Assignment 7

Munkhnaran Gankhuyag

October 14, 2017

```
library(dbConnect)
```

```
## Loading required package: RMySQL
```

```
## Loading required package: DBI
```

```
## Loading required package: gWidgets
```

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

```
library(gWidgets)
library(bitops)
library(RCurl)
library(jsonlite)
library(XML)
library(knitr)
```

I created my table in SQL and I connected through R.

```
con =dbConnect(MySQL(), user ='root',  host = 'localhost',
               dbname = 'CUNY')

dbListTables(con)
```

```
## [1] "books"
```

```
books <- "select * from books;"
books1 <- dbGetQuery(con,books)
```

```
kable(books1)
```

Title	Author	Year_Published	Category
Data Science for Business	Foster Provost, Tom Fawcett	2013	Business
OpenIntro Statistics	David M Diez, Christopher D Barr, Mine C¸ etinkaya	2015	Statistics
R for Data Science	Garrett Grolemund, Hadley Wickham	2017	Computer

First I converted my table in HTML format and uploaded in Github.

```
library(tableHTML)
```

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

```
tableHTML(books1)
```

Title

Author

Year_Published

Category

1

Data Science for Business

Foster Provost, Tom Fawcett

2013

Business

2

OpenIntro Statistics

David M Diez, Christopher D Barr, Mine C¸ etinkaya

2015

Statistics

3

R for Data Science

Garrett Grolemund, Hadley Wickham

2017

Computer

```
write_tableHTML(tableHTML(books1), file = "books.html")
```

Below is how structure of the table in hmtl format.

```
books_html <- "https://raw.githubusercontent.com/mikegankhuyag/607-HW/master/books.html"

books_html1 <- htmlParse(getURL(books_html))
books_html1
```

```
## <!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/REC-html40/loose
## <html><body><table style="border-collapse:collapse;" class="table_6542" border="1">
## <thead><tr>
## <th id="tableHTML_header_1"> </th>
##   <th id="tableHTML_header_2">Title</th>
##   <th id="tableHTML_header_3">Author</th>
##   <th id="tableHTML_header_4">Year_Published</th>
##   <th id="tableHTML_header_5">Category</th>
## </tr></thead>
## <tbody>
## <tr>
## <td id="tableHTML_rownames">1</td>
##   <td id="tableHTML_column_1">Data Science for Business</td>
##   <td id="tableHTML_column_2">Foster Provost, Tom Fawcett</td>
##   <td id="tableHTML_column_3">2013</td>
##   <td id="tableHTML_column_4">Business</td>
## </tr>
## <tr>
## <td id="tableHTML_rownames">2</td>
##   <td id="tableHTML_column_1">OpenIntro Statistics</td>
##   <td id="tableHTML_column_2">David M Diez, Christopher D Barr, Mine CÂ¸ etinkaya</td>
##   <td id="tableHTML_column_3">2015</td>
##   <td id="tableHTML_column_4">Statistics</td>
## </tr>
## <tr>
```

	Title	Author	Year_Published	Category
1	Data Science for Business	Foster Provost, Tom Fawcett	2013	Business
2	OpenIntro Statistics	David M Diez, Christopher D Barr, Mine CÂ¸ etinkaya	2015	Statistics
3	R for Data Science	Garrett Grolemund, Hadley Wickham	2017	Computer

```
## <td id="tableHTML_rownames">3</td>
##   <td id="tableHTML_column_1">R for Data Science</td>
##   <td id="tableHTML_column_2">Garrett Grolemund, Hadley Wickham</td>
##   <td id="tableHTML_column_3">2017</td>
##   <td id="tableHTML_column_4">Computer</td>
## </tr>
## </tbody>
## </table></body></html>
##
```

Below is the HMTL read as a table

```
books_html2 <- readHTMLTable(books_html1)
kable(books_html2)
```

XML was a little easier to understand, so I wrote the table below in the format.

```
books_xml <- "https://raw.githubusercontent.com/mikegankhuyag/607-HW/master/Books_xml_final.xml"
books_xml1 <- getURL(books_xml)
books_xml2 <- xmlParse(books_xml1)
books_xml2
```

```
## <?xml version="1.0"?>
## <Books>
##   <book category="Data Science">
##     <Title>Data Science for Business</Title>
##     <Author>Foster Provost, Tom Fawcett</Author>
##     <Year_Published>2013</Year_Published>
##     <Category>Business</Category>
##   </book>
##   <book category="Data Science">
##     <Title>OpenIntro Statistics</Title>
##     <Author>David M Diez, Christopher D Barr, Mine C&#xB8; etinkaya-Rundel</Author>
##     <Year_Published>2015</Year_Published>
##     <Category>Statistics</Category>
##   </book>
##   <book category="Data Science">
##     <Title>R for Data Science</Title>
##     <Author>Garrett Grolemund, Hadley Wickham</Author>
##     <Year_Published>2017</Year_Published>
##     <Category>Computer</Category>
##   </book>
## </Books>
##
```

Below is the XML language read as a table.

```
books_xml_table <- xmlToDataFrame(books_xml2, stringsAsFactors = FALSE)
kable(books_xml_table)
```

Title	Author	Year Published	Category
Data Science for Business	Foster Provost, Tom Fawcett	2013	Business
Data Science for Business	Foster Provost, Tom Fawcett	2015	Statistics
R for Data Science	Garrett Grolemund, Hadley Wickham	2017	Computer

Title	Author	Year_Published	Category
Data Science for Business	Foster Provost, Tom Fawcett	2013	Business
OpenIntro Statistics	David M Diez, Christopher D Barr, Mine CÂ¸ etinkaya-Rundel	2015	Statistics
R for Data Science	Garrett Grolemund, Hadley Wickham	2017	Computer

Below is the table written in JSN format.

```
books_jsn <- getURL("https://raw.githubusercontent.com/mikegankhuyag/607-HW/master/Books.json.json")
books_jsn
```

```
## [1] "{\"books\" :[\n    {\n    \"Title\" : \"Data Science for Business\",\n    \"Author\" : \"Foster
```

```
books_jsn2 <- fromJSON(books_jsn)
```

Below is the JSN language read as a table.

```
kable(books_jsn2)
```

Overall, I had a little trouble understanding HTML format. I used to play around with it back when Myspace was popular. I can see that JSN and XML are a lot more structured and easier to read. This was a great assignment to get a feel for the different languages.