

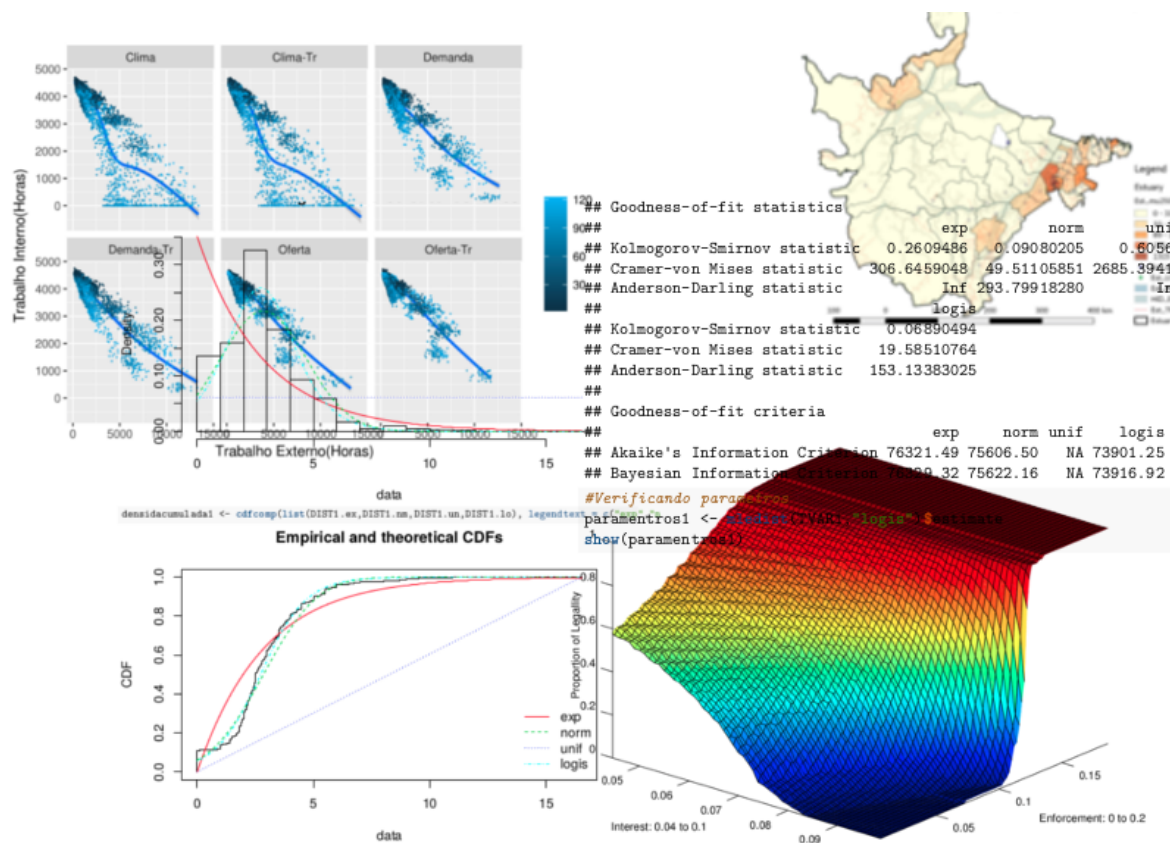
Construção de Modelos e Indicadores Econômicos

Introdução ao Tratamento e Análise de Dados em R

ou Data Science para todos!

S. Rivero, H. Farias and Equipe UFPA

27 de fevereiro de 2019



Introdução ao Tratamento e Análise de Dados em R

ou Data Science para todos!

S. Rivero and H. Farias

Programa de Pós-Graduação em Economia
Instituto de Ciências Sociais Aplicadas
Universidade Federal do Pará
Rua Augusto Correia, 1
Belém, Pará - 66.075-200

Equipe UFPA

Faculdade de Economia
Instituto de Ciências Sociais Aplicadas
Universidade Federal do Pará
Rua Augusto Correia, 1
Belém, Pará - 66.075-200

Produto: Cursos

Propriedade BANPARÁ e PPGE-UFPA
(Distribuição Restrita)

Prepared for Banpará

Under FADESP-NO-CONTRATO

Monitored by Banpará

Abstract: Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Disclaimer:

This document is an output from the Banpará Project

PPGE-UFPA report

Distribution Restrictions

©2019, All rights reserved

Sumário

Introdução	1
Introdução	1
1 Aula 1 - Instalando e Configurando o R e RStudio	3
1.1 Instalando o R e RStudio	3
1.1.1 Instalando o R	3
1.1.2 Instalando o R Studio	6
1.2 Checando a Instalação Existente e os Requisitos	6
1.3 Pacotes no R	6
1.3.1 O conceito de pacote e para que serve	6
1.3.2 Como sei que pacotes eu preciso?	6
1.3.3 Baixando os pacotes	7
1.3.4 Resolvendo problemas de compilação	7
1.3.5 Utilizando os pacotes no seu programa R	7
2 Aula 2 - Acessando e Utilizando Bases de Dados	8
2.1 O ciclo de tratamento e análise de dados	8
2.2 Tipos de Dados em R	8
2.3 Dataframes	8
2.4 Acessando Arquivos no computador	9
2.5 Acessando Bases de dados via APIs	9
2.6 Trabalhando com bases de dados muito grandes	9
3 Aula 3 - Limpando e organizando seus dados	11
3.1 O que é uma boa base de dados e que tipos de bases existem?	11
3.2 dplyr	11
3.3 tidyr	11
3.4 tidyverse	11
4 Aula 4 - Apresentando Resultados	12
4.1 Rmarkdown - Preparando o relatório enquanto você analisa os dados	12
4.2 Gerando dados sintetizados - Utilizando dataframes	12
4.3 Gerando Tabelas	12
4.3.1 xtable	12
4.3.2 stargazer	12
5 Aula 5 - Gerando estatísticas dos Dados	13
5.1 Estatísticas Descritivas	13
5.2 Correlação	13
5.3 Apresentando os Resultados em Gráficos	13
5.3.1 Gráficos Simples	13
5.3.2 Bibliotecas de Gráficos	13

6	Aula 6 - Apresentando Resultados	14
6.1	ggplot	14
6.2	shiny.....	14
7	Aula 7 - Utilizando modelos	15
7.1	Regressão.....	15
7.1.1	Executando a Regressão	15
7.1.2	Apresentando os resultados da Regressão	15
7.2	Gerando Dashboards	15
8	Aula 8 - Encerramento do Curso	16
9	Onde aprender mais?	17
	Onde aprender mais?	17
	Referências Bibliográficas	18

Introdução

O R é uma suíte integrada de software que permite a recuperação, o tratamento, e a análise de dados[VS11]. Pode se dizer que o R é um ambiente de tratamento de dados que permite ao usuário, além a análise de dados propriamente dita, escrever extensões e ampliar o seu escopo.

R é uma ferramenta de software livre que atende aos critérios da *Free Software Foundation* e tem uma licença **GNU**¹

Algumas das funcionalidade do R são[VS11]:

- Ferramentas para manuseio e armazenamento de dados
- Um conjunto de operadores que permitem o cálculo numérico e a manipulação de matrizes
- Um enorme conjunto de bibliotecas para análise de dados
- Ferramentas para apresentação gráfica de dados e resultados
- Uma linguagem de programação orientada a objetos e extensível
- A possibilidade de estender a linguagem, suas bibliotecas e funções

O R tem um conjunto extenso de ferramentas que permitem desde uma análise simples de regressão ou de aglomerados (*cluster*) até apresentações de resultados extraídos de bancos de dados em larga escala com ferramentas de busca *SQL*² apresentas em páginas web¹.

Neste texto, iremos apresentar um conjunto de ferramentas de uso livre, a sua maioria de código aberto, que permitirão a extração, tratamento, análise e apresentação de dados, tanto para análises estatísticas mais diretas quanto para a tomada de decisões estratégicas de negócios.

¹<https://www.gnu.org/>

²Structured Query Language - <https://pt.wikipedia.org/wiki/SQL>

Was U.S. Monetary Policy Too Tight or Too Loose?

Created by [Andrew Gillen](#)

Last Updated: May 15, 2016

This tool is designed to help determine how effective monetary policy has been. In the left panel, you can choose the Federal Reserve's target(s) and adjust options for each. The main panel will show whether monetary policy was too tight or too loose based on your choices.

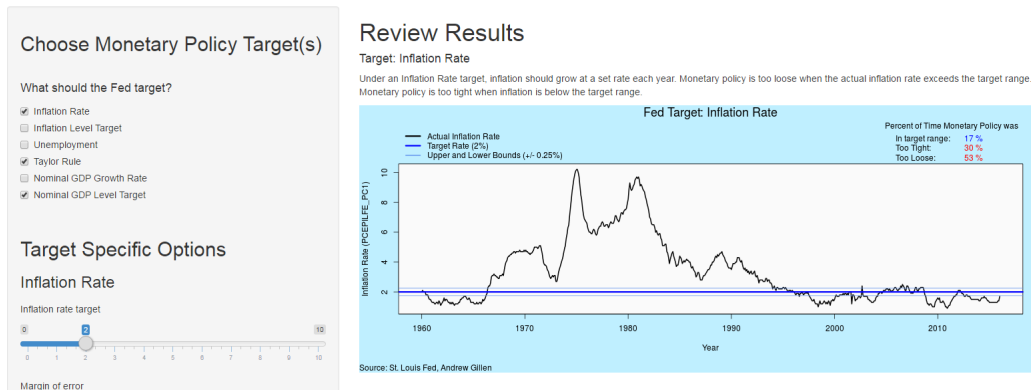


Figura 1. Exemplo de Dashboard de Política monetária usando R [Sho19]

1 Aula 1 - Instalando e Configurando o R e RStudio

Nesta aula os alunos aprenderão a baixar o R e RStudio bem como aprenderão a utilizar bibliotecas em R

1.1 Instalando o R e RStudio

1.1.1 Instalando o R

Para instalar o R é preciso acessar o site <https://www.r-project.org/>. Os passos estão enumerados a seguir:

Nesta página clicar em **download R** que acessa a página - <https://cran.r-project.org/mirrors.html> - (Figura 2)

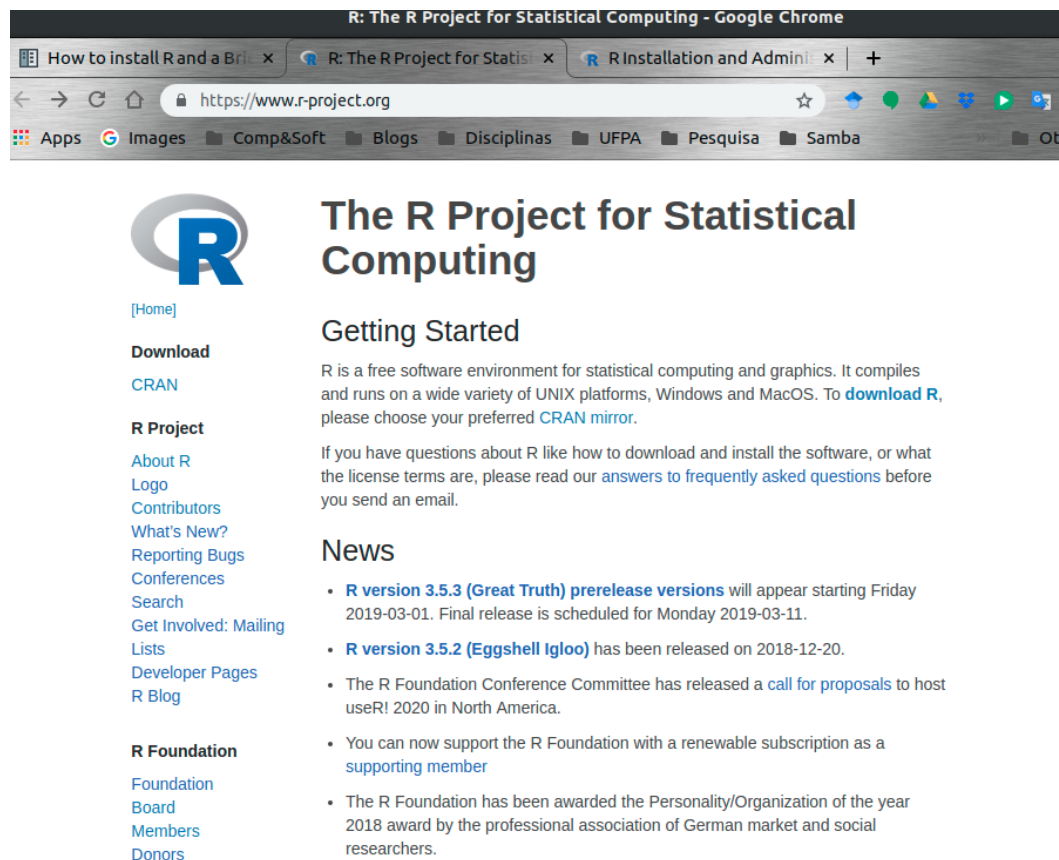


Figura 2. Nesta página clicar em download R

Nesta página você seleciona o espelho que utilizará para baixar o R. Há sites no Brasil ou sítios em nuvem o item 0 - Cloud redirecionará automatica-

mente para sítios apoiados pelo Rstudio - <https://cloud.r-project.org/> (Figura 3) - A diferença entre os prefixos *http* e *https* é que os sites com "s" utilizam encriptação. Muitas vezes é necessário utilizar o *https* em instalações cujos ambientes de TI exigem.

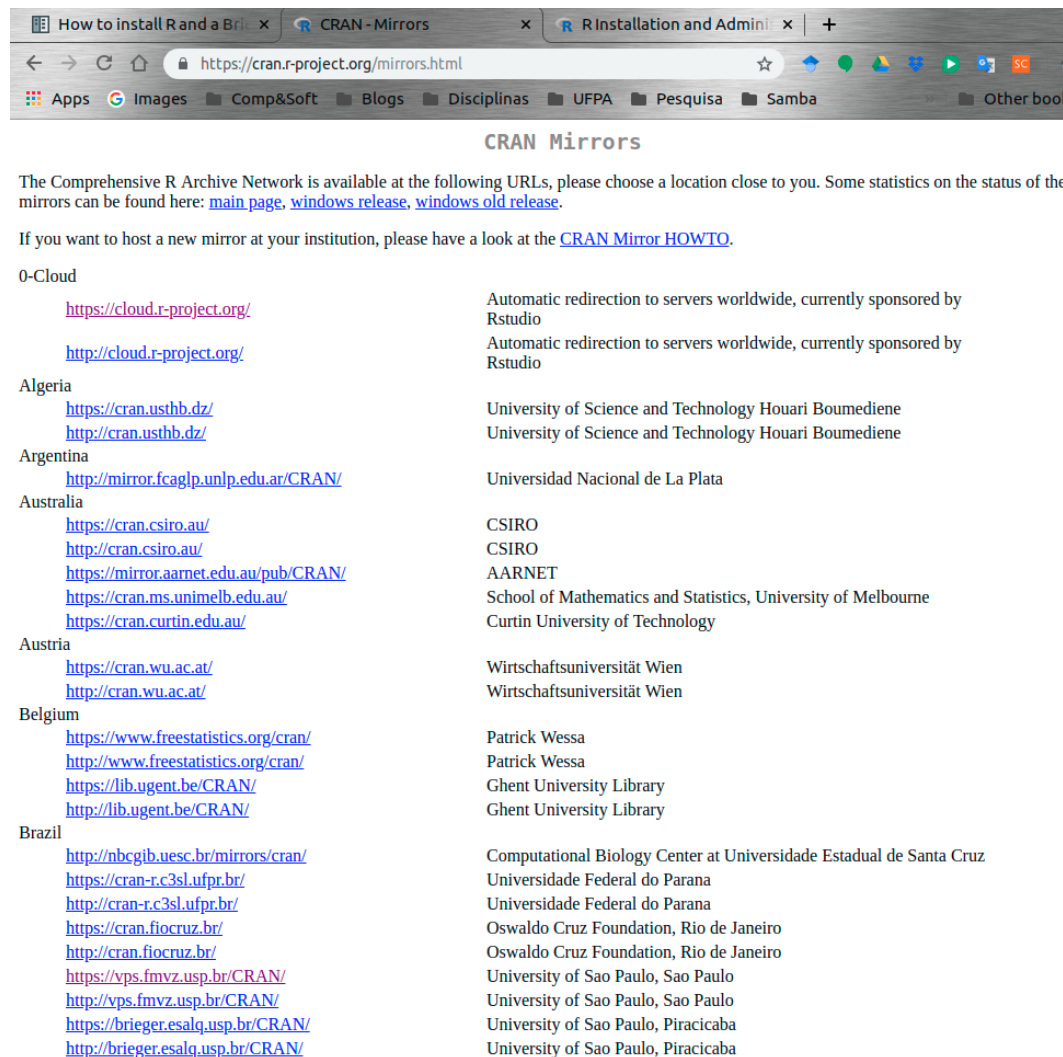


Figura 3. Aqui você seleciona o espelho que vai usar para baixar o R

Você poderá baixar o R (por exemplo em <https://cloud.r-project.org/>) ou em algum dos outros espelhos citados anteriormente, de acordo com seu sistema operacional (Figura 4) Cada sistema (*Linux*, *MacOSX* ou *Windows*) tem uma rotina diferente para instalação. Aqui vamos explicar o sistema operacional mais comum (*Windows*). Para outros sistemas, recomenda-se buscar instruções específicas em fóruns adequados.

Finalmente, para instalar o R, você pode executar o arquivo que está no link apresentado na Figura 5. Você baixará um arquivo **.exe** e, ao executá-lo, o instalador do *Windows* tornará o programa disponível para uso na sua máquina.

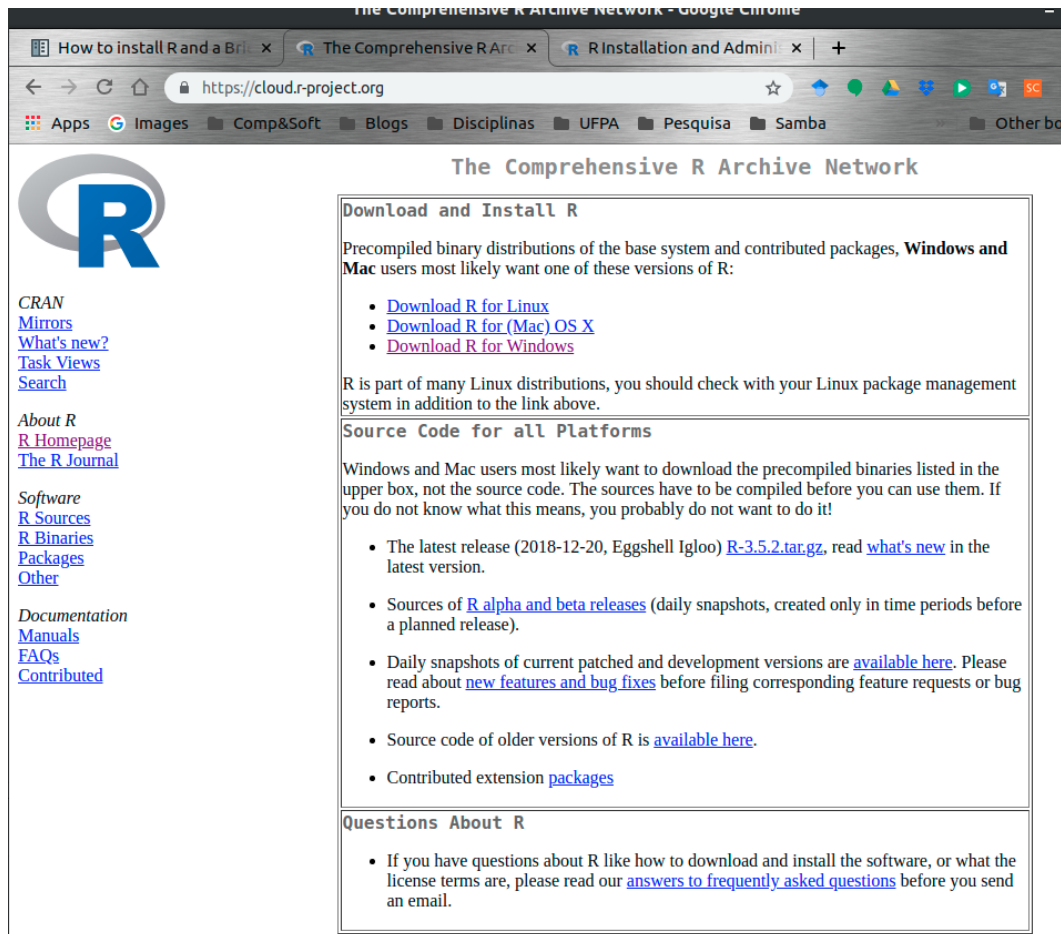


Figura 4. Sítio para baixar o R de acordo com seu sistema operacional

R-3.5.2 for Windows (32/64 bit)

[Download R 3.5.2 for Windows](#) (79 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Figura 5. Página do Download do R

Você pode conseguir mais informações sobre o R e os detalhes da instalação checando nos *FAQs*³ dos sítios que você estiver acessando <https://cloud.r-project.org/bin/windows/base/rw-FAQ.html>, ou em <https://cran.r-project.org/doc/manuals/R-admin.html>

1.1.2 Instalando o R Studio

<https://www.rstudio.com/products/rstudio/>

<http://web.cs.ucla.edu/~gulzar/rstudio/>

<https://www.rstudio.com/products/rstudio/download/>

<http://rprogramming.net/download-and-install-rstudio/>

1.2 Checando a Instalação Existente e os Requisitos

<https://stackoverflow.com/questions/11103189/how-to-find-out-which-package-version>

<https://www.r-bloggers.com/list-of-user-installed-r-packages-and-their-versions/>

<https://community.rstudio.com/t/reinstalling-packages-on-new-version-of-r/7670>

1.3 Pacotes no R

1.3.1 O conceito de pacote e para que serve

<https://www.datacamp.com/community/tutorials/r-packages-guide>

<https://www.rstudio.com/products/rpackages/>

<http://r-pkgs.had.co.nz/>

1.3.2 Como sei que pacotes eu preciso?

<https://blog.revolutionanalytics.com/2017/01/cran-10000.html>

https://cran.r-project.org/web/packages/available_packages_by_name.html

³Frequently Asked Questions

<https://cran.r-project.org/web/packages/>

1.3.3 Baixando os pacotes

<https://www.r-bloggers.com/installing-r-packages/>

<https://www.r-bloggers.com/how-to-install-and-include-an-r-package/>

http://kbroman.org/pkg_primer/pages/build.html

1.3.4 Resolvendo problemas de compilação

<https://stackoverflow.com/questions/23135703/package-install-error-compilation-fai>

<https://support.rstudio.com/hc/en-us/community/posts/200522573-Can-t-install-packa>

<http://mazamascience.com/WorkingWithData/?p=1185>

1.3.5 Utilizando os pacotes no seu programa R

<https://www.statmethods.net/interface/packages.html>

<https://www.dummies.com/programming/r/how-to-install-load-and-unload-packages-in-r>

2 Aula 2 - Acessando e Utilizando Bases de Dados

Apresentar o conceito de Dataframe, os tipos de dados utilizados no R e os principais comandos

2.1 O ciclo de tratamento e análise de dados

2.2 Tipos de Dados em R

<https://www.statmethods.net/input/datatypes.html>

<https://swcarpentry.github.io/r-novice-inflammation/13-supp-data-structures/>

https://www.tutorialspoint.com/r/r_data_types.htm

<http://www.r-tutor.com/r-introduction/basic-data-types>

<https://www.cyclismo.org/tutorial/R/types.html>

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/typeof.html>

2.3 Dataframes

https://www.tutorialspoint.com/r/r_data_frames.htm

<https://www.datamentor.io/r-programming/data-frame/>

<http://www.r-tutor.com/r-introduction/data-frame>

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/data.frame.html>

<https://www.tutorialgateway.org/data-frame-in-r/>

<https://datacarpentry.org/R-ecology-lesson/02-starting-with-data.html>

<https://www.statmethods.net/input/importingdata.html>

2.4 Acessando Arquivos no computador

https://www.datacamp.com/community/tutorials/r-data-import-tutorial?utm_source=adwords_ppc&utm_campaignid=1455363063&utm_adgroupid=65083631748&utm_device=c&utm_keyword=&utm_matchtype=b&utm_network=g&utm_adposition=1t1&utm_creative=332602034364&utm_targetid=dsa-473406573035&utm_loc_interest_ms=&utm_loc_physical_ms=1001610&gclid=Cj0KCQiA5NPjBRDDARIsAM9X1GLgVEiwAqyW9CPisvAqFv2mNXzwarSlIaAgdZEALw_wcB

<http://rprogramming.net/read-csv-in-r/>

<https://www.rdocumentation.org/packages/gdata/versions/2.18.0/topics/read.xls>

<https://stat.ethz.ch/R-manual/R-devel/library/utils/html/read.fwf.html>

<https://riptutorial.com/r/example/31447/importing-fixed-width-files>

2.5 Acessando Bases de dados via APIs

<https://www.r-bloggers.com/accessing-apis-from-r-and-a-little-r-programming/>

<https://cran.r-project.org/web/packages/htrr/vignettes/api-packages.html>

<https://zapier.com/learn/apis/>

<https://www.earthdatascience.org/courses/earth-analytics/get-data-using-apis/API-data-access-r/>

2.6 Trabalhando com bases de dados muito grandes

<http://dept.stat.lsa.umich.edu/~jerrick/courses/stat701/notes/sql.html>

<https://datacarpentry.org/R-ecology-lesson/05-r-and-databases.html>

<https://db.rstudio.com/>

http://www.columbia.edu/~sjm2186/EPIC_R/EPIC_R_BigData.pdf

<https://www.rstudio.com/resources/webinars/working-with-big-data-in-r/>

https://rpubs.com/msundar/large_data_analysis

3 Aula 3 - Limpando e organizando seus dados

3.1 O que é uma boa base de dados e que tipos de bases existem?

3.2 dplyr

3.3 tidyr

3.4 tidyverse

4 Aula 4 - Apresentando Resultados

4.1 Rmarkdown - Preparando o relatório enquanto você analisa os dados

4.2 Gerando dados sintetizados - Utilizando dataframes

4.3 Gerando Tabelas

4.3.1 xtable

4.3.2 stargazer

5 Aula 5 - Gerando estatísticas dos Dados

5.1 Estatísticas Descritivas

5.2 Correlação

5.3 Apresentando os Resultados em Gráficos

5.3.1 Gráficos Simples

5.3.2 Bibliotecas de Gráficos

6 Aula 6 - Apresentando Resultados

6.1 ggplot

6.2 shiny

7 Aula 7 - Utilizando modelos

7.1 Regressão

7.1.1 Executando a Regressão

7.1.2 Apresentando os resultados da Regressão

7.2 Gerando Dashboards

8 Aula 8 - Encerramento do Curso

9 Onde aprender mais?

Referências Bibliográficas

[Sho19] Show me shiny: Gallery of r web apps, 2019.

[VS11] W N Venables and D M Smith. *An Introduction to R*, volume 2. CRAN - Comprehensive R Archive Network, 2011.