

**PROJETOS DE ESTATÍSTICA
EM SALA DE AULA COM O R**

Conselho Editorial

Série Letra Capital Acadêmica

Ana Elizabeth Lole dos Santos (PUC-Rio)
Beatriz Anselmo Olinto (Unicentro-PR)
Carlos Roberto dos Anjos Candeiro (UFTM)
Claudio Cezar Henriques (UERJ)
Ezilda Maciel da Silva (UNIFESSPA)
João Luiz Pereira Domingues (UFF)
João Medeiros Filho (UCL)
Leonardo Agostini Fernandes (PUC-Rio)
Leonardo Santana da Silva (UFRJ)
Lina Boff (PUC-Rio)
Luciana Marino do Nascimento (UFRJ)
Maria Luiza Bustamante Pereira de Sá (UERJ)
Michela Rosa di Candia (UFRJ)
Olavo Luppi Silva (UFABC)
Orlando Alves dos Santos Junior (UFRJ)
Pierre Alves Costa (Unicentro-PR)
Rafael Soares Gonçalves (PUC-RIO)
Robert Segal (UFRJ)
Roberto Acízelo Quelhas de Souza (UERJ)
Sandro Ornellas (UFBA)
Sergio Azevedo (UENF)
Sérgio Tadeu Gonçalves Muniz (UTFPR)
Waldecir Gonzaga (PUC-Rio)

Alexandre Sousa da Silva, Bruno Francisco Teixeira Simões,
Davi da Silveira Barroso Alves, Felipe Rafael Ribeiro Melo,
Letícia Martins Raposo, Luciane de Souza Velasque,
Maria Beatriz Assunção Mendes da Cunha,
Maria Tereza Serrano Barbosa, Steven Dutt-Ross e
Vinicius Pinheiro Israel

PROJETOS DE ESTATÍSTICA EM SALA DE AULA COM O R

LETRAPITAL



Copyright © Alexandre Sousa da Silva, Bruno Francisco Teixeira Simões,
Davi da Silveira Barroso Alves, Felipe Rafael Ribeiro Melo,
Letícia Martins Raposo, Luciane de Souza Velasque,
Maria Beatriz Assunção Mendes da Cunha,
Maria Tereza Serrano Barbosa, Steven Dutt-Ross e
Vinicius Pinheiro Israel, 2024

*Todos os direitos reservados e protegidos pela Lei nº 9.610, de 19/02/1998.
Nenhuma parte deste livro pode ser reproduzida ou transmitida, sejam quais forem os
meios empregados, sem a autorização prévia e expressa do autor.*

EDITOR João Baptista Pinto

CAPA Dos autores

PROJETO GRÁFICO/EDITORAÇÃO Luiz Guimarães

REVISÃO Dos autores

CIP-BRASIL. CATALOGAÇÃO NA PUBLICAÇÃO
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

P958

PROJETOS DE ESTATÍSTICA EM SALA DE AULA COM O R / Alexandre Sousa da
Silva, Bruno Francisco Teixeira Simões, Davi da Silveira Barroso Alves, Felipe Rafael
Ribeiro Melo, Letícia Martins Raposo, Luciane de Souza Velasque, Maria Beatriz
Assunção Mendes da Cunha, Maria Tereza Serrano Barbosa, Steven Dutt-Ross e Vinicius
Pinheiro Israel - 1. ed. - Rio de Janeiro: Letra Capital, 2024.

244 p. : il. ; 23 cm.

Inclui bibliografia

ISBN 978-85-7785-969-6

1. Estatística. 2. Métodos estatísticos. I. Silva, Alexandre Sousa da.

24-92588

CDD: 519.5

CDU: 519.2

Gabriela Faray Ferreira Lopes - Bibliotecária - CRB-7/6643

LETRA CAPITAL EDITORA
Tels.: (21) 3553-2236 / 2215-3781 / 99380-1465
www.letracapital.com.br

Sumário

Apresentação	7
CAPÍTULO 1 - R Commander: uma alternativa à dificuldade (docente e discente) em usar o R em sala de aula.....	11
<i>Felipe Rafael Ribeiro Melo</i>	
CAPÍTULO 2 - Aquisição de microdados de sistemas de informação com R para projetos na área da saúde	43
<i>Davi da Silveira Barroso Alves</i>	
CAPÍTULO 3 - Aprendizagem de estatística baseada em problemas: um exemplo com dados sobre estilo de vida e acidente vascular cerebral	68
<i>Letícia Martins Raposo</i>	
CAPÍTULO 4 - Estatística cidadã: a COVID-19 como temática na formação de estudantes de medicina	102
<i>Luciane de Souza Velasque</i> <i>Maria Beatriz Assunção Mendes da Cunha</i>	
CAPÍTULO 5 - Diferenças de gênero no desempenho em matemática: uma análise dos microdados do enem 2019	125
<i>Alexandre Sousa da Silva</i>	

CAPÍTULO 6 - Casos de ensino em estatística: uma abordagem para aprender de forma mais aplicada com ênfase na programação com R.....	141
<i>Steven Dutt-Ross</i>	
CAPÍTULO 7 - Da elaboração de questionários à modelagem estatística: um projeto sobre a lei de cotas	165
<i>Maria Tereza Serrano Barbosa</i>	
CAPÍTULO 8 - O ensino da estatística multivariada: uma aplicação na identificação de padrões mundiais de alimentação.....	187
<i>Bruno Francisco Teixeira Simões</i> <i>Giovanna da Conceição Nepomuceno</i>	
CAPÍTULO 9 - Do ensino tradicional às metodologias ativas: uma aplicação no ensino de probabilidade nas ciências humanas	210
<i>Vinicius Pinheiro Israel</i>	
Biodata.....	239

Apresentação

É com muita alegria que disponibilizamos, para a comunidade de docentes que ensinam Estatística, o fruto da nossa experiência didática organizada e sistematizada neste livro. Apesar de já divulgada em eventos e artigos científicos, pretendemos, com esta nova publicação, que a nossa metodologia de ensino possa ser mais bem compreendida, multiplicada e aperfeiçoada pelos que se dedicam à tarefa de transmitir o pensamento estatístico e seus conceitos fundamentais.

A necessidade do conhecimento e do uso consciente da Estatística para a sociedade vem, a cada dia, ficando mais evidente. Ela está presente nas mais diversas áreas, sendo aplicada para a compreensão da estrutura distributiva da renda e riqueza na sociedade, se mostrando fundamental no entendimento e modelagem das incertezas no mercado financeiro, nas organizações ou nas disputas eleitorais, desempenhando um papel de centralidade no desenvolvimento de pesquisas e ações públicas nas áreas da saúde e do meio ambiente, assim como em outras áreas.

Durante a pandemia da COVID-19, a leitura das curvas epidêmicas, a compreensão das diferentes medidas de eficácia das vacinas e os resultados dos estudos que comprovaram a ineficiência da cloroquina no tratamento da doença fizeram com que a Estatística passasse a estar presente na mídia e nas conversas em família. Outros assuntos que também circularam na mídia, como a interpretação dos resultados das pesquisas eleitorais, com suas diferentes margens de erros e metodologias, e a divulgação e discussão a respeito de diferentes estatísticas sobre o desmatamento também deixaram claro o quanto esse conhecimento é fundamental para que profissionais das mais diferentes áreas tenham a capacidade de ler e interpretar a realidade social, política e ambiental por meio dos dados e possam atuar na formulação ou avaliação de políticas públicas.

As experiências aqui compartilhadas tiveram início em 2012, quando professores do atual Departamento de Métodos Quantitativos (DMQ) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) reuniram-se e decidiram, a partir de uma avaliação conjunta de suas disciplinas, reformular a metodologia de ensino utilizada em sala de aula. Apesar de a diversidade de formação e enfoque de cada professor, desde o início da mudança de metodologia, os estudantes das disciplinas de Estatística oferecidas pelo DMQ aprendem a ler e analisar uma base de dados por meio do programa R e desenvolvem um artigo a ser apresentado em formato pôster ou oral no final da disciplina.

Nesses dez anos, acumulamos muitas experiências para compartilhar, desde as dificuldades iniciais que tivemos ao renunciar às aulas tradicionais, a introdução de dinâmicas para incentivar a participação e criar uma relação mais dialógica com os estudantes, até os inúmeros casos de sucesso, com muitos trabalhos apresentados em eventos científicos, e a satisfação de vermos estudantes mais motivados e interessados em continuar os estudos em áreas afins à Estatística.

Durante esse período, alguns dos professores incorporaram a Educação Estatística como sua área de pesquisa e começaram a contribuir com propostas e reflexões didáticas para o Ensino Básico ou Superior, outros priorizaram desenvolver com os estudantes a capacidade de formular questões de pesquisa ou, ainda, de utilizar uma análise de dados para propor uma intervenção prática.

Essas multiplicidades de visões e atuações estão reproduzidas nos diversos capítulos em que cada professor compartilha um pouco das suas experiências. Os dois capítulos iniciais são preciosos para quem deseja experimentar nossa metodologia: o primeiro, escrito por Felipe Ribeiro, apresenta a interface R Commander (propiciada pelo pacote Rcmdr) e suas potencialidades, sendo fruto de sua experiência na autoria de duas apostilas sobre esse pacote e nas oficinas/minicursos para públicos diversos que tem tido a oportunidade de ministrar. O segundo, do Davi Alves, nos fornece um mapa da mina para a aquisição de

bancos de dados, com uma abordagem que avança em temas e conceitos relacionados aos desenhos de estudos em várias áreas da saúde. Os outros capítulos, em geral, demonstram como a metodologia pode ser aplicada na prática. Aqueles que fizerem uma leitura atenta podem discernir as múltiplas possibilidades. Os dois capítulos a seguir são direcionados principalmente aos cursos da área da saúde. O escrito pela Letícia Raposo utiliza uma base de dados de acidente vascular cerebral para descrever como é possível apresentar aos estudantes, de forma gradual, os conceitos estatísticos envolvidos, desde uma análise exploratória univariada, até a aplicação de diferentes testes de hipóteses por meio de *scripts* do R. O capítulo escrito em conjunto pela Maria Beatriz Assunção e Luciane Velasque, que têm muita experiência com estudantes de medicina, mostra também o caminho de uma análise estatística completa até a regressão logística, mas agora com dados da Covid-19 no Rio de Janeiro e utilizando a interface do R Commander. Alexandre Silva, no capítulo a seguir, dá alguns exemplos de temas que foram trabalhados pelas suas turmas de vários cursos e utiliza o ciclo investigativo para descrever uma análise de dados do ENEM, realizada junto ao curso de Licenciatura em Matemática, com vistas a identificar o efeito da renda e gênero nas notas de matemática e português. Steven Ross, que em suas aulas incentiva a escrita autoral dos estudantes e a elaboração de uma proposta de intervenção a partir de uma análise estatística, descreve um dos casos trabalhados com estudantes da Administração Pública, a partir de uma base de dados simulada, com o objetivo de avaliar se está havendo uma discriminação em uma determinada empresa. O capítulo de Maria Tereza apresenta o projeto que deu origem à mudança de metodologia e que foi realizado coletivamente pelos docentes do departamento na época. A proposta visava avaliar o quanto a Lei de Cotas impactaria o perfil dos estudantes na universidade. Nesse capítulo são descritas as etapas desde a elaboração do questionário até a comparação dos resultados obtidos em 2012 e 2015, em que uma tabela de contingência ou uma regressão logística permitem avaliar se houve mudança de opinião entre os períodos avaliados.

Os dois capítulos finais são, também, bem ilustrativos da diversidade de enfoques que podem abranger essa metodologia. Bruno Simões, com a colaboração de sua orientanda Giovanna Nepomuceno e empregando um banco de dados a respeito da disponibilidade de grupos alimentares em 171 países no ano de 2018, utiliza a técnica de estatística multivariada de Análise de Componentes Principais para deduzir padrões alimentares, consoantes com a teoria qualitativa da área da Nutrição e, por fim, mostra como representá-los em mapas-múndi para visualizar o espalhamento desses padrões no globo terrestre. Por fim, o último capítulo, escrito por Vinicius Israel, demonstra como os conceitos de probabilidade podem ser contextualizados em exemplos que motivem estudantes do Bacharelado em Ciência Política.

Ao compartilhar esses projetos desenvolvidos em sala de aula, queremos ampliar a utilização dessa metodologia, permitindo que ela se multiplique e seja aperfeiçoada pelos colegas docentes da área. Os bancos de dados e os códigos em R utilizados nos capítulos estão disponíveis em: <https://github.com/DMQ-UNIRIO/livro-projetos-com-o-R>. Nosso objetivo é contribuir para a formação de profissionais que conheçam a beleza da Estatística, sua importância para a sociedade, e que saibam identificar os limites e alcances das informações obtidas a partir de um conjunto de dados.

Os autores

R Commander: uma alternativa à dificuldade (docente e discente) em usar o R em sala de aula

Felipe Rafael Ribeiro Melo

1. Introdução

1.1. O *software* R

R é um *software* livre voltado principalmente para tratamento e análise de dados. De maneira mais estendida, pode ser visto como uma linguagem de programação especializada na gestão e análise de dados. Ele começou a ser desenvolvido pelos pesquisadores Robert Gentleman e Ross Ihaka (do Departamento de Estatística da Universidade de Auckland, na Nova Zelândia) em meados dos anos 1990. O nome do *software* se deve à letra inicial dos nomes dos pesquisadores.

As principais vantagens do *software* R seguem listadas abaixo:

- Gratuito;
- Potente e estável;
- É um programa que tem uma longa história, com mais de 20 anos de desenvolvimento;
- É apoiado por uma grande equipe de desenvolvedores em todo o mundo;
- Grande quantidade de pacotes disponíveis, os quais fornecem funcionalidades específicas;
- Metodologias de ponta desenvolvidas primeiramente em R e disponibilizadas em forma de pacotes;
- Amplamente utilizado no meio acadêmico em diversas áreas de conhecimento;
- Diversos materiais, tutoriais e fóruns de discussão disponíveis gratuitamente na Internet;
- Disponível para Windows, Linux e Mac.

Apesar dessas e de outras vantagens, o R ainda não possui oficialmente uma versão compatível para dispositivos móveis. Entretanto, o principal ponto negativo aos que lhe são apresentados reside em sua interface padrão pouco amigável, apta a receber apenas linhas de comando. Uma forma de contornar (ao menos parcialmente) essa última desvantagem é o foco deste capítulo: o uso de um “R com menus” propiciado pela interface R Commander, carregada por meio de um pacote chamado Rcmdr. Para o leitor pouco ou nada experiente com o *software* R, as últimas frases podem ter soado como escritas em um idioma desconhecido. Resumindo a ideia, o que está sendo passado aqui é a possibilidade real de se trabalhar com o R em turmas de graduação considerando que o docente **não** necessite ter conhecimento de programação na linguagem R e/ou não precise ensinar programação em R para seus alunos. Apesar da interface mencionada (R Commander) não lidar com tudo que o R é capaz de fazer, no que tange ao tratamento e à análise de dados, por meio de menus, as funcionalidades presentes nos menus do R Commander e dos seus pacotes adicionais contemplam uma gama enorme de possibilidades para alunos de graduação das mais diversas áreas, tais como Administração, Biblioteconomia, Ciências Biológicas, Nutrição, Engenharia de Produção, Sistemas de Informação, entre outras. O corpo docente do Departamento de Métodos Quantitativos (DMQ) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) tem utilizado o *software* R com sua interface R Commander nas disciplinas de Estatística, sejam para cursos de ciências humanas, biomédicas ou exatas. Contando com um *feedback* positivo de vários alunos em um histórico de quase dez anos de aplicação de uma metodologia de ensino de estatística com suporte do *software* R por meio da interface R Commander, este capítulo não traz apenas ideias que “podem vir a dar certo, mas que nunca foram testadas”. O propósito aqui é incentivar o uso de uma ferramenta que tem sido proveitosa ao longo dos últimos anos nas referidas turmas, trazendo aos discentes uma forma diferente de lidar com as disciplinas de estatística: um processo de aprendizagem

baseado no ato de decorar fórmulas e fazer contas em posse de apenas de lápis/caneta, papel e calculadora transforma-se em uma disciplina na qual busca-se a compreensão do que cada “fórmula” quer dizer, aonde ela “quer chegar” e quando pode/deve (ou não pode/não deve) ser utilizada, com devida argumentação, ficando o “trabalho braçal” (ou seja, as temidas contas) por conta do *software* R.

1.2. Objetivo

Apresentar o *software* R e a interface R Commander, gerada por meio do pacote Rcmdr, além de algumas funcionalidades básicas envolvendo tratamento e análise de conjuntos de dados.

1.3. Público-alvo

Este capítulo direciona-se a todos aqueles que tem interesse no tratamento e análise exploratória de dados utilizando o *software* R, mas possuem pouco ou nenhum conhecimento de programação na linguagem R, ou ainda que desconheçam interfaces mais “amigáveis” que pacotes do R podem oferecer. Em especial, o presente capítulo foi pensado para professores das mais diferentes áreas de conhecimento que tem interesse em trabalhar com seus alunos o tratamento e a análise de conjuntos de dados relevantes para a(s) disciplina(s) que ministram sem a necessidade de ensinar programação em R para sua(s) turma(s).

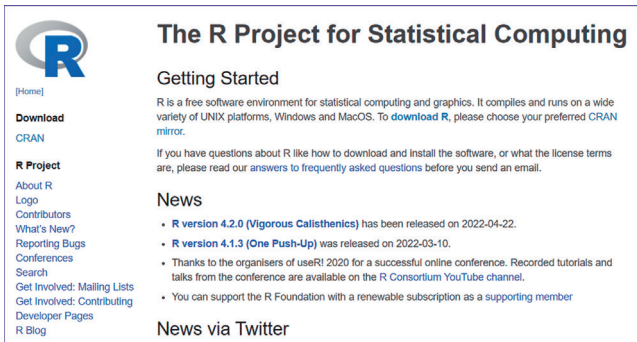
1.4. Habilidades a serem desenvolvidas

1. *Download* e instalação do *software* R.
2. Instalação e carregamento do pacote Rcmdr.
3. Importação de conjuntos de dados.
4. Algumas edições pertinentes em variáveis de um conjunto de dados, visando análises exploratórias básicas.
5. Edição/compilação de linhas de comando geradas por ações nos menus do R Commander.

2. Download e instalação do R (no Windows)

Para baixar o R, acesse <https://www.r-project.org/> e clique sequencialmente em:

1. CRAN (do lado esquerdo da tela);
2. um dos *mirrors* disponíveis (sugestão: <https://cran-r.c3sl.ufpr.br/>);
3. *Download R for Windows*;
4. *base*;
5. por fim, no *link* em destaque (algo do tipo no *Download R-4.2.0 for Windows*).



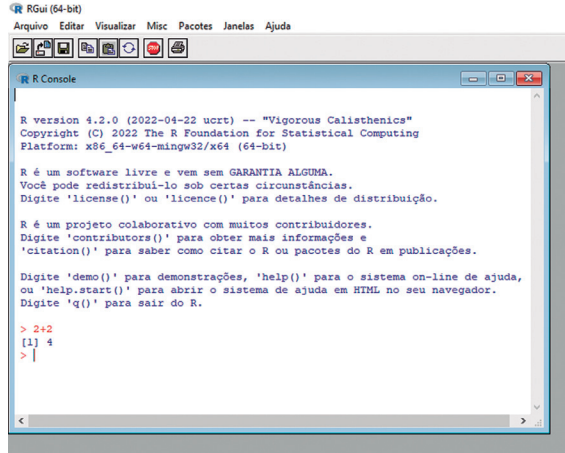
The screenshot shows the R Project website for Windows. On the left is a navigation menu with links: [Home], Download, CRAN, R Project, About R, Logo, Contributors, What's New?, Reporting Bugs, Conferences, Search, Get Involved: Mailing Lists, Get Involved: Contributing, Developer Pages, and R Blog. The main content area is titled "The R Project for Statistical Computing" and "Getting Started". It states that R is a free software environment for statistical computing and graphics, available for UNIX, Windows, and MacOS. It instructs users to download R from a preferred CRAN mirror. Below this, it provides frequently asked questions. A "News" section lists three items: R version 4.2.0 (Vigorous Calisthenics) released on 2022-04-22, R version 4.1.3 (One Push-Up) released on 2022-03-10, and a thank you to the organisers of useR! 2020. At the bottom, there is a "News via Twitter" link.

Um caminho mais rápido para baixar o R (do *mirror* sugerido acima) se dá ao acessar <https://cran-r.c3sl.ufpr.br/bin/windows/base/>, o que contempla os quatro primeiros passos, restando apenas clicar no *link* em destaque do Passo 5.

Após concluído o *download*, a instalação do R é simples. Por comodidade, ela pode ser feita clicando sequencialmente em “Próximo” após o duplo clique no arquivo executável baixado.

Ao abrir o R, você irá se deparar com a já mencionada “interface pouco amigável”: uma janela grande (*R Gui*) que contém uma janela menor chamada *R Console*, na qual há uma série de créditos e uma “seta vermelha”, que é o chamado *prompt de comando*. Ele indica que o R está apto a receber uma linha de comando (que pode ser algo tão simples como “2+2” ou uma linha de comando bastante extensa e complexa). Na

janela *R Console*, basta apertar *Enter* após concluir a digitação da linha de comando para que ela seja compilada (faça um teste com o “2+2” e ignore o número 1 entre colchetes ao lado do resultado).



```

R GUI (64-bit)
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda

R R Console

R version 4.2.0 (2022-04-22 ucrt) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> 2+2
[1] 4
> |

```

A principal desvantagem do R reside, portanto, na interface pouco “amigável”, apta a receber apenas linhas de comando. Todavia, isso pode ser contornado (ao menos parcialmente) por meio de algumas interfaces fornecidas por pacotes do R. Aqui, abordaremos a interface R Commander, que provém de um pacote do R chamado Rcmdr.

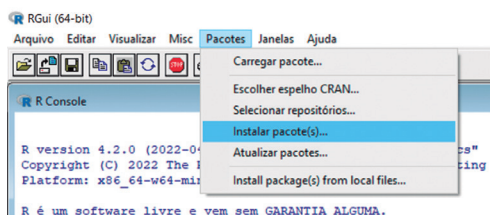
3. O pacote Rcmdr

Uma das características marcantes do *software* R é a grande quantidade de *pacotes* disponíveis, os quais fornecem funcionalidades específicas. Aqui, vamos explorar o pacote Rcmdr (uma forma abreviada de R Commander), 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*.

3.1 Instalando o pacote Rcmdr

A instalação de qualquer pacote no R pode ser feita por meio do menu *Pacotes* da janela *R Gui* ou por meio da linha de comando **install.packages**("nome do pacote") sem excluir as aspas, ou seja, o pacote Rcmdr pode ser instalado pela linha de comando **install.packages**("Rcmdr"). Uma vez instalado, ele pode ser utilizado a qualquer momento sem a necessidade de repetir esse processo. Segue o passo a passo para instalar o pacote Rcmdr via menu.

1. Acesse o menu *Pacotes* > *Instalar pacotes*;



2. Uma pequena janela com os CRAN *mirrors* disponíveis será aberta. Escolha um deles (sugestão: 0-Cloud) e clique em *OK*;
3. Uma nova janela com os pacotes disponíveis (no CRAN *mirror* escolhido) será aberta. Procure por Rcmdr, clique nele e em *OK*. Feito isso, pode ser que duas pequenas janelas com opções *Sim* e *Não* apareçam uma após a outra. Selecione *Sim* em ambas;
4. O procedimento de instalação do pacote será finalizado quando o prompt de comando (“seta vermelha”) tornar a aparecer na janela *R Console*.

Observação: Alternativamente, poderíamos ter escolhido não apenas o pacote Rcmdr no Passo 3 do processo de instalação acima, mas sim todos os pacotes que iniciam por “Rcmdr” de uma vez só antes de clicar em *OK*. Isso faz com que o Rcmdr e seus *plugins* sejam instalados simultaneamente. Contudo, o processo de instalação se torna mais demorado.


```

R R Console

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> utils::install.packages()
--- Please select a CRAN mirror for use in this session ---
tentando a URL 'https://cloud.r-project.org/bin/windows/contrib/4.2/Rcmdr_2.5
Content type 'application/zip' length 5318582 bytes (5.1 MB)
downloaded 5.1 MB

package 'Rcmdr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\09805174786\AppData\Local\Temp\RtmpySa2RR\downloaded_packages

>|

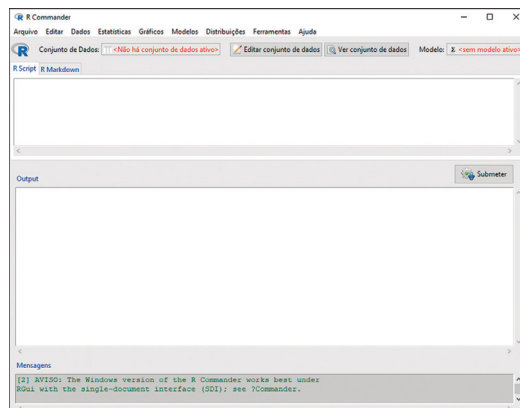
```

3.2 Carregando o pacote Rcmdr

Para usufruir de qualquer pacote instalado no R, é necessário *carregá-lo*. Para carregar o pacote Rcmdr, basta abrir o R, digitar uma das duas expressões abaixo¹ e apertar *Enter* em seguida.

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

Feito isso, uma nova janela será aberta: a janela *R Commander*, que é a interface “amigável” com a qual trabalharemos. Cabe ainda ressaltar que, 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*.



¹ Atente ao fato de que apenas uma letra de cada expressão é maiúscula e todas as outras são minúsculas. Isso deve ser respeitado, uma vez que o R reconhece uma mesma letra em suas versões minúscula e maiúscula como caracteres diferentes.

Observação: Se a janela *R Commander* for fechada, será necessário abrir uma nova sessão do R e carregar novamente o pacote `Rcmdr`, seja por meio da linha de comando `require(Rcmdr)` ou da linha de comando `library(Rcmdr)`.

Note que a janela *R Commander* é dividida em três janelas: *R Script*, *Output* e *Mensagens*. O propósito das duas primeiras ficará mais claro no decorrer do capítulo.

- Na janela *R Script*, são impressas sequencialmente as linhas de comando de tudo que é executado via menus.
- A janela *Output* é a janela de saída, na qual são exibidos os resultados de alguns comandos executados.
- A janela *Mensagens* exibe informações que o sistema julga relevante compartilhar com o usuário, incluindo alertas.

Os três principais menus do R Commander para quem está começando são:

- *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, testes de hipóteses, ajustes de modelos e análises multivariadas.

O foco maior neste capítulo será direcionado ao menu *Dados*, uma vez que os menus *Gráficos* e *Estatísticas* são mais intuitivos: enquanto o primeiro se refere à geração de diferentes visualizações gráficas, o segundo é dedicado a resultados descritivos, tais como distribuições de frequência, tabelas de contingência e resumos numéricos. Os resultados gerados pelo menu *Estatísticas* são exibidos na janela *Output*.

4. Importando conjuntos de dados

De posse de uma ferramenta poderosa para o tratamento e análise de dados como o R e de uma interface motivadora como o R Commander, como inserir os dados de interesse? Há três formas de se fazer isso:

1. **Digit**ar um conjunto de dados: menu *Dados > Novo conjunto de dados*;
2. **Carregar** conjunto de dados (na extensão .RData): menu *Dados > Carregar conjunto de dados*;
3. **Importar** um conjunto de dados (em outras extensões): menu *Dados > Importar arquivo de dados*.

Do ponto de vista de se trabalhar com dados secundários disponíveis em extensões diferentes de .RData (o que é muito comum), é necessário **importar** o conjunto de dados em questão, conforme será detalhado nas próximas subseções.

4.1 Importando conjuntos de dados – extensões .xls e .xlsx

Para ilustrar o procedimento de importação de conjunto de dados nestas extensões, acesse o arquivo “dataset_minicurso.xlsx” disponível na pasta do Capítulo1 no endereço: <https://github.com/DMQ-UNIRIO/livro-projetos-com-o-R> para baixar o arquivo nomeado dataset_minicurso.xlsx, que contém um conjunto de dados de alunos de uma turma fictícia dos últimos anos do ensino fundamental, com variáveis intrínsecas aos alunos e outras relacionadas ao número de faltas e notas numa disciplina qualquer, em cada um dos 4 bimestres do ano letivo. As figuras a seguir exibem recortes das duas planilhas contidas no arquivo: a planilha *Dados*, que traz o conjunto de dados em si; e a *Dicionario_de_variaveis*, que informa o que representa cada coluna da planilha anterior.

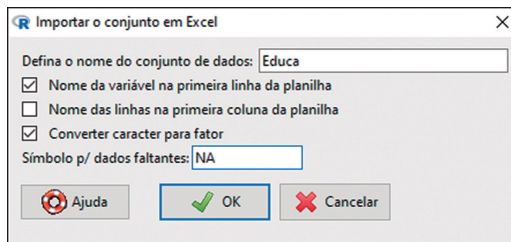
	A	B	C	D	E	F	G
1	NOME	SEXO	IDADE	REGIAO_RESID	MAIOR_INSTR_PAIS	FALTAS_BIM1	NOTA_BIM1
2	Ana Clara	1	14	Zona Oeste	Ensino Médio	1	5,5
3	Andressa	1	13	Baixada Fluminense	Ensino Fundamental	0	8,3
4	Arthur	2	12	Zona Norte	Ensino Médio	0	8,0
5	Áttila	2	13	Zona Oeste	Ensino Fundamental	3	4,2
6	Bianca	1	14	Zona 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
15	Guilherme	2	13	Zona Norte	Ensino Médio	0	6,8
16	Heitor	2	13	Zona Oeste	Ensino Médio	3	7,5
17	Henrique	2	14	Zona Norte		3	7,1
18	Isabele	1	14	Zona Oeste	Ensino Superior	2	7,5

	A	B	C
1	Nome da variável no conjunto de dados	Significado	
2	NOME	Nome do(a) aluno(a)	
3	SEXO	Sexo	1 = Feminino / 2 = Masculino
4	IDADE	Idade (em anos completos)	
5	REGIAO_RESID	Região de residência	Zona Norte / Zona Oeste / Baixada Fluminense
6	MAIOR_INSTR_PAIS	Maior nível de instrução dos pais	Sem instrução / Ensino Fundamental / Ensino Médio / Ensino Superior
7	FALTAS_BIM1	Total de faltas no 1º bimestre	
8	NOTA_BIM1	Nota na prova do 1º bimestre	
9	FALTAS_BIM2	Total de faltas no 2º bimestre	
10	NOTA_BIM2	Nota na prova do 2º bimestre	
11	FALTAS_BIM3	Total de faltas no 3º bimestre	
12	NOTA_BIM3	Nota na prova do 3º bimestre	
13	FALTAS_BIM4	Total de faltas no 4º bimestre	
14	NOTA_BIM4	Nota na prova do 4º bimestre	

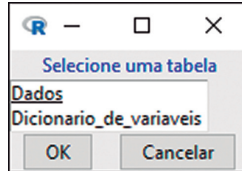
No R Commander, acesse o menu

Dados > Importar arquivos de dados > do arquivo Excel.

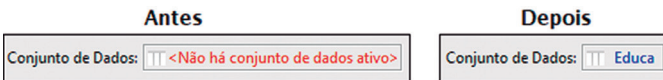
Uma pequena janela será aberta, na qual você deve dar um nome a este conjunto de dados **sem usar espaços**. No campo “Símbolo p/ dados faltantes”, é recomendável substituir por NA e, quanto às outras opções nessa pequena janela, podemos deixar como está.



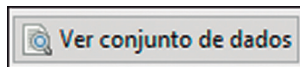
Após clicar em *OK* na janela acima, busque pelo arquivo baixado (*dataset_minicurso.xlsx*) e dê um duplo-clique nele. Por fim, não se esqueça de clicar em *Dados* na pequena janela que surge, uma vez que é nome da planilha onde está o conjunto de dados, e clique em *OK* para concluir o processo de importação.



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>”, acima da janela *R Script*. Portanto, o conjunto de dados importado passou a ser tomado como o **conjunto de dados ativo**. Por conjunto de dados ativo, entende-se o conjunto de dados (digitado, carregado ou importado) cujo nome está exposto nesse campo, na cor azul. Todas as funcionalidades do R Commander de tratamento e análise de dados tomarão como referência esse conjunto de dados. Mais à frente, será mostrado que, dentro de uma mesma sessão do R Commander, podem ser inseridos dois ou mais conjuntos de dados. Entretanto, apenas o que está com o nome escrito ao lado de “Conjunto de dados:” está na posição de conjunto de dados ativo, com os outros sendo “ignorados” pelos menus do R Commander.



Outra forma de conferir o sucesso da importação do conjunto de dados (nomeado aqui de *Educa*) é clicar na caixinha “Ver conjunto de dados”: o conjunto de dados ativo será exibido numa janela à parte a título de visualização, de forma que fechar essa janela *não* removerá o conjunto de dados.



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 esteja maximizada.

4.2 Importando conjuntos de dados – extensões .csv e .txt

Outras extensões comuns de arquivos contendo dados são as extensões .txt e .csv. Apesar desta última ser frequentemente visualizada por meio de *softwares* de planilhas, também é possível visualizá-la por meio de aplicativos de texto, como o Bloco de Notas. Portanto, arquivos em ambas as extensões podem ser abertos em programas como o Bloco de Notas ou similares.

Antes de iniciar, no *R Commander*, o procedimento de importação de conjunto de dados em alguma dessas extensões, é aconselhável abrir o arquivo em questão no Bloco de Notas (ou programa similar), no intuito de verificar:

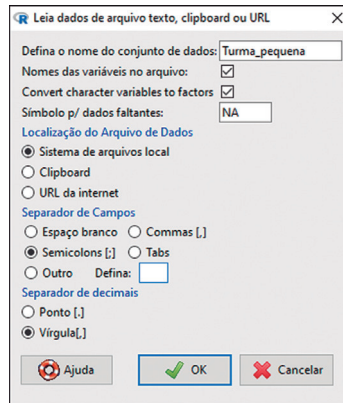
- qual caractere está sendo usado como separador de campos;
- qual caractere está sendo usado como separador de decimais.

O separador de decimais só é relevante quando o conjunto de dados possui pelo menos uma variável numérica que possa assumir valores não inteiros. Nesse caso, é necessário verificar se é usado ponto (.) ou vírgula (,) como caractere que separa a parte inteira dos decimais. Já a definição do separador de campos é sempre relevante, pois indica o caractere que “divide células” numa mesma linha do arquivo. A título de ilustração, acesse <https://github.com/DMQ-UNIRIO/livro-projetos-com-o-R> para baixar o arquivo Bancodedados.zip. Após concluído o *download*, extraia da pasta zipada o arquivo Turma3.txt, que consiste num pequeno conjunto de dados de uma turma de cinco alunos, contendo nome, sexo, número de faltas (numa determinada disciplina) e nota (numa prova da disciplina em questão) de cada um dos

alunos. Ao abrir o arquivo, verifica-se que o separador de campos é ponto e vírgula (;) e o separador de decimais é vírgula (.). Com essas informações, estamos em posição de importar o arquivo por meio do menu

Dados > Importar arquivos de dados > de arquivo texto, clipboard ou URL.

Tal como na importação de arquivos .xls/.xlsx, é necessário dar um nome para o conjunto de dados **sem usar espaços**. Feito isso e ajustando os separadores de campos e de decimais, basta clicar em *OK* e procurar pelo arquivo Turma3.txt, finalizando o processo de importação ao dar duplo clique no arquivo.



Note que, após realizar esse procedimento, o conjunto de dados ativo passa a ser este último (Turma_pequena), e não mais o anterior (Educa). No entanto, basta clicar no nome do conjunto de dados (no campo ao lado de “Conjunto de dados:”, acima da janela *R Script*) para alternar o conjunto de dados ativo.

4.3 Salvando/carregando conjuntos de dados

Qualquer conjunto de dados elencado como conjunto de dados ativo pode ser salvo no formato .RData por meio do menu *Dados > Conjunto de dados ativo > Salvar conjunto de dados ativo*.

Tal procedimento se torna interessante na medida que não será mais necessário importar novamente o conjunto de