In Class Lecture Week 11

Peyton Hall

03/21/2024

```r
rm(list=ls())
```

```r
# install.packages("rvest") # install r vest
library(rvest)
```

```r
# HTML Windows Note: right click --> "view page source"
# HTML MacBook Pro Note: Press "Develop" --> "show page source"
#                              (or shortcut: command + U) --> view HTML code
```

Data scraping

```r
# scrape the title of a webpage
url <- 'http://quotes.toscrape.com'
webpage <- read_html(url)

Title <- webpage%>%
  html_nodes("title")%>%
  html_text()
Title
```

```
## [1] "Quotes to Scrape"
```

```r
# extracting quotes
Quotes <- webpage%>%
  html_nodes(".quote .text") %>%
  html_text()
Quotes
```

```
##  [1] ""The world as we have created it is a process of our thinking. It cannot be changed without cha
##  [2] ""It is our choices, Harry, that show what we truly are, far more than our abilities.""
##  [3] ""There are only two ways to live your life. One is as though nothing is a miracle. The other is
##  [4] ""The person, be it gentleman or lady, who has not pleasure in a good novel, must be intolerably
##  [5] ""Imperfection is beauty, madness is genius and it’s better to be absolutely ridiculous than abs
##  [6] ""Try not to become a man of success. Rather become a man of value.""
##  [7] ""It is better to be hated for what you are than to be loved for what you are not.""
##  [8] ""I have not failed. I’ve just found 10,000 ways that won’t work.""
##  [9] ""A woman is like a tea bag; you never know how strong it is until it’s in hot water.""
## [10] ""A day without sunshine is like, you know, night.""
```

```
# extracting author names
Authors <- webpage %>%
  html_nodes(".quote .author") %>%
  html_text()
Authors
```

```
##  [1] "Albert Einstein"   "J.K. Rowling"      "Albert Einstein"
##  [4] "Jane Austen"       "Marilyn Monroe"    "Albert Einstein"
##  [7] "André Gide"        "Thomas A. Edison"  "Eleanor Roosevelt"
## [10] "Steve Martin"
```

```
FamousQuotes <- data.frame(Quotes, Authors)
FamousQuotes
```

```
##
## 1                  "The world as we have created it is a process of our thinking. It cannot be change
## 2                                                "It is our choices, Harry, that show what we truly a
## 3  "There are only two ways to live your life. One is as though nothing is a miracle. The other is as
## 4                             "The person, be it gentleman or lady, who has not pleasure in a good no
## 5                      "Imperfection is beauty, madness is genius and it’s better to be absolutely ri
## 6                                                                  "Try not to become a man of succes
## 7                                                   "It is better to be hated for what you are than t
## 8                                                                    "I have not failed. I’ve just fo
## 9                                                "A woman is like a tea bag; you never know how stron
## 10                                                                                  "A day without su
##              Authors
## 1    Albert Einstein
## 2       J.K. Rowling
## 3    Albert Einstein
## 4        Jane Austen
## 5     Marilyn Monroe
## 6    Albert Einstein
## 7         André Gide
## 8   Thomas A. Edison
## 9  Eleanor Roosevelt
## 10      Steve Martin
```

```
# Use functions html_nodes() and html_table() from the package of rvest

# Extracting table data
URL2 <- "https://www.metrostate.edu/academics/courses?combine=&cid=DATA&title"
CourseData <- read_html(URL2) %>%
  html_nodes("table") %>%
  html_table(fill = TRUE) %>%
  .[[1]]
CourseData
```

```
## # A tibble: 5 x 3
##   ‘Course ID\n  \n    Sort descending‘ ‘Course title and goal areas‘     Credits
##   <chr>                                <chr>                             <chr>
## 1 DATA 211                             Data Science and Visualization D~ 4
## 2 DATA 350I                            Data Science Internship DATA 350I 1-5
## 3 DATA 401                                  Applied Machine Learning DATA 401 4
## 4 DATA 499                                  Data Science Capstone DATA 499    4
## 5 DATA 611                                  Data Science and Analytics DATA ~ 3
```

```
# Microsoft excel introduction steps:
# open the table -> click on "one cell" -> pick the source and click on it ->
# press "submit"/ check -> press "new sheet" from "create your own pivot table"
# -> from there, it is the same as how it was on D2L
# Note: just submit the excel sheet
# Slides we worked with for excel: 25, 27, 29
```