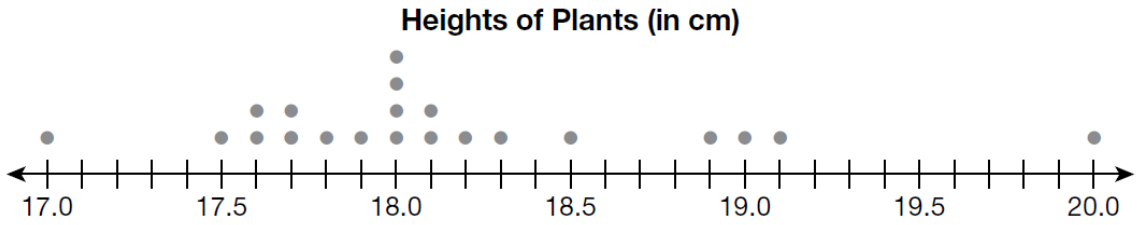
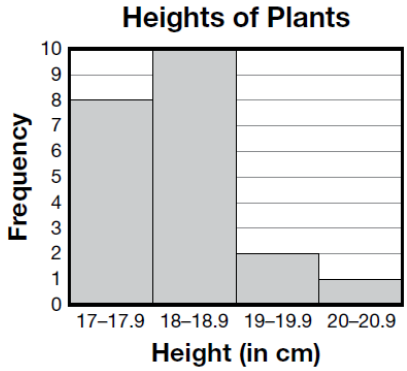
Univariate Statistics: Shape, Center, and Spread

Shape

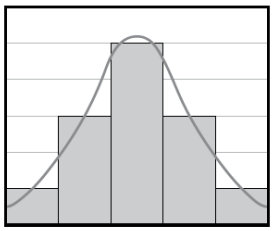
UNDERSTAND No matter what type of study you choose, it helps to organize your data in a data display. If the data are numerical and you have relatively few data points, a **dot plot** may be a good way to display them. The base of a dot plot is a number line that lists the possible values of the data. Each data point is represented by a dot placed over its value on the number line. A sample dot plot is shown below.

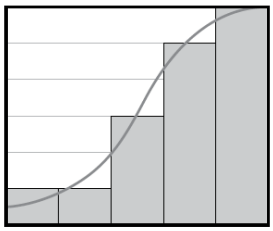


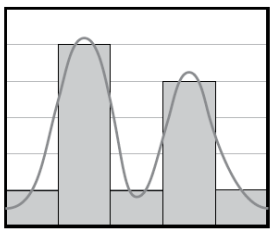
A **histogram**, another kind of data display, groups data points into ranges and shows how many of the data points fall in each range. All ranges or intervals in a histogram are of equal size, and they do not overlap. The sample histogram below shows the same data as in the dot plot above.



UNDERSTAND You can use the distribution of a data set, or its shape, to interpret it and to compare it to other data sets. Four kinds of distributions are described below.

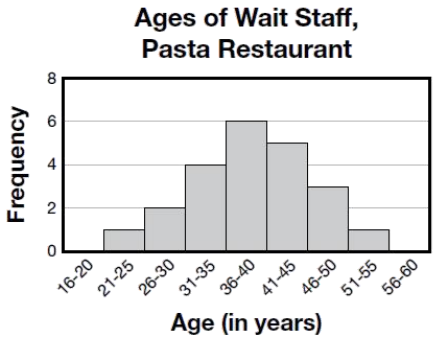

Normal Distribution

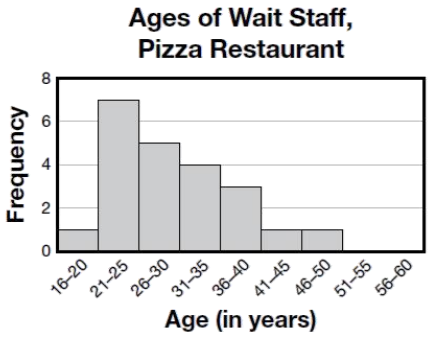

Skewed Distribution


Bimodal Distribution


Uniform Distribution

EXAMPLE B The two histograms below show the ages of wait staff at two restaurants.





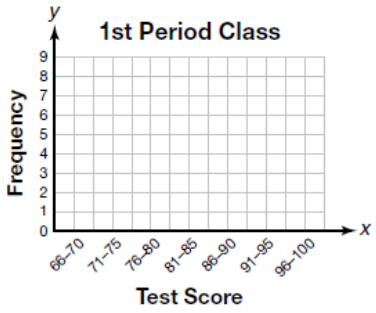
Identify the kind of distribution shown by each histogram. Use the shapes of the data sets to compare them.

For questions 3–5, use the given information. Create a histogram for each data set. Describe the distribution of each data set.

Students in the 1st and 2nd period biology classes took the same test. Their test scores are listed below.

3. 1st period test scores:

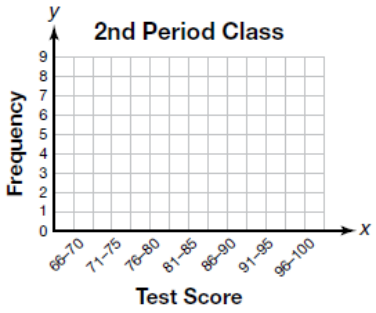
100, 91, 86, 73, 81, 100, 93, 94, 86,
86, 99, 93, 98, 84, 80, 97, 93,
87, 70, 97, 94, 88, 85, 96, 90



Distribution: _____

4. 2nd period test scores:

81, 87, 95, 85, 83, 82, 76, 68, 86,
83, 93, 87, 76, 87, 71, 100, 76,
91, 73, 80, 80, 84, 87, 88, 73



Distribution: _____

5. Compare and contrast the histograms for the biology classes in questions 4 and 5.

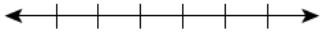
_____

_____

Create a dot plot for the given data. Describe the shape of the data.

6. Nathaniel opened 20 peanut shells and recorded the number of peanuts he found in each shell.

3, 2, 0, 1, 5, 2, 1, 2, 3, 1, 2, 2, 1, 2, 2, 3, 2, 3, 1, 2



_____

Fill in each blank with an appropriate word or phrase.

7. A _____ shows data points as dots above a number line.

8. A _____ shows how frequently data occur within certain ranges or intervals.

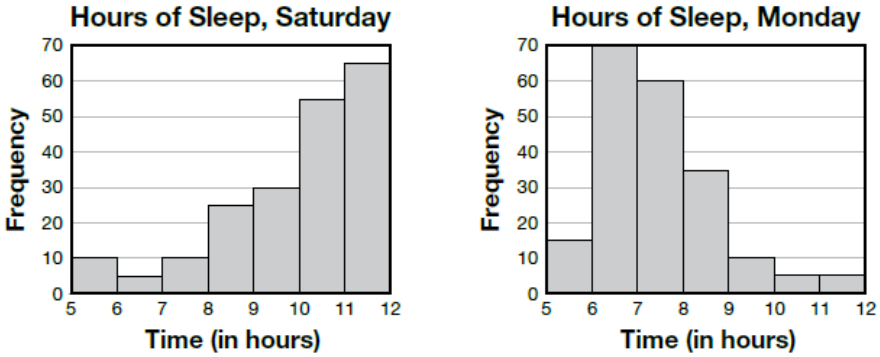
9. _____ used in a histogram must be equal.

10. A _____ distribution is symmetric and resembles a bell curve.

11. A _____ distribution has two distinct peaks.

12. A _____ distribution has a "tail" that extends more to one side of the graph than the other.

13. **COMPARE** Antoine surveyed 200 high school students to find out how many hours they slept this past Saturday night and how many hours they slept this past Monday night. The histograms show the data he collected.



Compare the distributions of data on the two graphs. How do they differ? For what reason could they be so different? Explain.

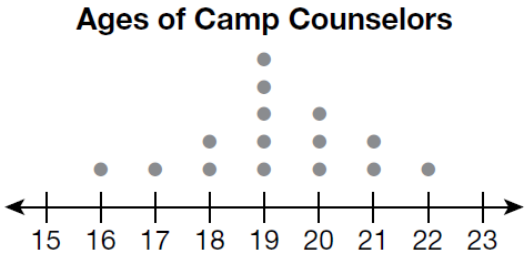
_____

_____

Center

UNDERSTAND One way to describe a data set is by using a **measure of center**. A measure of center, such as the mean, represents the average of a set of data and can be used to describe the set.

The dot plot below shows the ages of all Camp Sparrow counselors. By eyeballing the dot plot, you can see that the data are centered around the value 19. If someone asked you how old the camp counselors are, you would probably say around 19.



Each dot in the dot plot represents an element in the data set. All 15 elements from the data set are listed below.

16, 17, 18, 18, 19, 19, 19, 19, 19, 20, 20, 20, 21, 21, 22

A data set can have several kinds of measures of center. One measure of center is the **mean**, which is the average of the values in a data set. To calculate the mean, divide the sum of the elements by the total number of elements.

$$\text{mean} = \frac{16 + 17 + 18 + 18 + 19 + 19 + 19 + 19 + 19 + 20 + 20 + 20 + 21 + 21 + 22}{15}$$
$$= \frac{288}{15}$$
$$= 19.2$$

The mean age of camp counselors at Camp Sparrow is 19.2.

Another measure of center is the **median**. The median is the middle value when the elements are ordered from least to greatest. If there are an even number of elements, the median is the average (mean) of the middle two elements.

Since there are 15 camp counselors, and thus 15 ages, the middle number is the 8th value.

16, 17, 18, 18, 19, 19, 19, 19, 19, 20, 20, 20, 21, 21, 22

The median age of the camp counselors is 19.

Some data sets contain **outliers**. An outlier is a data point that is either much greater or far less than the rest of the data points. It lies far outside the group that contains the rest of the data. For example, a camp counselor who was 41 would be an outlier. An outlier can affect which measure of center best describes the data set. In general, when a data set contains an outlier, the median is a better measure of center than the mean.

Below are the scores that Justin earned on his last 8 homework assignments.

80, 95, 0, 90, 95, 80, 85, 90

What are his mean homework score and median homework score?

Find the mean score.

$$\text{mean} = \frac{80 + 95 + 0 + 90 + 95 + 80 + 85 + 90}{8}$$
$$\text{mean} = \frac{615}{8}$$

▶ mean = 76.875

To find the median score, first list the scores from least to greatest.

0, 80, 80, 85, 90, 90, 95, 95

Because there is an even number of terms in the set, find the mean of the two middle numbers.

The two middle numbers are 85 and 90.

0, 80, 80, 85, 90, 90, 95, 95

▶ $\text{median} = \frac{85 + 90}{2} = \frac{175}{2} = 87.5$

Find the mean and median for each data set.

3. 5, 25, 10, 15, 20

median: _____

mean: _____

4. 1, 7, 3, 2, 6

median: _____

mean: _____

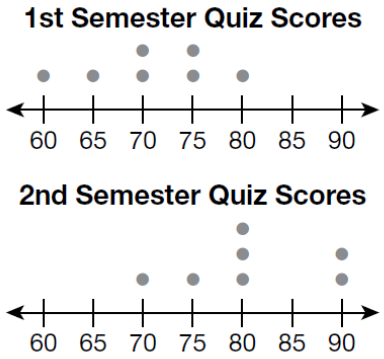
5. 10, 90, 10, 60, 40, 30

median: _____

mean: _____

Find the median for each data set or determine the interval in which the median must fall. Then compare the medians.

10. The dot plots show Kyla's Spanish quiz scores during the 1st and 2nd semesters.

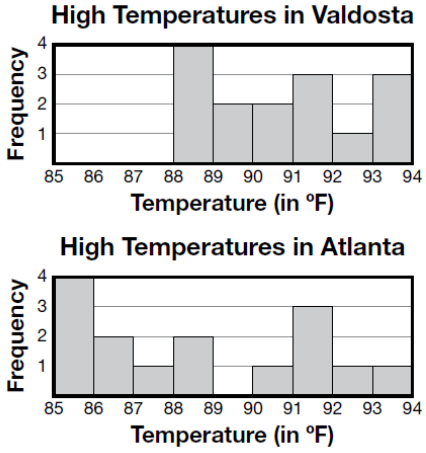


median score, 1st semester: _____

median score, 2nd semester: _____

Comparison: _____

11. The histograms show the daily high temperatures in two cities.



median high temperature, Valdosta: _____

median high temperature, Atlanta: _____

Comparison: _____

Find the mean for each data set. Then compare the means.

12. The tables show the number of ads that were sold by the actors and stage-crew members working on a school play.

Actor	Rajiv	Amy	Penny	Leonard	Adriel
Ads Sold	4	4	5	6	7

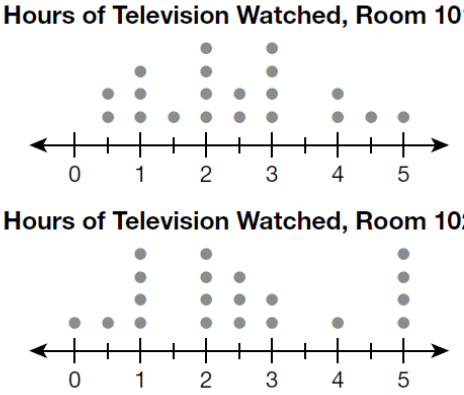
Crew Member	Tina	Ben	Ronny	Irene	Cris
Ads Sold	6	7	8	9	9

mean number sold, actors: _____

mean number sold, crew members: _____

Comparison: _____

13. The dot plots show the number of hours of television watched yesterday by students from two homerooms.



mean number of hours, Room 101: _____

mean number of hours, Room 102: _____

Comparison: _____

14. Which statement accurately compares the average weight of a puppy from the 2nd litter to the average weight of a puppy from the 1st litter?

Weights of Puppies (in ounces)

1st Litter	$3\frac{1}{2}$, 4, 4, $4\frac{1}{2}$
2nd Litter	$4\frac{1}{2}$, 5, 7, $7\frac{1}{2}$

A. The average weight is about the same for both litters.

B. The average weight of a puppy from the 2nd litter is about $\frac{1}{2}$ as great.

C. The average weight of a puppy from the 2nd litter is about $1\frac{1}{2}$ times as great.

D. The average weight of a puppy from the 2nd litter is about $2\frac{1}{2}$ times as great.

15. To compare two shipments, five packages from each shipment were chosen at random and weighed. Which measure or measures of center would be best to use if you wanted to compare the weight of a typical package from each shipment?

Weights of Packages (in pounds)

1st Shipment	2, 4, 6, 8, 10
2nd Shipment	3, 3, 5, 8, 50

A. Median would be the best measure of center.

B. Mean would be the best measure of center.

C. Median and mean would both be equally good measures of center.

D. Neither the mean nor the median would be a good measure of center.

Spread

UNDERSTAND The mean and **median** of a data set are used to measure where the center of a set of data lies. The **mean absolute deviation (MAD)** indicates how spread out, or how variable, data are.

The MAD measures how much the data points in a set vary from the mean, $\bar{x}$. To find the distance of a data point from the mean, subtract the mean from the data point, x: $x - \bar{x}$. Since distances are always positive, you must take the absolute value of that difference to find the absolute deviation from the mean. The MAD is the mean of all the absolute deviations in the data set.

Calculate the MAD of this data set: 5, 8, 9, 11, 12.

Calculate the mean, $\bar{x}$.

$$\bar{x} = \frac{5 + 8 + 9 + 11 + 12}{5} = \frac{45}{5} = 9$$

Find the absolute deviation of each data point from the mean. Use a table to organize your work.

Data Point, x	Deviation from Mean, $x - \bar{x}$	Absolute Deviation from Mean, $\lvert x - \bar{x} \rvert$
5	$5 - 9 = -4$	$\lvert -4 \rvert = 4$
8	$8 - 9 = -1$	$\lvert -1 \rvert = 1$
9	$9 - 9 = 0$	$\lvert 0 \rvert = 0$
11	$11 - 9 = 2$	$\lvert 2 \rvert = 2$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$

Calculate the mean of the absolute deviations.

$$\text{MAD} = \frac{4 + 1 + 0 + 2 + 3}{5} = \frac{10}{5} = 2$$

The mean absolute deviation is 2.

Use the data below for questions 1–4.

Heights (in inches) of Starting Players, Girls' Basketball Team	Heights (in inches) of Starting Players, Boys' Basketball Team
64, 66, 66, 68, 71	67, 67, 69, 70, 72

1. Calculate the mean and MAD of the heights of starting players for the girls' team. Use the table. Show your work.

$\bar{x}$ = _____

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
64		
66		
66		
68		
71		

MAD = _____

2. Calculate the mean and MAD of the heights of starting players for the boys' team. Use the table. Show your work.

$\bar{x}$ = _____

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
67		
67		
69		
70		
72		

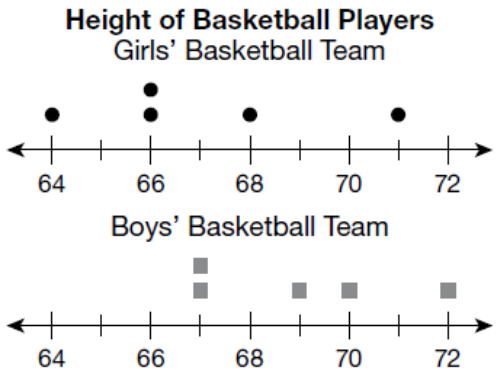
MAD = _____

3. On average, which team has taller starting players? Use the means you calculated above and the dot plots on the right.

_____

4. On which team are the heights of the starting players more variable? Use the MADs you calculated above and the dot plots on the right.

_____



Conduct the required calculations for the data sets below.

5. Set M: \$8.50, \$8.50, \$10.00

Set N: \$5.60, \$7.40, \$8.00

HINT: The mean is the sum of all the numbers in a set divided by the number of numbers in the set.

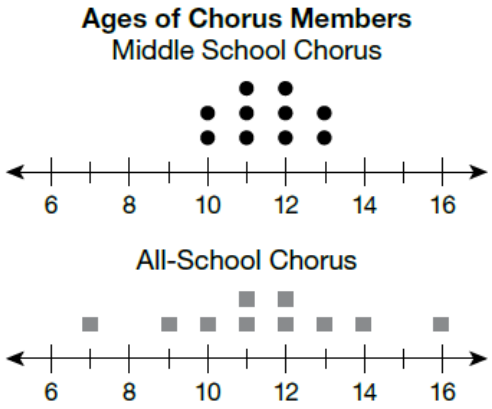
mean of Set M = _____ mean of Set N = _____

6. Set P: 54, 58, 90, 191, 142

Set Q: 96, 117, 107, 97, 118

MAD of Set P = _____ MAD of Set Q = _____

The dot plots show the ages of a random sample of students taken from a middle-school chorus and a sample of students taken from the all-school chorus in Lisa's hometown. Use these plots for questions 7–9. Show your work on a separate sheet of paper.



7. Calculate the mean age and MAD of members of the middle-school chorus.

mean = _____ MAD = _____

8. Calculate the mean age and MAD of members of the all-school chorus.

mean = _____ MAD = _____

9. **CONCLUDE** What conclusions can you draw about the average ages of chorus members, as well as the spread of the ages, in the two choruses? Use the measures you found above and the dot plots to support your answer.

_____

_____