# ASSIGNMENT 3
# PM569 Spatial Statistics, Fall 2018

Due October 19, 2018

## 1 Spatial Regression and Smoothing

We will use the Hurricane Harvey dataset for this question. Please use the grid created for you called windspGrid.csv for predictions.

a. Fit three different spatial models for atmospheric pressure with a covariate for wind speed using a) `lm()`, b) `gls()`, and c) `gam()`. For `gls()`, choose the best covariance/correlation structure between `corRatio`, `corGaus` and `corSpher` that includes an interaction term in the trend component. For `gam()`, use the default number of knots for the thin plate spline.
Describe the three modeling approaches and explain their differences. Where possible, compare the $\beta$ coefficients of x and y and their standard errors from each of the models. Also compare the model fits using `AIC()` on each model object. How are we treating the spatial structure in each? What are the pros and cons of each of these models?

b. Run `gam()` again using a larger number of knots (try 75% of the data size). Compare with your `gam()` results using the default settings. What is happening by increasing the number of knots?

c. Plot the semivariogram of the residuals from the `gam()` models, the `lm()` model, and your preferred `gls()` model. Examine and describe any residual spatial structure from these models.

d. Use `predict.gam()` and plot the predictions from the two `gam()` models (and their standard errors). Use `predict()` and plot the predictions from your preferred `gls()` model. Use the windspGrid data provided.

e. Interpolate atmospheric pressure using the inverse distance weighting (IDW) method. Make predictions in the windspGrid, and try two different values of $\rho$. Explain what changing $\rho$ does to your estimates. Discuss any advantages or disadvantages of the IDW method.

## 2 Areal Data: LA Neighborhoods and Tests of Spatial Association

a. Import the shapefile LAtracts into R and plot two maps: 1) showing median household income (MDHHI) and 2) showing the proportions of households with married couples with children (MARHH_CHD). You will have to caluclate proportions values (the HOUSEHOLDS variable represents the total number of households in a tract). Describe what break points you used for your map and whether you visualize any spatial clustering.

b. Construct rook, 2-NN and distance based adjacency matrices (choose a relatively small distance, can use `nbdists()` to examine distances). Plot the results on the map (separately) and describe the connectivity of the census tracts. Use the diffnb() function to examine the differences in the three matrices.

c. Look at the spatial correlogram of the rook matrix. What lag distance is best? Is 2-NN appropriate given this plot or should a different number of neighbours be used?

d. Calculate the global Moran's I for both variables given each of the 3 adjacency matrices and compare the results.