

# Exploring the Palmer Penguins Dataset

Daniel Perez

2026-01-14

## Introduction



In this report, we explore the Palmer Penguins data-set, which contains measurements of penguins from three species observed on islands in Antarctica.

We will:

- Inspect the structure of the data
- Summarize key variables
- Visualize GPA-like outcomes using a boxplot (body mass by species)

## Data preview

First 6 rows

```
## # A tibble: 6 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
## # i 2 more variables: sex <fct>, year <int>
```

## Data Wrangling

Some variables contain missing values.

For this analysis, we will focus on observations with non-missing body mass.

```
## # A tibble: 6 x 2
##   species body_mass_g
##   <fct>      <int>
## 1 Adelie      3750
## 2 Adelie      3800
## 3 Adelie      3250
## 4 Adelie      3450
## 5 Adelie      3650
## 6 Adelie      3625
```

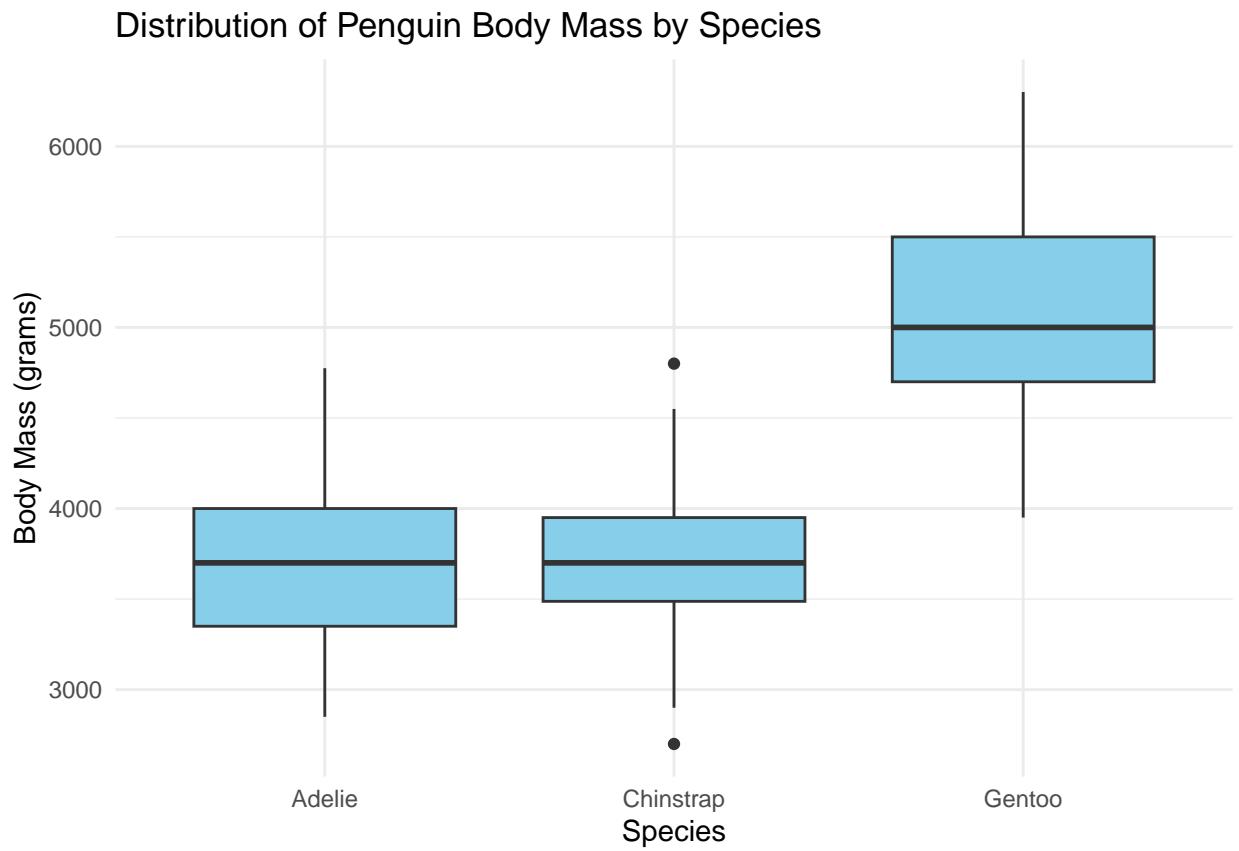
## Summary Statistics

Let's compute the average body mass by species.

```
## # A tibble: 3 x 4
##   species mean_body_mass sd_body_mass individuals
##   <fct>      <dbl>      <dbl>      <int>
## 1 Adelie      3701.        459.        151
## 2 Chinstrap   3733.        384.         68
## 3 Gentoo      5076.        504.        123
```

## Visualization: Body Mass by Species

We now visualize the distribution of body mass for each penguin species using a **boxplot**.



### Message to communicate:

Gentoo species have highest body mass.

*What else?*