Importing

Nathan Young

2022-12-26

Importing Data Into R

Inspired by issues faced while completing the Week 1 Quiz in coursera for Importing Data. I feel that a deeper dive into the functions and parameters will help understanding and make the experience less of a ‘black box’.

Downloading CSV file

You can download a csv file from the internet using `download.file()`, a url, a destination file path, and a method. Its good practice to include data in a project in a sub directory, so you can check for one and create it if necessary.

```r
if(!file.exists("data")) {dir.create("data")}
```

As an example of downloading a csv and then importing for use:

```r
fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
download.file(fileURL, destfile = "./data/NAME.csv")
Data1 <- read.table("./data/NAME.csv",sep=",",header=TRUE)
head(Data1[,1:5]) # too many columns, just look at a few
```

```
##   RT SERIALNO DIVISION PUMA REGION
## 1  H      186        8  700      4
## 2  H      306        8  700      4
## 3  H      395        8  100      4
## 4  H      506        8  700      4
## 5  H      835        8  800      4
## 6  H      989        8  700      4
```

You can then analyze the data columns in myriad ways. In the quiz, it asked to find houses valued at over 1,000,000.

```r
sum(Data1$VAL >= 24, na.rm = TRUE)
```

```
## [1] 53
```

Excel spreadsheet .xlsx

Looking into alternative to what was described in notes (xlsx package).
Tidyverse seems to be a popular alternative in which the 'readxl' package can be found. For purposes of the assessment, I viewed the file in the lower right pane and was able to use the built in RStudio syntax using read_excel.

Reading XML data

In the course quiz, I encounterd an error when trying to load a https URL and was able to load the data with the note path by changing the URL to http as below.

```r
library(XML)
fileURL <- "http://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Frestaurants.xml"
doc <- xmlTreeParse(fileURL,useInternal = TRUE)
# Data can then be extracted
rootNode <- xmlRoot(doc)
# and portions analyzed
zipcode_data <- xpathSApply(rootNode, "//zipcode", xmlValue)
sum(zipcode_data == 21231)
```

```
## [1] 127
```