Calculating and Graphing Correlations in R

Andy Grogan-Kaylor

2020-06-11

Contents

1 Palmer Penguins

This example uses the *Palmer Penguins* data set: https://github.com/allisonhorst/palmerpenguins.

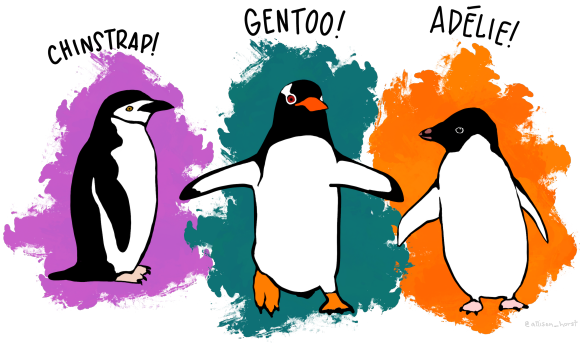


Figure 1: Palmer Penguins Illustration from @allison_horst

2 Look At The Data

```
library(palmerpenguins) # penguin data

library(pander) # nicely formatted tables

pander(head(penguins)) # nicely formatted table of top of data
```

Table 1: Table continues below

species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm
Adelie	Torgersen	39.1	18.7	181
Adelie	Torgersen	39.5	17.4	186
Adelie	Torgersen	40.3	18	195
Adelie	Torgersen	NA	NA	NA
Adelie	Torgersen	36.7	19.3	193
Adelie	Torgersen	39.3	20.6	190

body_mass_g	sex
3750	MALE
3800	FEMALE
3250	FEMALE
NA	NA
3450	FEMALE
3650	MALE

3 Calculate a Correlation

We calculate the correlation of *body mass* and *flipper length*.

> We need to use the option `use = "complete.obs"` to avoid an error message because some observations have missing data.

```
cor(penguins$body_mass_g,
    penguins$flipper_length_mm,
    use = "complete.obs")
```

```
## [1] 0.8712018
```

> There is some indication that penguins with higher body mass have longer flippers.

To get a more nicely formatted correlation value, we can read this correlation into a variable, and then print out this correlation as part of a sentence in inline code. See this RMarkdown document for how this is done, or take a look at this page from RStudio.

```
mycorrelation <- cor(penguins$body_mass_g,
                     penguins$flipper_length_mm,
                     use = "complete.obs")
```

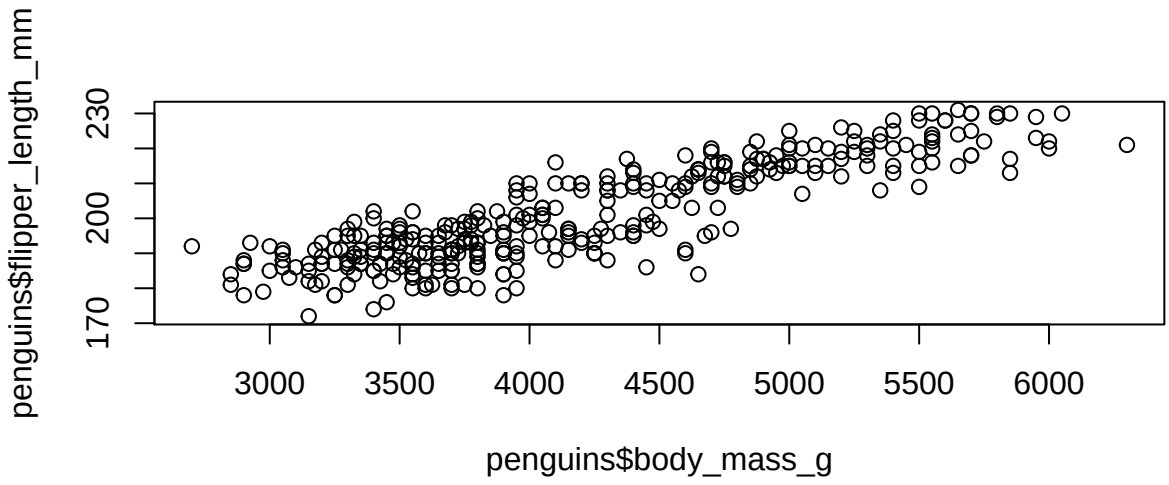
> The value of the correlation is 0.8712018.

4 Graphing

4.1 Base R

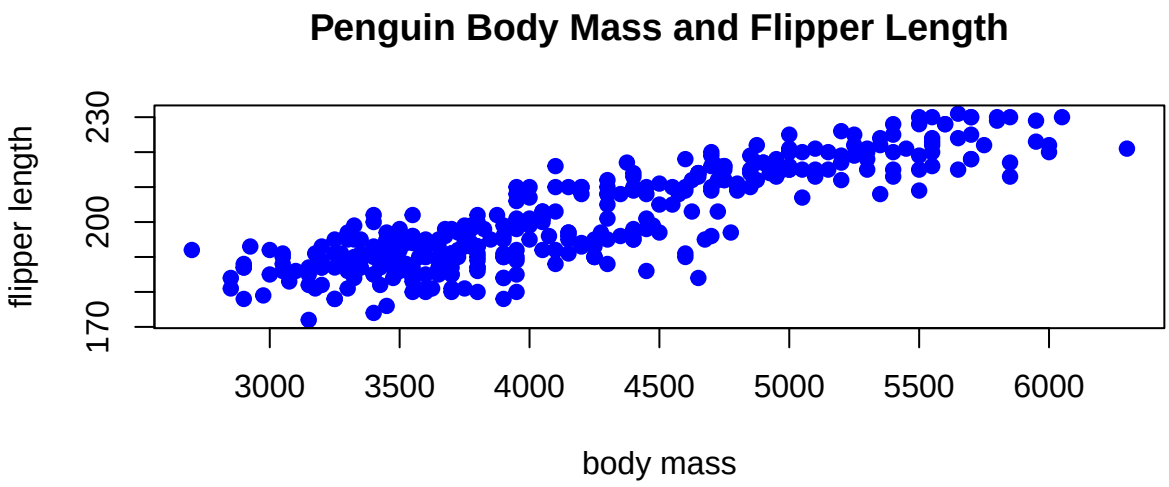
4.1.1 Basic Base R Plot

```
plot(penguins$body_mass_g,
    penguins$flipper_length_mm)
```



4.1.2 Advanced Base R Plot

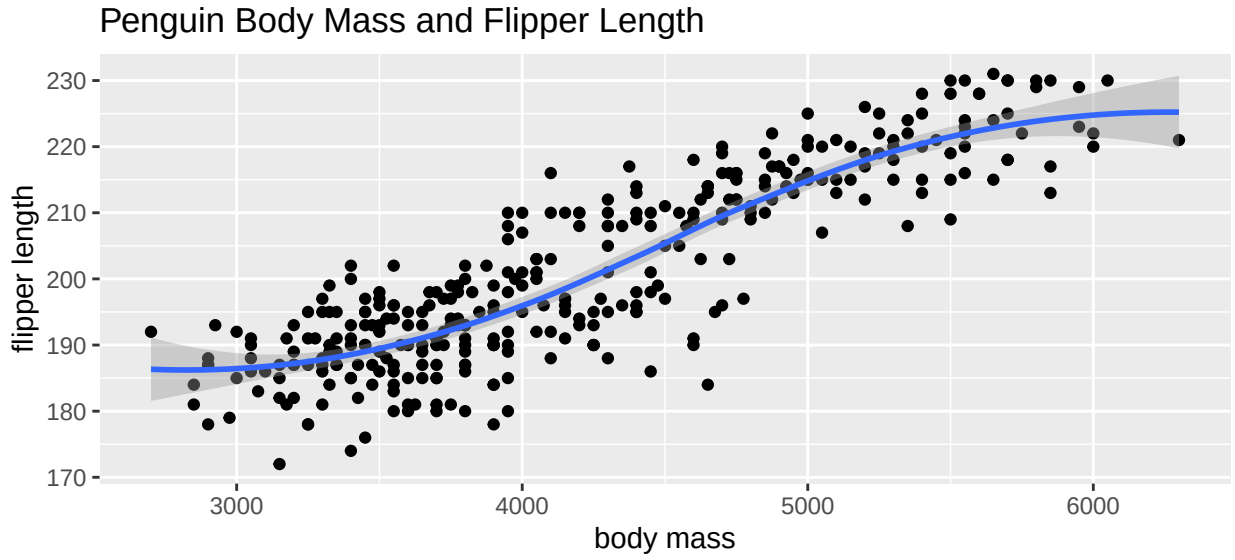
```
plot(penguins$body_mass_g,
    penguins$flipper_length_mm,
    col = "blue",
    pch = 19, # Plotting CHaracter
    xlab = "body mass",
    ylab = "flipper length",
    main = "Penguin Body Mass and Flipper Length")
```



4.2 ggplot

```
library(ggplot2)

ggplot(penguins,
       aes(x = body_mass_g,
           y = flipper_length_mm)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Penguin Body Mass and Flipper Length",
       x = "body mass",
       y = "flipper length")
```



5 Citation

Gorman KB, Williams TD, Fraser WR (2014). Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis). PLoS ONE 9(3): e90081. https://doi.org/10.1371/journal.pone.009008