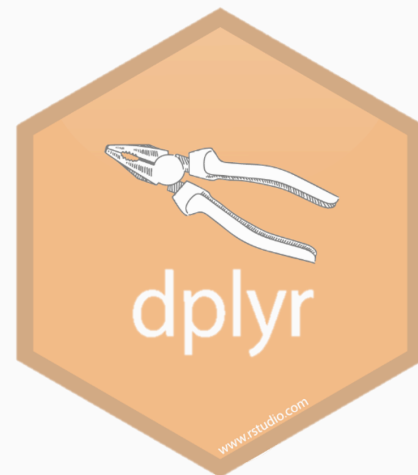


INTRODUÇÃO A PROJETOS NO R

Manipulação de dados com dplyr

Marie-Christine Rufener
macrufener@gmail.com



Conteúdo programático

- 1. Estrutura de dados
- 2. Operadores
- 3. Manipulação de dados
 - Funções base
 - Dplyr
- 4. Prática no R

Estrutura de dados

Estrutura de dados

Relembrando

O R opera em estrutura de *dados nomeados*

- Também conhecido por **objetos** ou **variáveis**
- Armazenamento & manipulação dos dados (ou resultados de análises)

```
# Operador de atribuição (à esquerda)
```

```
obj ← c(1, 2, 3)
```

```
# Operador de atribuição (à direita)
```

```
c(1, 2, 3) → obj2
```

```
# Operador de atribuição (igualdade)
```

```
obj3 = c(1, 2, 3)
```

Estrutura de dados

Classificação dos objetos

Os dados podem ser classificados conforme:

- Tipo de informação
 - Variável numérica, discreta, categórica, ...
- Dimensão da informação
 - 1D, 2D, 3D,

ATENÇÃO:

A manipulação do dados depende tanto do tipo de informação que está sendo tratado, como da sua dimensão

Estrutura de dados

Classificação dos objetos | Tipos

Os dados (i.e., variáveis) podem ser classificados de acordo com a natureza da informação sendo armazenada no objeto

Variável	Tipo	Exemplo
Integer	Formada por números inteiros	1, 300, -60
Numeric	Formada por números com casas decimais	3.2, 18.5, -0.9
Character	Formada por texto sem categorias e estrutura implícita	Gato, Dinamarca, CNPq
Factor	Formada por textos com categorias bem definidas	Manhã, Tarde, Noite
Logical	Formado por valores booleanos	TRUE, FALSE
Date	Formadas por valores temporais	2025, 2025-05-26, 2025-05-26 16:35
Missing	Não há nenhuma informação atribuída à variável	NA (Not Available)

Estrutura de dados

Classificação dos objetos | Tipos

Pode-se usar a função `class()` para avaliar a natureza do dado:

```
var1 ← c("olá")  
class(var1)
```

```
## [1] "character"
```

```
var2 ← pi  
class(var2)
```

```
## [1] "numeric"
```

```
var3 ← TRUE  
class(var3)
```

```
## [1] "logical"
```

Estrutura de dados

Classificação dos objetos | Tipos

Alternativamente, pode-se avaliar se o dado é de uma determinada natureza usando a sintaxe `is.NATUREZA_DO_OBJETO()`:

```
## Var1 é de natureza fator?  
is.factor(var1)
```

```
## [1] FALSE
```

```
## Var2 é de natureza numérica?  
is.numeric(var2)
```

```
## [1] TRUE
```

```
## Var3 é de natureza lógica?  
is.numeric(var3)
```

```
## [1] FALSE
```


Estrutura de dados

Classificação dos objetos | Tipos

De modo similar, é possível alterar a natureza do dado usando funções base conforme:

```
## Transformando caractere para fator  
periodo ← c("Manhã", "Tarde", "Noite")  
class(periodo)
```

```
## [1] "character"
```

```
periodo ← as.factor(periodo)  
class(periodo)
```

```
## [1] "factor"
```

Exemplos

```
as.numeric() as.integer()  
as.character() as.factor()  
as.logical()
```

Estrutura de dados

Classificação dos objetos | Dimensão

O dado também pode ser classificado de acordo com sua dimensão

- Quando contém apenas 1 elemento (1D):

Estrutura de dados

Classificação dos objetos | Dimensão

O dado também pode ser classificado de acordo com sua dimensão

- Quando contém apenas 1 elemento (1D):

Variável atômica (scalar)

Scalar



```
# Isso tudo é um scalar com  
# apenas 1 elemento
```

```
s1 ← 1 #integer
```

```
s2 ← "Oi" #character
```

```
s3 ← TRUE #logical
```

Estrutura de dados

Classificação dos objetos | Dimensão

O dado também pode ser classificado de acordo com sua dimensão

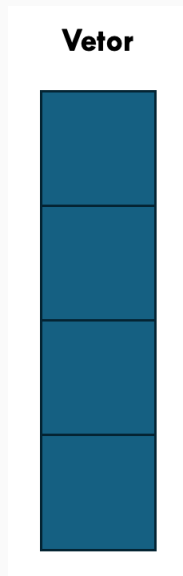
- Quando contém +1 elemento:
 - Classificação passa a depender da dimensão e tipo de dado
 - Vetor
 - Matriz
 - Array
 - Data frame
 - Lista

Estrutura de dados

Classificação dos objetos | Dimensão

Vetor (1D)

Conjunto de elementos do mesmo tipo (numérico, lógico, caractere etc.)



```
# Vetor numérico  
v1 ← c(1, 35, 110)  
  
## Vetor caractere  
v2 ← c("A", "B", "C")  
  
## Vetor lógico  
v3 ← c(TRUE, TRUE, FALSE)
```

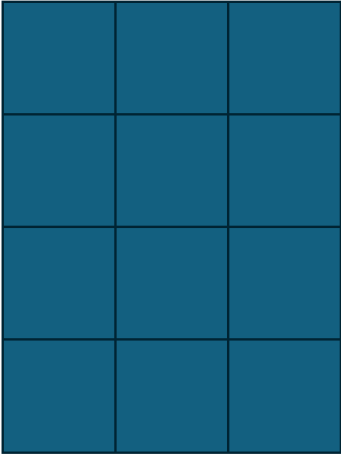
Estrutura de dados

Classificação dos objetos | Dimensão

Matriz (2D)

Representa dados no formato de tabela com linhas e colunas

Matriz




```
m <- matrix(data = letters[1:6],  
             nrow = 2,  
             ncol = 3)
```

m

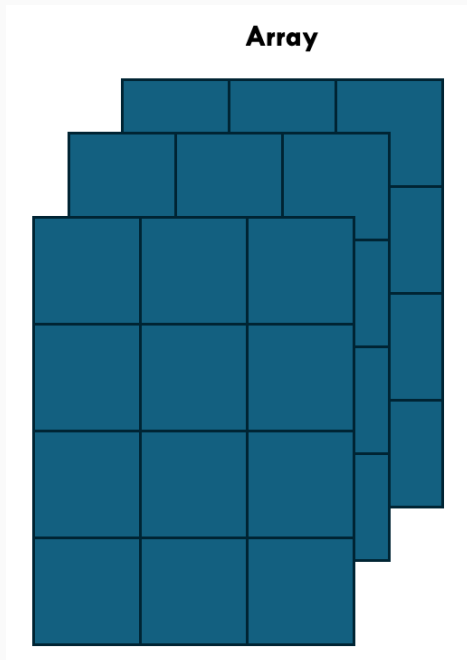
```
##      [,1] [,2] [,3]  
## [1,] "a"  "c"  "e"  
## [2,] "b"  "d"  "f"
```

Estrutura de dados

Classificação dos objetos | Dimensão

Array (+2D)

Extensão da matriz para +2D



```
# Array de 2 matrizes 2x2  
a ← array(1:8, dim = c(2, 2, 2))  
a
```

```
## , , 1  
##  
##      [,1] [,2]  
## [1,]    1    3  
## [2,]    2    4  
##  
## , , 2  
##  
##      [,1] [,2]  
## [1,]    5    7  
## [2,]    6    8
```

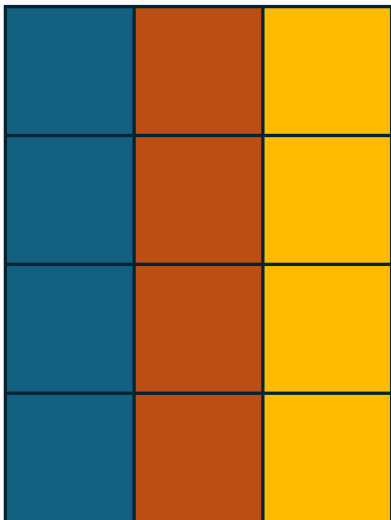
Estrutura de dados

Classificação dos objetos | Dimensão

Data frame (2D)

Similar a uma matriz, porém as colunas podem pertencer à variáveis de naturezas distintas

Data frame




```
df <- data.frame(Nome = c("Ana", "Beto")  
                  Idade = c(25, 16),  
                  Maioridade = c(TRUE, FA
```

```
df
```

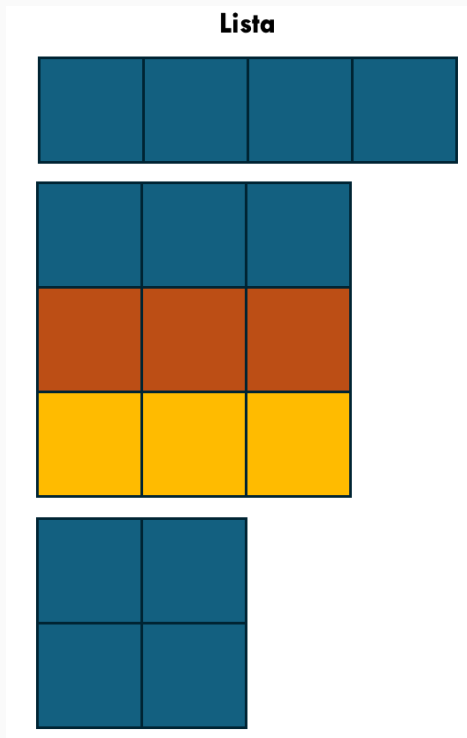
```
##   Nome  Idade Maioridade  
## 1  Ana    25         TRUE  
## 2 Beto    16        FALSE
```


Estrutura de dados

Classificação dos objetos | Dimensão

Lista (+2D)

Coleção de dados de qualquer tipo de informação e dimensão (inclusive outras listas)



```
l <- list(rep(1, 10), #vetor
          factor(c("vermelho", "azul",
                    matrix(c(1, 2), c(1, 2))) #matriz))

l

## [[1]]
## [1] 1 1 1 1 1 1 1 1 1 1
##
## [[2]]
## [1] vermelho azul verde
## Levels: azul verde vermelho
##
## [[3]]
##      [,1] [,2]
## [1,]    1    2
```

Operadores

Operadores

São símbolos (ou combinações de símbolos) que realizam operações específicas sobre variáveis

- usados para fazer *cálculos*, *comparações*, e diversos tipos de *manipulações*
- Podem ser agrupados em **4 tipos** básicos:
 - Aritméticos
 - Relacional
 - Lógico
 - Especiais

Operadores

Aritmético

Usados para fazer operações matemáticas

Operador	Descrição
+	Adição
-	Subtração
*	Multiplicação
/	Divisão
\wedge ou **	Potência
%%	Resto da divisão
%/%	Divisão inteira

Operadores

Relacional

Usados para comparar valores, retornando TRUE ou FALSE

Operador	Descrição
>	Maior
<	Menor
>=	Maior ou igual
<=	Menor ou igual
==	Igualdade
!=	Diferença

Operadores

Lógico

Usados para combinar ou inverter valores lógicos (TRUE/FALSE)

Operador	Descrição
!	Negação (Não)
&	E
	OU

Operadores

Especiais

Operador	Descrição
<code>%in%</code>	Testa se um elemento pertence a um vetor
<code>:</code>	Sequência de números inteiros
<code>%*%</code>	Produto matricial

Manipulação dos dados

Manipulação dos dados

Os dados podem ser manipulados por diversos motivos e objetivos:

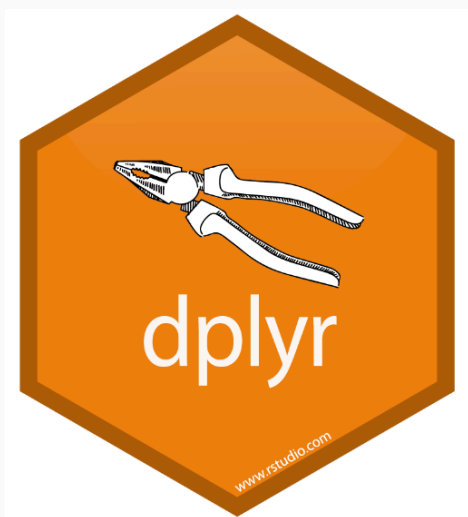
- Correção de erros (ortográficos, numéricos, ...)
- Organização
- Redução (filtragem ou agrupamento)
- Padronização para realizar certas análises
- Análise exploratória

Há várias formas de manipular os dados:

- Funções base
- Funções do pacote `dplyr`

Manipulação dos dados

O pacote dplyr



"É uma gramática de manipulação de dados, fornecendo um conjunto consistente de verbos que ajudam a resolver os desafios mais comuns de manipulação de dados"

Manipulação dos dados

O pacote dplyr | Funções

As principais funções usadas para manipular os dados são:

- `filter()`: filtra linha(s)
 - `select()`: seleciona coluna(s)
 - `arrange()`: ordena os dados
 - `mutate()`: cria e/ou modifica colunas
 - `group_by()`: agrupa os dados
 - `summarise()`: sumariza os dados
-

Funções especiais:

- `rename()`: renomeia as colunas
- `distinct()`: remove linhas repetidas
- `count()` / `n()`: contabiliza o número de linhas dentro de cada grupo
- `%>%`: pipe - executa as funções sequencialmente

Prática no R