

Module 2: Lesson 1

# Distributions and Their Shapes

## 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.

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.

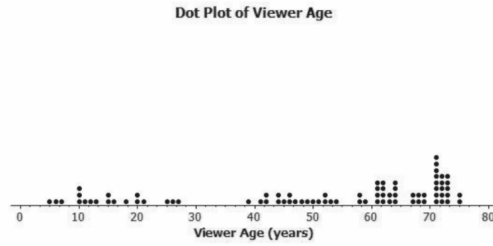
In this section of Module 2 we begin by focusing on analyzing graphs. Things to look out for may include:

1. Clusters of data
2. Large or small values of data
3. Shapes the data makes

## Types of Graphs

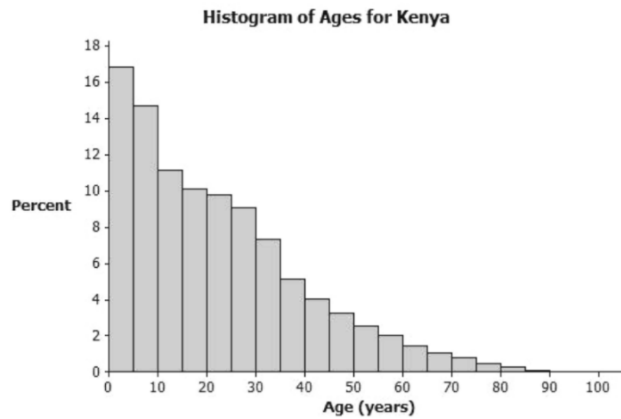
Data are often summarized by graphs; the graphs are the first indicator of variability in the data.

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

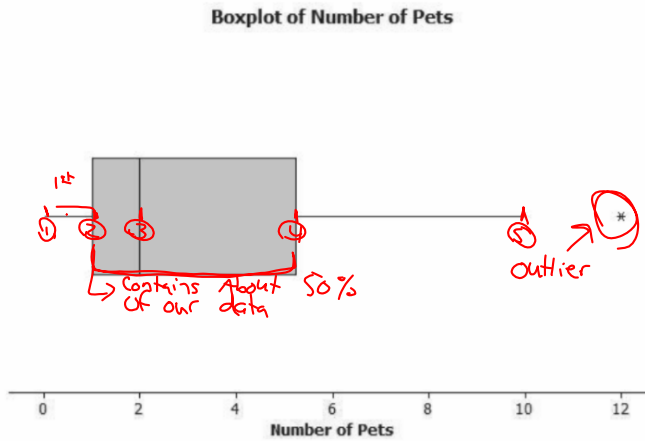


**HISTOGRAMS:** A graph of data that groups the data based on intervals and represents the data in each interval by a bar.

*Bar Graph*



**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.



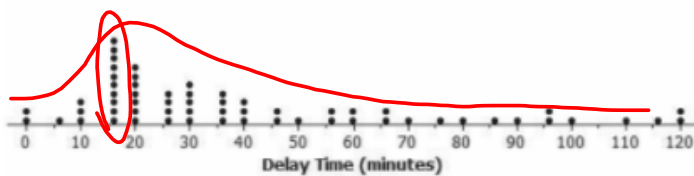
5-Number Summary

- ① Minimum - Smallest value
- ② Q2 - Middle value of our minimum + median
- ③ Median: Middle value
- ④ Q4 - Middle value of our median + maximum
- ⑤ Maximum - Largest value

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



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

- Most flight delays last 16 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?

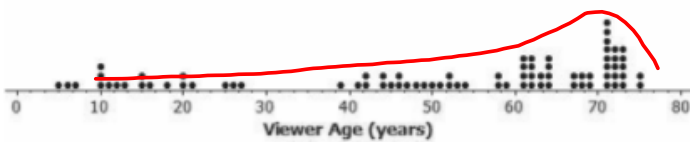
- Passengers
- Airport

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.

- Right - Skewed

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



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

◦ Most viewers are older  
(Around 60-70 years old)

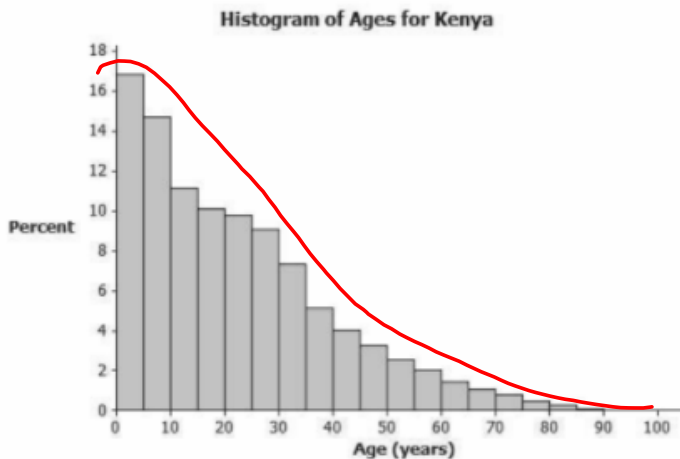
5. Can you think of a reason why the data presented by this graph provide important information? Who might be interested in this data distribution?

◦ Advertisers  
◦ TV Producers

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 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?

- Most of the population is fairly young (Deaths at a younger age)

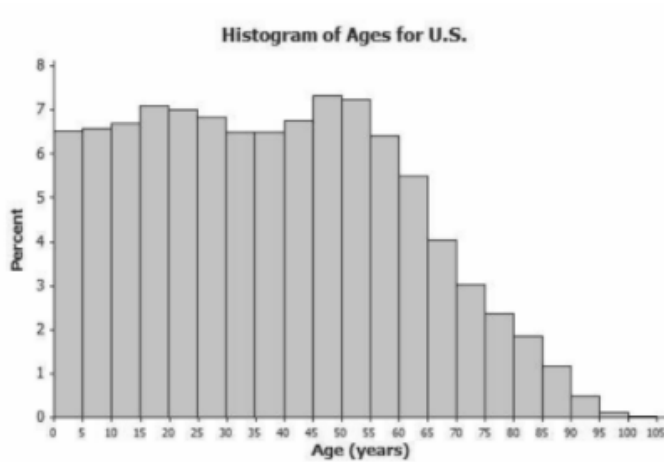
8. Why might we want to study the data represented by this graph?

- Medical researchers
- Geographers

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 to the right.

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



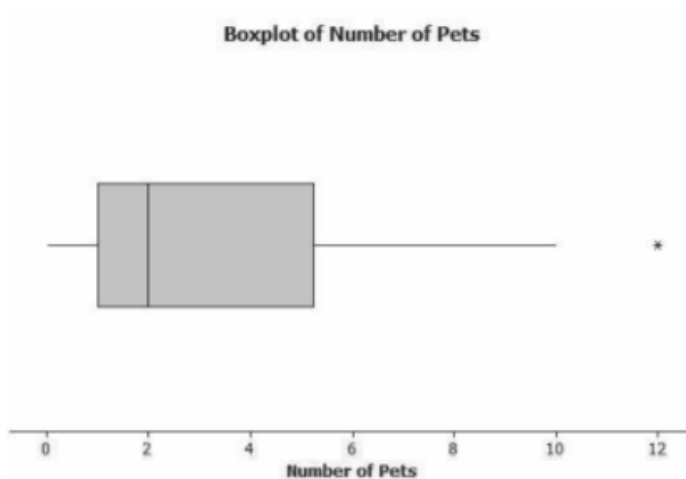
10. What do you think this graph is telling us about the population of the United States?

The percentage of population remain steady for the majority of the time, then it declines after about 60 years old.

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

Business companies, such as insurance companies, might use this data to determine insurance rates.

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?

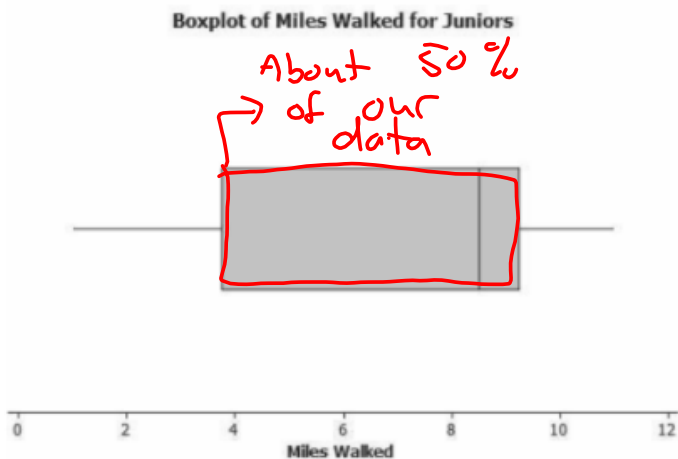
This data tells us that about 1/2 of the students that were interviewed own between 1 and 5 pets.

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

This might be important if the school is hosting an after-school event that revolves around pets. This information might give them some insight on how to prepare for such an event.



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?

about  $\frac{1}{2}$  walked between 3-9 miles.

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

The school band will be interested in this data because they will want to know how much money they raised for themselves and how much more they may need to reach a goal (if they had a goal in the first place).