

O OPERADOR PIPE

O operador `%>%` (pipe) foi uma das grandes revoluções recentes do R, tornando a leitura de códigos mais lógica, fácil e compreensível. Ele foi introduzido por Stefan Milton Bache no pacote `magrittr` e já existem diversos pacotes construídos para facilitar a sua utilização.

Para começar a utilizar o pipe, instale e carregue o pacote `magrittr`.

```
In [69]: install.packages("magrittr")  
library(magrittr)
```

Warning message:

"package 'magrittr' is in use and will not be installed"

A ideia do operador `%>%` (pipe) é bem simples: usar o valor resultante da expressão do lado esquerdo como primeiro argumento da função do lado direito.

Nos casos mais simples, o pipe parece não trazer grandes vantagens. Agora, veja como fica um caso mais etapas.

Vamos calcular a raiz quadrada da soma dos valores de 1 a 4. Primeiro, sem o pipe.

```
In [70]: x <- c(1, 2, 3, 4)  
         sqrt(sum(x))
```

3.16227766016838

Agora com o pipe.

```
In [71]: x %>% sum() %>% sqrt()
```

3.16227766016838

O caminho que o código `x %>% sum %>% sqrt` seguiu foi

- enviar o objeto `x` como argumento da função `sum()` e
- em seguida, enviar a saída da expressão `sum(x)` como argumento da função `sqrt()`.

Observe que escrevemos o código na mesma ordem das operações. A utilização de parênteses após o nome das funções não é necessário, mas recomendável.

Exemplo com o dataset airquality

Vamos fazer mais um exemplo do pipe, mas antes vamos conhecer o dataset airquality

In [72]: `head(airquality)`

A data.frame: 6 × 6

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6

Às vezes, queremos que o resultado do lado esquerdo vá para outro argumento do lado direito que não o primeiro. Para isso, utilizamos um `.` como marcador.

Queremos que o dataset seja recebido pelo segundo argumento (`data=`) da função `"lm"`.

```
In [73]: airquality %>%  
  na.omit %>%  
  lm(Ozone ~ Wind + Temp + Solar.R, data = .) %>%  
  summary
```

Call:

```
lm(formula = Ozone ~ Wind + Temp + Solar.R, data = .)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-40.485	-14.219	-3.551	10.097	95.619

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-64.34208	23.05472	-2.791	0.00623	**
Wind	-3.33359	0.65441	-5.094	1.52e-06	***
Temp	1.65209	0.25353	6.516	2.42e-09	***
Solar.R	0.05982	0.02319	2.580	0.01124	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.18 on 107 degrees of freedom

Multiple R-squared: 0.6059, Adjusted R-squared: 0.5948

F-statistic: 54.83 on 3 and 107 DF, p-value: < 2.2e-16

MANIPULAÇÃO DE DADOS

INTRODUÇÃO

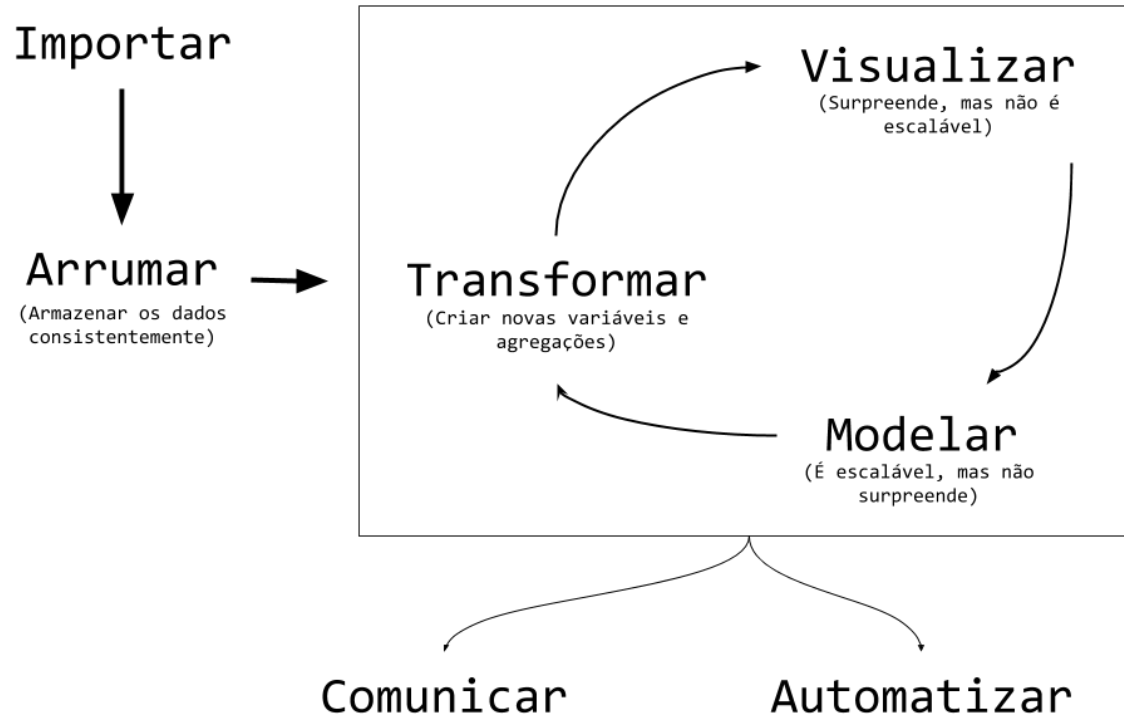
Esta AULA trataremos do tema manipulação de dados.

Trata-se de uma tarefa dolorosa e demorada, tomando muitas vezes a maior parte do tempo de uma análise estatística.

Essa etapa é essencial em qualquer análise de dados e, apesar de negligenciada pela academia, é decisiva para o sucesso de estudos aplicados.

Usualmente, o cientista de dados parte de uma base “crua” e a transforma até obter uma base de dados analítica, que, a menos de transformações simples, está preparada para passar por análises estatísticas.

A figura abaixo mostra a fase de “disputa” com os dados (data wrangling) para deixá-los no formato analítico.



Um conceito importante para obtenção de uma base analítica é o data tidying, ou arrumação de dados. Uma base é considerada tidy se

1. Cada linha da base representa uma observação.
2. Cada coluna da base representa uma variável.

A base de dados analítica é estruturada de tal forma que pode ser colocada diretamente em ambientes de modelagem estatística ou de visualização. Nem sempre uma base de dados analítica está no formato tidy, mas usualmente são necessários poucos passos para migrar de uma para outra. A filosofia tidy é a base do tidyverse.

Os principais pacotes encarregados da tarefa de estruturar os dados são o dplyr e o tidyr. Eles serão o tema desse tópico. Instale e carregue os pacotes utilizando:

```
In [1]: install.packages("dplyr")  
install.packages("tidyr")
```

Installing package into 'C:/Users/domin/OneDrive/Documentos/R/win-library/4.0'
(as 'lib' is unspecified)

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\domin\AppData\Local\Temp\RtmpYpNT33\downloaded_packages

Installing package into 'C:/Users/domin/OneDrive/Documentos/R/win-library/4.0'
(as 'lib' is unspecified)

package 'tidyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\domin\AppData\Local\Temp\RtmpYpNT33\downloaded_packages

```
In [2]: library(dplyr)
library(tidyr)
```

Warning message:

"package 'dplyr' was built under R version 4.0.3"

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

Warning message:

"package 'tidyr' was built under R version 4.0.3"

TRABALHANDO COM TIBBLES

Uma tibble nada mais é do que um `data.frame`, mas com um método de impressão mais adequado.

As tibbles são parte do pacote `tibble`. Assim, para começar a usá-las, instale e carregue o pacote.

```
In [78]: install.packages("tibble")  
library(tibble)
```

Warning message:

"package 'tibble' is in use and will not be installed"

vamos trabalhar com uma base de filmes do IMDB. Essa base pode ser baixada [clicando aqui \(https://github.com/curso-r/site-v2/raw/master/content/material/importacao/data/imdb.rds\)](https://github.com/curso-r/site-v2/raw/master/content/material/importacao/data/imdb.rds).

```
In [79]: install.packages("readr")
```

Warning message:

"package 'readr' is in use and will not be installed"

```
In [80]: library(readr)
```

```
In [3]: imdb <- readr::read_rds("imdb.rds")
```

Assim, utilizaremos o objeto `imdb` para acessar os dados.

In [4]: `head(imdb)`

A tibble: 6 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250
John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258
Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260

Veja que, no RStudio, apenas as dez primeiras linhas da tibble são impressas na tela. Além disso, as colunas que não couberem na largura do console serão omitidas. Também são apresentadas a dimensão da tabela e as classes de cada coluna.

Como estou usando o Jupyter Notebook a formatação pode ser um pouco diferente

In [5]: `head(imdb)`

A tibble: 6 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250
John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258
Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260

O PACOTE DPLYR

O dplyr é o pacote mais útil para realizar transformação de dados, aliando simplicidade e eficiência de uma forma elegante.

Os scripts em R que fazem uso inteligente dos verbos dplyr e as facilidades do operador pipe tendem a ficar mais legíveis e organizados sem perder velocidade de execução.

As principais funções do dplyr são:

- `filter()` - filtra linhas
- `select()` - seleciona colunas
- `arrange()` - ordena a base
- `mutate()` - cria/modifica colunas
- `group_by()` - agrupa a base
- `summarise()` - sumariza a base

Todas essas funções seguem as mesmas características:

- O input é sempre uma tibble e o output é sempre um tibble.
- Colocamos o tibble no primeiro argumento e o que queremos fazer nos outros argumentos.
- A utilização é facilitada com o emprego do operador %>%.
- O pacote faz uso extensivo de NSE (non standard evaluation).

As principais vantagens de se usar o dplyr em detrimento das funções do R base são:

- Manipular dados se torna uma tarefa muito mais simples.
- O código fica mais intuitivo de ser escrito e mais simples de ser lido.
- O pacote dplyr utiliza C e C++ por trás da maioria das funções, o que geralmente torna o código mais eficiente.
- É possível trabalhar com diferentes fontes de dados, como bases relacionais (SQL) e data.table. Agora, vamos avaliar com mais detalhes as principais funções do pacote dplyr.

FILTRANDO LINHAS

A função `filter()` filtra linhas. Ela é semelhante à função `subset()`, do R base. O código abaixo retorna apenas filmes com nota maior que nove.

```
In [6]: imdb %>%  
       filter(nota_imdb > 9)
```

A tibble: 3 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita	nota_imdb	likes_facebook
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<dbl>	<int>
The Shawshank Redemption	1994	Frank Darabont	142	Color	Crime Drama	USA	A partir de 18 anos	25000000	28341469	9.3	108000
The Godfather	1972	Francis Ford Coppola	175	Color	Crime Drama	USA	A partir de 18 anos	6000000	134821952	9.2	43000
Kickboxer: Vengeance	2016	John Stockwell	90	NA	Action	USA	Outros	17000000	NA	9.1	0

Para fazer várias condições, use os operadores lógicos & e | ou separe filtros entre vírgulas. Vamos selecionar filmes posteriores a 2010 e com nota maior que 8.5

```
In [7]: imdb %>%
  filter(ano > 2010 & nota_imdb > 8.5)
```

A tibble: 5 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita	nota_imdb	lik
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<dbl>	<ir
Interstellar	2014	Christopher Nolan	169	Color	Adventure Drama Sci-Fi	USA	A partir de 13 anos	165000000	187991439	8.6	34
Running Forever	2015	Mike Mayhall	88	Color	Family	USA	Outros	5000000	NA	8.6	49
A Beginner's Guide to Snuff	2016	Mitchell Altieri	87	Color	Comedy Horror Thriller	USA	Outros	NA	NA	8.7	8
Kickboxer: Vengeance	2016	John Stockwell	90	NA	Action	USA	Outros	17000000	NA	9.1	0
Butterfly Girl	2014	Cary Bell	78	Color	Documentary	USA	Outros	180000	NA	8.7	88

Agora vamos selecionar filems cuja receita foi maior que o orçamento OU que a nota foi maior que nove

```
In [8]: imdb %>%
  filter(receita > orcamento | nota_imdb > 9)
```

A tibble: 1762 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	recei
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237000000	7605
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300000000	3094
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250000000	4481
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258000000	3365
Avengers: Age of Ultron	2015	Joss Whedon	141	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000	4589
Batman v Superman: Dawn of Justice	2016	Zack Snyder	183	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000	3302
Pirates of the Caribbean: Dead Man's Chest	2006	Gore Verbinski	151	Color	Action Adventure Fantasy	USA	A partir de 13 anos	225000000	4230
Man of Steel	2013	Zack Snyder	143	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	225000000	2910
The Avengers	2012	Joss Whedon	173	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	220000000	6232
The Amazing Spider-Man	2012	Marc Webb	153	Color	Action Adventure Fantasy	USA	A partir de 13 anos	230000000	2620
The Hobbit: The Desolation of Smaug	2013	Peter Jackson	186	Color	Adventure Fantasy	USA	A partir de 13 anos	225000000	2583
Titanic	1997	James Cameron	194	Color	Drama Romance	USA	A partir de 13 anos	200000000	6586

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
Captain America: Civil War	2016	Anthony Russo	147	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000	4071
Jurassic World	2015	Colin Trevorrow	124	Color	Action Adventure Sci-Fi Thriller	USA	A partir de 13 anos	150000000	6521
Spider-Man 2	2004	Sam Raimi	135	Color	Action Adventure Fantasy Romance	USA	A partir de 13 anos	200000000	3733
Iron Man 3	2013	Shane Black	195	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000	4089
Alice in Wonderland	2010	Tim Burton	108	Color	Adventure Family Fantasy	USA	Livre	200000000	3341
Monsters University	2013	Dan Scanlon	104	Color	Adventure Animation Comedy Family Fantasy	USA	Outros	200000000	2684
Transformers: Revenge of the Fallen	2009	Michael Bay	150	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000	4020
Transformers: Age of Extinction	2014	Michael Bay	165	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	210000000	2454
Oz the Great and Powerful	2013	Sam Raimi	130	Color	Adventure Family Fantasy	USA	Livre	215000000	2349
The Amazing Spider-Man 2	2014	Marc Webb	142	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	200000000	2028
TRON: Legacy	2010	Joseph Kosinski	125	Color	Action Adventure Sci-Fi	USA	Livre	170000000	1720
Toy Story 3	2010	Lee Unkrich	103	Color	Adventure Animation Comedy Family Fantasy	USA	Outros	200000000	4149
Furious 7	2015	James Wan	140	Color	Action Crime Thriller	USA	A partir de 13 anos	190000000	3500
World War Z	2013	Marc Forster	123	Color	Action Adventure Horror Sci-Fi Thriller	USA	A partir de 13 anos	190000000	2023
X-Men: Days of Future Past	2014	Bryan Singer	149	Color	Action Adventure Fantasy Sci-Fi Thriller	USA	A partir de 13 anos	200000000	2339
Star Trek Into Darkness	2013	J.J. Abrams	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	190000000	2287
Transformers: Dark of the Moon	2011	Michael Bay	154	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	195000000	3523

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
Indiana Jones and the Kingdom of the Crystal Skull	2008	Steven Spielberg	122	Color	Action Adventure Fantasy	USA	A partir de 13 anos	185000000	317000000
...
Side Effects	2013	Steven Soderbergh	106	Color	Crime Drama Thriller	USA	A partir de 18 anos	30000000	321500000
She's Gotta Have It	1986	Spike Lee	88	Black and White	Comedy Romance	USA	A partir de 18 anos	175000	7137000
Another Earth	2011	Mike Cahill	92	Color	Drama Romance Sci-Fi	USA	A partir de 13 anos	100000	1316000
Sweet Sweetback's Baadasssss Song	1971	Melvin Van Peebles	97	Color	Crime Drama Thriller	USA	A partir de 18 anos	500000	1518000
Tadpole	2000	Gary Winick	78	Color	Comedy Drama Romance	USA	A partir de 13 anos	150000	2882000
The Texas Chain Saw Massacre	1974	Tobe Hooper	88	Color	Horror Thriller	USA	A partir de 18 anos	83532	3085000
Roger & Me	1989	Michael Moore	91	Color	Documentary	USA	A partir de 18 anos	160000	6706000
Your Sister's Sister	2011	Lynn Shelton	90	Color	Comedy Drama	USA	A partir de 18 anos	125000	1573000
Facing the Giants	2006	Alex Kendrick	111	Color	Drama Sport	USA	Livre	100000	1017000
The Gallows	2015	Travis Cluff	81	Color	Horror Thriller	USA	A partir de 18 anos	100000	2275000
Over the Hill to the Poorhouse	1920	Harry F. Millarde	110	Black and White	Crime Drama	USA	Outros	100000	3000000
Hollywood Shuffle	1987	Robert Townsend	81	Color	Comedy	USA	A partir de 18 anos	100000	5228000
The Lost Skeleton of Cadavra	2001	Larry Blamire	90	Black and White	Comedy Horror Sci-Fi	USA	Livre	40000	1105000

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
The Last House on the Left	2009	Dennis Iliadis	114	Color	Crime Horror Thriller	USA	A partir de 18 anos	15000000	3272
Pi	1998	Darren Aronofsky	84	Black and White	Drama Mystery Thriller	USA	A partir de 18 anos	60000	3216
20 Dates	1998	Myles Berkowitz	87	Color	Biography Comedy Romance	USA	A partir de 18 anos	60000	5367
Super Size Me	2004	Morgan Spurlock	100	Color	Comedy Documentary Drama	USA	Livre	65000	1152
The Brothers McMullen	1995	Edward Burns	98	Color	Comedy Drama Romance	USA	A partir de 18 anos	25000	1024
Tiny Furniture	2010	Lena Dunham	98	Color	Comedy Drama Romance	USA	Outros	65000	3898
George Washington	2000	David Gordon Green	90	Color	Drama	USA	Outros	42000	2418
Smiling Fish & Goat on Fire	1999	Kevin Jordan	90	Color	Comedy Romance	USA	A partir de 18 anos	40000	277
The Legend of God's Gun	2007	Mike Bruce	78	Color	Western	USA	A partir de 18 anos	30000	2437
Clerks	1994	Kevin Smith	102	Black and White	Comedy	USA	A partir de 18 anos	230000	3151
Slacker	1991	Richard Linklater	100	Black and White	Comedy Drama	USA	A partir de 18 anos	23000	1227
The Puffy Chair	2005	Jay Duplass	85	Color	Comedy Drama Romance	USA	A partir de 18 anos	15000	1924
Breaking Upwards	2009	Daryl Wein	88	Color	Romance	USA	Outros	15000	7638
Pink Flamingos	1972	John Waters	108	Color	Comedy Crime Horror	USA	A partir de 18 anos	10000	1804
Primer	2004	Shane Carruth	77	Color	Drama Sci-Fi Thriller	USA	A partir de 13 anos	7000	4247
El Mariachi	1992	Robert Rodriguez	81	Color	Action Crime Drama Romance Thriller	USA	A partir de 18 anos	7000	2040
My Date with Drew	2004	Jon Gunn	90	Color	Documentary	USA	Livre	1100	8522

O operador `%in%` é muito útil na hora de criar filtros. O resultado das operações com `%in%` é um vetor lógico o tamanho do vetor do elemento da esquerda, identificando quais elementos da esquerda batem com algum elemento da direita.

In [86]: `letters`

```
'a' · 'b' · 'c' · 'd' · 'e' · 'f' · 'g' · 'h' · 'i' · 'j' · 'k' · 'l' · 'm' · 'n' · 'o' · 'p' ·  
'q' · 'r' · 's' · 't' · 'u' · 'v' · 'w' · 'x' · 'y' · 'z'
```

In [87]: `letters %in% c("a", "e", "z")`

```
TRUE · FALSE · FALSE · FALSE · TRUE · FALSE · FALSE · FALSE · FALSE ·  
FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE ·  
FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE · TRUE
```

Agora vamos criar um filtro para os Filmes dos diretores Steven Spielberg e Quentin Tarantino

```
In [88]: imdb %>%
  filter(diretor %in% c("Steven Spielberg", "Quentin Tarantino")) %>%
  head()
```

A tibble: 6 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita	nota_im
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<dbl>
Indiana Jones and the Kingdom of the Crystal Skull	2008	Steven Spielberg	122	Color	Action Adventure Fantasy	USA	A partir de 13 anos	185000000	317011114	6.2
War of the Worlds	2005	Steven Spielberg	116	Color	Adventure Sci-Fi Thriller	USA	A partir de 13 anos	132000000	234277056	6.5
The Adventures of Tintin	2011	Steven Spielberg	107	Color	Action Adventure Family Mystery	USA	Livre	135000000	77564037	7.4
Minority Report	2002	Steven Spielberg	145	Color	Action Mystery Sci-Fi Thriller	USA	A partir de 13 anos	102000000	132014112	7.7
Django Unchained	2012	Quentin Tarantino	165	Color	Drama Western	USA	A partir de 18 anos	100000000	162804648	8.5
A.I. Artificial Intelligence	2001	Steven Spielberg	146	Color	Adventure Drama Sci-Fi	USA	A partir de 13 anos	100000000	78616689	7.1

Também podemos usar funções que retornam valores lógicos, como a `str_detect()`, do pacote `stringr`. Esse pacote possui funções para a manipulação de strings, e será abordado com mais detalhes nesta seção.

```
In [89]: library(stringr)

imdb %>%
  filter(str_detect(generos, "Action"))
```

A tibble: 861 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237000000
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300000000
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250000000
John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263700000
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258000000
Avengers: Age of Ultron	2015	Joss Whedon	141	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000
Batman v Superman: Dawn of Justice	2016	Zack Snyder	183	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000
Superman Returns	2006	Bryan Singer	169	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	209000000
Pirates of the Caribbean: Dead Man's Chest	2006	Gore Verbinski	151	Color	Action Adventure Fantasy	USA	A partir de 13 anos	225000000
The Lone Ranger	2013	Gore Verbinski	150	Color	Action Adventure Western	USA	A partir de 13 anos	215000000
Man of Steel	2013	Zack Snyder	143	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	225000000
The Chronicles of Narnia: Prince Caspian	2008	Andrew Adamson	150	Color	Action Adventure Family Fantasy	USA	Livre	225000000

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
The Avengers	2012	Joss Whedon	173	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	220000000
Pirates of the Caribbean: On Stranger Tides	2011	Rob Marshall	136	Color	Action Adventure Fantasy	USA	A partir de 13 anos	250000000
Men in Black 3	2012	Barry Sonnenfeld	106	Color	Action Adventure Comedy Family Fantasy Sci-Fi	USA	A partir de 13 anos	225000000
The Amazing Spider-Man	2012	Marc Webb	153	Color	Action Adventure Fantasy	USA	A partir de 13 anos	230000000
Robin Hood	2010	Ridley Scott	156	Color	Action Adventure Drama History	USA	A partir de 13 anos	200000000
Captain America: Civil War	2016	Anthony Russo	147	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	250000000
Battleship	2012	Peter Berg	131	Color	Action Adventure Sci-Fi Thriller	USA	A partir de 13 anos	209000000
Jurassic World	2015	Colin Trevorrow	124	Color	Action Adventure Sci-Fi Thriller	USA	A partir de 13 anos	150000000
Spider-Man 2	2004	Sam Raimi	135	Color	Action Adventure Fantasy Romance	USA	A partir de 13 anos	200000000
Iron Man 3	2013	Shane Black	195	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000
Transformers: Revenge of the Fallen	2009	Michael Bay	150	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000
Transformers: Age of Extinction	2014	Michael Bay	165	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	210000000
The Amazing Spider-Man 2	2014	Marc Webb	142	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	200000000
TRON: Legacy	2010	Joseph Kosinski	125	Color	Action Adventure Sci-Fi	USA	Livre	170000000
Green Lantern	2011	Martin Campbell	123	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000
Terminator Salvation	2009	McG	118	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	200000000
Furious 7	2015	James Wan	140	Color	Action Crime Thriller	USA	A partir de 13 anos	190000000

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
World War Z	2013	Marc Forster	123	Color	Action Adventure Horror Sci-Fi Thriller	USA	A partir de 13 anos	190000000
...
Sands of Iwo Jima	1949	Allan Dwan	109	Black and White	Action Drama Romance War	USA	Outros	1000000
Hardflip	2012	Johnny Remo	112	Color	Action Drama	USA	A partir de 13 anos	1000000
The Specials	2000	Craig Mazin	82	Color	Action Comedy Fantasy Sci-Fi	USA	A partir de 18 anos	1000000
The Helix... Loaded	2005	A. Raven Cruz	97	Color	Action Adventure Comedy Fantasy Sci-Fi	USA	A partir de 18 anos	1000000
Zombie Hunter	2013	K. King	93	Color	Action Comedy Horror	USA	Outros	1000000
Snitch	2013	Ric Roman Waugh	112	Color	Action Drama Thriller	USA	A partir de 13 anos	15000000
American Hero	2015	Nick Love	86	Color	Action Comedy Drama Sci-Fi	USA	Outros	990000
Saints and Soldiers	2003	Ryan Little	90	Color	Action Adventure Drama War	USA	A partir de 13 anos	780000
Aqua Teen Hunger Force Colon Movie Film for Theaters	2007	Matt Maiellaro	86	Color	Action Adventure Animation Comedy Fantasy Sci-Fi	USA	A partir de 18 anos	750000
Destiny	2014	Joseph Kosinski	NA	Color	Action Adventure Fantasy Sci-Fi	USA	Outros	NA
Mercy Streets	2000	Jon Gunn	106	Color	Action Crime Drama	USA	A partir de 13 anos	600000
Light from the Darkroom	2014	Lance McDaniel	90	Color	Action Drama Thriller	USA	A partir de 13 anos	600000
Subconscious	2015	Georgia Hilton	122	Color	Action Mystery Thriller	USA	A partir de 18 anos	NA
Fight to the Finish	2016	Warren Sheppard	94	Color	Action Romance Sport	USA	A partir de 13 anos	150000

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Born to Fly: Elizabeth Streb vs. Gravity	2014	Catherine Gund	82	Color	Action Biography Documentary Sport	USA	Outros	500000
The Toxic Avenger	1984	Michael Herz	91	Color	Action Comedy Horror Sci-Fi	USA	Outros	475000
Diamond Ruff	2015	Alec Asten	82	Color	Action	USA	A partir de 18 anos	500000
American Ninja 2: The Confrontation	1987	Sam Firstenberg	90	Color	Action Drama	USA	A partir de 18 anos	350000
Revolution	NA	NA	43	Color	Action Adventure Drama Sci-Fi	USA	Outros	NA
The Prophecy	1995	Gregory Widen	98	Color	Action Fantasy Horror Mystery Thriller	USA	A partir de 18 anos	8000000
Royal Kill	2009	Babar Ahmed	90	Color	Action Thriller	USA	A partir de 13 anos	350000
Death Race 2000	1975	Paul Bartel	80	Color	Action Comedy Sci-Fi Sport	USA	A partir de 18 anos	300000
Death Calls	2010	Ken Del Conte	90	Color	Action Adventure Mystery Romance Thriller	USA	A partir de 18 anos	290000
Amidst the Devil's Wings	2014	Daniel Columbie	90	Color	Action Crime Drama	USA	Outros	300000
Ink	2009	Jamin Winans	107	Color	Action Drama Fantasy Sci-Fi	USA	Outros	NA
Heroes of Dirt	2015	Eric Bugbee	98	Color	Action Drama Sport	USA	A partir de 13 anos	NA
Echo Dr.	2013	Patrick Ryan Sims	81	Color	Action Drama Sci-Fi Thriller	USA	Outros	100000
Queen Crab	2015	Brett Piper	80	Color	Action Sci-Fi Thriller	USA	Outros	75000
Sabotage	2014	David Ayer	109	Color	Action Crime Drama Thriller	USA	A partir de 18 anos	35000000
El Mariachi	1992	Robert Rodriguez	81	Color	Action Crime Drama Romance Thriller	USA	A partir de 18 anos	7000

SELECIONANDO COLUNAS

A função `select()` seleciona colunas (variáveis). É possível utilizar nomes, índices, intervalos de variáveis ou utilizar as funções `starts_with(x)`, `contains(x)`, `matches(x)`, `one_of(x)` para selecionar as variáveis.

```
In [90]: names(imdb)
```

```
'titulo' · 'ano' · 'diretor' · 'duracao' · 'cor' · 'generos' · 'pais' · 'classificacao' ·  
'orcamento' · 'receita' · 'nota_imdb' · 'likes_facebook' · 'ator_1' · 'ator_2' ·  
'ator_3'
```

```
In [91]: imdb %>%  
  select(titulo, ano, orcamento) %>%  
  head()
```

A tibble: 6 × 3

titulo	ano	orcamento
<chr>	<int>	<int>
Avatar	2009	237000000
Pirates of the Caribbean: At World's End	2007	300000000
The Dark Knight Rises	2012	250000000
John Carter	2012	263700000
Spider-Man 3	2007	258000000
Tangled	2010	260000000


```
In [92]: imdb %>%  
  select(starts_with("ator")) %>%  
  head()
```

A tibble: 6 × 3

ator_1	ator_2	ator_3
<chr>	<chr>	<chr>
CCH Pounder	Joel David Moore	Wes Studi
Johnny Depp	Orlando Bloom	Jack Davenport
Tom Hardy	Christian Bale	Joseph Gordon-Levitt
Daryl Sabara	Samantha Morton	Polly Walker
J.K. Simmons	James Franco	Kirsten Dunst
Brad Garrett	Donna Murphy	M.C. Gainey

O operador : pode ser usado para selecionar intervalos de colunas.

```
In [93]: imdb %>%  
  select(titulo, ator_1:ator_3) %>%  
  head()
```

A tibble: 6 × 4

titulo	ator_1	ator_2	ator_3
<chr>	<chr>	<chr>	<chr>
Avatar	CCH Pounder	Joel David Moore	Wes Studi
Pirates of the Caribbean: At World's End	Johnny Depp	Orlando Bloom	Jack Davenport
The Dark Knight Rises	Tom Hardy	Christian Bale	Joseph Gordon-Levitt
John Carter	Daryl Sabara	Samantha Morton	Polly Walker
Spider-Man 3	J.K. Simmons	James Franco	Kirsten Dunst
Tangled	Brad Garrett	Donna Murphy	M.C. Gainey

Para retirar colunas da base, base acrescentar um - antes da seleção.

```
In [94]: imdb %>%
  select(-ano, - diretor) %>%
  head()
```

A tibble: 6 × 13

titulo	duracao	cor	generos	pais	classificacao	orcamento	receita
<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
Avatar	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237000000	760505847
Pirates of the Caribbean: At World's End	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300000000	309404152
The Dark Knight Rises	164	Color	Action Thriller	USA	A partir de 13 anos	250000000	448130642
John Carter	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263700000	73058679
Spider-Man 3	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258000000	336530303
Tangled	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260000000	200807262

```
In [95]: imdb %>%
  select(-starts_with("ator")) %>%
  head()
```

A tibble: 6 × 12

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250
John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258
Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260

ORDENANDO A BASE

A função `arrange()` ordena a base segundo uma ou mais colunas. O argumento `desc=` pode ser utilizado para gerar uma ordem decrescente.

```
In [96]: imdb %>%
  arrange(orcamento) %>%
  select(orcamento, everything()) %>%
  head()
```

A tibble: 6 × 15

orcamento	titulo	ano	diretor	duracao	cor	generos	pais	classificacao	receita	nota_im
<int>	<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<dbl>
218	Tarnation	2003	Jonathan Caouette	88	Color	Biography Documentary	USA	Outros	592014	7.2
1100	My Date with Drew	2004	Jon Gunn	90	Color	Documentary	USA	Livre	85222	6.6
1400	A Plague So Pleasant	2013	Benjamin Roberds	76	Color	Drama Horror Thriller	USA	Outros	NA	6.3
3250	The Mongol King	2005	Anthony Vallone	84	Color	Crime Drama	USA	A partir de 13 anos	NA	7.8
7000	Primer	2004	Shane Carruth	77	Color	Drama Sci-Fi Thriller	USA	A partir de 13 anos	424760	7.0
7000	El Mariachi	1992	Robert Rodriguez	81	Color	Action Crime Drama Romance Thriller	USA	A partir de 18 anos	2040920	6.9


```
In [97]: imdb %>%
  arrange(desc(orçamento)) %>%
  select(orçamento, everything()) %>%
  head()
```

A tibble: 6 × 15

orçamento	titulo	ano	diretor	duracao	cor	generos	pais	class
<int>	<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>
300000000	Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A par 13 ar
263700000	John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A par 13 ar
260000000	Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre
258000000	Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A par 13 ar
258000000	Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A par 13 ar
250000000	The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A par 13 ar

```
In [98]: imdb %>%
  arrange(desc(ano), titulo) %>%
  select(titulo, ano, everything()) %>%
  head()
```

A tibble: 6 × 15

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento	receita	nota_imd
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<dbl>
10 Cloverfield Lane	2016	Dan Trachtenberg	104	Color	Drama Horror Mystery Sci- Fi Thriller	USA	A partir de 13 anos	15000000	71897215	7.3
13 Hours	2016	Michael Bay	144	Color	Action Drama Thriller War	USA	A partir de 18 anos	50000000	52822418	7.4
A Beginner's Guide to Snuff	2016	Mitchell Altieri	87	Color	Comedy Horror Thriller	USA	Outros	NA	NA	8.7
Alice Through the Looking Glass	2016	James Bobin	113	Color	Adventure Family Fantasy	USA	Livre	170000000	76846624	6.4
Allegiant	2016	Robert Schwentke	120	Color	Action Adventure Mystery Sci- Fi Thriller	USA	A partir de 13 anos	110000000	66002193	5.8
Alleluia! The Devil's Carnival	2016	Darren Lynn Bousman	97	Color	Horror Musical	USA	Outros	500000	NA	7.4

MUTATE()

A função `mutate()` cria ou modifica colunas. Ela é equivalente à função `transform()`, mas aceita várias novas colunas iterativamente. Novas variáveis devem ter o mesmo número de linhas da base original (ou comprimento 1).

A coluna "duracao" é sobrescrita

```
In [99]: imdb %>%  
  mutate(duracao = duracao/60) %>%  
  select(duracao) %>%  
  head()
```

A tibble: 6 ×

1

<u>duracao</u>
<u><dbl></u>
<u>2.966667</u>
<u>2.816667</u>
<u>2.733333</u>
<u>2.200000</u>
<u>2.600000</u>
<u>1.666667</u>

Criamos uma nova coluna na base

```
In [100]: imdb %>%  
  mutate(duracao_horas = duracao/60) %>%  
  select(duracao, duracao_horas) %>%  
  head()
```

A tibble: 6 × 2

duracao	duracao_horas
<int>	<dbl>
178	2.966667
169	2.816667
164	2.733333
132	2.200000
156	2.600000
100	1.666667

Podemos fazer diversas operações em um mesmo mutate.


```
In [101]: imdb %>%  
  mutate(  
    lucro = receita - orcamento,  
    resultado = ifelse(lucro < 0, "prejuizo", "lucro")  
  ) %>%  
  select(lucro, resultado) %>%  
  head()
```

A tibble: 6 × 2

lucro	resultado
<int>	<chr>
523505847	lucro
9404152	lucro
198130642	lucro
-190641321	prejuizo
78530303	lucro
-59192738	prejuizo

AGRUPANDO E SUMARIZANDO A BASE

A função `summarise()` sumariza a base. Ela aplica uma função às variáveis, retornando um vetor de tamanho 1. Ela é utilizada em conjunto da função `group_by()`. A função `n()` costuma ser bastante utilizada com a função `summarise()`.

```
In [102]: imdb %>%  
  summarise(media_orcamento = mean(orçamento, na.rm = TRUE))
```

A tibble: 1 × 1

<u>media_orcamento</u>
<u><dbl></u>
35755986

```
In [103]: imdb %>%  
  summarise(  
    media_orcamento = mean(orçamento, na.rm = TRUE),  
    mediana_orcamento = median(orçamento, na.rm = TRUE),  
    qtd = n(),  
    qtd_diretores = n_distinct(diretor)  
  )
```

A tibble: 1 × 4

media_orcamento	mediana_orcamento	qtd	qtd_diretores
<dbl>	<int>	<int>	<int>
35755986	20000000	3807	1813

```
In [104]: imdb %>%  
  group_by(ano) %>%  
  summarise(qtd_filmes = n()) %>%  
  head()
```

`summarise()` ungrouping output (override with `.groups` argument)

A tibble: 6 × 2

ano	qtd_filmes
<int>	<int>
1916	1
1920	1
1925	1
1929	1
1930	1
1932	1

O PACOTE TIDYR

O pacote tidyrr dispõe de funções úteis para deixar os seus dados no formato que você precisa para a análise. Na maioria das vezes, utilizamos para deixá-los tidy.

Outras, precisamos “bagunçá-los” um pouco para poder aplicar alguma função específica.

As principais funções deste pacote são a `gather()` e a `spread()`

GATHER()

A função `gather()` “empilha” o banco de dados. Ela é utilizada principalmente quando as colunas da base não representam nomes de variáveis, mas sim seus valores.

```
In [105]: imdb_gather <- imdb %>%  
  mutate(id = 1:n()) %>%  
  gather(  
    key = "importancia_ator",  
    value = "nome_ator",  
    ator_1, ator_2, ator_3  
  ) %>%  
  select(nome_ator, importancia_ator, everything())
```


In [106]: `head(imdb_gather)`

A tibble: 6 × 15

nome_ator	importancia_ator	titulo	ano	diretor	duracao	cor	generos
<chr>	<chr>	<chr>	<int>	<chr>	<int>	<chr>	<chr>
CCH Pounder	ator_1	Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi
Johnny Depp	ator_1	Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy
Tom Hardy	ator_1	The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller
Daryl Sabara	ator_1	John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi
J.K. Simmons	ator_1	Spider- Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance
Brad Garrett	ator_1	Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Roma

SPREAD()

A função `spread()` é essencialmente o inverso da `gather()`. Ela espalha uma variável nas colunas.

```
In [107]: imdb_spread <- imdb_gather %>%  
  spread(  
    key = importancia_ator,  
    value = nome_ator  
  )
```

In [108]: `head(imdb_spread)`

A tibble: 6 × 16

titulo	ano	diretor	duracao	cor	generos	pais	classificacao	orcamento
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<int>
Avatar	2009	James Cameron	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action Thriller	USA	A partir de 13 anos	250
John Carter	2012	Andrew Stanton	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263
Spider-Man 3	2007	Sam Raimi	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258
Tangled	2010	Nathan Greno	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260

OUTRAS FUNÇÕES DO TIDYR

A função `unite()` junta duas ou mais colunas usando algum separador (__, por exemplo).

A função `separate()` faz o inverso de `unite()`: transforma uma coluna em várias usando um separador.

In [109]:

```
imdb %>%  
  unite(  
    col = "titulo_diretor",  
    titulo, diretor,  
    sep = " - "  
  ) %>%  
  head()
```

A tibble: 6 × 14

titulo_diretor	ano	duracao	cor	generos	pais	classificacao	orcamento	re
<chr>	<int>	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<
Avatar - James Cameron	2009	178	Color	Action Adventure Fantasy Sci-Fi	USA	A partir de 13 anos	237000000	7
Pirates of the Caribbean: At World's End - Gore Verbinski	2007	169	Color	Action Adventure Fantasy	USA	A partir de 13 anos	300000000	3
The Dark Knight Rises - Christopher Nolan	2012	164	Color	Action Thriller	USA	A partir de 13 anos	250000000	4
John Carter - Andrew Stanton	2012	132	Color	Action Adventure Sci-Fi	USA	A partir de 13 anos	263700000	7
Spider-Man 3 - Sam Raimi	2007	156	Color	Action Adventure Romance	USA	A partir de 13 anos	258000000	3
Tangled - Nathan Greno	2010	100	Color	Adventure Animation Comedy Family Fantasy Musical Romance	USA	Livre	260000000	2

```
In [110]: imdb %>%
  separate(
    col = generos,
    into = c("genero_1", "genero_2", "genero_3"),
    sep = "\\|",
    extra = "drop"
  ) %>%
  head()
```

Warning message:

"Expected 3 pieces. Missing pieces filled with `NA` in 1537 rows [3, 19, 21, 42, 84, 88, 92, 102, 106, 111, 113, 129, 138, 147, 178, 183, 186, 219, 222, 233, ...]."

A tibble: 6 × 17

titulo	ano	diretor	duracao	cor	genero_1	genero_2	genero_3	pais	classificacao	orcamento	receita	nota
<chr>	<int>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<dbl>
Avatar	2009	James Cameron	178	Color	Action	Adventure	Fantasy	USA	A partir de 13 anos	237000000	760505847	7.9
Pirates of the Caribbean: At World's End	2007	Gore Verbinski	169	Color	Action	Adventure	Fantasy	USA	A partir de 13 anos	300000000	309404152	7.1
The Dark Knight Rises	2012	Christopher Nolan	164	Color	Action	Thriller	NA	USA	A partir de 13 anos	250000000	448130642	8.5
John Carter	2012	Andrew Stanton	132	Color	Action	Adventure	Sci-Fi	USA	A partir de 13 anos	263700000	73058679	6.6
Spider-Man 3	2007	Sam Raimi	156	Color	Action	Adventure	Romance	USA	A partir de 13 anos	258000000	336530303	6.2
Tangled	2010	Nathan Greno	100	Color	Adventure	Animation	Comedy	USA	Livre	260000000	200807262	7.8

Faça a atividade desta aula neste [link \(https://forms.gle/6tvkDosSgoPayCjL6\)](https://forms.gle/6tvkDosSgoPayCjL6).

<https://forms.gle/6tvkDosSgoPayCjL6> (<https://forms.gle/6tvkDosSgoPayCjL6>).