

Introdução à linguagem R para manipulação e visualização de dados

Semana de Estudos da Ecologia - Unesp Rio Claro

Maurício Vancine

10-11/11/2021

Resumo

O minicurso apresentará os fundamentos da linguagem R e IDE RStudio para a manipulação e visualização de dados através do *tidyverse*. O *tidyverse* é um conjunto de pacotes e funções que compartilham uma mesma estrutura visando ajudar os usuários a criar um código eficiente e ‘organizado’. Os pacotes são focados principalmente na organização de dados (leitura, filtragem, organização, seleção e resumo) e visualização de dados. Serão abordados os seguintes temas: (1) funcionamento da linguagem R, (2) estrutura e manipulação de objetos, (3) tidyverse e (4) visualização de dados. Ao final do minicurso espera-se que os participantes adquiram conhecimentos introdutórios do funcionamento da linguagem, assim como na manipulação e visualização de dados tabulares.

Carga horária

06 horas

Repositório do GitHub

<https://github.com/mauriciovancine/workshop-r-introduction>

Pré-requisitos

Conhecimentos básicos de programação são bem-vindos, mas não são exigidos.

Ministrante

Maurício Vancine é Bacharel em Ecologia, Mestre em Zoologia e Doutorando em Ecologia, Evolução e Biodiversidade pela UNESP de Rio Claro. Possui experiência em Ecologia Espacial, Ecologia da Paisagem, Modelos de Distribuição de Espécies e Ecologia de Anfíbios.

Mais informações:

- Website: <https://mauriciovancine.github.io/>
- Twitter: <https://twitter.com/mauriciovancine>
- GitHub: <https://github.com/mauriciovancine>

Conteúdo

1 Funcionamento da linguagem R (1 h)

1. Linguagem R
2. RStudio
3. Console
4. Scripts
5. Operadores
6. Objetos
7. Funções
8. Pacotes
9. Ajuda
10. Ambiente
11. Citações
12. Principais erros
13. Principal material de estudo

2 Estrutura e manipulação de objetos (2 h)

1. Atributos dos objetos
2. Manipulação de dados unidimensionais
3. Manipulação de dados multidimensionais
4. Valores faltantes e especiais
5. Diretório de trabalho
6. Importar dados
7. Conferência de dados importados
8. Exportar dados

3 tidyverse (3 h)

1. Contextualização
2. tidyverse
3. here
4. readr, readxl e writexl
5. tibble
6. magrittr (pipe - %>%)
7. tidyr
8. dplyr
9. stringr
10. forcats

4 Visualização de dados (2 h)

1. Contextualização
2. Principais pacotes para gráficos
3. Gramática dos gráficos
4. Principais livros e sites
5. Principais tipos de gráficos
6. Histograma e Densidade
7. Gráfico de setores
8. Gráfico de barras

9. Gráfico de caixas
10. Gráfico de dispersão
11. Gráfico pareado
12. Combinando gráficos
13. Gráficos animados
14. Gráficos interativos
15. Gráficos usando uma interface

Bibliografia

- Adler J. 2012. R in a nutshell: A desktop quick reference. O'Reilly Media, Inc.
- Allesina S, Wilmes M. 2019. Computing Skills for Biologists: A Toolbox. Princeton University Press.
- Beckerman AP, Childs DZ, Petchey OL. 2017. Getting started with R: an introduction for biologists. Oxford University Press.
- Braun WJ, Murdoch DJ. 2016. A first course in statistical programming with R. Cambridge University Press.
- Burns P. 2011. The R inferno. Lulu.com.
- Chambers J. 2008. Software for data analysis: programming with R. Springer Science & Business Media.
- Chambers JM. 2017. Extending R. Chapman and Hall/CRC.
- Campbell M. 2019. Learn RStudio IDE: Quick, Effective, and Productive Data Science. Springer.
- Chang W. 2018. R graphics cookbook: practical recipes for visualizing data. O'Reilly Media.
- Cotton R. 2013. Learning R: A Step-by-Step Function Guide to Data Analysis. O'Reilly Media, Inc.
- Cotton R. 2017. Testing R Code. Chapman and Hall/CRC.
- Crawley MJ. 2012. The R book. John Wiley & Sons.
- Davies TM. 2016. The Book of R: A first course in programming and statistics. No Starch Press.
- Engel C. 2019. Introduction to R. <https://cengel.github.io/R-intro/>
- Field A, Miles J, Field Z. 2012. Discovering statistics using R. Sage publications.
- Gandrud C. 2016. Reproducible research with R and R studio. Chapman and Hall/CRC.
- Gardener M. 2012. The essential R reference. John Wiley & Sons.
- Gentleman R. 2008. R programming for bioinformatics. Chapman and Hall/CRC.
- Gillespie C, Lovelace R. 2016. Efficient R programming: a practical guide to smarter programming. O'Reilly Media, Inc.
- Grolemund G. 2014. Hands-On Programming with R: Write Your Own Functions and Simulations. O'Reilly Media, Inc.
- Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media. <https://web.stanford.edu/~hastie/ElemStatLearn/>
- Healy K. 2018. Data visualization: a practical introduction. Princeton University Press.
- Hector A. 2015. The new statistics with R: an introduction for biologists. Oxford University Press.
- Hilfiger JJ. 2015. Graphing Data with R: An Introduction. O'Reilly Media, Inc.
- Holmes S, Huber W. 2018. Modern statistics for modern biology. Cambridge University Press. <http://web.stanford.edu/class/bios221/book/index.html>

- Irizarry RA, Love MI. 2016. Data Analysis for the Life Sciences with R. Chapman and Hall/CRC. <http://www.rwde2.com/files/rafa.pdf>
- Ismay C, Kim AY. 2019. Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse. CRC Press. <https://moderndive.com/>
- James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning (Vol. 112, p. 18). New York: springer. <http://faculty.marshall.usc.edu/gareth-james/ISL/>
- Jones O, Maillardet R, Robinson A. 2014. Introduction to scientific programming and simulation using R. Chapman and Hall/CRC.
- Kabacoff R. 2015. R in Action: Data Analysis and Graphics With R. Greenwich, CT.
- Lander JP. 2017. R for everyone: advanced analytics and graphics. Pearson Education.
- Maindonald J, Braun J. 2010. Data analysis and graphics using R: an example-based approach. Cambridge University Press.
- Matloff N. 2011. The art of R programming: A tour of statistical software design. No Starch Press.
- McGrath M. 2018. R for Data Analysis in easy steps - R Programming essentials. In Easy Steps Limited.
- Micheaux PL, Drouilhet R, Lique B. 2013. The R software The R Software: Fundamentals of Programming and Statistical Analysis. Springer Science & Business Media.
- Oliveira PF, Guerra S, McDonnell R. 2018. Ciência de dados com R – Introdução. IBPAD. <https://cdr.ibpad.com.br/>
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Racine JS. 2012. RStudio: a platform-independent IDE for R and Sweave. Journal of Applied Econometrics, 27(1), 167-172.
- Shahbaba B. 2012. Biostatistics with R: An Introduction to Statistics Through Biological Data. Springer
- Silge J, Robinson D. 2017. Text mining with R: A tidy approach. O'Reilly Media, Inc.
- Teetor P. 2019. R cookbook: proven recipes for data analysis, statistics, and graphics. O'Reilly Media, Inc.
- Wickham H. 2015. R packages: organize, test, document, and share your code. O'Reilly Media, Inc.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.
- Wickham H. 2019. Advanced r. Chapman and Hall/CRC.
- Wickham H, Grolemund G. 2016. R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>
- Wiley M, Wiley, JF. 2016. Advanced R Data Programming and the Cloud. Apress.
- Wilke CO. 2019. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media. <https://serialmentor.com/dataviz/>
- Xie Y, Allaire, JJ, Grolemund G. 2018. R markdown: The definitive guide. CRC Press. <https://bookdown.org/yihui/rmarkdown/>
- Zumel N, Mount J. 2014. Practical data science with R. Manning Publications Co.
- Zuur A, Ieno EN, Meesters E. 2009. A Beginner's Guide to R. Springer Science & Business Media.