Visualizando quantidades

Jairo Nicolau, usando Claus O. Wilke

atualização: 2022-05-02

Frequentemente encontramos datasets contendo quantidades simples

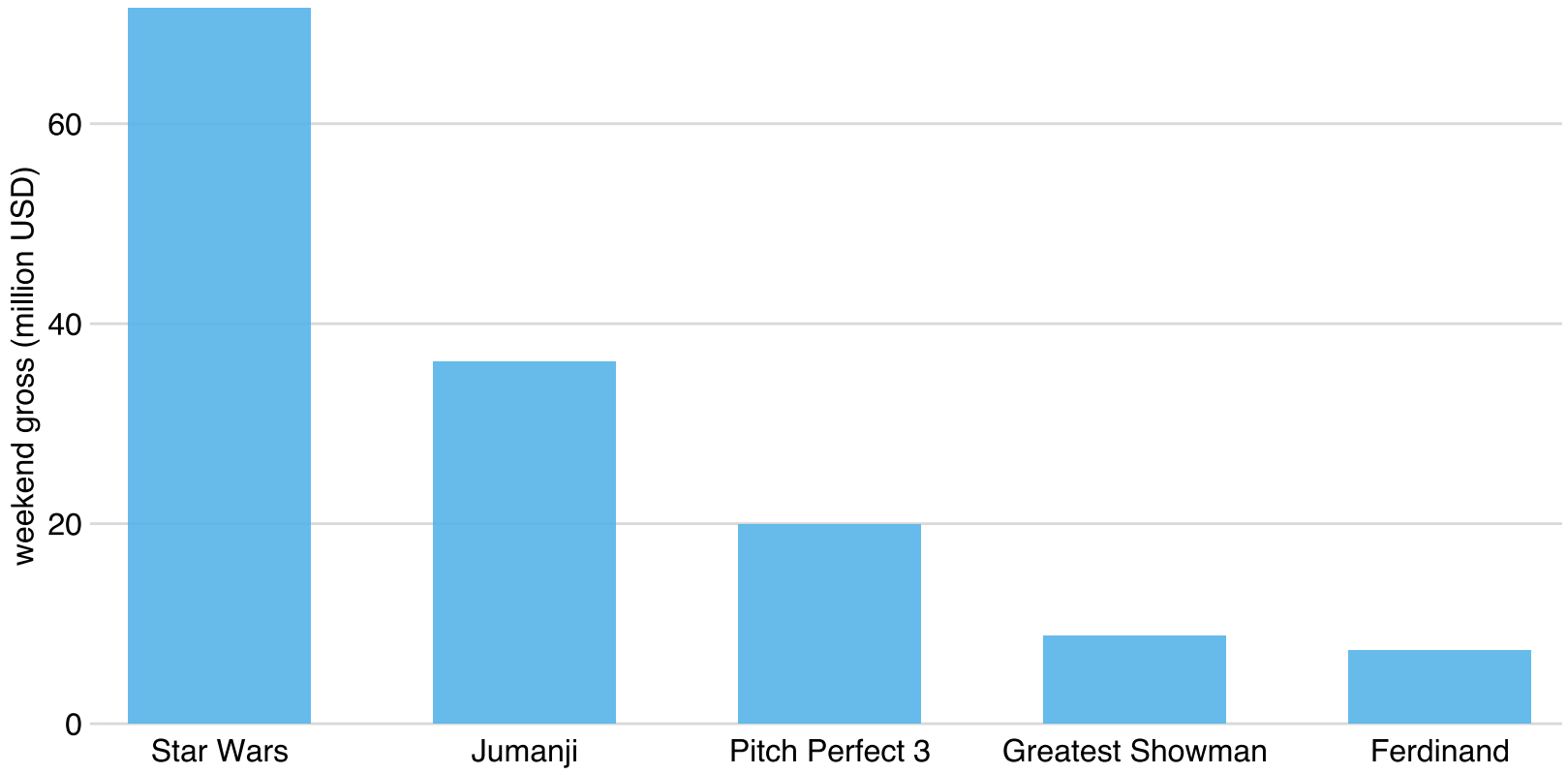
Frequentemente encontramos datasets contendo quantidades simples

Exemplo: filmes de maior bilheteria, dezembro/2017

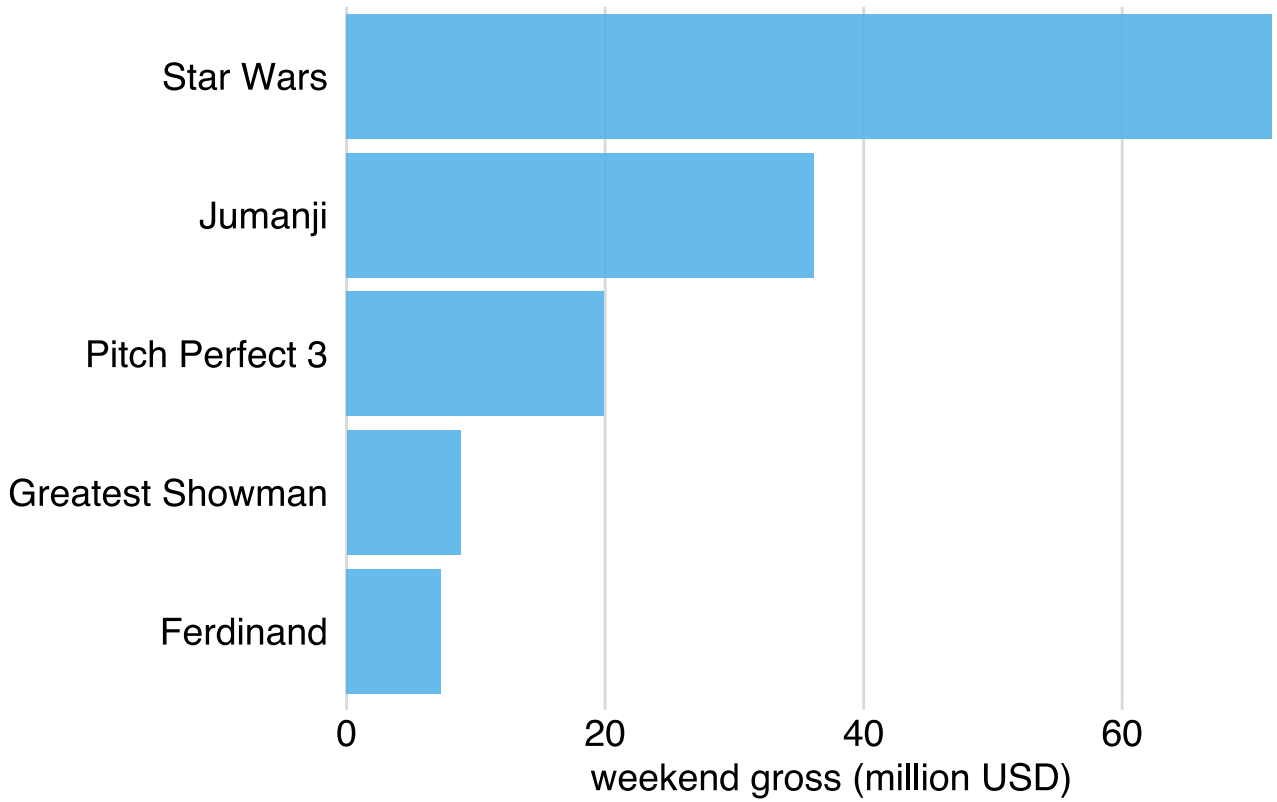
rank	title	amount
1	Star Wars	71.57
2	Jumanji	36.17
3	Pitch Perfect 3	19.93
4	Greatest Showman	8.81
5	Ferdinand	7.32

Data source: Box Office Mojo

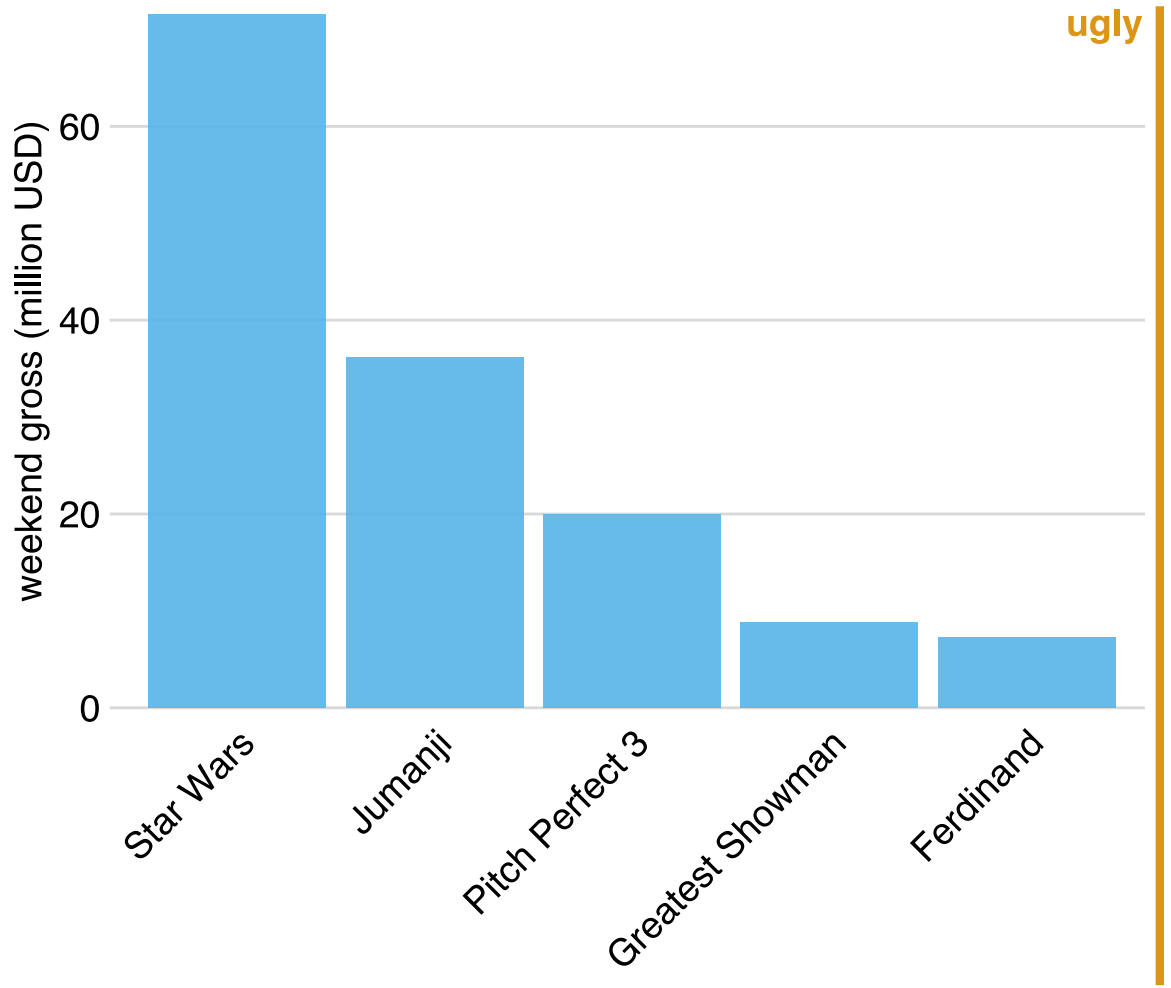
Visualizando quantidades com barras



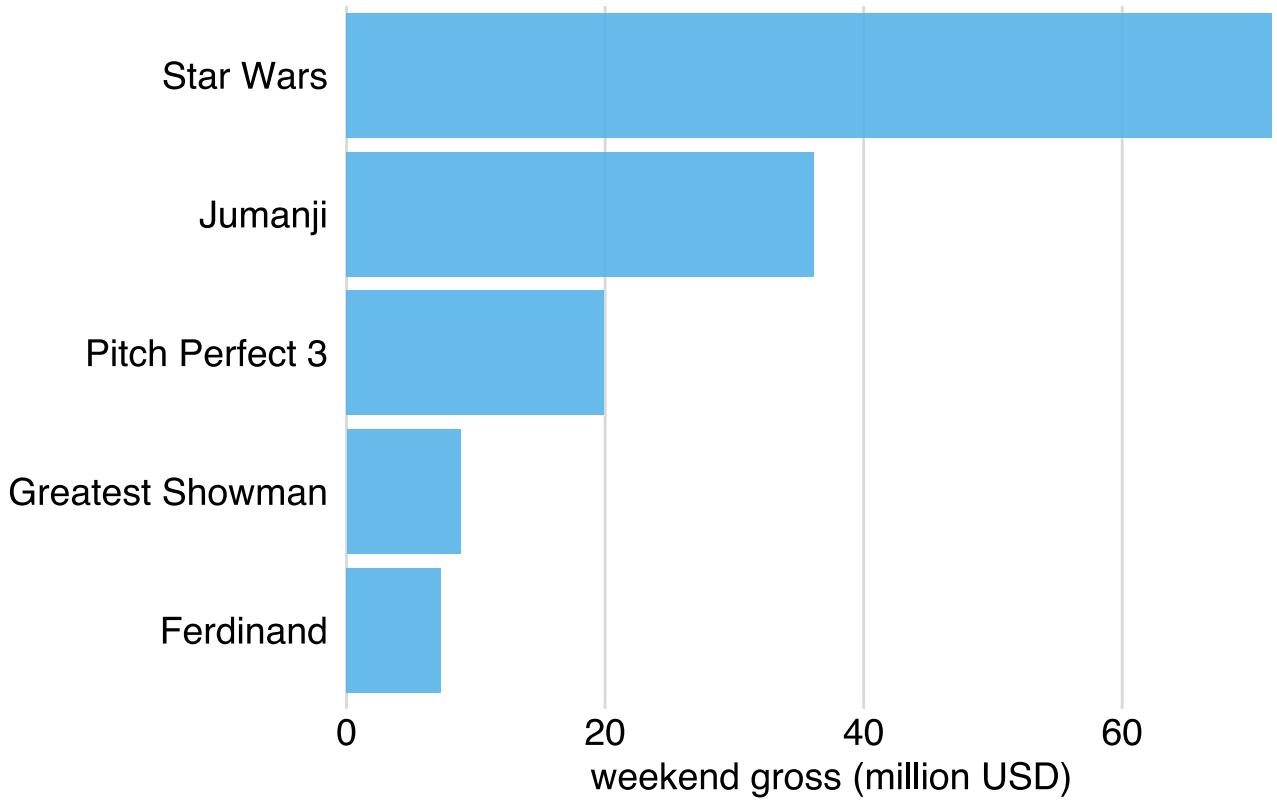
Barras também podem ser horizontais



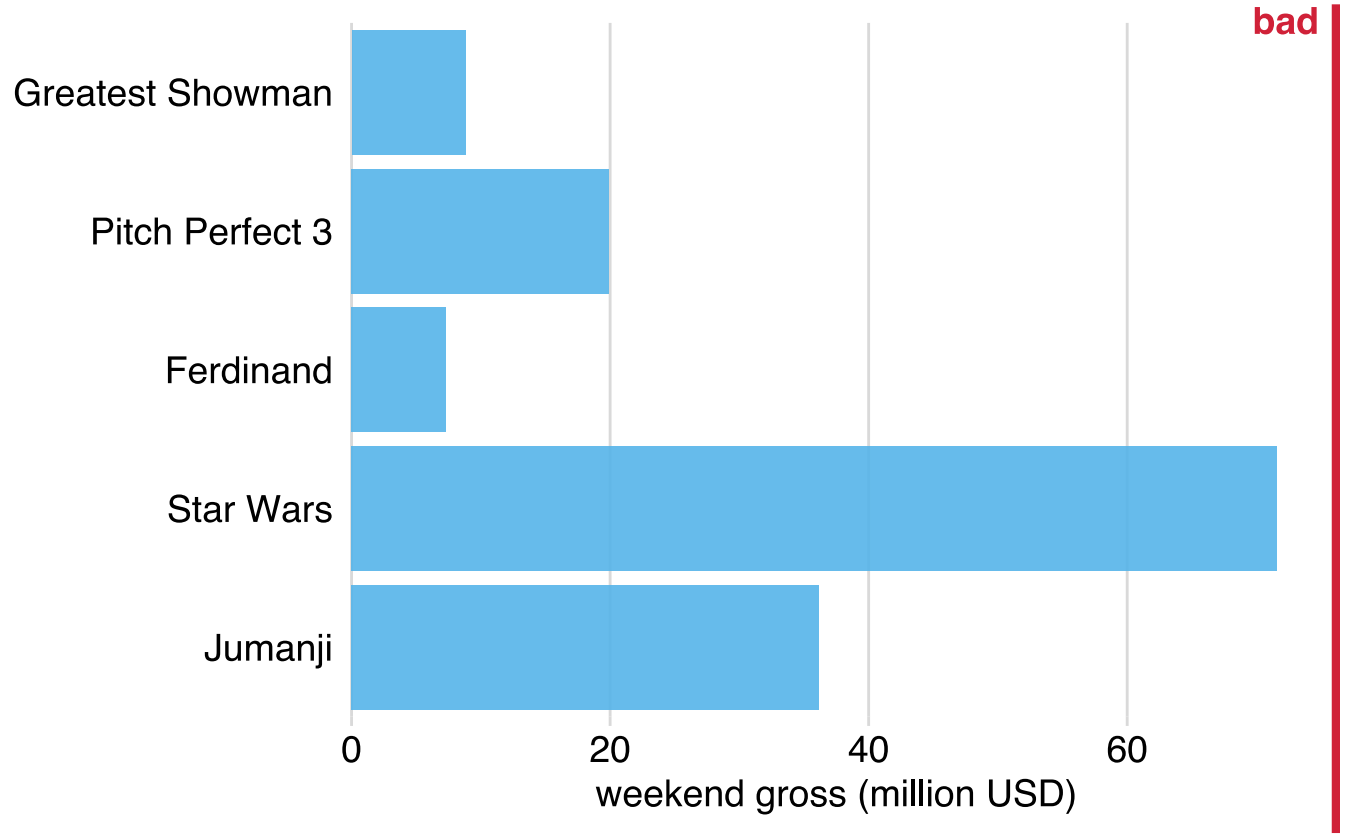
Evite labels na diagonal



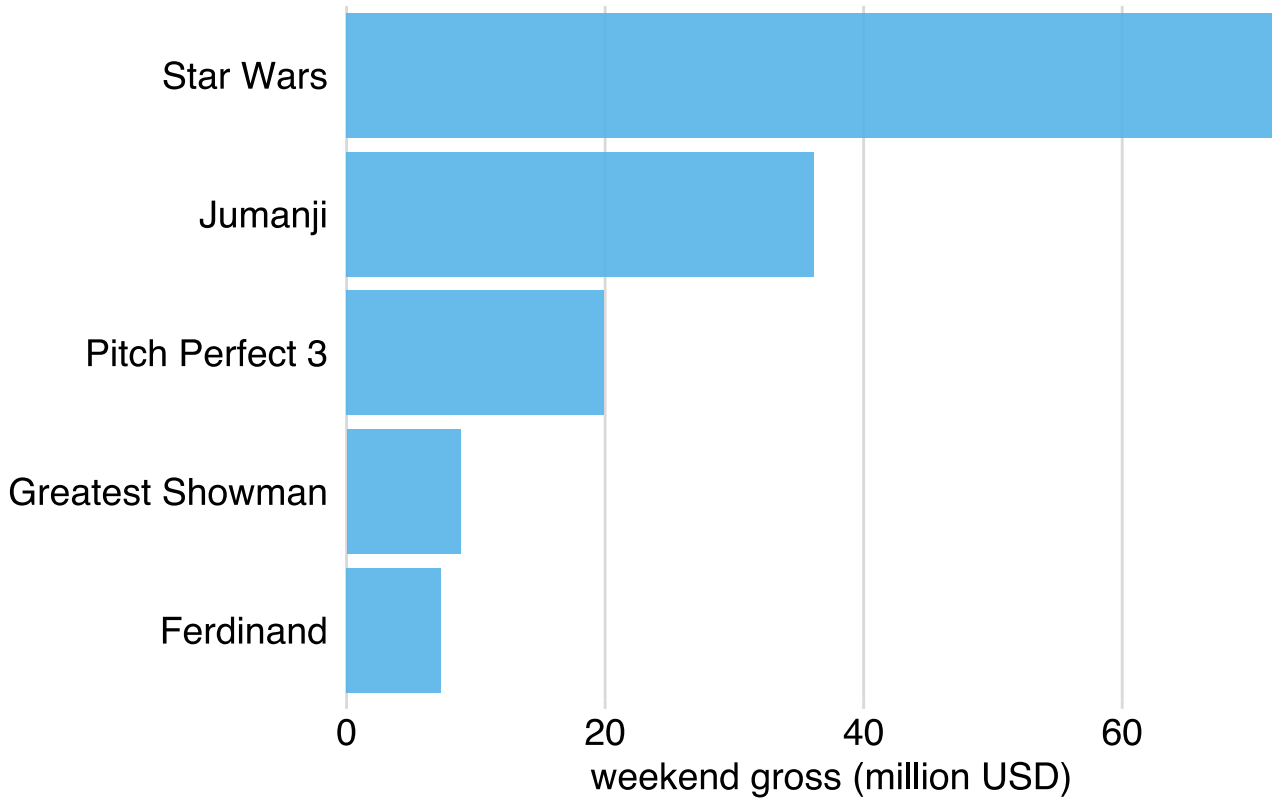
Evite labels na diagonal



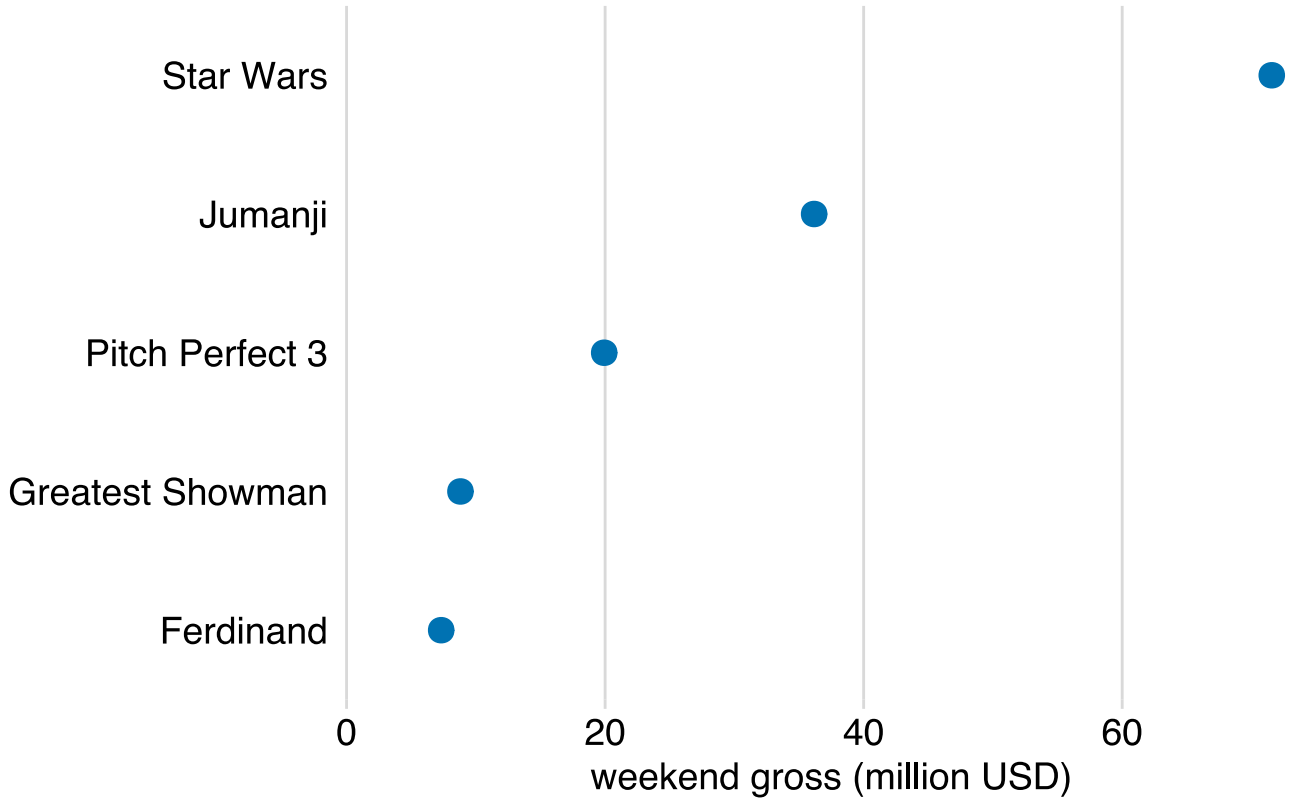
Preste atenção na ordem das barras



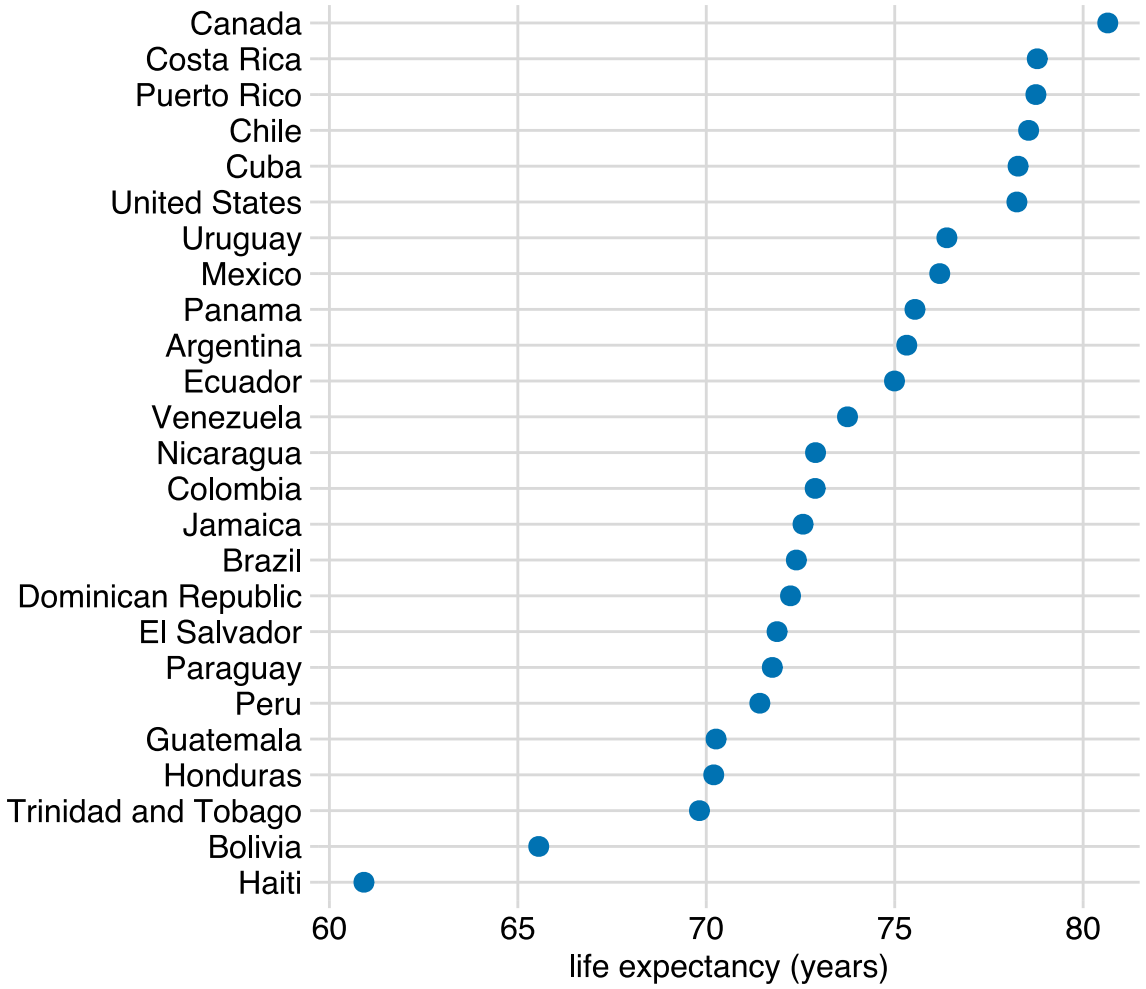
Preste atenção na ordem das barras



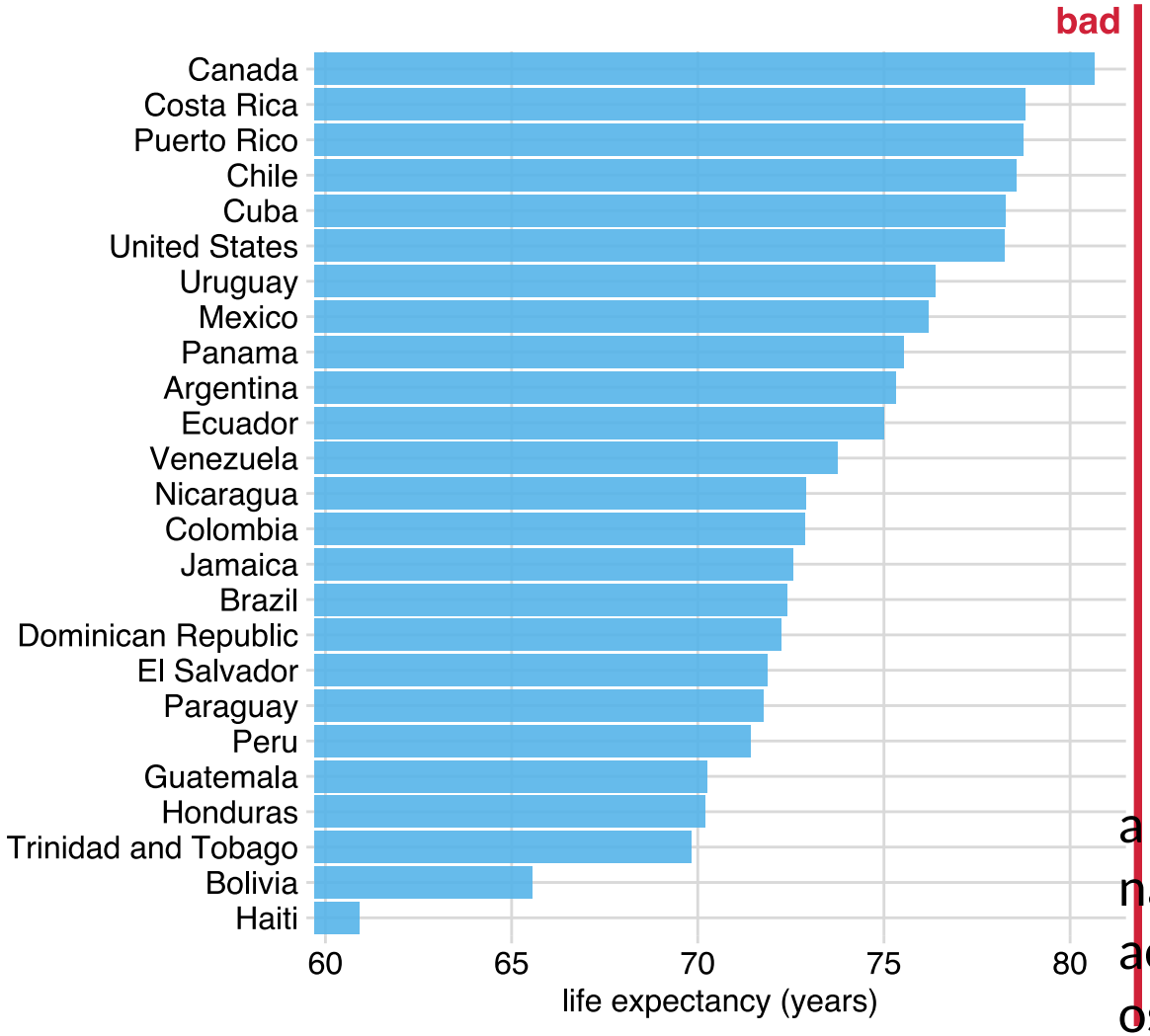
Podemos usar pontos no lugar de barras



Pontos são preferíveis se queremos truncar o eixo

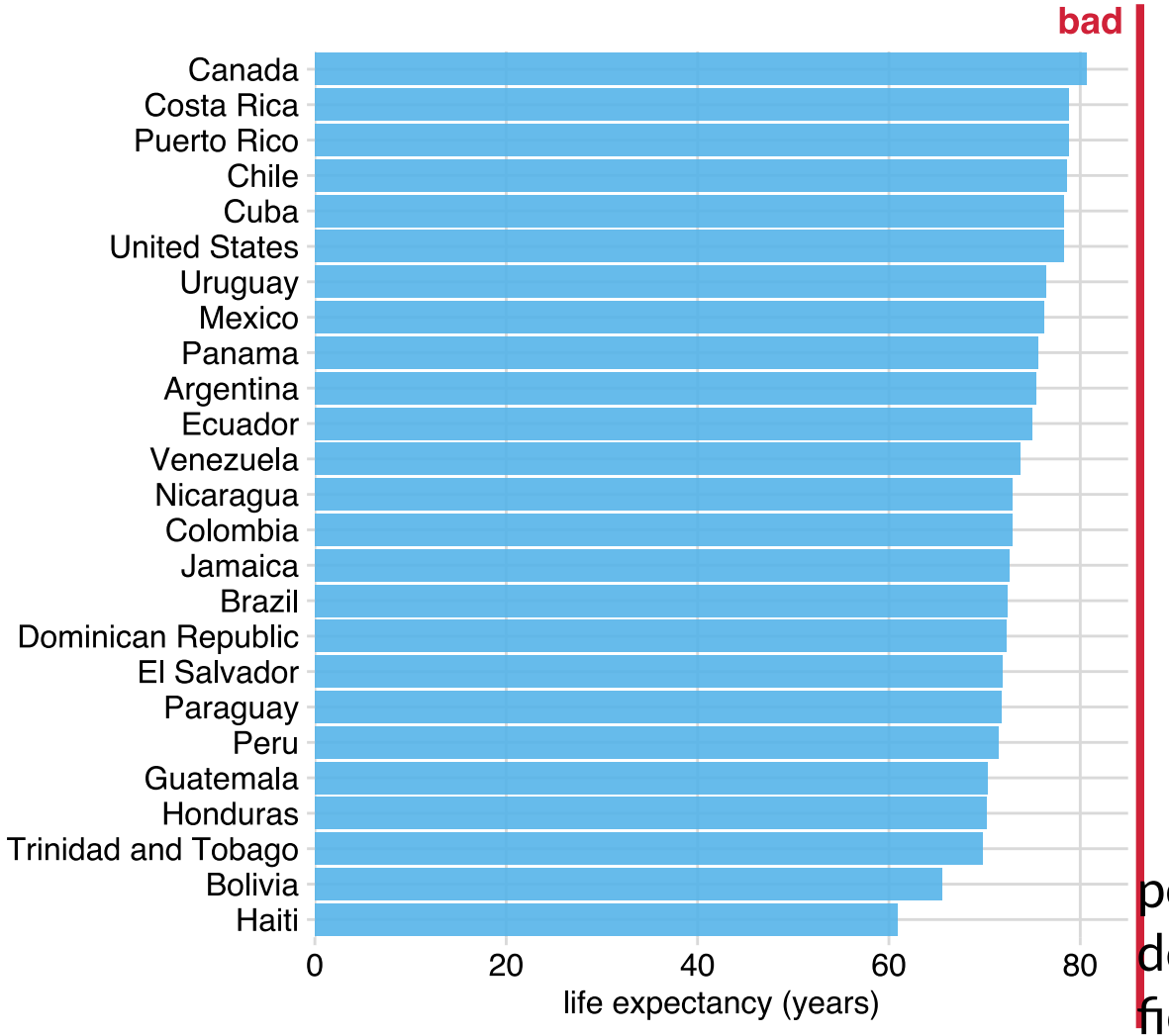


Pontos são preferíveis se queremos truncar o eixo



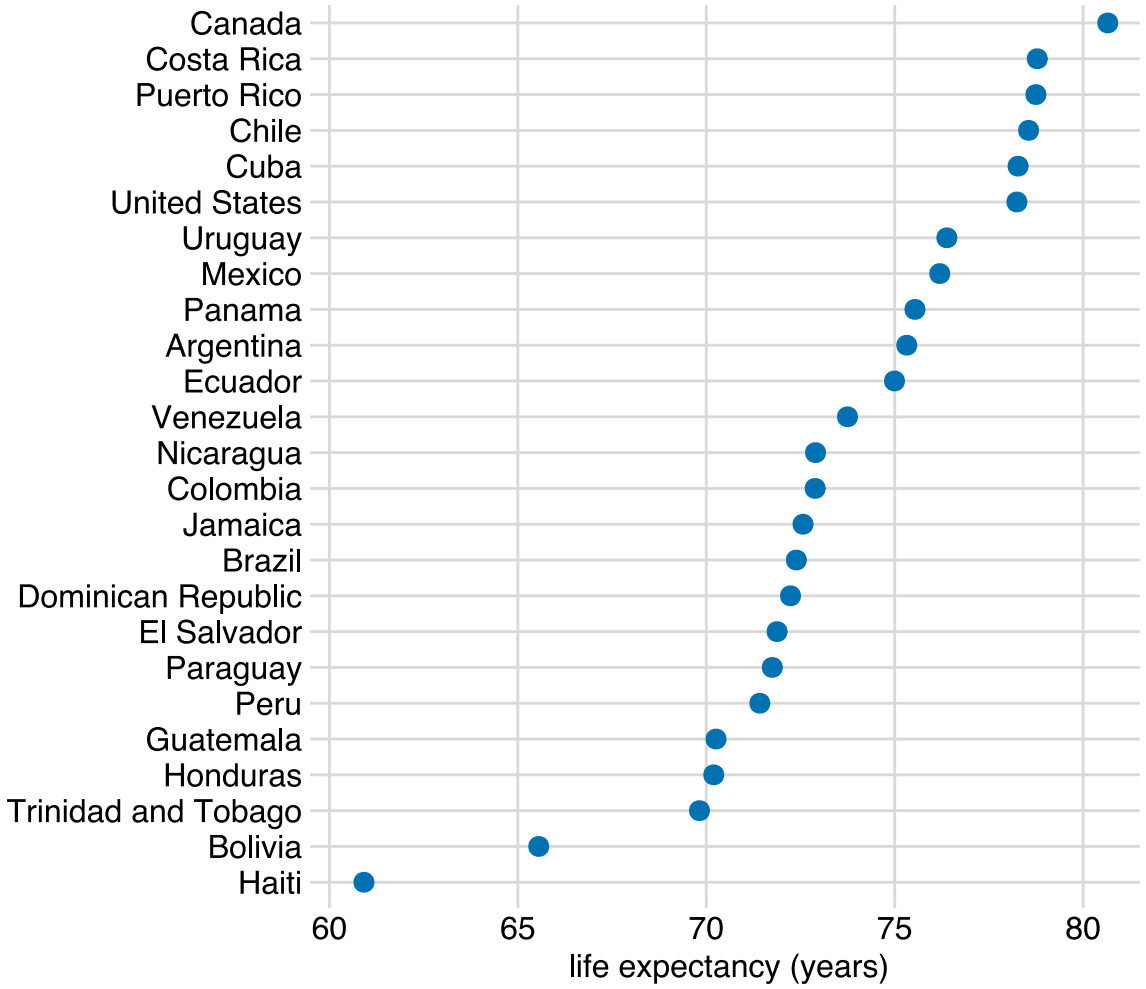
a extensão da barra
não representa
acuradamente
os valores da barra

Pontos são preferíveis se queremos truncar o eixo



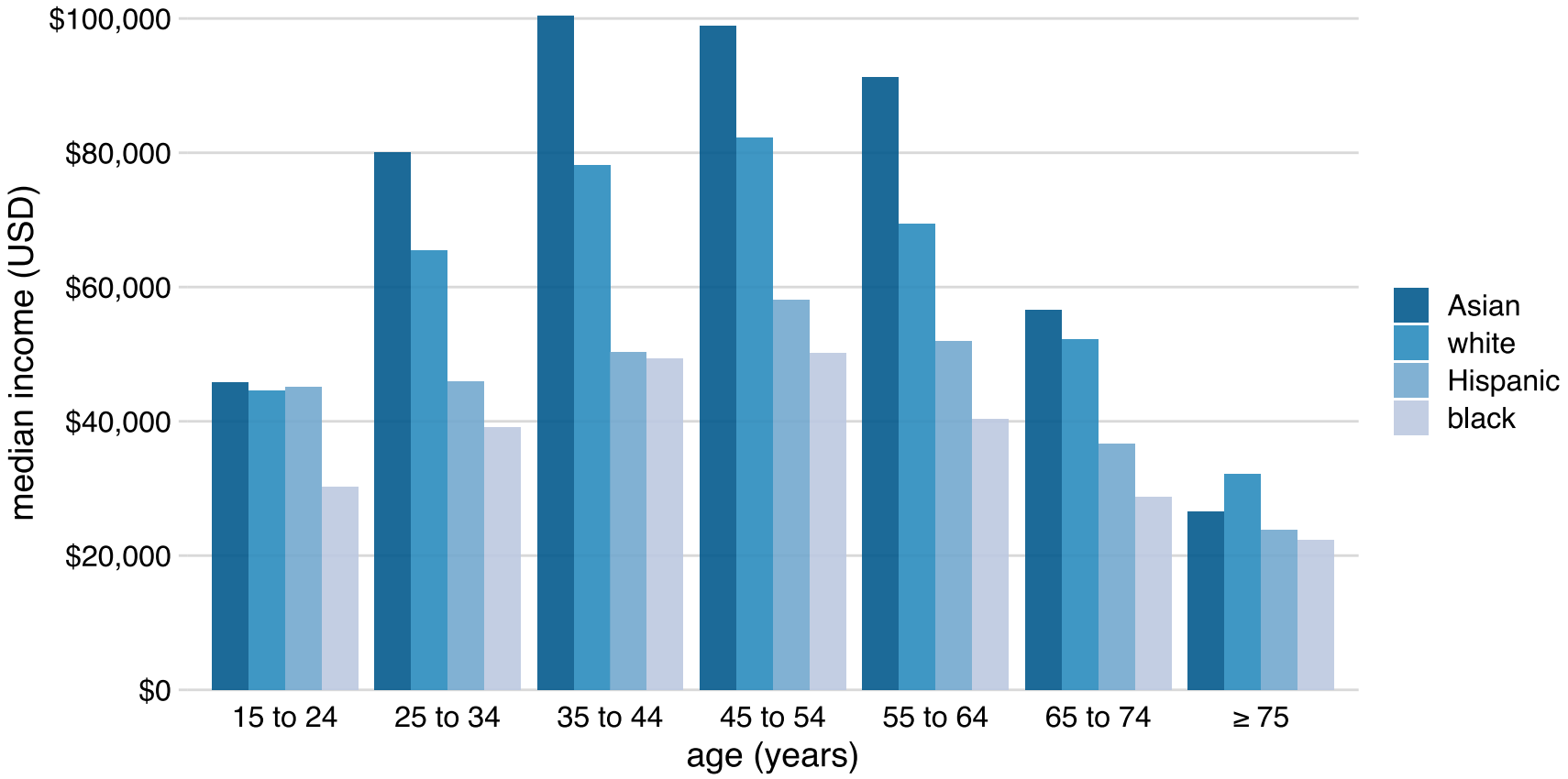
pontos chaves dos dados ficam obscurecidos

Pontos são preferíveis se queremos truncar o eixo



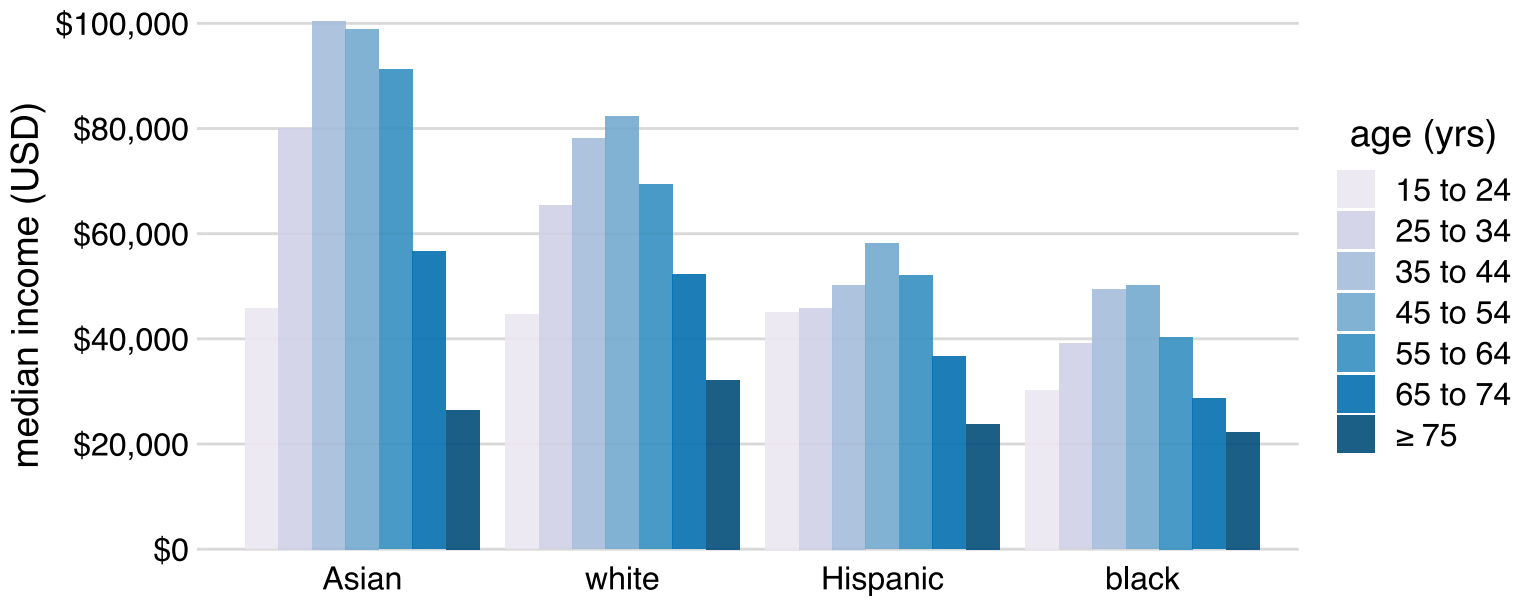
Barras agrupadas

Barras agrupadas para datasets com muitas dimensões



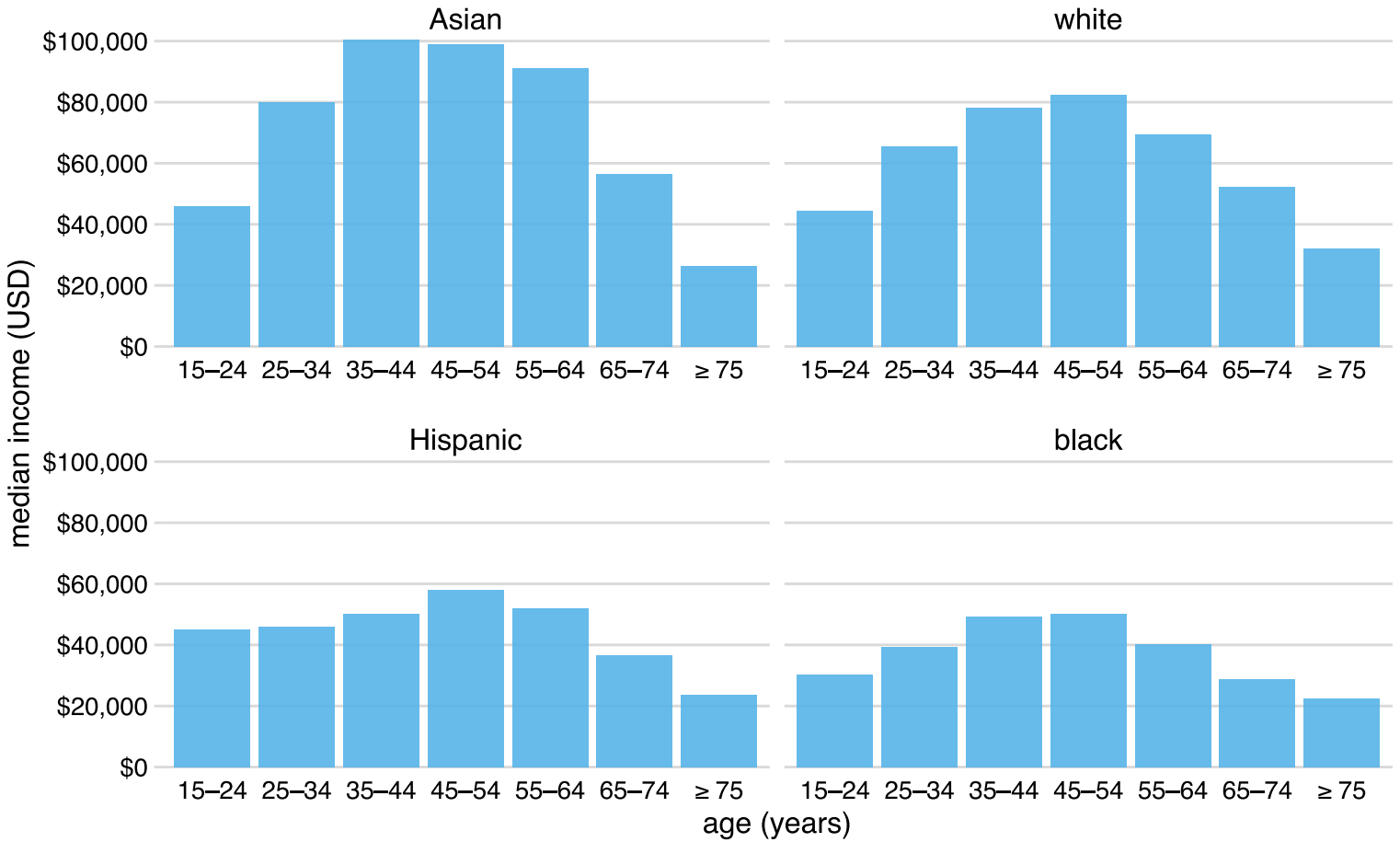
Data source: United States Census Bureau, 2016

Temos a liberdade de escolher a variável que queremos agrupar



Data source: United States Census Bureau, 2016

Podemos também usar painéis múltiplos (facets)



Data source: United States Census Bureau, 2016

Making bar plots in **ggplot2**

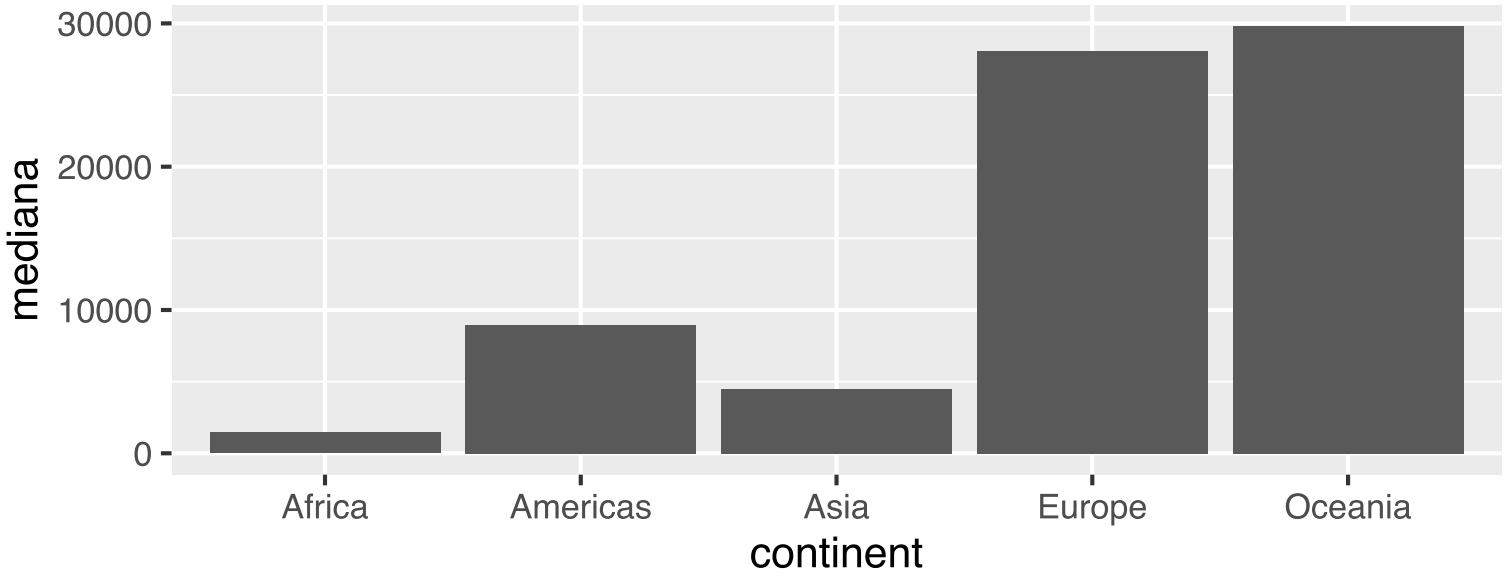
Carregando o Gapminder

```
library (gapminder)


df <- gapminder %>%
      filter(year == 2007) %>%
      group_by(continent) %>%
      summarise(mediana = median(gdpPercap))
head(df)
```

```
# A tibble: 5 × 2
  continent mediana
  <fct>       <dbl>
1 Africa      1452.
2 Americas    8948.
3 Asia        4471.
4 Europe     28054.
5 Oceania    29810.
```

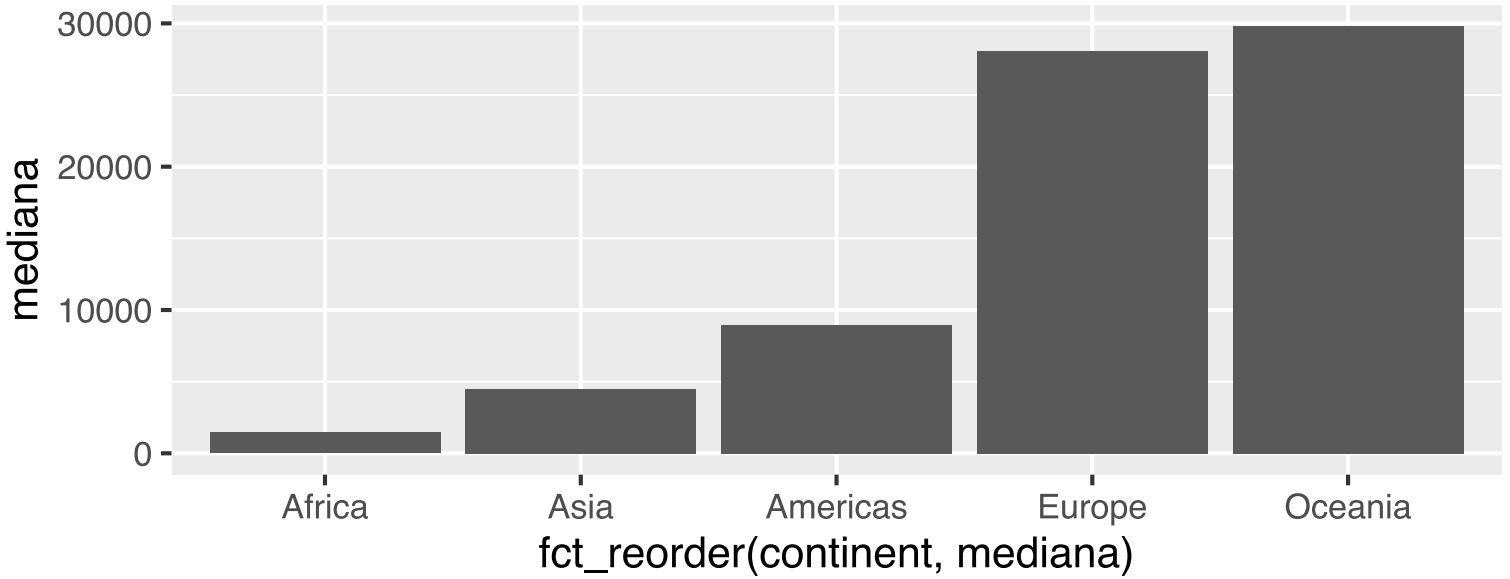
Visualize com um gráfico de barra

```
ggplot(df, aes(continent, mediana)) +
  geom_col()  # "col" é o atalho para coluna
```



Ordene pela variável quantitativa

```
ggplot(df, aes(fct_reorder(continent, mediana), mediana)) +
  geom_col()
```

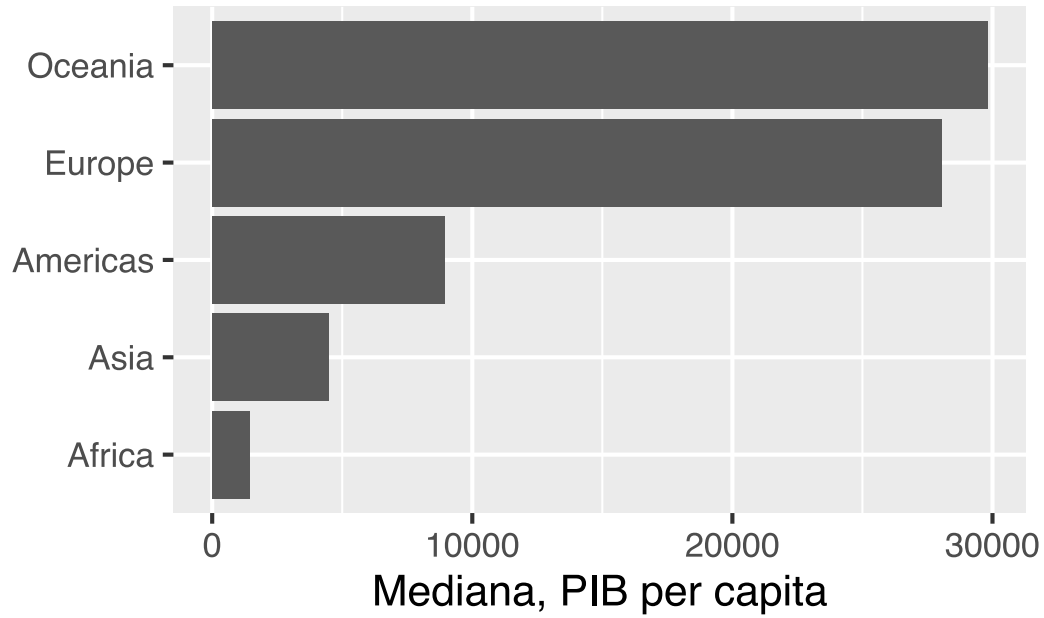


Ordene pela variável quantitativa, descendente

```
ggplot(df, aes(fct_reorder(continent, -mediana), mediana)) +
  geom_col() +
  xlab(NULL) # remova o label do eixo x
```

Vire os eixos x e y, customize o label do eixo x

```
ggplot(df, aes(mediana, fct_reorder(continent, mediana))) +
  geom_col() +
  xlab("Mediana, PIB per capita") +
  ylab(NULL)
```



Gráficos de barra com dados categóricos

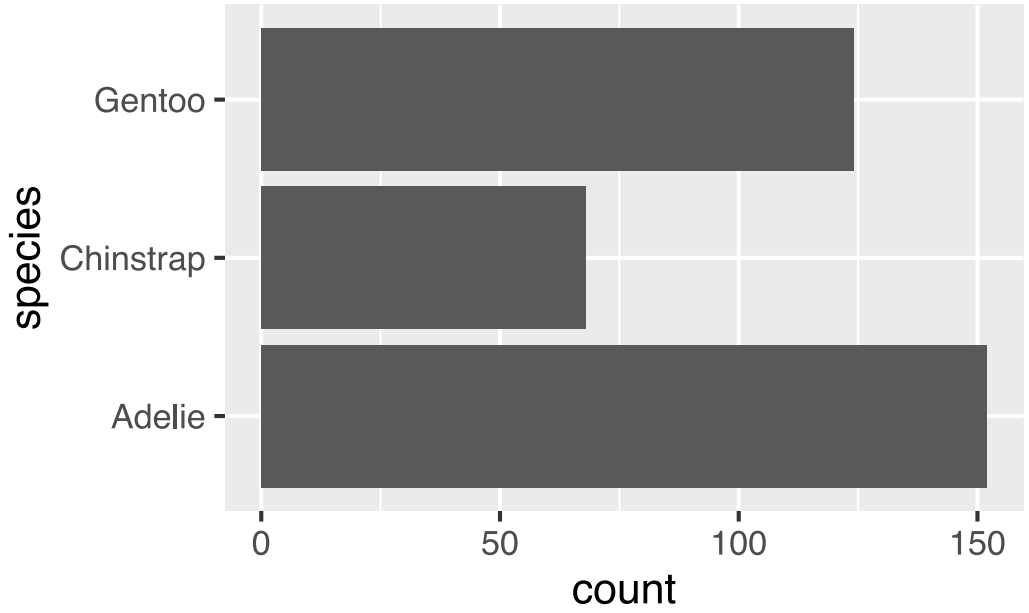
Objetivo: Visualizar o número de pinguins por espécie


```
library(palmerpenguins)
head(penguins)
```

```
# A tibble: 6 × 8
  species island bill_length_mm bill_depth_mm flipper_length_… body_mass_g sex
  <fct>   <fct>           <dbl>         <dbl>             <int>       <int> <fct>
1 Adelie  Torge…           39.1          18.7               181        3750 male
2 Adelie  Torge…           39.5          17.4               186        3800 fema…
3 Adelie  Torge…           40.3          18                 195        3250 fema…
4 Adelie  Torge…           NA            NA                  NA          NA <NA>
5 Adelie  Torge…           36.7          19.3               193        3450 fema…
6 Adelie  Torge…           39.3          20.6               190        3650 male
# … with 1 more variable: year <int>
```

Use `geom_bar()` para contar antes de fazer o gráfico

```
ggplot(penguins, aes(y = species)) + # note: nenhuma  aesthetic definida n
  geom_bar()
```

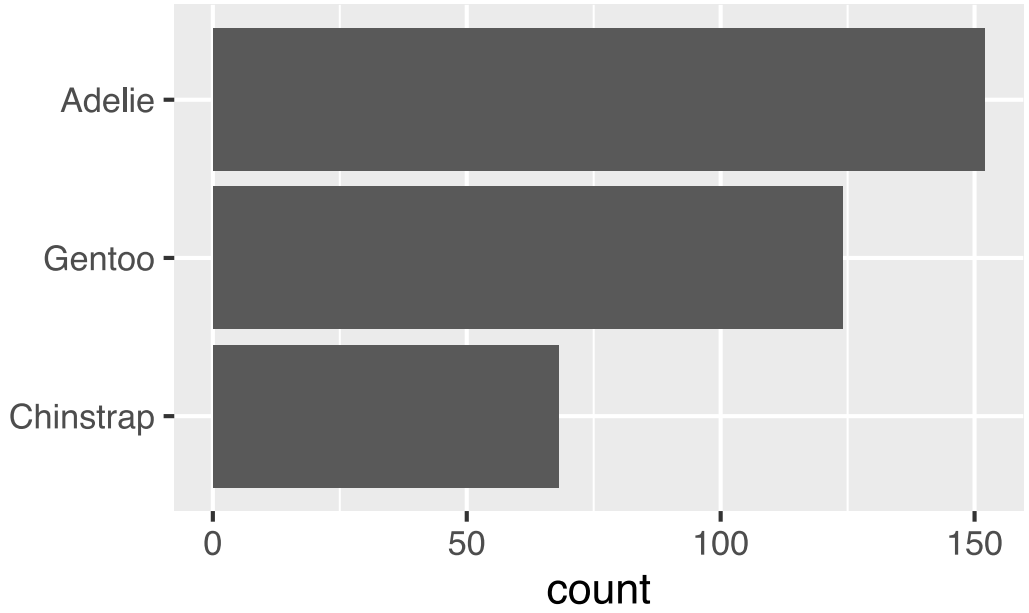


Colocando as barras em ordem

Colocando as barras em ordem

Opção 1: Manualmente, usando `fct_relevel()`

```
ggplot(penguins, aes(y = fct_relevel(species, "Chinstrap", "Gentoo", "Adelie"))) +
  geom_bar() +
  ylab(NULL)
```



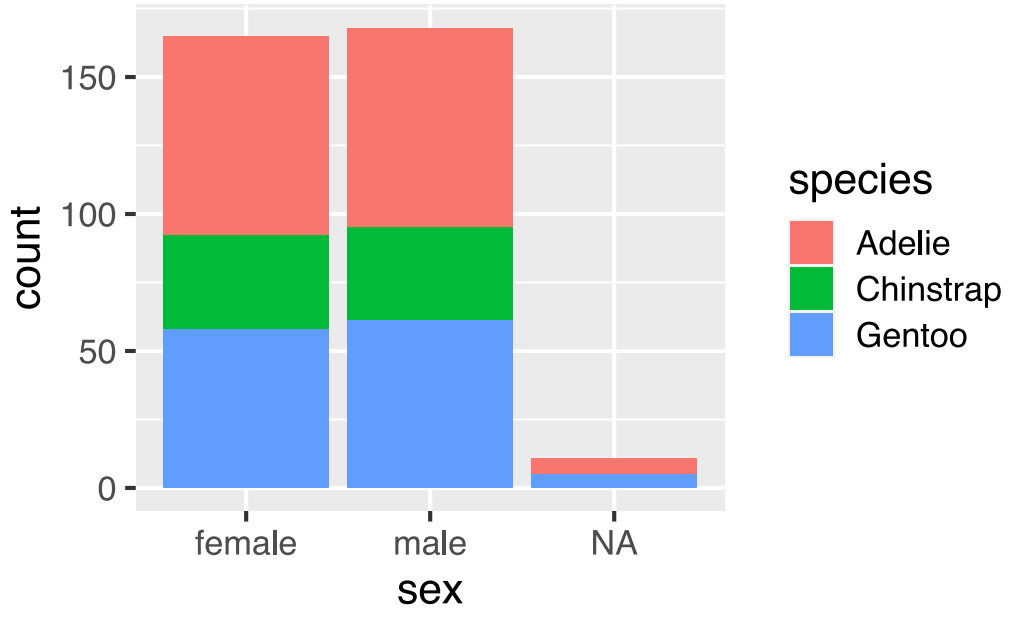
Colocando as barras em ordem

Opção 2: Usando `fct_rev()`, que inverte a ordem e do`fct_infreq()`, que ordena do menor para o maior, do pacote **forcats**

```
ggplot(penguins, aes(y = fct_rev(fct_infreq(species)))) +
  geom_bar() +
  ylab(NULL)
```

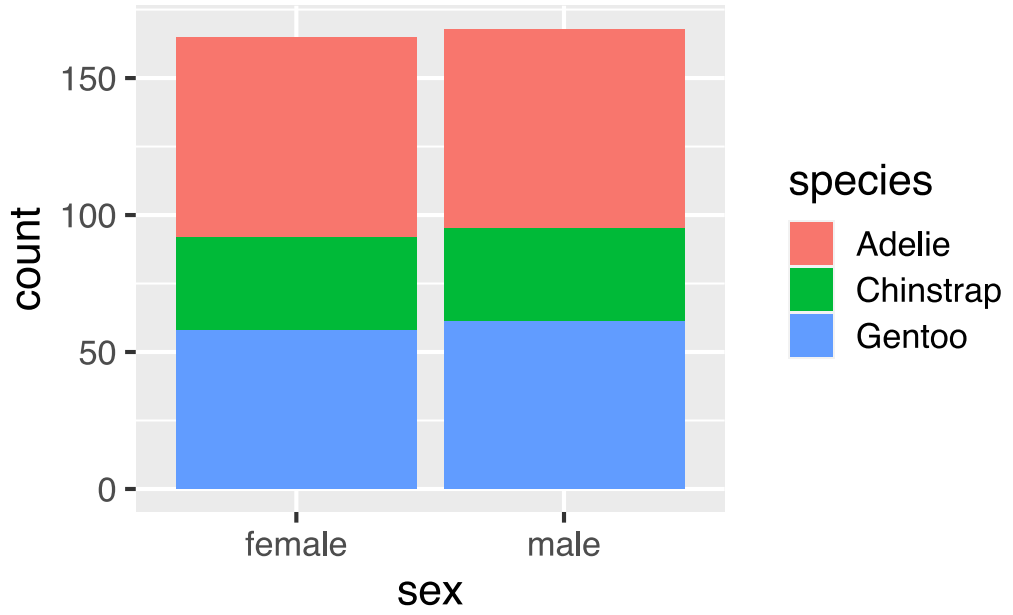
Mostre frequências por espécie e sexo

```
ggplot(penguins, aes(sex, fill = species)) +
  geom_bar()
```



Remova os valores missing (NAs)

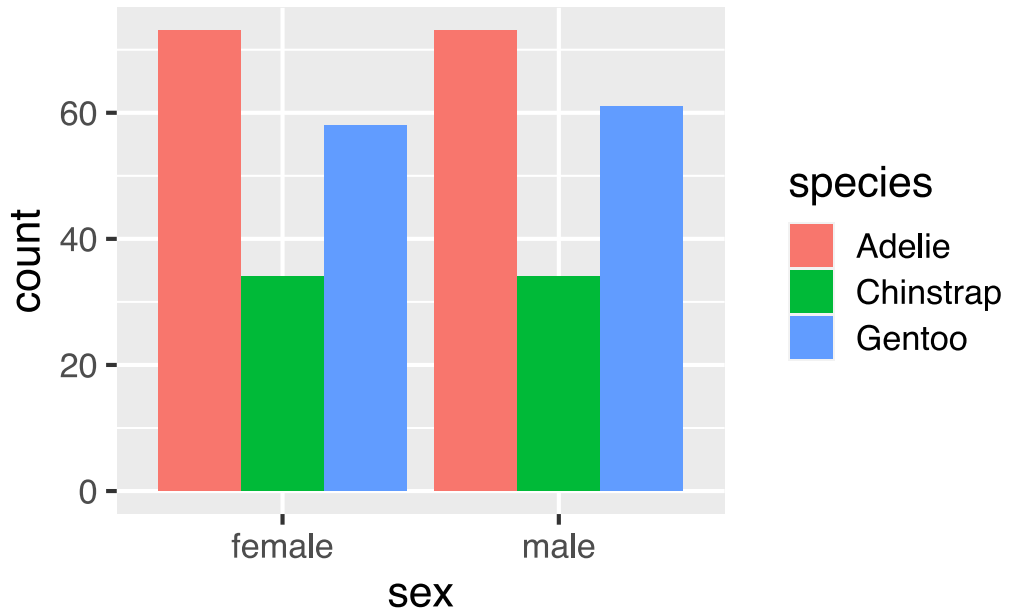
```
penguins_nomissing <- na.omit(penguins) # remove todas as linhas com valores missing

ggplot(penguins_nomissing, aes(sex, fill = species)) +
  geom_bar()
```



Positions define como os sub-grupos são mostrados

`position = "dodge"`: coloca as barras para subgroups side-by-side

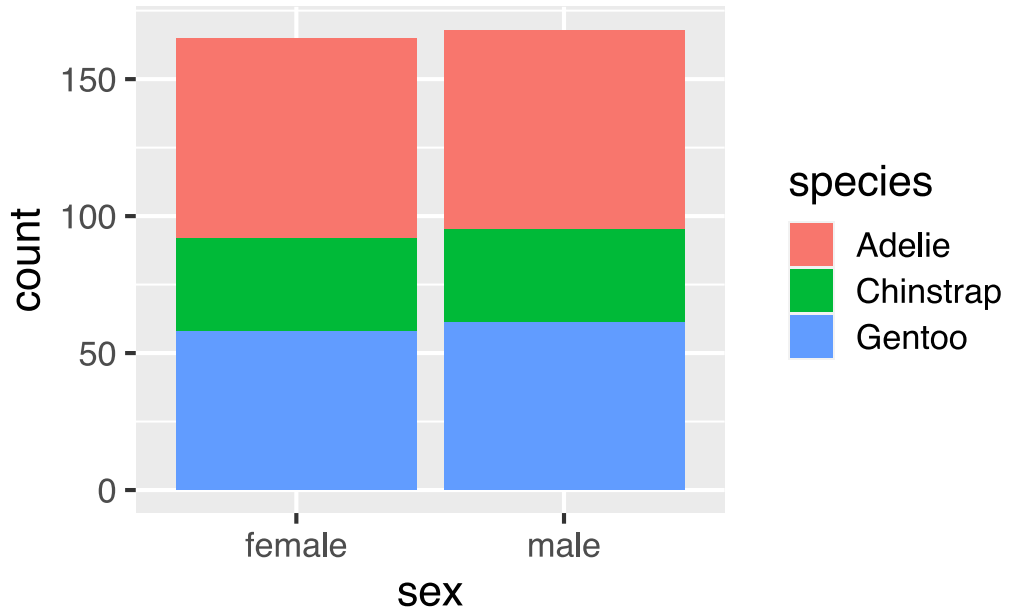
```
ggplot(penguins_nomissing, aes(sex, fill = species)) +
  geom_bar(position = "dodge")
```



Positions define como subgrupos são mostrados

`position = "stack"`: coloca barras dos subgroups no topo uma das outras

```
ggplot(penguins_nomissing, aes(sex, fill = species)) +
  geom_bar(position = "stack")
```



Positions define como subgrupos são mostrados

`position = "fill"`: como `"stack"`, mas a escala é de 100%

```
ggplot(penguins_nomissing, aes(sex, fill = species)) +
  geom_bar(position = "fill")
```