Assignment 5

Can Binay

2025-11-10

Dataset

```
data("penguins")
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>
```

```
#Cleaned dataaset
clean_peng <- na.omit(penguins)
head(clean_peng)
```

```
## # 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           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>
```

The `penguins` dataset (available through the `palmerpenguins` package and can be loaded after installing and loading the package) contains real-world data collected from three penguin species are Adélie, Chinstrap, and Gentoo living in the Palmer Archipelago of Antarctica. The measurements were taken between 2007 and 2009 consists many variables about the penguins such as `species`, `island`, `bill_length_mm`, and so on.

Drawing a plot for association

1. Draw a single plot showing association between the bill lengths and body mass of the penguins according to the their sex (60 pts).

```
ggplot(clean_peng, aes(x = bill_length_mm, y = body_mass_g, colour = factor(sex))) +
  geom_point() +
```

```
labs(x = "Bill Length",
     y= "Body Mass",
     title = "Visualization of the association between bill length and body mass by penguins gender",
     colour = "Gender") +
scale_colour_discrete(labels = c("Female", "Male"))
```

Visualization of the association between bill length and body mass by peng
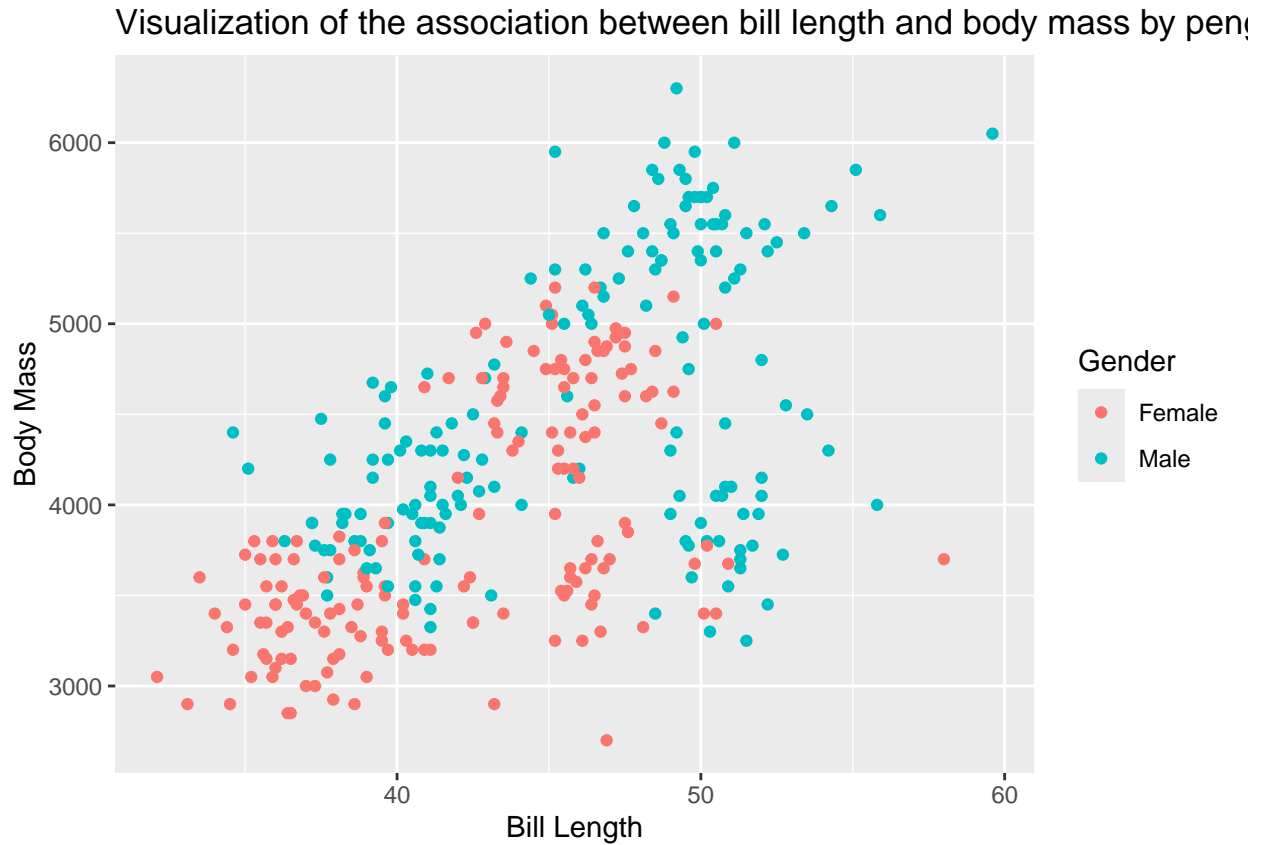
2. Interpret the plot (40 pts).

Interpretation

Penguin's body mass and penguin's bill length has a positive linear relationship by their sex.