

GAME OF THRONES

CSCI 403 PROJECT 9

May 4th, 2017

Taylor Chott, Erica Dettmer-Radtke, Grace Halbach, Carrie Kralovec

Colorado School of Mines



Contents

Introduction	3
Data	3
Methods	5
Technical Challenges	8
Conclusion	8
Reference	10
Tables	11

INTRODUCTION

Game of Thrones, a modern TV series based off *A Song of Ice and Fire*, George R. R. Martin's series of fantasy novels, is infamous for its unexpected and over-the-top death scenes. The overarching plot of the series is a power struggle between families (Stark, Arryn, Baratheon, Tully, Greyjoy, Lannister, Tyrell, Martell, and Targaryan) for control over a realm based on 1400s western European society. This struggle, and the talented actors who bring it to life, inspire strong feelings in viewers. Additionally, as *Game of Thrones* grows in popularity, more people provide ratings on websites like IMDb. The intersection of character death and episode ratings is the motivation for our project, which is a combination of database creation and basic statistical analysis. Through this project we will try to answer various questions of interest related to deaths and season ratings for the show *Game of Thrones*.

DATA

The data is especially interesting in the context of producers and marketing teams catering to a very invested fan-base in an attempt to keep ratings high and keep the show on the air. When organized appropriately, the data provides information about character deaths, character attributes (house loyalty, for example), total number of deaths, and ratings of episodes. Additionally, the data can also show relationships between the above categories. For example, querying the database can give information on patterns of deaths (Are there more deaths at the beginning of the season or end? Do later seasons have higher death tolls?) or the relationship between episode rating and the allegiances of killed characters.

The data was obtained from various websites. The tables containing data from the book, battles, character_deaths, and character_predictions were obtained from a website called Kaggle [1]. Kaggle is a platform that was created for predictive modeling and analytics competitions for which companies and researchers can post their data. This specific data set was pulled from other sources such as github and blogs.

The episode_ratings were obtained from IMDb [2]. This is a very reliable and commonly known online database that contains information about films, television programs and video games. This was extremely useful because it is well maintained and includes comprehensive ratings based off public opinion and votes rather than a single person.

The `screen_times` and the `got_death` tables were both originally obtained from data.world, a social network for data people to share cool and interesting data [3], [4]. From here, the tables came from slightly different sources. The screen times data also came from IMDb. The deaths data came from an article from Times Entertainment. All of the data for this project was easily accessible from online sources that we could then take and standardize for our database.

The data from Kaggle is released under CC0: Public Domain License. That is, the work from the these tables has been dedicated to the public domain. Data from IMDb and Data World are also available under creative commons release. From our searching, we were unable to find anymore restrictions on the data usage. Therefore we used the data assuming there were no restrictions that would prohibit us from using it for this educational based project.

The tables that we originally found to explore as described above are: `episode_rating`, `got_deaths`, `got_screentimes`, `battles` (book specific), `character_deaths` (book specific), `character_predictions` (book specific). The first three tables involving ratings, deaths, and screen times contain information from the HBO series. The next three tables contain information that is book specific only. For our purposes, we were more interested in data from the TV series. The `character_predictions` table contains a large amount of information that was generated using algorithms to try and predict when each character would die. This is interesting, but doesn't contain information that we can relate to the other tables. There is one column in this table that contains the allegiances that each character has to a specific house/family in the story. We were interested in this, so we decided to keep this table so that we could draw this aspect from it. The `battles` and `character_deaths` tables however did not have any use to us once we looked into them more. Most of our questions involved comparing things to the ratings that episodes had. We also found that `got_screentimes` contained information we did not necessarily care about when observing relationships about deaths. Thus we moved forward with the following tables: `episode_rating` and `got_deaths`.

The `episode_rating` table contains 5 different attributes. For all 6 seasons that have aired of *Game of Thrones* and each of the 10 episodes, it contains the season number, the episode number, episode name, rating from 1 – 10, and the total number of votes that contributed to that rating. The season and episode attributes are loaded as integers, the episode name is text, and the rating and votes are both numeric values. The primary key is `{season, episode}` to keep the unique constraint. From here, the `got_deaths` table is created with a foreign key referencing `episode_rating`. The deaths table also contains 5 attributes. There are 101 characters for which

there are recorded deaths. These characters are kept track of under the primary key *name*. For each character, we have the attributes role, what season they died in, which episode they died in, and execution (how they died). The season and episode are both integer values and the rest of the attributes are text. The episode and season are the attributes which the foreign key references the episode and season in the `episode_rating` table.

METHODS/RESULTS

Generic Stats

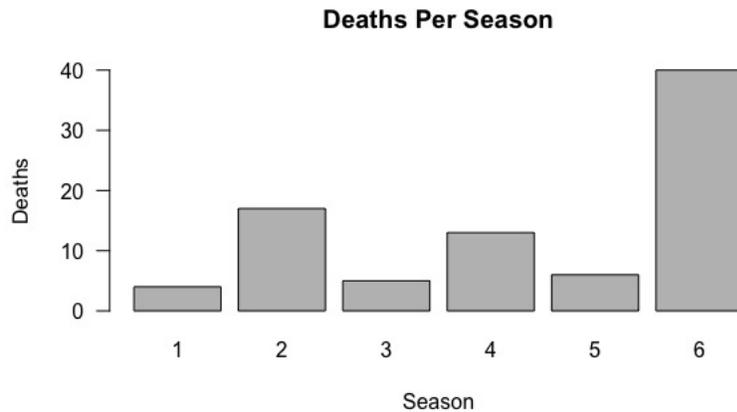
Through our query mash-ups of the tables, we were able to come up with various statistics on the data. For example, we found that there are episodes that do not include a named character dying. However, over the course of the entire show (60 episodes), there have been a total of 101 deaths of named characters, meaning the average episode contains approximately 1.83 deaths. Clearly this show is not for the faint of heart.

Compared to most shows, *Game of Thrones* has overall been rated fairly high. The lowest rated episode was season 5, episode 6. In this episode there were no deaths.



It appears that generally all the episodes have a range of about 0 – 4 deaths and most episodes are rated fairly high. The episode with the most deaths actually has one of the highest ratings. There seems to be other information like one really important death (major characters) over lots of deaths that make a rating increase. This data is difficult to find a relationship between

because there may be lots of other factors going into the rating. middle to high range and have a few deaths consistently occurring.



We were also curious which season had the most deaths. A histogram of this aggregated data can be seen above. Season 6 has the most deaths out of any season. It is interesting that the odd numbered seasons have less deaths overall. It could be interesting to see if this pattern persists in the next two seasons.

Deaths in a Given Episode

From our tables, we were able to create queries that could tell what episode or season that any given character died. From there, we decided to pull out all deaths that occurred in only the best episodes, or at least those with a rating above 9.5 (see Table 2). This table, which includes fan favorites such as Shae, Margaery Tyrell, and Wun Weg Wun Dar Wun, can be seen in the appendix.

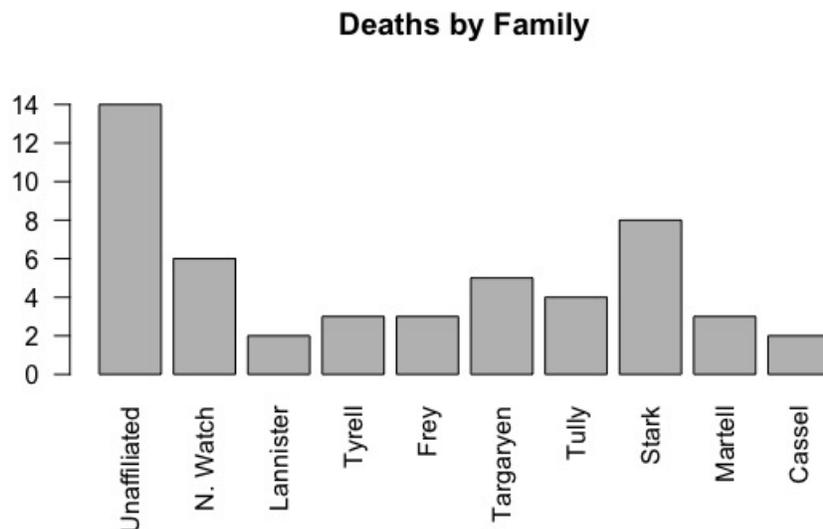
Deaths in Season Premiers versus Season Finales

From our observations, both season premiers and season finales contain a noticeable number of deaths. The query results can be seen in Table 3. We used our query skills to retrieve a table of the appropriate data and completed a t -test for the differences between means of small populations. Based on the data, we found a test statistic, $T = -29.33072$, with $\nu = 9$. Assuming $T \sim t_{\nu-1}$ we found a P -value of $= 9.483255e - 10 \times 10^{-10}$, which means we have statistically

significant evidence at a significance level of 99% that on average, more deaths occur in the season finales than in the season premiers. (The data that went into this calculation can be seen in the appendix, Table 3.)

Deaths By Family

One question of interest we had for this project was to find out which family or group had the most deaths. This could give us more insight for the show as to what family had the best chance of gaining the Iron Throne and ruling the Seven Kingdoms in the end. For this, we had to do a query that summed the number of deaths by affiliation throughout the entire history of the show. You can see the results of this query in Table 2. By the bar chart below, you can see that the most deaths are actually unaffiliated. However, the Stark family has experienced the next most number of deaths. Currently there are only four members of this family remaining alive in the show: Arya, Sansa, Bran, and the bastard Jon Snow. The powerhouse family that is tied for least deaths is the much disliked Lannister clan. They are tied with the the Cassel's but are much more prominent. This may give them an advantage going into the next season as they try to keep the throne.



Note that this graph does not include families or groups for which the number of deaths was less than or equal to one.

TECHNICAL CHALLENGES

The biggest obstacle in our completion of this project was copying the data into tables within the database. Our original COPY command produced the error:

```
ERROR:  must be superuser to COPY to or from a file
HINT:  Anyone can COPY to stdout or from stdin. psql's \copy command also
       works for anyone.
```

which we struggled to edit.

Prior to that we had smaller issues related to our *episode_rating* table. The original data was found in a table on IMDb. Due to our poor skills as Internet scrapers we were forced to copy and paste the data into an Excel file. This led to other issues. For example, the IMDb table identified season and episode in decimal format. Season 1 episode 10 was indicated by 1.10. When copying into Excel, the tool did not differentiate between 1.10 and 1.1 (Season 1 episodes 1 and 10). Thus, we had to create new columns for season and episode. We manually changed each of the episode 10's into the 10's and not the 1's that they showed. Our fix to this problem was not sophisticated, however due to the small size of the table (60 relations) it was sufficient.

Overall, each csv file needed some kind of modification in order to be loaded into the database in the correct format or so that they would contain the same format as another table so that we could effectively query them. Some of the names from *Game of Thrones* were spelled in different languages resulting in symbols that do not match the format the table was being loaded into. Further, in some of the columns when there was no value the term "unspecified" was used rather than leaving it blank. For columns that were not text, this caused an issue. We had to then go through and delete all of those values so that when loaded, the databases would reference those values as NULL.

However, once we were able to upload the tables properly, querying the tables was no issue.

CONCLUSION

Game of Thrones may be just a highly rated, death filled TV series. Through this project we have had the opportunity to answer our own questions regarding the deaths and ratings of the show to see how they might be related. We did this with our database knowledge and statistical skills.

The dataset provides plenty of options for future analysis. We could see how the next two seasons affect this data and if certain trends that we found continue. We also found a lot of data that we did not end up working with. There is potential to do more with the data that we have on alliances, for example, we could see if certain houses have more people associated with them than others. If we included the battle data, we could see if the houses with more affiliations actually win more battles. This data holds a lot of potential and it would be interesting to manipulate it and query it in more ways to draw more statistical conclusions about *Game of Thrones*.

Based on this project, you may be inclined to give this show another shot, or you may be discouraged from ever joining the bandwagon. However, we do hope that this project presented the skills we have learned through CSCI 403 in an entertaining manner.

REFERENCES

- [1] O'Neill, Myles . Game of Thrones | Kaggle. N.p., n.d. Web. 30 Apr. 2017.
<<https://www.kaggle.com/mylesoneill/game-of-thrones>>.
- [2] ""Game of Thrones" (2011)." IMDb. IMDb.com, n.d. Web. 30 Apr. 2017.
<<http://www.imdb.com/title/tt0944947/eprate>>.
- [3] Rininsland, Andrew. "Home." Data.world. N.p., n.d. Web. 30 Apr. 2017.
<<https://data.world/aendrew/game-of-thrones-screen-times>>.
- [4] Rininsland, Andrew. "Home." Data.world. N.p., n.d. Web. 30 Apr. 2017.
<<https://data.world/aendrew/game-of-thrones-deaths>>.

TABLES**Table 1:** Number of deaths per house.

house	count
Unaffiliated	14
Night's Watch	6
House Baratheon of Dragonstone	1
House Lannister	2
House Tyrell	3
House Lannister of Casterly Rock	1
House Thorne	1
Brotherhood without banners	1
House Arryn	1
House Mormont	1
House Frey	3
House Reed	1
House Trant	1
House Targaryen	5
Undying Ones	1
House Selmy	1
House Baratheon of King's Landing	1
House Nymeros Martell	1
House Baratheon	1
House Clegane	1
House Tully	4
House Stark	8
House Bolton	1
Peach	1
House Martell	3
House Cassel	2
House Royce of the Gates of the Moon	1
Thirteen	1
Thenn	1
Mance Rayder	1

Table 2: Deaths in episodes with ratings above 9.5.

season	episode	rating	name
1	9	9.6	Eddard Stark
3	9	9.9	Robb Stark
3	9	9.9	Catelyn Stark
3	9	9.9	Talisa Stark
4	2	9.7	Tansy
4	2	9.7	Joffrey Baratheon
4	8	9.7	The Mountain
4	8	9.7	Oberyn Martell
4	10	9.6	Jojen Reed
4	10	9.6	Tywin Lannister
4	10	9.6	Shae
5	8	9.9	Karsi
6	5	9.7	Three-Eyed Raven
6	5	9.7	Hodor
6	5	9.7	Leaf
6	9	9.9	Belicho Paenymion
6	9	9.9	Jon Umber
6	9	9.9	Ramsay Bolton
6	9	9.9	Razdal mo Eraz
6	9	9.9	Rickon Stark
6	9	9.9	Wun Weg Wun Dar Wun
6	10	9.9	Mace Tyrell
6	10	9.9	Lyanna Stark
6	10	9.9	Lothar Frey
6	10	9.9	Loras Tyrell
6	10	9.9	The High Sparrow
6	10	9.9	Lancel Lannister
6	10	9.9	Kevan Lannister
6	10	9.9	Tommen Baratheon
6	10	9.9	Margaery Tyrell
6	10	9.9	Grand Maester Pycelle
6	10	9.9	Walder Frey
6	10	9.9	Walder Rivers

Table 3: Deaths in season premiers and finales.

season	episode	death count
1	1	2
1	10	3
2	10	5
4	1	1
4	10	3
5	1	1
5	10	6
6	1	3
6	10	12

Table 4: Episodes with no deaths.

season	episode
1	2
1	3
1	4
2	1
2	4
2	8
2	9
3	1
3	2
3	7
3	8
3	10
4	3
4	4
4	6
5	2
5	5
5	6

Table 5: Number of named character deaths per episode (if the episode has deaths).

season	episode	rating	count
1	1	9	2
1	5	9.1	1
1	6	9.2	1
1	7	9.2	2
1	8	9	1
1	9	9.6	1
1	10	9.4	3
2	2	8.5	1
2	3	8.8	1
2	5	8.8	1
2	6	9	1
2	7	8.9	1
2	10	9.4	5
3	3	8.9	1
3	4	9.5	3
3	5	9	1
3	6	8.8	1
3	9	9.9	3
4	1	9.1	1
4	2	9.7	2
4	5	8.7	3
4	7	9.1	1
4	8	9.7	2
4	9	9.5	5
4	10	9.6	3
5	1	8.5	1
5	3	8.6	1
5	4	8.8	1
5	7	9	1
5	8	9.9	1
5	9	9.4	2
5	10	8.9	6
6	1	8.5	3
6	2	9.5	5
6	3	8.7	3
6	4	9.2	2
6	5	9.7	3
6	6	8.4	1
6	7	8.6	1
6	8	8.3	4
6	9	9.9	6
6	10	9.9	12

Table 6: Number of deaths per season

<u>season</u>	<u>count</u>
1	11
2	10
3	9
4	17
5	13
6	40