Penguin Clustering

Table of contents

Note that the numbering here is setup via both the YAML and reference doc.

# Penguins

## About the data

Let’s make a table:

from palmerpenguins import load\_penguins  
penguins = load\_penguins()  
penguins.head(10)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1.1: The first 10 rows of the Penguins dataset   |  | species | island | bill\_length\_mm | bill\_depth\_mm | flipper\_length\_mm | body\_mass\_g | sex | year | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 0 | Adelie | Torgersen | 39.1 | 18.7 | 181.0 | 3750.0 | male | 2007 | | 1 | Adelie | Torgersen | 39.5 | 17.4 | 186.0 | 3800.0 | female | 2007 | | 2 | Adelie | Torgersen | 40.3 | 18.0 | 195.0 | 3250.0 | female | 2007 | | 3 | Adelie | Torgersen | NaN | NaN | NaN | NaN | NaN | 2007 | | 4 | Adelie | Torgersen | 36.7 | 19.3 | 193.0 | 3450.0 | female | 2007 | | 5 | Adelie | Torgersen | 39.3 | 20.6 | 190.0 | 3650.0 | male | 2007 | | 6 | Adelie | Torgersen | 38.9 | 17.8 | 181.0 | 3625.0 | female | 2007 | | 7 | Adelie | Torgersen | 39.2 | 19.6 | 195.0 | 4675.0 | male | 2007 | | 8 | Adelie | Torgersen | 34.1 | 18.1 | 193.0 | 3475.0 | NaN | 2007 | | 9 | Adelie | Torgersen | 42.0 | 20.2 | 190.0 | 4250.0 | NaN | 2007 | |

## Visualising the data

### Bills and Flippers

from plotnine import ggplot, geom\_point, aes, stat\_smooth, facet\_wrap  
from palmerpenguins import load\_penguins  
(  
 ggplot(penguins, aes("bill\_length\_mm",  
 "flipper\_length\_mm", color="factor(species)"))  
 + geom\_point()  
)

/home/vscode/.local/lib/python3.13/site-packages/plotnine/layer.py:364: PlotnineWarning: geom\_point : Removed 2 rows containing missing values.

|  |
| --- |
| Figure 1.1: Scatter plot of bill length and flipper length |

## Modelling the data

We can…