Week11-twitter

Jiayin Qu

3/28/2020

Library Imports

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(twitteR)
library(stringr)
api <- 'e0y2pHHrYtT320wP8WSsPa0i7'
apiSecret <- 'Ewv63M79SR6ocpWGbW5lpOgiX3z8aMJddwqR5Fx70tjxo4lb4r'
access <- '1195948151179501569-THuTApWDbYUrbooDA0iSldWpgY63WT'
accessSecret <- 'OY5ESKILi1GgVivsbM1JeM3UmWLISO7ToADW8V9ZlZL4v'

setup_twitter_oauth(api, apiSecret, access, accessSecret)
```

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

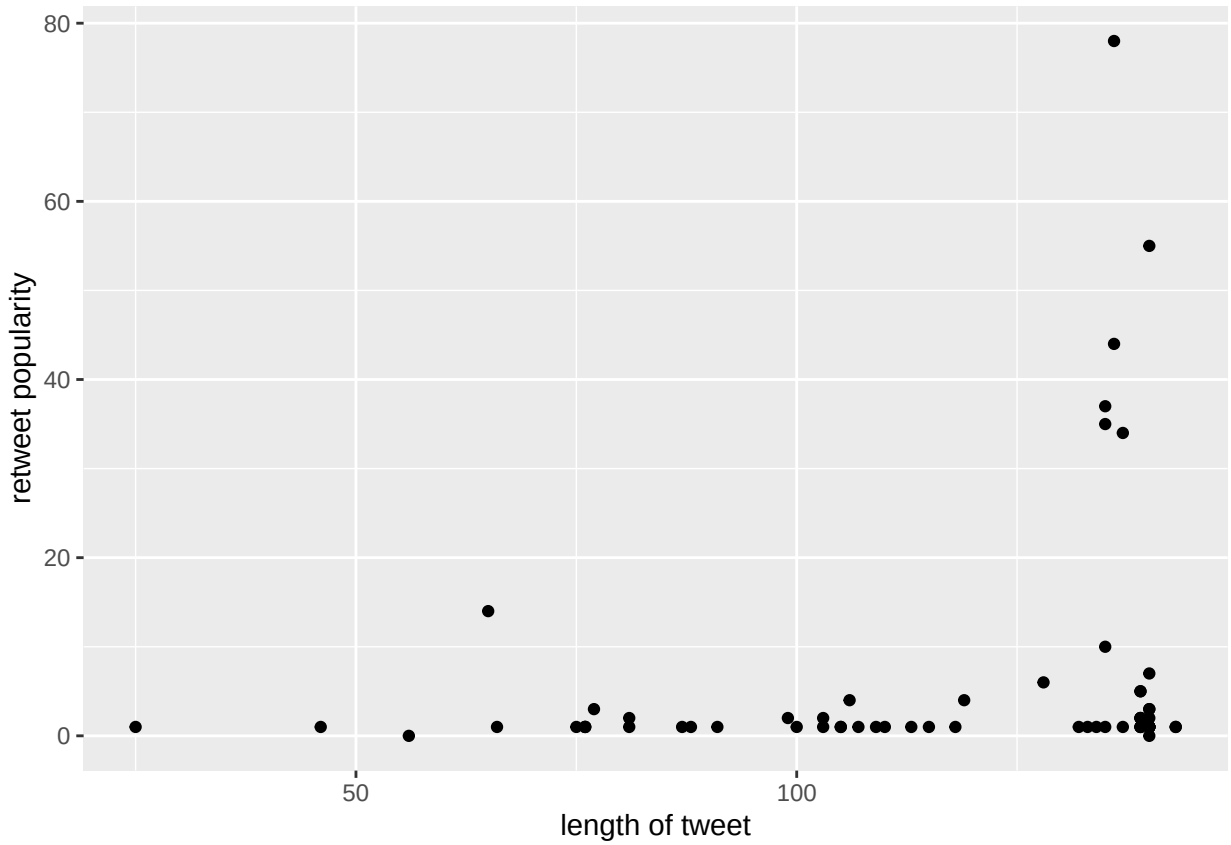
Data Import

```
tweets <- searchTwitter("#rstats", 1000)
tweets_clean <- strip_retweets(tweets)
tweets_tbl <- twListToDF(tweets_clean) %>%
  transmute(text,
            screenName,
            favoriteCount,
            retweetCount,
            len_tweets = str_count(text))
```

Visualization

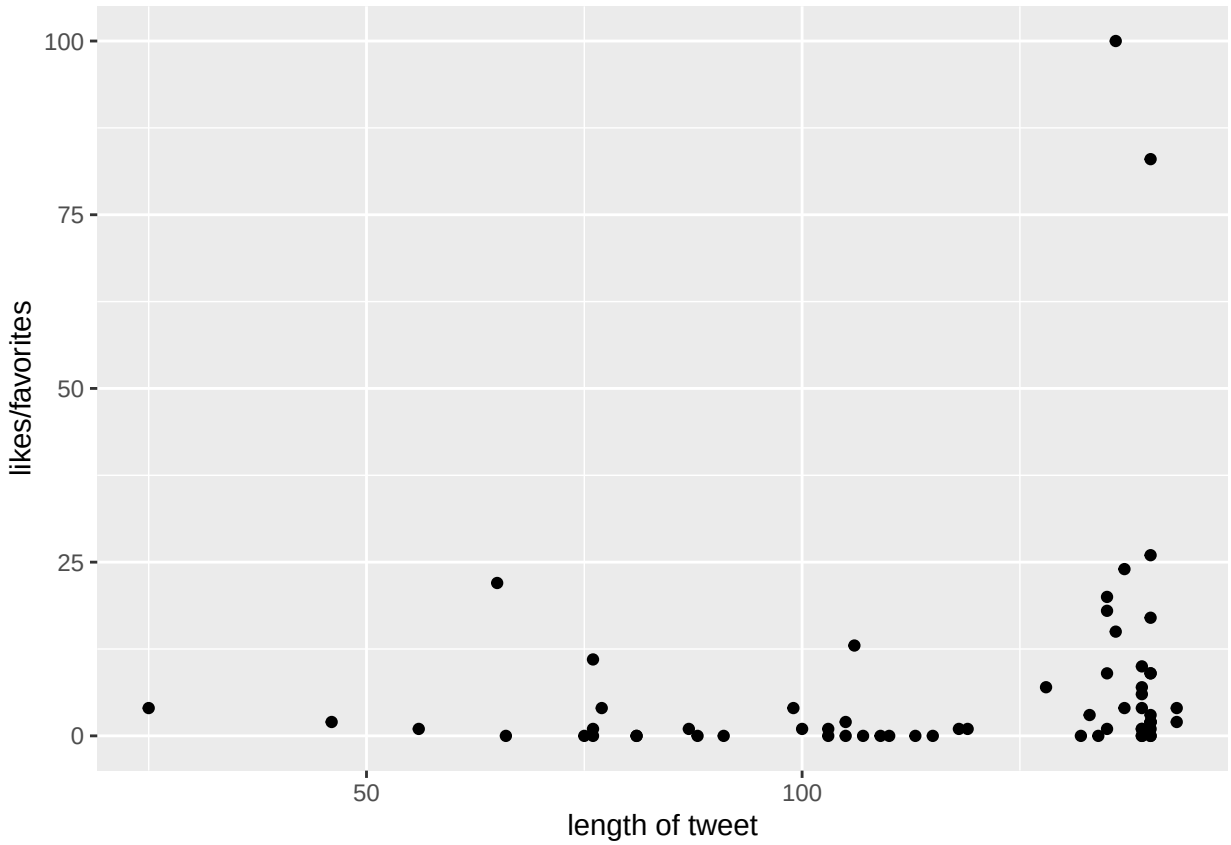
- visualize the relationship between length of tweet and retweet popularity

```
ggplot(tweets_tbl, aes(len_tweets, retweetCount)) +
  geom_point() +
  labs(x = "length of tweet", y = "retweet popularity")
```



- visualize the relationship between length of tweet and likes/favorites

```
ggplot(tweets_tbl, aes(len_tweets, favoriteCount)) +
  geom_point() +
  labs(x = "length of tweet", y = "likes/favorites")
```



Analysis

```
#calculate the relationship between length of tweet and other variables
corRetweet <- cor.test(tweets_tbl$len_tweets, tweets_tbl$retweetCount, method = "pearson")
corPop <- cor.test(tweets_tbl$len_tweets, tweets_tbl$favoriteCount, method = "pearson")
```

The correlation between length of tweet and retweet popularity was 0.2261786 (p>=.05), which is not statistically significant.

The correlation between length of tweet and likes/favorites was 0.2028331 (p>=.05), which is not statistically significant.