

Oficina CIF



Visualização de dados em Farmacologia: tipos de gráficos, como plotá-los e boas práticas para publicações

Tamires Martins

Aluna(o)

Índice

1	Introdução	1
2	Passos básicos com R	1
2.1	Como se ajudar	1
3	Como rodar o código?	1
3.1	A ordem importa (rode o código na sequência):	2
4	Os dados	2
4.1	Carregando pacotes	2
4.2	Importando os dados	3
4.3	Visualizando os dados	4
5	Os gráficos	8
5.1	Gráfico box plot simples usando ggplot2	8
5.2	Gráfico box plot simples usando tidypplots	9
6	Agora é a sua vez:	10

1 Introdução

Nesta prática, vamos explorar como visualizar dados da Farmacologia em R. Para isso, também vamos aprender passos básicos com a linguagem.

2 Passos básicos com R

2.1 Como se ajudar

Abrir a documentação de uma função é essencial para entender como ela funciona e quais parâmetros você pode usar. Existem várias maneiras de acessar a documentação no R:

```
1 ?mean
2 # ou
3 help(mean)
4 # ou clicar no nome da função apertar F1
```

3 Como rodar o código?

Coloque o cursor na linha que deseja rodar e aperte “Run” ou “Ctrl + Enter”. O resultado irá aparecer no console, faça o teste:

```
1 print("Olá, mundo!") # Imprimindo uma mensagem no console
```

```
[1] "Olá, mundo!"
```

```
1 1 + 1
```

```
[1] 2
```

3.1 A ordem importa (rode o código na sequência):

```
1 # tire a anotação da frente e tente rodar:
2 # numero
3
4 # Atribuindo um valor a um objeto
5 numero <- 10
6
7 numero
```

```
[1] 10
```

Para acessar a classe de um objeto:

```
1 class(numero)
```

```
[1] "numeric"
```

Os pacotes são conjuntos de funções. Para instalar um pacote, use `install.packages("nome_do_pacote")`. Para carregar um pacote já instalado, use `library(nome_do_pacote)`. Algumas funções são da base:

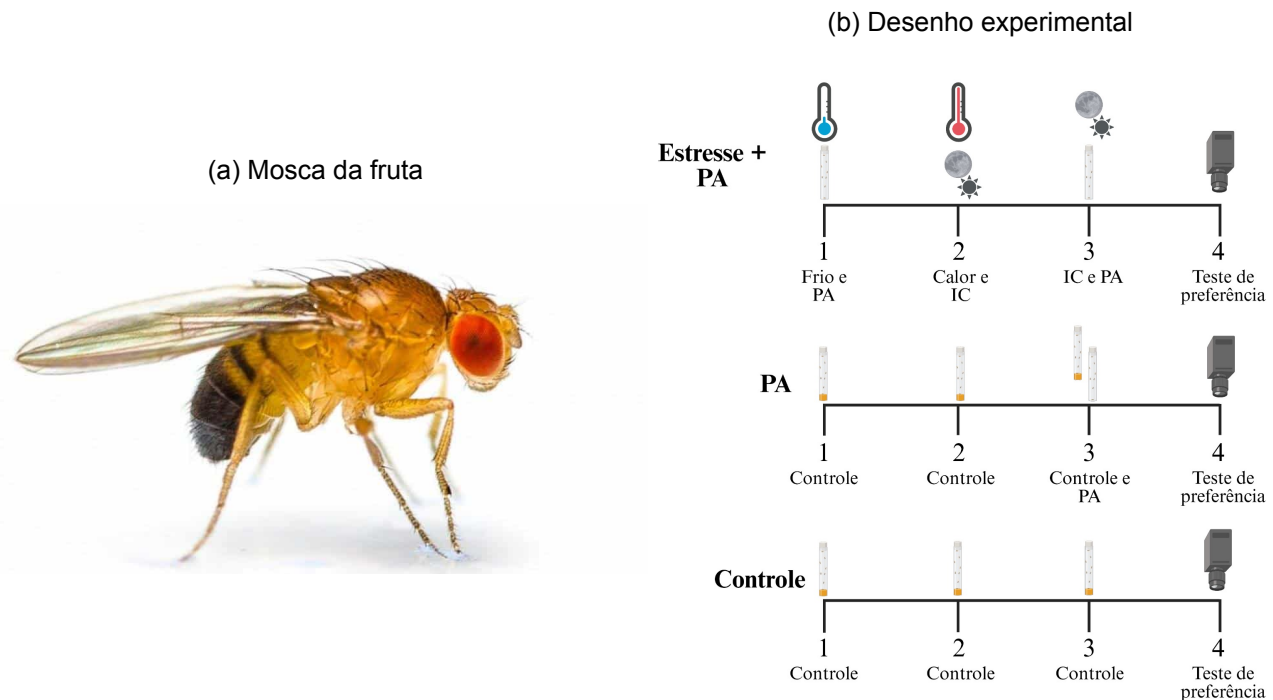
```
1 # mean() calcula a média
2
3 # A média de um vetor
4 meu_vetor <- c(1, 2, 3, 4, 5)
5 mean(meu_vetor)
```

```
[1] 3
```

4 Os dados

Os dados que iremos trabalhar são simulações baseadas em dados reais (Eckert et al. (não publicado)). O experimento consiste em avaliar comportamentos das moscas-das-frutas (*Drosophila melanogaster*) machos e fêmeas, em diferentes condições de tratamento.

Figura 1: Experimento com *Drosophila melanogaster*



4.1 Carregando pacotes

```
1 # seguido de texto são anotações, não código.
2
3 # Instale se ainda não tiver
4 # install.packages("ggplot2")
```

```

5 # install.packages("tidyplots")
6 # install.packages("patchwork")
7
8 library(ggplot2)
9 library(tidyplots)
10 library(patchwork)

```

4.2 Importando os dados

Os dados estão disponíveis no arquivo `dados_moscas.xlsx`, que deve estar na mesma pasta do arquivo `.qmd`.

```

1 # Importar os dados do arquivo Excel
2 dados <- readxl::read_excel("dados_moscas.xlsx")
3
4 dados

```

A tibble: 259 x 31

	video	animal_no_video	sexo	tratamento	animal_number	tempo	test_duration_s
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>
1	G1V3	2	M	controle	1	Total	2400
2	G1V2	8	M	controle	2	Total	2400
3	G1V1	1	M	controle	3	Total	2400
4	G1V1	3	M	controle	4	Total	2400
5	G1V3	12	M	controle	5	Total	2400
6	G2V1	2	M	controle	6	Total	2400
7	G2V3	13	M	controle	7	Total	2400
8	G2V4	1	M	controle	8	Total	2400
9	G2V4	3	M	controle	9	Total	2400
10	G3V1	7	M	controle	10	Total	2400

i 249 more rows

i 24 more variables: total_distance_travelled_m <dbl>,
total_time_mobile_s <dbl>, total_time_immobile_s <dbl>,
total_mobile_episodes <dbl>, total_immobile_episodes <dbl>,
number_of_entries_to_the_agar_zone <dbl>, time_in_the_agar_zone_s <dbl>,
distance_travelled_in_the_agar_zone_m <dbl>,
time_mobile_in_the_agar_zone_s <dbl>, ...

4.3 Visualizando os dados

```

1 dplyr::glimpse(dados)

```

Rows: 259

Columns: 31

\$ video <chr> "G1V3", "G1V2", "G1V1", "G1V~

\$ animal_no_video	<chr> "2", "8", "1", "3", "12", "2~
\$ sexo	<chr> "M", "M", "M", "M", "M", "M"~
\$ tratamento	<chr> "controle", "controle", "con~
\$ animal_number	<chr> "1", "2", "3", "4", "5", "6"~
\$ tempo	<chr> "Total", "Total", "Total", "~
\$ test_duration_s	<dbl> 2400, 2400, 2400, 2400, 2400~
\$ total_distance_travelled_m	<dbl> 20.703724, 14.925057, 7.7662~
\$ total_time_mobile_s	<dbl> 2238.402, 1796.914, 2277.511~
\$ total_time_immobile_s	<dbl> 161.5979, 603.0864, 122.4891~
\$ total_mobile_episodes	<dbl> 92, 189, 158, 61, 61, 144, 1~
\$ total_immobile_episodes	<dbl> 105, 178, 59, 172, 214, 130,~
\$ number_of_entries_to_the_agar_zone	<dbl> 63, 21, 81, 80, 21, 1, 60, 3~
\$ time_in_the_agar_zone_s	<dbl> 193.33418, 302.93306, 110.19~
\$ distance_travelled_in_the_agar_zone_m	<dbl> 0.92832895, 0.46652902, 0.89~
\$ time_mobile_in_the_agar_zone_s	<dbl> 140.29460, 297.49660, 77.520~
\$ time_immobile_in_the_agar_zone_s	<dbl> 53.039574, 5.436460, 32.6709~
\$ immobile_episodes_in_the_agar_zone	<dbl> 1, 7, 20, 9, 11, 1, 3, 5, 6,~
\$ number_of_entries_to_the_raia_zone	<dbl> 204, 174, 217, 247, 153, 300~
\$ time_in_the_raia_zone_s	<dbl> 1926.100, 1862.902, 2033.823~
\$ distance_travelled_in_the_raia_zone_m	<dbl> 19.313651, 13.688998, 6.3153~
\$ time_mobile_in_the_raia_zone_s	<dbl> 1887.455, 1296.854, 1950.409~
\$ time_immobile_in_the_raia_zone_s	<dbl> 38.64488, 566.04737, 83.4141~
\$ immobile_episodes_in_the_raia_zone	<dbl> 94, 163, 34, 159, 193, 120, ~
\$ number_of_entries_to_the_sacarose_zone	<dbl> 91, 92, 142, 62, 89, 11, 184~
\$ time_in_the_sacarose_zone_s	<dbl> 280.56567, 234.16512, 255.98~
\$ distance_travelled_in_the_sacarose_zone_m	<dbl> 0.46174388, 0.76953020, 0.55~
\$ time_mobile_in_the_sacarose_zone_s	<dbl> 210.65218, 202.56256, 249.58~
\$ time_immobile_in_the_sacarose_zone_s	<dbl> 69.913487, 31.602566, 6.4040~
\$ immobile_episodes_in_the_sacarose_zone	<dbl> 10, 8, 5, 4, 10, 9, 4, 6, 3,~
\$ pindex	<dbl> 1.4511954, 0.7729930, 2.3230~

```
1 skimr::skim(dados)
```

Tabela 1: Data summary

Name	dados
Number of rows	259
Number of columns	31
Column type frequency:	
character	6
numeric	25
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
video	0	1	4	4	0	22	0
animal_no_video	0	1	1	2	0	14	0
sexo	0	1	1	1	0	2	0
tratamento	0	1	4	8	0	5	0
animal_number	0	1	1	3	0	259	0
tempo	0	1	5	5	0	1	0

Variable type: numeric

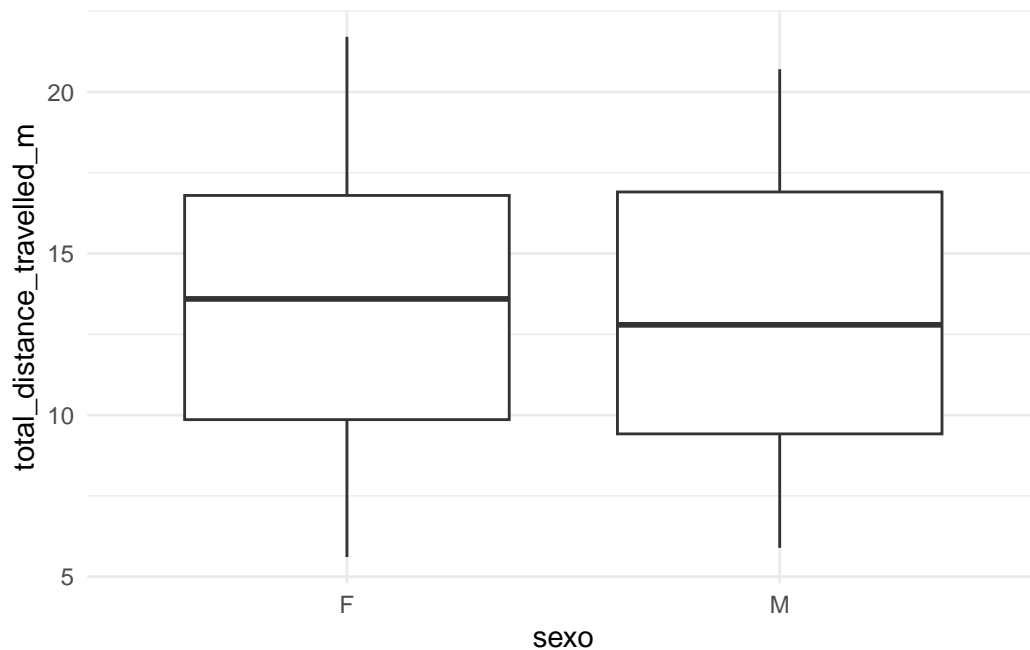
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
test_duration_s	0	1	2400.00	0.00	2400.00	2400.00	2400.00	2400.00	2400.00	□□□□□
total_distance_travelled_m	0	1	13.29	4.25	5.61	9.63	13.18	16.84	21.71	□□□□□
total_time_mobile_s	0	1	2001.32	179.09	1646.57	1841.93	2006.30	2154.65	2373.74	□□□□□
total_time_immobile_s	0	1	398.68	179.09	26.26	245.35	393.70	558.07	753.43	□□□□□
total_mobile_episodes	0	1	125.07	44.12	51.00	86.00	122.00	162.00	200.00	□□□□□
total_immobile_episodes	0	1	114.33	57.80	11.00	62.00	115.00	165.00	217.00	□□□□□

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
number_of_entries_to_the_agar_zone	0	1	46.41	29.16	1.00	20.50	45.00	72.00	99.00	□□□□□
time_in_the_agar_zone_s	0	1	242.72	115.71	14.44	146.97	239.52	330.19	482.63	□□□□□
distance_travelled_in_the_agar_zone_m	0	1	0.50	0.28	0.02	0.27	0.49	0.75	1.00	□□□□□
time_mobile_in_the_agar_zone_s	0	1	190.64	109.52	1.59	96.00	187.74	279.97	396.42	□□□□□
time_immobile_in_the_agar_zone_s	0	1	52.08	29.09	0.35	27.00	53.04	77.70	99.49	□□□□□
immobile_episodes_in_the_agar_zone	0	1	10.49	6.06	1.00	5.00	11.00	16.00	20.00	□□□□□
number_of_entries_to_the_raia_zone	0	1	193.84	58.55	100.00	147.00	183.00	250.00	300.00	□□□□□
time_in_the_raia_zone_s	0	1	1906.39	150.57	1510.82	1799.10	1902.12	2015.79	2280.83	□□□□□
distance_travelled_in_the_raia_zone_m	0	1	12.31	4.24	5.01	8.57	12.11	15.82	19.84	□□□□□
time_mobile_in_the_raia_zone_s	0	1	1609.04	232.60	951.70	1431.21	1621.05	1786.10	2142.30	□□□□□
time_immobile_in_the_raia_zone_s	0	1	297.35	178.04	2.75	137.88	295.15	451.37	598.57	□□□□□
immobile_episodes_in_the_raia_zone	0	1	98.52	56.69	1.00	46.50	99.00	145.50	199.00	□□□□□
number_of_entries_to_the_sacarose_zone	0	1	104.86	55.90	2.00	61.50	109.00	153.00	200.00	□□□□□
time_in_the_sacarose_zone_s	0	1	250.89	117.64	10.83	151.86	257.50	346.42	483.69	□□□□□
distance_travelled_in_the_sacarose_zone_m	0	1	0.48	0.28	0.00	0.23	0.46	0.73	1.00	□□□□□
time_mobile_in_the_sacarose_zone_s	0	1	201.64	115.06	1.22	92.13	218.22	292.37	399.75	□□□□□
time_immobile_in_the_sacarose_zone_s	0	1	49.25	28.08	0.25	23.78	50.44	71.73	99.74	□□□□□
immobile_episodes_in_the_sacarose_zone	0	1	5.32	3.07	0.00	2.50	5.00	8.00	10.00	□□□□□
pindex	0	1	1.62	2.07	0.02	0.58	1.06	1.87	19.62	□□□□□

5 Os gráficos

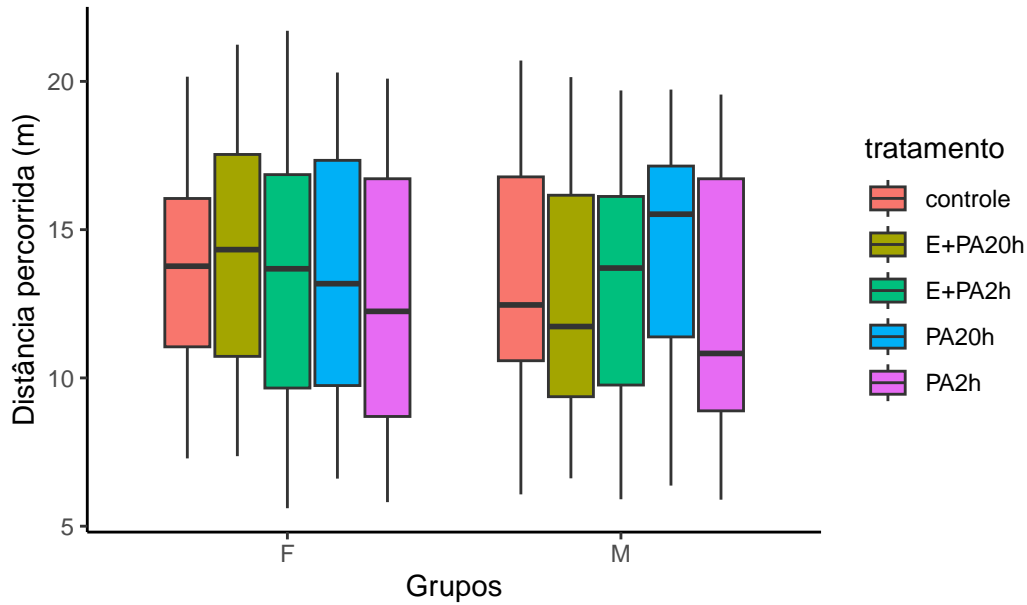
5.1 Gráfico box plot simples usando ggplot2

```
1 ggplot(dados, aes(x = sexo, y = total_distance_travelled_m)) +  
2   geom_boxplot() +  
3   theme_minimal()
```



```
1 # dando nome ao objeto, adicionando mais uma variável categórica, e mudando o tema:  
2 grafico <- ggplot(  
3   dados,  
4   aes(x = sexo, y = total_distance_travelled_m, fill = tratamento)) +  
5   geom_boxplot() +  
6   labs(  
7     title= "Quem caminha mais? machos ou fêmeas? com ggplot2",  
8     x = "Grupos",  
9     y = "Distância percorrida (m)") +  
10  theme_classic()  
11  
12 grafico
```


Quem caminha mais? machos ou fêmeas? com ggplot2

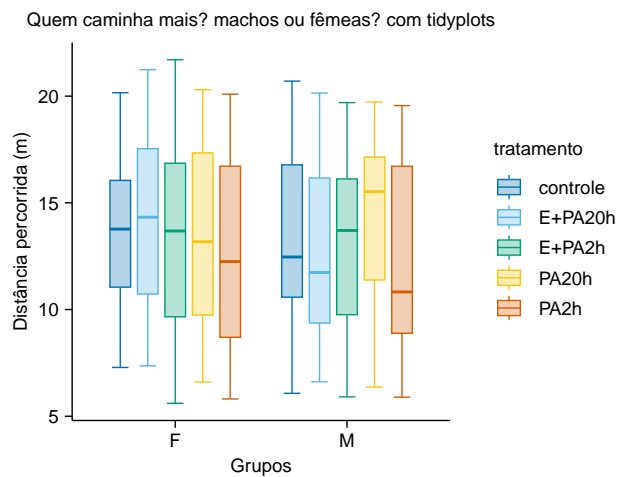


Depois de satisfeitos, podemos salvar:

```
1 ggsave(  
2   plot = grafico, # o default é o ultimo gráfico gerado  
3   filename = "ggplot2.png",  
4   path = "figuras",  
5   dpi = 600 # o default é 300  
6 )
```

5.2 Gráfico box plot simples usando tidypLOTS

```
1 grafico <- tidypLOT(  
2   dados,  
3   x = sexo,  
4   y = total_distance_travelled_m,  
5   fill = tratamento  
6 ) |>  
7 add_boxplot() |>  
8 adjust_title("Quem caminha mais? machos ou fêmeas? com tidypLOTS") |>  
9 adjust_x_axis_title("Grupos") |>  
10 adjust_y_axis_title("Distância percorrida (m)") |>  
11 theme_tidypLOT() |> # tema padrão do tidypLOTS  
12 save_plot("figuras/tidypLOT.png")
```



```

1 # ou salvando com ggplot2
2 ggsave(
3   plot = grafico,
4   filename = "tidypLOTS.png",
5   path = "figuras",
6   dpi = 600
7 )

```

6 Agora é a sua vez:

1. Faça uma pergunta sobre os dados.
2. Crie um gráfico que responda a essa pergunta.
3. Personalize o gráfico com títulos, rótulos e cores.
4. Salve o gráfico.

```

1 # [ADICIONE O CÓDIGO AQUI]

```