

Introdução ao R Commander

O objetivo deste tutorial é familiarizar o estudante com o software estatístico R Commander, uma interface gráfica para o R, desenvolvida para permitir ao usuário maior facilidade no uso da linguagem. O principal objetivo de se introduzir R Commander nesta disciplina é facilitar o acesso aos métodos estatísticos disponíveis na linguagem R. Assim o pesquisador deve aprender a usar os principais comandos, menus de organização e recuperação de dados em planilhas a partir do uso deste software.

Table of Contents

- [1 O que você deve saber antes de usar o R Commander](#)
- ▼ [2 Por que R Commander?](#)
 - [2.1 O R Commander:](#)
- ▼ [3 Abrindo o programa](#)
 - [3.1 No Linux](#)
 - [3.2 No Macintosh](#)
 - [3.3 No Windows](#)
- ▼ [4 Manipulação dos dados](#)
 - [4.1 Abrindo uma planilha de dados existente](#)
 - [4.2 Números: vírgulas ou pontos](#)
- [5 Visualizando os dados](#)
- [6 Salvando e recuperando dados](#)
- [7 Elementos da tela do Rcmdr](#)
- ▼ [8 Saindo e retornando ao R Commander](#)
 - [8.1 Desligando o R Commander](#)
 - [8.2 Voltando ao R Commander](#)
 - [8.3 Salvando e lendo scripts](#)
- ▼ [9 Manipulação da Planilha](#)
 - [9.1 Obtendo subconjunto dos dados](#)
 - [9.2 Escolhendo colunas](#)
 - ▼ [9.3 Escolhendo linhas](#)
 - [9.3.1 Variáveis numéricas](#)
 - [9.3.2 Variáveis textuais](#)
 - [9.3.3 Mais exemplos?](#)
 - [9.3.4 Veja as variáveis do data set](#)
 - [9.3.5 Removendo pacientes da planilha](#)
 - [9.3.6 Recodificando variáveis](#)
 - [9.3.7 Eliminando variáveis](#)
 - [9.3.8 Reordenando as colunas](#)
- ▼ [10 Para saber mais](#)
 - [10.1 Para iniciantes](#)
 - [10.2 Uso avançado](#)
- [11 Referências](#)

1 O que você deve saber antes de usar o R Commander

Para se instalar o **R Commander**, seu computador também precisa do R instalado.

Como o R Commander é disponível para diversas plataformas, presume-se que o estudante saiba utilizar seu computador. É necessário saber como baixar um arquivo da Internet (download), criar uma pasta (diretório) em seu computador, e encontrá-lo posteriormente. No tutorial de Minitab existe uma descrição das tarefas básicas que talvez lhe interesse observar.

Esta série de tutoriais foi inicialmente ilustrada com o R Commander versão 1.8-1, R na versão 2.14.1, rodando sob Linux (especificamente, Ubuntu, versão 12.04 LTS, em inglês). No entanto, espera-se que os usuários também explorem outras plataformas e versões de programas. Deverão ser capazes de utilizar

outros sistemas operacionais, com outras configurações de tela e/ou versões diferentes do R Commander e do R, e portanto haverá telas com aspectos ligeiramente diferentes, mas sem perturbar o entendimento do tutorial.

Caso encontre alguma discrepância maior, não hesite em nos alertar sobre as eventuais diferenças dos sistemas ou versões.

Para que todos usem o R Commander em **português** ou **inglês**, ao seu gosto, padronizamos usar os comandos nas duas versões na medida do possível. Por exemplo:



Estatísticas -> Médias -> Teste t para amostras independentes



Statistics -> Means -> Independent samples t-test

e, dentro do texto, colocaremos em **português** (e **inglês entre parênteses**).

2 Por que R Commander?

Recomendamos fortemente o aprendizado adicional do R Commander porque lhes permitirá aprofundar nos recursos de análises estatísticas complexas em R, pois pela escolha por meio de interface gráfica o R Commander mostrará os comandos em R no console, sem a necessidade de se consultar manuais ou outras fontes. R Commander (Rcmdr) e seus plugins são um programa ou pacotes R de código aberto e de uso livre. Pesquisadores ou alunos familiarizados com o R Commander poderão utilizá-los em suas tarefas, não dependendo de licenças institucionais ou da compra para uso pessoal. O R Commander também mostra, enquanto opera, como são os comandos R que constrói para você. No futuro, ou precisando de usos mais avançados, você poderá aprender R, que é uma linguagem de programação forte em estatística, mas com vários outros recursos. Programar em R é outro nível, mas as possibilidades são muito mais amplas.

2.1 O R Commander:

- Apresenta organização de menus semelhantes aos de aplicativos comumente utilizados, como processadores de texto (e.g. Word, LibreOffice Writer) e planilhas eletrônicas (e.g., Excel, Gnumeric, LibreOffice Calc), facilitando seu aprendizado; e
- Incorpora uma série de recursos de manipulação de dados, de confecção de gráfico e de rotinas de análise estatística mais importantes que irão satisfazer a maioria das necessidades de estudantes e pesquisadores na área biomédica.

Pretende-se introduzir o [R Commander \(http://www.rcommander.com/\)](http://www.rcommander.com/), comumente abreviado para **Rcmdr**, mostrando algumas rotinas de manipulação de arquivos que serão úteis para resolução de problemas práticos de análise de dados em medicina.

Não é possível esgotar todas as possibilidades e recursos oferecidos pelo R em poucas aulas, mas esperamos poder mostrar algumas novas perspectivas de uso prático de um pacote estatístico que também tem os principais recursos de um programa de planilha eletrônica, semelhante ao **Excel** (Microsoft), ou **OpenOffice/LibreOffice** (este um programa livre).

Para aqueles mais curiosos, sugere-se a leitura do [manual disponível na página dos autores \(http://socserv.mcmaster.ca/jfox/Misc/Rcmdr\)](http://socserv.mcmaster.ca/jfox/Misc/Rcmdr). Os testes estatísticos propriamente ditos serão introduzidos ao longo das aulas práticas subsequentes.

3 Abrindo o programa

3.1 No Linux

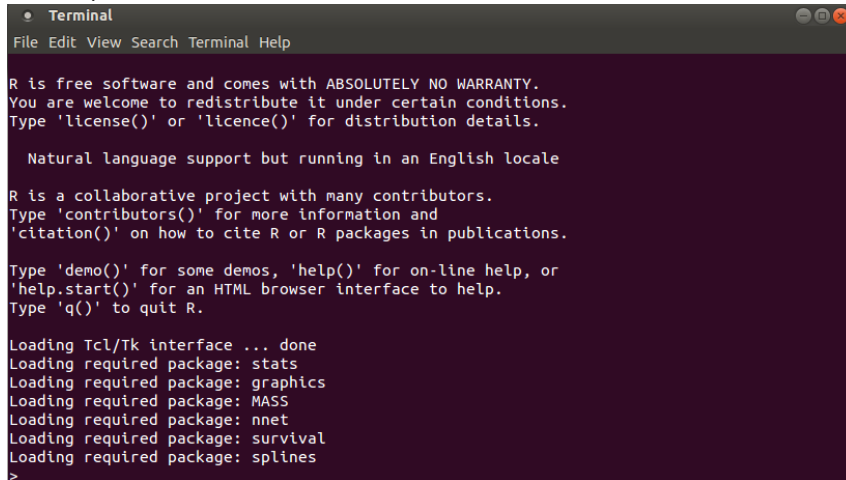
Inicie o R Commander. No Ubuntu, versão 12.04 LTS está em



???

 Applications -> Science -> R Commander

Um terminal abre-se primeiro e o R é ativado:



```
• Terminal
File Edit View Search Terminal Help

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

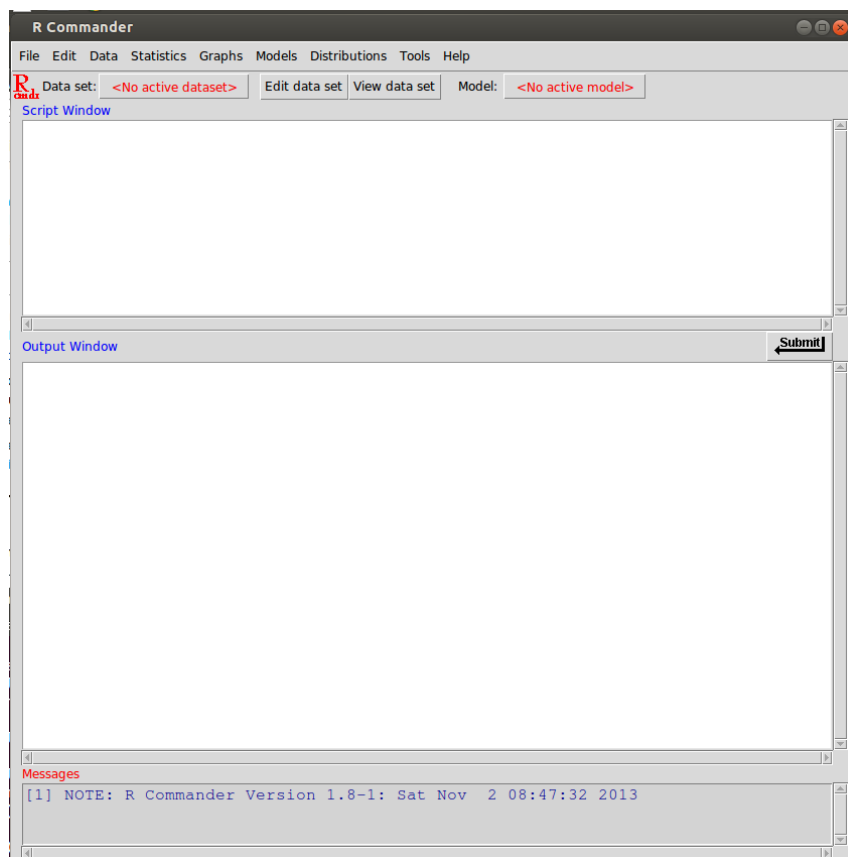
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Loading Tcl/Tk interface ... done
Loading required package: stats
Loading required package: graphics
Loading required package: MASS
Loading required package: nnet
Loading required package: survival
Loading required package: splines
>
```

em seguida aparece a interface gráfica do R Commander:



O terminal pode ser minimizado, mas não deve ser fechado. Este terminal é totalmente funcional e pode ser utilizado simultaneamente ao Rcmdr. Experimente escrever no terminal o comando `R version` após o sinal de ">" (este é o prompt do R). Deve obter algo como:

```
> version

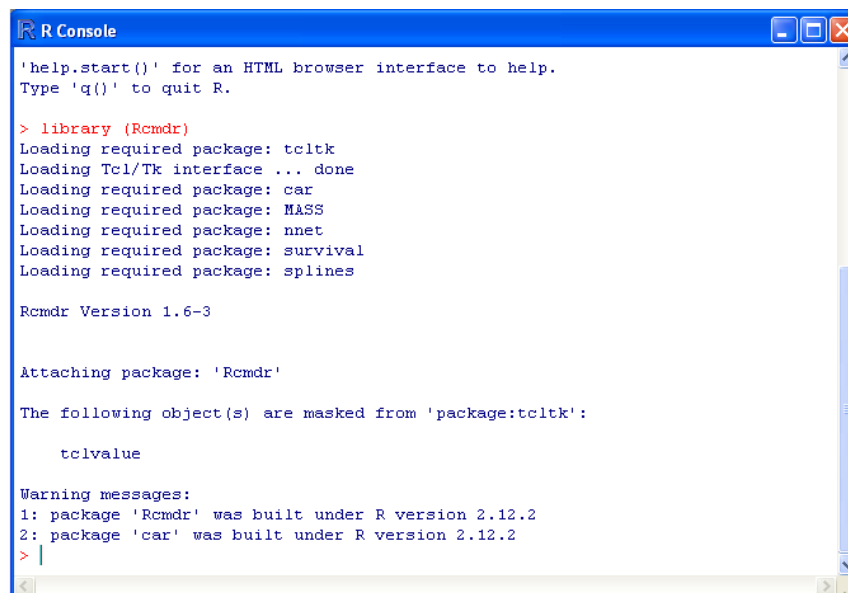
platform      i686-pc-linux-gnu
arch           i686
os             linux-gnu
system         i686, linux-gnu
status
major          2
minor          14.1
year           2011
month          12
day            22
svn rev        57956
language       R
version.string R version 2.14.1 (2011-12-22)
>
```

3.2 No Macintosh

A ser descrito por um usuário deste sistema operacional.

3.3 No Windows

Na área de trabalho do Windows encontre e utilize o acesso ao R, que abre a RGui, uma interface gráfica que facilita o uso do R e contém um terminal.



```
R R Console

'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(Rcmdr)
Loading required package: tcltk
Loading Tcl/Tk interface ... done
Loading required package: car
Loading required package: MASS
Loading required package: nnet
Loading required package: survival
Loading required package: splines

Rcmdr Version 1.6-3

Attaching package: 'Rcmdr'

The following object(s) are masked from 'package:tcltk':

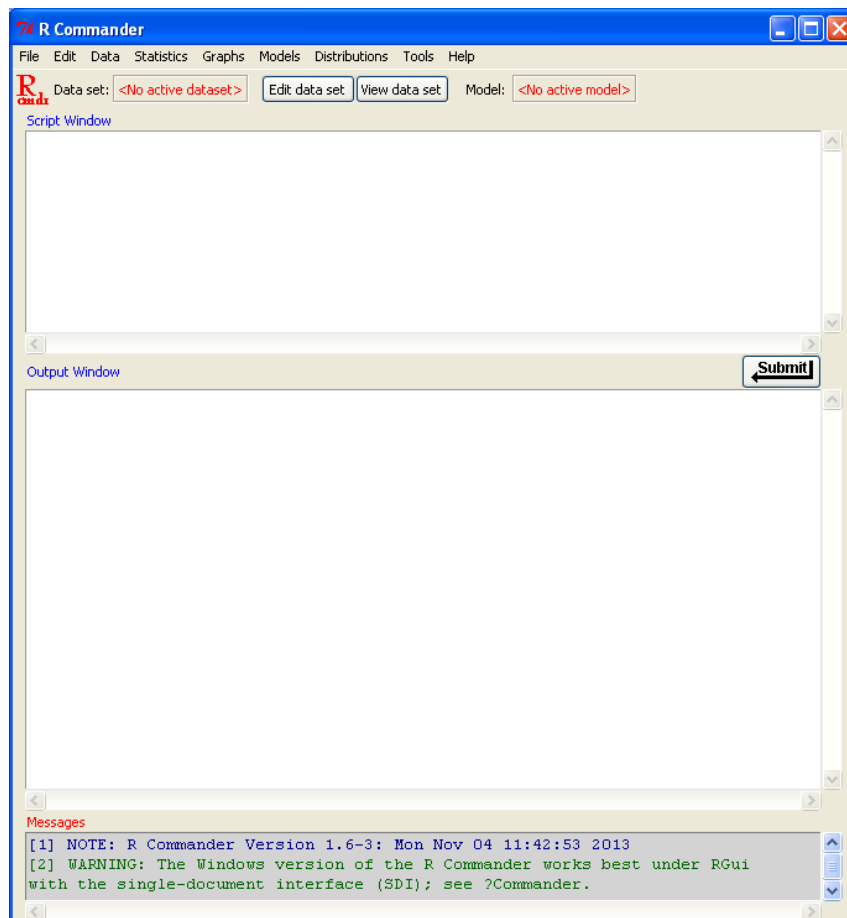
    tclvalue

Warning messages:
1: package 'Rcmdr' was built under R version 2.12.2
2: package 'car' was built under R version 2.12.2
> |
```

Neste terminal (como mostrado acima) digite o comando:

```
> library(Rcmdr)
```

A janela do R Commander deve abrir:



Da mesma forma que acontece em outras plataformas, o terminal é funcional (pode testar com o comando `version`, como descrito para a implementação em Linux). O terminal pode ser minimizado, mas não deve ser fechado.

4 Manipulação dos dados

4.1 Abrindo uma planilha de dados existente

Os dados que utilizaremos aqui estão no arquivo [exames1.txt](https://github.com/mdr5728r/2020/blob/master/Rcmdr_tutorial/Rcmdr_Intro/exames1.txt?raw=true) (https://github.com/mdr5728r/2020/blob/master/Rcmdr_tutorial/Rcmdr_Intro/exames1.txt?raw=true), que você deve baixar para seu computador (dependendo da configuração de seu computador e navegador, experimente usar o botão da direita do mouse para fazer este download). Este arquivo de dados é apresentado em texto puro, delimitado por "tabs", e precisa ser importado pelo Rcmdr, como descrito abaixo.

Este arquivo de dados tem o mesmo conteúdo da planilha Excel [exames1.xls](https://github.com/mdr5728r/2020/blob/master/Rcmdr_tutorial/Rcmdr_Intro/exames1.xls?raw=true) (https://github.com/mdr5728r/2020/blob/master/Rcmdr_tutorial/Rcmdr_Intro/exames1.xls?raw=true). O R Commander nas versões atualmente (2014) disponíveis para Linux e Macintosh não abrem planilhas Excel diretamente, mas (curiosamente) abrem arquivos do Minitab. Caso você tenha dados em Excel, pode gerar um arquivo texto, como fizemos.

Arquivos textos com delimitadores são a forma mais universal de transpor dados entre programas diferentes e, portanto, vale a pena saber lidar com eles. Ouvi dizer que na versão mais nova em Windows o Rcmdr importa planilhas Excel. Com a versão 2016 no Windows 10 foi possível importar planilha Excel. O procedimento é o mesmo: Dados -> Importar arquivo de dados... a última opção é arquivo do Excel.

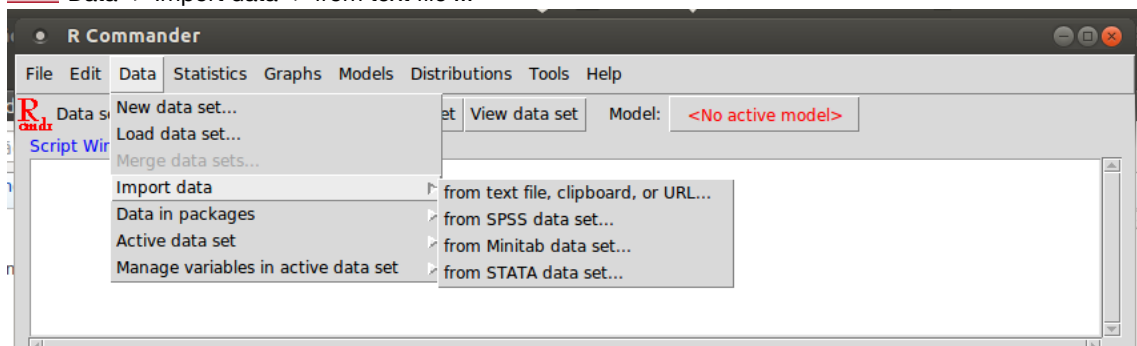
Acesse, no Rcmdr:



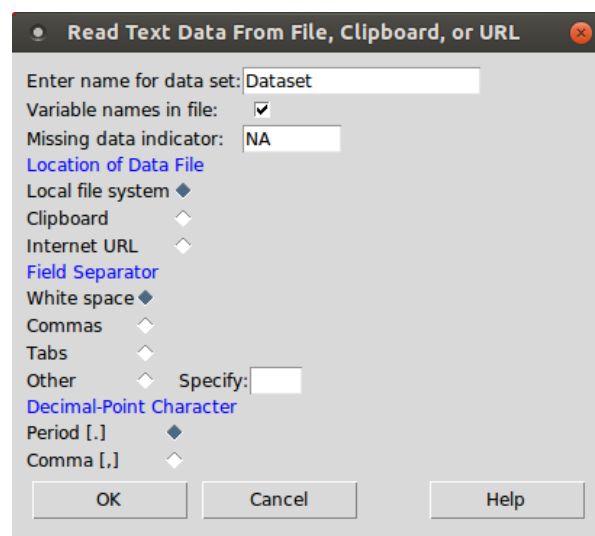
Dados -> Importar conjunto de dados -> de arquivo de texto ...



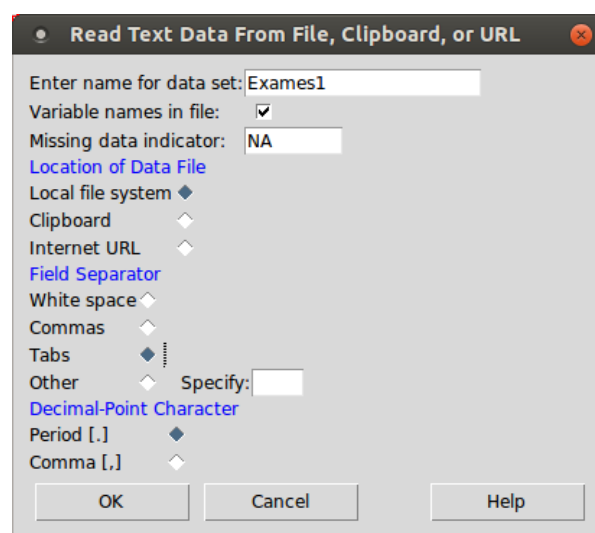
Data -> Import data -> from text file ...



A caixa de diálogo que aparece:



precisa ser configurada de acordo com o formato do arquivo texto:



Note as seguintes opções que fizemos:

- **Defina o nome do conjunto de dados (Enter name for data set):** Exames1 (apenas para que lembremos de qual arquivo os dados vieram);

- **Nome das variáveis no arquivo (Variable names in file):** selecionado (porque, neste arquivo, a primeira linha contém os nomes das variáveis, o que é sempre recomendado);
- **Separador de campos (Field separator):** Tabs (porque foi o delimitador que utilizamos ao gerar exames1.txt);
- **Separador de decimais (Decimal-point character):** Period (no caso, foi como exportamos os dados numéricos).

Para importar os dados: clique **[Ok]** e aponte o arquivo de dados que você baixou. Repare o comando R ecoado na **Output (Output Windows)**: este é o comando que você precisaria escrever em R, caso usasse a linha de comando, para importar este arquivo. É deste trabalho que o Rcmdr o poupa:

```
Exames1 <- read.table("/home/usuario/Aulas/MDR5728/Rcmdr_tutorial/Rcmdr_Intro/MPT0164_exames1.txt", header=TRUE, sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
```

Se tudo estiver correto, na tela de mensagens deve aparecer:



NOTA: Os dados Exames1 tem 705 linhas e 15 colunas.



NOTE: The dataset Exames1 has 705 rows and 15 columns.

4.2 Números: vírgulas ou pontos

O uso de vírgula ou de ponto como separador decimal dá problema entre computadores utilizando português ou inglês. Particularmente (opinião de Paulo Silveira), prefiro desrespeitar a gramática para trabalhar e **sempre** utilizar o ponto (trocando para vírgulas somente no que preciso publicar em português). Caso sua opção seja outra, saiba que isto depende da configuração do computador utilizado no momento da exportação dos dados para o arquivo **txt** e do que você selecionar no R Commander no momento de importar os dados para o data set. Verifique como está o seu computador. Veja opinião do Prof. Koichi Sameshima, algo divergente em Uma Pausa: cuidado no uso da vírgula decimal.

5 Visualizando os dados

Após importar os dados, um conjunto de dados (data set) com o nome que você escolheu está ativo. Você pode conferir se a importação dos dados deu certo clicando em **Ver conjunto de dados (View data set)** para abrir esta janela:

	PACIENTE	DN	IDADE	SEXO	CLIN	TPCLIN	D
1	1	9/10/50	52	M	ACMG	A	
2	2	7/4/52	50	M	ALMR	A	
3	3	2/4/56	47	F	ANCL	A	SEM DADOS CLIN
4	4	12/10/30	72	F	A2MV	A	EPILEPSIA + ANEMIA
5	5	7/3/50	52	F	ACMG	A	
6	6	6/17/60	42	M	A2MR	A	SEM DADOS CLIN
7	7	5/9/43	59	M	A2MV	A	POLIARTRITE
8	8	2/16/72	31	F	A2MR	A	CONT
9	9	7/3/39	63	M	AURG	A	
10	10	9/30/57	45	F	AGMI	A	COLICA R
11	11	3/13/58	44	M	ACMG	A	SEM DADOS CLIN
12	12	10/10/49	53	F	ALMR	A	
13	13	4/16/27	75	F	EGIN	E	SEM DADOS CLIN
14	14	3/26/40	62	M	ALMR	A	DADO CLINICO ILEG
15	15	6/4/60	42	F	TGIG	A	SEM DADOS CLIN
16	16	9/25/16	86	F	A2MV	A	SEM DADOS CLIN
17	17	2/12/82	21	F	ACMG	A	SEM DADOS CLIN
18	18	1/2/56	47	M	ACMG	A	HIPOTIREOID
19	19	12/22/25	77	F	U2CC	E	SEM DADOS CLIN
20	20	4/9/56	46	F	ACPL	A	POS-OPERAT
21	21	4/17/46	56	F	ELMR	E	AN
22	22	2/25/35	68	F	ALMR	A	
23	23	3/11/35	67	F	ADEG	A	SEM DADOS CLIN
24	24	5/29/30	72	F	EPSC	E	SEM DADOS CLIN
25	25	12/2/56	46	F	ANCL	A	SEM DADOS CLIN
26	26	7/5/52	50	F	ACMG	A	DOR ABDOM
27	27	2/2/84	19	F	AGAE	A	
28	28	9/29/53	49	F	ALMR	A	SEM DADOS CLIN
29	29	4/9/48	54	M	ACMG	A	SEM DADOS CLIN
30	30	5/6/86	16	F	AGAE	A	SEM DADOS CLIN

Nesta tela aparecem os dados importados. Há também barras de rolagens horizontal e vertical, para que você possa percorrê-los. Antes de prosseguir, identifique os elementos que compõem a janela em questão, pois é fundamental entender como os dados ficam organizados:

- note que os dados de cada paciente estão organizados em linhas, identificadas na coluna à esquerda enumeradas a partir de 1 (o número de cada linha coincide com o número de identificação do paciente, mas isto é somente coincidência);
- observe que cada coluna, denominada variável, é encimada por sua identificação, o nome da variável. É através deste identificador, importado da primeira linha do arquivo texto, que você referenciará as variáveis que desejar para executar os procedimentos adiante.

Dica: não utilize espaços ou símbolos especiais (o Rcmdr aceita "." se precisar de nomes compostos) para compor nome de identificadores; também recomenda-se iniciar os nomes das variáveis com uma letra.

Depois que verificar o sucesso da importação, sugere-se que esta janela seja fechada.

O Rcmdr trabalha com conjuntos de dados (data sets). Você pode ter mais que um conjunto de dados simultaneamente aberto em uma mesma seção do Rcmdr, mas os comandos são executados apenas com o conjunto ativo. A qualquer momento você pode alterar o conjunto ativo através de Dados -> conjunto de dados ativos (Data->Active data set) ou clicando sobre o botão que aparece contendo o nome do data set ativo.


Ao lado do botão **[Ver conjunto de dados]** (**[View data set]**) aparece **[Editar conjunto de dados]** (**[Edit data set]**), o qual obviamente permite modificar os dados. A interface de edição não é muito boa. Caso precise de alterações mais extensas, é melhor arrumar os dados em uma planilha e importar novamente. Há mais detalhes sobre este editor abaixo.

6 Salvando e recuperando dados

Este conjunto de dados pode ser salvo no formato do R Commander para usar mais tarde (assim não é necessário repetir a importação). Esta operação, bem como várias outras, encontra-se em:




Dados -> conjuntos de dados ativos -> salvar conjunto de dados ativos...

 Data->Active data set->Save active data set...

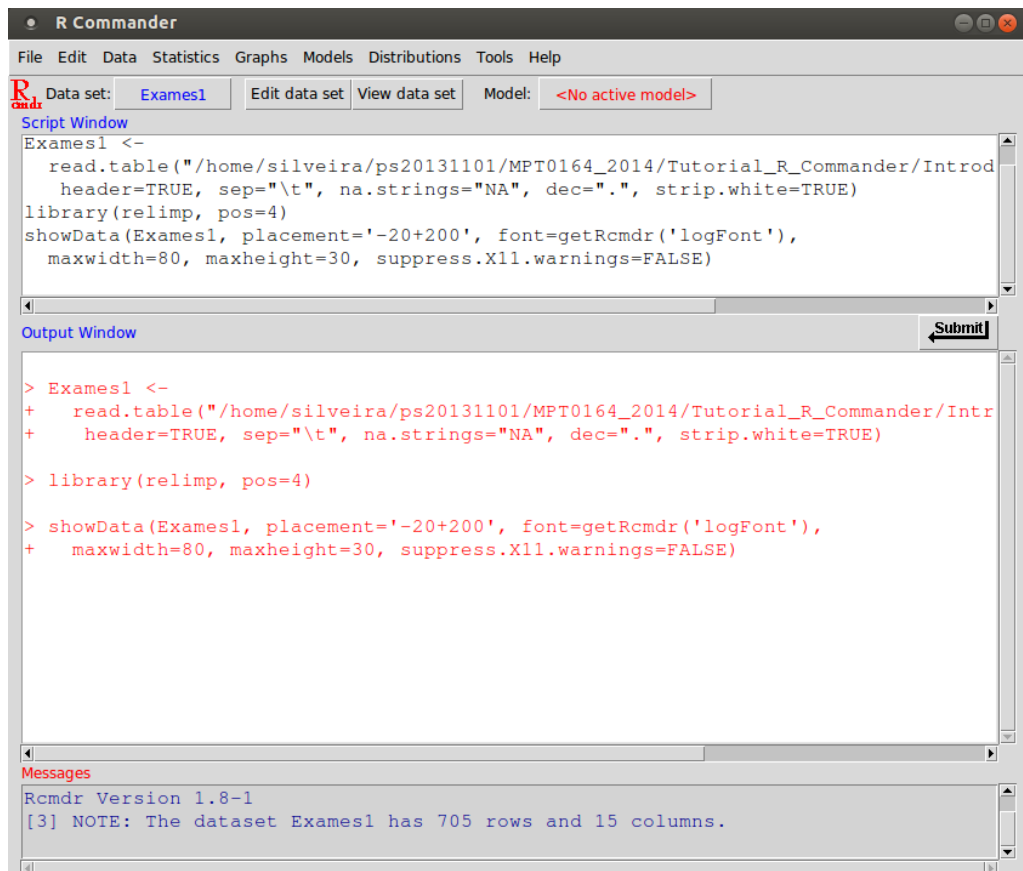
Escolha a pasta e grave o arquivo que recebe o nome do data set e a extensão Rdata: no exemplo o arquivo gerado é Exames1.Rdata. Por esta razão, procure escolher nomes com significado para seus conjuntos de dados; não os chame de dados, planilha, pacientes, ou coisas similares. Quando quiser voltar a trabalhar com o mesmo conjunto de dados, busque-o com:



 Dados -> Carregar conjunto de dados

 Data->Load data set...


7 Elementos da tela do Rcmdr


A esta altura, a janela do R Commander deve ter o seguinte aspecto:



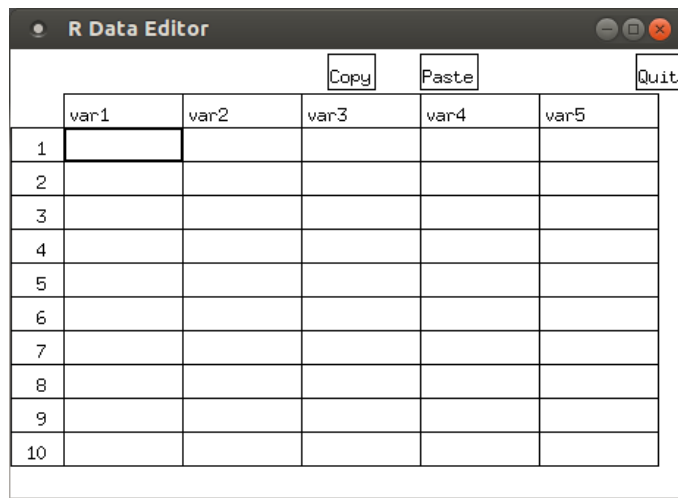
Identifique os elementos desta tela: menus de acesso aos comandos no topo, uma fileira de botões que indicam o data set ativo, permitem edição ou visualização dos dados, e uma indicação  (). Abaixo aparecem três janelas: script (*Script Window*), Output (*Output Window*) e Mensagens (*Messages*). Respectivamente mostram os comandos que você teria que executar se usasse um terminal (usando o R sem a interface gráfica do Rcmdr), o que este terminal ecoaria, e mensagens do sistema.

Preenchimento manual da planilha Vamos abrir um novo conjunto de dados usando o comando

 Dados ->Novo conjunto de dados...

 Data -> New data set... Como exercício, batize este conjunto com o nome de Colegas, pois armazenaremos dados de nomes de seus colegas, idade e cidade de nascimento. Portanto será uma planilha com três variáveis.

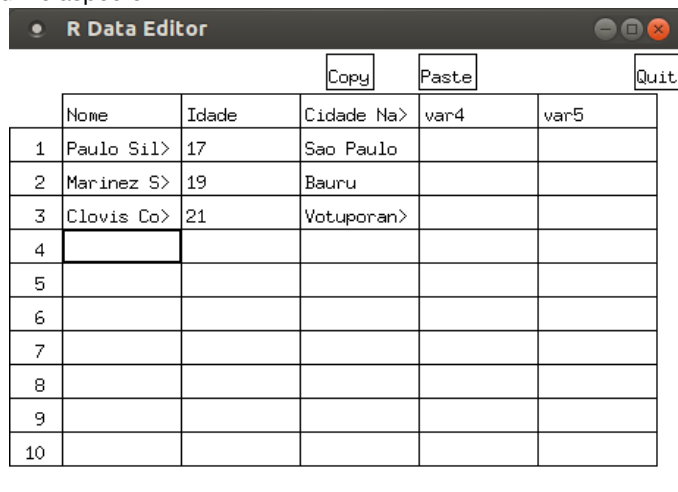
A interface é auto explicativa. Se necessário, alargue o tamanho da tela para aparecer, pelo menos, três colunas:



Renomeie as três primeiras colunas, respectivamente, como Nome, Idade e Cidade Nasc: clique sobre os nomes pré-definidos (var1, var2 ou var3), escolha nome da variável (Change name) e entre com os nomes de variáveis desejados. A seguir, entre com o seguinte pequeno conjunto de dados:

Pedro Souza	17	Sao Paulo
Marines Sanchez	19	Bauru
Clovis Concravo	21	Votuporanga

Na versão do Rcmdr que estou usando a interface não é das melhores, mas para pequenas entradas como esta é o suficiente. Entre as principais deficiências estão as colunas não se ajustam de acordo com o conteúdo e a acentuação em português não é aceita (repare em *Sao Paulo* sem ã ou *Clovis* sem ó). Entrar os dados e mudar de célula é simples: clique sobre a célula onde deseja entrar cada dado, e use as teclas direcionais (setas) para mudar para células vizinhas. Quando acabar de entrar os dados deve ver a tela com o seguinte aspecto:



Clique **Sair** ([Quit]). O conjunto de dados **Colegas** é criado já fica ativo. Pode conferir com **[Ver conjunto de dados]** ([View Data Set]).

Observação: na versão utilizada atualmente não há mais botão ([Quit]). Basta clicar em **X** no canto superior direito da janela.

Colegas			
	Nome	Idade	CidadeNasc
1	Pedro Souza	17	Sao Paulo
2	Marines Sanchez	19	Bauru
3	Clovis Concravo	21	Votuporanga

Note que o Rcmdr não admite espaço no nome das variáveis, de forma que a terceira coluna ficou denominada como **Cidade.Nasc**. Note também que **[Ver conjunto de dados] ([View Data Set])** ajusta automaticamente a largura das colunas ao conteúdo das células, facilitando a leitura.

Dica que vale para qualquer programa, não somente Rcmdr: não assuma que um programa tenha sido desenvolvido ou que está bem implementado em relação às acentuações da língua portuguesa; não as utilize antes de testar o programa de computador para dois aspectos: primeiro, apresentar consistência na comparação textual de palavras, por exemplo se galã é igual a gala e laco é igual laço; em segundo, se ordena consistentemente as palavras acentuadas e as que usam cedilhados. Um jeito seguro é não usar acentos nem cedilhas.

Não se esqueça de salvar sua conjunto de dados em um arquivo chamado **Colegas.Rdata**.

8 Saindo e retornando ao R Commander

8.1 Desligando o R Commander

Saber sair corretamente do programa é importante. O mais normal é fechar a interface gráfica e o terminal, utilizando:



Arquivo -> Sair -> Do Commander e do R



File -> Exit -> From R Commander and R

O programa perguntará se quer salvar o script e o output. Por enquanto responda **[Não] ([No])** para ambos; voltaremos a isto adiante.

8.2 Voltando ao R Commander

Nas seções anteriores você abriu dois data sets e os salvou. O primeiro, **Exames1**, foi obtido de um arquivo texto que geramos a partir de uma planilha do Excel; o segundo, **Colegas**, foi digitado por você. Respectivamente, devem ter sido salvos no formato do Rcmdr com os nomes **Exames1.Rdata** e **Colegas.Rdata**.

Reabra, portanto, o Rcmdr e utilize:



Dados -> Carregar conjunto de dados...



Data -> Load data set

aponte para leitura o arquivo **Exames1.Rdata**. A seguir, repita o procedimento para ler **Colegas.Rdata**. Repare a *Script Window* onde aparece

```
load("{aqui aparece o caminho para seu arquivo}/Exames1.RData")
load("{aqui aparece o caminho para seu arquivo}/Colegas.RData")
```

8.3 Salvando e lendo scripts

Este é um uso mais avançado e este exemplo é muito simples (mas ilustrativo).

Pois bem... utilize:



Arquivo -> salvar script como ...



File -> Save script as...

e dê o nome **Leitura_Exames_e_Colegas.R** para este script.

Feche completamente o programa (**Sair -> Do Commander e do R, Exit -> From Commander and R**) e abra novamente. Utilize



Arquivo -> Abrir arquivo c/ script



File -> Open script file...

e traga de volta o conteúdo da Janela script (Script Window) recuperando o arquivo

Leitura_Exames_e_Colegas.R. Repare que esta janela permite edição, de forma que você pode conferir e ajustar o script sempre que quiser. Quanto estiver pronto, note o botão **Submeter (Submit)** no canto inferior direito desta janela: clique nele para executar seu script. A janela mensagens (Messages) mostra:



ERRO: Nada foi selecionado.



ERROR: Nothing is selected.

Selecione, então, na Janela Script (Script Window) as duas linhas:

```
load("{aqui aparece o caminho para seu arquivo}/Exames1.RData")
load("{aqui aparece o caminho para seu arquivo}/Colegas.RData")
```

e clique em **Submeter (Submit)** novamente. Desta vez a Janela output (Output Window) reage, mostrando

```
> load("{aqui aparece o caminho para seu arquivo}/Exames1.RData")
> load("{aqui aparece o caminho para seu arquivo}/Colegas.RData")
```

Clique o botão da fileira superior onde aparece **<Não há conjunto de dados ativos>** (<No active dataset>) e escolha um dos dois *data sets* lidos. A janela Mensagens (Messages) acusará o sucesso dizendo:



NOTA: Os dados Exames1 tem 705 linhas e 15 colunas.



NOTE: The dataset Exames1 has 705 rows and 15 columns.

ou



NOTA: Os dados Colegas tem 3 linhas e 3 colunas.



NOTE: The dataset `Colegas` has 3 rows and 3 columns.

de acordo com o conjunto de dados que você deixou ativo.

A partir deste exemplo simples, sofisticue. Qualquer script que você salvar, reler e submeter repetirá uma sequência de ações. Com o que aprender adiante, poderá trabalhar no R Commander experimentando fazer alguma análise. Apague as linhas que não deram certo. faça coisas algo mais sofisticadas, como gráficos e análises estatísticas, teste com determinado conjunto de dados e, depois, adapte o script para repetir o trabalho em outros conjuntos de dados. Salve seus scripts "amadurecidos" e evitará um monte de (re)trabalho.

Similar, mas menos útil (pois não roda como um script), caso queira preservar um lembrete do que fez enquanto usava o Rcmdr, é salvar a Janela Output(Output Window) em um arquivo texto.

9 Manipulação da Planilha

Há uma quantidade de operações disponíveis nos menus do R Commander. Aqui ilustraremos algumas, que lhe permitirão:

- criar subsets
- fazer gráficos

9.1 Obtendo subconjunto dos dados

Importante: Antes de começar o exercício proposto aqui, observe que na versão xls, da qual os dados de **Exames1** vieram, a coluna de sódio sérico tem o nome de variável NA. Embora a importação ocorra sem problemas, o nome da variável entra em conflito com a palavra reservada do Rcmdr, **NA**, significando **Not Available** (confira com **[View data set]**: aparece um pequeno ponto ao lado de **NA**). É possível consertar o arquivo texto e reimportar os dados, mas a solução mais simples e rápida, agora que já importou os dados, é entrar no modo de preenchimento manual da planilha, alterar o nome da variável (por exemplo, para **SODIO** ou **Natremia** - evitando começar com **NA** em maiúsculas) e, então, salvar seu data set novamente, para preservar a modificação.

Outra maneira de fazer a mesma coisa é usar **Dados -> modificação de variáveis no conjunto de dados -> Renomear variáveis...** (**Data -> Manage variables in active data set -> Rename variables...**).

Uma vez consertado e ativo seu data set **Exames1**, utilize:

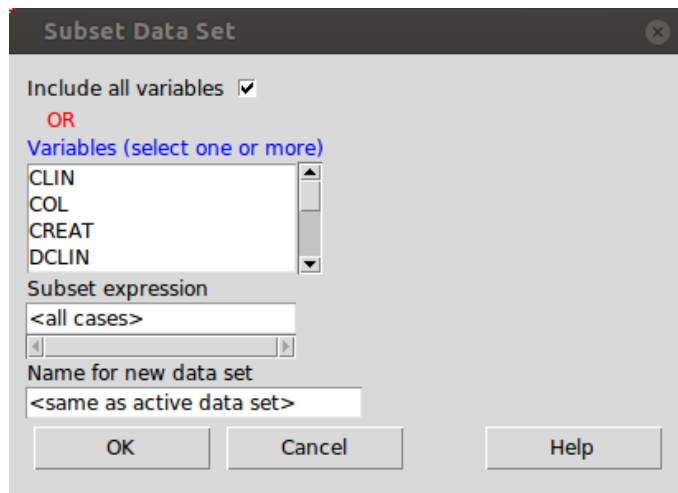


Dados -> Conjunto de dados ativos -> definir subconjunto de dados ativos...



Data -> Active data set -> Subset active data set...

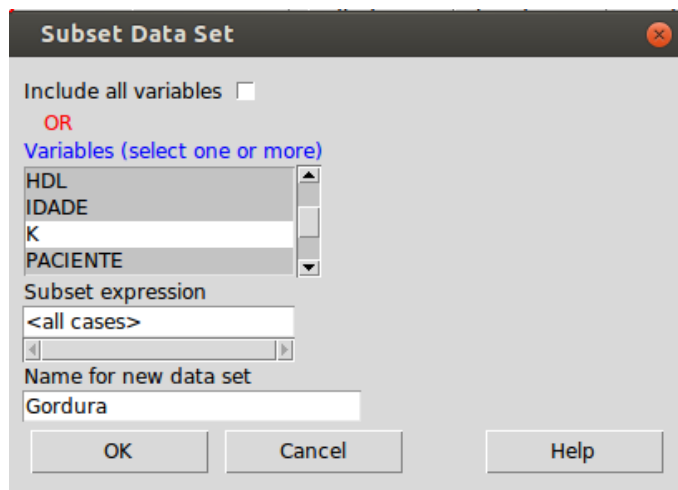
Observe a janela que se abre:



É possível "recortar" colunas ou linhas do data set.

9.2 Escolhendo colunas

Suponha que você queira apenas algumas variáveis em um novo data set, em um estudo que somente lhe interessa estudar colesterol (HDL e COL) e triglicérides (TRIG); para não perder a referência dos dados do paciente guardaremos também a identificação do PACIENTE, IDADE, SEXO. Basta desmarcar "Incluir todas as variáveis" ("Include all variables") e, segurando a tecla CTRL, apontar as variáveis de interesse, que aparecem em ordem alfabética: COL, HDL, IDADE, PACIENTE, SEXO e TRIG. Deixe Expressão (Subset expression) como está, () e escolha um nome para o novo data set (e.g. Gordura).



Clique **[OK]**. Na **Janela de Script (Script Window)** aparece o comando:

```
Gordura <- subset(Exames1, select=c(COL,HDL,IDADE,PACIENTE,SEXO,TRIG))
```

Na Output Window ecoa-se a execução do comando R:

Gordura <- subset(Exames1, select=c(COL,HDL,IDADE,PACIENTE,SEXO,TRIG)) Na janela mensagens (Message Window) o sistema notifica:



NOTA: os dados Gordura tem 705 linhas e 6 colunas.



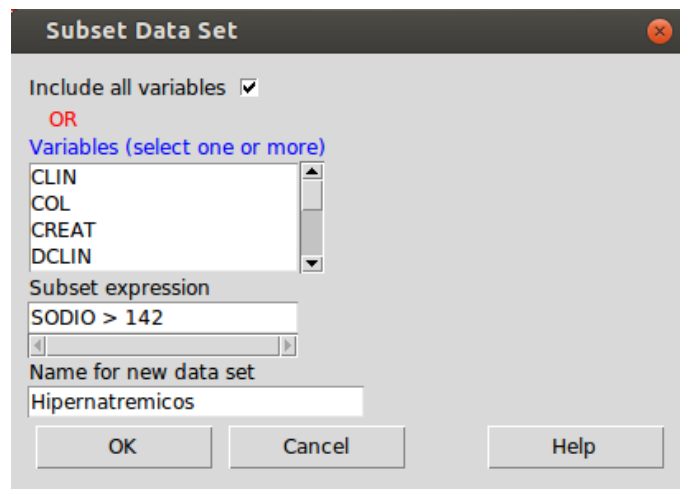
NOTE: The dataset Gordura has 705 rows and 6 columns.

o que confere, pois todos os pacientes foram transferidos (mesmo número de linhas do conjunto original), mas com apenas 6 colunas. Repare que o conjunto de dados ativo agora é Gordura; como sempre, você pode conferir com **[Ver conjunto de dados]** (**[View data set]**) e salvar seu conjunto de dados em um arquivo **Gordura.Rdata** para uso futuro.

9.3 Escolhendo linhas

9.3.1 Variáveis numéricas

Torne o data set **Exames1** novamente ativo e repita o processo para fazer um Subconjunto (subset). Podemos, por exemplo, separar os pacientes com hipernatremia (sódio sérico acima de 142 mEq/L). Desta vez vamos deixar assinalado **Incluir todas as variáveis** (**Include all variables**), portanto não há sentido em apontar nada no quadro de variáveis, e colocaremos em **Expressão** (**Subset expression**) a expressão **SODIO > 142**. Como sempre, escolheremos um nome para o novo data set que seja significativo (e.g. Hipernatremicos).



Ao clicar [OK] acontecerá algo imprevisto: uma mensagem de alerta (em verde) na Janela de mensagem (Message Window):



NOTA: Aviso em Ops.factor(SODIO, 142) : > não faz sentido para fatores



WARNING: Warning in Ops.factor(SODIO, 142) : > not meaningful for factors

Por que? Esta coluna de **SODIO**, que já tínhamos renomeado, ainda tem problemas. Neste caso, volte a ativar o conjunto de dados Exames1, entre no editor de dados clicando **[Editar conjunto de dados]** (**[Edit data set]**) e clique sobre o nome da variável **SODIO**. Há um asterisco em frente da palavra Character (character)

	SODIO	K	HDL
13	SODIO		43
13	Real		37
13	* Character		32
14	Change Name		62

compare com a coluna do potássio sérico:

	K	UNI	COL
5.1	K		84
4.1	* Real		23
4.9	Character		63
4.3	Change Name		14
4.8		55	267

que é *Real*. Respectivamente significa que **SODIO**, por alguma razão, está sendo tratada pelo Rcmdr como texto, enquanto K é numérica. Para corrigir, basta indicar que **SODIO** é *Real*, sair do editor e repetir a geração do sub set chamado **Hipernatremicos**. Desta vez, na Mensagens (Message Window), aparece:



NOTA: Os dados Hipernatremicos tem 25 linhas e 15 colunas.



NOTE: The dataset Hipernatremicos has 25 rows and 15 columns.

Podemos isolar uma "fatia" dos pacientes, por exemplo, os normonatremicos (com natremia igual ou superior a 120 e igual ou inferior a 142 mEq/L). Torne **Exames1** ativo e repita o procedimento, desta vez usando ??? (**Subset expression**) como **SODIO >= 120 & SODIO <= 142** (onde **&** é a sintaxe R para operador lógico **AND**). A resposta deve ser:



NOTA: OS dados Normonatremicos tem 678 linhas e 15 colunas.



NOTE: The dataset Normonatremicos has 678 rows and 15 columns. Note que $678 + 25 = 703$. Os outros dois pacientes devem ser os Hiponatremicos. Experimente criar um subset usando a **Expressão** (**Subset expression**) **SODIO < 120**. Obterá:



NOTA: Os dados Hiponatremicos tem 1 linhas e 15 colunas.



NOTE: The dataset Hiponatremicos has 1 rows and 15 columns. Faltou um paciente. Estranho, não? Reative Exames1 e verifique a coluna SODIO: o paciente da linha 121 tem o valor de SODIO faltando: ??? (NA).

Que tal verificar as demais colunas, se estão registradas como numéricas e textuais como deviam ser? Uma sugestão: dê uma olhada em GLIC na linha 144, que tem um * (aceitável para uma variável contendo texto), transforme GLIC em Real e olhe a linha 144 novamente (você deve aprender algo sobre o Rcmdr com isto).

9.3.2 Variáveis textuais

Dados textuais (i.e., character) também podem ser utilizados para separar um subset. Suponha que eu queira isolar as mulheres em um subset **Exames1_mulheres**. É o mesmo procedimento (lembre-se de tornar Exames1 novamente ativo), utilizando a **Expressão (Subset expression)** como **SEXO == "F"** (note o uso de dois sinais de igual, sintaxe do R para comparações lógicas pois o sinal de igual simples serve para atribuições, e F entre aspas porque é uma string), obtendo.



NOTA: Os dados Exames1_mulheres tem 448 linhas e 15 colunas. Flag of the United States.svg

NOTE: The dataset Exames1_mulheres has 448 rows and 15 columns.

9.3.3 Mais exemplos?

Enquanto executava as operações acima, você prestou atenção à Script (Script Window)? Se sim, deve ter visto os comandos:

```
Gordura <- subset(Exames1, select=c(COL,HDL,IDADE,PACIENTE,SEXO,TRIG))
Hipernatremicos <- subset(Exames1, subset=SODIO > 142)
Normonatremicos <- subset(Exames1, subset=SODIO >= 120 & SODIO <= 142)
Hiponatremicos <- subset(Exames1, subset=SODIO < 120)
Exames1_mulheres <- subset(Exames1, subset=SEXO == "F")
```

O que aparece após **subset=** são expressões R válidas. Em qualquer site que dê exemplos encontrará outros operadores (como |, barra vertical, que significa **or**). Poderá, com isto, descobrir como construir expressões mais complexas para gerar o subset que precisar.

Veja, por exemplo, [o link \(http://rprogramming.net/subset-data-in-r/\)](http://rprogramming.net/subset-data-in-r/).

Uma pequena lista dos operadores básicos[1], úteis para construir as expressões, é

Operadores para expressões em R

Operador	Significado
==	igual
!=	diferente
&	AND
	OR
is.na(nome_da_variável)	inclui os valores não disponíveis (NA) da variável
!is.na(nome_da_variável)	exclui os valores não disponíveis (NA) da variável
>	maior que
<	menor que
>=	maior ou igual que
<=	menor ou igual que

Parênteses também podem ser usados para a precedência de expressões. Por exemplo (absurdo, mas somente para exercitar), se quisermos um *subset* ou subconjunto das mulheres hipernatrêmicas e dos homens normonatrêmicos, poderíamos usar a expressão

```
(SODIO > 142 & SEXO == "F") | ((SODIO >= 120 & SODIO <=142) & SEXO == "M")
```

garantindo, desta forma, que todos os **AND** e **OR** aconteçam na ordem desejada. Se o executar obterá:

```
Exemplo_complicado <- subset(Exames1, subset=(SODIO > 142 & SEXO == "F")  
| ((SODIO >= 120 & SODIO <=142) & SEXO == "M"))
```

e



NOTA: Os dados Exemplo_complicado tem 257 linhas e 15 colunas.



NOTE: The dataset Exemplo_complicado has 257 rows and 15 columns.

Use Ver conjunto de dados ([View data set]) e confira; procure as poucas mulheres filtradas e verá que são as hipernatrêmicas.

Outras operações com o subset ativo Dois menus do Rcmdr permitem várias manipulações que podem ser úteis:



Dados -> conjunto de dados ativo



Data -> Active data set

e



Dados -> Modificação de variáveis no conjunto de dados ativos



Data -> Manage variables in active data set

Seus nomes são o que sugerem, de forma que sugerimos que os explore por sua conta. Aqui daremos apenas alguns exemplos.

Note: embora a planilha no Rcmdr tenha uma organização matricial em colunas e linhas, ela tem uma diferença fundamental em relação às planilhas do Excel, Gnumeric, OpenOffice, LibreOffice e similares: suas células comportam valores, mas não fórmulas. Devido a isso, as operações são realizadas com os conteúdos de colunas e o resultado armazenado em uma coluna escolhida na operação.

9.3.4 Veja as variáveis do data set

Use



Dados -> Conjunto de dados ativos -> variáveis em conjunto de dados ativo



Data -> Active data set -> Variables in active data set

Como sempre, veja o comando R na Script Window e a resposta na Output Window

9.3.5 Removendo pacientes da planilha

Em algumas situações a análise é perturbada por dados incorretamente preenchidos ou não preenchidos. Respectivamente experimente



Dados -> conjunto de dados ativos -> Remova linha(s) do conjunto de dados ativos...



Data -> Active data set -> Remove rows from active data set...

e



Dados -> Conjunto de dados ativos -> Remover observações com dados faltantes...



Data -> Active data set -> Remove cases with missing data...

Em ambos você tem a opção de fazer a operação no próprio data set em uso, o que pode não ser conveniente caso precise voltar atrás, ou criar um novo, que você pode salvar para uso futuro.

9.3.6 Recodificando variáveis

Com



Dados -> Modificação de variáveis de conjunto de dados ativo -> Recodificar variáveis...



Data -> Manage variables in active data set -> Recode variables...

você pode criar outras variáveis. Por exemplo, caso queiramos classificar as pessoas de acordo com suas idades, experimente:

1 selecionar a variável **IDADE**; 2 escolher o nome da nova variável, e.g. IDADE_classificacao; 3 preencher **Definições p/ recodificação (Enter recode directives)** com um critério de sua escolha, e.g.:

```
1:11 = "crianca"
12:17 = "adolescente"
18:25 = "adulto jovem"
26:64 = "adulto"
65:1000 = "idoso"
```

Observe que nova coluna foi adicionada ao seu *data set*.

Existem outras formas de definir as diretivas --- consulte **[Ajuda]** (**[Help]**) antes de clicar **[OK]**.

Experimente "esquecer" um pedaço das idades, por exemplo, parando os adultos aos 55 anos de idade, e observe como ficam os conteúdos da variável criada.

9.3.7 Eliminando variáveis

Com



Dados -> Modificação de variáveis de conjunto de dados ativo -> Apagar variáveis do conjunto de dados...



Data -> Manage variables in active data set -> Delete variables from data set...

você pode retirar colunas inteiras de seus dados, seja porque não quer determinadas colunas para sua análise, ou porque criou uma coluna erradamente. Experimente remover a **IDADE_classificacao** que acabou de criar.

9.3.8 Reordenando as colunas

Quando se utiliza o recurso "Definir sub-conjunto de dados ativo",



Dados -> Conjunto de dados ativos -> definir subconjunto de dados ativos...



Data -> Active data set -> Subset active data set...

as colunas do novo conjunto de dados serão automaticamente ordenadas na ordem alfabética das variáveis selecionadas.

No exemplo do *data set* **Gordura**, a tabela fica da seguinte forma:

	COL	HDL	IDADE	PACIENTE	SEXO	TRIG
1	184	43	52	1	M	151
2	223	37	50	2	M	119
3	163	32	47	3	F	46
4	214	62	72	4	F	150
5	267	55	52	5	F	139
6	191	32	42	6	M	202
7	129	30	59	7	M	103
8	130	33	31	8	F	71
9	227	53	63	9	M	124
10	147	52	45	10	F	81
11	216	32	44	11	M	426
12	259	38	53	12	F	338
13	143	38	75	13	F	79
14	198	31	62	14	M	114
15	217	51	42	15	F	148
16	269	64	86	16	F	187
17	205	70	21	17	F	59
18	202	38	47	18	M	153
19	36	9	77	19	F	34
20	202	59	46	20	F	98
21	469	48	56	21	F	406
22	193	43	68	22	F	106
23	226	56	67	23	F	81
24	165	48	72	24	F	109
25	185	62	46	25	F	128
26	330	55	50	26	F	189
27	141	51	19	27	F	42
28	144	45	49	28	F	86
29	263	45	54	29	M	54
30	171	45	16	30	F	123

Lembrando que para criar essa tabela o comando utilizado foi:

```
Gordura <- subset(Exames, select=c(COL,HDL,IDADE,PACIENTE,SEXO,TRIG))
library(relimp, pos=4)
```

Observe que a ordem das colunas segue a ordem das variáveis no comando (COL,HDL,IDADE,PACIENTE,SEXO,TRIG). Assim, para reordenar as colunas na tabela, basta alterar a ordem das variáveis no comando. Por exemplo:

```
Gordura <- subset(Exames, select=c(PACIENTE,SEXO,IDADE,COL,HDL,TRIG))
library(relimp, pos=4)
```

Resultando numa apresentação mais ordenada dos dados:

	PACIENTE	SEXO	IDADE	COL	HDL	TRIG
1	1	M	52	184	43	151
2	2	M	50	223	37	119
3	3	F	47	163	32	46
4	4	F	72	214	62	150
5	5	F	52	267	55	139
6	6	M	42	191	32	202
7	7	M	59	129	30	103
8	8	F	31	130	33	71
9	9	M	63	227	53	124
10	10	F	45	147	52	81
11	11	M	44	216	32	426
12	12	F	53	259	38	338
13	13	F	75	143	38	79
14	14	M	62	198	31	114
15	15	F	42	217	51	148
16	16	F	86	269	64	187
17	17	F	21	205	70	59
18	18	M	47	202	38	153
19	19	F	77	36	9	34
20	20	F	46	202	59	98
21	21	F	56	469	48	406
22	22	F	68	193	43	106
23	23	F	67	226	56	81
24	24	F	72	165	48	109
25	25	F	46	185	62	128
26	26	F	50	330	55	189
27	27	F	19	141	51	42
28	28	F	49	144	45	86
29	29	M	54	263	45	54
30	30	F	16	171	45	123

10 Para saber mais

Há muitos tutoriais e dicas de uso na Web. Quem encontrar e testar outros, pode no enviar sugestões colocando um breve comentário sobre sua experiência.

10.1 Para iniciantes

Getting Started with R Commander² (<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf>):

Trata-se de um mini-manual de 28 páginas disponível em PDF, escrito por um dos autores do Rcmdr. É indicado para quem vai usar o R Commander pela primeira vez ou precisa de uma recordação rápida sobre seu uso. Começa desde o início básico, explicando todos os elementos da interface do R Commander, e conclui com algumas manipulações básicas dos dados.

10.2 Uso avançado

Nenhum tutorial na lista (incluir o nome aqui) (incluir seu comentário aqui)

11 Referências

¹ David McCallum. R Commander HowTo (PDF). (http://www.junction404.com/R_Commander.pdf)

² Fox J. Getting Started with R Commander (<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf>)

