

Assignment 7

MG

10/10/2020

Week 7 Working with XML and JSON in R

Our assignment this week is to choose 3 books (I'm choosing sports books) and include the title, author(s) and two attributes of my choosing. I will then create three separate files in the HTML, XML, and JSON file formats and post them to my github. Finally, I will write R code and load the three separate files from github and compare the three imported dataframes to see if they are identical.

Pulling in the HTML code

```
bookhtml <- "https://raw.githubusercontent.com/mjgons/DATA607/master/Data607Books.html"
bookhtml2 <- getURL(bookhtml)
bookhtml3 <- readHTMLTable(bookhtml2, header=TRUE)
bookhtml3
```

```
## $'NULL'
##                                     Title
## 1   I Live for This: Baseball's Last True Believer
## 2       If These Walls Could Talk: San Jose Sharks
## 3 The San Jose Earthquakes: A Seismic Soccer Legacy
##                                     Author      Publisher Pages
## 1 Bill Plaschke, Tommy Lasorda      Houghton Mifflin Harcourt   256
## 2  Dan Rusanowsky, Ross McKeon      Triumph Books           320
## 3              Gary Singh History Press Library Editions   146
```

Pulling in the XML code

```
bookxml <- "https://raw.githubusercontent.com/mjgons/DATA607/master/Data607Books.xml"
bookxml2 <- getURL(bookxml)
bookxml3 <- xmlToDataFrame(bookxml2)
bookxml3
```

```
##                                     Title
## 1   I Live for This: Baseball's Last True Believer
## 2       If These Walls Could Talk: San Jose Sharks
## 3 The San Jose Earthquakes: A Seismic Soccer Legacy
##                                     Author      Publisher Pages
## 1 Bill Plaschke, Tommy Lasorda      Houghton Mifflin Harcourt   256
## 2  Dan Rusanowsky, Ross McKeon      Triumph Books           320
## 3              Gary Singh History Press Library Editions   146
```

Pulling in the JSON code

```
bookjson <- "https://raw.githubusercontent.com/mjgons/DATA607/master/Data607Books2.json"
bookjson2 <- getURL(bookjson)
bookjson3 <- fromJSON(bookjson2)
bookjson3
```

```
##                               Title
## 1   I Live for This: Baseball's Last True Believer
## 2       If These Walls Could Talk: San Jose Sharks
## 3 The San Jose Earthquakes: A Seismic Soccer Legacy
##                               Author                Publisher Pages
## 1 Bill Plaschke, Tommy Lasorda      Houghton Mifflin Harcourt   256
## 2  Dan Rusanowsky, Ross McKeon      Triumph Books           320
## 3              Gary Singh History Press Library Editions   146
```

Conclusion

I created the 3 filetypes: HTML, XML, and JSON by hand as I've never worked with these filetypes before. I then pushed the files to github and then pulled them into R.

The dataframes all look similar. I have a '\$NULL' in the html file but other than that, they all look the same.