

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

O Universo Tidyverse

Métodos Quantitativos Aplicados à Ciência Política

Frederico Bertholini

14.set.2020

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

1 Abrindo o R e conhecendo o software

2 Vetores

3 Operações matemáticas

4 data.frame

5 Pacotes

6 Importação de dados

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

7 O Universo tidyverse

8 Importação no tidyverse

9 Pacotes dplyr e tidyr

10 select

11 filter

12 mutate

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

13 summarise

14 arrange

O Universo
Tidyverse

Frederico
Bertholini

**Abrindo o R e
conhecendo o
software**

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Abrindo o R e conhecendo o software

Passos iniciais

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- R é um ambiente de desenvolvimento para cálculos estatísticos e gráficos (um software de **PROGRAMAÇÃO ESTATÍSTICA**)
- Tem dividido com o Python o posto de software mais popular entre cientistas sociais
- Vantagens
 - Software Livre;
 - Documentação completa e acessível;
 - Diversidade de arquivos;
 - Replicabilidade de rotinas

Lei de ouro

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Se é possível fazer no Excel, é possível fazer no R
- Entretanto, se é possível fazer no R, não necessariamente é possível fazer no Excel
- Vários cálculos estatísticos mais sofisticados estão disponíveis no R através de pacotes desenvolvidos pela comunidade

Habilidades necessárias

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Escrita:
 - Elementos (ex: numeric, character, factor...);
 - Funções básicas (ex: sum(), table(), sd());
 - Composição do script (ex: c(), for loop);
 - Sensibilidade
- Leitura:
 - Identificação de funções;
 - Diferentes soluções;
 - Pacotes intuitivos;
 - Alertas de erros;

Habilidade mais importante

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Devido à multiplicidade de soluções e de alertas de erros
- Também, ao universo de pacotes com diversas funções
- A habilidade mais importante é saber pesquisar corretamente no Google
 - Melhor em inglês pela quantidade de fóruns;
 - Melhor fórum: stackoverflow.com
- É importante adquirir autonomia com o software

RStudio

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- RStudio é um ambiente para desenvolvimento do R (IDE)
- Ao longo do curso utilizaremos o Rstudio
Utilizando nada dos botões ou possibilidades específicas do RStudio
Portanto, utilizaremos a programação somente

Aparência do RStudio

O Universo
Tidyverse

Frederico
Bertholini

Abriendo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

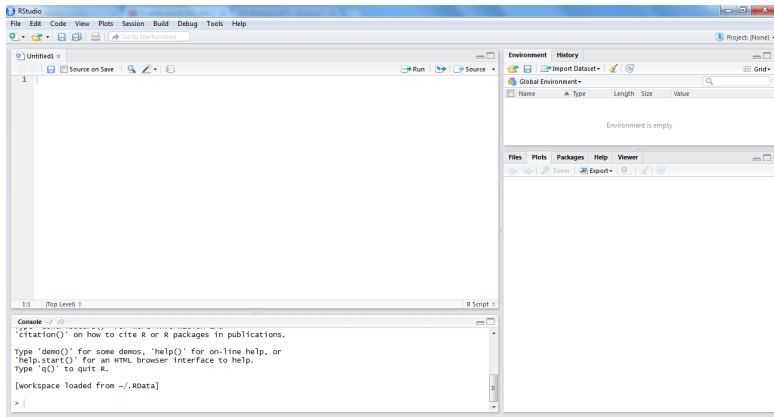


Figure 1: Aparência

Básico do básico

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para começar é importante apresentar um operador e um comando

- Hastags
Insere comentários sem gerar outputs
É importante para organizar e registrar dentro dos scripts

- Ctrl + enter (cmd + enter no mac)

Roda as funções e programações escritas nas linhas selecionadas
As linhas escritas no ambiente acima do console, onde é possível deixar registrado o script

Scripts são arquivos de texto

Fluxos: scripts e projetos

Como sai no R?

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

#Insere comentarios sem rodar o comando

Importante para organizar e

Erro: unexpected symbol in "Importante para"

##Registrar o que se vai fazer ou foi feito

- Quando o comando Ctrl + enter é acionado, o campo **Console** registra > e o comando escrito no script, em seguida

R como calculadora

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Uma função primordial e básica do R, como software estatístico, é de calculadora
- O R possui os operadores básicos da matemática como '+', '-', '*', e '/'

```
5 + 5
```

```
## [1] 10
```

```
5 - 3
```

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

```
4 * 9
```

```
## [1] 36
```

```
16 / 2
```

```
## [1] 8
```

R como calculadora

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Assim como na matemática, atenção em relação aos ()

```
(5 + 6) * 3
```

```
## [1] 33
```

```
5 + 6 * 3
```

```
## [1] 23
```

- Além das funções de exponencial e raiz quadrada

- Respectivamente, ^ e sqrt()

```
2 ^ 2
```

```
## [1] 4
```

```
sqrt(36)
```

```
## [1] 6
```

Lógica

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- O R permite também avaliações lógicas
- Ou seja, o software possui operadores lógicos afim de fazer testes lógicos com resultados de Verdadeiro ou Falso de acordo com a proposição
- Os principais operadores são ==, <, >, <=, >= e !=

```
5 == 5
```

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

```
5 <= 5 / 5
```

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

```
5 * 4 > 5
```

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

```
3 != 6
```

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


Lógica

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

■ Testamos também Verdade e Falsidade

```
TRUE == TRUE
```

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

```
TRUE <= FALSE
```

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

■ Assim como estamos caracteres

```
"Python" == "python"
```

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

```
"Stata" != "Sasta"
```

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

Operadores lógicos especiais

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Atenção para os operadores e e ou
- O primeiro, para ser verdade, precisa que todos os pressupostos sejam verdadeiros

```
(3 == 3) & (4 != 5)
```

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

- O ou, por sua vez, para ser verdade precisa que apenas 1 pressuposto seja verdadeiro

```
(3 != 3) | (4 != 5)
```

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

Atribuição

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Trata-se da famosa setinha que indica objetos (valores, vetores, dataframes) para alguma etiqueta
- Dessa maneira, podemos 'salvar' os objetos nas etiquetas para utilizarmos através dessas em qualquer momento ao longo do script
- Quando utilizadas em operações, as etiquetas representam aquilo que fora atribuído a elas
- Quando criamos a etiqueta, não geramos outputs, apenas quando rodamos diretamente a etiqueta

```
sorte <- 5
```

Regras do uso da setinha

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Atenção, letras maiúsculas e minúsculas importam

```
sorte <- 5
```

Sorte

Erro: objeto 'Sorte' não encontrado

- Também não podemos criar etiquetas que começam com números

```
15luck <- 15
```

Erro: unexpected symbol in "15luck"

- Cuidado com a utilização de etiquetas com o mesmo nome de funções, pode gerar confusão no script

Classes

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Em basicamente tudo que iremos fazer no R, a classe da informação importa
- Em termos elementares, ou ao nível dos valores, existem três grandes classes:
 - Numeric (numérico);
 - Logical (lógico);
 - Character ou factor (caracteres);
- Para obter a informação sobre a classe, iremos aprender nossa primeira função no R: `class()`

Númerico

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Numeric é a classe composta por valores numéricos

```
class(sorte)
```

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

- Objetos do tipo “numeric” permitem funções matemáticas como média, mediana etc...
- No caso de valores decimais, utilizamos ponto ao invés de vírgula

```
decimal <- 3,5
```

```
# Erro: ', ' inesperado in "decimal <- 3,"
```

```
decimal <- 3.5
```

Lógico

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Logical é a classe composta por TRUE, FALSE e NA

```
vdd <- TRUE
```

```
class(vdd)
```

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

- Não é preciso escrever sempre TRUE e FALSE, isto é, podemos resumir para T e F, respectivamente
- Por trás dos valores T e F, há valores numéricos correspondente a 1 e 0

```
T + F
```

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

Caracteres

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Characters é a classe composta por nomes
- Importante característica é que os nomes devem estar dentro de aspas, caso contrário, o R não reconhecerá como caracter

```
nome <- "Alvaro"
```

```
nome
```

```
## [1] "Alvaro"
```

```
class(nome)
```

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


Caracteres vs Fatores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Em termos estatísticos, não há diferença entre caracteres e fatores
- Fatores apresentam as categorias por dentro de vetor.
- Em termos de manipulação de dataframes, porém, veremos que esses dois objetos terão tratamento distintos, a começar pela importação da base e a definição de string tratada como factor

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Vetores

O que são vetores?

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Vetores são combinações de valores em uma estrutura unidimensional
- Ou seja, podem ser combinações de números, valores lógicos, nomes e várias outras possibilidades
- Por exemplo, combinação de valores pares: 2,4,6,8
- Ou a combinações de nomes começando por P: Pedro, Paula, Pietro, Paloma
- Combinação de valores lógico: TRUE, FALSE, TRUE, TRUE

Criação de vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para criar vetores é importante combinar valores
- A função necessária para a criação de vetores, portanto, é a função `c()`, *c* de *combine* ou *concatenate*

```
c(2,4,6,8)
```

```
## [1] 2 4 6 8
```

```
c("Pedro","Paula","Pietro","Paloma")
```

```
## [1] "Pedro" "Paula" "Pietro" "Paloma"
```

```
c(TRUE,FALSE,TRUE,FALSE)
```

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

Etiquetas para vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Assim como fizemos com os valores, atribuímos também etiquetas aos vetores
- Dessa maneira, podemos trabalhar com as sequências ao longo do script, assim como dar razão ou justificativa aos valores combinados

```
n.pares <- c(2,4,6,8)
```

```
nomes.com.p <-  
c("Pedro", "Paula", "Pietro", "Paloma")
```

```
valores.log <- c(TRUE, FALSE, TRUE, FALSE)
```

Classes dos vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Os vetores também possuem classes
- Essas categorias determinam as operações possíveis dentro de um vetor
- Para obter a informação sobre a classe de um vetor, utilizamos a função `class()`

```
class(n.pares)
```

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

```
class(nomes.com.p)
```

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

```
class(valores.log)
```

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

Comprimento de vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Os vetores podem ser medidos em relação ao seu comprimento
- O seu comprimento define a extensão do vetor, assim como quantos elementos estão presentes dentro da combinação de valores
- O comprimento pode ser medido pela função `length()`

```
length(n.pares)
```

```
## [1] 4
```

```
length(nomes.com.p)
```

```
## [1] 4
```

```
length(valores.log)
```

```
## [1] 4
```

Somatório de vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para vetores numérico podemos somar os valores de um vetor
- A função se chama `sum()` que corresponde ao \sum

```
sum(n.pares)
```

```
## [1] 20
```

- Em vetores com valores lógicos, com TRUE e FALSE, o `sum` soma o número de T que temos dentro de um vetor

```
sum(valores.log)
```

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


Somatório de vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- A função `sum()` pode se tornar ainda um contador de um teste
- Por exemplo, queremos saber quantos nomes são iguais ao de Pedro no vetor `'nomes.com.p'`

```
teste1 <- nomes.com.p == "Pedro"
```

```
sum(teste1)
```

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

```
teste1
```

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

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Queremos selecionar no vetor `nomes.com.p` o segundo elemento que é Paula
- Vejamos, primeiro, como se compõe o vetor de caracteres `nomes.com.p` e a lógica do posicionamento dentro do vetor

```
nomes.com.p
```

```
## [1] "Pedro" "Paula" "Pietro" "Paloma"
```

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Vamos testar se Paula se inclui dentro do vetor sem termos que ver no console
- As vezes, vetores são maiores do que temos como exemplo
- O teste lógico, portanto, utilizamos o operado `%in%`, que retorna TRUE, caso o valor esteja incluso no vetor

```
"Paula" %in% nomes.com.p
```

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

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Algumas operações se restringem à determinados elementos ou à um conjunto deles incluso dentro de um vetor
- Para realizar tais procedimentos precisamos saber escolher elementos
- Para isso, utilizaremos o operador '[' depois de apontar o vetor
- Mais especificamente: Vetor; []; Posição ou regra;

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para escolher Paula, portanto, vamos definir como a segunda posição dentro do vetor

```
nomes.com.p[2]
```

```
## [1] "Paula"
```

- O posicionamento dos nomes ao longo do vetor determina como seleciona-lo individualmente
- Porém, podemos selecionar através do valor

```
nomes.com.p[nomes.com.p == "Paula"]
```

```
## [1] "Paula"
```

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Seleccionamos também o último valor do vetor `nomes.com.p`
- Sabemos que o vetor possui 4 valores utilizando a função `length()`

```
length(nomes.com.p)
```

```
## [1] 4
```

- Assim o último valor é igual a 4
- Para seleccionarmos o último valor, ou definimos como a posição 4 ou apenas o `length()` do vetor

```
nomes.com.p[4]
```

```
## [1] "Paloma"
```

```
nomes.com.p[length(nomes.com.p)]
```

```
## [1] "Paloma"
```

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para seleção de elementos através de regras, podemos utilizar o vetor numérico de `n.pares`

```
n.pares <- c(2,4,6,8)
```

```
n.pares
```

```
## [1] 2 4 6 8
```

- O objetivo aqui é selecionar os elementos maiores que 5, que é a mediana do vetor

```
median(n.pares)
```

```
## [1] 5
```

Seleção de elementos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Seleccionamos, assim, os elementos acima da mediana do vetor `n.pares`, que são os valores 6 e 8

```
n.pares[n.pares > 5]
```

```
## [1] 6 8
```

- Ou ainda podemos definir o valor 5 como o valor da mediana
- O resultado é o mesmo, porém, deixa claro ao leitor do seu script que se trata dos valores acima da mediana do vetor `n.pares`

```
n.pares[n.pares > median(n.pares)]
```

```
## [1] 6 8
```


O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

**Operações
matemáticas**

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Operações matemáticas

Operações com vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Em vetores numéricos, podemos fazer operações matemáticas a partir do seu posicionamento
- Vamos calcular o alcance do vetor “n.pares”

#ultimo elemento

```
n.pares[length(n.pares)]
```

```
## [1] 8
```

#primeiro elemento

```
n.pares[1]
```

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

- Subtraindo esses dois elementos, temos o alcance do vetor

```
n.pares[length(n.pares)] - n.pares[1]
```

```
## [1] 6
```

Operações com vetores

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Ainda em vetores numéricos, vamos calcular vetores com regras
- Selecionamos elementos maiores que 5

```
n.pares[n.pares >= 5]
```

```
## [1] 6 8
```

- Queremos duplicar os valores inclusos dentro da condição

```
n.pares[n.pares >= 5] * 2
```

```
## [1] 12 16
```

- Por fim, operar um conjunto com um elemento

```
n.pares[n.pares >= 5] * n.pares[1]
```

```
## [1] 12 16
```

Conjuntos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Utilizamos a teoria de conjuntos no R com a função de manipular e identificar elementos comuns ou diferente entre vetores
- Já temos o vetor “n.pares” e vamos criar um vetor com números naturais

```
n.naturais <- c(0,1,2,3,4,5,6,7,8,9)
n.naturais
```

```
## [1] 0 1 2 3 4 5 6 7 8 9
```

- Se perguntarmos por valores que não estão no vetor “n.naturais”, o resultado será conjunto vazio

```
n.naturais[n.naturais == 10]
```

```
## numeric(0)
```

Conjuntos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Outro ponto importante é saber se os elementos estão contidos dentro de outro vetor
- A pergunta a se fazer é: os elementos do vetor “n.partidos” está contido no vetor “n.naturais”
- No R, contidos é igual a `%in%`

```
n.pares %in% n.naturais
```

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

- Isto é, o vetor “n.pares” está contido no vetor “n.naturais” e é um subconjunto
- Outra maneira, porém é utilizando a função `is.element()`

```
is.element(n.pares,n.naturais)
```

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

Conjuntos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Para encontrar os valores exclusivos de um conjunto, vamos utilizar a escrita da seleção de elementos
- Podemos incluir a notação de “!”, que quer dizer diferente
- Portanto, selecionamos os valores que estão contido no conjunto dos naturais, porém não no conjunto dos valores pares

```
n.naturais[!is.element(n.naturais,n.pares)]
```

```
## [1] 0 1 3 5 7 9
```

Conjuntos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Por fim, vamos testar se um valor está presente em um desses vetores
- Testaremos os valores 1, 11 e 21

```
1 %in% c(n.naturais,n.pares)
```

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

```
11 %in% c(n.naturais,n.pares)
```

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

```
21 %in% c(n.naturais,n.pares)
```

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

- Assim, não temos os valores 11 e 21 em nenhum dos vetores, entendidos como conjuntos

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

data.frame

data.frame

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Um data.frame é o mesmo que uma tabela do SQL ou uma planilha Excel
- seus dados provavelmente serão importados para um objeto data.frame
- data.frame's são listas especiais em que todos os elementos possuem o mesmo comprimento.
- Cada elemento dessa lista pode ser pensado como uma coluna da tabela - ou como uma variável. Uso do '\$'
- Seu comprimento representa o número de linhas - ou seja, de observações

data.frame

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Como data.frames's são listas, suas colunas podem ser de classes diferentes. Essa é a grande diferença entre data.frame's e matrizes.

Funções úteis:

```
head() # Mostra as primeiras 6 linhas.
```

```
tail() # Mostra as últimas 6 linhas.
```

```
dim() # Número de linhas e de colunas.
```

```
names() # Os nomes das colunas (variáveis).
```

```
str() # Estrutura do data.frame. Mostra, entre outras coisas,
```

```
cbind() # Acopla duas tabelas lado a lado.
```

```
rbind() # Empilha duas tabelas.
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Pacotes

O que são pacotes

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- O R possui diversas funções já instaladas dentro da sua programação
- Exemplos são `sum()`, `length()`, `class()`, `c()`
- Outras, porém, devem ser instaladas para que possam ser utilizadas pelos usuários
- A maneira com a qual instalamos novas funções, não definidas anteriormente no software, é através de pacotes
- Pacotes concentram diversas funções para diversas demandas
 - Importação de dados;
 - Organização de banco de dados;
 - Análises estatísticas específicas;
 - Gráficos diferenciados;

O que são pacotes

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- A instalação de qualquer pacote pode ser feita por dentro do R
- Para isso, porém, é preciso primeiro conexão com a internet, já que o R busca o novo pacote no repositório de pacotes
- A função para instalar pacote, portanto, é `install.packages()`
- O nome da nova função deve vir, primeiramente entre parenteses
- Podemos começar instalando o pacote para importação de bases de dados: `foreign`

```
install.packages("foreign")
```

- Após alguns segundos, e algumas mensagens no console, a instalação será efetivada

Ativar pacotes

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Cada pacote, inclusive o foreign, tem uma documentação disponível na internet
- Nessa documentação estão disponíveis as funções que o pacote possui, além do nome do seu criador
- As funções não ficam disponíveis assim que o pacote termina a instalação
- Para ativar as funções do pacote, é preciso utilizar a função `library()`

```
library(foreign)
```

- Repare, que uma vez instalado, o nome do pacote não precisa mais estar entre aspas

Ativar pacotes

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Portanto, para começarmos o procedimento de instalação do pacote, seguimos os passos a seguir:
Caso não esteja instalado, instalar o pacote através da função `install.packages()`;
Para ativar o pacote, utilizar a função `library()` sem as aspas no nome do pacote instalado

```
install.packages("foreign")
```

```
library(foreign)
```

- Uma vez instalado o pacote, não é preciso instalar mais a não ser que você reinstale o R

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

**Importação de
dados**

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Importação de dados

Passo a passo

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- A importação é uma das tarefas que demandam mais atenção no R
- É preciso ter um conhecimento prévio de como sua base externa está constituída
- Outra informação importante é a extensão do arquivo da base
- Primeiramente, a informação que deve ser dada ao software é onde está a base - diretório de trabalho
- A função necessária é `setwd()` que define o diretório da sua seção no R
- Dentro da função, iremos inserir o local do arquivo
Em caso de Windows, inverta as barras ou duplique;
Não se esqueça das aspas;

```
setwd("/Volumes/Macintosh HD/MQCP_IPOL_2020/Slides/aula 02
```

Passo a passo

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Esse diretório definido significa que as bases e os gráficos produzidos serão enviados para essa pasta
- Finalmente, vamos importar as bases de dados
- Primeiro, vamos importar a base de extensão txt com o nome baserm
- Não é preciso de pacote para esse procedimento

```
lines <- readLines("dados/baserm.txt")
```

```
baserm <- read.table(text = lines, sep = '\t')
```

- Repare que definimos a base dentro das aspas e com a extensão
- Na segunda linha, o primeiro argumento é o texto, o segundo argumento trata de como os dados estão separados, geralmente txt vem separado assim

Passo a passo

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- No pacote foreign, a forma mais genérica de importação da base é o `read.table()`
- Entretanto, o pacote apresenta uma série de especialidades, a depender da extensão em questão
- Para CSV, vimos que tem a `read.csv()`. Já para dta, base de origem do stata, temos a função `read.dta()`
- O pacote foreign não possui a extensão `xlsx` e `xls`, extensão muito encontrada e comum entre as bases de dados disponíveis
- Para isso, vamos instalar um novo pacote `readxl`

```
install.packages("readxl")
```

- Esse pacote disponibiliza as funções `read_xls()` e `read_xlsx()`

Passo a passo

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Vamos ativar as funções disponíveis no pacote readxl com a função library()

```
library(readxl)
```

- Vamos importar a base controle_cgu_municípios.xlsx

```
cgu <-  
read_xlsx("dados/controle_cgu_municípios.xlsx")
```

- Repare que acessamos apenas a primeira página da base
- Para acessarmos a segunda páginas, utilizamos o argumento sheet=2

```
cgu <-  
read_xlsx("dados/controle_cgu_municípios.xlsx",  
sheet = 2)
```

- Alguns sinais de alerta surgem, porém não se trata de erro

Importação por pacote

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

**Importação de
dados**

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Por fim, podemos importar dados através de pacotes
- Após ativar o pacote “ElectionsBR”, a função `legend_fed()` fica disponível para a importação
- Como toda a função, ou quase todas, possui argumentos importantes:
 - ‘year =’ se refere ao ano de extração
 - ‘uf =’ se refere à UF

Importação

- Vamos começar importando dados de coalizões pré-eleitorais (coligações) do DF em 2018, nos retornando um objeto em 'tbl_df' e data frame

```
library(electionsBR)
```

```
##
```

```
## To cite electionsBR in publications, use: citation('ele
```

```
## To learn more, visit: http://electionsbr.com
```

```
leg_df_2018 <- legend_fed(year = 2018,uf="DF")
```

```
## Processing the data...
```

```
## Warning: `as.tbl()` is deprecated as of dplyr 1.0.0.
```

```
## Please use `tibble::as_tibble()` instead.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_warnings()` to see where this war
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

Abriendo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Primeira coisa importante de se informar é a classe desses objetos
- Temos 4 objetos: baserm, cgu, educacao e pnad2018

```
class(baserm)
```

```
## [1] "data.frame"
```

```
class(cgu)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
class(leg_df_2018)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Uma visão completa da base é o comando `View()`
- Entretanto, cuidado, dependendo do tamanho da base, podemos travar o software

```
View(baserm)
```

- Repare no V maiúsculo, lembre-se que o R é bastante sensível na sua linguagem
- O `View()` abre uma nova aba com a base no formato de grade
- Podemos, assim, visualizar a base de dados na forma mais intuitiva

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Porém, para bases como a pnad2018, por exemplo, sabemos que é grande demais para sua visualização ser feita através do View()
- Algumas funções podem nos ajudar nessa tarefa
- A primeira é o dim, que as dimensões da base

```
dim(leg_df_2018)
```

```
## [1] 251 21
```

- O primeiro valor sempre retrata o número de linhas, ou observações, enquanto o segundo valor apresenta o número de colunas, ou variáveis
- A função ncol() e length() também indicam quantas colunas, ou variáveis estão presentes na base

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Outra função importante na visualização de bases de dados é a lista de nomes
- A função `names()` descreve as variáveis presentes na base
- Isso facilita no momento de selecionar as variáveis que entrarão na análise de vocês

```
names(cgu)
```

```
## [1] "COD.IBGE7" "REGIÃO"
## [3] "UF" "PORTE"
## [5] "MUNICÍPIO" "falha"
## [7] "tempo_falha_02" "tempo_falha_01"
## [9] "reicidência_falha" "ano_eleitoral"
## [11] "PERCENT_ganhador_2000" "PERCENT_ganhador_2004"
## [13] "PERCENT_ganhador_2008" "PERCENT_ganhador_2012"
## [15] "PERCENT_ganhador_médio" "competição_pol_alta"
## [17] "Ideologia1" "Ideologia2"
## [19] "Ideologia3" "ideologia 4"
## [21] "ideologia_media" "IDHM_2000"
```

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Outra função possível é o `str()`
- Essa função apresenta o nome das variáveis, a classe de cada uma delas e os primeiros valores

```
str(leg_df_2018)
```

```
## tibble [251 x 21] (S3: tbl_df/tbl/data.frame)
## $ DATA_GERACAO      : chr [1:251] "27/08/2020" "27/0
## $ HORA_GERACAO       : 'hms' num [1:251] 16:33:45 16:
##   ..- attr(*, "units")= chr "secs"
## $ ANO_ELEICAO        : num [1:251] 2018 2018 2018 201
## $ COD_TIPO_ELEICAO   : num [1:251] 2 2 2 2 2 2 2 2 2
## $ NM_TIPO_ELEICAO    : chr [1:251] "ELEIÇÃO ORDINÁRIA
## $ NUM_TURNO          : num [1:251] 1 1 1 1 1 1 1 1 1
## $ COD_ELEICAO        : num [1:251] 297 297 297 297 29
## $ DESCRICAO_ELEICAO  : chr [1:251] "Eleições Gerais E
## $ DATA_ELEICAO      : chr [1:251] "07/10/2018" "07/1
## $ SIGLA_UF           : chr [1:251] "DF" "DF" "DF" "DF
## $ SIGLA_UF           : chr [1:251] "DF" "DF" "DF" "DF"
```

Visualizando a base

O Universo
Tidyverse

Frederico
Bertholini

- Finalmente, a função `head()` e `tail()`
- A primeira função apresenta os primeiros valores de uma base de dados

```
head(baserm, 2)
```

```
##      sigla cod.ibge  estado anoeleitoral1990 anoeleitoral1991
## 1      AC        12   Acre                1
## 2      AL        27 Alagoas                1
##      anoeleitoral1993 anoeleitoral1994 anoeleitoral1995 anoeleitoral1996
## 1                0                1                0
## 2                0                1                0
##      anoeleitoral1997 anoeleitoral1998 anoeleitoral1999 anoeleitoral2000
## 1                0                1                0
## 2                0                1                0
##      anoeleitoral2001 anoeleitoral2002 anoeleitoral2003 anoeleitoral2004
## 1                0                1                0
## 2                0                1                0
##      anoeleitoral2005 anoeleitoral2006 anoeleitoral2007 anoeleitoral2008
```

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

**O Universo
Tidyverse**

**Frederico
Bertholini**

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

**O Universo
tidyverse**

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

O Universo tidyverse

Manifesto tidyverse

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

O tidyverse, também chamado por muitos de hadleyverse, é um conjunto de pacotes que, por compartilharem esses princípios do manifesto tidy, podem ser utilizados naturalmente em conjunto. Pode-se dizer que existe o R antes do tidyverse e o R depois do tidyverse.

Os princípios fundamentais do tidyverse são:

- Reutilizar estruturas de dados existentes.
- Organizar funções simples usando o pipe.
- Aderir à programação funcional.
- Projetado para ser usado por seres humanos.

Manifesto tidy

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Tidy Tools Manifesto <https://cran.r-project.org/web/packages/tidyverse/vignettes/manifesto.html>
- Tidy data vignette <https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html>
- Tidy Data paper <http://vita.had.co.nz/papers/tidy-data.pdf>
- Conjunto de pacotes <https://www.tidyverse.org/packages/>

Usando o pipe - O operador %>%

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

O operador %>% (pipe) foi uma das grandes revoluções recentes do R, tornando a leitura de códigos mais lógica, fácil e compreensível.

```
library(tidyverse)
library(magrittr)
```


Ideia

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

A ideia do operador `%>%` (pipe) é bem simples: usar o valor resultante da expressão do lado esquerdo como primeiro argumento da função do lado direito.

■ As duas linhas abaixo são equivalentes.

```
f(x, y)
```

```
x %>% f(y)
```

E se aumentarmos o código?

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Vamos calcular a raiz quadrada da soma dos valores de 1 a 4.

Primeiro, sem o pipe.

```
sqrt(sum(x))
```

```
## [1] 3.162278
```

Agora com o pipe.

```
x %>%  
  sum %>%  
  sqrt
```

```
## [1] 3.162278
```

E se realmente tivermos muitas funções aninhadas?

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

A utilização do pipe transforma um código confuso e difícil de ser lido em algo *simples e intuitivo*.

Receita de bolo - sem pipe

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Tente entender o que é preciso fazer.

```
esfrie(  
  asse(  
    coloque(  
      bata(  
        acrescente(  
          recipiente(rep("farinha", 2), "água",  
                      "fermento", "leite", "óleo"),  
          "farinha", até = "macio"),  
          duração = "3min"),  
        lugar = "forma", tipo = "grande",  
        untada = TRUE), duração = "50min"),  
    "geladeira", "20min")
```

Receita de bolo - com pipe

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Desistiu? Agora veja como fica escrevendo com o %>%:

```
recipiente(rep("farinha", 2), "água", "fermento", "leite",  
  crescente("farinha", até = "macio") %>%  
  bata(duração = "3min") %>%  
  coloque(lugar = "forma", tipo = "grande", untada = TRUE)  
  asse(duração = "50min") %>%  
  esfrie("geladeira", "20min"))
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

**Importação no
tidyverse**

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Importação no tidyverse

Importação com readr, readxl, haven e DBI

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

No tidyverse, geralmente

- Funções `read_<formato>` servem para ler um arquivo no formato `<formato>`
- Funções `write_<formato>` servem para escrever num arquivo com o formato `<formato>`

Arquivos de texto

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- csv, tsv, txt, ...
- Para esses aqui, usar o pacote readr
- Você também pode experimentar o `data.table::fread`

Arquivos binários

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- .RData, .rds, .feather, .fst
- .dta (Stata), .sas7bdat (SAS), .sav (SPSS)
- Ler com readr, haven, feather, fst.

Bancos de dados

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- MySQL, SQL Server, PostgreSQL, SQLite, ...
- Spark, MongoDB, Hive, ...
- Utilizar pacotes DBI e odbc

Tidy data e janitor

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Veremos mais à frente, mas `janitor::clean_names()` é uma ferramenta tidy

```
library(janitor)
```

```
##
```

```
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      chisq.test, fisher.test
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

**Pacotes dplyr e
tidyr**

select

filter

mutate

summarise

Pacotes dplyr e tidyr

Conjunto de dados

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

Vamos trabalhar com a base decisoes, que contém decisões do Tribunal de Justiça de São Paulo

```
decisoes <- read_rds("dados/decisoes.rds")  
glimpse(decisoes)
```

```
## Rows: 11,731  
## Columns: 9  
## $ `ID Decisão`      <chr> "11094999", "11093733", "11093  
## $ n_processo        <chr> "0057003-20.2017.8.26.0000", "  
## $ `Classe/Assunto`  <chr> "Habeas Corpus / Homicídio Sim  
## $ Município         <chr> "Cosmópolis", "São Paulo", "Ri  
## $ Câmara           <chr> "3ª Câmara de Direito Criminal  
## $ `Data decisão`    <chr> "19/12/2017", "19/12/2017", "1  
## $ `Data registro`   <chr> "19/12/2017", "19/12/2017", "1  
## $ Juiz              <chr> "Luiz Antonio Cardoso", "Luiz  
## $ `txt decisão`     <chr> NA, NA, NA, "Execução Penal -
```

```
decisoes <- read_rds("dados/decisoes.rds") %>%  
  janitor::clean_names() # com dois pontos eu não preciso
```

Características do dplyr

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

**Pacotes dplyr e
tidyr**

select

filter

mutate

summarise

- A utilização é facilitada com o emprego do operador `%>%`
- No primeiro argumento colocamos o `data.frame` ou o `tibble`, e nos outros argumentos colocamos o que queremos fazer.

As cinco funções principais do dplyr

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- **select**: selecionar colunas
- **filter**: filtrar linhas
- **mutate**: criar colunas
- **summarise**: sumarizar colunas
- **arrange**: ordenar linhas

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

select

select

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Utilizar `starts_with(x)`, `contains(x)`, `matches(x)`, `one_of(x)`, etc.
- Possível colocar nomes, índices, e intervalos de variáveis com `:`.

Em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%
```

```
select(id_decisao, n_processo, municipio, juiz)
```

```
## # A tibble: 11,731 x 4
```

```
##   id_decisao n_processo      municipio
```

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

```
## 1 11094999    0057003-20.2017.8.26.0000 Cosmópolis
```

```
## 2 11093733    0052762-03.2017.8.26.0000 São Paulo
```

```
## 3 11093677    0055169-79.2017.8.26.0000 Ribeirão Preto
```

```
## 4 11093270    9000580-82.2017.8.26.0032 Araçatuba
```

```
## 5 11093374    0052938-79.2017.8.26.0000 São Paulo
```

```
## 6 11093320    9000723-79.2017.8.26.0482 Presidente Prud
```

```
## 7 11091506    0003276-86.2015.8.26.0075 Bertioga
```

```
## 8 11093326    9000298-11.2017.8.26.0625 Taubaté
```

```
## 9 11092475    0004653-39.2015.8.26.0028 Aparecida
```

```
## 10 11093773    2221930-66.2017.8.26.0000 Jandira
```

```
## # ... with 11,721 more rows
```

Em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%
```

```
select(classe_assunto:id_decisao, juiz)
```

```
## # A tibble: 11,731 x 4
```

```
##   classe_assunto                                n_processo
```

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

```
## 1 Habeas Corpus / Homicídio Simples          0057003-20.2017.
```

```
## 2 Habeas Corpus / Roubo                       0052762-03.2017.
```

```
## 3 Habeas Corpus / DIREITO PENAL               0055169-79.2017.
```

```
## 4 Agravo de Execução Penal / Pena Pr~         9000580-82.2017.
```

```
## 5 Mandado de Segurança / Crimes do S~         0052938-79.2017.
```

```
## 6 Agravo de Execução Penal / Pena Pr~         9000723-79.2017.
```

```
## 7 Apelação / Tráfico de Drogas e Con~         0003276-86.2015.
```

```
## 8 Agravo de Execução Penal / Livrame~         9000298-11.2017.
```

```
## 9 Apelação / Tráfico de Drogas e Con~         0004653-39.2015.
```

```
## 10 Habeas Corpus / Furto Qualificado          2221930-66.2017.
```

```
## # ... with 11,721 more rows
```

Em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%  
  select(id_decisao, starts_with('data_'))
```

```
## # A tibble: 11,731 x 3  
##   id_decisao data_decisao data_registro  
##   <chr>      <chr>      <chr>  
## 1 11094999    19/12/2017    19/12/2017  
## 2 11093733    19/12/2017    19/12/2017  
## 3 11093677    19/12/2017    19/12/2017  
## 4 11093270    14/12/2017    19/12/2017  
## 5 11093374    14/12/2017    19/12/2017  
## 6 11093320    14/12/2017    19/12/2017  
## 7 11091506    14/12/2017    19/12/2017  
## 8 11093326    14/12/2017    19/12/2017  
## 9 11092475    14/12/2017    19/12/2017  
## 10 11093773    19/12/2017    19/12/2017  
## # ... with 11,721 more rows
```

Operações

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- selecione as colunas que acabam com "cisao".

```
decisoes %>%  
  select(ends_with("cisao"))
```

```
## # A tibble: 11,731 x 3
```

```
##       id_decisao data_decisao txt_decisao
```

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

```
##    1 11094999    19/12/2017    <NA>
```

```
##    2 11093733    19/12/2017    <NA>
```

```
##    3 11093677    19/12/2017    <NA>
```

```
##    4 11093270    14/12/2017    "Execução Penal - Comutação
```

```
##    5 11093374    14/12/2017    "Mandado de segurança - Impe
```

```
##    6 11093320    14/12/2017    "Execução Penal - Apuração d
```

```
##    7 11091506    14/12/2017    "Tráfico de entorpecentes -
```

```
##    8 11093326    14/12/2017    "Execução Penal - Pedido de
```

```
##    9 11092475    14/12/2017    "Tráfico de entorpecentes -
```

```
##   10 11093773    19/12/2017    <NA>
```

```
## # ... with 11,721 more rows
```

Operações

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- tire as colunas de texto = 'txt_decisao' e classe/assunto = 'classe_assunto'.
 - Dica: veja os exemplos de ?select em Drop variables ...

```
decisoes %>%
```

```
select(-classe_assunto, -txt_decisao)
```

```
## # A tibble: 11,731 x 7
```

```
##   id_decisao n_processo      municipio      camara      data_dec
```

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

```
## 1 11094999   0057003-20.~ Cosmópolis 3ª Câma~ 19/12/20
```

```
## 2 11093733   0052762-03.~ São Paulo 3ª Câma~ 19/12/20
```

```
## 3 11093677   0055169-79.~ Ribeirão ~ 3ª Câma~ 19/12/20
```

```
## 4 11093270   9000580-82.~ Araçatuba 8ª Câma~ 14/12/20
```

```
## 5 11093374   0052938-79.~ São Paulo 8ª Câma~ 14/12/20
```

```
## 6 11093320   9000723-79.~ President~ 8ª Câma~ 14/12/20
```

```
## 7 11091506   0003276-86.~ Bertioiga 8ª Câma~ 14/12/20
```

```
## 8 11093326   9000298-11.~ Taubaté   8ª Câma~ 14/12/20
```

```
## 9 11092475   0004653-39.~ Aparecida 8ª Câma~ 14/12/20
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

filter

filter

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Use , ou & para “e” e | para “ou”.
- Condições separadas por vírgulas é o mesmo que separar por &.

filter em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%
```

```
  select(n_processo, id_decisao, municipio, juiz) %>%  
  filter(municipio == 'São Paulo')
```

```
## # A tibble: 2,446 x 4
```

##	n_processo	id_decisao	municipio	juiz
##	<chr>	<chr>	<chr>	<chr>
##	1	0052762-03.2017.8.26.0000	11093733	São Paulo Luiz
##	2	0052938-79.2017.8.26.0000	11093374	São Paulo Grass
##	3	2214049-38.2017.8.26.0000	11093604	São Paulo Luiz
##	4	2227499-48.2017.8.26.0000	11093642	São Paulo Luiz
##	5	9002384-31.2017.8.26.0050	11093376	São Paulo Grass
##	6	0021158-39.2015.8.26.0050	11091508	São Paulo Grass
##	7	7005375-26.2015.8.26.0198	11091668	São Paulo Grass
##	8	9002039-65.2017.8.26.0050	11094451	São Paulo Grass
##	9	2203993-43.2017.8.26.0000	11094449	São Paulo Grass
##	10	0099423-21.2016.8.26.0050	11091474	São Paulo Grass
##	#	... with 2,436 more rows		

Dica: usar %in%

O Universo
Tidyverse

Frederico
Bertholini

Abriendo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
library(lubridate) # para trabalhar com as datas
#`day(dmy(data_decisao))` pega o dia da decisão.
```

decisoes %>%

```
select(id_decisao, municipio, data_decisao, juiz) %>%
# municipio igual a campinas ou jaú, OU dia da decisão m
filter(municipio %in% c('Campinas', 'Jaú') | day(dmy(dat
```

```
## # A tibble: 3,352 x 4
```

```
##      id_decisao municipio data_decisao juiz
##      <chr>      <chr>      <chr>      <chr>
##  1 11093272    Campinas  14/12/2017  Grassi Neto
##  2 11093359    Campinas  07/12/2017  Grassi Neto
##  3 11088333    Campinas  14/12/2017  Grassi Neto
##  4 11093018     Jaú      28/11/2017  Ivan Sartori
##  5 11089105     Jaú      14/12/2017  Ricardo Tucunduva
##  6 11089111    Campinas  14/12/2017  Ricardo Tucunduva
##  7 11091386    Santos    27/11/2017  Ivo de Almeida
##  8 11091385    Araçatuba 27/11/2017  Ivo de Almeida
```

Mais ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%  
  select(juiz) %>%  
  # filtra juizes que têm `Z` ou `z` no nome  
  filter(str_detect(juiz, regex("z", ignore_case = TRUE)))  
  # conta e ordena os juizes em ordem decrescente  
  count(juiz, sort = TRUE) %>%  
  head(5)
```

```
## # A tibble: 5 x 2  
##   juiz                                n  
##   <chr>                          <int>  
## 1 Gilberto Ferreira da Cruz      237  
## 2 Diniz Fernando                 198  
## 3 Sérgio Mazina Martins          173  
## 4 Luiz Antonio Cardoso           163  
## 5 Rachid Vaz de Almeida          150
```

Obs

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

A função `str_detect()` retorna `TRUE` se um elemento do vetor de textos é compatível com uma *expressão regular*. Estudaremos o pacote `stringr` e as funções `str_*` em outra aula.

- filtre apenas casos em que id_decisao não é NA

```
decisoes %>%
```

```
  filter(is.na(id_decisao))
```

```
## # A tibble: 65 x 9
```

```
##   id_decisao n_processo classe_assunto municipio camar
```

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

```
## 1 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 2 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 3 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 4 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 5 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 6 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 7 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 8 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 9 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## 10 <NA>      <NA>      <NA>      <NA>      <NA>
```

```
## # ... with 55 more rows, and 3 more variables: data_reg
```

■ filtre todas as decisões de 2018.

– Dica: função `lubridate::year()`

```
decisoes %>%
```

```
  filter(year(dmy(data_decisao)) == 2018)
```

```
## # A tibble: 314 x 9
```

```
##       id_decisao n_processo classe_assunto municipio camar
```

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

```
##    1 11107242    0009617-6~  Apelação / Ro~  São Paulo  2ª Câ
```

```
##    2 11107425    2227593-9~  Habeas Corpus~  Iepê      2ª Câ
```

```
##    3 11107492    0076977-2~  Embargos de D~  São Paulo  2ª Câ
```

```
##    4 11107361    0012191-3~  Agravo de Exe~  Campinas  2ª Câ
```

```
##    5 11107383    2218460-2~  Habeas Corpus~  Sorocaba  2ª Câ
```

```
##    6 11107331    0006928-6~  Agravo de Exe~  Sorocaba  2ª Câ
```

```
##    7 11107651    0000297-5~  Apelação / Tr~  Junqueir~  2ª Câ
```

```
##    8 11107485    2225548-1~  Habeas Corpus~  Nazaré P~  2ª Câ
```

```
##    9 11107335    0006934-7~  Agravo de Exe~  Sorocaba  2ª Câ
```

```
##   10 11107340    0006682-6~  Agravo de Exe~  Sorocaba  2ª Câ
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

mutate

mutate

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Aceita várias novas colunas iterativamente.
- Novas variáveis devem ter o mesmo `length` que o `nrow` do `bd` original ou 1.

mutate em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%
```

```
  select(n_processo, data_decisao, data_registro) %>%  
  mutate(tempo = dmy(data_registro) - dmy(data_decisao))
```

```
## # A tibble: 11,731 x 4
```

```
##       n_processo      data_decisao data_registro
```

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

```
##    1 0057003-20.2017.8.26.0000 19/12/2017 19/12/2017
```

```
##    2 0052762-03.2017.8.26.0000 19/12/2017 19/12/2017
```

```
##    3 0055169-79.2017.8.26.0000 19/12/2017 19/12/2017
```

```
##    4 9000580-82.2017.8.26.0032 14/12/2017 19/12/2017
```

```
##    5 0052938-79.2017.8.26.0000 14/12/2017 19/12/2017
```

```
##    6 9000723-79.2017.8.26.0482 14/12/2017 19/12/2017
```

```
##    7 0003276-86.2015.8.26.0075 14/12/2017 19/12/2017
```

```
##    8 9000298-11.2017.8.26.0625 14/12/2017 19/12/2017
```

```
##    9 0004653-39.2015.8.26.0028 14/12/2017 19/12/2017
```

```
##   10 2221930-66.2017.8.26.0000 19/12/2017 19/12/2017
```

```
## # ... with 11,721 more rows
```

- Crie uma coluna binária `drogas` que vale `TRUE` se no texto da decisão algo é falado de drogas e `FALSE` caso contrário. – Dica: `str_detect`

Obs.: Considere tanto a palavra 'droga' como seus sinônimos, ou algum exemplo de droga e retire os casos em que `txt_decisao` é vazio

```
decisoes %>%
  filter(!is.na(txt_decisao)) %>%
  mutate(txt_decisao = tolower(txt_decisao),
         droga = str_detect(txt_decisao,
                             "droga|entorpecente|psicotr[ó]pico|maconha|haxixe|coc
dplyr::select(n_processo, droga)
```

```
## # A tibble: 6,933 x 2
##   n_processo      droga
##   <chr>          <lg1>
## 1 9000580-82.2017.8.26.0032 FALSE
## 2 0052038-70.2017.8.26.0000 FALSE
```

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

summarise

summarise

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Retorna um vetor de tamanho 1 a partir de uma operação com as variáveis (aplicação de uma função).
- Geralmente é utilizado em conjunto com `group_by()`.
- Algumas funções importantes: `n()`, `n_distinct()`.

Em ação

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
decisoes %>%
  select(n_processo, municipio, data_decisao) %>%
  #       pega ano da decisão
  mutate(ano_julgamento = year(dmy(data_decisao)),
         # pega o ano do processo 0057003-20.2017.8.26.000
         ano_proc = str_sub(n_processo, 12, 15),
         # transforma o ano em inteiro
         ano_proc = as.numeric(ano_proc),
         # calcula o tempo em anos
         tempo_anos = ano_julgamento - ano_proc) %>%
  group_by(municipio) %>%
  summarise(n = n(),
            media_anos = mean(tempo_anos),
            min_anos = min(tempo_anos),
            max_anos = max(tempo_anos))
```

Resultado

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

```
## `summarise()` ungrouping output (override with `.groups`
```

```
## # A tibble: 315 x 5
```

```
##      municipio      n media_anos min_anos max_an
```

```
##      <chr>      <int>      <dbl>      <dbl>      <db
```

```
## 1 Adamantina      17      0.765      0
```

```
## 2 Aguaí           19      1.16      0
```

```
## 3 Águas de Lindóia  5      1.4      0
```

```
## 4 Agudos          8      3.25      0
```

```
## 5 Altinópolis      7      0.857      0
```

```
## 6 Americana       56      1.41      0
```

```
## 7 Américo Brasiliense 9      1.56      0
```

```
## 8 Amparo          9      2.11      0
```

```
## 9 Andradina       41      0.707      0
```

```
## 10 Angatuba        4      0.5      0
```

```
## # ... with 305 more rows
```

usando count()

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

A função `count()`, simplifica um `group_by %>% summarise %>% ungroup`:

decisoes %>%

```
count(juiz, sort = TRUE) %>%  
mutate(prop = n / sum(n),  
        prop = scales::percent(prop))
```

```
## # A tibble: 100 x 3
```

##	juiz	n	prop
##	<chr>	<int>	<chr>
##	1 Gilberto Ferreira da Cruz	237	2.0203%
##	2 Francisco Orlando	226	1.9265%
##	3 Diniz Fernando	198	1.6878%
##	4 Walter da Silva	183	1.5600%
##	5 De Paula Santos	182	1.5514%
##	6 Machado de Andrade	182	1.5514%
##	7 Newton Neves	180	1.5344%
##	8 Leme Garcia	179	1.5259%

+ fácil ainda

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

mas sem formato %

```
decisoes %>%
```

```
  count(juiz, sort = TRUE) %>%
```

```
  mutate(prop = prop.table(n))
```

```
## # A tibble: 100 x 3
```

##	juiz	n	prop
##	<chr>	<int>	<dbl>
##	1 Gilberto Ferreira da Cruz	237	0.0202
##	2 Francisco Orlando	226	0.0193
##	3 Diniz Fernando	198	0.0169
##	4 Walter da Silva	183	0.0156
##	5 De Paula Santos	182	0.0155
##	6 Machado de Andrade	182	0.0155
##	7 Newton Neves	180	0.0153
##	8 Leme Garcia	179	0.0153
##	9 Grassi Neto	177	0.0151
##	10 Figueiredo Gonçalves	176	0.0150

O Universo Tidyverse

**Frederico
Bertholini**

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

arrange

arrange

O Universo
Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

- Simplesmente ordena de acordo com as opções.
- Utilizar `desc()` para ordem decrescente ou o sinal de menos (-).

O Universo Tidyverse

Frederico
Bertholini

Abrindo o R e
conhecendo o
software

Vetores

Operações
matemáticas

data.frame

Pacotes

Importação de
dados

O Universo
tidyverse

Importação no
tidyverse

Pacotes dplyr e
tidyr

select

filter

mutate

summarise

■ Quem são os cinco relatores mais prolixos?

– Dica: use `str_length()` – Lembre-se da função `head()`

```
decisoes %>%  
  filter(!is.na(txt_decisao)) %>%  
  mutate(tamanho = str_length(txt_decisao)) %>%  
  group_by(juiz) %>%  
  summarise(n = n(),  
            tamanho_mediana = median(tamanho)) %>%  
  filter(n >= 10) %>%  
  arrange(desc(tamanho_mediana)) %>%  
  head()
```

```
## `summarise()` ungrouping output (override with `.groups`)
```

```
## # A tibble: 6 x 3
```

##	juiz	n	tamanho_mediana
##	<chr>	<int>	<dbl>
## 1	Airton Vieira	154	3146.
## 2	Ely Amioka	81	1847
## 3	Grassi Neto	141	1675
## 4	Alcides Malossi Junior	95	1541
## 5	Guilherme Augusto de Azevedo	77	1244