

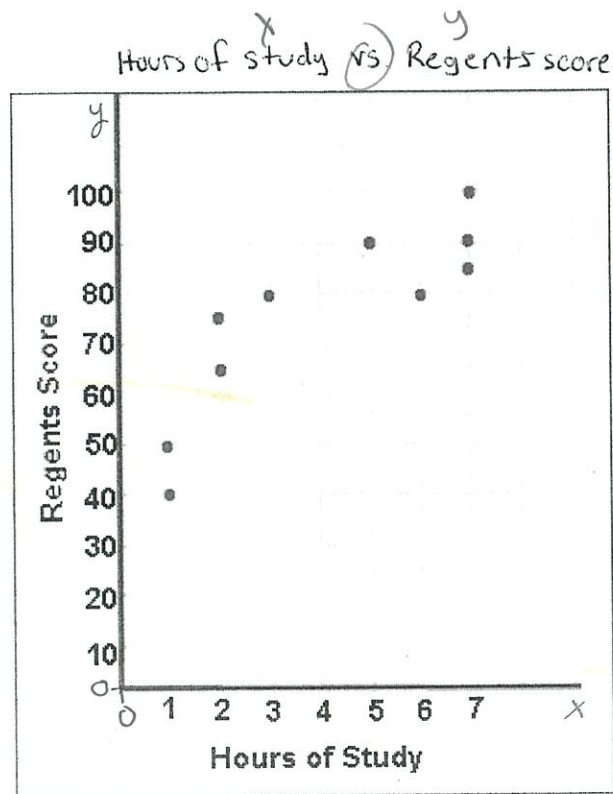
Scatter Plots and Correlation

↓
association/relationship

Have you ever been curious to know if one event affects another event? For example, if I study longer, will I get a better grade on my Regents exam? Statisticians and quality control technicians gather data to determine **correlations** (relationships) between such events. Scatter plots will often show at a glance whether a relationship exists between **bivariate** (two sets) data.

Let's decide if studying longer will affect Regents grades based upon a specific set of data. Given the data below, a scatter plot has been prepared to represent the data. Remember when making a scatter plot, do NOT connect the dots.

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



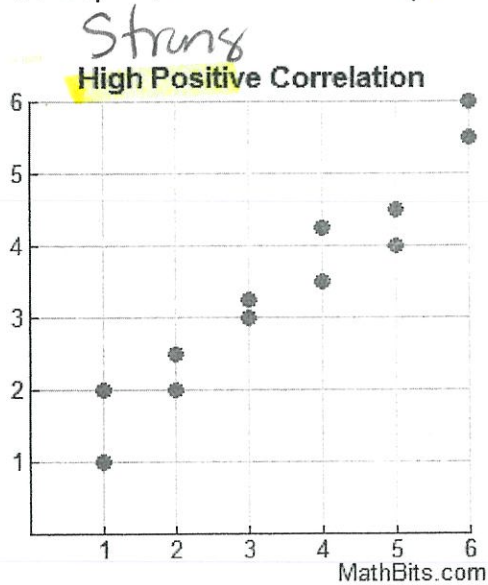
*Positive Correlation
 (Both increase)*

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 I study, the better grade I will get on my Regents examination.** !!

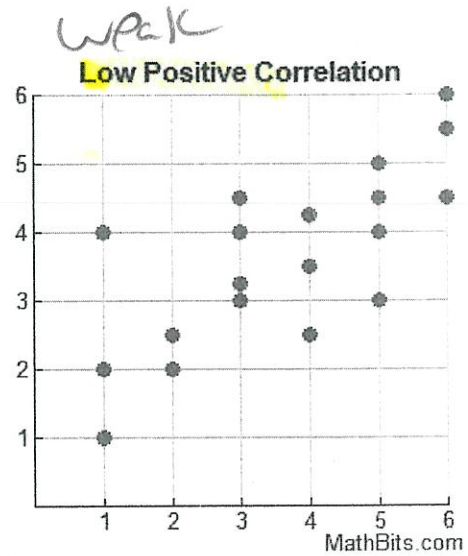
Correlation: Implies a relationship between the data and the study. It does not imply the reason for the relationship. There are three types of correlation. Correlation measures the strength of the linear association between two *quantitative* (number) variables.

Positive Linear Correlation

A positive correlation indicates the extent to which x and y data values increase at the same time. The y values will increase as the x values increase. The graph of such data will resemble a line rising from left to right. The slope of the line will be a positive number.



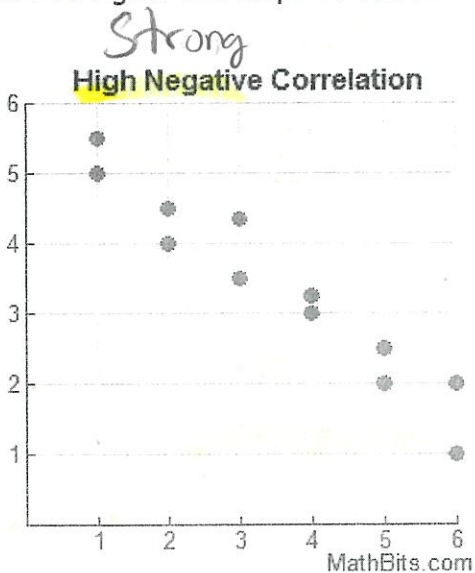
These data points can be described as clustering about a rising straight line with a positive slope. The extent of the positive relationship will be strong.



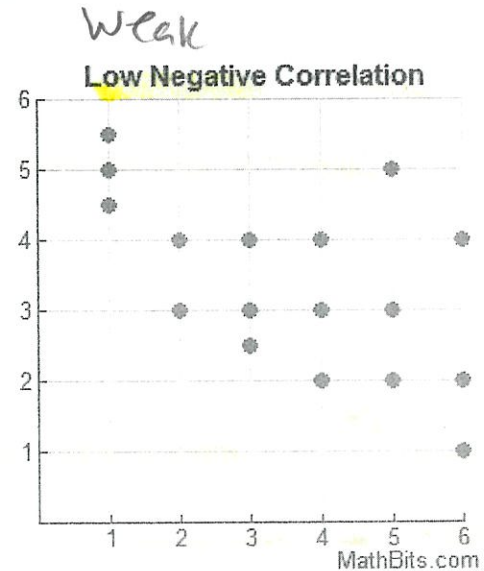
These data points are not clustered to clearly show a straight line. They "tend" to be rising, but the extent of the positive relationship will be less strong (weaker).

Negative Linear Correlation

A negative correlation indicates the extent to which one data value increases as the other decreases. The y values will decrease as the x values increase. The graph of such data will resemble a line falling from left to right. The slope of the line will be a negative number.



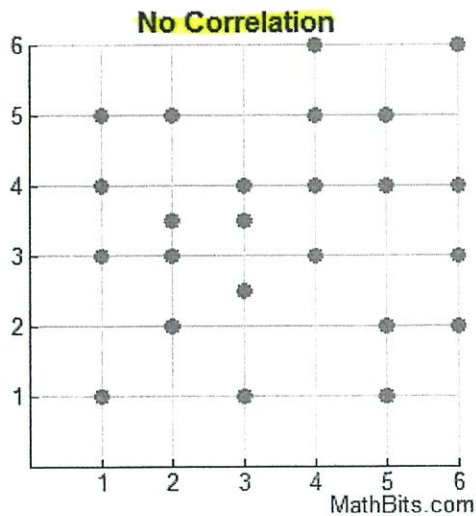
These data points can be described as clustering about a falling straight line with a negative slope. The extent of the negative relationship will be strong.



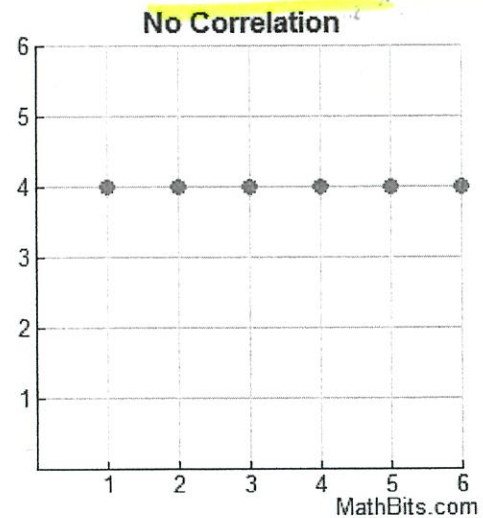
These data points are not clustered to clearly show a straight line. They "tend" to be falling, but the extent of the negative relationship will be less strong (weaker).

No Linear Correlation

If there is **no apparent relationship** between x and y , the data are said to have no correlation. The x and y values are referred to as being independent.



There is no way of knowing from these data points if the pattern is rising or falling. A straight line cannot be found. There is no implication of a relationship.



Be careful here! While a straight line passes through these points, the line is horizontal with a slope of zero (no change). This indicates that the value of x has no influence in changing the value of y .

BEWARE

It is fairly easy to find a situation where a change in one variable appears to predict a similar change in the other variable. When such situations are found, be careful not to assume that the change in one variable **causes** the change in the other variable. Just because there is a strong correlation between data, does not necessarily mean that one set of data is causing the affect that is occurring in the other set of data. **Correlation does not imply "causation"**. **Causation means one variable is causing change in another**. Keep in mind that there may be other factors influencing both variables in a similar manner, or it might simply be a coincidence.

Examples:

1) Which relationship can best be described as causal?

- a) number of students in a class and number of students with brown hair
- b) height and intelligence
- c) shoe size and running speed
- d) number of correct answers on a test and the test score.

2) Which situation describes a situation that is not a causal relationship?

- a) The rooster crows and the sun rises
- b) The more miles driven the more gasoline needed
- c) The more powerful the microwave the faster the food cooks
- d) The faster the pace of the runner the quicker the runner finishes

Graphing Scatter Plots

1) A teacher gave two grades to her students for the first school quarter. The first grade, x , was each student's homework grade and the second grade, y , was their test and quiz average. The data from eight of her students is shown in the chart below.

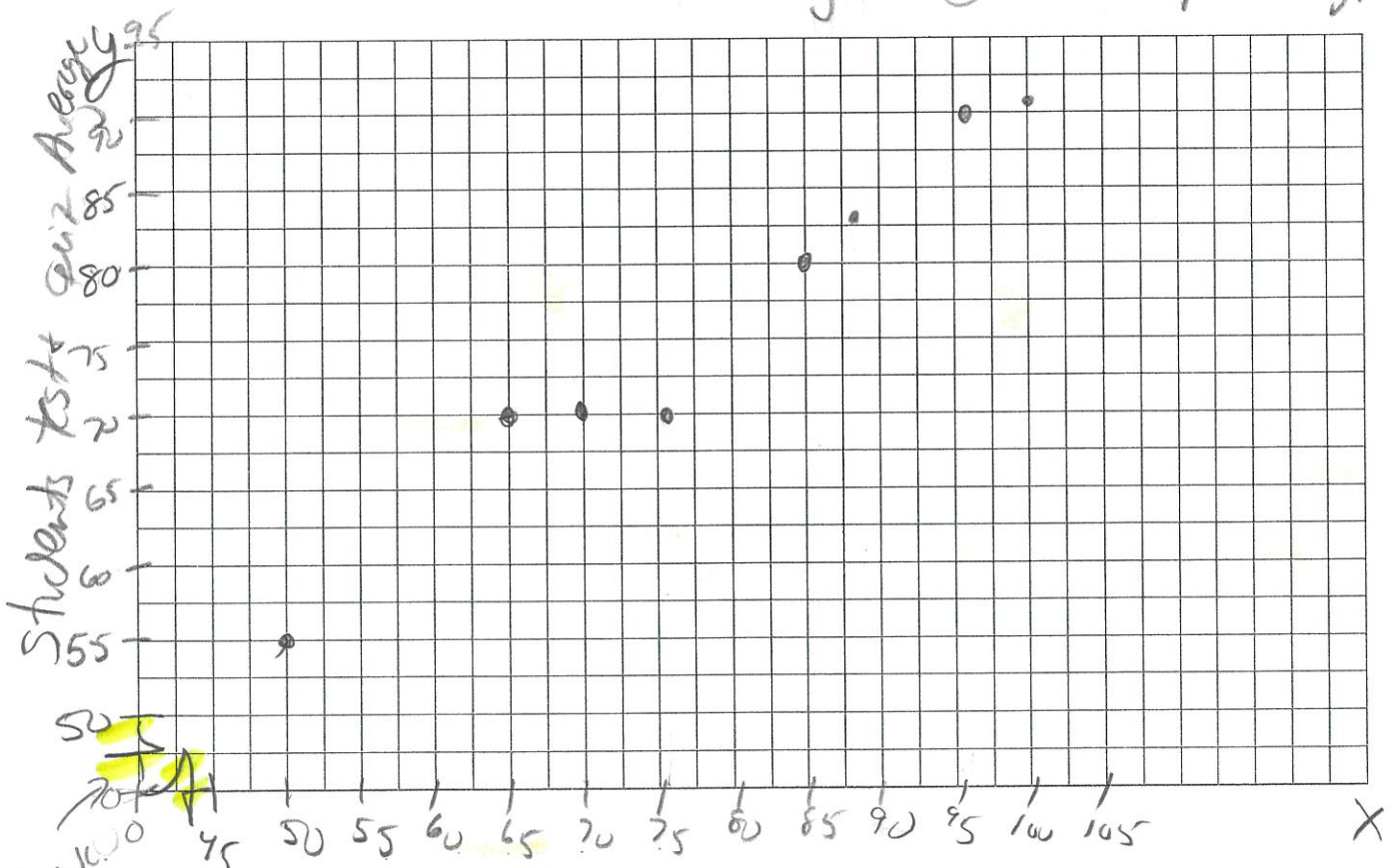
Homework grade \rightarrow	x	70	85	76	50	100	65	96	88
Test & quiz average \rightarrow	y	70	80	70	55	92	70	90	83

Don't connect the dots

a) Use this data to create a scatter plot

b) Determine the correlation *POSITIVE correlation:* As the student's homework grade increases, their test & quiz average also increases.

Students' Homework grade vs. test & quiz Avg.



*Breakdown
Can skip
numbers
when you use
a break*

Student's Homework grade

ATO graph on the calculator: 1) **STAT** **1: Edit** put x -values into L_1 , and y -values into L_2 2) **2nd** **y=** **1** **enter** 3) **ZOOM** **9: ZOOMSTAT** to see graph

2) The table below shows the IQ of 8 senior students and the number of hours of TV each student views per week.

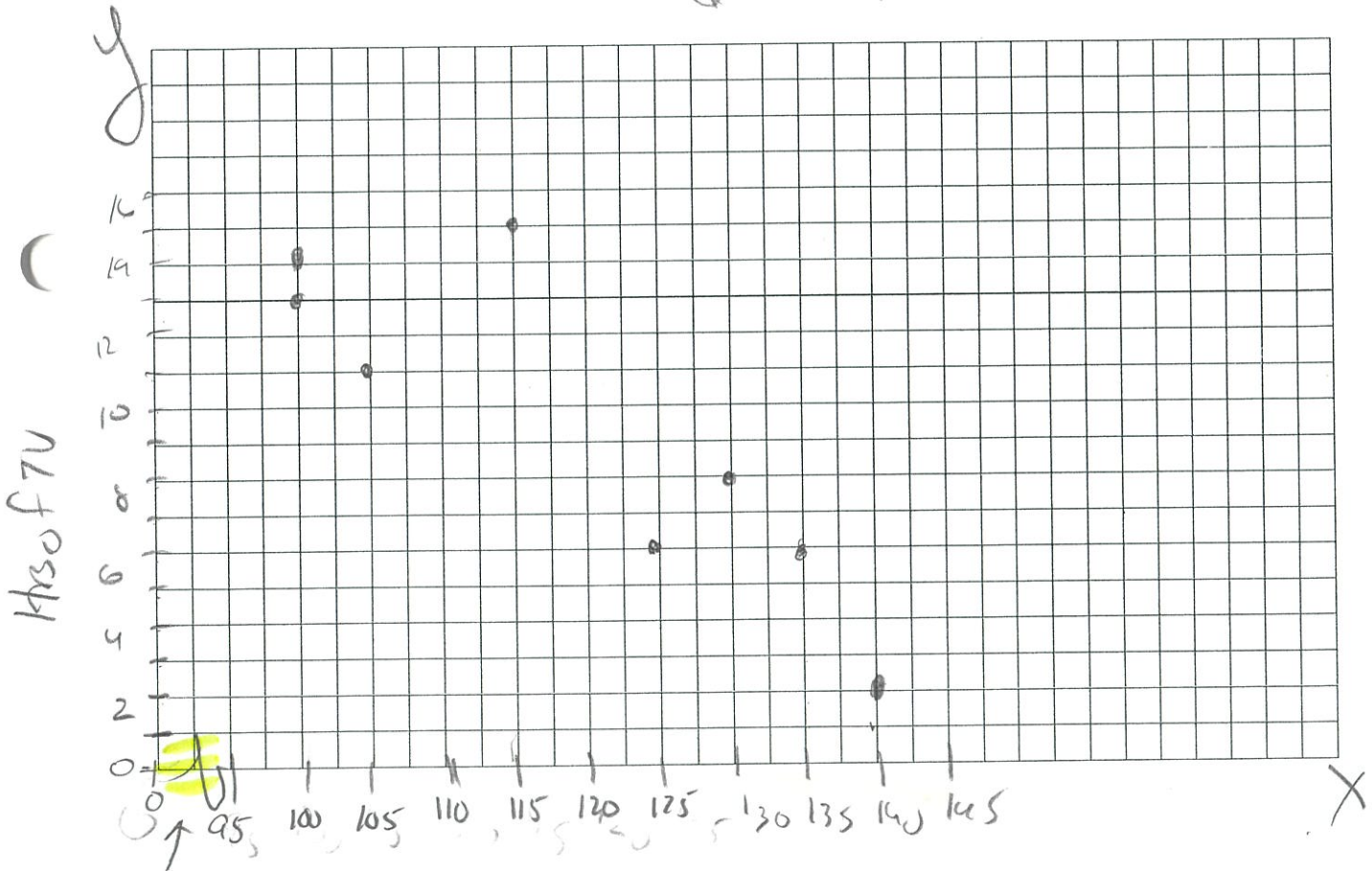
IQ (x)	105	125	135	100	115	130	140	100
Hrs. of TV (y)	11	7	6	13	15	8	2	14

A Dunt correct the dots

a) Use this data to create a scatter plot

b) Determine the correlation *Negative: As the student's IQ increases, the # of hrs of T.V. they view per week decreases.*

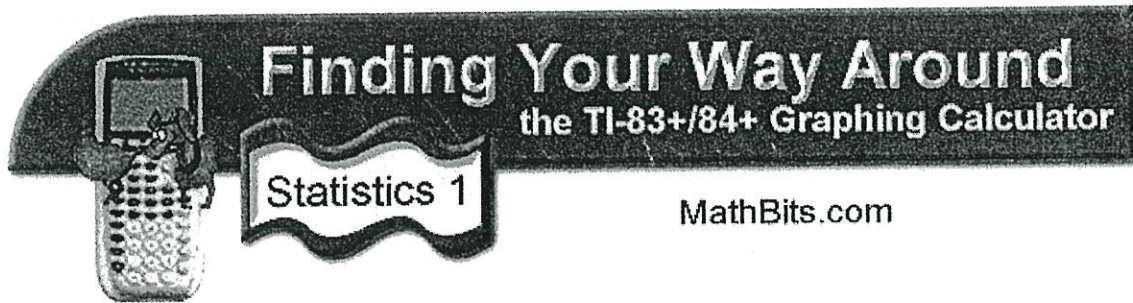
IQ vs. Hrs of TV



Use a

Break to skip #'s

IQ

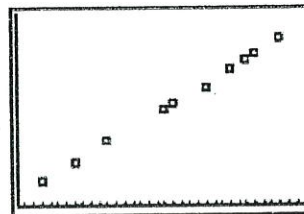


MathBits.com

Scatter Plots

A scatter plot is a graph used to determine whether there is a relationship between paired data.

In many real-life situations, scatter plots follow patterns that are approximately linear. If y tends to increase as x increases, then the paired data are said to be a **positive correlation**. If y tends to decrease as x increases, the paired data are said to be a **negative correlation**. If the points show no linear pattern, the paired data are said to have **relatively no correlation**.

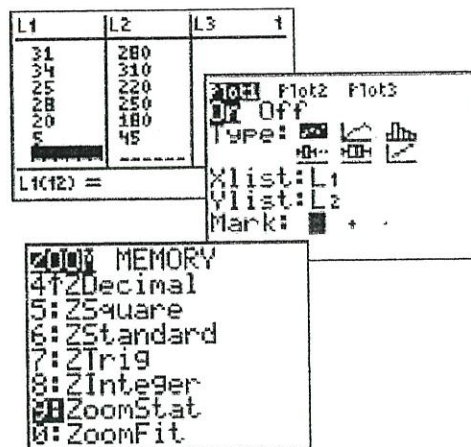


To set up a scatter plot:

Clear (or deactivate) any entries in $Y=$ before you begin.

1. Enter the X data values in L1. Enter the Y data values in L2, being careful that each X data value and its matching Y data value are entered on the same horizontal line.

(See [Basic Commands](#) for entering data.)



2. Activate the scatter plot. Press **2nd** **STATPLOT** and choose **#1 PLOT 1**. You will see the screen at the right. Be sure the plot is **ON**, the scatter plot icon is highlighted, and that the list of the X data values are next to **Xlist**, and the list of the Y data values are next to **Ylist**. Choose any of the three marks.

3. To see the scatter plot, press **ZOOM** and **#9 ZoomStat**. Hitting **TRACE** and right arrow will move along the data points.

4. To turn the scatter plot off, when you are finished with this problem:

Method 1: Go to the $Y=$ screen. Arrow up onto the **PLOT** highlighted at the top of the screen.

Press **ENTER** to turn it off.

Method 2: Go to **STAT PLOT** (above $Y=$). Choose your **PLOT** location. Arrow to **OFF**. Press **ENTER** to turn it off.