Lesson 1: Distributions and Their Shapes

Student Outcomes

- Students use informal language to describe the shape, center, and variability of a distribution based on a dot plot, histogram, or box plot.
- Students recognize that a first step in interpreting data is making sense of the context.
- Students make meaningful conjectures to connect data distributions to their contexts and the questions that could be answered by studying the distributions.

Lesson Notes

Students began their study of data in Grade 6 with dot plots, histograms, and box plots. In Grade 6, they learned how to construct a graph and how to summarize a distribution by its center and variability. This lesson looks back at the graphs students previously studied as an introduction to this module. Students are also asked to interpret what a graph communicates. They are reminded that a graph is not only a representation of data but also a summary of a data story. Each of the graphs presented in the exercises will be encountered in the lessons that follow in this module. This lesson asks students to start thinking about what the data indicate, how they might have been collected, and what they tell us.

Statistics is all about data. Without data to talk about or to analyze or to question, statistics would not exist. There is a story to be uncovered behind all data—a story that has characters, plots, and problems. The questions or problems addressed by the data and their story can be disappointing, exciting, or just plain ordinary. This module is about stories that begin with data.

Classwork

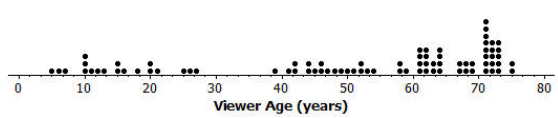
Example 1 (5 minutes): Graphs

Example 1

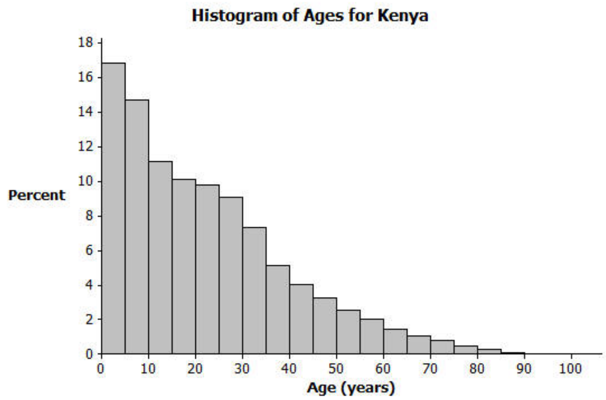
Data are often summarized by graphs; the graphs are the first indicator of variability in the data.

- **<u>Dot plots</u>: A plot of each data value on a scale or number line.**

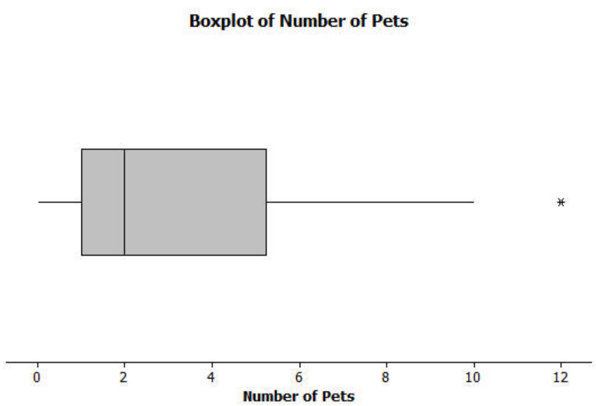
Dot Plot of Viewer Age



- **Histograms: A graph of data that groups the data based on intervals and represents the data in each interval by a bar.**



- **Box plots: A graph that provides a picture of the data ordered and divided into four intervals that each contains approximately 25% of the data.**



Review the different types of graphs that students have previously studied (dot plots, box plots, and histograms). Convey the following:

MP.2

- Think of each graph as telling a story.
- Graphs of distributions are often the starting point in understanding the variability in the data.
- The graphs in the following exercises will be analyzed in more detail in the lessons that follow.

Exercises 1–15 (25 minutes)

Spend a few minutes with students reading the opening paragraph. Discuss with them the graphs presented in the example. Ask them if they remember these graphs from their previous work with data and what they recall about these graphs. Allow time for students to read the exercises. Then, provide time for students to discuss the questions individually or in small groups for each set of graphs. Conduct a brief discussion with students after they have developed answers for the questions. The graphs and the questions are summarized in the teacher notes along with possible responses and discussion items that students might address.

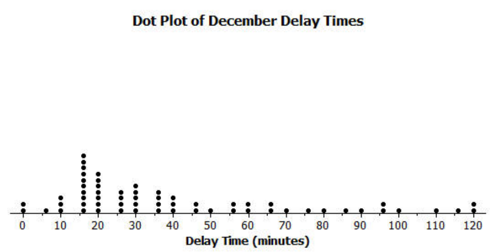
In most cases, the questions do not have exact answers. For this lesson, encourage students to make summaries based on what information the graphs convey about the data presented. More formal analysis of the data will be developed in the next set of lessons of this module.

Exercises 1–15

Answer the questions that accompany each graph to begin your understanding of the story behind the data.

Transportation officials collect data on flight delays (the number of minutes past the scheduled departure time that a flight takes off).

Consider the dot plot of the delay times for sixty BigAir flights during December 2012.



1. **What do you think this graph is telling us about the flight delays for these sixty flights?**

   Most flights are delayed for 15 minutes; some are delayed for a longer time.

2. **Can you think of a reason why the data presented by this graph provides important information? Who might be interested in this data distribution?**

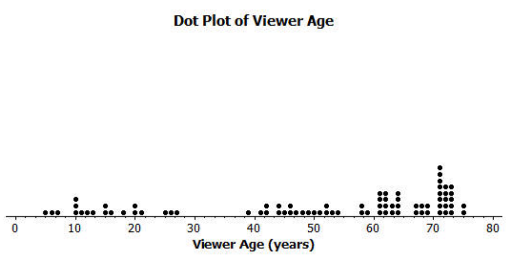
   If flights are late, travelers would not select this airline.

   BigAir and travelers using this airline would be interested in this information.

3. **Based on your previous work with dot plots, would you describe this dot plot as representing a symmetric or a skewed data distribution? (Recall that a skewed data distribution is not mound shaped.) Explain your answer.**

   Skewed; it has a tail to the right. (Students are introduced to this in Grade 6.)

A random sample of eighty viewers of a television show was selected. The dot plot below shows the distribution of the ages (in years) of these eighty viewers.



4. **What do you think this graph is telling us about the ages of the eighty viewers in this sample?**

   The typical age of viewers is between 60 and 70 years old; the show appeals to a wide range of ages.

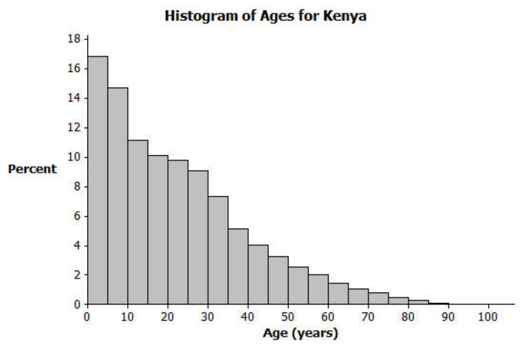
5. **Can you think of a reason why the data presented by this graph provides important information? Who might be interested in this data distribution?**

   This data is important in understanding the audience of the show. If the show is paid for by commercials, then the distribution is important for sponsors.

6. **Based on your previous work with dot plots, would you describe this dot plot as representing a symmetric or a skewed data distribution? Explain your answer.**

   Skewed; it has a tail to the left.

The following histogram represents the age distribution of the population of Kenya in 2010.



7. **What do you think this graph is telling us about the population of Kenya?**

   A large percentage of the people in Kenya are ages 10 or younger.

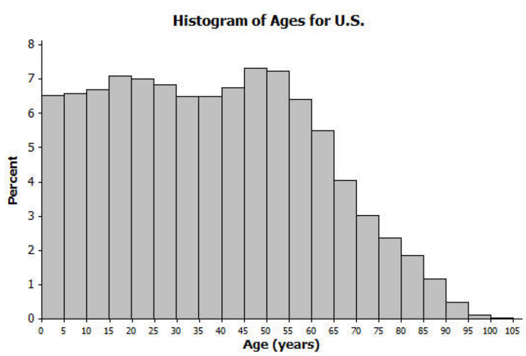
8. **Why might we want to study the data represented by this graph?**

   It tells us about Kenya and its challenges based on its population and demographics. It is important to understand the data because it may lead to finding the reason(s) as to why this data is occurring—which could lead to solutions.

9. **Based on your previous work with histograms, would you describe this histogram as representing a symmetrical or a skewed distribution? Explain your answer.**

   Skewed; it has a tail to the right.

The following histogram represents the age distribution of the population of the United States in 2010.



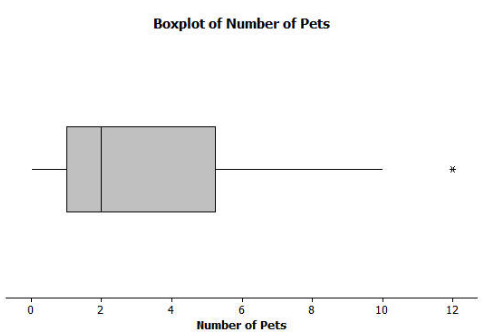
10. **What do you think this graph is telling us about the population of the United States?**

    The percentage of the population is about the same in each interval until the age range of 60 to 65 years old. Then, the percentages decline.

11. **Why might we want to study the data represented by this graph?**

    Population data are used to determine health care challenges (for 65 years and older) or education challenges (for 0 to 20 years old). Businesses (such as insurance companies) use this type of data.

Thirty students from River City High School were asked how many pets they owned. The following box plot was prepared from their answers.

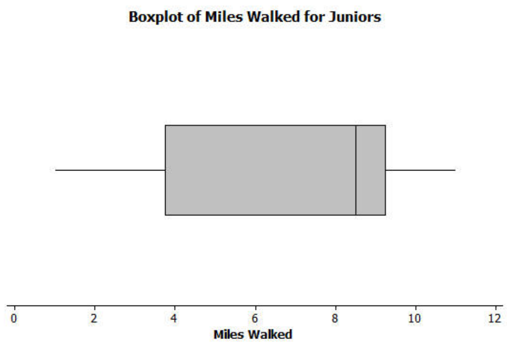


12. **What does the box plot tell us about the number of pets owned by the thirty students at River City High School?**

    50% of students own between 1 and 5 pets.

13. **Why might understanding the data behind this graph be important?**

    Understanding the data is important for planning special events involving pets and understanding interests of a group of people.

Twenty-two juniors from River City High School participated in a walkathon to raise money for the school band. The following box plot was constructed using the number of miles walked by each of the twenty-two juniors. 	14. What do you think the box plot tells us about the number of miles walked by the twenty-two juniors? *50% of the juniors walked between 4 and 9 miles; 25% walked between 1 and 4 miles; 25% walked between 9 and 11 miles.* 15. Why might understanding the data behind this graph be important? *It provides an indication of the participation of the junior class. Sponsors of the event may be interested in the data.*

Closing (5 minutes)

Pose at least two of the following questions; allow a few student responses for each.

- What are reasons that a scheduled airline flight might be delayed?
- What are some of the favorite television shows of the students in your class? List some of the most memorable commercials that are shown during those shows. In your opinion, do the commercials connect with the viewers?
- You walk into a store. You estimate that most of the customers are between fifty and sixty years old. What kind of store do you think it is?
- If you asked students in your class how many pets they owned, what do you think would be a typical value?
- You are selected to take a trip to Kenya. Do you think you will meet several people ninety or older? Why or why not?

> **Lesson Summary**
>
> **Statistics is about data. Graphs provide a representation of the data distribution and are used to understand the data and to answer questions about the distribution.**

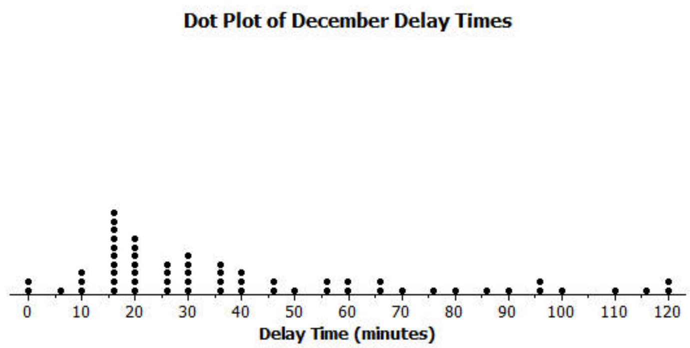
Exit Ticket (10 minutes)

Name ____ Date____________________

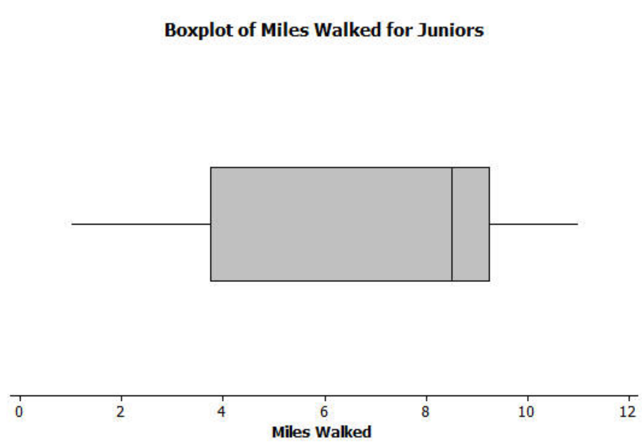
Lesson 1: Distributions and Their Shapes

Exit Ticket

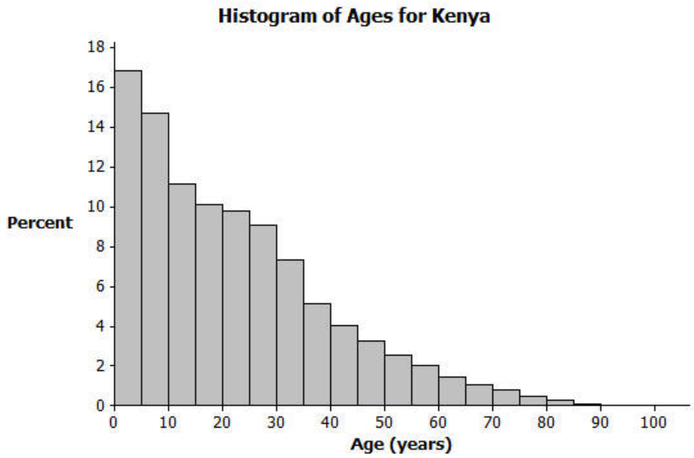
1. Sam said that a typical flight delay for the sixty BigAir flights was approximately one hour. Do you agree? Why or why not?



2. Sam said that 50% of the twenty-two juniors at River City High School who participated in the walkathon walked at least ten miles. Do you agree? Why or why not?

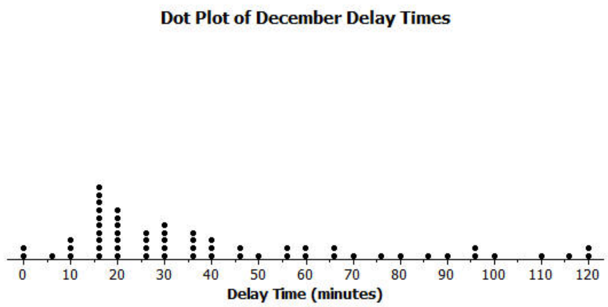


3. Sam said that young people from the ages of 0 to 10 years old make up nearly one-third of the Kenyan population. Do you agree? Why or why not?



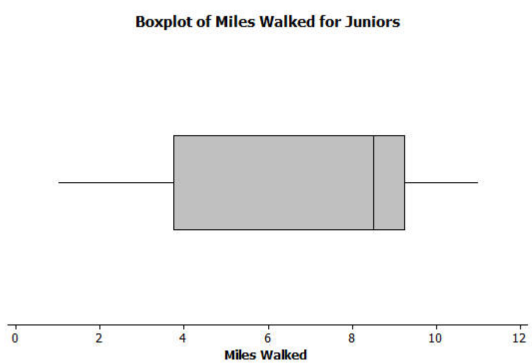
Exit Ticket Sample Solutions

1. **Sam said that a typical flight delay for the sixty BigAir flights was approximately one hour. Do you agree? Why or why not?**



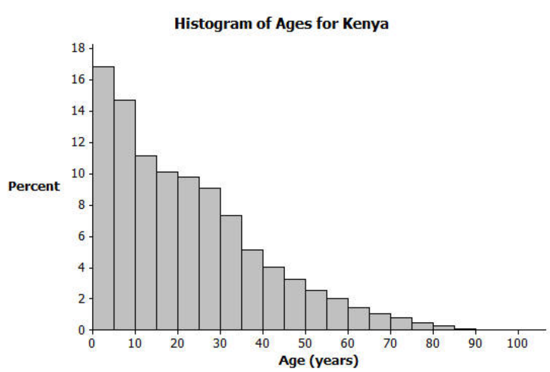
Most of the flight delays are less than 60 minutes; therefore, 60 minutes is not a typical description of how many minutes a flight is delayed.

2. **Sam said that 50% of the twenty-two juniors at River City High School who participated in the walkathon walked at least ten miles. Do you agree? Why or why not?**



It would not be accurate to indicate that 50% walked 10 or more miles. The upper quartile indicates that 25% of the 22 students walked 9 or more miles.

3. **Sam said that young people from the ages of 0 to 10 years old make up nearly one-third of the Kenyan population. Do you agree? Why or why not?**



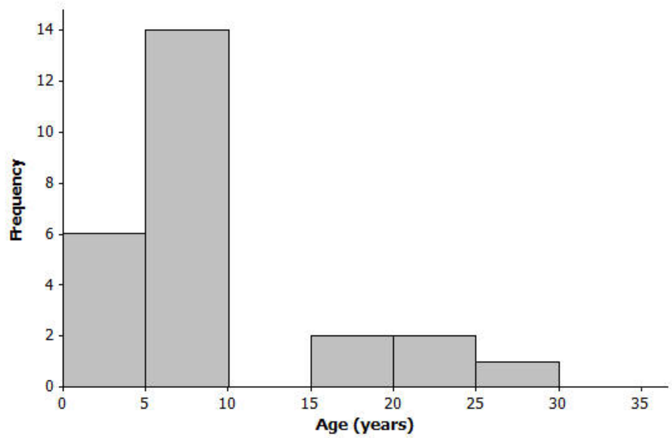
I do agree with Sam. The first two bars of the Kenya graph represent people between 0 and 10 years old. The first bar represents approximately 17% of the population (0–5 year olds), and the second bar represents approximately 15% of the population (5–10 year olds). Therefore, approximately 32%, or nearly one-third of the Kenyan population, is between 0 and 10 years old.

Problem Set Sample Solutions

1. **Twenty-five people were attending an event. The ages of the people are as follows:**

$$3, 3, 4, 4, 4, 4, 5, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 16, 17, 22, 22, 25.$$

a. **Create a histogram of the ages using the provided axes.**



b. **Would you describe your graph as symmetrical or skewed? Explain your choice.**

This graph is skewed with a tail to the right. Most of the ages are in the younger intervals.

c. **Identify a typical age of the twenty-five people.**

A typical age could be any age in the interval of 5 to 10, such as 6 or 7 years old.

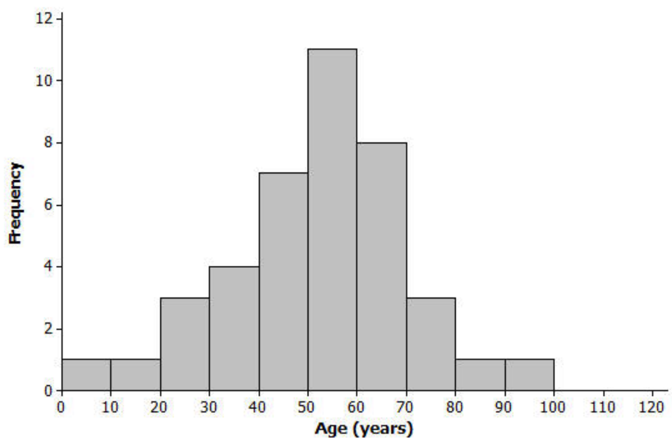
d. **What event do you think the twenty-five people were attending? Use your histogram to justify your conjecture.**

The 22 ages were obtained from a story time hour at a library. Most of the ages were the children attending the event. The older ages represent some of the caretakers and the storytellers. Discuss any conjectures in which the younger age intervals would likely represent most of the people attending the event, with some older people to help out with the event.

2. A different forty people were also attending an event. The ages of the people are as follows:

$$6, 13, 24, 27, 28, 32, 32, 34, 38, 42, 42, 43, 48, 49, 49, 49, 51, 52, 52, 53,$$
$$53, 53, 54, 55, 56, 57, 57, 60, 61, 61, 62, 66, 66, 66, 68, 70, 72, 78, 83, 97.$$

a. Create a histogram of the ages using the provided axes.



b. Would you describe your graph of ages as symmetrical or skewed? Explain your choice.

This histogram is nearly symmetrical.

c. Identify a typical age of the forty people.

A typical age is approximately 55 *years old.*

d. What event do you think the forty people were attending? Use your histogram to justify your conjecture.

The ages were obtained from people attending a family reunion. This is obviously not necessarily what you would expect from all family reunions. Discuss any conjectures in which a nearly symmetrical data distribution of ages could be a possibility.

e. How would you describe the differences in the two histograms?

The two age distributions differ primarily in shape (skewed and symmetrical) and in center (typical age).