

01

Relacionando tabelas parte 1

Transcrição

[00:00] Agora, a gente vai entender aqui no R, na nossa análise, como é que nós fazemos o relacionamento entre diferentes tipos de tabelas, diferentes tabelas. Dificilmente numa análise de dados, os dados vão estar numa única tabela, é muito comum executar esse relacionamento.

[00:20] Então, por exemplo, eu tenho uma tabela de cargos, quero relacionar com salários, despesa, eu quero relacionar com contas e assim por diante, então, para exemplificar esse processo, nós vamos criar algumas tabelas, duas tabelas bem simples.

[00:39] Então, as tabelas que a gente vai estar criando aqui, a gente vai chamar uma tabela dentro do R, chamada cargo, aonde nessa tabela, a gente vai criar... vai ter uma coluna chamada nome, que são os funcionários, uma outra coluna chamada cargo, onde tem o cargo desses funcionários.

[00:57] Então, essa primeira tabela e uma outra tabela, uma segunda tabela chamada local, aonde nós teremos o nome do funcionário e o local aonde trabalha. Então, a gente vai primeiramente fazer a criação dessas tabelas, para depois eu mostrar como é que vai funcionar esses relacionamentos.

[01:15] Então... Ah, antes de mais nada, também para nós executarmos o... vou abrir um chunk aqui, “Ctrl + Alt + I”. Então, vou... a gente vai precisar o quê? Carregar o pacote que vai ser necessário para fazer essa análise no R, então, carregar o pacote tidy.

[01:47] O pacote tidyverse, ele vai ser necessário para a gente executar depois esses relacionamentos entre tabelas. Então, a primeira coisa que a gente precisa fazer, só para lembrar, a gente coloca library e o respectivo pacote que é o tidyverse. Vou executar. Então, ele já está carregado.

[02:15] Caso vocês não tenham essa biblioteca instalada, lembrando que é só dar o install.packages e depois carrega a biblioteca. Então, na sequência aqui, vamos criar a primeira tabela. Então, vou deixar anotado aqui, criando as tabelas cargo e local. Para criar esses data frame, essas tabelas no R, como é que a gente faz?

[02:46] Eu vou criar aqui a tabela chamada cargo, ele vai ser um data frame, então data.frame, que é uma tabela simples e aí, a gente vai colocar o seguinte, vai criar um vetor aqui dentro, uma lista, como se fosse uma lista. Então, lembrando, eu já mostrei como é que a gente cria essa lista.

[03:16] Então, é só para a gente colocar aqui. Então, aqui eu vou colocar o André, fecha, vírgula, eu vou colocar o Marcos, tudo entre aspas também, abre aspas, Fernanda, vírgula, a Julia, abre e fecha aspas, Julia, colocar maiúscula sempre aqui, Julia.

[03:51] E aí, a gente fecha, vamos criar esses quatros, lembrando que a gente está criando exatamente: André, Marcos, Fernanda e Julia, primeiro cargo, então a primeira ponta vai ter isso. Aí, eu tenho agora uma segunda coluna, uma segunda lista, que eu vou estar criando aqui também, que são os cargos que ocupam.

[04:43] Então, primeiro cargo aqui: analista, o segundo aqui, também analista, terceiro, coordenador e o quarto, abre e fecha aspas, é uma gerente. Então, eu criei aqui essa primeira tabela. Aí, a gente precisa fazer... dar nome para esses campos. Então, o comando colnames, eu vou colocar aqui, aplicar na tabela cargos.

[04:58] Ah, vou colocar primeiro maiúsculo aqui, só para padronizar, Cargo e aí, vai receber o quê? Vai receber esses dois campos. Opa, abre e fecha aspas e aí, o primeiro campo é nome e o segundo campo que eu estou atribuindo aí é o cargo... Opa, abre e fecha aspas, cargo.

[05:26] Então aqui, eu estou criando a minha primeira tabela, vou colocar... colocando aqui, vou colocar (entre aqui), essa aqui é a primeira tabela, que a gente está chamando aí, está voltando lá de cargo. Eu vou executar. Opa, deu um erro aqui. Ah, está falando colname, cargo.

[05:53] Deixa eu executar de novo aqui por partes e aqui, colname, cargo, nome. Ah, o que que eu esqueci? Aqui também é um vetorzinho de lista, é uma lista. Aí, agora sim e aí, eu vou chamar a tabela cargo aqui já, para a gente visualizar. Cargo, "C" maiúsculo, vou executar.

[06:16] Então, eu criei essa primeira tabela, nome da pessoa, nome do funcionário, André, Marcos, Fernanda, Julia e o cargo, analista, analista, coordenador e gerente. Criamos essa primeira tabela aqui. Vamos criar a segunda tabela, "Ctrl + Alt + I", eu vou criar.

[06:37] Então, vou colocar aqui: entre a segunda tabela, que a segunda tabela o local, a gente está chamando de local. Então, o que a gente faz? Abre... chama local, a gente vai montar um data frame, data.frame e aí, a gente monta de novo aqueles vetores.

[07:09] Vetor lá. Esse primeiro vetor que é a lista, eu vou colocar aqui, vai ser... vai ter o funcionário Fabio, estou montando essa segunda, vai ter o funcionário Fabio, Marcos, Fernanda e Julia. Então, botei o Fabio, o Marcos, Fernanda e a Julia, sempre entre aspas, vírgula e a Julia.

[07:46] Então, primeiro vetor de lista. Esse primeiro campo aqui vai ser o nome e o segundo vai ser o local. Agora, eu vou colocar aqui, é um outro vetor que eu estou criando, que vai ter o local. O primeiro local aqui é comercial, área comercial, área, local, área, colocar onde a pessoa trabalha.

[08:08] Então, botei aqui comerciar, botei CRM, eu vou ter TI. Opa, abre e fecha aspas, TI e vou ter RH, abre e fecha aspas, RH. Criei o meu local. Atribuo o colnames, para dizer o nome... para criar as colunas, o nome das colunas na tabela que eu acabei de criar local e atribuo os respectivos... colunas, as linhas, o nome das colunas aqui entre aspas.

[08:42] Primeiro é nome e o segundo campo é local, segundo é local. Pronto, aí, eu vou chamar aqui a tabela, Local para imprimir e mostrar como é que ela foi feita. Então, criei essa segunda tabela, aonde eu tenho o nome dos funcionários e o local onde trabalham. Então, o próximo passo, o que que a gente quer fazer agora?

[09:21] A gente quer relacionar essas tabelas, de modo que eu quero, num primeiro relacionamento, saber, por exemplo, eu tenho duas tabelas, tanto cargo e local, eu quero saber... quero manter a primeira tabela, a tabela da esquerda, onde eu vou trazer o nome e o cargo e quero saber qual o local, aonde esses funcionário trabalham.

[09:48] Então, o primeiro relacionamento aqui é chamado de left join, eu mantendo a tabela da esquerda como referência, eu vou manter a tabela da esquerda e vou na segunda tabela que é local e trago as... aonde tem o relacionamento. O resultado desse relacionamento, vai ser o seguinte, vai ser isso aqui.

[10:09] Então, o resultado desse relacionamento vai... Eu mantive a tabela aqui de cargo e trouxe os resultados que eu consegui encontrar na tabela da direita, local, ou seja, então, por exemplo, o funcionário Marcos, eu sei que ele é analista e trabalha na área de CRM, ok, ele trouxe.

[10:32] A funcionária Fernanda é coordenadora, função de coordenadora e trabalha aonde? Ela trabalha como TI e a Julia, eu sei que é gerente e trabalha o RH. O André, ele não está nessa tabela local, ou seja, então ele vai aparecer o NA aqui, que é o valor ausente.

[10:50] Então, a gente vai construir aqui esse comando no R, exatamente com esse comando aqui, usando o comando left join, vamos fazer esse relacionamento e ver como é que fica. Então, “Ctrl + Alt + I”, aqui eu vou estar utilizando então esse comando, primeiro comando de relacionamento de left join, que eu faço relacionamento, mantendo como referência a minha esquerda.

[11:17] Então, a gente vai relacionar aqui a primeira tabela Cargo, Cargo em maiúsculo, tem que ser exatamente igual, deixa eu abrir aqui para ficar mais fácil. Então, aqui a gente vai colocar o pipe, lembrando como é que a gente coloca o pipe, “Ctrl + Shift + M”. Então, eu já sei que eu estou trabalhando com a minha tabela cargo.

[11:47] E aí, atribuo o meu comando `left_join`, já tem aqui, relacionando com a tabela local, que ela está aqui. Eu preciso colocar uma chave, by, quer seria a chave para relacionar essas duas tabelas, que é o quê? É o nome, coluna nome. Opa, aqui é minúsculo. Então, eu vou pegar e vou relacionar, vou executar aqui esse comando.

[12:23] Opa. Então, olha, vamos ver, o que que ele fez? Ah, ele apareceu que o André não encontrou, o Fernando, ele não encontrou, por que que ele não está encontrando? Vamos dar uma olhada aqui. Ah, já sei porquê que ele não encontrou, porque, o que que acontece?

[12:36] Quando a gente criou a nossa tabela aqui, uma tabela a gente colocou Fernanda e na outra, a gente colocou Fernando. Então, vamos corrigir isso aqui, porque deveria aparecer. Então, eu ou colocar aqui, só corrigir a minha tabela “A”, de cargo, Fernanda. Então aqui é Fernanda também.

[12:55] Vou executar de novo aqui essa tabela, ok, Fernanda, coordenador e aí, eu vou trazer, vou executar novamente o resultado aqui do nosso left join. Pronto, agora sim, agora ficou exatamente igual ao que tinha mostrado aqui no PPT. Então, aqui eu não encontro na tabela, quem? O André, por que que eu não encontro o André?

[13:21] Porque ele não está nessa tabela local aqui, que a gente criou. Então, ele vai trazer esse valor ausente. Então, esse é o primeiro relacionamento que a gente tem para mostrar, que é feito, exatamente, pela esquerda, left join. Então, por enquanto é isso aí, que eu vou mostrar para vocês.