Stats 380 a4 scho397

Sooyong Choi

19 October 2020

Question 1

```r
# file address is "C:/Users/rick9/Desktop/Stats 380/A4/summer.html"
processHTML = function(fileaddress) {
  data = readLines(fileaddress)
  first_extract = gregexpr("\\([a-zA-Z]+\\)</span>.?</th><td>[0-9]+</td><td>[
0-9]+</td><td>[0-9]+</td><td>[0-9]+", data)
length(unlist(first_extract)[which(unlist(first_extract) != -1)])
index = which(unlist(first_extract) != -1)
indexedData = data[index]
pattern = gregexpr("\\([a-zA-Z]+\\)</span>.?</th><td>[0-9]+</td><td>[0-9]+</t
d><td>[0-9]+</td><td>[0-9]+", indexedData)
extractedData = regmatches(indexedData, pattern)
# country label
country = gsub("[0-9]|<.+>", "", unlist(extractedData))
# medals
number = gregexpr("[0-9]+", unlist(extractedData))
medal_list = regmatches(unlist(extractedData), number)
medal = unlist(medal_list)
allData = data.frame(V1 = country, V2 = medal[seq(1, length(medal), 4)], V3 =
medal[seq(2, length(medal), 4)], V4 = medal[seq(3, length(medal), 4)], V5 = m
edal[seq(4, length(medal), 4)])
return(allData)

}

head(allData)
```

```
> processHTML("C:/Users/rick9/Desktop/Stats 380/A4/summer.html")
        V1 V2 V3 V4   V5
1   (USA) 46 37 38 121
2   (GBR) 27 23 17   67
3   (CHN) 26 18 26   70
4   (RUS) 19 17 20   56
5   (GER) 17 10 15   42
6   (JPN) 12  8 21   41
7   (FRA) 10 18 14   42
8   (KOR)  9  3  9   21
9   (ITA)  8 12  8   28
10  (AUS)  8 11 10   29
11  (NED)  8  7  4   19
12  (HUN)  8  3  4   15
13  (BRA)  7  6  6   19
14  (ESP)  7  4  6   17
15  (KEN)  6  6  1   13
```

```
> head(allData)
       V1 V2 V3 V4   V5
1 (USA) 46 37 38 121
2 (GBR) 27 23 17   67
3 (CHN) 26 18 26   70
4 (RUS) 19 17 20   56
5 (GER) 17 10 15   42
6 (JPN) 12  8 21   41
```

Question 2a

```
cell = function(n){
  counter = 0
  matrix = matrix(0, nrow = 3, ncol = 3)

  # checking whether the numbers in the matrix is 0 to 8 distinctively
  while (!all(sort(as.numeric(matrix)) == 0:8)) {
    matrix = matrix(0, nrow = 3, ncol = 3)
    row = sample(1:3, n, replace = TRUE)
    col = sample(1:3, n, replace = TRUE)

    for (i in 1:n){
    matrix[row[i], col[i]] = matrix[row[i], col[i]] + 1

    if (row[i] == 1){
      matrix[row[i]+1, col[i]] = matrix[row[i]+1, col[i]] + 1
    }
```

```
    if (row[i] == 2){
      matrix[row[i]-1, col[i]] = matrix[row[i]-1, col[i]] + 1
      matrix[row[i]+1, col[i]] = matrix[row[i]+1, col[i]] + 1
    }
    if (row[i] == 3){
      matrix[row[i]-1, col[i]] = matrix[row[i]-1, col[i]] + 1
    }
    if (col[i] == 1){
      matrix[row[i], col[i]+1] = matrix[row[i], col[i]+1] + 1
    }
    if (col[i] == 2){
      matrix[row[i], col[i]+1] = matrix[row[i], col[i]+1] + 1
      matrix[row[i], col[i]-1] = matrix[row[i], col[i]-1] + 1
    }
    if (col[i] == 3){
     matrix[row[i], col[i]-1] = matrix[row[i], col[i]-1] + 1
    }


  }
    counter = counter + 1
  }
print(counter)
matrix
}
```

```
> cell(9)
[1] 671
      [,1] [,2] [,3]
[1,]    3    7    0
[2,]    8    5    4
[3,]    2    6    1

> cell(10)
[1] 3175
      [,1] [,2] [,3]
[1,]    7    6    2
[2,]    8    5    1
[3,]    4    3    0
```

Question 2b

The minimum value of m that I end up with answer appears to be 9.