

Main_Analysis

2025-03-19

EDA

```
# Load necessary libraries
library(readr)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.3     v purrr     1.0.2
## vforcats    1.0.0     v stringr    1.5.0
## v ggplot2   3.5.1     v tibble     3.2.1
## v lubridate 1.9.3     v tidyv     1.3.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# 1. Load and Clean Data
csv_file <- "reddit_engagement_data.csv"
df <- read_csv(csv_file, show_col_types = FALSE)
df_cleaned <- df %>%
  drop_na()

df_cleaned <- df_cleaned %>%
  mutate(
    `Subreddit Name` = as.factor(`Subreddit Name`),
    `Contains Question` = as.factor(`Contains Question`),
    `Contains Emoji` = as.factor(`Contains Emoji`),
    `Contains Profanity` = as.factor(`Contains Profanity`),
    `Is Early Comment` = as.factor(`Is Early Comment`),
    `Comment Hour` = as.factor(`Comment Hour`),
    `Comment Day` = as.factor(`Comment Day`)
  )

names(df_cleaned) <- gsub(" ", "_", names(df_cleaned))
names(df_cleaned) <- gsub("[()]", "", names(df_cleaned))

# Optional: write cleaned data to file
#write_csv(df_cleaned, "reddit_cleaned.csv")

# 2. Basic Summary Statistics
summary(df_cleaned)

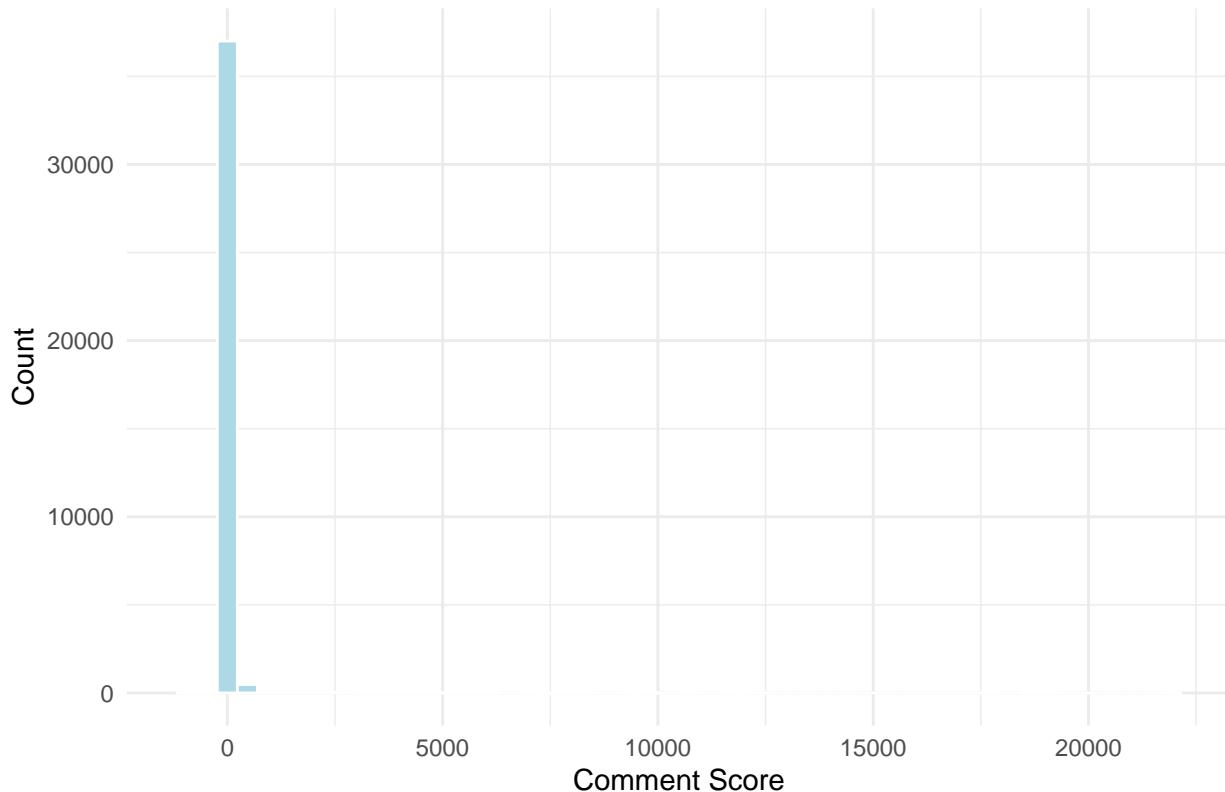
##  Comment_Score      Number_of_Replies      Text
##  Min. : -768.00    Min. : 0.0000    Length:37757
##  1st Qu.: 1.00      1st Qu.: 0.0000    Class :character
##  Median : 3.00      Median : 0.0000    Mode  :character
```

```

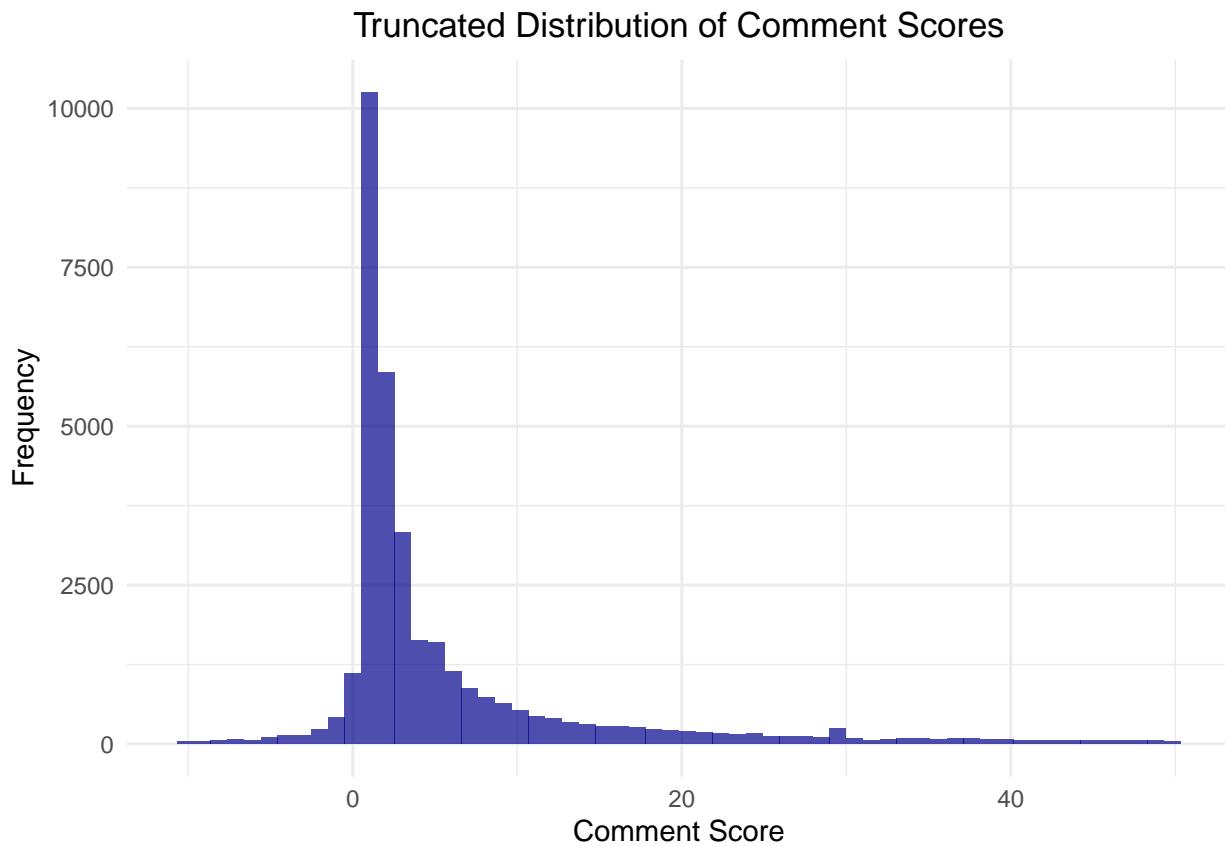
##  Mean    : 33.54   Mean    : 0.5763
##  3rd Qu.: 9.00    3rd Qu.: 1.0000
##  Max.   :22077.00  Max.   :50.0000
##
##          Subreddit_Name Sentiment_Score      Text_Length      Word_Count
##  technology       :13181   Min.   :-0.99810   Min.   : 1.0   Min.   : 1.0
##  AskReddit        : 5610    1st Qu.:-0.25000   1st Qu.: 49.0   1st Qu.: 9.0
##  gaming           : 4561    Median : 0.00000   Median : 103.0  Median : 19.0
##  datascience      : 3454    Mean    : 0.08411   Mean    : 181.7  Mean    : 32.2
##  unpopularopinion: 2696    3rd Qu.: 0.47670   3rd Qu.: 213.0  3rd Qu.: 38.0
##  movies            : 2448    Max.   : 0.99950   Max.   :5274.0   Max.   :898.0
##  (Other)          : 5807
##  Contains_Question Contains_Emoji Contains_Profanity Comment_Age_hours
##  0:15089          0:36851     0:32620           Min.   : 0.000
##  1:22668          1: 906      1: 5137          1st Qu.: 1.202
##  ##                         Median : 3.163
##  ##                         Mean   : 11.925
##  ##                         3rd Qu.: 7.487
##  ##                         Max.   :26582.484
##
##          Comment_Hour Comment_Day Is_Early_Comment Parent_Score      User_Karma
##  16   : 2482      0: 8893     0:29797           Min.   :-538   Min.   : -100
##  20   : 2473      1: 4804     1: 7960          1st Qu.: 11    1st Qu.: 3265
##  21   : 2421      2: 1011
##  0    : 2409      3: 1336
##  15   : 2374      4: 3522
##  19   : 2330      5:11065
##  (Other):23268   6: 7126           Median : 139   Median : 17766
##  Account_Age_days
##  Min.   : 0
##  1st Qu.: 769
##  Median :2137
##  Mean   :2326
##  3rd Qu.:3787
##  Max.   :7041
##
# 3. Full Distribution of Comment Score
ggplot(df_cleaned, aes(x = Comment_Score)) +
  geom_histogram(bins = 50, fill = "lightblue", color = "white") +
  labs(
    title = "Histogram of Comment Scores",
    x = "Comment Score",
    y = "Count"
  ) +
  theme_minimal()

```

Histogram of Comment Scores

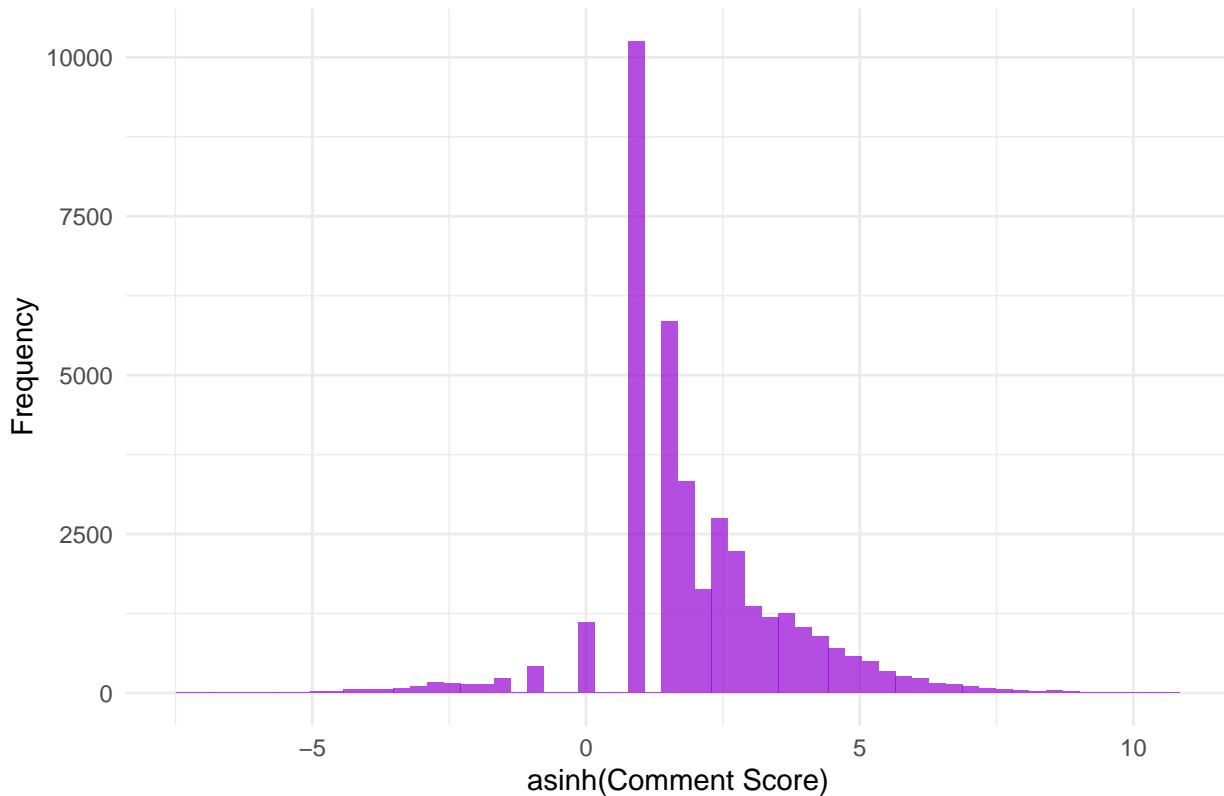


```
# 4. Zoomed-In Distribution (-10 to 50)
df_filtered <- df_cleaned %>% filter(Comment_Score >= -10 & Comment_Score <= 50)
ggplot(df_filtered, aes(x = (Comment_Score))) +
  geom_histogram(fill = "darkblue", bins = 60, alpha = 0.7) +
  theme_minimal() +
  labs(title = "Truncated Distribution of Comment Scores",
       x = "Comment Score",
       y = "Frequency") +
  theme(plot.title = element_text(hjust = 0.5))
```



```
## Inverse Hyperbolic Sine (IHS) Transformation Approach
# Distribution of Transformed Comment_Score using asinh()
ggplot(df_cleaned, aes(x = asinh(Comment_Score))) +
  geom_histogram(fill = "darkviolet", bins = 60, alpha = 0.7) +
  theme_minimal() +
  labs(title = "Distribution of IHS Transformed Comment Scores",
       x = "asinh(Comment Score)",
       y = "Frequency") +
  theme(plot.title = element_text(hjust = 0.5))
```

Distribution of IHS Transformed Comment Scores



Shifted Score Approaches

```
shift_df <- df_cleaned
lowest_score <- min(shift_df$Comment_Score)
shift_df$Comment_Score_shift <- shift_df$Comment_Score + abs(lowest_score) + 1
```

Log-Linear Multiple Linear Regression

```
library(MASS)

##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##     select
shift_df$log_score <- log(shift_df$Comment_Score_shift)

#Force Sentiment_Score to be part of the model
model_null_mlr <- lm(log_score ~ Sentiment_Score, data = shift_df)

#model_null_mlr <- lm(log_score ~ 1, data = shift_df)

model_full_mlr <- lm(log_score ~ Sentiment_Score + Text_Length + Word_Count +
    Contains_Question + Contains_Emoji + Contains_Profanity +
    Comment_Age_hours + Comment_Hour + Comment_Day +
```

```

    Is_Early_Comment + Parent_Score + User_Karma +
    Account_Age_days + Subreddit_Name, data = shift_df)

step_model_mlr <- stepAIC(model_null_mlr,
                           scope = list(lower = model_null_mlr, upper = model_full_mlr),
                           direction = "forward",
                           trace = TRUE)

## Start: AIC=-153293.9
## log_score ~ Sentiment_Score
##
##          Df Sum of Sq   RSS      AIC
## + Is_Early_Comment  1  20.3697 630.82 -154492
## + Subreddit_Name    9   4.9919 646.20 -153566
## + Comment_Hour     23   3.7885 647.41 -153468
## + Comment_Day       6   2.8175 648.38 -153446
## + Parent_Score      1   2.0501 649.14 -153411
## + User_Karma        1   1.4792 649.72 -153378
## + Account_Age_days 1   0.4559 650.74 -153318
## + Contains_Profanity 1   0.3441 650.85 -153312
## + Contains_Question 1   0.1606 651.03 -153301
## + Text_Length        1   0.1113 651.08 -153298
## + Word_Count         1   0.0839 651.11 -153297
## + Contains_Emoji     1   0.0738 651.12 -153296
## + Comment_Age_hours  1   0.0381 651.16 -153294
## <none>                  651.19 -153294
##
## Step: AIC=-154491.8
## log_score ~ Sentiment_Score + Is_Early_Comment
##
##          Df Sum of Sq   RSS      AIC
## + Subreddit_Name    9   5.5623 625.26 -154808
## + Comment_Day        6   4.7793 626.05 -154767
## + Comment_Hour     23   4.3179 626.51 -154705
## + Parent_Score       1   2.3810 628.44 -154633
## + User_Karma         1   1.4301 629.39 -154575
## + Account_Age_days  1   0.6903 630.13 -154531
## + Contains_Question 1   0.4233 630.40 -154515
## + Contains_Profanity 1   0.3571 630.47 -154511
## + Text_Length        1   0.3154 630.51 -154509
## + Word_Count         1   0.2817 630.54 -154507
## + Contains_Emoji     1   0.0897 630.73 -154495
## <none>                  630.82 -154492
## + Comment_Age_hours  1   0.0081 630.82 -154490
##
## Step: AIC=-154808.2
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name
##
##          Df Sum of Sq   RSS      AIC
## + Comment_Day        6   5.8560 619.41 -155151
## + Comment_Hour     23   4.7609 620.50 -155051
## + Parent_Score       1   1.6671 623.60 -154907
## + User_Karma         1   1.1853 624.08 -154878
## + Contains_Question 1   0.6788 624.58 -154847

```

```

## + Text_Length      1   0.4926 624.77 -154836
## + Account_Age_days 1   0.4788 624.78 -154835
## + Word_Count       1   0.4578 624.80 -154834
## + Contains_Profanity 1   0.3340 624.93 -154826
## + Contains_Emoji    1   0.1061 625.16 -154813
## <none>                  625.26 -154808
## + Comment_Age_hours  1   0.0024 625.26 -154806
##
## Step: AIC=-155151.5
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##             Comment_Day
##
##                               Df Sum of Sq   RSS     AIC
## + Comment_Hour          23  2.57240 616.83 -155263
## + Parent_Score           1   1.17067 618.24 -155221
## + User_Karma             1   0.98646 618.42 -155210
## + Contains_Question     1   0.36909 619.04 -155172
## + Account_Age_days      1   0.36852 619.04 -155172
## + Text_Length            1   0.35418 619.05 -155171
## + Contains_Profanity    1   0.33370 619.07 -155170
## + Word_Count             1   0.32249 619.08 -155169
## + Contains_Emoji         1   0.07030 619.34 -155154
## <none>                   619.41 -155151
## + Comment_Age_hours     1   0.00084 619.41 -155150
##
## Step: AIC=-155262.6
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##             Comment_Day + Comment_Hour
##
##                               Df Sum of Sq   RSS     AIC
## + Parent_Score           1   1.10507 615.73 -155328
## + User_Karma              1   0.98778 615.85 -155321
## + Account_Age_days        1   0.36318 616.47 -155283
## + Contains_Question       1   0.34942 616.48 -155282
## + Contains_Profanity      1   0.34057 616.49 -155281
## + Text_Length              1   0.34050 616.49 -155281
## + Word_Count              1   0.30821 616.53 -155279
## + Contains_Emoji           1   0.06042 616.77 -155264
## <none>                   616.83 -155263
## + Comment_Age_hours        1   0.00248 616.83 -155261
##
## Step: AIC=-155328.3
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##             Comment_Day + Comment_Hour + Parent_Score
##
##                               Df Sum of Sq   RSS     AIC
## + User_Karma              1   1.09398 614.63 -155393
## + Account_Age_days         1   0.42278 615.31 -155352
## + Text_Length              1   0.41516 615.31 -155352
## + Contains_Question        1   0.40618 615.32 -155351
## + Word_Count               1   0.37745 615.35 -155349
## + Contains_Profanity       1   0.32741 615.40 -155346
## + Contains_Emoji            1   0.06786 615.66 -155330
## <none>                   615.73 -155328

```

```

## + Comment_Age_hours 1 0.00247 615.73 -155326
##
## Step: AIC=-155393.5
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma
##
##                                     Df Sum of Sq   RSS      AIC
## + Text_Length                 1  0.37658 614.26 -155415
## + Contains_Question          1  0.36076 614.27 -155414
## + Word_Count                 1  0.34432 614.29 -155413
## + Contains_Profanity         1  0.31699 614.32 -155411
## + Account_Age_days           1  0.15384 614.48 -155401
## + Contains_Emoji             1  0.05104 614.58 -155395
## <none>                      614.63 -155393
## + Comment_Age_hours          1  0.00348 614.63 -155392
##
## Step: AIC=-155414.6
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma +
##           Text_Length
##
##                                     Df Sum of Sq   RSS      AIC
## + Contains_Profanity         1  0.202636 614.06 -155425
## + Contains_Question          1  0.169698 614.09 -155423
## + Account_Age_days           1  0.136698 614.12 -155421
## + Contains_Emoji             1  0.045987 614.21 -155415
## <none>                      614.26 -155415
## + Word_Count                 1  0.007519 614.25 -155413
## + Comment_Age_hours          1  0.003336 614.26 -155413
##
## Step: AIC=-155425.1
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma +
##           Text_Length + Contains_Profanity
##
##                                     Df Sum of Sq   RSS      AIC
## + Contains_Question          1  0.152906 613.90 -155432
## + Account_Age_days           1  0.135860 613.92 -155431
## + Contains_Emoji             1  0.047332 614.01 -155426
## <none>                      614.06 -155425
## + Word_Count                 1  0.014612 614.04 -155424
## + Comment_Age_hours          1  0.003121 614.05 -155423
##
## Step: AIC=-155432.5
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma +
##           Text_Length + Contains_Profanity + Contains_Question
##
##                                     Df Sum of Sq   RSS      AIC
## + Account_Age_days           1  0.131857 613.77 -155439
## + Contains_Emoji              1  0.045337 613.86 -155433
## <none>                      613.90 -155432
## + Word_Count                 1  0.030045 613.87 -155432
## + Comment_Age_hours          1  0.002889 613.90 -155431

```

```

## Step: AIC=-155438.6
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma +
##           Text_Length + Contains_Profanity + Contains_Question + Account_Age_days
##
##             Df Sum of Sq   RSS      AIC
## + Contains_Emoji    1  0.040676 613.73 -155439
## + Word_Count       1  0.032517 613.74 -155439
## <none>                  613.77 -155439
## + Comment_Age_hours 1  0.003000 613.77 -155437
##
## Step: AIC=-155439.1
## log_score ~ Sentiment_Score + Is_Early_Comment + Subreddit_Name +
##           Comment_Day + Comment_Hour + Parent_Score + User_Karma +
##           Text_Length + Contains_Profanity + Contains_Question + Account_Age_days +
##           Contains_Emoji
##
##             Df Sum of Sq   RSS      AIC
## <none>                  613.73 -155439
## + Word_Count       1 0.0312444 613.70 -155439
## + Comment_Age_hours 1 0.0030821 613.73 -155437
summary(step_model_mlr)

##
## Call:
## lm(formula = log_score ~ Sentiment_Score + Is_Early_Comment +
##     Subreddit_Name + Comment_Day + Comment_Hour + Parent_Score +
##     User_Karma + Text_Length + Contains_Profanity + Contains_Question +
##     Account_Age_days + Contains_Emoji, data = shift_df)
##
## Residuals:
##      Min        1Q        Median        3Q        Max 
## -6.7186 -0.0322 -0.0104  0.0055  3.3002 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               6.679e+00  3.998e-03 1670.610 < 2e-16 ***
## Sentiment_Score            2.357e-05  1.394e-03   0.017  0.986512  
## Is_Early_Comment1         6.148e-02  1.684e-03   36.508 < 2e-16 ***
## Subreddit_Namebooks        -5.658e-02 4.544e-03  -12.452 < 2e-16 ***
## Subreddit_Namedatascience -5.223e-02 3.465e-03  -15.073 < 2e-16 ***
## Subreddit_NameFloridaMan  -4.080e-02 4.774e-03  -8.546 < 2e-16 ***
## Subreddit_Namegaming        -4.659e-02 3.661e-03  -12.727 < 2e-16 ***
## Subreddit_NameHealth        -3.987e-02 4.274e-03  -9.330 < 2e-16 ***
## Subreddit_Namemovies        -4.715e-02 4.113e-03  -11.463 < 2e-16 ***
## Subreddit_NameShowerthoughts -3.406e-02 3.859e-03  -8.826 < 2e-16 ***
## Subreddit_Nametechnology   -3.742e-02 2.550e-03  -14.671 < 2e-16 ***
## Subreddit_Nameunpopularopinion -6.588e-02 3.890e-03  -16.935 < 2e-16 ***
## Comment_Day1                -3.084e-02 2.878e-03  -10.715 < 2e-16 ***
## Comment_Day2                -1.242e-03 4.756e-03  -0.261  0.794022  
## Comment_Day3                1.064e-04 4.283e-03   0.025  0.980190  
## Comment_Day4                1.505e-02 3.280e-03   4.588  4.48e-06 ***
## Comment_Day5                9.610e-04 2.624e-03   0.366  0.714157 

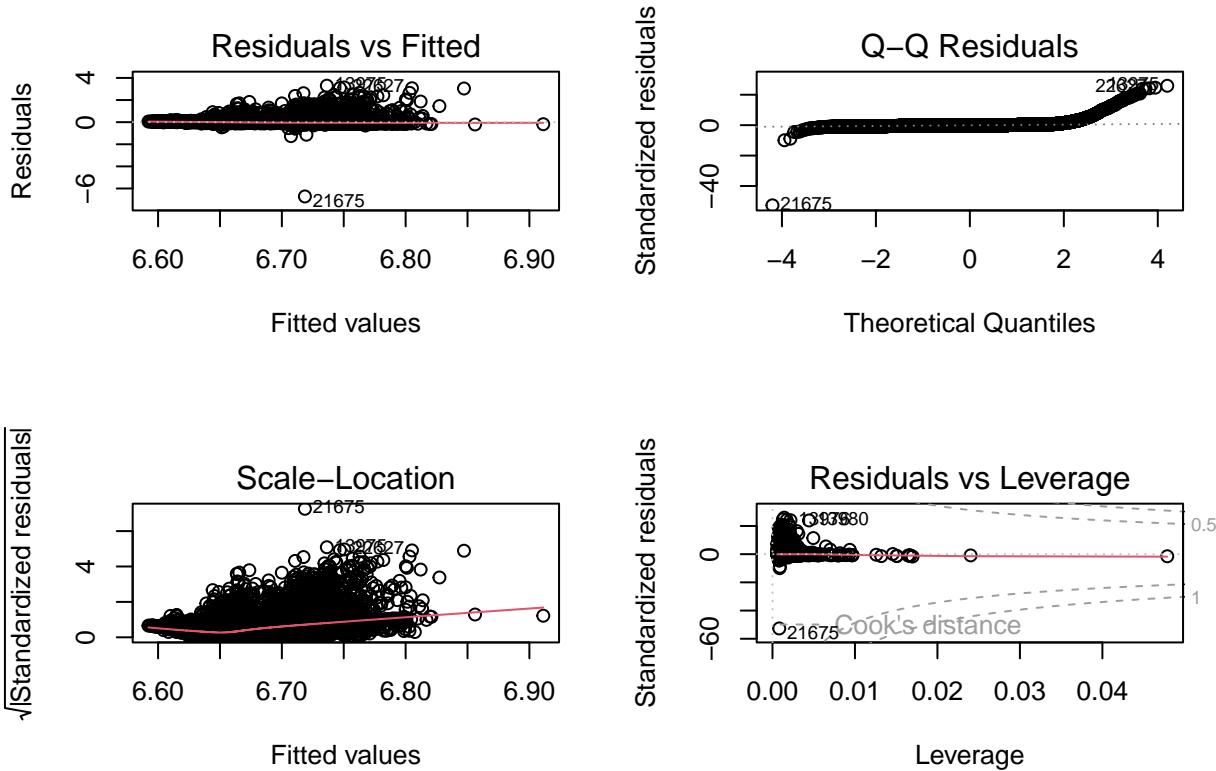
```

```

## Comment_Day6          6.070e-03  2.186e-03   2.777  0.005491 ***
## Comment_Hour1         -2.243e-03 3.839e-03  -0.584  0.559043
## Comment_Hour2          4.538e-03  4.288e-03   1.058  0.289986
## Comment_Hour3          2.431e-03  4.355e-03   0.558  0.576735
## Comment_Hour4          7.006e-03  4.502e-03   1.556  0.119676
## Comment_Hour5          8.141e-03  5.211e-03   1.562  0.118279
## Comment_Hour6          1.050e-02  5.962e-03   1.762  0.078063 .
## Comment_Hour7          7.605e-03  6.072e-03   1.252  0.210442
## Comment_Hour8          9.170e-03  6.532e-03   1.404  0.160357
## Comment_Hour9          8.533e-03  6.328e-03   1.349  0.177484
## Comment_Hour10         1.481e-02  6.159e-03   2.405  0.016171 *
## Comment_Hour11         4.322e-02  5.714e-03   7.564  4.01e-14 ***
## Comment_Hour12         3.605e-02  5.009e-03   7.197  6.28e-13 ***
## Comment_Hour13         1.395e-02  4.470e-03   3.121  0.001805 **
## Comment_Hour14         1.745e-02  4.154e-03   4.199  2.69e-05 ***
## Comment_Hour15         7.955e-03  3.903e-03   2.038  0.041566 *
## Comment_Hour16         1.129e-02  3.872e-03   2.915  0.003557 **
## Comment_Hour17         5.059e-03  4.005e-03   1.263  0.206560
## Comment_Hour18         4.578e-03  3.973e-03   1.152  0.249283
## Comment_Hour19         6.152e-04  3.914e-03   0.157  0.875098
## Comment_Hour20         8.401e-03  3.881e-03   2.165  0.030419 *
## Comment_Hour21        -6.023e-04  3.904e-03  -0.154  0.877403
## Comment_Hour22         9.236e-03  3.971e-03   2.326  0.020039 *
## Comment_Hour23         6.773e-03  4.204e-03   1.611  0.107158
## Parent_Score           6.311e-07  7.036e-08   8.970  < 2e-16 ***
## User_Karma              3.100e-08  4.483e-09   6.914  4.77e-12 ***
## Text_Length             7.568e-06  2.863e-06   2.644  0.008205 **
## Contains_Profanity1    6.777e-03  2.002e-03   3.386  0.000711 ***
## Contains_Question1     4.400e-03  1.464e-03   3.006  0.002645 **
## Account_Age_days       1.155e-06  4.133e-07   2.796  0.005183 **
## Contains_Emoji1        -6.825e-03  4.317e-03  -1.581  0.113912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1276 on 37709 degrees of freedom
## Multiple R-squared:  0.05767,   Adjusted R-squared:  0.0565
## F-statistic:  49.1 on 47 and 37709 DF,  p-value: < 2.2e-16

par(mfrow=c(2,2))
plot(step_model_mlr)

```



```
par(mfrow=c(1,1))
```

Gamma Regression with Log Link

```
#Force Sentiment_Score to be part of the model
model_null_gamma <- glm(Comment_Score_shift ~ Sentiment_Score, family = Gamma(link = "log"), data = shift_df)

#model_null_gamma <- glm(Comment_Score_shift ~ 1, family = Gamma(link = "log"), data = shift_df)

model_full_gamma <- glm(Comment_Score_shift ~ Sentiment_Score + Text_Length + Word_Count +
                          Contains_Question + Contains_Emoji + Contains_Profanity +
                          Comment_Age_hours + Comment_Hour + Comment_Day +
                          Is_Early_Comment + Parent_Score + User_Karma +
                          Account_Age_days + Subreddit_Name, family = Gamma(link = "log"), data = shift_df)

step_model_gamma <- stepAIC(model_null_gamma,
                               scope = list(lower = model_null_gamma, upper = model_full_gamma),
                               direction = "forward",
                               trace = TRUE)

## Start: AIC=480844.6
## Comment_Score_shift ~ Sentiment_Score
##
##          Df Deviance   AIC
## + Is_Early_Comment  1  1123.8 480442
## + Parent_Score      1  1169.2 480696
## + Comment_Hour     23  1165.9 480721
## + Subreddit_Name    9  1175.0 480744
## + Comment_Day       6  1181.8 480776
```

```

## + User_Karma      1  1187.4 480797
## + Account_Age_days 1  1193.2 480829
## + Contains_Profanity 1  1194.5 480837
## + Contains_Question 1  1195.7 480843
## + Contains_Emoji   1  1195.8 480844
## <none>           1196.3 480845
## + Comment_Age_hours 1  1196.2 480846
## + Text_Length     1  1196.2 480846
## + Word_Count      1  1196.3 480846
##
## Step: AIC=478472.8
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment
##
##          Df Deviance    AIC
## + Parent_Score   1  1098.6 478317
## + Comment_Hour  23  1093.9 478332
## + Comment_Day   6   1103.4 478357
## + Subreddit_Name 9   1103.7 478365
## + User_Karma    1   1116.2 478427
## + Account_Age_days 1   1120.0 478451
## + Contains_Profanity 1   1122.0 478464
## + Contains_Question 1   1122.3 478465
## + Contains_Emoji   1   1123.3 478472
## + Text_Length    1   1123.3 478472
## + Word_Count    1   1123.3 478472
## <none>           1123.8 478473
## + Comment_Age_hours 1   1123.7 478475
##
## Step: AIC=477614.5
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score
##
##          Df Deviance    AIC
## + Comment_Hour  23  1073.2 477488
## + Subreddit_Name 9   1082.1 477521
## + Comment_Day   6   1086.4 477544
## + User_Karma    1   1090.9 477564
## + Account_Age_days 1   1094.8 477591
## + Contains_Question 1   1096.9 477605
## + Contains_Profanity 1   1097.1 477606
## + Text_Length    1   1097.7 477611
## + Word_Count    1   1097.8 477611
## + Contains_Emoji   1   1098.1 477613
## <none>           1098.6 477614
## + Comment_Age_hours 1   1098.5 477616
##
## Step: AIC=476773.5
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
##          Comment_Hour
##
##          Df Deviance    AIC
## + Subreddit_Name 9   1055.2 476657
## + User_Karma     1   1066.5 476725
## + Comment_Day    6   1065.3 476727
## + Account_Age_days 1   1070.1 476753

```

```

## + Contains_Profanity 1 1071.7 476764
## + Contains_Question 1 1071.9 476766
## + Text_Length 1 1072.5 476771
## + Word_Count 1 1072.6 476771
## + Contains_Emoji 1 1072.9 476773
## <none> 1073.2 476774
## + Comment_Age_hours 1 1073.2 476775
##
## Step: AIC=476152.5
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
## Comment_Hour + Subreddit_Name
##
## Df Deviance AIC
## + Comment_Day 6 1044.3 476078
## + User_Karma 1 1049.0 476106
## + Account_Age_days 1 1052.2 476131
## + Contains_Question 1 1053.2 476138
## + Contains_Profanity 1 1053.8 476143
## + Text_Length 1 1054.2 476146
## + Word_Count 1 1054.3 476147
## + Contains_Emoji 1 1054.8 476151
## <none> 1055.2 476153
## + Comment_Age_hours 1 1055.2 476154
##
## Step: AIC=475768.4
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
## Comment_Hour + Subreddit_Name + Comment_Day
##
## Df Deviance AIC
## + User_Karma 1 1038.8 475726
## + Account_Age_days 1 1041.7 475749
## + Contains_Question 1 1042.8 475759
## + Contains_Profanity 1 1042.9 475759
## + Text_Length 1 1043.5 475764
## + Word_Count 1 1043.6 475765
## + Contains_Emoji 1 1043.9 475767
## <none> 1044.3 475768
## + Comment_Age_hours 1 1044.3 475770
##
## Step: AIC=475568.9
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
## Comment_Hour + Subreddit_Name + Comment_Day + User_Karma
##
## Df Deviance AIC
## + Contains_Profanity 1 1037.4 475560
## + Contains_Question 1 1037.5 475560
## + Account_Age_days 1 1037.7 475562
## + Text_Length 1 1038.0 475565
## + Word_Count 1 1038.1 475566
## + Contains_Emoji 1 1038.5 475568
## <none> 1038.8 475569
## + Comment_Age_hours 1 1038.7 475571
##
## Step: AIC=475522.2

```

```

## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
##   Comment_Hour + Subreddit_Name + Comment_Day + User_Karma +
##   Contains_Profanity
##
##               Df Deviance     AIC
## + Account_Age_days  1  1036.4 475516
## + Contains_Question 1  1036.4 475516
## + Text_Length        1  1037.0 475521
## + Word_Count         1  1037.1 475521
## + Contains_Emoji     1  1037.1 475522
## <none>                1037.4 475522
## + Comment_Age_hours  1  1037.4 475524
##
## Step:  AIC=475487
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
##   Comment_Hour + Subreddit_Name + Comment_Day + User_Karma +
##   Contains_Profanity + Account_Age_days
##
##               Df Deviance     AIC
## + Contains_Question 1  1035.4 475481
## + Text_Length        1  1036.0 475486
## + Word_Count         1  1036.1 475487
## + Contains_Emoji     1  1036.1 475487
## <none>                1036.4 475487
## + Comment_Age_hours  1  1036.4 475489
##
## Step:  AIC=475454
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
##   Comment_Hour + Subreddit_Name + Comment_Day + User_Karma +
##   Contains_Profanity + Account_Age_days + Contains_Question
##
##               Df Deviance     AIC
## + Contains_Emoji     1  1035.2 475454
## <none>                1035.4 475454
## + Text_Length        1  1035.4 475455
## + Word_Count         1  1035.4 475456
## + Comment_Age_hours  1  1035.4 475456
##
## Step:  AIC=475447
## Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment + Parent_Score +
##   Comment_Hour + Subreddit_Name + Comment_Day + User_Karma +
##   Contains_Profanity + Account_Age_days + Contains_Question +
##   Contains_Emoji
##
##               Df Deviance     AIC
## <none>                1035.2 475447
## + Text_Length        1  1035.1 475448
## + Word_Count         1  1035.2 475449
## + Comment_Age_hours  1  1035.2 475449
summary(step_model_gamma)

##
## Call:
## glm(formula = Comment_Score_shift ~ Sentiment_Score + Is_Early_Comment +

```

```

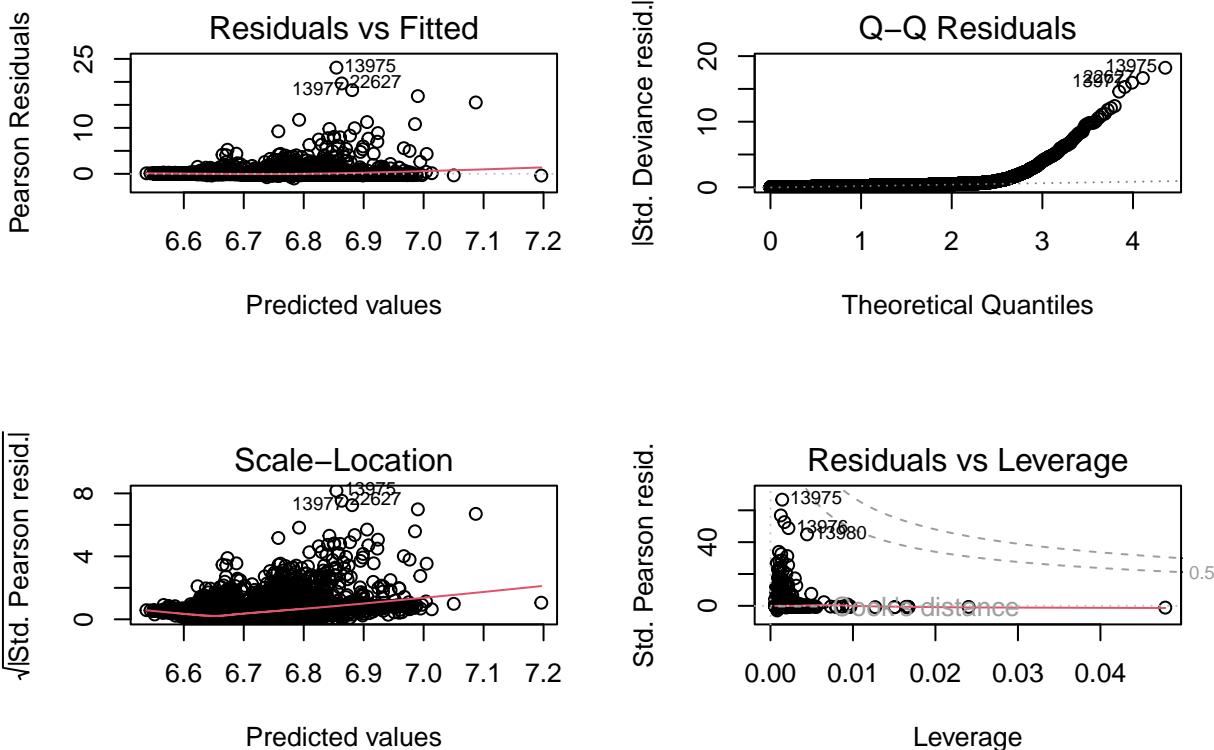
##      Parent_Score + Comment_Hour + Subreddit_Name + Comment_Day +
##      User_Karma + Contains_Profanity + Account_Age_days + Contains_Question +
##      Contains_Emoji, family = Gamma(link = "log"), data = shift_df)
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                6.694e+00  1.085e-02 616.929 < 2e-16 ***
## Sentiment_Score             5.842e-03  3.765e-03   1.552  0.12074
## Is_Early_Comment1          1.068e-01  4.573e-03  23.345 < 2e-16 ***
## Parent_Score                2.255e-06  1.909e-07 11.812 < 2e-16 ***
## Comment_Hour1              -2.638e-03  1.043e-02  -0.253  0.80027
## Comment_Hour2              5.960e-03  1.165e-02   0.512  0.60890
## Comment_Hour3              2.322e-03  1.183e-02   0.196  0.84442
## Comment_Hour4              1.195e-02  1.223e-02   0.977  0.32838
## Comment_Hour5              1.723e-02  1.416e-02   1.217  0.22355
## Comment_Hour6              2.213e-02  1.619e-02   1.367  0.17171
## Comment_Hour7              1.229e-02  1.650e-02   0.745  0.45634
## Comment_Hour8              2.129e-02  1.774e-02   1.200  0.23025
## Comment_Hour9              1.535e-02  1.719e-02   0.893  0.37181
## Comment_Hour10             3.175e-02  1.673e-02   1.898  0.05777 .
## Comment_Hour11             1.252e-01  1.552e-02   8.066 7.45e-16 ***
## Comment_Hour12             9.531e-02  1.361e-02   7.004 2.53e-12 ***
## Comment_Hour13             2.440e-02  1.214e-02   2.010  0.04447 *
## Comment_Hour14             3.200e-02  1.129e-02   2.836  0.00457 **
## Comment_Hour15             1.263e-02  1.060e-02   1.191  0.23353
## Comment_Hour16             2.068e-02  1.052e-02   1.966  0.04926 *
## Comment_Hour17             6.260e-03  1.088e-02   0.575  0.56507
## Comment_Hour18             5.877e-03  1.079e-02   0.544  0.58612
## Comment_Hour19             1.143e-03  1.063e-02   0.107  0.91441
## Comment_Hour20             2.641e-02  1.054e-02   2.505  0.01224 *
## Comment_Hour21             -3.152e-05 1.061e-02  -0.003  0.99763
## Comment_Hour22             2.319e-02  1.079e-02   2.150  0.03158 *
## Comment_Hour23             1.099e-02  1.142e-02   0.963  0.33579
## Subreddit_Namebooks        -1.042e-01  1.234e-02  -8.445 < 2e-16 ***
## Subreddit_Namedatascience -9.302e-02  9.393e-03  -9.903 < 2e-16 ***
## Subreddit_NameFloridaMan  -7.714e-02  1.296e-02  -5.950 2.70e-09 ***
## Subreddit_Namégaming       -8.150e-02  9.944e-03  -8.196 2.57e-16 ***
## Subreddit_NameHealth        -7.760e-02  1.161e-02  -6.685 2.35e-11 ***
## Subreddit_Namemovies        -9.312e-02  1.117e-02  -8.335 < 2e-16 ***
## Subreddit_NameShowerthoughts -6.983e-02  1.048e-02  -6.661 2.75e-11 ***
## Subreddit_Nametechnology    -7.829e-02  6.928e-03  -11.300 < 2e-16 ***
## Subreddit_Nameunpopularopinion -1.118e-01  1.057e-02  -10.579 < 2e-16 ***
## Comment_Day1                -5.346e-02  7.819e-03  -6.838 8.18e-12 ***
## Comment_Day2                -2.565e-03  1.292e-02  -0.198  0.84266
## Comment_Day3                -7.755e-04  1.164e-02  -0.067  0.94686
## Comment_Day4                2.821e-02  8.910e-03   3.166  0.00155 **
## Comment_Day5                2.579e-04  7.127e-03   0.036  0.97114
## Comment_Day6                9.170e-03  5.938e-03   1.544  0.12255
## User_Karma                  6.901e-08  1.218e-08   5.667 1.47e-08 ***
## Contains_Profanity1         1.579e-02  5.358e-03   2.947  0.00321 **
## Account_Age_days            3.138e-06  1.122e-06   2.796  0.00518 **
## Contains_Question1          1.047e-02  3.746e-03   2.796  0.00518 **
## Contains_Emoji1              -1.685e-02  1.173e-02  -1.437  0.15085
## ---

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.1201041)
##
##      Null deviance: 1196.3  on 37756  degrees of freedom
## Residual deviance: 1035.2  on 37710  degrees of freedom
## AIC: 475447
##
## Number of Fisher Scoring iterations: 6
par(mfrow=c(2,2))
plot(step_model_gamma)

```



```
par(mfrow=c(1,1))
```

Generalized Additive Model (GAM)

```

library(mgcv)

## Loading required package: nlme
##
## Attaching package: 'nlme'
## The following object is masked from 'package:dplyr':
##     collapse
## This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.
gam_model <- gam(Comment_Score_shift ~
                  s(User_Karma) +

```

```

Comment_Age_hours +
s(Word_Count) +
s(Account_Age_days) +
s(Text_Length) +
s(Parent_Score) +
s(Sentiment_Score) +
Is_Early_Comment +
Subreddit_Name +
Comment_Day +
Comment_Hour +
Contains_Question +
Contains_Emoji +
Contains_Profanity, data = shift_df)

summary(gam_model)

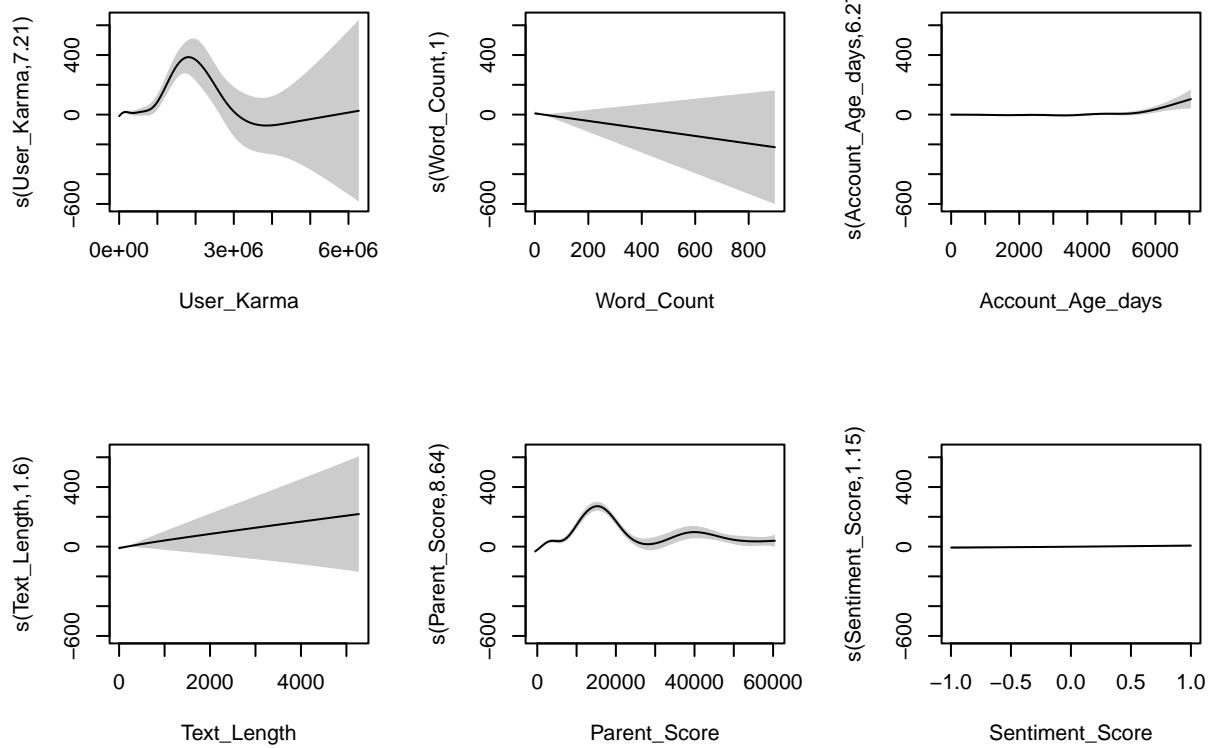
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Comment_Score_shift ~ s(User_Karma) + Comment_Age_hours + s(Word_Count) +
##   s(Account_Age_days) + s(Text_Length) + s(Parent_Score) +
##   s(Sentiment_Score) + Is_Early_Comment + Subreddit_Name +
##   Comment_Day + Comment_Hour + Contains_Question + Contains_Emoji +
##   Contains_Profanity
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 8.173e+02  1.030e+01  79.333 < 2e-16 ***
## Comment_Age_hours            -9.104e-04 6.914e-03 -0.132  0.89523
## Is_Early_Comment1           9.279e+01  4.415e+00 21.016 < 2e-16 ***
## Subreddit_Namebooks        -7.031e+01 1.200e+01 -5.861 4.65e-09 ***
## Subreddit_Namedatascience -5.655e+01 9.262e+00 -6.106 1.03e-09 ***
## Subreddit_NameFloridaMan   -5.321e+01 1.261e+01 -4.220 2.45e-05 ***
## Subreddit_Namegaming       -7.152e+01 9.686e+00 -7.384 1.57e-13 ***
## Subreddit_NameHealth       -5.055e+01 1.130e+01 -4.472 7.77e-06 ***
## Subreddit_Namemovies       -7.348e+01 1.081e+01 -6.798 1.07e-11 ***
## Subreddit_NameShowerthoughts -5.734e+01 1.022e+01 -5.612 2.01e-08 ***
## Subreddit_Nametechnology   -6.327e+01 6.754e+00 -9.369 < 2e-16 ***
## Subreddit_Nameunpopularopinion -7.757e+01 1.034e+01 -7.503 6.39e-14 ***
## Comment_Day1                -3.409e+01 7.605e+00 -4.482 7.40e-06 ***
## Comment_Day2                -5.764e-01 1.248e+01 -0.046  0.96316
## Comment_Day3                8.555e-02 1.125e+01  0.008  0.99393
## Comment_Day4                1.850e+01 8.702e+00  2.126  0.03349 *
## Comment_Day5                6.134e-01 6.981e+00  0.088  0.92998
## Comment_Day6                1.093e+01 5.895e+00  1.854  0.06374 .
## Comment_Hour1               5.213e-01 1.005e+01  0.052  0.95863
## Comment_Hour2               5.103e+00 1.123e+01  0.454  0.64965
## Comment_Hour3               5.524e+00 1.141e+01  0.484  0.62833
## Comment_Hour4               1.447e+01 1.180e+01  1.226  0.22003
## Comment_Hour5               1.747e+01 1.364e+01  1.281  0.20014
## Comment_Hour6               2.116e+01 1.560e+01  1.357  0.17476
## Comment_Hour7               1.385e+01 1.588e+01  0.873  0.38293

```

```

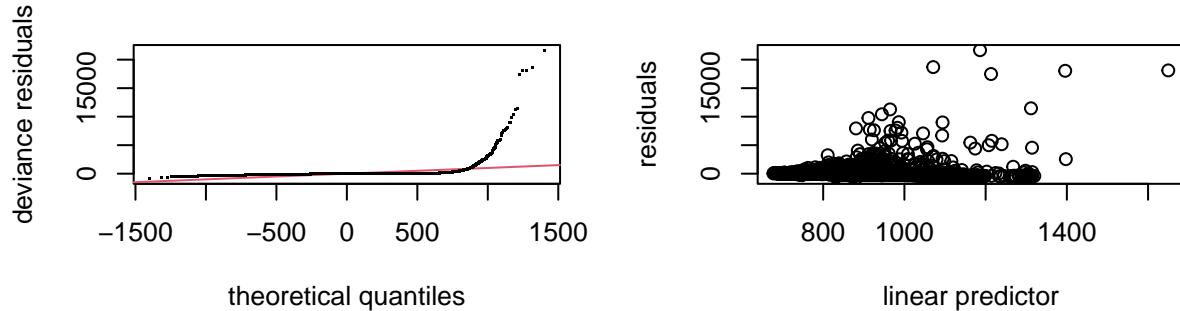
## Comment_Hour8          2.159e+01  1.708e+01  1.264  0.20614
## Comment_Hour9          1.780e+01  1.654e+01  1.076  0.28193
## Comment_Hour10         3.661e+01  1.611e+01  2.273  0.02305 *
## Comment_Hour11         1.247e+02  1.495e+01  8.337 < 2e-16 ***
## Comment_Hour12         8.540e+01  1.313e+01  6.505  7.88e-11 ***
## Comment_Hour13         2.220e+01  1.170e+01  1.897  0.05787 .
## Comment_Hour14         2.834e+01  1.087e+01  2.606  0.00917 **
## Comment_Hour15         1.020e+01  1.022e+01  0.998  0.31806
## Comment_Hour16         1.928e+01  1.015e+01  1.899  0.05755 .
## Comment_Hour17         6.119e+00  1.048e+01  0.584  0.55943
## Comment_Hour18         1.063e+01  1.040e+01  1.022  0.30677
## Comment_Hour19         6.202e+00  1.026e+01  0.604  0.54554
## Comment_Hour20         2.572e+01  1.017e+01  2.529  0.01145 *
## Comment_Hour21         6.312e+00  1.023e+01  0.617  0.53714
## Comment_Hour22         2.174e+01  1.041e+01  2.089  0.03671 *
## Comment_Hour23         1.466e+01  1.100e+01  1.333  0.18248
## Contains_Question1     5.121e+00  3.953e+00  1.295  0.19517
## Contains_Emoji1        -1.391e+01 1.130e+01 -1.231  0.21818
## Contains_Profanity1    1.059e+01  5.262e+00  2.013  0.04414 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                      edf Ref.df   F p-value
## s(User_Karma)      7.209 8.188 11.399 <2e-16 ***
## s(Word_Count)      1.000 1.000  1.318  0.2510
## s(Account_Age_days) 6.271 7.405  2.326  0.0159 *
## s(Text_Length)     1.602 2.013  1.053  0.3561
## s(Parent_Score)    8.645 8.954 63.793 <2e-16 ***
## s(Sentiment_Score) 1.150 1.285  2.443  0.0851 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0387  Deviance explained = 4.05%
## GCV = 1.1133e+05  Scale est. = 1.1113e+05 n = 37757
plot(gam_model, pages = 1, shade = TRUE)

```

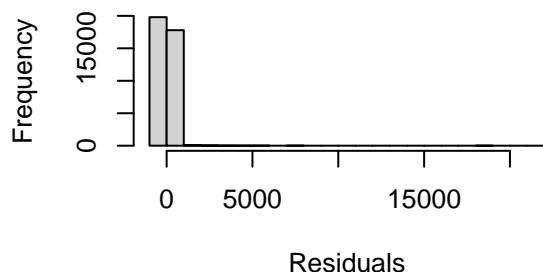


```
par(mfrow=c(2,2))
gam.check(gam_model)
```

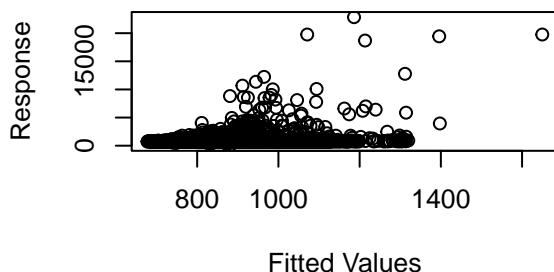
Resids vs. linear pred.



Histogram of residuals



Response vs. Fitted Values



```
##
## Method: GCV    Optimizer: magic
```

```

## Smoothing parameter selection converged after 20 iterations.
## The RMS GCV score gradient at convergence was 0.009353544 .
## The Hessian was positive definite.
## Model rank = 98 / 98
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##          k'  edf k-index p-value
## s(User_Karma)    9.00 7.21    1.00    0.29
## s(Word_Count)    9.00 1.00    1.01    0.81
## s(Account_Age_days) 9.00 6.27    1.01    0.83
## s(Text_Length)    9.00 1.60    1.00    0.74
## s(Parent_Score)   9.00 8.64    0.94 <2e-16 ***
## s(Sentiment_Score) 9.00 1.15    0.99    0.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(mfrow=c(1,1))

```

##Fitting Other Shifting Based Models (Not used in paper)

Poisson Model

```

full.model <- glm(Comment_Score_shift ~ Sentiment_Score + Text_Length + Word_Count +
                  Contains_Question + Contains_Emoji + Contains_Profanity +
                  Comment_Age_hours + Comment_Hour + Comment_Day +
                  Is_Early_Comment + Parent_Score + User_Karma +
                  Account_Age_days + Subreddit_Name, data = shift_df, family = poisson)

summary(full.model)

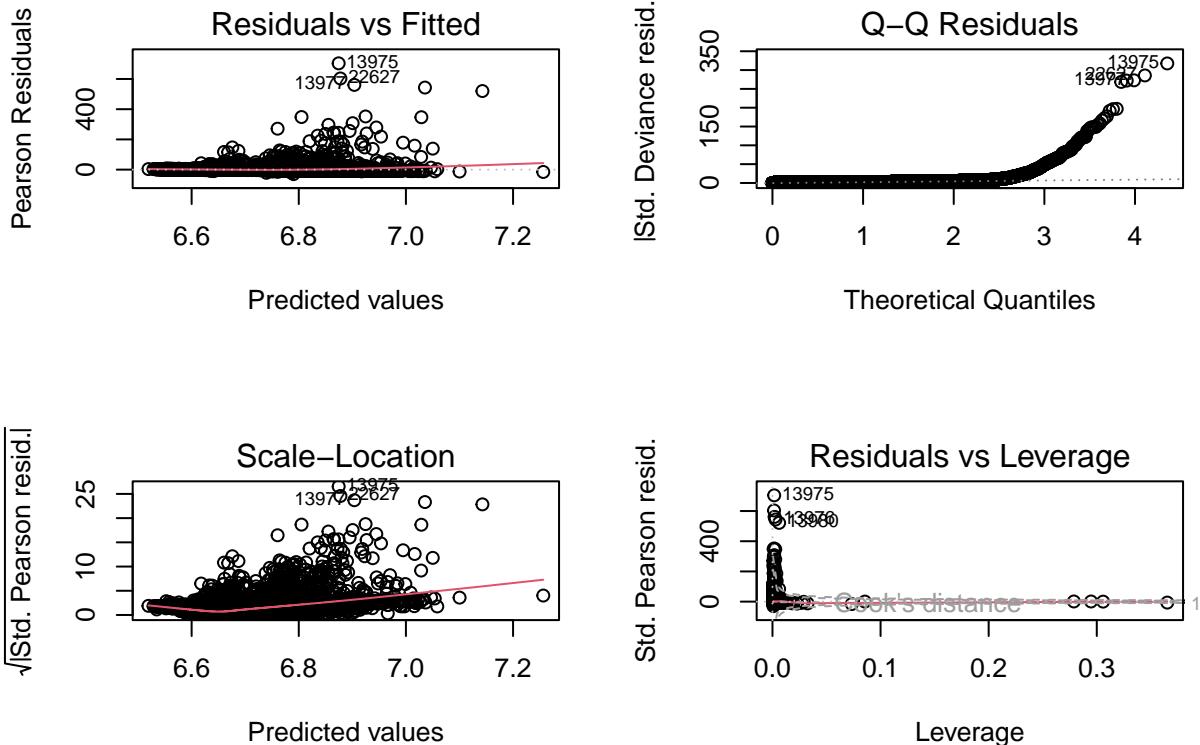
##
## Call:
## glm(formula = Comment_Score_shift ~ Sentiment_Score + Text_Length +
##       Word_Count + Contains_Question + Contains_Emoji + Contains_Profanity +
##       Comment_Age_hours + Comment_Hour + Comment_Day + Is_Early_Comment +
##       Parent_Score + User_Karma + Account_Age_days + Subreddit_Name,
##       family = poisson, data = shift_df)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           6.698e+00  1.107e-03 6049.528 < 2e-16 ***
## Sentiment_Score        6.940e-03  3.856e-04   18.001 < 2e-16 ***
## Text_Length            4.904e-05  3.764e-06   13.030 < 2e-16 ***
## Word_Count             -2.608e-04 2.210e-05  -11.802 < 2e-16 ***
## Contains_Question1    1.076e-02  4.086e-04   26.327 < 2e-16 ***
## Contains_Emoji1       -1.916e-02 1.208e-03  -15.859 < 2e-16 ***
## Contains_Profanity1   1.705e-02  5.514e-04   30.916 < 2e-16 ***
## Comment_Age_hours     -1.325e-06 7.519e-07  -1.763 0.07797 .
## Comment_Hour1          -3.681e-03 1.074e-03  -3.427 0.00061 ***
## Comment_Hour2          6.115e-03  1.196e-03   5.115 3.14e-07 ***
## Comment_Hour3          1.927e-03  1.215e-03   1.586 0.11271
## Comment_Hour4          1.347e-02  1.258e-03  10.708 < 2e-16 ***
## Comment_Hour5          1.830e-02  1.454e-03  12.591 < 2e-16 ***

```

```

## Comment_Hour6          2.568e-02  1.666e-03  15.419 < 2e-16 ***
## Comment_Hour7          1.399e-02  1.697e-03   8.240 < 2e-16 ***
## Comment_Hour8          2.512e-02  1.831e-03  13.716 < 2e-16 ***
## Comment_Hour9          1.732e-02  1.770e-03   9.780 < 2e-16 ***
## Comment_Hour10         3.737e-02  1.707e-03  21.888 < 2e-16 ***
## Comment_Hour11         1.491e-01  1.513e-03  98.520 < 2e-16 ***
## Comment_Hour12         1.038e-01  1.354e-03  76.632 < 2e-16 ***
## Comment_Hour13         2.662e-02  1.236e-03  21.534 < 2e-16 ***
## Comment_Hour14         3.409e-02  1.150e-03  29.645 < 2e-16 ***
## Comment_Hour15         1.172e-02  1.084e-03  10.814 < 2e-16 ***
## Comment_Hour16         2.232e-02  1.077e-03  20.717 < 2e-16 ***
## Comment_Hour17         5.801e-03  1.119e-03   5.183 2.19e-07 ***
## Comment_Hour18         5.626e-03  1.110e-03   5.069 3.99e-07 ***
## Comment_Hour19         4.980e-04  1.095e-03   0.455  0.64917
## Comment_Hour20         2.882e-02  1.078e-03  26.737 < 2e-16 ***
## Comment_Hour21        -1.067e-03  1.093e-03  -0.976  0.32915
## Comment_Hour22         2.532e-02  1.104e-03  22.930 < 2e-16 ***
## Comment_Hour23         1.065e-02  1.174e-03   9.070 < 2e-16 ***
## Comment_Day1           -6.099e-02  8.044e-04  -75.823 < 2e-16 ***
## Comment_Day2           -2.119e-03  1.332e-03  -1.591  0.11172
## Comment_Day3           -4.103e-04  1.201e-03  -0.342  0.73251
## Comment_Day4           3.111e-02  9.087e-04  34.238 < 2e-16 ***
## Comment_Day5           1.140e-03  7.300e-04   1.561  0.11841
## Comment_Day6           1.087e-02  6.047e-04  17.982 < 2e-16 ***
## Is_Early_Comment1      1.140e-01  4.531e-04  251.665 < 2e-16 ***
## Parent_Score            2.482e-06  1.863e-08  133.265 < 2e-16 ***
## User_Karma              7.762e-08  1.155e-09  67.179 < 2e-16 ***
## Account_Age_days        3.787e-06  1.140e-07  33.211 < 2e-16 ***
## Subreddit_Namebooks     -1.180e-01  1.269e-03  -92.917 < 2e-16 ***
## Subreddit_Namedatascience -1.059e-01  9.671e-04  -109.481 < 2e-16 ***
## Subreddit_NameFloridaMan -8.729e-02  1.343e-03  -64.988 < 2e-16 ***
## Subreddit_Namegaming     -9.381e-02  1.010e-03  -92.868 < 2e-16 ***
## Subreddit_NameHealth     -8.899e-02  1.193e-03  -74.602 < 2e-16 ***
## Subreddit_Namemovies     -1.070e-01  1.141e-03  -93.841 < 2e-16 ***
## Subreddit_NameShowerthoughts -8.081e-02  1.069e-03  -75.575 < 2e-16 ***
## Subreddit_Nametechnology -9.054e-02  6.998e-04  -129.387 < 2e-16 ***
## Subreddit_Nameunpopularopinion -1.250e-01  1.081e-03  -115.582 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 1790286  on 37756  degrees of freedom
## Residual deviance: 1644805  on 37707  degrees of freedom
## AIC: 1966218
##
## Number of Fisher Scoring iterations: 5
par(mfrow=c(2,2))
plot(full.model)

```



```
par(mfrow=c(1,1))
```

Negative Binomial Model

```
nb.model <- glm.nb(Comment_Score_shift ~ Sentiment_Score + Text_Length + Word_Count +
  Contains_Question + Contains_Emoji + Contains_Profanity +
  Comment_Age_hours + Comment_Hour + Comment_Day +
  Is_Early_Comment + Parent_Score + User_Karma +
  Account_Age_days + Subreddit_Name, data = shift_df)

summary(nb.model)
```

```
##
## Call:
## glm.nb(formula = Comment_Score_shift ~ Sentiment_Score + Text_Length +
##   Word_Count + Contains_Question + Contains_Emoji + Contains_Profanity +
##   Comment_Age_hours + Comment_Hour + Comment_Day + Is_Early_Comment +
##   Parent_Score + User_Karma + Account_Age_days + Subreddit_Name,
##   data = shift_df, init.theta = 37.35103336, link = log)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)               6.694e+00  5.248e-03 1275.573 < 2e-16 ***
## Sentiment_Score            5.643e-03  1.830e-03   3.083  0.00205 **
## Text_Length                4.991e-05  1.863e-05   2.679  0.00739 **
## Word_Count                -2.611e-04 1.091e-04  -2.392  0.01675 *
## Contains_Question1        1.002e-02  1.937e-03   5.173 2.30e-07 ***
## Contains_Emoji1           -1.657e-02 5.668e-03  -2.923  0.00347 **
## Contains_Profanity1       1.554e-02  2.632e-03   5.904 3.54e-09 ***
## Comment_Age_hours         -1.308e-06 3.474e-06  -0.377  0.70654
```

```

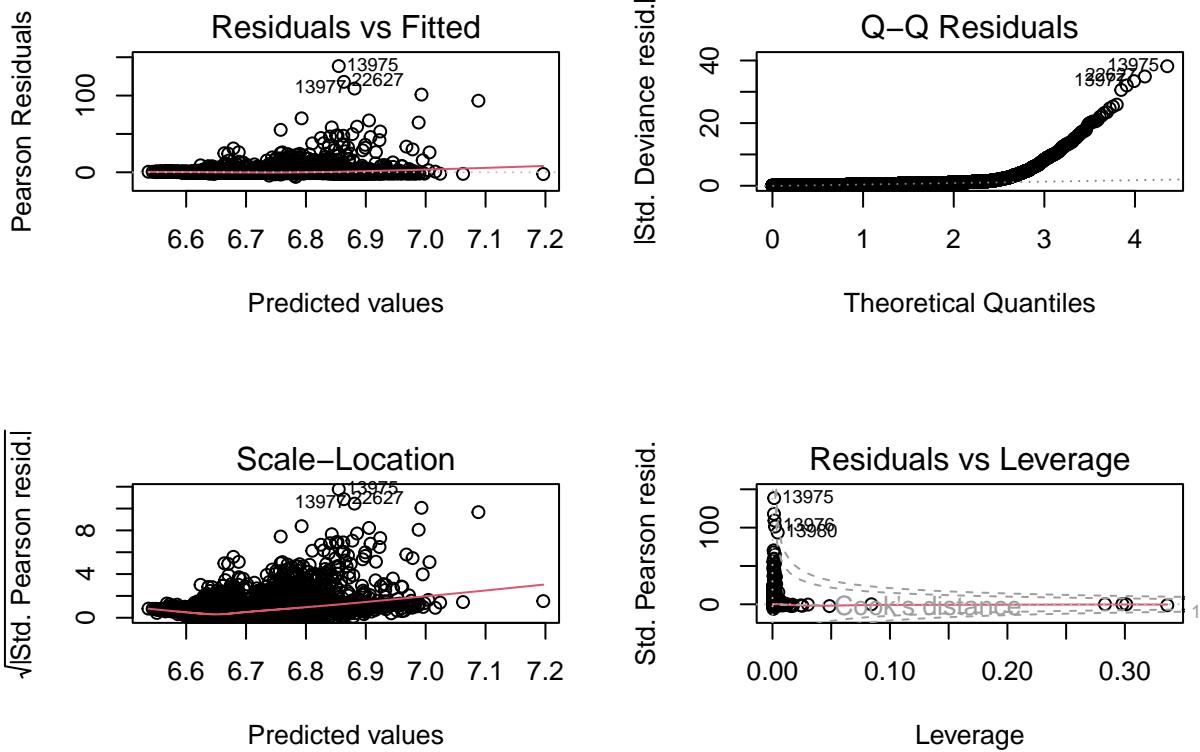
## Comment_Hour1          -2.668e-03  5.040e-03  -0.529  0.59649
## Comment_Hour2          6.027e-03  5.628e-03  1.071  0.28422
## Comment_Hour3          2.296e-03  5.717e-03  0.402  0.68791
## Comment_Hour4          1.187e-02  5.910e-03  2.009  0.04456 *
## Comment_Hour5          1.713e-02  6.841e-03  2.505  0.01226 *
## Comment_Hour6          2.241e-02  7.827e-03  2.864  0.00418 **
## Comment_Hour7          1.228e-02  7.971e-03  1.541  0.12330
## Comment_Hour8          2.147e-02  8.575e-03  2.504  0.01230 *
## Comment_Hour9          1.535e-02  8.307e-03  1.848  0.06463 .
## Comment_Hour10         3.199e-02  8.083e-03  3.958  7.57e-05 ***
## Comment_Hour11         1.260e-01  7.486e-03  16.829 < 2e-16 ***
## Comment_Hour12         9.567e-02  6.567e-03  14.568 < 2e-16 ***
## Comment_Hour13         2.451e-02  5.865e-03  4.180  2.92e-05 ***
## Comment_Hour14         3.214e-02  5.452e-03  5.895  3.74e-09 ***
## Comment_Hour15         1.257e-02  5.123e-03  2.453  0.01416 *
## Comment_Hour16         2.082e-02  5.083e-03  4.097  4.18e-05 ***
## Comment_Hour17         6.240e-03  5.258e-03  1.187  0.23533
## Comment_Hour18         5.860e-03  5.216e-03  1.124  0.26121
## Comment_Hour19         1.191e-03  5.137e-03  0.232  0.81669
## Comment_Hour20         2.662e-02  5.093e-03  5.227  1.73e-07 ***
## Comment_Hour21         5.097e-05  5.126e-03  0.010  0.99207
## Comment_Hour22         2.339e-02  5.212e-03  4.487  7.23e-06 ***
## Comment_Hour23         1.097e-02  5.518e-03  1.987  0.04690 *
## Comment_Day1           -5.362e-02  3.779e-03 -14.189 < 2e-16 ***
## Comment_Day2           -2.693e-03  6.245e-03 -0.431  0.66633
## Comment_Day3           -1.028e-03  5.625e-03 -0.183  0.85492
## Comment_Day4           2.825e-02  4.305e-03  6.562  5.32e-11 ***
## Comment_Day5           2.737e-04  3.444e-03  0.079  0.93665
## Comment_Day6           9.118e-03  2.869e-03  3.178  0.00148 **
## Is_Early_Comment1      1.070e-01  2.208e-03  48.436 < 2e-16 ***
## Parent_Score            2.270e-06  9.217e-08  24.632 < 2e-16 ***
## User_Karma              6.900e-08  5.870e-09  11.754 < 2e-16 ***
## Account_Age_days        3.157e-06  5.424e-07  5.821  5.83e-09 ***
## Subreddit_Namebooks     -1.046e-01  5.965e-03 -17.533 < 2e-16 ***
## Subreddit_Namedatascience -9.410e-02  4.551e-03 -20.678 < 2e-16 ***
## Subreddit_NameFloridaMan -7.721e-02  6.302e-03 -12.253 < 2e-16 ***
## Subreddit_Namégaming    -8.205e-02  4.804e-03 -17.079 < 2e-16 ***
## Subreddit_NameHealth    -7.842e-02  5.614e-03 -13.969 < 2e-16 ***
## Subreddit_Namemovies    -9.360e-02  5.397e-03 -17.341 < 2e-16 ***
## Subreddit_NameShowerthoughts -7.026e-02  5.066e-03 -13.869 < 2e-16 ***
## Subreddit_Nametechnology -7.916e-02  3.352e-03 -23.618 < 2e-16 ***
## Subreddit_Nameunpopularopinion -1.123e-01  5.106e-03 -21.987 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(37.351) family taken to be 1)
##
## Null deviance: 43130  on 37756  degrees of freedom
## Residual deviance: 37347  on 37707  degrees of freedom
## AIC: 475898
##
## Number of Fisher Scoring iterations: 1
##

```

```

##          Theta:  37.351
##      Std. Err.:  0.279
##
## 2 x log-likelihood: -475796.292
par(mfrow=c(2,2))
plot(nb.model)

```



```
par(mfrow=c(1,1))
```

Extra : Fitting IHS models (Not used in report)

Multiple Linear Regression (MLR)

```

library(MASS)

# Create the IHS-transformed response variable (using asinh)
df_cleaned <- df_cleaned %>%
  mutate(ihs_score = asinh(Comment_Score))

model_null <- lm(ihs_score ~ 1, data = df_cleaned)

model_full <- lm(ihs_score ~ Sentiment_Score + Text_Length + Word_Count +
  Contains_Question + Contains_Emoji + Contains_Profanity +
  Comment_Age_hours + Comment_Hour + Comment_Day +
  Is_Early_Comment + Parent_Score + User_Karma +
  Account_Age_days + Subreddit_Name , data = df_cleaned)

step_model <- stepAIC(model_null,
  scope = list(lower = model_null, upper = model_full),

```

```

    direction = "forward",
    trace = TRUE)

## Start: AIC=41395.11
## ihs_score ~ 1
##
##          Df Sum of Sq   RSS   AIC
## + Is_Early_Comment  1   5107.1 107903 39651
## + Subreddit_Name    9   4025.4 108984 40044
## + Comment_Day       6   1004.4 112005 41070
## + Comment_Hour      23   960.2 112050 41119
## + User_Karma        1   812.4 112197 41125
## + Account_Age_days  1   734.1 112276 41151
## + Contains_Profanity 1   132.8 112877 41353
## + Text_Length        1   98.4 112911 41364
## + Word_Count         1   94.3 112915 41366
## + Parent_Score        1   56.5 112953 41378
## + Sentiment_Score     1   54.1 112956 41379
## + Contains_Question   1   47.3 112962 41381
## + Comment_Age_hours   1   34.8 112975 41385
## + Contains_Emoji      1   20.4 112989 41390
## <none>                  113010 41395
##
## Step: AIC=39651.06
## ihs_score ~ Is_Early_Comment
##
##          Df Sum of Sq   RSS   AIC
## + Subreddit_Name    9   4482.4 103420 38067
## + Comment_Day       6   1603.4 106299 39098
## + Account_Age_days  1   874.9 107028 39346
## + User_Karma        1   793.2 107109 39374
## + Comment_Hour      23   896.8 107006 39382
## + Word_Count         1   186.6 107716 39588
## + Text_Length         1   186.6 107716 39588
## + Contains_Profanity 1   130.0 107773 39608
## + Contains_Question   1   118.4 107784 39612
## + Sentiment_Score     1   34.7 107868 39641
## + Parent_Score        1   34.0 107869 39641
## + Contains_Emoji      1   24.2 107878 39645
## + Comment_Age_hours   1   17.8 107885 39647
## <none>                  107903 39651
##
## Step: AIC=38067.09
## ihs_score ~ Is_Early_Comment + Subreddit_Name
##
##          Df Sum of Sq   RSS   AIC
## + Comment_Day       6   1839.93 101580 37401
## + Comment_Hour      23   1071.06 102349 37720
## + Account_Age_days  1   713.05 102707 37808
## + User_Karma        1   636.83 102783 37836
## + Word_Count         1   343.67 103077 37943
## + Text_Length         1   342.46 103078 37944
## + Contains_Question   1   302.42 103118 37959
## + Parent_Score        1   111.69 103309 38028

```

```

## + Contains_Profanity 1 100.83 103319 38032
## + Contains_Emoji 1 42.07 103378 38054
## + Comment_Age_hours 1 16.27 103404 38063
## <none> 103420 38067
## + Sentiment_Score 1 0.01 103420 38069
##
## Step: AIC=37401.32
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day
##
## Df Sum of Sq RSS AIC
## + Account_Age_days 1 639.49 100941 37165
## + User_Karma 1 563.02 101017 37193
## + Comment_Hour 23 520.64 101060 37253
## + Word_Count 1 274.25 101306 37301
## + Text_Length 1 273.50 101307 37302
## + Parent_Score 1 196.79 101384 37330
## + Contains_Question 1 188.07 101392 37333
## + Contains_Profanity 1 103.93 101476 37365
## + Contains_Emoji 1 30.18 101550 37392
## + Comment_Age_hours 1 14.11 101566 37398
## <none> 101580 37401
## + Sentiment_Score 1 0.70 101580 37403
##
## Step: AIC=37164.87
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
## Account_Age_days
##
## Df Sum of Sq RSS AIC
## + Comment_Hour 23 505.58 100435 37021
## + User_Karma 1 314.42 100626 37049
## + Text_Length 1 236.02 100705 37078
## + Word_Count 1 235.01 100706 37079
## + Parent_Score 1 166.29 100775 37105
## + Contains_Question 1 163.34 100778 37106
## + Contains_Profanity 1 99.17 100842 37130
## + Contains_Emoji 1 20.22 100921 37159
## + Comment_Age_hours 1 15.10 100926 37161
## <none> 100941 37165
## + Sentiment_Score 1 1.22 100940 37166
##
## Step: AIC=37021.28
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
## Account_Age_days + Comment_Hour
##
## Df Sum of Sq RSS AIC
## + User_Karma 1 310.704 100125 36906
## + Text_Length 1 225.081 100210 36939
## + Word_Count 1 223.249 100212 36939
## + Parent_Score 1 177.432 100258 36957
## + Contains_Question 1 153.854 100281 36965
## + Contains_Profanity 1 101.500 100334 36985
## + Contains_Emoji 1 19.138 100416 37016
## + Comment_Age_hours 1 17.512 100418 37017
## <none> 100435 37021

```

```

## + Sentiment_Score      1     1.945 100433 37023
##
## Step: AIC=36906.3
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##   Account_Age_days + Comment_Hour + User_Karma
##
##                                     Df Sum of Sq   RSS   AIC
## + Word_Count                  1   214.596 99910 36827
## + Text_Length                 1   214.465 99910 36827
## + Parent_Score                1   160.293 99964 36848
## + Contains_Question          1   141.618 99983 36855
## + Contains_Profanity         1    99.221 100025 36871
## + Comment_Age_hours          1    18.686 100106 36901
## + Contains_Emoji             1    15.571 100109 36902
## <none>                      100125 36906
## + Sentiment_Score            1     1.679 100123 36908
##
## Step: AIC=36827.28
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##   Account_Age_days + Comment_Hour + User_Karma + Word_Count
##
##                                     Df Sum of Sq   RSS   AIC
## + Parent_Score                1   141.218 99769 36776
## + Contains_Profanity          1    53.875 99856 36809
## + Contains_Question           1    48.961 99861 36811
## + Comment_Age_hours           1    18.521 99891 36822
## + Contains_Emoji              1    14.245 99896 36824
## + Sentiment_Score              1     6.171 99904 36827
## <none>                      99910 36827
## + Text_Length                 1     1.811 99908 36829
##
## Step: AIC=36775.88
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##   Account_Age_days + Comment_Hour + User_Karma + Word_Count +
##   Parent_Score
##
##                                     Df Sum of Sq   RSS   AIC
## + Contains_Profanity          1    58.542 99710 36756
## + Contains_Question           1    45.257 99723 36761
## + Comment_Age_hours           1    18.510 99750 36771
## + Contains_Emoji              1    13.373 99755 36773
## + Sentiment_Score              1     7.166 99762 36775
## <none>                      99769 36776
## + Text_Length                 1     1.431 99767 36777
##
## Step: AIC=36755.72
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##   Account_Age_days + Comment_Hour + User_Karma + Word_Count +
##   Parent_Score + Contains_Profanity
##
##                                     Df Sum of Sq   RSS   AIC
## + Contains_Question           1    41.215 99669 36742
## + Comment_Age_hours           1    18.086 99692 36751
## + Contains_Emoji              1    13.280 99697 36753

```

```

## <none>                               99710 36756
## + Text_Length            1      2.208 99708 36757
## + Sentiment_Score        1      1.204 99709 36757
##
## Step: AIC=36742.11
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##             Account_Age_days + Comment_Hour + User_Karma + Word_Count +
##             Parent_Score + Contains_Profanity + Contains_Question
##
##                                     Df Sum of Sq   RSS   AIC
## + Comment_Age_hours  1    17.7887 99651 36737
## + Contains_Emoji     1    12.7050 99656 36739
## <none>                           99669 36742
## + Text_Length          1     3.5255 99665 36743
## + Sentiment_Score      1     1.2043 99668 36744
##
## Step: AIC=36737.37
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##             Account_Age_days + Comment_Hour + User_Karma + Word_Count +
##             Parent_Score + Contains_Profanity + Contains_Question + Comment_Age_hours
##
##                                     Df Sum of Sq   RSS   AIC
## + Contains_Emoji     1    12.8039 99638 36735
## <none>                           99651 36737
## + Text_Length          1     3.5199 99648 36738
## + Sentiment_Score      1     1.1108 99650 36739
##
## Step: AIC=36734.52
## ihs_score ~ Is_Early_Comment + Subreddit_Name + Comment_Day +
##             Account_Age_days + Comment_Hour + User_Karma + Word_Count +
##             Parent_Score + Contains_Profanity + Contains_Question + Comment_Age_hours +
##             Contains_Emoji
##
##                                     Df Sum of Sq   RSS   AIC
## <none>                           99638 36735
## + Text_Length          1     3.2571 99635 36735
## + Sentiment_Score      1     0.8399 99638 36736

summary(step_model)

##
## Call:
## lm(formula = ihs_score ~ Is_Early_Comment + Subreddit_Name +
##     Comment_Day + Account_Age_days + Comment_Hour + User_Karma +
##     Word_Count + Parent_Score + Contains_Profanity + Contains_Question +
##     Comment_Age_hours + Contains_Emoji, data = df_cleaned)
##
## Residuals:
##     Min      1Q  Median      3Q      Max
## -9.9338 -0.8386 -0.1438  0.8108  7.7765
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               2.367e+00  5.095e-02 46.455 < 2e-16 ***
## Is_Early_Comment1         9.950e-01  2.146e-02 46.366 < 2e-16 ***

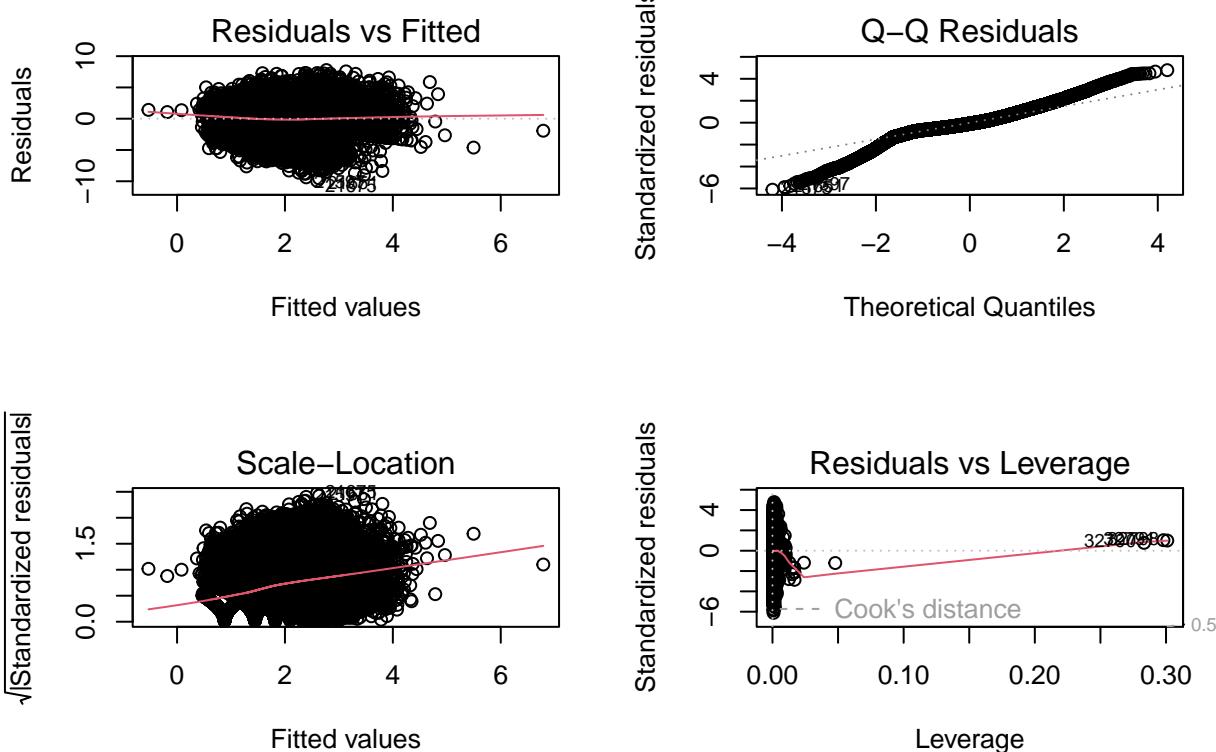
```

```

## Subreddit_Namebooks      -1.177e+00  5.779e-02 -20.371 < 2e-16 ***
## Subreddit_Namedatascience -1.257e+00  4.397e-02 -28.585 < 2e-16 ***
## Subreddit_NameFloridaMan -7.864e-01  6.116e-02 -12.858 < 2e-16 ***
## Subreddit_Namegaming     -1.242e+00  4.655e-02 -26.687 < 2e-16 ***
## Subreddit_NameHealth     -7.863e-01  5.446e-02 -14.439 < 2e-16 ***
## Subreddit_Namemovies     -8.641e-01  5.233e-02 -16.513 < 2e-16 ***
## Subreddit_NameShowerthoughts -9.620e-01  4.915e-02 -19.572 < 2e-16 ***
## Subreddit_Nametechnology -8.783e-01  3.250e-02 -27.024 < 2e-16 ***
## Subreddit_Nameunpopularopinion -1.796e+00  4.955e-02 -36.241 < 2e-16 ***
## Comment_Day1              -6.002e-01  3.667e-02 -16.369 < 2e-16 ***
## Comment_Day2              -1.369e-01  6.060e-02 -2.259  0.02388 *
## Comment_Day3              5.025e-02  5.457e-02  0.921  0.35714
## Comment_Day4              1.957e-01  4.178e-02  4.684  2.83e-06 ***
## Comment_Day5              -1.168e-02  3.342e-02 -0.349  0.72673
## Comment_Day6              1.785e-01  2.784e-02  6.411  1.47e-10 ***
## Account_Age_days          5.932e-05  5.266e-06 11.264 < 2e-16 ***
## Comment_Hour1             -3.384e-02  4.891e-02 -0.692  0.48912
## Comment_Hour2             1.681e-01  5.463e-02  3.076  0.00210 **
## Comment_Hour3             9.822e-02  5.549e-02  1.770  0.07674 .
## Comment_Hour4             -1.540e-01  5.736e-02 -2.685  0.00726 **
## Comment_Hour5             -5.594e-02  6.639e-02 -0.843  0.39944
## Comment_Hour6             1.034e-01  7.597e-02  1.361  0.17364
## Comment_Hour7             2.006e-01  7.737e-02  2.593  0.00953 **
## Comment_Hour8             5.053e-03  8.322e-02  0.061  0.95159
## Comment_Hour9             7.694e-02  8.062e-02  0.954  0.33992
## Comment_Hour10            1.023e-01  7.847e-02  1.303  0.19247
## Comment_Hour11            1.938e-01  7.280e-02  2.662  0.00778 **
## Comment_Hour12            4.901e-01  6.382e-02  7.678  1.65e-14 ***
## Comment_Hour13            2.684e-01  5.694e-02  4.713  2.46e-06 ***
## Comment_Hour14            1.493e-01  5.293e-02  2.820  0.00480 **
## Comment_Hour15            1.191e-01  4.973e-02  2.395  0.01662 *
## Comment_Hour16            2.370e-01  4.934e-02  4.804  1.56e-06 ***
## Comment_Hour17            1.552e-01  5.102e-02  3.042  0.00235 **
## Comment_Hour18            1.372e-01  5.062e-02  2.710  0.00674 **
## Comment_Hour19            1.097e-02  4.986e-02  0.220  0.82592
## Comment_Hour20            4.100e-02  4.945e-02  0.829  0.40699
## Comment_Hour21            -5.076e-02  4.974e-02 -1.021  0.30748
## Comment_Hour22            3.484e-02  5.060e-02  0.689  0.49107
## Comment_Hour23            3.180e-02  5.356e-02  0.594  0.55273
## User_Karma                5.862e-07  5.712e-08 10.263 < 2e-16 ***
## Word_Count                1.200e-03  2.123e-04  5.655  1.57e-08 ***
## Parent_Score               -6.554e-06  8.964e-07 -7.312  2.69e-13 ***
## Contains_Profanity1       1.130e-01  2.498e-02  4.523  6.13e-06 ***
## Contains_Question1        7.336e-02  1.878e-02  3.907  9.35e-05 ***
## Comment_Age_hours          -8.765e-05  3.369e-05 -2.602  0.00928 **
## Contains_Emoji1           -1.210e-01  5.496e-02 -2.201  0.02772 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.626 on 37709 degrees of freedom
## Multiple R-squared:  0.1183, Adjusted R-squared:  0.1172
## F-statistic: 107.7 on 47 and 37709 DF,  p-value: < 2.2e-16

```

```
par(mfrow = c(2, 2))
plot(step_model)
```



```
par(mfrow = c(1, 1))
```

GAM (2)

```
gam_model <- gam(ihs_score ~
  s(Comment_Age_hours) +
  s(User_Karma) +
  s(Word_Count) +
  s(Account_Age_days) +
  s(Sentiment_Score) +
  s(Text_Length) +
  s(Parent_Score) +
  Is_Early_Comment +
  Subreddit_Name +
  Comment_Day +
  Comment_Hour +
  Contains_Question +
  Contains_Emoji +
  Contains_Profanity,
  data = df_cleaned,
  select = TRUE)
```

```
summary(gam_model)
```

```
##  
## Family: gaussian
```

```

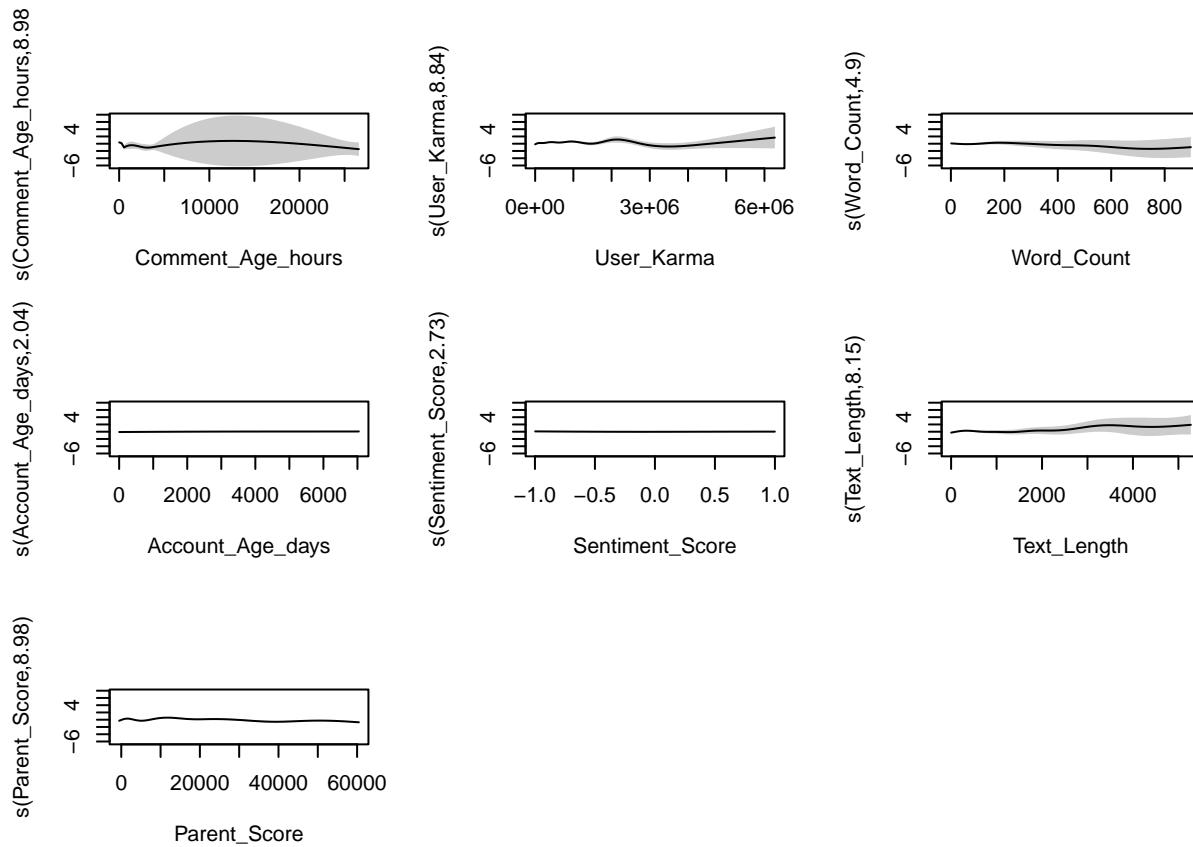
## Link function: identity
##
## Formula:
## ihs_score ~ s(Comment_Age_hours) + s(User_Karma) + s(Word_Count) +
##           s(Account_Age_days) + s(Sentiment_Score) + s(Text_Length) +
##           s(Parent_Score) + Is_Early_Comment + Subreddit_Name + Comment_Day +
##           Comment_Hour + Contains_Question + Contains_Emoji + Contains_Profanity
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.548822  0.049566 51.423 < 2e-16 ***
## Is_Early_Comment1          0.611499  0.023986 25.494 < 2e-16 ***
## Subreddit_Namebooks        -0.685955  0.058687 -11.688 < 2e-16 ***
## Subreddit_Namedatascience -0.710941  0.046308 -15.352 < 2e-16 ***
## Subreddit_NameFloridaMan   -0.257527  0.062489 -4.121 3.78e-05 ***
## Subreddit_Namegaming       -1.051062  0.046557 -22.576 < 2e-16 ***
## Subreddit_NameHealth       -0.416606  0.054699 -7.616 2.67e-14 ***
## Subreddit_Namemovies       -0.752069  0.051704 -14.546 < 2e-16 ***
## Subreddit_NameShowerthoughts -0.455162  0.050609 -8.994 < 2e-16 ***
## Subreddit_Nametechnology   -0.715357  0.032483 -22.023 < 2e-16 ***
## Subreddit_Nameunpopularopinion -1.635138  0.049407 -33.095 < 2e-16 ***
## Comment_Day1               -0.666891  0.036505 -18.269 < 2e-16 ***
## Comment_Day2               -0.260332  0.059691 -4.361 1.30e-05 ***
## Comment_Day3               -0.158676  0.053993 -2.939 0.003297 **
## Comment_Day4               -0.042052  0.042048 -1.000 0.317278
## Comment_Day5               -0.132572  0.033493 -3.958 7.57e-05 ***
## Comment_Day6               0.070347  0.028285  2.487 0.012884 *
## Comment_Hour1              -0.009843  0.047891 -0.206 0.837164
## Comment_Hour2              0.210029  0.053516  3.925 8.70e-05 ***
## Comment_Hour3              0.180999  0.054381  3.328 0.000875 ***
## Comment_Hour4              -0.113807  0.056215 -2.025 0.042925 *
## Comment_Hour5              0.044890  0.064984  0.691 0.489707
## Comment_Hour6              0.168761  0.074356  2.270 0.023235 *
## Comment_Hour7              0.289598  0.075660  3.828 0.000130 ***
## Comment_Hour8              0.126358  0.081415  1.552 0.120666
## Comment_Hour9              0.158567  0.078833  2.011 0.044288 *
## Comment_Hour10             0.266472  0.076839  3.468 0.000525 ***
## Comment_Hour11             0.287159  0.071287  4.028 5.63e-05 ***
## Comment_Hour12             0.512113  0.062578  8.184 2.84e-16 ***
## Comment_Hour13             0.297191  0.055765  5.329 9.91e-08 ***
## Comment_Hour14             0.155230  0.051821  2.996 0.002742 **
## Comment_Hour15             0.154662  0.048728  3.174 0.001505 **
## Comment_Hour16             0.210549  0.048436  4.347 1.38e-05 ***
## Comment_Hour17             0.116347  0.050006  2.327 0.019988 *
## Comment_Hour18             0.135241  0.049588  2.727 0.006388 **
## Comment_Hour19             0.014654  0.048940  0.299 0.764615
## Comment_Hour20             0.046126  0.048488  0.951 0.341459
## Comment_Hour21             -0.025190  0.048767 -0.517 0.605483
## Comment_Hour22             0.052457  0.049619  1.057 0.290426
## Comment_Hour23             0.030496  0.052429  0.582 0.560800
## Contains_Question1         -0.009240  0.020798 -0.444 0.656837
## Contains_Emoji1            -0.108283  0.053874 -2.010 0.044447 *
## Contains_Profanity1        0.071750  0.025373  2.828 0.004689 **
## ---

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df      F  p-value
## s(Comment_Age_hours) 8.982      9 132.155 < 2e-16 ***
## s(User_Karma)        8.836      9  32.686 < 2e-16 ***
## s(Word_Count)        4.895      9   0.974  0.0727 .
## s(Account_Age_days)  2.045      9   3.257 < 2e-16 ***
## s(Sentiment_Score)   2.726      9   0.907  0.0225 *
## s(Text_Length)       8.148      9   4.184 2.71e-06 ***
## s(Parent_Score)      8.978      9  39.359 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.158  Deviance explained =  16%
## GCV = 2.5258  Scale est. = 2.5199    n = 37757
plot(gam_model, pages = 1, shade = TRUE)

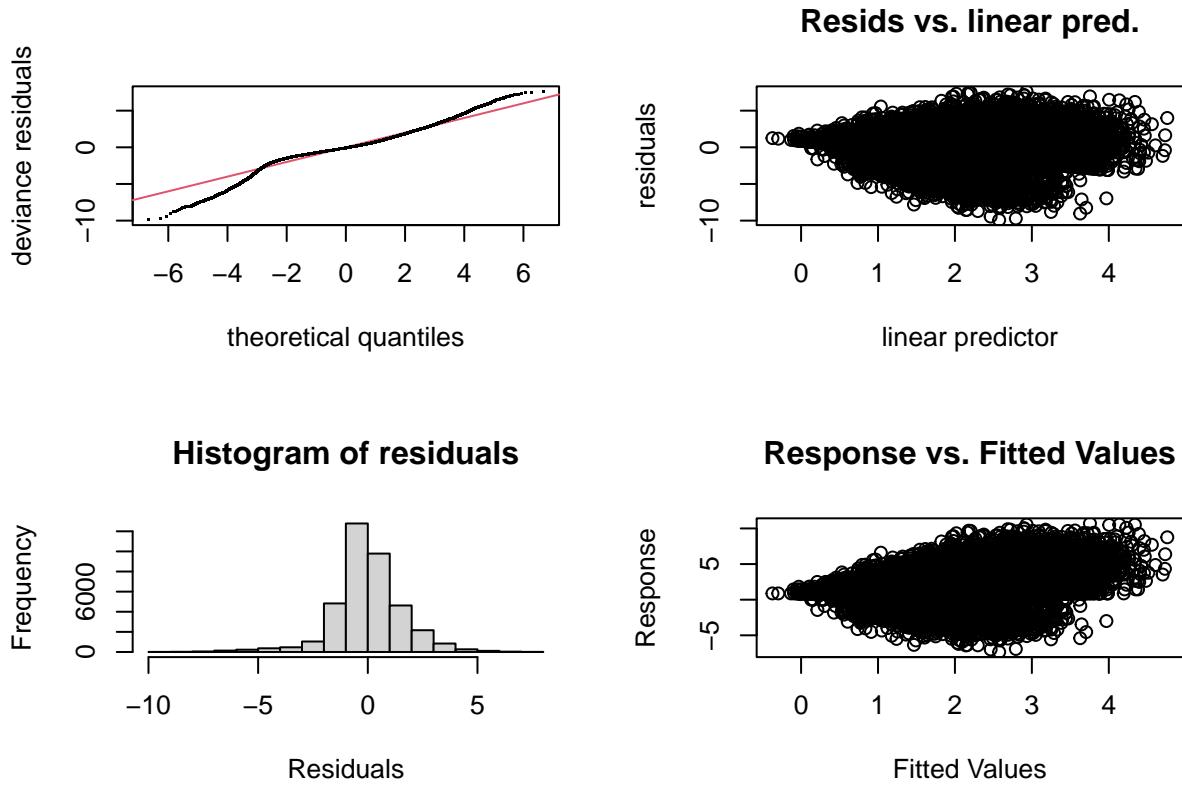
```



```

par(mfrow=c(2,2))
gam.check(gam_model)

```



```

## 
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 40 iterations.
## The RMS GCV score gradient at convergence was 4.390372e-06 .
## The Hessian was not positive definite.
## Model rank = 106 / 106
## 
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
## 
##          k'  edf k-index p-value
## s(Comment_Age_hours) 9.00 8.98    0.97  0.005 ***
## s(User_Karma)        9.00 8.84    0.96 <2e-16 ***
## s(Word_Count)        9.00 4.90    0.99  0.185
## s(Account_Age_days) 9.00 2.04    1.00  0.345
## s(Sentiment_Score)  9.00 2.73    0.99  0.300
## s(Text_Length)       9.00 8.15    1.00  0.425
## s(Parent_Score)      9.00 8.98    0.91 <2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(mfrow=c(1,1))

```