

Análise descritiva no RCommander (INI/FIOCRUZ)

Raquel de Vasconcellos Carvalhaes de Oliveira

Luan Noé da Silva

Colaboração na elaboração: Fabiano Marcos

Estatística Descritiva

Utilizaremos o banco de dados “birthwt” que foi um estudo realizado nos Estados Unidos com 189 bebês, a fim de investigar os fatores associados ao peso ao nascer.

low indicator of birth weight less than 2.5 kg.

age mother's age in years.

lwt mother's weight in pounds at last menstrual period.

race mother's race (1 = white, 2 = black, 3 = other).

smoke smoking status during pregnancy.

ptl number of previous premature labours.

ht history of hypertension.

ui presence of uterine irritability.

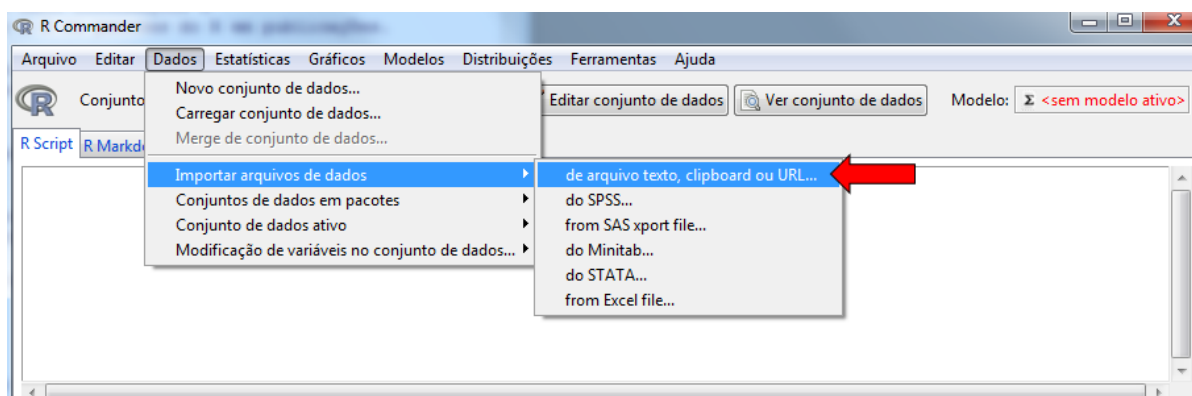
ftv number of physician visits during the first trimester.

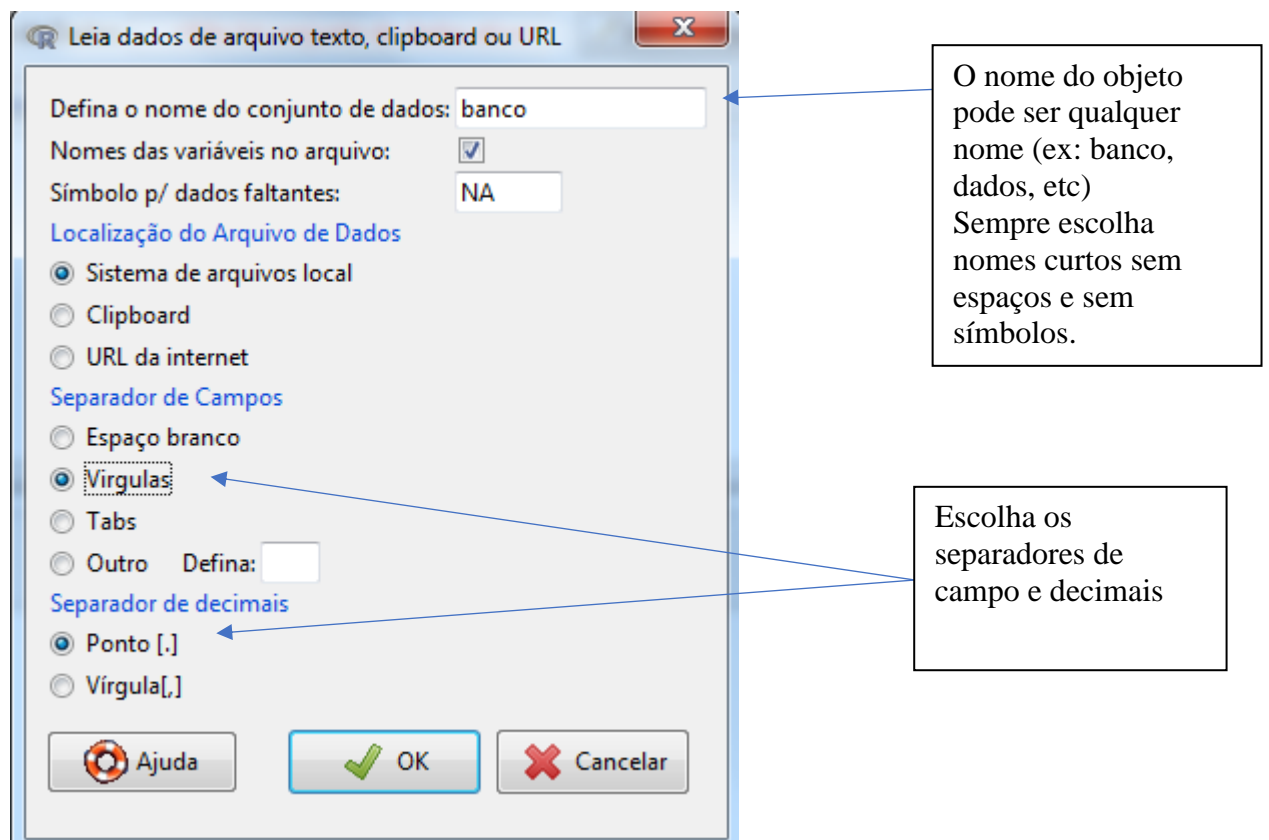
bwt birth weight in grams.

O banco de dados “birthwt” está no formato csv, que é um arquivo de texto, cujos campos são separados por “vírgula” ou “ponto e vírgula” (depende da configuração regional configurada no painel de controle do Windows) e os decimais separados por “ponto” ou “vírgula”. A importação de dados de arquivos com extensão .csv é realizada de forma similar aos demais arquivos.

Abra o arquivo .csv no bloco de notas e observe os separadores antes de importar.

Clique em “Dados-> Importar arquivos de dados->de arquivo texto, clipboard ou URL...”





Depois que o arquivo for carregado o status do “Conjunto de dados” muda e aparece o nome do arquivo que foi preenchido no campo “Defina o nome do conjunto de dados” na cor azul.

OBS: A importação acima funciona para bancos de outros formatos como SPSS, SAS, Stata e Excel (xls).

Após importar o banco de dados, deve-se checar se o banco de dados foi importado corretamente.

Observe as 6 primeiras linhas do banco de dados pelo seguinte comando na aba do R Script:

`head(banco)`

```
Output
> banco <- read.table("C:/Users/USUARIO/Desktop/birthwt.csv", header=TRUE, sep=";", na.str:
+ dec=".", strip.white=TRUE)
> head(banco)
  X low age lwt race smoke ptl ht ui ftv bwt
1 85  0 19 182   2    0  0  0  1  0 2523
2 86  0 33 155   3    0  0  0  0  3 2551
3 87  0 20 105   1    1  0  0  0  1 2557
4 88  0 21 108   1    1  0  0  1  2 2594
5 89  0 18 107   1    1  0  0  1  0 2600
6 91  0 21 124   3    0  0  0  0  0 2622
```

Observe as 10 primeiras linhas com todas as colunas.

`banco[1:10,]`

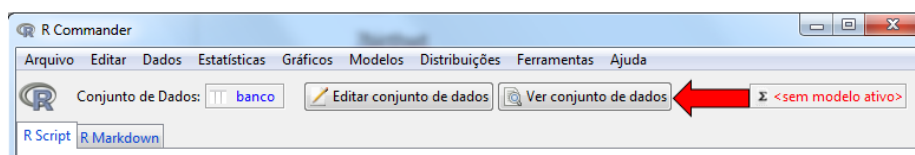
```
Output
> banco[1:10,]
  X low age lwt race smoke ptl ht ui ftv bwt
1 85  0 19 182   2    0  0  0  1  0 2523
2 86  0 33 155   3    0  0  0  0  3 2551
3 87  0 20 105   1    1  0  0  0  1 2557
4 88  0 21 108   1    1  0  0  1  2 2594
5 89  0 18 107   1    1  0  0  1  0 2600
6 91  0 21 124   3    0  0  0  0  0 2622
7 92  0 22 118   1    0  0  0  0  1 2637
8 93  0 17 103   3    0  0  0  0  1 2637
9 94  0 29 123   1    1  0  0  0  1 2663
10 95  0 26 113   1    1  0  0  0  0 2665

Mensagens
com a versão SDI da interface; veja ?Commander
[3] NOTA: Os dados banco tem 189 linhas e 11 colunas.
```

Pedindo a primeira coluna do banco de dados

```
Output
> banco[,1]
 [1] 85 86 87 88 89 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106
[22] 107 108 109 111 112 113 114 115 116 117 118 119 120 121 123 124 125 126 127 128 129
[43] 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
[64] 151 154 155 156 159 160 161 162 163 164 166 167 168 169 170 172 173 174 175 176 177
[85] 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 195 196 197 199 200 201
[106] 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222
[127] 223 224 225 226  4 10 11 13 15 16 17 18 19 20 22 23 24 25 26 27 28
[148] 29 30 31 32 33 34 35 36 37 40 42 43 44 45 46 47 49 50 51 52 54
[169] 56 57 59 60 61 62 63 65 67 68 69 71 75 76 77 78 79 81 82 83 84
```

Uma outra forma de fazer a observação dos dados é clicar no botão “Ver conjunto de dados”



Aparecerá uma janela mostrando todos os dados contidos no arquivo que foi carregado para o R Commander.

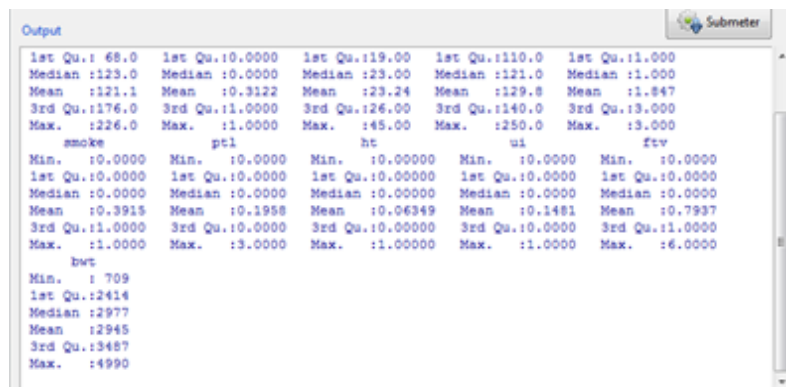
The image shows a window titled 'banco' displaying a data frame. The data is presented in a table with 30 rows and 11 columns. The columns are labeled X, low, age, lwt, race, smoke, ptl, ht, ui, ftv, and bwt. The rows contain numerical values for each variable. The first few rows are: Row 1: 85, 0, 19, 182, 2, 0, 0, 0, 1, 0, 2523; Row 2: 86, 0, 33, 155, 3, 0, 0, 0, 0, 3, 2551; Row 3: 87, 0, 20, 105, 1, 1, 0, 0, 0, 1, 2557; Row 4: 88, 0, 21, 108, 1, 1, 0, 0, 1, 2, 2594; Row 5: 89, 0, 18, 107, 1, 1, 0, 0, 1, 0, 2600; Row 6: 91, 0, 21, 124, 3, 0, 0, 0, 0, 0, 2622; Row 7: 92, 0, 22, 118, 1, 0, 0, 0, 0, 1, 2637; Row 8: 93, 0, 17, 103, 3, 0, 0, 0, 0, 1, 2637; Row 9: 94, 0, 29, 123, 1, 1, 0, 0, 0, 1, 2663; Row 10: 95, 0, 26, 113, 1, 1, 0, 0, 0, 0, 2665; Row 11: 96, 0, 19, 95, 3, 0, 0, 0, 0, 0, 2722; Row 12: 97, 0, 19, 150, 3, 0, 0, 0, 0, 1, 2733; Row 13: 98, 0, 22, 95, 3, 0, 0, 0, 1, 0, 2751; Row 14: 99, 0, 30, 107, 3, 0, 1, 0, 1, 2, 2750; Row 15: 100, 0, 18, 100, 1, 1, 0, 0, 0, 0, 2769; Row 16: 101, 0, 18, 100, 1, 1, 0, 0, 0, 0, 2769; Row 17: 102, 0, 15, 98, 2, 0, 0, 0, 0, 0, 2778; Row 18: 103, 0, 25, 118, 1, 1, 0, 0, 0, 3, 2782; Row 19: 104, 0, 20, 120, 3, 0, 0, 0, 1, 0, 2807; Row 20: 105, 0, 28, 120, 1, 1, 0, 0, 0, 1, 2821; Row 21: 106, 0, 32, 121, 3, 0, 0, 0, 0, 2, 2835; Row 22: 107, 0, 31, 100, 1, 0, 0, 0, 1, 3, 2835; Row 23: 108, 0, 36, 202, 1, 0, 0, 0, 0, 1, 2836; Row 24: 109, 0, 28, 120, 3, 0, 0, 0, 0, 0, 2863; Row 25: 111, 0, 25, 120, 3, 0, 0, 0, 1, 2, 2877; Row 26: 112, 0, 28, 167, 1, 0, 0, 0, 0, 0, 2877; Row 27: 113, 0, 17, 122, 1, 1, 0, 0, 0, 0, 2906; Row 28: 114, 0, 29, 150, 1, 0, 0, 0, 0, 2, 2920; Row 29: 115, 0, 26, 168, 2, 1, 0, 0, 0, 0, 2920; Row 30: 116, 0, 17, 113, 2, 0, 0, 0, 0, 1, 2920.

Observe a estrutura do banco de dados. Digite `str(banco)`

The image shows the R Output window with the command `str(banco)` executed. The output displays the structure of the data frame 'banco', which contains 189 observations of 11 variables. The variables and their data types are: X (integer), low (integer), age (integer), lwt (integer), race (integer), smoke (integer), ptl (integer), ht (integer), ui (integer), ftv (integer), and bwt (integer). The output shows the first few values for each variable, followed by an ellipsis indicating the rest of the data. For example, X starts with 85, 86, 87, 88, 89, 91, 92, 93, 94, 95, ... and bwt starts with 2523, 2551, 2557, 2594, 2600, 2622, 2637, 2637, 2663, 2665, ...

Para realizar um resumo estatístico do banco de dados, execute o seguinte comando:

`summary(banco)`

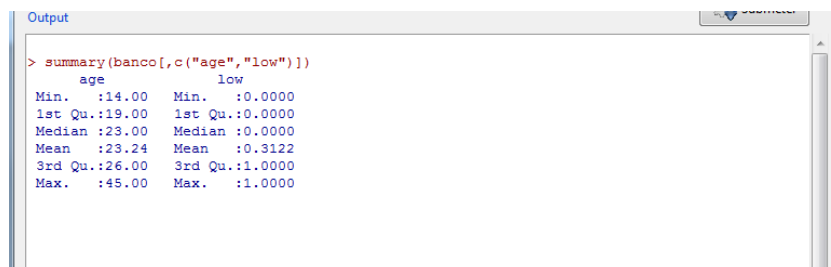


Output

1st Qu.: 66.0	1st Qu.: 0.0000	1st Qu.: 119.00	1st Qu.: 110.0	1st Qu.: 1.000
Median : 123.0	Median : 0.0000	Median : 23.00	Median : 121.0	Median : 1.000
Mean : 121.1	Mean : 0.3122	Mean : 23.24	Mean : 129.8	Mean : 1.847
3rd Qu.: 176.0	3rd Qu.: 1.0000	3rd Qu.: 126.00	3rd Qu.: 140.0	3rd Qu.: 3.000
Max. : 226.0	Max. : 1.0000	Max. : 145.00	Max. : 250.0	Max. : 3.000
smoke	ptl	ht	ui	ftv
Min. : 0.0000	Min. : 0.0000	Min. : 0.00000	Min. : 0.0000	Min. : 0.0000
1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.00000	1st Qu.: 0.0000	1st Qu.: 0.0000
Median : 0.0000	Median : 0.0000	Median : 0.00000	Median : 0.0000	Median : 0.0000
Mean : 0.3915	Mean : 0.1958	Mean : 0.06349	Mean : 0.1481	Mean : 0.7937
3rd Qu.: 1.0000	3rd Qu.: 0.0000	3rd Qu.: 0.00000	3rd Qu.: 0.0000	3rd Qu.: 1.0000
Max. : 1.0000	Max. : 3.0000	Max. : 1.00000	Max. : 1.0000	Max. : 6.0000
bwt				
Min. : 709				
1st Qu.: 2414				
Median : 2977				
Mean : 2945				
3rd Qu.: 3487				
Max. : 4990				

Observe a estrutura das variáveis smoke e low.

Sumário de um determinado conjunto de variáveis. `summary(banco[,c("age", "low")])`



Output

```
> summary(banco[,c("age", "low")])
```

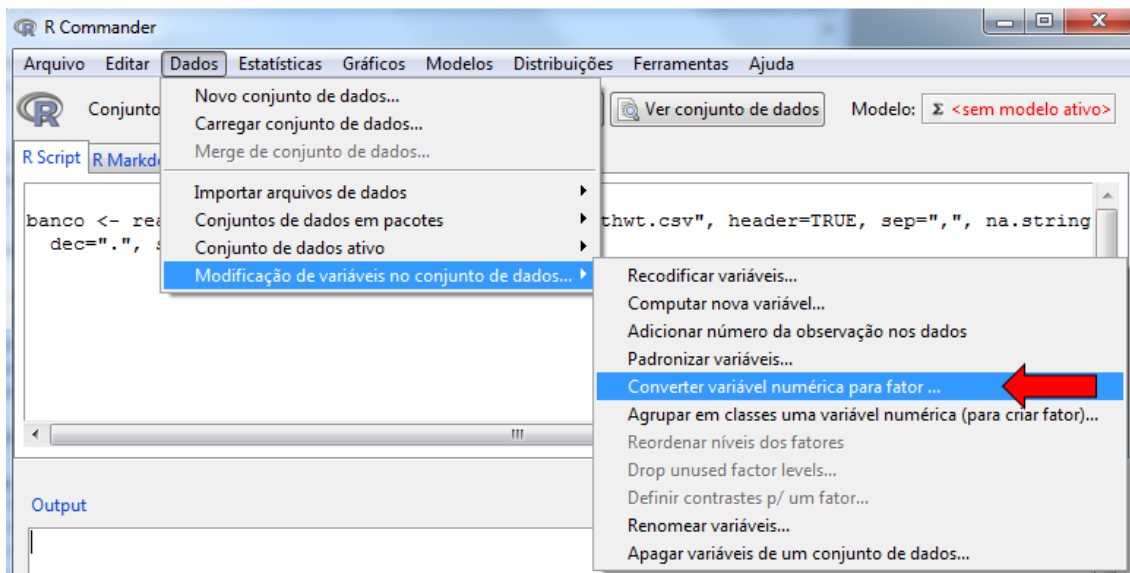
age	low
Min. :14.00	Min. :0.0000
1st Qu.:19.00	1st Qu.:0.0000
Median :23.00	Median :0.0000
Mean :23.24	Mean :0.3122
3rd Qu.:26.00	3rd Qu.:1.0000
Max. :45.00	Max. :1.0000

Conversão de Dados

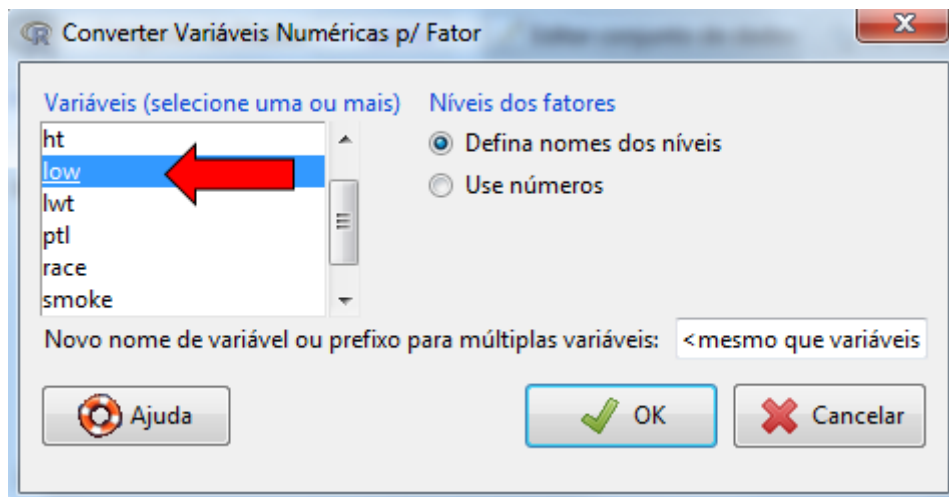
Muitas vezes é necessário converter variáveis numéricas para fator (variável qualitativa). Ao executar os comandos abaixo são realizadas as conversões dos dados das variáveis “low, race, smoke, ht e ui”.

No R Commander essa conversão pode ser feita da seguinte forma:

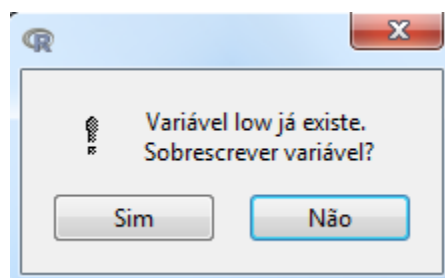
“Dados-> Modificação de Variáveis no conjunto de dados -> Converter variáveis numéricas par fator”



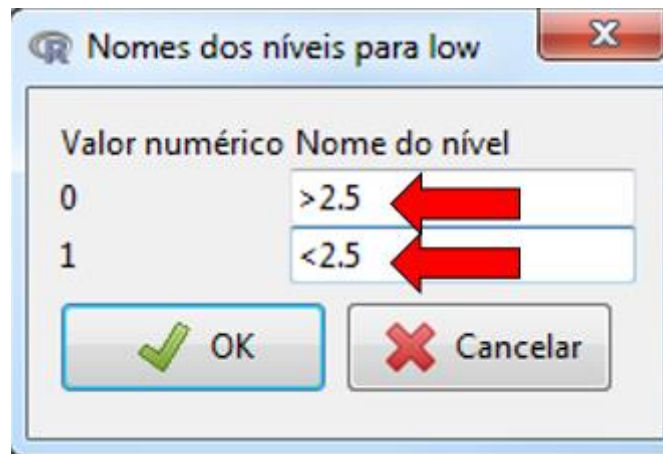
Vamos realizar a conversão para a variável “low” por exemplo. Na caixa de comando seleciona a variável.



E depois clique em OK e aparecerá a seguinte opção para redefinição da variável, clique em sim par sobrescrever.



Defina os rótulos de cada valor na caixa de diálogo. Após observar o dicionário de dados >2.5 e <2.5 , respectivamente. Clique em OK para a conversão ser concluída.



Repita os passos acima para as demais variáveis ou execute os comandos abaixo:

```
banco$low<-factor(banco$low,labels=c(">2.5","<2.5"))
banco$race<-factor(banco$race, labels=c("Branco","Preto","Outros"))
banco$smoke<-factor(banco$smoke,labels=c("Não","Sim"))
banco$ht<-factor(banco$ht,labels=c("Não","Sim"))
banco$ui<-factor(banco$ui,labels=c("Não","Sim"))
```

Resumos Estatísticos

Importante: Antes de realizar qualquer resumo do banco de dados verifique a estrutura do seu banco de dados pelo comando `str(banco)!!!!`

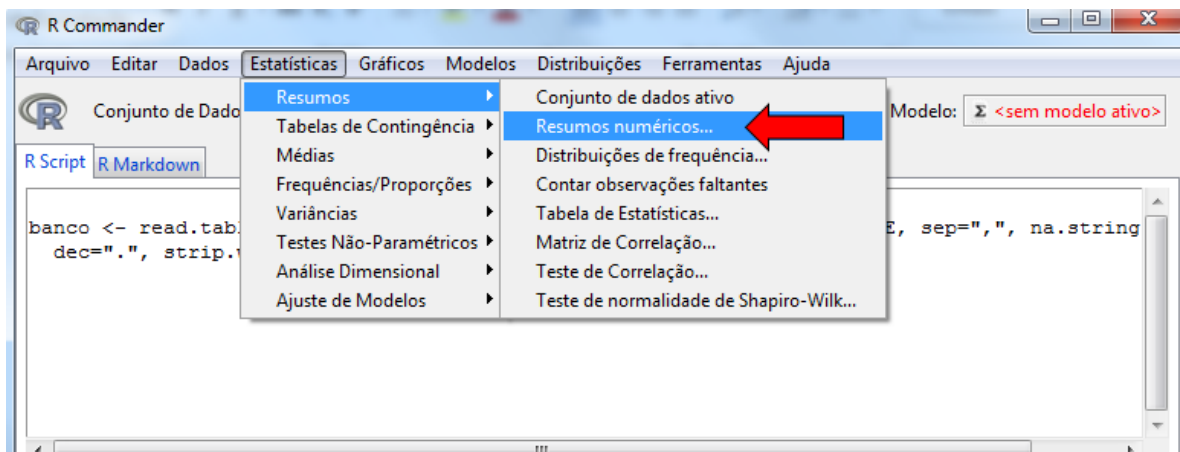
Obtendo o resumo da variável quantitativa “age” pelo script:

```
summary(banco$age)
```

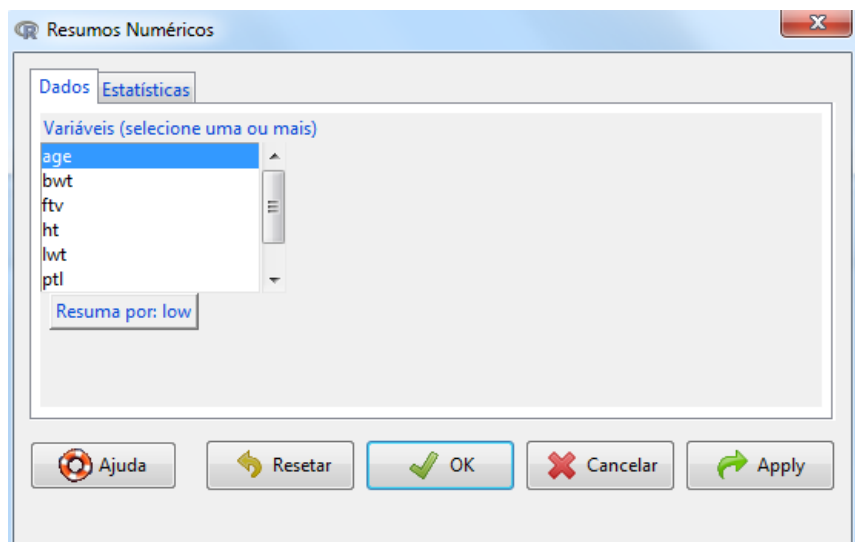
```
> summary(banco$age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 14.00  19.00   23.00   23.24  26.00   45.00
```

Uma outra forma se obter resumos estatísticos de variáveis quantitativas pelo R Commander:

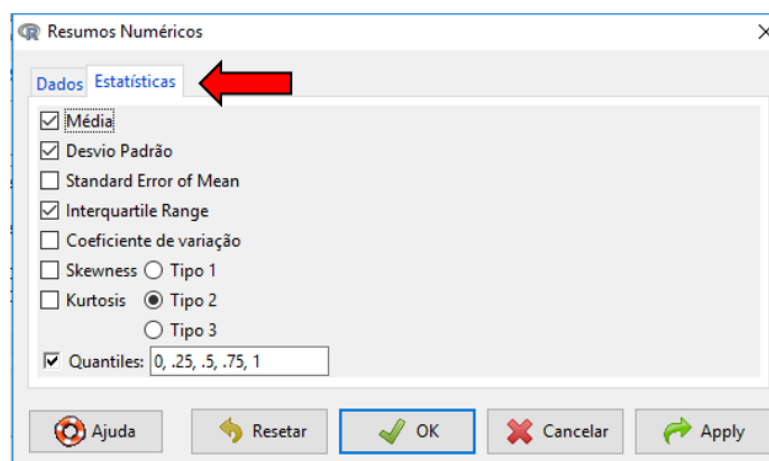
“Estatística -> Resumos -> Resumos Numéricos”



Selecione a(s) variável(is) que deseja realizar o resumo (Para selecionar mais de uma, utilize o CTRL).
OBS: Se desejar que os resultados saiam separados por alguma categoria de outra variável, clique no botão “Resuma por” e escolha a variável qualitativa correspondente a estratificação.



Clique na aba “Estatística” e escolha as medidas-sumário desejadas (exemplo: média, quartis, desvio-padrão) e dê OK.



Com a variável “age” temos o seguinte resultado:

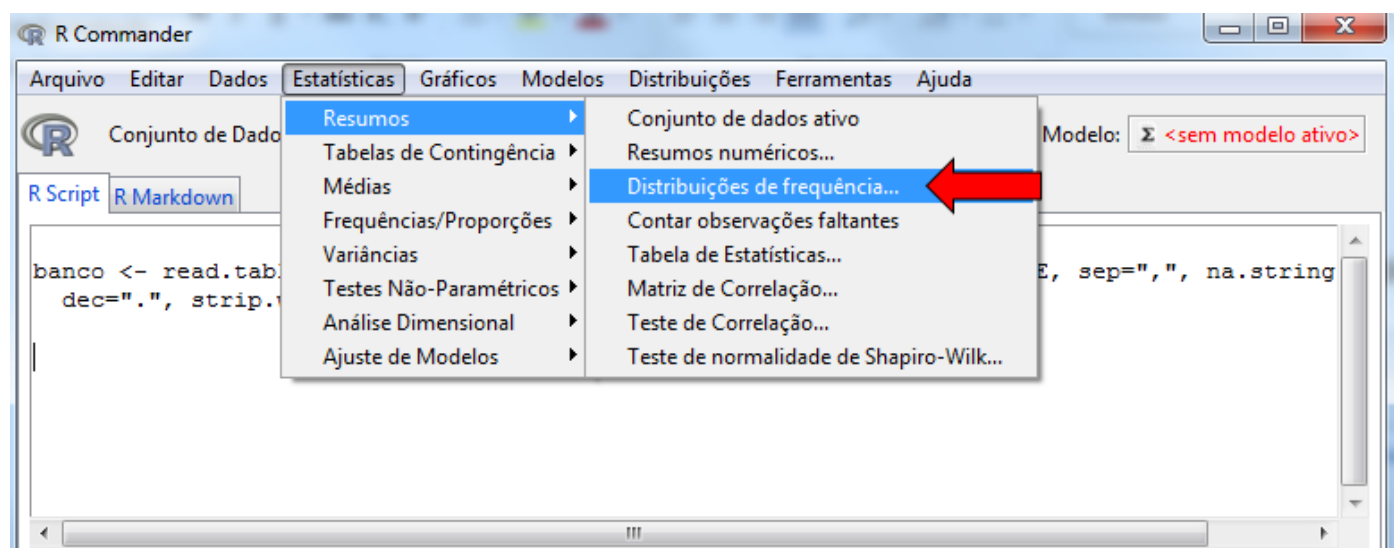
```
Output
```

```
> numSummary(banco[, "age"], statistics=c("mean", "sd", "IQR", "quantiles"),  
+ quantiles=c(0, .25, .5, .75, 1))  
  mean      sd IQR 0% 25% 50% 75% 100%   n  
23.2381 5.298678   7 14  19  23  26  45 189
```

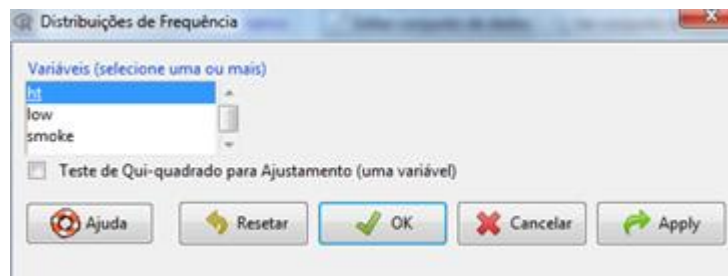
Tabelas de distribuição de frequências para variáveis qualitativas (ex:ht, low, smoke) podem ser realizadas pelo comando “table()” no script. Exemplo: O comando “table(banco\$ht)” cria uma distribuição de frequência para a variável “ht”.

No R Commander:

“Estatísticas-> Resumos -> Distribuições de frequência “



Feito isso escolhe-se a(s) variável(is) que se deseja realizar a distribuição de frequência. Vamos escolher a variável “ht” e dar OK.



Observação: Observe que a caixa de diálogo só mostra as variáveis definidas como qualitativas (factor), visto que não faz sentido tirar % de variáveis quantitativas.

```

Output
Submeter

> local({
+   .Table <- with(banco, table(ht))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table), 2))
+ })

counts:
ht
No yes
177 12

percentages:
ht
No yes
93.65 6.35

```

Outras medidas sumárias do banco de dados podem ser realizadas em subconjuntos de dados (filtros).

Exemplo: Vamos pedir o sumário da variável “age” das mães com filhos que nasceram com baixo peso.

`summary(banco$age[banco$low=="<2.5"])`

```

Output
Submeter

> summary(banco$age[banco$low=="<2.5"])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 14.00  19.50   22.00   22.31  25.00   34.00     1

```

Vamos pedir o sumário da variável “age” das mães com filhos que nasceram com peso normal:

`summary(banco$age[banco$low==">2.5"])`

Vamos agora pedir o sumário da variável “age” das mães com filhos com baixo peso e que fumaram:

`summary(banco$age[banco$low=="<2.5"&banco$smoke=="Sim"])` ou
`summary(banco$age[banco$low=="<2.5"&banco$smoke=="Sim"])`

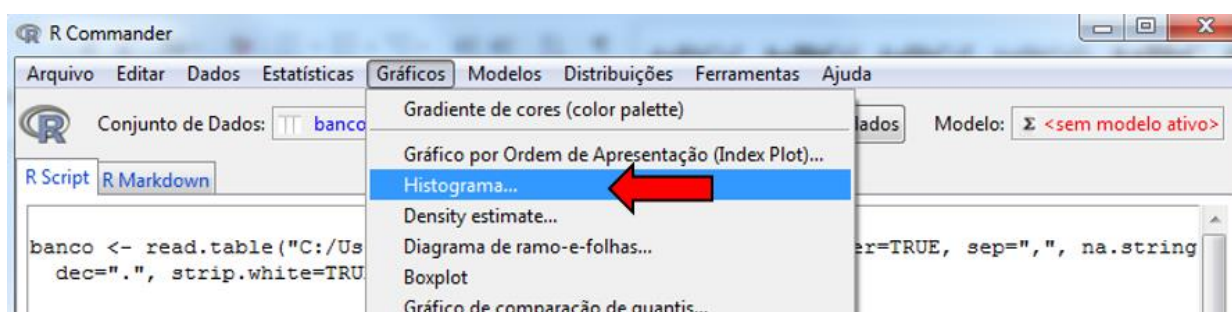
Gráficos

Histograma

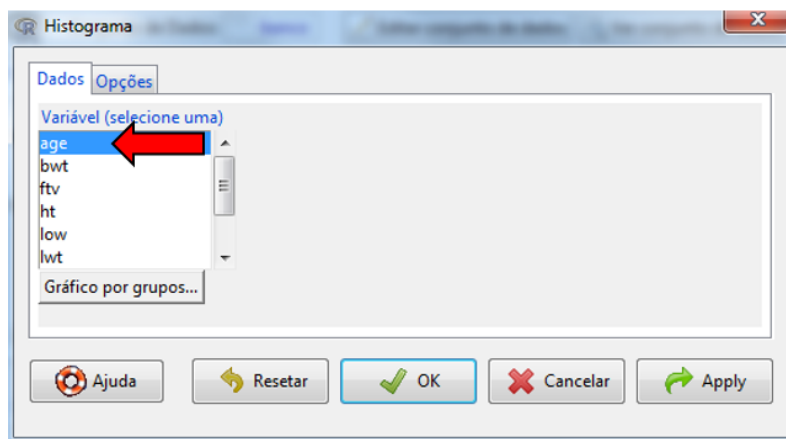
Para fazer um Histograma para a variável quantitativa “age” digite o comando na janela de script:

```
h1<-hist(banco$age,main="Histograma da variável idade")
```

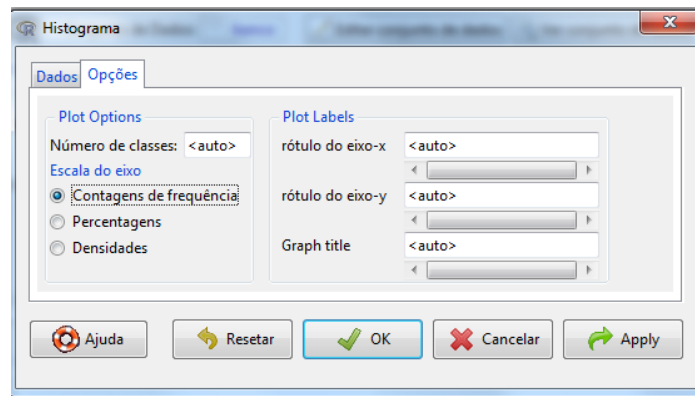
Ou faça o seguinte no menu do R Commander.



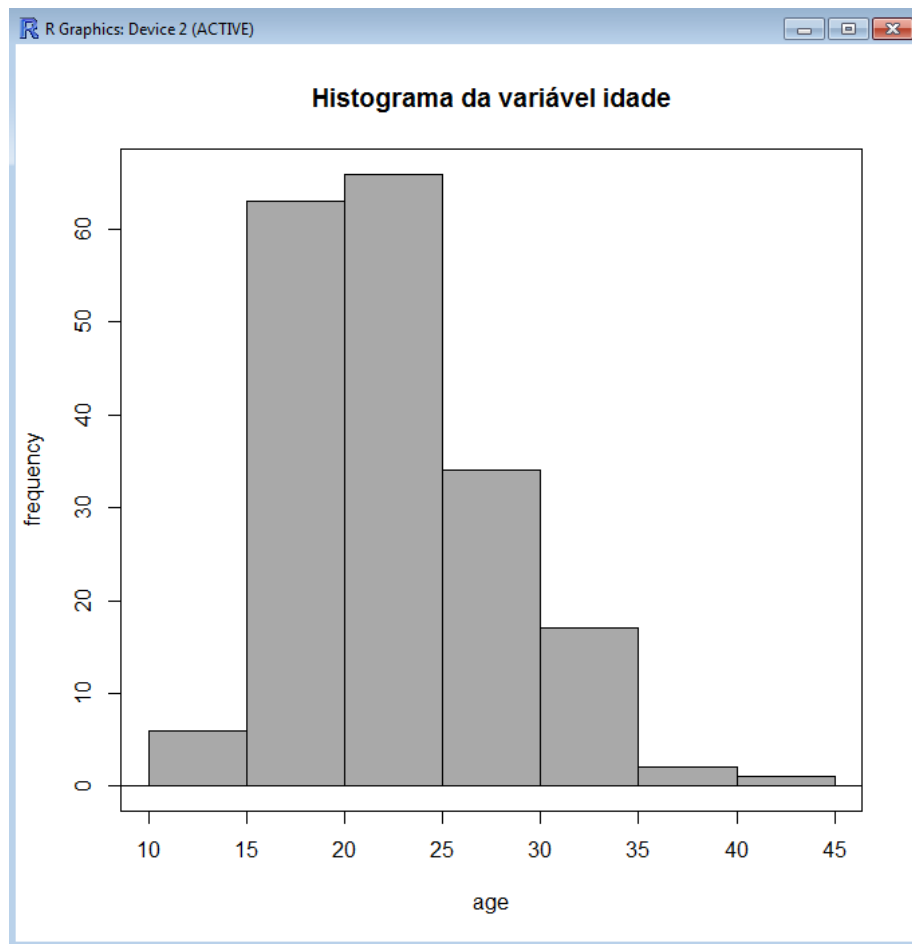
Ao clicar na opção do Histograma aparecerá a seguinte caixa de diálogo para realizar a escolha da variável, nosso exemplo, escolha a variável age.



Para criar uma legenda, modificar o rótulo dos eixos x e y ou mudar a escala do eixo, clique na aba “Opções” e parecerá a seguinte tela:



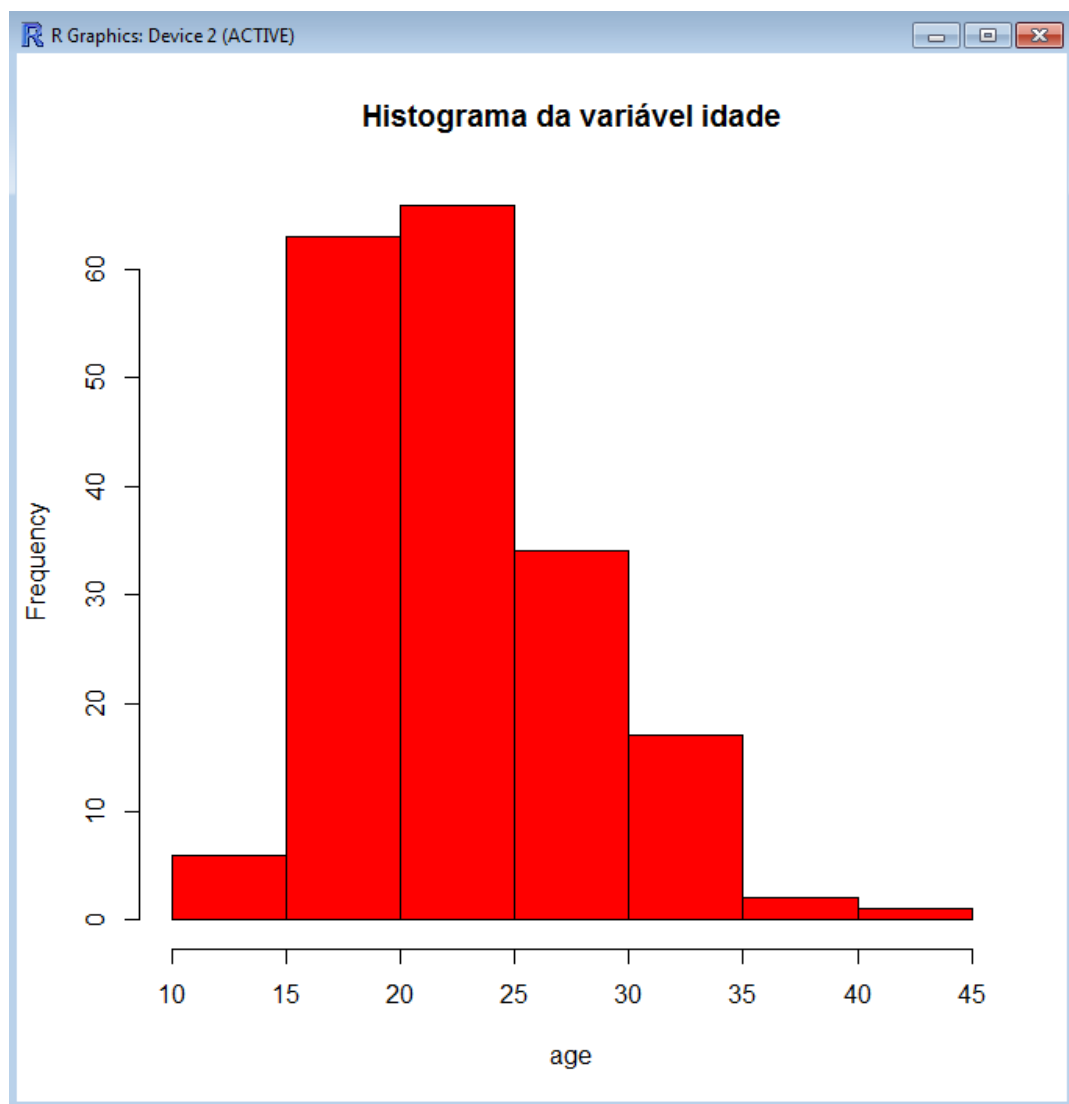
Clique em OK. Para o nosso exemplo aparecerá o seguinte gráfico:



OBS: Só é possível mudar as cores do gráfico modificando e executando o comando no script!

No exemplo abaixo, editamos o comando anterior para criar um histograma de cor vermelha.

```
h1<-hist(banco$age,main="Histograma da variável idade",xlab="age",col="red")
```



Histograma para variável age

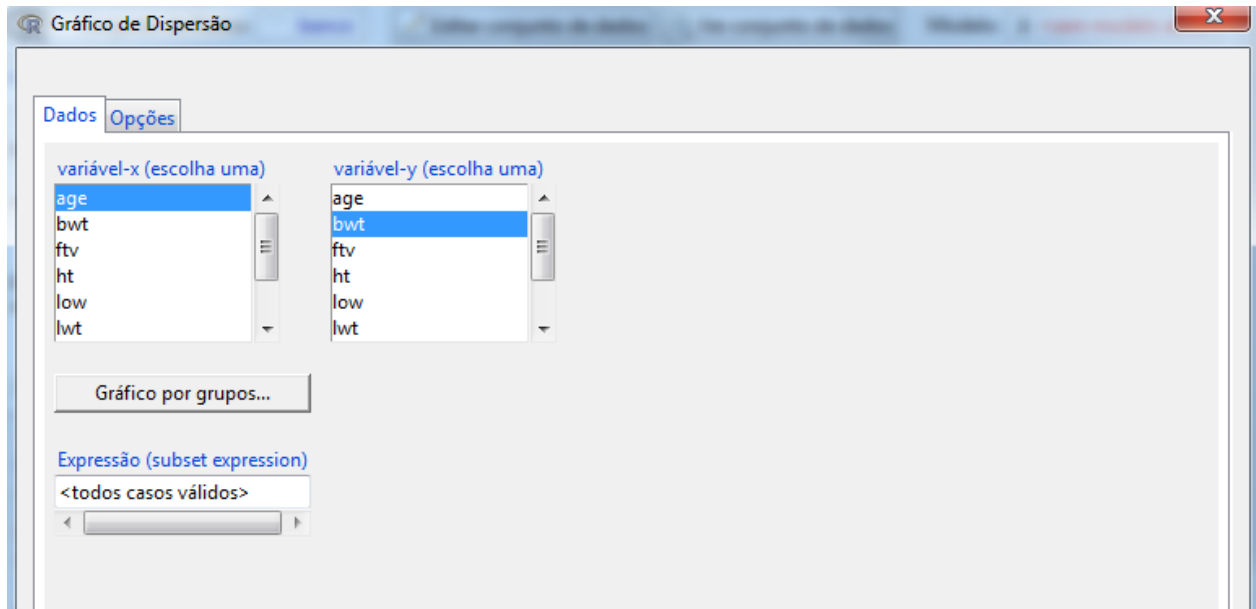
```
h2<-hist(banco$age,main="Histograma da variável idade", xlab="age",col="red",probability=TRUE)
```

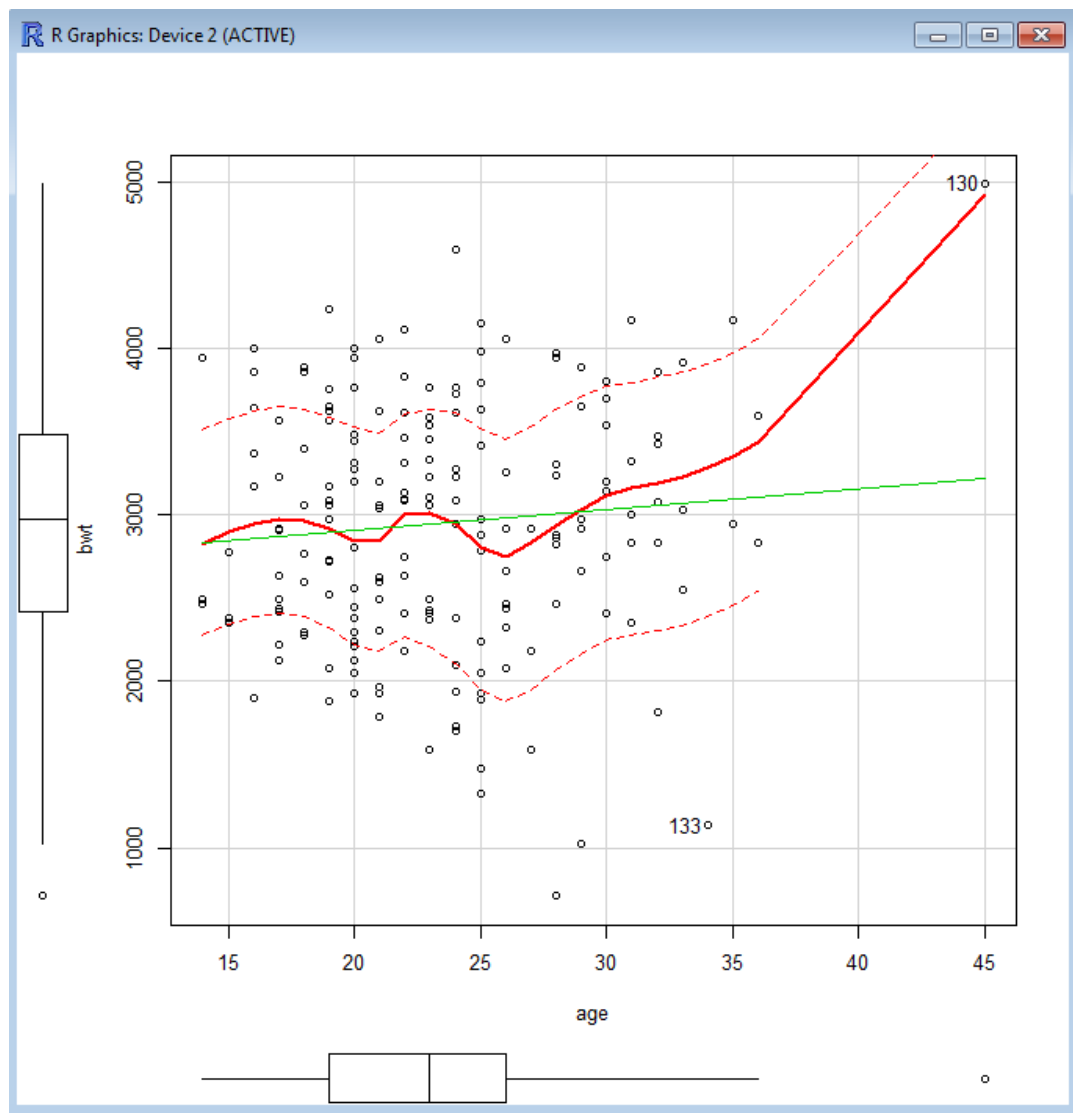
Gráfico de Dispersão

Digite no script `plot(banco$age,banco$bwt)`

ou clique no menu do R Commander: “Gráficos> Diagrama de dispersão”

Escolhas as variáveis desejadas para os eixos x e y e clique em Ok. No caso do nosso exemplo, queremos ver a relação entre as variáveis “age” e “bwt”.





Pergunta: O que acontece com o peso da criança quando a idade da mãe tende a crescer?

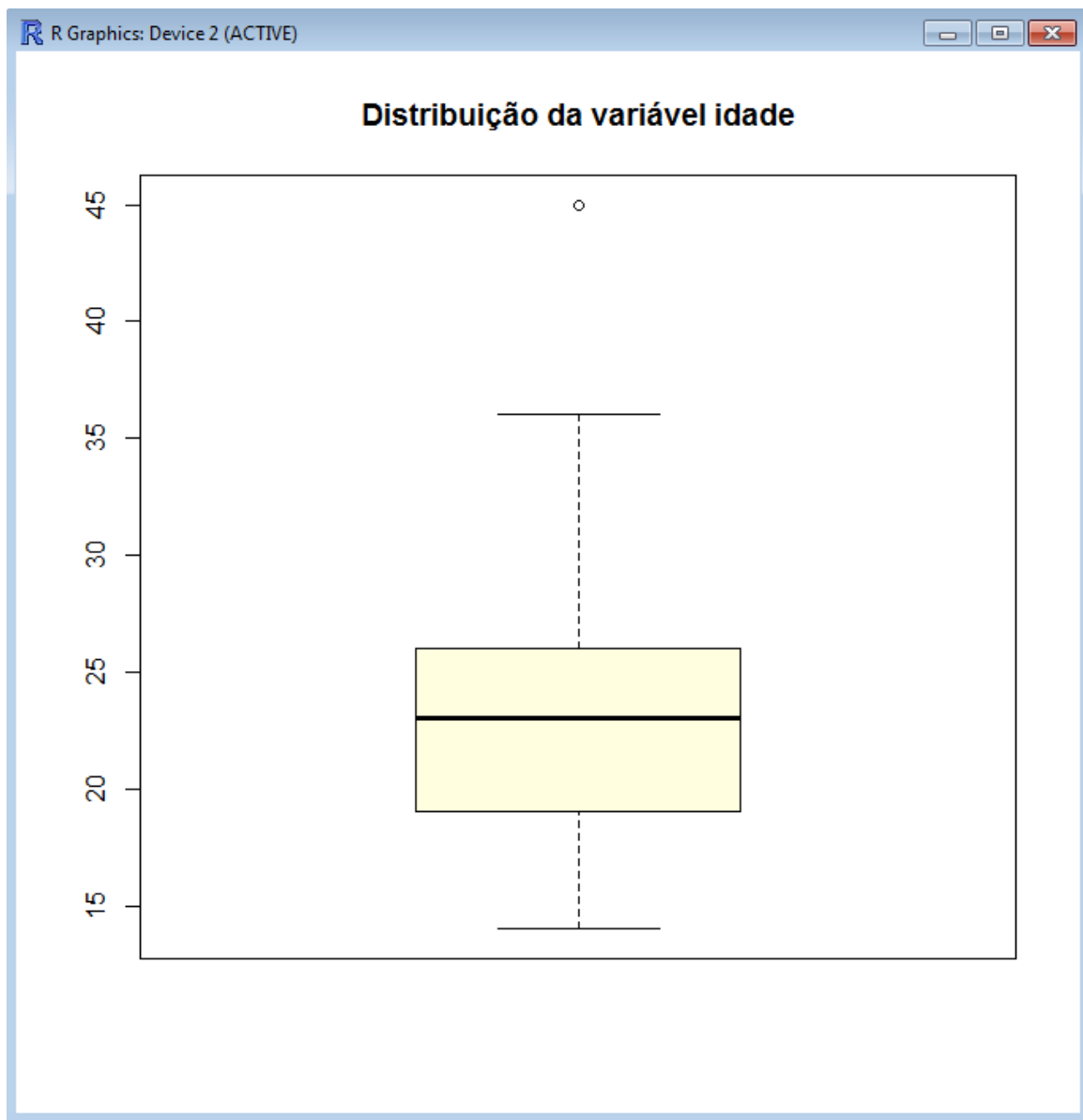
Box Plot

Para criar um boxplot da variável “age” por exemplo, faça o seguinte caminho no R Commander:

“Gráficos->Boxplot”

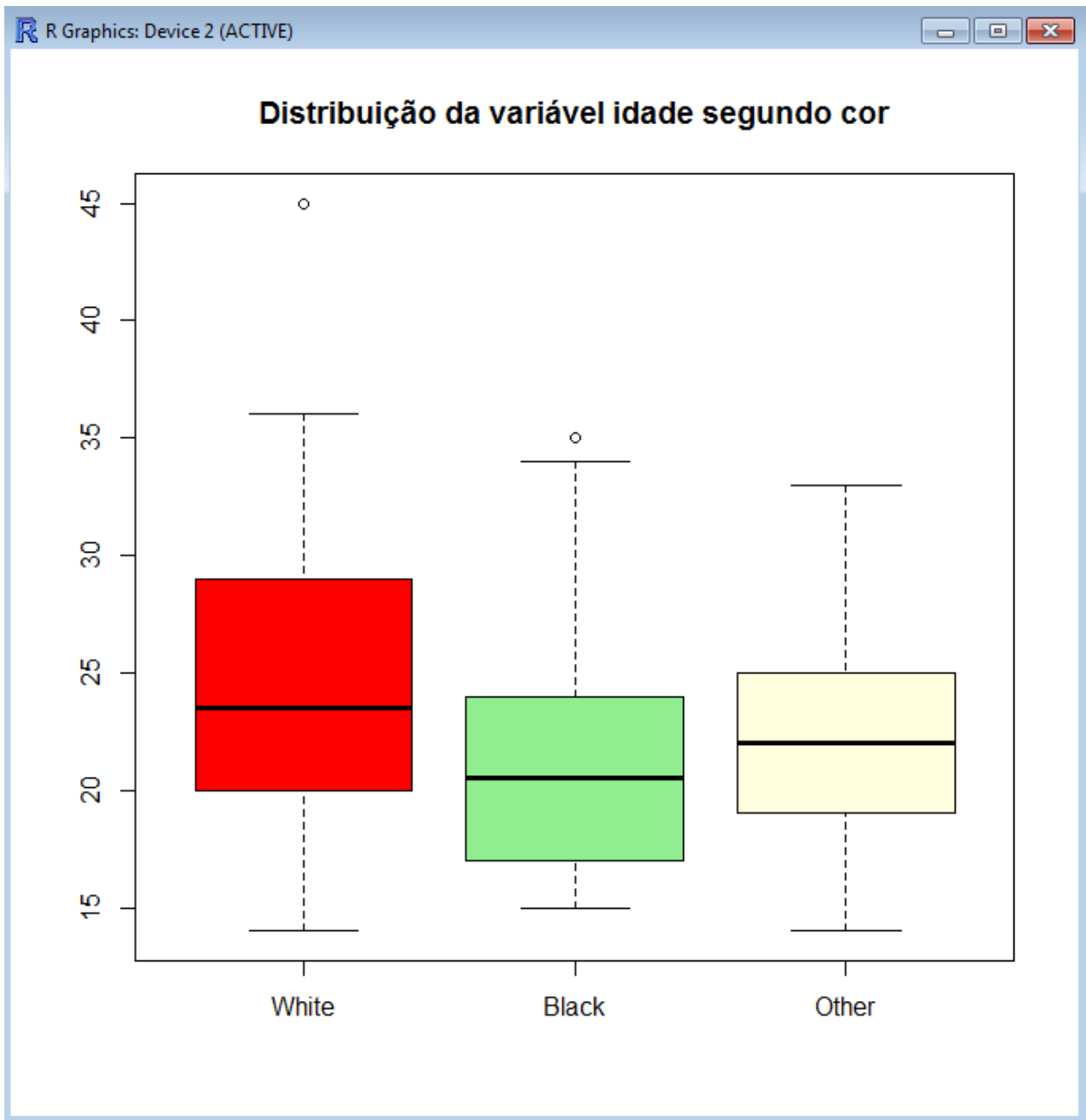
ou digite o seguinte comando no R script:

`boxplot(banco$age,main="Distribuição da variável idade", col="lightyellow")`



Boxplot da distribuição da variável quantitativa idade segundo a variável qualitativa cor. No R Script digite o seguinte comando:

```
boxplot(banco$age~banco$race,main="Distribuição da variável idade segundo cor",  
col=c("red","lightgreen","lightyellow"))
```



Pergunta: Parece ter diferença da idade nos grupos?

Gráfico de Setores (Pizza)

Para criar um gráfico de setores para variável “race” digite o comando no R Script:

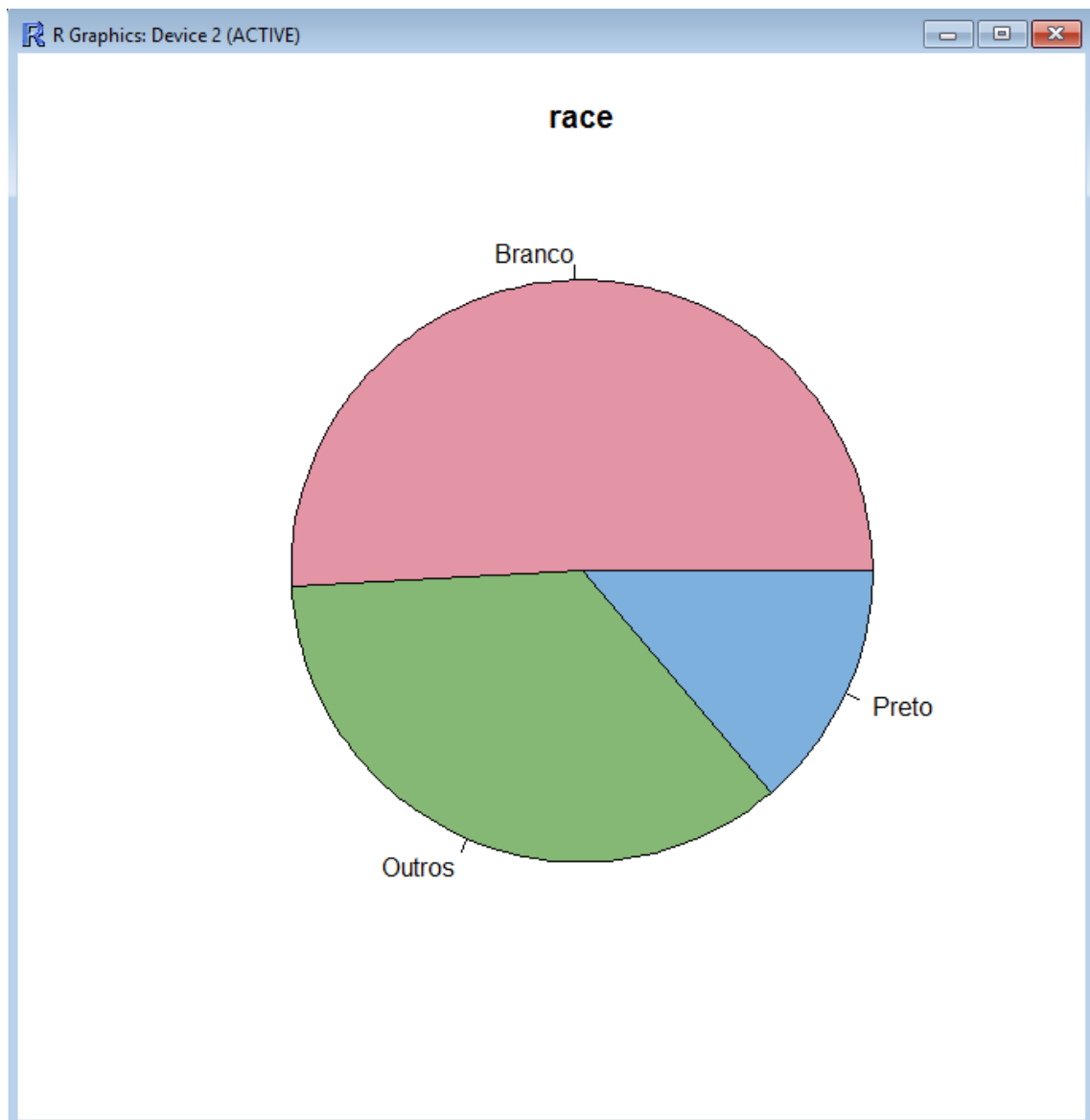
```
pie(table(banco$race))
```

ou faça o seguinte caminho no R Commander

“Gráficos > Gráfico de pizza “

Aparecerá a opção com todos os tipos de variáveis qualitativas, selecione a variável desejada e escolha as opções do gráfico.

Para o nosso exemplo o gráfico de setores para a variável “race” fica da seguinte forma:

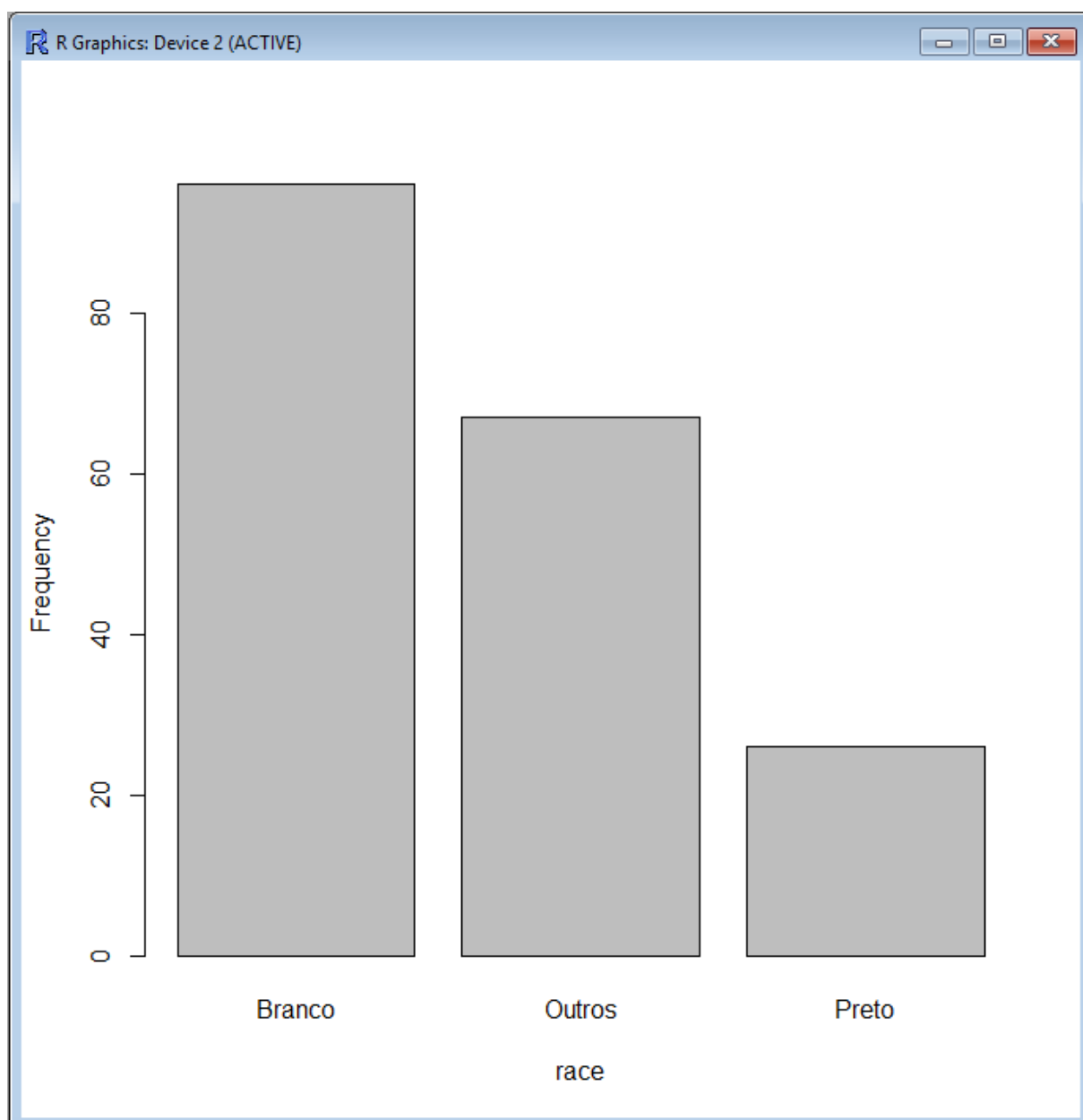


OBS: Só é possível colocar a porcentagem de cada categoria pelo R script!!! Editando os comandos anteriores:

```
pie(table(banco$race),main="Gráfico de setores: Raça")
text(locator(n=1),paste(round(prop.table(table(banco$race))[1],digits=2)*100,"%"))
text(locator(n=1),paste(round(prop.table(table(banco$race))[2],digits=2)*100,"%"))
text(locator(n=1),paste(round(prop.table(table(banco$race))[3],digits=2)*100,"%"))
```

Após executar o comando, clique em cada categoria para adicionar seus respectivos valores.

Da mesma maneira podemos criar um gráfico de barras para a variável “race” no R Commander



OBS: Caso no banco de dados não existe nenhuma variável do tipo fator (variáveis qualitativas) as opções de gráficos de setores e de barra ficam desabilitadas.

Gráfico de Linha

Esse banco de dados não possui dados com temporalidade. Assim, para fazer um gráfico de linha vamos criar dois vetores com informações de taxa de mortalidade e os anos de ocorrência, a fim de avaliar a evolução temporal de uma taxa ao longo do tempo. *OBS: Se os dados tivessem digitados previamente em uma planilha bastaria importar os dados.*

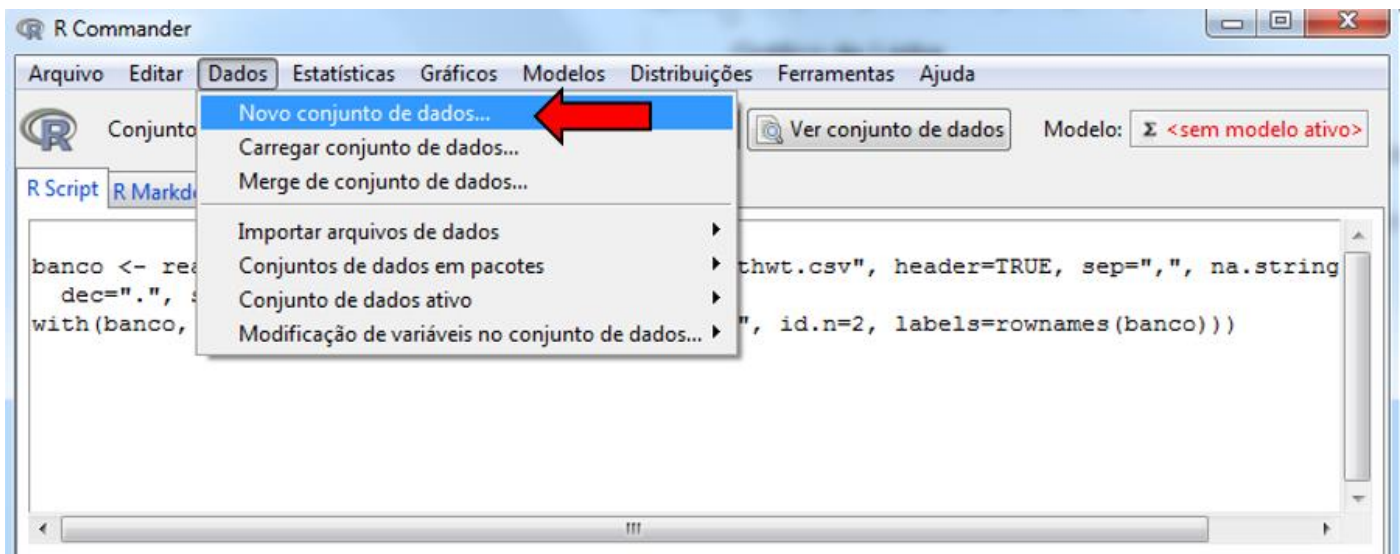
Na janela de script:

```
TM<-c(50,40,25,35,10,5)
anos<-c(1980,1985,1990,1995,2000,2005)
tabela<-as.data.frame(cbind(TM,anos))
tabela
```

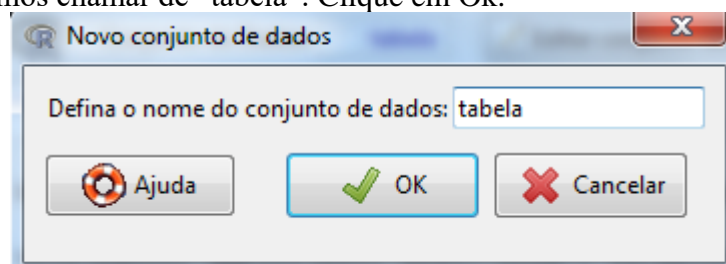
```
plot(tabela$anos, tabela$TM, type="l")
```

No menu do R Commander:

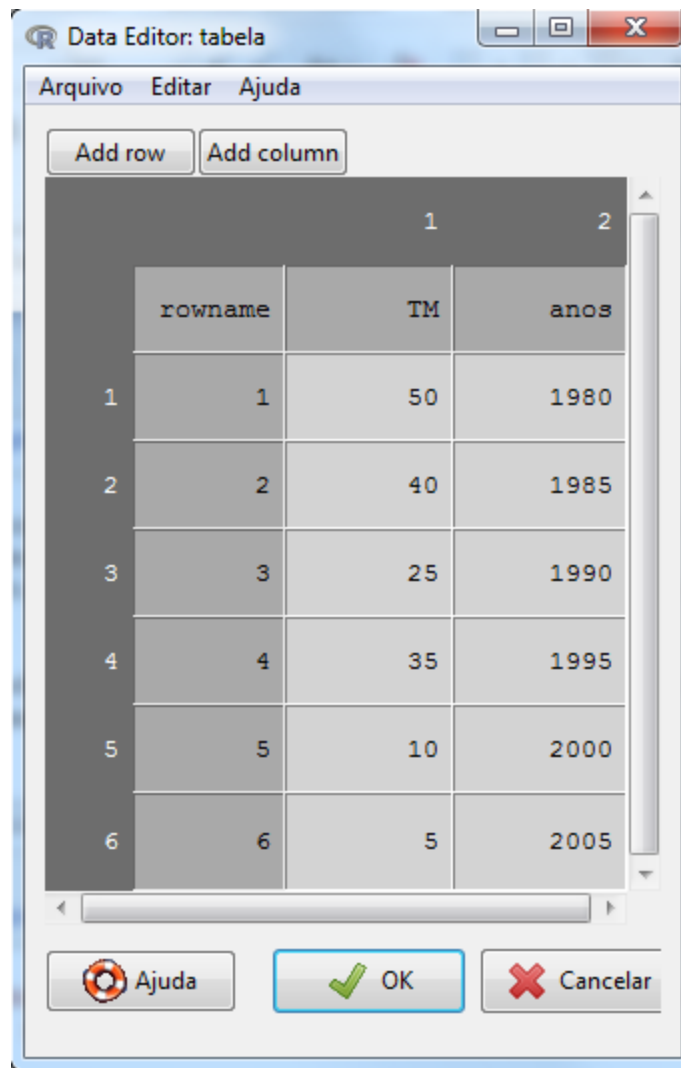
Clique em “Dados > Novo conjunto de dados”



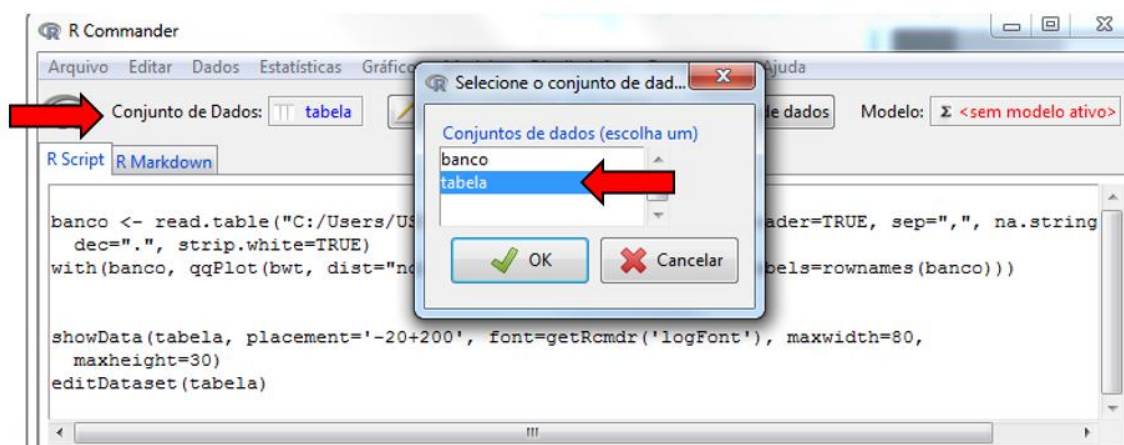
Então aparecerá a opção para definir o nome do novo conjunto de dados. Para o nosso exemplo vamos chamar de “tabela”. Clique em Ok.



Aparecerá uma opção para adicionar linhas e colunas. Adicione linhas e colunas conforme dado em nosso exemplo:

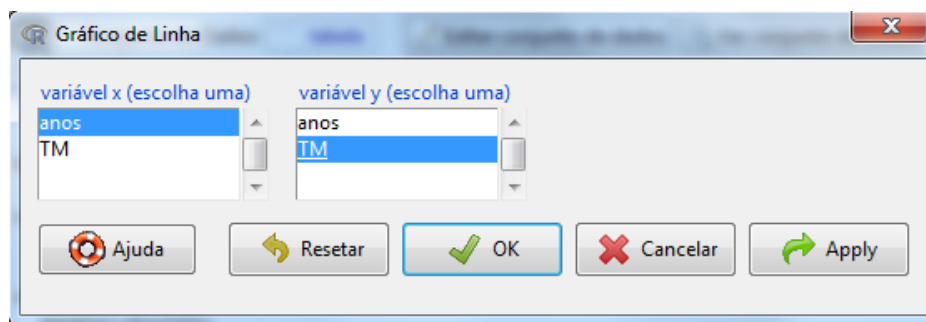


Atenção: Após carregar mais de um banco, deve-se informar ao R qual o banco de dados está ativo para o menu. A troca do banco de dados ativo pode ser feita clicando no campo “Conjunto de dados:” onde aparecerá as opções de banco. Para o nosso exemplo escolha o banco “tabela” que acabamos de criar.

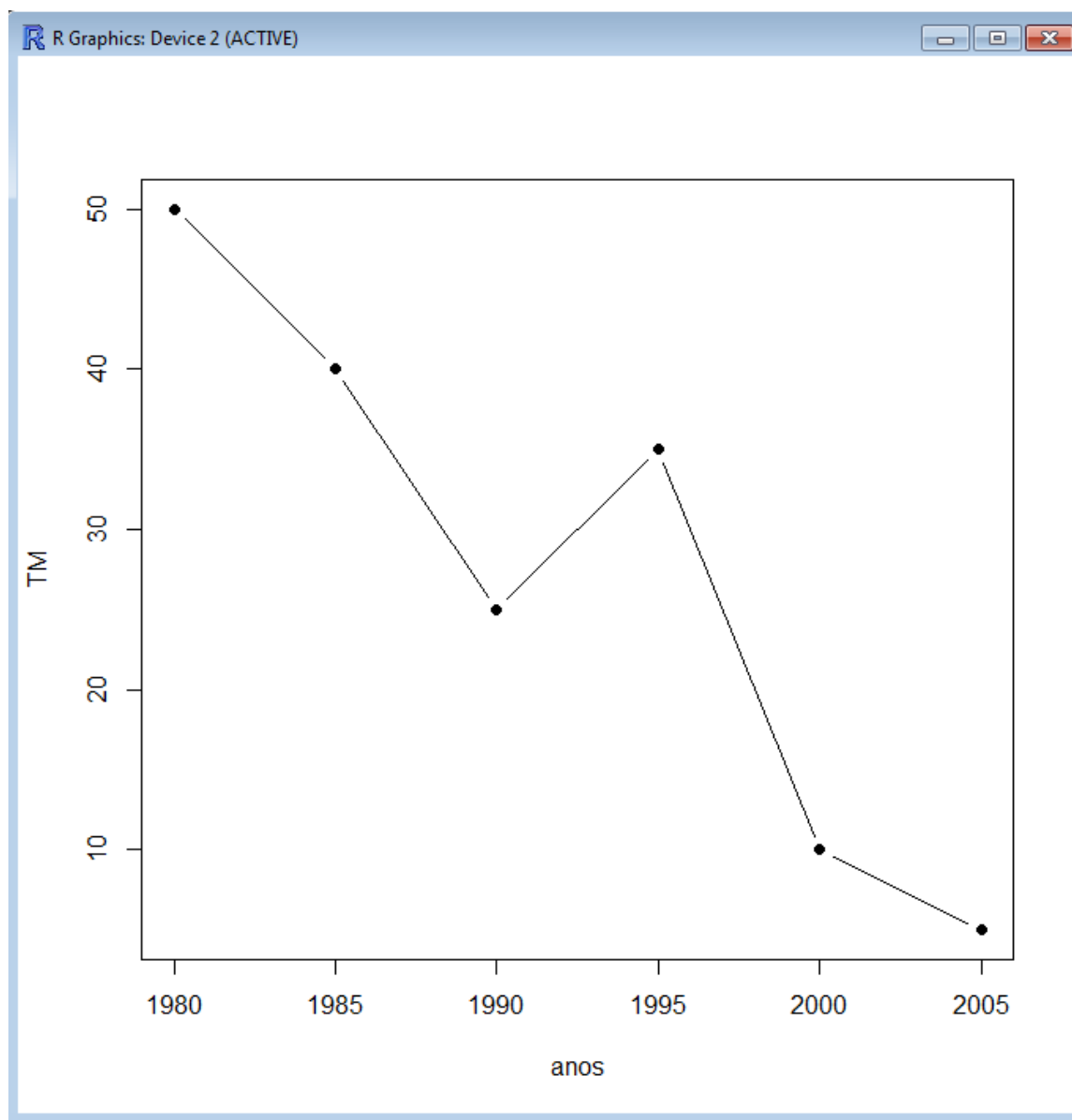


Agora vá em “Gráficos > Grafico de linha”

Selecione as variáveis desejadas, que nesse exemplo são “TM” e “anos”



Clique em ok e então parecerá o seguinte resultado:

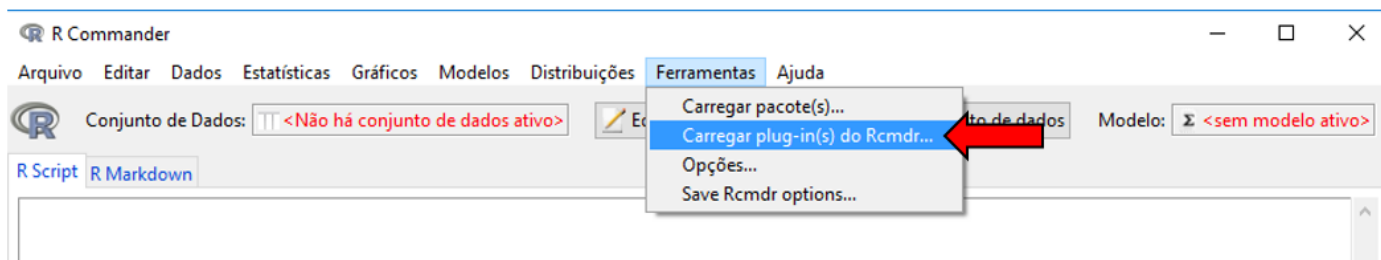


Carregando o Plugin EZR

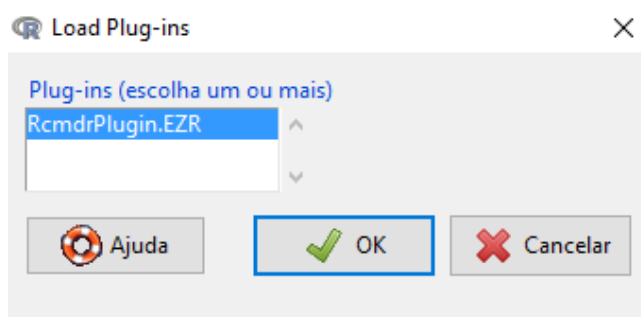
O plug-in “Easy R” EZR realiza diversas análises de forma mais fácil e com mais possibilidade de personalização. Além disso, permite a realização de comandos específicos à epidemiologia. O EZR modificará as opções de menu do R Commander. Antes de carregá-lo salve seu script e área de trabalho.

Após carregar o ambiente do R Commander, vamos carregar o “RcmdrPlugin.EZR” no seguinte caminho:

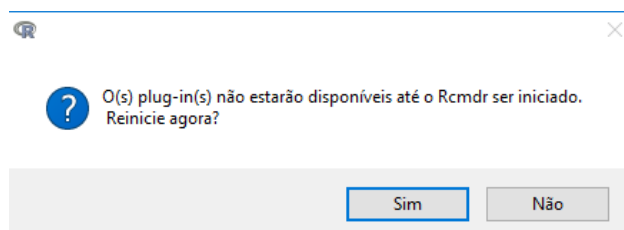
Ferramentas>Carregar plug-in(s) do Rcmdr..



Escolha o plugin desejado, no caso RcmdrPlugin.EZR e clique em OK



Aparecerá uma mensagem dizendo que o R Commander será reiniciado clique em “Sim”



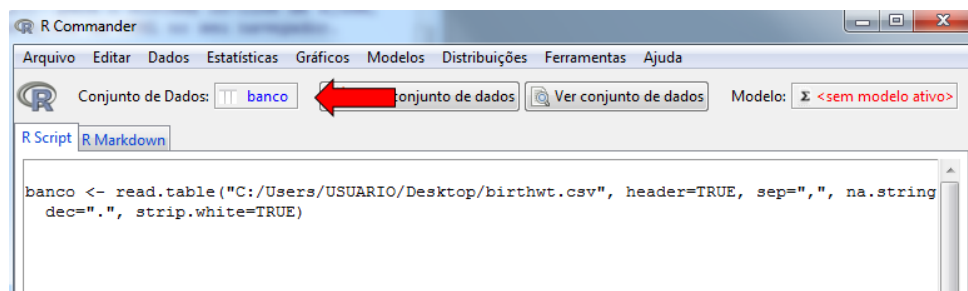
Então o R Commander irá reinicializar e abrir automaticamente. Clique no conjunto de dados ativo e verifique que todos os bancos de dados continuam na memória. Entretanto, o script é perdido, caso não esteja salvo previamente.

Dados para o R Commander no RcmdrPlugin.EZR

Note que antes de carregar ou criar um banco de dados, no campo conjunto de dados, está indicando a inexistência de qualquer informação ou dados a serem estudados.



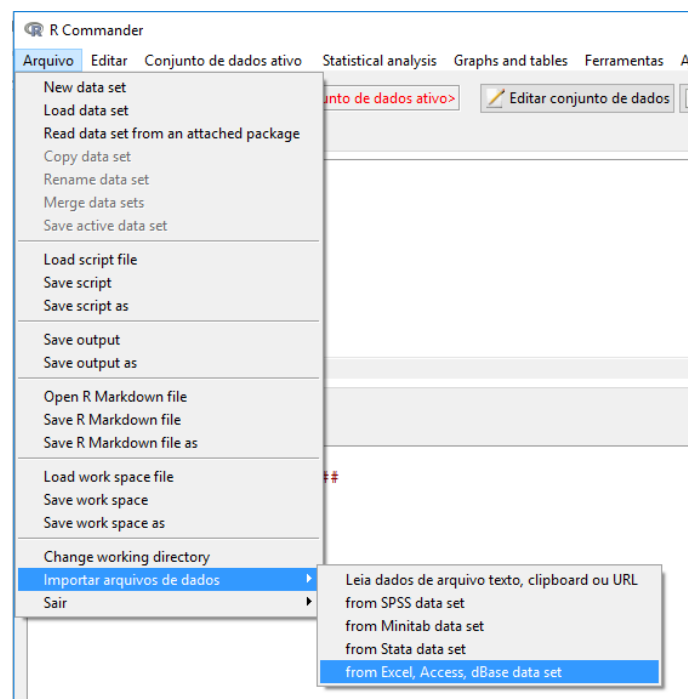
Esse status muda ao carregar/ou criar um banco conforme indicado no exemplo abaixo, em que o status do campo “Conjunto de Dados” mudou ao ser carregado um arquivo de dados nomeado como “banco”.



Para carregar um novo banco de dados, basta utilizar o menu do RCommander:

Arquivo> Importar arquivos de dados

E escolha o tipo de arquivo em que deseja importar.

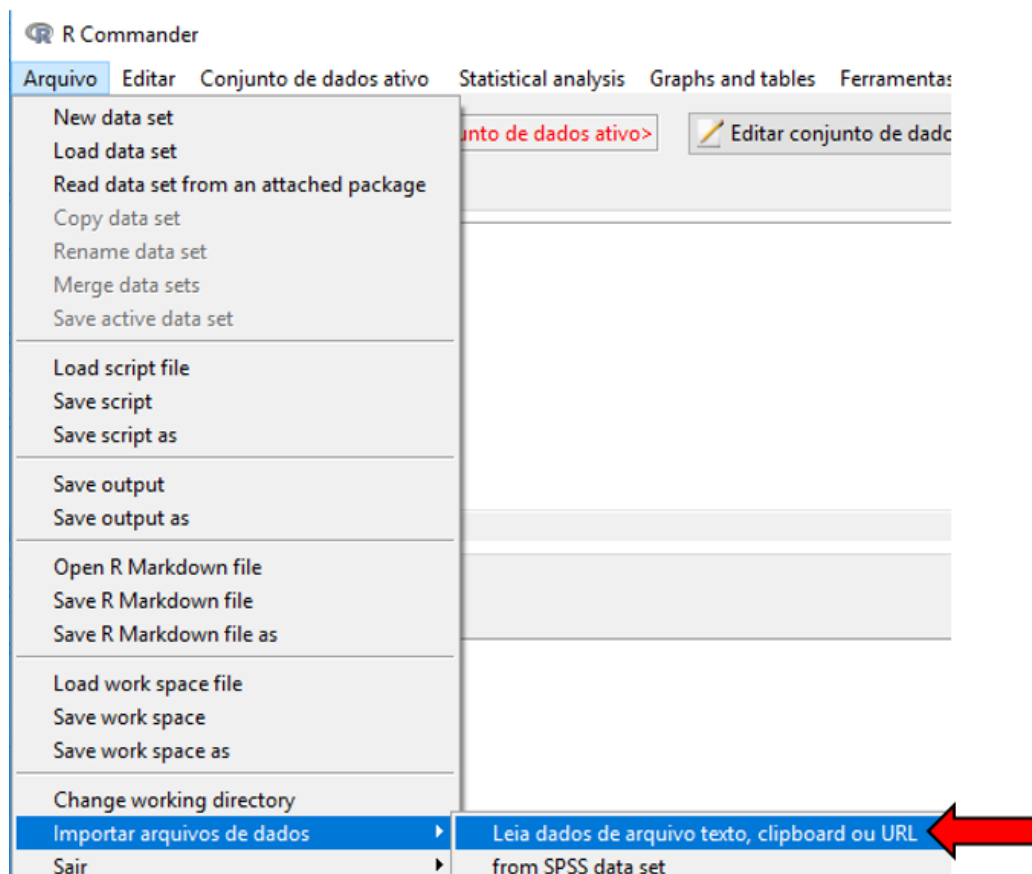


Análise Descritiva no EZR

Vamos agora realizar uma análise descritiva do banco de dados “*birthwt.csv*” utilizando as ferramentas do EZR.

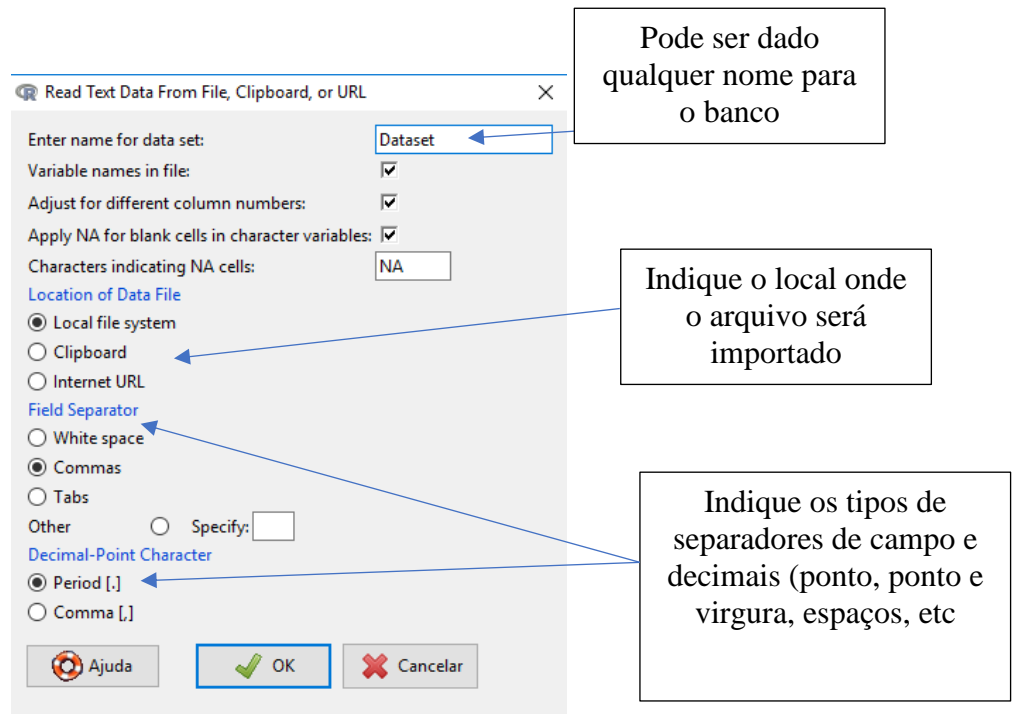
Primeiro vamos carregar o banco de dados para a nossa análise no EZR. Faça o seguinte caminho:

Arquivo > Importar arquivos de dados > Leia dados de arquivos texto, clipboard ou URL



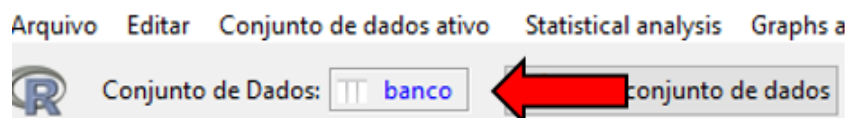
OBS: No exemplo acima esse caminho é utilizado para importação de arquivos do tipo csv, porém há outras extensões que pode ser importadas pelo mesmo caminho.

No caso do nosso exemplo aparecerá as seguintes opções para o arquivo:



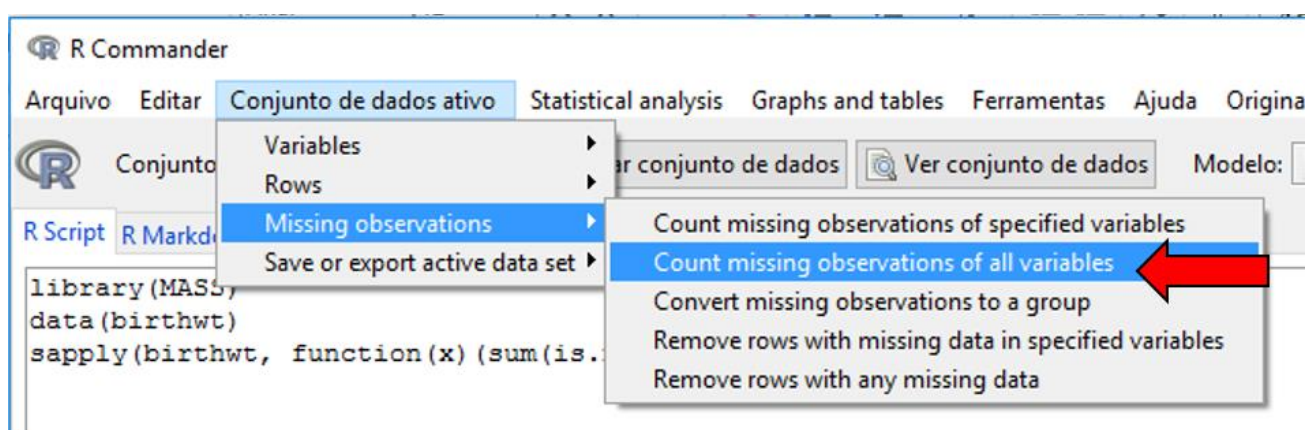
Selecione o nome que deseja dar ao seu banco, os separadores de campo e de decimais. Após indique o diretório onde o arquivo está salvo, selecione e clique em ok.

Para o nosso exemplo chamaremos o conjunto de dados de “banco” e após carregamento no o mesmo deve estar ativo no campo “Conjunto de dado:” conforme figura abaixo.



Verificando se há dados faltantes:

Conjunto de dados ativos > Missing observation > Count missing observation of all varablle

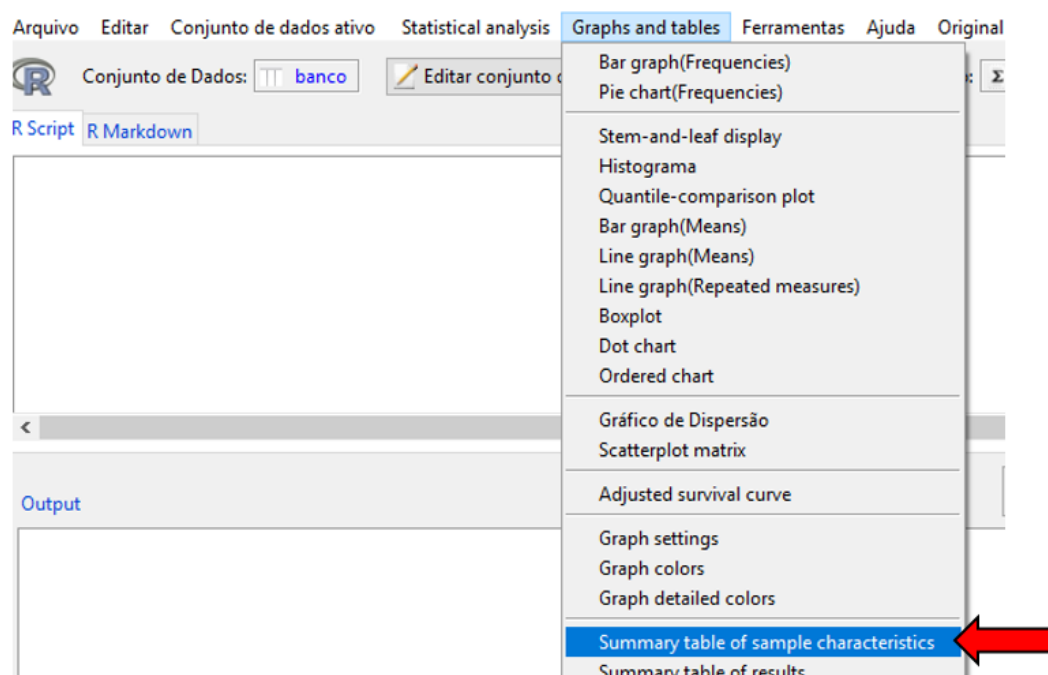


Aparecerá o seguinte resultado contabilizando se há ou não variáveis com dados faltantes (missings)

```
Output
> sapply(banco, function(x) (sum(is.na(x)))) # NA counts
      X      low      age      lwt      race      smoke
0      0      0      0      0      0      0
    ptl      ht      ui      ftv      bwt      cor
0      0      0      0      0      0
faixaetaria relação.fumo.peso
0      0
```

Análise Descritiva

Clique no menu em: **Graphs and tables > Summary table sample characteristic**



Aparecerá as seguintes opções conforme tela abaixo:

The screenshot shows a software dialog box titled "Summary table of sample characteristics". It contains several sections for variable selection and analysis options. Four red arrows with numbers 1 through 4 point to specific areas:

- 1** points to the "Grouping variable(pick 0 or 1)" section, which has a list of variables (age, bwt, ftv, ht, low, lwt, pti, race, smoke, ui) and a scroll bar.
- 2** points to the "Categorical variables" section, which also has a list of the same variables and a scroll bar.
- 3** points to the "Continuous variables (normal distribution)" section, which has a list of the same variables and a scroll bar.
- 4** points to the "Continuous variables (non-normal distribution)" section, which also has a list of the same variables and a scroll bar.

Below these sections are radio button options for "Test for categorical variables" (Chi-square, Fisher's exact, Automatic selection), "Range for non-normal categorical variables" (Minimum and maximum values, Interquartile ranges), and "Show explanation for continuous variables" (No, Yes). There are also options for "Output destination" (Clipboard, CSV file) and "Language" (English, Local). A text field for "Condition to limit samples for analysis" is present, with a default value "<all valid cases>". At the bottom are buttons for "Ajuda", "Resetar", "OK", "Cancelar", and "Apply".

1-Grouping variable: Variáveis de agrupamento são variáveis qualitativas utilizadas para agrupar ou estratificar observações. As variáveis de agrupamento são úteis para resumir ou visualizar dados por grupo. Neste campo, escolhe-se no máximo apenas uma variável.

2- Categorical variables – Para definir as variáveis qualitativas ou categóricas. Podem ser selecionadas uma ou mais variáveis.

3- Continuous variables (normal distribution) – Local para definir as variáveis quantitativas do tipo contínua com distribuição normal. Podem ser selecionadas uma ou mais variáveis.

3- Continuous variables (non-normal distribution) – Local para selecionar as variáveis tipo qualitativa contínua que não supõe distribuição normal. Pode ser seleciona uma ou mais variáveis.

Atenção: Os campos não pré-definem os tipos de variáveis. Em cada campo são mostradas todas as variáveis sem discriminar seu tipo (categórica, continua com ou sem distribuição normal). O usuário que deve selecionar corretamente as variáveis de forma adequada à cada campo.

Exemplo: Suponha que que estejamos interessados em verificar a relação da variável *low* com as demais variáveis, ou seja, verificar quais os fatores estão associados ao baixo peso ao nascer.

No campo Grouping variable selecione a variável *low* e nos demais campos selecione as variáveis de acordo com seu tipo e clique em OK.

Summary table of sample characteristics

Grouping variable(pick 0 or 1)

ftv
ht
low
lwt
ptl
race
relação.fumo.peso
smoke
ui
X

Click pressing Ctrl key to select multiple variables

Categorical variables

cor
faixaetaria
ftv
ht
low
lwt
ptl
race
relação.fumo.peso
smoke

Continuous variables (normal distribution)

age
bwt
cor
faixaetaria
ftv
ht
low
lwt
ptl
race

Continuous variables (non-normal distribution)

age
bwt
cor
faixaetaria
ftv
ht
low
lwt
ptl
race

Test for categorical variables

☐ Chi-square test with continuity correction
☐ Fisher's exact test
☒ Automatic selection

Range for non-normal categorical variables

☒ Minimum and maximum values
☐ Interquartile ranges

Show explanation for continuous variables

☒ No
☐ Yes

Output destination Language

☒ Clipboard ☒ English
☐ CSV file ☐ Local

Clipboard can be selected only in Windows.

Condition to limit samples for analysis. Ex1. age>50 & Sex==0 Ex2. age<50 | Sex==1

<all valid cases>

Ajuda Resetar **OK** Cancelar Apply

A análise será fará todos os testes, relacionado cada variável ao desfecho low (análises bivariadas). O resultado no output segue em forma de tabela conforme mostrado na figura abaixo.

Output

Submiter

```

> FinalTable <- rbind(n=row1, FinalTable)

> print(FinalTable[,2:length(FinalTable[1,])], quote=FALSE)

      Group      <2.5      >2.5      p.value
faixaetaria (%)  Menor de 20    15 (25.4)    36 (27.7)    0.162
                  21 a 29      40 (67.8)    71 (54.6)
                  31 a 39       4 ( 6.8)    22 (16.9)
                  40 ou mais    0 ( 0.0)     1 ( 0.8)
ht (%)          Não          52 (88.1)   125 (96.2)    0.052
                  Sim           7 (11.9)     5 ( 3.8)
race (%)        Branco       23 (39.0)    73 (56.2)    0.079
                  Outros        25 (42.4)    42 (32.3)
                  Preto         11 (18.6)    15 (11.5)
relação.fumo.peso (%) Não fuma e baixo peso 29 (49.2)     0 ( 0.0)    <0.001
                  Fuma e baixo peso 30 (50.8)     0 ( 0.0)
                  Fuma e peso normal  0 ( 0.0)    44 (33.8)
                  Não fuma e peso normal 0 ( 0.0)    86 (66.2)
smoke (%)       Não          29 (49.2)    86 (66.2)    0.036
                  Sim           30 (50.8)    44 (33.8)
lwt             122.14 (26.56)  133.30 (31.72) 0.020
age             22.00 [14.00, 34.00] 23.00 [14.00, 45.00] 0.247
ftv             0.00 [0.00, 4.00]  1.00 [0.00, 6.00] 0.238

> FinalTable <- rbind(row0, FinalTable)

> row0 <- rep("", length(colnames(FinalTable)))

```

O resultado também pode ser salvo para um arquivo csv, basta selecionar a opção “CSV file” antes de clicar em OK. Indique o diretório e o arquivo será salva automaticamente no local desejado.

Summary table of sample characteristics

Grouping variable(pick 0 or 1)

age

bwt

ftv

ht

low

lwt

ptl

race

smoke

ui

Click pressing Ctrl key to select multiple variables

Categorical variables

age

bwt

ftv

ht

low

lwt

ptl

race

smoke

ui

Continuous variables (normal distribution)

age

bwt

ftv

ht

low

lwt

ptl

race

smoke

ui

Continuous variables (non-normal distribution)

age

bwt

ftv

ht

low

lwt

ptl

race

smoke

ui

Test for categorical variables

☐ Chi-square test with continuity correction

☐ Fisher's exact test

☒ Automatic selection

Range for non-normal categorical variables

☒ Minimum and maximum values

☐ Interquartile ranges

Show explantation for continuous variables

☒ No

☐ Yes

Output destination

☐ Clipboard

☒ CSV file

Language

☒ English

☐ Local

Clipboard can be selected only in Windows.

Condition to limit samples for analysis. Ex1. age>50 & Sex==0 Ex2. age<50 | Sex==1

<all valid cases>

Ajuda

Resetar

OK

Cancelar

Apply