Declaration on Plagiarism

Name:	Rashmiranjan Das
Student Number:	19210554
Programme:	MCM1 (Data Analytics)
Module Code:	CA682
Assignment Title:	Data Visualisation
Submission Date:	13 Dec 2019
Module Coordinator:	Dr Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at http://www.dcu.ie/info/regulations/plagiarism.shtml, https://www4.dcu.ie/students/az/plagiarism and/or recommended in the assignment guidelines

Name:___Rashmiranjan Das_____ Date: 15-12-2019_____

Correlation between characters in Harry Potter and the Philosopher's Stone

Abstract:
Harry Potter is a fantasy novel series written by J. K. Rowling. The novel dives into the life of a young wizard name Harry Potter and his journey into the world of Magic. The story focuses on Harry's struggle to defeat Lord Voldemort.[1] The seven novels have also been adapted into eight movies. This aim of this project is to visualize the character interaction in the first book and movie i.e. Harry Potter and the Philosopher's Stone. Character dialogues were analyzed and their link with other characters was created.
As Harry potter and the philosophers stone is the first movie of the series. It focused on character introduction and interaction to strengthen the bond. Through this visualisation it was observed that Hagrid, who is viewed as a small supporting character has the greatest number of dialogues with Harry. Followed by interaction between Harry, Ron and Hermione, as their friendship is building in this movie. While the antagonist has one of the least dialogues in the movie.
I have used this opportunity to learn new languages and tools eg. D3js and Flourish

1. Dataset
As analysis were to be performed on books as well as the movies, data was scrapped from multiple sources. Book data was not readily available. It was extracted from data bricks file system. The data was in unstructured txt format. Movie dialogues were extracted from Kaggle public dataset. The Kaggle dataset wasn't clean and with very few contributions on Kaggle.[2] The data consisted of the speaker and each line spoken during the movie. The speaker and listener mapping were created through it. Data source for the second graph needed to be in a json format. Hence, a function was created to generate a json file of the character links. The json file is available on my Github link.
The data were extracted from multiple source and hence had VARIETY.

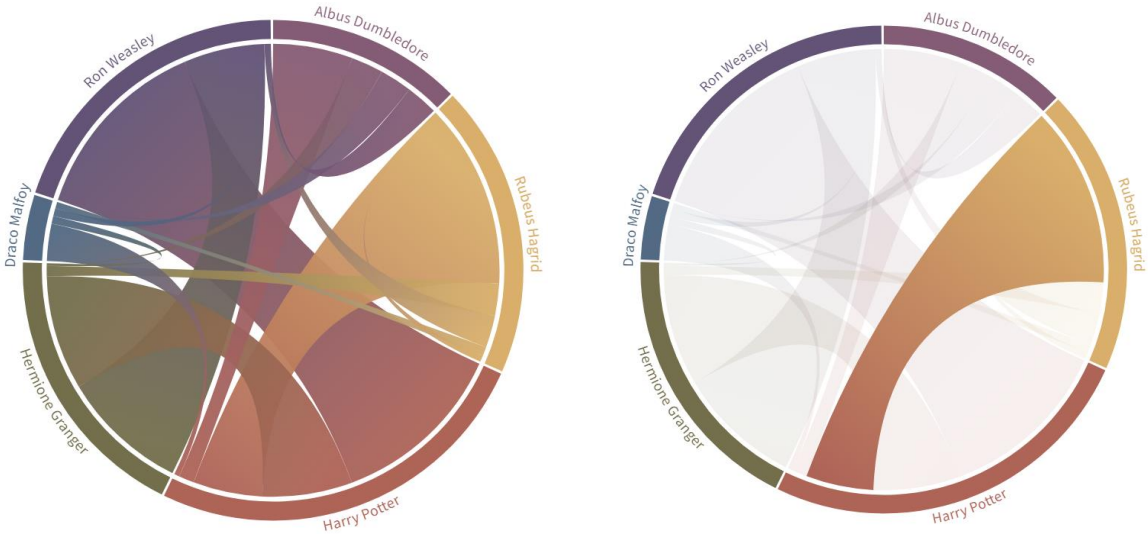
2. Data Exploration, Processing, Cleaning and/or Integration
The data extracted from Kaggle contained character dialogues. Generic characters like students, class, girl and guy had to be removed as they are not specific characters. Dialogues which are addressed to more than one person needed to be transposed into separate records. Dialogues which were addressed to houses for e.g. Gryffindor, Hufflepuff, Ravenclaw and Slytherin, had to be mapped to students of respective houses.

Graph 1: The speaker and listener of each scene were tagged together to create a new key. These tagged characters were grouped together along with their count.

Graph 2: The forced directed graph of every characters was created in D3.js.[3] D3 provides with barebones to work with, Hence each and every link and movements has to be coded. All characters needed to be connected to each other. An input file to the graph is a json file. A matrix was created to list every character who have interacted in the movie. This matrix was later converted into a json file. Every character is assigned a unique ID. Visual links were mapped using these links. Each character was assigned a group based on their relationship with harry. The json file is available on my Github mentioned below:
(https://github.com/Ranjandass/Visualization/blob/master/hp.json)

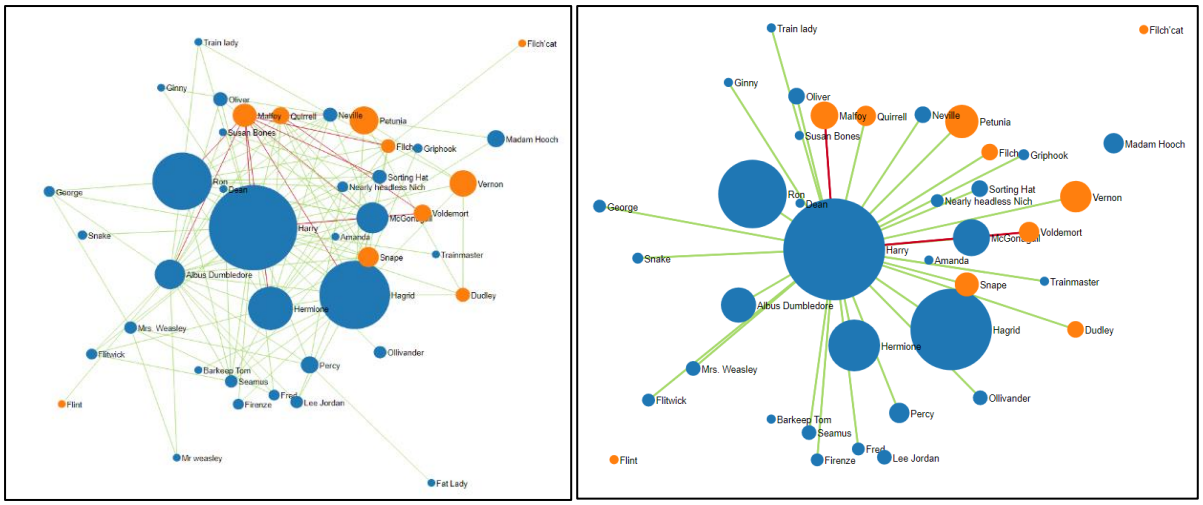
3. Visualisation

3.1 Chord Diagram: Flourish



This graph shows the interaction between top 6 characters in the movie. The link is based on the lines spoken between characters. The maximum number of lines conveyed is from Hagrid to Harry. The colour of the chart is set to antique to represent the era. Colour Green, Blue, Red and Yellow are used as they are colour of four houses in Hogwarts school of wizardry.
The visualisation is interactive, once hover on top of the chord. It highlights the selected chord and makes rest transparent. Number of lines spoken by each character is also displayed on selection.
(https://app.flourish.studio/visualisation/1099302/edit)

3.2 Force directed graph: D3.JS



1.Base Graph with all connections highlighted *2. On selection of single Node*

There are 40 characters in the movie. All the characters are related based on their interaction. Blue colour circles are positive character while orange is negative. The size of the circle is based on the lines spoken by that character. Radius of the circle is added by 4 so that even the smallest circle can be visible. The green lines denote healthy relation and the red lines

denote unhealthy relation. Names of the characters are displayed outside the circle (centre + radius).
The visualization is interactive and once a circle is selected all its links are shown and rest are hidden. Nodes can be selected and moved around using the force charge function which acts as gravity.[4]

4. Conclusion

The graphs created fulfils our objective to analyse the correlation between characters and their connectivity.
I have created the forced directed graph on characters from the first book. On availability of data of all the books. A graph can be crated for all the existing characters in the Harry potter universe.
Due to inexperience in D3js, I created a simpler graph and reused functionalities. The original idea was to create more complex graph but due to inexperience in loom-js.

5. References

[1] "*Harry Potter*," *Wikipedia*. 08-Dec-2019.
[2] "Harry Potter Dataset." [Online]. Available: https://kaggle.com/gulsahdemiryurek/harry-potter-dataset. [Accessed: 16-Dec-2019].
[3] "Harry Potter Characters." [Online]. Available: http://www.stencilled.me/post/2017-05-02-harry-potter-characters/. [Accessed: 16-Dec-2019].
[4] "Force Layout - D3 wiki." [Online]. Available: https://d3-wiki.readthedocs.io/zh_CN/master/Force-Layout/. [Accessed: 16-Dec-2019].