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

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

# Methods

In this section, we perform exploratory data analysis (EDA) and prepare the data for modeling as seen in [Figure 1](#fig-bill_length).

|  |
| --- |
| Figure 1: Bill Length |

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

# Model

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

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

# Results

We evaluate the performance of the model using the test dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Confusion matrix of predicted vs actual penguin species   | Prediction | Truth | Freq | | --- | --- | --- | | Adelie | Adelie | 36 | | Chinstrap | Adelie | 1 | | Gentoo | Adelie | 0 | | Adelie | Chinstrap | 0 | | Chinstrap | Chinstrap | 17 | | Gentoo | Chinstrap | 0 | | Adelie | Gentoo | 0 | | Chinstrap | Gentoo | 0 | | Gentoo | Gentoo | 30 | |

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

Source: [Article Notebook](https://lukeni777.github.io/final_practice_temp/report.qmd.html)

# 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.