PenguinsAnalysis

Biology3579

2024-12-06

Introduction

The genus Pygoscelis, which includes Adélie (*Pygoscelis adeliae*), Chinstrap (*Pygoscelis antarcticus*), and Gentoo (*Pygoscelis papua*) penguins, occupies overlapping ranges across the Western Antarctic Peninsula. Despite this sympatric distribution, these species exhibit distinct feeding ecologies, ranging from specialist to generalist strategies, allowing them to occupy different ecological niches and minimising competition[1]. Diet composition analysis of Chinstrap penguins shows a high degree of specialization, with their diet consisting almost exclusively of the krill species *Euphausia superba*[2]. In contrast, while Adélie penguins also rely heavily on krill, their diet is more varied, consisting of both *E. superba* and *E. crystallorophias* krill species, along with notothenioid fish[3]. Dietary analyses of Gentoo penguins reveal that this species has the most varied feeding habits, with a generalist diet consisting of a broad range of crustaceans and fish[1]. Notably, krill constitutes a smaller portion of their diet compared to other *Pygoscelis* species, highlighting a distinct difference in the feeding ecology of this species.

The bill morphology of penguins is a critical aspect of their feeding ecology, as it directly influences their ability to capture and consume different prey types. A study by Chávez-Hoffmeister (2020) found that penguin species display distinct bill shapes that are closely tied to their specific feeding strategies[4]. For example, krill-eating penguins tend to have wider beaks and broader jaws, which help them efficiently filter and capture krill. In contrast, fish-eating species possess narrower, more robust beaks designed to exert greater bite force, allowing them to grasp and hold onto slippery fish more effectively[4]. These morphological adaptations are essential for the species' feeding behavior, enabling them to exploit their preferred prey types. As such, each penguin species requires bill morphology that is finely tuned to its ecological niche, ensuring efficient foraging and survival in their specific environments.

Hypotheses

Based on the importance of bill morphology in feeding behaviors and the differences in feeding ecologies between the species, I hypothesize that the three penguin species will exhibit distinct bill morphologies that reflect their specific feeding strategies.

Methods

Data

To investigate the differences in bill morphology across penguin species, I will be utilizing the palmerpenguins dataset, which is derived from a study conducted by Gorman et al. (2014) on Pygoscelis penguins in the Palmer Archipelago, located west of the Antarctic Peninsula, near Anvers Island[5]. The data were collected from 2007 to 2009 across three islands: Biscoe (64° 48.9'S, 63° 46.9'W), Torgersen (64° 46.9'S, 64° 04.9'W), and Dream (64° 43.9'S, 64° 13.9'W). This dataset includes multiple measurements taken on three species of Pygoscelis penguins - Adélie, Chinstrap, and Gentoo - observed on these three islands, including body measurements (culmen length, culmen depth, flipper length, and body mass), sex information, nesting details (including clutch completion and egg laying dates), and stable isotope data (Δ15N and Δ13C) for 344 individual penguins across three species.

Statistical analyses

To investigate differences in bill morphology across Adélie, Chinstrap, and Gentoo penguins, I analyzed two key measurements: bill length and bill depth (which I also refer to as width). These metrics were used to assess morphological adaptations related to feeding ecology.

Ratio of Bill Length to Bill Depth The ratio of bill length to bill depth was used as an indicator of overall bill shape, as the two morphometric parameters for the penguins' bills provided in the dataset. A one-way ANOVA was conducted to compare the mean ratio across the three species, testing for significant differences in bill morphology. If significant results were found, Tukey's Honest Significant Difference (HSD) test was used for pairwise comparisons to identify where the species differences were coming from.

Individual Measurement Comparisons: Bill Length and Bill Depth To further examine morphological differences, I compared bill length and bill depth between the species using pairwise t-tests between each pair of penguin species (Adélie vs. Chinstrap, Adélie vs. Gentoo, and Chinstrap vs. Gentoo). Additionally, I calculated the 95% confidence intervals for the differences in means between each pair of species, which provide a precise estimate of the potential range for the true differences in bill morphology. These intervals help interpret the significance of the observed differences and whether they are meaningful in the context of ecological adaptation.

Analysis

Important note: *Comprehensive instructions, including all prerequisites and detailed guidelines for running this analysis and relevant code, are available in the README file of my GitHub repository for this project. Please refer to the README to ensure proper setup and usage.*

```
# Restoring project's environment
renv::restore()
```

```
## - The library is already synchronized with the lockfile.
```

```
# Source library loading fucntion
source(here::here("functions", "libraries.R"))
# Load necessary libraries
load_libraries()
```

```
#Load the raw data and save it
write.csv(penguins_raw, here("data", "penguins_raw.csv")) #To write data to csv
penguins_raw <- read.csv(here("data","penguins_raw.csv")) #Load data
```

```
#Clean the raw data and save it separately
source(here("functions", "cleaning_and_curating.R"))
penguins_clean <- cleaning_penguins(penguins_raw)
write_csv(penguins_clean, here("data", "penguins_clean.csv"))
```

```
#Curate the clean data and save it separately
source(here("functions", "cleaning_and_curating.R"))
#The curating_penguins_clean () funciton ...
analysis_data <- curating_penguins_clean(penguins_clean)
write_csv(analysis_data, here("data", "analysis_data.csv"))
```
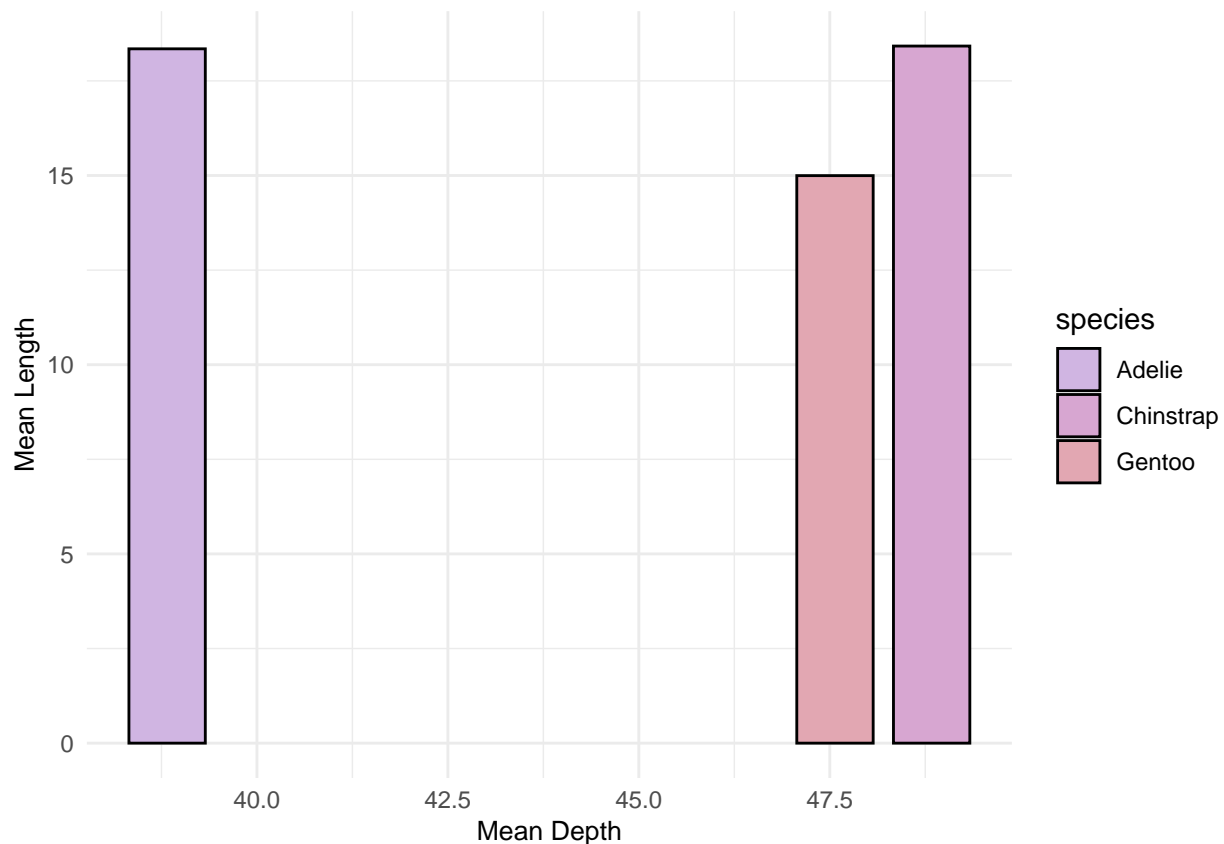
Exploratory analysis

QUESTION 1: Bad Figure

Figure 1: Bar chart of mean bill length vs. mean bill depth for three species of Pygoscelis *penguins.*

Reflection about bad figure

The figure displaying mean bill length and depth for Adélie, Chinstrap, and Gentoo penguins suffers from several design flaws that mislead the reader about the underlying data. The use of a bar chart to represent relationships between two continuous variables is inappropriate, as bar charts are better suited for categorical data and oversimplify the relationship being analyzed. A scatter plot would more effectively illustrate variation and trends between continuous measurements. Furthermore, the figure presents only summary statistics (mean values) without raw data points or indications of data distribution, preventing readers from assessing variability, outliers, or patterns within each species, which are critical for understanding the true complexity of the data.

Additionally, the figure lacks error bars or representations of variability, such as standard deviations or confidence intervals, which are essential for evaluating the precision of the means and the reliability of comparisons. Without them, the differences between species may appear more significant or consistent than they actually are.

Furthermore, some of the stylistic choices made for the graph make it harder to interpret the data. Firstly, the choice of similar colors for the three species which makes it difficult to differentiate between species at a glance; more contrasting colors or distinct patterns would improve clarity. In addition, the axis labels are vague - not indicating what exactly is being measured for length and depth - and fail to convey the units of measurement, detracting from the figure's interpretability.

Concerns about scientific reproducibility are increasingly prevalent[6]. Well-designed figures ensure that data is represented transparently, enabling others to interpret findings accurately and validate results. Poorly constructed visuals, on the other hand, can obscure key patterns or relationships, undermining the credibility of the research and impeding reproducibility.

QUESTION 2: Data Pipeline

```
# Specify location of functions
source(here("functions", "plotting.R"))
# Apply plot function
exploratory_figure <- plot_exploratory_figure(analysis_data)

# Suppress the null device output
invisible(capture.output(
```

```
# Apply save function
save_plot_svg(analysis_data,
              "figures/exploratory_figure.svg",
              size = 15,
              scaling = 1,
              plot_function = plot_exploratory_figure)
))
exploratory_figure # Show the plot
```
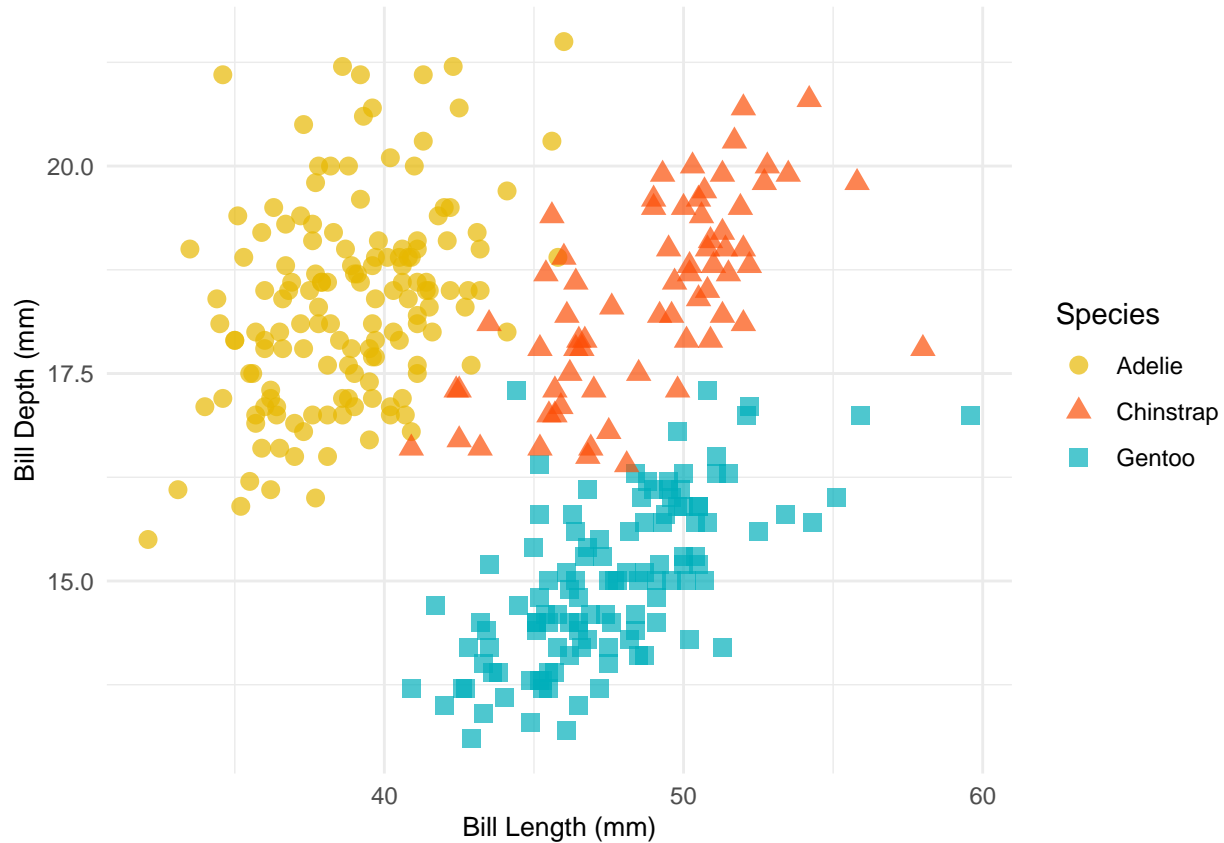

Figure 2: Scatterplot showing the distribution of bill lengths and bill depths for three species of Pygoscelis *penguins, with variability shown in bill length and depth across Adélie, Chinstrap, and Gentoo penguins.*

Figure 2 provides an exploratory visualisation of the distribution of bill length and bill depth among three species of *Pygoscelis* penguins, revealing distinct patterns in their morphospace. Adélie penguins tend to have relatively shorter but wider bills (greater bill depth), while Gentoo penguins exhibit longer and narrower bills. Chinstrap penguins occupy an intermediate space, showing bill depths similar to Adélie penguins but lengths closer to those of Gentoo penguins. This clear separation suggests potential species-specific differences in bill morphology.

However, it is important to note that this figure is purely exploratory. While it highlights potential patterns, no definitive conclusions about interspecies differences can be drawn without conducting appropriate statistical tests to assess the significance of these observations.

Statistical analysis

I first carried out an ANOVA, followed by Tukey-Kramer tests, to determine the general differences in bill morphology, using the bill length-to-depth ratio as a proxy for this. I then further investigated specific differences by testing bill length and bill depth separately to understand the variation between species. To provide more precise estimates of the species' bill characteristics, I also calculated 95% confidence intervals for both bill length and bill depth.

4

```r
# ANOVA to test for differences in bill morphology between penguin species
bill_morphology_model <- lm(bill_morphology ~ species, data = analysis_data)
anova(bill_morphology_model)
```

Differences in bill morphology

```
## Analysis of Variance Table
##
## Response: bill_morphology
##            Df Sum Sq Mean Sq F value    Pr(>F)
## species     2 73.175  36.587  1451.7 < 2.2e-16 ***
## Residuals 330  8.317   0.025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Tukey-Kramer test for pairwise comparisons of bill morphology between penguin species
TukeyHSD(aov(bill_morphology_model))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = bill_morphology_model)
##
## $species
##                        diff       lwr       upr p adj
## Chinstrap-Adelie 0.5322773 0.4774023 0.5871523     0
## Gentoo-Adelie    1.0551237 1.0089631 1.1012844     0
## Gentoo-Chinstrap 0.5228464 0.4660277 0.5796651     0
```

```r
# Pairwise t-tests for bill length between species with confidence intervals
bill_length_t_tests <- analysis_data %>%
  pairwise_t_test(
    bill_length_mm ~ species,      # Specify the variables
    paired = FALSE,                # Unpaired data
    p.adjust.method = "bonferroni" # Bonferroni correction for multiple tests
  )

# Print the result
print(bill_length_t_tests)
```

Differences in bill length

```
## # A tibble: 3 x 9
##   .y.          group1 group2    n1    n2        p p.signif     p.adj p.adj.signif
## * <chr>        <chr>  <chr>  <int> <int>    <dbl> <chr>        <dbl> <chr>
## 1 bill_length~ Adelie Chins~   146    68 2.52e-70 ****      7.57e-70 ****
## 2 bill_length~ Adelie Gentoo   146   119 1.05e-73 ****      3.16e-73 ****
## 3 bill_length~ Chins~ Gentoo    68   119 5.37e- 3 **        1.61e- 2 *
```

```r
# Pairwise t-tests for bill depth between species
bill_depth_t_tests <- analysis_data %>%
  pairwise_t_test(
    bill_depth_mm ~ species,        # Specify the variable and group
    paired = FALSE,                  # Specify the data is unpaired
    p.adjust.method = "bonferroni"  # Adjust p-values for multiple comparisons
  )

# Print the result
bill_depth_t_tests
```

Differences in bill depth

```
## # A tibble: 3 x 9
##    .y.          group1 group2    n1    n2        p p.signif    p.adj p.adj.signif
## * <chr>        <chr>  <chr>  <int> <int>    <dbl> <chr>       <dbl> <chr>
## 1 bill_depth_~ Adelie Chins~   146    68 6.57e- 1 ns          1     e+ 0 ns
## 2 bill_depth_~ Adelie Gentoo   146   119 6.62e-75 ****     1.99e-74 ****
## 3 bill_depth_~ Chins~ Gentoo    68   119 5.05e-59 ****     1.52e-58 ****
```

Results

All three studied penguin species show significant differences in length/depth ratios

```r
# Specify location of functions
source(here("functions", "plotting.R"))
# Apply plot function
bill_morphology_boxplot <- results_plot_1(analysis_data)

# Suppress the null device output
invisible(capture.output(
# Apply save function
save_plot_svg(analysis_data,
                "figures/bill_morphology_boxplot.svg",
                size = 15,
                scaling = 1,
                plot_function = results_plot_1)
))
bill_morphology_boxplot # Show the plot
```

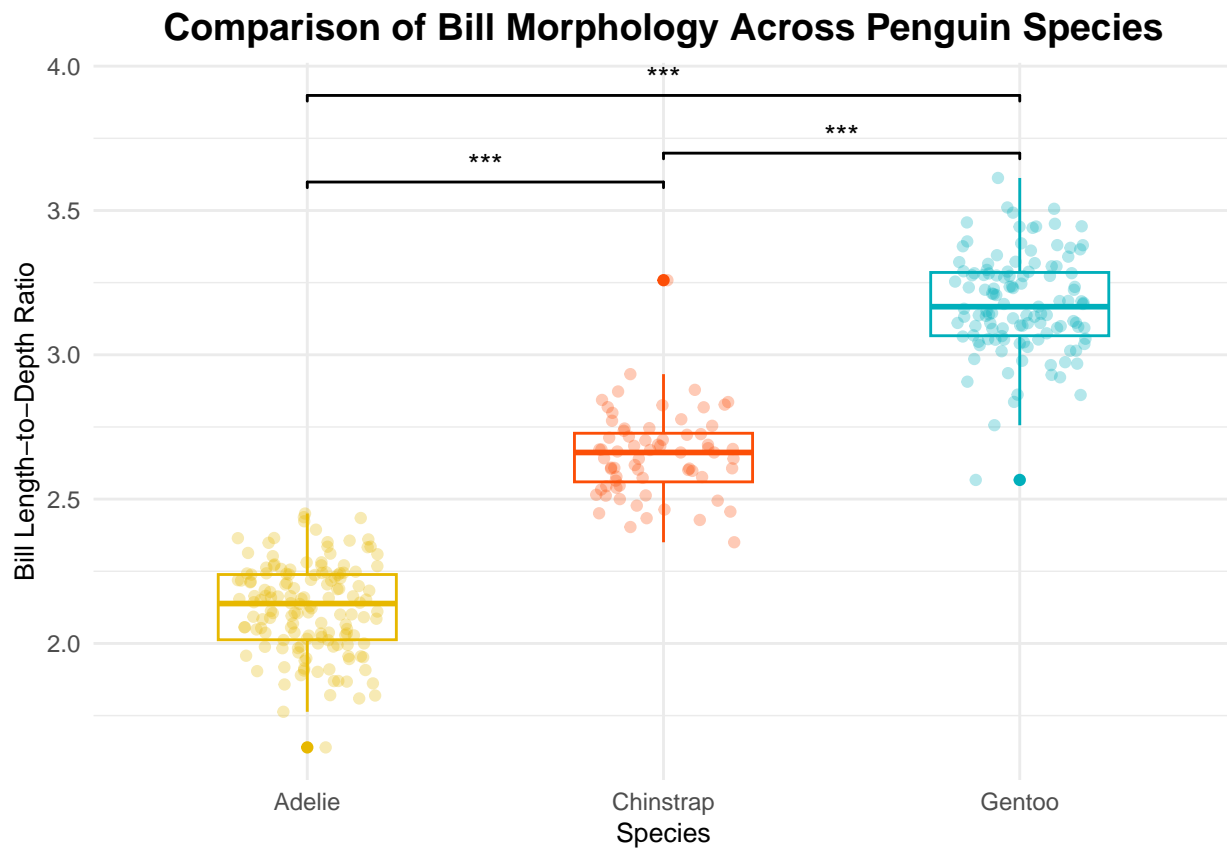# Comparison of Bill Morphology Across Penguin Species

Figure 3: Boxplot showing differences in bill morphology among Adélie, Chinstrap and Gentoo penguins. Statistical significance is indicated by asterisks (*** p < 0.001), showing highly significant differences between all species pairs.

To assess whether there are significant differences in bill morphology across the three penguin species (Adélie, Chinstrap, and Gentoo), a one-way ANOVA was performed on the combined bill length and depth measurements (bill morphology). The results from the ANOVA indicated that bill morphology significantly differed between the species ($F(2, 330) = 1451.7$, $p < 2.2e-16$), confirming the presence of interspecies variation. Additional post-hoc pairwise comparisons were conducted using Tukey's Honest Significant Difference (HSD) test, identified highly significant differences between all species pairs.

By analysis the specific bill length/depth ratios displayed in Figure 3, we can see that Gentoo penguins have the largest ratio, suggesting relatively longer and shallower bills, greatly contrasting to Adélie penguins have a smaller length/depth ratio, having shorter, deeper beaks. Instead, Chinstrap penguins seem to have an intermediate ratio, with their bill morphology falling between that of the Adélie and Gentoo penguins.

There is variability in bill length and depth across the three penguin species, with some overlap in their morphologies

```r
# Specify location of functions
source(here("functions", "plotting.R"))

# Apply plot functions
bill_length_boxplot <- results_plot_2(analysis_data)
bill_depth_boxplot <- results_plot_3(analysis_data)

# Create the multi-panel figure and display it
grid.arrange(bill_length_boxplot, bill_depth_boxplot, ncol = 2)

# Add labels to the panels
grid.text("A", x = 0.055, y = 0.95, gp = gpar(fontsize = 12))  # Adds label 'A' to the first graph
grid.text("B", x = 0.57, y = 0.95, gp = gpar(fontsize = 12))  # Adds label 'B' to the second graph
```
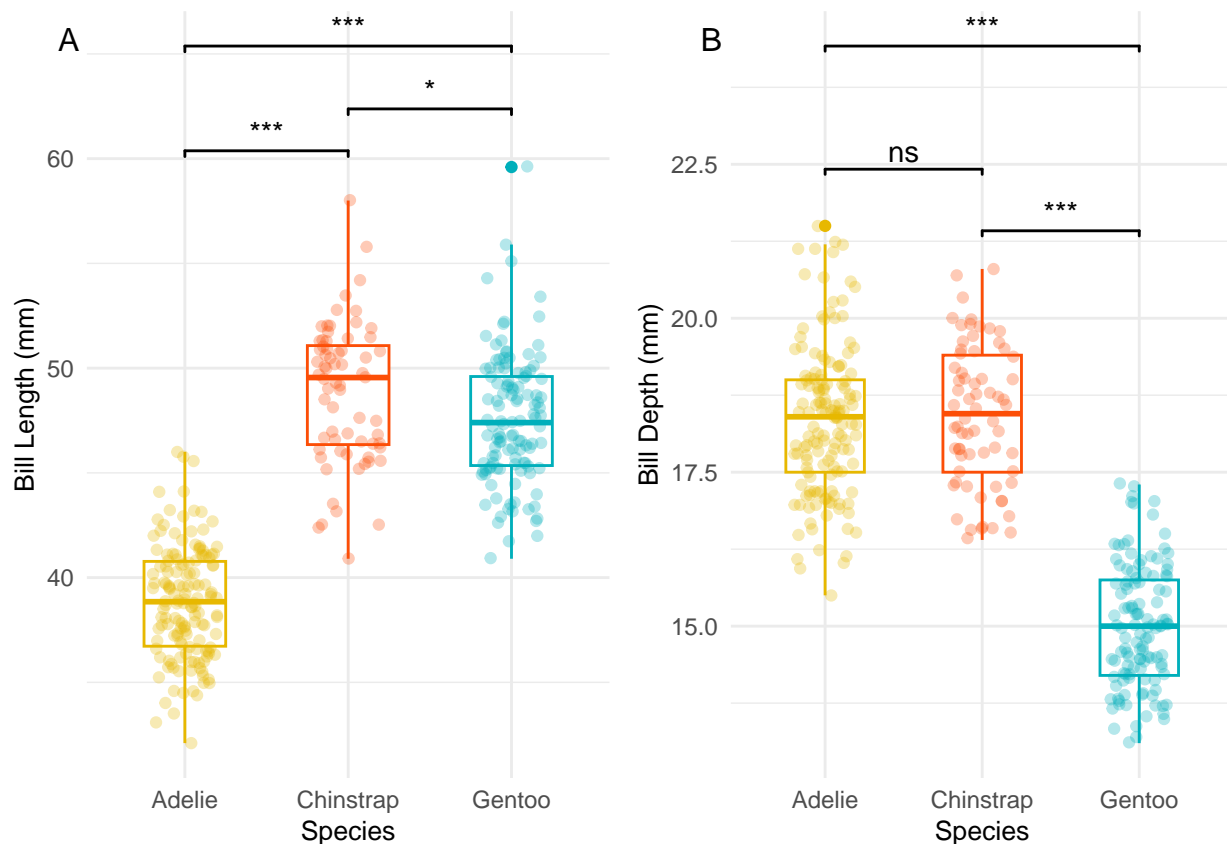
Figure 4: Boxplots showing differences in bill length (Figure 4a) and bill depth (Figure 4b) among Adélie, Chinstrap and Gentoo penguins. Statistical significance is indicated by asterisks (*** p < 0.001, ** p < 0.01, * p < 0.05, ns p > 0.05). Figure 4a shows differences in bill length across the three species. Figure 4b shows differences in bill depth across the three species.

To investigate further the differences in bill morphology, pairwise t-tests were conducted to assess whether there were significant differences in bill length and bill depth among the three species (Adélie, Chinstrap, and Gentoo). For bill length, the results showed significant differences between all species pairs. As shown in Figure 4a, Adélie penguins had highly significantly shorter bills compared to both Chinstrap (p < 2.52e-70) and Gentoo (p < 1.05e-73) penguins. The differences in bill length between Chinstrap and Gentoo penguins were less disparate but still significant. These findings indicate that bill length varies considerably between species.

Comparisons for bill depth across the species revealed that both Adélie and Chinstrap penguins exhibited significant differences in bill depth when compared to Gentoo penguins (Adélie vs. Gentoo: p < 6.62e-75, Chinstrap vs. Gentoo: p < 5.05e-59), but there were no significant differences between Adélie and Chinstrap penguins (p = 6.57e-01). From Figure 4b we can indeed see that Adélie and Chinstrap penguins have similarly wide bills, whereas Gentoo penguins have significantly narrower bills.

Conclusions

The results of this study reveal notable differences in bill morphology across the three Pygoscelis penguin species, highlighting potential relationships between bill structure and feeding strategies. All penguins show significant differences in bill length, which could indeed be reflective of different feeding habits. Additionally Gentoo penguins, in particular, exhibit distinctly narrower bills compared to the other two species, which may reflect their more diverse and generalized diet, consisting mainly of fish. The narrower bill may be better suited for capturing fish, which require more precise handling. On the other hand, Adélie and Chinstrap penguins show more similar bill depths that may be better adapted to capture krill, as the primary component of their diet. The similarity in bill depth between the two species might reflect a shared ecological niche, with both relying heavily on krill, a prey species that is heavily affected by climate change, particularly as warming oceans and retreating sea ice threaten the availability of krill in their habitats.

However, while the differences in bill morphology across the species are significant, the study does not provide direct evidence linking these morphological traits to dietary preferences. There may be other ecological factors at play, such as environmental influences and behavioural differences not associated to feeding (e.g social behaviours) which could also shape bill morphology.

Moreover, it is important to note that while the bill length-to-depth ratio has provided useful insights into bill morphology, it has limitations as a sole measure of feeding strategy. The relationship between bill morphology and feeding behavior is complex, and factors such as bill shape, size, and curvature could all play a role in how penguins capture and consume different prey types. Therefore, further research is needed to explore how these morphological traits interact with ecological factors and how penguin species might adapt to changing environmental conditions.

In conclusion, this study provides valuable insights into how bill morphology varies between three species of *Pygoscelis* penguins and provides promising insights into how these morphologies might reflect the different feeding strategies between these, but it also highlights the need for more comprehensive studies that take into account other ecological variables and the broader environmental context. Understanding these relationships is especially critical in the face of climate change, which is likely to affect the availability of prey species such as krill[7], and could enforce strong selection for shifting the dynamics between specialized and generalist feeders in penguin populations .

References

- [1] Polito, M., Trivelpiece, W., Patterson, W., Karnovsky, N., Reiss, C. and Emslie, S. (2015). Contrasting specialist and generalist patterns facilitate foraging niche partitioning in sympatric populations of Pygoscelis penguins. Marine Ecology Progress Series, 519, pp.221–237. doi:https://doi.org/10.3354/meps11095.

- [2] Ratcliffe☐N. and Trathan, P. (2011). Introduction A REVIEW OF THE DIET AND AT-SEA DISTRIBUTION OF PENGUINS BREEDING WITHIN THE CAMLR CONVENTION AREA. CCAMLR Science, [online] 18, pp.75–114. Available at: https: //www.ccamlr.org/en/system/files/science_journal_papers/Ratcliffe-Trathan_0.pdf [Accessed 11 Dec. 2024].

- [3] Tabassum, N., Lee, J.-H., Lee, S.-R., Kim, J.-U., Park, H., Kim, H.-W. and Kim, J.-H. (2022). Molecular Diet Analysis of Adélie Penguins (Pygoscelis adeliae) in the Ross Sea Using Fecal DNA. Biology, 11(2), p.182. doi:https://doi.org/10.3390/biology11020182.

- [4] Chávez-Hoffmeister, M. (2020). Bill disparity and feeding strategies among fossil and modern penguins. Paleobiology, [online] 46(2), pp.176–192. doi:https://doi.org/10.1017/pab.2020.10.

- [5] Gorman, K.B., Williams, T.D. and Fraser, W.R. (2014). Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis). PLoS ONE, 9(3), p.e90081. doi:https://doi.org/10.1371/journal.pone.0090081.

- [6] Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science, [online] 349(6251). doi:https://doi.org/10.1126/science.aac4716.

- [7] Clucas, G.V., Dunn, M.J., Dyke, G., Emslie, S.D., Levy, H., Naveen, R., Polito, M.J., Pybus, O.G., Rogers, A.D. and Hart, T. (2014). A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Scientific Reports, 4(1). doi:https://doi.org/10.1038/srep05024.

** QUESTION 3: Open Science**

My GitHub link: https://github.com/Biology3579/ReproducibleScienceAssignment.git

Note: for this section we worked in a three way system. I analysed Anonymous94394's code, whilst bleeddmagic's analysed my code.

Here are their respective repositories:

** Anonymous94394's GitHub link:** https://github.com/Anonymous94394/PalmerPenguins.git

** bleeddmagic's GitHub link:** https://github.com/bleeddmagic/PenguinsProject

Reflection after running Anonymous94394's

When I ran my partner's code, I found it to be well-organized and accessible, which greatly enhanced my understanding of the data pipeline. The code was structured into clearly defined sections with informative headings, making navigation intuitive. Each section was accompanied by comments explaining its purpose, which helped me follow the workflow and understand how the parts fit together. This thoughtful structure and clarity made the code easy to follow, even for someone unfamiliar with the dataset. A standout feature was the use of functions to break down complex tasks, which decluttered the main script and improved readability. This modular approach streamlined the process and allowed for easy reuse of code. I appreciated that the functions were called at the start, making their role within the script immediately clear. This separation of concerns not only enhanced organization but also made debugging and testing individual components straightforward. The use of the renv package to manage dependencies was another strong point. By ensuring access to the correct versions of required packages, the code was highly reproducible and avoided version-related errors. Additionally, listing all required libraries at the beginning of the script helped streamline the setup process and ensured that no dependencies were overlooked. These thoughtful design choices made setting up the environment much easier and faster. However, the code relies heavily on access to specific files in the repository, which is not clearly communicated. Without these files, the code cannot run properly. While this approach keeps the code itself uncluttered, the lack of a README.md file explaining these dependencies is a drawback. Adding a README file with setup instructions and a list of required files would improve accessibility and reproducibility, ensuring others can use the code without confusion.

The code executed without major issues, and I did not need to fix anything, though I had to ensure all necessary files and dependencies were in place. The use of renv simplified this process significantly, allowing the project to run seamlessly in its intended environment.

Overall, the code was well-prepared and reproducible, with only minor improvements needed to make it even more user-friendly and accessible to a broader audience.

Reflecting on my own code

What improvements did they suggest, and do you agree?

Based on my partner's feedback, one key suggestion was to place the methods and conclusions closer to the relevant code. I agree with this, as restructuring the code to include explanations directly beneath the relevant sections would improve readability. This would reduce the need to refer back to separate sections, making it easier for others to follow the analysis without constantly jumping between different parts of the code. Regarding the use of renv::init(), I made sure that I uploaded the updated renv.lock and properly specified its usage in the script, ensuring that the environment would be correctly initialized when other users run the code. This way, the necessary package versions and dependencies are accurately installed, and the environment remains consistent.

Another key suggestions was that my plotting functions could be made more flexible to allow for easier customization without having to dive into the plotting function itself. I agree with this suggestion. In my original code, I had static plotting functions that were designed to create specific figures, but they didn't allow for easy modifications like changing the axis labels or the variables being plotted.

Perhaps it would be more beneficial to design the function in a way that accepts dynamic arguments, allowing users to customise the plot directly from the analysis script without needing to modify the plotting function.

This could look like: # Function to create an exploratory figure with customizable parameters # plot_exploratory_figure <- function(data, x_col, y_col, x_label = " ", y_label =") { # ggplot(data, aes_string(x = x_col, y = y_col, color = color_map)) + # geom_point() + # labs(title ="Exploratory Figure", x = x_label, y = y_label) # # … # }

The user would then be able to customise specific parts of the function in the analysis script, (data, x-axis variable, y-axis variable, x-axis labels, and y- axis labels) and any other features added to this function defining section. While this change might make the code a bit longer, it does streamline the process of altering the figures, when having to make basic adjustments to the plots.

What did you learn about writing code for other people?

Working with other people and specifically having other people run my code, has been a valuable learning experience, particularly in understanding the importance of writing clear, reproducible, and accessible code for others. I have learned that while keeping code streamlined is important, it's equally essential to ensure it is flexible and reproducible across different environments. One of the challenges I encountered was ensuring that the environment was reproducible not only in my computer but for other people and their working systems too. I now truly appreciate the value of using tools like `renv` to manage package versions and environments, and `here` for consistent file paths. These tools ensure that the code runs smoothly regardless of the working system, making it more reproducible. Additionally, structuring functions to accept more customisable arguments makes the code more user-friendly and easier to adapt if necessary.