DESCRIBING DATA DISTRIBUTIONS GRAPHICALLY



The student is expected to use the graphical representation of numeric data to describe the center, spread, and shape of the data distribution.

TELL ME MORE ...

You can represent a data set graphically with a box plot, stem-and-leaf plot, histogram, or dot plot. One of the reasons that you would want to graphically display a data set is so that you can describe the center, spread, or shape of the data distribution.

The center of a data distribution involves where the center of the data set might be and is measured with central tendency such as mean or median. The spread of a data distribution involves how spread out the data values are and is measured by the interquartile range or range. The shape of the data distribution refers to peaks, gaps, or symmetry in how the data values are displayed.



EXAMPLES

EXAMPLE 1: The stem-and-leaf plot shows the size of hailstones, in inches, that were measured in a recent hailstorm in west Texas. Describe the center, spread, and distribution of the data set.

STEP 1 Describe the center of the data set.

- There are 15 data values, so the median will be the 8th data value when listed in order from least to greatest.
- The 8th data value is 1.5.

The center of the data set is around the median of 1.5 inches.

Stem	Leaf	
0	2258	
1	0025577899	
2		
3	2	
1 0 represents 1.0		
	-	
Stem	Leaf	
Stem 0	Leaf 2 2 5 8	
Stem 0 1	Leaf 2 2 5 8 0 0 2 5 7 7 8 9 9	
Stem 0 1 2	Leaf 2 2 5 8 0 0 2(5)5 7 7 8 9 9	
Stem 0 1 2 3	Leaf 2 2 5 8 0 0 2(5)5 7 7 8 9 9 2	

1 | 0 represents 1.0

STEP 2 Describe the spread of the data set.

- The minimum value is 0.2 inches and the maximum value is 3.2 inches.
- The range of the data set is 3.2 0.2 = 3 inches.

The data values are spread out over a 3 inch interval.

- **STEP 3** Describe the distribution of the data set.
 - The data values are clustered from 0 to 1.9 inches.
 - There is a gap between 1.9 and 3.2 inches.
 - The number of leaves in each stem is not symmetric.

The distribution is not symmetric. The data values are clustered from 0 to 1.9 inches and there is a gap between 1.9 and 3.2 inches.

EXAMPLE 2: The histogram shows the number of students who earned a certain number of points on a recent quiz in Mr. Bartos's math class. Describe the center, spread, and distribution of the data set.

STEP 1 Describe the center of the data set.

- Count the total number of students in the data set by reading the height of each bar in the histogram.
- $\bullet 3 + 1 + 3 + 5 + 7 + 10 = 29$
- The middle value is the 15th value.
- The middle value is found in the 30-35 bar.

The center of the data set is between 30 and 35 points scored.

- **STEP 2** Describe the spread of the data set.
 - In a histogram, values are reported as being within a range, so it is not possible to determine exact data values from the set in a histogram.
 - The minimum value could be as low as 0 points.
 - The maximum value could be as high as 40 points.
 - The range of the data set is 40 0 = 40 points.

The data values are spread out over as much as a 40-point interval.

- **STEP 3** Describe the distribution of the data set.
 - The data values are clustered toward the high end of the number of points scored.



- There is a gap between 5 and 15 points.
- The bars are not symmetric.
- Since the center is located on the right side of the distribution and data values extend to the left end, the distribution is skewed left from center.

The distribution is skewed left. There is a gap between 5 and 15 points.

EXAMPLE 3: The box plot shows data describing the number of people that rode on each of 50 buses for a recent rodeo. Describe the center, spread, and distribution of the data set.

STEP 1 Describe the center of the data set.

In a box plot, the median is the vertical line inside the box.



■ The median is 25 people.

The center of the data set is 25 people.

- **STEP 2** Describe the spread of the data set.
 - In a box plot, the box represents the middle half of the data set, between the first quartile and third quartile.
 - The difference between the first quartile and third quartile (the width of the box) is the **interquartile range**.
 - The interquartile range is 35 20 = 15 people, so the middle half of the data is spread between 20 people and 35 people.
 - The range of the data set is 60 10 = 50 people.

Half of the buses held between 20 and 35 people. The range of the number of people on each bus is 50 people.

- **STEP 3** Describe the distribution of the data set.
 - The right whisker is longer than the left whisker, so the data set is not symmetric.
 - More data values are on the left end of the data set, so the data set is skewed right.
 - A data value may be considered an **outlier** if it is more than 1.5 times the interquartile range above the third quartile or below the first quartile.
 - The maximum value is 60, which is 25 people greater than the third quartile. The interquartile range is 15 people.
 - 1.5(15) = 22.5, and 25 > 22.5, so the maximum value may be considered an outlier.

The distribution is skewed right. The maximum value, 60 people, is an outlier for the data set.

PRACTICE

For questions 1-4, identify whether the numeric data representation is symmetrical, skewed left, skewed right, or neither symmetrical nor skewed.





6. The stem-and-leaf plot represents the ages of 25 people responding to a survey about their internet use. Based on the stem-and-leaf plot, what is the median age of the survey responders?

Stem	Leaf	
0	8	
1	2489	
2	112559	
3	013478	
4	348	
5	45	
6	158	
1 2 represents 12		

8. The histogram shows the numbers of individual bags of trail mix with the amount of chocolate candies in each bag. What is the range of the possible number of chocolate candies contained in a bag of trail mix?



The histogram represents student responses to a survey about whether they watch the evening news. Which statement best describes the data shown in the histogram?

9.



- **A** The data are skewed to the right.
- **B** The data are skewed to the left.
- **C** The data are symmetrically distributed.
- **D** The data are not symmetrically distributed nor are they skewed.
- **10.** Kathleen collected data about the amount of time, in hours, college students spend listening to music each week. She graphed the data using the box plot shown. Which statement best describes the center of the data represented by the box plot?



- **F** The mean amount of time is
- 42.5 hours.
- **G** The median amount of time is 60 hours.
- H The amounts of time range by 40 hours.
- J Most students listen more than 60 hours.

 Students in Mr. Singleton's class were asked to estimate the number of hours per week spent on homework or reading. The dot plot represents the results.



Which statement best describes the data shown in the plot?

- **A** More than half the distribution is less than 7 hours.
- **B** The peak of the data is 14 hours.
- **C** The data distribution has no gaps.
- **D** The range of the data is 6 hours.
- **12.** The stem-and-leaf plot represents data collected about the length of time, in minutes, it takes students to get ready for school each morning.

Stem	Leaf
0	
1	5
2	0448
3	03359
4	145
5	
6	36
1 5 represents 15	

Which statement describes the data shown in the plot?

- **F** The data are mostly symmetrical.
- **G** The median time is 33 minutes.
- **H** The range is 60 minutes.
- **J** The mean time is 39 minutes.