

hw7

Yan Lin

October 15, 2017

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

```
## Warning: package 'jsonlite' was built under R version 3.4.2
```

```
## HTML file
```

```
theURL <- scan(file="https://raw.githubusercontent.com/shupa001/DATA607/master/books.html",what="charac
```

```
## parse HTML table
```

```
books.html <- readHTMLTable(theURL, header=TRUE, stringASFactors = FALSE)
```

```
## convert to data frame and change the column names
```

```
books.html <- as.data.frame(books.html)
```

```
colnames(books.html) <- c("Title","First Author","Second Author","Amazon Review","Best Quote/Interest")
```

```
books.html
```

```
##
##           Title      First Author
## 1           The Power of Now      Eckhart Tolle
## 2           How Not to Be Wrong Jordan Ellenberg
## 3 Dictionary of Finance and Investment Terms      John Downes
##      Second Author Amazon Review      Best Quote/Interest
## 1           4.5 All problems are illusions of the mind.
## 2           4.5           Improbable things happen a lot.
## 3 Jordan Goodman      5.0 Explanation of 5,000 + financial terms.
```

```
## XML file
```

```
xml.URL <- scan(file="https://raw.githubusercontent.com/shupa001/DATA607/master/books.xml",what="charac
```

```
## parse XML table
```

```
books.xml <- xmlParse(xml.URL)
```

```
books.root <- xmlRoot(books.xml)
```

```
books.xml <- xmlToDataFrame(books.root,stringsAsFactors = FALSE)
```

```
books.xml
```

```
##
##           Title      Author
## 1           The Power of Now      Eckhart Tolle
## 2           How Not to Be Wrong Jordan Ellenberg
## 3 Dictionary of Finance and Investment Terms      John Downes
##      Author2 Amazon_Review      Best_Quote_or_Interest
## 1           4.5 All problems are illusions of the mind.
## 2           4.5           Improbable things happen a lot.
## 3 Jordan Goodman      5.0 Explanation of 5,000 + financial terms.
```

```
## JSON file
```

```
json.URL <- scan(file="https://raw.githubusercontent.com/shupa001/DATA607/master/books.json",what="char
```

```
## parse JSON table
```

```
books.json <- fromJSON(json.URL)
```

```
## output  
books.json
```

```
## `$Favorite Books`
```

```
##                               Title                               Author  
## 1                      The Power of Now                      Eckhart Tolle  
## 2                      How Not to Be Wrong                      Jordan Ellenberg  
## 3 Dictionary of Finance and Investment Terms John Downes, Jordan Goodman  
##   Amazon Review                      Best Quote/Interest  
## 1                      4.5 All problems are illusions of the mind.  
## 2                      4.5      Improbable things happen a lot.  
## 3                      5.0 Explanation of 5,000 + financial terms.
```

```
## conclusion
```

```
##XML and HTML need seperate columns when it comes to different authors, but JSON can combine them together  
##It is much easier to parse JSON file, then XML, and then HTML.
```