

STAT 240

Dr. Lloyd T. Elliott
lloyd.elliott@sfu.ca

Debugging Code

- Types of errors:
 - Syntax errors
 - Runtime errors
 - Logic errors

Debugging in R

- Print statements
- Examine stack trace:
 - `options(error = browser)`
- Breakpoints & browser

Review

Finding Unique Combinations vs Counting Unique Combinations

- `unique(pokeLegGen)`
- `unique(pokeLegGenFire)`
- vs
- `table(pokeLegGen)`
- `table(pokeLegGenFire)`

Obtain an API key

- You can get your own API key: <https://developer.twitter.com/en/docs>
- This is used to download Twitter data through the Twitter API
- Getting an API key is not required for this course

Accessing Twitter

- `library(twitteR)`
- `consumer_key = "YOUR API KEY"`
- `consumer_secret = "YOUR API SECRET"`
- `access_token = "YOUR ACCESS TOKEN"`
- `access_secret = "ACCESS TOKEN SECRET"`
- `setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)`

access.r x

  Source on Save  Run  Source

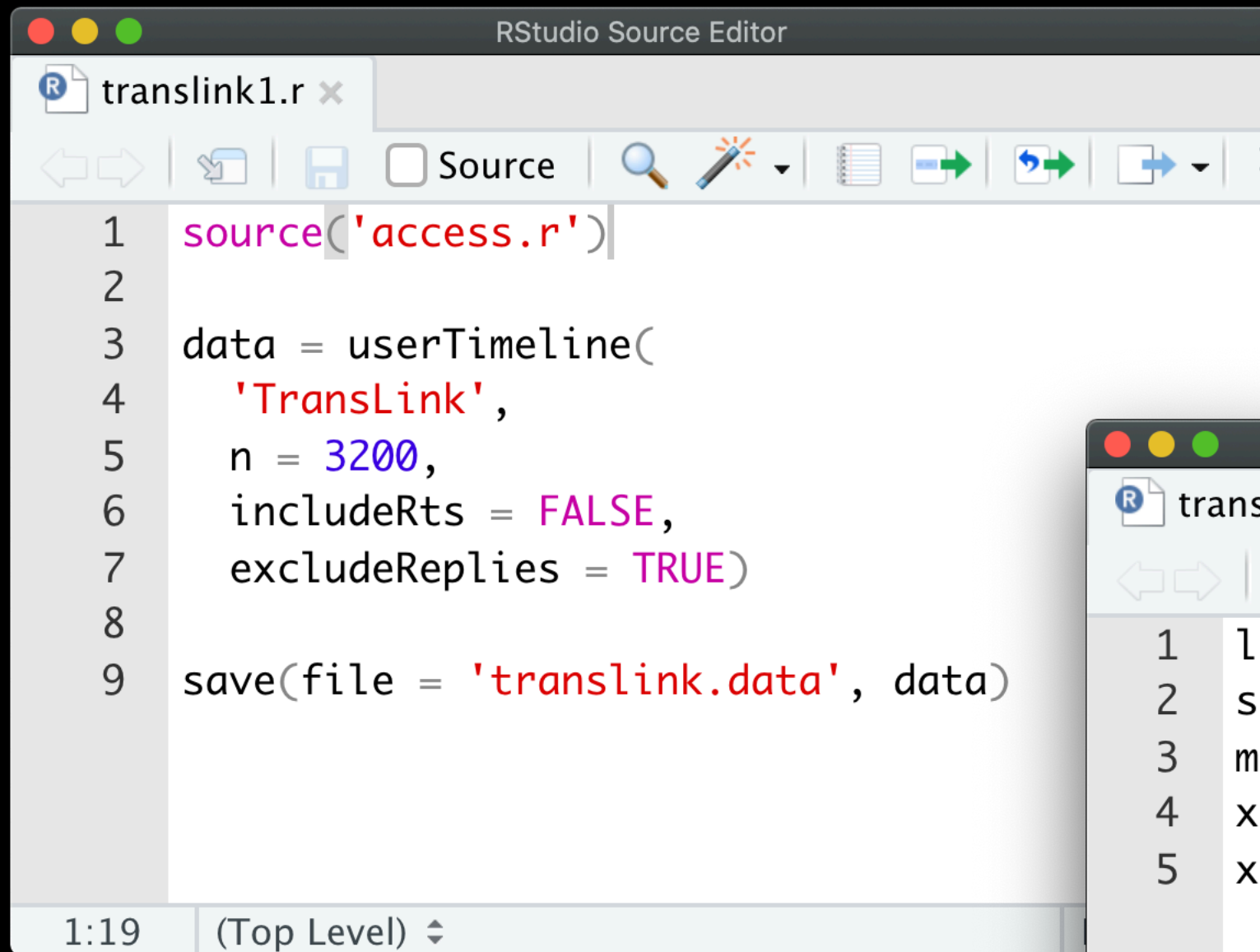
```
1 consumer_key = 'yH
2 consumer_secret =
3 access_token = '89
4 access_secret = '3
5
6 setup_twitter_oauth(
7   consumer_key,
8   consumer_secret,
9   access_token,
10  access_secret
11 )
12
13 rm(
14   c(
15     'consumer_key',
16     'consumer_secret',
17     'access_token',
18     'access_secret'
19   )
20 )
```

18:5

(Top Level) ▾

R Script ▾

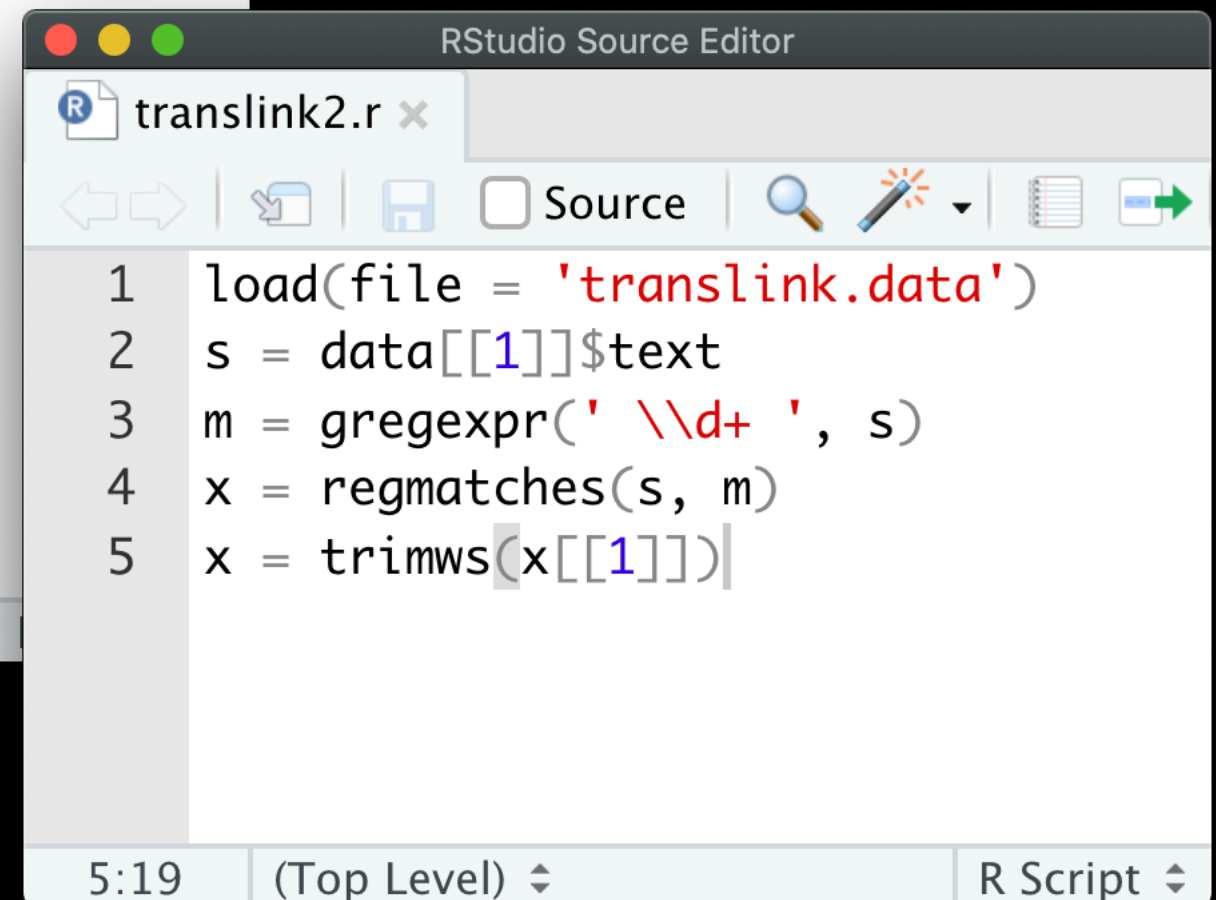
TransLink Example



RStudio Source Editor window titled 'translink1.r'. The code is as follows:

```
1 source('access.r')
2
3 data = userTimeline(
4   'TransLink',
5   n = 3200,
6   includeRts = FALSE,
7   excludeReplies = TRUE)
8
9 save(file = 'translink.data', data)
```

The status bar at the bottom shows '1:19' and '(Top Level)'.



RStudio Source Editor window titled 'translink2.r'. The code is as follows:

```
1 load(file = 'translink.data')
2 s = data[[1]]$text
3 m = gregexpr(' \\d+ ', s)
4 x = regmatches(s, m)
5 x = trimws(x[[1]])
```

The status bar at the bottom shows '5:19', '(Top Level)', and 'R Script'.

Further Reading

- *Automated data collection with R : a practical guide to web scraping and text mining* chapters 9.1.11 and 14.
- TwitterR manual: <https://cran.r-project.org/web/packages/twitterR/twitterR.pdf>

Web Scraping by Hand

- What is a website?
 - HTML/CSS/Javascript/backend
- Getting a website into R:
 - `library(RCurl)`
 - `url = 'https://www.reuters.com/article/us-baseball-mlb-lawsuit/astros-red-sox-major-league-baseball-urge-dismissal-of-sign-stealing-lawsuit-idUSKCN20H0SK'`
 - `s = getURL(url)`
 - `cat(s, file = 'article.txt')`

Extracting Website Text

- Two methods:
- Design regular expressions to extract text of article
 - Look at the string and find text that appears before and after the text of the article
 - Make a regular expression to extract all text between those markers
- Parse the HTML into a tree using an R library such as XML

Further Reading

- *Automated data collection with R : a practical guide to web scraping and text mining* chapters 8 and 9.
- TwitterR manual: <https://cran.r-project.org/web/packages/twitterR/twitterR.pdf>

Shiny Apps

- Shiny Apps tutorials:
 - <https://shiny.rstudio.com/tutorial/#video-tutorials>
 - <https://shiny.rstudio.com/tutorial/#written-tutorials>