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)
library(xml2)
library(rvest)
```

Read HTML Data

```
html_file <- 'http://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_5/books.html'

html_df <- read_html(html_file) %>%
  html_elements('table') %>%
  html_table()

html_df
```

```
## [[1]]
## # A tibble: 3 x 6
##   Title                  Author                  Year Publisher Pages ISBN
##   <chr>                  <chr>                  <int> <chr>     <int> <chr>
## 1 Treasure               Clive Cussler           1988 Simon an~   539 0-67~
## 2 The Hunt for Red October Tom Clancey           1984 Naval In~   387 0-87~
## 3 Command Authority      Tom Clancey, Mark Greaney 2013 G.P. Put~   736 978-~
```

Read XML Data

```
xml_file <- 'http://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_5/books.xml'

xml_df <- read_xml(xml_file) %>%
  xmlParse() %>%
  xmlToDataFrame()

xml_df
```

```
##                        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)

json_df
```

```
## $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
```