

Introdução ao R

Raimundo Marciano de Freitas Neto

2021-02-18

Contents

1	Prerequisites	5
2	O que é o R?	7
2.1	Scripts	10
3	Variáveis	11
3.1	Numéricas	13
3.2	Textuais	14
3.3	Lógicas	15
3.4	Datas	16
4	Estruturas de Dados	19
4.1	Vetores	19
4.2	Matrizes	20
4.3	Data frames	23
4.4	Listas	25
4.5	Factors	25
5	Trabalhando com variáveis no R Base	27
5.1	Manuseio de variáveis	27
5.2	Funções numéricas básicas	27
5.3	Funções textuais básicas	27
6	Criando sua própria função	29
6.1	Comentários	29

7	Methods	31
8	Applications	33
8.1	Example one	33
8.2	Example two	33
9	Final Words	35

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")  
# or the development version  
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading #.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.

Chapter 2

O que é o R?

R é uma linguagem de programação muito utilizada para o desenvolvimento de estatísticas e cálculos matemáticos. Entretanto, enquanto linguagem de programação, ela possui muitos usos, como a raspagem de sites, a automatização de download de arquivos, e a produção de dashboards e relatórios, sendo útil para a coleta de dados e para o reporte das análises, além de ser gratuita. Mas o que significa ser uma linguagem de programação? Basicamente, você irá dar ordens ao computador por meio de comandos. A vantagem de aprender a estruturar os comandos (ao invés de usar um programa estatístico já consolidado, como o Stata) é que você pode customizar a vontade suas análises. Os programas estatísticos trazem muitas opções pré-prontas. Por exemplo, no Stata é comum que você possa clicar em uma janela e solicitar um quadro com as estatísticas descritivas (média, moda, mediana) de um banco de dados. Entretanto, qualquer funcionalidade fora do padrão oferecido por aqueles pacotes precisará ser gerada pelo próprio usuário (e quando você usa algum recurso como o `ssc install` é porque alguém gastou um tempinho fazendo isso).

Aliás, o Python (outra linguagem de programação) possui funcionalidades bem similares ao R, embora o modo de escrever os programas seja bem diferente. Imagine que o Português e o Espanhol têm o mesmo propósito - a comunicação - e que mesmo tendo uma mesma raiz (e inclusive diversos vocábulos bem parecidos), possuem algumas regras estruturais bem diferentes, o que inclui, por exemplo, as regras de acentuação e o som produzidos pelas letras: qual a diferença de cajá (PT-BR) para caja (ES); qual a diferença de pastel (PT-BR) para pastel (ES)?

É preferível que o analista iniciante dedique-se a uma das duas (R ou Python). Posteriormente, aprender a outra (ou mais alguma que esteja despontando no mercado, como a Julia) será uma tarefa bem mais simples. Isso se deve ao fato de que diversos elementos da programação se repetem entre as diferentes linguagens, especialmente os conceitos. Um IF tem o mesmo propósito em R ou

em Python ou em Java ou em C++, embora a forma como você deve explicar ao computador o que fazer com esse IF será bem diferente em cada linguagem.

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 7.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

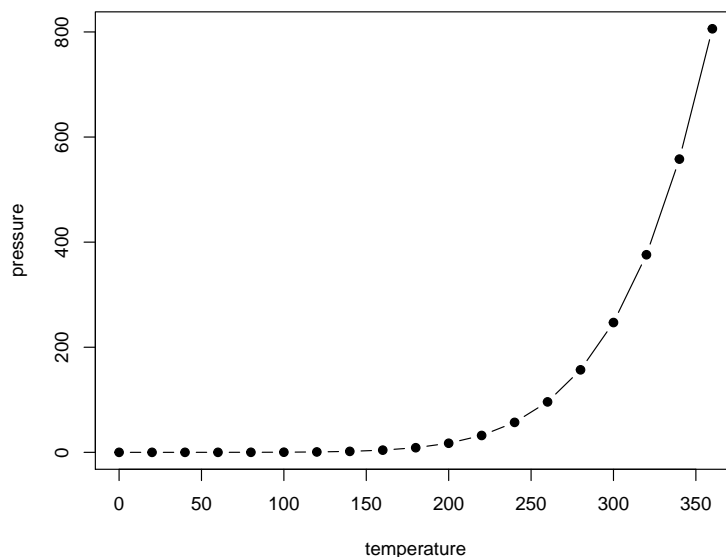


Figure 2.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

2.1 Scripts

Chapter 3

Variáveis

Os dados são inseridos no R como variáveis. Eles podem ser importados (trazidos de um arquivo para o R) ou digitados no próprio console. Quando você deseja armazenar um valor na memória, você deve atribuí-lo, usando o sinal = ou <=.

```
x = 1
print(x)
```

```
## [1] 1
```

```
y <- 1+1
y
```

```
## [1] 2
```

```
texto = "Esse é um texto. Observe as aspas duplas"
assim = 'Também podem ser usadas aspas simples'
sugestao = 'Use aspas simples sempre que "possível" para delimitar a variável. Os
textos em português costumam usadas aspas duplas.'
print(sugestao)
```

```
## [1] "Use aspas simples sempre que \"possível\" para delimitar a variável. Os \ntextos em portu
```

```
writeLines(sugestao)
```

```
## Use aspas simples sempre que "possível" para delimitar a variável. Os
## textos em português costumam usadas aspas duplas.
```

O R é uma linguagem que automaticamente reconhece o tipo do dado que foi inserido. Por uma questão de comparação, se você estivesse usando uma linguagem *tipada*, como o C++, você precisaria declarar explicitamente o tipo da variável que está sendo criada. Assim, seria um erro você criar uma variável usando

```
dois = 2
```

O C++ não entende o que você está querendo dizer com isso. Você precisaria informar que existe uma variável do tipo *integer* (simplesmente *int*) chamada *dois* e cujo valor é o número inteiro 2. Outra particularidade do R é que as linhas de código não precisam ser finalizadas com um ponto-e-vírgula (;), como acontece no C++.

```
int dois = 2;
```

Além disso, as variáveis em R são mutáveis, inclusive quanto ao tipo. Isso quer dizer que uma vez definidas (ou atribuídas), podem ter seus valores e tipo modificados. Assim, uma variável que antes tinha um número, pode passar a ter um texto ou um booleano.

```
# se a variavel receber um número (integer)
dois <- 2
# e depois receber um texto
dois <- "dois"
# seu conteúdo e seu tipo estarão de acordo com a última atribuição
print(dois)
```

```
## [1] "dois"
```

Por fim, para o R a indentação não é relevante. Há linguagens de programação em que a posição relativa das linhas de código é essencial para determinar se elas fazem parte de um bloco ou não. No R, os blocos são definidos com o uso de chaves ({ }). Ainda iremos explorar com detalhes o uso das condicionais, mas podemos ilustrar o seguinte caso: se o resultado for maior que zero, então, sim, tivemos lucro; caso contrário, não tivemos.

```
resultado <- 3000

#se o resultado for maior que zero
if (resultado > 0) {
  # informe: "Tivemos Lucro!"
  print("Tivemos Lucro!")
  # caso contrário
```

```

} else {
  # informe: "Deu ruim :( "
  print("Deu ruim :(")
}

```

```
## [1] "Tivemos Lucro!"
```

No R, esse alinhamento é apenas um facilitador de leitura do código. Se o código estiver organizado, as pessoas terão mais facilidade em entendê-lo. O código a seguir tem exatamente a mesma funcionalidade.

```

resultado <- 3000
if (resultado > 0) { print("Tivemos Lucro!") } else {print("Deu ruim :(")}

```

```
## [1] "Tivemos Lucro!"
```

Embora seja verdade que o espaço ocupado está menor, quando o código começa a adquirir um alto grau de complexidade, torna-se muito desejável que ele esteja melhor organizado e, preferencialmente, comentado.

No Python, onde a indentação faz diferença, o código precisaria seguir uma estrutura baseada em espaços, sendo desnecessário o uso das chaves.

```

#se o resultado for maior que zero
if (resultado > 0):
    # informe: "Tivemos Lucro!"
    print("Tivemos Lucro!")
else:
    # informe: "Deu ruim :( "
    print("Deu ruim :(")

```

No Python, portanto, o que importa para definir que o print("Tivemos Lucro!") está associado ao if(resultado > 0) é o fato de não haver nenhum espaço entre a margem da página e o if; e o fato de haver quatro espaços entre a margem da página e o print. Isso faz com que esse print esteja subordinado ao if; assim como, o segundo print (que também está mais recuado) é subordinado ao else (que não tem recuo).

3.1 Numéricas

Basicamente existem os inteiros (integers) e os ponto-flutuantes (float). O integer não admite casas decimais, sendo usado para eventos contáveis, como a

quantidade de vezes que algo ocorreu. O float admite casas decimais, sendo indicado para representar valores monetários.

Veja, por exemplo, o caso da idade. Se a idade for calculada como a diferença entre anos ($2021 - 1988 = 33$), o resultado será um número inteiro, que pode ser representado por um integer. Contudo, se uma idade como 12,5 anos for admitida, então será necessário trabalhar com float. Se você está lendo isso em fevereiro, pense na representação correta da idade de alguém que nasceu em dezembro.

Raramente, isso será uma preocupação. Como dito, o R faz a análise automática e, por padrão, categoriza os números dentro do tipo numeric, que aceita tanto integers como floats. Eventualmente, algum erro pode acontecer, como haver algum canto na planilha que está sendo importada em que foi digitado 2x20 ao invés de 2020. Quando o R tentasse ler esse valor (2x20), ele reconheceria um caractere que não é um número e automaticamente tentaria entender isso como sendo um texto.

3.2 Textuais

Para inserir um caractere, uma palavra ou uma frase, é necessário o uso de aspas simples ou duplas. Como no português é comum o uso de aspas duplas, é interessante o uso de aspas simples no R. Isso se deve ao fato de que se você precisar eventualmente incluir as aspas duplas, elas podem ficar dentro das aspas simples.

Por exemplo, é aceitável:

```
texto1 = 'Ela disse: "Ok".'  
texto2 = 'E eu respondi "Certo".'  
writeLines(c(texto1, texto2))
```

```
## Ela disse: "Ok".  
## E eu respondi "Certo".
```

Perceba que as aspas simples estão funcionando apenas como delimitadores. Estão dizendo que tudo que está dentro delas faz parte de um mesmo texto, inclusive as aspas duplas.

```
texto3 <- "Ela disse: "Ok"."
```

O texto3 seria problemático para o R. Assim que ele identifica as primeiras aspas, ele passa a esperar um texto. Quando aparecem mais aspas, ele entende que o texto foi encerrado e não pretende continuar lendo o que vem depois. Com isso, ele estranha que tenha mais texto logo após (Ok“.”)

3.2.1 Numérica ou textual?

O CNPJ de uma empresa é composto por 14 dígitos numéricos (0 a 9), separados por alguns marcadores (pontos, barras e traços). Por exemplo, pode ser 12.345.678/0001-19. Para guardar essa informação como uma variável numérica, você precisaria omitir as marcações, digitando 12345678000119. Entretanto, existe aqui um problema: o caso dos CNPJs iniciados por 0. Se há um 0 no início do número, ele é sempre ignorado. Por exemplo, 02 sempre será salvo na memória como 2, porque o zero à esquerda, na Matemática, não tem função.

Pensando nisso, pode ser mais interessante salvar o CNPJ como uma variável textual.

```
cnpj <- "12.345.678/0001-19"
```

Com isso não são perdidos os marcadores. Perceba, por fim, que o tratamento desse “número” como texto não tem implicações sobre a análise estatística dos dados. O CNPJ não é um número sobre o qual você realiza operações matemáticas. Você não calcula a média, a moda ou a mediana dos CNPJs e nem cria um gráfico de CNPJs. Ele é um identificador da empresa, tal qual o nome dela.

No caso de pessoas físicas, o mesmo raciocínio vale para o CPF.

Exceção: os dois últimos dígitos do CNPJ (e do CPF) são um código verificador. Alguns formulários cadastrais usam esses dois últimos dígitos para conferir se o campo foi preenchido corretamente. Essa conferência é feita a partir de um cálculo matemático; logo, há casos específicos em que mesmo o CNPJ pode ser usado para algum cálculo.

3.3 Lógicas

Em muitos casos, é importante ter variáveis que indiquem VERDADEIRO (TRUE) ou FALSO (FALSE). O tipo de variável que admite apenas esses dois valores lógicos é chamado de Booleana (boolean ou bool). Essas variáveis seguem a lógica de primeira ordem. Por exemplo, TRUE AND FALSE resulta em FALSE, enquanto TRUE OR FALSE resulta em TRUE.

```
# AND ou E é representado por &  
print(TRUE & FALSE)
```

```
## [1] FALSE
```

```
# OR ou OU é representado por |
print(TRUE | FALSE)
```

```
## [1] TRUE
```

Variáveis lógicas podem ter muitos usos, como parâmetros em funções ou filtros para listas de dados. Por exemplo, você pode querer que o usuário informe se ele quer imprimir o relatório dos valores ao final. Nesse caso, ele deveria informar que o valor do parâmetro `imprimir_relatorio` deveria ser `TRUE`. Caso ele não queira imprimir o relatório, deveria informar `imprimir_relatorio` como `FALSE`.

Também podem ser usados para comparações de igualdade (`==`) ou desigualdade (`!=`).

```
dois_numero <- 2
dois_texto <- "dois"
# Vamos testar de o numero dois é o mesmo que o dois "por extenso"
dois_numero == dois_texto
```

```
## [1] FALSE
```

```
# Vamos testar de o numero dois é diferente do dois "por extenso"
dois_numero != dois_texto
```

```
## [1] TRUE
```

O R não entende que “dois” é a forma “por extenso” do número 2. Para o R, há um número 2 e há uma palavra, cujo significado o R desconhece. Então, na comparação entre esse inteiro e essa string, o R entende que há uma diferença (de tipos). Por isso, ele afirma que são diferentes (ou não iguais).

```
dois_numero <- 2
dois_texto <- "2"
# Vamos testar de o numero dois é o mesmo que o dois "caractere"
# O resultado é verdadeiro porque apesar dos tipos serem diferentes, o conteúdo é o mesmo
dois_numero == dois_texto
```

```
## [1] TRUE
```

Existe outro operador de igualdade (`===`), mas que não será tratado agora.

3.4 Datas


```
today <- Sys.Date()  
print(today)
```

```
## [1] "2021-02-18"
```

```
today <- format(today, "%d/%m/%Y")  
print(today)
```

```
## [1] "18/02/2021"
```

```
today2 <- format(Sys.Date(), "%d/%m/%Y")  
print(today2)
```

```
## [1] "18/02/2021"
```


Chapter 4

Estruturas de Dados

Pode ser desejável juntar vários valores em um único objeto gravado na memória usada pelo R. Por exemplo, você pode querer criar um objeto que contenha o nome das empresas que tiveram as maiores ofertas públicas iniciais na bolsa de valores brasileira em 2020; ou você pode querer a relação dos nomes e dos valores. Dependendo dos tipos de dados envolvidos, a forma de fazer isso pode mudar.

4.1 Vetores

Quando a sua relação trata de apenas um conjunto de valores de mesmo tipo (todos numéricos, todos textuais ou todos lógicos), pode ser formada como um **vetor**. Por exemplo, poderia haver um vetor com o nome dos CEOs das empresas do setor bancário; ou um vetor com o salário desses CEOs.

Entretanto, se estivéssemos falando de uma relação envolvendo os nomes e os salários, seria inadequado usarmos o vetor, porque agora teríamos duas dimensões, além de dois tipos: uma mesma pessoa (o CEO) possui duas características: o nome (textual / string) e o salário (numérica / num).

A criação de vetores é feita com o uso do `c()`, em que os valores que devem ser atribuídos ao vetor são incluídos dentro dos parênteses e separados por vírgula.

```
vetorNumerico = c(1,2,3,4,5)
vetorLogico = c(TRUE, TRUE, FALSE,FALSE)
vetorTextual = c("Eu", "Nasci", "há", "10.000", "anos", "atrás")
```

Note que todos os vetores foram criados com dados de mesmo tipo. O vetor numérico tem apenas números inteiros (1 a 5); o vetor lógico tem apenas Verdadeiro (TRUE) e Falso (FALSE); o vetor de texto tem apenas elementos

textuais (mesmo o número está entre aspas, o que força que aqueles dígitos sejam lidos como “texto”).

Caso você tente misturar tipos, o R irá forçar a padronização. Por exemplo, o que aconteceria se você tentasse criar um vetor com números e textos?

```
vetorNumTex = c(1,2,"três","4",5)
print(vetorNumTex)
```

```
## [1] "1"      "2"      "três" "4"      "5"
```

Todos os elementos apresentados estão entre aspas. Como o R não tem uma forma de transformar “três” em um número, ele converte todos os outros números em texto. Por isso, 1 vira “1”, 2 vira “2” e 5 vira “5”.

Teste no seu console a seguinte operação: “3” + 1

O R informará um erro. “3” não é um número, mas um texto (por causa das aspas). Não é possível somar um texto com um número. É o mesmo que você tentar somar “banana” com 97. Não dá.

4.2 Matrizes

A matriz é um conjunto de dados do mesmo tipo, ordenados em linhas (filas horizontais) e colunas (filas verticais). É criada pelo `matrix()`, devendo ser informados os dados que irão compor a matriz (normalmente dispostos em um vetor), o número de linhas (`nrow`), o número de colunas (`ncol`) e se o preenchimento deve ser feito por linha (`byrow = TRUE`) ou por coluna (`byrow = FALSE`).

```
# Se a matriz dos inteiros de 1 a 6, dispostos em 3 linhas e 2 colunas, preenchida pela
mat = matrix(c(1:6), nrow = 3, ncol = 2, byrow = TRUE)
print(mat)
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    3    4
## [3,]    5    6
```

Ou seja, os números foram distribuídos até o fim da primeira linha, depois até o fim da segunda linha e depois até o fim da terceira (e última) linha.

```
# Se a matriz dos inteiros de 1 a 6, dispostos em 3 linhas e 2 colunas, preenchida pela ordem da
mat = matrix(c(1:6), nrow = 3, ncol = 2, byrow = FALSE)
print(mat)
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

Em contraste, aqui o preenchimento foi por colunas. Foi preenchida toda a primeira coluna e depois toda a segunda coluna.

Se já tivéssemos um vetor salvo na memória, poderíamos usá-lo para compor a matriz.

```
vetorParaMatriz <- c("eu", "nasci", "há", "dez mil", "anos", "atrás")
mat = matrix(vetorParaMatriz, nrow = 3, ncol = 2, byrow = TRUE)
print(mat)
```

```
##      [,1] [,2]
## [1,] "eu"  "nasci"
## [2,] "há"  "dez mil"
## [3,] "anos" "atrás"
```

Você reparou que aparecem valores entre colchetes e separados por vírgulas? Eles indicam a posição na matriz.

[m,n]	Posição	Elementos
[1,]	Primeira Linha	“eu”, “nasci”
[2,]	Segunda Linha	“há”, “dez mil”
[,1]	Primeira Coluna	“eu”, “há”, “anos”
[,2]	Segunda Coluna	“nasci”, “dez mil”, “atrás”
[1,1]	Primeira Linha, Primeira Coluna	“eu”
[1,2]	Primeira Linha, Segunda Coluna	“nasci”
[2,1]	Segunda Linha, Primeira Coluna	“há”
[2,2]	Segunda Linha, Segunda Coluna	“dez mil”

Eventualmente, pode ser interessante informar os nomes para as linhas e para as colunas. Por exemplo, se você estiver precisando mapear a distância entre as unidades de uma empresa, seria importante dizer quais são as filiais.

```
vetorFiliais <- c("Natal", "Parnamirim", "Macaíba")
vetorDistancias <- c(0, 10, 12, 10, 0, 2, 12, 2, 0)
```

Por isso, vamos incluir o argumento

```
dimnames = list(row_names, col_names)
```

dimnames significa nome das dimensões, que é informado por meio de dois vetores: o vetor com o nomes das linhas (*row_names*) e o vetor com o nome das colunas (*col_names*).

No nosso exemplo, as linhas e as colunas devem possuir os mesmos nomes (porque queremos estabelecer uma matriz com a distância entre cada unidade da empresa). Por isso, trocaremos *row_names* por *vetorFiliais* e **também** trocaremos *col_names* por *vetorFiliais*.

```
distanciaFiliais = matrix(vetorDistancias, nrow = 3, ncol = 3, byrow = TRUE,
                           dimnames = list(vetorFiliais, vetorFiliais))
```

```
distanciaFiliais
```

```
##           Natal Parnamirim Macaíba
## Natal           0           10          12
## Parnamirim      10           0           2
## Macaíba         12           2           0
```

Por que o elemento [1,1] da matriz foi 0? A distância de uma unidade para ela mesmo, por definição, é zero. Para solicitar a distância entre a filial Natal e a filial Parnamirim:

```
distanciaFiliais["Natal", "Parnamirim"]
```

```
## [1] 10
```

Perceba que, nessa matriz, a distância entre Natal - Parnamirim é a mesma que Parnamirim - Natal. Se pensarmos em termos de rodovias, isso pode não ser necessariamente verdade, pois em casos específicos pode ser necessário um deslocamento adicional em um dos sentidos (por exemplo, ir adiante x quilômetros para conseguir fazer um retorno). No caso dessa matriz, solicitar o valor como:

```
distanciaFiliais["Natal", "Parnamirim"]
```

ou

```
distanciaFiliais[“Parnamirim”, “Natal”]
```

produziria o mesmo resultado e, nesse caso, é irrelevante. Entretanto, é importante ter em mente que isso nem sempre será verdade. Além disso, se quiséssemos saber a distância da filial Natal para todas as demais, deveríamos informar apenas a linha, omitindo a informação após a vírgula.

```
distanciaFiliais[“Natal”,]
```

```
##      Natal Parnamirim   Macaíba
##           0         10        12
```

Alternativamente, sabendo que Natal é a primeira linha da matriz, também seria possível executar

```
distanciaFiliais[1,]
```

```
##      Natal Parnamirim   Macaíba
##           0         10        12
```

A segunda opção é particularmente importante quando as filas (linhas ou colunas) não estão nomeadas.

4.3 Data frames

O data frame é similar a matrix, mas admite colunas com tipos diferentes. Nesse caso, é possível incluir uma coluna para o nome de uma empresa (string), uma coluna para o CNPJ (string), uma coluna para o número de empregados (integer), uma coluna para o faturamento (float) e uma coluna para identificar se as demonstrações já foram encerradas ou se ainda estão pendentes (boolean).

O data frame é criado por meio do `data.frame()`, onde são passados os valores que irão entrar na estrutura. Uma possibilidade é passarmos vetores que representem as colunas. Assim, usaríamos um vetor para a empresa, um vetor para o CNPJ, um vetor para os empregados e um vetor para o faturamento.

Não é necessário informar a quantidade de linhas ou colunas, já que o R não precisará organizar os dados (no caso da matriz, foi necessário informar como ela deveria ser preenchida). Os vetores serão incluídos como colunas na mesma ordem em que forem aparecendo dentro do `data.frame()`.

```
# Ainda discutiremos o que são factors, mas por enquanto insiram esse argumento.
clientes <- data.frame(empresa = c("Marte S.A.", "Deimos S.A.", "Phobos S.A."),
                        cnpj = c("00.000.001/0001-00", "00.000.011/0001-10", "99.000.001.001.001/0001-55"),
                        empregados = c(10, 20, 25),
                        encerramento = c(FALSE, FALSE, TRUE),
                        stringsAsFactors = FALSE)

clientes
```

```
##      empresa      cnpj empregados encerramento
## 1 Marte S.A. 00.000.001/0001-00      10      FALSE
## 2 Deimos S.A. 00.000.011/0001-10      20      FALSE
## 3 Phobos S.A. 99.000.001/0001-55      25       TRUE
```

Por conta dessa permissão para uso de tipos diferentes em uma mesma estrutura, os data frames são consideravelmente importantes. Para solicitar uma fila específica (linha ou coluna) ou um elemento, é possível usar a mesma notação das matrizes.

```
clientes[1,]
```

```
##      empresa      cnpj empregados encerramento
## 1 Marte S.A. 00.000.001/0001-00      10      FALSE
```

```
# para gerar uma relação com os CNPJs:
clientes[,2]
```

```
## [1] "00.000.001/0001-00" "00.000.011/0001-10" "99.000.001/0001-55"
```

Da mesma forma que fizemos com as matrizes, podemos solicitar uma coluna com base em seu nome.

```
clientes[, "cnpj"]
```

```
## [1] "00.000.001/0001-00" "00.000.011/0001-10" "99.000.001/0001-55"
```

Perceba que nesse caso as linhas não foram nomeadas. No caso dos data frames, é bem comum que o identificador seja inserido como um dado

O data frame ainda admite uma forma adicional de solicitar uma coluna, informando o nome do objeto seguido do cifrão (\$) e do nome da coluna.


```
clientes$cnpj
```

```
## [1] "00.000.001/0001-00" "00.000.011/0001-10" "99.000.001/0001-55"
```

Perceba como a notação do R evolui. Quando resolvemos solicitar um `clientes$cnpj`, estamos usando um procedimento equivalente a `clientes[, "cnpj"]`. Como poderíamos fazer para usando a notação com cifrão, solicitar o terceiro elemento dessa relação?

```
clientes$cnpj[3]
```

```
## [1] "99.000.001/0001-55"
```

Não precisamos informar dois elementos dentro dos colchetes `[,]` porque um deles foi informado anteriormente (no caso, a coluna *cnpj*)! Assim, quando informarmos o `[3]` estamos automaticamente nos referindo à linha.

4.4 Listas

4.5 Factors

Chapter 5

Trabalhando com variáveis no R Base

5.1 Manuseio de variáveis

5.2 Funções numéricas básicas

5.3 Funções textuais básicas

Chapter 6

Criando sua própria função

6.1 Comentários

Chapter 7

Methods

We describe our methods in this chapter.

Chapter 8

Applications

Some *significant* applications are demonstrated in this chapter.

8.1 Example one

8.2 Example two

Chapter 9

Final Words

We have finished a nice book.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.21.