PenguinsAnalysis

Biology3579

2024-12-6

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. Diet composition analysis of Chinstrap penguins shows a high degree of specialization, with their diet consisting almost exclusively of the krill species *Euphausia superba*. 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. 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. Interestingly, 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 []. 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. 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.

Hypothesis

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. Specifically, since Gentoo penguins have a much more diverse diet compared to the other two species, I expect them to exhibit the most distinct bill morphology.

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. 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 analysis   To examine the differences in bill morphology across the three penguin speciesI will focus on two key measurements: bill length and bill depth. These measurements serve as indicators of bill morphology, providing insights into how adaptations to feeding strategies differ among species with varying diets.

1. Ratio of Bill Length to Bill Depth The first step in my analysis will involve examining the ratio of bill length to bill depth for each species. This ratio is a useful indicator of overall bill shape and can help highlight differences in bill morphology that are associated with each species' feeding ecology. For example, a more specialized feeder may have a differently proportioned bill compared to a more generalist species.

ANOVA tests assess whether there are any statistically significant differences in bill morphology, bill length, or bill depth between the species. The Tukey-Kramer test provides pairwise comparisons to identify which species pairs differ significantly from each other.

Statistical Test: One-Way ANOVA I will use one-way ANOVA to compare the mean ratio of bill length to bill depth across the three species (Adélie, Chinstrap, and Gentoo). ANOVA is appropriate for this analysis because I am comparing the means of a continuous variable (bill ratio) across more than two groups (species).

Rationale for ANOVA: ANOVA is ideal for testing if there are significant differences in the means of the bill ratio across the three species. If the ANOVA reveals a significant result, this would indicate that at least one species has a significantly different bill morphology in terms of its shape (ratio of length to depth).

Post-Hoc Analysis: If the ANOVA test shows significant differences, I will conduct post-hoc tests, such as Tukey's HSD (Honest Significant Difference), to determine which specific species pairs (e.g., Adélie vs. Gentoo) show significant differences in bill morphology. This is necessary to identify exactly where the differences lie between species.

2. Bill Length and Bill Depth Comparisons Following the ratio analysis, I will delve deeper into the individual measurements of bill length and bill depth. This will allow for a more detailed comparison of these specific morphological features across the species.

Statistical Test: One-Way ANOVA (for normally distributed data) I will first assess the normality of bill length and bill depth data for each species using tests such as the Shapiro-Wilk test. If the data are normally distributed, I will use one-way ANOVA to compare the means of bill length and bill depth across the three species.

Rationale for ANOVA: ANOVA is again used here to determine if there are significant differences in the bill length and bill depth among the species. A significant result would indicate that the species exhibit different morphologies in these specific aspects, which could be linked to their feeding behavior and ecological niches. I used this dataset to examine the culmen shapes of the different species, specifically, I will be focusing on two measurements - bill length and bill depth - as indicators of bill morphology.

To provide more precise estimates of the species' bill characteristics, I also calculated 95% confidence intervals for both bill length and bill depth. Confidence intervals are important because they offer a range within which the true population mean is likely to lie, allowing for a better understanding of the variability and uncertainty around the estimates. By including these intervals, I am able to give a more robust and complete picture of the data, helping to interpret the significance of the differences observed in the ANOVA and Tukey-Kramer tests.

Analysis

```
#renv::init()
#load all the required packages
library(here) # to specify the directory
library(tinytex)
library(renv) # to
library(janitor) # for cleaning the data
library(tidyverse) # for cleaning the data
library(palmerpenguins) # contains the dataset
library(dplyr) #for manipulating the data
library(ggplot2) #for making plots
library(lme4) #for making linear models
library(grid) #for making multi-panel figures
library(gridExtra) #for making multi-panel figures
library(ggsignif) #for adding significance stars to plots

#renv::snapshot()
```

Packages

All the necessary packages are store in the renv.lock file. The renv package manages the isolated environment for your R project, so the package versions are specific to the project rather than being installed globally. When you run renv::restore(), it will install the packages as described in the renv.lock file, ensuring that you and anyone else using the project have the same versions of the necessary packages. This process will make your project environment easily reproducible across different systems.

Note: install renv package (install.packages("renv") if not already installed.

```r
renv::restore() #to load up the necessary packages.
```

- The library is already synchronized with the lockfile.

```r
#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
glimpse(penguins_raw) # Quick summary of the structure of the raw data
```

```
## Rows: 344
## Columns: 18
## $ X                   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,~
## $ studyName           <chr> "PAL0708", "PAL0708", "PAL0708", "PAL0708", "PAL07~
## $ Sample.Number       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,~
## $ Species             <chr> "Adelie Penguin (Pygoscelis adeliae)", "Adelie Pen~
## $ Region              <chr> "Anvers", "Anvers", "Anvers", "Anvers", "Anvers", ~
## $ Island              <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen"~
## $ Stage               <chr> "Adult, 1 Egg Stage", "Adult, 1 Egg Stage", "Adult~
## $ Individual.ID       <chr> "N1A1", "N1A2", "N2A1", "N2A2", "N3A1", "N3A2", "N~
## $ Clutch.Completion   <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "N~
## $ Date.Egg            <chr> "2007-11-11", "2007-11-11", "2007-11-16", "2007-11~
## $ Culmen.Length..mm.  <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1~
## $ Culmen.Depth..mm.   <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1~
## $ Flipper.Length..mm. <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 1~
## $ Body.Mass..g.       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475~
## $ Sex                 <chr> "MALE", "FEMALE", "FEMALE", NA, "FEMALE", "MALE", ~
## $ Delta.15.N..o.oo.   <dbl> NA, 8.94956, 8.36821, NA, 8.76651, 8.66496, 9.1871~
## $ Delta.13.C..o.oo.   <dbl> NA, -24.69454, -25.33302, NA, -25.32426, -25.29805~
## $ Comments            <chr> "Not enough blood for isotopes.", NA, NA, "Adult n~
```

```r
#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"))
glimpse(penguins_clean) # Quick summary of clean data
```

```
## Rows: 344
## Columns: 18
## $ x                 <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1~
## $ study_name        <chr> "PAL0708", "PAL0708", "PAL0708", "PAL0708", "PAL0708~
## $ sample_number     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1~
## $ species           <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie", "A~
## $ region            <chr> "Anvers", "Anvers", "Anvers", "Anvers", "Anvers", "A~
## $ island            <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen", ~
## $ stage             <chr> "Adult, 1 Egg Stage", "Adult, 1 Egg Stage", "Adult, ~
## $ individual_id     <chr> "N1A1", "N1A2", "N2A1", "N2A2", "N3A1", "N3A2", "N4A~
## $ clutch_completion <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "No"~
## $ date_egg          <chr> "2007-11-11", "2007-11-11", "2007-11-16", "2007-11-1~
## $ culmen_length_mm  <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ~
## $ culmen_depth_mm   <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ~
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186~
## $ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ~
## $ sex               <chr> "MALE", "FEMALE", "FEMALE", NA, "FEMALE", "MALE", "F~
## $ delta_15_n_o_oo   <dbl> NA, 8.94956, 8.36821, NA, 8.76651, 8.66496, 9.18718,~
## $ delta_13_c_o_oo   <dbl> NA, -24.69454, -25.33302, NA, -25.32426, -25.29805, ~
## $ comments          <chr> "Not enough blood for isotopes.", NA, NA, "Adult not~
```

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

```
## Rows: 333
## Columns: 9
## $ x              <int> 1, 2, 3, 5, 6, 7, 8, 13, 14, 15, 16, 17, 18, 19, 20, 2~
## $ species        <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adelie~
## $ island         <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen", "T~
## $ individual_id  <chr> "N1A1", "N1A2", "N2A1", "N3A1", "N3A2", "N4A1", "N4A2"~
## $ bill_length_mm <dbl> 39.1, 39.5, 40.3, 36.7, 39.3, 38.9, 39.2, 41.1, 38.6, ~
## $ bill_depth_mm  <dbl> 18.7, 17.4, 18.0, 19.3, 20.6, 17.8, 19.6, 17.6, 21.2, ~
## $ body_mass_g    <dbl> 3750, 3800, 3250, 3450, 3650, 3625, 4675, 3200, 3800, ~
## $ sex            <fct> male, female, female, female, male, female, male, fema~
## $ bill_morphology <dbl> 2.090909, 2.270115, 2.238889, 1.901554, 1.907767, 2.18~
```
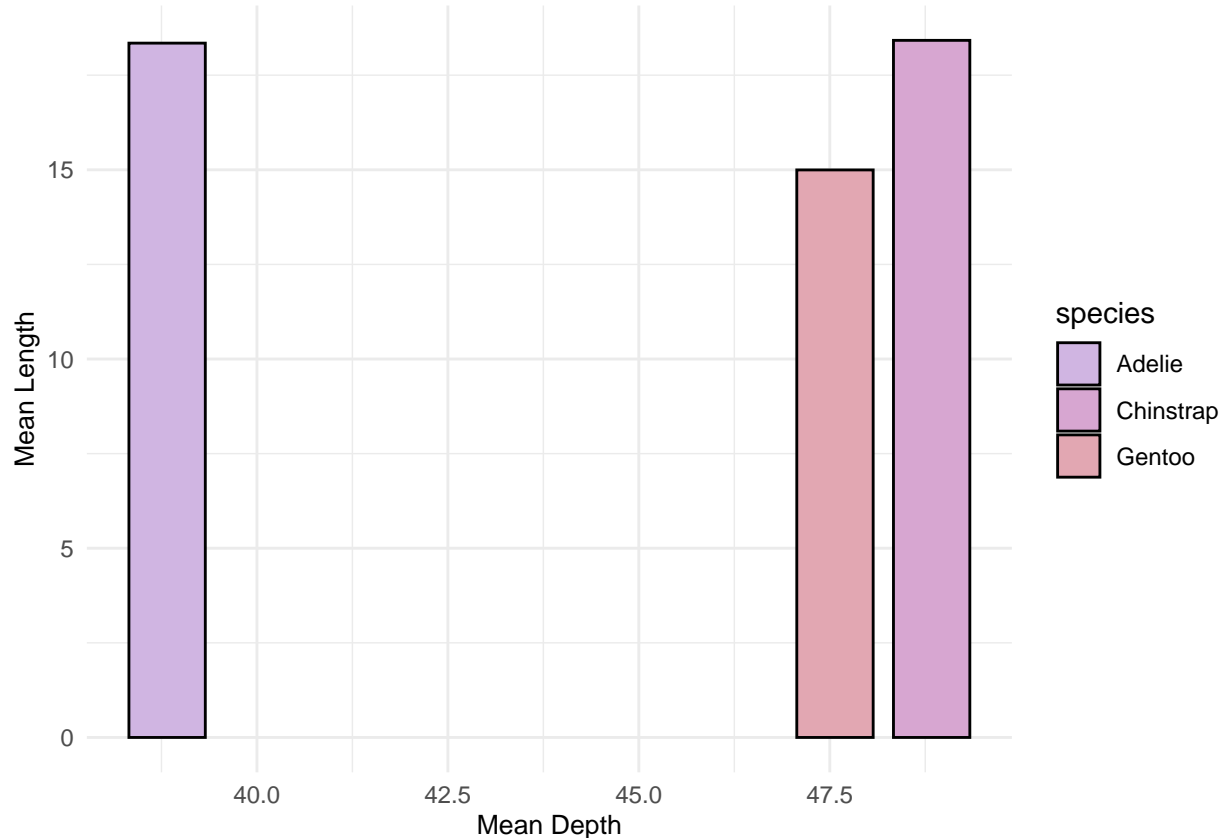
Exploratory analysis

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

b) Write about how your design choices mislead the reader about the underlying data (200-300 words).

This figure displays the mean bill length and depth for Adelie, Chinstrap, and Gentoo penguins. However, several design choices in this figure mislead the reader about the underlying data:

- A bar chart is a poor choice for displaying relationships between two continuous variables. Bar charts are typically used for categorical data, and applying them to continuous data (such as bill length and depth) oversimplifies the relationship. A scatter plot would be much more suited for representing .. []

4

- The figure only shows the summary statistics (mean values) for each species but does not include any raw data points or an indication of the data distribution. Without this, the reader cannot assess variability, outliers, or trends within each species, which are critical for understanding the data. This omission hides the true complexity and spread of the data. []

- The plot also lacks error bars or any indication of variability, such as standard deviation or confidence intervals. These are essential for understanding the precision of the mean values and the reliability of the comparisons between species. Without this information, the plot may give the false impression that the differences between species are more significant than they actually are []

- The colors used for the three species are all very similar and difficult to distinguish at a glance. This makes it hard for the viewer to quickly differentiate between species, even with a legend, and could cause confusion. More contrasting colors or patterns should have been used to make the distinction clearer.

- The axis labels are not informative and do not covey the units. Clearer labels would indicate what is being measured and …

Good exploratory figure

```
source(here("functions", "plotting.R"))
exploratory_figure <- plot_exploratory_figure(analysis_data)
exploratory_figure # Show the plot
```
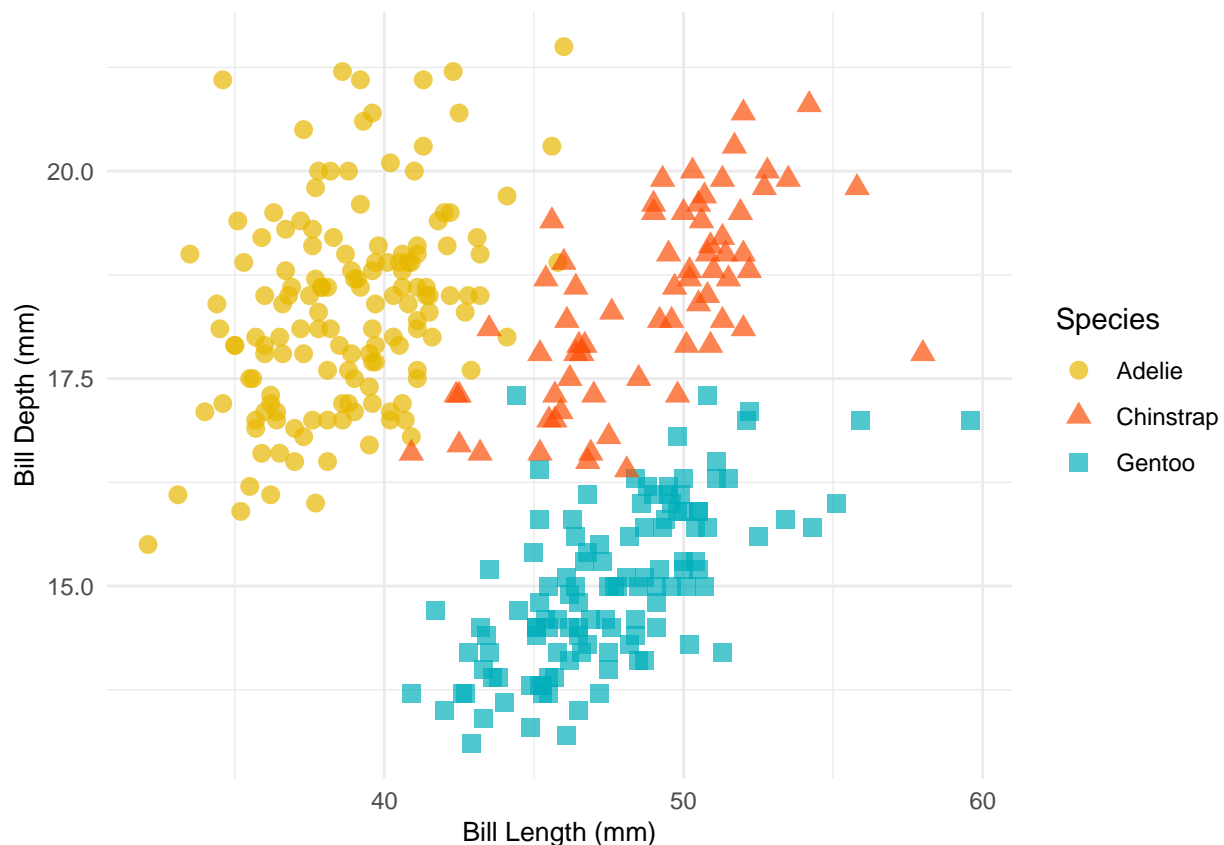


```
#Save plot
#gsave("exploratory_figure.png",
       #plot = exploratory_figure,
       #width = 8, height = 6, units = "in", dpi = 300)
```

Figure 2: Scatterplot showing the distribution of bill lengths and bill depths for three species of Pygoscelis penguins

Figure 2.A, we observe that the distribution of bill length and bill depth varies between the three penguin species. Adelie penguins tend to have smaller bill lengths and depths compared to the other species, while Chinstrap penguins show intermediate values. Gentoo penguins, on the other hand, display larger bills, with both the length and depth showing more variability across individuals…

The dsitrbution „,

Howver to fully deterine whetehr these differeces are we …

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.

```
# 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
```

```
# 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
```

```
# ANOVA to test for differences in bill length between penguin species
bill_length_model <- lm(bill_length_mm ~ species, data = analysis_data)
anova(bill_length_model)
```

Differences in bill length

```
## Analysis of Variance Table
##
## Response: bill_length_mm
##            Df Sum Sq Mean Sq F value    Pr(>F)
## species     2 7015.4  3507.7   397.3 < 2.2e-16 ***
## Residuals 330 2913.5     8.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Tukey-Kramer test for pairwise comparisons of bill length between penguin species
TukeyHSD(aov(bill_length_model))
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = bill_length_model)
##
## $species
##                         diff       lwr        upr       p adj
## Chinstrap-Adelie  10.009851  8.982789 11.0369128 0.0000000
## Gentoo-Adelie      8.744095  7.880135  9.6080546 0.0000000
## Gentoo-Chinstrap  -1.265756 -2.329197 -0.2023151 0.0148212
```

```r
# Pipe to calculate 95% confidence intervals for bill length
CIs_bill_length <- analysis_data %>%
  group_by(species) %>%
  summarise(
    mean_length = mean(bill_length_mm, na.rm = TRUE),
    se_length = sd(bill_length_mm, na.rm = TRUE) / sqrt(n()),
    lower_length_ci = mean_length - 1.96 * se_length,
    upper_length_ci = mean_length + 1.96 * se_length
  )

print(CIs_bill_length)
```

```
## # A tibble: 3 x 5
##   species    mean_length se_length lower_length_ci upper_length_ci
##   <fct>            <dbl>     <dbl>           <dbl>           <dbl>
## 1 Adelie            38.8     0.220            38.4            39.3
## 2 Chinstrap         48.8     0.405            48.0            49.6
## 3 Gentoo            47.6     0.285            47.0            48.1
```

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

Differences in bill depth

```
## Analysis of Variance Table
##
## Response: bill_depth_mm
##            Df Sum Sq Mean Sq F value    Pr(>F)
## species     2 870.79  435.39  344.83 < 2.2e-16 ***
## Residuals 330 416.67    1.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = bill_depth_model)
##
## $species
##                          diff       lwr        upr       p adj
## Chinstrap-Adelie   0.07332796 -0.315078  0.4617339 0.8968734
## Gentoo-Adelie     -3.35062162 -3.677347 -3.0238962 0.0000000
## Gentoo-Chinstrap  -3.42394958 -3.826113 -3.0217860 0.0000000
```

```r
# Pipe to calculate confidence intervals for bill depth
CIs_bill_depth <- analysis_data %>%
  group_by(species) %>%
  summarise(
    mean_depth = mean(bill_depth_mm, na.rm = TRUE),
    se_depth = sd(bill_depth_mm, na.rm = TRUE) / sqrt(n()),
    lower_depth_ci = mean_depth - 1.96 * se_depth,
    upper_depth_ci = mean_depth + 1.96 * se_depth
  )

print(CIs_bill_depth)
```

```
## # A tibble: 3 x 5
##   species   mean_depth se_depth lower_depth_ci upper_depth_ci
##   <fct>          <dbl>    <dbl>          <dbl>          <dbl>
## 1 Adelie          18.3   0.101           18.1           18.5
## 2 Chinstrap       18.4   0.138           18.2           18.7
## 3 Gentoo          15.0   0.0904          14.8           15.2
```

##Results

```r
source(here("functions", "plotting.R"))
results_figure_1 <- results_plot_1(analysis_data)

#Save plot
#gsave("exploratory_figure.png",
       #plot = exploratory_figure,
       #width = 8, height = 6, units = "in", dpi = 300)

results_figure_2 <- results_plot_2(analysis_data)

#Save plot
#gsave("exploratory_figure.png",
       #plot = exploratory_figure,
       #width = 8, height = 6, units = "in", dpi = 300)

#Make a multi-panel figure
grid.arrange(results_figure_1,  results_figure_2,  nrow = 2)

grid.text("A", x = 0.07, y = 0.95, gp = gpar(fontsize = 12)) #Adds label 'A' to first graph
grid.text("B", x = 0.57, y = 0.95, gp = gpar(fontsize = 12)) #Adds label 'B' to second graph
```
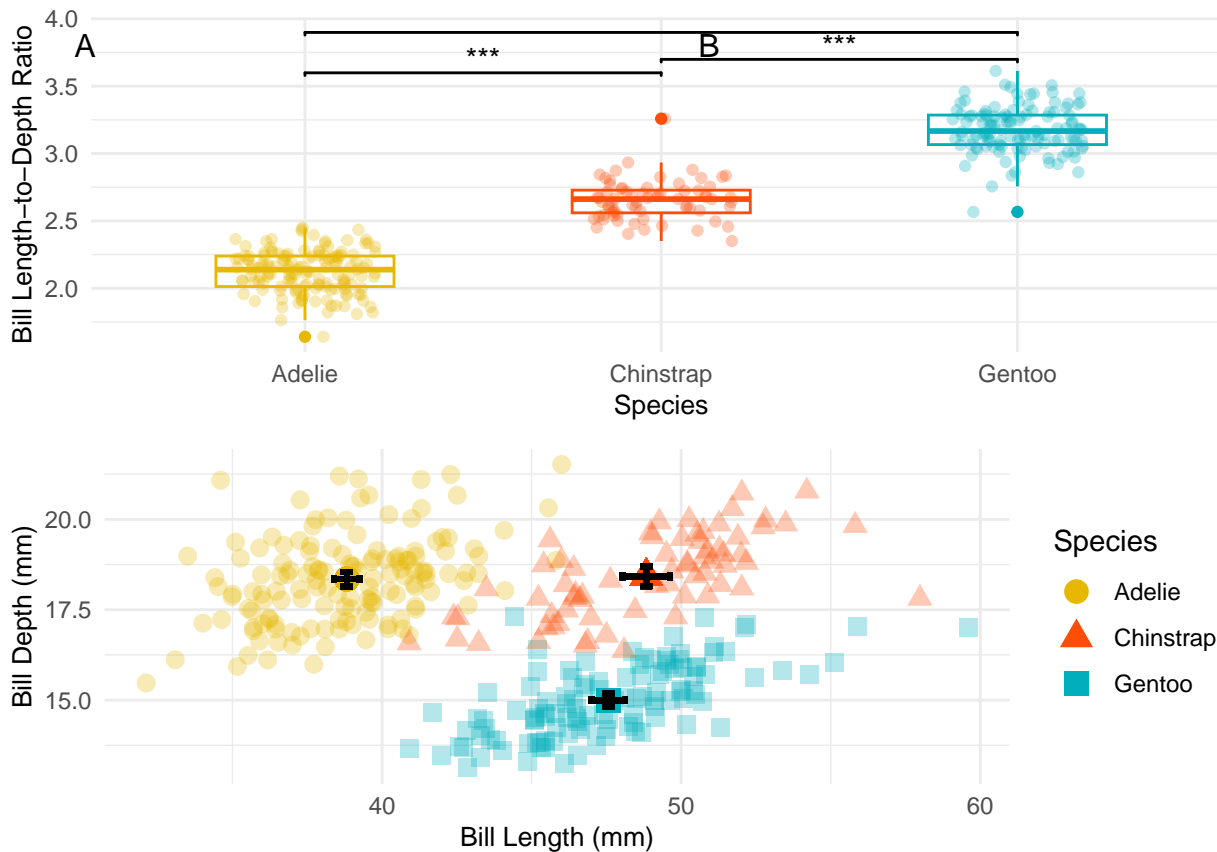
Figure 3. A plot …

Conclusion

Beak Shape Ratio: The ratio of culmen length to depth significantly differs across species. Gentoo penguins have the largest ratio, followed by Chinstrap penguins, and Adelie penguins have the smallest ratio.

Interpretation of Shape: Since the ratio is a measure of beak shape, this suggests that Gentoo penguins have relatively longer, shallower beaks, while Adelie penguins have shorter, deeper beaks.

Beak Length: Chinstrap and Gentoo penguins both have significantly longer beaks than Adelie penguins. Gentoo penguins have slightly shorter beaks than Chinstrap penguins, though this difference is still statistically significant. Beak Depth: Gentoo penguins have significantly deeper beaks compared to both Adelie and Chinstrap penguins. There is no significant difference in beak depth between Chinstrap and Adelie penguins. Are the results informative? Yes, these results are informative because:

They show clear patterns of differences in beak morphology (both length and depth) across species. The TukeyHSD results reveal specific comparisons, quantifying the size and direction of differences between species.

These findings align well with feeding habits.

As the cliamte changes, wih rapidly decling extent and duration of sea ice in Antartica, with correlated reductions in Antarctic krill6 (Euphausia superba), the main prey item for most meso- and top-predators in the Antarctic ecosystem will decrease. A recent article suggetsed that gentoo, which have the most diverse diets, will be able to adpat whilst chinsatrp and adelie, have been showing recent declines in. As such, nless these penguisna re able to adapt, they will perish (Clucas et al., 2014). Signficantylm I argue that bill mropholgoy might be amn impritnat determinant of diet, as it will speciesl;ay how they hunt and how ffectively they ctahc prey. of course mophological liamitaions and gentics will limit how quicly these can adapt and diactet whetehr tbe penguisna re able to survive at all.

References

…

Open Science

a) My GitHub link

https://github.com/Biology3579/ReproducibleScienceAssignment.git

b) Partner's GitHub link: https://github.com/Anonymous94394/PalmerPenguins.git

c) Reflect on your experience running their code. (300-500 words)

- *What elements of your partner's code helped you to understand their data pipeline?*

- *Did it run? Did you need to fix anything?*

- *What suggestions would you make for improving their code to make it more understandable or reproducible, and why?*

- *If you needed to alter your partner's figure using their code, do you think that would be easy or difficult, and why?*

d) Reflect on your own code based on your experience with your partner's code and their review of yours. (300-500 words)

- *What improvements did they suggest, and do you agree?*

- *What did you learn about writing code for other people?*