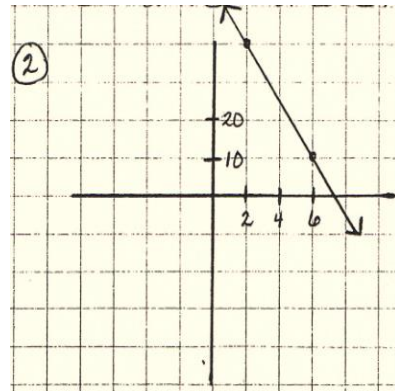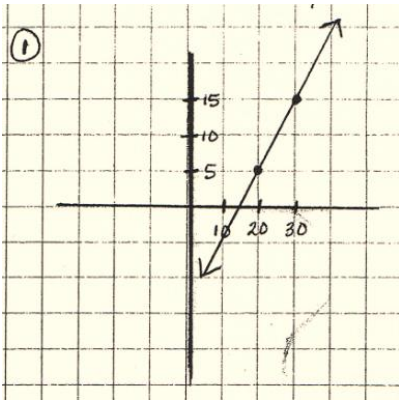# 2.7 Scatterplots, Correlation, and Line of Best Fit

Warm-Up: **Write an equation of the line drawn.**

(Do NOT count the slope – instead, using the two ordered pairs given, use the slope formula to determine the slope, then find b, and then write your equation.)

Vocabulary:

Scatterplot: a graph of a numerical set of data pairs, (x, y)

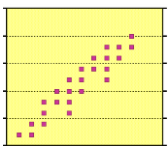**Correlation:** a mutual relationship between two or more things

Best Fit Lines: a line on a scatter plot which can be drawn to model the trend in a set of data
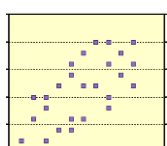   (also known as a "linear regression line")

Types of Correlation:

Positive Correlation	Negative Correlation	No Correlation
• When the data points generally **increase** to the right. • Line of best fit will have a **positive** slope.	• When the data points generally **decrease** to the right. • Line of best fit will have a **negative** slope.	• When the data points generally scatter about with no increase or decrease. • A line of best fit **CANNOT** be written for a graph with no correlation.

Strong positive correlation	Weak positive correlation	Strong negative correlation	Weak negative correlation
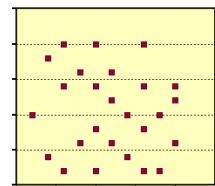
When drawing a line of best fit, the line should mimic the trend in data AND have about half the points above the line and half the points below the line.

Procedures for finding the Equation of the Line of Best Fit:

- **Using a ruler**, draw the line that best represents the data.
- Pick two points **ON THE LINE** (not necessarily data points), and calculate the slope using $m = \dfrac{y_1 - y_2}{x_1 - x_2}$.
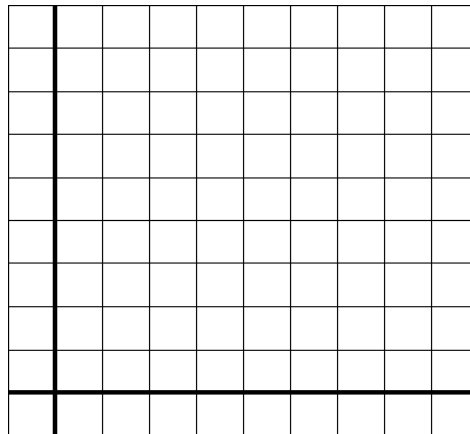- Plug one point and the slope (m) into $y = mx + b$ to find your y-intercept.
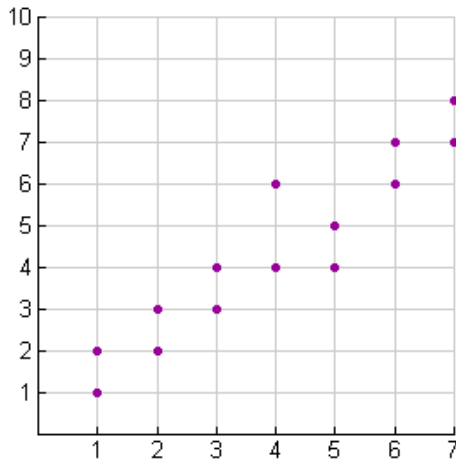- Write your equation and use it to predict values.

Examples:

1) Draw a scatter plot of the data. Then tell whether the data show a *positive correlation*, a *negative correlation*, or *relatively no correlation*.

(1, 7), (1, 5), (2, 3), (3, 2), (3, 6), (5, 5), (6, 4), (6, 8), (7, 6), (8, 2)

2.) Draw the line that best fits the data. Then, write an equation for your line.

3.) You are the student manager of your high school soccer team. You are working on the team's program guide and have recorded the height and weight of the eleven starting players in the table below.

Height (in)	72	70	71	70	69	67	68	73	66	74	76
Weight (lb)	180	170	180	175	160	155	155	180	150	185	200

A. Make a scatter plot of the data. Put height on the x-axis and weight on the y-axis.
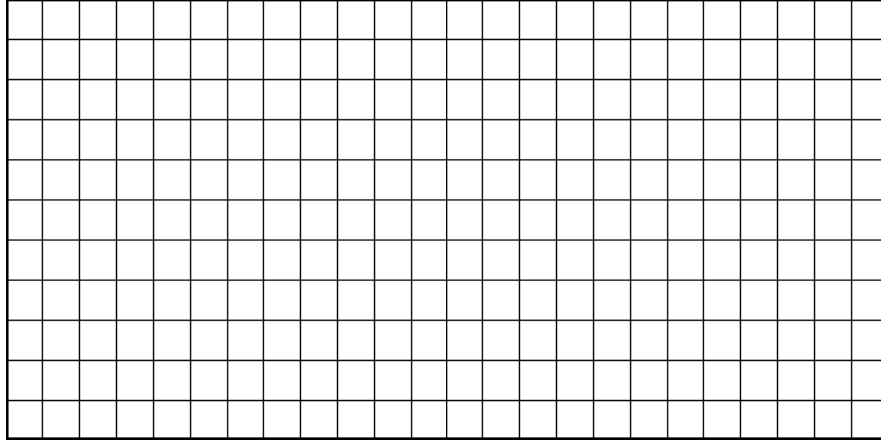
B. Find the equation of the line of best fit for the scatter plot.

C. Use the equation to estimate the weight of a player who is 60 inches tall and of one who is 71 inches tall.

D. In general, how does weight change as height changes?

4.) Shown below are the average values of a specific GMC Yukon.

   A. Make a scatter plot for the data. Put the age of the car on the x-axis and the value on the y-axis.

Age of car (years)	Value (in thousands)
1	65
2	58
3	51
4	45
5	40
6	33
7	25
8	20
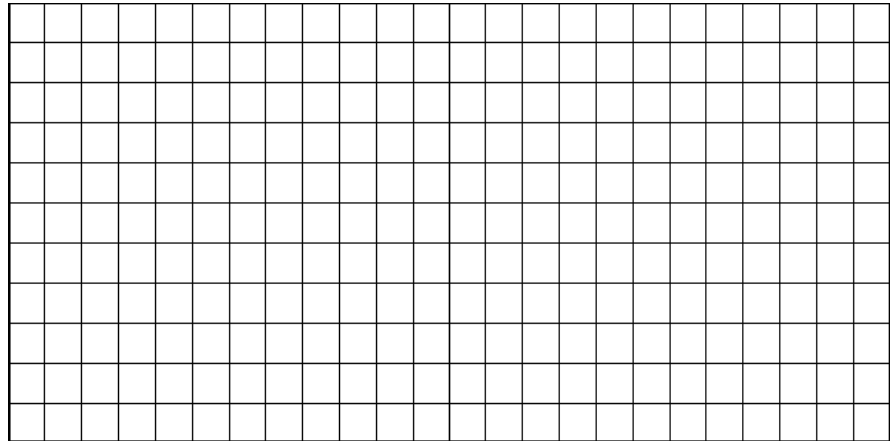
   B. Does the scatter plot show a relationship between the age and the value of the car?

   C. What does the slope represent? What does the y-intercept represent?

   D. Find the equation of the line of best fit for the scatter plot.

   E. Predict when the car is worthless.

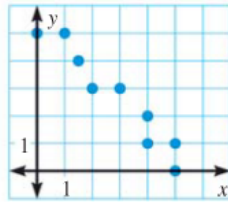**2.** Copy and complete the statement: The line that most closely models the data on a scatter plot is called the __?__ .

3. Does the scatter plot at the right show a *positive correlation*, a *negative correlation*, or *relatively no correlation*? Explain.

Internet Use The table shows estimates of the U.S. annual average total hours of consumer Internet use per person (ages 12 and over) since the year 2000.

Years since 2000, x	0	1	2	3	4	5	6
Hours, y	104	131	147	164	176	183	190

4. Draw a scatter plot of the data.

5. Approximate the best-fitting line for the data.

6. Use your best-fitting line to predict the hours of Internet use in 2010.
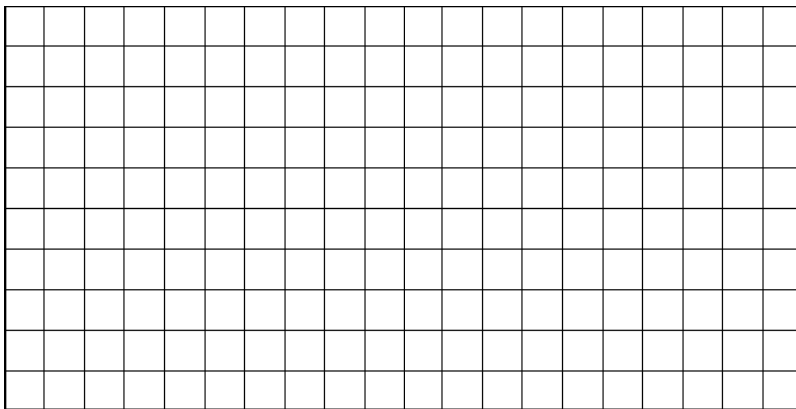
Describing Correlations Tell whether *x* and *y* show a *positive correlation*, a *negative correlation*, or *relatively no correlation*.

7.

8.

9.

Stars In Exercises 12–15, use the scatter plot, which represents stars of various types. The plot compares the magnitude, or brightness, of a star with its surface temperature, which influences the star's color.

Tell whether the stars plotted for the given class show a *positive correlation*, a *negative correlation*, or *relatively no correlation*.

12. blue supergiants

13. giants

14. main sequence

15. white dwarfs

Approximating Best-Fitting Lines Using the graph, approximate the best-fitting line for the data shown by the graph. Then, write the equation for your line of best fit.

18.

19.

31. Multiple Choice Which equation best models the data shown in the scatter plot?

(A) $y = 15$

(B) $y = x + 30$

(C) $y = -x + 30$

(D) $y = -\frac{4}{5}x + 33$