Lesson 9: Looking for Associations

Goals

- Calculate relative frequencies, and describe (orally and in writing) associations between variables using a relative frequency table.
- Coordinate (orally and in writing) two-way tables, bar graphs, and segmented bar graphs representing the same data.

Learning Targets

- I can identify the same data represented in a bar graph, a segmented bar graph, and a two-way table.
- I can use a two-way frequency table or relative frequency table to find associations among variables.

Lesson Narrative

In this lesson, students study categorical data displayed in two-way tables, bar graphs, and segmented bar graphs. The different graphical representations help students visualize the frequencies and relative frequencies, aiding them in making judgement about whether there is evidence of an association or not in the next lesson. While the concepts and structures in this lesson are not very complex, there are many new terms and representations, and students are given the opportunity to study them carefully so that they can make sense of them (MP1).

Alignments

Addressing

- 8.SP.A.4: Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?

Building Towards

- 8.SP.A.4: Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports
- Notice and Wonder
- Think Pair Share

Required Materials

Pre-printed cards, cut from copies of the blackline master

Required Preparation

Print and cut up cards from the Matching Representations blackline master. Prepare 1 set of cards for every 2 students.

Student Learning Goals

Let's look for associations in data.

9.1 Notice and Wonder: Bar Association

Warm Up: 5 minutes
The purpose of this warm-up is for students to become familiar with a bar graph by noticing and wondering things about it. While reading a bar graph is a review of a previous grade's work, it is an important for students to look for patterns of association in categorical data.

Building Towards

- 8.SP.A.4

Instructional Routines

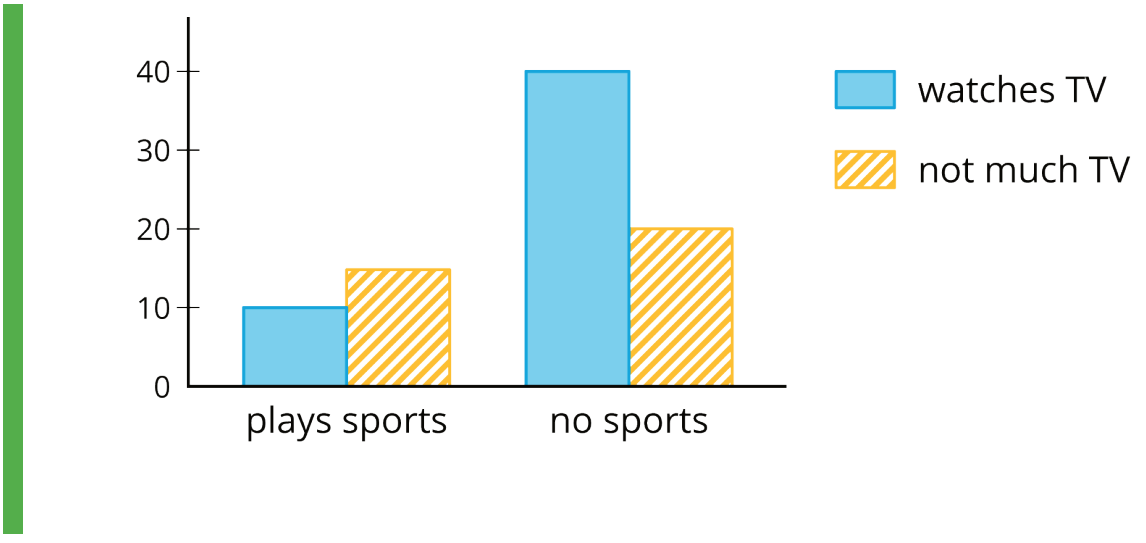
- Notice and Wonder

Launch

Tell students you are going to show them a graph and that their job is to find at least one thing they notice and one thing they wonder about the graph. Give students 1 minute of quiet think time followed by a whole-class discussion.

Student Task Statement

What do you notice? What do you wonder?



Student Response

Answers vary.

Students may notice:

- More people do not play sports than do.
- More people watch TV than watch not much TV.
- The red bar is taller than the blue for “plays sports” but the other way around for “no sports.”

Students may wonder:

- What does “watches TV” mean?
- What does “not much TV” mean?
- Who is represented in the poll?
- Can a person who watches some TV be in the blue and yellow bars?

Activity Synthesis

Ask students to share things they noticed and wondered. Record and display their responses for all to see. If students wondered who the data represents, tell them the data was collected from a survey of students. They were asked whether they played a sport or not and whether they watched more or less than 1 hour of TV each night.

If they did not wonder these things, tell them this information and then ask, “Based on the graph, do you think playing a sport and the amount of time watching TV are related?” (Yes. Among people who watch more than an hour of TV, there are a lot more who do not play a sport than people who do. For people who play a sport, more do not watch TV than do and this is reversed for those who do not play a sport.)

9.2 Card Sort: Matching Representations

20 minutes

In this activity students become familiar with two-way tables, clustered bar graphs, and segmented bar graphs by matching different situations. They label the diagrams to match the data given and create a table to match the data shown in one of the bar graphs.

Addressing

- 8.SP.A.4

Instructional Routines

- MLR2: Collect and Display
- MLR8: Discussion Supports

Launch

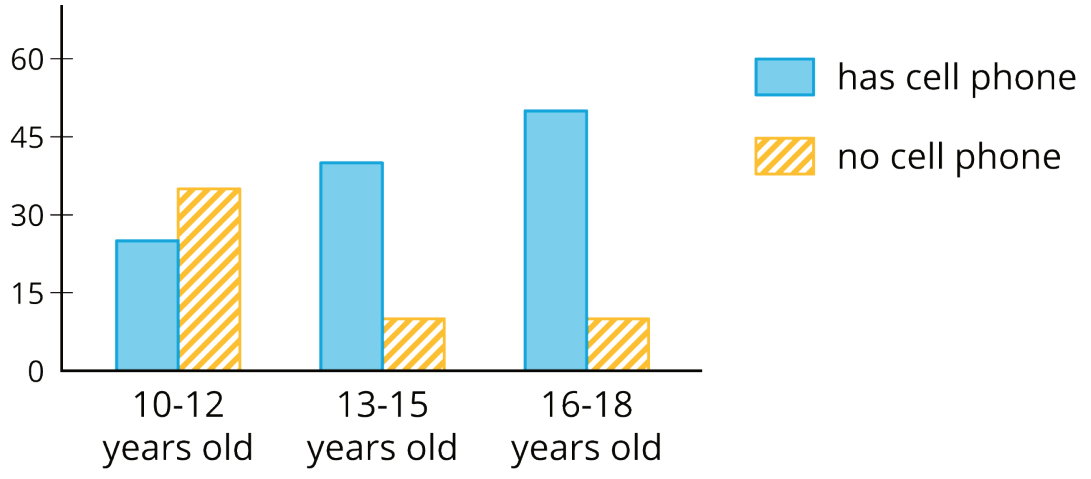
Arrange students in groups of 2–3. Display the two-way table from the activity for all to see. Use MLR 2 (Collect and Display) to gather student language used to describe two-way tables, bar graphs, and segmented bar graphs to be referenced and revised throughout later activities.

	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

Ask students:

- “Why do you think this is called a 'two-way table'?” (The table represents 2 variables. You can think of the people who were surveyed either by age group or by whether or not the person has a cell phone.)
- “What does the 25 mean?” (25 people 10 to 12 years old have cell phones.)
- “What does the 55 mean?” (55 of the people who were surveyed do not have a cell phone.)
- “How many people are represented in this table?” (170 people.)

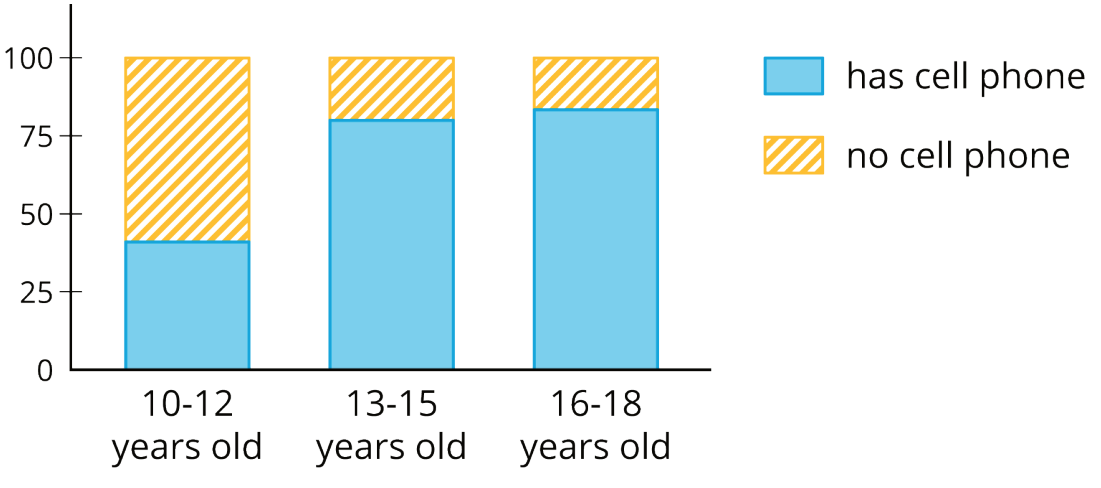
Display the bar graph from the activity for all to see.



Ask students:

- "Why do you think this is called a 'bar graph?'" (Each bar represents how many people are in each group named on the bottom.)
- "What do the blue bars represent? How do you know?" (The blue bars represent how many people in each group have cell phones. I can tell by the key on the right.)
- "Where does the 25 from the table show up in this bar graph?" (The first blue bar shows that 25 10–12 year olds have cell phones.)

Display the **segmented bar graph** from the activity for all to see.



Ask students:

- "Why do you think this is called a 'segmented bar graph?'" (Each bar is segmented into pieces.)
- "What do you think the vertical axis represents?" (The percentage in each section.)

Distribute 1 set of the pre-cut cards from the blackline master to each group.

Access for Students with Disabilities

Representation: Internalize Comprehension. Chunk this task into more manageable parts to differentiate the degree of difficulty or complexity by beginning with fewer cards. For example, give students a subset of the cards to start with and introduce the remaining cards once students have completed their initial set of matches.
Supports accessibility for: Conceptual processing; Organization

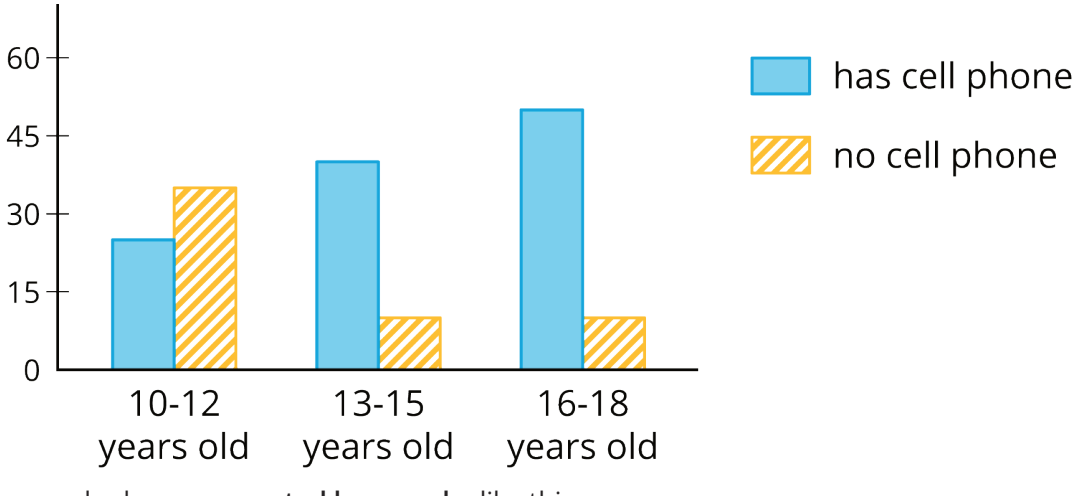
Student Task Statement

Your teacher will hand out some cards.

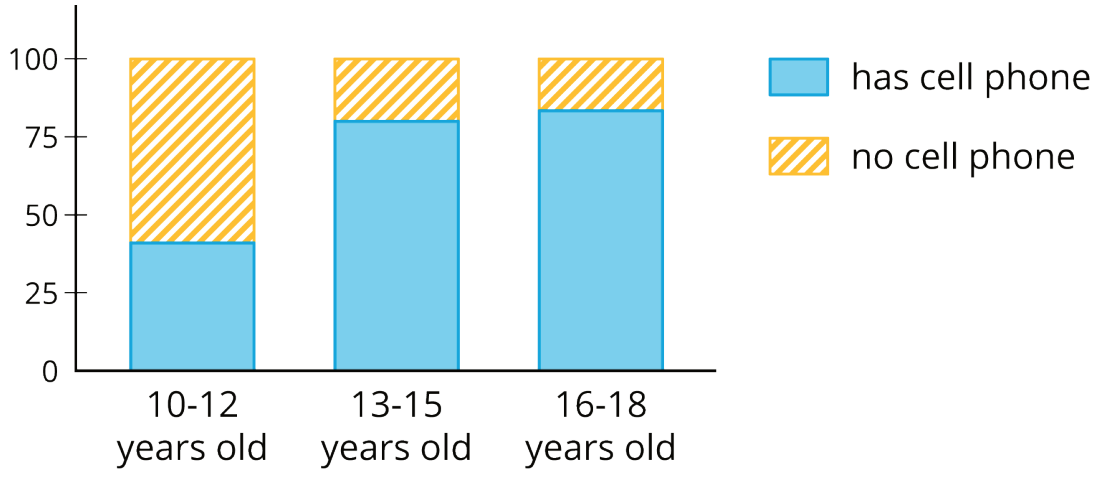
Some cards show **two-way tables** like this:

	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

Some cards show bar graphs like this:



Some cards show **segmented bar graphs** like this:



The bar graphs and segmented bar graphs have their labels removed.

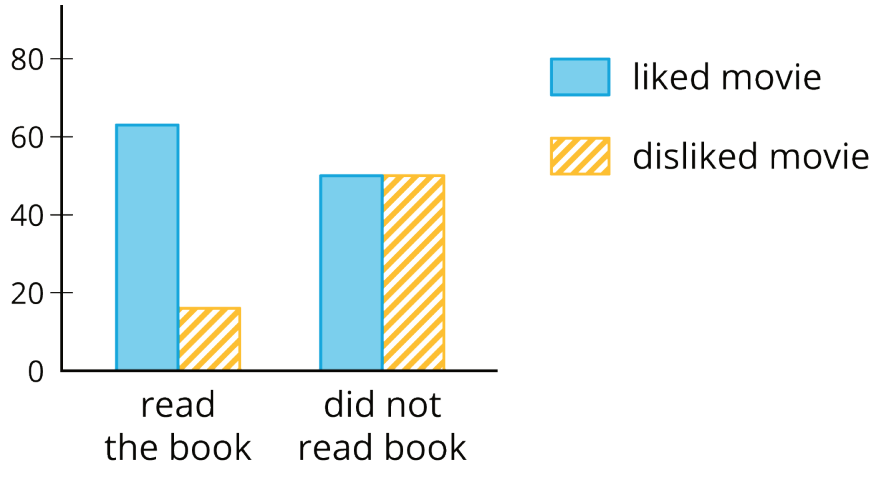
1. Put all the cards that describe the same situation in the same group.
2. One of the groups does not have a two-way table. Make a two-way table for the situation described by the graphs in the group.
3. Label the bar graphs and segmented bar graphs so that the categories represented by each bar are indicated.
4. Describe in your own words the kind of information shown by a segmented bar graph.

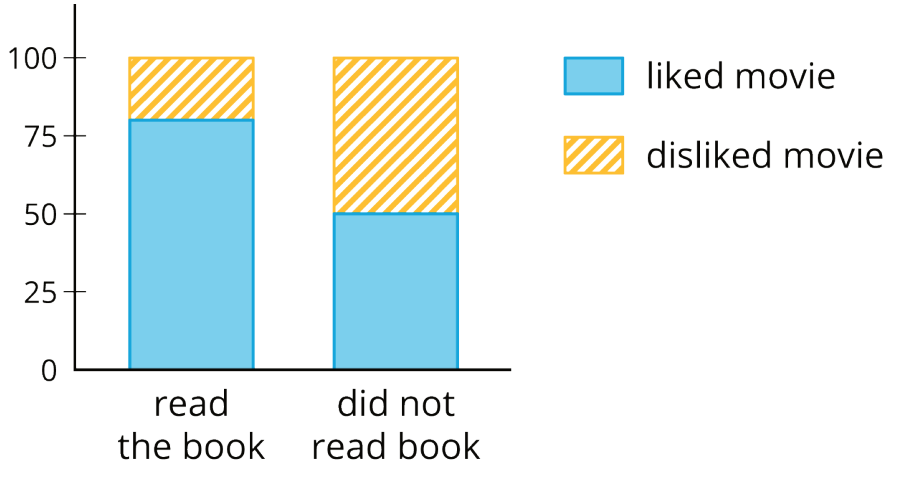
Student Response

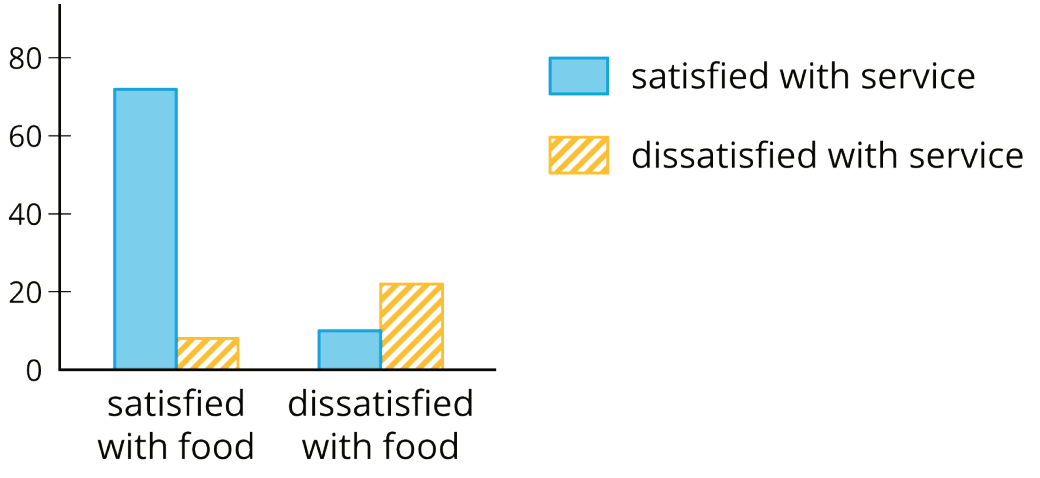
1. See blackline master for the groupings.
2. Answers vary. Sample response:

	ate breakfast	skipped breakfast	total
male	48	12	60
female	48	16	64
total	96	28	124

3. Answers vary. Possible response:





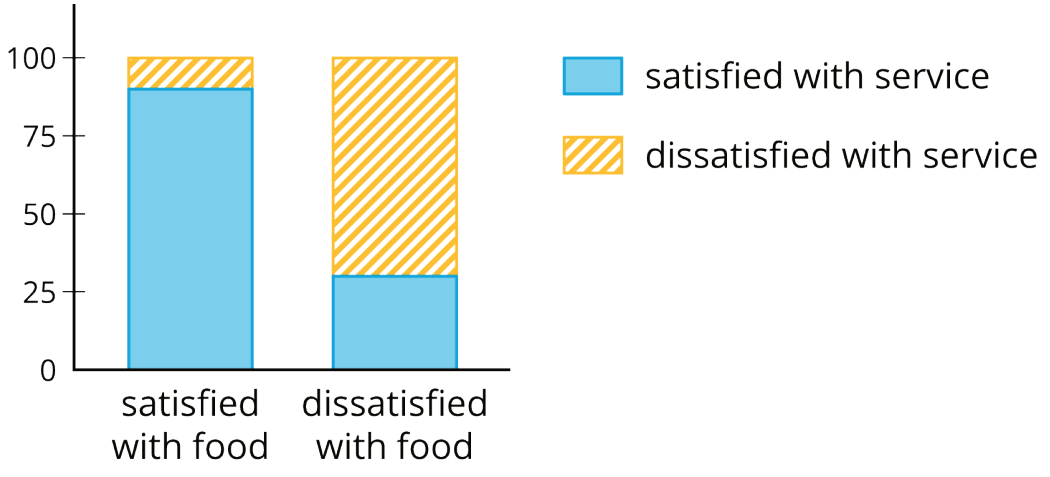


4. In a segmented bar graph, each bar represents 100% of a category. The bar is split into different subcategories that are stacked on top of each other so you can compare their percentages.

Are You Ready for More?

One of the segmented bar graphs is missing. Construct a segmented bar graph that matches the other representations.

Student Response



Activity Synthesis

The purpose of this activity is to help students understand the connections among the three representations of data. It also helps students see the importance of labeled visual representations.

To highlight the connections, ask:

- "What were some strategies you used to match the cards?" (The book table had the same number for those who liked the movie and those who did not like the movie among those who did not read the book, so I knew the two graphs would show equal sized bars.)
- "Were some cards easier to match than others? Explain." (The cards showing the actual numbers were easier to match than the cards showing the percentages.)
- "What strategy did you use to create the two-way data table?" (I started by thinking of two yes or no questions that could be asked and use those to label the rows and columns of the table.)

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support whole-class discussion. For each strategy that is shared, ask students to restate and/or revoice what they heard using mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped to clarify the original statement. This will provide more students with an opportunity to produce language as they interpret the reasoning of others.
Design Principle(s): Support sense-making

9.3 Building Another Type of Two-Way Table

10 minutes

In this activity, students create two-way tables displaying relative frequency. The relative frequency table converts the actual frequency data to percentages which can be useful when comparing groups that include different totals. Finally, students use the relative frequencies to look for a pattern in the data. In the following lesson, students will work with associations in categorical data more explicitly. In this activity, students should use an informal understanding of association to think about whether one is present in the data based on the relative frequency table.

Addressing

- 8.SP.A.4

Instructional Routines

- MLR2: Collect and Display
- Think Pair Share

Launch

Keep students in groups of 2–3. After an introduction to relative frequency tables, allow students 3 minutes quiet work time followed by partner discussion and whole-class discussion.

Display the table for all to see.

	watch the news daily	does not watch the news daily	total
younger than 18	30	80	110
18 or older	10	5	15
total	40	85	125

Ask students, "Based on this data, who is more likely to watch the news daily: someone who is younger than 18 or someone who is 18 or older?"

Tell students that, based on the numbers in the table, there are more younger people who watch the news (30) than older (10). On the other hand, the survey reached out to 110 young people and only 15 older people. Without looking at the whole table, that information may have been missed.

In cases like this, finding a relative frequency including percentages can be more helpful than looking at the actual frequency, which is what they are going to do now.

Use MLR 2 (Collect and Display) to co-create a graphic organizer with students comparing and contrasting the definitions and characteristics of “actual frequency” and “relative frequency” to be referenced and revised throughout this and later activities.

Access for English Language Learners

Conversing, Representing, Writing: MLR2 Collect and Display. As students work in groups, capture the vocabulary and phrases students use when comparing and contrasting the definitions and characteristics of "actual frequency" and "relative frequency." Listen for students who identify pros and cons for each representation and when one is preferable compared to the other. Record their language on a visual display that can be referenced in future discussions. This routine provides feedback to students that supports sense-making and increases meta-awareness of mathematical language.
Design Principle(s): Support sense-making; Maximize meta-awareness

Anticipated Misconceptions

Students may not calculate relative frequencies correctly. Look to see if they are dividing into the total for each row, instead of some other number in the row, or the total for the entire table.

Student Task Statement

Here is a two-way table that shows data about cell phone usage among children aged 10 to 18.

	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

1. Complete the table. In each row, the entries for "has cell phone" and "does not have cell phone" should have the total 100%. Round entries to the nearest percentage point.

	has cell phone	does not have cell phone	total
10 to 12 years old	42%		
13 to 15 years old			100%
16 to 18 years old		17%	

This is still a two-way table. Instead of showing *frequency*, this table shows **relative frequency**.

2. Two-way tables that show relative frequencies often don't include a "total" row at the bottom. Why?

3. Is there an association between age and cell phone use? How does the two-way table of relative frequencies help to illustrate this?

Student Response

1.

	has cell phone	does not have cell phone	total
10 to 12 years old	42%	58% since $35 \div 60 \approx 0.58$	100%
13 to 15 years old	80% since $40 \div 50 = 0.8$	20% since $10 \div 50 = 0.2$	100%
16 to 18 years old	83% since $50 \div 60 \approx 0.83$	17%	100%

2. The columns are not percentages of the same number, so you can't add them. The percentages in any given column are not likely to add up to 100%.

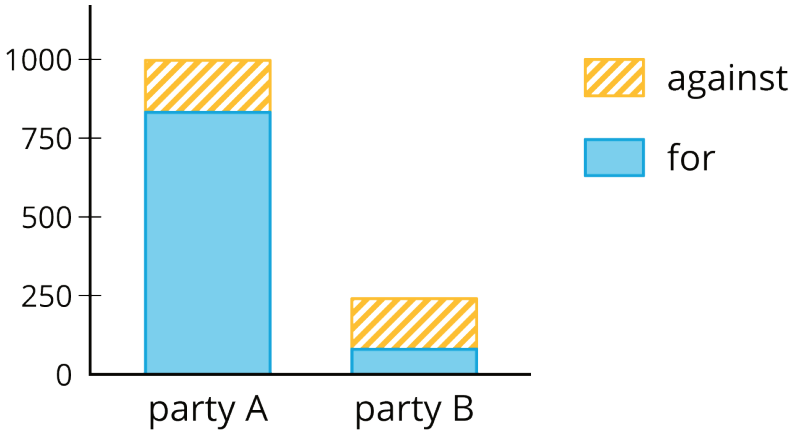
3. Yes, there is an association between age and cell phone use. A much higher percentage of children from 13 to 18 years old have cell phones than children from 10 to 12 years old do.

Are You Ready for More?

A pollster attends a rally and surveys many of the participants about whether they associate with political Party A or political Party B and whether they are for or against Proposition 3.14 going up for vote soon. The results are sorted into the table shown.

	for	against
party A	832	165
party B	80	160

- A news station reports these results by saying, "A poll shows that about the same number of people from both parties are voting against Proposition 3.14."
- A second news station shows this graphic.

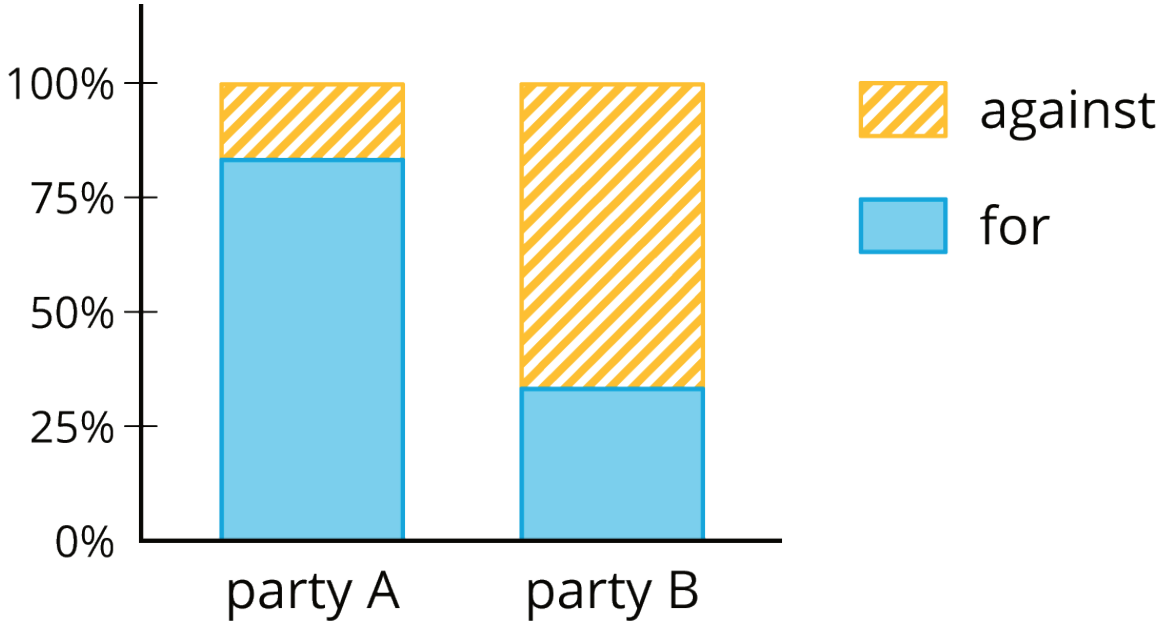


1. Are any of the news reports misleading? Explain your reasoning.
2. Create a headline, graphic, and short description that more accurately represents the data in the table.

Student Response

Answers vary. Sample response:

1. Yes, the first news station did not take into account that there were not as many people from party B that were polled. The graphic from the second news station is hard to understand because the bars from Party B are so much smaller than from Party A, so it is hard to tell what it is trying to say.
2. Headline: Party A is For Proposition 3.14 while Party B is Against



Description: Based on this poll, more than 80% of people in Party A are for Proposition 3.14, while only 33% of people in Party B share the same opinion. There seems to be a big divide along party lines for the proposition!

Activity Synthesis

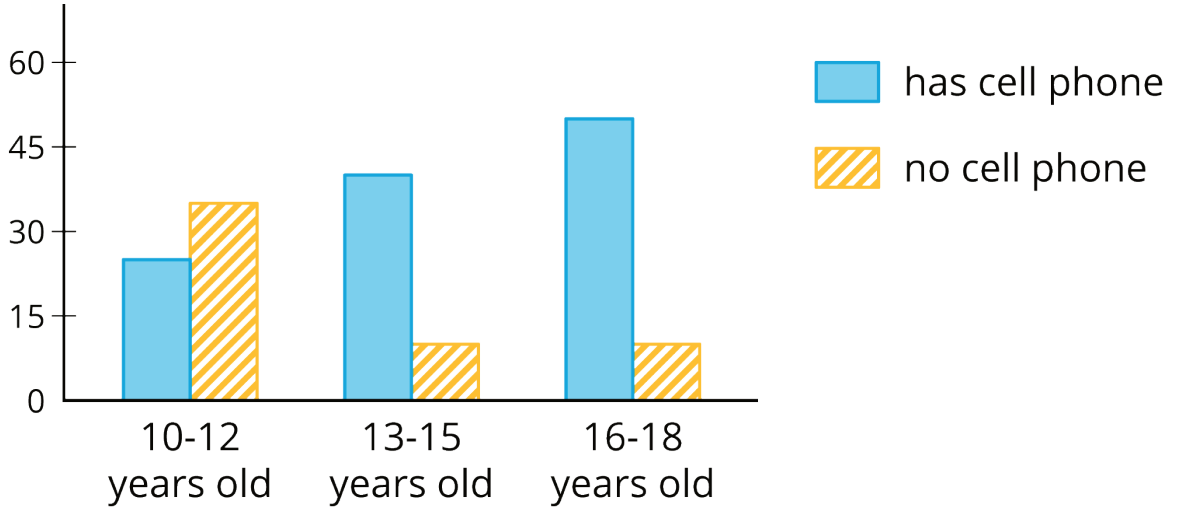
The purpose of this discussion is to help students see the usefulness of two-way tables that display relative frequency.

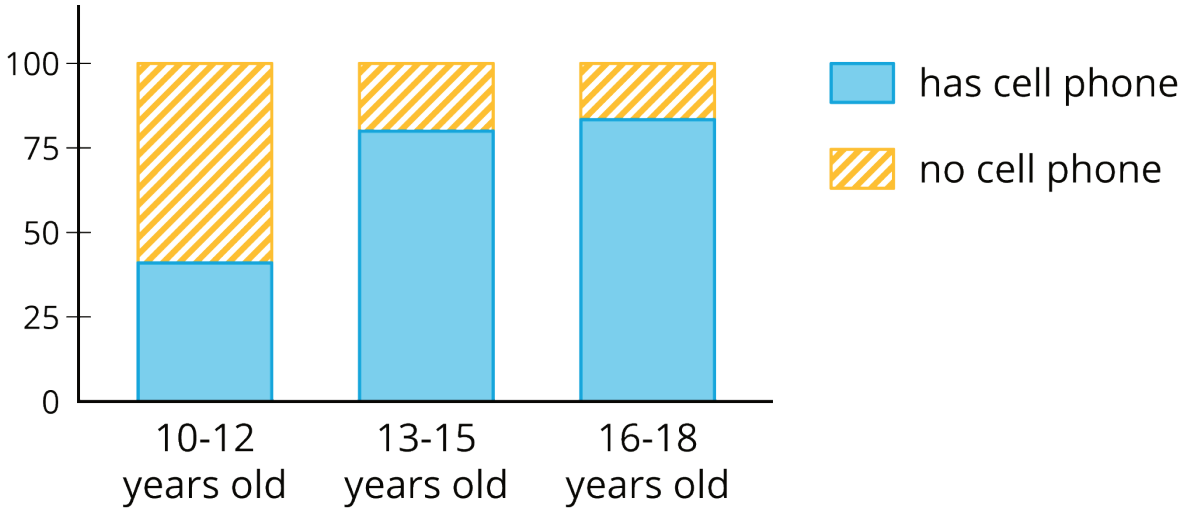
Some questions for discussion:

- "What does the 42% in table mean?" (42% of people 10 to 12 years old have a cell phone.)
- "What is the total of the percentages in the first column? What's wrong with that answer?" (205%. It doesn't make sense since it is greater than 100%. It is combining percentages of different wholes, so it doesn't make sense to add them.)
- "Did you say there was or was not an association between age and cell-phone use? Explain your reasoning." (There is an association since the percentages for the 10 to 12 year olds is very different from the other two age groups.)
- "What did the relative frequency table show you that was harder to see in the original two-way table?" (It was much easier to see the association with the percentages than the numbers. For example, the number of people without a cell phone for the older 2 age groups was the same, but the percentages were a little different.)

Lesson Synthesis

Display the graphs and tell students to refer to the data from the tables in the Building Another Type of Two-Way Table activity.





Consider asking these discussion questions to emphasize the main ideas from the lesson:

- "Which graphical representation do you prefer for the data?" (It is easier to see the association with the segmented bar graph, but the actual values are lost.)
- "In the original table, what did the number 40 represent? How is that group of people represented in the other table and the two graphs?" (There are 40 people who are 13 to 15 years old that have a cell phone. In the relative frequency table, this is represented by the 80%. In the bar graph, this is represented by the taller, blue bar for set of bars labeled 13–15 years old. In the segmented bar graph, this is represented by the bottom, blue segment of the bar labeled 13–15 years old.)
- "What values in the tables represent the same information as the tallest, yellow-striped bar in the bar graph?" (The tallest yellow-striped bar represents the number of people who are 10–12 years old who do not have cell phones. In the original table, this is the 35 value and in the relative frequency table this is about 58%.)

9.4 Guitar and Golf

Cool Down: 5 minutes
Students create their own data based on how they view the students at their school. This data is then converted into a relative frequency table to practice the skill needed to create segmented bar graphs.

Addressing

- 8.SP.A.4

Launch

Tell students that they will make up their own numbers to fill in the two-way table so that the data reflects what they think is reasonable for students at your school. Prepare a two-way table to collect the raw data from the class that will be used in a future lesson.

Student Task Statement

1. In a class of 25 students, some students play a sport, some play a musical instrument, some do both, some do neither. Complete the two-way table to show data that might come from this class.

	plays an instrument	does not play an instrument	total
plays a sport			
does not play a sport	5		
total			25

2. Using the entries from the previous table, complete this table so that it shows relative frequencies. Round entries to the nearest percentage point.

	plays an instrument	does not play an instrument	total
plays a sport			
does not play a sport			

Student Response

Answers vary. Sample response:

1.

	plays an instrument	does not play an instrument	total
plays a sport	2	16	18
does not play a sport	5	2	7
total	7	18	25

2.

	plays an instrument	does not play an instrument	total
plays a sport	11% since $2 \div 18 \approx 0.11$	89% since $16 \div 18 \approx 0.89$	100%
does not play a sport	71% since $5 \div 7 \approx 0.71$	29% since $2 \div 7 \approx 0.29$	100%

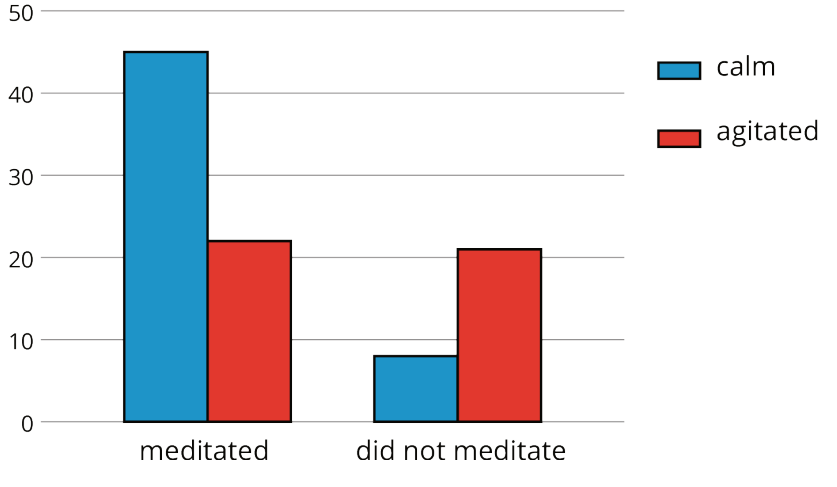
Student Lesson Summary

When we collect data by counting things in various categories, like red, blue, or yellow, we call the data *categorical data*, and we say that color is a *categorical variable*.

We can use **two-way tables** to investigate possible connections between two categorical variables. For example, this two-way table of frequencies shows the results of a study of meditation and state of mind of athletes before a track meet.

	meditated	did not meditate	total
calm	45	8	53
agitated	23	21	44
total	68	29	97

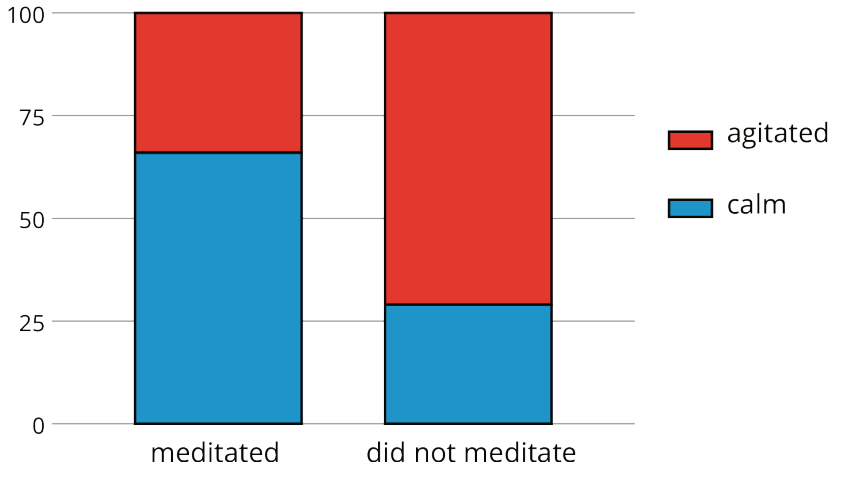
If we are interested in the question of whether there is an association between meditating and being calm, we might present the frequencies in a bar graph, grouping data about meditators and grouping data about non-meditators, so we can compare the numbers of calm and agitated athletes in each group.



Notice that the number of athletes who did not meditate is small compared to the number who meditated (29 as compared to 68, as shown in the table).

If we want to know the proportions of calm meditators and calm non-meditators, we can make a two-way table of **relative frequencies** and present the relative frequencies in a **segmented bar graph**.

	meditated	did not meditate
calm	66%	28%
agitated	34%	72%
total	100%	100%



Glossary

- relative frequency
- segmented bar graph
- two-way table

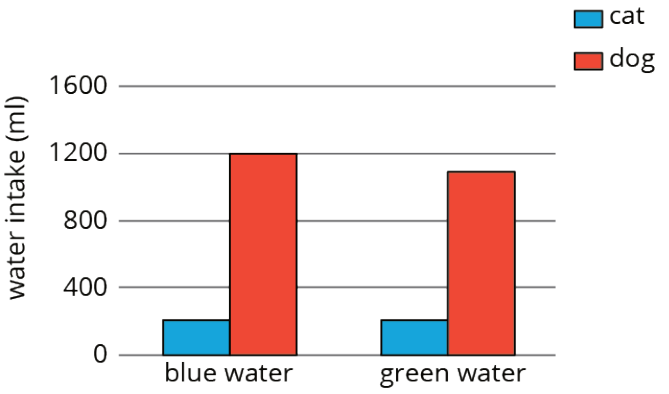
Lesson 9 Practice Problems

Problem 1

Statement

A scientist wants to know if the color of the water affects how much animals drink. The average amount of water each animal drinks was recorded in milliliters for a week and then graphed. Is there evidence to suggest an association between water color and animal?

	cat intake (ml)	dog intake (ml)	total (ml)
blue water	210	1200	1410
green water	200	1100	1300
total	410	2300	2710



Solution

No, the relative frequencies of the animals drinking each color of water are about the same, so there is no evidence of association.

Problem 2

Statement

A farmer brings his produce to the farmer's market and records whether people buy lettuce, apples, both, or something else.

	bought apples	did not buy apples
bought lettuce	14	58
did not buy lettuce	8	29

Make a table that shows the relative frequencies for each row. Use this table to decide if there is an association between buying lettuce and buying apples.

Solution

	bought apples	did not buy apples	total
bought lettuce	19% since $14 + 58 = 72$ and $14 \div 72 = 0.19\overline{4}$	81% since $58 \div 72 = 0.80\overline{5}$	100%
did not buy lettuce	22% since $8 + 29 = 37$ and $8 \div 37 = 0.\overline{216}$	78% since $29 \div 37 = 0.\overline{783}$	100%

Since the percentages in each column are close, there is not enough evidence to suggest an association between buying apples and lettuce.

Problem 3

Statement

Researchers at a media company want to study news-reading habits among different age groups. They tracked print and online subscription data and made a 2-way table.

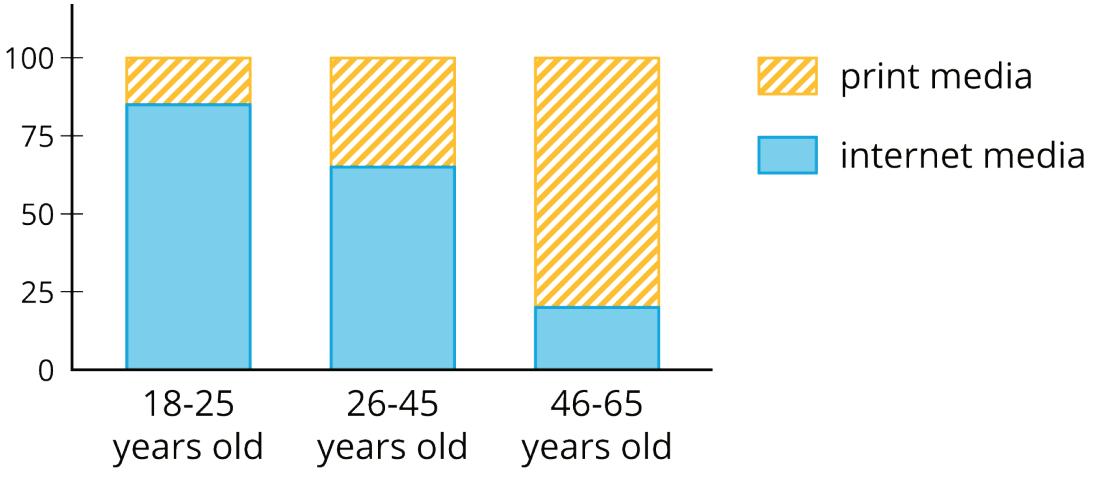
	internet media	print media
18-25 year olds	151	28
26-45 year olds	132	72
46-65 year olds	48	165

a. Create a segmented bar graph using one bar for each row of the table.

b. Is there an association between age groups and the method they use to read articles? Explain your reasoning.

Solution

a.

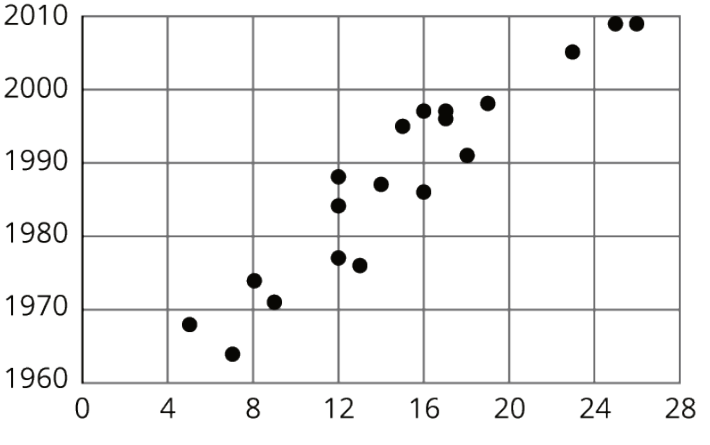


b. Yes. Explanations vary. Sample explanation: The segments of the bars are not very close to being the same size. Younger age groups use the internet articles much more than print articles, while the oldest age group reverses that pattern.

Problem 4

Statement

Using the data in the scatter plot, what is a reasonable slope of a model that fits this data?



A. -2.5

B. -1

C. 1

D. 2.5

Solution

D

(From Unit 6, Lesson 6.)