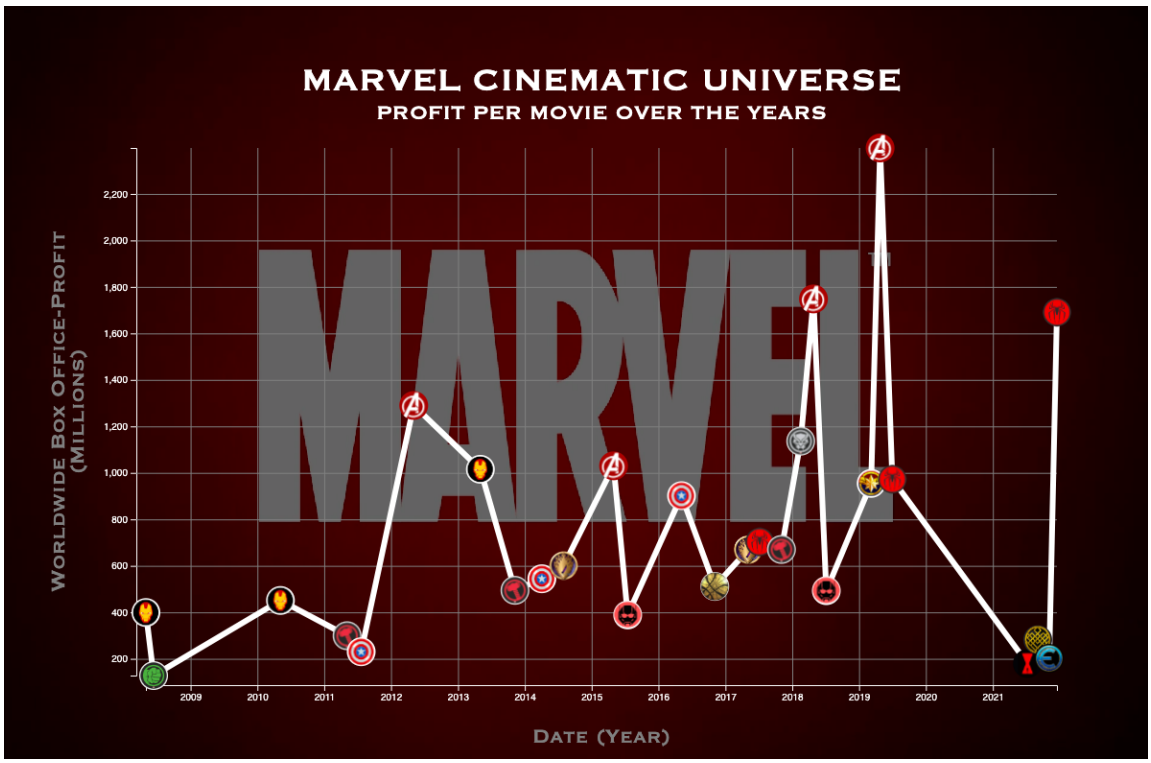
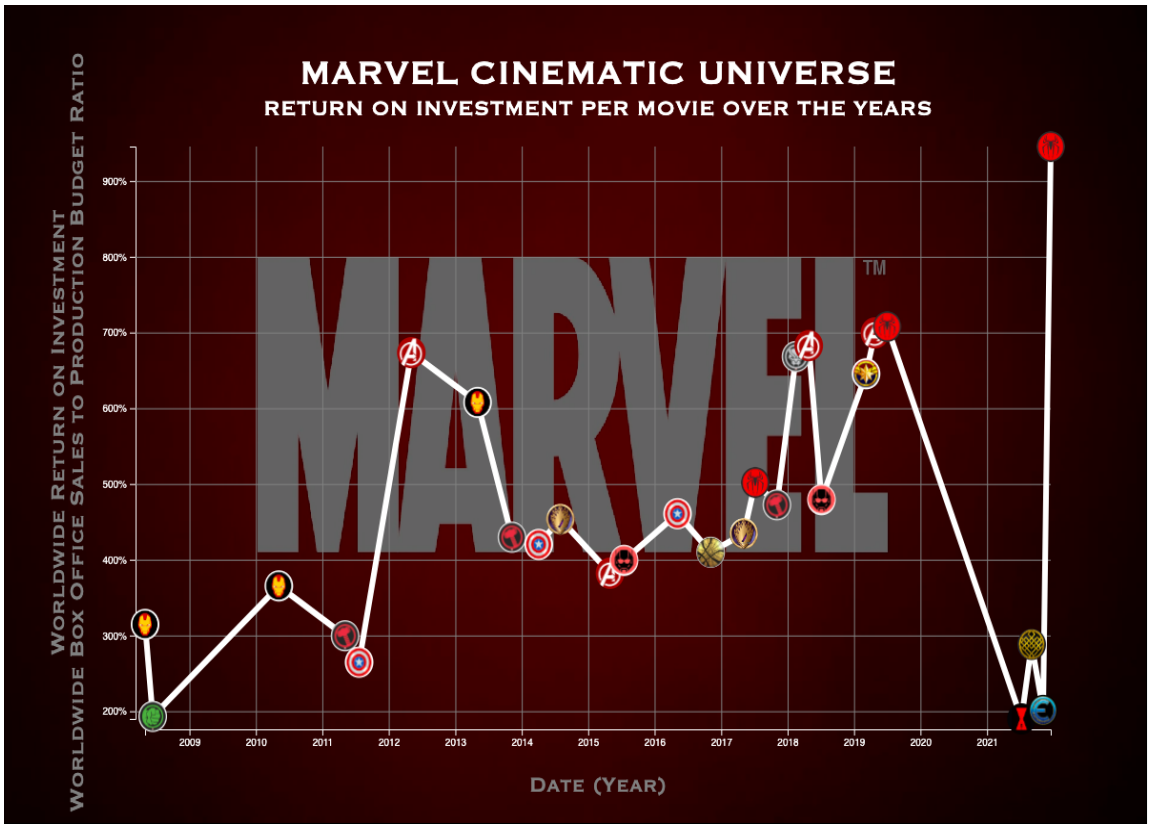
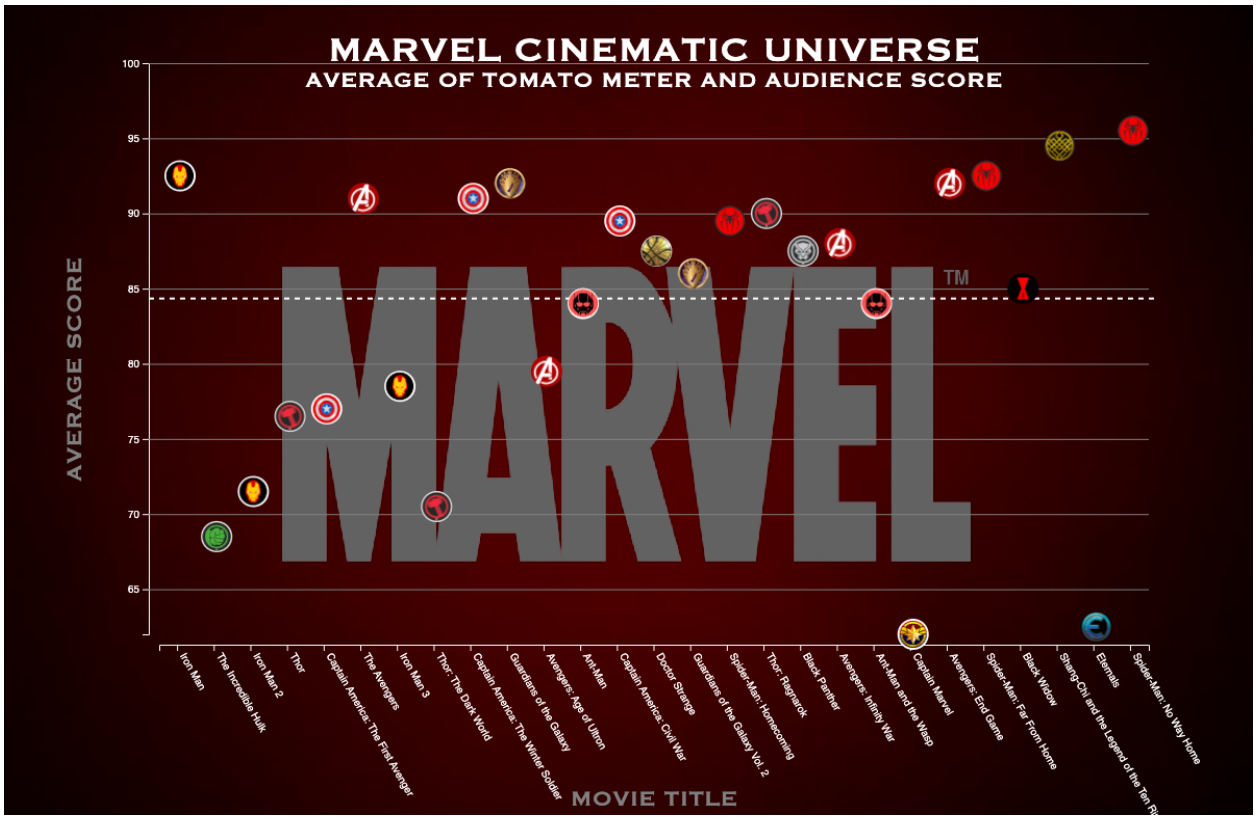
Visualization Project Final Report

Sabrina Li, Jackwin Hui, Vaidehi Raghu Raman







We took a Kaggle csv reporting Marvel movies throughout the years, their production budget, the worldwide box office revenue, tomato meter, and audience score to create our line graph and plot chart. Since everything was given in strings, we parsed the string dates into date objects and the production budget, worldwide box office, tomato meter, and audience score into numbers. We wanted to graph the profit made when canceling out the production value, so we did a simple subtraction and division to get the total amount made in millions. In addition to that, we took the names of the different movies and attached a marvel movie logo representing the series to each data point so it is easily referenceable. For our plot chart, we took the average rating for all the movies to create our line for reference and put the data points referencing each movie scattered along that line based on their rating.

For each of the graphs, we chose to have the time represented along the x-axis channel aligned by years of the movie's release date. This spanned over 2008 to 2022, starting with MCU's Phase 1 movie of *Iron Man* and ending with MCU's Phase 4 movie of *Spiderman: No Way Home*. This made the most sense from a design standpoint as we wanted to see how the MCU was doing over time and the different movies, whether that be the profit, return on investment, or their audience score.

We then decided to have the y-axis channel aligned by the respective data point we were comparing, which for the profit was the difference between the worldwide_box_office revenue and the production_budget, for the return on investment was the ratio between the

worldwide_box_office revenue and the production_budget, and for the audience score was the average of the tomato_meter and the audience_score. We found putting these datas along the y-axis would help show how they changed over time and between each MCU movie.

For each specific data point, we decided to represent each movie with their MCU logo, which should be quite familiar for most people familiar with the MCU movies. For the tomato_meter/audience_score graph, we also listed the names of the movies for people who may not be as familiar with the MCU movies.

The overall goal of the visualization is to give us more insights on how Marvel movies over the years compare to one another when it comes to money making, in regards to both profit (revenue minus costs) and return on investment (the ratio between revenue and costs), with the visualization of the average review score between the tomato meter and audience score for each movie providing some insights on the driving forces behind why some movies may have outperformed others.

Looking at just the visualization of profit per movie over the years, viewers can see some interesting patterns emerge. Standout data points are those that are abnormally high or low spikes relative to the movies around them; this includes all of the Avengers movies and *Spider-man: No Way Home* for movies with abnormally high profits and *Black Widow*, *Eternals*, and *Shang Chi and the Legend of the Ten Rings* with abnormally low profits. The fact that the movies with abnormally low profits are all clustered around the same release date suggests that their low performance in the profit sector may be driven by when they were released.

In addition, the viewer can utilize the visualization of average tomato meter and audience review scores to determine what some additional drivers behind these abnormalities may be since there is an expected correlation between the reviews a movie receives and its box office revenue. For example, compared to other movies released around the same time, *The Avengers* received significantly higher ratings which may have been a driving force for why it performed much better in terms of worldwide profit. In addition, *Spider-man: No Way Home* had the highest average review score of all the movies in the dataset, which likely contributed to its strong profit performance. Another key insight provided about these abnormalities from the visualization of review scores is that Eternals had a significantly low average review score, which likely helps explain its low performance in the profit visualization. Additionally, while Black Widow performed above the average review score for Marvel movies, it performed much lower than most of the other Marvel movies released around the same time period, which may have caused it to be below the expectations of fans and thus not gain as much profit.

Looking more closely at the return on investment visualization, the viewer can gain some additional insights on the abnormally high spikes in the profits visualization. For example, comparing the Avengers franchise across these two visualizations shows us that the third and fourth movies in the series–which had the two highest profits of all the Marvel movies–actually did not outperform the other movies as much (or were even on par or slightly lower than other movies released around the same time such as *Black Panther* and *Spider-man: Far from Home*) in terms of return on investment. This tells the viewer that despite acquiring significant box

office sales worldwide, the Avengers movies didn't actually make much more relative to the amount invested in them compared to other movies that made significantly less profit than they did.

Looking outside of these abnormally high and low spikes, we would expect the viewer to see a steady increase in the profits of Marvel films over time. Our visualization of the average review score once again conveys information on why this may be happening as from the graph, we can see a positive trend between time (since the movies are ordered by release date and in the same order as the other two visualizations) and average review score. This suggests that overtime, Marvel has been improving the quality of their movies, which likely accounts for some of the upward trend we see in profit per movie over the years.

Another interesting trend that viewers may notice is that movies that are part of the same franchise (i.e., Iron Man, Thor, Captain America, Spider-man, Avengers, Guardians of the Galaxy, and Ant-man), which can easily be seen as they have the same logos on the visualizations, tend to follow a pretty similar pattern in the profit per movie and return on investment per movie visualizations: the first movie in the series is the lowest performing and each new movie outperforms the last. This isn't particularly surprising, because of our previous discussion of Marvel potentially improving in quality over time, and because as more movies are released, the franchise is likely to gain more fans and thus more revenue in theaters.

What is surprising is that the Avengers franchise, which is likely one of the most well-known franchises that Marvel has, did not follow this pattern in a significant way. The second Marvel movie actually significantly underperformed the first one in both the profit and return on investment visualizations. Once again, our visualization of average review score provides a strong indication as to why this may be. The second Avengers movie had a below average review score that was also significantly lower than the scores of other movies released around the same time, which likely contributed to its underperformance compared to the first Avengers movie, which had a score that was much higher than the average score and also higher than the scores of other movies released around the same time frame.

Thus, these are some examples of the ways our viewers are able and likely to interpret our visualizations, providing insights on what some of the most and least money-making Marvel movies are and what trends exist amongst Marvel films in terms of money making, along with showing that average ratings that the movies receive from the tomato meter and audience scores can help explain some of these trends.

Team Contributions: Overall, every team member contributed to each of the visualizations, whether it be writing large amounts of code or making small tweaks. Some of the more specific contributions for each visualization are broken down by each team member as follows:

Sabrina

- Found and selected dataset to be used for our project
- Created the time parse function for release dates and created the date scale function ~10 min
- Wrote code to create axis lines ~ 10 min
- Wrote code to import data into the visualization and create the paths ~ 1.5 hours including debugging
- Adjusted code for tomato meter and audience scores to scale with graph ~ 1 hour including debugging
- Wrote the data selection section of report ~30 min

Jackwin

- Cleaned data and selected usable/relevant data columns (eliminated movie_duration, opening_weekend, domestic_box_office; kept production_budget, tomato_meter, audience_score, worldwide_box_office) ~ 15 minutes
- Parsed string data of $ values into usable numbers and created the profit scale function ~ 15 minutes
- Handled the CSS for the axis labels and title as well as the background ~ 2 hours
- Wrote the “design rationale” section of the report ~ 30 minutes

Vaidehi

- Looked for and downloaded pictures and logos of Marvel movies to be used as the dataset points ~30 minutes
- Researched and coded the logo images to be used as data points
    - This took the most time as there were many methods available online, some of which were more difficult to make work than others. It took approximately 2.5 hours to find, code, and debug a method that worked for us.
- Created the gridlines and axis labels on profit visualization ~5 minutes
- Wrote code to make chart and plot data for return on investment visualization ~45 minutes including debugging
- Wrote code to make chart and plot data for average tomato meter and audience score visualization ~1.5 hours including debugging
- Wrote the "story" section of the report ~2 hours