DS710 - Final Project

Jared Balkman

12/13/2020

```r
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE, dev = 'png')
```

TWITTER SENTIMENT ANALYSIS: POLARITY VS POPULARITY This project aims to investigate what, if any, relationship exists between the polarity, or positive/negative sentiment of a tweet, and that tweet's popularity as measured by likes and retweets.

First we'll load some libraries:

```r
library(readr)
library(dplyr)
library(ggformula)
library(car)
library(coefplot)
library(ggpubr)
```

The data to be analyzed consist of 9149 unique observations of 5 variables:

num_retweets: the number of retweets of an original tweet

num_likes: the number of times an original tweet was liked (aka favorited)

num_followers: the number of followers of the author of the original tweet

subjectivity: the subjectivity of the text of a tweet, as calculated by the textblob.subjectivity function. Values in range (0.0, 1.0) (0=objective, 1=subjective)

polarity: the polarity of the text of a tweet, as calculated by the SentimentIntensityAnalyzer.polarity_scores function from the vaderSentiment Python package. The function calculates separate scores for positive, neutral, and negative aspects of a tweet, and averages them to compute one compound score representing overall sentiment of a tweet. This compound score is what was used here. Values in range (-0.9884, 0.9992) (-1=negative, 0=neutral, 1=positive)

5546 rows came from original tweets. 3603 came from retweets, but in these cases the original tweet data for all attributes was collected from the 'retweeted_status' dictionary in the tweet Status object.

Load the dataset:

```r
data <- read_csv('twitter_data_for_r.csv')
head(data, 10)
```

```
## # A tibble: 10 x 6
##        X1 num_retweets num_likes num_followers subjectivity polarity
##     <dbl>        <dbl>     <dbl>         <dbl>        <dbl>    <dbl>
##   1     0            1        15          4654        0        0
##   2     1            1         2          3019        0.5      0
##   3     2            1         3           365        0.8     -0.751
##   4     3          870      2800        171687        0.440   -0.34
##   5     4            0         0             0        0.6     -0.542
```

```
## 6     5          0       0        131      0       0
## 7     6       3088    9975     131734      0.323   0.802
## 8     7          0       1       4676      0       0
## 9     8       7193  131552       5982      1       0.459
## 10    9        390    1474       1155      0.2     0.402
```

The tweets are created first (at which point, theoretically, they would take on their polarity and subjectivity scores), and the likes and retweets happen after, so for our analysis we want retweets and likes to be our dependent variables, and polarity and subjectivity our independent variables. Follower count we may wish to use as a control variable, since it may have an effect on number of likes and retweets. Since I am not confident performing multivariate regression (which would be called for since we have multiple independent and multiple dependent variables), I will attempt separate multiple regressions for likes and retweets.

My hypotheses are as follows:

H_0: There is no relationship between sentiment (measured by polarity and subjectivity) of a tweet and the popularity (measured by likes and retweets) of the tweet.

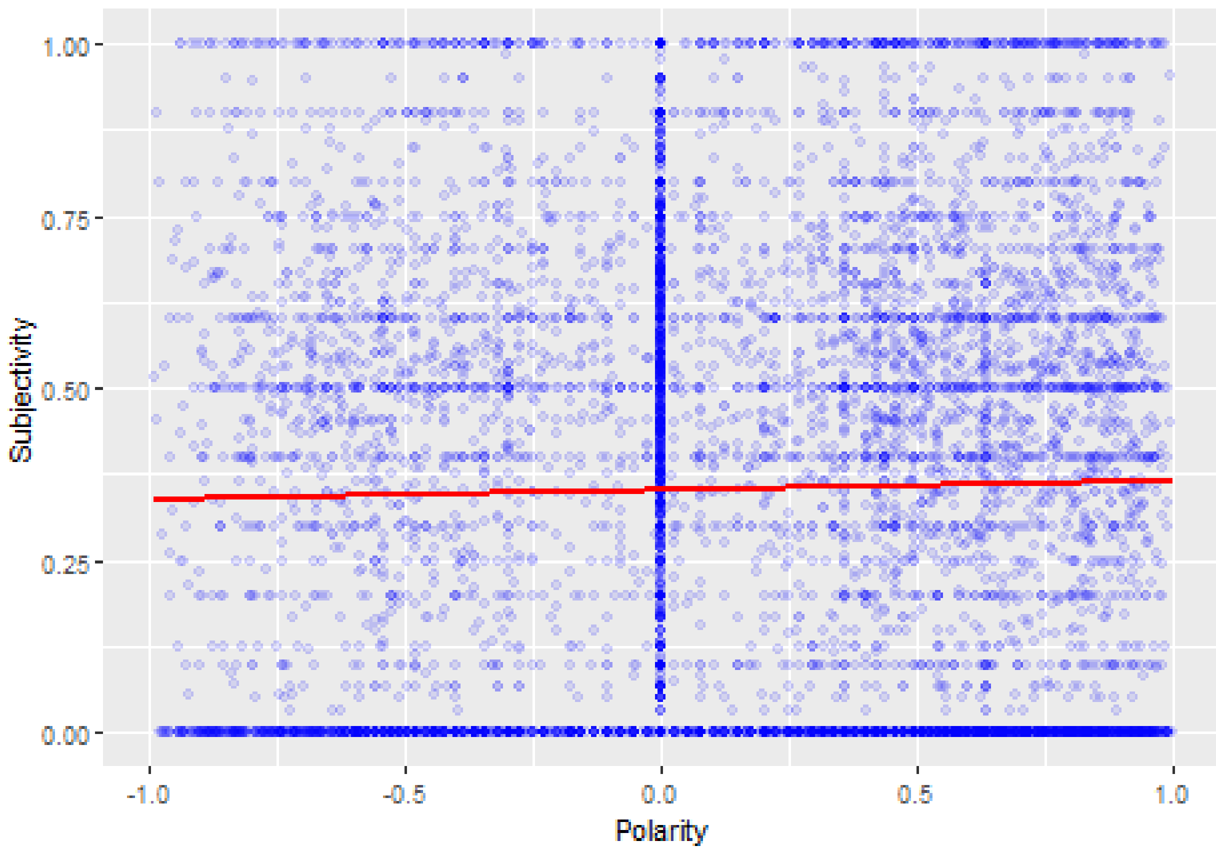
H_1: There is a statistically significant relationship between tweet sentiment and popularity.

We should first see if the assumption for multiple linear regression are satisfied. These assumptions include:

1) Linearity of relationship between y variable and x variables
2) Independence of x variables
3) Normal distribution of residuals
4) Homoscedasticity (variance of dependent variable remains constant as ind. var. changes)

First, it seems that it is rather difficult to test for the independence of two continuous variables; I couldn't find a good way to do this without binning the data, which seemed kind of artificial. We could look at a scatter plot with some correlation information to get an idea:

```
data %>%
  gf_point(subjectivity ~ polarity, alpha = 0.1, color = "blue",
           xlab = "Polarity", ylab = "Subjectivity") %>%
  gf_smooth(subjectivity ~ polarity, color = 'red')
```



And a correlation test:

```
cor.test(data$polarity, data$subjectivity)
```

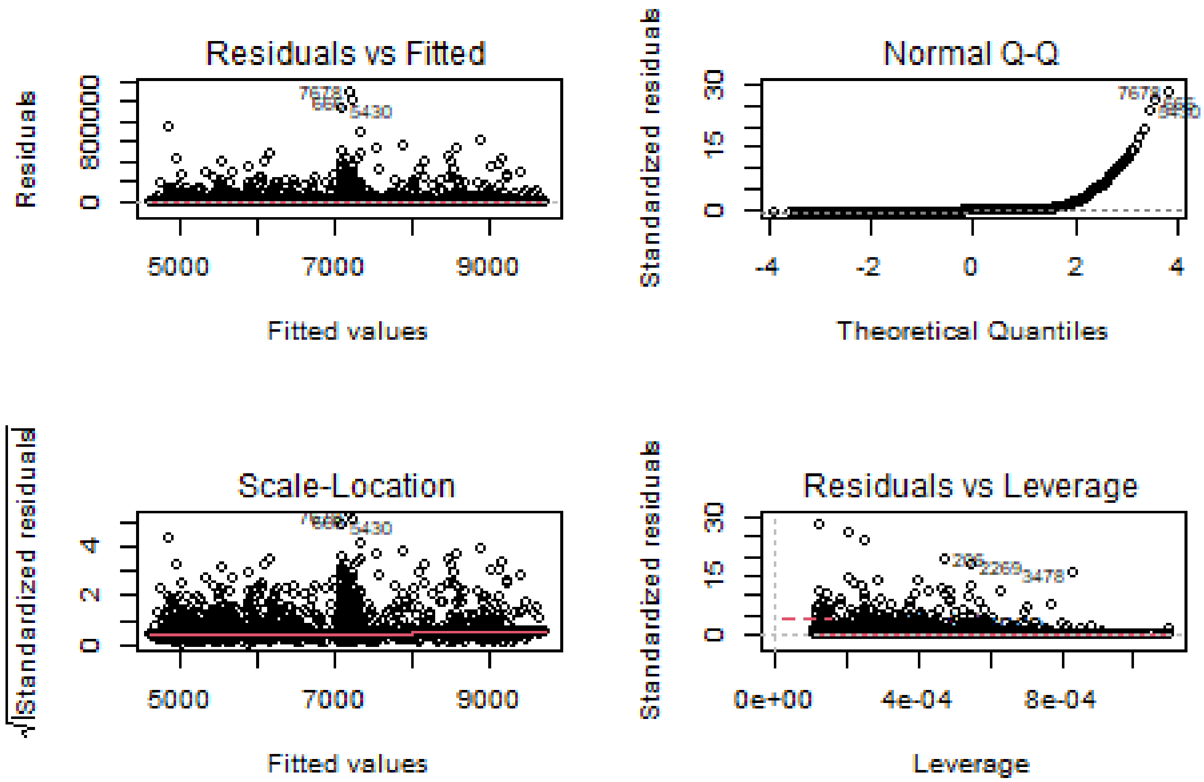
```
##
##  Pearson's product-moment correlation
##
## data:  data$polarity and data$subjectivity
## t = 1.9123, df = 9147, p-value = 0.05586
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0005003261  0.0404661080
## sample estimates:
##        cor
## 0.01999128
```

So we have a significant correlation, but it's awfully weak. Note also we have a lot of 0- and 1-values for subjectivity as well as 0-values for polarity, something I will look at more below.

We can examine assumptions 1), 3), and 4) using the plot() function, and also look at the coefficients:

```
model_likes <- lm(num_likes ~ polarity + subjectivity, data = data)

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



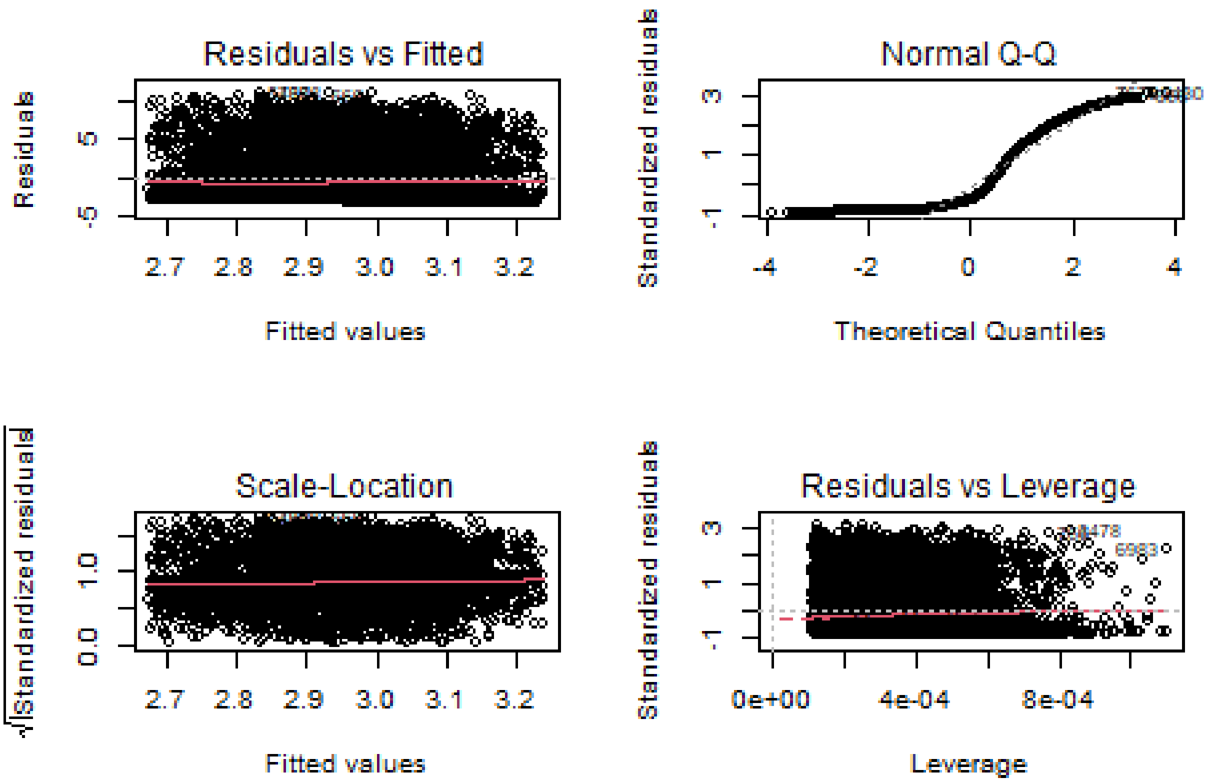

```r
summary(model_likes)
```

```
##
## Call:
## lm(formula = num_likes ~ polarity + subjectivity, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##   -9721   -7245   -6335   -5206 1091472
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7102.0      611.9  11.606  < 2e-16 ***
## polarity      -2449.9      838.6  -2.921  0.00349 **
## subjectivity    236.7     1248.3   0.190  0.84960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38910 on 9146 degrees of freedom
## Multiple R-squared:  0.0009342,  Adjusted R-squared:  0.0007157
## F-statistic: 4.276 on 2 and 9146 DF,  p-value: 0.01392
```

These plots indicate a transformation is in order. We'll try a log(x+1) transformation of the dependent variable to account for the exponential-looking skewing (and values of 0):

```r
model_likes <- lm(log1p(num_likes) ~ polarity + subjectivity, data = data)
```

```r
par(mfrow = c(2,2))
plot(model_likes)
```




```r
summary(model_likes)
```

```
##
## Call:
## lm(formula = log1p(num_likes) ~ polarity + subjectivity, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.228 -2.833 -1.815  2.389 10.978
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.83276    0.05524  51.281   <2e-16 ***
## polarity      0.16044    0.07570   2.119   0.0341 *
## subjectivity  0.24767    0.11269   2.198   0.0280 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.512 on 9146 degrees of freedom
## Multiple R-squared:  0.001039,   Adjusted R-squared:  0.0008204
## F-statistic: 4.756 on 2 and 9146 DF,  p-value: 0.008624
```

This looks better, but we might be violating the normality assumption. We can try adding in number of followers as a control variable since it would certainly seem to affect number of likes and retweets. First we can look at correlations of dependent variables with num_followers:

```r
cor.test(data$num_followers, log1p(data$num_likes))
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$num_followers and log1p(data$num_likes)
## t = 16.637, df = 9147, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1514188 0.1911984
## sample estimates:
##       cor
## 0.1713785
```

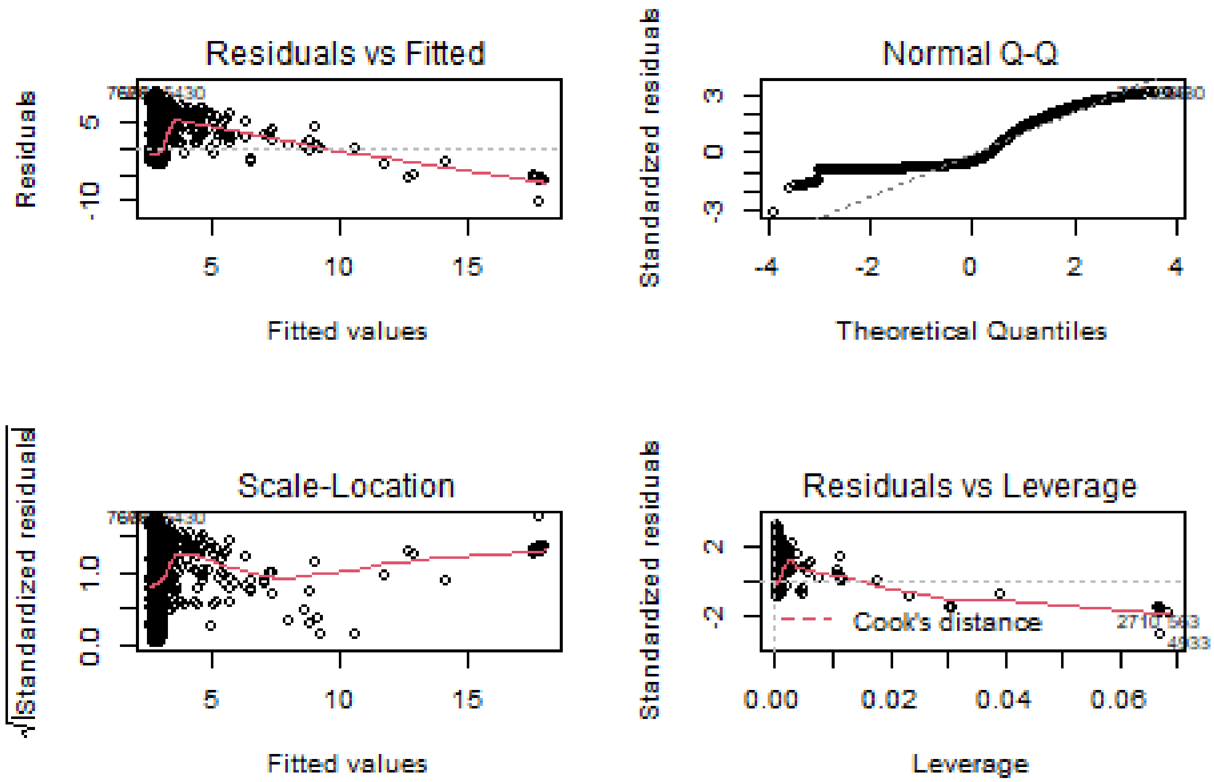
```r
cor.test(data$num_followers, log1p(data$num_retweets))
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$num_followers and log1p(data$num_retweets)
## t = 16.255, df = 9147, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1475744 0.1874071
## sample estimates:
##       cor
## 0.1675591
```

Both correlations with number of followers are significant with p-values near zero, so we should introduce this into the model:

```r
model_likes <- lm(log1p(num_likes) ~ polarity + subjectivity + num_followers, data = data)

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



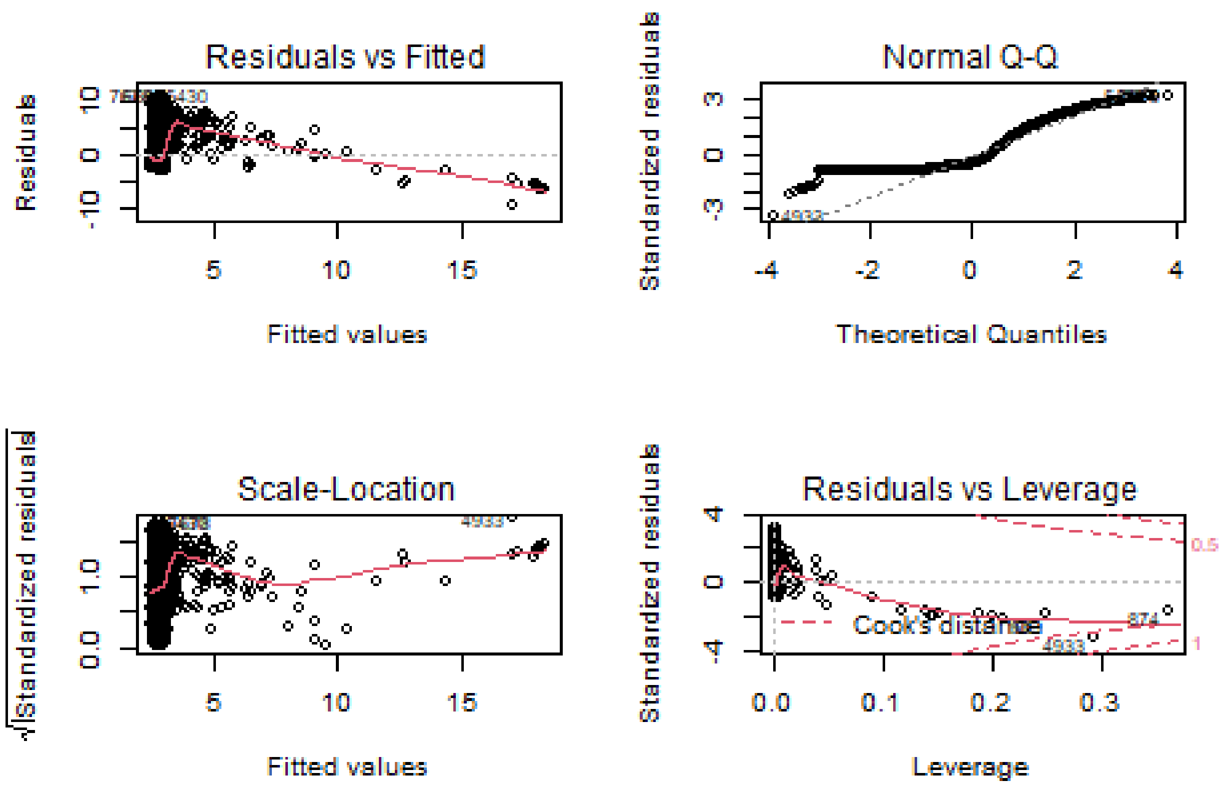

```r
summary(model_likes)
```

```
##
## Call:
## lm(formula = log1p(num_likes) ~ polarity + subjectivity + num_followers,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.421  -2.785  -1.773   2.349  11.030
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.785e+00  5.450e-02  51.099   <2e-16 ***
## polarity      1.516e-01  7.459e-02   2.033   0.0421 *
## subjectivity  2.357e-01  1.110e-01   2.123   0.0338 *
## num_followers 1.680e-07  1.011e-08  16.613   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.461 on 9145 degrees of freedom
## Multiple R-squared:  0.03031,    Adjusted R-squared:  0.02999
## F-statistic: 95.27 on 3 and 9145 DF,  p-value: < 2.2e-16
```

The coefficients are still significant, but the diagnostic plots look a little strange. We'll try one more model that accounts for possible interactions between independent variables:

```r
model_likes <- lm(log1p(num_likes) ~ polarity + subjectivity + num_followers +
                  polarity:subjectivity + polarity:num_followers +
                  subjectivity:num_followers, data = data)

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




```r
summary(model_likes)
```

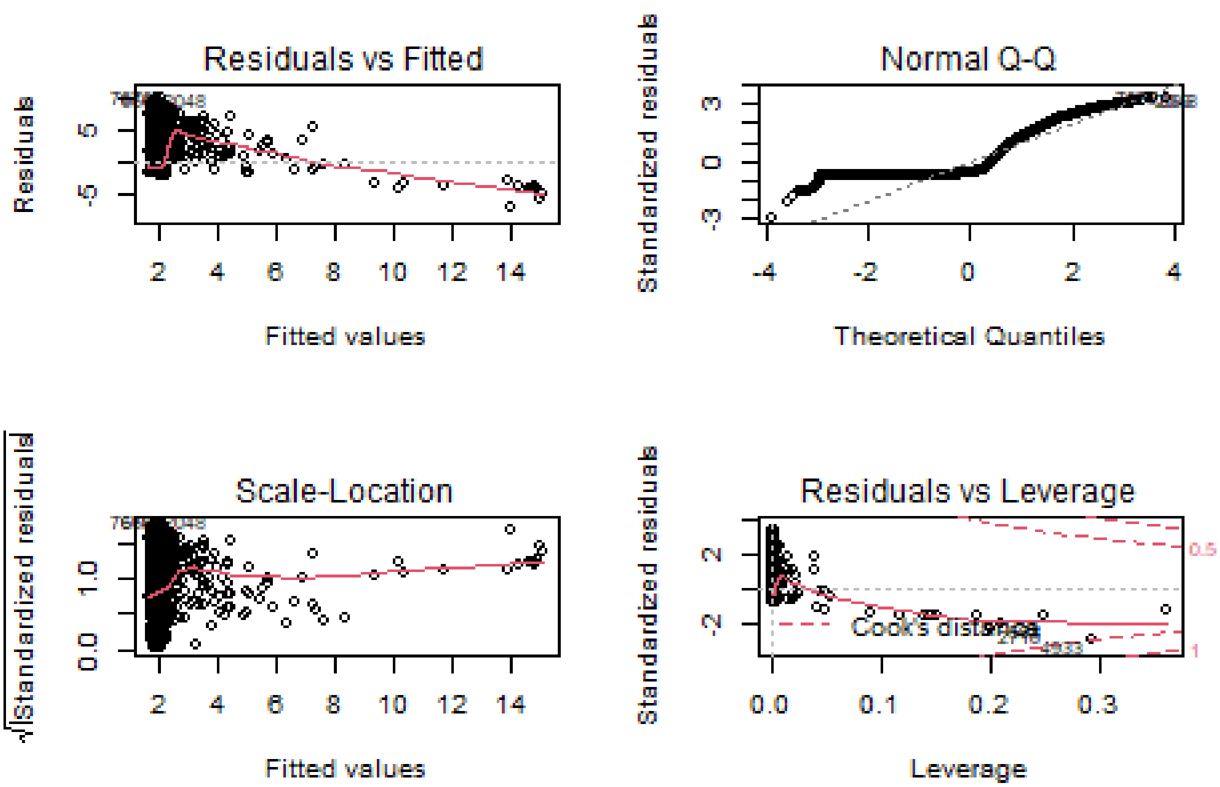
```
##
## Call:
## lm(formula = log1p(num_likes) ~ polarity + subjectivity + num_followers +
##     polarity:subjectivity + polarity:num_followers + subjectivity:num_followers,
##     data = data)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -9.628 -2.766 -1.797  2.359 11.026
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.766e+00  5.581e-02  49.561   <2e-16 ***
## polarity                   2.778e-01  1.103e-01   2.519   0.0118 *
## subjectivity               2.946e-01  1.167e-01   2.525   0.0116 *
## num_followers              1.721e-07  1.523e-08  11.299   <2e-16 ***
## polarity:subjectivity     -3.636e-01  2.305e-01  -1.578   0.1147
```

```
## polarity:num_followers      5.344e-09  2.001e-08   0.267   0.7894
## subjectivity:num_followers -1.346e-08  2.875e-08  -0.468   0.6396
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.461 on 9142 degrees of freedom
## Multiple R-squared:  0.0306, Adjusted R-squared:  0.02996
## F-statistic: 48.09 on 6 and 9142 DF,  p-value: < 2.2e-16
```

Here the diagnostic plots are pretty similar to the previous model. All the independent variables remain significant, and there are also significant interactions (subjectivity:num_followers is perhaps interesting). Let's proceed with retweets using this same model and see if there's anything different:

```r
model_retweets <- lm(log1p(num_retweets) ~ polarity + subjectivity + num_followers +
                     polarity:subjectivity + polarity:num_followers +
                     subjectivity:num_followers, data = data)

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




```r
summary(model_retweets)
```

```
##
## Call:
## lm(formula = log1p(num_retweets) ~ polarity + subjectivity +
##     num_followers + polarity:subjectivity + polarity:num_followers +
##     subjectivity:num_followers, data = data)
##
```

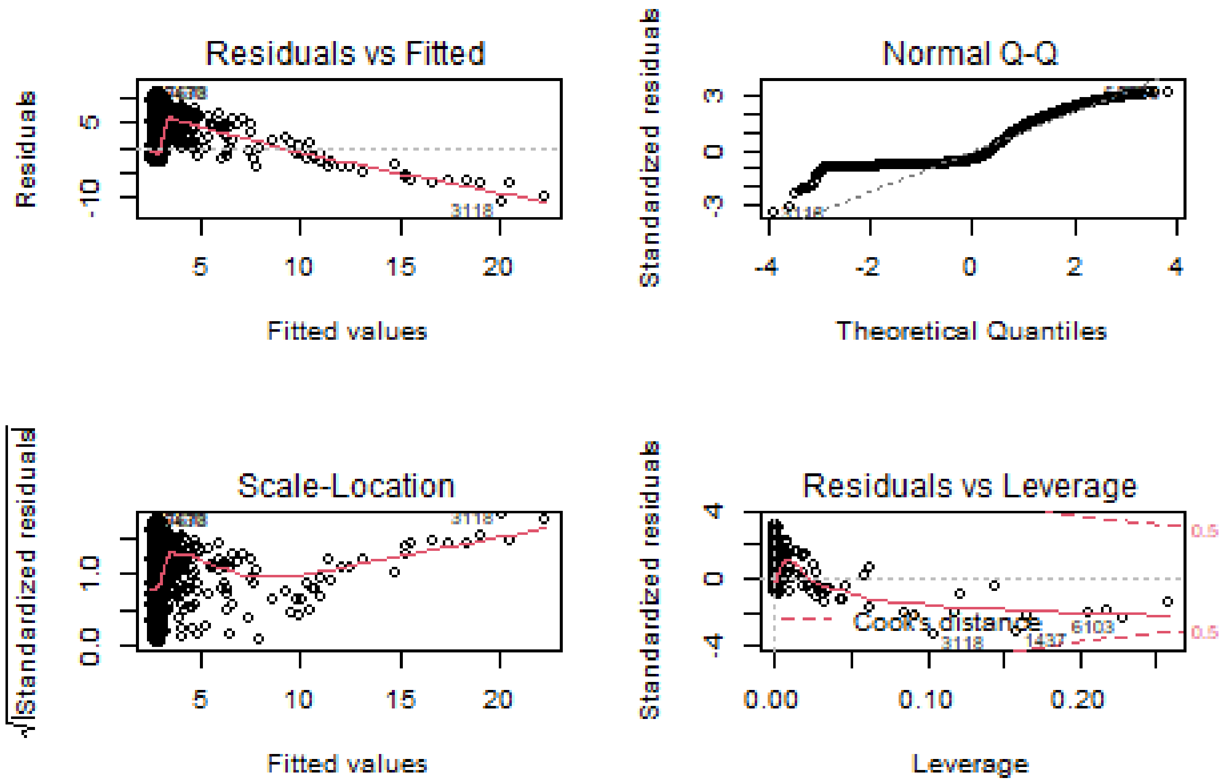
```
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.321 -2.086 -1.902  1.935  9.992
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.923e+00  4.797e-02  40.083  < 2e-16 ***
## polarity                     2.665e-01  9.480e-02   2.811  0.00494 **
## subjectivity                 2.787e-01  1.003e-01   2.778  0.00548 **
## num_followers                1.441e-07  1.309e-08  11.011  < 2e-16 ***
## polarity:subjectivity       -3.221e-01  1.981e-01  -1.626  0.10403
## polarity:num_followers       4.983e-09  1.720e-08   0.290  0.77209
## subjectivity:num_followers -1.062e-08  2.471e-08  -0.430  0.66752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.975 on 9142 degrees of freedom
## Multiple R-squared:  0.02963,    Adjusted R-squared:  0.02899
## F-statistic: 46.52 on 6 and 9142 DF,  p-value: < 2.2e-16
```

Perhaps not surprisingly, we have similar results for retweet count. One last thing to try is removing a few of the outliers. The rows below constitute three rounds of outlier identification and removal based on which points displayed their row numbers in the various diagnostic plots:

```r
data[c(18, 167, 563, 874, 1463, 1472, 1883, 2710, 2266, 4933, 6863, 7406),] = NA

model_likes <- lm(log1p(num_likes) ~ polarity + subjectivity + num_followers +
                  polarity:subjectivity + polarity:num_followers +
                  subjectivity:num_followers, data = data)

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



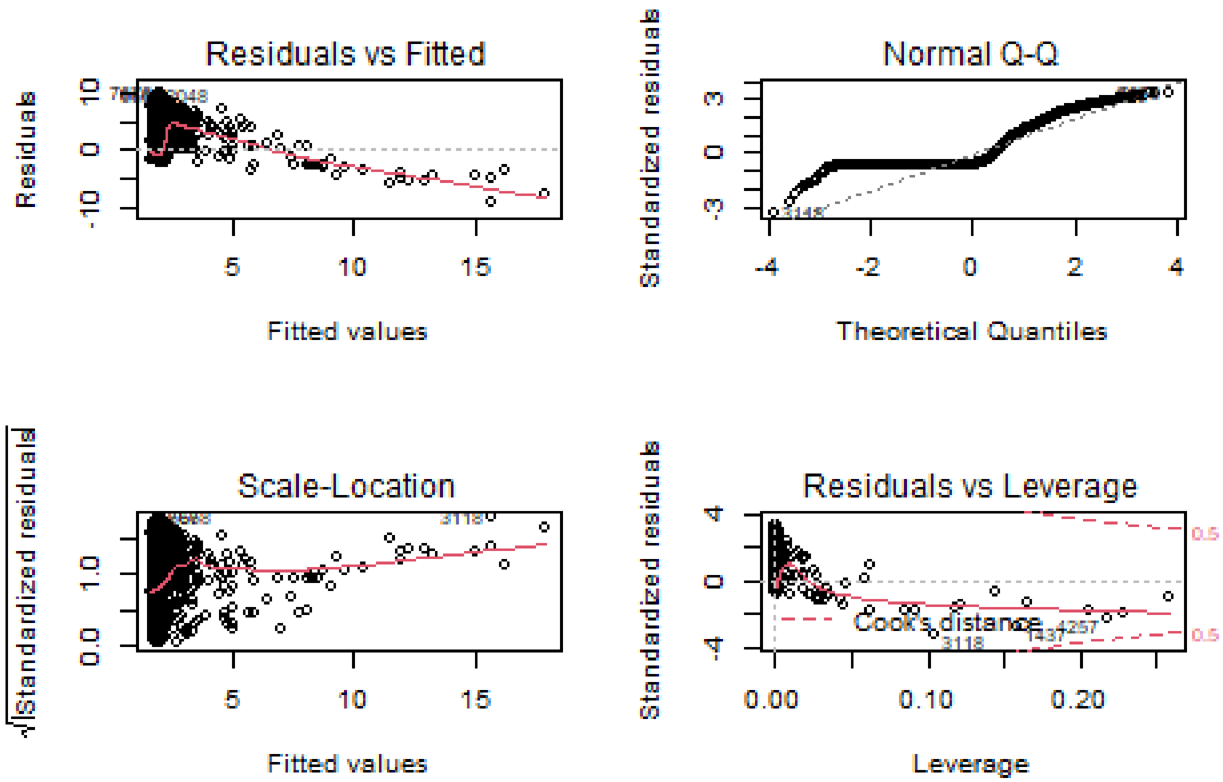

```r
summary(model_likes)
```

```
##
## Call:
## lm(formula = log1p(num_likes) ~ polarity + subjectivity + num_followers +
##     polarity:subjectivity + polarity:num_followers + subjectivity:num_followers,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.002  -2.765  -1.779   2.292  11.054
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.765e+00  5.548e-02  49.834  < 2e-16 ***
## polarity                    2.767e-01  1.098e-01   2.521   0.0117 *
## subjectivity                2.252e-01  1.162e-01   1.937   0.0527 .
## num_followers               1.765e-07  1.767e-08   9.989  < 2e-16 ***
## polarity:subjectivity      -4.051e-01  2.291e-01  -1.768   0.0771 .
## polarity:num_followers      4.821e-08  3.819e-08   1.262   0.2069
## subjectivity:num_followers  4.437e-07  6.068e-08   7.313 2.84e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.439 on 9130 degrees of freedom
##   (12 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.03801,    Adjusted R-squared:  0.03738
## F-statistic: 60.13 on 6 and 9130 DF,  p-value: < 2.2e-16
```

```
model_retweets <- lm(log1p(num_retweets) ~ polarity + subjectivity + num_followers +
                    polarity:subjectivity + polarity:num_followers +
                    subjectivity:num_followers, data = data)

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




```
summary(model_retweets)
```

```
##
## Call:
## lm(formula = log1p(num_retweets) ~ polarity + subjectivity +
##     num_followers + polarity:subjectivity + polarity:num_followers +
##     subjectivity:num_followers, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.257 -2.063 -1.902  1.930 10.013
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.922e+00  4.776e-02  40.245  < 2e-16 ***
## polarity                     2.632e-01  9.451e-02   2.785  0.00536 **
## subjectivity                 2.268e-01  1.001e-01   2.266  0.02349 *
```

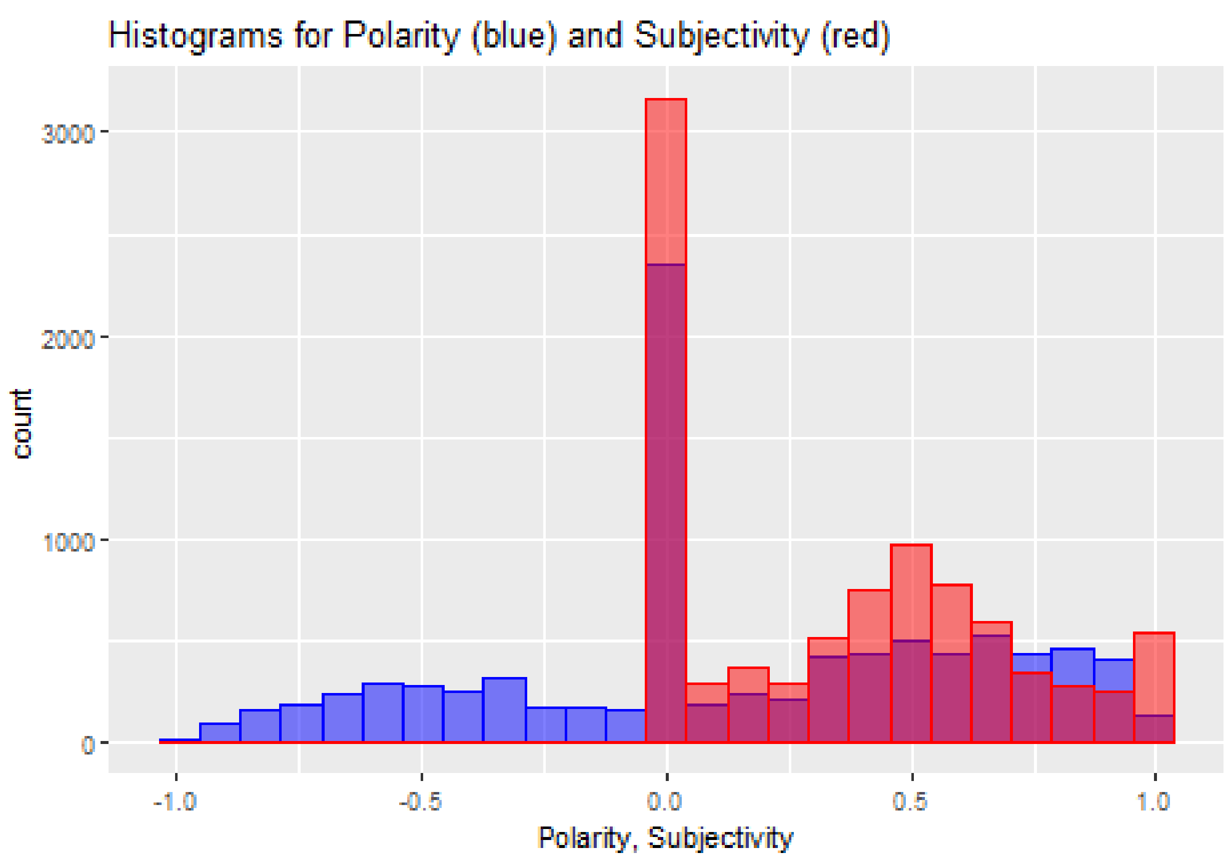
```
## num_followers              1.476e-07  1.521e-08   9.704  < 2e-16 ***
## polarity:subjectivity      -3.524e-01  1.973e-01  -1.786  0.07407 .
## polarity:num_followers      4.936e-08  3.288e-08   1.501  0.13336
## subjectivity:num_followers  3.280e-07  5.224e-08   6.279 3.57e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.961 on 9130 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.03362,    Adjusted R-squared:  0.03298
## F-statistic: 52.94 on 6 and 9130 DF,  p-value: < 2.2e-16
```

The profile of significant variables stays the same, with number of followers and intercept significant at the highest level (i.e. lowest p-value) as subjectivity:num_followers. We have evidence to reject the null hypothesis and claim that there is a statistically significant relationship between the sentiment of a tweet and its popularity.

On the other hand, we have some visual evidence in the diagnostic plots that complicates matters somewhat. We haven't really been able to get a good Q-Q plot, so we might be violating the normality assumption here. It's possible

So what about all those zeros for polarity and subjectivity (and ones for subjectivity)?

```r
data %>%
  gf_histogram(~polarity, color = 'blue', fill = 'blue') %>%
  gf_histogram(~subjectivity, color= 'red', fill = 'red') %>%
  gf_labs(title = 'Histograms for Polarity (blue) and Subjectivity (red)',
          x='Polarity, Subjectivity', y='count')
```



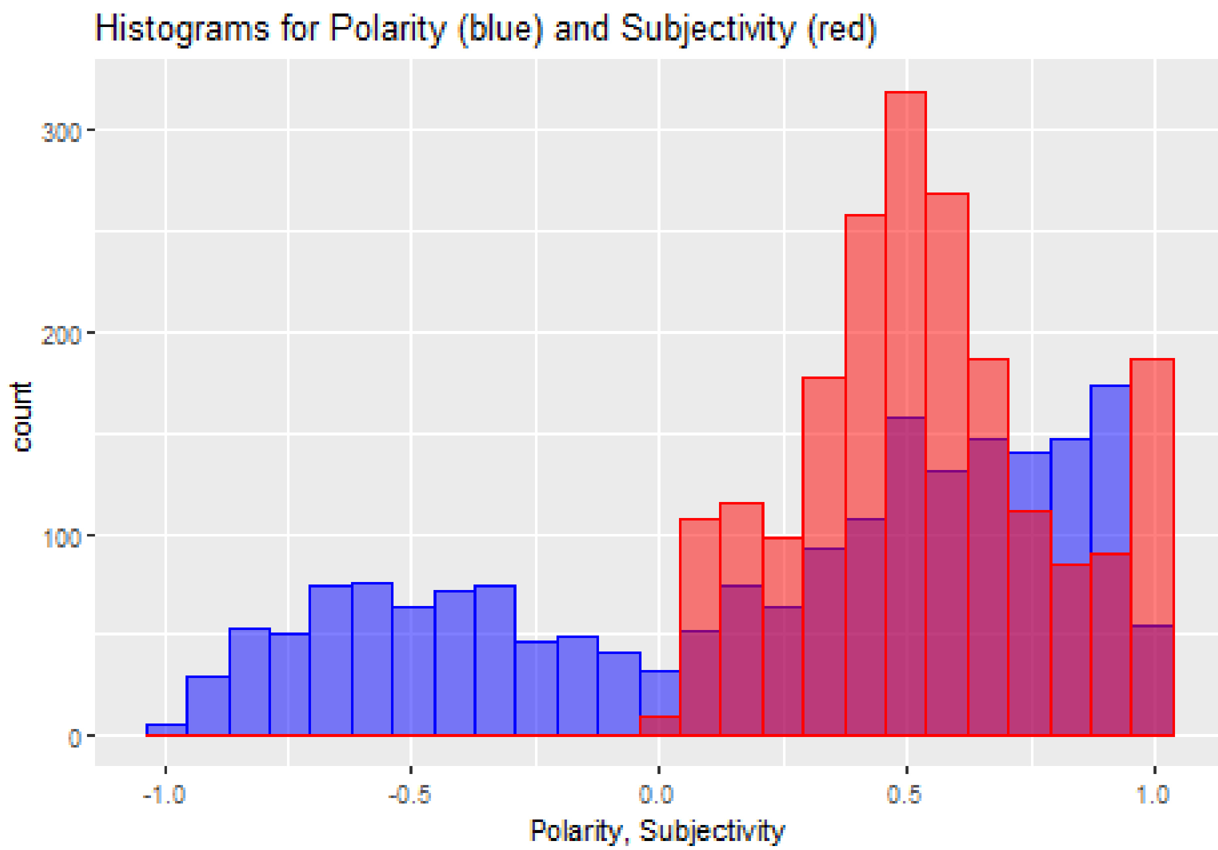
On the one hand, it doesn't seem right to remove them, since they are perfectly valid data points, tweets that were or weren't liked or retweeted with a polarity or subjectivity score happening to equal zero. On the other hand, we could make the argument that we aren't really interested as much in neutral tweets, but rather only the ones with some non-zero sentiment value. It's worth at least exploring (we'll hold off on the 1-values for subjectivity though).

```
non_zero_rows <- apply(data, 1, function(polarity) all(polarity != 0))
non_zero_rows <- apply(data, 1, function(subjectivity) all(subjectivity != 0))

non_zero_data <- data[non_zero_rows,]
non_zero_data
```

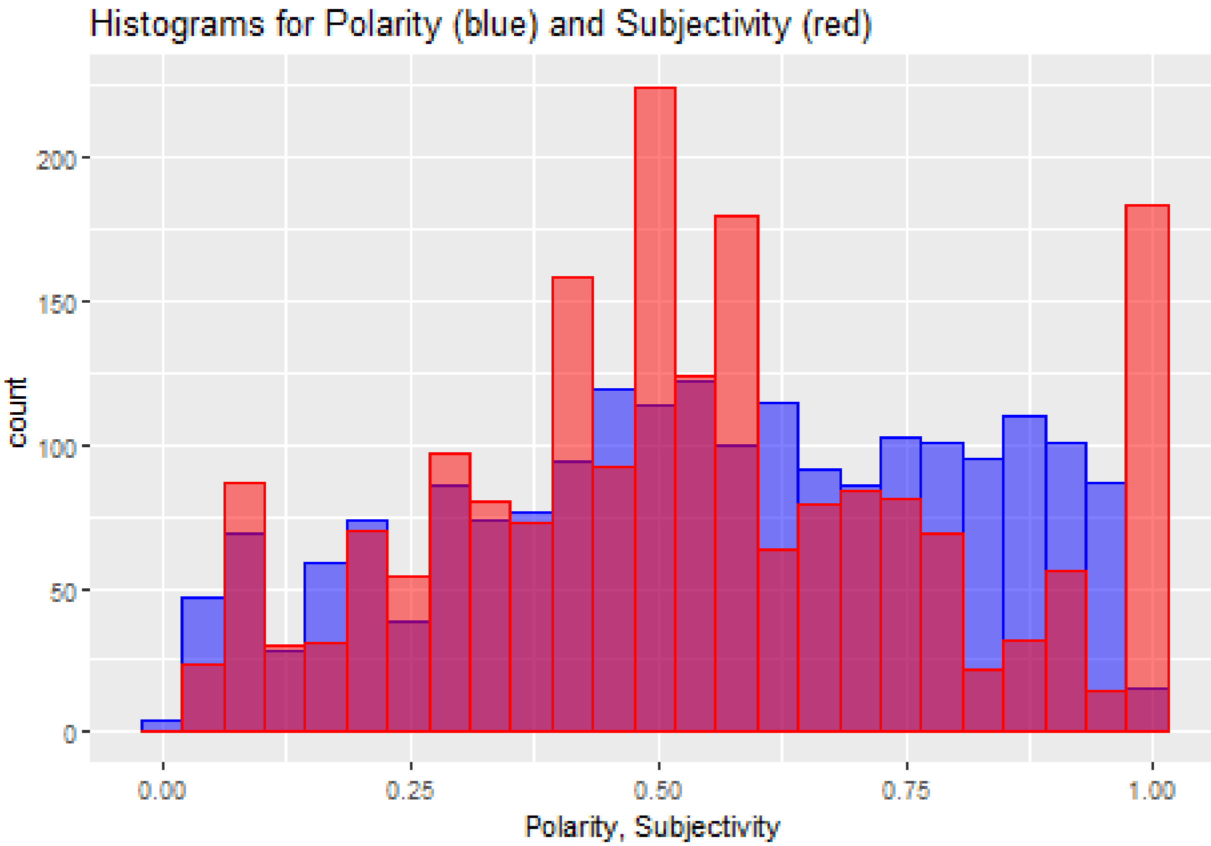
```
## # A tibble: 2,021 x 6
##       X1 num_retweets num_likes num_followers subjectivity polarity
##    <dbl>        <dbl>     <dbl>         <dbl>        <dbl>    <dbl>
##  1     2            1         3           365        0.8     -0.751
##  2     3          870      2800        171687        0.440   -0.34
##  3     6         3088      9975        131734        0.323    0.802
##  4     8         7193    131552          5982        1        0.459
##  5     9          390      1474          1155        0.2      0.402
##  6    12         1158      7502        502313        0.375   -0.103
##  7    15         3534      8211         51792        0.5      0.927
##  8    NA           NA        NA            NA       NA       NA
##  9    22          103       230           889        0.357    0.649
## 10    30            2        22           468        0.544    0.872
## # ... with 2,011 more rows
```

```
non_zero_data %>%
  gf_histogram(~polarity, color = 'blue', fill = 'blue') %>%
  gf_histogram(~subjectivity, color= 'red', fill = 'red') %>%
  gf_labs(title = 'Histograms for Polarity (blue) and Subjectivity (red)',
          x='Polarity, Subjectivity', y='count')
```



This definitely changes things. First, we lost over 7000 rows. Subjectivity looks like it could come from a normal distribution (aside from the excess 1-values), and polarity looks bimodal. But, since it looks symmetrical about 0, it might look pretty normal if we took the absolute value:

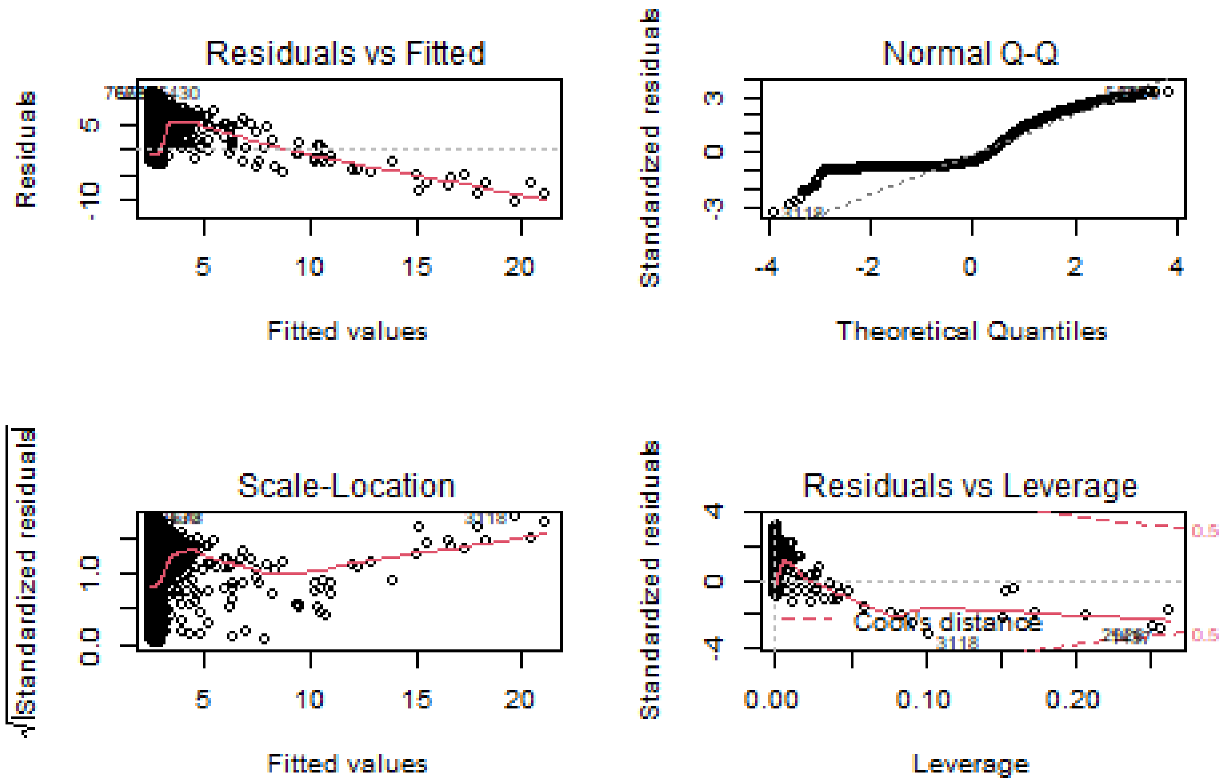
```
non_zero_data %>%
  gf_histogram(~abs(polarity), color = 'blue', fill = 'blue') %>%
  gf_histogram(~subjectivity, color= 'red', fill = 'red') %>%
  gf_labs(title = 'Histograms for Polarity (blue) and Subjectivity (red)',
          x='Polarity, Subjectivity', y='count')
```



This transformation also lines it up pretty well with subjectivity values. Let's see what happens with the model:

```
model_likes <- lm(log1p(num_likes) ~ abs(polarity) + subjectivity + num_followers +
                  abs(polarity):subjectivity + abs(polarity):num_followers +
                  subjectivity:num_followers, data = data)

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



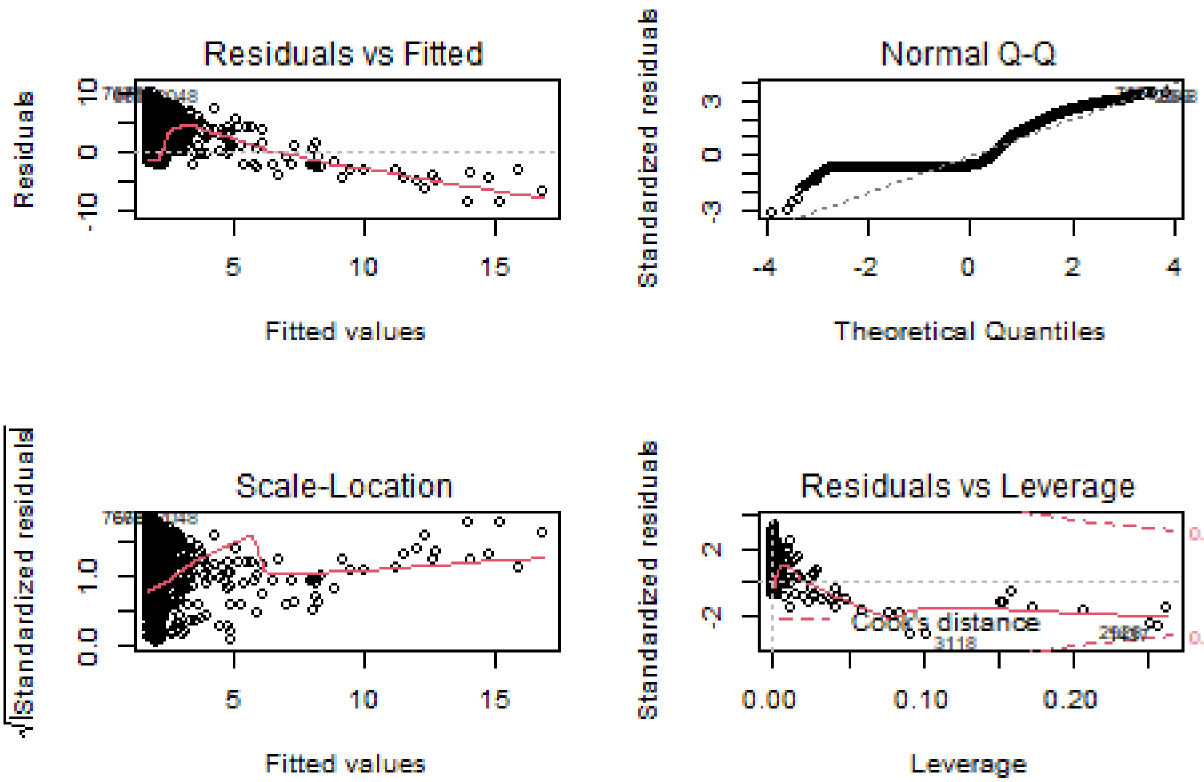

```r
summary(model_likes)
```

```
##
## Call:
## lm(formula = log1p(num_likes) ~ abs(polarity) + subjectivity +
##     num_followers + abs(polarity):subjectivity + abs(polarity):num_followers +
##     subjectivity:num_followers, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.483  -2.705  -1.753   2.298  11.192
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.582e+00  8.613e-02  29.980  < 2e-16 ***
## abs(polarity)               5.517e-01  1.705e-01   3.235  0.00122 **
## subjectivity                3.355e-01  1.784e-01   1.881  0.06001 .
## num_followers               1.662e-07  1.994e-08   8.335  < 2e-16 ***
## abs(polarity):subjectivity -4.149e-01  3.527e-01  -1.176  0.23949
## abs(polarity):num_followers 6.359e-08  5.237e-08   1.214  0.22470
## subjectivity:num_followers  4.262e-07  6.502e-08   6.554 5.91e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.438 on 9130 degrees of freedom
##   (12 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.03882,    Adjusted R-squared:  0.03819
## F-statistic: 61.45 on 6 and 9130 DF,  p-value: < 2.2e-16
```

```r
model_retweets <- lm(log1p(num_retweets) ~ abs(polarity) + subjectivity + num_followers +
                     abs(polarity):subjectivity + abs(polarity):num_followers +
                     subjectivity:num_followers, data = data)

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




```r
summary(model_retweets)
```

```
##
## Call:
## lm(formula = log1p(num_retweets) ~ abs(polarity) + subjectivity +
##     num_followers + abs(polarity):subjectivity + abs(polarity):num_followers +
##     subjectivity:num_followers, data = data)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -8.756 -2.050 -1.801  1.918 10.149
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.763e+00  7.415e-02  23.779  < 2e-16 ***
## abs(polarity)                4.886e-01  1.468e-01   3.328 0.000878 ***
## subjectivity                 2.864e-01  1.536e-01   1.865 0.062215 .
```

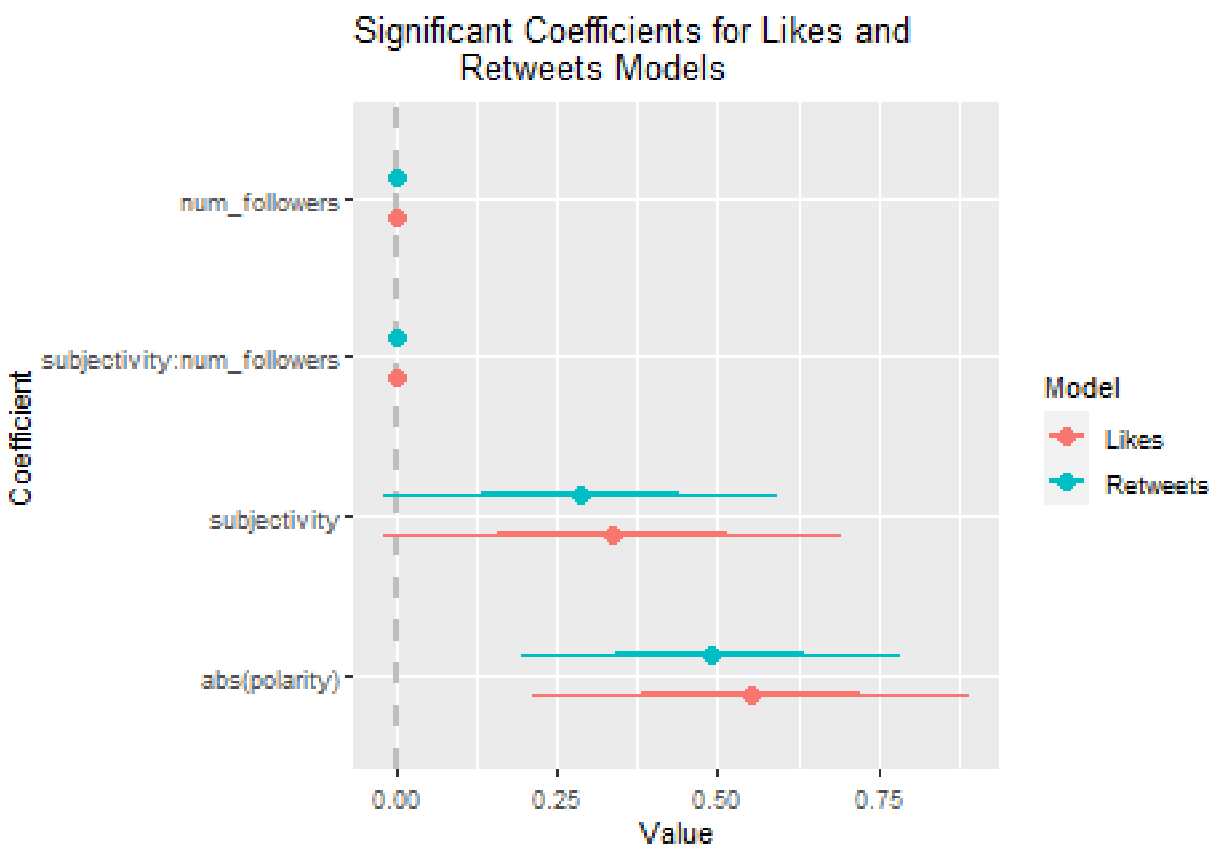
```
## num_followers                1.390e-07  1.717e-08   8.099 6.25e-16 ***
## abs(polarity):subjectivity  -2.689e-01  3.036e-01  -0.886 0.375891
## abs(polarity):num_followers  5.328e-08  4.508e-08   1.182 0.237364
## subjectivity:num_followers   3.170e-07  5.598e-08   5.663 1.53e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.959 on 9130 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.03447,    Adjusted R-squared:  0.03384
## F-statistic: 54.32 on 6 and 9130 DF,  p-value: < 2.2e-16
```

It turns out not much changed with our model, other than changes in the significance of abs(polarity) and subjectivity, and the loss of significance of the abs(polarity):subjectivity interaction. The coefficients for abs(polarity) and subjectivity are higher with 0-values removed (and the transformation on polarity, which we might call 'strength of polarity' or something along those lines).

Finally, we do still have evidence to reject the null hypothesis, and claim that there is a significant effect of polarity on the popularity of a tweet. But it should also be noted that we may not have landed on the best model for this dataset.

What remains is the best way to visualize this model. A good summation of the model might be done with a coefficient plot comprising the two models:

```r
multiplot(model_likes, model_retweets, title = 'Significant Coefficients for Likes and
          Retweets Models', dodgeHeight = 0.5, single=TRUE, numberAngle = 0, sort =
          'magnitude', decreasing = TRUE, names = c('Likes', 'Retweets'), coefficients =
          c('abs(polarity)', 'subjectivity', 'num_followers', 'subjectivity:num_followers'))
```



```r
#Export coefficient plot to pdf for the executive summary
pdf("Significant Coefficient Plot.pdf")
multiplot(model_likes, model_retweets, title = 'Significant Coefficients for Likes and
          Retweets Models', dodgeHeight = 0.5, single=TRUE, numberAngle = 0, sort =
          'magnitude', decreasing = TRUE, names = c('Likes', 'Retweets'), coefficients =
          c('abs(polarity)', 'subjectivity', 'num_followers', 'subjectivity:num_followers'))
dev.off()
```

```
## png
##   2
```