Assignment 2: Visualization by example

Jiwon MOk

2022-02-08

Exercise 1

i. How many rows and columns does this dataset have?

rows : 344, columns : 8

ii. What does a row in this dataset represent (i.e. what is the unit of observation)?

the number of penguins

iii. What are three categorical variables in the penguins dataset?

Gentoo, Chinstrap, Adelie

iv. What are four continuous variables in the penguins dataset?

bill_length_mm, bill_length_mm, flipper_length_mm, body_mass_g

v. Which variable in the penguins dataset could be treated as either continuous or categorical, depending on the context in which it is used?

categorial : sex, island, year continuous : bill_length_mm, flipper_length_mm, body_mass_g

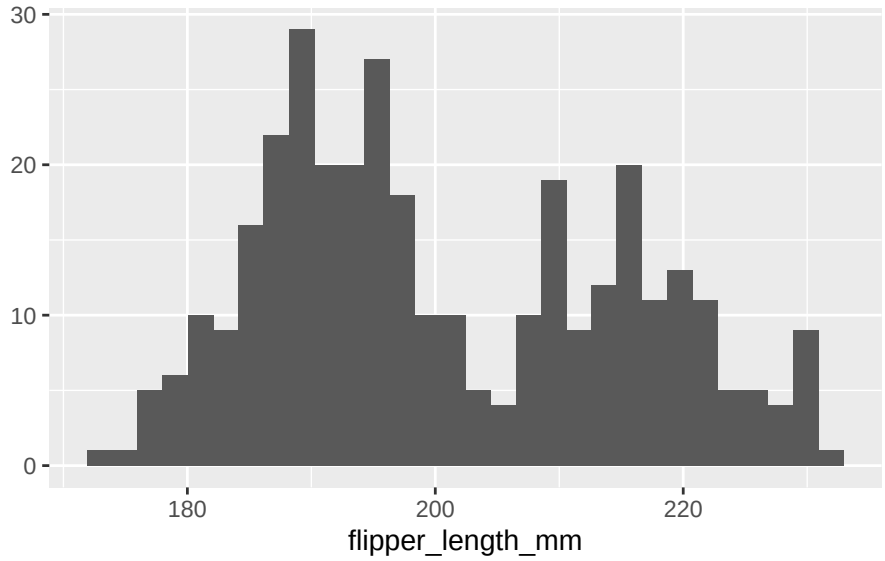
vi. What are the three species of penguin in the dataset? Gentoo, Chinstrap, Adelie

Exercise 2

```
qplot(x = flipper_length_mm, data = penguins)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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



i. **Which axis of the plot has the flipper length variable been plotted on?** x-axis

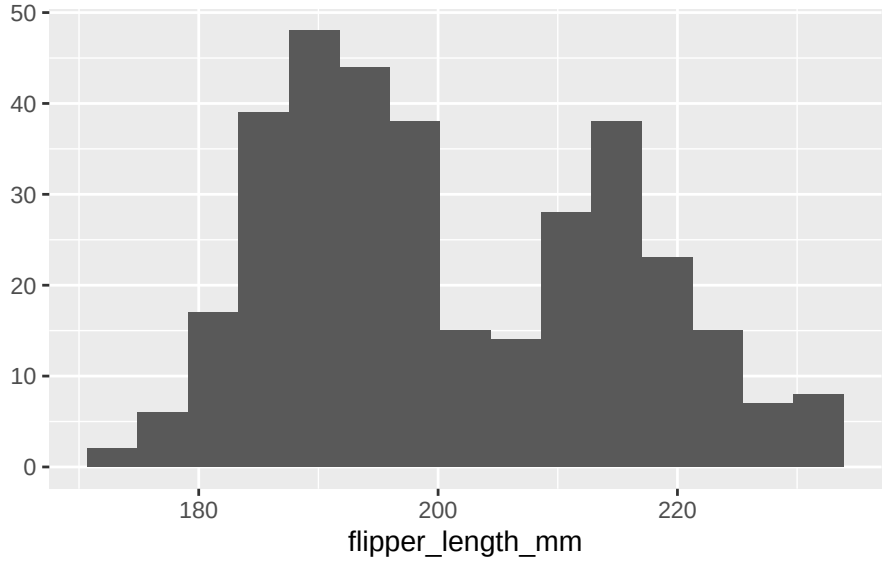
ii. **What do the numbers on the other axis of the graph represent?** the number of penguins

iii. **What is the modality of the distribution of flipper lengths?** bi-modal

Exercise 3

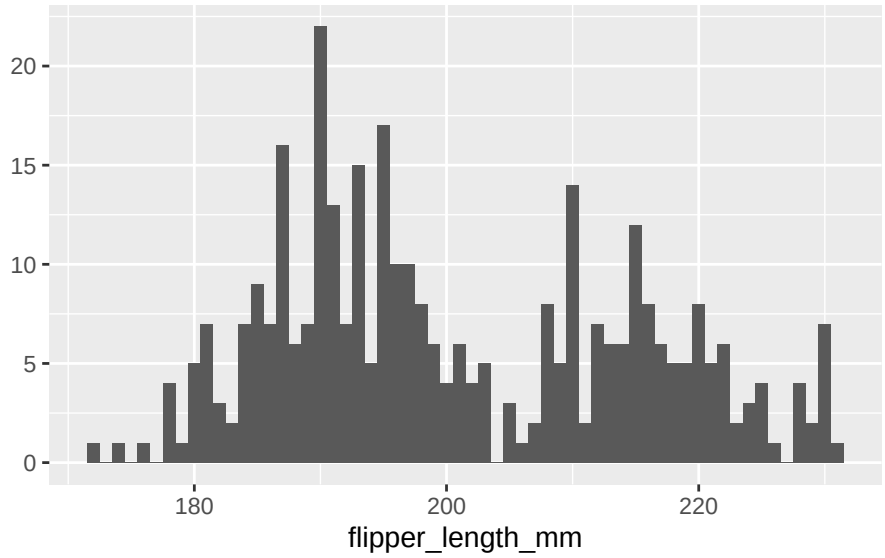
```
qplot(x = flipper_length_mm, bins = 15, data = penguins)
```

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




```
qplot(x = flipper_length_mm, binwidth=1, data = penguins)
```

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



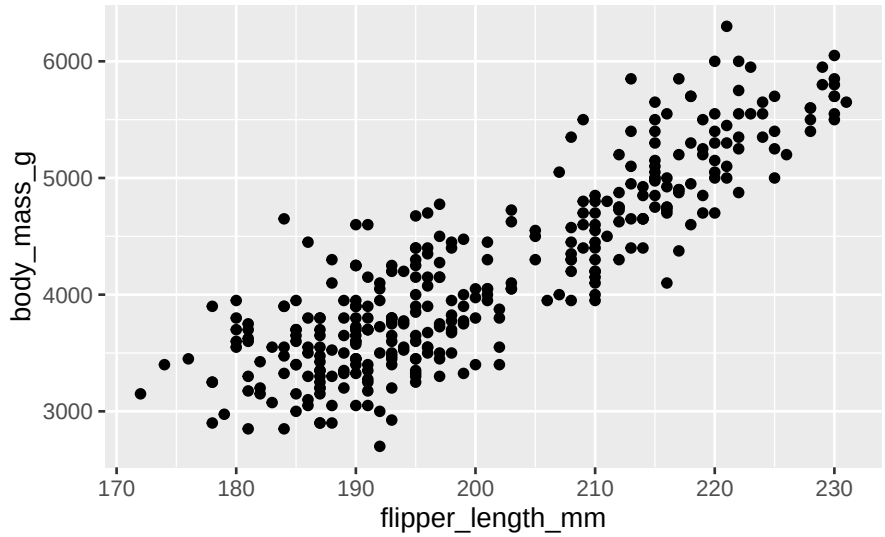
Are there more or less bins in this new histogram? Does that increase or decrease the smoothness of the distribution? Does that make the pattern of two peaks easier or harder to spot in this histogram?

- There seem to be more bins in the new histogram. It increases the smoothness and also it is easier to see the peaks as well.

Exercise 4

```
qplot(x = flipper_length_mm, y = body_mass_g, data = penguins)
```

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



i. What variable is on the y-axis of this scatter plot?

: body-mass-g

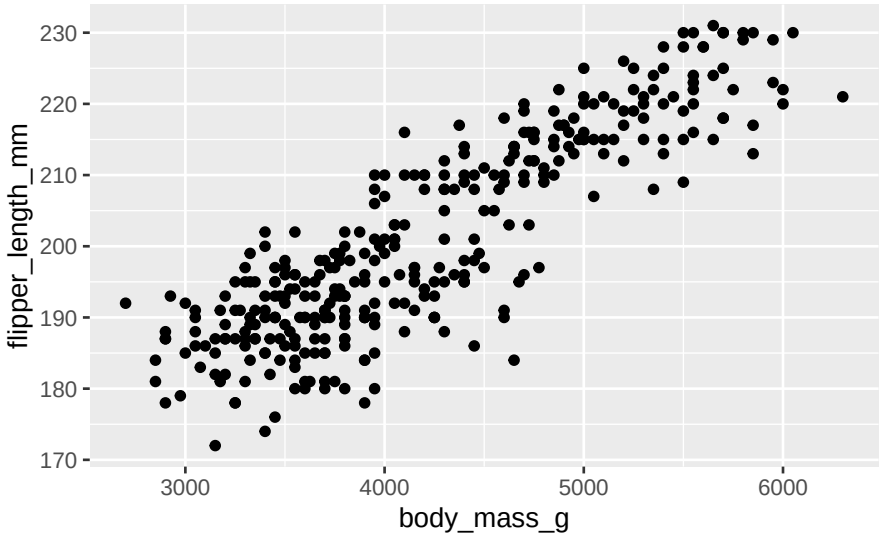
ii. Is there a relationship between these two variables, and if so, is it linear or non-linear?

: positive-correlation. linear graph

Exercise 5

```
qplot(y = flipper_length_mm, x = body_mass_g, data = penguins)
```

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



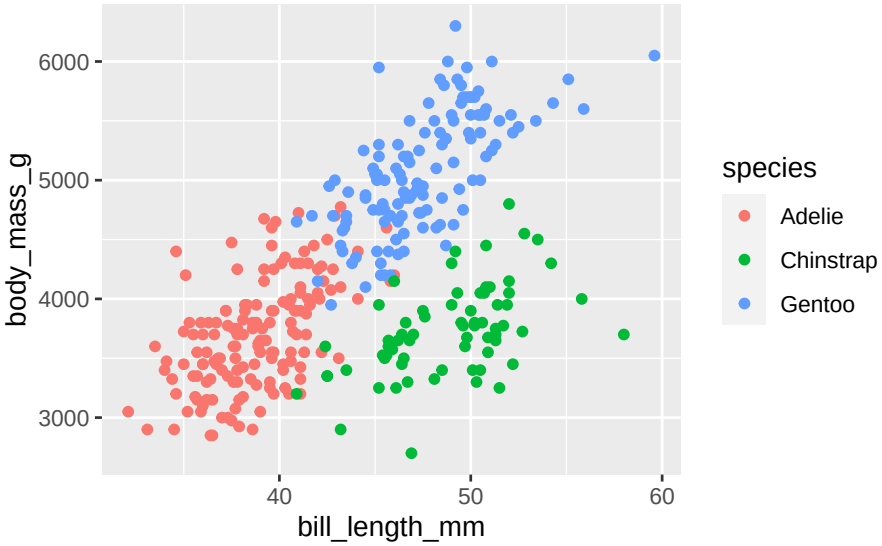
Does the correlation between bill length and body mass in this scatter plot look stronger or weaker than the correlation of the two variables in your previous scatter plot from Exercise 4?

They look simliar.

Exercise 6

```
qplot(
  x = bill_length_mm,
  y = body_mass_g,
  color = species,
  data = penguins
  )
```

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


body_mass_g
6000
5000
4000
3000
40
50
60
bill_length_mm
species
Adelie
Chinstrap
Gentoo