Y2 Project plan

1. Personal information

Title: Numerical data visualization library

Creator: Oona Sohlberg

Student number: 100752384

Study program: BioIT

Study year: 2nd

Date: 9.2.2024

2. General description and difficulty level

I will be doing a numerical data visualization library where users can plot line diagrams, histograms, and pie diagrams. I will be using PyQt6 library.

My program will stand out from others because I will make some algorithms in it. Users can, for example, sort histogram from smallest to largest. Also finding maximum and minimum values from diagram will be possible.

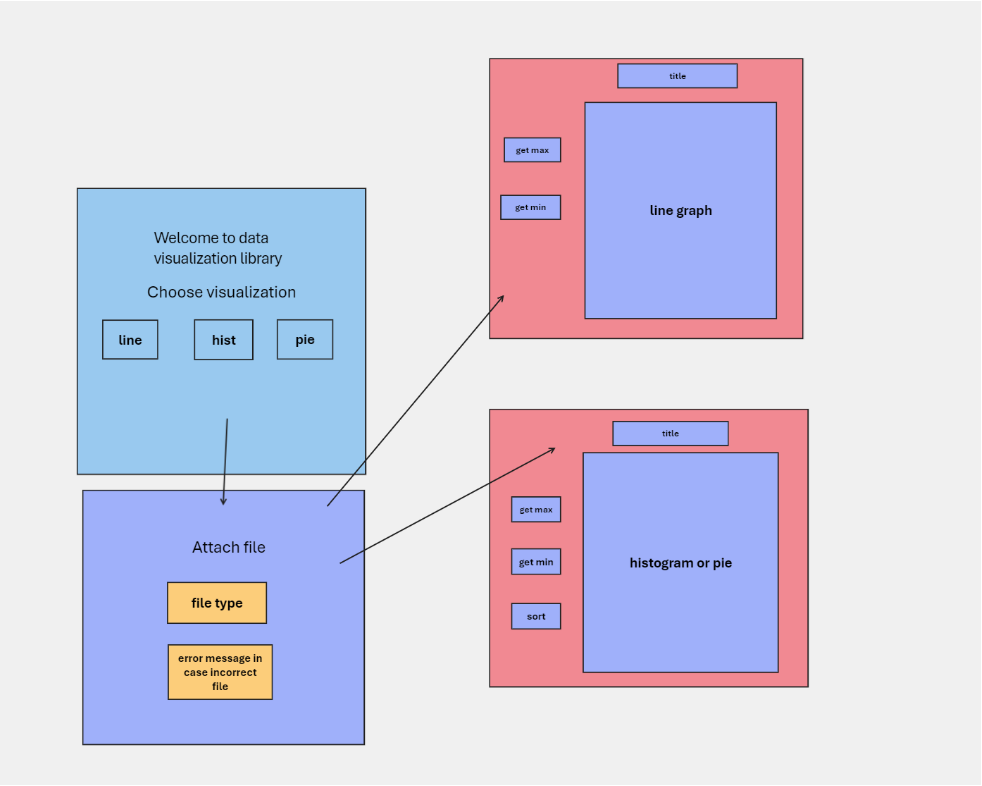
I am planning to do a hard program, but if it gets too difficult or I do not have enough time, I will do a medium hard.

3. Use case description and draft of user interface

First there will be a menu, where the user can select what kind of plot, he/she wants. After that the user adds a file. In the case of plot diagram, user must add a file which has data in format: x, y \n, where x and y are digits (float or int). Grid will be added to this kind of visualization. In the case of histogram or pie diagram, user must add a file where data is in format: (x = text, y), where text is some float, int or char and y is some digit. The user must first choose visualization type and then add the file. The user can add a title and titles for x- and y-axis. After that, the visualization will be drawn.

In case of wrong kind of input, the program will send an error message to user. It will be shown above the “drop a file” box.

The program knows a “sort” -command, which sorts the values of histogram or pie diagram from smallest to largest. It also knows get max and get min commands, which will get the maximum and minimum values (the value in x-axis corresponding to the y-value). These values will be shown below the buttons “get minimum value” and “get maximum value”. Points in line diagram will be colored differently for y-values that are below and above mean. There will be an explanation for that. As a new window opens, the previous one will close. So, in my program only one window will be open at a time. In each window, the menu window will be a back button, which will take the user back to the previous window.

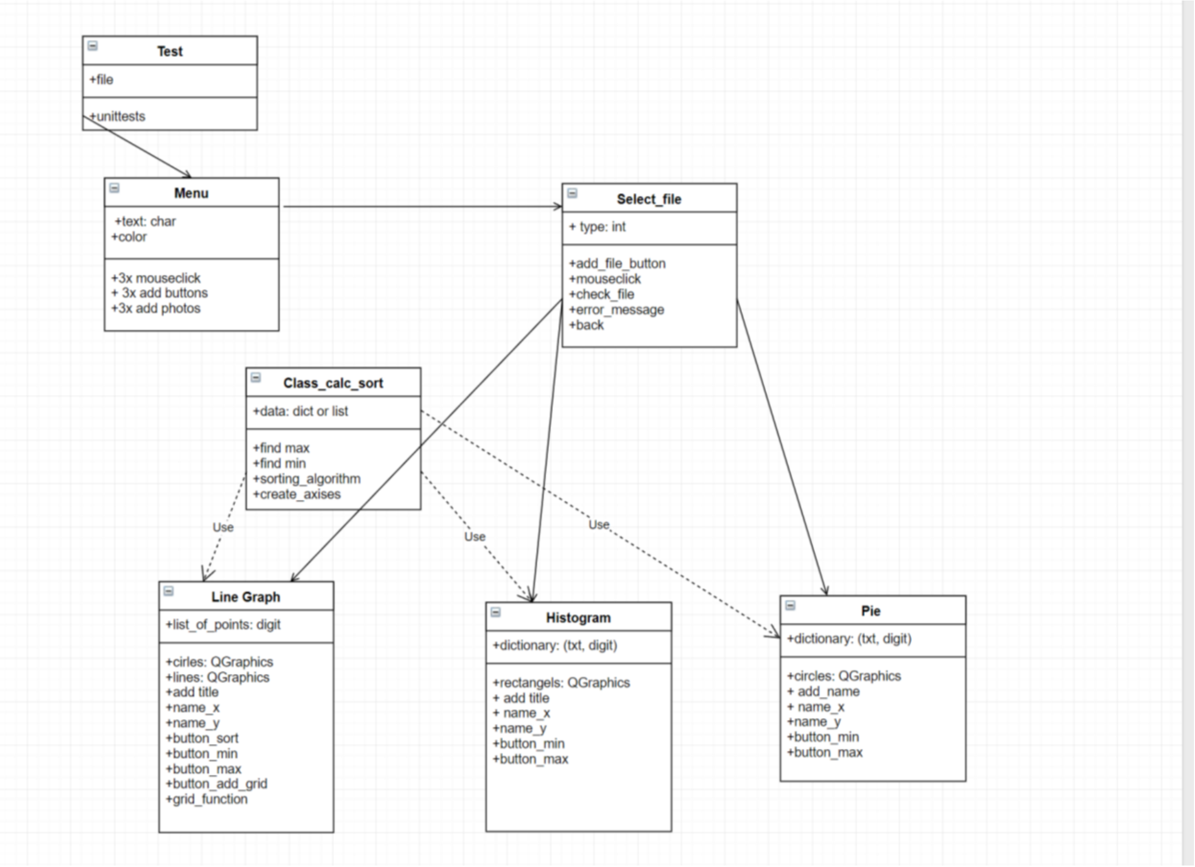


Example user case: User wants to plot line diagram of file, which contains Apple stock prices from year 2000 to 2022. Once the user opens my visualization program, he/she faces the menu. There the user thinks of a suitable visualization and ends up with histogram. (As there are supposed to be example pictures below or above each type). As some type is clicked, a new window will appear and the menu window closes. There will be a mouse click method for this case.

Then the user will attach a file. If the file is in expected format (There will be a check function for that), the visualization window will open. We now expect that the file is in correct format, so the stock prices will be integers or floating-point numbers.

Now the histogram is drawn. In the case of histogram, the user might be able to change the color from some number of colors. The histogram is drawn with pyQt6 rectangle and using loop or recursion. The user can set a grid, where he/she can see values more exactly (forgot to add a button for that, but it will be left to the graph). Now the user can check for example which year had the highest stock price by clicking the button. After all the user can exit to previous page (attach a new file) or exit the program.

4. Project structure plan



These are the main windows and functionalities of the program. Edit. Histogram won’t need a name to x-axis

5. Data structures

As I won't accept too big data, I think I don’t need to store it in abstract way. I will use pythons lists and dictionaries to store the data.

6. Files and file formats

I will load three photos (PNG) into my menu. User will attach a file where the data is to visualize. In the case of line diagram, it should be in format (x,y \n), where x and y are some digits. In the case of pie diagram or histogram, it should be in format (txt,y \n), where txt can be either text or digits and y should be digits. I won't accept (txt,y \n) -files over size 20, because the visualization wouldn’t be clear for too big data. (x,y \n) -files should be under size of 50.

7. Algorithms

As my data is quite small, I won’t use any fancy sorting or searching algorithms. For searching the min and max values, I will loop through the data. For sorting the data (as the size <= 20), I will use insertion sort, because it’s simple, stable and can be done for dictionaries. As I create the geometric items, grid and axis's, I will use loop or recursion to get the size I want.

8. Testing plan

I will be doing unittests, which test for example raising an error message when user’s input file is incorrect. I will test if the buttons in menu work correctly, so after they are pressed files should be in specific format and contain certain amount of data. I will check with tests that my algorithms work as they should. The size of geometric items and axes should be tested as well.

9. Libraries and other tools

PyQt

10. Schedule

12-16.2: Project plan (10h)

19.2-23.2: Start creating main window (10h)

24.2-1.3: Main window + possibly the second window (10h)

2.3-8.3: Second window (10h)

9.3-15.3: Second window + line graph window (10h)

16.3-22.3: Line graph+ algorithms + geometry (10h), 3th checkpoint

23.3-5.4: Line graph window should be working +testing for that+ start creating other graphs (20h)

6.4-12.4: Other graphs and testing, 6th checkpoint (10h)

13.4-(10.5, possibly earlier): Finishing all windows + testing + edit details (up to 40h)

11. Literature reference and links

I will be using internet and Y2 course materials as literature references.