

# Class06

Adam Bisharat

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 adequately 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]

```
# import data table (im doing this first for organizational purposes)
url <- "https://tinyurl.com/gradeinput"
gradebook <- read.csv(url)

# Select only the numeric columns of this data set
numeric_cols <- sapply(gradebook, is.numeric)
numeric_gradebook <- gradebook[, numeric_cols]

# Create the an appropriate grading formula which suits our needs and given data
grade <- function(x) {
  # Replace NA values with 0
  x[is.na(x)] <- 0

  # Drop the lowest score and calculate the mean of the remaining scores
  mean(x[-which.min(x)])
}

# Apply the grade function to each row of the numeric data and print the output
overall_grades <- apply(numeric_gradebook, 1, grade)

print(overall_grades)
```

```
[1] 91.75 82.50 84.25 84.25 88.25 89.00 94.00 93.75 87.75 79.00 86.00 91.75
[13] 92.25 87.75 78.75 89.50 88.00 94.50 82.75 82.75
```

```
#Once this was done I manually calculated a few averages to ensure everything worked properly
```

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

```
#Just going to use max finder functions to pull this data  
which.max(overall_grades)
```

```
[1] 18
```

```
max(overall_grades)
```

```
[1] 94.5
```

The top scoring student is student #18 with a grade of 94.50%

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

```
# Calculating the average score for each homework by taking the means of the columns  
homework_averages <- colMeans(numeric_gradebook)  
  
# Making NAs 0 and finding the homework with the lowest average score along with the names of  
numeric_gradebook[is.na(numeric_gradebook)] <- 0  
lowest_average <- min(homework_averages)  
toughest_homework <- names(which.min(homework_averages))  
  
print(lowest_average)
```

```
[1] NA
```

```
print(toughest_homework)
```

```
[1] "hw3"
```

Homework 2 is the hardest homework with an average score of 72.8

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]

```
# Calculating the correlation between each homework and the overall grades using the cor function
homework_correlations <- sapply(numeric_gradebook, function(hw_scores) {
  cor(hw_scores, overall_grades)
})

# Finding the homework with the highest correlation and its name (homework number)
max_corr <- max(homework_correlations)
most_predictive_hw <- names(which.max(homework_correlations))

print(max_corr)
```

```
[1] 0.6325982
```

```
print(most_predictive_hw)
```

```
[1] "hw5"
```

Q5. Make sure you save your Quarto document and can click the “Render” (or Rmarkdown”Knit”) button to generate a PDF format report without errors. Finally, submit your PDF to gradescope. [1pt]