

# Readme.rmd

Sergio Pedro R Oliveira

Março de 2022

## Objetivo

Estudo dirigido de linguagem R.

## Livro de referência

Utilizando a Linguagem R.  
Editora: ALTA BOOKS EDITORA

## Assuntos por capítulos e resumos

### Cap 1 - Instalação do R e Rstudio

### Cap 2 - Pacote base e funções estatísticas básicas

### Cap 3 - Principais pacotes

1. Principais pacotes:

- **Rmarkdown**  
Produção de relatorios(html, pdf, doc, md).
- **knitr**  
Interpretação e compilação do documento rmd.
- **data.table**  
Exploração de data.frames.
- **janitor**  
Limpeza de dados.
- **DescTools**  
Análise descritiva de dados.

- **tidyverse**  
conjunto de pacotes.
  - **readr**  
Importação e leitura de arquivos de dados.
  - **tibble**  
estruturação de `data.frame`.
  - **dplyr**  
Manipulação de `data.frame`.
  - **tidyr**  
Organização de `data.frame`.
  - **ggplot2**  
Visualização de dados, produção de gráficos.
  - **purrr**  
Manipulação de vetores e listas.

## 2. Instalação de pacotes:

- sintaxe de instalação:  
**install.packages**(*“nome do pacote”*)
- sintaxe de variáveis instalações simultâneas:  
**install.packages**(*c(“nome do pacote”, “nome do pacote”, ...)*, dependencies = **TRUE**)

## Cap 4 - R Markdown

## Cap 5 - Pacotes do Tidyverse e identificando/mudando tipos de variaveis

### 1. identificando/mudando tipos de variaveis

- i. identificando  
uso do **is**.
- ii. mudando o tipo de variavel:  
uso do **as**.

### 2. pacotes do Tidyverse:

- **readr**  
Leitura de dados.
- **tibble**  
Tipo de data.frame.
- **magrittr**  
Operador pipe ' $\%>\%$ ', concatena linhas de comando.
- **dplyr**  
Manipulação de dados.

#### i. manipulação de dados:

- *select*  
seleciona e retorna as colunas selecionadas da tabela.
- *pull*  
extrai uma coluna de uma tabela de dados e retorna ela como vetor.
- *filter*  
filtra linhas.
- *distinct*  
remove linhas com valores repetidos.
- *arrange*  
reordena ou combina linhas.
- *mutate*  
cria novas colunas.
- *transmute*  
cria novas colunas, mas não adiciona na base de dados.
- *summarise*  
sumariza valores.

- *group\_by*  
permite operações por grupo.
- *add\_column*  
adiciona novas colunas.
- *add\_row*  
adiciona novas linhas.
- *rename*  
renomeia uma coluna.

ii. combinando tabelas de dados:

- *bind\_cols*  
Une duas tabelas lado a lado. acrescenta numeração as colunas repetidas.  
É necessario que tenha o mesmo numero de linhas nas duas tabelas para fazer essa combinação.
- *bind\_rows*  
Une duas tabelas sobrepostas.  
Quando não há correspondencia o comando retorna **NA**.
- *inner\_join*  
A tabela final será o resultado da intersecção das duas colunas de x e y, que possuem pelo menos uma coluna em comum, a coluna chave.  
Junta duas colunas pela interseção.
- *left\_join*  
Une duas tabelas, definindo qual será a tabela principal e a unida a esquerda da outra. Esse fator muda a interpretação das linhas/registros correspondentes uma na outra, no caso, a tabela principal e tabela que será colocada a esquerda. É necessario que tenha pelo menos uma coluna em comum, uma coluna chave.
- *right\_join*  
Une duas tabelas, definindo qual será a tabela principal e a unida a direita da outra. Esse fator muda a interpretação das linhas/registros correspondentes uma na outra, no caso, a tabela principal e tabela que será colocada a direita. É necessario que tenha pelo menos uma coluna em comum, uma coluna chave.
- *full\_join*  
Une duas tabelas. Prestar atenção na junção das linhas/registros que formam novas informações, através da junção de correspondentes. É necessario que tenha pelo menos uma coluna em comum, uma coluna chave.
- *intersect*  
Retorna a interseção entre tabelas.
- *union*  
Retorna a união de tabelas.
- *setdiff*  
Retorna a diferença entre tabelas.
- *setequal*  
Esse comando verifica se duas tabelas de dados possuem linhas com os mesmos valores, independentemente da ordem em que tais valores se apresentem. retorna **TRUE**, se os

registros forem iguais, ou **FALSE**, se os registros forem diferentes.

- **tidyr**

Organização de dados.

- *pivot\_longer* ou *gather*

Converte a tabela de dados para o formato longo. (larga -> longo)

- *pivot\_wider* ou *spread*

Converte a tabela de dados para o formato larga. (longo -> larga)

- *separate*

Separa as respostas que estão em uma unica coluna para diversas colunas.

- *unite*

O comando unite é utilizado para unir duas ou mais colunas em uma unica coluna.

- *complete*

Completa as combinações de duas colunas, se não houver valor completa com *NA*.

- *drop\_na*

Elimina as linhas, especificadas ou não, com valor *NA*.

- *replace\_na*

Substitui o valor *NA* por outro valor especificado.

## Cap 6 - Pacote `data.table`

### 1. `data.table`

- Manipulando linhas
- Manipulando colunas
- Sumarizando dados
- Operando um subconjunto de dados
  - *lapply*
- modificando dados com `set`:
  - *set*  
modificando um valor.
  - *setnames*  
modificando nome da coluna.
  - *setorder*  
modificando ordem das linhas.
  - *setcolorder*  
modificando ordem das colunas.

## Cap 7 - Gráficos basicos e pacote ggplot2

### 1. Gráficos basicos:

- Gráfico de barras
- Gráfico circular (pizza)
- Gráfico de linhas
- Gráfico de dispersão
- Histograma
- Boxplot (diagrama de caixa)

### 2. Pacote **ggplot2**



## **Andamento dos Estudos**

### **Assunto em andamento:**

Atualmente estou estudando exercicios de gráficos basicos.

### **Em andamento:**

### **Vazios:**

### **Finalizando detalhes:**