

Software livre R

Curso Introdutório e Básico

Michelle Almeida/Maurício Lordêlo

Universidade Estadual de Feira de Santana - Bahia - Brasil

20 de agosto de 2022

Introdução ao uso do R

- Linguagem de programação livre, de código aberto, tem muitos recursos e uma ampla comunidade.
- Originalmente criado para computação estatística e gráficos (propósito específico)
- Conserva esse propósito até hoje, porém pode ser usado para outras tarefas não relacionada com análise estatística
- A lógica é:
 - Envia a instrução via linha de comando (CLI - *Command Line Interface*)
 - Recebe via REPL (*Read, Evaluate, Print and Loop*)
- Site oficial r-project.org
- Pode ser conectado com outros ambientes computacionais facilitando a interação

Introdução ao uso do R

- Desenvolvido por Robert Gentleman e Ross Ihaka (iniciais dos nomes) na University of Auckland, New Zealand em 1993
- Motivação: dar aulas de Estatística
- Com outros professores formaram o R CORE TEAM
- Esforço colaborativo com contribuição de pesquisadores de todo o mundo
- Para citar o R em publicações:
R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Introdução ao uso do R

- Multiplataforma: disponível para Windows, Linux e Mac OS X
- Formado por “pacotes” (*packages*)
- 18489 pacotes oficiais (dados de agosto/2022)
- Alguns pacotes já são inseridos na primeira instalação
- Outros pacotes podem ser instalados posteriormente
- Ampliam os recursos do R

Introdução ao uso do R

- Grande popularidade na área acadêmica e científica: universidades, institutos de pesquisa, órgãos governamentais, etc
- Com a *era da informação* ou (*Big Data*) : expansão para empresas privadas
- Construção de uma infinidade de tipos de gráficos
(<https://www.r-graph-gallery.com/>)
- Dentre outras opções: analisar textos, dados de redes sociais e conversas de aplicativos de mensagens; realizar edição de imagens; construir relatórios, artigos, livros, sites...; criar e reproduzir músicas além de analisar dados de plataformas (como Spotify), etc.

- IDE (*Integrated Development Environment*), traduzido como “Ambiente de Desenvolvimento Integrado”, criado para facilitar a “comunicação” do usuário com o R
- Em outubro, passará a ser chamado de **POSIT**
- É a IDE mais popular para trabalhar com R
- O R pode funcionar em outros ambientes
- R é linguagem de programação, RStudio é IDE (não confundir!)
- Mantida pela empresa RStudio: rstudio.com
- Devidamente preparada para trabalhar com todas as funcionalidades do R

O curso

- Os procedimentos que serão explorados:
 - criação de objetos;
 - entrada de dados;
 - leitura de dados fora do R;
 - instalação de pacotes (para aumentar as funcionalidades);
 - importação e preparação da base de dados;
 - selecionar, estruturar, filtrar, criar, converter, renomear, etc.
 - cálculo de medidas estatísticas
 - construção de gráficos (visualização é a ferramenta da comunicação)
 - apresentação em forma de relatório dinâmico (*Rmarkdown* passará a ser chamado **QUARTO**)

“O maior valor de um gráfico é quando ele nos força a perceber o que nós
jamais esperávamos ver”

“O gráfico simples trouxe mais informações à mente dos analistas de dados
do que qualquer outro dispositivo”

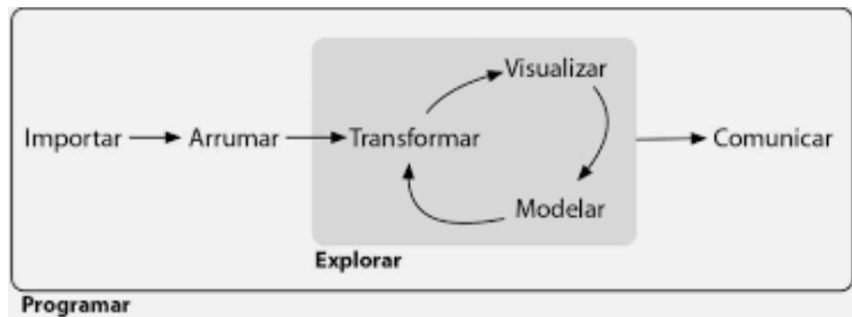
(John Tukey)

Pacote *tidyverse*

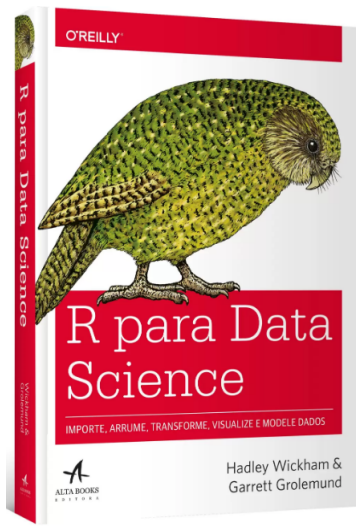


Ciência de dados (*Data Science*)

Dados: nova riqueza mundial



Ciência de dados



Para finalizar

- Dicas importantes
 - Aprendizado gradativo, de acordo com a necessidade, sem “atropelar” as etapas
 - Buscar ajuda para tópicos específicos no próprio R, em manuais, tutoriais, etc
 - Cada etapa vencida, pode ser usada em outras análises
 - Na programação qualquer detalhe pode fazer muita diferença (falta ou excesso de vírgula, parênteses, ponto, etc.)
 - Não desanime com as mensagens de erros e nem com o inglês
 - Para iniciantes (do zero): insira bastante comentários nos códigos pois ajudam no momento de reaproveitá-los
- Há um “mundo” de possibilidades a ser explorado, sempre novos desafios
- Detalhe: normalmente um procedimento realizado no R (como calcular medidas descritivas, criar um gráfico, realizar um teste de hipóteses, regressão linear, etc) pode ser realizado de mais de uma maneira