Class 06: R Functions

Sanket Garg (A59026686)

All about functions in R

Every function in R has atleast 3 thingsL - name (you pick it) - arguments (the input(s) to your function) - the body.

Today we will write a function to grade a class of student assignment scores (e.g. homework, etc).

First, I will work with a simplified vector input where I know what the answers should be.

```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)
```

Finding the average of student 1.

```r
mean(student1)
```

```
[1] 98.75
```

How can we drop the lowest score? I can use the min() function to find the lowest score (element in the vector) and the which.min() function to find the position of that lowest score.

```r
min(student1)
```

```
[1] 90
```

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

```
[1] 8
```

We can remove the lowest score using the '-' symbol.

```
student1
```

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

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

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

Let's put the use of 'which.min()', minus indexing and 'mean()' together to solve this baby!

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

```
[1] 100
```

Will this work for student2?

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

```
[1] NA
```

Using Variables for simplifying the analysis.

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

```
[1] NA
```

'na.rm=TRUE' is not optimal because it drops all the na values before performng the computation.

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

```
[1] 90
```

We can "mask" the NAs and change them to be 0. Rational here is that if you don't do the hw, then you should get zero points.

We can use the 'is.na()' to find where the missing homeworks are in the homework vector.

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

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

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

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

Let's put all the pieces together.

```
x <- student3
# Mask NA to 0
x[is.na(x)] <- 0
# Find the mean dropping the lowest score
mean(x[-which.min(x)])
```

```
[1] 12.85714
```

Turn this snippet into a function.

```
grade <- function(x){
  # this is where the body code lives

  # Mask NA to 0
  x[is.na(x)] <- 0
  # Find the mean dropping the lowest score
  mean(x[-which.min(x)])

}
```

We can use this function to grade any student.

```
grade(student3)
```

[1] 12.85714

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 adquately 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]

I need to read the gradebook csv file.

```
#row.name=1 changes column1 as the rownames
gradebook <- read.csv("https://tinyurl.com/gradeinput", row.names = 1)
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

A very useful function that Barry is forcing us to use is the 'apply()' function. How do we use it to take the 'grade()' function and apply it over the full gradebook?

```
ans <- apply(gradebook, 1, grade)
```

Q2. Using your grade() function and the supplied gradebook, Who is the top scoring student overall in the gradebook? [3pts]

```
which.max(ans)
```

```
student-18
       18
```

Q3. From your analysis of the gradebook, which homework was toughest on students (i.e. obtained the lowest scores overall? [2pts] Let's mask the na values to zero.

```
mask <- gradebook
mask[is.na(mask)] <- 0
hws <- apply(mask, 2, mean)
hws
```

```
  hw1   hw2   hw3   hw4   hw5
89.00 72.80 80.80 85.15 79.25
```

```
which.min(hws)
```

```
hw2
  2
```

Q4. From your analysis of the gradebook, which homework was most predictive of overall score (i.e. highest correlation with average grade score)? [1pt]

```
cor_values <- cor(ans, mask)
cor_values
```

```
           hw1       hw2       hw3       hw4       hw5
[1,] 0.4250204 0.176778 0.3042561 0.3810884 0.6325982
```

```
which.max(cor_values)
```

```
[1] 5
```

Now take the 'apply()' function and the 'cor()' function and run over the whole gradebook.

```
apply(mask,2,cor,y=ans)
```

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

```
which.max(apply(mask,2,cor,y=ans))
```

```
hw5
  5
```