

# Importação de dados com tidyverse :: FOLHA DE REFERÊNCIA



## Ler dados tabulados com readr

**read\_\***(file, col\_names = TRUE, col\_types = NULL, col\_select = NULL, id = NULL, locale, n\_max = Inf, skip = 0, na = c("", "NA"), guess\_max = min(1000, n\_max), show\_col\_types = TRUE) Ver ?**read\_delim**

A|B|C  
1|2|3  
4|5|NA

A	B	C
1	2	3
4	5	NA

**read\_delim**("file.txt", delim = "|") Lê arquivos com qualquer delimitador. Se o delimitador não for especificado, tenta adivinhar automaticamente.  
Para criar file.txt, execute: write\_file("A|B|C\n1|2|3\n4|5|NA", file = "file.txt")

A,B,C  
1,2,3  
4,5,NA

A	B	C
1	2	3
4	5	NA

**read\_csv**("file.csv") Lê arquivo separado com vírgula com ponto como separador decimal.  
write\_file("A,B,C\n1,2,3\n4,5,NA", file = "file.csv")

A;B;C  
1;5;2;3  
4;5;5;NA

A	B	C
1.5	2	3
4.5	5	NA

**read\_csv2**("file2.csv") Lê arquivo separado por ponto-e-vírgula com vírgula como separador decimal.  
write\_file("A;B;C\n1,5;2;3\n4,5;5;NA", file = "file2.csv")

A B C  
1 2 3  
4 5 NA

A	B	C
1	2	3
4	5	NA

**read\_tsv**("file.tsv") Lê arquivo separado com tab. Ver também **read\_table()**.  
**read\_fwf**("file.tsv", fwf\_widths(c(2, 2, NA))) Lê arquivo com largura fixa.  
write\_file("A\tB\tC\n1\t2\t3\n4\t5\tNA\n", file = "file.tsv")

### ARGUMENTOS ÚTEIS DE LEITURA

A	B	C
1	2	3
4	5	NA

#### Sem cabeçalho

**read\_csv**("file.csv", col\_names = FALSE)

1	2	3
4	5	NA

#### Pula linhas

**read\_csv**("file.csv", skip = 1)

x	y	z
A	B	C
1	2	3
4	5	NA

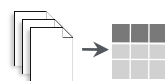
#### Fornecer cabeçalho

**read\_csv**("file.csv",  
col\_names = c("x", "y", "z"))

A	B	C
1	2	3

#### Lê subconjunto de linhas

**read\_csv**("file.csv", n\_max = 1)



#### Lê vários arquivos em um única tabela

**read\_csv**(c("f1.csv", "f2.csv", "f3.csv"),  
id = "origin\_file")

A	B	C
NA	2	3
4	5	NA

#### Lê valor como ausente (NA)

**read\_csv**("file.csv", na = c("1"))

A;B;C  
1,5;2;3,0

#### Especifica separador decimal

**read\_delim**("file2.csv", locale =  
locale(decimal\_mark = ","))

## Salvar dados com readr

**write\_\***(x, file, na = "NA", append, col\_names, quote, escape, eol, num\_threads, progress)

A	B	C
1	2	3
4	5	NA

A,B,C  
1,2,3  
4,5,NA

**write\_delim**(x, file, delim = " ") Grava arquivos com delimitador.

**write\_csv**(x, file) Grava arquivo separado por vírgula.

**write\_csv2**(x, file) Grava arquivo separado por ponto-e-vírgula.

**write\_tsv**(x, file) Grava arquivo separado por tab.

Um dos primeiros passos de um projeto é importar dados externos para o R. Os dados são frequentemente armazenados em formatos tabulados como arquivos .csv ou planilhas.



A página da frente desta folha de resumo mostra como importar e salvar arquivos texto usando o pacote readr.



O verso mostra como importar planilhas do Excel usando readxl ou planilhas do Google Sheets usando o googlesheets4.

### OUTROS FORMATOS DE DADOS

Experimente um desses pacotes para importar outros formatos de dados:

- **haven** – arquivos SPSS, Stata e SAS
- **DBI** – bando de dados
- **jsonlite** – json
- **xml2** – XML
- **httr** – Web APIs
- **rvest** – HTML (Web Scraping)
- **readr::read\_lines()** – dados texto

## Especificação de Coluna com readr

Especificação de coluna define qual o tipo de dado cada coluna de um arquivo será importada. Por padrão readr gera uma especificação quando o arquivo é importado e um resumo na saída.

**spec(x)** Extrai a especificação de coluna completa de um data frame importado.

```
spec(x)
# cols(
#   idade = col_integer(),
#   sexo = col_character(),
#   salário = col_double()
# )
```

idade é um inteiro

salário é um double (numérico)

sexo é um caractere

### ARGUMENTOS ÚTEIS PARA COLUNAS

#### Esconde as mensagens de especificação

**read\_\***(file, show\_col\_types = FALSE)

#### Seleciona colunas para importar

Use names, position, or selection helpers.

**read\_\***(file, col\_select = c(age, earn))

#### Adivinha tipo da coluna

To guess a column type, **read\_\***() looks at the first 1000 rows of data. Increase with **guess\_max**.  
**read\_\***(file, guess\_max = Inf)

### TIPOS DE COLUNAS

Cada tipo de coluna tem uma função e uma string de abreviação correspondente.

- **col\_logical()** - "l"
- **col\_integer()** - "i"
- **col\_double()** - "d"
- **col\_number()** - "n"
- **col\_character()** - "c"
- **col\_factor(levels, ordered = FALSE)** - "f"
- **col\_datetime(format = "")** - "T"
- **col\_date(format = "")** - "D"
- **col\_time(format = "")** - "t"
- **col\_skip()** - "-", "\_"
- **col\_guess()** - "?"

### DEFINE ESPECIFICAÇÃO DAS COLUNAS

#### Define o tipo padrão

**read\_csv**(  
file,  
col\_type = list(default = col\_double())  
)

#### Usa um tipo ou a string de abreviação

**read\_csv**(  
file,  
col\_type = list(x = col\_double(), y = "l", z = "\_")  
)

#### Usa um única string de abreviação

# col types: skip, guess, integer, logical, character  
**read\_csv**(  
file,  
col\_type = "\_?ilc"  
)

# Importando Planilhas

## com readxl

### LER ARQUIVOS EXCEL

	A	B	C	D	E
1	x1	x2	x3	x4	x5
2	x		z	8	
3	y	7		9	10

s1

**read\_excel(path, sheet = NULL, range = NULL)** Lê um arquivo .xls ou .xlsx baseado na extensão. Ver primeira página para mais argumentos de leitura. Ver **read\_xls()** e **read\_xlsx()**.  
`read_excel("excel_file.xlsx")`

### LER PLANILHAS

A	B	C	D	E

s1 s2 s3

**read\_excel(path, sheet = NULL)** Especifica qual planilha ler, por nome ou posição.  
`read_excel(path, sheet = 1)`  
`read_excel(path, sheet = "s1")`

s1	s2	s3
----	----	----

**excel\_sheets(path)** Pega os nomes das planilhas em um vetor.  
`excel_sheets("excel_file.xlsx")`

A	B	C	D	E

s1 s2 s3

Para ler várias planilhas:

1. Pegue do arquivo um vetor com nomes das planilhas.
2. Defina o vetor de nomes como nomes das planilhas.
3. Use `purrr::map_dfr()` para ler vários arquivos e gerar um único data frame.

```
path <- "caminho_do_arquivo.xlsx"
path %>% excel_sheets() %>%
  set_names() %>%
  map_dfr(read_excel, path = path)
```

### OUTROS PACOTES ÚTEIS PARA

Para funções de gravar dados em arquivos Excel, veja:

- **openxlsx**
- **writexl**

Para trabalhar com dados do Excel em formato não-tabular, veja:

- **tidyxl**



### ESPECIFICAÇÃO DE COLUNA - READXL

Especificação de coluna define o tipo de dado que cada coluna do arquivo terá após importada.

Use o argumento **col\_types** da **read\_excel()** para definir as especificações das colunas.

#### Adivinhar tipos de colunas

Para adivinhar o tipo da coluna, **read\_excel()** lê as 100 primeiras linhas. Aumente com o argumento **guess\_max**.

`read_excel(path, guess_max = Inf)`

#### Define todas as colunas com o mesmo tipo, ex. caractere

`read_excel(path, col_types = "text")`

#### Define cada coluna individualmente

```
read_excel(
  path,
  col_types = c("text", "guess", "guess", "numeric")
)
```

### TIPOS DE COLUNAS

logical	numeric	text	date	list
TRUE	2	hello	1947-01-08	hello
FALSE	3.45	world	1956-10-21	1

- skip
- guess
- logical
- numeric
- text
- date
- list

Use **list** para colunas que tem múltiplos tipos de dados. Veja sobre colunas de lista em **tidyr** e **purrr**.

### ESPECIFICAÇÃO DE CÉLULAS PARA READXL E GOOGLESHEETS4

	A	B	C	D	E
1	1	2	3	4	5
2	x		y	z	
3	6	7		9	10

s1

Use o argumento **range** da **readxl::read\_excel()** ou **googlesheets4::read\_sheet()** para ler um subconjunto de células de uma planilha.

`read_excel(path, range = "Sheet1!B1:D2")`  
`read_sheet(ss, range = "B1:D2")`

Use também o argumento **range** com funções de especificação de células **cell\_limits()**, **cell\_rows()**, **cell\_cols()** e **anchored()**.

## com googlesheets4

### LER PLANILHAS GOOGLE

	A	B	C	D	E
1	x1	x2	x3	x4	x5
2	x		z	8	
3	y	7		9	10

s1

**read\_sheet(ss, sheet = NULL, range = NULL)** Lê um arquivo com URL, um ID, ou um objeto drible do pacote googledrive. Veja página da frente para mais argumentos de leitura. Mesmo que **range\_read()**.

### METADATA DAS PLANILHAS

URLs estão na seguinte forma:  
`https://docs.google.com/spreadsheets/d/  
ID_ARQUIVO/edit#gid=ID_PLANILHA`

**gs4\_get(ss)** Lê o metadado da planilha.

**gs4\_find(...)** Lê dados de todos os arquivos de planilhas.

**sheet\_properties(ss)** Retorna um tibble com propriedades de cada planilha. Veja também **sheet\_names()**.

### GRAVAR PLANILHAS DO GOOGLE

	A	B	C
1	x	4	
2	y	5	
3	z	6	

s1

A	B	C	D
1			
2			

s1

x1	x2	x3
2	y	5
3	z	6

s1

**write\_sheet(data, ss = NULL, sheet = NULL)** Wgrava um data frame em um planilha nova ou já existente.

**gs4\_create(name, ..., sheets = NULL)** Cria uma nova planilha com um vetor de nomes, um data frame ou uma lista de data frame nomeada.

**sheet\_append(ss, data, sheet = 1)** Adiciona uma linha ao final da planilha.



### ESPECIFICAÇÃO DE COLUNAS - GOOGLESHEETS4

Especificação de coluna define o tipo de dado que cada coluna do arquivo terá após importada.

Use o argumento **col\_types** da **read\_sheet()/range\_read()** para definir as especificações das colunas.

#### Adivinhar tipos de colunas

Para adivinhar o tipo da coluna **read\_sheet()/range\_read()** lê as primeiras 100 linhas. Aumente com **guess\_max**.  
`read_sheet(path, guess_max = Inf)`

#### Define todas as colunas com o mesmo tipo, ex. caractere

`read_sheet(path, col_types = "c")`

#### Define cada coluna individualmente

# col types: skip, guess, integer, logical, character  
`read_sheets(ss, col_types = "?_?ilc")`

### TIPOS DE COLUNAS

I	n	c	D	L
TRUE	2	hello	1947-01-08	hello
FALSE	3.45	world	1956-10-21	1

- skip - "\_" or "-"
- guess - "?"
- logical - "l"
- integer - "i"
- double - "d"
- numeric - "n"
- date - "D"
- datetime - "T"
- character - "c"
- list-column - "L"
- cell - "C" Returns list of raw cell data.

Use **list** para colunas que tem múltiplos tipos de dados. Veja sobre colunas de lista em **tidyr** e **purrr**.

### OPERAÇÕES EM NÍVEL DE ARQUIVO

**googlesheets4** também oferece várias formas de modificar outros aspectos das planilhas (ex. congelar linhas, definir largura das colunas, gerenciar planilhas, etc). Veja [googlesheets4.tidyverse.org](https://tidyverse.org/packages/googlesheets4/) para maiores informações.

Para operações de arquivos (ex. renomear, compartilhar, mover para outra pasta, etc), veja mais sobre o pacote **googledrive** do tidyverse em: [googledrive.tidyverse.org](https://tidyverse.org/packages/googledrive/).

