1

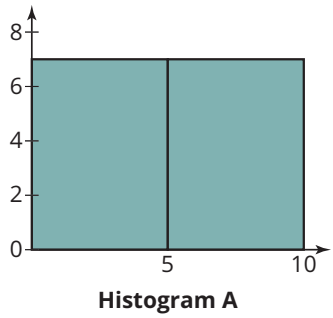
Represent!

Graphically Representing Data

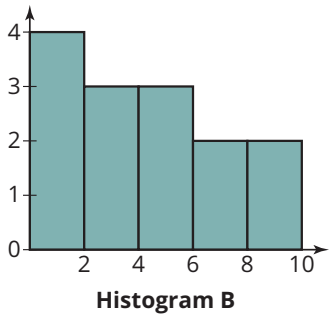
Warm Up

Consider the data set: 0.3, 0.7, 1.5, 1.9, 2.4, 3.0, 3.2, 5.3, 5.6, 5.8, 6.6, 7.5, 8.0, 9.1.

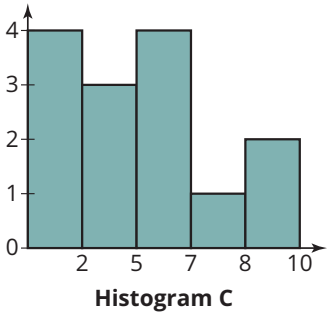
Determine which histogram best represents the data. Explain your reasoning.



Histogram A



Histogram B



Histogram C

Learning Goals

- Represent and interpret data displayed on dot plots, histograms, and box-and-whisker plots.
- Determine whether a dot plot, histogram, or box-and-whisker plot is the best way to display a data set.
- Compare the box-and-whisker plots of two different data sets.

Key Terms

- dot plot
- histogram
- bin
- frequency
- box-and-whisker plot
- five-number summary

You know how to represent and interpret data using dot plots, histograms, and box-and-whisker plots. How can you determine which representation is most appropriate given a data set?

Get the Lead Out

Ask yourself:

How is the statistical process similar to the modeling process?

Recall that there are four components of the statistical process:

- Formulating a statistical question.
- Collecting appropriate data.
- Analyzing the data graphically and numerically.
- Interpreting the results of the analysis.

An issue of concern for many cities is the level of lead found in the drinking water. Lead concentrations in drinking water should be less than 15 parts per billion (ppb). The system that supplies water to a city is required to collect samples of tap water from sites it services. A water system technician took samples of the amounts of lead in the water in one neighborhood of the city of Greenville and recorded the data in the table shown.

Site Number	Amount of Lead in Water (ppb)
1	11
2	22
3	6
4	10
5	8
6	3
7	12
8	5
9	10
10	4
11	4
12	7
13	7
14	11
15	7
16	5
17	1
18	13
19	4
20	7

1. Analyze the data collected using only the table. What conclusions can you draw about the amount of lead in the water at the different sites?

One way to better organize data in a table is to create a *dot plot*. A **dot plot** is a graph that shows how discrete data, or data that can be "counted," are distributed using a number line. Dot plots are best used to organize and display a small number of data points.

2. Construct a dot plot to represent the amount of lead in the water at each site. Make sure to label your dot plot.

3. What conclusions can you draw about the amount of lead in the water of the neighborhood in Greenville from your dot plot?

ACTIVITY
1.1

Histograms

According to the Environmental Protection Agency, if more than 10% of tap water samples are greater than or equal to 15 parts per billion (ppb), then action is required to reduce the levels.

1. According to the data in the Getting Started, is action required to reduce the amount of lead in the water in the neighborhood? Justify your response.

The fourth part of the statistical process is to interpret the results of your analysis.

The mayor of Greenville wants to analyze the lead levels for the water in the entire city. The frequency table displays the data for the amount of lead in drinking water samples taken from sites all over Greenville.

Amount of Lead in Water (ppb)	Frequency
0 up to 5	32
5 up to 10	48
10 up to 15	100
15 up to 20	47
20 up to 25	23

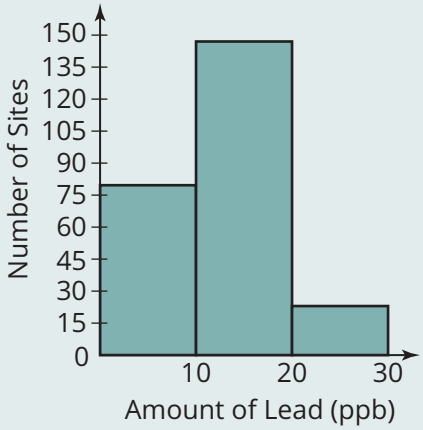
Another way to display quantitative data is to create a *histogram*. A **histogram** is a graphical way to display quantitative data using vertical bars. The width of a bar in a histogram represents an interval of data and is often referred to as a *bin*. The height of each bar indicates the **frequency**, which is the number of data values included in any given bin. Histograms are effective in displaying large amounts of continuous data, or data which can take any numerical value within a range.

2. Construct a histogram to display the data in the table.

A **bin** represents an interval of data instead of individual data values. The value shown on the left side of the bin is the least data value in the interval.

3. **What conclusions can you draw from the histogram about the amount of lead in the drinking water of Greenville?**

4. **Marcel created a histogram to display the same data and claimed that no action had to be taken to reduce lead levels in Greenville's drinking water.**



Is Marcel correct? Explain why or why not.

Ask yourself:

Do more than 10% of the samples have amounts of lead greater than or equal to 15 ppb?

5. **Does the water system management of Greenville need to take action to reduce the amount of lead in the water? Explain your reasoning.**

6. **Do you think a histogram is a good representation of the data? Explain your reasoning.**

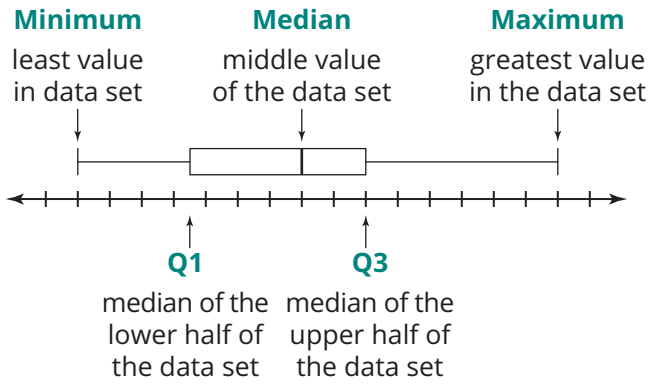
ACTIVITY
1.2 Box-and-Whisker Plots

The governor wants to compare the amount of lead in the drinking water of different cities in her state to see where improvements in the water systems should be made.

You can visually compare two large data sets at a glance using *box-and-whisker plots*. A **box-and-whisker plot** is a graphical representation that displays the distribution of quantitative data based on a *five-number summary*. The **five-number summary** consists of the minimum value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum value.

Worked Example

The five-number summary is used to create a box-and-whisker plot. Each vertical line of the box-and-whisker plot represents a value from the summary.



There are four sections of the graphical display: minimum to Q1, Q1 to median, median to Q3, and Q3 to maximum. Each section of the box-and-whisker plot represents 25 percent of the data set.

The five-number summaries of the data for the amount of lead in drinking water samples taken from the cities of Greenville and Oaktown are given.

Five-Number Summary	
Greenville	**Oaktown**
minimum = 1	minimum = 2
Q1 = 8	Q1 = 6
median = 12	median = 8
Q3 = 16	Q3 = 9
maximum = 22	maximum = 15

1. Construct a box-and-whisker plot of the data for each city on the same number line using the five-number summaries.

2. Suppose you work in the governor's office. Compare the data displayed in your box-and-whisker plots and write an analysis to present to the governor.

TALK the TALK

Dots, Bins, or Boxes?

Analyze each situation. Describe which representation you would use (dot plot, histogram, or box-and-whisker plot) to display and analyze the data set. Explain your reasoning.

1. **Nick collects data about the one hundred tallest buildings in the United States. He wants to quickly determine between which two heights the top 25% of the data fall.**

2. **Lily conducted a school survey to determine how many problems each math teacher assigned for homework on Friday.**

3. **The migration ranges of the white deer and mule deer populations in Montana are recorded and compared.**

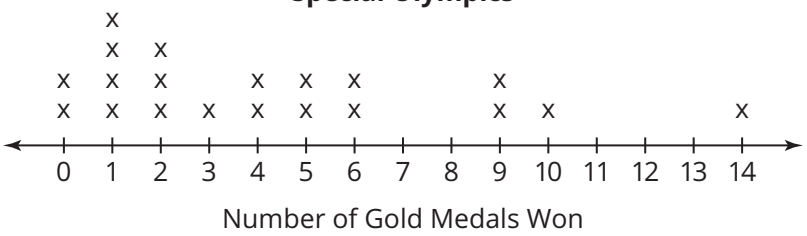
4. **A school district wanted to determine how many students in the entire district scored above 70 on a standardized test.**

NOTES

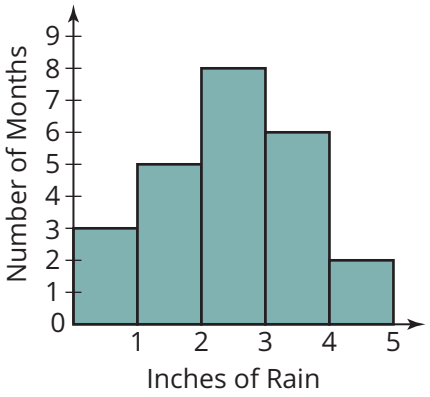
NOTES

5. What conclusions can you draw from each display?

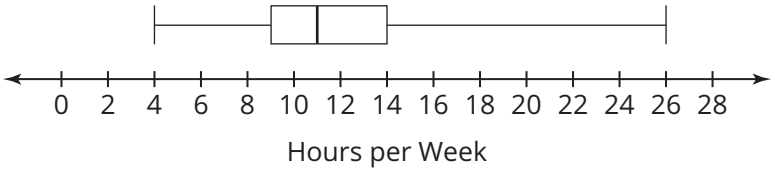
Participants Who Won Gold Medals at the Special Olympics



Rain in Collinsburg



Volunteers Hours at the Local Animal Shelter



Assignment

Write

Complete each statement.

1. A _____ is a graphical way to display quantitative data using vertical bars.
2. A _____ displays the data distribution based on a five number summary.
3. A _____ is a graph that shows how data are distributed using a number line.
4. For a set of data, the _____ consists of the minimum value, the first quartile, the median, the third quartile, and the maximum value.
5. The number of data values included in a given bin of a data set is called the _____.
6. The bar width in a histogram that represents an interval of data is often referred to as a _____.

Remember

A dot plot is useful for organizing a small number of data points. A histogram is effective in displaying large amounts of data. Box-and-whisker plots are effective for visually comparing two data sets.

Practice

1. Mr. Follweiller finished grading the quizzes for one of his Algebra 1 classes. The table shown is the recorded grades of the class.
   a. Mr. Follweiller is worried that his students may not have understood the material covered on the quiz. He would like to get a better idea of how the class did as a whole. Would you recommend that he make a dot plot, a box-and-whisker plot, or a histogram to display this data? Explain your reasoning.
   b. Construct a dot plot and histogram of the data in the table.
   c. What information does the dot plot provide that the histogram does not?
   d. The students argue that more than half the students failed the quiz, so they think Mr. Follweiller should let them retake it. A grade of 56 is failing. Construct a box-and-whisker plot of the data. Are the students correct? Explain your reasoning.

Student	Grade	Student	Grade
A	85	N	53
B	89	O	71
C	66	P	90
D	74	Q	65
E	77	R	55
F	72	S	98
G	64	T	53
H	55	U	62
I	61	V	55
J	52	W	64
K	81	X	62
L	61	Y	56
M	71	Z	87

Stretch

George bowls in tournaments on the weekends. He recorded the scores of each game for his last two tournaments. A perfect score is 300.

Tournament 1: 182, 197, 178, 272, 180, 188, 202, 179, 191
Tournament 2: 188, 195, 177, 192, 180, 187, 201, 183, 197

Calculate the five-number summary and IQR for the two tournaments. Interpret your findings.

Review

1. The table shows an example of a rabbit population.

Year	0	1	2	3	4	5	6
Population	4	11	29	79	213	577	1557

   a. Create a scatter plot of the data.
   b. What is the regression equation? Graph the equation on the grid with the scatter plot.
   c. How did you determine what type of function to use?
   d. What do you predict the rabbit population will be in the 20^{th} year? Explain your reasoning.