

Class 6: R Functions

Ayşe

2/4/2022

Learning about functions in 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)
```

Below is one way we can exclude the NA values when calculating mean

```
mean(na.omit(student2))
```

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

Below is a way to find the number of NA values in the vector student3

```
sum(is.na(student3))
```

```
## [1] 7
```

```
mean(student1)
```

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

Below, we are assigning the value 0 to NA values in the vector student2 - this is changing the vector itself. Then, we can find the mean of student2

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

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

Repeating for student 3

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

```
## [1] 11.25
```

In order to drop the lowest grade, we need to delete the minimum value in a vector, then get the mean. Let's try with student2

```
#student2 <-sort(student2, decreasing = FALSE)
#student2 <- student2[-1]
```

We can use this in the grade() function after we turn all NA values to 0.

```
grade <- function(x) {
  #make NA values into 0s
  x[is.na(x)]=0
  #sort grades from lowest to highest
  x <- sort(x, decreasing = FALSE)
  #remove the lowest (first) grade
  x <- x[-1]
  #calculate the mean of the remaining scores
  mean(x)
}
grade(student1)
```

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

Does the function work for student 2 and 3?

```
grade(student2)
```

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

```
grade(student3)
```

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

Test on a new student

```
student4 <- c(100, 100, NA, 100, 100)
grade(student4)
```

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

Alternatively, this is what was done in-class:

```
grade1 <- function(x){
  #find position of smallest value in the vector and drop that value
  x[is.na(x)]=0
  mean(x[-which.min(x)])
}
grade1(student1)
```

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

```
grade1(student2)
```

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

```
grade1(student3)
```

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

Import gradebook: I downloaded the file from this link <https://tinyurl.com/gradeinput>. The first column was made the row name.

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

Now, we can apply our function to the students in the gradebook Q2. Using your grade() function and the supplied gradebook, Who is the top scoring student overall in the gradebook?

```
student_scores <- apply(gradebook, MARGIN = 1, grade)
#top scoring student
which.max(student_scores)
```

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

Q3. From your analysis of the gradebook, which homework was toughest on students (i.e. obtained the lowest scores overall? For this, I used apply (by column) to convert NAs to 0s or just explude them, then found the mean of each homework, then found the index of the minimum value.

```
#if we make NAs excluded
hw_means_no_NAs <- apply(gradebook, 2, mean, na.rm=T)

which.min(hw_means_no_NAs)
```

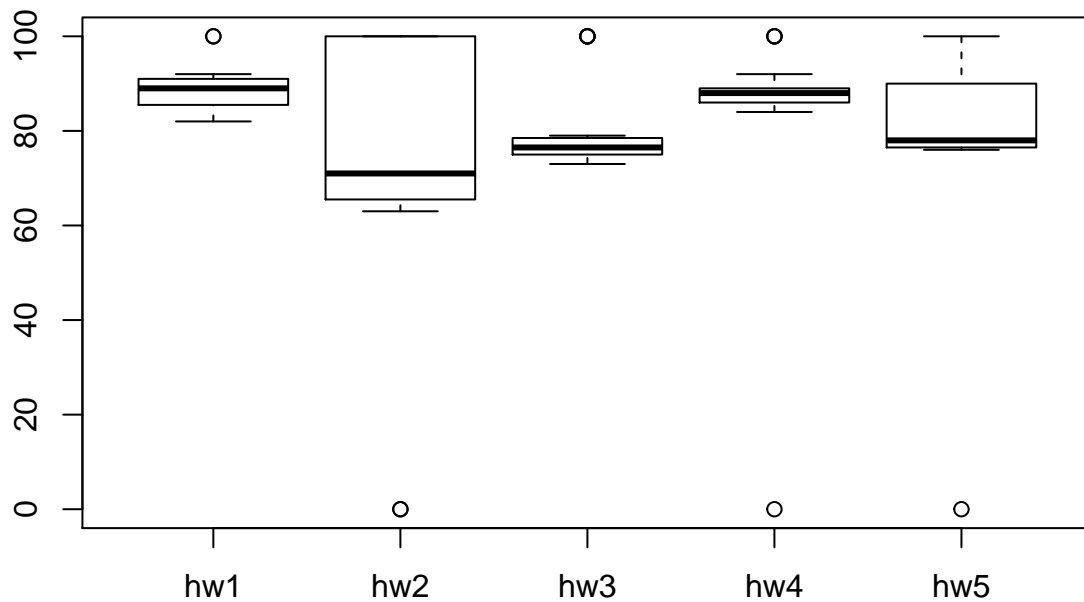
```
## hw3
##    3
```

```
#make NAs into 0s
gradebook[is.na(gradebook)]=0
#find mean grade for each hw assignment
hw_means <- apply(gradebook, MARGIN=2, mean)
#find which assignment had the lowest average score
which.min(hw_means)
```

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

```
#hw 2 was toughest on the students
```

```
boxplot(gradebook)
```



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

```
#install correlation package install.packages("ggpubr"), library("ggpubr")  
cor_hw1<-cor(gradebook[,1], student_scores, method = c("pearson", "kendall", "spearman"))  
print(cor_hw1)
```

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

apply to all HWs

```
cor_hw2<-cor(gradebook[,2], student_scores, method = c("pearson", "kendall", "spearman"))  
print(cor_hw2)
```

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

```
cor_hw3<-cor(gradebook[,3], student_scores, method = c("pearson", "kendall", "spearman"))  
print(cor_hw3)
```

```
## [1] 0.3042561
```

```
cor_hw4<-cor(gradebook[,4], student_scores, method = c("pearson", "kendall", "spearman"))  
print(cor_hw4)
```

```
## [1] 0.3810884
```

```
cor_hw5<-cor(gradebook[,5], student_scores, method = c("pearson", "kendall", "spearman"))  
print(cor_hw5)
```

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

```
#HW 5 was most prective of student scores.
```

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