Your deployed webpage should also include a write-up with the following components:

* A rationale for your design decisions. How did you choose your particular visual encodings and interaction techniques? What alternatives did you consider and how did you arrive at your ultimate choices?
* An overview of your development process. Describe how the work was split among the team members. Include a commentary on the development process, including answers to the following questions: Roughly how much time did you spend developing your application (in people-hours)? What aspects took the most time?

Marvel Superhero Trends Visualization

Marvel series have always been a great hit. Yet there seems to be thousands of superheroes for readers to memorize, as well as the complicated relationships between them. We want to understand more about the heroes’ demographic stereotypes as a group, not just the few main ones. And by analyzing the trends in creating Marvel characters, we also want to see whether there are changes in values widely recognized by the society through the years.

Here we used two datasets, both from Kaggle. One is FiveThirtyEight Comic Characters Dataset (<https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset> ) with characters’ basic information including the year he/she was created, appearance frequency in comic books, hair color, eye color, and gender. The other dataset is Marvel Universe the Social Network (<https://www.kaggle.com/csanhueza/the-marvel-universe-social-network#hero-network.csv> ) which has the counts of interactions between characters. We used the two datasets to complement each other.

Rational for Design Decisions

Since we have two datasets, one about basic demographic information and the other about interactions, we want to use two visualizations to answer two questions: is there a stereotype transformation through the years or were the characters just randomly assigned with some features? Were the most frequent interactions between characters with some dominant features (either the gender or eye or hair color- which we think might indicate the race of the character)?

Therefore, with the two questions, we wanted to use a timeline for the first visualization with each character represented by a line, starting from the year he/she was created and ending at the year he/she deceased. Then we realized that by using lines, the variance of eye and hair colors would be really hard to recognize since the lines might be very thin to accommodate such large dataset. Besides, in the different years, the amount of characters created varies. By using continuous lines, we wouldn’t be able to show the rise and fall, because the height of compiling lines would be the cumulative amount. So we changed our idea to using small circles to represent the heroes, and using colors to reflect the character’s hair and eye color. In this way, we sort the circles by the heroes’ gender or eye color, and the height of lined-up circles would reflect the amount of characters created in that single year.

For the second visualization, at first we wanted to use the kind of dynamic graph where you can drag and pull the circles. We thought linking the circles in a 3D space and allowing readers to drag them would visualize the relationships in a clear way. Then after going to the details, we found that that type of visualization might be over our capability, so we switched to an arch representation prototype which we also like a lot. So here the second visualization is the final product with circles lining up on the bottom showing the character’s hair color, and links between the circles with different widths showing the count of interactions.

In general, we’ve realized the visual effects that are important for answering our questions. And the final products share similar structures with our visions at the beginning.

Overview of Development Process

The work was divided evenly between group members. We process the whole work in three steps, data wrangling and cleaning, visualization one and two.

Qianhui took charge of the data wrangling part. The main problem of our datasets were that the names of superheroes didn’t match in the two datasets which took us a great amount of time matching the names later. And another challenge is that the datasets are to large to represent. The first one has more than 6000 characters and the second has more than 65,000 records of interactions. So here Qianhui filtered the data based on appearance frequency and found the appropriate number of records to keep for visualization.

In the first

Timeline

Data nature