From Penguins to Papers

Anna von Stein

Marcus Li

# Abstract

Reproducibility is a cornerstone of open science. In this demonstration, we explore how Quarto can be used to turn an exploratory data analysis into a transparent and reproducible scientific manuscript. Using the Palmer Penguins dataset, we test how bill length varies with depth, species, and sex. We show how Quarto enables a seamless integration of data, code, and narrative that is both executable and publication-ready.

# Introduction

The *palmerpenguins* Horst, Hill, & Gorman (2020) dataset has become a beloved alternative to the canonical iris dataset, offering richer ecological features and more realistic data imperfections. Bill length and depth are key morphological features in penguins, thought to relate to feeding ecology and sexual dimorphism. However, our true aim is not to contribute to penguin biology—but to demonstrate how to structure a computational notebook as a manuscript using [Quarto](https://quarto.org) Geert & Koßmann (2025).



Illustration of Adélie, Chinstrap, and Gentoo penguins. Artwork by Allison Horst.

Source: [Article Notebook](https://SORTEE.github.io/CodeClub/index-preview.html)

Quarto is a scientific and technical publishing system that combines the reproducibility of code with the clarity of writing. In this paper, we present a simple analysis of penguin morphology and show how Quarto can help you go from messy scripts to publication-ready documents.

# Methods

We used the palmerpenguins R package to access morphological data on three penguin species observed at the Palmer Station, Antarctica. Our cleaned dataset includes 333 individuals with complete measurements for bill length, bill depth, and sex.

Data were cleaned in scripts/clean-data.R and analyzed in scripts/analysis.R, with all outputs saved to the output/ folder. We modeled bill length as a function of bill depth, species, and sex using a linear model. The entire workflow is organized in a Quarto project with version-controlled scripts and data.

# Results

The model suggests that bill length increases with bill depth. Additionally, both species and sex have notable effects on bill size, with males generally exhibiting longer bills.

The figure below shows the estimated coefficients from the model:

|  |
| --- |
| Figure 1: Estimated coefficients from the model showing effects of bill depth, species, and sex on bill length. |

Source: [Article Notebook](https://SORTEE.github.io/CodeClub/index-preview.html)

# Discussion / Conclusions

Our mini-analysis illustrates how biological patterns (such as sexual dimorphism and inter-species differences) can be explored reproducibly using Quarto. But more importantly, this mock manuscript showcases how Quarto empowers researchers to write, code, and publish all in one place.

Every result shown here is fully traceable to the raw data, and every step of the analysis is version-controlled. This makes Quarto an excellent tool not only for individual projects but also for collaborative, transparent research workflows.

Future work may involve building more complex models or using Quarto to produce companion websites and teaching resources alongside publications.

# References

Geert, E. V., & Koßmann, L. (2025). Reproducible research reports with Quarto. *KU Leuven Open Science Day*.

Horst, A. M., Hill, A. P., & Gorman, K. B. (2020). *Palmerpenguins: Palmer archipelago (antarctica) penguin data*. Retrieved from <https://allisonhorst.github.io/palmerpenguins/>