

Pick Up the warm up

① Deaths among young people Among persons aged 15 to 24 years in the United States, the leading causes of death and number of deaths in a recent year were as follows: accidents, 12,015; homicide, 4651; suicide, 4559; cancer, 1594; heart disease, 984; congenital defects, 401.¹¹

What is the one piece of information needed to make a pie graph?

2) Frequent superpower? An on-line survey asked which superpower high-school students would choose to have—fly, freeze time, invisibility, super strength, or telepathy. Here are the responses from the 40 randomly selected students in the sample:

Fly Freeze time Telepathy Fly Telepathy Super strength Telepathy Telepathy Fly
 Super strength Invisibility Freeze time Fly Telepathy Freeze time Telepathy Super strength Fly
 Freeze time Telepathy Freeze time Freeze time Freeze time Fly Fly Freeze time
 Invisibility Fly Invisibility Telepathy Telepathy Fly Telepathy Fly Fly Telepathy Telepathy Fly

*make an appropriate graph to display the distribution of superpower preference. Describe what you see.

	Freq
Fly	14
Freeze time	8
Invisibility	3
super strength	3
telepathy	12

2) Frequent superpower? An on-line survey asked which superpower high-school students would choose to have—fly, freeze time, invisibility, super strength, or telepathy. Here are the responses from the 40 randomly selected students in the sample:

Fly Freeze time Telepathy Fly Telepathy Super strength Telepathy Telepathy Fly
 Super strength Invisibility Freeze time Fly Telepathy Freeze time Telepathy Super strength Fly
 Freeze time Telepathy Freeze time Freeze time Freeze time Fly Fly Freeze time
 Invisibility Fly Invisibility Telepathy Telepathy Fly Telepathy Fly Fly Telepathy Telepathy Fly

*make an appropriate graph to display the distribution of superpower preference. Describe what you see.

	Freq
Fly	14
Freeze time	8
Invisibility	3
super strength	3
telepathy	12

2) Frequent superpower? An on-line survey asked which superpower high-school students would choose to have—fly, freeze time, invisibility, super strength, or telepathy. Here are the responses from the 40 randomly selected students in the sample:

Fly Freeze time Telepathy Fly Telepathy Super strength Telepathy Telepathy Fly
 Super strength Invisibility Freeze time Fly Telepathy Freeze time Telepathy Super strength Fly
 Freeze time Telepathy Freeze time Freeze time Freeze time Fly Fly Freeze time
 Invisibility Fly Invisibility Telepathy Telepathy Fly Telepathy Fly Fly Telepathy Telepathy Fly

*make an appropriate graph to display the distribution of superpower preference. Describe what you see.

	Freq
Fly	14
Freeze time	8
Invisibility	3
super strength	3
telepathy	12

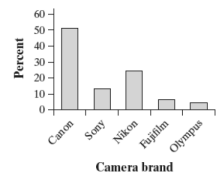
Were you aware that there is a video explanation 2 of the problems from the assignment (19 and 23)?

Symbol next to problem

1.13 First, a relative frequency table must be constructed.

The relative frequency bar graph is given below.

Camera brand	Relative Frequency
Canon	$23/45 = 0.511 = 51.1\%$
Sony	$6/45 = 0.133 = 13.3\%$
Nikon	$11/45 = 0.244 = 24.4\%$
Fujifilm	$3/45 = 0.067 = 6.7\%$
Olympus	$2/45 = 0.044 = 4.4\%$

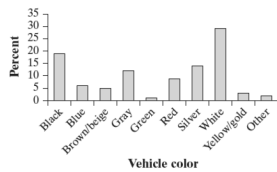


The most popular brand of camera among the 45 most recent purchases on the Internet auction site is Canon, followed by Nikon, Sony, Fujifilm, and Olympus.

Canon is the overwhelming favorite with over 50% of the customers purchasing this brand. Also noteworthy is that almost 25% of the customers purchased a Nikon camera.

1.15

- (a) The percent of cars with other colors is $100 - 19 - 6 - 5 - 12 - 1 - 9 - 14 - 29 - 3 = 2\%$.
 (b) A bar graph is given below.



The most popular color of vehicles sold that year was white, followed by black, silver, and gray. It appears that a majority of car buyers that year preferred vehicles that were shades of black and white.

- (c) It would be appropriate to make a pie chart of these data (including the other category) because the numbers in the table refer to parts of a single whole.

1.17

Estimates will vary, but should be close to 63% Mexican and 9% Puerto Rican.

1.19

The areas of the pictures should be proportional to the numbers of students they represent. As drawn, it appears that most of the students arrived by car but in reality, most came by bus (14 took the bus, 9 came in cars).

1.21

By starting the vertical scale at 12 instead of 0, it looks like the percent of binge-watchers who think that 5 to 6 episodes is too many to watch in one viewing session is almost 20 times higher than the percent of binge-watchers who think that 3 to 4 episodes is too many to watch in one viewing session. In truth, the percent of binge-watchers who think that 5 to 6 episodes is too many to watch in one viewing session (31%) is less than three times higher than the percent of binge-watchers who think that 3 to 4 episodes is too many to watch in one viewing session (13%). Similar arguments can be made for the relative sizes of the other categories represented in the bar graph.

(a) What proportion of subjects were given the control treatment?

Response	Treatment			Total
	"Smashed into"	"Hit"	Control	
Yes	16	7	6	29
No	34	43	44	121
Total	50	50	50	150

1.23

- (a) $50/150 = 0.333$. One-third of the 150 subjects were given the control treatment.
 (b) 10.7% said they saw broken glass at the accident; 89.3% said they did not; 14% said they saw broken glass at the accident.
 (c) Sixteen of the 150 subjects, or 10.67%, were given the "smashed into" treatment and said they saw broken glass at the accident.

(b) Find the distribution of responses about whether there was broken glass at the accident for the subjects in this study using relative frequencies.

Response	Treatment			Total
	"Smashed into"	"Hit"	Control	
Yes	16	7	6	29
No	34	43	44	121
Total	50	50	50	150

Out of the total number of subjects,

$$\frac{29}{150} = 0.193 = 19.3\% \text{ of the subjects said they saw glass}$$

$$\frac{121}{150} = 0.807 = 80.7\% \text{ of the subjects said they didn't see glass}$$

(c) What percent of the subjects were given the "smashed into" treatment and said they saw broken glass at the accident?

Response	Treatment			Total
	"Smashed into"	"Hit"	Control	
Yes	16	7	6	29
No	34	43	44	121
Total	50	50	50	150

Out of the total number of subjects,

$\frac{16}{150} = 0.107 = 10.7\%$ of the subjects were given the "Smashed into" treatment and said they saw glass.

Section 1.1 Continued
Corresponds to pp. 17-22
class notes

LEARNING TARGETS

By the end of this section, you should be able to:

- ✓ MAKE and INTERPRET bar graphs for categorical data.
- ✓ IDENTIFY what makes some graphs of categorical data misleading.
- ✓ CALCULATE marginal and joint relative frequencies from a two-way table.
- ✓ CALCULATE conditional relative frequencies from a two-way table.
- ✓ Use bar graphs to COMPARE distributions of categorical data.
- ✓ DESCRIBE the nature of the association between two categorical variables.

Marginal and joint relative frequencies do not tell us much about the relationship between environmental club membership and snowmobile use for the people in the sample.

		Environmental club		Total
		No	Yes	
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
Total		1221	305	1526

Marginal and joint relative frequencies do not tell us much about the relationship between environmental club membership and snowmobile use for the people in the sample.

		Environmental club		Total
		No	Yes	
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
Total		1221	305	1526

A **conditional relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition).

Marginal and joint relative frequencies do not tell us much about the relationship between environmental club membership and snowmobile use for the people in the sample.

		Environmental club		Total
		No	Yes	
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
Total		1221	305	1526

A **conditional relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition).

What percent of environmental club members in the sample are snowmobile owners?

Marginal and joint relative frequencies do not tell us much about the *relationship* between environmental club membership and snowmobile use for the people in the sample.

A **conditional relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition).

Snowmobile use	Environmental club			Total
	No	Yes		
Never used	445	212		657
Snowmobile renter	497	77		574
Snowmobile owner	279	16		295
Total	1221	305		1526

What percent of environmental club members in the sample are snowmobile owners?

tell us much about the *relationship* between environmental club membership and snowmobile use for the people in the sample.

frequency gives the percent or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition).

Snowmobile use	Environmental club			Total
	No	Yes		
Never used	445	212		657
Snowmobile renter	497	77		574
Snowmobile owner	279	16		295
Total	1221	305		1526

What percent of environmental club members in the sample are snowmobile owners?

$$\frac{16}{305} = 0.052 = 5.2\%$$

Snowmobile use	Environmental club			Total
	No	Yes		
Never used	445	212		657
Snowmobile renter	497	77		574
Snowmobile owner	279	16		295
Total	1221	305		1526

The distribution of snowmobile use among environmental club members is called the **conditional distribution** of snowmobile use among environmental club members.

Snowmobile use	Environmental club			Total
	No	Yes		
Never used	445	212		657
Snowmobile renter	497	77		574
Snowmobile owner	279	16		295
Total	1221	305		1526

Never: $\frac{212}{305} = 0.695$ or 69.5%
 Rent: $\frac{77}{305} = 0.252$ or 25.2%
 Own: $\frac{16}{305} = 0.052$ or 5.2%

The distribution of snowmobile use among environmental club members is called the **conditional distribution** of snowmobile use among environmental club members.

Snowmobile use	Environmental club			Total
	No	Yes		
Never used	445	212		657
Snowmobile renter	497	77		574
Snowmobile owner	279	16		295
Total	1221	305		1526

We can find the distribution of snowmobile use among the survey respondents who are not environmental club members in a similar way.

Snowmobile use	Not environmental club members	Environmental club members
Never	$\frac{445}{1221} = 0.364$ or 36.4%	$\frac{212}{305} = 0.695$ or 69.5%
Rent	$\frac{497}{1221} = 0.407$ or 40.7%	$\frac{77}{305} = 0.252$ or 25.2%
Own	$\frac{279}{1221} = 0.229$ or 22.9%	$\frac{16}{305} = 0.052$ or 5.2%

AP® Exam Tip

- ✓ When comparing groups of different sizes, be sure to use **relative frequencies** (percents or proportions) instead of **frequencies** (counts) when analyzing categorical data.
- ✓ Make sure to avoid statements like "More club members never use snowmobiles" when you mean "A greater percentage of club members never use snowmobiles."

2. The Pew Research Center asked a random sample of 2024 adult cell phone owners from the United States which type of cell phone they own: iPhone, Android, or other (including non-smart phones). Here are the results, broken down by age category:

	18-34	35-54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

	18-34	35-54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

- (a) What proportion of the sample use an iPhone?
 $467/2024 = 0.23$
- (b) What proportion of the sample use an iPhone and are 55+?
 $127/2024 = 0.063$
- (c) What proportion of the 55+ people in the sample use an iPhone?
 $127/870 = 0.146$
- (d) What proportion of the iPhone users in the sample are 55+?
 $127/467 = 0.272$

Association

There is an **association** between two variables if knowing the value of one variable helps us predict the value of the other.

4. How can you “see” an association between two categorical variables?

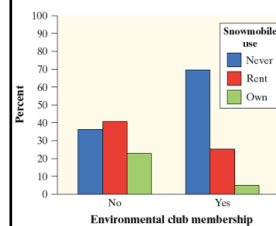
Make a side by side OR segmented bar chart

(showing the distribution of the response variable for each category of the explanatory variable.

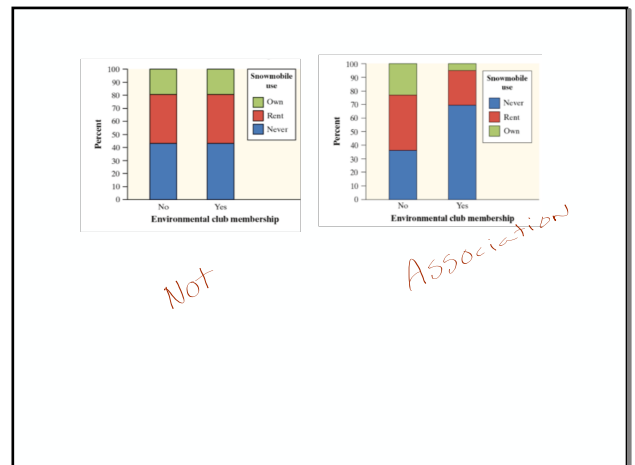
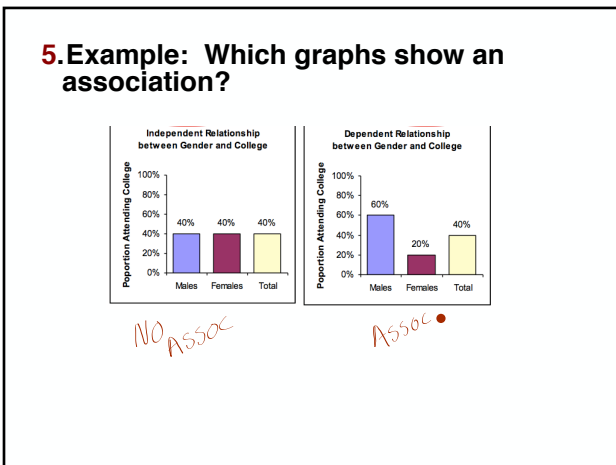
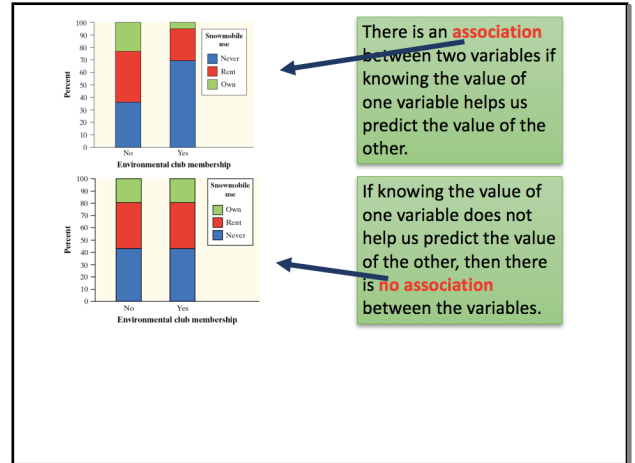
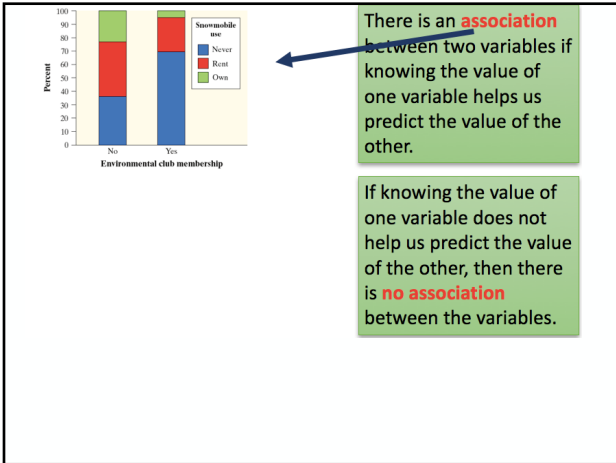
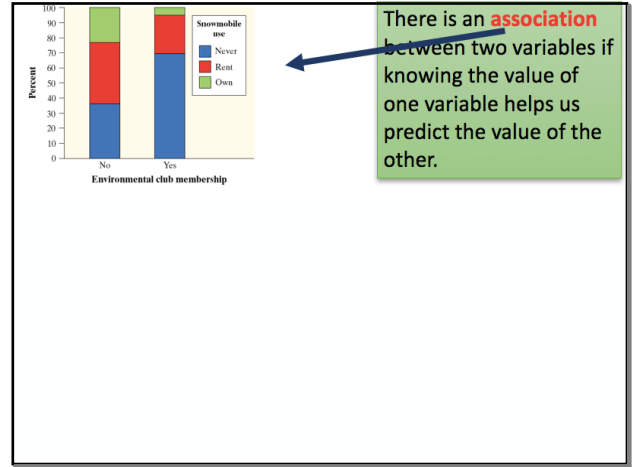
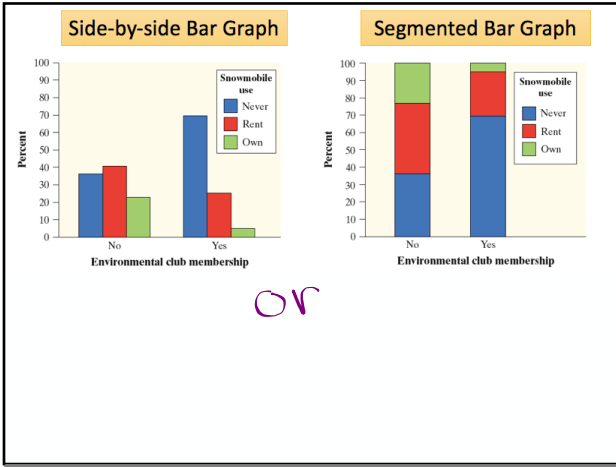
A **side-by-side bar graph** displays the distribution of a categorical variable for each value of another categorical variable. The bars are grouped together based on the values of one of the categorical variables and placed side by side.

A **segmented bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.

Side-by-side Bar Graph



A **segmented bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.



Relationships Between Two Categorical Variables

CAUTION!
Association does not necessarily imply causation!

EXAMPLE A Titanic Disaster
In 1912 the luxury liner *Titanic*, on its first voyage across the Atlantic, struck an iceberg and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table below gives information about adult passengers who lived and who died, by class of travel.

Class of Travel	Survival status	
	Survived	Died
First Class	197	122
Second Class	94	167
Third Class	151	476

(a) Find the **marginal distribution** of survival status. How could you display this distribution?

(b) Find the **marginal distribution** of class of travel. How could you display this distribution?

EXAMPLE A Titanic Disaster
In 1912 the luxury liner *Titanic*, on its first voyage across the Atlantic, struck an iceberg and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table below gives information about adult passengers who lived and who died, by class of travel.

Class of Travel	Survival status	
	Survived	Died
First Class	197	122
Second Class	94	167
Third Class	151	476

(a) Find the **marginal distribution** of survival status. How could you display this distribution?

First: $\frac{412}{1207} = .336$
Second: $\frac{765}{1207} = .634$

(b) Find the **marginal distribution** of class of travel. How could you display this distribution?

First: $\frac{319}{1207} = .264$
Second: $\frac{261}{1207} = .216$
Third: $\frac{627}{1207} = .518$

Class of Travel	Survival status	
	Survived	Died
First Class	197	122
Second Class	94	167
Third Class	151	476

442 765

(c) The movie *Titanic* with Leonardo DiCaprio and Kate Winslett suggested that first-class passengers received special treatment in boarding the lifeboats, while some other passengers were prevented from doing so (especially third-class passengers). Is there an **association** between class of travel and survival status? Use an appropriate **conditional distribution** to help answer this question.

$\frac{197}{442} =$ $\frac{94}{442} =$ $\frac{151}{442} =$
 $\frac{122}{765} =$ $\frac{167}{765} =$ $\frac{476}{765} =$

LEARNING TARGETS

After this section, you should be able to:

- ✓ MAKE and INTERPRET bar graphs for categorical data.
- ✓ IDENTIFY what makes some graphs of categorical data misleading.
- ✓ CALCULATE marginal and joint relative frequencies from a two-way table.
- ✓ CALCULATE conditional relative frequencies from a two-way table.
- ✓ Use bar graphs to COMPARE distributions of categorical data.
- ✓ DESCRIBE the nature of the association between two categorical variables.

B.B.
Moving Graphs

Reminder

We study/cover/investigate most concepts and skills in class, BUT NOT everything.

You are responsible for reading/studying the sections in the textbook.

Assignment

1.1 ... 27, 29, 33,
35, 40-43