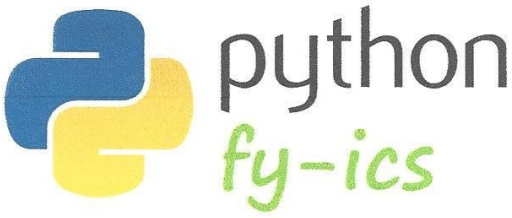


Assessed Work 2021/2022 Penguins

Summary
You are asked to write a Python 3 program that will read a CSV data file and output selected data in the forms of a bar chart and a scatterplot.

Data source

The CSV file holds data relating to the attributes of penguins, specifically three species of penguins:

Column	Column heading	Description	Type
1	id	Identifier unique to each sample	integer
2	species	Species	string
3	bill_length_mm	Length of the penguin's bill	float
4	bill_depth_mm	Depth of the penguin's bill	float
5	flipper_length_mm	Length of the penguin's flipper	float
6	body_mass_g/100	Weight of the penguin (in grams / 100)	float

`clean_penguins.csv` is available in Canvas. You should copy it into the directory in which your program resides.

Python configuration

This exercise requires the use of Python 3 together with matplotlib.

Reading data

Your program should read data from a CSV file using the csv module, opening the CSV file appropriately [lecture 9].

Reading data from the file into the program will require type casting from strings to floats [lecture 3].

Using list processing techniques (including list slicing) [lecture 8] will allow you to store the data in a reasonably convenient way. It is suggested that you read the CSV file into a list. Each element in the list would represent an individual penguin. Each individual record would contain the ID, species name and four tuples consisting of pairs of the kind of data observed and its value, for instance:

```
[…,
['97', 'Gentoo', ('bill_length_mm', 45.8), ('bill_depth_mm', 14.6),
('flipper_length_mm', 210.0), ('body_mass_g/100', 42.0)]
['98', 'Chinstrap', ('bill_length_mm', 50.9), ('bill_depth_mm', 19.1),
('flipper_length_mm', 196.0), ('body_mass_g/100', 35.5)],
…]
```

Choosing data records for your scatter plot and bar chart

The scatter plot requires the choice of a penguin's attribute by name and the bar chart requires the choice of an ID number, using input from the keyboard [lectures 2 and 3].

Outputting the data

The output uses the mathplotlib module [lecture 9] to represent the data. The required parameters are strings, floats and lists [lectures 3, 7, 8].

Offering the user the choice of a plot or a chart or to stop

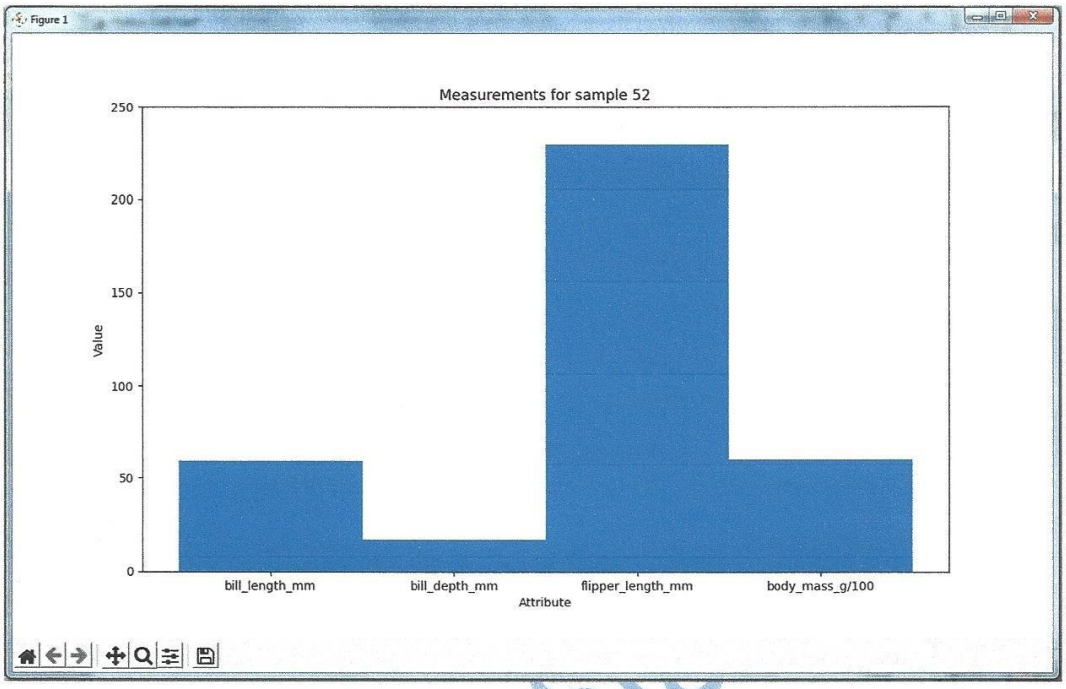
The user is allowed to set up another scatter plot or bar chart or to finish their session. This requires input from the user [lectures 2 & 3] and the use of iteration [lecture 6]. You will probably also need to use assignment with a conditional [lectures 4 & 5]

What does this program look like when it is running?

Here is an example of a program made earlier (with user input in bold):

```
Penguin attribute analysis

Please choose one of:
      1 - display penguin's measurements
      2 - display scatter plot of attribute's measurements
      3 - exit the system
Your choice? 1

Please enter penguin ID number? 52
```

At this point a bar chart was displayed:



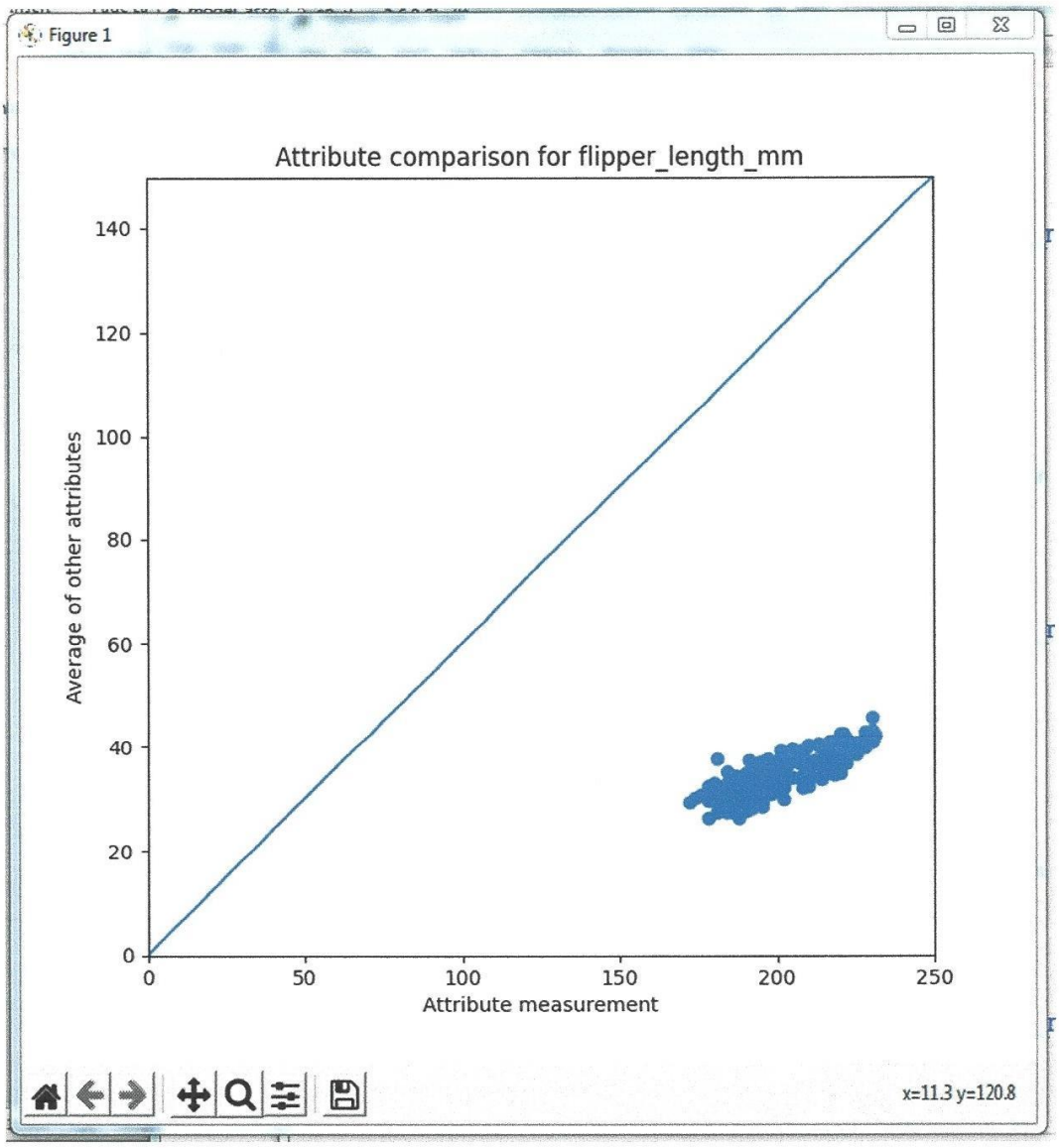
then processing resumed:

```
Penguin attribute analysis

Please choose one of:
  1 - display penguin's measurements
  2 - display scatter plot of attribute's measurements
  3 - exit the system
Your choice? 2
```

A choice of flipper length as the attribute against which the other numerical attributes are compared was made:

```
Attribute options are:
bill_length_mm    bill_depth_mm     flipper_length_mm body_mass_g/100

Please enter attribute? flipper_length_mm
```

At this point a scatter plot was displayed:



then processing resumed:

```
Penguin attribute analysis

Please choose one of:
  1 - display penguin's measurements
  2 - display scatter plot of attribute's measurements
  3 - exit the system
Your choice? 3
```

System requirements

The submitted code must be written in Python 3 in the 3.6.n version. The code that you submit will be tested using this version of Python.

What the assessors are looking for

This is "summative" work. There is a mark which contributes 30% of the final mark for *Elementary Computer Programming*. However, it is intended primarily to help you develop your skills.

The assessors are looking for a solution to the problem (ie a program that runs). There are marks for parts of the task so that, if you cannot complete all of the task, you should still submit the code you have developed.

A good solution should use appropriate constructs. (That means, for instance, that you use the more appropriate kind of loop if you have to use iteration.)

Output should be well presented – the use of English should be good and the layout consistent.

Duplicate code should be avoided.

10% of the mark for this exercise will be awarded for making the response to the user's input robust, for instance, to recognise invalid input for an ID number or a module name.

As university study is about becoming a self-sufficient learner capable of extending knowledge beyond lecture material, 10% of the mark for this exercise will be awarded for substantial use of a technique or technology not introduced in lectures 1 to 9. Two possible examples are the use of Python dictionaries (hash tables) to store penguin data instead of nested loops; use of a Tkinter Graphical User Interface instead of a text interface. (Be warned: Tkinter GUIs are difficult.)

Credit will be given for only one substantial extra technique – so a student using both dictionaries and a GUI would not get 20% (ie a possible total of 110%). The judgement of substantiality is reserved for the assessors and their judgement is final.

Lab support

The usual arrangements for assistance during lab sessions remain in place.

When the work is due

You should submit your program file by uploading to Canvas by 12 noon (BST) on Monday 16 May 2022.

Demonstrating your work

As part of the assessment process, there may be a demonstration as an opportunity for students to explain the working of their program. These will be held after the submission date.

What this assignment contributes to the module mark

The mark contributes 30% of the final mark for *Elementary Computer Programming*.

Feedback

Feedback will be given through a comment sheet on Canvas.

Plagiarism

It should go without saying that you should not copy code from the WWW – the exercise has been written to differ from what is available. Canvas's plagiarism detector will, of course, alert the markers to plagiarism. Equally, the use of "tutorial assistance" websites is not allowed – and these are monitored.

All files will be scanned by a plagiarism checker that specialises in detecting plagiarism in programs. (It is particularly good at spotting where a program has been altered by changing names of variables and functions and moving a few lines around.) If plagiarism is suspected it will be investigated and, if appropriate, reported.

As a guide: you will want to discuss the work with friends. If you discuss the general way in which the problem can be solved, your work will not be considered plagiarised. If you swap specific lines of code with friends and include them in your program, the plagiarism checker is very likely to detect them.

If you are not intending to plagiarise, follow these guidelines and don't worry: deliberate plagiarism tends to look very different from high-level discussion.

Hints on producing a solution

Don't start by trying to write the whole program in one go. Divide the assignment into parts and encapsulate these in user-defined functions. Some of the obvious parts include:

1. *Reading data from the CSV file*
   Test that you can do this by opening a CSV file and printing each row

2. *Producing a bar chart*
   You might develop your skills by writing a function with the appropriate parameters that displays a bar chart. You can then use this in your final program.

3. *Producing a scatter plot*
   You might develop your skills by writing a function with the appropriate parameters that displays a scatter plot. You can then use this in your final program.

4. *Obtaining the user's input to choose which action is to be performed.*

You are encouraged to talk through your solution with the demonstrators.

Some extra programming items that might be useful:

Appending a tuple to a list
If you have two data items that are paired, then it is easier to keep them together in a tuple. For instance, if you were storing family and given names in a list, you might append the next name to the list like this:

```
names.append(("Nelson", "Mandela"))
```

Setting the size of a Matplotlib window
Plot sizes should be set before anything else, even before setting title, axis labels and creating the plot. An example of the command is:

```
plt.subplots(figsize=(120, 70))
```

Formatting a bar chart
For a successful bar chart, you need to supply the number of bars required (using the range function); set the width of the columns and centre their labels; set the scale on the axes and insist that attributes are shown from (eg) 0 to 8. Here is code that will do that:

```
x_count = range(len(ylist))
plt.bar(x_count, ylist, width=1, align="center")

# set ticks and scale on y axis
plt.xticks(x_count, xlist)
plt.ylim(0, 8)
```

Adding a straight line to a scatterplot is easy
You simply plot the scatterplot and then the straight line graph:

```
# create scatter plot
plt.scatter(attribute_lengths, other_attribute_lengths)

# add x=y line
plt.plot([0, 10], [0, 10])
```

FINIS