summer school R programming 2025

2025-10-02

In this project I am using packages dplyr and ggplot2 to show the data of penguins.

```
library(tidyverse)
```

We check the observations of penguins

```
head(penguins)
```

```
## # A tibble: 6 × 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>
```

Here I am selecting the data of penguins with species, island, year and the data of penguins with all the informations except for sex and species

```
dplyr::select(penguins, species, island, year)
dplyr::select(penguins, -c(sex,species))
```

Show all the observations of Chinstrap penguins from 2007

```
dplyr::filter(penguins, year >= 2007& species == "Chinstrap")
```

```
## # A tibble: 68 × 8
##    species   island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##    <fct>     <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Chinstrap Dream            46.5          17.9               192        3500
##  2 Chinstrap Dream            50            19.5               196        3900
##  3 Chinstrap Dream            51.3          19.2               193        3650
##  4 Chinstrap Dream            45.4          18.7               188        3525
##  5 Chinstrap Dream            52.7          19.8               197        3725
##  6 Chinstrap Dream            45.2          17.8               198        3950
##  7 Chinstrap Dream            46.1          18.2               178        3250
##  8 Chinstrap Dream            51.3          18.2               197        3750
##  9 Chinstrap Dream            46            18.9               195        4150
## 10 Chinstrap Dream            51.3          19.9               198        3700
## # ℹ 58 more rows
## # ℹ 2 more variables: sex <fct>, year <int>
```

Show the observations of Chinstrap penguins in 2009, including species, sex and year.

```
penguins %>%
  dplyr::filter(year == 2009 & species == "Chinstrap") %>%
  dplyr::select(species, sex, year)
```

```
## # A tibble: 24 × 3
##    species   sex      year
##    <fct>     <fct>   <int>
##  1 Chinstrap female   2009
##  2 Chinstrap male     2009
##  3 Chinstrap female   2009
##  4 Chinstrap male     2009
##  5 Chinstrap male     2009
##  6 Chinstrap female   2009
##  7 Chinstrap female   2009
##  8 Chinstrap male     2009
##  9 Chinstrap female   2009
## 10 Chinstrap male     2009
## # i 14 more rows
```

Overwrite the weight of penguins in pounds.

```
penguins %>%
  dplyr::mutate(body_mass_kg = body_mass_g/1000,
               body_mass_lbs = body_mass_kg/453,
               body_mass_kg = "Overwrite")
```

```
## # A tibble: 344 × 10
##    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
##  7 Adelie  Torgersen           38.9          17.8               181        3625
##  8 Adelie  Torgersen           39.2          19.6               195        4675
##  9 Adelie  Torgersen           34.1          18.1               193        3475
## 10 Adelie  Torgersen           42            20.2               190        4250
## # i 334 more rows
## # i 4 more variables: sex <fct>, year <int>, body_mass_kg <chr>,
## #   body_mass_lbs <dbl>
```

Grouping the penguin by species and summarize their heaviest penguin.

```
penguins %>%
  dplyr::group_by(species) %>%
  dplyr::summarize(heaviest_penguin = max(body_mass_g, na.rm = T), .groups = "drop")
```

```
## # A tibble: 3 × 2
##   species   heaviest_penguin
##   <fct>                <int>
## 1 Adelie                4775
## 2 Chinstrap             4800
## 3 Gentoo                6300
```

Create a data frame arranged by body_mass_g of the penguins observed in the "Dream" island.

```
penguins %>%
  dplyr::filter(island == "Dream") %>%
  dplyr::arrange(desc(body_mass_g))
```

```
## # A tibble: 124 × 8
##    species   island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##    <fct>     <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Chinstrap Dream            52           20.7               210        4800
##  2 Adelie    Dream            39.8         19.1               184        4650
##  3 Adelie    Dream            39.6         18.8               190        4600
##  4 Chinstrap Dream            52.8         20                 205        4550
##  5 Chinstrap Dream            53.5         19.9               205        4500
##  6 Adelie    Dream            37.5         18.5               199        4475
##  7 Adelie    Dream            39.6         18.1               186        4450
##  8 Chinstrap Dream            50.8         18.5               201        4450
##  9 Adelie    Dream            44.1         19.7               196        4400
## 10 Chinstrap Dream            49.2         18.2               195        4400
## # ℹ 114 more rows
## # ℹ 2 more variables: sex <fct>, year <int>
```

Find out the longest observed bill of a Gentoo penguin in 2008.

```
penguins %>%
  dplyr::filter(species == "Gentoo" & year == 2008) %>%
  dplyr::arrange(desc(bill_length_mm))
```

```
## # A tibble: 46 × 8
##    species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>             <int>       <int>
##  1 Gentoo  Biscoe          54.3          15.7               231        5650
##  2 Gentoo  Biscoe          51.1          16.3               220        6000
##  3 Gentoo  Biscoe          50.7          15                 223        5550
##  4 Gentoo  Biscoe          50.5          15.9               222        5550
##  5 Gentoo  Biscoe          50.5          15.9               225        5400
##  6 Gentoo  Biscoe          50.4          15.3               224        5550
##  7 Gentoo  Biscoe          50.1          15                 225        5000
##  8 Gentoo  Biscoe          49.8          16.8               230        5700
##  9 Gentoo  Biscoe          49.6          16                 225        5700
## 10 Gentoo  Biscoe          49.6          15                 216        4750
## # ℹ 36 more rows
## # ℹ 2 more variables: sex <fct>, year <int>
```

Find out the shortest observed bill of a Gentoo penguin in 2008.

```
penguins %>%
  dplyr::filter(species == "Gentoo" & year == 2008) %>%
  dplyr::pull(bill_length_mm) %>%
  min(., na.rm = T)
```

```
## [1] 42.6
```

Subset rows 3, 5 and 6.

```
penguins %>% slice(c(3, 5, 6))
```

```
## # A tibble: 3 × 8
##   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
## 1 Adelie  Torgersen           40.3          18                 195        3250
## 2 Adelie  Torgersen           36.7          19.3               193        3450
## 3 Adelie  Torgersen           39.3          20.6               190        3650
## # i 2 more variables: sex <fct>, year <int>
```

returns vector with distinct species and returns vector with distinct species + count in data.

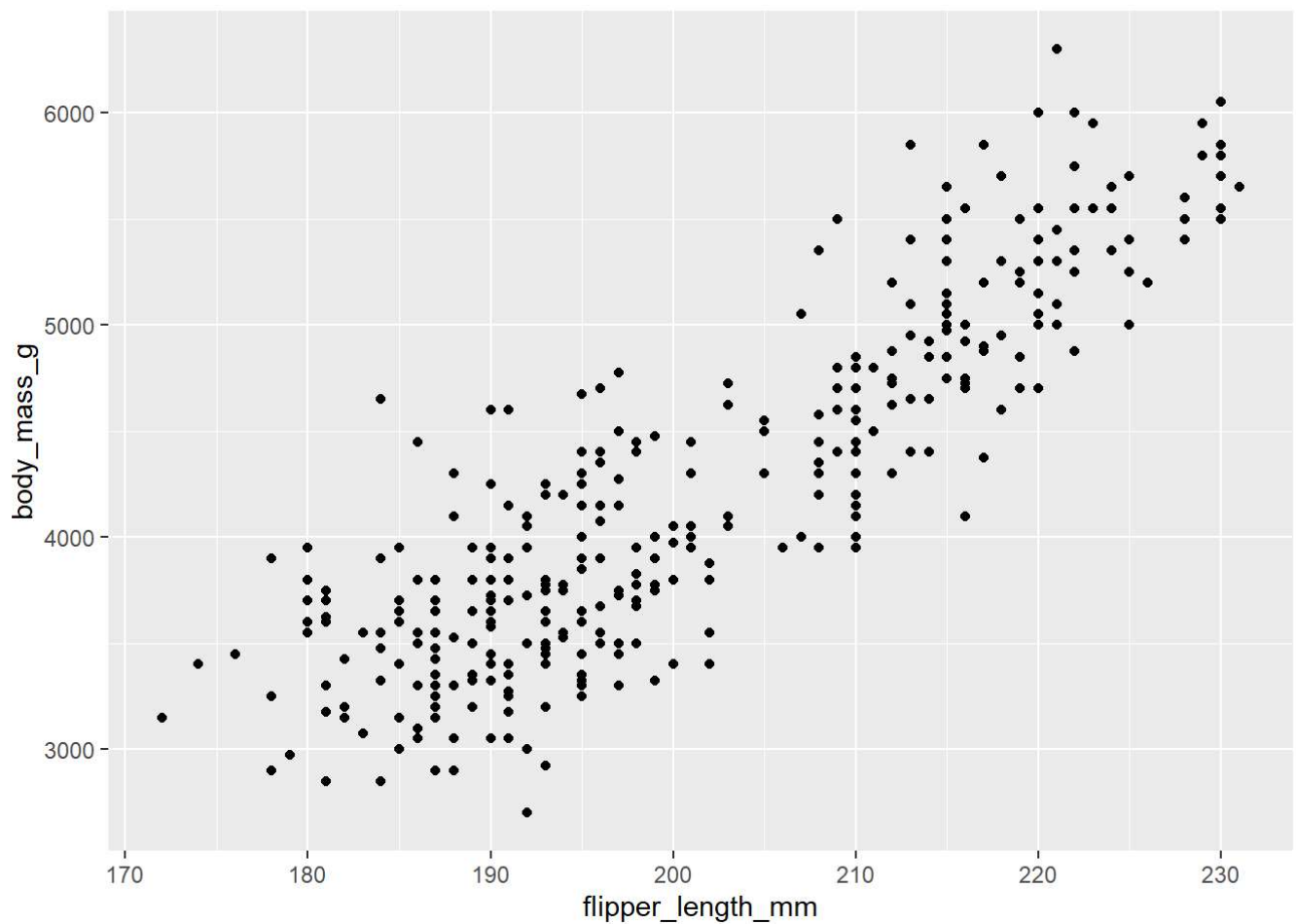```
penguins %>% dplyr::distinct(species)
```

```
## # A tibble: 3 × 1
##   species
##   <fct>
## 1 Adelie
## 2 Gentoo
## 3 Chinstrap
```

```
penguins %>% dplyr::count(species)
```

```
## # A tibble: 3 × 2
##   species      n
##   <fct>    <int>
## 1 Adelie     152
## 2 Chinstrap   68
## 3 Gentoo     124
```
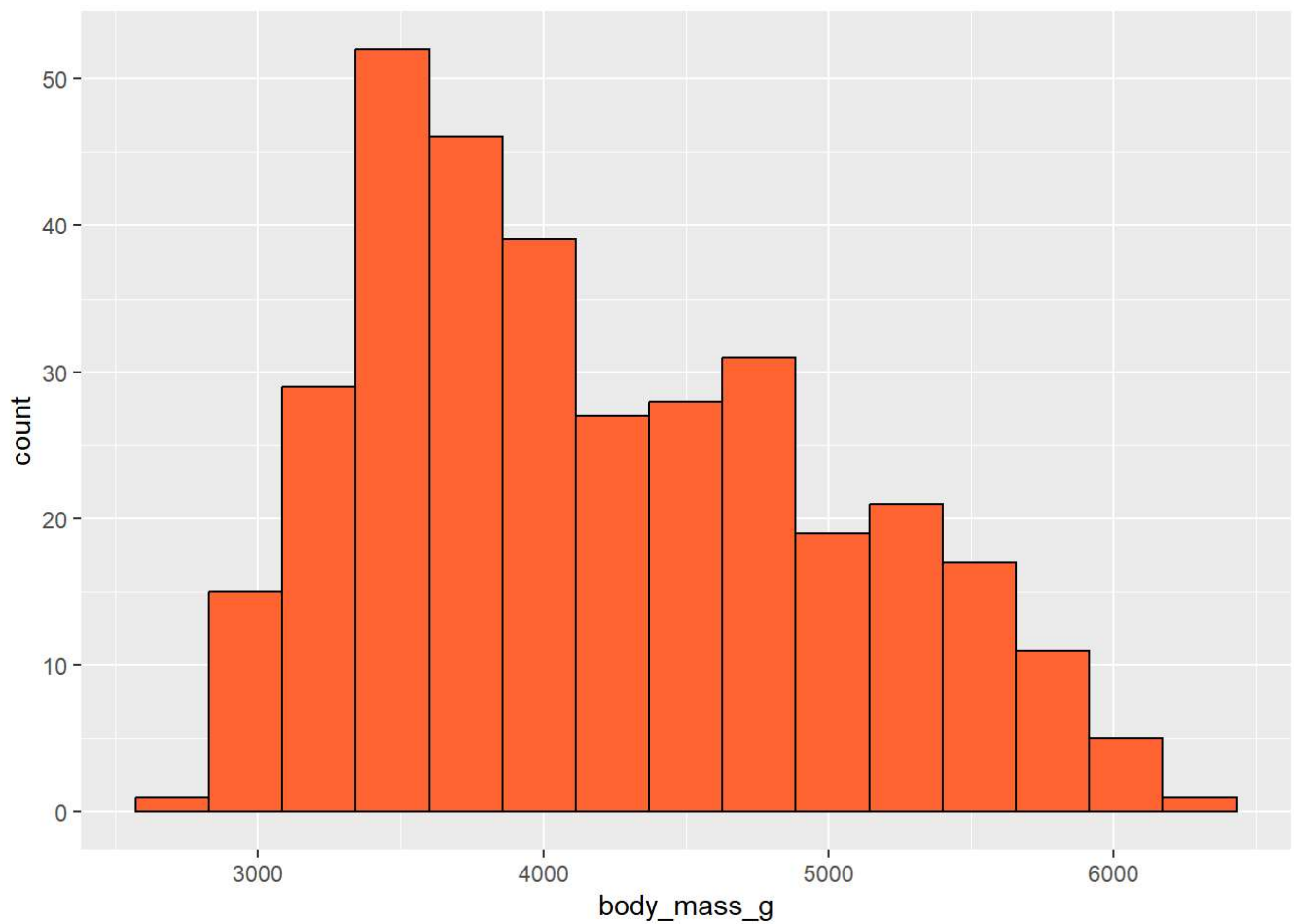
Showing the relationship between the length of a penguin's flipper and their weight.

```
ggplot(penguins, aes(x = flipper_length_mm, y=body_mass_g)) +
  geom_point()
```

Pesenting the weight distribution of penguins with histogram.

```
ggplot(penguins, aes(x = body_mass_g)) +
  geom_histogram(bins = 15, fill = "#FF6634", color = "black")
```

Showing the distribution of flipper length by species

```
ggplot(penguins, aes(x = species, y = flipper_length_mm)) +
  geom_boxplot() +
  theme_minimal() +
  labs(x = "Species",
       y = "flipper length (mm)")
```