



AULA USO DO R

O que é o R?

*"A chave para entender o R é que trata-se de uma linguagem.
Uma linguagem para manipular objetos."*

Bill Venables

- Um ambiente de programação
- **Código aberto: projeto colaborativo** (sob os termos da GNU da Free Software Foundation General Public License em forma de código fonte)
- Voltado para desenvolvimento e implementação pelo usuário.
- Ele compila e roda em uma ampla variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS.
- Grande variedade de ferramentas para manipulação de dados, cálculos e gráficos
- Alto padrão com atualização constante
- Comunidade ativa e crescente



Por que se chama R?

É a letra inicial do nome de dois de seus autores :

Robert Gentleman and **R**oss Ihaka

VANTAGENS



Ele fornece uma ampla variedade de técnicas estatísticas

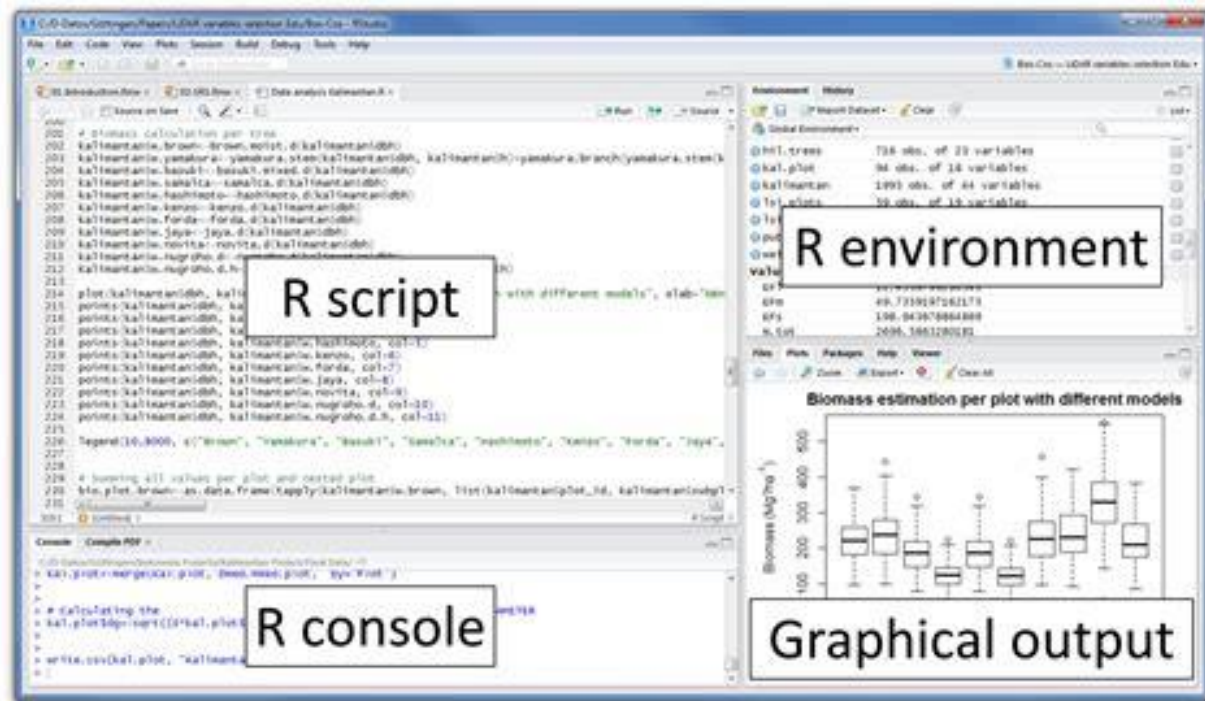
- ✓ modelos lineares e não lineares,
- ✓ testes estatísticos clássicos,
- ✓ análise de séries temporais,
- ✓ técnicas gráficas.

INSTALAÇÃO DO RSTUDIO



- <https://www.rstudio.com/>

RStudio é um [software livre](#) de [ambiente de desenvolvimento integrado](#) para [R](#) fazendo o uso deste mais fácil





INSTALAÇÃO DO R

O instalador encontra-se no portal: <http://www.r-project.org/>

- i) Acessar o site
- ii) Clicar em “CRAN MIRROR”
- iii) Clicar no link que se encontra em
Brazil - University of So Paulo, Sao Paulo
- iv) Clicar no link que representa seu sistema operacional

[Linux](#)

[MacOS X](#)

[Windows](#)

- v) Clicar em “base”
- vi) Download **R version ?**
- vii) Aguardar o download finalizar



INSTALAÇÃO

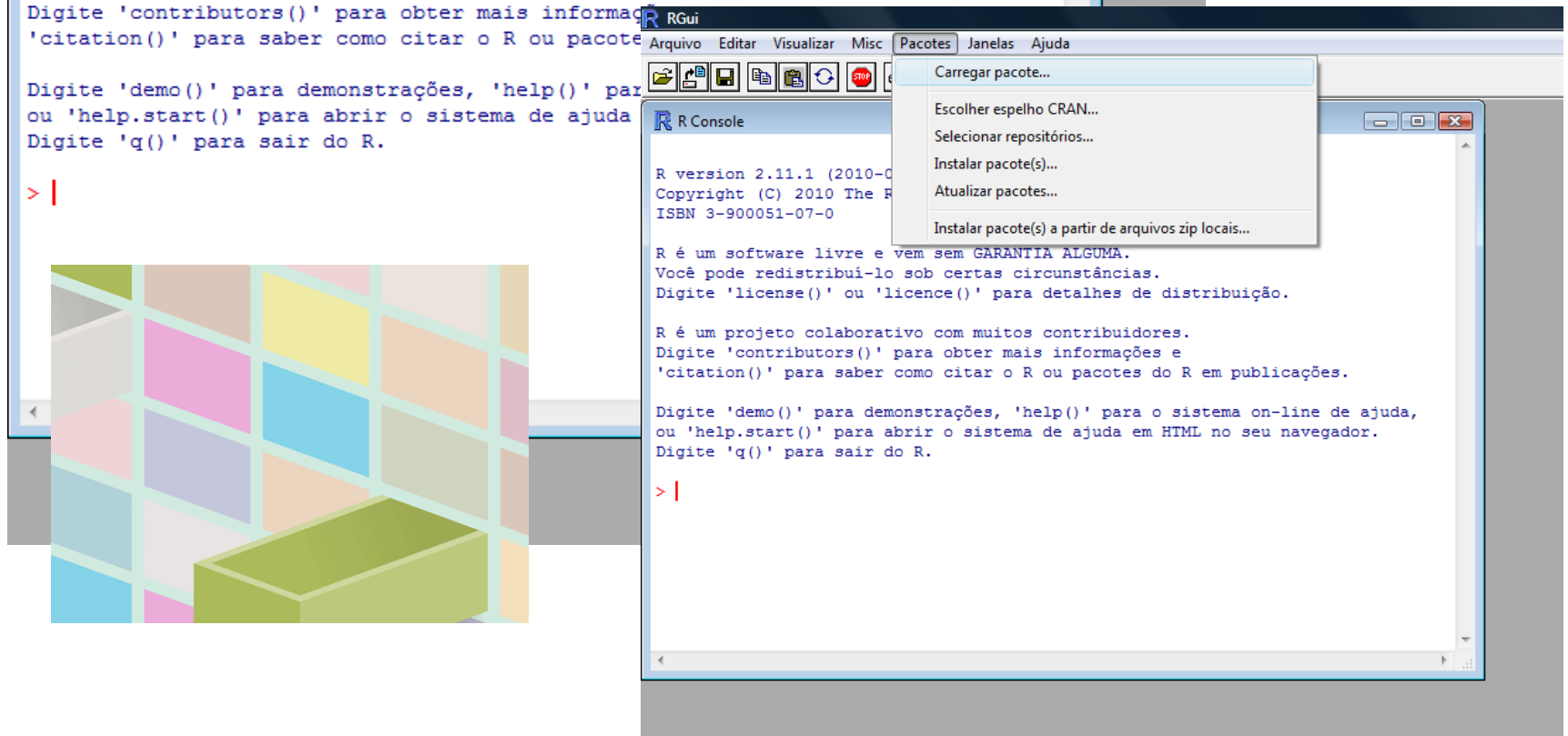
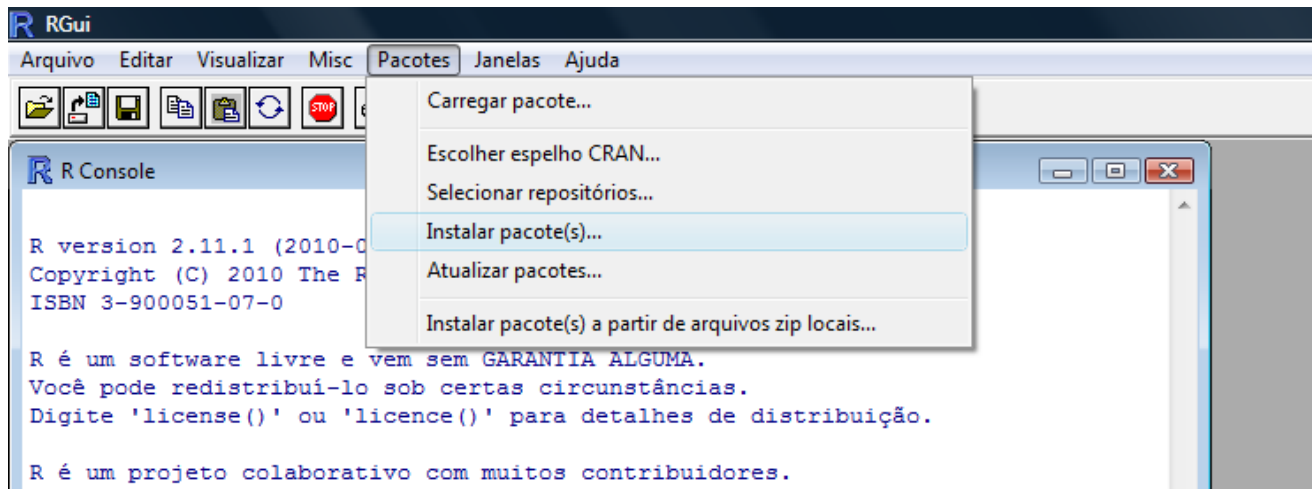
- i) Clicar no arquivo baixado
 - ii) Executar
 - iii) Caso abra uma janela de proteção, clicar em sim
 - iv) Selecionar o idioma (Português, Brasil)
 - v) Abrirá o Assistente de instalação, basta clicar em avançar até concluir
- Pronto, o R já está instalado na sua máquina, basta executá-lo.



PACOTES

O R funciona com um grupo de pacotes (Packages) alguns já formam parte da base e outros precisam ser instalados.





RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

```
1 attach(esteira)
2 require(rgl)
3 require(car)
4 library(rgl)
5 scatter3d( VO2 ~ IMC + carga |1, data = esteira)
6
7 m1<-lm(formula = VO2 ~ IMC + carga, data = esteira)
8 summary(m1)
9
```

43:17 (Top Level) R Script

Console Terminal Background Jobs

R 4.2.3 · ~/MAE0217/

R version 4.2.3 (2023-03-15 ucrt) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[workspace loaded from ~/MAE0217/.RData]

> |

Environment History Connections Tutorial

Import Dataset 37 MiB

R Global Environment

Data

apiclus1	183 obs. of 39 variables
apiclus2	126 obs. of 40 variables
apipop	6194 obs. of 37 variables
apisrs	200 obs. of 39 variables
apistrat	200 obs. of 39 variables
CI.out	4 obs. of 5 variables
data	1620 obs. of 2 variables

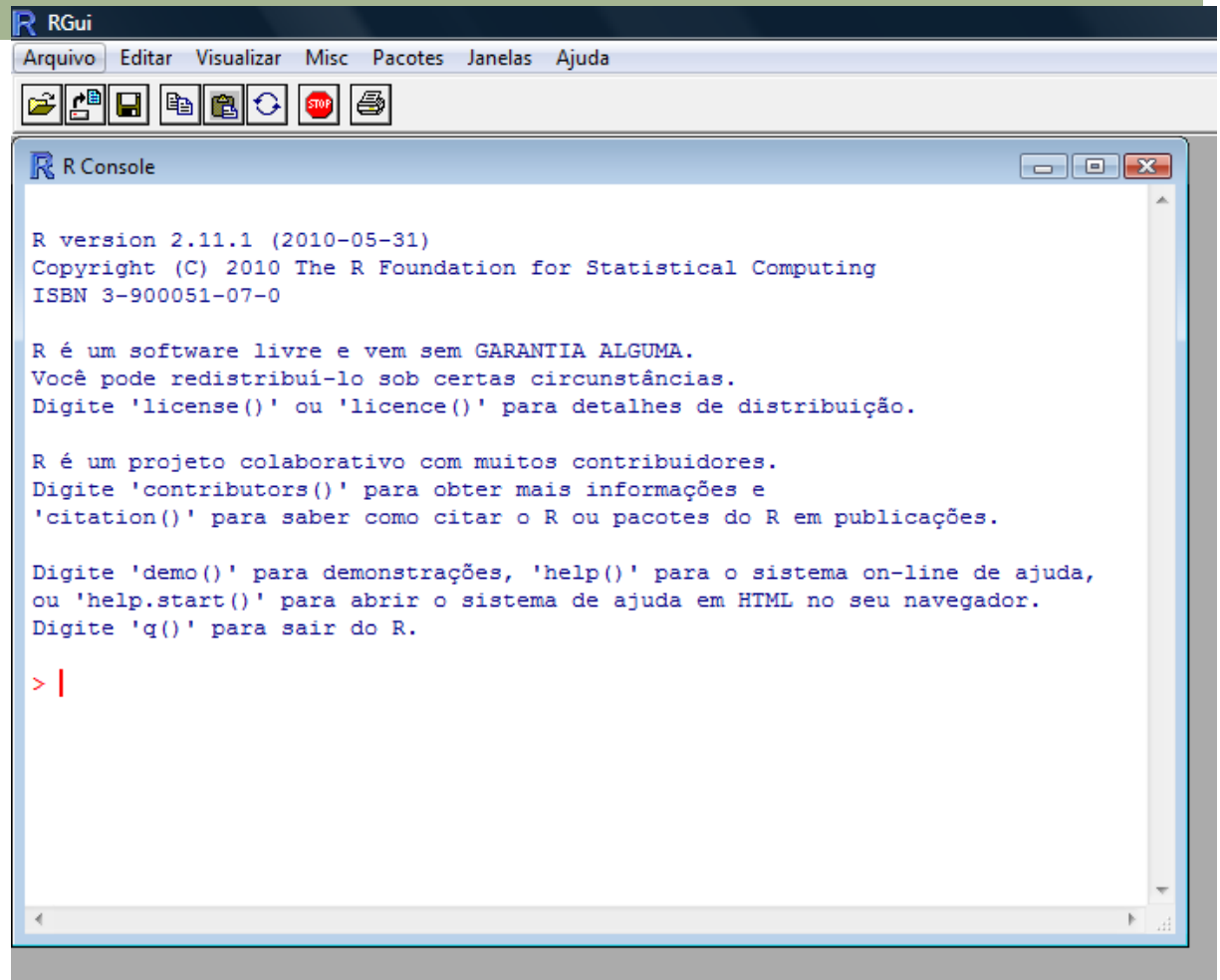
Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-5
<input type="checkbox"/> adabag	Applies Multiclass AdaBoost.M1, SAMME and Bagging	4.2
<input type="checkbox"/> ade4	Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences	1.7-19
<input type="checkbox"/> adehabitatLT	Analysis of Animal Movements	0.3.26
<input type="checkbox"/> adehabitatMA	Tools to Deal with Raster Maps	0.3.15
<input type="checkbox"/> ADER	Data Analysis in Ecology	1.4
<input type="checkbox"/> AEDForecasting	Change Point Analysis in ARIMA Forecasting	0.20.0
<input type="checkbox"/> AER	Applied Econometrics with R	1.2-10
<input type="checkbox"/> agridat	Agricultural Datasets	1.21
<input type="checkbox"/> amap	Another Multidimensional Analysis Package	0.8-18
<input type="checkbox"/> Amelia	A Program for Missing Data	1.8.0

PROMPT DE COMANDOS

- Ao abrir o R, você verá uma janela branca (R Console), onde podem ser digitados os comandos após o sinal **>** esse é o chamado “prompt de comandos”.



ACESSAR O MENU DE AJUDA DO R (HELP)

Para ver os arquivos de ajuda do R use `help(função)` ou `? função`.

A função para fazer anova é `aov`. Então vamos ver o help:

```
> help(aov) # abre o help sobre ANOVA
```

ou simplesmente

```
> ?aov
```

PARA PEDIR AJUDA

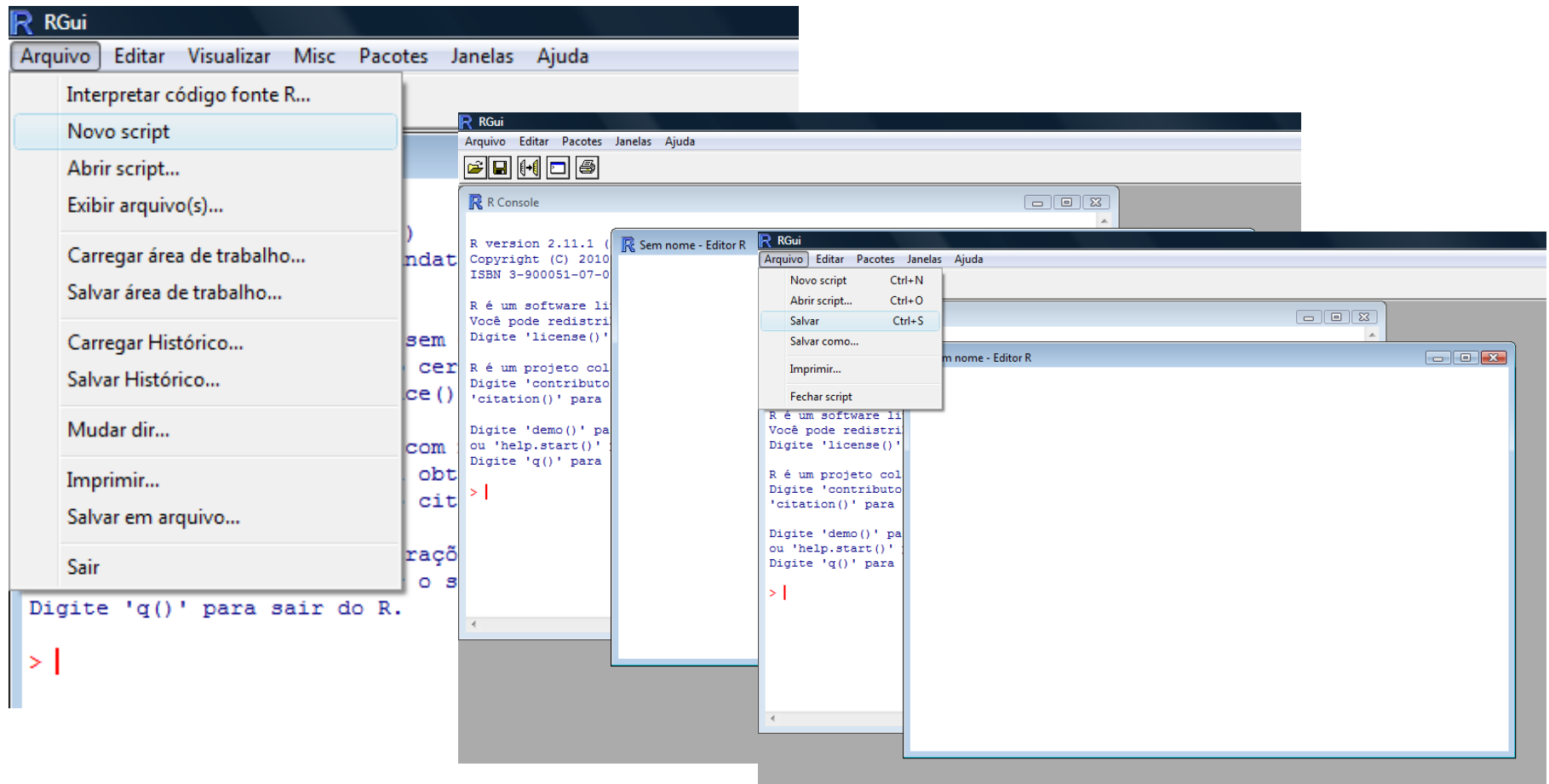
- Stack Overflow em Português é um site de perguntas e respostas para programadores profissionais e entusiastas.

<https://pt.stackoverflow.com/>



SCRIPTS

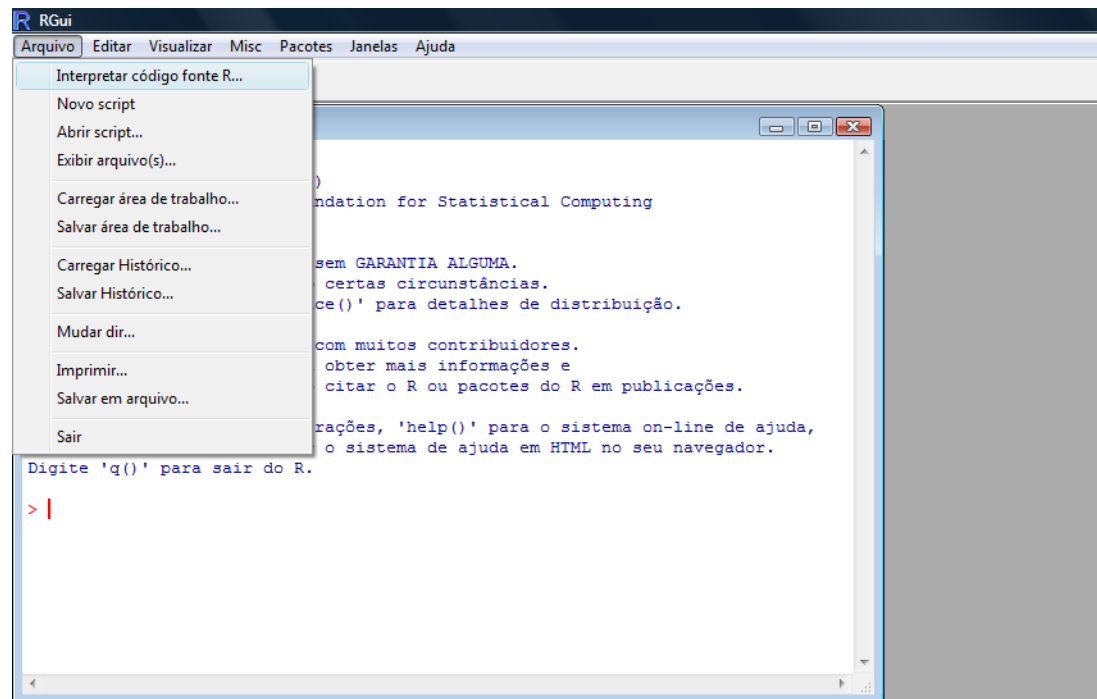
- Scripts são **arquivos** ou “**programas**” onde é salva uma lista de comandos, para não ter a necessidade de digitá-los no Prompt em cada nova sessão.
- Um script é um arquivo txt.
- Com o script você facilmente faz alterações e correções, além de salvar o script e poder refazer rapidamente suas análises após algum tempo.



SCRIPTS

Para executar um script basta digitar no prompt o comando:

```
>source ("local do arquivo")
```



ENTRADA DE DADOS

```
> Filhos = c(2,3,1,1,1,0,0,4)
```

```
> Filhos
```

```
[1] 2 3 1 1 1 0 0 4
```

```
> table(Filhos)
```

```
Filhos
```

```
0 1 2 3 4
```

```
2 3 1 1 1
```

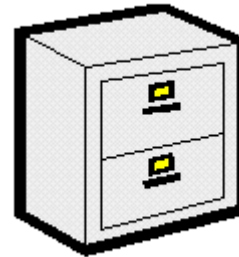


```
> attach(escore.inform)
```

```
> grupo
```

```
[1] G1 G1 G2 G2
```

```
Levels: G1 G2
```



Por fim, após a utilização do arquivo de dados, pode-se - aplica-se o detach

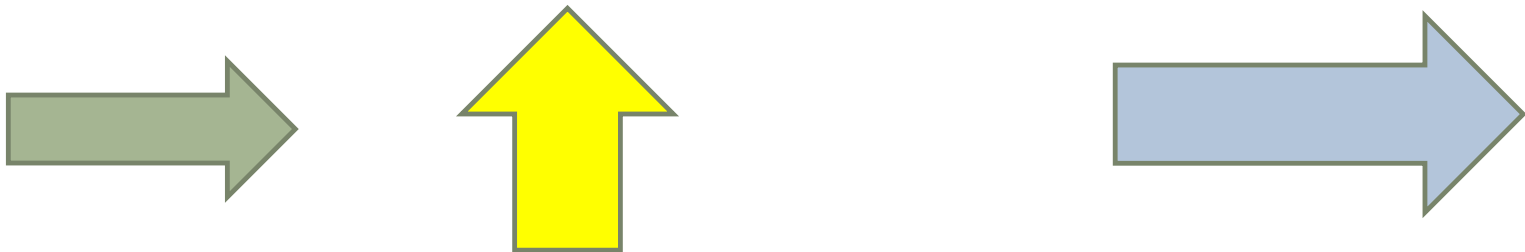
```
> detach(escore.inform)
```



RCMDR - PACOTE “R COMMANDER” — “MENUS” PARA O R

Alguns usuários acostumados com outros programas notarão de início a falta de "menus". Na medida que utilizam o programa, os usuários (ou boa parte deles) tendem a preferir o mecanismo de comandos pois é mais flexível e com mais recursos.

Entretanto, alguns iniciantes ou usuários esporádicos poderão ainda preferir algum tipo de "menu".



RCMDR - PACOTE “R COMMANDER” — “MENUS” PARA O R

O pacote Rcmdr foi desenvolvido por [John Fox](#) visando atender a esta demanda. Para utilizar este pacote basta instalá-lo e carregar com o comando `require(Rcmdr)` e o menu se abrirá automaticamente.

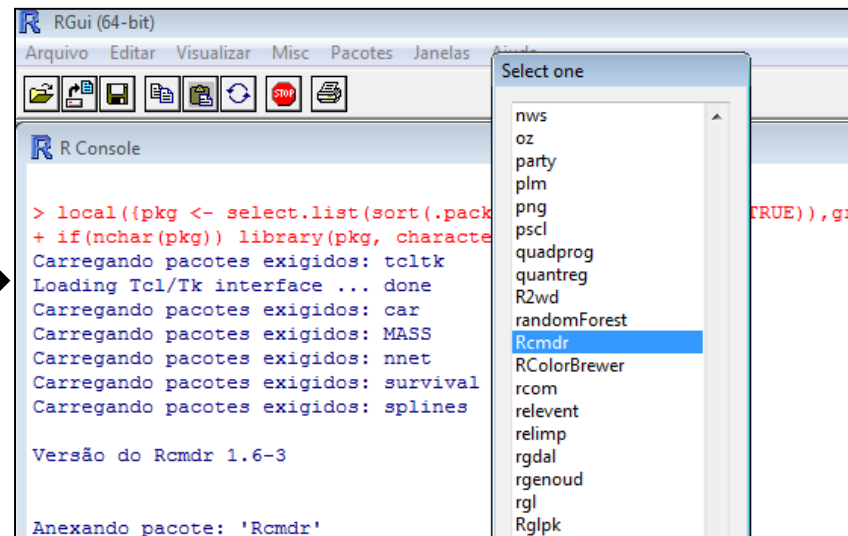
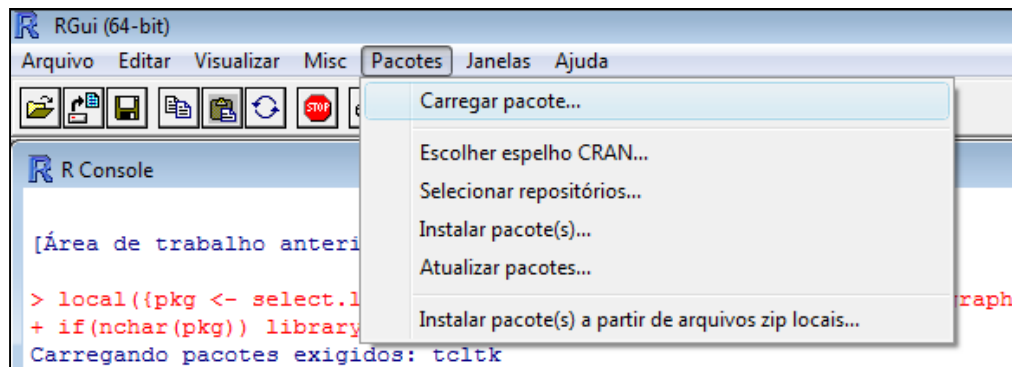
Atenção: Note que o Rcmdr não provê acesso a toda funcionalidade do R mas simplesmente a alguns procedimentos estatísticos mais usuais.

Maiores informações sobre este pacote podem ser encontradas na [página do Rcmdr](#).

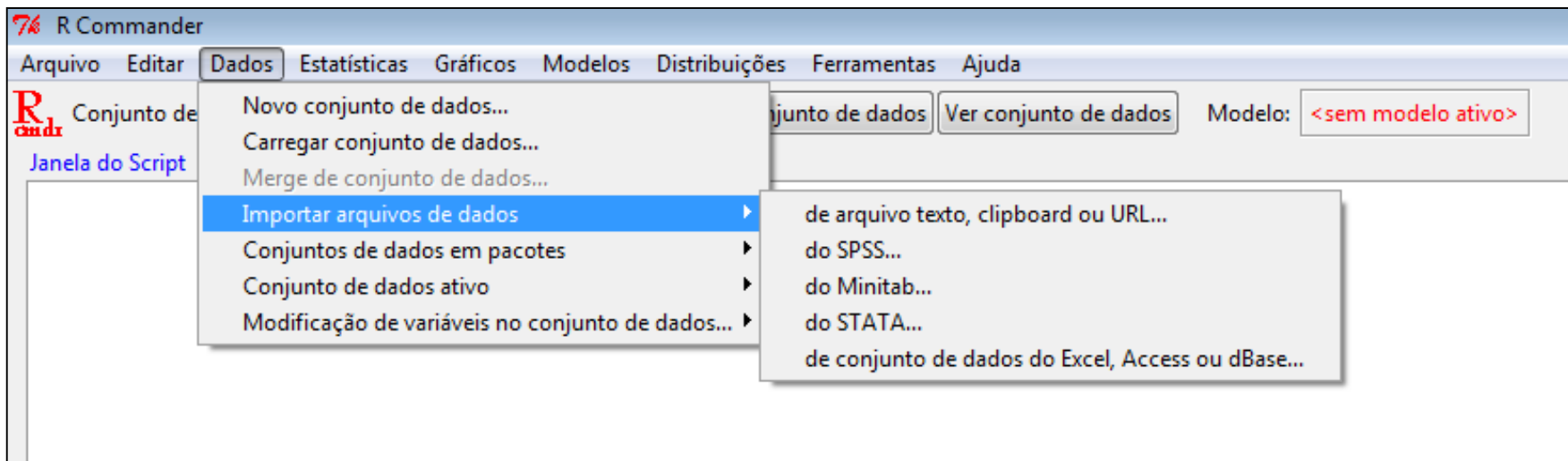
<http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/>

Arquivo : Carregando dados no R

(1) Carregar Rcmdr:

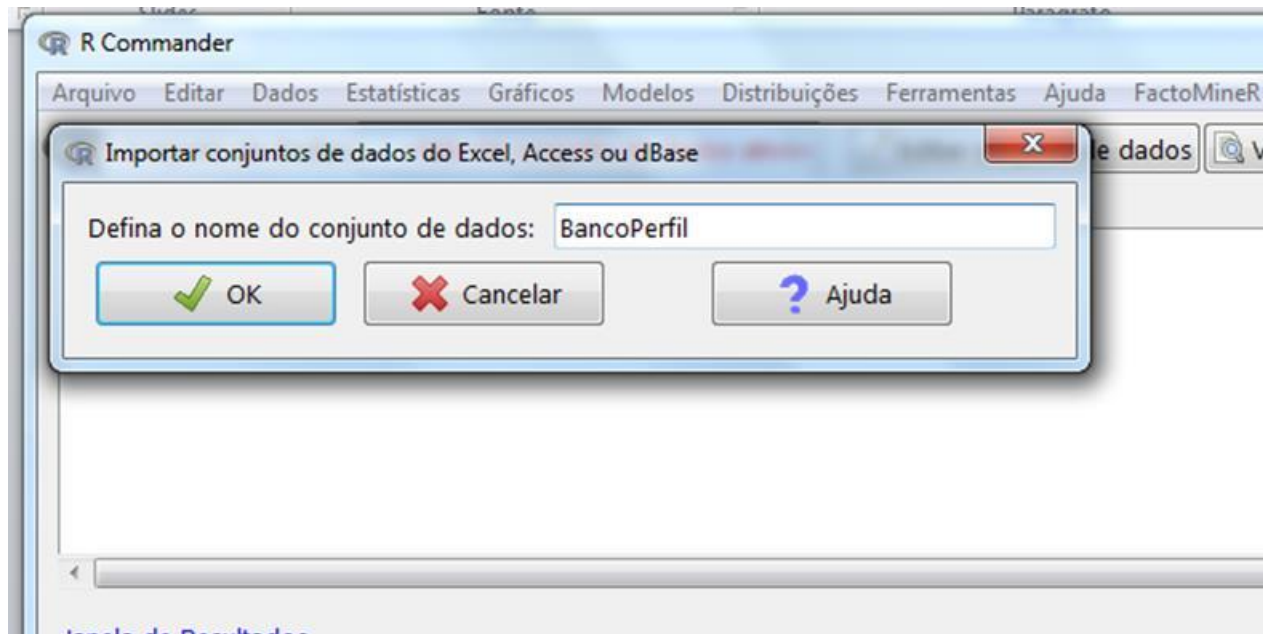


(2) Importar dados:



Arquivo Banco Perfil: Carregando dados no R

(3) Nomear o banco:



Arquivo Banco Perfil: Editar/Visualizar dados

R Commander

Arquivo Editar Dados Estatísticas Gráficos Modelos Distribuições Ferramentas Ajuda FactoMineR

Conjunto de Dados: BancoPerfil Editar conjunto de dados Ver conjunto de dados Modelo: <sem modelo ativo>

Janela do Script

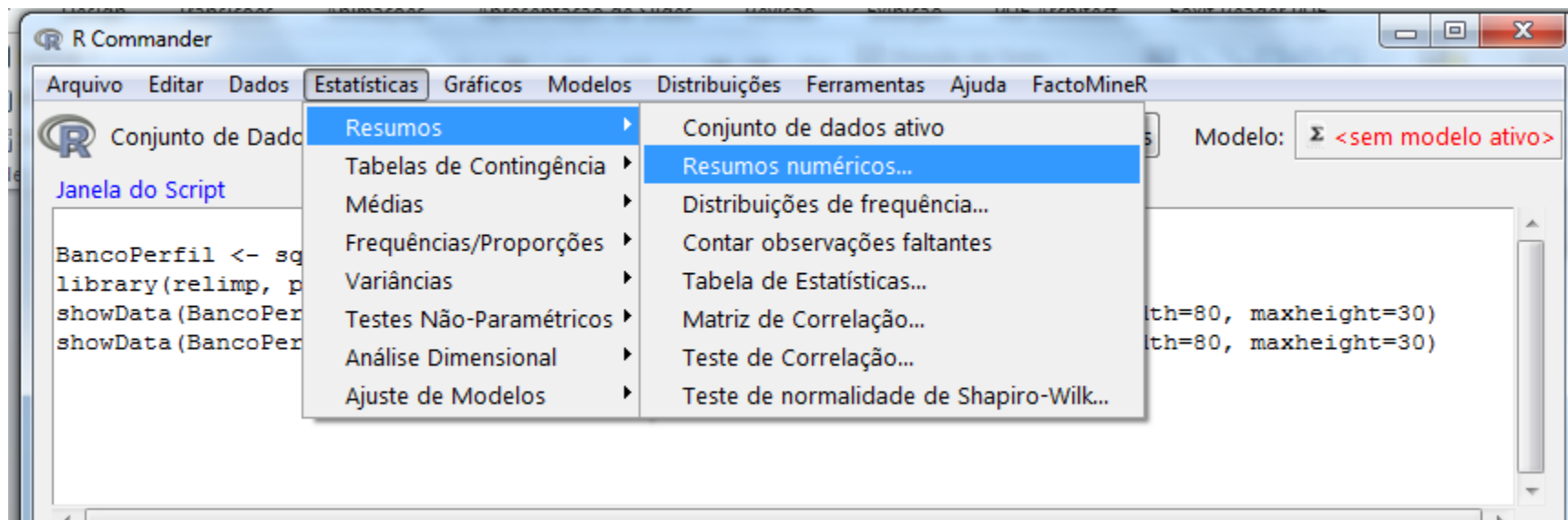
```
BancoPerf
library(r
showData(
showData(
```

Janela de Resposta

```
> BancoPe
> library
> showDat
> showDat
```

BancoPerfil

	id	data	sexo	nasc	idade	natural	classregião	nfilho	me
1	1	2011-05-25	2	1975-02-07	36	Jacobina	Nordeste	0	
2	3	2012-05-04	1	1960-09-14	51	Bambuí	Sudeste	0	
3	5	2012-05-04	2	1966-02-02	46	Campina Grande	Nordeste	0	
4	10	2012-05-04	1	1971-09-30	40	Registro	SP	0	
5	15	2012-05-04	2	1971-07-04	40	Itanhem	Nordeste	1	
6	17	2012-05-04	1	1962-02-10	50	Boa Esperança	Sudeste	3	
7	19	2012-05-04	2	1967-02-12	45	São Paulo	SP	2	
8	20	2011-05-20	2	1971-03-13	40	São José dos Pinhais	Sul	0	
9	21	2012-05-04	2	1962-11-21	49	Presidente Juscelino	Nordeste	1	
10	24	2012-05-04	2	1970-12-22	41	São Paulo	SP	0	
11	27	2012-05-04	1	1961-06-02	50	São Paulo	SP	1	
12	32	2012-05-22	2	1969-01-02	43	Diadema	SP	0	
13	34	2011-06-14	1	1975-02-25	36	Porteirinha	Sudeste	0	
14	38	2012-05-22	1	1952-10-18	59	São Paulo	SP	0	
15	41	2012-05-22	1	1970-06-24	41	União dos Palmares	Nordeste	1	
16	43	2012-05-22	1	1979-06-05	32	São Paulo	SP	0	
17	44	2012-05-22	2	1961-08-02	50	São Paulo	SP	2	
18	45	2012-05-22	1	1956-04-16	56	Ibirataia	Nordeste	3	
19	46	2012-05-25	1	1979-04-29	33	Araruama	Sudeste	0	



Estatísticas → Resumos Numéricos

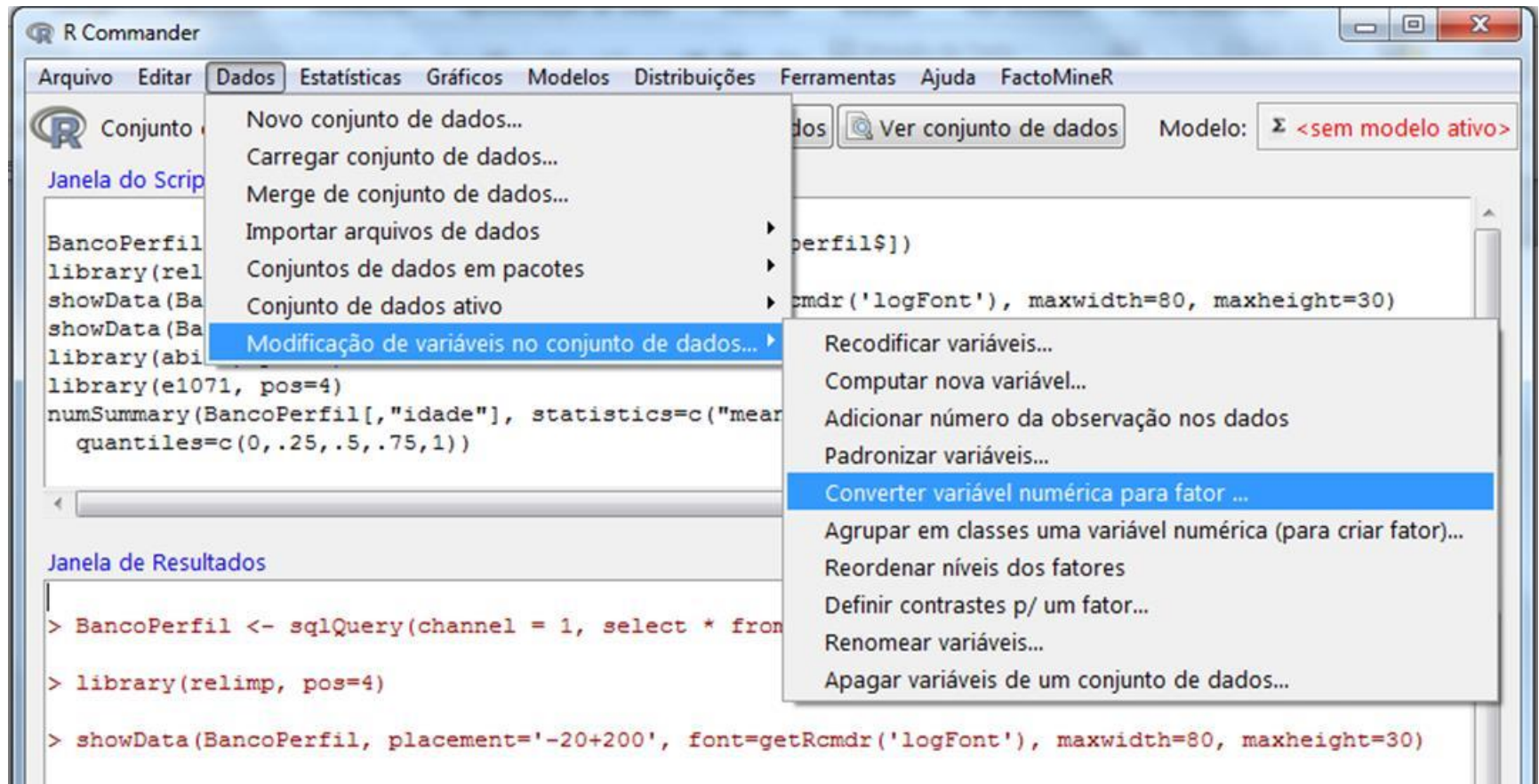
Arquivo *Banco Perfil*:

Medidas descritivas

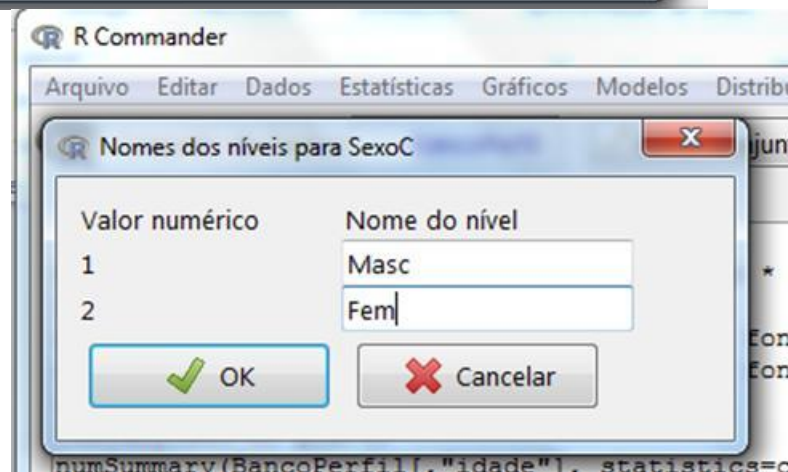
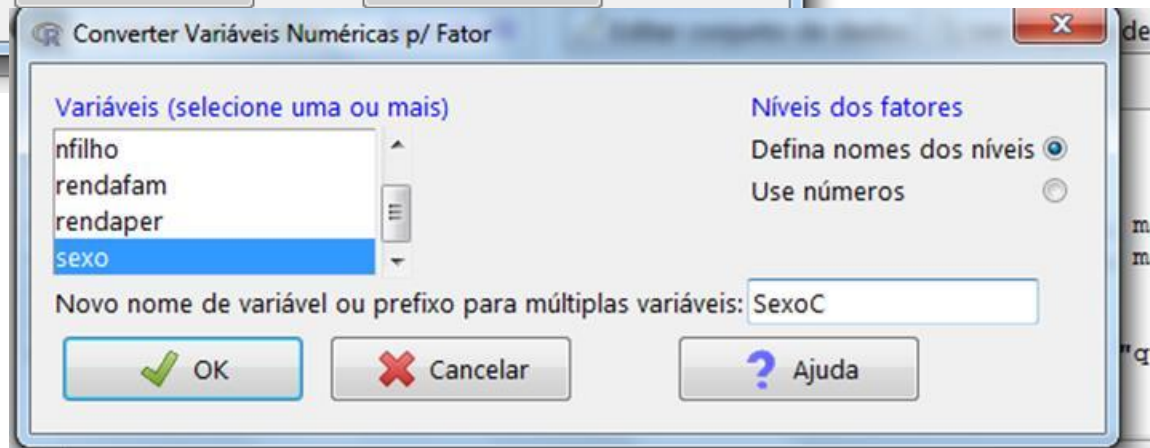
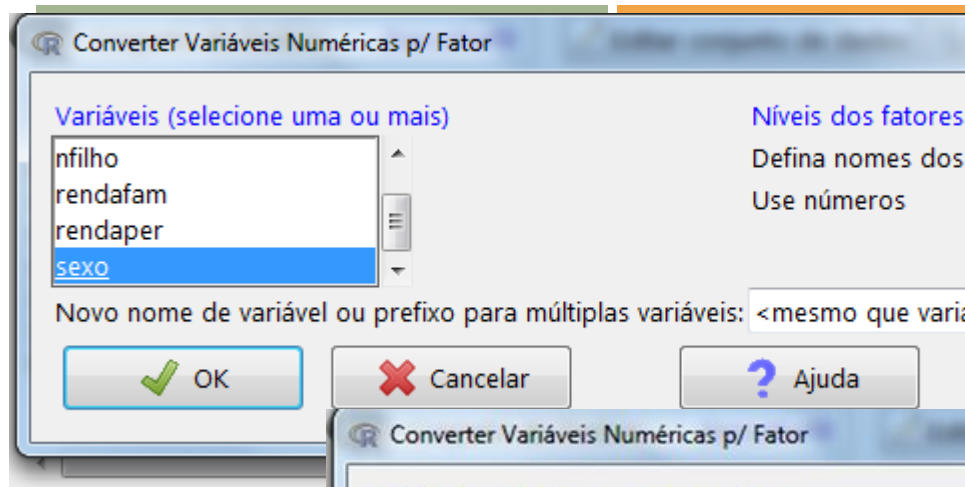
- Variável quantitativa Idade

```
> numSummary(BancoPerfil[, "idade"], statistics=c("mean", "sd", "IQR", "quantiles", "cv"),  
+   quantiles=c(0, .25, .5, .75, 1))  
      mean      sd IQR      cv 0% 25% 50% 75% 100%  n NA  
44.58427 7.133164  10 0.1599928 28  40  45  50   59 89  3
```

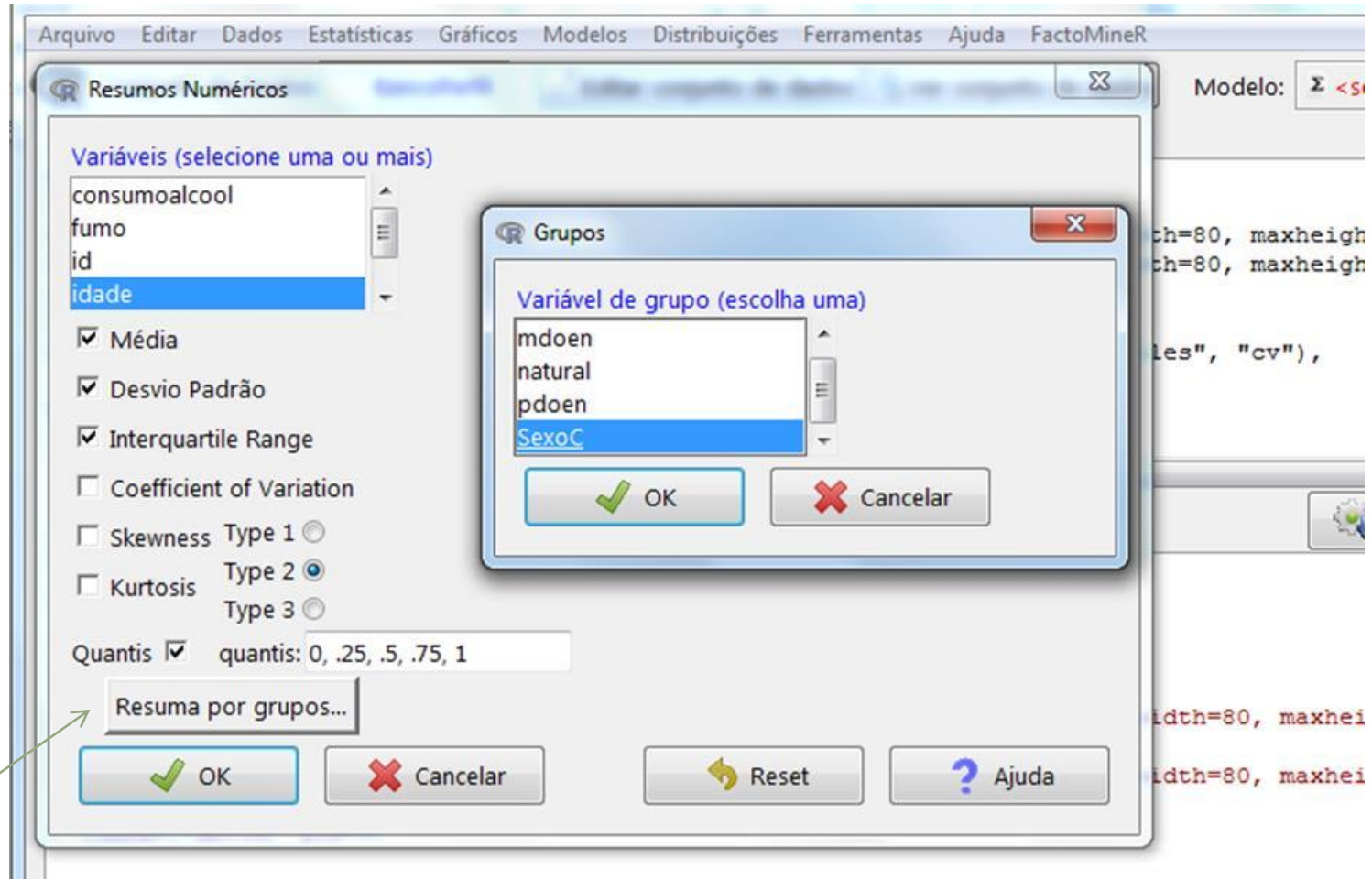
Recodificando



Recodificando



Medidas descritivas por Sexo



Estatísticas → Resumos Numéricos → Resuma por grupos

Medidas descritivas por Sexo

```
> numSummary(BancoPerfil[, "idade"], groups=BancoPerfil$SexoC, statistics=c("mean", "sd", "IQR",  
+ "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Masc	44.64286	7.950823	11.5	28	38.75	45	50.25	59	56
Fem	44.48485	5.590847	9.0	36	40.00	44	49.00	55	33

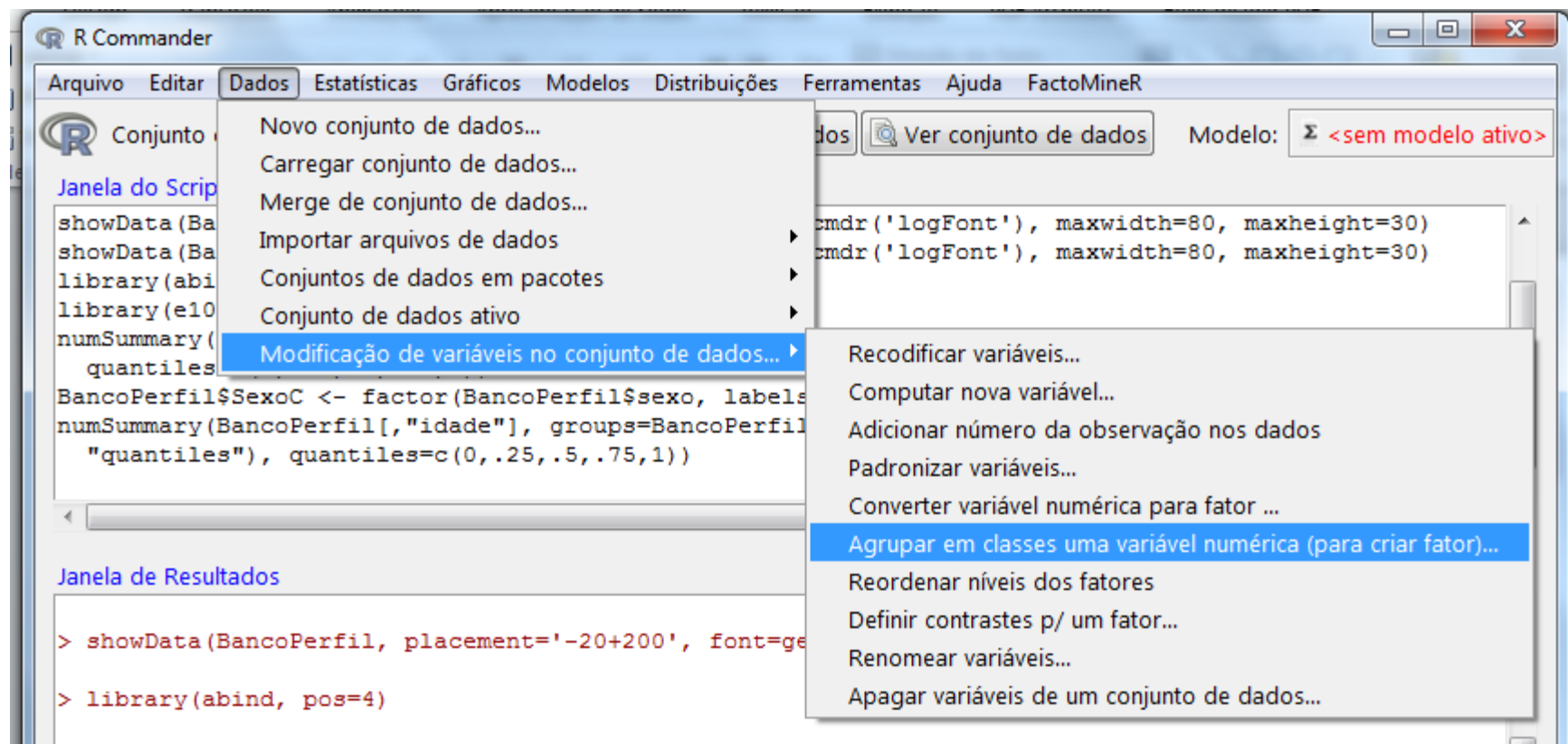
Exemplo I:

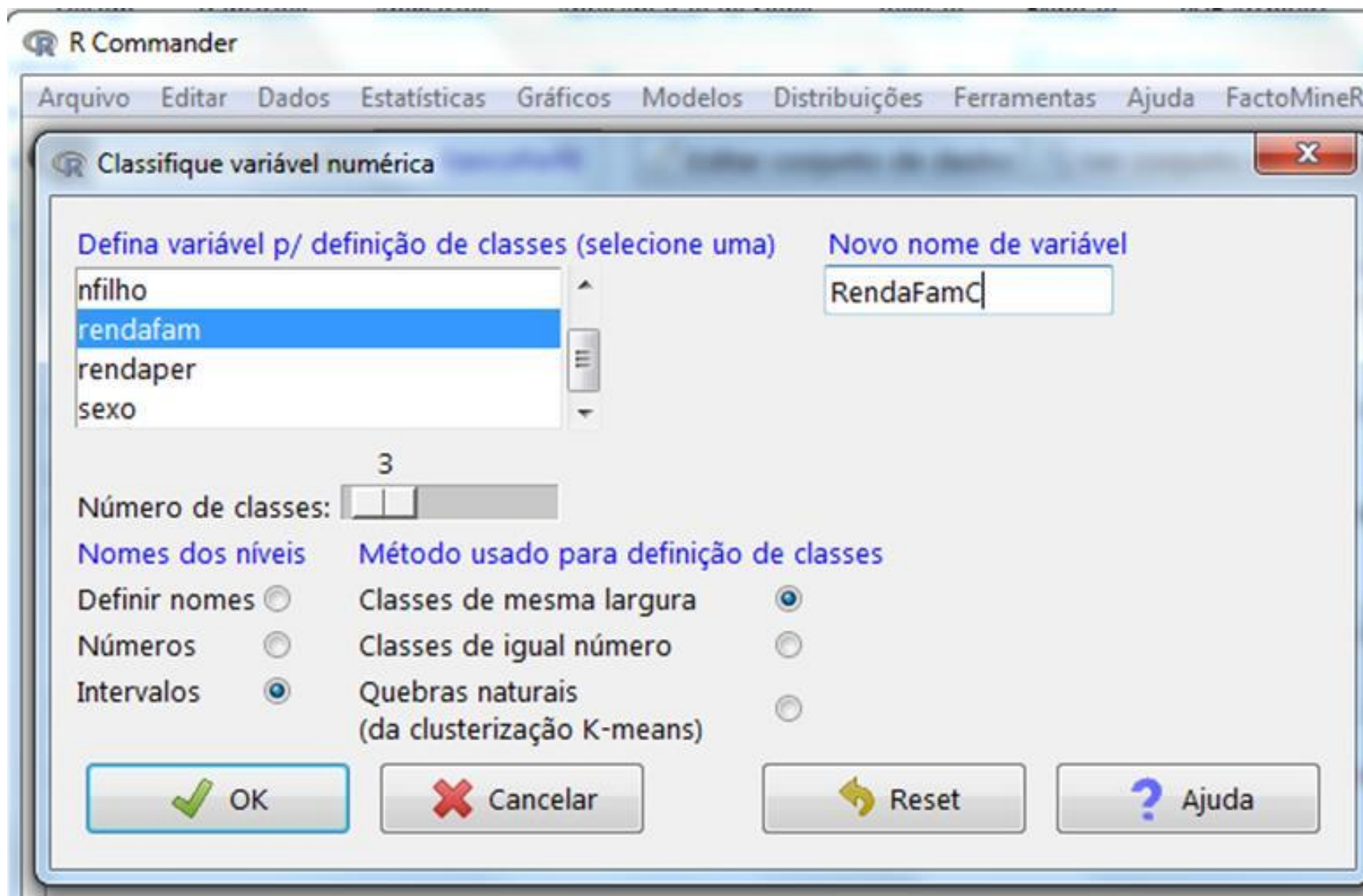
Renda Familiar

→ Variável
quantitativa →

Construir
intervalos
de classe

(I) Criar nova variável

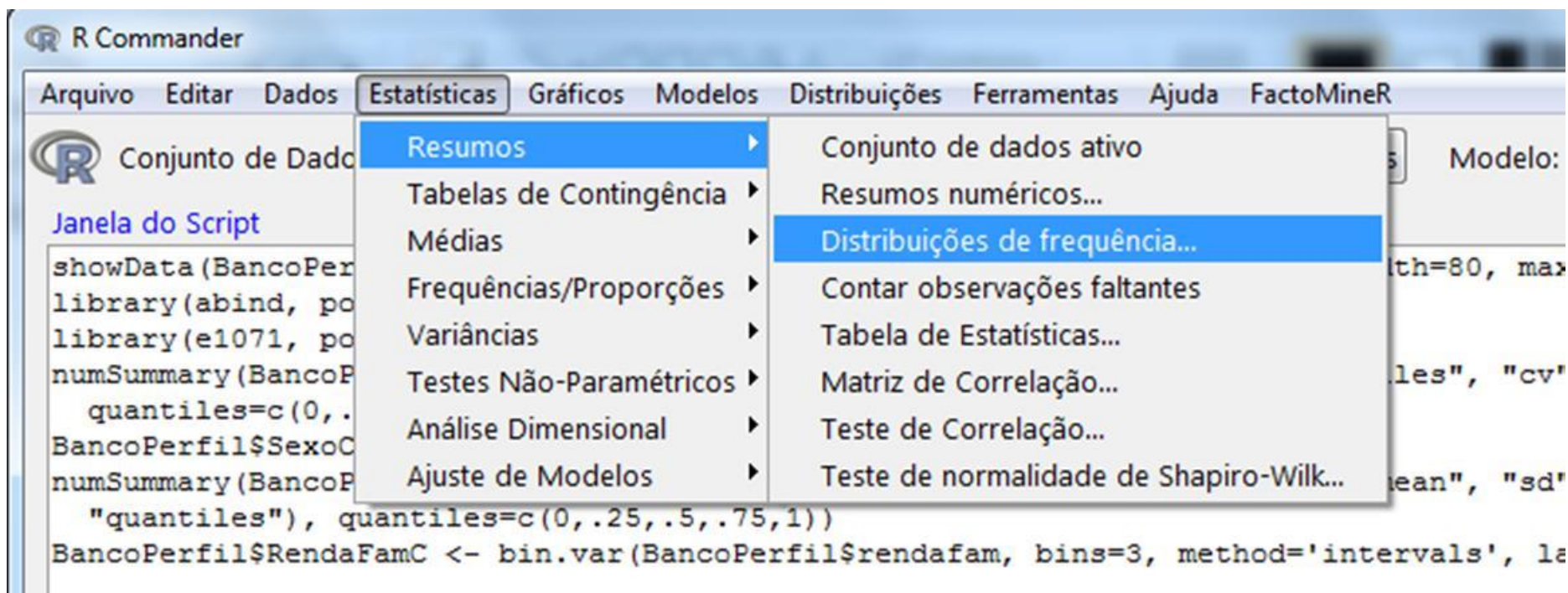




Exemplo 1:

Rcmdr:

(2) obter a distribuição de frequências da nova variável



(2) obter a distribuição de frequências da nova variável

```
> .Table <- table(BancoPerfil$RendaFamC)

> .Table  # counts for RendaFamC

      (486,5.33e+03] (5.33e+03,1.02e+04] (1.02e+04,1.5e+04]
                75                8                5

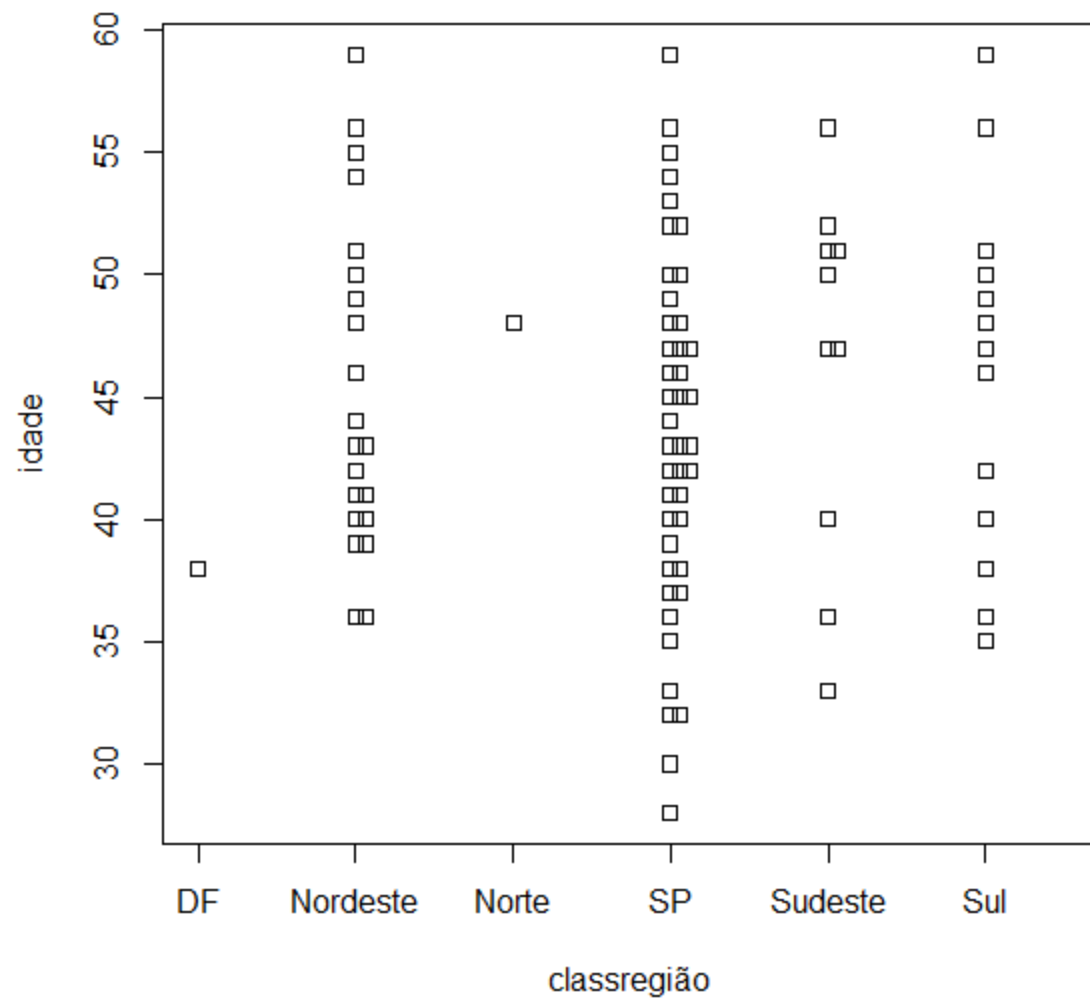
> round(100*.Table/sum(.Table), 2)  # percentages for RendaFamC

      (486,5.33e+03] (5.33e+03,1.02e+04] (1.02e+04,1.5e+04]
                85.23                9.09                5.68

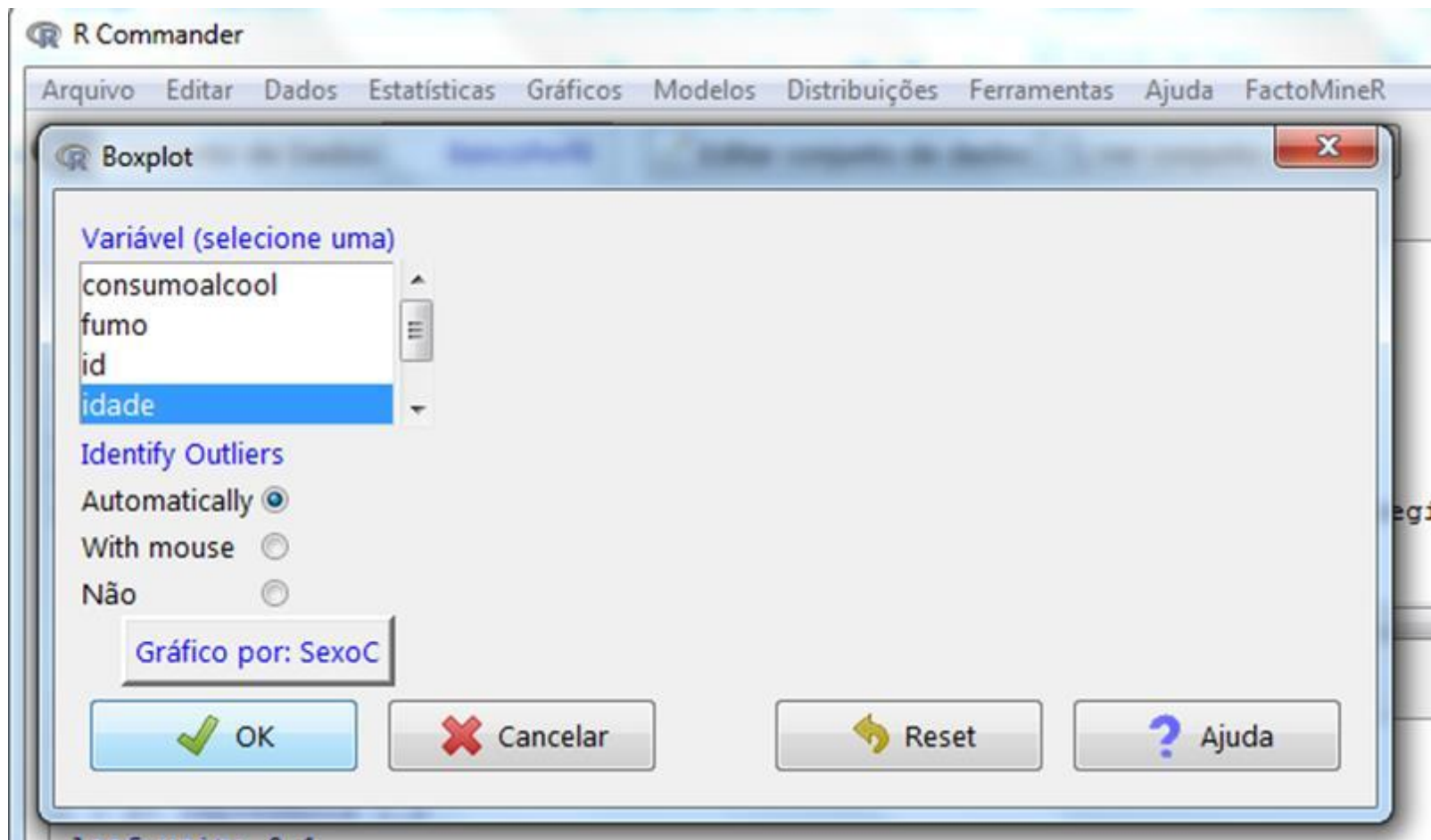
> remove(.Table)
```

Gráficos mais comuns para variáveis quantitativas

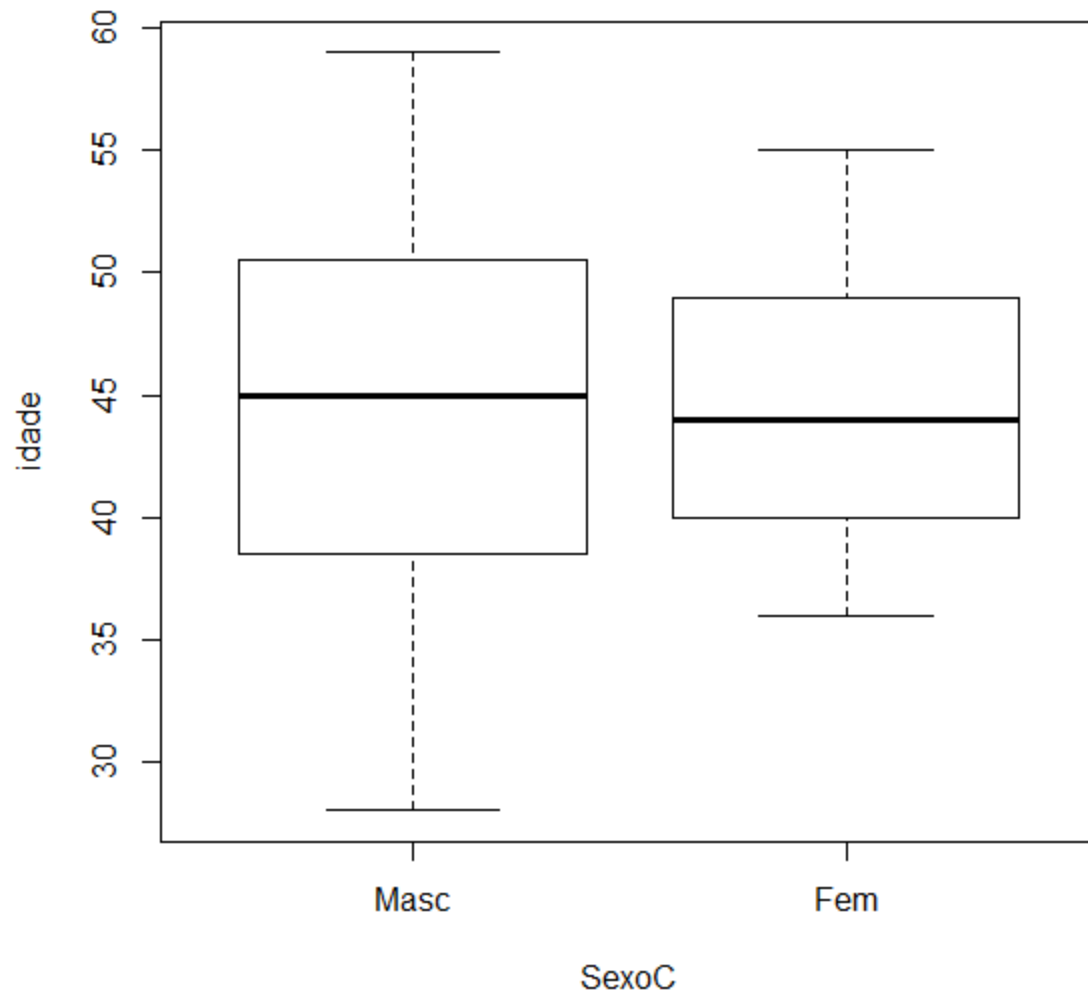
- “Strip Chart” ou “Dotplot”
- “Boxplot”
- Histograma



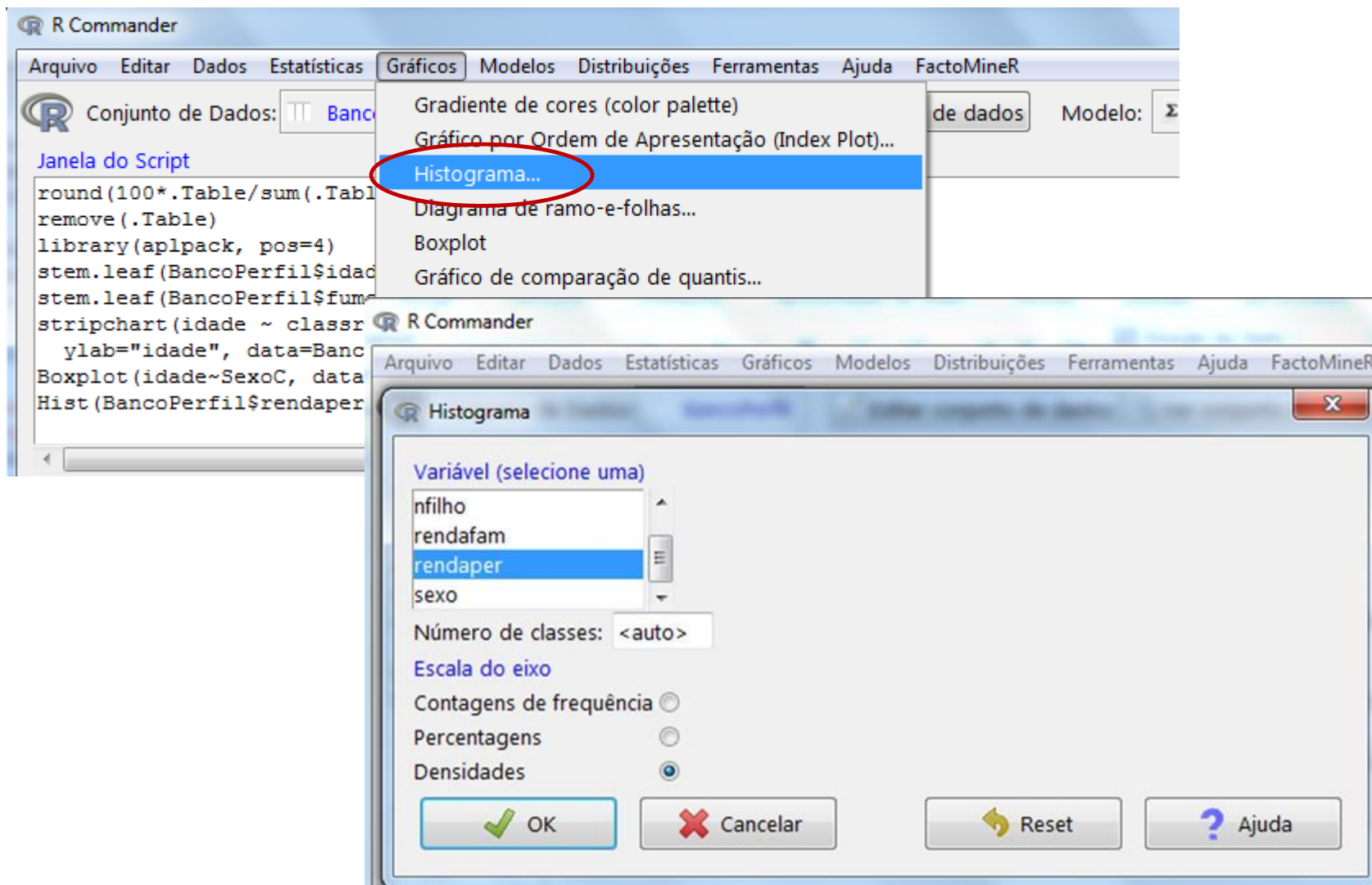
Boxplot

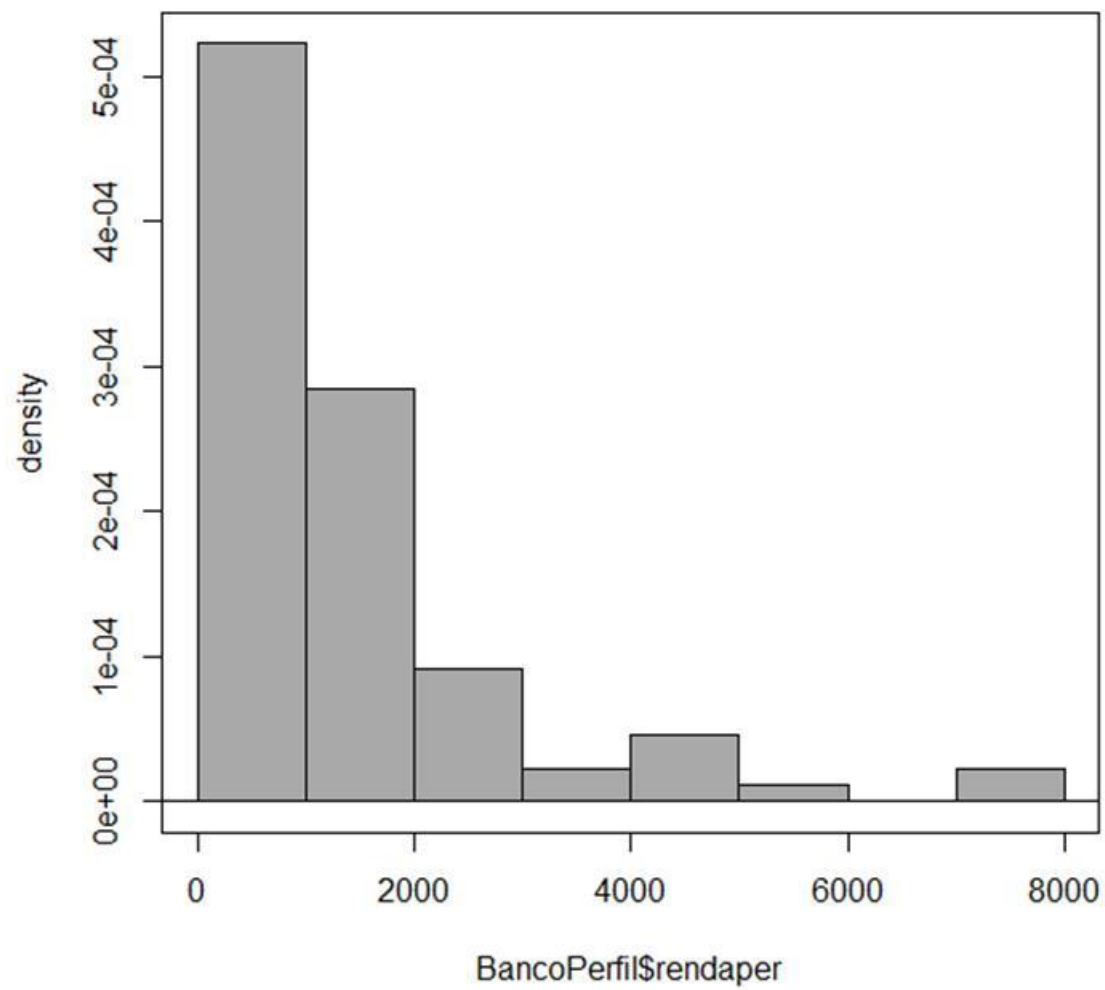


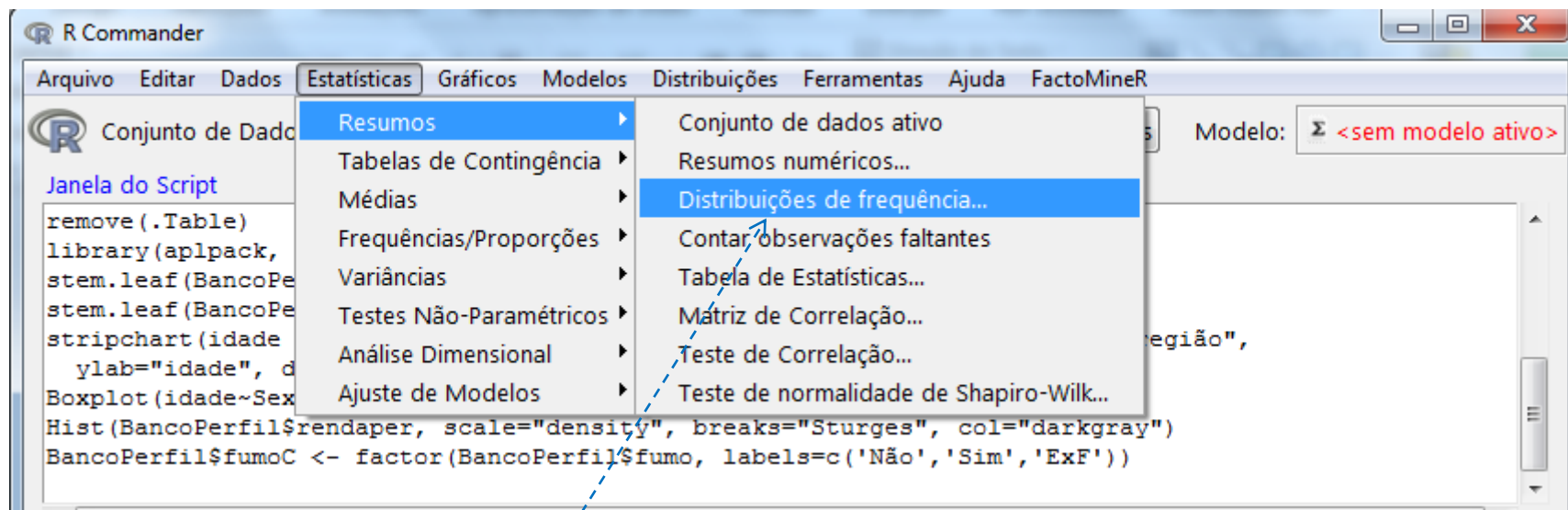
Boxplot



Gráficos → Histograma...

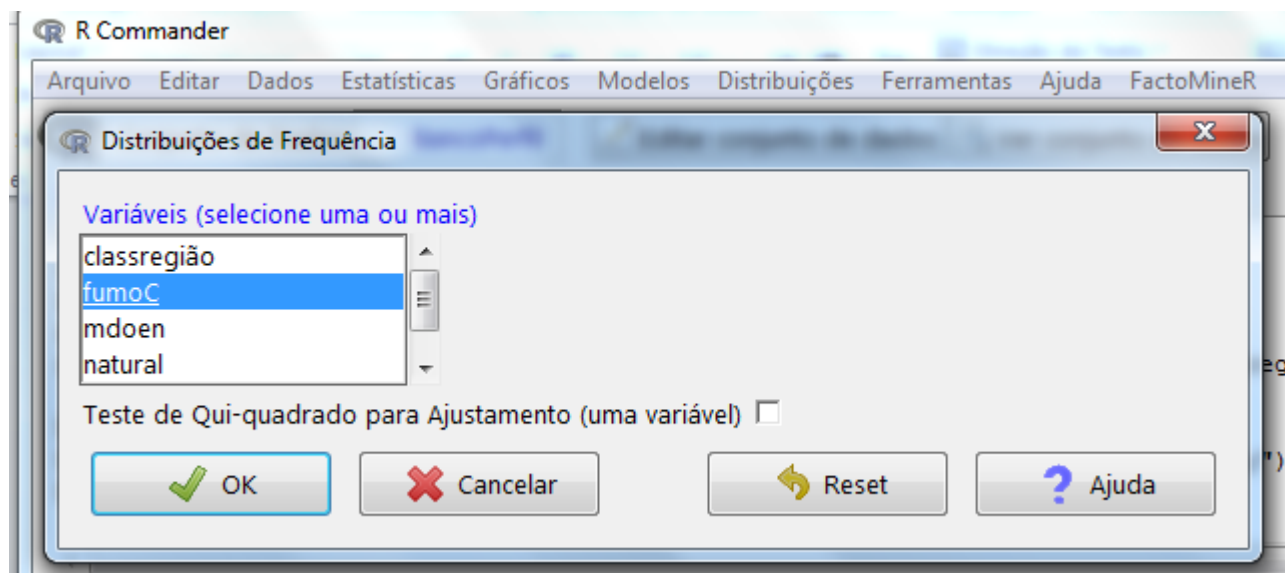






Cálculo de frequências e porcentagens

Variáveis qualitativas



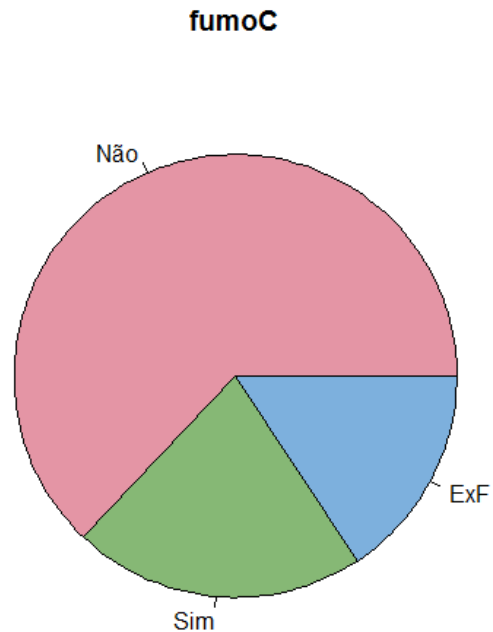
```
> .Table # counts for fumoC  
Não Sim ExF  
56 19 14  
  
> round(100*.Table/sum(.Table), 2) # percentages for fumoC  
Não Sim ExF  
62.92 21.35 15.73  
  
> remove(.Table)
```

Variáveis Qualitativas

Gráficos

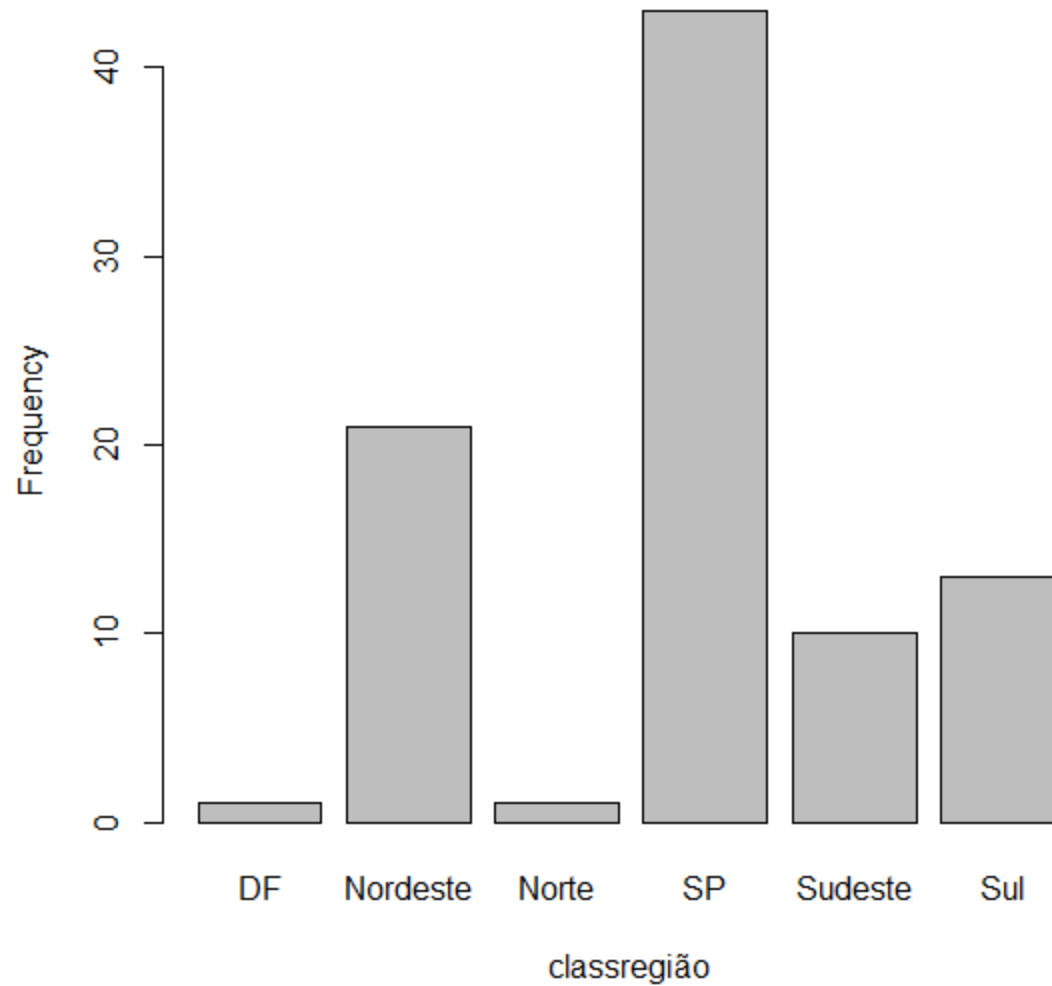
- Gráfico de setores
- Gráfico de barras

Gráfico de setores



Gráficos →
Gráfico de Pizza

Gráfico de barras



```
riqueza <- c(15,18,22,24,25,30,31,34,37,39,41,45)
area <- c(2,4.5,6,10,30,34,50,56,60,77.5,80,85)
area.cate <- rep(c("pequeno", "grande"), each=6)

par(las=1)
plot(riqueza~area, las=3)
```

Em geral, a informação que vem por último é a informação que o R vai tomar como verdadeira. Por exemplo, las controla a direção das legendas dos eixos (las= 1, legendas escritas sempre na horizontal, las=3, legendas sempre na vertical), sejam os números da escala ou o nome do eixo.

```
par(mfrow=c(1,2))
plot(riqueza~area)
boxplot(riqueza~area.cate)

hist(riqueza, breaks = c(10,20,25,30,50), freq = FALSE, col=c("pink"),
     main = "Histograma")
```

```
demo(graphics)
```