

Introdução ao R

Aula #02

Fernando de Souza

UFMG

2021-01-06

Roteiro

1. Criando Dados no R

- `cbind`
- `rbind`
- `data.frame`
- `tibble`

2. Exportando dados para o R

- `read.csv`
- `read.csv2`
- `read.txt`
- `haven`

3. Trabalhando com dados no R (I)

- seleção de dados
- `subset`
- novos comandos

Criando dados no R

É possível mesclar dois vetores do mesmo tamanho para gerar um banco de dados a partir dos comandos: *cbind*, *rbind* e **data.frame**

```
a <- c("A", "B", "C", "D", "E")
b <- c(65.8, 60, 41.8, 30, 23)
```

```
rbind(a, b)
```

```
##      [,1]      [,2] [,3]      [,4] [,5]
## a  "A"      "B"    "C"      "D"    "E"
## b  "65.8"   "60"   "41.8"   "30"   "23"
```

```
cbind(a, b)
```

```
##      a      b
## [1,] "A"   "65.8"
## [2,] "B"   "60"
## [3,] "C"   "41.8"
## [4,] "D"   "30"
## [5,] "E"   "23"
```

```
data.frame(a, b)
```

```
##      a      b
## 1 A  65.8
## 2 B  60.0
## 3 C  41.8
## 4 D  30.0
## 5 E  23.0
```

Criando dados no R

Para poder usar a tabela criada é preciso enderecar ela a um objeto

```
bd ← data.frame (a, b)  
bd
```

```
##      a      b  
## 1 A 65.8  
## 2 B 60.0  
## 3 C 41.8  
## 4 D 30.0  
## 5 E 23.0
```

Criando dados no R

Outra forma de criar data.frames é a partir do comando "tibble" no pacote "dplyr"

```
library (tidyverse)
```

```
bd2<- tibble (a, b)
```

```
bd2
```

```
## # A tibble: 5 x 2
```

```
##   a         b
```

```
##   <chr> <dbl>
```

```
## 1 A      65.8
```

```
## 2 B      60
```

```
## 3 C     41.8
```

```
## 4 D      30
```

```
## 5 E      23
```

Criando dados no R

Caso seja necessário é possível trocar o nome das colunas e das linhas a partir dos comandos "colnames" e "rownames"

```
colnames(bd)[1] ← "nomes"  
bd
```

```
##      nomes      b  
## 1      A 65.8  
## 2      B 60.0  
## 3      C 41.8  
## 4      D 30.0  
## 5      E 23.0
```

Criando dados no R

Agora que sabemos como criar um banco de dados dentro do R vamos aprender a importar um já existente.

Existem duas formas: por comando ou pelo botão "*Import Dataset*".

`read.csv ()` `read.csv2 ()` `read.txt ()`

Bancos de dados em outros formatos (.xlsx, .spss, .dat) podem ser abertos com ajuda de outros pacotes como o *haven* ou o *readxl*.

```
library (haven)
```

```
library (readxl)
```

```
dados ← read_sav ("https://github.com/neptune97/Metodologia-II/blob/main/Aula%2002/La  
#read_excel ()
```

Exploração de Banco de Dados no R

Seleção de colunas, linhas e elementos

Podemos selecionar colunas a partir do uso de colchetes ao lado do nome do banco de dados. Caso queria mais de uma coluna use o intervalo de colunas desejado e escreva-o dentro dos colchetes usando `:` como separador.

```
dados [1]
```

```
## # A tibble: 1,204 x 1
##       NUMINVES
##       <dbl+lbl>
## 1 2018 [2018]
## 2 2018 [2018]
## 3 2018 [2018]
## 4 2018 [2018]
## 5 2018 [2018]
## 6 2018 [2018]
## 7 2018 [2018]
## 8 2018 [2018]
## 9 2018 [2018]
## 10 2018 [2018]
## # ... with 1,194 more rows
```


Exploração de Banco de Dados no R

Para a seleção de uma linha colocasse dentro dos colchetes o número da linha seguido das colunas que devem entrar na seleção

```
dados [1, 1:2]
```

```
## # A tibble: 1 x 2
##   NUMINVES IDENPA
##   <dbl+lbl> <dbl+lbl>
## 1 2018 [2018] 76 [ Brasil]
```

Para selecionar elementos específicos usamos o comando **c()** após a virgula e dentro dos parenteses utilizamos a localização do elemento

```
dados [20, c(10:2)]
```

```
## # A tibble: 1 x 9
##   CODENC SEXO EDAD COMDIST TAMCIUD CIUDAD REG NUMENTRE
##   <dbl+> <dbl+l> <dbl> <dbl> <dbl+l> <dbl+lbl> <dbl+lbl> <dbl>
## 1 34 1 [Hom~ 20 3 6 [50.~ 7.60e7 [BR:~ 76001 [BR:~ 760020
## # ... with 1 more variable: IDENPA <dbl+lbl>
```

Exploração de Banco de Dados no R

Outra forma de solucionar uma coluna completa é a partir do comando banco de dados seguido de cifrao

```
dados$S5
```

```
## <labelled<double>[1204]>: S5 Religión del entrevistado
##      [1]  1  1  1  1  1 96  1  2  2  1  1  1  1  1  1  1  2  1  1  5  1  5  1
##     [25]  1  2 97  1  1  1 12  6  2  1  1  1 97 97  2  1 97 97  2  1  2  1  1  1
##     [49]  2  1  2  1 97  2  1  1  1  1  7  1  2  1 97  1  1  1  2 97  1  5  1  2
##     [73]  1 97  2  2  2  1  5  2  1  1  1 11 97  1 NA  1  2  1  2  1  1  1  1  1 97
##     [97]  1  1  1 97  1  1  1  1  1  1  1  1  2  1  2 97 NA  1  1  1  1  1  1  1
##    [121]  2  1 97  1 NA  1  2  1  1  2  1  1 97  1  1  1  1  1  1  1  3  1 97  1
##    [145] 97  2  1  2  1  2  1  5  1  2  5 97  1  1  5 97  5  1 10  5  2  1  1  1
##    [169]  1  1  2  1  1  1  1  1  1  1  1  1  1  2  1  1  1 97  1  1  2  1  1  1 97
##    [193] 97  1  1  1  2 97  1  1  1  2  1  1  1  1  1  1  1  1  1  1  2  1  2  1
##    [217]  2 97  1  1  2  1  2  1  5  5  1  1  5  1  5  1  5  5  1  1  2  1  1  1
##    [241]  1  1  1  1 97  1 97  1  1  2  1  1 10  2  2  2  1  2  1  2  1  1  5  1
##    [265]  1 97  1  1  2  1  1 NA  1  1  1  1  1  1 97  1  2  1  1  1  2 97  1  1
##    [289]  1  2  1  1 97  1  1  2  1  1  1 97  1  1  1  2  2  2 97  1  1  1  1  1
##    [313]  1  1 97  6  1  1 97 97  1  1  1  5 97  3  2 10  1 NA 12 97  1 97  1 96
##    [337]  1  2  2  1 12  1 97  1  3 96  5  1  5  5  1  1  1  1  1  1 NA  1 97  1
##    [361]  1  1  1  2  1  1  1  1  1  1  2  1  1  1  1  1  1  1  1  1 97  1  1
##    [385]  1  1  2  1  1  1  1  1  1  2  1 97  1  1  1  1  1  2  1  1  2  5  1
```

Exploração de Banco de Dados no R

Para selecionar elementos dentro do banco de dados que sejam semelhantes ou obedecem a uma regra/padrão utilizamos a seguinte forma:

banco de dados\$coluna operador elemento

```
bd$b = 30
```

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

Similar a essa solução existe a função subset que cria bancos de dados com especificidades basta que vc assinale o banco de dados e qual a regra para extrair o dado

```
dt <- subset (bd, b ≥ 40)
dt2 <- subset (dados, S10 ≥ 1)
```

```
summary (dt2$S10)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   5.000   10.000   9.483  13.000  17.000
```

Exploração de Banco de Dados no R

É possível ainda extrair dados sobre o banco utilizando alguns dos comandos que já sabemos

```
sum (bd$b)
```

```
## [1] 220.6
```

```
mean (bd$b)
```

```
## [1] 44.12
```

```
median (bd$b)
```

```
## [1] 41.8
```

```
sd (bd$b)
```

```
## [1] 18.52706
```

Exploração de Banco de Dados no R

E utilizar alguns novos:

```
length(bd) #número de colunas
```

```
## [1] 2
```

```
summary (bd) #resumo dos dados
```

```
##      nomes                b
## Length:5             Min.   :23.00
## Class :character     1st Qu.:30.00
## Mode  :character     Median :41.80
##                               Mean  :44.12
##                               3rd Qu.:60.00
##                               Max.   :65.80
```

```
head (bd) # mostra as primeiras linhas
```

```
##      nomes      b
## 1      A 65.8
## 2      B 60.0
## 3      C 41.8
## 4      D 30.0
## 5      E 23.0
```

```
tail (bd) # mostra as últimas linhas
```

```
##      nomes      b
## 1      A 65.8
## 2      B 60.0
## 3      C 41.8
## 4      D 30.0
## 5      E 23.0
```

Slides created via the R package **xaringan**.

The chakra comes from **remark.js**, **knitr**, and R Markdown.