R Notebook

```
library(twitteR)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ 
## x dplyr::filter()   masks stats::filter()
## x dplyr::id()       masks twitteR::id()
## x dplyr::lag()      masks stats::lag()
## x dplyr::location() masks twitteR::location()
```

Load necessary libraries.

```
# Obtain access to the Twitter API using appropriate authentication
api <- '7DPTSeLp9wqmBaoayNYpPJvLp'
apiSecret <- 'ftzB0U2ulWaYs32tNHn9lERkhbVUtbuAiP6o8Lec5XV7lNU7GM'
access <- '1012248144-7qLC5cfX6BVKS1a4lpnIHRndLSPOxiUA62RnEOw'
accessSecret <- 'Qbh5TneWdrQhEkmlDvzOGFKmHzkZVm1KcDzb0kOK13eBP'
setup_twitter_oauth(api, apiSecret, access, accessSecret)
```

```
## [1] "Using direct authentication"
```

```
# Query Twitter API using a function from the twitteR library to obtain a table
#with the 1000 most recent tweets with the hashtag 'rstats' and relevant
#information about the tweets
tweets <- searchTwitter("#rstats", 1000)
# Remove retweets
tweets_clean <- strip_retweets(tweets)
# Build a tibble with the name of the tweeter, the content of the tweet, and
#the number of both favorites and retweets
tweets_tbl <- twListToDF(tweets_clean) %>%
    select(screenName, text, favoriteCount, retweetCount)
```

```
rt.cor <- cor.test(nchar(tweets_tbl$text), tweets_tbl$retweetCount)
rt.cor
```

```
##
##  Pearson's product-moment correlation
##
## data:  nchar(tweets_tbl$text) and tweets_tbl$retweetCount
## t = 2.4686, df = 118, p-value = 0.015
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```

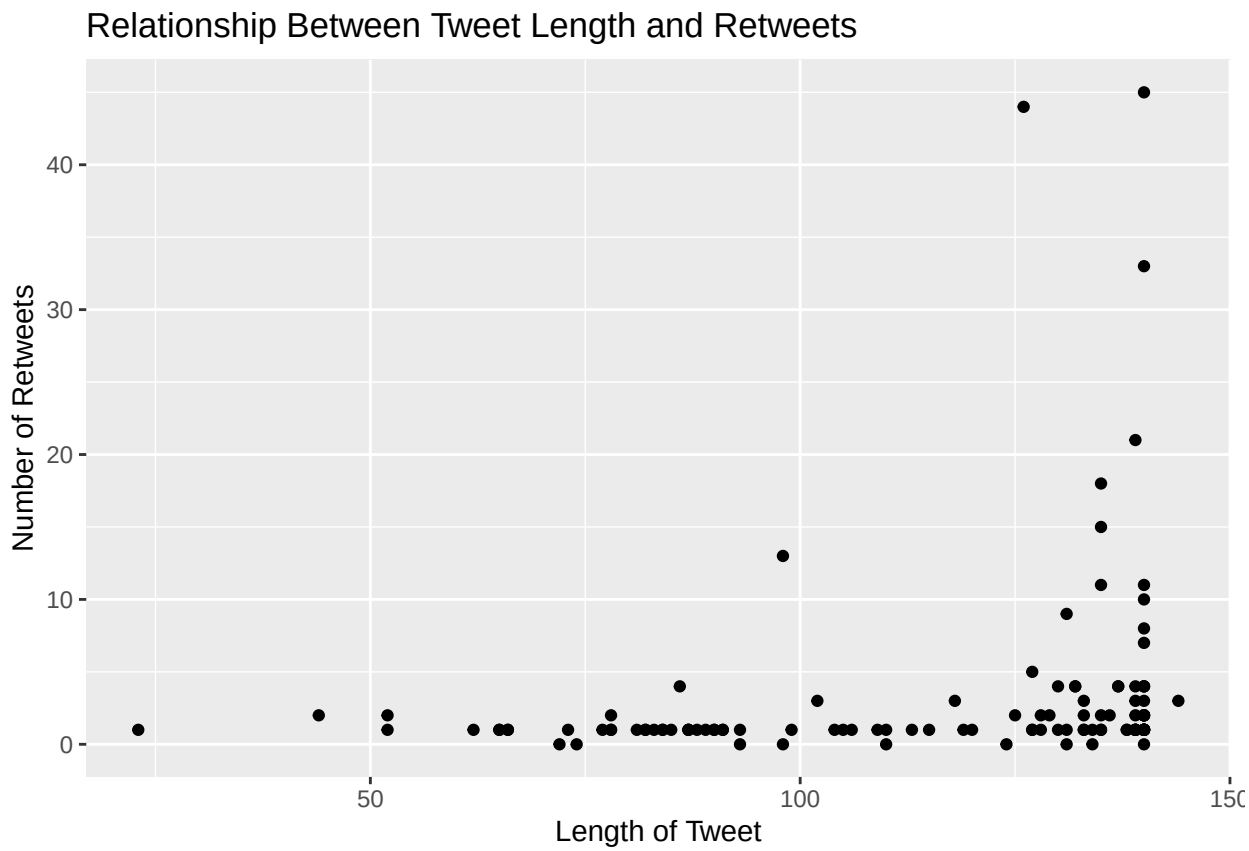
```
##  0.04410937 0.38552704
## sample estimates:
##       cor
## 0.2215987
```

```
fav.cor <- cor.test(nchar(tweets_tbl$text), tweets_tbl$favoriteCount)
fav.cor
```

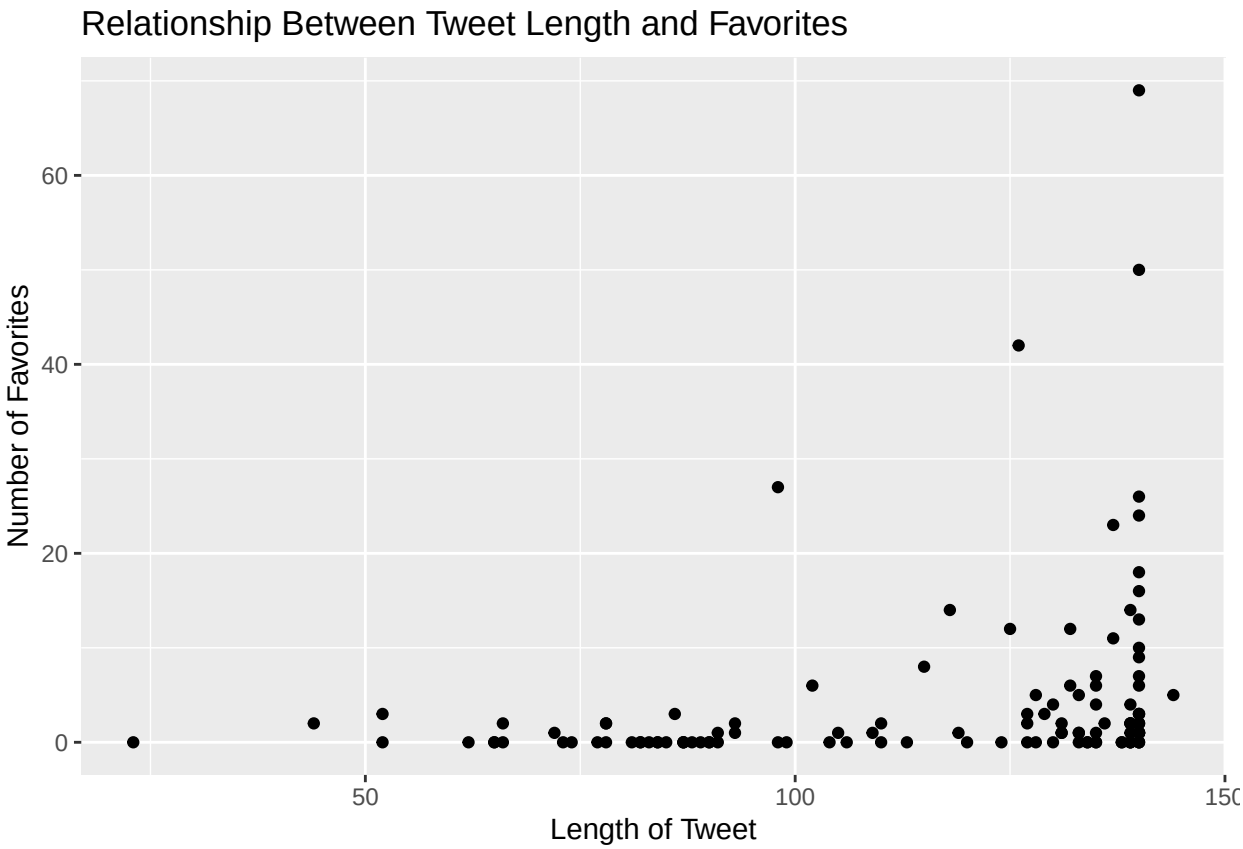
```
##
##  Pearson's product-moment correlation
##
## data:  nchar(tweets_tbl$text) and tweets_tbl$favoriteCount
## t = 2.9019, df = 118, p-value = 0.004427
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08267554 0.41799672
## sample estimates:
##       cor
## 0.2580922
```

Examine the relationship between length of tweet and both number of retweets and number of favorites. The correlation for length of tweet with number of retweets is 0.2215987, and the correlation for length of tweet with number of favorites is 0.2580922.

```
ggplot(tweets_tbl, aes(x = nchar(text), y = retweetCount)) +
           geom_point() +
    ggtitle("Relationship Between Tweet Length and Retweets") +
    xlab("Length of Tweet") +
    ylab("Number of Retweets")
```




```
ggplot(tweets_tbl, aes(x = nchar(text), y = favoriteCount)) +
           geom_point() +
    ggtitle("Relationship Between Tweet Length and Favorites") +
    xlab("Length of Tweet") +
    ylab("Number of Favorites")
```



These plots show the relationship between length of tweet and both number of retweets and number of favorites in the form of a scatterplot.