R Functions Lab

Mady Welch

Calculating grades with the lowest score dropped

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 can start by using the 'mean()' function to calculate an average.

```
mean(student1)
```

```
[1] 98.75
```

Find the lowest score in the vector using 'min()'

```
min(student1)
```

```
[1] 90
```

You can also use 'which.min()' to find the position in the vector the lowest score is in

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

```
[1] 8
```

Drop the lowest score from the vector

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

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

Use 'mean()' to calculate the new average grade

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

```
[1] 100
```

- This calculation of the mean does not work for student2 or student3 because of NA values in the vectors.

If we set na.rm to TRUE, it will drop any NA values in the vector

```
mean(student2[-which.min(student2)], na.rm = TRUE)
```

```
[1] 92.83333
```

- This function drops both the lowest score **and** the NA value.

What about student3?

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

```
[1] 90
```

- This function drops **every** NA value, so the average is only calculated using one score.

To fix this problem, we need to set the NA values equal to 0 before calculating the average.

```
student2[is.na(student2)] <- 0
student3[is.na(student3)] <- 0
```

Here are the **new** student2 and student3 score vectors:

```
student2
```

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

```
student3
```

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

We can now use the previous function used for student1 to calculate the grades for student2 and student3:

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

```
[1] 91
```

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

```
[1] 12.85714
```

This function is a little messy, let's clean it up a bit

```
x <- student1
x[is.na(x)] <- 0
mean(x[-which.min(x)])
```

```
[1] 100
```

Now we can calculate the average score for any student 'x' !

Now turn it into a function:

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

```
grade(student1)
```

```
[1] 100
```

```
grade(student2)
```

```
[1] 91
```

```
grade(student3)
```

[1] 12.85714

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

```
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
```

Use 'apply()' to use the gradebook data - 'apply(x, margin, function)'

```
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
```

- Now we have the average grades for all of the students

Save this answer as "results"

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

Now we can find the highest scoring student using 'which.max()'

```
which.max(results)
```

```
student-18
        18
```

But what was their score?

```
results[which.max(results)]
```

```
student-18
      94.5
```

Student 18 has the highest score

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

When we wanted to see the student scores, we looked at rows. But now we can look at the columns to see overall homework scores.

```
hwscores <- apply(gradebook, 2, sum, na.rm = TRUE)
hwscores
```

```
 hw1  hw2  hw3  hw4  hw5
1780 1456 1616 1703 1585
```

We can use 'which.min()' to see the lowest homework score:

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

```
 hw2
1456
```

Homework 2 was toughest on students

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

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

cor(mask[, 1], results)
```

[1] 0.4250204

That gave us one correlation, but how can we get all of them?

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

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

The highest correlation = the most predictive score

```
predict[which.max(predict)]
```

```
      hw5
0.6325982
```

Homework 5 was the most predictive of overall score