Data Analysis

Dr. Muhammad Saufi

2024-08-14

Contents

# 1. Setting the Environment

## 1.1 Loading the Packages

# Load the packages into R session   
library(tidyverse)

Warning: package 'tidyverse' was built under R version 4.4.1

── 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 3.5.1 ✔ 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()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# 2. Dataset

## 2.1 Reading the Dataset

# Import the dataset into workspace   
penguins <- read\_csv("Datasets/penguins.csv")

Rows: 340 Columns: 9  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (3): species, island, sex  
dbl (6): rowid, bill\_length\_mm, bill\_depth\_mm, flipper\_length\_mm, body\_mass\_...  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## 2.2 Exploring the Dataset

# Get a quick overview of dataset   
glimpse(penguins)

Rows: 340  
Columns: 9  
$ rowid <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1…  
$ species <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie", "A…  
$ island <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen", …  
$ 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 <dbl> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186…  
$ body\_mass\_g <dbl> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, …  
$ sex <chr> "male", "female", "female", NA, "female", "male", "f…  
$ year <dbl> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007…

## 2.3 Visualize the Dataset

ggplot(  
data = penguins,  
mapping = aes(  
x = flipper\_length\_mm,  
y = body\_mass\_g,  
color = species)  
) +  
geom\_point()