Penguin Project

Taylor Harris

2023-06-20

Installing and Loading Required Packages

Using tidyverse for manipulation on the palmerpenguins dataset.

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("palmerpenguins")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
```

```
## -- 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)
```

Exploring The Data

Creating a dataset and looking at the first few rows, the columns, and extra details about the penguin table.

```
penguin_df <- data(package = "palmerpenguins")

head(penguins)
```

```
## # A tibble: 6 x 8
##   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
## 1 Adelie  Torgersen           39.1          18.7               181        3750
## 2 Adelie  Torgersen           39.5          17.4               186        3800
## 3 Adelie  Torgersen           40.3          18                 195        3250
## 4 Adelie  Torgersen           NA            NA                  NA          NA
## 5 Adelie  Torgersen           36.7          19.3               193        3450
## 6 Adelie  Torgersen           39.3          20.6               190        3650
```

```
## # i 2 more variables: sex <fct>, year <int>
```

```
colnames(penguins)
```

```
## [1] "species"           "island"            "bill_length_mm"
## [4] "bill_depth_mm"     "flipper_length_mm" "body_mass_g"
## [7] "sex"               "year"
```

```
str(penguins)
```

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
##  $ species          : Factor w/ 3 levels "Adelie","Chinstrap",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ island           : Factor w/ 3 levels "Biscoe","Dream",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ bill_length_mm   : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ bill_depth_mm    : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
##  $ body_mass_g      : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
##  $ sex              : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
##  $ year             : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

By Island

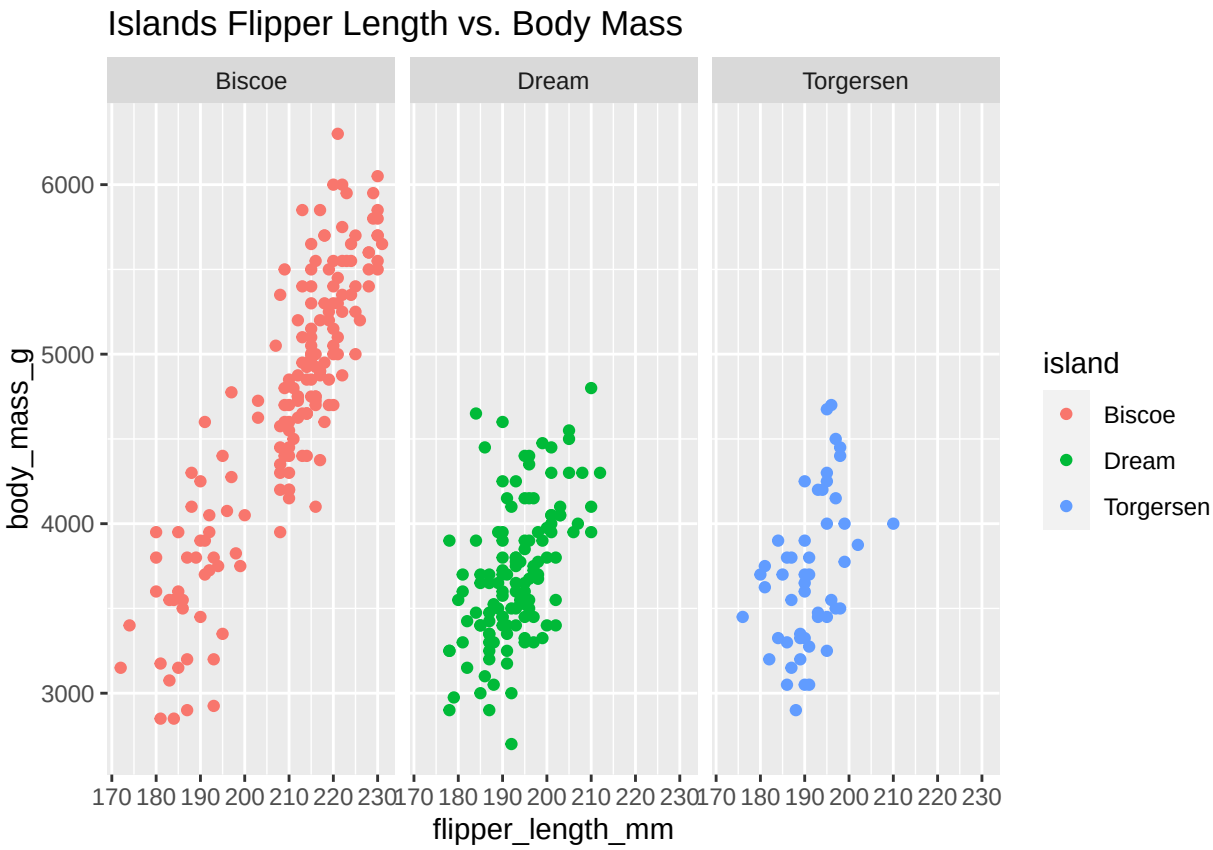
Here I looked at the average flipper length by island in a table and then I compared flipper lengths and body mass in each island.

```
avg_flipper_length_island <- penguins %>%
  group_by(island) %>%
  summarize(mean_flipper_length = mean(flipper_length_mm, na.rm=TRUE))
avg_flipper_length_island
```

```
## # A tibble: 3 x 2
##   island    mean_flipper_length
##   <fct>                   <dbl>
## 1 Biscoe                   210.
## 2 Dream                    193.
## 3 Torgersen                191.
```

```
ggplot(data = penguins) +
  geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g, color = island)) +
  facet_wrap(~island) +
  labs(title = "Islands Flipper Length vs. Body Mass")
```

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



By Species

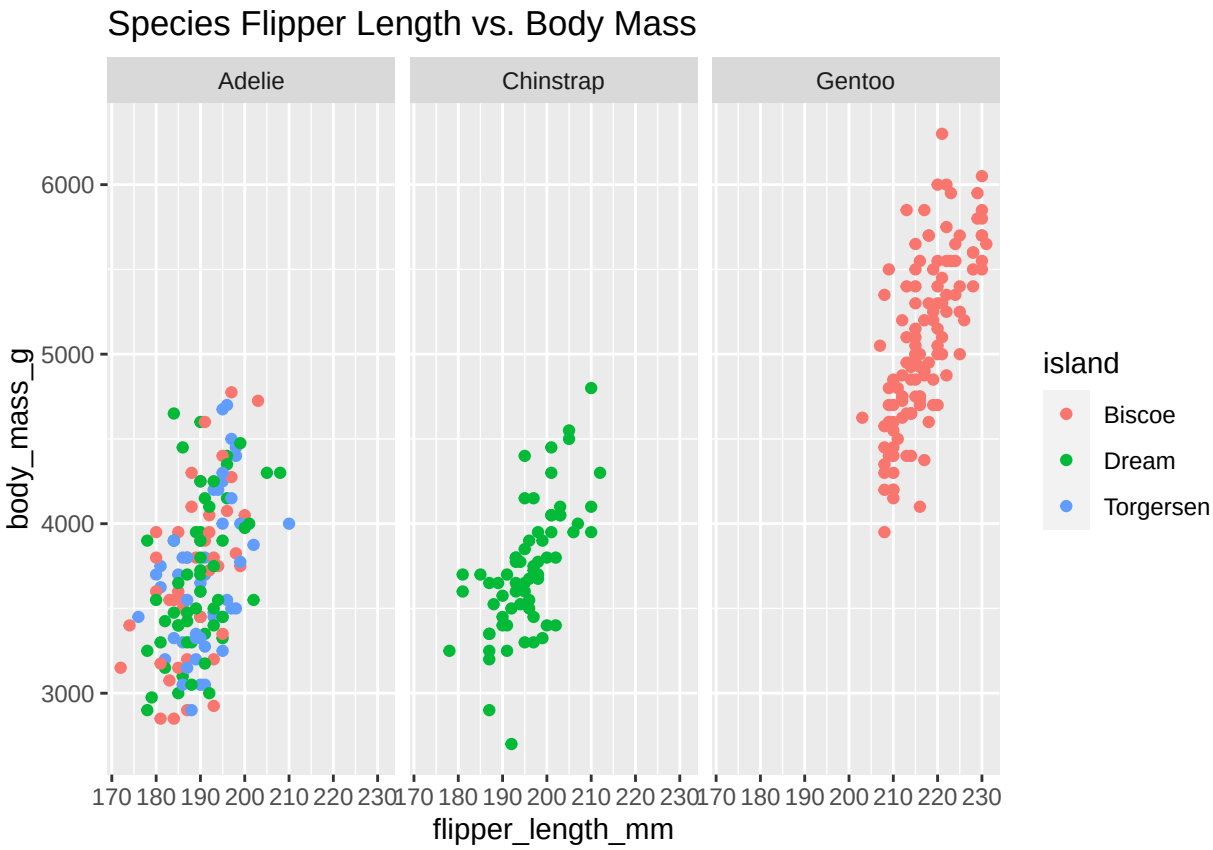
Here I looked at the average flipper length by species in a table and then I compared flipper lengths and body mass in each species

```
avg_flipper_length_species <- penguins %>%
  group_by(species) %>%
  summarize(mean_flipper_length = mean(flipper_length_mm, na.rm=TRUE))
avg_flipper_length_species
```

```
## # A tibble: 3 x 2
##   species   mean_flipper_length
##   <fct>                   <dbl>
## 1 Adelie                   190.
## 2 Chinstrap                196.
## 3 Gentoo                   217.
```

```
ggplot(data = penguins) +
  geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g, color = island)) +
  facet_wrap(~species) +
  labs(title = "Species Flipper Length vs. Body Mass")
```

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

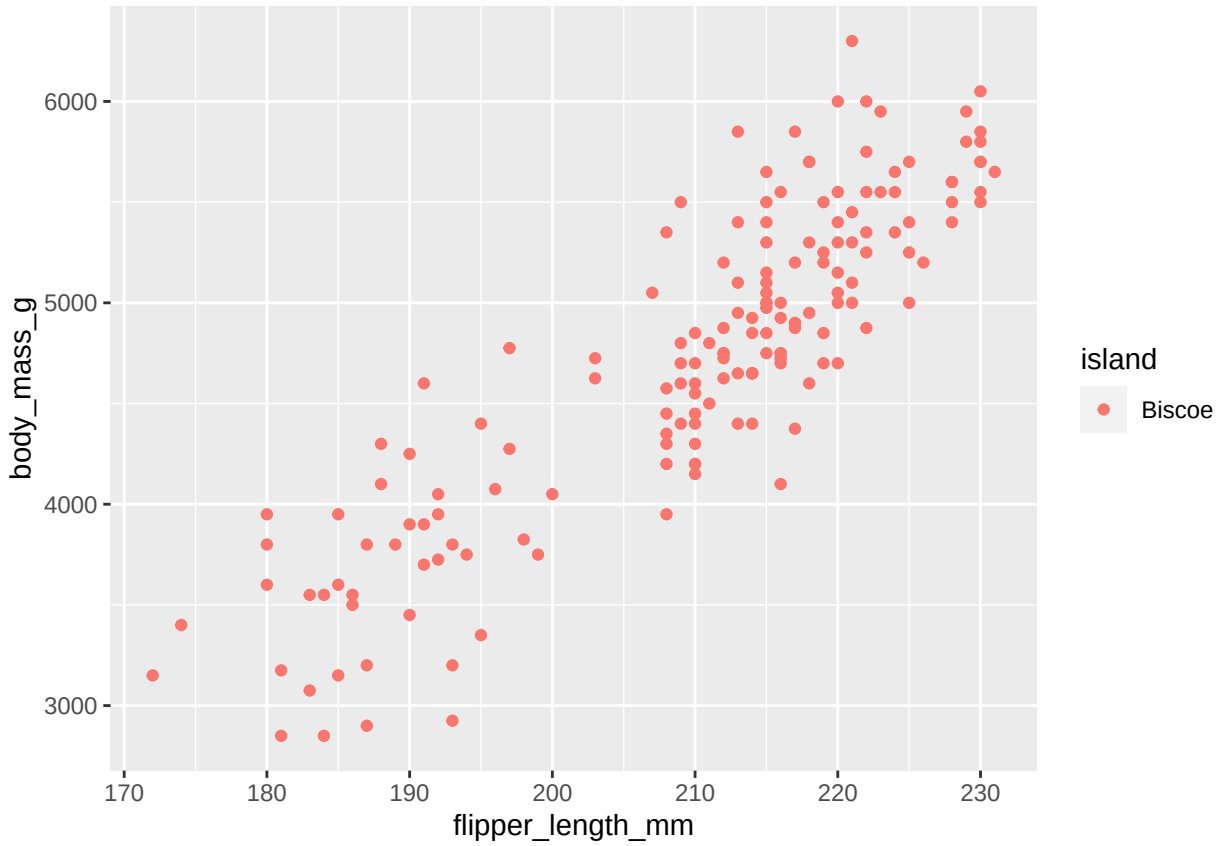


Better Look At The Distribution of Flipper Length and Body Mass On Biscoe Island

Biscoe island has the largest penguins so I zoomed in on that island to get a better idea of the distribution.

```
penguins %>%
  filter(island == "Biscoe") %>%
  ggplot(aes(x = flipper_length_mm, y = body_mass_g, color = island)) +
  geom_point()
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



Largest And Smallest Males And Females

The max and min body mass for each sex.

```
male_female_size <- penguins %>%
  group_by(sex) %>%
  summarize(max_weight_g = max(body_mass_g), min_weight_g = min(body_mass_g))
male_female_size
```

```
## # A tibble: 3 x 3
##   sex    max_weight_g min_weight_g
##   <fct>         <int>        <int>
## 1 female         5200         2700
## 2 male           6300         3250
## 3 <NA>             NA           NA
```

Species Per Island

The number of species per island

```
species_per_island <- penguins %>%
  group_by(island) %>%
  count(species)
species_per_island
```

```
## # A tibble: 5 x 3
## # Groups:   island [3]
##   island    species       n
```

```
##   <fct>     <fct>     <int>
## 1 Biscoe    Adelie       44
## 2 Biscoe    Gentoo      124
## 3 Dream     Adelie       56
## 4 Dream     Chinstrap    68
## 5 Torgersen Adelie       52
```