DATA ANALYSIS, STATISTICS, AND PROBABILITY

27 DATA DISPLAYS: TABLES AND BAR GRAPHS

10.D.1 Select, create, and interpret an appropriate graphical representation (e.g., scatterplot, table, stem-and-leaf plots, box-and-whisker plots, circle graph, line graph, and line plot) for a set of data.

KEY CONCEPTS YOU'LL NEED IN THIS LESSON

1. A **frequency table** (or interval table) must have **intervals** of equal size.
2. The intervals must account for all the data. Both the greatest and least values must be included in the table.
3. The intervals must not overlap, so that each score can be placed in only one interval.
4. A reasonable number of intervals, from approximately 5 to approximately 10, allows data to be grouped so that patterns can be seen.
5. A **frequency histogram** is a bar graph in which data from a frequency table is displayed. The bars are drawn next to one another to indicate that as one interval ends, the next begins.

1. FREQUENCY TABLES

11.8

EXAMPLE 1

The data below gives the amounts of money, in dollars, that 32 students estimate they spend during an average week.

10, 17, 20, 15, 22, 30, 25, 10,
20, 15, 24, 18, 21, 22, 25, 12,
13, 18, 20, 20, 14, 17, 25, 30,
6, 10, 16, 20, 14, 19, 28, 22

a. Complete a frequency table.
b. How many students estimate that they spend more than $20 per week?
c. In which interval does the greatest number of students fall?

STRATEGY: Use Key Concepts 1–4.

STEP 1: Choose an appropriate interval for the data.

An interval of $10 would give just 3 intervals: $1–$10, $11–$20, and $21–$30, not enough to get a good sense of the data.

An interval of $5 would give 6 intervals, a better number for the data.

STEP 2: Copy and complete a frequency table, using tallies.

Interval	Tallies	Frequency (Number)
1–5		0
6–10	////	4
11–15	~~////~~ /	6
16–20	~~////~~ ~~////~~ /	11
21–25	~~////~~ ///	8
26–30	///	3

STEP 3: Use the data in the table to find the number of students who estimate that they spend more than $20 per week.

The last two rows represent those students.
8 + 3 = 11, so 11 students estimate that they spend more than $20.

STEP 4: Read the table to find the interval in which the greatest number of students falls.

The interval $16 – $20 contains the greatest number of students.

2. FREQUENCY HISTOGRAMS

EXAMPLE 2

Display the data from Example 1 in a frequency histogram.

STRATEGY: Use Key Concept 5.

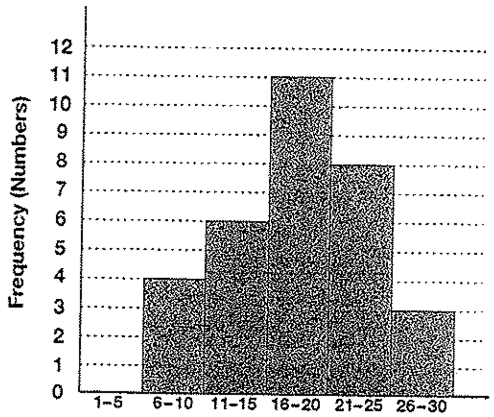
11.7

STEP 1: Give titles and values to the axes.

The horizontal axis represents the amounts of money (the intervals), each interval representing $5.

The vertical axis represents the number of students (the frequencies), each line representing 1 student.

Amounts Spent By Students



STEP 2: Draw the bars. Note that only the first interval, \$1 – \$5, has no bar. This is so because no students estimated within that range. All other bars are touching.

28 MEASURES OF CENTRAL TENDENCY

10.D.1 Use appropriate statistics (e.g., mean, median, range, and mode) to communicate information about the data. Use these notions to compare different sets of data.

KEY CONCEPTS YOU'LL NEED IN THIS LESSON

1. The **mean** is the average when the total of the scores is divided by the number of scores.
2. The **median** is the middle score when the scores are arranged numerically from least to greatest.
3. The **mode** is the score or scores that occur the greatest number of times. A set of data may have no mode.
4. Mean, median, and mode can be found for data in a frequency table.
   a. If the interval in the table is one number, all three measures can be found.
   b. If the interval in the table is greater than one number (a range of numbers), you can find the interval containing the mode and the median.

1. THE MEAN

11.8

EXAMPLE 1

Kim made telephone calls of the following number of minutes to a friend in New Jersey: 5, 8, 12, 16, 7, 10, and 8. What was the mean number of minutes her calls lasted, rounded to the nearest minute?

STRATEGY: Use Key Concept 1.

STEP 1: Add to find the sum of the scores (minutes).

$5 + 8 + 12 + 16 + 7 + 10 + 8 = 66$

STEP 2: Divide the sum by the number of scores (telephone calls).

$66 \div 7 \approx 9.43$

SOLUTION: Rounded to the nearest minute, Kim's calls lasted an average of 9 minutes.

EXAMPLE 2

Andrew's average on 4 tests is exactly 90. He received a grade of 90 on one test, 82 on another, and the same grade on each of the other two tests. What grade did he receive on each of the other two tests?

STRATEGY: **Use Key Concept 1, and an algebraic procedure.**

STEP 1: Assign a variable to the unknown test score.

Let x = the score Andrew received on each of two tests.

STEP 2: Set up and solve an equation that describes the data in the problem.

$$\frac{90 + 82 + \text{the same score on two tests}}{4} = 90$$

$$\frac{90 + 82 + 2 \cdot x}{4} = 90$$

$$\frac{172 + 2x}{4} = 90$$

$$172 + 2x = 360$$
$$2x = 188$$
$$x = 94$$

SOLUTION: **Andrew received a score of 94 on each of his other two tests.**

2. THE MEDIAN

EXAMPLE 3

The prices of 6 CDs rounded to the nearest dollar are:
$12, $15, $14, $17, $12, and $11.

What is the median price paid for the CDs?

STRATEGY: **Use Key Concept 2.**

STEP 1: Arrange the scores (prices) in order from least to greatest:

$11, $12, $12, $14, $15, $17

Since there are 6 scores, an even number, there is no middle score.

STEP 2: Find the average or mean of the two middle scores.

$11, $12, $12, $14, $15, $17

$$\frac{\$12 + \$14}{2} = \frac{\$26}{2} = \$13$$

SOLUTION: **The median price paid for the 6 CDs is $13.**

3. THE MODE

EXAMPLE 4

For the CDs described in Example 3, what is the mode price paid?

STRATEGY: **Use Key Concept 3.**

STEP 1: Examine the data for the most frequently occurring score (price).

$12 occurs more often than any other price.

SOLUTION: **The mode price paid for the CDs is $12.**

4. DATA IN A FREQUENCY TABLE AND MEASURES OF CENTRAL TENDENCY

EXAMPLE 5

The data in the table gives the number of blocks from school 14 students live.

Interval (Number of Blocks)	Frequency (Number of Students)
1	2
2	1
3	0
4	3
5	4
6	1
7	3

Find the mean, median, and mode distances.

STRATEGY: **Use Key Concept 4a, noting that in the table each interval represents a single number.**

Find the mean number of blocks students live from school.

STEP 1: In each row, multiply the interval (number of blocks) by the frequency (number of students) to find how many blocks are walked in all by students in that row.

STEP 2: Add the totals in the last column to find the total number of blocks walked by all students.

STEP 3: Divide that total by 14, the number of students.

Interval (blocks)	Frequency (Students)	Interval • Frequency
1	2	1 • 2 = 2
2	1	2 • 1 = 2
3	0	3 • 0 = 0
4	3	4 • 3 = 12
5	4	5 • 4 = 20
6	1	6 • 1 = 6
7	3	7 • 3 = 21

Total Frequency = 14   Total Blocks = 63

67 ÷ 14 = 4.5

SOLUTION: The average number of blocks walked by the 14 students is 4.5.

Find the median number of blocks.

Interval (blocks)	Frequency (Students)
1	2
2	1
3	0
4	3
5	4
6	1
7	3

STEP 1: Find the total frequency: the total frequency is 14.

STEP 2: Use the total frequency, 14 to determine which score is the median.

The median score is the middle when the scores are in numerical order. Since the scores are in numerical order in the table and there are an even number of students, the middle score is the average of the 7th and 8th scores.

STEP 3: Find the interval that contains the 7th and 8th scores.

Count from the top: 2 + 1 + 0 + 3 = 6, but 2 + 1 + 0 + 3 + 4 = 10, past the 8th score. So, the 7th and 8th scores must be in the interval of 5 blocks. There is no need to find the average of the scores because they are both 5.

SOLUTION: The median number of blocks walked is 5.

Find the mode number of blocks.

STEP: Find the interval that contains the greatest frequency.

The interval with the greatest frequency is 5, the modal interval.

SOLUTION: The mode number of blocks walked is 5.

EXAMPLE 6

The data in the table represents the number of minutes of math homework done one night by students in a class.

Interval (Number of Minutes)	Frequency (Number of Students)
10–19	3
20–29	5
30–39	8
40–49	6
50–59	7

Find the interval with the mode and the interval with the median.

STRATEGY: Use Key Concept 4b, noting that each interval is not a single number, but a range.

STEP 1: Find the interval containing the mode.

Find the interval with the greatest frequency. Since more students said they did between 30 and 39 minutes of math homework than any other length of time, the modal interval is 30–39.

STEP 2: Find the interval with the median.

Since there are 29 students, the median interval is the interval with the 15th student (14 students above and 14 students below).

Count the frequencies from the top of the table: 3 + 5 = 8, but 3 + 5 + 8 = 16. So the median must be somewhere in the interval 30–39.

SOLUTION: For the grouped data shown in the table, the modal interval and the interval with the median are both 30–39.

(Note that if the intervals are a range, it is impossible to tell which actual score is the mode or median. For example, every student in the range 30–39 may have done 31 minutes of homework, or the numbers may be spread throughout the range.)

29 DATA DISPLAYS: STEM-AND-LEAF PLOTS, BOX-AND-WHISKER PLOTS

10.D.1 Select, create, and interpret an appropriate graphical representation (e.g., scatterplot, table, stem-and-leaf plots, box-and-whisker plots, circle graph, line graph, and line plot) for a set of data and use appropriate statistics (e.g., mean, median, range, and mode) to communicate information about the data. Use these notions to compare different sets of data.

KEY CONCEPTS YOU'LL NEED IN THIS LESSON

STEM-AND-LEAF PLOTS (11.7)

1. A **stem-and-leaf plot** is similar to a bar graph. It is useful when you want to see the values of all the pieces of data that make up each bar. Stem-and-leaf plots are generally most valuable where all pieces of data have the same number of digits.

2. The digits in the greatest place value are used as the **stems**. The digits with the next greatest place value form the **leaves**.

In presidential elections, the ten states with the greatest populations in the year 2000 had the following numbers of electoral votes: 32, 22, 54, 33, 21, 18, 25, 15, 23, and 14. For the electoral votes of these states, a stem-and-leaf plot could be constructed as follows:

```
5 | 4
3 | 2 3
2 | 1 2 3 5
1 | 4 5 8        Key: Read 2 | 5 as 25
```

From the plot, it is possible to read the following:
- The state with the greatest population had 54 (5 | 4) electoral votes.
- There were 7 states with at least 20 electoral votes.
- There were 3 states with at least 30 electoral votes.
- There were no states with more than 33 but fewer than 54 electoral votes.

BOX-AND-WHISKER PLOTS (11.8)

3. **Box-and-whisker plots** show the following for a set of data:
   - median
   - range
   - upper and lower quartiles

4. • The **median** is the middle score when the scores are arranged in order from least to greatest. Fifty percent of the data points will be above the median, and 50% will be below the median.
   - The **range** is the difference between the greatest and least scores.
   - The **lower quartile (LQ)** is the median of the lower half of the data. For any set of data, 25% of the data will be below the lower quartile.
   - The **upper quartile (UQ)** is the median of the upper half of the data. For any set of data, 25% of the data will be above the upper quartile.

KEY CONCEPTS YOU'LL NEED IN THIS LESSON

The following data represent high temperatures in 12 American cities on August 25, 2001 in degrees Fahrenheit.

78, 89, 63, 99, 87, 90, 85, 98, 83, 104, 81, 100

The range is the difference between the greatest value (104) and the least (63):

$$104 - 63 = 41$$

The median is the middle score when the scores are arranged in order from least to greatest.

63, 78, 81, 83, 85, 87, 89, 90, 98, 99, 100, 104

Find the average of the two middle scores since 12 is an even number of scores.

$$\frac{87 + 89}{2} = 88$$

The lower quartile is the median of the lower half of the data.

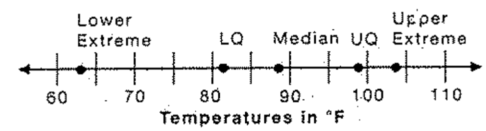
$$LQ = 63, 78, 81, 83, 85, 87 \Rightarrow \frac{81 + 83}{2} = 82$$

The upper quartile is the median of the upper half of the data.

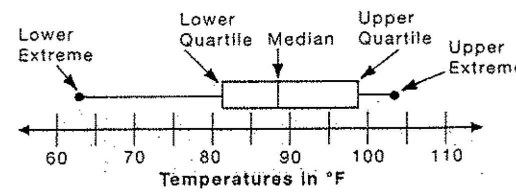
$$UP = 89, 90, 98, 99, 100, 104 \Rightarrow \frac{98 + 99}{2} = 98.5$$

Construct a box-and-whisker plot for the data as follows:

Draw a number line for the range of data. Mark the range values, the median, and the quartile values.



Complete the box-and-whisker plot.



The box connects the lower and upper quartiles. The median is the vertical line through the box. The whiskers extend out to the lower and upper extremes.

1. STEM-AND-LEAF PLOTS

EXAMPLE 1

The stem-and-leaf plot contains the selling prices in dollars of 20 different textbooks on sale in a college bookstore. Use the plot to answer the questions below it.

```
7 | 4
6 | 1 4 4 6
5 | 0 9
4 | 0 1 3 4 7
3 | 0 0 0 8 9 9
2 | 3 7
```

a. What was the range of prices of the books?
b. What was the median price?
c. How many books cost at least $50?
d. Was there a mode price? If so, what was it?
e. Explain how you would find the mean price of the books.

STRATEGY: Use Key Concepts 1 and 2.

STEP 1: To find the range, find the greatest and least values. Then find the difference between them.

Greatest value = $74 Least value = $23

74 − 23 = 51

STEP 2: Twenty is an even number. So, to find the median price, find the average of the two middle prices, the 10th and 11th scores.

41 + 43 = 84 84 ÷ 2 = 42

STEP 3: To find the number of books that cost at least $50, count the entries in the plot, beginning with 50.

50, 59, 61, 64, 64, 66, and 74 all represent prices of at least $50.

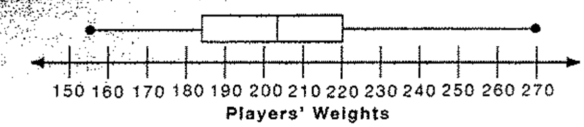
STEP 4: There were 3 books that cost $30 each. This value appears more than any other.

SOLUTION:
a. **The range of prices was $51.**
b. **The median price was $42.**
c. **There were 7 books that cost at least $50.**
d. **The mode price was $30.**
e. **To find the mean price, you would add the prices of all 20 books and then divide by 20, the number of books.**

2. BOX-AND-WHISKER PLOTS

EXAMPLE 2

The box-and-whisker plot contains data about the weights of the players on a men's high school football team. Use the plot to answer the questions below it.



a. What is the approximate median of the weights?
b. What is the approximate range of weights?
c. Estimate the upper and lower quartile weights.
d. What percent of the players weigh more than 220 pounds?

STRATEGY: Use Key Concepts 3 and 4.

STEP 1: Estimate the median. The vertical line through the box represents the median value. A good estimate for the median is approximately 203 pounds.

STEP 2: To find the range, identify the greatest and least values and find their difference.

270 − 155 = 115

STEP 3: The lower end of the box represents the lower quartile. Its value is approximately 184 pounds.

The upper end of the box represents the upper quartile. Its value is approximately 220 pounds.

STEP 4: Since 220 pounds represents the upper quartile, 25% of the players will weigh more than that value.

SOLUTION:
a. **The mean weight is approximately 203 pounds.**
b. **The range is approximately 115 pounds.**
c. **The lower quartile is approximately 184 pounds. The upper quartile is approximately 220 pounds.**
d. **Twenty-five percent of the players weigh more than 220 pounds.**

30 DATA DISPLAYS: CIRCLE GRAPHS AND LINE GRAPHS

10.D.1 Select, create, and interpret an appropriate graphical representation (e.g., scatterplot, table, stem-and-leaf plots, box-and-whisker plots, circle graph, line graph, and line plot) for a set of data.

KEY CONCEPTS YOU'LL NEED IN THIS LESSON

1. **Circle graphs** are used to show two relationships:
   - Between parts of a whole, and
   - Between parts and the whole.
2. The entire circle always represents 100% of the data.
3. The sum of the angles of all sectors of a circle graph is 360°. Therefore, the angle at the center of each sector (the central angle) can always be found using a proportion.

$$\frac{\text{central angle}}{360} = \frac{\text{percent of whole}}{100}$$

4. **Line graphs** are generally used to show change in data over time. The change has two components: amount and direction.
5. In a line graph, points represent the data. The lines connect these data points to show a trend. Trends may be either increasing or decreasing.
6. Double line graphs compare trends in two sets of data. The data represented by the two graphs are generally related to one another.

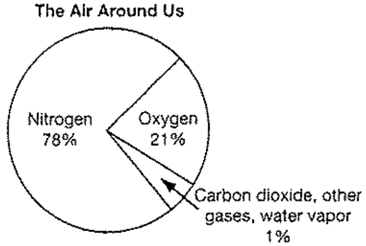
1. CIRCLE GRAPHS

EXAMPLE 1

The air we breathe is made up mainly of nitrogen and oxygen. The graph at the right shows the makeup of the air surrounding the Earth.

Human beings typically breathe about 6 quarts of air every minute. Approximately how many quarts of oxygen do we breathe in a 24-hour period?

The Air Around Us



STRATEGY: Use Key Concept 2.

STEP 1: Recall that to work with measures in different units, it is necessary to convert one measure to the other. Determine the number of minutes in 24 hours.

1 hr = 60 min, so 24 hr = 24 × 60 = 1,440 min

STEP 2: Find the number of quarts of air we breathe in 24 hr.

$6 \times 1{,}440 = \frac{8{,}640 \text{ quarts of air}}{24 \text{ hr}}$

STEP 3: Find the number of quarts of oxygen we breathe in 24 hr.

Recall that 21% can be represented as $\frac{21}{100}$ or 0.21.

Multiply: 0.21 × 8,640 = 1,814.4

SOLUTION: **So, we typically breathe about 1,814 quarts of oxygen every 24 hours.**

EXAMPLE 2

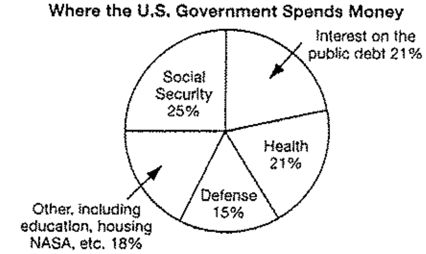
11.7

The graph at the right shows the ways the U.S. government spent money in 1999. The total amount spent was approximately $1.705 trillion or $1,705,000,000,000.

What is the measure of the central angle of each sector?

On which sector does the government spend approximately $250,000,000,000?

Where the U.S. Government Spends Money



STRATEGY: Use Key Concepts 3 and then 2.

STEP 1: Find the measure of each sector's central angle.

Social Security $\frac{25}{100} = \frac{x}{360}$

$x = 90°$

Interest on the Debt, Health $\frac{21}{100} = \frac{x}{360}$

$x = 75.6 \approx 76°$

Defense $\frac{15}{100} = \frac{x}{360}$

$x = 54°$

Other $\frac{18}{100} = \frac{x}{360}$

$x = 64.8 \approx 65°$

$90 + 75.6 + 75.6 + 54 + 64.8 = 360°$

STEP 2: Find the sector on which the government spends approximately $250,000,000.

Use an equation to represent the relationship:

$\$1{,}705{,}000{,}000{,}000x = \$250{,}000{,}000{,}000$

$x = \frac{250{,}000{,}000{,}000}{1{,}705{,}000{,}000{,}000}$

$x = 0.146\ldots$ or approximately 0.15 or 15%

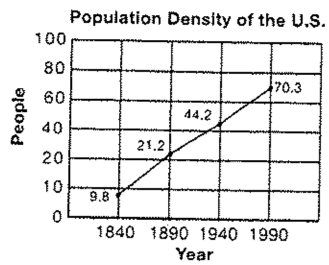
SOLUTION: **The U.S. government spent approximately $250,000,000,000 on defense in 1999.**

2. LINE GRAPHS

EXAMPLE 3

The graph at the right shows how the population density of the United States (number of people per square mile) grew from 1840 through 1990.

By what percent did the population density grow from 1840 through 1890?



STRATEGY: **Use Key Concept 4, using the change in absolute numbers to find the percent change.**

STEP 1: Find the change in the population density from 1840 until 1890: the number of people living (on average) per square mile in the country.

$21.2 - 9.8 = 11.4$

STEP 2: Find the percent change between the years.

To find the percent change, find the actual change (11.4). Then find the percent that it represents of the original, 9.8.

$\frac{11.4}{9.8} \approx 1.16$ $\qquad 1.16 \times 100 = 116\%$

SOLUTION: **Between 1840 and 1890, there was a change in the population density in the country of approximately 116%.**

Reading data in line graphs can often lead to understandings that are beyond what is actually recorded in the graph.

EXAMPLE 4

Of the years shown, the greatest percent growth in the total U.S. population occurred between 1840 and 1890, when the population more than tripled. However, the greatest increase in density was between 1890 and 1940. What conclusion can you draw about changes in the country during the two time periods?

STRATEGY: **Use Key Concepts 4 and 5, keeping in mind how the data in the graph relate to the question.**

STEP 1: Consider the meaning of population density: the number of people per square mile of area. If the population grows more quickly than the area, the density will increase. If the area grows more quickly than the population, the density will decrease.

Since the population increased by the greatest percent during 1840–1890, but the density increased most during another period, the area of the country must have expanded more during 1840–1890 than in the other period.

In fact, during the period 1840–1890, the area of the country increased from 1,792,552 mi^2 to 3,612,299 mi^2. From 1890–1940, the area increased from 3,612,299 mi^2 to 3,618,770 mi^2. [*Data is from the World Almanac and Book of Facts 2000, p. 386*]

SOLUTION: **One of the following conclusions can be drawn:**

- **From 1840–1890, the area of the country grew faster than the population when compared to 1890–1940, or**
- **From 1890–1940, the population of the country grew faster than the area when compared to 1840–1890.**

3. DOUBLE LINE GRAPHS

EXAMPLE 5

The graph compares how the percent of working men and women above the age of 65 has changed over the past 50 years.

What conclusions can you reach from studying the data in the graph?