

# **R Commander**

Facilitando o aprendizado da Estatística

George Amarante Melo

Jony Arrais Pinto Junior

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Instalação e abertura do R Commander</b>	<b>3</b>
2.1	Instalação . . . . .	3
2.2	Estrutura do R Commander . . . . .	5
<b>3</b>	<b>Leitura, criação e edição de bases de dados</b>	<b>7</b>
3.1	Criação e edição de bases de dados . . . . .	7
3.2	Leitura de bases de dados externas . . . . .	8
3.2.1	Submenu “Carregar conjunto de dados” . . . . .	8
3.2.2	Submenu “Importar arquivos de dados” . . . . .	8
3.2.3	Submenu “Conjunto de dados em pacotes” . . . . .	9
<b>4</b>	<b>Manipulação de variáveis no banco de dados ativo</b>	<b>12</b>
4.1	Recodificação e modificação de classe de variáveis . . . . .	12
<b>5</b>	<b>Medidas descritivas e tabelas</b>	<b>14</b>
5.1	Resumo da base de dados . . . . .	14
5.2	Comando “Tabelas de contingência” . . . . .	15
<b>6</b>	<b>Criação e personalização de gráficos</b>	<b>16</b>
6.1	Gráfico de barras . . . . .	16
6.2	Gráfico de setores . . . . .	17
6.3	Ramos-e-folhas . . . . .	18
6.4	Histograma . . . . .	19
6.5	Boxplot . . . . .	20
6.6	Diagrama de dispersão . . . . .	21
6.7	Salvando os gráficos gerados . . . . .	22

# 1 Introdução

Esta apostila tem por objetivo introduzir o R Commander, ferramenta disponibilizada pelo bastante conhecido software estatístico R.

É uma nova maneira de realizar análises estatísticas, tendo em vista que não é preciso conhecimento prévio na linguagem de programação R. Uma maneira mais prática e interativa de fazer algumas manipulações em bases de dados.

No entanto, não é uma ferramenta tão poderosa quanto o próprio R, já que muitas tarefas necessitam de algoritmos complexos e bem elaborados para que sejam feitas de maneira correta.

O foco deste material é mostrar as aplicações do R Commander no aprendizado da Estatística Básica e apresentá-lo como uma alternativa a aqueles que buscam fazer análises de dados sem ter conhecimento na linguagem de programação R; como, por exemplo, estudantes iniciantes em Estatística Básica e pessoas de outras áreas que precisem fazer alguma análise estatística.

Na elaboração desta apostila, o artigo de Umberto Guarnier Mignozzetti [1] foi usado como material auxiliar.

## 2 Instalação e abertura do R Commander

O R Commander é um pacote oferecido pelo software R. Portanto, é necessário ter o R instalado no seu computador. Com o software já instalado, podemos prosseguir com a instalação do Commander.

### 2.1 Instalação

A instalação pode ser feita a partir do RGui ou do RStudio. Aqui será utilizado o RGui, mas o processo ocorre de maneira parecida nos dois.

- Clique na aba **Pacotes** → **Instalar pacotes**;
- Aparecerá uma lista de espelhos CRAN. Selecione o mais próximo de onde você está;
- Em seguida, aparecerá uma lista dos pacotes disponíveis. Escolha o pacote **Rcmdr**. Isso iniciará a instalação do R Commander. Vale ressaltar que é preciso uma conexão ativa à internet, pois o R fará o download de alguns outros pacotes;
- Depois de terminada a instalação, digite **require(Rcmdr)** e aperte enter, e o R Commander aparecerá na tela do seu computador.

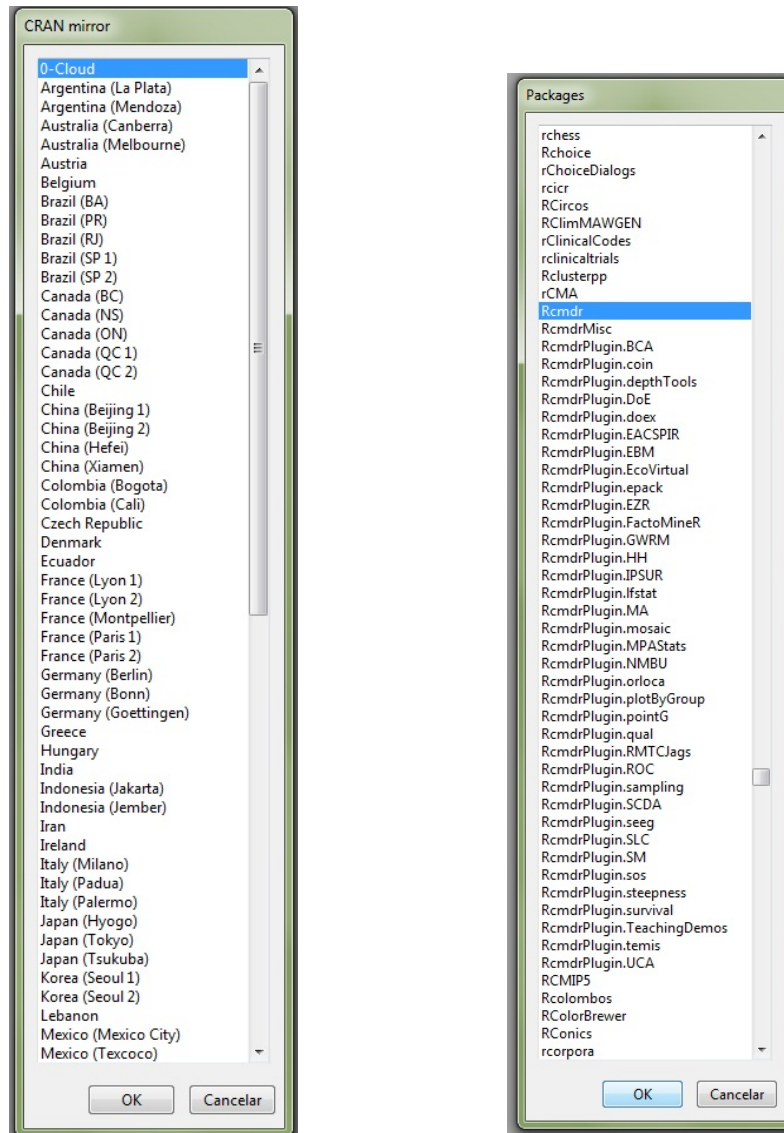


Figura 1: Lista de espelhos CRAN e pacotes disponíveis

*Observação: É possível que, na hora de escolher o espelho CRAN, apareça uma lista com muito poucos locais disponíveis, onde provavelmente o que você deseja não está lá. Se esse for o caso, clique na última opção (**HTTP mirrors**) e a lista com todos os espelhos disponíveis aparecerá.*

## 2.2 Estrutura do R Commander

Ao abrir o R Commander, será carregada a seguinte janela:

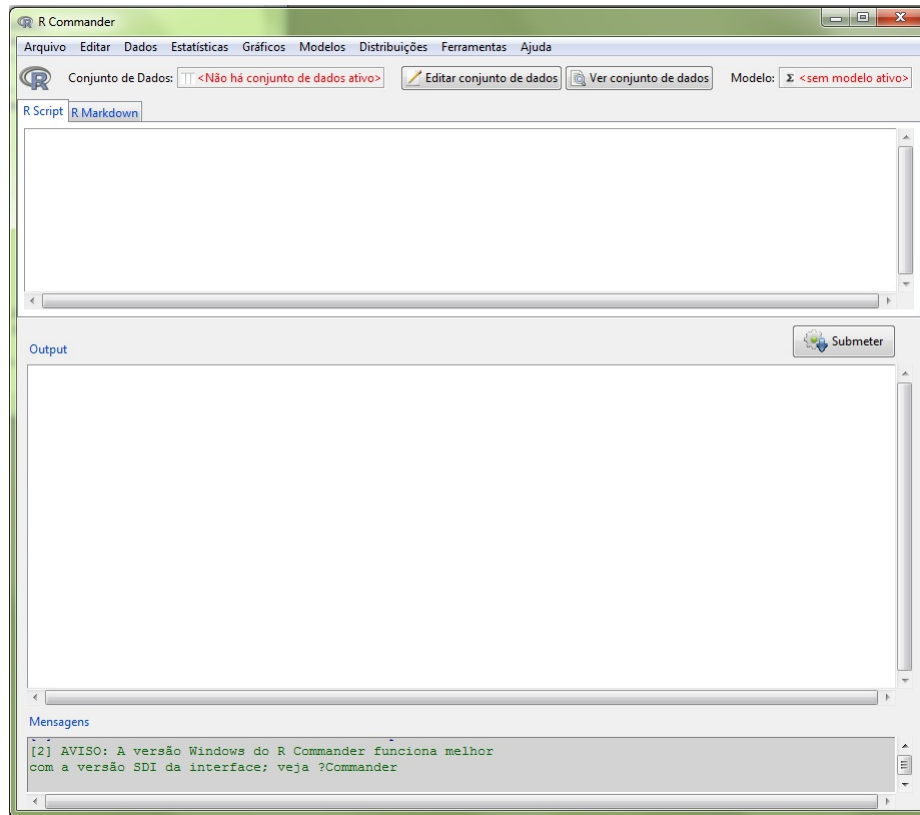


Figura 2: Estrutura geral do R Commander

Essa é a estrutura principal do programa. Na barra superior podemos ver alguns menus, os quais serão abordados mais detalhadamente ao decorrer da apostila. Vale lembrar que os menus **Arquivo** e **Editar** são para acessar opções gerais do programa, como abrir, salvar e modificar a estrutura de arquivos.

Abaixo dessa barra, encontra-se um menu relacionado a conjunto de dados e modelo. São menus gerais que permitem a visualização e edição do conjunto de dados e do modelo estatístico utilizados na ocasião.

A primeira área em branco é o espaço onde é visualizado o script do programa. É uma importante ferramenta, principalmente para aqueles que não conhecem a linguagem de

programação R. Isso porque a cada tarefa feita no programa, é reproduzida na janela de script a linha de comando que seria necessária para realizar essa mesma tarefa usando o R tradicional. Isso é importante, pois pode servir de aprendizado para aqueles que querem aprender a programar.

O botão **Submiter** tem a função de executar determinada parte da janela de script. Selecionando alguma linha de comando e submetendo-a, apenas ela será reproduzida. É possível usar em uma ou mais linhas de comando por vez.

Logo após se encontra a janela output. É a principal, onde os resultados de todas as tarefas são exibidos. Nessa janela aparecem tabelas, resumos, modelos, resultados de cálculos, entre outros. Os resultados de quase tudo que é feito no programa aparecem nessa área.

Por fim, a área de mensagens do R Commander. É um espaço destinado a avisos que o programa dá ao usuário. Tanto com relação ao banco de dados e variáveis quanto ao relato de erros; este último tipo de aviso é um dos mais importantes, pois indica ao usuário problemas ocorridos durante a compilação dos códigos derivados dos comandos.

### 3 Leitura, criação e edição de bases de dados

Nesta seção, serão abordados tópicos referentes à obtenção de bases de dados no R Commander.

#### 3.1 Criação e edição de bases de dados

É possível obter uma base de dados de diferentes maneiras no R Commander. A maneira mais trabalhosa e limitada é criando-a. Apesar de parecer pouco útil, pode ser um meio interessante quando se trata de bases pequenas.

Para isso, basta acessar o menu **Dados** → **Novo conjunto de dados**. Defina um nome para o novo conjunto de dados e a seguinte janela aparecerá:

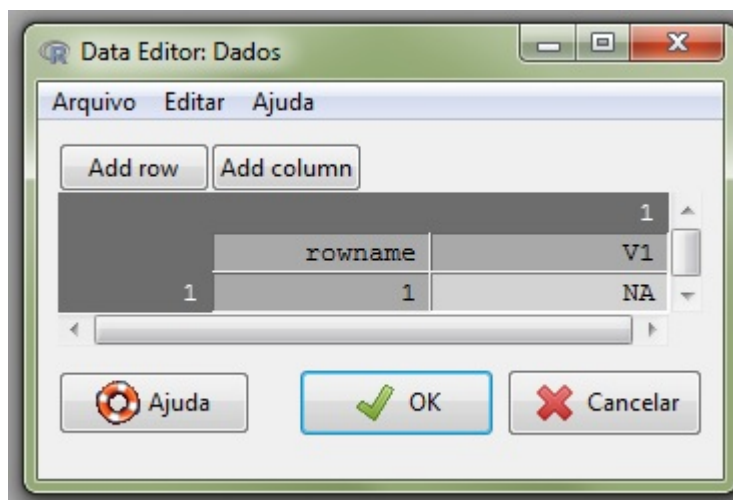


Figura 3: Janela para criação de uma base de dados

É possível adicionar linhas e colunas à planilha, com os botões **Add row** e **Add column**. Com o número de linhas e colunas definidos, basta clicar nas caselas para definir nomes dos indivíduos, das variáveis e escolher os valores desejados para cada observação. Feitos os ajustes, basta clicar em “Ok” e o banco de dados estará carregado.

Com isso, esse será o banco de dados ativo. Tendo um conjunto de dados ativo, é possível vê-lo em uma janela separada ou também editá-lo. Nesta última opção, aparecerá



uma janela, como a primeira usada para criar a base. As edições são feitas da mesma maneira como ocorreu na criação. Essas tarefas podem ser feitas recorrendo aos botões **Editar conjunto de dados** e **Ver conjunto de dados**.

## 3.2 Leitura de bases de dados externas

A maneira mais comum de se carregar uma base de dados - tanto no R quanto no R Commander - é carregando uma base externa. Bases em vários formatos podem ser lidas no programa, usando os comandos certos.

### 3.2.1 Submenu “Carregar conjunto de dados”

O primeiro comando a ser explicado é o comando **Carregar conjunto de dados**, do menu **Dados**. Com esse comando, é possível ler apenas bases de dados produzidas no R. É usado para carregar arquivos com formato “.RData”. Para carregar bases deste tipo basta que, após clicar nos menus mencionados anteriormente, o usuário localize no computador o banco de dados que será utilizado.

### 3.2.2 Submenu “Importar arquivos de dados”

Esta maneira de obtenção de bases de dados é a mais utilizada, tendo em vista que a maioria dos conjuntos de dados utilizados nas análises estatísticas são originários de softwares externos como Excel, SPSS, SAS e Calc.

Para ler bases de dados desses softwares externos, é necessário acessar o menu **Dados** → **Importar arquivos de dados**.

A primeira opção é “**de arquivo texto, clipboard ou url**”. Utilizada para ler bases de dados em formatos mais simples, como .txt e .csv. Selecionando essa opção, abrirá uma janela onde o usuário deve ajustar as especificações da base de dados a ser lida, como: nome, representação de dados faltantes, localização do arquivo, separador de campos (por exemplo, os arquivos .csv são usualmente separados por vírgula) e representação de decimais. Feito isso, basta clicar em “Ok” e localizar o arquivo no computador.

As outras opções podem ser utilizadas de maneira semelhante à primeira.

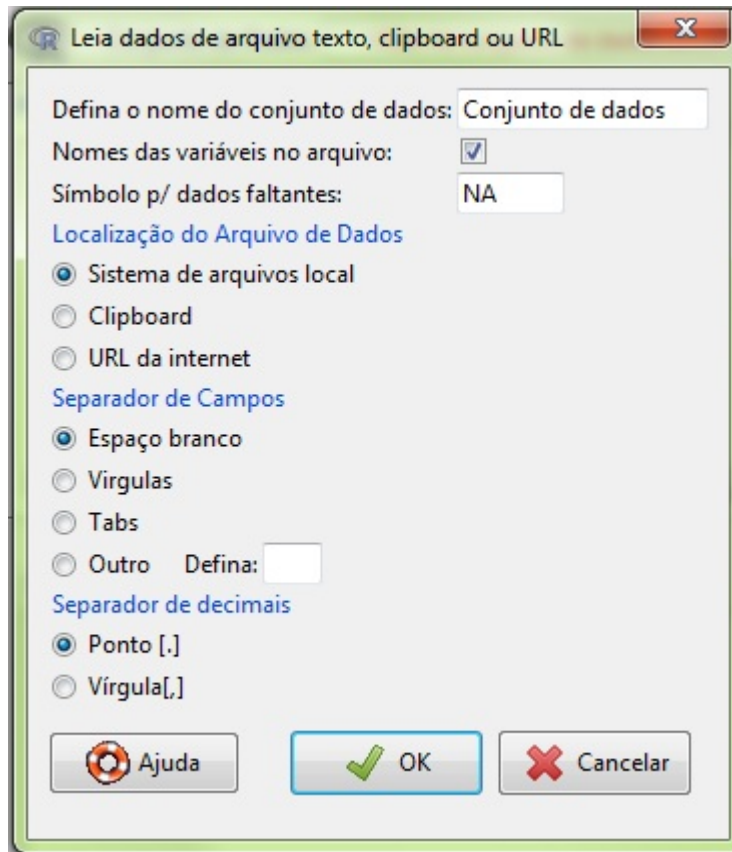


Figura 4: Submenu “Importar arquivos de dados”: exemplo

### 3.2.3 Submenu “Conjunto de dados em pacotes”

Este menu possibilita acessar bancos de dados previamente carregados no R. São bases disponibilizadas para que o usuário possa praticar usando dados reais, em vários contextos diferentes.

#### Comando “Listar conjuntos de dados em pacotes”

Acessando este comando, aparecerá uma janela no RGui com uma lista contendo os nomes todos os conjuntos de dados que o programa disponibiliza, além de um breve resumo sobre o que cada um contém.

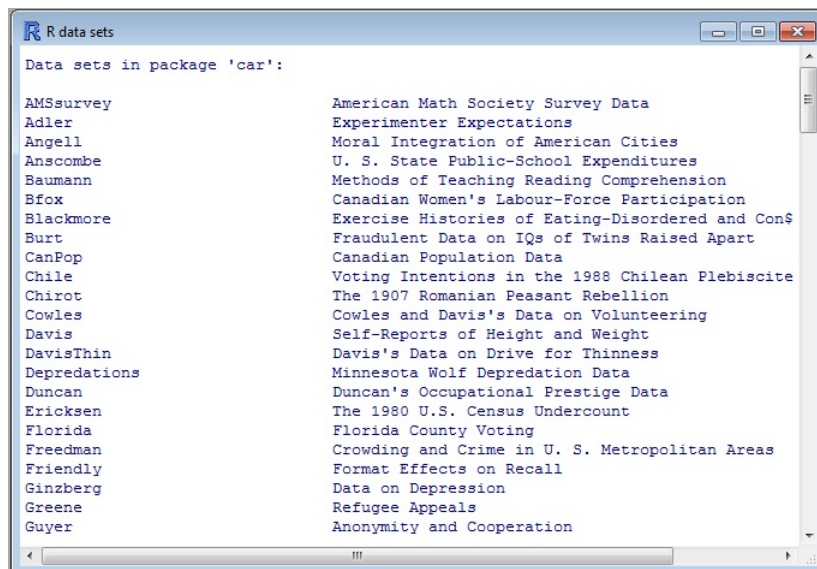


Figura 5: Comando “Listar conjuntos de dados em pacotes”

### Comando “Ler pacotes de dados attachado”

Aqui é possível carregar um conjunto de dados dentre os disponibilizados pelo R, fazendo com que o usuário possa fazer manipulações nessa base assim como seria feito no caso da leitura de dados externos.

Após clicar no comando, basta escolher um dos pacotes disponíveis, depois um dos conjuntos de dados pertencente ao pacote selecionado e, logo após, dar um nome à base escolhida (é possível também, utilizar a base com seu nome original, dando um clique duplo no nome do banco de dados). Em seguida, após clicar em “Ok”, o conjunto selecionado será o banco de dados ativo no Commander.

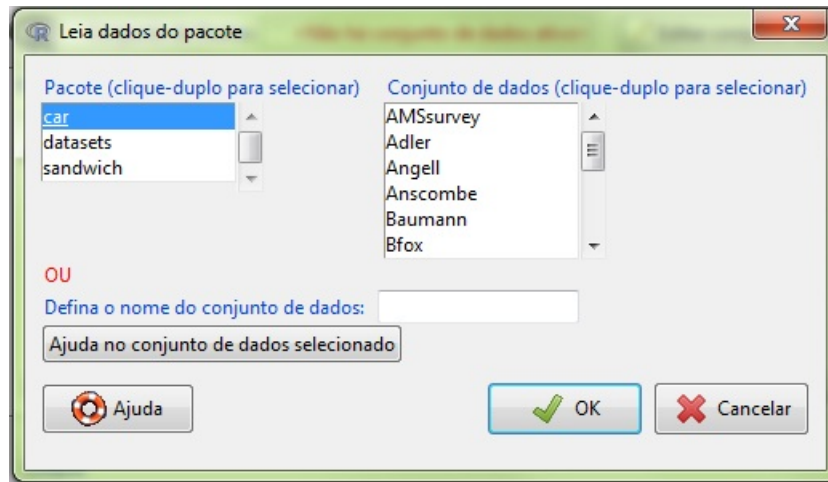


Figura 6: Comando “Ler pacote de dados attachado”

## 4 Manipulação de variáveis no banco de dados ativo

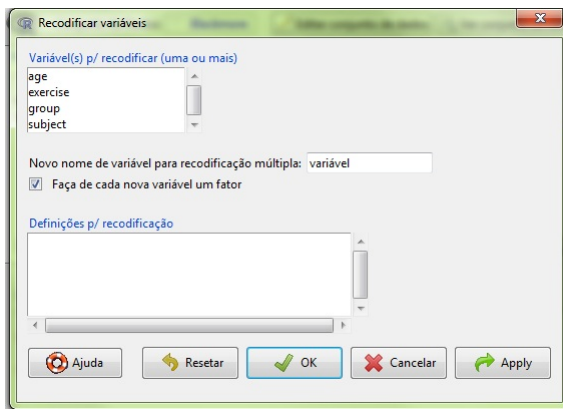
Manipular variáveis em um conjunto de dados é uma tarefa importante quando se deseja fazer análises estatísticas. O submenu **Modificação de variáveis no conjunto de dados** possibilita fazer diversas manipulações como, por exemplo, recodificar variáveis, trocar classe, excluir, renomear, entre outras.

### 4.1 Recodificação e modificação de classe de variáveis

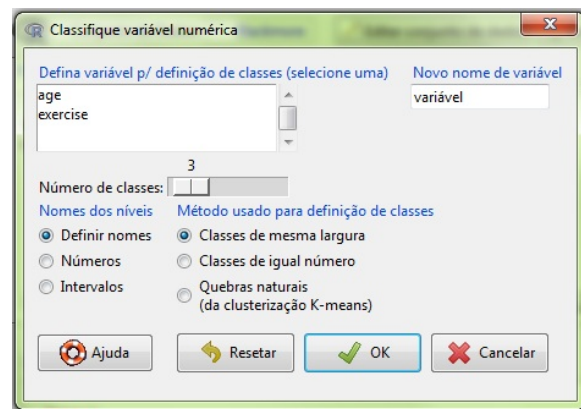
Existem diferentes maneiras de recodificar variáveis em um conjunto de dados no R Commander. Algumas são:

1. **DEFINIR, MANUALMENTE, CLASSES PARA UMA VARIÁVEL:** utilizando o comando **Recodificar variáveis**. É preciso selecionar a variável a ser recodificada, e dar um nome à nova variável. Após isso, é preciso digitar os critérios de recodificação que serão adotados. Há um texto explicando sobre como digitar esses critérios no botão “Ajuda”, no canto inferior esquerdo da janela.
2. **DEFINIR CLASSES AUTOMATICAMENTE:** o comando **Agrupar em classes uma variável numérica (ou fator)** define as classes automaticamente. Após selecionar o comando, escolha a variável a ser recodificada e o nome da nova variável. É possível selecionar o número desejado de classes e algumas especificações, como o método usado para definir os limites de cada classe (por exemplo, utilizar classes de mesma largura).
3. **CATEGORIZAR UMA VARIÁVEL NUMÉRICA:** utilizando o comando **Converter variável numérica para fator**. Ao se trabalhar com uma base de dados que tenha variáveis qualitativas, é comum que essas variáveis apareçam com uma codificação numérica, o que não significa que ela seja uma variável quantitativa. No entanto, para fazer com que o R Commander identifique uma variável deste tipo como qualitativa, é preciso transformá-la em fator. É nesse contexto que se insere o

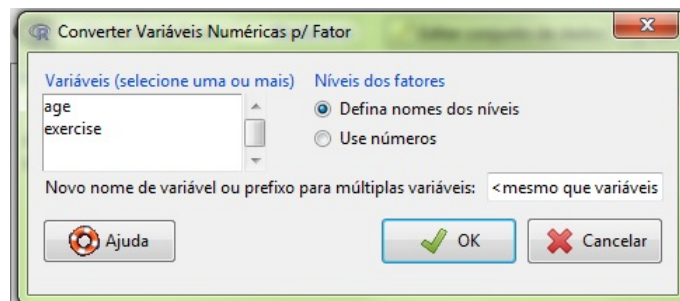
comando mencionado. Ao acessá-lo, basta escolher a variável a ser categorizada, a definição dos níveis dos fatores (nomes ou números) e o nome da nova variável (*obs.: se o nome da variável for o mesmo da variável a ser categorizada, a nova variável substituirá a primeira*).



(a) Comando do item 1



(b) Comando do item 2



(c) Comando do item 3

## 5 Medidas descritivas e tabelas

Uma etapa muito importante no processo de análise de dados é a obtenção de medidas descritivas e tabelas que possam trazer informações relevantes sobre um conjunto de dados. Os submenus a seguir são encontrados ao acessar o menu **Estatísticas**.

### 5.1 Resumo da base de dados

É possível obter diferentes tipos de resumo dos dados através do submenu **Resumos**. Na tabela a seguir são explicados alguns comandos deste submenu:

Tabela 1: Alguns comandos do submenu **Resumos**

Comando	Função
Conjunto de dados ativo	Retorna os valores dos quartis, da média e os valores mínimos e máximos no caso de variáveis quantitativas; para variáveis qualitativas, retorna o número de observações para cada classe da variável.
Resumos numéricos	Retorna uma tabela com algumas medidas descritivas como média, quantis, assimetria, desvio padrão para uma variável ou um grupo de variáveis. É possível definir quais medidas são mostradas.
Distribuições de frequência	Retorna uma tabela de distribuição de frequências para uma variável qualitativa.
Contar observações faltantes	Retorna o número de valores faltantes para cada variável.
Tabela de Estatísticas	Retorna uma tabela que informa a média, mediana, desvio padrão ou intervalo interquartilico (ou outra medida a definir) de uma variável quantitativa separada por fatores (classes) de uma variável qualitativa.

## 5.2 Comando “Tabelas de contingência”

Também é possível construir tabelas de dupla e multientrada, acessando o submenu **Tabelas de contingência**. Para a primeira, é preciso acessar **Estatísticas** → **Tabelas de contingência** → **Tabela de dupla entrada**. Uma janela aparecerá com as variáveis disponíveis a serem escolhidas; é possível personalizar a tabela mostrando porcentagens por linha, coluna e/ou total, além de alguns testes de hipótese.

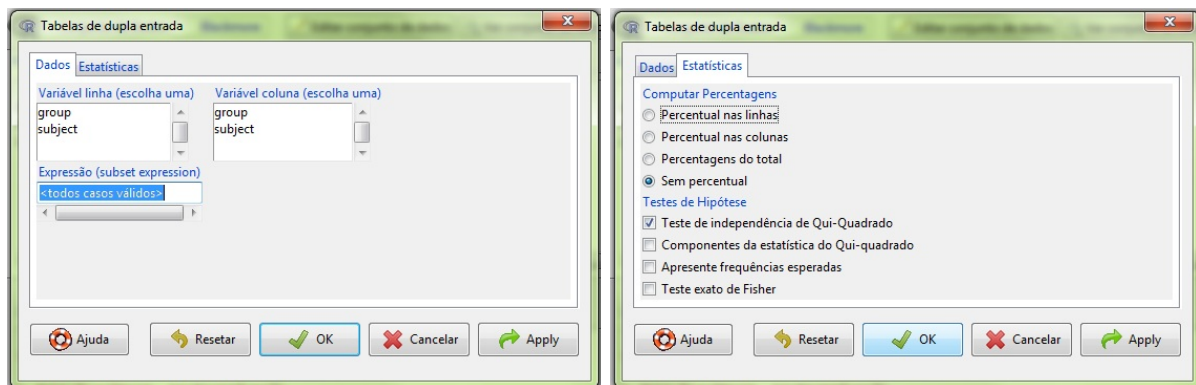


Figura 7: Criando uma tabela de dupla entrada

```
Output
+ .Table <- xtabs(~instruction+expectation, data=Adler)
+ cat("\nFrequency table:\n")
+ print(.Table)
+ .Test <- chisq.test(.Table, correct=FALSE)
+ print(.Test)
+ })

Frequency table:
          expectation
instruction HIGH LOW
GOOD          15  17
NONE          16  18
SCIENTIFIC   18  13

          Pearson's Chi-squared test

data:  .Table
X-squared = 1.0389, df = 2, p-value = 0.5948
```

Figura 8: Exemplo de tabela de dupla entrada



## 6 Criação e personalização de gráficos

O R Commander possibilita a criação de alguns gráficos que podem ajudar muito nas análises de dados. Essa tarefa pode ser feita acessando o menu **Gráficos**. Nesta seção, serão abordados os gráficos mais usados na Estatística Básica.

### 6.1 Gráfico de barras

Acessando o comando **Gráfico de Barras**, uma janela se abrirá onde deve-se escolher a variável qualitativa a ser representada. Também é possível personalizar o gráfico inserindo, por exemplo, rótulos nos eixos e legendas.

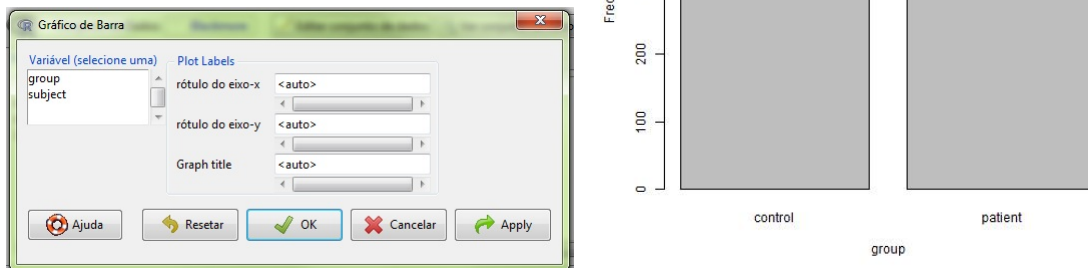


Figura 9: Gráfico de barras: criação e exemplo

## 6.2 Gráfico de setores

Este gráfico pode ser gerado acessando o comando **Gráfico de Pizza**. Uma janela simples se abrirá, onde é preciso escolher uma variável e é possível inserir rótulos e título.

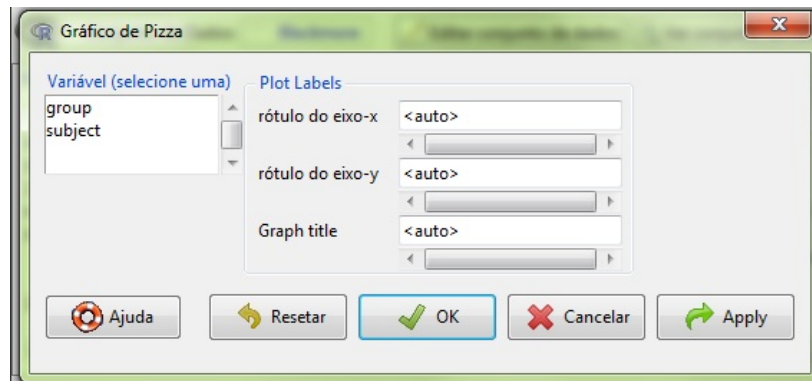


Gráfico de setores da variável Group

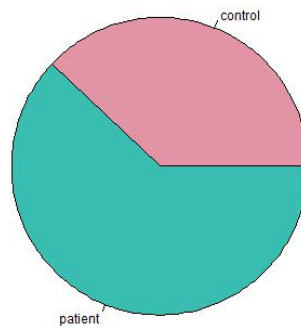


Figura 10: Gráfico de setores: criação e exemplo



## 6.4 Histograma

Ao acessar o comando **Histograma** uma janela se abrirá, onde é preciso definir a variável a ser representada. Além de mudar rótulos e título, é possível definir o número de classes do histograma, e também a escala do eixo vertical (contagem de frequência, porcentagens ou densidade).

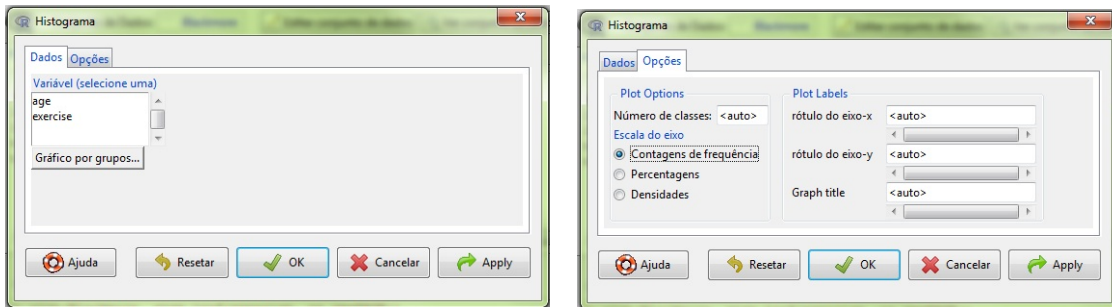


Figura 13: Histograma: criação



Figura 14: Histograma: exemplo



## 6.6 Diagrama de dispersão

Ao acessar o comando **Diagrama de dispersão**, a janela que se abrirá possibilita, depois de definir as variáveis a serem representadas, uma série de critérios para personalização do gráfico; como, por exemplo: tamanho dos pontos, inserção da reta de mínimos quadrados, mudança de título e rótulos, entre outros.

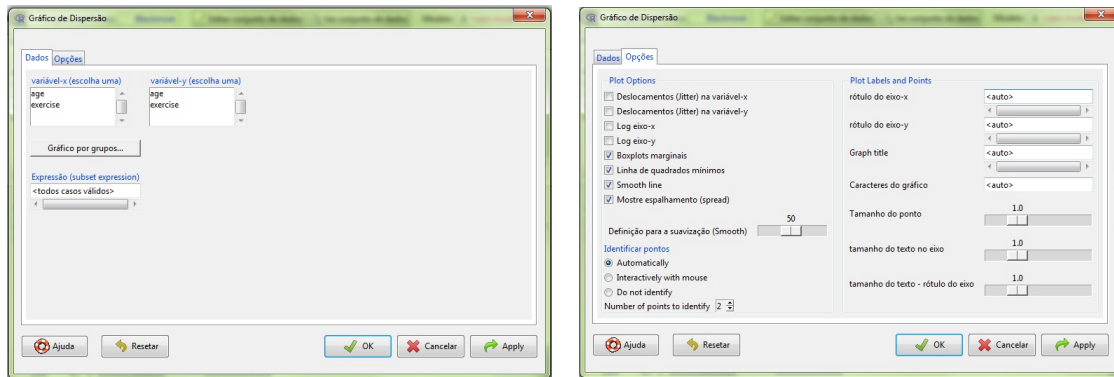


Figura 17: Gráfico de dispersão: criação

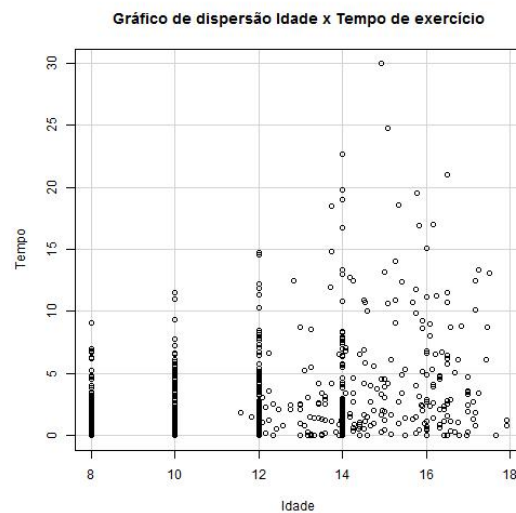


Figura 18: Gráfico de dispersão: exemplo

## 6.7 Salvando os gráficos gerados

Com exceção do gráfico de ramos-e-folhas, os gráficos gerados são mostrados numa nova janela, dentro do RGui. Para salvar um gráfico no computador, pode-se seguir os seguintes passos:

1. Logo após gerar o gráfico no RGui, acessar novamente o menu **Gráficos**;
2. Clicar no último comando da lista, **Salvar gráfico em arquivo**;
3. Escolher a extensão desejada para o arquivo;
4. Definir o diretório e algumas especificações do arquivo a ser salvo.

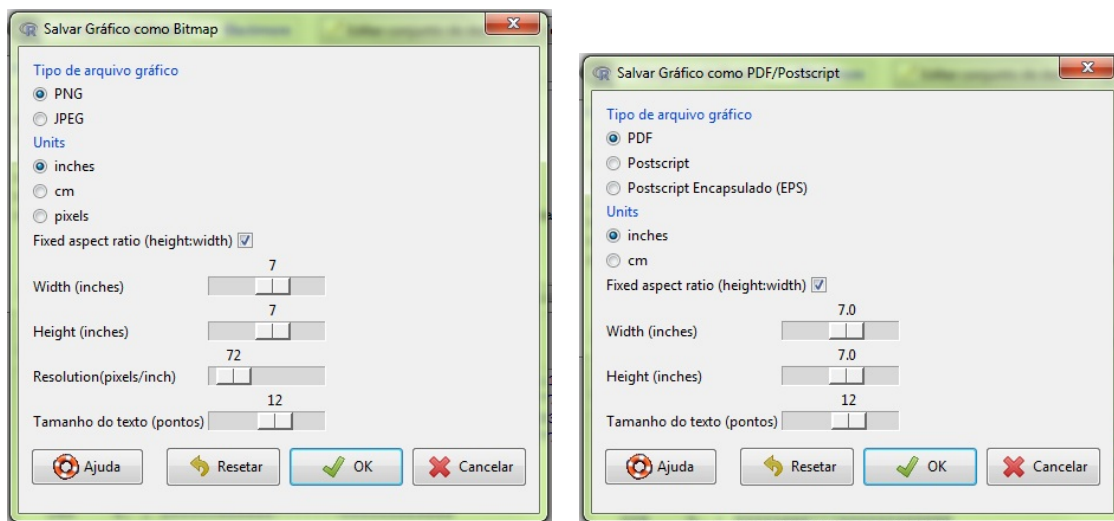


Figura 19: Opções para salvar os gráficos gerados

## Referências

- [1] Umberto Guarnier Mignozzetti. Introdução ao R Commander. 2009.