

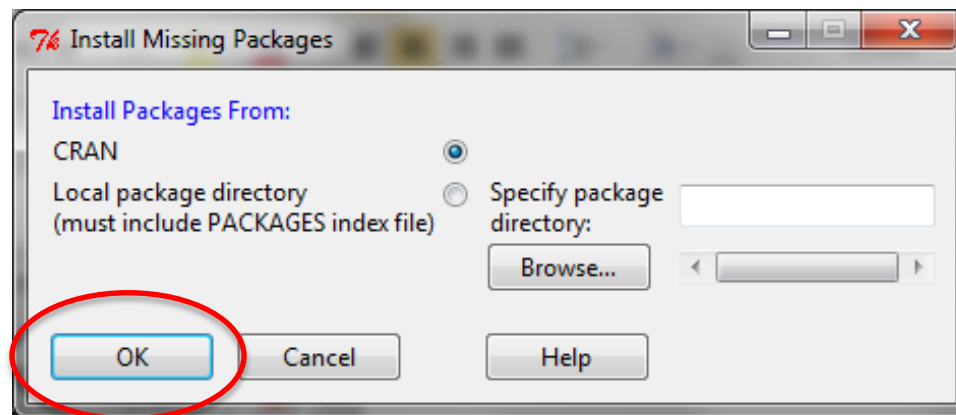
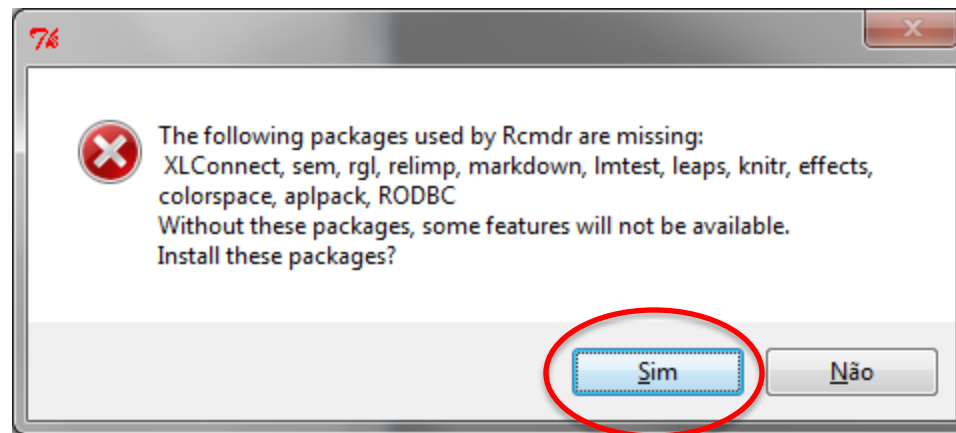
Introdução ao R Commander

Marcelo Lauretto

R Commander

- Interface gráfica para análises estatísticas básicas
 - Comandos R são gerados e executados via menus
- Pacote Rcmdr
 - Instalação: dentro do ambiente R:
 - `install.packages('Rcmdr')`
 - Chamada:
 - `library(Rcmdr)`
 - `Commander()`

- 1a chamada do Rcmdr:
 - Instalar pacotes adicionais



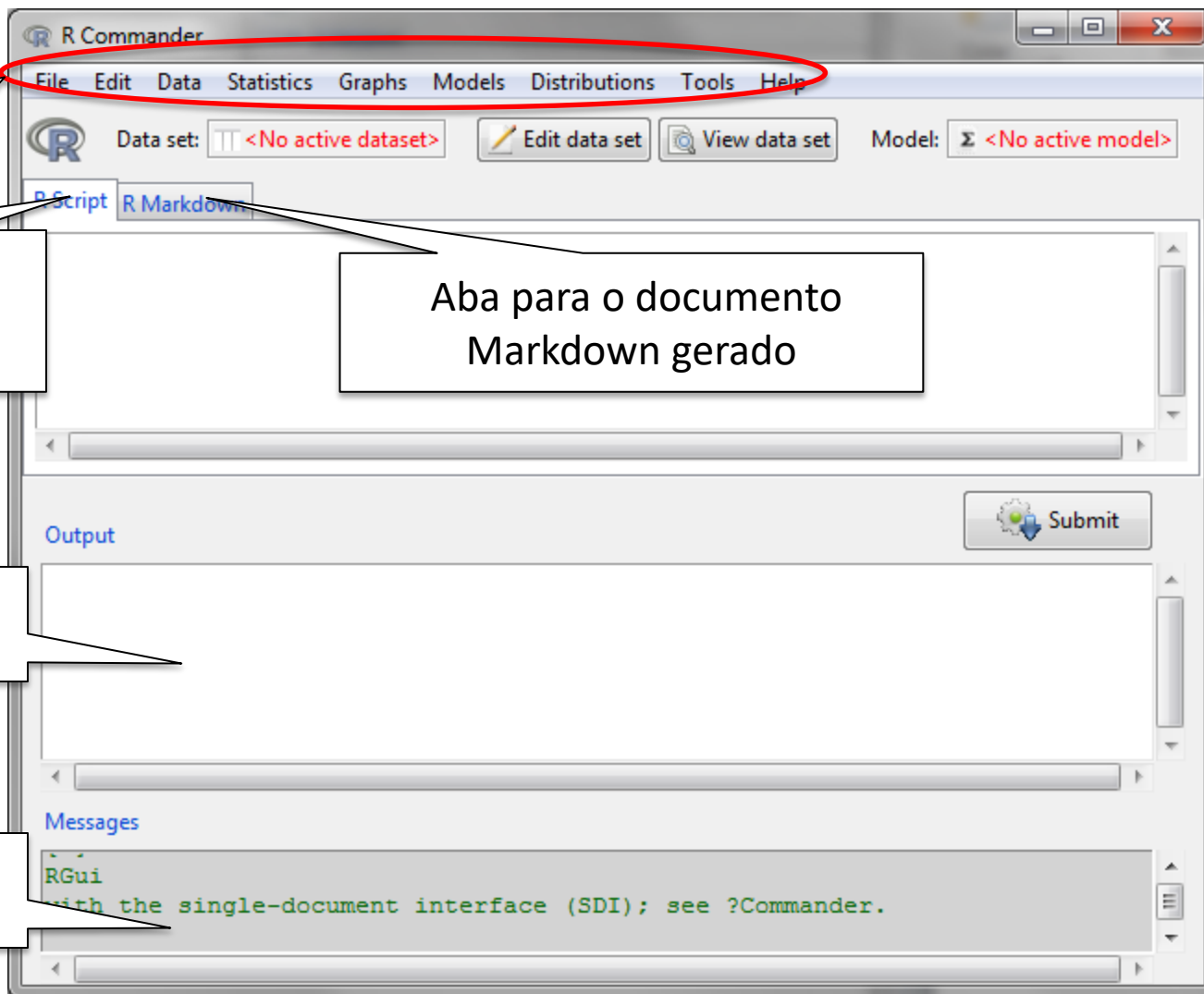
- Janela do R Commander:

Menu principal

Aba para o script de comandos R gerados

Resultados impressos

Mensagens de erro, avisos e outras notas



Aba para o documento Markdown gerado

Menu principal

- **File:** Opções para abrir e salvar diversos tipos de arquivos, e para alterar o diretório de trabalho do R – diretório no qual o R irá procurar ou gravar arquivos por default
- **Edit:** Opções para edição de texto, tais como Copy e Paste, bem como opções específicas para documentos *R Markdown* (discutidos adiante)
- **Data:** Submenus para importação, exportação e manipulação de dados
- **Statistics:** Submenus para vários tipos de análise estatística de dados (análise exploratória e inferência), incluindo ajustes de modelos estatísticos aos dados

- **Graphs:** Opções e submenus para a criação de gráficos estatísticos típicos
- **Models:** Submenus para realização de diversas operações usando modelos estatísticos previamente ajustados aos dados
- **Distributions:** Submenus para cálculo de probabilidades, gráficos e geração de variáveis para distribuições típicas
- **Tools:** Opções para carregar pacotes do R e suplementos (plug-ins) do R Commander
- **Help:** Ajuda e manuais básicos do R Commander, websites úteis, ajuda sobre data sets

Breve exemplo

Dataset Utilizado

- GSS.csv (livro R Commander)
<http://socserv.mcmaster.ca/jfox/Books/RCommander/>
- Pesquisa transversal periódica da população americana conduzida pelo National Opinion Research Center da Universidade de Chicago.
- Questão principal:
 - “... If a man and a woman have sex relations before marriage, do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?”

Leitura de dados

- Arquivo GSS.csv (U.S. General Social Survey)

Linha de cabeçalho (nomes das variáveis)

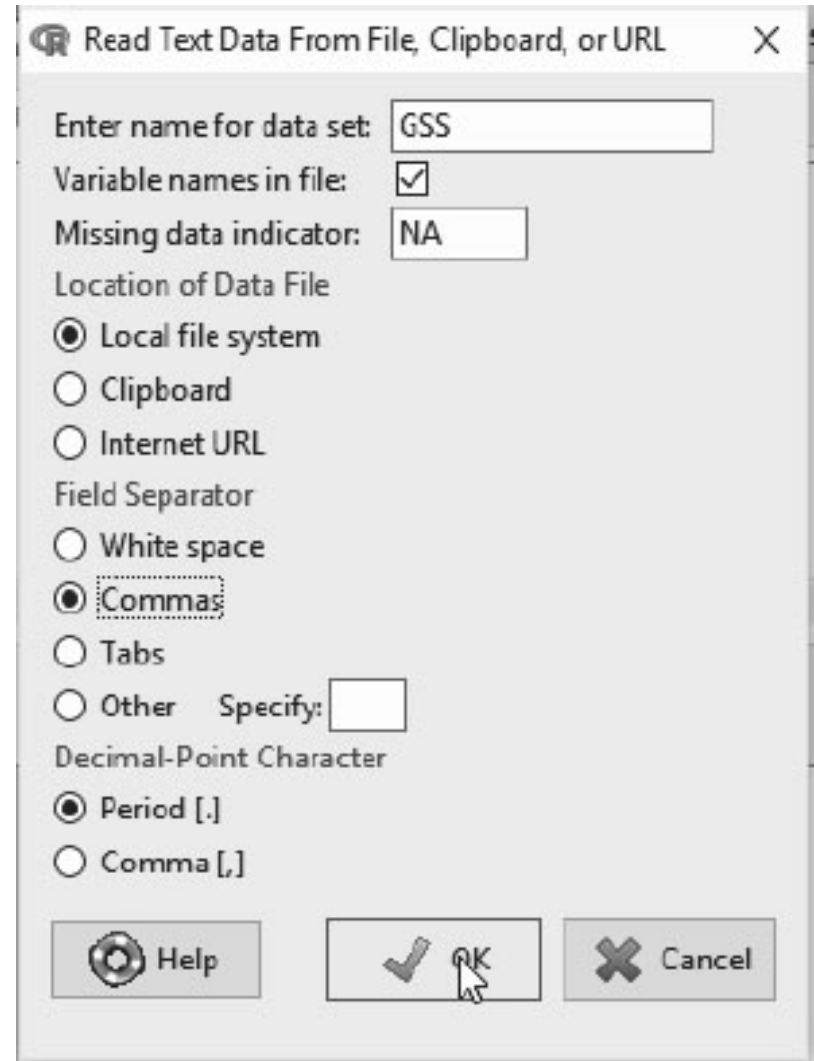
```
year,gender,premarital.sex,education,religion
1972,female,not wrong at all,post-secondary,Jewish
1972,male,always wrong,less than high school,Catholic
1972,female,always wrong,high school,Protestant
1972,female,always wrong,post-secondary,other
1972,female,sometimes wrong,high school,Protestant
. . .
2012,female,not wrong at all,post-secondary,none
2012,female,not wrong at all,high school,Catholic
2012,female,sometimes wrong,high school,Catholic
```

Registros (um por linha)

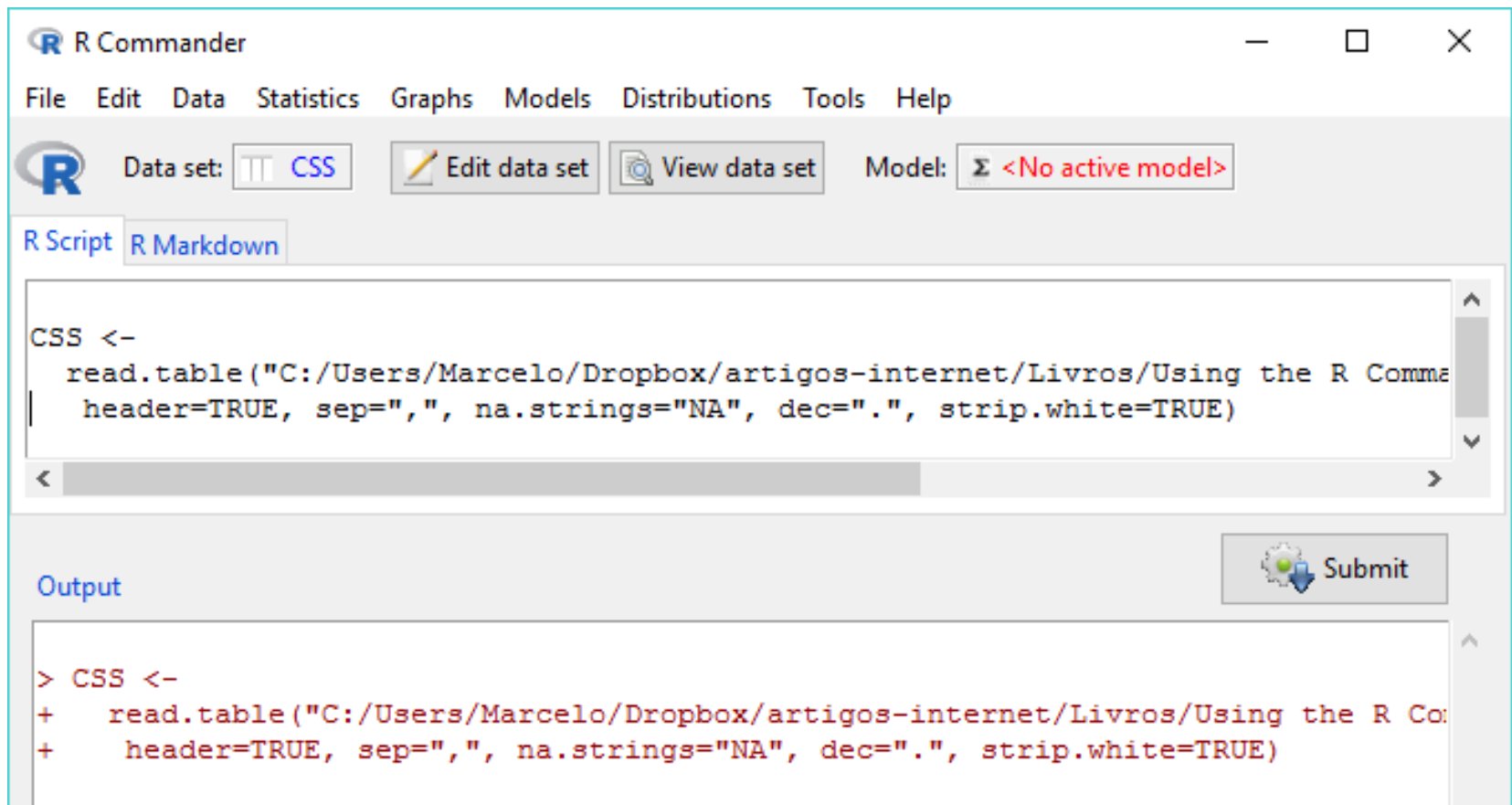
Separadores de campos (vírgulas)

- *Data > Import data > from text file*

- Campos importantes:
 - Nome do data set
 - Arquivo contém nomes de variáveis?
 - Representação dos dados faltantes
 - Separador de campos
 - Separador de casas decimais



- Cada operação realizada no R Commander gera uma ou mais linhas de comando R no quadro do Script.
 - Possibilidade de edição, customização, adaptação, replicação.
 - Dica: digite `?read.table` na janela de console do R



The screenshot displays the R Commander application window. The title bar reads "R Commander". The menu bar includes "File", "Edit", "Data", "Statistics", "Graphs", "Models", "Distributions", "Tools", and "Help". Below the menu bar, there is a toolbar with a "Data set:" dropdown showing "CSS", an "Edit data set" button, a "View data set" button, and a "Model:" dropdown showing "<No active model>". The main workspace is divided into two panes: "R Script" and "R Markdown". The "R Script" pane contains the following R code:

```
CSS <-  
  read.table("C:/Users/Marcelo/Dropbox/artigos-internet/Livros/Using the R Comma  
| header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

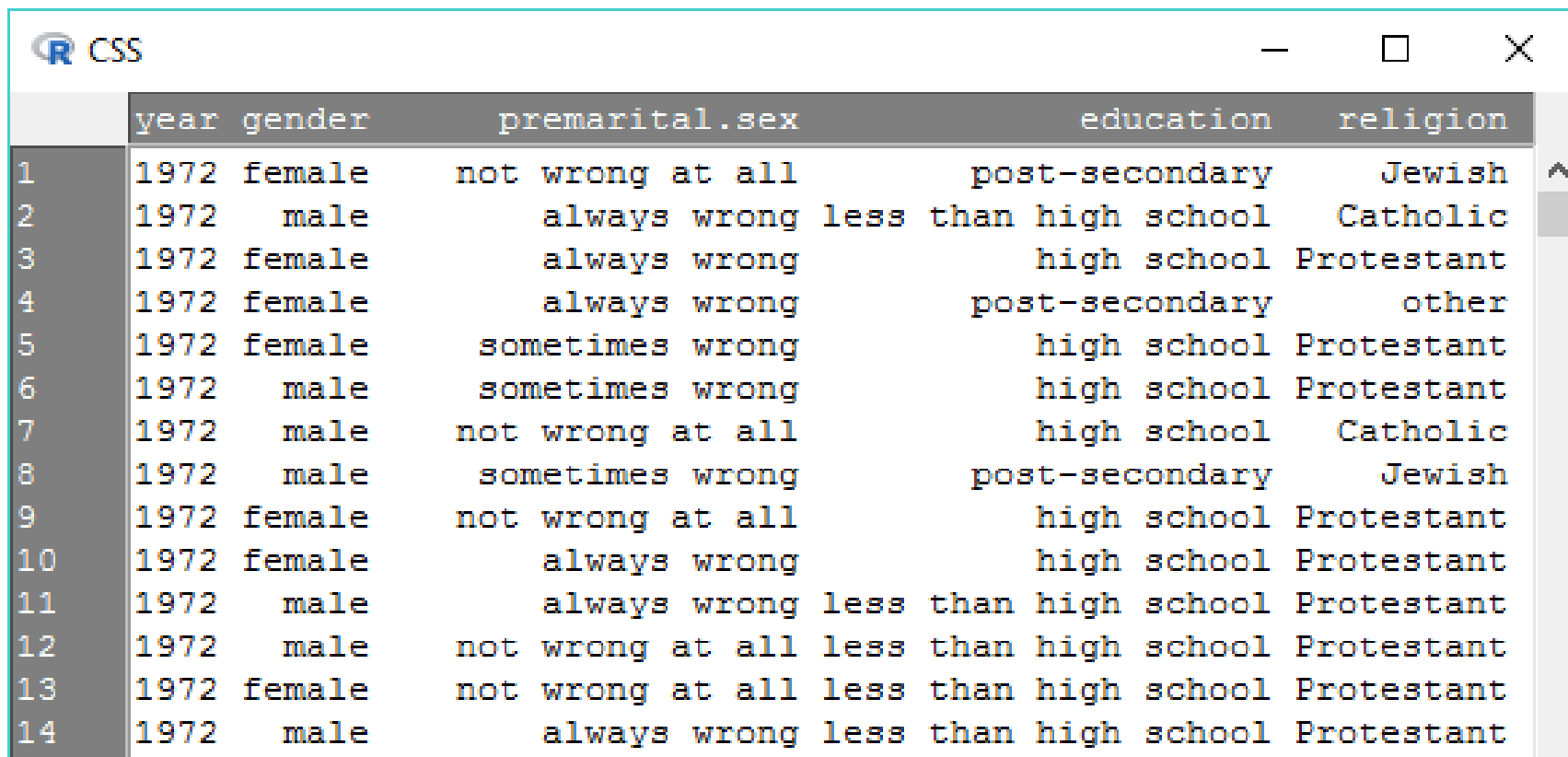
The "Output" pane at the bottom shows the execution of the code, with the prompt ">" and the code lines:

```
> CSS <-  
+   read.table("C:/Users/Marcelo/Dropbox/artigos-internet/Livros/Using the R Co  
+   header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

A "Submit" button with a gear icon is located in the top right corner of the Output pane.

Visualização e alteração de variáveis

- Botão *View data set*



The image shows a screenshot of an R console window titled "R CSS". The window displays a data set with the following columns: year, gender, premarital.sex, education, and religion. The data is presented in a table format with 14 rows. The first row is highlighted in grey.

	year	gender	premarital.sex	education	religion
1	1972	female	not wrong at all	post-secondary	Jewish
2	1972	male	always wrong	less than high school	Catholic
3	1972	female	always wrong	high school	Protestant
4	1972	female	always wrong	post-secondary	other
5	1972	female	sometimes wrong	high school	Protestant
6	1972	male	sometimes wrong	high school	Protestant
7	1972	male	not wrong at all	high school	Catholic
8	1972	male	sometimes wrong	post-secondary	Jewish
9	1972	female	not wrong at all	high school	Protestant
10	1972	female	always wrong	high school	Protestant
11	1972	male	always wrong	less than high school	Protestant
12	1972	male	not wrong at all	less than high school	Protestant
13	1972	female	not wrong at all	less than high school	Protestant
14	1972	male	always wrong	less than high school	Protestant

- *Statistics > Summaries > Active data set*

R Script R Markdown

```
read.table("C:/Users/Marcelo/Dropbox/artigos-internet/Livros/Using the R Comm
header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
fix(CSS)
summary(CSS)
```

Output

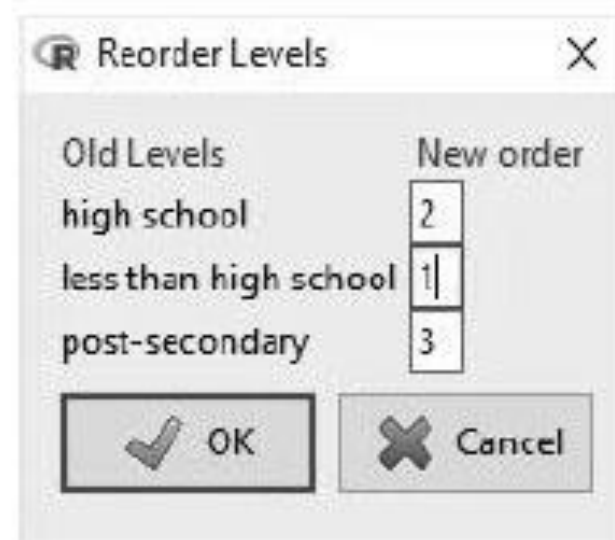
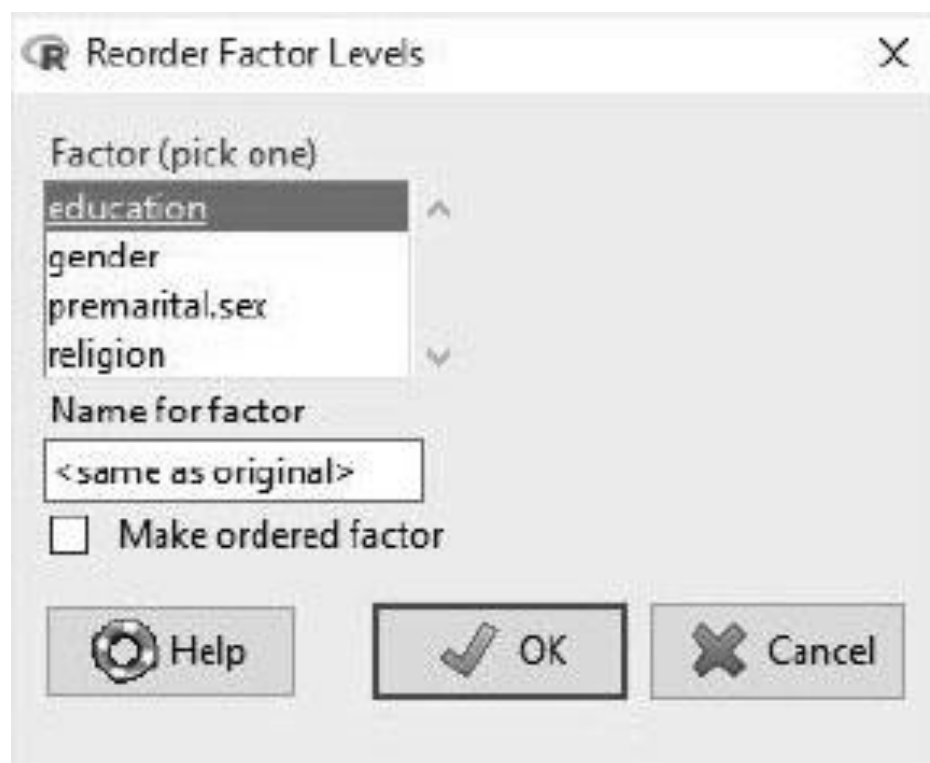
Submit

year	gender	premarital.sex
Min. :1972	female:18651	almost always wrong: 3177
1st Qu.:1982	male :14703	always wrong : 9196
Median :1991		not wrong at all :13965
Mean :1991		sometimes wrong : 7016
3rd Qu.:2000		
Max. :2012		

education	religion
high school :17251	Catholic : 8225
less than high school: 7508	Jewish : 714
post-secondary : 8595	none : 3579
	other : 1182
	Protestant:19654

Categorias fora da ordem natural

- *Data > Manage variables in active data set > Reorder factor levels*



```
education
less than high school: 7508
high school           :17251
post-secondary       : 8595
```

- Agrupamento dos anos na variável “década”:
- *Data > Manage variables in active data set > Recode variables*

Recode Variables

Variables to recode (pick one or more)

education
gender
premarital.sex
religion
year

New variable name or prefix for multiple recodes: decade

Make (each) new variable a factor

Enter recode directives

```
1970:1979 = "1970s"  
1980:1989 = "1970s"  
1990:1999 = "1970s"  
2000:hi = "2000s"
```

Help Reset OK Cancel Apply

Valores especiais **lo** e **hi** representam os valores mínimo e máximo de uma variável numérica

- Agrupamento das categorias “almost always wrong” e “always wrong” em uma única categoria
 - *Data > Manage variables in active data set > Recode variables*

```
premarital
not wrong at all:13965
sometimes wrong : 7016
wrong           :12373
```



Recode Variables

Variables to recode (pick one or more)

- decade
- education
- gender
- premarital.sex
- religion
- year

New variable name or prefix for multiple recodes: premarital

Make (each) new variable a factor

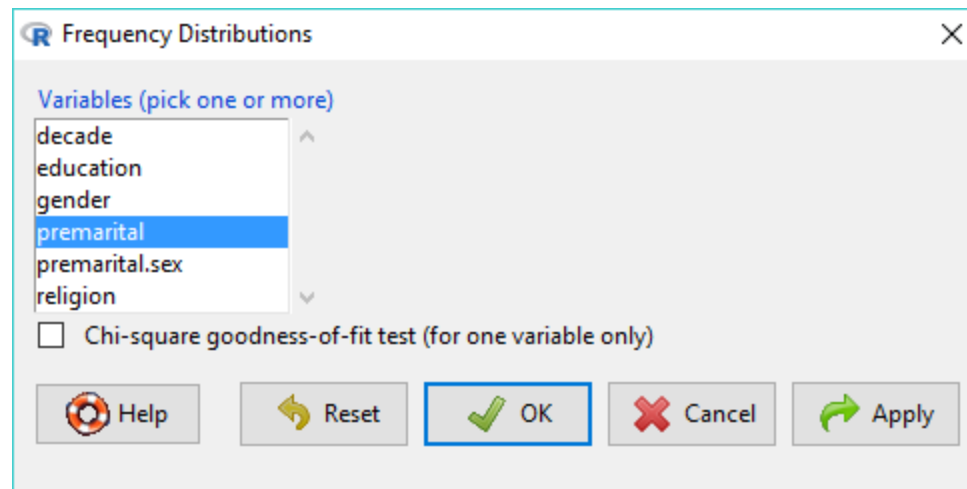
Enter recode directives

```
"always wrong", "always wrong" = "wrong"
```

Help Reset OK Cancel Apply

"almost always wrong", "always wrong" = "wrong"

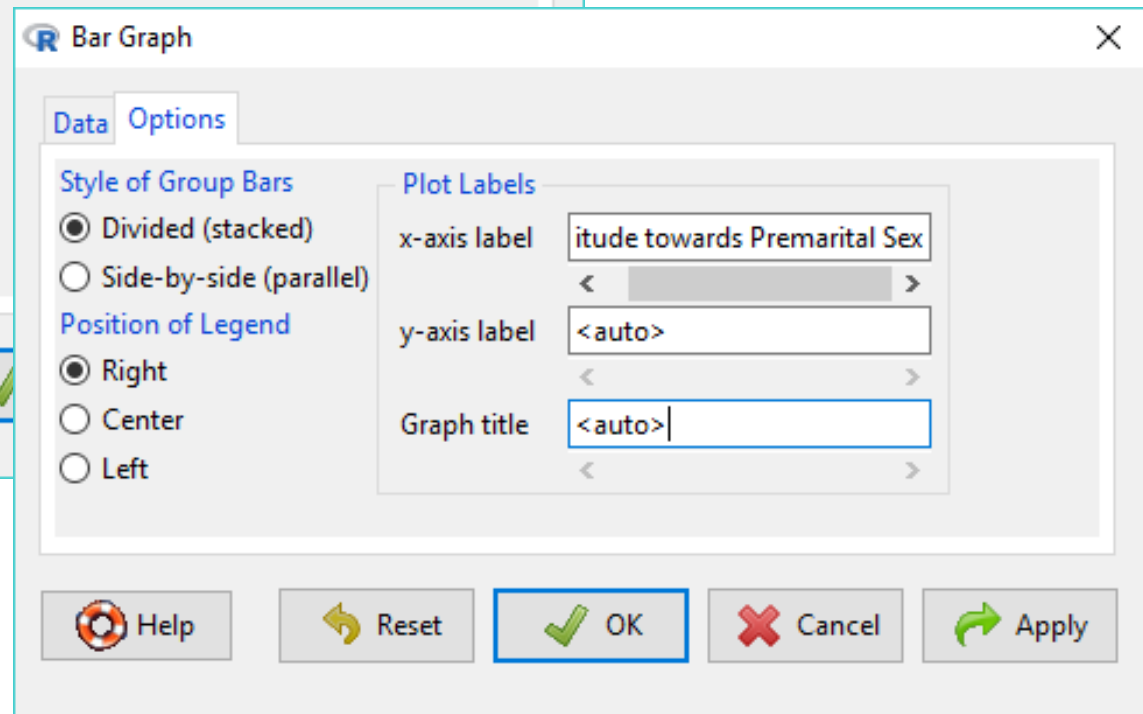
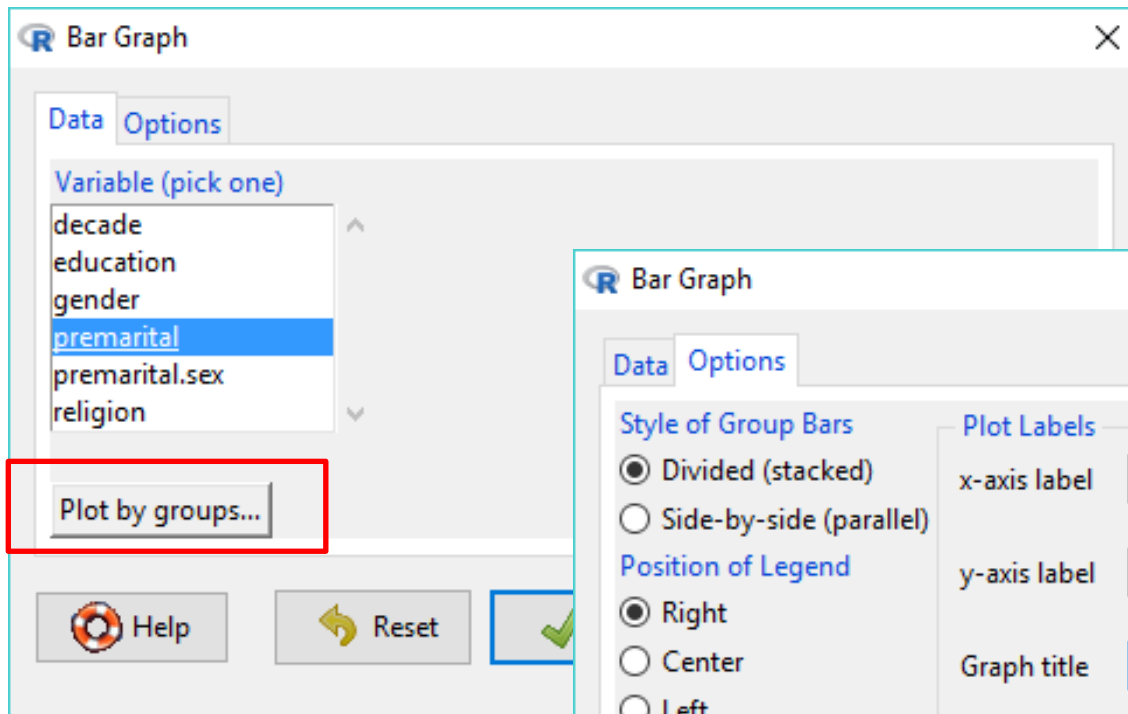
- Distribuição de frequências (tabela)
 - *Statistics > Summaries > Frequency distributions*



```
counts:
premarital
not wrong at all    sometimes wrong    wrong
                13965                7016                12373

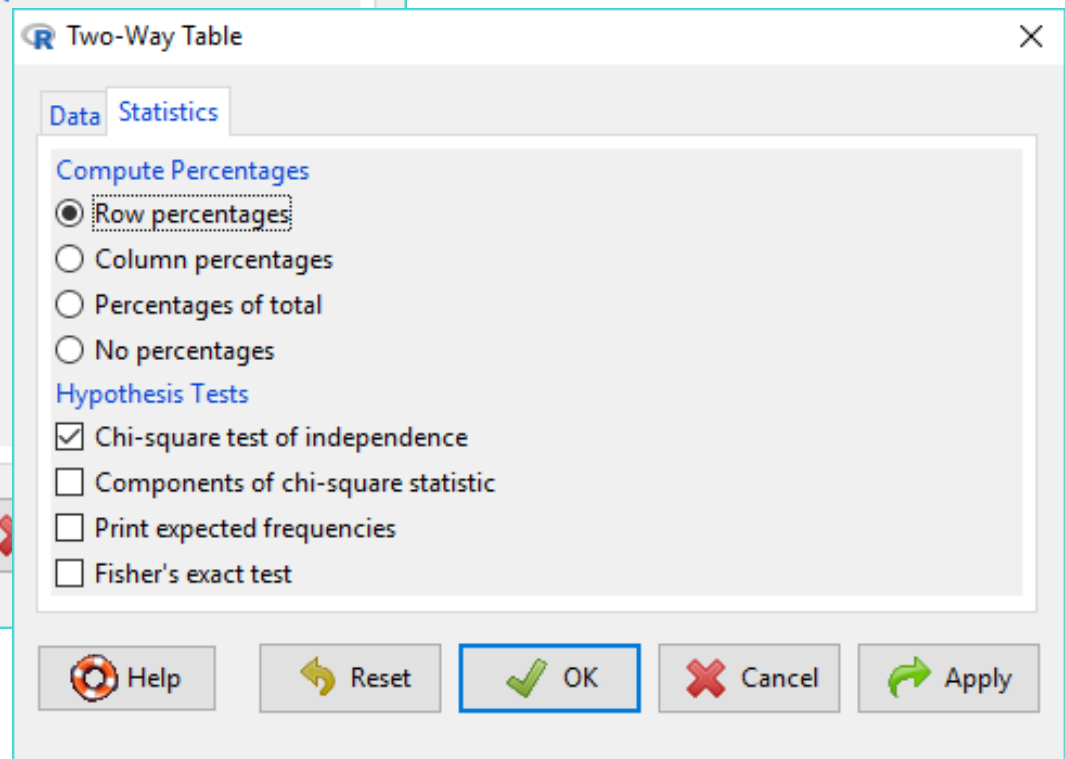
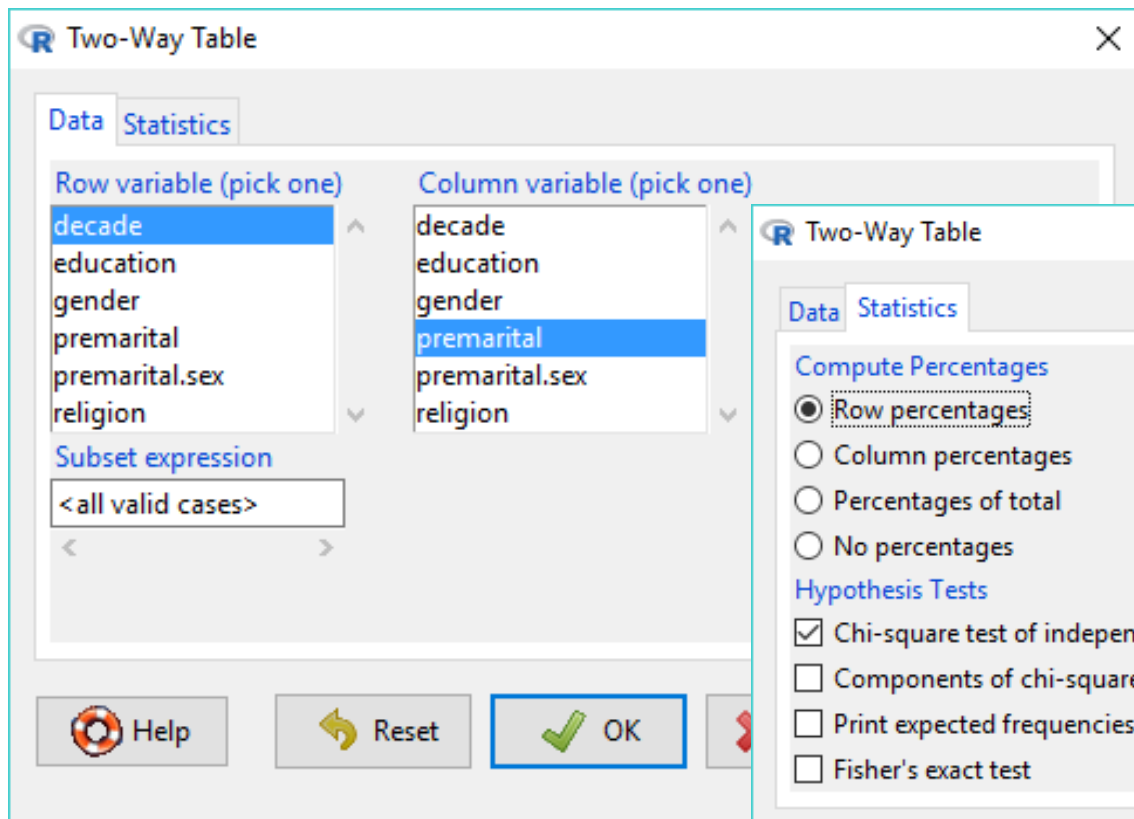
percentages:
premarital
not wrong at all    sometimes wrong    wrong
                41.87                21.03                37.10
```

- Gráfico de barras
 - *Graphs > Bar graph*



- Tabela de contingência

- *Statistics > Contingency tables > Two-way table*



- Tabela de contingência

- *Statistics > Contingency tables > Two-way table*

Frequency table:

		premarital		
decade	not wrong	at all	sometimes	wrong wrong
1970s		2423		1692 3207
1980s		3348		1789 3017
1990s		3647		1797 3035
2000s		4547		1738 3114

Row percentages:

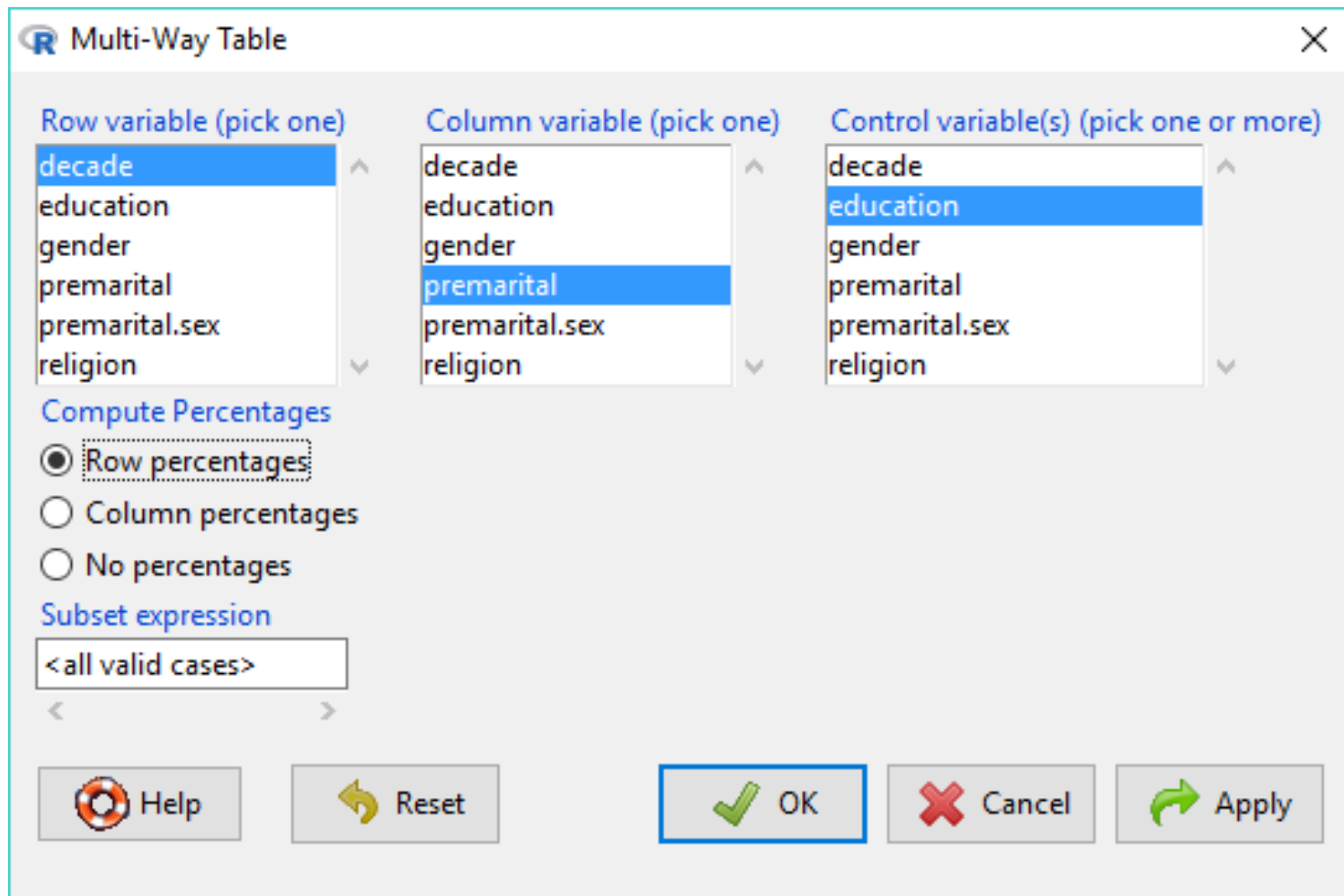
		premarital			Total	Count
decade	not wrong	at all	sometimes	wrong wrong		
1970s		33.1		23.1 43.8	100	7322
1980s		41.1		21.9 37.0	100	8154
1990s		43.0		21.2 35.8	100	8479
2000s		48.4		18.5 33.1	100	9399

Pearson's Chi-squared test

data: .Table

X-squared = 413.3, df = 6, p-value < 2.2e-16

- Tabela de contingência (múltiplas variáveis)
 - *Statistics > Contingency tables > Multi-way table*



- Tabela de contingência (múltiplas variáveis)
 - *Statistics > Contingency tables > Multi-way table*

Row percentages:

, , education = less than high school

		premarital			Total	Count
decade	not wrong	at all	sometimes	wrong	wrong	
1970s		26.6		19.5	53.8	99.9 2568
1980s		34.7		16.1	49.2	100.0 2182
1990s		36.8		17.0	46.2	100.0 1408
2000s		40.9		16.1	43.0	100.0 1350

, , education = high school

		premarital			Total	Count
decade	not wrong	at all	sometimes	wrong	wrong	
1970s		35.4		23.8	40.7	99.9 3630
1980s		42.3		23.2	34.6	100.1 4224
1990s		43.4		20.7	35.9	100.0 4545
2000s		48.4		17.7	33.9	100.0 4852

- Tabela de contingência sobre subconjuntos
 - *Statistics > Contingency tables > Two-way table*

The image displays two overlapping screenshots of the SPSS 'Two-Way Table' dialog box. The left screenshot shows the 'Data' tab, where the 'Row variable (pick one)' is 'education' and the 'Column variable (pick one)' is 'premarital'. The 'Subset expression' field is highlighted with a red box and contains the text 'less than high school'. The right screenshot shows the 'Statistics' tab, where the 'Compute Percentages' section has 'Row percentages' selected. Under 'Hypothesis Tests', the 'Chi-square test of independence' checkbox is checked. A callout box at the bottom left points to the 'Subset expression' field with the text 'education== "less than high school"'. Buttons for 'Reset', 'OK', 'Cancel', and 'Apply' are visible at the bottom of both dialog boxes.

education== "less than high school"

- Tabela de contingência sobre subconjuntos
 - *Statistics > Contingency tables > Two-way table*

```
Frequency table:
```

```
      premarital
decade  not wrong at all  sometimes wrong  wrong
1970s          684          502    1382
1980s          757          352    1073
1990s          518          239     651
2000s          552          217     581
```

```
Row percentages:
```

```
      premarital
decade  not wrong at all  sometimes wrong  wrong  Total  Count
1970s          26.6          19.5    53.8    99.9    2568
1980s          34.7          16.1    49.2   100.0    2182
1990s          36.8          17.0    46.2   100.0    1408
2000s          40.9          16.1    43.0   100.0    1350
```

```
      Pearson's Chi-squared test
```

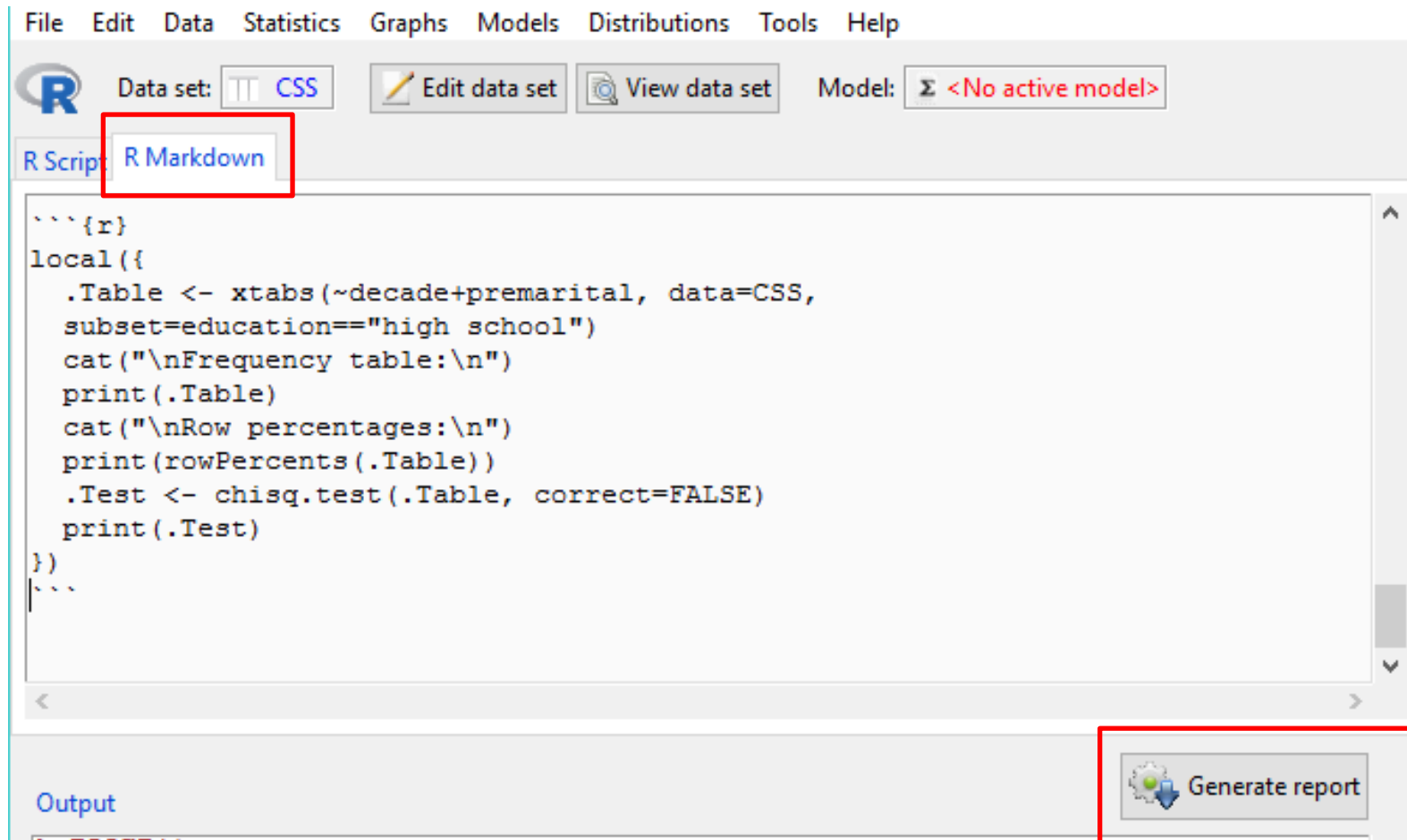
```
data: .Table
```

```
X-squared = 98.088, df = 6, p-value < 2.2e-16
```

Repetição para os demais níveis de escolaridade: pode-se editar o Script R e executar as linhas afetadas

- Exportação dos resultados
 - “Copy-and-paste” para um editor de textos
 - Cópia dos resultados gerados na tela do R Commander
 - Cuidado: usar fontes monoespaciaadas (p.ex. Courier)
 - Copy-and-Paste dos gráficos gerados na janela do ambiente R
 - Geração do relatório na aba R Markdown
 - Markdown: linguagem de marcadores
 - Gera relatórios permanentes e reproduzíveis, mesclando os comandos em R com comentários explicativos
 - Útil para refazer as análises com futuros data sets com a mesma estrutura e mesmas variáveis do data set atual

- Exportação dos resultados
 - Geração do relatório na aba R Markdown:



The screenshot displays the RStudio software interface. At the top, the menu bar includes 'File', 'Edit', 'Data', 'Statistics', 'Graphs', 'Models', 'Distributions', 'Tools', and 'Help'. Below the menu bar, the 'Data set:' field shows 'CSS' with an 'Edit data set' button and a 'View data set' button. The 'Model:' field shows 'Σ <No active model>'. The main workspace area has two tabs: 'R Script' and 'R Markdown', with the 'R Markdown' tab selected and highlighted by a red box. The code editor contains the following R code:

```
```{r}
local({
 .Table <- xtabs(~decade+premarital, data=CSS,
 subset=education=="high school")
 cat("\nFrequency table:\n")
 print(.Table)
 cat("\nRow percentages:\n")
 print(rowPercents(.Table))
 .Test <- chisq.test(.Table, correct=FALSE)
 print(.Test)
})
|```
```

At the bottom of the interface, the 'Output' pane is visible, and a 'Generate report' button with a gear icon is highlighted by a red box.

- Gravação e exportação de um data set
  - O R Commander permite gravar um data set em formato interno do R ou exportá-lo para um arquivo texto.
    - Conveniente caso um data set tenha sido criado ou alterado no R Commander e deverá ser utilizado para análises posteriores
  - Gravação em formato próprio:
    - *Data > Active data set > Save active data*
    - Mantém as ordem dos níveis dos fatores
    - Mais rápido de recarregar do que um arquivo texto grande
  - Exportação:
    - *Data > Active data set > Export active data set*
    - Não preserva todas as características (p.ex. ordem dos níveis dos fatores)
    - Útil para análise ou visualização em outros softwares