Q1b:

My graph is correct, but some design choices mislead the reader.

Firstly, there is no information presented about confounding variables, like species. This is misleading; penguin species is likely a confounding variable which impacts body mass, and the island(s) inhabited. This therefore may mislead a reader as the effects of different species are not explicitly shown, so a misleading correlation may be represented on my graph.

On this graph, I also only show a summary of the data, in this case maximum body mass in each category (i.e. on each island). This is misleading because important information may be hidden. For example, the maximum body mass for each island might be an outlier. This is also represented in the axis labels, which do not state maximum body mass but instead state only "body\_mass\_g".

The y-axis does not start from 0, but instead at 4000g, which is misleading as it makes differences in body mass between islands look bigger than they actually are.

Moreover, axis labels as specified in the .png plot are also very small, making it hard for a reader to understand what the graph is showing. Readers must look very closely at the axis labels to see the variables being compared. This is misleading because by not clearly showing your axis labels and any units associated with them, readers are more likely to miss key details about exactly what is being compared or what relationship is suggested by a graph between, which may cause them to believe a false correlation is present.

Finally, the colour scheme used is not at all accessible, using a red-green colour scheme which is not colour blind friendly. This means it is not easy to see what each colour represents for a colour-blind reader, and therefore makes my graph more misleading.

Q2a- Introduction (edited):

Many factors can impact beak morphology in birds, including foraging and thermoregulation. It has been shown that in many bird families, different species which inhabit different niches and have different foraging behaviours tend to evolve different bill shapes, which impacts variables like culmen length and depth (Friedman, 2019; Çakar, 2024). This pattern has been observed and tested in the birds of prey, but not in penguins (Çakar, 2024). The different sexes of many bird species also often have slightly different foraging behaviours outside of the breeding season, which could cause the evolution of different bill morphologies between the different sexes as each sex adapts to its different food sources (Gorman, 2014). Whether sex affects beak morphology is therefore an interesting question. It is known that species in the genus *Pygoscelis* display subtle sexual dimorphism in features like body mass, flipper length and bill size, with females usually being smaller than males. There is a possibility this will be reflected in the bill morphology of these penguins (Gorman, 2014; Polito, et al., 2012).

This analysis will use data from the Palmer Penguins dataset. This dataset contains data on 3 penguin species in the genus *Pygoscelis* (adelie, chinstrap and gentoo penguins) on a number of variables including body mass, culmen length, culmen depth, sex and clutch completion.

To study sexual dimorphism in beak shape in detail and simplify the analysis, this analysis will focus on only one penguin species (the chinstrap penguin- *Pygoscelis antarcticus*) to ensure only sex is examined. The chinstrap penguin has been chosen as it has been shown that it is the most dimorphic in culmen features, so will be an interesting species to study. It is hoped that findings from *P. antarcticus* can be extrapolated to other species within this genus, if not penguins as a whole.

## Hypotheses

Generic hypotheses on effect of sex:

Null hypothesis: The mean culmen length of each sex will not be significantly different from that of the other.

Alternative hypothesis: The mean culmen length of each sex will be significantly different from the other.

## Statistical Methods

In this analysis, an ANCOVA test will be conducted, to test whether sex affects culmen length, controlling for body mass. An ANCOVA is suitable for this analysis as there is a numerical response variable (culmen length), a categorical explanatory variable (sex) and a numerical covariate which must be controlled for (body mass). It is important to control for body mass as increasing body mass may be correlated with a longer beak.

Null hypotheses for ANCOVA on effect of sex in \*P. antarcticus\*:

1: There is no effect of sex on culmen length.

2: There is no effect of body mass on culmen length.

3: There is no interaction between sex and body mass.

Alternative hypotheses ANCOVA for effect of sex in \*P. antarcticus\*:

1: There is a significant effect of sex on culmen length.

2: There is a significant effect of body mass on culmen length.

3: There is a significant interaction between sex and body mass.

A two-sample t-test will also be conducted to assess whether sex affects body mass as well as culmen length.

Null hypothesis for t-test: There is no difference in mean body mass between male and female chinstrap penguins.

Alternative hypothesis for t-test: There is a statistically significant difference in mean body mass between male and female chinstrap penguins.

NB: The analyses presented here are not the same as in any referenced paper.

Results edited:

Sex w/ int:

* Adjusted R-squared = 0.4803 -> explains 48.03%% of variance.
* ANOVA finds no interaction of covariate. P=0.2953
* Sex has significant effect. p=<9.16\*10^-11
* Body mass has significant effect p=0.045

Sex w/o int:

* Adjusted R-squared = 0.4794 -> model explains 47.94% of variance.
* Sex has significant effect. p=<8.21\*10^-7
* Body mass does not have significant effect p=0.054

No massive difference between models (F = 1.1136).

Having conducted an ANCOVA analysis to test how sex affects culmen length when controlling for body mass, some interesting results were obtained.

The adjusted R-squared for the model including the interaction term is 0.4803, therefore 48.03% of variance is explained by the model. Sex has a significant effect on culmen length (p=<9.16\*10^-11), and as such the y-intercepts of the regression lines are significantly different. Body mass does seem to have a significant effect on culmen length (p=0.0450). There is no interaction effect between sex and body mass (p=0.2953), so the slopes of the regression lines are not significantly different.

Excluding the interaction term in the model, the adjusted R-squared for the model including the interaction term is 0.4794, therefore 47.94% of variance is explained by the model. Sex has a significant effect on culmen length (p=<8.6\*10^-11), and as such the y-intercepts of the regression lines are significantly different. Body mass does not seem to have a significant effect on culmen length (p=0.0540).

The difference between the models is minimal, but when the models are compared, the model without the interaction term has a marginally better fit (F=1.1136). Therefore, the findings from this model will be used to reject or not reject null hypotheses and draw conclusions from.

Using the results obtained from the model excluding the interaction term, it is possible to reject the following null hypotheses:

1: There is no effect of sex on culmen length.

It is not possible to reject the following null hypothesis:

2: There is no effect of body mass on culmen length.

3: There is no interaction effect between sex and body mass.

It should be stated that sex does also significantly affect body mass, as shown using a two-sample t-test (p=2.264\*10-6).

This therefore shows that sex does have an effect on culmen length in the chinstrap penguins, but body mass does not.

The fact that sex affects culmen length is unsurprising as it has been shown that sex affects bill morphology and \*P. antarcticus\* does display sexual dimorphism.

The fact that body mass does not also significantly affect culmen length is surprising as it is plausible that larger penguins would have longer culmens. While the effect was nearly significant, it was not less than the p < 0.05 cut off necessary to consider it significant. This may be because food is limited in the Antarctic ecosystem, so beaks are under some evolutionary constraint to ensure prey can still be caught, but this would require further research to verify.

It is unsurprising that sex affects body mass in this species independently, as this species has been shown to be dimorphic in body mass (Gorman, et al., 2014).

The above can also be seen on the results figure, where the slopes of each regression line are roughly parallel but each sex has a different y-intercept. It can also be seen that there are two distinct groups on the x-axis, with females generally having smaller body mass than males.

## Conclusions

In conclusion, it has been shown that culmen length is affected by sex, but not by body mass, and that there is no interaction effect between body mass and sex in *P. antarcticus*. Some of these results are surprising for reasons discussed above. This analysis does however still reaffirm the findings of previous research which stated that chinstrap penguins do display sexual dimorphism in their beak morphology (using culmen length as a proxy) and in their body mass.

Further research should be conducted to clarify why sexual dimorphism has evolved in the chinstrap penguins, and the genus *Pygoscelis* more widely. Moreover, the trophic ecology of male and female penguins in this genus should be studied to assess whether their trophic ecology differs greatly, and whether this can explain the variation in beak length and so beak morphology or whether some other factor has caused the evolution of sexual dimorphism. It would also be interesting to understand whether there is an evolutionary constraint which seems to make beak size consistent across chinstrap penguins regardless of body size. Lastly, it would be interesting to discover whether the findings of this analysis hold up in other penguin species or birds more widely.

References:

# References

Çakar, B. et al., 2024. Bill shape variation in selected species in birds of prey.. *Anatomia, Histologia, Embryologia,* 53(4), p. e13085.

Friedman, N. et al., 2019. Evolution of a multifunctional trait: shared effects of foraging ecology and thermoregulation on beak morphology, with consequences for song evolution.. *Proceedings of the Royal Society B,* 286(1917), p. 20192474.

Glen, S., 2020. *Misleading Graphs Part 1: Avoid These Common Mistakes.* [Online]   
Available at: https://www.datasciencecentral.com/misleading-graphs/  
[Accessed 11 December 2024].

Gorman, K., Williams, T. & Fraser, W., 2014. Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis). *PloS one,* 9(3), p. e90081.

Polito, M., Clucas, G., Hart, T. & Trivelpiece, W., 2012. A simplified method of determining the sex of Pygoscelis penguins using bill measurements.. *Marine Ornithology,* Volume 40, pp. 89-94.

Statistics How To, 2024. *Misleading Graphs: Real Life Examples.* [Online]   
Available at: https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/misleading-graphs/  
[Accessed 11th December 2024].

Q3:- for my partner

### 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?\*

The formatting of my partner’s code made it clear to see what each code chunk and function in each chunk does, clearly separating different parts to make code more understandable.

Steps for data analysis were well-presented such that it was clear how the analysis was structured. The reasons for carrying out certain steps in particular ways were well justified as well, making it clear what alternatives there might be to certain steps and why this person has used the methods they have.

The fact that detailed explanations of code were written above each code chunk also made the pipeline a lot more understandable, as it was very clear what each step was and why it was conducted.

My partner’s github repository was also well organised, with clearly labelled folders for things like functions, figures and data.

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

The code ran fine without needing to fix anything.

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

I would suggest that explanations of code could perhaps be more concise, cutting out unnecessary sentences or phrases to make the pipeline easier to understand overall. As it is, explanations are quite long and wordy which can make it more difficult to discern what steps are being taken, and therefore might make their analysis slightly less reproducible.

Some of the code chunks could however do with more annotations within the code itself, for example briefly explaining functions next to the line that they are in, to make it clearer what each part of the code does and why they are included. This would make it clearer why each part of the code is there and allow others to make changes more easily, but this is a minor issue and does not impact the reproducibility of this analysis, only the understanding of it.

They also should have put their cleaning function underneath the section of the cleaning.r file detailing the purpose of the script and the author. This would have made their cleaning file tidier. That said, this is again a minor issue which does not affect reproducibility.

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

If I had to alter my partner’s figure using their code, I believe I could accomplish this very easily. My partner has annotated their code to make their figure very well, such that it is easy to follow what each line does, track down where a change might need to be made and carry out that change accordingly. While I think this could be made easier with more in code annotations, on the whole I think it would be very easy to alter my partner’s figures.

c) for me:

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

My partner has extensively annotated their code, explaining what each line does in most cases. This helps understanding. By following the template, running their data follows a logical sequence. YAML headers also help understand what each chunk is for.

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

Broadly, yes the code ran. I first encountered an issue with the first library chunk, running `renv::restore()`, which should install the correct packages from the lockfile; however, I noticed some discrepancies. I got around this by just installing the most recent required packages with `renv::install()` . This works for now; however, isn't ideal for long-term reproducibility.

`renv::status()` revealed that "the lockfile was generated with R 4.4.1, but you're using R 4.2.2". This didn't seem to cause any issues and, if it did, it would fall on me to align the versions.

I move on to data preparation, which immediately gives an error:

```

> head(penguins\_data\_raw)

```

```

-Error in tempfile(pattern = "\_rs\_rdf\_", tmpdir = outputFolder, fileext = ".rdf") :

```

```

-temporary name too long

```

I circumvented this by re-extracting the repo to my Documents file, instead of a deeply nested OneDrive file. This fixed the issue.

When trying to run the data preparation chunk (line 134), it would not run when using the Run Current Chunk button. I was getting errors around text being part of the chunk. It was solved by adding a line break between the above section and the code chunk.

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

I find the in-code descriptions more difficult to read than the text sections. Personally, I think breaking the coding chunks into smaller segments and moving annotations to a text format would make the code more understandable. Additionally, where direct code annotations are preferable, such as for a minor annotation, I think the ease of understanding is improved if annotations immediately proceed the code, rather than all being found at the start of the coding sequence.

I think the code would be more readable if indents were smaller in some cases, such as with ggplot. The large indents followed by lines that over-run the line make following the logic more challenging to me.

When initially cleaning data, `na.omit()` is not included which causes some warning messages. It is used later when subsetting the data.

In the statistics chunk (line 337), the chunk produces two plated plots of Q-Q and Residuals vs Fitted. These seem almost identical and the cause for small difference between them is not obvious from the code. This does not make the user confident in the graph's reproducibility.

References are difficult to read in the bibliography and might have benefited from using Quarto's referencing capabilities.

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

I think the biggest obstacle to altering my partner's code would be the overall readability. Editing small, easily visible components like a title in a graph or colour would likely be easy. However, higher level, conceptual changes would likely be more difficult as a consequence of the formatting.

## 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?\*

My partner suggested I break up code chunks into smaller pieces with annotations written as text above, rather than within the chunk itself. I see the merit of this. Having in-code annotations can make code harder to read, however I wanted to keep to as few code chunks as possible so a user could run a specific chunk and get the desired output instead of running several smaller chunks. I also feel having bigger code chunks makes the document more easily navigated and more easily changed. For example, a user can easily identify my statistics chunk and change the model instead of having to sift through many code chunks to find what they are looking for, which can make my analysis more reproducible. That said, on my partner’s code the use of smaller code chunks did make their document easier to understand and so more reproducible.

They also suggested that annotations should immediately precede code instead of being at the top of a chunk. I agree; it would make my code clearer and therefore more reproducible. This was seen throughout my partner’s code and did help make their code a lot easier to understand.

They suggested having smaller indents. I also agree with this; smaller indents would ensure annotations do not run over a line, which would make my code clearer and more reproducible. While I had aimed to make my code clearer by formatting it with these indents, I think it had the opposite effect.

They critiqued me not using na.omit() at start when cleaning data. I see why this has been suggested, however as this does not affect the running of the code or the outputs generated, I would say this does not impact reproducibility.

They also critiqued having 2 sets of plated normality plots. I disagree with this suggestion. As the plots are only there to show that neither model violates the assumptions of the ANCOVA analysis I ran, and not to indicate a difference between the models, I do not believe this affects the reproducibility of my analysis.

Finally, my partner said that references were hard to read. I also disagree with this suggestion, as it is still obvious what the references are and they can still be found, so I would say this does not impact reproducibility.

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

From doing this exercise, I learnt the importance of clear annotation to explain what code does, as well as how to place those annotations within an r script to make code more understandable and therefore more reproducible. I also learnt how purely cosmetic formatting can actually make code harder to read and understand, which negatively impacts code reproducibility. There are many more factors which can impact reproducibility than I thought, and it was interesting trying to balance them all to ensure other people can read and interact with code that I am writing.