# IS607 Week7

*Sharon Morris*

*October 12, 2016*

**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?

**Load Library**

```
library(XML)
library(RCurl)
```

```
## Loading required package: bitops
```

**HTML**

```
# Get url
htmlURL <- "https://raw.githubusercontent.com/indianspice/IS607/master/books.hml"
htmlRead <- readLines( con = htmlURL)
```

```
# Extract table information
htmlTbl <- readHTMLTable(htmlRead, stringAsFactors = FALSE)
htmlTbl
```

```
## $`NULL`
##                                                               Book Title
## 1 Origins: George Washington Secret Six: The Spy Ring that Saved The American Revolution
## 2                                              Che Guevara: A Revolutionary Life
## 3                                                              Pocahontas
##                  Authors           Publisher          ISBN
## 1 Brian Kilmeade, Don Yeager          Sentinel 978-1595231031
## 2          Jon Lee Anderson       Grove Press 978-0802144119
## 3           Joseph Bruchac HMN Books for Young Readers 978-0152054650
```

**XML**

```
# Get xml code
xmlURL <- "https://raw.githubusercontent.com/indianspice/IS607/master/Week-7/books.xml"
xmlDta = getURL(xmlURL, ssl.verifypeer=0L, followlocation =1L)
xmlPrse = xmlParse(xmlDta)
xmlLst = xmlToList(xmlPrse)
xmlDf = as.data.frame(matrix(unlist(xmlLst), nrow = 4, ncol = 4, byrow = TRUE))
head(xmlDf)
```

```
##                                                                        V1
## 1 George Washington Secret Six: The Spy Ring that Saved The American Revolution
## 2                                                         978-1595231031
```

```
## 3                                                             Grove Press
## 4                                                          Joseph Bruchac
##                              V2                              V3
## 1            Brian Kilmeade                         Don Yeager
## 2                      1 Che Guevara: A Revolutionary Life
## 3              978-0802144119                             2
## 4 HMN Books for Young Readers              978-0152054650
##              V4
## 1       Sentinel
## 2 Jon Lee Anderson
## 3     Pocahontas
## 4             3
```

**JSON**

```r
library(jsonlite)
jsonUrl <- fromJSON("https://raw.githubusercontent.com/indianspice/IS607/master/Week-7/books.json")
jsonDta <- data.frame(jsonUrl, stringsAsFactors = FALSE)
jsonDta
```

```
##                                                 Books.I.Enjoyed.book
## 1 George Washington Secret Six: The Spy Ring that Saved The American Revolution
## 2                                     Che Guevara: A Revolutionary Life
## 3                                                          Pocahontas
##       Books.I.Enjoyed.author   Books.I.Enjoyed.publisher
## 1 Brian Kilmeade, Don Yeager                    Sentinel
## 2          Jon Lee Anderson                 Grove Press
## 3            Joseph Bruchac HMN Books for Young Readers
##   Books.I.Enjoyed.ISBN
## 1      978-1595231031
## 2      978-0802144119
## 3      978-0152054650
```