

class06

Katelyn Brown

2/6/2022

Class 6 R Functions

```
# 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 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>”

```
# Find average score of student1; straightforward mean
mean(student1)
```

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

```
# Find lowest score of student1 and what element this lowest score is in
min(student1)
```

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

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

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

Want to drop the lowest grade from the average overall grade.

```
# Exclude 8th element from student1
student1[-8]
```

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

Need to make the lowest grade more universal for other students.

```
# Exclude lowest grade from student vector
student1[-which.min(student1)]
```

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

```
# Find mean again
mean(student1[-which.min(student1)])
```

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

Need to apply mean to other students, with NA grades.

```
# Find mean of student2
mean(student2)
```

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

```
# Exclude NA from mean of student 2
mean(student2, na.rm = TRUE)
```

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

```
# Find mean of student3
mean(student3, na.rm = TRUE)
```

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

Now I need to convert NA to zero for fairness. First, I need to find where the NA values are.

```
x <- student2
which(is.na(x))
```

```
## [1] 2
```

```
which(is.na(student3))
```

```
## [1] 2 3 4 5 6 7 8
```

Now that I know what values are NA, I need to convert them to zero.

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

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

```
# Find mean of student2 now that NA is zero
mean(x)
```

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

```
# Need to exclude lowest score from mean
mean(x[-which.min(x)])
```

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

Now that it worked for student2, I am going to try to do it again for student3.

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

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

Now I can make the overall function.

Use what I have done so far to answer Q1.

```
# Name of the function is grade()
# Input of the function will be a vector
# Body of the function will be the working piece of the student grades to find
# the overall grade

#' Calculate the average score for a vector of student scores, dropping lowest
#' grade. Missing values are treated as zeroes.
#'
#' @param x A numeric vector of homework scores
#'
#' @return Average score
#' @export
#'
#' @examples
#' student <- c(100, NA, 90, 97)
#' grade(student)
#'
grade <- function(x) {
  # Treat missing values as 0
  x[is.na(x)] <- 0
  # Exclude lowest score from mean
  mean(x[-which.min(x)])
}
# Ran the chunk of code to save the grade() function to environment
```

Test the function for the students.

```
grade(student1)
```

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

```
grade(student2)
```

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

```
grade(student3)
```

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

The function works for the example students. Now, I need to apply the function to the “real” class data from the CSV format: “<https://tinyurl.com/gradeinput>”

```
# Upload gradebook  
url <- "https://tinyurl.com/gradeinput"  
gradebook <- read.csv(url, row.names = 1)  
# Ran chunk of code to save gradebook
```

```
# Apply the grade function to the whole class  
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?

```
# Use the apply function to find answer  
results <- apply(gradebook, 1, grade)  
results
```

```
## 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
```

```
# Find maximum from results  
which.max(results)
```

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

Student 18 had the highest score overall in the class.

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

```
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 NA 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 NA 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 NA
## student-16 92 100 74 89 77
## student-17 88 63 100 86 78
## student-18 91 NA 100 87 100
## student-19 91 68 75 86 79
## student-20 91 68 76 88 76
```

Find averages of the different homework assignments.

```
# Find mean of each homework assignment
ave.scores <- apply(gradebook, 2, mean, na.rm = TRUE)
ave.scores
```

```
##          hw1          hw2          hw3          hw4          hw5
## 89.00000 80.88889 80.80000 89.63158 83.42105
```

```
which.min(ave.scores)
```

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

```
# Find median of each homework assignment
med.scores <- apply(gradebook, 2, median, na.rm = TRUE)
med.scores
```

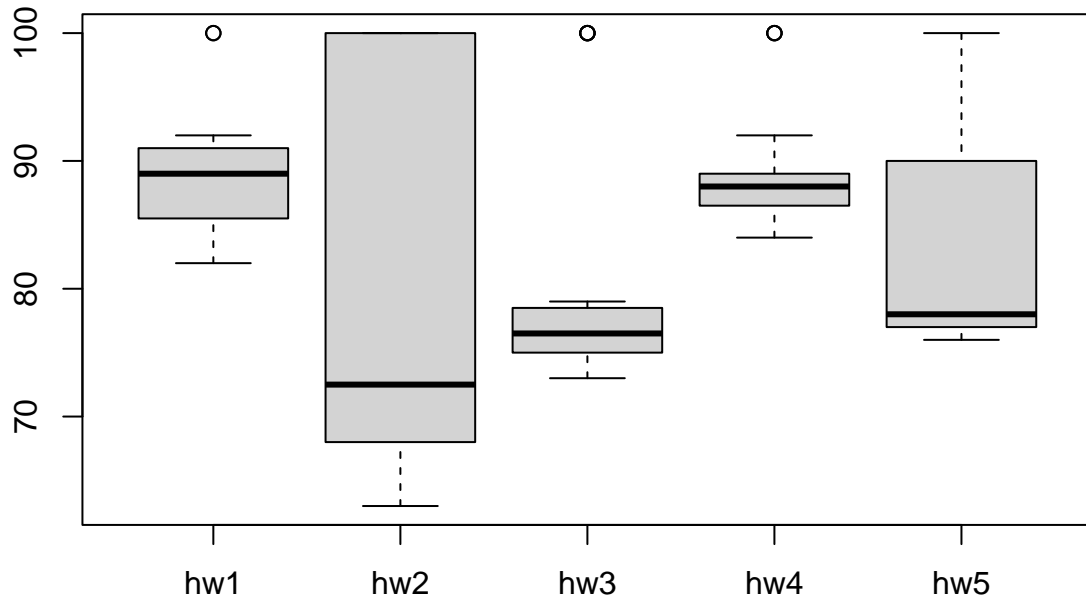
```
## hw1 hw2 hw3 hw4 hw5
## 89.0 72.5 76.5 88.0 78.0
```

```
which.min(med.scores)
```

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

Need to look at boxplot.

```
boxplot(gradebook)
```



Based on the average, homework 3 was toughest on students. However, based on the median and boxplot, it is clear that homework 2 was the toughest on students and had the greatest variation in scores.

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)?

Need to find if the final results, or average score for each student, are correlated with the results of the homework assignments, or the gradebook columns.

```
# Need to mask the gradebook for fairness
masked.gradebook <- gradebook
masked.gradebook[is.na(masked.gradebook)] <- 0
# Test masked.gradebook to make sure it was correct
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
```

```
# Use correlation function to look at correlation between grade and homework
cor(results, masked.gradebook$hw1)
```

```
## [1] 0.4250204
```

```
# Apply correlation function over each column (homework)
apply(masked.gradebook, 2, cor, x = results)
```

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

Homework 5 has the greatest correlation to the overall scores of students, because its correlation value is the highest and closest to 1.