



# 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 :

**R**obert 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

RStudio interface showing R code, workspace, and a faceted bar chart.

```

5- #####
6 setwd("C:\\Users\\Chris Rooney\\Documents\\My Dropbox\\CodeRepo\\R\\Vast")
7 pollData <- read.csv("data/poll.csv")
8 pollData <- na.omit(pollData)
9
10 # Plot a graph showing the submission of each participant
11 library(ggplot2)
12 library(scales)
13 library(gridextra)
14 library(xtable)
15 weeklyResults <- function(theData, column, type, interval, result) {
16   guessData <- subset(theData, theData$Name != result)
17   hline.result <- data.frame(z = column[theData$Name==result], Film = u
18   column.subset <- column[theData$Name!="Result"]
19   film.subset <- theData$Film[theData$Name!="Result"]
20   hline.mean <- aggregate(column.subset, list(Film=film.subset), mean)
21   hline.median <- aggregate(column.subset, list(Film=film.subset), medi
22   ggplot(data=guessData, aes_string(x="Name", y=type, fill="Name")) +
23     geom_bar(stat="identity", position="dodge", colour="black") +
24     facet_wrap(~Film) +
25
190:1 (Untitled) R Script

```

Workspace:

Data	Dimensions	Type
center	2048x3	double matrix
e1	2048x3	double matrix
e2	2048x3	double matrix
grey50	50 obs. of 4 variables	
lines	101 obs. of 5 variables	
mtcars	32 obs. of 11 variables	
pollData	206 obs. of 5 variables	
quine	146 obs. of 5 variables	
trans	3x3	double matrix
variable	50 obs. of 4 variables	
Values		
TwoN		numeric[101]
TwoTwoN		numeric[101]

Console:

```

> #
> # 01/03/13 #
> #
> #####
> printPlots(pollData, "2013-03-01", "no", 10, 1)
>
> #####
> #
> # 08/03/13 #
> #
> #####
> printPlots(pollData, "2013-03-08", "no", 10, 1)
> printoverviewTable(pollData, "2013-03-08")
Error: could not find function "printoverviewTable"
>
> #####
> #
> # 08/03/13 #
> #
> #####
> printPlots(pollData, "2013-03-15", "no", 10, 1)
> printoverviewTable(pollData, "2013-03-15")
Error: could not find function "printoverviewTable"
>
> createPerformanceTables(pollData, c("2013-02-22", "2013-03-01", "2013-03-08", "2013-03-15"))
>

```

Plots:

Figure: Faceted bar chart showing Rating (Y-axis, 0 to 8) versus Name (X-axis) for four movies: 21 and Over, Jack the Giant Slayer, Stoker, and The Last Exorcism Part II. The chart displays individual ratings for 15 participants (Bec, Bob, Chris, Efe, Irene, Jane, Kai, Leishi, Neesha, Peter, Phong, Rick, Simon, Tinni, Yang) across the four movies. A horizontal line indicates the median rating for each movie. The legend on the right identifies the participants by color.

# R-STUDIO

The image displays the R Studio interface with four main components highlighted by text boxes:

- R script:** The editor window on the left contains R code for data analysis, including loading data, creating a plot, and fitting models.
- R console:** The bottom-left window shows the execution of the R script, with output messages such as "Loading all values per plot and nested plot" and "Writing csv(kal.plot, 'kalteam...".
- R environment:** The top-right window shows the current environment with variables like `kal.trees` (738 obs. of 23 variables), `kal.plot` (84 obs. of 18 variables), `kal.teaman` (1993 obs. of 44 variables), `kal.plots` (39 obs. of 18 variables), `cal`, `calm`, `calm2`, and `calm3`.
- Graphical output:** The bottom-right window displays a box plot titled "Biomass estimation per plot with different models". The y-axis is labeled "Biomass (kg/ha)" and ranges from 100 to 600. The plot shows several box plots representing different models, with individual data points overlaid as small circles.



# 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

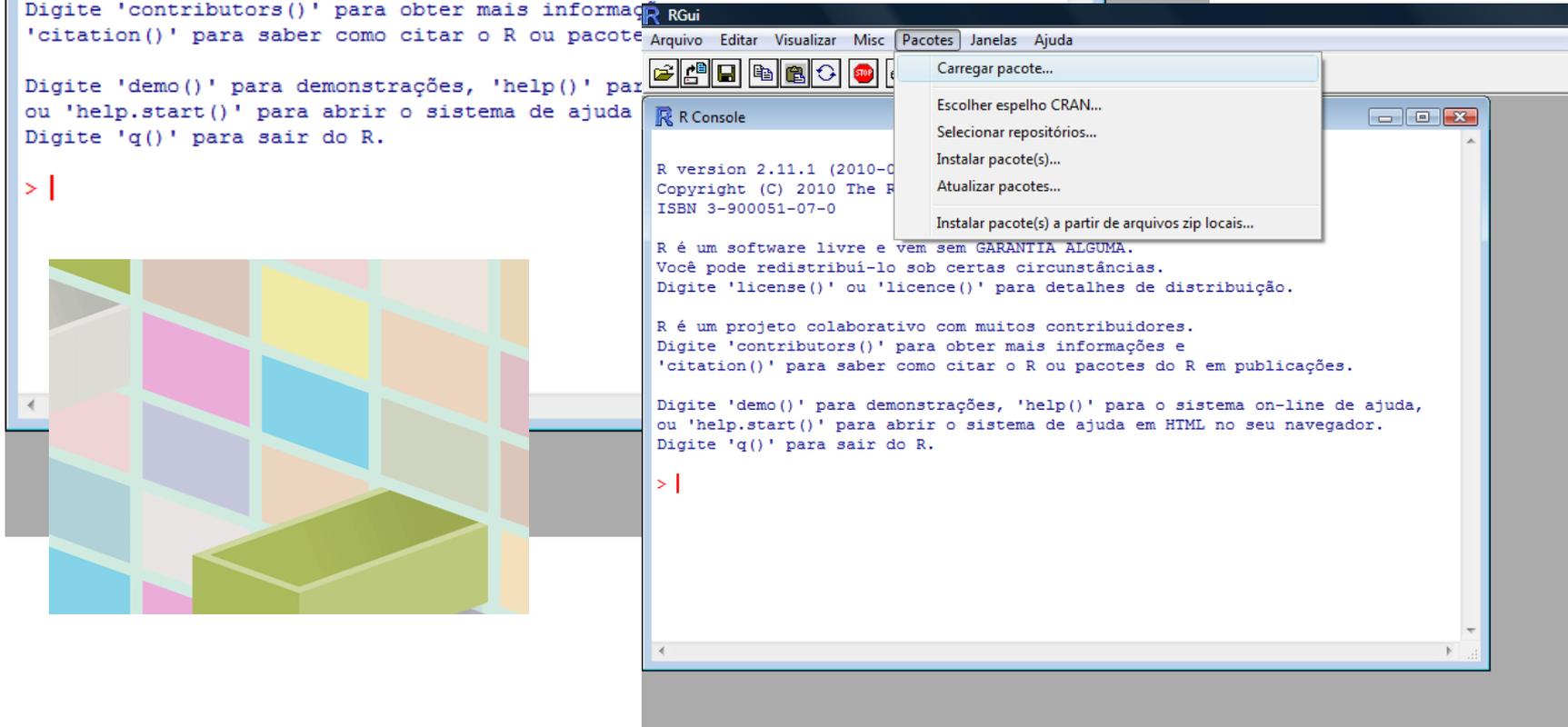
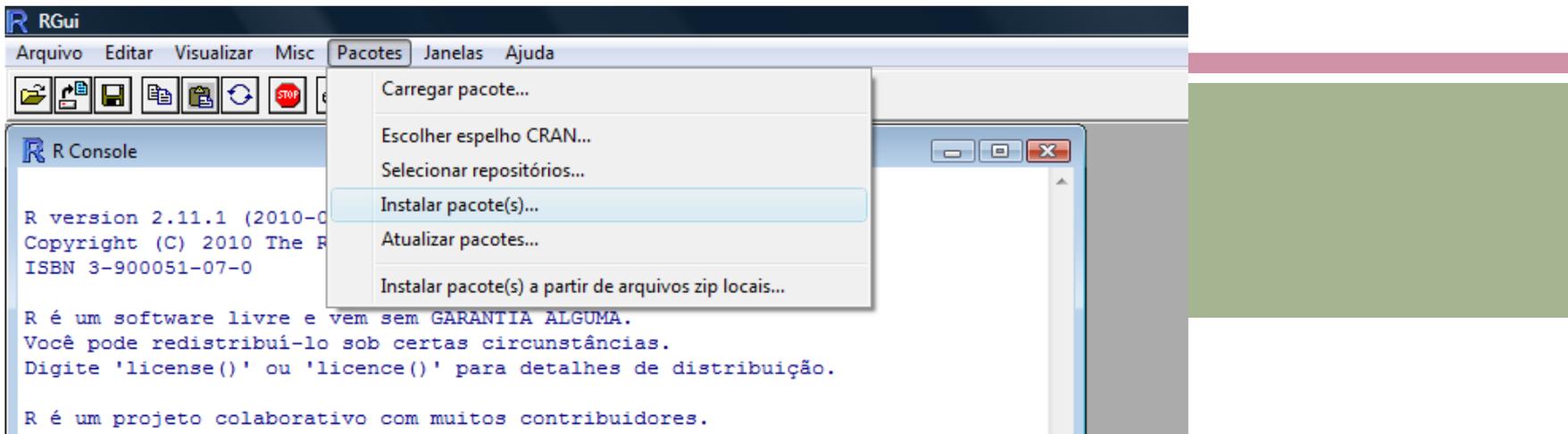
- i) Clicar no arquivo baixado
- ii) Executar
- iii) Caso abra uma janela de proteção, clicar em sim
- iv) Selecionar o idioma (Português, Brasil)
- v) Abrirá o Assistente de instalação, basta clicar em avançar até concluir  
Pronto, o R já está instalado na sua máquina, basta executá-lo.



# PACOTES

- 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

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
<b>User Library</b>		
<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

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.

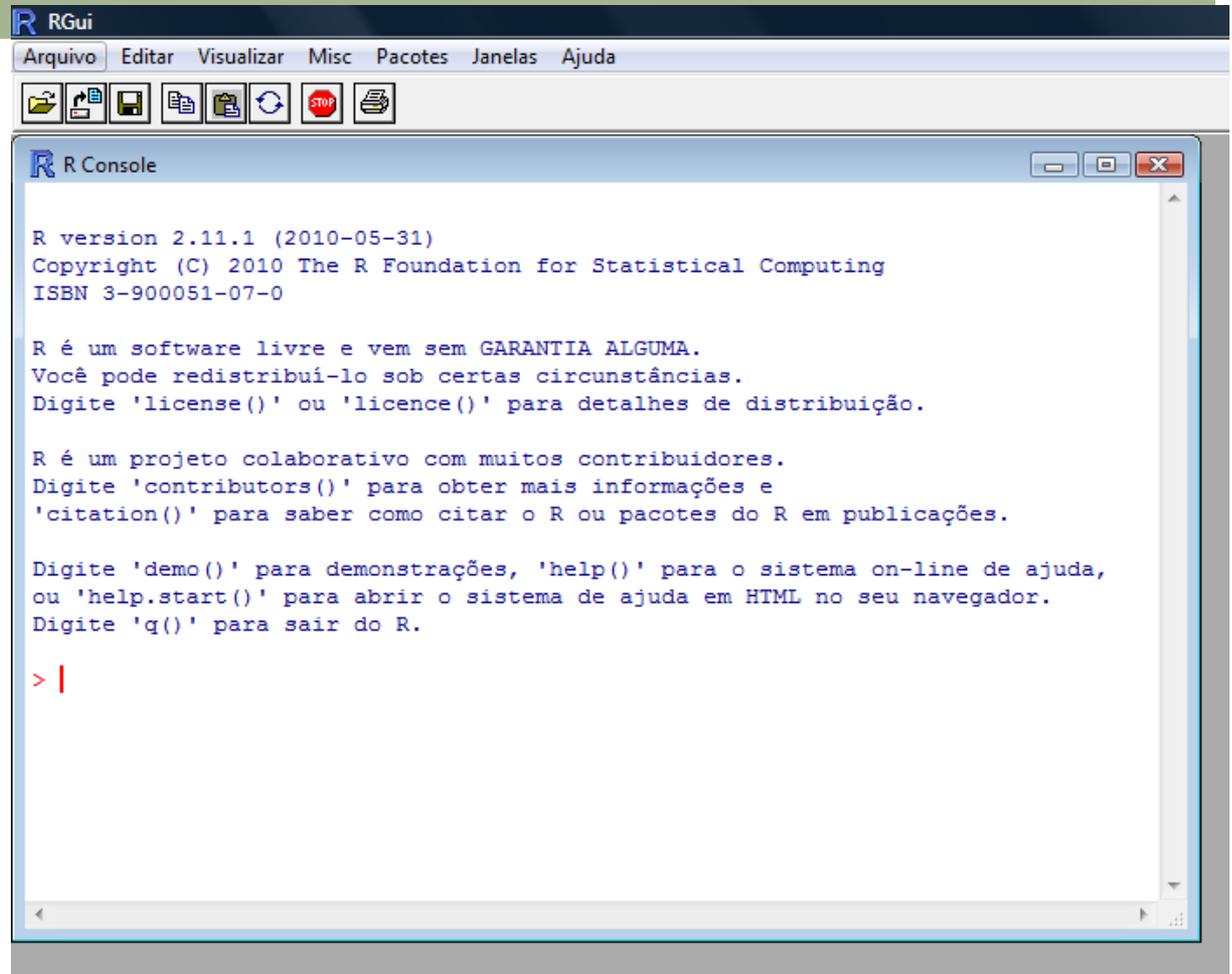
[workspace loaded from ~/MAE0217/.RData]

> |



# 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”.



```
R RGui
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda
R Console
R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> |
```

# 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

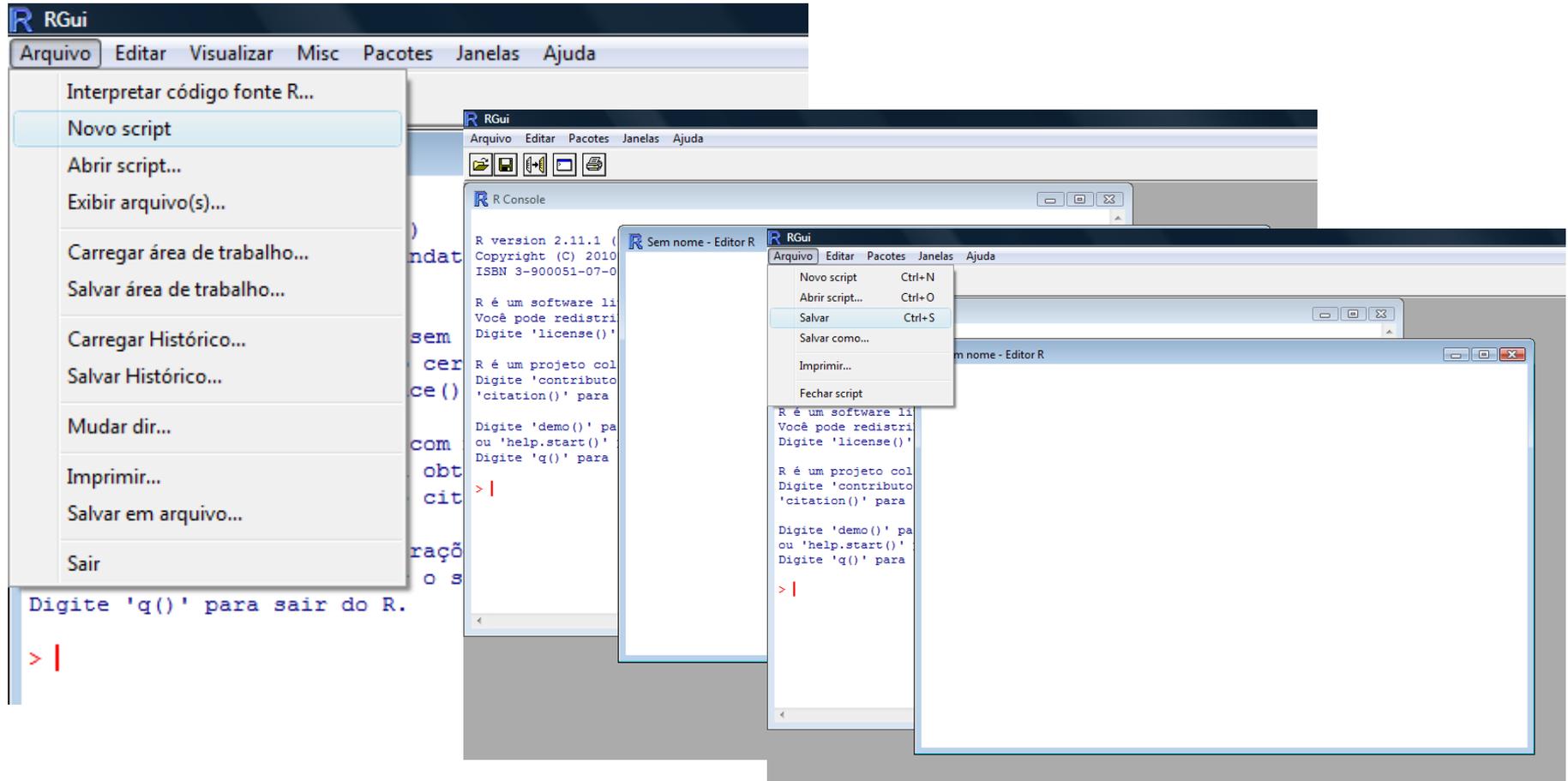
- Stack Overflow em Português é um site de perguntas e respostas para programadores profissionais e entusiastas.

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



# SCRIPTS

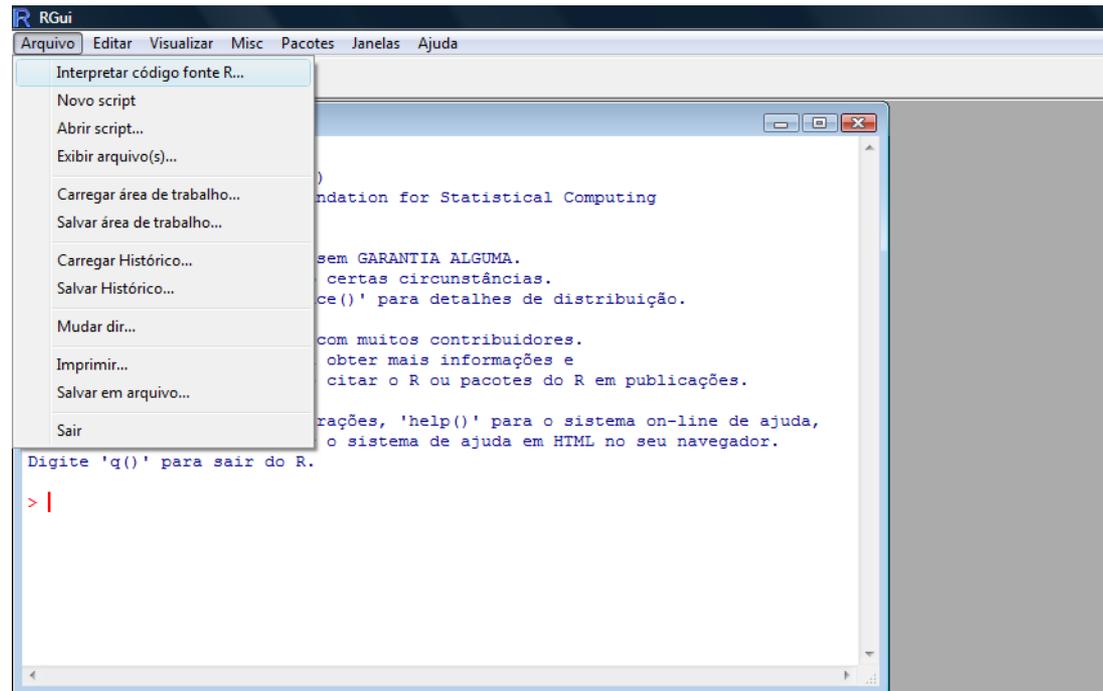
- Scripts são **arquivos** ou **“programas”** onde é salva uma lista de comandos, para não ter a necessidade de digitá-los no Prompt em cada nova sessão.
- Um script é um arquivo txt.
- Com o script você facilmente faz alterações e correções, além de salvar o script e poder refazer rapidamente suas análises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

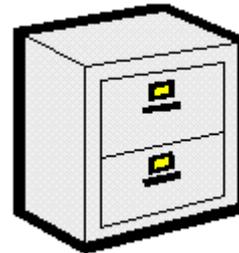


```
> attach(escore.inform)
```

```
> 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)
```

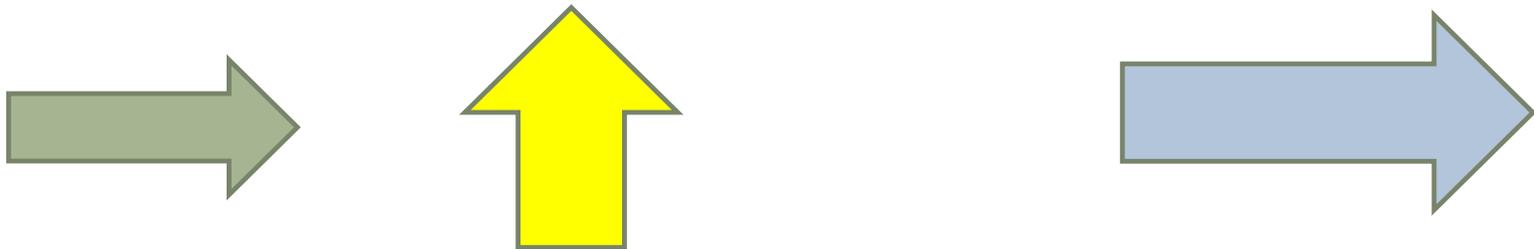


**KEEP  
CALM  
AND  
DON'T  
PANIC**

# 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.

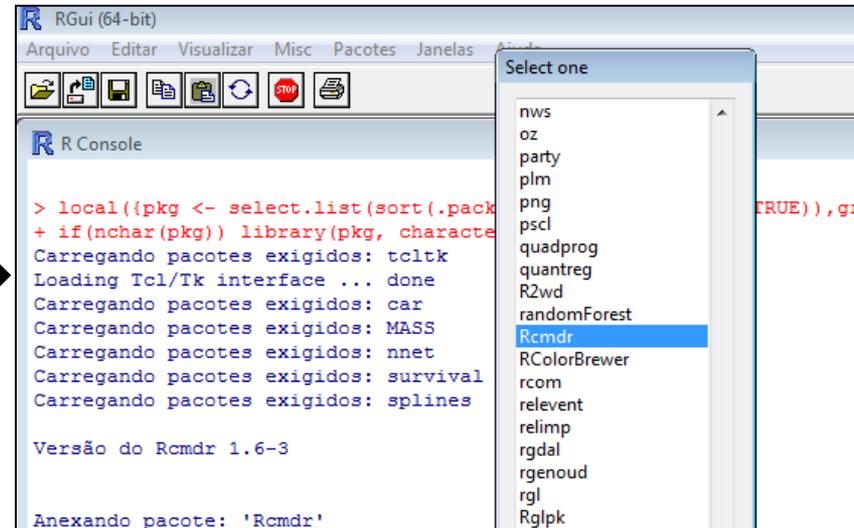
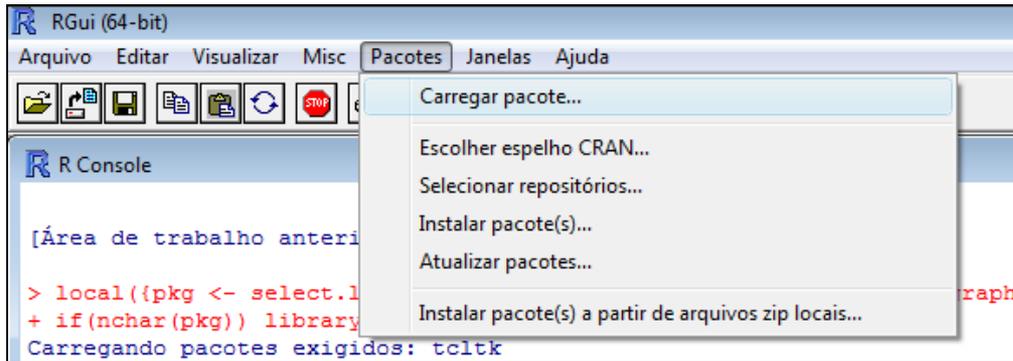
Atenção: Note que o Rcmdr não provê acesso a toda funcionalidade do R mas simplesmente a alguns procedimentos estatísticos mais usuais.

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

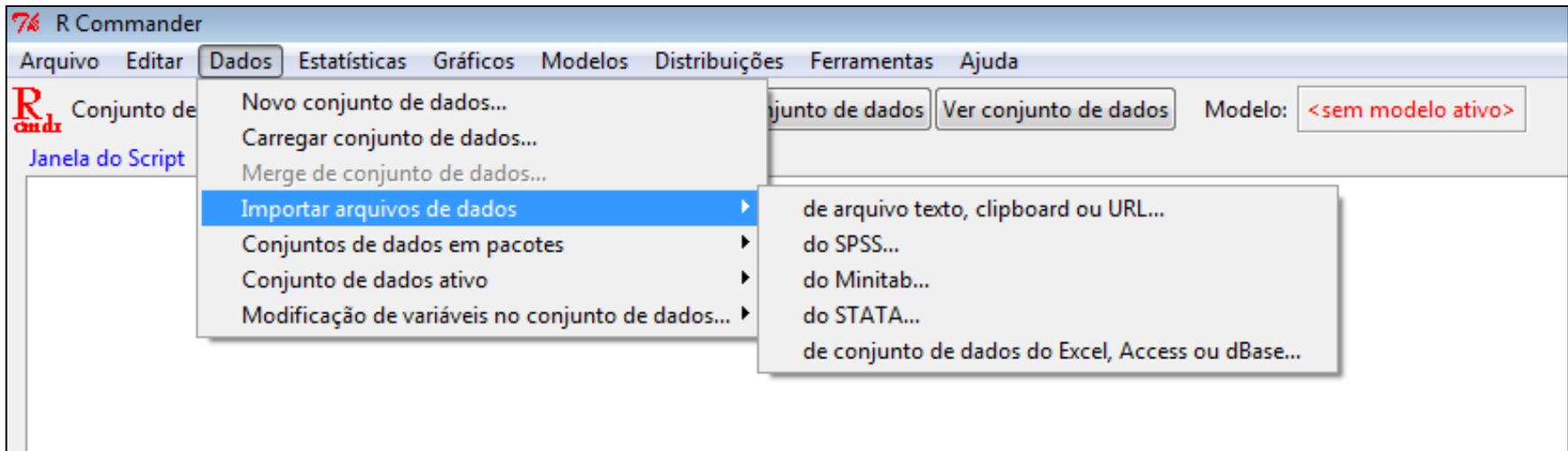
<http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/>

# Arquivo : Carregando dados no R

## (1) Carregar Rcmdr:

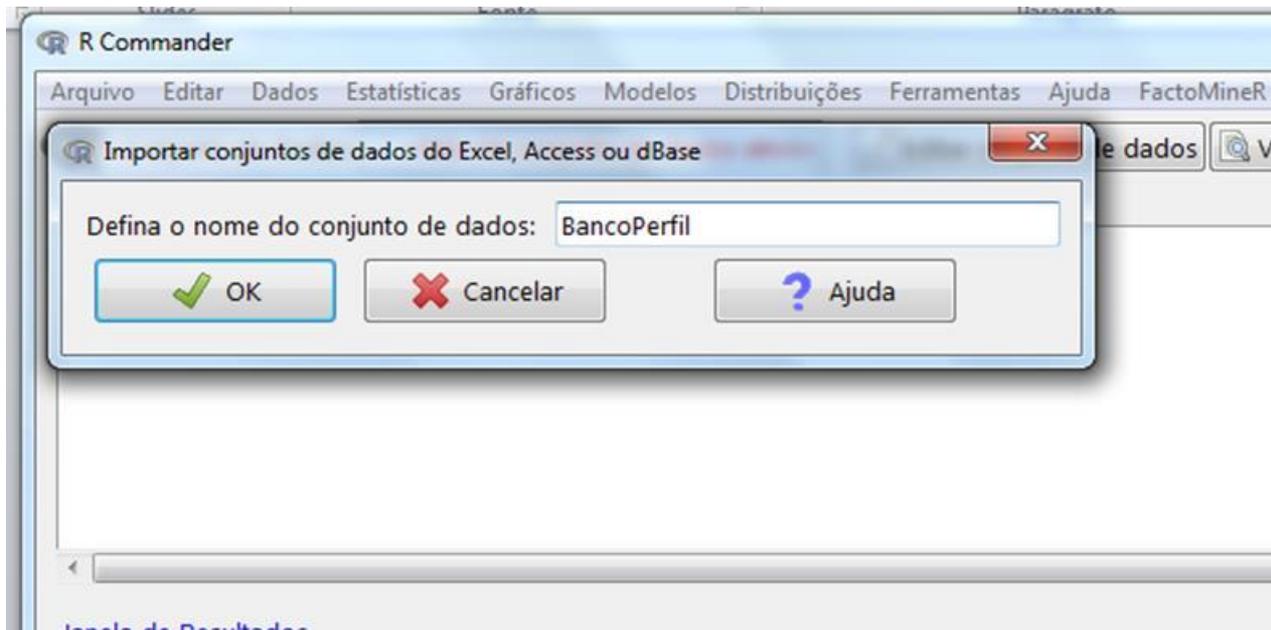


## (2) Importar dados:



# Arquivo Banco Perfil: Carregando dados no R

## (3) Nomear o banco:

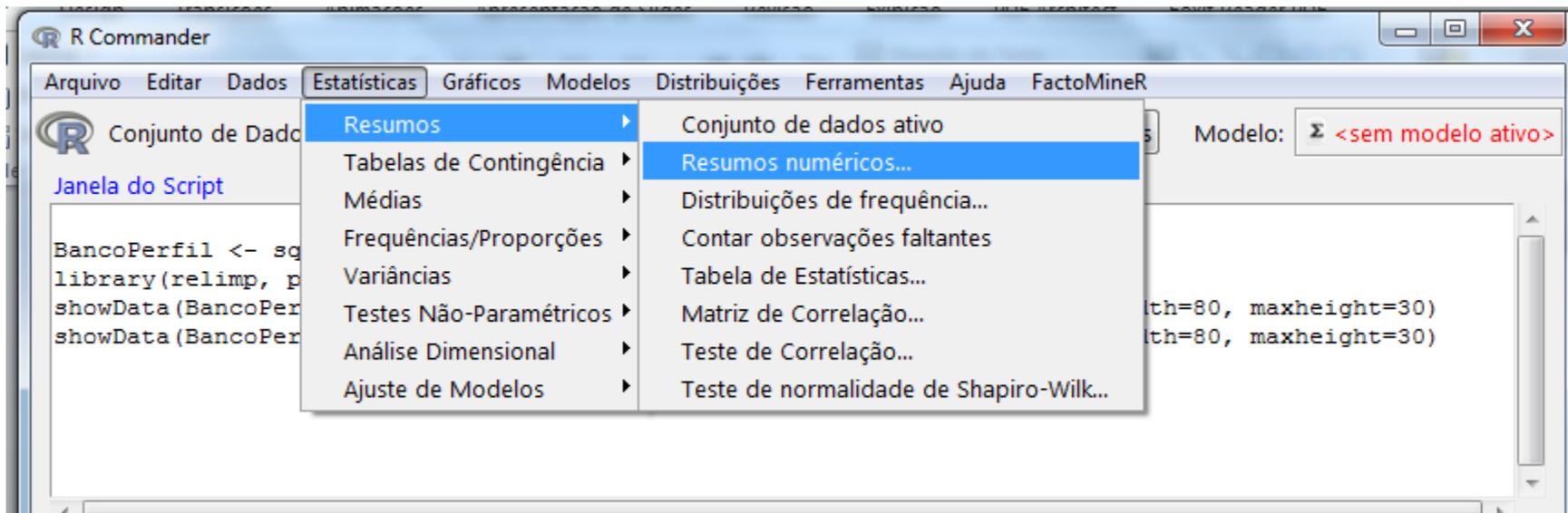


# Arquivo Banco Perfil: Editar/Visualizar dados



The screenshot shows the R Commander interface. The main window displays the 'BancoPerfil' dataset as a table with 19 rows and 10 columns. The columns are: id, data, sexo, nasc, idade, natural, classregião, nfilho, and me. The data is as follows:

	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	Araruana	Sudeste	0	



**Estatísticas → Resumos Numéricos**

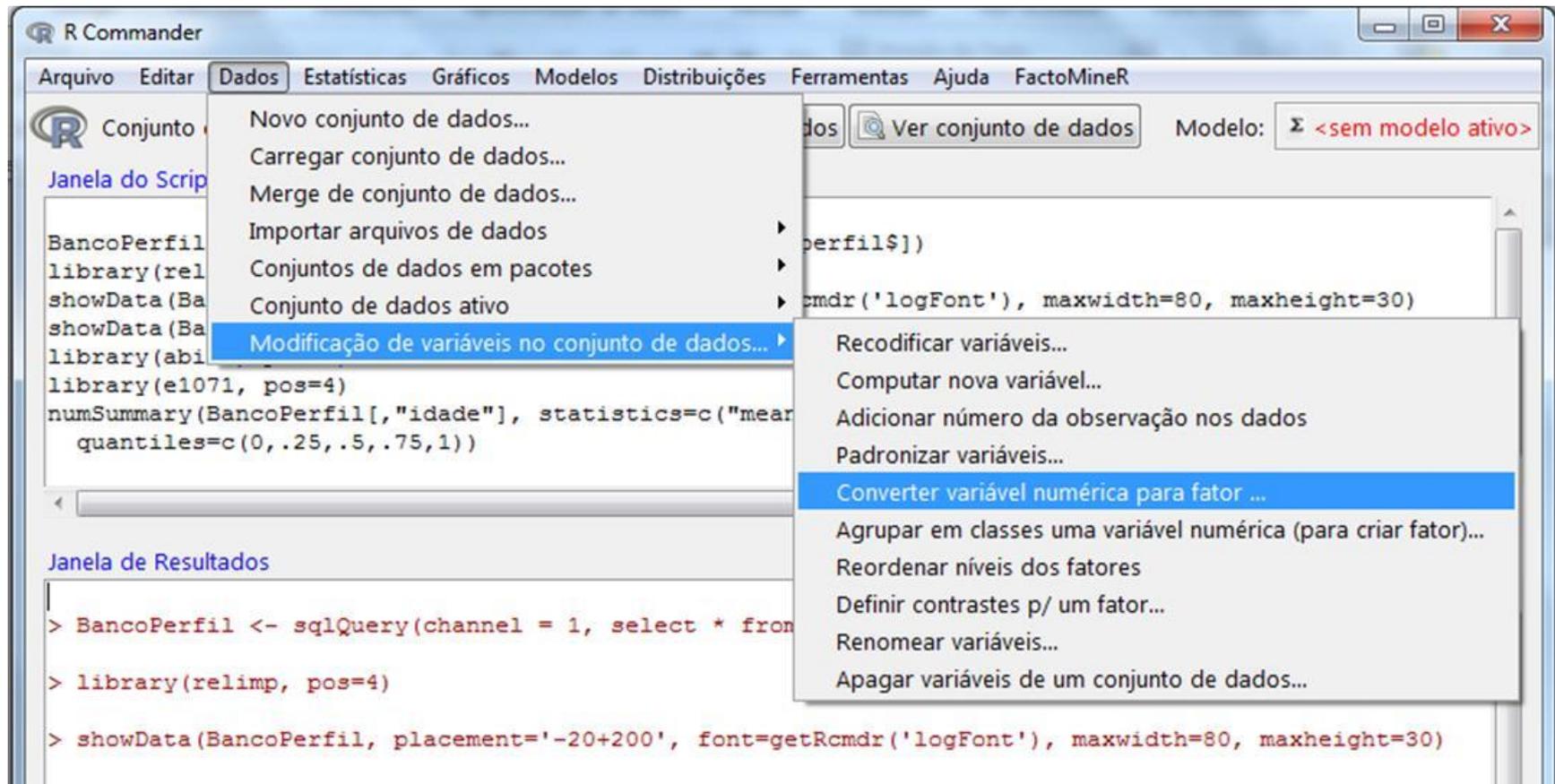
# Arquivo *Banco Perfil*:

## 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



R Commander

Arquivo Editar **Dados** Estatísticas Gráficos Modelos Distribuições Ferramentas Ajuda FactoMineR

Conjunto de dados: BancoPerfil Ver conjunto de dados Modelo:  $\Sigma$  <sem modelo ativo>

Janela do Script

```
BancoPerfil
library(reimp, pos=4)
showData(BancoPerfil, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, maxheight=30)
library(abi)
library(e1071, pos=4)
numSummary(BancoPerfil[, "idade"], statistics=c("mean", "sd", "var", "range", "min", "max", "quantiles=c(0, .25, .5, .75, 1)"))
```

Janela de Resultados

```
> BancoPerfil <- sqlQuery(channel = 1, select * from BancoPerfil)
> library(reimp, pos=4)
> showData(BancoPerfil, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, maxheight=30)
```

Novo conjunto de dados...  
Carregar conjunto de dados...  
Merge de conjunto de dados...  
Importar arquivos de dados  
Conjuntos de dados em pacotes  
Conjunto de dados ativo  
Modificação de variáveis no conjunto de dados...  
Recodificar variáveis...  
Computar nova variável...  
Adicionar número da observação nos dados  
Padronizar variáveis...  
Converter variável numérica para fator ...  
Agrupar em classes uma variável numérica (para criar fator)...  
Reordenar níveis dos fatores  
Definir contrastes p/ um fator...  
Renomear variáveis...  
Apagar variáveis de um conjunto de dados...

# Recodificando

The image illustrates the steps to recode a variable in R. It shows three overlapping dialog boxes from the R Commander interface.

**Top Dialog: Converter Variáveis Numéricas p/ Fator**

- Variáveis (selecione uma ou mais):** nfilho, rendafam, rendaper, **sexo** (selected).
- Níveis dos fatores:** Defina nomes dos níveis (selected), Use números.
- Novo nome de variável ou prefixo para múltiplas variáveis:** <mesmo que variáveis>
- Buttons: OK, Cancelar, Ajuda.

**Middle Dialog: Converter Variáveis Numéricas p/ Fator**

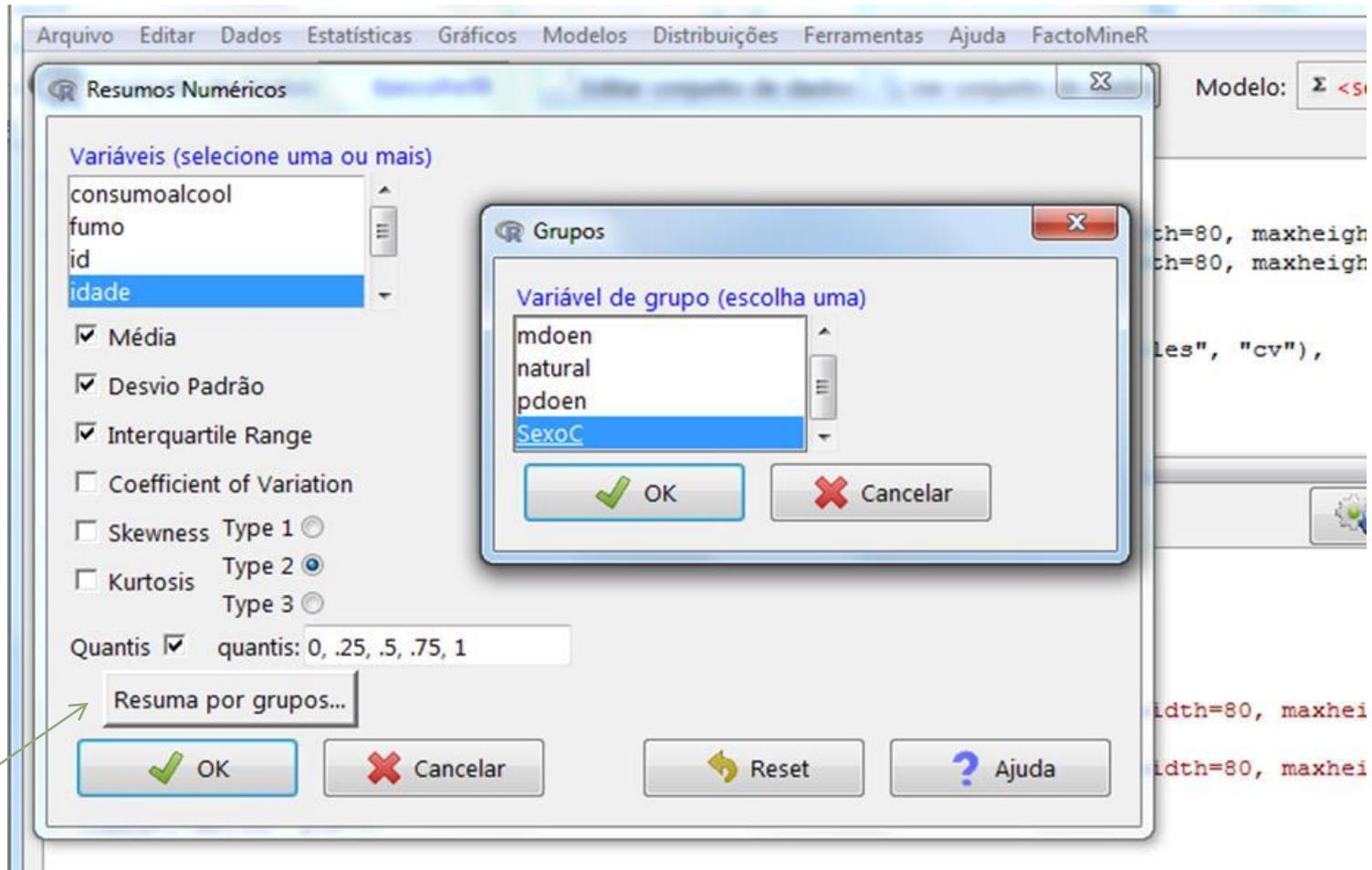
- Variáveis (selecione uma ou mais):** nfilho, rendafam, rendaper, **sexo** (selected).
- Níveis dos fatores:** Defina nomes dos níveis (selected), Use números.
- Novo nome de variável ou prefixo para múltiplas variáveis:** SexoC
- Buttons: OK, Cancelar, Ajuda.

**Bottom Dialog: Nomes dos níveis para SexoC**

Valor numérico	Nome do nível
1	Masc
2	Fem

Buttons: OK, Cancelar.

# Medidas descritivas por Sexo



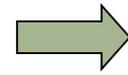
Estadísticas → Resumos Numéricos → Resuma por grupos

## 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



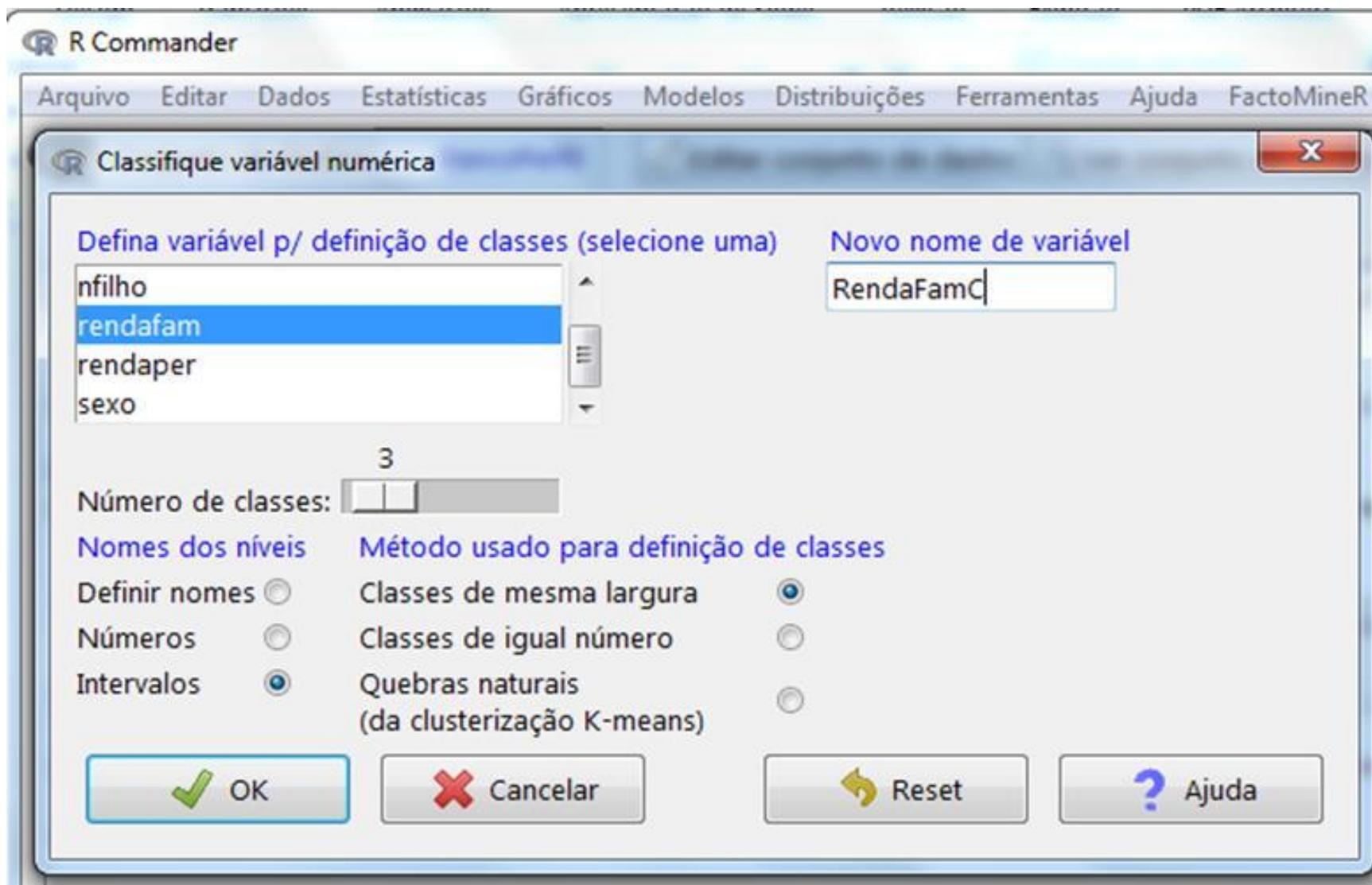
Construir  
intervalos  
de classe

(I) Criar nova variável

The screenshot shows the R Commander interface. The 'Dados' menu is open, and the option 'Agrupar em classes uma variável numérica (para criar fator)...' is highlighted. The main window displays R code for creating a factor variable from a numeric variable.

```
Novo conjunto de dados...  
Carregar conjunto de dados...  
Merge de conjunto de dados...  
Importar arquivos de dados  
Conjuntos de dados em pacotes  
Conjunto de dados ativo  
Modificação de variáveis no conjunto de dados...  
  Recodificar variáveis...  
  Computar nova variável...  
  Adicionar número da observação nos dados  
  Padronizar variáveis...  
  Converter variável numérica para fator ...  
  Agrupar em classes uma variável numérica (para criar fator)...  
  Reordenar níveis dos fatores  
  Definir contrastes p/ um fator...  
  Renomear variáveis...  
  Apagar variáveis de um conjunto de dados...
```

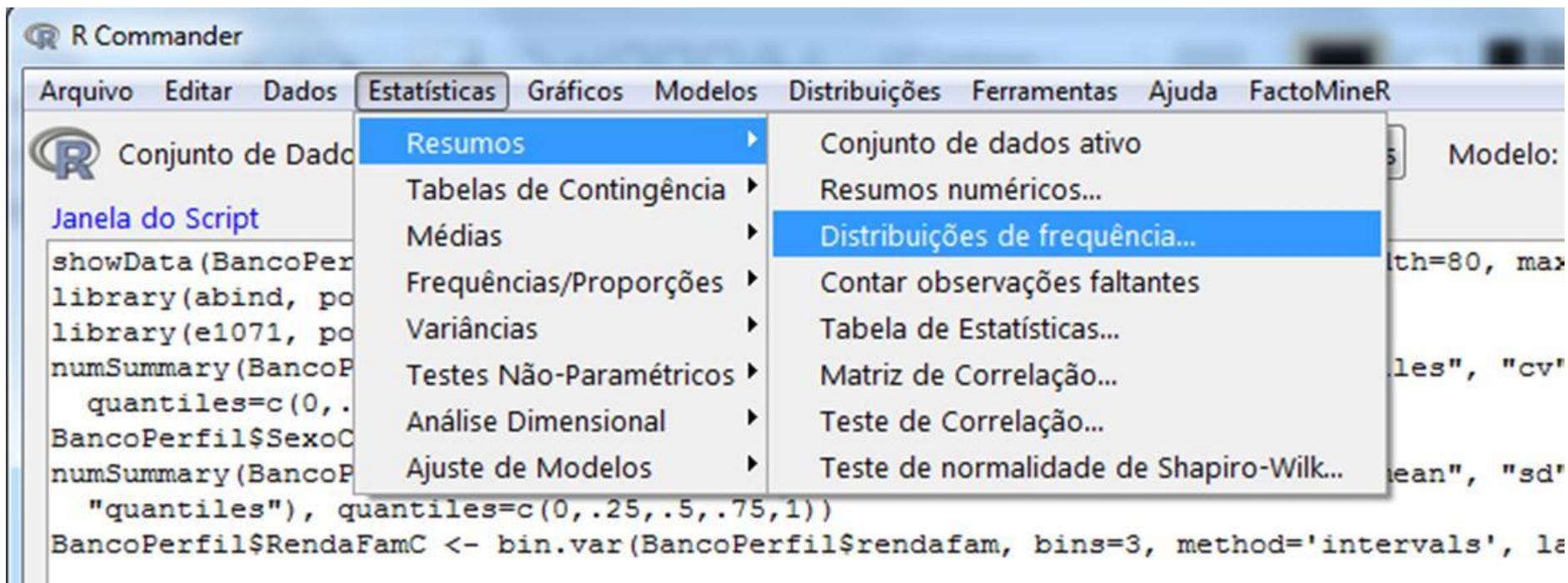
```
showData(BancoPerfil, placement='-20+200', font=ge)  
library(abind, pos=4)
```



# Exemplo 1:

Rcmdr:

(2) obter a distribuição de frequências da nova variável

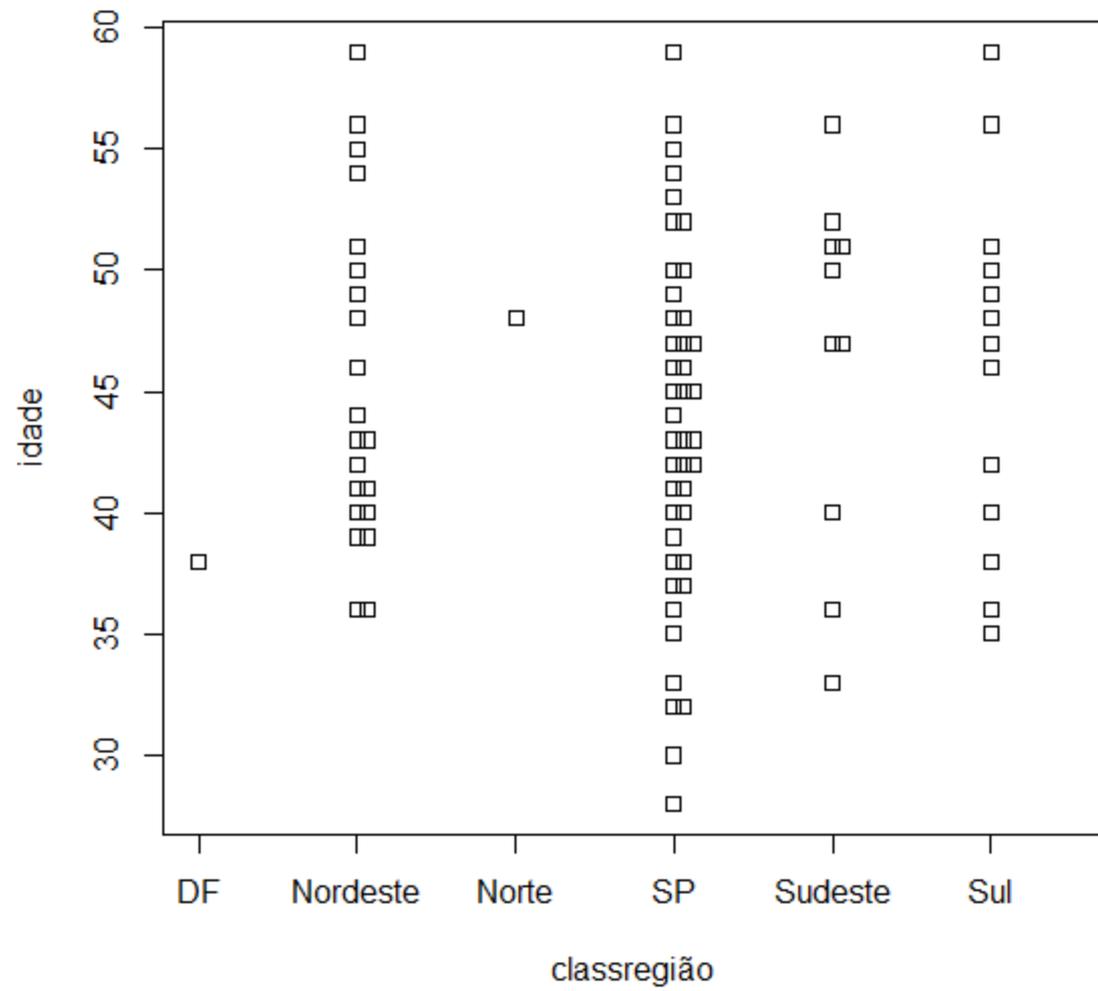


## (2) obter a distribuição de frequências da nova variável

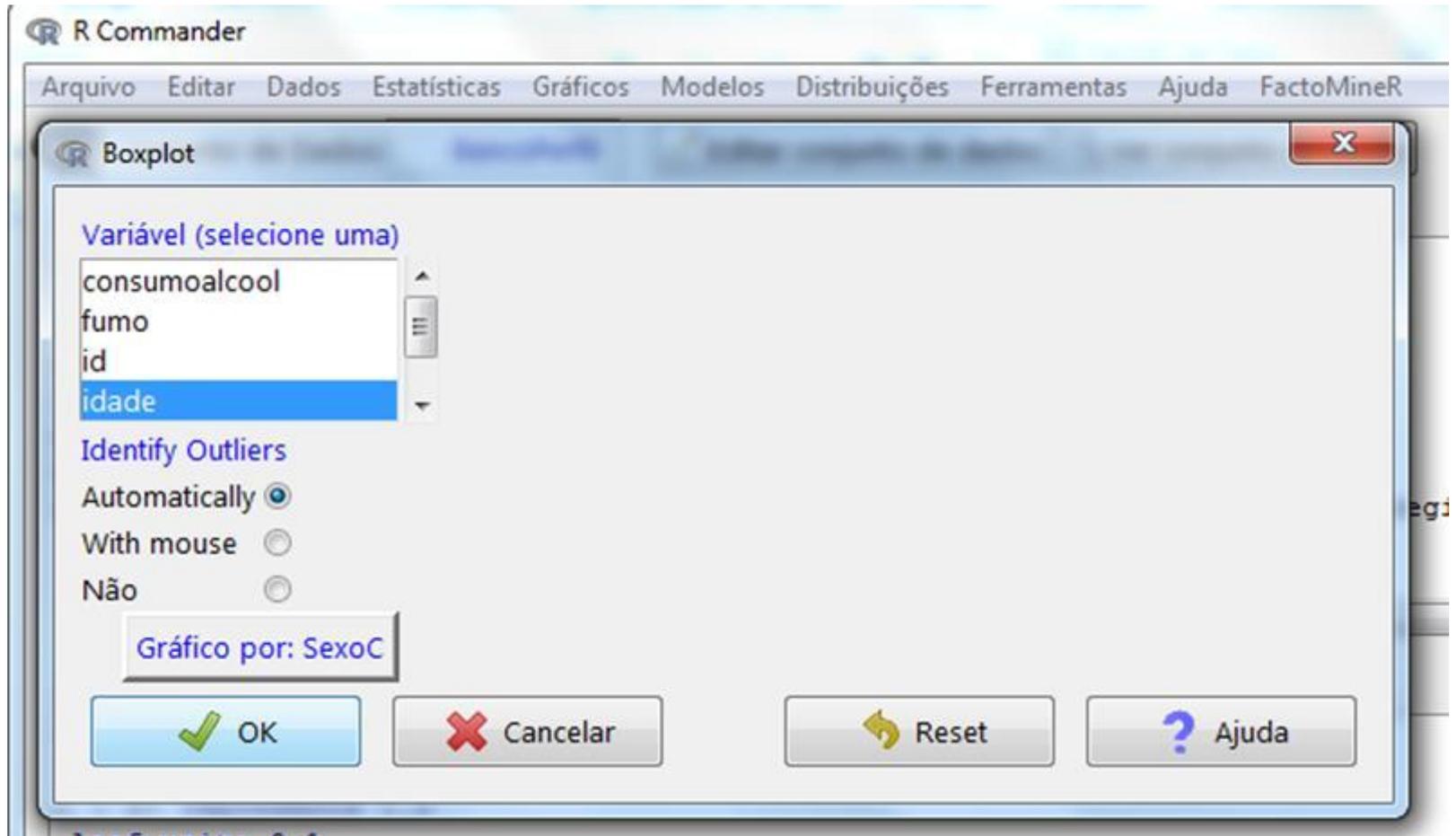
```
> .Table <- table(BancoPerfil$RendaFamC)
> .Table # counts for RendaFamC
      (486,5.33e+03] (5.33e+03,1.02e+04] (1.02e+04,1.5e+04]
                75                8                5
> round(100*.Table/sum(.Table), 2) # percentages for RendaFamC
      (486,5.33e+03] (5.33e+03,1.02e+04] (1.02e+04,1.5e+04]
                85.23                9.09                5.68
> remove(.Table)
```

# Gráficos mais comuns para variáveis quantitativas

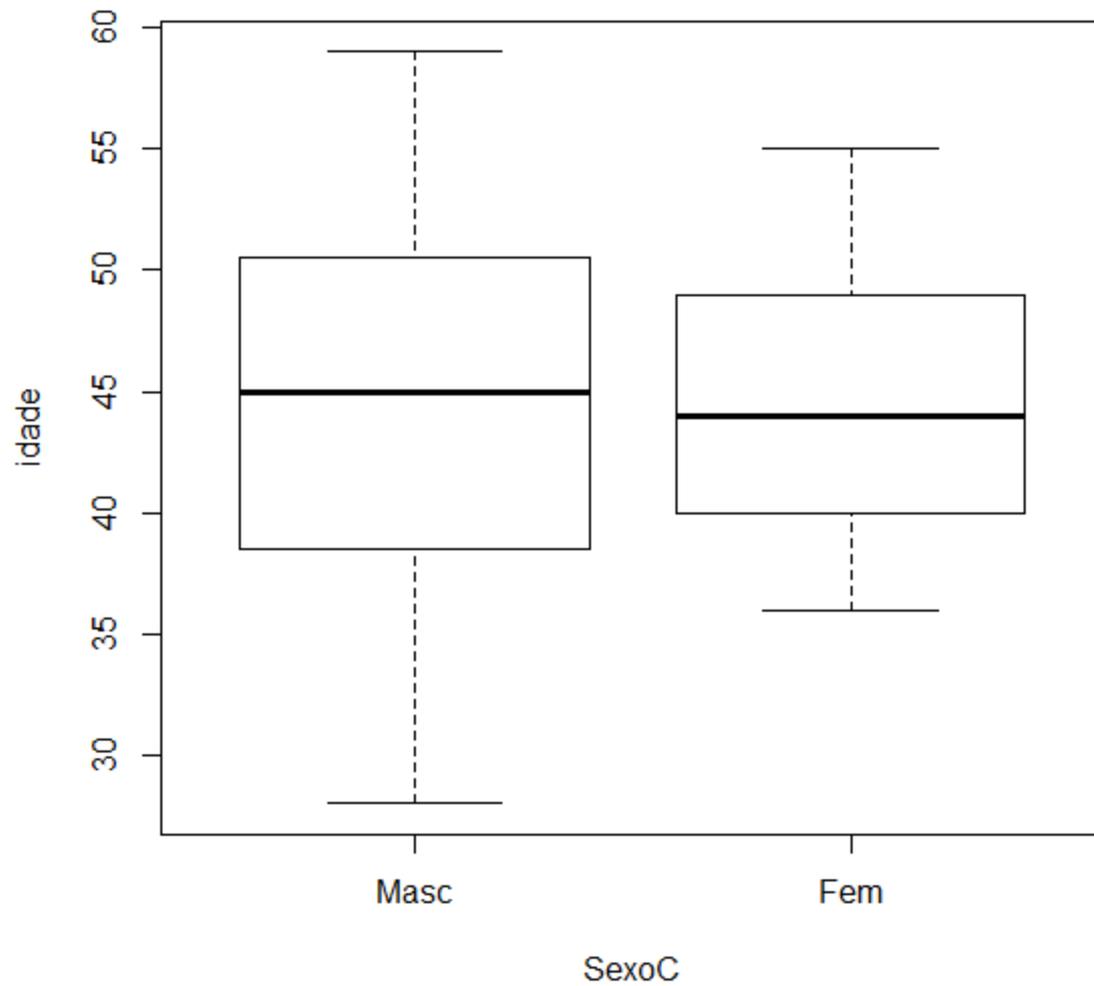
- “Strip Chart” ou “Dotplot”
- “Boxplot”
- Histograma



# Boxplot



# Boxplot



# Gráficos → Histograma...

The image shows two overlapping screenshots of the R Commander software interface. The top screenshot shows the 'Gráficos' (Graphics) menu with 'Histograma...' selected. The bottom screenshot shows the 'Histograma' dialog box with 'rendaper' selected as the variable and 'Densidades' selected as the scale.

**R Commander - Gráficos Menu**

- Gradiente de cores (color palette)
- Gráfico por Ordem de Apresentação (Index Plot)...
- Histograma...**
- Diagrama de ramo-e-folhas...
- Boxplot
- Gráfico de comparação de quantis...

**R Commander - Histograma Dialog**

Variável (selecione uma)

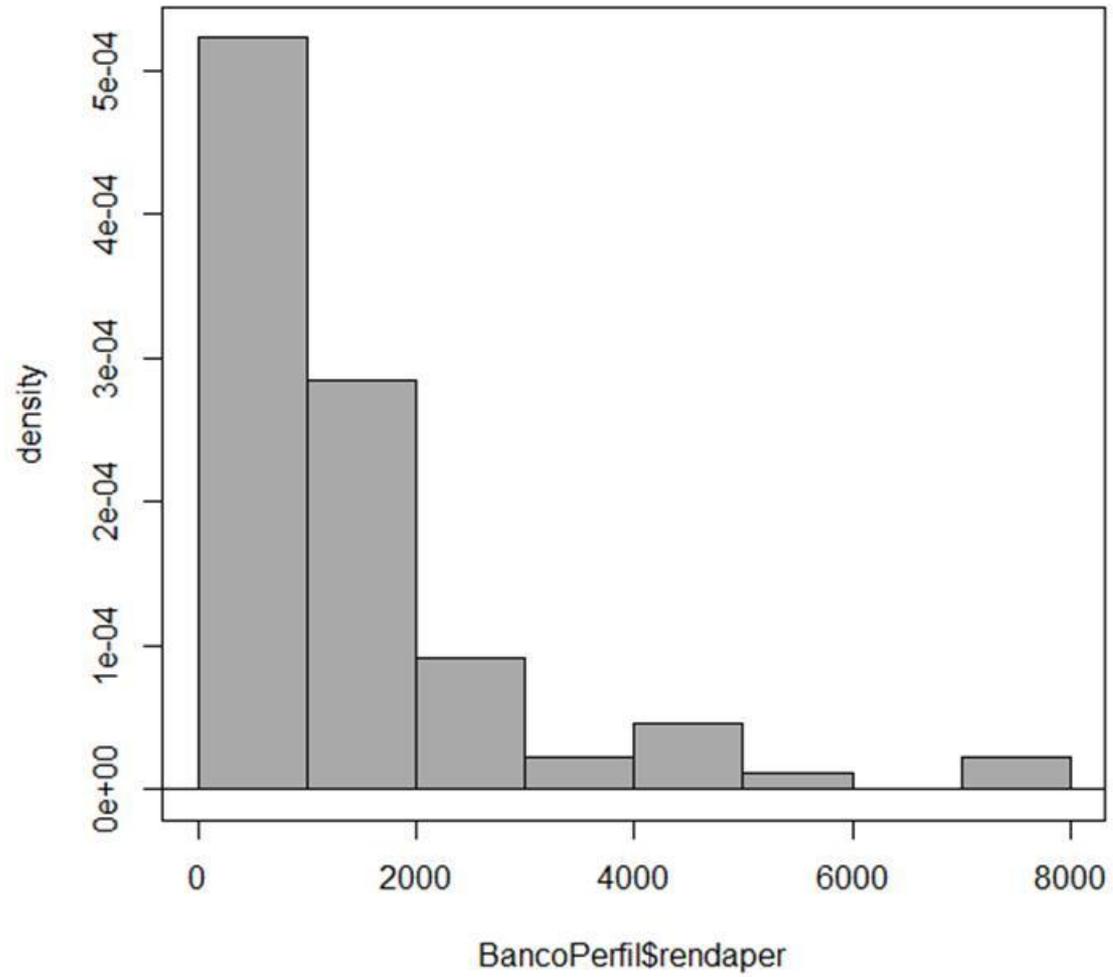
- nfilho
- rendafam
- rendaper**
- sexo

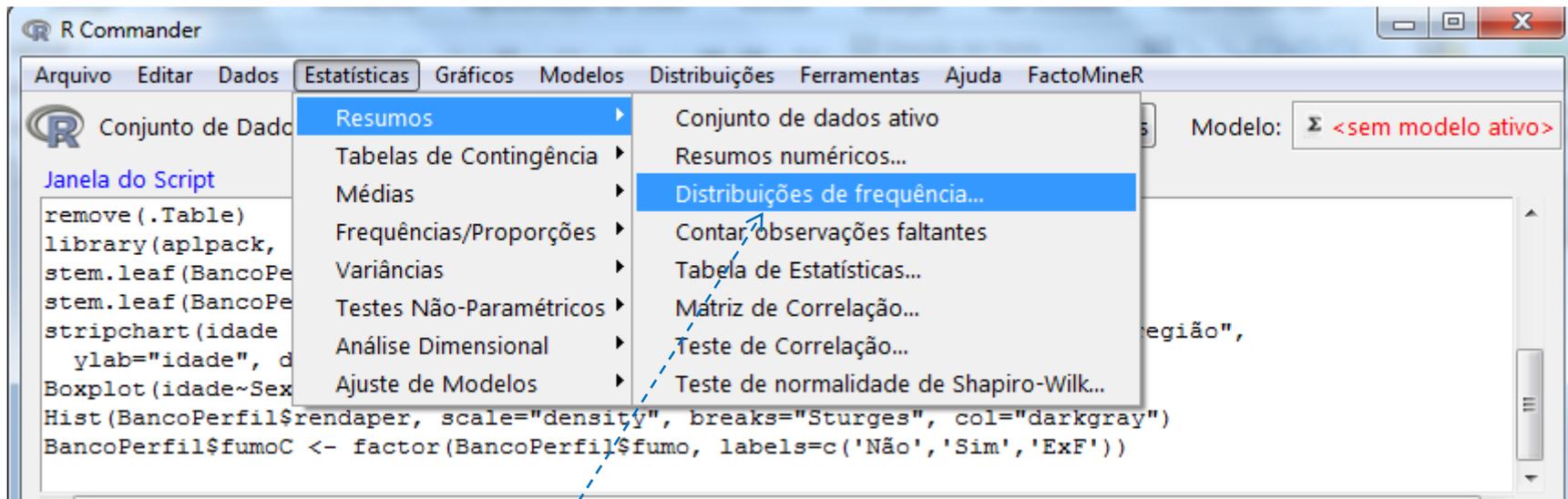
Número de classes: <auto>

Escala do eixo

- Contagens de frequência
- Percentagens
- Densidades

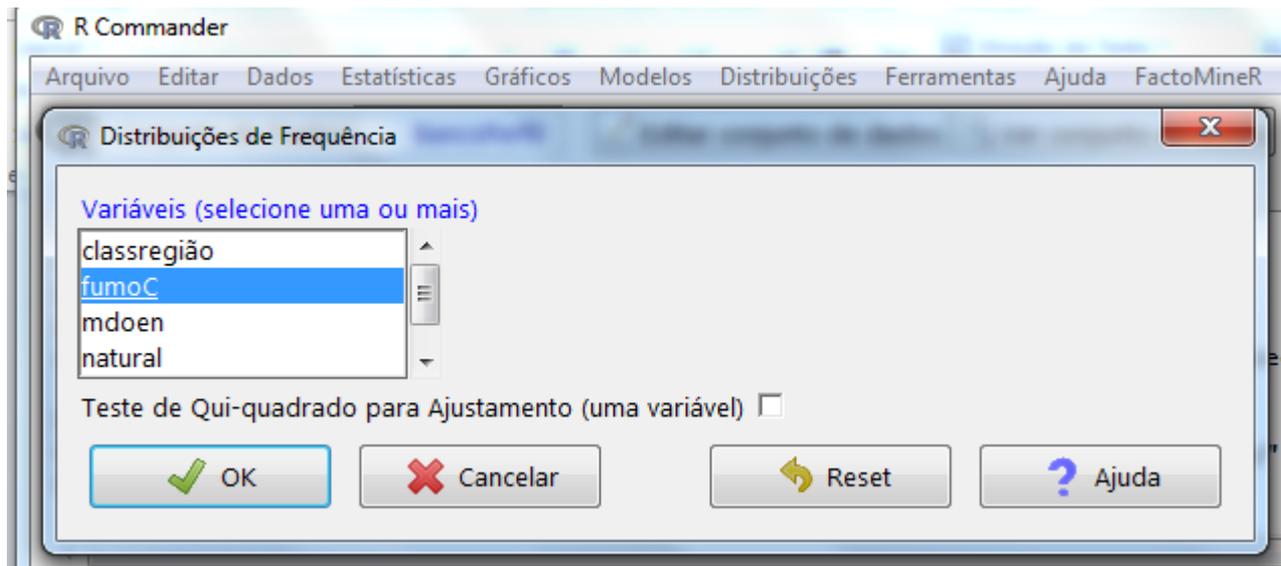
OK Cancelar Reset Ajuda





Cálculo de frequências e porcentagens

# Variáveis qualitativas



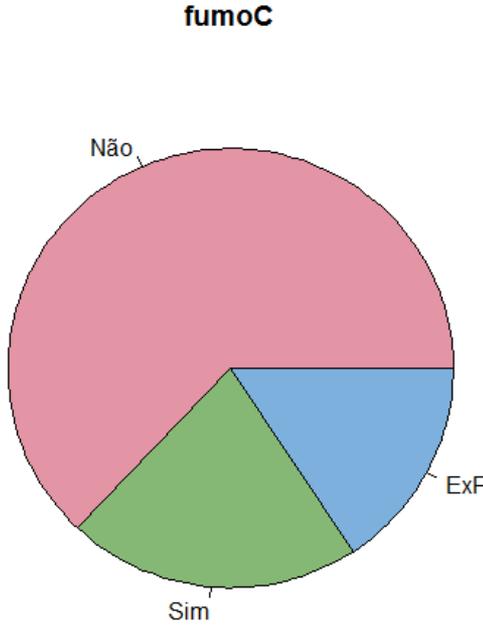
```
> .Table # counts for fumoC  
  
Não Sim ExF  
56 19 14  
  
> round(100*.Table/sum(.Table), 2) # percentages for fumoC  
  
Não Sim ExF  
62.92 21.35 15.73  
  
> remove(.Table)
```

# Variáveis Qualitativas

## Gráficos

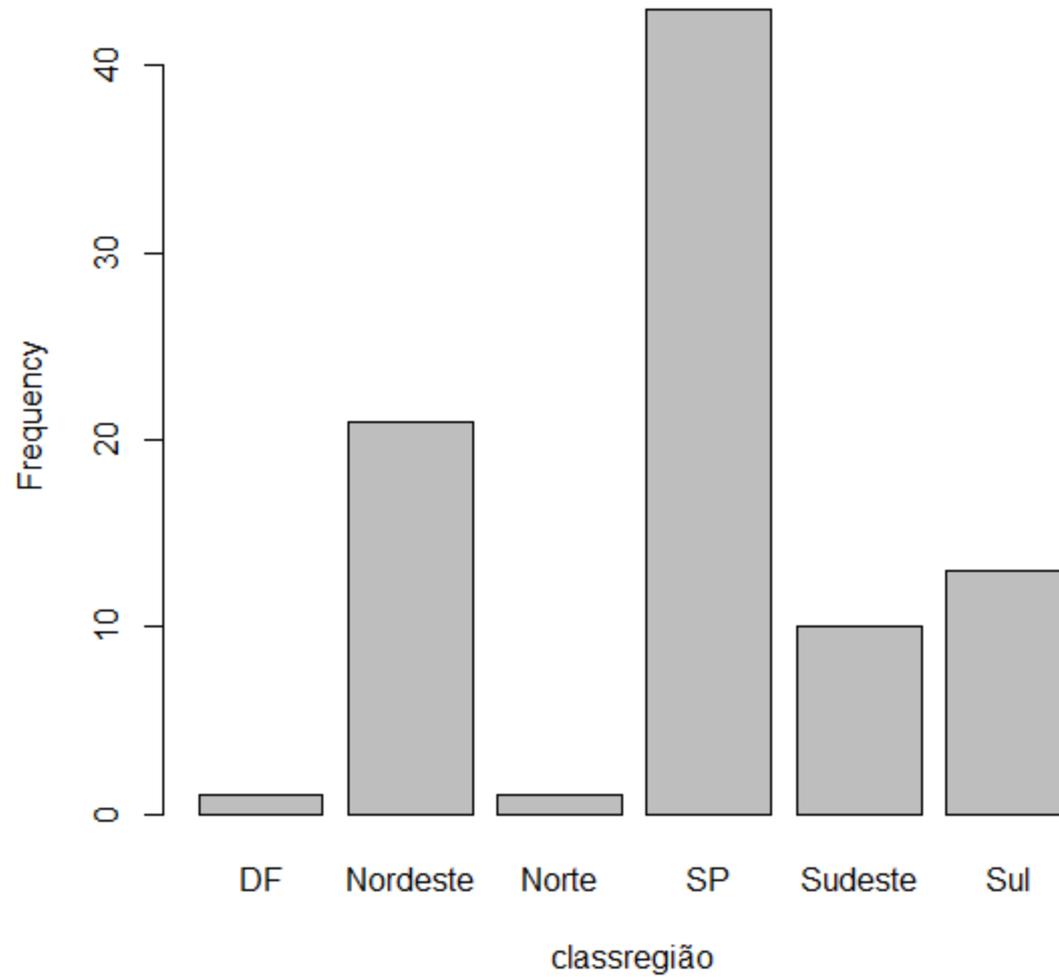
- Gráfico de setores
- Gráfico de barras

# 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)
```