Vall d'Hebron
Institut de Recerca

Practical exercises with R-commander

Curs d'Estadística Bàsica per a la Recerca Biomèdica

UEB – VHIR - GRBIO

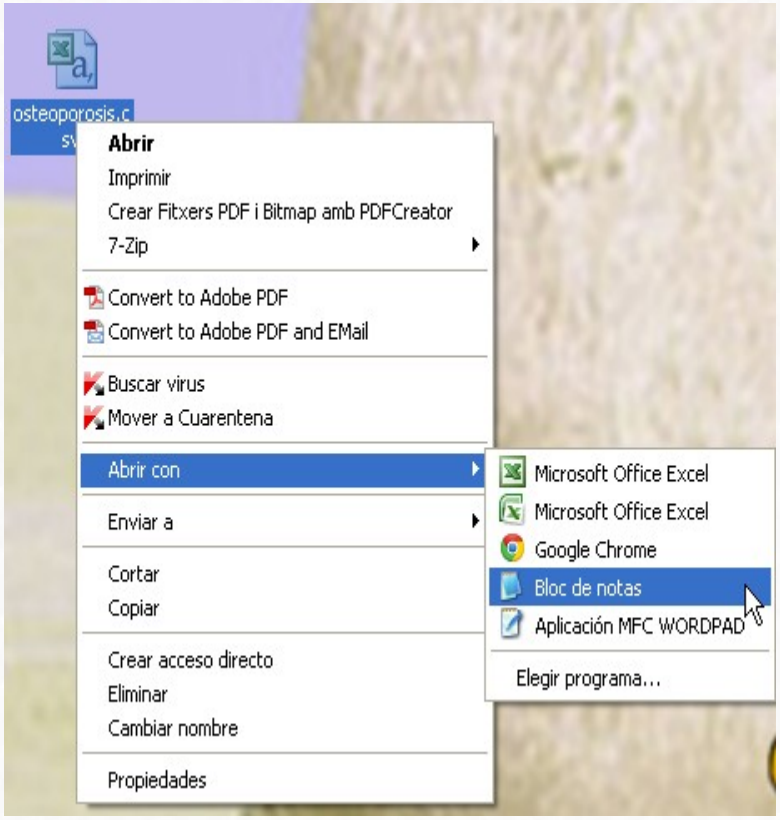
Practice. Loading data sets

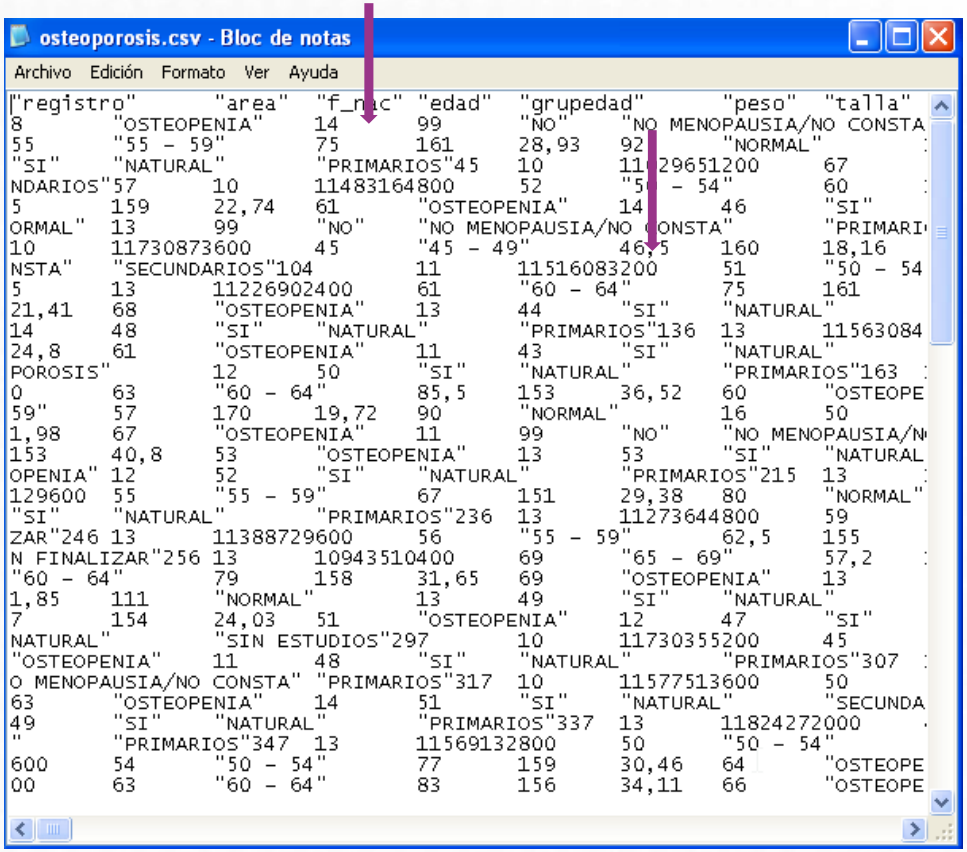
We have to load into Rcommander the dataset *osteoporosis.csv*

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*





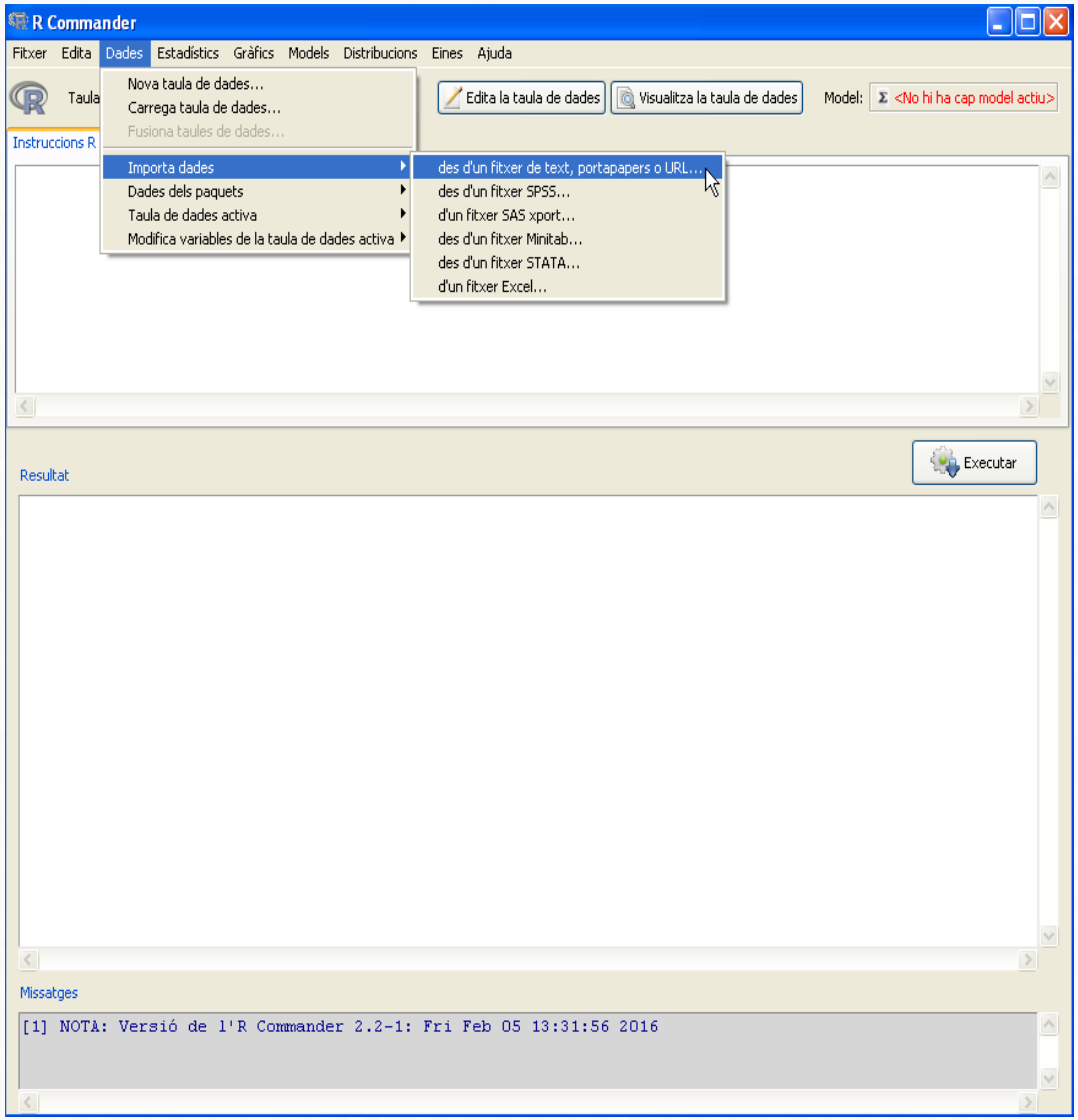
Practice. Loading data sets

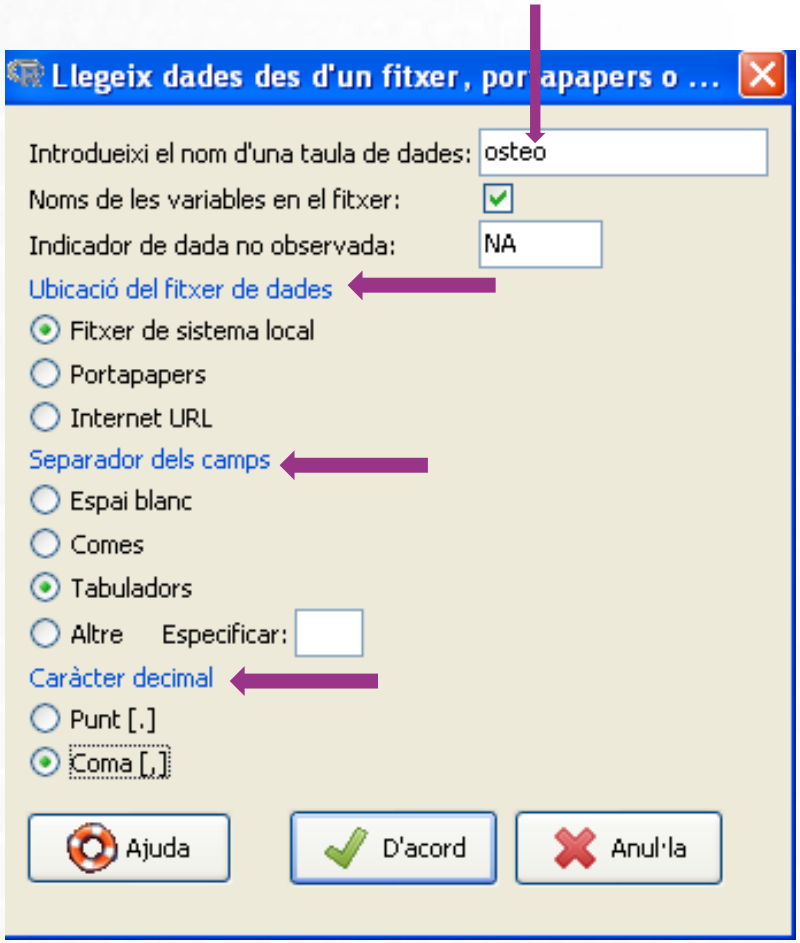
We have to load into Rcommander the dataset *osteoporosis.csv*

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

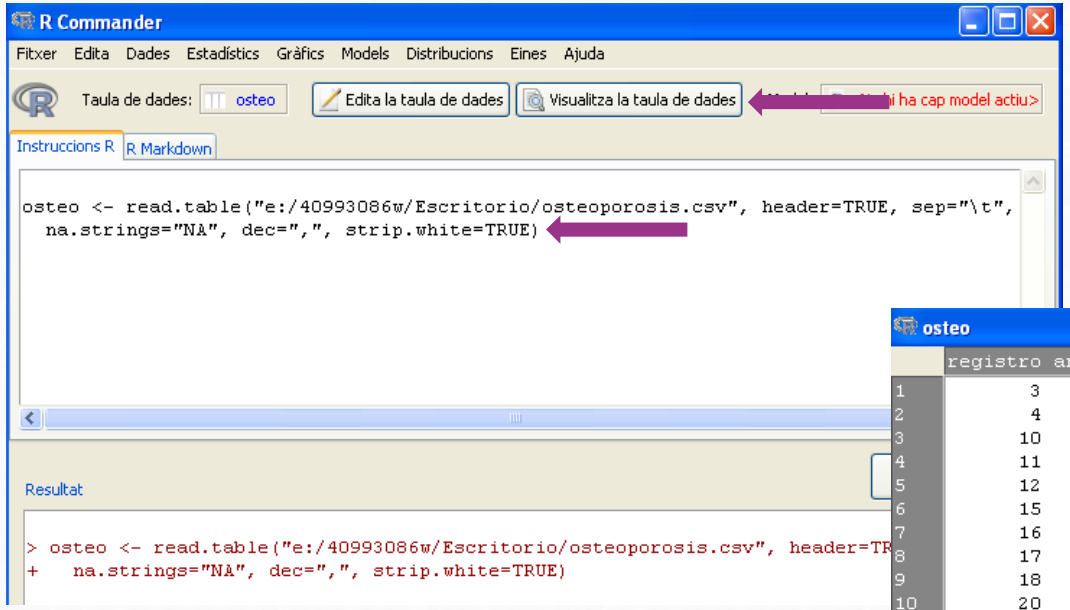
We have to import into Rcommander the dataset *osteoporosis.csv*





Practice. Loading data sets

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

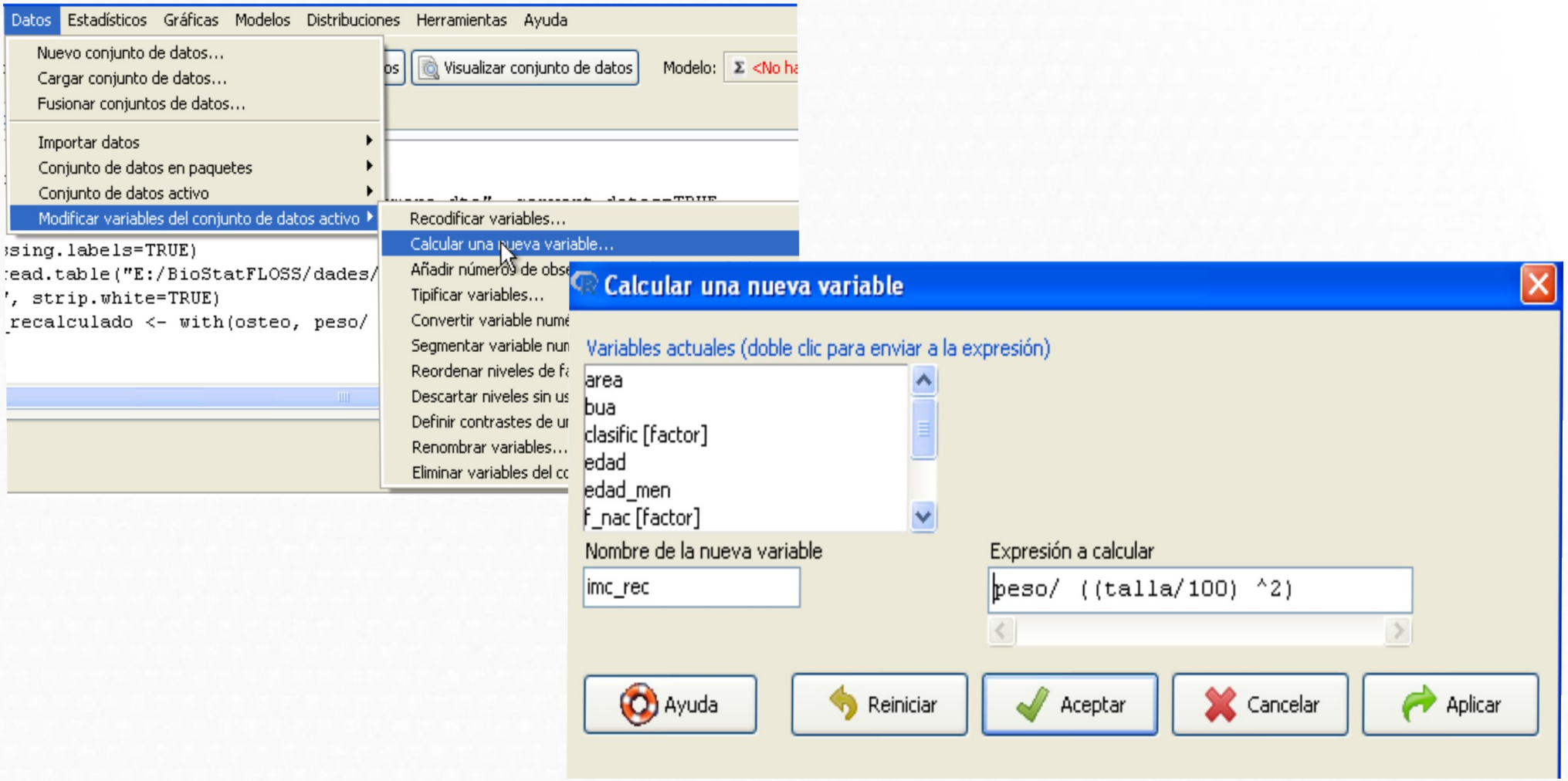


osteo

	registro	area	f_nac	edad	grupedad	peso	talla	imc	bua	clasific menar
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

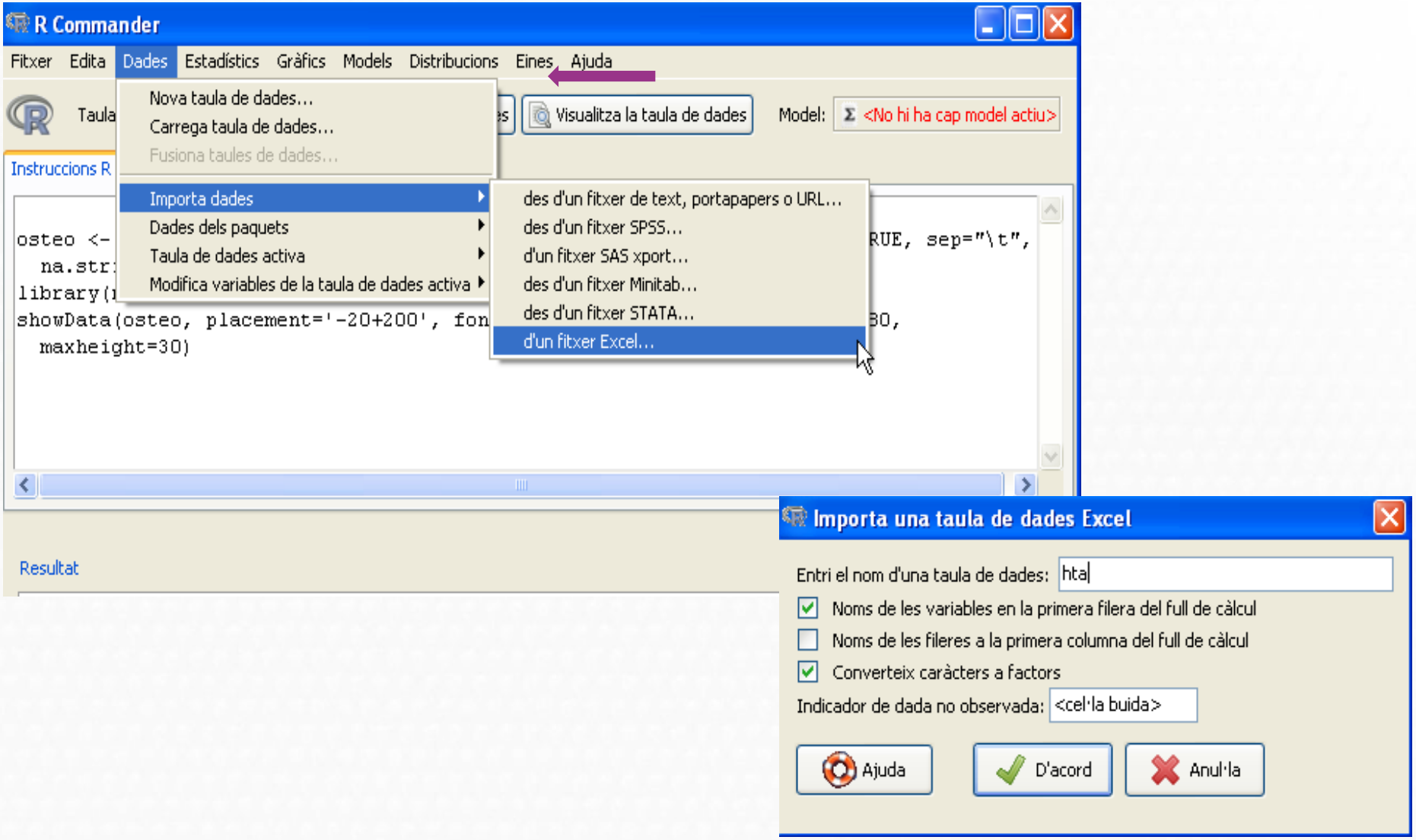
Practice. Adding variables to the data set

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



Practice. Loading data sets

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



Practice. Loading data sets

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

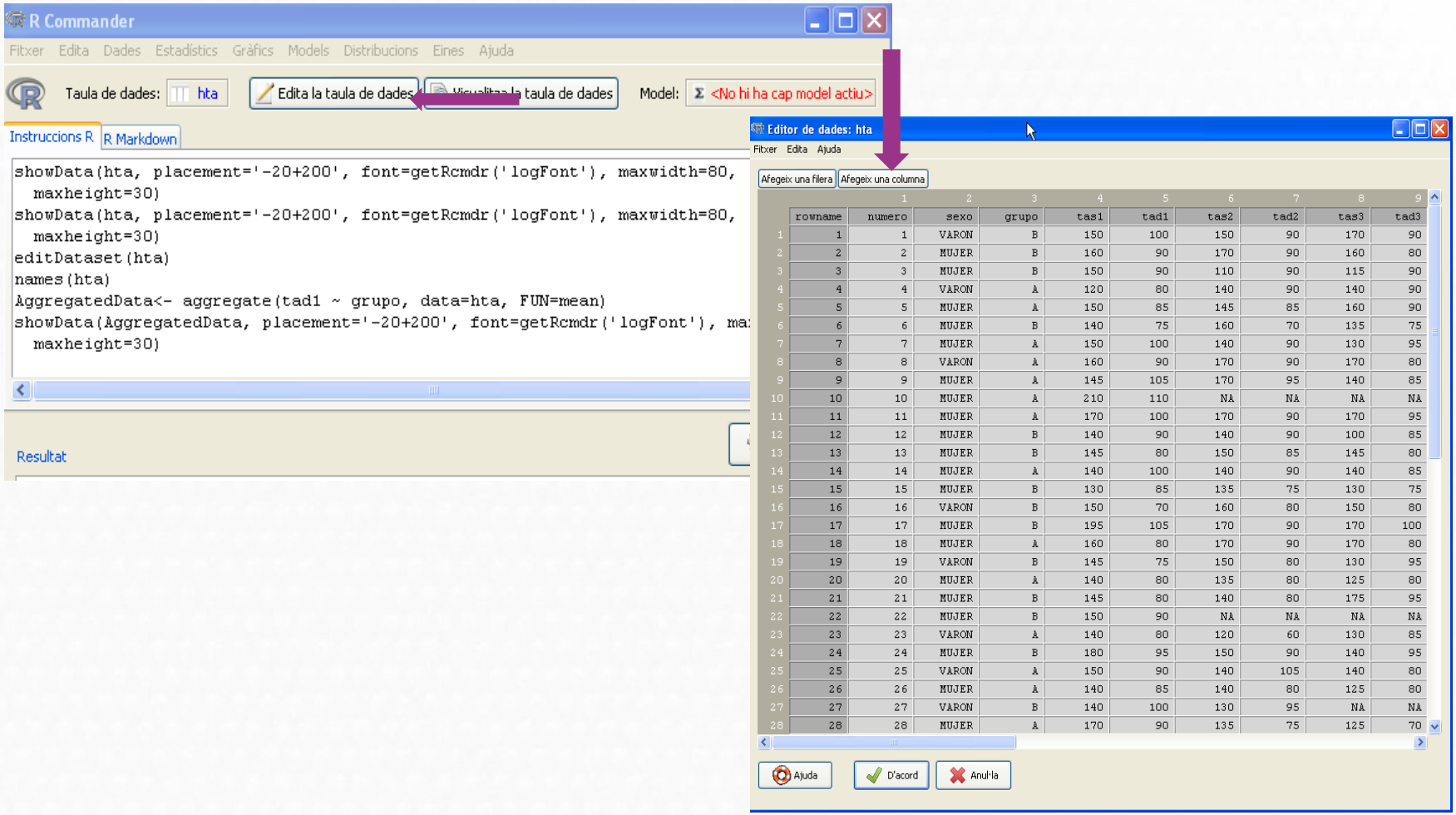
hta

	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

Vall d'Hebron
Institut de Recerca

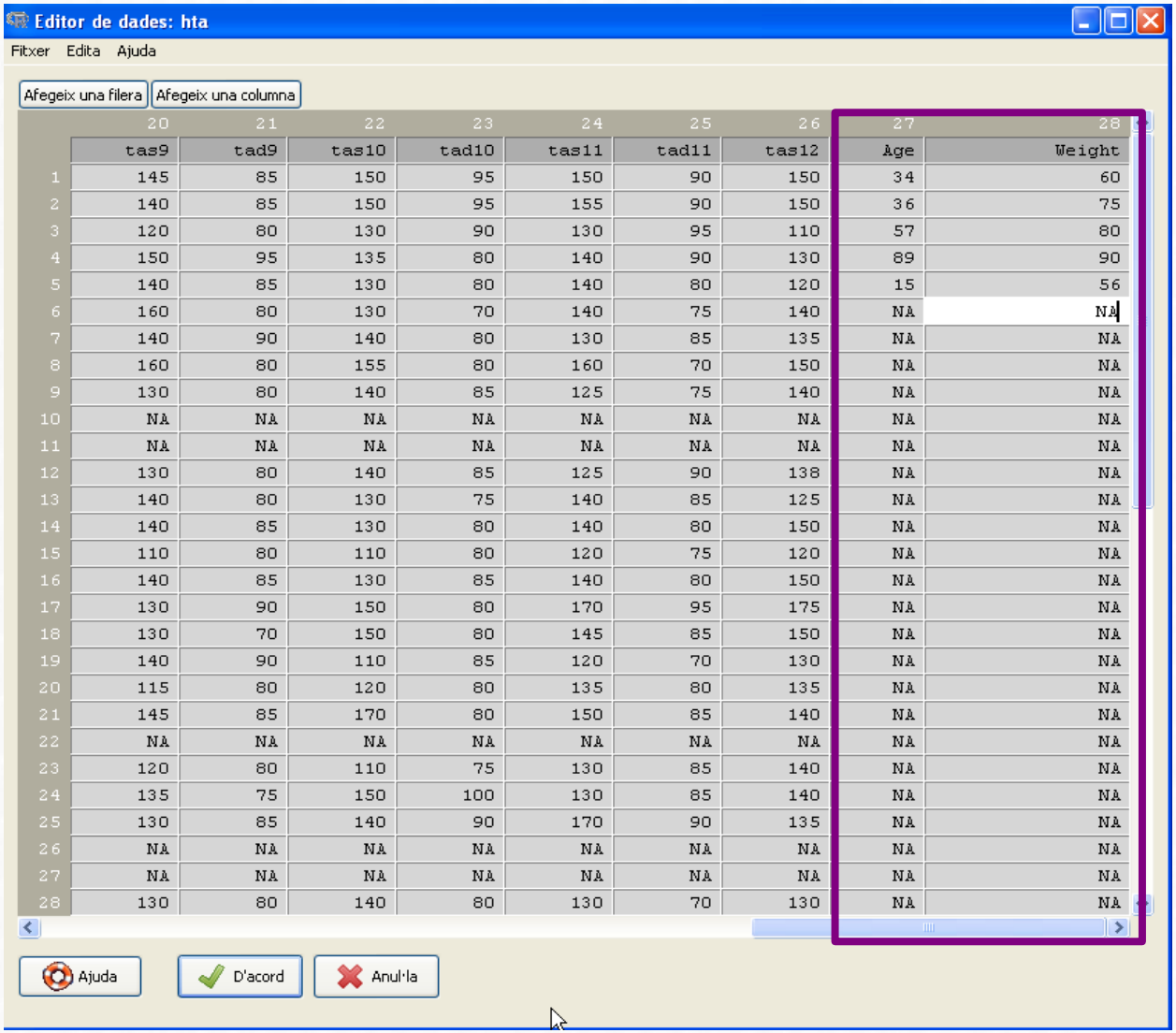
We have to add variables *age* and *weight* the dataset *hta.xls*



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



	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

Exercise

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”

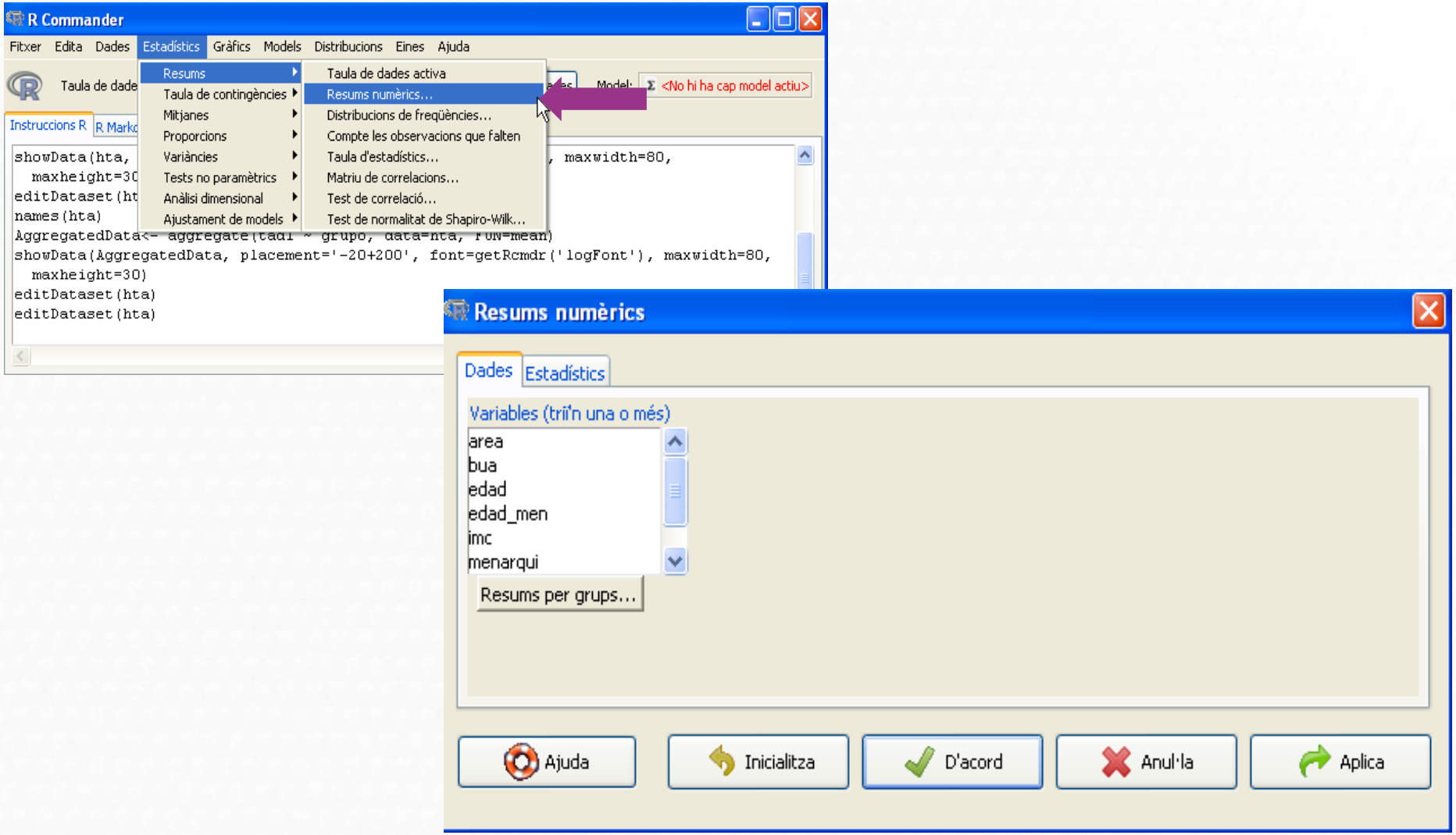
   25:50=”adult”

   51:75=”mid-old”

   76:100=”old”

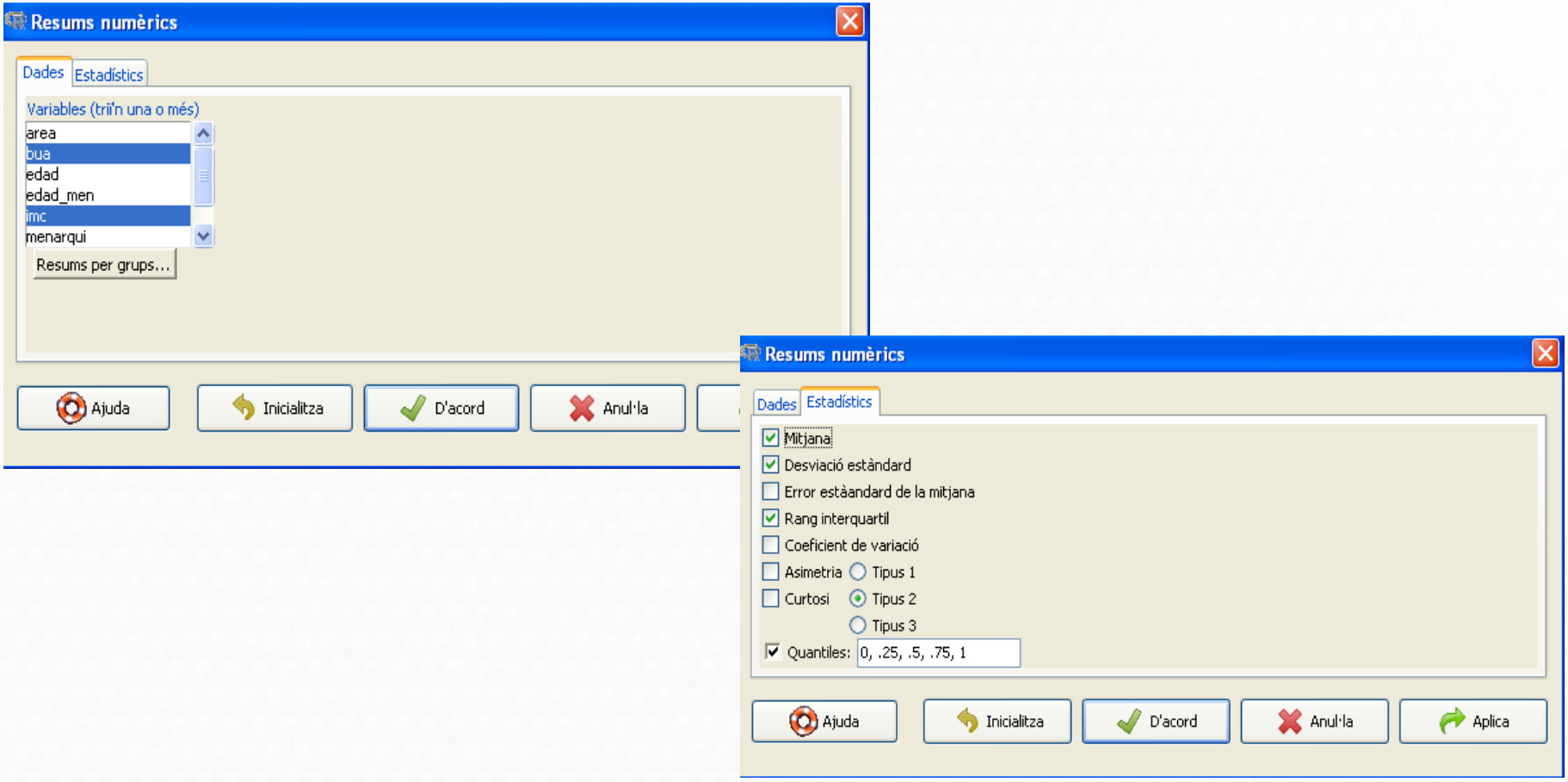
Practice. Statistics with Rcmdr

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



Practice. Statistics with Rcmdr

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



Practice. Statistics with Rcmdr

Vall d'Hebron Institut de Recerca

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

Executar

Resultat

```
> 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)

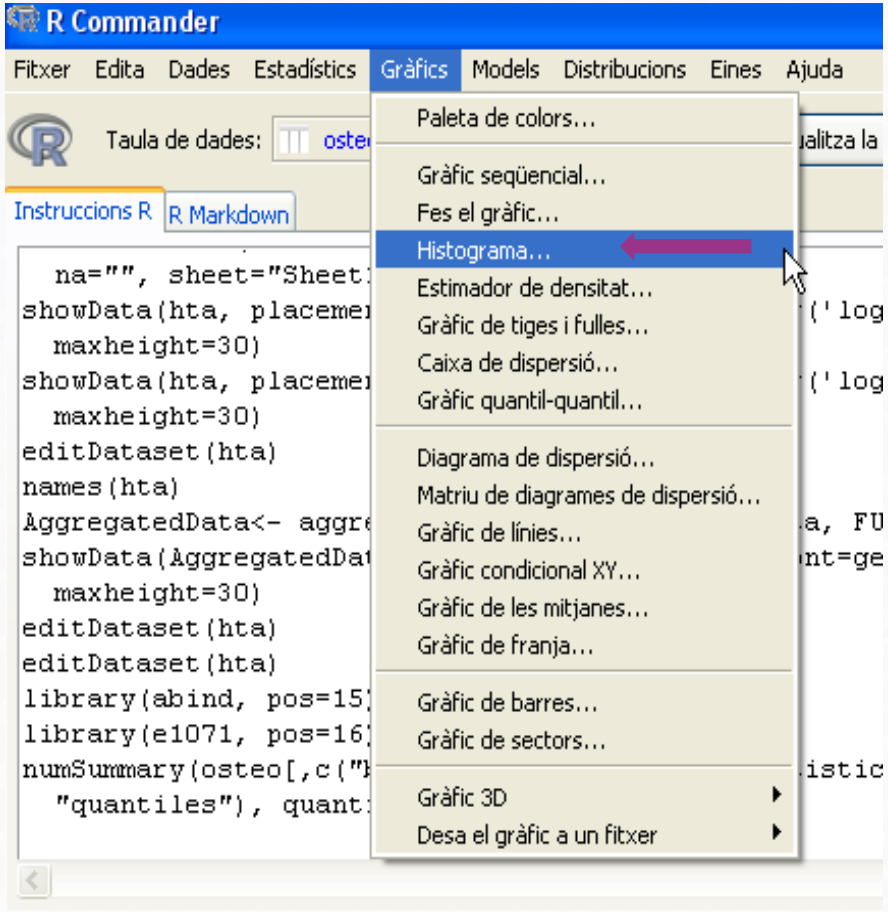
> library(e1071, pos=16)

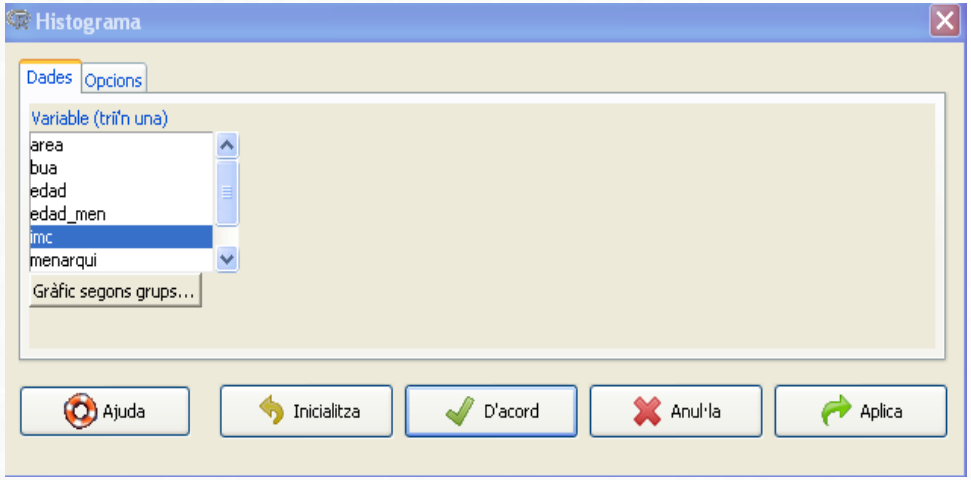
> numSummary(osteo[,c("bua", "imc", "peso")], statistics=c("mean", "sd", "IQR",
+   "quantiles"), quantiles=c(0,.25,.5,.75,1))
        mean        sd     IQR    0%     25%   50%   75%    100%    n
bua  73.29700 16.809323 22.0000 11.00 62.0000 72.00 84.00 136.00 1000
imc  28.10776  4.717925  6.0225 17.21 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

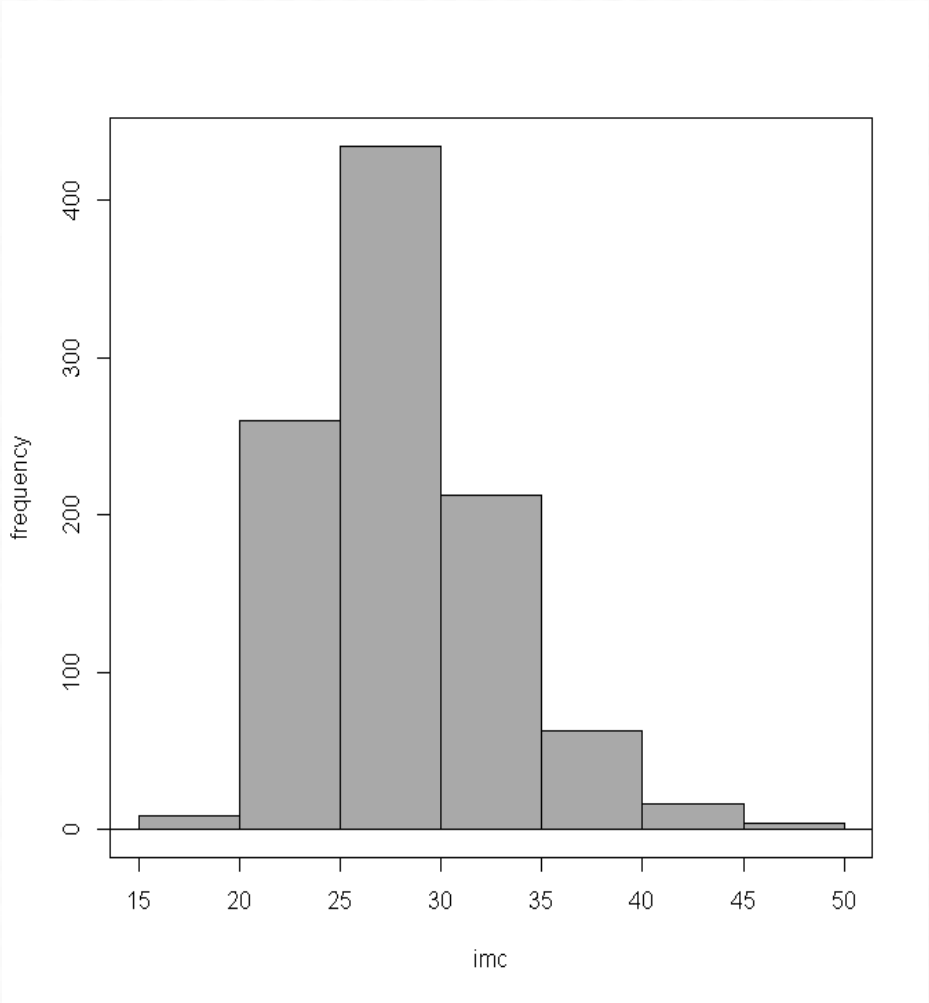
We want to see how the variable *imc* is distributed among the individuals





Practice. Graphics with Rcmdr

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



Practice. Running scripts with Rcmdr

Vall d'Hebron
Institut de Recerca

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

Instruccions R | R Markdown

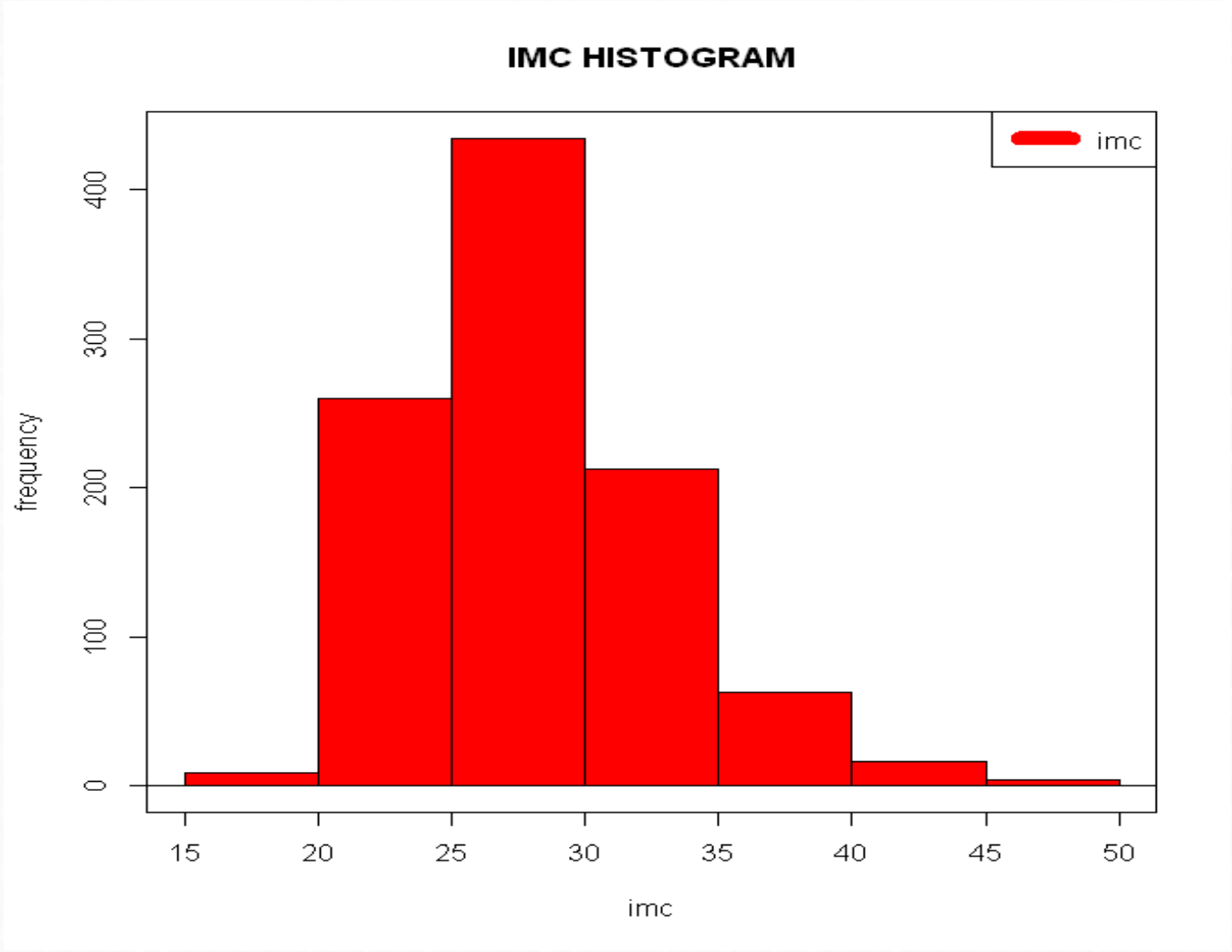
```
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
showData(hta, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
editDataset(hta)
names(hta)
AggregatedData<- aggregate(tad1 ~ grupo, data=hta, FUN=mean)
showData(AggregatedData, placement='-20+200',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

```
idth=80,

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

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
```

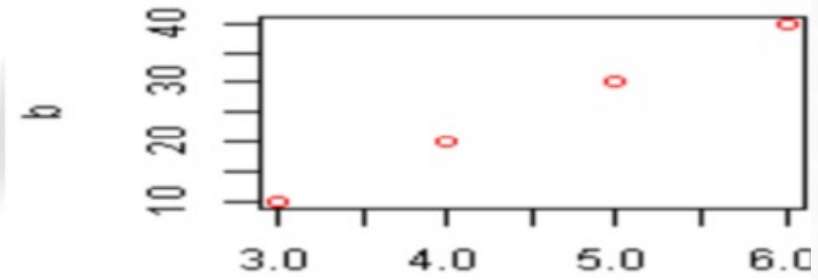
- media de b

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

```
[1] 25
```

- grafico a b

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



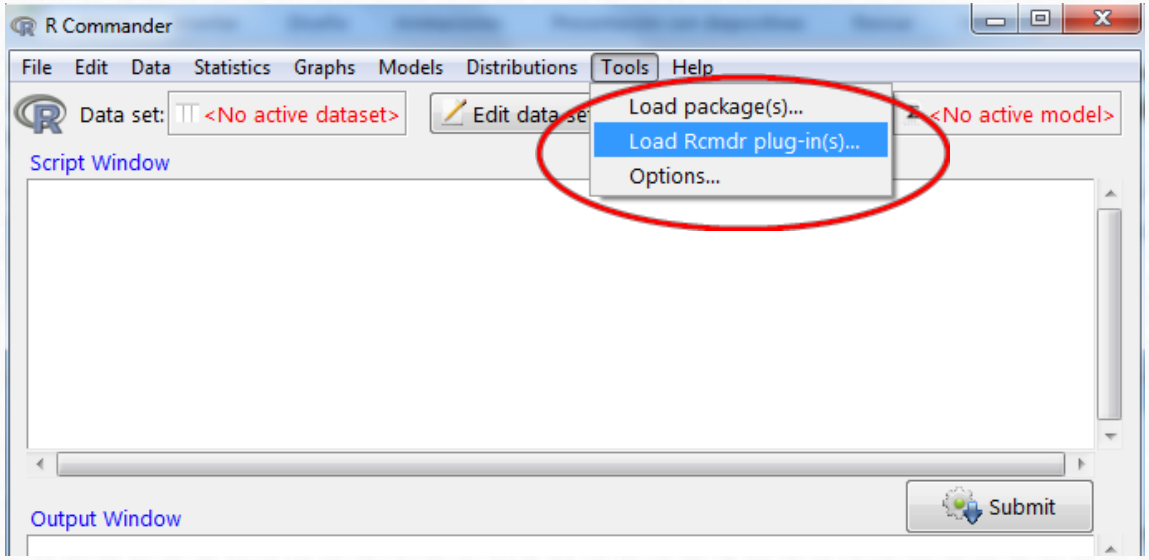
Exercise

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

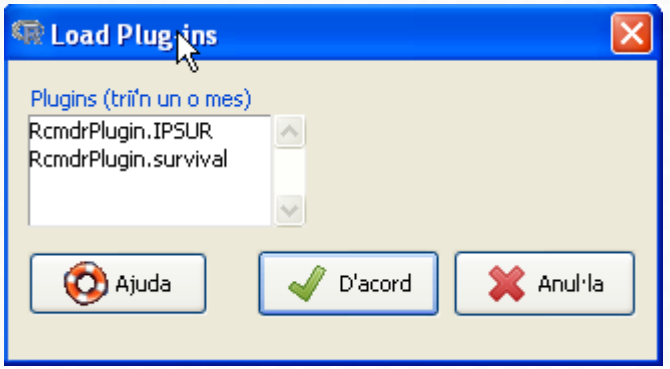
Rcommander's plugins installation

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…



Additional souRces

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