PROJECT PROPOSAL

TEAM: BOTHARA SHUBHAM (A20383550), ROY SUBHADEEP (A20358508)

INSTRUCTOR: Dr. DESPINA STASI

COURSE: MATH-564 Applied Statistics (Fall 2017)

SCHOOL: ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO

US County Premature Mortality Rate

Data sources:

http://www.countyhealthrankings.org/ranking-methods/exploring-data
http://opengovdata.pbworks.com/w/page/27141180/County%20Health%20Rankings

Description of the problem:

The County Health Rankings measure the health of all US counties and rank them within states. These ranks are estimated using composite scores from variety of health measures like quality of life, socio-economic factors, clinical care etc. The major goal of the rankings is to raise awareness about many factors that influence health and subsequently take corrective actions. A good indicator of the mortality measure for each county is YPLL (years of potential life lost) which indicates the cumulative years lost (typically calculated using predefined standards) due to premature death. Contribution to YPLL is the difference of premature deaths that are below 75 years of age as defined by the CDC. For example, if a person dies at the age of 72, then there is a contribution of 3 years towards the county YPLL. Any age of death above 75 is not counted towards YPLL as the criterion of expected age of 75 has supposedly been met.

YPLL data for each US county is available under general public license but without any supporting data. But, YPLL can be regressed if combined with external demographic data which well may act as a supporting evidence to a preventable action.

The YPLL for each 100,000 people, averaged across counties in the United States is between 8000 and 9000 depending on the year. The data file ypll.csv contains per-county YPLL's for the United States in 2011. The attached file additional_measures.csv contains demographic measures for each US counties.

We're going to clean the data, analyze which of the additional measures correlate strongly with our mortality measure, and fit/analyze regression models for YPLL.