CSci 3003: Introduction to Computing in Biology

Lab Assignment #6

10 points
Assigned: 10/31/19
Due: Thursday, 11/07/19 (before 11:55pm)

Goals of this lab:

- Become familiar with the R interface and RStudio.
- Run/modify a simple R script.

Part I: Instructions for installing R and RStudio

If you are using the CSE-Lab computers, R is already installed. You can move on to Part II. If you would like to work on your own computer, follow these instructions to install R and RStudio:

- At this point, we assume all of you are using anaconda to run Spyder and therefore should be able to install Rstudio (and with it, R) using anaconda-navigator. The figure below shows a snapshot of the navigator window just one click away from the installation of Rstudio.
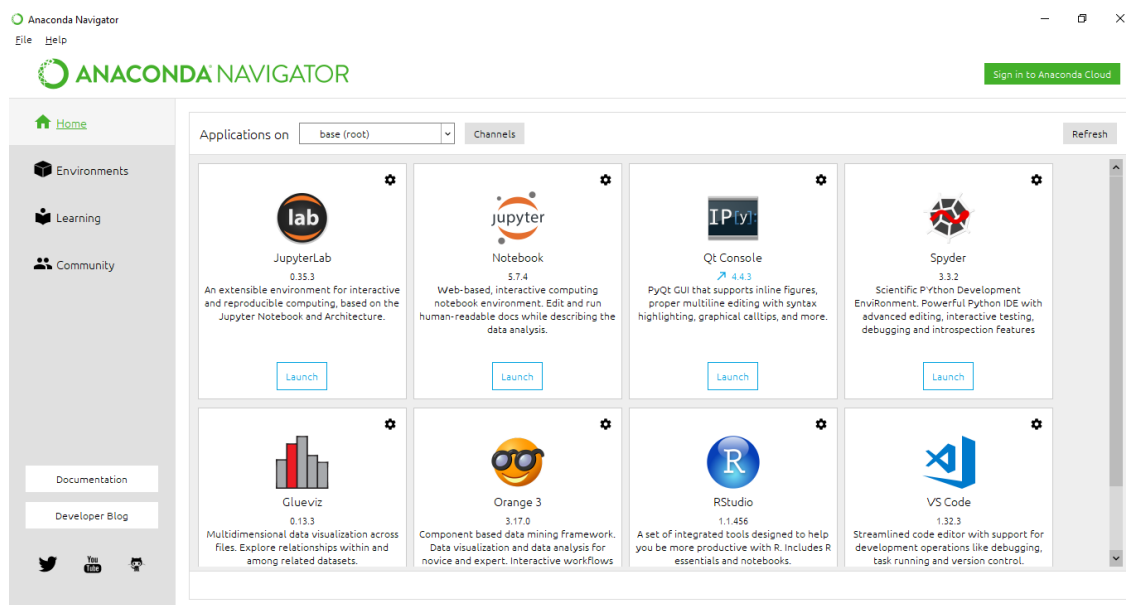

Figure: Rstudio installation. The navigator lists available software; click the Install button and change from the settings (button on the top-right corner) for a specific version.

Part II: Getting familiar with RStudio

RStudio is a development environment like Spyder. It enables you to write and run code all from the same program. To start RStudio from the shell can launch Rstudio from the anaconda navigator; to get to the graphical interface of the navigator, type in *"anaconda-navigator &"* in the shell or if you prefer to start rstudio directly, type in *"rstudio &"*.

Now that we have Rstudio working, let's move on to the assignment:

1. Download the R script for this lab and the data for today's lab from the course Moodle site.

   These data were derived from a microarray-based expression study: the first column is the microarray probe name, the second column is the gene name, the third column is a short description of the gene, the fourth column is the expression levels for a tumor sample #1, and the fifth column is the expression levels for tumor sample #2. These data are from a study in which samples were separated based on the estrogen receptor (ER) status (either positive or negative), which is an important clinical marker for determining treatment. Tumor one is from an ER+ patient and tumor two is from an ER- patient.

   Start RStudio using the instructions above. Be sure to change the current directory to the folder where you download the script and data file above by selecting the directory in Session -> Set Working Directory -> Choose Directory.

2. Open the script in RStudio by clicking File->Open File.

3. Modify the script where asked, get familiar with the R interface, and answer any of the questions. You can run your script a couple different ways. The "run" icon will run a selected chunk of code or one line per click. To run your script in its entirety, click the "source" icon. In total, you should answer ten different questions. Capture the output of your script, including any figures, and turn them in with your report.

Submit to Canvas

When you're finished with the lab, make a report of any questions you answered plus any requested output, and gather the scripts that you modified. Submit your homework file on the course Canvas website.