**Ecological Factors Associated with Self-Reported Mental Health Status**

Andrew S Cistola, Alyssa Berger

**Project Description**

Ecological factors such as socioeconomic status and health resource availability are known, relevant predictors of many health outcomes. However, the scope of research spatially identifying these factors in relation to mental health is relatively limited. There are many possible effects and interactions among these ecological factors and specific mental health conditions that have not been identified. Identifying which ecological factors are most influential in predicting mental health status would be vital information for preventative and targeted interventions.

**Specific Aims**

1. Identify local socioeconomic factors associated with a higher prevalence of poor self-reported mental health status
2. Identify area health resources that are associated with a lower prevalence of poor self-reported mental health status

**Research Design**

Specific Aim 1:

1. Collect estimated zip code level self-reported poor mental health status as percent of population from the 2020 release of the CDC and RWJF PLACES dataset (formerly 500 Cities project). This represents the outcome of interest.
2. Collect approximately 400 zip code level socio-economic variables as percent estimates in the detailed profile tables from the 2020 release of the US Census American Community Survey. These represent the possible predictors.
3. Connect outcomes with predictors, remove observations with missing values, impute missing data for predictor variables using median values, standard scale all variables.
4. Utilize open-source machine learning algorithms to identify variables with both high variation and high importance. Conduct cross-validated prediction to identify the smallest number of variables that will achieve the bets fitting model.
5. Use an artificial neural network with backwards propagation to predict zip codes in top quartile. Compare the predictive capability using (AUC test) all predictors, a random set of predictors, other known predictors (Neighborhood Deprivation Index), and the variables obtained in step 4.
6. Remove variables that are not consistent with any current theoretical construct test for OLS assumptions. Use domain knowledge to develop an appropriate mixed effects regression model for the purpose of identifying parameter estimates.

Specific AIM 2:

1. Collect estimated zip code level self-reported poor mental health status as percent of population from the 2020 release of the CDC and RWJF PLACES dataset (formerly 500 Cities project). This represents the outcome of interest.
2. Collect approximately 2000 county level health resource variables from the 2020 release of the HRSA Area Health resource File. These represent possible predictors.
3. Connect outcomes with predictors, remove observations with missing values, impute missing data for predictor variables using median values, standard scale all variables.
4. Using local Empirical Bayes smoothing and LISA quadrants, identify ‘hot and cold spot’ regions and assign nominal labels to contained counties. Use algorithms capable of multi-nominal prediction to identify health resources associated with each category.
5. Using zip code predictors, conduct geographic weighted regression to identify regions each predictor has significantly higher or lower coefficients. Assign nominal labels and use algorithms capable of multi-nominal prediction to identify health resources associated with each category.
6. Using domain knowledge create lists using the selected county variables associated with increased or decreased prevalence given the assumptions in step 4 and 5. Remove variables that are not consistent with any current theoretical construct and test for OLS assumptions.
7. For each county predictor create adjusted mixed-effects regression adjusting for zip code predictors that include the county predictor as an independent parameter (LISA) or as an interaction term (GWR). Compare statistically significant parameters estimates.