10/3/21

PROJECT 2021

Programming and Scripting

Due: last commit on or before April 30th, 2021.

This document contains instructions for Project 2021 for Programming and Scripting. You are not expected to know how to do the whole project from the beginning. Rather, we expect that you research ways to tackle the project and formulate your own submission based on your investigations. Remember, all students are bound by the GMIT's Quality Framework [2] including the Code of Student Conduct and the Policy on Plagiarism.

Problem statement

This project concerns the well-known Fisher's Iris data set [3]. You must research the data set and write documentation and code (in Python [1]) to investigate it. An online search for information on the data set will convince you that many people have investigated it previously. You are expected to be able to break this project into several smaller tasks that are easier to solve, and to plug these together after they have been completed.

You might do that for this project as follows:

1. Research the data set online and write a summary about it in your README.
2. Download the data set and add it to your repository.
3. Write a program called analysis.py that:
    - outputs a summary of each variable to a single text file,
    - saves a histogram of each variable to png files, and
    - outputs a scatter plot of each pair of variables.

It might help to suppose that your manager has asked you to investigate the data set, with a view to explaining it to your colleagues. Imagine that you are to give a presentation on the data set in a few weeks' time, where you explain what investigating a data set entails and how Python can be used to do it. You have not been asked to create a deck of presentation slides, but rather to present your code and its output to them.