

Types of Studies

There are many different ways of collecting data for analysis. Three common studies that can be used to collect data are survey, observation, and experiment.

A **survey** consists of asking people questions and recording their responses. For example, you might take a survey to determine the brand of toothpaste each student in your school uses.

An **observation** consists of recognizing and noticing a fact or an occurrence, without finding exact numbers. For example, you might make an observation to determine what color car drives by your house the most.

An **experiment** consists of performing a test or trial and recording your results. For example, you might perform an experiment to determine the number of times heads lands up if you flip a coin 100 times.

Classifying Data

Quantitative data include any information that can be collected that is numerical. For example, if you record the number that faces upward when rolling a number cube 20 times, you have a set of quantitative data.

Qualitative data include any information that can be collected that is not numerical. For example, if you record the month in which each of your classmates was born, you have a set of qualitative data.

Univariate data consist of samples or measurements that involve a single quantitative variable. For example, if you recorded the amount of rain that fell each month last year, you have a set of univariate data.

Bivariate data consist of paired samples or measurements that involve two quantitative variables. For example, if you recorded the times **and** the distances a person ran each day last week, you have a set of bivariate data.

Collecting Data

Random Sampling: Each member of the population is given an equal chance of being selected. The members are chosen independently of one another.

Cluster Sampling: The members of the population are chosen at random from a particular part of the population and are then polled in clusters.

Convenience Sampling: The population is chosen only because it is readily available.

Systematic Sampling: After the population has been ordered in some way, its members are chosen according to a pattern.

RETEACHING 1-1

COLLECT AND DISPLAY DATA

Taking a sample is a means of collecting information, or **data**, from a large group of people. Some ways of sampling an entire group, or population, are **random sampling**, **cluster sampling**, **systematic sampling**, and **convenience sampling**.

Example

To determine public opinion concerning an Arizona state senator, every tenth person visiting the Grand Canyon in Arizona was interviewed. What kind of sampling does this represent? Name a disadvantage of this kind of sampling for this situation.

Solution

convenience sampling—Many visitors to the Grand Canyon in Arizona are not Arizona residents. It would be better to interview Arizona residents whose names were chosen from voter lists in several Arizona cities.

EXERCISES

Is the sample appropriate for each situation? If not, what is its disadvantage? Suggest a better way of choosing the sample.

- | | |
|--|---|
| <p>1. To determine whether people like a new brand of bran muffin, ask every person who tastes a free sample at a health-food exposition.</p> <p>_____</p> <p>_____</p> <p>_____</p> | <p>2. To determine the quality of ball-point pens delivered to a school-supply store, refer to company records showing 4 out of every 500 pens were found defective.</p> <p>_____</p> <p>_____</p> <p>_____</p> |
| <p>3. To determine the average amount spent at a store in a day, poll everyone who leaves the store between 5:00 and 7:00 p.m.</p> <p>_____</p> <p>_____</p> <p>_____</p> | <p>4. To determine the proportion of students in your school who own cars, ask everyone in your math class.</p> <p>_____</p> <p>_____</p> <p>_____</p> |
| <p>5. To determine the "hottest" fall fashion color, count the number of garments of each color on racks in department stores in different parts of the city.</p> <p>_____</p> <p>_____</p> <p>_____</p> | <p>6. To determine whether town residents think there is too much traffic on local Route 44, ask all the homeowners who live along the route.</p> <p>_____</p> <p>_____</p> <p>_____</p> |

Mean, Median and Mode

Name _____

Where does one go to learn Spanish?

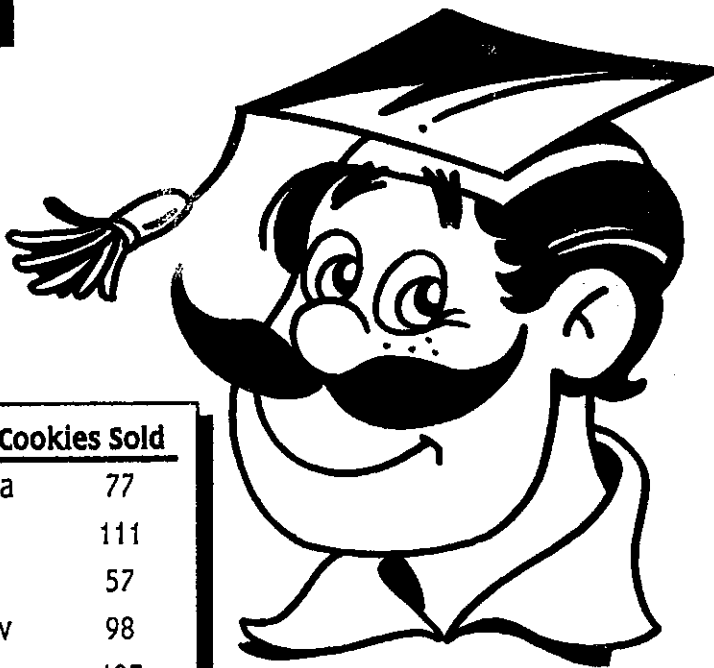
To find out, use the tables to answer the questions below. Then, put the corresponding letter above the answer at the bottom of the page.

Chuck's Test Scores

math	88
Spanish	73
art	94
science	84
history	73
drama	62

Boxes of Cookies Sold

Amanda	77
Joe	111
Jill	57
Andrew	98
Molly	107
Josh	44
Terry	35
Brad	107
Sarah	66



Birth Months of Students at WJHS

January	89
February	93
March	51
April	64
May	91
June	103
July	46
August	64
September	82
October	123
November	112
December	99

- I. What is the median of boxes of cookies sold?
- O. What is the mode of Chuck's scores?
- E. What is the mean of students born each month?
- H. What is the mode of cookies sold?
- R. What is the mean of Chuck's test scores?
- H. What is the mode of students born each month?
- S. What is the median of Chuck's scores?
- G. What is the median of students born each month?
- N. What is the mean of boxes of cookies sold?

78.5 84.75 78 73 79

64 77 90 107

Name _____

Missing data (averages)

1. Jacob's scores on 4 tests were 83, 86, 80, and 81. He has one more test to take, and he wants to earn an 85 average in the class. What score must he get to have an 85 average?

2. Money earned weekly for five weeks include: \$18, \$23, \$17, \$10, and \$16. Want to average \$20 per week. What do you need to earn during the sixth week to maintain an average of \$20 per week?

3. Mrs. Mesh lost a test in her class of 25 students. The mean score for the class was 83 and the sum of the other scores was 1980. What the was grade on the missing paper?

Find the missing number.

4. Mean = 12 9, 12, 15, 21, 4, _____

5. Mean = 62 30, 90, 60, 40, _____

6. Mean = 68.8 88, 56, 77, 32, _____

7. Mean = 15.75 13.8, 14.2, 19.6, _____

8. Mean = 69.25% 97%, 32%, 84%, _____

RETEACHING 1-3

STEM-AND-LEAF PLOTS

A stem-and-leaf plot can help you organize data so it can be easily analyzed.

Example

Make a stem-and-leaf plot for the high temperatures listed in the chart. Then write a description of the data.

MONDAY'S TEMPERATURES					
	High	Low		High	Low
Anchorage	70	54	Miami Beach	88	73
Boston	79	73	New York	89	73
Chicago	99	74	Richmond	97	77
Denver	84	60	St Ste Marie	75	56
Honolulu	87	75	Seattle	82	59
Houston	89	74	Washington	97	79
Los Angeles	76	63			

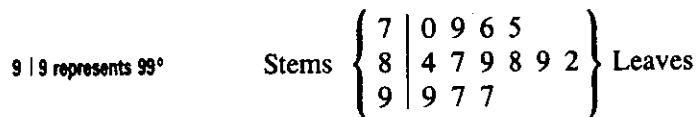
Solution

Step 1: Form the *stem*.

The high temperatures range from 70 to 99. Use the digits in the tens place, 7, 8, and 9, as the stems. Write them in a column. Draw a vertical line to the right.

Step 2: Form the *leaves*.

Show the first high temperature, 70° for Anchorage, by writing a "leaf," 0, next to the "stem," 7. To show 79° for Boston, write 9 next to the 0. Enter the rest of the leaves in the same way.



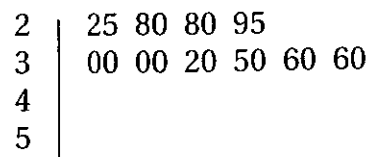
Step 3: The 70 is an *outlier* (extremely high or low value). *Clusters* (groups of values close to one another) appear in the high 80s and high 90s. There is a large *gap* (space between values) between 89 and 97.

EXERCISES

1. On a separate sheet of paper, make a stem-and-leaf plot for the low temperatures.
2. Write a description of the data noting any outliers, clusters, or gaps.

3. Complete the stem-and-leaf plot for advertised monthly rents for one-bedroom apartments. Use the numbers in the hundreds as the stem.

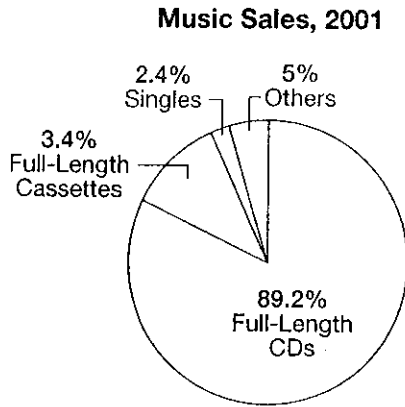
Rents for 1-Bedroom Apartments			
\$400	\$400	\$425	\$225
\$295	\$360	\$300	\$280
\$435	\$360	\$350	\$400
\$350	\$280	\$550	\$375
\$400	\$415	\$325	\$395
\$320	\$300	\$380	\$470



4. On a separate sheet of paper, describe the data to a person who is looking for a one-bedroom apartment.

EXTRA PRACTICE **1-4****CIRCLE GRAPHS**

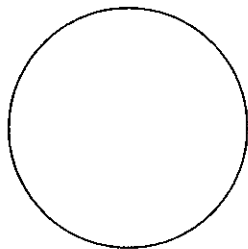
MUSIC For Exercises 1 and 2, use the circle graph below that shows data about music sales in 2001.



1. What angle corresponds to the sector labeled "Others" in the circle graph? Explain how you found your answer.

3. Explain how a circle graph could help you visualize the data in the table.

5. Draw a circle graph to represent the data.



INVESTMENTS For Exercises 3–6, use the table below that shows how Mr. Broussard has invested his money.

Investments	
Savings Account	\$60,000
Money Market Account	\$100,000
Mutual Funds	\$140,000
Stocks	\$500,000
Bonds	\$200,000

2. Use the circle graph to describe music sales in 2001.

4. Determine the percent of Mr. Broussard's total investments that each type of investment represents.

6. Use the circle graph you made in Exercise 5 to describe Mr. Broussard's investments.

RETEACHING 1-6

BAR GRAPHS AND LINE GRAPHS

A **bar graph** can be used to display data that you find in a list, see in a report, or read in a newspaper or book.

Example 1

Use the bar graph to answer the following questions.

- a. In which years did more than 120 banks fail?
- b. About how many more banks failed in 1993 than in 1994?

Solution

- a. 1991 and 1992
- b. $40 - 10 = 30$

To display data on a line graph points are plotted and connected in order.

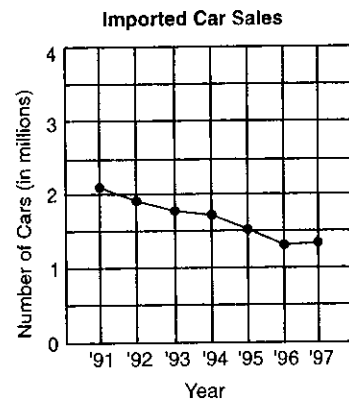
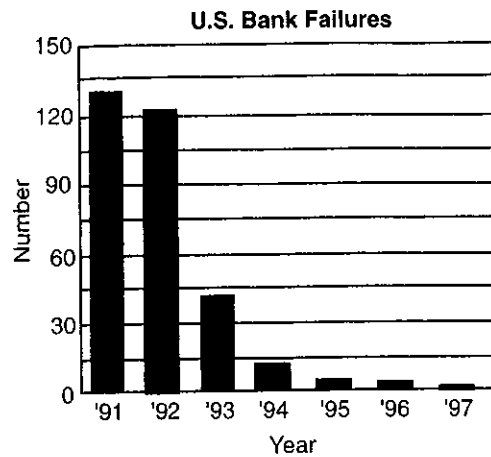
Example 2

Use the line graph to answer the questions.

- a. In which year were the most imported cars sold?
- b. About how many were sold in 1993?

Solution

- a. 1991
- b. about 1.8 million



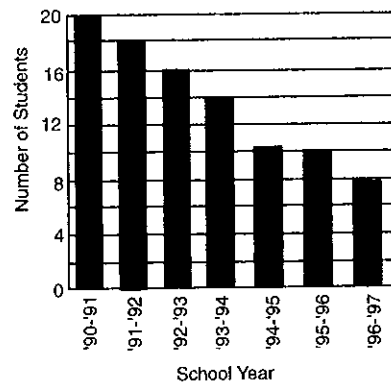
EXERCISES

Use the bar graph for Exercises 1–3.

- 1. Which two school years had about the same number of students per computer?

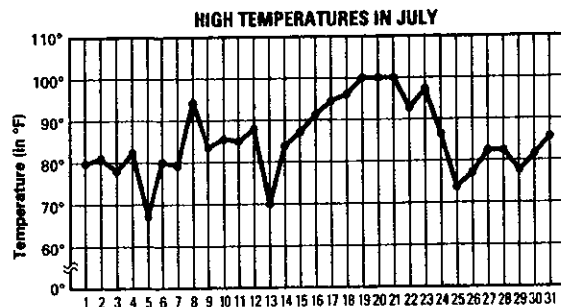
- 2. How many students were there per computer in '92–'93? _____
- 3. How many more students were there per computer in '95–'96 than in '96–'97?

Number of Students per Computer in Public Schools



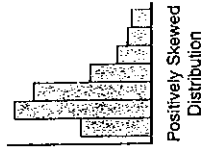
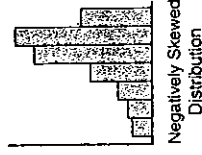
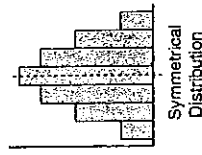
Use the weather graph for July to answer Exercises 4–6.

- 4. What was the lowest high temperature of the month? _____
- 5. On how many days was the high temperature above 90°? _____
- 6. What is the difference between the highest and the lowest high temperature? _____

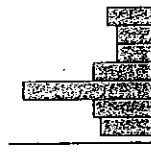


Shape of a histogram

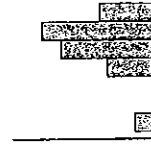
Although histograms can take on any shape, many histograms can be categorized into two different types: symmetrical (bell-shaped) and skewed. A histogram with a symmetrical distribution is symmetrical about a vertical line. A histogram with a skewed distribution is not symmetrical because it has a "tail" on one side or the other. A tail on the left side of the histogram shows a negatively skewed distribution, while a tail on the right side shows a positively skewed distribution.



Sometimes a histogram will have a bar that is significantly taller than the others. This bar represents a cluster. A cluster is a large set of data values that are very close together. The third bar from the left in the following histogram represents a cluster.

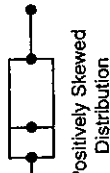
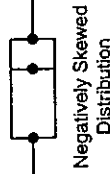
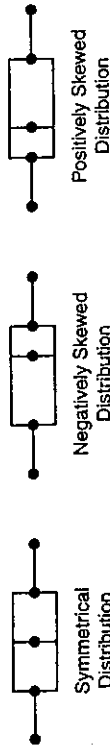


Sometimes a histogram will have a bar that is separate from the rest of the bars. In this situation, this bar represents an outlier. An outlier is a data value that is considerably smaller or larger than the rest of the data values. The first bar from the left in the following histogram represents an interval that contains an outlier.



Shape of a box-and-whisker plot

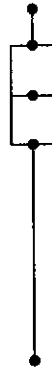
A box-and-whisker plot can show the symmetry, skewness, clusters, outliers, and spread of a distribution. A distribution is symmetrical if the median line is at or near the center of the box. A skewed distribution will have a median line that is far to the left or right of the center of the box. A line at the median that is far to the right of the box shows a negatively skewed distribution, while a median line that is far to the left of the box shows a positively skewed distribution.



Sometimes a box-and-whisker plot will show values that are considerably close to one another. These values represent a cluster. In the following box-and-whisker plot, the minimum value, lower quartile, and median are very close to one another. This means that there is a cluster of values between the minimum value and the median.

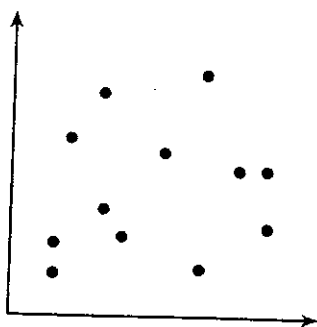


Sometimes a box-and-whisker plot will have a whisker that extends a great distance past a quartile. In this situation, the minimum or maximum value at the end of the whisker represents an outlier. In the following box-and-whisker plot, the left whisker extends far past the lower quartile. The minimum value at the end of the whisker is an outlier.




Test Your Skills

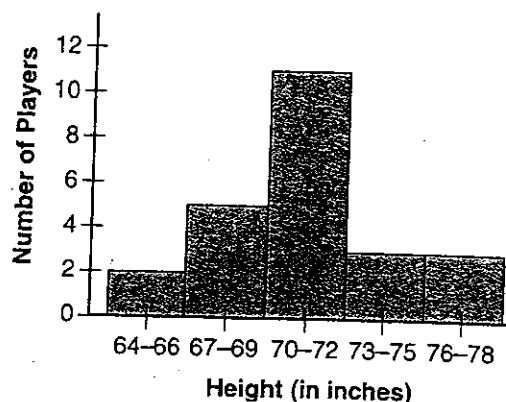
1. What type of correlation does the following scatterplot represent?



- A. no correlation
 B. perfect correlation
 C. positive correlation
 D. negative correlation
2. Which situation best provides a random sample that is **not** biased?
- A. determining the favorite food of people in your town by surveying customers at the pizza parlor
 B. determining the average height of students in your school by measuring every fourth student that enters school in the morning
 C. determining what sport is the most popular by surveying every tenth spectator as they arrive at a volleyball game
 D. determining the average number of hours ninth-graders in the U.S. watch TV by surveying every fifteenth student that enters your school on Wednesday morning

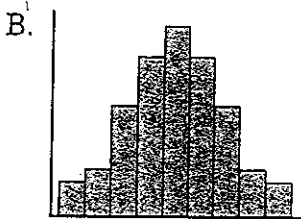
- Directions:** Use the following histogram to answer Numbers 3 and 4.

9th-Grade Basketball Players



3. What type of data is given in the histogram?
- A. bivariate and qualitative
 B. univariate and qualitative
 C. bivariate and quantitative
 D. univariate and quantitative
4. Which interval represents a cluster?
- A. 64-66
 B. 67-69
 C. 70-72
 D. 73-78

5. Which histogram shows a positively skewed distribution?



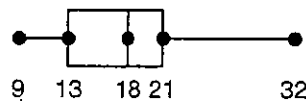
6. Which type of study would be **best** to use to determine the average distance you can throw a football?

- A. survey
- B. sample
- C. experiment
- D. observation

7. What type of display is used to represent bivariate data?

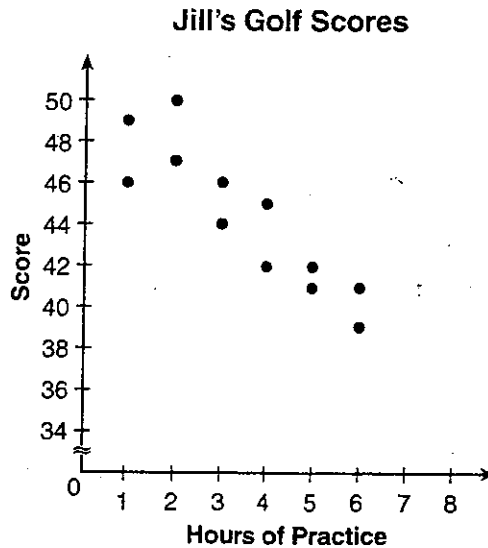
- A. histogram
- B. scatterplot
- C. circle graph
- D. box-and-whisker plot

8. What is the spread of the following distribution?



- A. 8
- B. 12
- C. 19
- D. 23

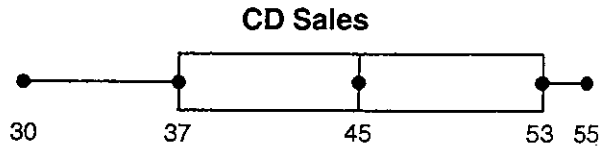
9. Draw a trend line on the following scatterplot.



What do you predict Jill's golf score will be if she practices for 8 hours?


Practice

Directions: Use the following box-and-whisker plot to answer Numbers 1 through 7.

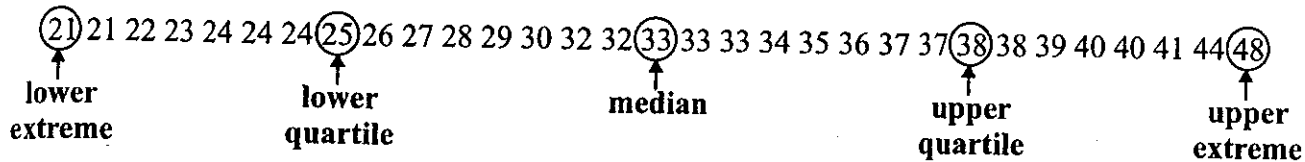


1. What is the median number of CDs sold? _____
 2. What is the minimum number of CDs sold? _____
 3. What are the lower and upper quartiles of the distribution?
 lower: _____ upper: _____
 4. Does the box-and-whisker plot represent a symmetrical, negatively skewed, or positively skewed distribution?

 5. In what range does there seem to be a cluster? _____
- | | |
|--|--|
| <ol style="list-style-type: none"> 6. What is the maximum number of CDs sold?
 A. 30
 B. 45
 C. 53
 D. 55 | <ol style="list-style-type: none"> 7. What is the spread of the distribution?
 A. 16
 B. 25
 C. 55
 D. 85 |
|--|--|

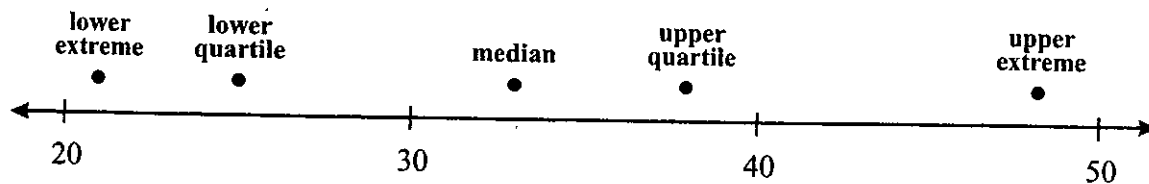
BOX-AND-WHISKER PLOTS

Box-and-whisker plots are used to summarize data as well as to display data. A box-and-whisker plot summarizes data using the median, upper and lower quartiles, and the lower and upper extreme values. Consider the data below—a list of employees' ages at the Acme Lumber Company:

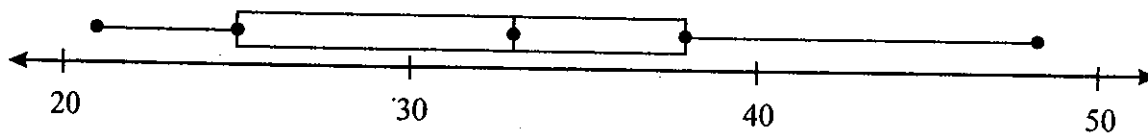


Step 1: Find the median, upper quartile, lower quartile, upper extreme, and lower extreme just like you did on the previous page.

Step 2: Plot the 5 data points found in step 1 above a number line as shown below.



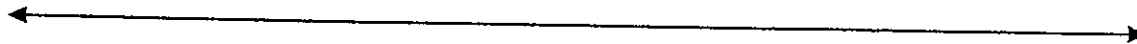
Step 3: Draw a box around the quartile values, and draw a vertical line through the median value. Draw whiskers from each quartile to the extreme value data points.



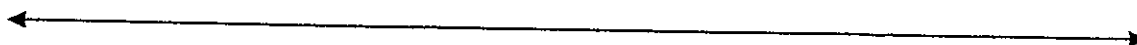
This box-and-whisker displays five types of information: lower extreme, lower quartile, median, upper quartile, and upper extreme.

Draw a box-and-whisker plot for the following sets of data.

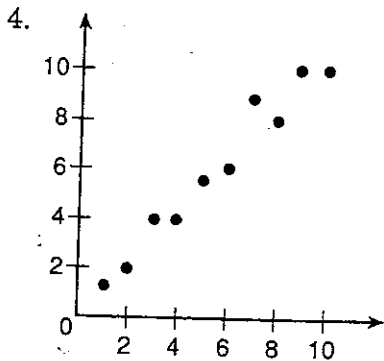
1.
10 12 12 15 16 17 19 21 22 22 25 27 31 35 36 37 38 38 41 43 45 50 51 56 57 58 59



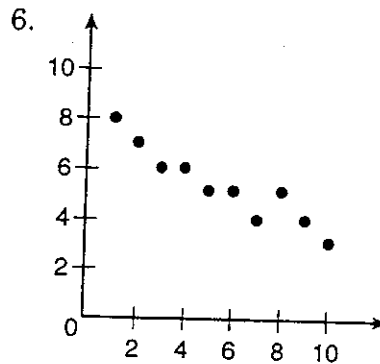
2.
5 5 6 7 9 9 10 11 12 15 15 16 17 18 19 19 20 22 24 26 27 27 30 31 31 35 37



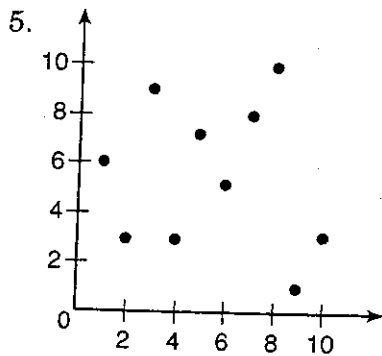
Directions: For Numbers 4 through 7, write whether each scatterplot shows positive, negative, or no correlation. If the scatterplot shows positive or negative correlation, draw a trend line that best fits the data.



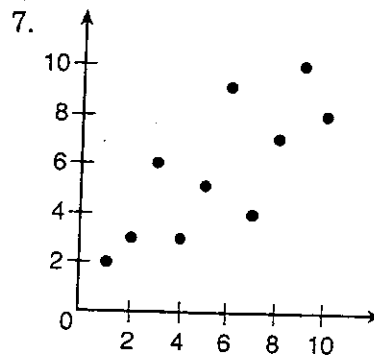
correlation: _____



correlation: _____



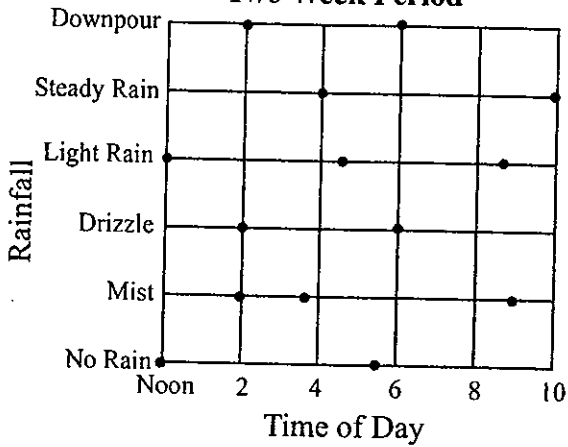
correlation: _____



correlation: _____

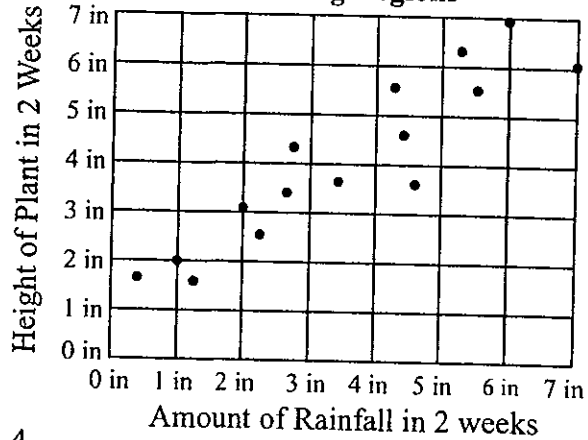
Examine each of the scatter plots below. On the line below each plot, write whether the relationship shown between the two variables is "positive", "negative", or "no relationship".

Two Week Period



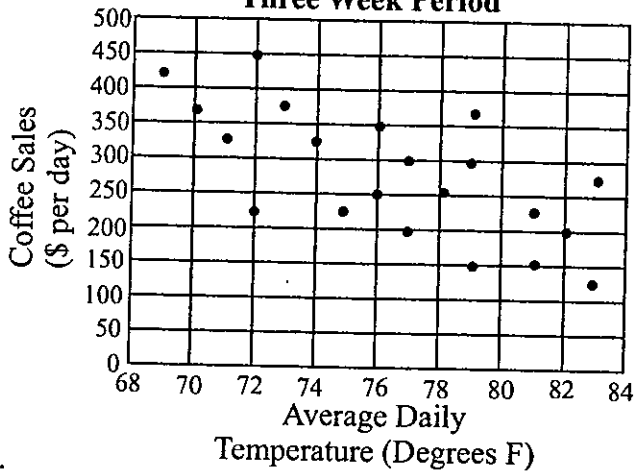
1. _____

15 Farming Regions



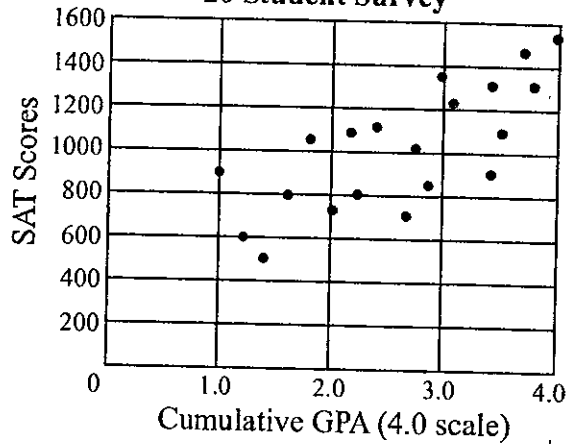
4. _____

Three Week Period



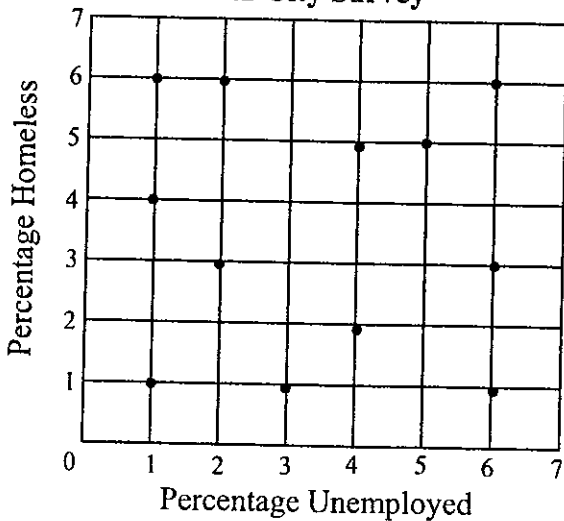
2. _____

20 Student Survey



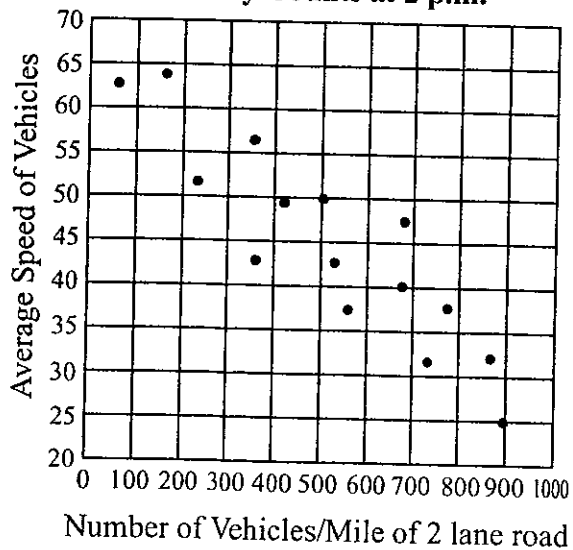
5. _____

12 City Survey



3. _____

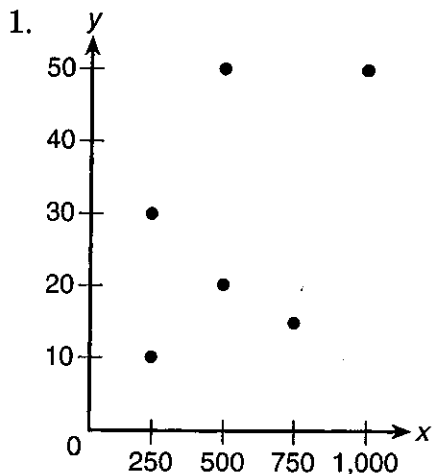
15 Day Counts at 2 p.m.



6. _____

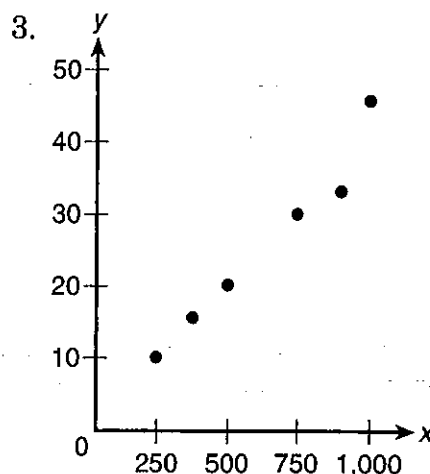
Practice

Directions: For Numbers 1 through 4, write whether each scatterplot shows positive, negative, or no correlation. If the scatterplot shows positive or negative correlation, draw a trend line and make the given prediction.



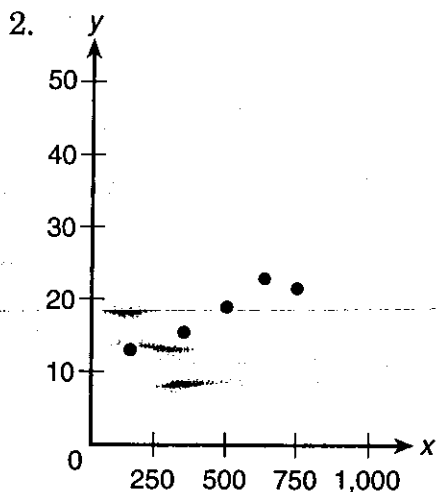
correlation: _____

prediction for y when x is 300: _____



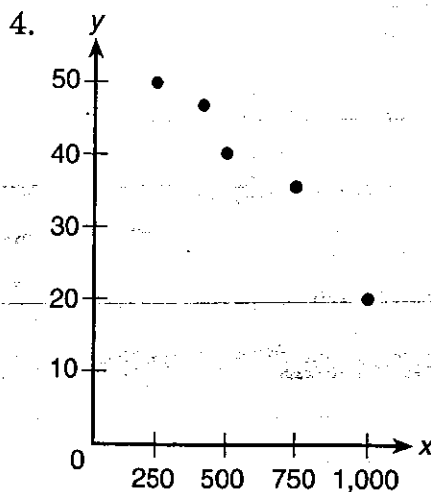
correlation: _____

prediction for y when x is 600: _____



correlation: _____

prediction for y when x is 1,000: _____



correlation: _____

prediction for y when x is 850: _____

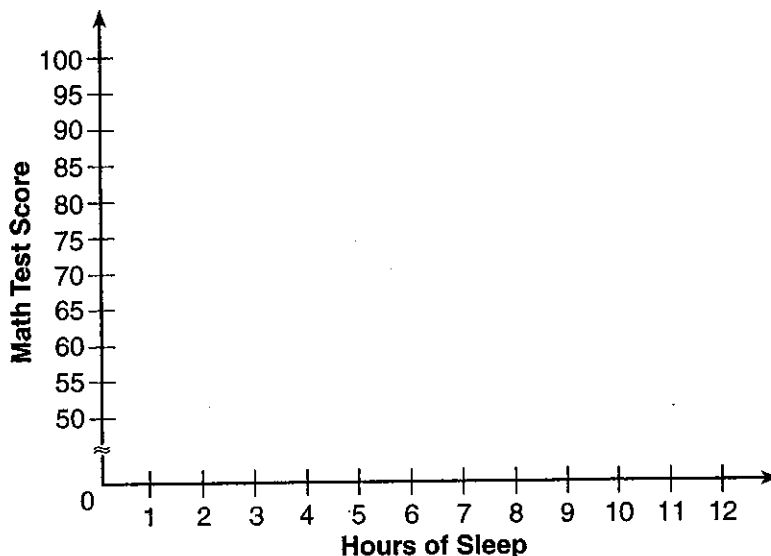
5. The following table shows the number of hours of sleep a group of 10th-grade students got the night before a math test and each student's score.

Math Score Versus Hours of Sleep

Hours of Sleep	9	4	7	6	6	8	9	8	5	7	8	7
Math Test Score	93	71	84	77	82	93	91	100	70	83	90	90

Use the information from the table to make a scatterplot of the data. Then draw a trend line through the data.

Math Score Versus Hours of Sleep



Directions: Use the scatterplot you created in Number 5 to answer Numbers 6 through 8.

6. Does the scatterplot show a positive or a negative correlation? _____

7. If a student got 2 hours of sleep, what is your prediction for his or her score?

8. Which statement about the scatterplot is **true**?

- A. It has no correlation.
- B. It has a weak correlation.
- C. It has a strong correlation.
- D. It has a perfect correlation.