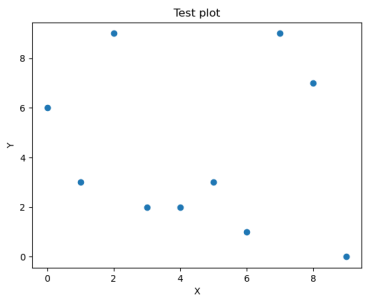
Data Science in Bioinformatics

Day 3

Questions marked with * are optional and may help you understand some topics more deeply.

1) Conda

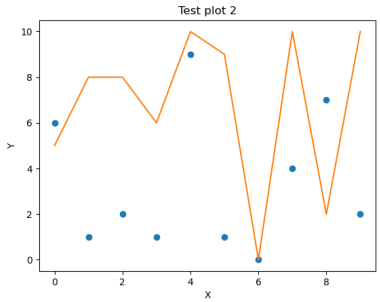
a. Why is it a good idea to use something like conda/mamba?
b. Install mambaforge using the link provided in the lecture slides. After successful installation you should see a “(base)” prompt in front of your terminal
c. To see which packages are installed run: conda list
d. Create a new environment called “plot” and install the package matplotlib via conda.
e. Change to your home directory.
f. Create a folder called “test”.
g. Switch to that folder via cd command.
h. Test the environment by creating a python file in the test folder. Write a small script to plot 10 randomly selected datapoints with a range between 0 and 10 into a diagram and save the resulting plot as a png-file. Please plot the datapoints as dots, add labels to the x- and y-axis and a title.



2) Git and GitHub

a. What is the difference between Git and GitHub? Why are both so important?
b. If you have never worked with GitHub before, please see https://docs.github.com/en/get-started/quickstart/hello-world for more information.
c. Please create a GitHub account if you don’t have one already.
d. Create a new private repository with a file called „test.py“.
e. Clone the repository to your local machine.
f. Create a new branch for the changes you want to make and call it “first-change”.
g. Change the „test.py“ file and add the code you tested your plotting environment on, to the file. Run the file in the command line of your chosen IDE to verify it works.

h. Stage and commit the changes made to the file and push the changes to your repository.
i. Change to GitHub and create a pull request for the changes currently made.
j. Merge the changes into your main branch.
k. Create an issue in your git repository stating that a second list of random numbers should be added to your file and be plotted. These values should be added as line in your existing plot.
l. Make the changes in your code, push the changes on GitHub again and then link the created issue to the opened pull request.



* 1. What is a “bare” git repository?
* 2. How would you set up a simple git server on a local network with no internet connectivity and only a git installation available?
The term “git server” in this case refers to a central location through which multiple people can collaborate by using git. Let’s assume every collaborator has access to the central server via ssh.
* 3. How can you make this setup more resilient against failure of the central server?
* 4. Read up on “Patterns for Managing Source Code Branches” by Martin Fowler (https://martinfowler.com/articles/branching-patterns.html)

3) Container and VMs

a. Under which circumstances and why is the use of docker containers a helpful tool?
b. How do containers differ from VMs and when do you use VMs instead of containers?
c. What is a docker image and why is it important for the creation of a container? And how are docker files involved?