

# Data Visualization Term Project Report

## Game of Thrones Data Visualization

Yash Agrawal                      Debraj Dey

---

### I. Abstract

Game of Thrones is an extremely popular TV series and an equally popular book series. It has extensive communications between the characters and hence, we chose the Game of Thrones dataset for our term data visualization project. We implemented the following techniques on this dataset:

1. Force Directed Graph
2. Interactive Web
3. Interactive Bubble Chart

### II. Data Gathering

Finding the dataset was a challenge for us more than the visualization part. The Game of Thrones dataset was available on an open source website named Kaggle. The dataset comprised of 3 files:

#### 1. **battles.csv**

The battles.csv file contains all the details about each war that has happened in the TV series as well as in the books. This file contained information such as name of the battle, year in which it happened, who the attacker king was, who the defending king was, a list of their allies and the outcome of the battle.

#### 2. **character\_deaths.csv**

This file contains the details about the death of each character and how it happened. It contains fields such as the name of the character, their allegiance, or the house they belong to, the year in which they died, etc.

#### 3. **character\_predictions.csv**

This file consists of predictions made on each character's death. It contains probabilities about the death of each character, how popular they are etc.

We gathered all this data, but for each visualization, we have had to make certain changes to the dataset. We extracted the relevant information from each table for the visualization and used only that data. For the Force directed graph, we only needed the house and character information, so from the `character_deaths.csv` file, we

only chose the respective fields and created a json file for the visualization. For the Interactive bubble, we decided to extract only the character name, book, and series in which the character died, the gender of the character and used only this information for our visualization.

### III. Visualization Goals

Our visualization goals were as follows:

1. Find out the extent of communication between characters and find out the characters who communicate the most throughout the series.
2. Game of Thrones has had a “few” deaths over its course. So, we wanted to analyze the deaths in the series by Year and by Gender.
3. To organize the characters according to the houses and allegiances so that we may know the strength of each house.

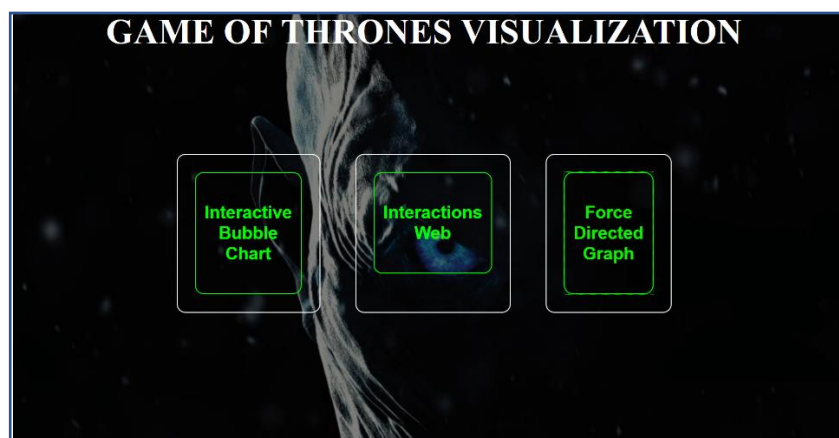
### IV. Technique and Results

The project was done using the D3 technology. There are extensive examples on the D3 website which served as guide for our project. D3 or Data Driven Documents is a JavaScript library that is used for creating visualizations.

Its basic components are:

- SVG
- Canvas
- HTML

The first thing we decided was to implement the HTML page which would have links to our visualizations. We also created a CSS file which we embedded into this HTML file for styling purposes.



*Figure 1: Opening page for our visualizations*

## 1. Interactive Web

To achieve our goal of finding the communication between characters, we thought that an undirected graph visualization was appropriately suited for task. Each character could be a node of a graph and the thickness of the edge between characters would represent the extent of their communication. The thicker the edge, the more those characters have communicated. This idea was taken from the example of the Les Misérables graph that was discussed in the class. Each node can be pulled out from the graph, and the graph will realign itself to better display the characters communications.

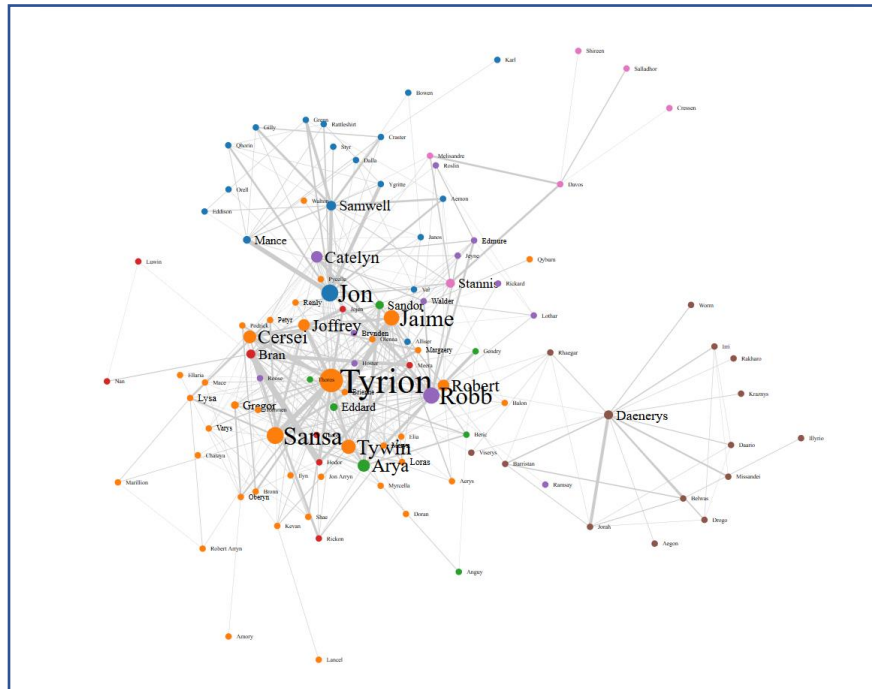
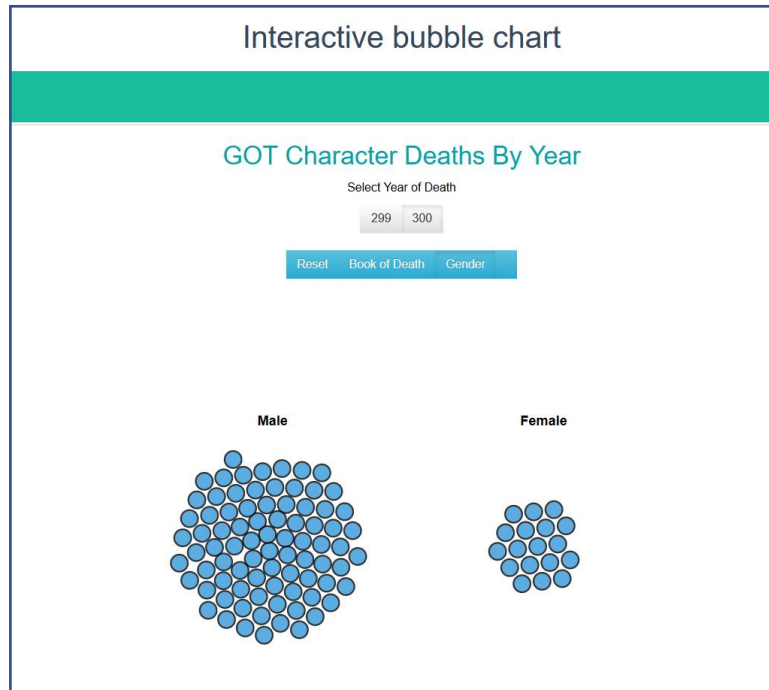


Figure 2: Interactive Web showing communication between characters

## 2. Interactive Bubble Chart

The next visualization technique is Interactive Bubble Chart. Here, our goal is to analyze the deaths in Game of Thrones, seemed to be best represented as clusters, since we had clear differentiating factors. Hence, we went for the bubble chart visualization since it would create clusters of the data and it would be appropriate for our analysis. To give information about each bubble, we also added the on mouse over functionality so that on hovering over the bubble, we get the information about the character. Also, we have taken two different Game of Thrones year 299 and 300. Another important features of this technique is the gender. It tells you the character which died of the series is either male or female. Of course, after analyzing the results we can see that more male passed away than females, due to different reasons like battles or

burglaries. Figure 3 shows the gender visualization. Also, to keep the bubbles intact, the reset button is used in this technique.



*Figure 3: Interactive Bubble showing deaths based on gender in the year 300*

### 3. Force Directed Graph

The final visualization technique is Force Directed Graph. To analyze the houses and its characters, we again chose to represent it as a graph. But this time, we chose to have each house as a node, and on clicking the node, each node expands to display its children. We chose this type of visualization since there are few houses, but each house has a lot of characters in it. So, having a static visualization of the graph would be too messy and too huge. It would just be unreadable. So instead, we give the user the choice of analyzing each house by itself. Also, this force directed graph is interactive and we can easily click on any edges and it will show up the other edges coming out of it. For example, if we click on house Greyjoy only the people associated with the house will show up. By this way we can hide the unwanted houses and nodes for analyzation purposes.

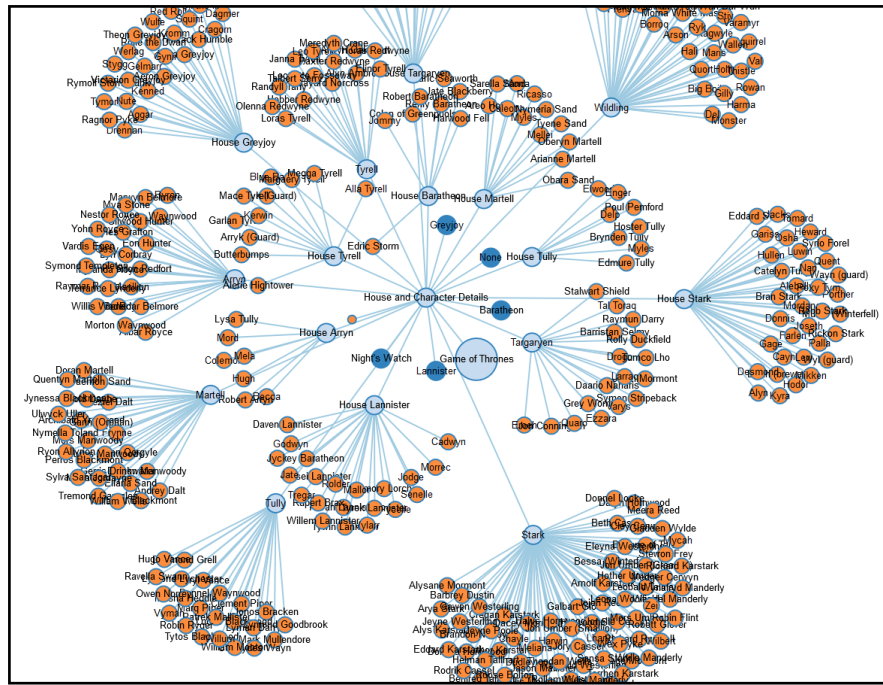


Figure 4: Graph showing the houses and its children

## V. Insights

There were a few insights we obtained from our visualization.

1. On analyzing the force directed graph, we realized that the survival of the house, does not depend on its size! House Baratheon was one of the strongest and most powerful houses at one point of time but is now almost extinct (known from the series and novel).
2. On analyzing the bubble chart, we realize that most of the deaths have occurred in the year 299 as compared to 300. This indicates that probably the year 300 has just set in or it could mean that these was a prolonged period of peace in 300 and the year 299 was a very battle intensive year.
3. Also, most people who have been killed have been males as compared to females. Even though we know that it is a very battle intensive series, where women are not fighting, we did not expect the number of women to be so many.
4. The interactive web reveals an interesting fact as well. Generally, the character with the most screen time is the one who has the most interactions. And we expected that Khaleesi or John Snow would be the biggest node. But, as it turns out, Tyron Lannister has the most interactions with people. This was a revealing fact to us. So, on further analysis we realized that Tyron is the only one who has had interactions with all the

major characters equally. John has extensive interactions with one set of people. Same goes for Khaleesi.

But Tyrion, interacts with both groups equally and hence is the largest node in the graph.

## **VI. Contribution**

Yash was responsible for collection of the dataset, modification of the data as required for the visualizations and creating the Force Directed Graph (as shown in Figure 4).

Debraj was responsible for creating the introductory HTML, the Interactive Web visualization, and the Interactive Bubble Visualization (as shown in Figure 1, 2 and 3). He also created the CSS file for the styling as required.

Both of us worked on the report, writing our individual parts and then combining them.

## **VII. Conclusion**

The goals that we had set for this project, were met by the visualizations we created. Also, useful insight was obtained by these visualizations, which revealed facts that went against popular notion.

## **VIII. References**

1. A dataset published on Kaggle April 2016 <https://www.kaggle.com/mylesoneill/game-of-thrones>.
2. The Official website Data Driven Documents <https://d3js.org/>
3. Deborah Mesquita “A gentle introduction to D3: how to build a reusable bubble chart.”  
<https://goo.gl/JmRqJF>