Import DataSet.txt, Run a mean using Sex as the category (use plyr package for this operation), Then write the resulting output to a file.

Next test the DataSet.txt dataframe for names whos name contains the letter i, then create a new data set with those names, Write those names to a file separated by comma's (CSV)

Hint(file.choose() and subset())

Step 1: Import dataset into R.

```
>x = read.table(file.choose(),header=TRUE,sep=",")
```

```
> x
         Name Age    Sex Grade
1        Raul  25   Male    80
2      Booker  18   Male    83
3       Lauri  21 Female    90
4      Leonie  21 Female    91
5     Sherlyn  22 Female    85
6     Mikaela  20 Female    69
7     Raphael  23   Male    91
8        Aiko  24 Female    97
9    Tiffaney  21 Female    78
10     Corina  23 Female    81
11 Petronila  23 Female    98
12     Alecia  20 Female    87
13    Shemika  23 Female    97
14     Fallon  22 Female    90
15    Deloris  21 Female    67
16     Randee  23 Female    91
17      Eboni  20 Female    84
18    Delfina  19 Female    93
19 Ernestina  19 Female    93
20       Milo  19   Male    67
>
```

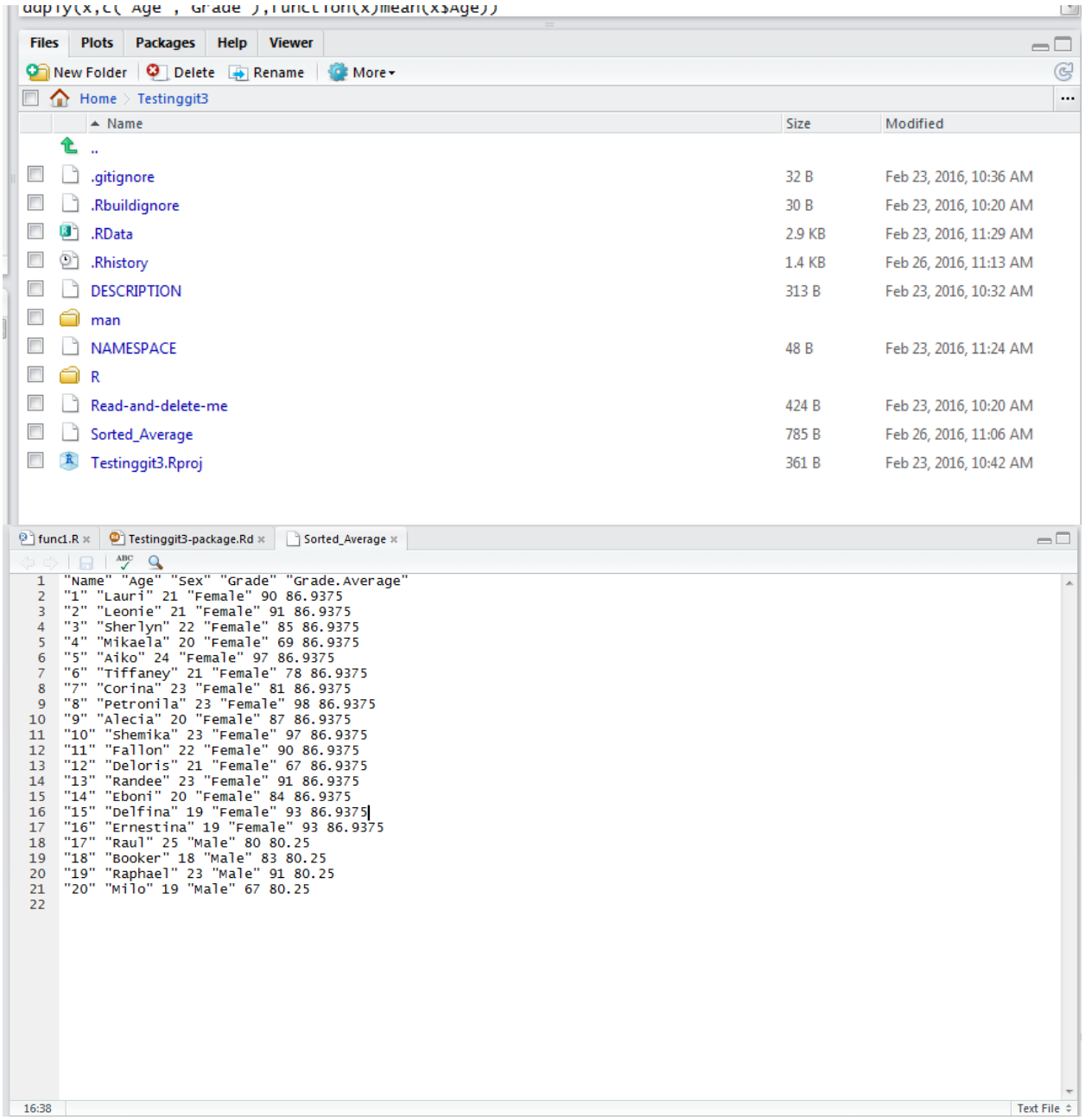
Step2: run plyer generates for the mean of both Age and Grade split by gendor

```
>install.packages("plyr")
>library(plyr)
>y = ddply(x,"Sex",transform, Grade.Average=mean(Grade))
```

```
> ddply(x,"Sex",transform, Grade.Average=mean(Grade))
         Name Age    Sex Grade 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
>
```

Step3: Print this to a file

```
>write.table(y,"Sorted_Average")
```

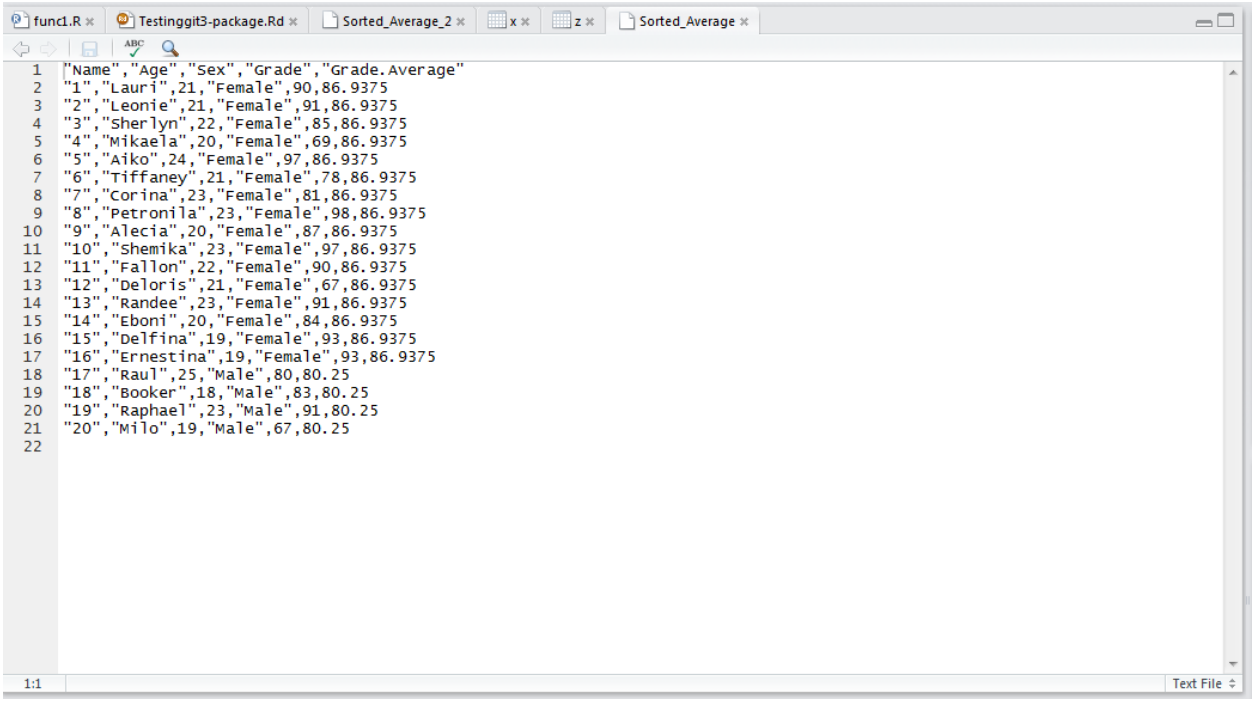


We notice that the format has “” around every value except numeric

We change the write function to generate the CSV

```
> write.table(y,"Sorted_Average",sep=",")
```

This create a CSV.



Step:4 Filter the names in the given list for names that contain the letter i.

First we must realize that the letter i is represented as both “l” and “i”

Therefore we must use the “[iI]” in grep()

In the past we have used the subset() command in R to subset a data.

x is where we stored the imported data set.

```
>newx = subset(x,grepl("[iI]",x$Name))
```

```
> subset(x ,grepl("i",x$Name))
        Name Age    Sex Grade
3      Lauri  21 Female    90
4     Leonie  21 Female    91
6    Mikaela  20 Female    69
8       Aiko  24 Female    97
9   Tiffaney  21 Female    78
10    Corina  23 Female    81
11 Petronila  23 Female    98
12    Alecia  20 Female    87
13   Shemika  23 Female    97
15   Deloris  21 Female    67
17     Eboni  20 Female    84
18   Delfina  19 Female    93
19 Ernestina  19 Female    93
20      Milo  19   Male    67
```

Then write this subset to a file

```
>write.table(newx,"DataSubset",sep=",")
```