**INPUT/OUTPUT STRING MANIPULATION AND PLYR PACKAGE**

**1. Step # 1 Import assignment 6 Data-set to R Download Import assignment 6 Data-set to R. Then, Run the commend "mean" using Sex as the category (use plyr package for this operation). Last commend in this step: write the resulting output to a file.**

**> <- read.table("<FileName>.txt", header = TRUE)**

**DATA SET**

**Name, Age, sex, Grade**

**Raul,25, Male,80**

**Booker,18, Male,83**

**Lauri,21, Female,90**

**Leonie,21, Female,91**

**Sherlyn,22, Female,85**

**Mikaela,20, Female,69**

**Raphael,23, Male,91**

**Aiko,24, Female,97**

**Tiffany,21, Female,78**

**Corina,23, Female,81**

**Petronila,23, Female,98**

**Alecia,20, Female,87**

**Shemika,23, Female,97**

**Fallon,22, Female,90**

**Deloris,21, Female,67**

**Randee,23, Female,91**

**Eboni,20, Female,84**

**Delfina,19, Female,93**

**Ernestina,19, Female,93**

**Milo,19, Male,67**

**Solution:**

**STEP1:  
CODE:**

#calculating the average of grades of males and females separately

Newdata <- ddply(data, .(Sex), transform, average = mean(Grade))

Newdata

#Writing the data into a new table

write.table(Newdata, "output.txt", sep = ",")

Newdata

Next, the average grades of males and females are calculated using the "ddply" function and stored in a new table called "New data". The data is then written into a new table "output.txt" using the "write. Table" function and displayed on the console.

**OUTPUT**

>

> #calculating the average of grades of male and female separately

> Newdata <- ddply(data, .(Sex), transform, average = mean(Grade))

> Newdata

Name Age Sex Grade average

1 Lauri 21 Female 90 86.9375

2 Leonie 21 Female 91 86.9375

3 Sherlyn 22 Female 85 86.9375

4 Mikaela 20 Female 69 86.9375

5 Aiko 24 Female 97 86.9375

6 Tiffaney 21 Female 78 86.9375

7 Corina 23 Female 81 86.9375

8 Petronila 23 Female 98 86.9375

9 Alecia 20 Female 87 86.9375

10 Shemika 23 Female 97 86.9375

11 Fallon 22 Female 90 86.9375

12 Deloris 21 Female 67 86.9375

13 Randee 23 Female 91 86.9375

14 Eboni 20 Female 84 86.9375

15 Delfina 19 Female 93 86.9375

16 Ernestina 19 Female 93 86.9375

17 Raul 25 Male 80 80.2500

18 Booker 18 Male 83 80.2500

19 Raphael 23 Male 91 80.2500

20 Milo 19 Male 67 80.2500

>

> #Writing the data into new table

> write.table(Newdata, "output.txt", sep = ",")

> Newdata

Name Age Sex Grade average

1 Lauri 21 Female 90 86.9375

2 Leonie 21 Female 91 86.9375

3 Sherlyn 22 Female 85 86.9375

4 Mikaela 20 Female 69 86.9375

5 Aiko 24 Female 97 86.9375

6 Tiffaney 21 Female 78 86.9375

7 Corina 23 Female 81 86.9375

8 Petronila 23 Female 98 86.9375

9 Alecia 20 Female 87 86.9375

10 Shemika 23 Female 97 86.9375

11 Fallon 22 Female 90 86.9375

12 Deloris 21 Female 67 86.9375

13 Randee 23 Female 91 86.9375

14 Eboni 20 Female 84 86.9375

15 Delfina 19 Female 93 86.9375

16 Ernestina 19 Female 93 86.9375

17 Raul 25 Male 80 80.2500

18 Booker 18 Male 83 80.2500

19 Raphael 23 Male 91 80.2500

20 Milo 19 Male 67 80.2500

**2.** Convert the data set to a dataframe for names whose name contains the letter i, then create a new data set with those names, Write those names to a file separated by commas (CSV)  
The comment

>write.table(students\_gendered\_mean, "Students\_Gendered\_Mean")

# Filter the original data set to include only data for which the student name

# Contained the letter i.

>i\_students <- subset(students, grepl("i", students$Name, ignore.case=T))

**Solution:**

**STEP2:  
CODE:**  
install.packages("plyr")

library(table)

library(plyr)

#Selecting the file

file <- file.choose("data.txt")

data <- read.table(file, header = TRUE, sep = ",")

data$Age <- as.numeric(data$Age)

data$Grade <- as.numeric(data$Grade)

data$Sex <- as.factor(data$Sex)

The R code selects a data file "data.txt" and reads it as a table using the "read. table" function. The data is modified to convert the Age and Grade columns to numeric and the Sex column to a factor variable. The "plyr" package is loaded to perform data manipulation tasks.

**OUTPUT**

> #Selecting the file

> file <- file.choose("data.txt")

> data <- read.table(file, header = TRUE, sep = ",")

> data$Age <- as.numeric(data$Age)

> data$Grade <- as.numeric(data$Grade)

> data$Sex <- as.factor(data$Sex)

3. Write the filtered data set and convert it to CSV file

Hint(file.Choose() and subset())

**Solution:**

**STEP3:  
CODE:**#Filtering the names with letter "i"

subset\_data <- subset(Newdata, grepl("i", data$Name))

#Writing it into new csv file

write.csv(subset\_data, "subset\_names.csv",row.names=TRUE)

subset\_data

Then, the names that contain the letter "i" are filtered from the "Newdata" table using the "subset" function and stored in a new table called "subset\_data". The filtered data is then written into a new csv file called "subset\_names.csv" using the "write.csv" function and displayed on the console.

**OUTPUT**

>

> #Filtering the names with letter "i"

> subset\_data <- subset(Newdata, grepl("i", data$Name))

>

> #Writing it into new csv file

> write.csv(subset\_data, "subset\_names.csv",row.names=TRUE)

> subset\_data

Name Age Sex Grade average

3 Sherlyn 22 Female 85 86.9375

4 Mikaela 20 Female 69 86.9375

6 Tiffaney 21 Female 78 86.9375

8 Petronila 23 Female 98 86.9375

9 Alecia 20 Female 87 86.9375

10 Shemika 23 Female 97 86.9375

11 Fallon 22 Female 90 86.9375

12 Deloris 21 Female 67 86.9375

13 Randee 23 Female 91 86.9375

15 Delfina 19 Female 93 86.9375

17 Raul 25 Male 80 80.2500

18 Booker 18 Male 83 80.2500

19 Raphael 23 Male 91 80.2500

20 Milo 19 Male 67 80.2500

In summary, this R code reads data from a file, performs some data manipulation tasks, calculates the average grades for males and females, writes the modified data into a new table, filters the data based on a condition, and writes the filtered data into a new CSV file.