Data Analysis

Dr. Muhammad Saufi

2024-08-14

Contents

1 Setting the Environment

1.1 Loading the Packages

```r
# 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 --
v dplyr     1.1.4      v readr     2.1.5
v forcats   1.0.0      v stringr   1.5.1
v ggplot2   3.5.1      v tibble    3.2.1
v lubridate 1.9.3      v tidyr     1.3.1
v purrr     1.0.2
-- 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 becom
```

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()
```

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