Losing your fear of using R for statistical analysis

Iara Shimizu[1,2], Juliana Carvalho Ferreira[2,3]

WHAT IS R?

R is a programming language widely used in health research, as it provides a vast collection of software packages that encompass a wide range of data analysis techniques to conduct from simple to complex statistical analyses, to create graphics and figures, and to design websites and apps.

Because it is a programming language, it requires input through a command line, which may seem intimidating to nonprogrammers. However, even if you have never programmed before, there are several tips to get started and gradually learn how to use R, including preset software packages that extend the capabilities of basic R and allow users to perform specialized tasks without having to write all of the code from scratch.

ADVANTAGES OF R

Being an open-source program, R facilitates changes in analyses and ensures reproducibility of results in new datasets, allowing you to document and share your analyses in a systematic and organized way, in addition to the possibility of creating reproducible reports that combine code, text, and visualizations in a single document.

For researchers who do not want to use commercial statistical software, R is an option because it is free and adaptable to different operating systems.

One of the greatest advantages of R is its online support, since it has a very extensive and active community of users around the world, who continually develop new functionalities for the program and offer solutions to the various questions that may arise.

GETTING STARTED WITH R AND RSTUDIO

To get started with R programming for statistical analysis in health research, you can follow these steps, summarized in Figure 1.

1. Installing R and RStudio—Install R and RStudio on your computer. R is a programming language used for statistical computing and graphics, and RStudio is a program designed for working with R programming language, providing a user-friendly interface. Visit the official R website[1] and download the appropriate version for your operating system. There is also the possibility of using an online version that does not require installation of any software.[2]

2. Importing data into R—This involves extracting data from a file or database and importing them into an R data frame. You can import data into R from various sources, such as CSV files, Excel files, or databases, and manage the data by filtering, sorting, merging, or transforming datasets. If you do not have data of your own, R has a list of open data sets that can be used to gain hands-on experience and improve programming skills.

3. Learning R coding basics—Get familiar with the basic language of R. There are several free online tutorials, books, and resources available for learning R, such as the "Hands-On Programming with R"[3] for programming beginners. To use the R program, you enter the instructions, known as commands, which direct R to perform a specific task, such as calculate the mean of a variable or perform a t-test to compare two groups. If you type a command that R does not recognize, it will return an error message. If that happens, do not panic! Read the error message to understand the problem, review the command that you have typed for any mistakes or syntax errors, or search for solutions online using help pages in R or in communities and forums.

4. Statistical analysis—R has a system with a varied number of packages designed specifically for statistical analysis, such as basic R functions, stats package, survival analysis package, and more specialized packages; packages not included in
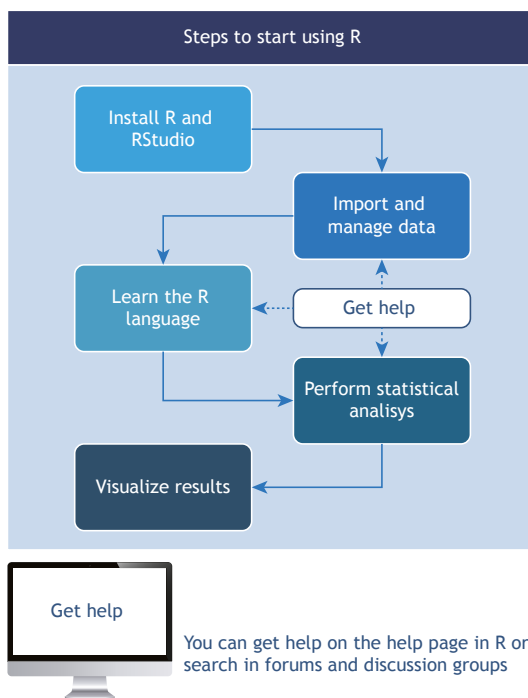

Figure 1. Steps to start using R.

1. Coordenação de Fisioterapia, Centro de Ciências da Saúde, Universidade Estadual do Piauí – UESPI – Teresina (PI) Brasil.
2. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.
3. Methods in Epidemiologic, Clinical, and Operations Research-MECOR-program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.

the basic software need to be installed. However, basic statistical analysis can be performed with a few simple and easy-to-learn commands.

5. Visualization of results—R provides a diverse range of packages that enable the generation of high-quality figures and charts. Researchers can create figures and charts that facilitate a deeper understanding of the data and aid in the communication of key results.

6. Getting help—Each R function comes with its own help page that you can access by typing the name of the function preceded by a question mark. R community forums and discussion groups allow you to submit a question or search through previously answered questions. Participating in the community will expose you to different perspectives, new techniques, and useful resources.

In conclusion, learning R programming for statistical analysis is like learning a new language: it may seem somewhat difficult in the beginning, but as you learn, it becomes easier. We recommend that, as a new user, you start with small projects to gradually build your skills and explore advanced techniques as you go.

REFERENCES

1. The R Foundation [homepage on the Internet]. Vienna, Austria: R Foundation for Statistical Computing; [cited 2023 Jun 17]. The R Project for Statistical Computing. Available from: https://www.r-project.org

2. posit Cloud [homepage on the Internet].Boston, MA: posit; c2023 [cited 2023 Jun 17]. Friction free data science. Available from: https://posit.cloud

3. Grolemund G. Hands-On Programming with R [monograph on the Internet]. Sebastopol, CA: O'Reilly Media; 2014. [cited 2023 Jun 17]. Available from: https://rstudio-education.github.io/hopr/