

Introduction to R and R-commander

Curs d'Estadística Bàsica per a la Recerca Biomèdica

Table of contents

- ▮ **Introduction to R (and S)**
- ▮ **Introduction to Rcmdr**
- ▮ **The Rcmdr menus**
- ▮ **Loading data sets**
- ▮ **Working with data sets**
- ▮ **Using R scripts**
- ▮ **Exporting results**
- ▮ **Additional sources**
- ▮ **Practice**

- R (and S) is a programming language and environment for **data exploration and manipulation, statistical computing** and **high-quality graphics**.
- Currently, S and R are the two most widely used languages in statistical research.
- S was originally developed at Bell Laboratories by John Chambers in 1976 and is a commercial language (S-Plus).
- R ("GNU" S) was implemented in 1993 by **Robert Gentleman** and **Ross Ihaka** as a free alternative to S.
 - It is currently developed and maintained by the *R Development Core Team*, a team of volunteer developers from across the globe including well-known statisticians (eg, J. Chambers, L. Terney, B. Ripley, D. Bates, etc.).
 - It is available from the R-Project website:

<http://www.r-project.org/>



Advantages

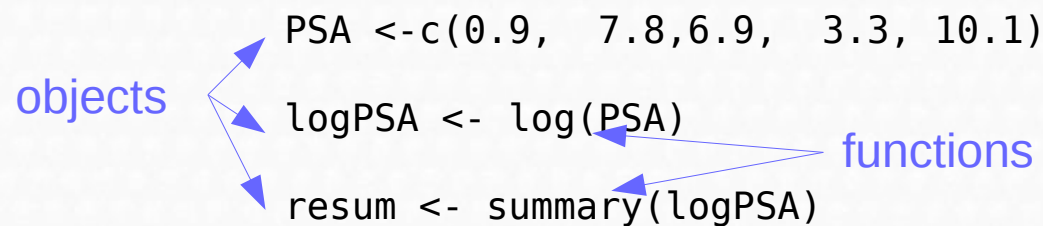
- It is free
- Multi-platform (Linux, Mac, Windows)
- Powerful in graphics generation
- Powerful statistical tool (top statistical methods)
- Is always growing in users and functionalities → Frequent updates
- Flexible, open source programming language
- Useful for repetitive tasks

Drawbacks

- It is a programming language (though incorporates GUIs as Rcmdr, Rstudio)
- Not so user friendly as desired
- Supporting documentation is of variable quality
- Frequent updates
- Works better with internet connection

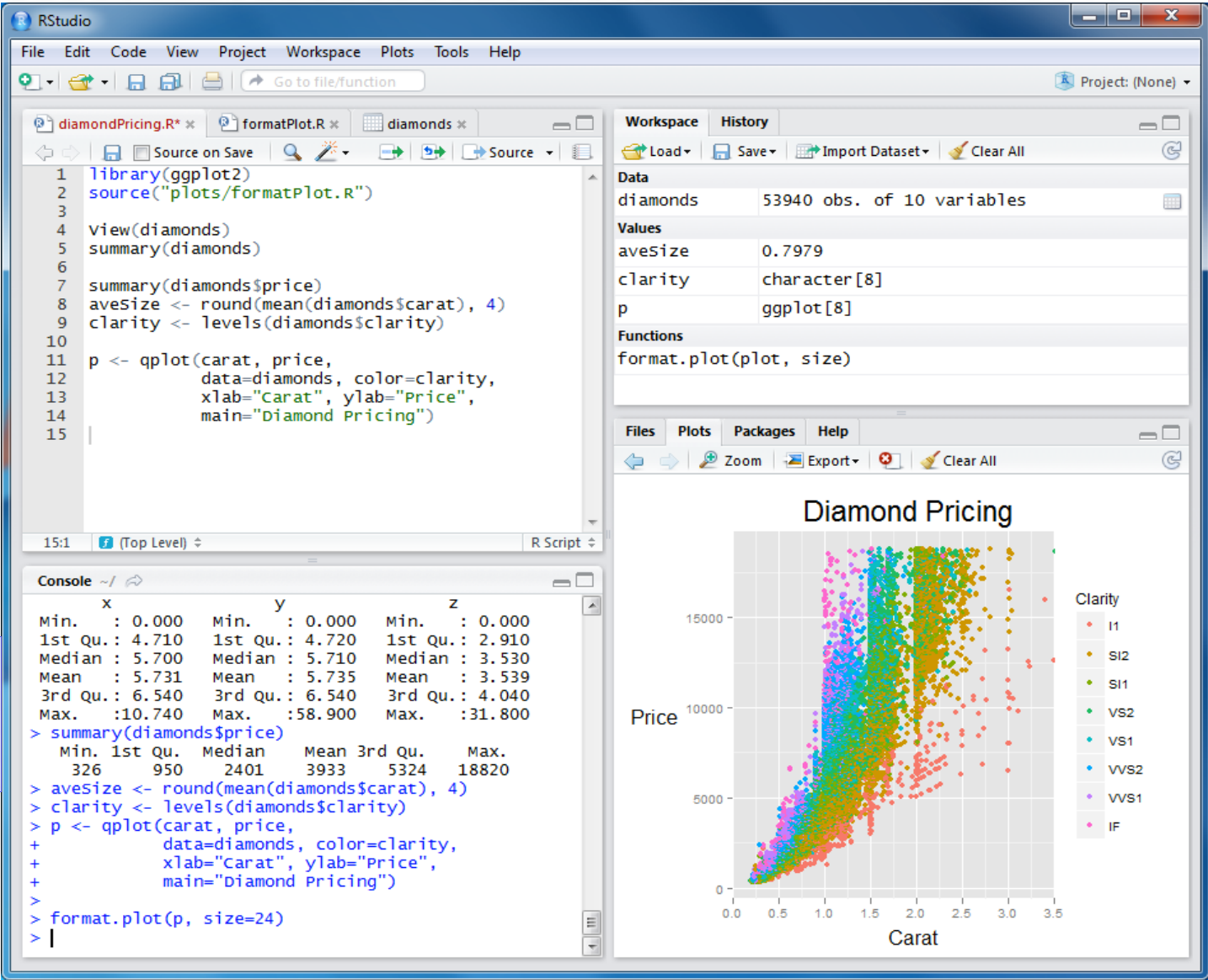
How it works

- R consists of a base system and additional packages that extend its functionalities.
- Language based on objects and functions



- Different interfaces:
 - R console
 - **RStudio**: Simplifies the use of R through a window system that gives access to code, console, data, graphics or help in a powerful and intuitive way
 - **RCommander**: allows the user to run basic functions using a menus interface. Offers a great way to begin familiarizing yourself with R and statistics within a GUI environment.

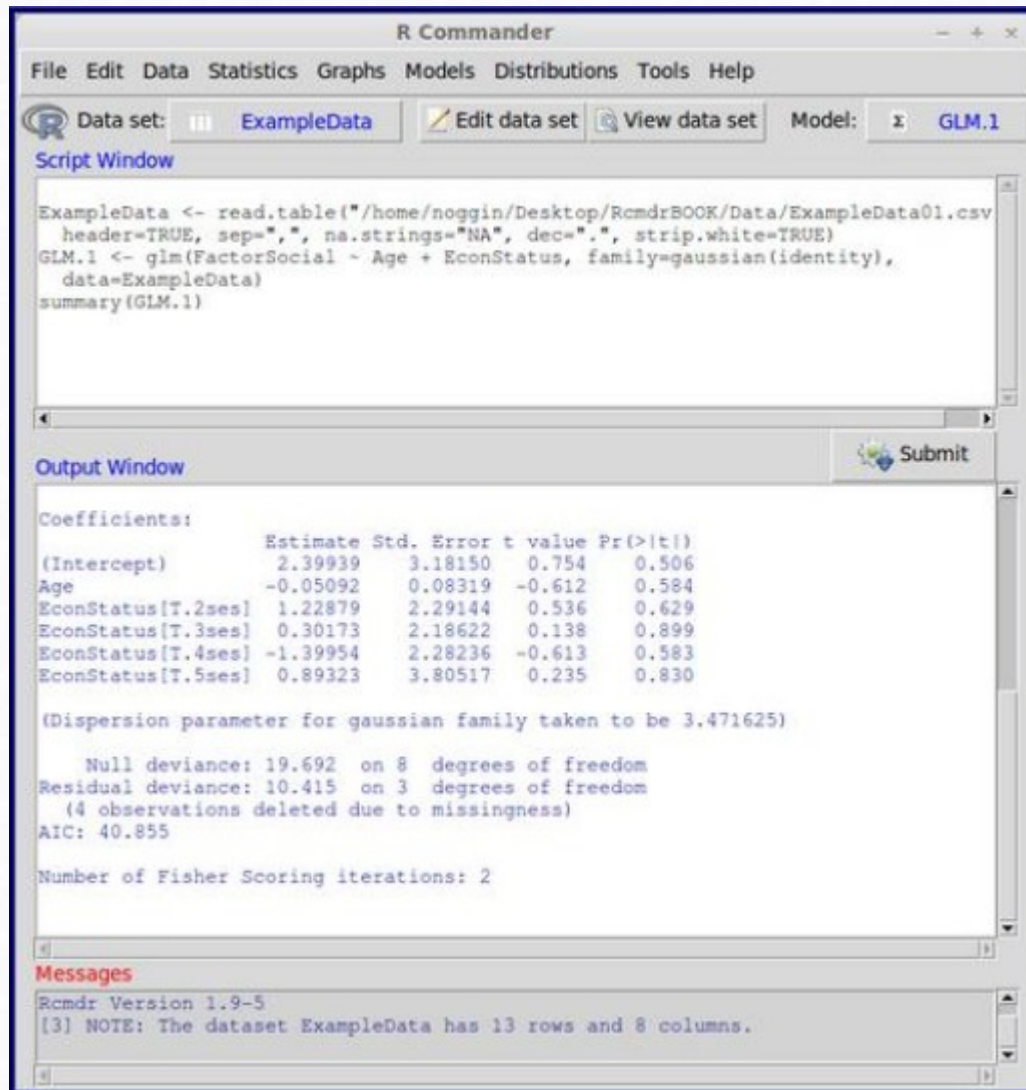
Source
-scripts
-text edit



**Environment
& History**

Console
-commands
-output

**Files, plots,
packages,
help**



The screenshot shows the R Commander application window. The top menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu bar, there are buttons for Data set (ExampleData), Edit data set, View data set, and Model (GLM.1). The Script Window contains the following R code:

```
ExampleData <- read.table("/home/noggin/Desktop/RcmdrBOOK/Data/ExampleData01.csv",
  header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
GLM.1 <- glm(FactorSocial ~ Age + EconStatus, family=gaussian(identity),
  data=ExampleData)
summary(GLM.1)
```

The Output Window displays the results of the GLM model:

```
Coefficients:
(Intercept)      2.39939    3.18150    0.754    0.506
Age             -0.05092    0.08319   -0.612    0.584
EconStatus[T.2ses] 1.22879    2.29144    0.536    0.629
EconStatus[T.3ses] 0.30173    2.18622    0.138    0.899
EconStatus[T.4ses] -1.39954    2.28236   -0.613    0.583
EconStatus[T.5ses] 0.89323    3.80517    0.235    0.830

(Dispersion parameter for gaussian family taken to be 3.471625)

Null deviance: 19.692  on 8  degrees of freedom
Residual deviance: 10.415  on 3  degrees of freedom
(4 observations deleted due to missingness)
AIC: 40.855

Number of Fisher Scoring iterations: 2
```

The Messages window at the bottom shows the following messages:

```
Rcmdr Version 1.9-5
[3] NOTE: The dataset ExampleData has 13 rows and 8 columns.
```

- Package to install after having R.
- Developed by a statistician, John Fox, to ease the use of R.
- Inspired in SPSS software.
- Menus accessible to use directly the functions and avoid the command line programming.
- Graphics are generated in the R interface.
- There are plug-ins with specific functionalities

Rcommander:

<http://www.rcommander.com/>

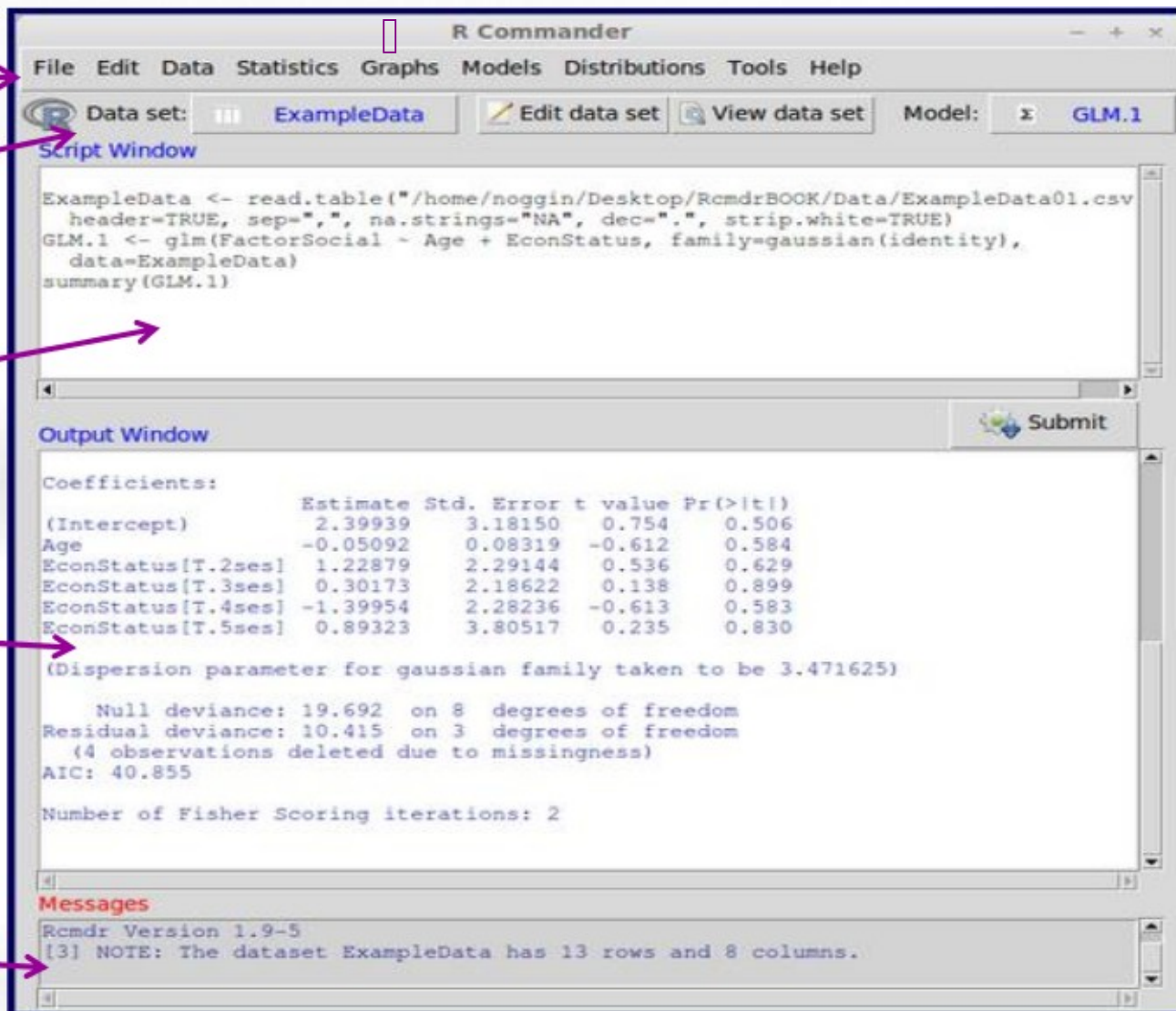
Menu bar

Tool bar

Command window

Text output window

Messages window



The screenshot shows the RCommander application window. The menu bar at the top includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu bar is a toolbar with buttons for Data set (ExampleData), Edit data set, and View data set. The main window is divided into three panes: a Script Window at the top containing R code, an Output Window in the middle displaying the results of the code execution, and a Messages window at the bottom showing status messages.

Script Window

```
ExampleData <- read.table("/home/noggin/Desktop/RcmdrBOOK/Data/ExampleData01.csv",
  header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
GLM.1 <- glm(FactorSocial ~ Age + EconStatus, family=gaussian(identity),
  data=ExampleData)
summary(GLM.1)
```

Output Window

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.39939	3.18150	0.754	0.506
Age	-0.05092	0.08319	-0.612	0.584
EconStatus[T.2ses]	1.22879	2.29144	0.536	0.629
EconStatus[T.3ses]	0.30173	2.18622	0.138	0.899
EconStatus[T.4ses]	-1.39954	2.28236	-0.613	0.583
EconStatus[T.5ses]	0.89323	3.80517	0.235	0.830

(Dispersion parameter for gaussian family taken to be 3.471625)

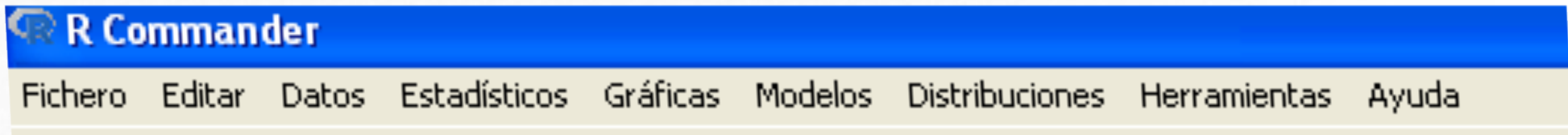
Null deviance: 19.692 on 8 degrees of freedom
Residual deviance: 10.415 on 3 degrees of freedom
(4 observations deleted due to missingness)
AIC: 40.855

Number of Fisher Scoring iterations: 2

Messages

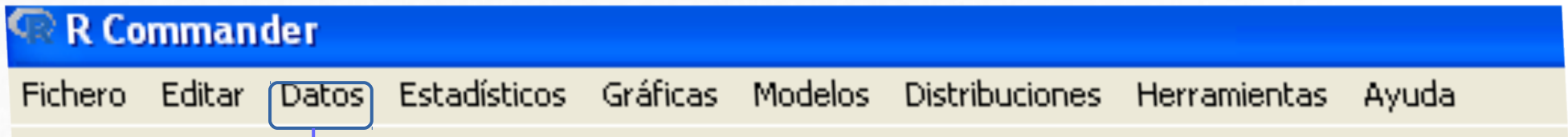
Rcmdr Version 1.9-5
[3] NOTE: The dataset ExampleData has 13 rows and 8 columns.

Menu bar



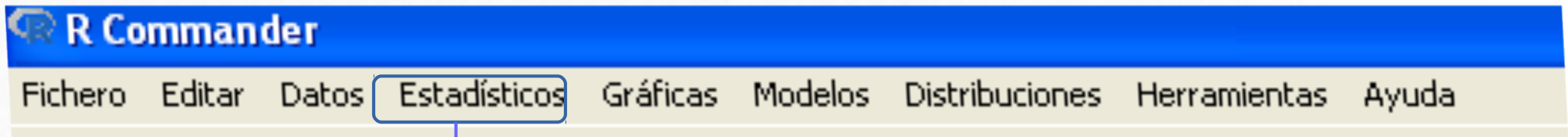
- **File:** contains options to load and save files, define settings and exit.
- **Edit:** options for editing output and log/script window contents.
- **Data:** options to read and modify data.
- **Statistics:** submenu containing options for basic statistical analysis
- **Graphs:** contains options for creating simple statistical graphs
- **Models:** options for obtaining numerical summaries, testing hypotheses and regression models.
- **Distributions:** options to calculate probabilities, obtain quantiles, and get plots of already known statistical distributions.
- **Help:** contains menus with info about how to work with R commander.

Menu bar



- Create / Load data
- Editing and inspection of data files.
- Data transformation / Creation of new variables.
- Selection of subsets of data or subgroups of variables.
- Conversion of numerical variables into factors.

Menu bar



- Summaries
- Contingency tables
- Medians
- Proportions
- Variants
- Non-parametric tests
- Dimensional analysis (A. Multivariate)
- Model adjustment (Regression)

Tool bar



- The program always works with a main set of data (active dataset).
- With the active "dataset" we can:
 - edit or visualize it
 - Do analysis
 - Build and use models
- At any moment we can change the active dataset.

Command window

R Script

R Markdown

```
Dataset <- read.table("E:/BioStatFLOSS/dades/osteoporosis.csv", header=TRUE,  
  sep="\t", na.strings="NA", dec="," , strip.white=TRUE)  
library(abind, pos=14)  
library(e1071, pos=15)  
summary(Dataset)
```

- The menu actions are converted into instructions in the Command window.

Text output window

Salida Ejecutar

```
> Dataset <- read.table("E:/BioStatFLOSS/dades/osteoporosis.csv", header=TRUE,
+   sep="\t", na.strings="NA", dec=".", strip.white=TRUE)

> library(abind, pos=14)

> library(e1071, pos=15)

> summary(Dataset)
```

registro	area	f_nac	edad	grupedad	peso	talla	imc	bua
Min. : 3.0	Min. :10.00	11509689600: 3	Min. :45.00	45 - 49:378	Min. : 44.00	Min. :138.0	Min. :17.21	Min. : 11.0
1st Qu.: 280.8	1st Qu.:10.00	11718518400: 3	1st Qu.:48.00	50 - 54:233	1st Qu.: 60.50	1st Qu.:153.0	1st Qu.:24.80	1st Qu.: 62.0
Median : 531.5	Median :11.00	11010297600: 2	Median :52.00	55 - 59:176	Median : 68.00	Median :157.0	Median :27.51	Median : 72.0
Mean : 529.9	Mean :11.58	11090822400: 2	Mean :53.42	60 - 64:129	Mean : 69.12	Mean :156.9	Mean :28.11	Mean : 73.3
3rd Qu.: 781.2	3rd Qu.:13.00	11098166400: 2	3rd Qu.:58.00	65 - 69: 84	3rd Qu.: 75.00	3rd Qu.:161.0	3rd Qu.:30.82	3rd Qu.: 84.0
Max. :1033.0	Max. :13.00	11181283200: 2	Max. :69.00		Max. :123.50	Max. :180.0	Max. :48.39	Max. :136.0

(Other) :986

clasific	menarqui	edad_men	menop	tipo_men	nivel_ed
NORMAL :469	Min. : 8.00	Min. :24.00	NO:303	AMBAS : 79	PRIMARIOS :467
OSTEOPENIA :467	1st Qu.:12.00	1st Qu.:46.00	SI:697	HISTERECTOMIA : 63	PRIMARIOS SIN FINALIZAR:212
OSTEOPOROSIS: 64	Median :13.00	Median :51.00		NATURAL :544	SECUNDARIOS :150
	Mean :12.71	Mean :63.04		NO MENOPAUSIA/NO CONSTA:303	SIN ESTUDIOS :122
	3rd Qu.:14.00	3rd Qu.:99.00		OVARIECTOMIA : 11	SUPERIORES : 49
	Max. :17.00	Max. :99.00			

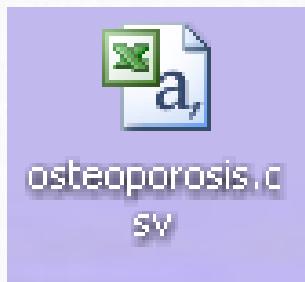
Message window

Mensajes

```
[1] NOTA: Versión de R Commander 2.3-1: Thu Jan 26 08:42:06 2017
[2] NOTA: El conjunto de datos Dataset tiene 1000 filas y 15 columnas.
```

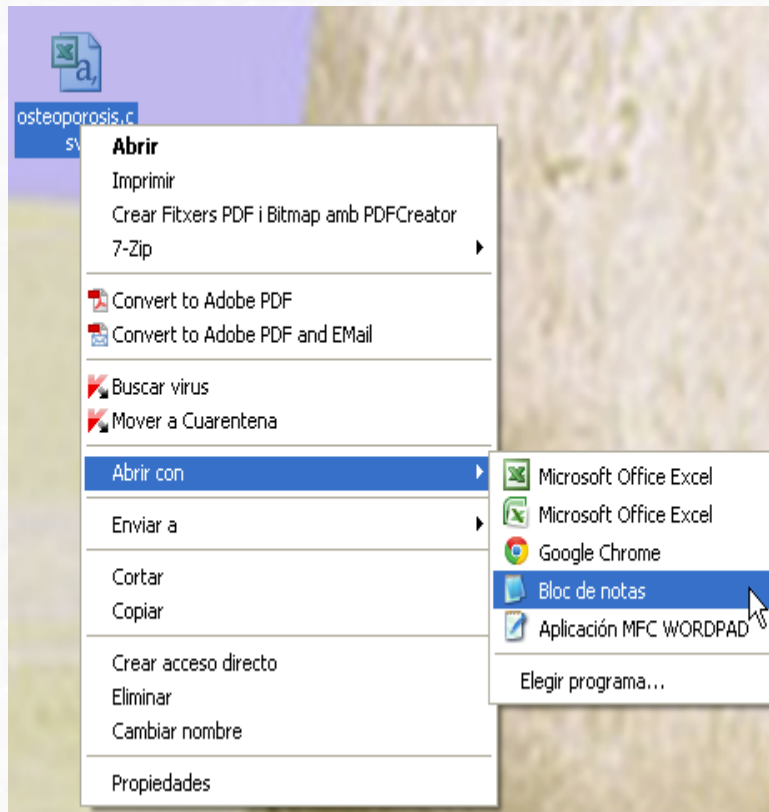

We have to load into Rcommander the dataset *osteoporosis.csv*

1. “.csv” -> “comma separated values” but could be separated with \tab, spaces, ; ,
2. First of all “look” the file with a file viewer like “wordpad”, “bloc de notas”,...



Practice. Loading data sets

We have to load into Rcommander the dataset *osteoporosis.csv*



The image shows a Notepad window titled 'osteoporosis.csv - Bloc de notas' displaying the contents of the CSV file. The data is organized into columns with headers: 'registro', 'area', 'f_nac', 'edad', 'grupedad', 'peso', and 'talla'. The first row of data is highlighted, and a red arrow points to the 'f_nac' column header.

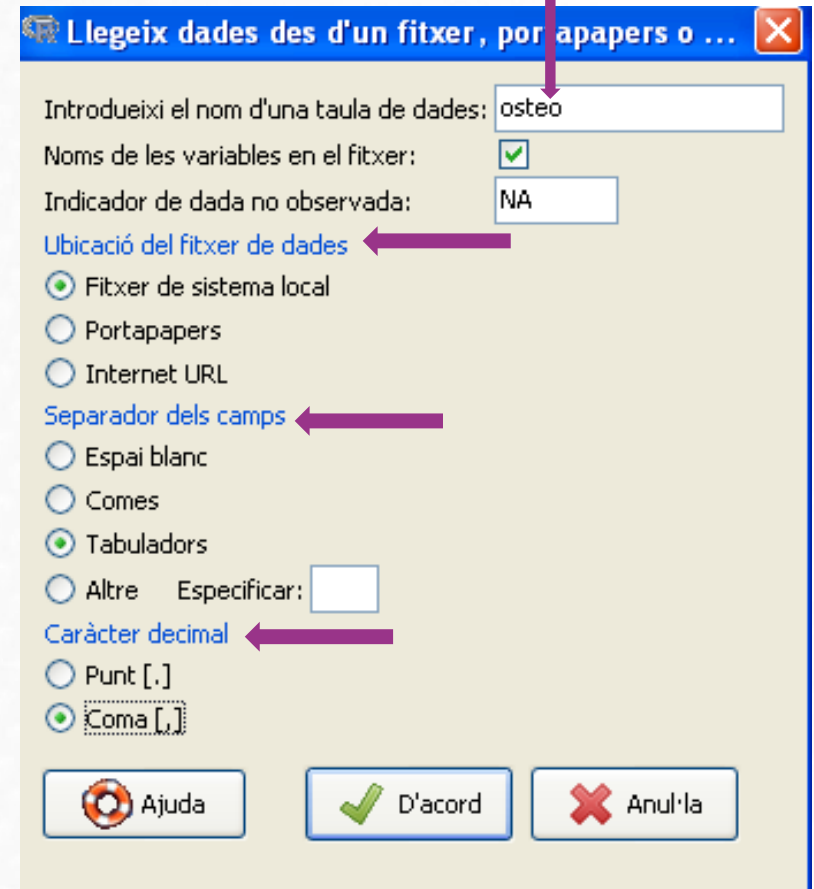
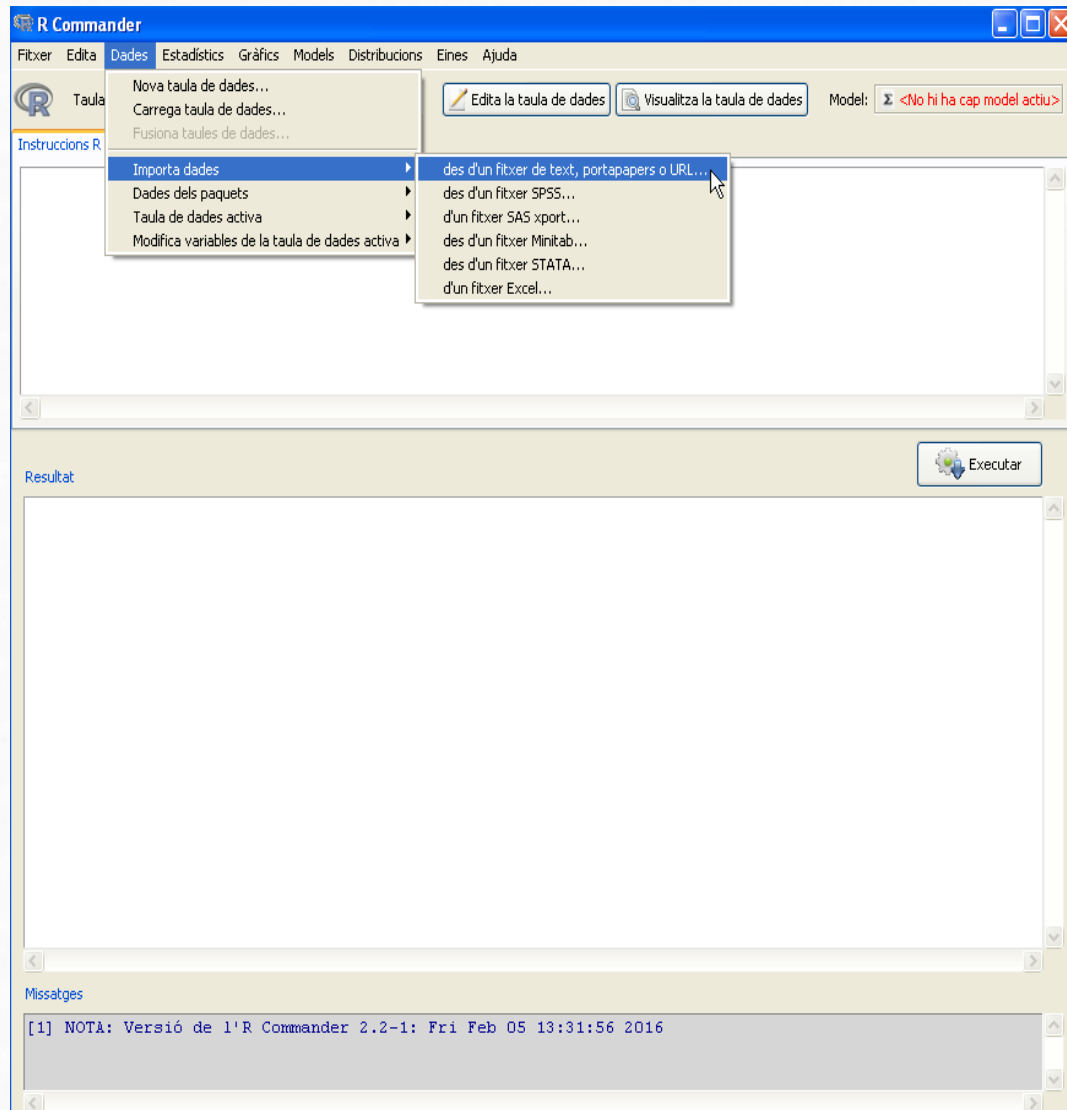
registro	area	f_nac	edad	grupedad	peso	talla
8	"OSTEOPENIA"	14	99	"NO"	"NO MENOPAUSIA/NO CONSTA"	"NORMAL"
55	"55 - 59"	75	161	28,93	92	"NORMAL"
"SI"	"NATURAL"	"PRIMARIOS"	45	10	11	29651200
NDARIOS"	57	10	11483164800	52	"5 - 54"	60
5	159	22,74	61	"OSTEOPENIA"	14	46
ORMAL"	13	99	"NO"	"NO MENOPAUSIA/NO CONSTA"	"PRIMARI"	
10	11730873600	45	"45 - 49"	46,5	160	18,16
NSTA"	"SECUNDARIOS"	104	11	11516083200	51	"50 - 54"
5	13	11226902400	61	"60 - 64"	75	161
21,41	68	"OSTEOPENIA"	13	44	"SI"	"NATURAL"
14	48	"SI"	"NATURAL"	"PRIMARIOS"	136	13
24,8	61	"OSTEOPENIA"	11	43	"SI"	"NATURAL"
POROSIS"	12	50	"SI"	"NATURAL"	"PRIMARIOS"	163
0	63	"60 - 64"	85,5	153	36,52	60
59"	57	170	19,72	90	"NORMAL"	16
1,98	67	"OSTEOPENIA"	11	99	"NO"	"NO MENOPAUSIA/N"
153	40,8	53	"OSTEOPENIA"	13	53	"SI"
OPENIA"	12	52	"SI"	"NATURAL"	"PRIMARIOS"	215
129600	55	"55 - 59"	67	151	29,38	80
"SI"	"NATURAL"	"PRIMARIOS"	236	13	11273644800	59
ZAR"	246	13	11388729600	56	"55 - 59"	62,5
N FINALIZAR"	256	13	10943510400	69	"65 - 69"	57,2
"60 - 64"	79	158	31,65	69	"OSTEOPENIA"	13
1,85	111	"NORMAL"	13	49	"SI"	"NATURAL"
7	154	24,03	51	"OSTEOPENIA"	12	47
NATURAL"	"SIN ESTUDIOS"	297	10	11730355200	45	
"OSTEOPENIA"	11	48	"SI"	"NATURAL"	"PRIMARIOS"	307
O MENOPAUSIA/NO CONSTA"	"PRIMARIOS"	317	10	11577513600	50	
63	"OSTEOPENIA"	14	51	"SI"	"NATURAL"	"SECUNDA"
49	"SI"	"NATURAL"	"PRIMARIOS"	337	13	11824272000
"	"PRIMARIOS"	347	13	11569132800	50	"50 - 54"
600	54	"50 - 54"	77	159	30,46	64
00	63	"60 - 64"	83	156	34,11	66

We have to load into Rcommander the dataset *osteoporosis.csv*

1. “.csv” -> “comma separated values” but could be separated with \tab, spaces, ; ,
2. First of all “look” the file with a file viewer like “wordpad”, “bloc de notas”,...
3. Once we know how:
 1. the data is separated
 2. the decimals are separated
4. we can open with Rcmdr

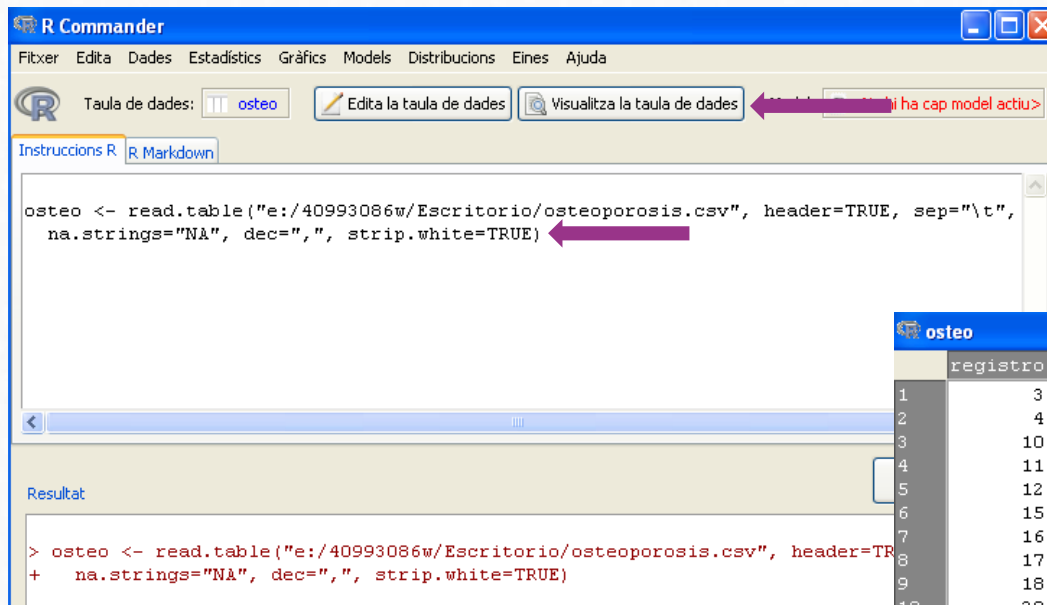
Practice. Loading data sets

We have to import into Rcommander the dataset *osteoporosis.csv*

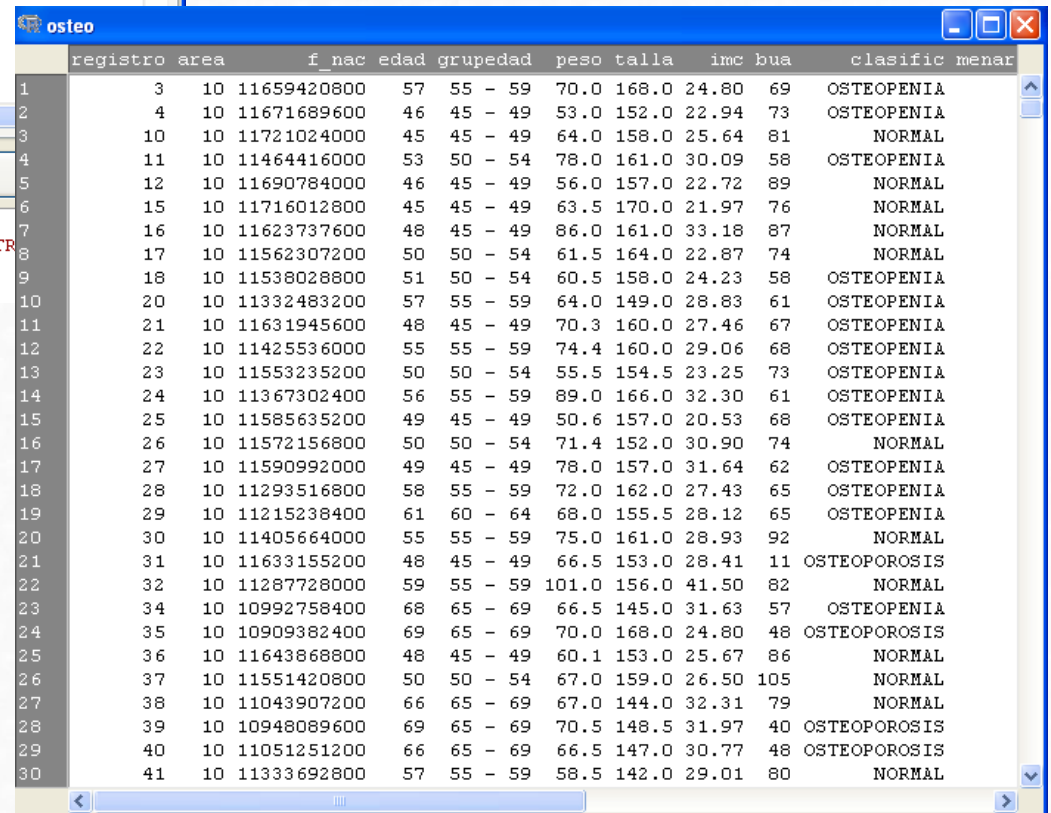


Practice. Loading data sets

We have to import into Rcommander the dataset *osteoporosis.csv*



```
osteoporosis <- read.table("e:/40993086w/Escritorio/osteoporosis.csv", header=TRUE, sep=";", as.is=TRUE, na.strings="NA", dec=".", strip.white=TRUE)
```



	registro	area	f_nac	edad	grupedad	peso	talla	imc	bua	clasific
1	3	10	11659420800	57	55 - 59	70.0	168.0	24.80	69	OSTEOPENIA
2	4	10	11671689600	46	45 - 49	53.0	152.0	22.94	73	OSTEOPENIA
3	10	10	11721024000	45	45 - 49	64.0	158.0	25.64	81	NORMAL
4	11	10	11464416000	53	50 - 54	78.0	161.0	30.09	58	OSTEOPENIA
5	12	10	11690784000	46	45 - 49	56.0	157.0	22.72	89	NORMAL
6	15	10	11716012800	45	45 - 49	63.5	170.0	21.97	76	NORMAL
7	16	10	11623737600	48	45 - 49	86.0	161.0	33.18	87	NORMAL
8	17	10	11562307200	50	50 - 54	61.5	164.0	22.87	74	NORMAL
9	18	10	11538028800	51	50 - 54	60.5	158.0	24.23	58	OSTEOPENIA
10	20	10	11332483200	57	55 - 59	64.0	149.0	28.83	61	OSTEOPENIA
11	21	10	11631945600	48	45 - 49	70.3	160.0	27.46	67	OSTEOPENIA
12	22	10	11425536000	55	55 - 59	74.4	160.0	29.06	68	OSTEOPENIA
13	23	10	11553235200	50	50 - 54	55.5	154.5	23.25	73	OSTEOPENIA
14	24	10	11367302400	56	55 - 59	89.0	166.0	32.30	61	OSTEOPENIA
15	25	10	11585635200	49	45 - 49	50.6	157.0	20.53	68	OSTEOPENIA
16	26	10	11572156800	50	50 - 54	71.4	152.0	30.90	74	NORMAL
17	27	10	11590992000	49	45 - 49	78.0	157.0	31.64	62	OSTEOPENIA
18	28	10	11293516800	58	55 - 59	72.0	162.0	27.43	65	OSTEOPENIA
19	29	10	11215238400	61	60 - 64	68.0	155.5	28.12	65	OSTEOPENIA
20	30	10	11405664000	55	55 - 59	75.0	161.0	28.93	92	NORMAL
21	31	10	11633155200	48	45 - 49	66.5	153.0	28.41	11	OSTEOPOROSIS
22	32	10	11287728000	59	55 - 59	101.0	156.0	41.50	82	NORMAL
23	34	10	10992758400	68	65 - 69	66.5	145.0	31.63	57	OSTEOPENIA
24	35	10	10909382400	69	65 - 69	70.0	168.0	24.80	48	OSTEOPOROSIS
25	36	10	11643868800	48	45 - 49	60.1	153.0	25.67	86	NORMAL
26	37	10	11551420800	50	50 - 54	67.0	159.0	26.50	105	NORMAL
27	38	10	11043907200	66	65 - 69	67.0	144.0	32.31	79	NORMAL
28	39	10	10948089600	69	65 - 69	70.5	148.5	31.97	40	OSTEOPOROSIS
29	40	10	11051251200	66	65 - 69	66.5	147.0	30.77	48	OSTEOPOROSIS
30	41	10	11333692800	57	55 - 59	58.5	142.0	29.01	80	NORMAL

Practice. Adding variables to the data set

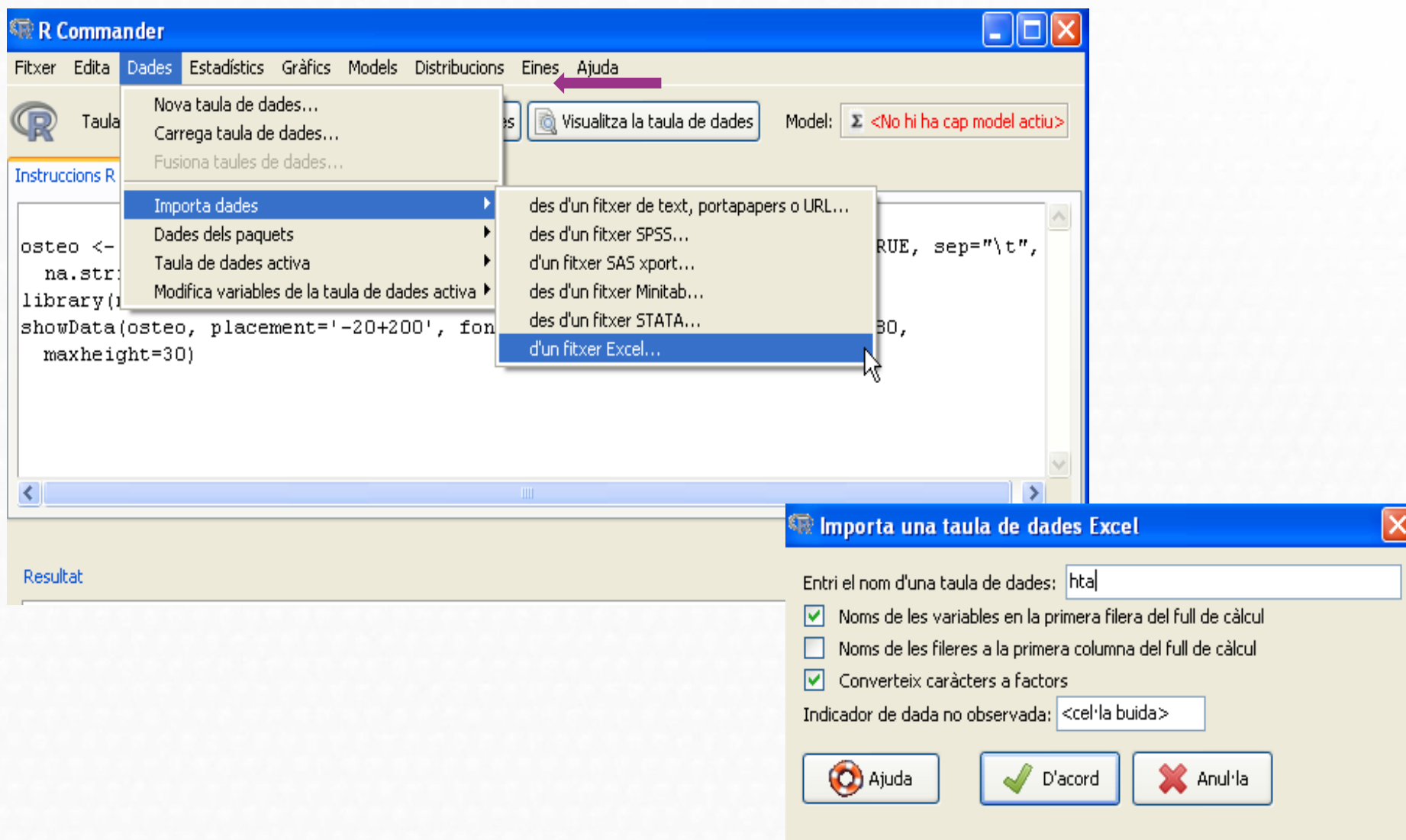
We have to add new variables to the dataset *osteoporosis.csv*

The screenshot shows the RStudio interface with the 'Calcular una nueva variable' dialog box open. The dialog box has a title bar with a red 'X' icon. It contains a list of variables on the left, including 'area', 'bua', 'clasific [factor]', 'edad', 'edad_men', 'f_nac [factor]', and 'Nombre de la nueva variable' with the value 'imc_rec'. On the right, the 'Expresión a calcular' field contains the formula 'peso/ ((talla/100) ^2)'. At the bottom, there are five buttons: 'Ayuda' (Help), 'Reiniciar' (Reset), 'Aceptar' (Accept), 'Cancelar' (Cancel), and 'Aplicar' (Apply). The background shows the RStudio menu bar with 'Datos', 'Estadísticos', 'Gráficas', 'Modelos', 'Distribuciones', 'Herramientas', and 'Ayuda'. A script editor is visible with R code:

```
using.labels=TRUE)
read.table("E:/BioStatFLOSS/dades/
", strip.white=TRUE)
recalculado <- with(osteo, peso/
```


Practice. Loading data sets

We have to import into Rcommander the dataset *hta.xls*



The screenshot shows the R Commander application window. The 'Dades' menu is open, and the 'Importa dades' option is selected, which has opened a sub-menu. In this sub-menu, the option 'd'un fitxer Excel...' is highlighted by the mouse. Below the main window, a dialog box titled 'Importa una taula de dades Excel' is open. This dialog box contains a text field for the dataset name, which is 'hta'. There are three checked options: 'Noms de les variables en la primera filera del full de càlcul', 'Converteix caràcters a factors', and 'Indicador de dada no observada: <cel·la buida>'. At the bottom of the dialog box, there are three buttons: 'Ajuda', 'D'acord', and 'Anul·la'.

R Commander

Fitxer Editar Dades Estadístics Gràfics Models Distributions Eines Ajuda

Nova taula de dades...
Carrega taula de dades...
Fusiona taules de dades...
Visualitza la taula de dades

Model: Σ <No hi ha cap model actiu>

Importa dades
Dades dels paquets
Taula de dades activa
Modifica variables de la taula de dades activa

des d'un fitxer de text, portapapers o URL...
des d'un fitxer SPSS...
d'un fitxer SAS xport...
des d'un fitxer Minitab...
des d'un fitxer STATA...
d'un fitxer Excel...

Resultat

Importa una taula de dades Excel

Entri el nom d'una taula de dades: hta

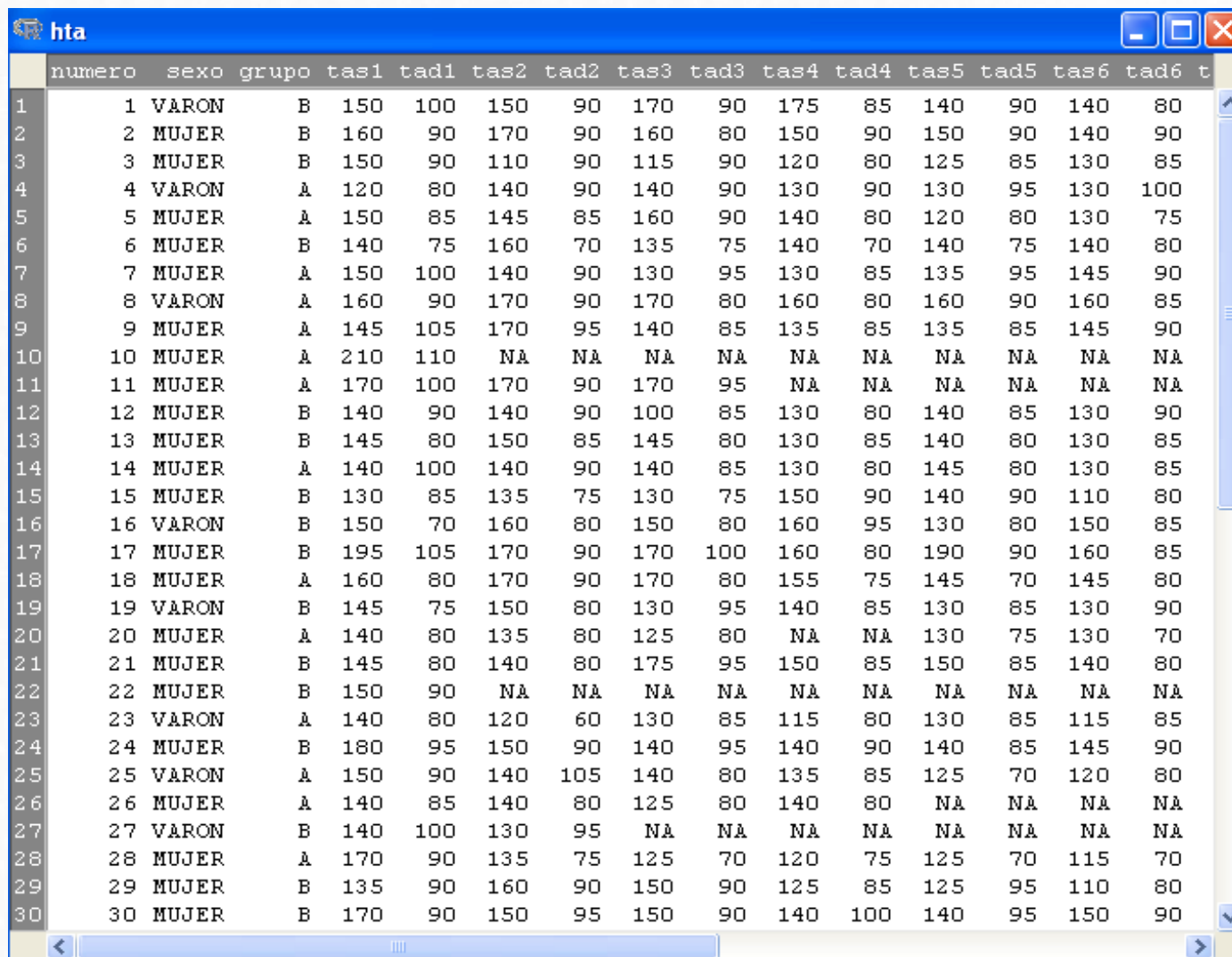
☒ Noms de les variables en la primera filera del full de càlcul
☐ Noms de les fileres a la primera columna del full de càlcul
☒ Converteix caràcters a factors

Indicador de dada no observada: <cel·la buida>

Ajuda D'acord Anul·la

Practice. Loading data sets

We have to import into Rcommander the dataset *hta.xls*



	numero	sexo	grupo	tas1	tad1	tas2	tad2	tas3	tad3	tas4	tad4	tas5	tad5	tas6	tad6	t
1	1	VARON	B	150	100	150	90	170	90	175	85	140	90	140	80	
2	2	MUJER	B	160	90	170	90	160	80	150	90	150	90	140	90	
3	3	MUJER	B	150	90	110	90	115	90	120	80	125	85	130	85	
4	4	VARON	A	120	80	140	90	140	90	130	90	130	95	130	100	
5	5	MUJER	A	150	85	145	85	160	90	140	80	120	80	130	75	
6	6	MUJER	B	140	75	160	70	135	75	140	70	140	75	140	80	
7	7	MUJER	A	150	100	140	90	130	95	130	85	135	95	145	90	
8	8	VARON	A	160	90	170	90	170	80	160	80	160	90	160	85	
9	9	MUJER	A	145	105	170	95	140	85	135	85	135	85	145	90	
10	10	MUJER	A	210	110	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
11	11	MUJER	A	170	100	170	90	170	95	NA	NA	NA	NA	NA	NA	
12	12	MUJER	B	140	90	140	90	100	85	130	80	140	85	130	90	
13	13	MUJER	B	145	80	150	85	145	80	130	85	140	80	130	85	
14	14	MUJER	A	140	100	140	90	140	85	130	80	145	80	130	85	
15	15	MUJER	B	130	85	135	75	130	75	150	90	140	90	110	80	
16	16	VARON	B	150	70	160	80	150	80	160	95	130	80	150	85	
17	17	MUJER	B	195	105	170	90	170	100	160	80	190	90	160	85	
18	18	MUJER	A	160	80	170	90	170	80	155	75	145	70	145	80	
19	19	VARON	B	145	75	150	80	130	95	140	85	130	85	130	90	
20	20	MUJER	A	140	80	135	80	125	80	NA	NA	130	75	130	70	
21	21	MUJER	B	145	80	140	80	175	95	150	85	150	85	140	80	
22	22	MUJER	B	150	90	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
23	23	VARON	A	140	80	120	60	130	85	115	80	130	85	115	85	
24	24	MUJER	B	180	95	150	90	140	95	140	90	140	85	145	90	
25	25	VARON	A	150	90	140	105	140	80	135	85	125	70	120	80	
26	26	MUJER	A	140	85	140	80	125	80	140	80	NA	NA	NA	NA	
27	27	VARON	B	140	100	130	95	NA	NA	NA	NA	NA	NA	NA	NA	
28	28	MUJER	A	170	90	135	75	125	70	120	75	125	70	115	70	
29	29	MUJER	B	135	90	160	90	150	90	125	85	125	95	110	80	
30	30	MUJER	B	170	90	150	95	150	90	140	100	140	95	150	90	

Practice. Adding variables to the data set

We have to add variables *age* and *weight* the dataset *hta.xls*

The image shows the R Commander interface. The main window displays the R console with the following code:

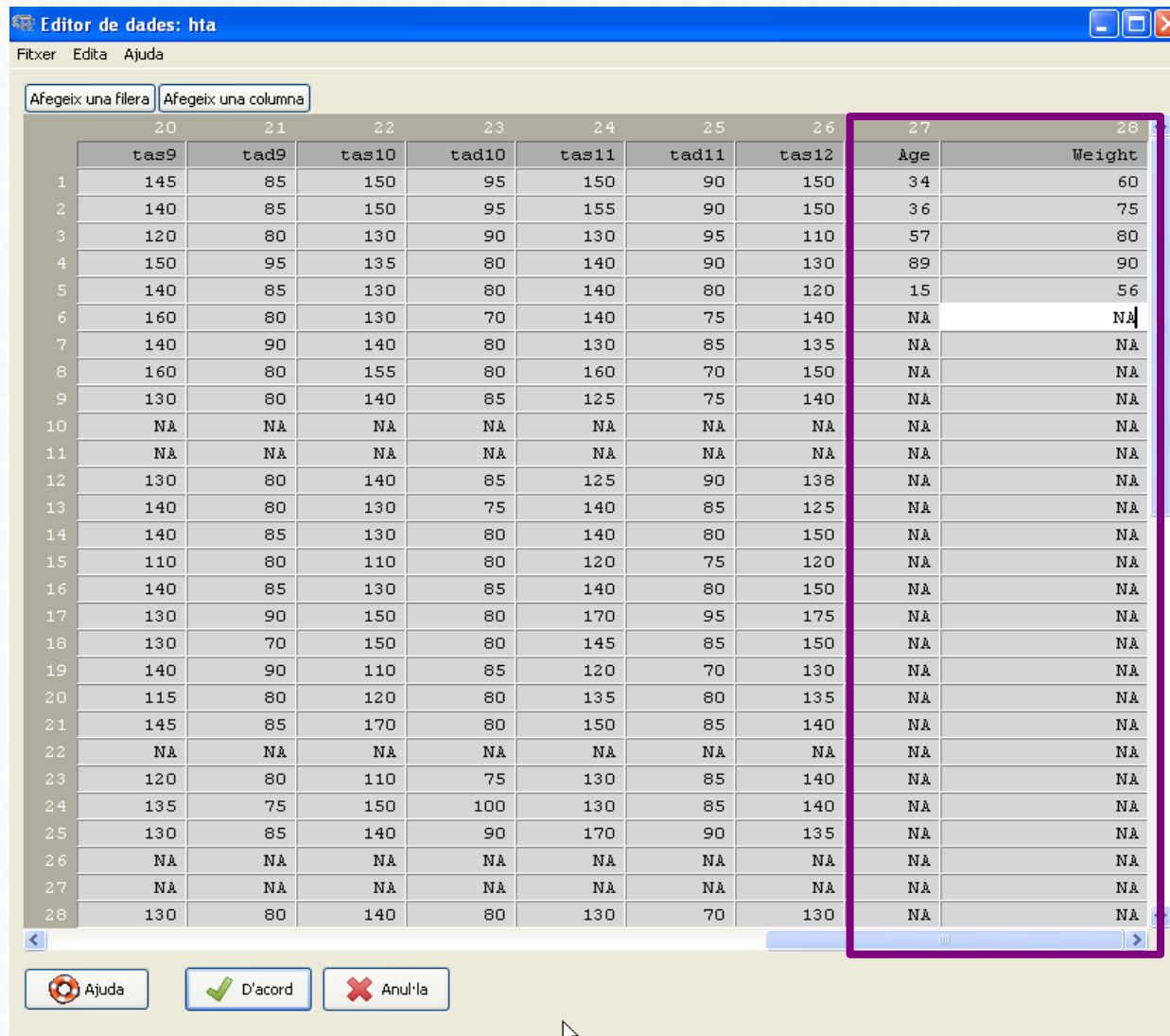
```
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
maxheight=30)
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
maxheight=30)
editDataset(hta)
names(hta)
AggregatedData<- aggregate(tad1 ~ grupo, data=hta, FUN=mean)
showData(AggregatedData, placement='-20+200', font=getRcmdr('logFont'), ma
maxheight=30)
```

The 'Resultat' pane is empty. The 'Editor de dades: hta' window is open, showing a table with 28 rows and 9 columns. The columns are: rowname, numero, sexo, grupo, tas1, tad1, tas2, tad2, tas3, and tad3. The data is as follows:

rowname	numero	sexo	grupo	tas1	tad1	tas2	tad2	tas3	tad3
1	1	VARON	B	150	100	150	90	170	90
2	2	MUJER	B	160	90	170	90	160	80
3	3	MUJER	B	150	90	110	90	115	90
4	4	VARON	A	120	80	140	90	140	90
5	5	MUJER	A	150	85	145	85	160	90
6	6	MUJER	B	140	75	160	70	135	75
7	7	MUJER	A	150	100	140	90	130	95
8	8	VARON	A	160	90	170	90	170	80
9	9	MUJER	A	145	105	170	95	140	85
10	10	MUJER	A	210	110	NA	NA	NA	NA
11	11	MUJER	A	170	100	170	90	170	95
12	12	MUJER	B	140	90	140	90	100	85
13	13	MUJER	B	145	80	150	85	145	80
14	14	MUJER	A	140	100	140	90	140	85
15	15	MUJER	B	130	85	135	75	130	75
16	16	VARON	B	150	70	160	80	150	80
17	17	MUJER	B	195	105	170	90	170	100
18	18	MUJER	A	160	80	170	90	170	80
19	19	VARON	B	145	75	150	80	130	95
20	20	MUJER	A	140	80	135	80	125	80
21	21	MUJER	B	145	80	140	80	175	95
22	22	MUJER	B	150	90	NA	NA	NA	NA
23	23	VARON	A	140	80	120	60	130	85
24	24	MUJER	B	180	95	150	90	140	95
25	25	VARON	A	150	90	140	105	140	80
26	26	MUJER	A	140	85	140	80	125	80
27	27	VARON	B	140	100	130	95	NA	NA
28	28	MUJER	A	170	90	135	75	125	70

Practice. Adding variables to the data set

We have to add variables *age* and *weight* the dataset *hta.xls*



Editor de dades: hta

Fitxer Edita Ajuda

Afegeix una filera Afegeix una columna

	20	21	22	23	24	25	26	27	28
	tas9	tad9	tas10	tad10	tas11	tad11	tas12	Age	Weight
1	145	85	150	95	150	90	150	34	60
2	140	85	150	95	155	90	150	36	75
3	120	80	130	90	130	95	110	57	80
4	150	95	135	80	140	90	130	89	90
5	140	85	130	80	140	80	120	15	56
6	160	80	130	70	140	75	140	NA	NA
7	140	90	140	80	130	85	135	NA	NA
8	160	80	155	80	160	70	150	NA	NA
9	130	80	140	85	125	75	140	NA	NA
10	NA	NA	NA	NA	NA	NA	NA	NA	NA
11	NA	NA	NA	NA	NA	NA	NA	NA	NA
12	130	80	140	85	125	90	138	NA	NA
13	140	80	130	75	140	85	125	NA	NA
14	140	85	130	80	140	80	150	NA	NA
15	110	80	110	80	120	75	120	NA	NA
16	140	85	130	85	140	80	150	NA	NA
17	130	90	150	80	170	95	175	NA	NA
18	130	70	150	80	145	85	150	NA	NA
19	140	90	110	85	120	70	130	NA	NA
20	115	80	120	80	135	80	135	NA	NA
21	145	85	170	80	150	85	140	NA	NA
22	NA	NA	NA	NA	NA	NA	NA	NA	NA
23	120	80	110	75	130	85	140	NA	NA
24	135	75	150	100	130	85	140	NA	NA
25	130	85	140	90	170	90	135	NA	NA
26	NA	NA	NA	NA	NA	NA	NA	NA	NA
27	NA	NA	NA	NA	NA	NA	NA	NA	NA
28	130	80	140	80	130	70	130	NA	NA

Ajuda D'acord Anul·la

1. Import into Rcmdr the dataset *demora.dta* (STATA file)
2. Convert numeric variable *dolor* to factor (rename levels accordingly, eg. 0=low, 1=medium, 2=high)
3. Remove cases with missing data (NA) in *demora* variable.
4. Recode variable *edad* in a new variable named *rango_edad* according to the following ranges:

0:25="young"

25:50="adult"

51:75="mid-old"

76:100="old"

Practice. Statistics with Rcmdr

We have to calculate some descriptive statistics with the dataset *osteo*.

The image shows two windows from the R Commander software. The top window is the main R Commander interface, and the bottom window is the 'Resums numèrics' (Numerical Summaries) dialog box.

R Commander Window:

- Menu: Estadístics
- Sub-menu: Resums
- Selected option: Resums numèrics...
- Code editor content:

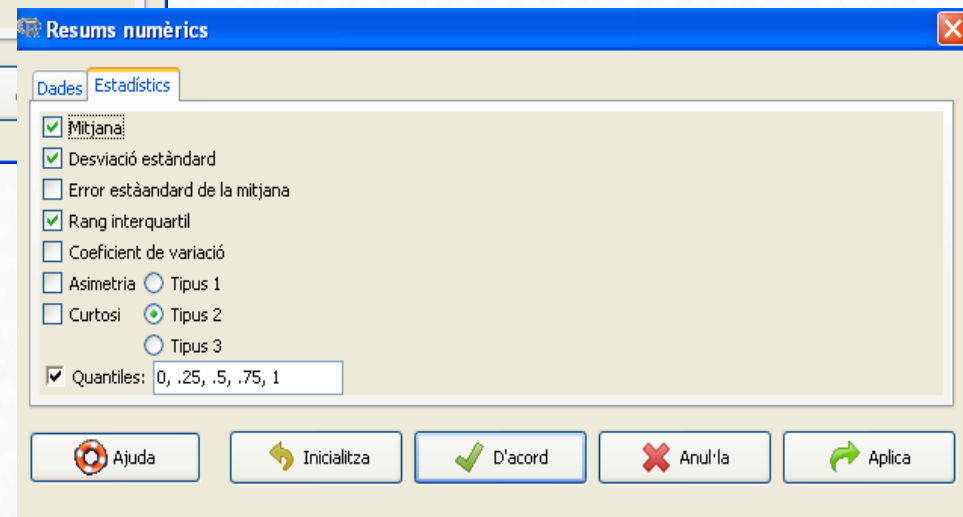
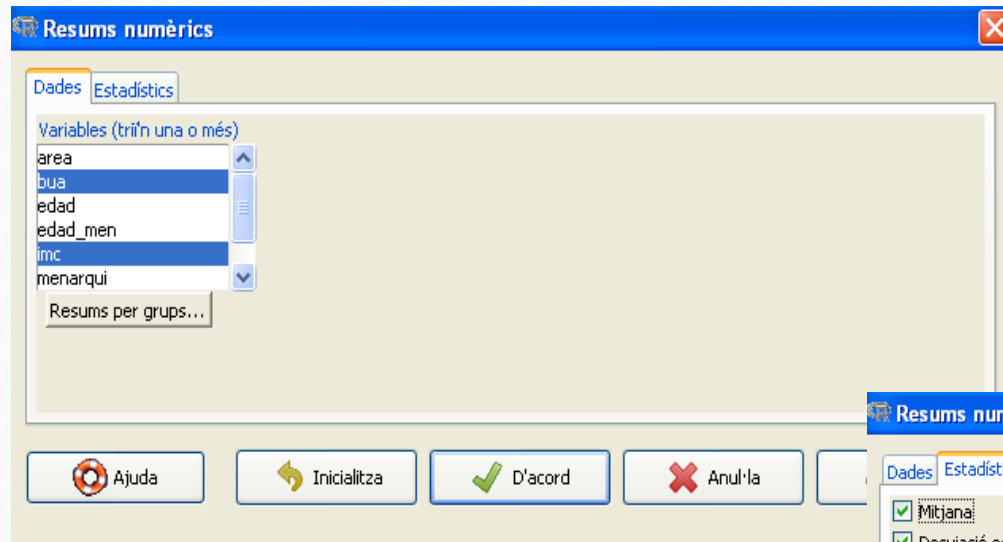
```
showData(hta,  
  maxheight=30)  
editDataset(hta,  
  names(hta))  
AggregatedData<- aggregate(cadi ~ grupo, data=nta, FUN=mean)  
showData(AggregatedData, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,  
  maxheight=30)  
editDataset(hta)  
editDataset(hta)
```

Resums numèrics Window:

- Tab: Estadístics
- Variables (tri'n una o més):
 - area
 - bua
 - edad
 - edad_men
 - imc
 - menarqui
- Button: Resums per grups...
- Buttons at the bottom: Ajuda, Inicialitza, D'acord, Anul·la, Aplica

Practice. Statistics with Rcmdr

We have to calculate some descriptive statistics with the dataset *osteo*.



Practice. Statistics with Rcmdr

We have to calculate some descriptive statistics with the dataset *osteo*.

Resultat

Executar

```
> names(hta)
[1] "numero" "sexo" "grupo" "tas1" "tad1" "tas2" "tad2" "tas3"
[9] "tad3" "tas4" "tad4" "tas5" "tad5" "tas6" "tad6" "tas7"
[17] "tad7" "tas8" "tad8" "tas9" "tad9" "tas10" "tad10" "tas11"
[25] "tad11" "tas12"

> AggregatedData<- aggregate(tad1 ~ grupo, data=hta, FUN=mean)

> showData(AggregatedData, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
+ maxheight=30)

> editDataset(hta)

> editDataset(hta)

> library(abind, pos=15)

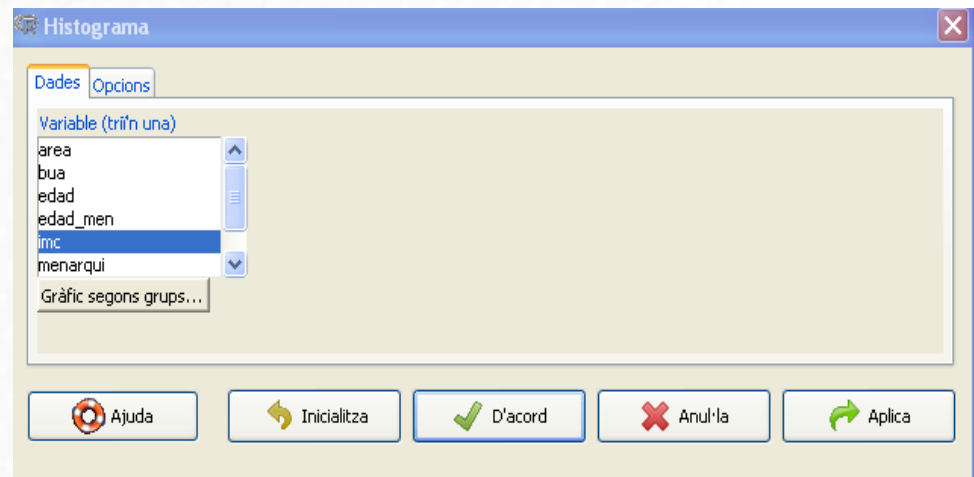
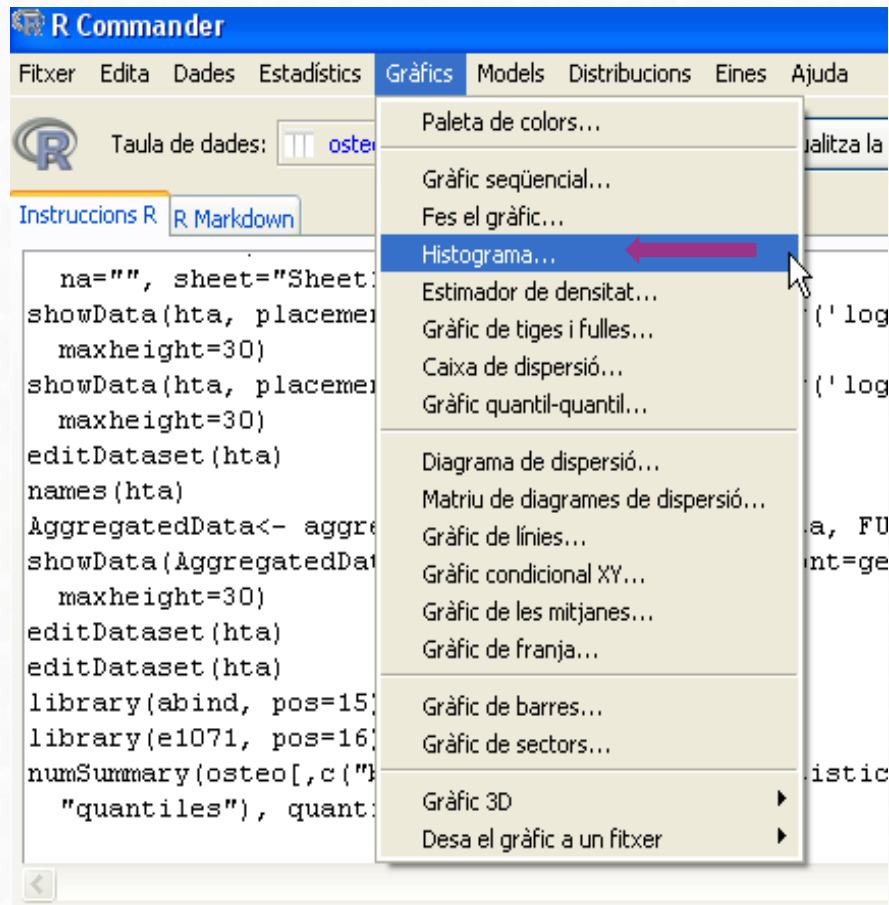
> library(e1071, pos=16)

> numSummary(osteo[,c("bua", "imc", "peso")], statistics=c("mean", "sd", "IQR",
+ "quantiles"), quantiles=c(0,.25,.5,.75,1))
  mean      sd      IQR    0%     25%    50%    75%   100%    n
bua  73.29700 16.809323 22.0000 11.00 62.0000 72.00 84.00 136.00 1000
imc  28.10776  4.717925  6.0225 17.21 24.7975 27.51 30.82  48.39 1000
peso 69.12280 11.643733 14.5000 44.00 60.5000 68.00 75.00 123.50 1000
```

Missatges

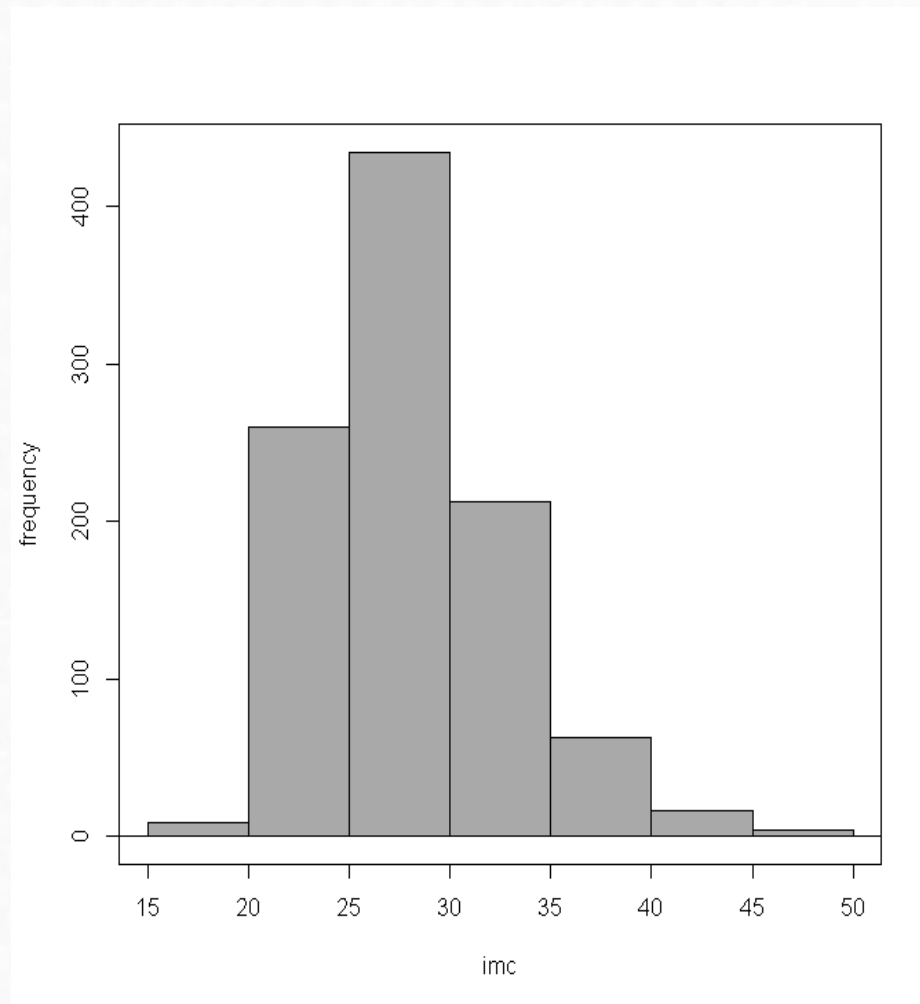
Practice. Graphics with Rcmdr

We want to see how the variable *imc* is distributed among the individuals



Practice. Graphics with Rcmdr

We want to see how the variable *imc* is distributed among the individuals



Practice. Running scripts with Rcmdr

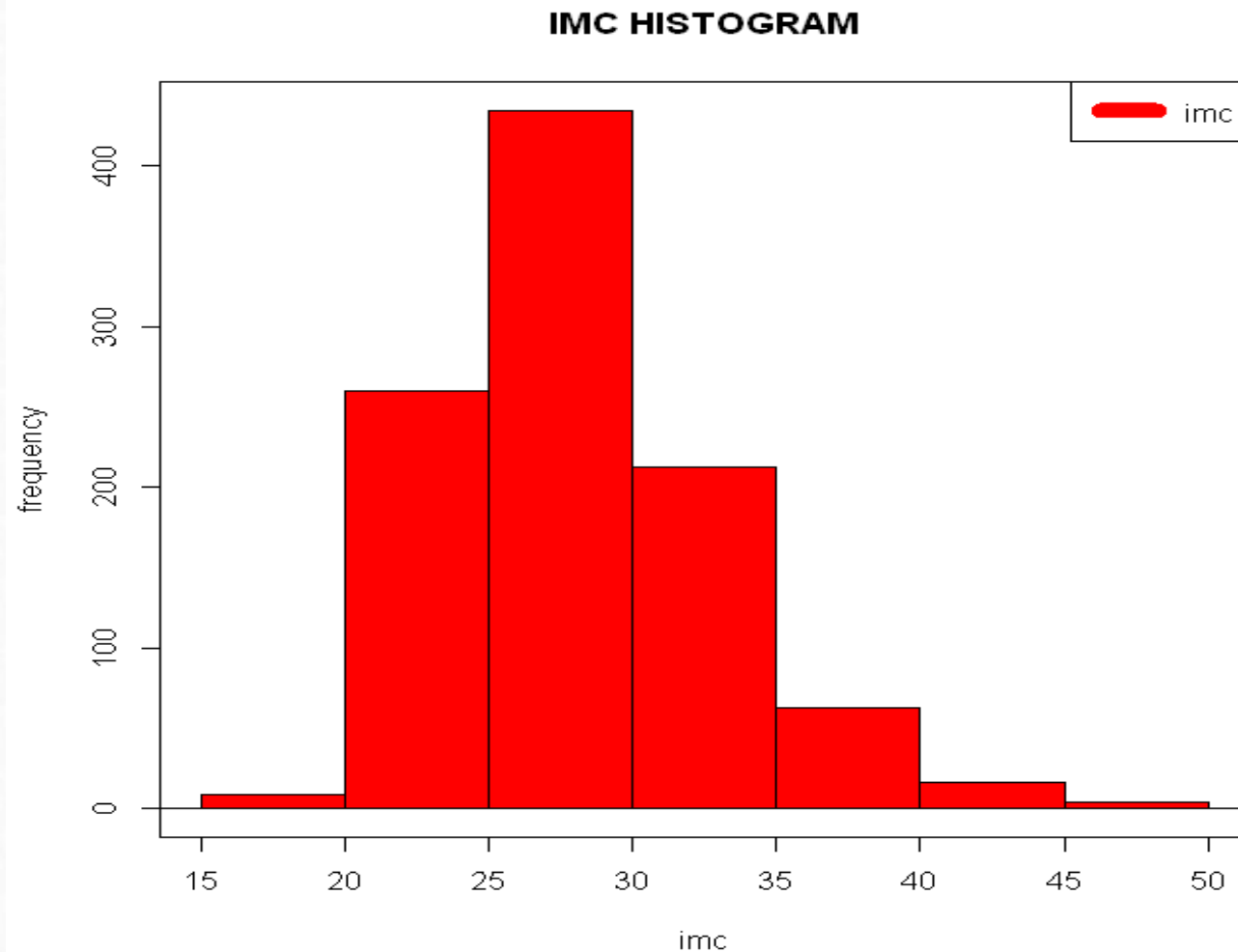
We want to change the graphic colour, add a legend and a main title

```
Instruccions R | R Markdown
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
editDataset(hta)
names(hta)
AggregatedData<- aggregate(tad1 ~ grupo, data=hta, FUN=mean)
showData(AggregatedData, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
editDataset(hta)
editDataset(hta)
library(abind, pos=15)
library(e1071, pos=16)
numSummary(osteo[,c("bua", "imc", "peso")], statistics=c("mean", "sd", "IQR",
  "quantiles"), quantiles=c(0, 25, 5, 75, 1))
with(osteo, Hist(imc, scale="frequency", breaks="Sturges", col="darkgray"))

showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
editDataset(hta)
names(hta)
AggregatedData<- aggregate(tad1 ~ grupo, data=hta, FUN=mean)
showData(AggregatedData, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
editDataset(hta)
editDataset(hta)
library(abind, pos=15)
library(e1071, pos=16)
numSummary(osteo[,c("bua", "imc", "peso")], statistics=c("mean", "sd", "IQR",
  "quantiles"), quantiles=c(0, 25, 5, 75, 1))
with(osteo, Hist(imc, scale="frequency", breaks="Sturges", col="red", main="IMC HISTOGRAM"))
legend("topright", "imc", col="red", lwd=10)
```

Practice. Running scripts with Rcmdr

We want to change the graphic colour, add a legend and a main title



Exporting results

Years ago, R-commander had the limitation that exporting data had to be done through saving plots and copy-pasting code to external software.

Now, Rcmdr integrates Markdown/Knitr results writing system, that allows the user to export all work (code, plots and results) to html pages in quick and easy way.

<https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

http://www.rstudio.com/ide/docs/authoring/using_markdown

<http://yihui.name/knitr/>

Practice. Export results as html

We want to export the results into an html report



The screenshot shows the R Commander application window. At the top, there are two tabs: 'Instruccions R' and 'R Markdown'. A purple arrow points to the 'R Markdown' tab. The main text area contains the following R Markdown template code:

```
<!-- R Commander Markdown Template -->

Replace with Main Title
=====

### Your Name

### `r as.character(Sys.Date())`

```{r echo=FALSE}
include this code chunk as-is to set options
knitr::opts_chunk$set(comment=NA, prompt=TRUE, out.width=750, fig.height=8, fig.width=8)
library(Rcmdr)
library(car)
library(RcmdrMisc)
```
```

At the bottom right of the window, there is a button labeled 'Genera informe' with a document icon. A purple arrow points to this button. In the bottom left corner, the word 'Resultat' is visible.

Practice. Export results as html

```
<!-- R Commander Markdown Template -->
```

```
Generacion de Informes con Rcommander y Markdown  
=====
```

```
Ricardo Gonzalo
```

```
### `r as.character(Sys.Date())`
```

```
```{r echo=FALSE}  
include this code chunk as-is to set options
opts_chunk$set(comment=NA, prompt=TRUE,
out.width=750, fig.height=8, fig.width=8)
library(Rcmdr)
```
```

```
Calculo de medias:  
- media de a
```

```
```{r calc1, echo=TRUE}  
a<-c(3,4,5,6)
mean(a)
```
```

```
- media de b  
```{r calc2, echo=TRUE}  
b<-c(10,20,30,40)
mean(b)
```
```

```
- grafico a b  
```{r plot1,fig.width=3,fig.height=3}  
plot(a,b,col="red")
```
```

Generacion de Informes con Rcommander y Markdown

Ricardo Gonzalo

2014-01-20

Calculo de medias:

- media de a

```
> a <- c(3, 4, 5, 6)  
> mean(a)
```

```
[1] 4.5
```

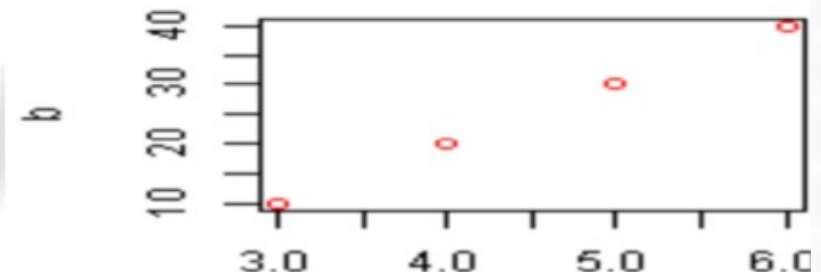
- media de b

```
> b <- c(10, 20, 30, 40)  
> mean(b)
```

```
[1] 25
```

- grafico a b

```
> plot(a, b, col = "red")
```



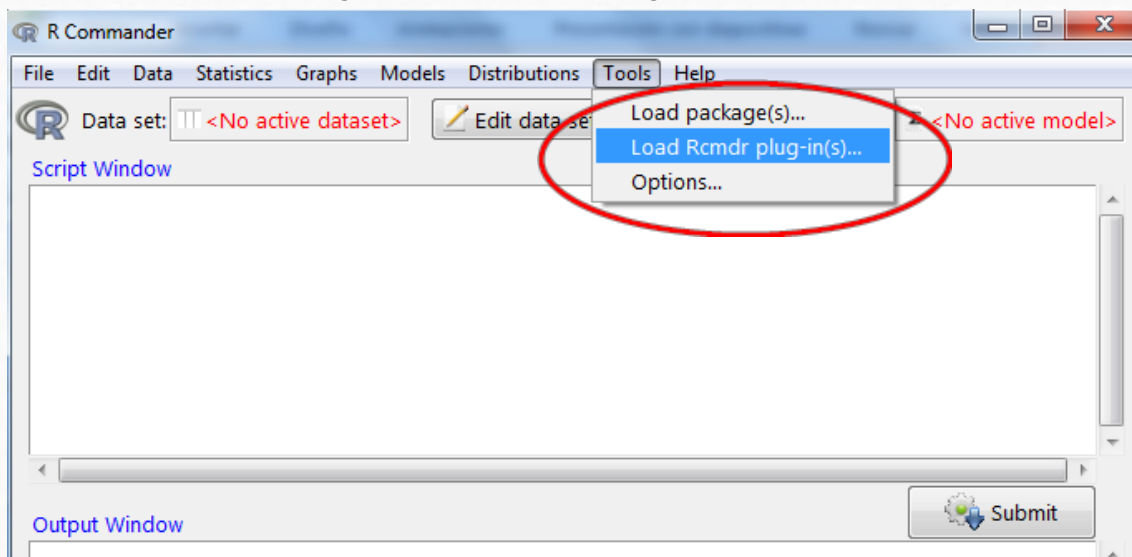
Exercise

1. Load into Rcmdr the dataset in prostate.csv
2. Calculate some descriptive statistics for some variables
3. Make a graphic from one or two variables
4. Generate the report with all the data

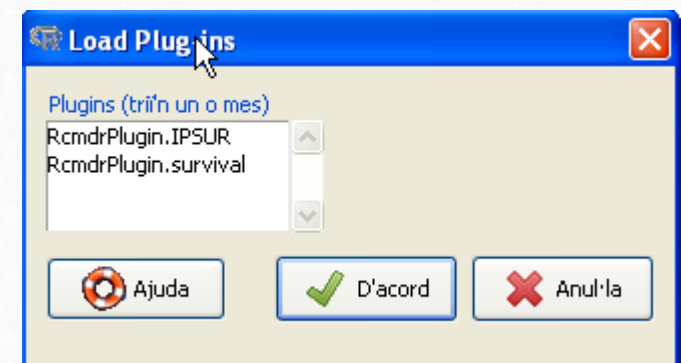
Rcommander's plugins installation

Plug-ins are additional libraries that we may want to have available in R-commander

To check if they are correctly installed...



You should find...



R manuals

- * Intro for beginners http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- * SimpleR <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>
- * Quick-R <http://www.statmethods.net/>
- * Basic statistics with R and R-commander <http://knuth.uca.es/ebrcmdr/>
- * Statistical methods with R and R-commander
<http://cran.r-project.org/doc/contrib/Saez-Castillo-RRCmdrv21.pdf>
- * Try R <http://tryr.codeschool.com/levels/1/challenges/1>

R books

- * Introductory Statistics with R
- * R for SPSS and SAS users