

Course Project - BIOL520I - Mock Manuscript

Harry Zhang^{a,*}

^a *University of British Columbia, 100 some street, Kelowna, 123456, B.C., Canada*

Abstract

This is a mock manuscript for course project BIOL520I Productivity and Reproducibility. This mini project explores penguin bill attributes in data package *palmerpenguins*.

Keywords: mock_project, course_project

1. Introduction

In this mini project for course BIOL520I Productivity and Reproducibility in Ecology and Evolution, the *palmerpenguins* data package was explored in R to understand the bill attributes of sampled penguin species. Tables and figures were created in this process. No in-depth analysis was conducted because the overall purpose of this mini project is to simply demonstrate a reproducible workflow.

2. Method

2.1. Data

The dataset used in this project is accessed from R data package *palmerpenguins* (Horst et al., 2020), originally published in Gorman et al. (2014).

2.2. Process

The dataset *penguins* was imported to dataframe from the data package. A copy of the raw, unaltered data was saved as a CSV file. The dataframe was then examined to understand the data structure completeness, where observations with missing bill attributes were dropped for simplicity in this demonstration. After cleaning up the data, I calculated simple descriptive statistics for bill dimensions (*bill_length_mm* and *bill_depth_mm* variables) and compared them across the three sampled species. Tables and boxplots were created to show these differences in bill dimensions. An additional histogram was created to show bill length distribution within the species “*Adelie*”. Finally, a scatterplot of bill dimensions was created across all three species with bill length on the x-axis and bill depth on the y-axis to visualize any potential correlation between the two variables. The scatter points were colored by species to aid identification of specific clustering in the variable space.

3. Results

The bill lengths by species is summarized in Table 1

*Corresponding author

Email address: harry.zhang@alumni.ubc.ca (Harry Zhang)

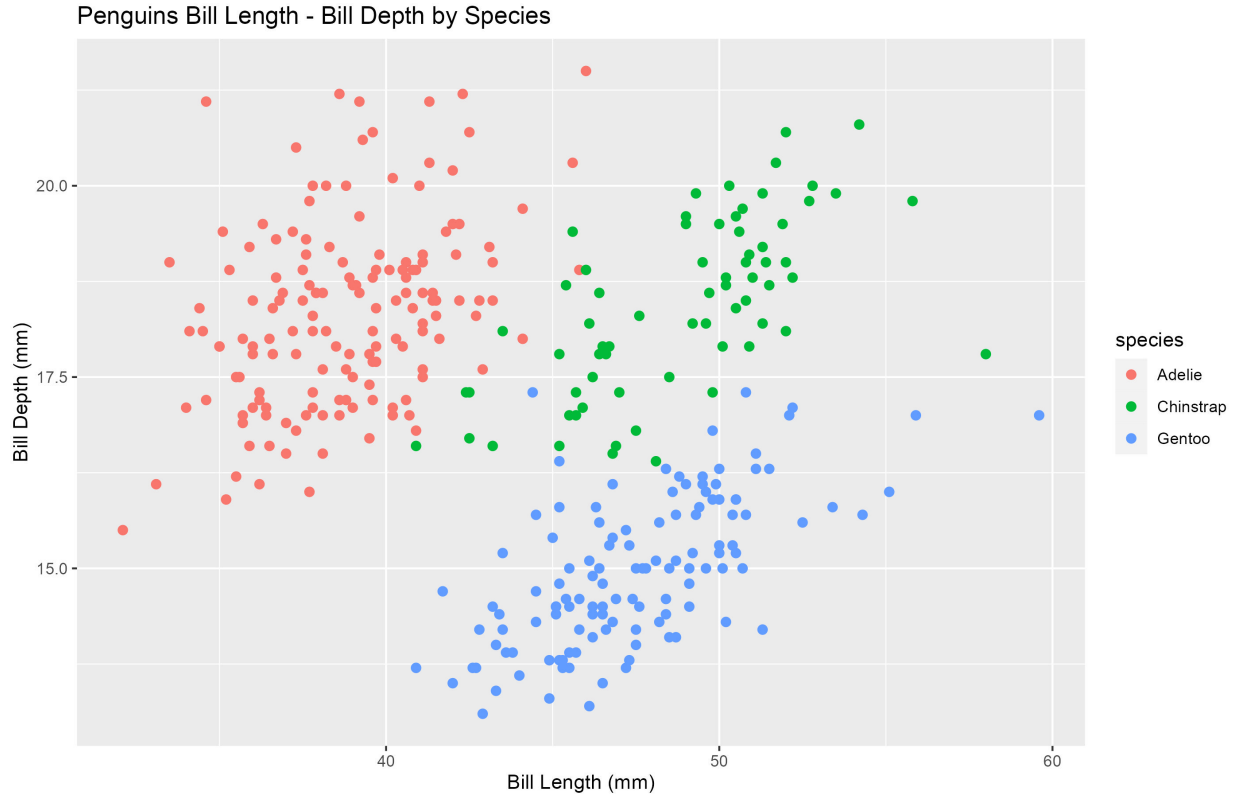


Figure 1: Scatterplot of Bill Length vs Bill Depth

Table 1: Bill Length by Species

species	min	mean	median	max
Adelie	32.1	38.79139	38.80	46.0
Gentoo	40.9	47.50488	47.30	59.6
Chinstrap	40.9	48.83382	49.55	58.0

The bill dimensions scatterplot (Figure 1) shows a visually detectable positive correlation between *bill_length_mm* and *bill_depth_mm* within species. There is specific clustering in the variable space.

4. Discussion

I have found a qualitatively positive correlation between bill length and bill depth in all three sampled penguin species. This correlation can be quantified using regression methods in future studies.

The bill attribute observations show a tendency of clustering by species in the bill dimension variable space. This suggests differences in bill sizes and shape characteristics across the three species. Further research on these characteristics may lead to quantitative methods to identify sample species using bill dimension measurements. # References {.unnumbered}

References

- Kristen B. Gorman, Tony D. Williams, and William R. Fraser. Ecological sexual dimorphism and environmental variability within a community of antarctic penguins (genus *pygoscelis*). *PLOS ONE*, 9(3):e90081, mar 2014. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0090081. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090081>.
- Allison M Horst, Alison Presmanes Hill, and Kristen B Gorman. palmerpenguins: Palmer archipelago (antarctica) penguin data. r package version 0.1.0., jul 2020. URL <https://allisonhorst.github.io/palmerpenguins/>.