Part I: Questions

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

A graphical user interface (GUI) allows users to interact with software or hardware through graphical elements or clickable icons. To utilize and navigate the software or devices, the user types instructions into a terminal when using a CLI, which is a console or text-based representation. A graphical user interface (GUI) allows users to interact with software or hardware through clickable icons.

What does the shell do?

The shell acts as an interface between the user and the kernel. The shell manages the interaction between you and the operating system by prompting you for input known as command and interpreting that input for the operating system, and then handling any resulting output from the operating system. Shells provide a way for you to communicate with the operating system

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

ls

What is the command used for changing directories?

cd

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

pwd

How do you get the manual for these commands?

man

What does the shell prompt look like?

$

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

ls -t

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

cd ~ or cd/

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

An absolute path specifies the complete path of a file from the root directory (/) while a relative path specifies the present working directory and never starts with a /.

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

. shows the current directory

.. moves up one directory

Why shouldn’t you put spaces in filenames?

CLI may read anything after a space as a different command

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

Use fullstop (.) or underscore (\_)

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

Opens a text editor. Yes, I use nano.

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

Matches any number of characters

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

List all .txt files

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

mv works silently and will remove the source file

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

rm works silently to delete files and you can’t recover them

What is the difference between > and >>?

> redirects output into a new file or overwrites an existing file while >> redirects output into an existing file by appending

What does head do?

Prints the first 10 lines of one or more files

What is the purpose of | (pipe)?

send the output of one command, program, or process to another command, program, or process for further processing.

Interpret the following command:

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

‘cat’ is used to read the contents of the file *huge\_file.fasta* sequentially and writes them into standard output. Duplicate lines are then removed from the file using the ‘uniq’ command and the first 5 lines of the output are displayed using the ‘head’ command. The output (the first 5 lines) is then finally redirected into a file called *proteins.fasta*.

What is a loop and when would you use it?

Loops are conditional statements used to perform repetitive tasks or problems programmatically.

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

$ for filename in \*.txt; do

> cat ${filename} >>fruits.txt

> done

For every filename with a .txt extension in the current directory, display the content of the file to the terminal and append the content of each file to a file called *fruits.txt*. The loop basically concatenates all txt files in the current directory into one file called *fruits.txt*.

What is a shell script?

A shell script is a txt file containing a sequence of commands for a unix-based operating system. It is a simple program to be run by shell or a CLI.

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

Allows the automation of performing batch jobs

When looking at a script what does # mean? Why would you use one?

# tells the shell to ignore. To add comments to a script

What does grep stand for and what does it do?

Global regular expression print. Lets you find a word by searching through all texts in a file or directory

What does find do?

Search for files within a specific directory

Part II - Writing Code

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.

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

amank@DESKTOP-0EFLLRQ MINGW64 ~

$ cd Desktop/

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop

$ ls

Draw 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.

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 your code here.

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop/BIOL792.2

$ touch file1.txt file2.txt

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop/BIOL792.2

$ ls \*.txt

file1.txt file2.txt

Download the file Hutia\_DNA.fasta 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?

$ cp ~/Downloads/Hutia\_DNA.fasta ~Desktop\FASTA\_files

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop

$ cp ../Downloads/Hutia\_DNA.fasta ../Desktop/FASTA\_files

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop

$ cd FASTA\_files/

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop/FASTA\_files

$ ls

Hutia\_DNA.fasta

$ grep -c ">" Hutia\_DNA.fasta

1191002

There are 1191002 sequences

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.

amank@DESKTOP-0EFLLRQ MINGW64 ~/Desktop/FASTA\_files

$ head -200 Hutia\_DNA.fasta | sort | uniq | wc -l > MyOutputFile.fasta && wc -l < MyOutputFile.fasta

1

1 unique line

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 filename in \*.txt

do

cp${filename} ~Desktop/FASTA\_files

done

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.

runpy.sh

#!/bin/bash #Write a shebang to let the CLI know that it should be run with bash shell

#Define the set of all the .fasta files in the current directory

FASTAfiles=(\*.fasta)

#use a for loop through each file in the set of .fasta files

for file in "${FASTAfiles[@]}"

do

#Run the 'find\_taxa.py' program on each file

python find\_taxa.py $file

done #end of the loop

#Print a message to confirm the program has been completed

echo "Program successful."

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 lines

$ grep -c 'GAGA' Hutia\_DNA.fasta