

# Class06: R Functions

Morgan Farrell

2/3/2022

Here we will write a function to grade student homework. We will start with a more simple input example- a vector of student homework scores

```
# 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)
```

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 adequately 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]

```
student1
```

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

The regular average

```
mean(student1)
```

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

To find the smallest value use 'which.min' it gives you the position

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

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

Use the 'which.min' function inside of 'student1' dataset gives us the value of the lowest score

```
student1[which.min(student1)]
```

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

To get everything but the min value

```
student1[-which.min(student1)]
```

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

Then take the mean of those results

```
mean(student1[-which.min(student1)])
```

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

This gives the average of all the scores minus the lowest value, but this is with a good student with all scores

```
student2
```

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

```
mean(student2[-which.min(student2)])
```

```
## [1] NA
```

Doesn't work because of that NA value, so we need to eliminate that to calculate the mean

This removes the NA, but there is no penalty for missing a homework assignment. Keep in mind student 3 as well.

```
mean(student2, na.rm = TRUE)
```

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

```
mean(student3, na.rm = TRUE)
```

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

Use the 'is.na' function, returns a logical vector

```
is.na(student2)
```

```
## [1] FALSE TRUE FALSE FALSE FALSE FALSE FALSE
```

We can use 'is.na' to replace the NA with 0

```
x <- student2  
x
```

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

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

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

Combine the remove NA function with the finding the mean minus the lowest score into one function. Can use the Code-> Extract Function->Name then need to execute the function for it to be in the memory.

```
x <- student3
grade <- function(x) {
  x[is.na(x)] <- 0
  mean(x[-which.min(x)])
}
```

```
grade(student1)
```

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

```
grade(student2)
```

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

```
grade(student3)
```

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

Your final function should be adequately 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]

```
## Calculate average score for vector of homework
## dropping the lowest homework score. Missing homework are counted as zeros.
## @param x Numeric vector of homework scores
##
## @return Average score
## @export
##
## @examples
## student <- c(100, NA, 90, 80)
## grade(student)
grade <- function(x) {
  #Map missing homework (NA) to 0
  x[is.na(x)] <- 0
  #Exclude the lowest homework score
  mean(x[-which.min(x)])
}
```

Now we can **grade the whole class**. First import the dataset aka gradebook

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

We have the answer for question 1

```
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]

First method is try sort, a bit lengthy

```
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
```

More direct to use 'which.max'

```
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]

```
#Try mean() first
hw.ave<-apply(gradebook, 2, mean, na.rm=TRUE)
which.min(hw.ave)
```

```
## hw3
##    3
```

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

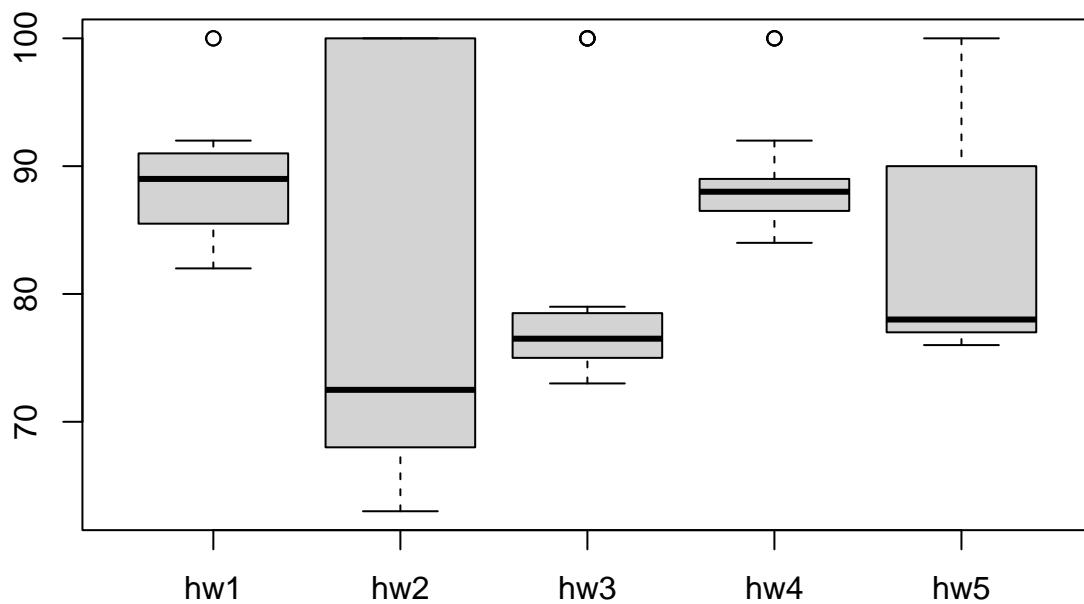
```
## hw2
##    2
```

```
#More comprehensive summary of the scores for each HW
apply(gradebook, 2, summary, na.rm=TRUE)
```

```
## $hw1
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      82.00  85.75   89.00   89.00   91.00   100.00
##
## $hw2
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      63.00  68.00   72.50   80.89   100.00   100.00         2
##
## $hw3
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      73.00  75.00   76.50   80.80   78.25   100.00
##
## $hw4
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      84.00  86.50   88.00   89.63   89.00   100.00         1
##
## $hw5
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      76.00  77.00   78.00   83.42   90.00   100.00         1
```

Visualize these results, making it easier to interpret

```
boxplot(gradebook)
```



Looks like homework 2 was the toughest for students, since it showed the lowest overall median and largest spread amongst students.

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$hw5)
```

```
## [1] 0.6325982
```

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

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

Homework 5 was the best predictor of performance overall.