1.1.1 Bar Graphs and Pie Charts

In this case, the *individuals* are the radio stations and the *variable* being measured is the kind of programming that each station broadcasts. The table on the left, which we call a **frequency table,** displays the counts (*frequencies*) of stations in each format category. On the right, we see a **relative frequency table** of the data that shows the percents (*relative frequencies*) of stations in each format category.

Frequency table

Format	Count of stations
Adult contemporary	1556
Adult standards	1196
Contemporary hit	569
Country	2066
News/Talk/Information	2179
Oldies	1060
Religious	2014
Rock	869
Spanish language	750
Other formats	1579
Total	**13,838**

Relative frequency table

Format	Percent of stations
Adult contemporary	11.2
Adult standards	8.6
Contemporary hit	4.1
Country	14.9
News/Talk/Information	15.7
Oldies	7.7
Religious	14.6
Rock	6.3
Spanish language	5.4
Other formats	11.4
Total	**99.9**

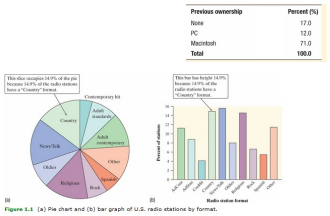
Aug 19-11:50 AM

2 most common graphs for categorical data:

Bar graph - use bars to represent the frequency of each category.

Pie graph - uses a slice of the pie to represent the data. Each slice is determined by taking the number for the category divided by the ***total number*** in the data set x 360°.

*don't make pie graphs if given %, it represents fractions rather than parts of a single whole.



Aug 19-11:56 AM

When creating graphs remember to:

- Label whether it be x and y axis or slices in a pie.
- Title the graph
- Draw it to scale - start with zero

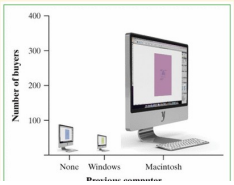
Aug 19-12:04 PM

1.1.2 Graphs: Good and Bad

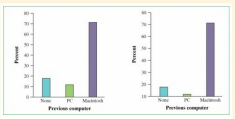
Example 4 Who Buys iMacs?

Beware the pictograph!

(a) Here's a clever graph of the data that uses pictures instead of the more traditional bars. How is this graph misleading?



(b) Two possible bar graphs of the data are shown below. Which one could be considered deceptive? Why?



Aug 19-12:07 PM

1.1.3 Two-Way Tables and Marginal Distributions

Day 2

Example 5 I'm Gonna Be Rich!

Two-way tables

A survey of 4826 randomly selected young adults (aged 19 to 25) asked, "What do you think the chances are you will have much more than a middle-class income at age 30?" The table below shows the responses.[6]

Young adults by gender and chance of getting rich

	Gender		
Opinion	Female	Male	Total
Almost no chance	96	98	**194**
Some chance but probably not	426	286	**712**
A 50-50 chance	696	720	**1416**
A good chance	663	758	**1421**
Almost certain	486	597	**1083**
Total	**2367**	**2459**	**4826**

This is a **two-way table** because it describes two categorical variables, gender and opinion about becoming rich. Opinion is the *row variable* because each row in the table describes young adults who held one of the five opinions about their chances. Because the opinions have a natural order from "Almost no chance" to "Almost certain," the rows are also in this order. Gender is the *column variable*. The entries in the table are the counts of individuals in each opinion-by-gender class.

Aug 19-12:11 PM

DEFINITION: Marginal distribution

The **marginal distribution** of one of the categorical variables in a two-way table of counts is the distribution of values of that variable among all individuals described by the table.

Young adults by gender and chance of getting rich

	Gender		
Opinion	Female	Male	Total
Almost no chance	96	98	194
Some chance but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

Aug 19-12:14 PM

(a) Use the data in the two-way table to calculate the marginal distribution (in percents) of opinions.

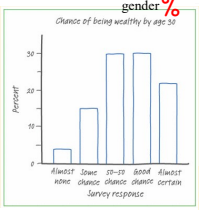
(b) Make a graph to display the marginal distribution. Describe what you see.

Young adults by gender and chance of getting rich

	Gender		
Opinion	Female	Male	Total
Almost no chance	96	98	194
Some chance but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

%

gender %



(b) Figure 1.3 is a bar graph of the distribution of opinion among these young adults. It seems that many young adults are optimistic about their future income. Over 50% of those who responded to the survey felt that they had "a good chance" or were "almost certain" to be rich by age 30.

Aug 19-12:17 PM

CHECK YOUR UNDERSTANDING

	Country	
Superpower	U.K.	U.S.
Fly	54	45
Freeze time	52	44
Invisibility	30	37
Superstrength	20	23
Telepathy	44	66

A random sample of 415 children aged 9 to 17 from the United Kingdom and the United States who completed a CensusAtSchool survey in a recent year was selected. Each student's country of origin was recorded along with which superpower they would most like to have: the ability to fly, ability to freeze time, invisibility, superstrength, or telepathy (ability to read minds). The data are summarized in the table.

1. Use the two-way table to calculate the marginal distribution (in percents) of superpower preferences.

Show Answer

2. Make a graph to display the marginal distribution. Describe what you see.

Show Answer

Aug 19-2:24 PM

1.1.4 Relationships between Categorical Variables: Conditional Distributions

The two-way table contains much more information than the two marginal distributions of opinion alone and gender alone. *Marginal distributions tell us nothing about the relationship between two variables.* To describe a relationship between two categorical variables, we must calculate some well-chosen percents from the counts given in the body of the table.

DEFINITION: Conditional distribution

A **conditional distribution** of a variable describes the values of that variable among individuals who have a specific value of another variable. There is a separate conditional distribution for each value of the other variable.

Young adults by gender and chance of getting rich

	Gender		
Opinion	Female	Male	Total
Almost no chance	96	98	194
Some chance but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

Female %

$\frac{96}{2367} = 4.1\%$

$\frac{426}{2367} = 18.0\%$

$\frac{696}{2367} = 29.4\%$

$\frac{663}{2367} = 28.0\%$

$\frac{486}{2367} = 20.5\%$

Male %

$\frac{98}{2459} = 4.0\%$

$\frac{286}{2459} = 11.6\%$

$\frac{720}{2459} = 29.3\%$

$\frac{758}{2459} = 30.8\%$

$\frac{597}{2459} = 24.3\%$

Aug 19-2:31 PM

We'll make a side-by-side bar graph to compare the opinions of males and females. Figure 1.6 displays the completed graph.

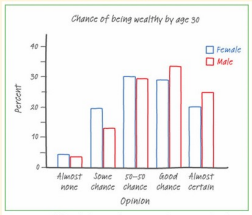


Figure 1.6 Side-by-side bar graph comparing the opinions of males and females.

Based on the sample data, men seem somewhat more optimistic about their future income than women. Men were less likely to say that they have "some chance but probably not" than women (11.6% vs. 18.0%). Men were more likely to say that they have "a good chance" (30.8% vs. 28.0%) or are "almost certain" (24.3% vs. 20.5%) to have much more than a middle-class income by age 30 than women were.

segmented bar graph

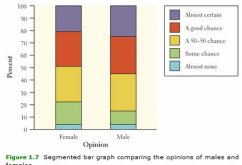


Figure 1.7 Segmented bar graph comparing the opinions of males and females.

Aug 19-2:40 PM

DEFINITION: Association

We say that there is an **association** between two variables if knowing the value of one variable helps predict the value of the other. If knowing the value of one variable does not help you predict the value of the other, then there is no association between the variables.

Both graphs provide evidence of an **association** between gender and opinion about future wealth in this sample of young adults. Men more often rated their chances of becoming rich in the two highest categories; women said "some chance but probably not" much more frequently.

Aug 19-2:45 PM