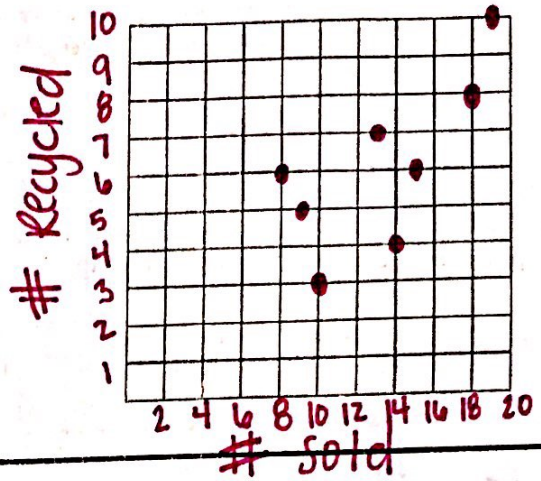
Scatter Plot & Correlation Notes

Name: _____ Date: _____

A **scatter plot** is often used to present **bivariate quantitative** data. Each variable is represented on an axis and the axes are labeled accordingly. It displays data as points on a grid using the associated numbers as coordinates or ordered pairs (x, y). The way the points are arranged by themselves in a scatter plot may or may not suggest a relationship between the two variables.

Make a scatter plot for this data:

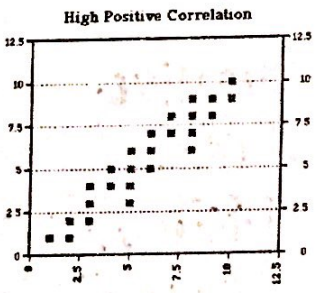
Beverage Can Recycling								
Number of Canned Beverages Sold (x)	18	15	19	8	10	13	9	14
Number of Cans Recycled (y)	8	6	10	6	3	7	5	4

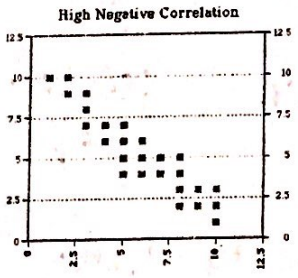


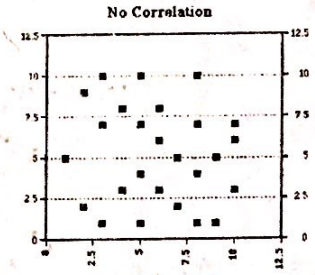
Correlation on a Scatterplot

Correlation can be categorized into 3 types: Positive Correlation, Negative Correlation, or NO Correlation. Within these three categories, correlation can have a range of strength. If correlation is **strong** (in either direction) then data point on a scatterplot will be closer to forming a line, correlation is **not a strong** the data points will be more spread out, but still trending in a positive or negative direction.

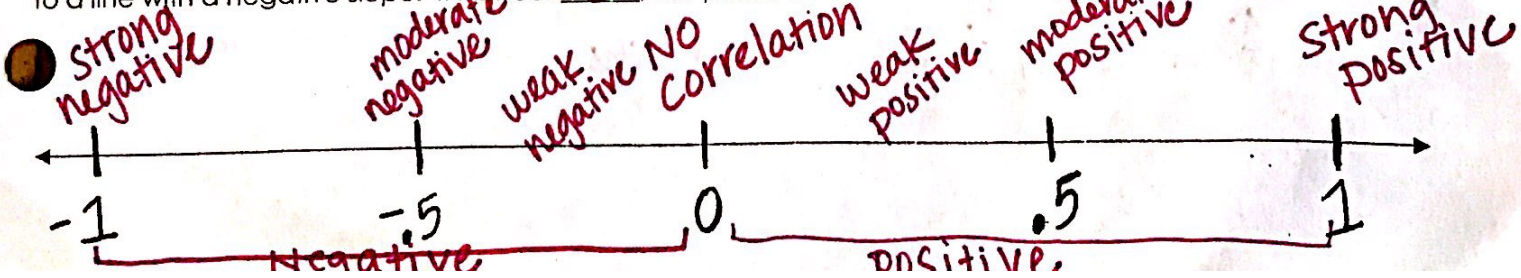
- If y tends to increase as x increases, then the data have Positive correlation.
- If y tends to decrease as x increases, then the data have negative correlation.
- If x and y seem to have no relationship, or the data show no pattern the data has NO correlation.







We use a **correlation coefficient**, denoted by r, to show the strength of correlation. This number can be estimated by looking at a graph or can be found exactly if a set of data is given. The correlation coefficient is a number from -1 to 1 that measures how well a line fits a set of data pairs (x, y). If r is near 1, the points lie close to a line with a positive slope. If r is near -1, the points lie close to a line with a negative slope. If r is near 0, the points do not lie close to any line.



Scanned with CamScanner

Correlation vs. Causation

Sometimes people assume that when there is correlation between two variables this must mean that one variable must cause another variable. When a scatter plot shows a correlation between two variables, even if it's a strong one, there **is *not* necessarily a cause-and-effect relationship**. Both variables could be related to some third variable that actually causes the apparent correlation. Also, an apparent correlation simply could be the result of chance.

For example: There is known correlation between shark attack and ice cream sales. More specifically as shark attacks increase so do ice cream sales. Do you think shark attacks cause people to buy ice cream? If not what do you think the factor that connects these two variables is?

Find the correlation coefficient in the calculator.

a.

Beverage Can Recycling								
Number of Canned Beverages Sold	18	15	19	8	10	13	9	14
Number of Cans Recycled	8	6	10	6	3	7	5	4

r = .71

b.

Hours (x)	1	2	3	4	5	6	7	8
Miles (y)	35	29	26	20	16	9	6	0

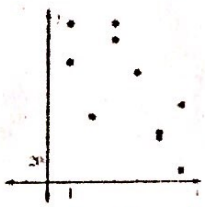
very strong r = -1

r = -.997

Practice Problems:

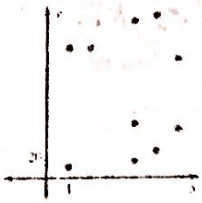
For each scatter plot, tell whether the data have a positive correlation, a negative correlation, or no correlation. Then, tell whether the correlation is closest to -1, -0.5, 0, 0.5, or 1.

1.



negative
r = -.5

2.



No correlation
r = 0

positive
r = 1

negative
r = -1

3. Positive, negative, or no correlation?

a. Amount of exercise and percent of body fat ___negative___

b. A person's age and the number of medical conditions they have ___positive___

c. Temperature and number of ice cream cones sold ___Positive___

d. The number of students at Lassiter and the number of dogs in Atlanta ___NO correlation___

e. Age of a tadpole and the length of its tail ___negative___

↑↑ or ↓↓ = Positive
↑↓ or ↓↑ = Negative

Scanned with CamScanner