CSCI 4180: Big Data Analytics
CSCI/DASC 6010: Big Data Analytics and Management
Spring 2023
Homework 1 – Linux power tools
Due Sunday, January 22, at 11:59 PM

Preliminary

Git Magic is a website that provide tips/tricks on how to use `git` version control. In this assignment we'll use the files from this website to get more familiar with Linux power tools. You can use the output from the `script` used in class, 02. Linux power tools.txt, to look at examples of commands used, along with their output.

1 Exercises (100 pts)

1. (5 pts) List the Linux command (`wget` or `curl`), complete with all options and arguments, used to download all the files from this website.

2. (15 pts) List the one line command used to count how many files were downloaded, along with its output (the number of files). Hint: you will have to use multiple commands, and pass the output from one command as input to the next command.

3. (30 pts) List the one line command used to count how many files of each type (extension) were downloaded, along with its output (count and file extension pairs).

4. (50 pts) List the one line command used to find the file with most numbers of lines, along with its output (lines count, and file name and path).

2 Extra credit (10 pts)

For extra credit you will need to use R Studio and the Rmd template file provided. You need to look through the output from the `script` used in class, 02. Linux power tools.txt, to find the examples.

List the commands shown in class, complete with all options and arguments, sample output (only partial if the output is too long), and a brief description of that command.

- If your last name starts with A–E, create slides only for the following commands: `cat`, `cd`, `cut`, `df`, `du`, and `exit`.

- If your last name starts with F–J, create slides only for the following commands: `file`, `find`, `for`, `grep`, `gunzip`, and `head`.

- If your last name starts with K–P, create slides only for the following commands: `ls`, `man`, `mv`, `open`, `paste`, and `pwd`.

- If your last name starts with R–Z, create slides only for the following commands: `rm`, `sed`, `tail`, `type`, `vi`, and `wget`.

Submit

Upload to Canvas the following:

- **Required:** Your answers to questions 1 through 4 (one PDF file).

- **Optional:** Your extra credit files (one Rmd and one PDF file).

If you're only providing your answers to questions 1 through 4 then upload the PDF file. Otherwise, compress all files (two PDF and one Rmd) into one ZIP file and upload this file.