

BIOC8145 Assignment #1
(100 pts total; Due Wednesday 4/1/20)

1. Go to the following Rivanna “Logging In” web site:
<https://www.rc.virginia.edu/userinfo/rivanna/login/>
2. Follow the instructions for Secure Shell Access (SSH) for your platform. For example, if you’re using Windows, they recommend downloading MobaXterm (ssh client). This is because Windows is a UNIX based operating system. Mac OSX and Linux users can access their terminals. For Mac users, go to your “Utilities” folder and click on “Terminal”. Try to log into Rivanna.
3. On your browser, go to the following Rivanna UNIX tutorials:
<https://workshops.rc.virginia.edu/lesson/unix-tutorial/>
4. Read the Introduction to the UNIX Operating System section and then work through the seven tutorials, typing commands in your terminal (executing commands on Rivanna!) as you go through them.
5. Briefly summarize the commands/results of the Exercises (which go up to Exercise 5a) and questions asked in the tutorial. It turns out that the science.txt file on Rivanna is different than the one referenced in the UNIX Tutorial 2. Replace the search words “science” and “Science” by “and” and “And” when you get to that part of the tutorial. Also, briefly summarize any problems that you encountered. (25 pts)
6. Answer the following additional questions (60 pts):
 - a. What are the three parts that make up UNIX?
 - b. Briefly describe the functionality of these three parts.
 - c. How is the UNIX file-system/set of directories structured?
 - d. From the unixstuff/backups directory, list three different commands that return you to the home directly.
 - e. Does the output from commands “grep -ivc and science.txt” and “grep -ic and science.txt” add up to the output of “wc -l science.txt”? Why or why not?
 - f. In Tutorial 2 Summary, they mistakenly replaced a command that they introduced in the body of the tutorial (i.e., they introduced a command in the body of the tutorial that’s missing in the summary and included a command in the summary that’s missing in the body of tutorial). What command in the body of the tutorial was replaced? What command was added in the summary? Do they do the same thing? Is there any difference in what they do (if so, briefly explain)?
 - g. Briefly define “standard input”, “standard output” and “standard error”.
 - h. Use pipes and grep to count how many lines from the command “apropos copy” contain the word “copy”.
 - i. In your unixstuff directory, create a new file, “biglist2” by executing the command “paste list1 list2 > biglist2”. Use the “cat” command to view the contents of biglist2. What’s in the file? Now, execute the commands “cut -f1 biglist2” and “cut -f2 biglist2”. What output does each of these commands produce?

- j. Execute the following set of commands and briefly explain the output:
"cat list1 list1"; "cat list1 list1 | sort"; "cat list1 list1 | sort | uniq".
What is the output if you skip "sort" in the last command (i.e., "cat list1 list1 | uniq")?
 - k. Type "echo \$PATH". Briefly describe the output in terms of the functionality of this environment variable.
 - l. Assume you had the program "units" in your home directory
"~/units174/bin", and you executed the command "export PATH=\$PATH:\$HOME/units174/bin". What would this enable you to do?
7. Go to the following SLURM job manager site:
<https://www.rc.virginia.edu/userinfo/rivanna/slurm/> Read this page which will instruct you on how to submit/run programs (especially computationally intensive ones including sequence alignment) and answer the following questions (15 pts):
- a. What is the default starting directory of a job? Why is this important to know? In the example SLURM job command file to run a serial R batch job (near the bottom of the page; assume it's named myRprog.slurm), which line of code would this effect? What script/program (shown in the command file) should you have in the directory from which you executed the command "sbatch myRprog.slurm"?
 - b. Will running sbatch myRprog.slurm work? Hint: execute the command "module load R". What happens when you do this? Second hint: go to the R and RStudio on Rivanna page:
<https://www.rc.virginia.edu/userinfo/rivanna/software/r/> and look at the instructions "Loading the R module". Execute the command "module load gcc R". Then type "R". You can type "q()" to exit. What happened when you typed "R"?
 - c. Write a script that would submit a single-core job to the cluster with the following options: 1 task, wallclock time no longer than 1 hr, output results to "my_analysis.out", standard partition request, charge account bioc8145 and execute the (non-existent) R program myRprogram.R. Assume you named this file/script R_job.slurm. How would you submit it and run your job? Don't actually execute this command. We'll do that shortly.
8. Also, download R onto your machine/laptop:
<http://cran.r-project.org/bin/windows/base/> (Windows)
<http://cran.r-project.org/bin/macosx/> (Mac)
You may also want to download R Studio (a nice working environment):
<http://www.rstudio.com/products/rstudio/download/>