

# 1. Data visualization

2024-07-10

```
library(ggplot2)
library(ggthemes)
```

```
custom_ggplot <- function(...) {
  ggplot(...) +
    scale_color_colorblind()
}
```

```
theme_set(theme_few())
set.seed(123)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v lubridate  1.9.3      v tibble    3.2.1
## v purrr      1.0.2      v tidyr     1.3.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(palmerpenguins)
```

```
glimpse(penguins)
```

```
## Rows: 344
```

```
## Columns: 8
```

```
## $ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie,
```

```
## $ island        <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgersen,
```

```
## $ 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, ~
```

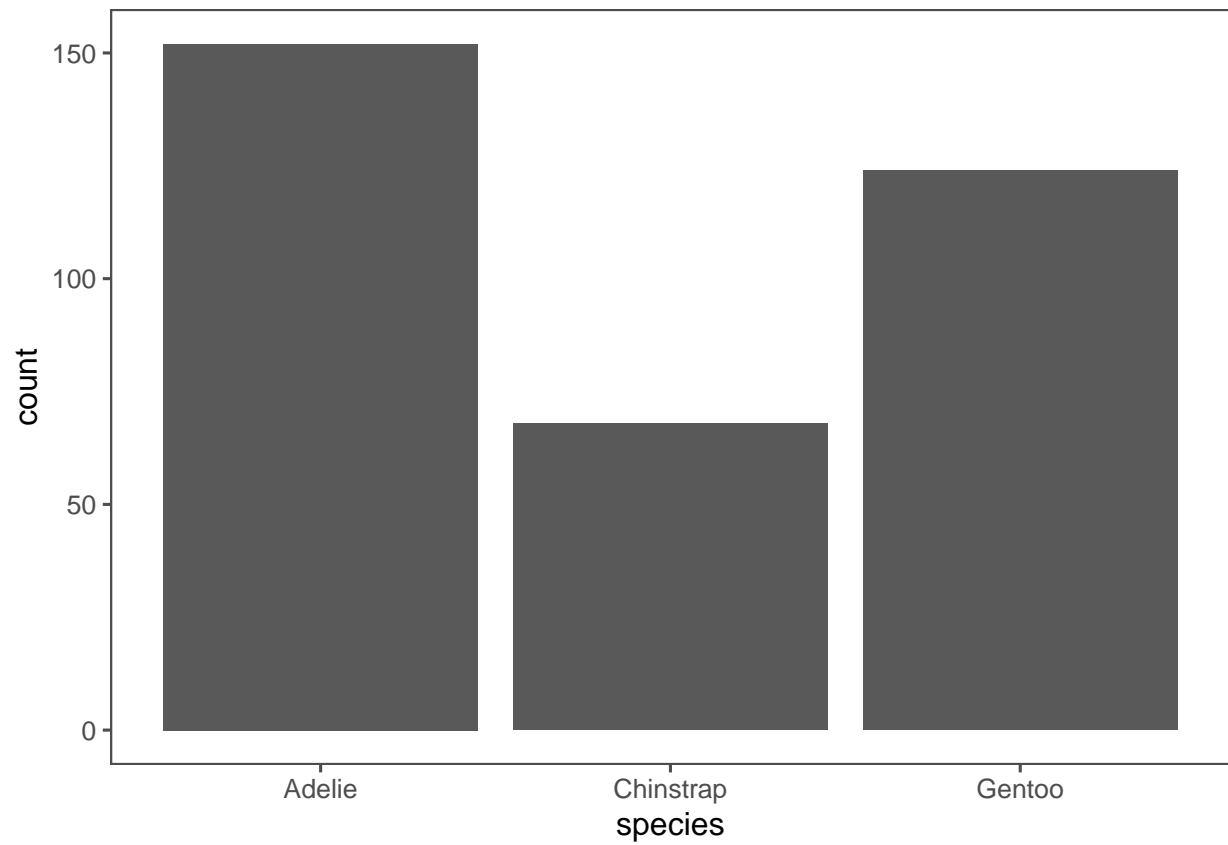
```
## $ sex            <fct> male, female, female, NA, female, male, female, male~
```

```
## $ year           <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007~
```

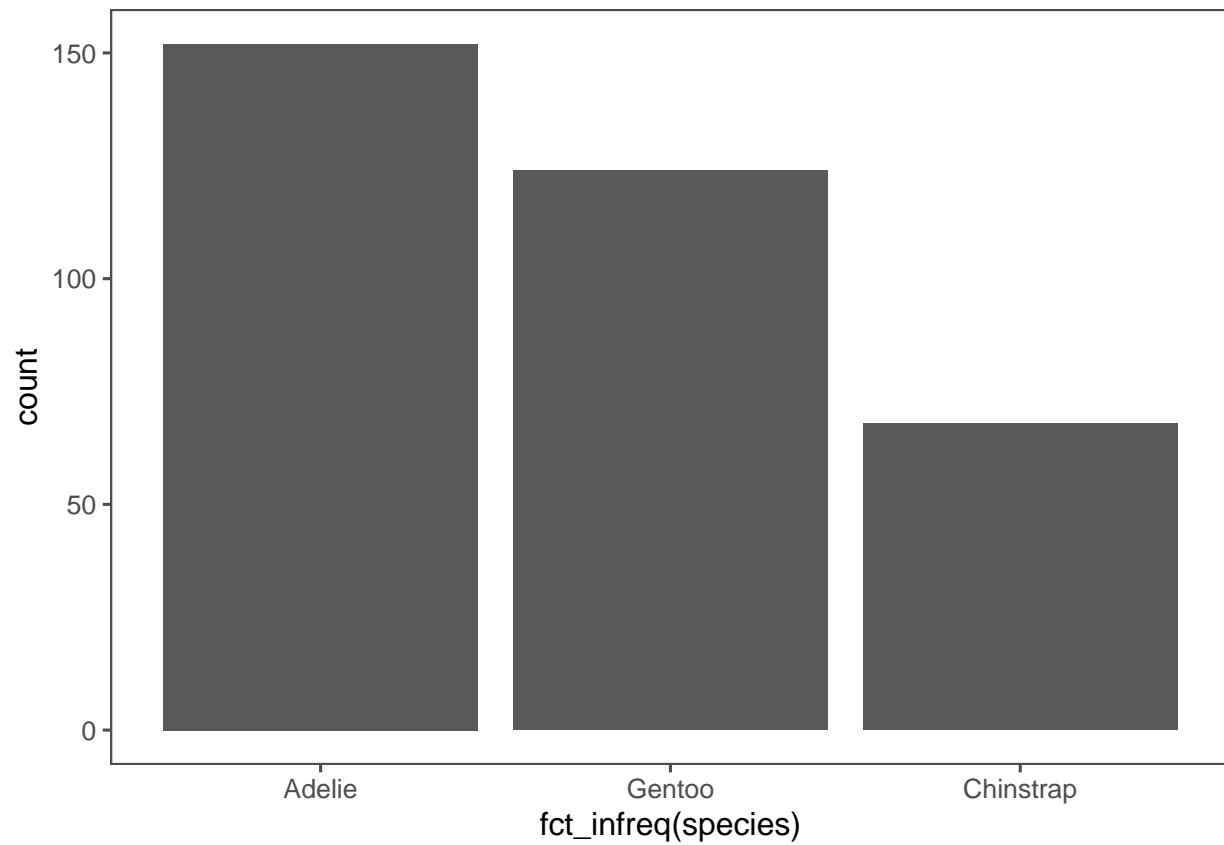
## 1.4 Visualizing distributions

### A categorical variable

```
ggplot(penguins, aes(x = species)) +
  geom_bar()
```



```
ggplot(penguins, aes(x = fct_infreq(species))) +  
  geom_bar()
```



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

```
## Warning: Removed 2 rows containing non-finite outside the scale range  
## (`stat_bin()`).
```

