

Pattern recognition of data by scatter plot

Scatter plots are similar to line graphs in that they use horizontal and vertical axes to **plot** data points. However, they have a very specific purpose. **Scatter plots show** how much one variable is affected by another. The relationship between two variables is called their correlation.

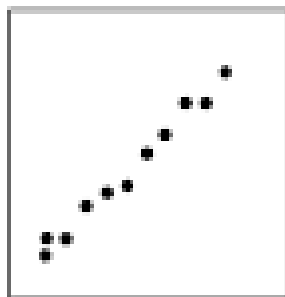
There are three types of correlation: positive, negative, and none (no correlation).

- Positive Correlation: as one variable increases so does the other. ...
- Negative Correlation: as one variable increases, the other decreases. ...
- No Correlation: there is no apparent relationship between the variables.

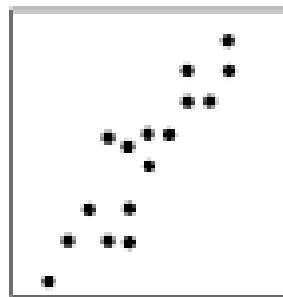
So, we plot data from two variables x and y and then we look at the pattern and see which of the following graph can be the best match then we have an evidence or lack of any relationship.

In the below image the first two from top right suggest a positive linear relationship that we call positive correlation, the top right shows no pattern therefore there is no correlation between the 2 variables

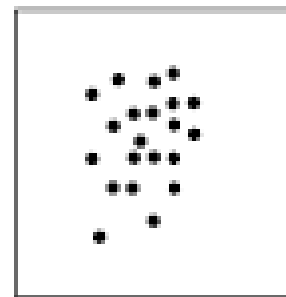
The two graphs from bottom left shows a negative correlation meaning as one increases the other one decreases and finally the bottom right suggests a curve type of relationship first decreases and then increases.



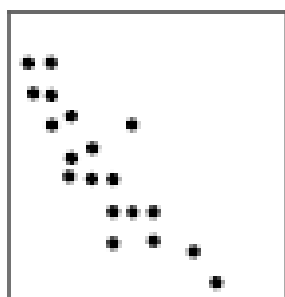
Strong positive correlation



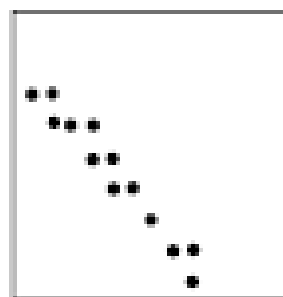
Moderate positive correlation



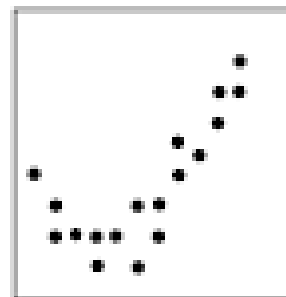
No correlation



Moderate negative correlation



Strong negative correlation



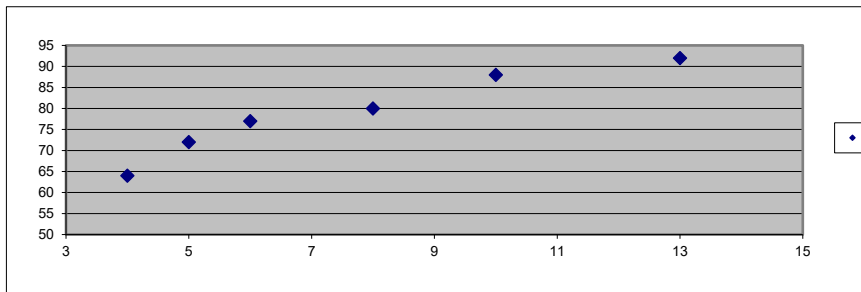
Curvilinear relationship

Example:

	X: hours of study	Y: test scores
1	5	72
2	10	88
3	13	92
4	8	80
5	6	77
6	4	64

Is there a relationship or correlation between hours of study and test scores?

To answer this question, we need to plot scatter plot and look at the pattern.



Conclusion the pattern of data points suggests that yes there is a strong correlation between these two variables meaning the high scores on the tests are tied to more study hours.

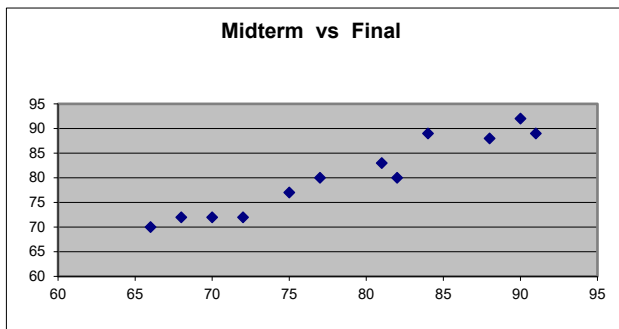
You can do scatter plots by hand or by using this website.

<http://www.endmemo.com/plot/scatterplot.php>

X= Midterm	75	68	82	91	84	77	72	88	90	66	70	81	59
Y= Final	77	72	80	89	89	80	72	88	92	70	72	83	66

Use the above the data and do a scatter plot to see what type of correlation if any exists?

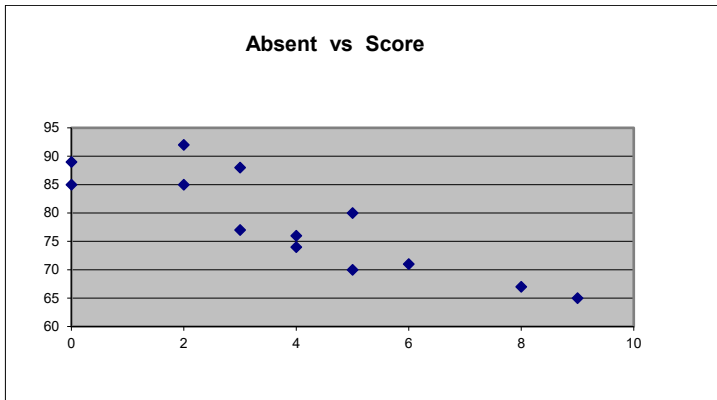
Is a higher score on final is correlated to higher score on the midterm? By looking at the scatter plot the answer is yes.



X = Number of times absent	2	3	5	2	6	0	4	3	9	5	0	4	8
Y = Average test scores	92	88	80	85	71	85	74	77	65	70	89	76	67

Use the above the data and do a scatter plot to see what type of correlation if any exists?

Is a higher score on final is correlated to higher number of absences? By looking at the scatter plot the answer is no, the higher scores on the test comes from fewer number of absences.



As was mention you can do these two scatterplots by hand or by using this website.

<http://www.endmemo.com/plot/scatterplot.php>