

Exercício Fixação manipulação de dados com dplyr

Marcelo Prudente

17 de setembro de 2018

Exercício 1: JOIN (merge)

- **Exercício 1:** o banco *dados_sociais* apresenta as Unidades da Federação, mas não especifica os nomes dos Estados ou mesmo as regiões a que pertencem esse estados. De acordo com os comando aprendidos, tente:
 1. Ler arquivo “uf.csv” no **R**. Salvar em um objeto com nome **uf**. Depois disso, encontre:
 - Qual a variável em comum entre os dois arquivos (*dados_sociais* e *uf*)?
 - Tente mesclar os dados de tal forma que todas observa??es do banco dados sociais contenham os nomes dos estados e suas respectivas regi?es. Crie um objeto chamado *dados_sociais_uf*.

Exercício 2: JOIN (merge)

- **Exercício 2:** O banco *dados_sociais* não traz a população Estadual para cada período. Assim, tente:
 1. Ler o arquivo “uf2.csv” no **R**.
 - Identifique as chaves de cruzamento.
 - Tente mesclar os dados de tal forma que todas observações do banco dados sociais contenham os nomes dos estados e suas respectivas populações em cada um dos períodos.

Exercício 3: comandos *dplyr* com base dados FIES

Baixar base do FIES

1. Selecionar diretório de trabalho: pasta dados.
2. Baixar dados do FIES na pasta dados.
3. carregar pacote *tidyverse* ou *dplyr*.

Observar dicionário de dados da base FIES

Em todas as análises das bases de dados, é importante observar o dicionário de dados disponibilizado pela instituição que produziu ou compilou a informação. Em geral, os dicionários identificam as informações de cada coluna como a descrição das variáveis, a classe (numérica, character, entre outras) e o comprimento de cada variável. Entretanto, no caso do FIES, o FNDE apenas fornece a descrição das variáveis.

Os dados originais do FIES para cada ano estão disponíveis no site dados abertos do FNDE, [aqui](#). O dicionário de dados pode ser encontrado **neste link**.

Exercícios de fixação dos comando básicos do *dplyr*:

1. Verifique a classe e a estrutura do objeto *fies*
2. Descubra o nome das colunas do banco FIES
3. Selecione variáveis que começam com o nome “DS”
4. Selecione variáveis que contém “CURSO” em seu nome
5. Selecione variáveis que terminam com “O”
6. Selecione algumas variáveis de interesse

7. Retire algumas variáveis do banco
8. Reordenar as variáveis
9. Renomear variável:
 - Sigla da UF como uf;
 - Valor da mensalidade como mens
 - Descrição do curso como curso
10. Filtrar bolsas do FIES do Distrito Federal:
 - quantas observações permanecem?

```
# dica
nrow()
```

11. filtrar bolsas do FIES de Alagoas e Sergipe - utilizar operadores lógicos
 - quantas observações permanecem?
12. filtrar bolsas do FIES de Alagoas e Sergipe - utilizar operador %in%
 - quantas observações permanecem?
13. filtrar bolsas do FIES de Alagoas e Sergipe para o curso de medicina
 - quantas observações permanecem?
14. filtrar bolsas do FIES de Alagoas e Sergipe para o curso de medicina e direito
 - quantas observações permanecem?
15. filtrar bolsas do fies cujo valor exceda R\$ 1.000,00
 - quantas observações permanecem?
16. filtrar bolsas do fies que não sejam do Estado de São Paulo ou Minas e com valores de mensalidade menores ou iguais que R\$ 600
 - quantas observações permanecem?
17. filtrar bolsas do fies cujos nomes dos Estados contenham a letra “S”

```
# Dica:
iris # banco de dados do sistema

filter(iris, grepl("set", Species))

filter(iris, grepl("v", Species))
```

18. ordenar os dados pelo valor da mensalidade (da maior para a menor e da menor para maior)
 - Qual a diferença?
19. Encontrar a média, o máximo e o mínimo da mensalidade por Estado.
20. Encontrar a média, o máximo e o mínimo da mensalidade por região
 - **Dica:** utilizar o comando join() para cruzar o banco fies com o banco uf.
21. Encontrar o valor total pago a título de mensalidade a cada IES

Exercício 4: JOIN (merge)

- **Exercício 4:** Imagine que você está diante de duas bases de dados administrativas. Na primeira, há o número do NIS de um cidadão. Na segunda, há o número do PIS.
 1. Ler os arquivos “nis_exemplo.csv” e “pis_exemplo.csv” no **R**. Certifique-se que todas as variáveis foram importadas corretamente.
 2. Identifique as chaves de cruzamento.
 3. Tente mesclar os dados de tal forma que o resultado identifique:
 - todos os caso coincidentes.
 - apenas os casos em que a base *nis* cruza com a base *pis*.
 - todos os casos não coincidentes.
 - todos os casos.
- Salve cada uma das consultas em objetos distintos.

Exercício 5: extrair contratos repetidos

- Com a base do FIES, encontre:
 1. há valores duplicados de contratos nesse banco?
 - Identifique o número de observações duplicadas e únicas.
 2. remover valores duplicados baseados em no código do contrato
 3. remover valores duplicados baseados em múltiplas variáveis: código do contrato e número do semestre
 4. encontrar a média, o máximo e o mínimo da mensalidade e da quantidade de semestres financiados por UF e por Região.
 5. Obter a média, mediana, máximo, mínimo e desvio padrão de todas as variáveis numéricas do banco fies

```
# Dica:  
library(ggplot2)  
mpg <- mpg  
summarise_if(mpg, is.numeric, funs(sum, sd))
```

6. transformar todas as variáveis de character em factor.

```
mpg  
mutate_if(mpg, is.integer, as.double)
```

7. obter os níveis de todas as variáveis que são fatores