Predicting Heart Disease Mortality

A Nationwide Socio-Economic View Across United States Counties

Christopher Fraher, July 2018

# Executive Summary

The data reviewed in this report were provided as part of the MPP Capstone Challenge, hosted by Microsoft, for the purpose of predicting heart disease mortality rates at the county level across the United States. The makeup of the 34 data elements provided for this task include socioeconomic indicators for 1,599 counties over two years, for a total of 3,198 observations across 34 features. The outcome measures the number of individuals who have expired in the rate per 100,000 and

Analysis on this dataset was performed using Tableau v10.3 and R Studio 1.1.453.

# Data Exploration

## Categorical Values

The data consist of primarily numeric components, most of which are percentage values with a few exceptions. Only 4 categorical values exist: area\_\_rucc, area\_\_urban\_influence, area\_\_urban\_influence, and yr.

