

ME115 - Linguagem R

Atividade Prática 08

1º semestre de 2023

Introdução

Nessa atividade, exploraremos os seguintes tópicos:

1. Aplicaremos a ideia: dado original \rightarrow seleção \rightarrow filtro (*pipe* ou `%>%`);
2. Principais verbos do pacote `dplyr`: `select()`, `filter()`, `arrange()`, `mutate()`, `summarize()`, e `group_by()`.

Antes de iniciar a atividade instale, se necessário, e carregue os pacotes `tidyverse` e `dslabs`. Note que ao carregar o `tidyverse`, vários pacotes são carregados, incluindo o `dplyr`. Veja quais são os demais.

```
library(tidyverse)
library(dslabs)
```

Nessa atividade, iremos trabalhar com o conjunto de dados `murders` do pacote `dslabs`. Carregue o conjunto de dados e use a função `glimpse()` do pacote `dplyr` para olhar sua estrutura. Compare com a função `str()` da base do R.

Atividade

1. Usando a função `mutate()`, adicione uma nova coluna chamada `rate` aos dados `murders` do pacote `dslabs`, dada por `rate = total/ population * 100000`.
2. A partir do conjunto de dados `murders` aumentado em (1), selecione as variáveis `state`, `region`, `rate` e os registros onde a taxa (`rate`) é maior que 0.6 usando as funções `select()` e `filter()`. Escreva seu código do modo tradicional, ou seja, sem usar o operador `%>%` (pipe).
3. A partir do conjunto de dados `murders` aumentado em (1), selecione as variáveis `state`, `region`, `rate` e os registros onde a taxa (`rate`) é maior que 0.6 usando as funções `select()` e `filter()`, agora com o operador `%>%`. Observe as diferenças entre os dois códigos.
4. Crie uma coluna chamada `rank` nos dados `murders` contendo o posto em ordem decrescente do estado de acordo com a taxa de assassinatos. Dica: `rank()`.
5. Calcule a média e o desvio padrão da taxa de assassinatos segundo a região e guarde o resultado no objeto `murder.by.region`. Qual a região mais segura? Dica: `group_by()` e `summarize()`.
6. Ordene o objeto `murder.by.region` em ordem decrescente de taxa de assassinato média.
7. Calcule quantidade de estados da região Sul (*South*) com taxa de assassinatos menor do que a média de assassinatos da mesma região.
8. Calcule a proporção de estados para cada região com taxa de assassinatos menor do que a média de assassinatos da respectiva região. Qual a região mais segura?
9. Crie uma nova coluna chamada `rank10` em `murders` usando `mutate()` tal, baseado na coluna `rank` criada em (4), ela seja 1 se o estado foi rankeado abaixo de 10 e 0 caso contrário. A seguir faça uma tabela classificando os estados abaixo da décima posição, por região. Qual região é mais segura?

Agradecimento

O material foi produzido pela Profa. Tatiana Benaglia para o curso de ME115.