Bioinformatics analysis with Python 1, BI121G, spring 2025

Assignment 3

Instructions

This assignment consists of 2 tasks related to the lecture material.

For task 1 you will get 0-15 points, as follows:

- 0 points if an answer is missing, is completely wrong or generates syntax errors.
- 1-10 points if the program is able to run without generating any syntax errors but contains major semantic errors, incorrect or missing implementations, or you are using badly chosen programming constructs. The code is at least sparsely described and does not contain any major incorrect descriptions of the code. The amount of points depends on the number of requirements that have been correctly implemented and how much of the code has been commented.
- 11-14 points if the program is able to run without generating any syntax errors but with some minor errors and minor use of badly chosen programming constructs. The code is well described and does not contain any major incorrect descriptions of the code. The amount of points depends on the number of requirements that have been correctly implemented and how much of the code has been commented.
- 15 points if the program is able to run without generating any syntax errors, includes an implementation of all requirements and contains no semantic errors. The program is well structured, for example, have explanatory variable and function names, and in detail and correctly described.

For task 2 you will get 0-5 points, as follows:

- 0 points if the answer is missing or if it is completely wrong.
- 1 point: The answer is given on a very general level and with no details referring back to the code.
- 2-4 points: Some of the answer is given on a detailed level with some references to the specifics in the code. The amount of points depends on the level of details and how much of the answer refers back to the code.
- 5 points: The answer is given on a detailed level with references to specifics in the code.

The maximum score for the assignment is 20 and you get a grade according to this scale:

F: 0-9, E: 10, D: 11-12, C: 13-16, B: 17-18, A: 19-20

Requirements for tasks

Your files must be uploaded in the assignments section of the course site.

- When the task requires you to implement a Python program, you have to submit your solution as a Python script, i.e., in a file with file extension ".py". Program solutions to tasks that have not been written in the same file and are not in .py-format will not be corrected.
- The Python program must include comments in your code which explains the important statements. From your description it should be clear what the code does.

The assignment is solely individual work since it is part of the examination for the course. It is not allowed to cooperate with other students or copy code from other sources when you answer the questions. It is not allowed to use an AI for answering any of the questions in the assignment. Any undue actions will be reported to the Disciplinary Committee.

If anything is unclear, then send me an email at angelica.lindlof@his.se.

1. Implement a program in Python that reads in an expression data file into a pandas data frame. The first column in the file should be set as the index of the data frame. The program should thereafter:

- generate a boxplot of each column and these boxplots should be plotted in the same figure.
- Print statistics for the data frame using the method `describe()`.
- ask the user for a gene name and make a line plot for that gene.

The program should also fulfill the following requirements:

- Include at least two functions developed by yourself for solving the problem.
- One function should return a variable.

The program should be tested on the file given in Canvas.

2. For the functions you developed by your own you should in this task motivate and describe:

- Why you decided to include these statements in each function.
- Why you decided to divide the program into these functions.
- Why you think these functions are good for solving the problem.