R Functions Lab Class 6

Nicholas

Writing Functions

Initiate sample student grades

```
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)
```

Create a grade() function.

```
grade <- function(n, na.rm = TRUE) {
  n <- sort(n, decreasing = T, na.last = T) #sort the vector from greatest to least
  n <- n[-length(n)] #remove the last element
  n[is.na(n)] <- 0 #this will set any NA values to 0 before calculating mean
  mean(n, na.rm = na.rm) #calculate mean ignoring NA
}
```

testing stuff

```
sort(student2)
```

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

```
sort(student2, decreasing = F, na.last = T)
```

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

Different ways to do the same thing exist

Instead of sorting we can use which.min to find the index of the lowest value and go vector[-which.mean(vector)]to remove the element at that vector.

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

Import grade book

```
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
```

Q2 Who is the top scoring student?

```
all_students <- apply(gradebook, 1, grade_class)
all_students[which.max(all_students)]
```

```
student-18
      94.5
```

Q3 What was the toughest homework?

```
homeworks <- apply(gradebook, 2, grade)
homeworks[which.min(homeworks)]
```

```
     hw2
76.63158
```

```r
#can also be done using sums
which.min(apply(gradebook, 2, sum, na.rm=TRUE))
```

```
hw2
  2
```

Q4 Which homework was most predictive of overall grade?

```r
#compare all student grades to homework grades to see which on is the least different
mask <- gradebook
mask[is.na(mask)] <- 0

cor(mask$hw5, all_students) #one column at a time
```

```
[1] 0.6325982
```

```r
#find the highest correlation out of all the homeworks
deviation <- apply(mask, 2, cor, y = all_students)
deviation
```

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

```r
deviation[which.max(deviation)]
```

```
      hw5
0.6325982
```