Manuscript for Data Analysis Project

Project 2

Chiyoko Onouye

Invalid Date

# 1. Summary/Abstract

*For Project 2, I used the cleaned Palmer Penguin data from project 1 and analyzed the data, creating graphs and figures to better visualize the data. Graphs were made based off of 3 main questions I wanted to answer about the Palmer Penguin dataset.*

# 2. Introduction

## 2.1 General Background Information

*Starting from Project 1, I used the Palmer Penguin dataset (****manual:2020?****) and cleaned it using Rstudio. This project–Project 2–aims to answer questions based on the clean data, again using Rstudio to make graphs and plots to help visualize the data.*

## 2.2 Description of data and data source

*The Palmer Penguins dataset is of morphometric data from three species of Antarctia penguins–Adelie Penguin (Pygoscelis adeliae), Gentoo penguin (Pygoscelis papua) and Chinstrap penguin (Pygoscelis antarctica)– such as: gender, flipper length, culmen length, culmen depth, island, and more.*

## 2.3 Questions/Hypotheses to be addressed

*Our research questions are:*

*1.Is there any differences seen between the two sex of penguins?*

*2.Are there really 3 groups of penguins based off the data?*

*3.Is there any differences seen between the Islands?*

# 3. Methods

*Using Rstudio, the Palmer Penguin set was cleaned by checking for errors and NA values in the data by using “skimr::skim”, “unique”, and view scatterplots to identify outliers.Then using the clean data, graphs and plots were made to help visualize the data.Packages required were: dplyr and tidyr for data processing and cleaning, skimr and lattice for nice visualization.*

## 3.1 Data acquisition, import and cleaning

*Data was obtained from an online source mbutler808(2023) Project-template[https://github.com/mbutler808/Project-template.git] using the penguins\_raw\_dirty.csv file. This data is based off the Palmer Penguin dataset(****manual:2020?****). I cloned this GitHub repository onto my computer and opened the Processingcode.R which can be found in the GitHub repository linked above under Project-template/Code/Processing\_code/processingcode.R. From here, I cleaned the data using the script written and saved the file as penguins.csv under Data/Processed\_data/. The script was checked by closing Rstudio and reopening it, deleting the Environment history, and running “source(”processingcode.R”) in a new terminal to make sure it runs without error and to make sure it has an output file. After, the penguins.csv file was loaded into a new R script using the path:data\_path <- “../../Data/Processed\_data/”.*

#Cleaning script

*Data was uploaded, checked for errors, and then plotted/graphed to help visualize the data. “plot” was used to look at the overall comparison between sex of the penguins and body mass (g), culmen length (mm), culmen depth (mm), and flipper length (mm). Then, each individual species was plotted to determine if there was a difference in sex within species. For the second question, “xyplot” was used to create scatterplots for each species next to each other while comparing body mass(g), culmen length (mm), culmen depth (mm), and flipper length (mm). Fianlly, to question the third question, both “plot” and “xyplot” were used to visualize data comparing the penguins to see if there was a difference between islands.*

#Anaylsis script

# 4. Results

## 4.1 Exploratory/Descriptive analysis

## 4.2 Sex

There appears to be a difference between female and male penguins, with males having higher body mass (g), culmen depth (mm), culmen length (mm), and flipper size (mm).

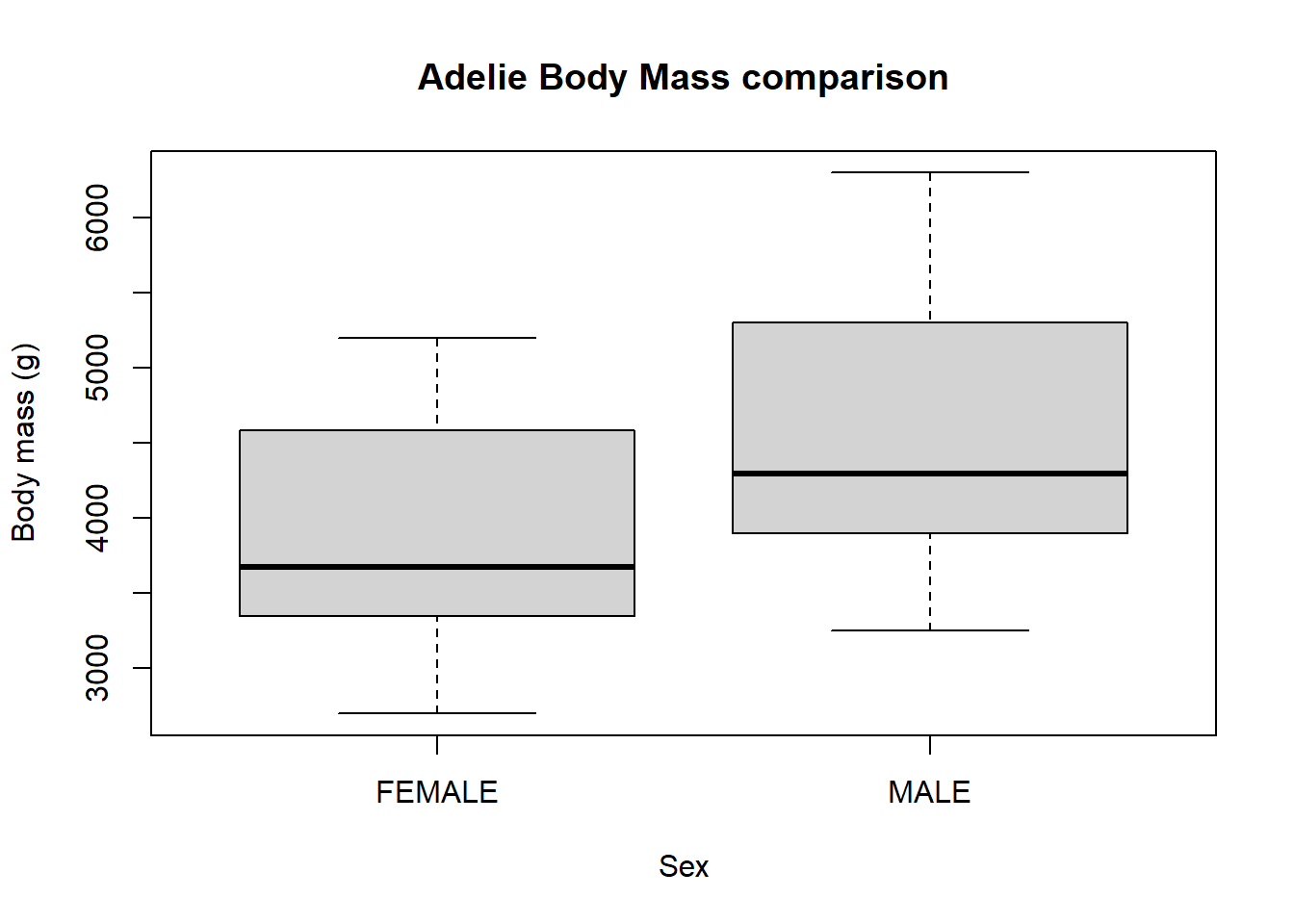
|  |
| --- |
| Figure 1: Comparison between gender and body mass, culmen depth, culmen length, and flipper length respectively. |

|  |
| --- |
| Figure 2: Comparison between gender and body mass, culmen depth, culmen length, and flipper length respectively. |

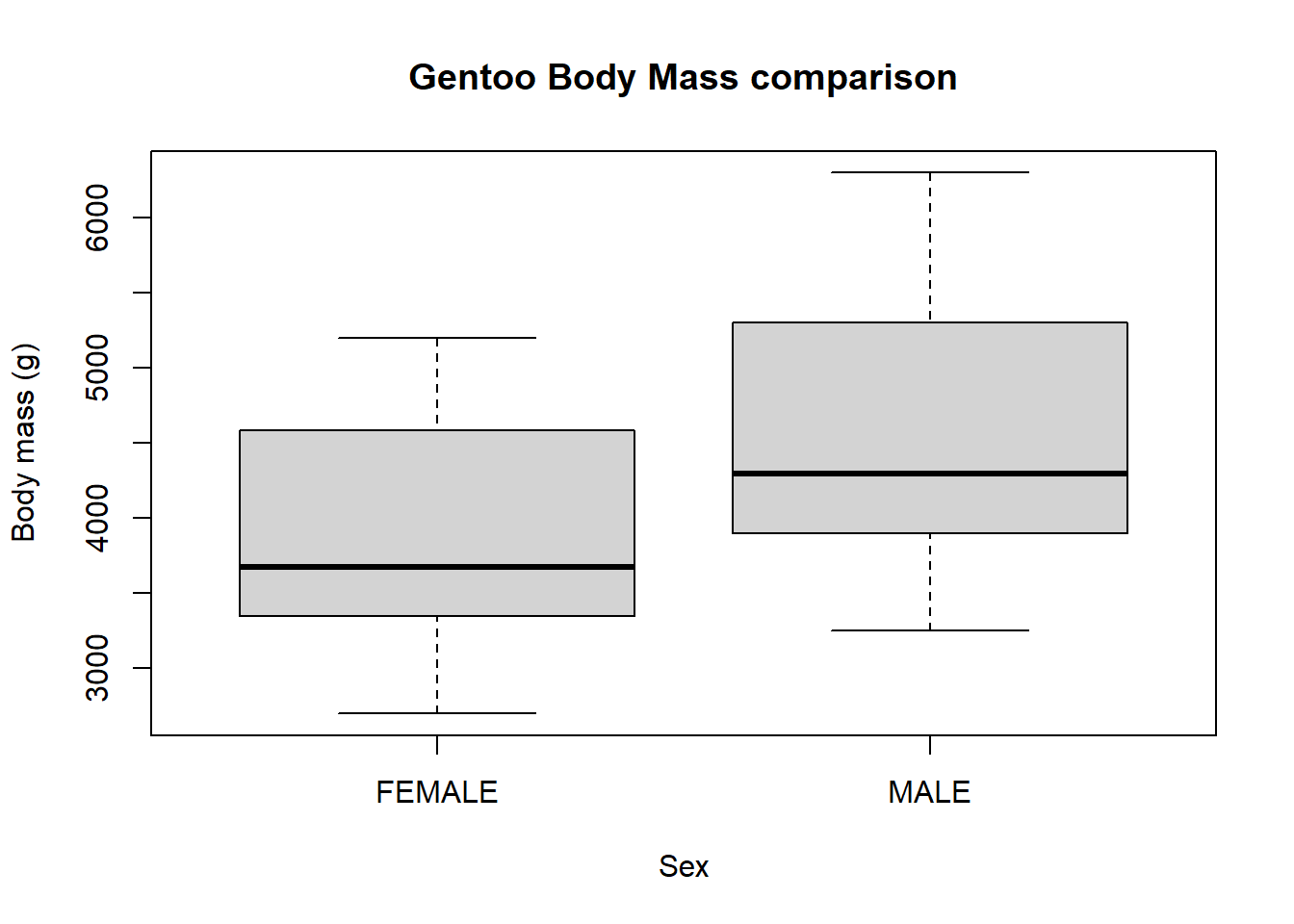
|  |
| --- |
| Figure 3: Comparison between gender and body mass, culmen depth, culmen length, and flipper length respectively. |

|  |
| --- |
| Figure 4: Comparison between gender and body mass, culmen depth, culmen length, and flipper length respectively. |

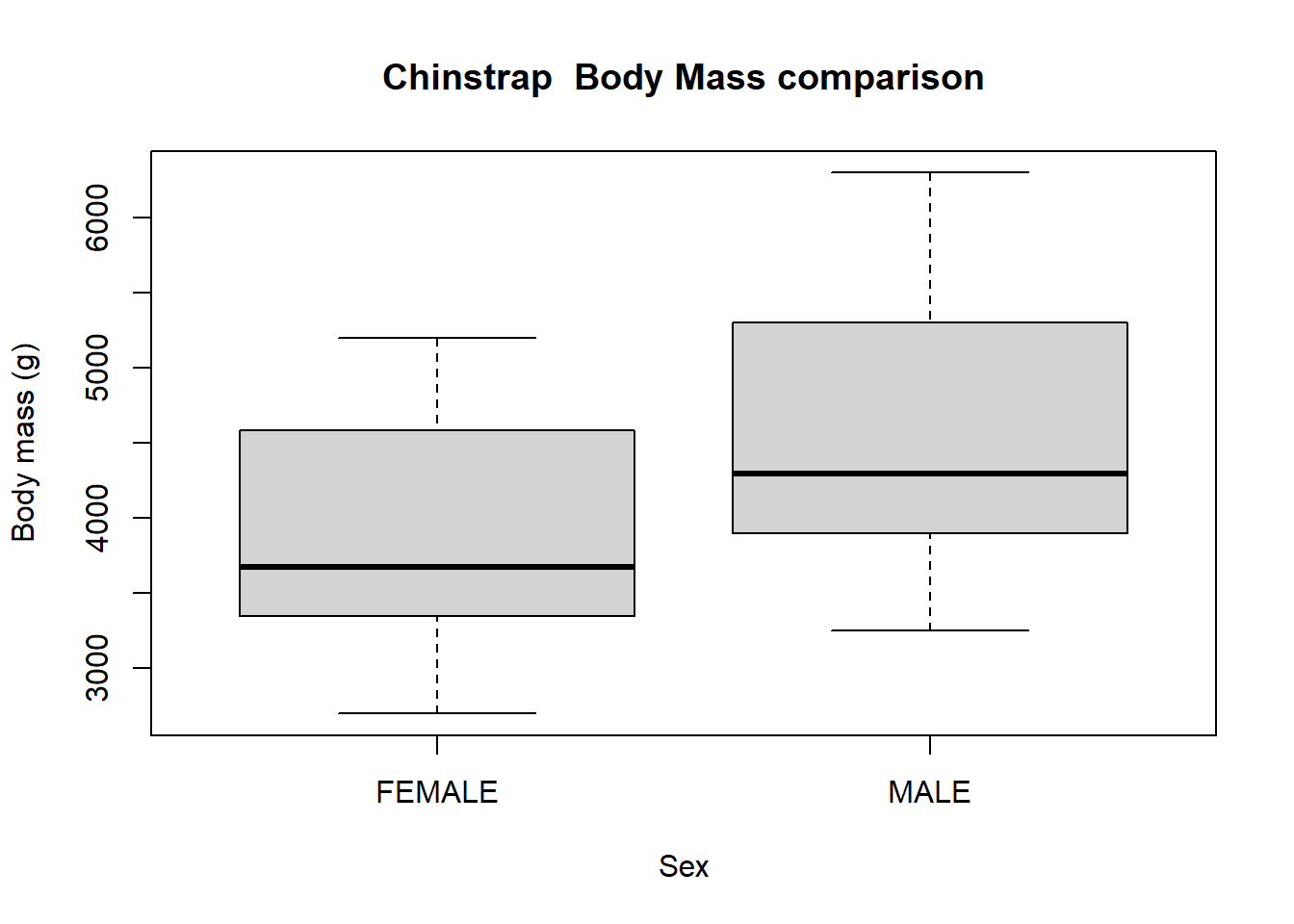
Difference between the sex of Adelie penguins.

{#fig-Adelie specific gender comparison}

Difference between the sex of Gentoo penguins.

{#fig-Gentoo specific gender comparison}

Difference between the sex of Chinstrap penguins.

{#fig-Chinstrap specific gender comparison}

The Gentoo penguin seem to have a higher body mass, culmen length, and flipper length when compare to the Adelie and Chinstrap.When looking at the culmen length and culmen depth, the Adelie have a lower culmen length but a higher culmen depth than both the Chinstrap and Gentoo(**?@fig-species\_comparison**).

|  |
| --- |
| Figure 5: Comparison of body mass, culmen length, culmen depth, and flipper length of all species |

|  |
| --- |
| Figure 6: Comparison of body mass, culmen length, culmen depth, and flipper length of all species |

|  |
| --- |
| Figure 7: Comparison of body mass, culmen length, culmen depth, and flipper length of all species |

|  |
| --- |
| Figure 8: Comparison of body mass, culmen length, culmen depth, and flipper length of all species |

|  |
| --- |
| Figure 9: Comparison of body mass, culmen length, culmen depth, and flipper length of all species |

|  |
| --- |
| Figure 10: Comparison of body mass, culmen length, culmen depth, and flipper length of all Islands. |

|  |
| --- |
| Figure 11: Comparison of body mass, culmen length, culmen depth, and flipper length of all Islands. |

|  |
| --- |
| Figure 12: Comparison of body mass, culmen length, culmen depth, and flipper length of all Islands. |

|  |
| --- |
| Figure 13: Comparison of body mass, culmen length, culmen depth, and flipper length of all Islands. |

|  |
| --- |
| Figure 14: Comparison of body mass, culmen length, culmen depth, and flipper length of all Islands. |

## 4.3 Further analyses

Based on the graphs created, there were some interesting points that would be worth looking further into. The body mass on Biscoe was higher than that of Dream Island and Torgersen.It would be interesting to look into this and identify if this was due to a certain species of penguins being more abundant on this island. We should also take into account the size of the island and the distance between all three islands. Similarly, it is noted that there was not an equal amount of each three species sampled in the original dataset which might skew the data in favor of one species over another.

Following the first idea, we noticed that the Gentoo Penguins have higher body mass than the other two species and noticed that Biscoe Island had the most amount of penguins sampled while Torgersen had the least amount sampled from. It would be interesting to look into the diet of the penguins since there was also a difference in culmen depth, culmen length, and flipper length between the penguins.

Finally, to ensure that there is a significant difference, we can actually perform an anova for these questions to identify if there is a difference. This would give us a statistical value which would help support our figures above.

# 5. Discussion

## 5.1 Summary and Interpretation

For this project, the cleaned data was taken and analyzed using Rstudio to make graphs and plots to help visualize the data. Three questions were posed at the beginning, 1. Is there a difference between the sex of penguins, 2. Are there really three groups of penguins–Adelie, Gentoo, and Chinstrap, and 3. Is there an difference between the islands? For all three questions, all continuous data was used to help analyze the data. Bar plots and comparative scatter plots were made to help visual data to help answer our questions.

## 5.2 Strengths and Limitations

This analysis looked at all the data presented to get an overall visual of the data to answer these questions. Further analysis should be done to better support these findings and more data is needed to help balance the species and island comparisons as they were not sampled from equally and having more data points in one over the other two might greatly skew the results.

## 5.3 Conclusions

To summarize, we found that 1. There is a difference between the sex of penguins, and more specifically, there is a difference between sex of all three species of penguins individually–with males having higher values in body mass, culmen length, culmen depth, and flipper length. 2. There are three distinct species of penguins based off the dataset, with Gentoo and Adelie having features that stick out, such as higher body mass and flipper length or higher culmen depth. 3. There is a difference between the islands as Biscoe Island had higher body mass than the other two. Dream Island also had a wider range of culmmen length when compared to Torgersen Island and Biscoe Island.

# 6. References

(**Manual?**{, title = {palmerpenguins: Palmer Archipelago (Antarctica) penguin data}, author = {Allison Marie Horst and Alison Presmanes Hill and Kristen B Gorman}, year = {2020}, note = {R package version 0.1.0}, doi = {10.5281/zenodo.3960218}, url = {https://allisonhorst.github.io/palmerpenguins/}, })