

Created by Josie Arteaga  
**WARMUP 12/**  $(11^{1111111111})$

- The manager of an ice cream shop studies its monthly sales figures. He compares the average air temperature with the amount of ice cream they sell. Do you think there would be a **positive, negative, or no correlation**? Explain.
- Do you think this correlation is a result of causation? If so, which variable causes the other? Explain.
- Which pair **best** represents a causation relationship?
  - a child's age and shoe size
  - the number of ice cream cones sold and the amount of sunscreen sold
  - the temperature at a football game and the number of hot drinks sold
  - the number of people attending a ballgame and the length of the ballgame

Read each description. Identify the variables in each situation and determine whether it describes a positive or negative correlation. Explain whether the correlation is a result of causation.

- A group of biologists is studying the population of wolves and the population of deer in a particular region. The biologists compared the populations each month for 2 years. After analyzing the data, the biologists found that as the population of wolves increases, the population of deer decreases.
 

The variables are the population of wolves and the population of deer. The greater population of wolves corresponds to a lower population of deer, so there is a **negative correlation**. Causation is possible but it is unknown which variable causes the other.
- Researchers at an auto insurance company are studying the ages of its policyholders and the number of accidents per 100 policyholders. The researchers compared each year of age from 16 to 65. After analyzing the data, the researchers found that as age increases, the number of accidents per 100 policyholders decreases.
 

The variables are the age in years and the number of accidents per 100 policyholders. The greater age corresponds to a lower number of accidents per 100 policyholders, so there is a **negative correlation**. Causation is possible. It is possible that the age of the driver affects safer driving and therefore fewer accidents because of experience and maturity. However it is also possible that other variables affect safe driving.

15. Educational researchers are investigating the relationship between the number of musical instruments a student plays and a student's grade in math. The researchers conducted a survey asking 110 students the number of musical instruments they play and went to the registrar's office to find the same 110 students' grades in math. The researchers found that students who play a greater number of musical instruments tend to have a greater average grade in math.

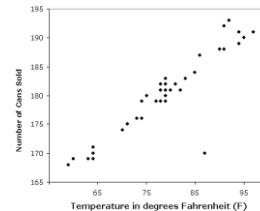
The variables are the number of musical instruments played and the grade in math. The greater number of musical instruments played corresponds to a higher grade in math, so there is a **positive correlation**. Causation is possible, but it is unknown which variable causes the other.

16. Researchers are studying the relationship between the median salary of a police officer in a city and the number of violent crimes per 1000 people. The researchers collected the police officers' median salary and the number of violent crimes per 1000 people in 84 cities. After analyzing the data, researchers found that a city with a greater police officers' median salary tends to have a greater number of violent crimes per 1000 people.

The variables are the police officers' median salary and the number of violent crimes per 1000 people. The greater police officers' median salary corresponds to a higher number of violent crimes per 1000 people, so there is a **positive correlation**. Causation is possible but it is unknown which variable causes the other.

### Clusters and Outliers

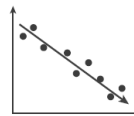
- Cluster – a bunch of points grouped together – indicates common values
- Outliers – values that are far away from the general pattern



### Correlation Coefficient

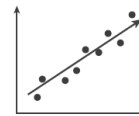
- The correlation coefficient, denoted by  $r$  varies from -1 to 1. It corresponds to the type of correlation.
- Strongly correlated data points have a value of  $r$  closer to 1 or -1.
- Weakly correlated data will have values closer to .5

Strong Negative Correlation



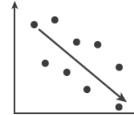
$r$  is close to -1

Strong Positive Correlation



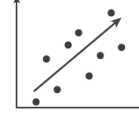
$r$  is close to 1

Weak Negative Correlation



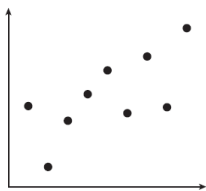
$r$  is close to -.5

Weak Positive Correlation



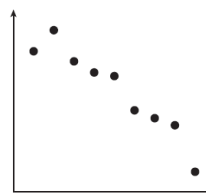
$r$  is close to .5

Estimate the value of the correlation coefficient  $r$



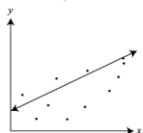
close to .5

Estimate the value of the correlation coefficient  $r$



close to -1

A scatter plot is shown. A student added the line, thinking it was a line of best fit for the data.



Which correlation coefficient **most accurately** describes the data's fit to a linear model?

- A -0.9
- B -0.5
- C 0.5
- D 0.9

C

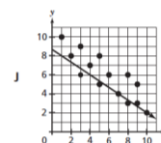
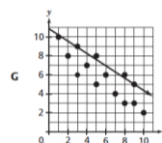
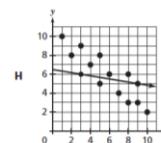
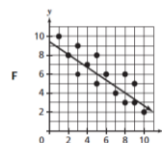
### Line of Best Fit

- A line that goes through the middle of the data
- Should have the same number of dots above and below it

### Line of Best Fit Application

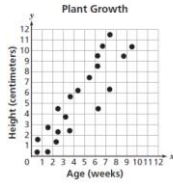
• <http://illuminations.nctm.org/Activity.aspx?id=4186>

Which graph shows the most accurate line of best fit for the given data?



F

The ages and heights of a number of different plants of the same species are recorded on the scatterplot.

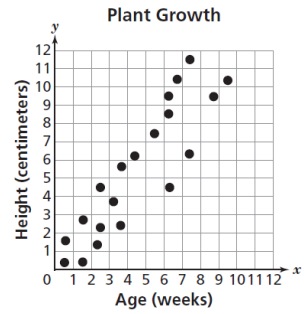


Which equation represents a line of best fit for this scatterplot?

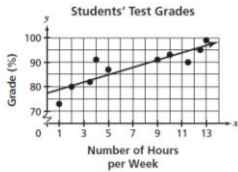
- F  $y = \frac{5}{7}x$
- G  $y = \frac{5}{6}x$
- H  $y = \frac{6}{5}x$
- J  $y = \frac{9}{5}x$

- F  $y = \frac{5}{7}x$
- G  $y = \frac{5}{6}x$
- H  $y = \frac{6}{5}x$
- J  $y = \frac{9}{5}x$

**H**



The scatterplot below shows the relationship between the test grades for 10 students and the numbers of hours they studied per week.



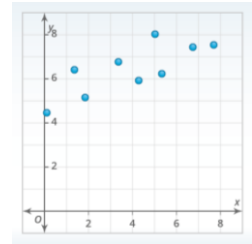
Based on the scatterplot, which is the best prediction of the test grade for a student who studied for 7 hours?

- F 98%
- G 91%
- H 88%
- J 82%

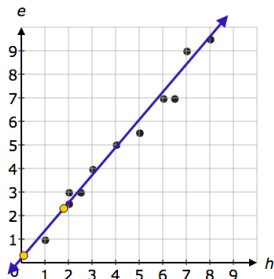
**H**

Which of these is the most appropriate line of best fit?

- A)  $y = x + 4$
- B)  $y = \frac{1}{5}x + 5$
- C)  $y = \frac{1}{3}x + 5$
- D)  $y = \frac{1}{4}x + 6$

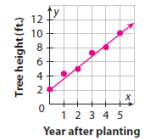


The scatter plot shows the number of eagles,  $e$ , observed during  $h$  hours of observations. Use the grid to graph the line of best fit.



Aoiffé plants a tree sapling in her yard and measures its height every year. Her measurements so far are shown. Make a scatter plot and find a line of fit if the variables have a correlation. What is the equation of your line of fit?

Years after Planting	Height (ft)
0	2.1
1	4.3
2	5
3	7.3
4	8.1
5	10.2



Choosing points: (0, 2) and (5, 10)

$$m = \frac{10 - 2}{5 - 0} = \frac{8}{5} \quad b = 2$$

The equation is of this line of fit is  $y = \frac{8}{5}x + 2$ .



## Homework

- Olympic Scatterplot