# Assignment 1: Command Line Interface (30 pt)

## GitHub (10 pt)

Create a repository, and list some information about yourself.

1. Go to GitHub and create a new repository called `assignment-1` without a README. (5 pt)
2. Clone the directory into your main GitHub directory.
3. Within that directory, create a new file empty file named `README.md`.
4. Open `README.md`.
5. Add your name and hit enter twice. Then write a short description about yourself, including things like your major, potential careers you are interested in, why you took this course, and your hobbies. (2 pt)
6. Stage and commit your changes before making more changes. (1 pt)
7. Copy the files for the assignment 1 bash portion from LATTE into this directory. (1 pt)
8. Stage and commit your changes before making more changes. (1 pt)

## Bash (20 pt)

The bash questions below will either give you a task or ask a question about bash and CLI. **Work in the local `assignment-1` repository you created.**

If the question gives you a task, write the bash command used to do this operation in `answers.txt`, a file you will create as part of this exercise.

These problems use [publicly available air quality data from the EPA](). We are using files with annual summaries of air quality measurements by metro area (CBSA) or by county.

1. Create a directory in your local `assignment-1` directory called `answers`. (1 pt)
2. Create a file called `answers.txt` in the `answers` directory. This is where you will be recording your answers to the rest of the questions for this assignment. (1 pt)
3. Edit the `answers.txt` file to record the commands used for questions 1 and 2. (1 pt)
4. List the absolute path to your `answers.txt` file. (1 pt)
5. What flag lets us use `ls` to list all files in a directory, including hidden files? (1 pt)
6. How can you use `ls` to list files outside of your current directory? (1 pt)
7. Why is `rm` a dangerous command? Describe a scenario where it could lead to unintended negative consequences. (2 pt)
8. Create a subdirectory in `answers` called `results`. (1 pt)
9. Create a script called `my_script.sh` that contains the following code: (2 pt)

```bash
#!/usr/bin/env bash

tsv="${1/.csv/.tsv}"

sed "s/,/\t/g" $1 > $tsv

awk -F '\t' '{print $1}' $tsv
```

(This code first takes in a name of csv. It then creates a variable named `tsv` that makes a filename replacing `.csv` with `.tsv`. It converts each csv into a tsv, a tab-deliminated data file. Finally, it prints the second column of each tsv.)

10. Change permissions on `my_script.sh` to be able to execute it. (1 pt)
11. Navigate to `data` directory in the `assignment1` directory from the `answers` directory in a single command. (1 pt)
12. Make a one-line pipe that returns the number of words in second largest file in the directory. (1 pt)
13. Loop through the .csv files in the data directory and run `my_script` on each of them. To run this script, you can run `/path/to/my_script.sh file.csv`. (2 pt)
14. .zip files use data compression to reduce file size. We can unzip files to get back the original versions of them with `unzip filename.zip`. Unzip all files in `data` with the `.zip` extension and store them in a directory called `unzipped-files`. (2 pt)
15. Remove all .zip files. (1 pt)
16. Go to `unzipped-files` and copy all files from the 2000s to the results subdirectory you made earlier. (1 pt)

Stage, commit, and push your changes. **Make sure to put a link to repository into the assignment submission on LATTE, and make sure the repo has all the changes you've made.**

Source of data:

https://aqs.epa.gov/aqsweb/airdata/download_files.html