

The background of the slide is a dark navy blue field densely populated with small, semi-transparent hexagons in various colors including red, orange, yellow, green, blue, and grey. These hexagons are scattered across the entire frame, creating a textured, starry effect.

Introdução ao tidyverse

O que é o tidyverse?

É uma coleção de pacotes do R voltados para a Ciência de Dados, em que cada pacote é especializado para um tipo de tarefa.

MANIPULAR E TRANSFORMAR DADOS



PROGRAMAÇÃO FUNCIONAL

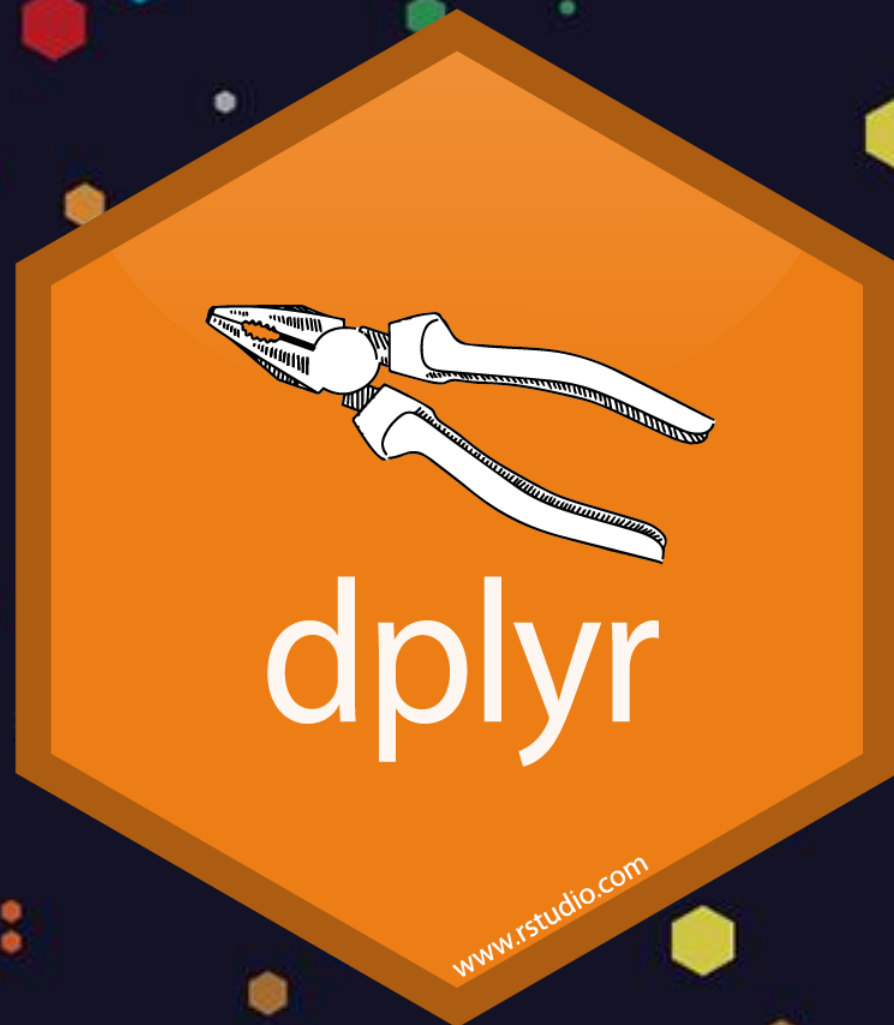


IMPORTAR DADOS



VISUALIZAR





- Dplyr é um pacote do R para manipulação de dados.
- Tem como base poucos verbos que resolvem boa parte dos problemas envolvendo manipulação de dados.

INSTALAÇÃO

Existem duas maneiras:

```
1.install.packages("tidyverse")
```

```
2.install.packages("dplyr")
```

```
require(tidyverse)
```

DADOS

The image shows the classic "Star Wars" logo in a yellow, outlined, blocky font. The word "STAR" is on the top line and "WARS" is on the bottom line, both in all caps. The logo is centered within a black rectangular box.

VERBOS

O **dplyr** é composto por 5 funções básicas que chamamos de verbos:

- **select()** : selecionar colunas
- **filter()** : filtrar registros
- **mutate()** : criar novas variáveis
- **summarise()** : resumir uma série de valores
- **arrange()** : ordenar o "data.frame" de acordo com valores de uma coluna

A função **group_by()** pode ser combinada com essas funções a fim de fazer operações por grupo.

```
> starwars
```

```
# A tibble: 87 x 13
```

	name	height	mass	hair_color	skin_color	eye_color	birth_year	gender	homeworld	species	films
	<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>	<lis>
1	Luke...	172	77	blond	fair	blue	19	male	Tatooine	Human	<chr...
2	C-3PO	167	75	NA	gold	yellow	112	NA	Tatooine	Droid	<chr...
3	R2-D2	96	32	NA	white, bl...	red	33	NA	Naboo	Droid	<chr...
4	Dart...	202	136	none	white	yellow	41.9	male	Tatooine	Human	<chr...
5	Leia...	150	49	brown	light	brown	19	female	Alderaan	Human	<chr...
6	Owen...	178	120	brown, gr...	light	blue	52	male	Tatooine	Human	<chr...
7	Beru...	165	75	brown	light	blue	47	female	Tatooine	Human	<chr...
8	R5-D4	97	32	NA	white, red	red	NA	NA	Tatooine	Droid	<chr...
9	Bigg...	183	84	black	light	brown	24	male	Tatooine	Human	<chr...
10	Obi-...	182	77	auburn, w...	fair	blue-gray	57	male	Stewjon	Human	<chr...

```
# ... with 77 more rows, and 2 more variables: vehicles <list>, starships <list>
```

```
> |
```

select()

Seleciona colunas, podendo ser indicadas pelos nomes ou pelo índice da coluna.

- Algumas maneiras de usar o select:

```
select(.data = starwars, name, height, mass)
```

```
select(starwars, "name", "height", "mass")
```

```
select(starwars, name:mass)
```

```
select(starwars, 1:3)
```

- Selecionar colunas, exceto a coluna *height*:

```
select(starwars, -height)
```


`filter()`

Recebe uma ou mais condições lógicas e retorna as linhas solicitadas. Aqui podemos utilizar mais de uma condição (usando `&` ou `|`).

- Exemplo: Suponha que desejamos filtrar os personagens que não possuem cabelo.

```
filter(.data = starwars, is.na(hair_color))
```

- Suponha que agora desejamos filtrar personagens sem cabelo e que pertencem a espécie Droids:

```
filter(starwars, is.na(hair_color), species == "Droid")
```

The background of the slide is dark blue with numerous small, colorful hexagons in shades of red, orange, yellow, green, and blue scattered across it.

`filter()`

- Agora desejamos filtrar por personagens com mais de 130 kg e selecionar as variáveis name, height, mass e homeworld:

```
select(filter(starwars, mass > 130), name:mass, homeworld)
```

NOTE QUE AGORA, AO ACRESCENTAR MUITAS CONDIÇÕES, A LEITURA DO CÓDIGO COMEÇA A FICAR MAIS COMPLICADA.



OPERADOR PIPE

- Este operador faz o encadeamento das funções do tidyverse.
- Pertence ao pacote [magrittr](#), mas ao carregar p dplyr você já consegue usá-lo.
- O *pipe* irá pegar o resultado da expressão a sua esquerda e colocar como primeiro argumento (por padrão) da expressão da direita.

Voltando ao exemplo anterior em que desejamos filtrar por personagens com mais de 130 kg e selecionar as variáveis name, height, mass e homeworld:

- Sem utilizar o "%>%":

```
select(filter(starwars, mass > 130), name:mass,  
homeworld)
```

- Utilizando o %>%:

```
starwars %>%  
  filter(mass > 130) %>%  
  select(name:mass, homeworld)
```



```
mutate(.data, ...)
```

Podemos criar novas colunas e essas novas colunas podem ser criadas em função das já existentes.

- Exemplo: Desejamos calcular o IMC dos personagens. Mas antes precisamos que a variável *altura esteja* em metros .

```
starwars %>%  
  mutate(altura_metros = height/100,  
         IMC = mass/(altura_metros^2)) %>%  
  select(name, IMC)
```

```
summarise(.data, ...)
```

Permite que usemos funções de sumarização.

- Exemplo: Desejamos saber a maior altura e a massa média.

```
starwars %>%  
  summarise(max_altura = max(height, na.rm = T),  
            massa_media = mean(mass, na.rm = T))
```



```
arrange (.data, ...)
```

Essa função nos permite ordenar o data set, de maneira crescente ou decrescente.

- Exemplo: Repetindo o exemplo anterior do IMC, agora queremos refazer o cálculo do IMC e usar essa coluna para ordenar, em ordem crescente.

```
starwars %>%  
  mutate(altura_metros = height/100,  
         IMC = mass/(altura_metros^2)) %>%  
  select(name, height, mass, IMC) %>%  
  arrange(IMC)
```

arrange()

```
# A tibble: 87 x 4
  name          height  mass  IMC
  <chr>         <int> <dbl> <dbl>
1 Wat Tambor    193    48  12.9
2 Adi Gallia    184    50  14.8
3 Sly Moore     178    48  15.1
4 Roos Tarpals  224    82  16.3
5 Padmé Amidala 165    45  16.5
6 Lama Su      229    88  16.8
7 Jar Jar Binks 196    66  17.2
8 Ayla Secura   178    55  17.4
9 Shaak Ti     178    57  18.0
10 Barriss Offee 166    50  18.1
# ... with 77 more rows
```

group_by()

Realiza operações por grupo.

- Exemplo: Desejamos calcular a maior altura e a massa média, por espécie.

```
starwars %>%  
  group_by(species) %>%  
  summarise(max_altura = max(height, na.rm = T),  
            massa_media = mean(mass, na.rm = T))
```

DADOS ALEATÓRIOS



HOME

ABOUT US

BLOG

CONTACT US

The background of the slide is a dark navy blue, densely populated with small, semi-transparent hexagons in various colors including red, orange, yellow, green, blue, and grey. These hexagons are scattered across the entire frame, creating a textured, starry effect.

tidyverse

Obrigada!!!