Analyzing Data Sets for Two Categorical Variables

10

10.1

Could You Participate in Our Survey?

Interpreting Frequency Distributions

LEARNING GOALS

In this lesson, you will:

- Construct and interpret frequency and frequency marginal distributions displayed in two-way tables for two-variable categorical data.
- Create and interpret graphs of frequency distributions displayed in two-way tables.

KEY TERMS

- categorical data
- two-way frequency table
- frequency distribution
- joint frequency
- frequency marginal distribution

In order for many businesses to be successful they need one thing: money. If they do not have money, they have two choices—make more money or cut back on spending. While businesses have the opportunity to make more money, government-funded programs, such as schools or public libraries, do not have this option. In turn, these programs must figure out ways to cut back on their spending. So what do they do? These places must first prioritize and determine what areas need the most funding. For example, a school might desperately need new computers. Once these decisions are made, they must look at their budget and figure out where they can cut costs so that they have the money needed for these priorities. Oftentimes in schools this may mean getting rid of extra-curricular activities or limiting the school lunch menu.

Have you noticed any changes in your school? Do you think these changes are the result of school budget cuts? Do you have any ideas on how schools could cut costs without eliminating any programs?

PROBLEM 1 Hot Lunch!

Ms. Seymour is the school cafeteria supervisor at Williams High School. She has been asked to cut her food budget for the upcoming school year. One idea she has is to cut the number of meal choices during the week. However, determining which meal to cut will not be an easy decision. Ms. Seymour wonders if there is a difference in students' favorite cafeteria meals by grade level. She decides to survey the students in Mr. Kolbe's gym class, which consists of 9th and 10th graders. She recorded the results of her survey in the table shown.

Grade	Favorite Meal
9	Salad bar
10	Burgers
10	Pizza
10	Chicken nuggets
10	Chicken nuggets
9	Burgers
10	Salad bar
9	Salad bar
10	Chicken nuggets
9	Burgers
10	Pizza
9	Salad bar
9	Burgers
10	Burgers
9	Chicken nuggets
9	Salad bar
10	Chicken nuggets
10	Chicken nuggets
10	Salad bar
10	Burgers
10	Salad bar
9	Burgers
9	Pizza
10	Chicken nuggets
10	Salad bar
9	Salad bar
10	Pizza
9	Pizza
10	Chicken nuggets
9	Pizza

1. Analyze Ms. Seymour's data table.

   a. Can you summarize her findings just by looking at her data table? Explain why or why not.

   b. Identify the variables of the data from Ms. Seymour's survey. Are the variables in the table numerical? Explain your reasoning.

Previously, you explored the relationship between two variables that had data values that were quantitative, or numerical. Data that can be grouped into categories, such as favorite meals, are called **categorical data**.

One method of organizing categorical data is to use a *two-way frequency table*. A **two-way frequency table** displays categorical data by representing the number of occurrences that fall into each group for two variables. On the table, one variable is divided into rows and the other is divided into columns.

2. Identify the groups for the variable, grade level. How many groups are there for this variable?

3. Identify the groups for the variable, favorite meal. How many groups are there for this variable?

4. Create a two-way frequency table of the data.

a. Enter the name of each group.

b. Record the favorite meal for each student in the appropriate row using tally marks. Then, write the frequency of each meal for each grade level.

Favorite Meals of Students

Grade Level				

5. What observations can you make from the data about the students' favorite meals?

The table you created is a *frequency distribution*. A **frequency distribution** displays the frequencies for categorical data in a two-way table. Each time you determined the frequency of one favorite meal of one of the grade levels, you recorded a *joint frequency*. Any frequency you record within the body of a two-way frequency table is known as a **joint frequency**.

A two-way frequency table is helpful in organizing each group's frequency in an efficient way. However, it is common to determine the total number of people surveyed just to ensure that a good survey was taken. Determining this total is also helpful to ensure that you recorded the data accurately within the table. For example, if you know 50 people took part in the survey, and the sum of the joint frequencies is 47, then you know that you are missing three data points from the data set.

6. Use the data from your frequency distribution to determine the total number of 9th graders and 10th graders, and to determine the total number of frequencies for each favorite meal category.

Favorite Meals of Students

Grade Level	Burgers	Chicken Nuggets	Pizza	Salad Bar	Total
9th grade					
10th grade					
Total					

You just created a *frequency marginal distribution* of the data by determining the totals for each group. A **frequency marginal distribution** displays the total of the frequencies of the rows or columns of a frequency distribution.

10

7. Analyze the frequency marginal distribution to answer each question.
   a. How many 9th graders participated in the survey?

   b. How many students prefer burgers?

   c. How many students prefer chicken nuggets?

   d. How many 10th graders participated in the survey?

   e. How many students prefer salad bar?

8. What do you notice about the total number of students who prefer burgers, chicken nuggets, pizza, and salad bar; and the total number of 9th and 10th graders? Can you use this observation to determine if you correctly determined the frequency distribution?

9. Use the frequency marginal distribution to answer each question.

   a. Which meal is the least favorite of all students?

   b. Which meal is the least favorite of 9th graders?

   c. Which meal is the most favorite of all students?

   d. Which meal is the most favorite of the 10th graders?

10

PROBLEM 2 Representing Data

While a two-way table shows a numerical summary of the data, a graph can help relay information about a survey in a visual way. Remember, every graph tells a story.

Recall that Ms. Seymour is trying to determine ways to cut the cafeteria budget for the upcoming school year. She has already gathered her data and organized it in a frequency distribution table. Ms. Seymour has a couple of ideas, but she would like to use a graph to visually display the ideas she has for cutting the cafeteria budget.

Ms. Seymour's data are displayed in the frequency distribution table shown.

Favorite Meals of Students

Grade Level	Burgers	Chicken Nuggets	Pizza	Salad Bar
9th grade	4	1	3	5
10th grade	3	7	3	4

1. Analyze the frequency distribution table.

   a. Determine which graphs would be appropriate to display Ms. Seymour's data. Justify your response.

   b. Determine which graphs would not be appropriate to display Ms. Seymour's data. Explain why these graphs would not be appropriate for displaying this data.

2. Construct a double bar graph of the frequencies. Let the *x*-axis represent the favorite meals, and let the *y*-axis represent the number of students.

3. What conclusions can you draw by examining the graph?

4. Use the graph to determine if you represented the data from the frequency distribution table accurately. Explain how you verified that the data in the graph matches the data in the frequency distribution table.

5. Construct a bar graph of the frequencies. This time, let the *x*-axis represent grade level.

10

6. What conclusions can you draw by examining the graph?

7. Does it matter which graph Ms. Seymour's uses to display her survey data? Why or why not.

PROBLEM 3 Putting It All Together

Ms. Seymour must decide on a plan for the upcoming school year. The principal of the school would like Ms. Seymour to present her data and a graph to justify her decision to cut costs.

1. Which meal choice would you cut according to the data? Explain why you would discontinue that meal choice. Then explain which graph you would recommend Ms. Seymour use when she presents her plan.

2. Ms. Seymour just thought of an idea, and she thinks it will help cut the cafeteria costs. She is recommending that two lunch periods be created: one for the 9th graders and one for the 10th graders. She thinks that if two lunch periods exist, she can keep all four meal choices, but just cook a lesser amount of certain choices; thus cutting costs.

   a. Do you think Ms. Seymour should present this idea to the principal? Use the data to justify your reasoning.

   b. Which graph would you recommend Ms. Seymour use to justify her solution? Explain your reasoning.

Be prepared to share your solutions and methods.

10

10.2

It's So Hot Outside!

Relative Frequency Distribution

LEARNING GOALS

In this lesson, you will:

- Construct and interpret relative frequency distribution and relative frequency marginal distributions displayed in two-way tables for categorical data.
- Analyze and use relative frequency marginal distributions to make decisions for a problem situation.

KEY TERMS

- relative frequency distribution
- relative frequency marginal distribution

Humans are warm-blooded mammals. Normal human body temperature can range from 97°F to 99°F so anything significantly higher or lower than that can cause major issues. Hypothermia and hyperthermia are two conditions that can occur when the body's temperature is greatly different from normal body temperature. Hypothermia occurs when the body's temperature drops below 95°F. This occurs when the body is exposed to low temperatures for an extended period of time. Hyperthermia is the opposite of hypothermia and occurs when the body is exposed to high temperatures for a prolonged period of time. Hyperthermia occurs when the body produces more heat than it can emit and the body's temperature climbs to over 100°F.

While it seems like that is not much of a change in temperature, your body is well equipped to regulate its temperature and you can actually be exposed to heat or cold for some time before experiencing any hyperthermia or hypothermia symptoms.

PROBLEM 1 What Do You Want to Do?

The Northpointe community outreach director wants to plan special summer activities for the members of Northpointe. He posts a survey on the local newspaper's website to gather information on the favorite activities of the community members. Participants identified their age and then chose from four given activities. The responses gathered from the survey are shown.

Activities Preferred During Hot Weather

	Sports	Movies	Reading	Walking	Total
Students Age 18 Years Old and Under	20	30	22	8	
Adults Age 19 Thru 50 Years Old	10	32	25	43	
Adults Over 50 Years Old	5	20	35	30	
Total					

1. Complete the frequency marginal distribution for the data given.

While the raw data provides some information, it is often more efficient to use percents when analyzing data. The relative frequencies of each data entry can provide that information. Representing the relative frequencies for joint data displayed in a two-way table is called a *relative frequency distribution*. The **relative frequency distribution** provides the ratio of occurrences in each category to the total number of occurrences. Displaying the relative frequencies for the rows or columns is called a *relative frequency marginal distribution*. The **relative frequency marginal distribution** provides the ratio of total occurrences for each category to the total number of occurrences.

2. Construct a relative frequency distribution and relative frequency marginal distribution of the data.

Activities Preferred During Hot Weather

	Sports	Movies	Reading	Walking	Total
Students Age 18 Years Old and Under					
Adults Age 19 Thru 50 Years Old					
Adults Over 50 Years Old					
Total					

3. After creating the relative frequency distribution and relative frequency marginal distribution, the students in Mr. Thomas's class made the following statements.

Marie

7.1% of students age 18 and under prefer playing sports in the hot weather.

Shane

1.07% of adults over age 50 prefer walking in the hot weather.

Isaac

29.3% of participants in the survey prefer watching movies or reading in the hot weather.

Olivia

More adults over 50 responded to the survey than any other age group.

Aaron

Playing sports is the least popular activity in the hot weather according to the survey results.

For each statement explain why the student is correct or incorrect. If the student is incorrect tell what the correct statement would be.

4. Which age group made up the smallest percent of people surveyed?

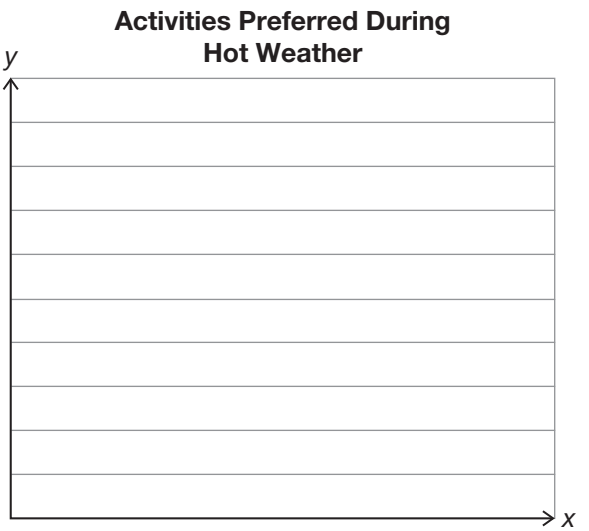
5. Which activity was preferred by the largest percent of people surveyed?

PROBLEM 2 How Does the Data Stack Up?

Previously, you used a bar graph to visually represent data. Another way to represent data is to use a stacked bar graph in which the bars are stacked on top of each other as opposed to sitting next to each other.

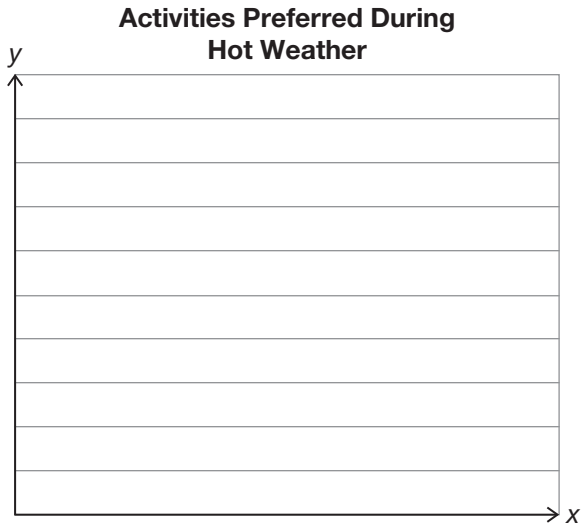
10

1. Construct a stacked bar graph of the relative frequency distribution. Let the *x*-axis represent age group.



2. What conclusions can you draw by examining the graph?

3. Construct a stacked bar graph of the relative frequency distribution. Let the *x*-axis represent the activities.



4. What conclusions can you draw by examining the graph?

5. Name some advantages of graphing the data by age group. Name some advantages of graphing the data by activity.

PROBLEM 3 So What Should We Do?

Now that the community outreach director has gathered this information, he wants to use it to plan different activities for the summer.

1. Analyze each activity shown. Determine whether you think the activity would be a good idea to have during the summer. Explain your reasoning based on the data.

   a. Start a walking club for community members age 19 to 50.

   b. Set up an over age 50 soccer tournament.

   c. Start an age 18 and under ultimate Frisbee league.

2. The community outreach director wants to offer one summer activity each week that will appeal to all ages of the community. Write a letter to the community outreach director recommending one activity and tell why the other activities may not be the best activities during the summer. Use the data to support your idea.

Be prepared to share your solutions and methods.

She Blinded Me with Science!

10.3

Relative Frequency Conditional Distribution

LEARNING GOALS

In this lesson, you will:

- Construct and interpret relative frequency conditional distributions displayed in two-way tables for categorical data.

KEY TERM

- relative frequency conditional distribution

Chances are pretty good you have taken some sort of science class every year since you started school. However, unlike elementary school science which was very general, you are probably now taking a more specific science class. The word science comes from the Latin word *scientia* which means knowledge and the study of science in the broadest sense has existed since humans began communicating knowledge to each other. In the Age of Enlightenment, which took place in the 17th and 18th centuries, there was rapid scientific advancement where scientists such as Descartes and Newton confirmed scientific thinking with experiments and mathematics. Today there are two major groups of sciences: natural sciences and social sciences. The natural sciences include topics such as astronomy, biology, chemistry, physics, and earth science. The social sciences include topics dealing with society and human behavior such as economics, geography, linguistics, and psychology.

Often scientists are belittled or questioned for their beliefs or ideas, such as when Galileo suggested the Earth traveled around the sun. However, without the work scientists have achieved we would have very little understanding about the world around us.

PROBLEM 1 Passing the Class

Mr. Lewis teaches three science classes at Matthews High School. He wants to compare the grades of the three classes of his students. He creates the following two-way frequency table shown.

Grades of Mr. Lewis's Science Students

Science Classes	A	B	C	D	F	Total
Biology	6	6	5	1	2	
Chemistry	4	8	12	4	2	
Physics	2	5	6	1	1	
Total						

10

1. Complete the frequency marginal distributions on Mr. Lewis's frequency table.

2. Complete the relative frequency and relative frequency marginal distributions for the data.

Grades of Mr. Lewis's Science Students

Science Classes	A	B	C	D	F	Total
Biology						
Chemistry						
Physics						
Total						

3. Write a paragraph interpreting the relative frequency distributions and relative frequency marginal distributions for the data.

4. Create a stacked bar grade to represent the percent of students passing in each class.

5. Campbell claims that the Chemistry class is the smartest because they have the greatest percent of students passing. Is Campbell's statement correct? Explain your reasoning.

PROBLEM 2 Which Is the Best?

Because each science class has a different number of students, the relative frequencies cannot help determine which class is doing the "best." Instead, you can use a *relative frequency conditional distribution* to determine this information. A **relative frequency conditional distribution** is the percent or ratio of occurrences of a category given the specific value of another category.

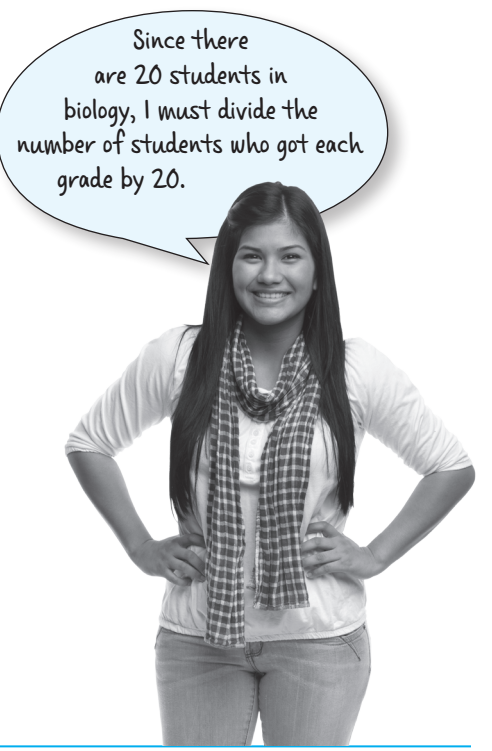
Let's construct a relative frequency conditional distribution of grades given the classes.

1. Use the information from Problem 1, Question 1 to determine the relative frequency for each grade given that particular class.

Grades of Mr. Lewis's Science Students

Science Classes	A	B	C	D	F	Total
Biology	$\frac{6}{20} = 30\%$					
Chemistry			$\frac{12}{30} = 40\%$			
Physics				$\frac{1}{15} \approx 6.7\%$		

2. Interpret the relative frequency conditional distributions of each class.



3. Use the relative frequency conditional distribution to answer each question.

   a. What percent of the biology students are passing?

   b. What percent of the chemistry students are passing?

   c. What percent of the physics students are passing?

   d. Which science class is doing the best according to their grades? Explain your reasoning.

   e. How does this compare to the statement Campbell made in Problem 1, *Passing the Class*, Question 5?

4. Which science class has the greatest percent of students failing?

Mr. Lewis also teaches two General Science classes. He wants to teach his students about a topic they are most interested in. He surveys his students and records the data in the table shown.

Science Classes	Matter	Plants and Animals	Astronomy	Anatomy	Genetics	Total
Class 1	5	3	10	3	4	25
Class 2	9	5	3	7	6	30
Total	14	8	13	10	10	55

5. Mr. Lewis wants to teach the same topic to both classes. Which topic would you recommend Mr. Lewis teach? Use the data to explain why you made your suggestion to Mr. Lewis.

Be prepared to share your solutions and methods.

10.4

Oh! Switch the Station!

Drawing Conclusions from Data

LEARNING GOALS

In this lesson, you will:

- Analyze different categorical data.
- Use categorical data to make decisions.

Have you ever gotten into a car, tuned to your favorite radio station, and then realized . . . it isn't the same station? It seems like overnight all of the on-air personalities and even the genre of music changed! Why does this happen? While there may be other factors, chances are your favorite radio station didn't have very high ratings. In the United States today there are over 10,000 AM and FM commercial radio stations. These stations earn most of their money from advertisements. Advertisers pay the radio stations to play their commercials in the hopes that people will hear the commercials and buy whatever they are selling. Companies want to make sure that there are many people listening to their commercials but if the ratings for a station are low, that means fewer people are listening. Companies will not give money to the station if the ratings are too low so, unfortunately for you, the radio station as you know it may be cancelled and something new will come on in its place in the hopes of attracting new listeners. Luckily today we have so many radio stations that you will probably be able to find a new favorite!

PROBLEM 1 Sifting Through Data to Make a Point

Andres is a new radio station general manager at KYWN. The radio station unfortunately has sagging ratings and low advertising. He has been charged with making the station more popular in the hopes that with more popularity, more companies will want to advertise on KYWN. The station owners have given him the authority to do anything to turn around the ratings; however, if things don't change, he'll be doing overnight radio in the Gobi Desert!

Andres is considering changing the genre of the radio station. Currently, the station features country music. However, if he changes the genre, what will the new genre be?

Andres wants to target one of the highest demographics in radio listening: teenagers. He decides to sponsor the next dance at Rawlings High School. KYWN will provide the food, drinks, and most importantly, the music for the dance. Prior to the dance, he surveys the students. He will use this data to determine the new genre of KYWN.

	Music Genre				
Grade Level	**Rock (Classic/ Alternative)**	**Classical**	**Hip-Hop/Rap**	**Dance**	**Country**
12				x	
9		x			
10				x	
10					x
9					x
11			x		
12		x			
10	x				
9					x
9				x	
10		x			
12				x	
11	x				
12				x	
11			x		
9	x				
9	x				
10					x
11	x				
9	x				
12				x	
12			x		
11			x		
10		x			

(*Continued*)

Music Genre					
Grade Level	Rock (Classic/ Alternative)	Classical	Hip-Hop/Rap	Dance	Country
9			x		
12			x		
11				x	
9			x		
10	x				
10	x				
12					x
9				x	
9				x	
9					x
10		x			
12	x				
12	x				
12		x			
10		x			
10			x		
10				x	
11			x		
9				x	
9					x
10		x			
10		x			
12				x	
11	x				
12	x				
11	x				
11		x			
12				x	
12				x	
12		x			
11			x		
11					x

1. Analyze the data Andres collected.

   a. In looking at the data, can you determine which music genre is most preferred by the Rawlings High School students?

   b. How would you advise Andres to organize the data he gathered?

2. Organize the data to help Andres determine which music genre is most popular at Rawlings High School according to the survey he conducted.

10

3. Analyze the table you created to organize Andres's data.

   a. How many students did Andres survey for the dance? How did you determine that you organized Andres data correctly?

   b. Can you determine which genre of music was the most popular from the table you created? Explain why or why not.

c. Do you think the results might be the same or different if Andres conducted another random survey at Rawlings High School? Explain your reasoning.

d. Based on the data you have analyzed, would you advise Andres to change the format of his station? If so, explain why. If not, explain why not. You may use graphs to better represent your advice to Andres.

PROBLEM 2 What if We Look at the Data *This* Way?

Data can be analyzed in many different ways, and then used to prove a point of an advertiser, a doctor, a scientist, or just about any occupation you can think of! How that data are interpreted can affect the decisions of people—for better or for worse.

Andres knows that one of the most sought after age groups is the age 18 to 35 group. Generally, if stations, magazines, blogs, and news websites can target and successfully attract viewers or users in this age range, they can then lure advertisers to buy air time or banners.

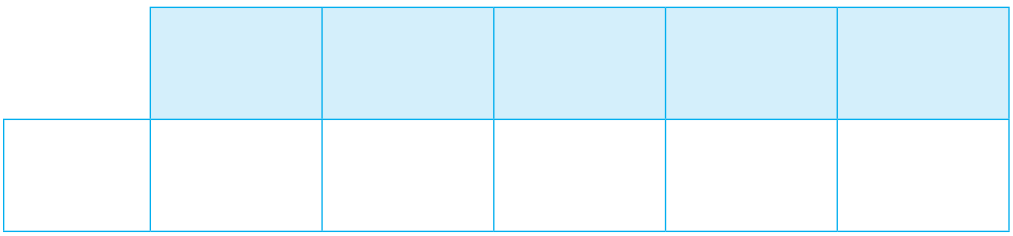
Knowing this, Andres decides to just use the data he gathered from the seniors he surveyed at Rawlings High School.

1. How could Andres use the table or graph(s) you created in Problem 1 to analyze the data for the seniors?

2. Analyze the tables and graphs you created.
   a. Do you think you can predict which music genre is most popular with the seniors at Rawlings High School using the data Andres collected? Explain why or why not.

   b. Predict (if possible) which music genre is the most popular for the Rawlings High School seniors. Explain how you came to your conclusion.

3. Suppose Andres decides to suggest a music format change for KYWN to dance music. What information would you advise Andres to use to strengthen his suggestion? You can use any of the data and can supply any graphs you think that will strengthen Andres's suggestion. Finally, explain why you chose the information.



4. Based on the information you analyzed regarding the seniors, would you change the advice you gave to Andres in Problem 1, Question 3, part (c)? Would you change KYWN's music format to match the Rawlings High School seniors' survey results? If yes, explain why using the data you have. If not, explain why not.

Be prepared to share your solutions and methods.

Chapter 10 Summary

KEY TERMS

- categorical data (10.1)
- two-way frequency table (10.1)
- frequency distribution (10.1)
- joint frequency (10.1)
- frequency marginal distribution (10.1)
- relative frequency distribution (10.2)
- relative frequency marginal distribution (10.2)
- relative frequency conditional distribution (10.3)

10.1 Creating a Two-Way Frequency Table to Analyze Frequency Distribution

Data that can be grouped into categories, such as favorite foods, are called categorical data. One method of organizing categorical data is in a two-way frequency table. A two-way frequency table displays categorical data by representing the number of occurrences that fall into each group for two variables. On the table, one variable is divided into rows and the other is divided into columns. A frequency distribution displays the frequencies for categorical data in a two-way table. Any frequencies recorded within the body of a two-way frequency table are known as joint frequencies.

Example

		Favorite Colors			
		Blue	**Red**	**Yellow**	**Green**
Gender	**Girls**	~~////~~ // 7	// 2	//// 4	/// 3
	Boys	// 2	~~////~~ /// 8	/// 3	/ 1

More girls liked the color blue than any other option and fewer girls liked red.

More boys liked red than any other option and fewer boys like green.

10.1 Creating and Analyzing a Frequency Marginal Distribution

A frequency marginal distribution displays the total of the frequencies of the rows or columns of a frequency distribution.

Example

		Favorite Colors				
		Blue	**Red**	**Yellow**	**Green**	**Total**
Gender	**Girls**	7	2	4	3	16
	Boys	2	8	3	1	14
	Total	9	10	7	4	30

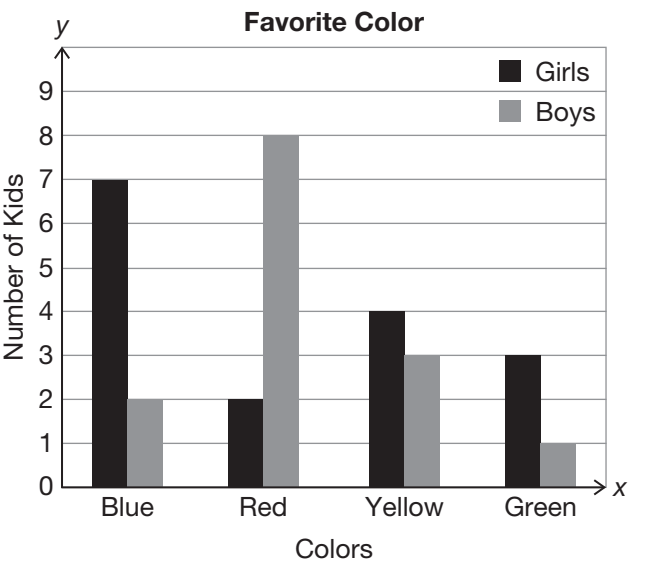
Thirty children participated in the survey. Ten children liked red best; it is the most popular favorite color of those polled. Four children liked green best; it is the least popular favorite color of those polled.

10

10.1 Graphing and Interpreting Graphs of Frequency Distributions

A graph can help relay information from a two-way frequency table in a visual way. A bar graph, a double bar graph, or a stacked bar graph are all good choices for displaying this type of data. Remember a key is necessary to identify what each bar represents.

Example



Red appears to be the most preferred by boys and blue is the most preferred by girls. Green appears to be the least favorite overall.

10.2 Creating and Analyzing Relative Frequency Distribution and Relative Frequency Marginal Distribution

A relative frequency is the ratio or percent of occurrences within a category to the total of the category. Representing the relative frequencies for joint data displayed in a two-way table is called a relative frequency distribution. The relative frequency distribution provides the ratio of occurrences in each category to the total number of occurrences. Displaying the relative frequencies for the rows or columns is called a relative frequency marginal distribution. The relative frequency marginal distribution provides the ratio of total occurrences for each category to the total number of occurrences.

Example

	Preferred Movie Genre				
	Animation	**Comedy**	**Drama**	**Horror**	**Total**
Movie Viewers ages 8 years and younger	$\frac{40}{240} \approx 16.7\%$	$\frac{18}{240} = 7.5\%$	$\frac{2}{240} \approx 0.8\%$	$\frac{0}{240} = 0\%$	$\frac{60}{240} = 25\%$
Movie Viewers ages 9 thru 16 years old	$\frac{20}{240} \approx 8.3\%$	$\frac{42}{240} = 17.5\%$	$\frac{13}{240} \approx 5.4\%$	$\frac{5}{240} \approx 2.1\%$	$\frac{80}{240} \approx 33.3\%$
Movie Viewers ages 17 and older	$\frac{5}{240} \approx 2.1\%$	$\frac{35}{240} \approx 14.6\%$	$\frac{32}{240} \approx 13.3\%$	$\frac{28}{240} \approx 11.7\%$	$\frac{100}{240} \approx 41.7\%$
Total	$\frac{65}{240} \approx 27.1\%$	$\frac{95}{240} \approx 39.6\%$	$\frac{47}{240} \approx 19.5\%$	$\frac{33}{240} \approx 13.8\%$	$\frac{240}{240} = 100\%$

Viewers ages 8 years and younger made up the smallest percent of participants.

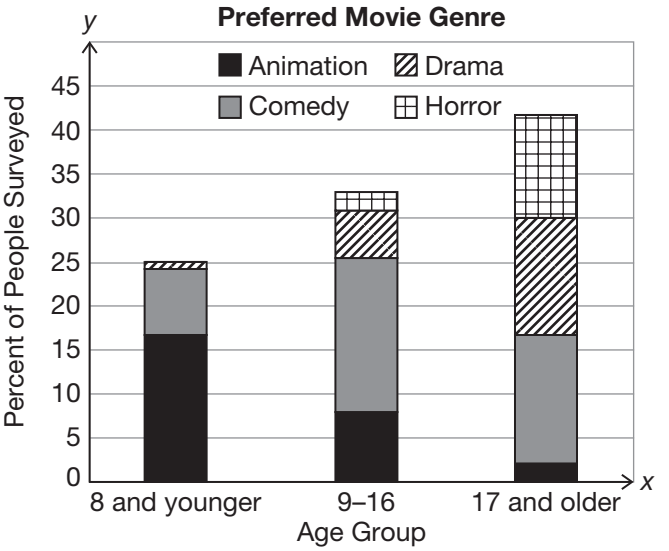
Horror movies are the least popular type of movie overall.

Comedies are preferred by 39.6% of participants.

10.2 Graphing and Interpreting Graphs of Relative Frequency Distributions

A graph can help visually represent data. Use a stacked bar graph to represent relative frequency distributions. A stacked bar graph is a graph in which the bars are stacked on top of each other as opposed to sitting next to each other.

Example



Animation appears to be the favorite genre of participants ages 8 and younger.

Comedy appears to be the favorite genre of participants ages 9 to 16.

Comedy, drama, and horror seem to be fairly evenly favored for participants ages 17 and older.

10.3 Creating and Analyzing a Relative Frequency Conditional Distribution

A relative frequency conditional distribution is the percent of proportion or occurrences of a category given the specific value of another category.

Example

	Preferred Movie Genre				
	Animation	**Comedy**	**Drama**	**Horror**	**Total**
Movie Viewers age 0–8	$\frac{40}{60} \approx 66.7\%$	$\frac{18}{60} = 30\%$	$\frac{2}{60} \approx 3.3\%$	$\frac{0}{60} = 0\%$	$\frac{60}{60} = 100\%$
Movie Viewers age 9–16	$\frac{20}{80} = 25\%$	$\frac{42}{80} = 52.5\%$	$\frac{13}{80} = 16.25\%$	$\frac{5}{80} = 6.25\%$	$\frac{80}{80} = 100\%$
Movie Viewers age 17+	$\frac{5}{100} = 5\%$	$\frac{35}{100} = 35\%$	$\frac{32}{100} = 32\%$	$\frac{28}{100} = 28\%$	$\frac{100}{100} = 100\%$

Of participants ages 8 and younger, only 3.3% preferred dramas.

Of participants ages 9 to 16, 52.5% preferred comedies.

Of participants ages 17 and older, only 5% preferred animated films.

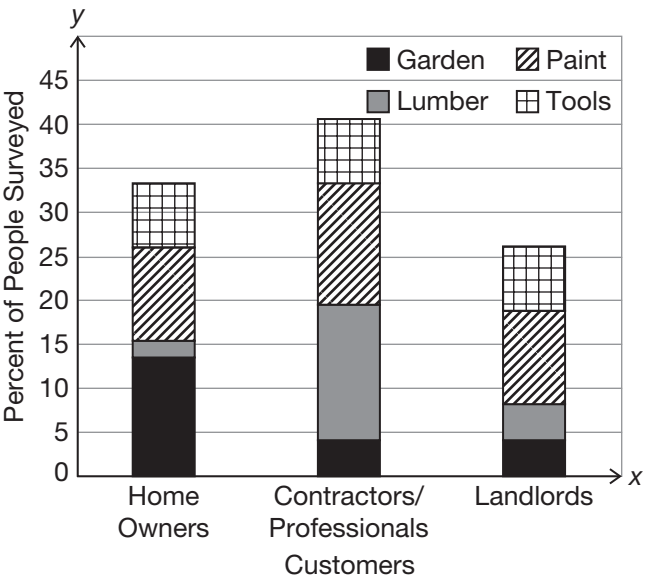
10.4 Drawing Conclusions from Data

Raw data can be organized into a relative frequency marginal distribution table and graph. To further examine the trends within certain categories instead of the overall group, a relative frequency conditional distribution table and graph can be used.

Example

A hardware store chain collected some data on the departments within their store that are most frequented by different types of customers so they could target their advertising.

	Garden	**Lumber**	**Paint**	**Tools**	**Total**
Home Owners	$\frac{83}{615} \approx 13.5\%$	$\frac{12}{615} \approx 2.0\%$	$\frac{65}{615} \approx 10.6\%$	$\frac{45}{615} \approx 7.3\%$	$\frac{205}{615} \approx 33.3\%$
Contractors/ Professionals	$\frac{25}{615} \approx 4.1\%$	$\frac{95}{615} \approx 15.4\%$	$\frac{85}{615} \approx 13.8\%$	$\frac{45}{615} \approx 7.3\%$	$\frac{250}{615} \approx 40.7\%$
Landlords	$\frac{25}{615} \approx 4.1\%$	$\frac{25}{615} \approx 4.1\%$	$\frac{65}{615} \approx 10.6\%$	$\frac{45}{615} \approx 7.3\%$	$\frac{160}{615} \approx 26.0\%$
Total	$\frac{133}{615} \approx 21.7\%$	$\frac{132}{615} \approx 21.4\%$	$\frac{215}{615} \approx 35.0\%$	$\frac{135}{615} \approx 22.0\%$	$\frac{615}{615} = 100\%$



10

From the relative frequency marginal distribution shown, it looks like most customers frequent the Paint department, so the store should focus their advertising on the paints and paint services.

The CEO of the company feels that contractors and landlords are more likely to have wholesale connections, so she wants to focus their advertising on home owners.

	Garden	Lumber	Paint	Tools	Total
Home Owners	$\frac{83}{205} \approx 40.5\%$	$\frac{12}{205} \approx 5.9\%$	$\frac{65}{205} \approx 31.7\%$	$\frac{45}{205} \approx 22.0\%$	$\frac{205}{205} \approx 100\%$

From the relative frequency conditional distribution shown, most home owners frequent the Garden department, so the store decides to focus their advertising on the Garden department.