

# Life Expectancy Analysis using Regression

## Description of the files

All files contain information at state level. In the US there are fifty states plus one federal district (District of Columbia) and five inhabited territories (Puerto Rico, American Samoa, Guam, Northern Mariana Islands, and US Virgin Islands).

### 1. Education file

The file "edu.csv" contains the following columns:

- name of the state
- % high school graduate or higher
- high School rank
- % bachelor degree or higher
- bachelor degree rank
- % advanced degree or higher
- advanced degree rank

We are not going to use the rank columns, so either don't load them or delete them afterwards.

### 2. Life Expectancy file

The file "lifeExpectancy.csv" contains the following columns:

- State: name of the state
- LifeExp2018: life expectancy (2017)
- LifeExp2010: life expectancy (2010)
- MaleLifeExp: male life expectancy
- FemLifeExp: female life expectancy

### 3. Crime file

The file "crime.csv" contains the following columns:

- name of the state
- population (total inhabitants) (2015)
- murders and non-negligent manslaughter (total deaths) (2015)
- murders (total deaths) (2015)
- gun murders (total deaths) (2015)
- gun ownership (percent) (2013)
- murders and non-negligent manslaughter rate (per 100,000) (2015)
- murder rate (per 100,000)
- gun murder rate (per 100,000)

To check if the numbers contain spurious characters, you can use regular expressions.

### 4. Area file

The file "area.csv" contains the following columns:

- State: name of the state
- TotalRank: total area rank
- TotalSqMi: total area in SqMi

- TotalKmQ: total area in KmQ
- LandRank: land area rank
- LandSqMi: land area in SqMi
- LandKmQ: land area in KmQ
- LandPer: land area percentage
- WaterRank: water area rank
- WaterSqMi: water area in SqMi
- WaterKmQ: water area in KmQ
- WaterPer: water area percentage

We are not going to use the rank columns, so either don't load them or delete them afterwards.

## 5. Income file

The "income" file is in Excel format. It contains the following columns:

- Rank: rank for income in 2017
- State: name of the state
- Income2017: median household income in 2017
- Income2016: median household income in 2016
- ...
- Income2007: median household income in 2007

We are not going to use the rank column, so either don't load it or delete it afterwards.

## 6. Region file

The region file is a text file containing the following columns:

- Name: name of the state
- Abb: abbreviation of the name of the state
- Region: the region that each state belongs to (Northeast, South, North Central, West)
- Division: state divisions (New England, Middle Atlantic, South Atlantic, East South Central, West South Central, East North Central, West North Central, Mountain, and Pacific)