# DATA ANALYTICS LAB

## EXERCISE 1

S Shyam Sundaram

19BCE1560

### Q1

Read the given table. Print the number of rows and columns. Order the table based on section.

**Code**

```
data<-read.csv(file='Score.csv')
stud<-data
stud<-stud[order(stud$Section, decreasing=FALSE),]
stud
dim(stud)
```

**Output**

```
> data<-read.csv(file='Score.csv')
> #Order by section
> stud<-data
> stud<-stud[order(stud$Section, decreasing=FALSE),]
> stud
      Student Section Test1 Test2 Final
1   Capalleti       1    94    91    87
3       Engles      1    95    97    97
6   Lundsford       1    92    40    86
7       Mcbane      1    75    78    72
8   Capalleti       1    94    91    87
10      Engles      1    95    97    97
13 Lundsford        1    92    40    86
14      Mcbane      1    75    78    72
16       Rebok      1    13    34    23
17 Capalleti        1    94    91    87
19      Engles      1    95    97    97
20      Mcbane      1    75    78    72
21 Capalleti        1    94    91    87
2       Dubose      2    51    65    91
4        Grant      2    63    75    80
5     Krupski       2    80    76    71
9       Dubose      2    51    65    91
11       Grant      2    63    75    80
12    Krupski       2    80    76    71
15       Suad       2    23    32    33
18      Dubose      2    51    65    91
22      Dubose      2    51    65    91
> dim(stud)
[1] 22  5
>
```

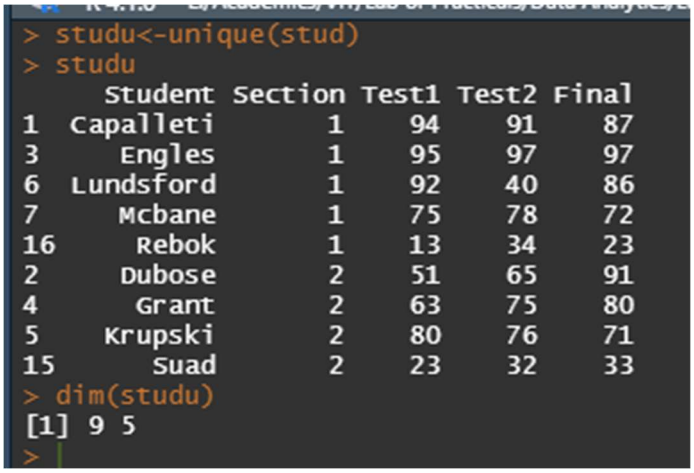
## Q2

Remove all duplicates and display the rows and columns.

**Code**

```
studu<-unique(stud)
studu
dim(studu)
```

**Output**



```
> studu<-unique(stud)
> studu
     Student Section Test1 Test2 Final
1  Capalleti       1    94    91    87
3     Engles       1    95    97    97
6  Lundsford       1    92    40    86
7     Mcbane       1    75    78    72
16     Rebok       1    13    34    23
2     Dubose       2    51    65    91
4      Grant       2    63    75    80
5    Krupski       2    80    76    71
15      Suad       2    23    32    33
> dim(studu)
[1] 9 5
>
```

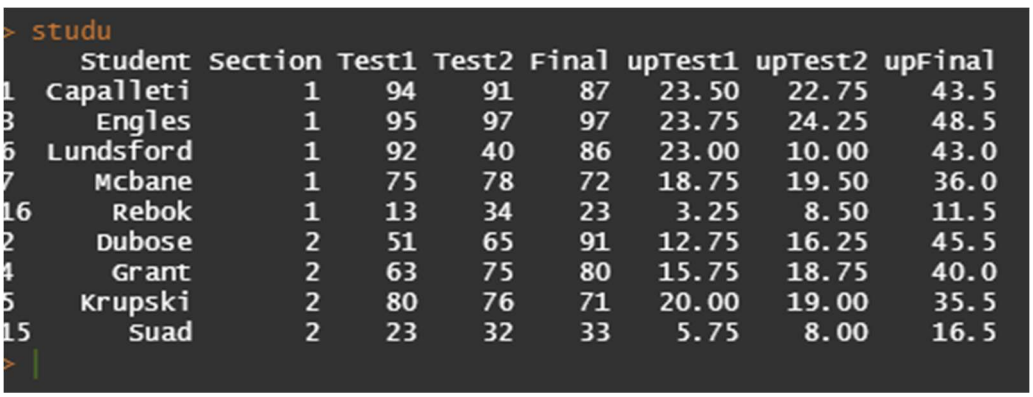
## Q3

Transform the student scores table given below and perform the following transformations. The students have been conducted two tests and a final exam. The weightage for Test 1&2 are 25 each. The weightage for the final exam is 50. Add columns for the updated test1, test2 and final exam by normalizing them according to their respective weights.

**Code**

```
studu <- transform(studu,upTest1 = (Test1/100)*25)
studu <- transform(studu,upTest2 = (Test2/100)*25)
studu <- transform(studu,upFinal = (Final/100)*50)
studu
```

**Output**



```
> studu
     Student Section Test1 Test2 Final upTest1 upTest2 upFinal
1  Capalleti       1    94    91    87   23.50   22.75    43.5
3     Engles       1    95    97    97   23.75   24.25    48.5
6  Lundsford       1    92    40    86   23.00   10.00    43.0
7     Mcbane       1    75    78    72   18.75   19.50    36.0
16     Rebok       1    13    34    23    3.25    8.50    11.5
2     Dubose       2    51    65    91   12.75   16.25    45.5
4      Grant       2    63    75    80   15.75   18.75    40.0
5    Krupski       2    80    76    71   20.00   19.00    35.5
15      Suad       2    23    32    33    5.75    8.00    16.5
>
```

## Q4

Sum the normalized columns to obtain the TotalMarks out of 100. Calculate the mean of the total marks.

**Code**

```
studu <- transform(studu,TotalMarks = upTest1+upTest2+upFinal)
studu
me<-mean(studu$TotalMarks)
me
```

**Output**

```
> studu <- transform(studu,TotalMarks = upTest1+upTest2+upFinal)
> studu
      Student Section Test1 Test2 Final upTest1 upTest2 upFinal TotalMarks
1   Capalleti       1    94    91    87   23.50   22.75    43.5      89.75
3      Engles       1    95    97    97   23.75   24.25    48.5      96.50
6   Lundsford       1    92    40    86   23.00   10.00    43.0      76.00
7      Mcbane       1    75    78    72   18.75   19.50    36.0      74.25
16      Rebok       1    13    34    23    3.25    8.50    11.5      23.25
2      Dubose       2    51    65    91   12.75   16.25    45.5      74.50
4       Grant       2    63    75    80   15.75   18.75    40.0      74.50
5     Krupski       2    80    76    71   20.00   19.00    35.5      74.50
15       Suad       2    23    32    33    5.75    8.00    16.5      30.25
> me<-mean(studu$TotalMarks)
> me
[1] 68.16667
```

## Q5

Add a new column Grade which splits the students into three categories and label them as Above Average (AA), Below Average (BA), Fail(FAIL). Note that students below 50 are considered FAIL.

**Code**

```
studu$Grades <- cut(studu$TotalMarks,breaks=c(-Inf,50,me,Inf), labels=c("FAIL","BA","AA"))
studu
```

**Output**

```
> studu$Grades <- cut(studu$TotalMarks,breaks=c(-Inf,50,me,Inf), labels=c("FAIL","BA","AA"))
> studu
      Student Section Test1 Test2 Final upTest1 upTest2 upFinal TotalMarks Grades
1   Capalleti       1    94    91    87   23.50   22.75    43.5      89.75     AA
3      Engles       1    95    97    97   23.75   24.25    48.5      96.50     AA
6   Lundsford       1    92    40    86   23.00   10.00    43.0      76.00     AA
7      Mcbane       1    75    78    72   18.75   19.50    36.0      74.25     AA
16      Rebok       1    13    34    23    3.25    8.50    11.5      23.25   FAIL
2      Dubose       2    51    65    91   12.75   16.25    45.5      74.50     AA
4       Grant       2    63    75    80   15.75   18.75    40.0      74.50     AA
5     Krupski       2    80    76    71   20.00   19.00    35.5      74.50     AA
15       Suad       2    23    32    33    5.75    8.00    16.5      30.25   FAIL
>
```