STAT 29000 Project 8

Topics: Loops in Bash Scripts

Motivation: Now that we know how to create bash scripts, the natural thing to learn next is how to write loops inside of bash scripts. That way, we can input (say) one parameter to a script, and then run a sequence of bash commands on many files, with just one command in the UNIX terminal. This is a very handy skill to know!

Context: There are many ways to write bash scripts. Although we do not cover every single type of loop, having an introduction to this concept will enable us to write other kinds of loops, whenever we have such a need.

Scope: Bash scripts contribute to systematic, reuseable, reproducible analysis. They allow us to do a lot of analysis with only a little bit of code. This ability becomes even stronger, once we get familiar with loops in bash scripts.

Question 1: Analyzing flights across all years, for any airport

Before you start question 1, please consider the examples given here:

`/class/datamine/data/examples/stat29000project8examples.txt`

Use this template to submit your solutions:

`/class/datamine/data/examples/stat29000project8tempate.txt`

1a. Write a bash script that takes 1 parameter as input (such as `IND` or `ORD`) and outputs the number of airplane flights that originate from that airport each year. You can either use your code from Project 7, Question 2, as part of your solution, or you can make an all new bash script. Each line of your output should contain the year and the number of flights that originate from that airport, like this.

If we gave the script the filename `myscript.sh` and put it into our home directory, and we type, for instance, `~/myscript.sh IND` it will output the following:

```
1987,8817

1988,37399

1989,40567
```

. . . etc. . .

Question 2: Analyzing number of donations, per year, from any State

2a. Write a bash script that takes 1 parameter as input (such as `IN` or `IL`) and does the following: For each even-numbered year from 1980 to 2020, print a list of the number of donations from that state, in each of these years. Each line of your output should contain the year and the number of donations from that state in that year, like this.

If we gave the script the filename `myscript.sh` and put it into our home directory, and we type, for instance, `~/myscript.sh IN` it will output the following:

```
1980,4606

1982,2274

1984,2161
```

. . . etc. . .

Question 3: Analyzing the average of the total cost of yellow taxi cab rides, within each month

3a. Write a bash script that does the following: For each month from January 2017 through December 2018, print the average of the total cost of a yellow taxi cab ride. Each line of your output should contain the year, the month, and the average of the total cost, like this:

```
2017,01,15.5267
2017,02,15.6324
2017,03,16.2237
```

. . . etc. . .

Project Submission:

Submit your solutions for the project at this URL: https://classroom.github.com/a/py2fJYnX using the instructions found in the GitHub Classroom instructions folder on Blackboard.