

HOMWORK 7

HUGEN 2070 · BIOINFORMATICS FOR HUMAN GENETICS · FALL 2019

For each question, please show both the answer to the question and the Unix command or pipeline that you used to get the answer. For this assignment do not use `plink` itself unless instructed to. Also, do not use `nano` to manually edit the files. Do not use R.

You may be asked to use options that were not detailed in the PowerPoint slide decks. Please consult the commands' man pages or use Google to try to figure out the answer before asking for help.

You do not need to submit the files created as part of the assignment. Merely answer the questions about them and submit the code used to create them or arrive at the answer to the question. Each question has 'code', 'answer', or 'code and answer' after them to indicate what I'm looking for.

Your submission should be a plain text file (`hw7_<lastname>.txt`) in this format:

#Question 1(a)

#Answer

#Question 1(b)

Code

.

.

.

#Question 3

Code

#Answer

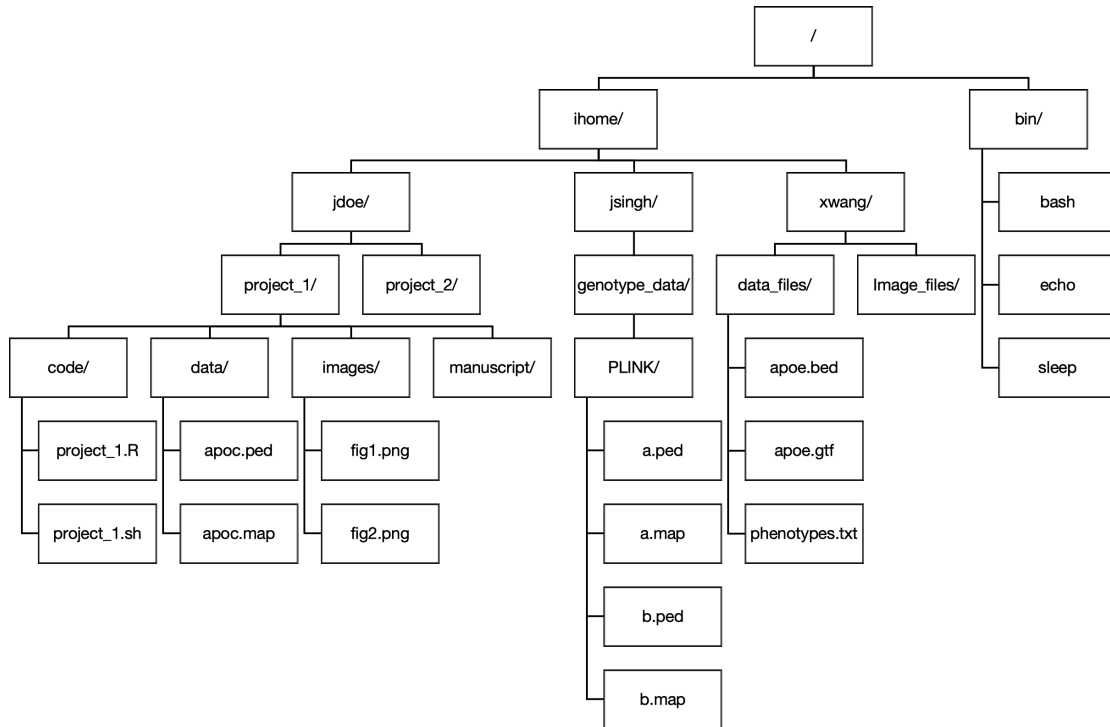
#Question 4

.

.

.

1. Examine this file hierarchy. The home directory is /ihome/jdoe/.



- You just logged in. What would running `pwd` return as your working directory?
(answer)
- Change the working directory to `data/`.
(code)
- Change the working directory to `images/`.
(code)
- Go to the parent directory of `images/`.
(code)
- Make a new directory here called `new_images/`.
(code)
- Copy `fig1.png` to the new directory.
(code)
- Remove `fig1.png` from `images/`.
(code)
- Move `fig2.png` to `manuscript/`.
(code)
- Return to the home directory.
(code)

2. You are working with `cat` and you want to put the output and any error messages into a single file called `cat_output_and_errors.txt`. What code would you use to do this? (The files you are cat'ing are called `a.txt`, `b.txt` and `c.txt`.)
(code)
3. You started an `awk` command in the background and realize that you made a mistake with it and want to end the process. `ps` shows the output below. How would you end the process?
(code)

```
PID TTY          TIME CMD
2057 pts/1      00:00:00 bash
3245 pts/1      00:00:30 awk
3246 pts/1      00:00:00 ps
```

The rest of the questions use plink files. You cannot use plink itself. Copy these files to your home directory:

```
cp /bgfs/rminster/hugen2070-2019f/data/PLINK/hapmap1* ~
```

4. How many participants are there in `hapmap1.ped`?
(code and answer)
5. What is the size of `hapmap1.ped` in megabytes?
(code and answer)
6. How many SNPs are there in `hapmap1.map`?
(code and answer)
7. What is the size of `hapmap1.map` in megabytes?
(code and answer)
8. What is the delimiter being used in `hapmap1.ped`?
(code and answer)
9. Create a file called `hapmap1.fam` that has only the first six columns of `hapmap1.ped`.
(code)
10. How many SNPs in `hapmap1.map` are on chromosome 15?
(code and answer)
11. Which line of `hapmap1.fam` contains data about individual `hcb215 1`?
(code and answer)
12. Extract the line that contains data about individual `hcb215 1` from `hapmap1.fam`.
(code)