

MPS814 - Tópicos Especiais em Epidemiologia

R para epidemiologia

Rodrigo Citton Padilha dos Reis
rodrigocpdosreis@gmail.com

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE PÚBLICA

Belo Horizonte
Junho de 2017

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.
- ▶ Informação (dados) é coletada para investigar uma questão de pesquisa.
- ▶ Os métodos e ferramentas de bioestatística são usados para analisar os dados, resumir a informação e gerar conhecimento para auxiliar na tomada de decisão.

Epidemiologia

Epidemiologia é o estudo da distribuição e determinantes de saúde, doença, ou danos em populações humanas e da aplicação deste estudo para o controle de problemas de saúde.

► Exemplos:

- ▶ Sistema nacional de vigilância (câncer, AIDS, a ocorrência de Dengue).
- ▶ Estudo de coorte para investigar a associação entre o uso de celulares e o desenvolvimento de tumores cerebrais.
- ▶ Levantamento de indivíduos que tomaram inibidores da COX-2.

Bioestatística

Estatística é a ciência que lida com a variação dos dados para obter resultados e conclusões confiáveis.

Bioestatística é a aplicação da estatística a problemas nas ciências biológicas, médicas e da saúde.

- ▶ Exemplos:
 - ▶ Cálculo de taxas de incidência de câncer ajustadas por idade para determinar as tendências ao longo do tempo e localidade.
 - ▶ Cálculo de medidas estatísticas do risco de desenvolver tumores cerebrais após o uso de telefone celular ajustando para possíveis variáveis de confusão.
 - ▶ Quantificação da relação entre o uso de inibidores de COX-2 e a qualidade de vida.

O papel dos métodos quantitativos na saúde pública

1. Abordar uma questão de saúde pública

- ▶ Gerar uma **hipótese** com base
 - ▶ na fundamentação científica
 - ▶ em observações e evidências ainda não testadas
 - ▶ em resultados de estudos prévios
- ▶ Exemplos de hipóteses
 - ▶ O risco de desenvolver câncer de pulmão permanece constante nos últimos cinco anos.
 - ▶ O uso do telefone celular está associado com o desenvolvimento de tumor cerebral.
 - ▶ Vioxx aumenta o risco de doença cardíaca.

O papel dos métodos quantitativos na saúde pública

2. Conduzir um estudo

- ▶ **Estudo tipo survey** é utilizado para estimar a extensão da doença na população.
- ▶ **Estudo de vigilância** é projetado para monitorar ou detectar doenças específicas.
- ▶ **Estudos observacionais** investigam associação entre uma exposição e uma doença (desfecho ou evento de interesse).
 - ▶ Alocação “natural” dos indivíduos em grupos expostos ou não expostos.
- ▶ **Estudos experimentais** também investigam a associação entre a exposição, usualmente um tratamento terapêutico, e a doença.
 - ▶ Os indivíduos são “intencionalmente” alocados em grupos de tratamento pelos investigadores.

O papel dos métodos quantitativos na saúde pública

3. Coletar dados

- ▶ Fatos numéricos, medidas, ou observações obtidas a partir de uma investigação para responder a uma pergunta.
- ▶ Exemplos
 - ▶ O número de casos de câncer pulmonar de 1960-2000 no Brasil.
 - ▶ O número de mortes por doenças cardiovasculares em homens e mulheres a partir do ano 2000.
 - ▶ O número de pessoas com ataques cardíacos entre os indivíduos que utilizaram o Vioxx antes de 2004.

O papel dos métodos quantitativos na saúde pública

4. Descrever os dados/observações

- ▶ **Métodos estatísticos descritivos** fornecem uma avaliação exploratória dos dados de um estudo.
 - ▶ Técnicas de análise exploratória de dados.
 - ▶ Organização e resumo dos dados: tabelas, gráficos, medidas resumo.

O papel dos métodos quantitativos na saúde pública

5. Avaliar a força da evidência a favor/contra a hipótese

- ▶ **Métodos estatísticos inferenciais** fornecem uma análise de dados de “confirmatória”.
 - ▶ Permitem generalizar conclusões a partir de parte dos dados de um grupo (**amostra**) para todo o grupo (**população**).
 - ▶ Permitem avaliar a força da evidência.
 - ▶ Realização de comparações.
 - ▶ Possibilitam fazer previsões.
 - ▶ A partir dos resultados, novas perguntas podem ser feitas; sugerem pesquisas futuras.

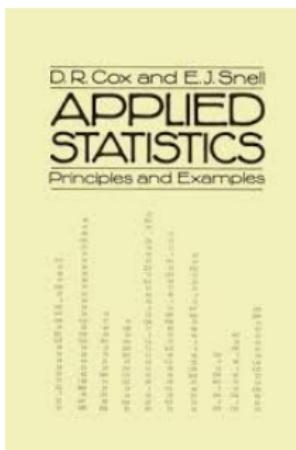
O papel dos métodos quantitativos na saúde pública

6. Recomendar intervenções ou programas preventivos

- ▶ Os resultados do estudo irão provar ou refutar a hipótese.
- ▶ Os resultados do estudo serão publicados em periódicos científicos e/ou divulgados ao público por outros meios.
- ▶ Como consequência, a política ou ação pode variar desde o desenvolvimento de programas de regulamentação específica até mudanças comportamentais pessoais gerais.

O papel dos métodos quantitativos na saúde pública

Fases de análise



1. Manipulação inicial dos dados.

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

2. Análise preliminar.

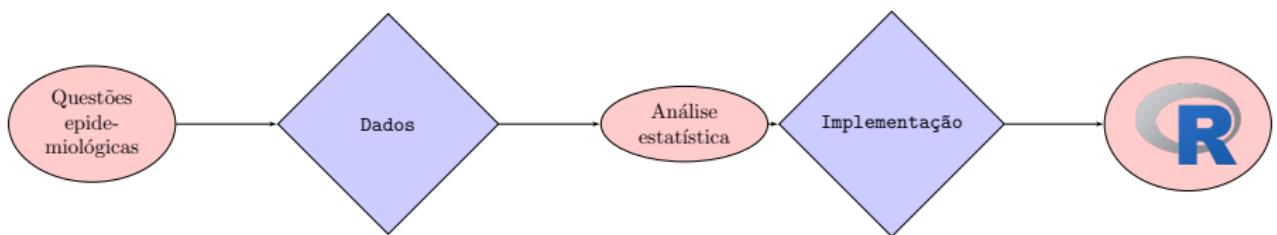
- ▶ Conhecimento dos dados, identificação de outliers, investigação preliminar.

3. Análise definitiva.

- ▶ Disponibiliza a base para as conclusões.

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

O papel dos métodos quantitativos na saúde pública



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

- ▶ **Supervisão:** Professora Sandhi M. Barreto
- ▶ **Sala de aula:** Laboratório 535
- ▶ **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 do pacote base do R
- ▶ Gráficos avançados
- ▶ 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.



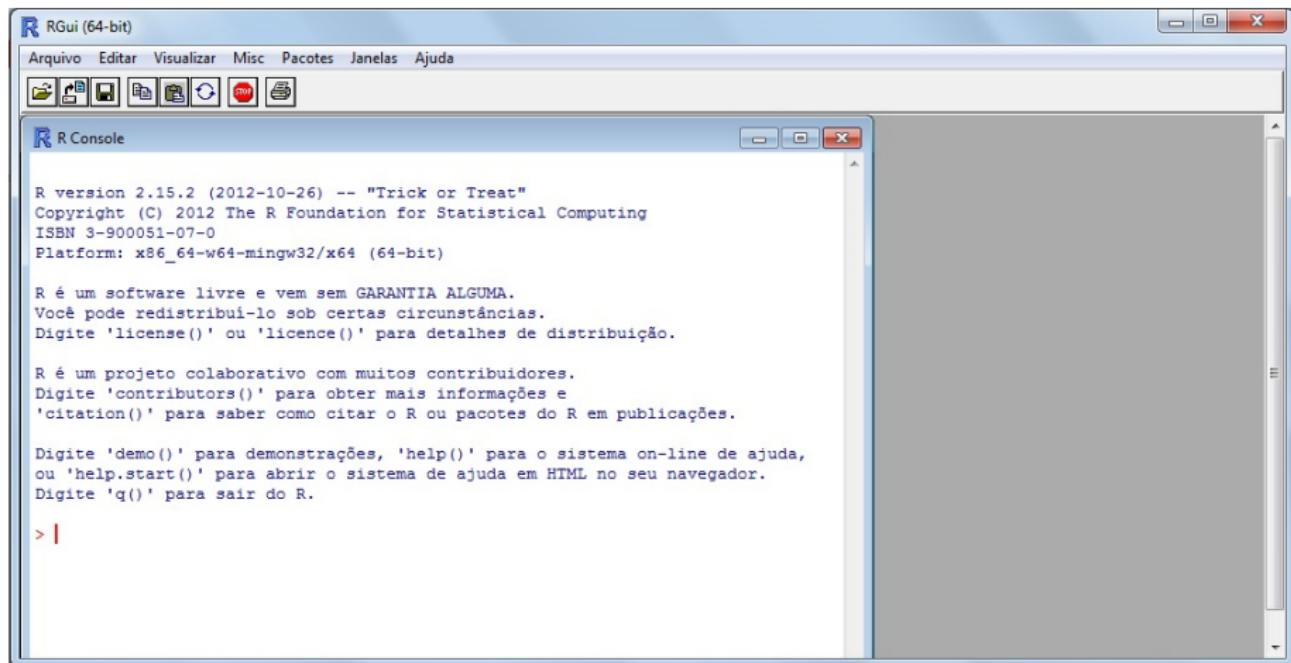
Breve histórico

- ▶ 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.

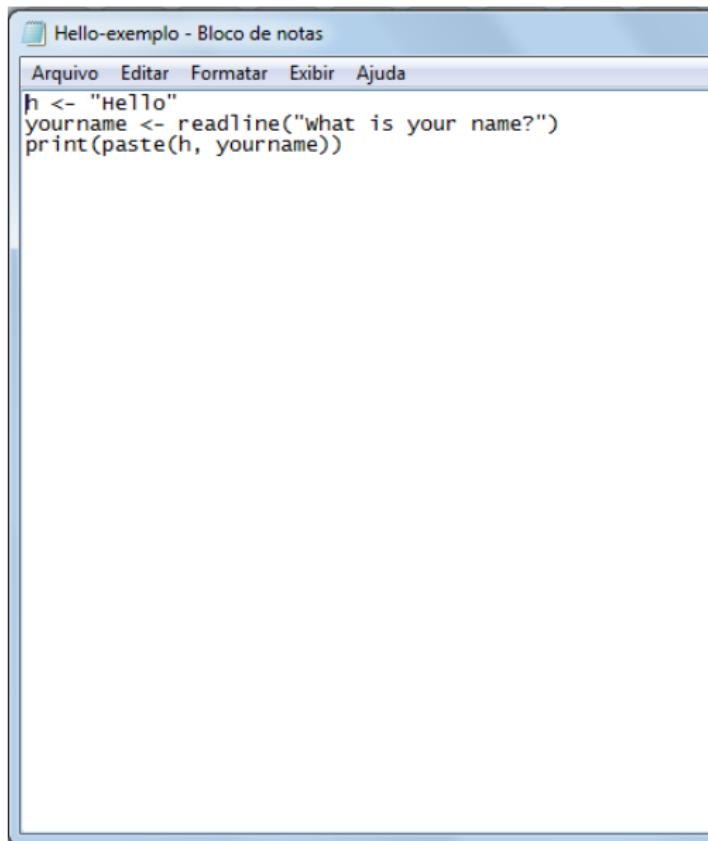
Interface do R



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'**.
 - ▶ **ATENÇÃO:** salve seus scripts em um pen drive; **os computadores do laboratório apagam** todo arquivo salvo no seu diretório quando desligados.
 - ▶ Utilização de editores de texto: **bloco de notas**, **Notepad ++**, **Tinn-R**, etc.
 - ▶ Interfaces de R para usuários: **RStudio**.

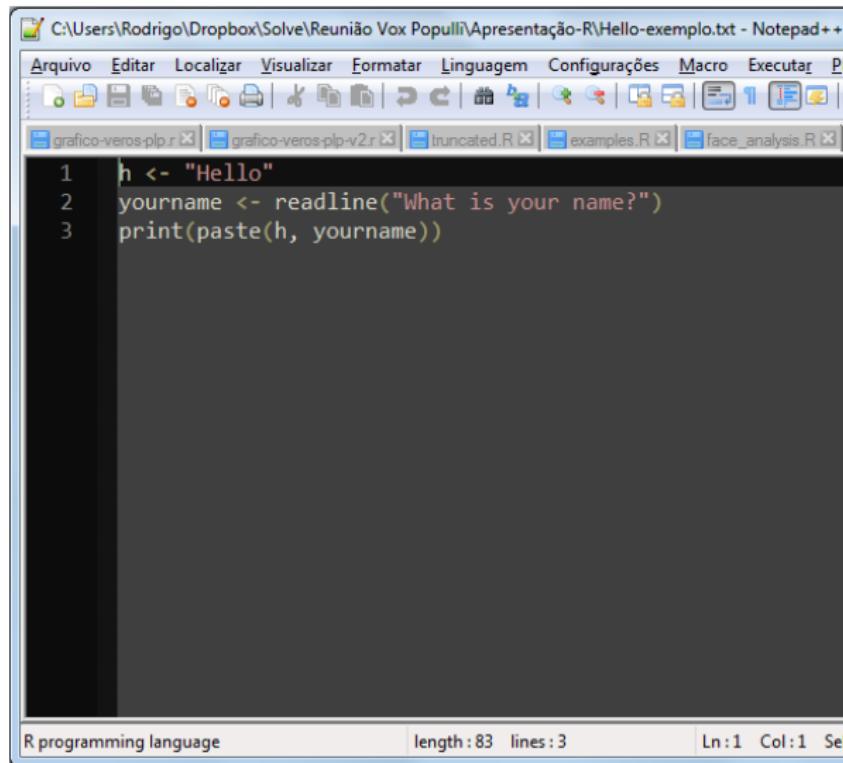
Editores de texto



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

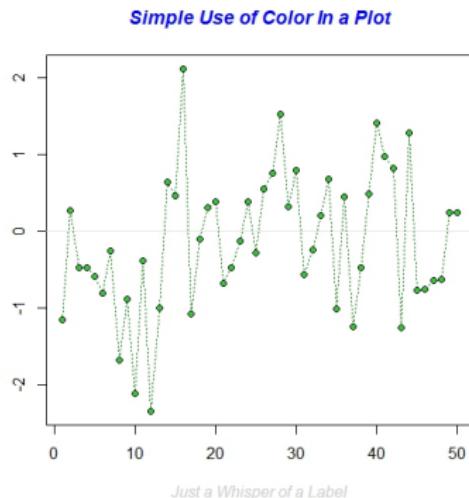
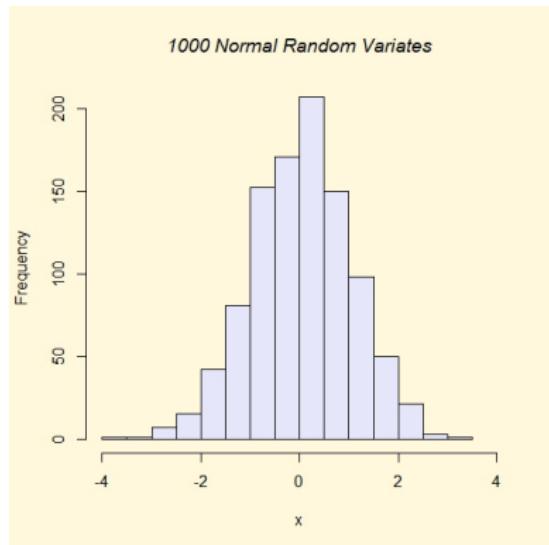
The figure is a scatter plot with a fitted logistic regression curve. The x-axis is labeled 'V1' and ranges from approximately 1.5 to 10.0. The y-axis is labeled 'classn' and ranges from 0.0 to 0.9. The data points are small black dots with some horizontal jitter, and they follow a clear S-shaped trend. A thick blue line represents the logistic regression model, which starts near 0.05 at V1=1.5 and approaches 0.95 as V1 increases towards 10.0. 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

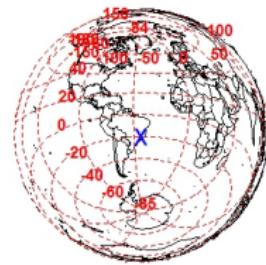
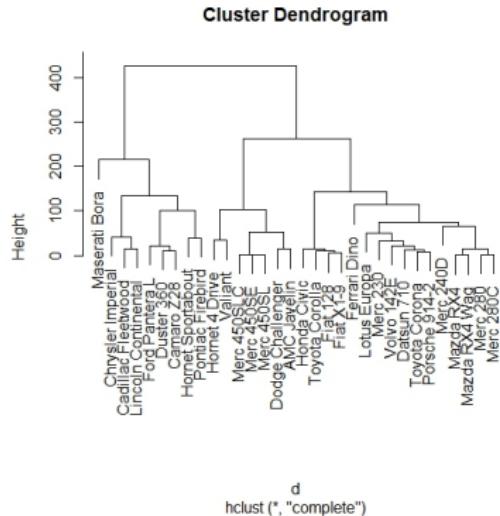
Resumindo: você pode usar o R em todas as etapas de uma análise de dados!

Gráficos do R

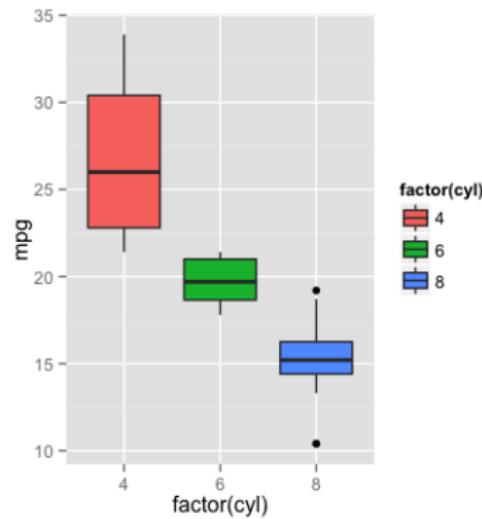
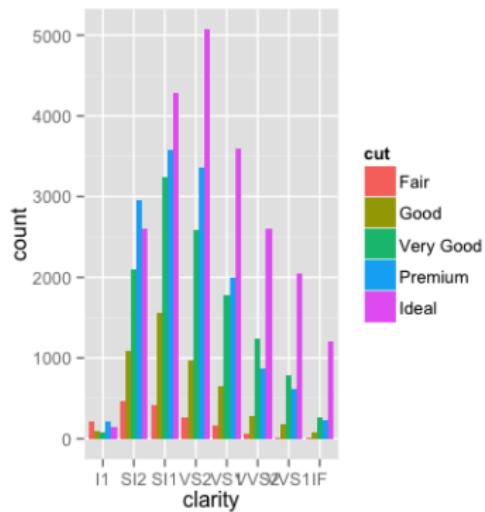


Just a Whisper of a Label

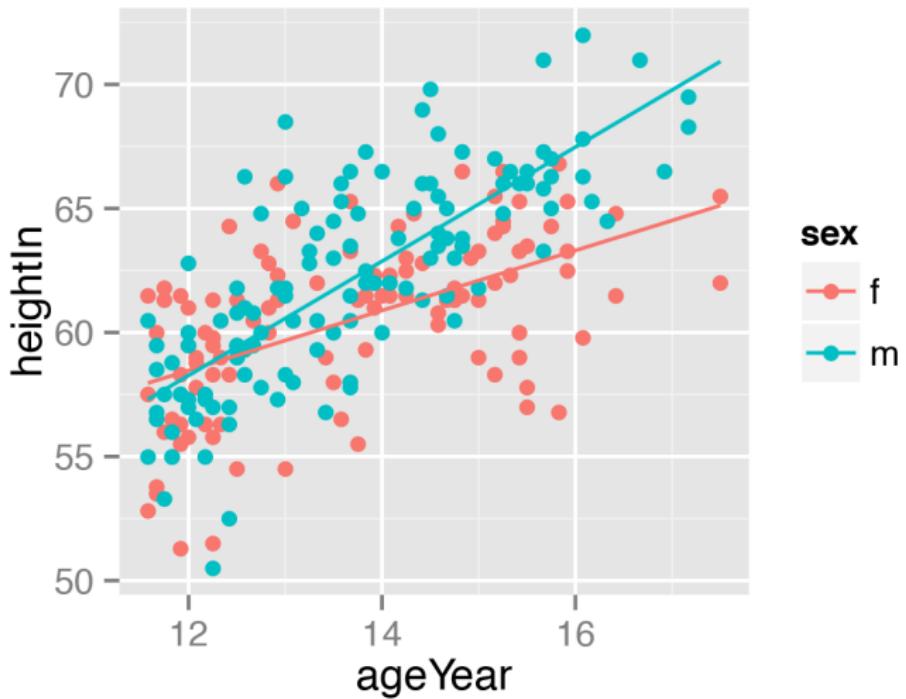
Gráficos do R



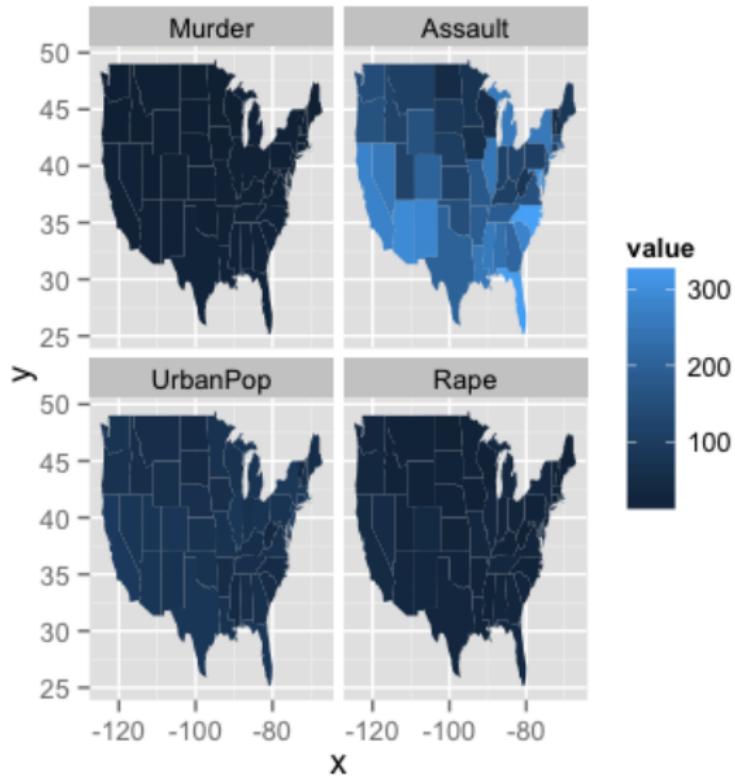
Gráficos do R



Gráficos do R



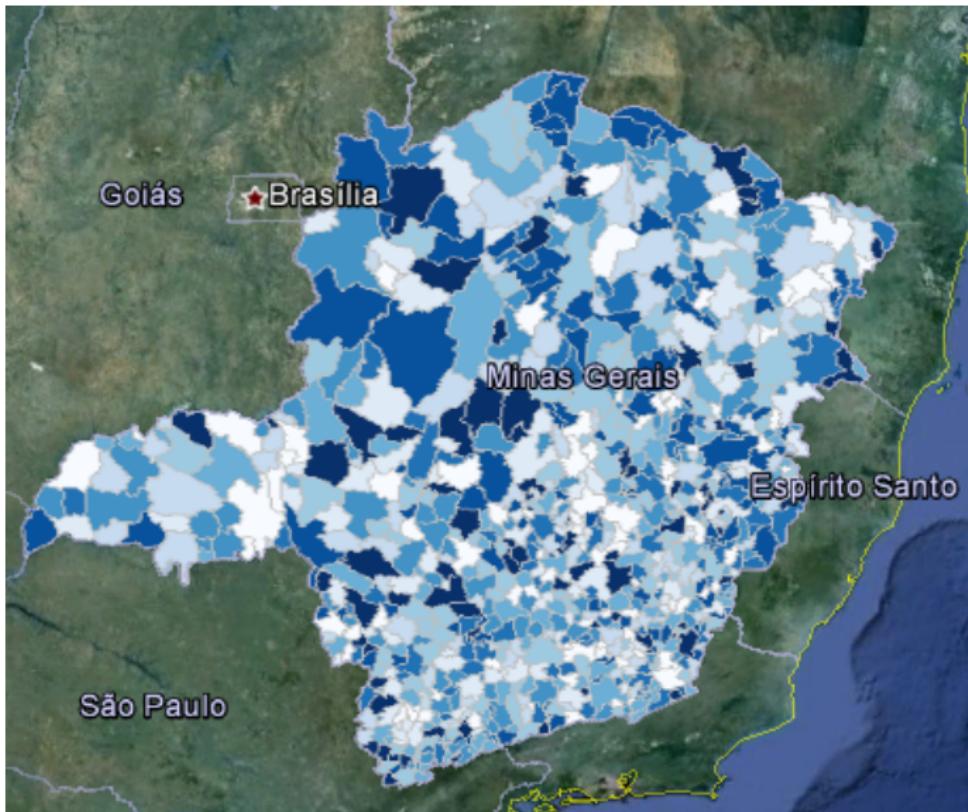
Gráficos do R



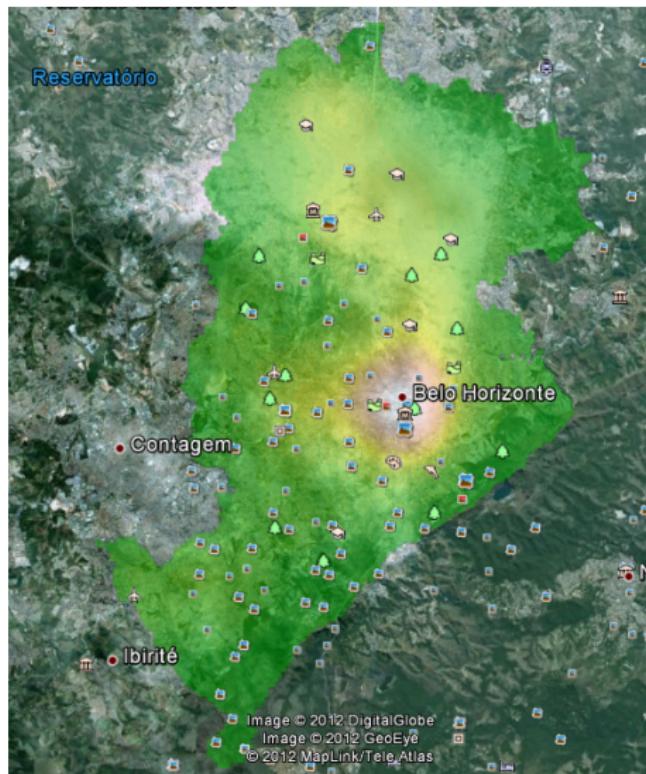
Gráficos do R



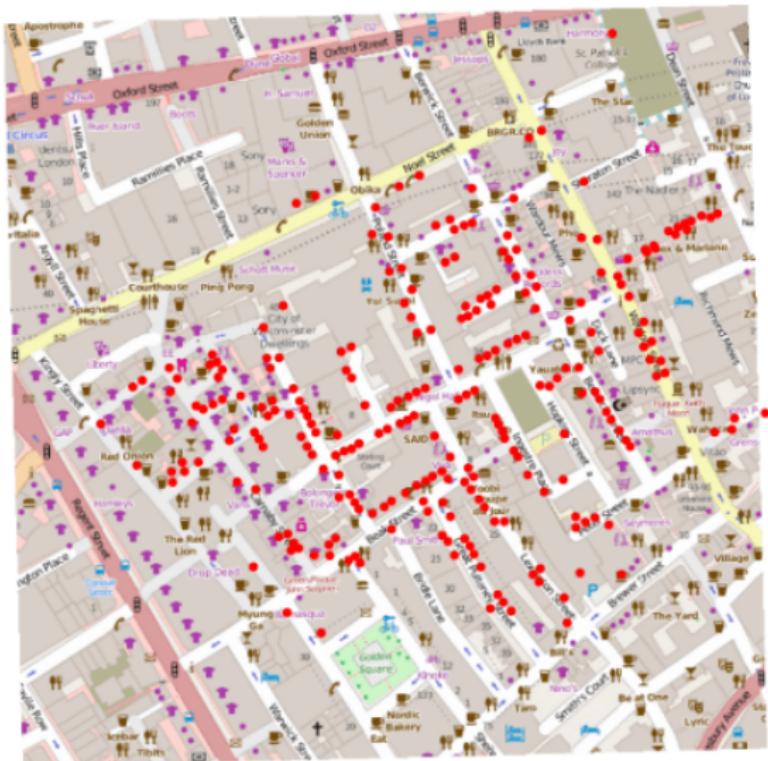
Gráficos do R



Gráficos do R



Gráficos do R



Gráficos interativos do R

- ▶ *Série temporal: ações da Apple*
- ▶ *Motion Chart*

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: Shiny

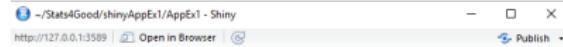
- ▶ Shiny é um pacote do R que torna mais fácil a construção de **aplicações web interativas** (apps) diretamente do R.
 - ▶ Permite a criação e compartilhamento de aplicativos.
 - ▶ Espera **nenhum conhecimento** de tecnologias web como HTML, CSS ou JavaScript (mas você pode aproveitá-las, caso as conheça)
 - ▶ 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.
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 3.3.1 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/>. 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 0.99.903 - 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.
- ▶ As funcionalidades do R, podem ser ampliadas carregando estes pacotes, tornando-o um software muito poderoso, capaz de realizar inúmeras tarefas:
 - ▶ Análise multivariada;
 - ▶ Análise Bayesiana;
 - ▶ Manipulação de dados;
 - ▶ Gráficos a nível de publicação.

Exemplos de pacotes

- ▶ maptools: Funções para leitura, exportação e manipulação de estruturas espaciais.
- ▶ epibasix: Ferramentas elementares para análise de problemas epidemiológicos.
- ▶ survey: Análise de pequisas com planos amostrais complexos.
- ▶ ggplot2: Criação de gráficos elegantes.
- ▶ survival: Análise de dados de sobrevivência .
- ▶ O R possui mais de **10.751** pacotes, e milhares de funções.
- ▶ **Resumindo:** para o R não existe o "**SE É POSSÍVEL FAZER**", mas sim, "**"COMO É POSSÍVEL FAZER"**!

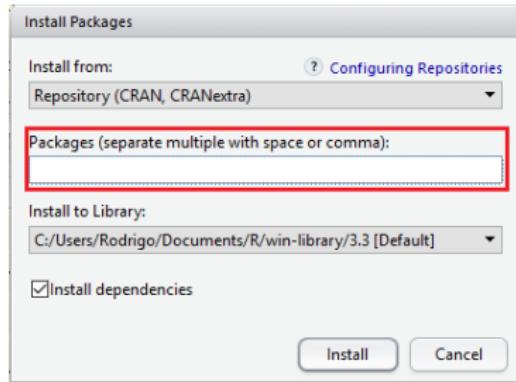
Instalando pacotes

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

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

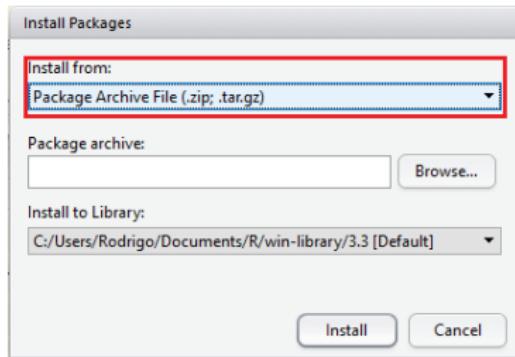
- ▶ Além da opção de comando, também podemos instalar pacotes utilizando o menu **Tools** do RStudio, opção **Install packages ...** e preenchendo com o(s) nome(s) do(s) pacote(s):

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
<http://tryr.codeschool.com/>
- ▶ Para algumas demonstrações da capacidade gráfica do R:

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

Algumas referências

