STAT 545: HW 1

Introduction

All code is also included in working .R files corresponding to each problem. Changes and new code are included in this pdf knit of my notebook for easy view.

Problem 1

We simply want to add out final score to our file. Thankfully, the calculations and final data frame is already made. All we need to do is bind final letter grade with our score file and write it into a csv file. This is the edited lines of code in "grade_P1.R".

```
#bind the final score and letter grade together
finalscore = cbind(score, grades)
finalscore

# We just add the write csv code line
write.csv(finalscore, "final_grades.csv", row.names = FALSE)
```

And here is our score and finalscore:

```
score
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_maxscore_100
## 1      Alice                  38                45              85
## 2        Bob                  47                42              90
## 3    Charlie                   0                40              92
## 4       Dave                  42                37              75
## 5       Erin                  40                48              98
##   Project_maxscore_100
## 1                   90
## 2                   95
## 3                    0
## 4                   92
## 5                   92
```

```
finalscore
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_maxscore_100
## 1      Alice                  38                45              85
## 2        Bob                  47                42              90
## 3    Charlie                   0                40              92
## 4       Dave                  42                37              75
```

```
## 5       Erin                  40                48              98
##   Project_maxscore_100 grades
## 1                   90      B
## 2                   95      A
## 3                    0      F
## 4                   92      B
## 5                   92      A
```

Problem 2

First, we have to read in our new file "scores_new.csv". Now we add up all 5 homework and change the component-wise score calculations to make it all equal.

So here, all we did was:

```
#Calculate HW total
score$HW_maxscore_100 = score$HW_1 + score$HW_2 + score$HW_3 + score$HW_4 + score$HW_5

#Calculate component-wise scores according to each equal
hw.component = (score$HW_maxscore_100/100)*25
```

Check our work:

```
score
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_1 HW_2 HW_3 HW_4 HW_5
## 1      Alice                  38                45   15   20   20   20   10
## 2        Bob                  47                42   20   20   20   15   15
## 3    Charlie                   0                40   20   16   16   20   20
## 4       Dave                  42                37    0   20   20   20   15
## 5       Erin                  40                48   20   20   20   20   18
##   Project_maxscore_100 HW_maxscore_100
## 1                   90              85
## 2                   95              90
## 3                    0              92
## 4                   92              75
## 5                   92              98
```

```
finalscore
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_1 HW_2 HW_3 HW_4 HW_5
## 1      Alice                  38                45   15   20   20   20   10
## 2        Bob                  47                42   20   20   20   15   15
## 3    Charlie                   0                40   20   16   16   20   20
## 4       Dave                  42                37    0   20   20   20   15
## 5       Erin                  40                48   20   20   20   20   18
##   Project_maxscore_100 HW_maxscore_100 grades
## 1                   90              85      B
## 2                   95              90      A
## 3                    0              92      F
## 4                   92              75      B
## 5                   92              98      A
```

Problem 3

This is R file of Problem 3. We choose to ignore NA for median score, instead of using 0 like we have earlier. It makes more sense to use median rather than setting to 0 when talking about medians so that it is not skewed by 0 values.

```
# Clear the workspace
rm(list=ls())

# Reading a data frame from an Excel file
score=read.csv("scores_new.csv", header =T)

# Simply use apply and median with na.rm=TRUE to find our homework medians
# First select the columns we want to use
hw_cols <- c("HW_1", "HW_2", "HW_3", "HW_4", "HW_5")
hw_medians = apply(score[, hw_cols], 2, median, na.rm = TRUE)
print(hw_medians)
```

```
## HW_1 HW_2 HW_3 HW_4 HW_5
##   20   20   20   20   15
```

Ok, next let us calculate Alice's median homework score, ignoring NA.

```
# Let us only look at Alice's row
alice_row = score[score$First.Name == "Alice", ]
# Calculate median (using same hw columns as before)
alice_median = median(as.numeric(alice_row[, hw_cols]), na.rm = TRUE)
print(alice_median)
```

```
## [1] 20
```

Alice's median homework score is 20.

Problem 4

We do a very similar process to previously, but instead of doing it for each homework (columns), we apply to each student (row).

```
# Clear the workspace
rm(list=ls())

# Reading a data frame from an Excel file
score=read.csv("scores_new.csv", header =T)

# Simply use apply and median with na.rm=TRUE to find our student medians
# First select the columns we want to use
hw_cols <- c("HW_1", "HW_2", "HW_3", "HW_4", "HW_5")
student_medians = apply(score[, hw_cols], 1, median, na.rm = TRUE)

# Add scores to data using column bind
```

```r
score_with_medians = cbind(score, Median_HW = student_medians)

# Now we sort using order
score_with_medians_final <- score_with_medians[order(score_with_medians$Median_HW, decreasing = TRUE), ]
```

Let us check our work:

```r
print(score_with_medians)
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_1 HW_2 HW_3 HW_4 HW_5
## 1      Alice                  38                45   15   20   20   20   10
## 2        Bob                  47                42   20   20   20   15   15
## 3    Charlie                  NA                40   20   16   16   20   20
## 4       Dave                  42                37   NA   20   20   20   15
## 5       Erin                  40                48   20   20   20   20   18
##   Project_maxscore_100 Median_HW
## 1                   90        20
## 2                   95        20
## 3                   NA        20
## 4                   92        20
## 5                   92        20
```

```r
print(score_with_medians_final)
```

```
##   First.Name Midterm_maxscore_50 Final_maxscore_50 HW_1 HW_2 HW_3 HW_4 HW_5
## 1      Alice                  38                45   15   20   20   20   10
## 2        Bob                  47                42   20   20   20   15   15
## 3    Charlie                  NA                40   20   16   16   20   20
## 4       Dave                  42                37   NA   20   20   20   15
## 5       Erin                  40                48   20   20   20   20   18
##   Project_maxscore_100 Median_HW
## 1                   90        20
## 2                   95        20
## 3                   NA        20
## 4                   92        20
## 5                   92        20
```

Looks the same because they all have the same median. Code works even if this is not the case.