

Data607-wk7

Joe Rovalino

12/14/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Assignment – Working with XML and JSON in R

Pick three of your favorite books on one of your favorite subjects. At least one of the books should have more than one author. For each book, include the title, authors, and two or three other attributes that you find interesting. Take the information that you’ve selected about these three books, and separately create three files which store the book’s information in HTML (using an html table), XML, and JSON formats (e.g. “books.html”, “books.xml”, and “books.json”). To help you better understand the different file structures, I’d prefer that you create each of these files “by hand” unless you’re already very comfortable with the file formats. Write R code, using your packages of choice, to load the information from each of the three sources into separate R data frames. Are the three data frames identical? Your deliverable is the three source files and the R code. If you can, package your assignment solution up into an .Rmd file and publish to rpubs.com. [This will also require finding a way to make your three text files accessible from the web].

##Library Load

```
#Library
library(XML)
library(rjson)
library(httr)
library(bitops)
library(RCurl)
```

html

```
url<-readHTMLTable(getURL("https://raw.githubusercontent.com/jrovalino/data607-wk7/master/wk7_books.htm
url<-lapply(url[[1]], function(x) {unlist(x)})
dfhtml<-as.data.frame(url)
dfhtml
```

```
##                                Book.Title                                Author.s.
## 1 Think and Grow Rich Napoleon Hill Classic                        Napoleon Hill
## 2   How To Win Friends and Influence People                        Dale Carnegie
## 3      Reminiscences of a Stock Operator Edwin Lefevre, Roger Lowenstein
##   Year.Published      Publisher      ISBN
## 1              2019 Sound Wisdom LLC 193787950X
## 2              2009   Simon Schuster 1439167346
## 3              2006         Wiley 0471770884
```

##xml

```
url2<-xmlInternalTreeParse(getURL("https://raw.githubusercontent.com/jrovalino/data607-wk7/master/wk7_b
#str(url2)
url2<-xmlSApply(xmlRoot(url2), function(x) xmlSApply(x, xmlValue))
url2<-data.frame(t(url2), row.names = NULL)
url2
```

```
##                                title                                author
## 1 Think and Grow Rich Napoleon Hill Classic                      Napoleon Hill
## 2   How To Win Friends and Influence People                      Dale Carnegie
## 3      Reminiscences of a Stock Operator Edwin Lefevre, Roger Lowenstein
##   published_year      publisher      isbn
## 1           2019 Sound Wisdom LLC 193787950X
## 2           2009   Simon Schuster 1439167346
## 3           2006           Wiley 0471770884
```

```
##JSON
```

```
#https://stackoverflow.com/questions/37139697/extract-data-from-json-with-lapply
```

```
url3<-fromJSON(file = "https://raw.githubusercontent.com/jrovalino/data607-wk7/master/wk7_books.json")
#str(url3)
url3<-lapply(url3[[1]], function(x) {unlist(x)})
url3<-as.data.frame(do.call("rbind", url3))
url3
```

```
##                                V1
## title      Think and Grow Rich Napoleon Hill Classic
## author                                Napoleon Hill
## published_year                                2019
## publisher                                Sound Wisdom LLC
## isbn                                193787950X
```