

R para epidemiologia

Rodrigo Citton P. dos Reis
rodrigocpdosreis@gmail.com

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
DEST, PPGEPI, ELSA-BRASIL

Porto Alegre, 2018

Bioestatística e epidemiologia no paradigma da saúde pública

Métodos quantitativos

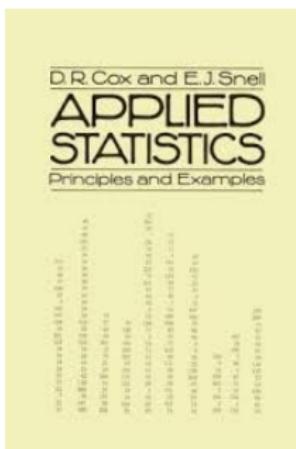
- ▶ A **epidemiologia** e a **bioestatística** são ciências básicas para a saúde pública.
 - ▶ Investigações de saúde pública, geralmente, usam **métodos quantitativos** que combinam as duas disciplinas.
- ▶ A epidemiologia está preocupada com a compreensão do desenvolvimento de doenças e os métodos utilizados para investigar a etiologia, a progressão e tratamento de doenças.
- ▶ Para investigar uma questão de pesquisa **dados** são necessários.
 - ▶ Algumas vezes os dados já estão disponíveis, em outras novos dados precisam ser coletados.

Métodos quantitativos

- ▶ Dados apresentam varibilidade.
- ▶ A **estatística** é a ciência que lida com a variação dos dados para obter resultados e conclusões confiáveis.
- ▶ A **bioestatística** é a aplicação da estatística a problemas nas ciências biológicas, médicas e da saúde.
 - ▶ Os métodos e ferramentas de bioestatística são usados para analisar os dados, resumir a informação e gerar conhecimento para auxiliar a tomada de decisão.

Analisando dados

Fases de análise



1. Manipulação inicial dos dados.

- ▶ Limpeza dos dados.
- ▶ Criação, transformação e recodificação de variáveis.

2. Análise preliminar.

- ▶ Conhecimento dos dados, identificação de outliers, investigação preliminar.

3. Análise definitiva.

- ▶ Disponibiliza a base para as conclusões.

4. Apresentação das conclusões de forma precisa, concisa e lúcida.

Analisando dados



Objetivos e organização do curso

Objetivos

- ▶ Neste curso apresentaremos os conceitos básicos do R.
- ▶ A partir da formalização destes fundamentos iremos explorar algumas ferramentas em R úteis para a **análise de dados epidemiológicos**.

Organização

- ▶ **Sala de aula:** Laboratório 502 (Inst. Psicologia)
- ▶ **Avaliação (?)**: listas de exercícios (**para casa**), atividades em sala de aula e avaliação final.

Conteúdo

- ▶ Conceitos básicos de R
- ▶ Leitura de bancos de dados
- ▶ Estatística básica (testes de hipóteses e intervalos de confiança)
- ▶ Comparação de grupos
- ▶ Gráficos
- ▶ Modelos de regressão:
 - ▶ Linear
 - ▶ Logístico
 - ▶ Poisson
- ▶ Apresentação dos resultados do ajuste de modelos de regressão
- ▶ Rmarkdown para elaboração de relatórios dinâmicos
- ▶ Tópicos especiais

Leituras



- ▶ **Hills, M., Plummer, M.** *A short introduction to R for Epidemiology*, Bendix Cartensen, 2014.
- ▶ **Culhane, A.** *Introduction to Programming and Statistical Modelling in R*, 2013.

O que é o R?

O que é o R?

- ▶ O R é uma linguagem de programação desenvolvida para:
 - ▶ Manipulação de dados;
 - ▶ Análise estatística;
 - ▶ Visualização de dados.
- ▶ O que diferencia o R de outras ferramentas de análise de dados?
 - ▶ Desenvolvido por estatísticos;
 - ▶ É um software livre;
 - ▶ É extensível através de pacotes.



Breve histórico

- ▶ **R** é a versão livre, de código aberto, e gratuita do **S**.
 - ▶ Nos anos 1980 o **S** foi desenvolvido nos **Laboratórios Bell**, por **John Chambers**, para análise de dados e geração de gráficos.



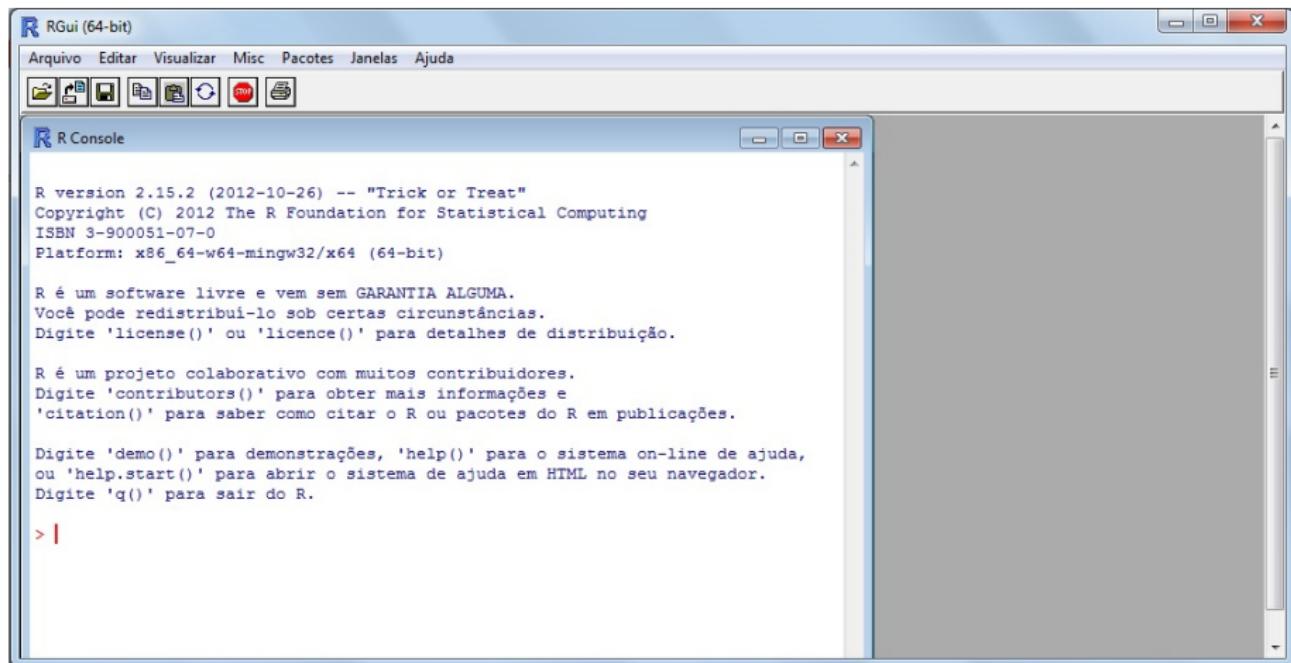
Breve histórico

- ▶ O **R** foi inicialmente escrito no começo dos anos 1990.
 - ▶ **Robert Gentleman** e **Ross Ihaka** no Dep. de Estatística da Universidade de Auckland.
 - ▶ O nome **R** se dá em parte por reconhecer a influência do **S** e por ser a inicial dos nomes **Robert** e **Ross**.



- ▶ Desde 1997 possui um grupo de 20 desenvolvedores.
 - ▶ A cada 6 meses uma nova versão é disponibilizada contendo atualizações.

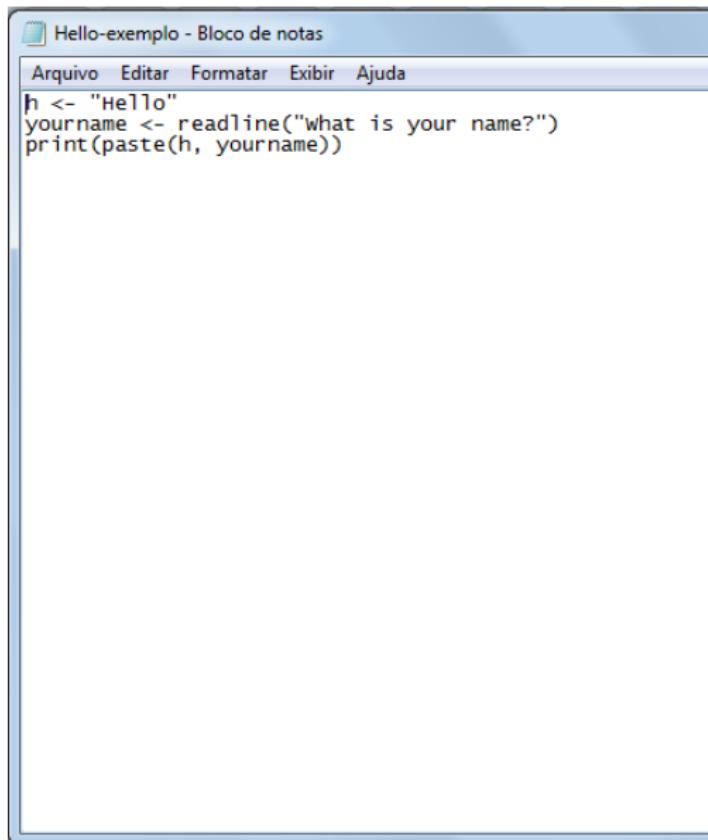
Interface do R



Como trabalhar com o R?

- ▶ Por ser uma linguagem de programação, o **R** realiza suas tarefas através de **funções e operadores**.
 - ▶ A criação de **scripts** (rotinas) é **a melhor prática para se trabalhar com o 'R'**.
 - ▶ **OBSERVAÇÃO:** sempre salve seus scripts (em um pen drive, dropbox ou e-mail); você pode querer utilizá-los novamente no futuro.
 - ▶ Utilização de editores de texto: **bloco de notas, Notepad ++, Tinn-R**, etc.
 - ▶ Interfaces de R para usuários: **RStudio**.

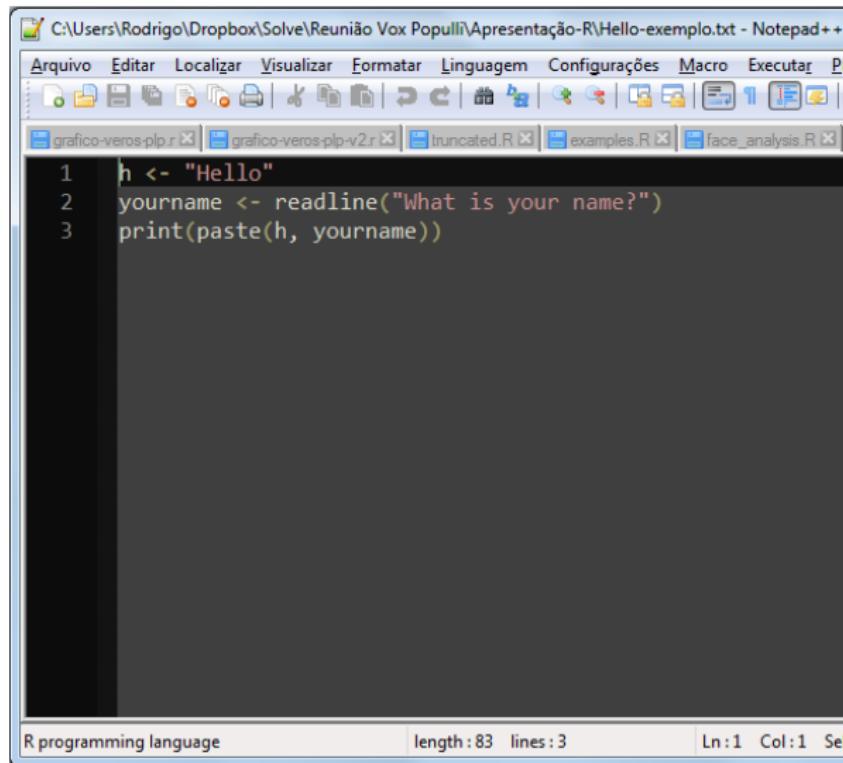
Editores de texto



The screenshot shows a Windows-style notepad window titled "Hello-exemplo - Bloco de notas". The menu bar includes "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The main text area contains the following R code:

```
h <- "Hello"  
yourname <- readline("what is your name?")  
print(paste(h, yourname))
```

Editores de texto



A screenshot of the Notepad++ text editor. The title bar reads "C:\Users\Rodrigo\Dropbox\Solve\Reunião Vox Populi\Apresentação-R\Hello-exemplo.txt - Notepad++". The menu bar includes Arquivo, Editar, Localizar, Visualizar, Formatar, Linguagem, Configurações, Macro, Executar, and Ajuda. The toolbar contains icons for file operations like Open, Save, Print, and Find. The status bar at the bottom shows "R programming language", "length : 83 lines : 3", "Ln : 1 Col : 1 Sel", and a small icon.

```
1 h <- "Hello"
2 yourname <- readline("What is your name?")
3 print(paste(h, yourname))
```

Interface do RStudio

RStudio

File Edit Code View Plots Session Project Build Tools Help

Project: (None)

NPP1984-grid-approx-samp.R AirCond1963-Jeffreys-grid-approx-sampling-eta.R Untitled1*

Source on Save Go to file/function

```
1 rm(list=ls())
2 
3 library(ggplot2)
4 library(MASS) # For the data set
5 
6 b <- biopsy
7 b$classn[b$class=="benign"] <- 0
8 b$classn[b$class=="malignant"] <- 1
9 ggplot(b, aes(x~V1, y~classn)) +
10   geom_point(position=position_jitter(width=0.3, height=0.06), alpha=0.4,
11   shape=21, size=1.5) +
12   stat_smooth(method=glm, family=binomial)
13 |
```

13:1 (Top Level) R Script

Console

```
> rm(list=ls())
> 
> library(ggplot2)
> library(MASS) # For the data set
> 
> b <- biopsy
> b$classn[b$class=="benign"] <- 0
> b$classn[b$class=="malignant"] <- 1
> ggplot(b, aes(x~V1, y~classn)) +
+   geom_point(position=position_jitter(width=0.3, height=0.06), alpha=0.4,
+   shape=21, size=1.5) +
+   stat_smooth(method=glm, family=binomial)
> |
```

Workspace History

Data b 699 obs. of 12 variables

Files Plots Packages Help

Zoom Export Clear All

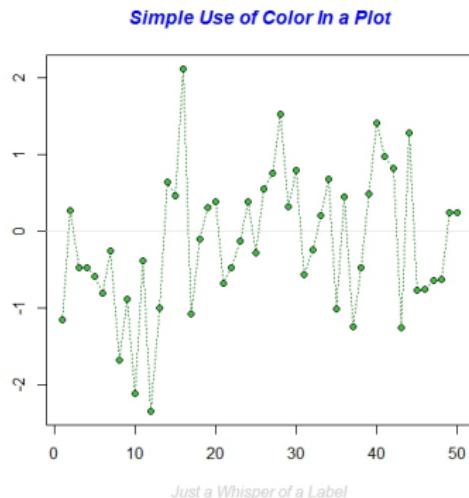
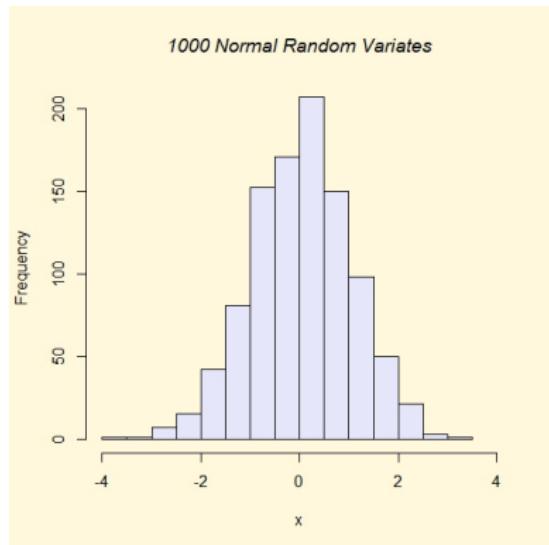
The figure displays a scatter plot of the variable V1 against classn. The x-axis (V1) has major ticks at 2.5, 5.0, 7.5, and 10.0. The y-axis (classn) has major ticks at 0.0, 0.3, 0.6, and 0.9. The data points are represented by small black dots with a slight horizontal jitter. A thick blue line represents the logistic regression fit, starting near (1.5, 0.05) and increasing monotonically to approach 1.0 as V1 increases. A light gray shaded area surrounds the blue line, indicating the confidence interval of the fit.

Você pode usar o R para

- ▶ Importação e exportação de dados
- ▶ Manipulação de dados
 - ▶ Transformação e recodificação de variáveis
 - ▶ Aplicação de filtros
- ▶ Visualização de dados
 - ▶ Diversos gráficos
 - ▶ Mapas
 - ▶ Gráficos e mapas interativos
- ▶ Análise de dados
 - ▶ Análise descritiva
 - ▶ Ajuste de modelos
 - ▶ Técnicas multivariadas
 - ▶ Análise de amostras complexas
- ▶ Geração de relatórios
 - ▶ Relatórios nos formatos: pdf, HTML, Word, Power Point

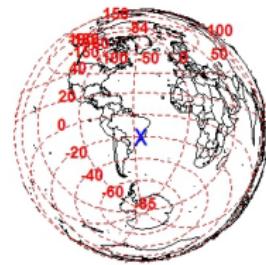
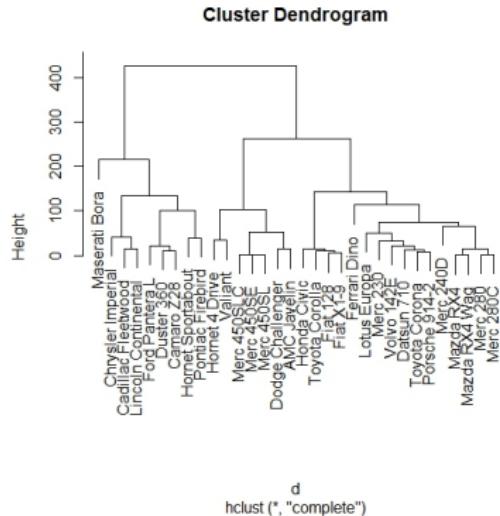
Resumindo: você pode usar o R em todas as etapas de uma análise de dados!

Gráficos do R

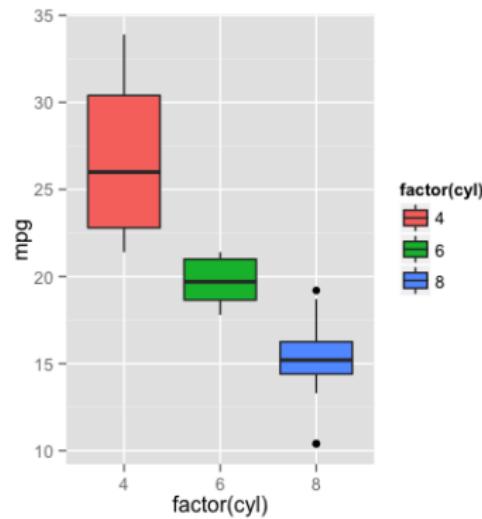
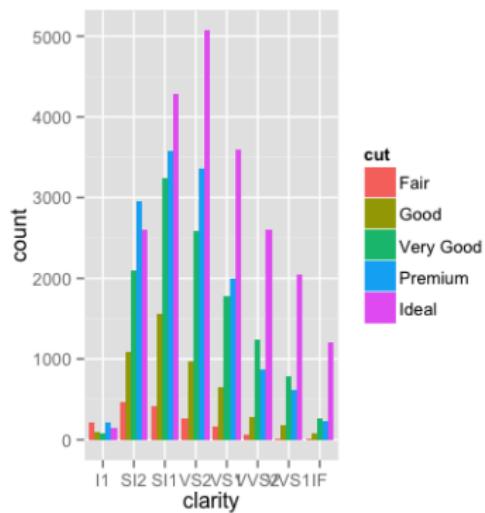


Just a Whisper of a Label

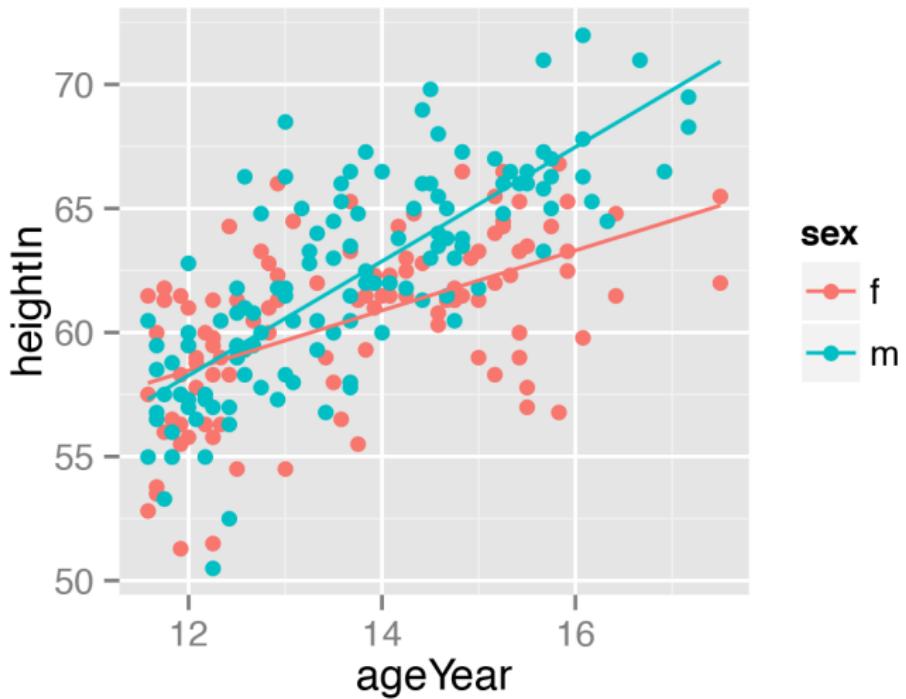
Gráficos do R



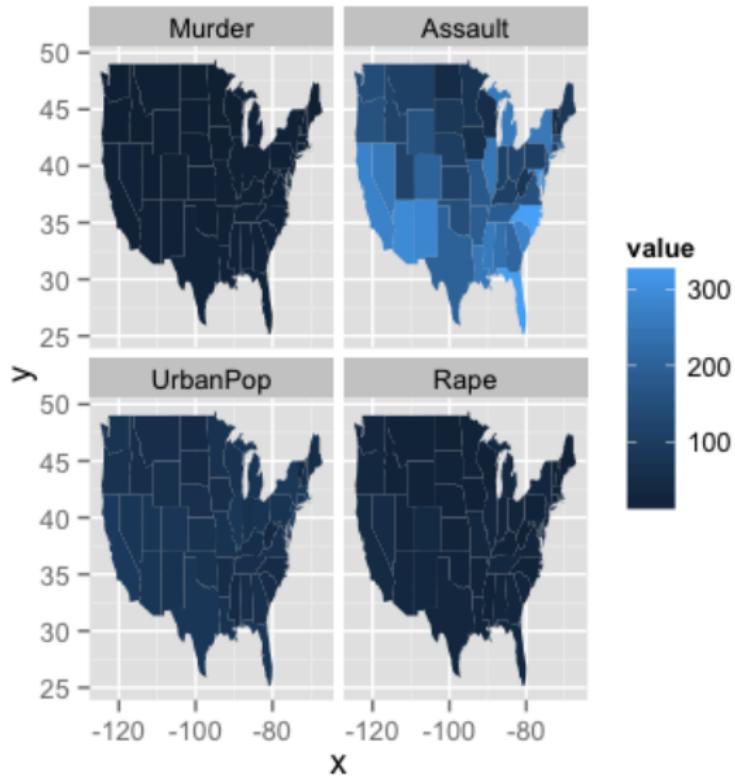
Gráficos do R



Gráficos do R



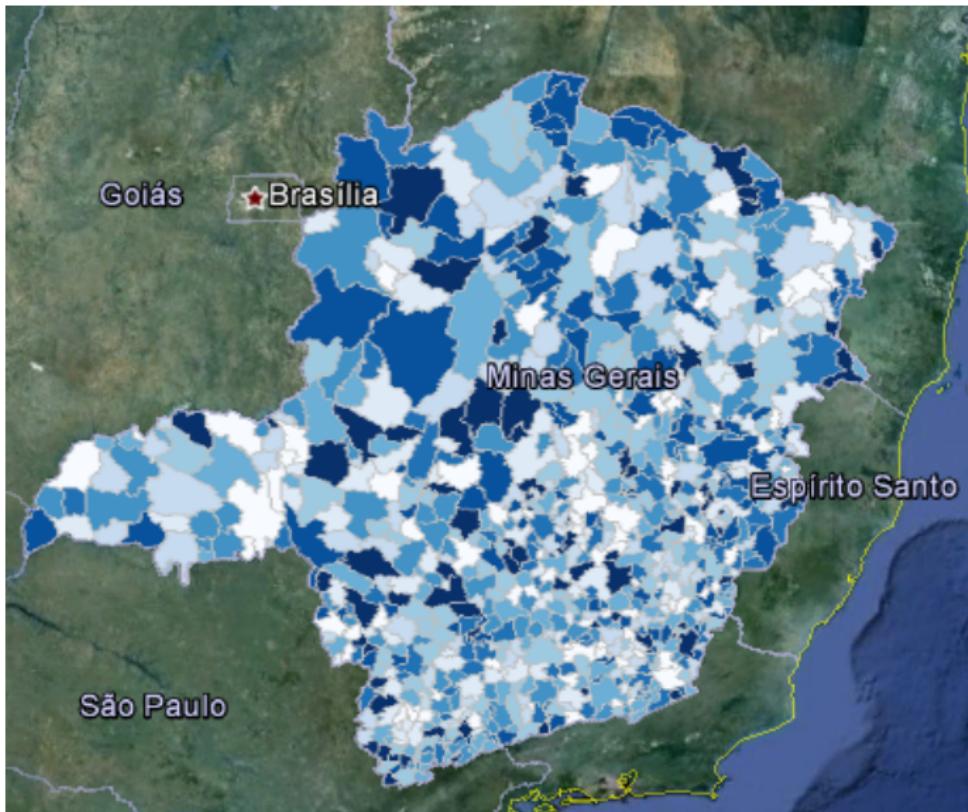
Gráficos do R



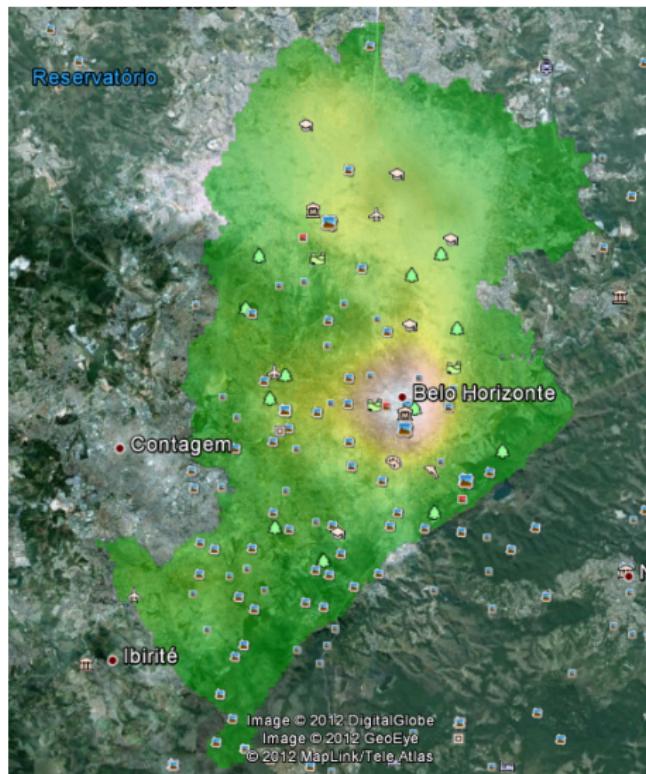
Gráficos do R



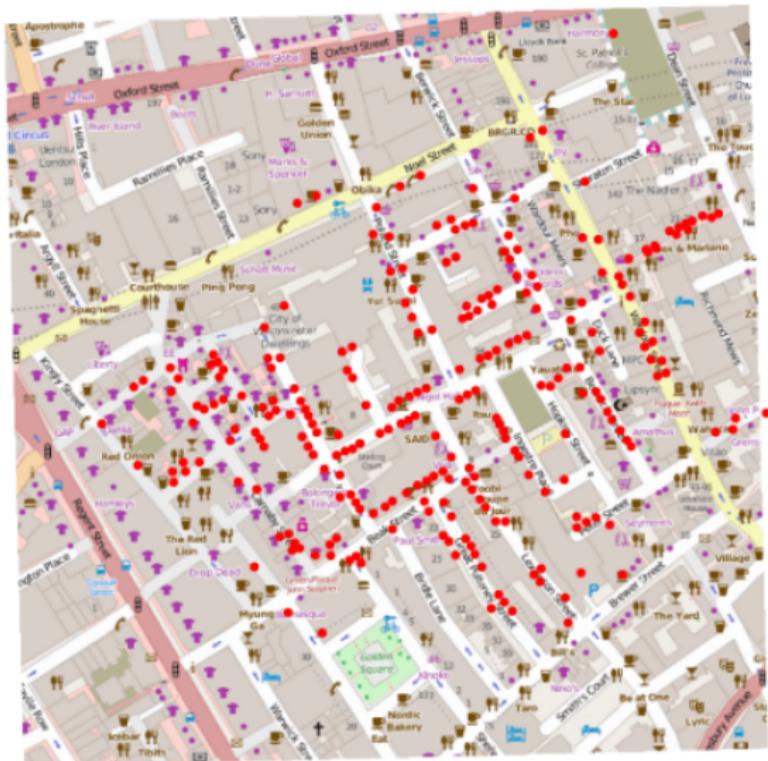
Gráficos do R



Gráficos do R



Gráficos do R



Comunicação de resultados através do R: R Markdown



1. Produz **documentos dinâmicos** em R.
2. Documentos R Markdown são completamente **reproduzíveis**.
3. R Markdown suporta dezenas de formatos de saída, incluindo **HTML, PDF, MS Word, Beamer, dashboards, aplicações shiny, artigos científicos** e muito mais.

Comunicação de resultados através do R: CompareGroups

Características dos grupos do estudo

	Total N=6324	Control N=2042	MDN N=2100	MDV N=2182	p-valor
Age	67.0 (6.17)	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Sex: Female	3645 (57.6%)	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	<0.001
Smoking:					0.444
Never	3892 (61.5%)	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	858 (13.6%)	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	1574 (24.9%)	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Waist circumference	100 [93.0;107]	101 [94.0;108]	100 [93.0;107]	100 [93.0;107]	0.085
Hormone-replacement therapy	97 (2.80%)	31 (2.64%)	30 (2.81%)	36 (2.95%)	0.898

Comunicação de resultados através do R: stargazer

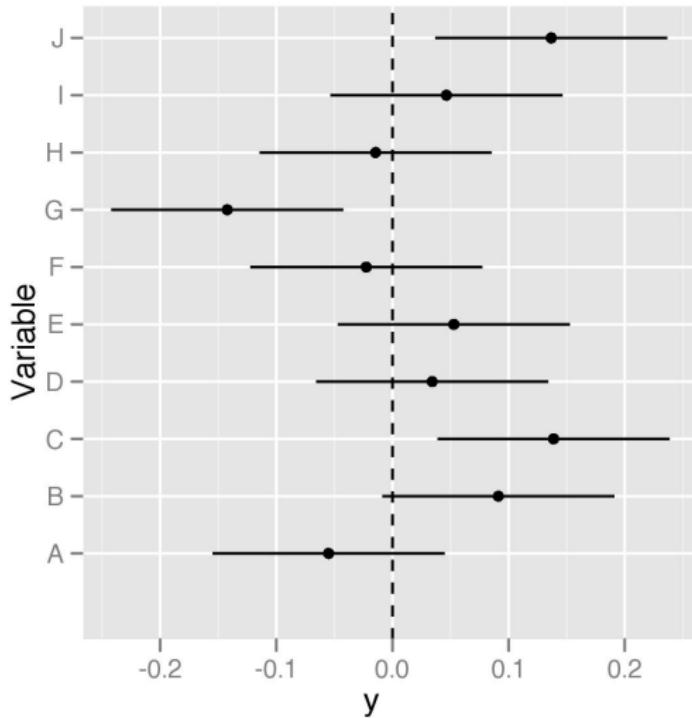
Estimativas dos efeitos fixos dos modelos simples.

	Variável resposta				
	Média de cinza				
	(1)	(2)	(3)	(4)	(5)
time1	4.190** (0.364, 8.016)	4.183** (0.355, 8.011)	4.190** (0.363, 8.017)	4.199** (0.372, 8.026)	4.191** (0.364, 8.019)
time2	9.155*** (4.789, 13.521)	9.138*** (4.768, 13.508)	9.161*** (4.791, 13.532)	9.081*** (4.712, 13.450)	9.178*** (4.808, 13.549)
forca.de.mordida	0.096*** (0.041, 0.150)				
idade		-1.241** (-2.376, -0.105)			
sexoFeminino			-6.492 (-27.707, 14.722)		
provisorioSim				16.420* (-0.556, 33.396)	
archMandibula					9.322 (-6.396, 25.040)
Constant	51.023*** (24.326, 77.721)	172.271*** (101.403, 243.139)	100.214*** (81.940, 118.489)	90.139*** (79.631, 100.646)	90.109*** (76.930, 103.287)
Observations	319	319	319	319	319

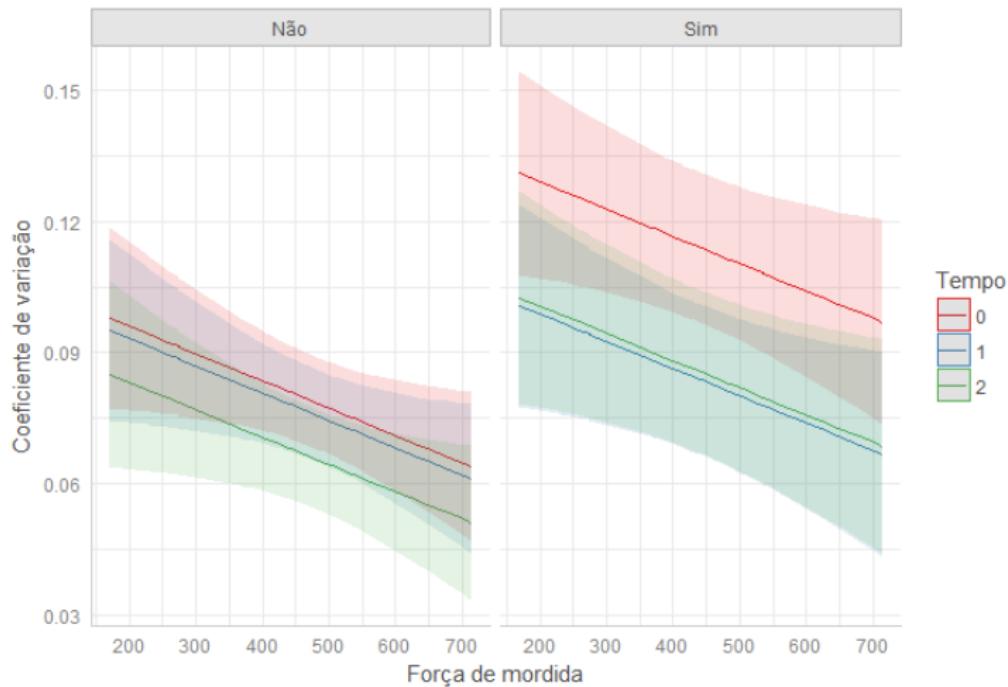
Note:

p<0.1; p<0.05; p<0.01

Comunicação de resultados através do R



Comunicação de resultados através do R



Comunicação de resultados através do R: Shiny

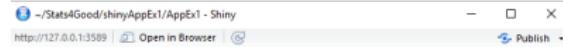
- ▶ Shiny é um pacote do R que torna mais fácil a construção de **aplicações web interativas** (apps) diretamente do R.
 - ▶ Permite a criação e compartilhamento de aplicativos.
 - ▶ Espera **nenhum conhecimento** de tecnologias web como HTML, CSS ou JavaScript (mas você pode aproveitá-las, caso as conheça)
 - ▶ Um aplicativo Shiny consiste em duas partes: uma **interface de usuário** (UI) e um **servidor**.

Shiny

```
# Run the application  
shinyApp(ui = ui, server = server)
```



Old Faithful Geyser Data



Old Faithful Geyser Data

Introdução ao R

Baixando e instalando o R

Para instalação do R acesse o site <https://www.r-project.org/>:

1. Em **Download** clique em CRAN.

- ▶ O **CRAN** (*The Comprehensive R Archive Network*) é uma rede de servidores ftp e web em todo o mundo que armazena versões de código e documentação idênticas e atualizadas para o R.

2. Escolha um repositório de sua preferência, por exemplo, Universidade Federal do Paraná (<http://cran-r.c3sl.ufpr.br/>).

3. Em **Download and Install R** clique no link adequado para o seu sistema operacional (no caso de Windows, clique no link **Download R for Windows**).

4. Clique no link **base** (no caso do sistema operacional ser Windows).

5. Finalmente clique no link para baixar o arquivo executável (a versão mais atual **Download R 3.5.1 for Windows**).

Após baixar o arquivo executável, abra-o e siga as etapas de instalação conforme as configurações padrões.

Baixando e instalando o RStudio

Para instalação do RStudio acesse o site

[https://www.rstudio.com/products/rstudio/download/.](https://www.rstudio.com/products/rstudio/download/)

- ▶ Em **Installers for Supported Platforms** baixe a versão mais recente do instalador do RStudio de acordo com o seu sistema operacional (no caso de Windows clique no link **RStudio 1.1.456 - Windows Vista/7/8/10**).

Pacotes

- ▶ Assim como a maioria dos softwares estatísticos, o R possui os seus “módulos”, mais conhecidos como **pacotes** do R.
- ▶ **Pacote:** é uma coleção de funções do R; os pacotes também são gratuitos e disponibilizados no **CRAN**.
- ▶ Um pacote inclui: **funções** do R, **conjuntos de dados** (utilizados em exemplos das funções), arquivo com **ajuda (help)**, e uma **descrição** do pacote.
- ▶ Atualmente, o repositório oficial do R possui mais de 12.000 pacotes disponíveis.
- ▶ As funcionalidades do R, podem ser ampliadas carregando estes pacotes, tornando-o um software muito poderoso, capaz de realizar inúmeras tarefas.

Pacotes

- ▶ Alguns exemplos destas tarefas e alguns destes pacotes são listados abaixo:
 - ▶ **Importação e exportação de dados**
 - ▶ `foreign`, `readr`, `haven`
 - ▶ **Manipulação de dados**
 - ▶ Transformação e recodificação de variáveis: `reshape2`, `stringr`
 - ▶ **Visualização de dados**
 - ▶ Diversos gráficos: `graphics`, `ggplot2`, `ggthemes`
 - ▶ Mapas: `ggmap`
 - ▶ Gráficos e mapas interativos: `plotly`
 - ▶ **Análise de dados**
 - ▶ Análise descritiva: `compareGroups`
 - ▶ Ajuste de modelos: `stats`, `survival`
 - ▶ Análise de amostras complexas: `survey`

Pacotes

- ▶ **Geração de relatórios**

- ▶ Relatórios nos formatos pdf, HTML, Word, Power Point: `knitr`,
`rmarkdown`, `officer`

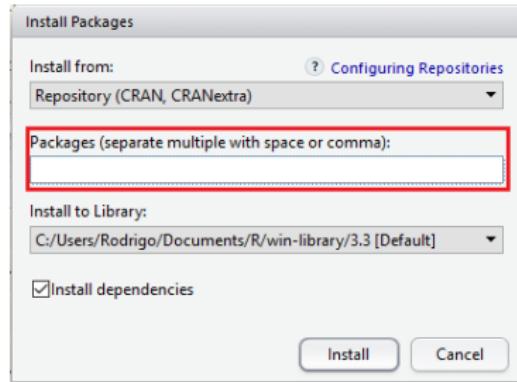
Instalando pacotes

- ▶ Para **instalação de um pacote**, basta um simples comando.

```
install.packages("survey")
```

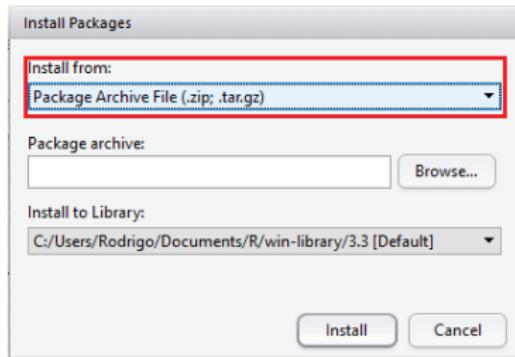
- ▶ Além da opção de comando, também podemos instalar pacotes utilizando o menu **Tools** do RStudio, opção **Install packages ...** e preenchendo com o(s) nome(s) do(s) pacote(s):

Instalando pacotes



- ▶ Outra opção é instalar o pacote a partir de seu arquivos fonte ([.zip](#) ou [.tar.gz](#)):
 - ▶ Para isso, obtenha o arquivo fonte do pacote (geralmente através do [CRAN](#)) e no menu **Tools** do RStudio, opção **Install packages ...** em **Install from** escolha a seguinte opção:

Instalando pacotes



Após a instalação do pacote, temos que **carregar o pacote** para nossa área de trabalho para podermos usufruir de suas funções.

```
library("survey")
require("survey")
```

Obtendo ajuda no R

- ▶ Para conhecer quais as funções disponíveis no pacote, faça:

```
help(package = "survey")
```

- ▶ Para pedir ajuda de uma determinada função:

```
?glm  
help("glm")
```

- ▶ Obtendo ajuda na internet:

```
help.search("t.test")
```

Obtendo ajuda no R

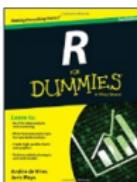
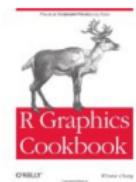
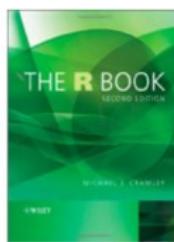
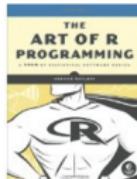
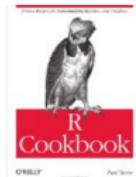
- ▶ Procurando por alguma função, mas esqueci o nome:

```
apropos("lm")
```

- ▶ Para todas as outras dúvidas existe o **Google!**
- ▶ Ver também <http://www.r-bloggers.com/> e <https://rstudio.cloud/>
- ▶ Para algumas demonstrações da capacidade gráfica do R:

```
demo(graphics)
demo(persp)
demo(Hershey)
demo(plotmath)
```

Algumas referências



Bons estudos!

