Chemistry 438/550: Computational Chemistry, Spring 2020

Homework #2, due 12 Feb.

1. Log into amarel, and execute the following:

   ```
   git clone https://github.com/dacase/rdbstat.git
   ```

   We will discuss in class what this does, but you should see a new folder called *rdbstat* in your home directory

2. Now, build the programs you just downloaded:

   ```
   cd rdbstat
   cat README.md            # just to get an overview of what is here
   cd src && make install   # builds the program
   ```

3. add the new programs to your PATH variable:

   ```
   export PATH="~/rdbstat/bin:$PATH"
   ```

   You can add the above command to the `.bash_profile` file in your home directory (note the "." at the beginning of the filename). Then your PATH will modified every time you log into amarel.

4. Back in your home directory, use a text editor to create an RDB table. It should have at least 10 records and 5 fields (columns). Use a mixture of text and numeric fields, but have at least two numeric fields. The actual data can be made up, or could a piece of some data you have for a project you have worked on. There should be one column (an "index" column) where all the entries are unique, that is, there are no duplicate entries in that column. For this exercise, do not have any "blank" entries: every element in the table should have a valid value. Name your file `lastname.rdb`, where *lastname* is your last name. Check the validity of your file by running this command:

   ```
   valid -w < lastname.rdb
   ```

   Work on the file until you get an "RDBtable OK" response.

5. Choose two of your your numeric columns and see how they are correlated by doing this:

   ```
   column < lastname.rdb  col1  col2  | headchg -del | pair
   ```

   In the command above, "col1" and col2" should be the names of the two numerical columns you have chosen for your example. Look at the output, and try to understand as much as you can (it's pretty straightforward). If you are not familiar with a Pearson correlation coefficient, visit the Wikipedia page on this (https://en.wikipedia.org/wiki/Pearson_correlation_coefficient). Re-run the command, redirecting *stdout* to a file called `lastname.pair`.

6. Modify the command in section 5, piping the output of the pair program into an AWK command that will print just the minimum value of Column 1. *Harder*: have the AWK script instead output just the value of correlation coefficient. (Note: there is nothing to hand in for this part, but you should do it anyway.)

7. "Hand in" your work in this way:

   ```
   cp lastname.rdb lastname.pair /home/dacase/ccb550/homework2
   ```