Palmer Penguins Report

Ebraheem Amashah

1/12/2022

Setting up my environment

Notes: Setting up the R environment by loading "tidyverse" and "palmerpenguins" packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(palmerpenguins)
```

Exploratory Analysis

Notes: I find it helpful to go through column names and some values before manipulating the data. I find this saves me a lot of time and confusion down the line.

```
colnames(penguins)
```

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

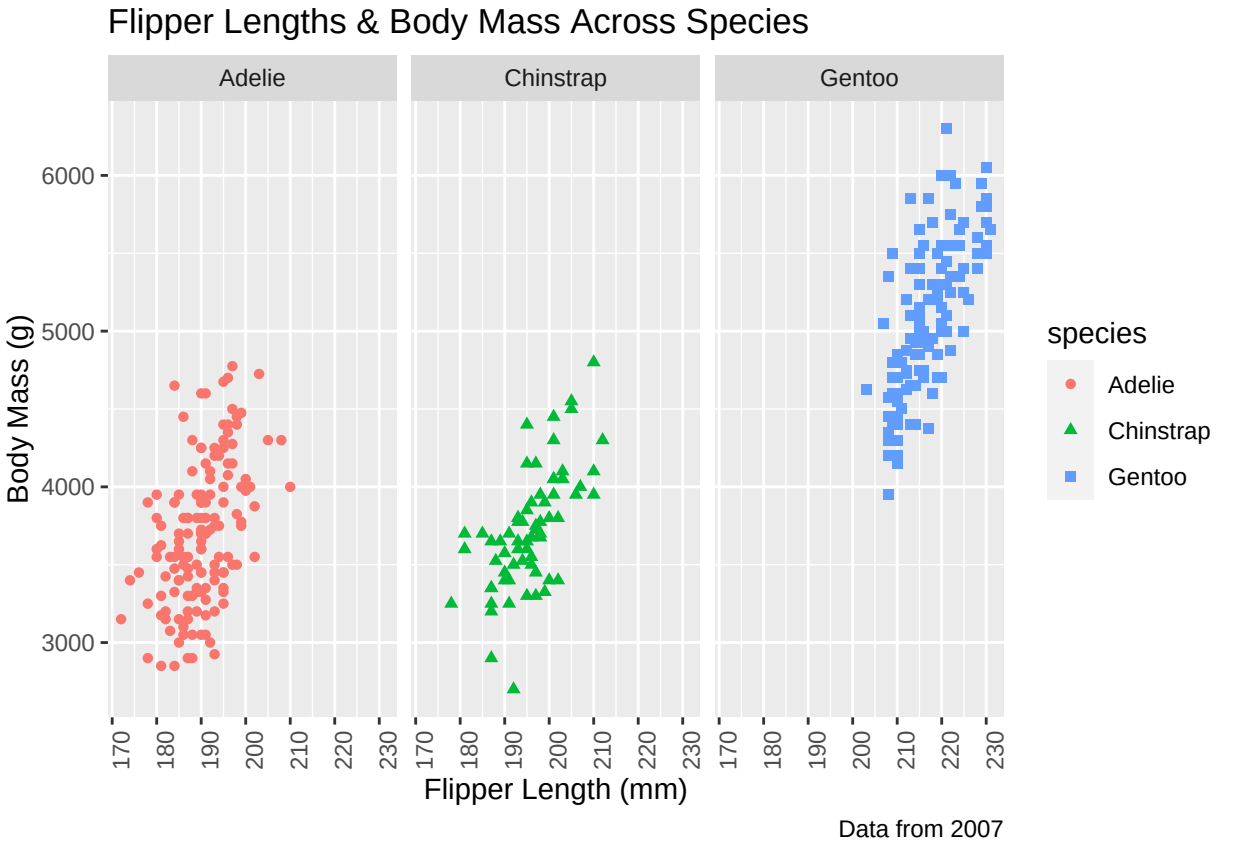
```
head(penguins)
```

```
## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>           <dbl>         <dbl>            <int>       <int> <fct>
## 1 Adelie  Torge~           39.1          18.7              181        3750 male
## 2 Adelie  Torge~           39.5          17.4              186        3800 fema~
## 3 Adelie  Torge~           40.3          18                195        3250 fema~
## 4 Adelie  Torge~           NA            NA                 NA          NA <NA>
## 5 Adelie  Torge~           36.7          19.3              193        3450 fema~
## 6 Adelie  Torge~           39.3          20.6              190        3650 male
## # ... with 1 more variable: year <int>
```

Visualization

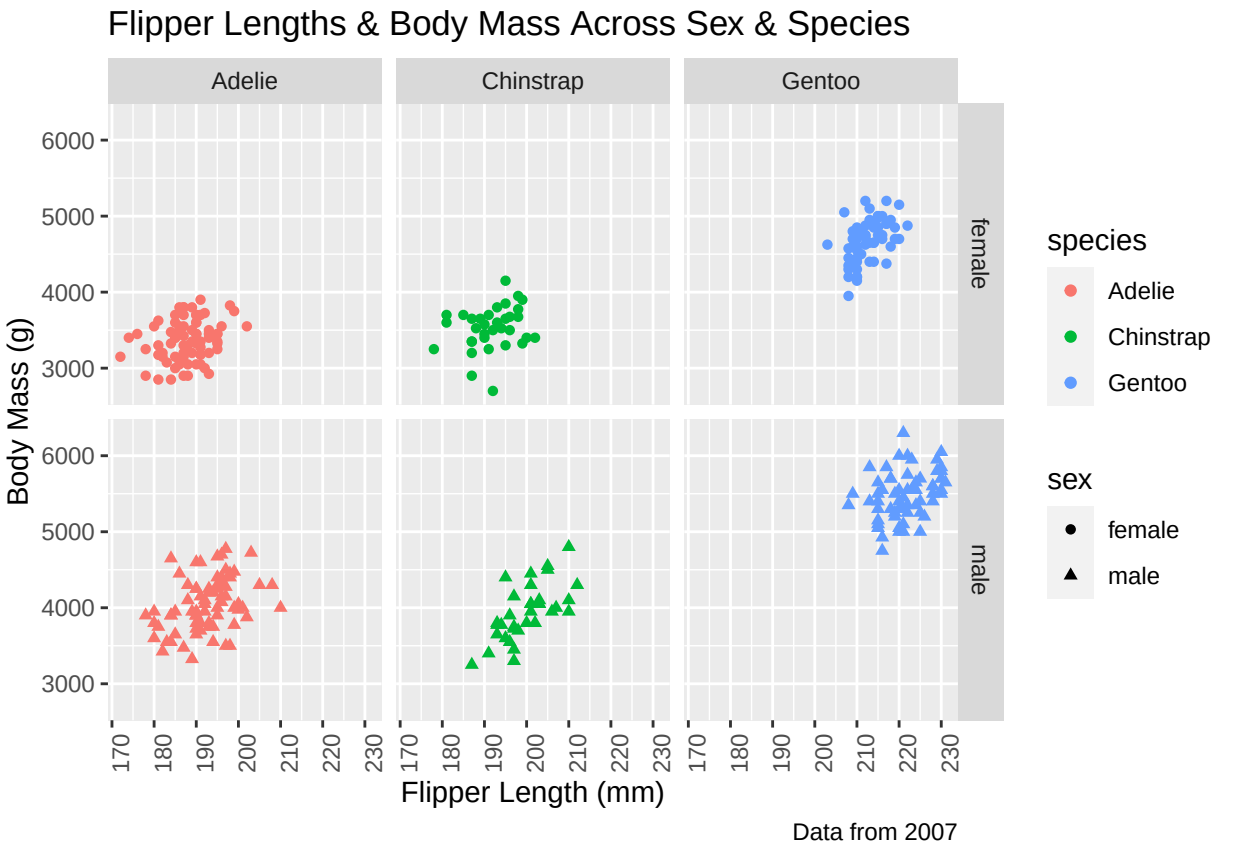
Notes: I decided to examine the relationship between *body mass* and *flipper length* across the different *species*. We have some null values in the *sex* column that we can get rid of.

```
penguins %>%
  drop_na(sex) %>%
  ggplot(mapping = aes(x = flipper_length_mm, y = body_mass_g,
  color = species, shape = species))+
  geom_point() + facet_wrap(~species) +
  labs(title = "Flipper Lengths & Body Mass Across Species", x =  "Flipper Length (mm)", y = "Body Mass
  theme(axis.text.x = element_text(angle = 90))
```



Next, I decided to explore the impact of adding *sex* in the equation. It could be insightful to make comparisons across different sexes along with *species*.

```
penguins %>%
  drop_na(sex) %>%
  ggplot(mapping = aes(x= flipper_length_mm, y = body_mass_g,
  color = species, shape = sex))+
  geom_point() + facet_grid(sex~species) +
  labs(title = "Flipper Lengths & Body Mass Across Sex & Species",
  x = "Flipper Length (mm)", y = "Body Mass (g)", caption = "Data from 2007") +
  theme(axis.text.x = element_text(angle = 90))
```



Observations and Findings

So far we can see that the male Gentoo penguins are the largest, both in terms of body mass and flipper lengths and the Adelie species in the smallest. We can use a quick statistical analysis code to confirm.

```
penguins %>%
  drop_na() %>%
  group_by(species) %>%
  summarize(mean_body_mass = mean(body_mass_g),
  mean_flipper_length = mean(flipper_length_mm))
```

```
## # A tibble: 3 x 3
##   species   mean_body_mass mean_flipper_length
##   <fct>              <dbl>               <dbl>
## 1 Adelie             3706.                190.
## 2 Chinstrap          3733.                196.
## 3 Gentoo             5092.                217.
```

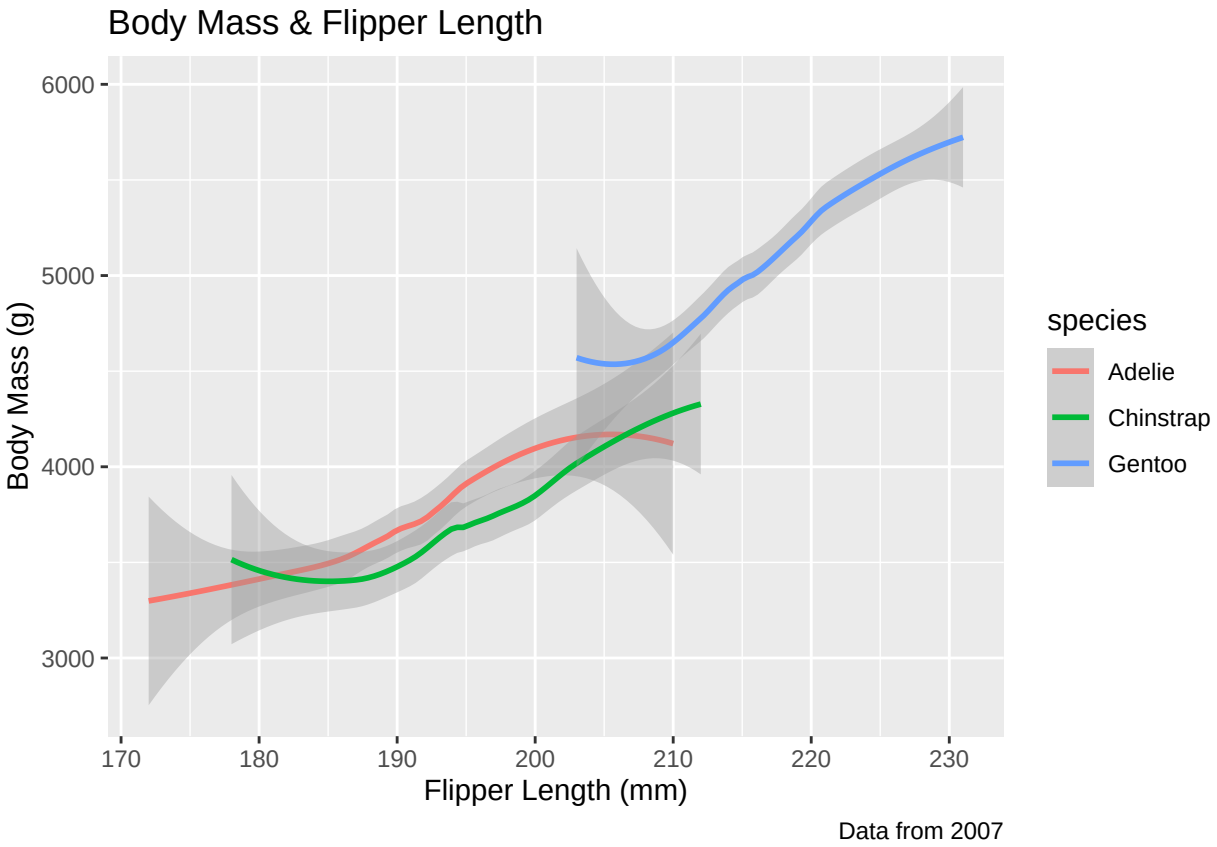
We can now be assured our initial observations are correct. The largest *species* by both *flipper length* and *body mass* is the Gentoos and the smallest *species* are the Adelie penguins. but is it safe to assume a positive correlation between *body mass* and *flipper length*?

```
ggplot(data = penguins) + geom_smooth(mapping = aes(x = flipper_length_mm, y = body_mass_g, line_type =
```

```
## Warning: Ignoring unknown aesthetics: line_type

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

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



This graph clear shows the positive relationship between *body mass* and *flipper length* present across all species.