Analysis: Sentiment Analysis of President Trump's Tweets and the SPY

Moko Sharma

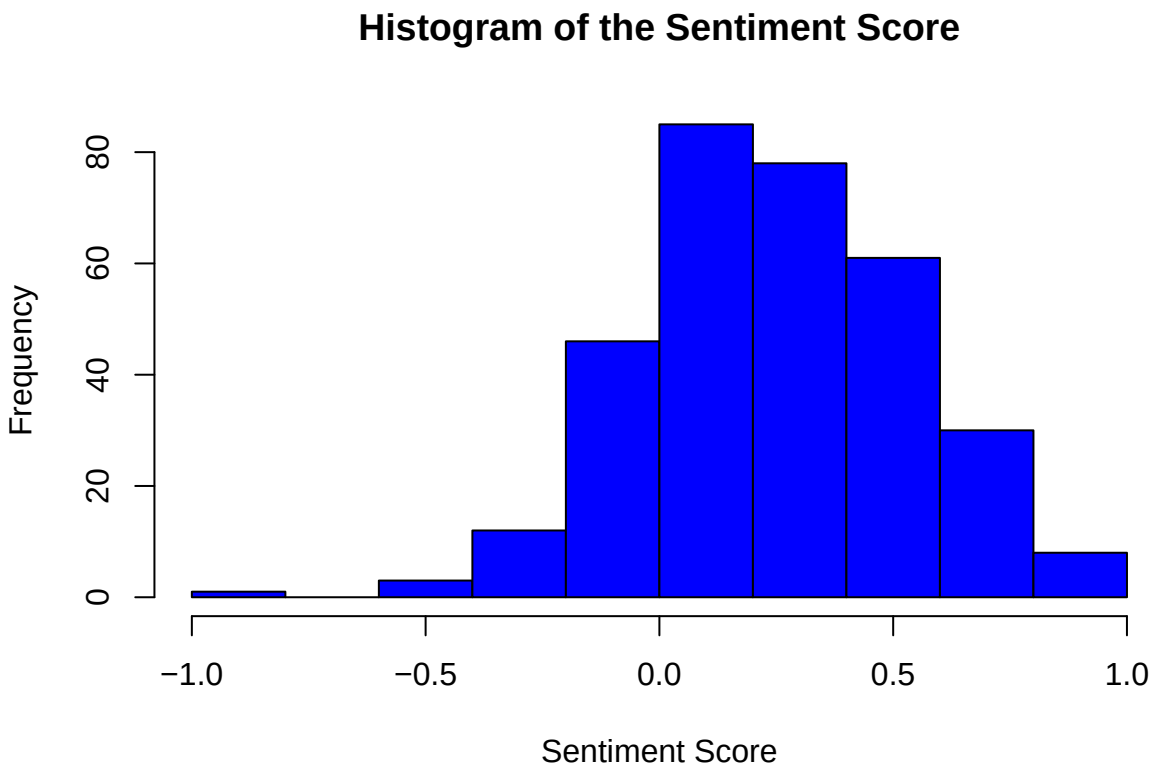
Reading in the combined Sentiment Analysis and SPY Dataframe via csv.

```
combined_df <- read.csv( '../../data/Combined_SentimentAndSPY.csv' )
summary( combined_df )
```

```
##          date          open            high            low
##  2016-12-30:  1   Min.   :219.6   Min.   :219.7   Min.   :217.6
##  2017-01-03:  1   1st Qu.:233.8   1st Qu.:234.5   1st Qu.:233.2
##  2017-01-04:  1   Median :243.9   Median :244.1   Median :243.0
##  2017-01-05:  1   Mean   :247.6   Mean   :248.4   Mean   :246.6
##  2017-01-06:  1   3rd Qu.:262.4   3rd Qu.:264.5   3rd Qu.:259.9
##  2017-01-09:  1   Max.   :284.8   Max.   :285.5   Max.   :283.4
##  (Other)   :318
##      close           volume           Name         range
##  Min.   :218.4   Min.   : 27856514   SPY:324   Min.   : 0.3171
##  1st Qu.:233.9   1st Qu.: 57361597             1st Qu.: 0.9484
##  Median :243.7   Median : 71742358             Median : 1.3129
##  Mean   :247.6   Mean   : 80817908             Mean   : 1.8621
##  3rd Qu.:261.8   3rd Qu.: 91498087             3rd Qu.: 2.0301
##  Max.   :285.4   Max.   :355026782             Max.   :12.4900
##
##  Sentiment.Score
##  Min.   :-0.94600
##  1st Qu.: 0.04018
##  Median : 0.23736
##  Mean   : 0.24154
##  3rd Qu.: 0.44511
##  Max.   : 0.91750
##
```

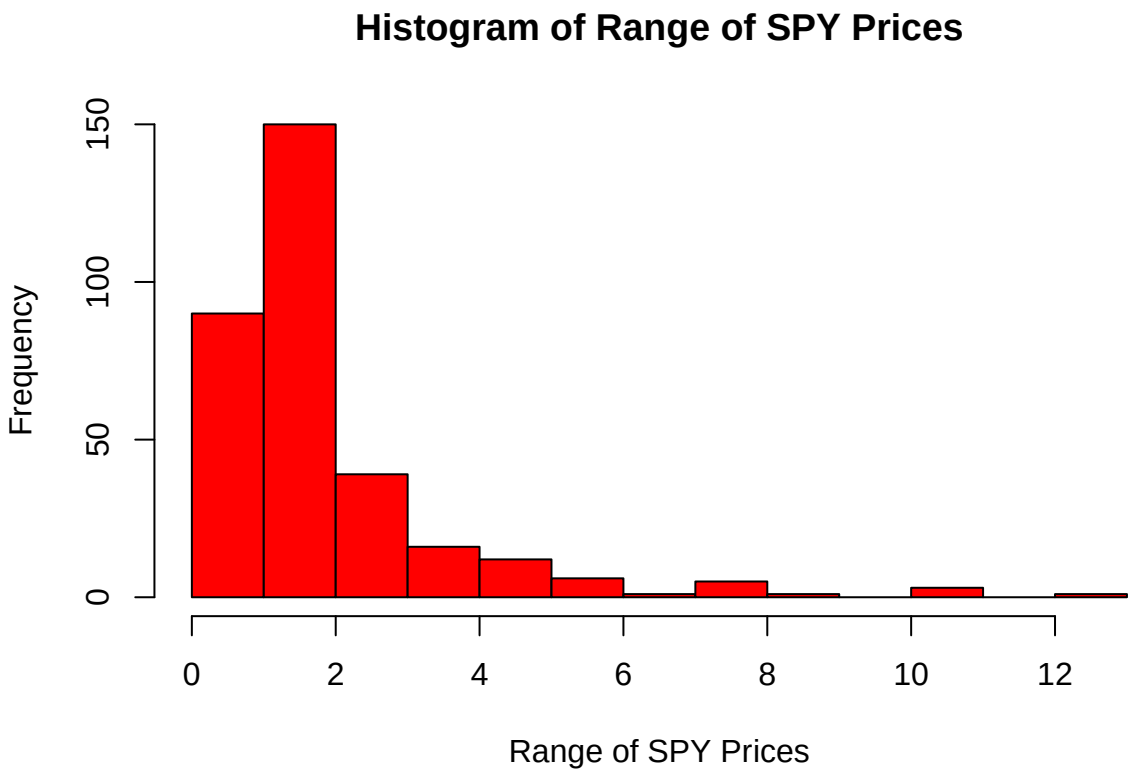
Histogram of the Sentiment Score

```
hist( x    = combined_df$Sentiment.Score,
      main = "Histogram of the Sentiment Score",
      xlab = "Sentiment Score",
      col  = 'blue' )
```



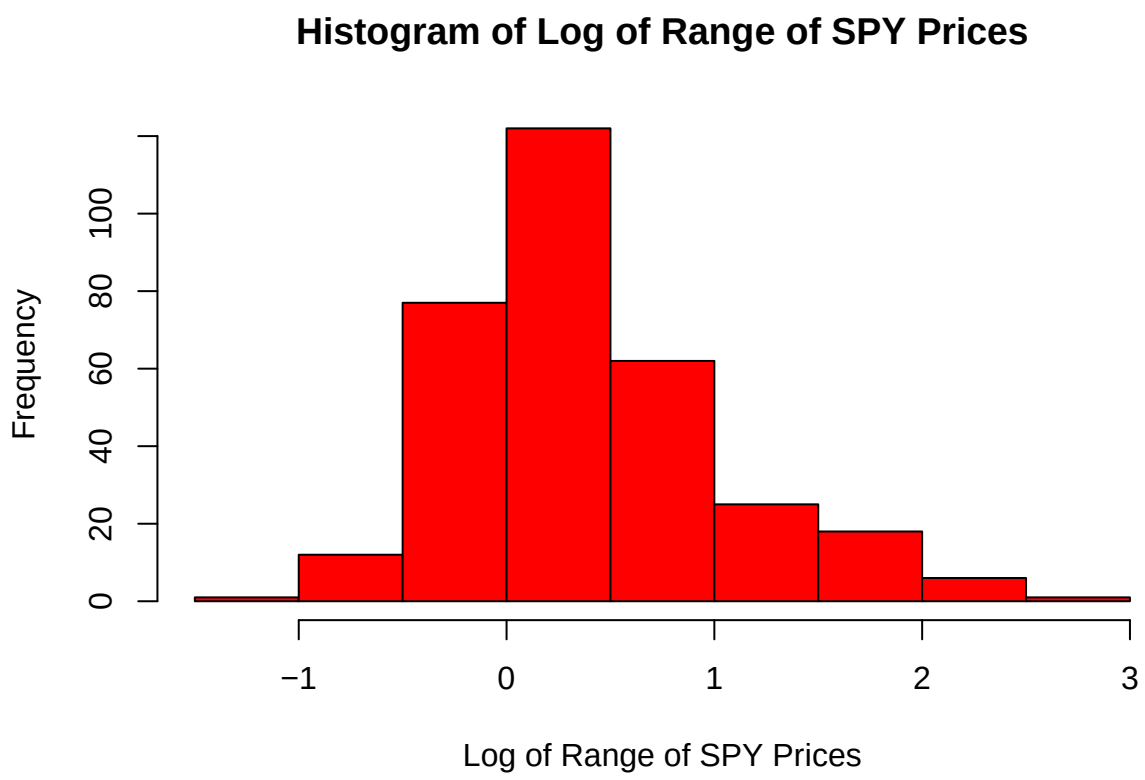
Histogram of the Range

```
hist( x    = combined_df$range,
      main = 'Histogram of Range of SPY Prices',
      xlab = 'Range of SPY Prices',
      col  = 'red' )
```



Log Transforming the Range Because of Skewness

```
log_range <- log( combined_df$range )
hist( x    = log_range,
      main = 'Histogram of Log of Range of SPY Prices',
      xlab = 'Log of Range of SPY Prices',
      col  = 'red' )
```

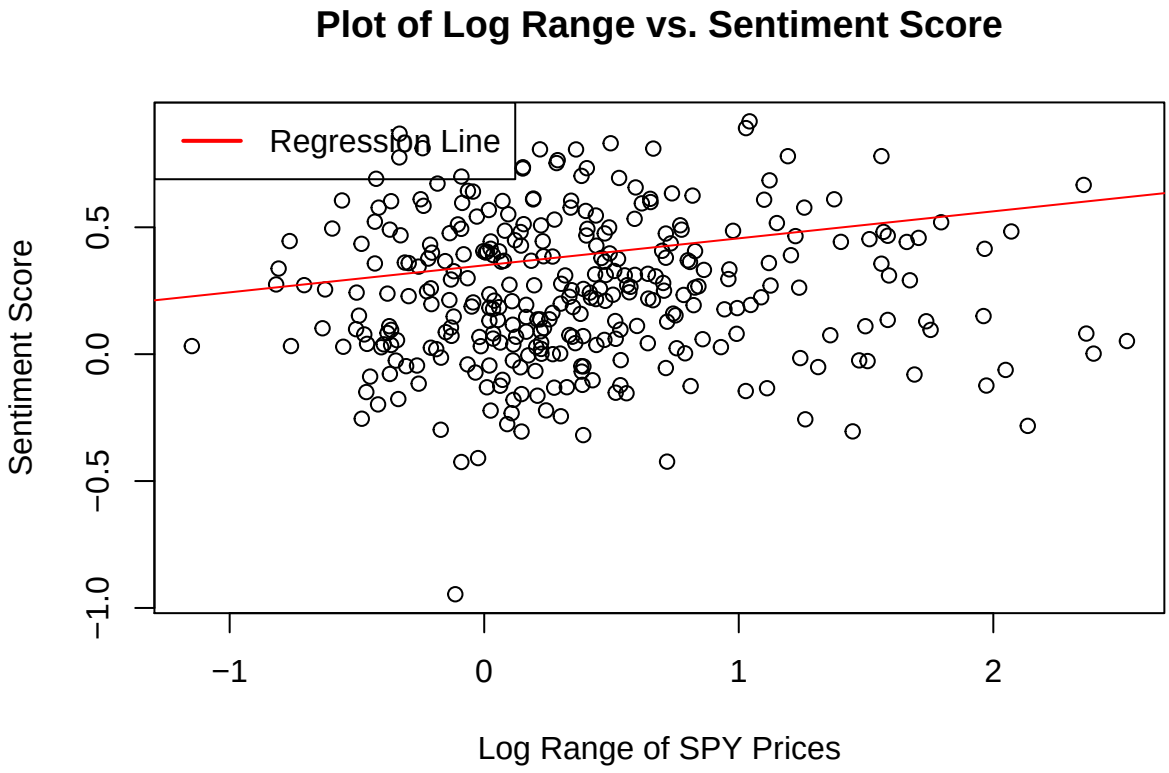



```
plot( x    = log_range,
      y    = combined_df$Sentiment.Score,
      main = "Plot of Log Range vs. Sentiment Score",
      xlab = 'Log Range of SPY Prices',
      ylab = "Sentiment Score" )

# Draw Regression Line
regression.model <- lm( log_range ~ combined_df$Sentiment.Score )

abline( a   = regression.model,
        col = 'red' )

legend( x = "topleft",
        c("Regression Line"),
        col = c("red"),
        lwd = c( 2 ))
```




```
cor.test( x = log_range, y = combined_df$Sentiment.Score )
```

```
##
##  Pearson's product-moment correlation
##
## data:  log_range and combined_df$Sentiment.Score
## t = 0.84227, df = 322, p-value = 0.4003
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.06239293  0.15505413
## sample estimates:
##        cor
## 0.04688604
```