State data  
  
Description  
Data set related to the 50 states of the United States of America (rows).  
  
Dataset contains the following statistics in the respective columns:  
  
Population: population estimate as of July 1, 1975  
  
Income: per capita income (1974)  
  
Illiteracy: illiteracy (1970, percent of population)  
  
Life Exp: life expectancy in years (1969-71)  
  
Murder: murder and non-negligent manslaughter rate per 100,000 population (1976)  
  
HS Grad: percent high-school graduates (1970)  
  
Frost: mean number of days with minimum temperature below freezing (1931-1960) in capital or large city  
  
Area: land area in square miles  
  
**TASK FOR THE PROJECT**  
  
Provide a **qualitative description** of the variables in the dataset and of **their distribution**, using histograms, pie charts, tables or other graphical instruments.

Then answer specifically to the following questions:

1. Divide the states into two groups, one with **low illiteracy percentage**, and one **with high illiteracy percentage**. Is there a significant difference in the **mean murder rate in the two groups**? *Median?*

2. Divide the states into two groups, one with **low per capita income**, and one with **high per capita** **income**. Is there a significant difference in the **mean murder rate in the two groups**? *Median?*

3. Compute the **mean population density for each state** (=Population/Area) and divide the states into two groups, one with **low population density**, and one with **high population density**. Is there a significant difference in the **mean murder rate in the two groups**? *Median?*

4. Discuss from the results of the previous points **which of the considered variables (illiteracy, income, population density) have more influence in reducing or increasing the murder rate**, taking care also of the presence of ***possible correlations between the variables***.

1. Frost and lif.exp as other two correlated var to murder
2. Due modi (o più) di dividere in due i campioni: a) mediana b) quantile scelto da noi c) metà del range ( (2.8+0.5)/2 = 1.65)
3. Se le var dei due campioni sono diversi il t-test deve essere aggiustato come un Welch’s test\*

\*approfondisci [Welch’s test](https://en.wikipedia.org/wiki/Welch%27s_t-test)