Analysis of Popular Comic Characters

Shaunak Sen

ss8n18@soton.ac.uk

# Data Story Summary

The topic I have chosen for the Data Story is **Analysis of Popular Comic Characters**. Through the data story, I plan on asking some questions and answering them through analysis and visualization. I also plan on promoting user interaction in my story so that users can explore further and answer questions that are of interest to them.

Few questions I pose in my data story are:

1. How is the popularity of comic characters related to their Character Alignment and Sex?
2. How is the popularity of comic characters impacted by Character Alignment, Sex and Alive field (whether the character is alive or not) - Take a deeper look?
3. What are the factors which affect Popularity - Strength, Durability, Intelligence, and Speed?
4. Who are the most powerful comic characters? - Which of them are super villains?
5. How many of the most powerful comic characters are female?
6. How popular are the most powerful characters?
7. Which are some most common superpowers? - Which are some of the rarest superpowers?
8. Which characters apart from Spider Man can sense danger?

While posing these questions I have kept in mind the **Flow and Engagement** narrative design patterns. Questions are posed such that there is a **Gradual Reveal** of information. We start with a basic overview of a topic and explore the results until we reach our conclusions. **Exploration** is promoted by the use of visualizations, with which users can easily interact and derive their own insights and conclusions (e.g. Bubble Chart, Heat Map and Word Cloud used in the story all support user interaction).

# Dataset Summary

Here you should briefly summarise the data sources you used, describe where you retrieved them from, describe how you linked together multiple datasets (if applicable), and describe any processing you needed to do to the data.

## Data Sources

The Data Sources used are as follows:

### <https://github.com/fivethirtyeight/data/blob/master/comic-characters/marvel-wikia-data.csv>

### <https://www.kaggle.com/claudiodavi/superhero-set>

Additionally, I scraped information about each character in the first dataset from the website: <http://marvel.wikia.com/>. For each character, the superhero name and attribute scores like strength, speed, fighting skills, intelligence and energy projection were scraped.

## Data Preprocessing

In the first dataset, most of the data extracted was text based, and required a lot of pre processing, like removal of leading and trailing whitespaces, converting them to appropriate case, removing special characters etc. After cleaning, the dataset was filtered to get the top hundred most popular superheroes to be used for the analysis. The number of appearances of the character in comic books has been considered as the indicator to popularity of that character. Also, few of the most popular characters had missing values which I looked up from the website and filled in.

The second dataset did not have any missing values or garbage text data. However, the data was in a wide format. Each superpower had its own columns. The data was converted from this wide format to a long format, where a new column called “Super power possessed” and populated.

This process was done to help create the Word Cloud as well as the Heat Map visualizations.

# Visualisations

## Bar Chart to understand how Character Alignment and Sex impact Popularity

### Description

Provide a brief description of the visualisation you have used (include an image or screenshot). Describe how interaction has been used appropriately.

### Justification

Discuss why the chosen visualisation is suitable for your narrative, and the data. Describe how the visualisation makes the data readable by the intended audience, avoids bias or confusion, and emphasises the key data. You should justify why (and how) you are using each visualisation with regard to concepts covered in the course; make sure you cite any relevant papers.

### Narrative Design Patterns

Describe which narrative design patterns [1] your visualisation uses, describe why you chose to use them, and how they support your story.

### Strengths and Weaknesses

Describe the strengths and weaknesses of your visualisation; again, make reference to concepts covered in the course (making sure you cite any relevant papers).

### Improvements

Describe how you could modify your visualisations to make them even better suited to their task, and how they could tell your story better. What additional data sources could be incorporated to improve the story?

## (Title of) Visualisation 2

### Description

### Justification

### Narrative Design Patterns

### Strengths and Weaknesses

### Improvements

## (Title of) Visualisation 3

### Description

### Justification

### Narrative Design Patterns

### Strengths and Weaknesses

### Improvements

# Conclusion

Conclude your report by summarising your data story, the key features you tried to get across to your audience, and how your visualisations accomplish that. You can also include any lessons you feel you have learned through completing this course.

##### References

References should be cited in text as consecutive numbers, within brackets (i.e. [1], [2], [3], etc.). References should be in [standard IEEE style](https://www.bath.ac.uk/publications/library-guides-to-citing-referencing/attachments/ieee-style-guide.pdf). Make sure to properly cite **all** work that is not your own. An example reference section is show below.

1. B. Bach, M. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engehardt, U. Köppen, and B Tversky. “Narrative design patterns for data-driven storytelling.” In Data-Driven Storytelling, N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale, Eds. CRC Press, USA, 2018, ch. 5, pp. 107–134.