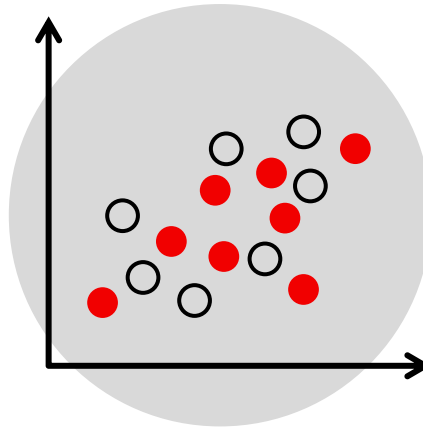


Continuous Improvement Toolkit

Scatter Diagram

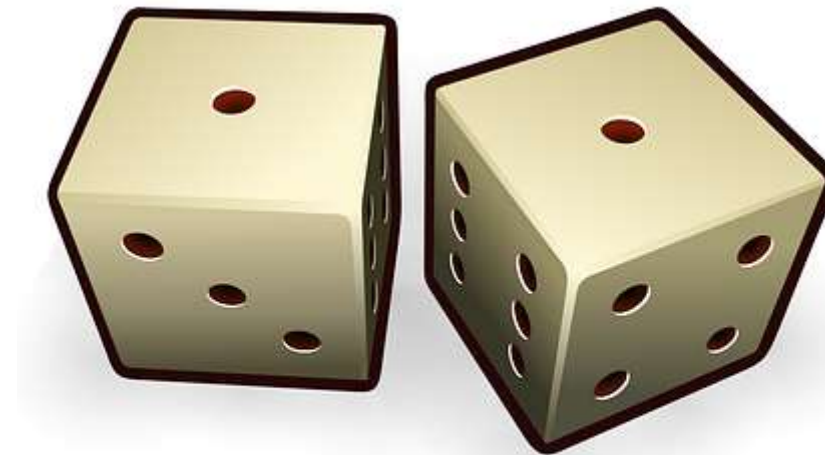


The Continuous Improvement Map



Scatter Diagram

Many situations require the investigating whether a **relationship** exists between two or more variables



Scatter Diagram

A line manager may want to check the relationship between the number of **training hours** and **employee productivity**



Or if the number of **defects** is a function of the **experience** of the person causing it

Scatter Diagram

Other Examples

The relationship between **equipment downtime** and its **cost of maintenance**



Scatter Diagram

Other Examples

The relationship between **driving speed** and **fuel consumption**



Scatter Diagram

Other Examples

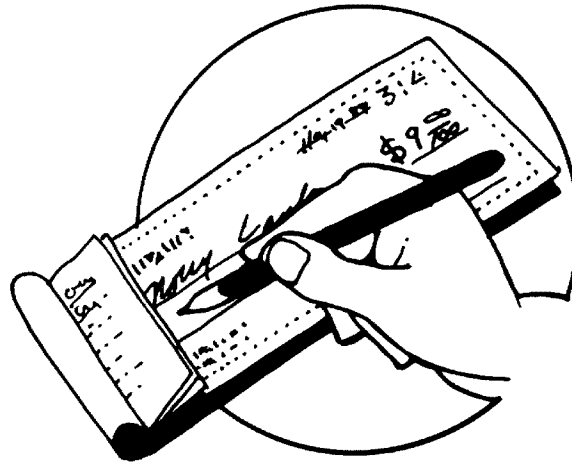
The relationship between the **number of people working on a shift** and the **average answer time** in a call center



Scatter Diagram

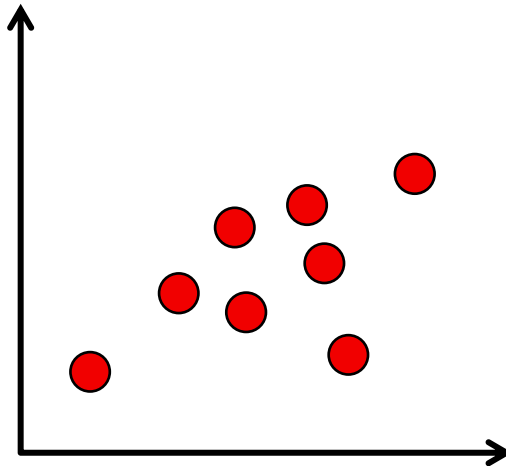
Other Examples

The relationship between the **number of years of education** someone has and the **annual income** of that person



Scatter Diagram

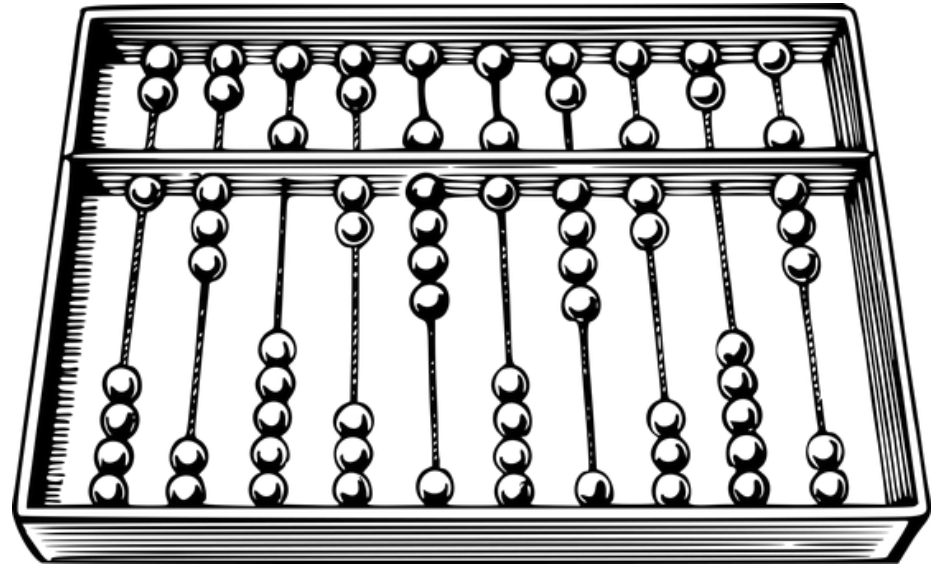
A diagram that shows whether two variables are **correlated** or related to each other



It shows patterns in the relationship that cannot be seen by just looking at the data

Scatter Diagram

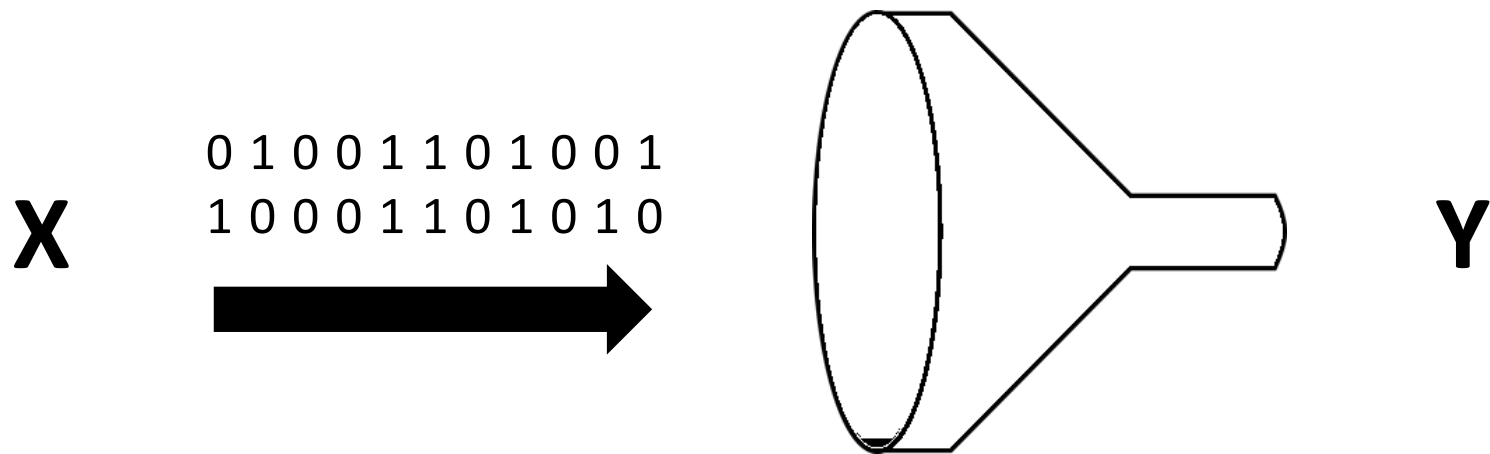
It works with both **continuous** and **count** data



Scatter Diagram

Primarily used to visually investigate the relationship between **two variables**

Often an output and an input variables



Scatter Diagram

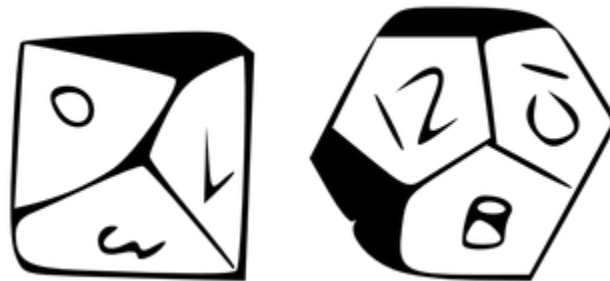
This is useful to **verify** that a change in one variable can affect the other variable



It helps detecting the **primary factors** that are really causing a problem and hence eliminating non-critical factors from consideration

Scatter Diagram

Often used as a **first step** when analyzing and communicating the correlation between pairs of variables



Before conducting advanced statistical techniques

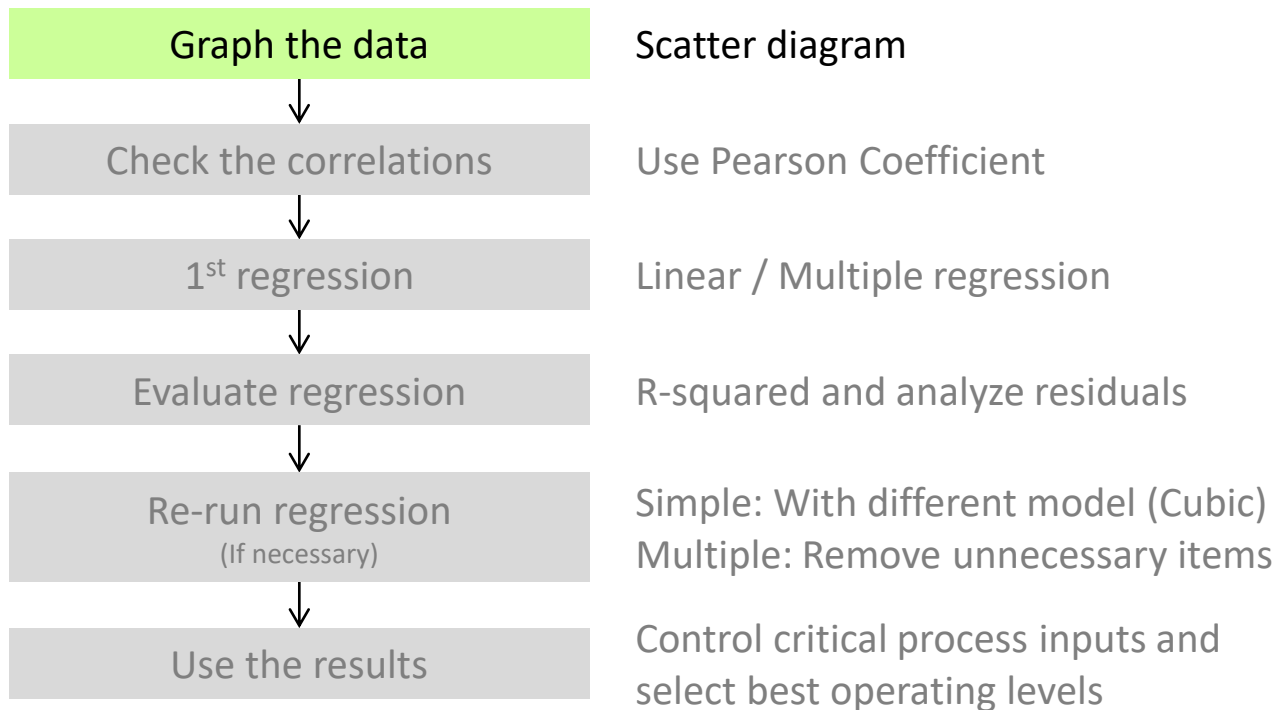
Scatter Diagram

Often used with **statistical tools** (such as regression) to support or reject hypotheses about the data



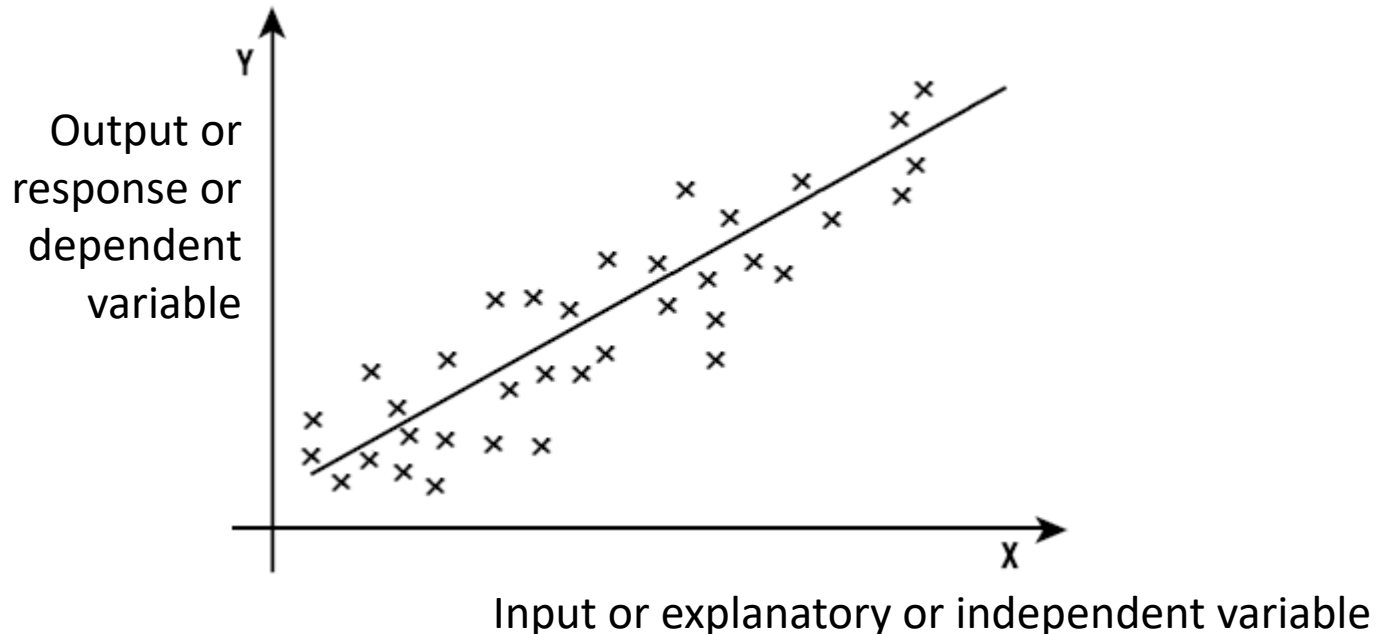
Scatter Diagram

Where do scatter diagrams fit?



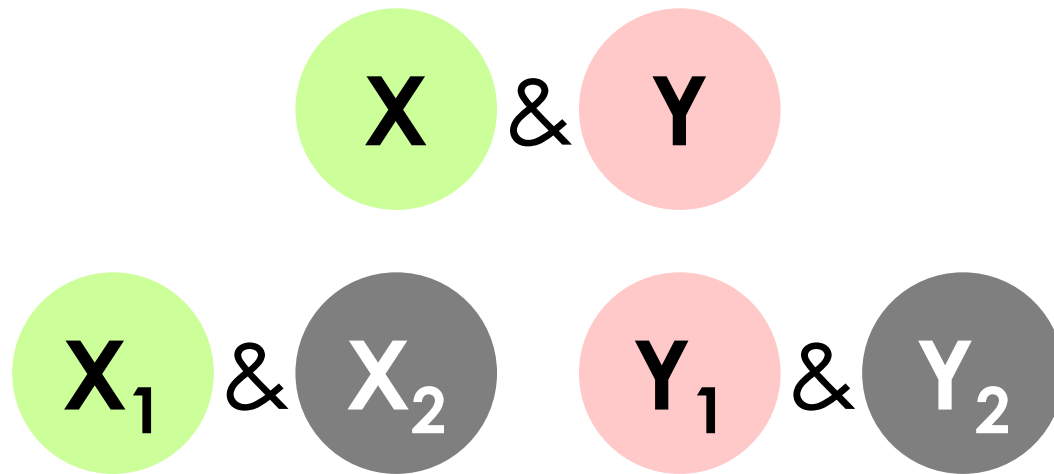
Scatter Diagram

The input variable is normally placed on the **horizontal axis** while the output variable is placed on the **vertical axis**



Scatter Diagram

You may also compare **two input or output variables** to each other

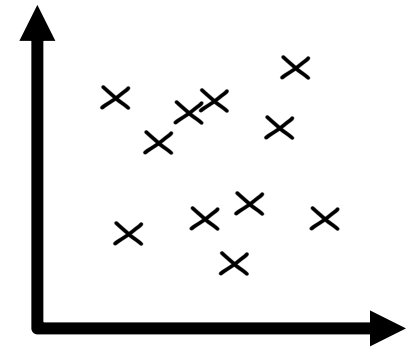


In this case, it doesn't matter which variable goes on the horizontal axis and which goes on the vertical axis

Scatter Diagram

Scatter diagrams can indicate **several types** of correlation

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

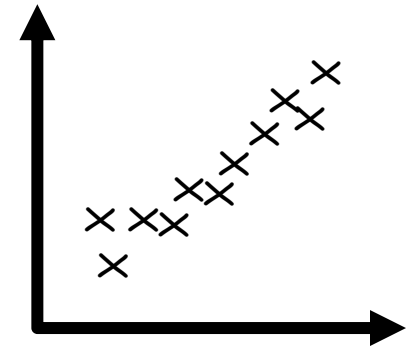


Note that the width of the scattered pattern reflects the strength of the relationship

Scatter Diagram

Scatter diagrams 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.

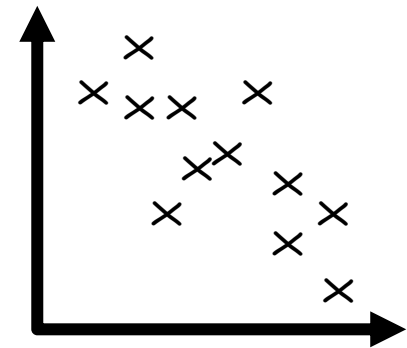


The fitted line slopes from bottom left to top right

Scatter Diagram

Scatter diagrams can indicate **several types** of correlation

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

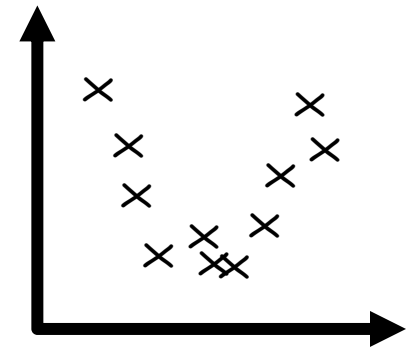


The fitted line slopes from upper left to lower right

Scatter Diagram

Scatter diagrams can indicate **several types** of correlation

Scatter diagrams can also indicate **nonlinear** relationships between variables.

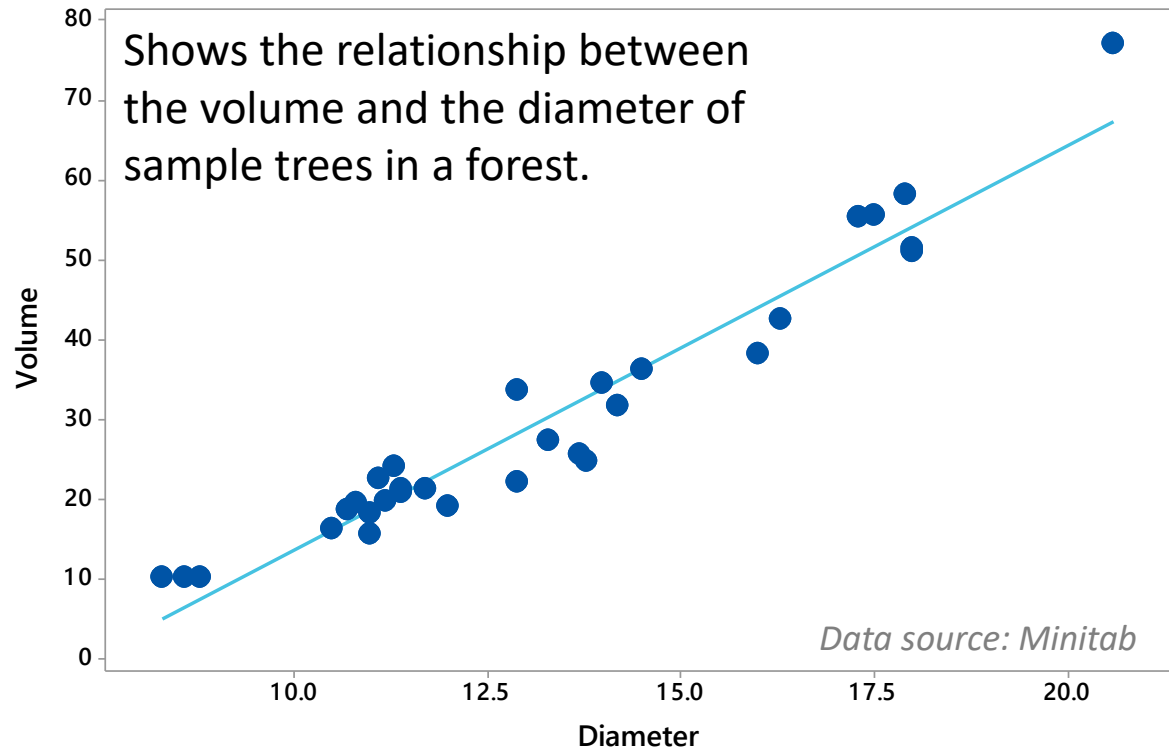


Scatter Diagram

Example – Forest Trees

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

Both are output variables

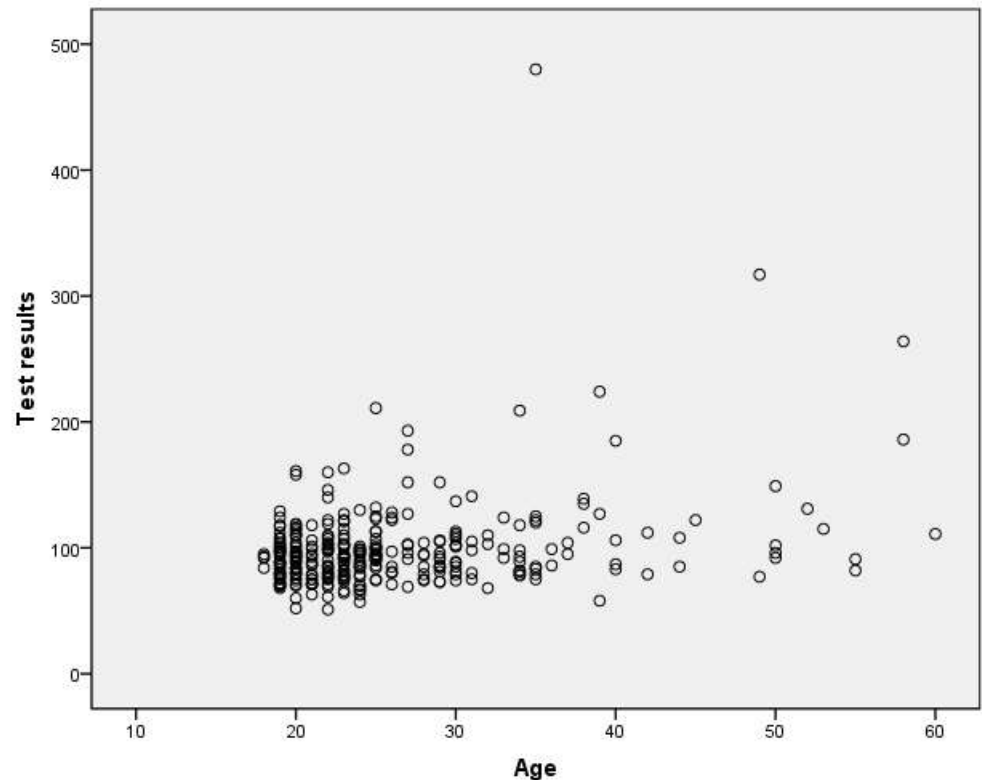


Scatter Diagram

Example – Presence of Diabetes at a Workplace

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

The population was generally **young** (75.8% were below thirty).

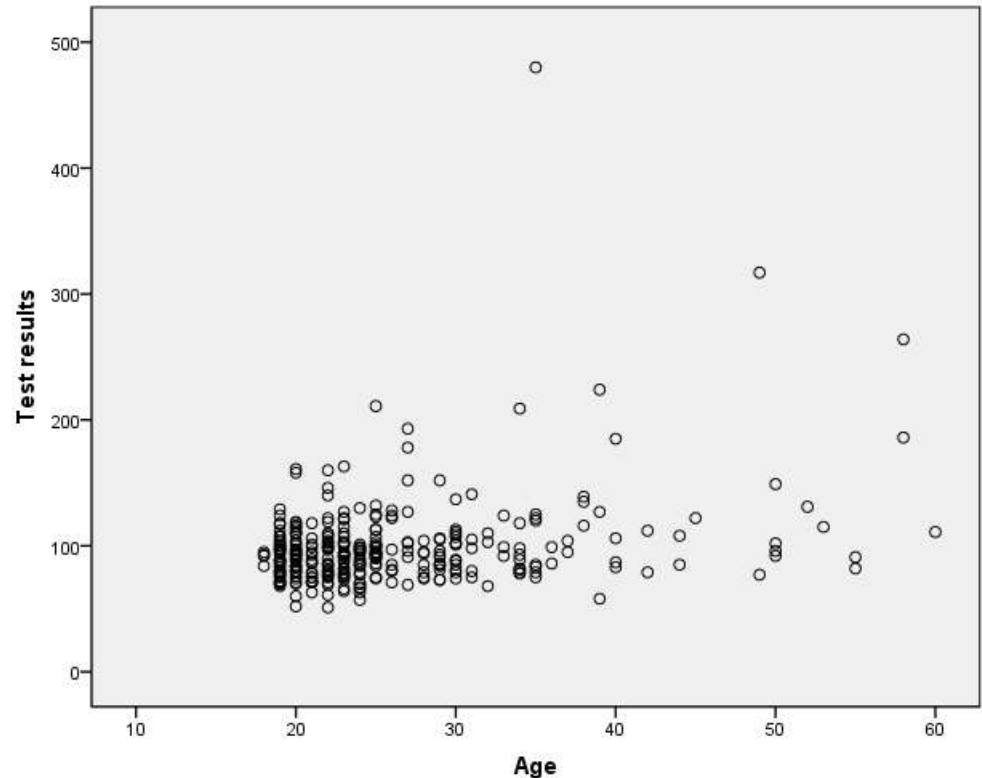


Scatter Diagram

Example – Presence of Diabetes

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

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

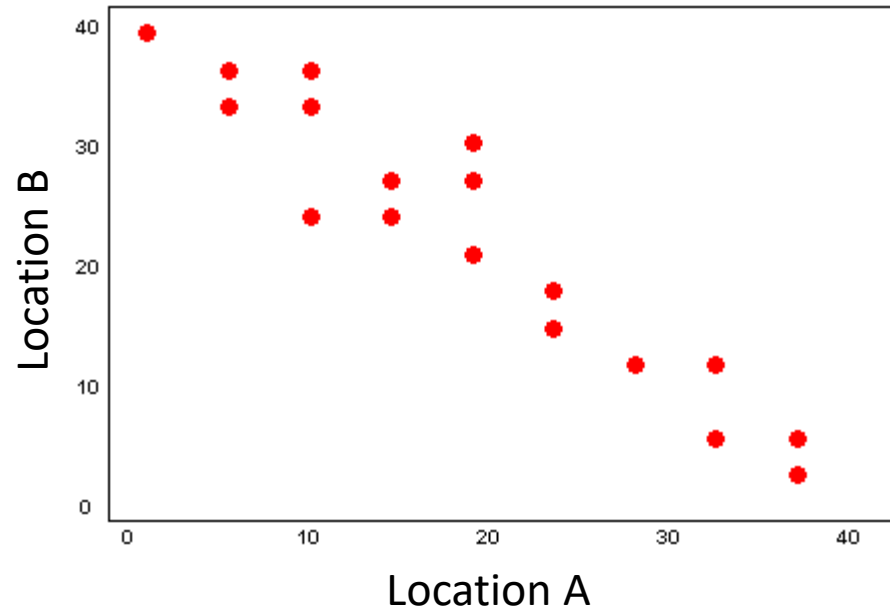


Scatter Diagram

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.

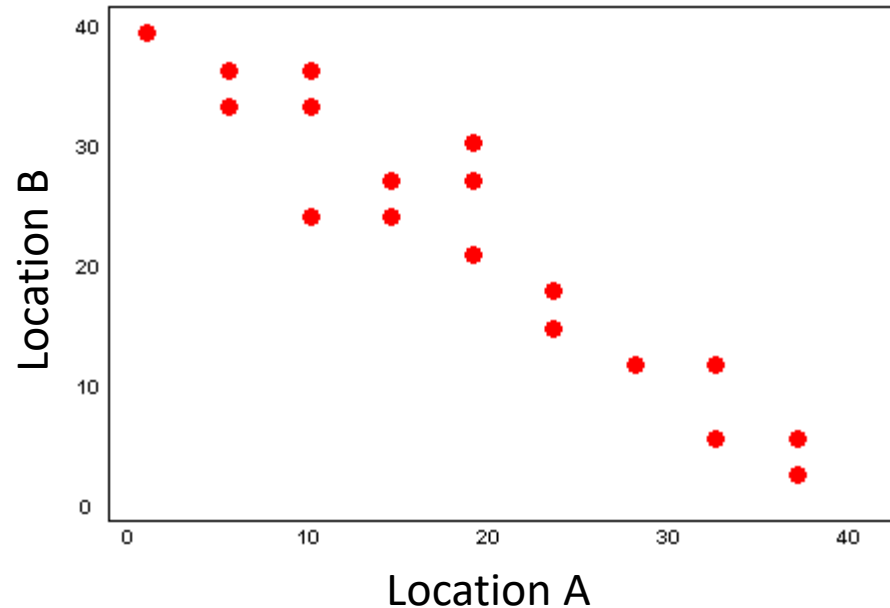


Scatter Diagram

Example – The amount of sales per month generated at two locations

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.



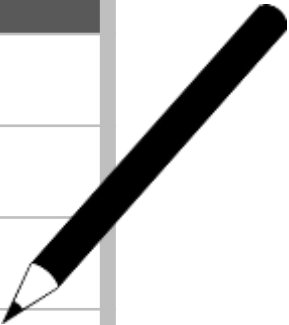
Scatter Diagram

How to Construct a Scatter Diagram

Collect the two paired sets of data

Create a summary table of the data

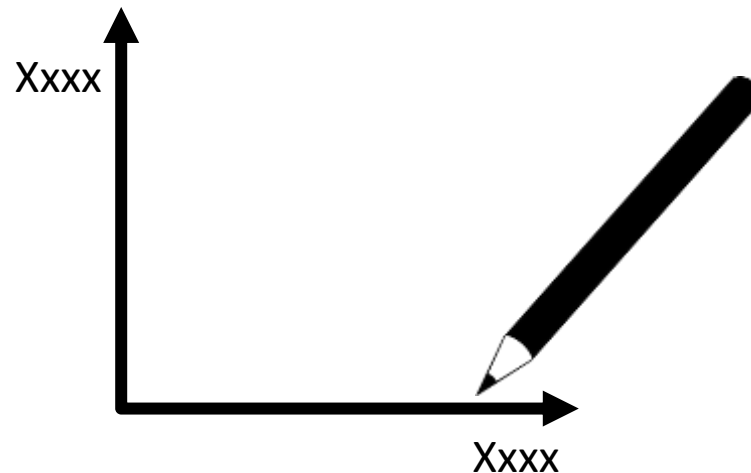
Variable 1	Variable 2



Scatter Diagram

How to Construct a Scatter Diagram

Draw and **label** the horizontal and vertical axes with variable names and scale values

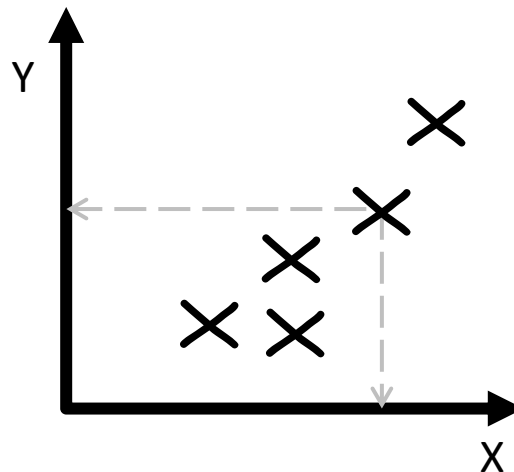


Scatter Diagram

How to Construct a Scatter Diagram

Plot the data pairs on the diagram by placing a dot at the **intersection** of each data pair

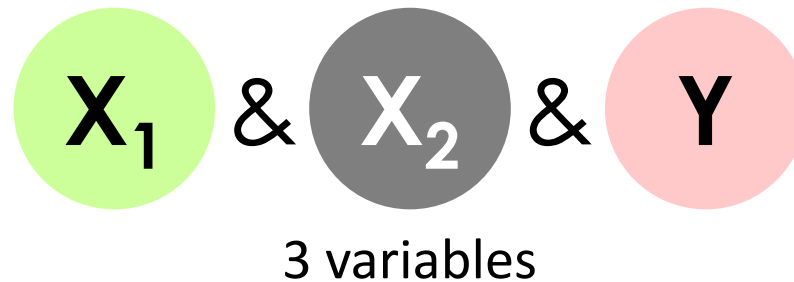
Look at how the pattern appears and how the two variables vary together



Scatter Diagram

Matrix Plots

Summarizes the relationship between **pairs of multiple variables** in one graph

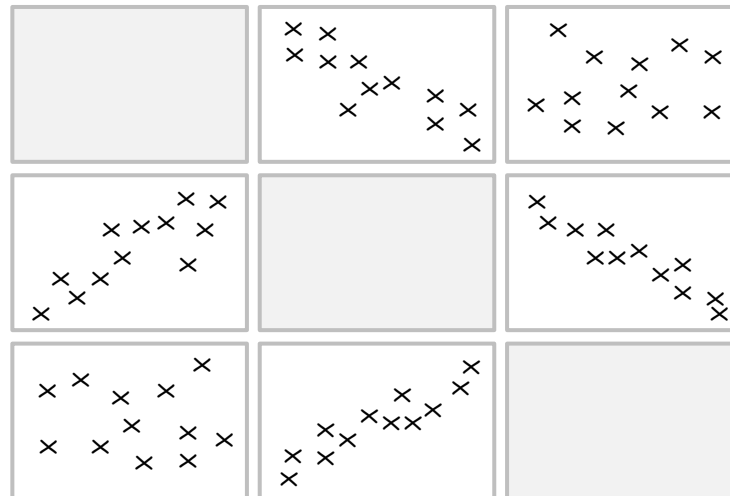


Allows to visually assess the variables that might be related

Scatter Diagram

Matrix Plots

Produces a scatter diagram for every combination of variables

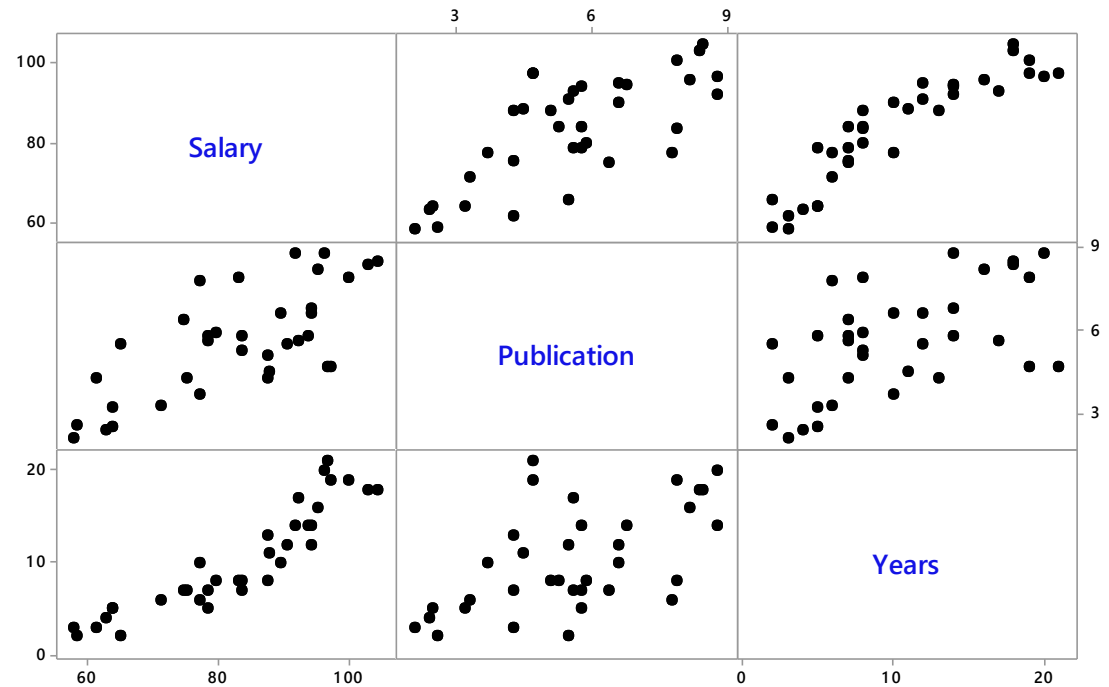


Potential correlations between pairs of variables can then be identified

Scatter Diagram

Matrix 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

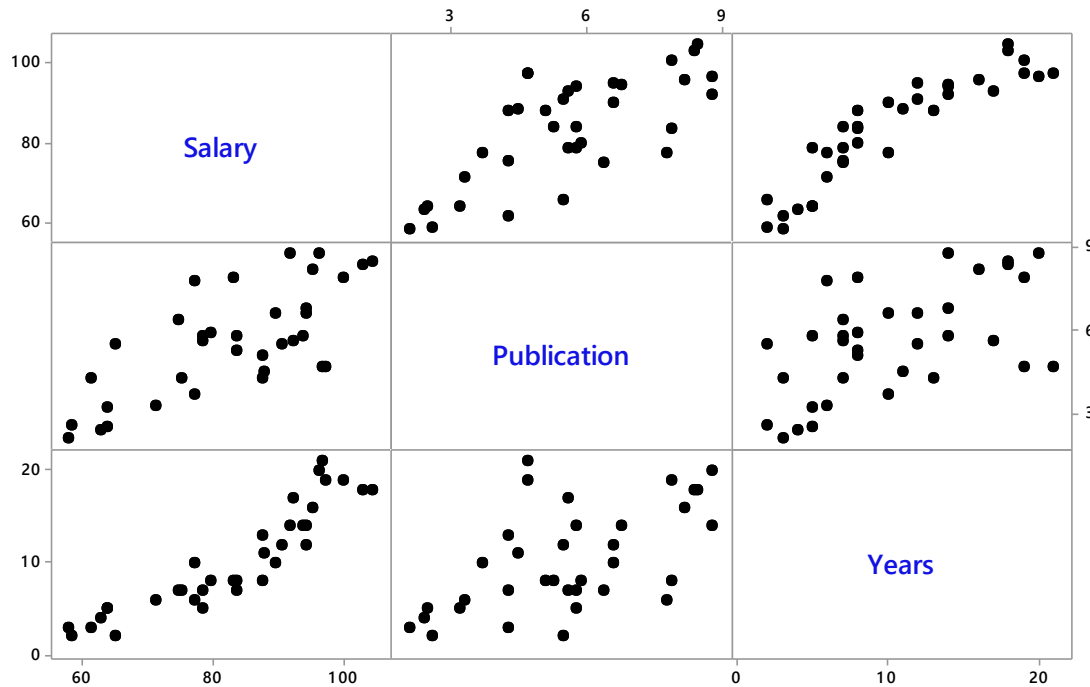


of publications doesn't appear to be correlated with the years of experience

There may be a relationship between the years of experience and salaries

Scatter Diagram

Matrix Plots – Example



Is there a correlation between the **number of publication** and **salaries**?

Scatter Diagram

Further Information – One of the 7 Basic Tools of Quality

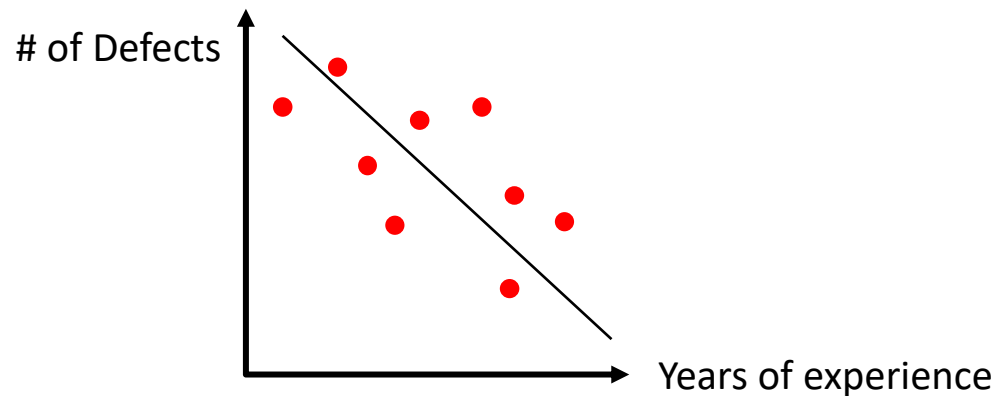


Scatter Diagram

Further Information

When the relationship is not so clear, **Correlation** can be used to help validate if a relationship exists between the variables

Regression techniques go a step further by defining the relationship in a mathematical format



Scatter Diagram

Further Information

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

$$Y = f(x)$$

Scatter Diagram

Further Information

You can also illustrate a **stratification factor** in scatter diagrams

For example, the relationship between a process output and a process input for two different settings

