

Readme.rmd

Sergio Pedro R Oliveira

2022-05-19

Contents

1	Objetivo	1
2	Livro de referência	1
3	Assuntos por capitulos e resumos	2
3.1	Cap 1 - Instalação do R e Rstudio	2
3.2	Cap 2 - Pacote base e funções estatísticas básicas	2
3.3	Cap 3 - Principais pacotes	2
3.4	Cap 4 - R Markdown	4
3.5	Cap 5 - Pacotes do Tidyverse e identificando/mudando tipos de variaveis	6
3.6	Cap 6 - Pacote data.table	9
3.7	Cap 7 - Gráficos basicos e pacote ggplot2	10
4	Andamento dos Estudos	13
4.1	Assunto em andamento:	13
4.2	Em andamento:	13
4.3	Vazios:	13
4.4	Finalizando detalhes:	13

1 Objetivo

Estudo dirigido de linguagem R.

2 Livro de referência

Utilizando a Linguagem R.
Editora: ALTA BOOKS EDITORA

3 Assuntos por capítulos e resumos

3.1 Cap 1 - Instalação do R e Rstudio

3.2 Cap 2 - Pacote base e funções estatísticas básicas

3.3 Cap 3 - Principais pacotes

1. Principais pacotes:

- **Rmarkdown**
Produção de relatórios(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**)

3.4 Cap 4 - R Markdown

1. Preâmbulo:

- *title*: “Titulo desejado”
- *author*: “Nome dos autores”
- *date*: “Data do dia da compilação”, para adicionar a data atual pode usar a função através do chunk “`r Sys.Date()`”
- *output*: o tipo de saída, podem ser:
 - Documentos:
 - * *pdf_document*
 - * *md_document*
 - * *html_document*
 - * *word_document*
 - * *odt_document*
 - * *rtf_document*
 - Apresentação:
 - * *powerpoint_presentation*
 - * *ioslides_presentation*
 - * *beamer_presentation*
 - mais:
 - * *flexdashboard::flex_dashboard*
 - * *github_document*

2. Chunks/codigos embutidos:

3. Titulos e subtítulos:

4. Listas e blocos de citação:

5. Inserir tabelas:

6. Fontes:

7. Hiperlinks e imagens:

- Hiperlinks

- Imagens

8. Letras gregas:

9. Fórmulas:

- Subscritos e superescritos

- Sublinhados, sobrelinhas e vetores

- Frações, matrizes e chavetas

- Expressões

- Sinais e setas

3.5 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, atraves 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.

3.6 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.

3.7 Cap 7 - Gráficos basicos e pacote ggplot2

1. Gráficos basicos:

- Gráfico de barras
barplot
- Gráfico circular (pizza)
pie
- Gráfico de linhas
plot
 - Para adicionar mais linhas no grafico.
lines
- Gráfico de dispersão
 - Para obter a correlação.
cor(x,y)
 - Para obter o coeficiente da reta de regressão.
lm(y ~x)\$coef
 - Adiciona a reta tracejada.
abline
- Histograma
hist
- **Boxplot** (diagrama de caixa)

2. Pacote **ggplot2**

- Constroi diversos tipos de graficos a partir da mesma estrutura de componentes:
 - *data*: referente ao banco de dados.
 - *geom_forma*: um rol de tipos possiveis de representação dos dados.
 - *coord_system*: referente ao sistema de coordenadas, que podem ser cartesianas, polares e projeção de mapas.

i) O que precisa para fazer o grafico?

- A. Um nome de objeto para guardar o grafico (uma variavel).
- B. A base de dados que será utilizada para a plotagem.

ggplot(*data=nome_da_base*)

- C. Descrever como as variaveis serão utilizadas na plotagem:

aes(*x=...*, *y=...*, ...)

- D. Especificar o tipo de grafico:

geom_forma(...)

- E. Utilizar o operador “+” para adicionar camadas ao objeto **ggplot** criado.
- F. Pacotes auxiliares como *ggthemes* e *grid*, dentre outros.

ii) Quais formatos podemos utilizar no ggplot2 - geom_forma?

```
##                                forma
## 1          geom_area ou geom_ribbon
## 2          geom_bar ou geom_col
## 3          geom_bar+coord_polar
## 4          geom_boxplot
## 5          geom_curve
## 6          geom_density
## 7          geom_dotplot
## 8          geom_histogram
## 9  geom_line, geom_abline, geom_hline, geom_vline
## 10         geom_point
## 11         geom_qq ou geom_qq_line
## 12         geom_tile, geom_rect ou geom_raster
## 13         geom_violin
##                                tipo_de_grafico
## 1  Produz um grafico para visualizar área sob a curva ou entre curvas.
## 2          Produz um grafico de colunas do vetor x.
## 3          Produz um grafico circular (Pizza).
## 4          Produz o boxplot de x.
## 5          Produz um grafico em curva.
## 6          Produz um grafico da densidade de x.
## 7          Produz um grafico de pontos.
## 8          Produz um histograma do vetor x.
## 9          Produz um grafico de linhas
## 10         Produz um grafico de dispersão entre x e y.
## 11         plota os quantis de x usando como base a curva normal.
## 12         Produz uma grade de retangulos.
## 13         Produz um grafico em forma de violino.
```

iii) Nome dos argumentos para adicionar efeito em graficos do pacote ggplot2.

```
##                                funcao
## 1          autoplot
## 2          coord_cartesian
## 3          coord_fixed
## 4          coord_flip
## 5          coord_polar
## 6          geom_blank
## 7          geom_jitter
## 8          geom_smooth
## 9          geom_text
## 10 scale_fill_(=brewer ou grey ou gradient)
## 11         scale_*_contínuos
## 12         scale_*_discrete
## 13         scale_*_manual
##                                efeito_no_grafico
## 1  Produz um grafico apropriado para o tipo de variavel
## 2          Coordenada cartesiana
## 3  Coordenada cartesiana com razão entre eixo x e y fixada
```

```
## 4             Inverte a posição dos eixos x e y
## 5             Coordenada polar
## 6             Janela em branco
## 7             Produz um efeito jitter
## 8             Produz uma curva suavizada
## 9             Aplica texto a janela grafica
## 10            Define a escala de cores
## 11            Define parametros para o eixo x ou y continuos
## 12            Define parametros para o eixo x ou y discreto
## 13            Define parametros para os eixos manualmente
```

- Definindo um tema para o grafico **ggplot**.

- *theme_gray*
Fundo cinza e linhas grandes brancas.
- *theme_bw*
O classico preto e branco. Otimo para projetor.
- *theme_linedraw*
Linhas pretasde varias larguras num fundo branco. semelhante ao theme_bw.
- *theme_light*
Semelhante ao theme_linedraw, porem com as linhas mais cinza claro, para dar atenção aos dados.
- *theme_dark*
Versão escura do theme_light, com o fundo escuro, util para criar linhas finas coloridas.
- *theme_minimal*
Um tema minimalista sem anotações de fundo.
- *theme_classic*
Tema classico, com linhas do eixo x e y, sem linhas de grade.
- *theme_void*
Um tema completamente vazio.

4 Andamento dos Estudos

4.1 Assunto em andamento:

Atualmente estou estudando pacote ggplot2.

4.2 Em andamento:

4.3 Vazios:

4.4 Finalizando detalhes: