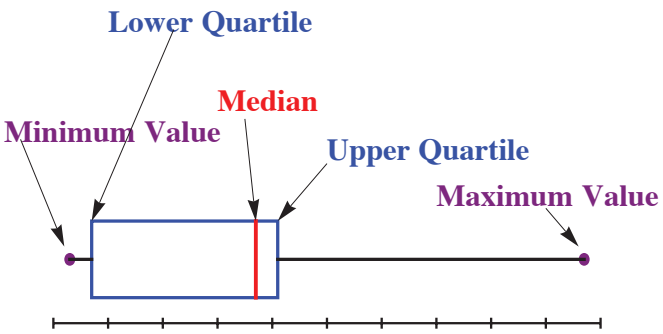


Box-and-Whisker Plots

ABOUT THIS LESSON

In this lesson, students create and interpret box-and-whisker plots (boxplots) in context of various situations. They determine the five-number summary, and use the interquartile range, IQR, to determine whether data points are outliers. Modified boxplots, boxplots that represent outliers with points beyond the endpoints of the whisker, are used in the activity since this is the preferred method in statistics. Students construct, interpret, and compare the distributions of parallel boxplots.

OBJECTIVES

Students will

- determine the five-number summary for data sets.
- construct standard and modified boxplots.
- read and interpret the distribution of data represented in boxplots.
- compare data represented in parallel boxplots.

LEVEL

Grade 6 or Grade 7 in a unit on graphical displays

MODULE/CONNECTION TO AP*

Graphical Displays and Distributions

**Advanced Placement and AP are registered trademarks of the College Entrance Examination Board. The College Board was not involved in the production of this product.*

MODALITY

NMSI emphasizes using multiple representations to connect various approaches to a situation in order to increase student understanding. The lesson provides multiple strategies and models for using those representations indicated by the darkened points of the star to introduce, explore, and reinforce mathematical concepts and to enhance conceptual understanding.

P – Physical
V – Verbal
A – Analytical
N – Numerical
G – Graphical

TEACHER PAGES

COMMON CORE STATE STANDARDS FOR MATHEMATICAL CONTENT

This lesson addresses the following Common Core State Standards for Mathematical Content. The lesson requires that students recall and apply each of these standards rather than providing the initial introduction to the specific skill.

Targeted Standards

6.SP.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
See questions 1a, 2b, 3a

6.SP.5c: Summarize numerical data sets in relation to their context, such as by
(c) giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data was gathered.
See questions 1a, 1c-d, 1f-h, 2a-b, 3a, 3c-e

7.SP.4: Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*
See questions 2c-e, 4a-c

COMMON CORE STATE STANDARDS FOR MATHEMATICAL PRACTICE

These standards describe a variety of instructional practices based on processes and proficiencies that are critical for mathematics instruction. NMSI incorporates these important processes and proficiencies to help students develop knowledge and understanding and to assist them in making important connections across grade levels. This lesson allows teachers to address the following Common Core State Standards for Mathematical Practice.

MP.1: Make sense of problems and persevere in solving them.
Students make a human boxplot to model the effect of an extreme value on a statistical measure.

MP.2: Reason abstractly and quantitatively.
In question 3, students recognize the effect of an extreme value on the position of the mean in a data set without actually calculating the mean, and then they numerically compute the largest salary that would not be an outlier.

MP.3: Construct viable arguments and critique the reasoning of others.
In question 2e, students must use characteristics of the parallel boxplots to justify that minivans tend to have more cargo volume than SUVs.

FOUNDATIONAL SKILLS

The following skills lay the foundation for concepts included in this lesson:

- Calculate measures of central tendency
- Scale and use a number line to plot data

ASSESSMENTS

The following formative assessment is embedded in this lesson:

- Students engage in independent practice.

The following additional assessments are located on our website:

- Graphical Displays and Distributions – 6th Grade Free Response Questions
- Graphical Displays and Distributions – 6th Grade Multiple Choice Questions
- Graphical Displays and Distributions – 7th Grade Free Response Questions
- Graphical Displays and Distributions – 7th Grade Multiple Choice Questions

MATERIALS AND RESOURCES

- Student Activity pages
- Graphing calculators (optional)

TEACHING SUGGESTIONS

Before beginning the lesson, tell students that the purpose of a boxplot is to determine the shape of the data and easily compare and contrast two or more data sets. A technique for actively engaging students is to construct a human boxplot. Ask for 9 volunteers to arrange themselves at the front of the classroom by some physical characteristic (hair length or height are possibilities). Ask the class to identify which student is the median (5th of the 9 ordered students). Have that student turn 90°(to eliminate him/herself from the other data points), and have the student hold a sign for the median. Now there are 4 students on either side of the median. Ask the class to identify Q_1 (between the 2nd and 3rd students) and Q_3 (between the 7th and 8th students) and ask for two more volunteers to hold the signs for Q_1 and Q_3 in the space between them (indicating that Q_1 and Q_3 are not actual data points). The first and last students in the arrangements can hold signs for the minimum and maximum. In this way, students review the vocabulary and process of calculating a five-number summary while they can "see" the construction of a standard boxplot. String or yarn can be used to connect the minimum and maximum values to the "box" so that students can visualize the whiskers.

In question 1a, the boxplot is provided. Although this type of graphical display is very different from a standard coordinate graph, a consistent scale is required. Like a scatterplot or a histogram, a scale is always required on the axis below the graph. Have the students label the five-number summary on the graph before they determine the scale of the horizontal axis.

Outliers refer to any number which is at the extremes of data sets. Part (g) in question 1 asks students to determine if the minimum and maximum cargo volumes are outliers. For a boxplot, an outlier is any data point(s) that is located more than 1.5 times the length of the IQR $(Q_3 - Q_1)$ from either Q_1 or Q_3. Students use the formula, Q_1–1.5(IQR) and Q_3+1.5(IQR) to determine outliers. Knowledge of the existence of outliers often alerts statisticians to unusual features in the data and reminds them to examine the data more carefully.

In question 2a, students compute the five-number summary for 45 data points. Using a graphing calculator to determine these values will save time and allow students to use technology for creating the graphical display. To determine these values, enter the data in L1 using STAT, Edit. Calculate the five-number summary by pressing STAT, CALC, 1-Var Stats, ENTER. Scroll down the screen until the values are displayed. If they do not have access to a graphing calculator, provide these values for the students so they can create the boxplot. (Minimum = 16; Q_1 = 34.75; median = 40; Q_3 = 47; Maximum = 84).

Since students are asked to construct modified boxplots in question 2b, discuss how a modified boxplot differs from a standard boxplot. The boxplot in question 1a is a standard boxplot. In a standard boxplot the whiskers of the graph are drawn to the minimum and maximum data values, and the graph does not display outliers. A modified boxplot displays any outliers that are present in the data. Outliers are represented with points beyond the whiskers and the endpoints of the whiskers represent the largest and smallest values in the data that are <u>not</u> outliers. Both plots display the five-number summary (minimum value, lower quartile (Q_1), median, upper quartile (Q_3), and maximum value). A graphing calculator can be set to display either a standard boxplot or a modified boxplot. On the statplot menu of the TI-84, the first boxplot option is a modified boxplot and the second boxplot option is a standard boxplot.

Comparing two graphical displays is an important concept in statistics. The students create the cargo volume for the SUVs and minivans. The purpose of these questions is to help students understand that these graphical displays can be used to make

decisions. Question 2e provides an opportunity to discuss this connection. In question 4, students are given parallel boxplots and asked more direct questions about the two graphs. When comparing graphs, students must use comparative language, such as "greater than, less than, or equal," and refer to both of the situations.

Determining the scale for the boxplot in question 3 may present a challenge for students. Using technology allows them to explore various scales for the data. If graphing calculators are not used, encourage students to share their graphs and how they determined the scale. As an extension to this question, ask five students to create a physical boxplot of the data in the hall. The physical representation of the outlier is very revealing.

You may wish to support this activity with TI-Nspire™ technology. See *Finding the Five Number Summary of a Data Set* and *Investigating Data Using Box Plots* in the NMSI TI-Nspire Skill Builders.

Suggested modifications for additional scaffolding include the following:

1g Provide the method to calculate IQR and outliers, $\text{IQR} = Q_3 - Q_1$; outliers: $Q_1 - 1.5(\text{IQR}); Q_3 + 1.5(\text{IQR})$.

2a Provide step-by-step instructions for calculating the five-number summary using a graphing calculator. If a graphing calculator is not available, provide the student with the five-number summary.

3a Modify the question by providing the boxplot and having the student calculate the five number summary and label the data on a provided boxplot.

4b Modify the question to ask which of the following statements is true: “The median iron composition is lower than the median magnesium composition." or "The highest iron composition and magnesium composition are about equal.”

NMSI CONTENT PROGRESSION CHART

In the spirit of NMSI's goal to connect mathematics across grade levels, a Content Progression Chart for each module demonstrates how specific skills build and develop from sixth grade through pre-calculus in an accelerated program that enables students to take college-level courses in high school, using a faster pace to compress content. In this sequence, Grades 6, 7, 8, and Algebra 1 are compacted into three courses. Grade 6 includes all of the Grade 6 content and some of the content from Grade 7, Grade 7 contains the remainder of the Grade 7 content and some of the content from Grade 8, and Algebra 1 includes the remainder of the content from Grade 8 and all of the Algebra 1 content.

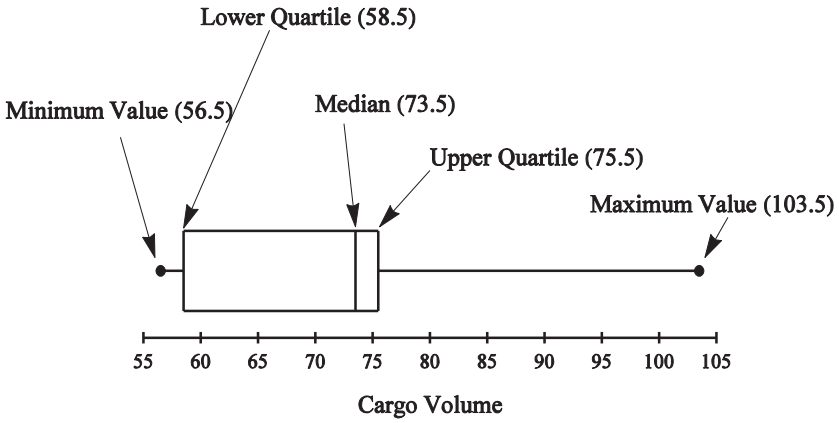
The complete Content Progression Chart for this module is provided on our website and at the beginning of the training manual. This portion of the chart illustrates how the skills included in this particular lesson develop as students advance through this accelerated course sequence.

6th Grade Skills/Objectives	7th Grade Skills/Objectives	Algebra 1 Skills/Objectives	Geometry Skills/Objectives	Algebra 2 Skills/Objectives	Pre-Calculus Skills/Objectives
Create, interpret, and compare dotplots (line plots), stemplots, and bar graphs.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.	Create, interpret, and compare dotplots (line plots), stemplots, bar graphs, histograms, and boxplots.
Calculate the mean, median, mode, range, and mean absolute deviation from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, range, and mean absolute deviation from tabular or graphical data or data presented in paragraph form	Calculate the mean, median, mode, range, and standard deviation from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, range, and standard deviation from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, range, and standard deviation from tabular or graphical data or data presented in paragraph form.	Calculate the mean, median, mode, range, and standard deviation from tabular or graphical data or data presented in paragraph form.

Box-and-Whisker Plots

Answers

1. a.



b. The boxplot gives one an idea of the shape of the data, specifically, how symmetrical the data appears. It also allows one to see if the data is generally spread out or if it is close together.

c. The median is identified by the middle of the "box." The mean cannot be calculated from the boxplot alone because the individual data points are unknown.

d. Range and IQR. The boxplot reveals the minimum and maximum points that allow a direct calculation of the range. Q_3 and Q_1 may be used to calculate the IQR.

e. No. The true weakness of the boxplot is that gaps and clusters are hidden. The actual data points are not revealed in this type of graph.

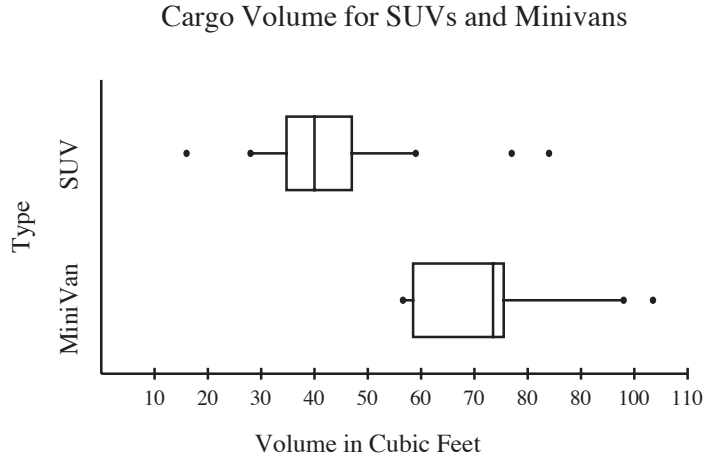
f. The data that lies between the upper and lower quartile is the middle 50% of the data.

g. IQR = 75.5-58.5=17
$Q_1 - 1.5(\text{IQR}) = 58.5\text{-}1.5(17) = 33$
$Q_3 + 1.5(\text{IQR}) = 75.5 + 1.5(17) = 101$
The maximum is an outlier since it is greater than 101. The minimum is not an outlier since it is not less than 33.

h. Lower quartile –1.5(IQR) = 33, so any value less than 33 is an outlier. Yes, it would be an outlier.

2. a. Minimum = 16; Q_1 = 34.75; median = 40; Q_3 = 47; Maximum = 84

b.



c. There are clearly two SUVs (marked as outliers) that have more cargo room than half of the minivans. If one had a graphical display of the actual data (like a stemplot), one could identify exact numbers to verify the claim.

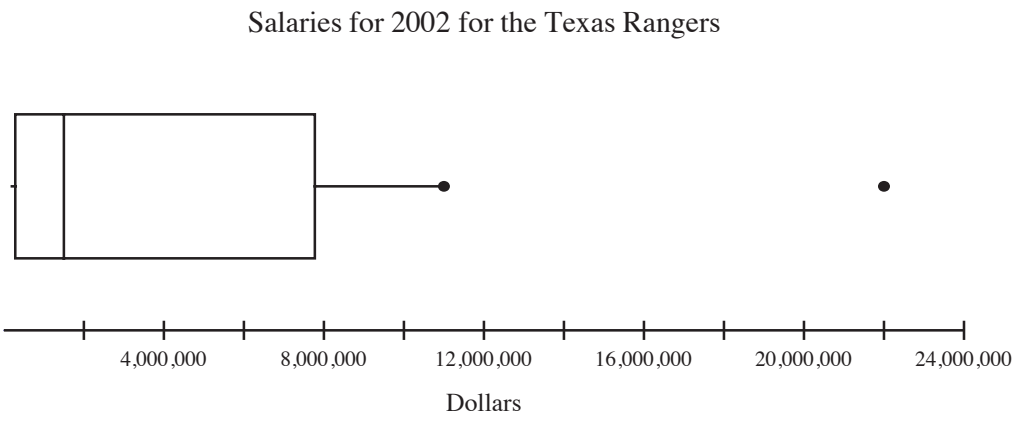
d. Yes. The middle of the minivan boxplot is the median that is higher than every point on the SUV boxplot that is not an outlier.

e. Answers vary. One can say that 75% of the minivans have more cargo room than every SUV that is not an outlier. One can also say the minivan median is closer in value to the upper quartile while the SUV median seems to be slightly closer to the lower quartile.

f. A parallel boxplot reveals exactly the same amount of information for both data sets, regardless of the amount of data in the set.

g. Since there are significantly more data points for the SUV than for the minivan, it may be difficult to draw conclusions from a stemplot. The stemplot would display the exact values for all the cargo volumes. Student answers will vary as to the preferred graphical display, but the boxplot makes it easier to compare the vehicles.

3. a.



Five number summary: 200,000; 285,000; 1,500,000; 7,733,333; 22,000,000.
Since 22,000,000 is an outlier, the highest salary that is not an outlier is 11,000,000.

b. A scale gives one an idea of the spread of the data. In addition, all graphs should include a scale. Labeling the five-number summary allows one to quickly identify those important numbers on the boxplot.

c. The minimum value and lower quartile are very close compared to the spread of the rest of the data. The small lower whisker means the lower 25% of the data is very close together. If a boxplot has the lowest number that is repeated and comprises 25% or more of the data, the boxplot will not reveal a lower whisker. So, in this problem, if Danny Kolb and Doug Davis both made $200,000 like Hank Blalock and Jovanny Cedeno, the lowest value and lower quartile would both be $200,000.

d. Yes. IQR = 7,733,333 – 285,000 = 7,448,333; 1.5(IQR) = 11,172,499.50;
Upper quartile + 1.5(IQR) = 18,905,832.50
Any salary above $18,905,832.50 is an outlier. Rodriguez's salary affects only the mean, not the median. Student should circle somewhere above (to the right of) the median.

e. If Rodriguez made $18,905,832, his salary would not be an outlier.

4. a. The median percent of iron is approximately 1.5% higher than the median percent of magnesium.

b. Answers will vary. Sample answers: The median iron composition is higher than the median magnesium composition. The median iron composition is higher than 75% (upper quartile) of the magnesium. The highest iron composition and magnesium composition are about equal.

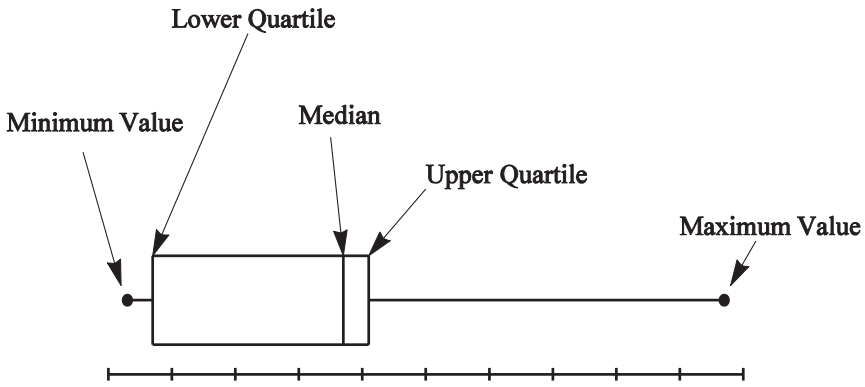
c. The iron IQR is about 5% while the magnesium IQR is a little less than 4%. To obtain these answers, approximate the upper and lower quartiles before subtracting to calculate the IQR.

Box-and-Whisker Plots

1. Karma's family is interested in buying a new minivan. Karma is involved with mountain bike racing, so the cargo volume is a very important consideration in buying the minivan. The table contains the cargo space of minivans available in 2002 measured in cubic feet.

Cargo Volume	Cargo Volume
98.0 cu. ft.	57.0 cu. ft.
75.5	56.5
73.5	56.5
73.5	75.5
67.0	75.5
98.0	63.0
67.0	103.5
58.5	

a. Determine the five-number summary and complete the boxplot provided, labeling the scale appropriately and indicating the value of each labeled point.



b. How does the boxplot help one visualize the data?

c. What measure(s) of central tendency can easily be identify from the boxplot? If only the boxplot is provided, what measure(s) of central tendency cannot be determined? Explain.

d. What measure(s) of spread can be easily calculated from the boxplot?

e. Sometimes data has gaps and clusters. Gaps occur when the data has large spaces between data points. Clusters occur when the data has many numbers that are very close together. Will a boxplot reveal gaps and clusters in the data? Explain.

f. Explain the significance of the data that lies between the upper and lower quartile.

g. Is the maximum or minimum value an outlier? Justify the answer by showing the process used to determine when a number is an outlier for a boxplot.

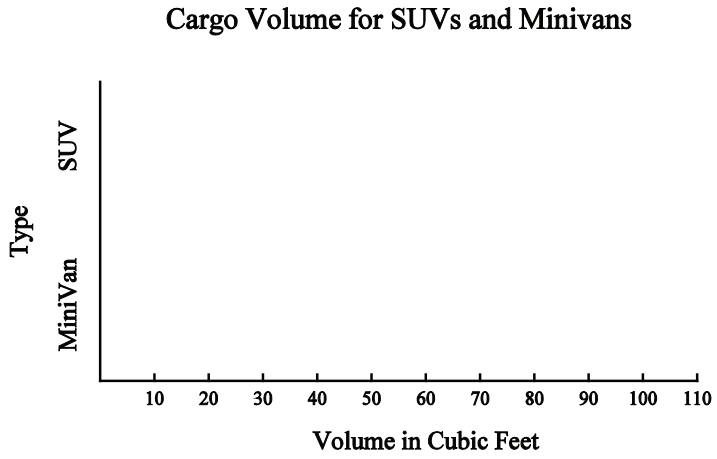
h. Suppose a vehicle with 31 cubic feet of cargo volume is classified as a minivan. Would it be considered an outlier?

2. Karma's family decides to choose one of the minivans that has the median cargo volume. Karma explains their purchase to her friend Roberto. He boasts that his family's SUV has 77 cubic feet of cargo room. Karma and Roberto argue about whether minivans or sports-utility vehicles tend to have more cargo room. Roberto searches the internet and locates data showing the cargo volumes, in cubic feet, of the 45 sports-utility vehicles available in 2002 as listed in the *Consumer Reports Auto Issue*, April 2002.

42	35.5	45	59	40	77	59	28	39
47.5	33.5	84	45.5	39	59	77	32	33.5
34.5	39.5	43	37	35	16	30	46.5	29
50	37.5	34.5	45.5	45.5	45	33	33.5	43.5
49	37	36	37.5	44	41.5	50	36.5	59

a. Using technology, calculate the five-number summary of the sports-utility vehicle data. If you do not have access to the appropriate technology, your teacher will provide the values of the 5-number summary.

b. Re-draw your boxplot from question 1 as a modified boxplot on the grid provided. Add a modified boxplot to display the SUV data.



c. Do all minivans have more cargo volume than sports-utility vehicles? Explain how the graph supports the answer. Explain how one could use graphical displays of the actual data to also support the claim.

d. Can one say that the minivans that are in the upper half of the boxplot have larger cargo volume than every SUV that is not an outlier? Explain using the graphs to support the answer.

e. What are the main points that Karma should make in trying to convince Roberto that minivans tend to have more cargo volume? Be specific.

f. What advantage does the boxplot provide in comparing the SUV data that has 45 data points to the minivan data that has 15 points?

g. What information is needed to draw a back-to-back stemplot for the data? What information could be determined from the stemplot that would not be available with the boxplot? Which graphical display would you prefer for this data set? Explain your position.

3. Andrew's cousin, Samuel, is visiting from Australia. He is interested in baseball and would like to know more information about baseball players and their annual salaries. Andrew and Samuel discover the following salary information about some of their favorite players.

Baseball Player	Salary	Baseball Player	Salary
Hank Blalock	$200,000	Dave Burba	$2,000,000
Frank Catalanotto	$2,475,000	Jovanny Cedeno	$200,000
Doug Davis	$310,000	Carl Everett	$8,666,666
Juan Gonzalez	$11,000,000	Rusty Greer	$6,800,000
Bill Haselman	$800,000	Danny Kolb	$260,000
Dan Miceli	$1,000,000	Alex Rodriguez	$22,000,000

a. Create a modified boxplot for the data. Be sure to title the plot and label the five-number summary.

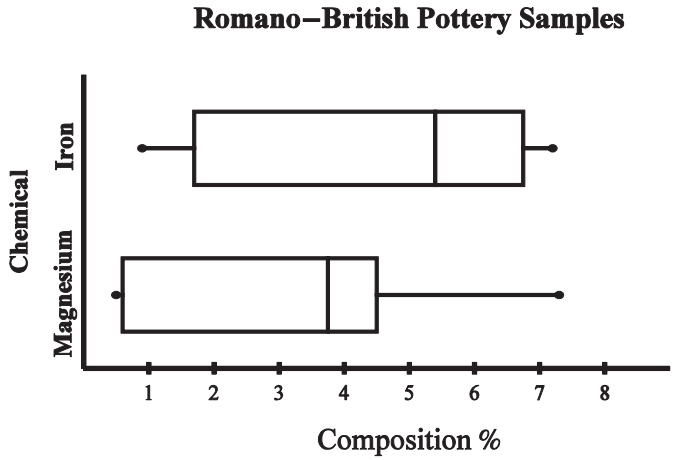
b. Why is it important to include a scale when displaying a boxplot? Why is it useful to label the five-number summary?

c. Why is the lower whisker so short? What does the length indicate about the data? Is it possible for a boxplot to have no lower whisker? Explain.

d. Is Alex Rodriguez an outlier? How does his salary affect the median salary? How does his salary affect the mean salary? Without actually calculating the mean salary, circle on the boxplot where the mean might be located.

e. What is the maximum salary, in whole numbers of dollars, that Alex Rodriguez could receive so that his salary would not be an outlier?

4. Twenty-six samples of Romano-British pottery were found at four different kiln sites. The percentage of oxides of two metals, magnesium and iron, measured by atomic absorption spectrophotometry, are displayed in the given boxplots.



a. Approximately how much larger is the median percent of iron than the median percent of magnesium?

b. Based on the graphs, make two statements about the differences in the percent of iron and magnesium.

c. Which oxide has a larger IQR? Explain how to determine the answer.