Dear Family,

The next unit in your child's mathematics class this year is ***Data About Us: Statistics***. Its focus is data investigation, and it teaches students to organize, display, analyze, and interpret data. Your child will learn to make and interpret many different types of data displays and to compute statistics to help describe data.

UNIT GOALS

The unit provides opportunities for students to ask questions about themselves, and then to collect data to help answer these questions. Students explore the lengths of their names, the distances they live from school, the numbers of times they can jump rope, the numbers of pets they have, their heights, and the lengths of their left feet.

Your child will learn to make line plots, bar graphs, coordinate graphs, and stem-and-leaf plots and to interpret patterns shown in these displays. Your child will also learn to compute the mode, median, mean, and range of a data set and to use these statistics to describe data and to make predictions.

HELPING WITH HOMEWORK

You can help with homework and encourage sound mathematical habits as your child studies this unit by asking questions such as:

- What is the question being asked?
- How do you want to organize the data?
- Which representation is best to use to analyze the distribution of the data?
- How can you use graphs and statistics to describe a data distribution or to compare two data distributions in order to answer the original question?
- How do you think the data were collected?
- Why are these data represented using this kind of graph?

In your child's notebook, you can find worked-out examples from problems done in class, notes on the mathematics of the unit, and descriptions of the vocabulary words.

HAVING CONVERSATIONS ABOUT THE MATHEMATICS IN *DATA ABOUT US*

You can help your child with his or her work for this unit in several ways:

- Look with your child for uses of data in magazines, newspapers, and on TV.
- Point out examples of graphical displays and ask your child questions about the information shown.
- Ask your child about the data studied in class. What were the typical values (mode, median, or mean) for these data?
- Look over your child's homework and make sure all questions are answered and that explanations are clear.

A few important mathematical ideas that your child will learn in *Data About Us* are given on the back. As always, if you have any questions or concerns about this unit or your child's progress in class, please feel free to call.

Important Concepts and Examples

Representing Data Distributions and Reading Data Representations

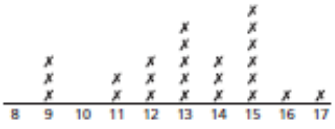
Statisticians use data representations such as line plots, bar graphs, stem-and-leaf plots, and coordinate graphs to describe and analyze their data.

READING STANDARD DATA REPRESENTATIONS

- *Reading the data* involves "lifting" information from a graph to answer explicit questions.
- *Reading between the data* includes the interpretation and integration of information presented in a graph.
- *Reading beyond the data* involves extending, predicting, or inferring from data to answer implicit questions.

LINE PLOT Each case is represented as an "X" positioned over a labeled number line.

Name Lengths of Ms. Jee's Students



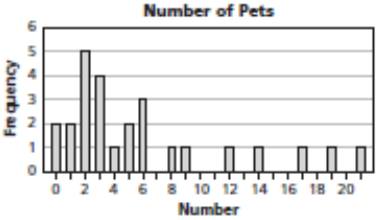
STEM-AND-LEAF PLOT
A plot that permits students to group data in intervals (usually by 10s).

Movies Watched

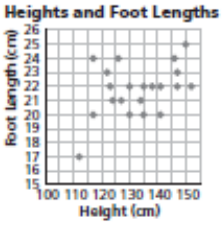
0	3 3 5 6 6 7
1	1 5 6 8
2	

Key: 1 \| 5 means 15 movies

FREQUENCY BAR GRAPH
A bar's height is not the value of an individual case but rather the number (frequency) of cases that all have that value.



SCATTERPLOT
The relationship between two variables is explored by plotting data values on a Cartesian coordinate system.



Using Measures of Center (Mode, Median, Mean)

MODE The mode is the value that occurs with greatest frequency in a set of data.

14 students said that they had the following number of siblings: 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 5, 6. The mode is 2.

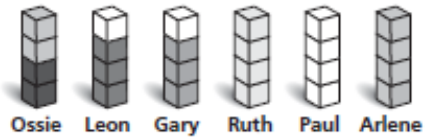
MEDIAN The median value marks the location that separates an ordered set of data in half.

The median for the data set 3, 4, 4, 7, 8, 9 is 5 , the number halfway between 4 and 7. For 4, 5, 5, and 7, the median is 5.

MEAN We emphasize the fair share (or evening out) interpretation of mean (average).

The mean (average) number of people in these households is 4. There are 24 people shared among 6 households.

BEFORE		AFTER	
Ossie	2 people	Ossie	4
Leon	3 people	Leon	4
Gary	3 people	Gary	4
Ruth	4 people	Ruth	4
Paul	6 people	Paul	4
Arlene	6 people	Arlene	4
Total	24 people		24 people



Using Measures of Variability

Measures of variability are used to describe how widely spread or closely clustered the individual data values are.

RANGE The range depends on only two values, the greatest and the smallest.

Distinguishing Different Types of Data

NUMERICAL DATA are values that are counts or measures (pulse, height). We can use mean, median, mode, and range as summary statistics.

CATEGORICAL DATA are data sets that are responses representing categories (favorite color, month of birth, etc.). We can use only the mode as the summary statistic.

Investigation 1

ACE Assignment Choices

Differentiated Instruction

Problem 1.1
Core 1, 22–25

Problem 1.2
Core 2, 5–12
Other *Applications* 3, 4; *Connections* 26–28; unassigned choices from previous problems

Problem 1.3
Core 13
Other *Connections* 29, 30; unassigned choices from previous problems

Problem 1.4
Core 14–20, 31
Other *Connections* 32, *Extensions* 40–43; unassigned choices from previous problems

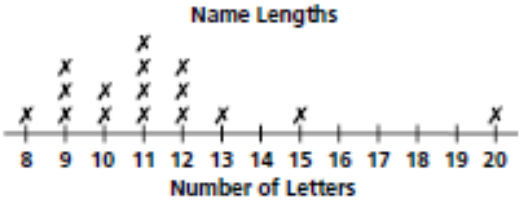
Problem 1.5
Core 21
Other *Extensions* 33–39; unassigned choices from previous problems

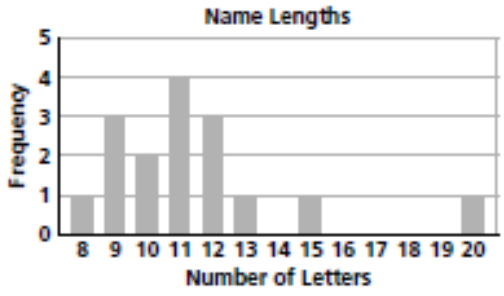
Adapted For suggestions about adapting Exercises 3–6 and other ACE exercises, see the CMP *Special Needs Handbook*.

Applications

1. Name Lengths

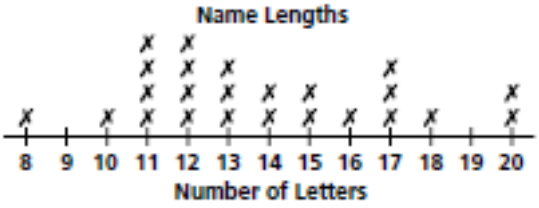
Name	Number of Letters
Ben Foster	9
Ava Baker	8
Lucas Fuentes	12
Juan Norinda	11
Ron Weaver	9
Bryan Wong	9
Toby Vanhook	11
Katrina Roberson	15
Rosita Ramirez	13
Kimberly Pace	12
Paula Wheeler	12
Darnell Fay	10
Jeremy Yosho	11
Cora Harris	10
Corey Brooks	11
Tijuana Degraffenreid	20



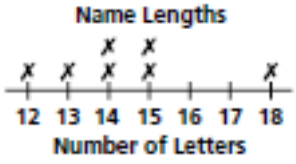


2. The median name length is 11 letters, and the range of the data is 12 letters. A name length of 20 letters is somewhat unusual. The typical number of letters is clustered around the median in an interval of 8–13 letters or 9–12 letters. The mode is the same as the median in this example, although 9 letters and 12 letters occur almost as frequently as the mode.

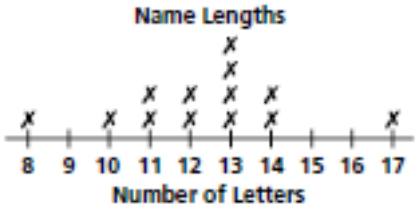
3. Possible answer:



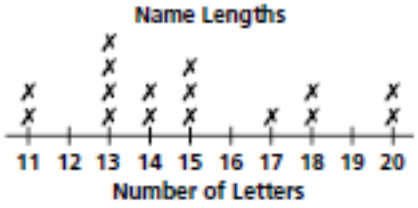
4. Possible answer:



5. Possible answer:



6. Possible answer:



7. Possible answer: The data distribution from Ms. Campo's class shows that the typical number of letters is clustered around the median in an interval of 10–17 or 12–16 letters. This is slightly higher than the data distribution from Mr. Young's class. The data for Ms. Campo's class vary from 10 to 19 letters; the data for Mr. Young's class vary from 8 to 20 letters.

8. C

9. H

10. 27 students; the bar for each number represents the number of students with that name length, so adding the bar heights (1 + 2 + 4 + 3 + 4 + 7 + 3 + 2 + 1) gives the total number of students.

11. 9 letters

12. 14 letters; there are 27 name lengths, so the median occurs at the fourteenth name length, which is 14 letters.

13. a. (Figure 2)

    b. The median will change because three pieces of the new data are above the original median and only one is below.

14. numerical

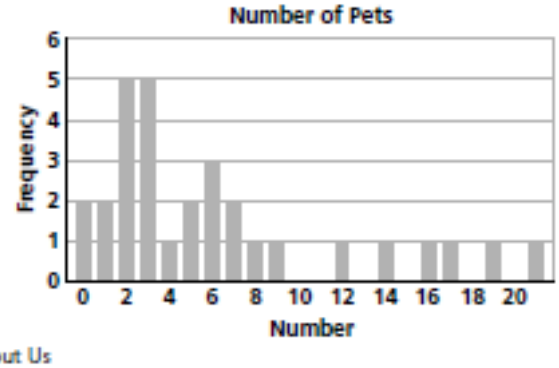
15. categorical

16. categorical

17. categorical (NOTE: The question asks for "yes" or "no," not for a number.)

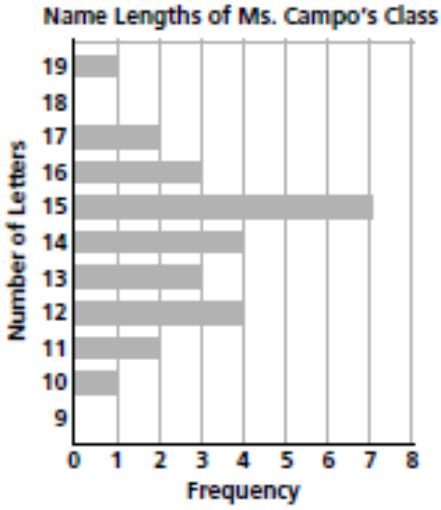
18. categorical

Figure 2



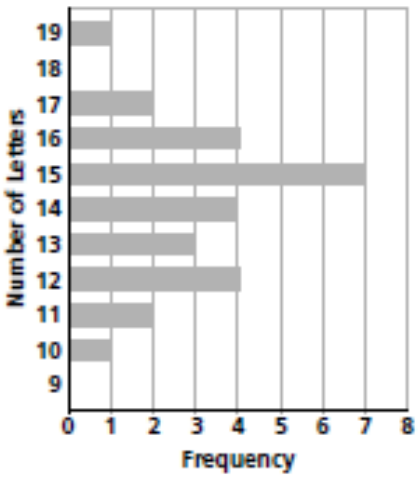
19. numerical **20.** numerical

21.



a. The median name length is 14 letters. It is the same data, just represented horizontally, so the median remains the same. (There are 27 data values; the median is the fourteenth data value, which happens to be 14 letters, in an ordered list of the data.)

b.



The median name length is $14\frac{1}{2}$ letters (There are now 28 data values; the median is the average of the fourteenth and fifteenth data values, which are 14 and 15 letters, in an ordered list of the data values.)

Connections

22. Possible answer: Graphs A and B; Graph C has values that are 0; since the students are children, their families could not have 0 children. Some students might argue that Graph A is not correct because it is unlikely that there will be a lot of families with 5, 7, and 8 children.

23. Possible answer: Graphs A and B; Graph A and Graph B are labeled 1 through 12 on the horizontal axis. However, some students might argue that in Graph B, it is unlikely that 10 students were born in February. Graph C only has labels from 0 to 9 on the horizontal axis.

24. Any of the graphs could show numbers of pizza toppings. Some students might argue that Graph C is correct because most of their friends like two or three toppings.

25. Possible answers:

Graph A: Birth Months of Students, Birth Month, Frequency

Graph B: Number of Children in Students' Families, Number of Children, Frequency

Graph C: Number of Pizza Toppings, Number of Toppings, Frequency

26. The median is the number that separates the ordered data in half. The number of people that consume 5 juice drinks in one day is near the upper end of the data, so 5 cannot be the median.

27. There are 100 students, so the median is between the fiftieth and fifty-first ordered data values. A total of 39 students consumed 0 or 1 juice drink in one day. This means the median is greater than 1 juice drink, because the fiftieth value will be in the bar that represents 2 juice drinks in one day.

28. B

29. a. $\frac{29}{100}$ b. $\frac{16}{100} = 16\%$

30. The total number of juice drinks students consumed is determined by evaluating each bar of the graph:

7 people $\times$ 0 juice drinks $=$ 0 juice drinks

32 people $\times$ 1 juice drink $=$ 32 juice drinks

29 people $\times$ 2 juice drinks $=$ 58 juice drinks

16 people $\times$ 3 juice drinks $=$ 48 juice drinks

6 people × 4 juice drinks = 24 juice drinks

5 people × 5 juice drinks = 25 juice drinks

3 people × 6 juice drinks = 18 juice drinks

1 person × 7 juice drinks = 7 juice drinks

1 person × 10 juice drinks = 10 juice drinks

So, 100 students consumed a total of 222 juice drinks in one day.

31. Numerical; the answer to the question, "How many juice drinks do you consume in one day?", is a number.

32. **a.** Half of all rats live less than $2\frac{1}{2}$ years, and half live longer than $2\frac{1}{2}$ years.

b. If Alex knew the greatest age of a rat, he would know how much longer his rat could possibly live.

Extensions

33. The bar height, 7, represents the number of stickers left. Because there were 12 to begin with, 5 have been sold.

34. The bar height, 4, represents the number of street signs left. Because there were 12 signs to begin with, 8 have been sold.

35. The bar graphs show that the number of stickers remaining is less than the number of street signs remaining. Students may want to debate this because of the "peaks" in the data. You will want to remind them that the bar graphs show the number remaining, not the number sold.

36. The store has collected $144 from the sale of 96 name stickers.

37. The most stickers, 12, have been sold for Amanda. The fewest stickers, 2, have been sold for Ana.

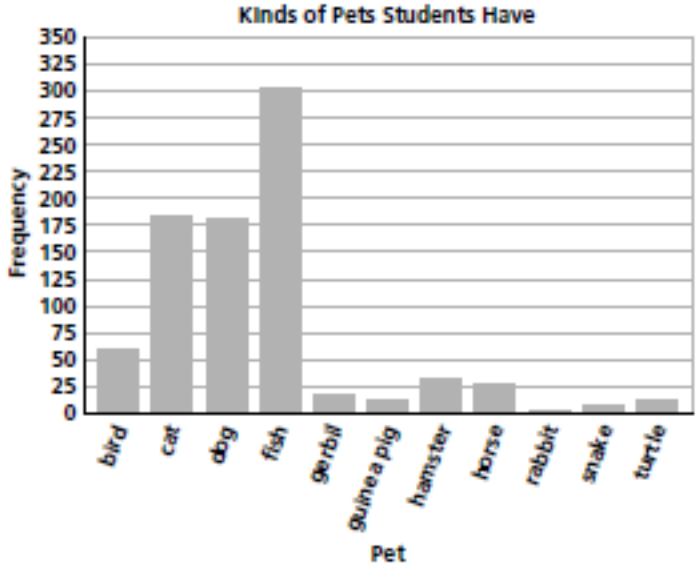
38. For Amy, the bars for stickers and signs are the same height. This shows that the numbers of stickers and signs sold are the same.

39. The stacked bars allow us to look at the data for stickers and street signs together. For example, Amanda has 0 stickers and 11 street signs left, while Alicia has 7 stickers and 4 street signs remaining. These names have the same total number of items remaining because the stacked bars are the same height. Allison and Amber are the most popular names because their stacked bars are the shortest. Ana is the least popular name because it has the highest stacked bar.

40. Possible answer: (Figure 3)

The challenge for students will be developing the scale for the vertical axis. Because of the range of the data (2 to 303 pets), the scale probably needs to be numbered by at least tens or twenties.

Figure 3



41. Possible answer: Fish occur the most frequently, followed by cats and dogs. In Problem 1.4, dogs occur most frequently, followed by cats, but the numbers are much smaller. The remaining pets are not like those of the students in Problem 1.4. Many of these pets are "indoor" pets. In Problem 1.4, many of the pets were "outdoor" pets that would live on a farm or in more rural areas.

42. Agree, because $\frac{61 + 184 + 180}{841} = \frac{425}{841} \approx 50\%$.

43. Answers will vary. Some students may immediately respond that 841 people were surveyed, indicating that each person surveyed had one pet. Other students may note that this response does not take into account that it is likely that some people surveyed had no pets or had more than one pet. This may lead students to look back at the data from Problem 1.4, where they know both the total numbers of pets and the number of people surveyed. From these data, students might find the median number of pets per person to be $3\frac{1}{2}$. Then they might divide 841 pets by 3.5 per student to get the possible number surveyed (about 240 students). Some students may raise a concern that the data from Problem 1.4 may reflect a special group of students who live in the country, and therefore, often have more pets; perhaps these particular data do not reflect the kinds of people surveyed for Problem 1.4. Students may have other strategies as well.

Possible Answers to Mathematical Reflections

1. A table of data, a line plot, and a bar graph are all tools for organizing and visualizing data. All three indicate the possible values of the item being measured (for example, 10 letters, 11 letters). A line plot and a bar graph indicate the number of times each value occurs.

   A line plot has a horizontal axis that shows the possible values with marks above the numbers indicating the number of times each value occurs.

   Like the line plot, a bar graph has a horizontal axis showing the possible values. Instead of using marks, the number of times a value occurs is indicated by the height of a bar over the value. A vertical axis indicates the frequency, corresponding to the height of each bar.

   A line plot is usually vertical, but a bar graph can be vertical or horizontal.

2. The mode is the value in a data set that occurs most frequently. There may be more than one mode, and a mode may occur at any location in the data. The mode can describe both categorical and numerical data. In categorical data, the mode would tell you which category occurs most frequently, and in numerical data the mode tells you which numerical value occurs most frequently.

3. The median is the value that divides an ordered set of data in half; half the data are below the median, and half the data are above the median. The median is not easily affected by the addition of very high or very low values. A median can only be used with numerical data because categorical data cannot be ordered.

4. The mode and the median for a set of data may or may not be the same. For the data set 1, 2, 3, 3, 4, 5, 6, both the median and the mode are 3. For the data set 1, 1, 1, 3, 5, 6, 6, the mode is 1 and the median is 3.

5. The range indicates how spread out the data are. Combined with a measure of center such as the median or the mode, the range helps to give a picture of the data. For example, if you know a data set has a median of 20, you know where the middle of the data set is. If, in addition, you know the range is 4 (or 60), you have a much better idea of what the data may look like. Range can only be used to describe numerical data.

6. The range is the difference between the least and greatest data values.

7. The mode, median, and range can be used to describe what is typical about a data set. You can also give an interval in which most of the data values fall. Giving information about the shape of the data (peaks, gaps, clusters) also helps describe what is typical.

Investigation 2

ACE Assignment Choices

Problem 2.1
Core 1–4

Problem 2.2
Core 5–7, 10, 13
Other *Extensions* 14; unassigned choices from previous problems

Problem 2.3
Core 8, 11
Other unassigned choices from previous problems

Problem 2.4
Core 12
Other *Applications* 9, *Extensions* 15; unassigned choices from previous problems

Adapted For suggestions about adapting Exercise 8 and other ACE exercises, see the CMP *Special Needs Handbook*.
Connecting to Prior Units 11, 12: *Bits and Pieces I*

Applications

1. A
2. H
3. 26 students; you can count the number of leaves on the stem plot. Each value represents one student.
4. Answers will vary. Students may find the median, which is 18.5 min. They may offer other alternatives as well; however, they must provide clear reasoning for their responses.
5. **Student Ages**

6	8
7	3 6 8
8	0 1 2
9	0 9
10	1 1 3 5 8
11	3 4
12	0 0 9
13	2 2 2 8
14	0 4 5 6 8 9
15	2

Key: 7 \| 6 means 76 months

6. 6.7 to 7.4 yr, or about $6\frac{1}{2}$ to $7\frac{1}{2}$ yr; divide the number of months by 12 to convert to years.
7. $113\frac{1}{2}$ mo (about $9\frac{1}{2}$ yr); there are 30 data values, so the median is the value halfway between the fifteenth and sixteenth values (113 and 114).

8. **a.** (Figure 2)

   b. You can locate the youngest student (the furthest to the left on the horizontal axis) and the shortest student (the closest to the bottom on the vertical axis). You can quickly see that the youngest student is the shortest student.

   c. In general, as students get older, their heights increase.

   d. People stop growing in their late teens or early twenties. The graph would level out at this time, and we would not see much increase afterward.

9. **a.** The graph indicates that, in general, taller people have longer foot lengths. However, knowing a person's foot length will not definitively tell you that person's height.

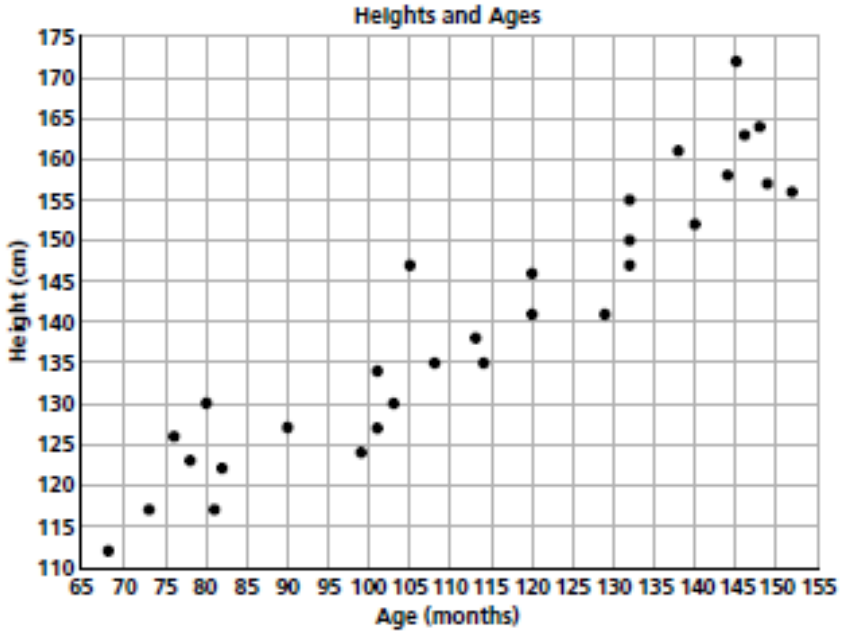
   b. The median height is 141 cm. (NOTE: If students use the table they will find a median height of $139\frac{1}{2}$ cm. This is because the graph takes the heights of only 29 students out of the 30 students on the table. Only one student with the height of 127 cm and the foot length of 21 cm is represented on the graph). The median foot length is 22 cm. Dividing 141 by 22, we see that the median height is a little more than 6.4 times the median foot length.

   c. Answers will vary. Height is generally about 6 to $6\frac{1}{2}$ times foot length.

   d. The answers to parts (b) and (c) show that a person's height is generally 6 to $6\frac{1}{2}$ times his or her foot length, so we can use foot length to estimate height. However, we cannot know the exact height for certain.

   e. If you started the graph at 0, the data points would be shifted horizontally to the right about 20 spaces and vertically upward about 14 spaces. There would be no data points in the lower left-hand region of the graph.

Figure 2



Connections

10. a.

Student Heights

11	2 7 7
12	2 3 4 6 7 7
13	0 0 4 5 5 8
14	1 1 6 7 7
15	0 2 5 6 7 8
16	1 3 4
17	2

Key: 13 | 5 means 135 centimeters

b. The least value is 112, and the greatest value is 172. These points give us the endpoints with which you can begin and end your line plot. To complete your line plot, put an X over each value in the line plot that is found in the data.

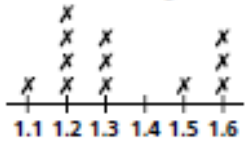
c. The least value is 112, and the greatest value is 172. These values give us the endpoints with which you can begin and end your bar graph. To complete your bar graph, make a bar over each value in the bar graph whose height aligns with the frequency of those values.

d. By making a stem-and-leaf plot, you do not have to draw a bar for each height, so the stem-and-leaf plot is more compact.

11. a. (Figure 3)

b.

Student Heights



c. Some students may say that the median of 1.3 meters is typical.

12. a. 8 hours

b. 75%; according to the pie chart, Harold spends about 2 hours on math, 2 hours on science, and 2 hours on history. So he spends 2 + 2 + 2, or 6 out of 8 hours on those subjects altogether. $\frac{6}{8} = 75\%$

Figure 3

Student Heights and Foot Lengths

Age (mo)	Height (cm)	Height (m)	Rounded Height (tenth of a meter)	Foot Length (cm)	Foot Length (m)
76	126	1.26	1.3	24	0.24
73	117	1.17	1.2	24	0.24
68	112	1.12	1.1	17	0.17
78	123	1.23	1.2	22	0.22
81	117	1.17	1.2	20	0.20
82	122	1.22	1.2	23	0.23
80	130	1.30	1.3	22	0.22
90	127	1.27	1.3	21	0.21
138	161	1.61	1.6	28	0.28
152	156	1.56	1.6	30	0.30
149	157	1.57	1.6	27	0.27
132	150	1.50	1.5	25	0.25

Extensions

13. The graph below (Figure 4) uses G and B in place of actual numbers of jumps (refer to earlier stem plot for numbers), giving girls' and boys' data on the same plot. Students will have made separate plots, but you may use this summary graph as a way to show how we can modify stem plots to give different information.

    Generally speaking, girls performed slightly better in Mr. Costo's class than the girls in Mrs. Reid's class. It seems as though Mr. Costo's class has three outliers for girls. One girl in Mr. Costo's class jumped rope more times than anyone else in either class. The boys performed similarly in both classes. We can see an outlier for boys in each class.

14. One way to answer this question is to show all the girls' data on one side of the stem plot and all the boys' data on the other side, as shown below. (Figure 5)

    The boys' data clusters at the lower end of the stem plot. The girls' data is spread out with more of the data showing larger numbers of jumps. So the girls did better.

Figure 4

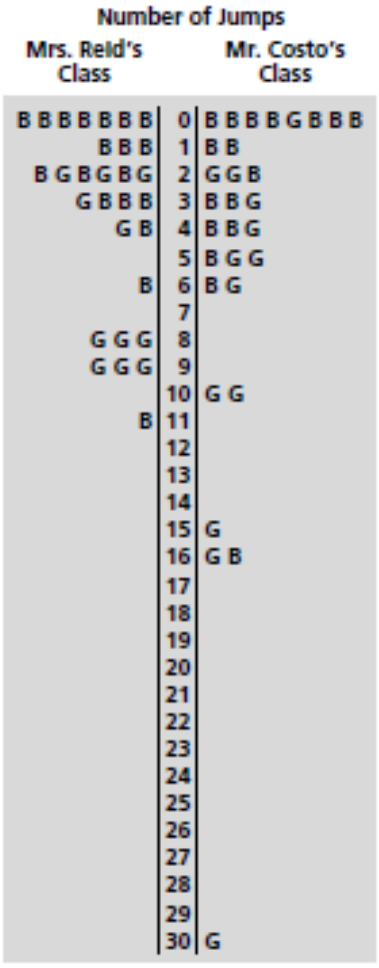
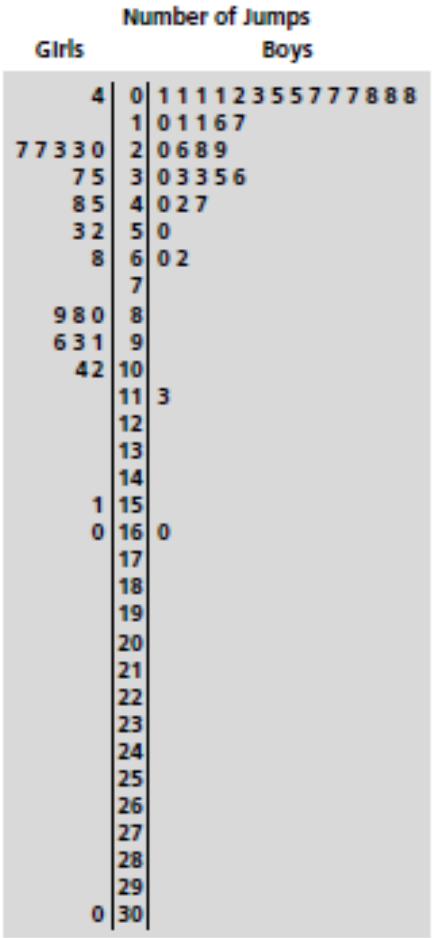


Figure 5



15. **a.** The actual counts vary from 309 to 607 seeds, so the range is 298 seeds. The graph shows that the actual counts fall within a smaller range compared to the guesses. The median is $458\frac{1}{2}$ seeds (halfway between 455 and 462).

b. The guesses vary from 200 to 2,000 seeds, so the range is 1,800 seeds. The graph shows that the guesses are much more spread out than the actual counts. The median is $642\frac{1}{2}$ seeds (halfway between 630 and 655).

c. (Figure 6)

d. Points on or near the line represent guesses that are very close or equal to the actual counts.

e. Points above the line represent guesses that are larger than the actual counts.

f. Points below the line represent guesses that are smaller than the actual counts.

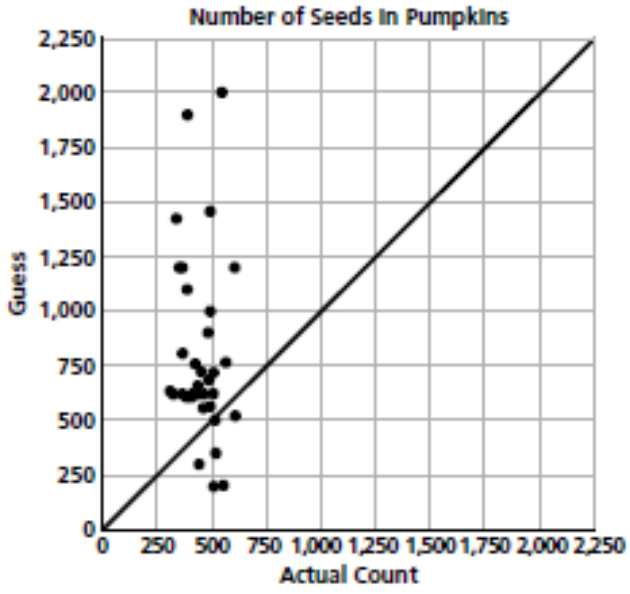
g. In general, the guesses are larger than the actual counts. The median for the guesses is $642\frac{1}{2}$ with a range of 1,800 seeds. The median of the actual counts is $458\frac{1}{2}$ with a range of 298 seeds. The median for the actual counts is much smaller than the median for the guesses. The range for the guesses spans 1,800 ÷ 298, or about 6 times as many values.

h. Possible answer: You could change the scale on the horizontal axis to go from 0 to 750 scaling by 100.

Possible Answers to Mathematical Reflections

1. These can easily be found on the stem plot. Find the median by first determining the number of data values in the data set and then finding half of that number. For example, if there are 46 data values, the median lies between the twenty-third and twenty-fourth values. If there are 45 data values, the median is the twenty-third value. As long as you count consecutively from the least value or the greatest value, you may count from either end of the data displayed in the stem plot to locate the median. Find the range by arranging all leaves in ascending order and then finding the difference between the greatest and least values in the data set.

Figure 6



2. To place a point, start at $(0, 0)$ and move to the right, along the horizontal axis (x-axis), the number of units given by the first coordinate. Then move up, along the vertical axis (y-axis), the number of units given by the second coordinate.
3. You assign the first measure for each pair to the x-axis, and the second measure to the y-axis. Then, you consider the spread of the data as you set up the scale of each axis.
4. A stem-and-leaf plot is more useful than a line plot or bar graph for data that are spread out. For this type of data, grouping by intervals allows us to see patterns in the data. Line plots are quickly constructed graphs that can be used when you want to "sketch" a data set. If there is a great number of data items, the bar graph is a more useful tool because its vertical scale is adjustable.

Investigation 3

ACE Assignment Choices

Problem 3.1

Core 1–4
Other *Connections* 7, 8, 19

Problem 3.2

Core 5, 10
Other *Applications* 6; *Extensions* 20, 21; unassigned choices from previous problems

Problem 3.3

Core 11, 17, 23
Other *Connections* 9, 12–16, 18; *Extensions* 22; unassigned choices from previous problems

Adapted For suggestions about adapting Exercise 6 and other ACE exercises, see the CMP *Special Needs Handbook*.
Connecting to Prior Units 7, 8, 10, 12–16: *Bits and Pieces I*

Applications

1. **a.** 3; order the data from least to greatest. The median is the value that separates the data in half.

   b. Yes, six households have 3 children. The median is located using the data values. The only time the median will not be one of the data values is when it is determined by finding the mean of two middle values that are not the same.

2. **a.** 4; you can add the data values together and divide by the number of data values to get the mean. Or, you can find the mean by making stacks of cubes for each of the households and then evening out the stacks so there are 16 households, each with 4 members. The mean tells you the value that each data item would have if all the data had the same value.

   b. There are no squares over the number 4 on the line plot, which means there are no households in the data set with four children. This is possible because there are households with more than four children and households with less than four children to balance each other.

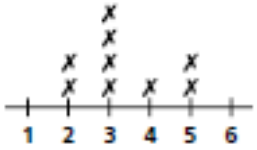
3. C

4. **a.** Possible line plot:

   b. Possible line plot:

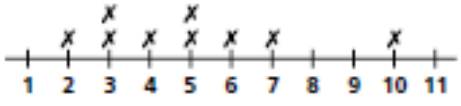
   c. Answers will vary.

5. Possible line plot:



   For nine households to have a mean of $3\frac{1}{3}$ people, there would have to be a total of $9 \times 3\frac{1}{3}$, or 30 people.

6. Possible answer:



Connections

7. $\frac{3}{4}$ hour. One way students can think of this problem is by using blocks that each represent one fourth, then making towers that correspond to the data values. For example, $\frac{3}{4}$ would be represented by a tower of three blocks, $\frac{1}{2}$ by a tower of two blocks, and so on. By distributing the blocks evenly, students can see that all the towers will have three blocks, which represents a mean of $\frac{3}{4}$.
8. G
9. **a.** 32 oz per player;
   5,760 oz ÷ (18 × 10) players = 32 oz
   b. The mean, because it represents the total amount of water evenly shared among the 180 players.
10. The typical price of a box of granola bars is $2.66, and there are nine different brands of granola. So the total cost of nine boxes (one of each brand) is $23.94. You have to price the boxes so the total cost is $23.94. You could have the nine brands all priced at $2.66, or have just a few priced at $2.66, or have no brands priced at $2.66. Here is one possibility: $2.70, $2.78, $2.98, $2.34, $2.58, $2.70, $2.50, $2.58, $2.78
11. **a.** The mean tells Ralph that if all the rabbits in the data lived to be the same age, that age would be 7 years. What actually happens is that some of the rabbits don't live to 7 years and some of the rabbits live beyond 7 years.
    b. Knowing the spread would give Ralph more information about the possible life span of his rabbit.
12. **a.** Sabrina and Diego danced $3\frac{3}{4}$ hours and Marcus danced $2\frac{1}{4}$ hours.
    b. The mean is less than the median. The median is $3\frac{3}{4}$ hours, and mean is less than $3\frac{3}{4}$ hours because $2\frac{1}{4}$ hours decreases the amount of hours each person danced.
13. No, some children may have watched videos for 39 minutes, but most children spent less or more time watching videos.
14. About 67%
15. About $\frac{3}{4}$ or 0.75 of an hour
16. $2\frac{1}{3}$ hours or about 2.3 hours
17. **a.** Mayor Phillips determined the mean income. The total income is $32,000; dividing by the number of incomes, 16, gives $2,000 per week. Lily Jackson found the median income. There are a total of 16 values, so the median is between the eight and ninth values. The eighth value is $0 and the ninth value is $200, so the median is $100. Ronnie Ruis looked at the mode, which is $0. Each of their computations is correct.
    b. No; no one earns $2,000 per week.
    c. No; no one earns $100 per week.
    d. Yes; eight people earn $0 per week.
    e. $200 is a good answer. Possible explanation: The people who have $0 incomes are probably children, so the people who earn $200 and the person who earns $30,600 are the residents who are employed. The "typical" income is either the median or the mode since the mean is greatly affected by the one large income.
    f. The mode is $200. The median is $200. The mean is $1,640.
18. **a.** Possible answer: The data are skewed to the lower values.

Movies Watched

0	1 1 1 2 2 2 3 3 4 4 4 5 8 8 9
1	0 1 2 3 5 5 7
2	0 5
3	0

Key: 1 | 5 means 15 movies

    b. 9 movies; add to find the total number of the movies watched (225). Then divide the total by the number of students (25).
    c. The mean is 9 movies, and the data vary from 1 to 30 movies. Since the mean is closer to the low end of the data values, more students fall in the low end of the data values.
    d. The median number of movies watched is 8. The mean is greater than the median because the large values pull the mean up, but have less influence on the median.

19. a. A, B, and D because they are divisible by 6.

b. 2 pens ($12 \div 6 = 2$), 3 pens ($18 \div 6 = 3$), or 8 pens ($48 \div 6 = 8$)

c. I agree because an average can be found by sharing the total amount of pens evenly among all students.

Extensions

20. Answers will vary. Pay attention to the students' reasoning. Generally, data reported in newspapers use the mean.

21. There are 365 days in a year. This means the average third grader watches $1{,}170 \div 365$, or about $3\frac{1}{5}$ hours of television per day.

22. a. Mrs. Reid's class:

mean: $\approx 38\frac{1}{2}$ ($1{,}157 \div 30$); median: 28

Mr. Costo's class:

mean: $54\frac{2}{5}$ ($1{,}632 \div 30$); median: 34

The mean is greater than the median for each class, because there are some greater values in each set of data.

b. They should use the median, because it is greater than the mode.

c. The median decreases by 1 to 33 jumps. The median is the middle data value, so it is not changed much by removing the greatest data value.

d. The mean is now ($1{,}332 \div 29$) or $45\frac{27}{29}$ or 45.931. The mean decreases by a little more than 8 jumps. It decreases because the greatest value was removed, which had a big influence on the mean.

e. Mrs. Reid's class's mean and median are still less than each of the same statistics for Mr. Costo's class, so Mrs. Reid's class cannot make a valid claim that they did better.

23. a. 14

b. I. Yes, 1 is an outlier.

II. 12.375

III. The new mean is lower. Possible explanation: By adding the 1 to the data, the mean decreases because, like with the cube stacks in Problem 4.1, cubes need to be added to the stack of one, and this would decrease the heights of the other stacks.

Possible Answers to Mathematical Reflections

1. Possible method: Add together all the values. Divide the sum by the number of values. This method works because the sum of the values tells us how much is to be shared or "evened out." The number of values is the number of parts into which the total must be divided. Division gives the number in each part.

2. a. They are measures of center because they are good indicators of a typical value.

b. The mode is the data value that occurs most frequently. The median is the middle value that separates an ordered set of data in half. The mean is the "balance point," or the value that each item would have if all the data had the same value.

c. The median is not greatly affected by outliers in the data.

3. The least and greatest values give the upper and lower boundaries of the data set, while the range gives the distance between these two values. A measure of center gives a typical value in the data.

4. a. The student is correct. Finding the mode just involves counting the most frequently occurring data value within a data set, so it can be used with both numerical and categorical data. Finding the mean requires dividing a sum by the number of values, and finding the median requires ordering data values from least to greatest. Both cannot be done with categorical data.

b. No; finding range depends on being able to identify the least and greatest values. You cannot order categorical data in a logical way.

Mathematics of the Unit

Pearson Prentice Hall Professional Development

Overview

Exploring statistics as a process of data investigation involves a set of four interrelated components (Graham, 1987):

- Posing the question: formulating the key question(s) to explore and deciding what data to collect to address the question(s)
- Collecting the data: deciding how to collect the data as well as actually collecting it
- Analyzing the data: organizing, representing, summarizing, and describing the data and looking for patterns in the data
- Interpreting the results: predicting, comparing, and identifying relationships and using the results from the analyses to make decisions about the original question(s)

This dynamic process often involves moving back and forth among the four interconnected components. For example, collecting the data and, after some analysis, deciding to refine the question and gather additional data. It may involve spending time working within a single component. For example, creating several different representations of the data, some in earlier stages of the process and others at a later time, before selecting the representation(s) to be used for final presentation of the data.

In many of the problems, data are provided. We assume students have had experience collecting data as part of statistical investigations. If they have not, we encourage you to have your class collect their own data for some of the problems. The problems can be applied either to the data provided or to data collected by students.

Even if your students have already had experience collecting data, they may be interested in investigating data about their class. Students will feel empowered if they have the opportunity to use the process of data investigation to explore questions that are of interest to them. Keep in mind that collecting data is time-consuming, so carefully choose the problems for which you will have students generate data.

Summary of Investigations

Investigation 1

Looking at Data

This first investigation develops some introductory statistical techniques that will be used throughout *Data About Us*. It focuses on describing, interpreting, and comparing distributions. A discussion about the origin of names is used, providing an opportunity to integrate social studies. In addition, students consider lengths of names, and compare distributions of lengths of names from two data sets that are provided and their class's data. Students are introduced to or review the use of tables, line plots, and bar graphs to represent data; ways to describe the shape of a distribution; and the use of measures of center (the mode and median), spread, and range to characterize a distribution.

Students are also introduced to types of data, with a focus on categorical and numerical data. They consider two tables and graphs of data that relate to two questions, one that involves numerical data and one that involves categorical data. Finally, they experiment with using and making horizontal and vertical bar graphs.

Investigation 2

Using Graphs to Explore Data

This investigation first focuses on developing strategies for grouping and displaying data in intervals using stem-and-leaf plots. Data that are collected are often quite spread out or have a great deal of variability. A line plot or bar graph may not be very useful for displaying such data in order to see patterns in the distributions (e.g., clusters, gaps). Students need strategies for grouping and displaying data in equal intervals. The stem-and-leaf plot (or stem plot) is a useful tool for grouping data in intervals of 10, and it helps students see patterns in the data. Students use a stem-and-leaf plot to examine two given data sets. The first data set is about time and distance required for students in a particular class to travel to school. The second data set is about how many times each student in two different classes jumped rope without stopping.

Students then use coordinate graphs to display pairs of data. They begin by collecting data about the lengths of their arm spans and their heights. Using these data, they make a coordinate graph and sketch the $y = x$ line so they can discuss people who are above, on, or below the line and what this means in terms of the relationship between arm span and height (that is, are most people's arm span and height similar?). They return to the travel time and distance data set and look at a coordinate graph that shows a student's travel time paired with distance traveled in order to discuss whether there is a relationship between travel time and distance traveled (that is, does it take someone who travels farther more time to get to school?).

Investigation 3

What Do We Mean by *Mean*?

This investigation focuses on developing the concept of mean. The "average" number of people in the families of students in a class provides the setting. The notion of "evening out" or "balancing" the distribution at a point (the mean) located on the horizontal axis is modeled by using cubes and stick-on notes. These models support development of the algorithm for finding the mean: adding up all the numbers and dividing by the number of items.

Mathematics Background

In *Data About Us*, several big ideas about statistics are explored. On the next page is a concept map that provides some insights into the overall relationships among these and other important concepts. The shaded portions of the diagram and highlighted graph names are central statistical ideas that are emphasized in *Data About Us*.

Different Types of Data

Questions in real life often result in answers that involve one of two general kinds of data: categorical data or numerical data. Knowing the type of data helps us to determine the most appropriate measures of center and displays to use for the data.

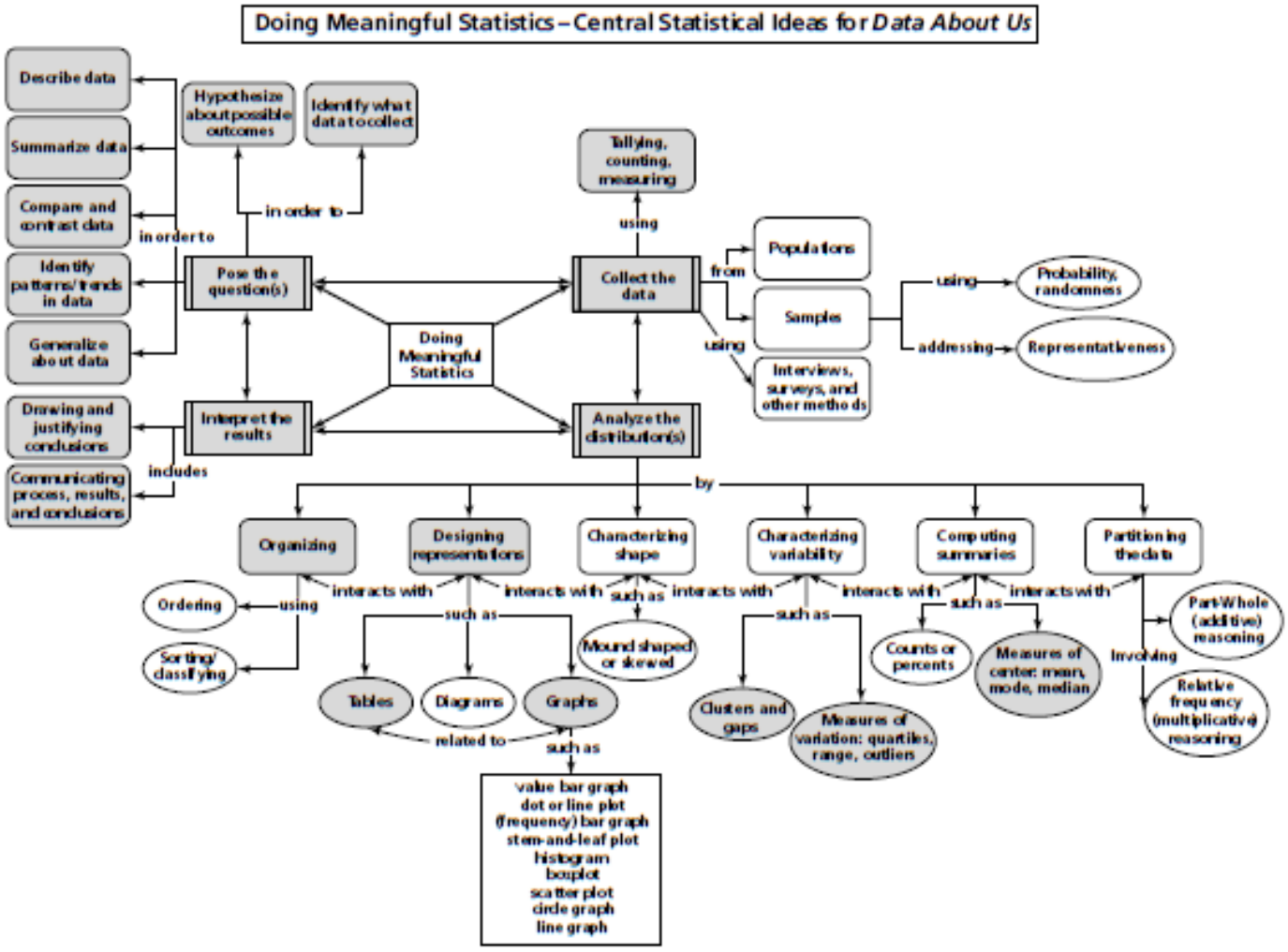
Numerical Data

- We can collect data about family size and organize them by using frequencies of how many families have zero children, one child, two children, and so on.
- We can collect data about pulse rates and organize them using intervals by using frequencies of how many people have pulse rates in the intervals of 60 to 69 beats, 70 to 79 beats, and so on.
- We can collect data about height and organize them into intervals by using frequencies of how many people are from 40 to 44 inches tall, 45 to 49 inches tall, and so on.
- We can collect data about time spent sleeping in one day and organize them by frequencies of how many people slept 7 hours, $7\frac{1}{2}$ hours, 8 hours, and so on.
- We can collect data about responses to a question such as, "On a scale of 1 to 5 with 1 as 'low interest,' rate your interest in participating in the school's field day" and organize them by using frequencies of how many people indicated each of the ratings 1, 2, 3, 4, or 5.
- We can use the mean, median, mode, and range as summary statistics on any numerical data.

Categorical Data

- We can collect data about birth years and organize them by using frequencies of how many people were born in 1980, 1981, 1982, and so on.
- We can collect data about favorite type of book to read and organize them by using frequencies of how many people like mysteries, adventure stories, science fiction, and so on.
- We can collect data about hobbies and organize them by using frequencies of how many people collect stamps, build models, knit, and so on.
- Mode is the only summary statistic we can use on categorical data.

At times, categorical data seem to be organized like numerical data. A bar graph of birth months may employ numbers to represent months. For example, 1 is used for January, 2 is used for February, and 3 is used for March. However, we cannot perform numerical operations using months of the year, because months represented numerically are actually categories with a number label representing the category.


Doing Meaningful Statistics – Central Statistical Ideas for *Data About Us*
Doing Meaningful Statistics
Describe data
Summarize data
Compare and contrast data
Identify patterns/trends in data
Generalize about data
in order to
Hypothesize about possible outcomes
Identify what data to collect
in order to
Pose the question(s)
Tallying, counting, measuring
using
Collect the data
from
Populations
Samples
using
Interviews, surveys, and other methods
using
Probability, randomness
addressing
Representativeness
Drawing and justifying conclusions
Communicating process, results, and conclusions
includes
Interpret the results
Analyze the distribution(s)
by
Organizing
Designing representations
Characterizing shape
Characterizing variability
Computing summaries
Partitioning the data
Ordering
using
Sorting/classifying
interacts with
interacts with
such as
Mound shaped or skewed
such as
Tables
Diagrams
Graphs
related to
such as
value bar graph
dot or line plot
(frequency) bar graph
stem-and-leaf plot
histogram
boxplot
scatter plot
circle graph
line graph
interacts with
interacts with
such as
Clusters and gaps
Measures of variation: quartiles, range, outliers
interacts with
interacts with
such as
Counts or percents
Measures of center: mean, mode, median
interacts with
Involving
Part/Whole (additive) reasoning
Relative frequency (multiplicative) reasoning

Distribution

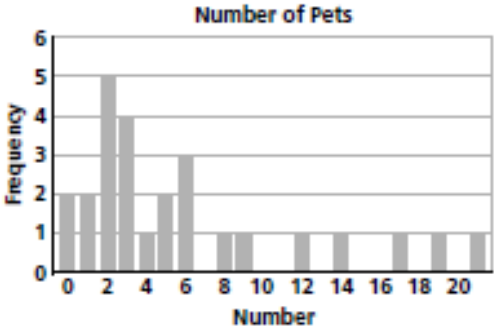
The distribution of data refers to the way data occur in a data set. We often use graphs to help us see how data are distributed. A distribution (data as a whole versus individual data values) has characteristics that can be described using statistics such as measures of center or range.

When students work with data, they are often interested in the individual cases, particularly if the data are about themselves. However, statisticians like to look at the overall distribution of a data set and are not interested in individual cases.

We use graphs to help clarify a distribution of data. Distributions (unlike individual cases) have properties such as measures of central tendency (i.e., mean, median, mode) or variation (e.g., outliers, range) and shape (e.g., clusters, gaps).

There appear to be several general ways students think about data:

- Students focus on each data value. For example, they may focus on individual name lengths. They may not see that a group of cases may be related (e.g., several name lengths cluster around lengths of 8 to 10 letters). This kind of thinking is more characteristic of young children. However, when looking at outliers, a focus on individual data values is necessary. How might we explain a name length of 1,019 letters if this data value was part of the data set?
- Students focus on subsets of data values that may be the same or similar like a category or a cluster. This is easier for students when using categorical data (e.g., more students chose dogs as their favorite kind of pet). If students are using numerical data, they might notice clusters (e.g., the number of pets students have at home in the interval of 2 to 3).
- Students view all the data values as an "object" or distribution (see graph of number of pets below). Students look for features of the distribution that are not features of any of the individual data values (e.g., shape, range, clusters). In looking at the distribution of the number of pets students have, we can see that data are clustered at one end with a kind of tail going off to the right that accounts for several cases in which students have more than six pets.



Data Reduction

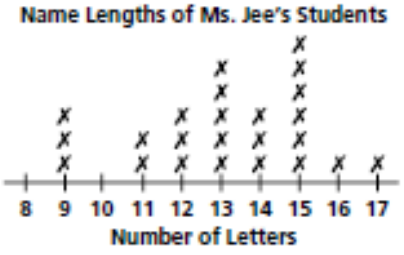
Statisticians use the term "data reduction" to describe what they do when they use representations or statistics during the analysis part of the process of statistical investigation.

Standard Graphs

Representations in the K–12 curriculum that are addressed in *Data About Us* include the following:

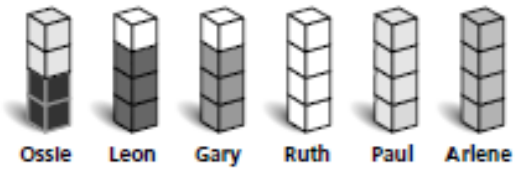
Line Plot

Each case is represented as an "X" positioned over a labeled number line.



Frequency Bar Graph

A bar's height is not the value of an individual case but rather the number (frequency) of cases that all have that value. (See Number of Pets graph above.)



After:

Ossie	4 people
Leon	4 people
Gary	4 people
Ruth	4 people
Paul	4 people
Arlene	4 people
Total	24 people

Mode is the value that occurs with greatest frequency in a set of data.

Median is the value that marks the location that separates an ordered set of data into two equal-sized groups, with the same number of values before the median and after the median.

Although there is one median in a set of data, there may be more than one mode.

In a data set with an even number of values, where the two middle values differ by more than one, the median is the midpoint between these values. For example, the median for the data set 3, 4, 4, 7, 8, 9 is $5\frac{1}{2}$, the number that is the midpoint between 4 and 7. If the two middle values are the same number, the median is the value of that number. For example, for the data set 3, 4, 5, 5, 7, 8, the median falls between the two 5's, so the median is 5.

Measures of Variation

Measures of variation establish the degree of variability or scatter of the individual data values and their deviations from (or differences from) the measures of center. In *Data About Us,* students use *range* as one measure of variation. Range is the difference between the least and the greatest data values. In addition, students are encouraged to talk about where data cluster and where there are "holes" in the data as further ways to comment about variation.

Covariation

Covariation is a way of characterizing a kind of relationship between two variables. It means that information about values from one variable helps us to understand and explain or predict values of the other variable. In *Data About Us,* students are asked to think about whether changes in one variable (e.g., time traveled to school) might be related to changes in another variable (e.g., distance traveled to school). The primary goal for this unit is to review using a coordinate graph to represent data. More formal work with covariation continues in the other data units.