

## PERGUNTAS REALIZADAS DURANTE A AULA E RESPONDIDAS POR E-MAIL

**Curso:** MBA em Data Science e Analytics

**Disciplina:** Data Wrangling II

**Data:** 16/11/2021

**Renata Pamella Costa Teixeira**

Ainda não absorve 100%. O objetivo dos dois é trocar nomes, pq em um usamos mutate e outro recode? `rename(grupo_renda="Income group") %>% mutate(grupo_renda = recode(grupo_renda,`

Renata, em apertada síntese:

mutate: adiciona novas variáveis e preserva as existentes; e

recode: substitui valores numéricos com base em sua posição ou nome.

**Ianca Almeida**

Por favor. Por que no comando `-> select(everything(), -`Country Code`)` %>% o Country Code está entre aspas simples? E não fica verdinho?

Ianca, está indicando que estamos tratando de um texto com um espaço. Pedimos que verifique as aulas introdutórias de estatística e de programação, por gentileza. Caso as dúvidas persistam, peço que nos acione no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Cleverson Henrique De Freitas**

`bind_cols` e `bind_rows` seria semelhante aos `'cbind'` e `'rbind'`?

Cleverson, são semelhantes. As funções `rbind()` e `cbind()` possibilitam a criação de matrizes com a união de linhas (`row`) e colunas (`column`) respectivamente.

**Carlos Eduardo Rocha de Queiroz**

`Inner` = semi??? qual a diferença na prática?

Carlos, em apertada síntese:

`inner`: cria um novo dataset com as observações que estão em X e Y, ou seja, a intersecção de dois conjuntos.

`semi`: mantém em X as observações que coincidem com Y. No entanto, a função `semi` não realiza o merge. Ou seja, a função de `semi` é apenas comparar datasets

**Aquila José Gonçalves Delfino**

E como seria a estrutura da segunda função se eu usar o pipe? `>>> Join(dataset_nicial, dataset_merge, by="var1") %>%` e a outra função eu colocaria oq?

Aquila, deveríamos, nesse caso, extrair o primeiro argumento da função, conforme linha 447 do script.

**Vanessa Zeni da Silva**

O seminário join não é um join?

Vanessa, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

### Vitor Zambon De Souza

Professor, o que acontece quando a chave se repete em alguma base? irá duplica as linhas? Vitor, é uma boa pergunta e um excelente exercício para se tentar realizar. As boas soluções em programação partem de ideias e, algumas vezes, de tentativa e erro. Recomendamos que faça esse exercício, transformando sua dúvida em uma programação e testando. Acreditamos que irá agregar muito ao conhecimento a elaboração de rotinas para testar as soluções pensadas. Caso tenha dúvidas na execução, nos acione no e-mail da monitoria. Iremos prontamente lhe auxiliar.

### Gabriel Toledo Peregrino

Nessa correção do "Breaking Bad", como ficaria apenas 1 linha usando pipe? Gabriel, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

### Sidnei Augusto Tavares

Error: Problem with `mutate()` column `melhoresTOTTEN`. i `melhoresTOTTEN = cut(...)` x lengths of 'breaks' and 'labels' differ Run `rlang::last\_error()` to see where the error occurred.

Sidnei, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

### Ulisses Teixeira Neto

Como usar o slice sem utilizar a proporção? Por exemplo trazer a maior distância Ulisses, posicione o cursor no nome do argumento ou função em que tem dúvidas, aperte a tecla F1, aparecerá uma documentação referente ao argumento ou função, com um informativo detalhado sobre os objetivos, argumentos e características de cada função. Caso ainda assim as dúvidas persistam, nos acione no e-mail da monitoria. Teremos prazer em auxiliar.

### Débora Fernandes Garcia

não entendi muito bem a função probs

Débora, posicione o cursor no nome do argumento ou função em que tem dúvidas, aperte a tecla F1, aparecerá uma documentação referente ao argumento ou função, com um informativo detalhado sobre os objetivos, argumentos e características de cada função. Caso ainda assim as dúvidas persistam, nos acione no e-mail da monitoria. Teremos prazer em auxiliar.

### Thiago Molina Capelão

Dá p usar argumentos lógicos com slice? Retorna algum erro?

Thiago, posicione o cursor no nome do argumento ou função em que tem dúvidas, aperte a tecla F1, aparecerá uma documentação referente ao argumento ou função, com um informativo detalhado sobre os objetivos, argumentos e características de cada função. Caso ainda assim as dúvidas persistam, nos acione no e-mail da monitoria. Teremos prazer em auxiliar.

### João Carlos Simonetti

Professor, por que é feito o filtro depois do agrupamento (linhas 402 à 404 no script 1.1), e em seguida os dados são desagrupados?

João, os agrupamentos ficam salvos na memória do R e os cálculos (a partir do agrupamento) são feitos a partir dos dados agrupados. Por isso, ao final da análise que pretendíamos realizar com os dados agrupados, precisamos desagrupar esses dados para seguir com a análise. Na maioria das vezes criamos um objeto com os dados agrupados, para nossa visualização da análise, para posteriormente procedermos com o desagrupamento.

**Ianca Almeida**

Por favor! Não estou muito certa dessa pergunta, mas quando eu uso ( ) e quando eu uso [ ]? Ianca, pedimos que assista às aulas introdutórias de programação, bem como as aulas introdutórias de estatística. Caso ainda persista a dúvida, pedimos que, por gentileza, nos acione no e-mail da monitoria, iremos prontamente lhe auxiliar.

**Jeferson Galvão Trindade**

Em termos de performance, o filtro por slicing tem a mesma eficiência que o filter, ou há maior consumo de recursos em uma delas?

Jeferson, são funções com objetivos parecidos e que necessitam de recursos computacionais parecidos. Não se tratam de funções muito pesadas.

**Adalberto de Souza**

na função between os extremo estão contidos no intervalo?

Sim, Adalberto.

**Matheus Augusto Cândido**

Se eu não usasse o na.rm, ainda precisaria do droplevels?

Sim, Matheus.

**Luis Abel da Silva Filho**

Está dando erro aqui: completo - completo %>% mutate(IMDB = ajuste\_imdb, + ROTTEN = ajuste\_rotten) Error: Problem with `mutate()` column `IMDB`. i `IMDB = ajuste\_imdb`

Luis, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

**Nilton Sergio Novack Junior**

completo - completo %>% mutate(IMDB = ajuste\_imdb, ROTTEN = ajuste\_rotten) .. e se no meio do caminho o db mudar a ordem das observações? o IMDB não está referenciado?

Nilton, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Andrew Kazuo da Silva**

Existe algum comando que identifica o código automaticamente?

Andrew, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Gabriel Henrique Mageste Costa**

Como unir as bases e criar uma nova "coluna" com o nome da base. Ex: adicionar um novo campo indicando se é série ou filme?

Poderia utilizar uma função para criação de variáveis (por exemplo: mutate), adicionando níveis para cada uma das ocorrências (filmes e séries).

**Patricia Egri Wissinievski**

na junção dos arquivos filmes e series, se excluimos a coluna X de Series, como ele gerou o conteúdo do X ao importar as linhas no COMPLETO?

Patrícia, não entendemos especificamente de qual linha do script está falando. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Nilton Sergio Novack Junior**

existe função para validar os nomes das colunas de dois bancos de dados antes de executar o bind\_rows (pensando em db com muitas variáveis..)?

Nilton, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Julia Rodrigues de Sousa**

Quando faz o blind.rows já exclui o cabeçalho da planilha que foi posta em baixo?

Julia, sim. O comando junta as bases com essa junção já feita de modo estruturado.

**Joao Vitor do Carmo Marques Dias**

Não está baixando o banco de dados de Series, aparece que não tem o arquivo ou o diretório

João, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

**Paulo Ricardo da Silva Xavier**

Dado incompatível (texto) e NULL são desconsiderados ao utilizar na.rm = T na funcao?

Paulo, a função na.rm remove os dados considerados como missing. Quando da importação da base, temos a opção de informar os missing da base (por exemplo células vazias, mas poderia ser qualquer valor que informássemos ao R), são esses dados que a função na.rm irá retirar.

**Carolline Leite Lima Nascimento**

o ungroup na função desagrupa apenas a base original? o novo objeto salvo permanece agrupado?

O ungroup desagrupa somente a base que foi agrupada, Caroline.

**Marco Gomes Baptista**

Se o pipe faz na sequencia em está escrito, então o ungroup() após o group() não anularia o agrupamento? Linha 95 do exercício 2

Exatamente. Marco, os agrupamentos ficam salvos na memória do R e os cálculos (a partir do agrupamento) são feitos a partir dos dados agrupados. Por isso, ao final da análise que

pretendíamos realizar com os dados agrupados, precisamos desagrupar esses dados para seguir com a análise. Na maioria das vezes criamos um objeto com os dados agrupados, para nossa visualização da análise, para posteriormente procedermos com o desagrupamento.

**Tawane Yara Nunes**

Professor, o `na.rm()` considera as observações removidas pelo `na.rm`?

Sim, Tawane.

**Felipe Oliveira Rodrigues**

Professor, pq na base covid tivemos que desagrupar , sendo que a base covid em si não havia sido agrupada??

Felipe, não entendemos especificamente de qual linha do script está falando. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar. No exemplo dado pelo professor, provavelmente o procedimento está sendo feito para fins didáticos.

**Uilson Vicente Bonfim Júnior**

Na linha 95, você desagrupou para poder utilizar a `base_covid` sem correr o risco de trabalhar com ela agrupada. A `base_covid` é salva no modo agrupada ainda que tenha sido usada dentro de uma função?

Uilson, caso seja efetuada uma programação para criar ou salvar a nova base, sim. No caso foi criado um objeto com a nova base.

**Patrícia Sales Rocha Pombo**

Esses dois últimos comandos que ele pediu para darmos, deu erro. `Base_covid` e `Base_quartis`. O que pode ser? Error: Must group by variables found in ``data``. \* Column ``grupos`` is not found. Run ``rlang``:

Patrícia, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

**Thiago Molina Capelão**

pq nessa dos países da 2.1, a tabela não teve a contagem de países refeita? na aula passada você excluiu as linhas "global" e "outros".

Thiago, as operações são feitas na memória do R. A base de dados em Excel (ou qualquer outro formato) permanece a mesma, a não ser que efetuemos algum comando para escrever novamente a base de dados no diretório com alguma função de algum pacote específico.

**Liège Pilling Baptista da Costa**

O `group_by` é comumente usado com outras funções além do `summarise`?

Sim, Liège. Pode ser usado inclusive para a construção de plots específicos para grupos de observações, ou ainda para a construção de modelos estatísticos para grupos específicos.

**Marcelo Viana de Aguiar**

Podemos criar combinações que não tem dados ou o `summarise` com `group_by` só retorna valores para as combinações que possuem informações no banco de dados?

Marcelo, estamos agrupando informações existentes na base de dados. Para a criação ou junção de observações ou variáveis fora do banco de dados temos outras ferramentas.

**Rogério Yokomizo**

na linha 342 como foi definido o grupo a partir dos dois períodos?

Por intermédio do comando `presnete` na linha 333, Rogério:

```
base_grupo <- group_by(nova_base, periodo)
```

Agrupamos por período (manhã, tarde).

**Fabiola Verissimo Da Silva Pereira**

Se fizermos um `groupby()` por perfil, período, ele faz o `summarize` nesta ordem? Existe relevância na ordem dos critérios?

Fabiola, é uma boa pergunta e um excelente exercício para se tentar realizar. As boas soluções em programação partem de ideias e, algumas vezes, de tentativa e erro.

Recomendamos que faça esse exercício, transformando sua dúvida em uma programação e testando. Acreditamos que irá agregar muito ao conhecimento a elaboração de rotinas para testar as soluções pensadas. Caso tenha problemas na execução, peço que nos acione pelo e-mail da monitoria.

**Melissa Siqueira Leonardo Nogueira Cruz**

Porque no comando da linha 358, onde se agrupa por período e perfil a coluna observações, nome dos alunos, foi ignorada?

Melissa, nessa parte do script o professor está solicitando estatísticas descritivas, agrupando por meio de duas características (período e perfil), não está utilizando as demais variáveis presentes no dataset.

**Renata Pamella Costa Teixeira**

Como ficaria o código sem o `base_sem_grupo` ?

Renata, não entendemos a sua pergunta, não sabemos o momento específico da aula ao qual está se referindo. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Thiago Sartori**

O `glimpse` mostra que o dataset está agrupado, mas somente mostrou o "período". Qual função mostra o `group by` completo?

Thiago, não entendemos a sua pergunta. Caso queira, peço que, por gentileza, nos encaminhe sua dúvida de modo um pouco mais detalhado no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Guilherme Dan Yamada**

Qual comando pode ser usado para identificar (via código) se um dataset está agrupado ou não?

Guilherme, não sei se entendi a sua pergunta. Vou tentar responder e, caso não seja o que perguntou, nos acione no e-mail da monitoria, por gentileza. Caso queira testar se o objeto que está trabalhando está agrupado, bastaria clicar no objeto no Global Environment, esta informação se encontra descrita nas características gerais da base.

**Victor Cañete De Mendonça**

Não localizei onde eu vejo se a base está agrupada ou não, onde fica?

Victor, não sei se entendi a sua pergunta. Vou tentar responder e, caso não seja o que perguntou, nos acione no e-mail da monitoria, por gentileza. Caso queria testar se o objeto que está trabalhando está agrupado, bastaria clicar no objeto no Global Enviroment, esta informação se encontra descrita nas características gerais da base.

**Fernando Pepeu Jardim Braga**

não consigo usar a função relocate. o rstudio diz que não acha a mesma, mesmo com o tidyverse instalado e o dplyr tb, mesmo tentando com dplyr::relocate

Fernando, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

**Fernando Pepeu Jardim Braga**

não consigo usar a função relocate. o rstudio diz que não acha a mesma, mesmo com o tidyverse instalado e o dplyr tb, mesmo usando dplyr::relocate

Fernando, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.

**Aquila José Gonçalves Delfino**

Como posso dar um SELECT, usando o everything() trazendo a coluna 1 duas vezes? Eu pensei em select(data\_set, coluna1, everything()) mas se fizer isso o everything() ignora a coluna 1 :c

Aquila, é uma boa pergunta e um excelente exercício para se tentar realizar. As boas soluções em programação partem de ideias e, algumas vezes, de tentativa e erro.

Recomendamos que faça esse exercício, transformando sua dúvida em uma programação e testando. Acreditamos que irá agregar muito ao conhecimento a elaboração de rotinas para testar as soluções pensadas. Teste a sua solução, caso tenha dúvidas, nos acione no e-mail da monitoria. Iremos prontamente lhe auxiliar.

**Aquila José Gonçalves Delfino**

Eu tenho uma duvida da aula passada, por que no comando select, o everything() desconsidera as variaveis que ja foram solicitadas? Nao deveria o everything() trazer todas as variáveis?

Aquila, solicitamos que envie as dúvidas de aulas passadas pelo e-mail, por gentileza.

**Emerson Fioravante Tomazi**

mas se fosse uma variavel numerica (uma faixa de 1 até 10 para perfil agressivo, 11 a 20 para moderado, etc....nós usariamos a função recode:? De qual fora seria?

Emerson, respondida ao vivo pelo professor.

**Priscila Schall**

boa noite, pessoal. rodei o comando da linha 95, mas não excluiu os resultados NA da base. Na linha 103 também, a base carregou com NAs. Será que preciso ajustar algo aqui?

Priscila, nos encaminhe um print do erro no e-mail da monitoria, juntamente com as informações do seu sistema operacional. Iremos prontamente lhe auxiliar.



**Anderson José Mendes**

Existe algum comando que mostre os agrupamentos que tem em uma base?

Anderson, não sei se entendi a sua pergunta. Vou tentar responder e, caso não seja o que perguntou, nos acione no e-mail da monitoria, por gentileza. Caso queria testar se o objeto que está trabalhando está agrupado, bastaria clicar no objeto no Global Environment, esta informação se encontra descrita nas características gerais da base.

**Alex Ananias Da Silva**

Qual a maneira de se fazer o desagrupamento sem o pipe?

Alex, operações sem o pipe, geralmente, levam o argumento que inicialmente está antes do pipe para dentro da função.

Hernandes Matias Junior 124.799.11640