class06 R Functions

Mina Wu (A59013200)

2024-01-26

Our first simple function

All functions in R have 3 parts. They have:

- a name
- input arguments(none, one or more)
- a body

A function to add two numbers

```
sillyadd <- function(x,y=1){
  x+y
}
```

Let me try out this function

```
sillyadd(100)
```

```
[1] 101
```

Let's do something more useful

Paste the student's grade in to the chuck.

```
# 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)
```

I will begin by getting a skateboard solution to my tesla problem.

```r
mean(student1)
```

```
[1] 98.75
```

which.min gives the position of the minimum value in the vector.

```r
min(student1)
```

```
[1] 90
```

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

```
[1] 8
```

```r
x <- student3
x
```

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

```r
# Find lowest value
ind <- which.min(x)
ind
```

```
[1] 1
```

```r
#Exclude lowest value and find mean and replace Na with 0.
x[is.na(x)] <- 0
mean (x[-ind], na.rm= TRUE)
```

```
[1] 0
```

Find and replace the NA values with zero

```r
x<- 1:5
x
```

```
[1] 1 2 3 4 5
```

```r
x[x==3]<- 10000
x
```

```
[1]     1     2 10000     4     5
```

```r
x<- student2
x
```

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

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

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

Turn it into function called grade

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

Test out the grade function with student1,2, or 3.

```r
grade(student1)
```

```
[1] 100
```

Read a class gradebook CSV file from here: "https://tinyurl.com/gradeinput"

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

Now use our `grade()` function to grade the whole class... We can "apply" our new `grade()` function over whether the rows or the columns of the gradebook, with MARGIN=1 or MAR-GIN=2. Use MARGIN1, which is for the rows.

```
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
```

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

```
#which student have the highest score
which.max(results)
```

```
student-18
        18
```

```
# what is the highest score
max(results)
```

```
[1] 94.5
```

Q3. From your analysis of the gradebook, which homework was toughest on students (i.e. obtained the lowest scores overall? [2pts].

```
apply(gradebook,2, grade)
```

```
    hw1      hw2      hw3      hw4      hw5
89.36842 76.63158 81.21053 89.63158 83.42105
```

Use apply, but it gives NA values

```
apply(gradebook, 2, mean)
```

```
 hw1  hw2  hw3  hw4  hw5
89.0   NA 80.8   NA   NA
```

Rewrite the grade function to drop the lowest score.

```
grade <- function(x, drop.lowest= TRUE){
  x[is.na(x)] <- 0

  if (drop.lowest){
    ans <- mean (x[-which.min(x)])
  }
  else{
    ans<- mean(x)
  }
    ans
}
```

Remove Na using na.rm = TRUE. HW3 is the toughest assignment.

```
toughest_assignment<- apply(gradebook, 2, mean, na.rm=T)
which.min(toughest_assignment)
```
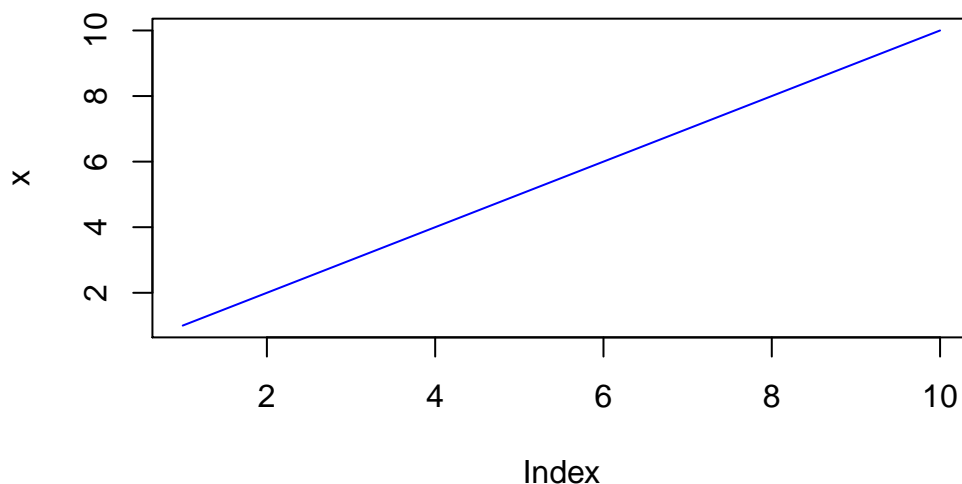
```
hw3
  3
```

```
#HW3 is the toughest assignment.
```

Side question: about ... and what it does in a function

```
plotme<- function(x,...){
  plot(x,...)
}
plotme (1:10, col= "blue", type= "l")
```


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]

mask the Na score to 0

```
mask <- gradebook
mask[is.na(mask)] <- 0
mask
```

```
          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
```

give us correlation on hw3 and hw5

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

[1] 0.6325982

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

[1] 0.3042561

Use apply to look for the correlation of each hw.

```
correlation <- apply(mask, 2, cor,y=results)
correlation
```

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

```
which.max(correlation)
```

hw5
  5

#HW5 is the predictive of overall score