

EPI02 - Fundamentos de Bioestatística

Apresentações

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

Porto Alegre, 2024



O professor

Olá!

- ▶ Desde outubro de 2017 eu sou Professor do Departamento de Estatística e faço parte do Corpo Docente do Programa de Pós Graduação em Epidemiologia da Universidade Federal do Rio Grande do sul (UFRGS).
- ▶ Eu me formei Bacharel em Estatística pelo Departamento de Estatística da UFRGS em 2008, e Mestre (2010) e Doutor (2014) em Estatística pelo Programa de Pós Graduação em Estatística da Universidade Federal de Minas Gerais (UFMG).
 - ▶ Período de pós-doutoramento no Programa de Pós-Graduação em Saúde Pública da UFMG (2015-2017).

Olá!

- ▶ Na graduação eu tenho me concentrado nas disciplinas de Estatística Descritiva, Amostragem e Modelos para dados correlacionados.
 - ▶ Na pós-graduação: Seminários de Doutorado II e III e Análise de Sobrevida.

A disciplina

Objetivos

- ▶ Identificar técnicas básicas para descrição de dados;
- ▶ Desenvolver os princípios básicos da inferência estatística;
- ▶ Auxiliar na escolha de testes estatísticos apropriados às várias situações básicas de pesquisa.

Organização

- ▶ **Disciplina:** Fundamentos de Bioestatística
- ▶ **Turma:** U
- ▶ **Modalidade:** Ensino presencial
- ▶ **Professor:** Rodrigo Citton Padilha dos Reis
 - ▶ e-mail: citton.padilha@ufrgs.br ou rodrigocpdosreis@gmail.com
 - ▶ Sala: B215 do Instituto de Matemática e Estatística; 518 do Instituto de Psicologia (Centro de Dados ELSA-Brasil)
- ▶ **Monitora:** Letícia Menegotto

Aulas e material didático

- ▶ **Aulas** (teóricas e práticas)
 - ▶ Exposição e **discussão** dos conteúdos
 - ▶ Faremos leituras de artigos e capítulos de livros
 - ▶ Exemplos
- ▶ **Notas de aula**
 - ▶ Slides
 - ▶ Arquivos de rotinas em R
- ▶ **Exercícios**
 - ▶ Listas de exercícios
 - ▶ Para casa
- ▶ **Canais de comunicação:**
 - ▶ Durante as aulas
 - ▶ Moodle: aulas, materiais, listas de exercícios e **fórum geral**
 - ▶ e-mail do professor

Aulas e material didático

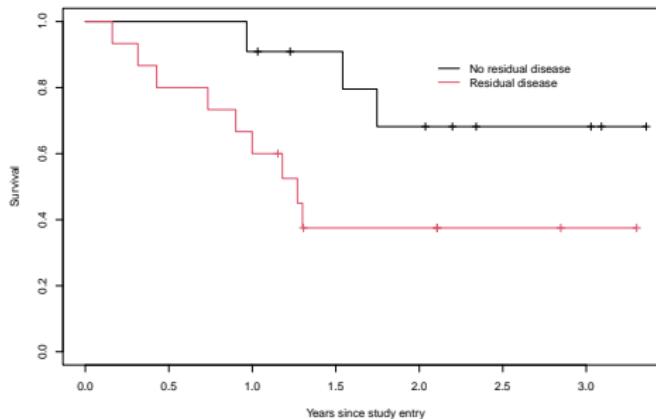
- ▶ **Aulas:** quintas, das 14h às 16h, na **Sala 527 (Lab. de Informática)** do **Instituto de Psicologia, 5ºandar** - Campus Saúde

Aulas e material didático



- ▶ Exemplos e exercícios com o apoio do computador:
 - ▶ R e RStudio

```
library(survival)
fit1 <- survfit(Surv(futime,fustat) ~ resid.ds,
                  data = ovarian)
```



Conteúdo programático

1. Variabilidade e intervalos de confiança

- ▶ Variabilidade e o papel da estatística
- ▶ Conceitos básicos
- ▶ Levantamentos estatísticos
- ▶ Descrevendo a variabilidade I: técnicas descritivas
- ▶ Descrevendo a variabilidade II: intervalos de confiança
- ▶ Estimando probabilidades: riscos e prevalências

2. comparando riscos e chances

- ▶ Estimando riscos relativos
- ▶ Estimando chances e razões de chances
- ▶ Comparando duas probabilidades ou chances

3. Taxas e estudos caso-controle

- ▶ Medidas de frequência de doença: estimando taxas
- ▶ Estudos caso-controle

Conteúdo programático

4. Ajuste para confundimento usando análise estratificada
 - ▶ Confundimento
 - ▶ Interação: modificação de efeito
5. Padronização e riscos atribuíveis
 - ▶ Método da padronização direta
 - ▶ Método da padronização indireta
 - ▶ Risco atribuível populacional
6. Introdução à regressão logística
 - ▶ Regressão logística simples
 - ▶ Regressão logística múltipla

Avaliação

- ▶ Serão realizadas três atividades de avaliação: P_1 , P_2 e P_3 ;
- ▶ Cada atividade de avaliação vale 10 pontos.
- ▶ Será realizada uma prova presencial e individual como atividade de recuperação (*PR*):
 - ▶ Para os(as) alunos(as) que não atingirem o conceito mínimo;
 - ▶ **Esta prova abrange todo o conteúdo da disciplina.**

Avaliação

$$MF = \frac{P_1 + P_2 + P_3}{3}$$

- ▶ **A:** $9 \leq MF \leq 10$
- ▶ **B:** $7,5 \leq MF < 9$
- ▶ **C:** $6 \leq MF < 7,5$
- ▶ **D:** $MF < 6$
- ▶ **FF:** se o(a) aluno(a) tiver frequência inferior a 75% da carga horária prevista no plano da disciplina

Avaliação

- Se $MF < 6$ e frequência mínima de 75%, o(a) aluno(a) poderá realizar a prova de recuperação e neste caso

$$MF' = MF \times 0,4 + PR \times 0,6$$

- C: $MF' \geq 6$
- D: $MF' < 6$

Referências bibliográficas



Principais

Kahn, H. A. e Sempos, C. T. **Statistical Methods in Epidemiology**. New York: Oxford University Press, 1989.

Kleinbaum, D. G., Kupper, L. L., Morgenstern, H. **Epidemiologic Research: Principles and Quantitative Methods**. New York: John Wiley & Sons, 1982.

Complementares

Callegari-Jacques, S. M. **Bioestatística: Princípios e aplicações**. Porto Alegre: Artmed, 2003.

Woodward, M. **Epidemiology: Study Design and Data Analysis**, Third Edition. Boca Raton: CRC Press, 2014.

Epidemiologia e Bioestatística

Epidemiologia e Bioestatística

Epidemiologia¹

"é o estudo da distribuição e de determinantes de estados ou eventos relacionados com a saúde em populações especificadas e com a aplicação desse estudo para controlar problemas de saúde".

- O que deve ser visto com atenção nessa definição é que inclui tanto as descrições do conteúdo da disciplina quanto proposta ou aplicação para as quais as investigações epidemiológicas são realizadas.

Epidemiologia e Bioestatística

Quais são os objetivos específicos da epidemiologia?

1. Identificar a **etiologia** ou a **causa de uma doença** e os fatores de risco relevantes.
2. Determinar a extensão da doença encontrada em uma comunidade.
3. Estudar a história natural e o prognóstico da doença.
4. Avaliar medidas preventivas e terapêuticas e modelos novos ou existentes de assistência à saúde.
5. Fornecer fundamentos para o desenvolvimento de políticas públicas relacionando problemas ambientais, questões genéticas e outras no que diz respeito à prevenção de doenças e promoção da saúde.

Epidemiologia e Bioestatística

- ▶ A **epidemiologia** e a **bioestatística** são ciências básicas para a saúde pública (saúde coletiva).
 - ▶ Investigações de saúde pública, geralmente, usam **métodos quantitativos** que combinam as duas disciplinas.
- ▶ Para investigar uma questão de pesquisa **dados** são necessários.
 - ▶ Algumas vezes os dados já estão disponíveis, em outras novos dados precisam ser coletados.

Epidemiologia e Bioestatística

- ▶ Dados apresentam **varabilidade**.
- ▶ 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.

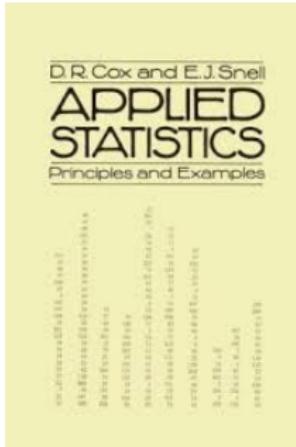
¹Porta M: A Dictionary of Epidemiology, 5th ed. New York, Oxford University Press, 2008.

Analisando dados



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



Um breve exemplo: prevalência do diabetes

Definição e diagnóstico do diabetes

Medição da doença

a proporção de prevalência

Instrumento de medida

Flower City

População e amostra

Censo e amostragem

Parâmetros e estatísticas

A variação amostral

Variação inter- e intraindivíduo

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.



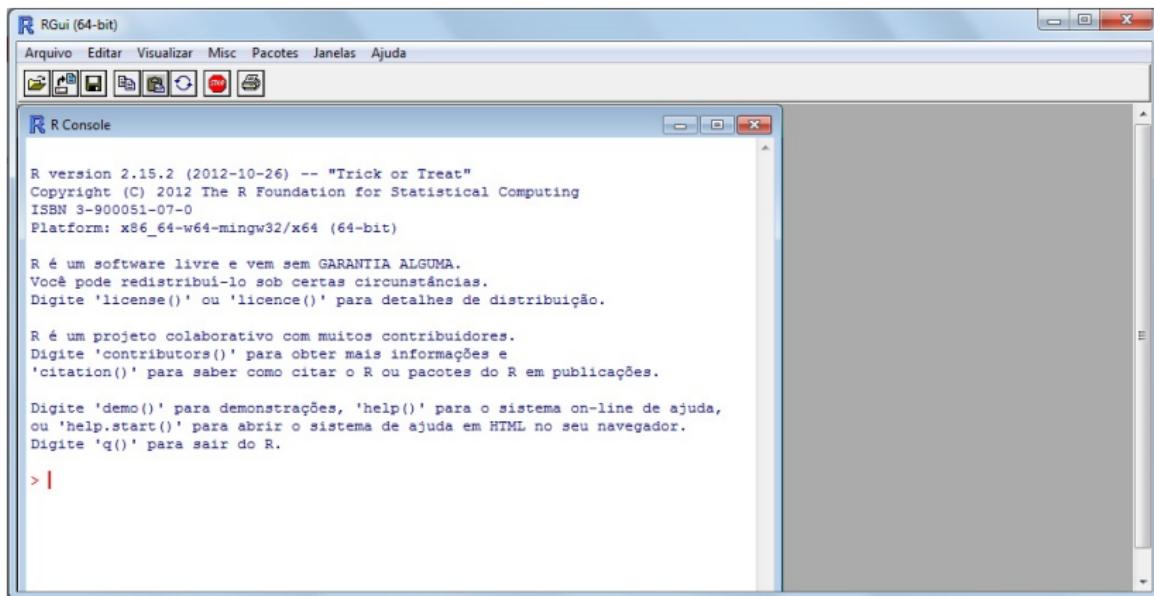
Breve histórico

- ▶ 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.

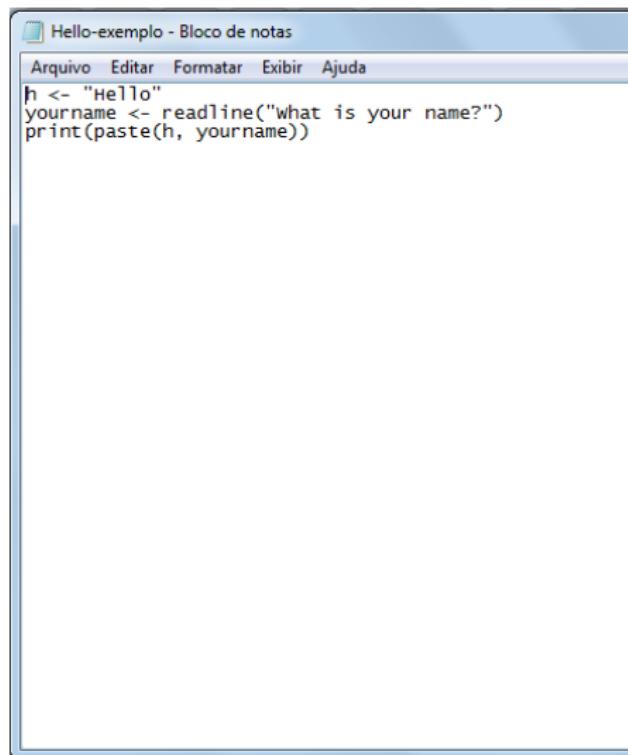
Interface do R



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**.

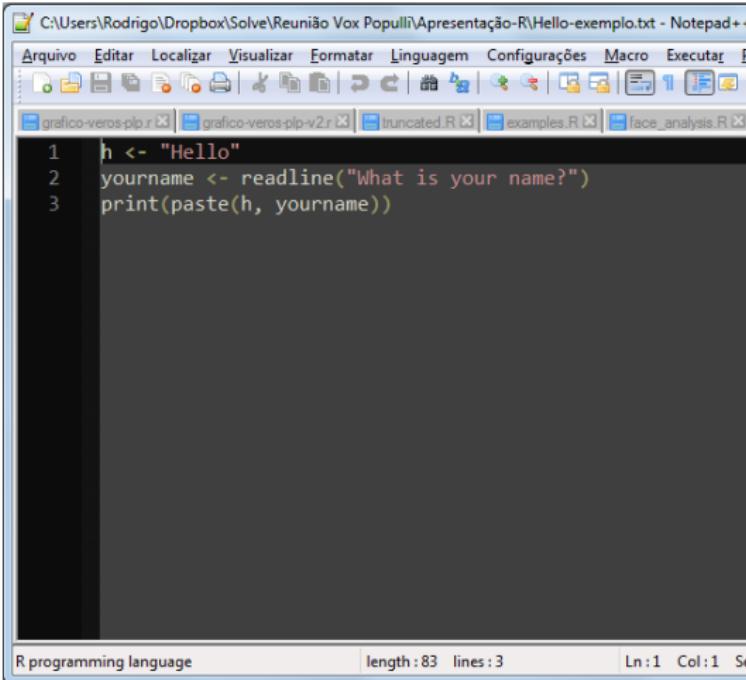
Editores de texto



The image shows a screenshot of a Windows Notepad window. The title bar reads "Hello-exemplo - Bloco de notas". The menu bar includes "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The main content 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. Below the menu is a toolbar with various icons. The main window displays the following R code:

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

The status bar at the bottom shows "R programming language", "length : 83 lines : 3", and "Ln : 1 Col : 1 Sel".

Interface do RStudio

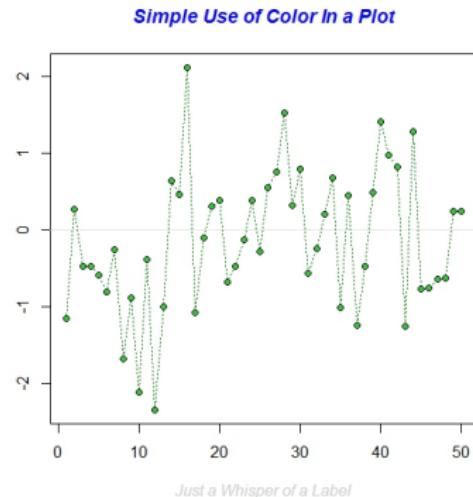
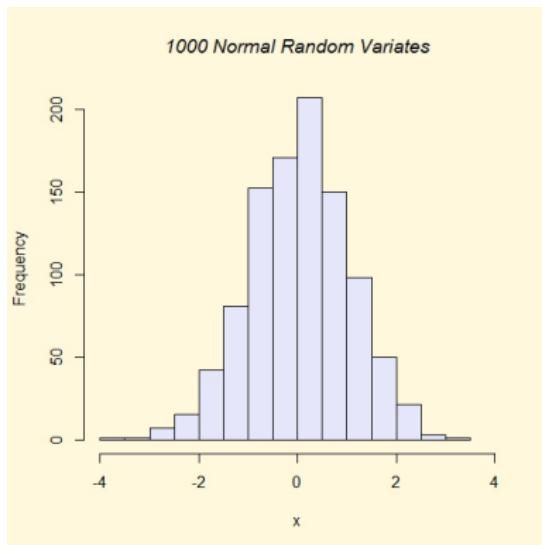
The screenshot displays the RStudio environment with several panes:

- Source Pane:** Shows R code for generating a scatter plot. The code uses the ggplot2 package to create a scatter plot of 'classn' vs 'V1'. It includes a 'geom_point' layer with jittered data points and a 'stat_smooth' layer using a binomial family.
- Console Pane:** Displays the same R code as the Source pane, indicating it was run successfully.
- Workspace Pane:** Shows the dataset 'b' loaded, containing 699 observations and 12 variables.
- Plots Pane:** Contains a scatter plot of 'classn' (y-axis, 0.0 to 1.0) versus 'V1' (x-axis, 2.0 to 10.0). The plot shows a clear positive trend with a blue smoothing line and a gray shaded area representing the confidence interval.

Você pode usar o R para

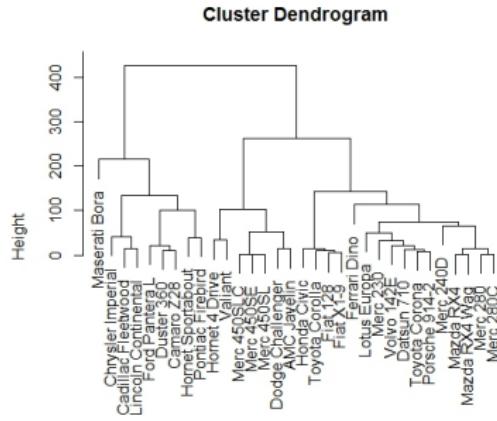
- ▶ Importação e exportação de dados
- ▶ Manipulação de dados
 - ▶ Transformação e recodificação de variáveis
 - ▶ Aplicação de filtros
- ▶ Visualização de dados
 - ▶ Diversos gráficos
 - ▶ Mapas
 - ▶ Gráficos e mapas interativos
- ▶ Análise de dados
 - ▶ Análise descritiva
 - ▶ Ajuste de modelos
 - ▶ Técnicas multivariadas
 - ▶ Análise de amostras complexas

Gráficos do R

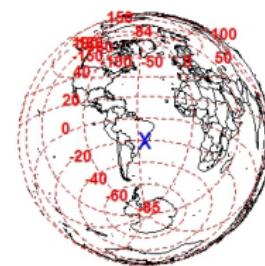


Just a Whisper of a Label

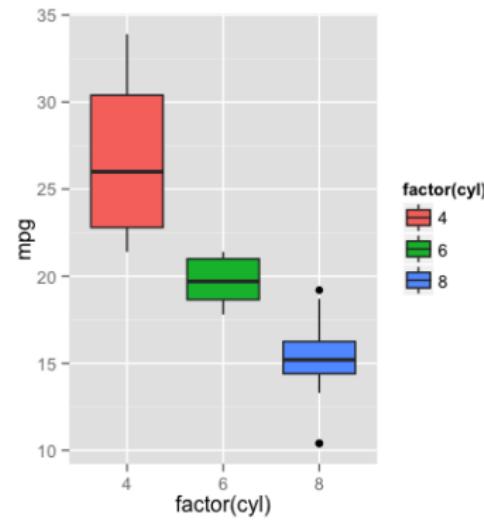
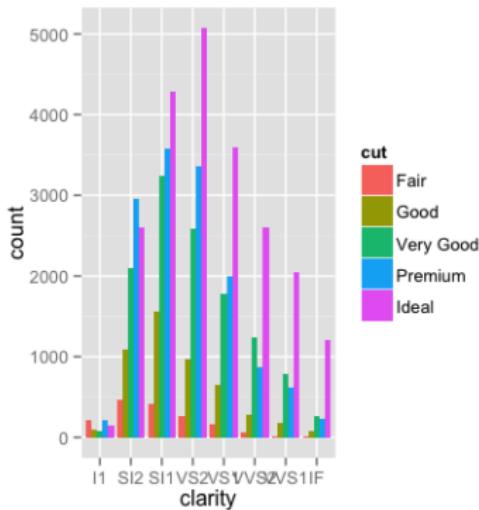
Gráficos do R



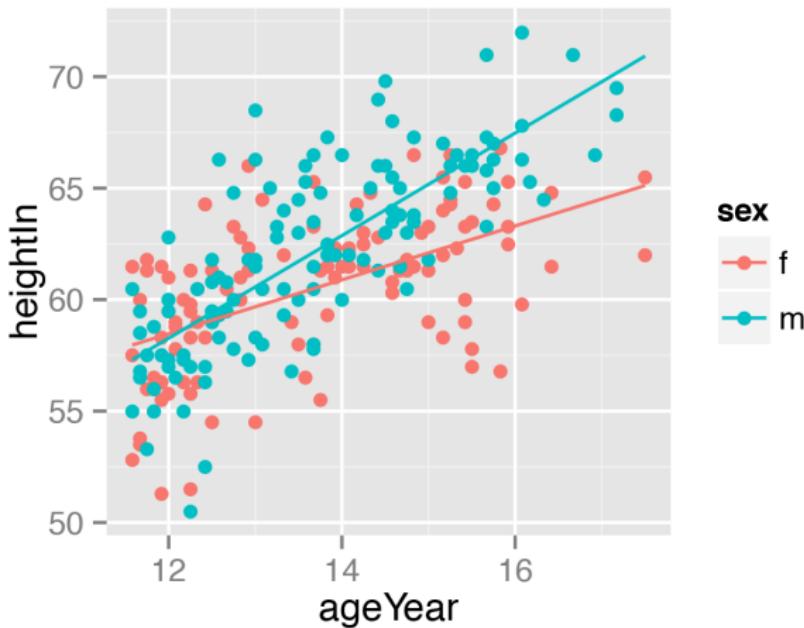
d
hclust (*, "complete")



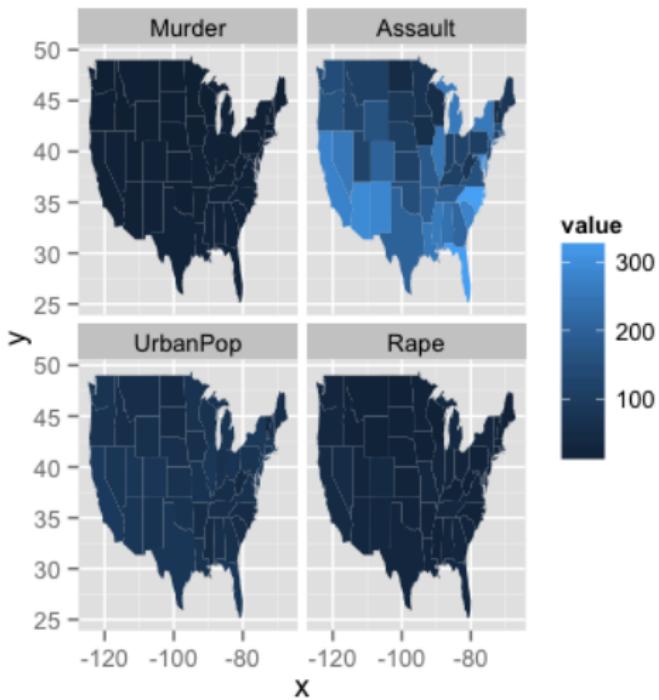
Gráficos do R



Gráficos do R



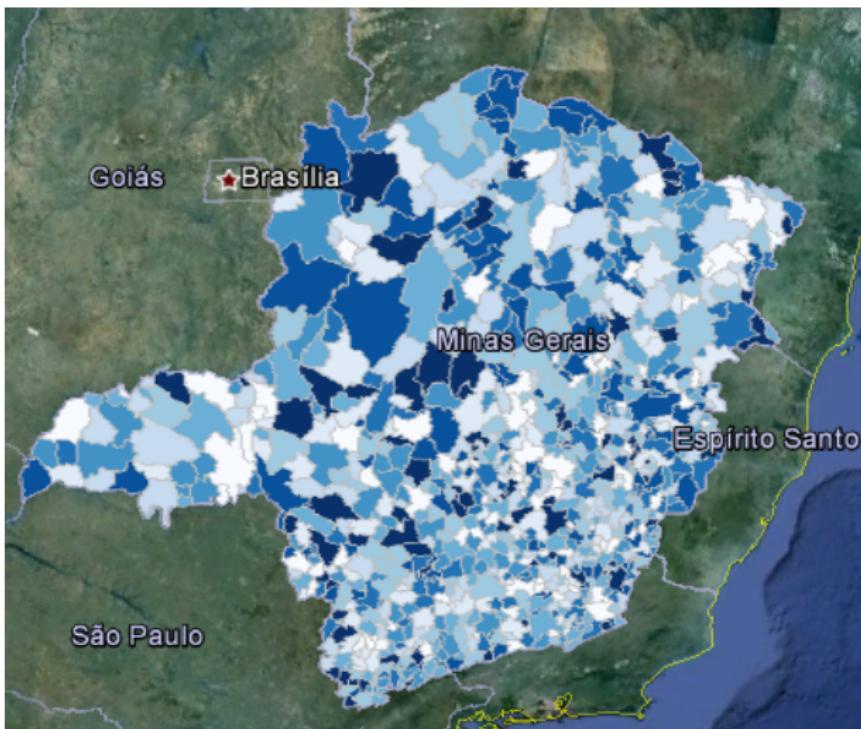
Gráficos do R



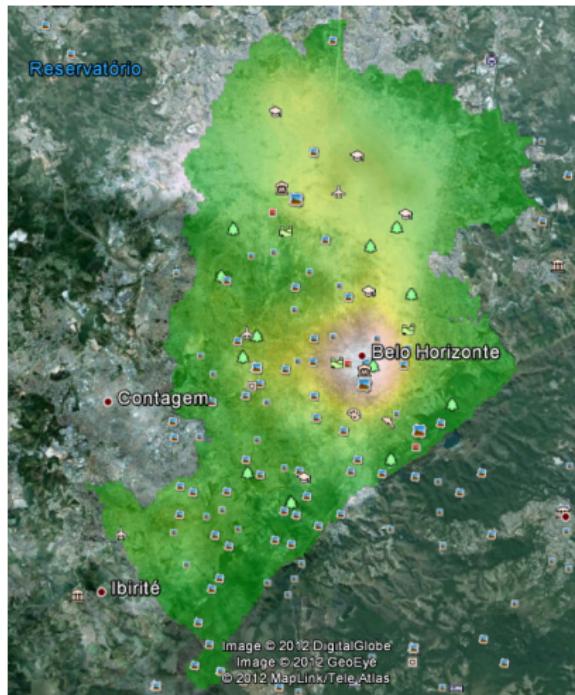
Gráficos do R



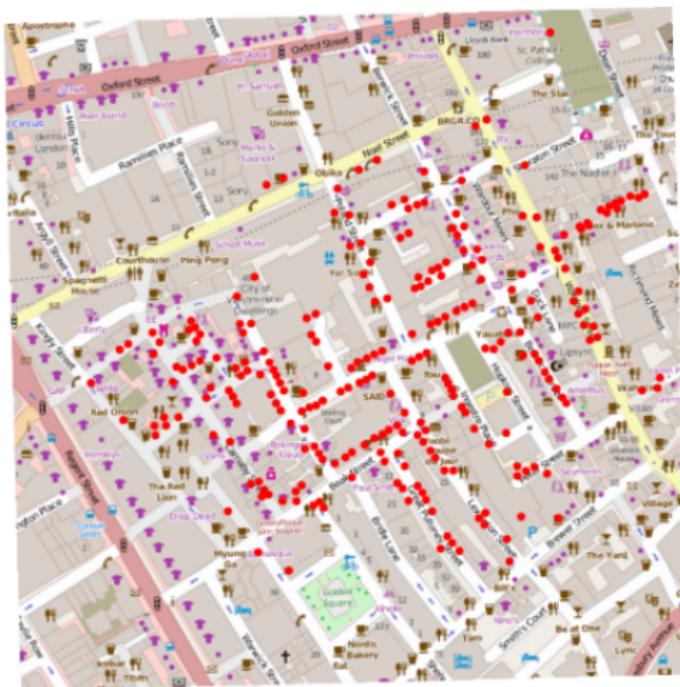
Gráficos do R



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

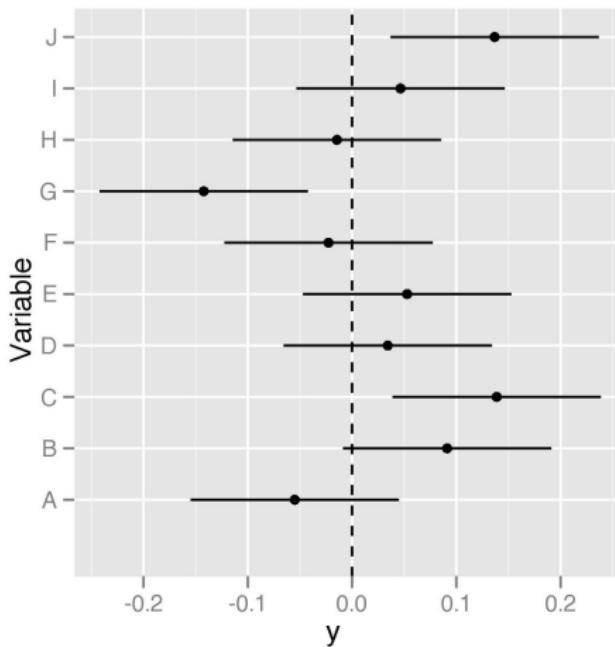
Estimativas dos efeitos fixos dos modelos simples.

	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)	
archMandíbula					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

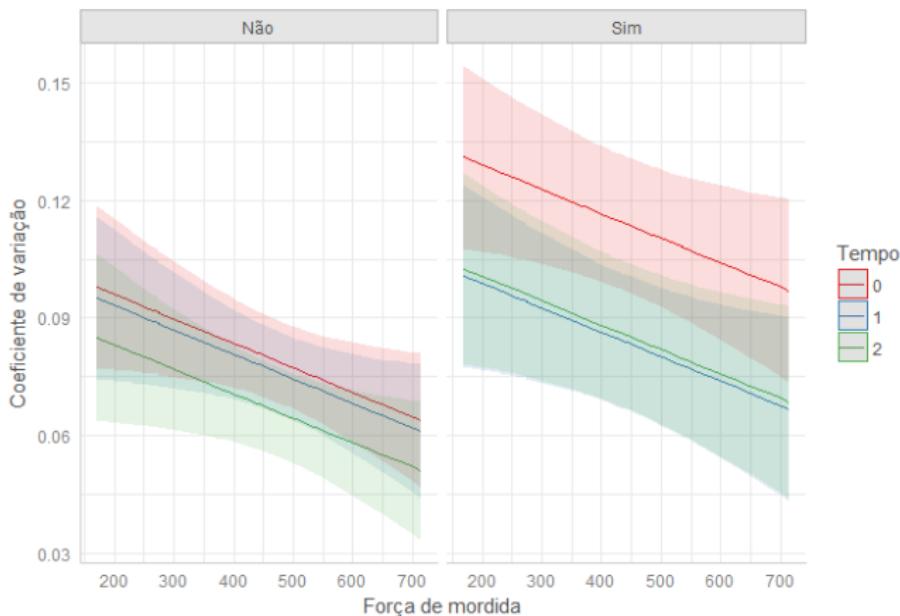
Note:

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

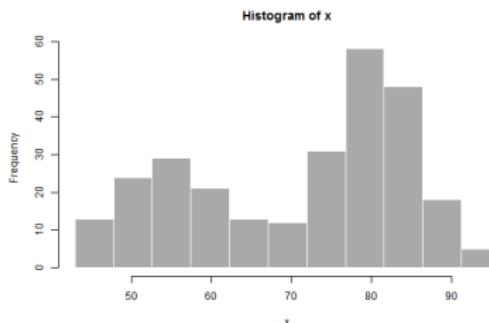
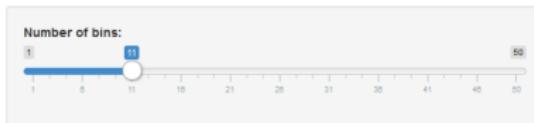
- ▶ Shiny é um pacote do R que torna mais fácil a construção de **aplicações web interativas** (apps) diretamente do R.
 - ▶ Permite a criação e compartilhamento de aplicativos.
 - ▶ Espera **nenhum conhecimento** de tecnologias web como HTML, CSS ou JavaScript (mas você pode aproveitá-las, caso as conheça)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

Shiny

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

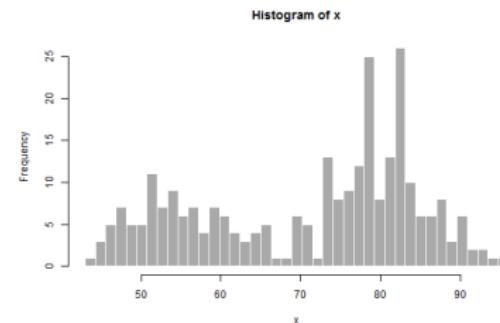
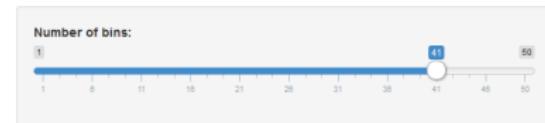
~/Stats4Good/shinyAppEx1/AppEx1 - Shiny
http://127.0.0.1:3589 | Open in Browser | Publish

Old Faithful Geyser Data



~/Stats4Good/shinyAppEx1/AppEx1 - Shiny
http://127.0.0.1:3589 | Open in Browser | Publish

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 4.3.3 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 **Install RStudio**, clique no botão **DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS** para baixar a versão mais recente do instalador do RStudio (no caso de Windows, **RSTUDIO-2023.12.1-402**).

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 20.500 pacotes disponíveis.
- ▶ As funcionalidades do R, podem ser ampliadas carregando estes pacotes, tornando-o um software muito poderoso, capaz de realizar inúmeras tarefas.

Pacotes

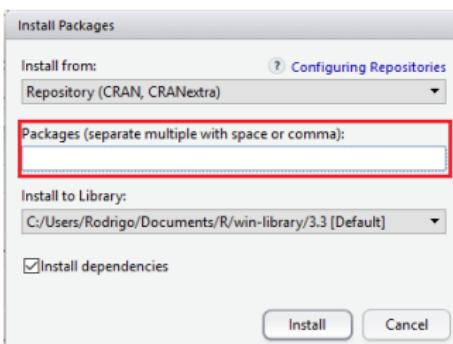
- ▶ Alguns exemplos destas tarefas e alguns destes pacotes são listados abaixo:
 - ▶ Importação e exportação de dados
 - ▶ foreign, readr, haven
 - ▶ Manipulação de dados
 - ▶ Transformação e recodificação de variáveis: reshape2, stringr
 - ▶ Visualização de dados
 - ▶ Diversos gráficos: graphics, ggplot2,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

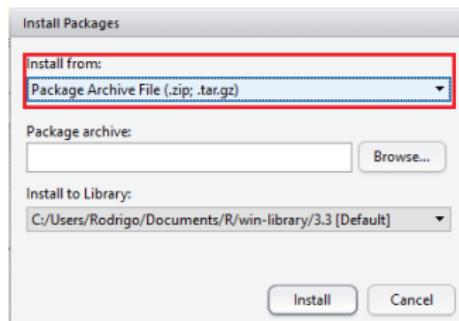
- ▶ Além da opção de comando, também podemos instalar pacotes utilizando o menu **Tools** do RStudio, opção **Install packages ...** e preenchendo com o(s) nome(s) do(s) pacote(s):



- ▶ Outra opção é instalar o pacote a partir de seu arquivos fonte (**.zip** ou **.tar.gz**):

Instalando pacotes

- ▶ 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:



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

