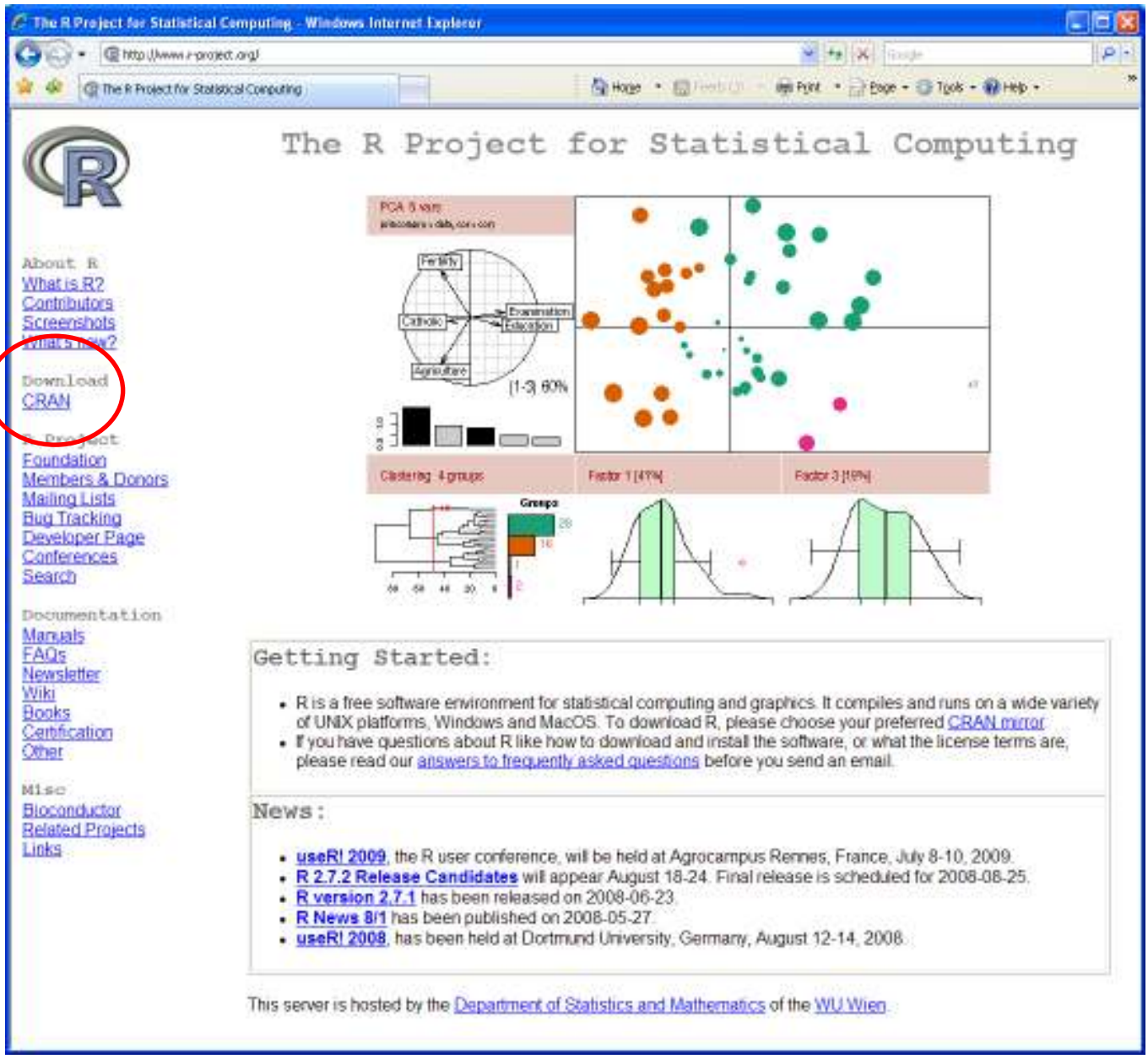
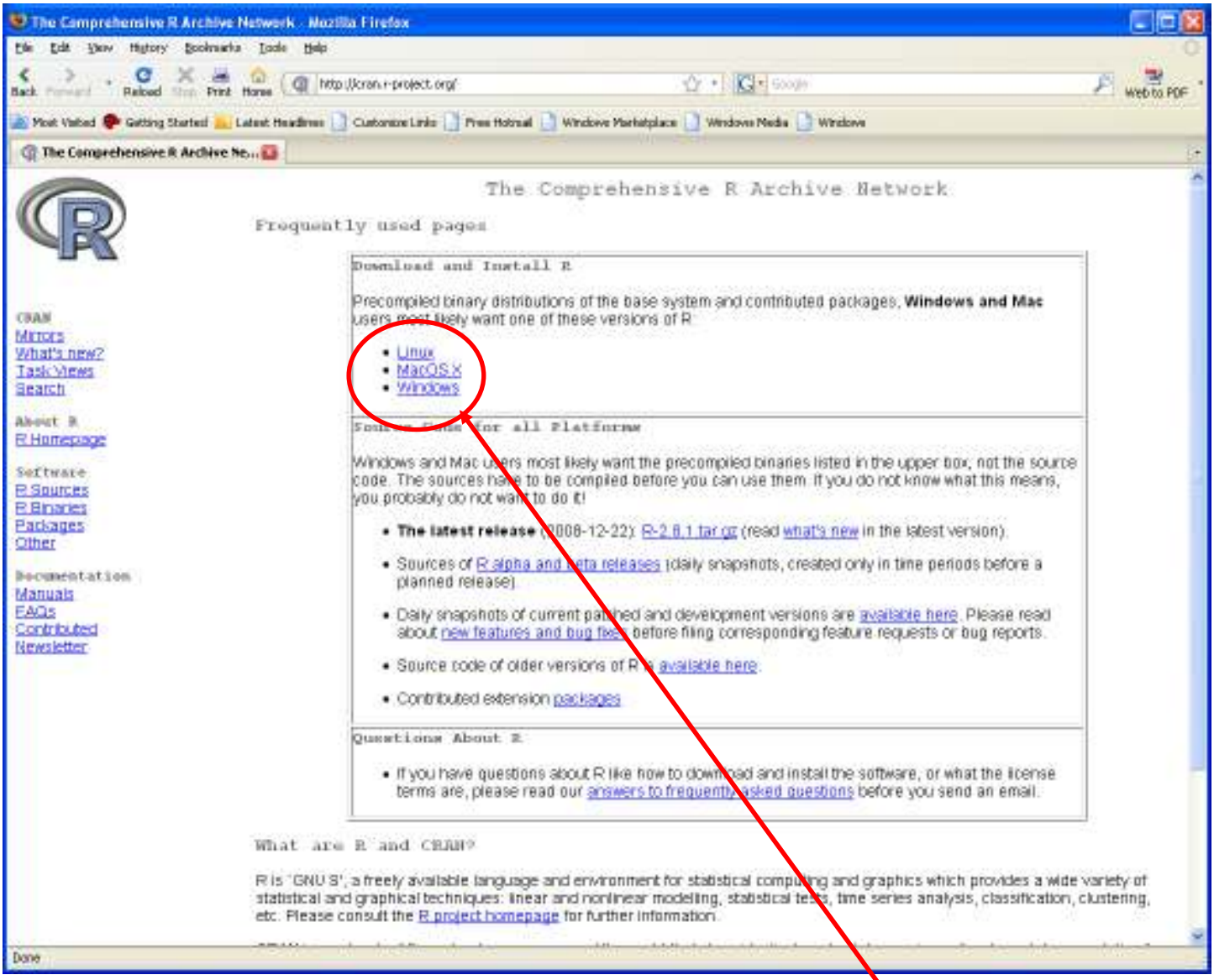
R Lecture for Students

Learning about R: go to http://www.r-project.org/

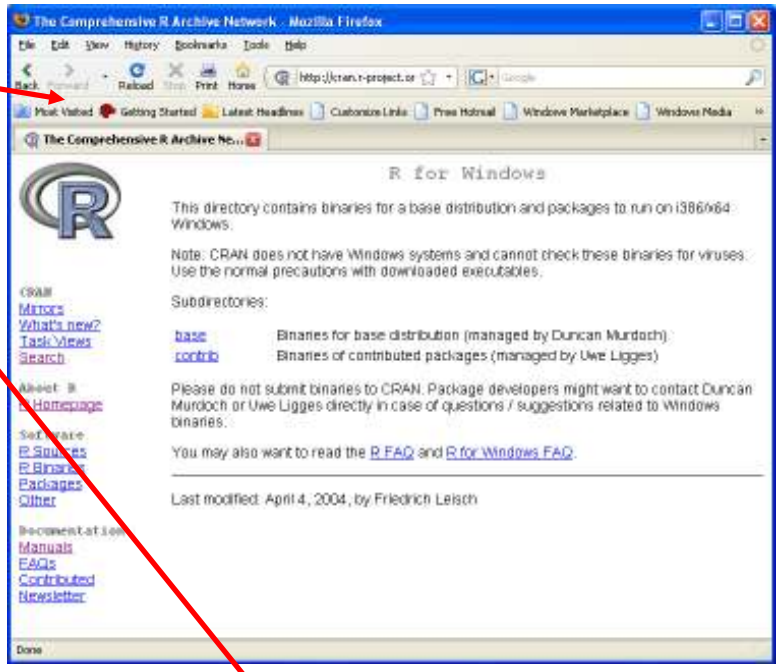


Getting R from CRAN website: http://cran.r-project.org/



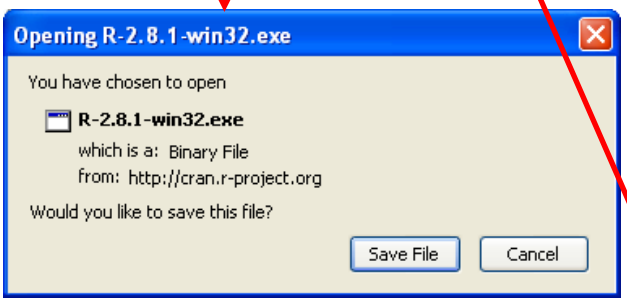
To download and install R on your computer, click the appropriate Operating System you have and follow the directions.

For example, for Windows users, clicking the Windows link get you to this:

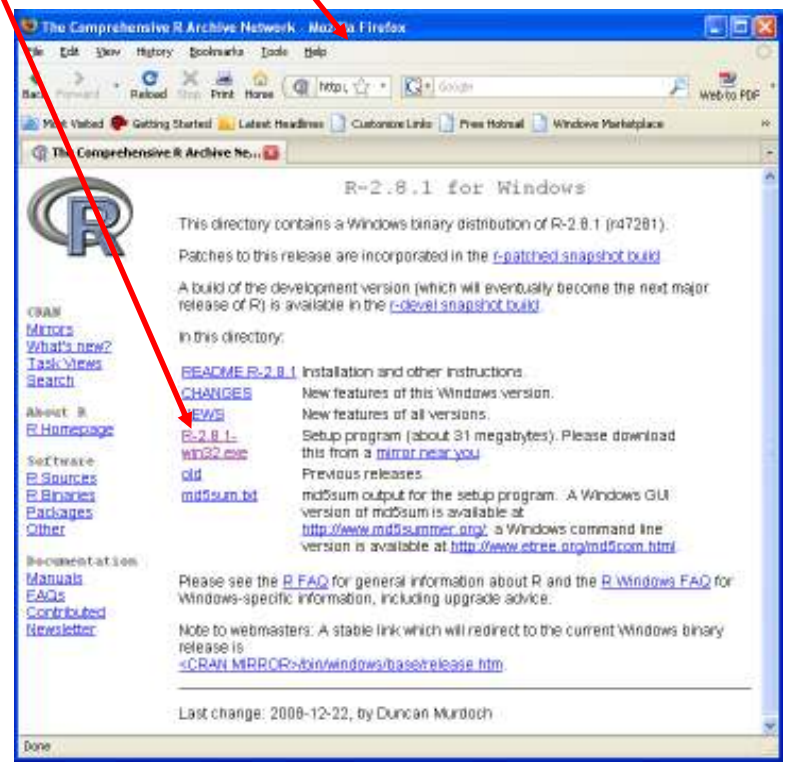


Clicking on the base link gets you to:

Finally, clicking on the R-2.8.1-win32.exe link, gets you



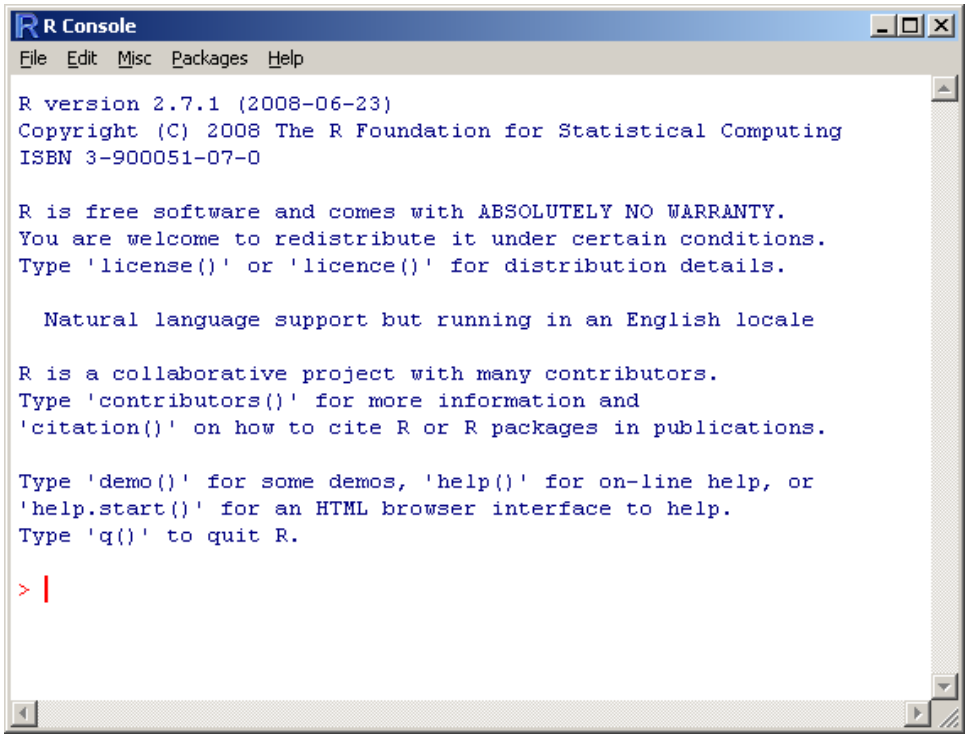
Click the Save File button and save this file to your hard drive. You will be asked where you want to save it. Once you save it, navigate to the file using Windows Explorer and double click on it. You be asked if you want to Run it, click the Run button.



Windows will then set up R. During the setup, when you are asked to choose either MDI or SDI, choose the SDI option. Accept all of the other defaults during the setup, since you can change them later.

Starting R (double-click the R icon or select R from the START menu)

R Startup Screen (aka, R Console Window)

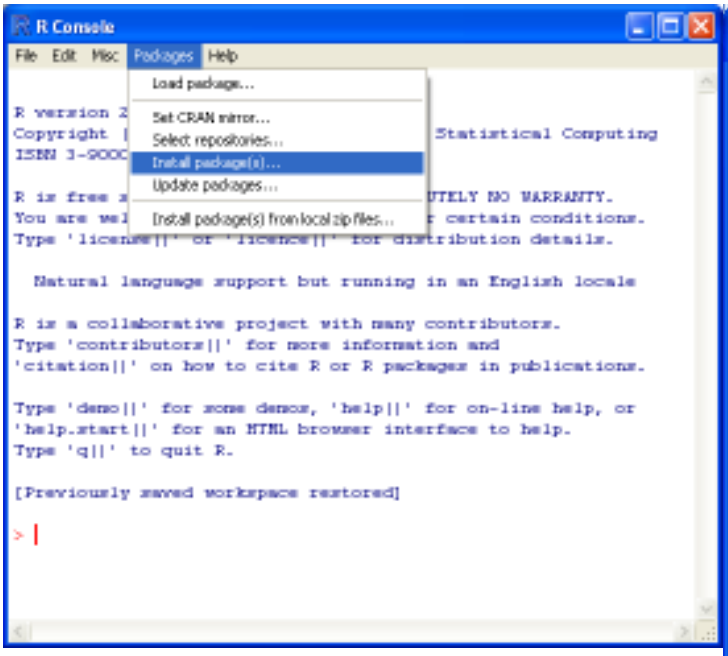


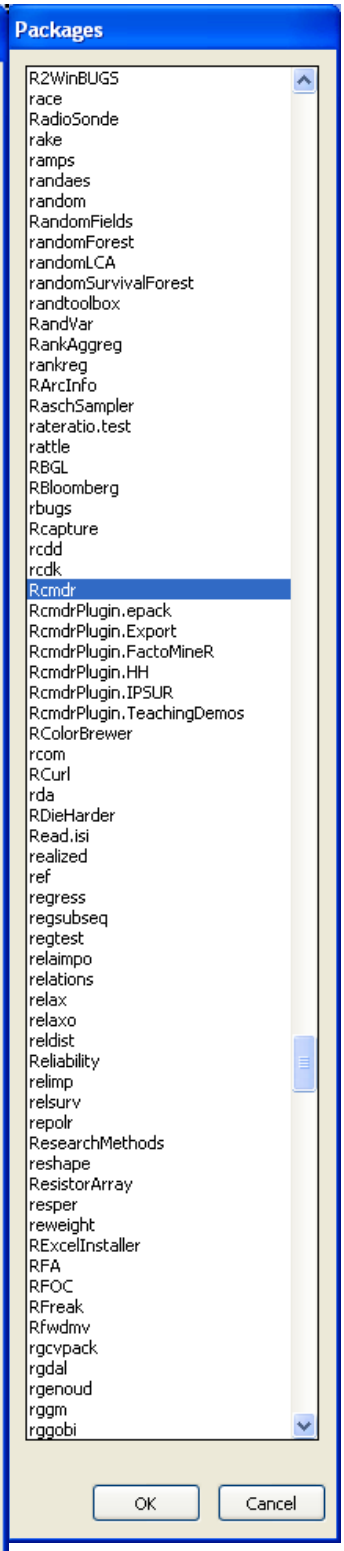
R is a command line system and the R Console is where commands are entered and output occurs.

RCommander, seen next, eliminates knowing commands and allows you to "point-and-click" to produce results.

Installing the Rcmdr package (becomes part of the R Library but not yet usable)

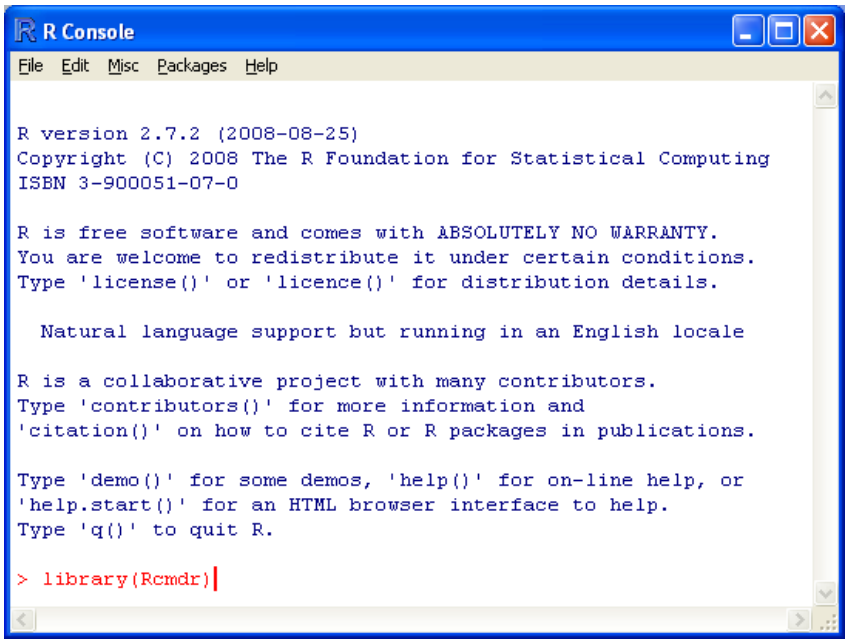
From here to here to here



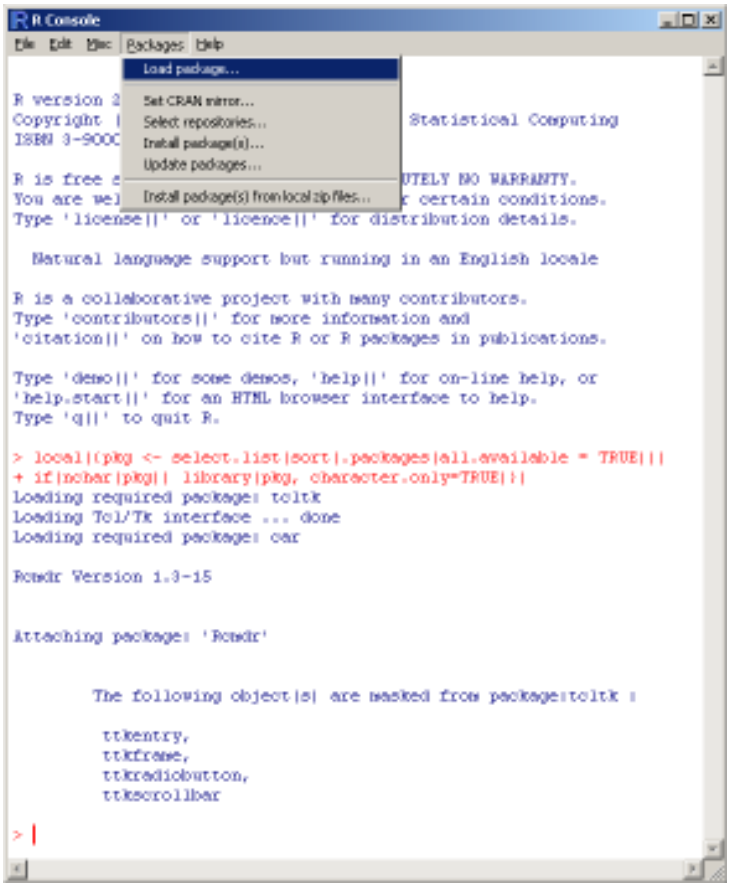


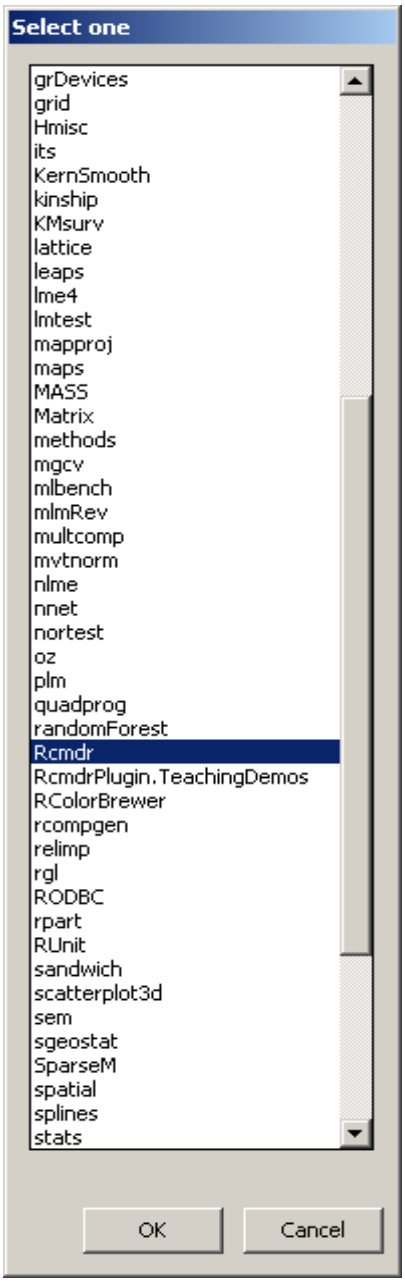
Starting RCommander

Once RCommander has been installed it is part of the R Library, BUT you still need to "start" RCommander by either using the library(Rcmdr) command



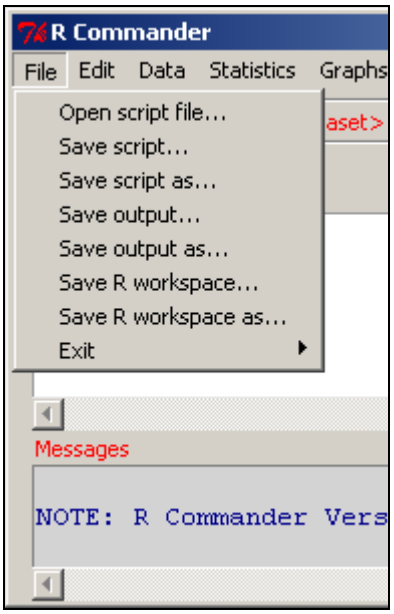
OR Load package... menu:

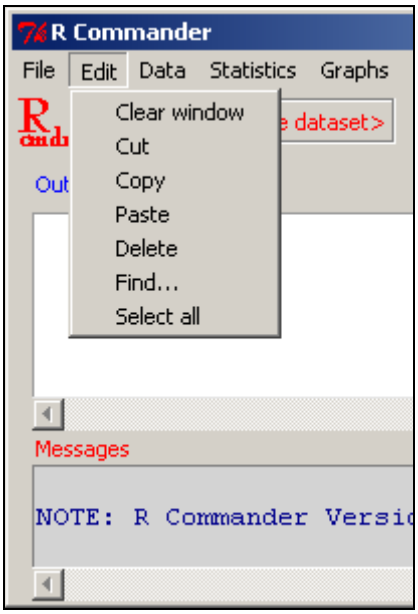


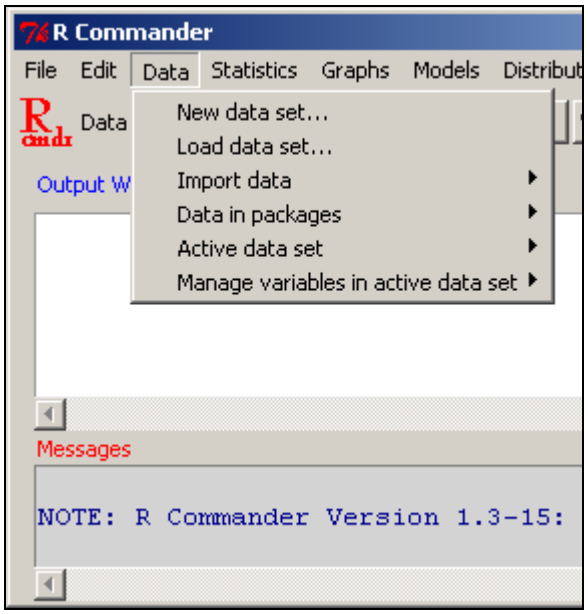


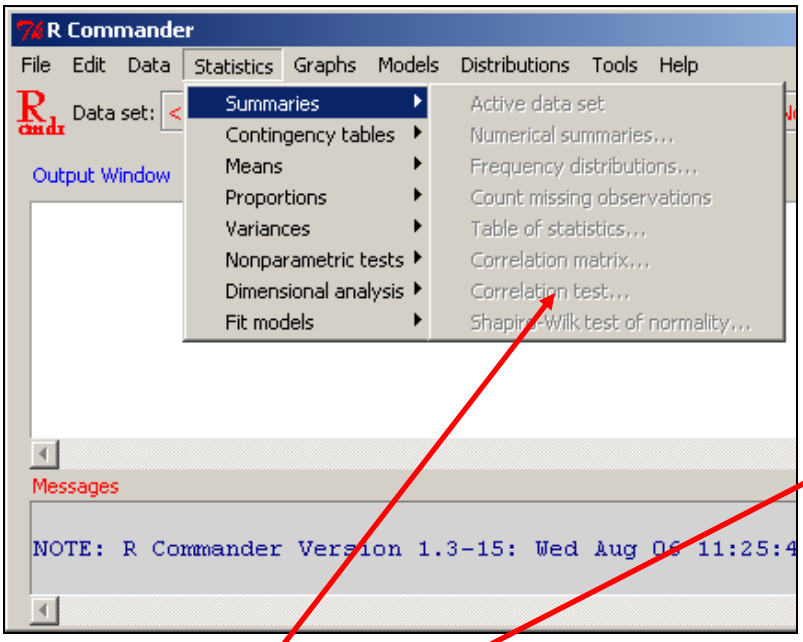
RCommander Window

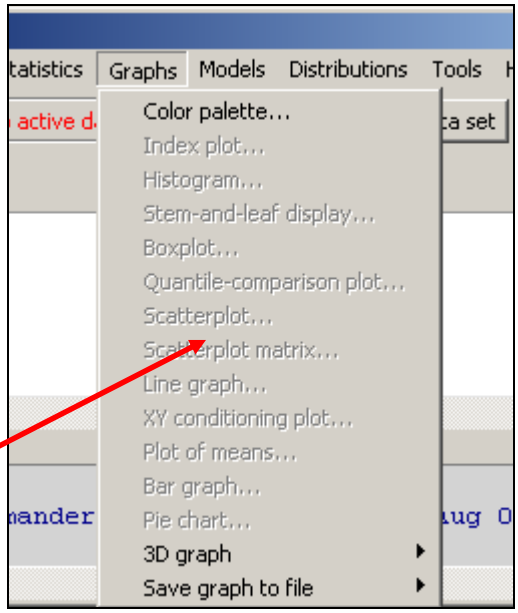
RCommander Menus



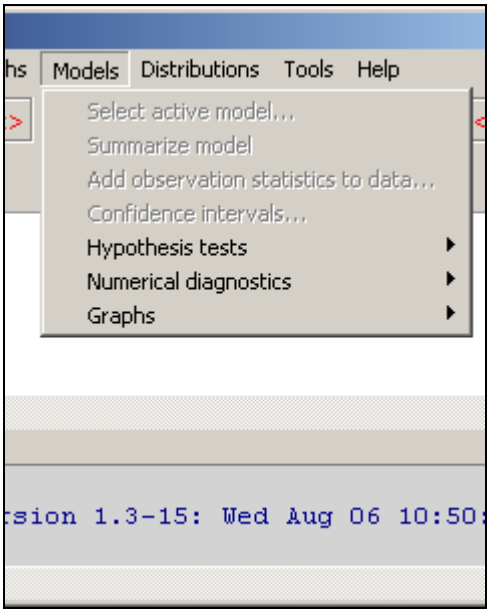


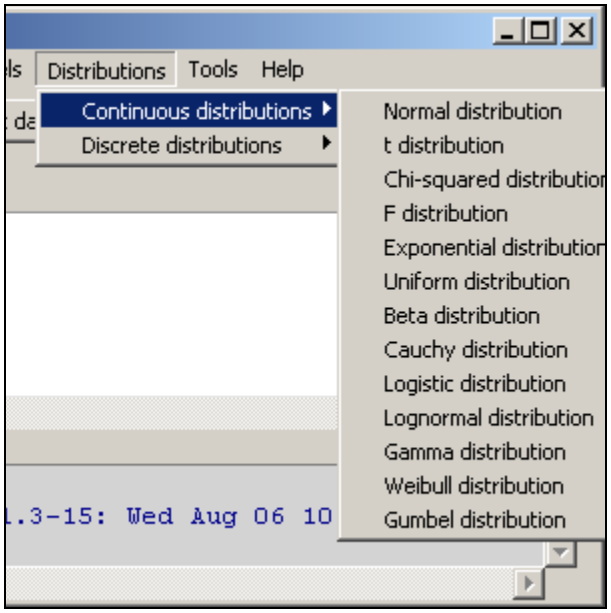


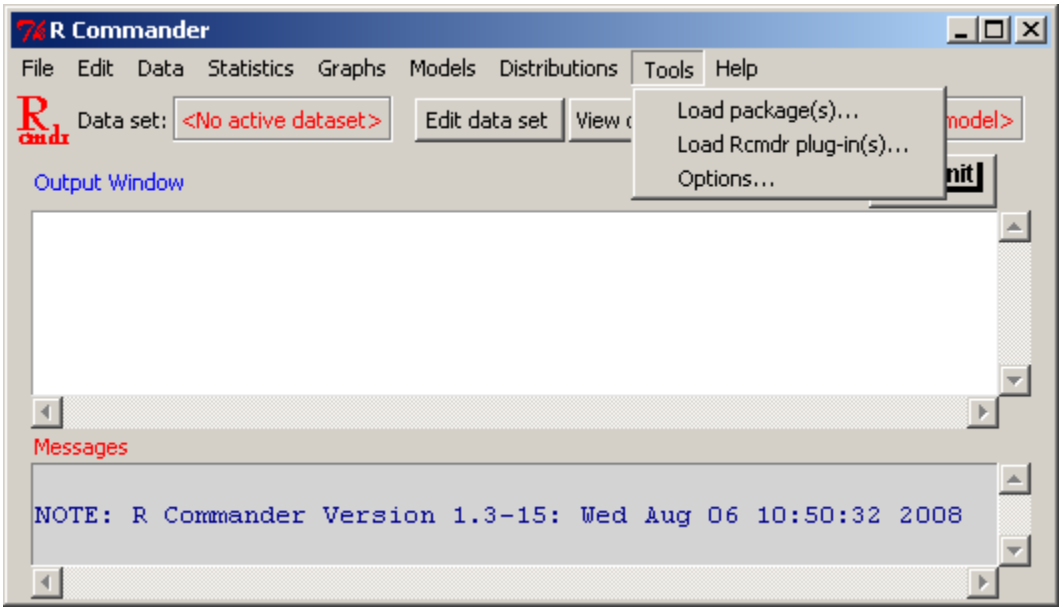




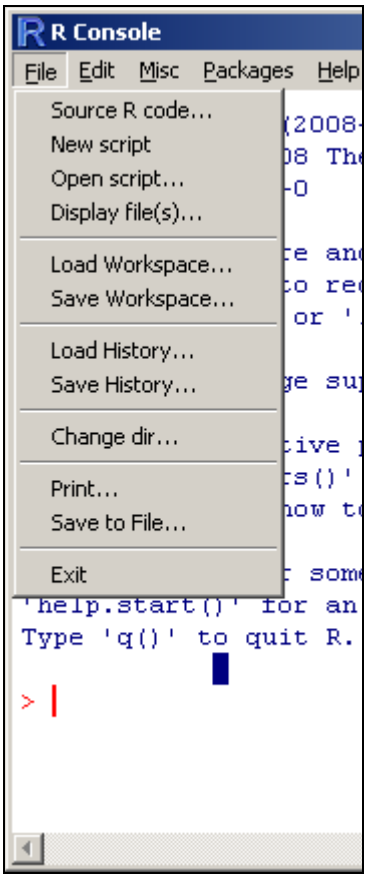
NOTE!

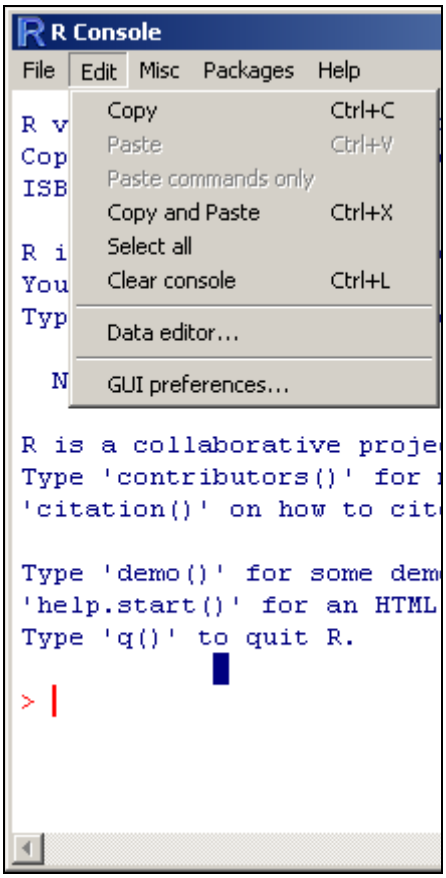

Models
Distributions
Tools
Help
Select active model...
Summarize model
Add observation statistics to data...
Confidence intervals...
Hypothesis tests
Numerical diagnostics
Graphs
sion 1.3-15: Wed Aug 06 10:50:

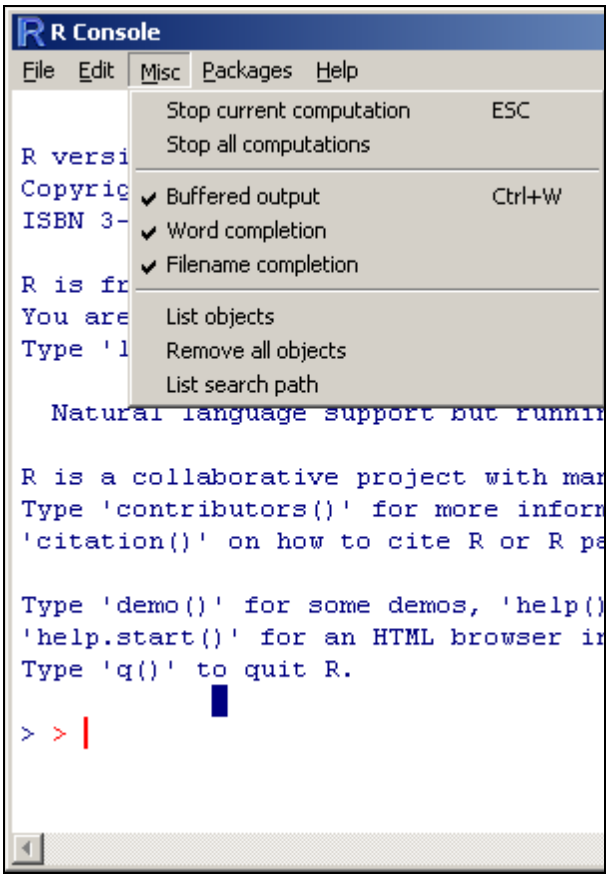

Distributions
Tools
Help
Continuous distributions
Discrete distributions
Normal distribution
t distribution
Chi-squared distributio
F distribution
Exponential distributio
Uniform distribution
Beta distribution
Cauchy distribution
Logistic distribution
Lognormal distribution
Gamma distribution
Weibull distribution
Gumbel distribution
.3-15: Wed Aug 06 10

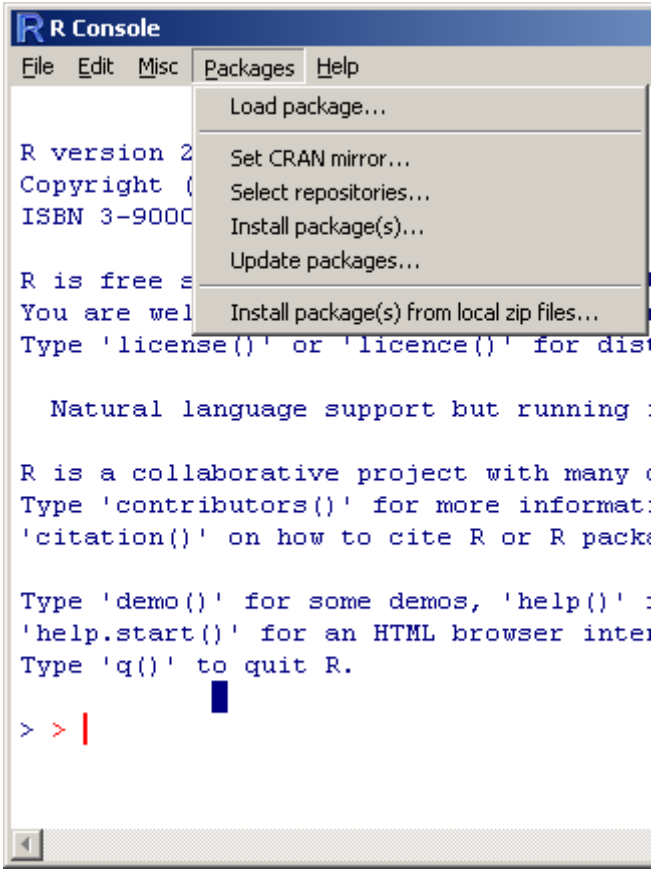

R Commander
File
Edit
Data
Statistics
Graphs
Models
Distributions
Tools
Help
Data set:
<No active dataset>
Edit data set
Load package(s)...
Load Rcmdr plug-in(s)...
Options...
Output Window
Messages
NOTE: R Commander Version 1.3-15: Wed Aug 06 10:50:32 2008

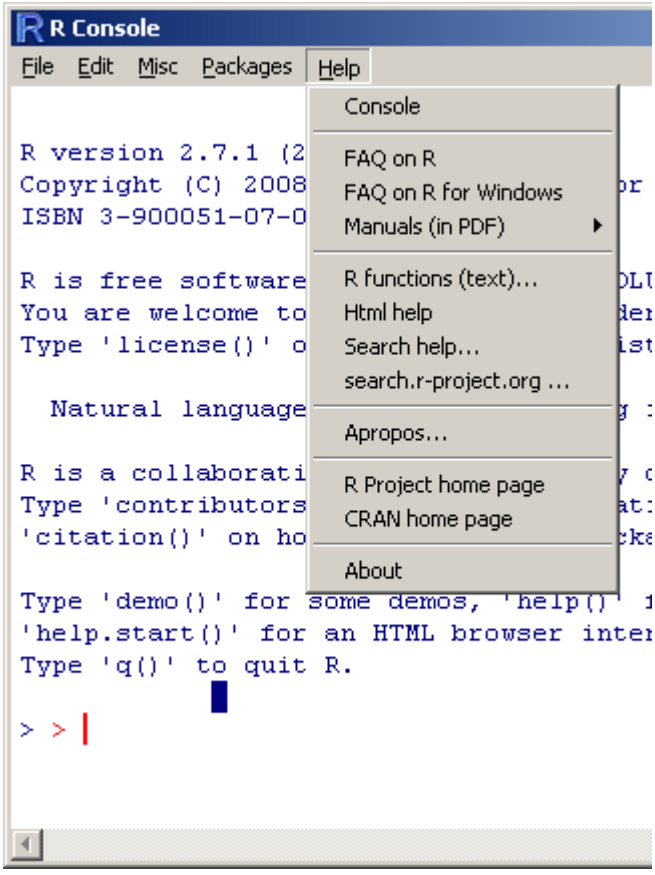
R Console Menus



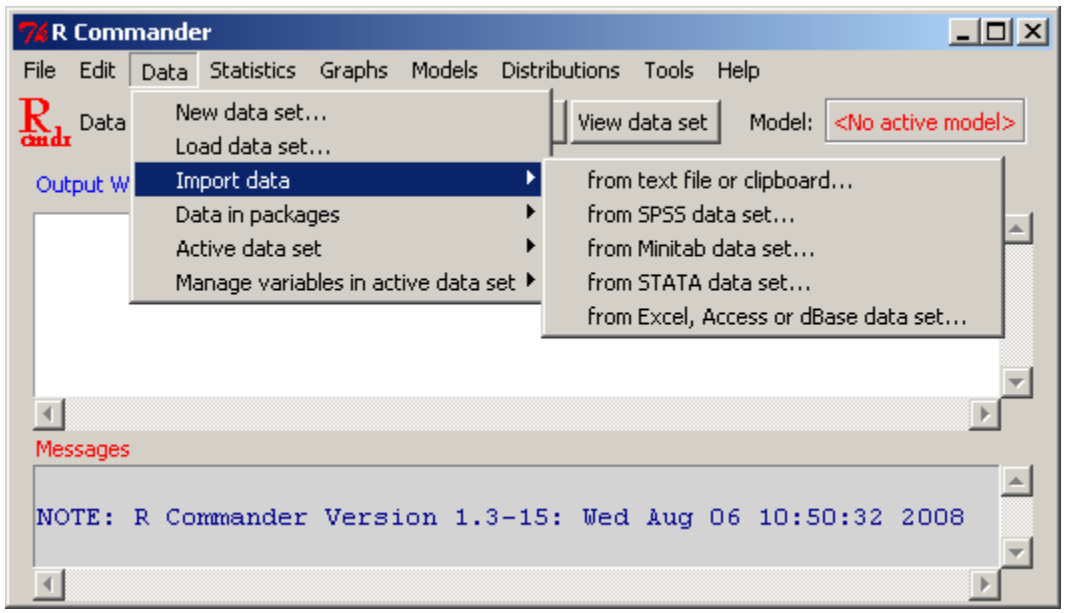








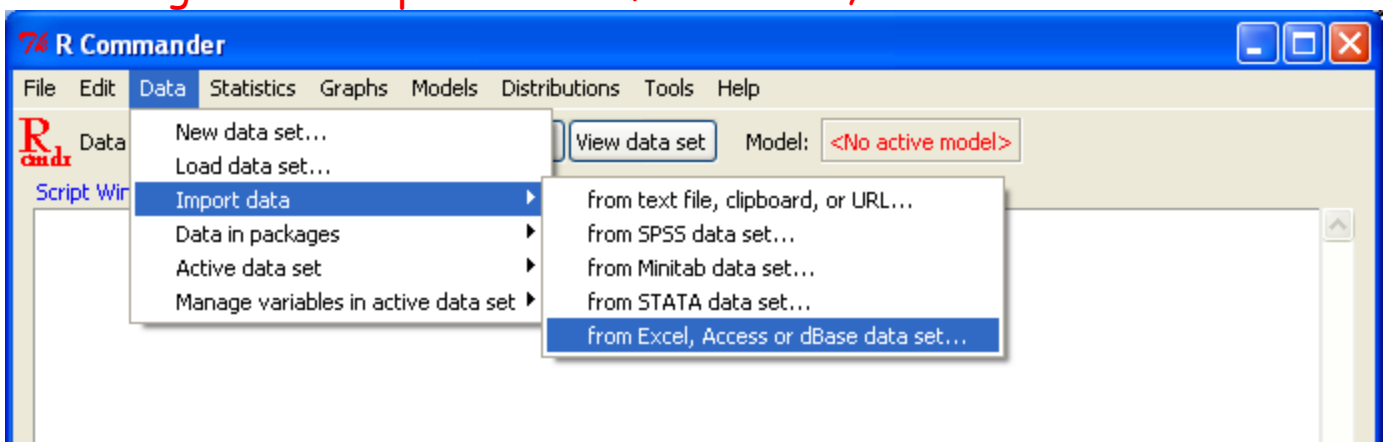
Getting Data Into R Using RCommander



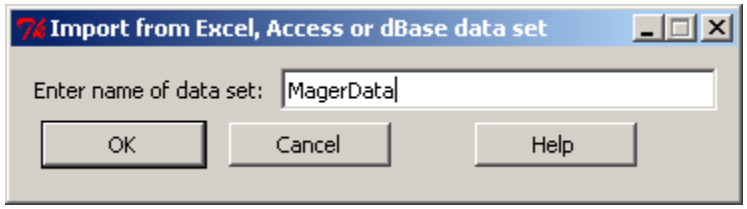
Importing Data from an Excel Spreadsheet

Jay Mager (Zoo) Michigan Loon Behavior Data

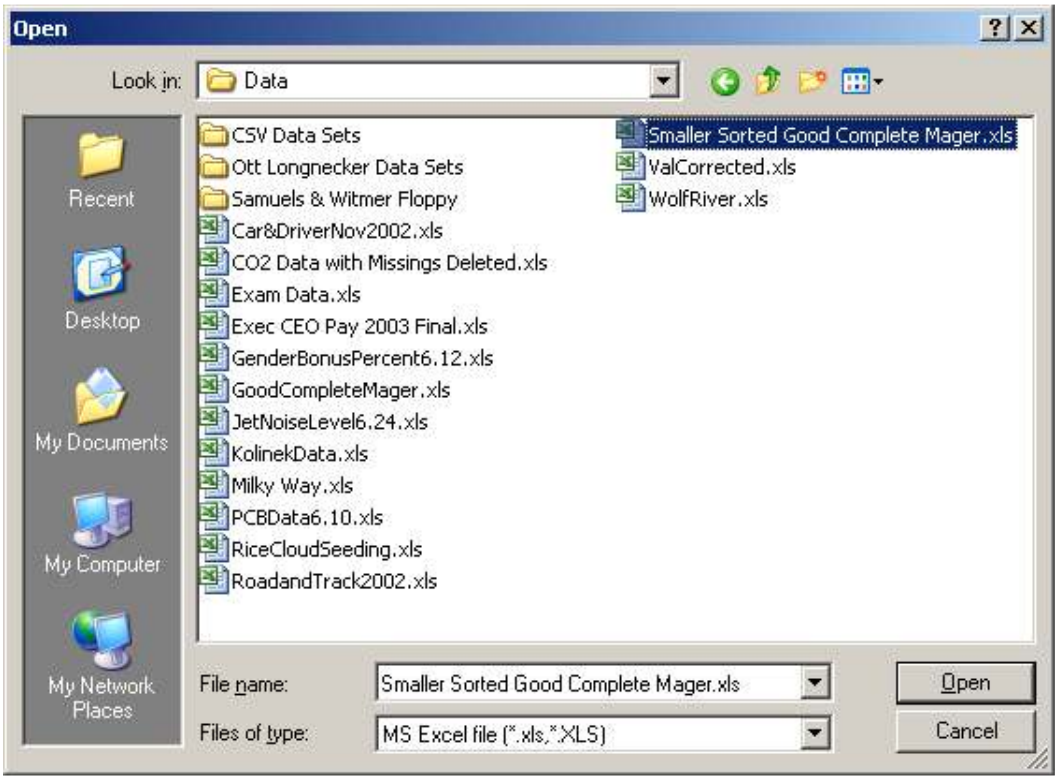
Selecting Data > Import data > from Excel, Access or dBase data set...



Yields



and



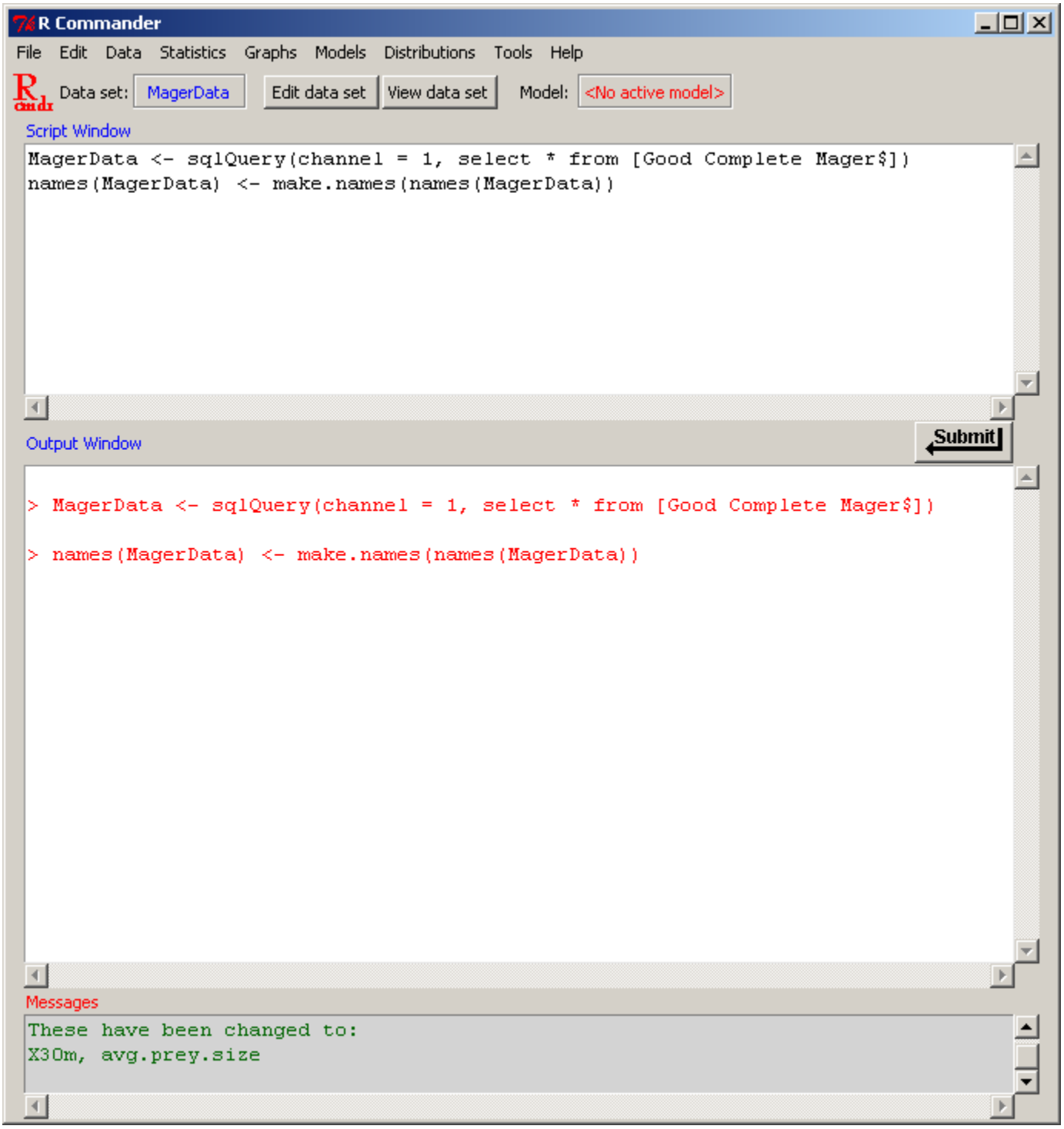
and

Notes re the Jay Mager data set. Each variable value of the behavior is the % of total time observed for that observation period that the behavior was observed.

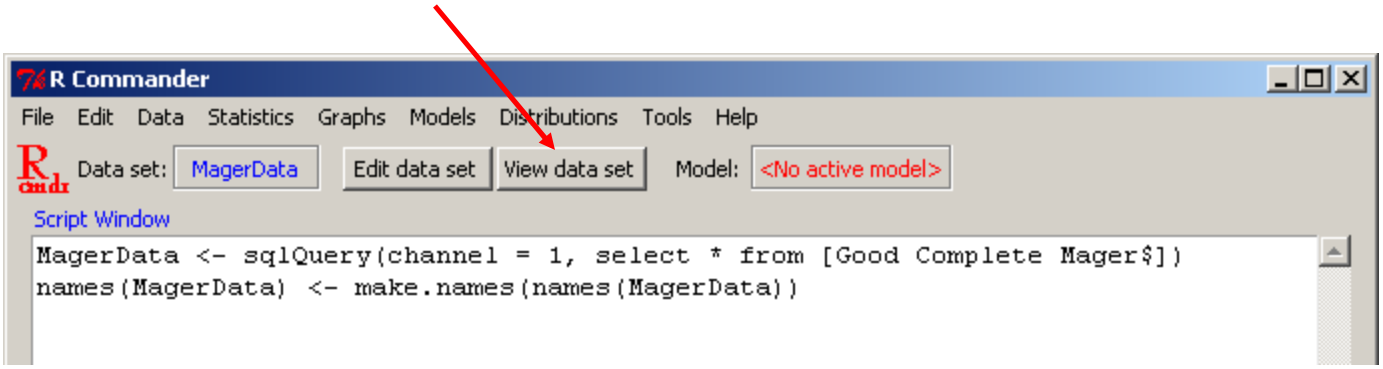
SF = foraging under water (for under 1min) for prey
LOC = locomotion other diving including flying, surface swimming, and walking on land
REST = floating or drifting on surface of water, either sleeping or awake
MAIN = preening, waterproofing of feathers, or bathing
SOC = social gatherings of loons, including circle swimming and jerk-diving
INC = incubation, sitting on eggs, regardless of other activity
NB = nest building or maintenance, when material is brought from under water or land and added to nest
CF = chick feeding, foraging for chicks and presenting chicks with prey
AG = aggression and alertness towards other loons and individuals
CARRY = carrying chicks on their backs

NOTES:

- Script
- Output
- R Commands
- "Active" Data Set



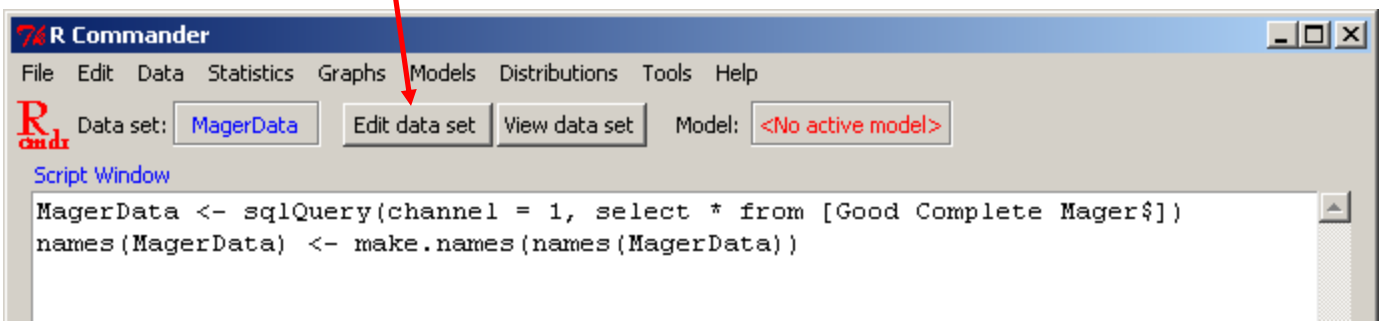
"View data set" button

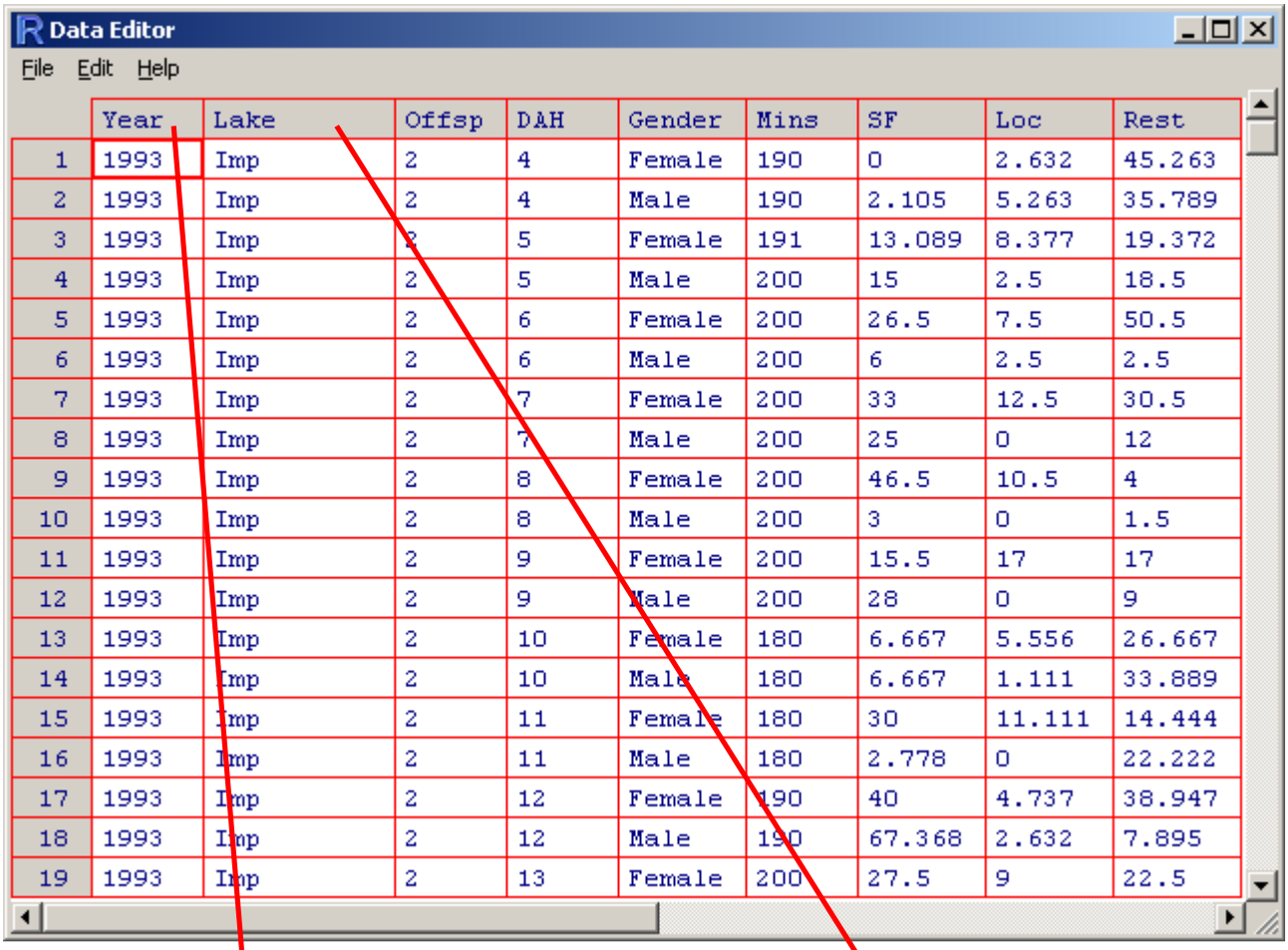


MagerData

	Year	Lake	Offsp	DAH	Gender	Mins	SF	Loc	Rest	Main	Soc	Inc
1	1993	Imp	2	4	Female	190	0.000	2.632	45.263	2.105	0.000	0.000
2	1993	Imp	2	4	Male	190	2.105	5.263	35.789	8.421	0.000	0.000
3	1993	Imp	2	5	Female	191	13.089	8.377	19.372	5.759	0.000	0.000
4	1993	Imp	2	5	Male	200	15.000	2.500	18.500	8.000	0.000	0.000
5	1993	Imp	2	6	Female	200	26.500	7.500	50.500	6.500	0.000	0.000
6	1993	Imp	2	6	Male	200	6.000	2.500	2.500	2.500	0.000	0.000
7	1993	Imp	2	7	Female	200	33.000	12.500	30.500	12.500	0.000	0.000
8	1993	Imp	2	7	Male	200	25.000	0.000	12.000	11.000	0.000	0.000
9	1993	Imp	2	8	Female	200	46.500	10.500	4.000	24.000	0.000	0.000
10	1993	Imp	2	8	Male	200	3.000	0.000	1.500	0.000	0.000	0.000
11	1993	Imp	2	9	Female	200	15.500	17.000	17.000	12.500	0.000	0.000
12	1993	Imp	2	9	Male	200	28.000	0.000	9.000	8.500	0.000	0.000
13	1993	Imp	2	10	Female	180	6.667	5.556	26.667	5.556	0.000	0.000
14	1993	Imp	2	10	Male	180	6.667	1.111	33.889	4.444	0.000	0.000
15	1993	Imp	2	11	Female	180	30.000	11.111	14.444	8.333	0.000	0.000
16	1993	Imp	2	11	Male	180	2.778	0.000	22.222	20.556	0.000	0.000
17	1993	Imp	2	12	Female	190	40.000	4.737	38.947	4.737	0.000	0.000
18	1993	Imp	2	12	Male	190	67.368	2.632	7.895	13.158	0.000	0.000
19	1993	Imp	2	13	Female	200	27.500	9.000	22.500	15.500	0.000	0.000
20	1993	Imp	2	13	Male	200	32.000	11.000	38.500	7.000	0.000	0.000
21	1993	Imp	2	17	Female	200	51.000	11.000	18.000	11.500	0.000	0.000
22	1993	Imp	2	17	Male	200	24.000	7.500	37.500	17.000	0.000	0.000
23	1993	Imp	2	18	Female	200	29.500	8.500	26.000	15.500	0.000	0.000
24	1993	Imp	2	18	Male	200	22.500	7.000	26.500	20.500	0.000	0.000
25	1993	Imp	2	19	Female	200	37.500	14.500	26.000	9.000	3.000	0.000
26	1993	Imp	2	19	Male	200	10.000	3.000	46.000	13.500	13.000	0.000
27	1993	Imp	2	20	Female	200	47.500	3.000	15.500	14.000	0.000	0.000
28	1993	Imp	2	20	Male	200	56.000	0.000	14.500	12.000	0.000	0.000
29	1993	Imp	2	21	Female	190	45.263	19.474	24.211	9.474	0.000	0.000
30	1993	Imp	2	21	Male	190	33.684	22.105	36.842	7.368	0.000	0.000

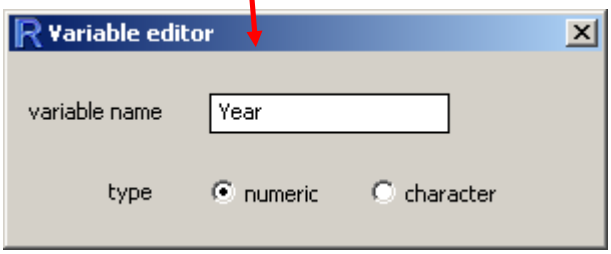
"Edit data set" button

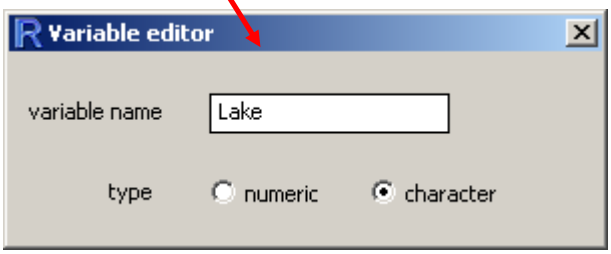




	Year	Lake	Offsp	DAH	Gender	Mins	SF	Loc	Rest
1	1993	Imp	2	4	Female	190	0	2.632	45.263
2	1993	Imp	2	4	Male	190	2.105	5.263	35.789
3	1993	Imp	2	5	Female	191	13.089	8.377	19.372
4	1993	Imp	2	5	Male	200	15	2.5	18.5
5	1993	Imp	2	6	Female	200	26.5	7.5	50.5
6	1993	Imp	2	6	Male	200	6	2.5	2.5
7	1993	Imp	2	7	Female	200	33	12.5	30.5
8	1993	Imp	2	7	Male	200	25	0	12
9	1993	Imp	2	8	Female	200	46.5	10.5	4
10	1993	Imp	2	8	Male	200	3	0	1.5
11	1993	Imp	2	9	Female	200	15.5	17	17
12	1993	Imp	2	9	Male	200	28	0	9
13	1993	Imp	2	10	Female	180	6.667	5.556	26.667
14	1993	Imp	2	10	Male	180	6.667	1.111	33.889
15	1993	Imp	2	11	Female	180	30	11.111	14.444
16	1993	Imp	2	11	Male	180	2.778	0	22.222
17	1993	Imp	2	12	Female	190	40	4.737	38.947
18	1993	Imp	2	12	Male	190	67.368	2.632	7.895
19	1993	Imp	2	13	Female	200	27.5	9	22.5

Kinds of Data in R

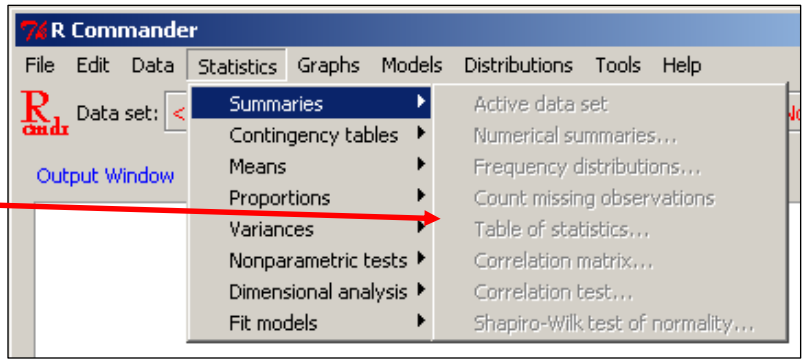


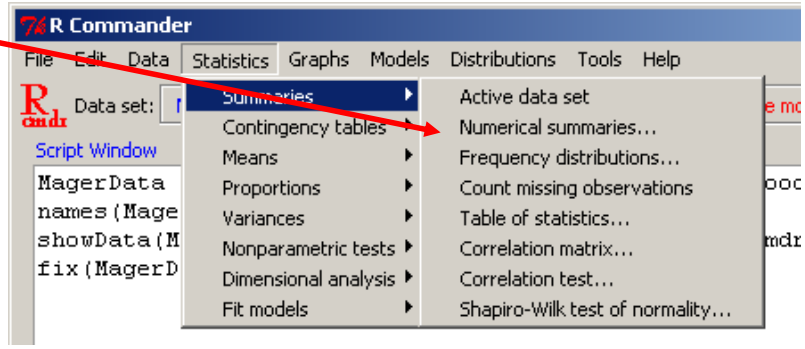


Now What Can We Do?

Recall what options we had without an active dataset. NONE!

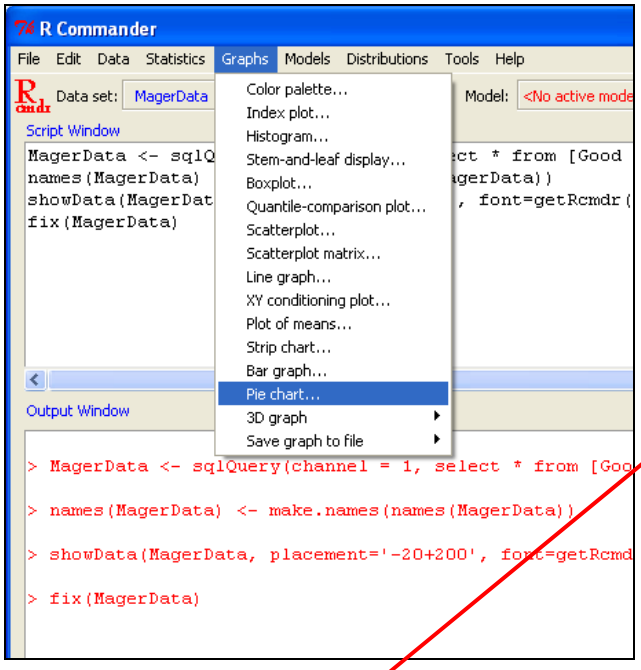
But now with an active data set, we have options!

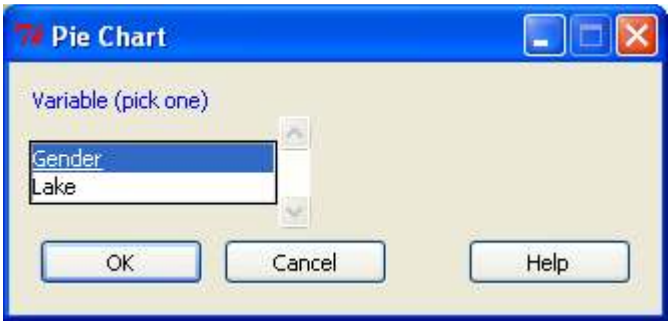


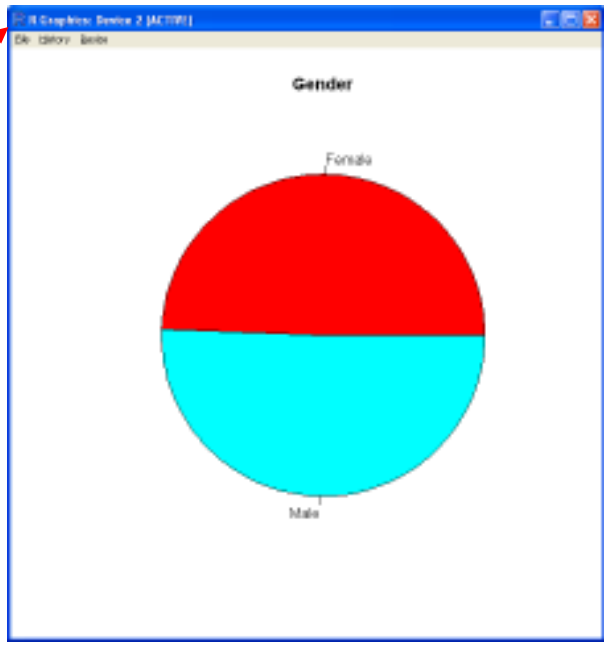


Summarizing CATEGORICAL Data (Graphically and Numerically)

GRAPHICAL Summary of CATEGORICAL Data (Pie chart/Bar graph)

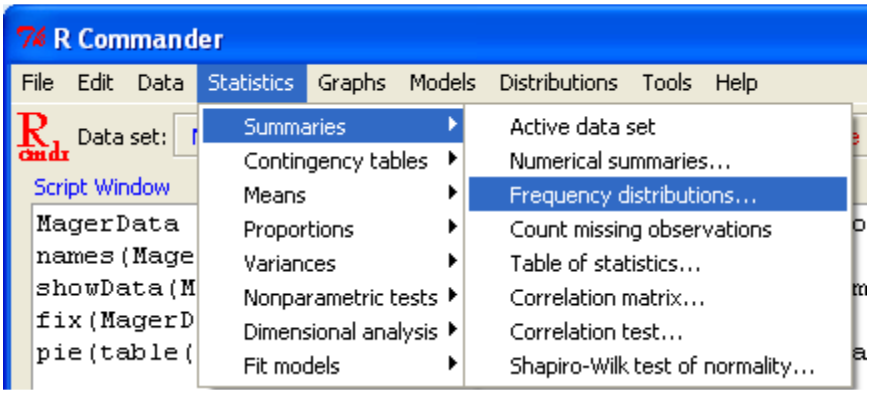


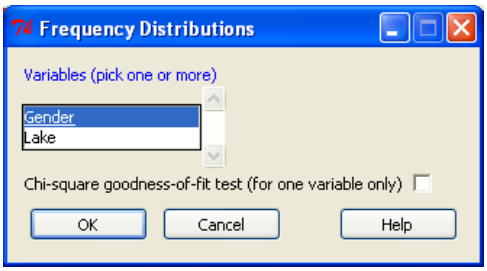


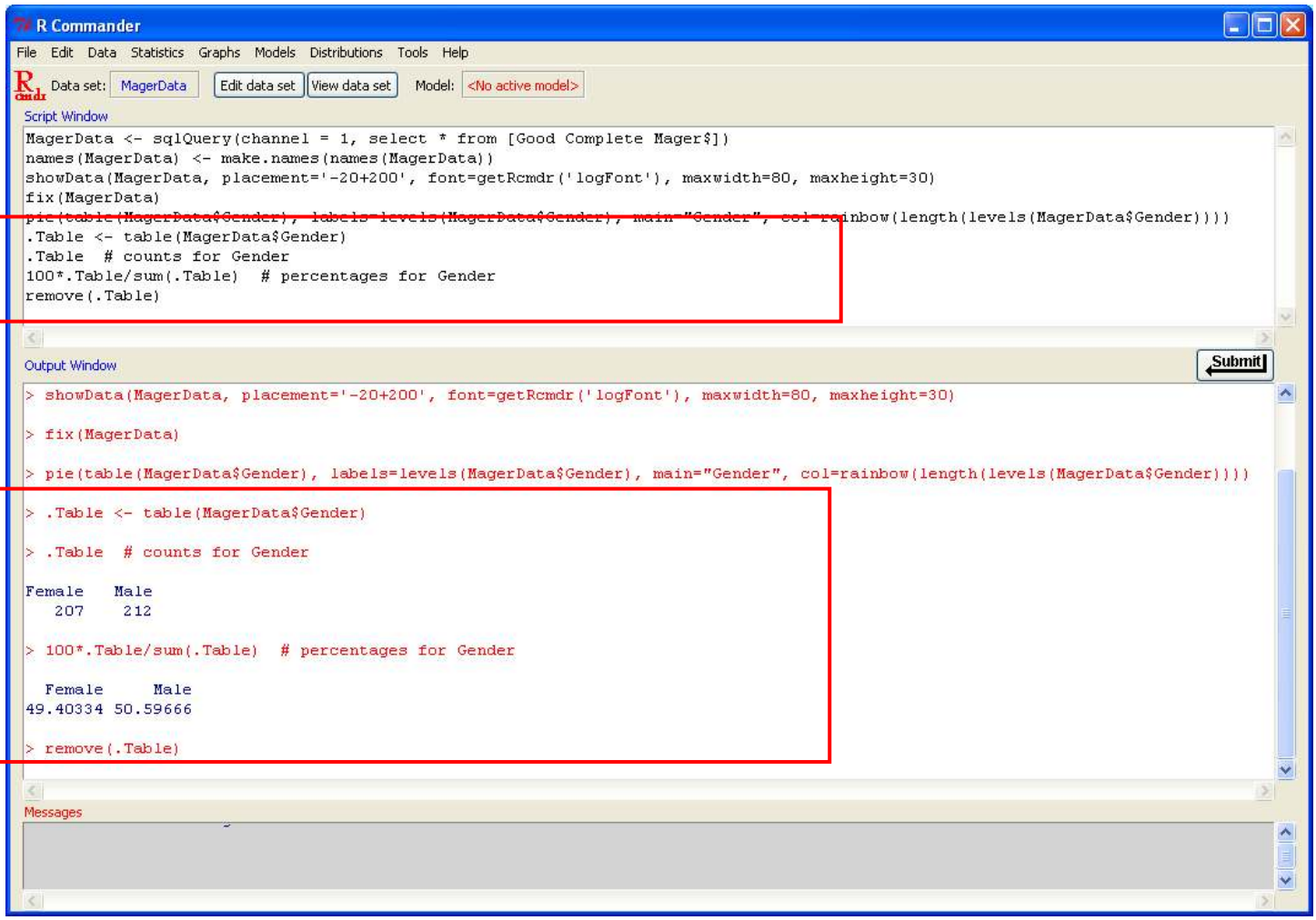


High resolution graphs will show up in the R Graphics window.

NUMERICAL Summary of CATEGORICAL Data (Counts, %, & Tables)



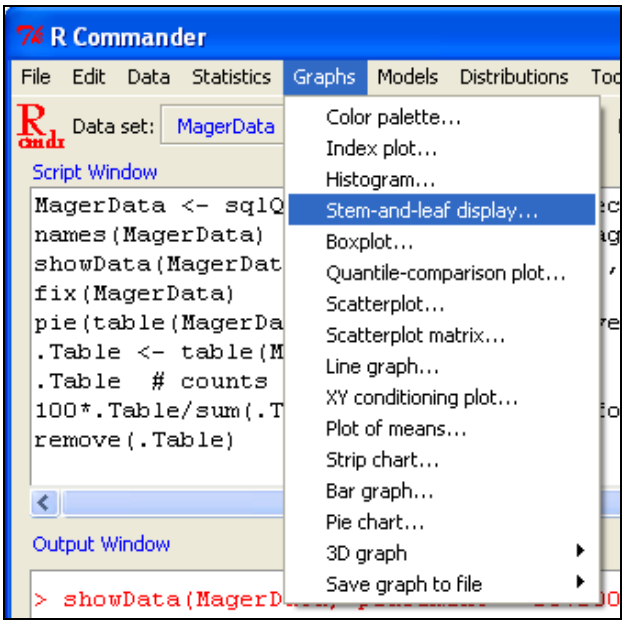


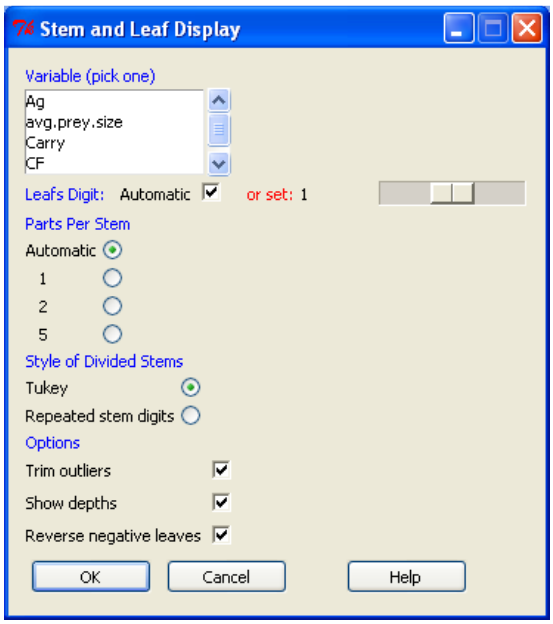


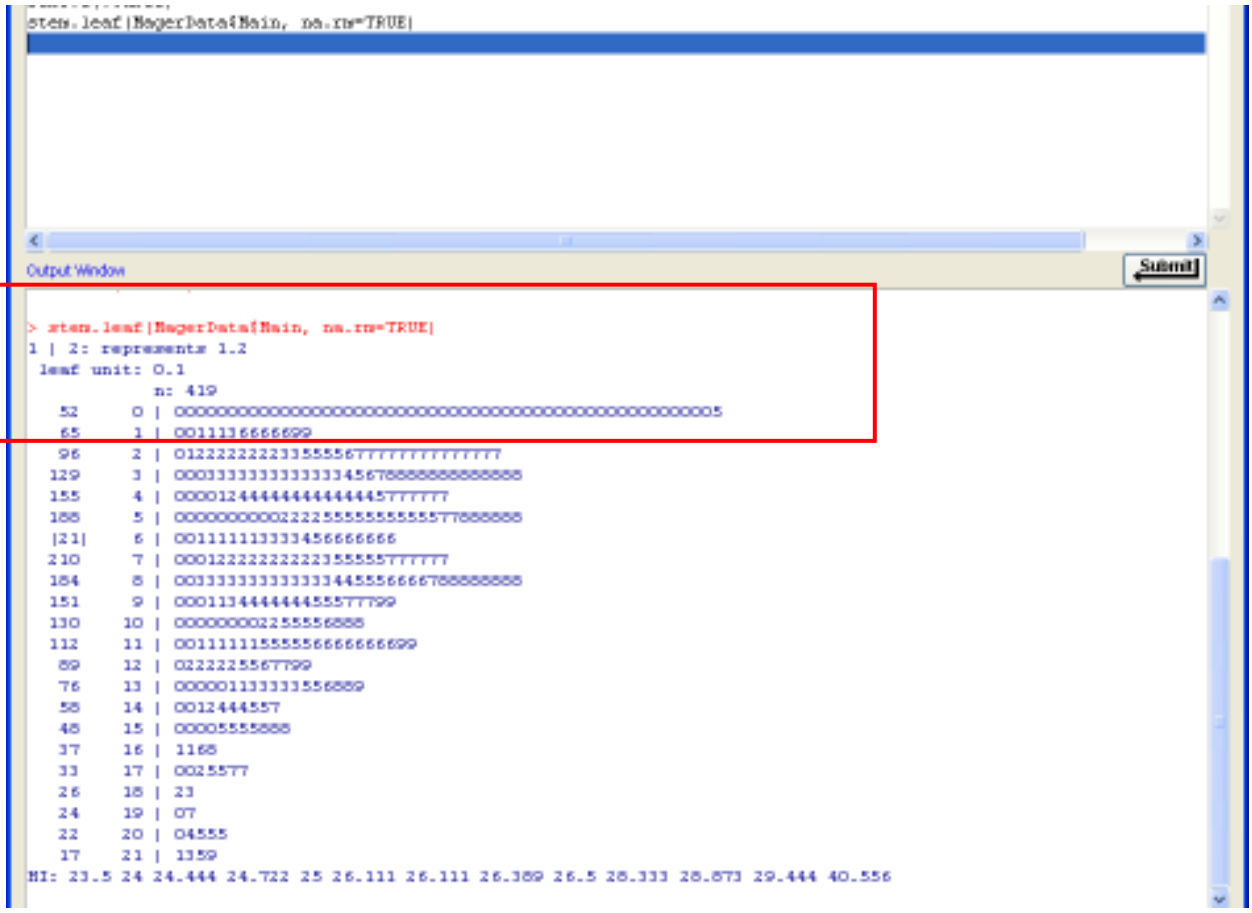
Summarizing NUMERICAL Data (Graphically and Numerically)

Graphical Summary of Numerical Data (Stems, Histogram, Boxplot…)

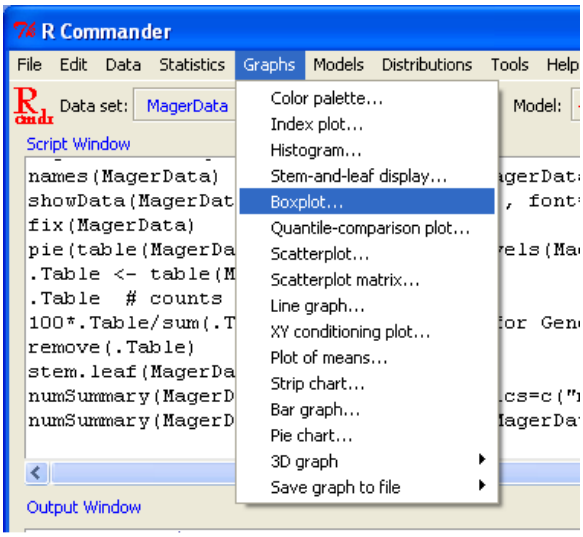
Stem and Leaf Plot

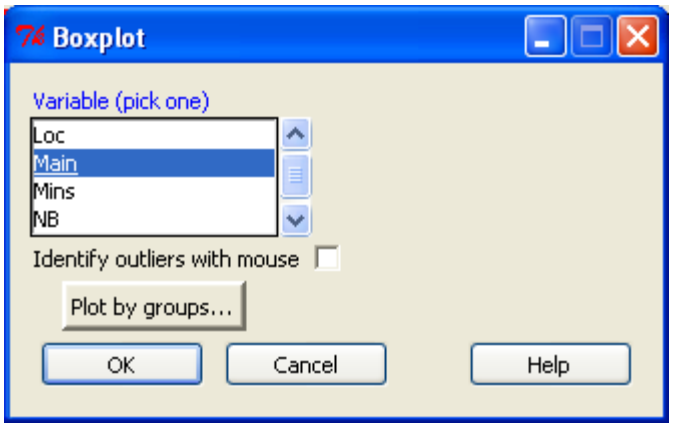


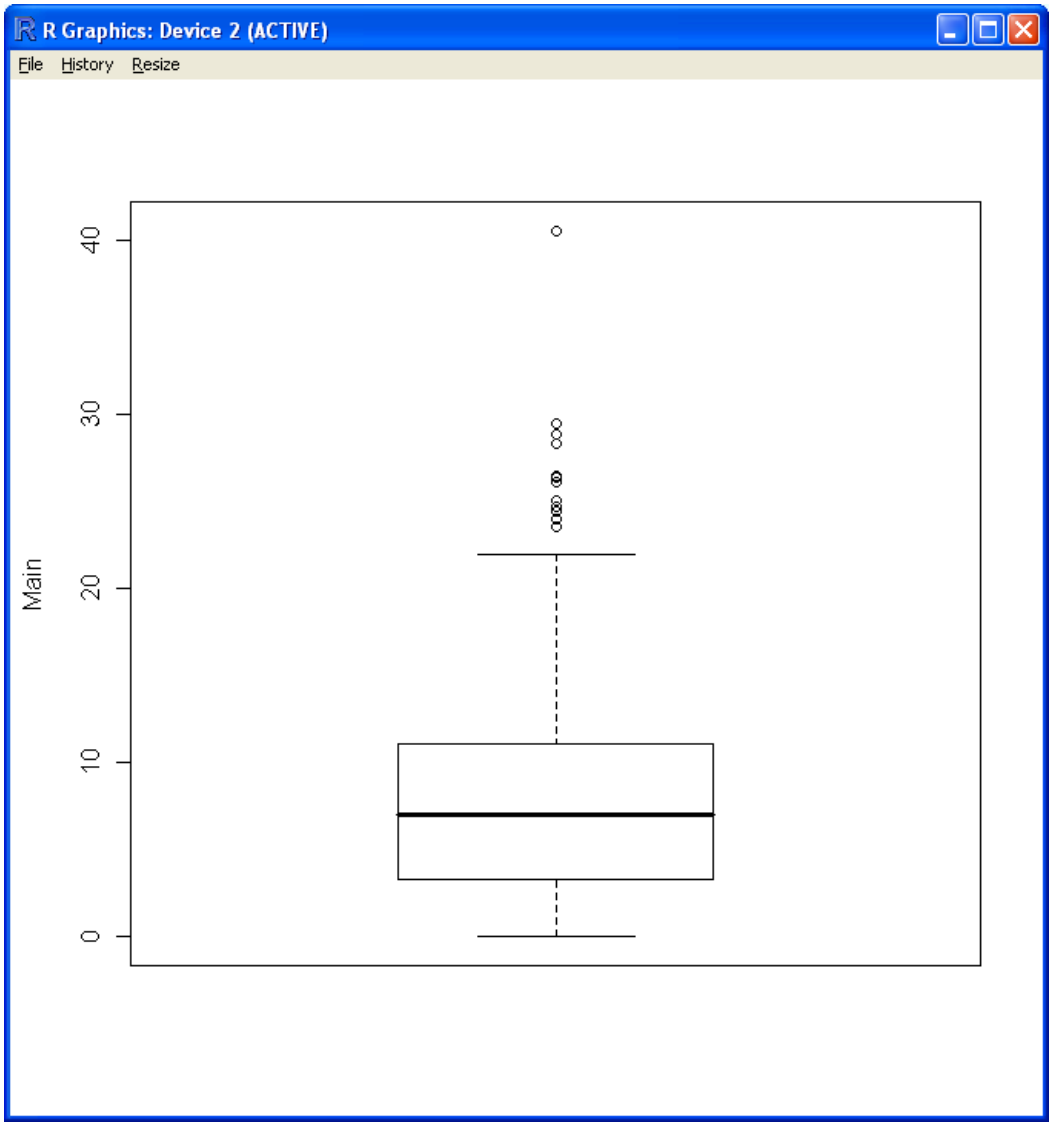




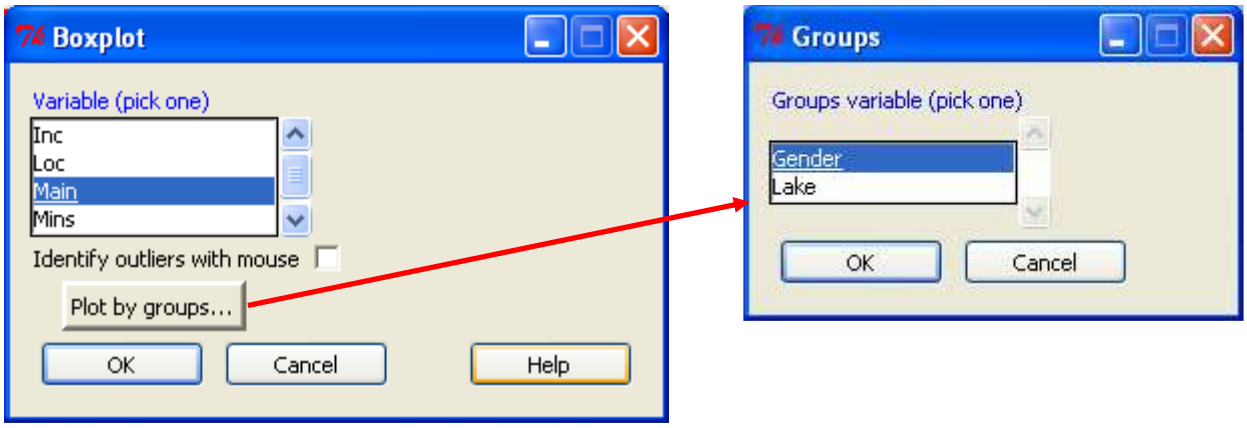
Boxplot

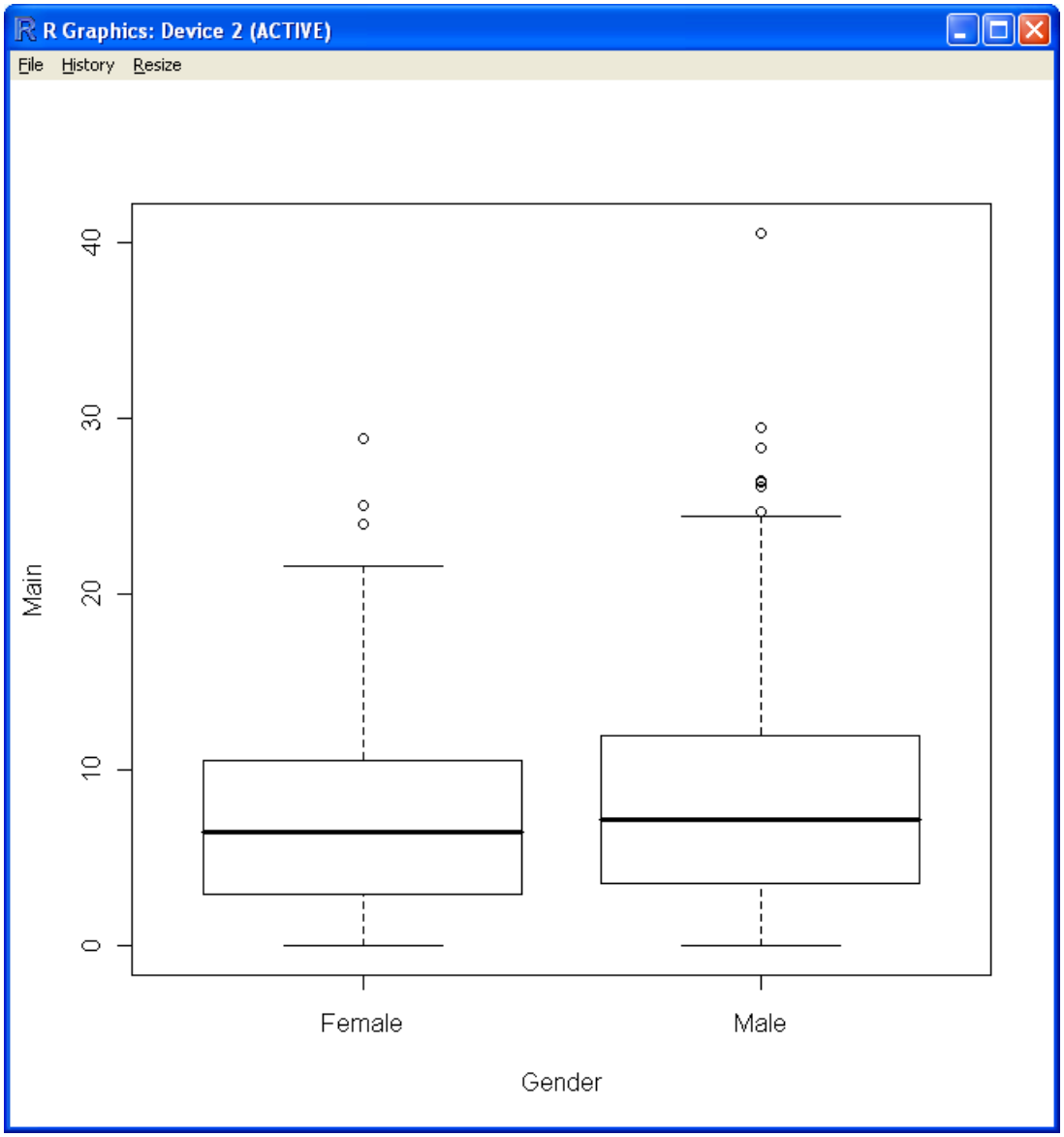

R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: MagerData
Model:
Script Window
names(MagerData)
showData(MagerDat
fix(MagerData)
pie(table(MagerDa
.Table <- table(M
.Table # counts
100*.Table/sum(.T
remove(.Table)
stem.leaf(MagerDa
numSummary(MagerD
numSummary(MagerD
Output Window
Color palette...
Index plot...
Histogram...
Stem-and-leaf display...
Boxplot...
Quantile-comparison plot...
Scatterplot...
Scatterplot matrix...
Line graph...
XY conditioning plot...
Plot of means...
Strip chart...
Bar graph...
Pie chart...
3D graph
Save graph to file


Boxplot
Variable (pick one)
Loc
Main
Mins
NB
Identify outliers with mouse
Plot by groups...
OK
Cancel
Help

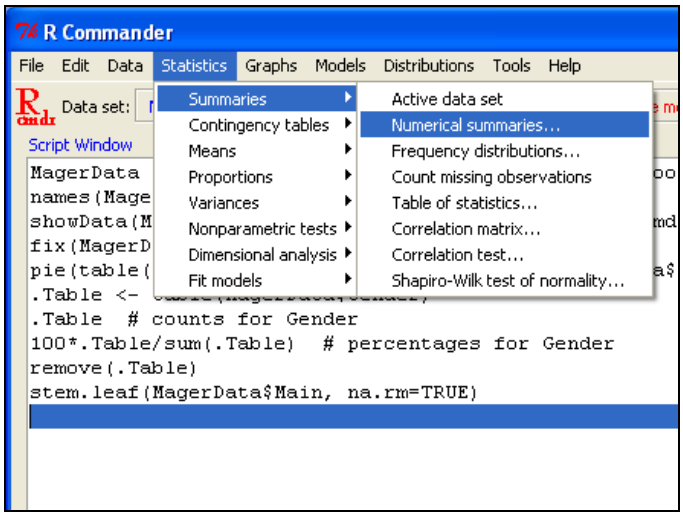

R Graphics: Device 2 (ACTIVE)
File History Resize
Main
0
10
20
30
40

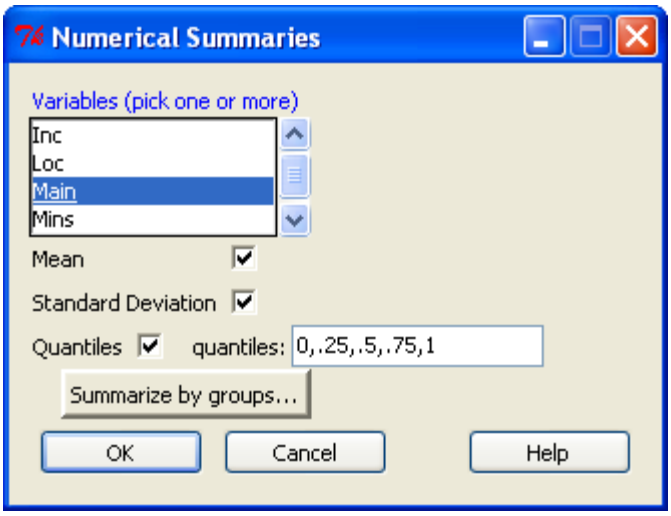
Side by Side Boxplot

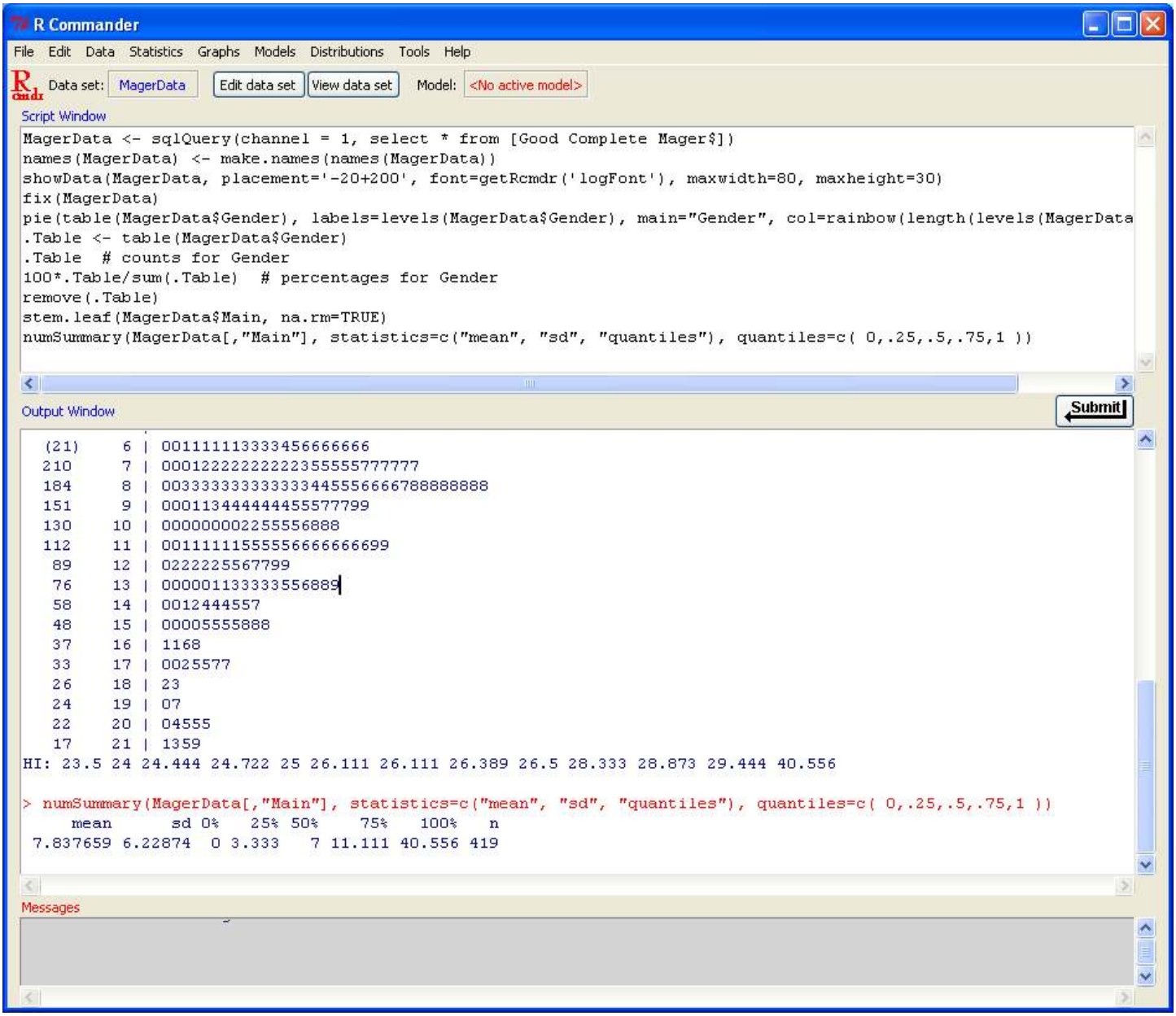




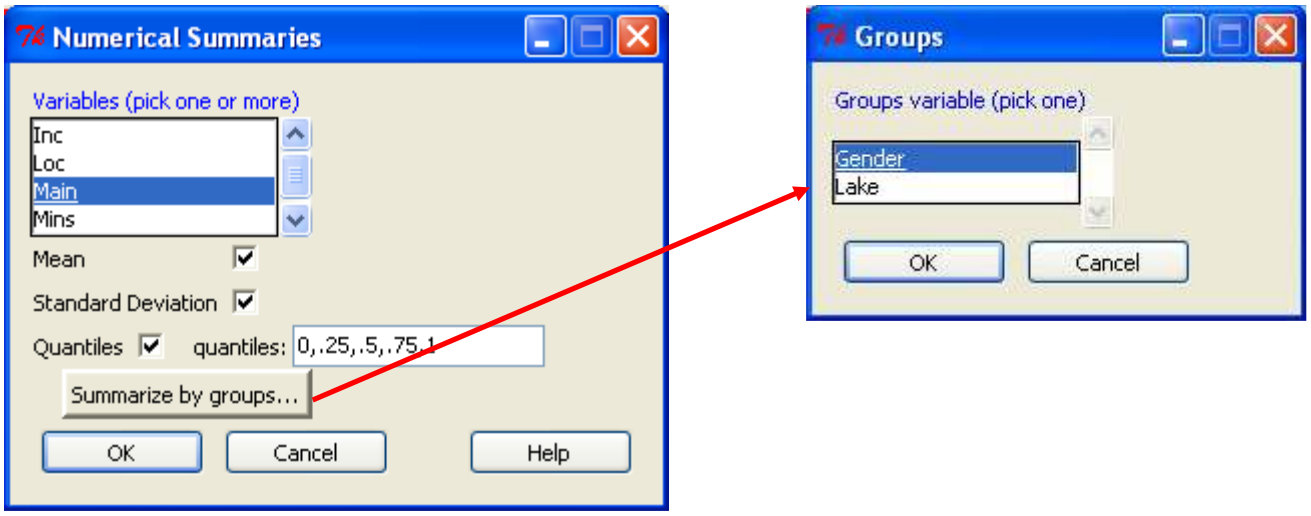
Numerical Summary of Numerical Data

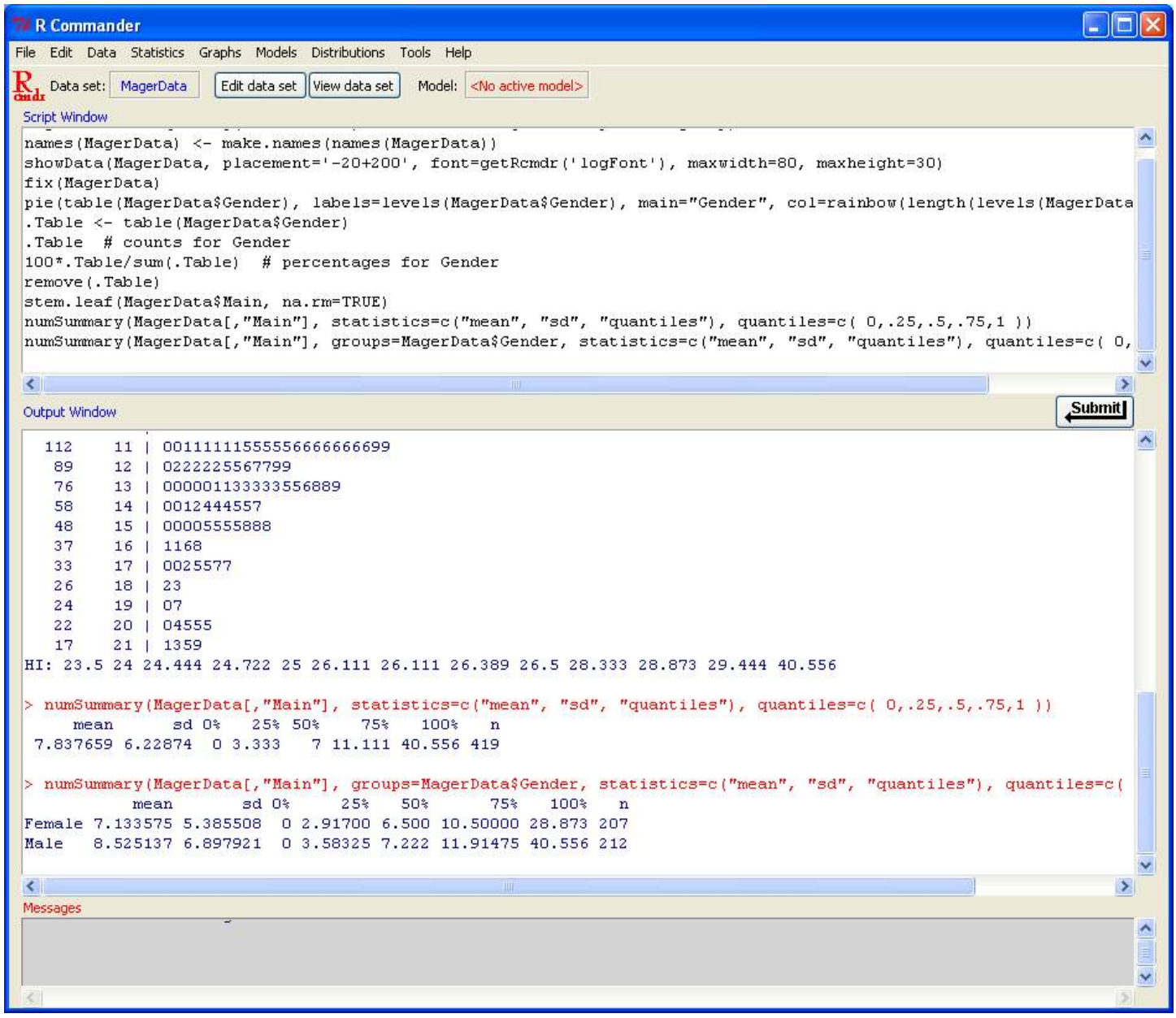






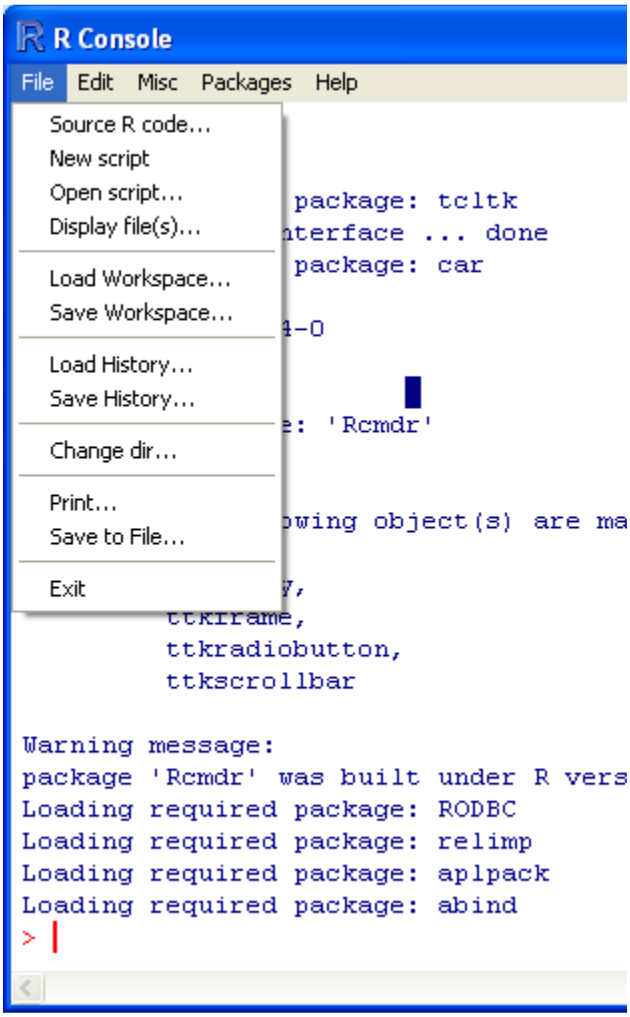
Numerical Summary of Numerical Data By a Categorical Variable





Saving Data, Commands, Output (.RData, .R, and .txt), and Graphs

Saving in the R Console



Workspaces are saved as:

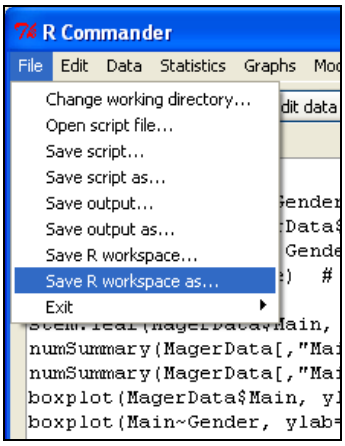
valid filename.RData

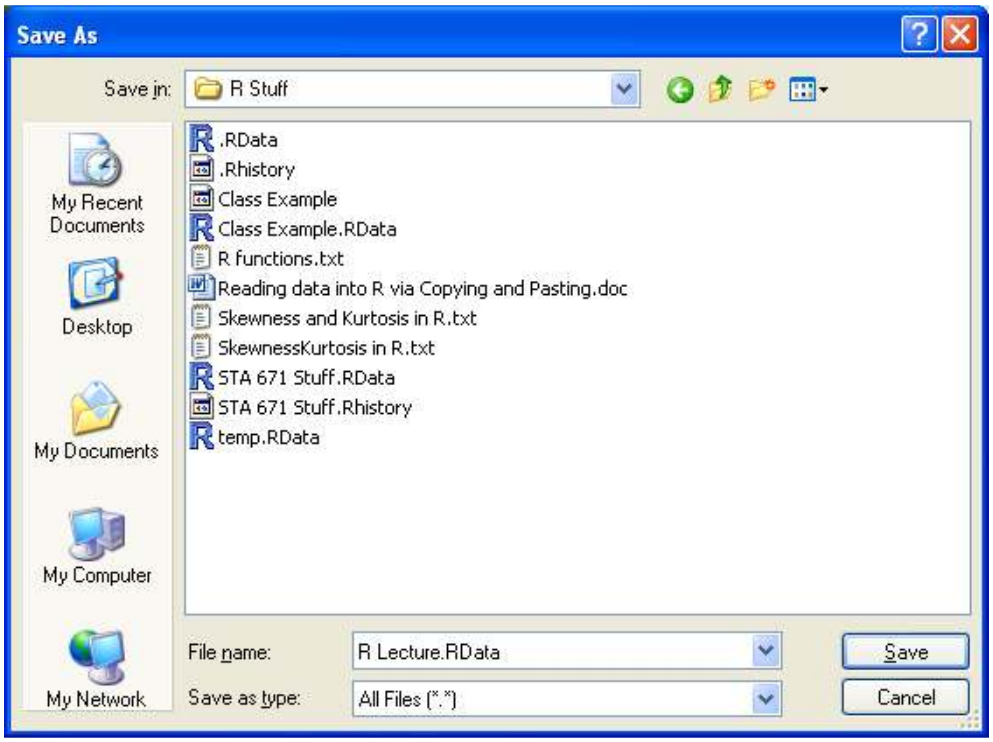
Histories are saved as:

valid filename.RHistory

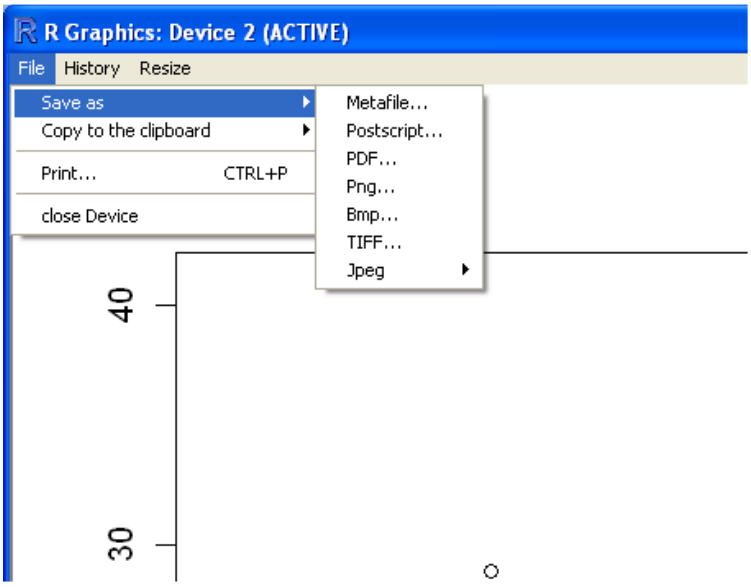
Saving in R Commander

Saving Commands (script), Output, or Data and Objects (Workspace)



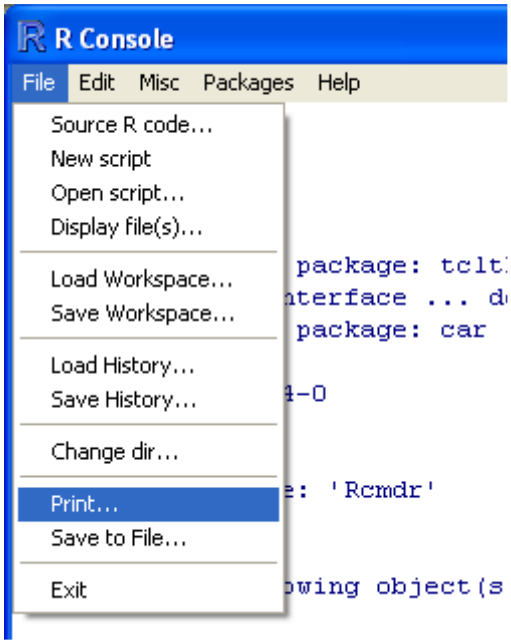


Saving in R Graphics

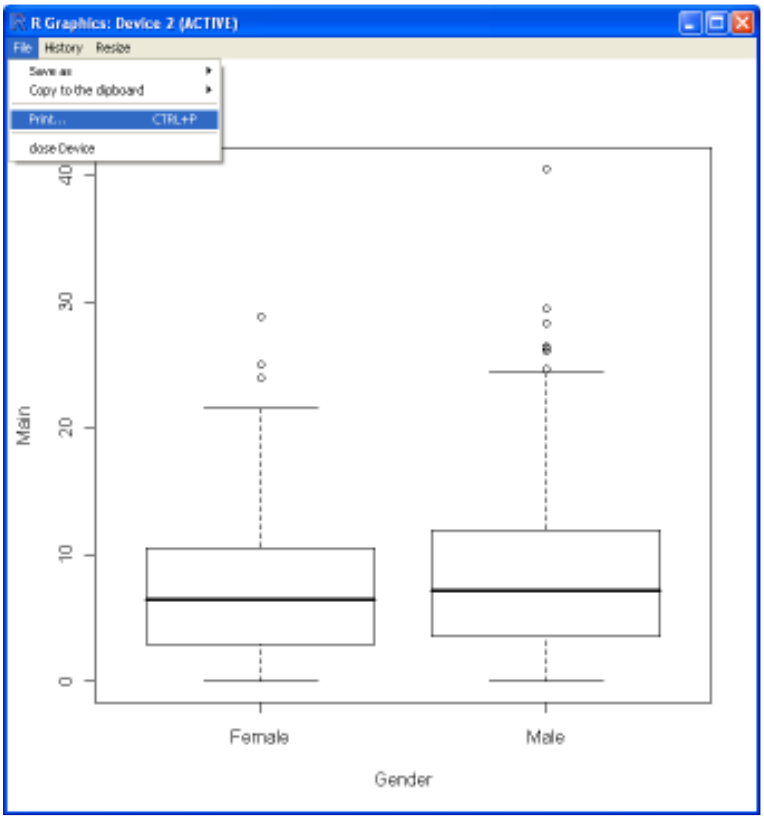


Printing

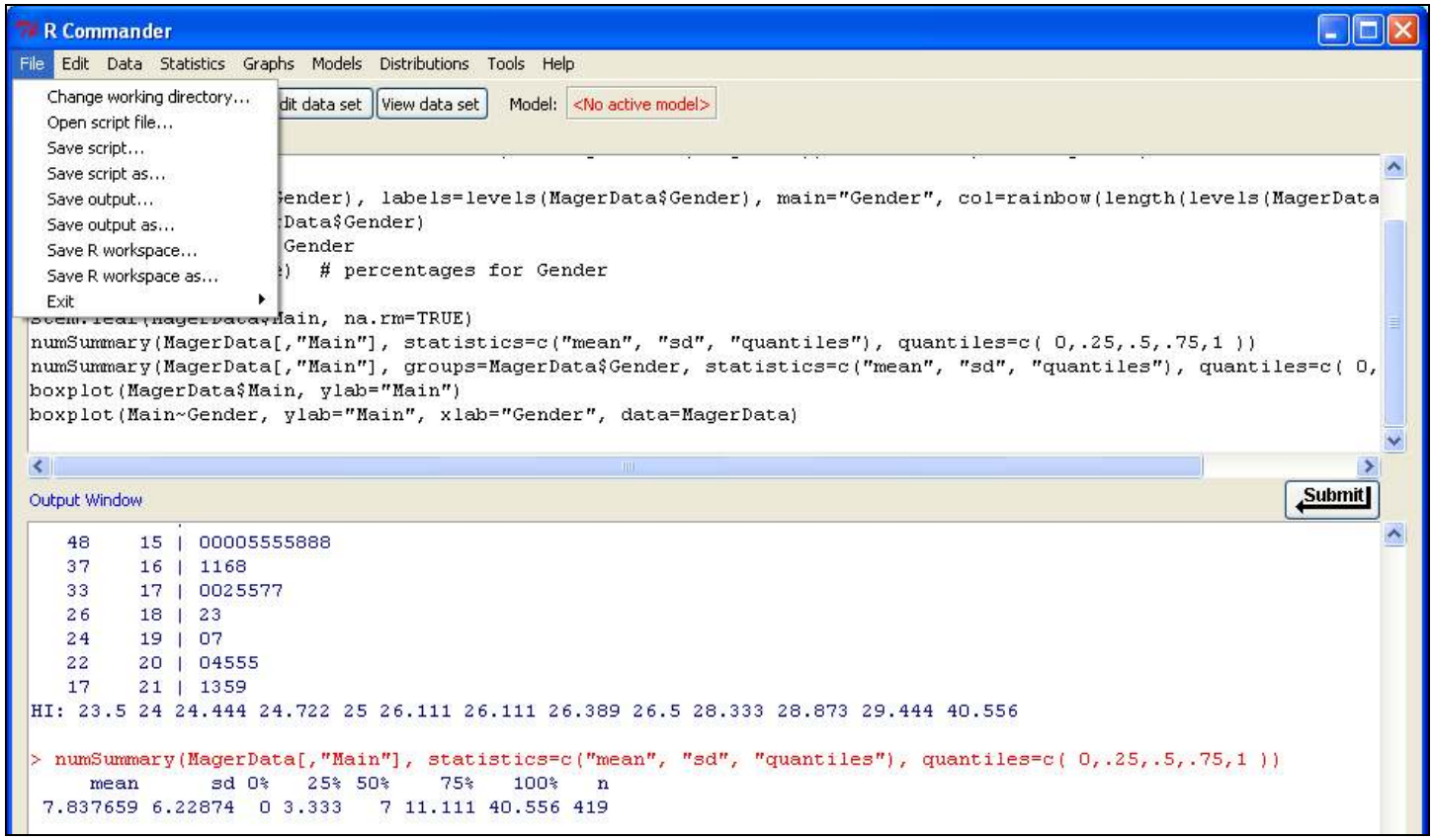
Printing in the R Console: **PRINTS EVERYTHING IN THE WINDOW or just the selected/highlighted output.**



Printing the **R Graphics Window** [OR copying contents to clipboard to paste elsewhere or save]



Printing in the **R Commander Window**



"Things" in R

Objects	Defn	Analogies
vector	• collection of elements • elements MUST be all of the same type (numeric or categorical)	• similar to a vector seen in math classes, but vectors in R are dimensionless and they don't have to be numbers
matrix	• rectangular collection of data (n rows with k columns) • elements MUST be all of the same type (numeric or categorical)	• same as a matrix seen in math classes, but now they don't have to be numbers
data.frame	• rectangular collection of data (n rows with k columns) • elements can be mixed (numeric or categorical) types	• similar to a spreadsheet in Excel
list	• collection of matrices and/or vectors of different lengths	• kind of like a wallet or purse that contains lots of dissimilar stuff, such as $, ID's, photos, …
function	• commands or operations to perform or do stuff in R • all of the R Packages are made up of functions that will do things that are not part of the base R package	• the square root key on a calculator • average or standard deviation button on a calculator • think of program written in C or Fortran

Examples

OBJECT	EXAMPLE

vector

```
> numvector = c(1, 2, 3, 4) # C() IS A "COMBINE" OPERATION IN R
> numvector
[1] 1 2 3 4
> catvector = c(joe, mary, bill)
Error: object "joe" not found
> catvector = c("joe", "mary", "bill")
> catvector
[1] "joe"  "mary" "bill"
> vector = c(1, 2, "joe", "mary")
> vector
[1] "1"    "2"    "joe"  "mary"
```

VECTOR ELEMENTS ARE ALL OF THE SAME TYPE!

matrix

```
> nummatrix = matrix (c(1, 2, 3, 4, 5, 6), nrow=3, ncol=2)
> nummatrix
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
> nummatrix = matrix (c(1, 2, 3, 4, 5, 6), nrow=3, ncol=2, byrow=FALSE)
> nummatrix
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
> nummatrix = matrix (c(1, 2, 3, 4, 5, 6), nrow=3, ncol=2, byrow=TRUE)
> nummatrix
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6

> catmatrix = matrix(c("joe", "mary", "bill", "sue"), nrow=2, ncol=2)
> catmatrix
     [,1]   [,2]
[1,] "joe"  "bill"
[2,] "mary" "sue"
> catmatrix = matrix(c("joe", "mary", "bill", "sue"), nrow=2, ncol=2,
byrow=TRUE)
> catmatrix
     [,1]   [,2]
[1,] "joe"  "mary"
[2,] "bill" "sue"
```

MATRIX ELEMENTS ARE ALL OF THE SAME TYPE!

data.frame

list

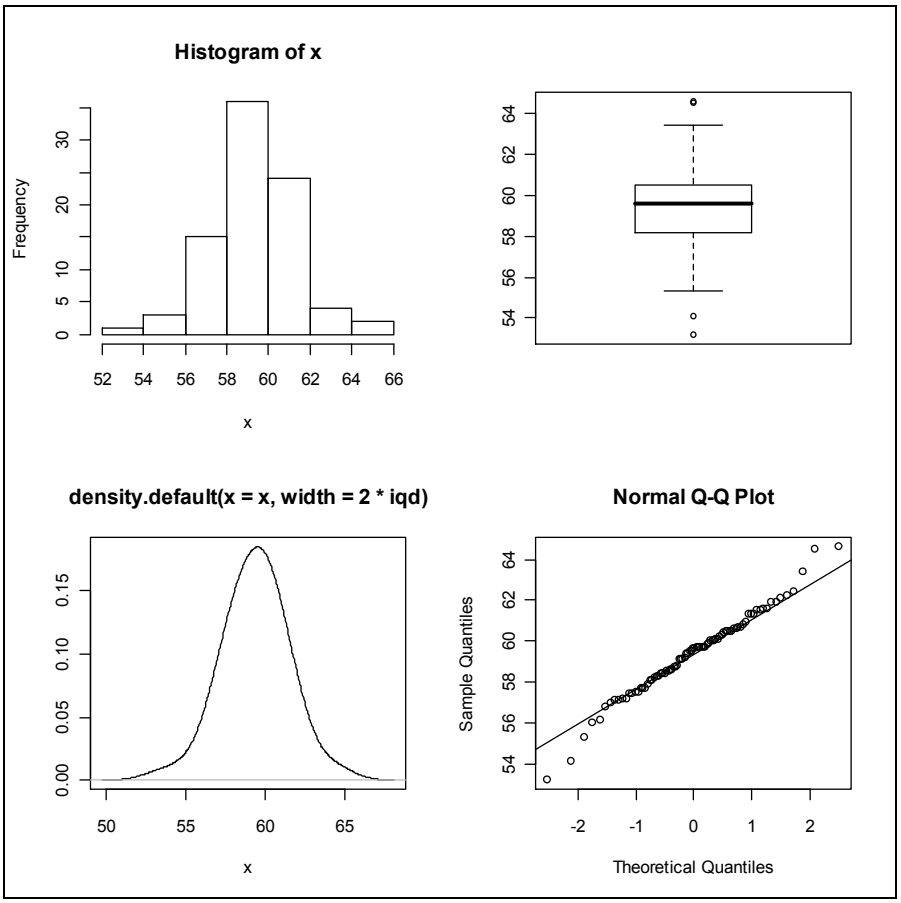
function

```
> eda.shape = function(x)
+ {
+   par(mfrow = c(2,2))
+   hist(x)
+   boxplot(x)
+   iqd <- summary(x)[5] -summary(x)[2]
+   plot(density(x,width=2*iqd),xlab="x",ylab="",type="l")
+   qqnorm(x)
+   qqline(x)
+   par(mfrow=c(1,1))
+ }
```

This function will create a Exploratory Data Analysis (EDA) display of the data in the object "x". This EDA display is not part of the base R package, but this function will take things that R does and create the display.

The EDA display of the Milky Way candy weight data would look like:

```
> eda.shape(Milky.Way$CandyWgt)
```



How to determine what Objects are in the Workspace?

In the workspace that loaded and then saved after doing the above examples, use the ls() R command to obtain a listing of all objects in the current workspace. Here's what you get:

```
> ls()
 [1] "AcornData"              "Alcohol"                "BDeehr"
 [4] "CAQuakes"               "cardata"                "catmatrix"
 [7] "catvector"              "Cheese"                 "CO2.Data"
[10] "CoalGasOilConsump"      "CoalGasPetro"           "CoalMercury"
[13] "COPolution"             "CoralHeadDensity"       "Cruise.Line"
[16] "DRusselBirdBanding"     "eda.shape"              "ex3.1"
[19] "ex3.2"                  "ex5.55"                 "ex5.57"
[22] "ex5.58"                 "Exam.Data"              "example5.17"
[25] "example5.18"            "example6.5"             "exp10"
[28] "health.care"            "JBFishData"             "JConstible"
[31] "JetNoiseLevel"          "Kolinek"                "Koper.Spider"
[34] "Mager"                  "MercuryFromFires"       "MercuryInBass"
[37] "Milky.Way"              "NewZealandRiverLengths" "NOAAWaterData"
[40] "Nuclear.Data"           "nummatrix"              "numvector"
[43] "OhioSchoolBusCrashes"   "PaperSci"               "PCBLevels"
[46] "Pottery"                "Pottery.reduced"        "RiceCloudSeeding"
[49] "SleepInMammals"         "Spider.Data"            "STA671"
[52] "Supers.Salary"          "SWOhioSpeedTraps"       "table3.1"
[55] "USArrests"              "VDSouva"                "vector"
[58] "VHorobik"               "WolfRiverPolution"
>
```

Notice that the "workspace" includes: vectors, matrices, data.frames, and functions!