Readme.rmd

Sergio Pedro R Oliveira

2022-05-19

Table of Contents

# 1 Objetivo

Estudo dirigido de linguagem R.

# 2 Livro de referência

Utilizando a Linguagem R.  
Editora: ALTA BOOKS EDITORA

# 3 Assuntos por capitulos 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 relatorios(html, pdf, doc, md).
* **knitr**  
  Interpretação e compilação do documento rmd.
* **data.table**  
  Exploração de data.frames.
* **janitor**  
  Limpeza de dados.
* **DescTools**  
  Analise 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.
  + **purr**  
    Manipulação de vetores e listas.

1. Instalação de pacotes:

* sintaxe de instalação:  
  **install.packages**(“*nome do pacote*”)
* sintaxe de variais instalações simultaneas:  
  **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 atraves do chunk “r Sys.Date()”
* *output*: o tipo de saida, 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*

1. Chunks/codigos embutidos:
2. Titulos e subtitulos:
3. Listas e blocos de citação:
4. Inserir tabelas:
5. Fontes:
6. Hiperlinks e imagens:

* Hiperlinks
* Imagens

1. Letras gregas:
2. 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  
   1. identificando  
      uso do **is**.
   2. 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.  
  1. munipulação 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.
  2. 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)

1. 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.

1. 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.
2. 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.

1. 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\_\*\_continuos  
## 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: