

## Data Display

### Scatter Plots

**Introduction:** There are many ways to organize and display data. A scatter plot is a display used when the purpose is to order a data set to show how one variable is affected by another. A vertical axis is used to display one of the variables and a horizontal axis displays the other.

**Instruction:** A scatter plot displays data that represents a correlation between factors that appear to be related. For instance, the average temperature on a day might affect the number of cold water bottles sold at a vending machine or the experience of a driver might affect the number of automobile accidents he is involved in. If there is a correlation, the data points often make that apparent and a line is sometimes drawn through the middle of the data points to represent the general relationship.

**Example 1:** Display the data from the following chart using a scatter plot.

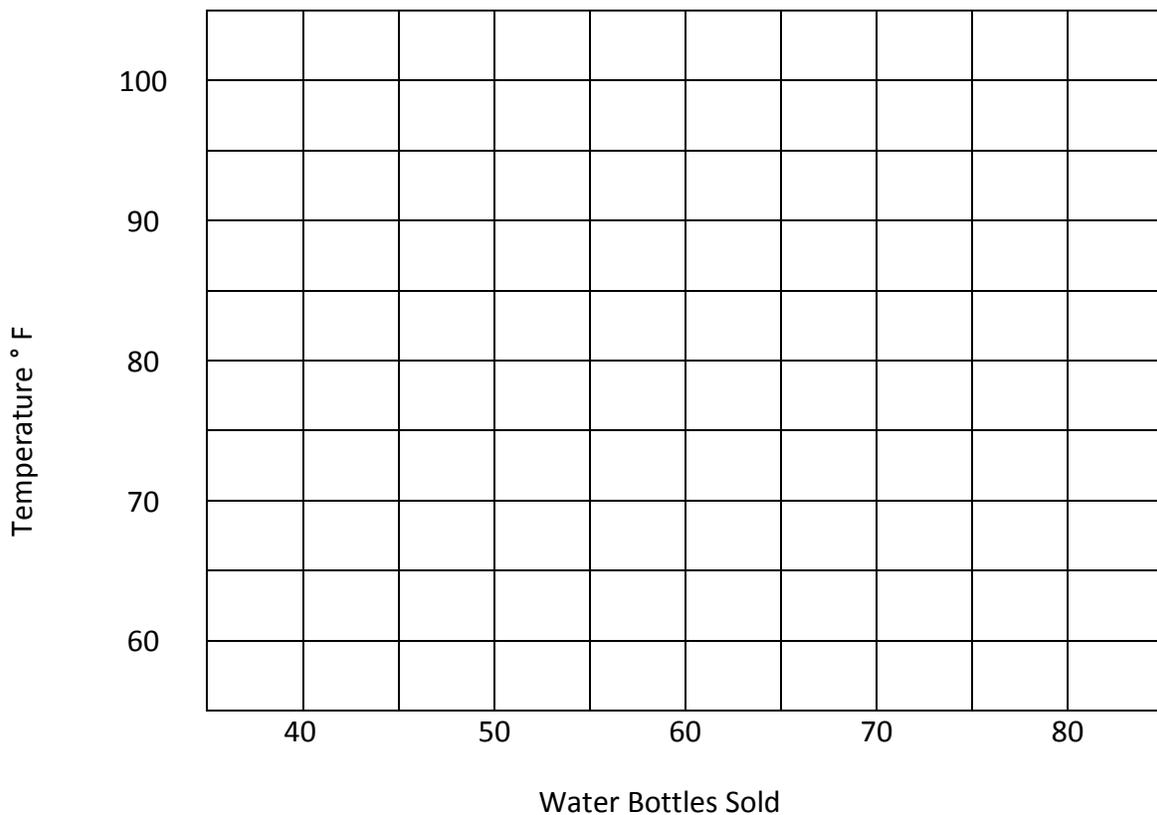
Tom’s family runs a vending machine business. One of their vending machines sells chilled water bottles at a ballpark. Tom wants to know how temperatures affect sales, so he keeps track of the average daily temperature for fourteen days when there is a ball game at the park and the water bottle sales for the same days. His data is entered on the following chart.

Top line = average temperature for the day

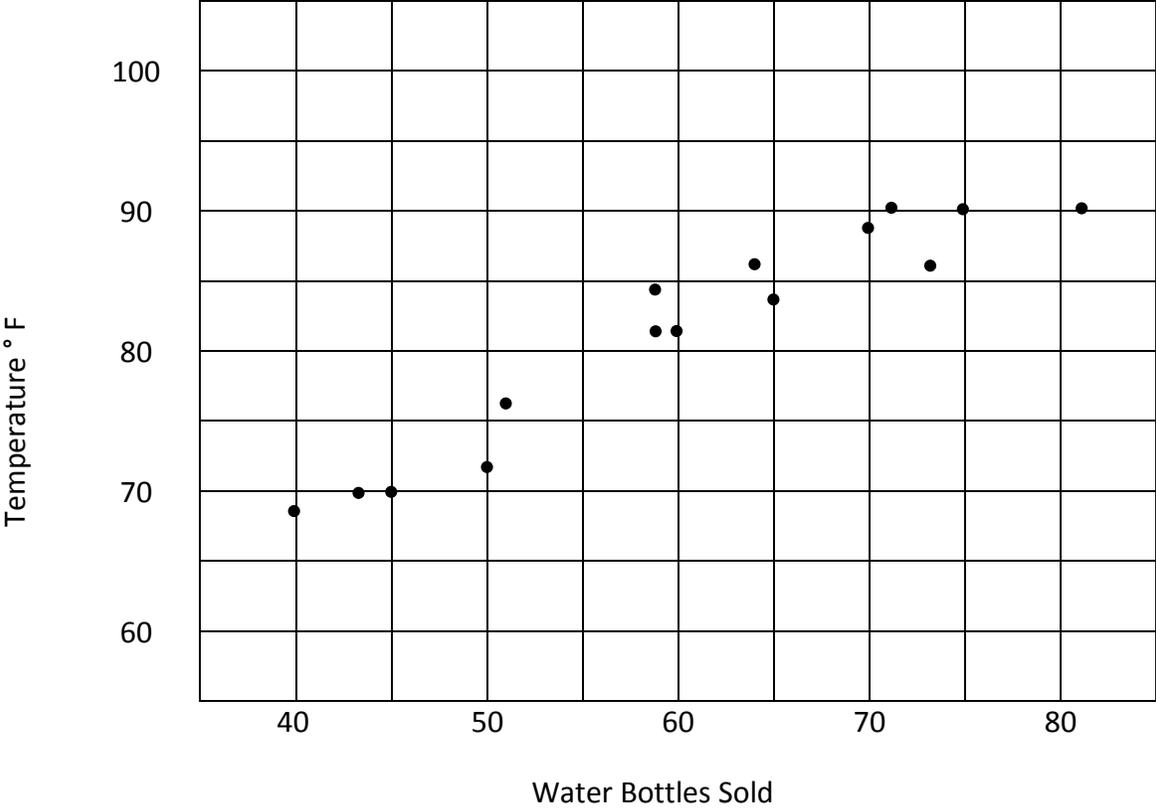
Bottom line = number of water bottles sold on that day

82	84	90	88	86	72	70	68	70	76	82	84	90	90	86
60	58	72	70	64	50	45	40	43	52	58	65	75	81	73

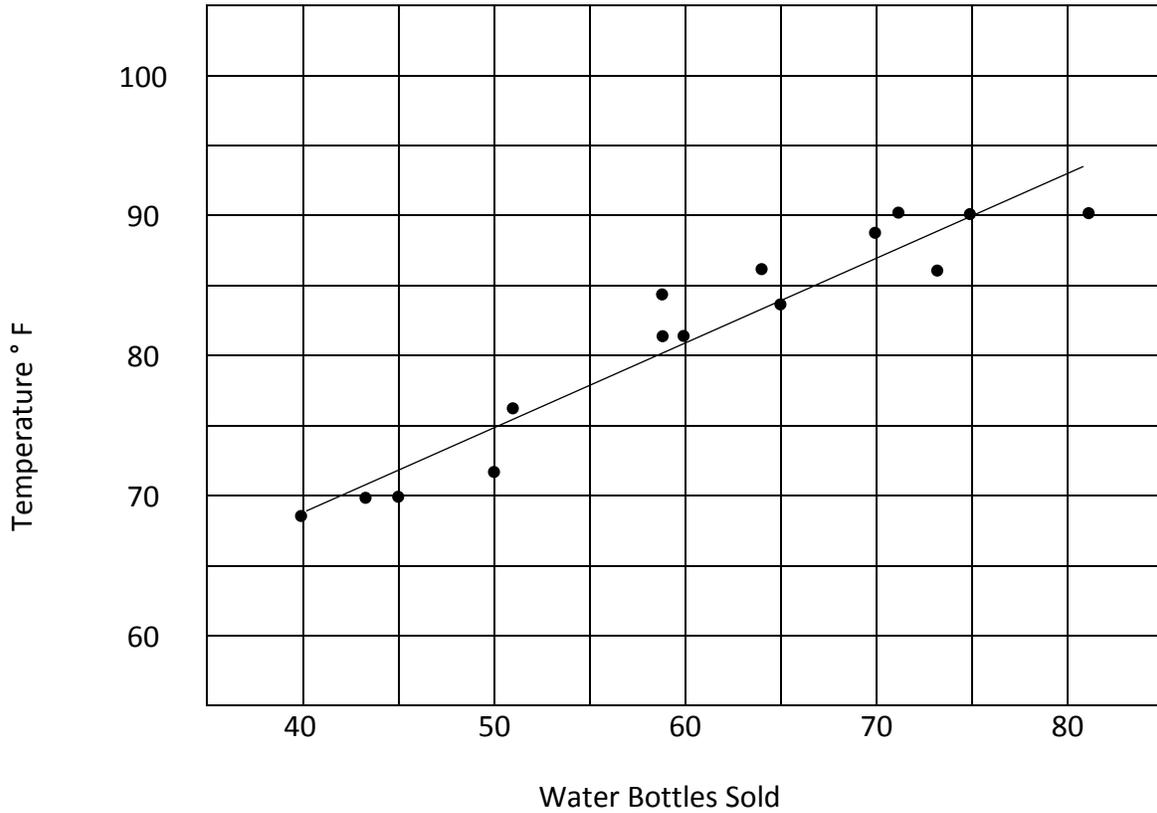
Step 1: Decide which values will be represented by the vertical scale and which values will be represented by the horizontal scale. Decide what increments will be represented on each scale. We’re going to plot temperatures on the vertical scale in 5 degree increments. We’re going to plot water bottle sales on the horizontal scale in units of five.



Step 2: Plot each of the data points from the set by placing a mark on the plot where the temperature and number of water bottles intersect.



Step 3: If desired, draw a line through the middle of the data points to represent the general relationship between the values.



**Q:** What does this scatter plot seem to indicate about the relationship between average daily temperature and sales of cold water bottles?

**A:** Sales increase as temperature increases.

**Example 2:** Display the data from the following chart using a scatter plot.

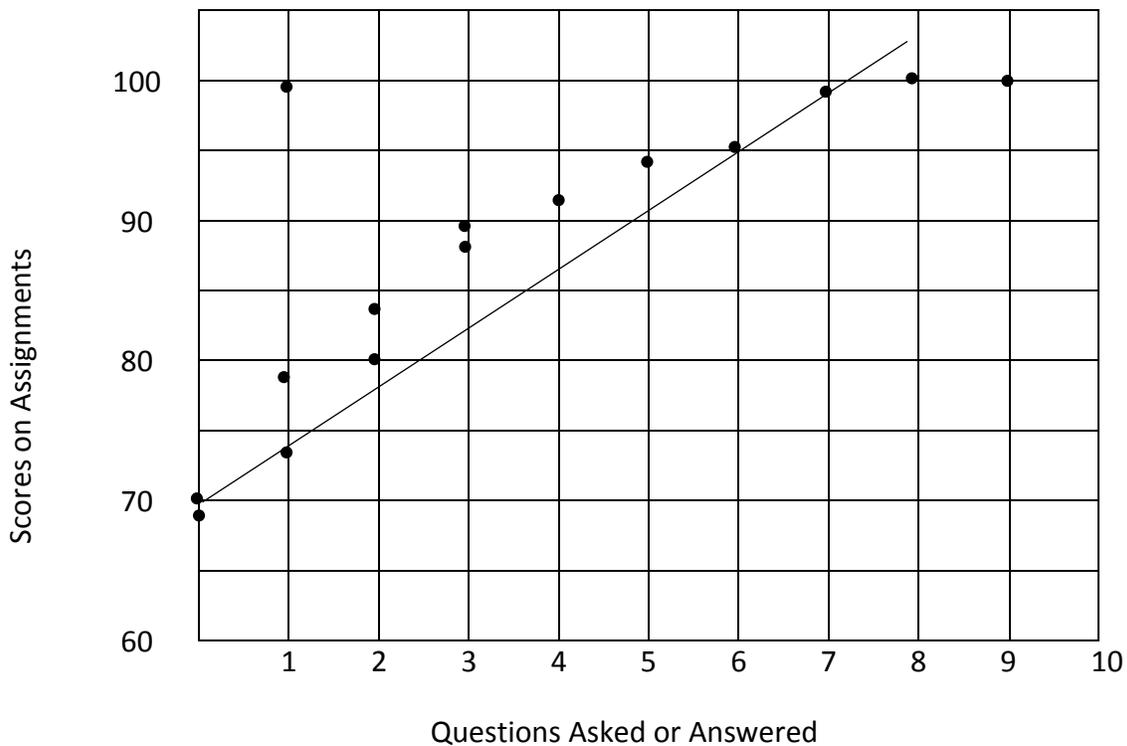
Mrs. Phillips thinks there is a relationship between the number of questions that students ask or answer during a lesson and their scores on assignments. For fifteen days, she selects a different student in her class, keeps track of the number of questions that student asks or answers during the math lesson for the day, and records that student's grade (as a percent correct) for the day's math assignment.

Top line = number of questions asked or answered during the lesson

Bottom line = percent grade on the assignment for the day

5	3	7	1	2	2	0	1	8	6	9	3	0	1	4
94	90	98	100	84	80	70	78	100	95	100	87	68	74	92

Step 1: Decide which values will be represented by the vertical scale and which values will be represented by the horizontal scale. Decide what increments will be represented on each scale. We're going to plot grades on the vertical scale in 5 percent increments. Since all of the grades are above sixty, we're going to start the scale at sixty to save space on the graph. We're going to plot questions on the horizontal scale in units of one.



Step 2: Plot each of the data points from the set by placing a mark on the plot where the number of questions and assignment score intersect.

Step 3: If desired, draw a line through the middle of the data points to represent the general relationship between the values.

**Q:** What does this scatter plot seem to indicate about the relationship between questions asked or answered and scores on assignments?

**A:** Students who participate more tend to get higher grades.

### **Extend the Problem**

**Q:** Is there a data point that doesn't seem to fit with the rest of the data?

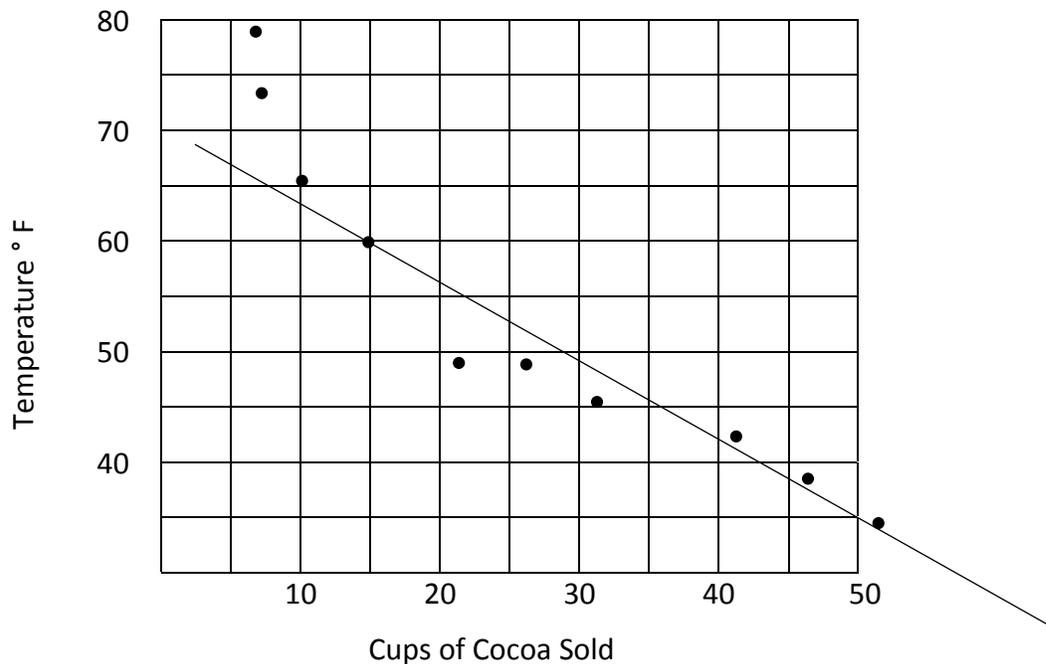
**A:** Yes, the student with only one question but a score of 100. This data point is called an *outlier*. It doesn't seem to fit with the rest of the data. The general trend observed in the data can still be valid even when there are a few outliers.

**Q:** We decided to start our vertical scale at 60 because there were no grades below 60. That's a legitimate choice, but it might make the graph misleading. How might the data on this graph give a different impression if the vertical scale started at 0 instead of 60?

**A:** If the scale started at 0, the differences in grades would not look as large. The smaller scale (60–100) makes the differences look bigger than they would look if the full scale from 0–100 were used. It's important to understand the scale used for a graph, because graph makers sometimes intentionally use a scale that distorts the visual impression created by the graph.

## Interpreting a Scatter Plot

**Example 3:** Consider this scatter plot and answer the questions related to it. The plot represents concession stand sales of hot cocoa at soccer games during September and October. The vertical axis displays temperature at game time and the horizontal axis displays cups of cocoa sold.



**Q:** Did sales of hot cocoa increase or decrease with higher temperatures?

**A:** decrease

**Q:** How many cups of cocoa were sold on the day the temperature was 80 degrees?

**A:** 5

**Q:** Which temperature has more than one data point?

**A:** 50 degrees

**Q:** How many cups of cocoa, altogether, were sold on the days when game time temperature was 50 degrees?

**A:** 45

**Q:** How many cups of cocoa, altogether, were sold on the days when the temperature was above 70 degrees?

**A:** 12 [Since the graph does not clearly mark single units, answers of 11 or 13 are close enough.]

### **Extend the Problem**

**Q:** If the temperature at game time were 45 degrees, how many cups of cocoa do you estimate the concession stand would sell? Explain your answer. Do you think your answer is within 10 cups of actual sales?

**A:** 35—based on the line that represents the general trend of the data or based on the average of the sales when the temperatures were 43 and 47. Answers should be between 30–40: the trend has some variation, but not very much. An answer greater than 40 or less than 30 would not be reasonable based on the data.

**Q:** If the temperature dropped below 30 degrees, do you think the concession stand would sell more than 50 cups of cocoa or fewer than 50 cups of cocoa?

**A:** More than 50 cups, based on the trend in sales and the line that represents the trend.