

Exercícios 2

2020-03-29

Pacotes para a vida

Você não pode rodar essas linhas comentadas abaixo dentro de um chunk. Terá de rodá-las dentro do console

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

Com os pacotes instalados, você poderá carregar - los com o library

```
library(tidyverse)  
library(lubridate)
```

Exercícios importação e exportação de dados

- 1- Importe o arquivo Salaries.csv com a função read.csv()
- 2- Importe o mesmo arquivo agora com a função read_csv()
- 3- Agora exporte o arquivo com a função write_csv

Lembre-se de que você deverá dar um nome para o novo arquivo

Dica do capítulo: Sempre de preferência as funções do readr como read_csv e write_csv. Elas são mais simples e rápidas.

Exercícios do dplyr

- 1- Importe o arquivo sales.csv

```
df1 <- read_csv("dados/sales.csv")[,-1]
```

- 1- Visualize - o com a função head()

```
head(df1)
```

```
## # A tibble: 6 x 10  
##   QUANTITYORDERED PRICEEACH ORDERDATE MONTH_ID YEAR_ID PRODUCTLINE CUSTOMERNAME  
##           <dbl>      <dbl> <chr>          <dbl>   <dbl> <chr>          <chr>  
## 1              30       95.7 2/24/200~         2    2003 Motorcycles Land of Toy~  
## 2              34       81.4 5/7/2003~         5    2003 Motorcycles Reims Colle~  
## 3              41       94.7 7/1/2003~         7    2003 Motorcycles Lyon Souven~  
## 4              45       83.3 8/25/200~         8    2003 Motorcycles Toys4GrownU~  
## 5              49      100 10/10/20~        10    2003 Motorcycles Corporate G~  
## 6              36       96.7 10/28/20~        10    2003 Motorcycles Technics St~  
## # ... with 3 more variables: CITY <chr>, STATE <chr>, COUNTRY <chr>
```

3 -

Imagine que eu queira agora uma tabela que me diga quais são os 3 estados dos EUA que tem maior receitas com o produto de motocicletas. Como faríamos? Vou te dar um passo a passo:

A) As colunas que precisaremos selecionar são: COUNTRY,PRODUCTLINE,PRICEEACH,QUANTITYORDERED,STATE

Será que a função select nos ajuda nessa seleção?

```
df2 <- df1 %>% select(COUNTRY,PRODUCTLINE,PRICEEACH,QUANTITYORDERED,STATE)
```

Não se esqueça de olhar sempre como os seus dados ficaram depois da função, use head()

B) Agora precisamos filtrar o país, que no caso é os EUA (Está como USA).

Será que a função filter nos ajuda nessa filtragem?

```
df2 <- df2 %>% filter(COUNTRY == "USA")
```

C) Agora precisamos filtrar o produto, que no caso é a motocicleta (Está como Motorcycles).

```
df2 <- df2 %>% filter(PRODUCTLINE == "Motorcycles")
```

D) Agora precisamos criar uma coluna que nos dê a receita de cada transação, para isso temos que multiplicar o preço pela quantidade vendida.

Será que a função mutate pode nos ajudar?

```
df2 <- df2 %>% mutate(receita = PRICEEACH*QUANTITYORDERED)
```

E) Agora podemos resumir a tabela, na qual queremos agrupar por estado e calcular a receita total de cada um deles.

Será que group_by e summarize nos ajudariam?

```
df2 <- df2 %>% group_by(STATE) %>%  
  summarise(receita_total = sum(receita))
```

F) Agora queremos arranjar as receitas em ordem decrescente,

Será que a função arrange(desc()) nos ajudaria?

```
df2 <- df2 %>% arrange(desc(receita_total))
```

G) Agora queremos pegar somente os 3 maiores,

Será que a função slice nos ajudaria?

```
df2 %>% slice(1:3)
```

```
## # A tibble: 3 x 2  
##   STATE receita_total  
##   <chr>          <dbl>  
## 1 CA             129761.  
## 2 NY              82315.  
## 3 MA              58219.
```

Exercícios do tidyr

1- Quando carregamos o pacote tidyverse, ele carrega um data frame chamado population. De um head() nele e observe - o

2- Veja que ele está no formato long, como poderíamos fazer para transforma - lo no formato wide? Dica: Use `spread()`

3- Agora que está no formato long, como fazemos para transforma -lo no formato wide? Dica: Use a função `gather()`

Exercicios com lubridate

1- Vou criar umas datas para utilizarmos como exemplo, porém elas serão do tipo character ainda

```
d1 <- "20010104"  
d2 <- "01/12/2009"  
d3 <- "12,3,2007"
```

2- Usando as funções que aprendemos, utilize - as para formatar corretamente as datas a cima. Padrão correto: yyyy-mm-dd

Se quisessemos extrair apenas o dia de uma data, como faríamos?

Observe o exemplo abaixo:

```
# Data  
d5 <- "20090202"  
  
# Transformando em date  
d5 <- ymd(d5)  
  
# Extraindo o dia  
day(d5)
```

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

Para extrair dia,mês e ano. Usando as funções `day()`,`month()` e `year()` respectivamente.

Faça isso para uma das datas que você transformou no primeiro exercicio

```
# dia  
  
# mês  
  
# ano
```