

## Lesson 35: Scatter Plots

A **scatter plot** is a graph that plots the values of **two variables as ordered pairs**. You can then analyze the scatter plot by examining how **closely the points come to forming a straight line**.

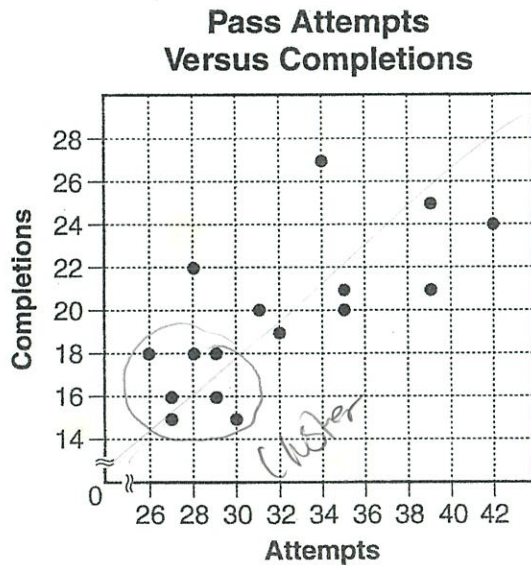
### ▶ Example

The following table shows the number of passes Brett attempted and completed in 16 football games.

**Pass Attempts Versus Completions**

<b>Attempts</b>	28	31	39	35	34	35	27	32	29	26	42	27	39	28	29	30
<b>Completions</b>	22	20	25	20	27	21	16	19	16	18	24	15	21	18	18	15

The following scatter plot displays the data from the table.



*Weak Positive Correlation*

From the scatter plot you can tell that the number of **completions increases as the number of pass attempts increase**.

The closer the data points are to forming a straight line, the greater the correlation between the variables. The grouping of points along the line is called **clustering**. When the data points are not clustered closely together along this imaginary line, there is not a strong association between the data sets.

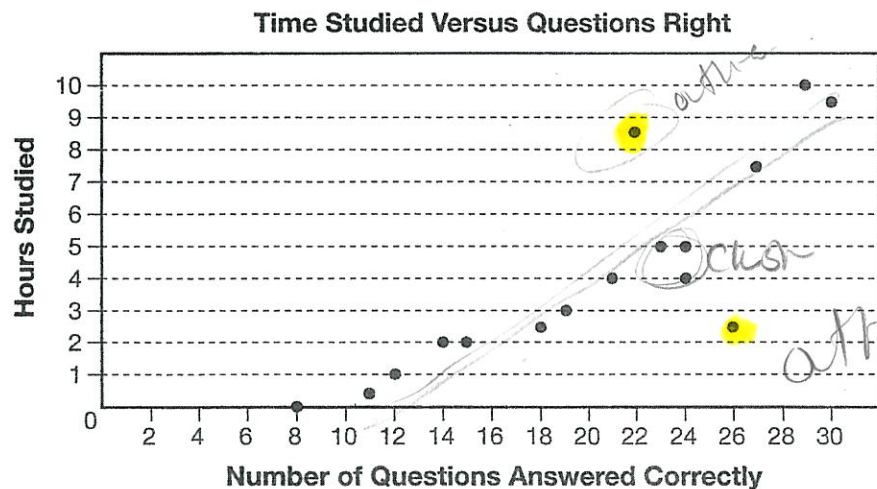
However, even scatter plots with closely clustered data points can have a few points that are far from the line. These points are called **outliers** because they lie outside the general pattern of the data. Plotting data points in a scatter plot helps identify outliers because the outliers will visibly stick out from the other data points.

### Example

The following table and scatter plot shows the number of hours 16 students studied for a test and the number of questions they got right in a 30-question test.

**Time Studied Versus Questions Right**

Questions Right	29	14	12	15	19	24	24	27	11	22	26	8	18	30	23	21
Hours Studied	10	2	1	2	3	4	5	7.5	0.5	8.5	2.5	0	2.5	9.5	5	4



The data points are mostly clustered together in a line that extends from (8, 0) to (30, 9.5). However, there are two points that are far from this line: (22, 8.5) is very high above the line and (26, 2.5) is very low beneath the line on the scatter plot. These are the outliers of the data.



\* Bivariate data - Data involving two different variables (can be measured) which may have a relationship with one another.

### Ex's of Bivariate Data

- 1) - The number of hours a student studies vs. his/her final examination scores.
- 2) - Lowering cholesterol vs. lowering the risk of heart disease.
- 3) - The number of females in Computer Programming vs. their scores in Mathematics

Ex: Which situation should be analyzed using bivariate data?

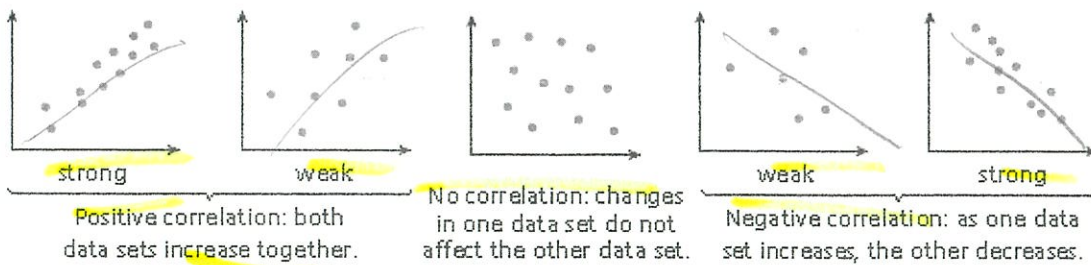
Choose:

- a) Ms. Saleem keeps a list of the amount of time her daughter spends on her social studies homework.
- b) Mr. Benjamin tries to see if his students' shoe sizes are directly related to their heights.
- c) Mr. DeStefan records his customers' best video game scores during the summer.
- d) Mr. Chan keeps track of his daughter's algebra grades for the quarter.

Correlation (aka association) - the relationship between events

Causation: one event causes a 2nd event

ex snow storm causes a day off

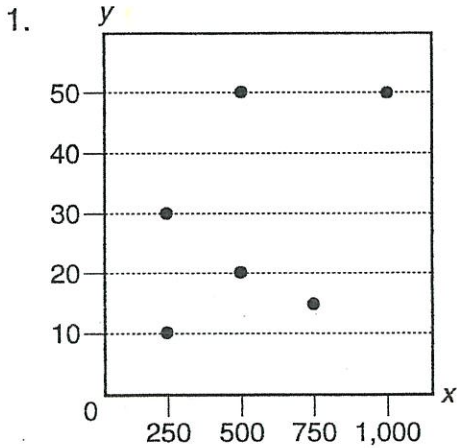


CCSS: 8.SP.1, 8.SP.2

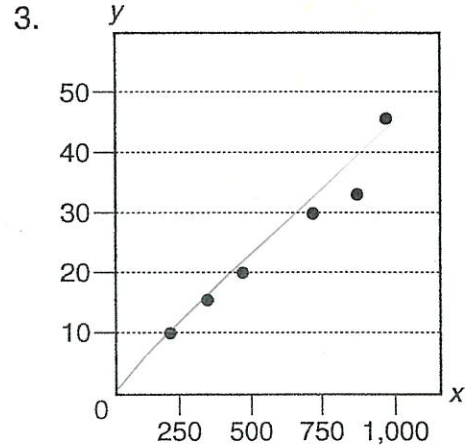
# Practice

Directions: For questions 1 through 4, write whether each scatter plot shows a positive, negative, or no association.

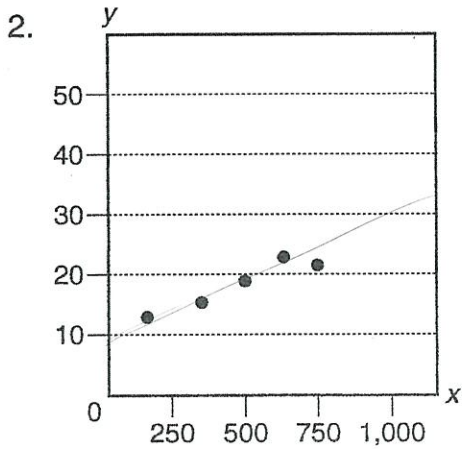
*Correlation (Slope)*



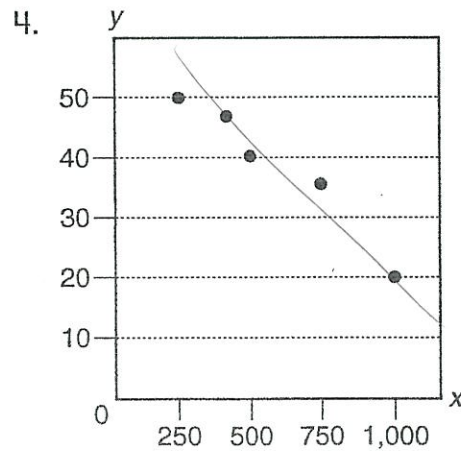
association: NO correlation  
*Correlation*



association: Positive



association: Positive



association: Negative



## Linear and Nonlinear Association *(Correlation)*

Most scatter plots, where the variables have a strong association, follow a straight trend line. These scatter plots have a **linear association** because the ordered pairs of the data sets follow a straight line. However, some scatter plots show an association with a curved trend line. These scatter plots have a **nonlinear association** because the relationship of the data sets does not follow a straight line. Nonlinear associations may be represented by a curved line.

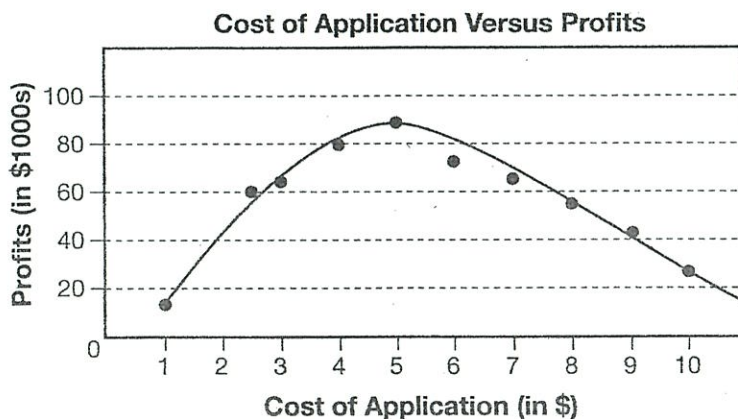
### ▶ Example

A phone application company is deciding how much to charge for its program. The company has decided to charge an amount from \$1 to \$10. The company knows that fewer people will buy the application if it is more expensive. However, it also knows that some customers won't buy a program if it's too cheap. The company predicts the profits based on different prices of the application in the following table.

**Cost of Application Versus Profits**

Cost (in \$)	1	2.5	3	4	5	6	7	8	9	10
Profits (in \$1000s)	15	60	63	80	88	75	64	55	42	24

The following scatter plot shows the data from the table with a trend line.

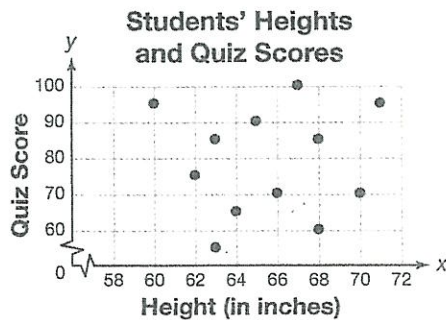


There is an association between the cost of the application and the profits that the company expects to make based on that cost. At first the company makes more money based on the higher cost of the application. But at some point the profits begin to decrease. The trend line forms an upside-down U shape. This is an example of a scatter plot with a **nonlinear association**.

# Practice

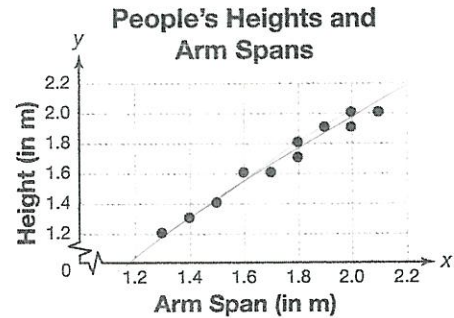
For each scatter plot, describe the association shown as **linear** or **nonlinear**. If no association is shown, state that. If the association is linear, identify it as **positive** or **negative**.

1.



- NO correlation

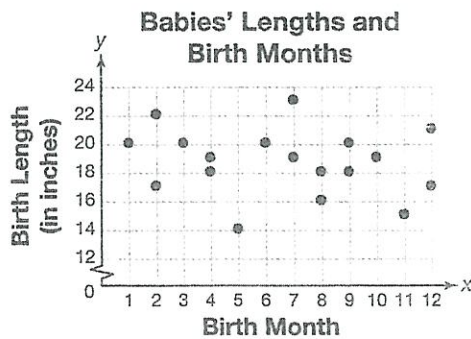
2.



- positive correlation  
- linear

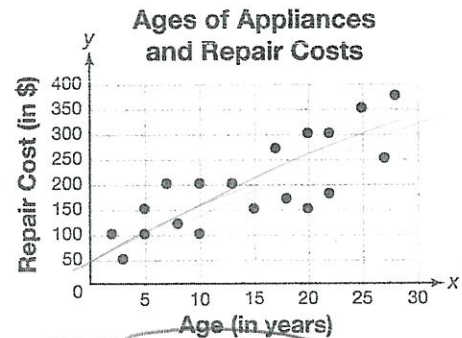
Choose the best answer.

3. The scatter plot shows the lengths of several babies and the numbers representing the months in which they were born. Which best describes the association, if any, that is shown?



- A. positive association  
B. negative association  
**C. no association**  
D. nonlinear association

4. The scatter plot shows the ages of appliances and the costs of repairing them. Which best describes the association, if any, that is shown?



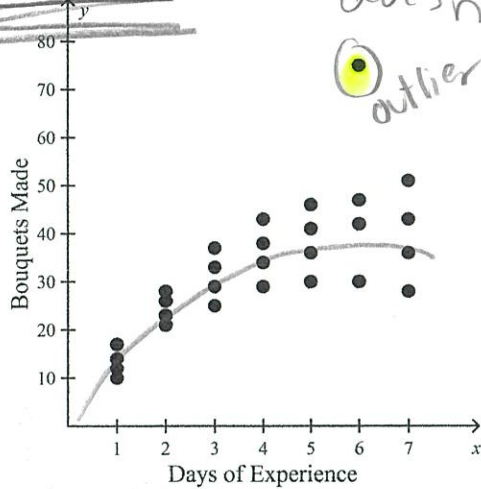
- A. positive association**  
B. negative association  
C. no association  
D. nonlinear association



**Multiple Choice**

Identify the choice that best completes the statement or answers the question.

1. A floral delivery company conducts a study to measure the effect of worker experience on productivity. Tell whether the scatter plot appears to have a linear or non-linear pattern of association. Describe any clustering and identify outliers.



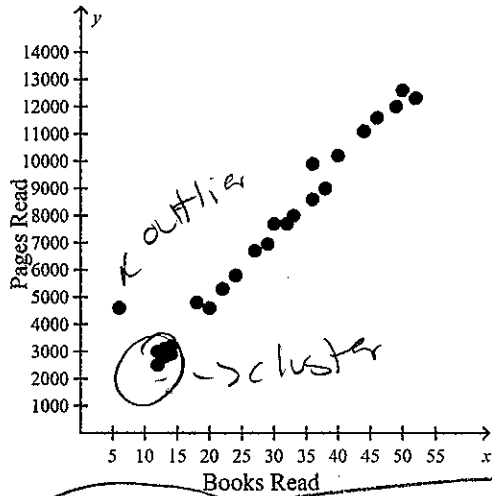
does not belong/not by the others  
outlier

grouping

NON-linear

- a. The pattern of association appears to be linear.  
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.  
There do not appear to be any outliers.
- b. The pattern of association appears to be non-linear.  
There appears to be clustering of the data points at 6 and 7 days. Before that, the results are less clustered.  
There do not appear to be any outliers.
- c. The pattern of association appears to be non-linear.  
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.  
The point near (6, 75) appears to be an outlier.**
- d. The pattern of association appears to be linear.  
There appears to be clustering of the data points at 1 and 2 days. After that, the results become less clustered.  
The point near (6, 75) appears to be an outlier.

2. A book club has 25 members. The scatter plot compares the number of books read to the total number of pages read for each club member. Tell whether the scatter plot appears to have a linear or non-linear pattern of association. Describe any clustering and identify outliers.



- poss. correlation  
- linear

- a. The pattern of association appears to be linear.  
There appears to be clustering of the data points between 12 and 14 books. After that, the results become less clustered.  
The point near (6, 4800) appears to be an outlier.
- b. The pattern of association appears to be linear.  
There appears to be clustering of the data points between 40 and 50 books. After that, the results become less clustered.  
There do not appear to be any outliers.
- c. The pattern of association appears to be non-linear.  
There appears to be clustering of the data points between 12 and 14 books. After that, the results become less clustered.  
The point near (6, 4800) appears to be an outlier.
- d. The pattern of association appears to be non-linear.  
There appears to be clustering of the data points between 12 and 14 books. After that, the results become less clustered.  
There do not appear to be any outliers.