



Aula 3

Semabio 2024

Carolina Musso

IB/UnB



- Estrutura:
 - Aula 1: Instalação e Intro Rmarkdown
 - Aula 2 (hoje): Rmarkdown, pacote dplyr
 - Aula 3: Gráficos ggplot, Tabelas resumo e testes estatísticos

Quiz

Revisando ...

O Rmarkdown tem 3 partes

- Cabeçalho: YAML
- Texto: Markdown
- Código: chunk e inline

```
1 |---
2 |title: "Meu primeiro arquivo automático"
3 |author: "Meu nome"
4 |date: "`r format(Sys.Date(), '%d/%m/%Y')`"
5 |output:
6 |  word_document: default
7 |  html_document: default
8 |---
9 |
10 | ```${r sets, echo=F, message=F, warning=F }
11 | # COMENTÁRIOS INICIAIS ----
12 |
13 | # tá vendo que eu estou escrevendo em uma parte que tem um fundo cinza? Isso é chamado de "chunk" que
14 | # significa pedaço em inglês. Um "chunk" é onde a gente vai escrever os códigos de R mesmo. Depois vamos
15 | # voltar a esses assuntos com mais detalhes!
16 |
17 | # PRE-SETS ----
18 | rm(list=ls()) # limpa o ambiente
19 |
20 | pasta=paste0("Saidas-Exercicios-", Sys.Date()) #cria uma pasta para as saídas do dia
21 | if(!dir.exists(pasta)) dir.create(pasta) #cria uma pasta para as saídas do dia
22 |
23 |
24 |
25 |
26 |
27 |
28 |
29 |
30 |
31 | ```${r pacotes, echo=F, message=F, warning=F }
32 |
33 |
34 |
35 |
36 |
37 | ## Aqui é o subtítulo que você quiser
38 |
39 | Daí, aqui você escreve o que você achar interessante sobre os dados. Pode interpretar, enfim...
40 |
```


Termos

- Funções
- Pacotes
- Objetos
- Variáveis

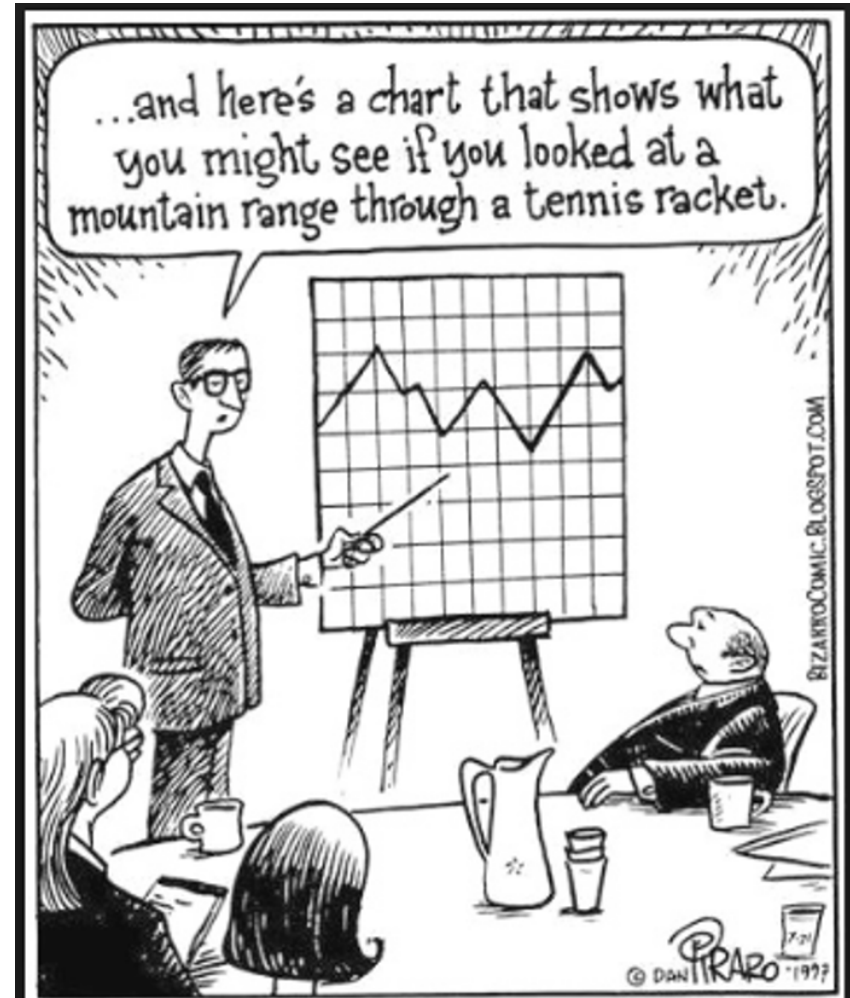
Gráficos

Leitura para aprofundamento

- EpiHandbook, Cap.30
- Curso-R, Livro, cap. 13.8; Curso Visualização
- DataCamp, Introdução a visualização com ggplot
- R4DS *O livro começa com gráficos!*
- GGPLOT

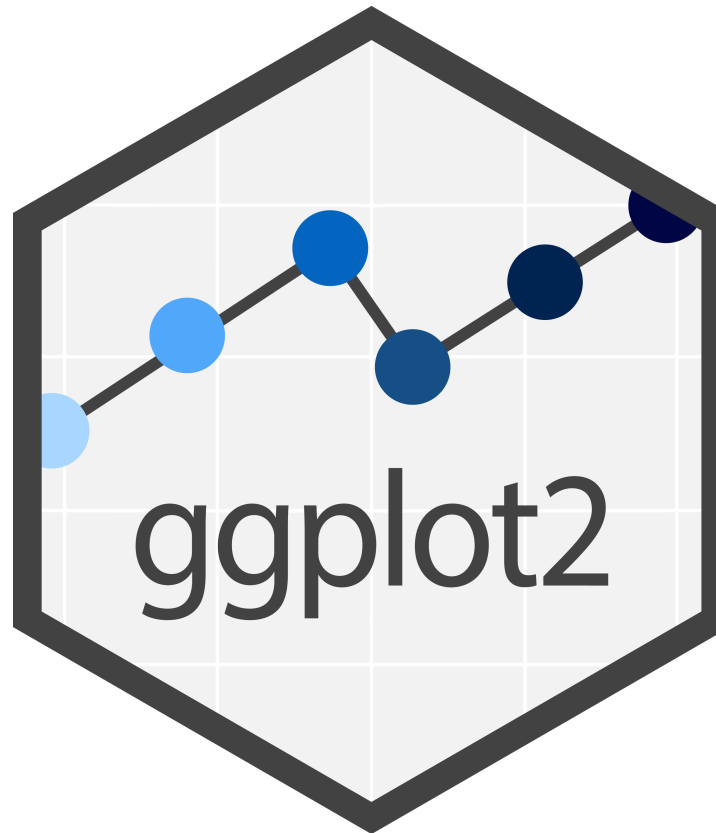
O poder da visualização

- DataViz, para os íntimos
 - Muito importante já na parte de exploração dos dados
 - Ajudam ao cérebro a dar sentido em todas aquelas informações
 - Ajuda a levantar hipóteses
 - Importante na apresentação final dos resultados



ggplot2

Mas onde está o carregamento desse pacote no nosso software?



O que vamos (começar) aprender hoje

- Uma forma mais rápida de visualizar (poucas linhas de código)
 - Sem (muitas) firulas visuais
- Não vamos focar, no momento, em deixar os gráficos perfeitos.
- Adivinhem: o ggplot faz as duas coisas!
 - Data-to-Viz
 - R Gallery
 - Bruno Miotto

A lógica do ggplot2

- Utiliza a Gramática dos gráficos (daí o gg de ggplot): você precisa de atributos estéticos (cor, tamanho...) e atributos geométricos (pontos, linhas, colunas...).
- São utilizadas camadas: Sobreposição de elementos visuais.

A lógica do ggplot

Primeira camada

Segunda camada

Terceira camada

Firulas

O “canvas” vazio

```
1 iris_media_area %>%  
2   ggplot()
```


Lógica do ggplot

As camadas são adicionadas com um “+”

- Diferença + para %>%
- Há as partes *estáticas* e partes que *dinâmicas* (que respondem aos dados)
 - estão dentro do aes().
- Cuidado com as aspas (principalmente nas opções de parâmetros)!
- Conheça as variáveis (numéricas ou categóricas)
- É preciso ter no mínimo a função ggplot() e uma função de geometria geom_?()
- Há várias formas de escrever a mesma coisa!

Várias formas de escrever a mesma coisa

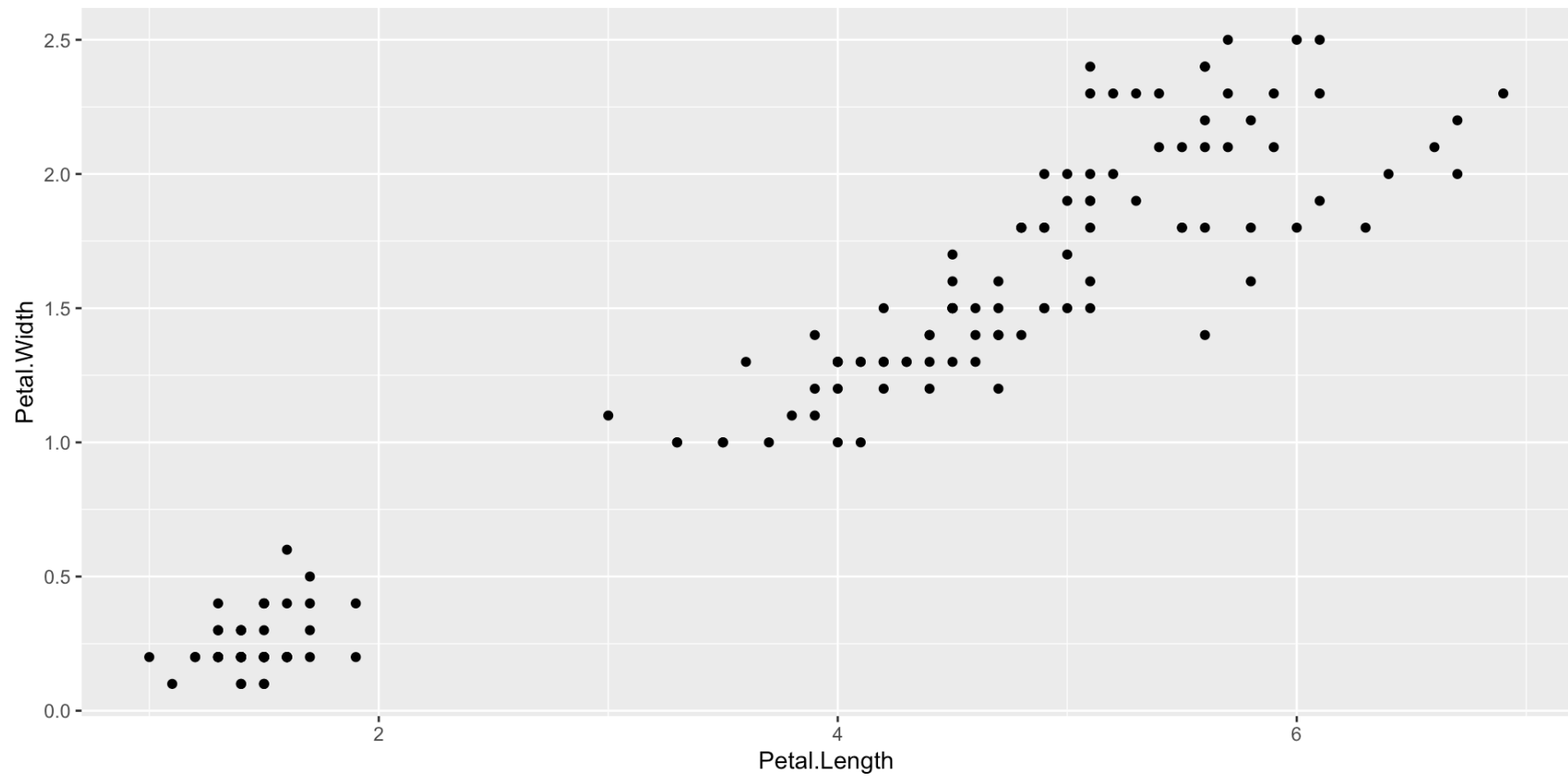
Opção 1

Opção 2

Opção 3

Opção 4

```
1 ggplot(data = iris) +  
2   geom_point(mapping = aes(x = Petal.Length, y = Petal.Width))
```



Customizações

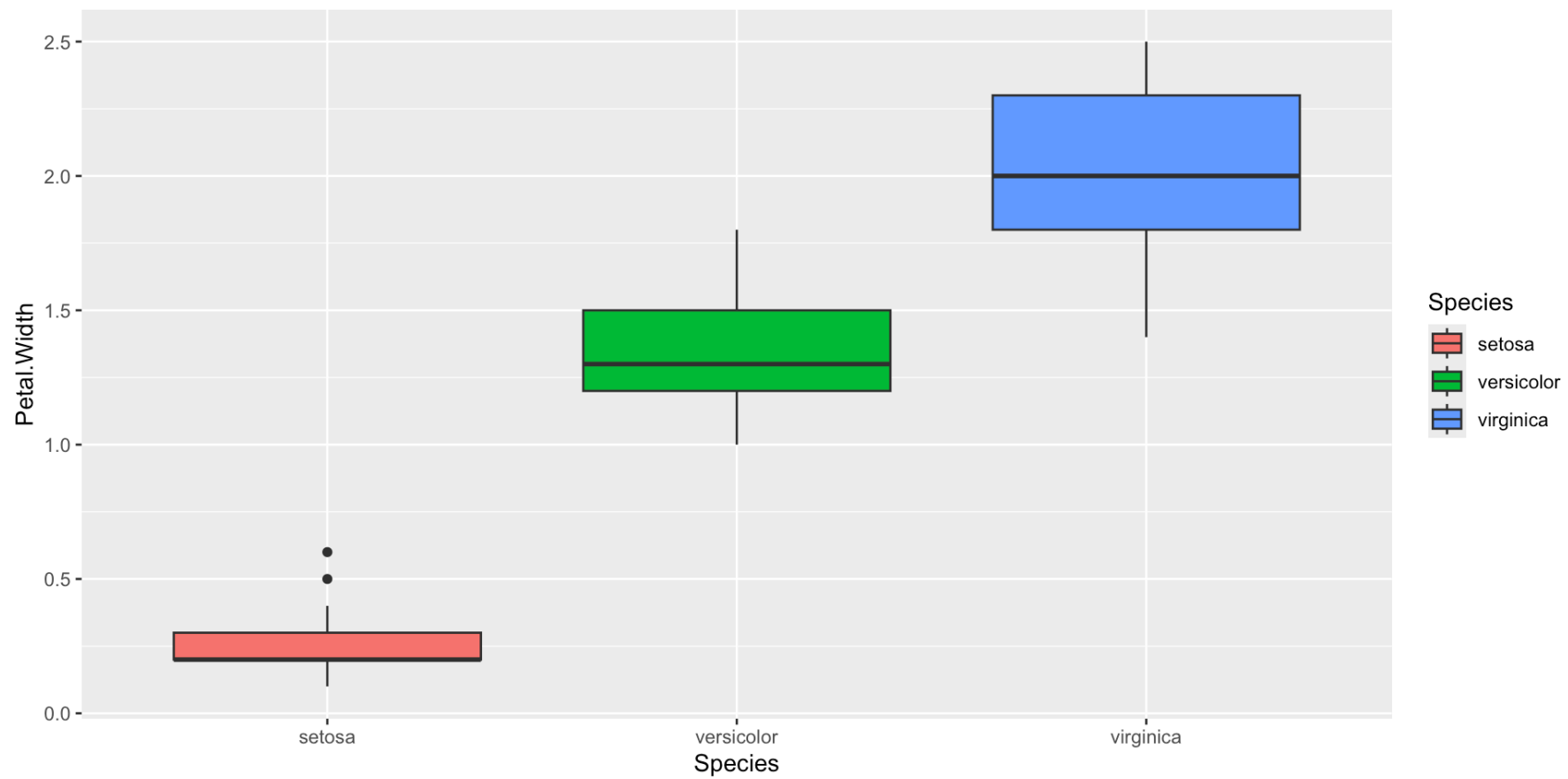
[illegible]

Desafio (#Forms)

- O que acontece se o color estiver fora do aes()?
- Em um gráfico de colunas ou boxplot
 - Diferença do **fill**= dentro ou fora do **aes()** ?

```
1 graf <- ggplot(iris) +  
2   aes(x = Species, y = Petal.Width, fill=Species) +  
3   geom_boxplot()
```

```
1 graf
```



PRÁTICA

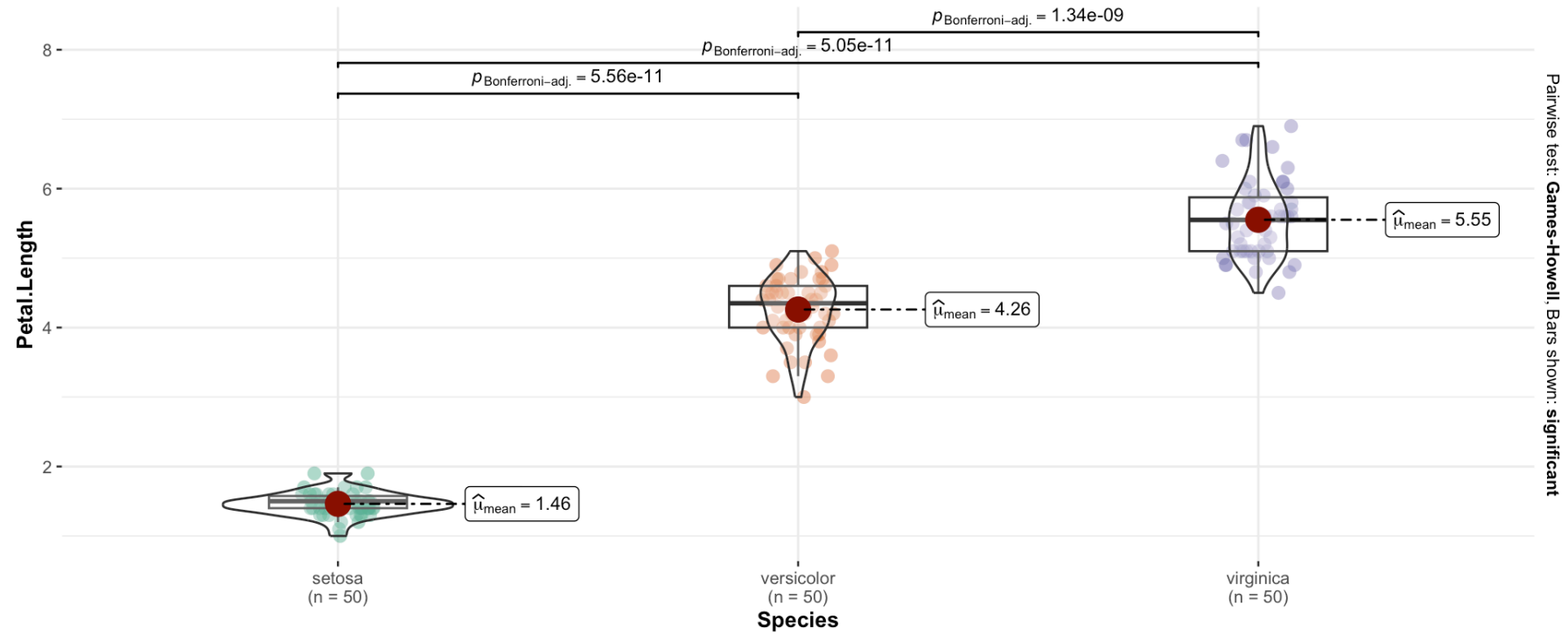
pacote ggstatsplot

- Documentação

```
1 ggbetweenstats(  
2   data = iris,  
3   x = Species,  
4   y = Petal.Length,  
5   p.adjust.method = "bonferroni",  
6   title = "Comprimento das Pétalas entre Espécies",  
7   type = "parametric",  
8  
9   bf.message = TRUE  
10 )
```

Comprimento das Pétalas entre Espécies

$F_{\text{Welch}}(2, 78.07) = 1828.09$, $p = 2.69\text{e-}66$, $\hat{\omega}_p^2 = 0.98$, $\text{CI}_{95\%} [0.97, 1.00]$, $n_{\text{obs}} = 150$



$\log_e(\text{BF}_{01}) = -201.27$, $\hat{R}_{\text{Bayesian}}^2 = 0.94$, $\text{CI}_{95\%}^{\text{HDI}} [0.94, 0.94]$, $r_{\text{Cauchy}}^{\text{JZS}} = 0.71$

Pacotes para tabelas

- DT (iterativos em html)
- kable, kableExtra (mais “LateX friendly”)
- formattable, gt, stargazer...
- gtsummary

flextable

- Ótimo para gerar os arquivos em word!

```
1 tab1 <- iris_media_area %>%  
2 flextable()
```


Lógica...

Gosto de pensar que também funciona como “camadas”.

- alinhamento
- fonte
- largura
- cores...

width : Largura

```
1 flextable( iris_media_area) %>%  
2   width(width=3)
```

Species	area_media	area_dp
setosa	0.182800	0.09057729
versicolor	2.860200	0.68420143
virginica	4.963387	0.66677756

```
1 flextable(iris_media_area) %>%  
2   width(width=1) %>%  
3   width (j=1, width=2)
```

Species	area_media	area_dp
setosa	0.182800	0.09057729
versicolor	2.860200	0.68420143
virginica	4.963387	0.66677756

bold: Negrito

- podemos usar a mesma lógica para as colunas
- Mas com o negrito, e outras formatações internas podemos também trabalhar nas **linhas**.
- **j** eram as colunas ...
- **i** são as linhas

```
1 flextable(iris_media_area) %>%  
2   width(width=1) %>%  
3   width(j=1, width=3) %>%  
4   bold(i=1)
```

Species	area_media	area_dp
setosa	0.182800	0.09057729
versicolor	2.860200	0.68420143
virginica	4.963387	0.66677756

```
1 flextable(iris_media_area) %>%
```

```
2 width(width=1) %>%
3 width(j=1, width=3) %>%
4 bold(part= "header")
```

Species	area_media	area_dp
setosa	0.182800	0.09057729
versicolor	2.860200	0.68420143
virginica	4.963387	0.66677756

alinhamento

```
1 flextable(iris_media_area) %>%  
2   width(width=1) %>%  
3   width (j=1, width=2) %>%  
4   bold (part = "header") %>%  
5   align(align="center", part="all") %>%  
6   align(align="left", j=1, part="all")
```

Species	area_media	area_dp
setosa	0.182800	0.09057729
versicolor	2.860200	0.68420143
virginica	4.963387	0.66677756

Mais firulas

```
1 flextable(iris_media_area) %>%
2   width(width=1) %>%
3   width (j=1, width=0.5) %>%
4   bold (part = "header") %>%
5   align(align="center", part="all") %>%
6   align(align="left", j=1, part="all") %>%
7   vline(j=1, border=fp_border(color = "purple",
8                               style = "dotted",
9                               width = 1) ) %>%
10  set_header_labels(
11    Species="Espécie") %>%
12  compose(
13    i = 1, j = c(2,3), part = "header",
14    value = c(as_paragraph(
15      "Área Média / cm",
16      as_sup("2")),
17      as_paragraph(
18        "Desvio Padrão / cm",
19        as_sup("2")))
```

Espécie	Área Média / cm ²	Desvio Padrão / cm ²
setosa	0,18	0,09
versicolor	2,86	0,68

gtsummary

- Relatorios Clínicos com gtsummary
- Documentação . . .

gtsummary

```
1 iris_filt <- iris %>%
2   mutate(Species=as.character(Species)) %>%
3   filter(Species!="setosa")
4
5 iris_filt %>%
6   select(-Species) %>%
7   tbl_summary()
```

Characteristic	N = 100 ¹
Sepal.Length	6.30 (5.80, 6.70)
Sepal.Width	2.90 (2.70, 3.03)
Petal.Length	4.90 (4.38, 5.53)
Petal.Width	1.60 (1.30, 2.00)
¹ Median (IQR)	

```
1 iris_filt %>%
2   tbl_summary(by= Species)
```

Characteristic	versicolor, N = 50 ¹	virginica, N = 50 ¹
Sepal.Length	5.90 (5.60, 6.30)	6.50 (6.23, 6.90)
Sepal.Width	2.80 (2.53, 3.00)	3.00 (2.80, 3.18)
Petal.Length	4.35 (4.00, 4.60)	5.55 (5.10, 5.88)

Petal.Width	1.30 (1.20, 1.50)	2.00 (1.80, 2.30)
¹ Median (IQR)		

gtsummary

```
1 iris_filt %>%
2   tbl_summary(by= Species,
3               statistic = list(all_continuous() ~ "{mean} ({sd})")
```

Characteristic	versicolor, N = 50 ¹	virginica, N = 50 ¹
Sepal.Length	5.94 (0.52)	6.59 (0.64)
Sepal.Width	2.77 (0.31)	2.97 (0.32)
Petal.Length	4.26 (0.47)	5.55 (0.55)
Petal.Width	1.33 (0.20)	2.03 (0.27)
¹ Mean (SD)		

```
1 iris_filt %>%
2   tbl_summary(by= Species,
3               statistic = list(
4                 all_continuous() ~ "{mean} ({sd})") ) %>%
5   add_p() %>%
6   bold_labels() %>%
7   italicize_levels() %>%
8   bold_p(t = 0.8)
```

Characteristic	versicolor, N = 50 ¹	virginica, N = 50 ¹	p-value ²
Sepal.Length	5.94 (0.52)	6.59 (0.64)	<0.001

Sepal.Width	2.77 (0.31)	2.97 (0.32)	0.005
Petal.Length	4.26 (0.47)	5.55 (0.55)	<0.001
Petal.Width	1.33 (0.20)	2.03 (0.27)	<0.001

¹ Mean (SD)

² Wilcoxon rank sum test

```

1 iris_filt %>%
2   tbl_summary(by= Species,
3               statistic = list(
4                 all_continuous() ~ "{mean} ({sd}))" ) ) %>%
5   add_p(test= everything() ~"t.test") %>%
6   bold_labels() %>%
7   italicize_levels() %>%
8   bold_p(t = 0.8)

```

Characteristic	versicolor, N = 50 ¹	virginica, N = 50 ¹	p-value ²
Sepal.Length	5.94 (0.52)	6.59 (0.64)	<0.001
Sepal.Width	2.77 (0.31)	2.97 (0.32)	0.002
Petal.Length	4.26 (0.47)	5.55 (0.55)	<0.001
Petal.Width	1.33 (0.20)	2.03 (0.27)	<0.001

¹ Mean (SD)

² Welch Two Sample t-test

Prática Tabelas

Obrigada!

