

# Fifty Shades of Grey

*"How could a book with so much bad attention,  
like Fifty Shades of grey, become a bestseller?"*



Course: From Objects to Data  
Submission Date: 23 – 10 – 2015  
Teacher: Marijn Koolen  
Students:  
Cindy Looijen  
Halil Ibrahim Kara  
Tijmen van Willigen  
Tobias Bekendam  
Loraine Hoogendoorn

# Research Question

For this report we have decided to use the following research question: *“How could a book with a lot of bad attention, like Fifty Shades of grey, become a bestseller?”*.

To answer this question we came up with the following sub questions:

- Is the overall opinion of the book indeed negative or is this just an assumption made?
- What is the general motivation of the negative/positive reviews?
- What is the motivation for people to start reading this book?

# Approach

In finding an answer to the first sub question, we wanted to look at three things. First of all, we had a look at the proportion of positive and negative votes using unix scripts (number 1). Secondly, we looked at the usefulness of the reviews, also with unix scripts (numbers 2 & 3). With these results we wanted to check the findings of the distribution of the reviews. Besides that, we wanted to know whether we could speak of an instant negative hype or not. For this we used unix commands (numbers 4,5 & 6) and looked at the ratings just after publishment. With the answers to these questions we could form a general idea about the overall opinion of the book.

To find out what the general motivation of people is to either like or dislike the book, we divided all the reviews in two parts using unix scripts (numbers 7 & 8). Now we have one file with all the ratings lower than 3, a second file with all the ratings higher than 3. By counting the words in both groups of reviews with a unix script (number 9), some frequently reoccurring words could be identified per file. We also visualized these words in a wordcloud, using the tool <http://tagcrowd.com/>. Next, we have created a word tree, using the tool <https://www.jasondavies.com/wordtree/?source=5ede12648b2941c476779cb549d4704d> for the positive wordcloud and <https://www.jasondavies.com/wordtree/?source=5943308abfde281c4d31a56bc9bc3861> for the negative wordcloud. These tools helped us to close read on the words that were most frequent in both files, since these tools showed us the most frequent word combinations made with these words. Now we had a good idea of what people were actually saying about the book.

For the question why people actually started reading this book, we will look at the contexts of motivational sentences like "*i started reading this book...*" or "*... decided to read ...*". We will look for these sentences/words and their surrounding words using unix commands (numbers 10-15). By reading these motivations and their surrounding words we got some insight in the general motivation of people to read this book. This gave us new search terms to look for using unix commands (numbers 16-21). With these results we got a better insight in the general motivation of people to read this book.

With all of this information combined we could eventually formulate a conclusion to our main question.

## Analysis and interpretation of our results

### Is the overall opinion of the book indeed negative or is this just an assumption made?

First we wanted to find out whether the overall opinion about the book was really negative or whether this was just an assumption most people made. With the results from script 1 we could conclude that most people rate the book a 1.0 and therefore we could assume that the overall opinion about the book is really negative. See figure 1 for the overall rating behaviour.

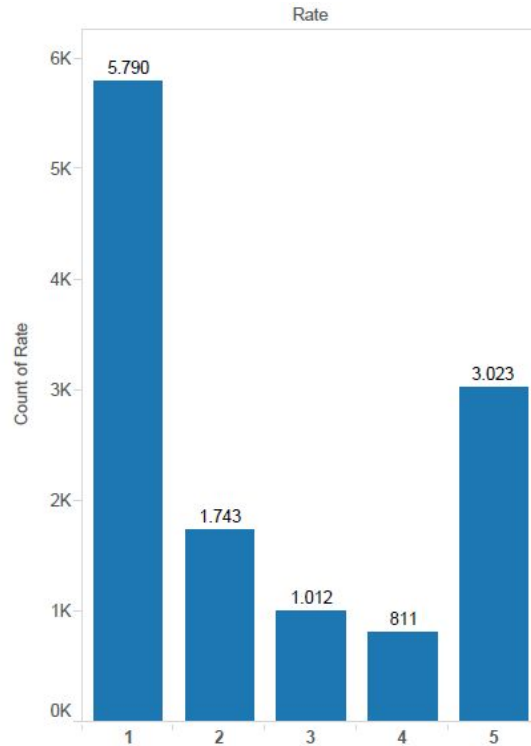


Figure 1 - Rating behaviour (made using script 1)

By looking at the usefulness of the reviews we got the following results, using script number 3.

|     |       |
|-----|-------|
| 1.0 | 1013  |
| 1.0 | 1239  |
| 1.0 | 15906 |
| 1.0 | 1698  |
| 1.0 | 208   |
| 1.0 | 213   |
| 1.0 | 219   |
| 1.0 | 280   |
| 1.0 | 290   |
| 1.0 | 311   |
| 1.0 | 317   |
| 1.0 | 3211  |
| 1.0 | 336   |
| 1.0 | 4265  |

|     |       |
|-----|-------|
| 1.0 | 441   |
| 1.0 | 446   |
| 1.0 | 490   |
| 1.0 | 503   |
| 1.0 | 574   |
| 1.0 | 6083  |
| 1.0 | 653   |
| 1.0 | 720   |
| 1.0 | 789   |
| 1.0 | 930   |
| 1.0 | 969   |
| 2.0 | 23311 |
| 3.0 | 10755 |
| 3.0 | 1550  |
| 4.0 | 707   |
| 5.0 | 323   |

From this list we could conclude that the reviews with a 4.0 and a 5.0 rating are not considered to be that useful. 12 305 people thought a review with a rating of 3.0 was a useful review, which would be the people with a neutral opinion about the book. 23.311 people voted a review with a rating of 2.0 as a useful review and 42.104 people voted a review with a rating of 1.0 as a useful review. This shows us that most of the votes for a useful review are given to reviews with a rating of 1.0 and 2.0. This confirms our findings to the above question. The review which got the most votes as a useful review is a review with a rating of 2.0, after that a review with a 1.0 got the most votes, and third in line is a review with a 3.0 rating. This information also confirms the findings above.

Besides that, we found out, using scripts 4,5 and 6, that in the first months after publishment of the book already more negative reviews were given than positive reviews (see figure 2). Vintage Books acquired the publishing rights of the book in march 2012. The ratings that were given in this month are:

|     |         |     |
|-----|---------|-----|
| 103 | 2012-03 | 1.0 |
| 36  | 2012-03 | 2.0 |
| 13  | 2012-03 | 3.0 |
| 2   | 2012-03 | 4.0 |
| 3   | 2012-03 | 5.0 |

In the following months the book was given the following ratings:

|     |         |     |
|-----|---------|-----|
| 417 | 2012-04 | 1.0 |
| 108 | 2012-04 | 2.0 |
| 50  | 2012-04 | 3.0 |
| 60  | 2012-04 | 4.0 |
| 233 | 2012-04 | 5.0 |
| 794 | 2012-05 | 1.0 |
| 210 | 2012-05 | 2.0 |
| 104 | 2012-05 | 3.0 |
| 122 | 2012-05 | 4.0 |
| 451 | 2012-05 | 5.0 |
| 768 | 2012-06 | 1.0 |
| 194 | 2012-06 | 2.0 |
| 91  | 2012-06 | 3.0 |
| 78  | 2012-06 | 4.0 |

|     |         |     |
|-----|---------|-----|
| 393 | 2012-06 | 5.0 |
| 795 | 2012-07 | 1.0 |
| 183 | 2012-07 | 2.0 |
| 64  | 2012-07 | 3.0 |
| 51  | 2012-07 | 4.0 |
| 221 | 2012-07 | 5.0 |

From this list we can conclude that in the following months the book got more positive reviews. So in these months there are people who thought the book had something good in it. However, the negative rating of a 1.0 remains the rating that is given the most. Therefore, we could not say that the book has been more popular shortly after release. When using script number 6 we found out that most reviews were written in May, June and July 2012.

Looking at these results we can therefore conclude there has been an instant negative hype about the book.

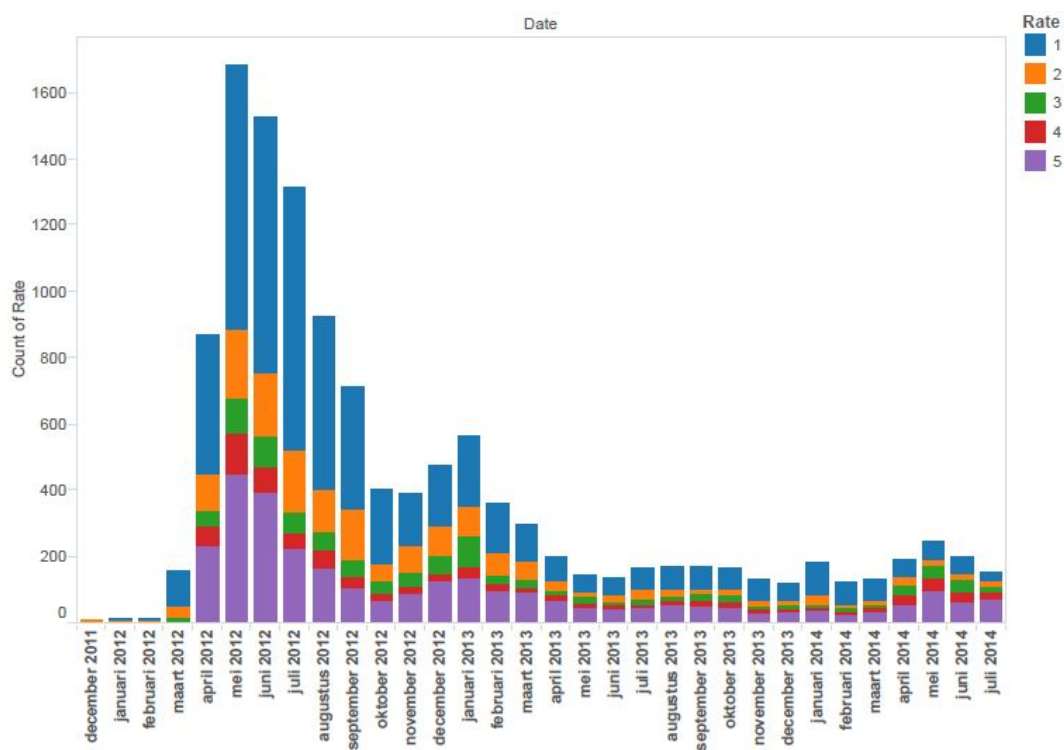


Figure 2. Rating per month (made using script 5)

### What is the general motivation of the negative/positive reviews?

With the tool <http://tagcrowd.com/> we have created a wordcloud of the file with the ratings 4.0 and 5.0 (figure 3) and a wordcloud of the file with the ratings 1.0 and 2.0 (figure 5). In figure 3 you can see that the words 'Love' and 'Story' are used very often. By looking at the context of these words, using a word tree (figure 4) and other means of close reading, we concluded that most people like the story of the book and especially the love story. People often mention how they could not stop reading the book, describing the story as absorbing, addicting and dramatic, full of surprises, twists and turns. Some people also disagree with the negative comments on the plausibility of the plot and



unbelievable characters and bad sex scenes. Some readers mention unrealistic sex scenes and describe the book as a 'misinterpretation of s&m relationship', some readers find the book not erotic enough. Another thing that stood out was the use of the word 'God' (figure 7). We noticed that the word God was used a lot in a religious way. People seem to be praying to God to forbid these kind of books to be published.



Figure 5. Negative reviews word cloud (made at <http://tagcrowd.com/>)

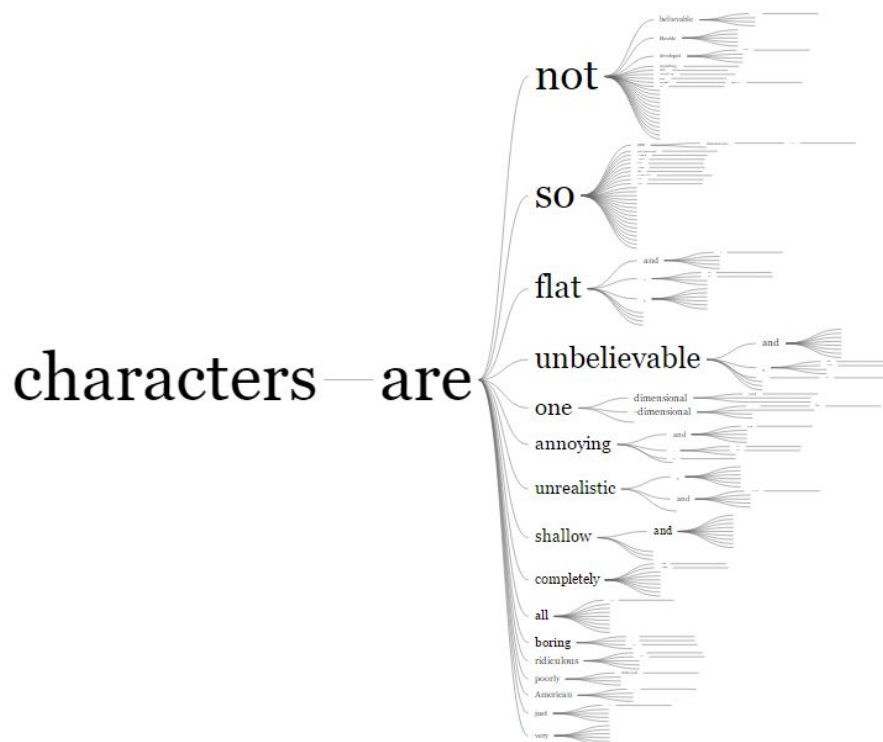
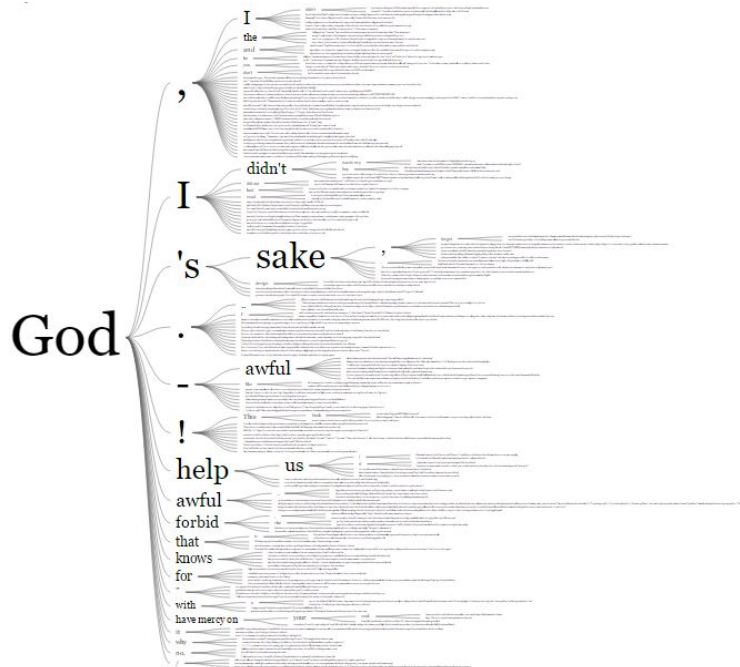


Figure 6. Word tree 'Characters' - negative ratings





## What is the motivation for people to start reading this book?

To find out about the reasons for people reading the book we decided to first of all look for words and sentences like: '*I read*' and '*I read this because*' (scripts 12-15). We noticed that these search terms do not really give the results you want. The word 'read' can be used in too many different contexts, such as 'worst thing I have ever read'. Weirdly enough, 'bought' gives a lot of results which lead to motivations for reading the book (script 11). Apparently, people are more inclined to give their motivation on money spent than time spent.

What we found out while close reading some of the results we got from using the first egrep commands is that there are several reasons given for reading this book. Words like “hype”, “fuzz”, “buzz” and “hoopla” seem to be used a lot. We then decided to search for these words using scripts (numbers 16-19). From these results we found out that most people say they read this book because of the popularity. Most of them say they wanted to see what the hype was about.

This is however still not an explanation for the book getting popular in the first place. People also say that all their friends loved the book or that someone recommended it to them. Next to that, people state that they read it because there were so many good reviews. This while most reviews in our file have a 1.0 score. Somehow this doesn't add up.

This made us wonder why people would even write reviews? Probably because they want to inform people on a book so that they can decide whether to read it for themselves. Most reviews in our file have a score of 1.0, and we found out they were mostly written as warnings (script 21). There are some cases of people literally warning their fellow reader. Most of them however contain some kind of implicit warning, like ‘the writing is bad’ or ‘this sex is dull’.

So if bad reviews are like warnings, for what kind of things would people warn each other when writing a review? We decided to look at the words from the word clouds made and especially the word sex. Using script 20, we found out 'sex' is being used in a lot of different contexts. People mostly want to warn future readers about sex.

When you reason like this it seems like sex is in fact a big reason for people being curious about the book. Especially since the reviewers probably warn other people of things they were themselves tempted by when picking up the book. This made us think that most people actually read this book because of the sexual content.

**Conclusion: "How could a book with a lot of bad attention, like Fifty Shades of grey, become a bestseller?"**

First of all, we have confirmed that the overall opinion about the book is negative and that we could speak of an instant negative hype. The things negative reviewers mostly dislike about the book are the poor writing, unbelievable characters and bad sex scenes. The things positive reviewers mostly like about the book are the story and especially the love story. While looking at why people started to read this book we confirmed that most people wanted to know what the hype was about. So the popularity of the book would be a good motivation for people to read the book. Besides that, we believe a lot of people also started to read the book because of its sexuality. Sex is still a kind of taboo in our community which ascribes to the fact people do not literally say they read the book because of the sexual content. However, we did find out that a lot of negative reviews contain warnings for the sex scenes. We suspect this would also be something people were interested in. In the end, this is also what makes the book stand out from almost all of the other books published.

## Reflection on lessons learned

First of all we found out that there were reviews about random products in the dataset, which were written before the book was even published. These results could falsify our findings and we have thought of different ways to filter them out. We tried different scripts but in the end we would always also remove some valid reviews. This is why we decided to keep the data like we go it.

There were some problems finding the right searchterm when searching for reasons the reviewers read the book. Since a lot of terms were too general to get useful results some close reading was needed in order to find the right terms.

Also, there seemed to be differences in word counts by different tools; this seemed to be due to different ways of interpreting similarities. The unix command line would for example read '*don't*' as '*don*' and '*t*'. Also words starting with a capital letter were counted as different words as the same words without capitals.

Something interesting we found out by doing this research is that there is not really a way of searching through reviews without doing a lot of close reading. There is no searchterm that can exclude every useless usage of that searchterm. Because people use language in a lot of different ways you can never really make sense of a text without looking at the context of the words used. We found ourselves pondering over this while staring at the command line, trying to come up with a regular expression that would perfectly capture people's motivation for reading the book.