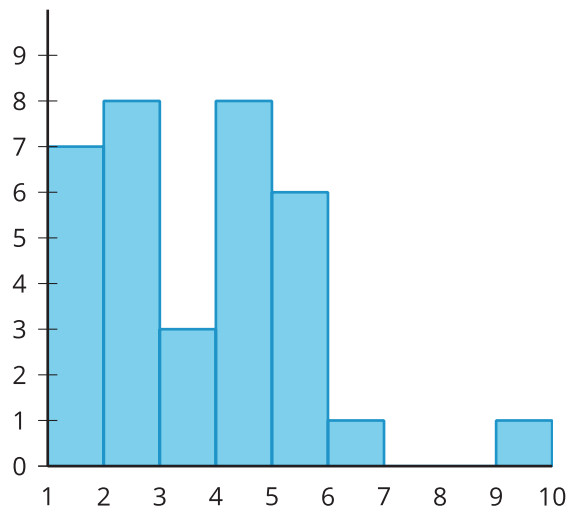# Unit 1 Lesson 14: Outliers
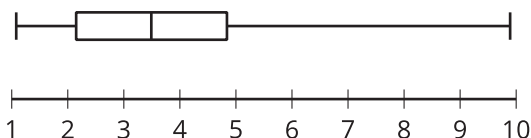
1 Health Care Spending (Warm up)

Student Task Statement

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.


per capita health spending by country (thousands of dollars)

per capita health spending by country (thousands of dollars)

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.

Activity Synthesis
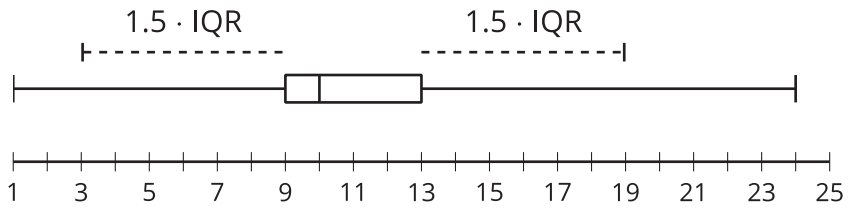
2 Investigating Outliers

Student Task Statement

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.

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?

3 Origins of Outliers

Student Task Statement

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?

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.

number of siblings
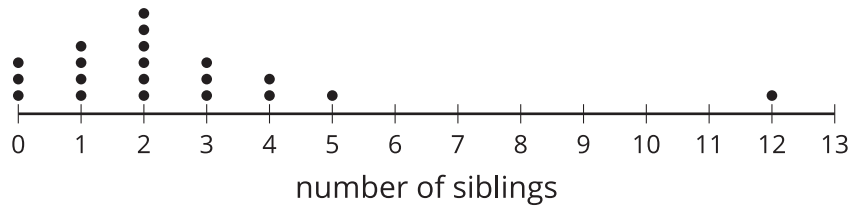
    c. In a science class, 11 groups of students are synthesizing biodiesel. At the end of the experiment, each group recorded the mass in grams of the biodiesel they synthesized. The masses of biodiesel are

$$0, 1.245, 1.292, 1.375, 1.383, 1.412, 1.435, 1.471, 1.482, 1.501, 1.532$$