# CE 263 Midterm

## Scalable Spatial Analytics

In this midterm we will work on predicting single family homes sale prices. This task will require multiple skills you have acquired in the course (spatial data handling, exploratory analysis, model selection) in order to build a predictive model for a real world multivariate spatially distributed dataset.

**Data.** The dataset available to you in **housing_midterm_trn.csv** contains 20357 records of single family homes sold in Lucas County, Ohio, in 1993-1998. There are coordinates, date of sale (and derived dummy variable for the year of sale), as well as multiple variables providing essential information about each house. Some properties of the dataset are highlighted in the **Midterm_2016.ipynb**, which is also provided to you.

**Problem**. You are required to build a predictive model for the house prices, and apply it to predict the sale prices of 5000 houses contained in the file **housing_midterm_locs.csv**. It contains all the attributes of each house in the same format as the training dataset.

The accuracy of your model will be assessed in terms of the MSE.

Your submisison has to contain two parts:

1. Your predictions submitted via **Kaggle**.
2. Your code and a **brief pdf** report describing your approach.

Good luck!