

Penguin Data Analysis

Raj Mishra

2025-06-21

Penguin Data Analysis

Loading Data

In this section, I load the `tidyverse` and `palmerpenguins` packages and use the built-in `penguins` dataset.

```
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.2      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(palmerpenguins)

data("penguins")
```

Exploring the Data

Here, I generate a summary and structure of the `penguins` dataset to understand the variables and any missing values.

```
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
```

```
glimpse(penguins)
```

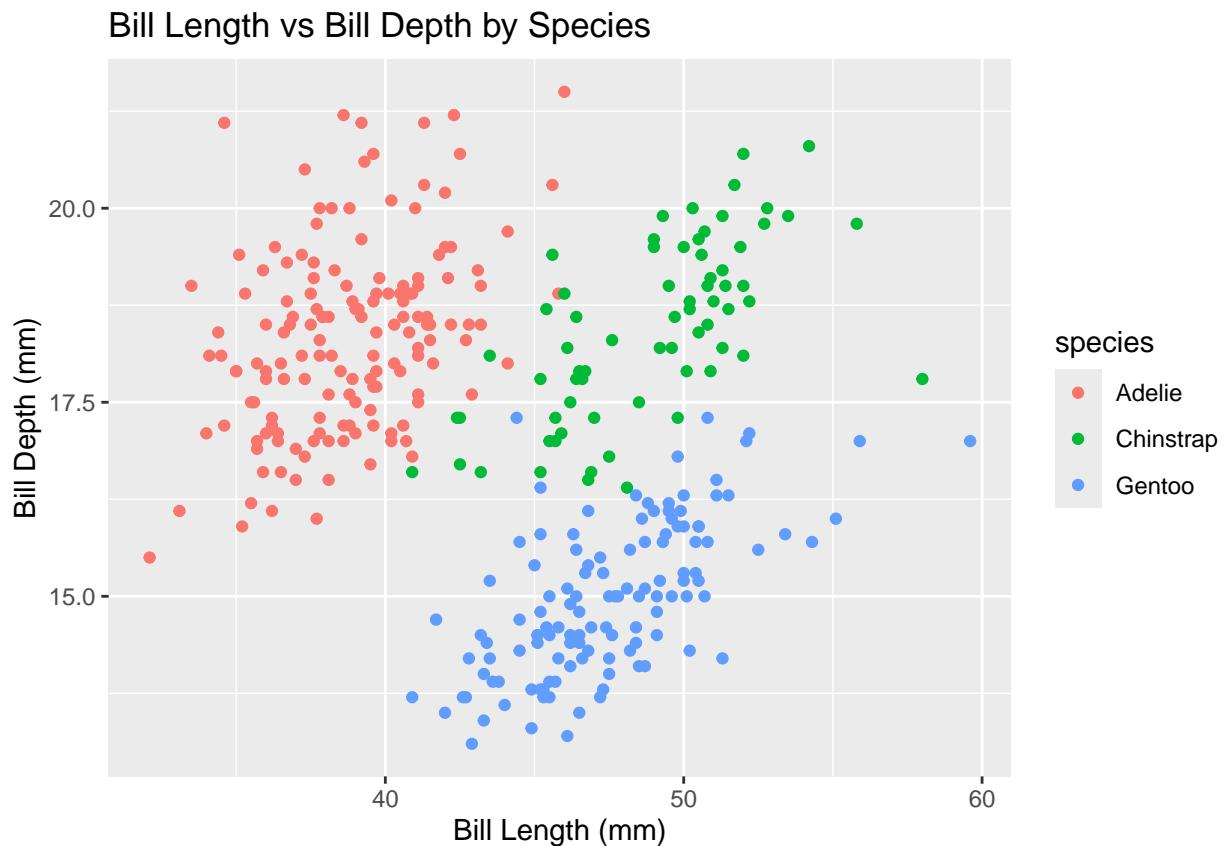
```
## Rows: 344
## Columns: 8
## $ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel~
## $ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse~
## $ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ~
## $ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ~
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186~
## $ body_mass_g   <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ~
## $ sex           <fct> male, female, female, NA, female, male, female, male~
## $ year          <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007~
```

Visualizing the Data

This scatterplot shows the relationship between bill length and bill depth across penguin species.

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

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



Conclusion

In this analysis, I explored the Palmer Penguins dataset and visualized the relationship between physical measurements. The scatterplot highlights how species differ in bill dimensions, which is useful for species classification.