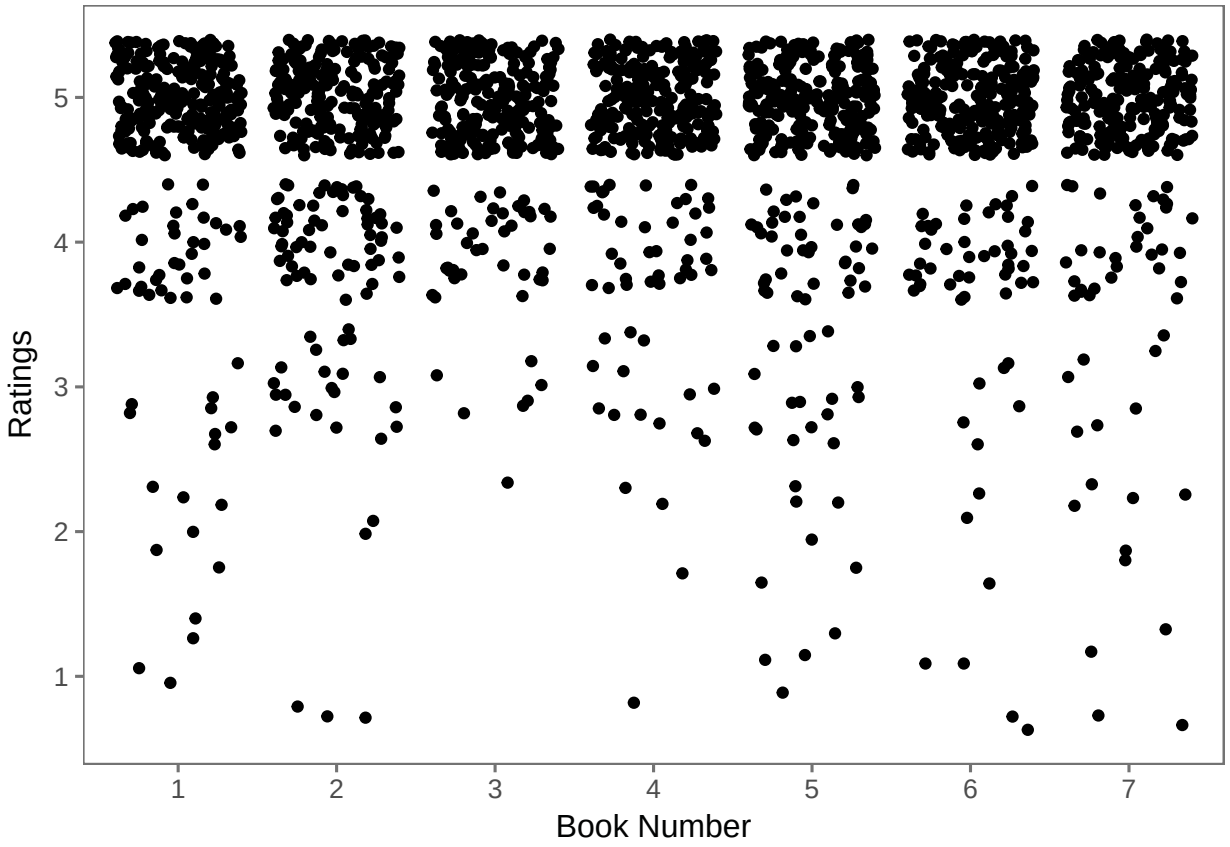
HarryPotter

Steven Tran

April 21, 2018

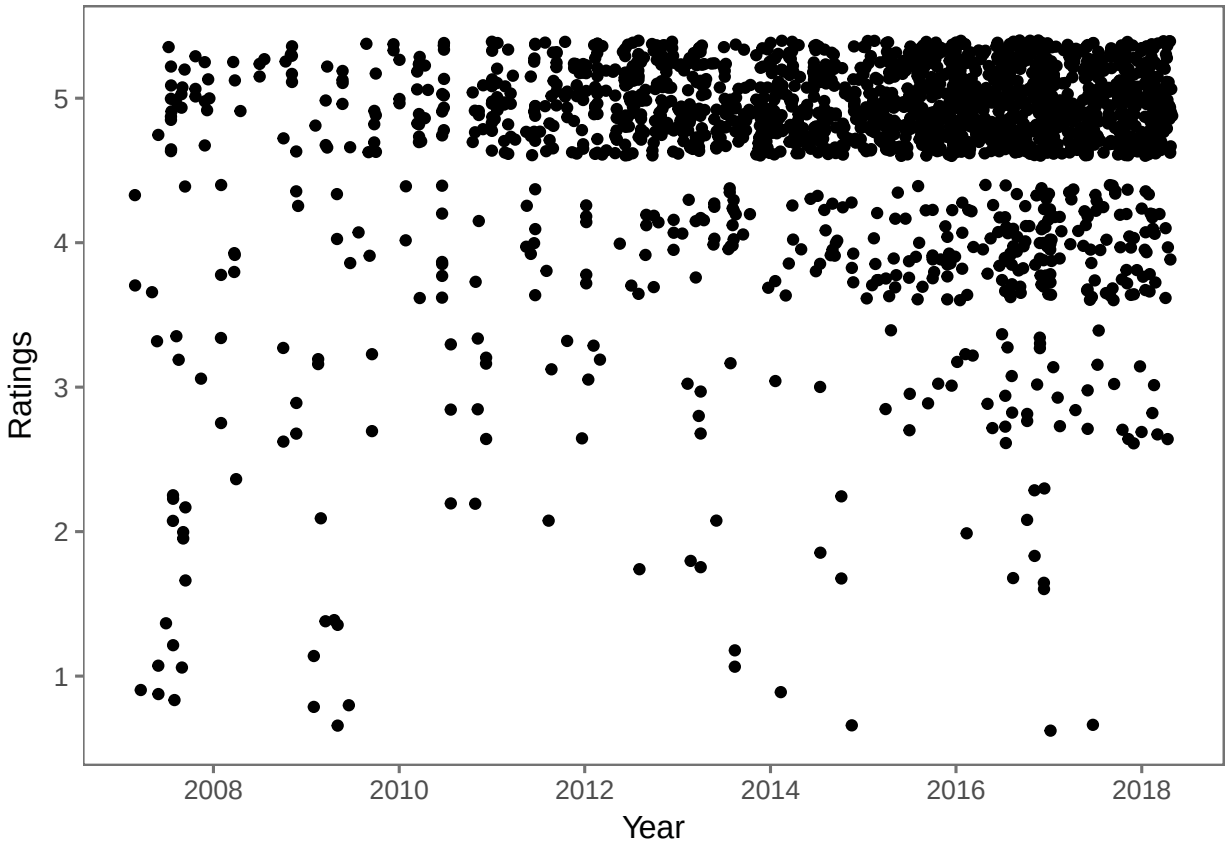
The Harry Potter series is a beloved series about Harry and his newfound experiences in the wizarding world. There, he discovers how magic has continued to hide from non-magical humans and to grow both for good and evil. It is also there that Harry discovers the cause of death of his parents and his mortal enemy, Voldemort. More chilling is the prophecy that states clearly that neither can live while the other survives.



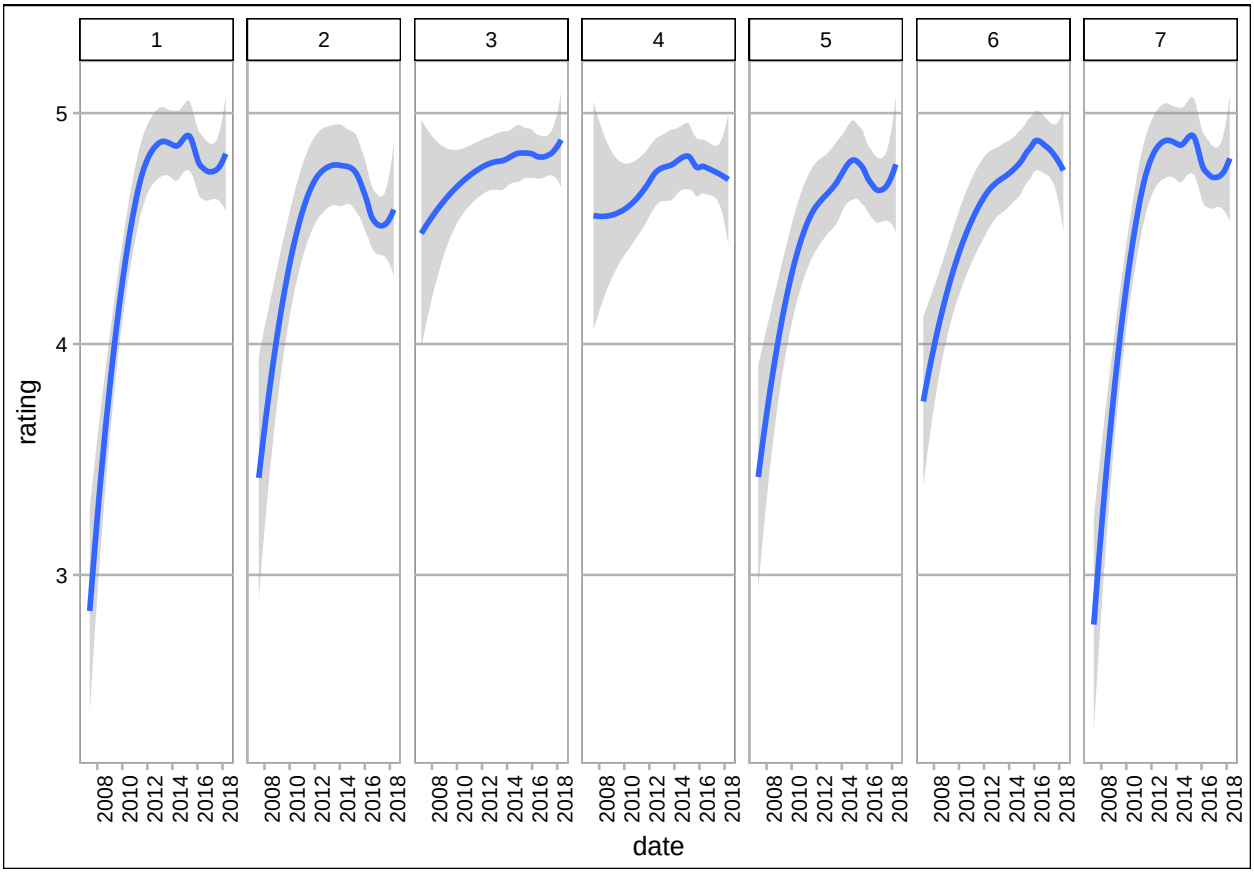
The above graph is the result of a relatively small section of the reviews on the Goodreads website. As you can see, each book in the series consists of mostly good reviews, with most critics giving it a four or five out of five.

```
## Analysis of Variance Table
##
## Response: rating
##                          Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(bookNumber)     6   7.78 1.29689  2.6827 0.0135 *
## Residuals              2004 968.80 0.48343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova also states there is too little evidence to say that any of the books has any different ratings than the others.



Here, we observe overall that most of the five star ratings comes from more recent years that the old ones. Thisis obvious by the dense clustering of five star ratings in the most recent years. The series clearly didn't pick up steam until around 2012. However, we need to keep in mind that the entire original series was completely published in 2007, so it is interesting to note that it took around five years until it started to gain the popularity it has today.



Here we have separate plots of each book's rating over time. They each show that most of them started out with decent reviews before surging to a five star rating.

```
##
##  Welch Two Sample t-test
##
## data:  ratings2007$rating and ratings2018$rating
## t = -4.5801, df = 48.1, p-value = 3.303e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.5735515 -0.6135061
## sample estimates:
## mean of x mean of y
##  3.673913  4.767442
```

This t-test confirms that it is highly likely that all ratings in 2007 and the ratings in 2018 are different from each other.

```
##
##  Welch Two Sample t-test
##
## data:  book1ratings2007$rating and book1ratings2018$rating
## t = -2.4128, df = 6.2737, p-value = 0.05059
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.275580983  0.005739714
## sample estimates:
## mean of x mean of y
```

```
##  3.142857  4.777778
```

```
##
##  Welch Two Sample t-test
##
## data:  book2ratings2007$rating and book2ratings2018$rating
## t = -1.0329, df = 3.1359, p-value = 0.3746
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.006784  2.006784
## sample estimates:
## mean of x mean of y
##       3.5       4.5
```

```
##
##  Welch Two Sample t-test
##
## data:  book3ratings2007$rating and book3ratings2018$rating
## t = -0.63562, df = 1.0739, p-value = 0.6335
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.828587  5.181528
## sample estimates:
## mean of x mean of y
##  4.500000  4.823529
```

```
##
##  Welch Two Sample t-test
##
## data:  book4ratings2007$rating and book4ratings2018$rating
## t = 1.3887, df = 12, p-value = 0.1902
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1312903  0.5928287
## sample estimates:
## mean of x mean of y
##  5.000000  4.769231
```

```
##
##  Welch Two Sample t-test
##
## data:  book5ratings2007$rating and book5ratings2018$rating
## t = -2.4992, df = 7.3074, p-value = 0.03967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.91692272 -0.09307728
## sample estimates:
## mean of x mean of y
##     3.375     4.880
```

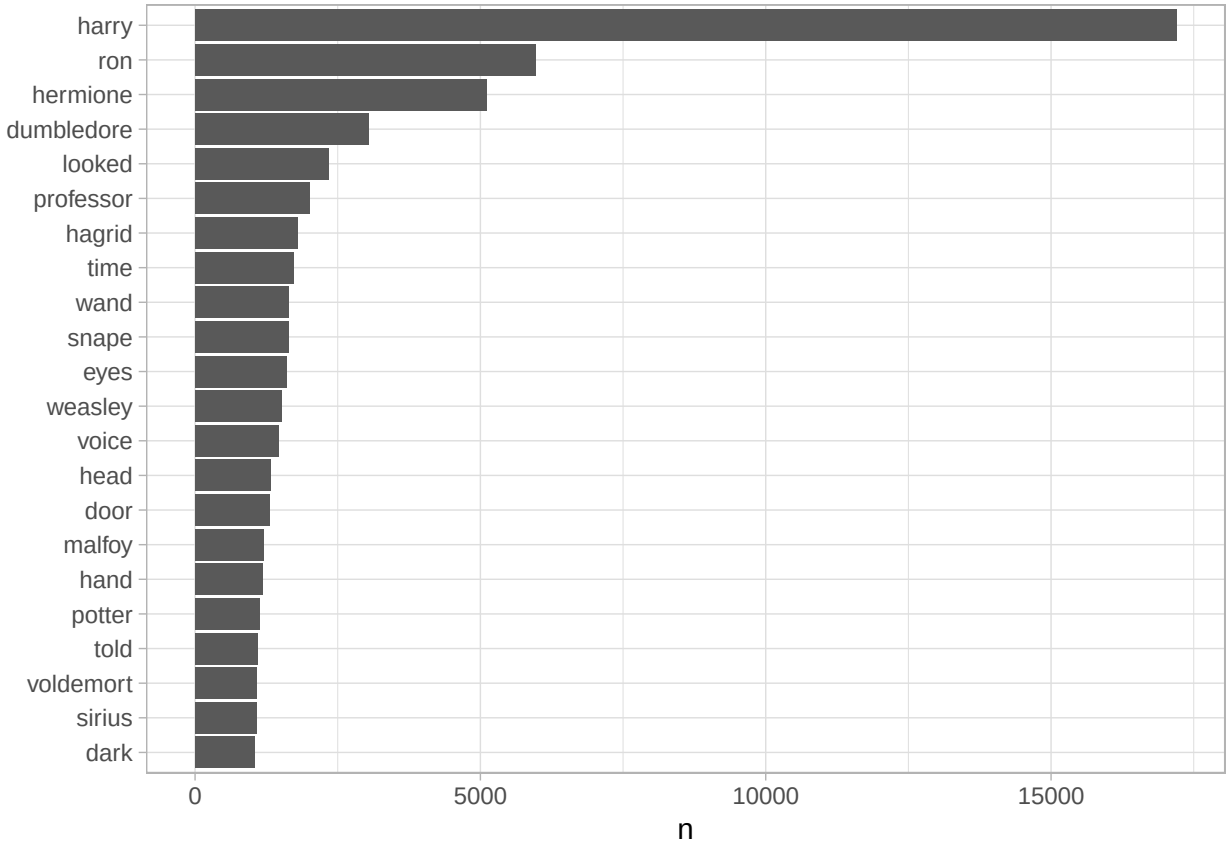
```
##
##  Welch Two Sample t-test
##
## data:  book6ratings2007$rating and book6ratings2018$rating
## t = -2.3711, df = 12.662, p-value = 0.03433
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -2.00655891 -0.09060708
## sample estimates:
## mean of x mean of y
##  3.846154  4.894737
```

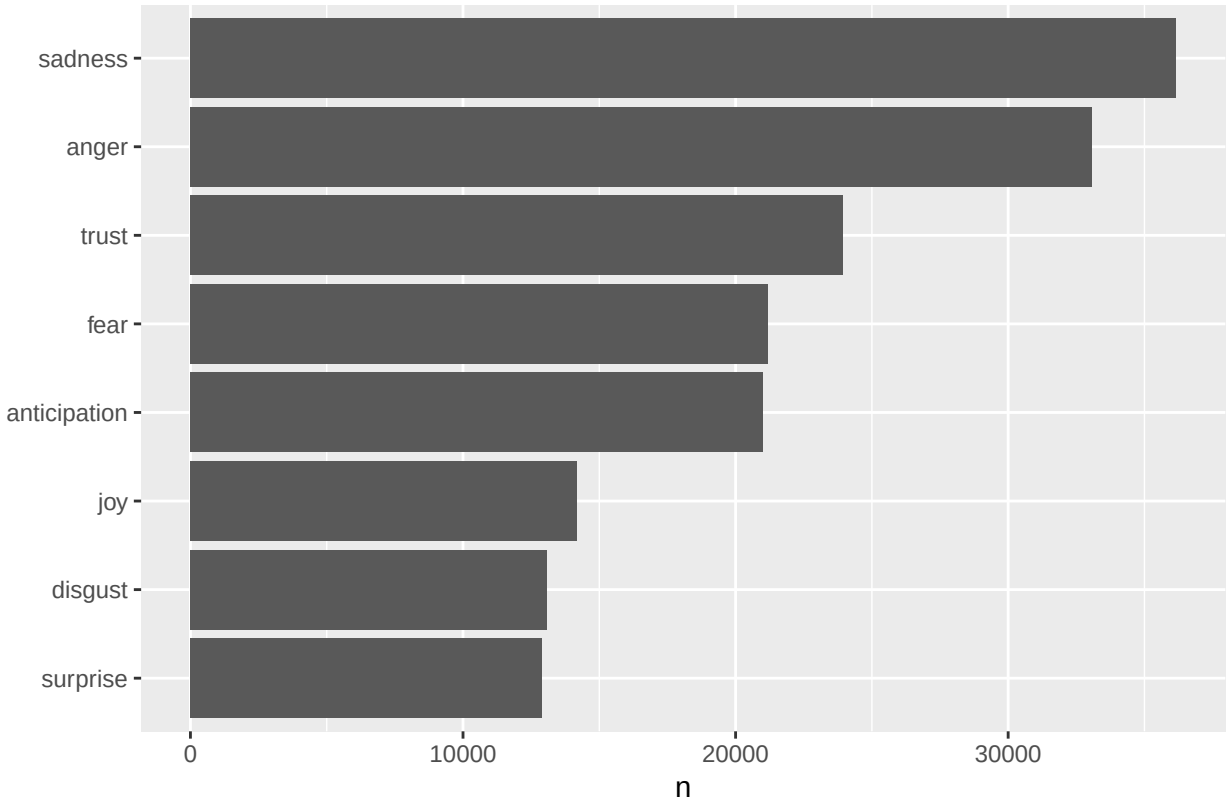
```
##
##  Welch Two Sample t-test
##
## data:  book7ratings2007$rating and book7ratings2018$rating
## t = -2.3376, df = 6.3765, p-value = 0.05554
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.23182082  0.05086844
## sample estimates:
## mean of x mean of y
##  3.142857  4.733333
```

However, separating the ratings by book, we can see that there isn't enough evidence to suggest that the ratings for most of the books to have changed their ratings. Only books five and six have a high chance that its ratings in 2007 are different than in 2018.

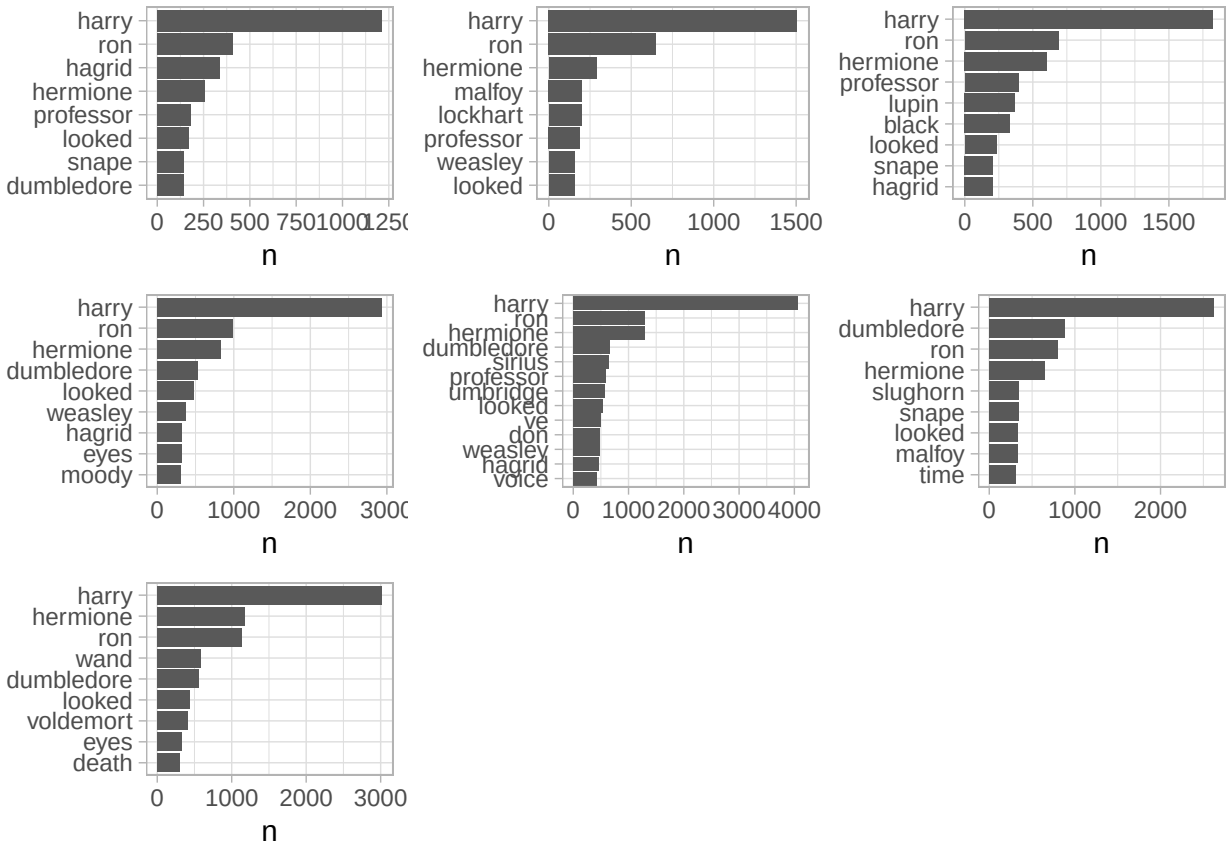
Now we can look at word count and sentiment analysis.



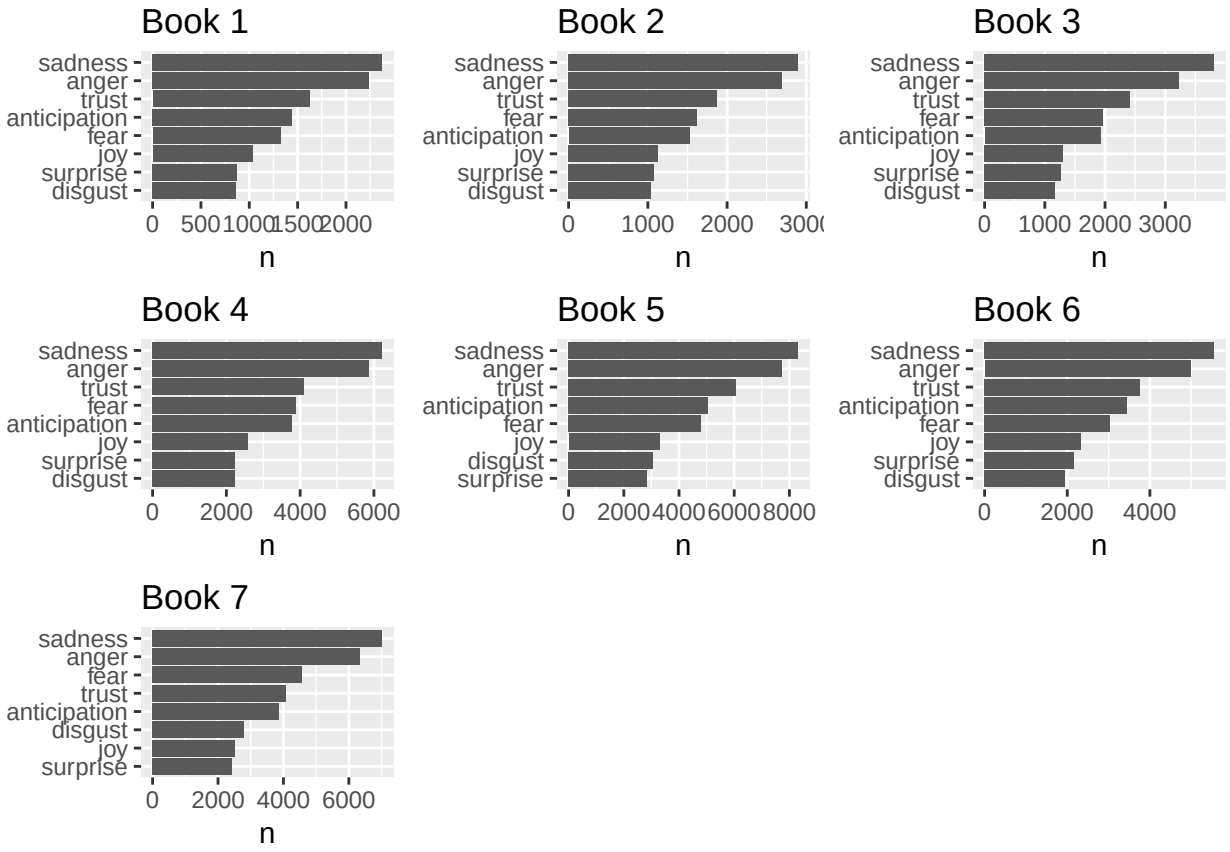
Here we see that the main character's name is the most used word. This should have been obvious since the text is written with an unreliable narrator focused on Harry. His friends' names are next in line since he interacts with them the most. The words "looked" and "eyes" emphasize Rowling's focus on sight and possible indicate that Rowling is trying to state that not everything is as it seems in the magical world. After that, we have many last names like "Weasley", "Malfoy", and "Potter."



With all the losses and the suffering in these books, it is no wonder that the Harry Potter series is doused in negativity. There's always the sadness that comes with loss, which is then followed swiftly by anger at the offender. There is also the fear that Voldemort inflicted upon the wizarding world and the disgust that pureblood wizards express at all others. I do, however, believe that this negativity causes the reader to keep turning the page in hopes of Harry overcoming these emotions and doing his best to fight against evil.



Here we have the word counts for the Harry Potter series by books. Unsurprisingly, Harry is still the most used name and the people whom he interacted with the most follow far behind. The main difference between each book is only the order of the names below the first. Other than names, the theme of sight is still prevelant in most of the books.



Separated by books, there is not much difference in results. Sadness, anger, trust, and anticipation are always at the top of the list of sentiments when it comes to the Harry Potter Series.