

Github introduction

1. Create a Github.com account
2. Add your account to [google sheet](#)

UNIX practice

1. Login to HPCC
2. Create a folder

Data processing

Go into your bigdata folder. If you have not used the cluster before then you will be in the **gen220** project. Or you may have your own lab bigdata folder.

```
cd ~/bigdata # this should work but if it doesn't
cd /bigdata/gen220/$USER # will go into your bigdata folder for the class
# if the above doesn't work you are likely already in a lab group on HPCC
cd /bigdata/$GROUP/$USER # should work since $USER is your login and $GROUP is your primary
```

Now you want to make a folder for your work for this class

```
# you can make a folder for GEN220
mkdir gen220
go into that folder
cd gen220
# now use git to checkout the class data folder
git clone https://github.com/biodataprogram/GEN220_data.git
# now go into this folder
cd GEN220_data
```

Look around in the folder. Go into the **tabular** folder where I've stored some tab or comma delimited data. You will write a set of bash commands to answer these questions. You can solve everything with the following unix commands: **cut, sort, uniq, tail or head, wc, sed** ## Airports 1. how many lines are in the airport codes file? 1. how many airports are listed as closed? 1. how many airports are in Antarctica (iso_country code AQ)? How many are heliports? 1. what is the name of the airport located in highest elevation, and what is its elevation? 1. what is name of the airport located in the second lowest elevation? 1. what country has the most airports? List the top 5 countries with the most airports