```
Lab52.R ×    Lab51.R ×    position1 ×

  Source on Save                                          Run      Source
 1  library(rvest)
 2  library(RCurl)
 3  library(xml2)                              10/10  10
 4  library(stringr)
 5
 6  course_url = "http://www.sfu.ca/outlines.html?2021/spring/macm/201/d300"
 7  course_page = readLines(course_url)
 8  position1 = grep('<h3',course_page)
 9  heading = trimws(course_page[position1])
10  heading = str_replace_all(heading,"<.*?>","")
11  heading = str_c(heading[1], heading[2], sep = "     ")
12  heading
13
14
15  position2 = grep('<h1', course_page)
16  course_code1 = trimws(course_page[position2])
17  course_code = course_code1[2]
18  neglect = str_locate(course_code, "-")
19  neglect1 = str_locate(course_code, "<span>")
20  course_code = substr(course_code, neglect+1,neglect1-1)
21  course_code = trimws(course_code)
22  course_code
23
24  course = function()
25 - {
26      position2 = grep('<h1', course_page)
27      course_code1 = trimws(course_page[position2])
28      course_code = course_code1[2]
29      neglect = str_locate(course_code, "-")
30      neglect1 = str_locate(course code, "<span>")

15:1   (Top Level)                                         R Script

Console  C:/Users/chopr/OneDrive/Desktop/STAT 240/
> heading
[1] "Class Number: 3620      Delivery Method: In Person"
> |
```

```
23
24  course = function()
25 ▾ {
26      position2 = grep('<h1', course_page)
27      course_code1 = trimws(course_page[position2])
28      course_code = course_code1[2]
29      neglect = str_locate(course_code, "-")
30      neglect1 = str_locate(course_code, "<span>")
31      course_code = substr(course_code, neglect+1,neglect1-1)
32      course_code = trimws(course_code)
33      course_code
34
35      position3 = grep('<h4>Instructor:', course_page)
36      instructor1 = course_page[position3 + 1]
37      instructor1 = trimws(instructor1)
38      instructor1 = str_replace_all(instructor1,"<.*?>", "")
39      instructor1
40
41      Data = list(course = course_code, instructor = instructor1)
42      Data
43 ▴ }
44  course()$instructor
45  course()$course
46  |
```

46:1   (Top Level) ⬦

Console C:/Users/chopr/OneDrive/Desktop/STAT 240/ ⬈

```
> course = function()
+ {
+    position2 = grep('<h1', course_page)
+    course_code1 = trimws(course_page[position2])
+    course_code = course_code1[2]
+    neglect = str_locate(course_code, "-")
+    neglect1 = str_locate(course_code, "<span>")
+    course_code = substr(course_code, neglect+1,neglect1-1)
+    course_code = trimws(course_code)
+    course_code
+
+    position3 = grep('<h4>Instructor:', course_page)
+    instructor1 = course_page[position3 + 1]
+    instructor1 = trimws(instructor1)
+    instructor1 = str_replace_all(instructor1,"<.*?>", "")
+    instructor1
+
+    Data = list(course = course_code, instructor = instructor1)
+    Data
+ }
> course()$instructor
[1] "Lloyd Elliott"
> course()$course
[1] "STAT 240"
> |
```

```
 5
 6  course_url = "https://www.sfu.ca/outlines.html?2021/spring/stat/240/d100"
 7  course_page = readLines(course_url)
 8  position1 = grep('<h3',course_page)
 9  heading = trimws(course_page[position1])
10  heading = str_replace_all(heading,"<.*?>","")
11  heading = str_c(heading[1], heading[2], sep = "      ")
12  heading
13
14
15  position2 = grep('<h1', course_page)
16  course_code1 = trimws(course_page[position2])
17  course_code = course_code1[2]
18  neglect = str_locate(course_code, "-")
19  neglect1 = str_locate(course_code, "<span>")
20  course_code = substr(course_code, neglect+1,neglect1-1)
21  course_code = trimws(course_code)
22  course_code
23
24  course = function()
25 ▾ {
26    position2 = grep('<h1', course_page)
27    course_code1 = trimws(course_page[position2])
28    course_code = course_code1[2]
29    neglect = str_locate(course_code, "-")
30    neglect1 = str_locate(course_code, "<span>")
```

24:1   🔳 course() ⇕

Console C:/Users/chopr/OneDrive/Desktop/STAT 240/ ⤴

```
> position2 = grep('<h1', course_page)
> course_code1 = trimws(course_page[position2])
> course_code = course_code1[2]
> neglect = str_locate(course_code, "-")
> neglect1 = str_locate(course_code, "<span>")
> course_code = substr(course_code, neglect+1,neglect1-1)
> course_code = trimws(course_code)
> course_code
[1] "STAT 240"
> |
```

```
23
24  course = function()
25 ▾ {
26      position2 = grep('<h1', course_page)
27      course_code1 = trimws(course_page[position2])
28      course_code = course_code1[2]
29      neglect = str_locate(course_code, "-")
30      neglect1 = str_locate(course_code, "<span>")
31      course_code = substr(course_code, neglect+1,neglect1-1)
32      course_code = trimws(course_code)
33      course_code
34
35      position3 = grep('<h4>Instructor:', course_page)
36      instructor1 = course_page[position3 + 1]
37      instructor1 = trimws(instructor1)
38      instructor1 = str_replace_all(instructor1,"<.*?>", "")
39      instructor1
40
41      Data = list(course = course_code, instructor = instructor1)
42      Data
43 ▴ }
44  course()$instructor
45  course()$course
46  |
```
46:1   (Top Level) ⬍

Console C:/Users/chopr/OneDrive/Desktop/STAT 240/ ⟳

```
> course = function()
+ {
+    position2 = grep('<h1', course_page)
+    course_code1 = trimws(course_page[position2])
+    course_code = course_code1[2]
+    neglect = str_locate(course_code, "-")
+    neglect1 = str_locate(course_code, "<span>")
+    course_code = substr(course_code, neglect+1,neglect1-1)
+    course_code = trimws(course_code)
+    course_code
+
+    position3 = grep('<h4>Instructor:', course_page)
+    instructor1 = course_page[position3 + 1]
+    instructor1 = trimws(instructor1)
+    instructor1 = str_replace_all(instructor1,"<.*?>", "")
+    instructor1
+
+    Data = list(course = course_code, instructor = instructor1)
+    Data
+ }
> course()$instructor
[1] "Amarpreet Rattan"
> course()$course
[1] "MACM 201"
> |
```

```
14
15  position2 = grep('<h1', course_page)
16  course_code1 = trimws(course_page[position2])
17  course_code = course_code1[2]
18  neglect = str_locate(course_code, "-")
19  neglect1 = str_locate(course_code, "<span>")
20  course_code = substr(course_code, neglect+1,neglect1-1)
21  course_code = trimws(course_code)
22  course_code
23
24  course = function()
25 ▾ {
26    position2 = grep('<h1', course_page)
27    course_code1 = trimws(course_page[position2])
28    course_code = course_code1[2]
29    neglect = str_locate(course_code, "-")
30    neglect1 = str_locate(course_code, "<span>")
```
24:1   🗊 course() ⬍

Console C:/Users/chopr/OneDrive/Desktop/STAT 240/ 🔗

```
> position2 = grep('<h1', course_page)
> course_code1 = trimws(course_page[position2])
> course_code = course_code1[2]
> neglect = str_locate(course_code, "-")
> neglect1 = str_locate(course_code, "<span>")
> course_code = substr(course_code, neglect+1,neglect1-1)
> course_code = trimws(course_code)
> course_code
[1] "MACM 201"
> |
```

```r
 3  library(xml2)
 4  library(stringr)
 5
 6  boxoffice = function()
 7 ▾{
 8      movie_url = "https://www.imdb.com/chart/boxoffice"
 9      movie_table = read_html(movie_url)
10      table1 = html_table(html_nodes(movie_table,"table"))
11      Name1 = table1[[1]]$Title
12      BoxOffice1 = table1[[1]]$Gross
13      Week1 = table1[[1]]$Weeks
14      BoxOffice2 = vector()
15      PerWeek1 = vector()
16      BoxOffice2 = substr(BoxOffice1, 2,str_length(BoxOffice1)-1)
17
18      BoxOffice2 = as.numeric(BoxOffice2)
19
20      PerWeek1 = BoxOffice2/Week1
21      PerWeek1 = str_c("$", PerWeek1)
22      PerWeek1 = str_c(PerWeek1,"M")
23
24
25      Data = list(Name = Name1, BoxOffice = BoxOffice1, PerWeek = PerWeek1)
26      Data
27 ▴}
28  boxoffice()$Name
29  boxoffice()$BoxOffice
30  boxoffice()$PerWeek
31
```

```
32:1    (Top Level) ≎                                                                R
```

```
Console  C:/Users/chopr/OneDrive/Desktop/STAT 240/
+
+       BoxOffice2 = as.numeric(BoxOffice2)
+
+       PerWeek1 = BoxOffice2/Week1
+       PerWeek1 = str_c("$", PerWeek1)
+       PerWeek1 = str_c(PerWeek1,"M")
+
+
+       Data = list(Name = Name1, BoxOffice = BoxOffice1, PerWeek = PerWeek1)
+       Data
+ }
> boxoffice()$Name
 [1] "Raya and the Last Dragon"    "Tom and Jerry"
 [3] "Chaos Walking"               "Boogie"
 [5] "The Croods: A New Age"       "The Little Things"
 [7] "Wonder Woman 1984"           "The Marksman"
 [9] "Judas and the Black Messiah" "Monster Hunter"
> boxoffice()$BoxOffice
 [1] "$8.5M"  "$23.0M" "$3.8M"  "$1.2M"  "$53.6M" "$13.7M" "$44.4M" "$13.0M" "$4.5M"
[10] "$14.4M"
> boxoffice()$PerWeek
 [1] "$8.5M"              "$11.5M"             "$3.8M"              "$1.2M"
 [5] "$3.57333333333333M" "$2.28333333333333M" "$4.03636363636364M" "$1.625M"
 [9] "$1.125M"            "$1.2M"
> |
```