Class 6: R functions Lab

Patricia Chen A16138722

4/25/23

Task: Create a new RStudio project for class 6 session. Answer Questions 1-5 (Note that Q4 is optional but Q5 is not).

Questions:

> Q1. Write a function grade() to determine an overall grade from a vector of student homework assignment scores dropping the lowest single score. If a student misses a homework (i.e. has an NA value) this can be used as a score to be potentially dropped. Your final function should be adquately explained with code comments and be able to work on an example class gradebook such as this one in CSV format: "https://tinyurl.com/gradeinput" [3pts]

Answer:

```r
# Example input vectors
student1 <- c(100, 100, 100, 100, 100, 100, 100, 90)
student2 <- c(100, NA, 90, 90, 90, 90, 97, 80)
student3 <- c(90, NA, NA, NA, NA, NA, NA, NA)
```

```r
# Straightforward mean()
mean(student1)
```

```
[1] 98.75
```

First identify the lowest score in order to drop it.

```r
min(student1)
```

```
[1] 90
```

```
# Identify the lowest element of the vector.
which.min(student1)
```

```
[1] 8
```

Exclude the lowest score from mean() calculation.

```
# Returns all except for the sixth element of the vector
student1[-6]
```

```
[1] 100 100 100 100 100 100  90
```

Function which.mean() return all other elements of the vector.

```
# First working snippet
mean(student1[-which.min(student1)])
```

```
[1] 100
```

To find out if this will work for other students Approach 1: use na.rm=TRUE argument for mean, but it would not be a good approach, since the yielded results are not equal

```
# student2 <- c(100, NA, 90, 90, 90, 90, 97, 80)
mean(student2[-which.min(student2)])
```

```
[1] NA
```

Approach 2: mask all NA values with zero. 1. Find the NA elements of the vector. To find the NA elements use:

```
x <- student2

which(is.na(x))
```

```
[1] 2
```

2. Mask the identified NA elements, and replace them with zeros:

```
x[which(is.na(x))] <- 0
```

3. Drop the lowest score:

```
mean(x[-which.min(x)])
```

```
[1] 91
```

Our working snippet now:

```
# student3 <- c(90, NA, NA, NA, NA, NA, NA, NA)
x <- student3
x[which(is.na(x))] <- 0
mean(x[-which.min(x)])
```

```
[1] 12.85714
```

Turn working snippet into a function

3 main parts of a function: 1. Function name, which is `grade()` 2. Function input arguments, student scores vector 3. Function body, the working snippet

Using RStudio, select `Code -> Extract Function`

```
grade <- function(x) {
  x[which(is.na(x))] <- 0
  mean(x[-which.min(x)])
}
```

```
grade(student1)
```

```
[1] 100
```

```
grade(student2)
```

```
[1] 91
```

```r
grade(student3)
```

```
[1] 12.85714
```

Comments to explain how to use this function.

```r
#' The average scores for a vector of student homework assignment scores is calculated aft
#' Missing values (NA) treated as 0
#' @param x A numeric vector of homework scores
#' @return Average score
#' @export
#' @examples
#' student <- c(100, NA, 90, 97)
#' grade(student)
#'
grade <- function(x) {
  # NA replaced with 0
  x[which(is.na(x))] <- 0
  # Lowest score excluded from mean
  mean(x[-which.min(x)])
}
```

Function on our whole class data from this CSV format: "https://tinyurl.com/gradeinput"

```r
url <- "https://tinyurl.com/gradeinput"
gradebook <- read.csv(url, row.names=1)
```

```r
apply(gradebook, 1, grade)
```

```
 student-1  student-2  student-3  student-4  student-5  student-6  student-7 
     91.75      82.50      84.25      84.25      88.25      89.00      94.00 
 student-8  student-9 student-10 student-11 student-12 student-13 student-14 
     93.75      87.75      79.00      86.00      91.75      92.25      87.75 
student-15 student-16 student-17 student-18 student-19 student-20 
     78.75      89.50      88.00      94.50      82.75      82.75 
```

> Q2. Using your grade() function and the supplied gradebook, Who is the top scoring student overall in the gradebook? [3pts]

Run the apply() function

```r
results <- apply(gradebook, 1, grade)
sort(results, decreasing = TRUE)
```

```
student-18  student-7  student-8 student-13  student-1 student-12 student-16 
     94.50      94.00      93.75      92.25      91.75      91.75      89.50 
 student-6  student-5 student-17  student-9 student-14 student-11  student-3 
     89.00      88.25      88.00      87.75      87.75      86.00      84.25 
 student-4 student-19 student-20  student-2 student-10 student-15 
     84.25      82.75      82.75      82.50      79.00      78.75 
```

```r
which.max(results)
```

```
student-18 
        18 
```

Answer: The top scoring student overall is Student 18.

> Q3. From your analysis of the gradebook, which homework was toughest on students (i.e. obtained the lowest scores overall? [2pts]

```r
ave.scores <- apply(gradebook, 2, mean, na.rm=TRUE)

which.min(ave.scores)
```

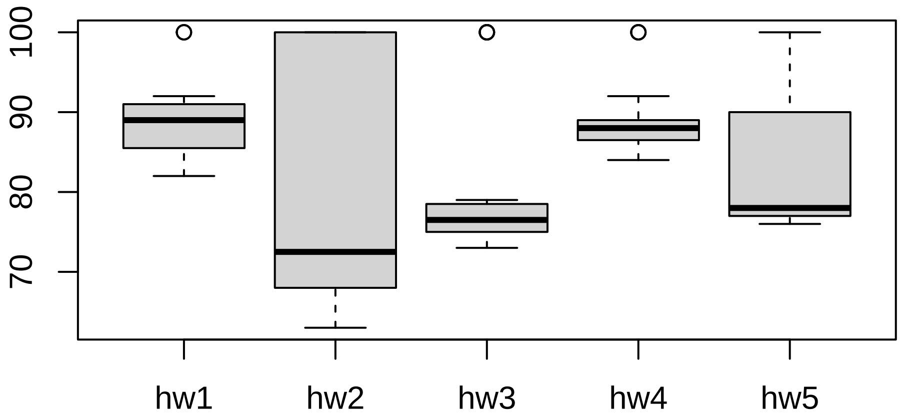
```
hw3 
  3 
```

```r
med.scores <- apply(gradebook, 2, median, na.rm=TRUE)

which.min(med.scores)
```

```
hw2 
  2 
```

```r
boxplot(gradebook)
```



Answer: According to our yielded bar plot and data, the toughest homework on the students is HW2.

> Q4. Optional Extension: From your analysis of the gradebook, which homework was most predictive of overall score (i.e. highest correlation with average grade score)? [1pt]

```r
masked.gradebook <- gradebook
masked.gradebook[ is.na(masked.gradebook) ] <- 0
masked.gradebook
```

```
           hw1 hw2 hw3 hw4 hw5
student-1  100  73 100  88  79
student-2   85  64  78  89  78
student-3   83  69  77 100  77
student-4   88   0  73 100  76
student-5   88 100  75  86  79
student-6   89  78 100  89  77
student-7   89 100  74  87 100
student-8   89 100  76  86 100
student-9   86 100  77  88  77
student-10  89  72  79   0  76
```

```
student-11  82  66  78  84 100
student-12 100  70  75  92 100
student-13  89 100  76 100  80
student-14  85 100  77  89  76
student-15  85  65  76  89   0
student-16  92 100  74  89  77
student-17  88  63 100  86  78
student-18  91   0 100  87 100
student-19  91  68  75  86  79
student-20  91  68  76  88  76
```

```
cor(results, masked.gradebook$hw5)
```

```
[1] 0.6325982
```

```
apply(masked.gradebook, 2, cor, x=results)
```

```
      hw1       hw2       hw3       hw4       hw5 
0.4250204 0.1767780 0.3042561 0.3810884 0.6325982 
```

Answer: According to the analysis of the gradebook, the homework most predictive of the overall score is Hw5.

> Q5. Make sure you save your Quarto document and can click the “Render” (or Rmark- down”Knit”) button to generate a PDF foramt report without errors. Finally, submit your PDF to gradescope. [1pt]