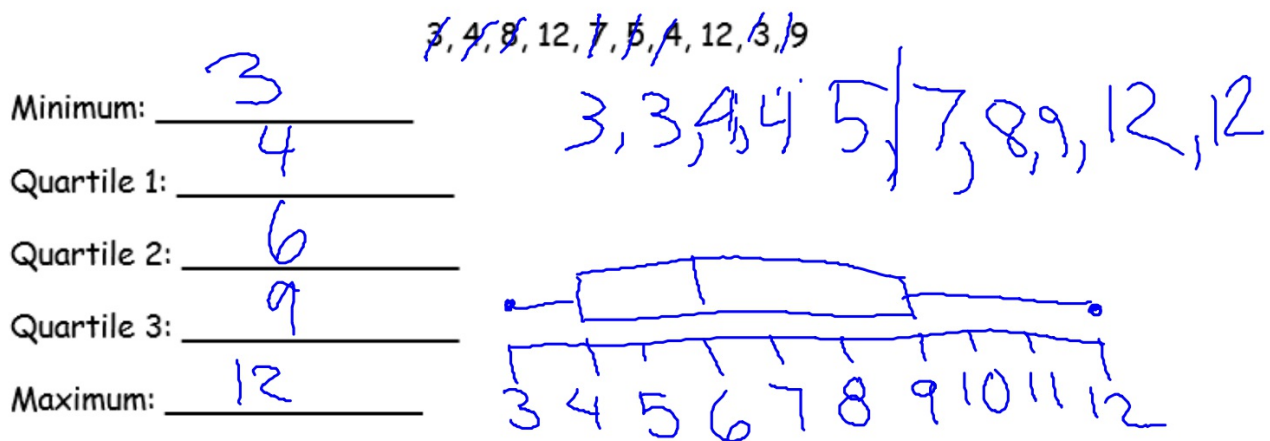


Content and Language Objectives:

Students will calculate quartiles, Interquartile ranges (IQR's) and five number summaries of data sets and draw conclusions and justify in writing those conclusions, after:

- a.) creating and interpreting box plots of data sets
- b.) defining outliers
- c.) making a chart with a partner about the characteristics of box plots, interquartile ranges, and outliers

The number of runs scored by a softball game in 10 games is given. Use the data to make a box-and-whisker plot.



You work for a company and you want a raise in your pay. You find out the following information:

Annual Income	Number of Employees
\$105,000	1
60,000	3
30,000	1
28,000	5
21,000	10

$$\begin{array}{r}
 105,000 \\
 3 \cdot 60,000 = 180,000 \\
 30,000 \\
 140,000 \\
 210,000 \\
 \hline
 665,000 \\
 20
 \end{array}$$

Find the information below:

Mean 33,250 Mode 21,000

In negotiations:

The owner of the company will probably use the mean to describe the company. Why?

It is a larger value than the mode, and it doesn't represent the incomes fairly

The union leader (person negotiating for the workers) will probably use the mode to describe the company. Why?

50% of the employees earn \$21,000

A statistician would probably use the mode to describe the company. Why? Best representation of data.

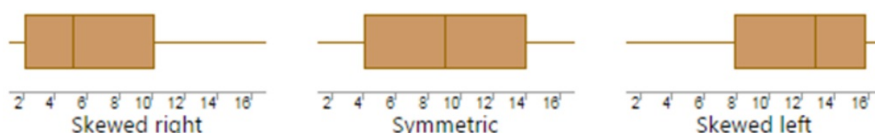
How to Interpret a Boxplot

Here is how to read a boxplot. The median is indicated by the vertical line that runs down the center of the box. In the boxplot above, the median is about 400.

Additionally, boxplots display two common measures of the variability or spread in a data set.

- **Range.** If you are interested in the spread of *all* the data, it is represented on a boxplot by the horizontal distance between the smallest value and the largest value, including any outliers. In the boxplot above, data values range from about -700 (the smallest outlier) to 1700 (the largest outlier), so the range is 2400. If you ignore outliers, the range is illustrated by the distance between the opposite ends of the whiskers - about 1000 in the boxplot above.
- **Interquartile range (IQR).** The middle half of a data set falls within the interquartile range. In a boxplot, the interquartile range is represented by the width of the box (Q3 minus Q1). In the chart above, the interquartile range is equal to 600 minus 300 or about 300.

And finally, boxplots often provide information about the shape of a data set. The examples below show some common patterns.



Each of the above boxplots illustrates a different **skewness** pattern. If most of the observations are concentrated on the low end of the scale, the distribution is skewed right; and vice versa. If a distribution is symmetric, the observations will be evenly split at the median, as shown above in the middle figure.

Problem 1

Consider the boxplot below.



Which of the following statements are true?

- ☒ I. The distribution is skewed right.
- ☐ II. The interquartile range is about 8.
- ☒ III. The median is about 10.

(A) I only

(B) II only

(C) III only

(D) I and III

(E) II and III

In Mr. Aye's class, the students' grades for the last test were: 89, 76, 90, 87, 89, 65, 77, 95, 89, 76, 50, 79, 82, and 88.

While going over the test with the students, Mr. Aye discovered he made a mistake—everyone actually had a question correct that he had marked wrong, so everyone's score should be four points higher. Calculate the original measures of center and spread and the new measures of center and spread. Fill in the table with the information.

Statistic	Original Data	Data + 4
Mean		
Median		
Mode		
Range		
Lower Quartile <i>Q₁</i>		
Upper Quartile <i>Q₃</i>		

Draw the box-and-whisker plots side-by-side for both sets of data.



We can also use the IQR to determine whether a number is an outlier of a data set:

A Test for Outliers:

A data point is considered to be an outlier if it lies more than 1.5 interquartile ranges below Q_1 (i.e., the number is less than $Q_1 - 1.5(IQR)$) or 1.5 interquartile ranges above Q_3 (i.e., the number is greater than $Q_3 + 1.5(IQR)$).

The following data are the numbers of home runs Barry Bonds hit in his first 16 seasons, sorted:

16 19 24 24 | 25 33 33 34 | 34 37 37 40 | 42 46 49 73

- a. Create a five-number summary of this data.

min - 16
 Q_1 - 24.5
med - 34
 Q_3 - 41
max - 73

$$IQR = Q_3 - Q_1 = 41 - 24.5 = 16.5$$

$$1.5(16.5) = 24.75$$

$$1.5(IQR)$$

- b. We suspect that Bonds' 73-home-run season is an outlier. Is it?

$$Q_3 + 1.5(IQR)$$

$$41 + 24.75 = 65.75$$

73 is greater than 65.75 so
it's an outlier.

- c. For good measure, is the 16-home-run season an outlier? Give specific calculations.

$$Q_1 - 1.5(IQR)$$

$$24.5 - 24.75 = -0.25$$

16 is not an outlier

Make a box-and-whisker plot

Statistic	Original Data	Data + 4
Mean	80.9	84.9
Median	84.5	88.5
Mode	89	93
Range	45	45
Lower Quartile	76	80
Upper Quartile	89	93

Draw the box-and-whisker plots side-by-side for both sets of data.

