Chapter 6

Part A: Importing Data Sets from the Internet

Tao, Kara

2020-09-27

Contents

Learning outcomes

(1) Knowing how to import data into R.
(2) Being able to download data through APIs.

Files are as follows:

- `Ch6_1 2.Rmd`: This is the `.Rmd` file used to compile the pdf of this class.

https://www.neonscience.org/resources/data-tutorials?type=All&field_ds_tags_tid=All&field_ds_languages_tid=1215&page=7

Data format

.Rdata: The best way to store objects from R is with .RData files.

6.3 Importing Data Sets from the Internet

(a) Before you can analyze and visualize data, you have to get that data into R.

(b) read.csv function is commonly used.

(c) read.csv("a filepath or an URL")

6.3.1 Data from non-secure (http) URLs

What is a URL?

A URL can be typed into your browser's address bar.

import with read.csv

```
#https://raw.githubusercontent.com/karanavock/Rep_Sci/master/DB8.csv
#Open a connection.
url("https://raw.githubusercontent.com/karanavock/Rep_Sci/master/DB8.csv")
  #Modes: "r"(read), "w"(write)
#import the URL
URL_data<-read.csv("https://raw.githubusercontent.com/karanavock/Rep_Sci/master/DB8.csv")
# what kind of object is it?
class(URL_data) # A data frame is a table
#check the head of the data
head(URL_data)
summary(URL_data)
URL_complete<-URL_data[complete.cases(URL_data),]
```

6.3.1 Data from non-secure (http) URLs

import with read.table

```
URL_data_2<-read.table("https://raw.githubusercontent.com/karanavock/Rep_Sci/master/DB8.csv")
head(URL_data_2)
```

```
##
## 1 SeedlingID,Height,Width,Intersect,Plant_Area,Image_ID_Portrait,Processed_Image_ID_Portrait,Date,Pr
## 2                                                                                        1125,2.9,3.
## 3                                                                                           661,5.1
## 4                                                                                            644,4.
## 5                                                                                             864,1
```

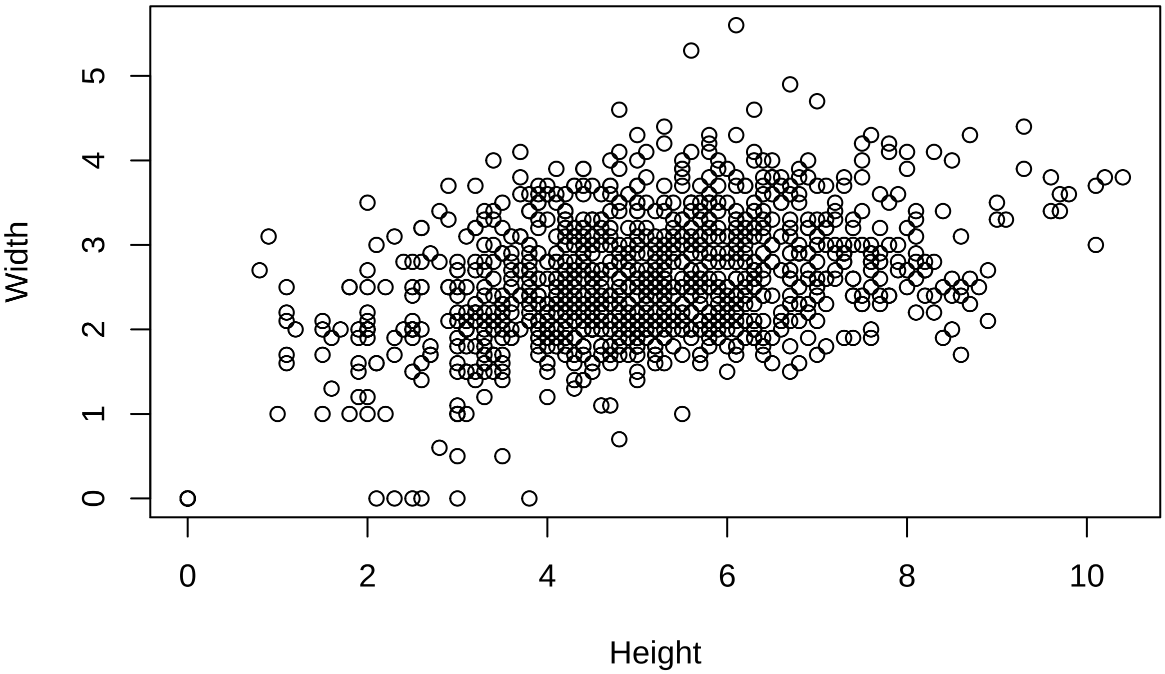

Figure 1: Plot of width in relation to height.

```
## 6                                                                                   1070,4.2,2
```

download.file is used to download a file from the Internet.

download.file from https://www.neonscience.org/data/about-data/spatial-data-maps

```
#download.file(url, destfile)
#download.file("https://www.neonscience.org/sites/default/files/NEONAquaticWatershed.zip",destfile="/Us
#setwd()
url="https://www.neonscience.org/sites/default/files/NEONAquaticWatershed.zip"
destfile="NEONAquaticWatershed.zip"
download.file(url, destfile)
```

```
download.file("https://mikethetesternz.files.wordpress.com/2019/02/apinotipa.png",destfile="IPA2.png")
```

download.file

```
#download.file("https://cshperspectives.cshlp.org/content/8/9/a023218.full.pdf",destfile="/Users/owner/
```

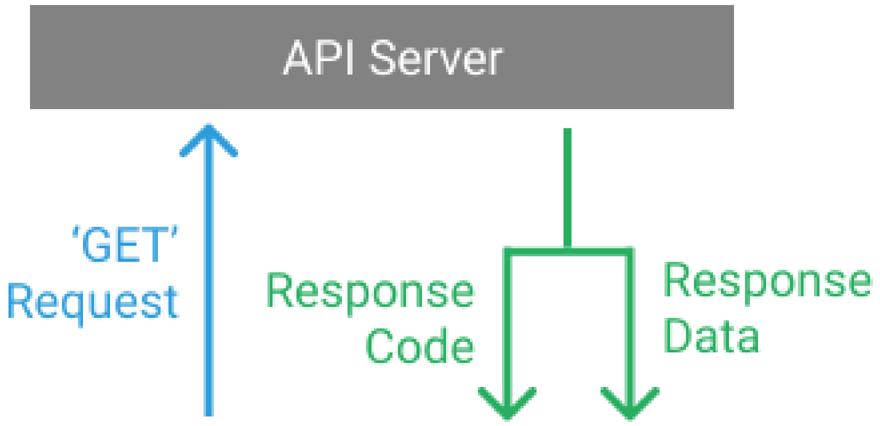

Figure 2: API Request https://www.dataquest.io/blog/r-api-tutorial/

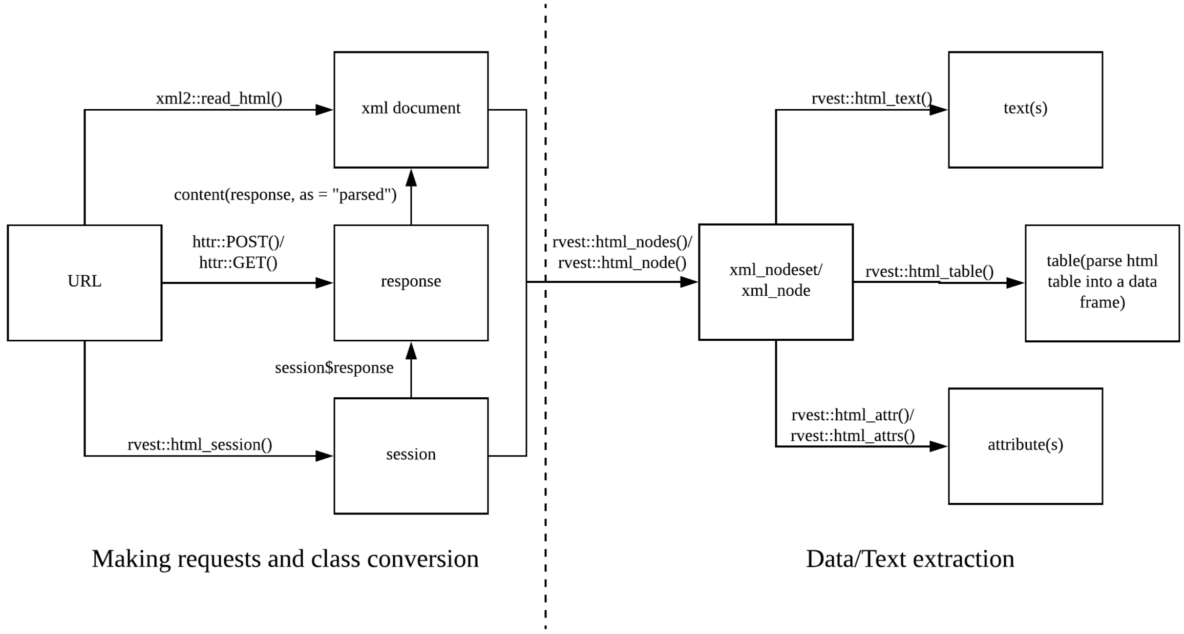

Figure 3: API Request https://github.com/yusuzech/r-web-scraping-cheat-sheet

source_data

Data retrieval from internet: (1) web scraping or (2) web APIs.

HTML (Right-click the page and click on “View Page Source,”)

Some websites have web APIs.

6.3.4 Data APIs & feeds

The term API is an acronym, and it stands for Application Programming Interface.

APIs offer users a polished way to request clean and curated data from a website.

To work with APIs in R, we need to bring in some libraries.

httr::GET (request to the server)

```r
# install neonUtilities - can skip if already installed
#install.packages("neonUtilities")
# load neonUtilities
library(neonUtilities)
```

The identifier of the NEON data product: https://data.neonscience.org/data-products/explore

```r
zipsByProduct(dpID="DP1.20093.001", #Chemical properties of surface water
              site="BIGC", #Upper Big Creek, CA
              startdate = "2019-01", enddate = "2019-04",
              check.size = FALSE #R would ask you to approve the file size
              #,package = "basic", avg = "all", savepath = NA,load = F
  )
```

```
## Downloading files totaling approximately 0.10178 MB

## Warning in dir.create(filepath): '/Users/owner/Desktop/EEB603/Ch6/
## filesToStack20093' already exists

## Downloading 3 files
##   |                                                                              |
## 3 files downloaded to /Users/owner/Desktop/EEB603/Ch6/filesToStack20093
```

To sum up

read.csv

download.file

APIs

References

source_DropboxData source_data