

# Dimensions of penguins

Monika B

## Introduction

This is a first project fully executed in R with support from RStudio and Quarto.

Here, we'll look into analyzing the dimensions of Adelie, Chinstrap and Gentoo penguins from the original [data set](#).



Figure 1: Penguins ahoy!

## Let's get the party started!

We should start with importing all necessary packages for the code, which should include tidyverse and palmerpenguins .

```
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 ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- 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
```

```
library(palmerpenguins)
```

We will use the penguins dataset from the palmerpenguins package.

## Species

To plot the species, load the dataset and configure ggplot:

```
penguins
```

```
# A tibble: 344 x 8
  species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>         <dbl>         <dbl>           <int>         <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
6 Adelie Torgersen      39.3           20.6             190           3650
7 Adelie Torgersen      38.9           17.8             181           3625
8 Adelie Torgersen      39.2           19.6             195           4675
9 Adelie Torgersen      34.1           18.1             193           3475
10 Adelie Torgersen      42            20.2             190           4250
# i 334 more rows
# i 2 more variables: sex <fct>, year <int>
```

Type in the glimpse command to list the available columns:

```
glimpse(penguins)
```

```
Rows: 344
Columns: 8
$ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel~
$ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse~
$ 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~
```

It seems that the columns

- species
- flipper\_length\_mm
- body\_mass\_g

are exactly what we need for the visualisation!

## First plot in R!

Now that we know what we want, customise the ggplot command to visualize the scatter plot (Figure 2):

```
ggplot(
  data = penguins,
  mapping = aes(x = flipper_length_mm, y = body_mass_g)
) +
  geom_point(aes(color = species, shape = species)) +
  geom_smooth(method = "lm") +
  labs(
    title = "Body mass and flipper length",
    subtitle = "Dimensions for Adelie, Chinstrap, and Gentoo Penguins",
    x = "Flipper length (mm)", y = "Body mass (g)",
    color = "Species", shape = "Species"
  )
```

`geom\_smooth()` using formula = 'y ~ x'

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

```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_point()`).
```

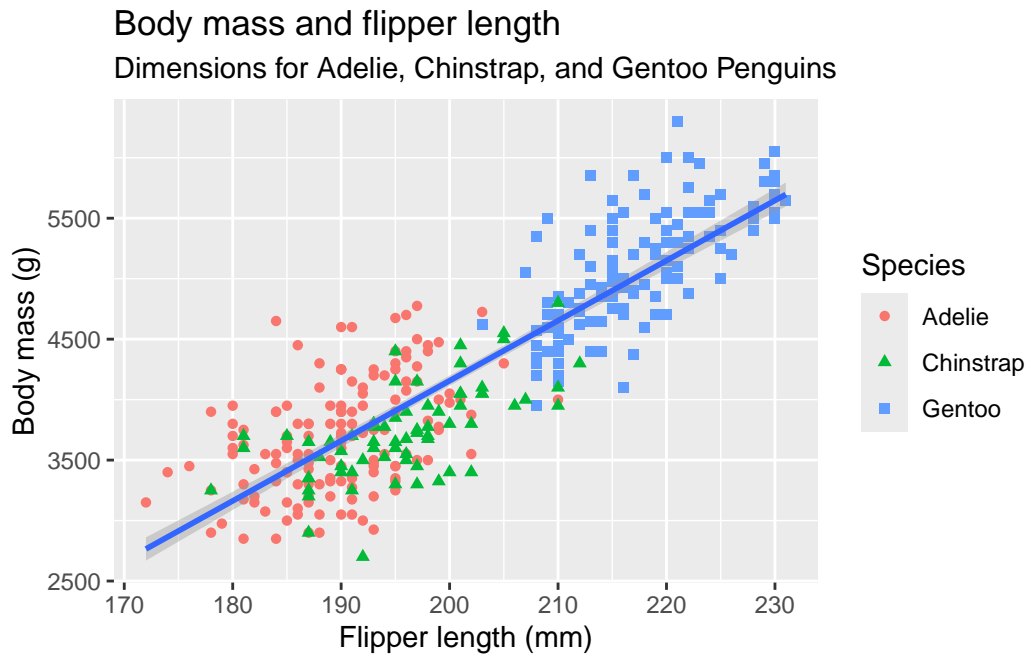


Figure 2: Scatter plot of the relationship between the body mass and the flipper length

## Summary

From here, we can summarise that the [Gentoo penguins](#) are heavier and bigger, therefore their flippers are longer, whereas the [Adelie penguins](#) are smaller with short flippers.

**Very cute!**