

R project for analysis palmer penguins dataset

Shrouk 2024-03-11

Install and load the required packages

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —  
## ✓ dplyr      1.1.4      ✓ readr      2.1.5  
## ✓ forcats    1.0.0      ✓ stringr    1.5.1  
## ✓ ggplot2    3.5.0      ✓ tibble     3.2.1  
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1  
## ✓ purrr      1.0.2
```

```
## — Conflicts ————— tidyverse_conflicts() —  
## ✗ dplyr::filter() masks stats::filter()  
## ✗ dplyr::lag()     masks stats::lag()  
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
```

```
library(ggplot2)
```

Load the Palmer Penguins dataset

```
library(palmerpenguins)
data("penguins")
```

Display the first few rows of the dataset

```
head(penguins)
```

```
## # 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          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>
```

Summary statistics

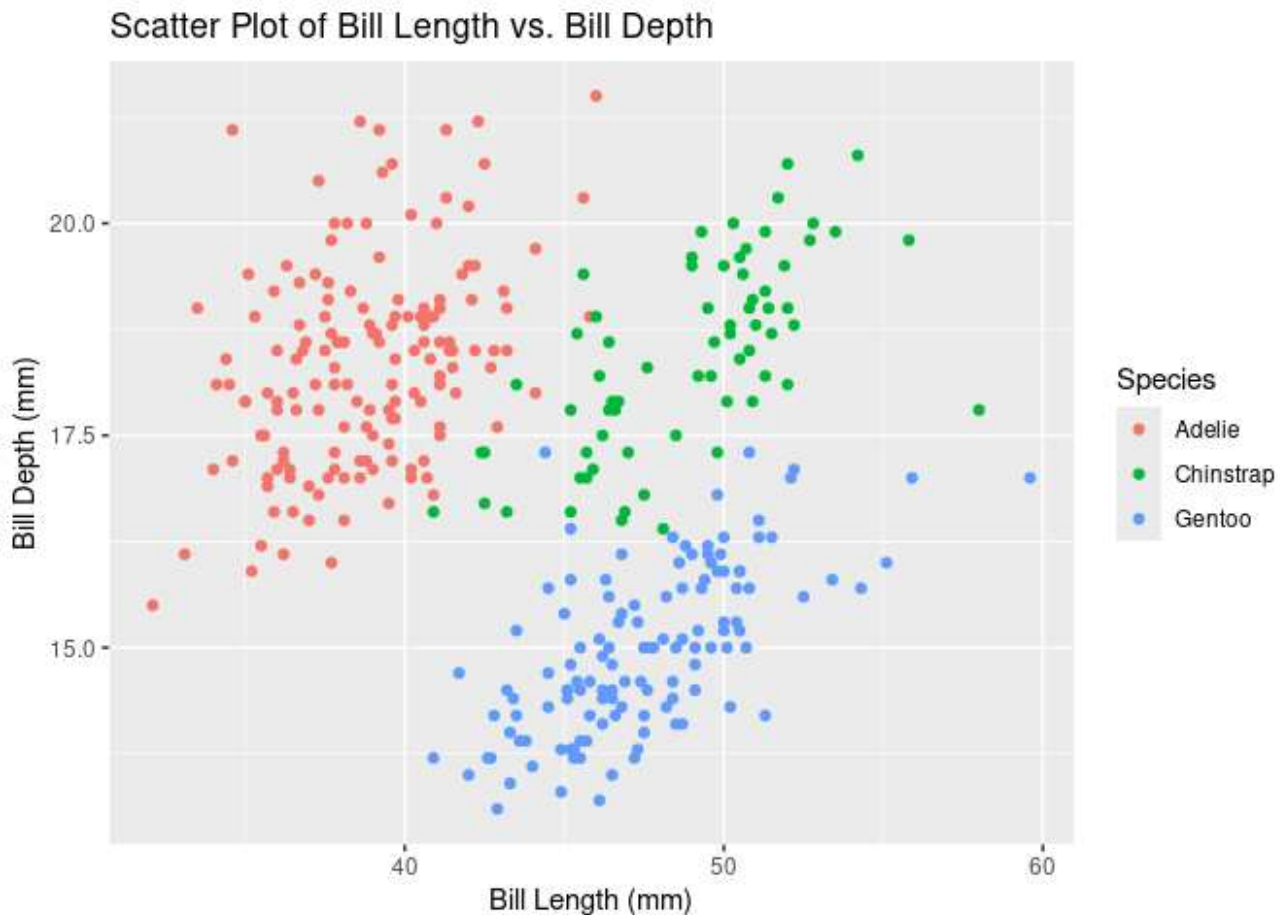
```
summary(penguins)
```

```
##      species      island  bill_length_mm  bill_depth_mm
##  Adelie    :152  Biscoe    :168  Min.      :32.10  Min.      :13.10
##  Chinstrap: 68  Dream     :124  1st Qu.:39.23  1st Qu.:15.60
##  Gentoo    :124  Torgersen: 52  Median :44.45  Median :17.30
##                                     Mean      :43.92  Mean      :17.15
##                                     3rd Qu.:48.50  3rd Qu.:18.70
##                                     Max.      :59.60  Max.      :21.50
##                                     NA's      :2      NA's      :2
## flipper_length_mm  body_mass_g      sex      year
##  Min.      :172.0    Min.      :2700  female:165  Min.      :2007
##  1st Qu.:190.0    1st Qu.:3550  male   :168  1st Qu.:2007
##  Median :197.0    Median :4050  NA's   : 11  Median :2008
##  Mean      :200.9    Mean      :4202                      Mean      :2008
##  3rd Qu.:213.0    3rd Qu.:4750                      3rd Qu.:2009
##  Max.      :231.0    Max.      :6300                      Max.      :2009
##  NA's      :2      NA's      :2
```

Visualization: Scatter plot of Bill Length and Bill Depth

```
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, color = species)) +  
  geom_point() +  
  labs(title = "Scatter Plot of Bill Length vs. Bill Depth",  
        x = "Bill Length (mm)",  
        y = "Bill Depth (mm)",  
        color = "Species")
```

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



Visualization: Boxplot of Flipper Length by Species

```
ggplot(penguins, aes(x = species, y = flipper_length_mm, fill = species)) +  
  geom_boxplot() +  
  labs(title = "Boxplot of Flipper Length by Species",  
        x = "Species",
```

```
y = "Flipper Length (mm)",  
fill = "Species")
```

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

