

Chapter 4 – Summarizing Data Numerically
Section 3 – Boxplots

Objectives

1. Find quartiles of some data.
2. Find the interquartile range of some data.
3. Construct a boxplot to describe a distribution without outliers.
4. Identify outliers.
5. Construct a boxplot to describe a distribution with one or more outliers.
6. Compare the boxplots of two groups of data.

Vocabulary

1. quartile
2. interquartile range
3. boxplot
4. whisker
5. left/right fence
6. outlier

Lesson/Activity

OBJECTIVE 1 – Find quartiles of some data.

Definition: First quartile, second quartile, and third quartile

The first quartile, second quartile, and third quartile are the 25th, 50th, and 75th percentiles, respectively.

We use the symbols Q_1 , Q_2 , and Q_3 to stand for the first, second, and third quartiles, respectively.

1. Students in one of the author's statistics classes were surveyed about the number of hours they spending watching videos, movies, and television shows on a weekday. Here are the anonymous responses (in hours) of ten of the students: 4, 10, 3, 2, 0, 2, 1, 2, 4, 7 (Source: J. Lehmann).
 - a. Construct a histogram to determine whether the mean or the median is the better measure of the center. Explain.
 - b. Find Q_2 (the median).
 - c. Find Q_1 .
 - d. Find Q_3 .
 - e. Use technology to verify the results you found in Parts (b), (c), and (d).
 - f. Describe the meaning of the results you found in Parts (b), (c), and (d).

If some data have an odd number of data values, do not include the median in the lower half of the data when finding Q_1 . Likewise, do not include the median in the upper half of the data when finding Q_3 .

OBJECTIVE 2 – Find the interquartile range of some data.

Definition: Interquartile Range

The **interquartile range (IQR)** is given by $IQR = Q_3 - Q_1$.

The IQR measures the spread of the middle 50% of the observations (from Q_1 to Q_3).

2. In Problem 1, we found $Q_1 = 2$, $Q_2 = 2.5$, and $Q_3 = 4$ for the data about the number of hours students spend watching videos, movies, and television shows per weekday: 4, 10, 3, 2, 0, 2, 1, 2, 4, 7.
- Calculate the IQR.
 - Calculate the range.
 - What do the IQR and the range mean in this situation?
 - Discuss whether the IQR and the range are sensitive to the outlier 10 hours.

The IQR is resistant to outliers and the range is sensitive to outliers.

OBJECTIVE 3 – Construct a boxplot to describe a distribution without outliers.

3. The number of annual shark attacks in Hawaii are shown in the following table for the years 2001–2014.

3	6	5	3	4	3	7
<u>1</u>	<u>3</u>	<u>4</u>	<u>3</u>	<u>10</u>	<u>13</u>	<u>7</u>

Source: Florida Museum of Natural History

- Construct a boxplot.
- Which is longer, the left part of the box or the right part? Which is longer, the left whisker or the right whisker? Assuming the distribution is unimodal, is the distribution skewed left, skewed right, or symmetric?
- Refer to the boxplot to estimate the range. What does it mean in this situation?
- Refer to the boxplot to estimate the IQR. What does it mean in this situation?

If one whisker is longer than the other, that does *not* mean more observations are represented by the longer whisker. It means that the observations represented by the longer whisker are more spread out than the approximately equal number of observations represented by the shorter whisker.

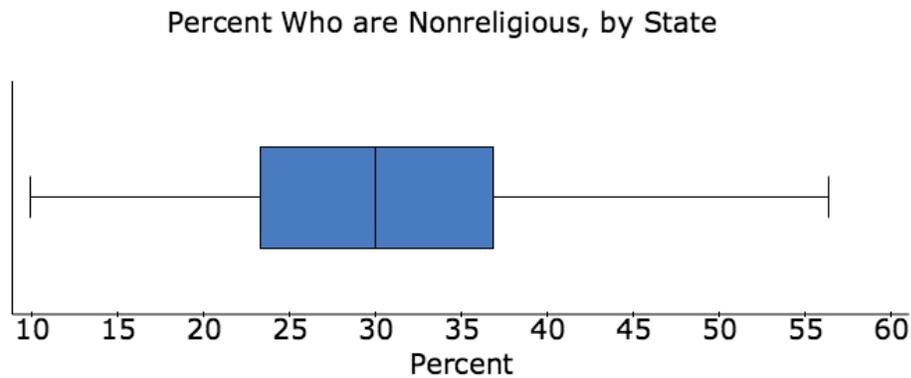
Meaning of Boxplots That Describe Distributions without Outliers

- A boxplot that describes a distribution without outliers consists of four parts: the left whisker, the left part of the box, the right part of the box, and the right whisker. Each part represents 25% of the observations.
- The full length of the box is equal to the IQR.

The following two statements apply to unimodal distributions:

- If the left part of the box is at least as long as the right part and the left whisker is quite a bit longer than the right whisker, then the distribution is skewed left.
- If the right part of the box is at least as long as the left part, and the right whisker is quite a bit longer than the left whisker, then the distribution is skewed right.

4. The percentages of residents who are nonreligious are described by the boxplot for the fifty states.



Source: The Gallup Organization

- It turns out the distribution is unimodal. On the basis of the boxplot, is the distribution symmetric, skewed left, or skewed right? Explain.
- Estimate the range. What does it mean in this situation?
- Estimate the IQR. What does it mean in this situation?

OBJECTIVE 4 – Identify outliers.

Definition: Fences

The **left fence** and the **right fence** of some data values are given by

$$\text{left fence} = Q_1 - 1.5 \cdot \text{IQR}$$

$$\text{right fence} = Q_3 + 1.5 \cdot \text{IQR}$$

Definition: Outlier

An **outlier** is an observation that is less than the left fence or greater than the right fence.

We say that the outlier lies outside the fences.

5. In Problems 1 and 2, we found $Q_1 = 2$, $Q_3 = 4$, and $\text{IQR} = 2$ for the data about the number of hours students spend watching videos, movies, and television shows per weekday:
4, 10, 3, 2, 0, 2, 1, 2, 4, 7.

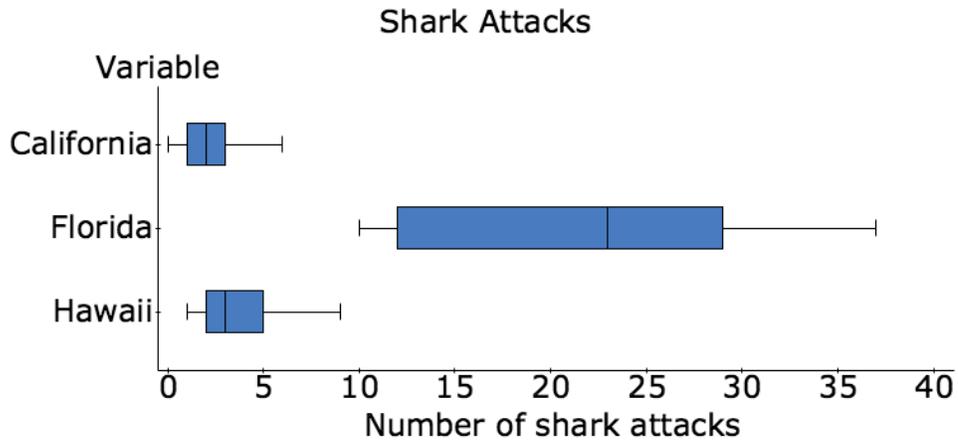
Identify and interpret any outliers.

OBJECTIVE 5 – Construct a boxplot to describe a distribution with one or more outliers.

- Construct a boxplot that describes the data about the number of hours students spend watching videos, movies, and television shows per weekday: 4, 10, 3, 2, 0, 2, 1, 2, 4, 7.
- Which is longer, the left part of the box or the right part? Which is longer the left whisker or the right whisker? What do your responses tell you about the shape of the distribution?

OBJECTIVE 6 – Compare the boxplots of two groups of data.

8. The number of annual shark attacks in California, Florida, and Hawaii are described by the following boxplots for the years 2001–2014.



Source: Florida Museum of Natural History

- Compare the centers of the three distributions. What does this mean in this situation?
- Compare the spreads of the three distributions. What does this mean in this situation?
- Compare the IQR of Florida's distribution to the range of California's distribution. What does this mean in this situation?

Homework/Assessment

1, 3, 5, 11, 13, 15, 17, 21, 23, 25, 29, 33, 41, 45, 47