

# LAB\_06

RUNQI ZHANG

## Table of contents

### 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)
```

“student\_homework” import

```
student_homework = read.csv("C:\\Users\\zhang\\OneDrive\\ \\BIMM 143 - Bioinformatics Lab\\
View(student_homework)
nrow(student_homework)
```

```
[1] 20
```

```
ncol(student_homework)
```

```
[1] 6
```

## Practice

package import

```
library(FSA)
```

```
## FSA v0.9.3. See citation('FSA') if used in publication.
## Run fishR() for related website and fishR('IFAR') for related book.
```

```
library(tidyverse)
```

```
-- Attaching packages ----- tidyverse 1.3.2 --
```

```
v ggplot2 3.3.6      v purrr  0.3.5
v tibble  3.1.8      v dplyr  1.0.10
v tidyr   1.2.1      v stringr 1.4.1
v readr   2.1.3      v forcats 0.5.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
```

```
library(ggplot2)
library("gridExtra")
```

Attaching package: 'gridExtra'

The following object is masked from 'package:dplyr':

combine

replace NA with 0

```
student_homework[is.na(student_homework)] <- 0
```

calculate the sum and min for each student

```
student_min <- apply(student_homework[,-1], 1, min)
student_min
```

```
[1] 73 64 69  0 75 77 74 76 77  0 66 70 76 76  0 74 63  0 68 68
```

```
student_sum <- apply(student_homework[,-1], 1, sum)
student_sum
```

```
[1] 440 394 406 337 428 433 450 451 428 316 410 437 445 427 315 432 415 378 399
[20] 399
```

## subtract the total score by their lowest score

```
adjusted_sum <- student_sum - student_min
adjusted_sum
```

```
[1] 367 330 337 337 353 356 376 375 351 316 344 367 369 351 315 358 352 378 331
[20] 331
```

#Average = total/ #assignment one col contains student name, one col is disposed, #remaining scores = ncol-2

```
student_mean <- adjusted_sum/(ncol(student_homework)-2)
student_mean
```

```
[1] 91.75 82.50 84.25 84.25 88.25 89.00 94.00 93.75 87.75 79.00 86.00 91.75
[13] 92.25 87.75 78.75 89.50 88.00 94.50 82.75 82.75
```

#transpose col to row and append it to the original csv document

```
student_mean <- t(t(student_mean))
nrow(student_mean)
```

```
[1] 20
```

```
student_final <- cbind(student_homework, student_mean)
```

## Q1:

**#integrate the codes above into a R function named grade()** 1. takes into a parameter, namely the gradebook 2. replace NA with 0 3. calculate the total and min for each student 4. calculate the adjusted score, subtract total by min 5. calculate the final score (average), divide adjusted total by n-2

```
grade <- function(grade_raw) {  
  grade_raw[is.na(grade_raw)] <- 0 #replace NA with numerical 0  
  
  #student's sum subtracted by min, and transposed into rows  
  grade_adjusted = t( t( apply(grade_raw[,-1], 1, sum) - apply(grade_raw[,-1], 1, min) )  
  
  grade_mean = grade_adjusted / (ncol(grade_raw)-2) #calculate the mean  
  grade_final = cbind(grade_raw, grade_mean) #integrate the mean into a final document  
  return(grade_final)  
}
```

#test run

```
test_doc= read.csv("C:\\Users\\zhang\\OneDrive\\ \\BIMM 143 - Bioinformatics Lab\\Lab_06\\  
grade(test_doc)
```

	X	hw1	hw2	hw3	hw4	hw5	grade_mean
1	student-1	100	73	100	88	79	91.75
2	student-2	85	64	78	89	78	82.50
3	student-3	83	69	77	100	77	84.25
4	student-4	88	0	73	100	76	84.25
5	student-5	88	100	75	86	79	88.25
6	student-6	89	78	100	89	77	89.00
7	student-7	89	100	74	87	100	94.00
8	student-8	89	100	76	86	100	93.75
9	student-9	86	100	77	88	77	87.75
10	student-10	89	72	79	0	76	79.00
11	student-11	82	66	78	84	100	86.00
12	student-12	100	70	75	92	100	91.75
13	student-13	89	100	76	100	80	92.25
14	student-14	85	100	77	89	76	87.75
15	student-15	85	65	76	89	0	78.75
16	student-16	92	100	74	89	77	89.50
17	student-17	88	63	100	86	78	88.00
18	student-18	91	0	100	87	100	94.50

```
19 student-19 91 68 75 86 79      82.75
20 student-20 91 68 76 88 76      82.75
```

**Q2: Using your grade() function and the supplied gradebook, Who is the top scoring student**

overall in the gradebook?

**A2:** student-18 is the top scoring student with an final score of 94.5

```
max(grade(student_homework)$grade_mean)
```

```
[1] 94.5
```

```
which.max(grade(student_homework)$grade_mean)
```

```
[1] 18
```

**Q3: From your analysis of the gradebook, which homework was toughest on students (i.e. obtained the lowest scores overall?)**

**A3:** homework#2 is the toughest because it has the lowest average and the lowest median

**Q4: From your analysis of the gradebook, which homework was most predictive of overall score (i.e. highest correlation with average grade score)?**

**A4:** HW5 was most predictive of final score because it shows the highest correlation.

```
student_final = grade(student_homework)
cor(student_final$grade_mean, student_final[2:6])
```