

lab06

Ebony Michelle Argaez (PID: A96056556)

Grades of sample students

```
#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>”

```
student1
```

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

Determining Average using ‘`mean()`’.

```
mean(student1)
```

```
[1] 98.75
```

Determining minimum score of student 1

```
#to determine the minimum value
min(student1)
```

```
[1] 90
```

to determine position of minimum value

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

```
[1] 8
```

getting the value of 8th position

```
student1[8]
```

```
[1] 90
```

getting minimum value

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

```
[1] 90
```

to remove a specific position, in this case position 8, the min val

```
student1[-8]
```

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

to get everything but min value

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

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

now taking the mean of student 1

```
#first sln  
mean(student1[-which.min(student1)])
```

```
[1] 100
```

Now doing this for student2 & student 3

```
student2
```

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

```
#student 2 has NA value
```

```
#this would work for student 2 but not student 3  
mean(student2, na.rm=TRUE)
```

```
[1] 91
```

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

```
[1] 90
```

```
#this student only submitted 1 assignment
```

Need to convert all NA's to 0, using 'na.rm=True' is unfair.

```
#converting NA of student 2  
student2= replace(student2, is.na(student2), 0)  
student2
```

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

```
#converting NA of student 2  
student3
```

```
[1] 90 NA NA NA NA NA NA NA
```

```
s2=student2  
s2[is.na(s2)] = 0  
s2
```

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

```
#finding mean of student 2
mean(s2[-which.min(s2)])
```

```
[1] 91
```

```
#converting NA of student 3
student3
```

```
[1] 90 NA NA NA NA NA NA NA
```

```
s3=student3
s3[is.na(s3)] = 0
s3
```

```
[1] 90 0 0 0 0 0 0 0
```

```
mean(s3[-which.min(s3)])
```

```
[1] 12.85714
```

Create function to determine mean and removing lowest score

```
#' Calculate average score for vector of HW scores
#' dropping the lowest score
#' Missing values (NA) will be treated as zero
#' @param numeric vector of hw scores
#'
#' @return average score
#' @export
#'
#' @examples
#' student <- c(100, NA, 90, 80)
#' grade(student)
grade <- function(x) {
  x[is.na(x)] = 0 #converts NA/missing hw to 0
  mean(x[-which.min(x)]) #determines lowest value and excludes it
}
```

Checking if grade function works with example students

```
grade(student1)
```

```
[1] 100
```

```
grade(student2)
```

```
[1] 91
```

```
grade(student3)
```

```
[1] 12.85714
```

```
#it works ^_^
```

Apply function to example class gradebook:

```
url <- "https://tinyurl.com/gradeinput"
gradebook <- read.csv(url, row.names=1) #starts with hw column not student ID column

#View(gradebook)

results <- apply(gradebook, MARGIN=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	

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

```
#gives me student with highest score
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?)

Calculate summary stat for each HW or column

```
#determining average of each hw

avg_hw <- apply(gradebook, 2, mean, na.rm=TRUE)
avg_hw
```

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

```
#determining hw with lowest mean
which.min(avg_hw)
```

```
hw3
3
```

```
#determining median of each hw

median_hw <- apply(gradebook, 2, median, na.rm=TRUE)
median_hw
```

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

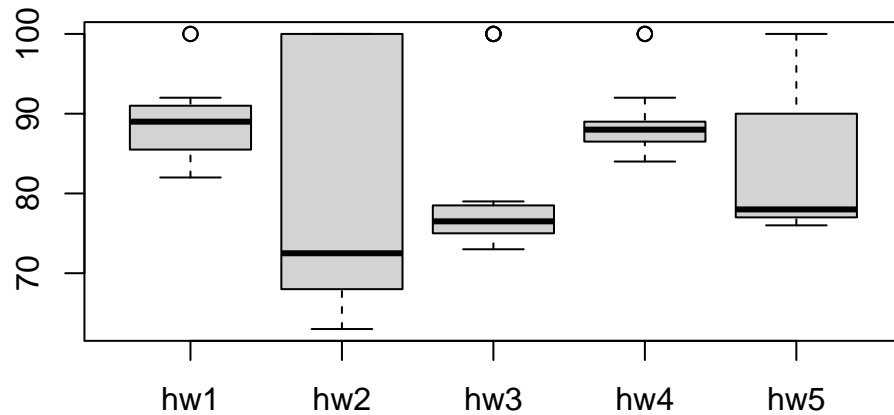
```
#determining hw with lowest median
which.min(median_hw)
```

```
hw2
2
```

When determining mean and median, you get different results.

Plotting data to see differences.

```
boxplot(gradebook)
```



Based on the plot, hw 1 and 4 most people scored high based on the mean; students ranged in scores for hw 2, hw 3 the average score was low; and hw 5 there was a

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

```
cor(results, gradebook$hw1)
```

```
[1] 0.4250204
```

```
cor(results, gradebook$hw2)
```

```
[1] NA
```

```
cor(results, gradebook$hw3)
```

```
[1] 0.3042561
```

```
cor(results, gradebook$hw4)
```

```
[1] NA
```

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

```
[1] NA
```

I get NAs.

```
#changing NAs
gradebook[is.na(gradebook)] <-0

#checking to see if this works for all hw's
cor(results, gradebook$hw1)
```

```
[1] 0.4250204
```

```
cor(results, gradebook$hw2)
```

```
[1] 0.176778
```

```
cor(results, gradebook$hw3)
```

```
[1] 0.3042561
```

```
cor(results, gradebook$hw4)
```

```
[1] 0.3810884
```

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

```
[1] 0.6325982
```



```
#using apply to check cor of all hw's  
apply(gradebook, 2, cor, x=results)
```

hw1	hw2	hw3	hw4	hw5
0.4250204	0.1767780	0.3042561	0.3810884	0.6325982

HW 2 has a poor correlation.