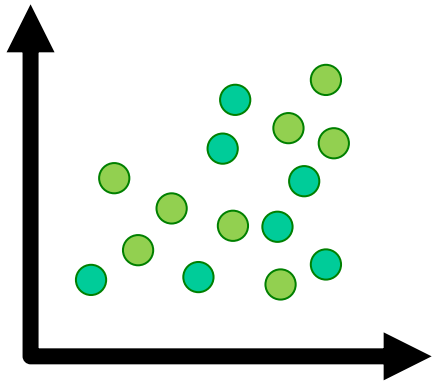
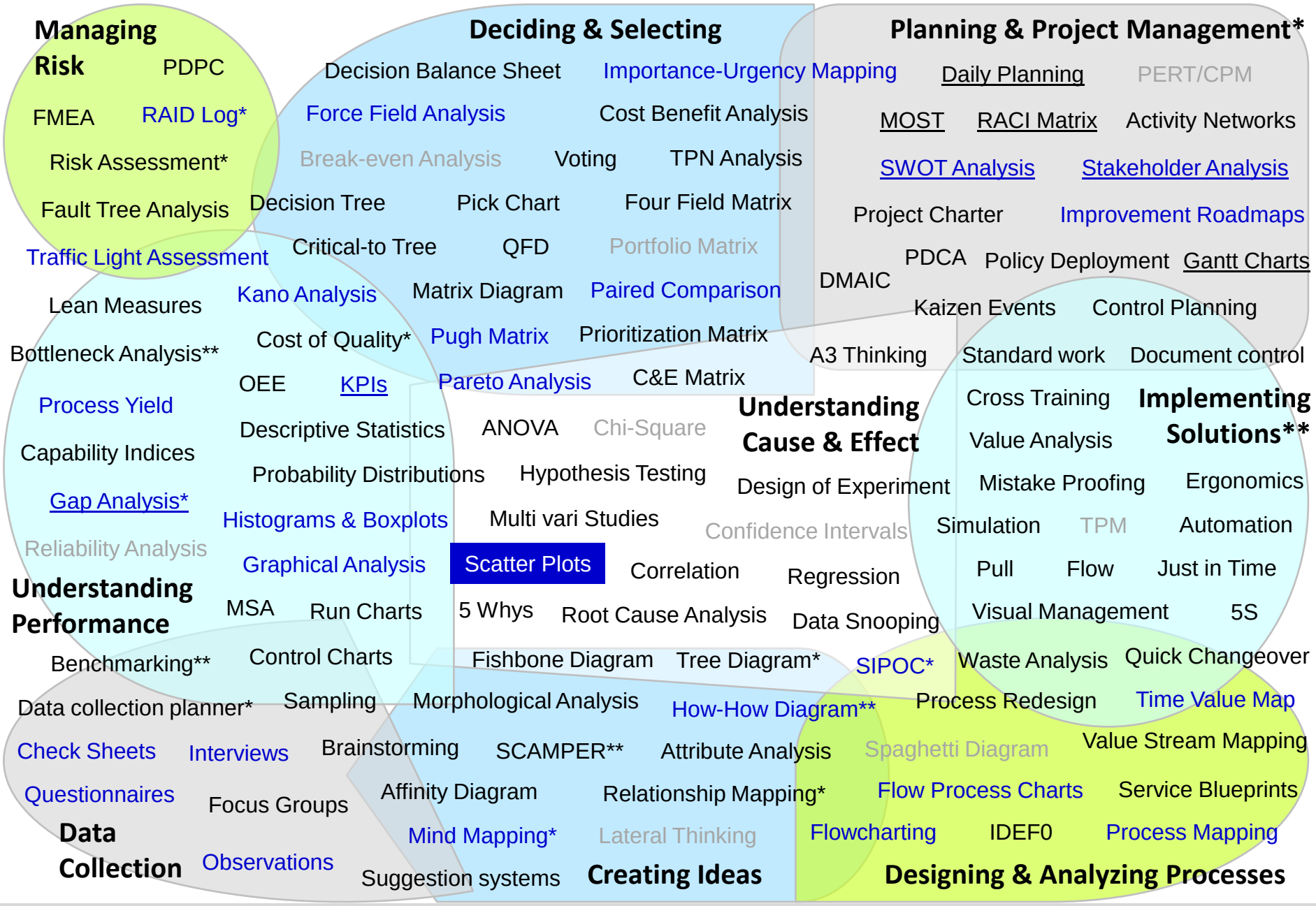
Continuous Improvement Toolkit

Scatter Plots

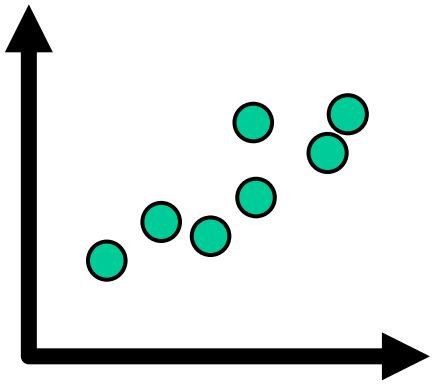


The Continuous Improvement Map



- Scatter Plots

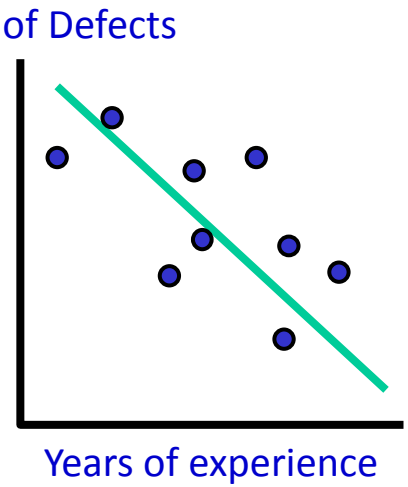
- A diagram that shows whether two variables are correlated.
- Shows patterns in the relationship that cannot be seen by just looking at the data.
- Used as a first step in analyzing correlation between pairs of variables before conducting advanced statistical analyses.
- Works with both **continuous** and **count** data.



- Scatter Plots

Example:

- A line manager for example may want to check the relationship between:
  - The number of **training hours** and **employee productivity**.
  - The number of **defects** and the **experience of the staff**.
  - The equipment **downtime** and its **cost of maintenance**.



- Scatter Plots

Other Examples:

- **The relationship between:**
  - Driving speed and fuel consumption.
  - The number of people working on a shift and the average answer time in a call center.
  - The number of years of education someone has and the annual income of that person.

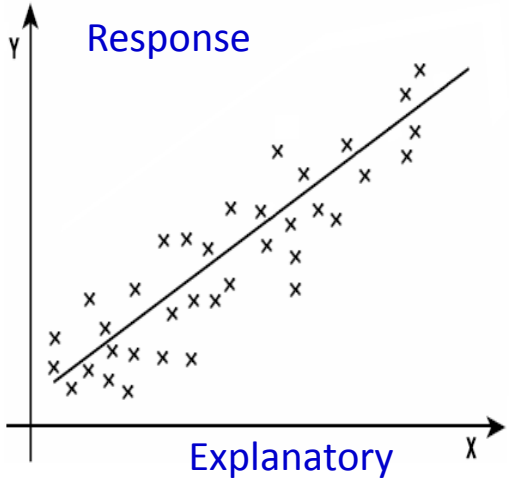
- Scatter Plots

- Used to visually investigate the relationship between two variables.
- Used to verify that a change in one variable can affect the other variable.
- Helps detecting the primary factors that are really causing a problem.
- Helps eliminating non-critical factors from consideration.
- Used to determine the strength of the relationship.
- Used with statistical tools to support or reject hypotheses about the data.

- Scatter Plots

- When comparing an input with an output variable.
  - The **explanatory variable** is normally placed on the horizontal axis.
  - The **response variable** is placed on the vertical axis.

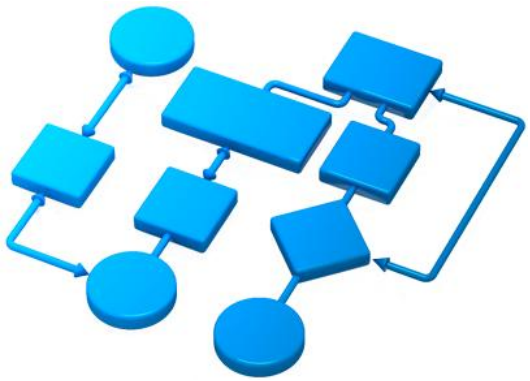
You may also compare two input or output variables



- Scatter Plots

How to Construct a Scatter Plot?

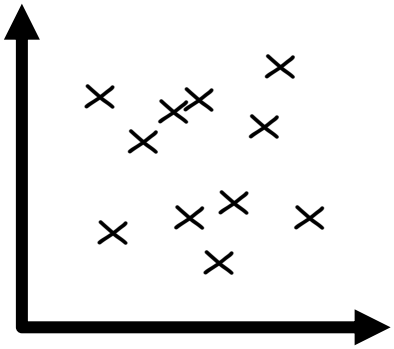
- Collect the two paired sets of data.
- Create a summary table of the data.
- Draw and label the horizontal and vertical axes.
- Plot the data pairs on the diagram by placing a dot at the intersection of each data pair.
- Look at how the two variables vary together.



- Scatter Plots

- Scatter plots can indicate several types of correlation:

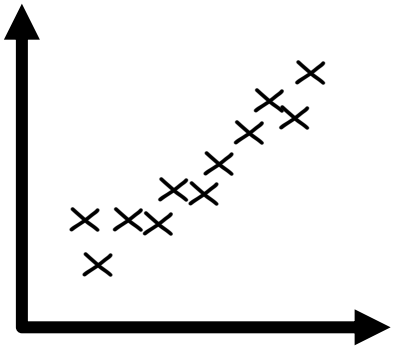
No correlation when the data points are scattered randomly without showing any particular pattern



- Scatter Plots

- Scatter plots can indicate several types of correlation:

A **positive correlation** occurs when the values of one variable increase as the values of the other also increase

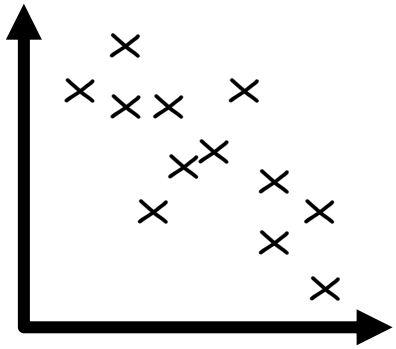


The fitted line slopes from bottom left to top right

- Scatter Plots

- Scatter plots can indicate several types of correlation:

A **negative correlation** occurs when the values of one variable increase as the values of the other decrease

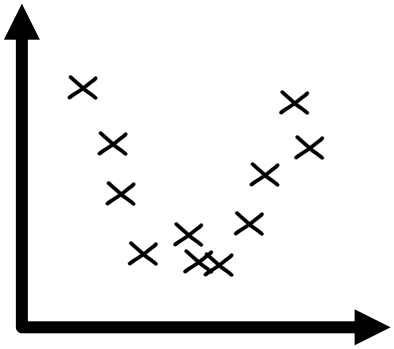


The fitted line slopes from upper left to lower right

- Scatter Plots

- Scatter plots can indicate several types of correlation:

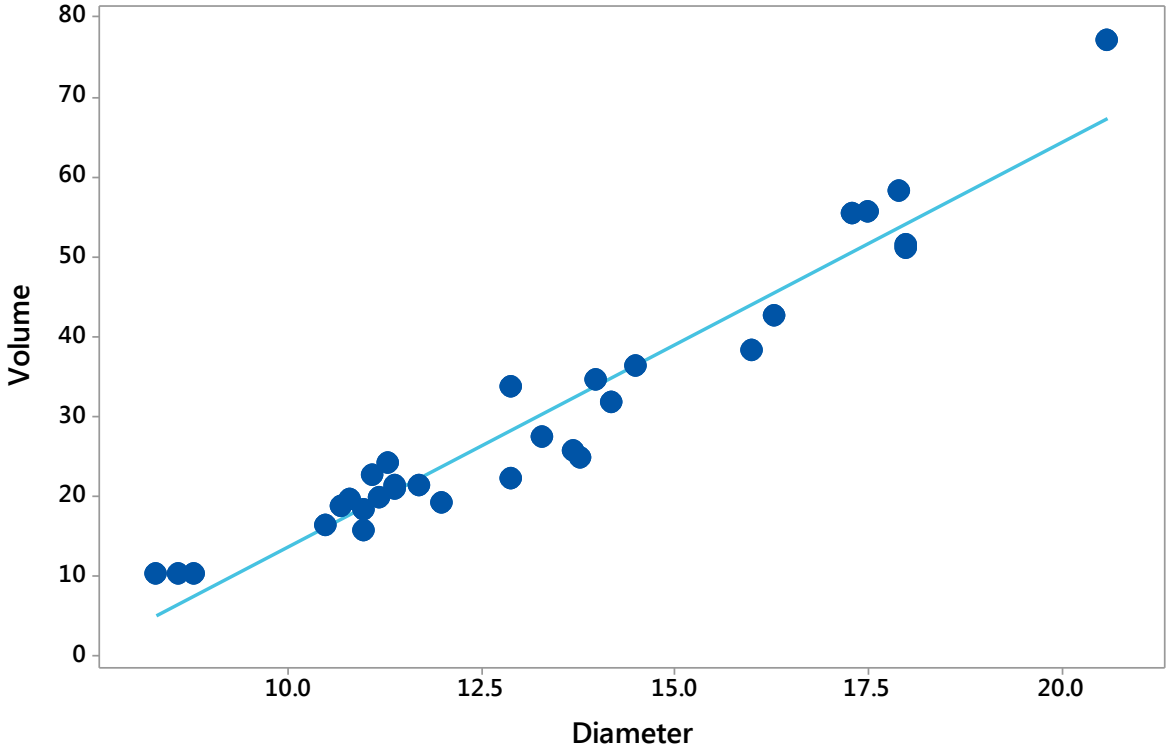
Scatter plots can also indicate **nonlinear relationships** between variables



- Scatter Plots

Example:

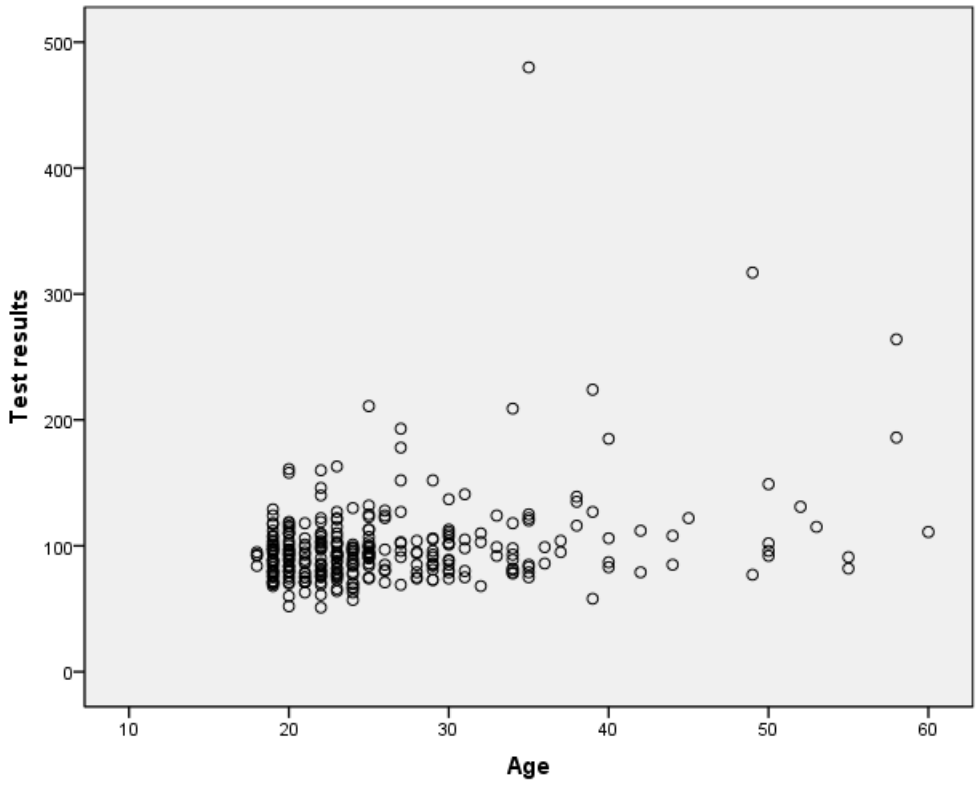
Diameter	Height	Volume
8.3	70	10.3
8.6	65	10.3
8.8	63	10.2
10.5	72	16.4
10.7	81	18.8
10.8	83	19.7
11	66	15.6
11	75	18.2
11.1	80	22.6
11.2	75	19.9
11.3	79	24.2
11.4	76	21



The **volume** and the **diameter** of sample trees in a forest

- Scatter Plots

Example – An analysis that was conducted for diagnosing the presence of diabetes at a workplace.



The population is generally young (75.8% are below thirty).

This scatter plot illustrates that there is no obvious relationship between age and glucose levels.

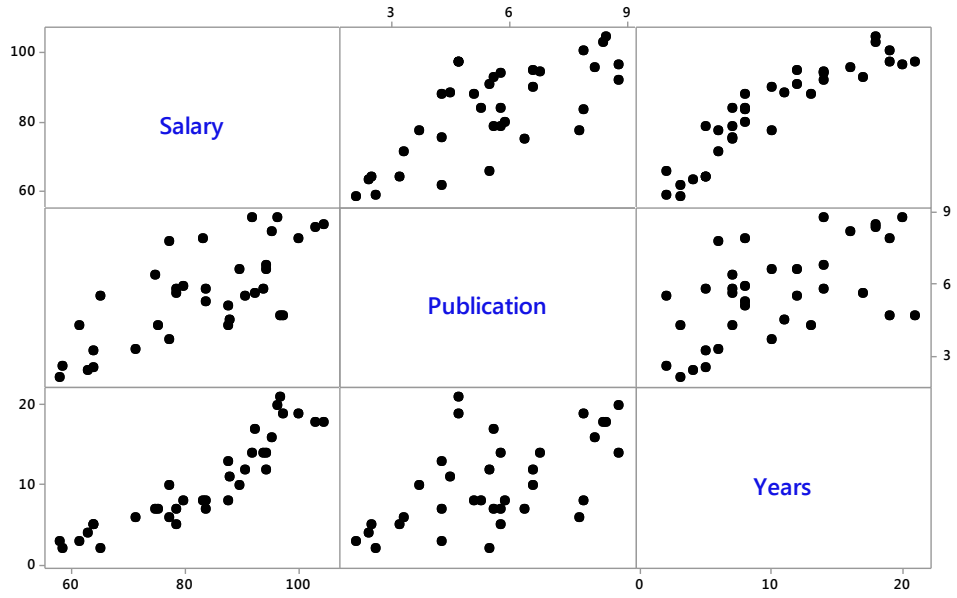
High glucose levels are found in all ages above twenty, and normal glucose levels are found in higher ages.

- Scatter Plots

Matrix Plots:

- Summarizes the relationship between several variables.
- Produces a scatter plot for every combination of variables.
- Allows to visually assess the variables that might be related.

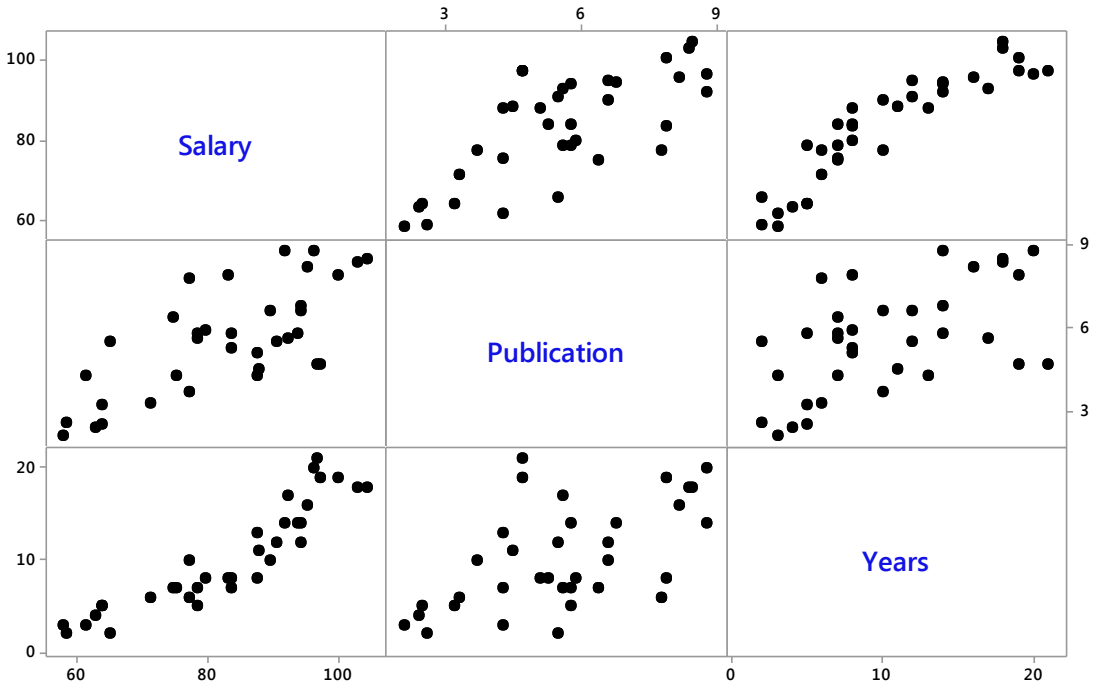
Potential correlations can then be identified



- Scatter Plots

Example:

Salary	Gender	Publication	Years
88	F	4.5	11
77.3	M	7.8	6
75.3	M	4.3	7
96.4	M	8.8	20
87.7	M	5.1	8
58.1	F	2.1	3
63.1	F	2.4	4
58.5	M	2.6	2
95.4	F	8.2	16
92	F	8.8	14
94.5	M	6.6	12
103	M	8.4	18



There is a relationship between the years of experience and salaries

The number of publications does not appear to be correlated with the years of experience

- Scatter Plots

Further Information:

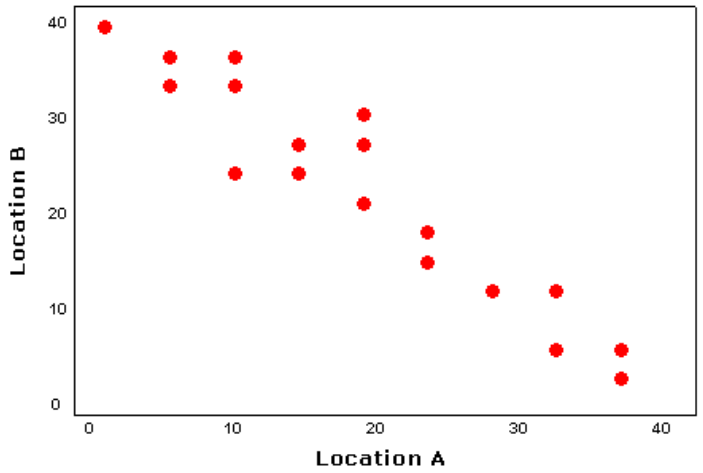
- When the relationship is not so clear, **Correlation** can be used to help determine if a relationship exists between the variables. **Regression** techniques go a step further by defining the relationship in a mathematical format.
- Be careful before concluding that there is a direct cause-and-effect relationship between the variables. There might be a third factor that is causing the change in the two variables.
- You can also illustrate a stratification factor in the scatter plot. For example, the relationship between a process output and a process input for two different settings.

-- Scatter Plots

Example - The amount of sales per month generated at two locations:

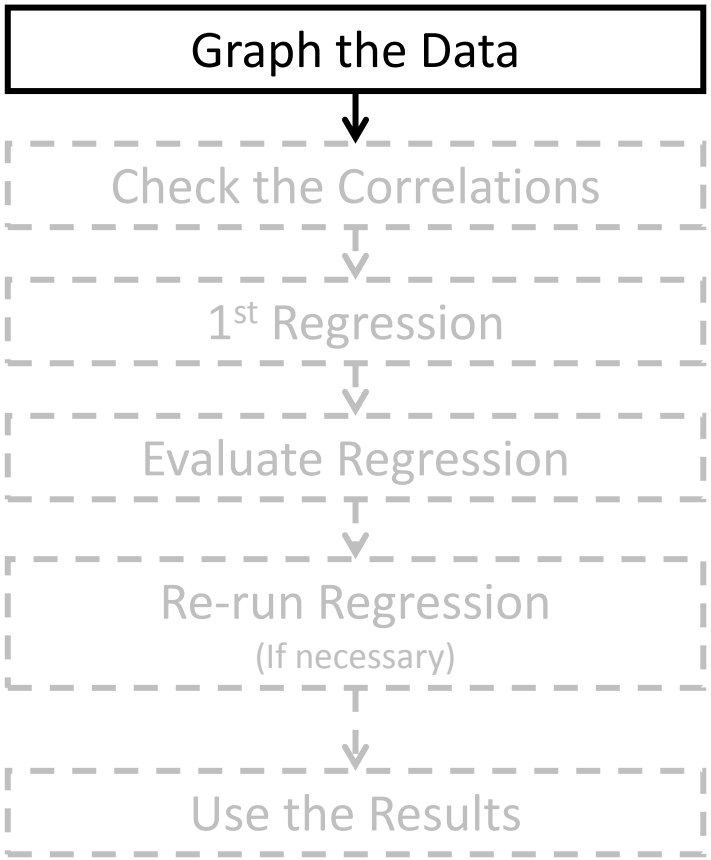
- The plotted points form a negative slope.
- The sales at location B is inversely related to the sales at location A.
- Does this mean that location A caused the decrease in sales at location B, or vice versa?

Answer: Not necessarily, unless the two locations are direct competitors.



- Scatter Plots

Where Does It Fit?



Scatter plot

Use Pearson Coefficient

Linear / Multiple regression

R-squared & analyze residuals

Simple: With different model (Cubic)
Multiple: Remove unnecessary items

Control critical process inputs & select best operating levels.