**Exploratory Data Analysis (EDA)**

EDA is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. It is about knowing your data, gaining a certain amount of familiarity with the data, before one starts to extract insights from it.

# Where is EDA in the Machine Learning Process:

# 

**Objectives of EDA:**

* To gain an understanding of data and find clues from the data
* To formulate assumptions and hypothesis for our modelling
* To check the quality of data for further processing and cleaning if necessary

**Data Visualization with Seaborn**

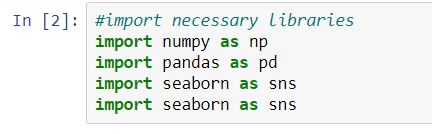
Seaborn is a Python data visualization library based on [matplotlib](https://matplotlib.org/). It provides a high-level interface for drawing attractive and informative statistical graphics.

**Installation of Seaborn:**

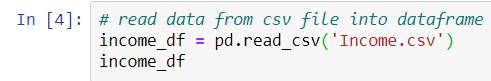
* pip install seaborn
* In anaconda seaborn is already inbuilt package. Can be accessed using the below statement
  + import seaborn

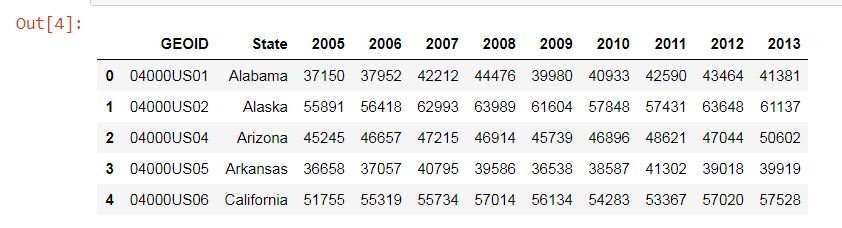
## Scatter plot:

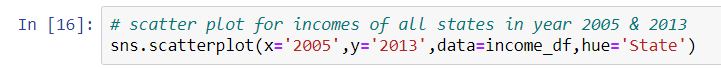
There are several ways to draw a scatter plot in seaborn. The most basic, which should be used when both variables are numeric, is the scatterplot() function.

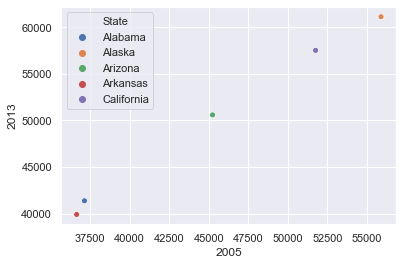


Dataset: [Income.csv](https://drive.google.com/file/d/1j--coT6x2PZXSZHehlfMmA2tSMYsvnrv/view?usp=sharing)









* x, y 🡪 Input data variables must be numeric.

Can pass data directly or reference columns in data.

* hue 🡪 Grouping variable that will produce points with colors.

Can be either categorical or numeric.