!pip install palmerpenguins

Collecting palmerpenguins  
 Downloading palmerpenguins-0.1.4-py3-none-any.whl (17 kB)  
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from palmerpenguins) (1.5.3)  
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from palmerpenguins) (1.23.5)  
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->palmerpenguins) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->palmerpenguins) (2023.3.post1)  
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas->palmerpenguins) (1.16.0)  
Installing collected packages: palmerpenguins  
Successfully installed palmerpenguins-0.1.4

import pandas as pd

a = "prit"  
type (a)

str

a = ['Prit' 'Shah']  
type(a)

list

import numpy as np  
ar = np.array(['prit','shah'])  
type(ar)

numpy.ndarray

rob = (["a",1], ["b",2,3])  
rob[1][2]

3

ar = np.array([1,'Shah'])  
ar[1]

'Shah'

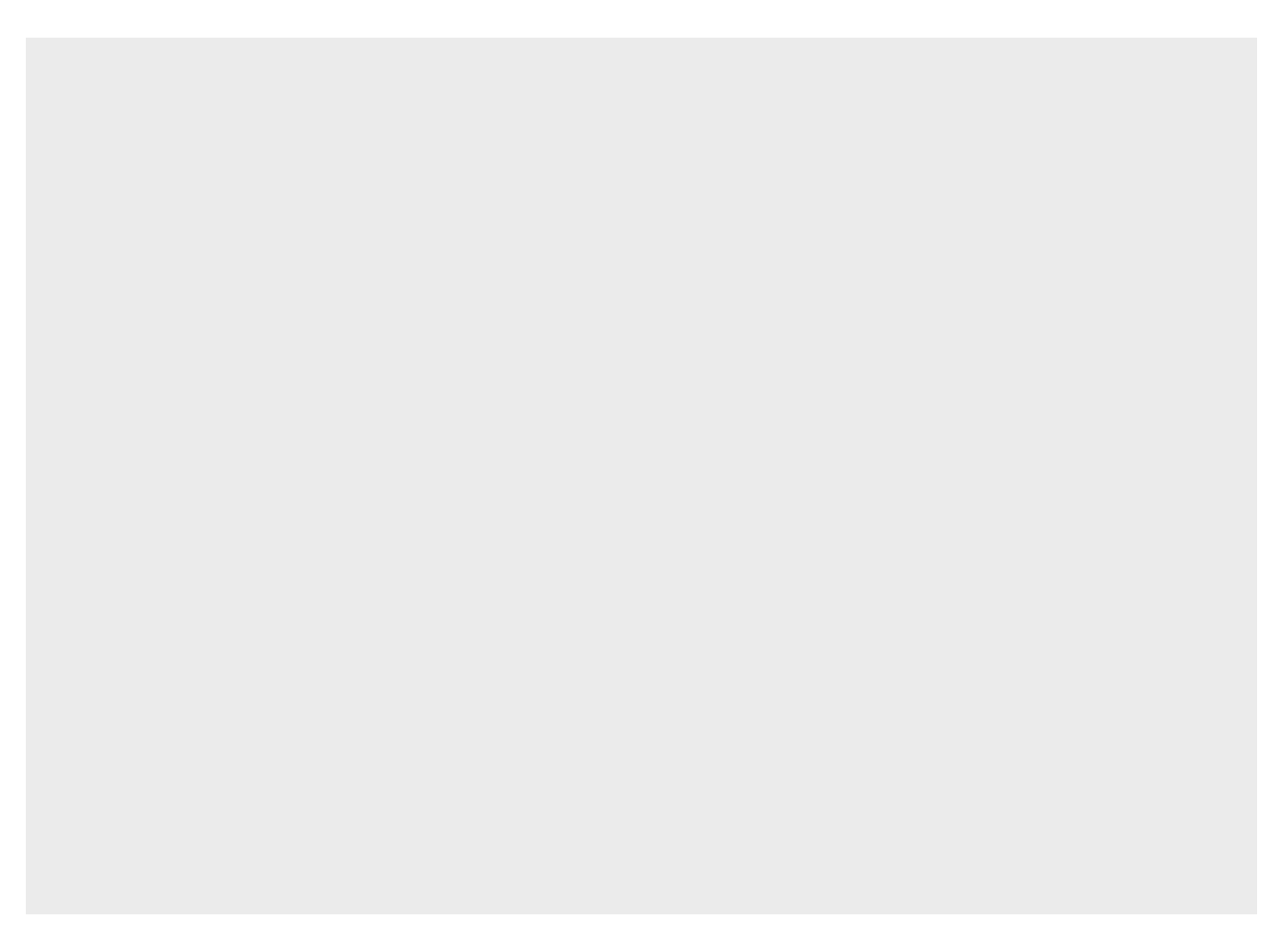
rob = (["a",1], ["b",2,3])  
pd.DataFrame(rob)

|  | 0 | 1 | 2 |
| --- | --- | --- | --- |
| 0 | a | 1 | NaN |
| 1 | b | 2 | 3.0 |

rob = (["a",1], ["b",2,3])  
rob\_df = pd.DataFrame(rob)  
rob\_df.describe()

|  | 1 | 2 |
| --- | --- | --- |
| count | 2.000000 | 1.0 |
| mean | 1.500000 | 3.0 |
| std | 0.707107 | NaN |
| min | 1.000000 | 3.0 |
| 25% | 1.250000 | 3.0 |
| 50% | 1.500000 | 3.0 |
| 75% | 1.750000 | 3.0 |
| max | 2.000000 | 3.0 |

from plotnine import ggplot, aes  
(ggplot(aes()))



<Figure Size: (640 x 480)>

New Packages to Install:

from palmerpenguins import load\_penguins  
load\_penguins()

|  | 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 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 339 | Chinstrap | Dream | 55.8 | 19.8 | 207.0 | 4000.0 | male | 2009 |
| 340 | Chinstrap | Dream | 43.5 | 18.1 | 202.0 | 3400.0 | female | 2009 |
| 341 | Chinstrap | Dream | 49.6 | 18.2 | 193.0 | 3775.0 | male | 2009 |
| 342 | Chinstrap | Dream | 50.8 | 19.0 | 210.0 | 4100.0 | male | 2009 |
| 343 | Chinstrap | Dream | 50.2 | 18.7 | 198.0 | 3775.0 | female | 2009 |

pen = load\_penguins()  
print(pen)

species island bill\_length\_mm bill\_depth\_mm flipper\_length\_mm \  
0 Adelie Torgersen 39.1 18.7 181.0   
1 Adelie Torgersen 39.5 17.4 186.0   
2 Adelie Torgersen 40.3 18.0 195.0   
3 Adelie Torgersen NaN NaN NaN   
4 Adelie Torgersen 36.7 19.3 193.0   
.. ... ... ... ... ...   
339 Chinstrap Dream 55.8 19.8 207.0   
340 Chinstrap Dream 43.5 18.1 202.0   
341 Chinstrap Dream 49.6 18.2 193.0   
342 Chinstrap Dream 50.8 19.0 210.0   
343 Chinstrap Dream 50.2 18.7 198.0   
  
 body\_mass\_g sex year   
0 3750.0 male 2007   
1 3800.0 female 2007   
2 3250.0 female 2007   
3 NaN NaN 2007   
4 3450.0 female 2007   
.. ... ... ...   
339 4000.0 male 2009   
340 3400.0 female 2009   
341 3775.0 male 2009   
342 4100.0 male 2009   
343 3775.0 female 2009   
  
[344 rows x 8 columns]

you can write #pen.species, pen.describe, pen[‘species’]

pen.species

0 Adelie  
1 Adelie  
2 Adelie  
3 Adelie  
4 Adelie  
 ...   
339 Chinstrap  
340 Chinstrap  
341 Chinstrap  
342 Chinstrap  
343 Chinstrap  
Name: species, Length: 344, dtype: object

adelies\_xy = (pen['species'] =='Adelie')  
adelies\_xy

0 True  
1 True  
2 True  
3 True  
4 True  
 ...   
339 False  
340 False  
341 False  
342 False  
343 False  
Name: species, Length: 344, dtype: bool

Only True

pen[adelies\_xy]

|  | 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 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 147 | Adelie | Dream | 36.6 | 18.4 | 184.0 | 3475.0 | female | 2009 |
| 148 | Adelie | Dream | 36.0 | 17.8 | 195.0 | 3450.0 | female | 2009 |
| 149 | Adelie | Dream | 37.8 | 18.1 | 193.0 | 3750.0 | male | 2009 |
| 150 | Adelie | Dream | 36.0 | 17.1 | 187.0 | 3700.0 | female | 2009 |
| 151 | Adelie | Dream | 41.5 | 18.5 | 201.0 | 4000.0 | male | 2009 |

Only False

pen[~adelies\_xy]

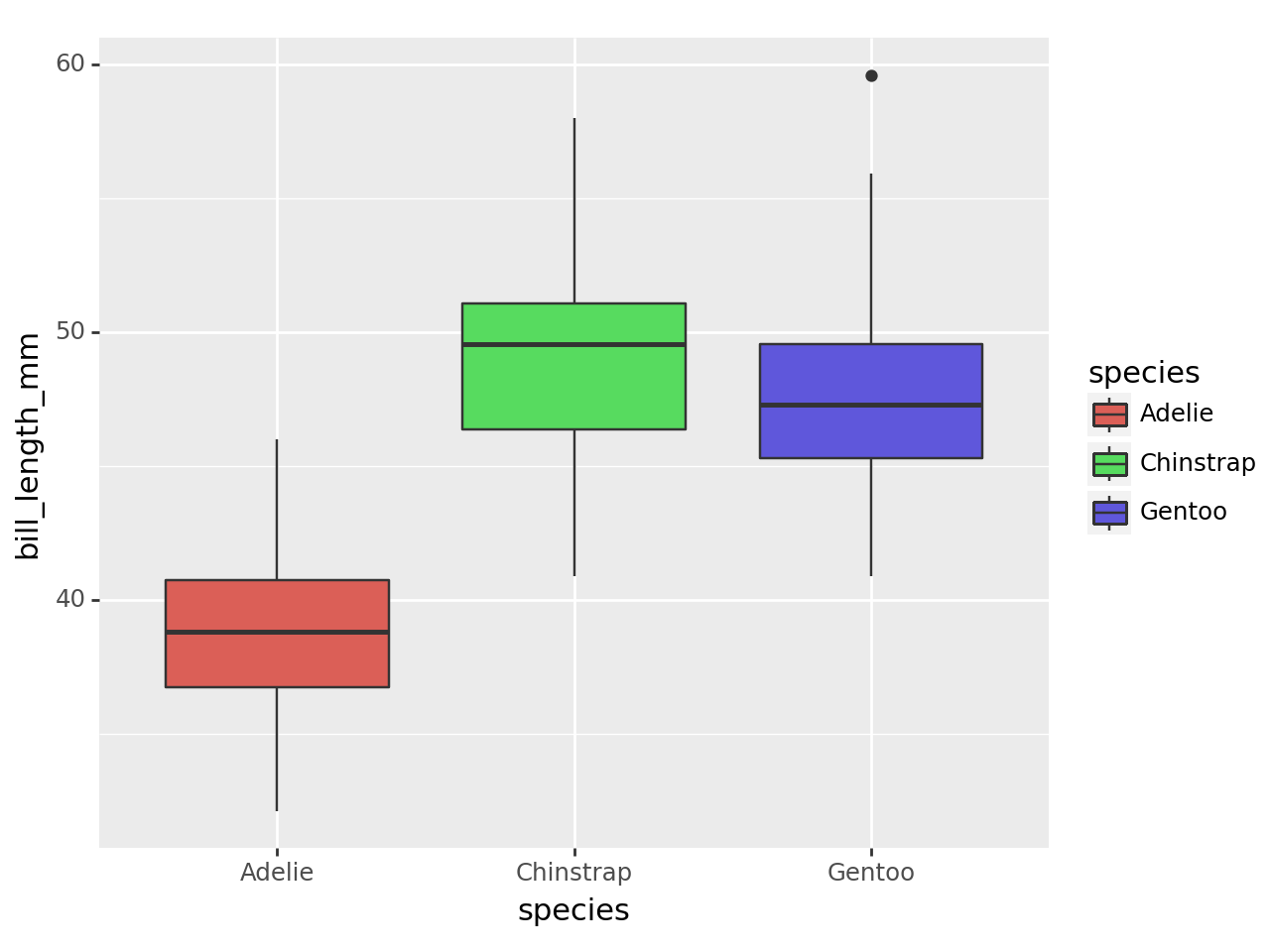
|  | species | island | bill\_length\_mm | bill\_depth\_mm | flipper\_length\_mm | body\_mass\_g | sex | year |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 152 | Gentoo | Biscoe | 46.1 | 13.2 | 211.0 | 4500.0 | female | 2007 |
| 153 | Gentoo | Biscoe | 50.0 | 16.3 | 230.0 | 5700.0 | male | 2007 |
| 154 | Gentoo | Biscoe | 48.7 | 14.1 | 210.0 | 4450.0 | female | 2007 |
| 155 | Gentoo | Biscoe | 50.0 | 15.2 | 218.0 | 5700.0 | male | 2007 |
| 156 | Gentoo | Biscoe | 47.6 | 14.5 | 215.0 | 5400.0 | male | 2007 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 339 | Chinstrap | Dream | 55.8 | 19.8 | 207.0 | 4000.0 | male | 2009 |
| 340 | Chinstrap | Dream | 43.5 | 18.1 | 202.0 | 3400.0 | female | 2009 |
| 341 | Chinstrap | Dream | 49.6 | 18.2 | 193.0 | 3775.0 | male | 2009 |
| 342 | Chinstrap | Dream | 50.8 | 19.0 | 210.0 | 4100.0 | male | 2009 |
| 343 | Chinstrap | Dream | 50.2 | 18.7 | 198.0 | 3775.0 | female | 2009 |

To plot the Graph and you need Data, Geometry- what kind of plot, Mapping/ Aesthetic —- which varibales create the plot.

import pandas as pd  
from palmerpenguins import load\_penguins  
from plotnine import ggplot, geom\_point, aes, geom\_boxplot

(ggplot(pen, aes(x = "species", y = "bill\_length\_mm", fill = "species"))  
+ geom\_boxplot()  
)

/usr/local/lib/python3.10/dist-packages/plotnine/layer.py:284: PlotnineWarning: stat\_boxplot : Removed 2 rows containing non-finite values.



<Figure Size: (640 x 480)>

1. Scales - numeric/ log, colors
2. Theme: Visual Element axis lebels/ticks annotations Font & Color
3. Facats