CS 4120/5120 project. Kresman

This activity takes **time**! Consider spending time before, during, and after the break, thanks☺

Implement a program (C, C++, Java, Python, C#, MATLAB, R, etc.) to experimentally evaluate sorting performance across a suite of algorithms on synthetic (randomly generated) datasets and compare performance against theoretical worst-case.

* Algorithms: **I**nsertion sort; **~~H~~**~~eap sort~~; and **Q**uick sort
* Data set size: 5,000; 25,000; and 50,000.

Program command line option (one of the following):

* ‘D’ just generates the datasets (10 sets for **each** size noted above) and dumps them to a file(s).
* ‘S’ sorts the dataset using the algorithms and produces comma separated output (\*). For this option, the (input) file names may come from command line or hard-coded.

Note that neither option above has any terminal input!

Use Excel (or if you choose to, you can do within the same program) to post process the comma separated output file - generate graphs to visually depict algorithm behavior.

Finally, explain the program organization and interpret the results in a 2-page write-up (word doc).

(\*) Format of comma separated output shown below. For theoretical numbers below, compute worst case numbers.

{I H Q},{data set # 1 – 10}, {size}, {# of comparisons}, {alg cpu time}, {# theoretical comparisons}

For example, a typical line: H,3,50000,12345, 0.912359, youShouldKnowThisNumber

The program should be documented, readable, and organized with clearly identified modules for the various items.

The write-up must discuss theory and experimental results and provide informal justification for match/anomaly. As well, the write-up explains how to run the program.

Canvas Submission requirements (5 files, all under **ONE** zip): (1) Documented program listing with run snapshots; (2) Data set 1; (3) comma separated output file; (4) one pdf file with all graphs; and (5) write-up.

**Notes:**

* **You can work in groups (CS 5120** group size < 2; **CS 4120** group size < 4). Group has either 4120 or 5120

student, not a mix of the two. Sorry, no brownie points if you decide to work on your own☹

* ONLY **ONE** submission/group. Comments @ top includes student names, and description of who did what.
* **CS 4620 - drop** either Quick or Heap sort (drop ONE of these 2) in the implementation!

Grading rubric:

(4) Data generation

(8) run/implementation/documentation of each algorithm with count of comparisons (4 points for each alg)

(3) comma separated output

(3) CPU time computation

(6) graphs

(12) write-up.

Questions:

Are their 10 data sets for each size of 5,25,50 thousand?

10 different files?

Ask to explain what the submission needs to be. Do the datasets need to be all in one file or can their be 30 different files with the datasets.

For the build what is needed:

Main needs to accept argV from file call

Main that has two options from argv. D and S.

D will create the datasets

S will sort the data sets and report them to the terminal output

3 files for dataset