**Emmanuel Adeyemi**

**Part I:**

1. What is a Graphical User Interface (GUI) and how does that differ from the Command Line Interface (CLI)?

A GUI is a medium for interacting with a computer, usually through clicks. A GUI will typically contain dropdowns and other selectable menus. This differs from the CLI, which interacts with the computer via specific command inputs in the command line. In addition, the GUI is often a product of a programmer or creator’s reasoning in terms of available options, therefore, the GUI will often present less options than the CLI – which contains the full range of available options.

1. What does the shell do?

The shell is a program that runs other programs.

1. What is the command used for listing things in a directory?

ls

1. What is the command used for changing directories?

cd

1. What command would you use to get your current working directory?

pwd

1. How do you get the manual for these commands?

man [command]

1. What does the shell prompt look like?

 (I installed “oh my zsh”)

1. How would you list things in a directory in chronological order?

ls -t

1. Name two ways to get to the ‘home’ or ‘root’ directory.

cd ~ or cd

1. What is the difference between an absolute path and a relative path?

Absolute path is specified from the root up to the directory, folder, or file of interest. This path also never changes regardless of current directory. Relative paths constantly change, with respect to the current directory.

1. What are the two relative path directories we talked about and what do they mean?

“.” and “..”

**.**  means the current directory

.. means one directory up from the current directory

12. Why shouldn’t you put spaces in filenames?

Because shell interprets spaces as arguments provided to the file

13. Name a way to have multiple readable words in a filename without spaces?

Using camel cases or underscores

14. What is the program nano, what does it do? Do you use nano? If not what do you use?

Nano is a text editor embedded in the terminal. It can be used to edit several formats of already-existing files, and to create new files. I use nano.

15. \* is a form of a wildcard. What does it mean?

This is a regular expression which, depending on the preceding command, could include or ignore every occurrence of a specified pattern.

16. The following code would match what? ls \*.txt

This code will list all files that have a .txt extension in the directory

17. Why do we need to be careful with the mv command?

This command is similar to cut, it moves the specified file to a different directory and leaves no duplicate copy in the directory from which it was moved.

18 .Why do we need to be careful with the rm command?

rm commands deletes a file permanently

19. What is the difference between > and >>?

These both redirect the output of a command to an output file. However, the first one creates a file every time it executes a command while the second one first searches the directory for the output file and concatenates the output from the command, rather than overwriting the output file, if one already exists.

20. What does head do?

With no further arguments, this command prints the first five lines of the specified file.

21. What is the purpose of | (pipe)?

A pipe helps to run multiple commands in a single line of code. The output of the first code serves as input for the next code.

22. Interpret the following command:

cat huge\_file.fasta | uniq | head -n 5 >proteins.fasta

We’re printing the contents of huge\_file.fasta to screen, then the next command is parsing the output and getting rid of repeated lines (displaying only a single occurrence of such lines). Finally, the first five lines of output which contains only one occurrence of repeated line is stored in a file called proteins.fasta

23. What is a loop and when would you use it?

A loop is an iterative command that goes through a list of items and performs a particular task or a set of tasks. Loops are ideal for performing repetitive tasks on a huge dataset.

24. In a directory with the following files. What would the following loop do?

$ for filename in \*.txt; do

> cat ${filename} >>fruits.txt

> done

This loop will go through every file with a .txt extension in the directory and concatenate each file’s content to fruits.txt

25. What is a shell script?

A shell script is a file that contains a line of codes that can be run on the command line, by the shell interpreter, to execute a particular task. A shell script can contain python codes (.py extension) or linux code (.sh extension) so far as the first line in the script contains “#!/usr/python3” or “#!/usr/bash”

26. What are the benefits of writing and running a script over typing the code in?

Writing scripts allows the coder to document/make useful notes as they go along the blocks of code.

27. When looking at a script what does # me**an? Why would you use one?**

This means the following text is not a line of code and will be ignored by the interpreter. The hash symbol is used when making a note or a comment.

**What does grep stand for and what does it do?**

Grep stands for global regular expression print. It is a regular expression that searches and returns lines that contain the specified pattern in the provided file(s).

**What does find do?**

Finds the provided directory or file. For a directory or folder, all the contents get listed.

As part of this exercise please open the shell and practice moving around into different files. Best practice is to try a little bit every day. Try to challenge yourself by not using the mouse. For each of the questions below type your code and the result from the shell prompt.

1. Open the shell and change directories to the desktop and list the contents, paste your code here.

cd Desktop

ls

1. Diagram

   Description automatically generatedDraw the file structure from the root directory into one of the folders on the desktop. You can submit a pdf with a drawing. Select three locations and type the commands for changing directories to those locations use a combination of absolute and relative paths.
2. On your computer create a directory on your desktop. Create two files without opening them that have file extensions .txt. Then use wildcards to list the files in that directory. Paste you code here.

mkdir BIO792

touch file.txt file1.txt

ls \*.txt

1. Download the file [Hutia\_DNA.fasta](https://www.dropbox.com/s/07y610c8zt63nt5/Hutia_DNA.fasta?dl=0) from the Github repo. Make a new directory on the desktop. Move the file into that directory. List the contents of the directory. How many lines are in this file? Fasta files are a type of data file that holds DNA sequences. They are all formatted the same. The first line starts with a > and a name followed by a line break, and then DNA sequence followed by a line break. The next line starts with an > and another name followed by a line break and the DNA sequence. With that in mind, how could you use linux commands to determine how many sequences are in this file? How many sequences are in this file?

There are 2382004 lines in the file

grep “>” | wc -l

there are 1191002 sequences

1. Build a single line of code that would take the first 100 sequences of this file and put them in a new file called ‘MyOutputFile.fasta’. Use a combination of wc, sort and uniq in a single line of code to tell me how many unique lines are in this file. Paste your commands and the answer here.

Head -n 100 Hutia\_DNA.fasta >> MyOutoutFile.fasta

uniq MyOutputFile.fasta | sort | wc -l

100

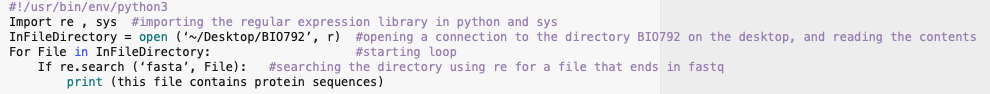
1. Write code that would create a loop to copy all the files in one directory ending in .txt to another directory. Paste the code here.

For file in Desktop/\*.txt

do cp $file Documents/BIO792

done

1. Write a bash script with in-line documentation (hint #) to show how to run a made up python program (script) called 'find\_taxa.py' on set of files ending in ‘.fasta’. Show in the script how to run the program on each of those files. What is the name of your script? Type the script here.

(This python script didn’t work when I tried to run it)

1. From the Hutia\_DNA.fasta file tell me how on how many lines do we find the pattern ‘GAGA’. What was the code used to find this?

304806

Grep ‘GAGA’ Hutia\_DNA.fasta | wc -l