data.R

USER

2023-11-10

```r
# Load the RedditExtractor package
library(RedditExtractoR)
```

```
## Warning: package 'RedditExtractoR' was built under R version
## 4.3.2
```

```r
library(SentimentAnalysis)
```

```
## Warning: package 'SentimentAnalysis' was built under R
## version 4.3.2
```

```
##
## Attaching package: 'SentimentAnalysis'
```

```
## The following object is masked from 'package:base':
##
##     write
```

```r
library(SnowballC)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```r
library(ggpmisc)
```

```
## Warning: package 'ggpmisc' was built under R version 4.3.2
```

```
## Loading required package: ggpp
```

```
## Warning: package 'ggpp' was built under R version 4.3.2
```

```
## Registered S3 methods overwritten by 'ggpp':
##   method                from
##   heightDetails.titleGrob ggplot2
##   widthDetails.titleGrob  ggplot2
```

```
##
## Attaching package: 'ggpp'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
## Registered S3 method overwritten by 'ggpmisc':
##   method                from
##   as.character.polynomial polynom
```

```
##
## Attaching package: 'ggpmisc'
```

```
## The following object is masked from 'package:ggpp':
##
##     quadrant_example.df
```

```r
wsb_threads <- find_thread_urls(subreddit = "wallstreetbets", sort_by = "top", period = "year")
```

```
## parsing URLs on page 1...
## parsing URLs on page 2...
## parsing URLs on page 3...
## parsing URLs on page 4...
## parsing URLs on page 5...
## parsing URLs on page 6...
## parsing URLs on page 7...
## parsing URLs on page 8...
## parsing URLs on page 9...
## parsing URLs on page 10...
```

```r
# view the results
head(wsb_threads)
```

```
##                                                                date_utc
## It\031s not gambling its called investing.                  2023-08-02
## FED won\031t STFU part 2! Mr. Powell is watching you!        2023-01-09
## Soft landing?                                                2022-12-18
## All Hail ChatGPT                                             2023-02-15
## Skynet                                                       2023-05-18
## Puts on the US government ?                                  2023-09-10
##                                                                timestamp
## It\031s not gambling its called investing.                  1690988078
## FED won\031t STFU part 2! Mr. Powell is watching you!        1673222667
## Soft landing?                                                1671368306
## All Hail ChatGPT                                             1676504502
## Skynet                                                       1684442012
## Puts on the US government ?                                  1694354330
##
## It\031s not gambling its called investing.                             It\031s not gambling its called inve
## FED won\031t STFU part 2! Mr. Powell is watching you! FED won\031t STFU part 2! Mr. Powell is watchin
## Soft landing?                                                                              Soft la
## All Hail ChatGPT                                                                           All Hail (
## Skynet
## Puts on the US government ?                                                         Puts on the US goverr
##                                                              text
## It\031s not gambling its called investing.
## FED won\031t STFU part 2! Mr. Powell is watching you!
## Soft landing?
## All Hail ChatGPT
## Skynet
## Puts on the US government ?
##                                                             subreddit
## It\031s not gambling its called investing.        wallstreetbets
## FED won\031t STFU part 2! Mr. Powell is watching you! wallstreetbets
## Soft landing?                                     wallstreetbets
## All Hail ChatGPT                                  wallstreetbets
## Skynet                                            wallstreetbets
## Puts on the US government ?                       wallstreetbets
##                                                              comments
## It\031s not gambling its called investing.              60
## FED won\031t STFU part 2! Mr. Powell is watching you!    642
## Soft landing?                                            524
## All Hail ChatGPT                                         132
## Skynet                                                   108
## Puts on the US government ?                              522
##
## It\031s not gambling its called investing.                    https://www.reddit.com/r/wallstreetbet
## FED won\031t STFU part 2! Mr. Powell is watching you! https://www.reddit.com/r/wallstreetbets/commen
## Soft landing?                                                            https://www
## All Hail ChatGPT                                                      https://www.reddi
## Skynet                                                                      https:,
## Puts on the US government ?                                https://www.reddit.com/r/v
```

```
wsb_comments <- get_thread_content(wsb_threads$url[1:15])

# changing text format
wsb_comments$comments$comment <- enc2utf8(wsb_comments$comments$comment)
```

```
# run sentiment analysis
wsb_sentiment <- analyzeSentiment(wsb_comments$comments$comment)

# combine data
wsb_dataset <- bind_cols(wsb_comments$comments, wsb_sentiment)

# plot histogram of sentiment
ggplot(wsb_dataset, aes(x = SentimentGI)) + geom_histogram() + labs(title = "sentiment for wallstreetbe
```
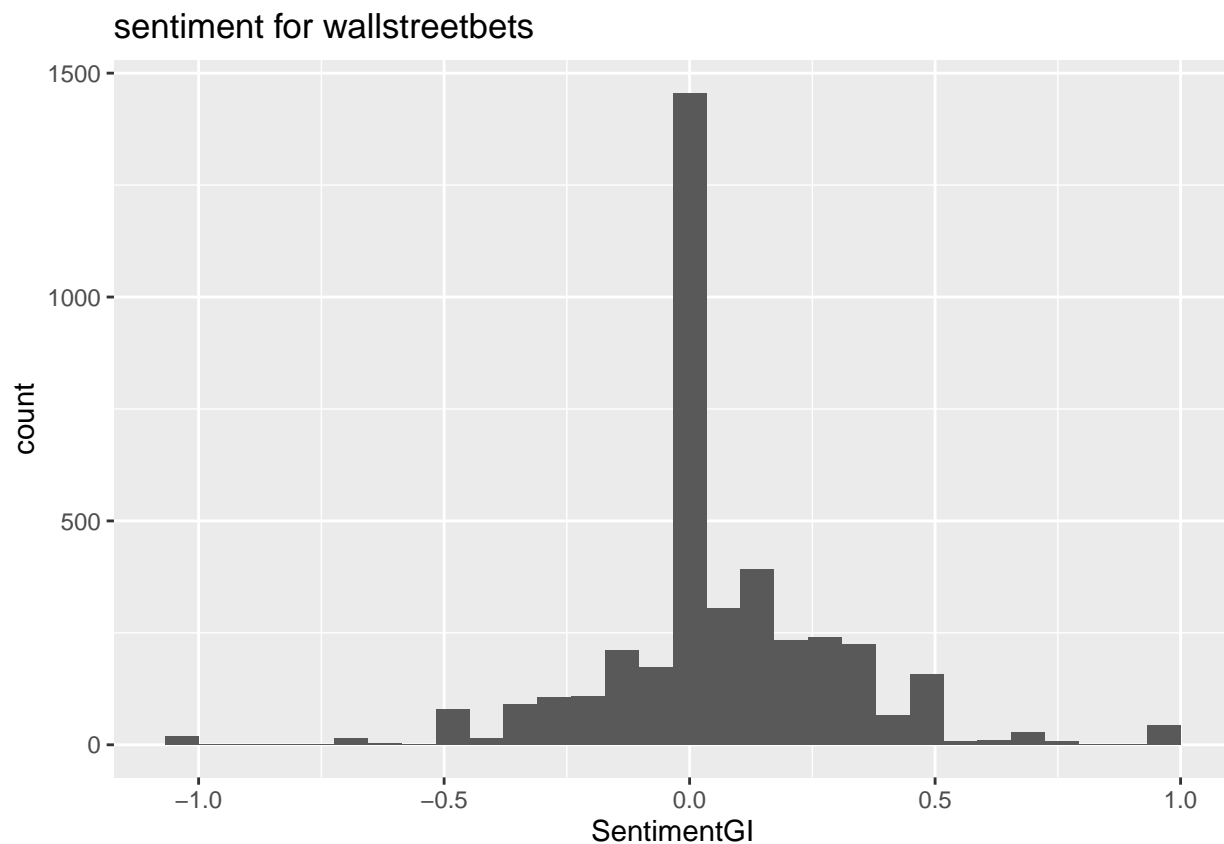
```
## 'stat_bin()' using 'bins = 30'. Pick better value with
## 'binwidth'.
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_bin()').
```
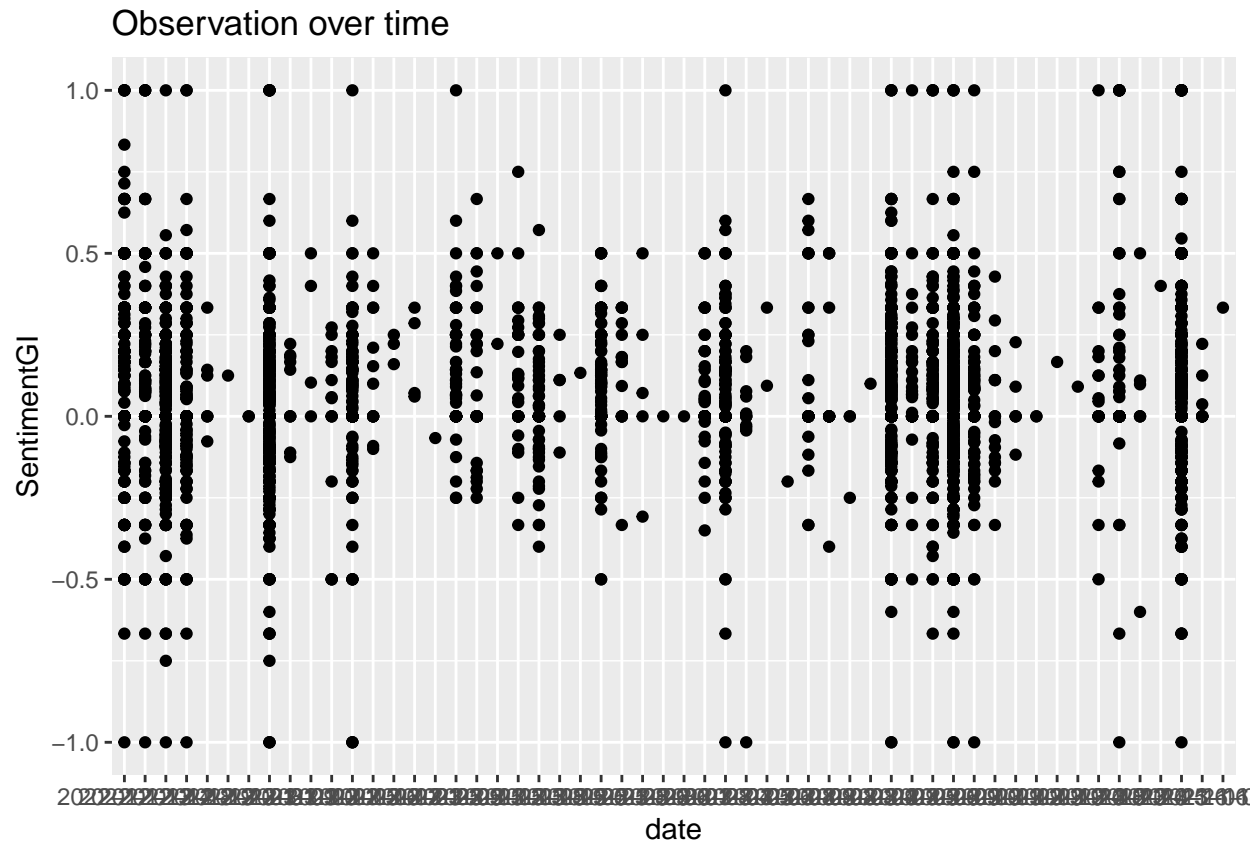


```
# plot sentiment vs day of posting
# scatterplot
ggplot(wsb_dataset, aes(x = date, y = SentimentGI)) + geom_point() + labs(title = "Observation over time
```

```
## Warning: Removed 39 rows containing missing values
## ('geom_point()').
```
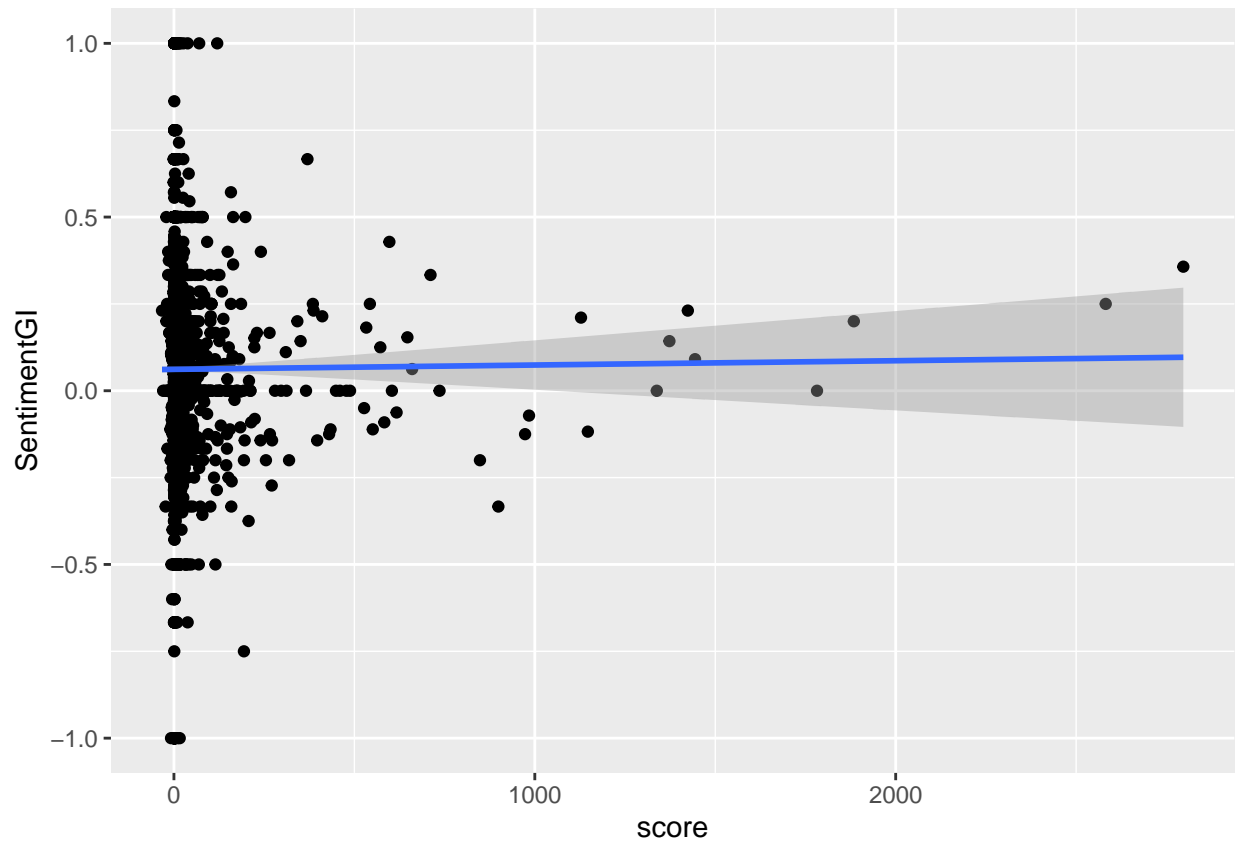
Observation over time


```r
# plot sentiment by the comment score
ggplot(wsb_dataset, aes(x = score, y = SentimentGI)) + geom_point() + geom_smooth(method = lm )
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_smooth()').
```

```
## Warning: Removed 39 rows containing missing values
## ('geom_point()').
```

```r
# hexplot of density with linear model
ggplot(wsb_dataset, aes(x = score, y = SentimentGI)) + geom_hex(bins = 20, color = "white") +
  scale_fill_gradient(low = 'blue', high = 'orange') +
  labs(title = "sentiment by comment score") +
  geom_smooth(method = lm, formula = y ~ poly(x,2)) +
  theme_minimal() +
  stat_poly_eq(
    aes(label = paste(..eq.label.., ..adj.rr.label.., sep = "~~~~" )),
    formula = y ~ poly(x,2), parse = TRUE
  )
```
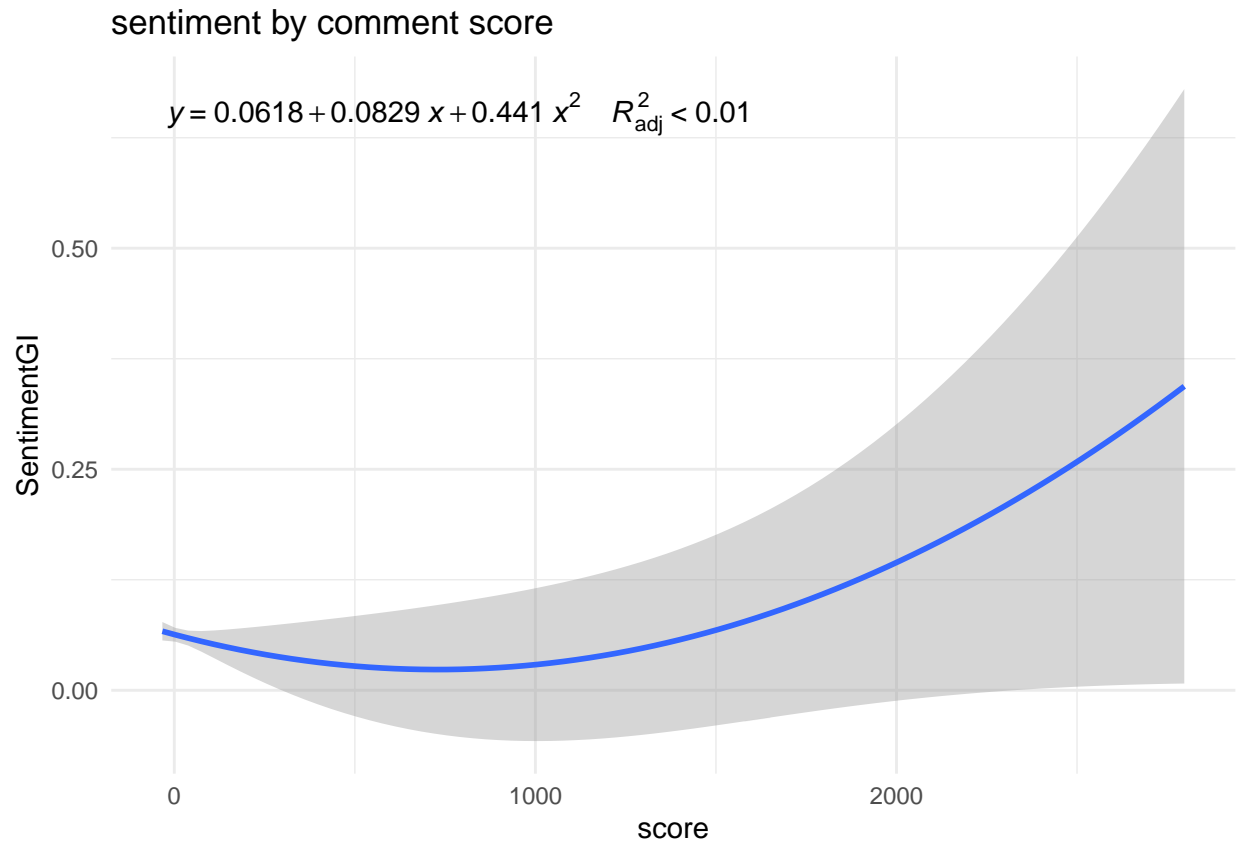
```
## Warning: The dot-dot notation (`..eq.label..`) was deprecated in
## ggplot2 3.4.0.
## i Please use `after_stat(eq.label)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this
## warning was generated.
```

```
## Warning: Removed 39 rows containing non-finite values
## (`stat_binhex()`).
```

```
## Warning: Computation failed in `stat_binhex()`
## Caused by error in `compute_group()`:
## ! The package "hexbin" is required for `stat_binhex()`
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_smooth()').
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_poly_eq()').
```

sentiment by comment score

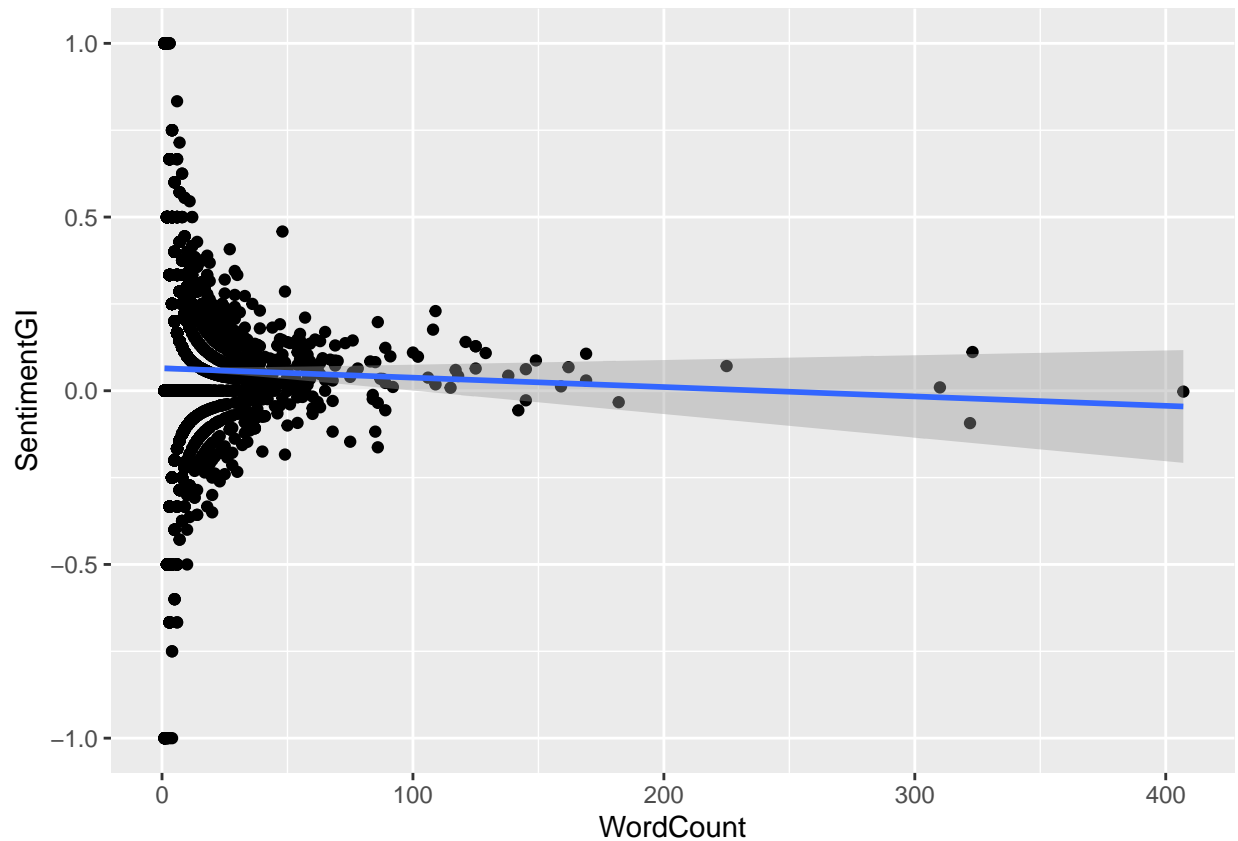$$y = 0.0618 + 0.0829\,x + 0.441\,x^2 \quad R^2_{adj} < 0.01$$


```
# plot sentiment by word count
# scatterplot
ggplot(wsb_dataset, aes(x = WordCount, y = SentimentGI)) + geom_point() + geom_smooth(method = lm )
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_smooth()').
```

```
## Warning: Removed 39 rows containing missing values
## ('geom_point()').
```

```
# hexplot of density with linear model
ggplot(wsb_dataset, aes(x = WordCount, y = SentimentGI)) + geom_hex(bins = 20, color = "white") +
  scale_fill_gradient(low = 'blue', high = 'orange') +
  labs(title = "sentiment by comment score") +
  geom_smooth(method = lm, formula = y ~ poly(x,2)) +
  theme_minimal() +
  stat_poly_eq(
    aes(label = paste(..eq.label.., ..adj.rr.label.., sep = "~~~~" )),
    formula = y ~ poly(x,2), parse = TRUE
  )
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_binhex()').
```

```
## Warning: Computation failed in 'stat_binhex()'
## Caused by error in 'compute_group()':
## ! The package "hexbin" is required for 'stat_binhex()'
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_smooth()').
```

```
## Warning: Removed 39 rows containing non-finite values
## ('stat_poly_eq()').
```

sentiment by comment score

$y = 0.0618 - 0.318\,x + 0.179\,x^2 \quad R^2_{adj} < 0.01$