

CAPÍTULO 1

Introdução ao ambiente R

Jéssica Karina da Silva Pachú, José Bruno Malaquias, Tatiane Caroline Grella

<https://doi.org/10.4322/mp.2020-12.c1>

Resumen

O R apresenta uma gama de ferramentas como testes paramétricos e não paramétricos, análise de regressão linear e não linear, análises de sobrevivência, análises multivariadas, produção gráfica, dentre outros. É importante mencionar que o R não é simplesmente um programa estatístico. No presente capítulo, apresentamos aspectos introdutórios ao ambiente R, como instalação do programa, carregamento e instalação de pacotes. Os comandos a serem executados no R estão sendo apresentados com a cor azul.

Palabras clave: linguagem R; apresentação; instalação; carregamento; pacotes.

1. Introdução

A estatística pode ser definida como a matemática aplicada aos dados de observação [1]. Através da Estatística conseguimos descrições claras, sintéticas e objetivas [2].

Em uma perspectiva entomológica, a estatística é uma condição essencial pois trabalha com métodos científicos para coleta, organização, resumo e apresentação de dados e também para obtenção de conclusões e a tomada de decisões.

As dificuldades enfrentadas por estudantes e pesquisadores entomologistas que precisam fazer uso da inferência estatística parecem decorrer da falta de livros didáticos, escritos em linguagem a eles adequada, livre de algebrismos usualmente empregados pelos estatísticos matemáticos. Para sanar essa lacuna, o presente livro tem como objetivo divulgar comandos básicos para análise descritiva, análise de variância e análise de deviance com dados entomológicos em linguagem R. A estratégia utilizada para alcançar esse objetivo consistiu na preparação de scripts com linguagem simples e com exemplos entomológicos.

O programa estatístico que tem mais auxiliado em análises estatísticas de dados biológicos é aquela baseada no ambiente R, que é uma linguagem de programação estatística, de acesso aberto, e que vem se tornando cada vez mais popular.

Em virtude de suas características, o *R Development Core Team* classifica como ambiente R [3].

O R é um software livre para análise de dados. Isto significa que ele pode ser utilizado, copiado, distribuído, alterado e melhorado de forma livre. Foi desenvolvido em 1996, com os professores de estatística Ross Ihaka e Robert Gentleman, eles trabalhavam e

eram colegas no departamento de estatística na Universidade de Auckland, os dois compartilhavam o interesse por estatística computacional e viram a necessidade de um melhor ambiente de software para o laboratório da área. Então, eles desenvolveram uma nova linguagem computacional, similar a linguagem S [3].

O software é um ambiente colaborativo internacional de desenvolvimento e pesquisa mantido formalmente pela R Foundation, disponibiliza uma ampla variedade de técnicas estatísticas e gráficas, incluindo modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento e outras. Você pode “baixá-lo” da Internet, de forma gratuita (www.r-project.org). Um conjunto básico de pacotes vem embutido na instalação do R, com muitos outros disponíveis na rede de distribuição do R (CRAN). Pesquisas e levantamentos com profissionais da área mostram que a popularidade do R aumentou substancialmente nos últimos anos [4]. Em 2017, o R já possui mais de 10.000 pacotes disponíveis [5]. Um pacote é um conjunto de funções que você consegue acessar em seus códigos, bastando importá-lo. Existem pacotes para resolver vários problemas como construção de gráficos, análise descritiva, análise de variância etc.

2. Motivos para utilizar o R

- Todos os códigos são abertos, reproduzíveis e adaptáveis;
- Compatível com Windows, Mac, Linux e etc;
- Utilizado para criar outros softwares;
- Empresas privadas e públicas estão se atentando ao R;
- Link com C, C++, Java entre outros;
- Ótima documentação (alterações de versões e help's);
- Comunidade acadêmica disseminando conhecimento;
- Altamente extensível;
- É continuamente suportado pela comunidade, com rápida correção de *bugs* e adição de novas funcionalidades.

3. Instalação do R

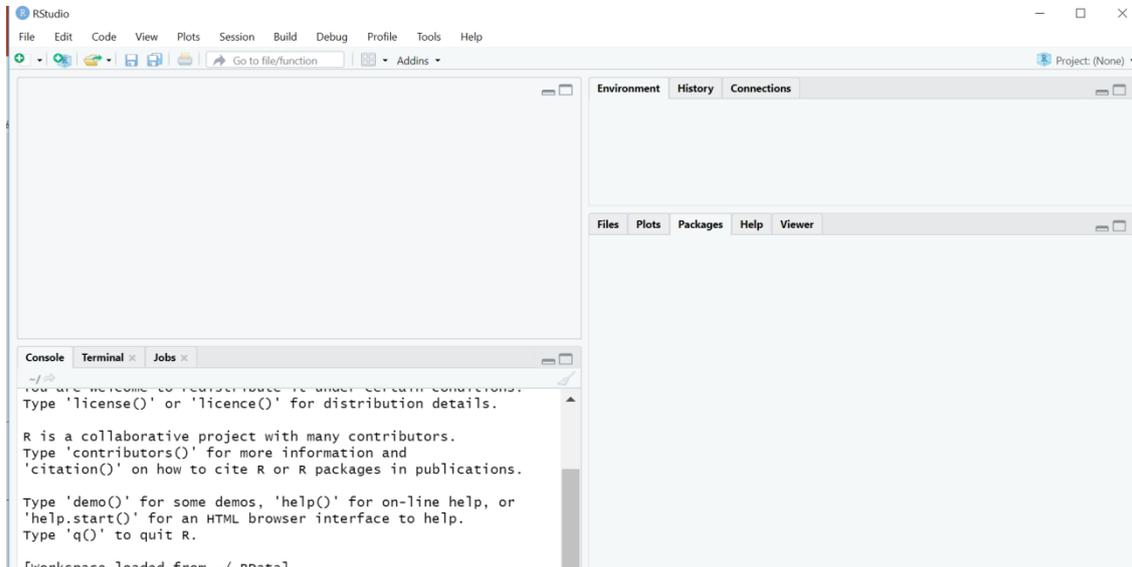
- Clique em CRAN;
- Escolha o espelho de sua preferência (CRAN mirrors);
- Clique em Windows;
- Clique em base e salve o arquivo do R para Windows;
- Depois é só executar o arquivo.

4. Console do software “R”



5. Usando o R como um ambiente de desenvolvimento integrado (IDEs)

RStudio é uma interface integrada para gráficos e cálculos estatísticos. Um acessório escrito na linguagem C++, está disponível em duas edições: RStudio Desktop, que roda localmente como um aplicativo desktop padrão; e RStudio Server, que permite acessar o RStudio usando um navegador web.



Para **baixar o RStudio** clique em:

<https://www.rstudio.com/products/rstudio/download/> e escolha opção *Rstudio desktop*.

6. Script

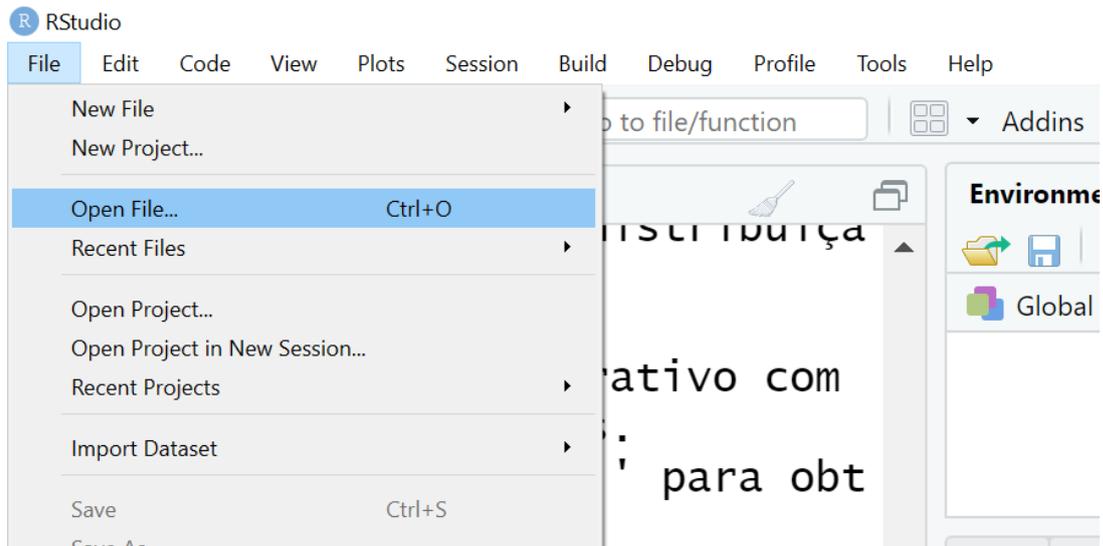
Para abrir e praticar no Rstudio com um novo script, vá em:

- File > New File > R Script

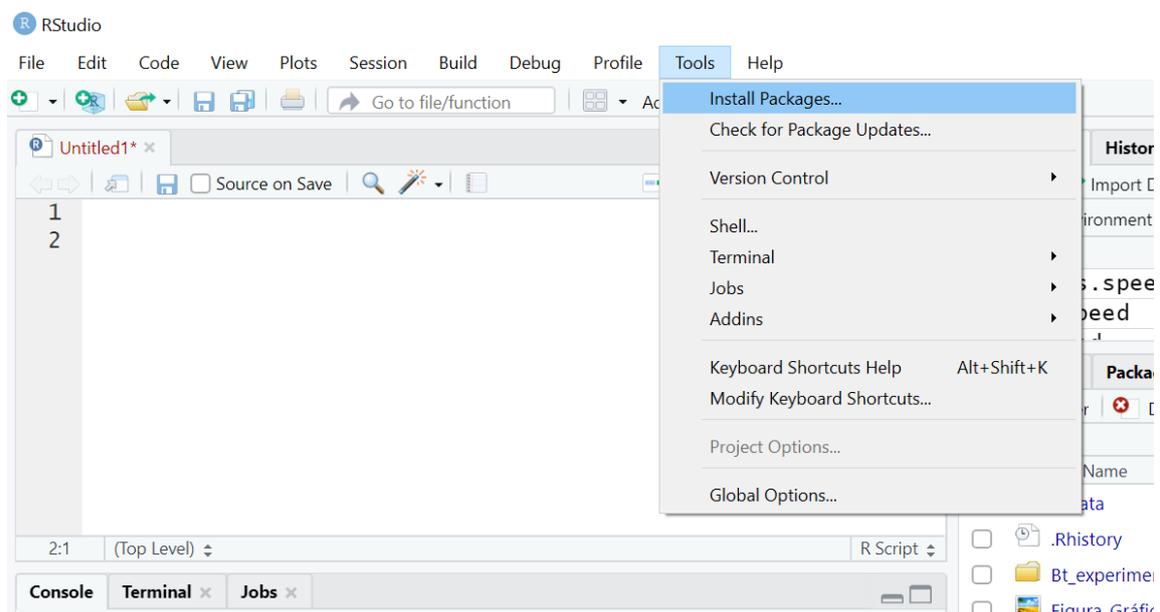
Se preferir abrir um script que já tenha salvo, vá em:

- Open File > selecione o arquivo > Abrir

No script você pode digitar comandos a serem executados e também comentários. Os comandos são escritos como no console e tudo que é escrito após **#** são considerados apenas como comentários.

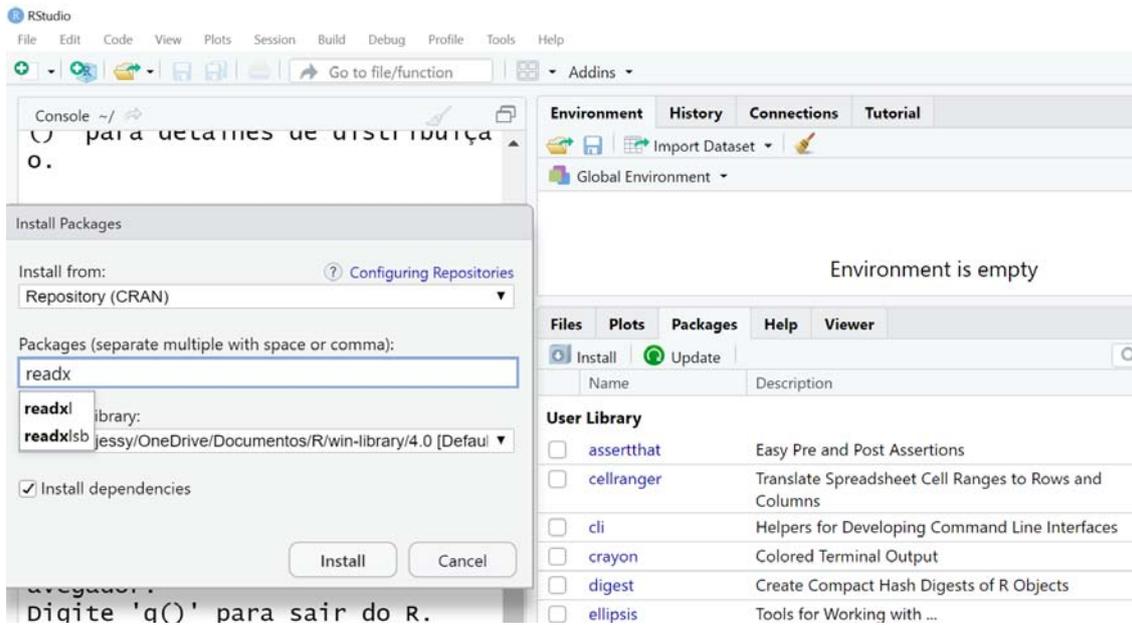


O R é um programa que ocupa pouco espaço e memória e geralmente roda rápido, isso porque os pacotes vêm na instalação base, ou seja, configurações mínimas para funcionar. Para realizar tarefas mais complexas pode ser necessário instalar pacotes adicionais (packages), esse é um dos grandes benefícios do R, seu grande acervo de pacotes. Qualquer pessoa pode enviar um pacote e todo código enviado está disponível na internet. Existe, porém, um processo de avaliação que o código passa e certas normas rígidas devem ser respeitadas sobre o formato do código, o manual do usuário e a forma de atualização do pacote.



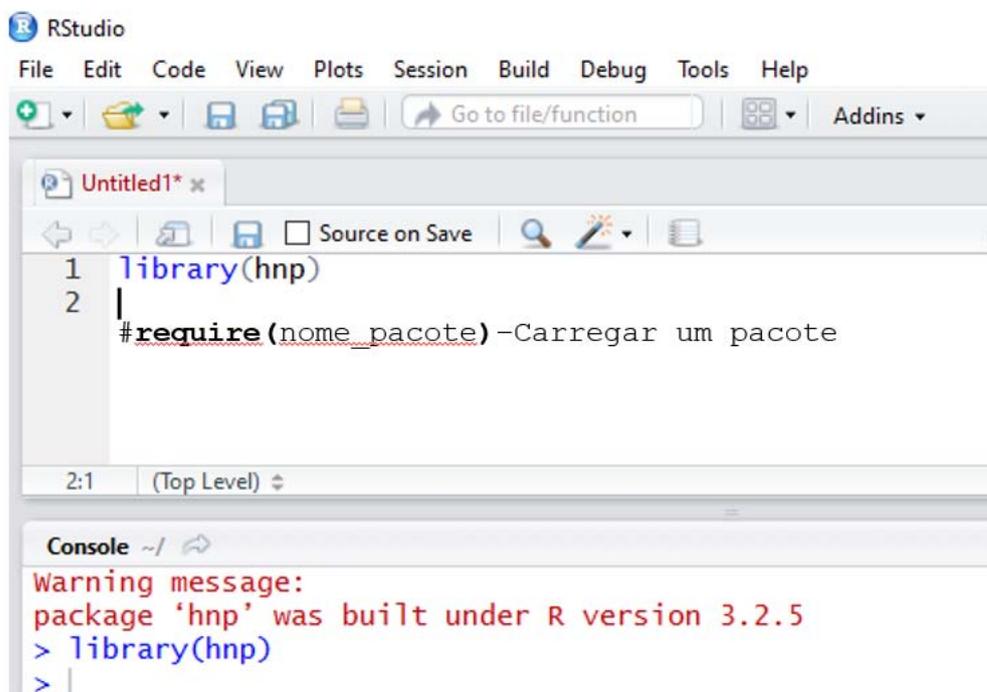
Outra forma é, no lado direito da tela, clicar em:

- Packages>Install> digitar o nome do pacote e instalar.



7. Usando o R como um ambiente de desenvolvimento integrado (IDEs)

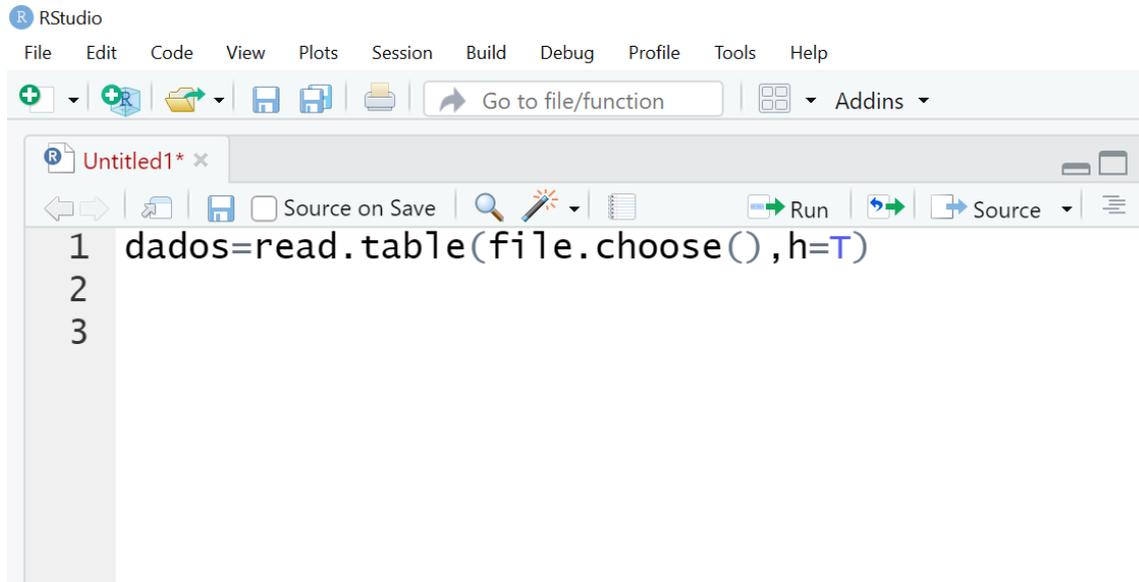
Para carregar um pacote dentro de um script, utilizamos a função `library()` ou `require()`, dentro do parêntese o nome do pacote que deverá ser carregado.



8. Lendo um arquivo

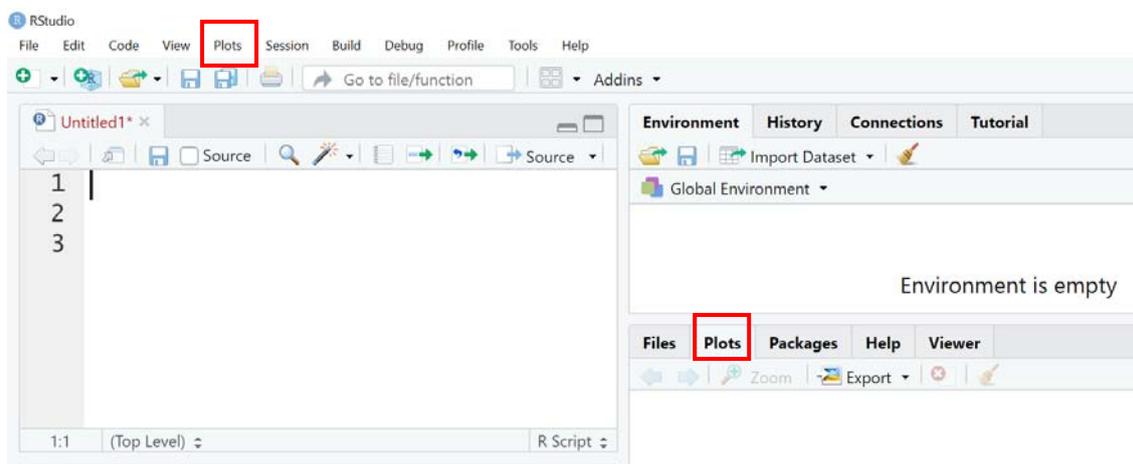
Para ler um banco de dados utilize a função `read.table` para fazer a importação dos dados.

A linha de comando exemplificada `dados=read.table(file.choose(),h=T)` direciona você à pasta que contém o seu banco de dados no ambiente R (dados), a função para fazer a importação dos dados (read.table) e a função para escolher a pasta na qual seus dados estão (file.choose).



9. Visualizando um Plot

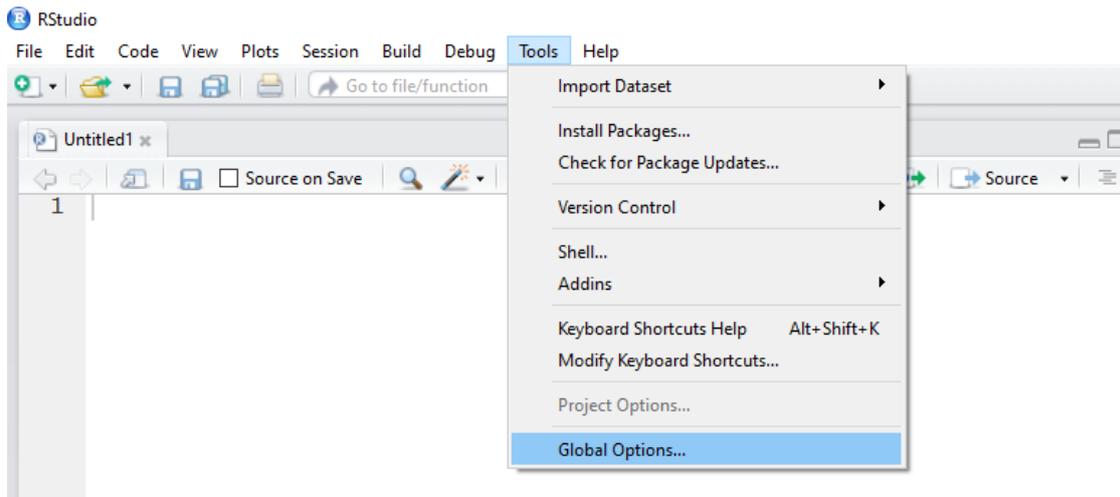
Para visualizar seus gráficos, clique na aba plots na linha superior do console ou na aba plots no lado direito, dessa maneira você pode usar a vassoura para limpar ou zoom para aumentar o seu plot.



10. Customize seu Rstudio

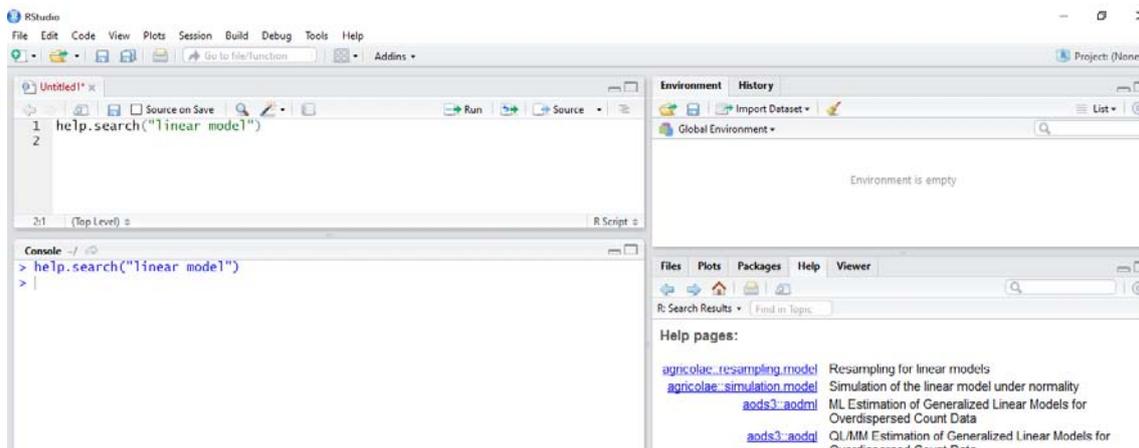
Nessa opção você pode alterar e personalizar o Rstudio de acordo com suas preferências, por exemplo, modificar a fonte, o tamanho, a cor etc. Clique em:

- Tools> Global Options>



11. Buscando ajuda

Existem algumas maneiras para obter ajuda sobre o funcionamento de comandos e pacotes. Por exemplo, clicando na aba Help ou digitando a linha de comando `help.search()`



12. Algumas funções e comandos básicos

Para visualizar o banco de dados

View ()

Limpar o console

Ctrl+L

Para ler as linhas de comando

Ctrl+Enter ou Ctrl+R

Mudança de diretório

Ctrl+Shift+H:

getwd()

Para visualizar os nomes dos arquivos que tem dentro da pasta escolhida para trabalhar:

`list.files()`

13. Referências dos pacotes utilizados

- [1] Pimentel-Gomes F. Curso de estatística experimental. 15. Ed. Piracicaba: FEALQ, 2009.
- [2] Nazareth H. Curso básico de estatística. Ática, São Paulo, 2003.
- [3] Peternelli L.A., Mello M.P. Conhecendo o R: uma visão estatística. Viçosa: UFV, v. 1, 2011.
- [4] Muenchen R.A. The popularity of data analysis software. URL <http://r4stats.com/popularity>, 2012.
- [5] Smith D. CRAN now has 10,000 R packages. Here's how to find the ones you need, 2017.

14. Referências recomendadas

Borcard D., Gillet F., Legendre P. Numerical ecology with R. Springer, 2018.

British Ecological Society. A guide to data management in ecology and evolution. 2014.

Crawley M.J. The R book. John Wiley & Sons, 2012.

Matloff Norman. The art of R programming: A tour of statistical software design. No Starch Press, 2011.

Peternelli L.A., Mello M.P. Conhecendo o R: uma visão estatística. Viçosa: UFV, v. 1, 2011.

Autores

Jéssica Karina da Silva Pachú*, Departamento de Entomologia e Acarologia - LEA Avenida Pádua Dias, 11 - CEP 13418-900 - Piracicaba/SP, Brasil

José Bruno Malaquias, Instituto de Biociências - Câmpus de Botucatu. R. Prof. Dr. Antônio Celso Wagner Zanin, 250 - Distrito de Rubião Junior - Botucatu/SP, Brasil - CEP 18618-689

Tatiane Caroline Grella, Biologia Celular e Molecular, Universidade Estadual Paulista "Júlio de Mesquita Filho" – Laboratório de Ecotoxicologia e Conservação de Abelhas (LECA) - Avenida 24 A,1515 - Bela Vista - CEP 13506-900 – Rio Claro/SP, Brasil

* Autor para correspondência: jessikapachu@gmail.com