

CENTRO UNIVERSITÁRIO IBMEC

CAMPUS BARRA

**MBA EM IA, DATA SCIENCE E BIG
DATA PARA NEGÓCIOS**

ESTATÍSTICA COM R

AULA 1



Sumário

1. R e RStudio	3
2. História do R	3
3. CRAN.....	3
3.1. Funções do CRAN	3
3.2. Como Instalar Pacotes do CRAN:	4
3.3. Verificação e Manutenção de Pacotes:	4
3.4. Espelhos do CRAN:	4
3.5. Importância do CRAN na Comunidade:	4
4. Download e Atualização do Programa.....	4
5. Layout do R e RStudio	5
6. Funções e Operadores	5
7. Pacotes.....	6
8. Scripts	6
9. Datasets do R.....	6
10. Citação do R e Seus Pacotes.....	7
11. O R Como Calculadora	7
12. Menus de Ajuda.....	7
13. Criação de Objetos no R.....	7
14. Remoção e Listagem de Objetos no R	8
15. Importação de Dados.....	8
16. Manipulação de Vetores, Data Frames e Matrizes	8
17. Gerando Sequências, Números Aleatórios e Repetições.....	9

1. R e RStudio

O **R** é uma linguagem de programação amplamente utilizada para estatística, ciência de dados e machine learning. Sua popularidade se dá por ser uma linguagem open-source e por possuir uma comunidade ativa, sempre contribuindo com novos pacotes e atualizações.

O **RStudio** é um ambiente de desenvolvimento integrado (IDE) que facilita a escrita e execução de códigos em R. Ele oferece ferramentas visuais que tornam o processo de análise de dados mais intuitivo.

Entre as principais características do R e RStudio estão:

- **Interface gráfica do RStudio:** Ajuda a organizar scripts, gráficos e objetos.
- **Versatilidade do R:** Capaz de realizar desde cálculos simples até complexas modelagens estatísticas.
- **Extensão do R através de pacotes:** Bibliotecas criadas para ampliar as funções nativas.

2. História do R

O R foi criado por **Ross Ihaka** e **Robert Gentleman** no início dos anos 90, como uma implementação da linguagem **S**, desenvolvida nos laboratórios da AT&T. Com o tempo, o R cresceu rapidamente em popularidade na comunidade acadêmica e se expandiu para a indústria devido à sua capacidade de lidar com grandes volumes de dados de forma eficiente e com excelente suporte para visualizações gráficas.



Figura 1: Robert Gentleman e Ross Ihaka.

3. CRAN

O **CRAN** (Comprehensive R Archive Network) é o principal repositório de pacotes e softwares relacionados ao R. Ele foi criado para disponibilizar uma ampla gama de recursos que ampliam as funcionalidades da linguagem R, permitindo que a comunidade de usuários contribua e acesse pacotes desenvolvidos por outros.

3.1. Funções do CRAN

1. **Distribuição do R:** O CRAN é o local oficial para baixar o R e suas atualizações para diferentes sistemas operacionais (Windows, Mac, Linux).
2. **Repositório de Pacotes:** O CRAN hospeda milhares de pacotes que ampliam as funcionalidades do R. Pacotes são coleções de funções, dados e documentação criados por usuários da comunidade. Eles são submetidos ao CRAN, onde passam por verificações para garantir compatibilidade, estabilidade e conformidade com os padrões da plataforma.
3. **Documentação e Manuais:** No CRAN, você também encontra manuais e documentação completa para o uso do R, incluindo guias de instalação, tutoriais e referências sobre os pacotes.

4. **Distribuição Global:** O CRAN é replicado em diversos servidores ao redor do mundo, chamados de **espelhos (mirrors)**. Isso garante que os usuários em diferentes regiões possam acessar o R e os pacotes com rapidez e eficiência.

3.2. Como Instalar Pacotes do CRAN:

Para instalar um pacote do CRAN, utilizamos a função `install.packages()`, que baixa e instala automaticamente o pacote diretamente do repositório. Por exemplo:

```
install.packages("ggplot2")
```

Isso baixa o pacote **ggplot2** diretamente do CRAN e o torna disponível para uso local.

3.3. Verificação e Manutenção de Pacotes:

Todos os pacotes submetidos ao CRAN passam por um rigoroso processo de verificação. Isso garante que os pacotes sejam compatíveis com a versão mais recente do R e que funcionem corretamente em diferentes sistemas operacionais. O CRAN também solicita que os mantenedores dos pacotes atualizem e corrijam erros regularmente, garantindo um repositório confiável e atualizado.

3.4. Espelhos do CRAN:

Ao redor do mundo, existem diferentes espelhos do CRAN para facilitar o acesso de diferentes regiões geográficas. Durante a instalação do R ou de pacotes, o R automaticamente seleciona o melhor espelho disponível, mas você também pode escolher manualmente um espelho próximo à sua localização:

```
chooseCRANmirror()
```

3.5. Importância do CRAN na Comunidade:

O CRAN desempenha um papel central no crescimento da comunidade R, pois oferece uma plataforma que permite a colaboração global. Pesquisadores, desenvolvedores e analistas de dados podem contribuir com pacotes que são imediatamente acessíveis a milhões de usuários em todo o mundo, facilitando a partilha de novas técnicas e ferramentas.

Em resumo, o CRAN é mais do que apenas um repositório de pacotes; é um ecossistema robusto que suporta a evolução contínua do R, garantindo que a comunidade tenha acesso a ferramentas de última geração para ciência de dados, estatísticas e muito mais.

4. Download e Atualização do Programa

1. **Baixar o R:**
 - Acesse o site oficial: <https://cran.r-project.org>
 - Selecione o sistema operacional correto (Windows, Mac ou Linux) e siga as instruções de instalação.
2. **Baixar o RStudio:**
 - Acesse: <https://rstudio.com>
 - Baixe a versão gratuita e instale após o R estar configurado.
3. **Atualização:**
 - Para verificar se há pacotes desatualizados e atualizá-los, use:

```
update.packages()
```

5. Layout do R e RStudio

O RStudio é dividido em **quatro painéis principais**:

1. **Editor de Script (superior esquerdo):**
 - Escreva e edite códigos em R. Salve scripts para reutilização posterior.
2. **Console (inferior esquerdo):**
 - Área onde os comandos são executados diretamente.
3. **Environment/History (superior direito):**
 - **Environment:** Exibe todos os objetos (vetores, data frames, funções) criados durante a sessão.
 - **History:** Exibe o histórico de comandos executados.
4. **Files/Plots/Packages/Help (inferior direito):**
 - **Files:** Navegue pelos arquivos do seu computador.
 - **Plots:** Exibe gráficos gerados.
 - **Packages:** Gerencia os pacotes instalados.
 - **Help:** Acessa a documentação e ajuda do R.

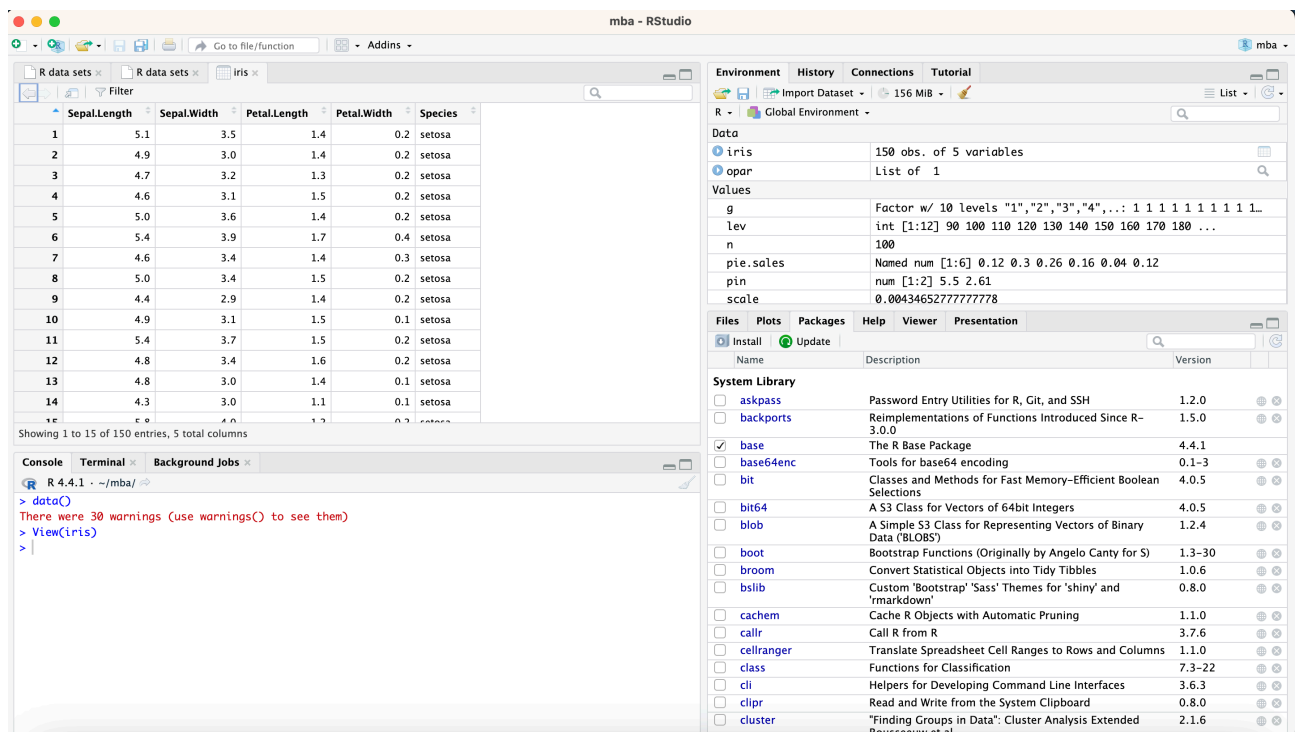


Figura 2: Perspectiva do RStudio.

6. Funções e Operadores

Operadores Matemáticos:

- **Atribuição de valor:** `<-` ou `=`
- **Operações aritméticas:** `+`, `-`, `*`, `/`, `^`

Operadores Lógicos:

- **Igualdade:** `==`
- **Diferença:** `!=`
- **Maior/Menor que:** `>`, `<`

- **Maior/Menor ou igual:** `>=`, `<=`
- **E lógico:** `&`
- **Ou lógico:** `|`

Exemplo:

```
x <- 10 # Atribuição de valor
y <- x + 5 # Soma
```

7. Pacotes

Pacotes são coleções de funções e dados que ampliam as funcionalidades básicas do R. Existem pacotes específicos para diferentes tipos de análises, como visualização de dados, machine learning e manipulação de dados.

- **Instalar pacotes:**

```
install.packages("nome_do_pacote")
```

- **Carregar pacotes:**

```
library(nome_do_pacote)
```

- **Ver pacotes instalados:**

```
installed.packages()
```

8. Scripts

Scripts são documentos onde podemos escrever múltiplas linhas de código e salvá-las para uso posterior. No RStudio, os scripts podem ser criados e executados de maneira simples.

- **Criando um novo script:**
 - `Ctrl + Shift + N`
- **Executando comandos de um script:**
 - Selecione uma linha e pressione `Ctrl + Enter`.

Exemplo de script simples:

```
x <- 10
y <- 5
z <- x + y
print(z) # Mostra o resultado de z
```

9. Datasets do R

O R contém vários datasets incorporados que podem ser usados para experimentos e aprendizado. O dataset mais conhecido é o **iris**, que contém dados sobre diferentes espécies de plantas.

- **Carregar e visualizar datasets:**

```
data(iris)
View(iris)
```

- **Outros datasets:**

```
data(mtcars)
data(airquality)
```

10. Citação do R e Seus Pacotes

É importante citar o R e os pacotes utilizados em suas análises. O R possui uma função que gera automaticamente a citação correta.

- **Citando o R:**

```
citation()
```

- **Citando um pacote específico:**

```
citation("ggplot2")
```

11. O R Como Calculadora

Além de ser uma poderosa ferramenta de análise de dados, o R pode ser usado como uma calculadora.

- **Operações básicas:**

```
3 + 5 # Soma
10 - 7 # Subtração
4 * 6 # Multiplicação
20 / 4 # Divisão
2^3 # Exponenciação
```

- **Funções matemáticas avançadas:**

```
sqrt(16) # Raiz quadrada
log(10) # Logaritmo natural
```

12. Menus de Ajuda

O R possui uma ampla documentação integrada que pode ser acessada facilmente:

- **Ajuda sobre uma função:**

```
?mean
```

- **Busca por um termo específico:**

```
help.search("anova")
```

13. Criação de Objetos no R

- **Criando um objeto:**

```
x <- 10
y <- c(1, 2, 3, 4, 5) # Criando um vetor
```

- **Acessando elementos de um vetor:**

```
y[2] # Acessa
```

14. Remoção e Listagem de Objetos no R

- Para listar os objetos criados no ambiente:

```
ls()
```

- Para remover um objeto:

```
rm(x)
```

15. Importação de Dados

Para importar arquivos de diferentes formatos:

- **CSV:**

```
dados <- read.csv("caminho_do_arquivo.csv")
```

- **Excel:** (necessário instalar o pacote `readxl`)

```
library(readxl)
dados <- read_excel("caminho_do_arquivo.xlsx")
```

16. Manipulação de Vetores, Data Frames e Matrizes

Vetores:

- Criando um vetor:

```
v <- c(1, 2, 3, 4, 5)
caracteres <- c("a", "b", "c", "d")
```

- Acessando elementos:

```
v[2] # Segundo elemento
```

Data Frames:

- Criando um data frame:

```
df <- data.frame(nome = c("Ana", "João"), idade = c(28, 35))
```

- Acessando colunas:

```
df$nome
```

Matrizes:

- Criando uma matriz:

```
m <- matrix(1:9, nrow = 3, ncol = 3)
```


17. Gerando Sequências, Números Aleatórios e Repetições

- **Sequências:**

```
seq(1, 10, by = 2) # Sequência de 1 a 10 com incremento de 2
```

- **Números aleatórios:**

```
runif(5, min = 0, max = 10) # 5 números aleatórios entre 0 e 10
```

- **Repetições:**

```
rep(2, times = 5) # Repete o número 2, 5 vezes
```

Referências:

- Manual oficial do R: <https://cran.r-project.org/manuals.html>
- Documentação do RStudio: <https://support.rstudio.com/hc/en-us>