Reproducible Figures Assignment

2024-11-26

QUESTION 01: Data Visualisation for Science Communication

Loading libraries

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(palmerpenguins)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(here)
```

```
## here() starts at /cloud/project
```

```
library(ragg)
library(svglite)
```

Loading functions

```
#Cleaning function file
source(here("functions", "cleaning.R"))

#Functions to produce exploratory figures
source(here("functions", "exploratory_figure_functions.R"))

#Function file to save figures as .png and .svg files
source(here("functions", "png_and_svg_functions.R"))
```

Loading data set

```
#Loading and visualising raw dataset
penguins_raw
```

```
## # A tibble: 344 x 17
##    studyName `Sample Number` Species        Region Island Stage `Individual ID`
##    <chr>               <dbl> <chr>          <chr>  <chr>  <chr> <chr>
##  1 PAL0708                 1 Adelie Penguin~ Anvers Torge~ Adul~ N1A1
##  2 PAL0708                 2 Adelie Penguin~ Anvers Torge~ Adul~ N1A2
##  3 PAL0708                 3 Adelie Penguin~ Anvers Torge~ Adul~ N2A1
##  4 PAL0708                 4 Adelie Penguin~ Anvers Torge~ Adul~ N2A2
##  5 PAL0708                 5 Adelie Penguin~ Anvers Torge~ Adul~ N3A1
##  6 PAL0708                 6 Adelie Penguin~ Anvers Torge~ Adul~ N3A2
##  7 PAL0708                 7 Adelie Penguin~ Anvers Torge~ Adul~ N4A1
##  8 PAL0708                 8 Adelie Penguin~ Anvers Torge~ Adul~ N4A2
##  9 PAL0708                 9 Adelie Penguin~ Anvers Torge~ Adul~ N5A1
## 10 PAL0708                10 Adelie Penguin~ Anvers Torge~ Adul~ N5A2
## # i 334 more rows
## # i 10 more variables: `Clutch Completion` <chr>, `Date Egg` <date>,
## #   `Culmen Length (mm)` <dbl>, `Culmen Depth (mm)` <dbl>,
## #   `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>, Sex <chr>,
## #   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>, Comments <chr>
```

```r
#Saving raw datast
write_csv(penguins_raw, here("data", "penguins_raw.csv"))
```

Cleaning raw data set

```r
#clean data
penguins_clean <- penguins_raw %>%
  clean_column_names() %>%
  remove_columns("comments") %>%
  remove_columns("delta") %>%
  remove_empty_columns_rows() %>%
  shorten_species()

#save and load clean data
write_csv(penguins_clean, here("data", "penguins_clean.csv"))

penguins_clean <- read_csv(here("data", "penguins_clean.csv"), show_col_types = FALSE)

#visualise the cleaned data
colnames(penguins_clean)
```

```
##  [1] "study_name"        "sample_number"     "species"
##  [4] "region"            "island"            "stage"
##  [7] "individual_id"     "clutch_completion" "date_egg"
## [10] "culmen_length_mm"  "culmen_depth_mm"   "flipper_length_mm"
## [13] "body_mass_g"       "sex"
```

```r
penguins_clean
```

```
## # A tibble: 344 x 14
##    study_name sample_number species region island    stage        individual_id
##    <chr>              <dbl> <chr>   <chr>  <chr>     <chr>        <chr>
##  1 PAL0708                1 Adelie  Anvers Torgersen Adult, 1 Egg~ N1A1
##  2 PAL0708                2 Adelie  Anvers Torgersen Adult, 1 Egg~ N1A2
##  3 PAL0708                3 Adelie  Anvers Torgersen Adult, 1 Egg~ N2A1
##  4 PAL0708                4 Adelie  Anvers Torgersen Adult, 1 Egg~ N2A2
##  5 PAL0708                5 Adelie  Anvers Torgersen Adult, 1 Egg~ N3A1
```

```
##  6 PAL0708                 6 Adelie  Anvers Torgersen Adult, 1 Egg~ N3A2
##  7 PAL0708                 7 Adelie  Anvers Torgersen Adult, 1 Egg~ N4A1
##  8 PAL0708                 8 Adelie  Anvers Torgersen Adult, 1 Egg~ N4A2
##  9 PAL0708                 9 Adelie  Anvers Torgersen Adult, 1 Egg~ N5A1
## 10 PAL0708                10 Adelie  Anvers Torgersen Adult, 1 Egg~ N5A2
## # i 334 more rows
## # i 7 more variables: clutch_completion <chr>, date_egg <date>,
## #   culmen_length_mm <dbl>, culmen_depth_mm <dbl>, flipper_length_mm <dbl>,
## #   body_mass_g <dbl>, sex <chr>
```

Making scatterplot (Culmen Length vs Culmen Depth)

```r
#Subset columns and removing NA values
species_culmen_length_and_depth <- penguins_clean %>% select(species, culmen_length_mm, culmen_depth_mm
  drop_na()

#Visualising modified data set
species_culmen_length_and_depth
```

```
## # A tibble: 342 x 3
##    species culmen_length_mm culmen_depth_mm
##    <chr>              <dbl>           <dbl>
##  1 Adelie              39.1            18.7
##  2 Adelie              39.5            17.4
##  3 Adelie              40.3            18
##  4 Adelie              36.7            19.3
##  5 Adelie              39.3            20.6
##  6 Adelie              38.9            17.8
##  7 Adelie              39.2            19.6
##  8 Adelie              34.1            18.1
##  9 Adelie              42              20.2
## 10 Adelie              37.8            17.1
## # i 332 more rows
```

```r
#Saving modified data
write_csv(penguins_clean, here("data", "species_culmen_length_and_depth.csv"))
```

Making scatter plot of culmen length vs culmen depth for three species

```r
#Scatter plot for three species function
scatterplot <- three_species_scatterplot(species_culmen_length_and_depth,
                        culmen_length_mm, culmen_depth_mm, species,
                        "Culmen Length (mm)", "Culmen Depth (mm)")

#Visualising scatter plot
scatterplot
```
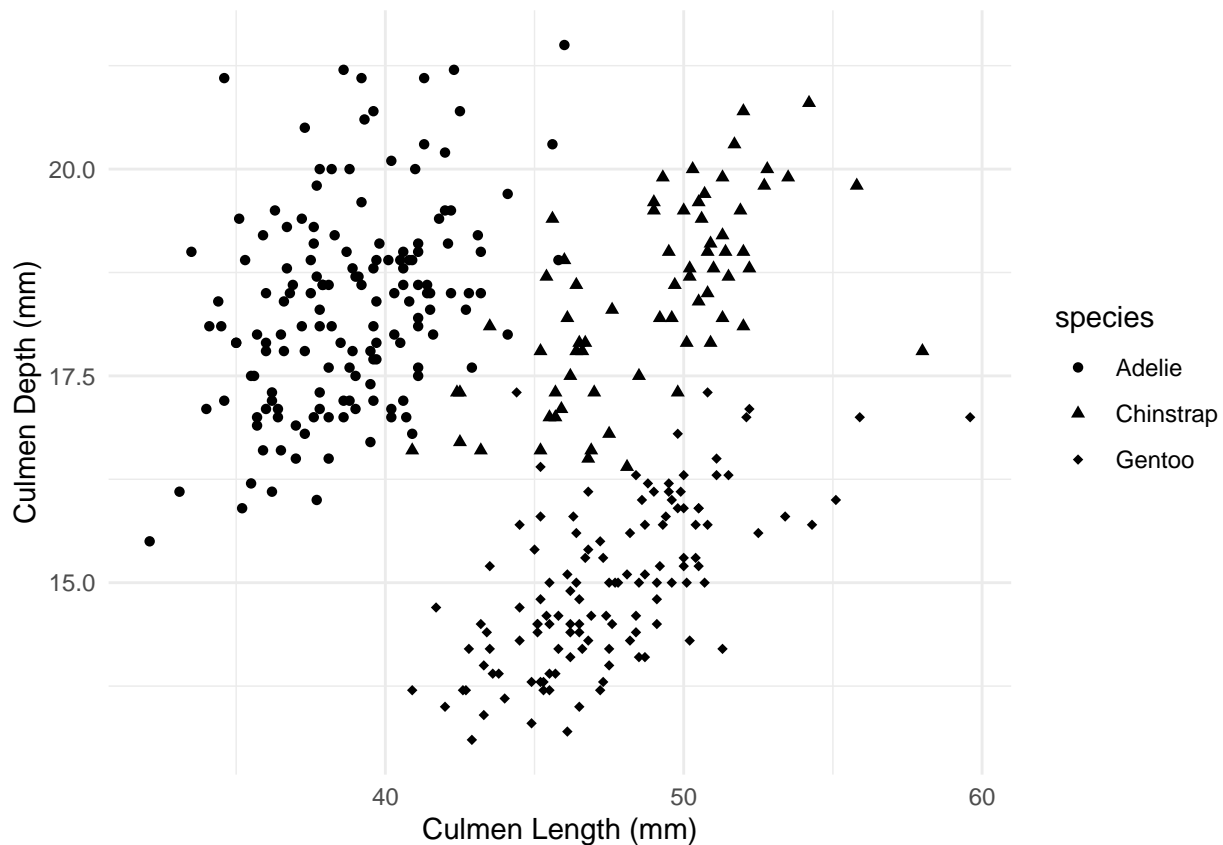
```
#Saving scatterplot as .png
save_as_png(species_culmen_length_and_depth,
            here("figures", "clvscd_scatterplot_report.png"),
            scatterplot,
            size = 20, res = 300, scaling = 2)
```

```
## pdf
##   2
```

```
#Saving scatterplot as .svg
save_as_svg(species_culmen_length_and_depth,
            here("figures", "clvscd_scatterplot_report.svg"),
            scatterplot,
            size = 20, scaling = 2)
```

```
## pdf
##   2
```

In this graph I have plotted Culmen Depth (mm) against Culmen Length (mm) for the three different penguin species. However, this figure badly communicates the data because it is difficult to distinguish the species data points from each other. Increasing the size of the figure would have helped to separate the data points from each other so that they aren't overlapping as much. Additionally, instead of using the same colour scheme for each data point, it would have been better to have assigned each species with a species specific colour. Ideally the colours picked would have been dissimilar from each other and colour blind friendly too (e.g. blue, orange and violet). Furthermore, the shapes chosen to represent each species are too similar, especially the shapes chosen to represent the Adelie and Gentoo species. To make it easier to distinguish between the data points, it would have been better to have chosen a cross to represent the Gentoo species. This would ensure that all species data points look different to each other. Finally, a trend line for each

species with a matching colour scheme would have been useful too. As long as the data points were more transparent than the trend line itself, this would make it easier to see how culmen depth varies with culmen length especially for each species in parts of the figure where there is an overlap of species data points.

QUESTION 2: Data Pipeline

INTRODUCTION

"Sexual dimorphism" means that the two sexes of a species differ in external appearances or other features (Ralls & Mesnich., 2009). This includes differences in size, where one of the sexes are considerably larger than the other. In this study, I aim to investigate whether this sexual size dimorphism is apparent in the Gentoo penguin (Pygoscelis papua) using body mass data acquired from the "palmerpenguins" data set. If there was no size difference, we would expect the coefficient from a linear model analysis to be zero. If there was a size difference, we would expect the coefficient from a linear model analysis to be significantly different form zero.

STATISTICAL METHODS

Loading libraries

```r
library(tidyverse)
library(palmerpenguins)
library(janitor)
library(here)
library(ragg)
library(svglite)
library(broom)
```

Loading functions

```r
#Cleaning function file
source(here("functions", "cleaning.R"))

#Function file to save figures as .png and .svg files
source(here("functions", "png_and_svg_functions.R"))

#Functions to produce exploratory figures
source(here("functions", "exploratory_figure_functions.R"))

#Linear model figure function
source(here("functions", "linear_model_figure.R"))

#Diagnostic plot function
source(here("functions", "diagnostic_plot_function.R"))
```

Loading the data

```r
#Loading raw data set
penguins_raw
```

```
## # A tibble: 344 x 17
##    studyName `Sample Number` Species         Region Island Stage `Individual ID`
##    <chr>               <dbl> <chr>           <chr>  <chr>  <chr> <chr>
## 1 PAL0708                 1 Adelie Penguin~ Anvers Torge~ Adul~ N1A1
## 2 PAL0708                 2 Adelie Penguin~ Anvers Torge~ Adul~ N1A2
## 3 PAL0708                 3 Adelie Penguin~ Anvers Torge~ Adul~ N2A1
## 4 PAL0708                 4 Adelie Penguin~ Anvers Torge~ Adul~ N2A2
## 5 PAL0708                 5 Adelie Penguin~ Anvers Torge~ Adul~ N3A1
## 6 PAL0708                 6 Adelie Penguin~ Anvers Torge~ Adul~ N3A2
## 7 PAL0708                 7 Adelie Penguin~ Anvers Torge~ Adul~ N4A1
```

```
##  8 PAL0708                  8 Adelie Penguin~ Anvers Torge~ Adul~ N4A2
##  9 PAL0708                  9 Adelie Penguin~ Anvers Torge~ Adul~ N5A1
## 10 PAL0708                 10 Adelie Penguin~ Anvers Torge~ Adul~ N5A2
## # i 334 more rows
## # i 10 more variables: `Clutch Completion` <chr>, `Date Egg` <date>,
## #   `Culmen Length (mm)` <dbl>, `Culmen Depth (mm)` <dbl>,
## #   `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>, Sex <chr>,
## #   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>, Comments <chr>
```

```r
#Saving raw data set
write_csv(penguins_raw, here("data", "penguins_raw.csv"))
```

Cleaning raw data set

```r
#clean data
penguins_clean <- penguins_raw %>%
  clean_column_names() %>%
  remove_columns("comments") %>%
  remove_columns("delta") %>%
  remove_empty_columns_rows() %>%
  shorten_species()

#save and load clean data
write_csv(penguins_clean, here("data", "penguins_clean.csv"))

penguins_clean <- read_csv(here("data", "penguins_clean.csv"), show_col_types = FALSE)

#visualise the cleaned data
colnames(penguins_clean)
```

```
##  [1] "study_name"       "sample_number"    "species"
##  [4] "region"           "island"           "stage"
##  [7] "individual_id"    "clutch_completion" "date_egg"
## [10] "culmen_length_mm" "culmen_depth_mm"  "flipper_length_mm"
## [13] "body_mass_g"      "sex"
```

```r
penguins_clean
```

```
## # A tibble: 344 x 14
##    study_name sample_number species region island    stage          individual_id
##    <chr>              <dbl> <chr>   <chr>  <chr>     <chr>          <chr>
##  1 PAL0708                1 Adelie  Anvers Torgersen Adult, 1 Egg~ N1A1
##  2 PAL0708                2 Adelie  Anvers Torgersen Adult, 1 Egg~ N1A2
##  3 PAL0708                3 Adelie  Anvers Torgersen Adult, 1 Egg~ N2A1
##  4 PAL0708                4 Adelie  Anvers Torgersen Adult, 1 Egg~ N2A2
##  5 PAL0708                5 Adelie  Anvers Torgersen Adult, 1 Egg~ N3A1
##  6 PAL0708                6 Adelie  Anvers Torgersen Adult, 1 Egg~ N3A2
##  7 PAL0708                7 Adelie  Anvers Torgersen Adult, 1 Egg~ N4A1
##  8 PAL0708                8 Adelie  Anvers Torgersen Adult, 1 Egg~ N4A2
##  9 PAL0708                9 Adelie  Anvers Torgersen Adult, 1 Egg~ N5A1
## 10 PAL0708               10 Adelie  Anvers Torgersen Adult, 1 Egg~ N5A2
## # i 334 more rows
## # i 7 more variables: clutch_completion <chr>, date_egg <date>,
## #   culmen_length_mm <dbl>, culmen_depth_mm <dbl>, flipper_length_mm <dbl>,
## #   body_mass_g <dbl>, sex <chr>
```

Cleaning and subsetting the data

```
#Clean and subset the data. Filter by Gentoo species and subset species, body mass and sex
clean_gentoo_body_mass_sex <- penguins_clean %>%
  filter(species == "Gentoo") %>%
  select(species, body_mass_g, sex) %>%
  remove_NA()

#Visualise the data
head(clean_gentoo_body_mass_sex)
```

```
## # A tibble: 6 x 3
##   species body_mass_g sex
##   <chr>         <dbl> <chr>
## 1 Gentoo         4500 FEMALE
## 2 Gentoo         5700 MALE
## 3 Gentoo         4450 FEMALE
## 4 Gentoo         5700 MALE
## 5 Gentoo         5400 MALE
## 6 Gentoo         4550 FEMALE
```

```
#Save cleaned data set
write_csv(clean_gentoo_body_mass_sex, here("data", "clean_gentoo_body_mass.csv"))
```

Creating an Exploratory Figure - Histogram

I first visualized the differences in body mass between the sexes using a histogram. The histogram suggests that the modal body mass of male Gentoo penguins is approximately 250 grams heavier than females, while the mass range appears similar for both sexes, around 1500 grams. Aside from two male outliers and one female outlier, the body mass data seems to exhibit little skew and appears approximately normally distributed. This suggests that a linear model would be an appropriate statistical test to compare body mass differences between the sexes.

```
#Histogram function
histogram <- overlapping_histogram_comparing_sexes(clean_gentoo_body_mass_sex, body_mass_g, sex, "Body

#Visualising histogram
histogram
```
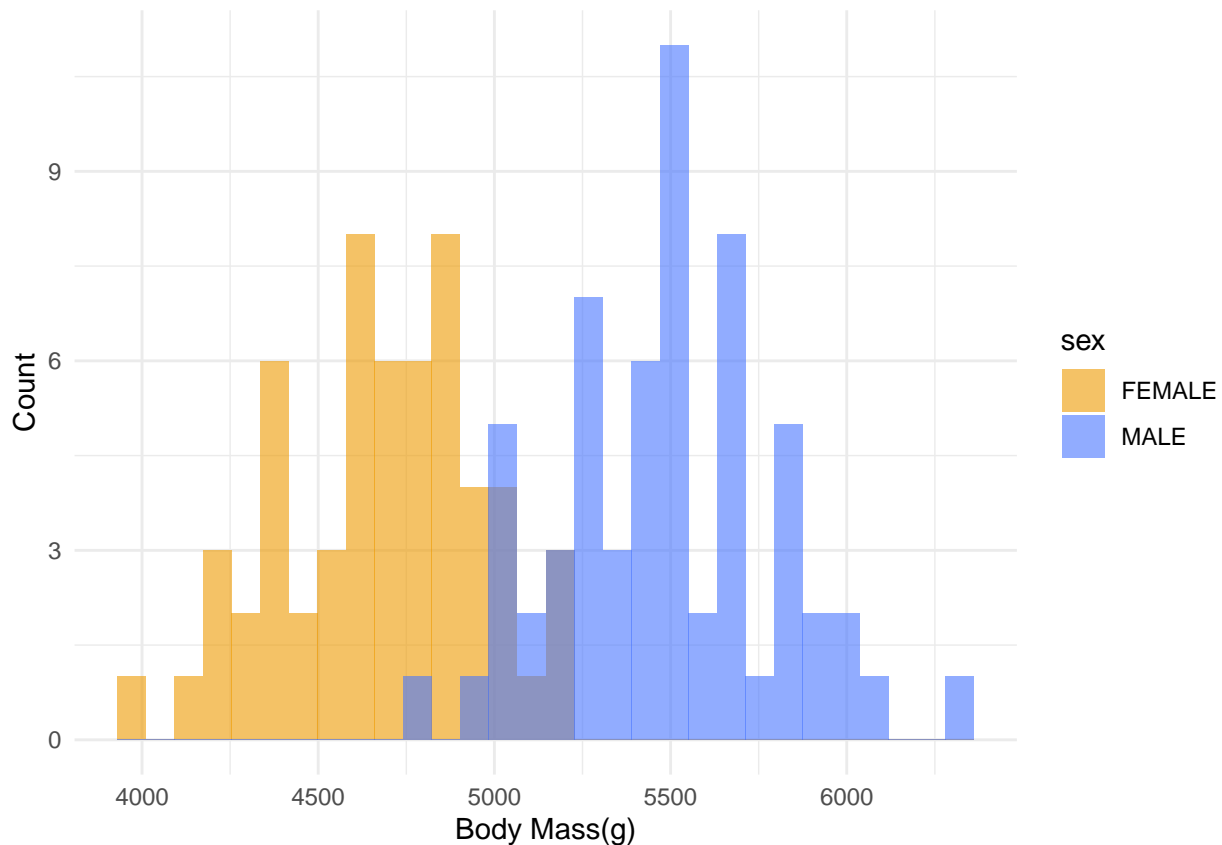
```
#Saving histogram as .png
save_as_png(clean_gentoo_body_mass_sex,
            here("figures", "histogram.png"),
            histogram,
            size = 20, res = 300, scaling = 2)
```

```
## pdf
##   2
```

```
#Saving histogram as .svg
save_as_svg(clean_gentoo_body_mass_sex,
            here("figures", "histogram.svg"),
            histogram,
            size = 20, scaling = 2)
```

```
## pdf
##   2
```

Defining the linear model

```
#Defining linear model
linear_model <- lm(body_mass_g ~ sex, data = clean_gentoo_body_mass_sex)
```

Checking for assumptions of linear model analysis

Before conducting the linear model analysis, it is essential to verify that the data meets the assumptions of: 1) independence, 2) normality, 3) homoscedasticity, and 4) absence of influential outliers. Although I lack information on the data collection methods, I assume that the data points are independent of each other. To check for the remaining assumptions, I examined diagnostic plots displayed in a multi-panel grid. The Q-Q residuals plot indicates that the residuals are normally distributed, as the data points closely follow

the straight line. The residuals versus fitted values and scale-location plots show random scatter and a horizontal line, suggesting homoscedasticity. Finally, the residuals versus leverage plot indicates that the outliers identified in the histogram have a small Cook's distance, implying they are not influential.

```
#Diagnostic plots
#Displaying diagnostic plots in a multi-panel grid and adjusting margin size
par(mfrow = c(2, 2),
    mar = c(4, 4, 2, 2))


#Diagnostic plot function
diagnostic_plot_function()
```
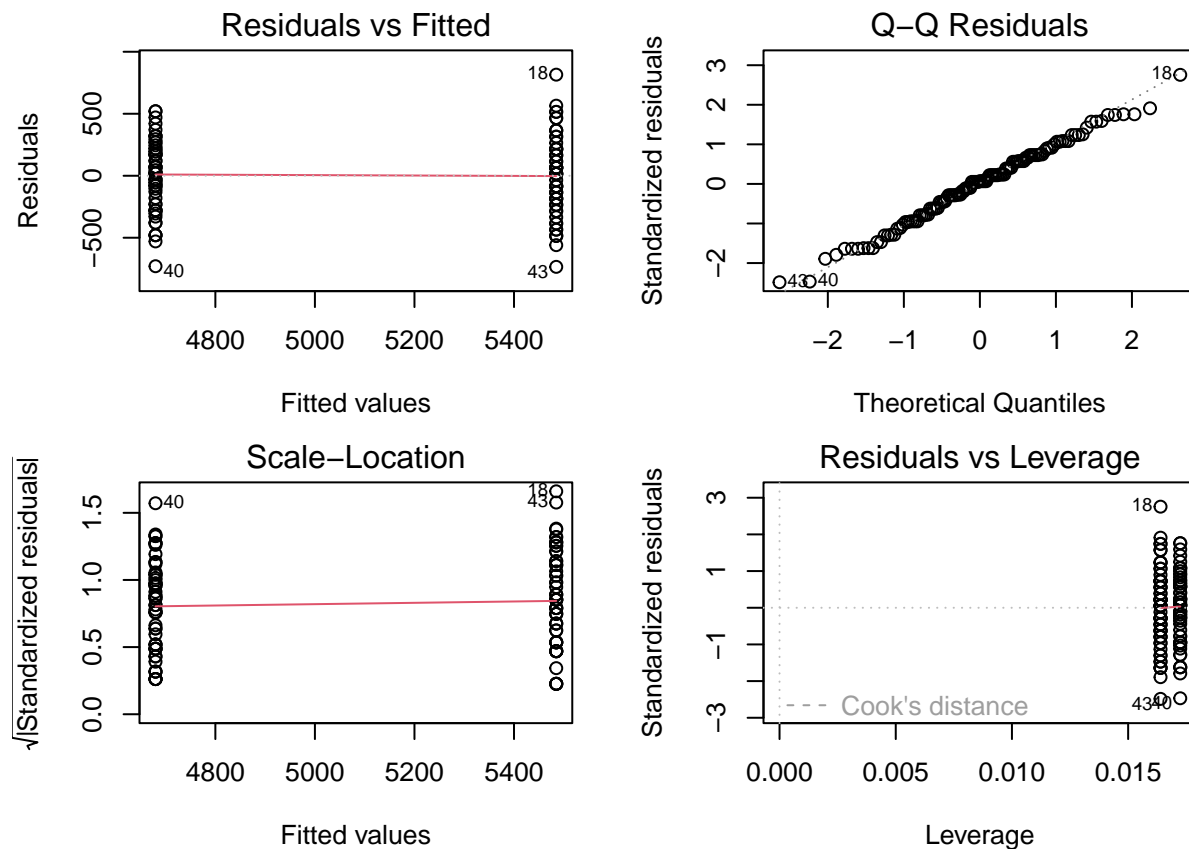


```
# Saving the diagnostic plot as a .png
save_diagnostic_plot_as_png(
  here("figures", "diagnostic_plot.png"),
  diagnostic_plot_function,
  size = 20,
  res = 300,
  scaling = 2
)
```

```
## pdf
##   2
```

```
# Saving the diagnostic plot as a .svg
save_diagnostic_plot_as_svg(
  here("figures", "diagnostic_plot.svg"),
  diagnostic_plot_function,
```

```
    size = 20,
    scaling = 2
)
```

pdf
2

RESULTS

Given that all the assumptions for the linear model were met, I proceeded with the linear model analysis. The null hypothesis is: if there was no size difference, we would expect the coefficient from a linear model analysis to be zero. The alternative hypothesis is: if there was a size difference, we would expect the coefficient from a linear model analysis to be significantly different form zero. The results from the linear model show that male Gentoo penguins are, on average, 805.09 grams heavier than females. This difference is statistically significant (p < 0.05) so we can reject the null hypothesis. The figure displaying the results of the linear model analysis also clearly show a regression line with an increasing slope, indicating an increase in body mass from female to male sex. Additionally, the multiple R-squared value of 0.6494 suggests that sex explains a substantial portion of the variance in body mass, though other factors likely contribute as well.

```
#Linear model analysis
summary(linear_model)
```

```
##
## Call:
## lm(formula = body_mass_g ~ sex, data = clean_gentoo_body_mass_sex)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -734.84 -207.29   20.26  215.16  815.16
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4679.74      39.15  119.52   <2e-16 ***
## sexMALE       805.09      54.69   14.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 298.2 on 117 degrees of freedom
## Multiple R-squared:  0.6494, Adjusted R-squared:  0.6464
## F-statistic: 216.7 on 1 and 117 DF,  p-value: < 2.2e-16
```

```
#Linear model analysis summary table
#Firstly, converting model into a clean structure and extracting key details
linear_model_summary <- tidy(linear_model)

#Summary table - estimate, SE, and statistic rounded to three decimal points and p-values to 4 significc
linear_model_summary_table <- linear_model_summary %>%
  select(term, estimate, std.error, statistic, p.value) %>%
  mutate(
    estimate = round(estimate, 3),
    std.error = round(std.error, 3),
    statistic = round(statistic, 3),
    p.value = format(p.value, scientific = TRUE, digits = 4))

print(linear_model_summary_table)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl> <chr>
## 1 (Intercept)    4680.      39.2     120.  "3.888e-124"
## 2 sexMALE         805.      54.7      14.7 " 2.134e-28"
```
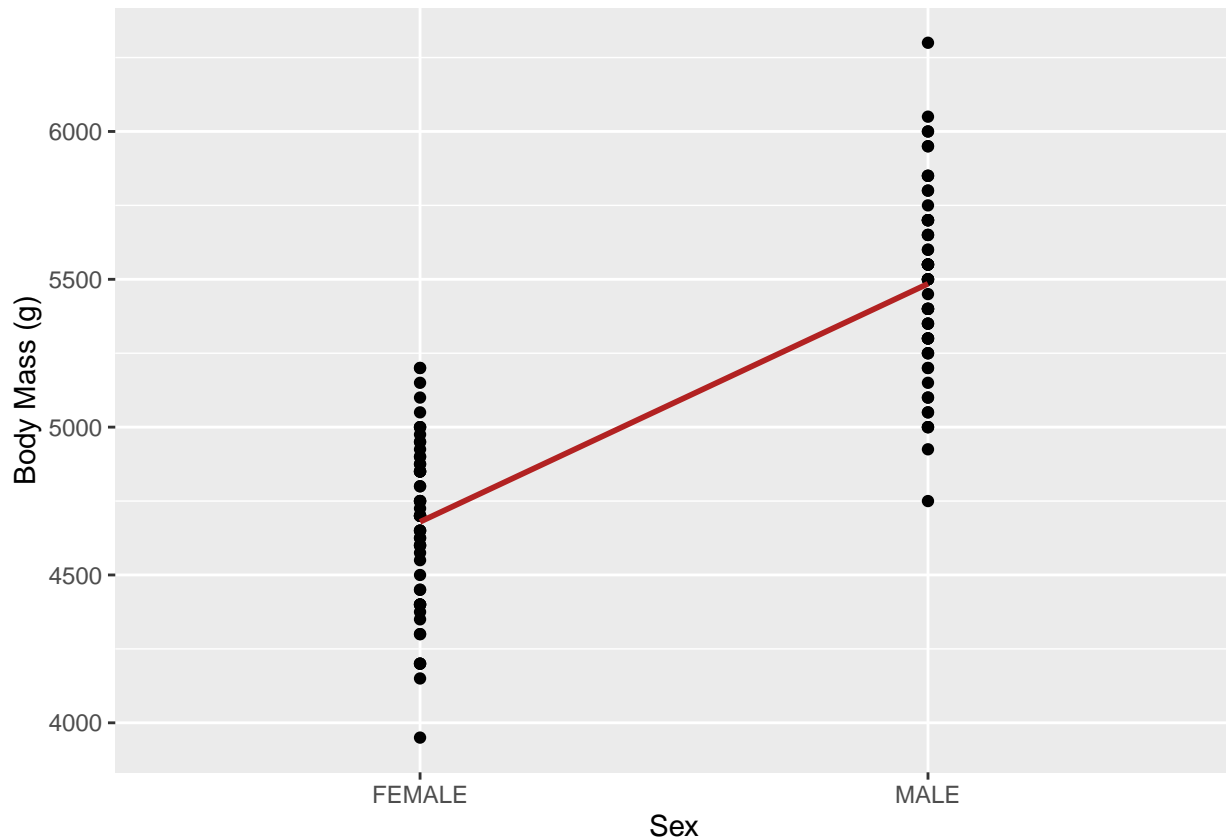
```
#Results figure function
linear_model_results_figure <-linear_model_figure(clean_gentoo_body_mass_sex, sex, body_mass_g, "Sex",

linear_model_results_figure
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
#Saving results figure as .png
save_as_png(clean_gentoo_body_mass_sex,
          here("figures", "results_figure.png"),
          linear_model_results_figure,
          size = 20, res = 300, scaling = 2)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## pdf
##   2
```

```
#Saving results figure as .svg
save_as_svg(clean_gentoo_body_mass_sex,
          here("figures", "results_figure.svg"),
          linear_model_results_figure,
          size = 20, scaling = 2)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## pdf
##   2
```

DISCUSSION

In this study, I examined whether sexual size dimorphism is present in Gentoo penguins by analysing body mass data from the palmerpenguins data set. The results from the linear model analysis that I carried out after verifying it's assumptions indicate a significant difference in body mass between male and female Gentoo penguins. On average, the males were 805.09 grams heavier. Therefore this result supports the hypothesis that sexual dimorphism is present in this species

This size difference is consistent with other studies looking at different penguin species (Poisbleau et al., 2010; Koehn et al., 2016). Furthermore, the large multiple R-squared value (0.6494) suggests that sex is an important factor in determining the body mass in Gentoo penguins. However, it is important to note that other variables, such as food availability, determining body mass have not been accounted for. Hence, future studies could include additional covariates to better explain the variability in body mass.

Another potential area for future research could be to explore the potential reasons behind this sexual size dimorphism. For example, the breeding ecology of these penguins could mean that males are better competitors for mating opportunities when they have a larger body mass. As a result, sexual selection over evolutionary time may have selected for males to have a larger body size in this species. Investigating the relationship between body mass and a potential behavioural traits as just described could provide additional insights into the evolutionary pressures that shape size dimorphism in this species.

Finally it is important to note that although the linear model met most of the necessary assumptions there is still the main limitation that I have presumed independence even though the data could have been collected in a way that could introduce biases between individual measurements.

CONCLUSION

This study confirms the presence of sexual size dimorphism in the Gentoo penguin species, where males are significantly larger than females by an average of 805.09 grams. The linear model analysis shown that this variation was largely explained by sex with a high multiple R-squared value (0.6494) but it should be noted that other factors not considered in this study will affect body mass. Whilst most of the assumptions for the linear model were met, independence had to be presumed based off the lack of knowledge on the methodology carried out to obtain body mass measurements for the different sexes. This is therefore a key limitation of the study. Future research should incorporate other covariates that would affect body size, such as food availability, as well as looking at the evolutionary drivers of sexual size dimorphism in this species. Overall, this study provides evidence for the presence of sexual size dimorphism in the Gentoo penguin species.

REFERENCES

Koehn, L.E., Hard, J.J., Akst, E.P., Boersma, D.P. (2016). Natural selection on morphology varies among years and by sex in Magellanic Penguins (Spheniscus magellanicus), The Auk, 133(4): pp 783-805. https://doi.org/10.1642/AUK-16-50.1

Poisbleau, M., Demongin, L., van Noordwijk, H.J., Strange, I.J., Quillfeldt, P. (2010). Sexual Dimorphism and use of Morphological Measurements to Sex Adults, Immatures and Chicks of Rockhopper Penguins, Ardea, 98(2): pp 217-224. https://doi.org/10.5253/078.098.0212

Ralls, K., Mesnick, S. (2009) Sexual Dimorphism, Enclyclopedia of Marine Mammals, pp 1005-1011. http://dx.doi.org/10.1016/B978-0-12-373553-9.00233-9

QUESTION 3: Open Science

   a) GitHub

GitHub link: https://github.com/oxfordstudent02/ReproducableFigures

   b) Share your repo with a partner and try to run their data pipeline

Partner's GitHub link: https://github.com/benaiahhoong/ReproducibleAssignment

c) Reflect on your experience running their code

Elements of my partner's code that helped me to understand and run their data pipeline include that they broke their coding up into logical chunks and provided a concise explanation for what each section was being used to code for. This was a good way of setting their work out as it provided a clear step by step process that could be followed along with relative ease. In addition to headings at the top of their code, my partner also annotated most lines of their code with brief explanations for as to what that line was coding for. This is practical because it means that if I wanted to change a variable in a figure for example, I would be able to locate the line of code faster than if there were no annotations. When clearing the data, my partner also made use of piping to prevent overwriting. Furthermore, they also used cleaning functions which helped to improve the readability of the code.

My partners code ran well and nothing was needed to be fixed. Suggestions I would make for improving their code to make it more understandable or reproducible include changing the code in the statistics section, where my partner could have directly filtered the adelie_data inside the pipe chain instead of creating a separate variable in order to improve the efficiency of his code. Additionally, my partner could have made use of more functions in their project to make their project more scalable, say if they were working in a larger team. For example, if they had made a function for producing a scatterplot it would have made it easier for somebody else to produce a similar scatterplot but looking at the relationship between different variables. On top of that, including functions would improve the readability of the code because it allows you to abstract complex tasks into a simple function call.

The code used by my partner to produce the figures in their assignment are clearly separated form other chunks of code which makes it easy to identify which lines of code I would be working with. For most lines of code there is adequate commenting that tells me which part of the code contributes to which feature of the figure. Therefore if I was unhappy with a particular aspect of a figure I would be able to quickly find the line of code that I would need to alter. Although custom functions are useful in general it can be slightly hard to get to use with at first. However, looking at their function file it is clear that my partner has annotated each function well so it would not be to hard and overtime using the function itself would save more time than having to learn how to use it. Overall, I think that if I needed to alter my partners figure using their code it would be quite easy.

d) Reflect on your own code based on your experience with your partner's code and their review of yours

My partner thought that because I only used some of my functions once it would be better to include the source code for the function in the main code instead as it would better explain the input used to generate the figures. I disagree with this improvement because it is but a minor inconvenience to check the functions file and I think that the improved readability of the code as a result of using a function outweighs the potential downsides. Another improvement suggested by my partner related to how it might not seem apparent to how changes in the element of the function may correspond to changes in the output. I do agree with this comment because I could have commented in my code whether the x axes or the y axes came first and similar could be said for the x axes and y axes labels. Whilst this problem still could have been solved quickly by looking at the function file, it would be slightly more efficient to improve my comments in the main code.

I learned that in order to write code for other people organisation is important. For example, it is important to simply organise your files in a logical manner to make it easy to open and store items. It is also important to use a clear, modular structure and annotate my lines of code so that people can understand the step by step process going on in addition to making any changes they might seem necessary for their analysis. Learning how to use piping and creating functions helped me to develop techniques to improve the readability of my code. Furthermore, learning how to write functions helped me to build the skills needed to make my coding more reusable for other people because it helps to avoid rewriting code for similar figures. Using print statements in the function can make it extra clear to what the function is carrying out. Before this assignment I also didn't realise that by using renv, we are able to keep a record of all the libraries we have installed in our code so that others are able to quickly access them once again making our code more reusable.