Pick Up
the Warm Up

Pull out the last sheet from yesterday's handout, specifically #2

* Note: The answer to 236 in the book is incorrect.

1 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

(a) What proportion of the sample use an iPhone?

(b) What proportion of the sample use an iPhone and are 55+?

(c) What proportion of the 55+ people in the sample use an iPhone?

(d) What proportion of the iPhone users in the sample are 55+?

1 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

(a) What proportion of the sample use an iPhone?

Marginal relative freq $\frac{467}{2024} = .23$

(b) What proportion of the sample use an iPhone and are 55+?

joint relative freq $\frac{127}{2024} = .06$

(c) What proportion of the 55+ people in the sample use an iPhone?

cond rel freq $\frac{127}{870} = .15$

(d) What proportion of the iPhone users in the sample are 55+?

$\frac{127}{467} = .27$

2. Which question above is asking for *conditional relative frequency*?

Joint relative frequency? b

Marginal relative frequency? a

c and d

3. What does it mean for two variables to have an association?

Association

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

4. What can you make that allows you to "see" if there is an association between two categorical variables?

Side by Side bar graphs
or
Segmented bar graphs
or
Mosaic Plots

Make a side by side OR segmented bar chart

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

FROM Yesterday's handout

2. An article in the Journal of the American Medical Association reports the results of a study designed to see if the herb St. John's wort is effective in treating moderately severe cases of depression. The study involved 338 patients who were being treated for major depression. The subjects were randomly assigned to receive one of three treatments: St. John's wort, Zoloft (a prescription drug), or placebo (an inactive treatment) for an 8-week period. The two way table summarizes the data from the experiment.

		Treatment		
		St. John's wort	Zoloft	Placebo
Change in depression	Full response	27	27	37
	Partial response	16	26	13
	No response	70	56	66

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

b. Find the distribution of change in depression for the subjects in this study using relative frequencies.

		Treatment		
		St. John's wort	Zoloft	Placebo
Change in depression	Full response	27	27	37
	Partial response	16	26	13
	No response	70	56	66

c. What percent of subjects took Zoloft and showed a full response?

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

b. Find the distribution of change in depression for the subjects in this study using relative frequencies.

		Treatment			
		St. John's wort	Zoloft	Placebo	
Change in depression	Full response	27	27	37	91
	Partial response	16	26	13	55
	No response	70	56	66	192
		113	109	116	338

c. What percent of subjects took Zoloft and showed a full response?

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

b. Find the distribution of change in depression for the subjects in this study using relative frequencies.

St. Johns $\frac{27}{113}$

$\frac{16}{113}$

$\frac{70}{113}$

c. What percent of subjects took Zoloft and showed a full response?

$\frac{27}{338} = .08 = 8\%$

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

b. Find the distribution of change in depression for the subjects in this study using relative frequencies.

St. Johns 27/113 Zoloft 27/109
16/113 26/109
70/113 56/109

etc.

	Treatment			
	St. John's wort	Zoloft	Placebo	
Change in depression: Full response	27	27	37	91
Partial response	16	26	13	55
No response	70	56	66	192
	113	109	116	338

c. What percent of subjects took Zoloft and showed a full response?

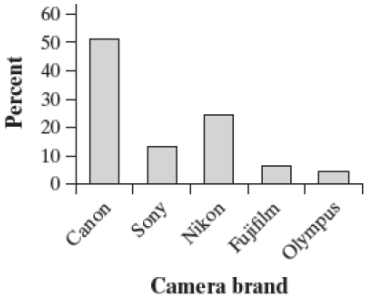
$\frac{27}{338} = .08 = 8\%$

13

1.13
First, a relative frequency table must be constructed.

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%

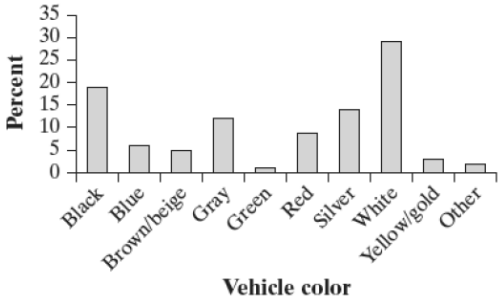
The relative frequency bar graph is given below.



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?

	Treatment			
Response		"Smashed into"	"Hit"	Control	Total
	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.

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

Out of the total number of subjects,

$^{29}/_{150}$ = 0.193 = 19.3% of the subjects said they saw glass

$^{121}/_{150}$ = 0.807 = 80.7% 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?

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

Out of the total number of subjects,

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

Today: **watch a few things that will lead us up to Mosaic Plots.**

You will receive some notes later.

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

		Environmental club		
		No	Yes	Total
Snowmobile use	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: No	Yes	Total
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%

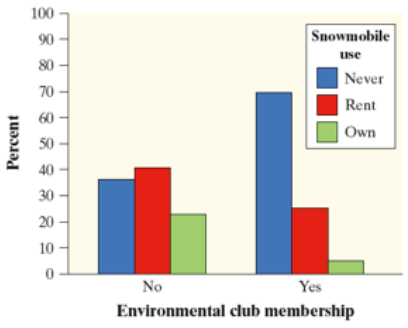
The distribution of snowmobile use among environmental club members is called the **conditional distribution** of snowmobile use among environmental club members.

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

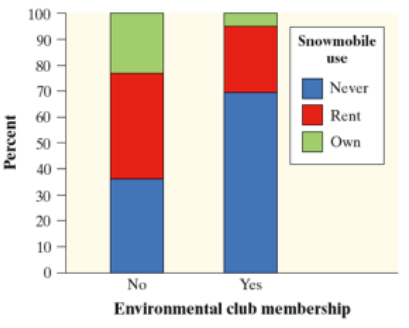
We can find the distribution of snowmobile use among the survey respondents who are not environmental club members in a similar way.

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%

Side-by-side Bar Graph



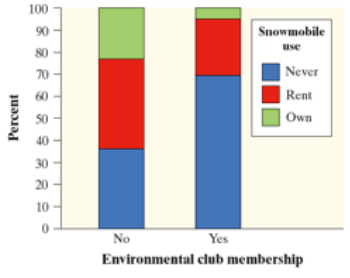
Segmented Bar Graph



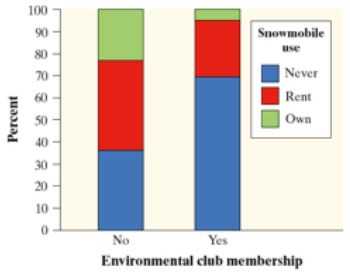
There is an **association** between two variables if knowing the value of one variable helps us predict the value of the other.

CAUTION:
Association does not necessarily imply causation!

sociation
ariables if
lue of
ps us
of the

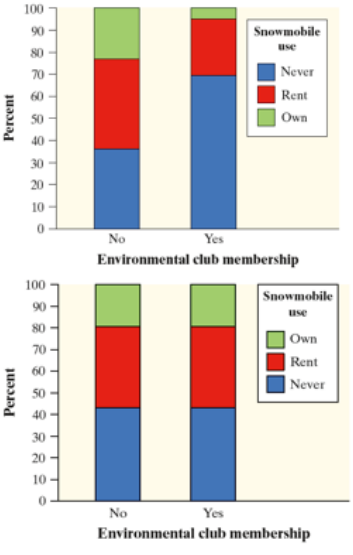


There is an **association** between two variables if knowing the value of one variable helps us predict the value of the other.



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

If knowing the value of one variable does not help us predict the value of the other, then there is **no association** between the variables.



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

If knowing the value of one variable does not help us predict the value of the other, then there is **no association** between the variables.

There is a modified version of a segmented bar graph that actually makes more sense. It is called a **Mosaic Plot**.

Notice the overall membership size of the No and Yes Groups. They are quite different yet the segmented bar graphs would suggest they are about the same.

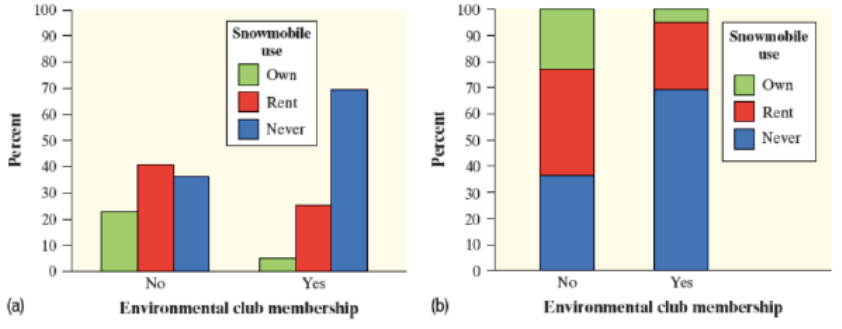
Snowmobile use	Environmental club		
	No	Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

Now you can Pick up the notes about Mosaic Plots

Yellowstone National Park staff surveyed a random sample of 1526 winter visitors to the park. They asked each person whether he or she belonged to an environmental club (like the Sierra Club). Respondents were also asked whether they owned, rented, or had never used a snowmobile. Here is a two-way table summarizing the results.

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

Figure 1.3 compares the distributions of snowmobile use for Yellowstone National Park visitors who are environmental club members and those who are not environmental club members with (a) a **side-by-side bar graph**, (b) a **segmented bar graph**, and (c) a **mosaic plot**. Notice that the segmented bar graph can be obtained by stacking the bars in the side-by-side bar graph for each of the two environmental club membership categories (no and yes). The bar widths in the mosaic plot are proportional to the number of survey respondents who are (305) and are not (1221) environmental club members.



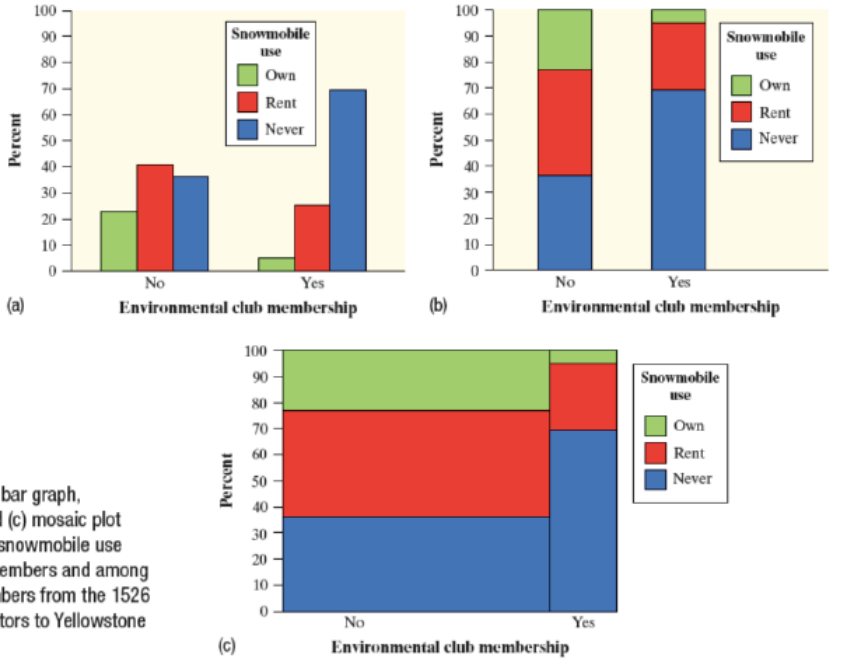
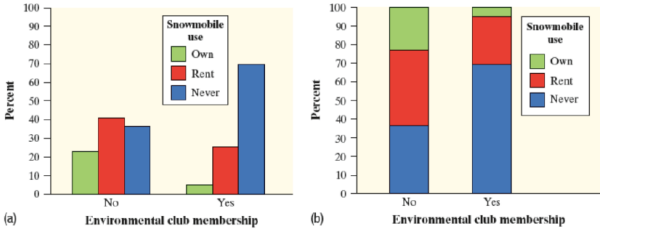


FIGURE 1.3 (a) Side-by-side bar graph, (b) segmented bar graph, and (c) mosaic plot displaying the distribution of snowmobile use among environmental club members and among non-environmental club members from the 1526 randomly selected winter visitors to Yellowstone National Park.

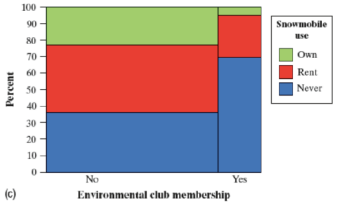
A **mosaic plot** is a modified segmented bar graph in which the width of each rectangle is proportional to the number of individuals in the corresponding category.

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

Notice how much bigger the "No" group is than the "Yes" group.



Yet the bars add up to the same size (100%)



1221 is about 4 times bigger than 305

and the width of the "No" group is about 4 times wider.

Exercise 1

1. Body image A random sample of 1200 U.S. college students was asked, "What is your perception of your own body? Do you feel that you are overweight, underweight, or about right?" The two-way table summarizes the data on perceived body image by gender.

		Gender		
		Female	Male	Total
Body image	About right	560	295	855
	Overweight	163	72	235
	Underweight	37	73	110
	Total	760	440	1200

(a) Of the respondents who felt that their body weight was about right, what proportion were female?

(b) Of the female respondents, what percent felt that their body weight was about right?

1. Body image A random sample of 1200 U.S. college students was asked, "What is your perception of your own body? Do you feel that you are overweight, underweight, or about right?" The two-way table summarizes the data on perceived body image by gender.

		Gender		
		Female	Male	Total
Body image	About right	560	295	855
	Overweight	163	72	235
	Underweight	37	73	110
	Total	760	440	1200

(a) Of the respondents who felt that their body weight was about right, what proportion were female?

$\frac{560}{855} = 0.655$ proportion

65.5% percent

(b) Of the female respondents, what percent felt that their body weight was about right?

1. Body image A random sample of 1200 U.S. college students was asked, "What is your perception of your own body? Do you feel that you are overweight, underweight, or about right?" The two-way table summarizes the data on perceived body image by gender.

		Gender		
		Female	Male	Total
Body image	About right	560	295	855
	Overweight	163	72	235
	Underweight	37	73	110
	Total	760	440	1200

(a) Of the respondents who felt that their body weight was about right, what proportion were female?

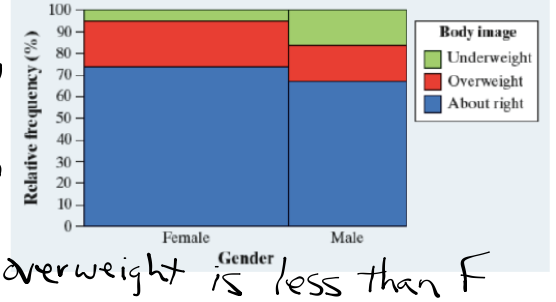
$\frac{560}{855} = 0.655$ proportion

65.5% percent

(b) Of the female respondents, what percent felt that their body weight was about right?

$\frac{560}{760} = .737 = 73.7\%$

(c) The mosaic plot displays the distribution of perceived body image by gender. Describe what this graph reveals about the association between these two variables for the 1200 college students in the sample.



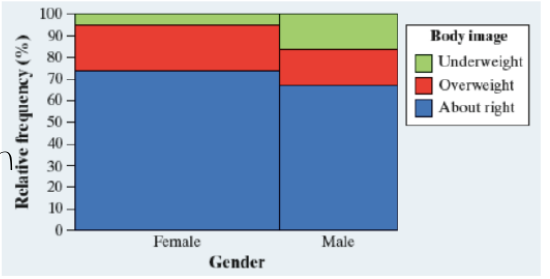
Is an association

More F than males

freq. for men feeling overweight is less than F

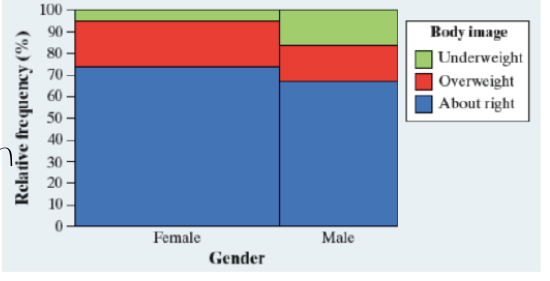
Females less likely to feel underweight.

(c) The mosaic plot displays the distribution of perceived body image by gender. Describe what this graph reveals about the association between these two variables for the 1200 college students in the sample.



Based upon the Mosaic Plot, there is an association between perceived body image and gender.

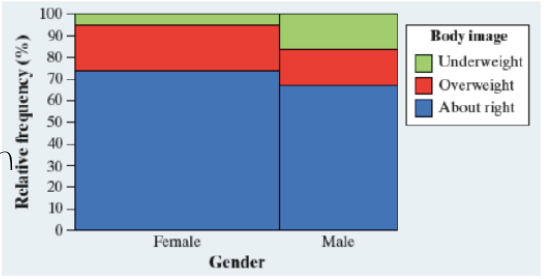
(c) The mosaic plot displays the distribution of perceived body image by gender. Describe what this graph reveals about the association between these two variables for the 1200 college students in the sample.



Based upon the Mosaic Plot, there is an association between perceived body image and gender.

Males are about 4 times as likely as females to perceive that they are under weight. Males are also less likely to perceive they are over weight.

(c) The mosaic plot displays the distribution of perceived body image by gender. Describe what this graph reveals about the association between these two variables for the 1200 college students in the sample.



Based upon the Mosaic Plot, there is an association between perceived body image and gender.

Males are about 4 times as likely as females to perceive that they are under weight. Males are also less likely to perceive they are over weight, however the overwhelming majority of both genders perceive that their body image is about right.

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

Study pp. 17-22