IM 9–12 MATH v.I

Unit 1

Sequences and Functions

ALGEBRA 1

Lesson 14

Outliers

Illustrative Mathematics®

Kendall Hunt

Learning Goal

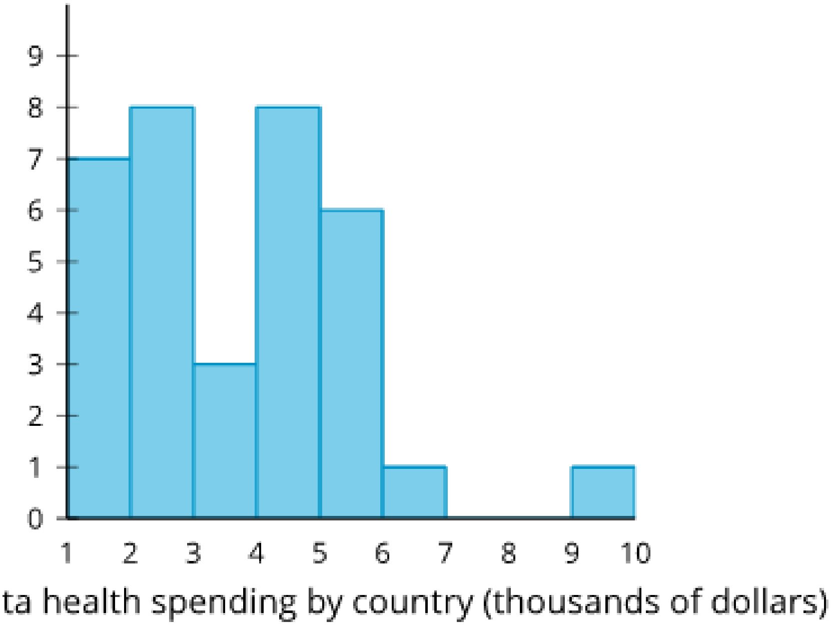
Let’s investigate outliers and how to deal with them.

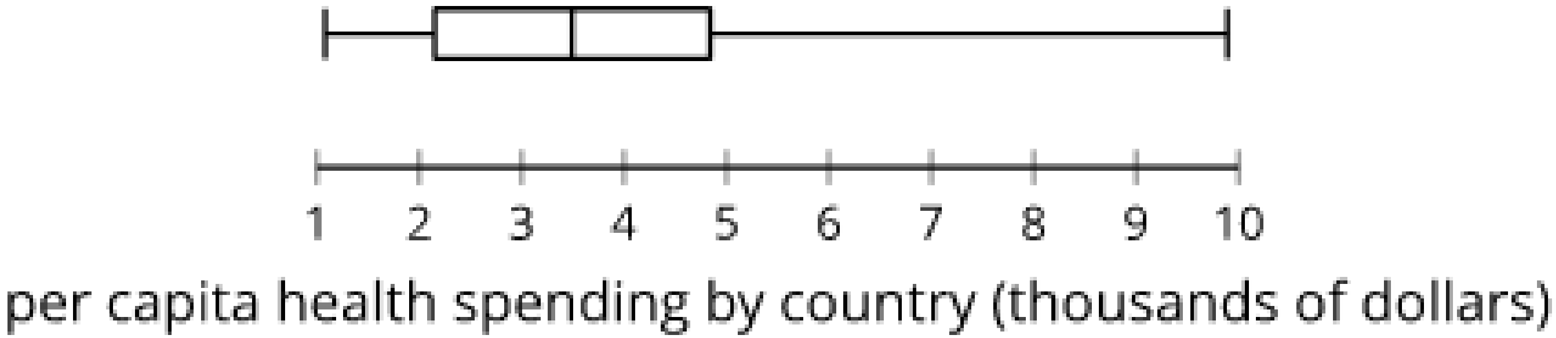
Algebra 1

Health Care Spending

Warm-up: Notice and Wonder

- What is one thing you notice?
- What is one thing you wonder?



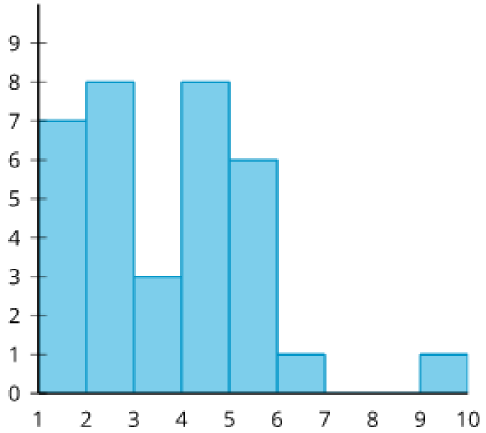


Health Care Spending

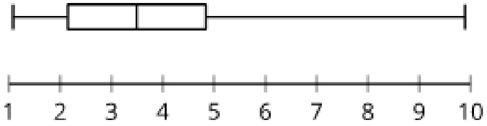
Warm-up: Notice and Wonder

The histogram and box plot show the average amount of money, in thousands of dollars, spent on each person in the country (per capita spending) for health care in 34 countries.

1. One value in the set is an outlier. Which one is it? What is its approximate value?
2. By one rule for deciding, a value is an outlier if it is more than 1.5 times the IQR greater than Q3. Show on the box plot whether or not your value meets this definition of outlier.



per capita health spending by country (thousands of dollars)



per capita health spending by country (thousands of dollars)

Health Care Spending

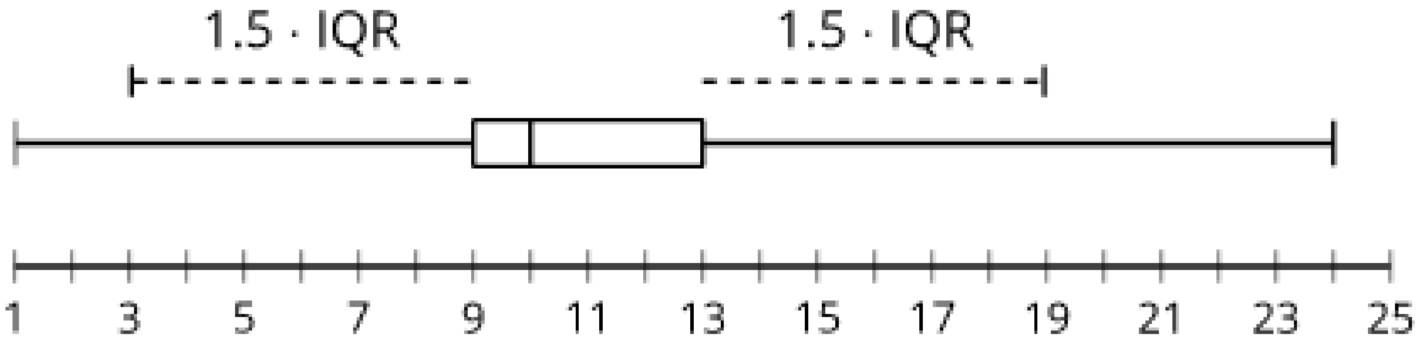
Warm-up: Notice and Wonder

- Values in a data set that are greatly different from the rest of the data are called outliers. The precise meaning of greatly different will be different for different situations. For example, a possible $4,000 difference in this graph does seem like a lot, but if the data represented the entire budgets of these countries in the billions or trillions of dollars (rather than spending on each member of the population for healthcare), it would not be a great difference.
- Using the IQR to determine outliers helps to adjust the difference to the variability of the bulk of the middle data. Using 1.5 times the IQR allows for some variability on the ends of the distribution to be considered usual.

Health Care Spending

Warm-up: Notice and Wonder

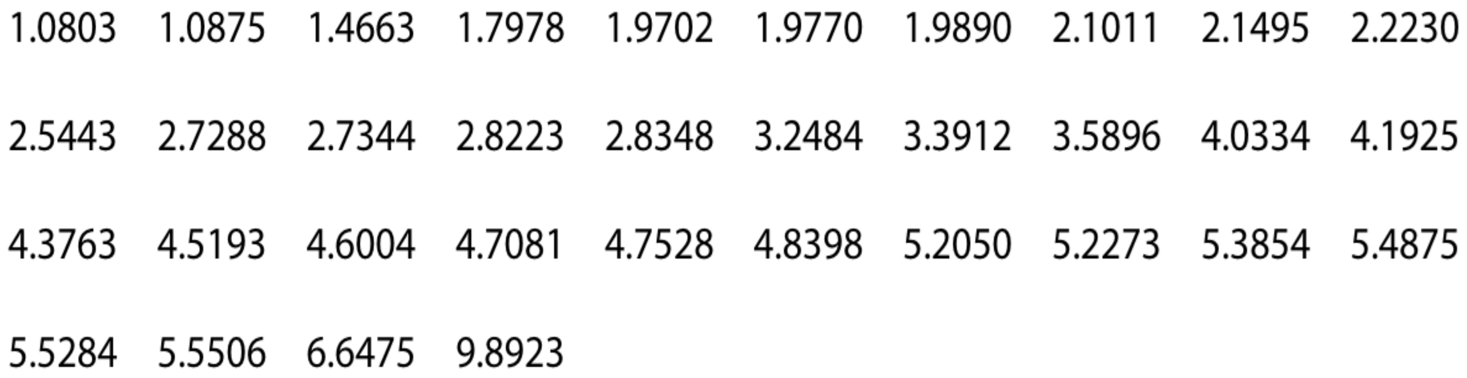
- It is also possible for there to be values that are unusually low compared to the rest of the data set. Consider this box plot that displays Q1 – 1.5 • IQR. The minimum value for this data set should be considered an outlier.



- For the purposes of this unit, a value will be considered an outlier for a data set if it is greater than Q3 + 1.5 • IQR or less than Q1 - 1.5 • IQR. These formulas compare extreme values to the middle half of the data to determine if the value should be considered an outlier.

Investigating Outliers

Here is the data set used to create the histogram and box plot from the warm-up.



1.0803	1.0875	1.4663	1.7978	1.9702	1.9770	1.9890	2.1011	2.1495	2.2230
2.5443	2.7288	2.7344	2.8223	2.8348	3.2484	3.3912	3.5896	4.0334	4.1925
4.3763	4.5193	4.6004	4.7081	4.7528	4.8398	5.2050	5.2273	5.3854	5.4875
5.5284	5.5506	6.6475	9.8923						

1. Use technology to find the mean, standard deviation, and five-number summary.
2. The maximum value in this data set represents the spending for the United States. Should the per capita health spending for the United States be considered an outlier? Explain your reasoning.

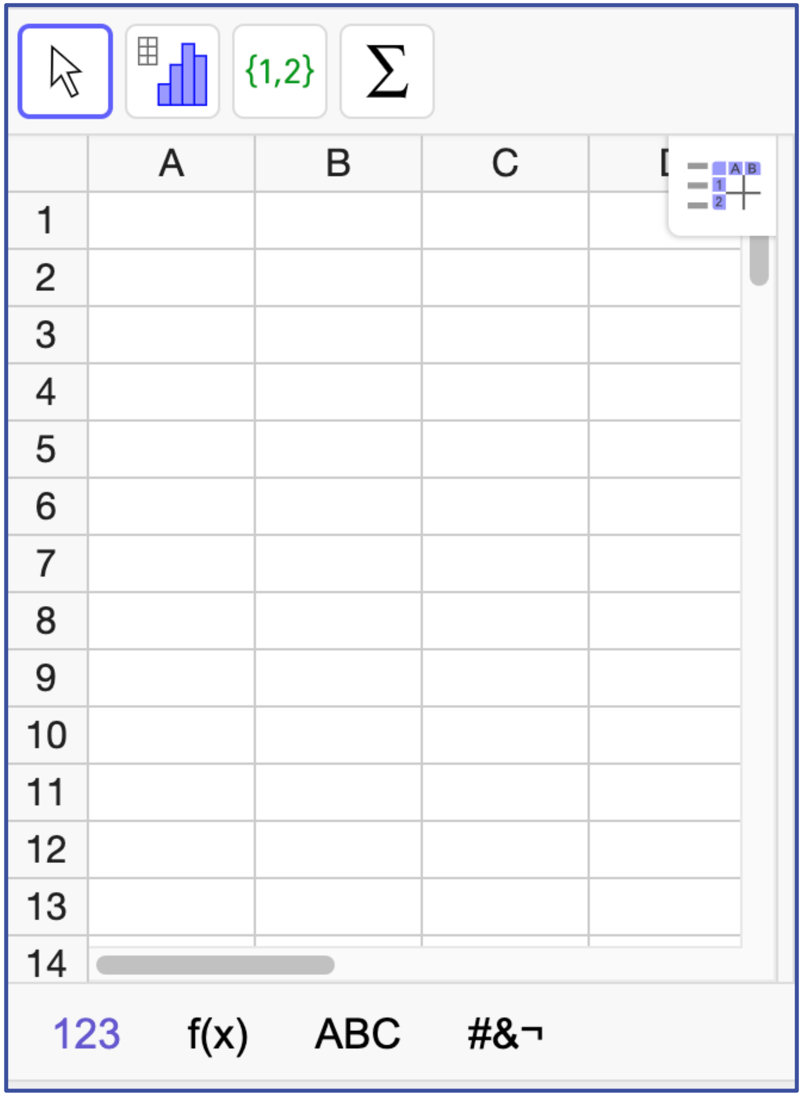
Investigating Outliers

3. Although outliers should not be removed without considering their cause, it is important to see how influential outliers can be for various statistics. Remove the value for the United States from the data set.

   a. Use technology to calculate the new mean, standard deviation, and five-number summary.

   b. How do the mean, standard deviation, median, and interquartile range of the data set with the outlier removed compare to the same summary statistics of the original data set?

Investigating Outliers



Investigating Outliers

- Do you think that 9.8923 should be eliminated from the data set? Why or why not?
- Which measure of center is more greatly affected by the inclusion of extreme values, the mean or median? Explain your reasoning.
- Which measure of variability is more greatly affected by the inclusion of extreme values, the standard deviation or the interquartile range? Explain your reasoning.

Origins of Outliers

1. The number of property crime (such as theft) reports is collected for 50 colleges in California. Some summary statistics are given:

15 17 27 31 33 39 39 45 46 48 49 51 52 59 72 72 75 77 77

83 86 88 91 99 103 112 136 139 145 145 175 193 198 213 230

256 258 260 288 289 337 344 418 424 442 464 555 593 699 768

- mean: 191.1 reports
- minimum: 15 reports
- Q1: 52 reports
- median: 107.5 reports
- Q3: 260 reports
- maximum: 768 reports

   a. Are any of the values outliers? Explain or show your reasoning.

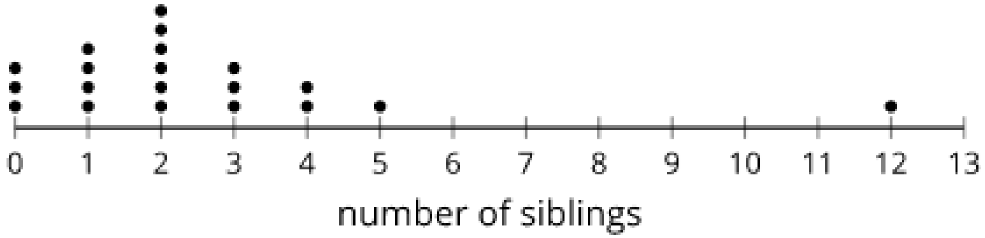
   b. If there are any outliers, why do you think they might exist? Should they be included in an analysis of the data?

Origins of Outliers

2. The situations described here each have an outlier. For each situation, how would you determine if it is appropriate to keep or remove the outlier when analyzing the data? Discuss your reasoning with your partner.

    a. A number cube has sides labelled 1–6. After rolling 15 times, Tyler records his data: 1, 1, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6, 20

    b. The dot plot represents the distribution of the number of siblings reported by a group of 20 people.



    a. In a science class, 12 groups of students are synthesizing biodiesel. At the end of the experiment, each group recorded the mass in grams of the biodiesel they synthesized. The masses of biodiesel are 0, 1.245, 1.292, 1.375, 1.383, 1.412, 1.435, 1.471, 1.482, 1.501, 1.532

Origins of Outliers

- Why is it important to analyze the source of outliers?
- What are reasons to keep an outlier in a data set?
- What are reasons to remove an outlier from a data set?
- What could be done about the 3 outliers for the college crime data to account for school size as the source of the outliers?
- How do you know that a value is an outlier?

Outliers

Lesson Synthesis

- What is an outlier?
- Why are outliers important to notice in a data set?
- How do outliers affect measures of center?
- How do outliers affect measures of variability?
- Why would you eliminate an outlier?

Unit 1 • Lesson 14

Learning Targets

- I can find values that are outliers, investigate their source, and figure out what to do with them.
- I can tell how an outlier will impact mean, median, IQR, or standard deviation

Algebra 1

Illustrative Mathematics®

Kendall Hunt

Expecting Outlier

Cool-down

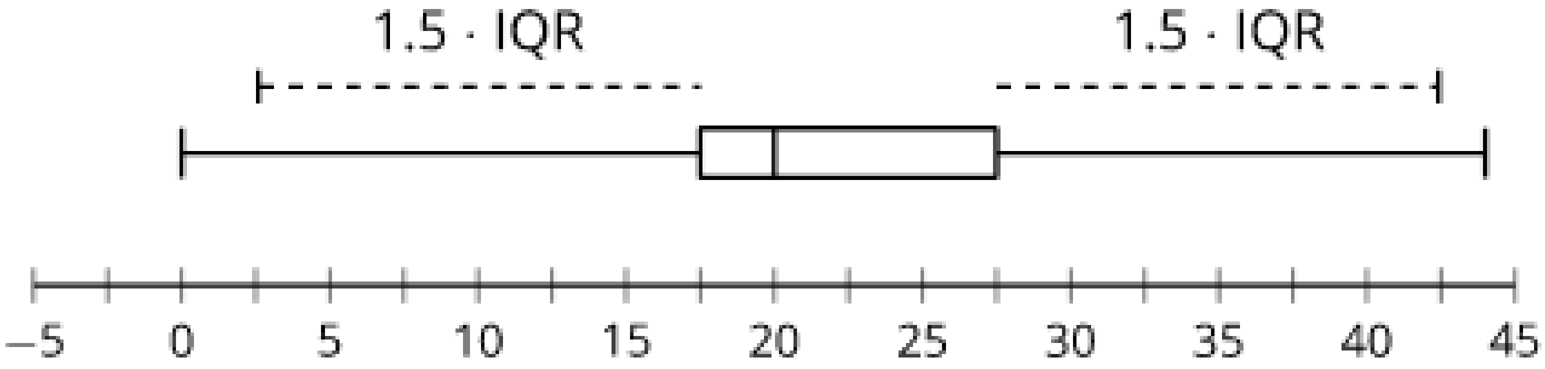
A group of 20 students are asked to report the number of pets they keep in their house. The results are 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 4, 4, 4, 21

- mean: 2.4 pets
- standard deviation: 4.47 pets
- Q1: 0.5 pets
- median: 1 pet
- Q3: 2.5 pets

1. Would any of these values be considered outliers? Explain your reasoning.
2. After being told that they should not count any fish in the report, the value of 3 becomes a 2 and the value of 21 becomes 1. Would these changes affect the median, mean, standard deviation, or interquartile range? If so, would each measure decrease or increase from their original values?

outlier

A data value that is unusual in that it differs quite a bit from the other values in the data set. In the box plot shown, the minimum, 0, and the maximum, 44, are both outliers.



standard deviation

A measure of the variability, or spread, of a distribution, calculated by a method similar to the method for calculating the MAD (mean absolute deviation). The exact method is studied in more advanced courses.

statistic

A quantity that is calculated from sample data, such as mean, median, or MAD (mean absolute deviation).

This slide deck is copyright 2020 by Kendall Hunt Publishing, https://im.kendallhunt.com/, and is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), https://creativecommons.org/licenses/by-nc/4.0/.

Illustrative Mathematics®

Kendall Hunt