CSCI 4502/5502: Data Mining
Homework 2

Due at **12:30pm on Thursday, Feb 5, 2015**. Submit two files electronically at moodle: (1) "**Last Name_FirstName_Homework2.py**" containing your python source code for Question 1 and (2) "**LastName_FirstName_Homework2.pdf**" containing your plots for Question 2. Make sure to include your name, student id, and the Honor Code Pledge (`http://honorcode.colorado.edu/student-information/honor-code-pledge`).

1. Stock quotes analysis. Go to `http://www.nasdaq.com/quotes/`, enter a symbol in step 1 (e.g., "HD") and click "Historical Quotes" in step 2. On the next page, you can select different time frame (e.g., "3 Years"), view the data, and download the data as a .csv file using the link at the bottom of the page. Write a python program that takes two command line variables $< filename1 >$ and $< filename2 >$ (which have the same format as the .csv file), and does the following analysis:

   (a) Using the "Volume" attribute data in $< filename1 >$, compute the corresponding normalized values using min-max normalization with new range $[0, 1.0]$. Output two values per line: the original volume and the normalized volume, separated by the $< tab >$ key.

   (b) Using the "Open" attribute data in $< filename2 >$, compute the corresponding normalized values using z-score normalization. Output two values per line: the original open value and the normalized open value, separated by the $< tab >$ key.

   (c) Compute the following correlation coefficients and output the three values in a single line, separated by the $< tab >$ key:

      i. between the "High" attribute in $< filename1 >$ and the "Low" attribute in $< filename1 >$.

      ii. between the "High" attribute in $< filename2 >$ and the "Low" attribute in $< filename2 >$.

      iii. between the "Close/Last" attribute in $< filename1 >$ and the "Close/Last" attribute in $< filename2 >$.

2. Using the website above, obtain the 3-year data for "HD" and plot the following:

   (a) A single plot showing the temporal change of the "High" and "Low" attributes.

   (b) A boxplot for the "Open" and "Close/Last" attributes.

   (c) The 10-bin equal-width histogram for the "Volume" attribute.

   (d) Any plot that interests you (using this HD dataset).