CIS 2033 Fall 2014: Computational Prob and Stat
Lab 5 Assignment (Due on Dec.5th before class)

The Matlab sample data sets "cities" contains ratings for 9 categories, which are rated by citizens in 329 cities. You can load the data by using 'load cities' in Matlab. Analysing the data using numeric and graphic summaries is very helpful for discovering the potential relations between different categories and for better understanding the city.

Question Choose 3 categories of your interest, e.g. education, crimes. Plot the histograms for ratings of each category with bin size 10, 20, 50, 100 respectively. Summarize your observation in one paragraph.

Question Choose 3 categories of your interest. Calculate the following 7 sample statistics: sample mean, sample median, sample standard deviation, sample MAD, lower quantile, upper quantile, IOR. Put the values in a 3 by 7 table, with categories as the rows and statistics as the columns. Summarize your observation in one paragraph.

Question Choose 3 pairs of categories of your interest, e.g. education vs crime, recreation vs arts, economics vs education. Plot the scatter plots for the ratings of each pair respectively. Summarize your observation in one paragraph.

Question Calculate the pairwise correlation among the ratings for the 9 categories. You should get a 9 by 9 matrix C, with element $C(i, j)$ denoting the correlation between category i and category j. Find out which two categories are most positively correlated and which two are most negatively correlated. Summarize your observation in one paragraph. (Hint: corrcoef function can take the matrix as an input.)