

ReproducibleResearching

C'est Moi, Evidemment

04/12/2023

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

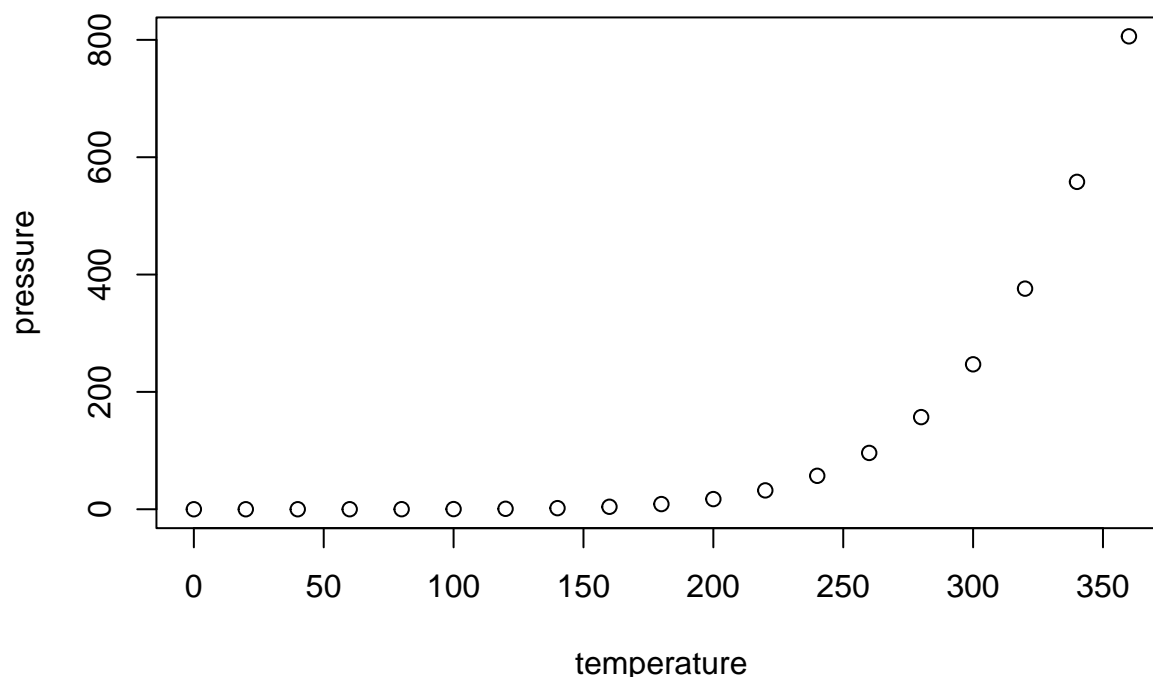
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

QUESTION 01: Data Visualisation for Science Communication a) Provide your figure here:

```
## loading packages
library(tinytex)
install.packages(c("ggplot2", "palmerpenguins", "janitor", "dplyr", repos = "http://cran.us.r-project.org"))
```

```
## Warning: package 'http://cran.us.r-project.org' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
##
## There is a binary version available but the source version is later:
## binary source needs_compilation
## dplyr 1.1.3 1.1.4 TRUE
##
##
## The downloaded binary packages are in
## /var/folders/tr/6jq1jcx50nb9n64nnf6qwspw0000gn/T//RtmpAz8JsE/downloaded_packages
```

```
## installing the source package 'dplyr'
```

```
install.packages("knitr")
```

```
##  
##   There is a binary version available but the source version is later:  
##       binary source needs_compilation  
## knitr   1.44   1.45                FALSE
```

```
## installing the source package 'knitr'
```

```
install.packages("latexpdf")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/tr/6jq1jcx50nb9n64nnf6qwspw0000gn/T//RtmpAz8JsE/downloaded_packages
```

```
install.packages("car")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/tr/6jq1jcx50nb9n64nnf6qwspw0000gn/T//RtmpAz8JsE/downloaded_packages
```

```
library(car)
```

```
## Loading required package: carData
```

```
library(latexpdf)  
library(knitr)  
library(ggplot2)  
library(palmerpenguins)  
library(janitor)
```

```
##  
## Attaching package: 'janitor'
```

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

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':  
##  
##   recode
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
## checking the penguin data loaded
head(penguins_raw)
```

```
## # A tibble: 6 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
## # 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>
```

```
write.csv(penguins_raw, "data/penguins_raw.csv")
names(penguins_raw)
```

```
## [1] "studyName"      "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"             "Delta 15 N (o/oo)"
## [16] "Delta 13 C (o/oo)" "Comments"
```

```
## my functions that will be used
```

```
clean_column_names <- function(penguins_data) {
  penguins_data %>%
    clean_names()
}
```

```
shorten_species <- function(penguins_data) {
  penguins_data %>%
    mutate(species = case_when(
      species == "Adelie Penguin (Pygoscelis adeliae)" ~ "Adelie",
      species == "Chinstrap penguin (Pygoscelis antarctica)" ~ "Chinstrap",
      species == "Gentoo penguin (Pygoscelis papua)" ~ "Gentoo"
    ))
}
```

```
remove_empty_columns_rows <- function(penguins_data) {
  penguins_data %>%
```

```

    remove_empty(c("rows", "cols"))
  }

subset_columns <- function(penguins_data, column_names) {
  penguins_data %>%
    select(all_of(column_names))
}

filter_by_species <- function(penguins_data, selected_species) {
  penguins_data %>%
    filter(species == selected_species)
}

remove_NA <- function(penguins_data) {
  penguins_data %>%
    na.omit()
}

## cleaning the penguin data
penguins_clean <- penguins_raw %>%
  clean_column_names() %>%
  shorten_species() %>%
  remove_empty_columns_rows

write.csv(penguins_clean, "Data/penguins_clean.csv")
names(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"             "delta_15_n_o_oo"
## [16] "delta_13_c_o_oo" "comments"

#View(penguins_clean)

## creating my graph
bad_culmen_scatter <- ggplot(data = penguins_clean,
                             aes(x = culmen_length_mm,
                                 y = culmen_depth_mm)) +
  geom_point(color = "blue", alpha = 0.5, show.legend = FALSE) +
  geom_smooth(method = "lm", se = FALSE) +
  xlab("c_length") +
  ylab("c_depth") +
  scale_x_continuous(limits = c(0, 200), breaks = seq(0, 200, by = 100)) +
  scale_y_continuous(limits = c(10, 25), breaks = seq(10, 25, by = 7)) +
  theme_minimal() # You can customize the theme if needed

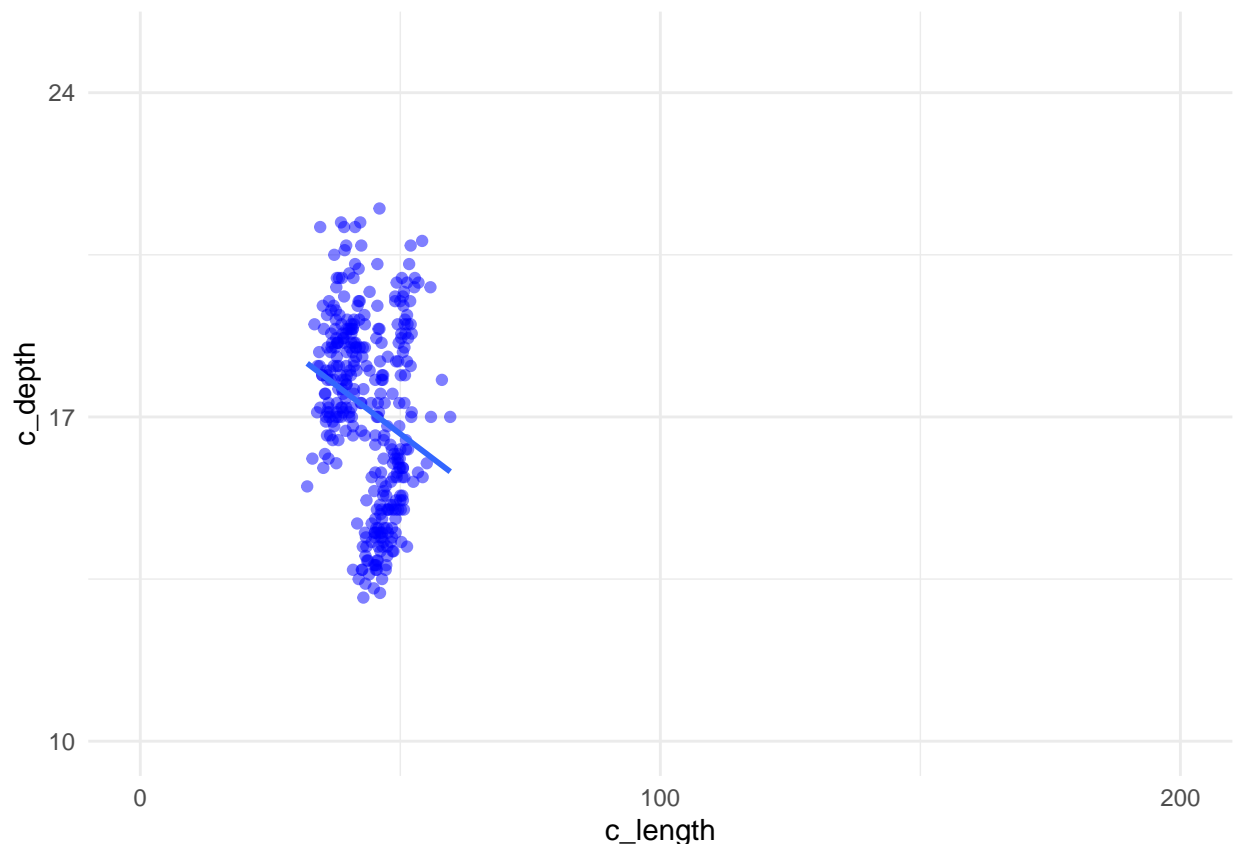
bad_culmen_scatter

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

```

```
## Warning: Removed 2 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 2 rows containing missing values ('geom_point()').
```



b) Write about how your bad design choices misled the reader about the underlying data. Firstly and perhaps most importantly, I have chosen to ignore the fact that the data is actually comprised of three species – instead representing all the datapoints the same colour and without a legend so that they appear to be the same. This obscures the truth that the data actually contains three separate positive correlations and suggests that it is in fact one negative correlation between the variables – a misconception which I have developed through the addition of a linear regression line which supports this. Additionally, I have made the points the same colour as the line, hiding it slightly, and changed their opacity, making them harder to see and distinguish. Further, I have truncated the x and y axis as Dreissen et al. (2022) show that truncated axis can skew an observers interpretations about the data they are presented with. This has led to the data appearing to be more tightly clustered on the x axis and further spread on the y than they would if represented truly, obscuring the true correlation. I have also used large gaps between the gridlines so it is harder to tell where in them the points are actually representing and made the y axis gridlines a multiple of 7, which further complicates reading the data. Finally, I have also changed the axis labels, removing the (mm) so that it is unclear what is actually being measured, furthering this by abbreviating the labels to “c_length” and “c_depth” so that it is unclear what part of the penguin is actually being measured – the culmen. Overall, the graph is hard to read, and even when read correctly misrepresents the true dataset.

```
##QUESTION 2: Data Pipeline
```

```
###Introduction:
```

Loading the data: Before I begin, I must load the data so the first line of my code does so from the ‘palmerpenguins’ package. Then, I save my raw data to a safe csv file so that I, or anybody else, can always

refer back to it and it is not lost as I begin to edit it. Finally, I check the names of the columns in the dataset.

Cleaning the data: Here, I call a function which I defined earlier to clean the data, removing superfluous columns and renaming the remaining columns to be more code friendly. Additionally, I shorten the species names from, for example, “Adelie Penguin (*Pygoscelis adeliae*)” to simply “Adelie” and remove any empty columns and rows for ease of data handling.

Creating an exploratory figure: I was interested to see if there was a relationship between culmen length and depth, so created a scatterplot using ggplot to see the data. Crucially, unlike in the bad graph earlier, I separated the points based on species, which left me with a scatter graph that appeared to be showing three separate positive correlations, one for each species – each very similar. This, I thought was worth exploring further to see if these correlations were statistically backed, or if they appeared to be so but were not actually correlated.

```
#load the data
```

```
library(palmerpenguins)
write.csv(penguins_raw, "data/penguins_raw.csv")
names(penguins_raw)
```

```
## [1] "studyName"      "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"              "Delta 15 N (o/oo)"
## [16] "Delta 13 C (o/oo)" "Comments"
```

```
#clean the data
```

```
penguins_clean <- penguins_raw %>%
  clean_column_names() %>%
  shorten_species() %>%
  remove_empty_columns_rows()

write.csv(penguins_clean, "Data/penguins_clean.csv")
names(penguins_clean)
```

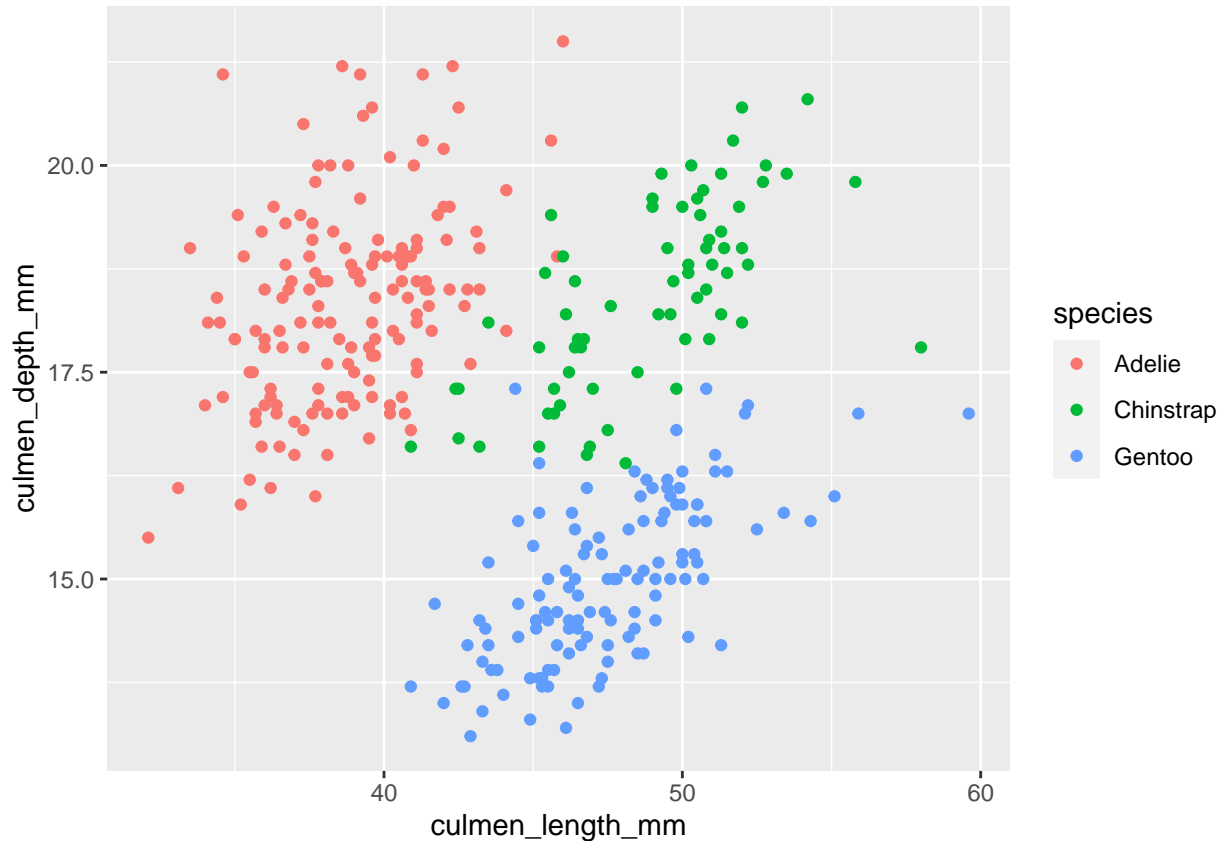
```
## [1] "study_name"      "sample_number"    "species"
## [4] "region"          "island"           "stage"
## [7] "individual_id"   "clutch_completion" "date_egg"
## [10] "culmen_length_mm" "culmen_depth_mm"  "flipper_length_mm"
## [13] "body_mass_g"     "sex"              "delta_15_n_o_oo"
## [16] "delta_13_c_o_oo" "comments"
```

```
#create an exploratory figure
```

```
culmen_scatter <- ggplot(data = penguins_clean,
  aes(x = culmen_length_mm,
      y = culmen_depth_mm)) +
  geom_point(aes(color = species), show.legend = TRUE)

culmen_scatter
```

```
## Warning: Removed 2 rows containing missing values ('geom_point()').
```



```
#save the figure
ggsave("figures/culmenfigure_vector.svg", culmen_scatter, width = 5.9, height = 5.9)
```

```
## Warning: Removed 2 rows containing missing values ('geom_point()').
```

###Hypotheses: 1. Culmen length and depth are positively correlated in all three species separately 2. The correlation is not 1:1 for depth:length across all species

##Stats Method: Hypothesis 1: First I grouped all of the data by species, allowing me to do a separate linear regression for each species and thus to determine if each showed a positive correlation. I then completed a linear regression for each species to work out if there was such a correlation, if it was statistically significant and how much of the variance in the data it explains.

Hypothesis 2: To test whether the correlation was 1:1 in nature I fitted a linear mixed effects model to the data. This once again tested for a positive correlation, accounting for variability caused by species, and looked at how this changed as the intercept, culmen depth, increased.

```
# run a statistical test
Adelie_data <- subset(penguins_clean, species == "Adelie")
Adelie_data
```

```
## # A tibble: 152 x 17
##   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 142 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>
```

```
Gentoo_data <- subset(penguins_clean, species == "Gentoo")
Gentoo_data
```

```
## # A tibble: 124 x 17
##   study_name sample_number species region island stage      individual_id
##   <chr>          <dbl> <chr>   <chr>  <chr>  <chr>      <chr>
## 1 PAL0708          1 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N31A1
## 2 PAL0708          2 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N31A2
## 3 PAL0708          3 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N32A1
## 4 PAL0708          4 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N32A2
## 5 PAL0708          5 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N33A1
## 6 PAL0708          6 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N33A2
## 7 PAL0708          7 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N34A1
## 8 PAL0708          8 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N34A2
## 9 PAL0708          9 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N35A1
## 10 PAL0708         10 Gentoo  Anvers Biscoe Adult, 1 Egg St~ N35A2
## # i 114 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>
```

```
Chinstrap_data <- subset(penguins_clean, species == "Chinstrap")
Chinstrap_data
```

```
## # A tibble: 68 x 17
##   study_name sample_number species region island stage      individual_id
##   <chr>          <dbl> <chr>   <chr>  <chr>  <chr>      <chr>
## 1 PAL0708          1 Chinstrap Anvers Dream Adult, 1 Egg ~ N61A1
## 2 PAL0708          2 Chinstrap Anvers Dream Adult, 1 Egg ~ N61A2
## 3 PAL0708          3 Chinstrap Anvers Dream Adult, 1 Egg ~ N62A1
## 4 PAL0708          4 Chinstrap Anvers Dream Adult, 1 Egg ~ N62A2
## 5 PAL0708          5 Chinstrap Anvers Dream Adult, 1 Egg ~ N64A1
## 6 PAL0708          6 Chinstrap Anvers Dream Adult, 1 Egg ~ N64A2
## 7 PAL0708          7 Chinstrap Anvers Dream Adult, 1 Egg ~ N66A1
## 8 PAL0708          8 Chinstrap Anvers Dream Adult, 1 Egg ~ N66A2
## 9 PAL0708          9 Chinstrap Anvers Dream Adult, 1 Egg ~ N67A1
## 10 PAL0708         10 Chinstrap Anvers Dream Adult, 1 Egg ~ N67A2
## # i 58 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>
```

```
lm_Adelie <- lm(culmen_length_mm ~ culmen_depth_mm, data = Adelie_data)
summary(lm_Adelie)
```

```
##
## Call:
## lm(formula = culmen_length_mm ~ culmen_depth_mm, data = Adelie_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5513 -1.8016  0.0055  1.6771  6.5341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      23.068      3.034   7.603 3.01e-12 ***
## culmen_depth_mm    0.857      0.165   5.193 6.67e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 149 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.1533, Adjusted R-squared:  0.1476
## F-statistic: 26.97 on 1 and 149 DF, p-value: 6.674e-07
```

```
lm_Gentoo <- lm(culmen_length_mm ~ culmen_depth_mm, data = Gentoo_data)
summary(lm_Gentoo)
```

```
##
## Call:
## lm(formula = culmen_length_mm ~ culmen_depth_mm, data = Gentoo_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.7888 -1.4097  0.1361  1.3882  8.0174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      17.2295      3.2818   5.250 6.60e-07 ***
## culmen_depth_mm    2.0208      0.2186   9.245 1.02e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.369 on 121 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.4139, Adjusted R-squared:  0.4091
## F-statistic: 85.46 on 1 and 121 DF, p-value: 1.016e-15
```

```
lm_Chinstrap <- lm(culmen_length_mm ~ culmen_depth_mm, data = Chinstrap_data)
summary(lm_Chinstrap)
```

```
##
## Call:
## lm(formula = culmen_length_mm ~ culmen_depth_mm, data = Chinstrap_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.1163 -1.2641 -0.1254  1.4807 10.3590
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.428      5.057   2.655  0.00992 **
## culmen_depth_mm    1.922      0.274   7.015 1.53e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.547 on 66 degrees of freedom
## Multiple R-squared:  0.4271, Adjusted R-squared:  0.4184
## F-statistic: 49.21 on 1 and 66 DF,  p-value: 1.526e-09
```

```
install.packages("lme4")
```

```
## also installing the dependency 'RcppEigen'
```

```
##
##   There are binary versions available but the source versions are later:
##           binary      source needs_compilation
## RcppEigen 0.3.3.9.3 0.3.3.9.4             TRUE
## lme4       1.1-34   1.1-35.1              TRUE
```

```
## installing the source packages 'RcppEigen', 'lme4'
```

```
## Warning in install.packages("lme4"): installation of package 'RcppEigen' had
## non-zero exit status
```

```
## Warning in install.packages("lme4"): installation of package 'lme4' had
## non-zero exit status
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
# Fit mixed-effects model
mixed_model <- lmer(culmen_length_mm ~ culmen_depth_mm + (1 | species), data = penguins_clean)

# Display mixed-effects model summary
summary(mixed_model)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: culmen_length_mm ~ culmen_depth_mm + (1 | species)
## Data: penguins_clean
##
## REML criterion at convergence: 1618
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.1860 -0.6286  0.0352  0.6720  3.9847
##
## Random effects:
## Groups Name Variance Std.Dev.
## species (Intercept) 48.322  6.951
## Residual          6.338  2.518
## Number of obs: 342, groups: species, 3
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    21.0953     4.5328  4.654
## culmen_depth_mm  1.3883     0.1219 11.393
##
## Correlation of Fixed Effects:
##              (Intr)
## clmn_dpth_m -0.464
```

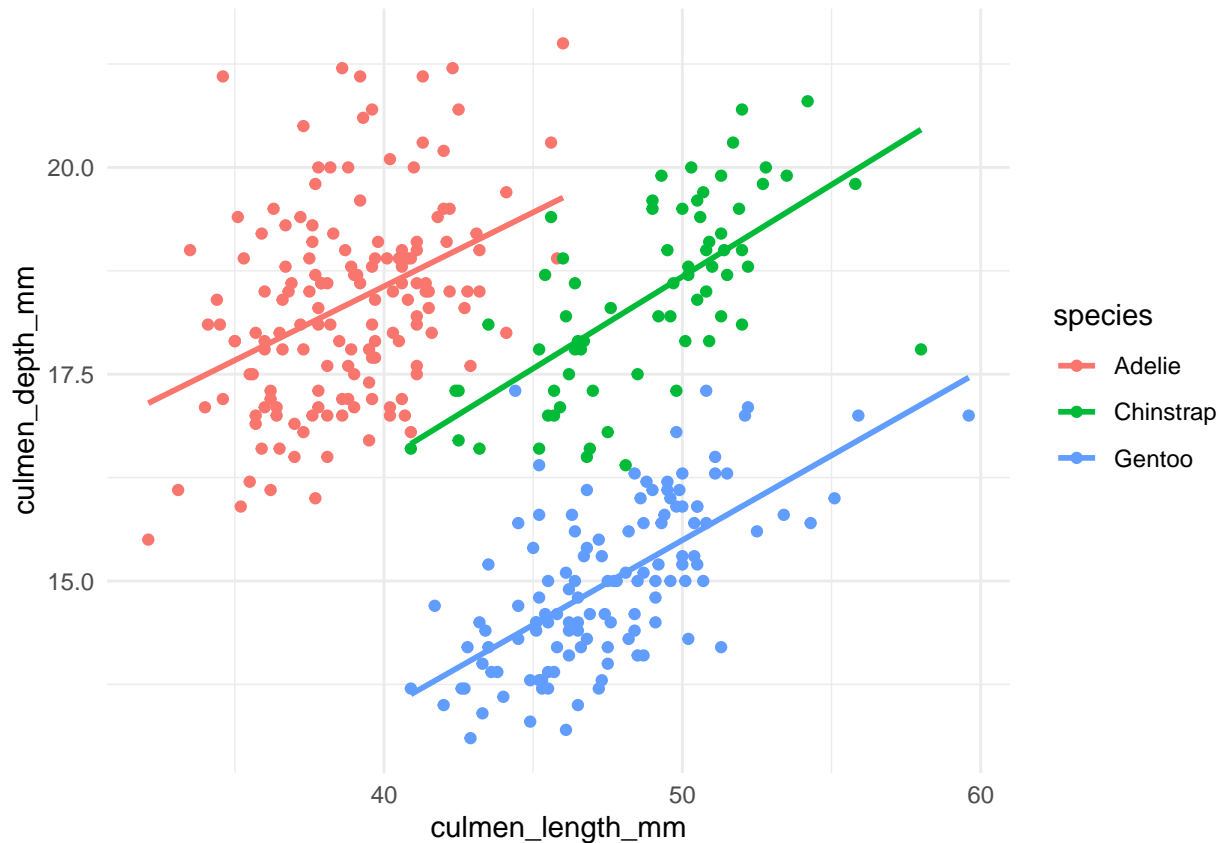
```
# create a results figure
culmen_scatter2 <- ggplot(data = penguins_clean, aes(x = culmen_length_mm, y = culmen_depth_mm)) +
  geom_point(aes(color = species), show.legend = TRUE) +
  geom_smooth(aes(group = species, color = species), method = "lm", se = FALSE) +
  theme_minimal()

culmen_scatter2
```

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

```
## Warning: Removed 2 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 2 rows containing missing values ('geom_point()').
```



```
# save the figure
ggsave("figures/culmenfigure2_vector.svg", culmen_scatter2, width = 5.9, height = 5.9)
```

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

```
## Warning: Removed 2 rows containing non-finite values ('stat_smooth()').
```

```
## Removed 2 rows containing missing values ('geom_point()').
```

##Results: Hypothesis 1: The linear regression on the Adelie data shows that there is a statistically significant positive correlation between culmen length and depth (p-value: 6.674e-07); however, looking at the adjusted R-squared you can see that this only explains around 15% of the variability in the data (Adjusted R-squared: 0.1476). Looking at the Gentoo data, there is once again a statistically significant interaction, this time with p-value: 1.016e-15. With Adjusted R-squared: 0.4091, this suggests that the interaction explains around 41% of the data, much more than for the Adelie penguins. Finally, Chinstrap penguins show statistically significant correlation again, this time with p-value: 1.526e-09 and Adjusted R-squared: 0.4184 suggests that around 42% of the data is explained by this, in line with Gentoo data, once again much more than for Adelie.

Hypothesis 2: From the adjusted R squared values, it is clear that there is a difference in the amount of the data explained by each linear regression; however, each does still show a statistically significant positive linear correlation. This is backed up by the linear mixed effects model which shows that there is a significant positive relationship between culmen depth and length while accounting for random variability between different species. It also shows that species accounts for most of the variance in the data, further confirming the idea that each species has its own very distinct correlation. Finally, the model shows that as culmen depth increases, its effect on culmen length decreases, suggesting that they do not scale at a 1:1 ratio – but instead a 1:1.3383 ratio with length scaling faster than depth.

Discussion: Firstly, it is clear that each species has a very different bill morphology with much of the variance in the data being explained by these differences in species (Variance = 48.322). This may be due to differences in ecological niche and the subsequent optimal beak morphology for foraging in each. The relationship between culmen length and depth explained more variation in both Chinstrap and Gentoo penguins than in Adelie penguins, which may be related to differences in habitat, with Adelies being found primarily on the Antarctic continent whilst Gentoos and Chinstraps are both found on sub-Antarctic islands as well as the north of the continent. Potentially, if beak morphology impacts likelihood of catching prey, then this will be more important in Chinstraps and Gentoos as they will be hunting in likely more fruitful waters; whereas, Adelies may struggle more to find food, being based more so on the continent, and thereby many other factors more associated with their ability to hunt for prey will be more important in determining morphology. The positive correlation between culmen length and depth makes sense biologically, as one would expect larger penguins to have larger bills overall, with increased size across both areas. This study does not correlate this data with penguin size but future studies could explore this area. The proportionally bigger increase in length than depth may be due to longer beaks providing more of an adaptive advantage than deeper beaks, and therefore penguins putting more resources into increasing length. This could be due to longer beaks increasing the range at which a penguin is able to catch prey. However, much of the variance in the data is still not explained by the models and this should be looked into, potentially exploring the relationship between bill morphology and prey type, diving behaviour or other species specific foraging habits. Understanding how bill morphology relates to foraging behaviour may help future efforts to conserve these species under climate change, with these Antarctic residents likely to be disproportionately impacted over the coming years.

Conclusion: Overall, it is clear that culmen length and depth are correlated and vary across species. Future studies should explore how bill morphology relates to factors such as prey preference, prey abundance, foraging strategies and climatic effects such as ocean temperature to help inform future conservation strategies for these penguins.