Homework 1

STAT40830-Adv Data Prog with R

Elise Landman (22205684)

2024-06-15

Introduction

This analysis explores the **relationship between penguin species** and flipper length using the *palmerpenguins* dataset which is available as part of the base R package *palmerpenguins*. The dataset contains measurements of various penguin species collected at Palmer Station, Antarctica.

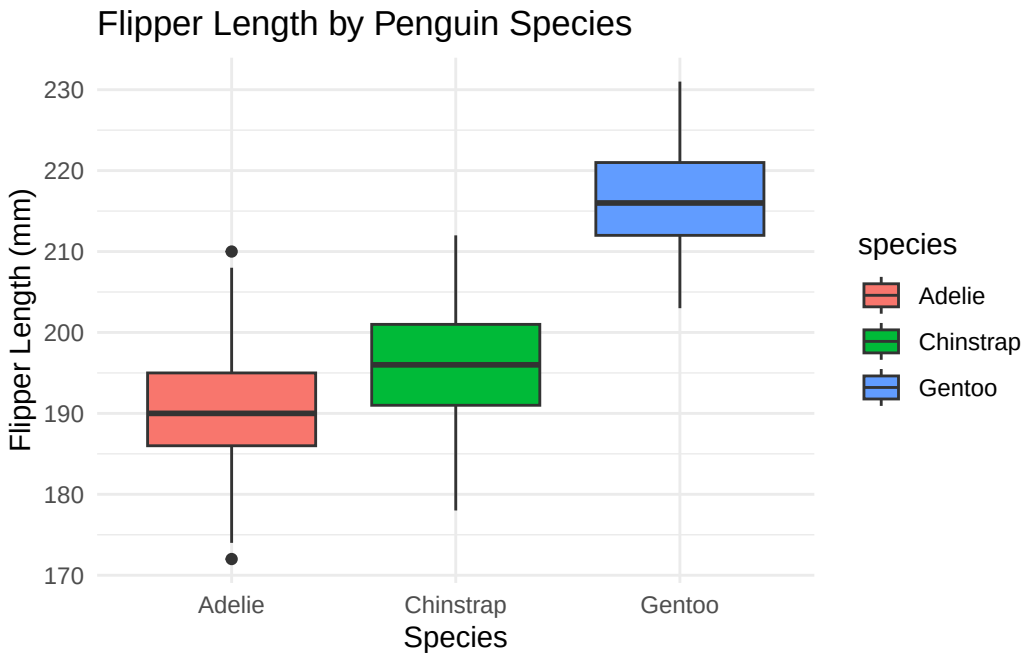
Below the first 6 rows of the dataset are displayed:

```
# 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>
```

We can see that the dataset contains 8 columns.

Data and Visualization

We can plot the data and visually analyze it using the R package *ggplot2*:



The above graph provides insights into the differences and similarities in flipper lengths among the penguin species. It depicts three box plots, each representing the flipper length in milimeters for a specific penguin species.

We can observe that Gentoo penguins have the longest median flipper length, followed by Chinstrap and Adelie penguins. We can also observe that in the dataset, only the Adalie Species contain outliers.