Start Your Day the Right Way
Graphically Representing Data

Vocabulary
Choose the term that best completes each statement.

<table>
<thead>
<tr>
<th>dot plot</th>
<th>five number summary</th>
<th>data distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>symmetric</td>
<td>discrete data</td>
<td>skewed left</td>
</tr>
<tr>
<td>histogram</td>
<td>skewed right</td>
<td>frequency</td>
</tr>
<tr>
<td>box-and-whisker plot</td>
<td>bin</td>
<td>continuous</td>
</tr>
</tbody>
</table>

1. A(n) _______ histogram _______ is a graphical way to display quantitative data using vertical bars.

2. A data distribution is _______ skewed right _______ if the peak of the data is to the left side of the graph with only a few data points to the right side of the graph.

3. Discrete data _______ are data that have only a finite number of values or data that can be “counted.”

4. A(n) _______ box-and-whisker plot _______ displays the data distribution based on a five number summary.

5. The overall shape of a graph which shows the way in which data are spread out or clustered together is called the _______ data distribution _______.

6. Continuous data _______ are data which can take any numerical value within a range.

7. A data distribution is _______ skewed left _______ if the peak of the data is to the right side of the graph with only a few data points to the left side of the graph.

8. A(n) _______ dot plot _______ is a graph that shows how discrete data are distributed using a number line.

9. For a set of data, the _______ five number summary _______ consists of the minimum value, the first quartile, the median, the third quartile, and the maximum value.

10. A data distribution is _______ symmetric _______ if the peak of the data is in the middle of the graph. The left and right sides of the graph are nearly mirror images of each other.

11. The number of data values included in a given bin of a data set is called _______ frequency _______.

12. The bar width in a histogram that represents an interval of data is often referred to as a _______ bin _______.
Problem Set

Construct the graphical display for each given data set. Describe the distribution of the data.

1. Construct a dot plot to display the scores on a recent math quiz. The data are 12, 14, 8, 13, 12, 14, 15, 10, 13, 12, 0, 14, 11, 14, 13, and 10.

![Dot plot for math quiz scores]

The data are skewed left.

2. Construct a dot plot to display the number of canned goods donated by each student during a charity event. The data are 15, 18, 18, 22, 13, 15, 19, 17, 18, 17, 16, 10, 17, 20, 19, 25, 17, 18, 19, and 16.

![Dot plot for canned goods donated]

The data are symmetric.

3. Construct a dot plot to display the number of items purchased by a number of randomly chosen customers at a toy store. The data are 2, 4, 3, 7, 12, 3, 1, 5, 6, 3, 4, 2, 4, 3, 7, 14, 10, 3, 5, and 9.

![Dot plot for items purchased]

The data are skewed right.
4. Construct a box-and-whisker plot to display the number of pets owned by a number of randomly chosen students. The data are 2, 0, 5, 1, 2, 1, 0, 8, 4, 3, 9, 1, 2, 3, and 1.

The data are skewed right.

5. Construct a box-and-whisker plot to display the scores on a recent science test. The data are 90, 95, 100, 70, 85, 65, 90, 80, 65, 70, 75, 80, 85, 80, 60, 80, 75, and 85.

The data are symmetric.

6. Construct a box-and-whisker plot to display the number of miles from school that a number of randomly chosen students live. The data are 5, 10, 15, 12, 1, 14, 9, 15, 3, 10, 12, 15, 8, 14, 13, and 2.

The data are skewed left.
7. Construct a histogram to display the circumferences of the pumpkins in the Jeffers’ family pumpkin crop. The data are 22.1, 35.6, 15.8, 36.9, 40.0, 28.5, 38.4, 20.4, 25.8, 34.1, 39.9, 42.2, 24.3, 22.7, 19.8, 27.9, 22.2, 34.3, 40.4, 20.6, 38.2, and 18.1. Use $10 \leq x < 20$ as the first interval.

The data are symmetric.

8. Construct a histogram to display the scores on a recent English quiz. The data are 18, 45, 20, 32, 9, 35, 49, 28, 25, 19, 5, 30, 22, 24, and 14. Use $0 \leq x < 10$ as the first interval.

The data are symmetric.
Analyze the given dot plot which displays the number of home runs by each of the girls on the softball team this season. Use the dot plot to answer each question.

9. Describe the distribution of the data in the dot plot and explain what it means in terms of the problem situation.
   The data are skewed right, because a majority of the data values are on the left of the plot and only a few of the data values are on the right of the plot. This means that a majority of the players on the softball team hit a small number of home runs, while only a few players on the team hit a large number of home runs.

10. How many players are on the softball team?
    There are 14 players on the softball team.

11. How many players hit more than 2 home runs?
    Five players hit more than 2 home runs.

12. How many players hit at least 1 home run?
    Ten players hit at least 1 home run.

13. How many players hit more than 1 and fewer than 9 home runs?
    Six players hit more than 1 and fewer than 9 home runs.

14. How many players scored more than 12 home runs?
    No players scored more than 12 home runs.
Analyze the given box-and-whisker plot which displays the heights of 40 randomly chosen adults. Use the box-and-whisker plot to answer each question.

Heights of 40 Randomly Chosen Adults

15. What is the height range of the middle 50 percent of the surveyed adults?
   The middle 50 percent of the surveyed adults are at least 65 inches and at most 72 inches tall.

16. How many of the surveyed adults are exactly 68 inches tall?
   It is not possible to determine the number of surveyed adults who are exactly 68 inches tall.

17. What percent of the surveyed adults are 68 inches tall or shorter?
   Fifty percent of the surveyed adults are 68 inches tall or shorter.

18. What is the height of the tallest adult surveyed?
   The tallest adult surveyed is 79 inches tall.

19. How many of the surveyed adults are at least 58 inches tall?
   One hundred percent of the surveyed adults are at least 58 inches tall. Therefore, all 40 of the surveyed adults are at least 58 inches tall.

20. Describe the distribution of the data in the box-and-whisker plot and explain what it means in terms of the problem situation.
   The data are symmetric, because the data are clustered around the center of the data values. Twenty-five percent of the data are in the lowest 7 inch range and twenty-five percent of the data are in the highest 7 inch range. Fifty percent of the data are in the middle 7 inch range.
Analyze the given histogram which displays the ACT composite score of several randomly chosen students. Use the histogram to answer each question.

21. How many students are represented by the histogram?
   There are a total of 31 students represented by the histogram.

22. Describe the distribution of the data in the histogram and explain what it means in terms of the problem situation.
   The data are symmetric, because a majority of the values are in the middle while the remaining values are spread out approximately evenly on both sides of the middle. Most of the students scored in the middle range of values while a smaller number of students had a very low or very high score.
23. How many of the students had an ACT composite score of exactly 25?  
   It is not possible to determine the number of students who scored exactly 25.

24. How many of the students had an ACT composite score of at least 20?  
   Twenty students had an ACT composite score of at least 20.

25. How many of the students had an ACT composite score less than 30?  
   Twenty-six students had an ACT composite score less than 30.

26. How many more students had an ACT composite score between 15 and 20 than had a composite score between 30 and 35?  
   Two more students had an ACT composite score between 15 and 20 than had a composite score between 30 and 35.
Which Measure Is Better?
Determining the Best Measure of Center for a Data Set

Vocabulary
Define each term in your own words.

1. statistics
   The numerical characteristics of a data set.

2. measure of central tendency
   A statistic that describes the “center” of a data set.

Problem Set
Create a dot plot of each given data set. Calculate the mean and median. Determine which measure of center best describes each data set.

1. The data are 1, 3, 2, 0, 7, 2, 1, 10, 1, 12, 1, 2, 0, 3, and 4.

\[
\bar{x} = \frac{\sum x}{n}
\]

\[
= \frac{0 + 0 + 1 + 1 + 1 + 1 + 2 + 2 + 7 + 10 + 12}{15}
\]

\[
\approx 3.27
\]

The mean is approximately 3.27 and the median is 2. The median is the best measure of center because the data are skewed right.
2. The data are 7, 2, 9, 9, 10, 12, 17, 10, 6, 11, 9, 10, 8, 11, and 8.

\[ \bar{x} = \frac{\sum x}{n} \]

\[ = \frac{2 + 6 + 7 + 8 + 8 + 9 + 9 + 10 + 10 + 10 + 11 + 11 + 12 + 17}{15} \]

\[ \approx 9.27 \]

The mean is approximately 9.27 and the median is 9. The mean is the best measure of center because the data are symmetric.

3. The data are 4, 0, 13, 15, 14, 10, 13, 8, 13, 12, 11, 13, 14, 1, 15, 13, 14, 12, 10, and 7.

\[ \bar{x} = \frac{\sum x}{n} \]

\[ = \frac{0 + 1 + 4 + 7 + 8 + 10 + 10 + 11 + 12 + 12 + 13 + 13 + 13 + 13 + 13 + 14 + 14 + 14 + 15 + 15}{20} \]

\[ = 10.6 \]

The mean is 10.6 and the median is 12.5. The median is the best measure of center because the data are skewed left.

4. The data are 50, 50, 40, 70, 60, 50, 20, 50, 80, 40, 60, 40, and 50.

\[ \bar{x} = \frac{\sum x}{n} \]

\[ = \frac{20 + 40 + 40 + 40 + 50 + 50 + 50 + 50 + 50 + 60 + 60 + 60 + 70 + 80}{13} \]

\[ \approx 50.77 \]

The mean is approximately 50.77 and the median is 50. The mean is the best measure of center because the data are symmetric.
5. The data are 40, 45, 48, 49, 50, 49, 47, 50, 49, 42, 49, 50, 48, 50, and 47.

\[ \bar{x} = \frac{\sum x}{n} \]
\[ = \frac{40 + 42 + 45 + 47 + 47 + 48 + 48 + 49 + 49 + 49 + 50 + 50 + 50 + 50}{15} \]
\[ \approx 47.53 \]

The mean is approximately 47.53 and the median is 49. The median is the best measure of center because the data are skewed left.

6. The data are 13, 12, 12, 11, 17, 10, 11, 12, 14, 20, 15, 12, 18, 13, 12, 17, 14, and 11.

\[ \bar{x} = \frac{\sum x}{n} \]
\[ = \frac{10 + 11 + 11 + 11 + 12 + 12 + 12 + 12 + 13 + 13 + 13 + 14 + 14 + 15 + 17 + 17 + 18 + 20}{18} \]
\[ \approx 13.56 \]

The mean is approximately 13.56 and the median is 12.5. The median is the best measure of center because the data are skewed right.
Determine which measure of center best describes the data in each given data display. Then determine the mean and median, if possible. If it is not possible, explain why not.

7. Average Annual Snowfall (inches)

Average Annual Snowfall in Select U.S. Cities

The mean is the best measure of center to describe the data because the data are symmetric. The mean and median cannot be determined because the data values are not given.

8. Math Quiz Scores

The median is the best measure of center to describe the data because the data are skewed left.

\[ \bar{x} = \frac{\sum x}{n} \]

\[ = \frac{0 + 5 + 8 + 10 + 11 + 11 + 12 + 12 + 12 + 13 + 13 + 13 + 13 + 14 + 14 + 15}{16} \]

\[ = 11 \]

The mean math quiz score is 11 and the median math quiz score is 12.
9. The median is the best measure of center to describe the data because the data are skewed right. The median number of movies watched last month is 6. The mean cannot be determined because the data values are not given.

10. The mean is the best measure of center to describe the data because the data are symmetric. The mean and median cannot be determined, because the data values are not given.
11. **Fishing Derby Results**

The median is the best measure of center to describe the data because the data are skewed right.

\[ \bar{x} = \frac{\sum x}{n} \]

\[ = \frac{0 + 0 + 1 + 1 + 2 + 2 + 2 + 2 + 3 + 3 + 4 + 5 + 8 + 11}{14} \]

\[ \approx 3.14 \]

The mean number of fish caught is approximately 3.14 and the median number of fish caught is 2.

12. **Results of Diving Expedition**

The median is the best measure of center to describe the data because the data are skewed left. The median number of sharks sighted is 11. The mean cannot be determined, because the data values are not given.
You Are Too Far Away!
Calculating IQR and Identifying Outliers

Vocabulary
Match each definition to its corresponding term.

1. interquartile range (IQR)
   a. A value calculated using the formula $Q_1 - (IQR \cdot 1.5)$.
   b. A value calculated using the formula $Q_3 - (IQR \cdot 1.5)$.

2. outlier
   b. A value calculated by subtracting $Q_1$ from $Q_3$.
   d. A data value that is significantly greater than or less than the other values in a data set.

3. lower fence
   a. $Q_3 + (IQR \cdot 1.5)$.
   c. A value calculated using the formula $Q_3 + (IQR \cdot 1.5)$.

4. upper fence
   c. $Q_3 + (IQR \cdot 1.5)$.
   d. A data value that is significantly greater than or less than the other values in a data set.
Problem Set

Calculate the IQR of each given data set. Determine whether there are any outliers in each set and list them.

1. The data are 4, 4, 5, 5, 8, 9, 10, 10, 12, 12, 16, 20, and 30.
   
   \[ Q_1 = 5, \quad Q_3 = 14 \]
   
   \[ \text{IQR} = Q_3 - Q_1 \]
   \[ = 14 - 5 \]
   \[ = 9 \]

   **Lower Fence:**
   \[ Q_1 - (\text{IQR} \cdot 1.5) = 5 - (9 \cdot 1.5) \]
   \[ = 5 - 13.5 \]
   \[ = -8.5 \]

   **Upper Fence:**
   \[ Q_3 + (\text{IQR} \cdot 1.5) = 14 + (9 \cdot 1.5) \]
   \[ = 14 + 13.5 \]
   \[ = 27.5 \]

   The value 30 is an outlier because it is greater than the upper fence.

2. The data are 0, 3, 10, 16, 16, 18, 20, 21, 22, 24, 25, 25, 27, 30, 35, and 41.
   
   \[ Q_1 = 16, \quad Q_3 = 26 \]
   
   \[ \text{IQR} = Q_3 - Q_1 \]
   \[ = 26 - 16 \]
   \[ = 10 \]

   **Lower Fence:**
   \[ Q_1 - (\text{IQR} \cdot 1.5) = 16 - (10 \cdot 1.5) \]
   \[ = 16 - 15 \]
   \[ = 1 \]

   **Upper Fence:**
   \[ Q_3 + (\text{IQR} \cdot 1.5) = 26 + (10 \cdot 1.5) \]
   \[ = 26 + 15 \]
   \[ = 41 \]

   The value 0 is an outlier because it is less than the lower fence.
3. The data are 9, 15, 26, 30, 32, 32, 35, 36, 38, 40, 40, 45, and 59.

\[ Q_1 = 28, \quad Q_3 = 40 \]

\[ \text{IQR} = Q_3 - Q_1 
\]
\[ = 40 - 28 
\]
\[ = 12 \]

Lower Fence:
\[ Q_1 - (\text{IQR} \cdot 1.5) = 28 - (12 \cdot 1.5) \]
\[ = 28 - 18 \]
\[ = 10 \]

Upper Fence:
\[ Q_3 + (\text{IQR} \cdot 1.5) = 40 + (12 \cdot 1.5) \]
\[ = 40 + 18 \]
\[ = 58 \]

The value 9 is an outlier because it is less than the lower fence. The value 59 is an outlier because it is greater than the upper fence.

4. The data are 18, 25, 30, 32, 33, 33, 35, 38, 39, 40, 42, 43, 44, 48, and 55.

\[ Q_1 = 32, \quad Q_3 = 43 \]

\[ \text{IQR} = Q_3 - Q_1 
\]
\[ = 43 - 32 
\]
\[ = 11 \]

Lower Fence:
\[ Q_1 - (\text{IQR} \cdot 1.5) = 32 - (11 \cdot 1.5) \]
\[ = 32 - 16.5 \]
\[ = 15.5 \]

Upper Fence:
\[ Q_3 + (\text{IQR} \cdot 1.5) = 43 + (11 \cdot 1.5) \]
\[ = 43 + 16.5 \]
\[ = 59.5 \]

There are no outliers in the data set because there are no values less than the lower fence or greater than the upper fence.
5. The data are 22, 19, 20, 21, 25, 10, 18, 28, 32, 24, and 25.
   \[Q1 = 18.5, \quad Q3 = 25\]
   \[IQR = Q3 - Q1 = 25 - 18.5 = 6.5\]

   Lower Fence: \[Q1 - (IQR \cdot 1.5) = 18.5 - (6.5 \cdot 1.5) = 18.5 - 9.75 = 8.75\]
   Upper Fence: \[Q3 + (IQR \cdot 1.5) = 25 + (6.5 \cdot 1.5) = 25 + 9.75 = 34.75\]

   The value 8 is an outlier because it is less than the lower fence.

6. The data are 60, 55, 70, 80, 20, 60, 105, 65, 75, 100, 55, 15, 115, 65, 70, 45, and 60.
   \[Q1 = 55, \quad Q3 = 77.5\]
   \[IQR = Q3 - Q1 = 77.5 - 55 = 22.5\]

   Lower Fence: \[Q1 - (IQR \cdot 1.5) = 55 - (22.5 \cdot 1.5) = 55 - 33.75 = 21.25\]
   Upper Fence: \[Q3 + (IQR \cdot 1.5) = 77.5 + (22.5 \cdot 1.5) = 77.5 + 33.75 = 111.25\]

   The values 15 and 20 are outliers because they are less than the lower fence. The value 115 is an outlier because it is greater than the upper fence.
Calculate the IQR of the data set represented in each box-and-whisker plot and determine whether there are any outliers in each data set.

7.

\[ Q1 = 7, \quad Q3 = 10 \]
\[ IQR = Q3 - Q1 \]
\[ = 10 - 7 \]
\[ = 3 \]

**Lower Fence:**
\[ Q1 - (IQR \cdot 1.5) = 7 - (3 \cdot 1.5) \]
\[ = 7 - 4.5 \]
\[ = 2.5 \]

**Upper Fence:**
\[ Q3 + (IQR \cdot 1.5) = 10 + (3 \cdot 1.5) \]
\[ = 10 + 4.5 \]
\[ = 14.5 \]

There is at least 1 outlier less than the lower fence because the minimum value of the data set is 1.

8.

\[ Q1 = 16, \quad Q3 = 20 \]
\[ IQR = Q3 - Q1 \]
\[ = 20 - 16 \]
\[ = 4 \]

**Lower Fence:**
\[ Q1 - (IQR \cdot 1.5) = 16 - (4 \cdot 1.5) \]
\[ = 16 - 6 \]
\[ = 10 \]

**Upper Fence:**
\[ Q3 + (IQR \cdot 1.5) = 20 + (4 \cdot 1.5) \]
\[ = 20 + 6 \]
\[ = 26 \]

There are no outliers in the data set.
9.

Q1 = 45, Q3 = 60
IQR = Q3 − Q1
   = 60 − 45
   = 15

Lower Fence:
Q1 − (IQR · 1.5) = 45 − (15 · 1.5)
   = 45 − 22.5
   = 22.5

Upper Fence:
Q3 + (IQR · 1.5) = 60 + (15 · 1.5)
   = 60 + 22.5
   = 82.5

There is at least 1 outlier less than the lower fence because the minimum value of the data set is 15. There is at least 1 outlier greater than the upper fence because the maximum value of the data set is 90.

10.

Q1 = 7, Q3 = 12
IQR = Q3 − Q1
   = 12 − 7
   = 5

Lower Fence:
Q1 − (IQR · 1.5) = 7 − (5 · 1.5)
   = 7 − 7.5
   = −0.5

Upper Fence:
Q3 + (IQR · 1.5) = 12 + (5 · 1.5)
   = 12 + 7.5
   = 19.5

There is at least 1 outlier greater than the upper fence because the maximum value of the data set is 20.
11. 

\[ Q1 = 350, \quad Q3 = 550 \]
\[ IQR = Q3 - Q1 \]
\[ = 550 - 350 \]
\[ = 200 \]

**Lower Fence:**
\[ Q1 - (IQR \cdot 1.5) = 350 - (200 \cdot 1.5) \]
\[ = 350 - 300 \]
\[ = 50 \]

**Upper Fence:**
\[ Q3 + (IQR \cdot 1.5) = 550 + (200 \cdot 1.5) \]
\[ = 550 + 300 \]
\[ = 850 \]

There is at least 1 outlier less than the lower fence because the minimum value of the data set is 0.

12. 

\[ Q1 = 4, \quad Q3 = 10 \]
\[ IQR = Q3 - Q1 \]
\[ = 10 - 4 \]
\[ = 6 \]

**Lower Fence:**
\[ Q1 - (IQR \cdot 1.5) = 4 - (6 \cdot 1.5) \]
\[ = 4 - 9 \]
\[ = -5 \]

**Upper Fence:**
\[ Q3 + (IQR \cdot 1.5) = 10 + (6 \cdot 1.5) \]
\[ = 10 + 9 \]
\[ = 19 \]

There are no outliers in the data set.
Whose Scores Are Better?  
Calculating and Interpreting Standard Deviation

Vocabulary

Define each term in your own words.

1. standard deviation
   Standard deviation is a measure of how spread out the data are from the mean. A lower standard deviation represents data that are more tightly clustered. A higher standard deviation represents data that are more spread out from the mean.

2. normal distribution
   A collection of many data points that form a bell-shaped curve.

Problem Set

Calculate the mean and the standard deviation of each data set without the use of a calculator.

1. The data are 0, 3, 6, 7, and 9.

   \[ \bar{x} = \frac{0 + 3 + 6 + 7 + 9}{5} = \frac{25}{5} = 5 \]

   \[ \sigma = \sqrt{\frac{25 + 4 + 1 + 4 + 16}{5}} = \sqrt{\frac{50}{5}} = \sqrt{10} \approx 3.16 \]

   \( (x_1 - \bar{x})^2 = (0 - 5)^2 = 25 \)
   \( (x_2 - \bar{x})^2 = (3 - 5)^2 = 4 \)
   \( (x_3 - \bar{x})^2 = (6 - 5)^2 = 1 \)
   \( (x_4 - \bar{x})^2 = (7 - 5)^2 = 4 \)
   \( (x_5 - \bar{x})^2 = (9 - 5)^2 = 16 \)

   The mean is 5. The standard deviation is approximately 3.16.
2. The data are 6, 8, 9, 10, 10, and 11.
\[
\bar{x} = \frac{6 + 8 + 9 + 10 + 10 + 11}{6} = \frac{54}{6} = 9
\]
\[
\sigma = \sqrt{\frac{9 + 1 + 0 + 1 + 1 + 4}{6}} = \sqrt{\frac{16}{6}} = \sqrt{2.67} = 1.63
\]
\[
(x_1 - \bar{x})^2 = (6 - 9)^2 = 9
\]
\[
(x_2 - \bar{x})^2 = (8 - 9)^2 = 1
\]
\[
(x_3 - \bar{x})^2 = (9 - 9)^2 = 0
\]
\[
(x_4 - \bar{x})^2 = (10 - 9)^2 = 1
\]
\[
(x_5 - \bar{x})^2 = (10 - 9)^2 = 1
\]
\[
(x_6 - \bar{x})^2 = (11 - 9)^2 = 4
\]
The mean is 9. The standard deviation is approximately 1.63.

3. The data are 1, 5, 10, 15, 16, 20, and 24.
\[
\bar{x} = \frac{1 + 5 + 10 + 15 + 16 + 20 + 24}{7} = \frac{91}{7} = 13
\]
\[
\sigma = \sqrt{\frac{144 + 64 + 9 + 4 + 9 + 49 + 121}{7}} = \sqrt{\frac{400}{7}} = \sqrt{57.14} = 7.56
\]
\[
(x_1 - \bar{x})^2 = (1 - 13)^2 = 144
\]
\[
(x_2 - \bar{x})^2 = (5 - 13)^2 = 64
\]
\[
(x_3 - \bar{x})^2 = (10 - 13)^2 = 9
\]
\[
(x_4 - \bar{x})^2 = (15 - 13)^2 = 4
\]
\[
(x_5 - \bar{x})^2 = (16 - 13)^2 = 9
\]
\[
(x_6 - \bar{x})^2 = (20 - 13)^2 = 49
\]
\[
(x_7 - \bar{x})^2 = (24 - 13)^2 = 121
\]
The mean is 13. The standard deviation is approximately 7.56.
4. The data are 13, 14, 15, 15, 16, 16, 17, and 18.

\[ \bar{x} = \frac{13 + 14 + 15 + 15 + 16 + 16 + 17 + 18}{8} = \frac{124}{8} = 15.5 \]

\[ (x_1 - \bar{x})^2 = (13 - 15.5)^2 = 6.25 \]
\[ (x_2 - \bar{x})^2 = (14 - 15.5)^2 = 2.25 \]
\[ (x_3 - \bar{x})^2 = (15 - 15.5)^2 = 0.25 \]
\[ (x_4 - \bar{x})^2 = (15 - 15.5)^2 = 0.25 \]
\[ (x_5 - \bar{x})^2 = (16 - 15.5)^2 = 0.25 \]
\[ (x_6 - \bar{x})^2 = (16 - 15.5)^2 = 0.25 \]
\[ (x_7 - \bar{x})^2 = (17 - 15.5)^2 = 2.25 \]
\[ (x_8 - \bar{x})^2 = (18 - 15.5)^2 = 6.25 \]

The mean is 15.5. The standard deviation is 1.5.

5. The data are represented by a dot plot.

\[ \bar{x} = \frac{2 + 3 + 3 + 4 + 4 + 4 + 5 + 5 + 6}{9} = \frac{36}{9} = 4 \]

\[ (x_1 - \bar{x})^2 = (2 - 4)^2 = 4 \]
\[ (x_2 - \bar{x})^2 = (3 - 4)^2 = 1 \]
\[ (x_3 - \bar{x})^2 = (3 - 4)^2 = 1 \]
\[ (x_4 - \bar{x})^2 = (4 - 4)^2 = 0 \]
\[ (x_5 - \bar{x})^2 = (4 - 4)^2 = 0 \]
\[ (x_6 - \bar{x})^2 = (4 - 4)^2 = 0 \]
\[ (x_7 - \bar{x})^2 = (5 - 4)^2 = 1 \]
\[ (x_8 - \bar{x})^2 = (5 - 4)^2 = 1 \]
\[ (x_9 - \bar{x})^2 = (6 - 4)^2 = 4 \]

The mean is 4. The standard deviation is approximately 1.15.
6. The data are represented by a dot plot.

\[ \bar{x} = \frac{0 + 2 + 5 + 5 + 6 + 8 + 10 + 12}{8} = \frac{48}{8} = 6 \]

\[
\begin{align*}
(x_1 - \bar{x})^2 &= (0 - 6)^2 = 36 \\
(x_2 - \bar{x})^2 &= (2 - 6)^2 = 16 \\
(x_3 - \bar{x})^2 &= (5 - 6)^2 = 1 \\
(x_4 - \bar{x})^2 &= (5 - 6)^2 = 1 \\
(x_5 - \bar{x})^2 &= (6 - 6)^2 = 0 \\
(x_6 - \bar{x})^2 &= (8 - 6)^2 = 4 \\
(x_7 - \bar{x})^2 &= (10 - 6)^2 = 16 \\
(x_8 - \bar{x})^2 &= (12 - 6)^2 = 36
\end{align*}
\]

\[\sigma = \sqrt{\frac{36 + 16 + 1 + 1 + 0 + 4 + 16 + 36}{8}} = \sqrt{\frac{110}{8}} \approx 3.71\]

The mean is 6. The standard deviation is approximately 3.71.

Calculate the mean and the standard deviation of each given data set using a graphing calculator.

7. The data are 1, 3, 4, 6, 6, 8, 9, 10, and 12.
   The mean is approximately 6.56. The standard deviation is approximately 3.34.

8. The data are 18, 20, 24, 25, 26, 28, 30, 32, and 35.
   The mean is 26.4. The standard deviation is approximately 4.90.

9. The data are 102, 103, 103, 104, 104, 105, 105, 106, 106, and 107.
   The mean is approximately 104.45. The standard deviation is approximately 1.44.

10. The data are 3.5, 4, 5.5, 6, 6, 7, 7.5, 8, 9.5, and 10.5.
    The mean is 6.75. The standard deviation is approximately 2.11.
11. The data are represented by a dot plot.

```
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
X   X   X   X   X   X   X   X   X   X   X
```

The mean is 7.9. The standard deviation is approximately 3.42.

12. The data are represented by a dot plot.

```
40 41 42 43 44 45 46
X   X   X   X   X   X   X
```

The mean is approximately 43.07. The standard deviation is approximately 1.58.
Putting the Pieces Together
Analyzing and Interpreting Data

Vocabulary

For each problem situation, identify whether a stem-and-leaf plot or a side-by-side stem-and-leaf plot would be appropriate. Explain your choice for each.

1. For a history project, Roberto is comparing the ages of the U.S. Presidents at inauguration and at death.
   For this situation, it would be best to use a side-by-side stem-and-leaf plot because Roberto is comparing two data sets.

2. During the Summer Olympic Games, Karen keeps track of the number of gold medals won by the various countries participating.
   For this situation, it would be best to use a stem-and-leaf plot because Karen is only ordering one set of numerical data.

Problem Set

Construct a box-and-whisker plot of each given data set and include any outliers. Calculate the most appropriate measure of center and spread for each data set based on the data distribution.

1. The data are 0, 2, 3, 4, 4, 5, 5, 5, 6, 6, 8, and 9.
   The most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric.
   The mean is 4.75 and the standard deviation is approximately 2.35.

2. The data are 1, 6, 9, 12, 14, 15, 17, 17, 17, 18, 18, 18, 19, and 20.
   The most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed left.
   The median is 17 and the IQR is 6.
3. The data are 50, 53, 57, 58, 58, 59, 59, 60, 60, 60, 61, 61, 62, 63, and 67.

The most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric.
The mean is 59.2 and the standard deviation is approximately 3.85.

4. The data are 20, 20, 20, 21, 21, 21, 22, 22, 23, 24, 25, 28, and 30.

The most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed right.
The median is 22 and the IQR is 4.

5. The data are 80, 85, 90, 30, 70, 90, 95, 10, 100, 70, 80, 55, 50, 95, 65, and 90.

The most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed left.
The median is 80 and the IQR is 30.

6. The data are 7, 11, 10, 13, 0, 3, 10, 9, 17, 11, 10, 20, 9, 8, and 12.

The most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric.
The mean is 10 and the standard deviation is approximately 4.68.
Two data sets are given in a side-by-side stem-and-leaf plot. Calculate the most appropriate measure of center and spread for each set based on the data distribution.

7. **Data Set 1**
   - 8 7 5 3
   - 8 4 1
   - 2 2
   - 0 3

   **Data Set 2**
   - 0 3 4 6 9
   - 1 4 5
   - 2
   - 2

   Key: 2|5 = 25

   For each data set, the most appropriate measure of center is the median and the most appropriate measure of spread is the IQR, because the data are skewed right.
   - For Data Set 1, the median is 11 and the IQR is 14.
   - For Data Set 2, the median is 11 and the IQR is 15.

8. **Data Set 1**
   - 9 8 7
   - 8 7 5 5 3
   - 2 1 0

   **Data Set 2**
   - 1 8 9
   - 0 2 4 4 6 9
   - 3 1 3

   Key: 2|0 = 20

   For each data set, the most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric.
   - For Data Set 1, the mean is 25 and the standard deviation is approximately 5.01.
   - For Data Set 2, the mean is 24.6 and the standard deviation is approximately 4.86.
For Data Set 1, the most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric. For Data Set 1, the mean is 69.9 and the standard deviation is approximately 8.40.

For Data Set 2, the most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed left. For Data Set 2, the median is 74 and the IQR is 16.5.

For each data set, the most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed left.

For Data Set 1, the median is 39.5 and the IQR is 14.

For Data Set 2, the median is 38 and the IQR is 17.5.
11. | Data Set 1 | Data Set 2 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9 9 9 8 8 8 3 7 9</td>
<td></td>
</tr>
<tr>
<td>8 6 5 2 4 0 3 4 5 8</td>
<td></td>
</tr>
<tr>
<td>5 1 5 2 5 6 6</td>
<td></td>
</tr>
<tr>
<td>0 6 1 2</td>
<td></td>
</tr>
</tbody>
</table>

Key: 4|0 = 40

For Data Set 1, the most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed right. For Data Set 1, the median is 42 and the IQR is 11.

For Data Set 2, the most appropriate measure of center is the mean, and the most appropriate measure of spread is the standard deviation because the data are symmetric. For Data Set 2, the mean is 49.5 and the standard deviation is approximately 7.98.

12. | Data Set 1 | Data Set 2 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7 5 3 3 2 1 10 1 1 3 5 8 8 9</td>
<td></td>
</tr>
<tr>
<td>8 5 2 11 1 3 5</td>
<td></td>
</tr>
<tr>
<td>9 4 12 5</td>
<td></td>
</tr>
<tr>
<td>2 13 3</td>
<td></td>
</tr>
<tr>
<td>0 14</td>
<td></td>
</tr>
</tbody>
</table>

Key: 12|5 = 125

For each data set, the most appropriate measure of center is the median, and the most appropriate measure of spread is the IQR because the data are skewed right. For Data Set 1, the median is 112 and the IQR is 23.5. For Data Set 2, the median is 108.5 and the IQR is 10.