STAT 240 - Assignment 6

Problem 2

```r
# vector of website links
# stat 330 & 350 course outlines are not available for the current term
courses = c('https://www.sfu.ca/outlines.html?2017/spring/evsc/100/d100',
            'https://www.sfu.ca/outlines.html?2018/fall/stat/452/d100',
            'https://www.sfu.ca/outlines.html?2022/spring/stat/100/d100',
            'https://www.sfu.ca/outlines.html?2022/spring/stat/201/d900',
            'https://www.sfu.ca/outlines.html?2022/spring/stat/203/d100',
            'https://www.sfu.ca/outlines.html?2022/spring/stat/270/d100',
            'https://www.sfu.ca/outlines.html?2021/fall/stat/330/d100',
            'https://www.sfu.ca/outlines.html?2021/fall/stat/350/d100')
# scraper
# vectors of cols
course_number = vector(mode="character", length=length(courses))
course_title = vector(mode="character", length=length(courses))
course_instructor = vector(mode="character", length=length(courses))
course_time_location = vector(mode="character", length=length(courses))
class_number = vector(mode="character", length=length(courses))
class_delivery = vector(mode="character", length=length(courses))
# for-loop through vector of links
for(i in 1:length(courses)) {
  course_page = readLines(courses[i])
  # extract course
  cnum =  grep('<h1 id="name"',course_page,value=TRUE)
  cnum = gsub("<[^<>]+>", " ", cnum)
  cnum = gsub("\\s{2,}", " ", cnum)
  course_number[i] = cnum
  # extract course title
  ctitle = grep('<h2 id="title"',course_page)
  ctitle = course_page[ctitle + 1]
  ctitle = gsub("\\s{2,}", " ", ctitle)
  course_title[i] = ctitle
  # extract course instructor
  cins = grep('<h4>Instructor',course_page)
  cins = course_page[cins + 1]
  cins = gsub("<[^<>]+>", " ", cins)
  cins = gsub("\\s{2,}", " ", cins)
  course_instructor[i] = cins
  # extract course time & location
  ctime_l = grep('<h4>Course Times',course_page)
  ctime_l = course_page[ctime_l + 1]
  ctime_l = gsub("<[^<>]+>", " ", ctime_l)
  ctime_l = gsub("\\s{2,}", " ", ctime_l)
  ctime_l = gsub("&ndash;", "-", ctime_l)
```

```r
  course_time_location[i] = ctime_l
  # extract class number and delivery
  heading3 = grep("<h3",course_page,value=TRUE)
  heading3 = gsub("<[^<>]+>", " ", heading3)
  heading3 = gsub("\\s{2,}", " ", heading3)
  class_number[i] = heading3[1]
  class_delivery[i] = heading3[2]
}
# data-frame from vectors
result_df = data.frame(courseNumber=course_number, courseTitle=course_title,
                       courseInstructor=course_instructor,
                       courseTimeLocation=course_time_location,
                       classNumber=class_number, classDelivery=class_delivery)
result_df
```

```
##                    courseNumber
## 1  Spring 2017 - EVSC 100 D100
## 2     Fall 2018 - STAT 452 D100
## 3  Spring 2022 - STAT 100 D100
## 4  Spring 2022 - STAT 201 D900
## 5  Spring 2022 - STAT 203 D100
## 6  Spring 2022 - STAT 270 D100
## 7     Fall 2021 - STAT 330 D100
## 8     Fall 2021 - STAT 350 D100
##                                      courseTitle    courseInstructor
## 1              Introduction to Environmental Science  Marnie Branfireun
## 2                 Statistical Learning and Prediction        Brad McNeney
## 3                          Chance and Data Analysis  Richard Lockhart
## 4                 Statistics for the Life Sciences              Wei Lin
## 5  Introduction to Statistics for the Social Sciences    Gamage Perera
## 6          Introduction to Probability and Statistics      Derek Bingham
## 7             Introduction to Mathematical Statistics  Liangliang Wang
## 8                 Linear Models in Applied Statistics    Payman Nickchi
##                              courseTimeLocation         classNumber
## 1            Fr 2:30 PM - 4:20 PM SUR 5240, Surrey   Class Number: 8909
## 2          Mo 9:30 AM - 10:20 AM SSCK 9500, Burnaby   Class Number: 4648
## 3           Mo 2:30 PM - 4:20 PM SSCC 9001, Burnaby   Class Number: 6694
## 4           Mo 12:30 PM - 1:20 PM SRYC 2600, Surrey   Class Number: 6704
## 5         Mo 10:30 AM - 12:20 PM SSCC 9002, Burnaby   Class Number: 6697
## 6  Mo, We, Fr 9:30 AM - 10:20 AM WMC 3520, Burnaby   Class Number: 6688
## 7         Mo 10:30 AM - 12:20 PM SSCC 9002, Burnaby   Class Number: 5042
## 8           Mo 8:30 AM - 9:20 AM WMC 3210, Burnaby   Class Number: 5062
##               classDelivery
## 1  Delivery Method: In Person
## 2  Delivery Method: In Person
## 3  Delivery Method: In Person
## 4  Delivery Method: In Person
## 5  Delivery Method: In Person
## 6  Delivery Method: In Person
## 7  Delivery Method: In Person
## 8  Delivery Method: In Person
```