

# R Commander: um R com menus

Felipe Rafael Ribeiro Melo

2º Educa-SER | 27 de maio de 2022

# Motivação

- Apesar das diversas vantagens oferecidas pelo *software* R, sua interface pouco amigável, apta a receber apenas linhas de comando, causa desmotivação em usuários iniciantes, sobretudo aos que tem dificuldade com lógica de programação.
- Entretanto, o R possui uma grande quantidade de ***pacotes***, e alguns destes pacotes fornecem interfaces gráficas nas quais é possível realizar análise de dados sem a necessidade de inserir linhas de comando, tais como os pacotes:
  - Rcmdr;
  - esquisse;
  - radiant;
  - MLstudio.

# O pacote Rcmdr

- Rcmdr (uma forma abreviada de R Commander) é um pacote disponível para o *software* R que possibilita a execução de várias funcionalidades voltadas ao tratamento e à análise de conjuntos de dados por meio de menus, sem a necessidade de digitar linhas de comando, fornecendo uma interface “point & click”.
- Para usufruir do que um pacote do R fornece, são necessárias duas etapas: *instalação* do pacote e *carregamento* do pacote.
- A instalação de um pacote precisa ser feita *uma única vez*; já o carregamento *deve sempre ser feito a cada sessão do R que se deseje utilizá-lo*.

# Instalação do pacote Rcmdr

- No R, acesse o menu *Pacotes* → *Instalar pacotes(s)*.
- Escolha algum *mirror* (sugestão: 0-Cloud) e dê OK.
- Na janela seguinte:
  - **para instalar apenas o pacote Rcmdr**: selecione Rcmdr e dê OK.

OU

- **para instalar o pacote Rcmdr e seus “plug-ins”**: selecione todos os pacotes que começam com “Rcmdr” (incluindo o próprio pacote Rcmdr) e dê OK - este processo de instalação é mais demorado.


# Instalação via linha de comando

- A instalação de qualquer pacote também pode ser realizada via linha de comando.
- O caminho de instalar apenas o Rcmdr (sem seus “plug-ins”) exposto no *slide* anterior pode ser substituído simplesmente por digitar a linha de comando

```
install.packages("Rcmdr")
```

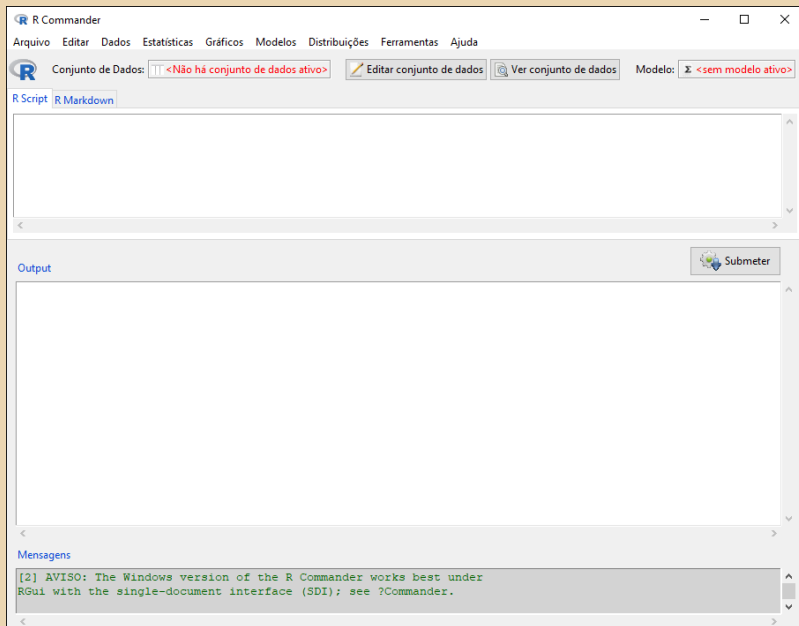
e apertar  em seguida.

# Carregando o pacote Rcmdr

- Conforme já mencionado, a instalação do pacote Rcmdr, tal como de qualquer outro pacote, só precisa ser feita uma vez.
- Para ter acesso a interface que o pacote Rcmdr fornece, é necessário **carregá-lo**. Para tal, basta abrir o R, digitar uma das duas linhas abaixo e apertar  em seguida:

```
require(Rcmdr)  
library(Rcmdr)
```

- Na primeira vez que o pacote Rcmdr for carregado, pode surgir uma pequena janela com opções *Sim* e *Não*. Clique em *Sim* e, na janela seguinte, em OK.



# E se a janela *R Commander* for fechada acidentalmente?

Para reabri-la, será necessário abrir uma nova sessão do R e carregar novamente o pacote Rcmdr, por meio da linha de comando

```
require(Rcmdr)
```

ou da linha de comando

```
library(Rcmdr)
```



A janela *R Commander* é dividida em três janelas:

- ***R Script***: Nesta janela são impressas as linhas de comando de tudo que é executado via menus. A última linha de comando nesta janela se refere ao último comando executado, a penúltima linha de comando se refere ao penúltimo comando executado, e assim sucessivamente.
- ***Output***: Esta é a janela de saída, na qual são exibidos os resultados de alguns comandos executados.
- ***Mensagens***: Informações que o sistema julga relevante compartilhar com o usuário, incluindo alertas.

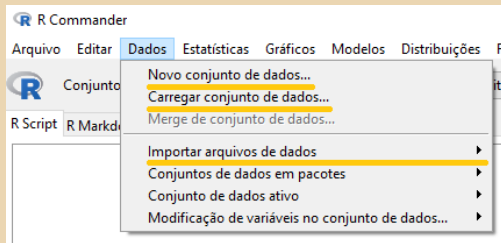
# Os três principais menus para quem está começando no R Commander:

- *Dados*, para tudo que se refere à geração, importação, carregamento e/ou edição de conjunto de dados;
- *Gráficos*, para a geração de visualizações gráficas;
- *Estatísticas*, para a geração de resultados que não sejam gráficos, como distribuições de frequências, tabelas de contingência, resumos numéricos, intervalos de confiança e testes de hipóteses.

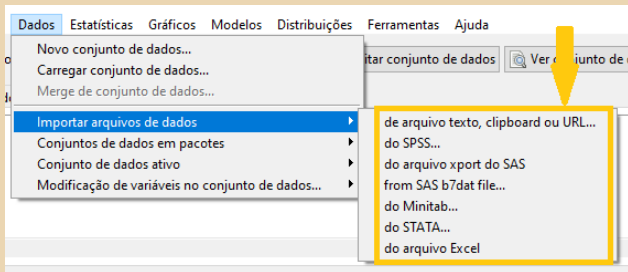
# Inserindo conjunto de dados

Há três formas de inserir conjunto de dados via R Commander.

- 1 ***Novo*** conjunto de dados: **digitar** diretamente no R.
- 2 ***Carregar*** conjunto de dados: para quando o conjunto de dados de interesse está **no formato .RData**.
- 3 ***Importar*** arquivo de dados: para quando o conjunto de dados de interesse está **em outros formatos**.



- Como é comum lidarmos com dados secundários em formatos diferentes de .RData, frequentemente os dados são inseridos via *importação*.



- Após a importação, o conjunto de dados pode ser facilmente salvo no formato .RData e, como isso, ele pode ser recuperado em qualquer sessão futura por meio da opção de *carregamento*, sem a necessidade de importá-lo novamente.

# Importando conjunto de dados em .xls/.xlsx

- Usaremos aqui o conjunto de dados presente no arquivo `dataset_minicurso.xlsx`, disponível em [http://gae.uniriotec.br/7/educaser/dataset\\_minicurso.xlsx](http://gae.uniriotec.br/7/educaser/dataset_minicurso.xlsx).
- Consiste num conjunto de dados de alunos de uma turma do final do ensino fundamental, com variáveis intrínsecas aos alunos e outras variáveis relacionadas ao número de faltas e notas numa disciplina qualquer, em cada um dos 4 bimestres do ano letivo.

	A	B	C	D	E	F	G
1	NOME	SEXO	IDADE	REGIAO_RESID	MAIOR_INSTR_PAIS	FALTAS_BIM1	NOTA_BIM1
2	Ana Clara	1	14	Zona Oeste	Ensino Médio	1	5,5
3	Andressa	1	13	Baixada Fluminense	Ensino Fundamental	0	8,3
4	Arthur	2	12	Zona Norte	Ensino Médio	0	8,0
5	Átila	2	13	Zona Oeste	Ensino Fundamental	3	4,2
6	Bianca	1	14	Zona Oeste	Ensino Médio	0	9,5
7	Breno	2	14	Zona Oeste	Ensino Médio	0	8,0
8	Bruna	1	13	Zona Norte	Ensino Superior	0	9,1
9	Cauan	2	13	Zona Norte	Ensino Médio	2	5,4
10	Daniel	2	12	Zona Oeste	Ensino Médio	1	8,3
11	Ellen	1	12	Zona Norte	Sem instrução	1	7,6
12	Fabício	2	13	Baixada Fluminense	Ensino Fundamental	0	6,7
13	Fernanda	1	13	Zona Norte	Ensino Médio	0	8,5
14	Gabriela	1	14	Zona Norte	Ensino Médio	0	9,0

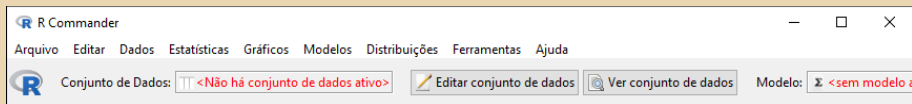
- Menu para a importação de um conjunto de dados em “arquivo Excel” (formatos .xls ou .xlsx) no R Commander:

*Dados → Importar arquivos de dados → do arquivo Excel*

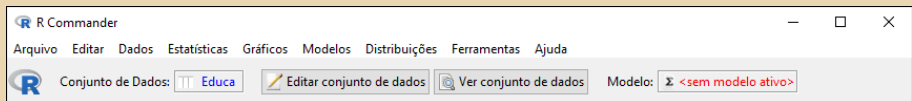
- Dê um nome ao conjunto de dados sem usar espaços (sugestão: Educa), substitua *<casela vazia>* por *NA* e clique em OK.
- Busque pelo arquivo `dataset_minicurso.xlsx` e, após duplo-clique, não esqueça de clicar em *Dados* na pequena janela que surge, pois é esse o nome da planilha onde está o conjunto de dados dentro do arquivo selecionado na importação.

O nome dado ao conjunto de dados importado aparece escrito, na cor azul, ao lado de “Conjunto de dados:”, onde estava inicialmente escrito, na cor vermelha: “<Não há conjunto de dados ativo>”.

- Antes da importação:

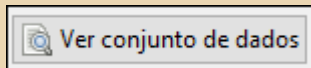


- Após a importação:





- O conjunto de dados cujo nome está escrito na cor azul é o chamado *conjunto de dados ativo*.
- Clicando na caixinha “Ver conjunto de dados”, o conjunto de dados ativo será exibido numa janela à parte, à título de visualização. Ou seja, fechar esta janela não remove o conjunto de dados ativo.



- **IMPORTANTE:** Se a janela *R Commander* estiver maximizada, a janela à parte que mostra o conjunto de dados não ficará visível. Portanto, antes de clicar em “Ver conjunto de dados”, restaure o tamanho da janela *R Commander*, caso esta esteja maximizada.

# Importando conjuntos de dados em mais formatos

O menu *Dados* → *Importar arquivos de dados* possibilita a importação de conjuntos de dados em outros formatos diferentes de .xls ou .xlsx:

Formatos	Submenu
*.txt ; *.dat ; *.csv	<i>de arquivo texto, clipboard ou URL</i>
*.por ; *.sav	<i>do SPSS</i>
*.xpt ; *.xport	<i>do arquivo xport do SAS</i>
*.sas7bdat	<i>from SAS b7dat file</i>
*.mtp	<i>do Minitab</i>
*.dta	<i>do STATA</i>

# Salvando e carregando conjunto de dados

- Para salvar o conjunto de dados (que está elencado como o conjunto de dados ativo): menu

*Dados → Conjunto de dados ativo  
→ Salvar conjunto de dados ativo*

- Feito isto, não é mais necessário importar o conjunto de dados, e após a atual sessão do R Commander ser fechada - acidentalmente ou não - este conjunto de dados pode ser facilmente recuperado via menu de carregamento:

*Dados → Carregar conjunto de dados*

- Ainda, é aconselhável salvar o conjunto de dados a cada modificação que for feita nele, de forma a não precisar refazê-las em sessões futuras.

# Variável categórica x variável numérica

- A primeira coluna do conjunto de dados que importamos expressa a identificação dos alunos.
- Todas as outras colunas expressam ***variáveis***, que são interpretadas como categóricas ou como numéricas.
  - **Variáveis categóricas:** Região de residência e Maior nível de instrução dos pais.
  - **Variáveis numéricas:** Sexo, Idade, Faltas (em cada um dos bimestres) e Nota (em cada um dos bimestres).
- No R Commander, há funcionalidades que estão disponíveis apenas para variáveis categóricas e outras que estão disponíveis apenas para variáveis numéricas.

# Funciona apenas para variável categórica

Resultado	Menu
Gráfico de Barras	<i>Gráficos → Gráfico de Barras</i>
Gráfico de Pizza	<i>Gráficos → Gráfico de Pizza</i>
Distribuição de Frequências	<i>Estatísticas → Resumos → Distribuições de frequência</i>
Tabela de dupla entrada (tabela de contingência com 2 variáveis)	<i>Estatísticas → Tabelas de Contingência → Tabela de dupla entrada</i>

# Funciona apenas para variável numérica

Resultado	Menu
Histograma	<i>Gráficos → Histograma</i>
Gráfico de “hastes”	<i>Gráficos → Plot discrete numeric variable</i>
Boxplot	<i>Gráficos → Boxplot</i>
Densidades	<i>Gráficos → Estimativa de densidade</i>
Medidas descritivas (média, quartis, etc.)	<i>Estatísticas → Resumos → Resumos numéricos</i>
Coeficiente(s) de correlação	<i>Estatísticas → Resumos → Matriz de correlação</i>

# Três questões importantes a serem levantadas:

1. Como fazer o R entender a ordenação natural de variável categórica do tipo ordinal?
2. Como decodificar variável categórica expressa de forma numérica, fazendo com que o R a “entenda” como categórica?
3. Como obter distribuição de frequências de uma variável numérica?

# 1. Reordenando categorias

- No conjunto de dados *Educa*, o “maior nível de instrução dos pais” (MAIOR\_INSTR\_PAIS) é uma variável categórica ordinal. Uma forma de verificar como seus atributos estão ordenados se dá por meio de um gráfico de barras, disponível no menu

*Gráficos → Gráfico de Barras*


- Neste menu, apenas clique em MAIOR\_INSTR\_PAIS e, em seguida, em OK.
- Gráficos gerados pelos menus do R Commander não são exibidos na janela *R Commander*, mas sim na janela do R (*R Gui*), numa sub-janela denominada *R Graphics*.





- Para que os atributos da variável MAIOR\_INSTR\_PAIS passem a seguir a ordenação “Sem instrução” / “Ensino fundamental” / “Ensino médio” / “Ensino superior” em vez da ordem alfabética, acesse o menu

*Dados → Modificação de variáveis no conjunto de dados  
→ Reordenar níveis dos fatores*

- Escolha a variável MAIOR\_INSTR\_PAIS, clique em OK e confirme em seguida o fato de que a “nova” variável irá sobrescrever a antiga.
- Renumere os campos conforme a nova ordenação desejada e clique em OK.

 Reordenar níveis ✕

Níveis antigos	Nova ordem
Ensino Fundamental	2
Ensino Médio	3
Ensino Superior	4
Sem instrução	1

 OK  Cancelar

- Uma forma de verificar que a reordenação foi, de fato, executada, é refazer o gráfico de barras de MAIOR\_INSTR\_PAIS. Aproveitemos para “caprichar” um pouco mais nele.
- Acesse o meu *Gráficos → Gráfico de Barras*, selecione MAIOR\_INSTR\_PAIS e, antes de clicar em OK, clique na aba *Opções* e faça as seguintes edições:
  - rótulo do eixo-x:
  - rótulo do eixo-y: Frequência
  - Título do gráfico: Maior nível de instrução dos pais
  - desmarque a caixinha *Show counts or percentages in bars*.

# Salvando gráficos

- Note que este último gráfico gerado (o mais “caprichado”) eliminou o anterior.
- Portanto, para não perder um gráfico gerado, é necessário salvá-lo (como um arquivo de imagem ou um arquivo pdf).
- Para tal, basta clicar sobre o gráfico e, **ainda na janela *R Gui***, acessar o menu

*Arquivo → Salvar como*

e escolher o formato desejado (sugestão: Png).

## 2. Variável de natureza categórica codificada por números

- Note que a variável SEXO do nosso conjunto de dados está rotulada por números (1 para feminino e 2 para masculino). Desta forma, o R Commander a vê como uma variável numérica.
- Ao acessar qualquer menu de funcionalidades restritas a variáveis categóricas (como gráfico de barras, gráfico de pizza, distribuições de frequências), a variável SEXO não estará listada.

- Para que o R entenda tal variável como categórica, será necessário recorrer ao menu

*Dados → Modificação de variáveis no conjunto de dados  
→ Converter variável numérica para fator*

e gerar uma nova coluna onde 1 se torne “feminino” e 2 se torne “masculino” (ou ainda sobrescrever a variável SEXO original deste conjunto de dados).

- No menu acima citado, selecione a variável SEXO, dê OK e confirme que a “nova” variável sobrescreverá a original. Na janela seguinte, digite *Feminino* ao lado de 1, digite *Masculino* ao lado de 2 e clique em OK.
- Clique em “Ver conjunto de dados” (sem a janela *R Commander* estar maximizada) para verificar a alteração.

- Após o procedimento acima, ao acessar qualquer menu de funcionalidades restrito a variáveis categóricas, a variável SEXO estará listada. Um destes menus, que ainda não exploramos, é o que gera gráficos de pizza:

*Gráficos → Gráfico de Pizza*

- Neste menu, selecione SEXO e, em *Título do gráfico*, escreva seu título de preferência.

### 3. Distribuição de frequências para variáveis numéricas

- Uma vez que o menu *Estatísticas* → *Resumos* → *Distribuições de frequência* funciona apenas para variáveis categóricas, como fazer, portanto, a distribuição de frequência das idades dos alunos, uma vez que a variável IDADE tem, de fato, natureza numérica?
- **Alternativa:** usar o mesmo menu de conversão com o qual substituímos 1 e 2 por Feminino e Masculino na variável SEXO.
  - ⇒ A diferença é que, dessa vez, não queremos “perder” a coluna IDADE na sua forma original, pois ela expressa, de fato, uma variável de natureza numérica.



- Acessar o menu *Dados* → *Modificação de variáveis no conjunto de dados* → *Converter variável numérica para fator*, escolher variável IDADE, marcar *Use números* e dar um nome diferente de IDADE no campo *Novo nome de variável* (sugestão: IDADE\_catég).
- Uma nova coluna (com o nome escolhido acima) é gerada no conjunto de dados, “espelhando” os valores da variável IDADE, mas com o R interpretando esses caracteres não como números, mas sim como meros “rótulos”.
- Fazer a distribuição de frequências da variável IDADE\_catég no menu *Estatísticas* → *Resumos* → *Distribuições de frequência*.

Apesar da alternativa acima ser razoável para a geração de distribuições de frequências, o mesmo não se pode dizer para gráficos de barras. No conjunto de dados que dispomos, há alunos com 12, 13, 14 e 16 anos, mas não há alunos com 15 anos. O gráfico de barras de IDADE\_categ não entende que a magnitude de salto de 14 para 16 é diferente da magnitude de salto de 12 para 13 e de 13 para 14, uma vez que interpreta estes números apenas como “rótulos”.



**Solução:** usar gráficos propícios para variáveis numéricas, tais como gráfico de “hastes” (menu *Gráficos → Plot discrete numeric variable*) ou ainda gráfico de pontos (menu *Gráficos → Gráfico de pontos*).

# Explorando a janela *R Script*

- Linhas de comando são escritas sequencialmente na janela *R Script* a cada ação “finalizada” nos menus do R Commander.
- Todo início de linha de comando na janela *R Script* começa “colado” à margem.
- Linhas de comando muito compridas podem ocupar mais de uma linha na janela *R Script*. Quando uma linha não começa “colada” na margem, ela é continuação da linha acima - ou seja, ainda é a mesma linha de comando da linha acima.

- Para ilustrar, gere um gráfico de barras da variável REGIAO\_RESID por meio do menu *Gráficos → Gráfico de Barras* (com título pertinente).
- Considere que o último comando executado antes do gráfico sugerido acima foi o gráfico de “hastes” das idades. As três linhas na figura abaixo representam o final da janela *R Script*, onde são ilustradas apenas duas linhas de comando, uma vez que a terceira linha é a continuação da segunda linha, formando uma única linha de comando.

```
with(Educa, discretePlot(IDADE, scale="frequency"))  
with(Educa, Barplot(REGIAO_RESID, xlab="", ylab="Frequência",  
  main="Região de residência", label.bars=FALSE))
```





- Uma das vantagens da disponibilização de linhas de comando na janela *R Script* reside no fato de podermos usar sua “estrutura base” para realizar modificações pertinentes que não podem ser feitas pelos menus do R Commander.
- Marcando toda a linha de comando de interesse (como se fóssemos copiá-la) e clicando em seguida no botão *Submeter* (atalho:  +  ), tal linha de comando - modificada ou não - é executada.
- Vamos ilustrar esta ideia modificando a cor das barras do último gráfico de barras gerado.

- Antes dos dois “fecha parênteses” no final da linha de comando associada a este gráfico de barras na janela *R Script*, digite

`,col="red"`

de forma que fique algo do tipo

```
with(Educa, Barplot(REGIAO_RESID, xlab="", ylab="Frequência",  
  main="Região de residência", label.bars=FALSE,col="red"))
```


- Marque toda a linha de comando como se fosse copiá-la, mas em vez de  +  , use  +  ou clique em *Submeter*.
- Confira na janela *R Graphics* (dentro da janela *R Gui*) o gráfico de barras com barras vermelhas.

# E para deixar cada uma das três barras com uma cor diferente?

- No lugar de "red", usar um vetor de três cores.

- Por exemplo, no lugar de `,col="red"`, usar

`,col=c("blue","yellow","green")`

Para ver a lista com os nomes das cores que o R disponibiliza, digite `colors()` na janela *R Script* e submeta-a, ou então digite-a na mesma janela que você digitou `require(Rcmdr)` ou `library(Rcmdr)` e aperte .

# Referências para o “Rcmdr básico”

- Acesse <http://gae.uniriotec.br/7/material.html>.
- Apostila introdutória “Introdução ao R Commander - Notas de Aula” e apostila complementar “R Commander: um pouco além dos menus gráficos”.
- Conjuntos de dados utilizados em exemplos e exercícios destas duas apostilas.
- Apostilas revisadas pela última vez em fevereiro de 2019 (versão do R: 3.5.2 / Versão do Rcmdr: 2.5-1).
- Em essência, elas podem ser bem aproveitadas, ainda que atualizações do pacote Rcmdr tenham acrescentado novas possibilidades não abordadas nas apostilas (e talvez removido outras possibilidades).



# Referências Bibliográficas



Fox, J. (2022). Package ‘RcmdrMisc’. Disponível em <https://cran.r-project.org/web/packages/RcmdrMisc/RcmdrMisc.pdf>



Fox, J., & Bouchet-Valat, M. (2021). Rcmdr: R Commander. R package version 2.7-2.



R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.



Ribeiro Melo, F.R. (2019). Introdução ao R Commander – Notas de Aula. Disponível em [http://gae.uniriotec.br/7/Notas\\_de\\_aula\\_Rcmdr.pdf](http://gae.uniriotec.br/7/Notas_de_aula_Rcmdr.pdf).



Ribeiro Melo, F.R. (2019). R Commander: um pouco além dos menus gráficos. Disponível em [http://gae.uniriotec.br/7/Rcmdr\\_alem\\_dos\\_menus.pdf](http://gae.uniriotec.br/7/Rcmdr_alem_dos_menus.pdf).



Terrón, A., Cabellero, P. & Alcaraz, L. (2011). Estadística Básica con R-Commander. Disponível em <http://www.bubok.es/libros/203887/Estadistica-basica-con-RCommander>.



AlternativeTo (2020). *GUIs to save from typing R code*. Disponível em <https://alternativeto.net/list/2063/guis-to-save-from-typing-r-code/>

*MUITO OBRIGADO !!!*

