PROGRAMAÇÃO PARA CIÊNCIAS SOCIAIS

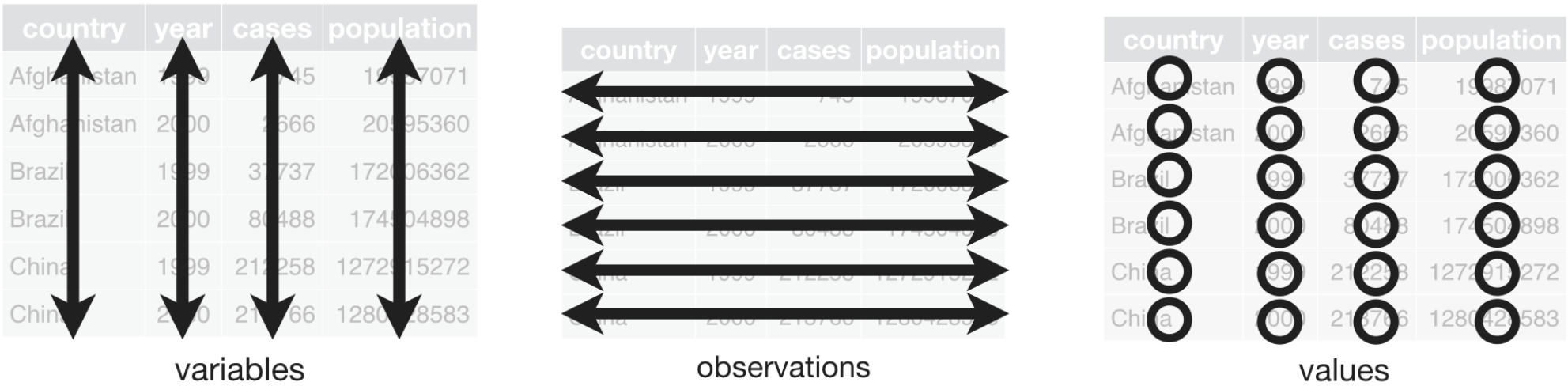
VAMOS CONHECER O TIDYVERSE?

O que é tidyverse?

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

MAS O QUE É TIDY?

HADLEY WICKHAM (2014)



	country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPercap <dbl>
44	Angola	Africa	1987	39.906	7874230	2430.2083
15	Albania	Europe	1962	64.820	1728137	2312.8890
53	Argentina	Americas	1972	67.065	24779799	9443.0385
136	Bolivia	Americas	1967	45.032	4040665	2586.8861
139	Bolivia	Americas	1982	53.859	5642224	3156.5105
41	Angola	Africa	1972	37.928	5894858	5473.2880
91	Bahrain	Asia	1982	69.052	377967	19211.1473
112	Belgium	Europe	1967	70.940	9556500	13149.0412
97	Bangladesh	Asia	1952	37.484	46886859	684.2442
52	Argentina	Americas	1967	65.634	22934225	8052.9530

1-10 of 1,704 rows Previous 1 2 3 4 5 6 ... 100 Next

DATA FRAME TIDY

DATA FRAME NÃO TIDY

variable <fctr>	value <chr>	1952 <int>	1957 <int>	1962 <int>	1967 <int>	1972 <int>	1977 <int>	1982 <int>	1987 <int>
country	Afghanistan	12	12	12	12	12	12	12	12
country	Albania	12	12	12	12	12	12	12	12
country	Algeria	12	12	12	12	12	12	12	12
country	Angola	12	12	12	12	12	12	12	12
country	Argentina	12	12	12	12	12	12	12	12
country	Australia	12	12	12	12	12	12	12	12
country	Austria	12	12	12	12	12	12	12	12
country	Bahrain	12	12	12	12	12	12	12	12
country	Bangladesh	12	12	12	12	12	12	12	12
country	Belgium	12	12	12	12	12	12	12	12

1-10 of 443 rows | 1-10 of 14 columns Previous 1 2 3 4 5 6 ... 45 Next

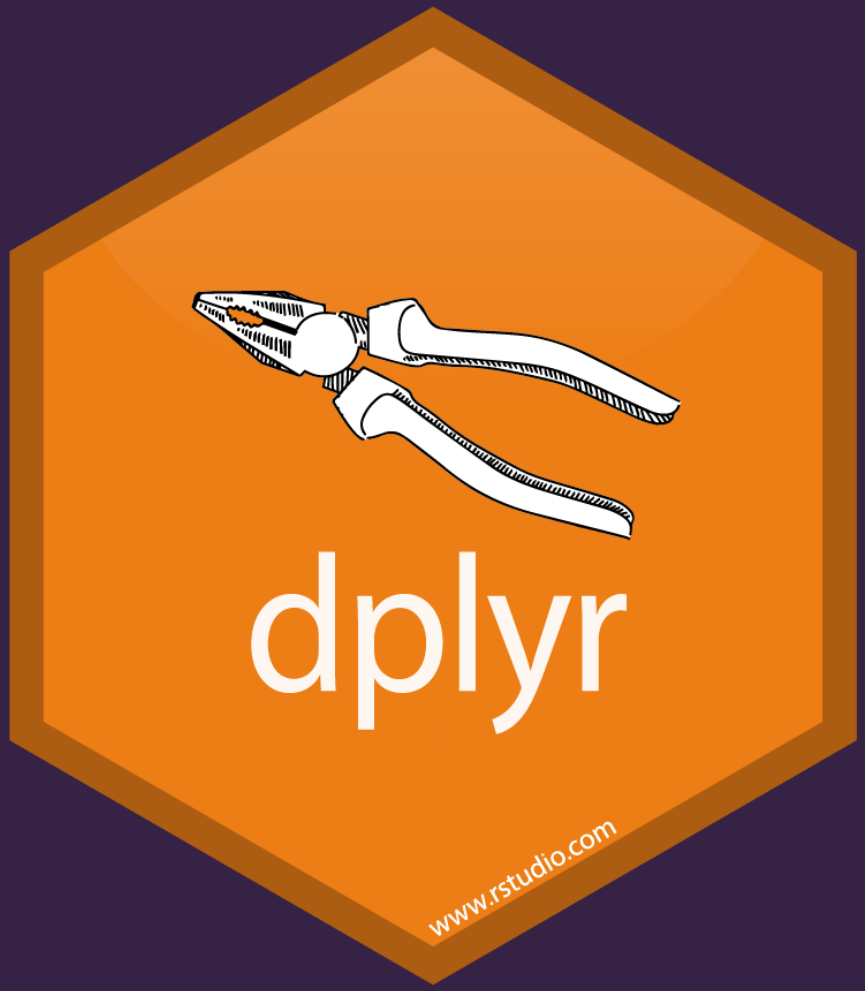
NÚCLEO DO TIDYVERSE

O TIDYVERSE TAMBÉM TEM OUTRAS FERRAMENTAS

Abra o R e digite o código abaixo!

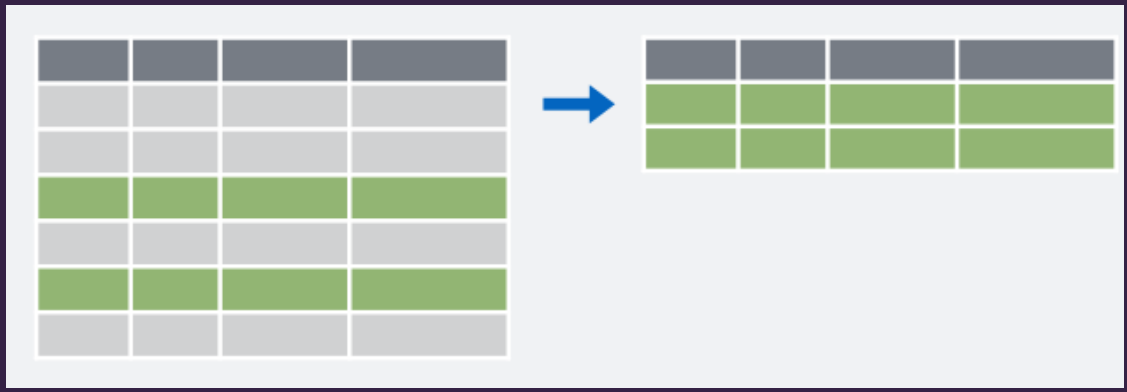
```
install.packages("tidyverse")
```

OS VERBOS DO DPLYR



FILTER()

Utilizado para filtrar casos



```
filter(gapminder, year == 1952)
filter(gapminder, year == 1952 & gdpPercap >= 1000)
filter(gapminder, country %in% c("Austria", "Albania", "Kuwait"))
```

TESTES LÓGICOS

`x < y`	Less than
`x > y`	Greater than
`x == y`	Equal to
`x <= y`	Less than or equal to
`x >= y`	Greater than or equal to
`x != y`	Not equal to
`x %in% y`	Group membership
`is.na(x)`	Is NA
`!is.na(x)`	Is not NA

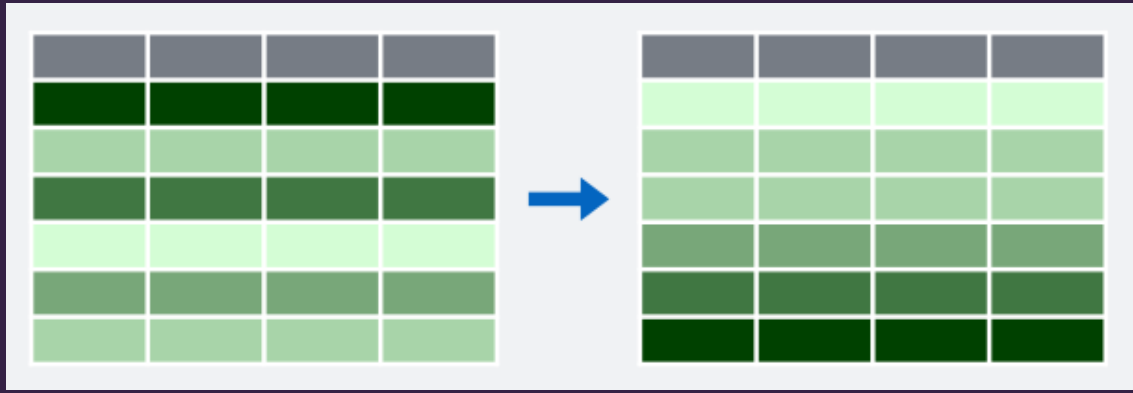
SELECT()

Utilizado para extrair variáveis

```
select(gapminder, year, lifeExp)
select(gapminder, -continent)
select(gapminder, starts_with("c"))
```

ARRANGE()

Utilizado para arranjar variáveis em ordem crescente ou descrescente



```
arrange(gapminder, year)
arrange(gapminder, desc(pop))
arrange(gapminder, year, desc(pop))
```

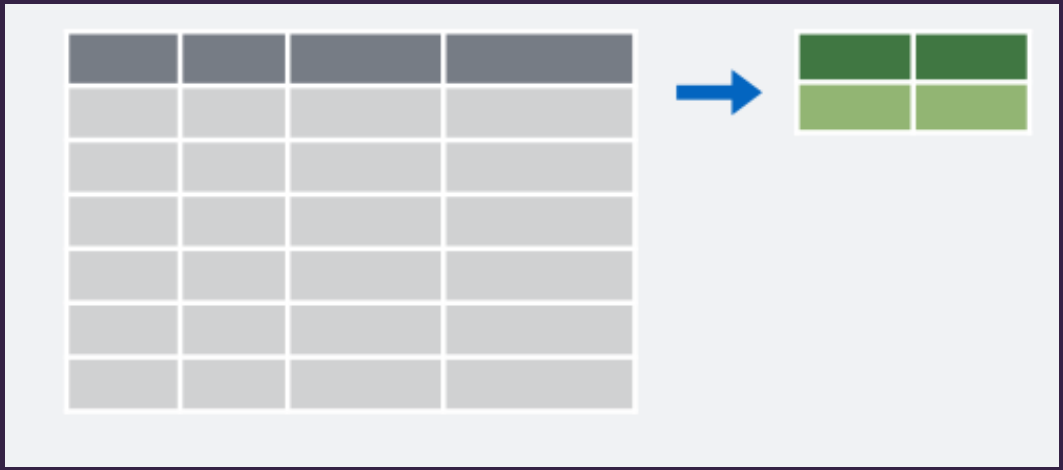
MUTATE()

Utilizado para criar novas variáveis

```
mutate(gapminder, gdp = pop * gdpPercap)
mutate(gapminder, continent = stringr::str_to_upper(continent))
mutate(gapminder, pop_life = pop / lifeExp)
```

GROUP_BY() E SUMMARISE()

Utilizado para sumarizar variáveis



```
gapminder %>%
  group_by(continent) %>%
  summarise(media_lifeExp = mean(lifeExp))
```

pipe operator

```
library(dplyr)
library(datasets)

data("iris")

summarise(
 select(
  mutate(
   filter(
    iris,
    Species == "virginica"),
   Sepal_LW = Sepal.Length/Sepal.Width,
   Petal_LW = Petal.Length/Petal.Width),
  Sepal_LW, Petal_LW),
 mean_Sepal = mean(Sepal_LW),
 mean_Petal = mean(Petal_LW)
 )
```

O que está acontecendo aqui?

summarise(select(mutate(filter(...))))

%>%

```
iris %>%
 filter(Species == "virginica") %>%
 mutate(Sepal_LW = Sepal.Length/Sepal.Width,
        Petal_LW = Petal.Length/Petal.Width) %>%
 select(Sepal_LW, Petal_LW) %>%
 summarise(mean_Sepal = mean(Sepal_LW),
           mean_Petal = mean(Petal_LW))
```