Extension Analyze Data Distribution

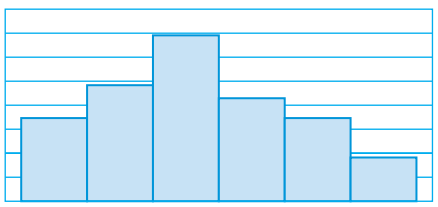
GOAL Choose an appropriate display, measure of central tendency, and measure of spread based on the shape of a data distribution.

COMMON CORE

CC.9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).*

When you are presenting a set of data, you should consider the distribution of the data before deciding what type of measure of central tendency and graph to use for the data.

DATA THAT ARE CLOSELY GROUPED Use a histogram to display the data. Use the mean as a measure of central tendency. Use standard deviation as a measure of the spread.



DATA VALUES THAT ARE SPREAD OUT Use a box-and-whisker plot to display the data. Use the median as a measure of central tendency. Use the interquartile range as a measure of the spread.

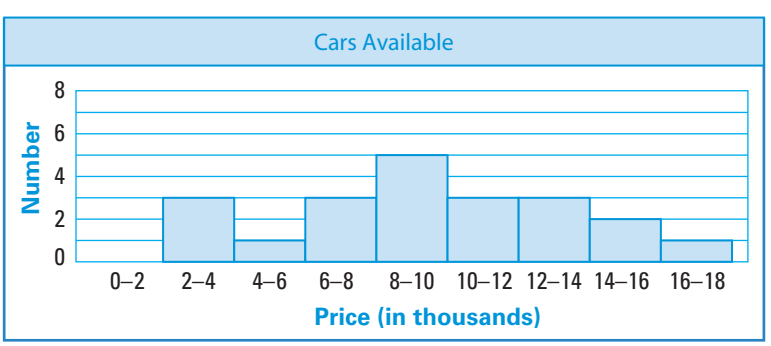
EXAMPLE 1 Choose a display for data

A used car dealer has 21 cars for sale at the prices shown in the table. Choose an appropriate display, measure of central tendency, and measure of spread for this data set.

$2150	$2800	$3500	$5100	$6050	$7100	$7250
$8000	$8850	$9100	$9225	$9900	$10,200	$10,800
$11,750	$12,200	$12,640	$13,020	$14,700	$15,500	$16,400

Solution

The data are close together with no outliers. Use a histogram. The center of the data can be represented by the mean, which is $9,345. The spread can be represented by the standard deviation, which is about $3946.



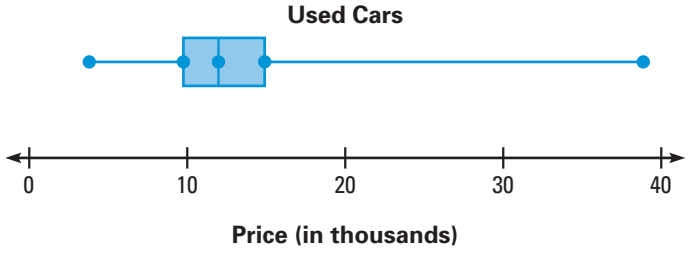
EXAMPLE 2 Choose a display for data

Another used car dealer has 24 cars for sale at the prices shown in the table. Choose an appropriate display, measure of central tendency, and measure of spread for this data set.

$3,800	$5,100	$7,100	$7,250	$8,850	$9,225	$9,900	$10,200
$10,500	$10,800	$11,400	$11,750	$12,200	$12,350	$12,640	$13,020
$13,890	$14,700	$15,500	$15,990	$17,000	$17,800	$22,900	$38,775

Solution

The data value $38,775 appears to be an outlier. Use a box-and-whisker plot to display the data. The outlier will affect the mean and standard deviation, so they do not represent the data well. The median is $11,975. The interquartile range is $5537.50.



PRACTICE

For Exercises 1–6, choose an appropriate display, measure of central tendency, and measure of spread for the data set. Explain your reasoning.

1. **QUIZ SCORES** The scores on the first quiz in Mr. Stuart's math class were 6, 9, 10, 12, 12, 13, 14, 14, 15, 15, 15, 16, 16, 17, 17, 17, 17, 18, 18, 18, 19, 19, 19, 20, and 20.
2. **FOOTBALL** The points scored by twenty of the top 25 college football teams on Saturday, September 25, 2010 were 24, 73, 37, 42, 17, 31, 70, 35, 10, 20, 37, 65, 22, 31, 20, 24, 12, 27, 14, and 34.
3. **RUNNING** The time (in minutes) it took twenty freshmen to run the mile in physical education class were 7, 7.5, 8, 8, 8.2, 8.4, 8.5, 9, 9, 9, 9.6, 9.8, 10, 10.5, 10.5, 10.8, 11.2, 11.5, 11.7, and 12 minutes.
4. **HOMEWORK** The numbers of hours that twenty-five students spent doing homework last week were 1, 8, 8, 8.5, 9, 9.5, 9.5, 10, 10, 10, 10, 10, 10.5, 10.5, 10.5, 11, 11, 11, 11, 11.5, 11.5, 12, 12, 12, and 12.
5. **COOKIES** The numbers of cookies in 20 boxes at a bake sale are 16, 16, 18, 18, 20, 20, 24, 24, 24, 24, 26, 28, 28, 30, 30, 30, 30, 36, 36, and 36.
6. **BASEBALL** The attendance at a professional baseball team's home games during September are shown in the table.

39,555	31,424	40,788	31,647	31,596	33,623
36,364	37,285	34,481	36,553	39,316	38,057