

# Introdução

R, história, pacotes, RStudio e primeiro contato

Prof. Me. Lineu Alberto Cavazani de Freitas

Métodos Estatísticos em Pesquisa Científica  
Apoio computacional em linguagem R

Departamento de Estatística  
Laboratório de Estatística e Geoinformação



# Estatística e o desenvolvimento computacional

- ▶ A popularização da Estatística se dá graças ao **desenvolvimento computacional**.
- ▶ Os computadores pessoais tornaram os métodos estatísticos mais acessíveis ao público geral por meio de **softwares** que implementam as metodologias.
- ▶ Devido ao avanço computacional, houve um aumento considerável na capacidade de produzir e armazenar dados provenientes das mais diversas fontes.

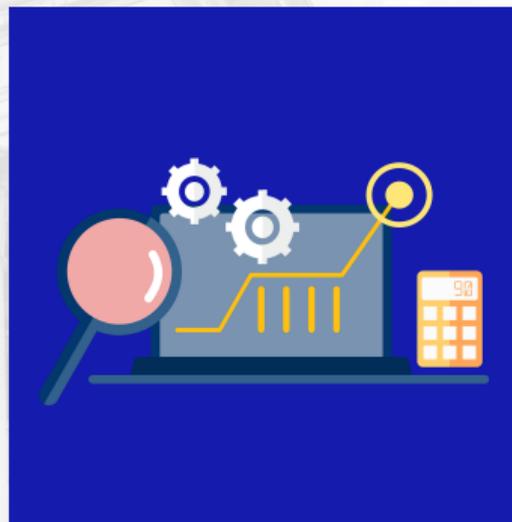


Figura 1. Extraído de pixabay.com.

# Estatística e o desenvolvimento computacional

- ▶ Graças ao avanço computacional podemos lidar com a manipulação de **grandes conjuntos de dados**.
- ▶ Este grande volume de dados também força o **desenvolvimento dos métodos estatísticos** e softwares para análise de dados.
- ▶ A capacidade computacional atual também desperta o interesse por **métodos estatísticos computacionalmente intensivos**.



Figura 2. Extraído de pixabay.com.

# Ferramentas para análises estatísticas

Existem diversas ferramentas disponíveis:

- ▶ R;
- ▶ Python;
- ▶ SAS;
- ▶ Spss;
- ▶ Biostat;
- ▶ Minitab;
- ▶ Tableau;
- ▶ Stata;
- ▶ E diversas outras.

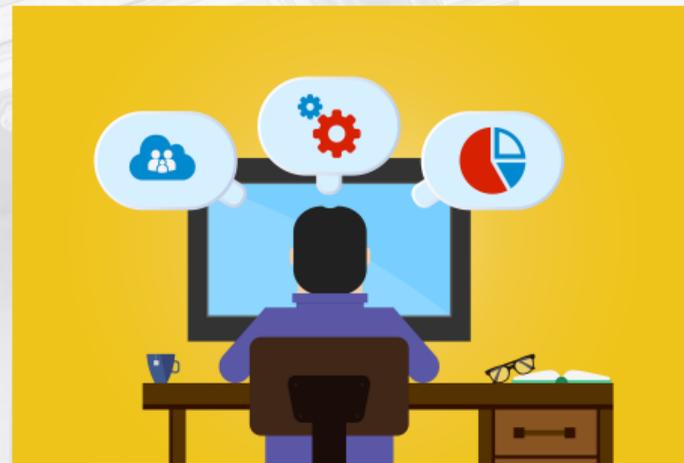


Figura 3. Extraído de pixabay.com.

# Ferramentas para análises estatísticas

Existem diversas ferramentas disponíveis:

- ▶ **R**;
- ▶ Python;
- ▶ SAS;
- ▶ Spss;
- ▶ Biostat;
- ▶ Minitab;
- ▶ Tableau;
- ▶ Stata;
- ▶ E diversas outras.



Figura 4. Logo do R.

- ▶ R é uma linguagem e ambiente para **computação estatística** e **gráficos**.
- ▶ É **livre** e de **código aberto**.
  - ▶ Livre (free): usuários tem liberdade de:
    1. executar como desejar e para qualquer propósito.
    2. estudar o funcionamento e adapta-lo à necessidades específicas.
    3. distribuir cópias de versões originais e modificadas.
  - ▶ Código aberto (open source): o acesso ao código fonte é gratuito.

- ▶ Muito popular no meio **acadêmico** e tem uso cada vez maior no meio **corporativo**.
  - ▶ É acessível e gratuito.
  - ▶ Tem diversas técnicas e aplicações possíveis.
- ▶ Tem potencial uso em todas as etapas do processo de análise de dados.
  - ▶ Obtenção, importação, manipulação e tratamento.
  - ▶ Análise exploratória.
  - ▶ Ajuste de modelos estatísticos, modelos de aprendizado de máquina, dentre outros.
  - ▶ Elaboração de relatórios dinâmicos e reproduzíveis.

# Antes do R

- ▶ O R sucedeu a linguagem S, de **John Chambers**, Rick Becker, Trevor Hastie, Allan Wilks e outros.
- ▶ A linguagem S foi iniciada em 1976 e disponibilizada publicamente no início dos anos 80 já como um ambiente de análise estatística.
- ▶ Uma das principais limitações da linguagem S era que ela só estava disponível em um **pacote comercial**: o S-PLUS. O que motivou a criação do R.
- ▶ Quando o R foi desenvolvido a sintaxe era muito semelhante ao S, pensando na migração dos usuários de uma linguagem para outra.

# História do R

- ▶ Em **1991** o R foi criado por Ross Ihaka e Robert Gentleman (“**R & R**”) no Departamento de Estatística da Universidade de Auckland.
- ▶ Em **1993** o R foi **anunciado publicamente** pela primeira vez.
- ▶ Em **1995** o R passou a usar a **GNU General Public License**, o que tornou o R um software livre e de código aberto.



Figura 5. Robert Gentleman e Ross Ihaka.

# História do R

- ▶ Em **1996** foi publicado o **artigo** em que Ross e Robert descrevem sua proposta: “*R: A language for data analysis and graphics.*” no Journal of Computational and Graphical Statistics.
- ▶ Ainda em **1996** foram criadas as listas de discussão **R-help** e **R-devel**.

## **R: A Language for Data Analysis and Graphics**

ROSS IHAKA and Robert GENTLEMAN

In this article we discuss our experience designing and implementing a statistical computing language. In developing this new language, we sought to combine what we felt were useful features from two existing computer languages. We feel that the new language provides advantages in the areas of portability, computational efficiency, memory management, and scoping.

**Key Words:** Computer language; Statistical computing.

Figura 6. Artigo original do R.

# História do R

- ▶ Em **1997** foi formado o **R Core Team**: um grupo de aproximadamente 20 desenvolvedores que mantêm, gerenciam, controlam o código fonte e orientam a evolução da linguagem.
- ▶ Os membros do R Core Team fundaram a **R Foundation**: uma organização sem fins lucrativos que trabalha no interesse público para dar suporte ao R.
- ▶ Os direitos autorais do código-fonte primário do R pertencem à R Foundation e são publicados sob a GNU General Public License versão 2.0.

# História do R

- ▶ Em **2000** o R 1.0.0 foi lançado.
- ▶ O R base está disponível para instalação no **Comprehensive R Archive Network**, também conhecido como **CRAN**.

"The release of a current major version indicates that we believe that R has reached a level of stability and maturity that makes it suitable for production use. Also, the release of 1.0.0 marks that the base language and the API for extension writers will remain stable for the foreseeable future. In addition we have taken the opportunity to tie up as many loose ends as we could."

For the R Core Team,  
Peter D.

Here's the relevant bit of the NEWS file:

CHANGES IN R VERSION 1.0.0

Figura 7. R 1.0.0.

# O que é o R

- ▶ O R é uma **linguagem de programação**.
- ▶ Você faz a análise de dados escrevendo **funções e scripts**, não apontando, clicando e arrastando caixas.
- ▶ Para quem nunca programou, parece assustador. Mas o R é fácil de aprender e guiado a análise de dados.
- ▶ É possível instalar e usar o R nos principais sistemas operacionais.
- ▶ Assim como vários outros softwares livres e de código aberto, o R tem lançamentos frequentes de **versões**.
- ▶ A comunidade R é altamente ativa com usuários no mundo todo que contribuem, desenvolvem pacotes e ajudam uns aos outros por meio de materiais online como listas de discussão e tutoriais.

# R e os pacotes

- ▶ **Pacotes R** são coleções de funções R, dados e código compilado.
- ▶ O R já vem com um conjunto de pacotes por padrão e outros podem ser adicionados para estender os recursos.
- ▶ Os pacotes hoje disponíveis são o resultado de anos de colaboração de pessoas de todo o mundo.
- ▶ Uma das funções do CRAN é hospedar diversos pacotes complementares.



Figura 8. Extraído de pixabay.com.

# R-base

- ▶ O R “base” contém o pacote básico necessário para executar o R e as funções mais fundamentais.
- ▶ As funções já vem disponíveis, prontas para chamada e uso.
- ▶ É composto por **15** pacotes:

1. base
2. compiler
3. datasets
4. grDevices
5. graphics
6. grid
7. methods
8. parallel
9. splines
10. stats
11. stats4
12. tcltk
13. tools
14. translations
15. utils.

# R-recommended

- ▶ Um segundo grupo de pacotes que já vem com a instalação do R são os pacotes “recomendados”.
- ▶ Apesar de já instaladas, estas bibliotecas precisam ser chamadas para que seja possível usar as funções.
- ▶ É composto por outros **15** pacotes.

1. KernSmooth

2. MASS

3. Matrix

4. boot

5. class

6. cluster

7. codetools

8. foreign

9. lattice

10. mgcv

11. nlme

12. nnet

13. rpart

14. spatial

15. survival.

# Outros pacotes

- ▶ Você pode facilmente obter e instalar pacotes além dos 30 que já vem com a instalação tradicional do R.
- ▶ A principal fonte de pacotes é o próprio CRAN, que hoje conta com **mais de 19000 pacotes**.
- ▶ Fontes secundárias envolvem páginas web e repositórios como github, onde desenvolvedores mantêm pacotes em desenvolvimento.

Available CRAN Packages By Name	
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<a href="#">A3</a>	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
<a href="#">AalenJohansen</a>	Conditional Aalen-Johansen Estimation
<a href="#">AATools</a>	Reliability and Scoring Routines for the Approach-Avoidance Task
<a href="#">ABACUS</a>	Apps Based Activities for Communicating and Understanding Statistics
<a href="#">abbreviate</a>	Readable String Abbreviation
<a href="#">abbyyR</a>	Access to Abbyy Optical Character Recognition (OCR) API
<a href="#">abc</a>	Tools for Approximate Bayesian Computation (ABC)
<a href="#">abc.data</a>	Data Only: Tools for Approximate Bayesian Computation (ABC)
<a href="#">ABC.RAP</a>	Array Based CpG Region Analysis Pipeline
<a href="#">ABCanalysis</a>	Computed ABC Analysis
<a href="#">abclass</a>	Angle-Based Large-Margin Classifiers
<a href="#">ABCoptim</a>	Implementation of Artificial Bee Colony (ABC) Optimization
<a href="#">ABCp2</a>	Approximate Bayesian Computational Model for Estimating P2
<a href="#">abcrf</a>	Approximate Bayesian Computation via Random Forests
<a href="#">abcrlda</a>	Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis
<a href="#">abctools</a>	Tools for ABC Analyses
<a href="#">abd</a>	The Analysis of Biological Data
<a href="#">abdiv</a>	Alpha and Beta Diversity Measures
<a href="#">abe</a>	Augmented Backward Elimination
<a href="#">abess</a>	Fast Best Subset Selection
<a href="#">ablasso</a>	Adaptive Bayesian Graphical Lasso
<a href="#">ABHgenotypeR</a>	Easy Visualization of ABH Genotypes
<a href="#">abind</a>	Combine Multidimensional Arrays
<a href="#">abjData</a>	Databases Used Routinely by the Brazilian Jurimetrics Association

Figura 9. Lista de pacotes disponíveis por nome no CRAN.

# Limitações

- ▶ O R é essencialmente baseado em tecnologia antiga (sistema S da Bell Labs).
- ▶ Por ser baseado em tecnologia antiga o R sofre com suporte para algumas ferramentas modernas como gráficos dinâmicos.
- ▶ O R demanda memória física, os objetos geralmente são armazenados em memória. É algo que vem sendo melhorado aos poucos.
- ▶ O R depende dos seus usuários. Os novos recursos dependem do interesse da comunidade. A partir do momento em que o interesse acabar, o R ficará estagnado.

# Em resumo

O R fornece

- ▶ Diversos recursos de Estatística.
- ▶ Diversos recursos gráficos.
- ▶ Uma vasta coleção de pacotes oficiais e não oficiais.
- ▶ Uma linguagem de programação bem desenvolvida, simples e eficaz.
- ▶ Possibilidade de instalação e uso em uma ampla variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS.
- ▶ Possibilidade de uso de códigos C, C++ e Fortran para tarefas computacionalmente intensivas.
- ▶ Documentação padronizada.

# IDEs e Editores

- ▶ Existem softwares adicionais úteis para ajudar a programar de forma mais rápida e eficiente.
- ▶ As IDE's (*Integrated Development Environment*) são softwares que oferecem algumas facilidades para se programar em determinada linguagem.
- ▶ Já os editores tendem a ser úteis para múltiplas linguagens e fornecem mais alternativas de customização.
- ▶ Para trabalhar em R, dentre IDEs e editores, o RStudio IDE é a opção mais famosa.



Figura 10. Logo do RStudio.

# Links

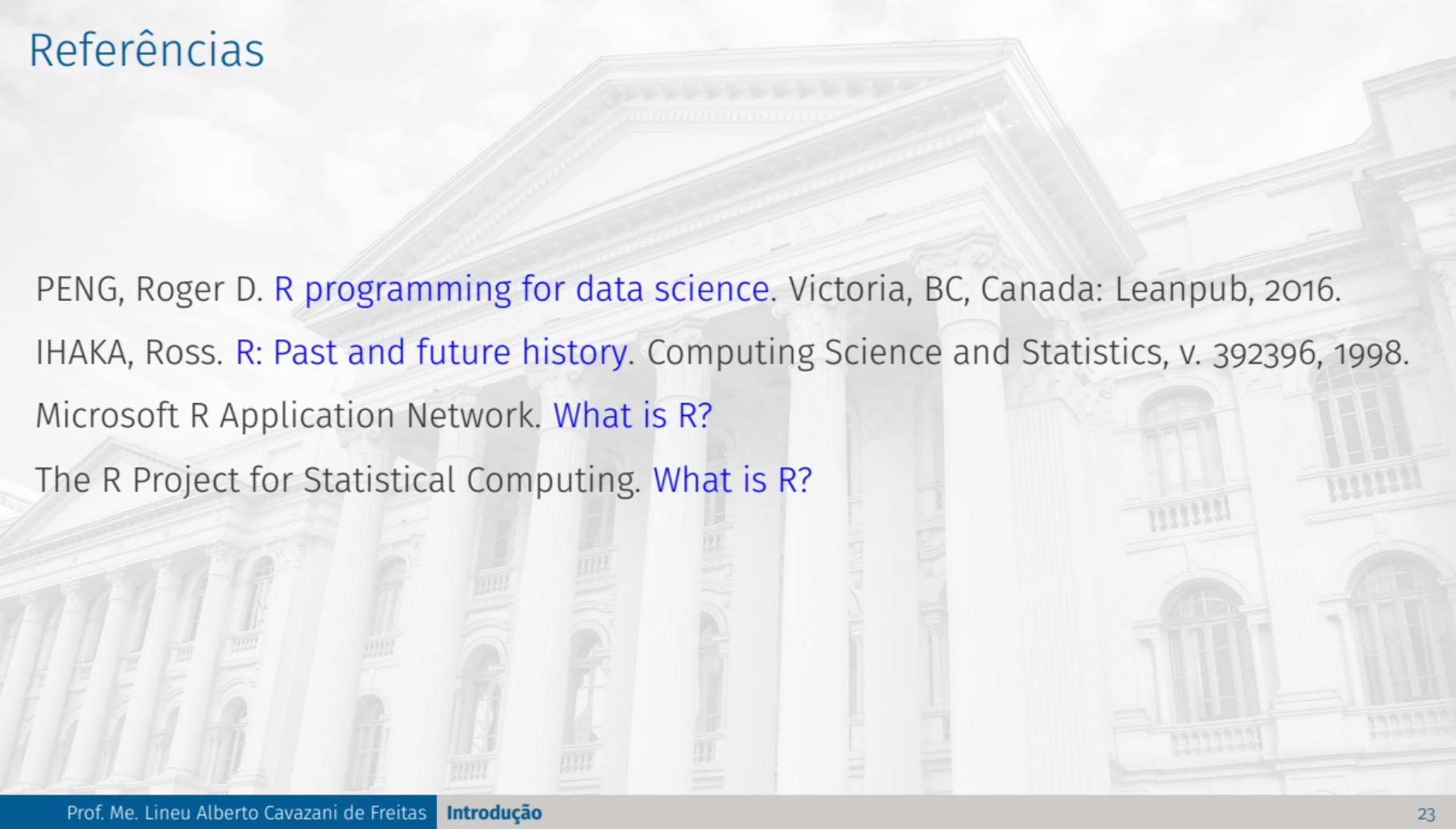


- ▶ [The Comprehensive R Archive Network.](#)
- ▶ [RStudio Desktop.](#)
- ▶ [Posit Cloud.](#)

# Materiais para aprender R

- ▶ Mayer, FP; Bonat, WH; Zeviani, WM; Krainski, EK; Ribeiro Jr., PJ. [Estatística Computacional com R](#). DEST/UFPR, 2018.
- ▶ Ribeiro Jr., PJ. [Introdução ao Ambiente Estatístico R](#). 2011.
- ▶ Horton, NJ; Pruiim, R; Kaplan, DT. [A Student's Guide to R](#). 2015.
- ▶ Maindonald, JH. [Using R for Data Analysis and Graphics](#). 2008.
- ▶ Paradis, E. [R for Beginners](#). 2005.

# Referências



PENG, Roger D. [R programming for data science](#). Victoria, BC, Canada: Leanpub, 2016.

IHAKA, Ross. [R: Past and future history](#). Computing Science and Statistics, v. 392396, 1998.

Microsoft R Application Network. [What is R?](#)

The R Project for Statistical Computing. [What is R?](#)