Game of Thrones: IMDb Scraping and Analysis

Patricia Tang

6/10/2019

Introduction

Game of Thrones was one of the most popular television shows of all time. And, given the recent dropoff in quality of the show's eighth and final season, many individuals are attempting to determine what exactly makes a good Game of Thrones episode so compelling and what makes a bad Game of Thrones episode so revolting. In order to examine this, I used IMDb's database of Game of Thrones episodes and their respective content keywords, synopsis information, and creator information to answer the following question: what parameters affect user rating and sentiment for a Game of Thrones episode, and what factors go into the making of a highly-rated episode?

More concretely, I first use frequency analysis to examine the changing state of Game of Thrones over time as well as trends in content and authorship. Next, I use t-tests, traditional linear regression, regression discontinuity design, and traditional multiple regression to explore how general content, plot details, authorship, and the availability of book material to adapt to screen affect the rating of a Game of Thrones episode, as well as to determine which of these factors is most important in the making of a good Game of Thrones episode.

Scraping the Data

This section contains code for scraping the IMDb website for Game of Thrones, including clicking from the main directory into individual episode pages and keyword pages. This was done to generate information for analysis.

Included below is code for scraping general information about Game of Thrones episodes from IMDb, the Internet Movie Database. This includes: season, episode title, episode number (within the season, not in general), a text description of each episode (kind of like a synopsis), rating score (user-generated based on IMDb users), episode airdate, and pre-link info (for later).

```
get_info <- function(x){
  page <- read_html(paste0("https://www.imdb.com/title/tt0944947/episodes?season=",x))
  season <- x
  ep_title <- page %>% html_nodes("#episodes_content strong a") %>%
    html_text()
  ep_num <- page %>%
   html_nodes("#episodes_content strong a") %>%
    html_attr("href") %>%
    gsub(".*ep", "", .)
  text <- page %>%
    html_nodes(".item_description") %>%
    html_text() %>%
  gsub("\\n ", "", .)
  rating <- page %>%
    html_nodes(".ipl-rating-star.small .ipl-rating-star__rating") %>%
    html_text() %>% as.numeric()
  airdate <- page %>% html_nodes(".airdate") %>%
```

```
    html_text() %>%
    dmy()
  links <- page %>%
   html_nodes("#episodes_content strong a") %>%
    html_attr('href')
  return(data.frame(season,
  ep_title,
  ep_num,
  text,
  rating,
  airdate,
  links,
  stringsAsFactors = F))
}

Thrones <- lapply(1:8, get_info) %>% bind_rows()
```

Included below is code for filling out links to episode pages via the aforementioned pre-link info.

```
# to get full links
for(i in 1:73){
  Thrones$links[i] <- paste("https://www.imdb.com", Thrones$links[i], sep ='')
}
```

Next, I created a new set of links per page in order to click through to the separate keyword pages per episode. The following code was created after analyzing IMDb's URL formats for these pages.

```
# to get href links for keyword scraping
Thrones$keywordlinks <- substr(Thrones$links, start=1, stop=37)
```

```
# to make full links for all keyword links
for(i in 1:73){
  Thrones$keywordlinks[i] <- paste(Thrones$keywordlinks[i],
  "keywords?ref_=tt_stry_kw",
  sep ='')
  # "keywords?ref_=tt_stry_kw" returns keyword page for ALL IMDb pages here
}
```

Included below is code scraping keyword data.

```
# to get keyword data
for(i in 1:73){
  link = as.character(Thrones$keywordlinks[i])
  Thrones$keywords[i] <- read_html(link) %>%
  html_nodes("div.sodatext") %>%
    html_text %>%
    paste(collapse = "\n\n")
}
```

I then create new links in order to find a new page on the IMDb site per episode for casting information; this code was created after analyzing the URLs of these cast pages per episode.

```
# to get directors/writing credits
Thrones$creditlinks <- substr(Thrones$links, start=1, stop=37)
for(i in 1:73){
  Thrones$creditlinks[i] <- paste(Thrones$creditlinks[i],
```

```
  "fullcredits?ref_=tt_ov_wr",
  sep ='')
  # "fullcredits?ref_=tt_ov_wr" returns keyword page for ALL IMDb pages here
}
```

Included below is code scraping for directorial credits. Since the teams on Game of Thrones stay relatively static with two showrunners consistently heading writing and production teams staying fairly stable, directors can be seen as varying sources of authorship per episode.

```
# to get directorial credit.
for(i in 1:73){
  link = as.character(Thrones$creditlinks[i])
  Thrones$director[i] <- read_html(link) %>%
    html_nodes(".simpleCreditsTable:nth-child(2)") %>%
    html_text %>% paste(collapse = "\n\n")
}
```

Prepping for Text Analysis

Here, I prep DTMs for keyword and text documents. This is used later in frequency analysis.

IMDb lists keywords one-by-one in a list without grammar or punctuation, so there is no need here to do additional preparations such as removing punctuation.

```
# DTM for keywords
ThronesCorpus <- VCorpus(VectorSource(Thrones$keywords)) %>%
  tm_map(removeWords, stopwords("english"))
KeywordDTM = DocumentTermMatrix(ThronesCorpus)
```

```
KeywordDTMclean = KeywordDTM %>% tidy()
```

The DTM for synopsis text is prepared differently from that of the keywords due to the complete sentences that consist of synopsis text.

```
# DTM for text
TextCorpus <- VCorpus(VectorSource(Thrones$text)) %>%
  tm_map(removePunctuation) %>% tm_map(tolower) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(PlainTextDocument)
TextDTM = DocumentTermMatrix(TextCorpus)
TextDTM_copy = TextDTM %>% removeSparseTerms(.99)
TextDTMclean = TextDTM_copy %>% tidy()
tfidf <- TextDTMclean %>%bind_tf_idf(term,document, count)
```

Frequency Analysis

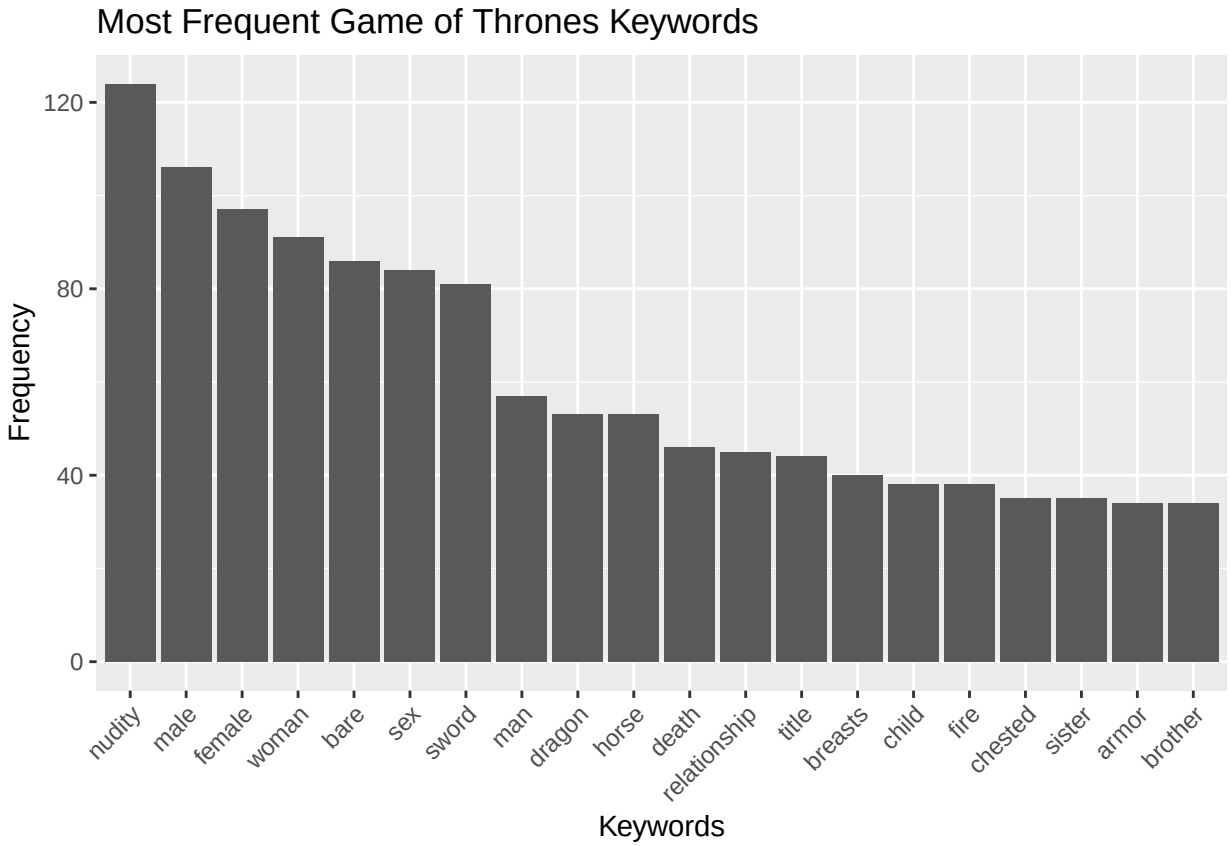
In order to characterize the IMDb dataset, I used frequency analysis to determine key trends and characteristics of Game of Thrones in regards to content, authorship, and rating over time.

Frequency Analysis: Keywords

Below, I analyze the most frequent-occuring Game of Thrones keywords (from IMDb's list of keywords per episode, NOT synopsis information) in order to derive what might characterize Game of Thrones as a TV show.

```
# Keywords
Keyword_Matrix <-as.matrix(KeywordDTM)
```

```r
freq <- sort(colSums(Keyword_Matrix), decreasing = TRUE)
freqdf <-as.data.frame(freq)
top <- head(freq, 20)


top_df <- data.frame(word = names(top), freq = top)
Keywordfreq <- ggplot(data=top_df, aes(x=reorder(word, -freq), y=freq)) +
  geom_bar(stat="identity") +
  xlab("Keywords") +
  ylab("Frequency") +
  theme(legend.title = element_text(size = 3),
  legend.text = element_text(size = 0.01),
  axis.text.x=element_text(angle=45,hjust=1,vjust=1)) +
  ggtitle("Most Frequent Game of Thrones Keywords")
Keywordfreq
```



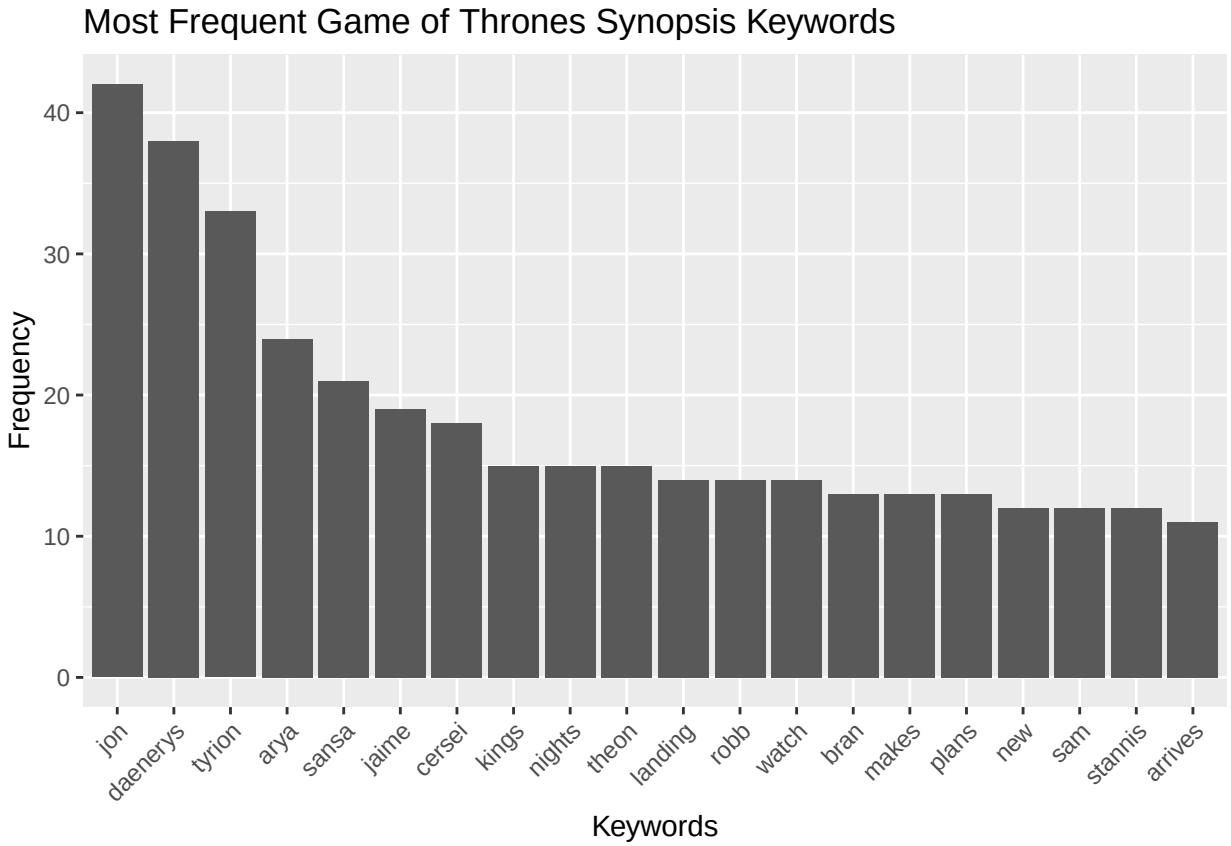
Accordingly, "nudity" is the most frequent keyword. For a TV show that is known for what is commonly known as "sexposition," where characters explain plot details while surrounded by gratuitious nudity, this is expected; however, one might be surprised at the sheer amount of nudity on the show as demonstrated by the frequency of "nudity" as a keyword. "Sex" is also a close sixth, corroborating this idea that Game of Thrones is known for "sexposition."

"Dragon," "horse" and "death" also rank fairly high in terms of frequency. This makes sense: in Game of Thrones, there are a lot of horses. Additionally, Game of Thrones is known for dragons. And, given that Danearys is a main character and is known for her interactions with both symbolic and real dragons, it makes sense that dragons seem to appear quite a lot in Game of Thrones, whether or not they actually appear actually on screen.

Frequency Analysis: Synopsis

Next, I analyze keywords from the synopsis information from each episode, scraped as the "text" field. This information was pre-processed so that information would be extractable. The importance of synopsis information differs from keyword data in that specific character appearances and storylines can be tracked. As such, the frequency of appearance of character names in the frequency analysis here denotes their relative importance.

```
# Synopsis
TextMatrix <-as.matrix(TextDTM)
freq2 <- sort(colSums(TextMatrix), decreasing = TRUE)
freqdf2 <-as.data.frame(freq2)
top2 <- head(freq2, 20)


top_df2 <- data.frame(word = names(top2), freq = top2)
Textfreq <- ggplot(data=top_df2, aes(x=reorder(word, -freq), y=freq)) +
  geom_bar(stat="identity") +
  xlab("Keywords") +
  ylab("Frequency") +
  theme(legend.title = element_text(size = 3),
  legend.text = element_text(size = 0.01),
  axis.text.x=element_text(angle=45,hjust=1,vjust=1)) +
  ggtitle("Most Frequent Game of Thrones Synopsis Keywords")
Textfreq
```



As shown above, character names are at the top of the frequency analysis. This analysis is fairly insightful in regards to determing what the show thinks are the most important characters since if a certain character is more important, their plotline would be tracked within the synopsis information more often. Accordingly, Jon and Dany are the two most frequent occurrences, signifying that the show emphasizes them most. This makes sense, since they seem to be the two main characters of the Game of Thrones series.

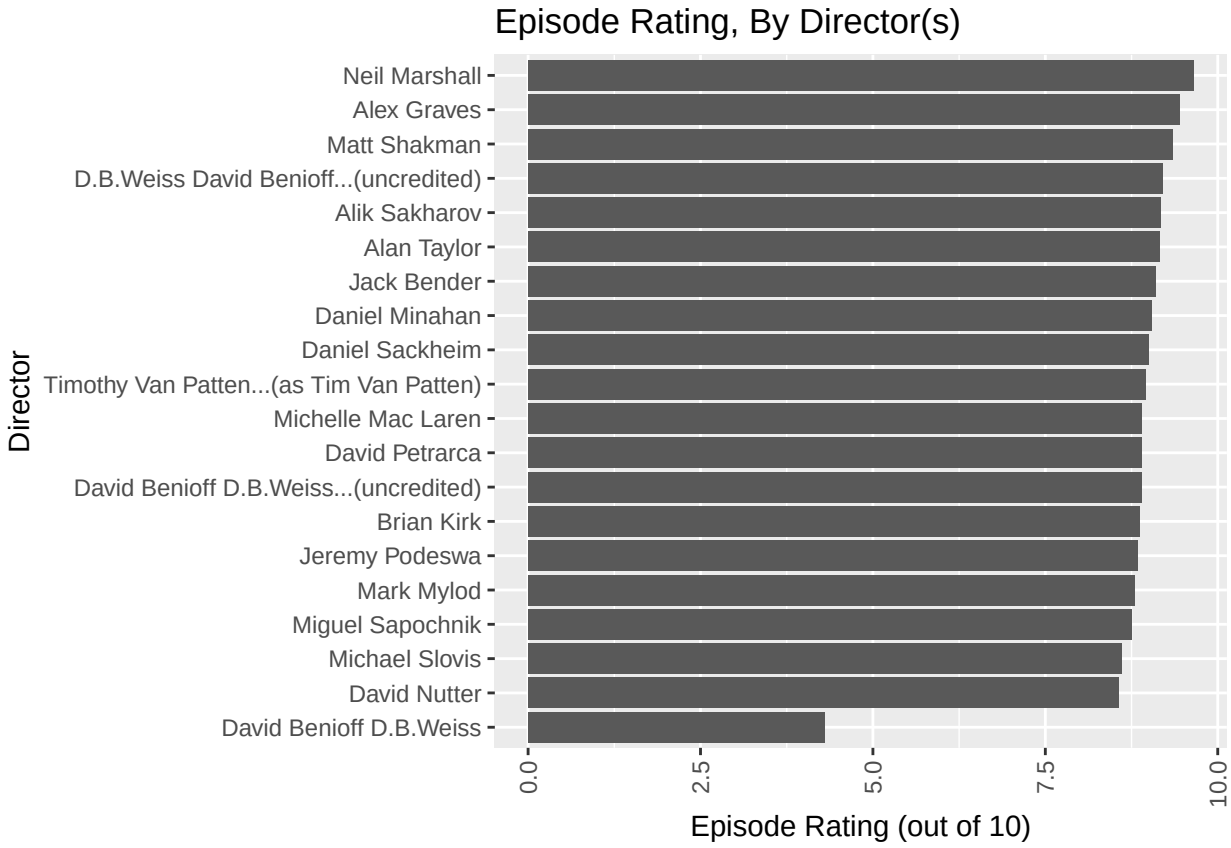
Interestingly, the top results in this analysis all happen to be characters who live all the way up to Season 8; some even live all the way to the end of it. Thus, the results here might be biased since other main characters that the show particularly focused on (notably, Robb) died seasons ago and therefore would not be mentioned.

Directors: Disparities in Authorship

Next, I examine the difference in ratings between directors to assess the difference in authorship quality between different directors.

```
# Ratings by director
perdirect <- ddply(Thrones, .(director), summarize,  rating=mean(rating))
perdirect$director <- sapply(perdirect$director,
function(x) { gsub("\n", "", x) })
perdirect$director <- sapply(perdirect$director,
function(x) { gsub("[\r\n]", "", x) })
perdirect$director <- sapply(perdirect$director,
function(x) { gsub(" ", "", x) })
perdirect$director <- sapply(perdirect$director,
function(x) {gsub("([[:lower:]])(?=[[:upper:]])", "\\1 ", x, perl = TRUE)})
```

```
Directrat <- ggplot(data=perdirect, aes(x=reorder(director, rating), y=rating)) +
  geom_bar(stat="identity") +
  xlab("Director") +
  ylab("Episode Rating (out of 10)") +
  theme(legend.title = element_text(size = 3),
  legend.text = element_text(size = 0.1),
  axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  ggtitle("Episode Rating, By Director(s)") + coord_flip()
Directrat
```



It is important to note that showrunners David Benioff and D.B. Weiss have three different ways they have been credited. This is actually fairly significant: in the entries where (uncredited) is mentioned, Benioff and Weiss were not directing alone. However, the entry that has their names only indicates the ratings of episodes they directed alone.

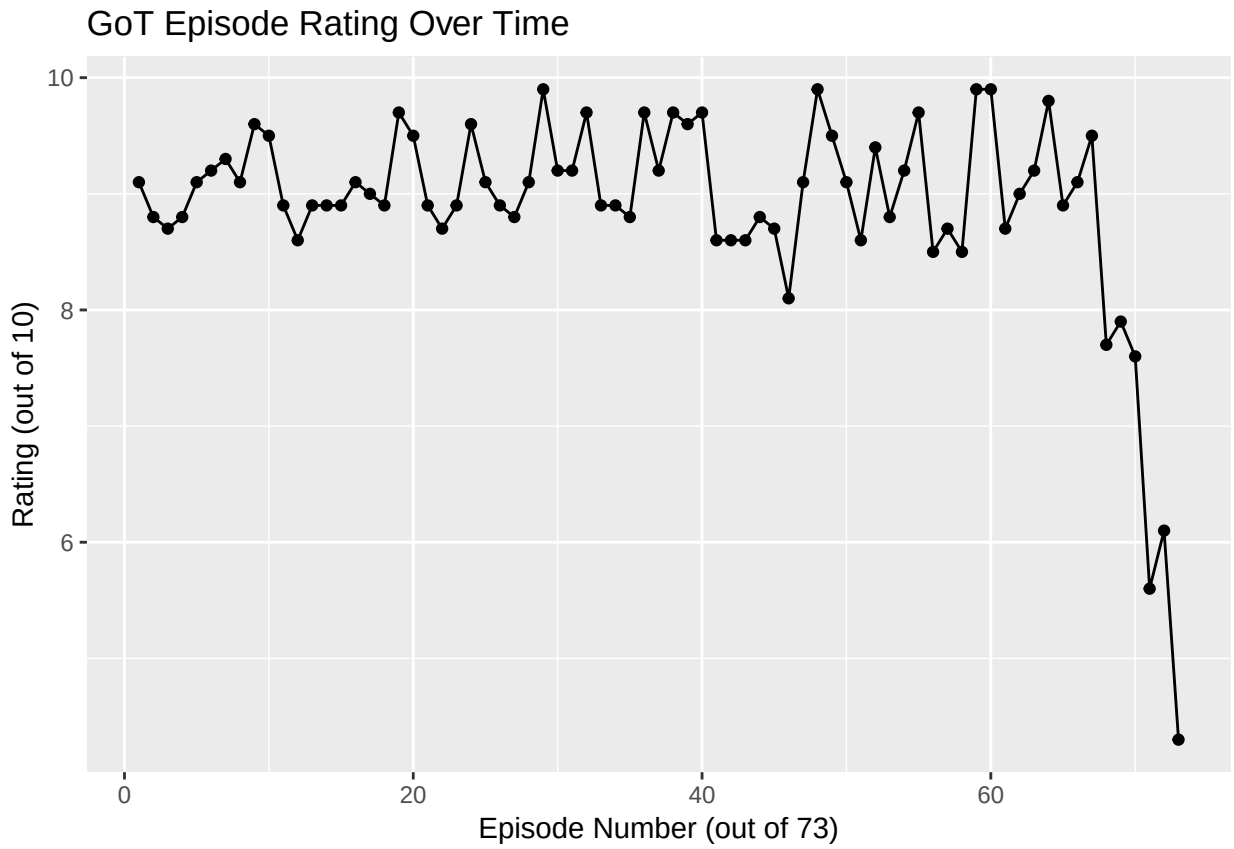
Clearly, Benioff and Weiss' individually-directed episodes have much lower ratings than the episodes where they directed with other uncredited directorial personnel. From this, it can be extracted that Benioff and Weiss' involvement seems to have a detrimental effect on episode quality. I come back to this later to assess the statistical significance of this using a regression model.

Neil Marshall, Alex Graves, and Matt Shakman clearly directed the highest-rated episodes. I also come back to this later and determine if there is a statistically significant difference in regards to their authorship through t-testing.

Episode Quality Over Time: A Significant Drop?

Next, I track the episode quality (indicated by IMDb user rating) over time.

```
Thrones$epnumeral <- c(1:73)

Bestepisodes <- ggplot(data=Thrones, aes(x=epnumeral, y=rating)) +

  geom_line() +

  geom_point() +

  ggtitle("GoT Episode Rating Over Time") +

  xlab("Episode Number (out of 73)") +

  ylab("Rating (out of 10)") +

  theme(legend.title = element_text(size = 3),

  legend.text = element_text(size = 0.01))

Bestepisodes
```



Note the drop in rating for the end of the series; this evidences fans' disappointment with the ending of Game of Thrones. The precipitious fall of user ratings indicates the episodes in Game of Thrones' Season 8, indicating a stark disparity in quality between Season 8 of the series and the rest of the show.

However, showrunners Benioff and Weiss deviated from the books much earlier than Season 8; I examine the effect of this deviation later in a regression discontinuity model.

Building The "Character Score"

Given that synopsis information is not representative of which characters show up in a given episode since synopsis information often leaves much out, it seems invalid to run t-tests per character to determine the statistical significance of their appearances. Thus, in order to determine whether or not character appearances are correlated with episode rating, I create an aggregate variable that I call a character "score" indicating how many of these popular characters are present in a given episode's synopsis. While this might at first be apparent as an indicator for fan investment in a given episode, it might also serve as an indicator of quality; if a given episode features a number of characters, it might also involve more interconnected storylines and more character-heavy storylines involving political intrigue and interactions.

```
Thrones$jon <- ifelse(str_detect(Thrones$text, "Jon") == "TRUE", "1", "0")
Thrones$dany <- ifelse(str_detect(Thrones$text, "Daenerys") == "TRUE", "1", "0")
Thrones$tyrion <- ifelse(str_detect(Thrones$text, "Tyrion") == "TRUE", "1", "0")
Thrones$arya <- ifelse(str_detect(Thrones$text, "Arya") == "TRUE", "1", "0")
Thrones$sansa <- ifelse(str_detect(Thrones$text, "Sansa") == "TRUE", "1", "0")
Thrones$jaime <- ifelse(str_detect(Thrones$text, "Jaime") == "TRUE", "1", "0")
Thrones$cersei <- ifelse(str_detect(Thrones$text, "Cersei") == "TRUE", "1", "0")
Thrones$theon <- ifelse(str_detect(Thrones$text, "Theon") == "TRUE", "1", "0")
```

```
Thrones$robb <- ifelse(str_detect(Thrones$text, "Robb") == "TRUE", "1", "0")
Thrones$bran <- ifelse(str_detect(Thrones$text, "Bran") == "TRUE", "1", "0")
Thrones$sam <- ifelse(str_detect(Thrones$text, "Sam") == "TRUE", "1", "0")
Thrones$stannis <- ifelse(str_detect(Thrones$text, "Stannis") == "TRUE", "1", "0")
```

```
Thrones$jon <- as.numeric(Thrones$jon)
Thrones$dany <- as.numeric(Thrones$dany)
Thrones$tyrion <- as.numeric(Thrones$tyrion)
Thrones$arya <- as.numeric(Thrones$arya)
Thrones$sansa <- as.numeric(Thrones$sansa)
Thrones$jaime <- as.numeric(Thrones$jaime)
Thrones$cersei <- as.numeric(Thrones$cersei)
Thrones$theon <- as.numeric(Thrones$theon)
Thrones$robb <- as.numeric(Thrones$robb)
Thrones$bran <- as.numeric(Thrones$bran)
Thrones$sam <- as.numeric(Thrones$sam)
Thrones$stannis <- as.numeric(Thrones$stannis)
Thrones$charascore <- Thrones$jon + Thrones$dany + Thrones$tyrion +
  Thrones$arya +  Thrones$sansa + Thrones$jaime +  Thrones$cersei +
  Thrones$theon + Thrones$robb +  Thrones$bran + Thrones$sam +
  Thrones$stannis
```

I then use this character score as an indicator variable for the number of major characters in a given episode; will use this within regression analysis.

Authorial Studies

Next, I examine using t-tests whether or not there is a statistically significant difference in terms of rating if a top-rated director did or didn't direct an episode.

Here, I assume for simplicity that the "top-rated" directors are the top 3 directors by episode rating: Neil Marshall, Alex Graves, and Matt Shakman.

Note for t-tests: I assume a p-value significance level of 0.05.

Neil Marshall

```
Thrones$neilm <- ifelse(str_detect(Thrones$director, "Neil Marshall") == "TRUE",
"1",
"0")
t.test(Thrones$rating~Thrones$neilm)
```

```
##
##  Welch Two Sample t-test
##
## data:  Thrones$rating by Thrones$neilm
## t = -6.5049, df = 24.211, p-value = 9.564e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.0156791 -0.5265744
## sample estimates:
## mean in group 0 mean in group 1
##        8.878873        9.650000
```

We see a statistically significant difference here; thus, we use IMDb information to

delve deeper into which episodes Marshall directed in order to determine the source of this correlation. As it turns out, Marshall only directed two episodes: "Blackwater" (The fabled Battle of the Blackwater) and "The Watchers on the Wall" (Wildling assault on Castle Black and Ygritte's death). These are both storyline-defining and iconic episodes in Game of Thrones; they are also episodes that focus solely on the battles within them. Thus, Marshall directs episodes with major "key moments" and spectacles in them and without much complexity, potentially contributing to this correlation.

Alex Graves

```
Thrones$alexg <- ifelse(str_detect(Thrones$director, "Alex Graves") == "TRUE",
"1",
"0")
t.test(Thrones$rating~Thrones$alexg)
```

```
##
##  Welch Two Sample t-test
##
## data:  Thrones$rating by Thrones$alexg
## t = -3.2584, df = 12.431, p-value = 0.006565
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9984276 -0.2000798
## sample estimates:
## mean in group 0 mean in group 1
##        8.850746        9.450000
```

We see a statistically significant difference here; thus, we use IMDb information to

delve deeper into which episodes Graves directed in order to determine the source of this correlation. IMDb states that Graves directed key storyline-defining and iconic episodes in Game of Thrones, such as "And Now His Watch Has Ended" (Dany purchases the Unsullied) and "The Lion and the Rose" (Joffrey's death). Thus, Graves directs episodes with major "key moments" in them, potentially contributing to this correlation.

Matt Shakman

```
Thrones$matts <- ifelse(str_detect(Thrones$director, "Matt Shakman") == "TRUE",
"1",
"0")
t.test(Thrones$rating~Thrones$matts)
```

```
##
##  Welch Two Sample t-test
##
## data:  Thrones$rating by Thrones$matts
## t = -0.99988, df = 1.118, p-value = 0.4856
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.060569  4.135217
## sample estimates:
## mean in group 0 mean in group 1
##        8.887324        9.350000
```

No statistically significant difference.

Regression Analysis

Lastly, I use regression analysis to determine how general content, plot details, authorship, and the availability of book material to adapt to screen affect the rating of a Game of Thrones episode, as well as to determine which of these factors is most important in the making of a good Game of Thrones episode.

Simple Linear Regression: Character Score

First, I run a simple linear regression on character score (as defined previously) and episode rating.

Note: for this section, I assume a significance level of 0.05.

```
Charalinear <- lm(Thrones$rating ~ Thrones$charascore, data = Thrones)
summary(Charalinear)
```

```
##
## Call:
## lm(formula = Thrones$rating ~ Thrones$charascore, data = Thrones)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0647 -0.3753  0.0283  0.4247  1.4389
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         8.16113    0.23922   34.12  < 2e-16 ***
## Thrones$charascore  0.20354    0.06004    3.39  0.00114 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8427 on 71 degrees of freedom
## Multiple R-squared:  0.1393, Adjusted R-squared:  0.1272
## F-statistic: 11.49 on 1 and 71 DF,  p-value: 0.001144
```

There is a statistically significant relationship between character score (as defined earlier) and rating. Thus, while individual characters did not seem to affect rating (as per chi-squared test), rating seems to be positively correlated with the number of characters highlighted in a given episode.

Regression Discontinuity: Rating By Availability of Book Material

In order to determine whether or not the availability of book material changed episode ratings and thus quality, I used a regresion discontinuity model. Using a 2015 article from Forbes (https://www.forbes.com/sites/insertcoin/2015/04/12/why-game-of-thrones-finally-outrunning-the-books-is-a-good-thing/#6bfc36c7524b) as a source, I defined the cutoff point to be the break between Seasons 4 and 5.

```
reg.1=RDestimate(rating~epnumeral,data=Thrones,cutpoint = 41)
```

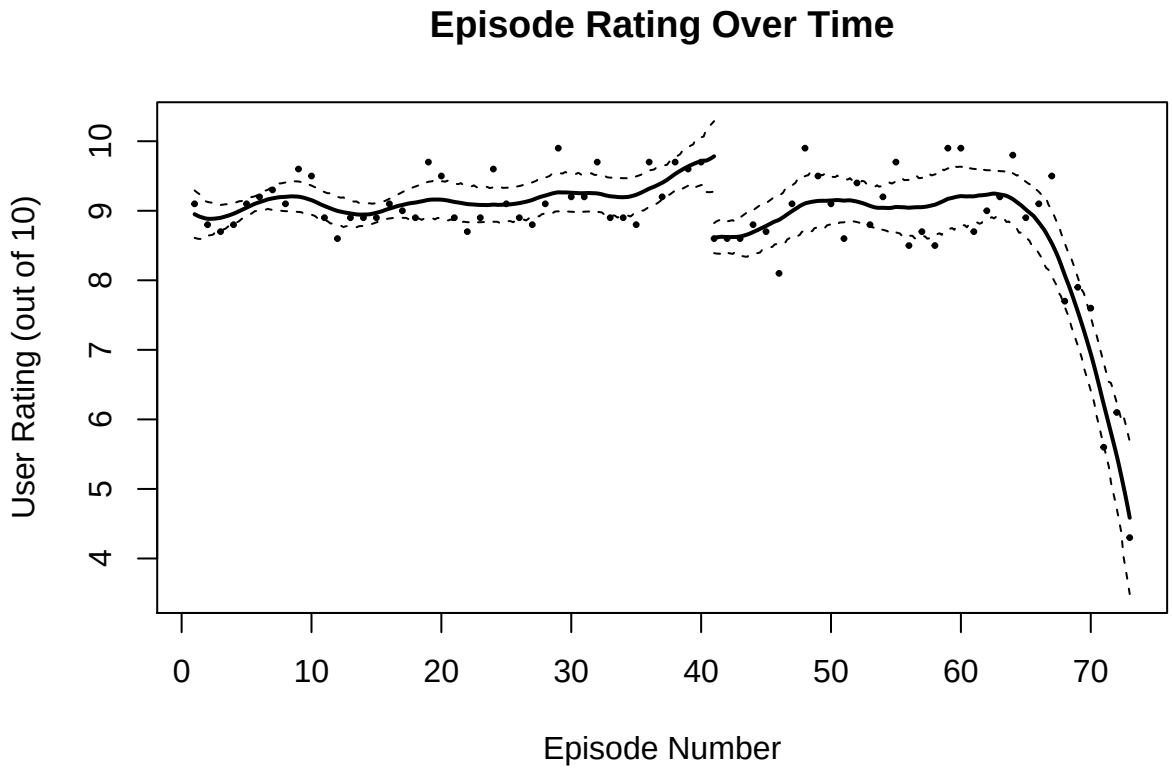
```
## Warning in summary.lm(x): essentially perfect fit: summary may be
## unreliable
```

```
plot(reg.1)
```

```
## Warning in summary.lm(x): essentially perfect fit: summary may be
## unreliable
```

```
title(main="Episode Rating Over Time",
xlab="Episode Number",
```

```
ylab="User Rating (out of 10)")
```




```
summary(reg.1)
```

```
##
## Call:
## RDestimate(formula = rating ~ epnumeral, data = Thrones, cutpoint = 41)
##
## Type:
## sharp

## Warning in summary.lm(x): essentially perfect fit: summary may be
## unreliable

## Estimates:
##            Bandwidth  Observations  Estimate  Std. Error  z value
## LATE        5.396     11            -1.176    0.1006      -11.69
```

```
## Half-BW     2.698      5            -1.200    0.0000        -Inf
## Double-BW  10.791     21            -1.303    0.1243      -10.48
##            Pr(>|z|)
## LATE       1.457e-31  ***
## Half-BW    0.000e+00  ***
## Double-BW  1.071e-25  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable
## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable

## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable

## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable

## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable

## Warning in summary.lm(object$model[[i]]): essentially perfect fit: summary
## may be unreliable
## F-statistics:
##            F          Num. DoF  Denom. DoF  p
```

```
## LATE       4.198e+01  3          7          1.527e-04
## Half-BW    2.169e+30  3          1          8.882e-16
## Double-BW  9.491e+00  3         17          1.306e-03
```

By this, we see a statistically significant difference in mean ratings, where episodes after Season 4 on average had lower mean ratings. Thus, episode rating is correlated with the availability of book material.

Multiple Regression: What Is The Biggest Factor?

Next, I use multiple regression to answer the following question: what is the biggest factor in terms of episode quality - character score (how many characters that people like are in an episode), whether or not an episode was directed by an outside talent (not the writers of the show themselves), or whether or not book material was still there to guide the writing team (as defined by article)? "'

```
# To determine involvement of directors.
# Since they are credited together usually, we can use one of the two
Thrones$DandD <- ifelse(str_detect(Thrones$director, "D.B. Weiss") == "TRUE", "1", "0")
Thrones$DandD <- as.numeric(Thrones$DandD)
```

```
# Creating a dummy variable
# all episodes after season 4 finale will be counted as "departed from the books"
for(i in 1:73){
  Thrones$depart <- ifelse(Thrones$epnumeral>41, "1", "0")
}
```

```
# Linear model
Thronesfactors <- lm(rating ~ depart + charascore + DandD, data = Thrones)
summary(Thronesfactors)
```

```
##
## Call:
## lm(formula = rating ~ depart + charascore + DandD, data = Thrones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.17558 -0.40910 -0.03704  0.46296  1.42025
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.40769    0.20819  40.385  < 2e-16 ***
## depart1     -0.75729    0.17356  -4.363 4.39e-05 ***
## charascore   0.24312    0.05254   4.628 1.68e-05 ***
## DandD       -1.41794    0.42327  -3.350  0.00131 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7154 on 69 degrees of freedom
## Multiple R-squared:  0.3972, Adjusted R-squared:  0.371
## F-statistic: 15.15 on 3 and 69 DF,  p-value: 1.124e-07
```

All factors here are statistically significant here but, clearly, the directorial involvement of Benioff and Weiss is the greatest factor in terms of rating as determined by the magnitude of coefficient in this model. Thus, there is a correlation between Benioff and Weiss making artistic choices about a Game of Thrones episode and the subsequent quality of said episode.

Conclusion

Through examining content, plot details, authorship, and the availability of source material, I determine that all factors here are more or less correlated with episode rating to some extent, and that the greatest factor in terms of episode rating was the directorial involvement of Benioff and Weiss is the greatest factor in terms of rating as determined by the magnitude of coefficient in this model. This lends credence to recent criticism of Benioff and Weiss' artistic merit, and shows that increases in their involvement within Game of Thrones is correlated with decreases in episode quality.