

R para análise de dados em Psicologia

Uma Introdução

Francisco Pablo Huascar Aragão Pinheiro

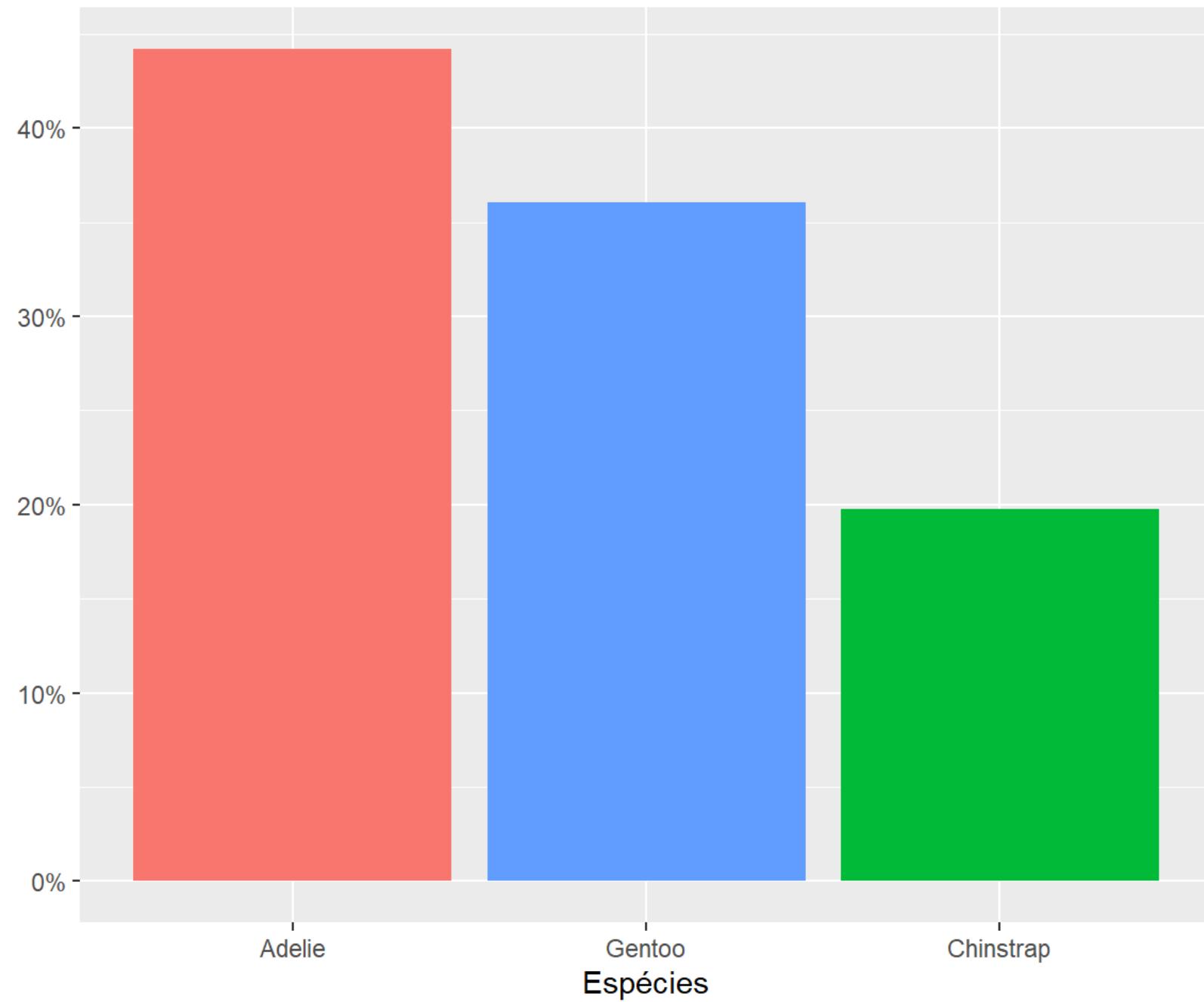


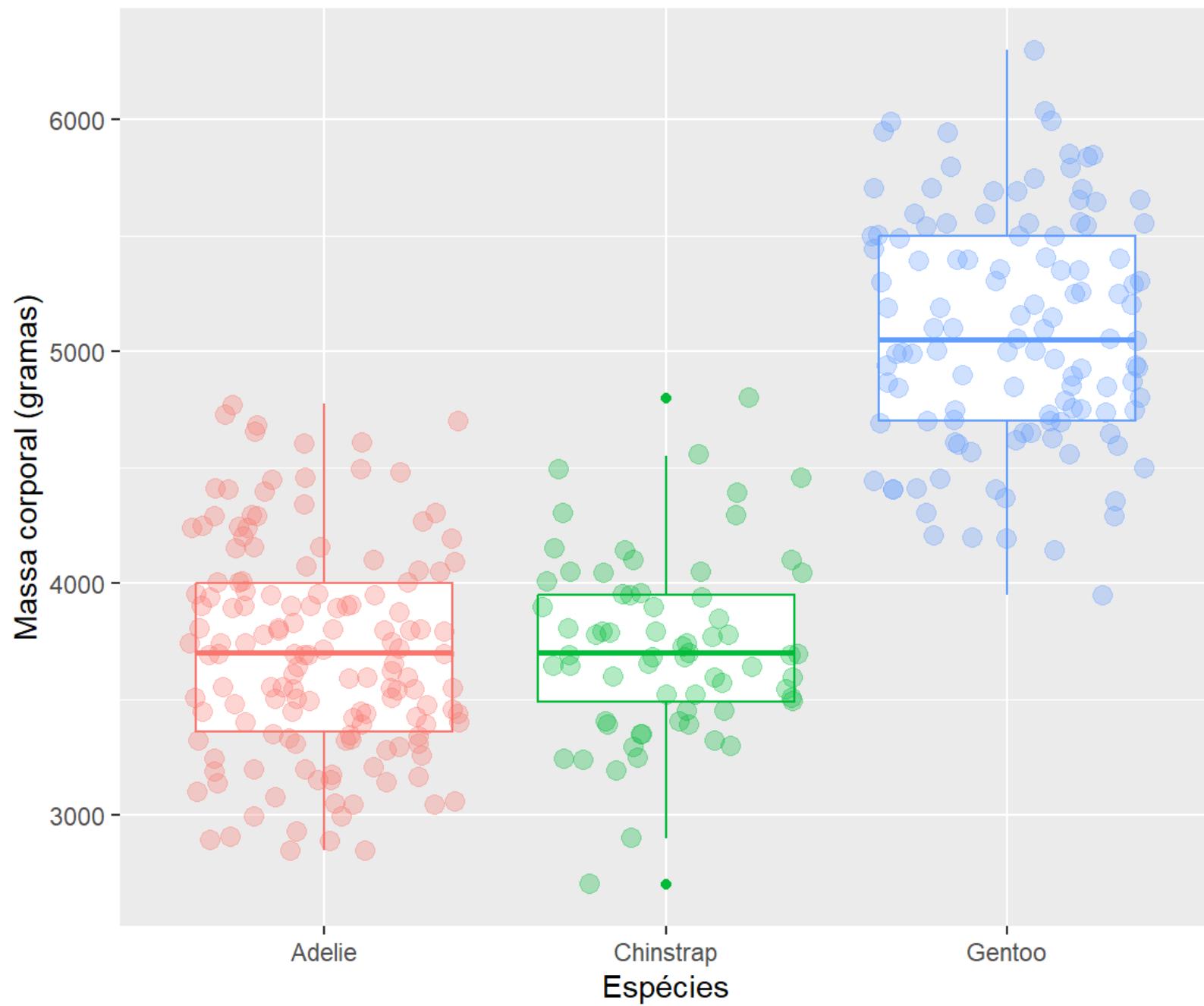
Quem sou eu



- Psicólogo
- Doutor em educação
- Professor do campus Sobral da UFC
- Recentemente, um entusiasta do R

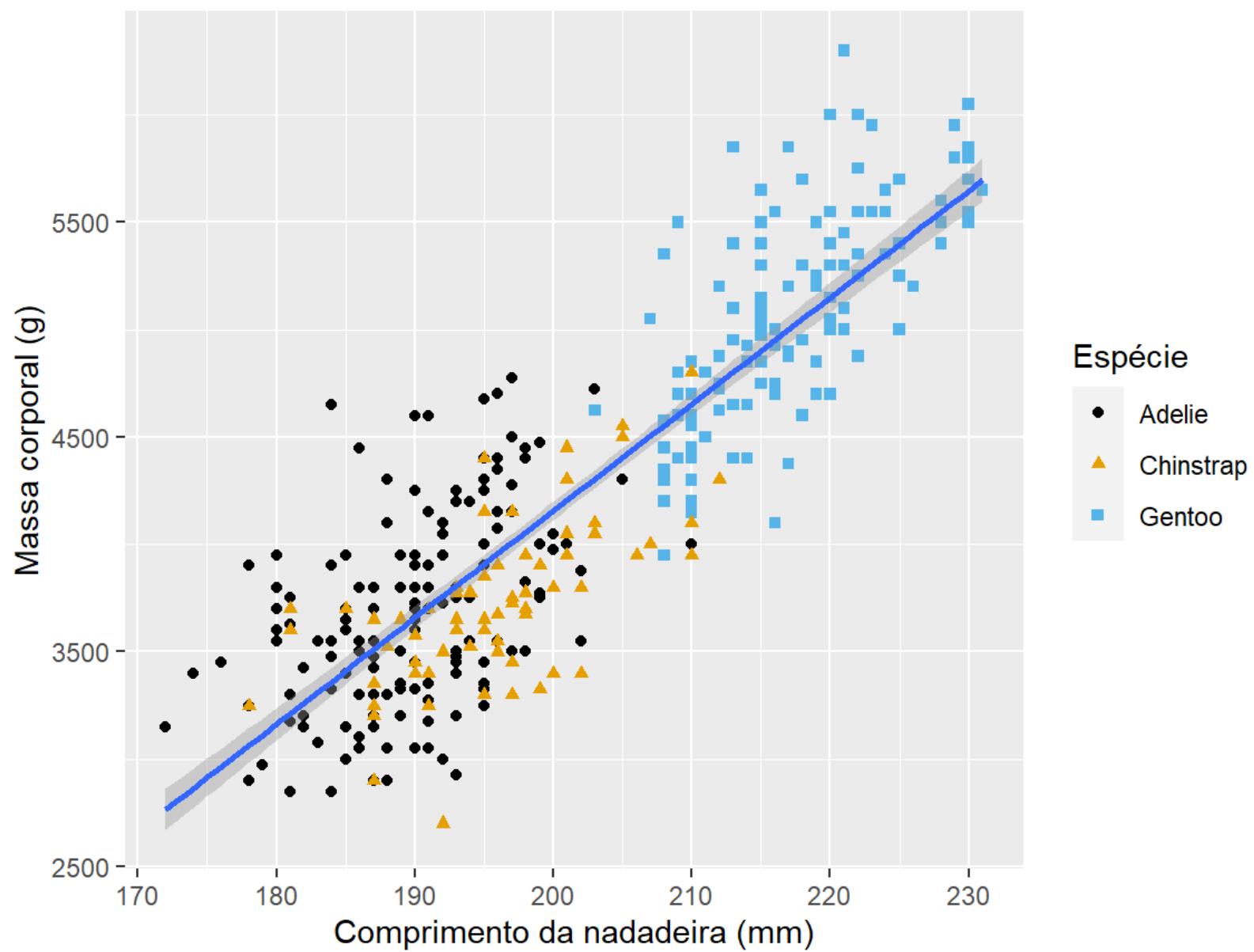
Gráficos

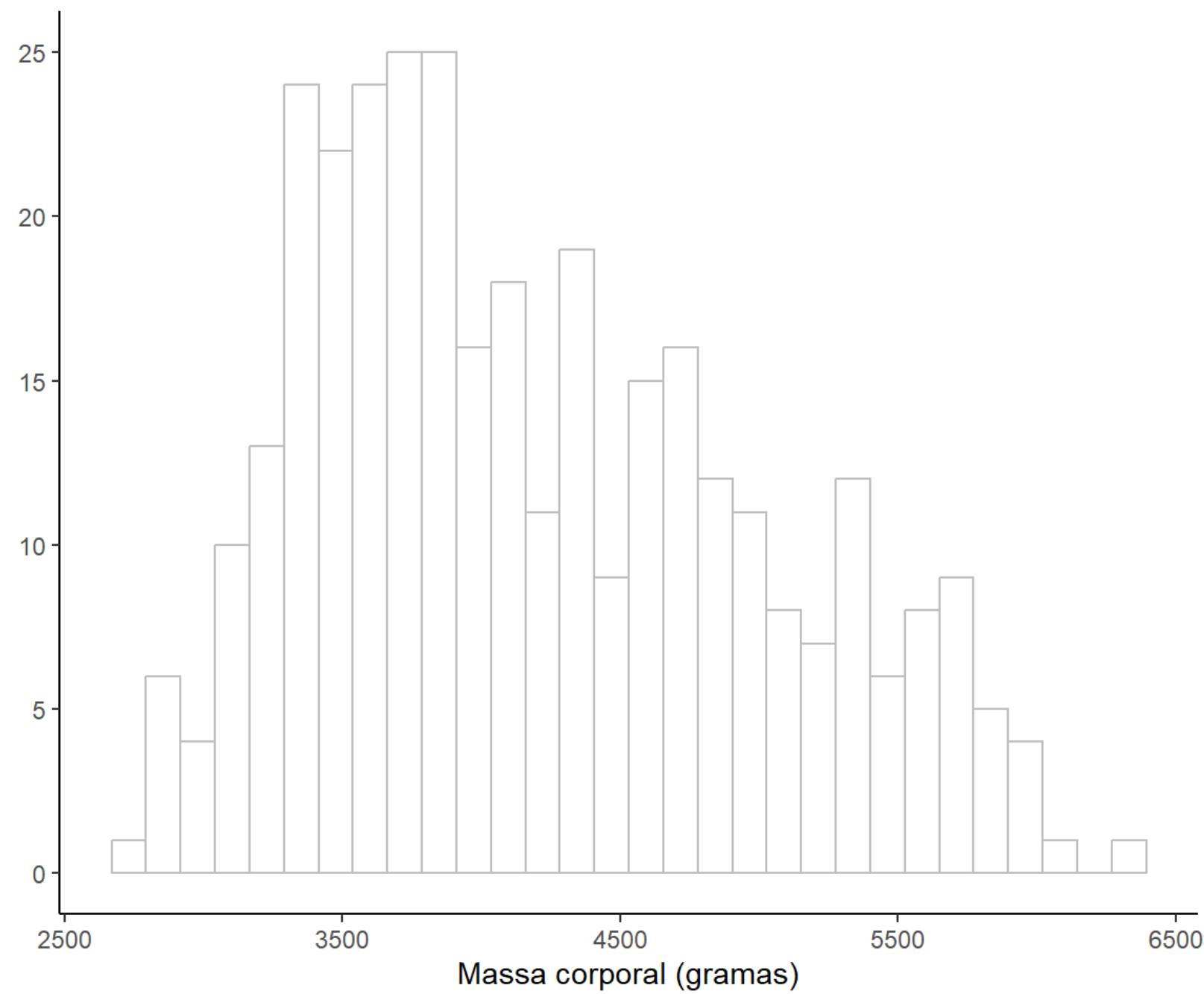




Massa corporal e comprimento da nadadeira

Dimensões para Adelie, Chinstrap e Gentoo

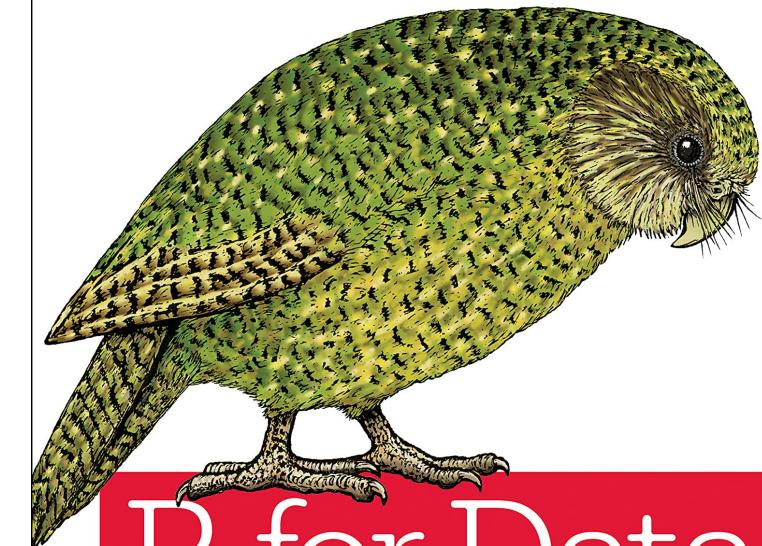




R for Data Science

Livro base para o mini curso

O'REILLY®



R for Data Science

IMPORT, TIDY, TRANSFORM, VISUALIZE, AND MODEL DATA

Hadley Wickham &
Garrett Grolemund

R for Data Science

- Neste [link](#) você pode acessar a primeira edição do livro:

<https://r4ds.had.co.nz/>

- E [aqui](#) você pode acessar a segunda edição que ainda está sendo finalizada:

<https://r4ds.hadley.nz/>

ggplot2: Elegant Graphics for Data Analysis

ggplot2: Elegant Graphics for Data Analysis

- Neste [link](#) você pode acessar a terceira edição do livro que ainda está sendo finalizada:

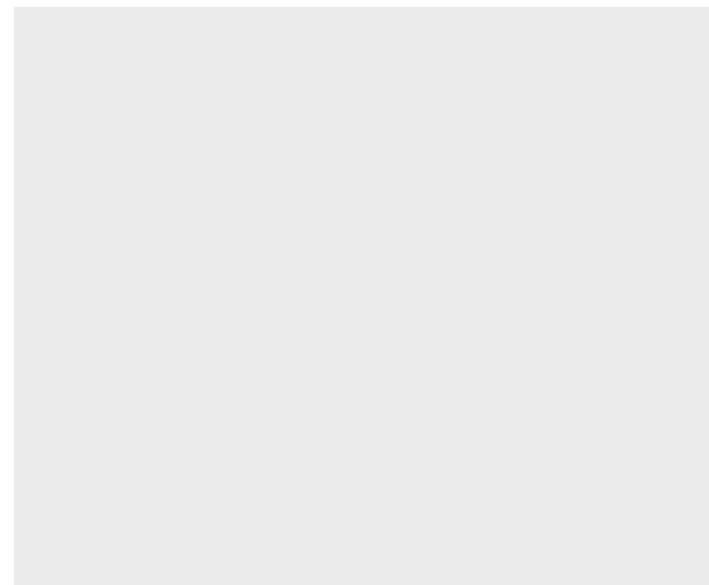
<https://ggplot2-book.org/index.html>

2 Visualização de Dados

2.1 Introdução

- **ggplot2:** grammar of graphics (gramática dos gráficos)
 - Sistema coerente para descrever e construir gráficos

data mapping geom_



Pacotes

```
1 library(tidyverse)
```

there is no package called 'tidyverse'

```
1 install.packages("tidyverse")
2 library(tidyverse)
```

```
1 library(palmerpenguins)
2 library(ggthemes)
```



2.2 Primeiros Passos

- Os pinguins com nadadeiras mais longas pesam mais ou menos que os pinguins com nadadeiras mais curtas?
- Como é a relação entre o comprimento da nadadeira e a massa corporal? É positivo? Negativo? Linear? Não linear?
- A relação varia de acordo com a espécie do pinguim?
- O que nos diz a ilha onde mora o pinguim?
- Vamos criar visualizações que podemos usar para responder a essas perguntas.

O banco de Dados dos penguins

- Bando de dados: uma coleção retangular de variáveis (nas colunas) e observações (nas linhas)
- *penguins*: 344 casos/observações
- Os dados foram coletados e disponibilizados pela Dra. Kristen Gorman e a Estação Palmer, Antártica LTER

Algumas Definições

- **Variável:** uma quantidade, qualidade ou propriedade que você pode medir
- **Valor:** o estado de uma variável quando você a mede
 - O valor de uma variável pode mudar de medição para medição

Algumas Definições

- **Observação/caso:** conjunto de medições feitas em condições semelhantes
 - Todas as medições em uma observação são feitas ao mesmo tempo e no mesmo objeto
- Uma observação/caso conterá vários valores, cada um associado a uma variável diferente.
- **penguins**
 - Variável: atributo de todos os pinguins
 - Observação: todos os atributos de um único pinguim

Dados Tabulares

- Conjunto de valores, cada um associado a uma variável e a uma observação.
- Os dados tabulares são organizados (“tidy”) se cada valor for colocado em sua própria “célula”, cada variável em sua própria coluna e cada observação em sua própria linha

Tidy data

country	year	cases	population
Afghanistan	1990	745	1987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

variables

country	year	cases	population
Afghanistan	1990	745	1987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

observations

country	year	cases	population
Afghanistan	1990	745	1987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	128042583

values

tibble

```
1 penguins

# A tibble: 344 × 8
  species island      bill_length_mm bill_depth_mm flipper_length_mm
  <fct>   <fct>          <dbl>           <dbl>             <int>
body_mass_g
  <int>
  1 Adelie Torgersen      39.1            18.7            181
3750
  2 Adelie Torgersen      39.5            17.4            186
3800
  3 Adelie Torgersen      40.3            18               195
3250
  4 Adelie Torgersen       NA              NA              NA
NA
  5 Adelie Torgersen      36.7            19.3            193
3450
  # ... with 344 more rows, and 1 more variable:
  #   name: sex <fct>
```

- Oito colunas (variáveis)

Variáveis

```
1 glimpse(penguins)
```

```
Rows: 344  
Columns: 8  
$ species <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie,  
Adel...  
$ island <fct> Torgersen, Torgersen, Torgersen, Torgersen,  
Torgers...  
$ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1,  
...  
$ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1,  
...  
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190,  
186...  
$ body_mass_g <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475,  
...  
$ sex <fct> male, female, female, NA, female, male, female,  
... - 7 -
```

Variáveis

1 View (penguins)

encontro_1_cap_3.qmd x encontro_2_cap_2.qmd* penguins

Filter

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
1	Adelie	Torgersen	39.1	18.7	181	3750	male	2007
2	Adelie	Torgersen	39.5	17.4	186	3800	female	2007
3	Adelie	Torgersen	40.3	18.0	195	3250	female	2007
4	Adelie	Torgersen	NA	NA	NA	NA	NA	2007
5	Adelie	Torgersen	36.7	19.3	193	3450	female	2007
6	Adelie	Torgersen	39.3	20.6	190	3650	male	2007
7	Adelie	Torgersen	38.9	17.8	181	3625	female	2007
8	Adelie	Torgersen	39.2	19.6	195	4675	male	2007
9	Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
10	Adelie	Torgersen	42.0	20.2	190	4250	NA	2007
11	Adelie	Torgersen	37.8	17.1	186	3300	NA	2007
12	Adelie	Torgersen	37.8	17.3	180	3700	NA	2007
13	Adelie	Torgersen	41.1	17.6	182	3200	female	2007
14	Adelie	Torgersen	38.6	21.2	191	3800	male	2007
15	Adelie	Torgersen	34.6	21.1	198	4400	male	2007
16	Adelie	Torgersen	36.6	17.8	185	3700	female	2007
17	Adelie	Torgersen	38.7	19.0	195	3450	female	2007
18	Adelie	Torgersen	42.5	20.7	197	4500	male	2007
19	Adelie	Torgersen	34.4	18.4	184	3325	female	2007
20	Adelie	Torgersen	46.0	21.5	194	4200	male	2007
21	Adelie	Biscoe	37.8	18.3	174	3400	female	2007
22	Adelie	Biscoe	37.7	18.7	180	3600	male	2007

Showing 1 to 23 of 344 entries, 8 total columns

Variáveis no banco de dados dos pinguins

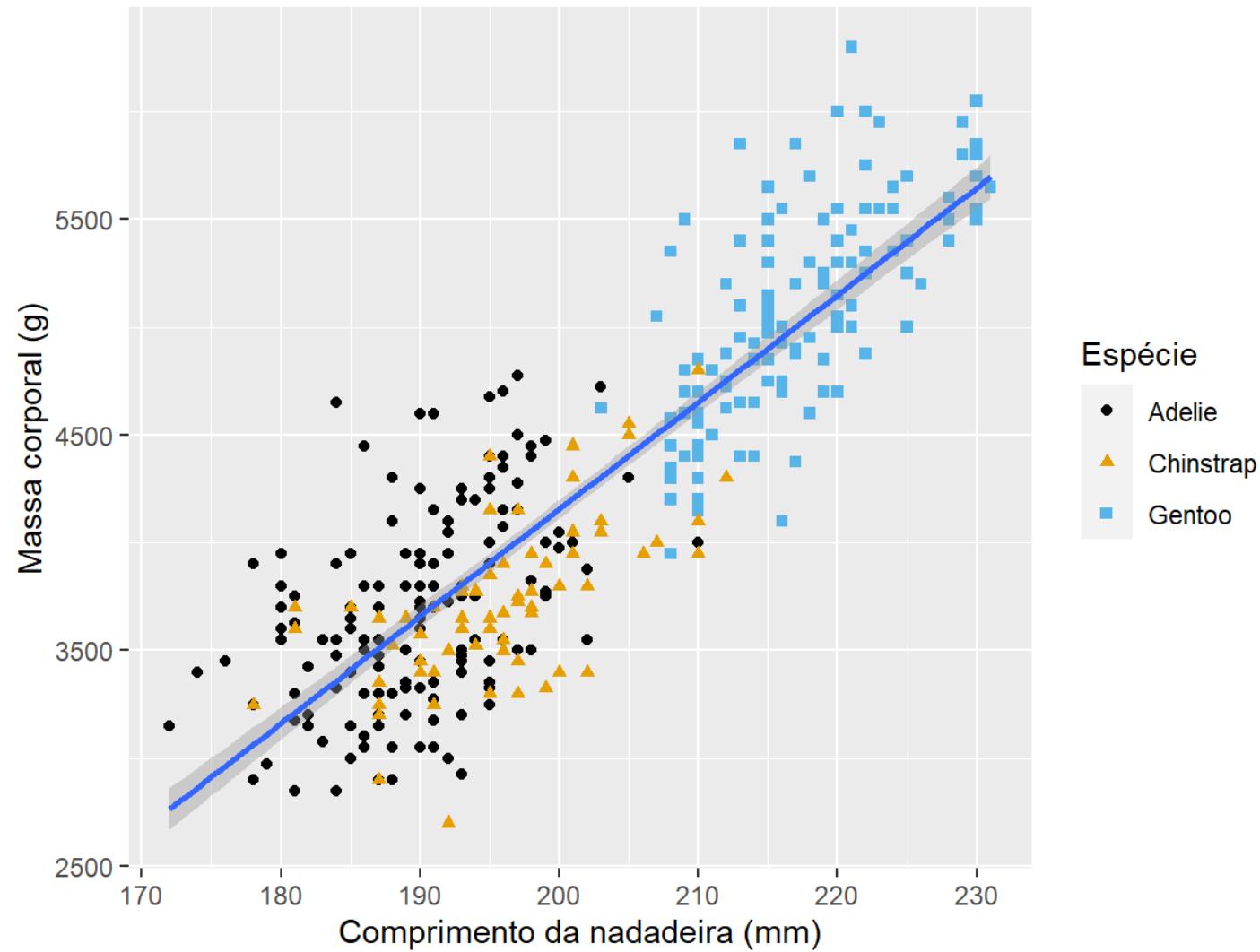
1. `species`: uma espécie de pinguim (Adelie, Chinstrap ou Gentoo)
2. `flipper_length_mm`: comprimento da nadadeira de um pinguim, em milímetros
3. `body_mass_g`: massa corporal de um pinguim, em gramas

Para saber mais sobre o banco de dados:

```
1 ?penguins
```

Objetivo final

Massa corporal e comprimento da nadadeira
Dimensões para Adelie, Chinstrap e Gentoo

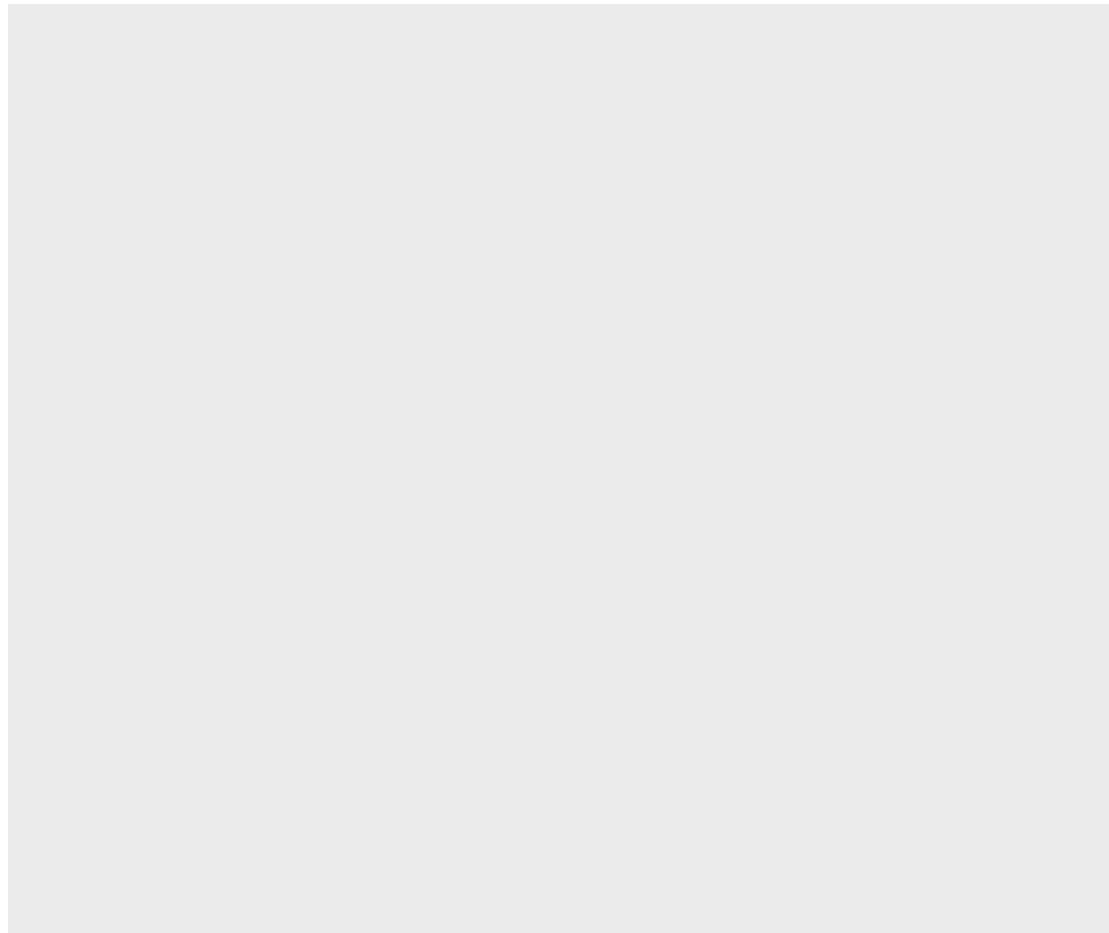


Criando um ggplot camada por camada

Banco de dados

data

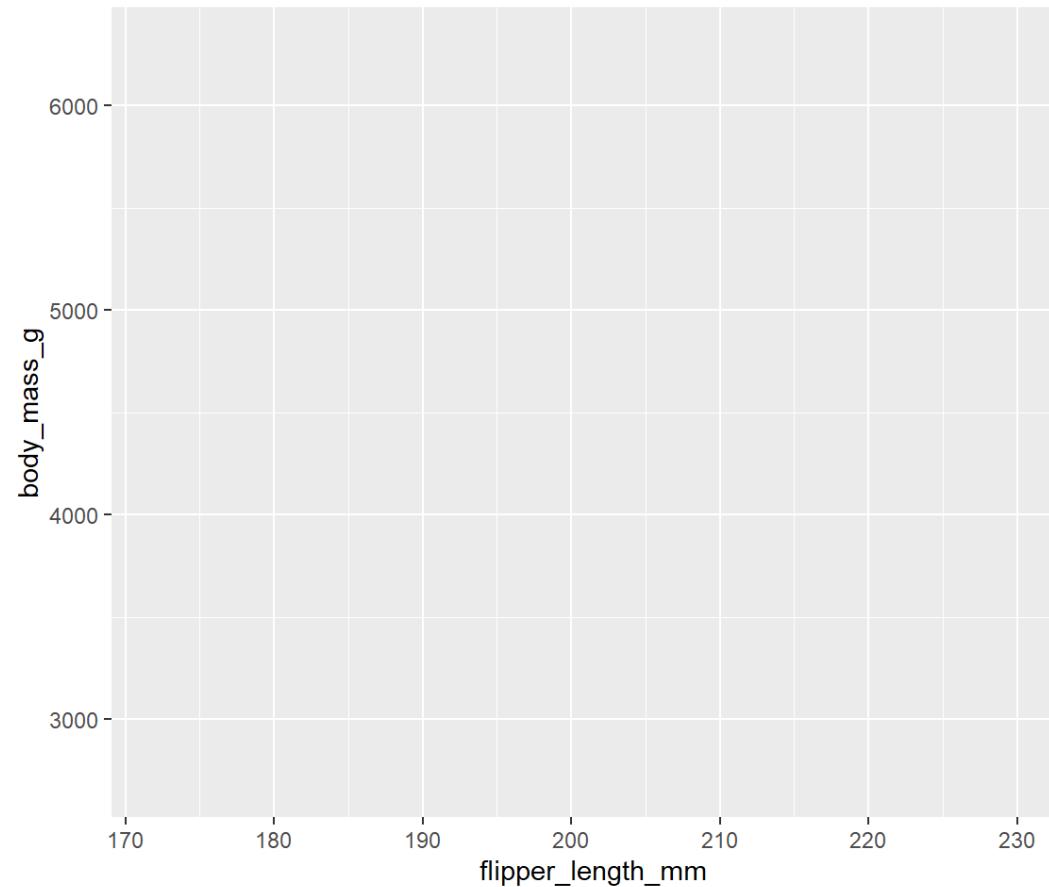
```
1 ggplot(data = penguins)
```



Representação visual das informações

mapping e aes(aesthetics - estética)

```
1 ggplot(data = penguins,  
2         mapping = aes(x = flipper_length_mm, y = body_mass_g))
```

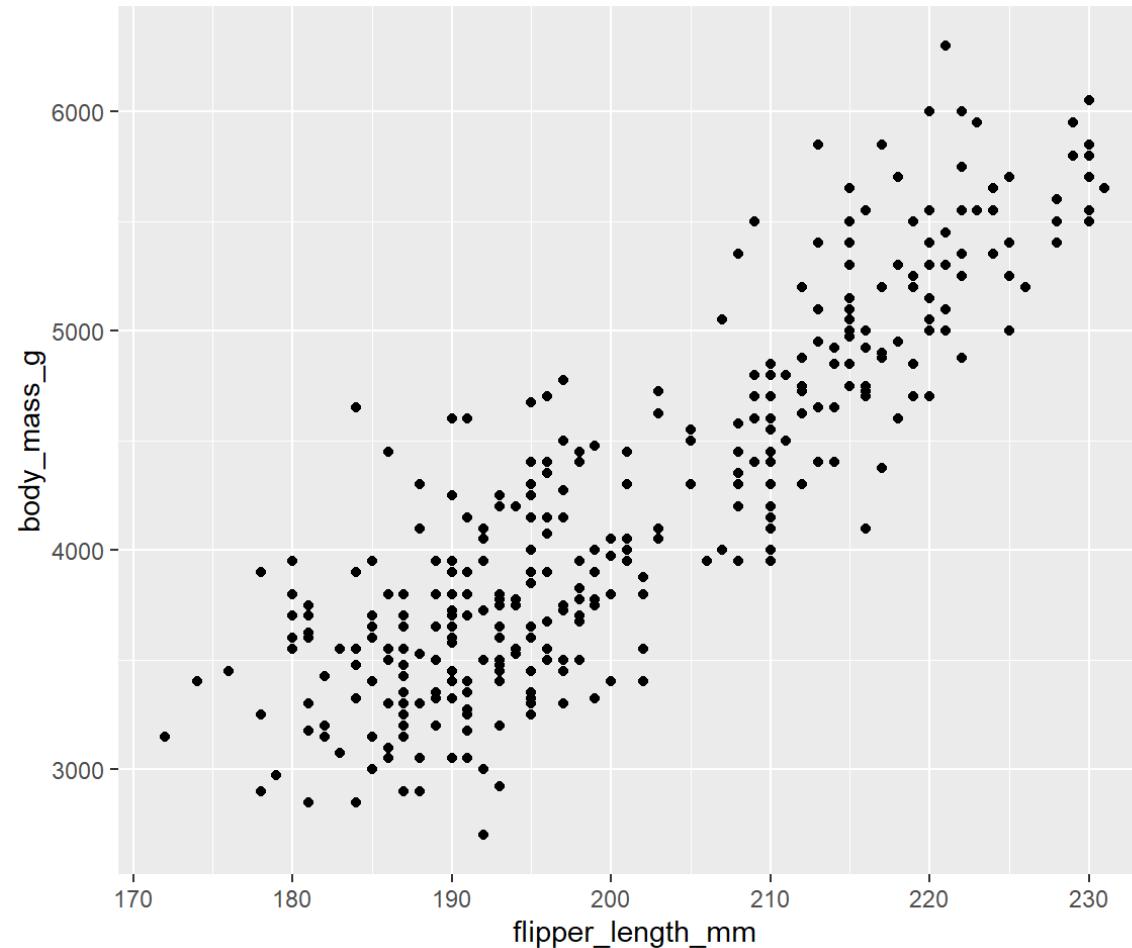


Como representar as informações?

- **geom**: o objeto geométrico que um gráfico usa para representar dados
- Funções que sempre começam com `geom_`
- Gráficos de barras: `geom_bar()`
- Gráficos de linhas: `geom_line()`
- Gráficos de caixa de bigodes: `geom_boxplot()`
- Gráficos de dispersão: `geom_point()`

geom_point()

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm, y = body_mass_g)) +
3   geom_point()
```



Mensagem de aviso (Warning message):

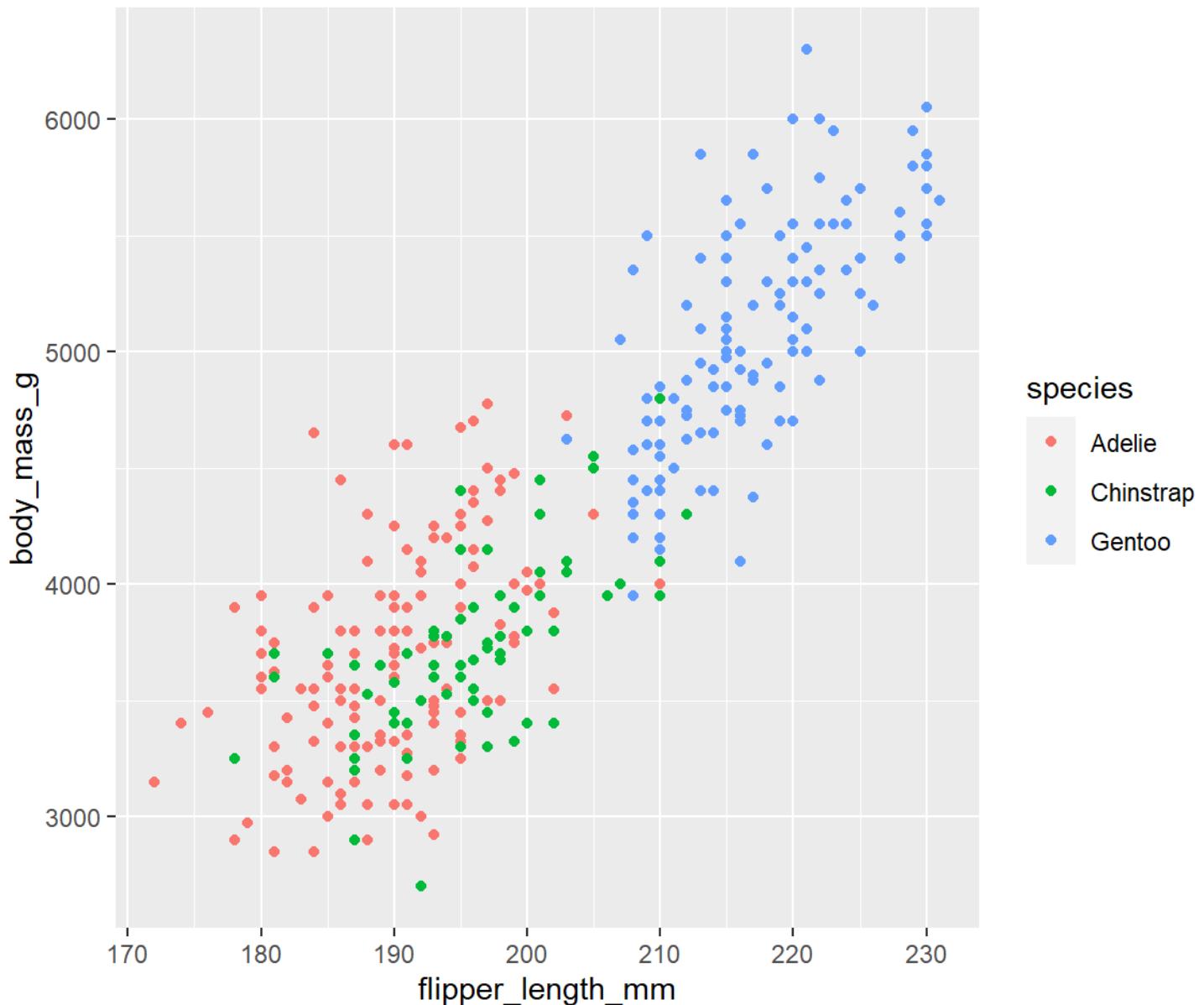
Removed 2 rows containing missing values
(`geom_point()`).

Adicionando uma estética (aesthetics)

color

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g,
4                 color = species)) +
5   geom_point()
```

Adicionando uma estética (aesthetics)

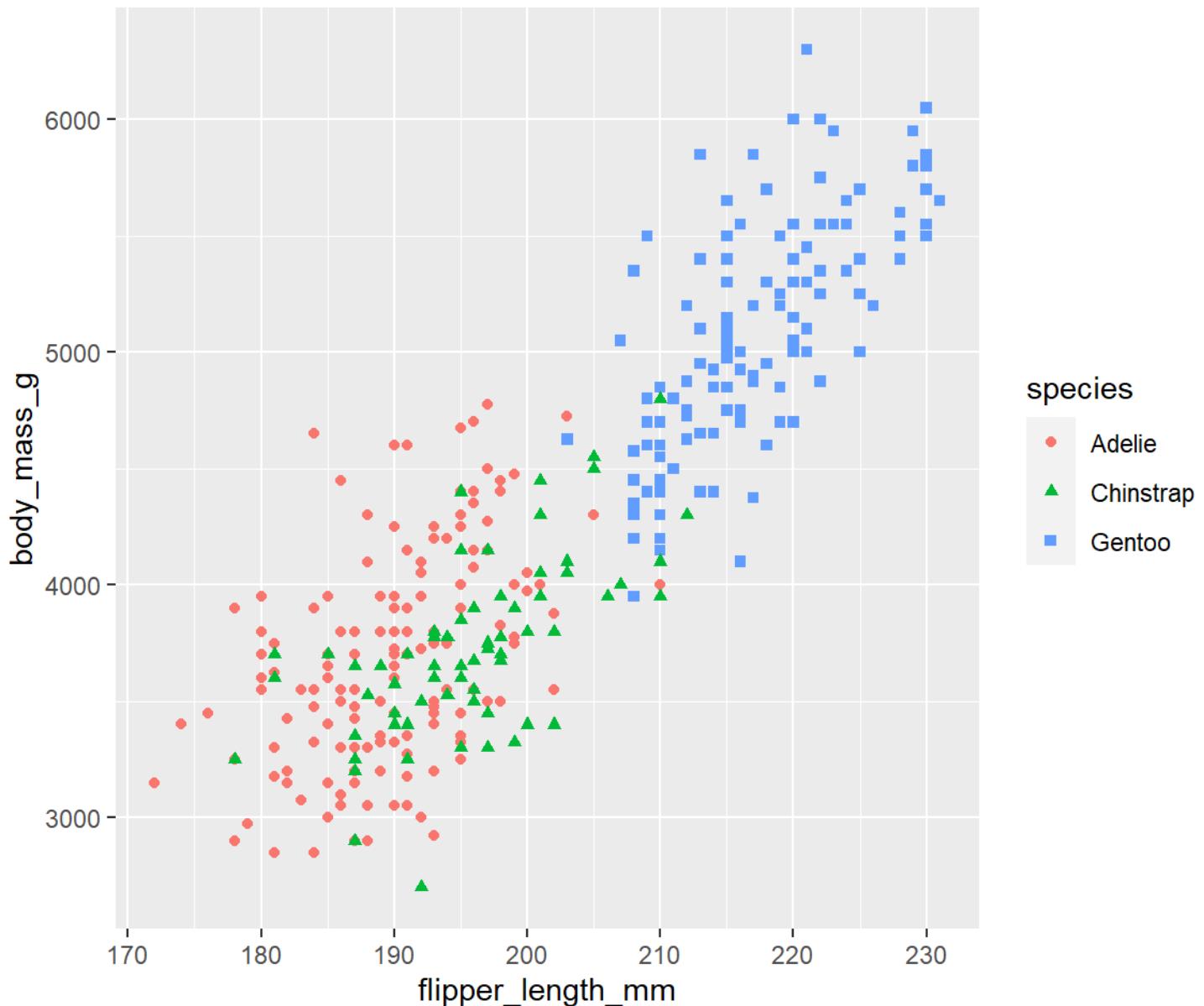


Adicionando uma estética (aesthetics)

shape

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g,
4                 color = species,
5                 shape = species)) +
6   geom_point()
```

Adicionando uma estética (aesthetics)

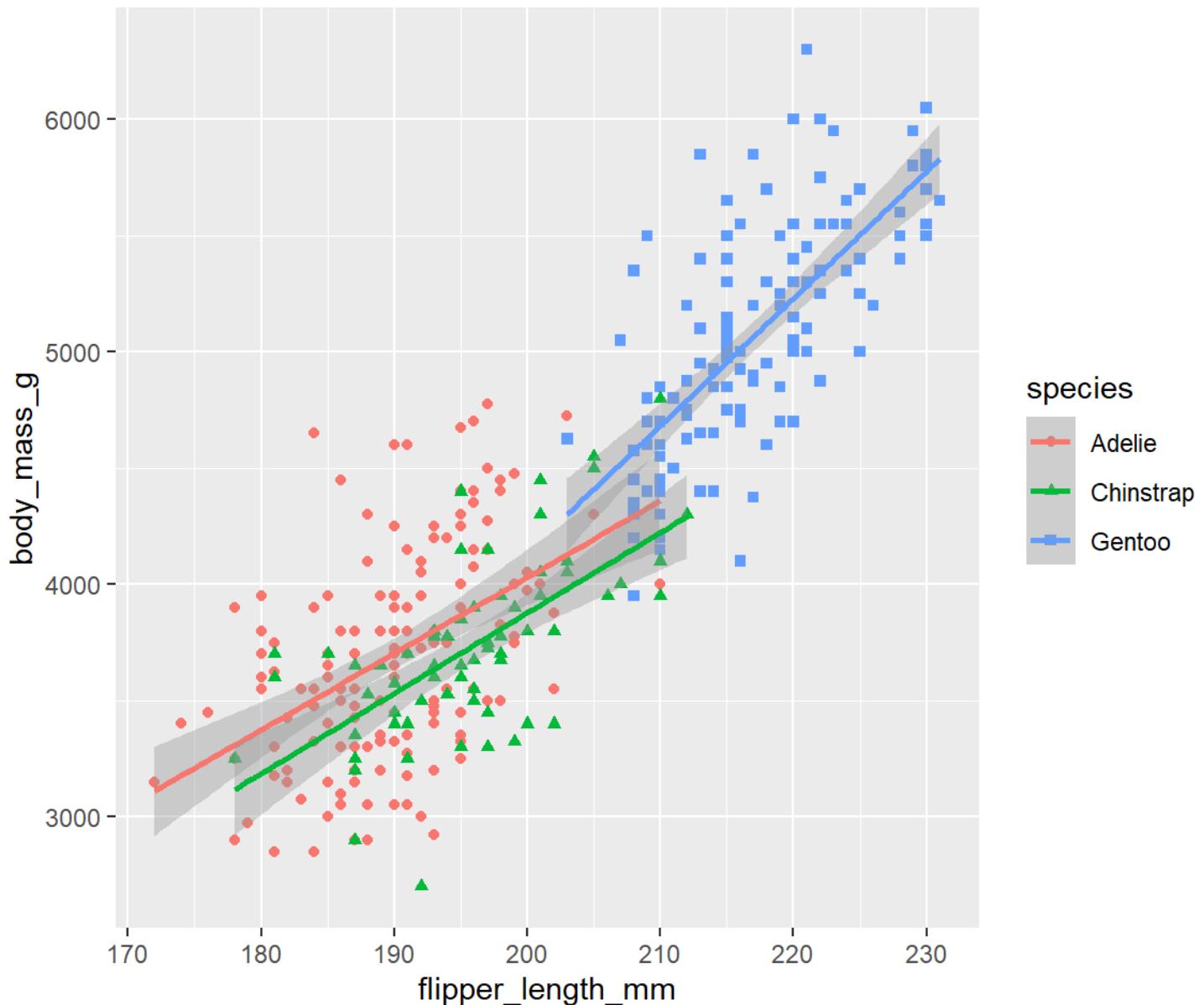


Adicionando uma camada (layer)

geom_smooth()

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g,
4                 color = species,
5                 shape = species)) +
6   geom_point() +
7   geom_smooth(method = "lm")
```

Adicionando uma camada (layer)

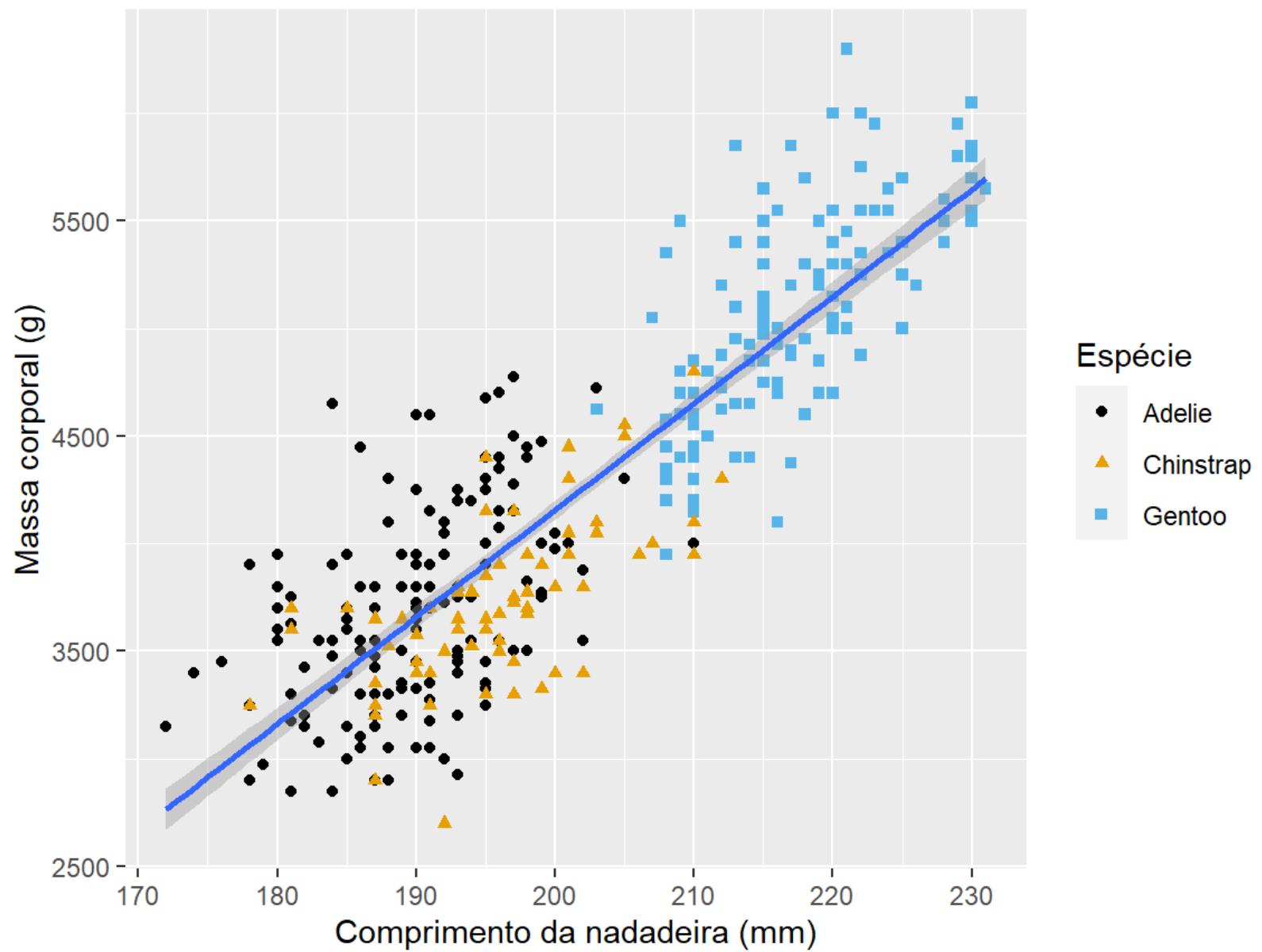


species

- Adelie
- Chinstrap
- Gentoo

Massa corporal e comprimento da nadadeira

Dimensões para Adelie, Chinstrap e Gentoo

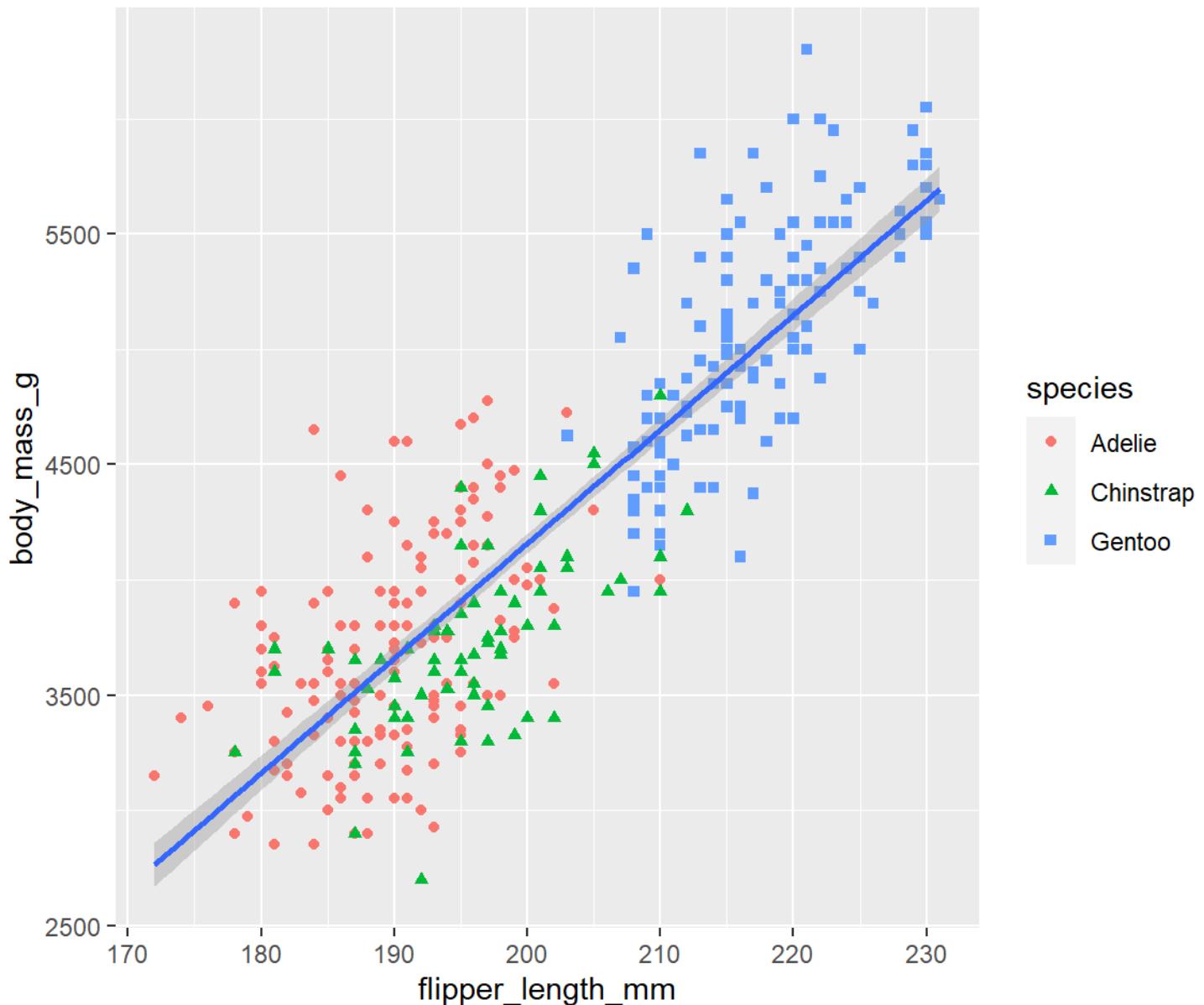


Adicionando uma camada (layer)

geom_smooth()

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g)) +
4   geom_point(aes(color = species,
5               shape = species)) +
6   geom_smooth(method = "lm")
```

Adicionando uma camada (layer)



Adicionando rótulos (labels)

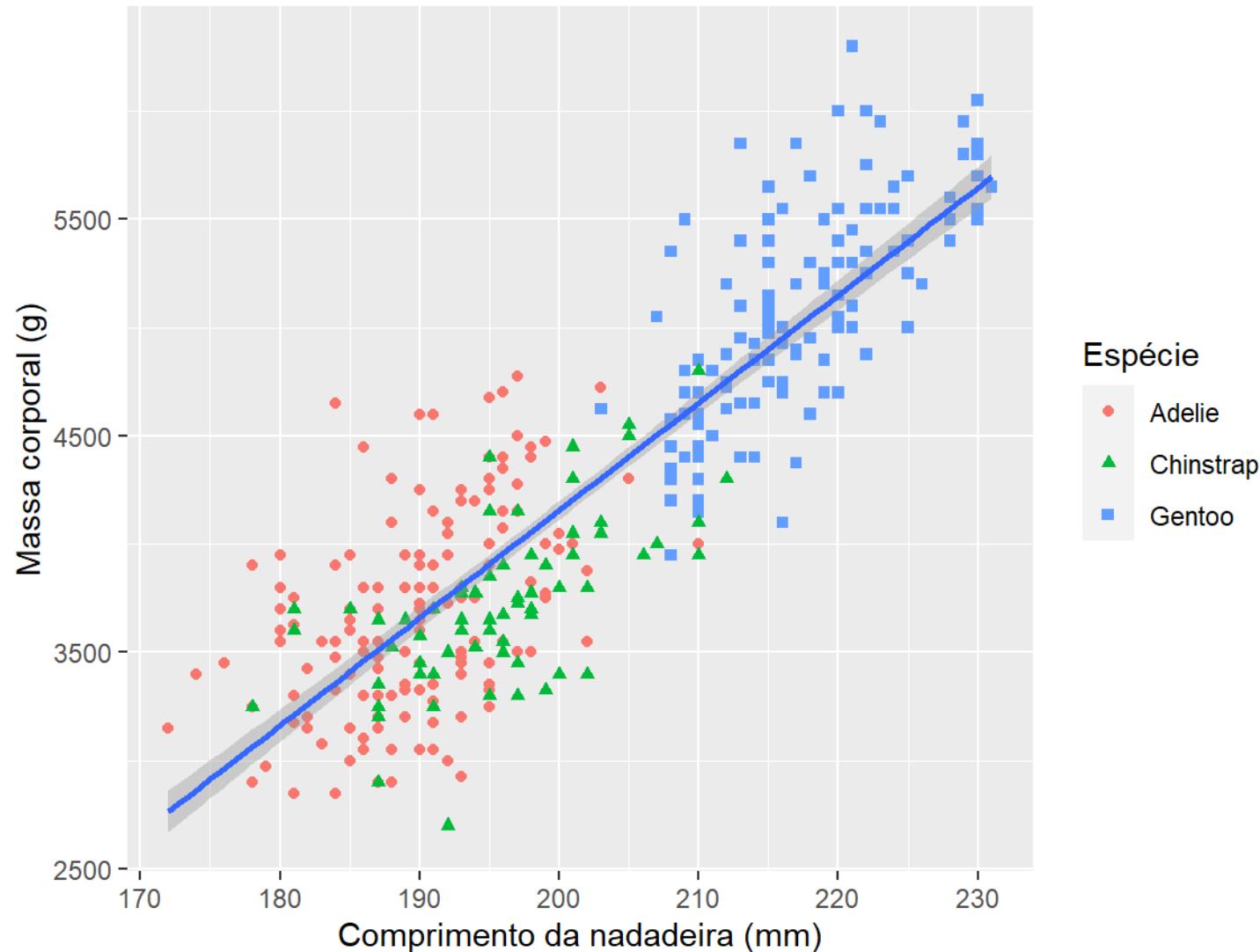
labs()

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g)) +
4   geom_point(aes(color = species,
5               shape = species)) +
6   geom_smooth(method = "lm") +
7   labs(title = "Massa corporal e comprimento da nadadeira",
8        subtitle = "Dimensões para Adelie, Chinstrap e Gentoo",
9        x = "Comprimento da nadadeira (mm)",
10       y = "Massa corporal (g)",
11       color = "Espécie",
12       shape = "Espécie")
```

Adicionando rótulos (labels)

Massa corporal e comprimento da nadadeira

Dimensões para Adelie, Chinstrap e Gentoo



Cores para daltônicos

ggthemes

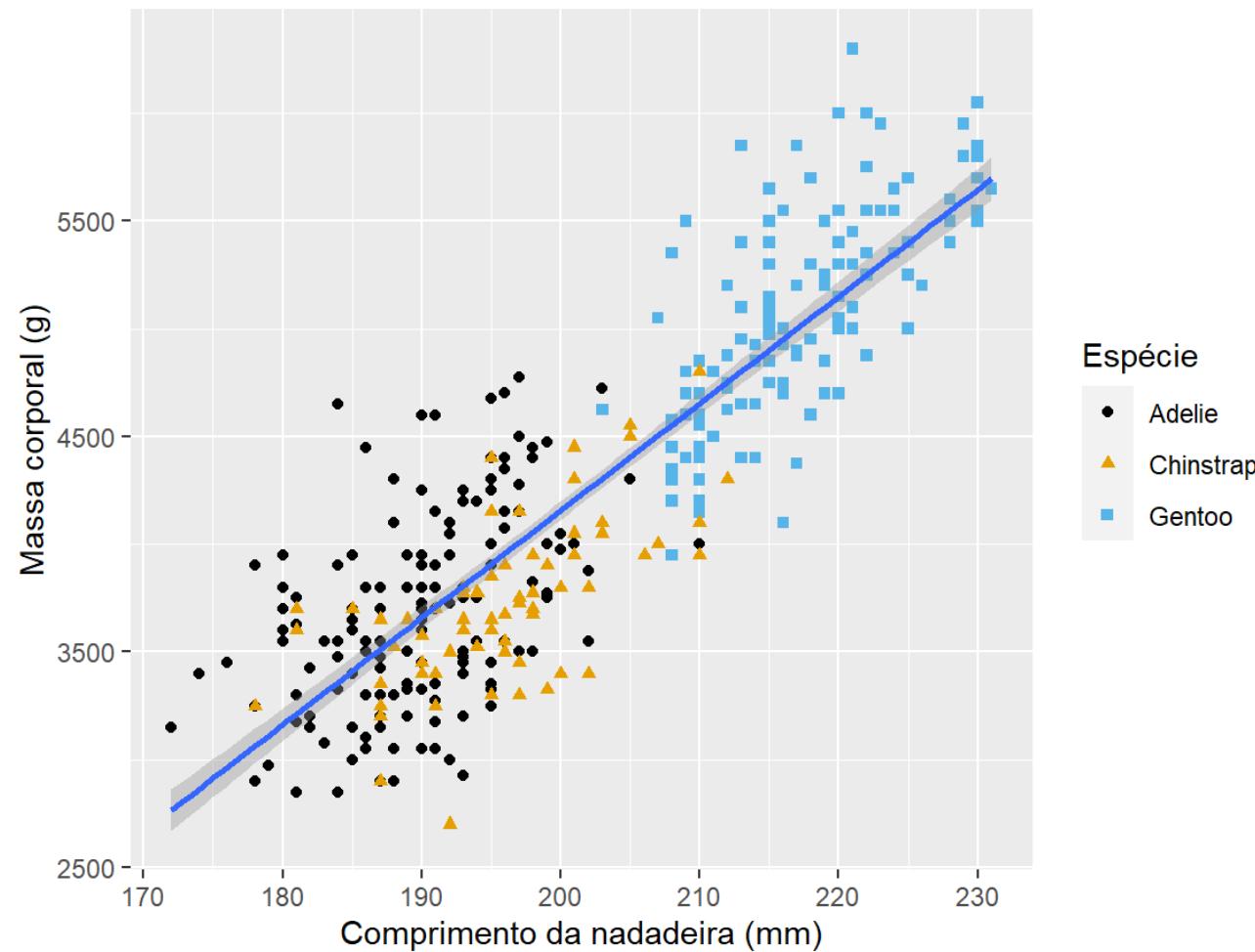
```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g)) +
4   geom_point(aes(color = species,
5               shape = species)) +
6   geom_smooth(method = "lm") +
7   labs(title = "Massa corporal e comprimento da nadadeira",
8       subtitle = "Dimensões para Adelie, Chinstrap e Gentoo",
9       x = "Comprimento da nadadeira (mm)",
10      y = "Massa corporal (g)",
11      color = "Espécie",
12      shape = "Espécie") +
13   scale_color_colorblind()
```

Cores para daltônicos

ggthemes

Massa corporal e comprimento da nadadeira

Dimensões para Adelie, Chinstrap e Gentoo



Exercícios

Qual o número de linhas no banco de dados dos [penguins](#)?
Quantas colunas?

Resposta

```
1 library(tidyverse)
2 library(palmerpenguins)
3
4 glimpse(penguins)
```

```
Rows: 344
Columns: 8
$ species           <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie,
Adel...
$ island            <fct> Torgersen, Torgersen, Torgersen, Torgersen,
Torgerser...
$ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1,
...
$ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1,
...
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190,
186...
$ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475,
...
$ sex               <fct> male, female, female, NA, female, male, female,
```

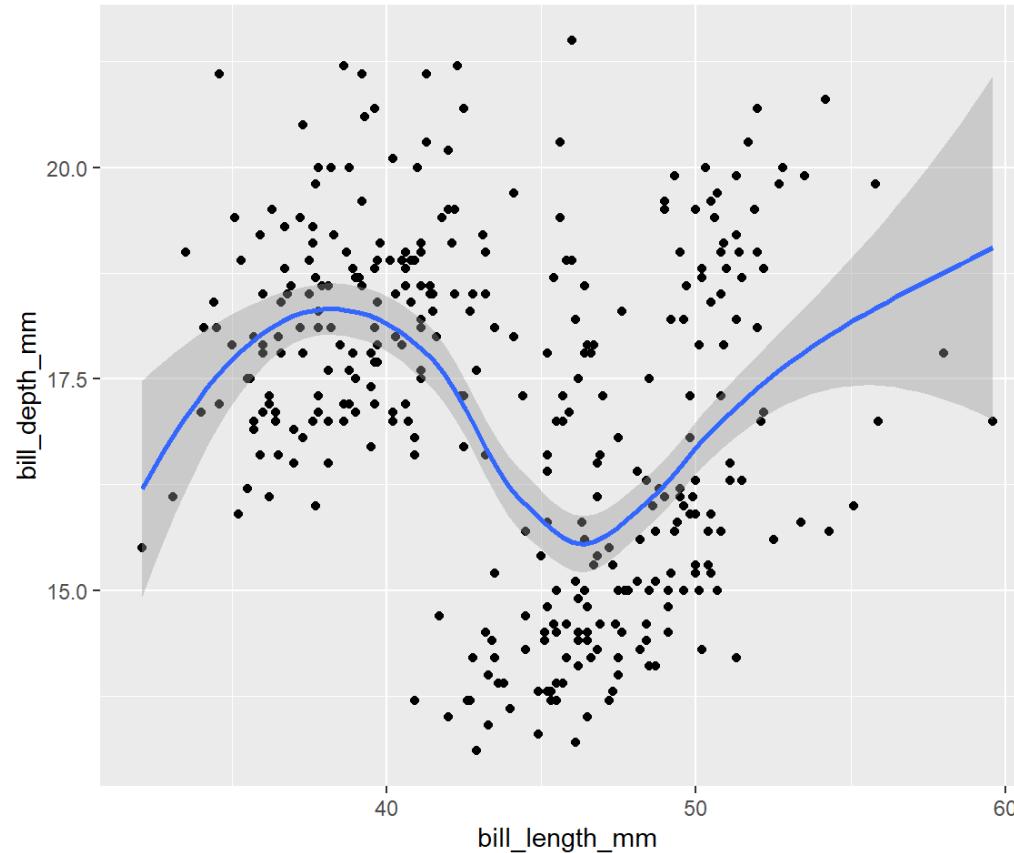
Exercícios

Faça um gráfico de dispersão de `bill_depth_mm` vs. `bill_length_mm`. Ou seja, faça um gráfico de dispersão com `bill_depth_mm` no eixo y e `bill_length_mm` no eixo x. Descreva a relação entre essas duas variáveis.

`bill` = bico

Resposta

```
1 ggplot(data = penguins,
2         aes(x = bill_length_mm,
3              y = bill_depth_mm)) +
4   geom_point() +
5   geom_smooth()
```

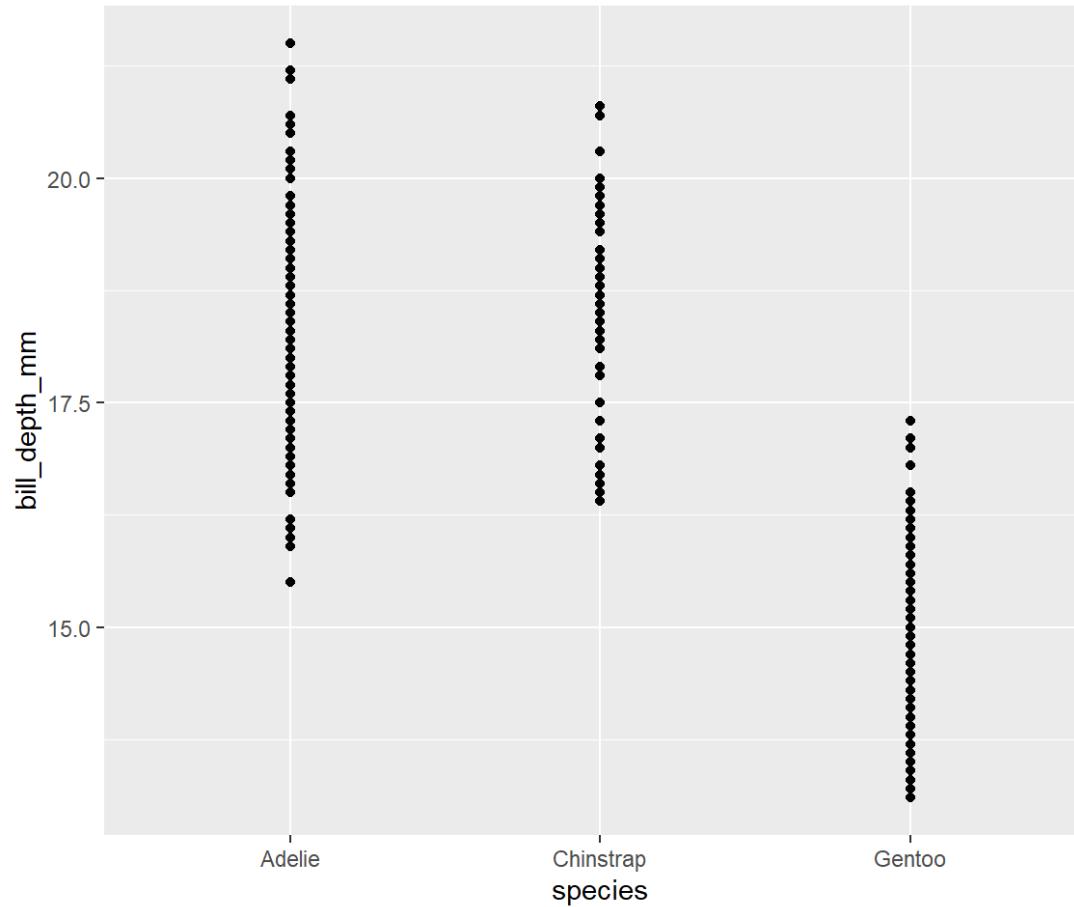


Exercícios

O que acontece se você fizer um gráfico de dispersão de `species` versus `bill_depth_mm`? O que poderia ser uma melhor escolha de geom?

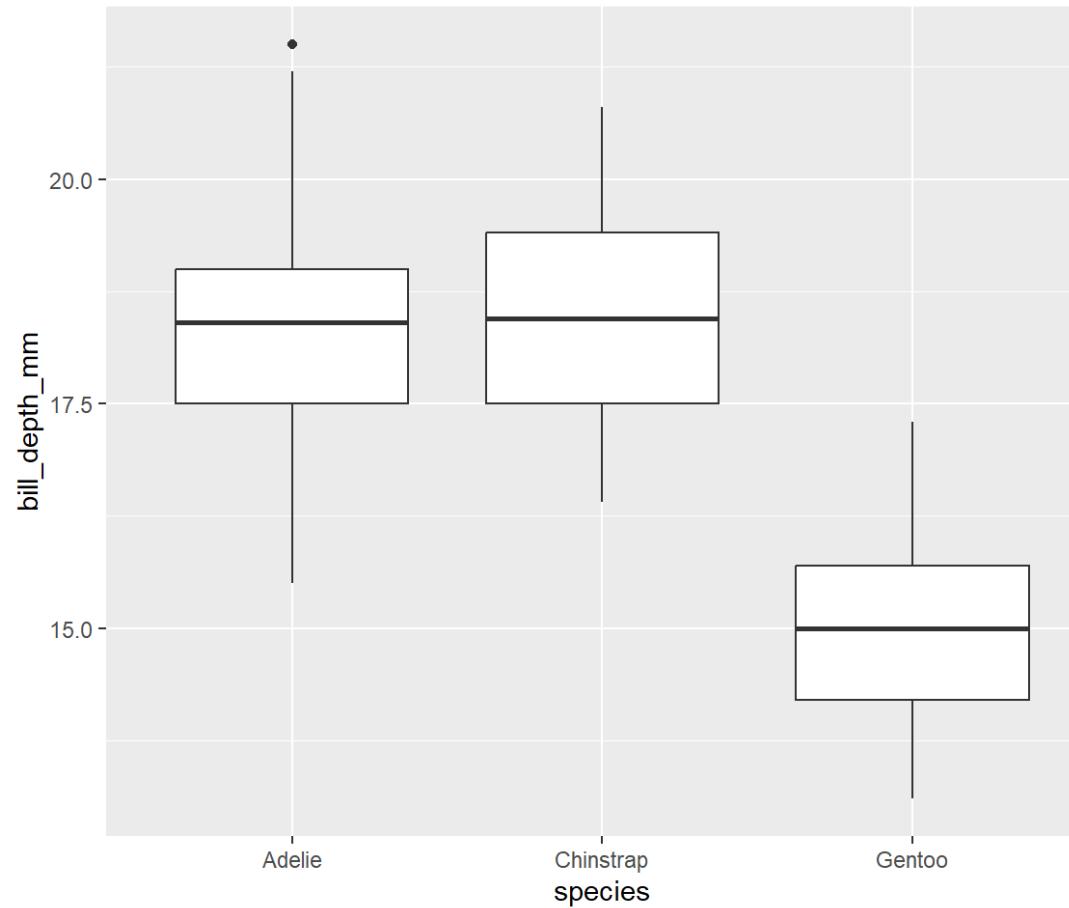
Resposta

```
1 ggplot(data = penguins,
2         aes(x = species,
3               y = bill_depth_mm)) +
4   geom_point()
```



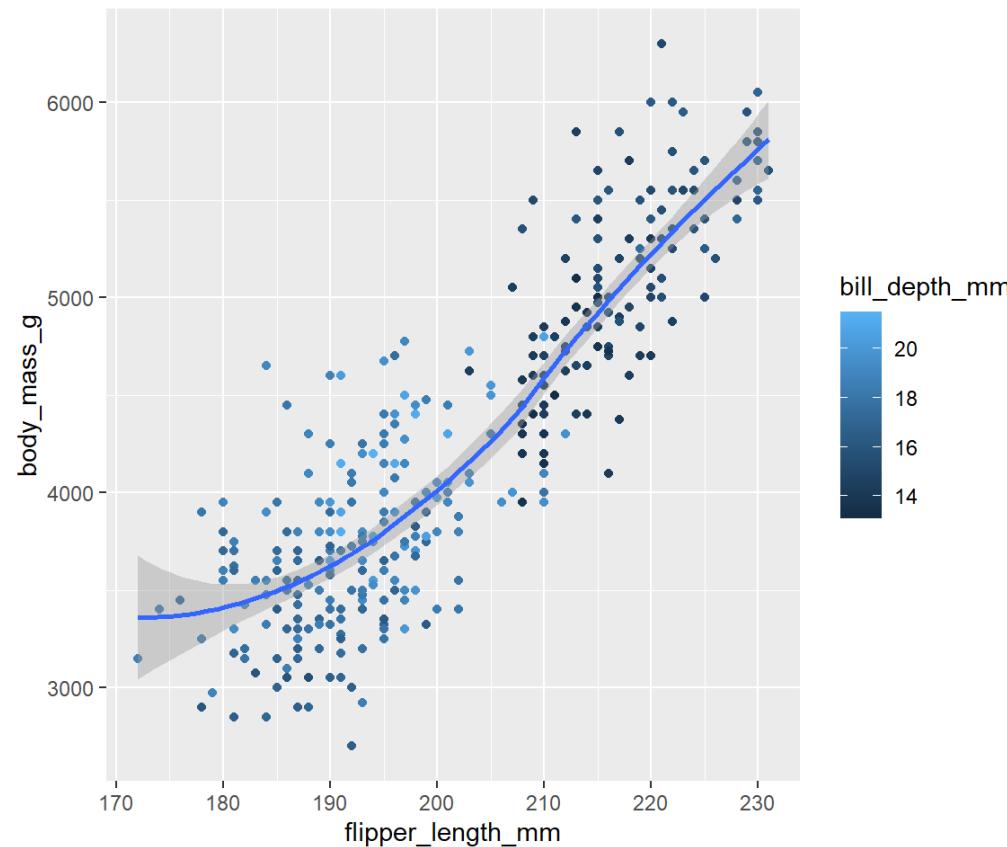
Resposta

```
1 ggplot(data = penguins,
2         aes(x = species,
3               y = bill_depth_mm)) +
4   geom_boxplot()
```



Exercícios

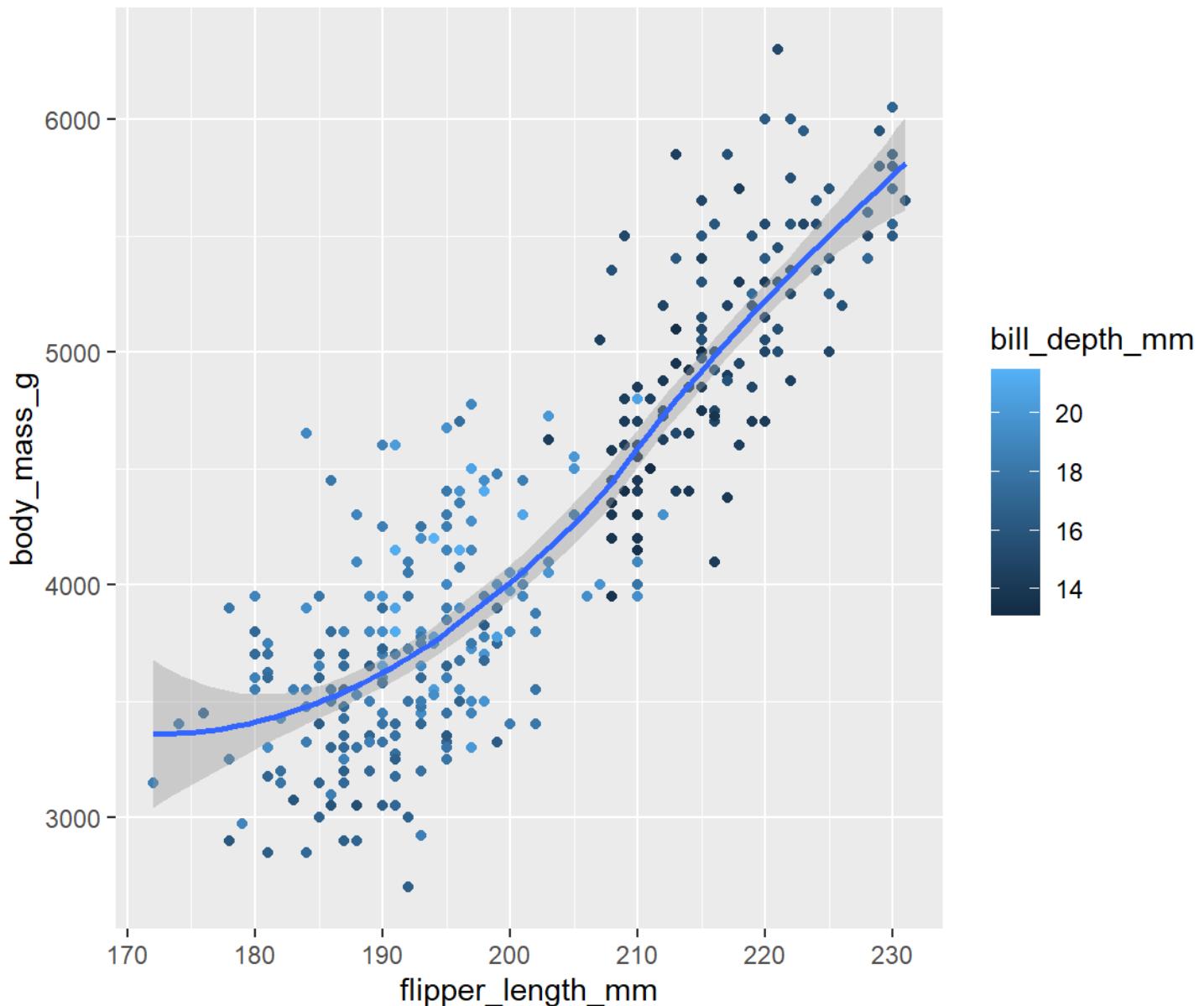
Recrie a seguinte visualização. Para qual estética o bill_depth_mm deve ser mapeado? E deve ser mapeado no nível global ou no nível geom?



Resposta

```
1 ggplot(data = penguins,
2         aes(x = flipper_length_mm,
3              y = body_mass_g)) +
4   geom_point(aes(color = bill_depth_mm)) +
5   geom_smooth()
```

Resposta



2.3 Chamadas do ggplot2

- `data`
- `mapping(aes(x = ... , y = ...))`

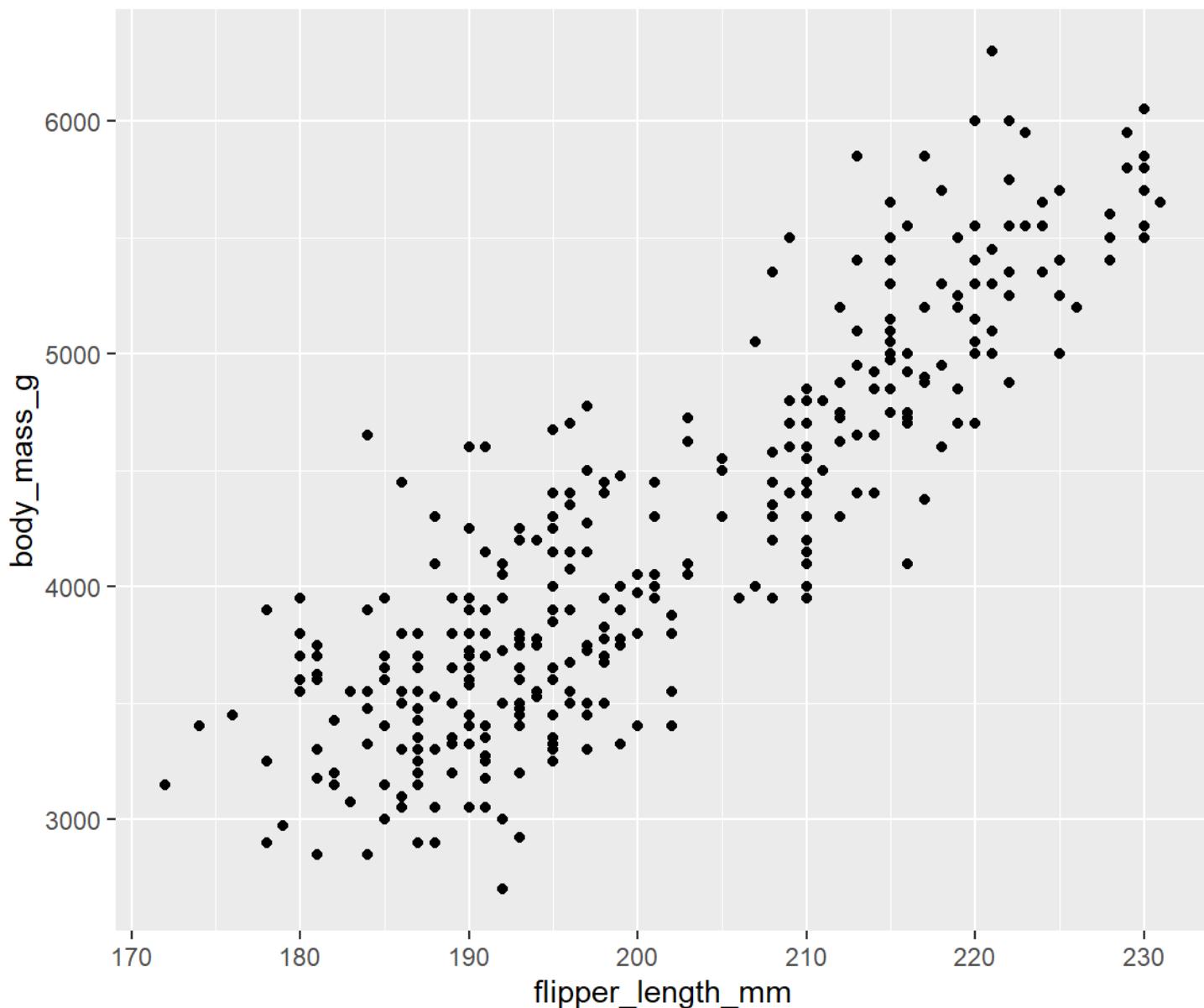
Código completo

Código com omissões

Código usando pipe | >

```
1 ggplot(data = penguins,
2   mapping = aes(x = flipper_length_mm,
3                 y = body_mass_g)) +
4   geom_point()
```

2.3 Chamadas do ggplot2



2.4 Visualizando distribuições

- A visualização vai depender do tipo de variável
 - Categórica
 - Numérica

Variável categórica

- Os números são usados para classificar objetos
- $1 \neq 2 \neq 3$
- Grupo: frutas
 - categorias
 - 1 = laranja
 - 2 = maçã
 - 3 = uva
 - laranja \neq maçã \neq uva

Variável categórica: Gráfico de Barras (geom_bar())

Altura das barras: número de observações em cada valor de x

```
1 ggplot(penguins,  
2       aes(species)) +  
3     geom_bar()
```

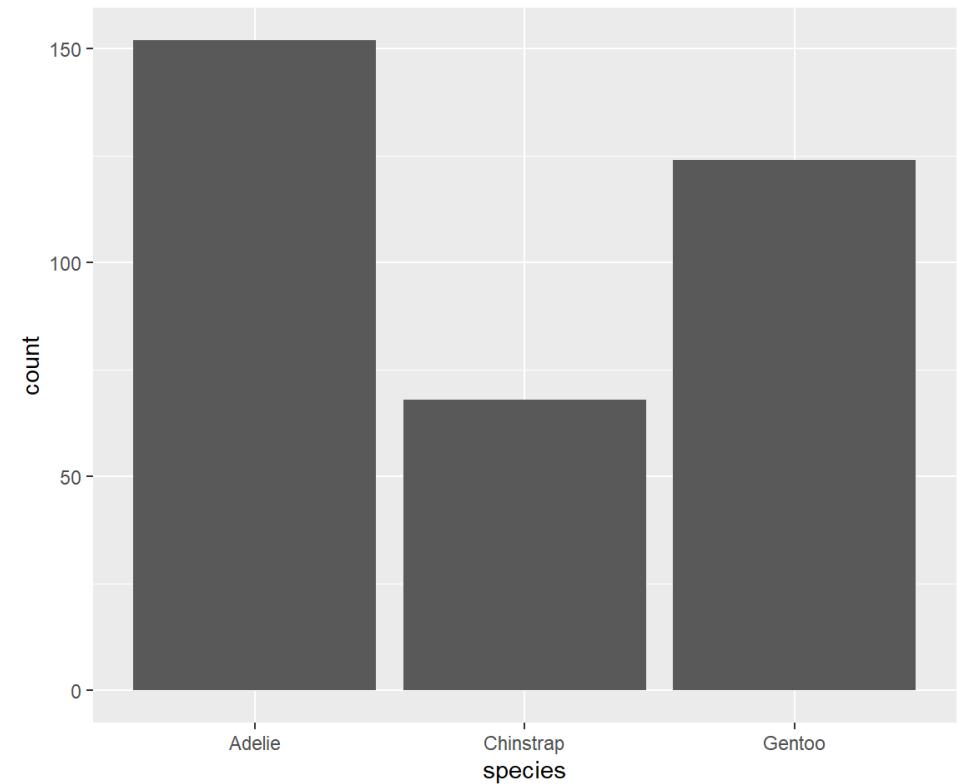


Gráfico de Barras (`geom_bar()`): variáveis ordendas pela frequência

```
1 ggplot(penguins,  
2       aes(fct_infreq(species))) +  
3   geom_bar()
```

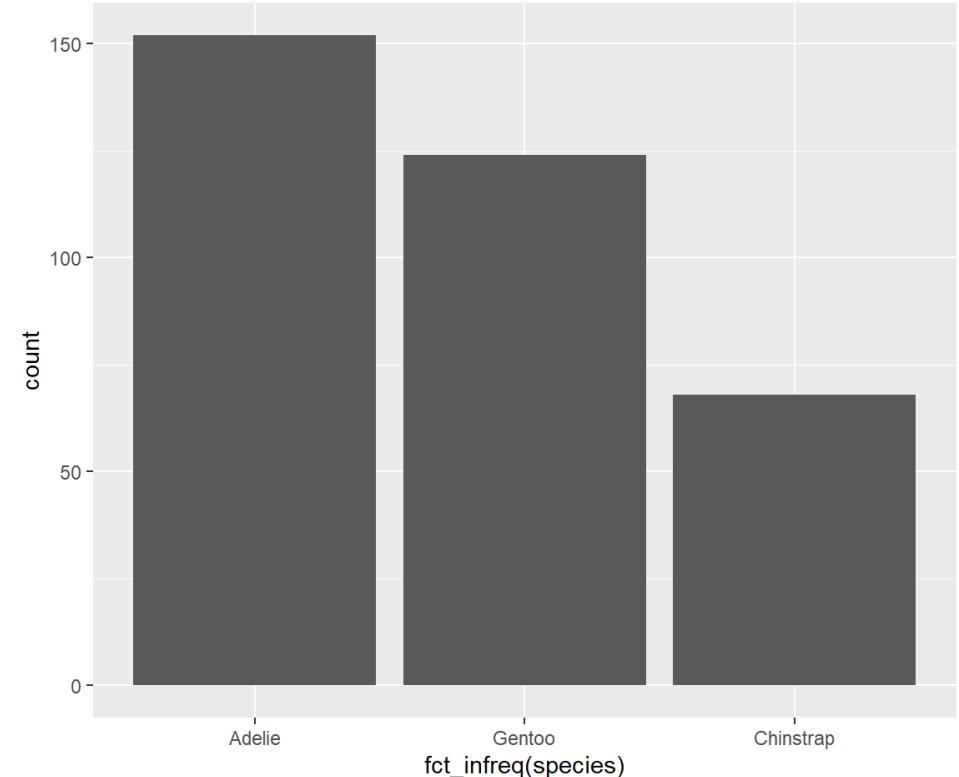


Gráfico de Barras (`geom_bar()`): proporções

Código para percentuais

Gráfico com percentuais

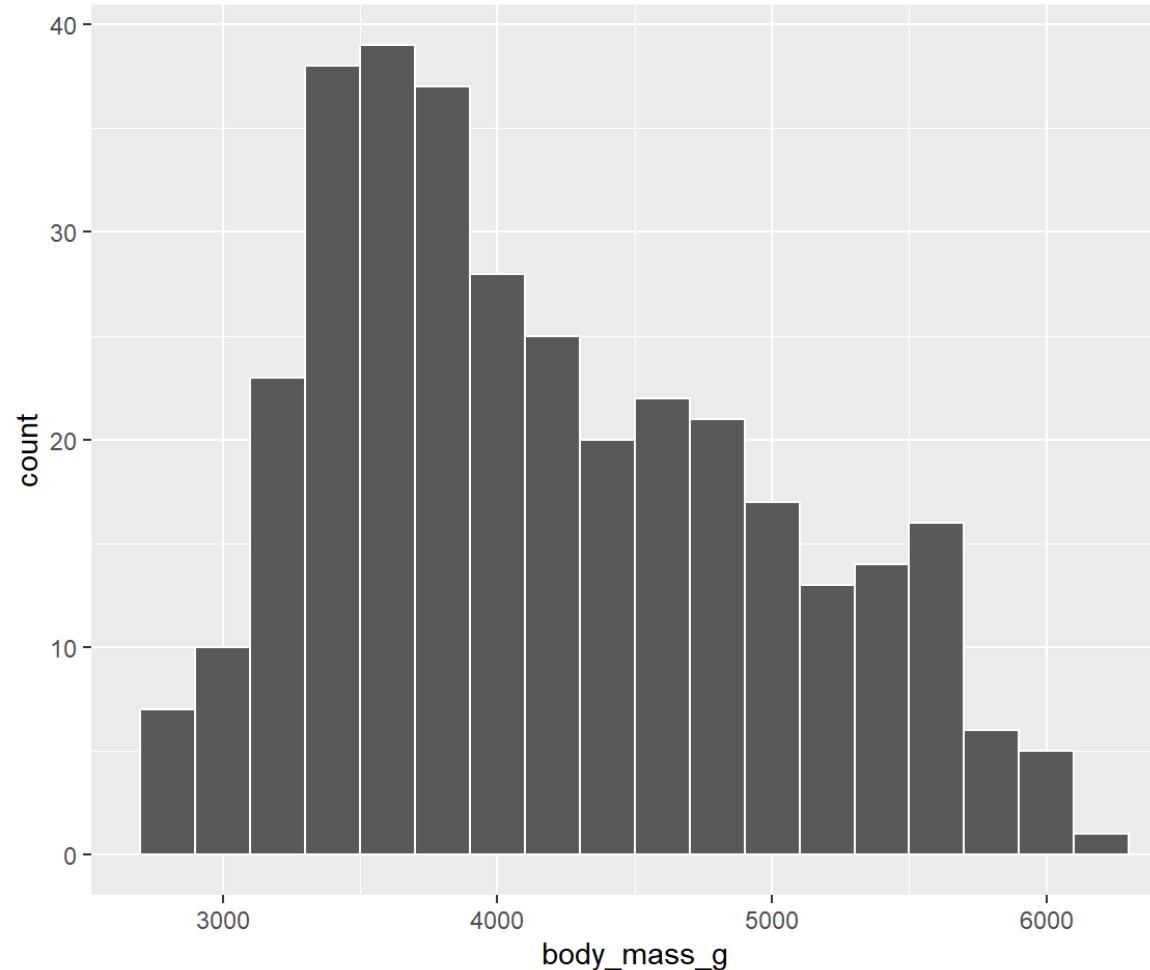
```
1 ggplot(penguins,
2     aes(fct_infreq(species),
3         y = after_stat(count/sum(count)))) +
4     geom_bar() +
5     geom_text(aes(label = percent(after_stat(count/sum(count)))),
6               stat = "count",
7               nudge_y = 0.03) +
8     scale_y_continuous(labels = percent)
```

Variável numérica

- $1 > 2 > 3$
 - $2 - 1 = 1$
 - $3 - 2 = 1$
 - $3 - 1 = 2$
- Discreta
 - pessoas: 1, 2, 3, ou 4
 - $1/2$ (0,5) pessoa?
- Contínua
 - comprimento: 1,23 cm, 1,234 cm, 1,2347 cm ...
 - valores infinitos entre 2 números

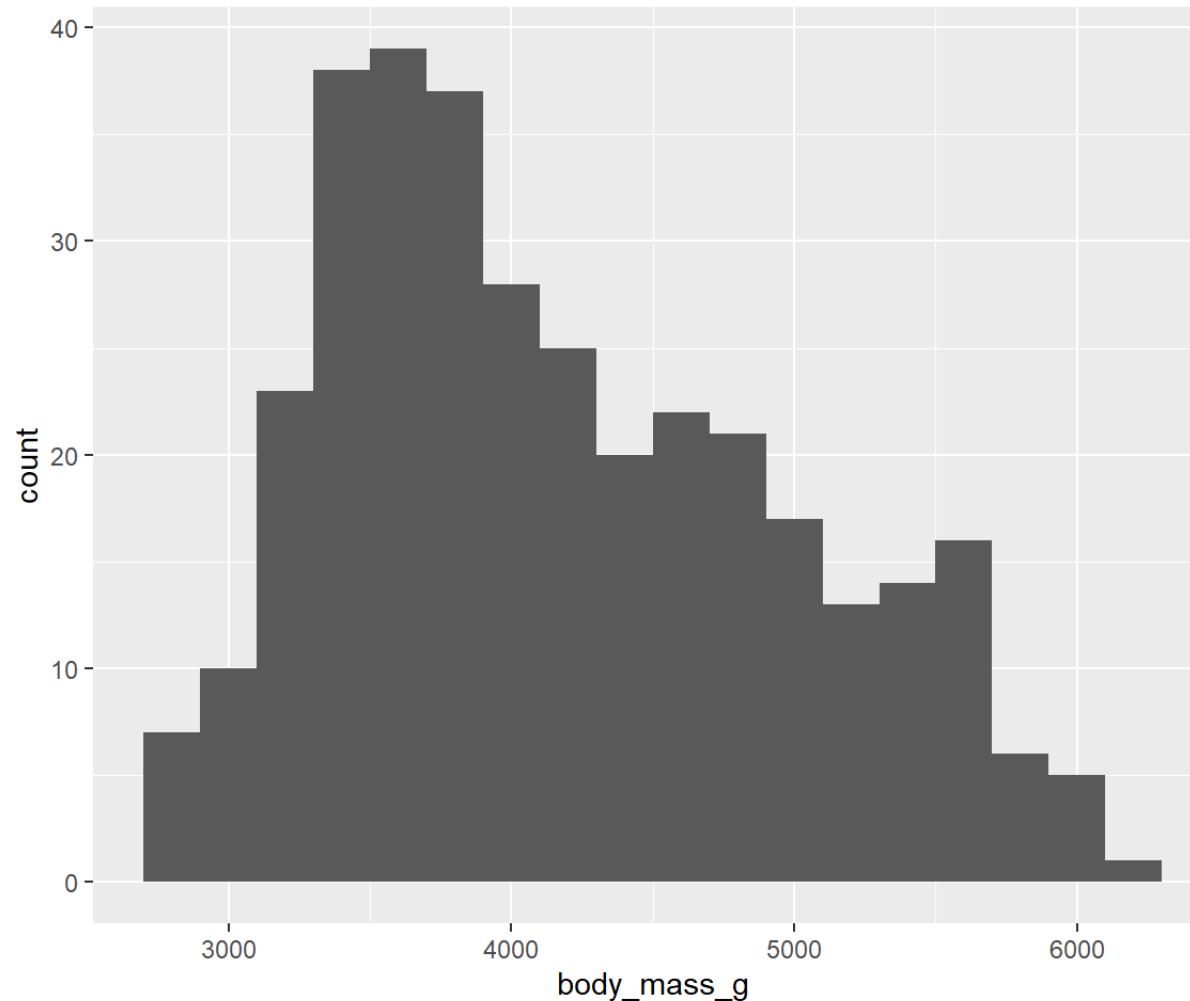
Histograma (`geom_histogram()`)

- Divide o eixo x em caixas igualmente espaçadas
- Usa a altura de uma barra para exibir o número de observações em cada caixa



Histograma (geom_histogram())

```
1 ggplot(penguins, aes(x = body_mass_g)) +  
2   geom_histogram(binwidth = 200)
```

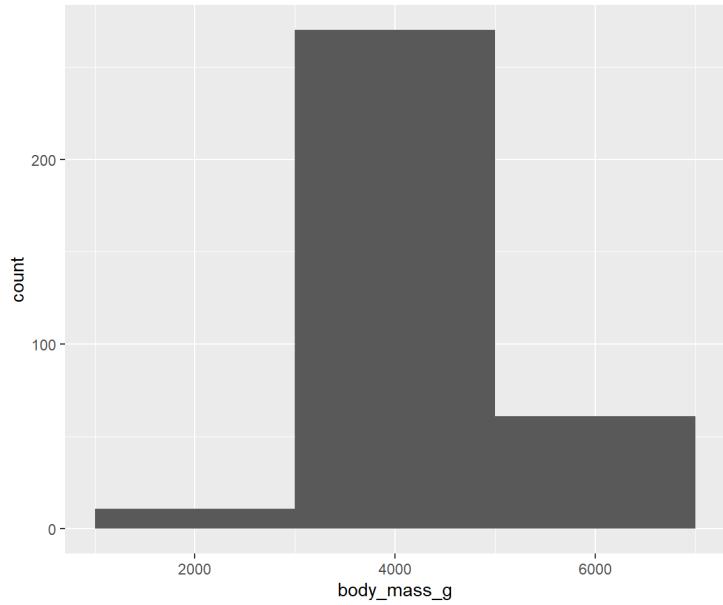


Histograma (`geom_histogram()`): `binwidth()`

- Define a largura dos intervalos

`binwidth(20)`

`binwidth(2000)`

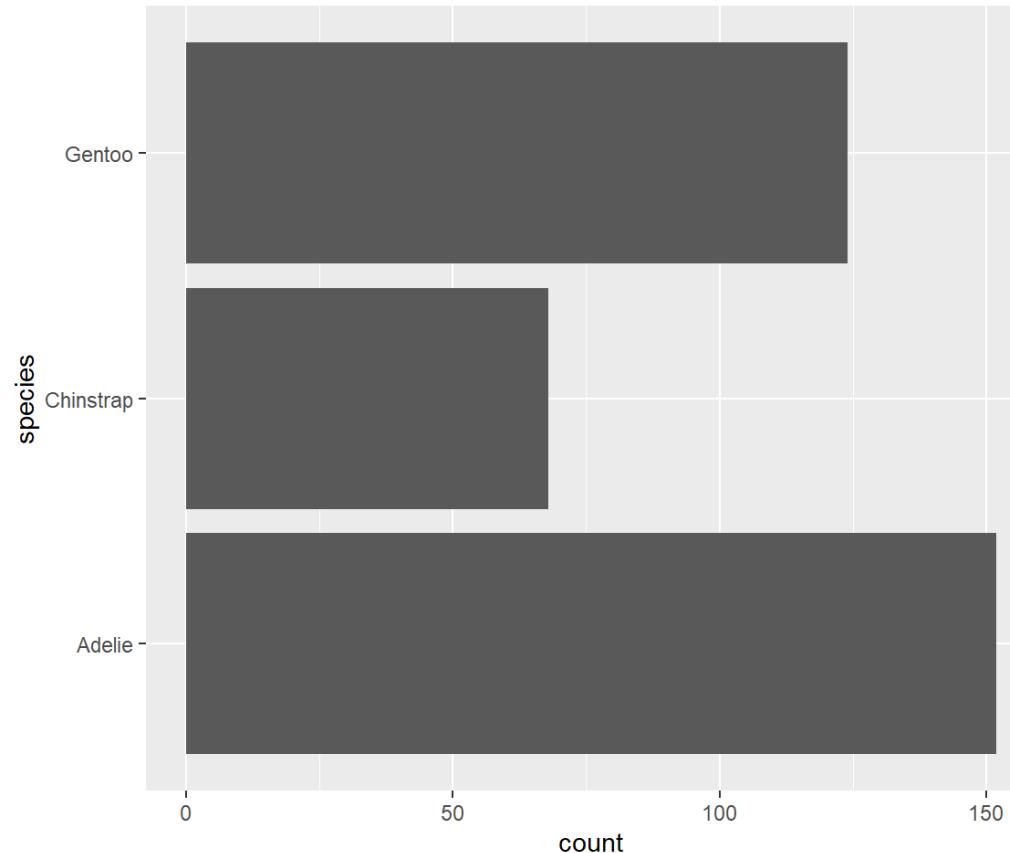


Exercícios

Faça um gráfico de barras das espécies (*species*) de pinguins, onde você atribui espécies (*species*) ao eixo y. Como esse gráfico é diferente?

Resposta

```
1 ggplot(penguins,  
2       aes(y = species)) +  
3     geom_bar()
```



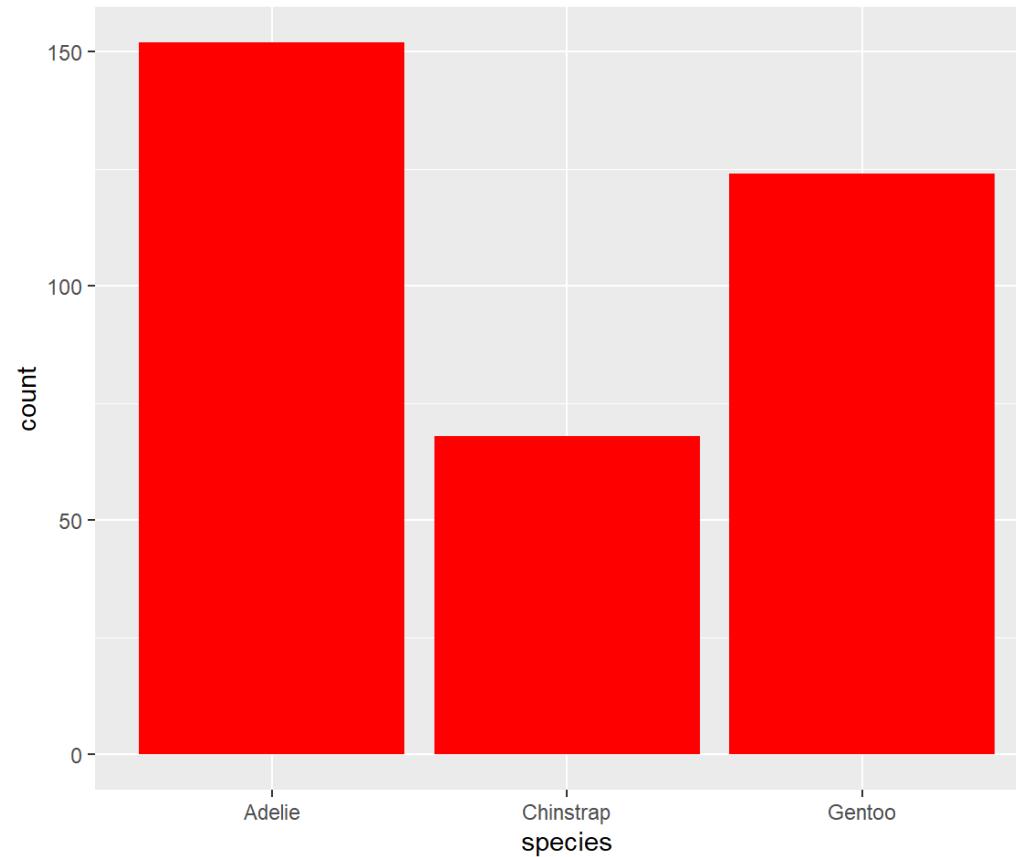
Exercícios

Em que os dois gráficos a seguir são diferentes? Qual estética, cor ou preenchimento, é mais útil para mudar a cor das barras?

```
1 ggplot(penguins, aes(species)) +  
2   geom_bar(color = "red")  
3  
4 ggplot(penguins, aes(species)) +  
5   geom_bar(fill = "red")
```

Resposta

```
1 ggplot(penguins, aes(species)) +  
2   geom_bar(fill = "red")
```



Exercícios

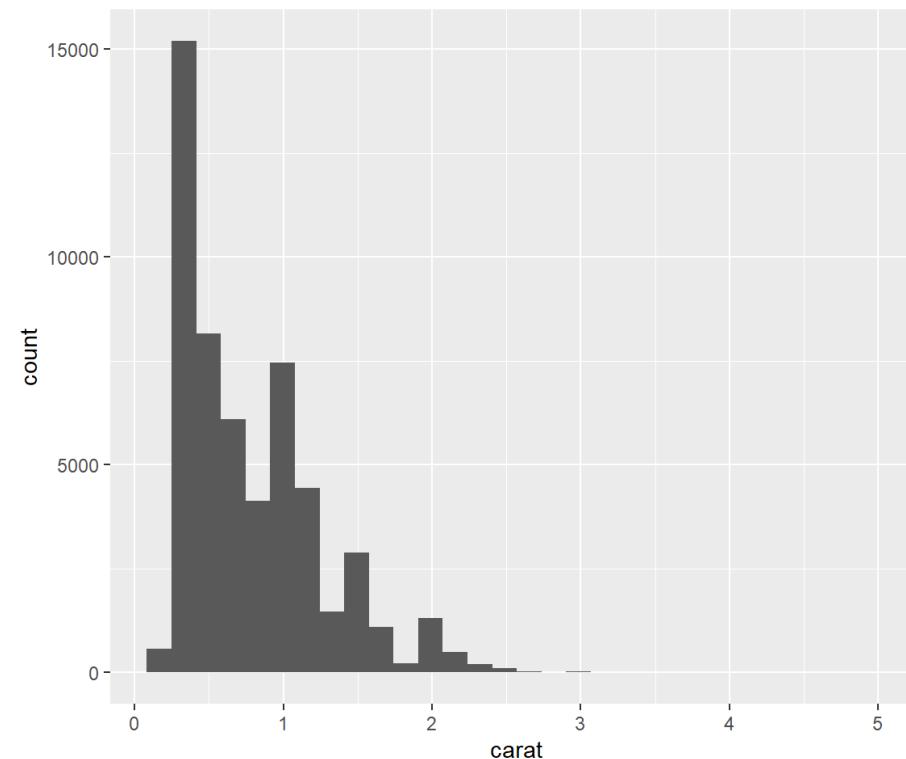
Faça um histograma da variável quilate (`carat`) no conjunto de dados de diamantes que está disponível quando você carrega o pacote `tidyverse`. Experimente diferentes larguras de caixa (`bin`). Qual largura de `bin` revela os padrões mais interessantes?

Resposta

bin padrão (30)

histograma base R

```
1 ggplot(diamonds,  
2         aes(carat)) +  
3     geom_histogram()
```



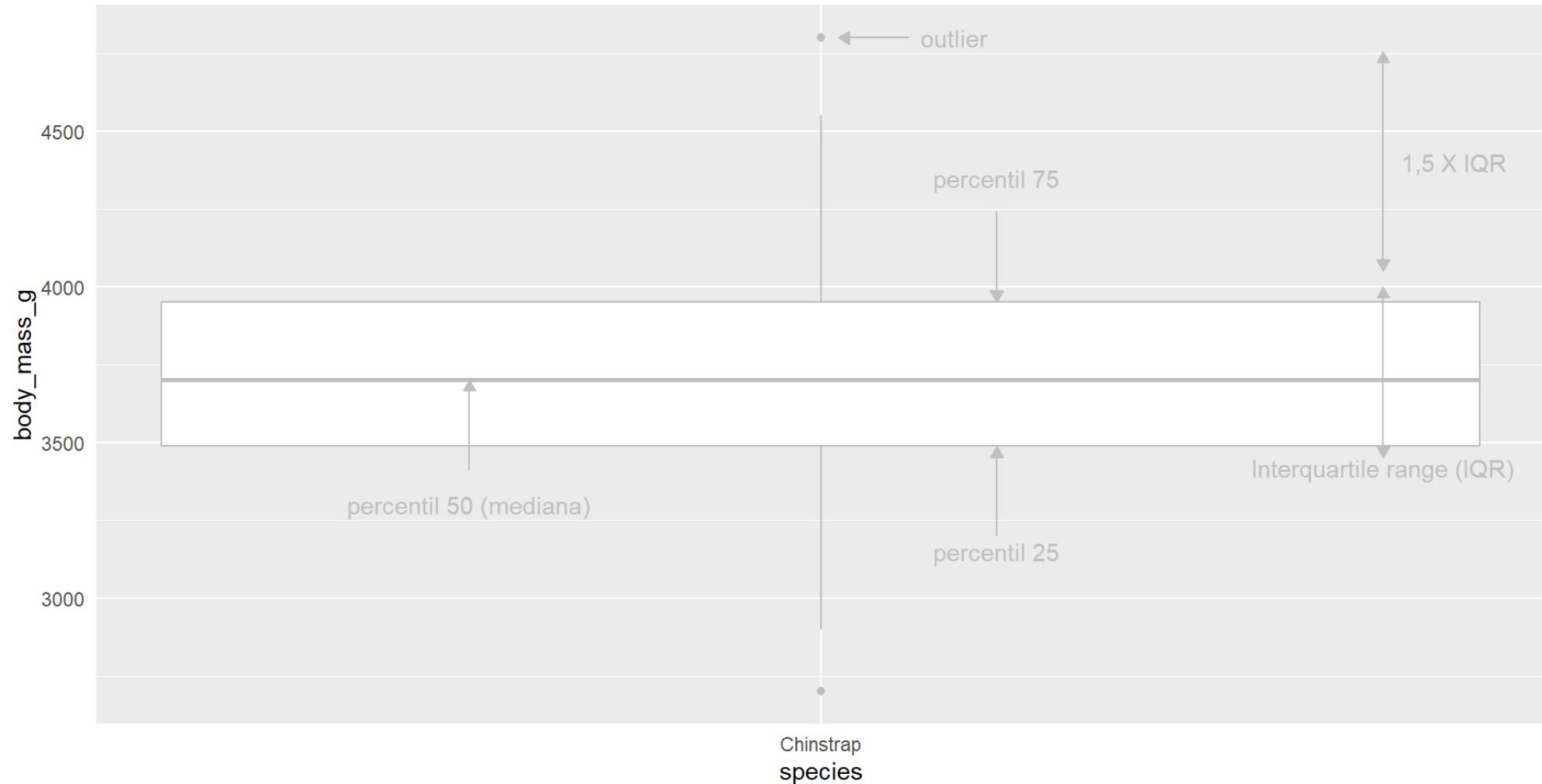
2.5 Visualizando relacionamentos

- 2 variáveis
 - categórica x numérica
 - categórica x categórica
 - numérica x numérica
- 3 ou mais variáveis

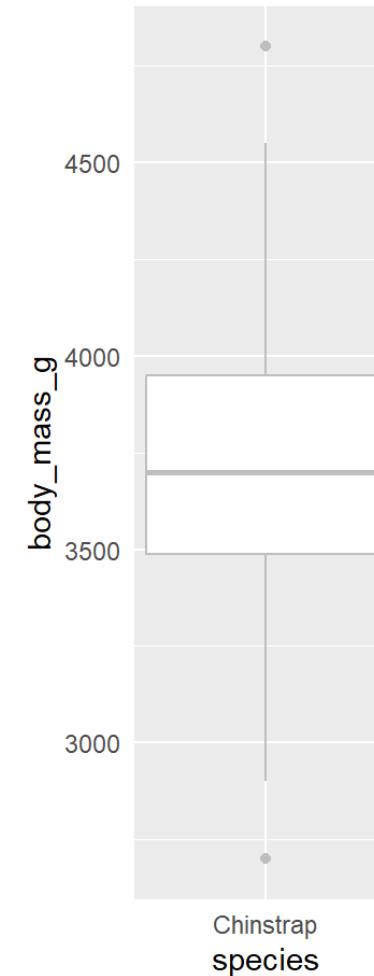
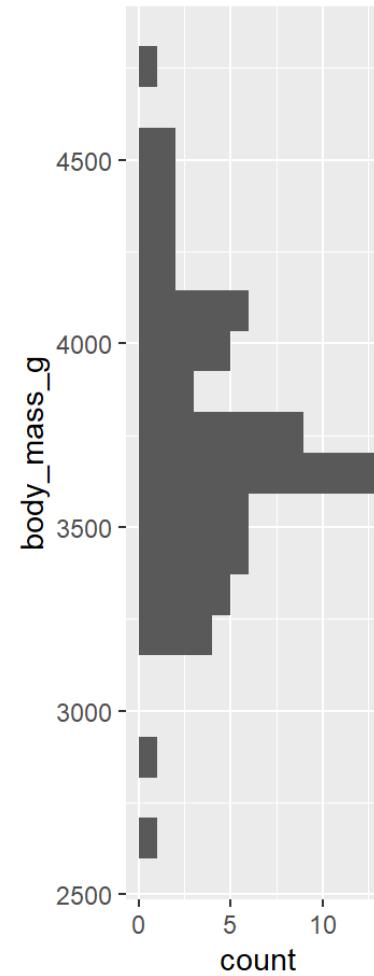
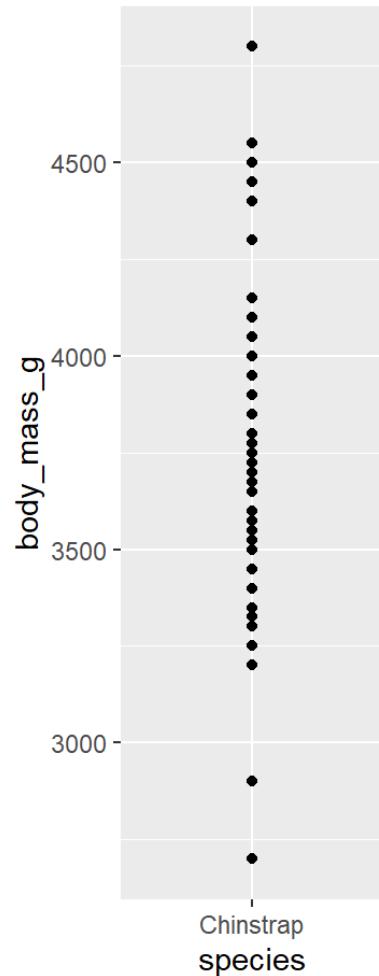
Variáveis numéricas e categóricas

- Diagrama de caixa (`geom_boxplot()`)
 - Atalho visual para medidas de posição (percentis) que descrevem uma distribuição
 - Útil para identificar possíveis outliers

Entendendo um diagrama de caixa (`geom_boxplot()`)



Entendendo um diagrama de caixa (`geom_boxplot()`)



Duas variáveis categóricas

- Relacionamento entre ilha (*island*) e espécie (*species*)
 - Distribuição das espécies dentro de cada ilha

Tabela de Frequência

```
1 penguins |>
2   group_by(island, species) |>
3   summarise(n = n()) |>
4   pivot_wider(names_from = species,
5                 values_from = n) |>
6   rowwise() |>
7   mutate(total_island = sum(c_across(Adelie:Chinstrap), na.rm = T)) |>
8   ungroup() |>
9   kable() |>
10  kable_minimal()
```

island	Adelie	Gentoo	Chinstrap	total_island
Biscoe	44	124	NA	168
Dream	56	NA	68	124
Torgersen	52	NA	NA	52

Gráfico de barras empilhados (stacked bar plots)

```
1 ggplot(penguins,  
2         aes(island,  
3                  fill = species)) +  
4     geom_bar()
```

Gráfico de barras empilhados (stacked bar plots)

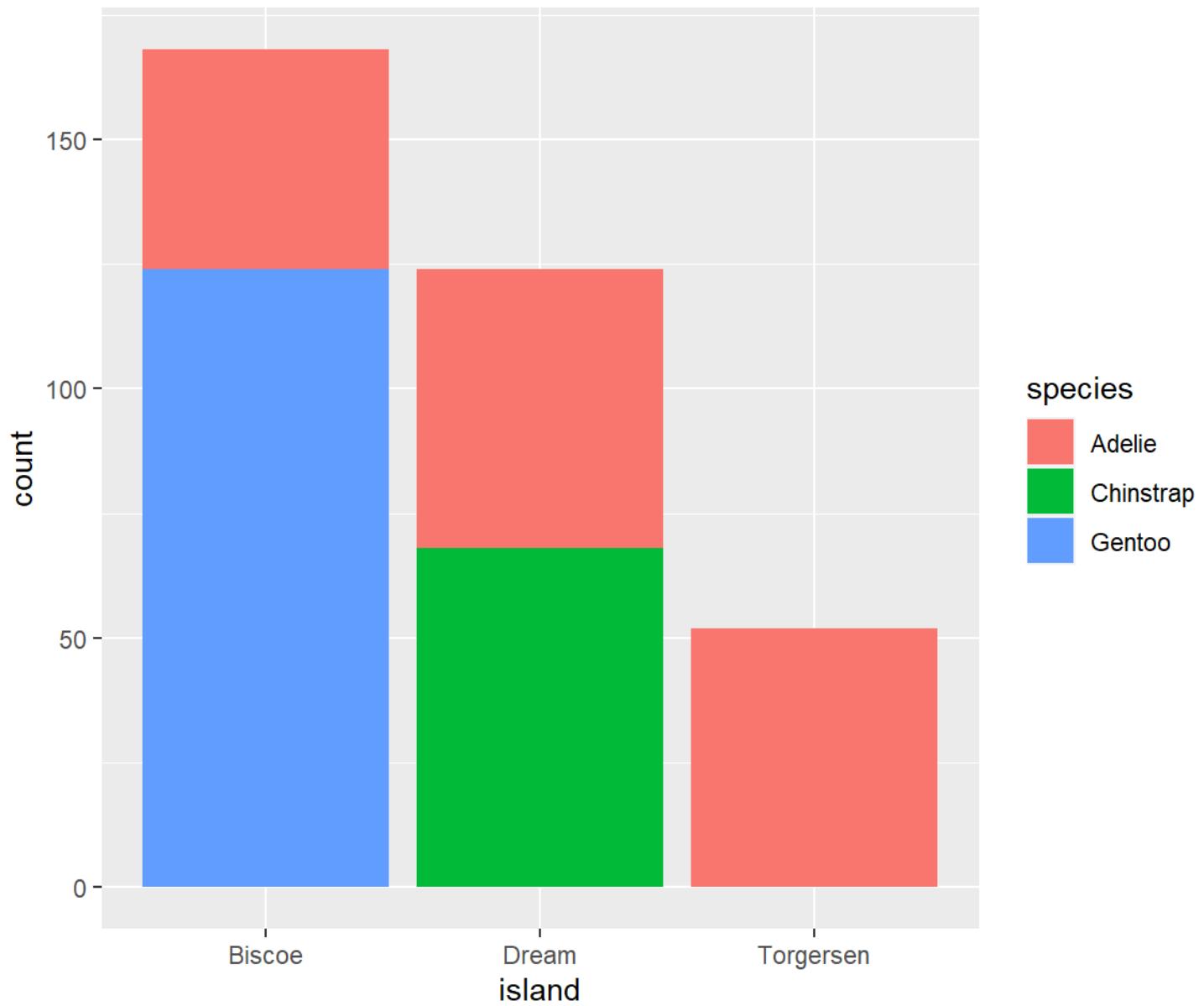


Gráfico de frequência relativa (relative frequency plot)

- `position = "fill"`
- útil para comparar a distribuição de espécies nas ilhas
 - não é afetado pelo número desigual de pinguins nas ilhas

```
1 ggplot(penguins,
2       aes(island,
3             fill = species)) +
4   geom_bar(position = "fill")
```

Gráfico de frequência relativa (relative frequency plot)

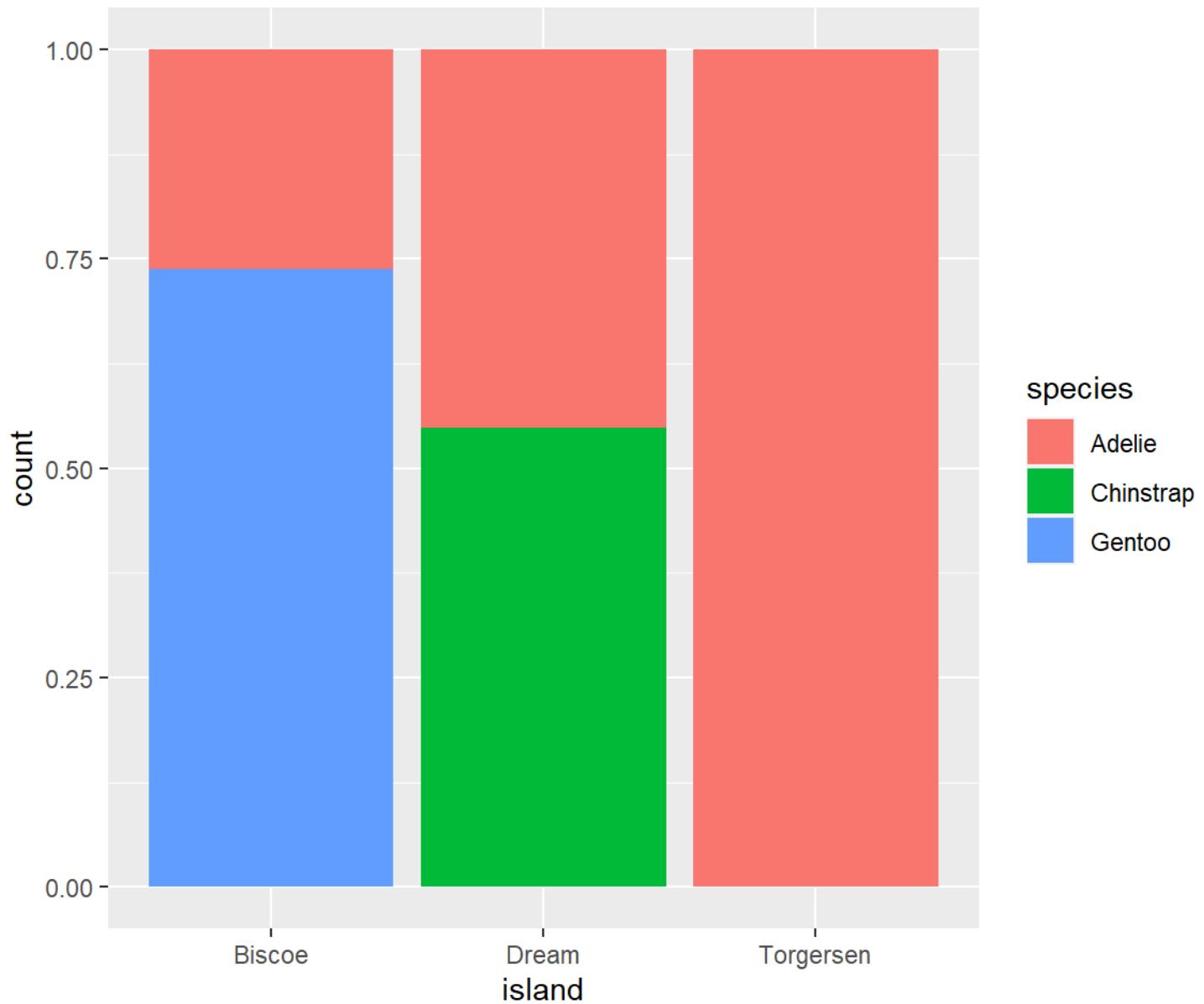
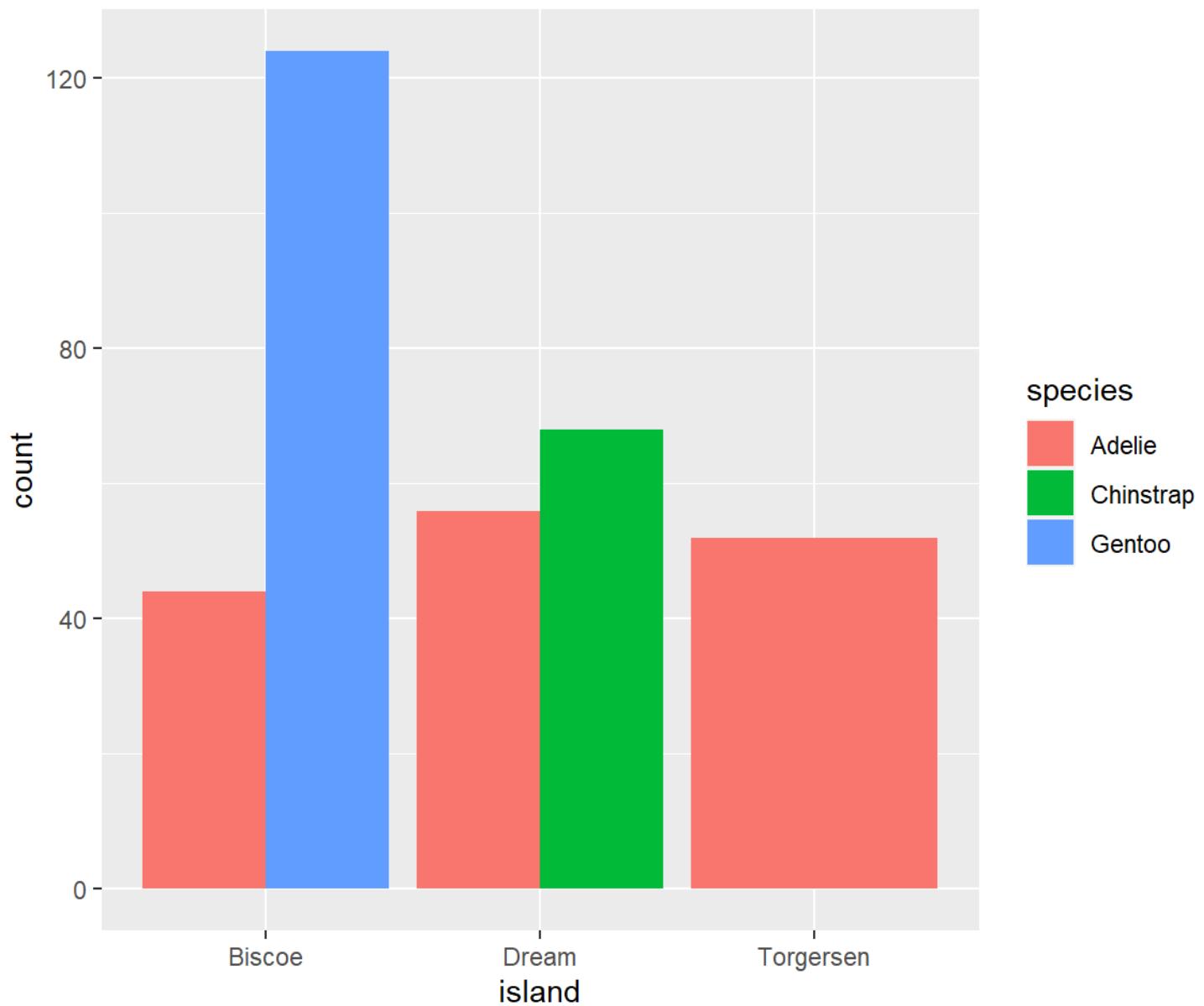


Grafico de barras

- `position = "dodge"`

```
1 ggplot(penguins,  
2       aes(island,  
3             fill = species)) +  
4   geom_bar(position = "dodge")
```

Grafico de barras

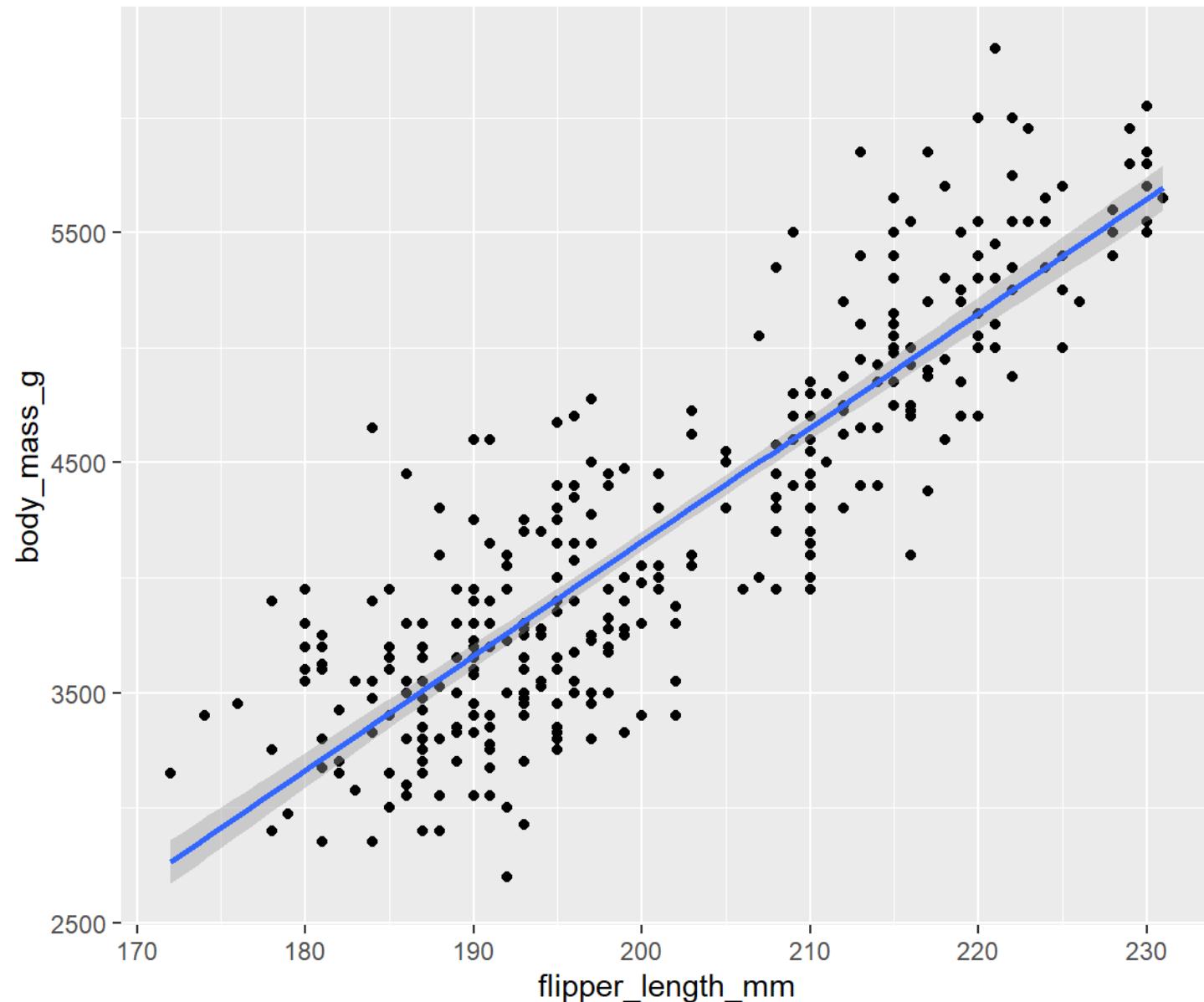


Duas variáveis numéricas

- `geom_point()`
- `geom_smooth()`

```
1 ggplot(penguins,
2     aes(flipper_length_mm,
3         body_mass_g)) +
4     geom_point() +
5     geom_smooth(method = "lm")
```

Duas variáveis numéricas

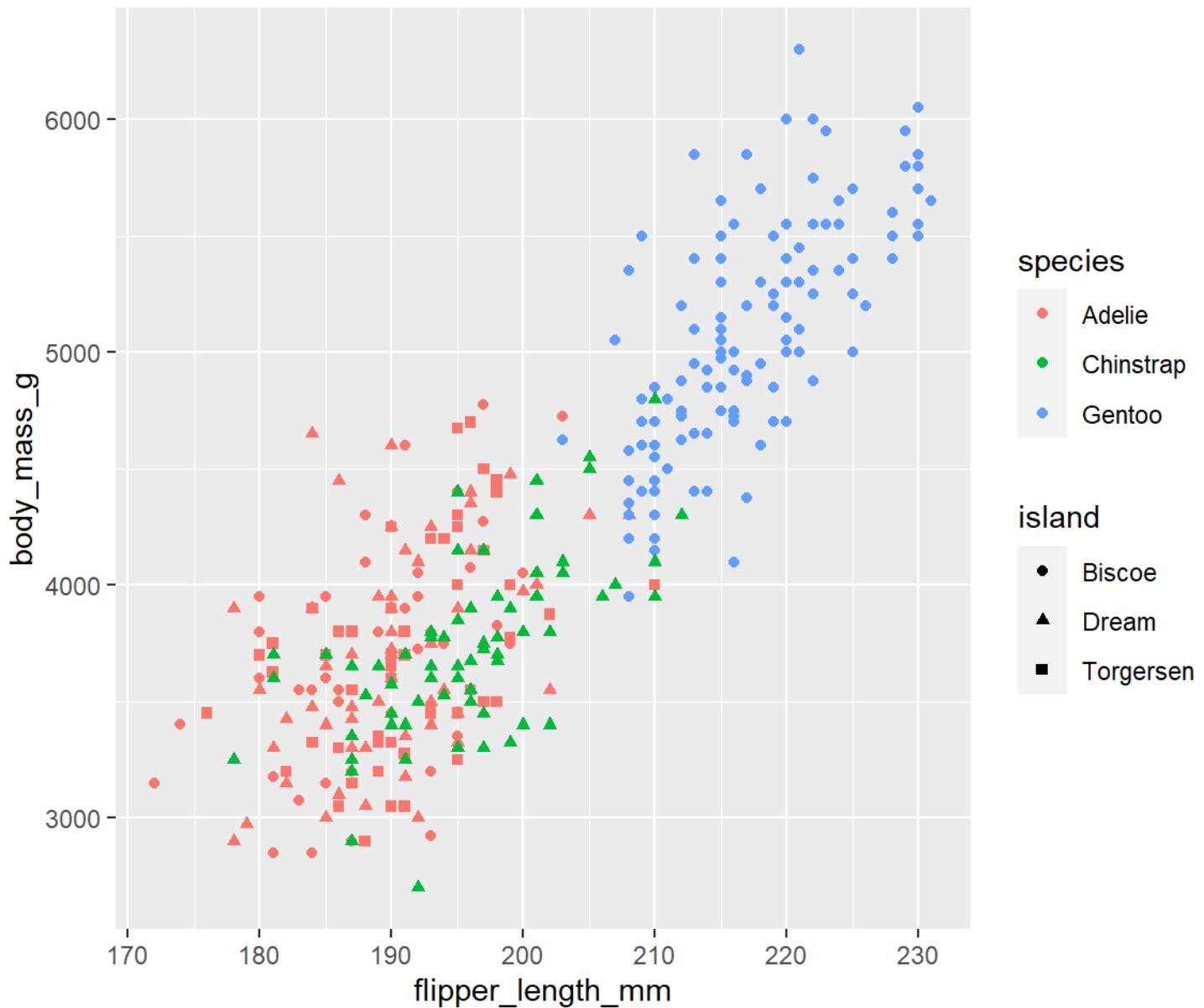


Três ou mais variáveis

- Podem ser mapeadas para diferentes estéticas (aesthetics)
- `color` = espécies (`species`)
- `shape` = ilhas (`island`)

```
1 ggplot(penguins,
2       aes(flipper_length_mm,
3             body_mass_g)) +
4   geom_point(aes(color = species,
5               shape = island))
```

Três ou mais variáveis

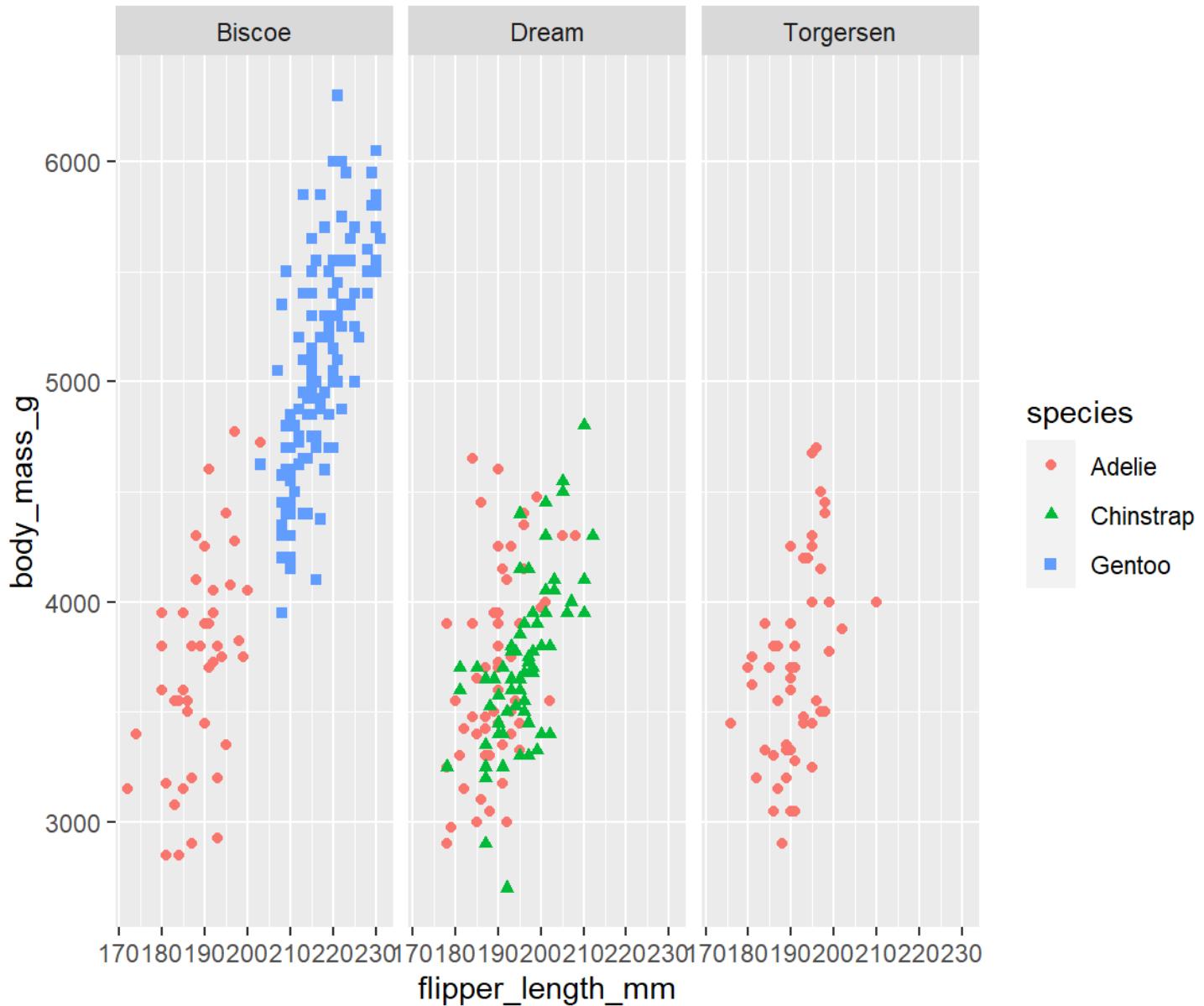


facets

- Muitas variáveis nas estéticas pode deixar o gráfico confuso
- **facet_**

```
1 ggplot(penguins,
2     aes(flipper_length_mm,
3         body_mass_g)) +
4     geom_point(aes(color = species,
5                 shape = species)) +
6     facet_wrap(~ island)
```

facets



species

- Adelie
- Chinstrap
- Gentoo

Exercícios

O banco de dados `mpg` que acompanha o pacote `ggplot2` contém 234 observações coletadas pela Agência de Proteção Ambiental dos EUA em 38 modelos de carros. Quais variáveis em `mpg` são categóricas? Quais variáveis são numéricas? (Dica: digite `?mpg` para ler a documentação do banco de dados.) Como você pode ver essas informações ao executar `mpg`?

Reposta

```
1 glimpse(mpg)
```

```
Rows: 234
Columns: 11
$ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi",
"..."
$ model          <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro",
"..."
$ displ          <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0,
2...
$ year           <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999,
200...
$ cyl            <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8,
...
$ trans          <chr> "auto(15)", "manual(m5)", "manual(m6)", "auto(av)",
"auto...
$ drv             <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4",
"4"
```

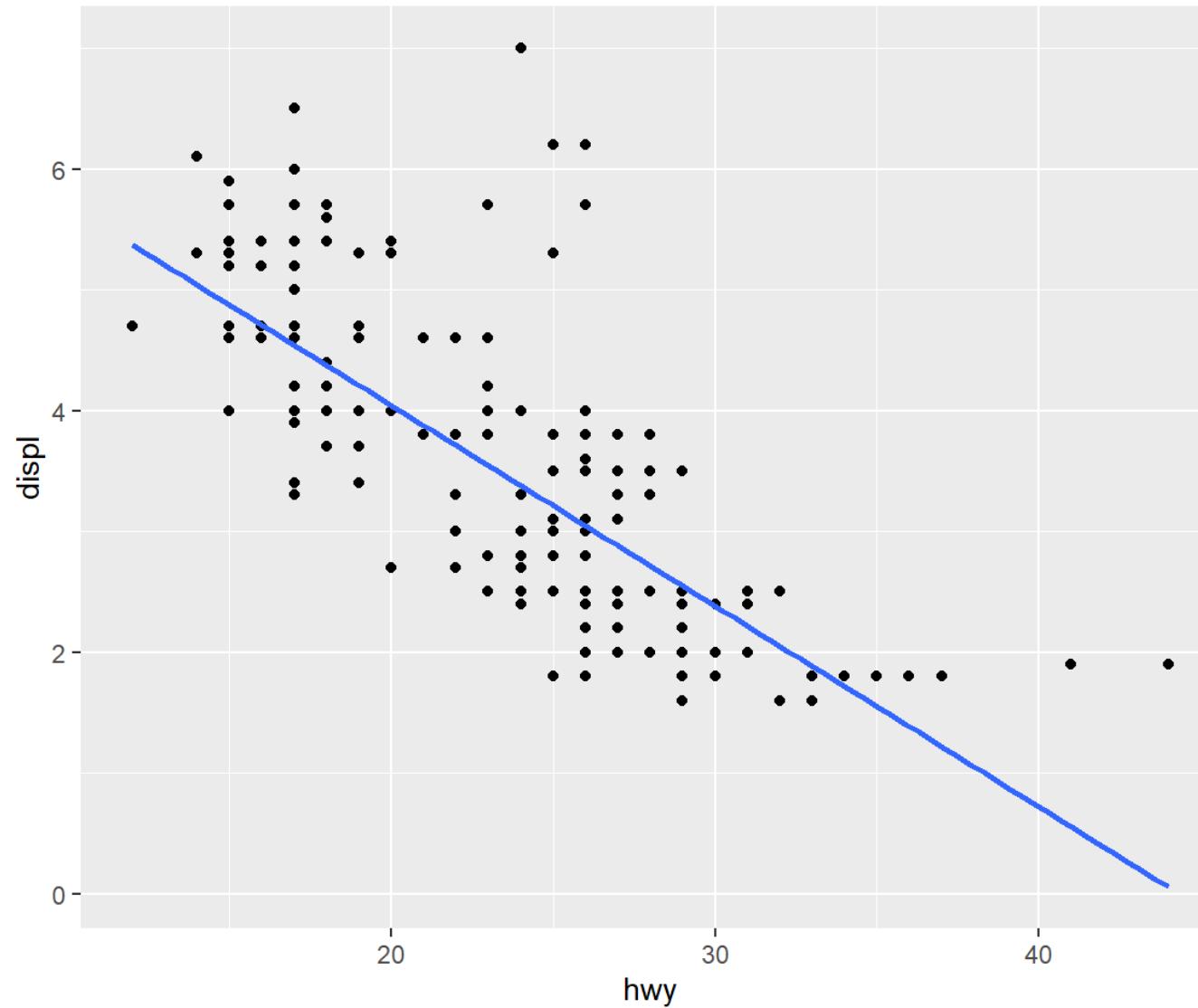
Exercícios

Faça um gráfico de dispersão de hwy vs. displ usando o banco de dados mpg. Em seguida, mapeie uma terceira variável numérica para `color`, depois para `size`, então para `color` e `size` ao mesmo tempo, então para `shape`. Como essas estéticas se comportam de maneira diferente para variáveis categóricas e numéricas?

Resposta

```
1 p <-
2   ggplot(mpg,
3     aes(hwy,
4       displ)) +
5   geom_point() +
6   geom_smooth(method = "lm",
7     se = F)
8
9 p
```

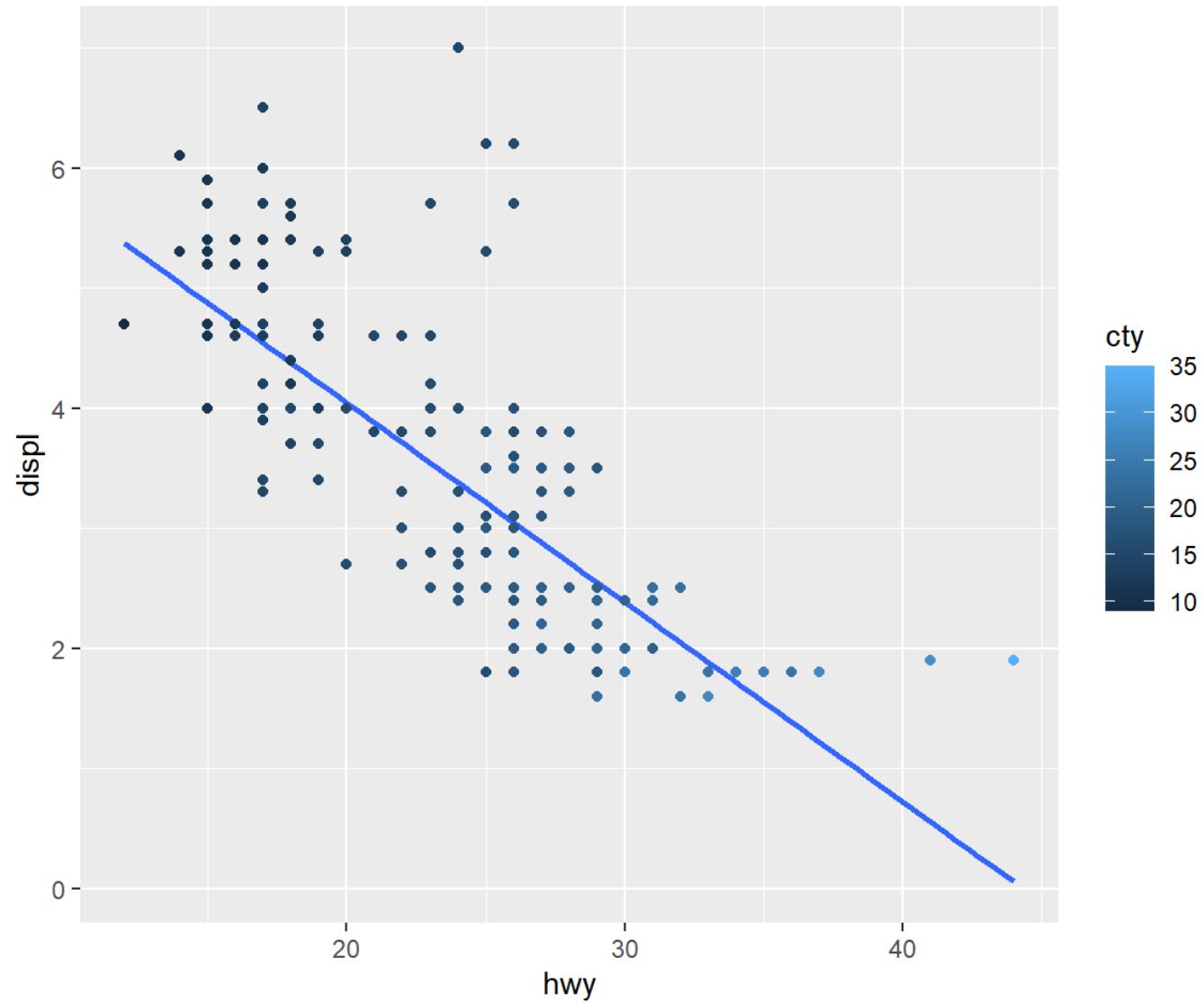
Resposta



Resposta

```
1 p + geom_point(aes(color = cty))
```

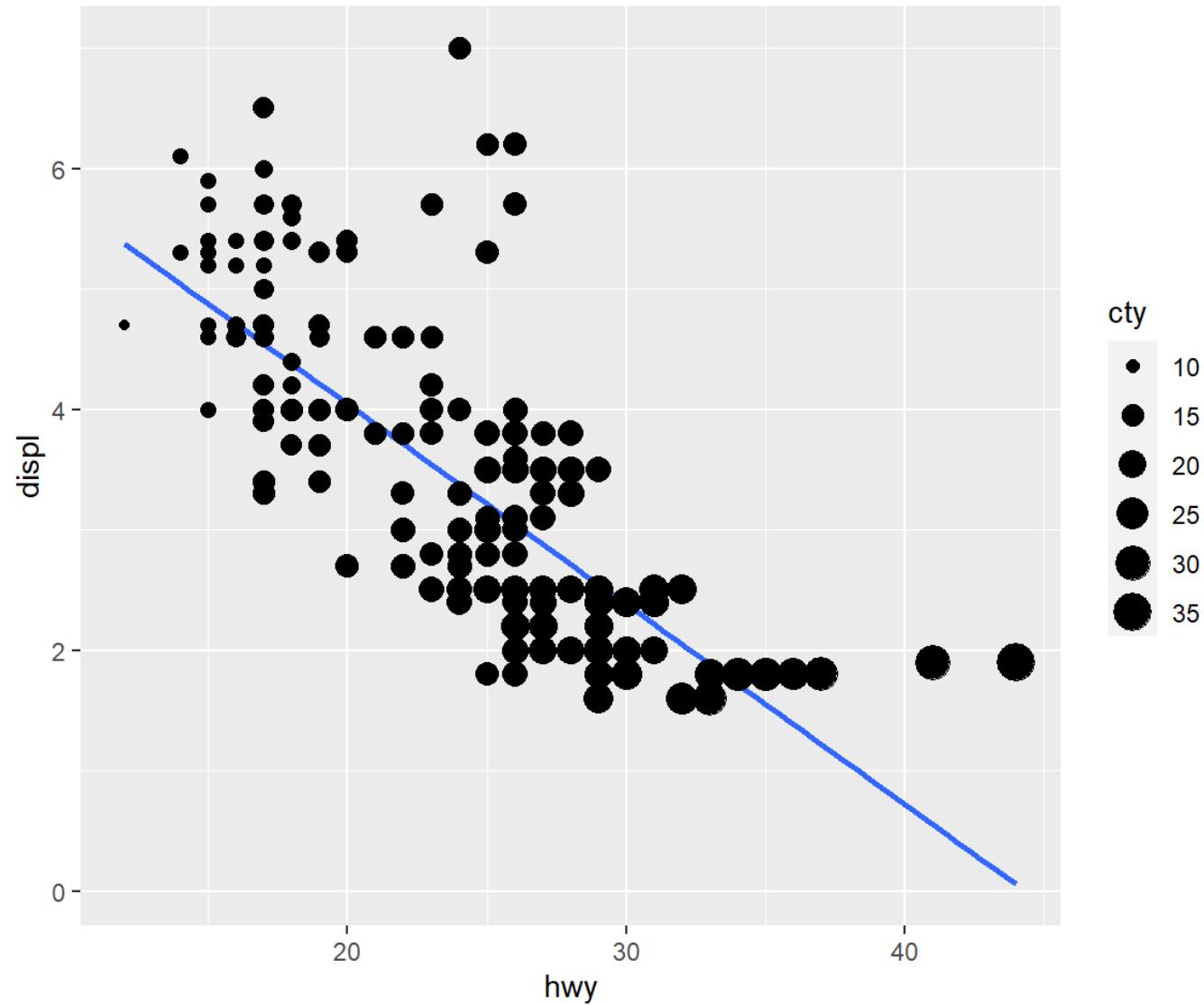
Resposta



Resposta

```
1 p + geom_point(aes(size = cty))
```

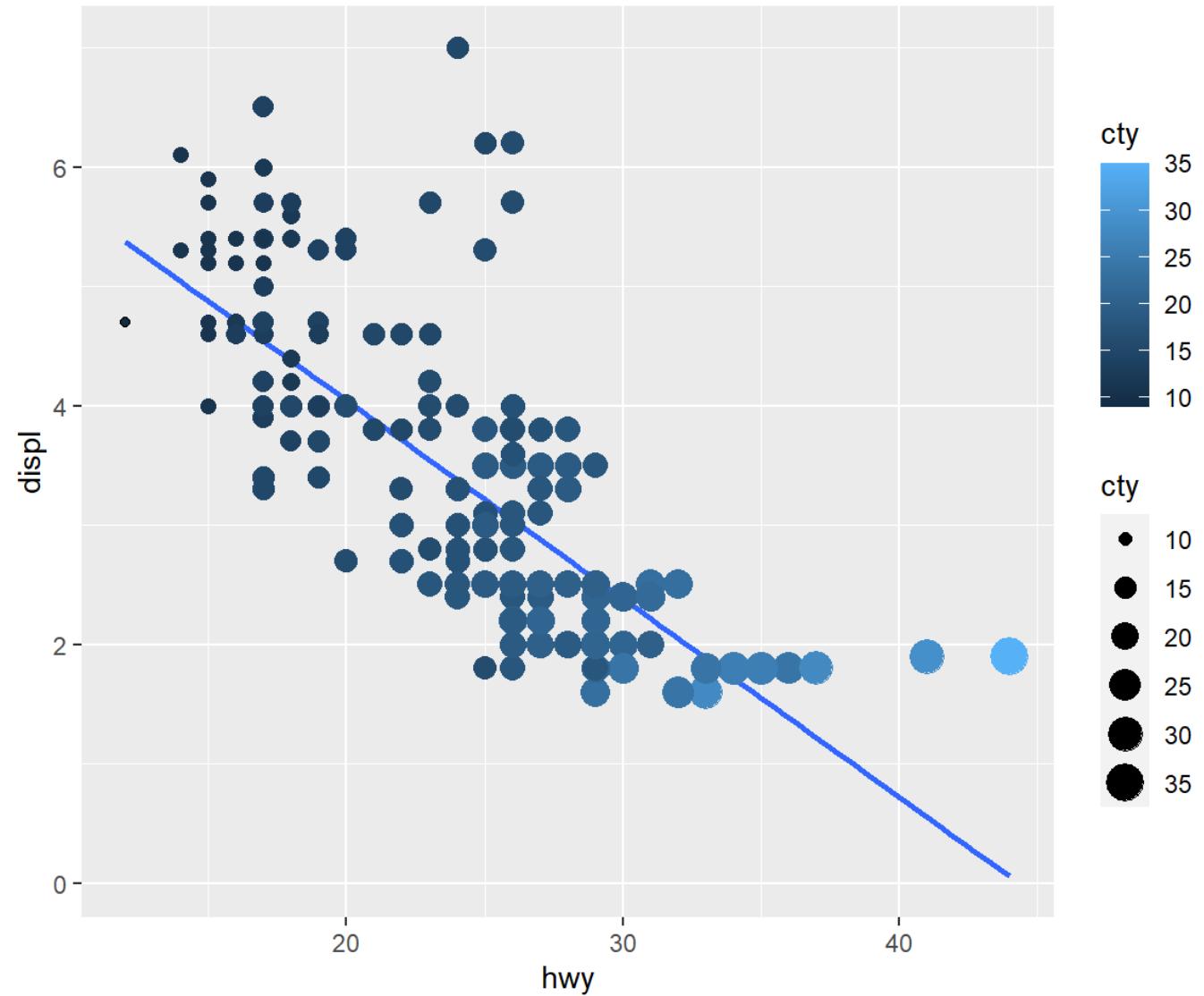
Resposta



Resposta

```
1 p + geom_point(aes(color = cty,  
2                   size = cty))
```

Resposta



Resposta

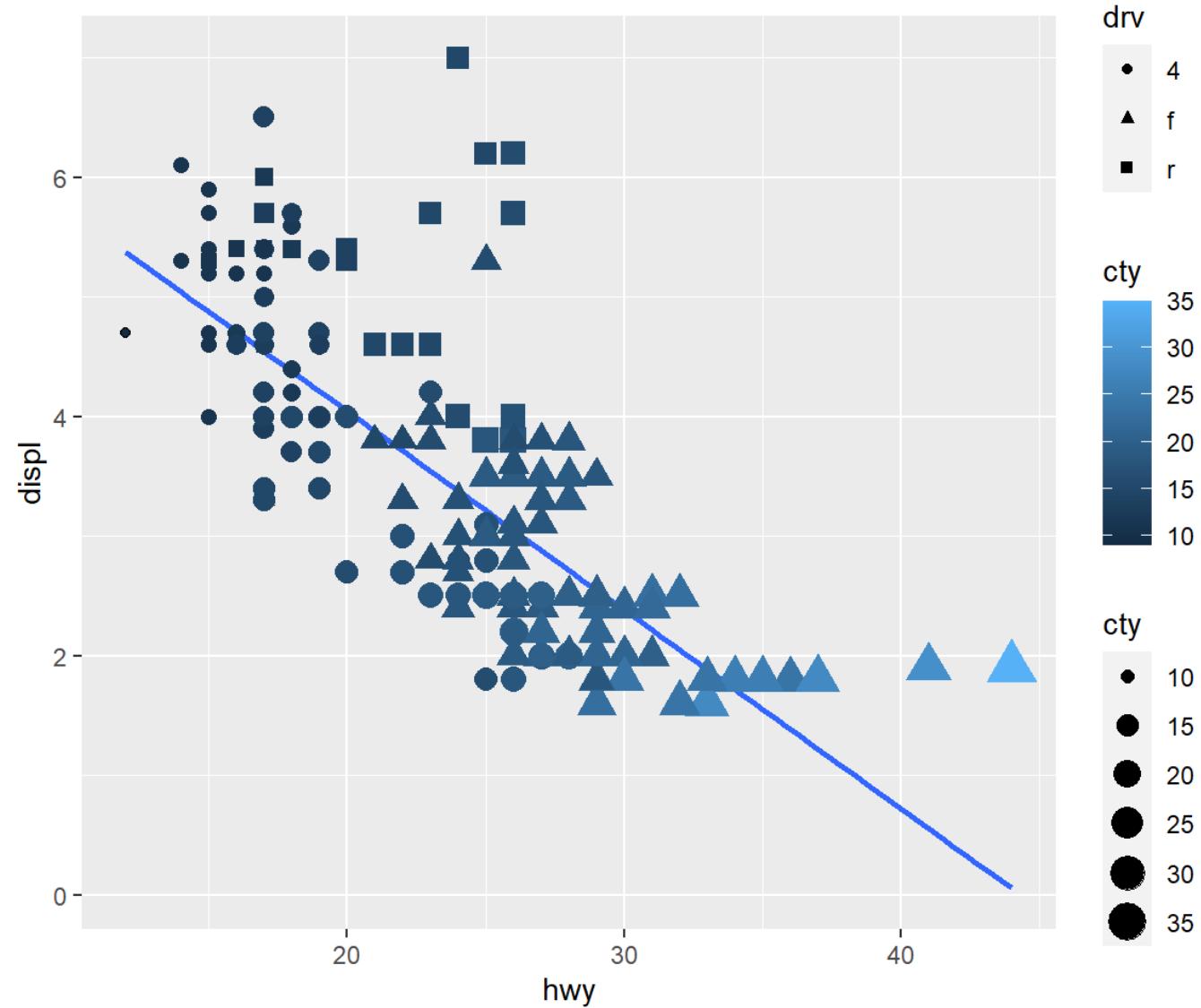
```
1 p + geom_point(aes(shape = cty))
```

Error in `geom_point()`:! Problem while computing aesthetics. **i** Error occurred in the 2nd layer. Caused by error in `scale_f()`:! A continuous variable cannot be mapped to the shape aesthetic **i** choose a different aesthetic or use `scale_shape_binned()` Run `rlang::last_error()` to see where the error occurred.

Resposta

```
1 p +  
2   geom_point(aes(color = cty,  
3                 size = cty,  
4                 shape = drv))
```

Resposta



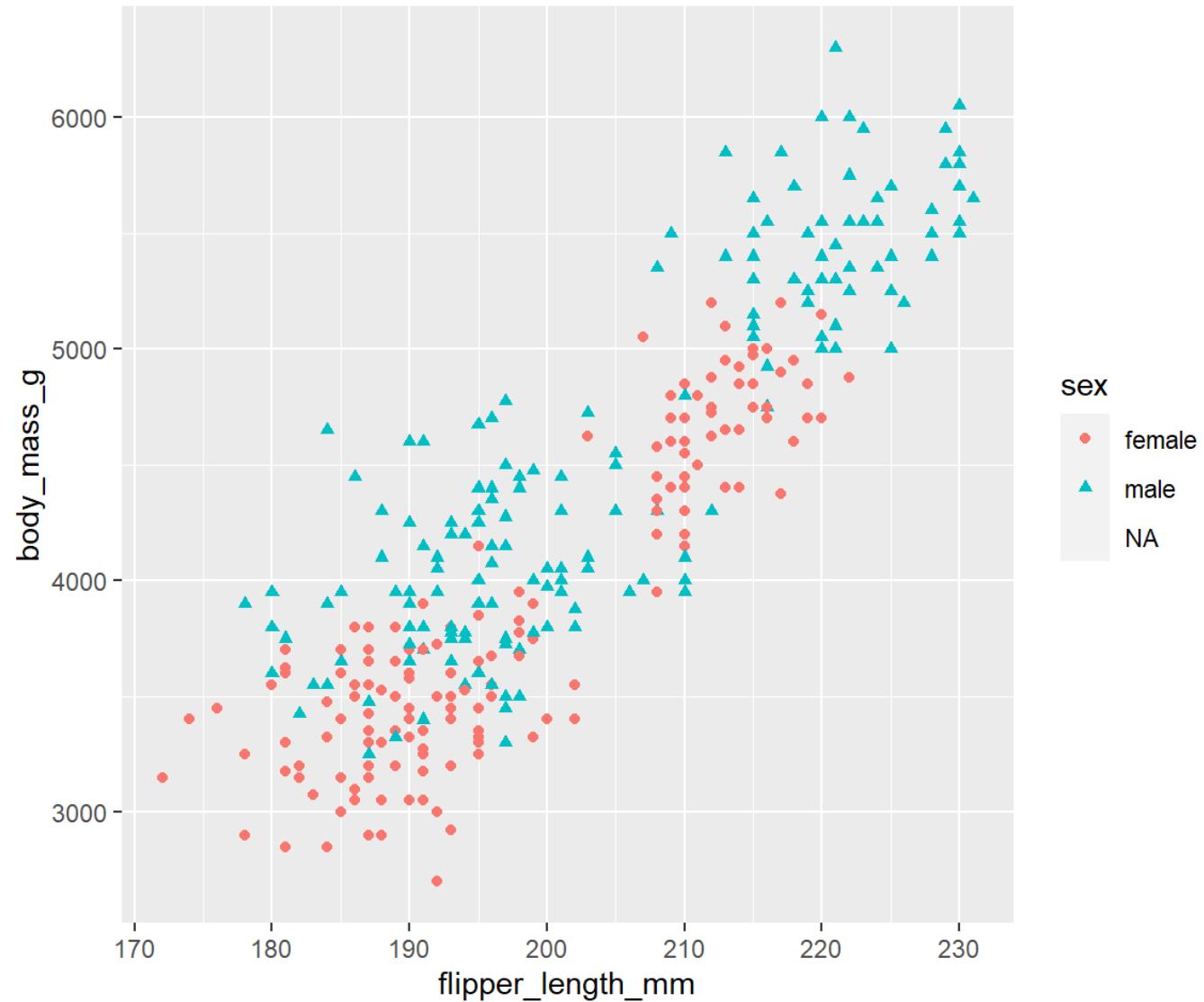
Exercícios

O que acontece se você mapear a mesma variável para várias estéticas?

Resposta

```
1 ggplot(penguins,
2       aes(flipper_length_mm,
3             body_mass_g)) +
4   geom_point(aes(color = sex,
5               shape = sex))
```

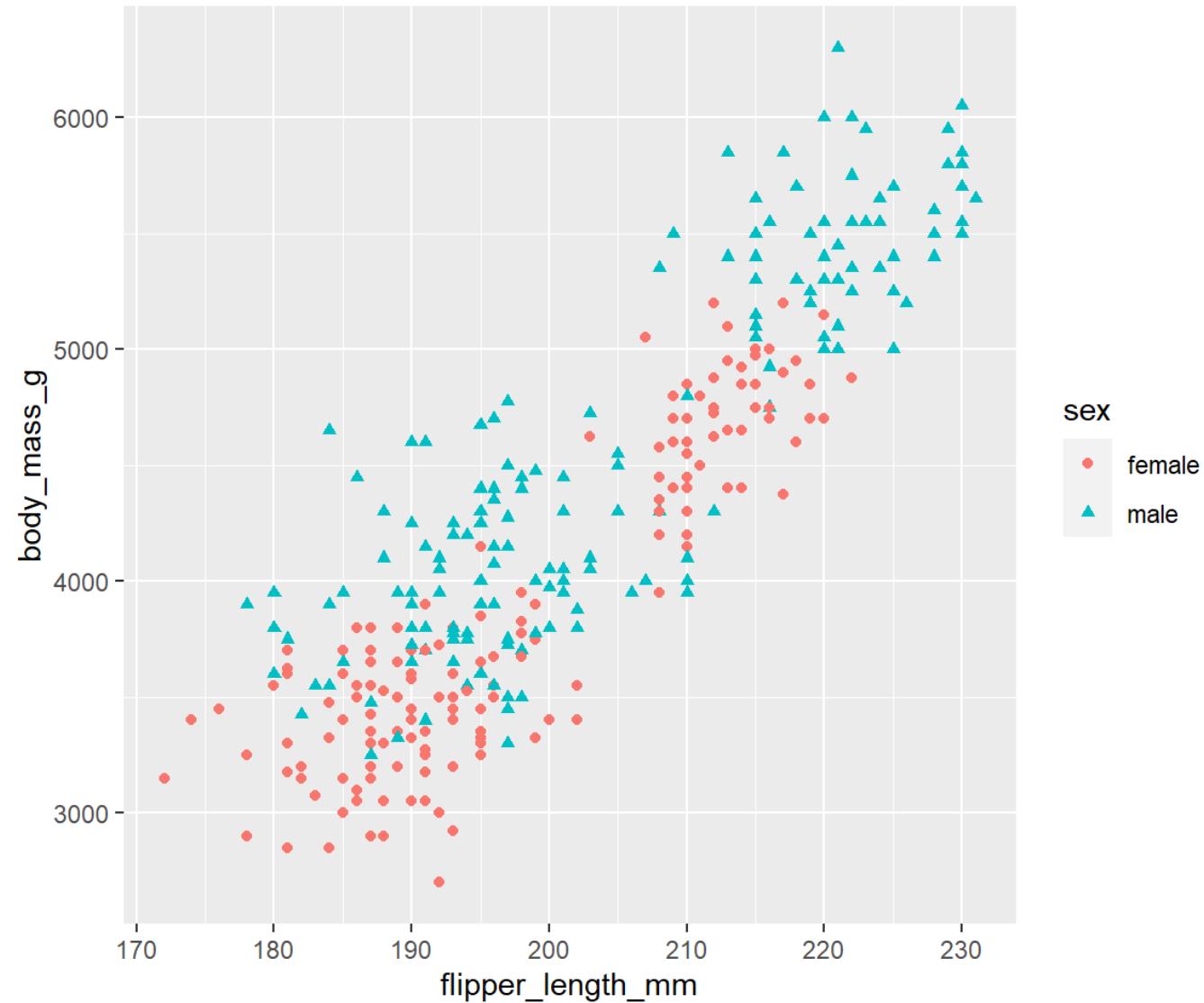
Resposta



Resposta

```
1 penguins |>
2   drop_na() |>
3   ggplot(aes(flipper_length_mm,
4             body_mass_g)) +
5   geom_point(aes(color = sex,
6               shape = sex))
```

Resposta



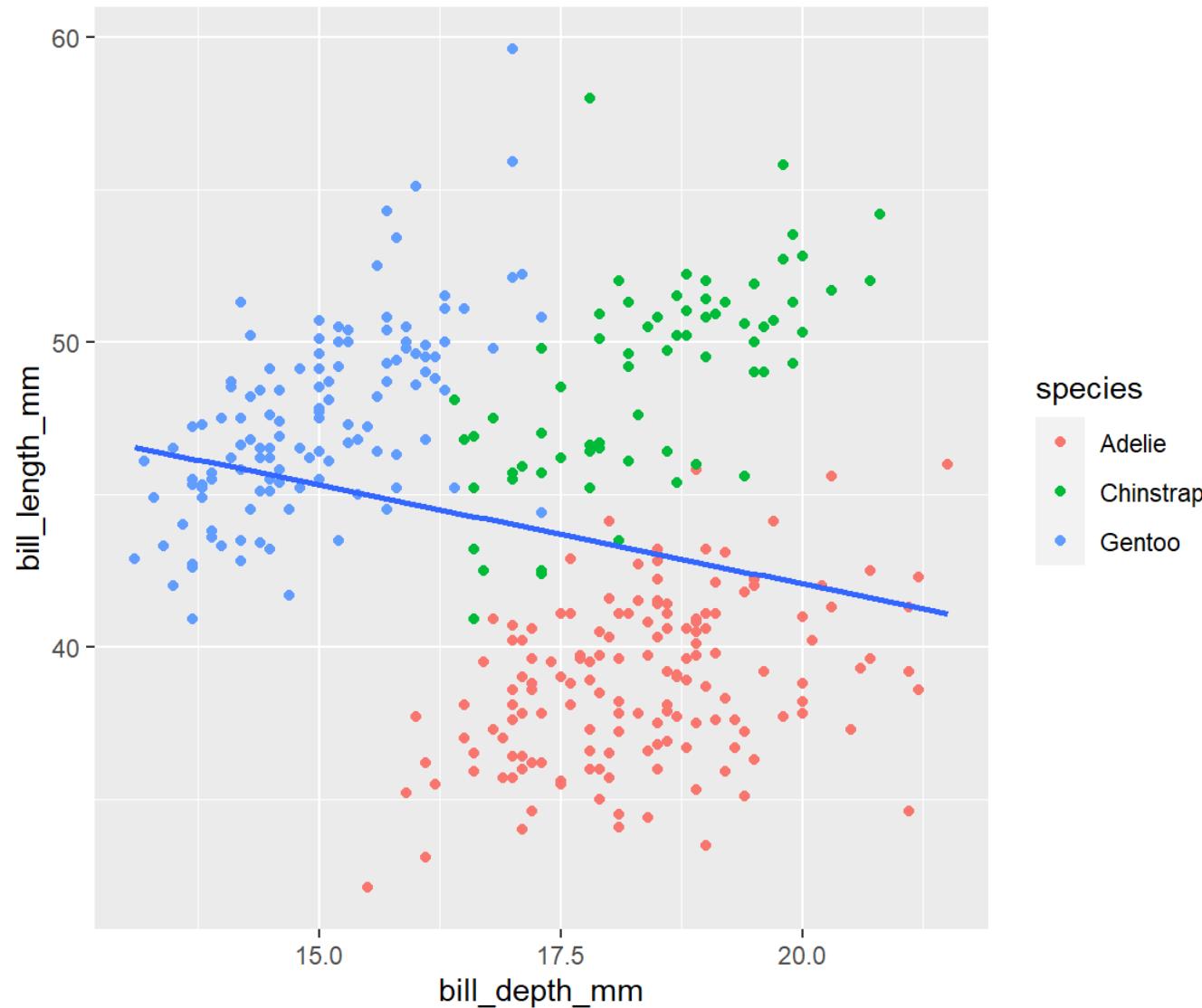
Exercícios

Faça um gráfico de dispersão de bill_depth_mm vs. bill_length_mm. Você deve colorir os pontos por espécies. O que a adição de coloração por espécie revela sobre a relação entre essas duas variáveis? Que tal facetar por espécie?

Resposta

```
1 ggplot(penguins,
2     aes(bill_depth_mm,
3         bill_length_mm)) +
4     geom_point(aes(color = species)) +
5     geom_smooth(method = "lm",
6                 se = F)
```

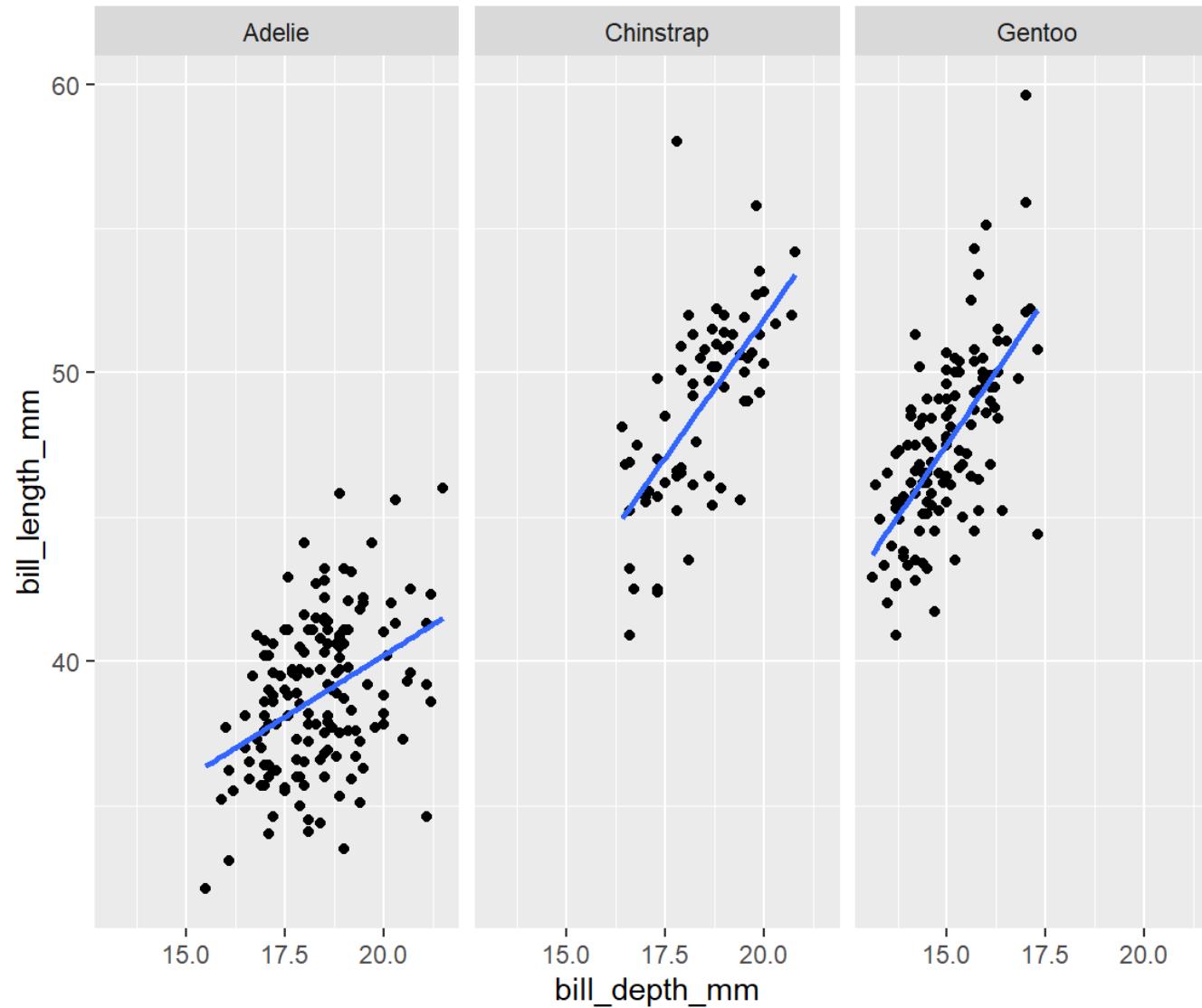
Resposta



Resposta

```
1 ggplot(penguins,
2     aes(bill_depth_mm,
3         bill_length_mm)) +
4     geom_point() +
5     geom_smooth(method = "lm",
6                 se = F) +
7     facet_wrap(~ species)
```

Resposta



2.5 Salvando seus gráficos

- `ggsave()`
- Gráficos ficam salvos no diretório de trabalho
 - `getwd()`
- É possível especificar `width` (largura) e `height`
- `?ggsave()`

Código

```
1 histograma <-
2   ggplot(penguins,
3     aes(body_mass_g)) +
4     geom_histogram(color = "gray",
5                   fill = "white") +
6     theme_classic()
7
8 ggsave(filename = "histograma.png", dpi = 600)
```

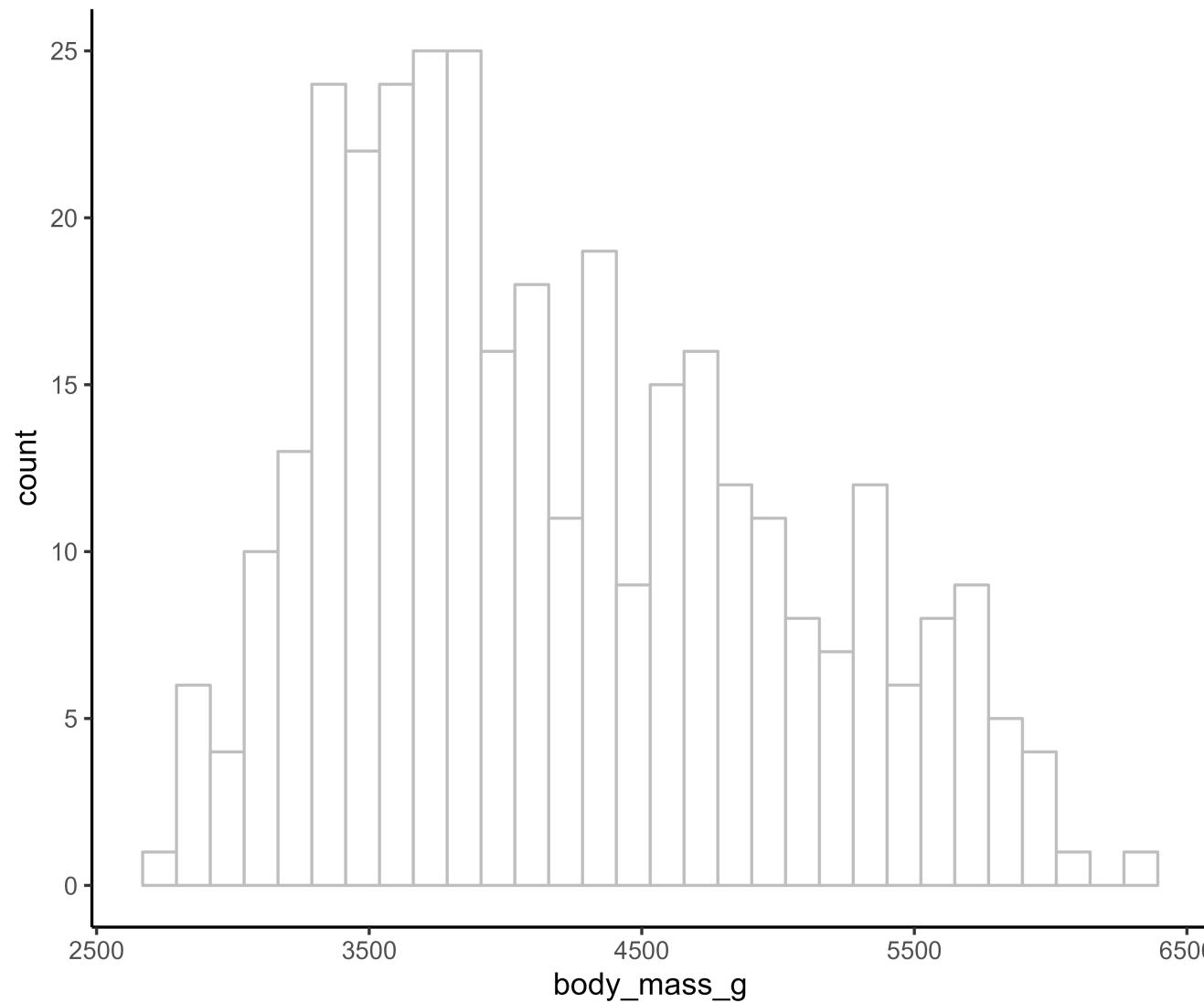
Resultado

The screenshot shows the RStudio interface with the 'Files' tab selected in the top navigation bar. The file explorer sidebar indicates the current location is 'Home > projetos_r > curso_r'. The main area displays a list of files and folders:

	Name	Size	Modified
..	..		
✓	.gitignore	677 B	Apr 6, 2023, 11:26 AM
✓	.Rhistory	9.3 KB	Apr 6, 2023, 7:09 AM
✓	curso_r.Rproj	218 B	Apr 6, 2023, 7:09 AM
✓	encontro_1_cap_3_files		
✓	encontro_1_cap_3.html	97.6 KB	Apr 4, 2023, 6:05 AM
✓	encontro_1_cap_3.qmd	15.1 KB	Apr 2, 2023, 6:58 AM
✓	encontro_2_cap_2_files		
✓	encontro_2_cap_2.html	116.9 KB	Apr 6, 2023, 5:19 PM
✓	encontro_2_cap_2.qmd	32.7 KB	Apr 6, 2023, 5:19 PM
✓	images		
✓	README.md	311 B	Mar 21, 2023, 10:34 PM
✓	slides_pdf		
✓	histograma.png	35.3 KB	Apr 6, 2023, 5:19 PM

The file 'histograma.png' is highlighted with a red border.

Gráfico como imagem



Compartilhar conhecimento sempre!

O código e as imagens utilizados para a construção desse slide estão disponíveis [aqui](#), assim como os slides em formato PDF:

https://github.com/pablo-huascar/curso_r

