

DATA607 - Assignment Week 7

Working with XML and JSON in R

Donald Butler

10/10/2021

Introduction

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.

Create file representations of the selected books in the formats: HTML, XML, and JSON.

Are the three data frames identical?

Load required R Libraries

```
library(tidyverse)
library(XML)
library(jsonlite)
```

Read HTML Data

```
html_file <- 'https://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_5/books.html'
(html_df <- XML::readHTMLTable(html_file, as.data.frame = TRUE))
```

```
## named list()
```

Read XML Data

```
xml_file <- 'https://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_5/books.xml'
xml_file <- 'C:/Users/dab31/OneDrive/Documents/GitHub/DATA607/Homework/Assignment_5/books.xml'
(xml_df <- xmlToDataFrame(xml_file))
```

##	title	authors	year	publisher
## 1	Treasure	Clive Cussler	1988	Simon and Schuster
## 2	The Hunt for Red October	Tom Clancey	1984	Naval Institute Press

```
## 3      Command Authority Tom Clancey, Mark Greaney 2013    G.P. Putnam's Sons
##  pages                      ISBN
## 1    539              0-671-62132
## 2    387              0-870-212850
## 3    736 978-0-39-916077-9
```

Read JSON Data

```
json_file <- 'https://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_5/books.json'
(json_df <- fromJSON(json_file))
```

```
## $books
##           title                      authors year      publisher
## 1      Treasure                Clive Cussler 1988    Simon and Schuster
## 2 The Hunt for Red October          Tom Clancey 1984  Naval Institute Press
## 3      Command Authority Tom Clancey, Mark Greaney 2013    G.P. Putnam's Sons
##  pages                      ISBN
## 1    539              0-671-62132
## 2    387              0-870-212850
## 3    736 978-0-39-916077-9
```