# Trends within r/DestructiveReaders from Dec/2018 to September/2019

*Aleks Wittkamp*

*October 3, 2019*

For years now my hobby has been writing short fiction. For much of that time I've been active on the writing critique subreddit r/DestructiveReaders. In the past, some users on the subreddit have analyzed trends related to submission wordcount and the amount of likes and comments each submission receives. It's been about three years since the last time somebody did this. I'm studying stats and computer science and I find data interesting. I thought I could do another round of analysis.

To start, I used the beautifulsoup module in Python to scrape the subreddit. Reddit's API restricts access to more than the most recent 900 posts, so that was the size of my dataset. For each post, I scraped the title, genre, wordcount, number of comments, number of likes, and the date of submission. I was able to get complete data on nearly every post, though there were some exceptions, described below.

There were posts whose author accounts have been deleted, in which case I made the author 'deleted'.

There were posts whose genre wasn't provided, in which case I made the genre 'other'; posts whose genre was strange but contained a word like 'fantasy' somewhere, in which case I simplified it; and over a hundred posts whose genres were strange and not easily parsed, in which case I marked them 'other'.

One last corner case is that some of the posts are meta posts made for community discussion. I included these in the dataset with the genre 'meta', but excluded them from much of my analysis.

The end result is 900 posts' worth of data going back to the end of November 2018, however I've excluded the data from November so as to make it a clean 9 months' worth.

Please find my python scraping code and the dataset csv elsewhere in this file.

Find a sampling of the data below.

| | Title | Author | Upvotes | Comments | WordCount | Genre | Date |
|---|---|---|---|---|---|---|---|
| 13 | [1818] The Order of the Bell: A Vision and a Reunion | md_reddit | 8 | 3 | 1818 | fantasy | 2019-09-30 |
| 14 | [2,581] SheMancer: Chapter One | TempestheDragon | 3 | 8 | 2581 | fantasy | 2019-09-30 |
| 15 | [2212] Bolim Two MilesScifiThriller | useful_idiot118 | 4 | 18 | 2212 | other | 2019-09-30 |
| 16 | [1024]Samsheer | rksomayaji | 3 | 4 | 1024 | other | 2019-09-30 |
| 17 | [808] Rerun | vjuntiaesthetics | 13 | 4 | 808 | scifi | 2019-09-29 |
| 18 | [2125] Understanding UnderstandingFictive NonFiction | wowdor | 2 | 13 | 2125 | other | 2019-09-29 |

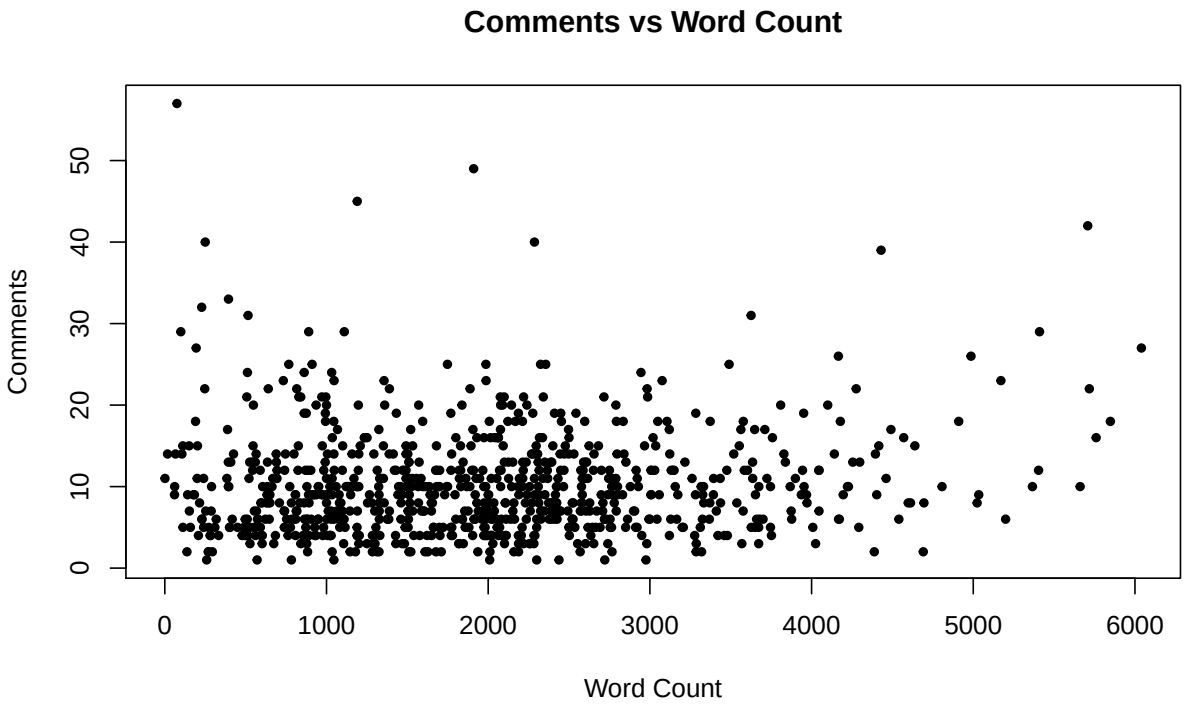
And now a quick summary.

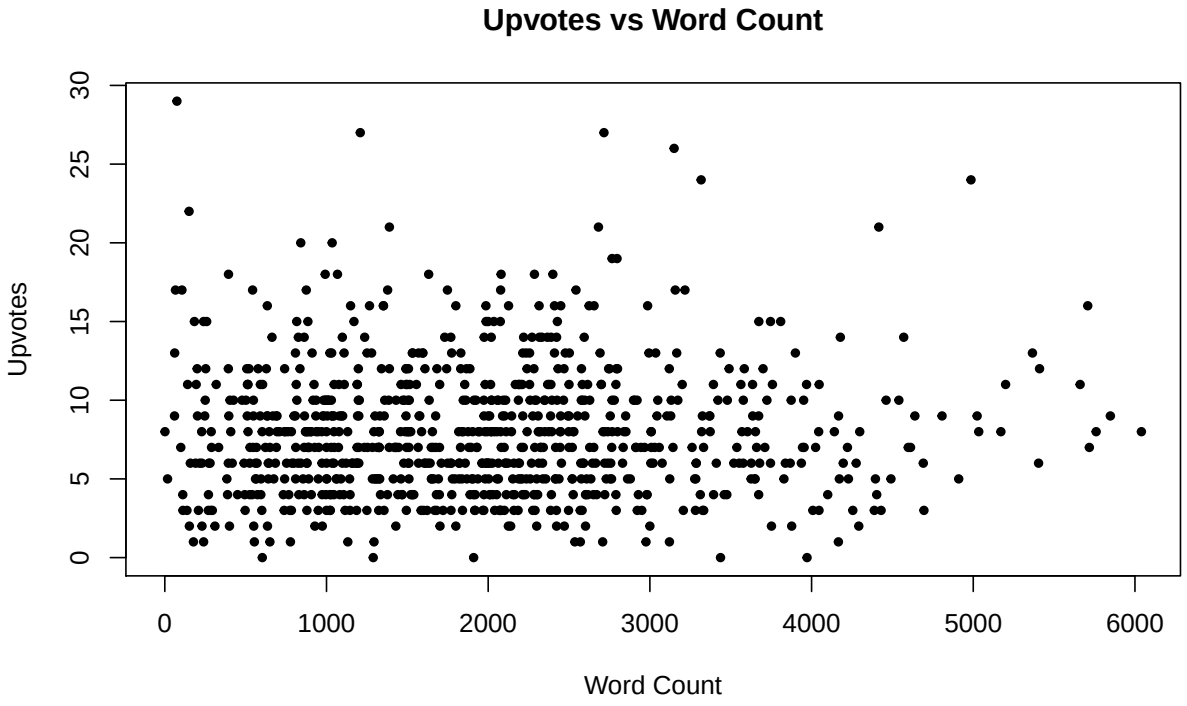
| Title | Author | Upvotes | Comments | WordCount | Genre | Date |
|---|---|---|---|---|---|---|
| [ 158 ] The Garden : 1 | md_reddit : 77 | Min. : 0.000 | Min. : 1.0 | Min. : 1 | other :514 | Min. :2018-12-01 |
| [1,002] Greydogs : 1 | OldestTaskmaster : 17 | 1st Qu.: 5.000 | 1st Qu.: 6.0 | 1st Qu.:1046 | fantasy :167 | 1st Qu.:2019-02-21 |
| [1,190] The Executive Suite : 1 | deleted : 15 | Median : 7.000 | Median : 9.0 | Median :1912 | scifi : 39 | Median :2019-05-09 |
| [1,197] Pennies : 1 | Jraywang : 15 | Mean : 7.951 | Mean :10.2 | Mean :1966 | horror : 27 | Mean :2019-05-06 |
| [1,330] Proem : 1 | TheManWhoWas-Tuesday: 14 | 3rd Qu.:10.000 | 3rd Qu.:13.0 | 3rd Qu.:2602 | contemporary: 23 | 3rd Qu.:2019-07-20 |
| [1000] Behind the Looking Glass : 1 | RustyMoth : 12 | Max. :29.000 | Max. :57.0 | Max. :6040 | ya : 14 | Max. :2019-09-30 |
| (Other) :829 | (Other) :685 | NA | NA | NA | (Other) : 51 | NA |

Over the last 9 months, there were 835 stories submitted, consisting of 1641900 words, and these were submitted by 416 authors. These stories received a combined total of 8518 comments and 6639 upvotes.

The most salient question on my mind going into this was whether there was a correlation between the length of a story and the numbers of comments and upvotes it receives. Common wisdom on the subreddit says that longer posts receive less attention because people are less willing to engage with the longer piece.

Here are the relevant scatterplots.




Fitting a linear model in R to these pairings either returns, in the case of Comments, a statistically significant slope of 0, while in the case of Upvotes, a p-value so large that we reject the hypothesis that Upvotes and Word Count are correlated. We have thus found no correlation between Word Count and Upvotes/Comments.

Unsurprisingly, an analysis of Upvotes and Comments does show a statistically significant positive correlation. It makes sense that people either comment on the posts they like or like the posts they comment on.

I did more testing here to answer questions that I thought would be relevant to the subreddit. I determined which days of the week were the most busiest and I came up with leaderboards ranking those users who did things like post the most or receive the most upvotes. However for this posting on github, I'll leave off here.

My takeaways from doing this are a better understanding of R and beautifulsoup as well as a deeper appreciation for the value to a data scientist in knowing which questions are worth asking of the data. I gave a lot of information in the post I made to reddit, but community engagement was limited to only a few of the statistics. I would have benefited from asking the community ahead of time to tell me which sorts of questions they'd like the answers to. Then I could have made sure that the data I was gathering was able to answer those questions. For example, one of the questions I was asked was whether I could determine how many of the comments constituted story feedback, and with my dataset I wasn't able to find that out. In hindsight, I should have scraped the comments that the posts received.

Beyond all this, this project has given me some experience in asking myself what sorts of questions I can ask of the data I have available. I'm very glad I did this. I'll try to do something like it again.

Thanks for reading. Have a great day!