STAT 240 - Assignment 6

Problem 1

```
course_url = "https://www.sfu.ca/outlines.html?2022/spring/stat/240/d100"
course_page = readLines(course_url)
```

1

```
# extract all h3 heading
(heading3 = grep("<h3",course_page,value=TRUE))
```

```
## [1] "                       <h3 id=\"class-number\">Class Number: 6708</h3>"
## [2] "                       <h3 id=\"delivery-method\">Delivery Method: In Person</h3>"
```

```
# remove html formatting pieces and replace them white space
(heading3a = gsub("<[^<>]+>", " ", heading3))
```

```
## [1] "                        Class Number: 6708 "
## [2] "                        Delivery Method: In Person "
```

```
# replace extra white spaces
gsub("\\s{2,}", " ", heading3a)
```

```
## [1] " Class Number: 6708 "        " Delivery Method: In Person "
```

2

```
# extract course number
cnum =  grep('<h1 id="name"',course_page,value=TRUE)
cnum = gsub("<[^<>]+>", " ", cnum)
cnum = gsub("\\s{2,}", " ", cnum)
cnum
```

```
## [1] " Spring 2022 - STAT 240 D100 "
```

3

```r
# extract course title
ctitle = grep('<h2 id="title"',course_page)
ctitle = course_page[ctitle + 1]
ctitle = gsub("\\s{2,}", " ", ctitle)
ctitle
```

```
## [1] " Introduction to Data Science"
```

4

```r
# extract instructor
cins = grep('<h4>Instructor',course_page)
cins = course_page[cins + 1]
cins = gsub("<[^<>]+>", " ", cins)
cins = gsub("\\s{2,}", " ", cins)
cins
```

```
## [1] " Lloyd Elliott "
```

5

```r
# 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)
ctime_l
```

```
## [1] " Mo 12:30 PM - 2:20 PM AQ 3182, Burnaby "
```