class06BGGN213

Phoebe LI

2/7/2022

Here we will write a function to grade some student homework We will start with a more simple example

```r
# Example input vectors to start with
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
student1
```

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

The regular average will be returned by mean function.

```r
mean(student1)
```

```
## [1] 98.75
```

To find the position of the smallest value

```r
which.min(student1)
```

```
## [1] 8
```

so this will give us the min value

```r
# The first solution
mean(student1[-which.min(student1)])
```

```
## [1] 100
```

How about NA? it is a bad idea to use na.rm=T

```r
# For students who have NA as their score
# Change NA=0
x<- student2
x
```

```
## [1] 100  NA  90  90  90  90  97  80
```

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

```
## [1] 100   0  90  90  90  90  97  80
```

```r
mean(x)
```

```
## [1] 79.625
```

Find the average for student 3

```r
x<- student3
x[is.na(x)]<-0
mean(x[-which.min(x)])
```

```
## [1] 12.85714
```

Let's make a function!! REMEBER TO EXTACT THE FUNCTION

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

Let's grade students 1,2,3

```r
grade(student1)
```

```
## [1] 100
```

```r
grade(student2)
```

```
## [1] 91
```

```r
grade(student3)
```

```
## [1] 12.85714
```

We can grade the Whole class now

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.

```r
#' Map missing HW, NA value to 0
#' We exclude the lowest one
#'
#' @param x Numeric vector of home
#' @return Average score
#' @example
#'    student<-c(100, NA, 90, 80)
#'    grade(student)
#'
grade <- function(x) {
    #map missing HW, NA value to 0
    x[is.na(x)]<-0
    # We exclude the lowest one
    mean(x[-which.min(x)])
}
```

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]

```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]

```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
```

Or you can find the one has highest score

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

```
## student-18
##         18
```

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

```r
# Let's try average
hw.ave<-apply(gradebook, 2, mean, na.rm=TRUE)
which.min(hw.ave)
```
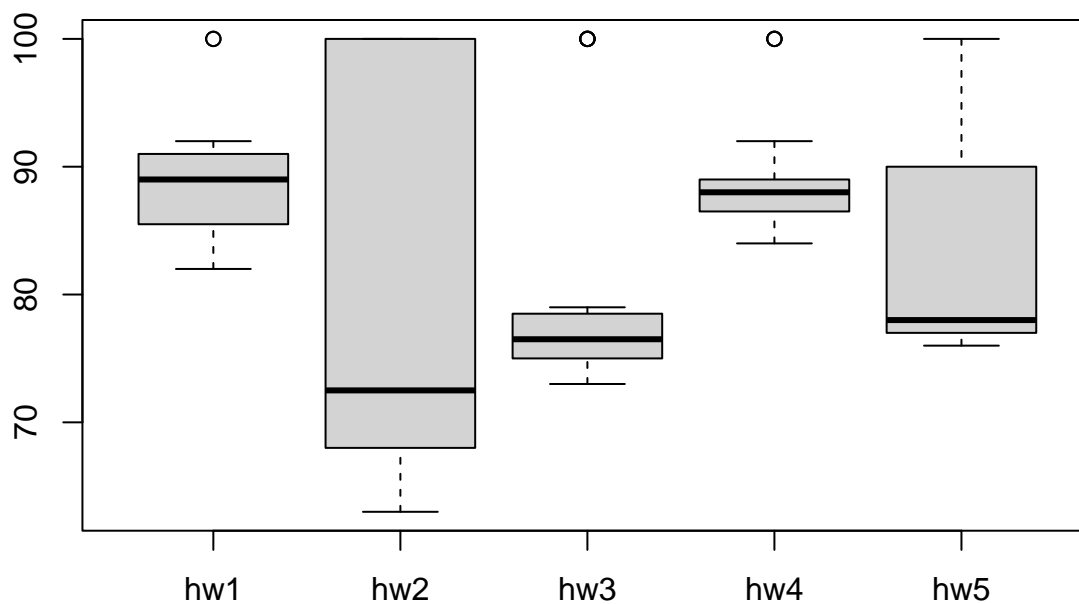
```
## hw3
##   3
```

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

```
## hw2
##   2
```

So, which one I should use??? Let's plot it!!!

```r
boxplot(gradebook)
```


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]

```
gradebook[ is.na(gradebook)] <-0
cor(results, gradebook$hw2)
```

```
## [1] 0.176778
```

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

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

Q5. Make sure you save your Rmarkdown document and can click the "Knit" button to generate a PDF foramt report without errors. Finally, submit your PDF to gradescope. [1pt]