Lesson 8: Describing Distributions on Histograms

Goals

- Compare and contrast (orally) bar graphs and histograms, recognizing that descriptions of shape, center, and spread don't pertain to bar graphs.
- Describe (orally and in writing) the overall shape and features of a distribution represented on a histogram, including peaks, clusters, gaps, and symmetry.
- Identify histograms that display distributions with specific features.

Learning Targets

- I can describe the shape and features of a histogram and explain what they mean in the context of the data.
- I can distinguish histograms and bar graphs.

Lesson Narrative

In this lesson, students explore various shapes and features of a distribution displayed in a histogram. They use the structure (MP7) to look for symmetry, peaks, clusters, gaps, and any unusual values in histograms. Students also begin to consider how these features might affect how we characterize a data set. For example, how might we describe what is typical in a distribution that shows symmetry? What about in a distribution that has one peak that is not symmetrical? This work is informal, but helps to prepare students to better understand measures of center and spread later in the unit. Students also distinguish between the uses and construction of bar graphs and histograms in this lesson.

Alignments

Addressing

- 6.SP.A.2: Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Building Towards

- 6.SP.B.5.d: Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- MLR2: Collect and Display
- MLR5: Co-Craft Questions

- Which One Doesn’t Belong?

Required Materials

Pre-printed cards, cut from copies of the blackline master

Required Preparation

Print and cut up cards from the Sorting Histograms blackline master. Prepare 1 set of cards for every 3–4 students.

The Getting to School activity requires students to use data previously collected on their travel methods and times. Organize the data into the tables in the blackline master ahead of time or allow time for students to do it themselves. Either make a copy for every 2 students, or display the completed tables for all to see during the activity.

Student Learning Goals

Let's describe distributions displayed in histograms.

8.1 Which One Doesn’t Belong: Histograms

Warm Up: 5 minutes

This warm-up encourages students to make sense of histograms in terms of center and spread. It prompts students to hold mathematical conversations and explain their reasoning (MP3), and gives the teacher the opportunity to hear how students compare data sets represented by histograms.

Addressing

- 6.SP.A.2

Instructional Routines

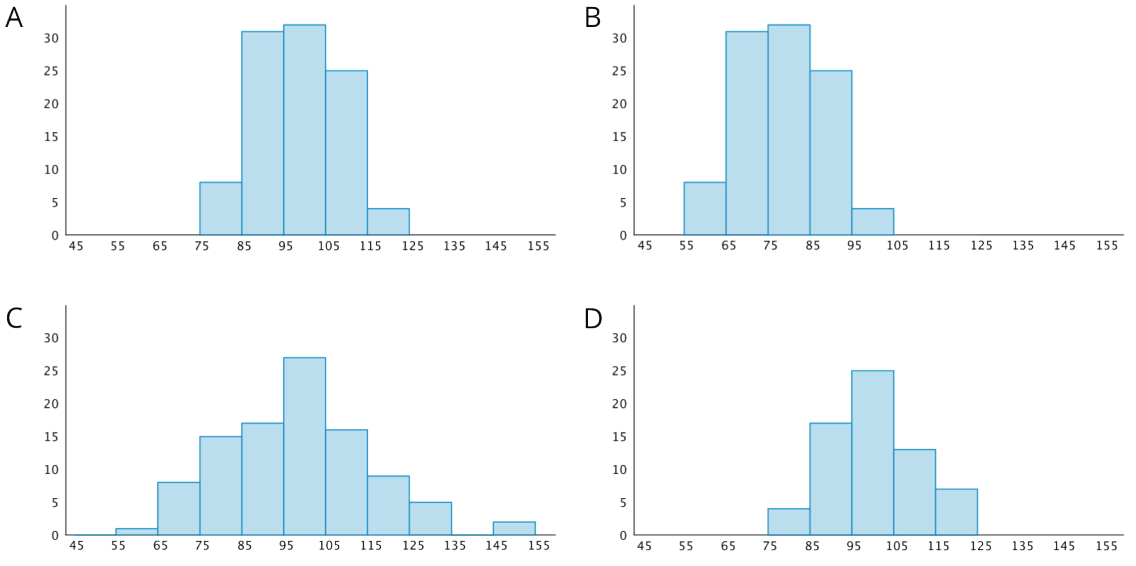
- Which One Doesn’t Belong?

Launch

Arrange students in groups of 2–4. Display the images for all to see. Give students 1 minute of quiet think time and ask them to indicate when they have noticed one image that does not belong and can explain why. When the minute is up, give students 2 minutes to share their thinking with their small group, and then, together, find a reason that each image doesn't belong.

Student Task Statement

Which histogram does not belong? Be prepared to explain your reasoning.



Student Response

Answers vary. Sample responses:

- Histogram B does not belong. Unlike the others, its distribution is not centered around 100.
- Histogram C does not belong. The spread of the data is much wider than that of the other histograms.
- Histogram D does not belong. It represents a smaller set of data compared to the others.

Activity Synthesis

Ask students to share one reason their group decided a particular image does not belong. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer to the question of which one does not belong, attend to students' explanations and ensure the reasons given are reasonable.

If students use terms that are essential in this unit (such as center, spread, distribution, frequency, etc.), ask them to explain their meanings in their own words; these are opportunities to reinforce their understanding of the terms and to note any misconceptions. If students give unsubstantiated claims, ask them to substantiate them.

8.2 Sorting Histograms

20 minutes

This activity is designed to expand both students' exposure to various features of distributions and the language they could use to describe distributions. Students sort histograms based on features such as symmetry, gaps, clusters, and unusual values. In earlier grades, students used the term

“symmetry” to describe geometric figures (4.G.3); here they use it to describe the shape of a distribution.

Addressing

- 6.SP.A.2

Building Towards

- 6.SP.B.5.d

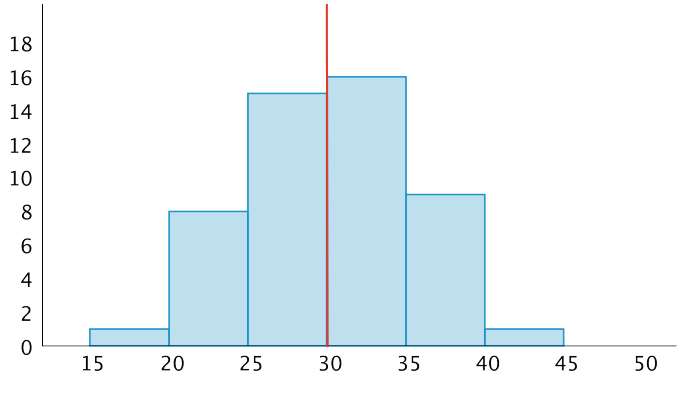
Instructional Routines

- MLR5: Co-Craft Questions

Launch

Display the image of the histogram here for all to see. Explain to students that a diagram of a distribution—a dot plot or a histogram—is described as symmetrical if you can draw a line on the diagram and the parts on one side of the line mirror the parts on the other side. Many distributions are not perfectly symmetrical, but are close to or approximately symmetrical.

The histogram here shows an approximately symmetrical distribution. When a line is drawn at the center (such as the line at 30) the two sides are roughly mirror images. If you were to fold the histogram at the line, the two sides would be close to matching.



Tell students that symmetry is used to describe distributions, and that they will now look for symmetry and other features of distributions in a dozen histograms.

Arrange students in groups of 3–4. Give each group one set of pre-cut cards from the blackline master. Ask students to study the histograms and identify the features described on their task statement. Give groups 10–12 minutes to complete the activity. Explain that each group will need to discuss their work with another group after the first question and before completing the rest of the activity.

Access for English Language Learners

Representing, Conversing: MLR5 Co-Craft Questions. Begin the launch by inviting students to write mathematical questions about the histogram displayed. Give students 1–2 minutes to share their questions with a partner. Listen for students that use terms center, spread, distribution, or frequency in their questions. Invite students to share their questions with the class, and to compare the language of their questions before continuing with the activity.
Design Principle(s): Optimize output (for generalization); Cultivate conversation

Student Task Statement

1. Your teacher will give your group a set of histogram cards. Sort them into two piles—one for histograms that are approximately symmetrical, and another for those that are not.
2. Discuss your sorting decisions with another group. Do both groups agree which cards should go in each pile? If not, discuss the reasons behind the differences and see if you can reach agreement. Record your final decisions.
    - Histograms that are approximately symmetrical:
    - Histograms that are not approximately symmetrical:
3. Histograms are also described by how many major peaks they have. Histogram A is an example of a distribution with a single peak that is not symmetrical.

    Which other histograms have this feature?
4. Some histograms have a gap, a space between two bars where there are no data points. For example, if some students in a class have 7 or more siblings, but the rest of the students have 0, 1, or 2 siblings, the histogram for this data set would show gaps between the bars because no students have 3, 4, 5, or 6 siblings.

    Which histograms do you think show one or more gaps?
5. Sometimes there are a few data points in a data set that are far from the center. Histogram A is an example of a distribution with this feature.

    Which other histograms have this feature?

Student Response

Answers vary. Sample responses:

1. No answers required.

2. Approximately symmetrical: Histograms F, I, and J. Students may consider Histograms C, E and K as approximately symmetrical as well. Not approximately symmetrical: Histograms other than the ones previously listed.
3. One peak, not symmetrical: Histograms B. Some may also include Histogram L.
4. With gaps: Histograms A, D, E, G, K and L.
5. With values far from the center: Histograms A, E, G and L. Some may consider including Histogram D, but because there are quite a few data points in the upper group, this really looks more like a distribution that just has a gap.

Activity Synthesis

Students will have had a chance to discuss the different features of a distribution in small groups. Use the whole-class discussion to prompt students to think about what the features might mean, and whether or how they affect the way we characterize a distribution. Remind students that we have been using the center of a distribution to talk about what is typical in a group. Discuss some of these questions:

- "Look at the histograms that you think show symmetry. When a distribution is approximately symmetrical, where might its center be?"
- "Now look at the histograms that you think are not approximately symmetrical. Where might its center be? How might we describe what is typical of a group that has one peak that is not symmetrical (such as that in Histogram B)?"
- "Look at the histograms that show gaps. How might a gap (such as that in Histogram K) affect our description of what is typical in a group?"
- "Look at the histograms that have values that are far away from other values. Do unusual values (such as those in Histogram G) affect our description of center and spread? If those unusual values weren't there, would our description of center and spread change?"

Expect students' answers to be very informal. The goal of the discussion is to raise students' awareness that the shape and features of distributions may affect how we characterize the data. This experience provides a conceptual foundation that would help students make sense of measures of center (mean and median) and measures of spread (mean absolute deviation, interquartile range, and range) later.

8.3 Getting to School

10 minutes

In this activity, students draw a bar graph and histogram, then describe the distributions shown on each display. Although the two visual displays may appear similar at first glance, there are important distinctions between the representations. Students notice differences in how we might characterize distributions in bar graphs and those in histograms, including how we describe typical

values or categories. Along the way, students consolidate their understanding about categorical and numerical data.

Addressing

- 6.SP.A.2
- 6.SP.B.4

Instructional Routines

- MLR2: Collect and Display

Launch

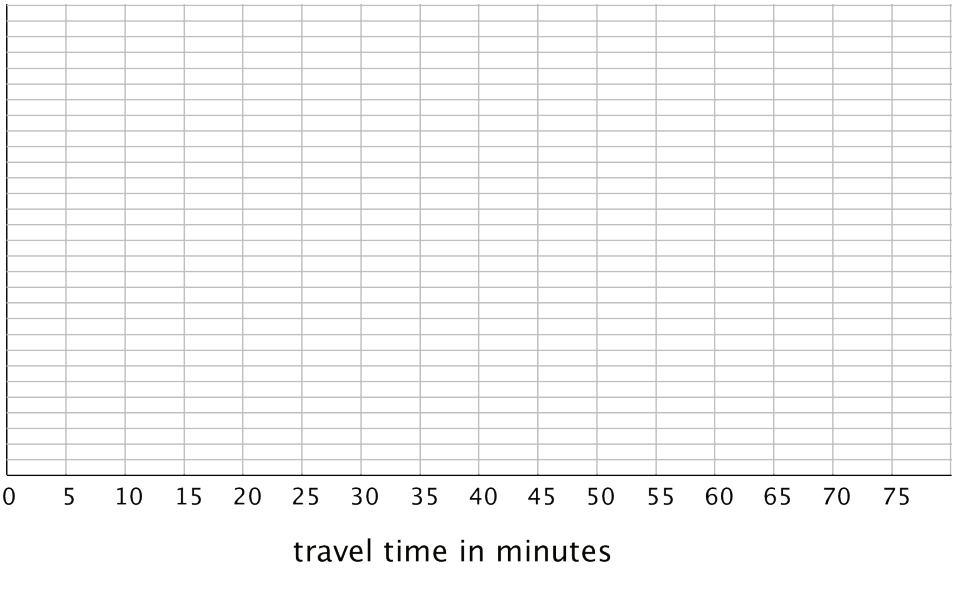
Students will need the data on their travel methods and times, collected at the beginning of the unit. Distribute or display the data collected for these questions from the survey given earlier in the unit. Alternatively, complete the tables in the blackline master ahead of time.

Arrange students in groups of 2. Give one copy of the blackline master to each group of students. Display the data from the prior survey or the completed frequency tables for all to see or give a copy to each group of 2 students. Give students 5–6 minutes to complete the activity. Ask one partner to create a bar graph to represent the data on the class's travel methods and the other to create a histogram to represent the data on travel times, and then answer the questions together.

Student Task Statement

Your teacher will provide you with some data that your class collected the other day.

1. Use the data to draw a histogram that shows your class's travel times.



2. Describe the distribution of travel times. Comment on the center and spread of the data, as well as the shape and features.

3. Use the data on methods of travel to draw a bar graph. Include labels for the horizontal axis.

4. Describe what you learned about your class's methods of transportation to school. Comment on any patterns you noticed.

5. Compare the histogram and the bar graph that you drew. How are they the same? How are they different?

Student Response

1. Answers and graphs vary based on class data.
2. Answers vary. Students should describe the distribution in terms of center, spread, shape, or other previously discussed features of distribution.
3. Answers and graphs vary based on class data.
4. Answers vary. Students may identify categories that are most or least prevalent. They may also describe the distribution in terms fraction or percentage of data values, or point out both common and unusual characteristics
5. Answers vary. Sample responses:
    - Bar graphs and histograms both use heights of bars to show frequency—the more frequently a value appears, the taller the bar that represents it.
    - The graphs are different in that histograms show numerical data and group data values into "bins" (so we cannot tell how many of a particular value there are). Bar graphs show categorical data, so the horizontal axis shows different categories or labels, rather than a number line. The frequency tells us exactly how many times a particular category appears in the data set.

Are You Ready for More?

Use one of these suggestions (or make up your own). Research data to create a histogram. Then, describe the distribution.

- Heights of 30 athletes from multiple sports
- Heights of 30 athletes from the same sport
- High temperatures for each day of the last month in a city you would like to visit
- Prices for all the menu items at a local restaurant

Student Response

Answers vary.

Activity Synthesis

The purpose of the discussion is for students to recognize the differences between histograms and bar graphs.

Select one student to show a completed histogram and another to show a completed bar graph. Then, solicit several observations about how the two graphical displays compare. Ask questions such as:

- "How are the bar graphs and histograms alike? How are they different?"
- "Can we use a bar graph to display the data on travel times? Why or why not?"
- "Can we use a histogram to display the data on methods of travel? Why or why not?"

Next, select a few other students to share their descriptions of the distributions shown on each type of display. Then, ask questions such as:

- "How are your descriptions of the distribution for travel methods different than those for travel times?"
- "Can you talk about the shape of a distribution shown on a bar graph? Can you talk about its center and spread? Why or why not?"

Students should recognize that only the distribution of numerical data can be described in terms of shape, center, or spread. We cannot analyze these features for a distribution of a categorical data on a bar graph because a bar graph does *not* use a number line. This means the bars can be drawn anywhere, in any order, and with any kind of spacing, so shape, center, and spread would have no meaning.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Provide students with a graphic organizer to record differences between histograms and bar graphs. Consider creating a shared display and keeping it visible throughout the unit.
Supports accessibility for: Language; Organization

Access for English Language Learners

Representing, Conversing: MLR2 Collect and Display. As students discuss the similarities and differences between histograms and bar graphs, record and display phrases you hear in a chart, such as a Venn diagram. Use this as a reference when highlighting or revoicing academic vocabulary from this unit (frequency, bins, numerical data, categorical data). This will help students connect unique and shared characteristics of histograms and bar graphs while reinforcing mathematical language.
Design Principle(s): Maximize meta-awareness

Lesson Synthesis

In this lesson, we look at the shapes and features of distributions that are represented by histograms.

- "What does it mean for a histogram to have symmetry?"
- "What is a 'peak' in a distribution? Is it always in the middle, or can it be to one side?"
- "Can a distribution have more than one peak?"
- "What does it mean for a histogram to show a cluster (or more than one clusters)?"
- "What does it mean for a histogram to show a gap?"

We also contrast bar graphs and histograms.

- "When do we use a histogram and when do we use a bar graph?"
- "What are the major differences between how a histogram is drawn and how a bar graph is drawn?"

8.4 Point Spread

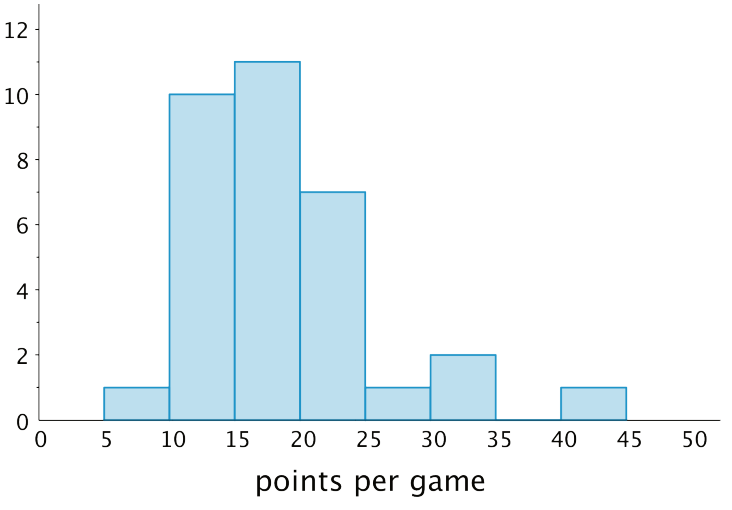
Cool Down: 5 minutes

Addressing

- 6.SP.A.2

Student Task Statement

Here is a histogram that shows the number of points scored by a college basketball player during the 2008 season. Describe the shape and features of the data.

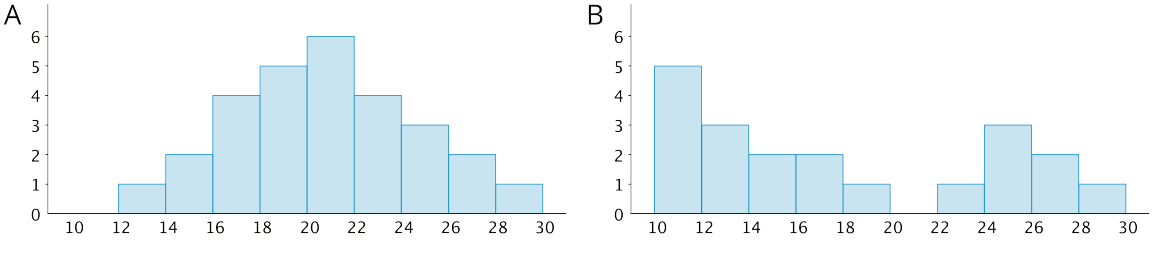


Student Response

Answers vary. Sample response: The data are not symmetrical, there is a peak on the left. The histogram shows a gap between 35 and 40, so there is no game where the player scored 35, 36, 37, 38 or 39 points. There was one game that was unusually high scoring, between 40 and 44 points. The peak is between 15 and 20 points. The center is around 20 points and the data is spread out pretty far above this center.

Student Lesson Summary

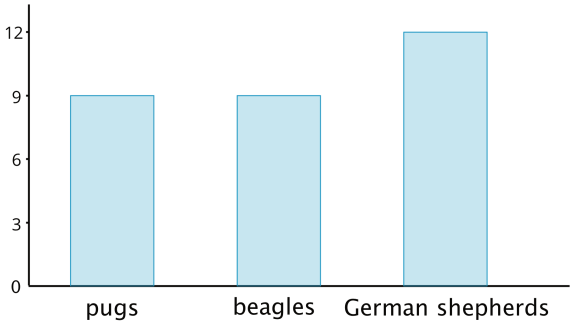
We can describe the shape and features of the distribution shown on a histogram. Here are two distributions with very different shapes and features.

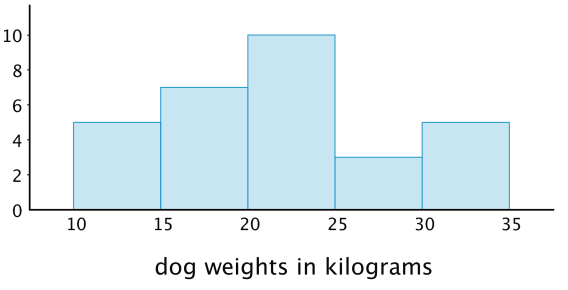


- Histogram A is very symmetrical and has a peak near 21. Histogram B is not symmetrical and has two peaks, one near 11 and one near 25.
- Histogram B has two clusters. A cluster forms when many data points are near a particular value (or a neighborhood of values) on a number line.

- Histogram B also has a gap between 20 and 22. A gap shows a location with no data values.

Here is a bar graph showing the breeds of 30 dogs and a histogram for their weights.





Bar graphs and histograms may seem alike, but they are very different.

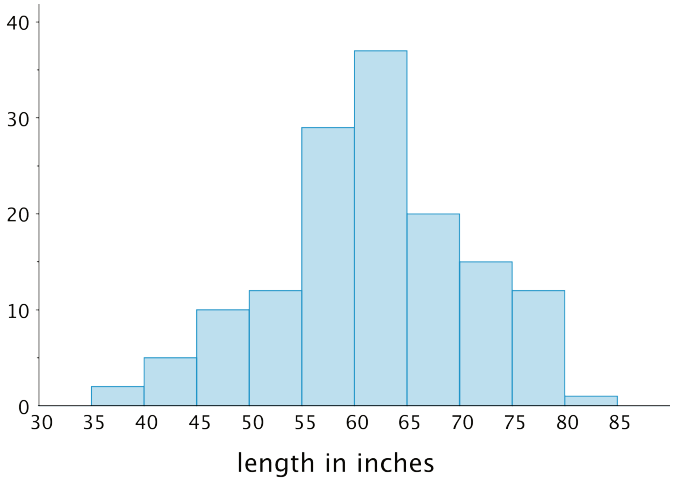
- Bar graphs represent categorical data. Histograms represent numerical data.
- Bar graphs have spaces between the bars. Histograms show a space between bars *only* when no data values fall between the bars.
- Bars in a bar graph can be in any order. Histograms must be in numerical order.
- In a bar graph, the number of bars depends on the number of categories. In a histogram, we choose how many bars to use.

Lesson 8 Practice Problems

Problem 1

Statement

The histogram summarizes the data on the body lengths of 143 wild bears. Describe the distribution of body lengths. Be sure to comment on shape, center, and spread.



Solution

Answers vary. Sample response: The distribution of body lengths is approximately symmetrical. A typical body length for the bears in the group studied is about 60 inches. There is a lot of variability in the body lengths of the bears, with the shortest length being somewhere between 35 and 40 inches and the longest length being somewhere between 80 and 85 inches.

Problem 2

Statement

Which data set is more likely to produce a histogram with a symmetric distribution? Explain your reasoning.

- Data on the number of seconds on a track of music in a pop album.
- Data on the number of seconds spent talking on the phone yesterday by everyone in the school.

Solution

Data on the number of seconds on a track of music in a pop album. Explanations vary. Sample explanation: Most pop songs are around the same amount of time, but most people in the school will not talk much on the phone while a few people will talk a lot so there will be a peak near zero and a few very long times for some people.

Problem 3

Statement

Evaluate the expression $4x^3$ for each value of x.

a. 1

b. 2

c. $\frac{1}{2}$

Solution

a. 4

b. 32

c. $\frac{1}{2}$ or equivalent

(From Unit 6, Lesson 15.)

Problem 4

Statement

Decide if each data set might produce one or more gaps when represented by a histogram. For each data set that you think might produce gaps, briefly describe or give an example of how the values in the data set might do so.

a. The ages of students in a sixth-grade class.

b. The ages of people in an elementary school.

c. The ages of people eating in a family restaurant.

d. The ages of people who watch football.

e. The ages of runners in a marathon.

Solution

a. No.

b. Yes. Sample reasoning: The data set might show a lot of observations that are between 5 and 12 years old (the students' ages) and older than 20 years (the ages of staff and teachers), but no observations between 12 and 20.

c. Yes. Sample reasoning: The data set might show a lot of observations of children, then their parents. Observations between 12 and 25 are less likely.

d. No.

e. Yes. Sample reasoning: Most of the runners might be adults between 24 and 60 years of age, but there might be a few runners who are in their late teens or older runners in their 80's.

Problem 5

Statement

Jada drank 12 ounces of water from her bottle. This is 60% of the water the bottle holds.

a. Write an equation to represent this situation. Explain the meaning of any variables you use.

b. How much water does the bottle hold?

Solution

a. Answers vary. Sample responses: $12 = \frac{60}{100}b$ or $12 = 0.6b$, where b is the number of ounces of water the bottle holds

b. $b = 20$

(From Unit 6, Lesson 7.)