

Conhecendo o R

Quatro coisas importantes

Ou o que eu quero que vocês saiam daqui sabendo...

- O que é o R
- Como instalar o R
- Algumas coisas que se faz no R
- Onde aprender o R

Sumário

- O R: seu contexto e particularidades
- Um pouco sobre sua linguagem de programação
- Ferramentas e recursos

O R: seu contexto e particularidades

“R is ‘GNU S’, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.”

- site do cran.r-project.org

Um pouco da história do R

Ou tudo que eu sei sobre a história dele...



O R foi criado em 1993 na Universidade de Auckland por Ross Ihaka and Robert Gentleman como uma implementação da linguagem de programação “S”.

A linguagem “S” por sua vez foi criada em 1976 por John Chambers enquanto trabalhava na Bell Labs.

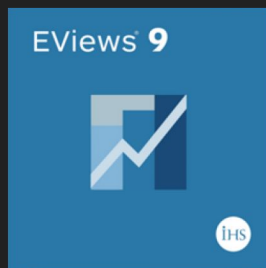
A maioria das funções do R é escrita em R; operações que demandam mais esforço computacional são implementadas em C, C++ e Fortran.

O CRAN e a R-Foundation tem um papel essencial na divulgação do software R.

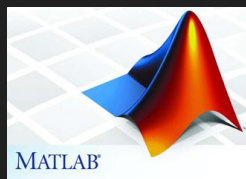
"to turn ideas into software, quickly and faithfully"

- John Chambers

Alguns dos competidores e suas vantagens



THE
POWER
TO KNOW®



Uma comparação bem feita das capacidades estatísticas dos principais desses softwares pode ser encontrada em: http://stanfordphd.com/Statistical_Software.html

Pacotes pacotes e pacotes...

Na comunidade de desenvolvedores do R, a contribuição coletiva para melhorar o software é constante.

Esses pacotes variam desde implementações de modelos estatísticos, manipulação de dados, criação de gráficos, desenvolvimento de aplicações, lidar com fontes e tipos de dados específicos...

Em 2009¹ haviam aproximadamente 2000 pacotes disponíveis no site do CRAN-R.

1. Fonte: R Journal Vol 1/2

E ontem a noite:

8078

pacotes disponíveis no CRAN-R!!

Como instalar...

O arquivo para instalação do R, usando o mirror aqui da USP está disponível no seguinte link:

<http://www.vps.fmvz.usp.br/CRAN/>

Além disso, um guia completo para a instalação do R em português foi elaborado por dois ex-alunos aqui do IME e também esta disponível via CRAN-R no seguinte link:

<https://cran.r-project.org/doc/contrib/ltano-installation.pdf>

Rstudio



Um IDE muito usado para deixar a utilização da linguagem R um pouco mais amigável com o usuário (vulgo nós).

Também contribuem de forma intensiva na criação de muitos pacotes úteis na vida com o R. Alguns deles são:



Todos esses pacotes estão na rede CRAN-R.

Um pouco sobre sua linguagem de programação

Suas estruturas de dados

Homogênea	Heterogênea
vector	list
matrix	data frame
array	

Coisas que eu fico impressionado

Limpando dados com `dplyr`, `tidyr`, `lubridate`: dados do UCI HAR Data Set*

Criação de funções funcionais: Hospital Compare Downloadable Database

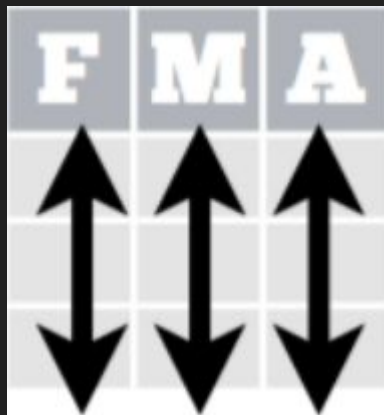
Subsetting: `airquality`

Tabelas: `airquality`

Visualizações: `Diamonds` + `airquality`

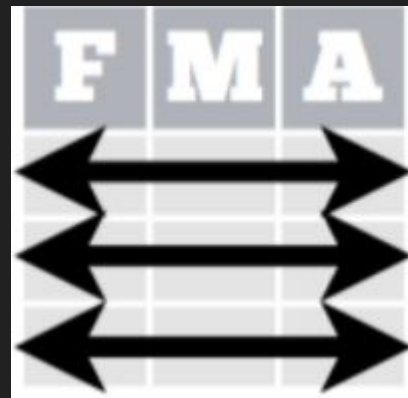
Dados limpos/Tidy data

Em um banco
de dados limpo:



Cada **variável** tem sua
coluna

&



Cada **observação** está
em uma **linha**

“Tidy data sets are all alike; every messy data set is messy in its own way.” - Hadley Wickham

O artigo de Hadley Wickham, da Rice University “Tidy data” publicado no “*Journal of Statistical Software*” pode ser encontrado em: <http://vita.had.co.nz/papers/tidy-data.pdf>

Coisas que eu fico impressionado

Limpando dados com dplyr, tidyr, lubridate: dados do UCI HAR Data Set*

Criação de funções funcionais: Hospital Compare Downloadable Database

Subsetting: airquality

Tabelas: airquality

Visualizações: Diamonds + airquality

Ferramentas e recursos

Open source e o Git Hub



GitHub é um Serviço de Web Hosting Compartilhado para projetos que usam o programa de controle de versões Git.

Ótima ferramenta para compartilhamento de programa, com uma comunidade de usuários de R muito ativa.

Os códigos usados ao vivo estão disponíveis em minha página:

<https://github.com/GuiMarthe/IntroR>

Swirl

Aprendendo R no R...



Criado por Nick Carchedi e Brain Caffo do departamento de Bioestatística da Johns Hopkins University.

É um pacote (open source) do R que se propõe a ensinar a linguagem de programação do R no próprio R.

Mais informações de como instalar o swirl e os cursos disponíveis no seguinte link:

<http://swirlstats.com/students.html>

Referências



Contúdo em português no CRAN disponível [aqui](#):

- “Bioestatística usando R” by Colin Robert Beasley (Data).
- “Introdução à Biometria utilizando R” by Leandro R. Monteiro and José Louvise Gomes-Jr (Data).
- “Introdução à Programação em R” by Luis Torgo .
- “Tópicos de Estatística utilizando R” by Fernando Itano (2007-08-21).
- “Guia de instalação do R” by Fernando Itano (2007-08-21).
- “Estatística aplicada à ecologia usando o R” by Diogo Borges Provete (Materials, 2011-08-08)
- “Introdução ao uso do programa R” by Victor Lemes Landeiro (Materials, 2011-08-25).

5 livros para aprender R. Post no [blog](#) do Instituto Brasileiro de Pesquisa e análise de Dados.

Livro do Carlos Cinelli “Introdução à análise de dados com R” (em construção). Disponível [aqui](#)

Referências

Referências em inglês:

[Introduction to R](#) - W. N. Venables, D. M. Smith and the R Core Team

[R Books For Undergraduate Students](#) - Daniel Hocking

[Awesome R](#) - uma lista com mais de 400 referências no Git Hub - ?qinwf

[R-bloggers](#) - bastante coisa nesse portal de curadoria de blogs de R

[R twotorials](#) - uma coleção de videos com menos de 3 min ensinando R

[Quick R](#) - um web livro muito bem recomendado

[TryR!](#) - uma aplicação de web que ensina R (do Code Academy)

[Data Camp](#) - uma plataforma de ensino de R e Python para análise de dados

Quatro coisas importantes

Ou o que eu quero que vocês saiam daqui sabendo...

- O que é o R
- Como instalar o R
- Algumas coisas que se faz no R
- Onde aprender o R

Valeu!!