CS 3300 Project 1 Design Document

Michael Huang (mxh2), Abraham Meller (abm79), and Mahak Goel (mg785)

Introduction

Game of Thrones is a TV show like no other. Set in the fictional continents of Westeros and Essos, the show features an unprecedented amount of violence and we were curious about how deaths of major characters impact the popularity of the show. More specifically, how do the number of deaths in an episode and relative importance of the deceased characters contribute to the likeability and popularity of that episode?

A. Data Collection

Our variables include: the episode number, the likeability or popularity of the episode, the importances of the characters who died, the number of characters that died in the episode, and the house affiliation of the characters who died. In the section we describe how we collected the data for each of these variables.

- ***Likeability and popularity of the episode:*** The first thing we needed was the TV rating for every episode. At first we were going to use the actual TV rating, i.e. the viewership as the rating for that episode. However, we learned that the TV rating is just the number of people watching the episode when it first airs, which is not a true measure of the quality, popularity, or likeability of the episode because the people watching it when it first airs don't know anything about the episode. Therefore, we decided to use the IMDB rating for that episode as a measure of the popularity and likeability of that episode. It's a better measure because both users and critics rate the episode on IMDB after watching it, and IMDB displays the average of all of these views. This makes it a better measure because it accounts for informed perspectives and for a larger variety of them. We manually collected the IMDB ratings off the website because there are only 40 episodes.
- ***Importance of the character:*** For the importance of each character, we initially quantified it as the number of times their name was mentioned on the Game of Thrones wiki, divided by the total number of name mentions of all characters up to the point of that character's death. We collected this data by parsing through the Game of Thrones wiki, using a python parser. We made our visualization using this measure and saw a decent correlation, but we felt that the number of name mentions in a wiki is a very subjective measure because different Game of Thrones summaries can mention each character a different amount of times, and we wanted a more absolute measure of the importance of the character who dies. We then decided to use the cumulative screen time of the character up to the point of their death to quantify their importance because that's a more objective measure. We got the screen time data from a reddit page (the link is in the references section at the bottom) and copy and pasted the data into our json that contains all of our data.

- ***House affiliations of the characters:*** We used the same Game of Thrones wiki for the house affiliation of each character. We manually surfed the website and filled the data in our json manually.

You may notice that our graph only shows about 50 character deaths across the seasons, even though there are more deaths in *Game of Thrones* overall. The reason we only include about 50 characters is that we don't have the data for the rest of the characters who died because they had screen times below a minute and are presumably unimportant to the plot. We feel that this pruning of data was necessary to avoid clutter and get a better representation of how the number of important characters who died effects the rating of the show. Otherwise, we would have a lot of data markers for insignificant dead characters on the x-axis that may distract from the conclusions the visualization is trying to share.

We compiled all of this data into two json objects called episode_info.json and character_info.json. The episode_info.json has the IMDB rating for each episode and the character_info.json has the screen time and the house affiliation for each character who died in each episode. We use these two jsons for our visualization.

B. Visualizing the Data

Below we describe the different components of our visualization and how they were executed

- ***Episode Number:*** The x-axis of our visualization corresponds to the episode number. This serves as the x-axis for all components of the graph, specifically speaking, the IMDB rating of the episode and screen time of the characters who died in that episode.
- ***IMDB rating (likeability and popularity of the episode):*** The y-axis on the left of our graph presents the IMDB rating of each of our episodes. We plotted this as a line for any given point on the line where *x* is an integer between 1 and 40, i.e. *x* is an episode number, the y position of the point on the line is the IMDB rating for that episode. We developed this using a d3 line for which we set the x value to the episode number from our episode_rating.json and y value to the rating from episode_rating.json. We set the interpretation to linear in order to connect each point by a line.
- ***Screen time of the character who died (importance of the character):*** The y-axis to the right of the visualization shows the scale for the cumulative screen time of each character who died. It happens that coincidentally this screen time varies from almost 0 minutes to almost 100 minutes. Each dot on the graph represents a character who died; the x-coordinate of the dot is the episode in which they died and the y-coordinate is total screen time of the character, ie. their importance. We did this by using svg circles by selecting all circles and setting their attributes so that they have radii of 7 pixels, and their x-coordinate comes from the episode number in the character_info.json and y-coordinate is the the screen time also from the character_info.json.
- ***House affiliation of the character:*** We colored each dying character's dot according to their house affiliation. The legend on the right shows what house each color

represents. We implemented this by making an array in which the attribute is the house affiliation and the key is the color for that house. We then set the fill color attribute for each circle to the color for the house affiliation in the character_info.json by representing the array of colors we created.

- ***Seasons:*** We made a different colored rectangle as a background for the graph for each of the different seasons to be able to analyze trends within and across seasons.
- ***Axes and labels:*** We made the axes using SVG axes, and the labels and title by appending SVG text. We downloaded the Game of Thrones theme font from online and set the font family attribute to "GameOfThrones", the name of the font, for the title.

C. The Story

Our main purpose in making this visualization was to see how the importance of the characters who died and the number of important characters who died correlates with how much people like a given episode. What's interesting to see is that there is somewhat of a correlation. Notice that in a given episode, if a character who has a cumulative screen time above 60 minutes dies, the IMDB rating for that episode spikes up compared to the rating for the other episodes in the seasons. The most clear examples of this are:

- Ned Stark: Screen Time: > 90 min, Episode Number: 9, IMDB rating: 9.1
- Rob Stark and Catelyn Stark: Screen Time: almost 90 min, Episode Number: 29, IMDB rating: 9.8
- Jaime Lannister: Screen Time: almost 70 min, Episode Number: 32, IMDB rating: 9.6
- Tywin Lannister: Screen Time: > 80 min, Episode Number: 40, IMDB rating: 9.2

What's interesting about all of these characters is that their either from the House Stark or House Lannister, which are two of the most important houses in the show and the two houses that have a strong rivalry against each other. Ned Stark and Catelyn Stark were married and the heads of House of Starks, and their son Rob was trying to seize the throne. Tywin Lannister was the head of the Lannister house and his nephew Jaimie Lannister was a kid who currently held the throne, so these were all important people within their houses and within the show overall. Therefore, it's interesting to see that the episodes in which they died had a high rating because of the major plot twist.

Another thing that we can analyze in this graph is the trend of ratings within a season and across seasons. We can see that the except for season two, the second to last episode is the highest rated episode of the season, even higher than the last episode of the season. This might be because *Game of Thrones* tends to have a major plot twist at the second to last episode of the season and the last episode depicts the consequences.

One thing that is surprising about the visualization is episode 19 (9th episode of season 2). The episode didn't have any very importance deaths but it's still has a high rating of 9.3. This goes to show that death is not the only factor that makes a *Game of Thrones* episode good. Episode 36 (the 6th episode of season 4) is surprising in the same way with a rating of 9.4 and no non-negligible deaths.

It's also interesting to compare the screen times of the different characters who died. We can see that the Starks are clearly the one of the most important houses of the show, because they have the highest screen time and Ned Stark, who died so early on in the show in episode 9, has the highest screen time of all characters even though the show hadn't aired for a long time.

Overall, we conclude that the deaths of not very important characters don't necessarily mean a lower rating, but deaths of multiple important characters tend to contribute to a higher rating. For example we can see that in the last 4 episodes of season 4 many characters with a screen time above 20 died and the ratings of all these episodes are above 8.7.

References

- IMDB: http://www.imdb.com/title/tt0944947/
- Game of Thrones Wiki: http://gameofthrones.wikia.com/wiki/Game_of_Thrones_Wiki
- Reddit: http://www.reddit.com/r/gameofthrones/comments/2acke9/s4heres_a_breakdown_of_each_characters_screen/