**Linux Exercises Part 2 Name: Ryan Kyger**

This week we went over:

[Working with Files and Directories](http://swcarpentry.github.io/shell-novice/03-create/index.html) and [Pipes and Filters](http://swcarpentry.github.io/shell-novice/04-pipefilter/index.html)

**Part I: Short Answer**

1. *Why are we learning the shell? What are the benefits of the shell over the GUI?*

You can perform processes on hundreds of files. Shell is often faster than a GUI.

1. *Why is it not recommended to put spaces in filenames?*

The shell will expect a command after a space.

1. *Name a way to have multiple readable words in a filename but no spaces?*

You can use dashes: “file-name.txt”

1. *What is the program nano, what does it do?*

Nano is a command line text editor.

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

It means any text/string.

1. *ls \*.txt would match what?*

This would match all files that end in “.txt”

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

“mv” can move or rename files

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

“rm” permanently deletes files.

1. *What is the difference between > and >>?*

Output to a file: “>”

Append to a file: “>>”

1. *What does head do?*

Lists the first 10 lines of a file.

1. *What is the purpose of | (pipe)?*

The purpose is to send the output of one command to another command.

1. *Interpret the following command:*

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

Read the fasta file, find all unique lines, get the first 5 lines, export output to new fasta file.

**Part II: Practice**

1. *As part of this exercise please open the shell and practice moving around into different files, copying files and listing files with wildcards. Best practice is to try a little bit every day. Again, try to challenge yourself by not using the mouse.*
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. Type or your commands here.*

mkdir thesis

touch draft.txt

touch abstract.txt

ls \*.txt

1. *Download the file Frog\_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 DNA data file. 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 many sequences are in this file?***

Number of lines: 2382004 command used: less Hutia\_DNA.fasta | wc -l

Number of sequences: 1191002 command used: grep ">" Hutia\_DNA.fasta | wc -l

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 200 Hutia\_DNA.fasta > MyOutputFile.fasta

less MyOutputFile.fasta | sort | uniq | wc -l number of lines: 199