# MULTIPLE LINEAR REGRSSION MODELLING CASE STUDY

## Predict the life expectancy by state using the state statistics

### About the data

We will be examining the "**state**" dataset, which has data from the 1970s on all fifty US states. For each state, the dataset includes the population, per capita income, illiteracy rate, murder rate, high school graduation rate, average number of frost days, area, latitude and longitude, division the state belongs to, region the state belongs to, and two-letter abbreviation

This dataset has 50 observations (one for each US state) and the following 15 variables:

- **Population** - the population estimate of the state in 1975
- **Income** - per capita income in 1974
- **Illiteracy** - illiteracy rates in 1970, as a percent of the population
- **Life.Exp** - the life expectancy in years of residents of the state in 1970
- **Murder** - the murder and non-negligent manslaughter rate per 100,000 population in 1976
- **HS.Grad** - percent of high-school graduates in 1970
- **Frost** - the mean number of days with minimum temperature below freezing from 1931–1960 in the capital or a large city of the state
- **Area** - the land area (in square miles) of the state
- **state.abb** - a 2-letter abreviation for each state
- **state.area** - the area of each state, in square miles
- **x** - the longitude of the center of the state
- **y** - the latitude of the center of the state
- **state.division** - the division each state belongs to (New England, Middle Atlantic, South Atlantic, East South Central, West South Central, East North Central, West North Central, Mountain, or Pacific)
- **state.name** - the full names of each state
- **state.region** - the region each state belong to (Northeast, South, North Central, or West)

## Problem Statement:

Build an analytical model to **predict the life expectancy by state using the state statistics**. Build the model with all potential variables included (Population, Income, Illiteracy, Murder, HS.Grad, Frost, and Area). Note that you should use the variable "Area" in your model, NOT the variable "state.area".