

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Organiza  o dos dados em R

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Revisão

Sinal de atribuição	<code><-</code>
Vetor numérico	<code>c(1,2,3,4)</code>
Vetor lógico	<code>c(TRUE, FALSE, NA)</code>
Vetor de caracteres	<code>c("My", "name", "is")</code>
Ajuda	<code>?c</code>
Operações aritméticas	<code>c(1.1, 9, 3.14)*2+100 # 102.20 118.00 106.28</code>
Reciclagem	<code>c(1,2,3,4) + c(0,10) # 1 12 3 14</code>

Revisão

Sequência de números	<code>1:20; 15:1; seq(1,20, by=0.5); seq(5,10,length=30)</code>
Replicar números	<code>rep(0, times=40); rep(c(0,1,2), times=10)</code>
Tamanho do vetor	<code>length(rep(0, times=40))</code>
Juntar palavras	<code>paste(c("My", "code"), collapse = " ")</code>

If e Else

```
a <- 300

if( a >= 500 ) {
    print("escopo 1")
} else if(a < 100) {
    print("escopo 2")
} else {
    print("escopo 3")
}
```

If e Else

Para lidar com vetores...

```
ifelse(variavel >= 500,  
       'executa essa tarefa se  
TRUE', 'executa outra se  
FALSE')
```

```
ifelse(1:20 >= 10, 0, 10)
```

If e Else

Para lidar com vetores...

```
a <- 839
```

```
c <- ifelse(a >= 10000,  
  'VALOR ALTO', ifelse(a <  
    10000 & a >= 1000, 'VALOR  
    MEDIO', 'VALOR BAIXO'))
```

```
c
```

Functions

```
func1 <- function(par1, par2){  
  # sequência de tarefas  
  return(result)  
}  
  
# chamada da função  
func1(a,b)
```

Matrizes e Data Frame

- Tabela bidimensional;
- Coleção de elementos (numérico, caracteres, lógicos);
- Número determinado de linhas e colunas;

Matriz

- Os elementos são do mesmo tipo.

Data Frame

- Os elementos podem ser de tipos diferentes.

Matrizes e Data Frame

Criando matrizes `m <- matrix(1:20, 4, 5)`

`m <- c(1:20); dim(m) <- c(4,5)`

Criando Data Frame `patients <- c("Bill","Gina","Kelly","Sean")`
`data.frame(patients, m)`

Combinar colunas `cbind(m, m)`

Combinar linhas `rbind(m, m)`

Nomear colunas `colnames(v)`

Nomear linhas `rownames(v)`

Matrizes e Data Frame

```
m <- matrix(1:20, 4, 5)
```

```
m
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	5	9	13	17
[2,]	2	6	10	14	18
[3,]	3	7	11	15	19
[4,]	4	8	12	16	20

```
m[2,3] # elemento da 2ª linha
```

```
# e 3ª coluna
```

```
m[1,] # primeira linha
```

```
m[,1] # primeira coluna
```

Matrizes e Data Frame

```
> m <- matrix(1:20, 4, 5, byrow = TRUE)
```

```
> m
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	2	3	4	5
[2,]	6	7	8	9	10
[3,]	11	12	13	14	15
[4,]	16	17	18	19	20

Matrizes e Data Frame

```
> m <- matrix(1:20, 4, 5)
> colnames(m) <- c("a", "b", "c", "d", "e")
> m
```

	a	b	c	d	e
[1,]	1	2	3	4	5
[2,]	6	7	8	9	10
[3,]	11	12	13	14	15
[4,]	16	17	18	19	20

```
# elemento da 2ª L e 3ª C
m[2,3]
m[2,"c"]

# primeira linha
m[1,]

# primeira coluna
m[,1]
m[, "a"]
```

Abrindo dados em R

Organização dos dados em R

```
read.csv(file, header = TRUE,  
          sep = ",", quote = "\"", dec  
          = ".", fill = TRUE,  
          comment.char = "", ...)
```

Revisão

<code>ls()</code>	Listar as variáveis no ambiente;
<code>class(v)</code>	Verifica a classe de um objeto <code>v</code> ;
<code>dim(v)</code>	Mostra as dimensões de uma tabela <code>v</code> (linha x coluna);
<code>nrow(v)</code>	Mostra o número de linhas em <code>v</code> ;
<code>ncol(v)</code>	Mostra o número de colunas em <code>v</code> ;
<code>object.size(v)</code>	Mostra o espaço ocupado na memória por <code>v</code> ;
<code>head(v)</code>	Mostra as 6 primeiras linhas de <code>v</code> ;
<code>tail(v)</code>	Mostra as 6 últimas linhas de <code>v</code> ;
<code>summary(v)</code>	Mostra um breve sumário de <code>v</code> ;
<code>table(v)</code>	Conta o número de ocorrência de cada categoria em <code>v</code> ;
<code>str(v)</code>	Mostra a estrutura de <code>v</code> ;

iris.data e iris.name

Source: <https://archive.ics.uci.edu/ml/datasets/iris>



Dados de comprimento e largura de pétalas e sépalas de três variedades de íris;

- iris.data → contém os dados;
- Iris.names → contém algumas descrições sobre os dados;

Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	2411032

iris.data e iris.name



Exercício:

1. Veja a documentação da função `read.csv()` e carregue os dados do `iris.data`;
2. Veja a descrição de cada coluna dos dados em `iris.names` e nomeie as colunas;
3. Explore os dados utilizando as funções vistas nesta aula e responda:
 - a. Qual a dimensão da tabela?
 - b. Quais são os tipos de dados de cada coluna?
 - c. Quanto de memória estes dados estão ocupando?
 - d. Quantas amostras estão presentes em cada classe de Íris?
 - e. Tem algum dado faltante na tabela?

Para casa

`library(Swirl)`

Swirl:

- 7: Matrices and Data Frames
 - 8: Logic
 - 9: Functions
 - 10: lapply and sapply
 - 11: vapply and tapply
 - 12: Looking at Data
-