

R Take-Home Assignment 2: Palmer Penguins (Functions & Visualization)

Your Name

2025-10-01

Overview

In this assignment you will explore the **Palmer Penguins dataset**, which contains measurements for three penguin species living in the Palmer Archipelago in Antarctica. Researchers collected data on their bill size, flipper length, body mass, sex, and island of origin. The dataset is a favourite in data science because it provides a simple but realistic setting to practise exploring biological variation.

You will use this dataset to practise creating functions in R and visualising data in meaningful ways. Along the way, you will think like a biologist: Do larger penguins weigh more? Do species differ in size? And can we summarise this variation in a reusable way?

Preparation

You will need: `dplyr`, `ggplot2`, `palmerpenguins`.

```
library(dplyr)
library(ggplot2)
library(palmerpenguins)
```

1. Load and inspect the data

Before beginning any analysis, the first step is always to bring the data into R and take a first look. By knowing how many rows we have, which species are included, and what variables are measured, we can plan the kinds of questions we might answer.

Task: Load the `penguins` dataset from the `palmerpenguins` package. Then:

- Report the number of rows in the dataset.
- List the unique species included.
- Identify which variables are numeric.

```
data(penguins)
nrow(penguins)
```

```
## [1] 344
```

```
unique(penguins$species)
```

```
## [1] Adelie    Gentoo    Chinstrap
## Levels: Adelie Chinstrap Gentoo
```

```
sapply(penguins, is.numeric)
```

```
##           species           island  bill_length_mm  bill_depth_mm
##           FALSE           FALSE             TRUE             TRUE
```

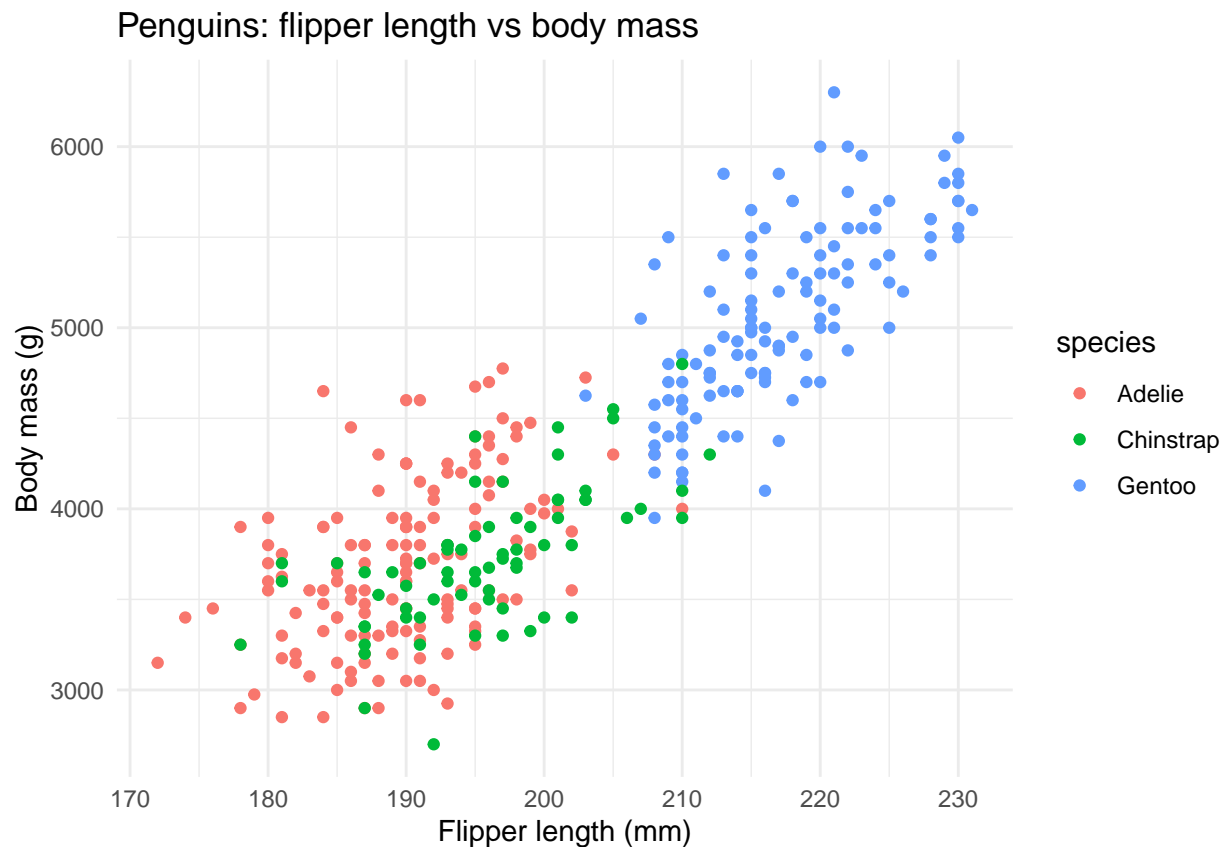
```
## flipper_length_mm    body_mass_g    sex    year
##                   TRUE          TRUE    FALSE  TRUE
```

2. Plot flipper length vs body mass

Biologists often wonder whether certain traits scale together. For penguins, larger flippers might be expected to support a heavier body. If this is true, we should see a clear relationship between flipper length and body mass. By colouring the points by species, we can also see whether the relationship holds consistently across Adelie, Chinstrap, and Gentoo penguins.

Task: Make a scatterplot of `flipper_length_mm` (x-axis) against `body_mass_g` (y-axis). Colour the points by species. Add axis labels and a title.

```
ggplot(penguins, aes(x=flipper_length_mm, y=body_mass_g, color=species)) +
  geom_point() +
  labs(x="Flipper length (mm)", y="Body mass (g)", title="Penguins: flipper length vs body mass") +
  theme_minimal()
```



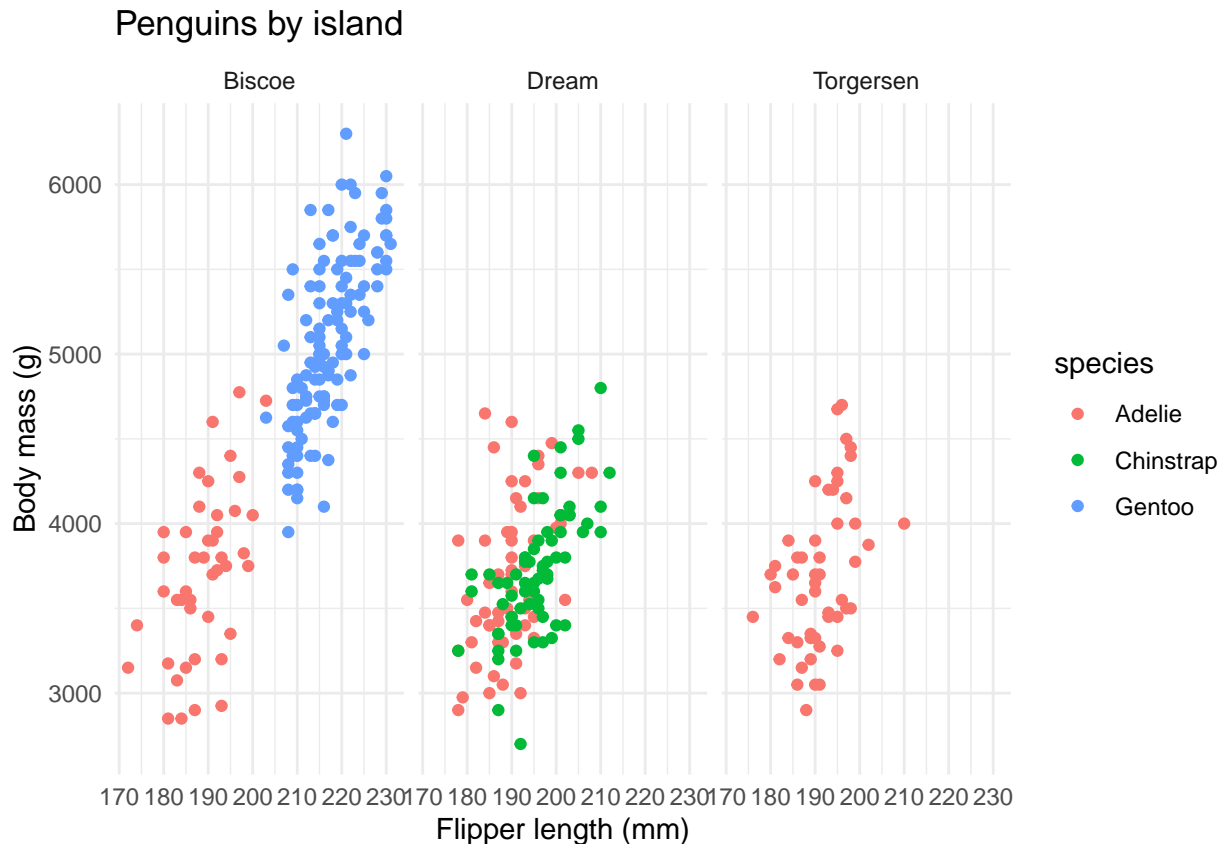
3. Facet by island

Species sometimes live on different islands, and environmental conditions (like food availability) can influence body size. To check whether penguins differ not just by species but also by location, we can break the plot into separate panels for each island.

Task: Recreate the scatterplot from Question 2, but facet it by `island` so that each island has its own panel.

```
ggplot(penguins, aes(x=flipper_length_mm, y=body_mass_g, color=species)) +
  geom_point() +
```

```
facet_wrap(~ island) +
labs(x="Flipper length (mm)", y="Body mass (g)", title="Penguins by island") +
theme_minimal()
```



4. Write a summary function

Analyses are more powerful when we can repeat them easily. Suppose we want to compare average bill length, flipper length, and body mass for each species. Instead of writing separate code each time, we can create a function that computes these summaries for any species we provide.

Task: Write a function called `species_summary(sp)` that takes a species name as input and returns:

- the mean bill length (mm),
- the mean flipper length (mm),
- the mean body mass (g).

```
species_summary <- function(sp){
  penguins %>%
    filter(species == sp) %>%
    summarise(mean_bill = mean(bill_length_mm, na.rm=TRUE),
              mean_flipper = mean(flipper_length_mm, na.rm=TRUE),
              mean_mass = mean(body_mass_g, na.rm=TRUE))
  ) %>% mutate(species = sp) %>% select(species, everything())
}
species_summary("Adelie")
```

```
## # A tibble: 1 x 4
##   species mean_bill mean_flipper mean_mass
```

```
##      <chr>      <dbl>      <dbl>      <dbl>
## 1 Adelie      38.8        190.       3701.
```

5. Apply your function

Now that you have written your function, you can quickly generate summaries for all species and combine them into one table. This gives a compact overview of how species differ.

Task: Apply your `species_summary()` function to all three species in the dataset. Combine the results into a single data frame with one row per species.

```
bind_rows(lapply(unique(penguins$species), species_summary))
```

```
## # A tibble: 3 x 4
##   species mean_bill mean_flipper mean_mass
##   <fct>      <dbl>      <dbl>      <dbl>
## 1 Adelie      38.8        190.       3701.
## 2 Gentoo      47.5        217.       5076.
## 3 Chinstrap  48.8        196.       3733.
```

6. Reflection

Finally, think about what these analyses reveal. Did larger flippers usually go hand in hand with heavier bodies? Were differences between species clear? Did location appear to matter?

Task: Write 3–4 sentences describing the main species differences and any island effects you observed.

Write your reflection here...