

Introdução ao *R*: utilizando o *RStudio*



Pedro Gomes Andrade
pedrogandrade@yahoo.com.br



R

O R é uma das principais ferramentas da atualidade para análise de dados.

- Gama de possibilidades de aplicação;
- Software gratuito;
- Colaborativo; e
- Grande comunidade.



R

Linguagem de programação orientada a objetos: Se baseia em variáveis (ou dados, funções, etc.), armazenados através de objetos em sua memória.

- É recomendado que não se utilize de acentos, caracteres especiais e cedilhas, para dar nome aos objetos.
- No caso de bases de dados procure utilizar letras minúsculas para nomear bases de dados.



R

Se estivermos interessados em associar o valor 5.5 ao objeto `x` (variável) podemos utilizar o comando `<-` que significa “atribua a”, ou o sinal `=`. Textos devem ser inseridos entre aspas.

Para realizar comentários basta utilizar `#`; para help de funções conhecidas: `?`, para funções desconhecidas: `help.search("...")`.

```
x <- 5.5      #atribui ao objeto x o valor 5.5
```

```
x = 5.5
```

```
y <- "Bom dia"
```

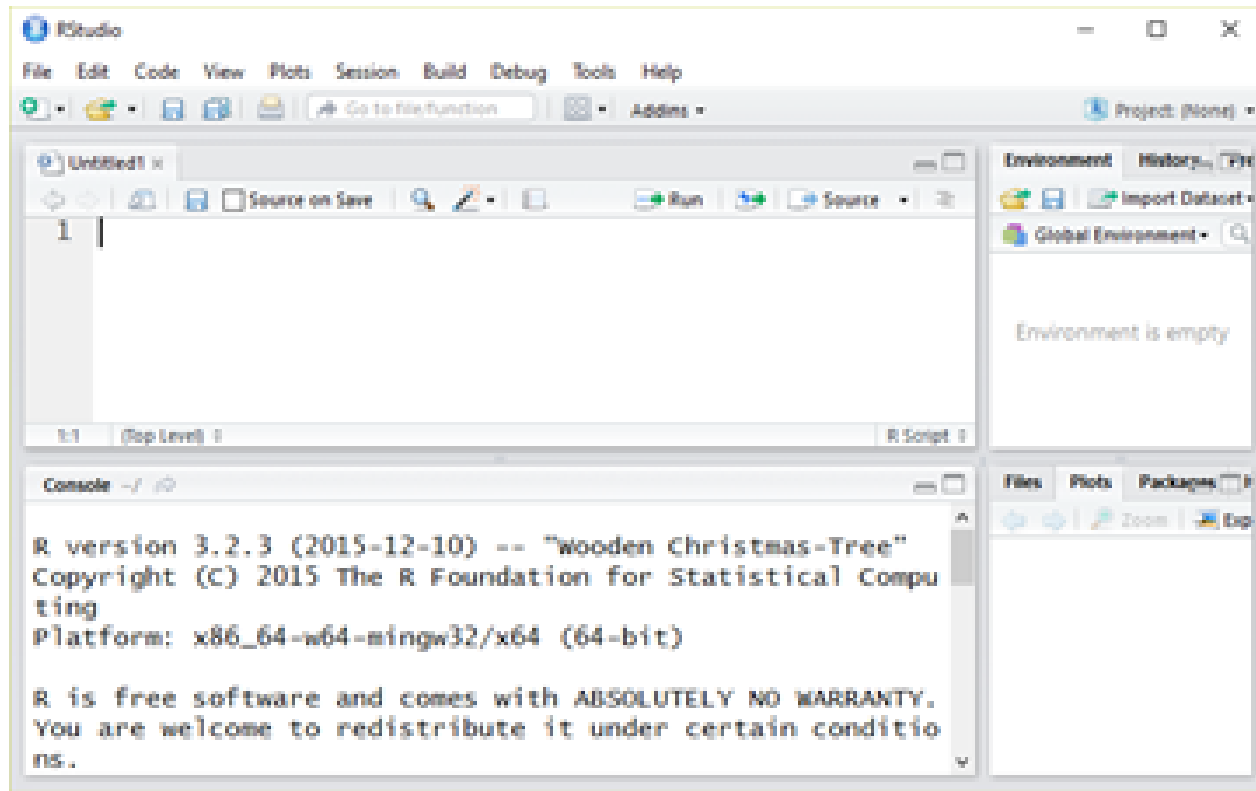
```
?sum
```



Cran: *Comprehensive R Archive Network*



RStudio (editor de *scripts*)





RStudio (editor de *scripts*)

O sinal de **>** indica o *R* está pronto para receber comandos.

O sinal **+** indica que o comando está incompleto.

Para mandar que o *R* compile o programa é necessário basta clicar em **run** ou *ctrl/+r*.

Após o comando realizado é possível cancelá-lo com **esc** ou clicando em **stop**.

Exemplo prático



Atlas do Desenvolvimento
Humano no Brasil 2013



Atlas do Desenvolvimento
Humano no Brasil 2013



| GEOCODIGO | ANO | REGIAO | UF | UFN | Codmun6 | Município | ESPVIDA | FECTOT | MORT1 | (...) |
|-----------|-------|--------|-------|----------|---------|-----------------------|---------|--------|-------|-------|
| 1100015 | 2010 | 1 | 11 | Rondônia | 110001 | ALTA FLORESTA D'OESTE | 70.75 | 2.24 | 23.8 | (...) |
| 1100023 | 2010 | 1 | 11 | Rondônia | 110002 | ARIQUEMES | 73.36 | 1.91 | 19.2 | (...) |
| 1100031 | 2010 | 1 | 11 | Rondônia | 110003 | CABIXI | 70.39 | 2.15 | 24.8 | (...) |
| 1100049 | 2010 | 1 | 11 | Rondônia | 110004 | CACOAL | 74.27 | 1.9 | 14.3 | (...) |
| 1100056 | 2010 | 1 | 11 | Rondônia | 110005 | CEREJEIRAS | 72.94 | 2.12 | 18.1 | (...) |
| (...) | (...) | (...) | (...) | (...) | (...) | (...) | (...) | (...) | (...) | (...) |



| SIGLA | DEFINIÇÃO |
|------------|--|
| GEOCODIGO | Código utilizado pelo IBGE para identificação do município (com dígito verificador). |
| REGIAO | Indica a Grande Região Geográfica: 1 Norte; 2 Nordeste; 3 Sudeste; 4 Sul 5 Centro Oeste |
| UF | Código utilizado pelo IBGE para identificação do estado. |
| UFN | Nome da Unidade da Federação. |
| ESPVIDA | Número médio de anos que as pessoas deverão viver a partir do nascimento, se permanecerem constantes ao longo da vida o nível e o padrão de mortalidade por idade prevalentes no ano do Censo. |
| T_ANALF15M | Taxa de analfabetismo - 15 anos ou mais |
| (...) | (...) |



Vamos praticar?

1º passo: modificar o diretório

2º passo: leitura de dados

3º passo: manipulação da base de dados

4º passo: análise descritiva



Aula Introdução
ao R



Exercício 1

Agora estamos interessados em analisar apenas o Sudeste e apenas as variáveis:

GEOCODIGO, REGIAO, UF, UFN, ESPVIDA, T_ANALF15M, RDPC e Município

- a) Crie uma base de dados reduzida, apenas com o Sudeste.
- b) Quantos municípios pertencem a região Sudeste?
- c) Quais municípios possuem expectativa de vida acima de 78 anos?
- d) Quantos municípios estão abaixo da média da taxa de analf. da região?
- e) Crie uma variável indicadora para os municípios que estão abaixo da média da renda domiciliar per capita da região Sudeste e calcule a média de expectativa de vida para estes municípios (FAREMOS JUNTOS).



Exercício 2

Vamos retornar a base de dados original, com todas as grandes regiões. Selecione alguma variável quantitativa do seu interesse, monte uma nova base de dados com esta variável e adicione:

GEOCODIGO, REGIAO, UF, UFN

a) Calcule o esquema de 5 números por UF da variável de interesse.

dica: min, quantile, max

b) Exporte os resultados para o Excel.

c) Faça o histograma e boxplot.

dica: hist e boxplot