

# Aula prática 1- Manipulação de banco de dados (INI/FIOCRUZ)

Raquel de Vasconcellos Carvalhaes de Oliveira

Luan Nôe da Silva

Colaboração: Fabiano Marcos

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.

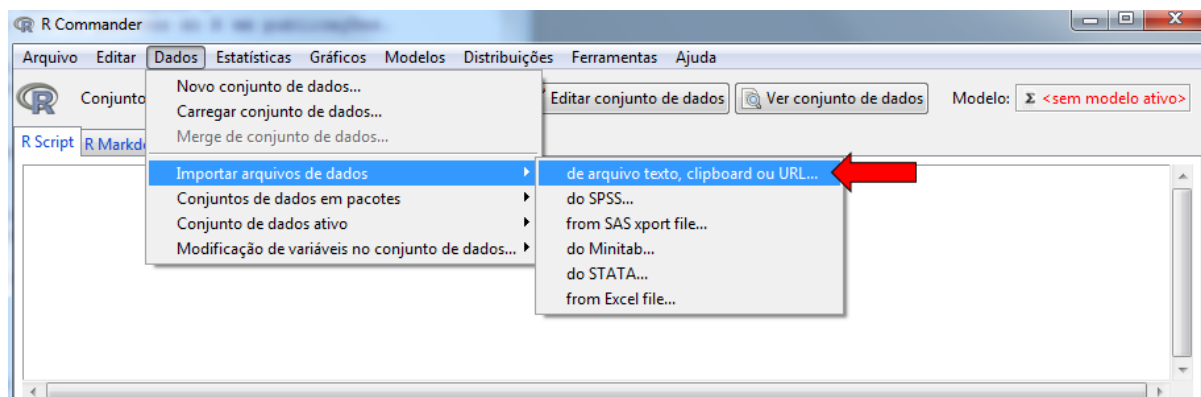
ftv number of physician visits during the first trimester.

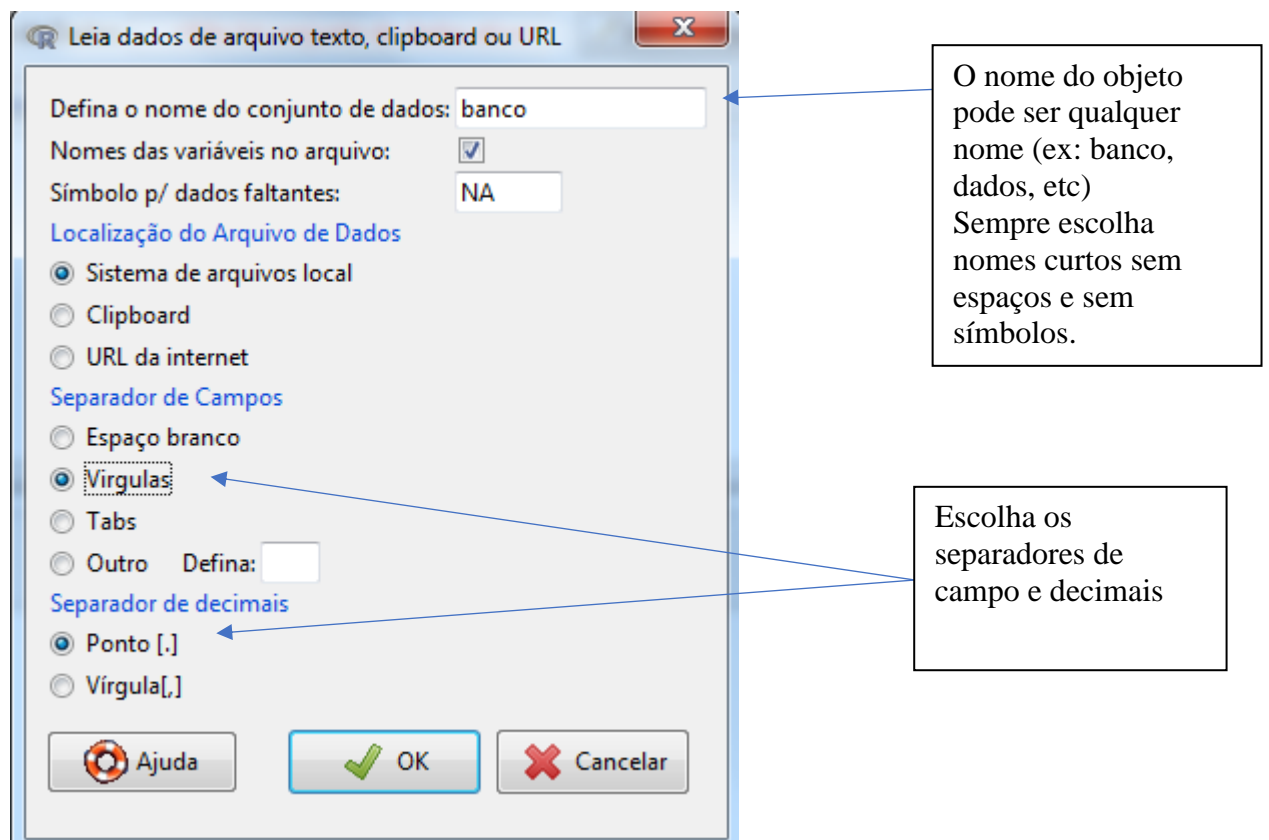
bwt birth weight in grams.

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

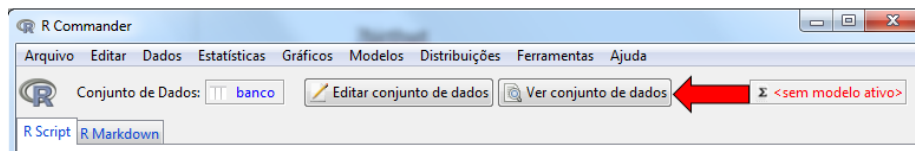
Mensagens
com a versão SDI da interface; veja ?Commander
[3] NOTA: Os dados banco tem 189 linhas e 11 colunas.
```

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 with 30 rows and 11 columns. The columns are labeled: X, low, age, lwt, race, smoke, ptl, ht, ui, ftv, and bwt. The data is as follows:

|    | 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 |
| 11 | 96  | 0   | 19  | 95  | 3    | 0     | 0   | 0  | 0  | 0   | 2722 |
| 12 | 97  | 0   | 19  | 150 | 3    | 0     | 0   | 0  | 0  | 1   | 2733 |
| 13 | 98  | 0   | 22  | 95  | 3    | 0     | 0   | 1  | 0  | 0   | 2751 |
| 14 | 99  | 0   | 30  | 107 | 3    | 0     | 1   | 0  | 1  | 2   | 2750 |
| 15 | 100 | 0   | 18  | 100 | 1    | 1     | 0   | 0  | 0  | 0   | 2769 |
| 16 | 101 | 0   | 18  | 100 | 1    | 1     | 0   | 0  | 0  | 0   | 2769 |
| 17 | 102 | 0   | 15  | 98  | 2    | 0     | 0   | 0  | 0  | 0   | 2778 |
| 18 | 103 | 0   | 25  | 118 | 1    | 1     | 0   | 0  | 0  | 3   | 2782 |
| 19 | 104 | 0   | 20  | 120 | 3    | 0     | 0   | 0  | 1  | 0   | 2807 |
| 20 | 105 | 0   | 28  | 120 | 1    | 1     | 0   | 0  | 0  | 1   | 2821 |
| 21 | 106 | 0   | 32  | 121 | 3    | 0     | 0   | 0  | 0  | 2   | 2835 |
| 22 | 107 | 0   | 31  | 100 | 1    | 0     | 0   | 0  | 1  | 3   | 2835 |
| 23 | 108 | 0   | 36  | 202 | 1    | 0     | 0   | 0  | 0  | 1   | 2836 |
| 24 | 109 | 0   | 28  | 120 | 3    | 0     | 0   | 0  | 0  | 0   | 2863 |
| 25 | 111 | 0   | 25  | 120 | 3    | 0     | 0   | 0  | 1  | 2   | 2877 |
| 26 | 112 | 0   | 28  | 167 | 1    | 0     | 0   | 0  | 0  | 0   | 2877 |
| 27 | 113 | 0   | 17  | 122 | 1    | 1     | 0   | 0  | 0  | 0   | 2906 |
| 28 | 114 | 0   | 29  | 150 | 1    | 0     | 0   | 0  | 0  | 2   | 2920 |
| 29 | 115 | 0   | 26  | 168 | 2    | 1     | 0   | 0  | 0  | 0   | 2920 |
| 30 | 116 | 0   | 17  | 113 | 2    | 0     | 0   | 0  | 0  | 1   | 2920 |

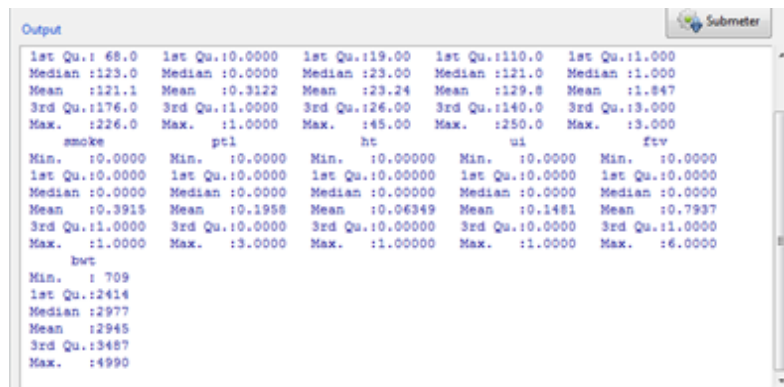
Observe a estrutura do banco de dados. Digite `str(banco)`

The image shows the R Output window with the command `str(banco)` executed. The output shows the structure of the data frame 'banco', which has 189 observations and 11 variables. The variables 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 is as follows:

```
> str(banco)
'data.frame': 189 obs. of  11 variables:
 $ X   : int  85 86 87 88 89 91 92 93 94 95 ...
 $ low : int  0 0 0 0 0 0 0 0 0 0 ...
 $ age : int  19 33 20 21 18 21 22 17 29 26 ...
 $ lwt : int  182 155 105 108 107 124 118 103 123 113 ...
 $ race: int   2 3 1 1 1 3 1 3 1 1 ...
 $ smoke: int  0 0 1 1 1 0 0 0 1 1 ...
 $ ptl  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ ht   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ ui   : int  1 0 0 1 1 0 0 0 0 0 ...
 $ ftv  : int  0 3 1 2 0 0 1 1 1 0 ...
 $ bwt  : int  2523 2551 2557 2594 2600 2622 2637 2637 2663 2665 ...
```

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

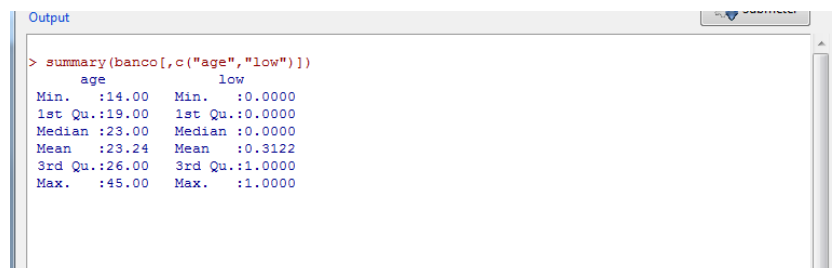
`summary(banco)`



| 1st Qu.: | 68.0   | 1st Qu.:0.0000 | 1st Qu.:119.00  | 1st Qu.:1210.0 | 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.:1140.0 | 3rd Qu.:3.000  |
| Max. :   | 226.0  | Max. :1.0000   | Max. :45.00     | Max. :250.0    | Max. :3.000    |
| smoke    |        |                |                 |                |                |
| 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.: | 3457   |                |                 |                |                |
| Max. :   | 4990   |                |                 |                |                |

Observe a estrutura das variáveis smoke e low.

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



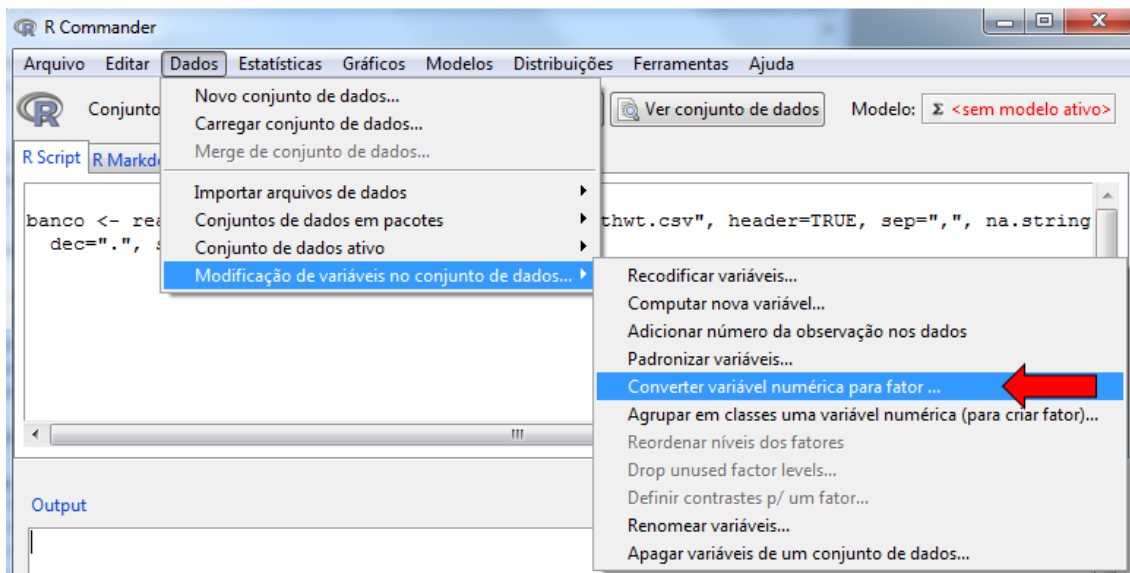
| age      |       | low      |        |
|----------|-------|----------|--------|
| Min. :   | 14.00 | Min. :   | 0.0000 |
| 1st Qu.: | 19.00 | 1st Qu.: | 0.0000 |
| Median : | 23.00 | Median : | 0.0000 |
| Mean :   | 23.24 | Mean :   | 0.3122 |
| 3rd Qu.: | 26.00 | 3rd Qu.: | 1.0000 |
| Max. :   | 45.00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

## Conversão de Dados

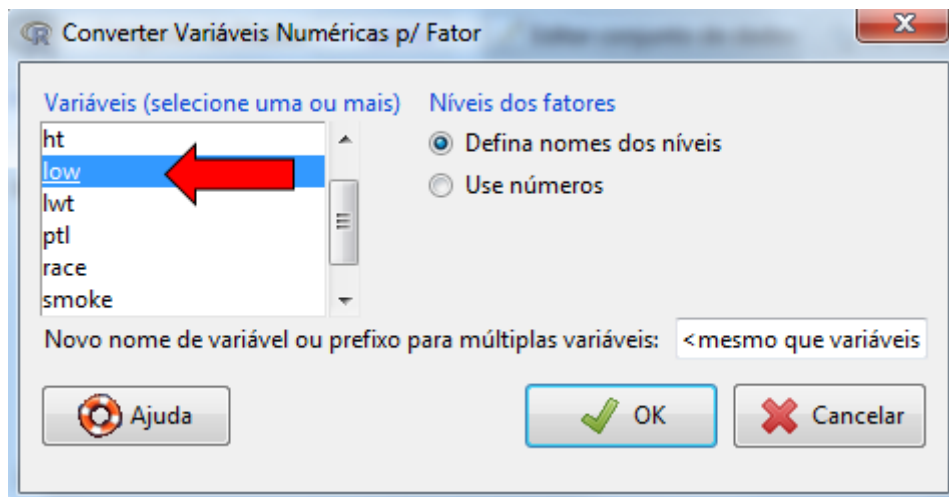
Muitas vezes é necessário converter variáveis numéricas para fator (variável qualitativa). Ao executar os comandos abaixo são realizadas as conversões dos dados das variáveis “low,race,smoke,ht e ui”.

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

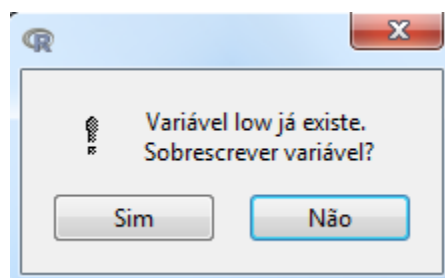
“Dados-> Modificação de Variáveis no conjunto de dados -> Converter variáveis numéricas par fator”



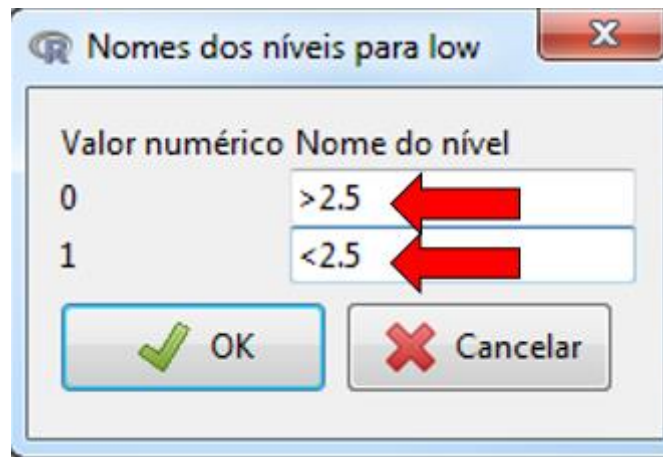
Vamos realizar a conversão para a variável “low” por exemplo. Na caixa de comando seleciona a variável.



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"))
```

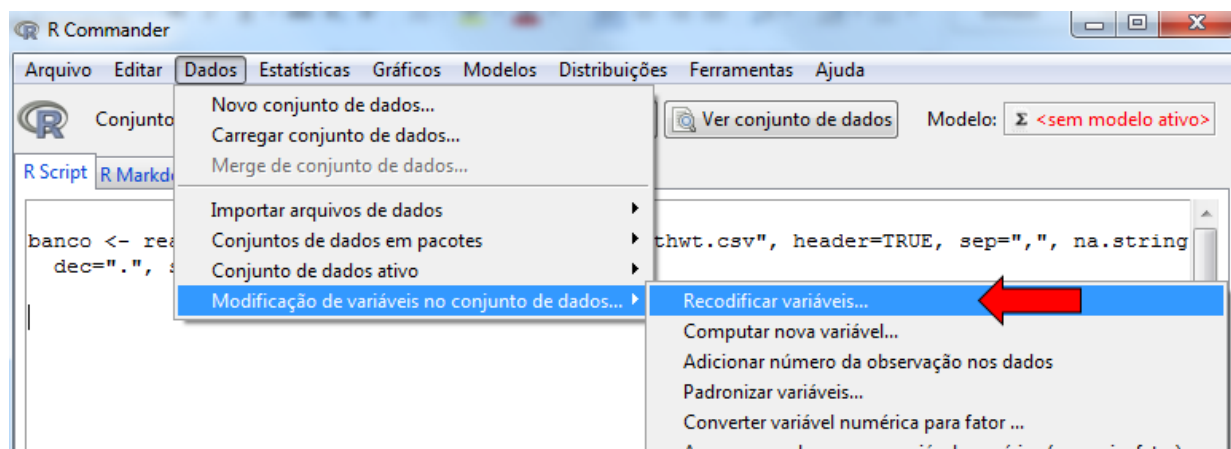
## Recodificação de Variáveis

Suponha que verificamos que a informação sobre a mulher que teve 3 partos prematuros está equivocada e desejamos recodificá-la como caso ignorado. O comando a ser executado é o seguinte referente à variável “ptl”

```
banco$ptl[banco$ptl==3]<-NA
```

No R Commander pode ser feito da seguinte forma:

“Dados-> Modificação de Variáveis no conjunto de dados -> Recodificar variáveis..”

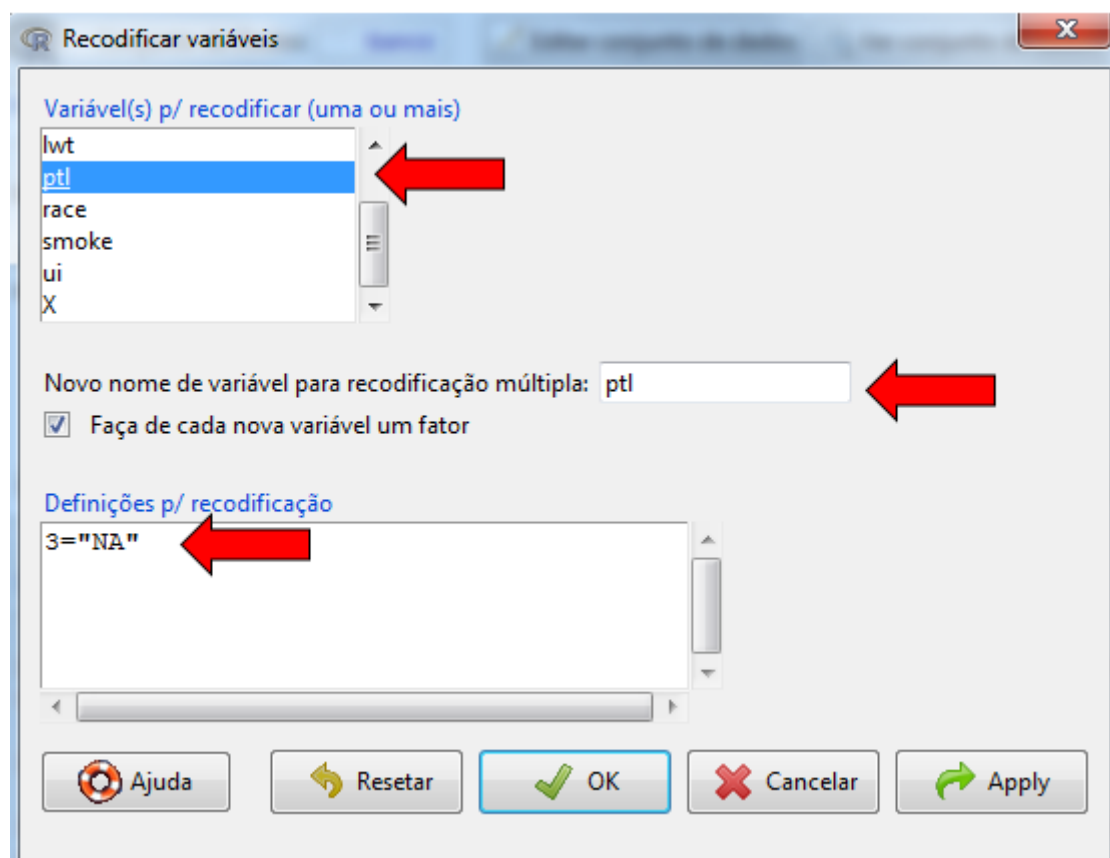


E no campo “Definições p/ recodificação” indicar o que deve ser mudado.

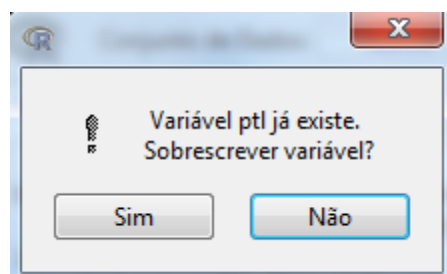
No caso do nosso exemplo **3=NA** faça:

- Na caixa de seleção da variável escolha a variável “ptl”
- No campo “Novo nome de variável para recodificação múltipla” escreva também o nome da variável. Neste caso “ptl” novamente.
- Na caixa “Definições p/ recodificação” escreva a forma que deseja recodificar os valores da variável.





Clique em OK e aparecerá uma caixa de diálogo pergunta se deseja sobrescrever os valores da variável. Clique em sim e então será feita a recodificação.



Obs: Recodificação de variáveis no R Commander funciona apenas para variável numérica. Quando a recodificação for de uma variável factor, só poderá ser feito pelo script.

## Agrupando Variável

Muitas vezes é interessante agrupar variáveis. Uma estratégia é a criação de uma nova variável baseada no agrupamento dos valores de uma outra variável, a fim de manter todas as informações iniciais do banco.

Por exemplo, vamos agrupar os valores “Pretos” e “Outros” encontrados na variável “race” criando uma nova variável chamada “cor” que será o resultado desses dois valores.

Antes de agrupar vamos checar a contagem dos valores da variável “race” pelo comando “table”

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

Output

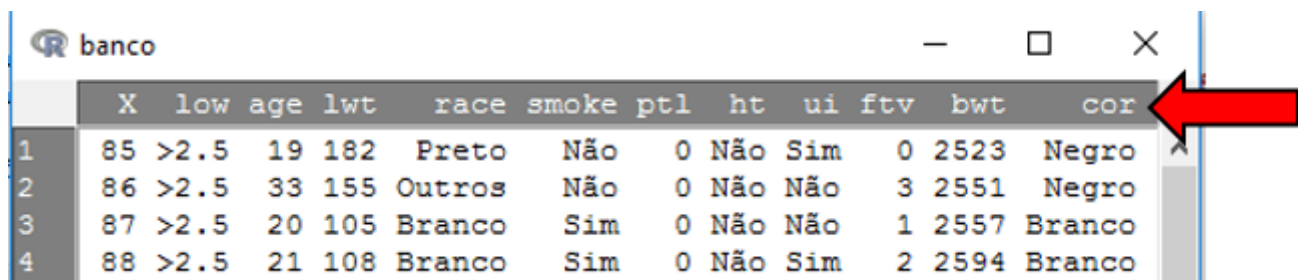
```
> table(banco$race)

Branco  Preto  Outros
    96     26     67
```

O comando que cria a nova variável:

```
banco$cor[banco$race=="Branco"]=1
banco$cor[banco$race=="Preto"|banco$race=="Outros"]=2
banco$cor<-factor(banco$cor,lab=c("Branco","Negro"))
```

Note que após realizar o comando foi criada a nova variável “cor” e os valores da variável “race” agrupados em “Branco” e “Negro”



|   | X  | low  | age | lwt | race   | smoke | ptl | ht  | ui  | ftv | bwt  | cor    |
|---|----|------|-----|-----|--------|-------|-----|-----|-----|-----|------|--------|
| 1 | 85 | >2.5 | 19  | 182 | Preto  | Não   | 0   | Não | Sim | 0   | 2523 | Negro  |
| 2 | 86 | >2.5 | 33  | 155 | Outros | Não   | 0   | Não | Não | 3   | 2551 | Negro  |
| 3 | 87 | >2.5 | 20  | 105 | Branco | Sim   | 0   | Não | Não | 1   | 2557 | Branco |
| 4 | 88 | >2.5 | 21  | 108 | Branco | Sim   | 0   | Não | Sim | 2   | 2594 | Branco |

Sempre que agrupar e criar novas variáveis execute o comando “table” para verificar a contagem.

```
table(banco$cor)
```

## Output

```
> table(banco$cor)
```

```
Branco  Negro  
    96    93
```

Agrupando a variável “age” (idade) em faixa etária.

Suponha que desejamos agrupar a variável idade em faixa etária para realizar um outro tipo de análise. Sendo assim, vamos criar uma nova variável chamada “faixaetaria” e utilizar os valores da variável “age”.

O comando a ser executado:


```
banco$faixaetaria[banco$age<20]=1
```

```
banco$faixaetaria[banco$age>=20&banco$age<30]=2
```

```
banco$faixaetaria[banco$age>=30&banco$age<40]=3
```

```
banco$faixaetaria[banco$age>=40]=4
```

```
banco$faixaetaria<-factor(banco$faixaetaria,lab=c(" Menor de 20 ", "21 a 29 ", "31 a 39", "40 ou mais"))
```



| id | bwt  | cor    | faixaetaria |
|----|------|--------|-------------|
| 0  | 2523 | Negro  | Menor de 20 |
| 3  | 2551 | Negro  | 31 a 39     |
| 1  | 2557 | Branco | 21 a 29     |
| 2  | 2594 | Branco | 21 a 29     |
| 0  | 2600 | Branco | Menor de 20 |
| 0  | 2622 | Negro  | 21 a 29     |
| 1  | 2637 | Branco | 21 a 29     |

Execute também a função “table” para conferir a contagem da nova variável “faixaetaria”

```
table(banco$faixaetaria)
```

```
> table(banco$faixaetaria)
```

```
Menor de 20    21 a 29    31 a 39    40 ou mais  
      51         111         26         1  
> |
```

```
table(banco$age,banco$faixaetaria)
```


Suponha que estamos interessados em criar uma variável que agrupe a informação do peso da criança (baixo peso e peso normal) em função do status da mãe (ser fumante ou não):

Mãe fumante e Criança com peso normal  
 Mãe fumante e Criança com baixo peso  
 Mãe não fumante e Criança com peso normal  
 Mãe não fumante e Criança com baixo peso

```
banco$relação.fumo.peso[banco$smoke=="Sim" & banco$low=="<2.5"]<-1 #fuma e baixo peso
banco$relação.fumo.peso[banco$smoke=="Sim" & banco$low==">2.5"]<-2 #fuma e peso normal
banco$relação.fumo.peso[banco$smoke=="Não" & banco$low=="<2.5"]<-3 # não fuma e baixo peso
banco$relação.fumo.peso[banco$smoke=="Não" & banco$low==">2.5"]<-4 # não fuma e peso normal
```

```
banco$relação.fumo.peso<-factor(banco$relação.fumo.peso,lab=c("Fuma e baixo peso","Fuma e peso
normal"," Não fuma e baixo peso","Não fuma e peso normal"))
```

Novamente será criado uma nova variável, porém agora utilizado as informações das variáveis “smoke” e “low” observamos seus valores simultaneamente.



| bwt  | cor    | faixaetaria | relação.fumo.peso      |
|------|--------|-------------|------------------------|
| 2523 | Negro  | Menor de 20 | Não fuma e peso normal |
| 2551 | Negro  | 31 a 39     | Não fuma e peso normal |
| 2557 | Branco | 21 a 29     | Fuma e peso normal     |
| 2594 | Branco | 21 a 29     | Fuma e peso normal     |
| 2600 | Branco | Menor de 20 | Fuma e peso normal     |
| 2622 | Negro  | 21 a 29     | Não fuma e peso normal |
| 2637 | Branco | 21 a 29     | Não fuma e peso normal |
| 2637 | Negro  | Menor de 20 | Não fuma e peso normal |
| 2663 | Branco | 21 a 29     | Fuma e peso normal     |
| 2665 | Branco | 21 a 29     | Fuma e peso normal     |
| 2722 | Negro  | Menor de 20 | Não fuma e peso normal |

Sempre que uma variável for agrupada execute o comando “table”

```
table(banco$relação.fumo.peso)
```

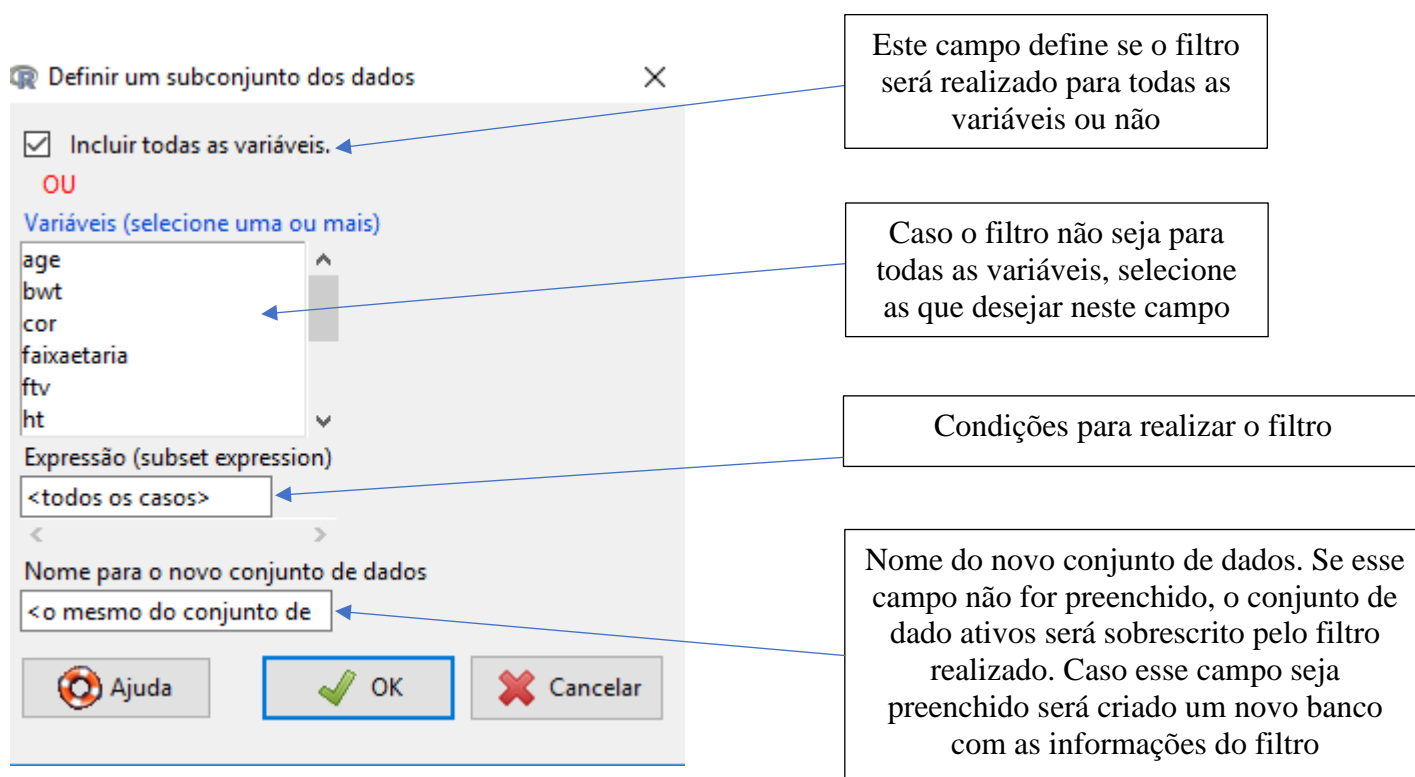
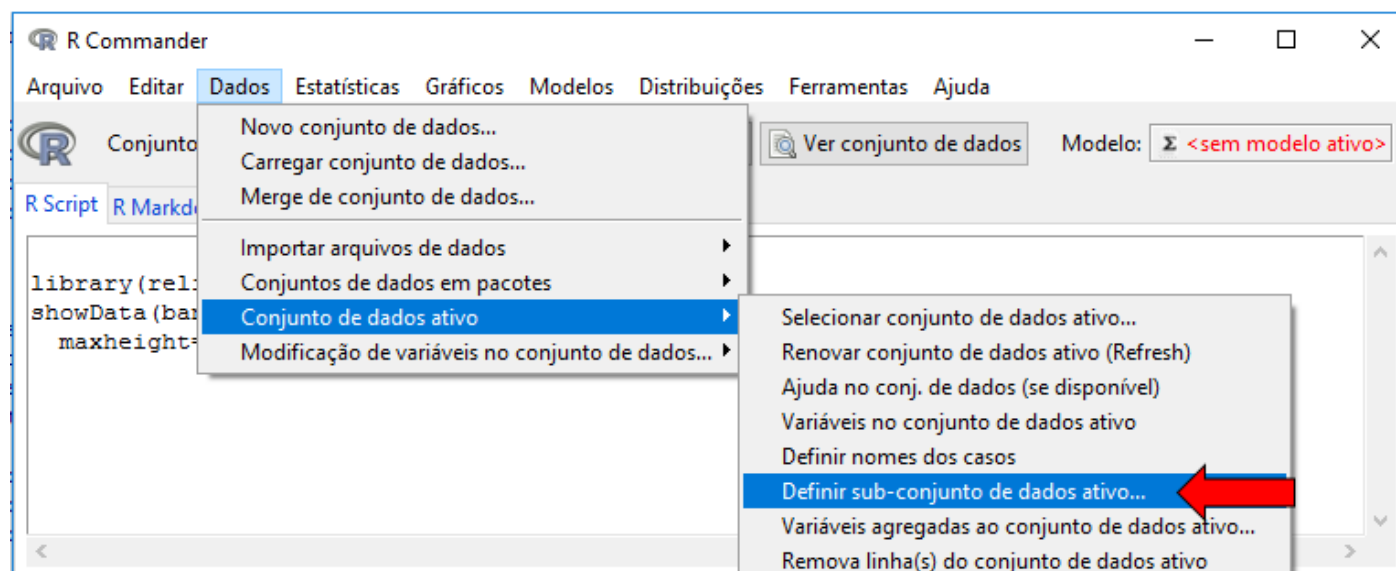
```
> table(banco$relação.fumo.peso)

      Fuma e baixo peso      Fuma e peso normal      Não fuma e baixo peso
                30                44                29
Não fuma e peso normal
                86
> |
```

## Filtros

Para criar um subconjunto do banco ativo, pode-se realizar um filtro desse banco. Para realizar o filtro, deve-se escolher se quer reescrever o banco de dados (substituir o banco ativo pelo filtro) ou criar um “banco novo” apenas com as informações filtradas.

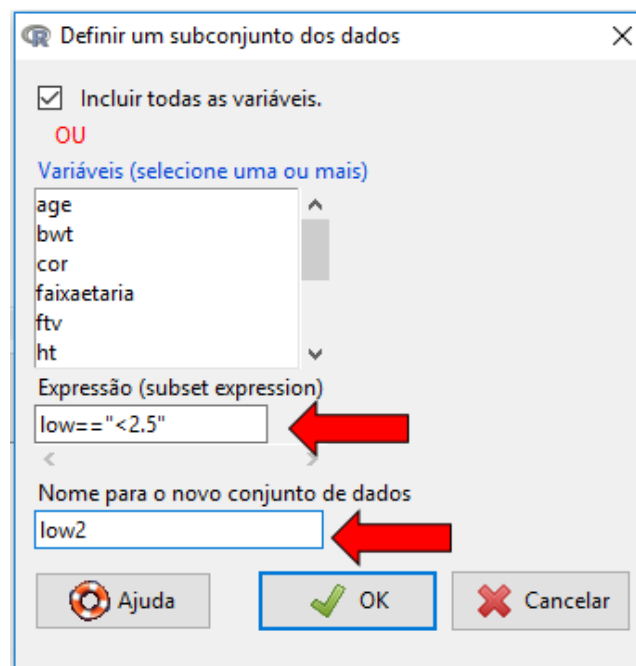
Dados > Conjunto de dados ativos > Definir sub-conjunto de dados ativos



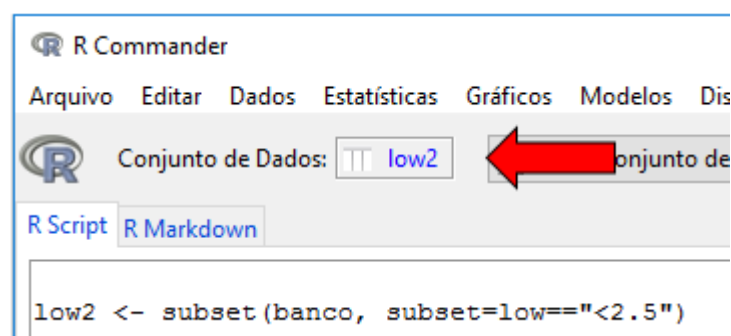
Exemplo:

Criando um filtro com os dados das crianças que nasceram com baixo peso, ou seja, as crianças com peso abaixo de 2.5. Vamos criar também um novo banco chamado “low2”

Dados > Conjunto de dados ativos > Definir sub-conjunto de dados ativos



Após clicar em “ok” será criado um novo banco e que resultado do filtro:



|     | X  | low  | age | lwt | race   | smoke | ptl | ht  | ui  | ftv | bwt  | cor    | faixaetaria | rel     |
|-----|----|------|-----|-----|--------|-------|-----|-----|-----|-----|------|--------|-------------|---------|
| 131 | 4  | <2.5 | 28  | 120 | Outros | Sim   | 1   | Não | Sim | 0   | 709  | Negro  | 21 a 29     | Fum     |
| 132 | 10 | <2.5 | 29  | 130 | Branco | Não   | 0   | Não | Sim | 2   | 1021 | Branco | 21 a 29     | Não fum |
| 133 | 11 | <2.5 | 34  | 187 | Preto  | Sim   | 0   | Sim | Não | 0   | 1135 | Negro  | 31 a 39     | Fum     |
| 134 | 13 | <2.5 | 25  | 105 | Outros | Não   | 1   | Sim | Não | 0   | 1330 | Negro  | 21 a 29     | Não fum |
| 135 | 15 | <2.5 | 25  | 85  | Outros | Não   | 0   | Não | Sim | 0   | 1474 | Negro  | 21 a 29     | Não fum |
| 136 | 16 | <2.5 | 27  | 150 | Outros | Não   | 0   | Não | Não | 0   | 1588 | Negro  | 21 a 29     | Não fum |
| 137 | 17 | <2.5 | 23  | 97  | Outros | Não   | 0   | Não | Sim | 1   | 1588 | Negro  | 21 a 29     | Não fum |
| 138 | 18 | <2.5 | 24  | 128 | Preto  | Não   | 1   | Não | Não | 1   | 1701 | Negro  | 21 a 29     | Não fum |
| 139 | 19 | <2.5 | 24  | 132 | Outros | Não   | 0   | Sim | Não | 0   | 1729 | Negro  | 21 a 29     | Não fum |
| 140 | 20 | <2.5 | 21  | 165 | Branco | Sim   | 0   | Sim | Não | 1   | 1790 | Branco | 21 a 29     | Fum     |
| 141 | 22 | <2.5 | 32  | 105 | Branco | Sim   | 0   | Não | Não | 0   | 1818 | Branco | 31 a 39     | Fum     |
| 142 | 23 | <2.5 | 19  | 91  | Branco | Sim   | 2   | Não | Sim | 0   | 1885 | Branco | Menor de 20 | Fum     |
| 143 | 24 | <2.5 | 25  | 115 | Outros | Não   | 0   | Não | Não | 0   | 1893 | Negro  | 21 a 29     | Não fum |
| 144 | 25 | <2.5 | 16  | 130 | Outros | Não   | 0   | Não | Não | 1   | 1899 | Negro  | Menor de 20 | Não fum |
| 145 | 26 | <2.5 | 25  | 92  | Branco | Sim   | 0   | Não | Não | 0   | 1928 | Branco | 21 a 29     | Fum     |
| 146 | 27 | <2.5 | 20  | 150 | Branco | Sim   | 0   | Não | Não | 2   | 1928 | Branco | 21 a 29     | Fum     |
| 147 | 28 | <2.5 | 21  | 200 | Preto  | Não   | 0   | Não | Sim | 2   | 1928 | Negro  | 21 a 29     | Não fum |
| 148 | 29 | <2.5 | 24  | 155 | Branco | Sim   | 1   | Não | Não | 0   | 1936 | Branco | 21 a 29     | Fum     |
| 149 | 30 | <2.5 | 21  | 103 | Outros | Não   | 0   | Não | Não | 0   | 1970 | Negro  | 21 a 29     | Não fum |
| 150 | 31 | <2.5 | 20  | 125 | Outros | Não   | 0   | Não | Sim | 0   | 2055 | Negro  | 21 a 29     | Não fum |
| 151 | 32 | <2.5 | 25  | 89  | Outros | Não   | 2   | Não | Não | 1   | 2055 | Negro  | 21 a 29     | Não fum |
| 152 | 33 | <2.5 | 19  | 102 | Branco | Não   | 0   | Não | Não | 2   | 2082 | Branco | Menor de 20 | Não fum |
| 153 | 34 | <2.5 | 19  | 112 | Branco | Sim   | 0   | Não | Sim | 0   | 2084 | Branco | Menor de 20 | Fum     |
| 154 | 35 | <2.5 | 26  | 117 | Branco | Sim   | 1   | Não | Não | 0   | 2084 | Branco | 21 a 29     | Fum     |
| 155 | 36 | <2.5 | 24  | 138 | Branco | Não   | 0   | Não | Não | 0   | 2100 | Branco | 21 a 29     | Não fum |
| 156 | 37 | <2.5 | 17  | 130 | Outros | Sim   | 1   | Não | Sim | 0   | 2125 | Negro  | Menor de 20 | Fum     |
| 157 | 40 | <2.5 | 20  | 120 | Preto  | Sim   | 0   | Não | Não | 3   | 2126 | Negro  | 21 a 29     | Fum     |
| 158 | 42 | <2.5 | 22  | 130 | Branco | Sim   | 1   | Não | Sim | 1   | 2187 | Branco | 21 a 29     | Fum     |
| 159 | 43 | <2.5 | 27  | 130 | Preto  | Não   | 0   | Não | Sim | 0   | 2187 | Negro  | 21 a 29     | Não fum |
| 160 | 44 | <2.5 | 20  | 80  | Outros | Sim   | 0   | Não | Sim | 0   | 2211 | Negro  | 21 a 29     | Fum     |

Os filtros podem ser realizados com duas ou mais condições

`banco$low=="<2.5"&banco$smoke=="Sim"`

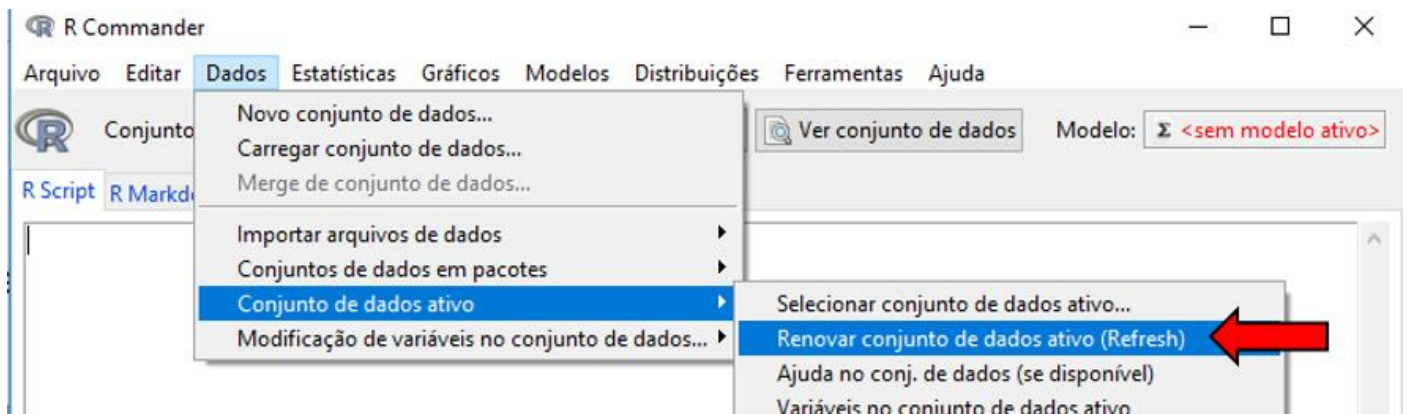
ou

`low=="<2.5"&smoke=="Sim"`

Importante!! Após qualquer mudança significativa e/ou alterações feita ao banco de dados através dos comandos no script do R, temos que renovar o conjunto de dados dando um “Refresh”:

Dados> Conjunto de dados ativo > Renovar conjunto de dados ativo (Refresh)

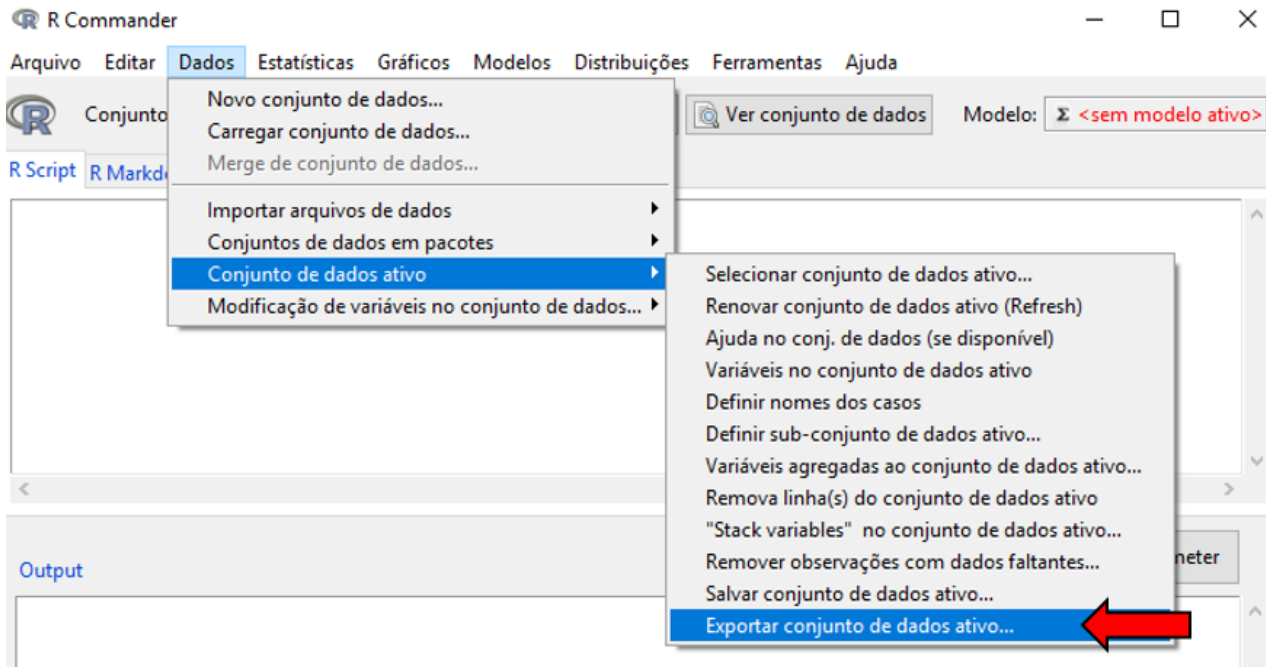




Isso evita futuros erros ao realizar a análise do banco!

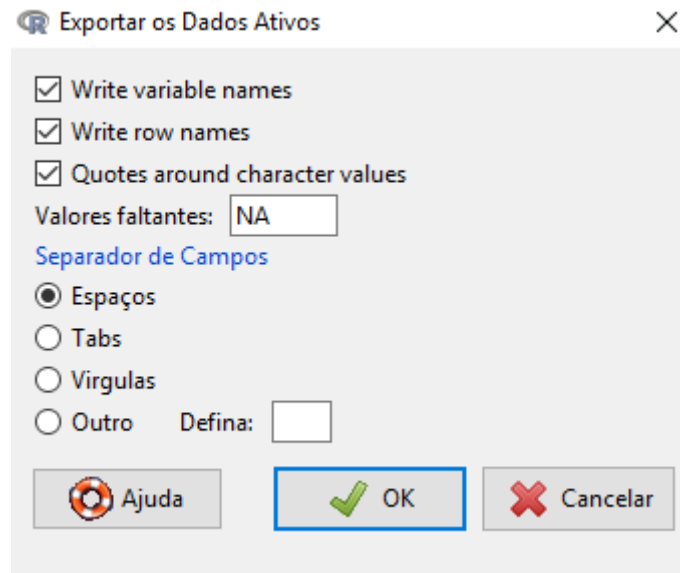
Também é possível exportar e salvar as alterações realizadas para um arquivo externo tipo csv.

“Dados> Conjunto de dados ativo > Exportar conjunto de dados ativo...”



Escolha o tipo de separador de campo e clique em ok.

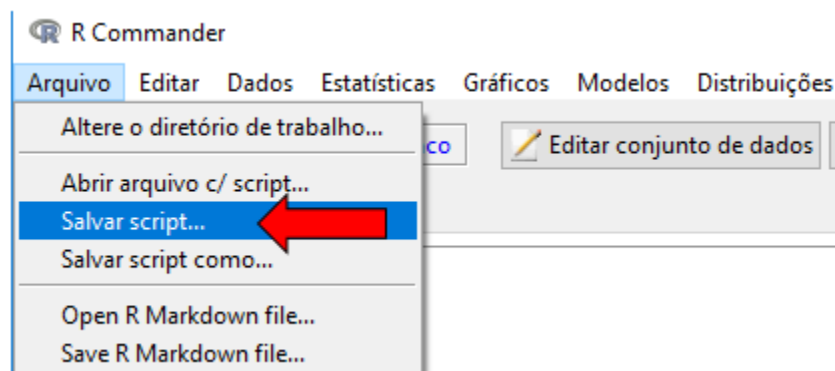




## Salvando a área de trabalho

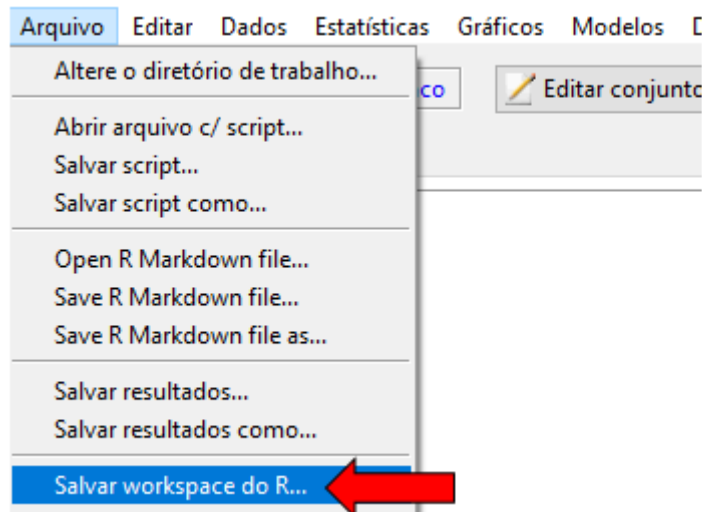
Após toda manipulação do banco, o script e área de trabalho podem ser salvos para guardar todas as funções e objetos criados, respectivamente.

Para salvar o script: “Arquivo > Salvar script”



Escolha o local onde o arquivo será salvo e clique em OK

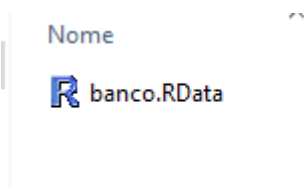
Para salvar a área de trabalho: “Arquivo > Salvar workspace do R”



Escolha o local onde o arquivo será salvo, dê um nome para o arquivo e clique em OK. Será salvo um arquivo tipo .RData.

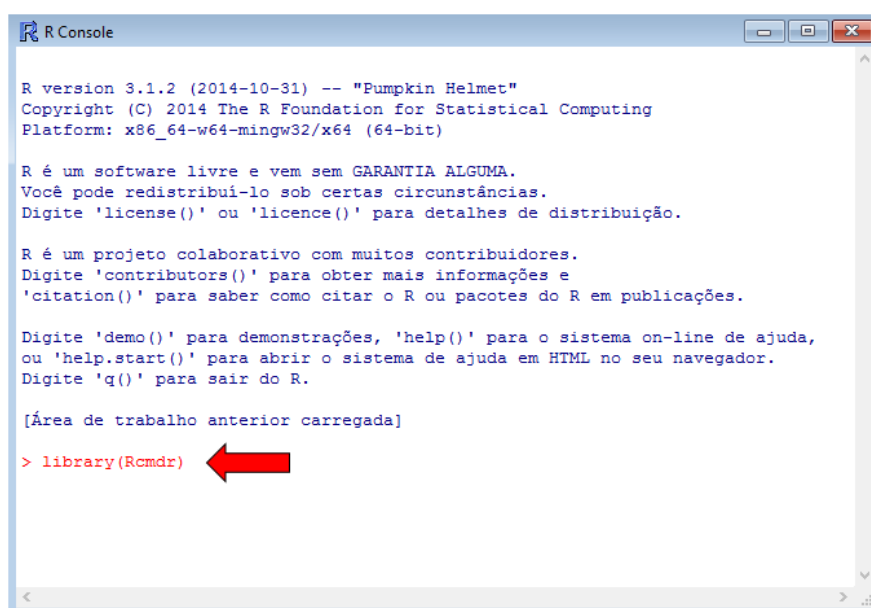
Em uma próxima sessão, podem ser recuperados o script e a área de trabalho salvos.

Para reabrir a área de trabalho, vá no R em “Arquivo-Carregar área de trabalho”, escolha o local onde o arquivo foi salvo dê um duplo click no arquivo salvo.

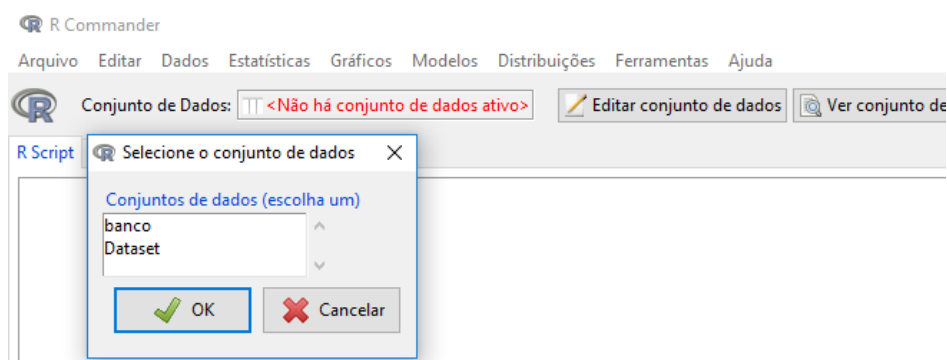


Aparecerá o comando no R: **load("C:\\Users\\xxx\\banco.RData")**

OBS: Após abrir carregar a área de trabalho deve-se carregar novamente o R Commander.



Ao carregar o R Commander, note que todos os bancos que estavam criados na sessão anterior já estão carregados, com todas as modificações (agregações e etc) feitas antes de salvar.



|   | t | race   | smoke | ptl | ht  | ui  | ftv | bwt  | cor    | faixaetaria | relação.fumo.peso      |
|---|---|--------|-------|-----|-----|-----|-----|------|--------|-------------|------------------------|
| 1 | 2 | Preto  | Não   | 0   | Não | Sim | 0   | 2523 | Negro  | Menor de 20 | Não fuma e peso normal |
| 2 | 5 | Outros | Não   | 0   | Não | Não | 3   | 2551 | Negro  | 31 a 39     | Não fuma e peso normal |
| 3 | 5 | Branco | Sim   | 0   | Não | Não | 1   | 2557 | Branco | 21 a 29     | Fuma e peso normal     |
| 4 | 8 | Branco | Sim   | 0   | Não | Sim | 2   | 2594 | Branco | 21 a 29     | Fuma e peso normal     |