assignment 2 clarification

Dhara Shah

February 5, 2020

What to submit and how it would be graded

I am attaching the template for the answers on iCollege. It is essentially a folder with six bash files, and the file 4HKD.pdb. Your job is to fill in the answer of your questions in each of these files, pertaining to the appropriate question number, where it says "fill in your answer here".
Once you fill in the answers, you will be uploading **this entire folder** on iCollege.

To grade your assignment, I will

- Download the entire folder
- Replace 4HKD.pdb with a fresh instance and different file than you got each file with

```
bash q1.sh
bash q2.sh
.........
bash q6.sh
```

After each run, I will plugin some test cases for counting line numbers, or min-max or average values etc, and will be matching those to expected answers.

I will NOT be reading your code, only the output of your test cases. So, there is no partial credit for "writing the code".

Grade distribution

- Total: 10 points
- Q1 and Q6: 1 point each
- Q2 to Q5: 2 points each

Possible spelling mistakes

This is just a hint page, so, for choosing strings to grep, please verify the spellings from Assignment 2. Also, please check the spellings and syntax of your commands.

Some clarifications and hints

Question 1:

You want all the lines form the 4HKD.pdb file that do not contain "ATOM", "CONECT", "HETATM","TER", and "END". We will be calling each line of 4HKD.pdb file a **record** for this assignment.
Hint: What do you get with the following command?

```
cat 4HKD.pdb | grep "ATOM" | grep "END"
cat 4HKD.pdb | grep -v "ATOM"
```

Question 2:

First, find the lines that have "HETATM" and "MSE" both. Replace "HETATM" with "ATOM " (make sure to put two spaces after ATOM!) and "MSE" with "MET" only in these lines, in a way that the file 4HKD.pdb reflects the changes afterward.
Hint:

- What does this command do?

  ```
  echo "Hi" | sed 's/Hi/Dhara'
  ....(some bash commands outputting some lines of 4HKD.pdb)
  ```

Question 3:

First, filter all the lines starting with "ATOM " from the 4HKD.pdb. Then, take the columns 7, 8, 9, and output their maximum and minimum values. When I run your answer in file q3.sh, it should give me the output in the following format:
Column 7 min = (your answer)
Column 7 max = (your answer)
.....
Column 9 max = (your answer)
Hint:

- How many "inbuilt" variables awk does have? Which of these would indicate the line number?
- Sometimes it is great to use END in awk without BEGIN. What does this END do?
- Note that awk is a functional programming language, and for doing column operations, so if you wish, you can get away with an if-else statement :)

- if you don't know what I am talking about, as far as if-else statement is concern, figure out how to loop through each line of the given column, and find min. In this case, you may want to use a for loop.
- There are two ways to do this: either write an awk function that takes the column number as an input and calculates min/max of that column, or if you don't want to write such a function, then write 6 different awk scripts, one for each line.

What gets you a zero: hard-coding min and max values in your code

Question 4:

Hint: If you figured previous question, this is the same methodology, but different function. Upon running the code, you it should print
Column 7 average = (your answer)
.....
Column 9 average = (your answer)

Question 5:

Hint: you do the same thing that you did in Question 2.

Question 6:

Hint: Awk and bash both have sort functions :)

Asking questions to me regarding this assignment

This week, there are no office hours in person, but I will be available through email dshah8@student.gsu.edu. Please utilize Dr Harrison's office hours for in-person questions.

Please be mindful that I cannot give you any additional hints if they are not given to the entire class. Also, I cannot verify your answers before the submissions, or see what you are doing wrong, while you are live coding. However, I can help you with the errors.