Tutorial 6: Refactoring R Code

Introduction

In this tutorial, you will refactor the code into separate scripts corresponding to each section. The dataset we will use comes from the `palmerpenguins` package, which contains measurements of penguins from three species.

Load Libraries and Data

Methods

In this section, we perform exploratory data analysis (EDA) and prepare the data for modeling.

From Table 1 and Table 2 we get a rough idea of the data.

Model

We will fit a classification model using `tidymodels` to predict the species of a penguin based on its physical characteristics.

Results

We evaluate the performance of the model using the test dataset in Table 3

Table 1: Glimpsed Summary

```
# A tibble: 333 x 8
   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
   <chr>   <chr>              <dbl>         <dbl>             <dbl>       <dbl>
 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
 7 Adelie  Torgersen           39.2          19.6               195        4675
 8 Adelie  Torgersen           41.1          17.6               182        3200
 9 Adelie  Torgersen           38.6          21.2               191        3800
10 Adelie  Torgersen           34.6          21.1               198        4400
# i 323 more rows
# i 2 more variables: sex <chr>, year <dbl>
```

Table 2: Summarized Summary

```
# A tibble: 1 x 2
  mean_bill_length mean_bill_depth
             <dbl>           <dbl>
1             44.0            17.2
```

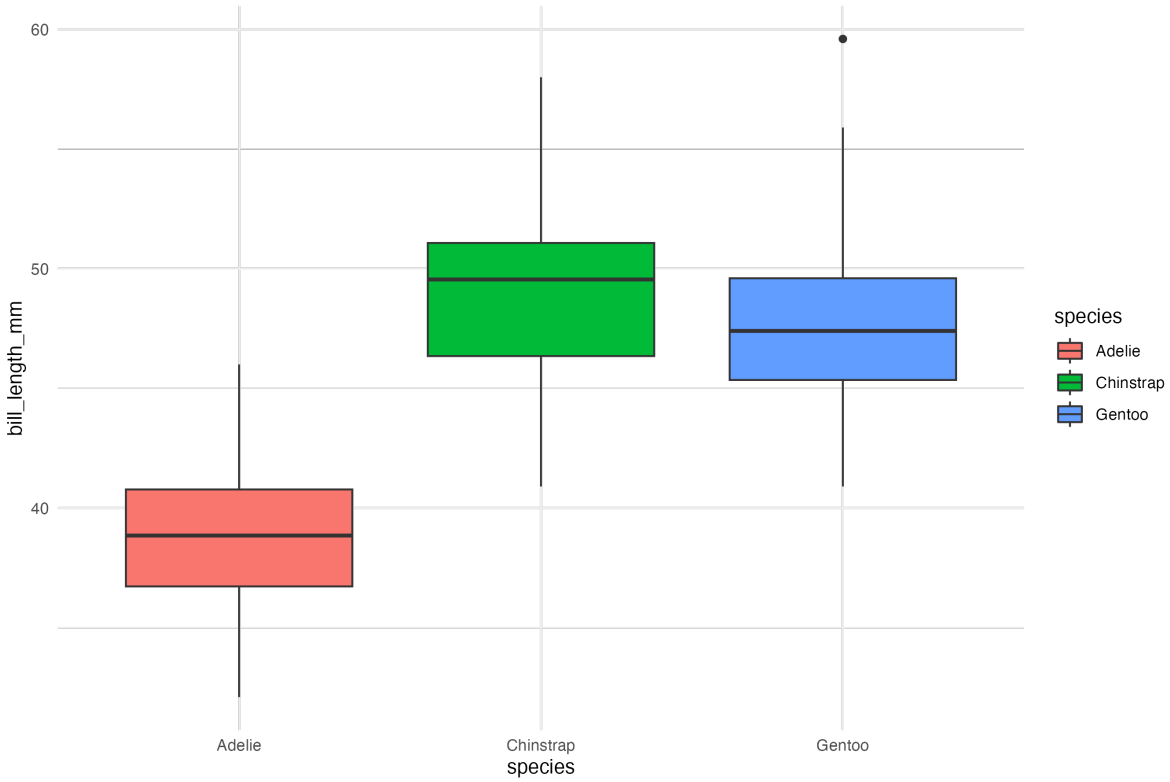
60
50
40
bill_length_mm
Adelie
Chinstrap
Gentoo
species
species
Adelie
Chinstrap
Gentoo

Figure 1: EDA

Table 3: Confusion Matrix Table

	Adelie	Chinstrap	Gentoo
Adelie	36	0	0
Chinstrap	1	17	0
Gentoo	0	0	30

Conclusion

In this tutorial, we:

- Loaded and cleaned the `palmerpenguins` dataset.
- Performed exploratory data analysis.
- Built a k-Nearest Neighbors classification model using `tidymodels`.
- Evaluated the model's performance.