



Enap

Análise de Dados em Linguagem R

Módulo

2 Fundamentos da Linguagem R



Fundação Escola Nacional de Administração Pública

Presidente

Diogo Godinho Ramos Costa

Diretor de Desenvolvimento Profissional

Paulo Marques

Coordenador-Geral de Educação a Distância

Carlos Eduardo dos Santos

Equipe responsável

Ana Carla Gualberto Cardoso (Diagramação, 2020).

Ana Paula Medeiros Araújo (Direção e Produção Gráfica, 2020).

Douglas Gomes Ferreira (Conteudista, 2020).

Guilherme Teles da Mota (Implementação Rise, 2020).

Iara da Paixão Corrêa Teixeira (Designer Instrucional, 2020).

Juliana Bermudez Souto de Oliveira (Revisão Textual, 2020).

Larisse Padua da Silva (Produção Audiovisual, 2020).

Michelli Batista Lopes (Produção Audiovisual e Implementação, 2020).

Patrick Coelho (Implementação Moodle, 2020).

Sheila Rodrigues de Freitas (Coordenação Web, 2020).

Desenvolvimento do curso realizado no âmbito do acordo de Cooperação Técnica FUB / CDT / Laboratório LatITUDE e Enap.

Curso produzido em Brasília, 2020.



Enap, 2020

Enap Escola Nacional de Administração Pública

Diretoria de Educação Continuada

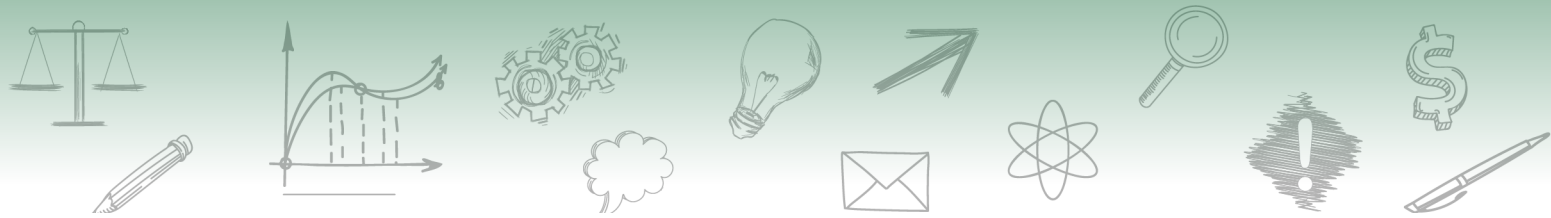
SAIS - Área 2-A - 70610-900 — Brasília, DF



Sumário

Unidade 1 - Preparação do Ambiente de Análise de Dados	5
1.1 Instalando o R	5
1.2 Instalando o RStudio	6
1.3 Conhecendo o RStudio.....	6
Unidade 2 - Panorama da Linguagem R.....	8
2.1. Criando o seu primeiro script.....	8
2.2 Comentando seu código	8
2.3 Documentação da função	8
2.4 Instalando pacotes.....	8
2.5 Estrutura de dados.....	9
2.6 Estruturas de repetição.....	10
2.7 Controle de fluxo.....	10
2.8 Criando funções	10
2.9 Utilizando funções <i>apply</i>	10
2.10 Criando gráficos	11
Unidade 3 - Manipulação de Dados com R.....	11
3.1. Utilizando junções.....	11
3.2 Seleccionando os dados	12
3.3 Transformando os dados.....	12
Referências.....	13





Módulo 2 Fundamentos da Linguagem R

DESTAQUE

Ao final deste módulo, você terá aptidão para executar a instalação dos programas R e RStudio, usar suas estruturas, os tipos de dados e as funções da linguagem R, além de reconhecer as principais funções relacionadas à seleção e transformação de dados nessa linguagem.

Unidade 1 - Preparação do Ambiente de Análise de Dados

Nesta unidade você vai acompanhar o passo a passo para a instalação dos programas R e RStudio. Para tanto, vamos utilizar dois vídeos com a demonstração da instalação dos programas.


1.1 Instalando o R

Primeiramente, vamos relembrar o conceito de linguagem R:

A linguagem R é uma linguagem de programação usada para análise estatística e produção de gráficos. Trata-se de um software gratuito com código aberto e com uma linguagem acessível, o que facilita a sua usabilidade.

Para podermos trabalhar com o R, é necessário fazer o download do programa no computador. Portanto, vamos abrir o navegador e, na barra de endereço, digitar o link: <https://cran.r-project.org/>. Em seguida, devemos escolher a versão de acordo com o sistema operacional utilizado e executar o arquivo de instalação.

No vídeo a seguir, você poderá acompanhar os passos para instalação do R.

 **Vídeo 1:** [Instalando o R](#)

Após instalar o R em seu computador, é necessário instalar o RStudio, que será mostrado no próximo tópico.



1.2 Instalando o RStudio

Antes de aprendermos a instalar o RStudio, é importante saber que ele **é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos**. Pode-se afirmar ainda que o RStudio possui uma interface gráfica que simplifica o uso da linguagem R na manipulação de dados, na criação de gráficos e nos cálculos estatísticos.

Depois dessa breve explicação, vamos aprender a instalar o programa RStudio. Para isso, vamos abrir o navegador e, na barra de endereço, digitar o link: <https://rstudio.com/products/rstudio/download/#download>. A página exibida apresentará uma lista com os principais navegadores. Em seguida, devemos escolher a versão compatível com o sistema operacional utilizado e executar o arquivo de instalação.

No vídeo a seguir, você poderá acompanhar os passos para instalação do RStudio.

 **Vídeo 2:** [Instalando RStudio](#)

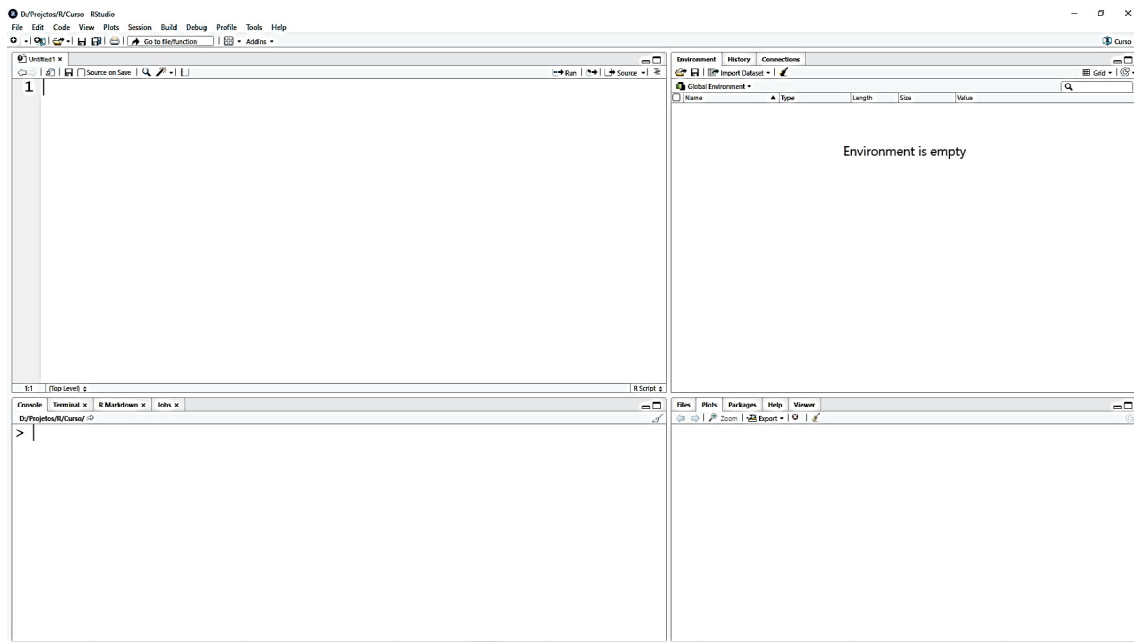
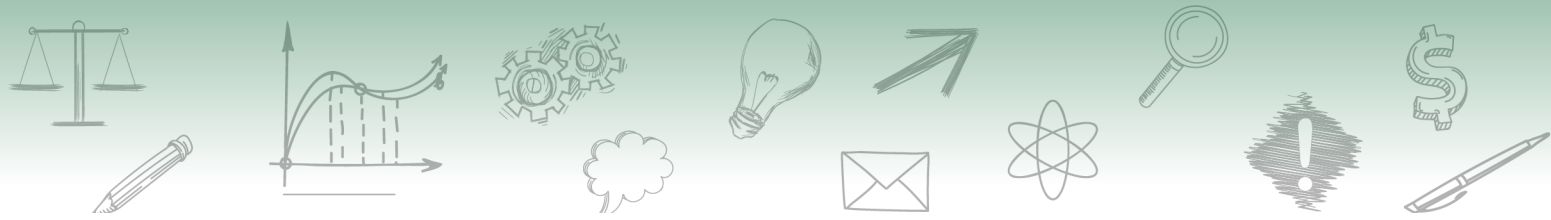
Agora você possui uma ferramenta para auxiliar o trabalho com a infinidade de dados produzidos no dia a dia e nas organizações.

1.3 Conhecendo o RStudio

Após a instalação do programa RStudio, para usá-lo efetivamente, é necessário conhecer a sua estrutura.

O RStudio é dividido em quatro janelas ou quadrantes: “Source”, “Console”, “Environment” e “Visualização”.

A imagem a seguir apresenta essa divisão:



Neste momento, vamos conhecer cada uma dessas janelas, com uma breve descrição.

A primeira janela, localizada no primeiro quadrante (parte superior à esquerda), é denominada “Source” e será o local onde o código deve ser inserido.

A segunda janela, localizada no segundo quadrante (parte inferior à esquerda), é denominada “Console” e exibe o resultado do que é executado na janela “Source”.

A terceira janela, localizada no terceiro quadrante (parte superior à direita), é denominada “Environment” e exibe objetos que estão carregados na memória. Nela também é possível remover ou salvar os objetos.

A última janela, localizada no quarto quadrante (parte inferior à direita), é denominada “Visualização” e é composta pelas abas: “Files”, “Plots”, “Packages” e “Help”. A seguir, acompanhe a funcionalidade de cada uma:

“Files”

Exibe a estrutura de arquivos do projeto.

“Plots”

Exibe os gráficos.

“Packages”

Permite instalar, atualizar e pesquisar pacotes.

“Help”

Exibe a documentação de uma determinada função ou pacote quando a função “help()” é executada.

Após conhecer a estrutura do Rstudio, que tal praticar um pouco?



Unidade 2 - Panorama da Linguagem R

DESTAQUE

Ao final desta unidade, depois de conhecer os tipos de dados e as funções da linguagem R, você terá aptidão para utilizar suas estruturas e construir gráficos com as ferramentas disponíveis no R.

2.1. Criando o seu primeiro script

Para demonstrar a criação de um script, acompanhe os passos descritos no vídeo a seguir:

 **Vídeo 3:** [Criando seu primeiro script](#)

Após aprender a criar um script, vamos avançar um pouco mais!

2.2 Comentando seu código

No próximo vídeo, é apresentada uma maneira para criação de comentário por meio de um símbolo muito usado na linguagem da internet. Você faz ideia de que símbolo é esse?

 **Vídeo 4:** [Comentando o seu Código](#)

Você acertou qual era o símbolo? E já tinha imaginado que o caractere *hashtag* seria usado em linguagem de dados?

2.3 Documentação da função

A essa altura, você pode estar com dúvidas relacionadas ao RStudio. No vídeo a seguir, é demonstrado como utilizar o recurso de ajuda. Acompanhe:

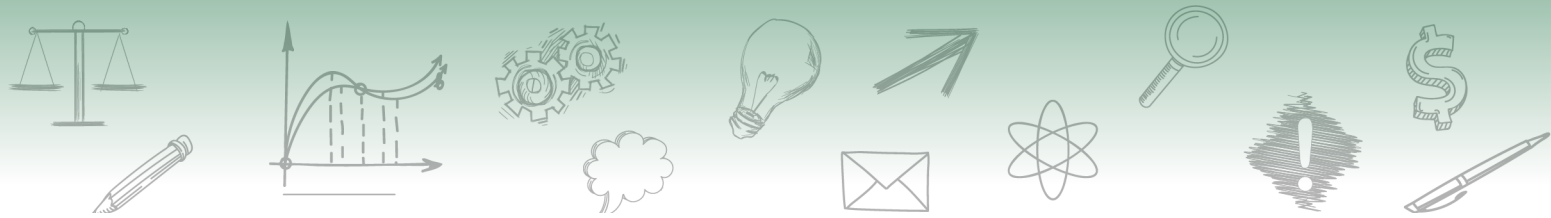
 **Vídeo 5:** [Documentação da Função](#)

2.4 Instalando pacotes

Uma das grandes vantagens do R são os pacotes disponíveis. Alguns vêm pré-instalados e são encontrados na aba “Packages”.

Para usarmos as funções que um pacote oferece, primeiro é necessário instalar o pacote desejado por meio da função “install.packages()” e depois carregá-lo utilizando a função “library()”.

O vídeo a seguir explica como realizar esses passos.



🎥 Vídeo 6: [Instalando pacotes](#)

Você percebeu que o ponto final foi usado várias vezes? Na linguagem R é comum utilizá-lo para representar o espaço, diferente de outras linguagens, que utilizam o *underscore* ou o padrão “CamelCase”.

SAIBA MAIS

É possível consultar os pacotes disponíveis na página oficial do Comprehensive R Archive Network (CRAN), no documento [Available CRAN Packages By Name](#).

2.5 Estrutura de dados

Na linguagem R temos diversas estruturas que permitem armazenar os dados produzidos. Nesse tópico, vamos conhecer essas estruturas e suas principais funções.

Vetor

Armazena um conjunto de valores ordenados chamados de elementos. Todos os elementos de um vetor devem ser do mesmo tipo.

Fator

Armazena valores categóricos (nominal ou ordinal).

Lista

É um tipo especial de vetor, porém é possível adicionar elementos de tipos diferentes.

Data frame

É utilizado para organizar elementos em linhas e colunas, semelhante a uma planilha ou base de dados. Além disso, pode ser formado por listas ou vetores.

Matriz

Armazena dados tabulares semelhante ao *data frame*, porém só aceita um tipo de dado.

O vídeo a seguir demonstra como essas estruturas funcionam.

🎥 Vídeo 7: [Estruturas de dados](#)

Esse vídeo é fundamental para aprender mais sobre a linguagem de programação R. Em nosso curso, limitamos a apresentar as funções mais utilizadas, no entanto, existem inúmeras outras que podem auxiliar o seu trabalho. Esperamos que você se aventure na busca de outras funções que atendam suas necessidades.



2.6 Estruturas de repetição

Nesse tópico, vamos ensinar como utilizar as estruturas de repetição na linguagem R, pegando duas como exemplo: *For* e *While*.

A função “for()” é utilizada para acessar cada elemento de uma sequência (vetor, coluna de matriz ou *data frame* matriz), enquanto a função “while()” é utilizada para repetir uma operação por um determinado número de vezes.

O vídeo a seguir exemplifica como utilizar essas estruturas.

 **Vídeo 8:** [Estruturas de repetição](#)

As funções aprendidas nesse vídeo são muito úteis. Você imagina em quais situações pode aplicá-las?

2.7 Controle de fluxo


O programa linguagem R possui uma função que permite testar condições e controlar o fluxo do script. Acompanhe no vídeo a seguir:

 **Vídeo 9:** [Controle de fluxo](#)

2.8 Criando funções

O R já vem com várias funções instaladas e, com elas, você pode executar muitas tarefas, conforme mostramos até agora.

Ao criar suas funções, você facilita o seu trabalho e diminui o tempo de execução de suas análises. O vídeo a seguir demonstra como criar funções no R.

 **Vídeo 10:** [Criando funções](#)

Agora que ensinamos como criar funções no R, você pode criar várias outras para facilitar suas análises. Vá em frente!

2.9 Utilizando funções *apply*

Você reparou no título desse tópico? Percebeu que a palavra funções está no plural? Sabe por quê?



Agora, vamos aprender a utilizar mais uma função do R, ou melhor, um grupo de funções. A função “apply”, na verdade, é um conjunto de funções, ou seja, ela possui algumas variações, tais como: “lapply”, “tapply”, “sapply” e outras.

Para aprender como utilizá-las, acompanhe o vídeo a seguir:

 **Vídeo 11:** [Utilizando funções apply](#)

2.10 Criando gráficos

Chegamos no ápice da nossa unidade de estudo: a criação de gráficos. Os gráficos facilitam a análise e a interpretação dos dados, além de possibilitar a apresentação dos dados e subsidiar a tomada de decisão com maior segurança.

No vídeo a seguir, são apresentadas as ferramentas para a criação de gráficos no R.

 **Vídeo 12:** [Criando gráficos](#)

Com esse vídeo sobre criação de gráficos em linguagem R, concluímos esta unidade. Tivemos a oportunidade de conhecer e aplicar algumas funções e utilizamos as ferramentas que possibilitam a construção desses gráficos.

Agora, você pode treinar usando outros dados disponíveis na sua organização ou em vários sites governamentais ou não.

Unidade 3 - Manipulação de Dados com R

DESTAQUE

Ao final desta unidade, você deverá ser capaz de reconhecer as principais funções relacionadas à união, seleção e transformação de dados em linguagem R.

3.1. Utilizando junções

A linguagem R permite unir dois conjuntos de dados em um único lugar. É isso que vamos abordar nesse tópico.

No vídeo a seguir, ensinamos como utilizar as funções “joins” em R. Acompanhe:

 **Vídeo 13:** [Utilizando junções/joins](#)

O que achou dessas funções? Sente-se apto a usá-las em suas análises?



3.2 Selecionando os dados

Nesse tópico, vamos aprender a trabalhar com a seleção de dados no R. Uma parte considerável dos problemas relacionados à manipulação de dados pode ser solucionada com o uso das funções do “dplyr”, por isso é um dos pacotes mais comuns, além da simplicidade na sua execução. No vídeo a seguir, apresentamos como utilizar esse pacote.

 **Vídeo 14:** [Selecionando os dados](#)

Terminamos mais um tópico em que você pôde aprender a selecionar dados, uma etapa importante na manipulação de dados.

3.3 Transformando os dados

Para manipular dados na linguagem R, além de utilizar as funções “joins” e selecionar os dados, você precisa aprender a transformar esses dados, a fim de obter informações mais compatíveis com a sua necessidade.

É isso que abordaremos no próximo vídeo. Acompanhe:

 **Vídeo 15:** [Transformando os dados](#)

Com a abordagem acerca da manipulação de dados em R, terminamos este módulo. Esperamos que os conhecimentos aqui apreendidos possam facilitar suas próximas análises de dados.



Referências

Unidade 1 - Preparação do Ambiente de Análise de Dados

RSTUDIO. Boston: RStudio, c2020. Disponível em: <https://rstudio.com/products/rstudio/download/>. Acesso em: 10 ago. 2020.

WIRTSCHAFTS UNIVERSITÄT WIEN. **The Comprehensive R Archive Network**. Wien: Institute for Statistics and Mathematics, 2020. Disponível em: <https://cran.r-project.org/>. Acesso em: 10 ago. 2020.

Unidade 2 - Panorama da Linguagem R

APPLY(), Lapply(), Sapply(), Tapply() Function in R with Examples. **Guru 99**, [s. /], [2020]. Disponível em: <https://www.guru99.com/r-apply-sapply-tapply.html>. Acesso em: 10 ago. 2020.

ENDER, P. **Multivariate Statistical Techniques: Matrix Operations in R**. Westwood, LA: UCLA, 2007. Disponível em: <http://www.philender.com/courses/multivariate/notes/matr.html>. Acesso em: 10 ago. 2020.

FANARA, C. R Tutorial on the Apply Family of Functions. **R-bloggers**, [s. /], 28 jul. 2015. Disponível em: <https://www.r-bloggers.com/r-tutorial-on-the-apply-family-of-functions/>. Acesso em: 10 ago. 2020.

GILLESPIE, C. S. Installing R packages. **R-bloggers**, [s. /], 6 nov. 2010. Disponível em: <https://www.r-bloggers.com/installing-r-packages/>. Acesso em: 10 ago. 2020.

PLOT: Generic X-Y Plotting. **RDocumentation**, [s. /], [2020]. Disponível em: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/plot>. Acesso em: 10 ago. 2020.

R DATA Frame. **DataMentor**, Nepal, [2020]. Disponível em: <https://www.datamentor.io/r-programming/data-frame/>. Acesso em: 10 ago. 2020.

R HISTOGRAMS. **DataMentor**, Nepal, [2020]. Disponível em: <https://www.datamentor.io/r-programming/histogram/>. Acesso em: 10 ago. 2020.

R WHILE Loop. **DataMentor**, Nepal, [2020]. Disponível em: <https://www.datamentor.io/r-programming/while-loop/>. Acesso em: 10 ago. 2020.

THEUWISSEN, M. How to write the first for loop in R. **R-bloggers**, [s. /], 2 dez. 2015. Disponível em: <https://www.r-bloggers.com/how-to-write-the-first-for-loop-in-r/>. Acesso em: 10 ago. 2020.

TOTH, M. Detailed Guide to the Bar Chart in R with ggplot. **R-bloggers**, [s. /], 1 maio 2019. Disponível em: <https://www.r-bloggers.com/detailed-guide-to-the-bar-chart-in-r-with-ggplot/>. Acesso em: 10 ago. 2020.



VRIES, A.; MEYS, J. How to Work with Ordered Factors in R. **Dummies**, [s. l.], c2020. Disponível em: <https://www.dummies.com/programming/r/how-to-work-with-ordered-factors-in-r/>. Acesso em: 10 ago. 2020.

WICKHAM, H. **Advanced R**. 2. ed. Boca Ratón: CRC Press, 2019. Disponível em: <https://adv-r.hadley.nz/>. Acesso em: 10 ago. 2020.

YAU, C. **R Tutorial: An R Introduction to Statistics**. [S. l.: S. n.], 2020. Disponível em: <http://www.r-tutor.com/r-introduction>. Acesso em: 10 ago. 2020.

Unidade 3 - Manipulação de Dados com R

BOEHMKE, B. Reshaping your data with tidyr. **UC Business Analytics R Programming Guide**, Cincinnati, [2018]. Disponível em: <https://uc-r.github.io/tidyr>. Acesso em: 11 ago. 2020.

WICKHAM, H. Gather columns into key-value pairs. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://tidyr.tidyverse.org/reference/gather.html>. Acesso em: 10 ago. 2020.

WICKHAM, H. Separate a character column into multiple columns with a regular expression or numeric locations. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://tidyr.tidyverse.org/reference/separate.html>. Acesso em: 11 ago. 2020.

WICKHAM, H. Unite multiple columns into one by pasting strings together. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://tidyr.tidyverse.org/reference/unite.html>. Acesso em: 11 ago. 2020.

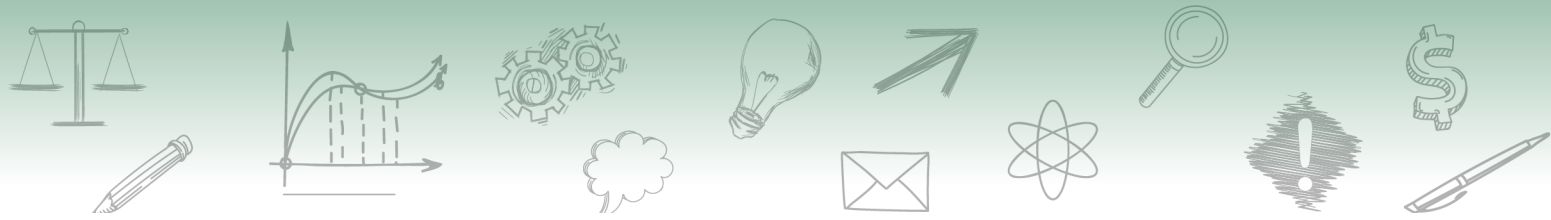
WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Arrange rows by column values. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://dplyr.tidyverse.org/reference/arrange.html>. Acesso em: 10 ago. 2020.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Create, modify, and delete columns. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://dplyr.tidyverse.org/reference/mutate.html>. Acesso em: 10 ago. 2020.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Group by one or more variables. **Tidyverse**, [s. l.], [2020]. Disponível em: https://dplyr.tidyverse.org/reference/group_by.html

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Join two tbls together. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://dplyr.tidyverse.org/reference/join.html>. Acesso em: 10 ago. 2020.

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Subset columns using their names and types. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://dplyr.tidyverse.org/reference/select.html>. Acesso em: 10 ago. 2020.



WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. Subset rows using their positions. **Tidyverse**, [s. l.], [2020]. Disponível em: <https://dplyr.tidyverse.org/reference/slice.html>. Acesso em: 10 ago. 2020.