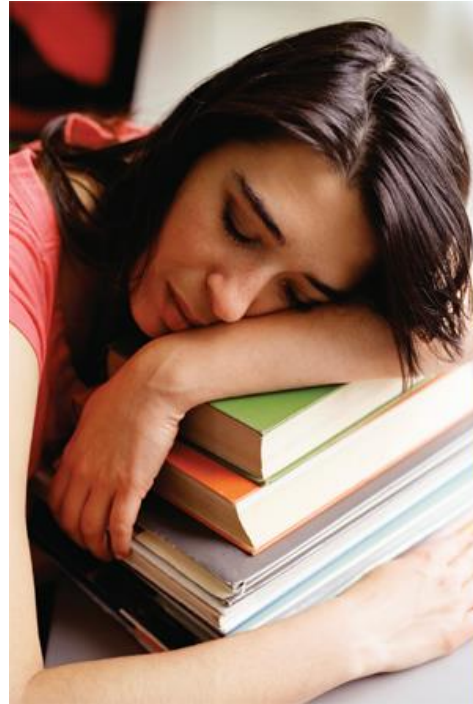


2

Descriptive Analysis and Presentation of Single-Variable Data



2.1 Graphs, Pareto Diagrams, and Stem-and-Leaf Displays



Qualitative Data

Qualitative Data

Pie charts (circle graphs) and bar graphs Graphs that are used to summarize **qualitative**, or attribute, or categorical data.

Pie charts (circle graphs) show the amount of data that belong to each category as a proportional part of a circle.

Bar graphs show the amount of data that belong to each category as a proportionally sized rectangular area.

Example 1 – *Graphing Qualitative Data*

Table 2.1 lists the number of cases of each type of operation performed at General Hospital last year.

Type of Operation	Number of Cases
Thoracic	20
Bones and joints	45
Eye, ear, nose, and throat	58
General	98
Abdominal	115
Urologic	74
Proctologic	65
Neurosurgery	23
<i>Total</i>	498

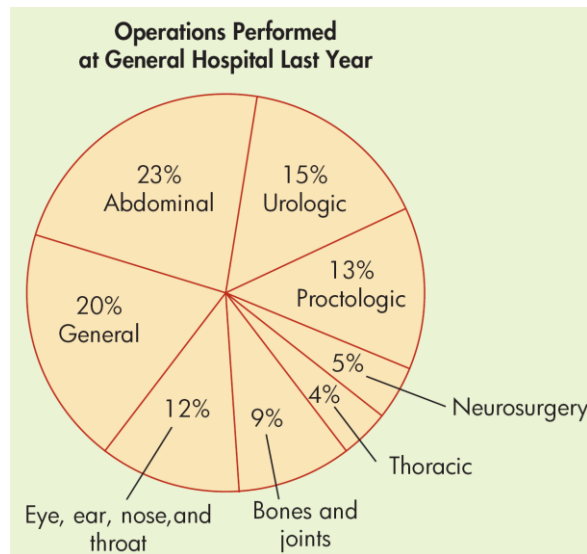
Operations Performed at General Hospital Last Year [TA02-01]

Table 2.1

Example 1 – *Graphing Qualitative Data*

cont'd

The data in Table 2.1 are displayed on a pie chart in Figure 2.1, with each type of operation represented by a relative proportion of a circle, found by dividing the number of cases by the total sample size, namely, 498.



Pie Chart

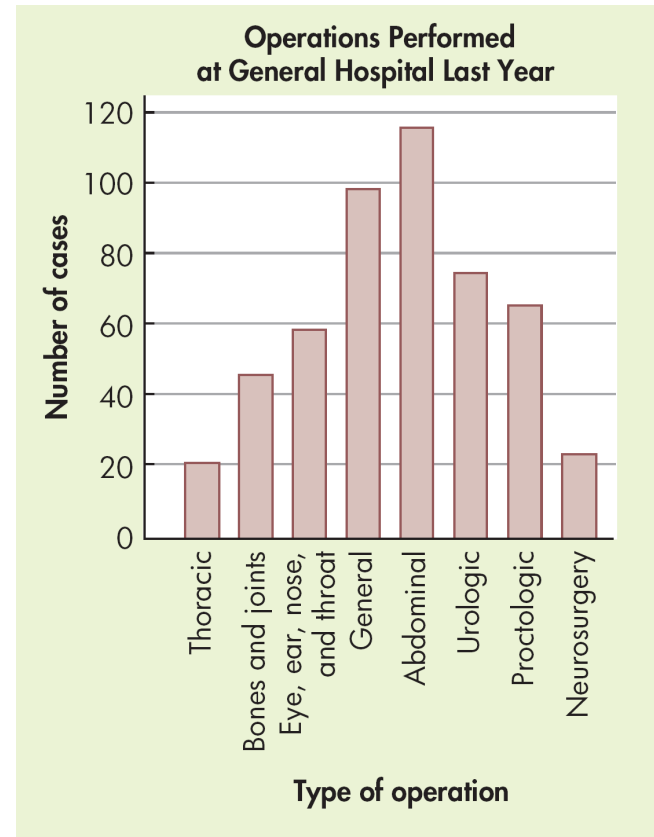
Figure 2.1

Example 1 – *Graphing Qualitative Data*

cont'd

The proportions are then reported as percentages (for example, 25% is 1/4 of the circle). Figure 2.2 displays the same “type of operation” data but in the form of a bar graph.

Bar graphs of attribute data should be drawn with a space between bars of equal width.



Bar Graph

Figure 2.2

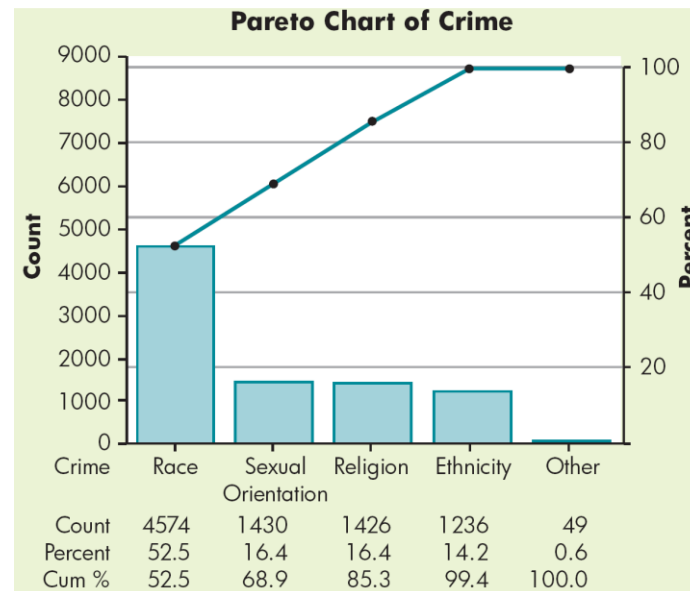
Qualitative Data

Pareto diagram A bar graph with the bars arranged from the most numerous category to the least numerous category. It includes a line graph displaying the cumulative percentages and counts for the bars.

The Pareto diagram is popular in quality-control applications. A Pareto diagram of types of defects will show the ones that have the greatest effect on the defective rate in order of effect. It is then easy to see which defects should be targeted to most effectively lower the defective rate.

Example 2 – *Pareto Diagram of Hate Crimes*

The FBI reported the number of hate crimes by category for 2003 (<http://www.fbi.gov/>). The Pareto diagram in Figure 2.3 shows the 8715 categorized hate crimes, their percentages, and cumulative percentages.



Pareto Diagram

Figure 2.3



Quantitative Data

Quantitative Data

One major reason for constructing a graph of **quantitative data** is to display its *distribution*.

Distribution The pattern of variability displayed by the data of a variable. The distribution displays the frequency of each value of the variable.

Dotplot display Displays the data of a sample by representing each data value with a dot positioned along a scale. This scale can be either horizontal or vertical. The frequency of the values is represented along the other scale.

Quantitative Data

Stem-and-leaf display Displays the data of a sample using the actual digits that make up the data values. Each numerical value is divided into two parts: The leading digit(s) becomes the stem, and the trailing digit(s) becomes the leaf. The stems are located along the main axis, and a leaf for each data value is located so as to display the distribution of the data.

Example 5 – Overlapping Distributions

A random sample of 50 college students was selected. Their weights were obtained from their medical records. The resulting data are listed in Table 2.3.

Student	1	2	3	4	5	6	7	8	9	10
Male/Female	F	M	F	M	M	F	F	M	M	F
Weight	98	150	108	158	162	112	118	167	170	120
Student	11	12	13	14	15	16	17	18	19	20
Male/Female	M	M	M	F	F	M	F	M	M	F
Weight	177	186	191	128	135	195	137	205	190	120
Student	21	22	23	24	25	26	27	28	29	30
Male/Female	M	M	F	M	F	F	M	M	M	M
Weight	188	176	118	168	115	115	162	157	154	148
Student	31	32	33	34	35	36	37	38	39	40
Male/Female	F	M	M	F	M	F	M	F	M	M
Weight	101	143	145	108	155	110	154	116	161	165
Student	41	42	43	44	45	46	47	48	49	50
Male/Female	F	M	F	M	M	F	F	M	M	M
Weight	142	184	120	170	195	132	129	215	176	183

Weights of 50 College Students [TA02-03]

Table 2.3

Example 5 – Overlapping Distributions cont'd

Notice that the weights range from 98 to 215 pounds. Let's group the weights on stems of 10 units using the hundreds and the tens digits as stems and the units digit as the leaf (see Figure 2.7).

The leaves have been arranged in numerical order. Close inspection of Figure 2.7 suggests that two overlapping distributions may be involved.

Weights of 50 College Students (lb)	
$N = 50$	Leaf Unit = 1.0
9	8
10	1 8 8
11	0 2 5 5 6 8 8
12	0 0 0 8 9
13	2 5 7
14	2 3 5 8
15	0 4 4 5 7 8
16	1 2 2 5 7 8
17	0 0 6 6 7
18	3 4 6 8
19	0 1 5 5
20	5
21	5

Stem-and-Leaf Display

Figure 2.7

Example 5 – Overlapping Distributions cont'd

That is exactly what we have: a distribution of female weights and a distribution of male weights.

Figure 2.8 shows a “back-to-back” stem-and-leaf display of this set of data and makes it obvious that two distinct distributions are involved.

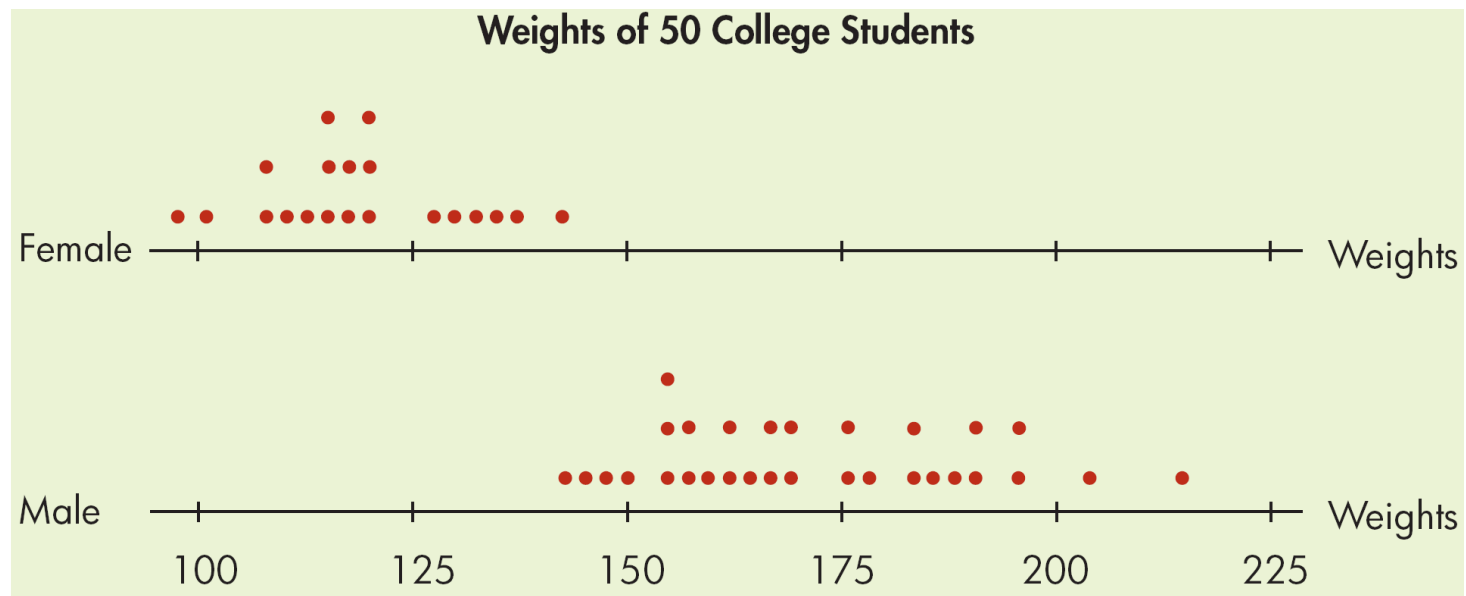
Weights of 50 College Students (lb)		
Female		Male
	8	09
	1 8 8	10
0 2 5 5 6 8 8	11	
0 0 0 8 9	12	
2 5 7	13	
2	14	3 5 8
	15	0 4 4 5 7 8
	16	1 2 2 5 7 8
	17	0 0 6 6 7
	18	3 4 6 8
	19	0 1 5 5
	20	5
	21	5

“Back-to-Back” Stem-and-Leaf Display

Figure 2.8

Example 5 – *Overlapping Distributions* cont'd

Figure 2.9, a “side-by-side” dotplot (same scale) of the same 50 weight data, shows the same distinction between the two subsets.



Dotplots with Common Scale

Figure 2.9

Example 5 – *Overlapping Distributions* cont'd

Based on the information shown in Figures 2.8 and 2.9, and on what we know about people's weight, it seems reasonable to conclude that female college students weigh less than male college students.