

# Scatter Plot

A Scatter Plot (also called scatter diagram) is a **static** data display used to investigate whether a relationship exists between two variables that both relate to the same "event." The scatter plot graphs pairs of numerical data, one variable on each axis, to look for a relationship between them. The shape of these plots demonstrates at a glance whether a relationship (**correlation**) exists between two sets of data; if the variables are correlated, the points will fall along a line or curve. The better the correlation, the tighter the points will hug the line. Points distributed all over the graph indicate that no relationship exists.

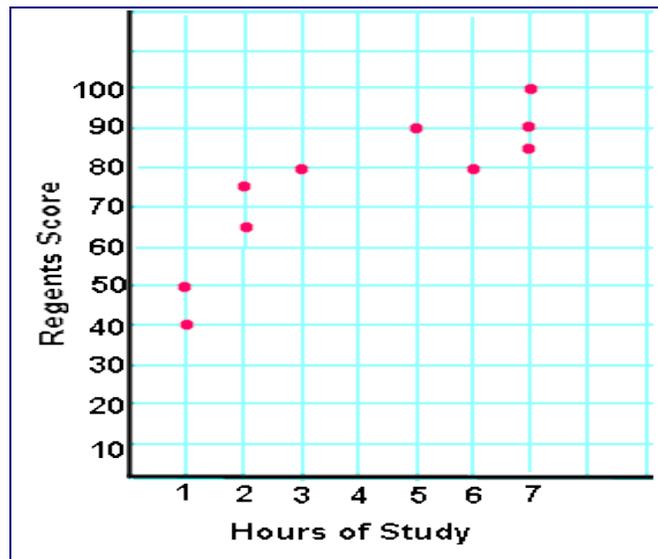
## When to Use a Scatter Diagram

- When you have paired numerical data. Drawing a scatter diagram is often the first step in looking for a relationship between variables
- When trying to determine whether the two variables are related, such as...
  - When trying to identify potential root causes of problems
  - After brainstorming causes and effects using a fishbone diagram, to determine objectively whether a particular cause and effect are related
  - When determining whether two effects that appear to be related both occur with the same cause

## Scatter Plot Example

Note: When making a scatter plot, do NOT connect the dots.

Study Hours	Exam Score
3	80
5	90
2	75
6	80
7	90
1	50
2	65
7	85
1	40
7	100

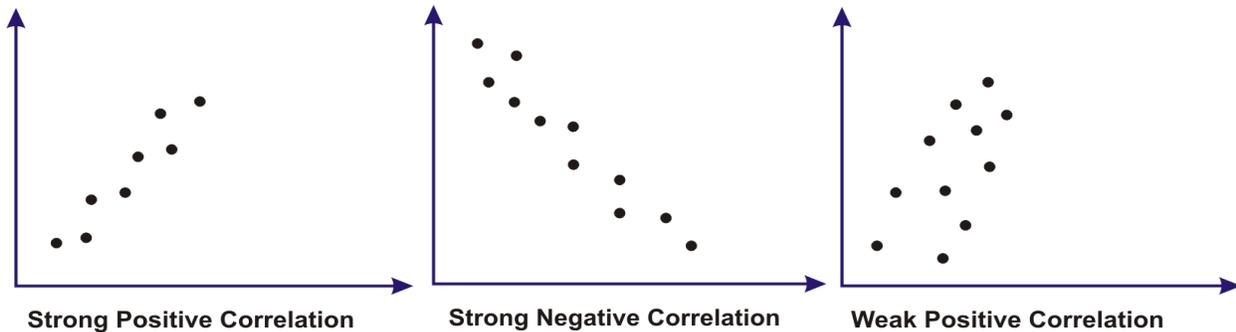


The data displayed on the graph resembles a line rising from left to right. Since the slope of the line is positive, there is a **positive correlation** between the two sets of data. This means that according to this set of data, the longer you study, the better grade you will get on your exam.

If the slope of the line had been negative (falling from left to right), a **negative correlation** would exist since the slope of the line would have been negative. Under a negative correlation, the longer you study, the worse grade you would get on your exam.

If the plot on the graph is scattered in such a way that it does not approximate a line (it does not appear to rise or fall), there is **no correlation** between the sets of data. No correlation means that the data just doesn't show if studying longer has any effect on exam scores.

The following graphs illustrate the various types of correlations:



## Warning!!

### Correlation does not necessarily mean Causation.

Just because there is a strong correlation between data, does not necessarily imply that one set of data is causing the effect that is occurring in the other set of data. Both may be influenced by other factors.