

assignment 2 clarification

Dhara Shah

February 12, 2020

Possible correct answers

Question 1:

Find the records that do not have "ATOM", "CONECT", "HETATM", "TER" and "END"

Answer 1

```
cat 4HKD.pdb | grep -v "ATOM\|CONECT\|HETATM\|TER\|END"
```

Question 2:

Get the records that have "HETATM" and "MSE", replace both the tokens with "ATOM " and "MET" respectively. Save the result as "file2".

Answer 2

```
cat 4HKD.pdb | grep "HETATM" | grep "MSE" |  
sed 's/HETATM/ATOM_/ ' | sed 's/MSE/MET/ ' > file2
```

Question 3:

The original questions was "Get all the records from 4HKD.pdb starting with ATOM, as well as the HETATM records that you replaced in Q2. Get the max and min values of columns 7,8,9". However, we simplified it to : "Get all the records from "file2" starting with "ATOM ", and find max and min values of columns 7,8 and 9."

Answer 3

Case 1: Desired answer without modification

```
awk '{  
if ($0 ~ ^HETATM && $0 ~ MSE) print ($0);  
if ($0 ~ ^ATOM_) print ($0);  
}' 4HKD.pdb > file2.2  
cat file2.2 | sort -nk 7 | head -n1 | awk '{print "Column_7_min_=", $7} '  
cat file2.2 | sort -nrk 7 | head -n1 | awk '{print "Column_7_max_=", $7} '  
cat file2.2 | sort -nk 8 | head -n1 | awk '{print "Column_8_min_=", $8} '  
cat file2.2 | sort -nrk 8 | head -n1 | awk '{print "Column_8_max_=", $8} '  
cat file2.2 | sort -nk 9 | head -n1 | awk '{print "Column_9_min_=", $9} '  
cat file2.2 | sort -nrk 9 | head -n1 | awk '{print "Column_9_max_=", $9} '
```

Case 2: Desired answer with simplification: You did not write the awk block and applied it on file2 from previous output

Case 3: I also allowed if you did not write the awk block and applied it on 4HKD.pdb from previous output

Question 4:

Original question: "Get all the records from 4HKD.pdb starting with ATOM, as well as the modifications from Q2, and find the average value of columns 7,8,9." However, we also allowed the simplification "Get all the records from file2 and get the average of columns 7,8,9"

Answer 4 **Case 1** Desired answer without modification:

Take file2.2 from Q3. and

```
cat file2 | awk '{sum+=_$_7}_END{print "Column_7_average=",sum/NR}'
cat file2 | awk '{sum+=_$_8}_END{print "Column_8_average=",sum/NR}'
cat file2 | awk '{sum+=_$_9}_END{print "Column_9_average=",sum/NR}'
```

Case 2: Desired answer with simplification: block above applied on file2

Case 3: I also allowed if you applied above block on 4HKD.pdb starting with

```
cat 4HKD.pdb | grep "^ATOM "
```

Question 5:

Original question: Get all the "HOH" tokens to change to "WAT" in-place.

Simplified question: Save the changes in "file3".

Answer 5

Since there are no "HOH" lines either in file2 or file2.2, obviously this operation can done only on 4HKD.pdb.

Case 1: Desired answer without modification

```
sed -i 's\HOH\WAT\' 4HKD.pdb
```

Case 2: Desired answer with simplification

```
sed 's\HOH\WAT\' 4HKD.pdb > file3
```

Question 6:

Original question was, sort all the records that start with "ATOM" with respect to its 11th position, after replacement from Q2 is done in 4HKD.pdb. Modified question is, sort all the records that start with "ATOM" with respect to its 11th position in file3.

Answer 6

Case 1: Desired answer without modification

```
cat file2.2 | grep "^ATOM" | sort -nk 11
```

Case 1: Desired answer with simplification

```
cat file3 instead from above case
```

Rubric

Question 1:

Points: 1.5

Penalties

- **1.1** Total number of lines is not 446: -0.5 pt
- **1.2** First line is not "HEADER TRANSFERASE ..": -0.5 pt
- **1.3** Last line is not "SCALE3 0.000000 0.000000 0.004781 0.00000": -0.5 pt
- **1.4** Output is saved in a file, but follows everything else without penalty: -0.5 pt
- **1.5** Hard coded any values: -1.5 (zero for this question)
- **1.6** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 to -1.5 pts

Question 2:

Points: 1

Penalties

- **2.1** Total number of lines not equal to 32: -0.5 pts
- **2.2** The first line doesn't match at all indents of the following: (you forgot two spaces after ATOM) :-0.5 pt
"ATOM 129 N MET A 450 .."
- **2.3** Hard coded any values: -1 pts (zero for this question)
- **2.4** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 to -1
- **2.5** The filename is not "file2":-0.5

Question 3:

Points: 3

Penalties

- **3.1** The output is not either the entire column of Case1/3, or the entire column of Case2 as below: (you have a few answer form either of the columns showing up, or completely different answers) -3 pts (you get 0 for this answer)
- **3.2** Output is generally consistent with one or other column, with some mistakes: -0.5 pt for each wrong answer

- **3.3** Hard coded any values: -3 pts (zero for this question)
- **3.4** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 to -3

	Case 1, 3	Case 2
Column 7 min =	-8.866	-10.909
Column 7 max =	1.354	11.011
Column 8 min =	16.830	9.343
Column 8 max =	47.100	52.712
Column 9 min =	78.833	75.346
Column 9 max =	122.510	132.824

Question 4:

Points: 3

Penalties

- **4.1** Output is neither completely case 1 or 2 or 3: -3 pts (you get 0 for this question)
- **4.2** Output is generally consistent with one of the cases: -1 pt per wrong answer
- **4.3** Hard coded any values: -3 pts (zero for this question)
- **4.4** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 to -3 pts

	Case 2	Case 1	Case 3
Column 7 average =	-3.69047	-2.37467	-2.26269
Column 8 average =	29.7547	29.9458	29.962
Column 9 average =	104.936	107.511	107.73

Question 5:

Points: 0.5

Penalties

- **5.1** If there are no lines in file3: -0.5 pts (you get 0 for this question)
- **5.2** If any of the lines in file3 contains "HOH": -0.5 pt (you get 0 for this question)
- **5.3** Hard coded any values: -0.5 pts (zero for this question)
- **5.4** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 pts
- **5.5** The filename is not "file3" or output not saved: -0.5 pts (zero for this question)

Question 6:

Points: 1 Turns out all the values in field 11 are 1 :) so we compare the lines!

Penalties

- **6.1** If the first line of the output is not "ATOM 74 C..." :-0.5 pts
- **6.2** If the last line of the file is not "ATOM 4 O ...":-0.5 pts
- **6.3** Hard coded any values: -1 pts (zero for this question)
- **6.4** Passes the test cases, but something wrong with the code: Grader's discretion - -0.5 to -1 pts
- **6.5** Wrong answer -1 pt (0 pts for this question)