# Week 10 - Twitter

## R Studio API Code

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

## Libraries

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

## Data Import and Cleaning

```
api <- "USXSQ96P2LuIVPA5ilppBDofX"
apiSecret <- "0ePRPNq4HXUxNEEeRVznqUasgeFwKAcMJ1Wqu7Nh8vrx8BjGxK"
access <- "1243552285424340994-jeL8kxj8zRP95M9khVPCptH6NaWp7m"
accessSecret <- "mMoP3SJ2dGY67j1yT9CLLeXTQ60X48ThtHjq1nJOPds6O"
setup_twitter_oauth(api, apiSecret,access,accessSecret)
```

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

```
tweets <- searchTwitter("#rstats", 1000)
tweets_tbl <- twListToDF(tweets) %>%
  dplyr::select(screenName, text, favoriteCount, retweetCount)
```

## Analysis

```
tweets_tbl <- tweets_tbl %>%
  rowwise() %>%
  mutate(tweetlength = str_length(text))
```

## Correlation between length of tweet and retweet popularity

```
cor.test(tweets_tbl$tweetlength, tweets_tbl$retweetCount)
```

```
##
##  Pearson's product-moment correlation
##
## data:  tweets_tbl$tweetlength and tweets_tbl$retweetCount
```

```
## t = 5.6545, df = 998, p-value = 2.04e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1154577 0.2356093
## sample estimates:
##       cor
## 0.1761897
```

## Correlation between length of tweet and likes/favorites

```r
cor.test(tweets_tbl$tweetlength, tweets_tbl$favoriteCount)
```

```
##
##  Pearson's product-moment correlation
##
## data:  tweets_tbl$tweetlength and tweets_tbl$favoriteCount
## t = -1.8393, df = 998, p-value = 0.06617
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.119684907  0.003884001
## sample estimates:
##         cor
## -0.05812308
```

# Visualization

```r
tweets_tbl %>%
  ggplot(aes(x = tweetlength, y = retweetCount)) +
  geom_point() +
  geom_smooth(method = "lm", se = F) +
  labs(title = "Tweet length and retweet popularity") +
  theme_classic()
```

## Tweet length and retweet popularity
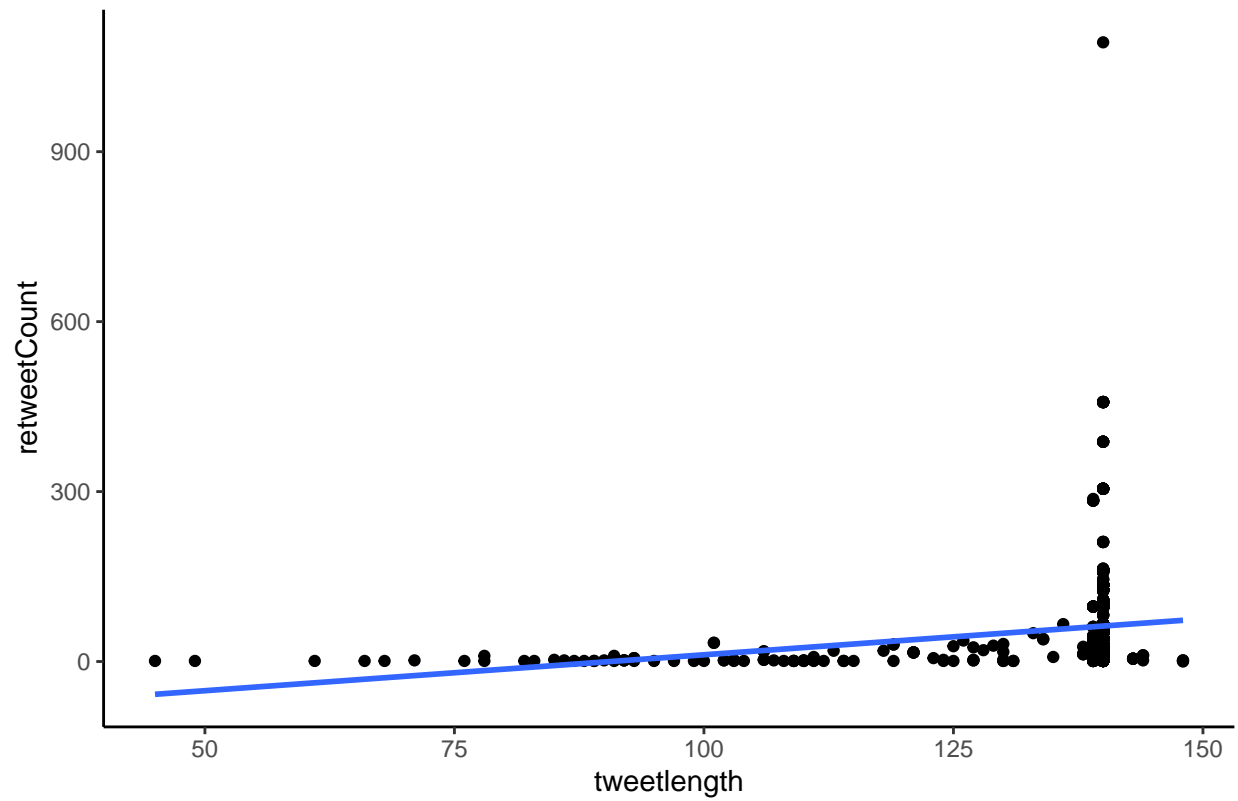


```
tweets_tbl %>%
  ggplot(aes(x = tweetlength, y = favoriteCount)) +
  geom_point() +
  geom_smooth(method = "lm", se = F) +
  labs(title = "Tweet length and likes/favorites") +
  theme_classic()
```

Tweet length and likes/favorites