
Palmer Penguins Analysis Project

Ahmed Kamel

2025-12-02

1. Project Introduction 🐧 This project explores the Palmer Penguins dataset to analyze the relationship between various physical characteristics of penguin species using R. We will focus on species distribution, sexual dimorphism, and correlation between body measurements.

2. Loading Libraries & Data First, we load the necessary libraries for visualization and analysis.

```
library(palmerpenguins)
library(ggplot2)
library(RColorBrewer)
library(pheatmap)

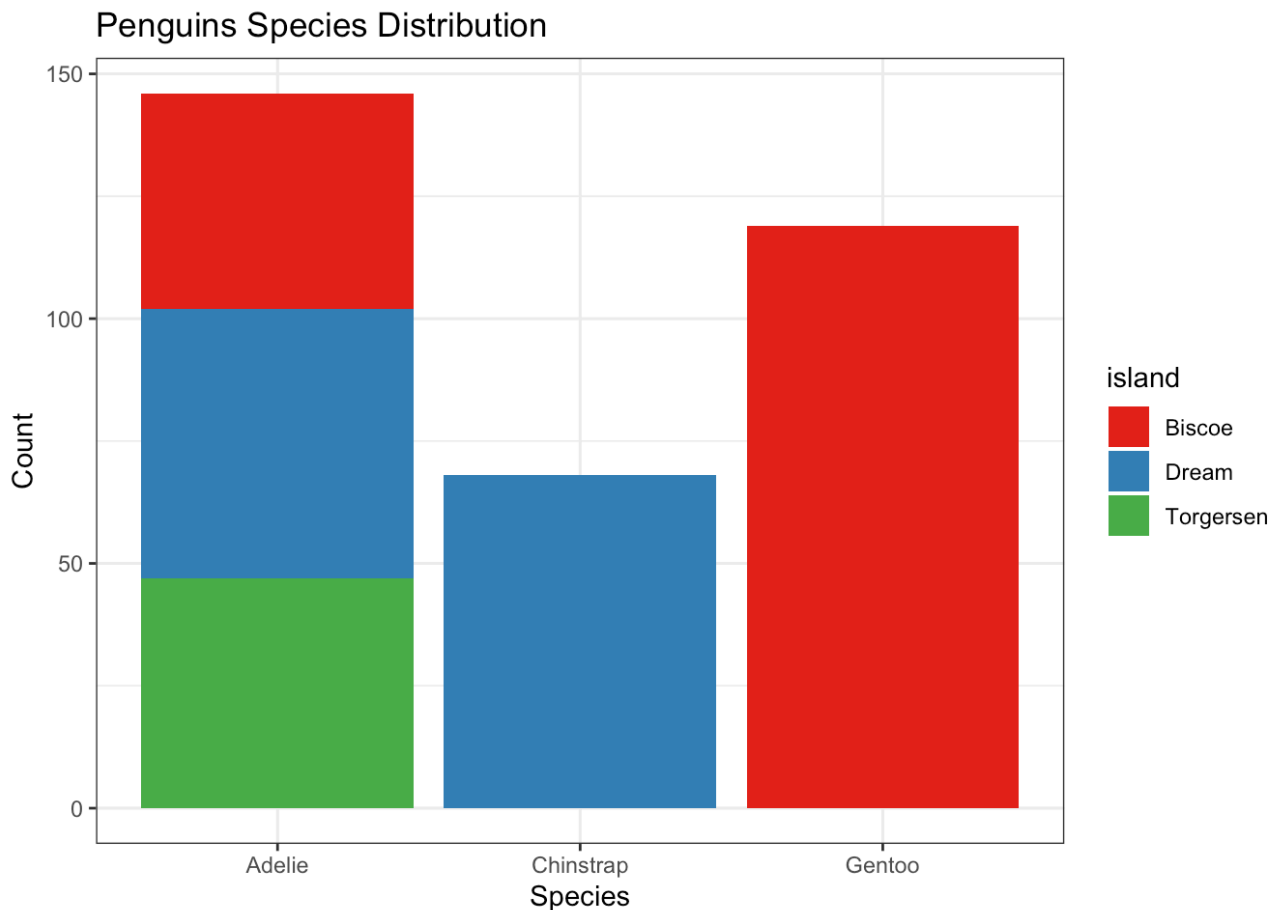
# Load and Inspect Data
Data = penguins
Data_clean = na.omit(Data)
head(Data_clean)
```

```
## # A tibble: 6 × 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         36.7          19.3           193
3450
## 5 Adelie  Torgersen         39.3          20.6           190
3650
## 6 Adelie  Torgersen         38.9          17.8           181
3625
## # i 2 more variables: sex <fct>, year <int>
```

3. Univariate Analysis

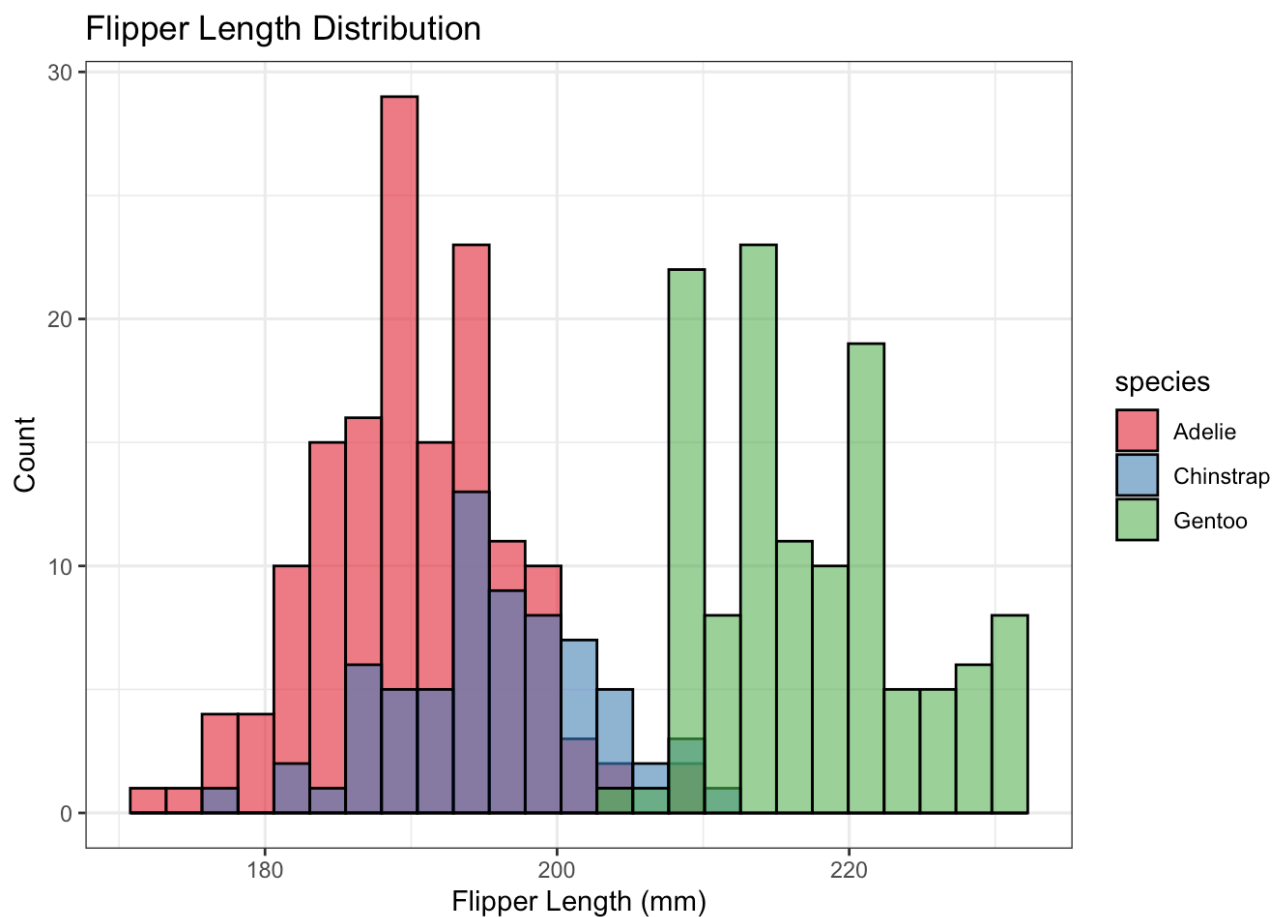
Species Distribution Which species is the most common across the islands?

```
ggplot(data = Data_clean, mapping = aes(x = species, fill = island)) +
  geom_bar() +
  theme_bw() +
  scale_fill_brewer(palette = "Set1") +
  xlab("Species") +
  ylab("Count") +
  ggtitle("Penguins Species Distribution")
```



Flipper Length Distribution Analyzing the spread of flipper lengths reveals distinct groups.

```
ggplot(data = Data_clean, mapping = aes(x = flipper_length_mm, fill = species)) +
  geom_histogram(bins = 25, color = "black", alpha = 0.6, position = "identity") +
  theme_bw() +
  scale_fill_brewer(palette = "Set1") +
  xlab("Flipper Length (mm)") +
  ylab("Count") +
  ggtitle("Flipper Length Distribution")
```

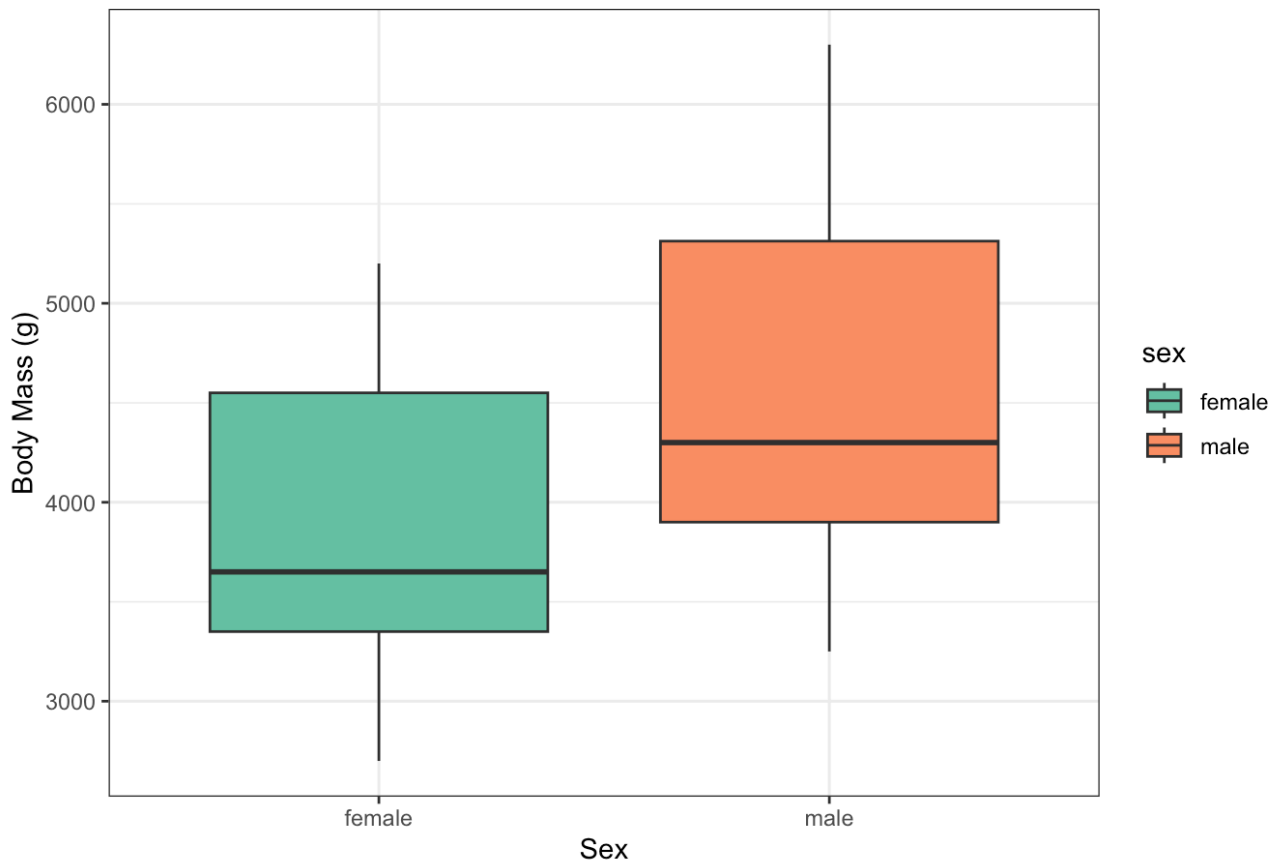


4. Bivariate Analysis

Sexual Dimorphism (Body Mass) Is there a significant difference in mass between male and female penguins?

```
ggplot(data = Data_clean, mapping = aes(x = sex, y = body_mass_g, fill = sex)) +
  geom_boxplot() +
  theme_bw() +
  scale_fill_brewer(palette = "Set2") +
  xlab("Sex") +
  ylab("Body Mass (g)") +
  ggtitle("Body Mass Comparison by Sex")
```

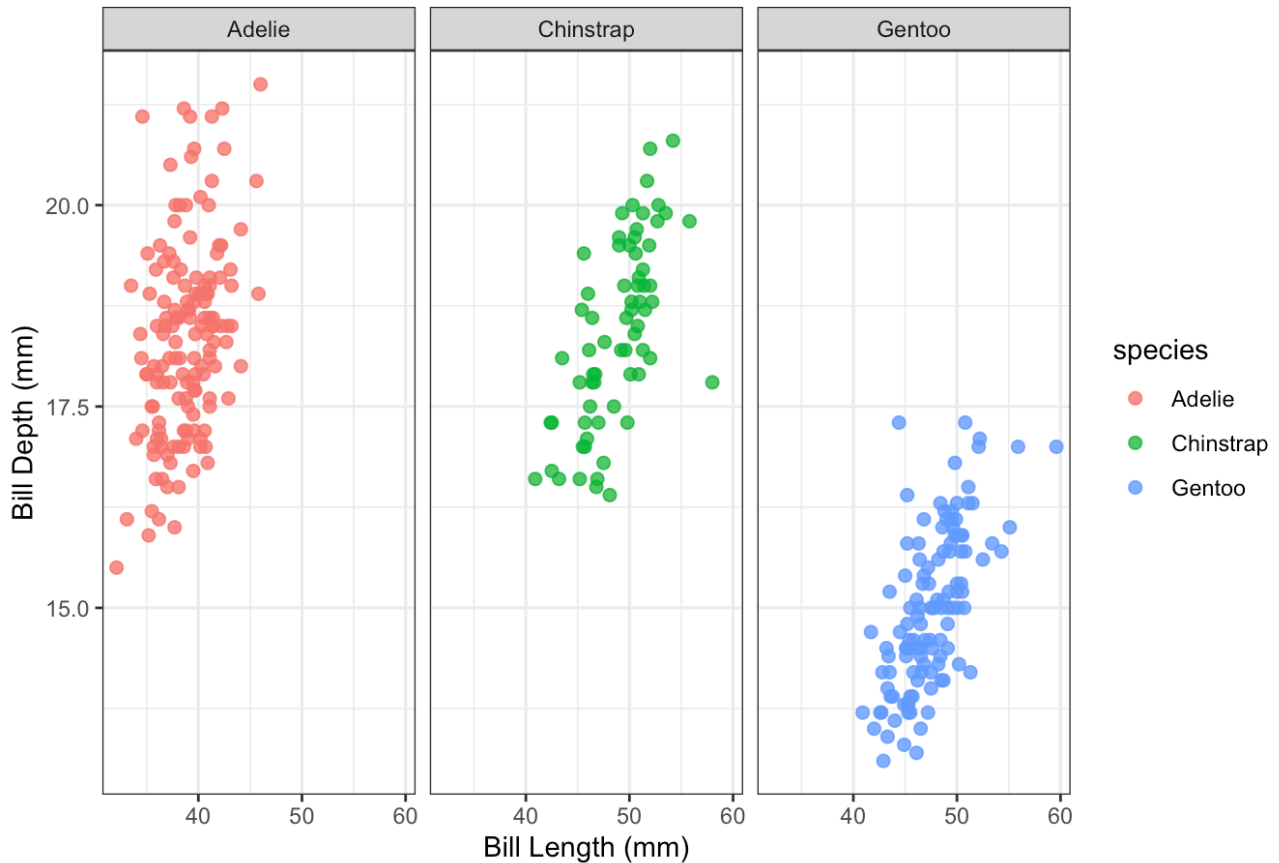
Body Mass Comparison by Sex



Bill Dimensions (Simpson's Paradox) Investigating the relationship between bill length and depth.

```
ggplot(data = Data_clean, mapping = aes(x = bill_length_mm, y = bill_depth_mm, colour = species)) +  
  geom_point(size = 2, alpha = 0.8) +  
  theme_bw() +  
  xlab("Bill Length (mm)") +  
  ylab("Bill Depth (mm)") +  
  ggtitle("Relationship Between Bill Length & Bill Depth") +  
  facet_grid(~species)
```

Relationship Between Bill Length & Bill Depth



5. Multivariate Analysis (Correlation) Finally, we look at the correlation between all numeric variables.

```
# Prepare Data
numeric_col <- Data_clean[, c("bill_length_mm", "bill_depth_mm",
                              "flipper_length_mm", "body_mass_g")]
cor_matrix <- cor(numeric_col)

# Plot Heatmap
pheatmap(mat = cor_matrix,
         display_numbers = TRUE,
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         color = colorRampPalette(brewer.pal(n = 7, name = "RdBu"))(100),
         main = "Correlation Heatmap")
```

Correlation Heatmap

