Organizing Qualitative Data

Some situations generate an overwhelming amount of data. We can often make a large or complicated set of data more compact and easier to understand by organizing it in a table, chart, or graph. In this section, we examine some of the most important ways to organize qualitative data. In the next section, we do that for quantitative data.

Frequency Distributions:

Recall that qualitative data are values of a qualitative (non-numerically valued) variable. One way of organizing qualitative data is to construct a table that gives the number of times each distinct value occurs. The number of times a particular distinct value occurs is called its frequency (or count).

Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

To Construct a Frequency Distribution of Qualitative Data

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.

Step 3 Count the tallies for each distinct value and record the totals in the third column of the table.

Example:

Political Party Affiliations Professor Weiss asked his introductory statistics students to state their political party affiliations as Democratic (D), Republican (R), or Other (O). The responses of the 40 students in the class are given in Table 2.1. Determine a frequency distribution of these data.

Solution We apply Procedure 2.1.

TABLE 2.1

Political party affiliations of the students in introductory statistics

D	R	O	R	R	R	R	R
D	O	R	D	O	O	R	D
D	R	O	D	R	R	O	R
D	O	D	D	D	R	O	D
O	R	D	R	R	R	R	D

Step 1 List the distinct values of the observations in the data set in the first column of a table.

The distinct values of the observations are Democratic, Republican, and Other, which we list in the first column of Table 2.2.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.

The first affiliation listed in Table 2.1 is Democratic, calling for a tally mark in the Democratic row of Table 2.2. The complete results of the tallying procedure are shown in the second column of Table 2.2.

Step 3 Count the tallies for each distinct value and record the totals in the third column of the table.

Counting the tallies in the second column of Table 2.2 gives the frequencies in the third column of Table 2.2. The first and third columns of Table 2.2 provide a frequency distribution for the data in Table 2.1.

TABLE 2.2

Table for constructing a frequency distribution for the political party affiliation data in Table 2.1

Party	Tally	Frequency
Democratic	ЦН ЦН |||	13
Republican	ЦН ЦН ЦН |||	18
Other	ЦН ||||	9
		40

Interpretation From Table 2.2, we see that, of the 40 students in the class, 13 are Democrats, 18 are Republicans, and 9 are Other.

By simply glancing at Table 2.2, we can easily obtain various pieces of useful information. For instance, we see that more students in the class are Republicans than any other political party affiliation.

Relative-Frequency Distributions:

Relative-Frequency Distribution of Qualitative Data

A **relative-frequency distribution** of qualitative data is a listing of the distinct values and their relative frequencies.

In addition to the frequency that a particular distinct value occurs, we are often interested in the **relative frequency,** which is the ratio of the frequency to the total number of observations:

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Number of observations}}.$$

For instance, as we see from Table 2.2, the relative frequency of Democrats in Professor Weiss's introductory statistics class is

$$\text{Relative frequency of Democrats} = \frac{\text{Frequency of Democrats}}{\text{Number of observations}} = \frac{13}{40} = 0.325.$$

In terms of percentages, 32.5% of the students in Professor Weiss's introductory statistics class are Democrats. We see that a relative frequency is just a **percentage** expressed as a decimal.

As you might expect, a relative-frequency distribution of qualitative data is similar to a frequency distribution, except that we use relative frequencies instead of frequencies.

> ## To Construct a Relative-Frequency Distribution of Qualitative Data
>
> **Step 1** Obtain a frequency distribution of the data.
>
> **Step 2** Divide each frequency by the total number of observations.

Example:

Political Party Affiliations Refer to Example 2.5 on page 40. Construct a relative-frequency distribution of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1.

Solution We apply Procedure 2.2.

Step 1 Obtain a frequency distribution of the data.

We obtained a frequency distribution of the data in Example 2.5; specifically, see the first and third columns of Table 2.2 on page 41.

Step 2 Divide each frequency by the total number of observations.

Dividing each entry in the third column of Table 2.2 by the total number of observations, 40, we obtain the relative frequencies displayed in the second column of Table 2.3. The two columns of Table 2.3 provide a relative-frequency distribution for the data in Table 2.1.

Party	Relative frequency	
Democratic	0.325	← 13/40
Republican	0.450	← 18/40
Other	0.225	← 9/40
	1.000	

Interpretation From Table 2.3, we see that 32.5% of the students in Professor Weiss's introductory statistics class are Democrats, 45.0% are Republicans, and 22.5% are Other.

Graphical Presentation of Qualitative Data

Two common methods for graphically displaying qualitative data are pie charts and bar charts. We begin with pie charts.

Pie Chart:

Pie Chart

A **pie chart** is a disk divided into wedge-shaped pieces proportional to the relative frequencies of the qualitative data.

To Construct a Pie Chart

Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

Step 2 Divide a disk into wedge-shaped pieces proportional to the relative frequencies.

Step 3 Label the slices with the distinct values and their relative frequencies.

Example:

Political Party Affiliations Construct a pie chart of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1 on page 40.

Solution We apply Procedure 2.3.

Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.
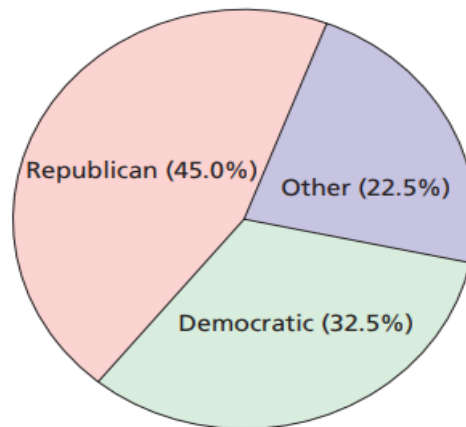
We obtained a relative-frequency distribution of the data in Example 2.6. See the columns of Table 2.3.

FIGURE 2.2
Pie chart of the political party
affiliation data in Table 2.1

Political Party Affiliations

Step 2 Divide a disk into wedge-shaped pieces proportional to the relative frequencies.

Referring to the second column of Table 2.3, we see that, in this case, we need to divide a disk into three wedge-shaped pieces that comprise 32.5%, 45.0%, and 22.5% of the disk. We do so by using a protractor and the fact that there are 360° in a circle. Thus, for instance, the first piece of the disk is obtained by marking off 117° (32.5% of 360°). See the three wedges in Fig. 2.2.

Step 3 Label the slices with the distinct values and their relative frequencies.

Referring again to the relative-frequency distribution in Table 2.3, we label the slices as shown in Fig. 2.2. Notice that we expressed the relative frequencies as percentages. Either method (decimal or percentage) is acceptable.

Bar Graphs:

Frequencies, relative frequencies, or percentages can be used to label a bar chart. Although we primarily use relative frequencies, some of our applications employ frequencies or percentages.

Bar Chart

A **bar chart** displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value. The bars should be positioned so that they do not touch each other.

Example:

Political Party Affiliations Construct a bar chart of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1 on page 40.

Solution We apply Procedure 2.4.

Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

We obtained a relative-frequency distribution of the data in Example 2.6. See the columns of Table 2.3 on page 42.

Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.

See the horizontal and vertical axes in Fig. 2.3.

Step 3 For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.

Referring to the second column of Table 2.3, we see that, in this case, we need three vertical bars of heights 0.325, 0.450, and 0.225, respectively. See the three bars in Fig. 2.3.

Step 4 Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative frequency."

FIGURE 2.3

Bar chart of the political party
affiliation data in Table 2.1
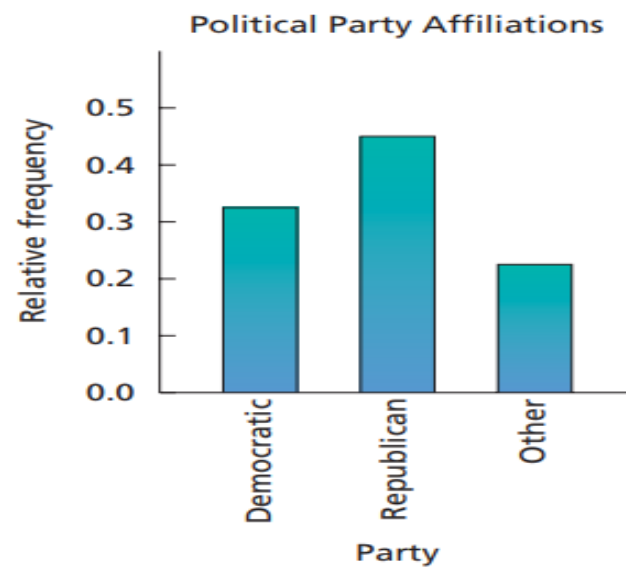
Political Party Affiliations

Practice Problems

2.20 Colleges of Students. The following table provides data on college for the students in one section of the course Introduction to Computer Science during one semester at Arizona State University. In the table, we use the abbreviations BUS for Business, ENG for Engineering and Applied Sciences, and LIB for Liberal Arts and Sciences.

ENG	ENG	BUS	BUS	ENG
LIB	LIB	ENG	ENG	ENG
BUS	BUS	ENG	BUS	ENG
LIB	BUS	BUS	BUS	ENG
ENG	ENG	LIB	ENG	BUS

2.21 Class Levels. Earlier in this section, we considered the political party affiliations of the students in Professor Weiss's introductory statistics course. The class levels of those students are as follows, where Fr, So, Jr, and Sr denote freshman, sophomore, junior, and senior, respectively.

So	So	Jr	Fr	Jr	So	Jr	So
So	So	Sr	So	Jr	Jr	Sr	Fr
Jr	Jr	So	Jr	Fr	Sr	Jr	So
Jr	Fr	Fr	Jr	Sr	So	Sr	Sr
So	Jr	So	Sr	So	So	Fr	So

2.22 U.S. Regions. The U.S. Census Bureau divides the states in the United States into four regions: Northeast (NE), Midwest (MW), South (SO), and West (WE). The following table gives the region of each of the 50 states.

SO	WE	WE	MW	NE	WE	WE	SO	MW	SO
WE	NE	WE	SO	MW	MW	NE	WE	SO	WE
WE	SO	MW	SO	MW	WE	SO	NE	SO	SO
SO	SO	MW	NE	SO	NE	MW	NE	WE	MW
WE	SO	MW	SO	MW	NE	MW	SO	NE	WE

2.23 Road Rage. The report *Controlling Road Rage: A Literature Review and Pilot Study* was prepared for the AAA Foundation for Traffic Safety by D. Rathbone and J. Huckabee. The authors discuss the results of a literature review and pilot study on how to prevent aggressive driving and road rage. As described in the study, *road rage* is criminal behavior by motorists characterized by uncontrolled anger that results in violence or threatened violence on the road. One of the goals of the study was to determine when road rage occurs most often. The days on which 69 road rage incidents occurred are presented in the following table.

F	F	Tu	Tu	F	Su	F	F	Tu	F
Tu	Sa	Sa	F	Sa	Tu	W	W	Th	Th
Th	Sa	M	Tu	Th	Su	W	Th	W	Tu
Tu	F	Th	Th	F	W	F	Th	F	Sa
F	W	W	F	Tu	W	W	Th	M	M
F	Su	Tu	F	W	Su	W	Th	M	Tu
F	W	Th	M	Su	Sa	Sa	F	F	