



WORKING WITH WEB DATA IN R

Introduction: Working With Web Data in R

Oliver Keyes & Charlotte Wickham
Instructors



Working with Web Data in R

- Downloading files and using specialised packages to get data from web



Working with Web Data in R

- Downloading files and using specialised packages to get data from web
- `httr` package to query APIs using `GET()` and `POST()`



Working with Web Data in R

- Downloading files and using specialised packages to get data from web
- httr package to query APIs using GET() and POST()
- JSON and XML: data formats commonly returned



Working with Web Data in R

- Downloading files and using specialised 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

- Import functions often take URLs as arguments

```
read.csv("local-file.csv")
```



Importing data from a URL

- Import functions often take URLs as arguments

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

Importing data from a URL

- Import functions often take URLs as arguments

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

- To avoid re-downloading, download.file()

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




WORKING WITH WEB DATA IN R

Let's practice!



WORKING WITH WEB DATA IN R

Understanding Application Programming Interfaces

Oliver Keyes
Instructor



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

- pageviews is an API client for number of views of Wikipedia pages

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



WORKING WITH WEB DATA IN R

Let's practice!



WORKING WITH WEB DATA IN R

Access Tokens and APIs

Charlotte Wickham
Instructor



API etiquette

- Overwhelming the API means you can't use it
- Overwhelming the API means nobody *e/se* 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



WORKING WITH WEB DATA IN R

Let's practice!