# MATH 189 Homework 7

## Due March 3rd, 2023

Places Rated Almanac rated 329 communities in the United States according to the following nine criteria:

1. Climate and Terrain

2. Housing

3. Health Care & the Environment

4. Crime

5. Transportation

6. Education

7. The Arts

8. Recreation

9. Economics

The rating results can be found in Places_Rated.txt. In this dataset, the first 9 columns represent the above 9 variables. The 10th column is the index of communities, ranging from 1 to 329. Note that, except for housing and crime, the higher the score is the better condition the community has. Analyze this dataset according to the following steps.

1. Calculate the eigenvalues and eigenvectors of the covariance matrix of **standardized** data (each column has mean 0 and variance 1). Calculate the proportion of total variance explained, by each eigenvector and the cumulative proportion of total variance explained by the first $k$ (=1,…,9) eigenvectors.

Report your results in scree plot and cumulative plot. Next, repeat the above steps to **raw** data, and draw scree plot and cumulative plot. Compare the results obtained from **raw** and **standardized** data.

2. Apply principal component analysis to the **standardized** data. Choose the number of principal components ($k$) according to the scree plot you obtained in Part 1. Report the corresponding principal component loading vectors. Visualize the dataset by projecting the observations onto the plane spanned by the first two principal components.