

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	Tiffany	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")
```

The screenshot shows the RStudio interface. The top pane displays a file explorer for the 'Testinggit3' directory. The bottom pane shows a text editor with the contents of the 'Sorted_Average' file.

File Explorer (Testinggit3):

Name	Size	Modified
..		
.gitignore	32 B	Feb 23, 2016, 10:36 AM
.Rbuildignore	30 B	Feb 23, 2016, 10:20 AM
.RData	2.9 KB	Feb 23, 2016, 11:29 AM
.Rhistory	1.4 KB	Feb 26, 2016, 11:13 AM
DESCRIPTION	313 B	Feb 23, 2016, 10:32 AM
man		
NAMESPACE	48 B	Feb 23, 2016, 11:24 AM
R		
Read-and-delete-me	424 B	Feb 23, 2016, 10:20 AM
Sorted_Average	785 B	Feb 26, 2016, 11:06 AM
Testinggit3.Rproj	361 B	Feb 23, 2016, 10:42 AM

Text Editor (Sorted_Average):

```
1 "Name" "Age" "Sex" "Grade" "Grade.Average"
2 "1" "Lauri" 21 "Female" 90 86.9375
3 "2" "Leonie" 21 "Female" 91 86.9375
4 "3" "Sherlyn" 22 "Female" 85 86.9375
5 "4" "Mikaela" 20 "Female" 69 86.9375
6 "5" "Aiko" 24 "Female" 97 86.9375
7 "6" "Tiffany" 21 "Female" 78 86.9375
8 "7" "Corina" 23 "Female" 81 86.9375
9 "8" "Petronila" 23 "Female" 98 86.9375
10 "9" "Alecia" 20 "Female" 87 86.9375
11 "10" "Shemika" 23 "Female" 97 86.9375
12 "11" "Fallon" 22 "Female" 90 86.9375
13 "12" "Deloris" 21 "Female" 67 86.9375
14 "13" "Ranee" 23 "Female" 91 86.9375
15 "14" "Ebony" 20 "Female" 84 86.9375
16 "15" "Delfina" 19 "Female" 93 86.9375
17 "16" "Ernestina" 19 "Female" 93 86.9375
18 "17" "Raul" 25 "Male" 80 80.25
19 "18" "Booker" 18 "Male" 83 80.25
20 "19" "Raphael" 23 "Male" 91 80.25
21 "20" "Milo" 19 "Male" 67 80.25
22
```

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.

```
func1.R x Testinggit3-package.Rd x Sorted_Average_2 x x z x Sorted_Average x
1 "Name","Age","Sex","Grade","Grade.Average"
2 "1","Lauri",21,"Female",90,86.9375
3 "2","Leonie",21,"Female",91,86.9375
4 "3","Sherlyn",22,"Female",85,86.9375
5 "4","Mikaela",20,"Female",69,86.9375
6 "5","Aiko",24,"Female",97,86.9375
7 "6","Tiffaney",21,"Female",78,86.9375
8 "7","Corina",23,"Female",81,86.9375
9 "8","Petronila",23,"Female",98,86.9375
10 "9","Alecia",20,"Female",87,86.9375
11 "10","Shemika",23,"Female",97,86.9375
12 "11","Fallon",22,"Female",90,86.9375
13 "12","Deloris",21,"Female",67,86.9375
14 "13","Randee",23,"Female",91,86.9375
15 "14","Eboni",20,"Female",84,86.9375
16 "15","Delfina",19,"Female",93,86.9375
17 "16","Ernestina",19,"Female",93,86.9375
18 "17","Raul",25,"Male",80,80.25
19 "18","Booker",18,"Male",83,80.25
20 "19","Raphael",23,"Male",91,80.25
21 "20","Milo",19,"Male",67,80.25
22
```

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 "I" 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.

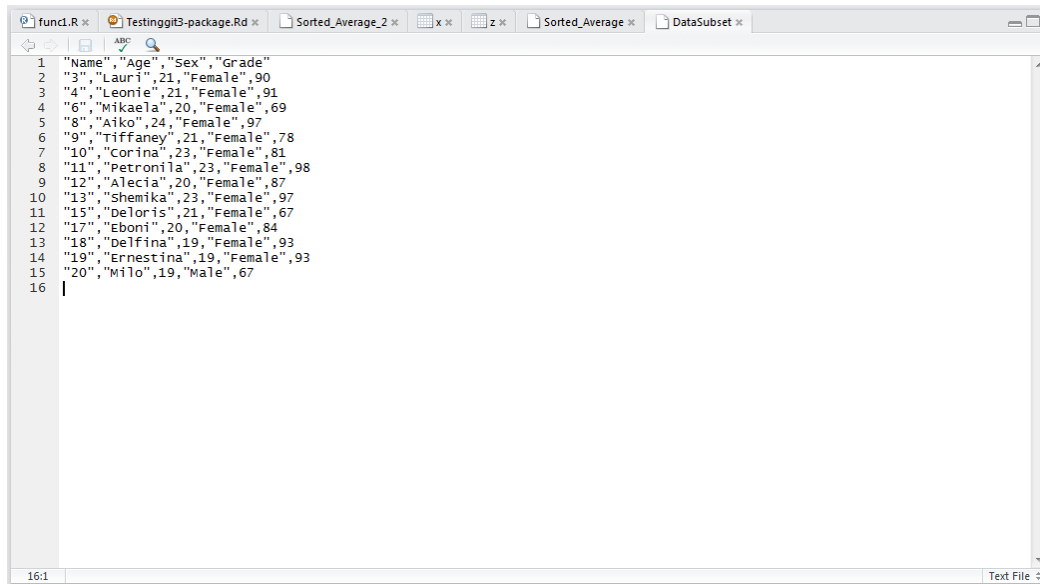
x is where we stored the imported data set.

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

```
> subset(x ,grep("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=",")
```



The screenshot shows an RStudio editor window with the following tabs: `func1.R`, `Testinggit3-package.Rd`, `Sorted_Average_2`, `x`, `z`, `Sorted_Average`, and `DataSubset`. The active tab is `func1.R`, which contains the following R code:

```
1 "Name", "Age", "Sex", "Grade"
2 "3", "Lauri", 21, "Female", 90
3 "4", "Leonie", 21, "Female", 91
4 "6", "Mikaela", 20, "Female", 69
5 "8", "Aiko", 24, "Female", 97
6 "9", "Tiffany", 21, "Female", 78
7 "10", "Corrina", 23, "Female", 81
8 "11", "Petronilla", 23, "Female", 98
9 "12", "Alecia", 20, "Female", 87
10 "13", "Shemika", 23, "Female", 97
11 "15", "Deloris", 21, "Female", 67
12 "17", "Ebony", 20, "Female", 84
13 "18", "Delfina", 19, "Female", 93
14 "19", "Ernestina", 19, "Female", 93
15 "20", "Milo", 19, "Male", 67
16 |
```

The status bar at the bottom left shows `16:1` and the bottom right shows `Text File`.