Assignment 6

Matthew Roland

2023-10-12

Loading the HTML, JSON, & XML files into R

```
###HTML

##This will load the table as a list containing the head and body. To fix this, I will need a handy bit

books_html <- read_html("https://raw.githubusercontent.com/Mattr5541/DATA-607/main/Week%206/books.html")

##Now the lists will be condensed into a singular list containing the table laid out in a more organize

books_html <- html_table(books_html)



books_html <- as.data.frame(books_html[[1]])
kable(books_html)
```

Title	Author	Genre	Release	Sales (estimate)	Adapted into Movie Format (y/n)
DUNE	Frank Herbert	Sci-Fi	1965	20,000,000	y
The Three Body Problem	Liu Cixin	Sci-Fi	2008	8,000,000	n
The Long Earth	Terry Pratchett and Stephen Baxter	Sci-Fi	2012	100,000,000	n

```
###JSON

##THis code will load the json table into R. However, it seems to convert the sales values into scienti
books_json <- fromJSON("https://raw.githubusercontent.com/Mattr5541/DATA-607/main/Week%206/books_new.js

books_json$`Sales (estimate)` <- format(books_json$`Sales (estimate)`, scientific = F)

kable(books_json)
```

Title	Author	Genre	Release	Sales (estimate)	Adapted into Movie Format (y/n)
DUNE	Frank Herbert	Sci-Fi	1965	20000000	y
The Three Body Problem	Liu Cixin	Sci-Fi	2008	8000000	n
The Long Earth	Terry Pratchett and Stephen Baxter	Sci-Fi	2012	100000000	n

```r
###XML

##And finally, let's load in an XML table

books_xml <- read_xml("https://raw.githubusercontent.com/Mattr5541/DATA-607/main/Week%206/books.xml")

##But it saved every element in the schema as a list, so I'll have to do something that's a little less

Title <- xml_text(xml_find_all(books_xml, "//Title"))
Author <- xml_text(xml_find_all(books_xml, "//Author"))
Genre <- xml_text(xml_find_all(books_xml, "//Genre"))
Release <- xml_text(xml_find_all(books_xml, "//Release"))
`Sales (estimate)` <- xml_text(xml_find_all(books_xml, "//Sales"))
Adapted_Into_Movie_Format <- xml_text(xml_find_all(books_xml, "//Adapted_Into_Movie_Format"))

books_xml <- data.frame(Title = Title,
                Author = Author,
                Genre = Genre,
                Release = as.numeric(Release),
                `Sales (estimate)` = `Sales (estimate)`, Adapted_Into_Movie_Format = Adapted_Into_Movie_

kable(books_xml)
```

Title	Author	Genre	Release	Sales..estimate.	Adapted_Into_Movie_Format
DUNE	Frank Herbert	Sci-Fi	1965	20,000,000	y
The Three Body Problem	Liu Cixin	Sci-Fi	2008	8,000,000	n
The Long Earth	Terry Pratchett and Stephen Baxter	Sci-Fi	2012	100,000,000	n

So, to summarize the results of this exercise, the three dataframes were not entirely identical, and required some cleaning for standardization purposes. However, they were all rather similar.