matplotlib_scatter_graphs_workbook_answers

September 21, 2020

1 MATPLOTLIB SCATTER GRAPHS WORKBOOK ANSWERS

Remember, there are different ways to write code to get the same answer, so your answer can be correct and different to the answer example!

If you feel stuck and want some in person help, then have a look at the events page to join in a workshop https://swamphen.co.uk/events.

In [ ]:
```
# upload the butterflies data set
from google.colab import files
uploaded = files.upload()
```

In [ ]:
```
# read the butterflies data set in to this colab sheet
import csv
with open('ButterfliesBc.csv') as data:
    reader = csv.reader(data, delimiter = ',')
    for i in reader:
        my_list = list(reader)

print(my_list)
```

In [ ]:
```
# split the data into 3 lists.
# 0th element is temperature (C), needs to be changed into a float
# 1st element is wing length (mm), needs to be changed into a float
# 2nd element is sex, needs to be changed to 0 or 1

temp = []
wing = []
sex = []

for element in my_list:
    temp.append(float(element[0]))
    wing.append(float(element[1]))
    if element[2] == 'Female':
        sex.append(0)
    else:
        sex.append(1)

print(temp, wing, sex)
```

```
In [ ]: # import matplotlib
        import matplotlib.pyplot as plt

In [ ]: # create a scatter plot with green dots for temp
        plt.plot(temp, 'g.')

In [ ]: # what does this tell you about the data?

        # repeated set of measurements of male and female
        # should be done with 2 axes

In [ ]: # create a scatter plot with red stars for wings
        plt.plot(wing, 'r*')

In [ ]: # what does this tell you about the data

        # two sets of data, slight overlap

In [ ]: #create a scatter plot of temperature against wingspan with blue circles
        # label the axes
        plt.scatter(temp, wing, color = 'b', marker = 'o')
        plt.xlabel('temperature (C)')
        plt.ylabel('wing span (mm)')

In [ ]: # what does this tell you about the data?

        # no link between wingspan and temperature
        # data not evenly taken across temperature range

In [ ]: # does there look to be any outliers?

        # the two at 4 deg C

In [ ]: # look at the relationship between wingspan and sex with red diamonds
        # add in axis labels
        plt.scatter(wing, sex, color = 'r', marker = 'd')
        plt.xlabel('temperature (C)')
        plt.ylabel('sex (0 = female, 1 = male)')

In [ ]: # what does this tell you about the relationship between wingspan and sex

        # overlap between sizes, but female butterflies tend to be larger

In [ ]: # does there look to be any outliers?

        # the point at 19.5deg C

In [ ]: # do any of the points you identified as outliers match between the two graphs?
        # do you think there are any outliers in this data set?

        # no
```