r/AITA Polarization and Popularity

Peter Kress

2022/04/16

Contents

```
##################)
### Author: Peter Kress
### Date: 2022/04/06
### Purpose: Analyze AITA posts and comments
##################)
```

Is Polarization Popular on AITA?

I examined the top comments from top AITA subreddit posts from 2018-2019 to explore whether popular posts are associated with more engaged and polarized comments section.

AITA is a subreddit seeking to give access to crowdsourced social judgement to clarify those sticky situations where we aren't quiite sure if we're being an A-hole.

The analysis is comprised of three main steps:

- Determining basic descriptive facts about the top posts
- Determining if more intense and balanced posts are more popular
- Determining if more polarizing posts more popular

Eventually, we seek to extend this analysis by exploring how post characteristics (e.g. age of poster, family vs relationship content) may impact community responses to determine which biases manifest in this social judgement context. Such an analysis would build off the analysis in Alice Wu 2019 (here: https://scholar.harvard.edu/files/alicewu/files/wu_ejr_paper_2019.pdf) and Ferrer et al 2020 (here: https://arxiv.org/pdf/2008.02754.pdf).

Basic Descriptive Facts

We want to explore the overall distributions of the key variables in this analysis, and confirm that the data adhere to our expectations. We focus this analysis on comments/replies, score, intensity, and balance.

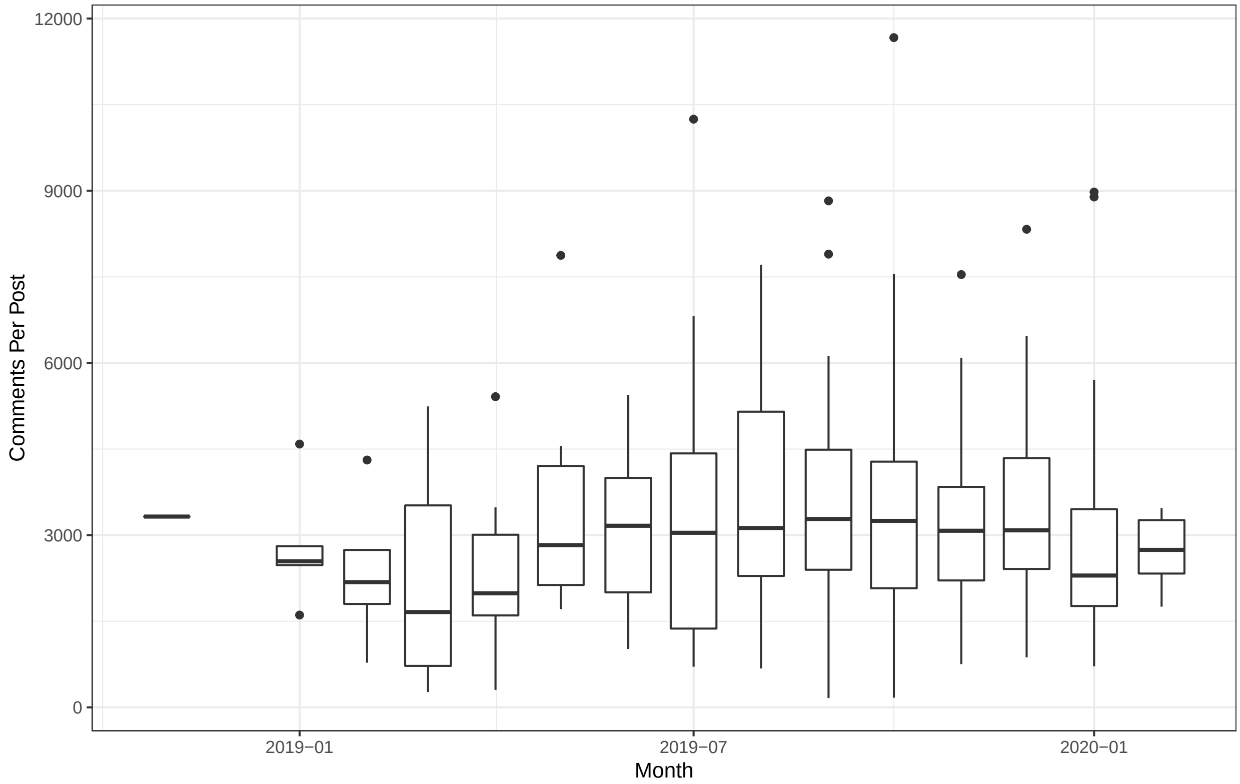
- Comments/replies refers to the number of comments or replies that a given post or comment receives.
- Score refers to the net upvotes a post or comment receives.
- Intensity is the ratio of the sum of words from 8 emotions to total words in a given post or comment. The values are calculated based on matching the words in the post or comment to the NRC dictionary (Saif Mohammad's NRC Emotion lexicon, see http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm). A post with a higher intensity has more emotionally laden words.
- Balance is the ratio of the positive valance value to the sum of the positive and negative valance values. Again, the values are derived from the NRC dictionary. A balanced post will have a balance of 0.5, indicating that there are as many positive words as there are negative words.

Overall

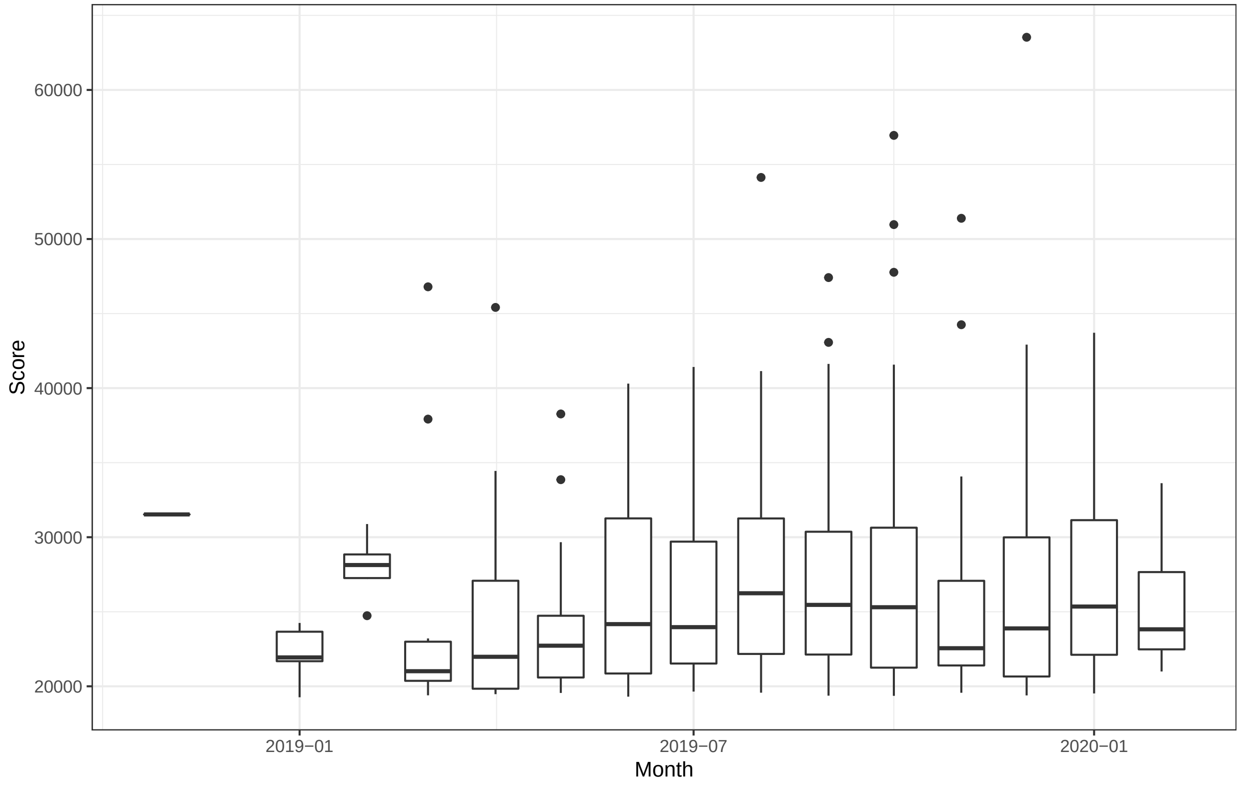
First, we consider the data posts overall by checking if censoring over time is a big driver of comment counts or score. The following plots indicate that censoring isn't a driving issue.

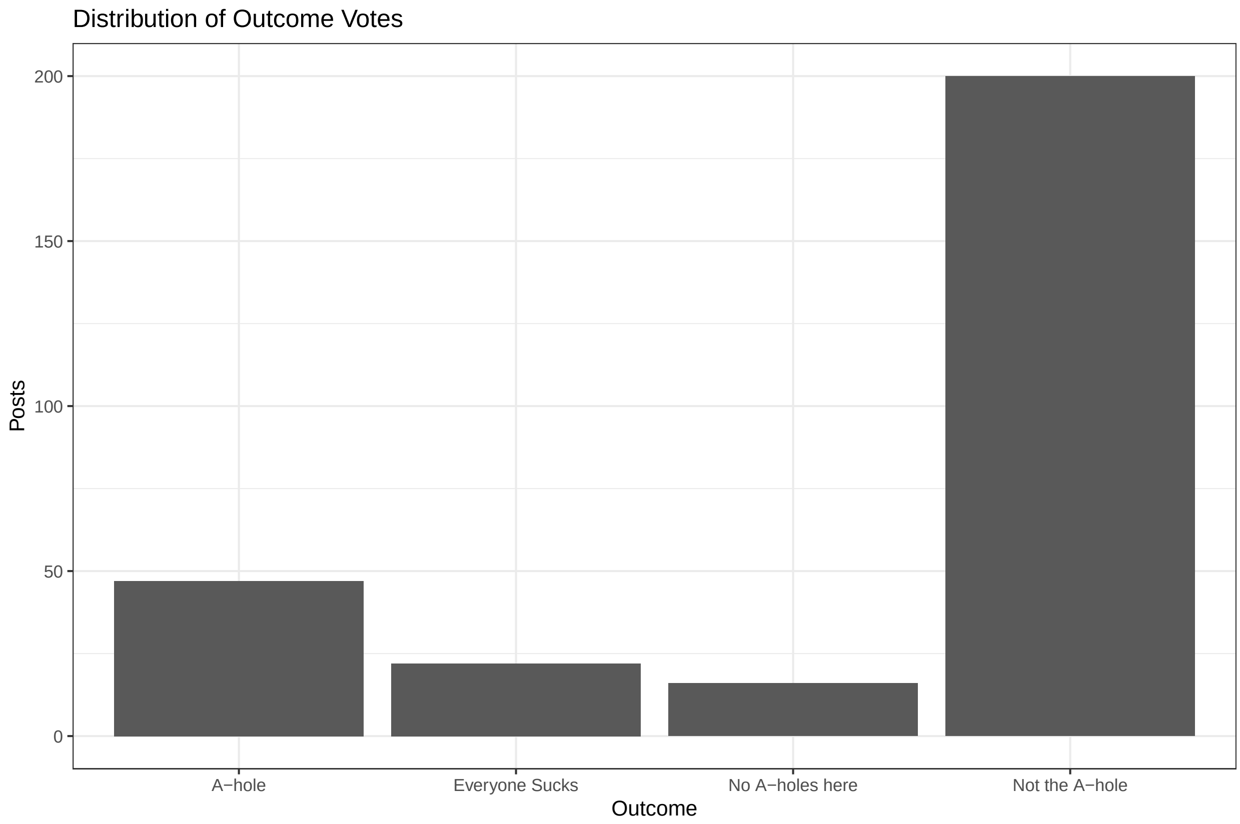
We also note that the vast majority of top posts are "Not the A-hole."

Distribution of Comments per Post over time



Distribution of Score over time

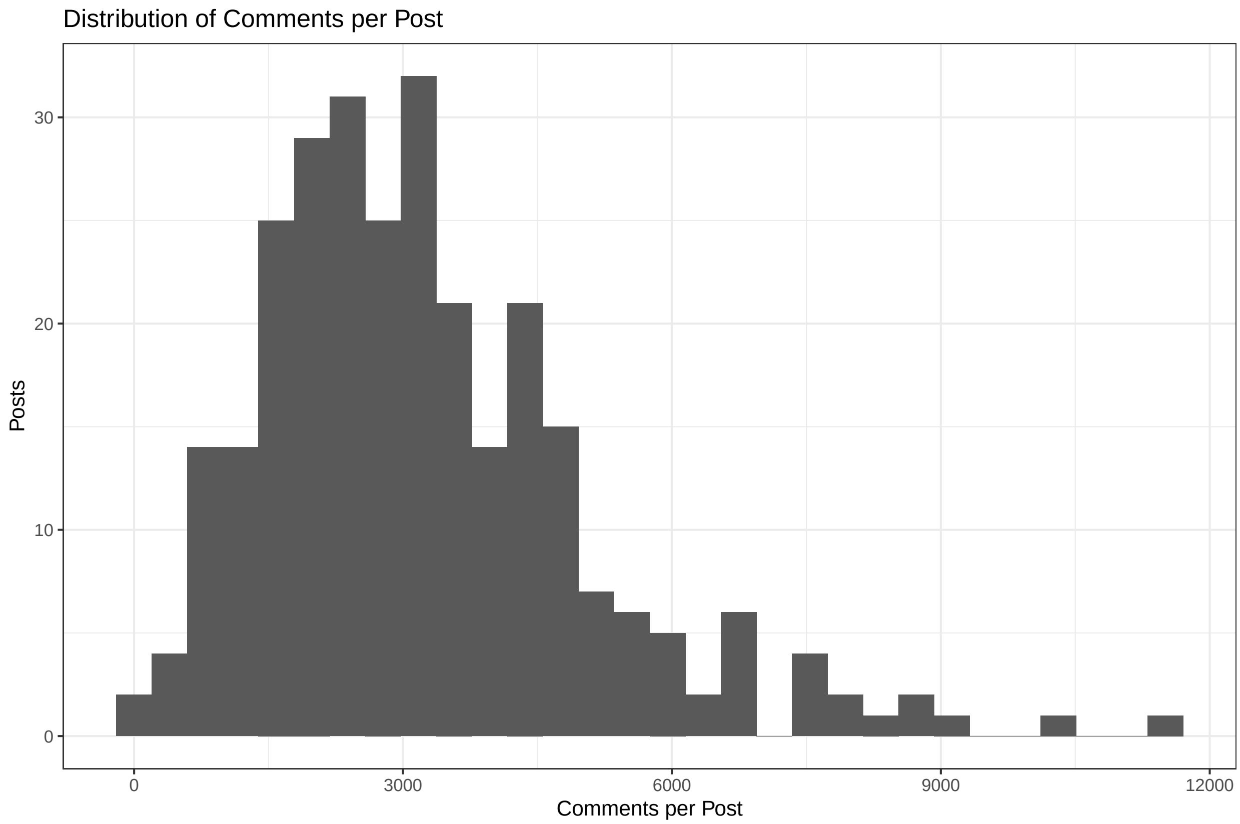


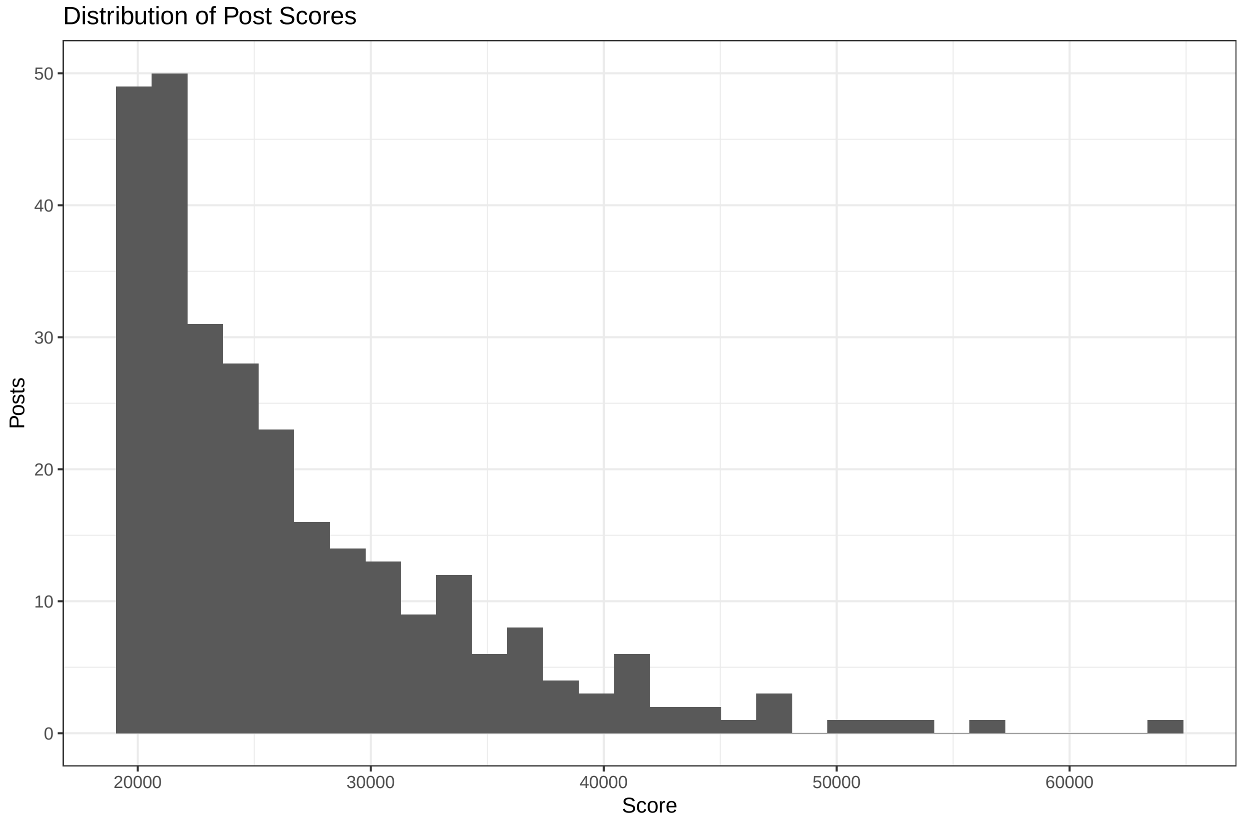


Distributions of key variables in Posts

We consider the distributions of post comments, score, intensity and balance to identify outliers or observations that should be dropped.

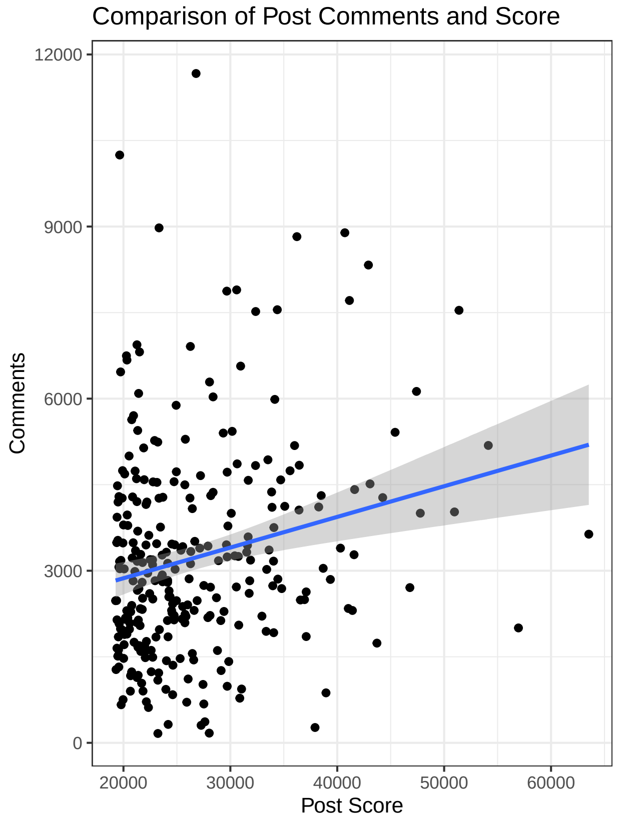
We don't see anything unusual among the non-deleted posts.

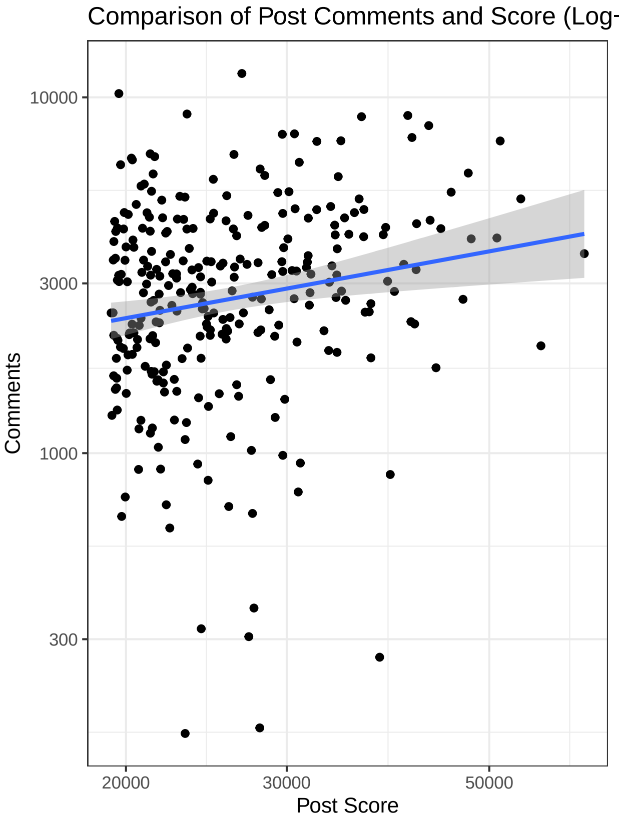




Having checked the marginal distributions of comments and score, we also want to consider the joint distribution.

For both level-level and log-log, comments and score are correlated which is as we might expect. A 1 point increase in score is associated with a 0.05 increase in comments, and a 1 percent increase in score is associated with a 0.47 percent increase in comments.



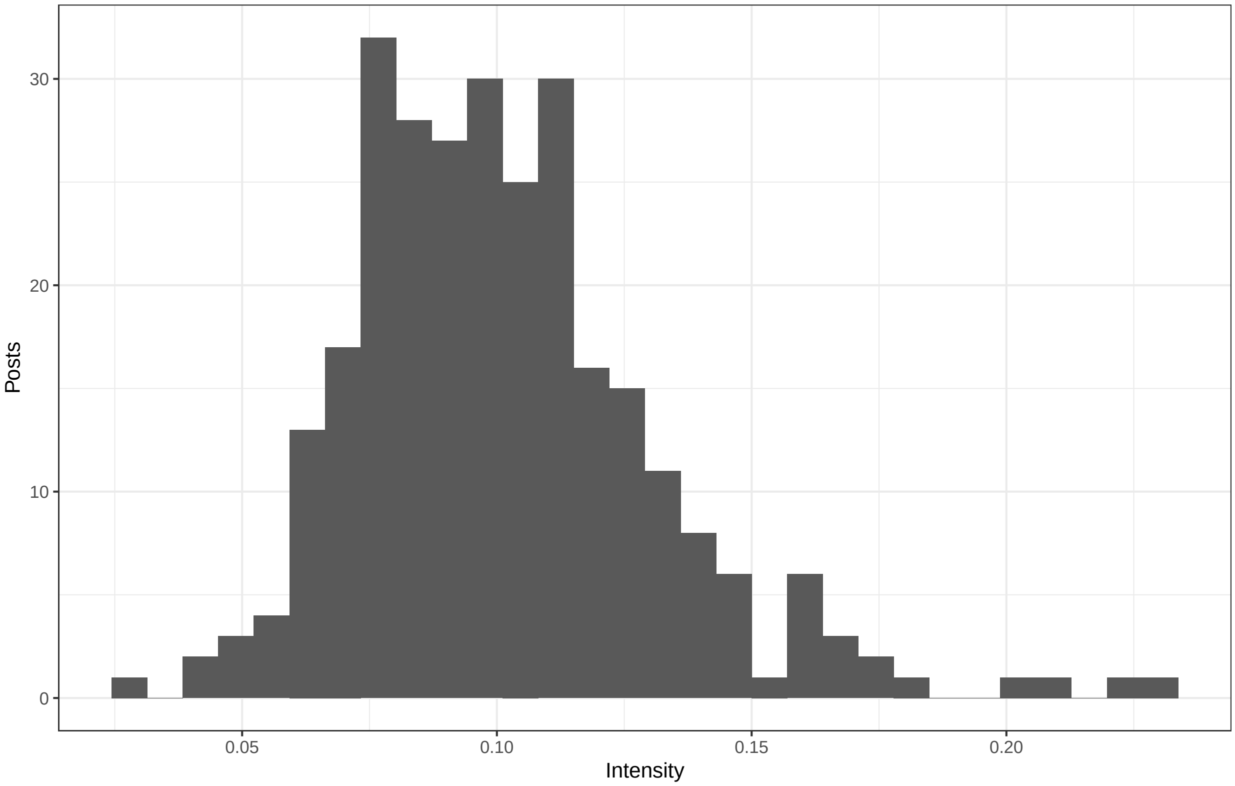


```
##                           lev_lev          log_log
## Dependent Var.:       num_comments log(num_comments)
##
## (Intercept)      1,799.8*** (394.9)    3.111. (1.581)
## score            0.0535*** (0.0142)
## log(score)                          0.4716** (0.1555)
## _______________ __________________ _________________
## S.E. type                      IID               IID
## Observations                   285               285
## R2                         0.04766           0.03148
## Adj. R2                    0.04429           0.02805
```

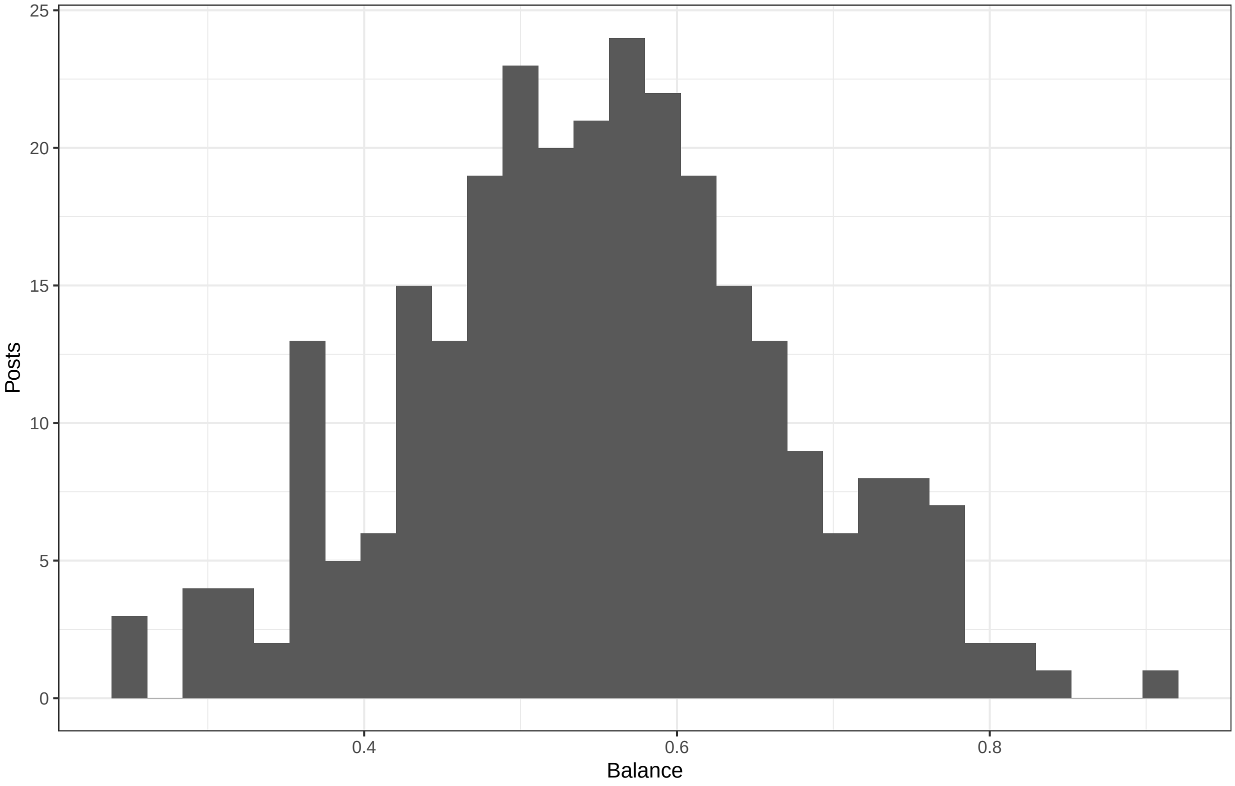
Lastly, we want to check the distributions of intensity and balance.

We observe that intensity is somewhat right skewed, so most posts tend to be less intense than the most extreme posts.

We observe that balance is fairly evenly distributed, but is centered above 0.5. This indicates that balance varies by post, but tends to be a bit more positive than negative.



Distribution of Post Intensity

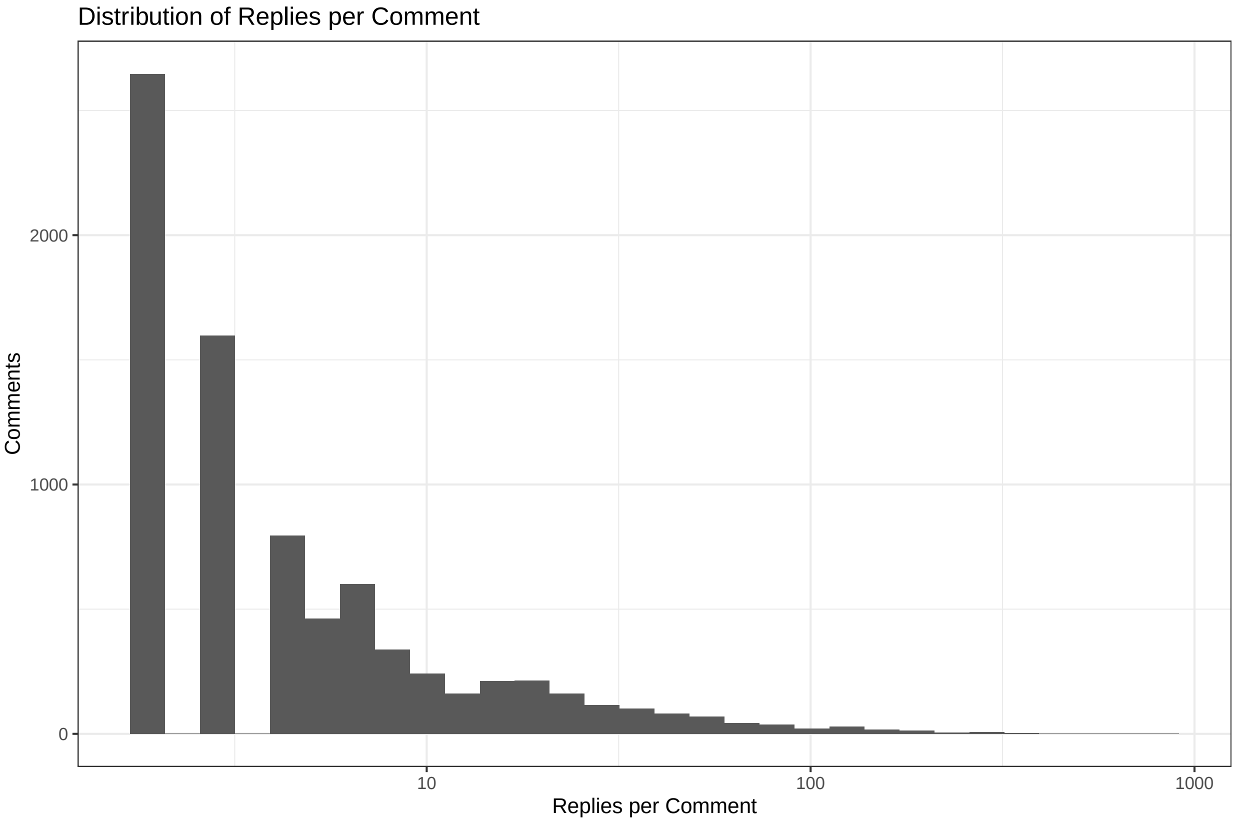


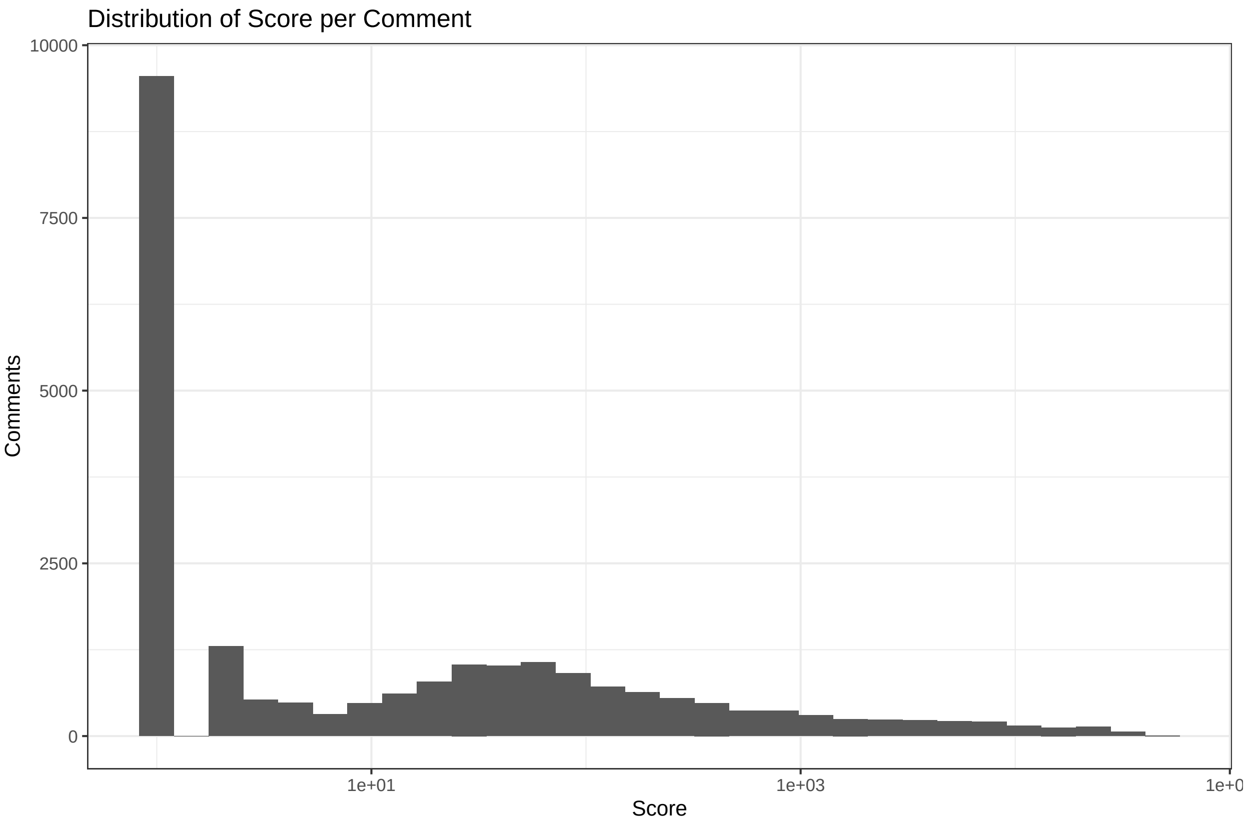
Distribution of Post Balance

Distributions of key variables in Comments

We consider the distribution of comment replies, score, intensity and balance to identify outliers or observations that should be dropped.

We notice that an enormous share of comments have 1 upvote. Since this may be a self-voted value and is therefore unrelated to a replies impact on other people, we don't consider single upvote comments when investigating comment scores.

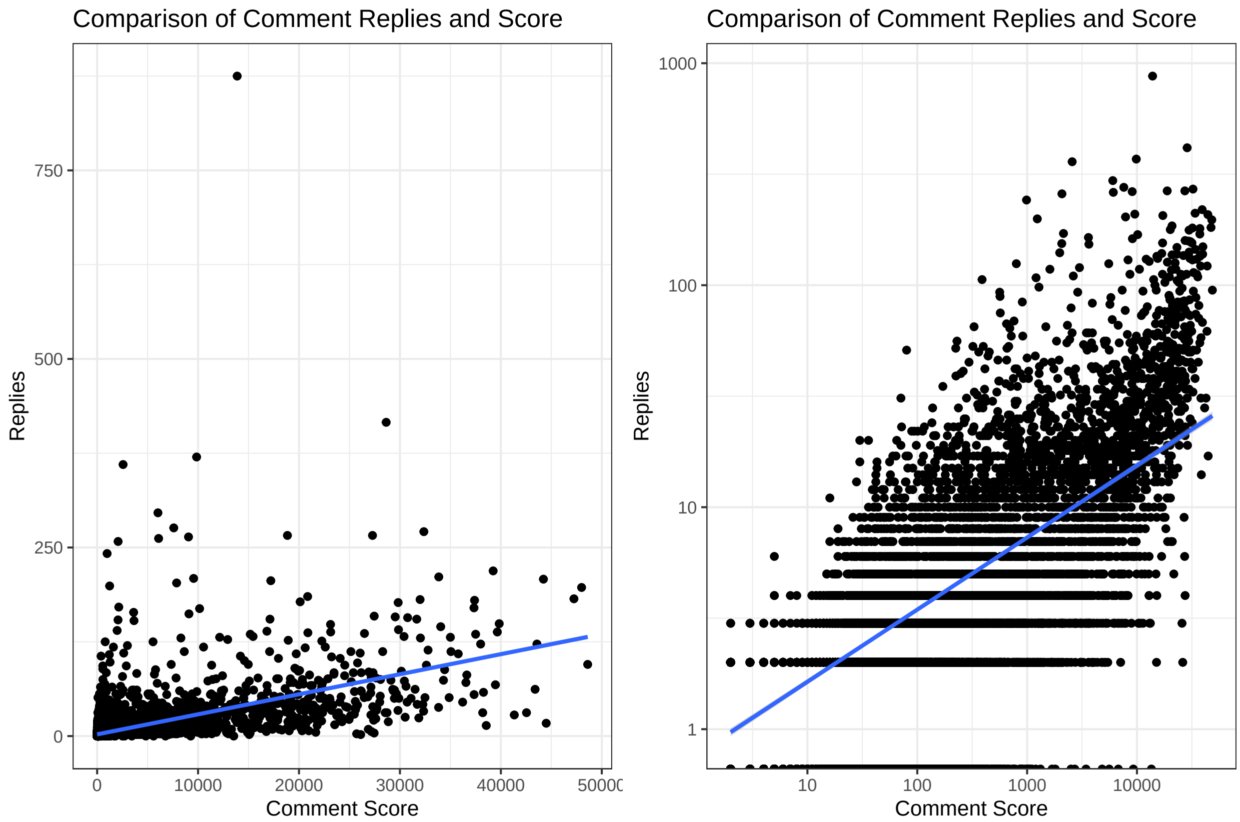




Having checked the marginal distributions of replies and score, we also want to consider the joint distribution.

For both level-level and log-log, replies and score are correlated which is as we might expect. A 1 point increase in score is associated with a 0.003 increase in replies, and a 1 percent increase in score is associated with a 0.32 percent increase in comments.

Including post fixed effects, we observe similar correlations: 0.003 and 0.34 respectively.


Comparison of Comment Replies and Score

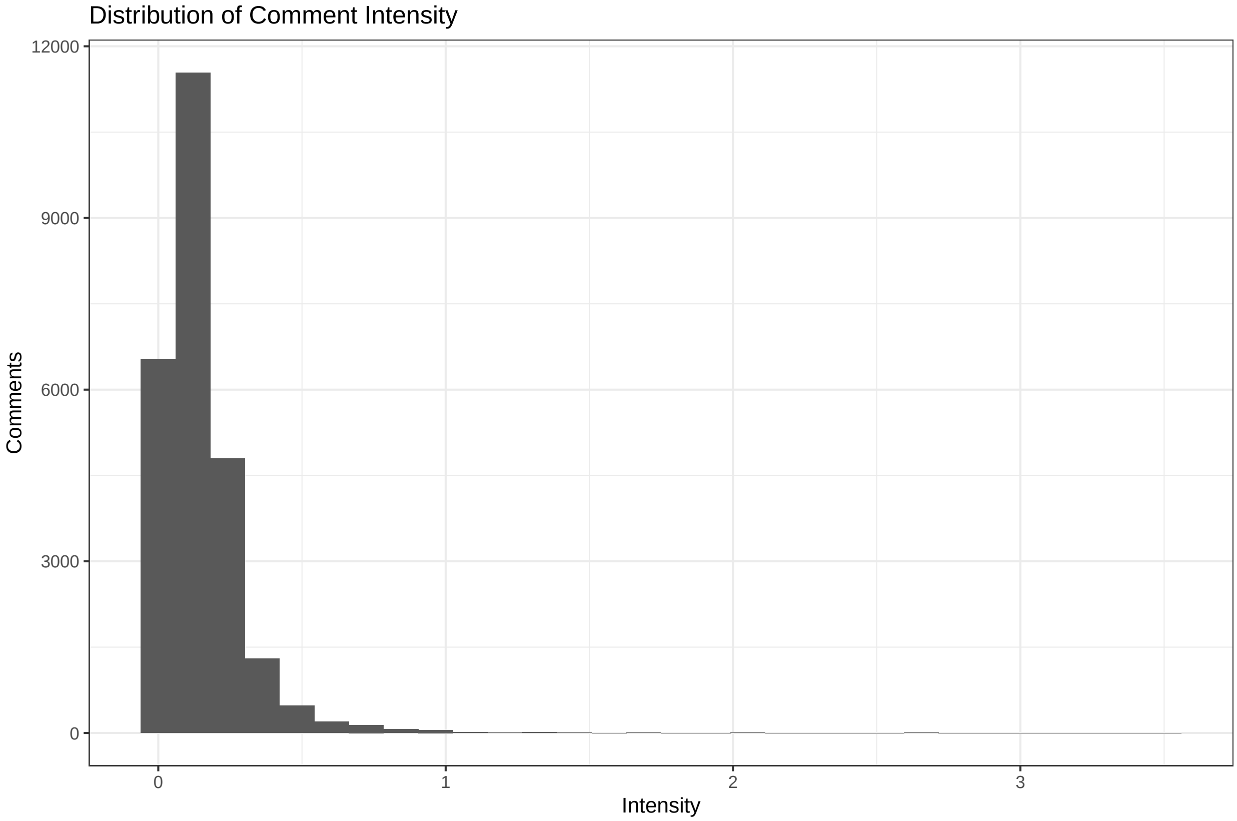
Comparison of Comment Replies and Score

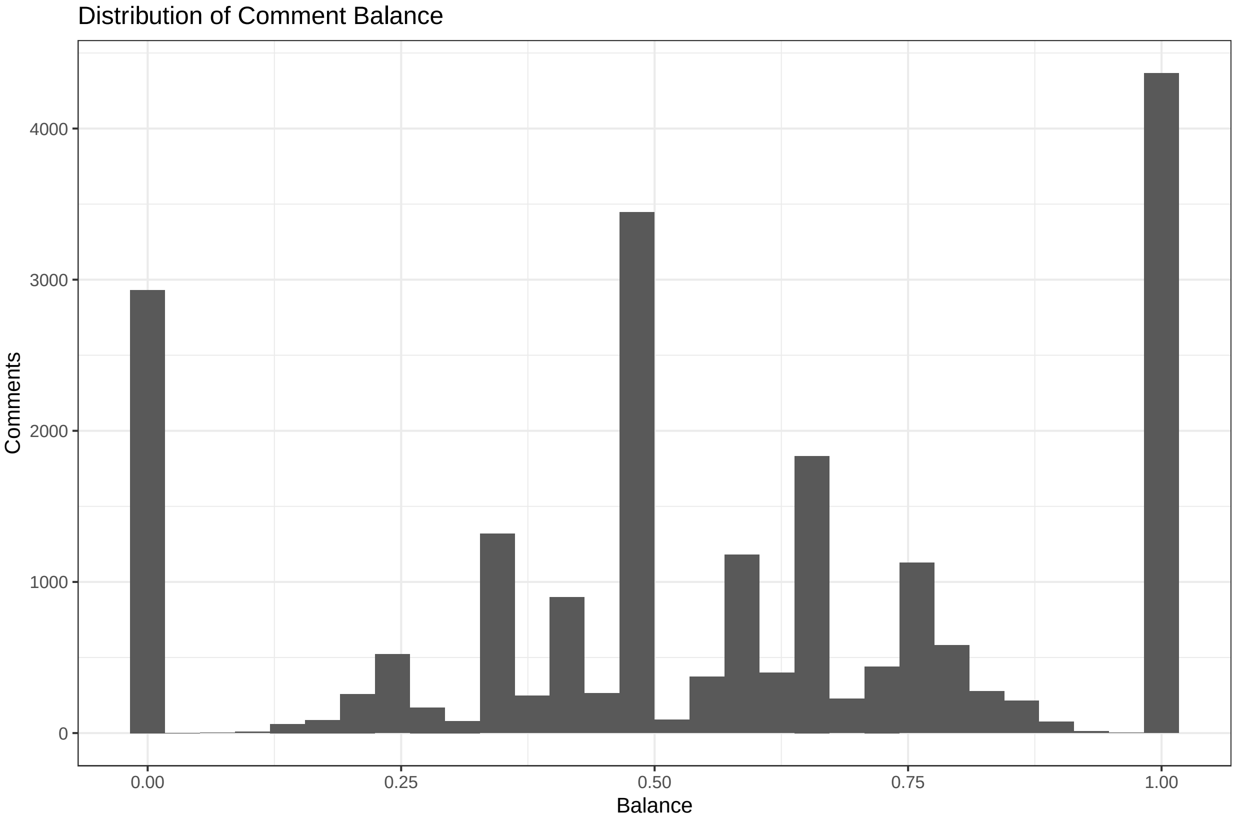
```
##                                    lev_lev                  log_log
## Dependent Var.:    reply_count_comment log(reply_count_comment)
##
## (Intercept)          2.276*** (0.1315)      -0.2562*** (0.0232)
## score_comment      0.0027*** (3.16e-5)
## log(score_comment)                           0.3249*** (0.0039)
## Fixed-Effects:     ------------------- ------------------------
## id                                  No                       No
## __________________ ___________________ ________________________
## S.E. type                          IID                      IID
## Observations                    13,550                    7,545
## R2                             0.34276                  0.47560
## Within R2                           --                       --
##                        lev_lev_post_fe          log_log_post_fe
## Dependent Var.:    reply_count_comment log(reply_count_comment)
##
## (Intercept)
## score_comment        0.0026*** (0.0001)
## log(score_comment)                           0.3352*** (0.0069)
## Fixed-Effects:     ------------------- ------------------------
## id                                 Yes                      Yes
## __________________ ___________________ ________________________
## S.E. type                       by: id                   by: id
## Observations                    13,550                    7,545
## R2                             0.39341                  0.56281
## Within R2                      0.35060                  0.51896
```

Lastly, we want to check the distributions of intensity and balance.

We observe that intensity ranges greatly and is severely right skewed. Most comments tend to be very unintense, but some are very intense. Note that values above 1 come from comments with words that appear in multiple emotions.

We observe that balance is fairly evenly distributed, but is concentrated at 0, 0.5, and 1, as well as 1/3, 2/3, 1/4, 3/4, and other fractions. This is because most comments are much shorter than posts, and so often have few if any valance (positive or negative) words.





Post Intensity, Balance and Popularity

We now turn to the relationship between post intensity and popularity. We expect that more intense posts are likely to be more popular since many forum posters don't engage unless moved emotionally. Post intensity measures the emotional impact of a post, so we expect more intense posts to correspond to more engaging posts.

We also consider whether balanced or unbalanced posts are more popular. On the one hand, balanced posts are likely to be more moderate in tone and thereby less engaging. On the other hand, unbalanced posts may be alienating or unambiguous, rendering them uninteresting.

```
## Intensity and Popularity ----
```

Intensity and Popularity

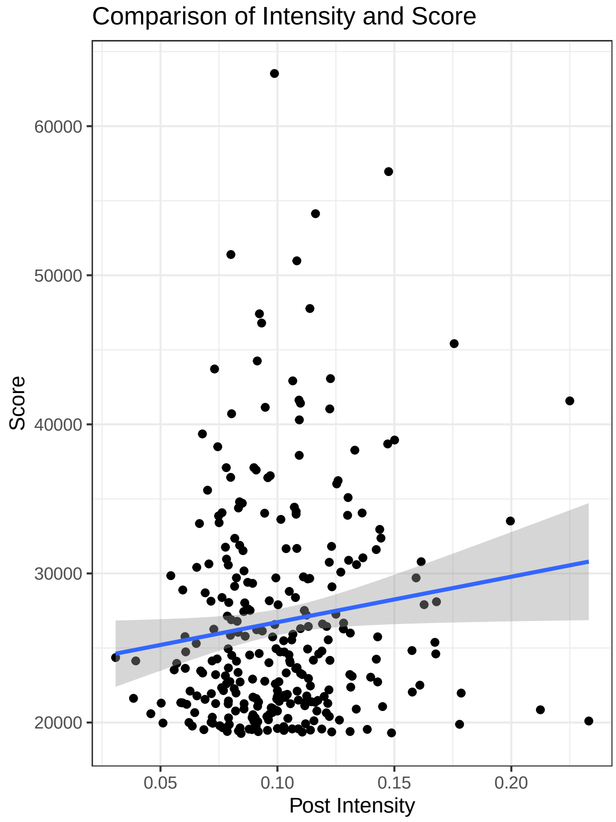
We measure intensity, as described above, based on the share of words in a post that correspond to emotional responses in the NRC emotion lexicon.

We run a regression of post score on intensity, and find some correlation but no obvious trend. Similarly, post comments appear mostly unrelated to post intensity. As a result, the role of intensity and score remains inconclusive.

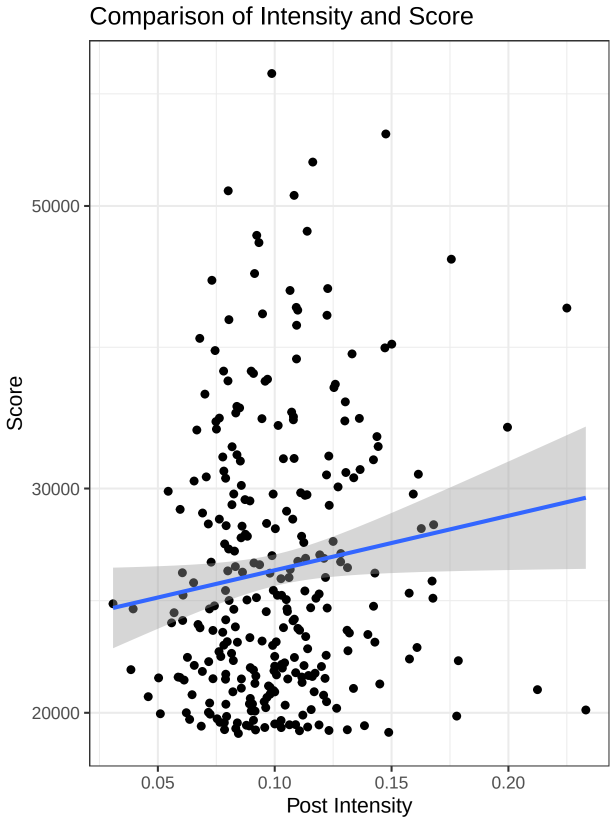
One interesting takeaway is that the least intense posts (intensity<=0.06) are all low score/comments, while the some of the highest intensity posts have high scores and many commments (intensity>=0.15). Additionally, almost all the very high score/comment posts are in the middle intenstiy range.

Perhaps there is some negative engagement response associated with overly bland or intense posts, though with the current lack of correlation it's hard to say.

Further analysis with a larger sample size and evaluation of specific emotions may give more insight into the relationship of emotional intensity and post popularity.

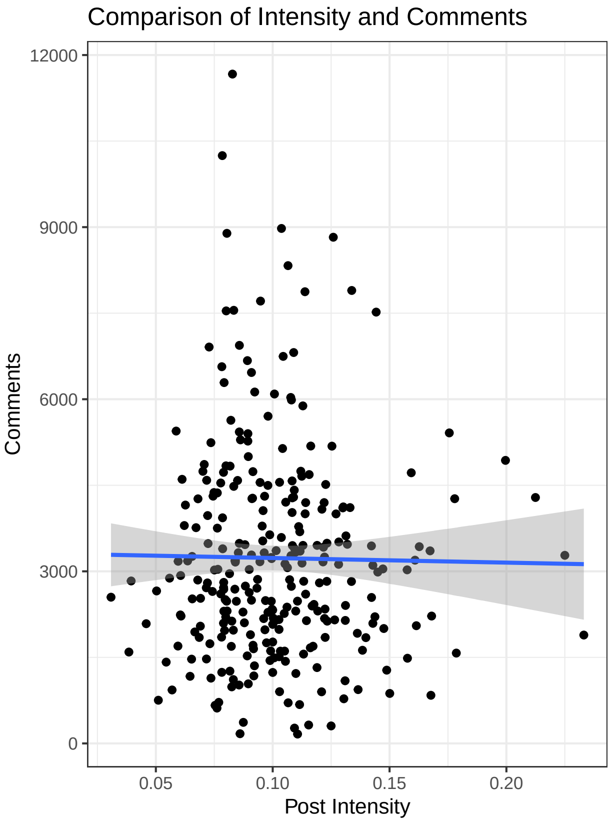


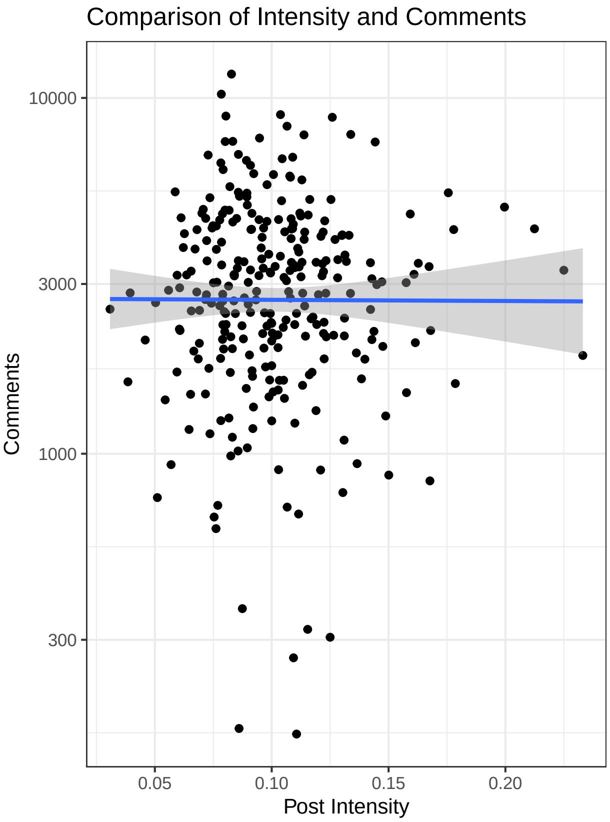
Comparison of Intensity and Score



Comparison of Intensity and Score

```
##                                lev_lev          log_lev
## Dependent Var.:                  score       log(score)
##
## (Intercept)     23,678.3*** (1,558.8) 10.06*** (0.0511)
## intensity        30,513.7* (14,754.3)  0.9851* (0.4834)
## _______________ _____________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.01489           0.01446
## Adj. R2                       0.01141           0.01098
```





```
##                           lev_lev           log_lev
## Dependent Var.:       num_comments log(num_comments)
##
## (Intercept)     3,312.6*** (384.6) 7.911*** (0.1368)
## intensity         -804.5 (3,640.5)   -0.0753 (1.294)
## _______________ __________________ _________________
## S.E. type                      IID               IID
## Observations                   285               285
## R2                         0.00017            1.2e-5
## Adj. R2                   -0.00336          -0.00352
```

Balance and Popularity

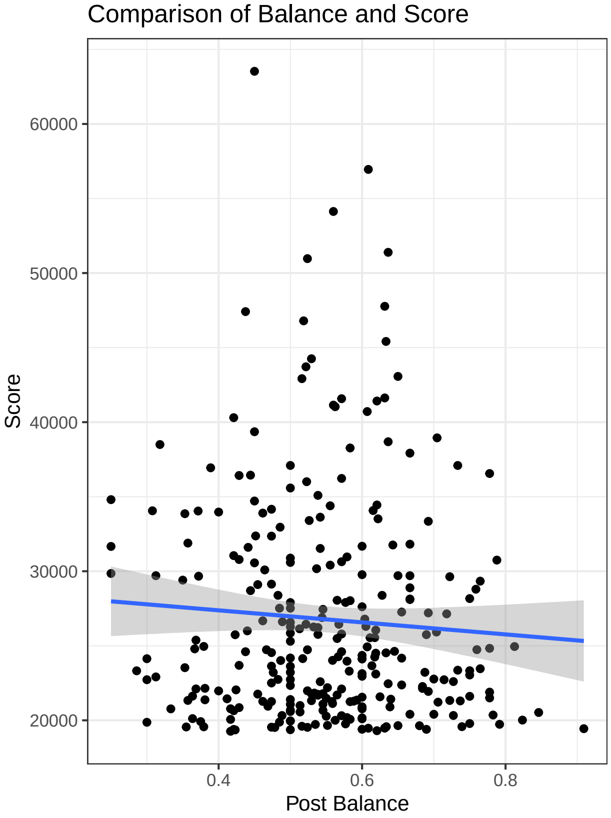
We measure balance, as described above, based on the share of words in a post that correspond to a valance response in the NRC emotion lexicon.

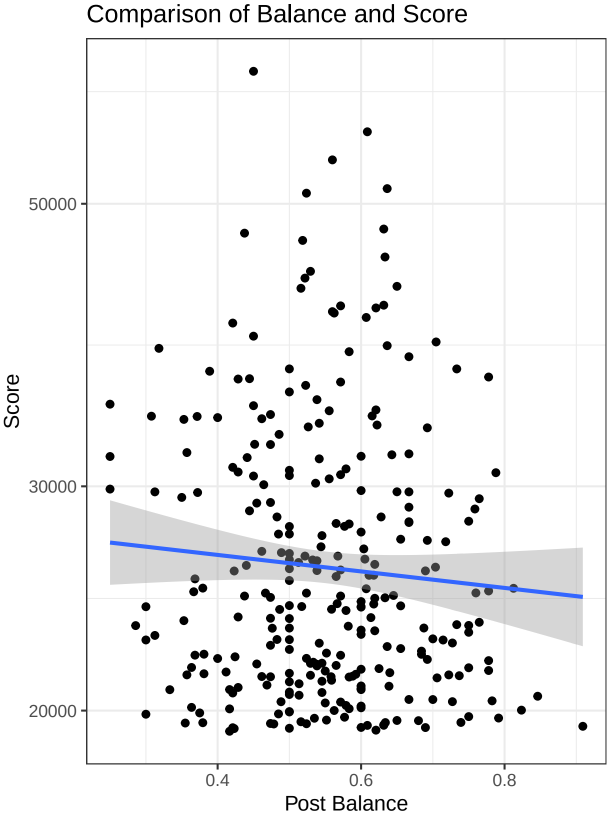
We run a regression of post score on balance, and find some correlation but no obvious trend. Similarly, post comments appear mostly unrelated to post balance. As a result, the role of balance and score remains inconclusive.

One interesting takeaway is that almost all the most positive posts have low scores, while very negative posts have more positive scores. Additionally, nearly all the highest scored/most commented as well as the least commented posts are in the middle range of balance.

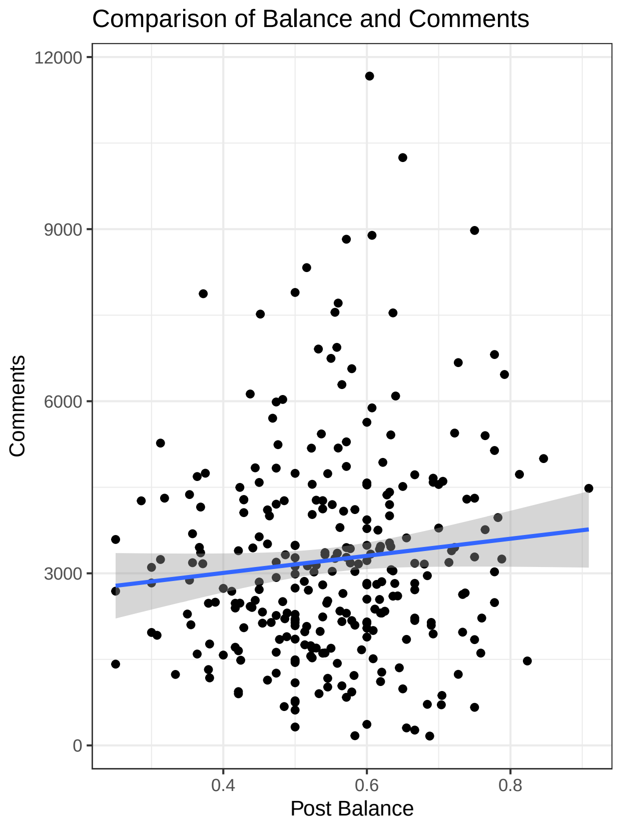
Perhaps there is some negative engagement response associated with overly positive or negative posts, though with the current lack of correlation it's hard to say.

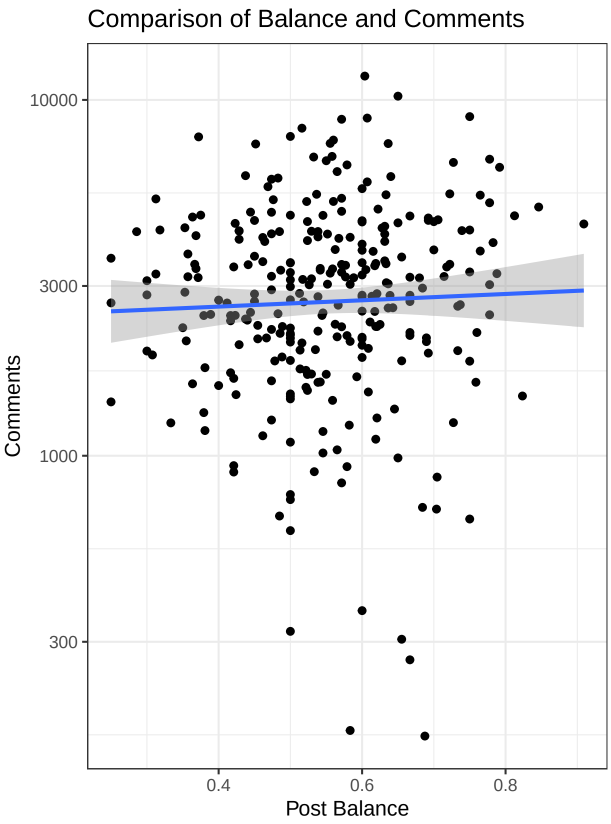
Further analysis with a larger sample size may give more insight into the relationship of emotional balance and post popularity.





```
##                               lev_lev           log_lev
## Dependent Var.:                 score        log(score)
##
## (Intercept)     28,991.0*** (2,062.5) 10.24*** (0.0675)
## balance            -4,034.0 (3,658.4)  -0.1491 (0.1198)
## _______________ _____________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.00428           0.00545
## Adj. R2                       0.00076           0.00193
```





```
##                              lev_lev           log_lev
## Dependent Var.:       num_comments log(num_comments)
##
## (Intercept)     2,410.5*** (503.8) 7.791*** (0.1798)
## balance           1,490.6. (893.6)   0.2043 (0.3190)
## _______________ __________________ _________________
## S.E. type                      IID               IID
## Observations                   285               285
## R2                         0.00974           0.00145
## Adj. R2                    0.00624          -0.00208
```

Polarization and Popularity

We now turn to the relationship between polarization and popularity. We expect that more polarizing posts are likely to be more popular since many forum posters don't engage unless moved emotionally. Comment polarization measures the emotional impact of a post, so we expect more polarization to correspond to more engaging posts.

While polarization may be a good measure of emotional engagement, other explanations may exist. For example, non-polarized posts from particularly humorous or outlandish posts may also excel on the platform. This analysis seeks to determine whether polarization is indeed associated with popularity on AITA, which gives some indication into whether the most engaging posts are divisive.

We measure polarization in two ways: emotional polarization of comments and voting breakdowns of comments.

```
## Emotional Polarization ----
```

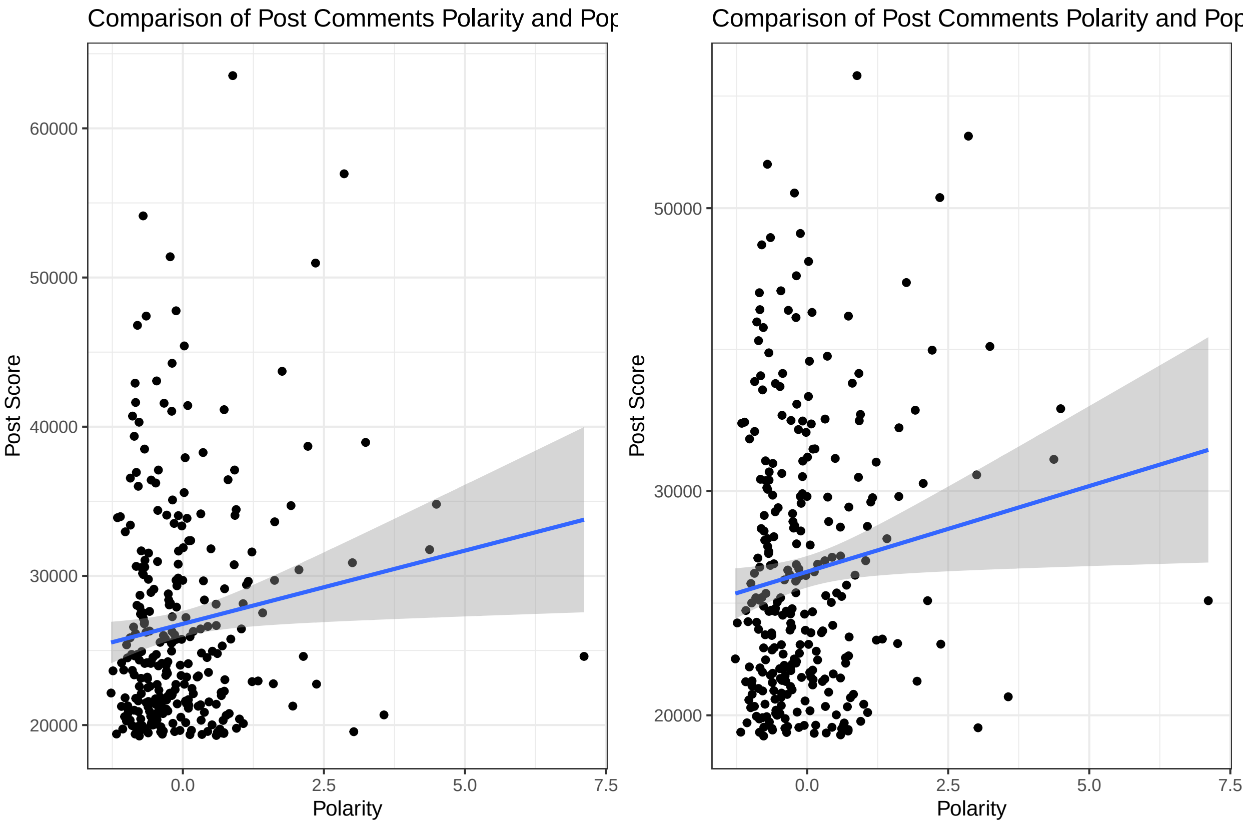
Emotional Polarization and Popularity

We construct a normalized index of polarization for each post based on intensity and balance of its comments. A post is highly polarized if there are many intense comments that disagree in terms of balance. We capture this polarity using standard deviation of the product of intensity and polarization.

We see some association between polarity and score. However, repeating the same analysis with number of comments at the measure of popularity indicates that more comments are negatively associated with polarity. This suggests that comments are a confounder for the effect of polarity on popularity. Conditioning on comments, we see that polarity is strongly associated with higher scores.

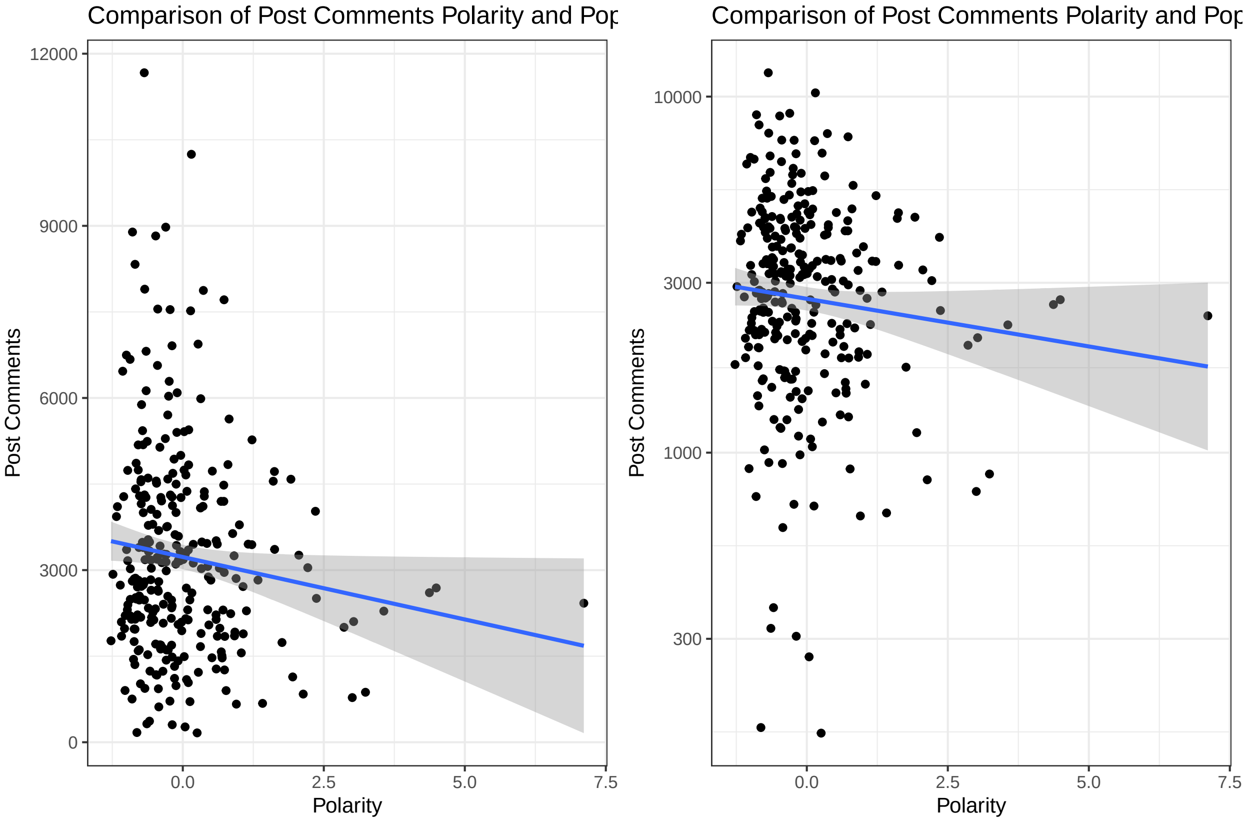
Since standard deviation may be affected by outliers and sample sizes we try two robustness checks: removing the top/bottom 5% of polarizing comments and using IQR instead of standard deviation. These don't impact the results. See the appendix for the estimates.

Results of Polarity on Score



```
##                                lev_lev            log_lev
## Dependent Var.:                  score         log(score)
##
## (Intercept)     26,783.7*** (441.0) 10.16*** (0.0145)
## sd_norm              982.3* (438.3)  0.0310* (0.0144)
## --------------- ------------------- -----------------
## S.E. type                        IID              IID
## Observations                      285              285
## R2                            0.01744          0.01614
## Adj. R2                       0.01397          0.01267
```

Results of Polarity on Number of Comments



```
##                            lev_lev           log_lev
## Dependent Var.:       num_comments log(num_comments)
##
## (Intercept)     3,228.1*** (108.2) 7.903*** (0.0386)
## sd_norm            -217.6* (107.5)  -0.0616 (0.0383)
## _______________ __________________ _________________
## S.E. type                      IID               IID
## Observations                   285               285
## R2                         0.01426           0.00906
## Adj. R2                    0.01078           0.00555
```

Results of Polarity on Score, Controlling for Number of Comments

```
fixest::etable(lev_lev, log_lev)
```

```
##                                    lev_lev            log_lev
## Dependent Var.:                     score        log(score)
##
## (Intercept)       23,653.8*** (873.8) 9.594*** (0.1737)
## sd_norm             1,193.2** (429.5)  0.0354* (0.0142)
## num_comments       0.9696*** (0.2358)
## log(num_comments)                     0.0720** (0.0219)
## _________________ ___________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.07303           0.05239
## Adj. R2                       0.06645           0.04567
```

```
## Voting Polarization ----
```

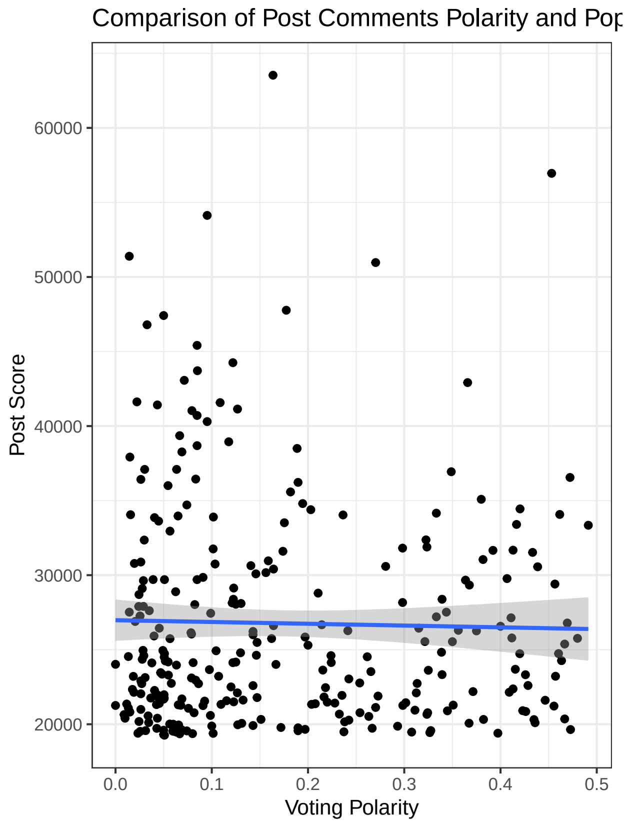
Voting Polarization and Popularity

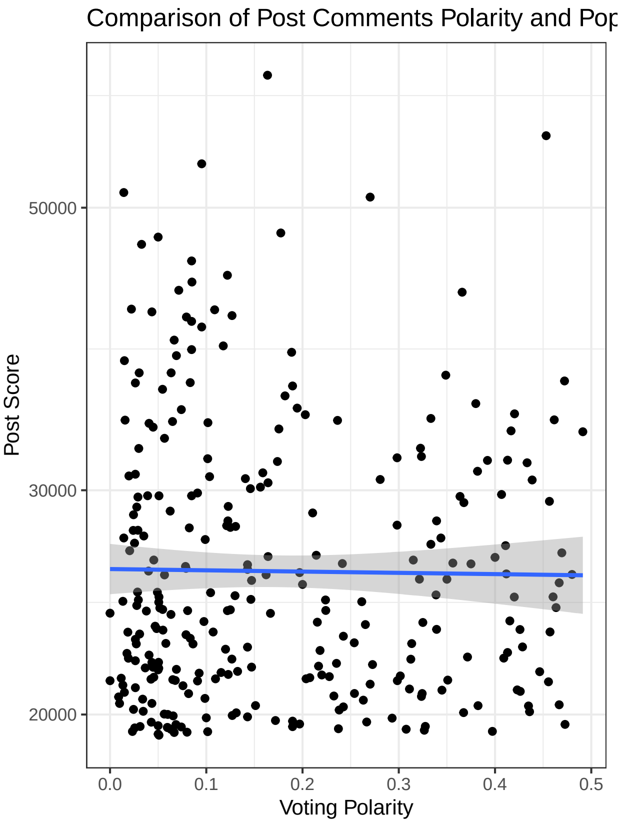
We construct an index of polarization for each post based on the share of votes that are NTA or YTA. A post is highly polarized if there are many votes that disagree.

Work in Progress…

Measure based on aggreement and YTA

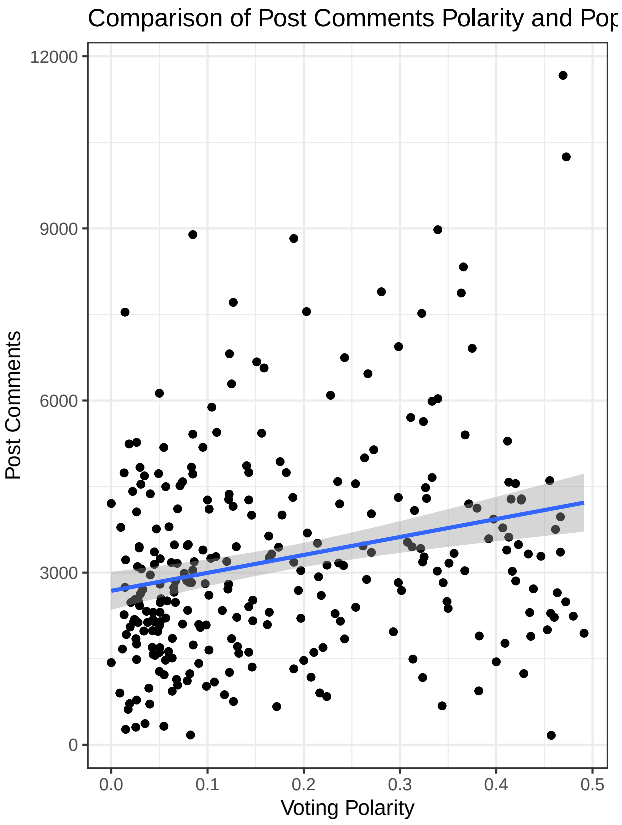
Results of Polarity on Score

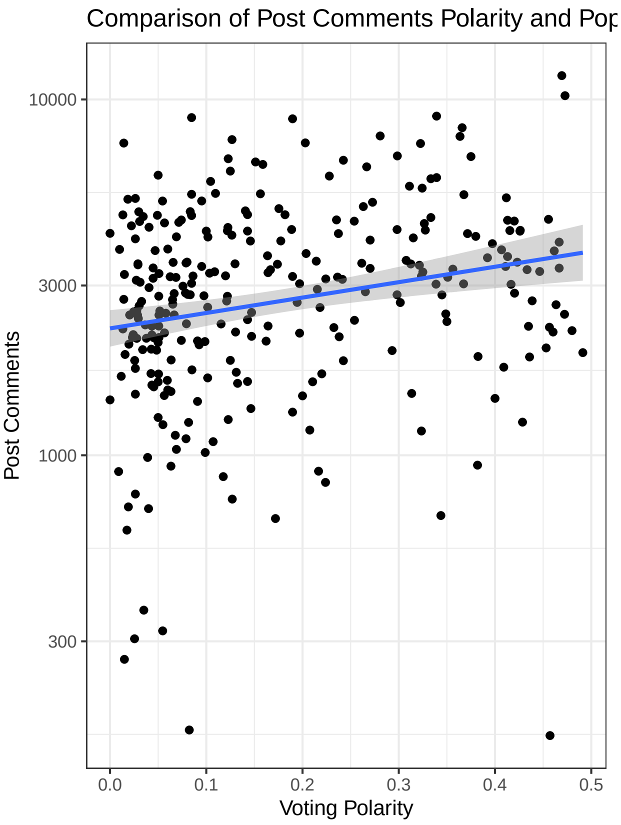




```
##                                 lev_lev           log_lev
## Dependent Var.:                    score        log(score)
##
## (Intercept)       26,980.0*** (705.1) 10.17*** (0.0231)
## vote_polarization  -1,197.2 (3,119.9)  -0.0228 (0.1022)
## _________________ ___________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.00052           0.00018
## Adj. R2                      -0.00301          -0.00336
```

Results of Polarity on Score





```
##                              lev_lev            log_lev
## Dependent Var.:         num_comments  log(num_comments)
##
## (Intercept)       2,682.7*** (167.5)  7.729*** (0.0599)
## vote_polarization 3,127.4*** (741.4) 0.9971*** (0.2652)
## _________________ __________________ __________________
## S.E. type                        IID                IID
## Observations                     285                285
## R2                           0.05916            0.04758
## Adj. R2                      0.05583            0.04422
```

Results of Polarity on Score, Controlling for Number of Comments

```
fixest::etable(lev_lev, log_lev)
```

```
##                                lev_lev           log_lev
## Dependent Var.:                  score        log(score)
##
## (Intercept)       24,373.9*** (948.8) 9.616*** (0.1758)
## vote_polarization  -4,235.4 (3,135.2)  -0.0939 (0.1031)
## num_comments       0.9715*** (0.2438)
## log(num_comments)                     0.0712** (0.0226)
## _________________ ___________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.05378           0.03431
## Adj. R2                       0.04707           0.02746
```

```
##################)
# Appendix ----
##################)
```

Appendix

Emotional Polarity Robustness

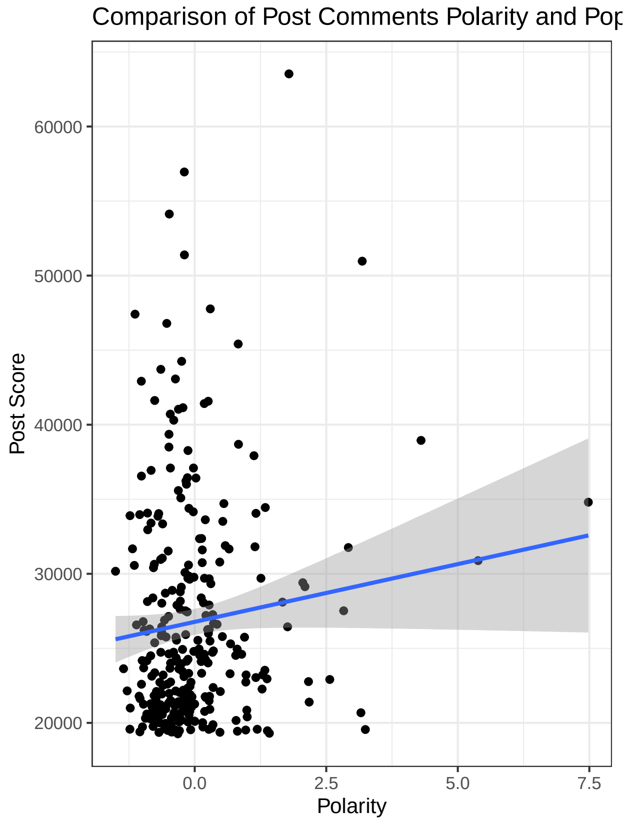
```
post_polar = top_posts[
  ,`:=`(quant_min = quantile(intensity_comment*balance_comment, .05, na.rm = T)
        , quant_max = quantile(intensity_comment*balance_comment, .95, na.rm = T))
  , id
][
  between(intensity_comment*balance_comment, quant_min, quant_max)
  , .(iqr_ind = IQR(intensity_comment*balance_comment, na.rm = T)
      , sd_ind = sd(intensity_comment*balance_comment, na.rm = T)
      , score = first(score)
      , num_comments = first(num_comments))
  , .(id, selftext, intensity, balance)
][
  , `:=`(iqr_norm = (iqr_ind - mean(iqr_ind))/sd(iqr_ind)
         , sd_norm = (sd_ind - mean(sd_ind))/sd(sd_ind))
][
  , `:=`(iqr_shift = iqr_norm - min(iqr_norm)
         , sd_shift = sd_norm - min(sd_norm))
]

lev_lev_plot = post_polar %>%
  ggplot()+
  geom_point(aes(x = sd_norm,y=score))+
  geom_smooth(aes(x = sd_norm,y=score), method = "lm")+
  theme_bw()+
  labs(x = "Polarity", y = "Post Score"
       , title = "Comparison of Post Comments Polarity and Popularity")

log_lev_plot = post_polar %>%
  ggplot()+
  geom_point(aes(x = sd_norm,y=score))+
  geom_smooth(aes(x = sd_norm,y=score), method = "lm")+
  theme_bw()+
  labs(x = "Polarity", y = "Post Score"
       , title = "Comparison of Post Comments Polarity and Popularity")+
  scale_y_log10()

lev_lev = post_polar %>%
  fixest::feols(score~sd_norm)
```

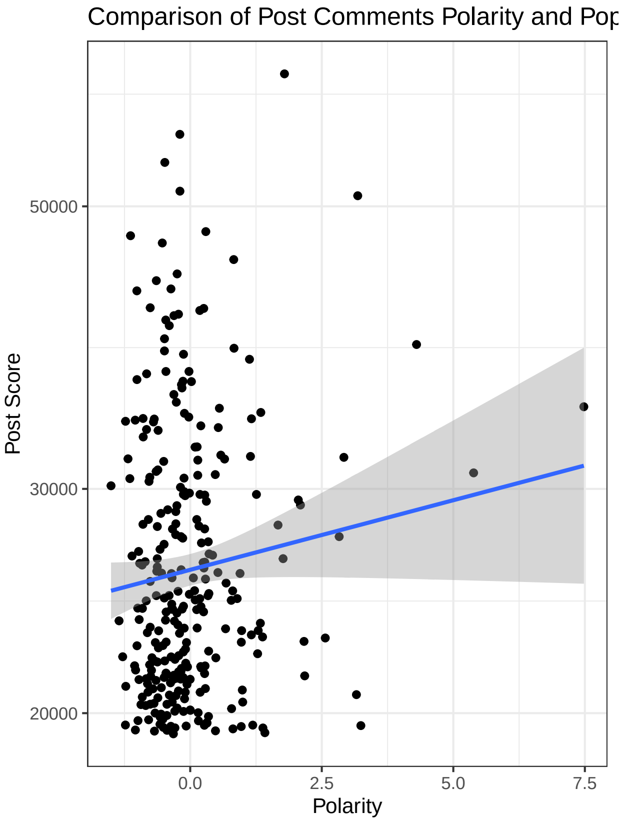
```
## NOTE: 15 observations removed because of NA values (LHS: 15).
```

```
log_lev = post_polar %>%
  fixest::feols(log(score)~sd_norm)
```

```
## NOTE: 15 observations removed because of NA values (LHS: 15).
```

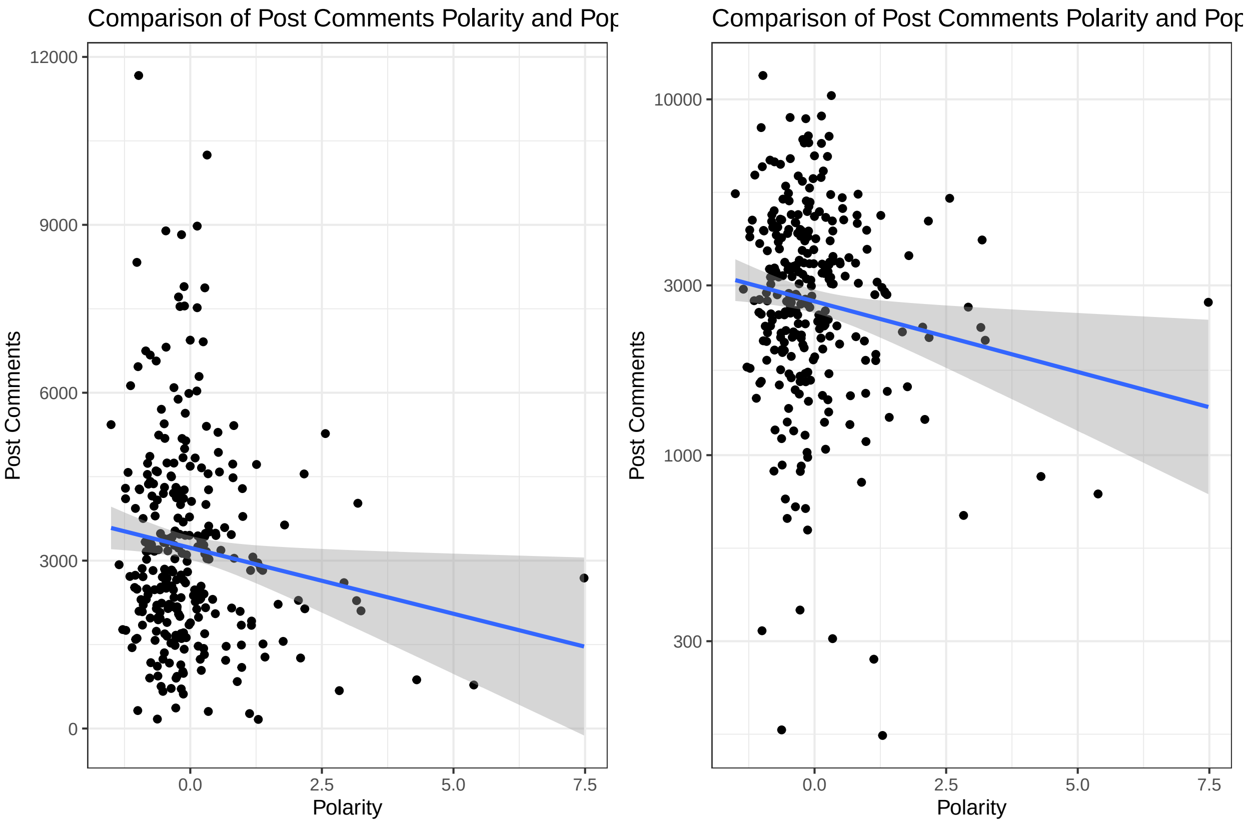
Results of Polarity on Score





```
##                             lev_lev           log_lev
## Dependent Var.:               score        log(score)
##
## (Intercept)     26,776.1*** (442.5) 10.16*** (0.0145)
## sd_norm              774.7. (437.8)  0.0251. (0.0143)
## _______________ ___________________ _________________
## S.E. type                       IID               IID
## Observations                    285               285
## R2                          0.01094           0.01075
## Adj. R2                     0.00745           0.00726
```

Results of Polarity on Number of Comments



```
##                            lev_lev           log_lev
## Dependent Var.:       num_comments log(num_comments)
##
## (Intercept)     3,229.3*** (108.0) 7.903*** (0.0384)
## sd_norm            -235.7* (106.9) -0.0913* (0.0379)
## _______________ __________________ _________________
## S.E. type                      IID               IID
## Observations                   285               285
## R2                         0.01689           0.02007
## Adj. R2                    0.01342           0.01660
```

Results of Polarity on Score, Controlling for Number of Comments

```
fixest::etable(lev_lev, log_lev)
```

```
##                               lev_lev            log_lev
## Dependent Var.:                 score         log(score)
##
## (Intercept)       23,665.8*** (878.3)  9.580*** (0.1751)
## sd_norm              1,001.7* (429.9)   0.0319* (0.0142)
## num_comments       0.9631*** (0.2370)
## log(num_comments)                     0.0737*** (0.0221)
## _________________ ___________________ __________________
## S.E. type                         IID                IID
## Observations                      285                285
## R2                            0.06565            0.04841
## Adj. R2                       0.05902            0.04166
```

```
post_polar = top_posts[
  ,`:=`(quant_min = quantile(intensity_comment*balance_comment, .05, na.rm = T)
        , quant_max = quantile(intensity_comment*balance_comment, .95, na.rm = T))
  , id
][
  # between(intensity_comment*balance_comment, quant_min, quant_max)
  , .(iqr_ind = IQR(intensity_comment*balance_comment, na.rm = T)
      , sd_ind = sd(intensity_comment*balance_comment, na.rm = T)
      , score = first(score)
      , num_comments = first(num_comments))
  , .(id, selftext, intensity, balance)
][
  , `:=`(iqr_norm = (iqr_ind - mean(iqr_ind))/sd(iqr_ind)
         , sd_norm = (sd_ind - mean(sd_ind))/sd(sd_ind))
][
  , `:=`(iqr_shift = iqr_norm - min(iqr_norm)
         , sd_shift = sd_norm - min(sd_norm))
]

lev_lev_plot = post_polar %>%
  ggplot()+
  geom_point(aes(x = iqr_norm,y=score))+
  geom_smooth(aes(x = iqr_norm,y=score), method = "lm")+
  theme_bw()+
  labs(x = "Polarity", y = "Post Score"
       , title = "Comparison of Post Comments Polarity and Popularity")

log_lev_plot = post_polar %>%
  ggplot()+
  geom_point(aes(x = iqr_norm,y=score))+
  geom_smooth(aes(x = iqr_norm,y=score), method = "lm")+
  theme_bw()+
  labs(x = "Polarity", y = "Post Score"
       , title = "Comparison of Post Comments Polarity and Popularity")+
  scale_y_log10()
```

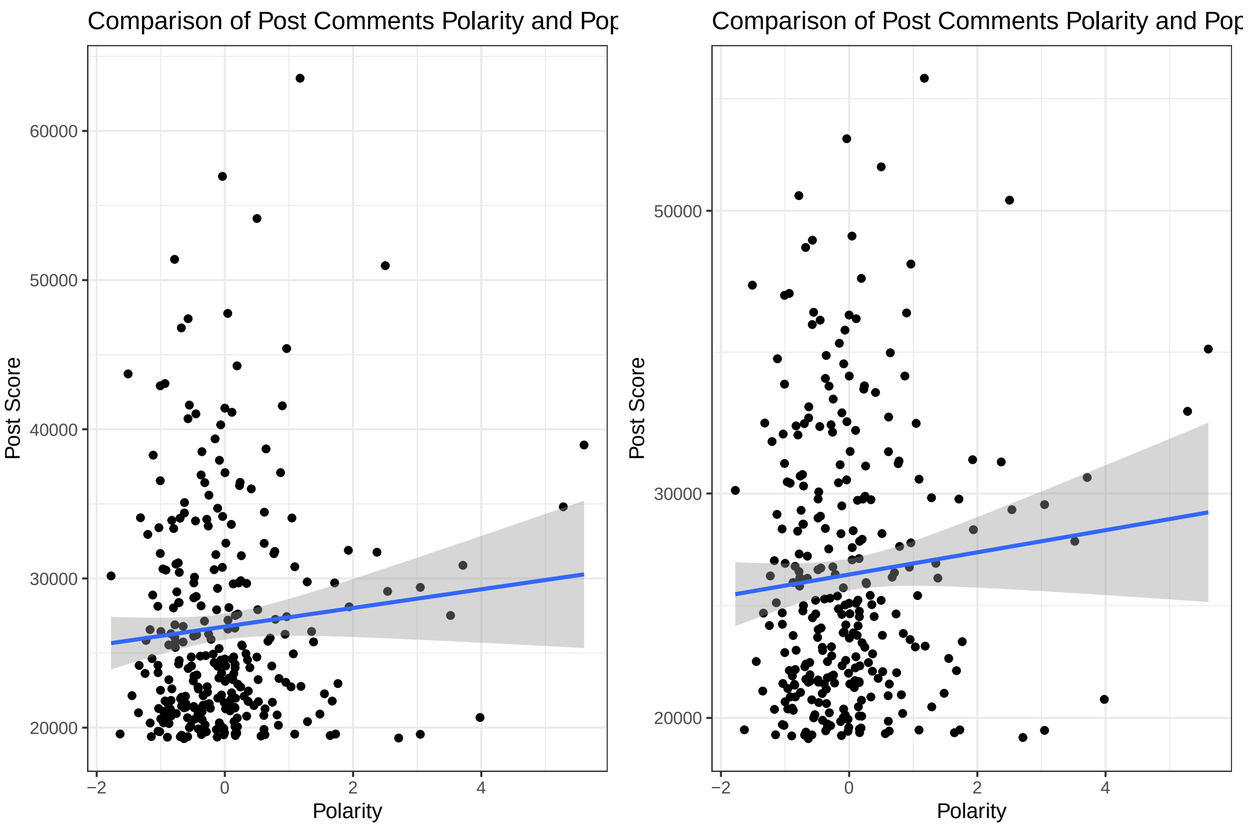
```
lev_lev = post_polar %>%
  fixest::feols(score~iqr_norm)
```

```
## NOTE: 15 observations removed because of NA values (LHS: 15).
```

```
log_lev = post_polar %>%
  fixest::feols(log(score)~iqr_norm)
```

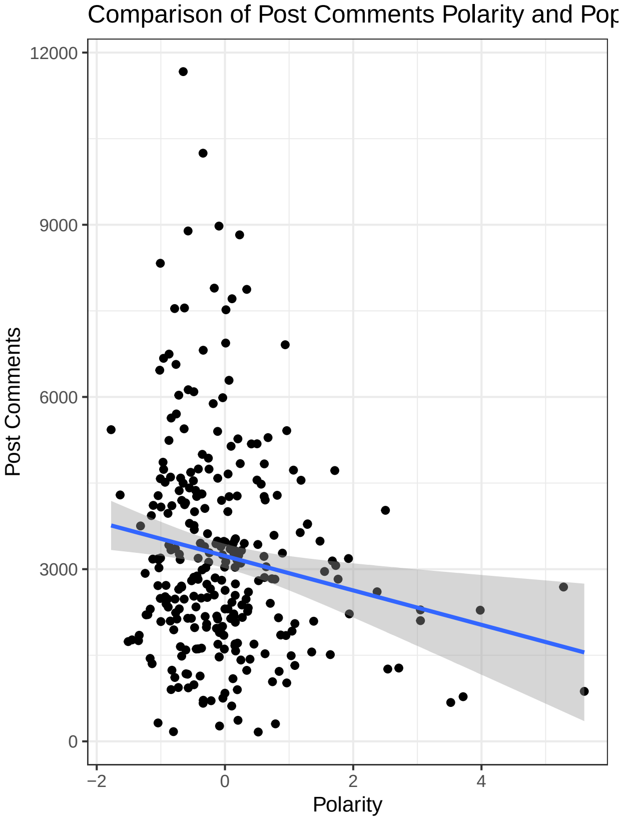
```
## NOTE: 15 observations removed because of NA values (LHS: 15).
```

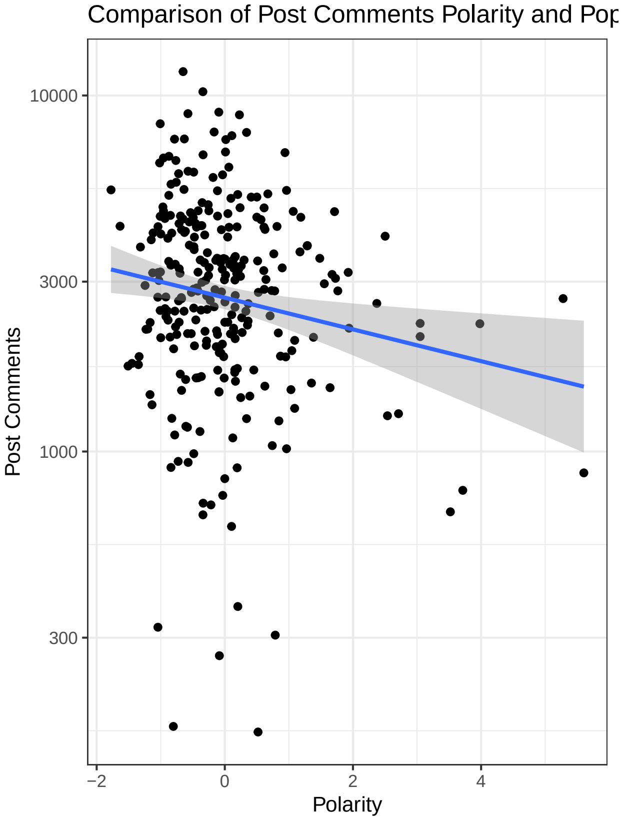
Results of Polarity on Score



```
##                             lev_lev           log_lev
## Dependent Var.:               score        log(score)
##
## (Intercept)     26,774.1*** (443.3) 10.16*** (0.0145)
## iqr_norm              625.0 (440.4)   0.0200 (0.0144)
## _______________ ___________________ _________________
## S.E. type                       IID               IID
## Observations                    285               285
## R2                          0.00706           0.00676
## Adj. R2                     0.00356           0.00325
```

Results of Polarity on Number of Comments





```
##                            lev_lev            log_lev
## Dependent Var.:       num_comments  log(num_comments)
##
## (Intercept)     3,229.2*** (107.5)  7.903*** (0.0382)
## iqr_norm          -299.7** (106.8) -0.1027** (0.0380)
## _______________ __________________ __________________
## S.E. type                      IID                IID
## Observations                   285                285
## R2                         0.02708            0.02517
## Adj. R2                    0.02364            0.02173
```

Results of Polarity on Score, Controlling for Number of Comments

```
fixest::etable(lev_lev, log_lev)
```

```
##                               lev_lev           log_lev
## Dependent Var.:                 score        log(score)
##
## (Intercept)       23,628.2*** (883.3) 9.582*** (0.1759)
## iqr_norm               916.9* (434.7)  0.0276. (0.0144)
## num_comments       0.9742*** (0.2387)
## log(num_comments)                     0.0735** (0.0222)
## _________________ __________________ _________________
## S.E. type                         IID               IID
## Observations                      285               285
## R2                            0.06245           0.04398
## Adj. R2                       0.05581           0.03720
```