SHAPE, CENTER, AND SPREAD OF DATA notes

shape of data - tells us how the data is distributed

center of data - gives us the middle of the data set
    *can be found by mean or median

spread of data - tells how close together or spread out the data set is
    *can be found by range, interquartile range, mean absolute deviation, or standard deviation

peak - point where the data is higher than the other parts of the data set
    *can be found by mode
*one peak = unimodal   *two peaks = bimodal *multiple peaks= multimodal *no peaks = uniform

Example: Students were surveyed on how many minutes they spend reading each night.
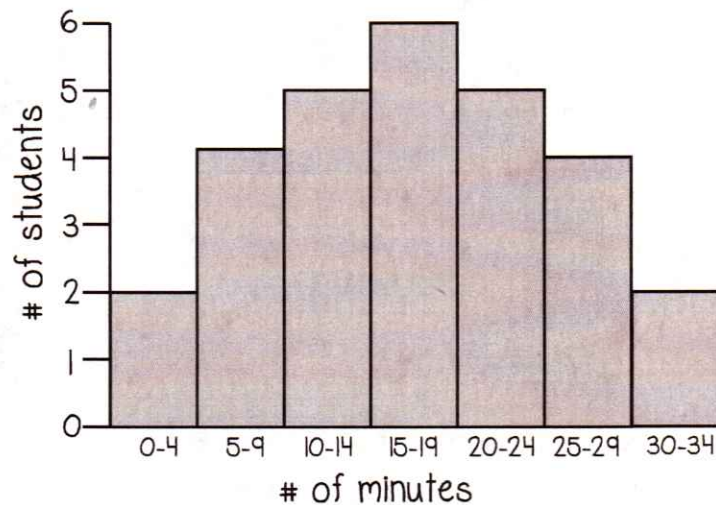
How would you describe the shape of the data?
symmetric

How many peaks does the graph have?
one (unimodal)

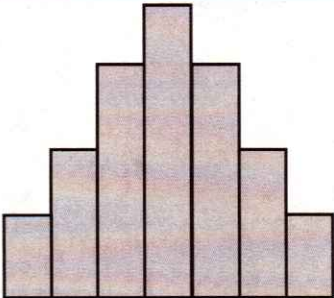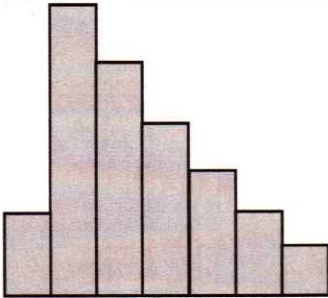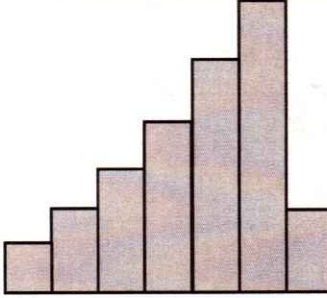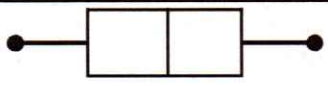Estimate the center (mean or median) of this data set.
≈17
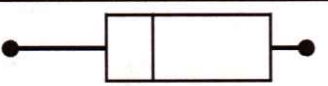
How would you describe the spread of data?
The data is relatively spread out.

SHAPE OF DATA *notes continued*

Data sets can be categorized by how they are <u>shaped.</u> Here are 3 possible shapes of a data set.

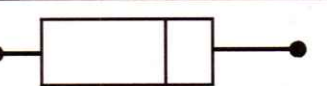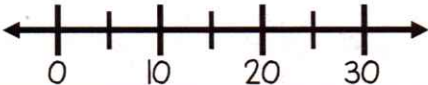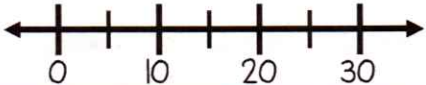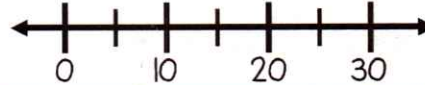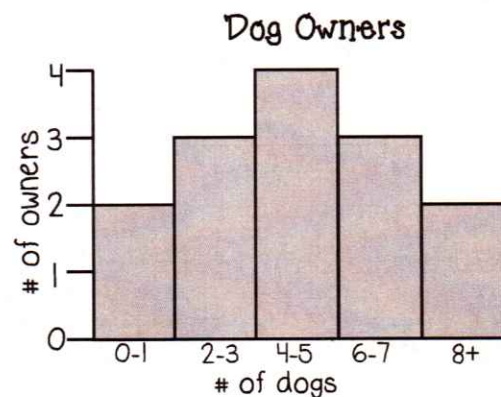| SYMMETRIC	SKEWED RIGHT	SKEWED LEFT
• bell-shaped • mean and median are the same or very close	• positive skew • mean > median • long tail on right • also called positive skew	• negative skew • mean < median • long tail on left • also called negative skew

Comparing Data Sets

People were asked if they owned cats or dogs. The data is organized into the two histograms below.
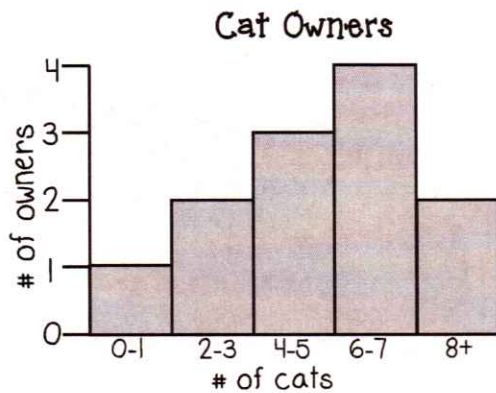
1. What is the shape for the cat owner histogram? <mark>skew left</mark>

2. What is the shape of the dog owner histogram? <mark>symmetric</mark>

3. How do the spread of the two data sets compare? <mark>they both have about the same spread</mark>

4. How do their peaks compare? <mark>they are both unimodal</mark>

   Cat owners peak at a higher # of cats than dog owners do at # of dogs

SHAPE, CENTER, AND SPREAD practice
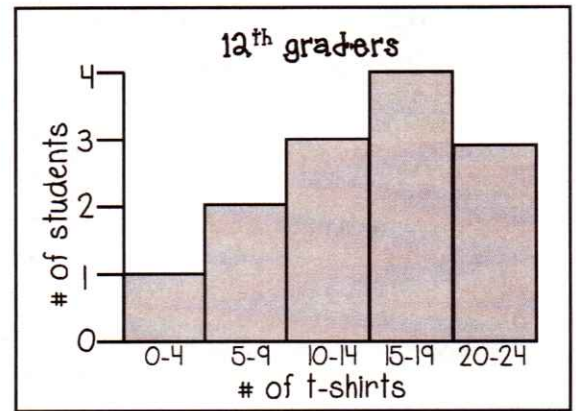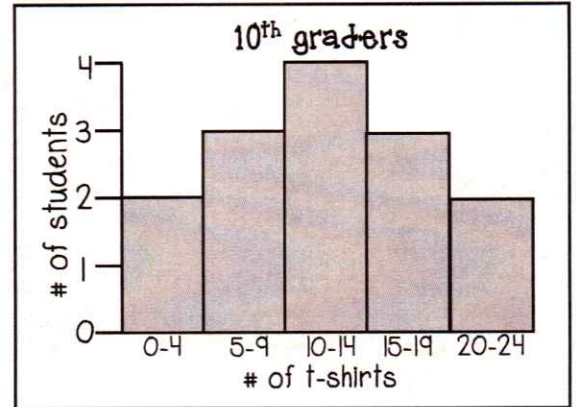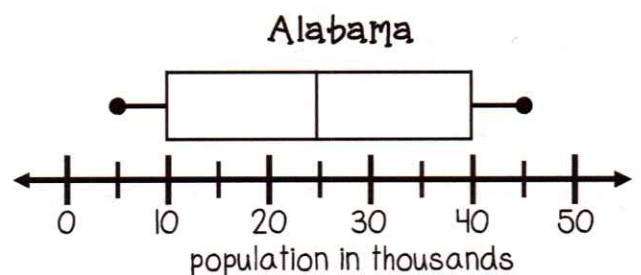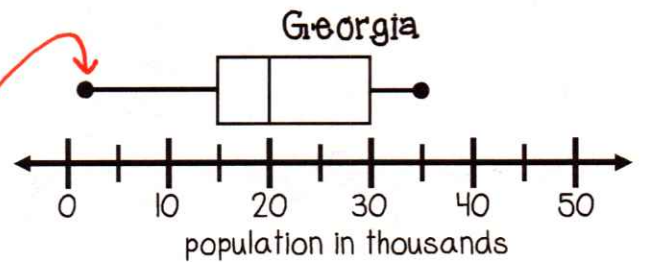
Use the histograms to answer questions #1-6. A group of 10th and 12th graders were surveyed about how many t-shirts they own.

1. Which data set is symmetric? 10th

2. Give a possible reason for the difference in their shapes.
12th graders have had longer to collect t-shirts
* answers can vary

3. How do the centers compare? Which is greater?
12th graders have a higher center

4. How do the spreads of each data set compare?
They are about the same

5. What is the shape of the 12th grade data set?
skewed left

6. Are the sets unimodal, bimodal, or uniform?
both sets are unimodal

10th graders

(histogram: x-axis "# of t-shirts" with categories 0-4, 5-9, 10-14, 15-19, 20-24; y-axis "# of students" 0 to 4; bars: 0-4=2, 5-9=3, 10-14=4, 15-19=3, 20-24=2)

12th graders

(histogram: x-axis "# of t-shirts" with categories 0-4, 5-9, 10-14, 15-19, 20-24; y-axis "# of students" 0 to 4; bars: 0-4=1, 5-9=2, 10-14=3, 15-19=4, 20-24=3)

Use the box and whisker plots to answer questions #7-12. The population of 15 rural towns in Georgia and 15 rural towns in Alabama are organized and displayed using the box and whisker plots below.

7. Which state's data set has a greater center?
Alabama

8. How do the shapes of each data set compare?
Georgia is skewed right. Alabama is symmetric.

9. Does either data set have an outlier?
Georgia at around 2,000
   probably an outlier since left whisker is so long

10. Which state's data set has a larger spread?
Alabama (range is 40)

11. True or False: The 15 rural towns in Georgia are mostly between 15,000 and 30,000. True

12. Which state's data set has a smaller range?
Georgia

Georgia

(box and whisker plot: x-axis "population in thousands" 0 to 50)

Alabama

(box and whisker plot: x-axis "population in thousands" 0 to 50)

SHAPE, CENTER, AND SPREAD *practice 2*

Use the dot plots to answer questions #1-6. A group of freshmen and sophomores were asked how many hours they spend on homework each week.

1. How do the shapes of each data set compare?
They are both symmetric (other than the outlier)

2. Do either group of students have an outlier?
Freshmen

Freshmen

time in hours

3. Which data set has a greater center?
Sophomores

4. Which group has a wider spread of data?
Sophomores
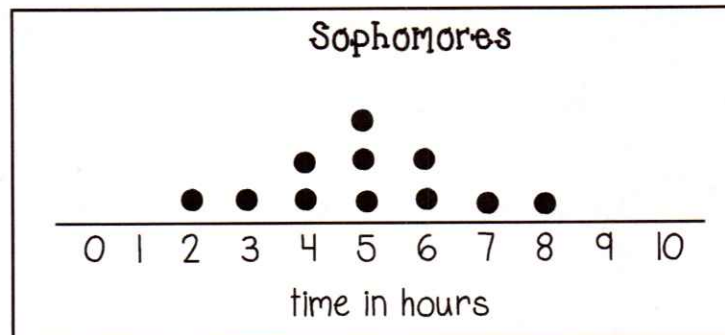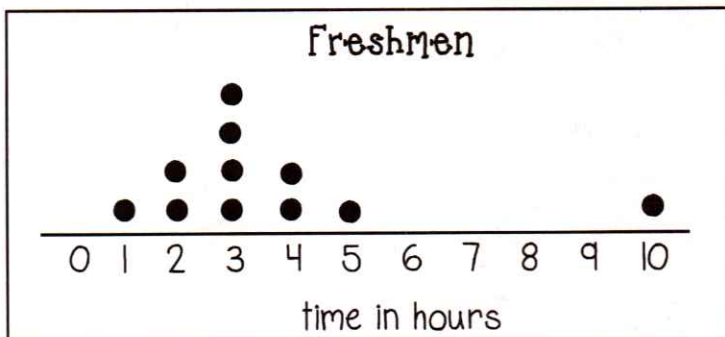
5. Give one possible reason for the difference in centers for these data sets.
Sophomores have harder classes
*answers can vary

Sophomores

time in hours

6. True or False: The data set for the sophomores is uniform. False - it's unimodal
↳uniform is when all the same height

Use the stem and leaf plots to answer questions #7-12. As people walked into a mall, a surveyor asked their age. She did this on a Tuesday and again on a Saturday from 1 p.m. - 3 p.m.

7. How do the shapes compare?
Tuesday doesn't have a distinct shape. Saturday is symmetric.
   Tuesday is mostly symmetric & bimodal

8. How do the peaks compare?
Tuesday is bimodal. Saturday is unimodal.

9. How do the centers compare?
They are both about the same (30s-40s). Tuesday's mean is a little higher.

Tuesday

Stem	Leaf
1	
2	0, 2, 2, 3, 4, 4
3	3, 5
4	2, 6
5	1, 3
6	0, 1, 3, 5, 5

10. Did the surveyor leave an age range out? Do you think this affected the data sets? Yes 0-10 years old. It could possiblly affect the data.

11. True or False: The ages of people who attended the mall on Saturday were evenly distributed. True

12. Give one possible reason for the shape of Tuesday's data.
College kids and retirees were at the mall.

Saturday

Stem	Leaf
1	1, 3, 5, 7, 7, 9
2	0, 2, 4, 8, 8, 9, 9
3	1, 3, 4, 5, 7, 7, 8, 9, 9
4	2, 3, 4, 5, 6, 7, 8
5	1, 3, 4, 5, 7, 7
6	0, 1, 5, 6