

Importing Data Sets Into R—

There are essentially 2 ways to import data sets

If you make an EXCEL data file or an SPSS data file, and you want to use that file in an R application, save it as a **.csv** extension (comma separated version). When you save, **.csv** will be one of the options you can save the file in.

Previously we have used the file **elecbill.sav** in our lab. I saved it as **elecbill.csv**, and stored it in the same place as I have my R stored. [On your home computer R is automatically stored in My Documents folder, so put the .csv file in the same place; on a lab computer, place the **.csv** file on the Desktop.]

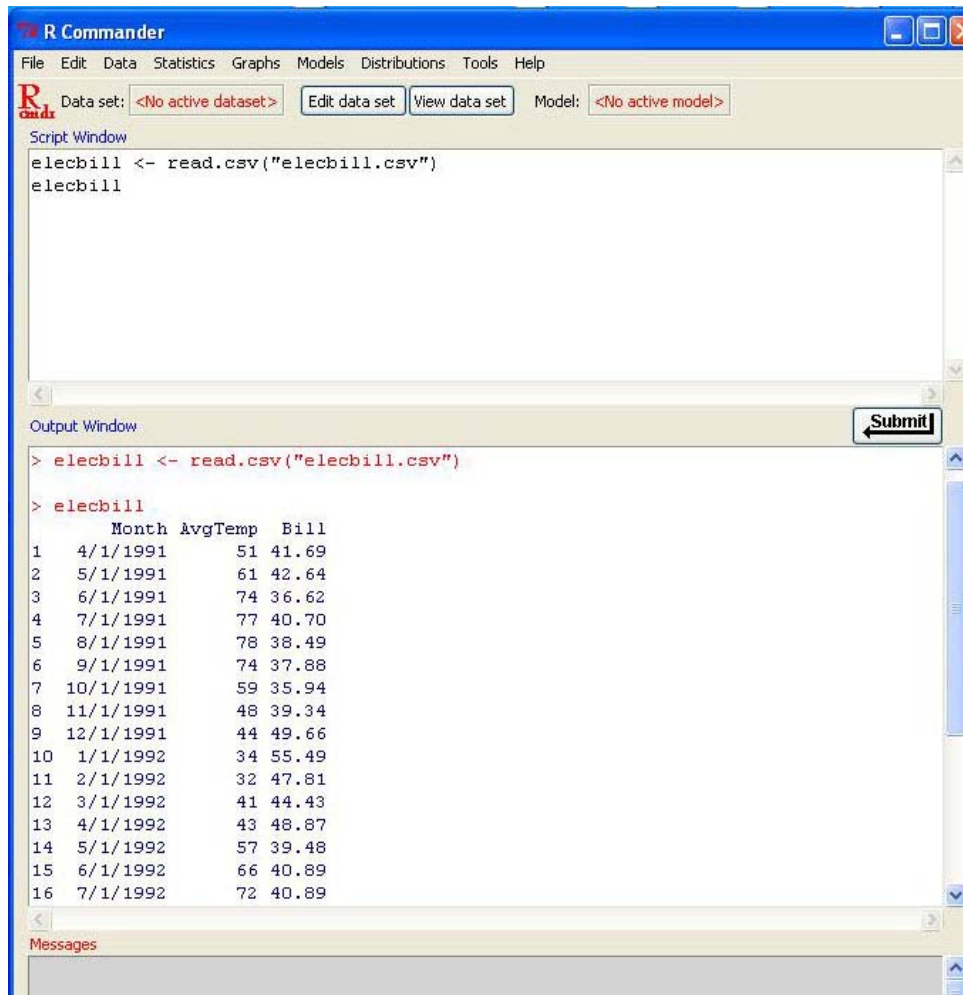


The icon saved as a **.csv** looks like this.

Now open R-Commander and type the command:

```
elecbill <- read.csv("elecbill.csv", header=TRUE)
```

as shown below. Note that the data file is labeled elecbill, is imported into R, and the column headers have been saved.



The screenshot shows the R Commander interface. The Script Window contains the following code:

```
elecbill <- read.csv("elecbill.csv")
elecbill
```

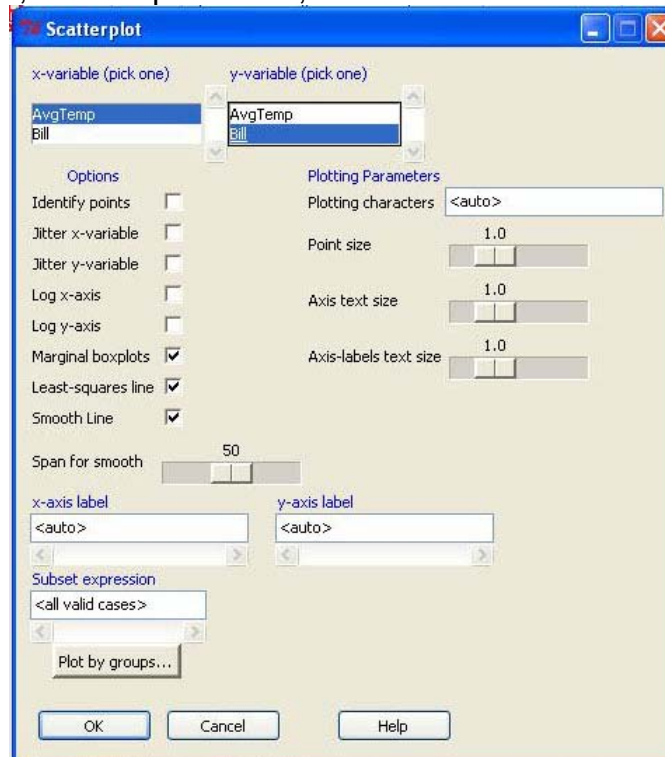
The Output Window shows the result of the command:

```
> elecbill <- read.csv("elecbill.csv")
> elecbill
      Month AvgTemp  Bill
1  4/1/1991      51 41.69
2  5/1/1991      61 42.64
3  6/1/1991      74 36.62
4  7/1/1991      77 40.70
5  8/1/1991      78 38.49
6  9/1/1991      74 37.88
7 10/1/1991      59 35.94
8 11/1/1991      48 39.34
9 12/1/1991      44 49.66
10 1/1/1992      34 55.49
11 2/1/1992      32 47.81
12 3/1/1992      41 44.43
13 4/1/1992      43 48.87
14 5/1/1992      57 39.48
15 6/1/1992      66 40.89
16 7/1/1992      72 40.89
```

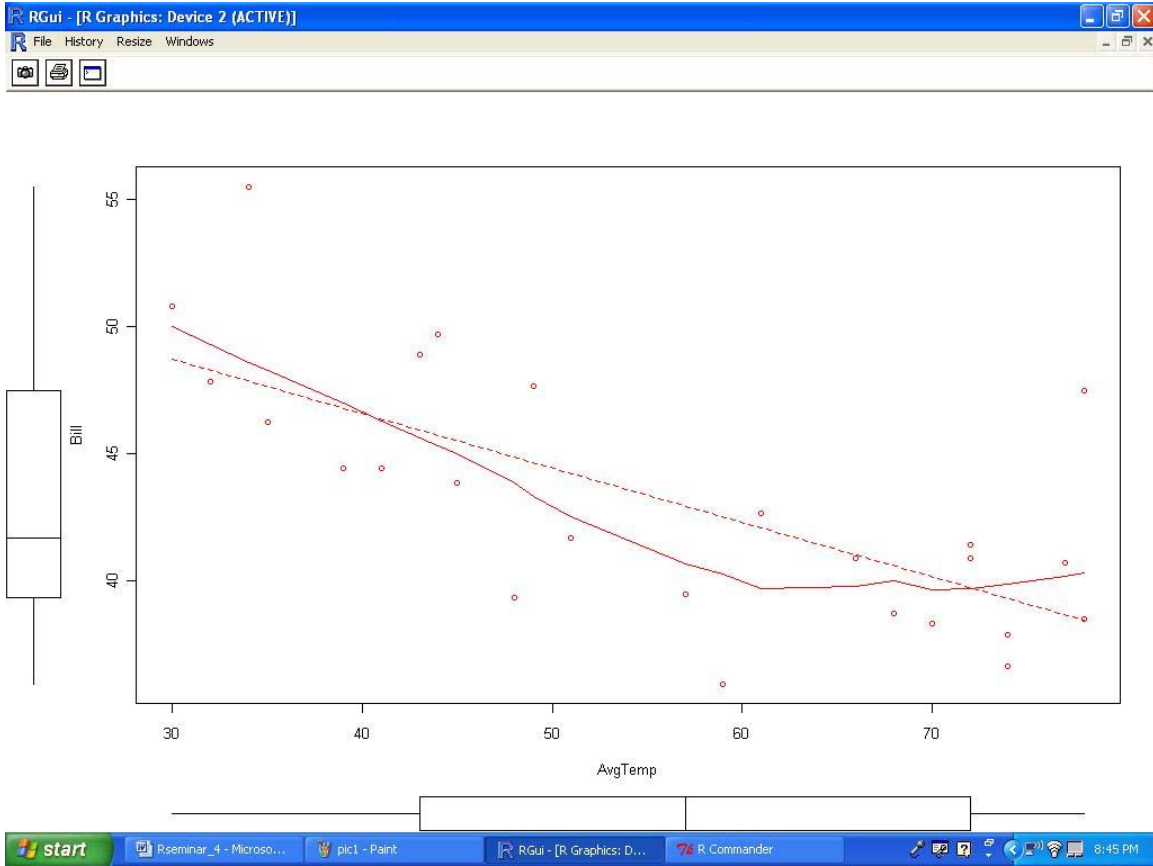
Now click on DATA ACTIVE DATA SET SELECT ACTIVE DATA SET, to select elecbill as the “active data set”.

We want to make a scatterplot and find the line of best fit, plot that line as well as plot the “loess” curve of best fit.

So, click on GRAPH SCATTERPLOT and get the window below. Be sure to click on all items shown: that is, pick AvgTemp as x variable, Bill as y variable, Marginal boxplots, Least-squares line, and Smooth line checked.

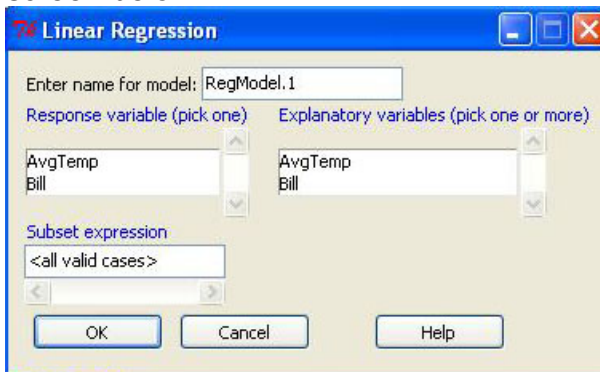


Now, open up the graphics window from Rgui, and see the scatterplot shown.



Note that the points, best fit line, “lowess” line, and boxplots of the x and y distributions is shown.

Now, click on **STATISTICS** **FIT MODEL** **LINEAR REGRESSION**, to get the screen below.



Again, pick Bill as response variable, and AvgTemp as explanatory variable, to get the output information on the regression shown below.

```
Output Window Submit
Call:
lm(formula = Bill ~ AvgTemp, data = elecbill)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5769 -1.9354 -0.4797  2.0316  9.0154

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  55.13140    2.67855   20.583 2.58e-16 ***
AvgTemp      -0.21380    0.04611   -4.637 0.000115 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.662 on 23 degrees of freedom
(3 observations deleted due to missingness)
Multiple R-squared:  0.4832, Adjusted R-squared:  0.4607
F-statistic:  21.5 on 1 and 23 DF,  p-value: 0.000115
```

Note that pushing these menu items and clicking on buttons will generate code which shows up in the script window as well as output window of R-Commander. So, you can see what the commands are which you would have had to type into the script window of Rgui if you had not had R-Commander.

I will leave to you more experimentation with R-Commander. You can import text files which have been saved in columns, with column headers, where columns are separated by spaces, just like you imported the **.csv** file above.

The text files are saved as **.txt**, as Jon saves a lot of his data files. To get them into R or R-Commander, you would type:

```
elecbill <- read.table("electbill.txt" header=TRUE)
```

Then do the same steps as shown above to graph the data, or to compute statistics (like we did our linear regression).

Finally, I hear there is another “user friendly” R feature, similar to R-Commander, which you can download for free, just like you did R and R-Commander. It is called Tinn-R (which you can Google) and I know nothing about it. You may play with it in your spare time.