Palmer Penguins Report

from palmerpenguins import load\_penguins  
import pandas as pd  
import numpy as np  
  
penguins = load\_penguins()  
penguins\_cleaned = penguins.dropna()

The *palmer penguins* dataset contains information about 344 penguins, after **discarding** rows with NA values there are 333 penguins left.

Image Credit: Allison Horst

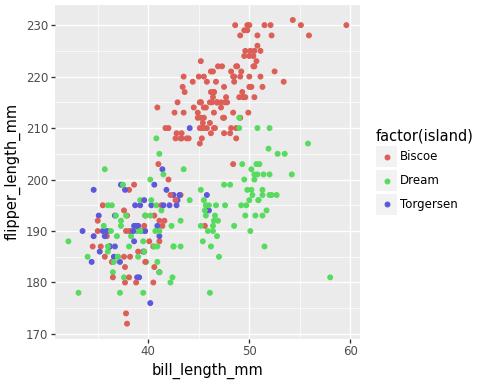
penguins\_cleaned.head()

|  | 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 |
| 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 |

The penguins are differentiated by both island and species, but there is clearer clustering by species when looking at the variables: bill\_length\_mm and flipper\_length\_mm.

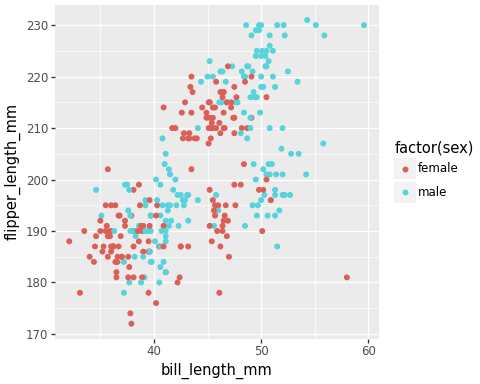
### Island

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



### Sex

(  
 ggplot(penguins\_cleaned, aes("bill\_length\_mm",  
 "flipper\_length\_mm", color="factor(sex)"))  
 + geom\_point()  
)



### Species

(  
 ggplot(penguins\_cleaned, aes("bill\_length\_mm",  
 "flipper\_length\_mm", color="factor(species)"))  
 + geom\_point()  
)

