HW 1
SDSC 5002

Instructions for HW1. When the path to an answer involves coding, please show the Python codes and proper output. When plot a graph, remember to **add title, labels for x and y axes. Also, add legend,** if necessary.

1. Suppose you have a set of data $x_1, x_2, \ldots\ldots, x_n$. Show that the sum of deviations (not squared) is just 0.

2. For the following set of data

   2, 3, 7, 7, 10, 9, 7, 10, 6, 10, 3, 10, 20, 3, 10, 8, 5, 1, 5

   provide:

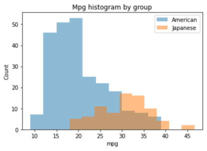
   a) Tukey's 5-number summaries
   b) Interquartile range
   c) A box plot (mark the important values on the plot.)

3. **Python problem:** Looking at the dataset Housing.csv. The variables are
   - medv: median home price in different neighborhoods
   - crim: per capita crime rate
   - rm: average number of rooms per dwelling
   - zn: proportion of large lots (zoned for > 25, 000 feet)
   - river: whether a home is near a river (0: No, 1: yes)
   - ptratio: pupil-teacher ratio by town

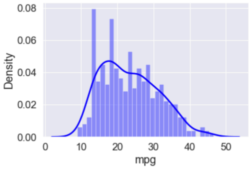
   Questions:

   a) Show first 8 rows of the dataset.
   b) Given the median and mean of **medv**.
   c) Calculate 1st and 3rd quantile of **crim**.
   d) Plot histogram of **crime**.
   e) On the same plot, draw histograms of **medv** near a river and not near a river, respectively.

Recall the example in the tutorial:



f) Provide correlation matrix and the corresponding heatmap for the dataset.
g) Fit a density curve using 10 bins for **medv**.
Recall the example in the tutorial:


Density
mpg