

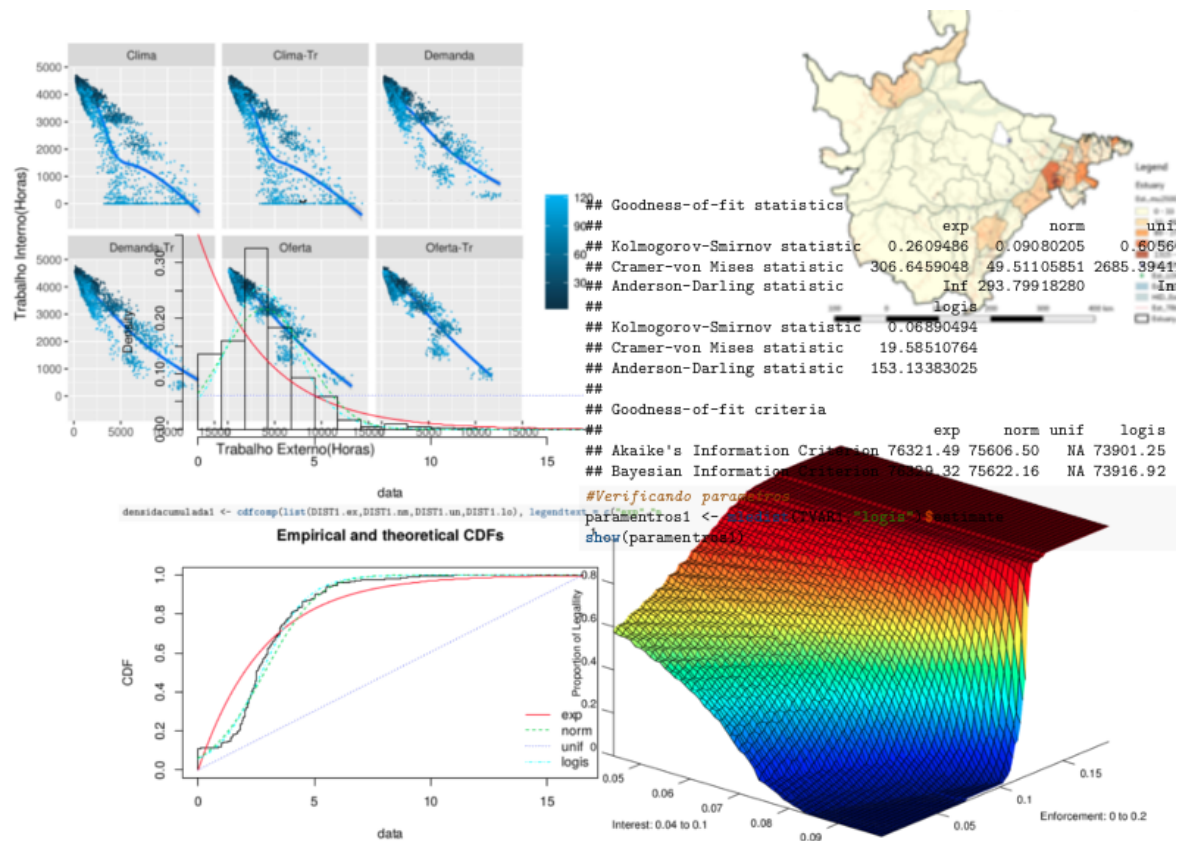
Construção de Modelos e Indicadores Econômicos

# Introdução ao Tratamento e Análise de Dados em R

ou Data Science para todos!

S. Rivero, H. Farias and Equipe UFPA

26 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

1	Introdução .....	1
2	Aula 1 - Instalando e Configurando o R e RStudio.....	2
2.1	Checando a Instalação Existente e os Requisitos.....	2
2.2	Instalando o R e RStudio .....	2
2.2.1	Instalando o R .....	2
2.2.2	Instalando o R Studio .....	2
2.3	Pacotes no R.....	2
2.3.1	O conceito de pacote e para que serve .....	3
2.3.2	Como sei que pacotes eu preciso?.....	3
2.3.3	Baixando os pacotes .....	3
2.3.4	Resolvendo problemas de compilação .....	3
2.3.5	Utilizando os pacotes no seu programa R .....	3
3	Aula 2 - Acessando e Utilizando Bases de Dados .....	4
3.1	O ciclo de tratamento e análise de dados.....	4
3.2	Tipos de Dados em R .....	4
3.3	Dataframes .....	4
3.4	Acessando Arquivos no computador .....	5
3.5	Acessando Bases de dados via APIs .....	5
3.6	Trabalhando com bases de dados muito grandes .....	5
4	Aula 3 - Limpando e organizando seus dados.....	7
4.1	O que é uma boa base de dados e que tipos de bases existem?.....	7
4.2	dplyr .....	7
4.3	tidyr.....	7
4.4	tidyverse .....	7
5	Aula 4 - Apresentando Resultados.....	8
5.1	Rmarkdown - Preparando o relatório enquanto você analisa os dados .....	8
5.2	Gerando dados sintetizados - Utilizando dataframes .....	8
5.3	Gerando Tabelas.....	8
5.3.1	xtable.....	8
5.3.2	stargazer .....	8
6	Aula 5 - Gerando estatísticas dos Dados .....	9
6.1	Estatísticas Descritivas.....	9
6.2	Correlação .....	9
6.3	Apresentando os Resultados em Gráficos.....	9
6.3.1	Gráficos Simples .....	9
6.3.2	Bibliotecas de Gráficos .....	9
7	Aula 6 - Apresentando Resultados.....	10

---

7.1	ggplot .....	10
7.2	shiny .....	10
8	Aula 7 - Utilizando modelos .....	11
8.1	Regressão .....	11
8.1.1	Executando a Regressão .....	11
8.1.2	Apresentando os resultados da Regressão .....	11
8.2	Gerando Dashboards .....	11
9	Aula 8 - Encerramento do Curso .....	12
10	Onde aprender mais? .....	13
	Referências Bibliográficas .....	14

# 1 Introdução

Para que serve o R?

Como se pode utilizar o conjunto de pacotes R para Análise de Dados e BI?

Uma configuração adequada para usar ferramentas de BI

## 2 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

### 2.1 Checando a Instalação Existente e os Requisitos

1

<https://stackoverflow.com/questions/11103189/how-to-find-out-which-packages-are-installed-on-my-system>

<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>

### 2.2 Instalando o R e RStudio

#### 2.2.1 Instalando o R

<https://a-little-book-of-r-for-bioinformatics.readthedocs.io/en/latest/src/installr.html>

<https://cran.r-project.org/doc/manuals/R-admin.html>

#### 2.2.2 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/>



## 2.3 Pacotes no R

### 2.3.1 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/>

### 2.3.2 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](https://cran.r-project.org/web/packages/available_packages_by_name.html)

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

### 2.3.3 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](http://kbroman.org/pkg_primer/pages/build.html)

### 2.3.4 Resolvendo problemas de compilação

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

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

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

### 2.3.5 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-p>

## 3 Aula 2 - Acessando e Utilizando Bases de Dados

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

### 3.1 O ciclo de tratamento e análise de dados

### 3.2 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](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>

### 3.3 Dataframes

[https://www.tutorialspoint.com/r/r\\_data\\_frames.htm](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>

### 3.4 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\\_adpostion=1t1&utm\\_creative=332602034364&utm\\_targetid=dsa-473406573035&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=1001610&gclid=Cj0KCQiA5NPjBRDDARIsAM9X1GLkgYWeIgVEiwAqyW9CPisvAqFv2mNXzwarSlIaAgdZEALw\\_wcB](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_adpostion=1t1&utm_creative=332602034364&utm_targetid=dsa-473406573035&utm_loc_interest_ms=&utm_loc_physical_ms=1001610&gclid=Cj0KCQiA5NPjBRDDARIsAM9X1GLkgYWeIgVEiwAqyW9CPisvAqFv2mNXzwarSlIaAgdZEALw_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>

### 3.5 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/httr/vignettes/api-packages.html>

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

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

### 3.6 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](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](https://rpubs.com/msundar/large_data_analysis)

## 4 Aula 3 - Limpando e organizando seus dados

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

4.2 dplyr

4.3 tidyr

4.4 tidyverse

## 5 Aula 4 - Apresentando Resultados

5.1 Rmarkdown - Preparando o relatório enquanto você analisa os dados

5.2 Gerando dados sintetizados - Utilizando dataframes

5.3 Gerando Tabelas

5.3.1 xtable

5.3.2 stargazer

## 6 Aula 5 - Gerando estatísticas dos Dados

### 6.1 Estatísticas Descritivas

### 6.2 Correlação

### 6.3 Apresentando os Resultados em Gráficos

#### 6.3.1 Gráficos Simples

#### 6.3.2 Bibliotecas de Gráficos

## 7 Aula 6 - Apresentando Resultados

### 7.1 ggplot

### 7.2 shiny



## 8 Aula 7 - Utilizando modelos

### 8.1 Regressão

#### 8.1.1 Executando a Regressão

#### 8.1.2 Apresentando os resultados da Regressão

### 8.2 Gerando Dashboards

## 9 Aula 8 - Encerramento do Curso

## 10 Onde aprender mais?

## Referências Bibliográficas