Introduction: Working With Web Data in R

WORKING WITH WEB DATA IN R

Oliver Keyes & Charlotte Wickham
Instructors

datacamp

Working with Web Data in R

- Downloading files and using specialized packages to get data from web
- `httr` package to query APIs using `GET()` and `POST()`
- JSON and XML: data formats commonly returned
- CSS to navigate and extract data from webpages

Importing data from a URL

```
read.csv("http://website.url/remote-file.csv")
```

Downloading data from a URL

```
download.file(
    url = "http://website.url/remote-file.csv",
    destfile = "local-file.csv"
)
```

Let's practice!

WORKING WITH WEB DATA IN R

Understanding Application Programming Interfaces

WORKING WITH WEB DATA IN R

Oliver Keyes
Instructor

datacamp

Application Programming Interfaces

- "Websites, but for machines"
- Can be used to expose data automatically
- Lets you make queries for specific bits of that data

API Clients

- Native (in R!) interfaces to APIs
- Hides API complexity
- Lets you read data in as R objects

Using API Clients

- Always use a client if you can
- Find them by googling "CRAN [name of website]"
- Only write code you have to write

pageviews

```
library(pageviews)
article_pageviews(article = "R_(programming_language)")
```

Let's practice!

WORKING WITH WEB DATA IN R

Access Tokens and APIs

WORKING WITH WEB DATA IN R

Charlotte Wickham
Instructor

datacamp

API etiquette

- Overwhelming the API means you can't use it
- Overwhelming the API means nobody *else* can use it
- APIs issue "access tokens" to control and identify use

Getting access tokens

- Usually requires registering your email address
- Sometimes providing an explanation
- Example: **https://www.wordnik.com/** which requires both!

birdnik

- `birdnik` a package that wraps the Wordnik API
- Provide API key in `key` argument in `birdnik` functions

Let's practice!

WORKING WITH WEB DATA IN R