Module 2, Assignment 1

Ellen Bledsoe

2022-09-29

Assignment Details

Purpose

The goal of this assignment is to assess your ability to produce and interpret histograms and scatter plots.

Task

Write R code which produces the correct answers and correctly interpret the plots produced.

Criteria for Success

- Code is within the provided code chunks
- Code is commented with brief descriptions of what the code does
- Code chunks run without errors
- Code produces the correct result
  - Code that produces the correct answer will receive full credit
  - Code attempts with logical direction will receive partial credit
- Written answers address the questions in sufficient detail

Due Date

October 11 at midnight MST

Assignment Questions

For this assignment, we are going to be making plots! We are going to use a data set called `penguins` from the `palmerpenguins` package.

Most of the code you will need to complete this assignment is code we used in the first lesson of this module, 1_FoodGoneBad.

1. Load both the `palmerpenguins` package and the `tidyverse` package into the workspace. (2 points)

```
library(palmerpenguins)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

When we use data from a data package, it doesn't automatically show up in our environment. Run this code chunk so it does show up in the environment.

```
penguins <- penguins
```

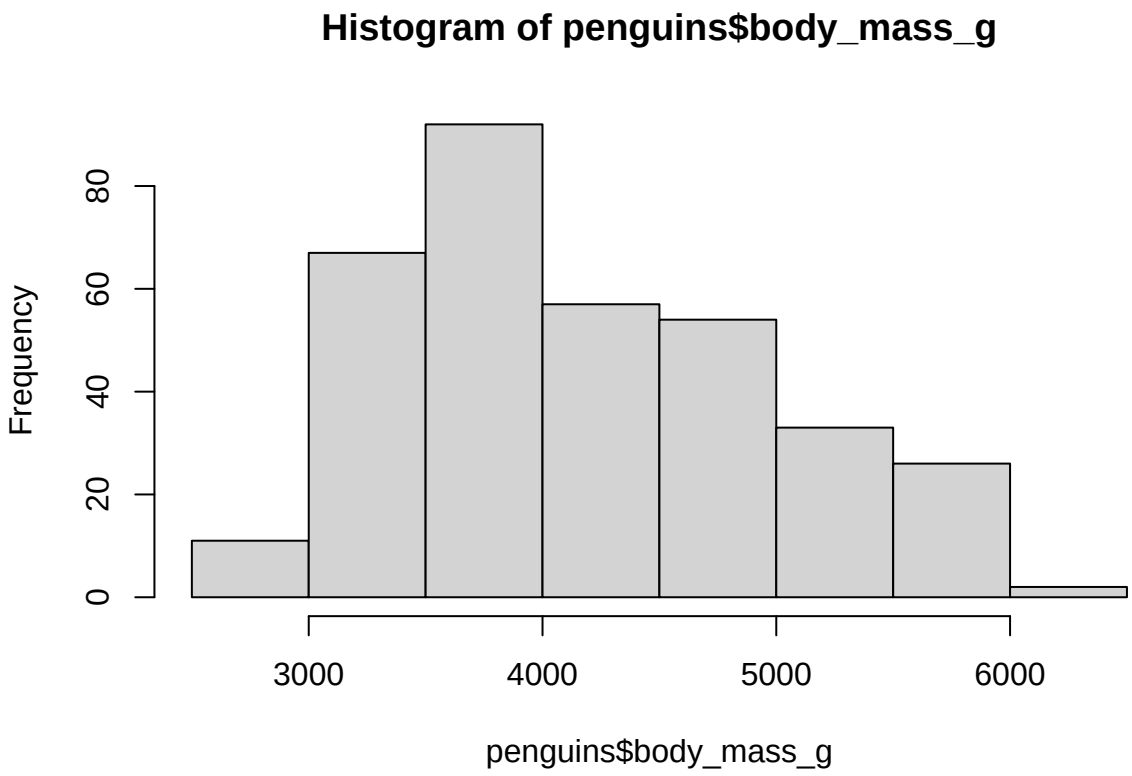
2. Use the `head()` function to take a look at the `penguins` data frame. (1 points)

```
head(penguins)
```

```
## # A tibble: 6 x 8
##   species island    bill_length_mm bill_depth_mm flipper_l~1 body_~2 sex    year
##   <fct>   <fct>              <dbl>         <dbl>       <int>   <int> <fct> <int>
## 1 Adelie  Torgersen           39.1          18.7         181    3750 male   2007
## 2 Adelie  Torgersen           39.5          17.4         186    3800 fema~  2007
## 3 Adelie  Torgersen           40.3          18           195    3250 fema~  2007
## 4 Adelie  Torgersen           NA            NA            NA      NA <NA>   2007
## 5 Adelie  Torgersen           36.7          19.3         193    3450 fema~  2007
## 6 Adelie  Torgersen           39.3          20.6         190    3650 male   2007
## # ... with abbreviated variable names 1: flipper_length_mm, 2: body_mass_g
```

3. Make a histogram of the body mass column. (2 points)

```
hist(penguins$body_mass_g)
```



4. In 2-3 sentences, describe what the histogram is telling you. I'm not necessarily looking for technical answers, but I want you to practice interpreting what histograms are telling you. (Examples: Are there even numbers of each body mass or different? Where is the peak? Are there lots of heavy penguins or not?) (3 points)

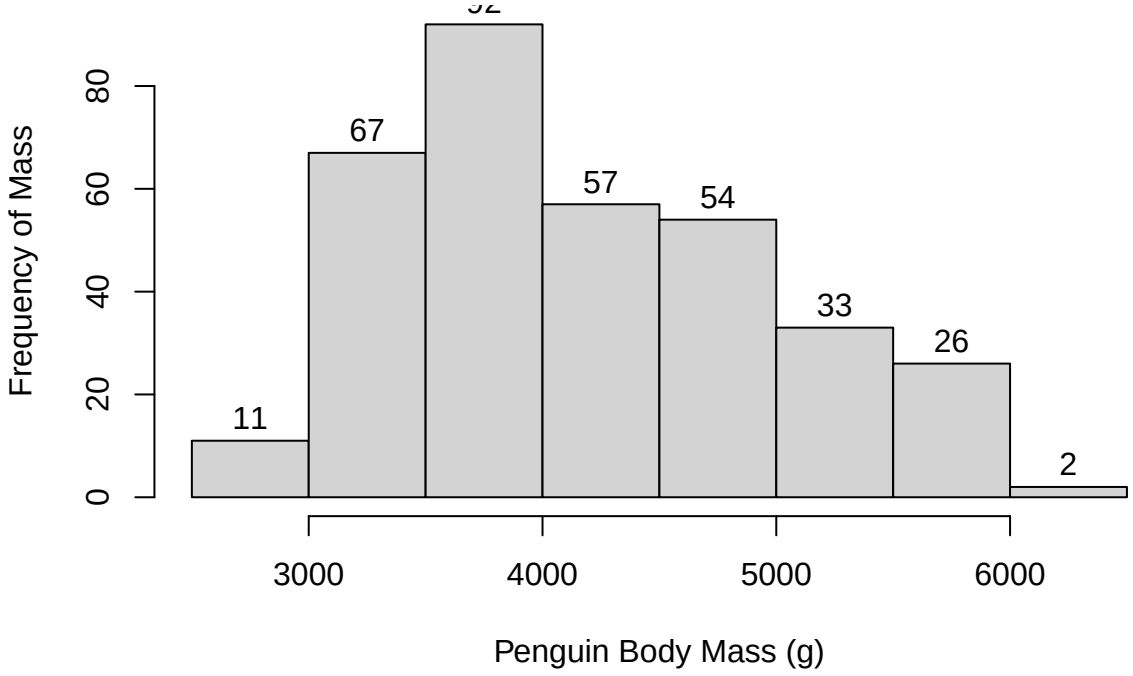
Answer:

5. Let's make the histogram in question 4 a bit easier for others to understand. (4 points)

Make the following changes:

- remove the main title (set the "main" argument to `NULL`)
- make the x-axis (horizontal) label easier to understand
- for the y-axis (vertical) label, add more detail ("Frequency" of what?)
- add total counts above each bin (this is controlled by the `labels` argument)

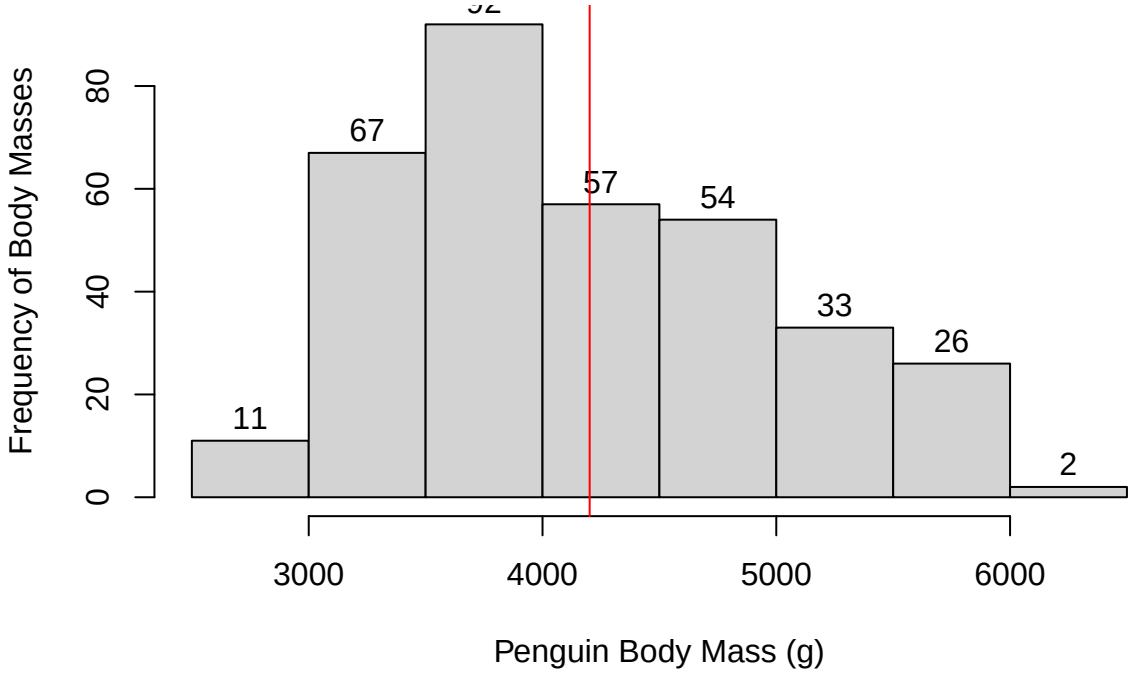
```
hist(penguins$body_mass_g, main = NULL,
     xlab = "Penguin Body Mass (g)",
     ylab = "Frequency of Mass",
     labels = TRUE)
```



6. Copy the code you wrote for question 5 and paste it below. Using the `abline()` function, add a vertical line to this histogram that represents the average penguin body mass. Make that line red. (2 points)

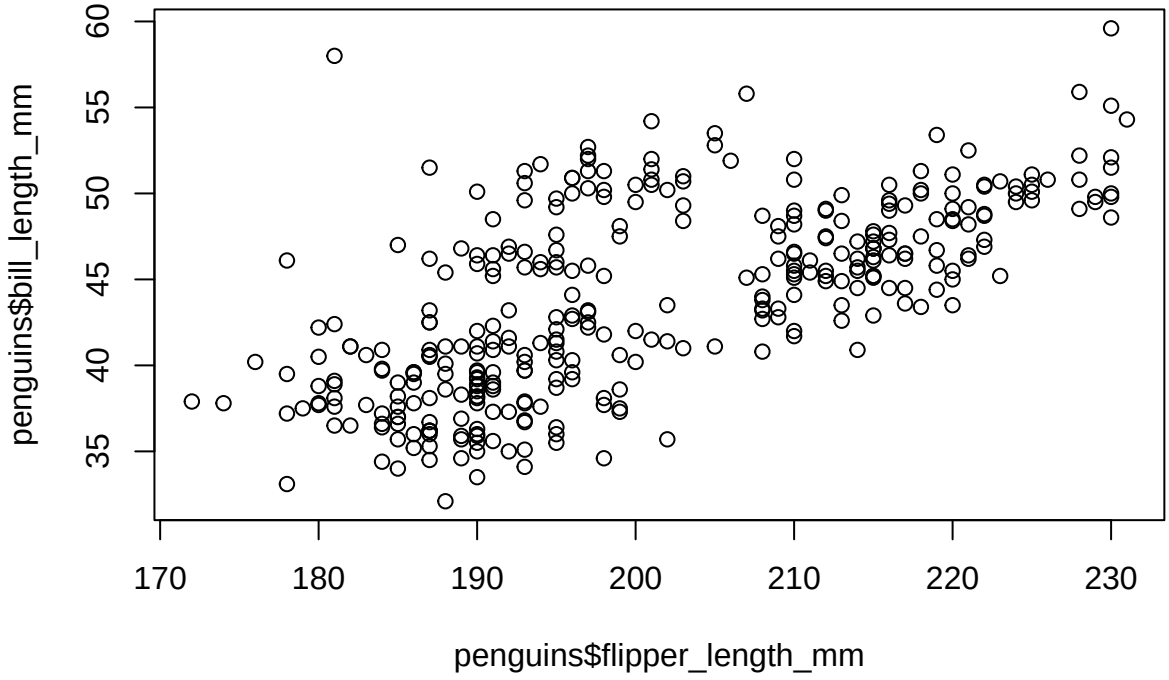
   Hint: you will need to include the `na.rm = TRUE` argument in the `mean()` function for the line to appear.

```
hist(penguins$body_mass_g, main = NULL,
     xlab = "Penguin Body Mass (g)",
     ylab = "Frequency of Body Masses",
     labels = TRUE)
abline(v = mean(penguins$body_mass_g, na.rm = TRUE), col = "red")
```



7. Make a scatter plot with flipper length on the x-axis (horizontal) and bill length on the y-axis (vertical). (2 points)

```
plot(x = penguins$flipper_length_mm, y = penguins$bill_length_mm)
```

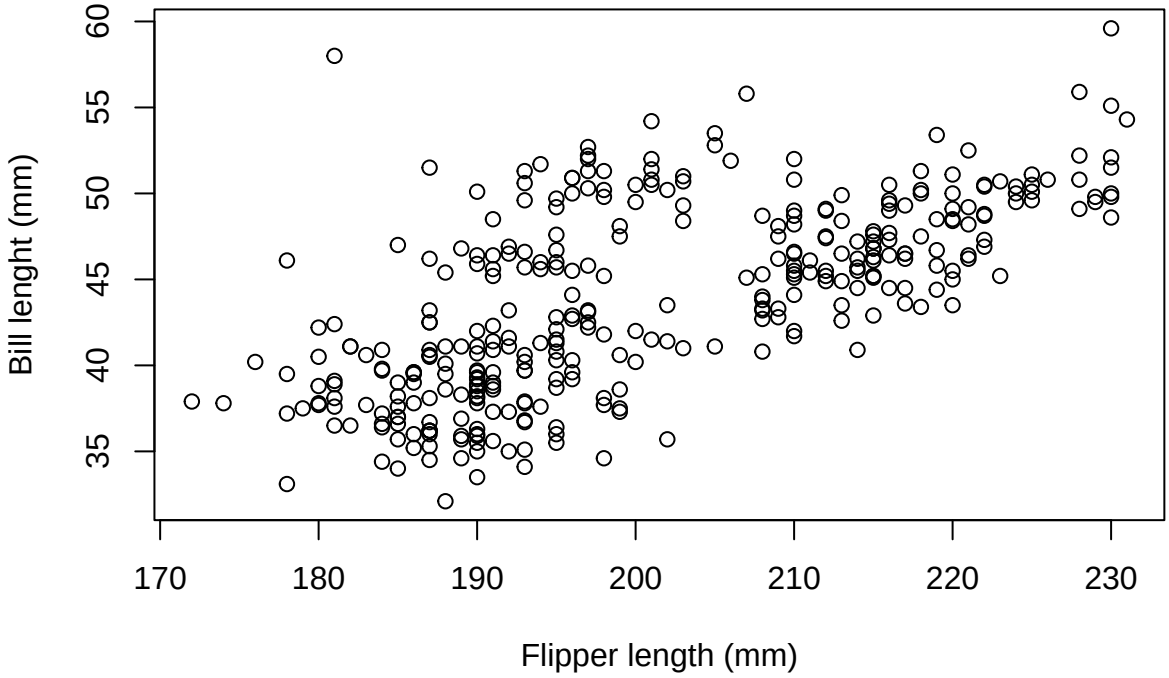


8. Write 1-2 sentences interpreting this plot. (Examples: Is there a positive relationship or a negative relationship? As flipper length increases, does bill length tend to increase, decrease, or stay the same? Is a penguin with a long flipper likely to have a long bill, too?) (2 points)

Answer:

9. As with the histogram above, we want to make this plot easier for others to understand. Change the axis labels so that they are clearer. (2 points)

```
plot(penguins$flipper_length_mm, penguins$bill_length_mm,
     xlab = "Flipper length (mm)",
     ylab = "Bill lenght (mm)")
```


Bill lenght (mm)
35
40
45
50
55
60
170
180
190
200
210
220
230
Flipper length (mm)