Lab Exercise

R V Abhishek

2025-09-23

Data Visualisation with the Palmer penguins Dataset

Variables in the dataset: bill_length, bill_depth, flipper_length, body_mass, species, island, sex

```
# Load necessary libraries
library(palmerpenguins)
```

```
##
## Attaching package: 'palmerpenguins'

## The following objects are masked from 'package:datasets':
##
##     penguins, penguins_raw
```

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

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

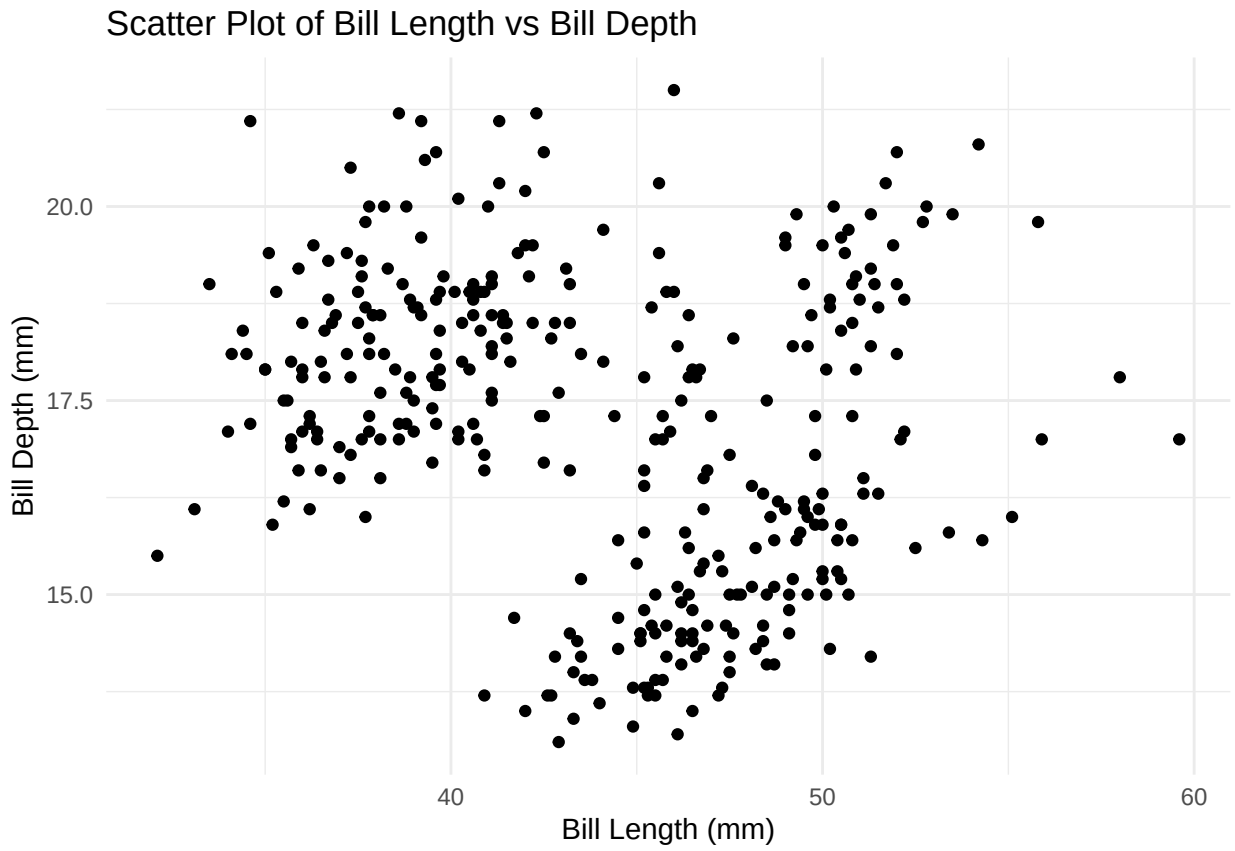
```
# Load the penguins dataset
data("penguins")

# Display the first few rows of the dataset
head(penguins)
```

```
## # 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>
```

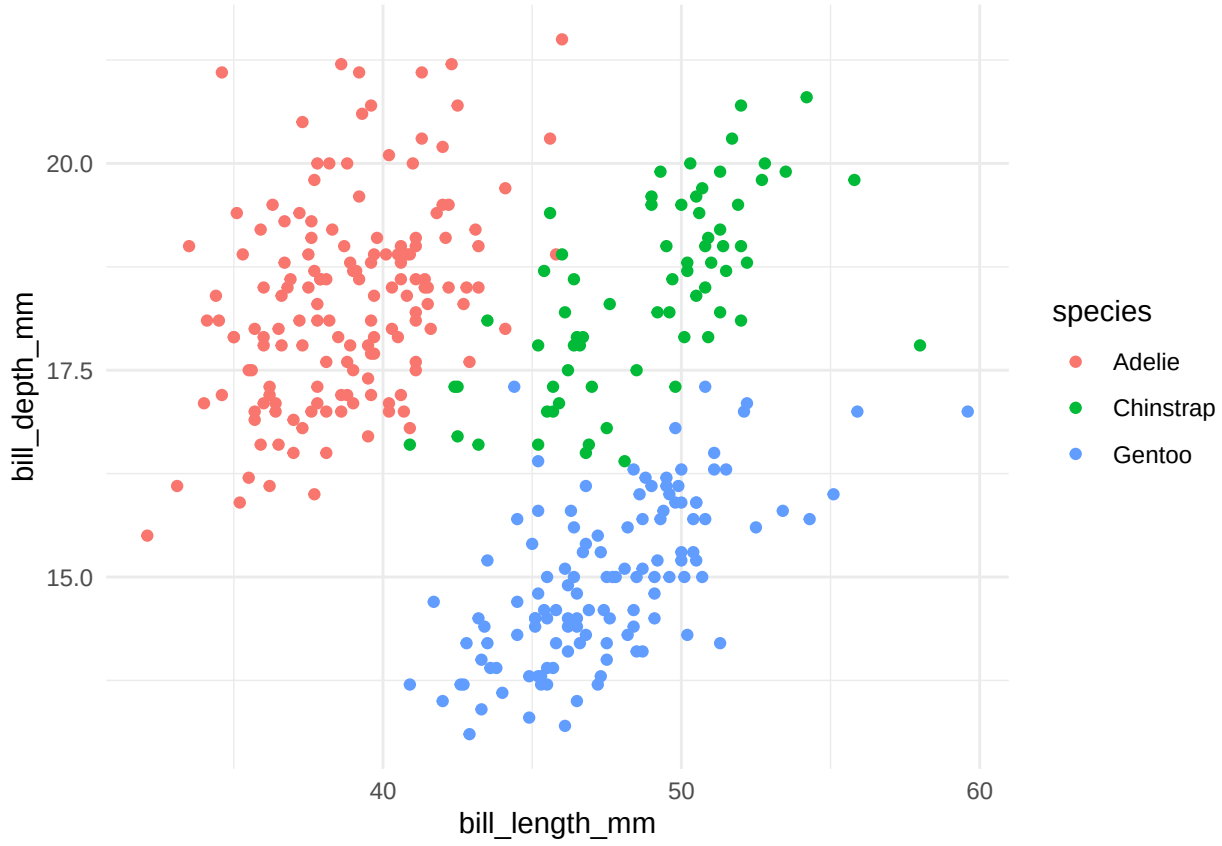
```r
# Q1. Scatter plot of bill.length.mm vs bill.depth.mm
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm)) +
  geom_point() +
  labs(title = "Scatter Plot of Bill Length vs Bill Depth",
       x = "Bill Length (mm)",
       y = "Bill Depth (mm)") +
  theme_minimal()
```

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



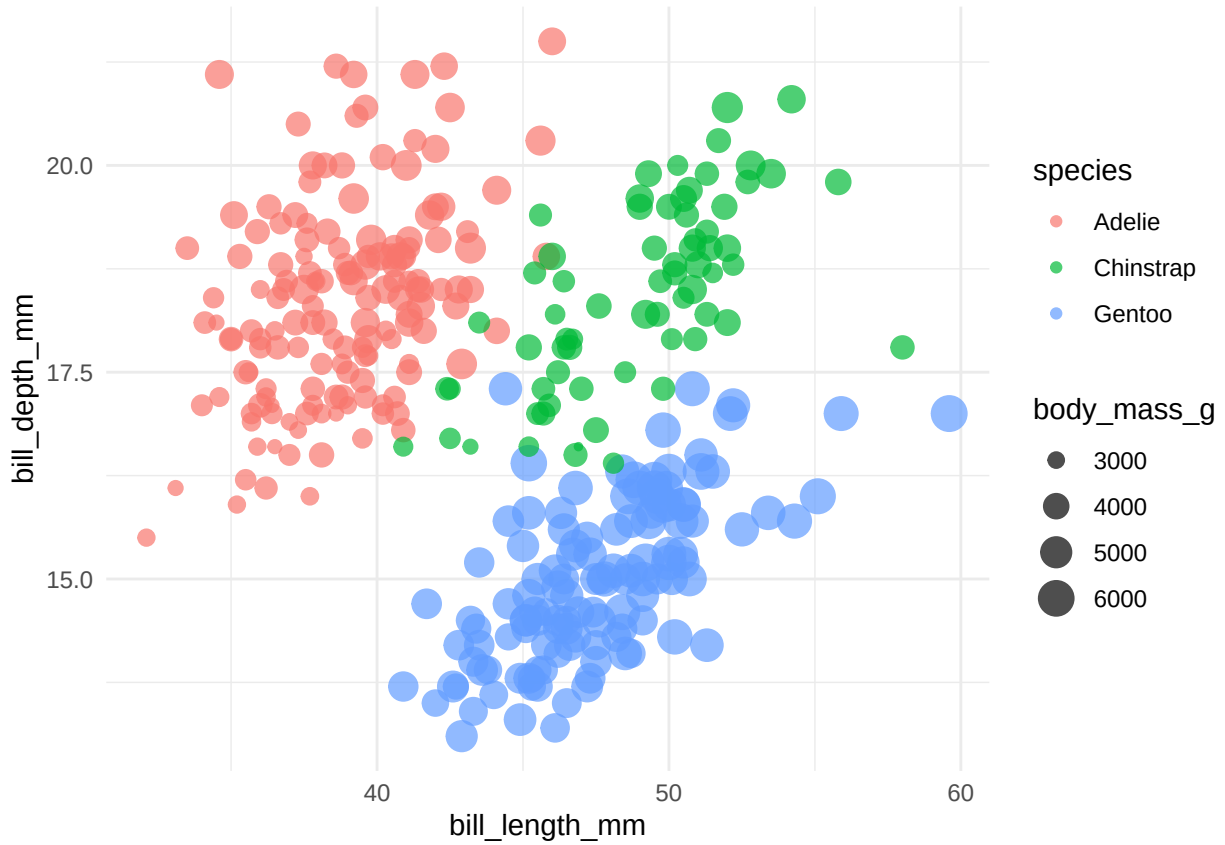

```r
# Q2. Scatter plot colored by species
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, color = species)) +
  geom_point() +
  theme_minimal()
```

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



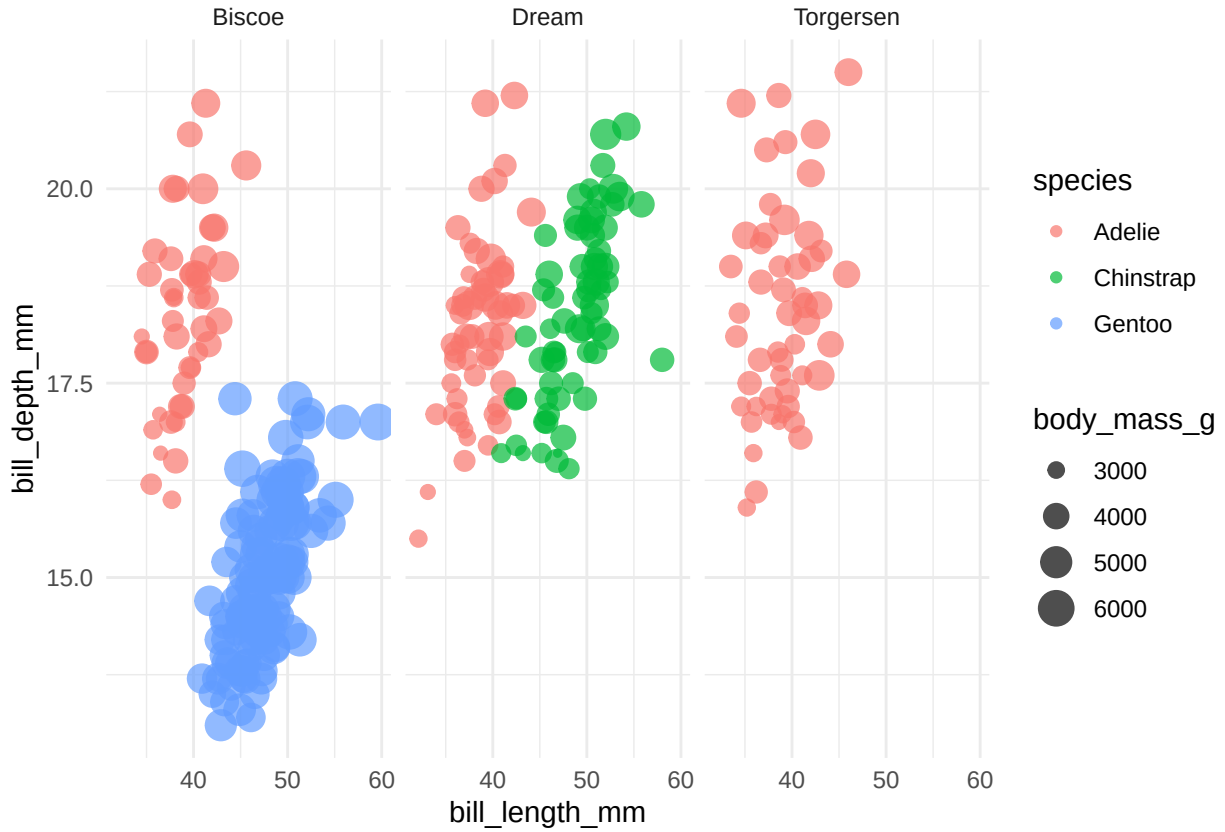

```
# Q3. Change point size based on body mass
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, color = species, size = body_mass_g)) +
  geom_point(alpha = 0.7) +
  theme_minimal()
```

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



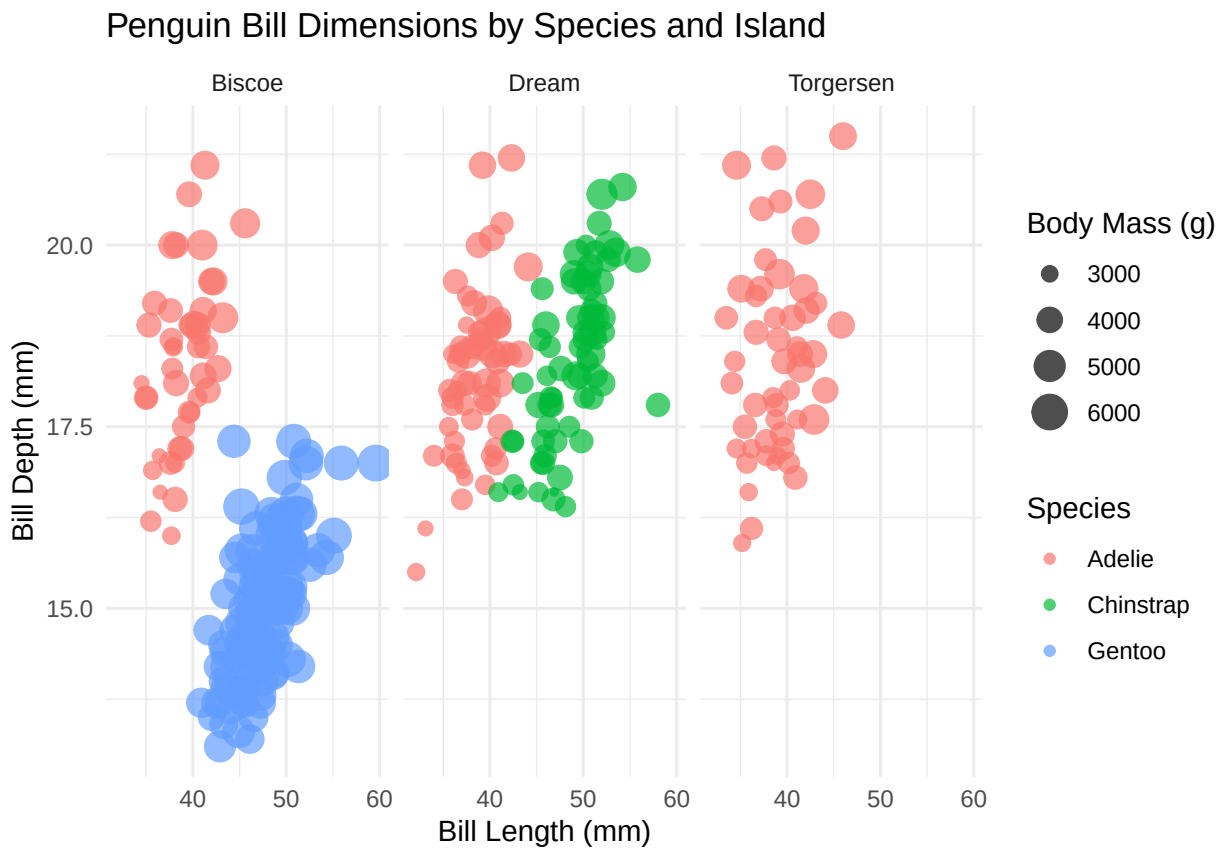

```
# Q4. Facet by island
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, color = species, size = body_mass_g)) +
  geom_point(alpha = 0.7) +
  theme_minimal() +
  facet_wrap(~ island)
```

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



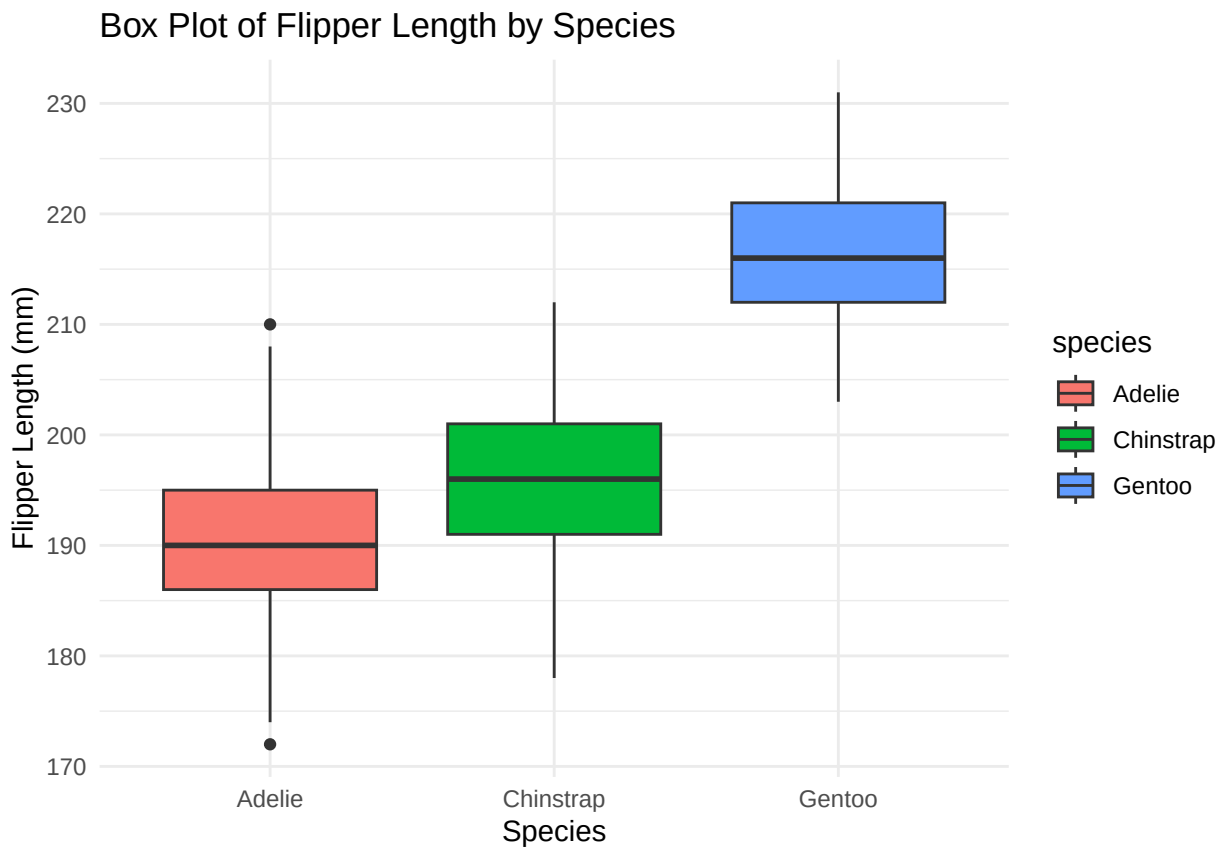

```
# Q5. Add axis title and meaningful labels
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, color = species, size = body_mass_g)) +
  geom_point(alpha = 0.7) +
  theme_minimal() +
  facet_wrap(~ island) +
  labs(title = "Penguin Bill Dimensions by Species and Island",
       x = "Bill Length (mm)",
       y = "Bill Depth (mm)",
       color = "Species",
       size = "Body Mass (g)")
```

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



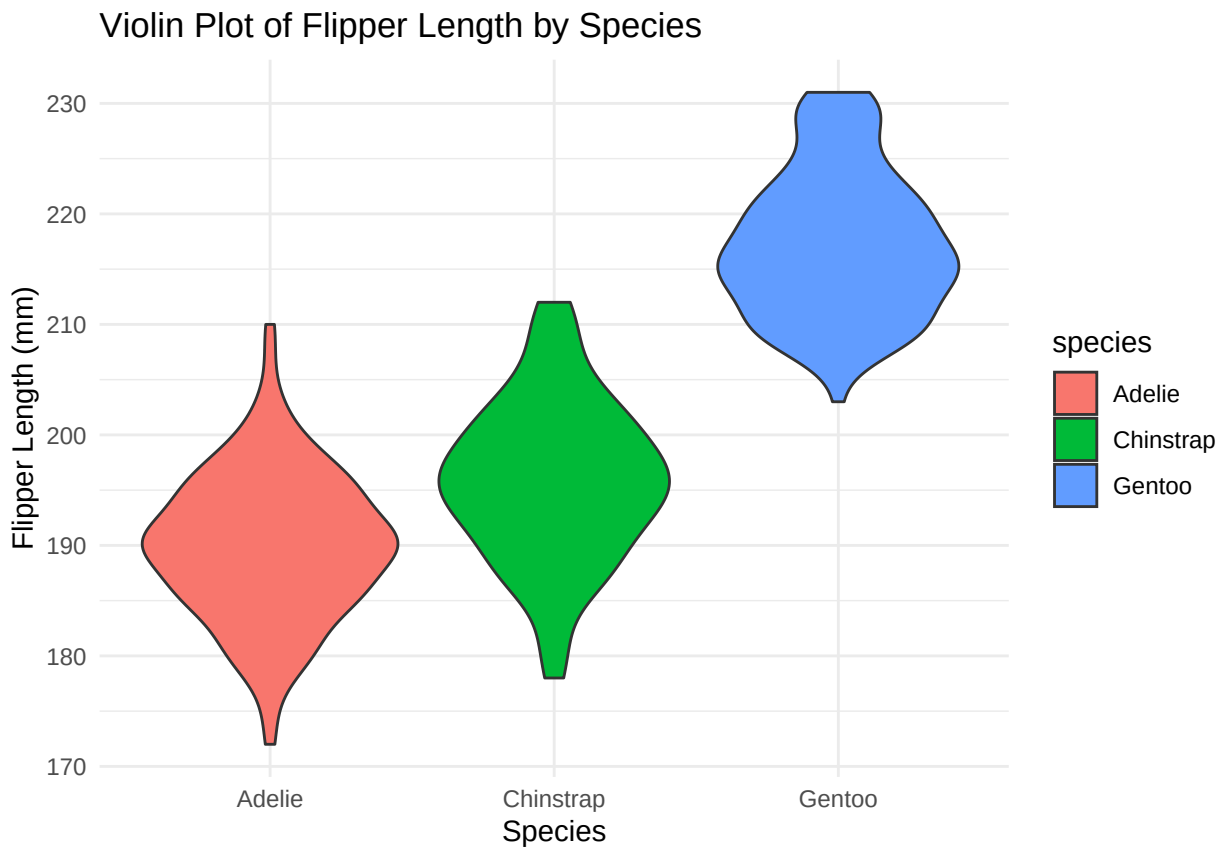

```
# Q6. Box plot of flipper length accross species
ggplot(penguins, aes(x = species, y = flipper_length_mm, fill = species)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Box Plot of Flipper Length by Species",
       x = "Species",
       y = "Flipper Length (mm)") +
  theme(legend)
```

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



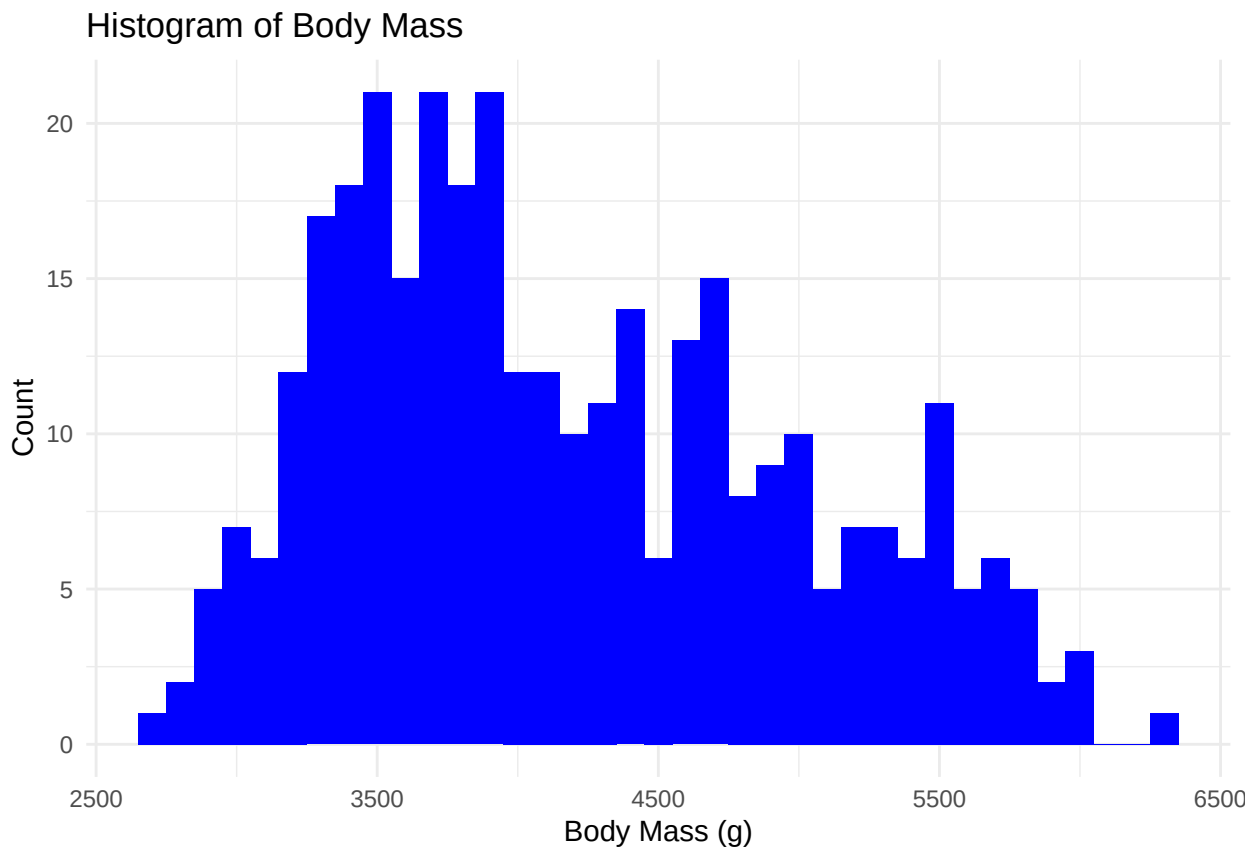

```
# Q7. Violin plot instead of box plot
ggplot(penguins, aes(x = species, y = flipper_length_mm, fill = species)) +
  geom_violin() +
  theme_minimal() +
  labs(title = "Violin Plot of Flipper Length by Species",
       x = "Species",
       y = "Flipper Length (mm)") +
  theme(legend)
```

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



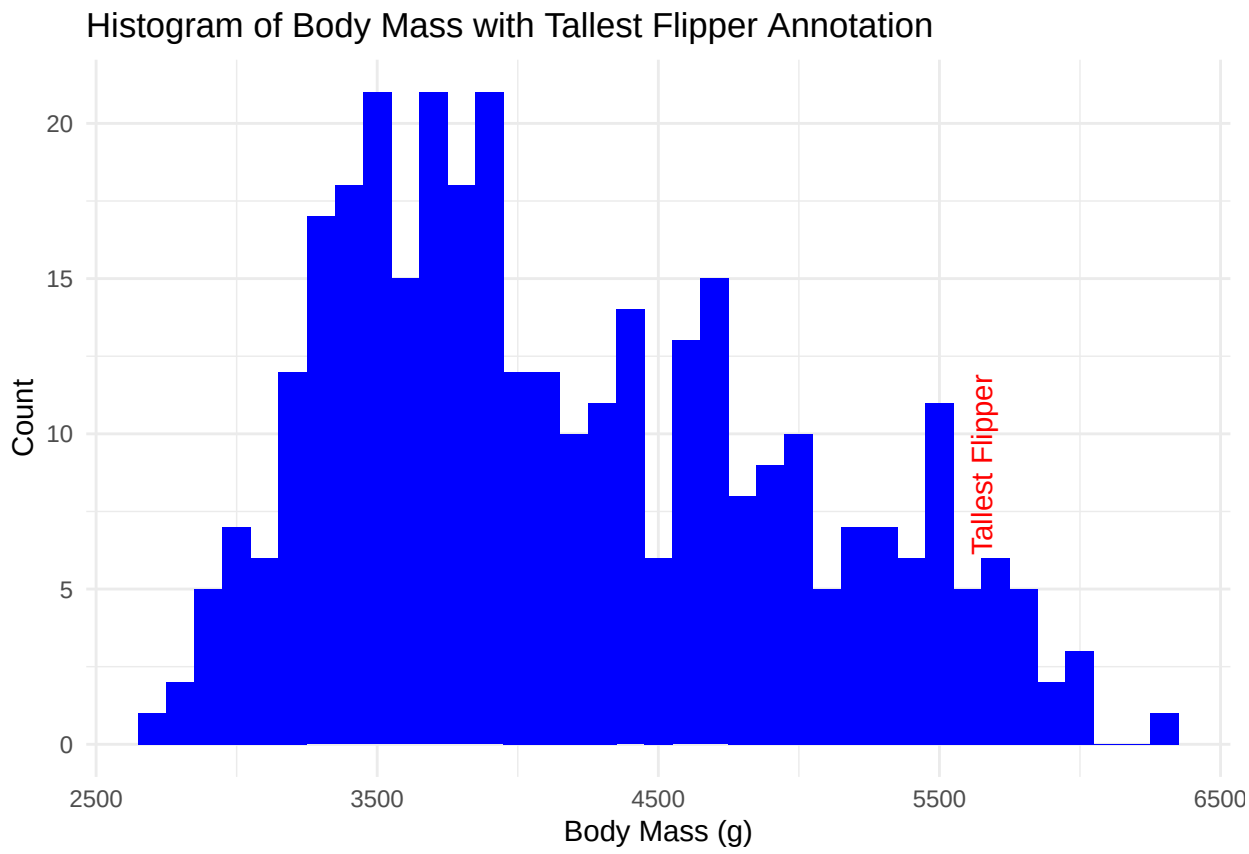

```
# Q8. Histogram of body mass
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 100, fill = "blue") +
  theme_minimal() +
  labs(title = "Histogram of Body Mass",
       x = "Body Mass (g)",
       y = "Count")
```

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




```
# Q9. Annotate tallest Penguin flipper range on the histogram
tallest_flipper <- penguins %>%
  filter(flipper_length_mm == max(flipper_length_mm, na.rm = TRUE)) %>%
  select(body_mass_g)
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 100, fill = "blue") +
  theme_minimal() +
  labs(title = "Histogram of Body Mass with Tallest Flipper Annotation",
       x = "Body Mass (g)",
       y = "Count") +
  annotate("text", x = tallest_flipper$body_mass_g, y = 9, label = "Tallest Flipper",
           angle = 90, color = "red")
```

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

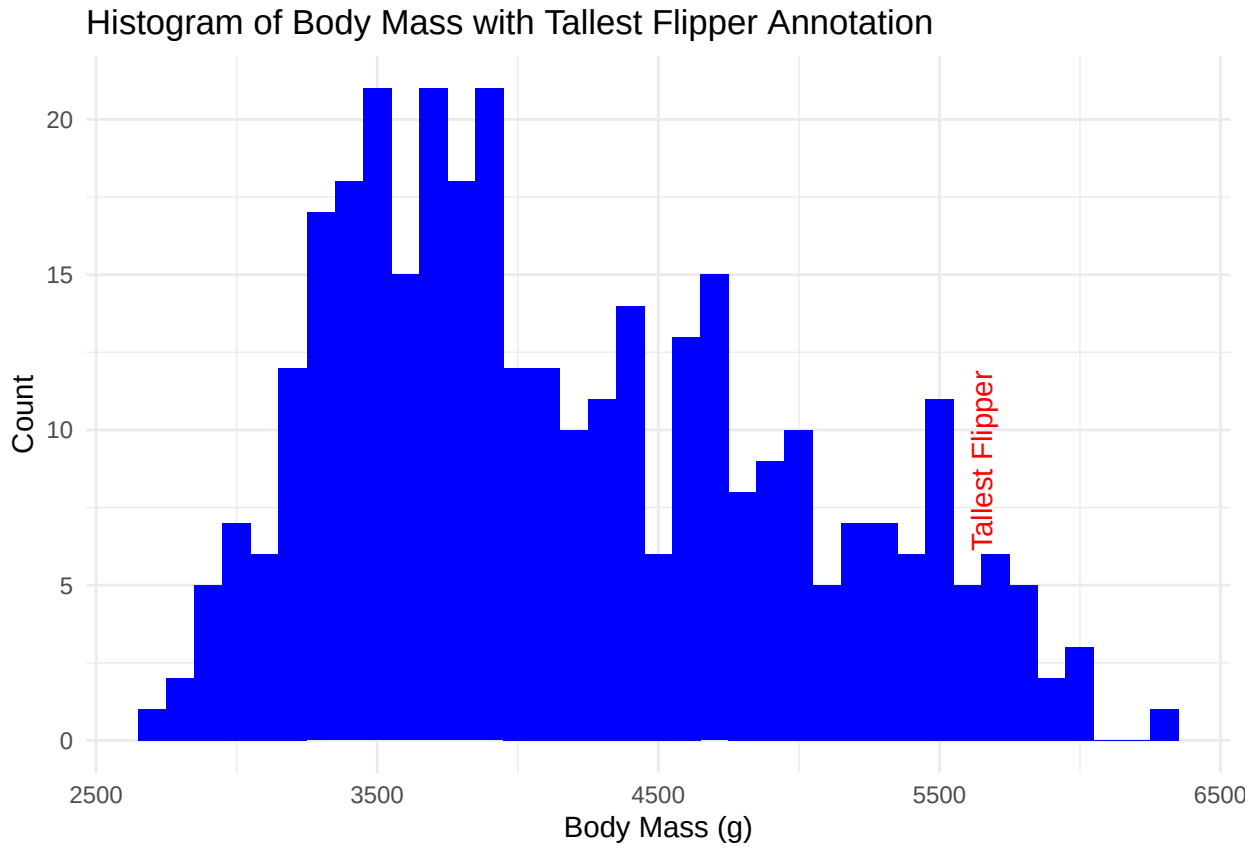



```
# Q10. Save with ggsave
plot <- ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm,
                             color = species, size = body_mass_g)) +
  geom_point(alpha = 0.7) +
  theme_minimal() +
  facet_wrap(~ island) +
  labs(title = "Penguin Bill Dimensions by Species and Island",
       x = "Bill Length (mm)",
       y = "Bill Depth (mm)",
       color = "Species",
       size = "Body Mass (g)")
ggsave("penguin_bill_dimensions.png", plot = plot, width = 10, height = 6)
```

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

```
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 100, fill = "blue") +
  theme_minimal() +
  labs(title = "Histogram of Body Mass with Tallest Flipper Annotation",
       x = "Body Mass (g)",
       y = "Count") +
  annotate("text", x = tallest_flipper$body_mass_g, y = 9, label = "Tallest Flipper",
           angle = 90, color = "red")
```

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




```
ggsave("penguin_body_mass_histogram.png", width = 8, height = 5)
```

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