Palmer Penguins Report

Arief

2022-06-27

Setting up my environment

Notes: setting up my R environment by loading the 'tidyverse', 'palmerpenguins', "janitor" and "ggpubr" packages

```
library(tidyverse)
library(palmerpenguins)
library(janitor)
library(ggpubr)
```

Cleaning data

Setting all the column title to lowercase and clean the name so that only underscore, numbers and letters left for easy analysis. Also I remove NA to get a more accurate analysis

```
penguins1 <- rename_with(penguins, tolower)

penguins_clean <- clean_names(penguins1) %>%
  drop_na()
```

Checking if there is a different in male and female number for each species of penguins, to see if there is a bias in the data

```
penguins_clean %>%
  group_by(species) %>%
  count(sex)
```

```
## # A tibble: 6 x 3
## # Groups:   species [3]
##   species   sex        n
##   <fct>     <fct>  <int>
## 1 Adelie    female    73
## 2 Adelie    male      73
## 3 Chinstrap female    34
## 4 Chinstrap male      34
## 5 Gentoo    female    58
## 6 Gentoo    male      61
```

The number of male and female is about the same for each species, so there is no sex bias in the data.

Analysing data

Here we look at what is the mean, max and min of bill and flipper length of penguins grouping them by species

```
penguins_clean %>%
  group_by(species) %>%
  summarise(mean_bill_length_mm = mean(bill_length_mm),
            max_bill_length_mm = max(bill_length_mm),
            min_bill_length_mm = min(bill_length_mm))
```

```
## # A tibble: 3 x 4
##   species   mean_bill_length_mm max_bill_length_mm min_bill_length_mm
##   <fct>                   <dbl>              <dbl>              <dbl>
## 1 Adelie                   38.8               46                 32.1
## 2 Chinstrap                48.8               58                 40.9
## 3 Gentoo                   47.6               59.6               40.9
```

```
penguins_clean %>%
  group_by(species) %>%
  summarise(mean_flipper_length_mm = mean(flipper_length_mm),
            max_flipper_length_mm = max(flipper_length_mm),
            min_flipper_length_mm = min(flipper_length_mm))
```

```
## # A tibble: 3 x 4
##   species   mean_flipper_length_mm max_flipper_length_mm min_flipper_length_mm
##   <fct>                      <dbl>                 <int>                 <int>
## 1 Adelie                      190.                   210                   172
## 2 Chinstrap                   196.                   212                   178
## 3 Gentoo                      217.                   231                   203
```

I found that only species Adelie are in the three island, so here we see if penguin of the same species have different mean body mass when they located in different island.

```
penguins_clean %>%
  filter(species == 'Adelie') %>%
  group_by(island) %>%
  summarise(mean_body_mass_g = mean(body_mass_g))
```

```
## # A tibble: 3 x 2
##   island    mean_body_mass_g
##   <fct>                <dbl>
## 1 Biscoe               3710.
## 2 Dream                3701.
## 3 Torgersen            3709.
```

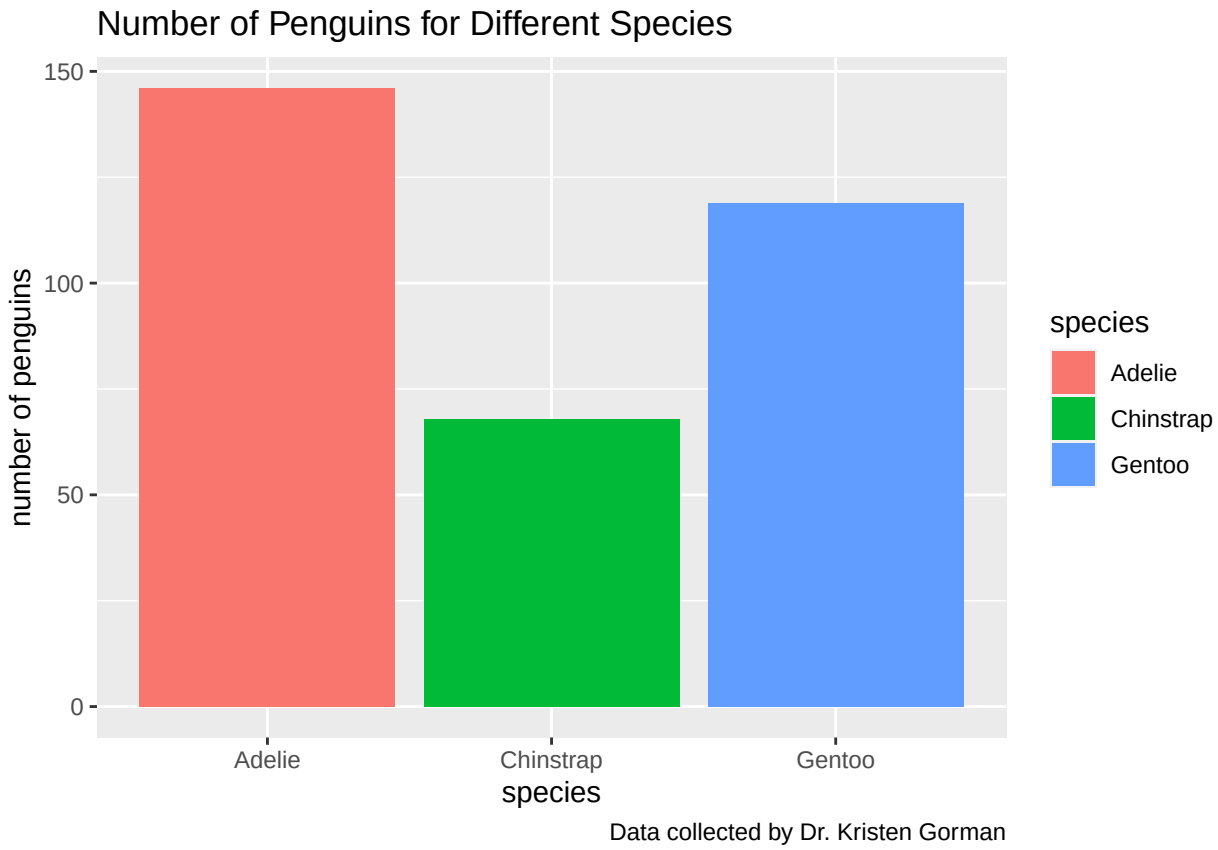
Visualisations

Here we will go through a series of visualisations

Number of penguins of each species

We will look at how many number of penguins for each species

```
ggplot(data=penguins_clean) +
  geom_bar(mapping = aes(x = species, fill = species)) +
  labs(title = "Number of Penguins for Different Species", y = "number of penguins",
       caption = "Data collected by Dr. Kristen Gorman")
```

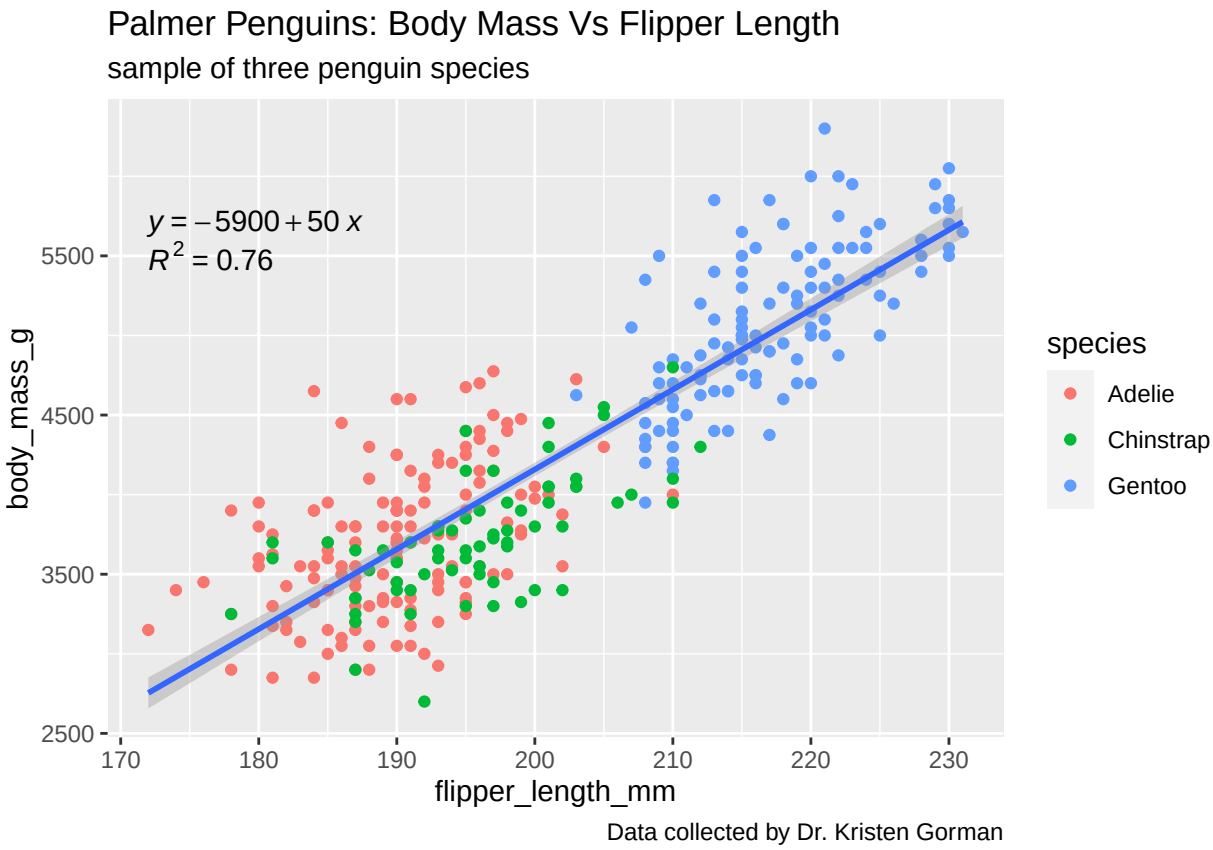


Body mass and flipper length

Here we plot body mass in gram(g) against flipper length is millimeter(mm)

```
ggplot(data=penguins_clean, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species)) +
  geom_smooth(method = "lm") +
  stat_regline_equation(label.y=5700, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y=5500, aes(label = ..rr.label..)) +
  labs(title = "Palmer Penguins: Body Mass Vs Flipper Length",
       subtitle = "sample of three penguin species",
       caption = "Data collected by Dr. Kristen Gorman")
```

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

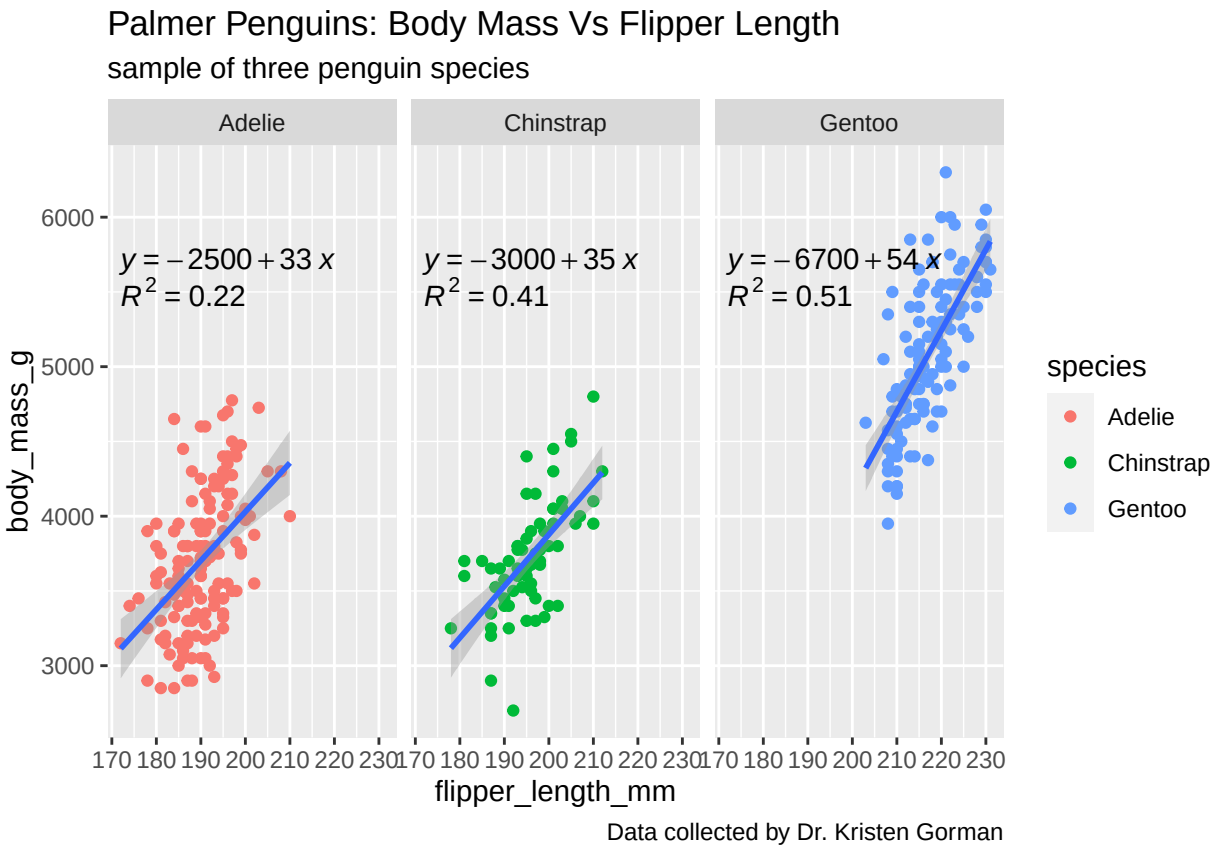


Body mass and flipper length of each species

Here we separate the plot into species

```
ggplot(data=penguins_clean, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species)) +
  geom_smooth(method = "lm") +
  facet_wrap(~species) +
  stat_regline_equation(label.y=5700, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y=5500, aes(label = ..rr.label..))+
  labs(title = "Palmer Penguins: Body Mass Vs Flipper Length",
       subtitle = "sample of three penguin species",
       caption = "Data collected by Dr. Kristen Gorman")
```

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

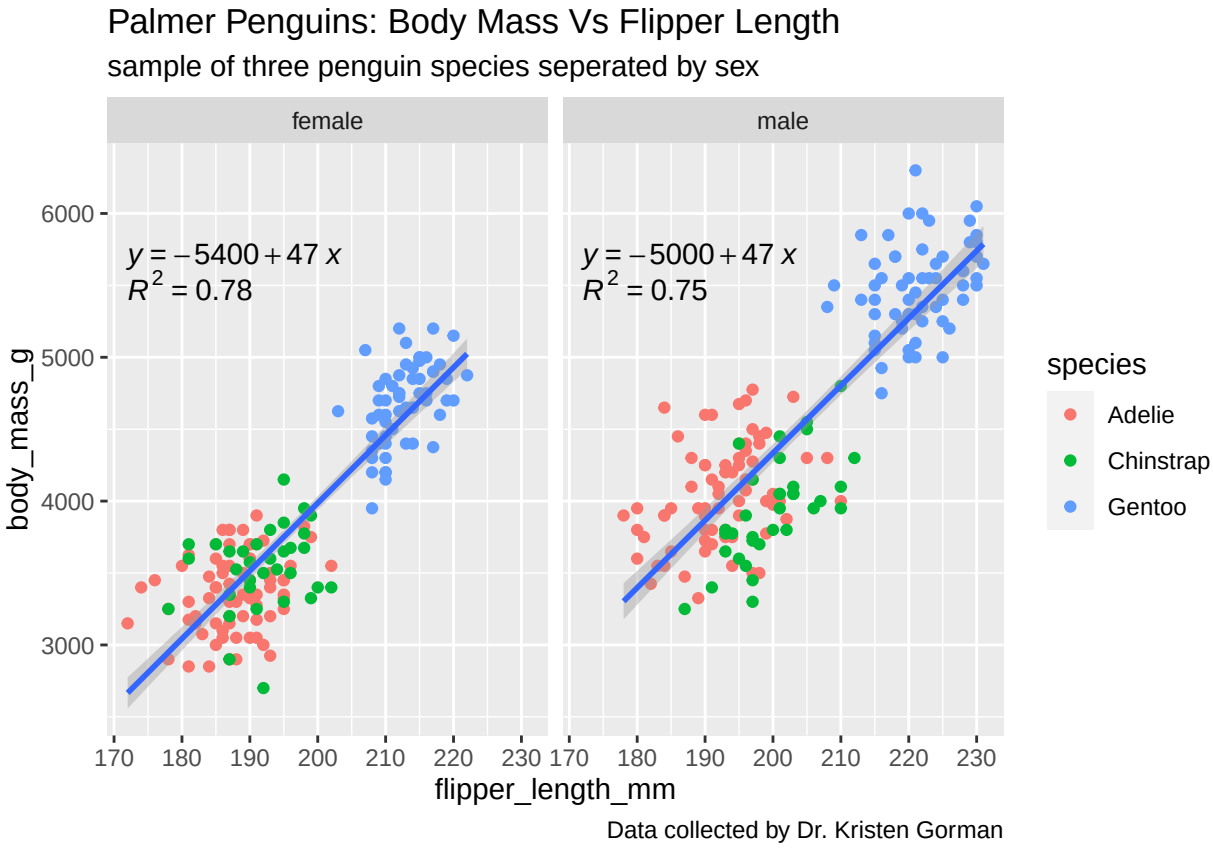


Body mass and flipper length seperated by sex

Here we seperate the plot of body mass against flipper length to sex

```
ggplot(data=penguins_clean, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species)) +
  geom_smooth(method = "lm") +
  facet_wrap(~sex) +
  stat_regline_equation(label.y=5700, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y=5500, aes(label = ..rr.label..))+
  labs(title = "Palmer Penguins: Body Mass Vs Flipper Length",
       subtitle = "sample of three penguin species seperated by sex",
       caption = "Data collected by Dr. Kristen Gorman")
```

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


Palmer Penguins: Body Mass Vs Flipper Length
sample of three penguin species seperated by sex
female
male
y = −5400 + 47 x
R2 = 0.78
y = −5000 + 47 x
R2 = 0.75
body_mass_g
6000
5000
4000
3000
170
180
190
200
210
220
230
flipper_length_mm
Data collected by Dr. Kristen Gorman
species
Adelie
Chinstrap
Gentoo