# Investigating Patterns of Brigading on Reddit

Adam Kaderabek

Fundamentals of Computing and Data Display

University of Michigan

December 16, 2020

## Abstract

Norm-violation in social media occurs at both the community and individual levels. There is a growing focus on the impacts of collective social action in online communities. Much of the research on inter-community conflict as a form of online collective action has aimed to identify network relationships amongst antagonists. Collectively antagonistic action can be deployed as a measure of oppression or social reform, this begs the question of its implementation, tactics, and outcomes. In this paper, I look at the online phenomenon known as "brigading" and its prevalence on Reddit. There is no doubt that brigading exists in the realm of social media phenomena but existing research has primarily focused on extreme norm-violation in online behavior. As not all collective social action is an extension of extremism (whether in favor or against the status quo), it is the goal of this paper to identify brigading can be observed metrically using data available through the Pushshift API and what, if any, evidence is available regarding the level of social acceptability.

# Contents

# 1  Introduction

## 1.1  Background

Parallel to our analog existence, online communities also set standards for normative behavior and these social norms are interwoven throughout the community governance. Similar to our in-person social interactions, online social norms are also emergent, or that is to say, they are dependent on dynamic relationships among the community members and these norms will vary between communities (Chandrasekharan et al. 2018). This formalization of social norms exists on two levels, one which is the interpretation of community members and the enforcement of these norms by community leaders; known as moderators in the world of Reddit (Matias 2016).

Although there is a plethora of norm-violating behavior in online communities it is important to recognize some behaviors such as trolling, baiting, or fisking, are exclusively individualistic. Within the realm of collective action, there are also several types of norm-violating phenomena that occur in the Reddit-sphere. The most commonly accepted of these norm-violations include mass spamming, flame wars, brigading, crapflooding, and shitposting (Kumar et al. 2018).

## 1.2  Properties of Brigading on Reddit

In this paper, I focus specifically on brigading. *Brigading occurs when a large number of computer users deliberately give a ranking to something (such as a movie or restaurant) or vote in a poll so as to create a result that likely would not have occurred if only the usual sample of respondents had voted or ranked*[1]. Brigading as an act of norm-violating collective action presents several challenges for researchers.

### 1.2.1  User Anonymity

As is the case in most off-line communities, participants are generally norm-compliant, but, anonymity can reduce the inhibitions of users (Suler 2004). The creation of a Reddit account only requires a valid email address. This allows anti-social actors tactical optionality, such that, they can engage in norm-violating behavior under a legitimate, yet anonymous user name or engage in maleficence through the creation of 'throw-away' accounts which evade the risks of being banned from Reddit or retaliation by other users. In combination with user anonymity, the lack of available information on author associations makes it difficult to identify social collusion.

Reddit does not explicitly convey the network relationships of communities or members. Depending on how inference is done and what data drove it, different networks may emerge

---

[1]brigading. 2020. In *Merriam-Webster.com.* Retrieved December 1, 2020, from https://www.merriam-webster.com/words-at-play/brigading-online-poll-meaning

(Datta, Phelan, and Adar 2017). There are methods for identifying users and their frequented communities, however, when coupled with the above challenges of anonymity and moderation, those relationships can become ambiguous and they do not allow for evidencing collaborative social action without additional context.

### 1.2.2 Community Moderation

The second challenge Reddit presents for researchers relates not to the norm-violation *per se*, but the enforcement of community norms in the form of moderation. Moderators are the protectors of the realm in the world of Reddit. They are volunteers and members of their respective communities who take on the obligation of enforcing the community norms. In addition to the human component of moderation, subreddits also employ 'auto-moderation', or the use of bots that target possibly controversial submission or comments and quarantine or remove them from public view.

When a submission is removed, moderators can choose to explain why the submission or comment was removed. The behavior of moderators can be as varied as the community rules they are tasked with enforcing; however, moderators can provide feedback to the content creator by commenting on the removed post and explaining the reason for removal, flair[2] the removed post, or send a private message to the submitter (Jhaver et al. 2019).

### 1.2.3 Event Temporality

Lastly, it is important to understand the temporal nature of brigading. Brigading is a context dependent phenomenon in that it must both be motivated and executed within a specific subreddit at a specific point in time. The most common definition of brigading refers specifically to "down-vote brigading" in which a brigade attempts to down-vote a submission to impede or entirely prevent it from increased visibility accompanied by increasing up-votes. Further complications arise from the inability to compare up-and down-votes directly through API requested data which is an intentional aspect of the Reddit API design.

## 2 Data

## 2.1 Pushshift API

Reddit data can be accessed directly through the Reddit API but the limitations on API requests can limit the particular data available and the speed at which it can be retrieved. Pushshift is a social media data collection, analysis, and archiving platform that since 2015 has collected Reddit data and made it available to researchers. Pushshift's Reddit dataset is updated in real-time and includes historical data back to Reddit's inception (Baumgartner

---

[2]Post flair allows mods and community members to create visual flags for tagged content. It can be used to help categorize post content types, flag content about a certain subtopic of the community, tag a post with removal reasons, or anything else redditors think might be useful.

et al. 2020). Since that its inception, the Pushshift Reddit dataset has drawn significant attention from researchers; as of December 2020, Google Scholar indexes just shy of 300 publications using Pushshift data.

## 2.2 Sample Data

I reviewed submissions including the keyword "brigading" for all Reddit posts from January 1, 2020, through December 14, 2020. Excluding posts made by moderators, there were 5,493 submissions made with a total of 224,979 top-level comments. These submissions were made by 3,857 individual Redditors across 1,864 unique subreddits. These numbers inflate greatly with the inclusion of "brigade" and "brigaded" however these forms of the word overlap extensively with gaming and military subreddits using the traditional semantics of the words so they have been excluded from my investigation. The rates of keyword submissions and comments are displayed by month in figure 1.
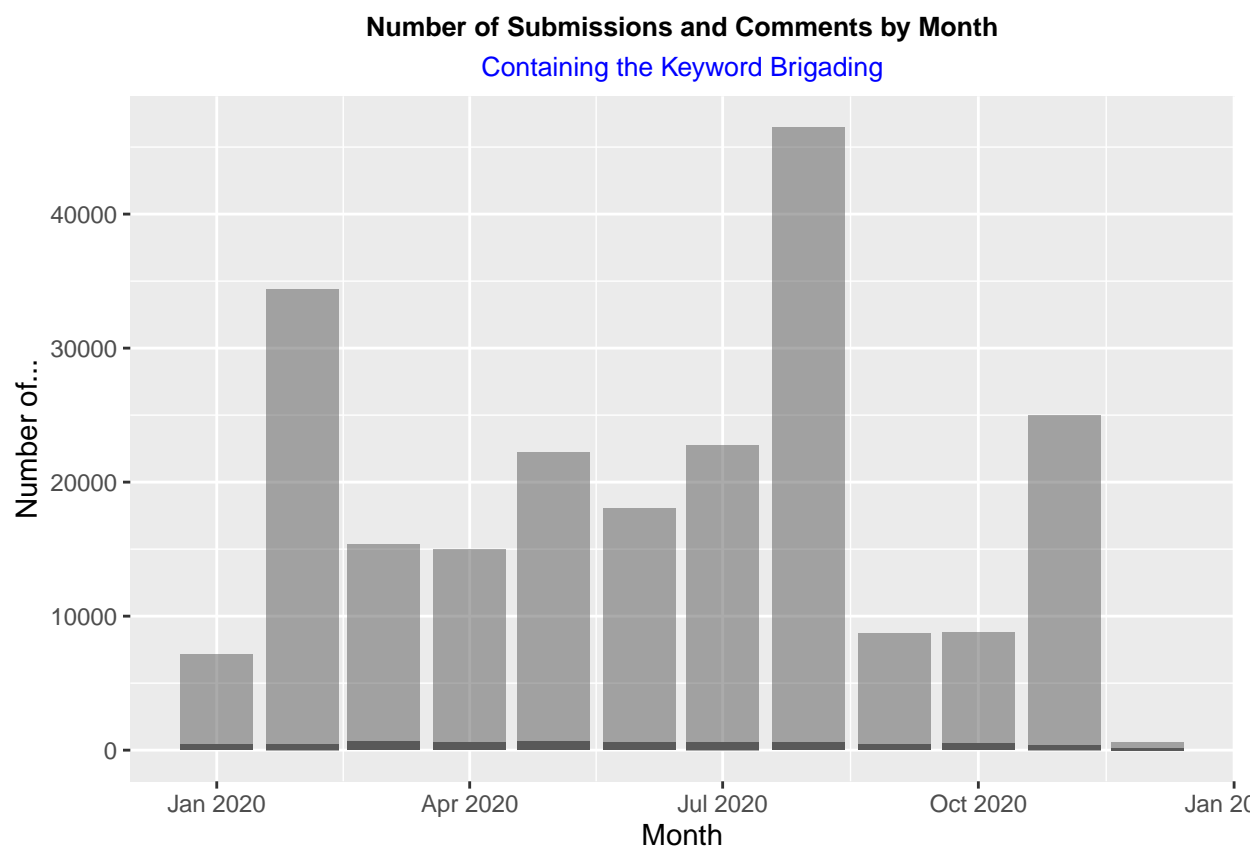


Figure 1: Number of Keyword Submissions and Comments by Month

# 3 Method

I partitioned all submissions posted by Reddit users and moderators/admins in my analysis. In total, 5.73% ($n_{modsubmissions} = 344$) of the 5,527 submissions were posted by moderators or admins. Of the 224,353 comments which were made across all submissions, less than 1.5% ($n_{modcomments} = 2,995$) were made by moderators or admins. The submissions relate largely to discussions on subreddit rules, changes in rules, or moderator lead discussions on alleged instances of brigading, whereas the comments may be similar in content they are attached to specific user submissions.

I focused on specific subsets of the data to identify patterns or bias arising from both user and moderator, that subset is comment-related behaviors occurring in the first thirty-six hours after a submission was posted. Any collaborative action requires a time line. I also investigated the patterns between submissions with removed comments compared to submissions without removed comments. Figure 2 displays the number of comments against the number of removed comments for 713 submissions accounting for the 10,963 removed comments.
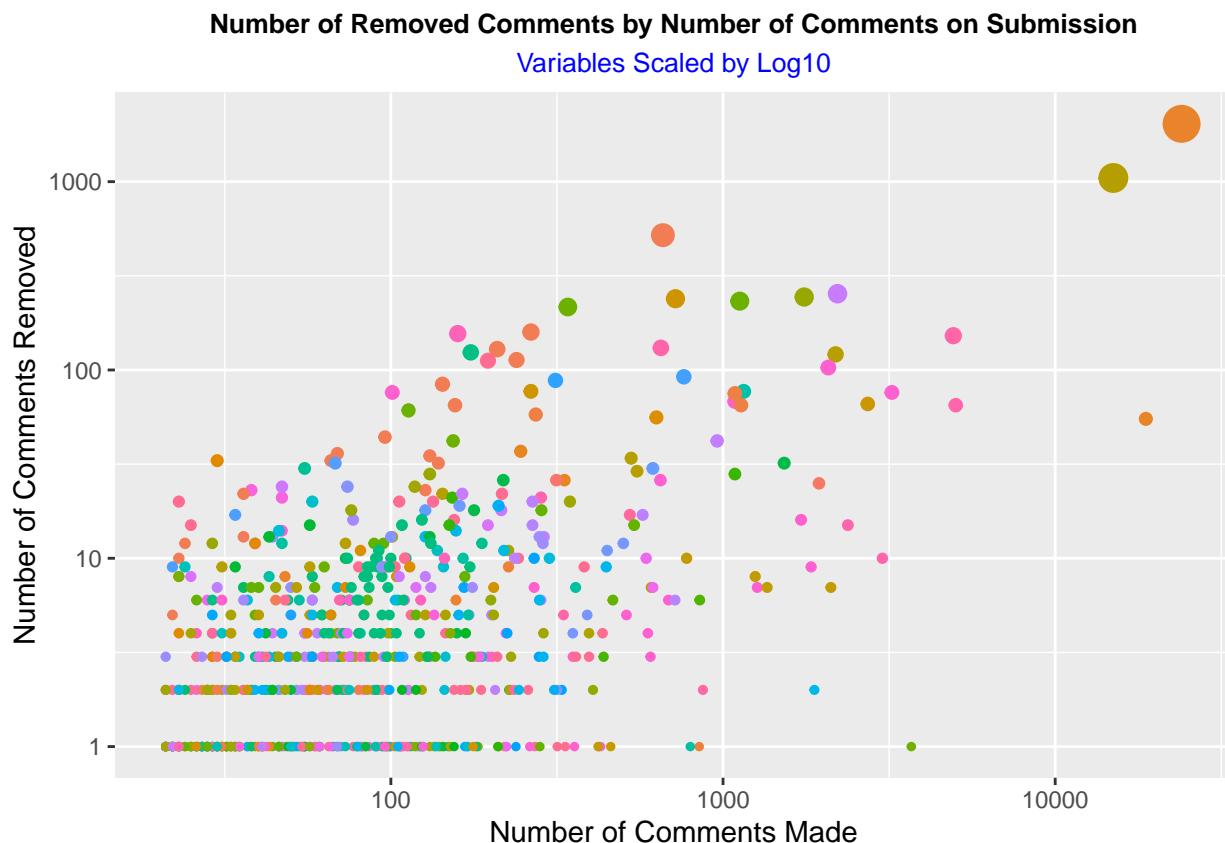


Figure 2: Number of Removed Comments by Number of Submission Comments

# 4 Results

There were 144 submissions which maintained at least 10 removed comments; of these submissions only 5 had been removed. As one might expect there does appear to be a linear relationship between total number of comments and the number of removed comments; however, as is visible in figure 2, it appears the relationship is less linear below a threshold of 10 removed comments. The removed submissions (excluding the deleted authors) were made by 432 different users. Only one user had more than 5 removed submissions.

Redditors posted 5,493 submissions and 221,358 comments. There were 72 submissions by deleted authors, 103 by auto-moderators, and 403 which were posted by two self-identified bots - removalbot posted 306 times, and MarkdownShadowBot posted 97 times. There were 562 removed submissions and 10,963 removed comments. I also looked at the mean comment score for submissions based on the number of removed comments attached to the submission and found a unique distribution of scores which is displayed in figure 3. Excluding the outliers we see a consistent flattening of the mean comment score as the number of removed comments increases.
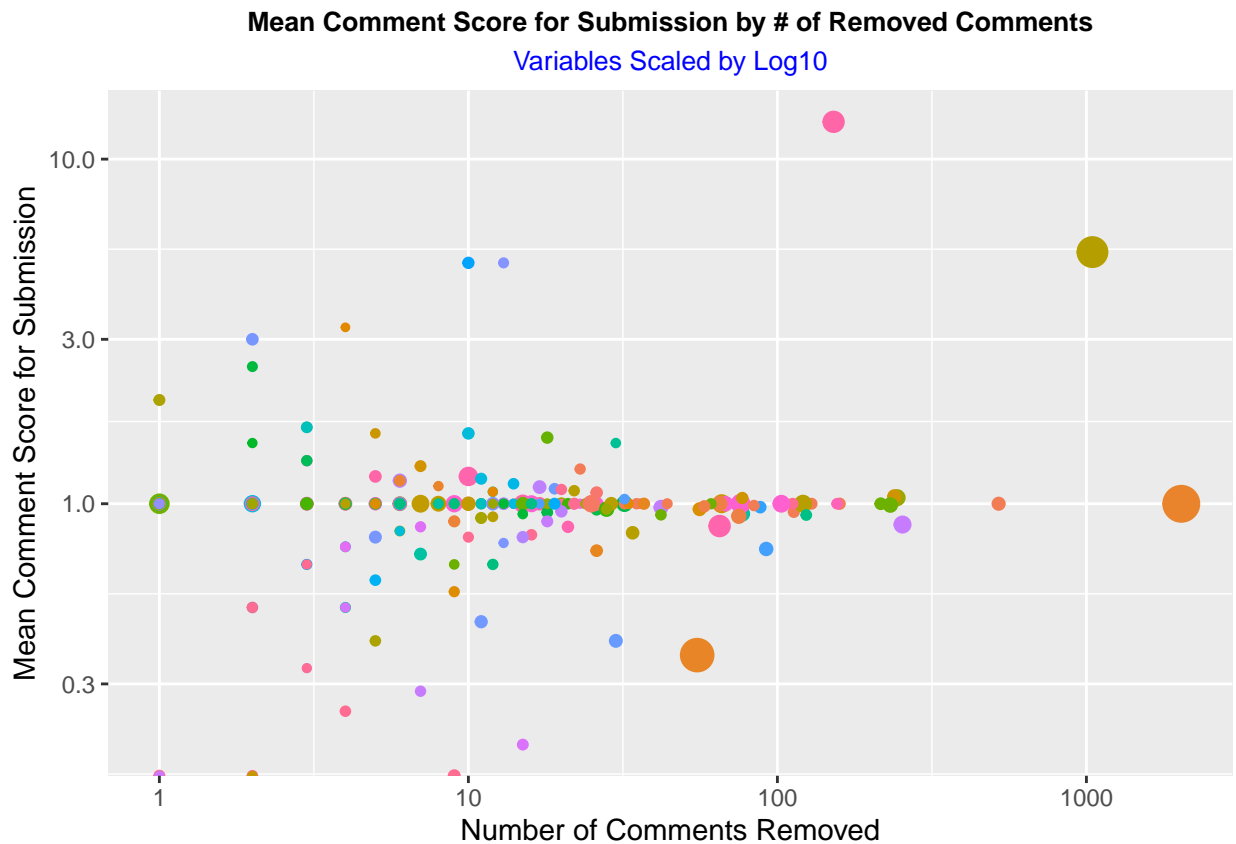


Figure 3: Number of Removed Comments by Average Comment Score

Compared to the mean comment scores for the submissions with removed comments the mean scores for the entire dataset appear to have the same level of variance. The distribution of mean comment score by number of comments is presented in figure 4.
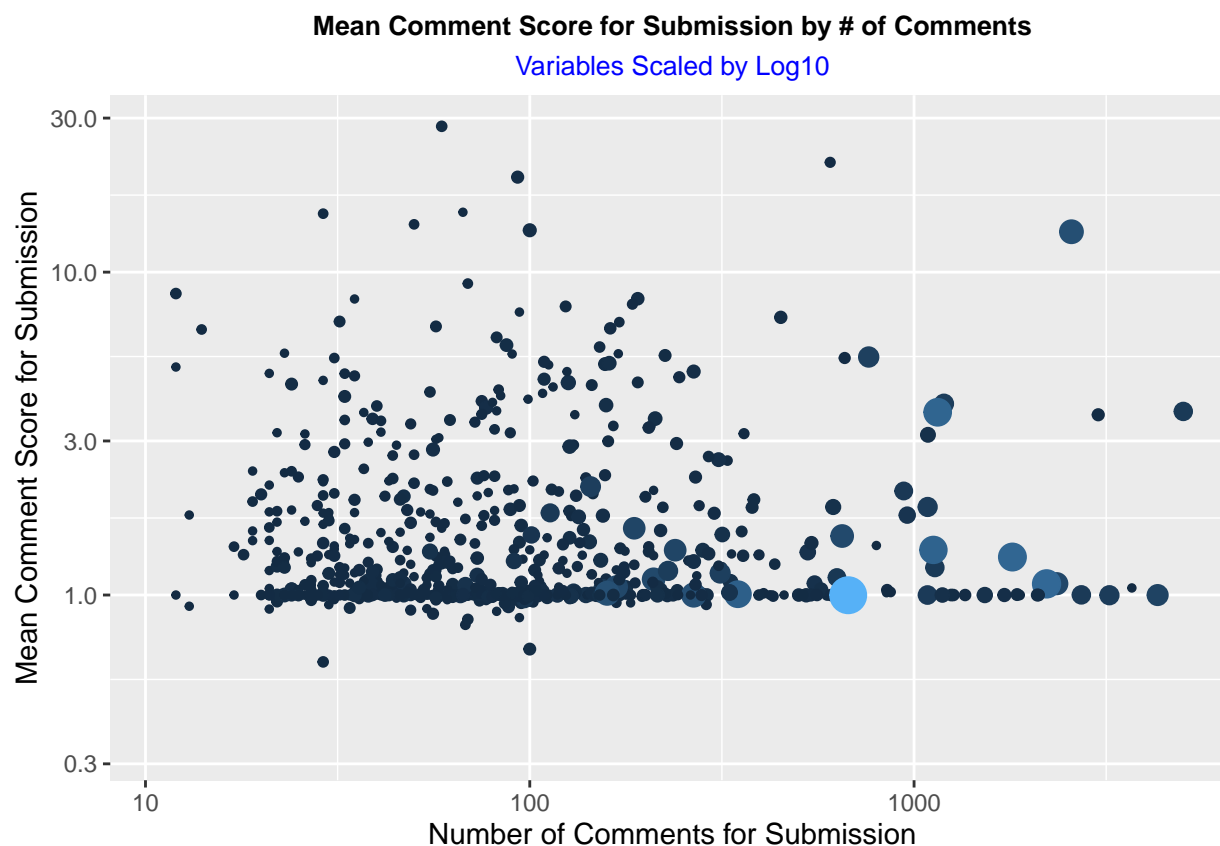


Figure 4: Number of Comments by Average Comment Score for All Submissions

## 4.1 Analysis

Metrically observing brigading was difficult to do with the available data. Brigading is most narrowly defined as "down-vote brigading" in which participants excessive their freedom to vote on Reddit content. No measure of down-votes is available through PushShift or the Reddit API. Scores are supposed to identify the difference between the number of up-votes and the number of down-votes, however Reddit masks these values and makes periodic adjustments to their scoring algorithm. The Reddit API does allow retrieval of the number of up-votes and it also maintains accurate up-vote ratios which are not fully accurate within the PushShift data.

The hypothesis that removed comments could be indicative of brigading a specific submission could not be fully tested since there is no ability to identify the content of the comment once

removed. Some additional information could be available through the collection of second-tier comments where moderators may have provided commentary when removing a particular comment. It is unlikely the peripheral data would be complete enough to be conclusive of a collaborative brigading event however.

# 5  Discussion

In-depth analysis allowing for inference or modeling of brigading on Reddit will face the same challenges discussed in this paper, however, the inclusion of multiple data points related to voting behavior could help overcome these limitations. Qualitative analysis, including sentiment analysis and text mining, could identify whether or not brigading presents a bias towards antagonistic conflict or whether it may also function as an act of resistance. Because moderation practices vary significantly by community it is difficult to make assumptions about the motivations or how actors may identify successful outcomes but most communities discuss the topic of brigading openly and in diverse context.

This investigation was limited due to the ambiguity of the data available, but I was able to identify key properties of the phenomenon of brigading as well as relevant variables that could aid future exploration. The review of commentary and censure, if directed at specific historical instances of alleged brigading, could also provide new insights although such evidence may also be community/context dependent.

# 6  References

Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. "The Pushshift Reddit Dataset." Pp. 830–39 in *Proceedings of the international aaai conference on web and social media.* Vol. 14.

Chandrasekharan, Eshwar, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. "The Internet's Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales." *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW):1–25.

Datta, Srayan, Chanda Phelan, and Eytan Adar. 2017. "Identifying Misaligned Inter-Group Links and Communities." *Proceedings of the ACM on Human-Computer Interaction* 1(CSCW):1–23.

Jhaver, Shagun, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. ""Did You Suspect the Post Would Be Removed?" Understanding User Reactions to Content Removals on Reddit." *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW):1–33.

Kumar, Srijan, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. "Community Interaction and Conflict on the Web." Pp. 933–43 in *Proceedings of the 2018 world wide web conference.*

Matias, J. Nathan. 2016. "Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout." Pp. 1138–51 in *Proceedings of the 2016 chi conference on human factors in computing systems.*

Suler, John. 2004. "The Online Disinhibition Effect." *Cyberpsychology & Behavior* 7(3):321–26.

# 7 Appendix

## 7.1 R Code Chunks

```r
# Extract moderator/admin authored posts from all submissions
moderator_submissions_df <- subs_df %>%
  filter(distinguished != "")
```

```r
# Extract moderator/admin authored posts from all submissions
user_submissions_df <- subs_df %>%
  filter(distinguished == "") %>%
  select(author = ï..author, 2:16)
```

```r
# Extract moderator/admin removed posts from all submissions
removed_submissions_df <- subs_df %>%
  filter(removed_by_category != "")
```

```r
masterblaster <- subs_df %>%
  left_join(comments_df, by = "id") %>%
  select(sub_ID = id, sub_date, sub_month = Month.x, sub_week = Week.x,
         subreddit = subreddit.x, subreddit_subscribers, sub_author = ï..author,
         sub_AuthType = authorType.x,  sub_title = title, selftext,
         num_comments, sub_removed = removed.x, removed_by_category,
         comment_date, comment_month = Month.y, comment_week = Week.y,
         comment_author = author, comment_AuthType = authorType.y,
         comment = body, is_submitter, author_premium, comment_score = score,
         banned_at_utc,  comment_removed = removed.y, comment_locked = locked,
         comment_nofollow = no_follow)
```

```r
# Range_df <- subset(masterblaster, comment_removed==1)
moderator_comments_df <- subset(masterblaster, comment_AuthType!="user")
```

```r
moderator_comments_df <- comments_df %>%
  filter(distinguished != "")
```

```r
moderator_comments_ByModBySubR <- moderator_comments_df %>%
  group_by(subreddit, body, author, .add = TRUE) %>%
  summarise(count = n())
```