# Lesson 1: Distributions and Their Shapes

Classwork

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.
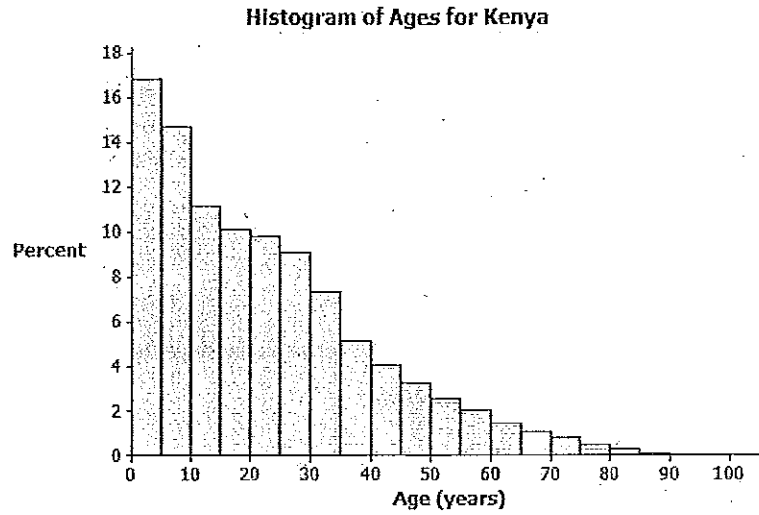
Example: Graphs

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

- DOT PLOTS: A plot of each data value on a scale or number line.

Dot Plot of Viewer Age
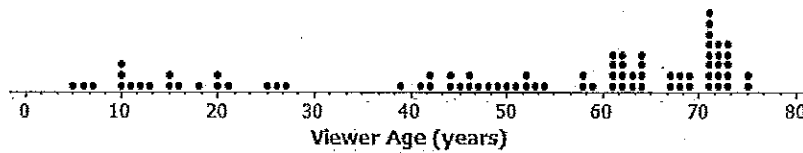
- **HISTOGRAMS:** A graph of data that groups the data based on intervals and represents the data in each interval by a bar.

Histogram of Ages for Kenya

- **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.

Boxplot of Number of Pets
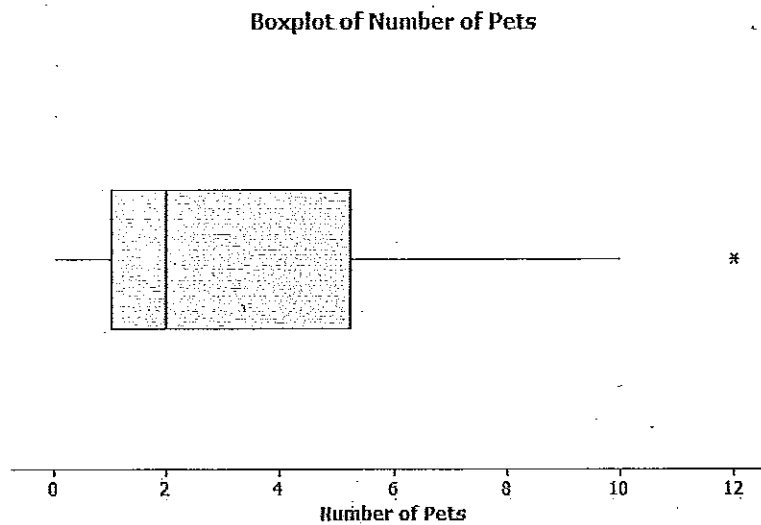
Exercises _There's Always More than 1 good answer_

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.

Dot Plot of December Delay Times

Delay Time (minutes)
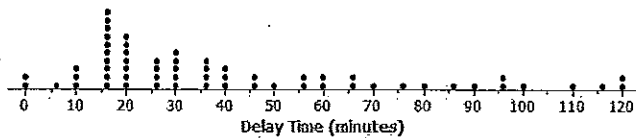
1. What do you think this graph is telling us about the flight delays for these sixty flights?

   Most delays are less than 40 minutes

2. Can you think of a reason why the data presented by this graph provide important information? Who might be interested in this data distribution?
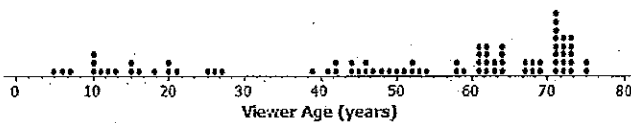
   People who want to fly with BigAir want to know what delays to expect.

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. Most data is on the left with a long "tail" to the right

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.

Dot Plot of Viewer Age

Viewer Age (years)

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

*Most viewers are older, but there are some young viewers too.*

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

Advertisers might want to know who watches a program so they can target ads.

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. Most data is on the right & there's a gap in the middle.

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

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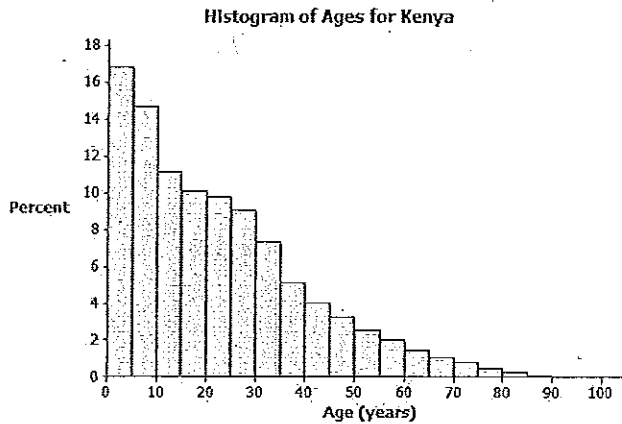
7. What do you think this graph is telling us about the population of Kenya?

Most Kenyans are young.

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

Look at disease trends or net worth compared to age.

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. Most data is on the left.

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

Histogram of Ages for U.S.
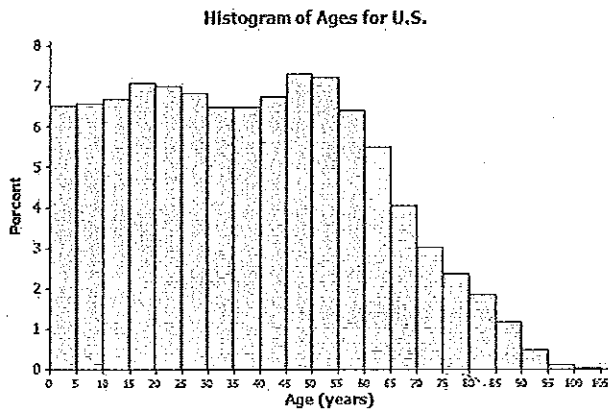
10. What do you think this graph is telling us about the population of the United States?

People live longer in the US than Kenya.

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

Demographic trends.

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

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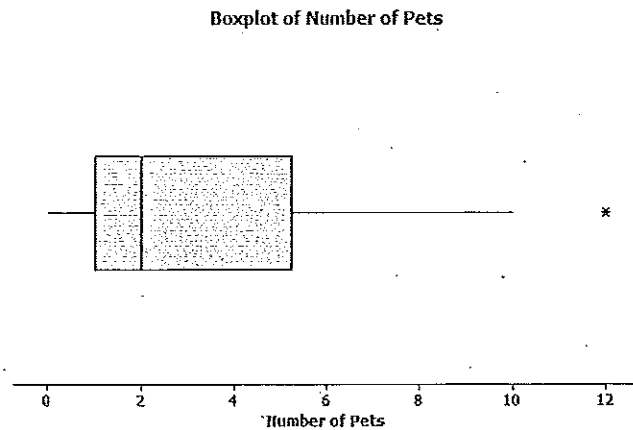
12. What does the box plot tell us about the number of pets owned by the thirty students at River City High School?

Most students have a few pets. One has 12.

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

To know what students are interested in.

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.

Boxplot of Miles Walked for Juniors
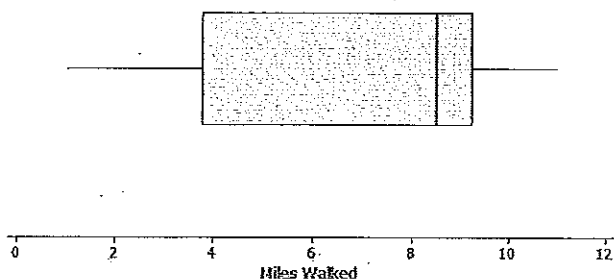
Miles Walked

14. What do you think the box plot tells us about the number of miles walked by the twenty-two juniors?

Half of the juniors walked between 4 & 9 miles.

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

To know the fitness/motivation of students and raise more $ next time.


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.

Problem Set

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

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

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

   Histogram for Event 1
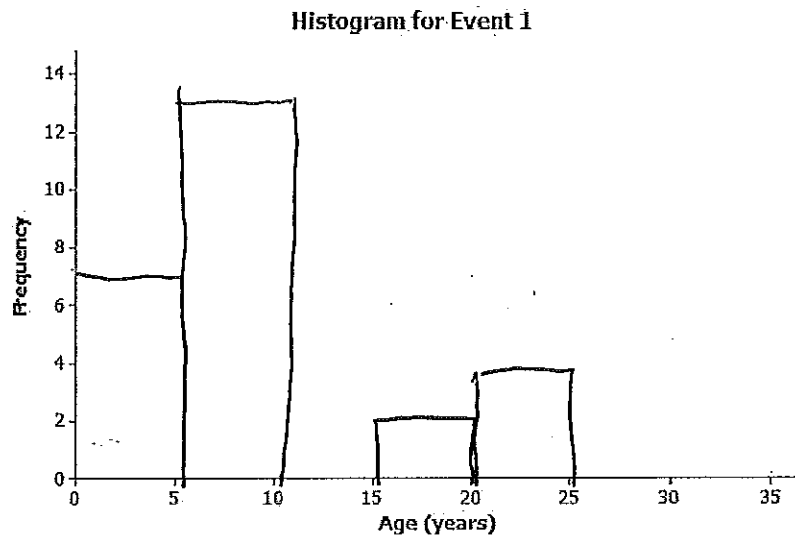
   b. Would you describe your graph as symmetrical or skewed? Explain your choice. *Skewed*

   c. Identify a typical age of the twenty-five people. *6 or 7*

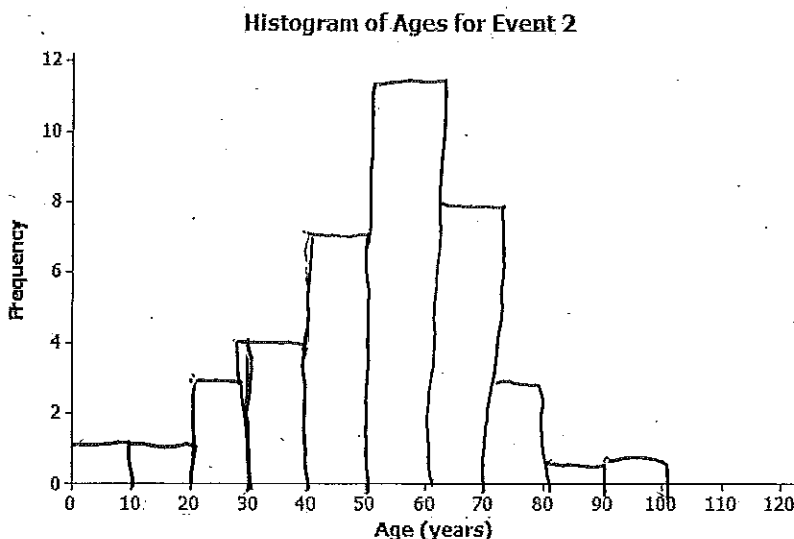   d. What event do you think the twenty-five people were attending? Use your histogram to justify your conjecture. *Soccer camp? School carnival? Lots of little kids and a few young adults/teens.*

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.

Histogram of Ages for Event 2

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

c. Identify a typical age of the forty people. *55*

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

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

d) Going to the theatre. Mostly middle aged people with some kids & old people.

e) Skewed vs Symmetric. Data split in 2 pieces versus concentrated in the center.

Lesson 2: Describing the Center of a Distribution

Classwork

In previous work with data distributions, you learned how to derive the mean and the median of a data distribution. This lesson builds on your previous work with a center.

Exploratory Challenge/Exercises 1–9

Consider the following three sets of data.

Data Set 1: Pet owners

Students from River City High School were randomly selected and asked, "How many pets do you currently own?" The results are recorded below.

0	0	0	0	1	1	1	1	1	1	1	1	1	1	2
2	2	2	3	3	4	5	5	6	6	7	8	9	10	12

Data Set 2: Length of the east hallway at River City High School

Twenty students were selected to measure the length of the east hallway. Two marks were made on the hallway's floor: one at the front of the hallway and one at the end of the hallway. Each student was given a meter stick and asked to use the meter stick to determine the length between the marks to the nearest tenth of a meter. The results are recorded below.

8.2	8.3	8.3	8.4	8.4	8.5	8.5	8.5	8.5	8.5
8.6	8.6	8.6	8.6	8.7	8.7	8.8	8.8	8.9	8.9

Data Set 3: Age of cars

Twenty-five car owners were asked the age of their cars in years. The results are recorded below.

0	1	2	2	3	4	5	5	6	6	6	7	7
7	7	7	7	8	8	8	8	8	8	8	8	