R Notebook

Dana Tomeh

# Rstudio API code

# Libraries

library(twitteR)  
library(ggplot2)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ tibble 2.1.3 ✓ dplyr 0.8.4  
## ✓ tidyr 1.0.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0  
## ✓ purrr 0.3.3

## ── Conflicts ────────────────────────────────────────────────────────────────────────────────────── tidyverse\_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()

# Data Import and Cleaning

Pulled the api, secret api, access, and secret access codes from twitter in orger to set up the twitter authorization. I then Pulled the first 1000 tweets that mention the #rstats. Removed all of those that were retweets and not orginal tweets. Then created a dataframe out of the twitter list that includes the tweet itself, the user who wrote it’s screen name, the number of times it was favorited, and the number of times it was retweeted.

api <- "zkHmLKzPFWPc2HfCDbjUU8HP2"  
apiSecret <- "JZGwN3tqcQracHJwb2io97wRsLKFO6h1EY8MyEGz3RjGBMwXEx"  
access <- "48436651-M7DDZdx7f5BHIIHB4zlH34tMSKxDNLU0XOdD5FKN7"  
accessSecret <- "YR2BAv1bXDY9ejt0eaoBtK1y4aeup0Z67XmqtTezIbmF3"  
  
setup\_twitter\_oauth(api, apiSecret, access, accessSecret)

## [1] "Using direct authentication"

tweets <-searchTwitter("#rstats", 1000)  
tweets\_clean <- strip\_retweets(tweets)  
tweets\_tbl <- twListToDF(tweets\_clean) %>%   
 select(text,screenName,favoriteCount,retweetCount)

#Analysis

In this section a correlation test was run between the length of the number of characters in a tweet and the number of retweets.

retweet.cor <- cor.test(nchar(tweets\_tbl$text), tweets\_tbl$retweetCount)

The correlation between the length of the tweet and the number of retweets is 0.1585448 with a p value of 0.1628471

In this section a correlation test was run between the length of the number of characters in a tweet and the number of favorites.

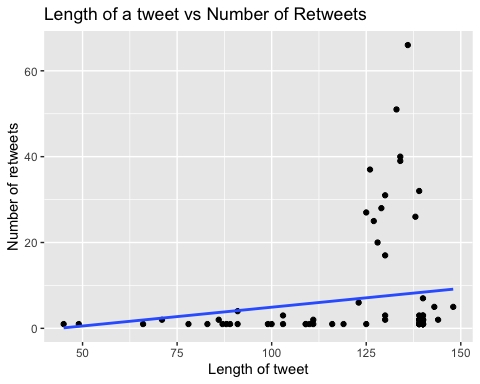
fav.cor <-cor.test(nchar(tweets\_tbl$text), tweets\_tbl$favoriteCount)

The correlation between the length of the tweet and the number of retweets is 0.1255186 with a p value of 0.2703672

# Visualization

Created a scatterplot of the relationship between the length of a tweet and the number of retweets using ggplot from the package ggplot2. Superimposed a regression line on top of the points using geom\_smooth with the method set to “lm”.

ggplot(tweets\_tbl, aes(nchar(tweets\_tbl$text), tweets\_tbl$retweetCount))+  
 geom\_point() +  
 geom\_smooth(method="lm", se=FALSE) +   
 labs( title = "Length of a tweet vs Number of Retweets",   
 x="Length of tweet",  
 y="Number of retweets")



Created a scatterplot of the relationship between the length of a tweet and the number of favorites using ggplot from the package ggplot2. Superimposed a regression line on top of the points using geom\_smooth with the method set to “lm”.

ggplot(tweets\_tbl, aes(nchar(tweets\_tbl$text), tweets\_tbl$favoriteCount))+  
 geom\_point() +   
 geom\_smooth(method="lm", se=FALSE) +   
 labs( title = "Length of a tweet vs Number of Favorites",   
 x="Length of tweet",  
 y="Number of Favorites")

