AN INTRODUCTION TO STATISTICS

WITH

DATA ANALYSIS

by

SHELLEY RASMUSSEN

Department of Mathematical Sciences
Olney 428T
University of Massachusetts/Lowell
Lowell, MA 01854

Originally published by Brooks/Cole Publishing Company,
Division of Wadsworth, Inc.

ISBN 0-534-13578-1

© 1992 by Wadsworth, Inc.

© 2006 by Shelley Rasmussen

Permission is granted by the author for non-for-profit educational use.

Shelley_Rasmussen@uml.edu

Minitab is a statistical package, a computer program that performs many statistical procedures. The versions of Minitab now available for use on personal computers are menu-driven and much easier to use than the main-frame version originally discussed in this text. Those sections are not included in this online edition of the text. At this time, the most recent version is Minitab 15, available at very reasonable prices for purchase or rental from:

www.e-academy.com/minitab

System Requirements

Processor:	PC with a 1 GHz 32- or 64-bit processor
Memory:	512 MB or more of available RAM
Disk Space:	125 MB free space available
Operating System:	Microsoft Windows 2000, XP, or Vista.
Display:	A display capable of 1024 X 768 or higher resolution
Software:	Adobe Acrobat Reader 5.0 or higher for Meet Minitab

CHAPTER 4

Studying Two Variables at a Time

IN THIS CHAPTER

- Two-way frequency table or contingency table
- Plots of a quantitative variable over levels of a qualitative variable
- Dot chart
- Scatterplot
- Logarithmic transformation

Looking at individual variables can be interesting, but analysis should not stop there. In fact, we went beyond individual variables in Chapters 2 and 3 when we compared plots and descriptive statistics for life expectancy and population growth across economic categories. Such analyses address our second goal: to study each indicator separately within economic categories and compare results across economic categories. We continue now by discussing a number of ways to study relationships between two variables. Let's begin with two-way frequency tables, for studying relationships between two qualitative variables.

4-1

Two-Way Frequency Tables for Studying the Relationship Between Two Qualitative Variables

In making projections of future population sizes, the World Bank classifies countries as having primary school enrollments for females greater than or less than 70% (World Bank, 1987, page 281). Analysts make separate projections for the two groups of countries. To see how many countries are in each group, we might look at a frequency table such as Table 4-1. This table shows the number of countries with high female primary school enrollments, low enrollments, and missing enrollment information.

From female primary school enrollment, we have created an ordinal qualitative variable with two possible values: low (less than or equal to 70%) and high (greater than 70%). About 22% of the 128 countries have low female primary school enrollments as defined by this new variable; about 11% have missing values. Do these percentages vary across economic categories? To find out, we can count the number of countries with low female primary school enrollments, the number with high enrollments, and the number with missing information, within each economic category. We might then arrange these counts in a table that has columns defined by level of female primary school enrollment and rows defined by economic categories. (Alternatively, columns could be defined by economic category and rows by enrollment level.)

In general, when we count the number of cases within each combination of values for two qualitative variables and arrange these counts in a two-dimen-

TABLE 4-1 Countries classified by number of females enrolled in primary school in 1984 as percentage of 6–11-year age group.

1984 primary school enrollment for females	Number of countries	Percentage of countries
Less than or equal to 70%	28	22
Greater than 70%	86	67
Missing	14	11
Total	128	100

sional table with rows defined by one variable and columns by another, we create a *two-way frequency table.* We use such a table to look for association between the two variables.

> A **two-way frequency table**, or two-dimensional contingency table, is a tabular display of the number of cases within each combination of categories of two qualitative variables.

If the distribution of frequencies across categories of one variable depends on the category of the other variable, we say the two variables are *associated.* If the distribution of frequencies across categories of one variable is about the same for each category of the other variable, we say the two variables are *independent,* or not associated.

> We say two qualitative variables are **associated** if the distribution across categories of one variable depends on the category of the other variable.
>
> We say two qualitative variables are **independent** or **not associated** if the distribution across categories of one variable is about the same for all categories of the other variable.

Table 4-2 is a two-way frequency table for economic category and female primary school enrollment. Examining the table, we see that the distribution of counts across levels of female primary school enrollment depends on economic category. Equivalently, the distribution of counts across economic categories is different for the different levels of female primary school enrollment. Based on this table, we say there is an association between the two variables: Lower female primary school enrollments and more missing values are associated with lower economic categories; higher female primary school enrollments and fewer missing values are associated with greater economic development as measured by economic category.

Do lower female primary school enrollments tend to be associated with lower levels of contraception use among the World Bank countries? We address this question by examining the two-way frequency table in Table 4-3. In

TABLE 4-2 Countries classified by economic category and number of females enrolled in primary school in 1984 as percentage of 6–11-year age group.

	1984 primary school enrollment for females			
Economic category	≤70%	>70%	Missing	Total
Low income	18	13	6	37
Lower-middle income	9	24	3	36
Upper-middle income	0	22	1	23
High-income oil exporting	1	2	1	4
Industrial market	0	18	1	19
Nonmember	0	7	2	9
Total	28	86	14	128

TABLE 4-3 Countries classified by number of females in primary school in 1984 as percentage of age group and percentage of married women of child-bearing age using contraception in 1984. (Thirty-nine countries are excluded because of missing values on one or both of these variables.) The number in parentheses to the right of a frequency is the percentage of the row total. The number in parentheses below a frequency is the percentage of the column total.

Percent contra-ception use	Female primary school enrollment		
	≤70%	>70%	Total
≤35%	23 (53) (96)	20 (47) (31)	43 (100) (48)
>35%	1 (2) (4)	45 (98) (69)	46 (100) (52)
Total	24 (27) (100)	65 (73) (100)	89 (100) (100)

this table, female school enrollment has two categories: low (≤70%) and high (>70%). Percent contraception use also has two categories: low (≤35%) and high (>35%). The cutoff value of 35% defining the categories for contraception use divides the countries with nonmissing information into approximately equal sized groups. Note that Table 4-3 is based on the 89 countries with nonmissing information on both variables.

For interpreting a two-way frequency table, it can be helpful to show each frequency as a percentage of its row total and as a percentage of its column total. In Table 4-3, these percentages are displayed along with the frequencies. The number in parentheses to the right of a frequency is the percentage of the row total. The number in parentheses below a frequency is the percentage of the column total. These row and column percentages help us see the strength of the association between female primary school enrollment and contraception use in Table 4-3. As the first column shows, all but one of the countries with lower enrollments also have lower contraception use. The second column shows that 69% of countries with higher enrollments also have greater contraception use. Looking at the table row-wise, we see that more than half the countries with lower contraception use have lower enrollments. All but one country with greater contraception use also have higher female primary school enrollments.

We have to be careful interpreting Table 4-3. Thirty-nine (30%) of the 128 countries have missing information on one or both variables. Also, we have taken two quantitative variables and created two new variables, each with just a low and a high category. *How we select the dividing point for low and high can affect what we see in the frequency table.* We might choose several ways to divide the variables and see if the same relationship holds in each analysis. Or, we might decide not to categorize the indicators at all, and instead examine the relationship between them using scatterplots, as discussed in Section 4-3.

We will look at female primary school enrollment and contraception use again. But first, we consider ways to study the relationship between a quantitative variable and a qualitative variable.

4-2

Tables and Graphs for Studying the Relationship Between a Quantitative Variable and a Qualitative Variable

We saw two modes or peaks when we looked at the plot of fertility rates in Figure 2-15. (Recall that fertility rate is an estimate of the average number of children per woman in a country.) One peak was around 2 children per woman and another around 6.5 children per woman. Are lower fertility rates associated with higher levels of economic development? Let's see how the quantitative variable fertility rate is related to the qualitative variable economic category.

Table 4-4 shows some descriptive statistics for fertility rates, overall and within economic categories. We see that both the mean and median fertility rate decrease as economic category increases from low-income to lower-middle-income to upper-middle-income to industrial market. The high-income oil exporters are on a par with the low-income countries. The non-members are close to the upper-middle-income group.

Figure 4-1 shows a dot plot of fertility rates for each economic category. The axes are lined up, with the same scale for each plot, making visual comparisons easy. These plots clearly show the relationship between fertility rate and economic category. The industrial market countries have the lowest fertility rates and the least spread in the values. The other economic categories all show much higher fertility rates in general, and greater variation among the values.

We might choose to show a box plot of the quantitative variable for each level of the qualitative variable. Such a display allows easy comparisons of

TABLE 4-4 Descriptive statistics for 1985 total fertility rates, overall and within economic categories. The summary statistics are the number in the group, *n;* the mean; median; standard deviation, SD; minimum; maximum. Two low-income countries (Afghanistan, Kampuchea) and one lower-middle-income country (Lebanon) are excluded because of missing values.

Economic category	*n*	Mean	Median	SD	Minimum	Maximum
Low income	35	6.1	6.4	1.2	2.3	8.0
Lower-middle income	35	5.1	5.4	1.4	2.5	6.9
Upper-middle income	23	3.4	2.9	1.6	1.7	6.7
High-income oil exporting	4	6.4	6.5	1.0	5.2	7.2
Industrial market	19	1.8	1.7	.3	1.3	2.6
Nonmembers	9	3.2	2.3	1.6	1.8	6.4
Overall	125	4.5	4.7	2.0	1.3	8.0

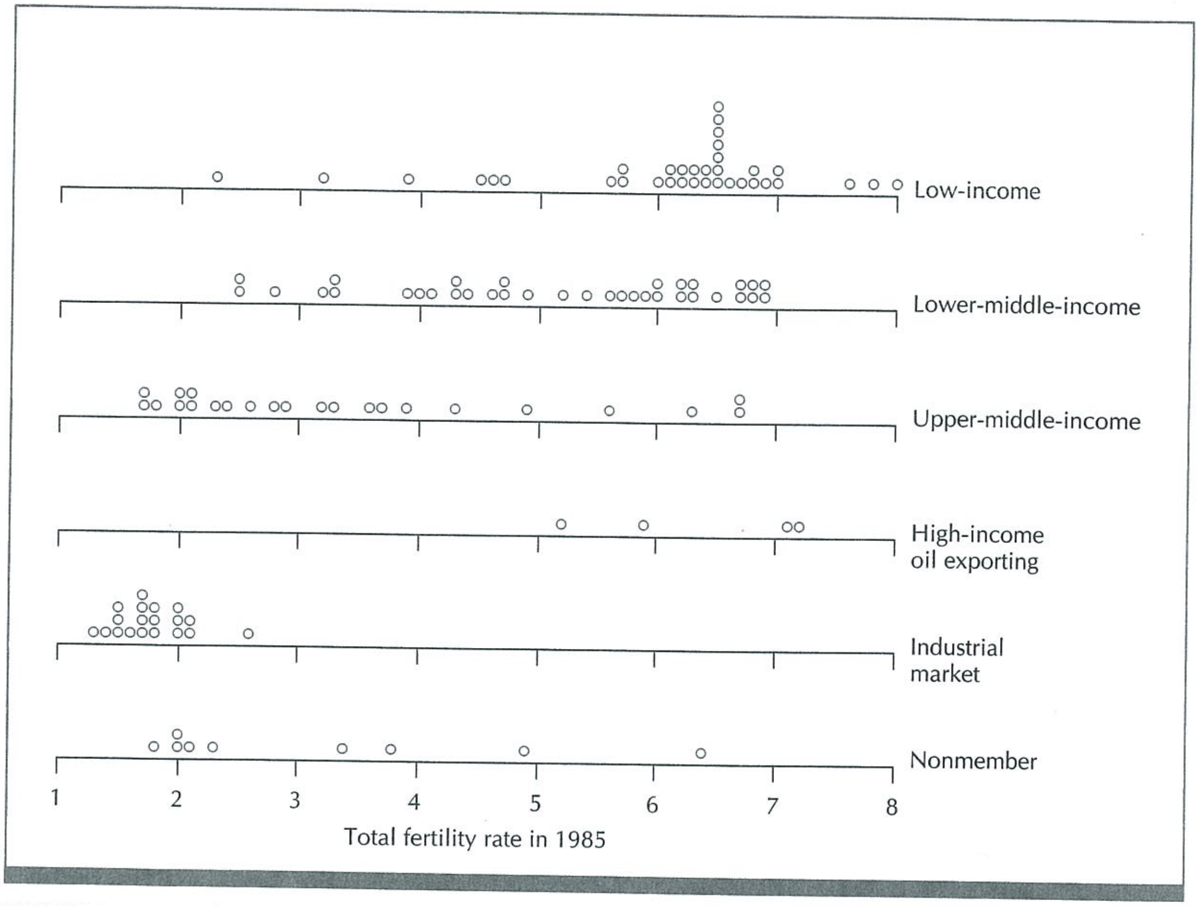

FIGURE 4-1 Dot plot of 1985 total fertility rate for each economic category. Two low-income countries and one lower-middle-income country are excluded because of missing values.

center, spread, and extreme values of the quantitative variable within each level of the qualitative variable. Figure 4-2 displays a box plot of fertility rate for each economic category.

A less informative (but often used) type of plot is shown in Figure 4-3. Here the mean fertility rate, the mean plus one standard deviation, and the mean minus one standard deviation are plotted for each economic category. This type of plot can be misleading. It suggests symmetry when the distribution of values may not be symmetrical at all (as we see in Figures 4-1 and 4-2). We get no information about skewness or concentrations of values. Displays such as those in Figures 4-1 and 4-2 are much more informative.

Even less informative and potentially more misleading than Figure 4-3 is a graph that shows only the means of the quantitative variable across levels of the qualitative variable (Figure 4-4). Some authors connect the means to the

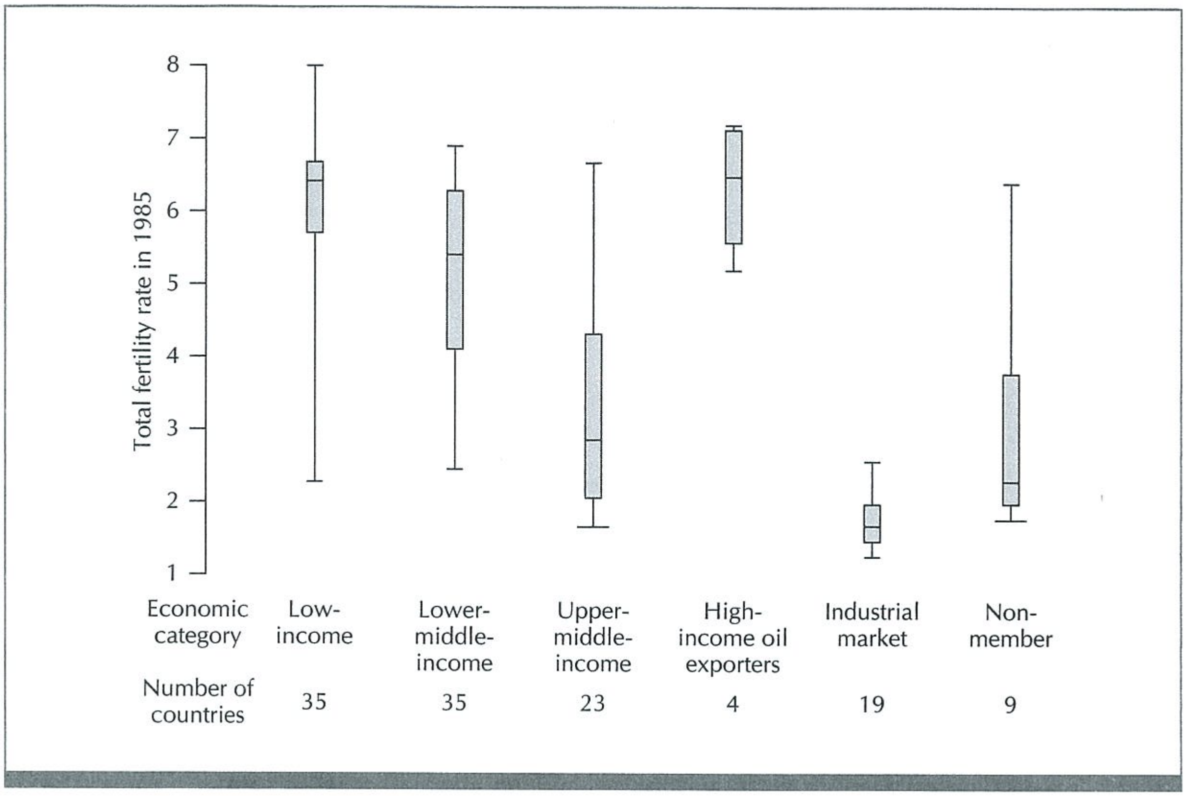

FIGURE 4-2 Box plot of 1985 total fertility rate for each economic category. Two low-income countries and one lower-middle-income country are excluded because of missing values.

horizontal axis with a line or bar, making the graph look like a histogram or frequency plot. The visual impression of differences between means is completely dependent on the scale chosen for the quantitative variable. Since no measures of variation are provided, we cannot make meaningful comparisons across categories. This is not a useful graph.

Dot Charts

Using box plots, histograms, and dot plots, we can compare the distribution of a quantitative variable across levels of a qualitative variable. Sometimes we want to compare totals (or percentages) of a quantitative variable across levels of a qualitative variable. A good graphical tool for such a comparison is the dot chart.

A *dot chart* is a graphical display of totals (or percentages) for a quantitative variable at each level of a qualitative variable (Cleveland, 1985). Examples of dot charts are given in Figures 4-5 and 4-6. Levels of the qualitative

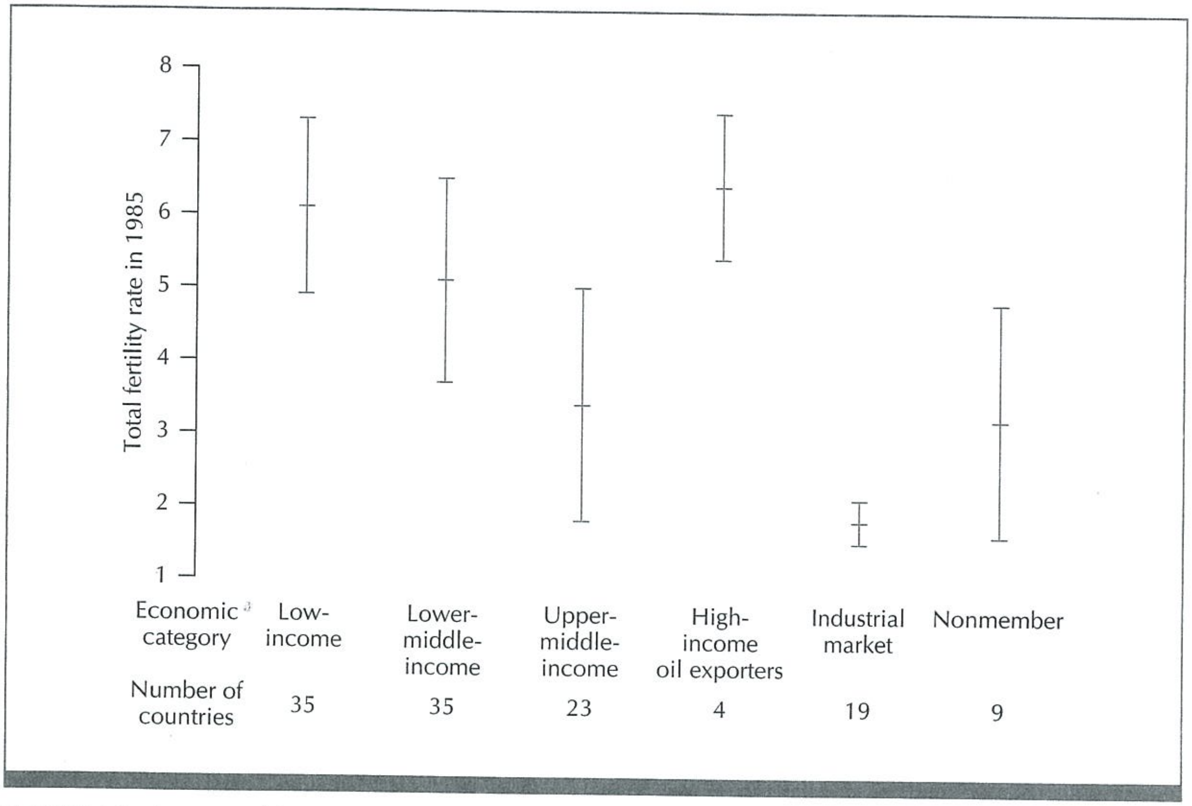

FIGURE 4-3 Mean total fertility rate in 1985, mean plus one standard deviation, and mean minus one standard deviation for each economic category. (Two low-income countries and one lower-middle-income country are excluded because of missing values.) *This type of plot is of limited usefulness.*

variable are listed at the left of the chart. For each category, a horizontal line illustrates the value (and relative value or percentage of total) of the quantitative variable. The axis at the bottom of the chart shows the scale for the quantitative variable; the axis at the top shows percentage of total.

A **dot chart** is a graphical tool for comparing totals (or percentages) for a quantitative variable across levels of a qualitative variable. We make visual comparisons along a straight axis.

Figure 4-5 displays population size for each of five economic categories (nonmember nations are excluded). The lower axis shows population in millions of people and the upper axis is percentage of total population among the countries included in the chart. A dot indicates the population size for each economic category. A line helps us connect the economic category label with the corresponding dot.

The dot chart in Figure 4-6 displays total gross domestic product (an estimate of the value of goods and services produced by a country, in millions

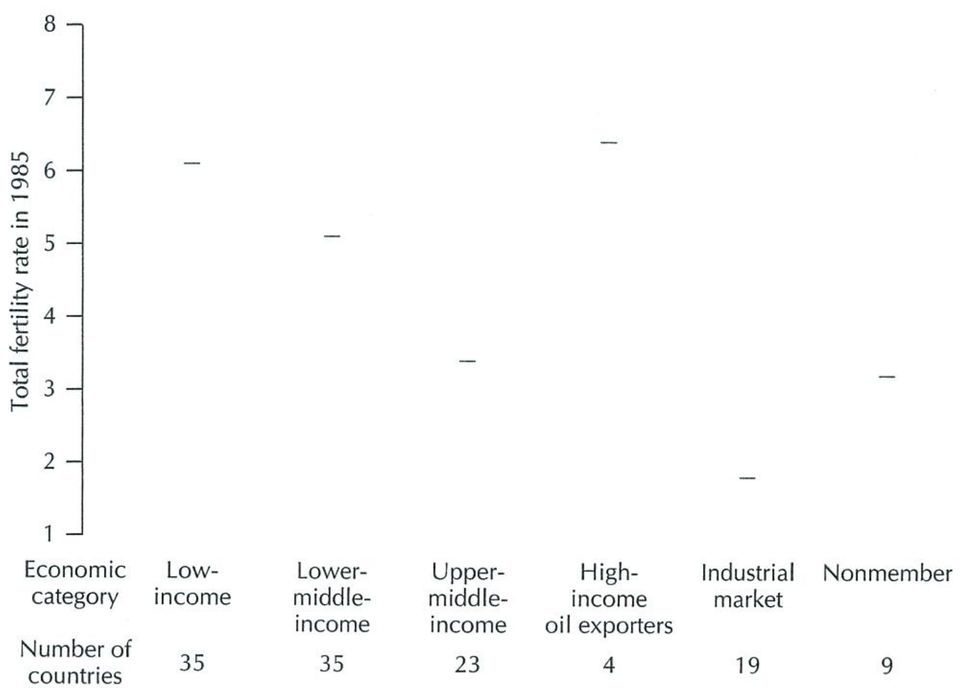

FIGURE 4-4 Mean total fertility rate in 1985 for each economic category. (Two low-income countries and one lower-middle-income country are excluded because of missing values.) *Please do not use this type of graph!*

of U.S. dollars) for each of five economic categories, again excluding nonmembers. The scale for gross domestic product is shown at the bottom. The upper axis gives percentage of total gross domestic product among the countries included in the plot.

Figure 4-5 shows that the low-income countries comprise over 50% of the total population of countries included in the chart; the industrial market countries represent less than 20% of the population total. Figure 4-6 shows that the low-income countries account for less than 10% of the total gross domestic product of countries included in the chart, while the industrial market countries account for over 70% of this total. We must be careful in comparing the two graphs because they are based on different numbers of countries. (The two charts could have been based only on countries with nonmissing information on both variables.) In addition, gross domestic product is difficult to compare across countries, especially across economic categories. However, these two dot charts give an overwhelming impression that the low-income countries include most of the population and the industrial market countries most of the material wealth.

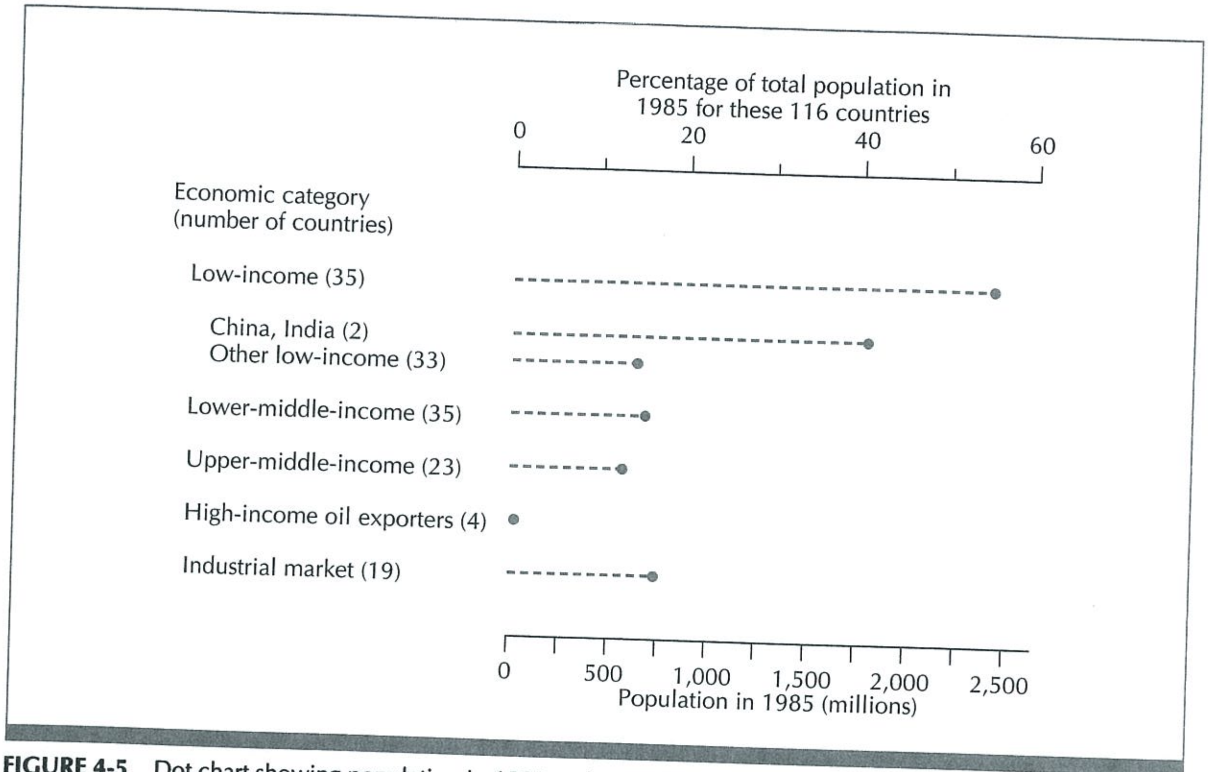

FIGURE 4-5 Dot chart showing population in 1985 and percentage of total, by economic category.

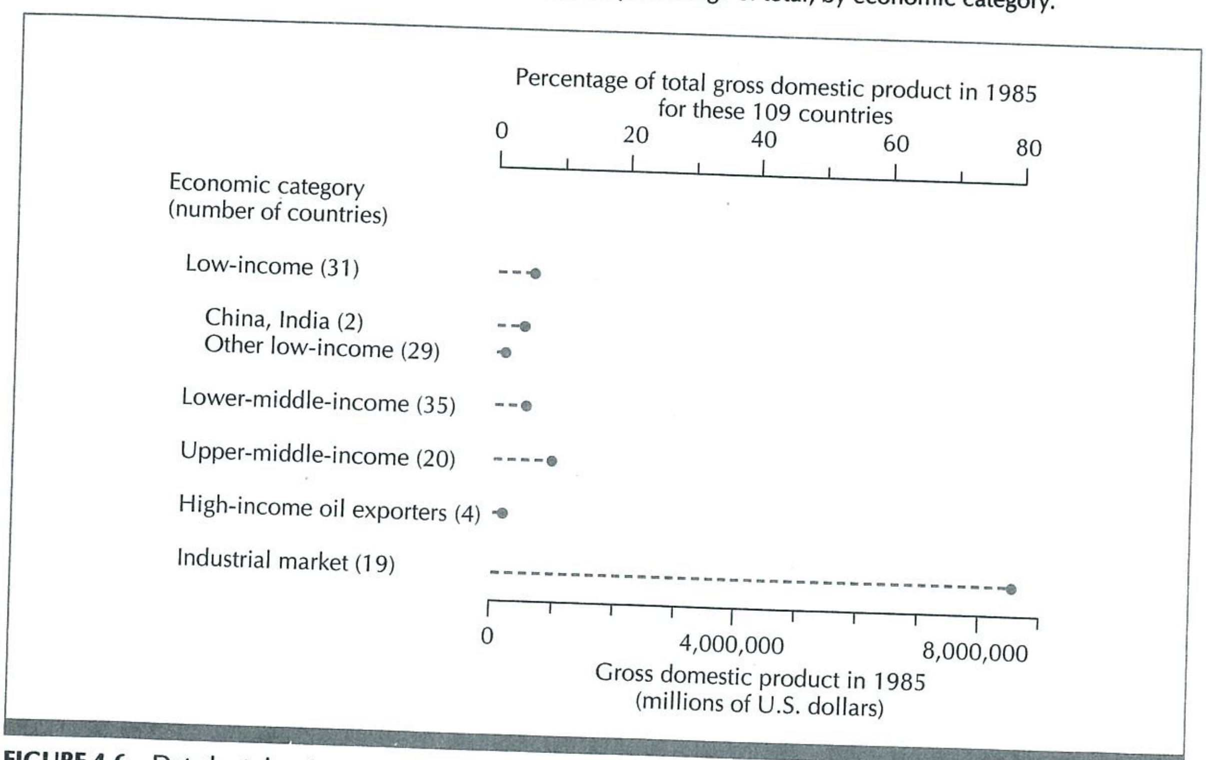

FIGURE 4-6 Dot chart showing gross domestic product in 1985 and percentage of total, by economic category.

When used to display frequencies and percentages across categories of a single qualitative variable, a dot chart is the same as a frequency plot (see Exercise 2-23). Dot charts present the same kind of information commonly displayed in pie charts for many business and social science applications. As a graphical tool, the dot chart is easier to interpret because we make comparisons of totals or percentages along a straight axis. A pie chart requires more difficult visual comparisons of angles. For a discussion of the advantages of the dot chart over the pie chart for visual interpretations of data, see Cleveland (1985).

Section 4-3 looks at ways to examine the relationship between two quantitative variables.

4-3 Scatterplots for Studying the Relationship Between Two Quantitative Variables

Life expectancies are shorter among low-income countries than among industrial market nations (Figure 2-18). How do life expectancies in general vary with gross national product among the World Bank countries? We can examine the relationship between these two quantitative variables using a scatterplot.

A *scatterplot* or scattergram is a two-dimensional graphical display of two quantitative variables. The scale for one variable is on the vertical axis, the scale for the other variable on the horizontal axis. Each case is represented by a point on the plot positioned according to the values of the two variables.

> A **scatterplot** or scattergram is a two-dimensional graphical display of two quantitative variables, the scale for one variable on the vertical axis and the scale for the other on the horizontal axis. Each case is represented by a point on the plot positioned according to the values of the two variables.

Gross national product per capita and life expectancy are listed in Table 4-5 for the four high-income oil exporters. We see that the two countries with greater per capita gross national products also have longer life expectancies. Figure 4-7 shows a scatterplot for these data. Each of the four points on the graph represents a high-income oil exporting nation. Since we are interested in seeing how life expectancies vary with per capita gross national product, we follow tradition and plot per capita gross national product along the horizontal axis and life expectancy along the vertical axis. (The variable on the horizontal axis is often called the independent variable; the variable on the vertical axis, the dependent variable. *Such designations are merely conventions; they do not imply any cause-and-effect relationship between the two variables.*)

The scatterplot in Figure 4-7 uses a *range frame* (Tufte, 1983, page 130): Each axis extends from the minimum to the maximum value of the corresponding variable. The vertical axis extends from 60 to 72 years, the horizontal axis from 7,170 to 19,270 U.S. dollars. Note that the axes do not intersect. Such

TABLE 4-5 Per capita gross national product in 1985 and life expectancy at birth in 1985 for four high-income oil exporting nations.

Nation	Per capita gross national product in 1985 (U.S. dollars)	Life expectancy at birth in 1985 (years)
Libya	7,170	60
Saudi Arabia	8,850	62
Kuwait	14,480	72
United Arab Emirates	19,270	70

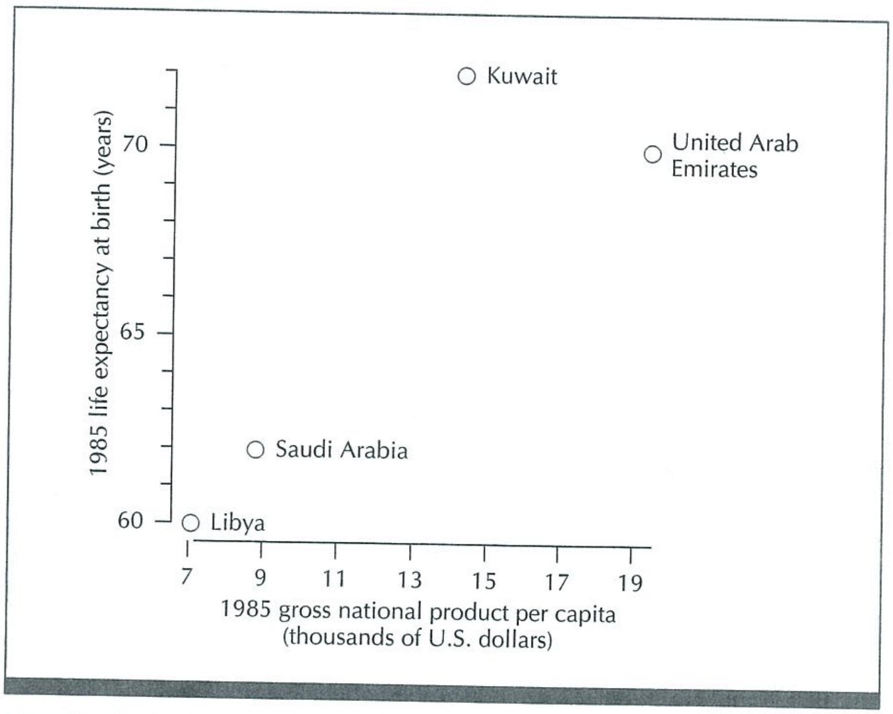

FIGURE 4-7 Scatterplot of 1985 life expectancy at birth and 1985 gross national product per capita for four high-income oil exporting nations. The range of values for each variable is indicated by the line on the corresponding axis.

a range frame allows efficient use of the plotting portion of the graph, while clearly showing the range of each variable. With a range frame, we avoid the often misleading perception of an origin where axes meet (see Exercise 4-1).

In a **range frame** for a scatterplot, each axis extends from the minimum to the maximum value for the corresponding variable.

Sometimes on a scatterplot we provide some identifying information near each plotted point. In Figure 4-7, country name is indicated near each point. Such labels can be useful in plots without too many points (when labeling might lead to confusing clutter).

Our overall impression from Figure 4-7 is that high-income oil exporters with greater per capita gross national products tend to have longer life expectancies. This is a very small group of countries. We would not necessarily expect to see the same relationship within another group of countries or in our data set as a whole. Figure 4-8 shows a scatterplot of life expectancy and gross

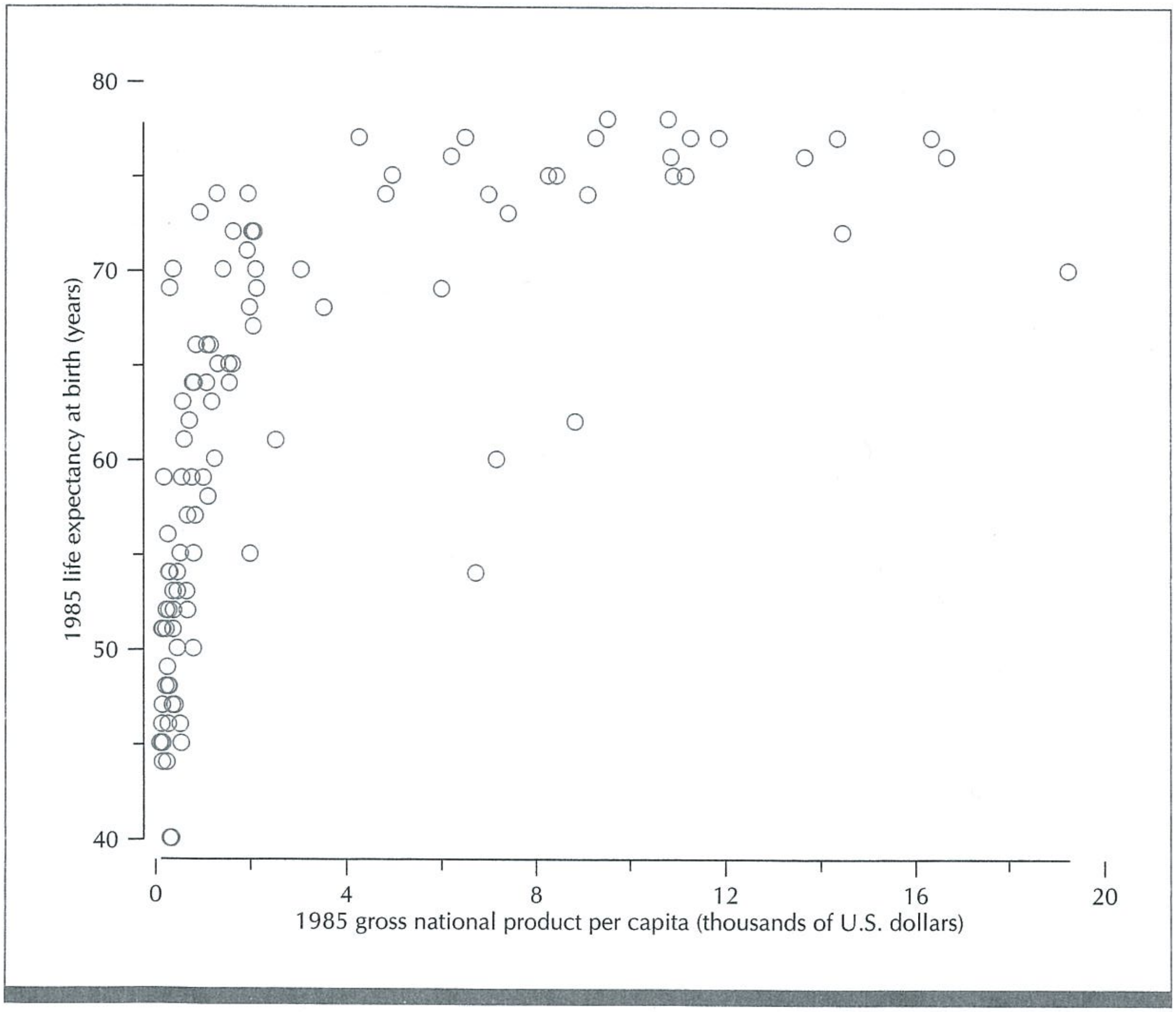

FIGURE 4-8 Scatterplot of 1985 life expectancy at birth and 1985 gross national product per capita for 109 countries. Nineteen countries are excluded because of missing values. The range of values for each variable is indicated by the line on the corresponding axis.

national product, based on the 109 countries with information on both variables. Each country in this plot is represented by an open circle. Darker circles indicate more than one case with the same (or very similar) values. For suggestions on dealing with overlap in scatterplots, see Cleveland (1985).

Figure 4-8 shows that countries with greater per capita gross national products tend to have longer life expectancies. For countries with per capita gross national products less than $2,000, small differences in income are associated with much larger differences in life expectancy than is true for other countries. Also, more than half the countries have per capita incomes less than $2,000. (The median per capita gross national product is $1,010.) This creates congestion in the left portion of the plot. It would be nice if we could stretch out the gross national product scale for lower incomes and compress the scale for higher incomes. One such change of scale uses the logarithm of per capita gross national product. We call such a change of scale a *transformation*. When we transform the scale of a variable, we hope to make interpretations easier in our analysis.

> A **transformation** of a variable is a mathematical manipulation of each value of the variable. When we make a transformation, we transform the original scale of measurement for the variable to a new scale. For example, the logarithmic transformation involves taking the logarithm of each value of a variable, the square root transformation involves taking the square root of each value of a variable, and a power transformation involves taking a power (such as the square or cube) of each value of a variable.

We will use logarithm base-10 to transform the values of per capita gross national product to a new scale. Recall that the logarithm base-10 of a constant is the number you raise 10 to in order to get the constant back. For example, the logarithm base-10 of 1,000, denoted $\log_{10}(1{,}000)$, is 3 since $10^3 = 1{,}000$. Similarly, $\log_{10}(10{,}000) = 4$ since $10^4 = 10{,}000$ and $\log_{10}(5{,}000) = 3.69897$ since $10^{3.69897} = 5{,}000$.

> To make the **logarithmic transformation** of a variable, take the logarithm of each value of the variable.
>
> The **logarithm base-10,** denoted $\log_{10}$, of a constant is the number you raise 10 to in order to get the constant back. For example, $\log_{10}(10{,}000) = 4$ since $10^4 = 10{,}000$.

A scatterplot of life expectancy and the logarithm of per capita gross national product is shown in Figure 4-9. The scale for the logarithm base-10 of per capita gross national product is shown on the top of the plot. The same scale is shown at the base of the plot, but with actual per capita gross national products indicated on the logarithmic scale.

What does Figure 4-9 show? We see that on the logarithmic scale, the distance between gross national products of $100 and $1,000 is the same as the distance between $1,000 and $10,000. We have indeed stretched out the lower incomes and compressed the higher incomes. This relieves the congestion we saw in Figure 4-8 and gives us a clearer picture of the relationship between

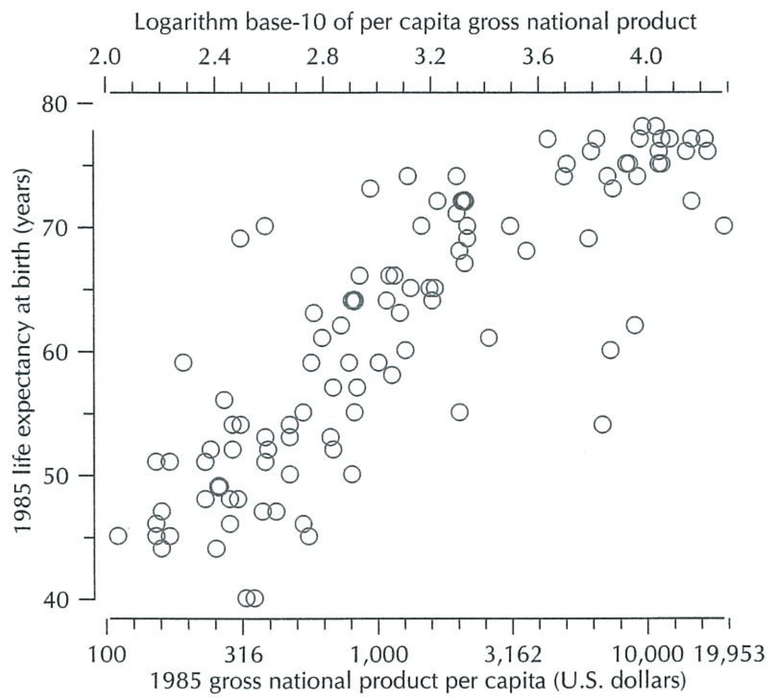

FIGURE 4-9 Scatterplot of 1985 life expectancy at birth and the logarithm of 1985 gross national product per capita for 109 countries. Nineteen countries are excluded because of missing values. The range of values for each variable is indicated by the line on the corresponding axis.

gross national product and life expectancy. We see that average life expectancy increases in roughly a linear fashion as the logarithm of per capita gross national product increases.

In Figure 4-10, we have added a dot plot for life expectancy on the left vertical axis. We have also added a dot plot for gross national product (on the logarithmic scale) on the bottom axis. This is a variation on the dot-dash-plot suggested by Tufte (1983, page 133). We get a visual impression of the distribution of each variable separately as well as the relationship between the two variables.

With a transformation, we get a clearer picture of the relationship between gross national product and life expectancy. Sometimes it is helpful to transform both variables, making two changes of scale in a scatterplot (see Exercise 4-26). Trial and error determines whether a transformation of one or both variables helps to illustrate the relationship between two variables.

As another example, Figure 4-11 shows a scatterplot of primary school enrollment and gross national product. We again see congestion of points at low income levels. Transforming gross national product by taking the logarithm base-10, we obtain the scatterplot in Figure 4-12. We see an increasing relationship between gross national product and primary school enrollments for countries with per capita gross national products less than $1,000. This increasing relationship does not exist for countries with per capita gross national products over $1,000. For these countries, primary school enrollments level off, fluctuating around 100%.

The examples so far have shown positive or increasing relationships

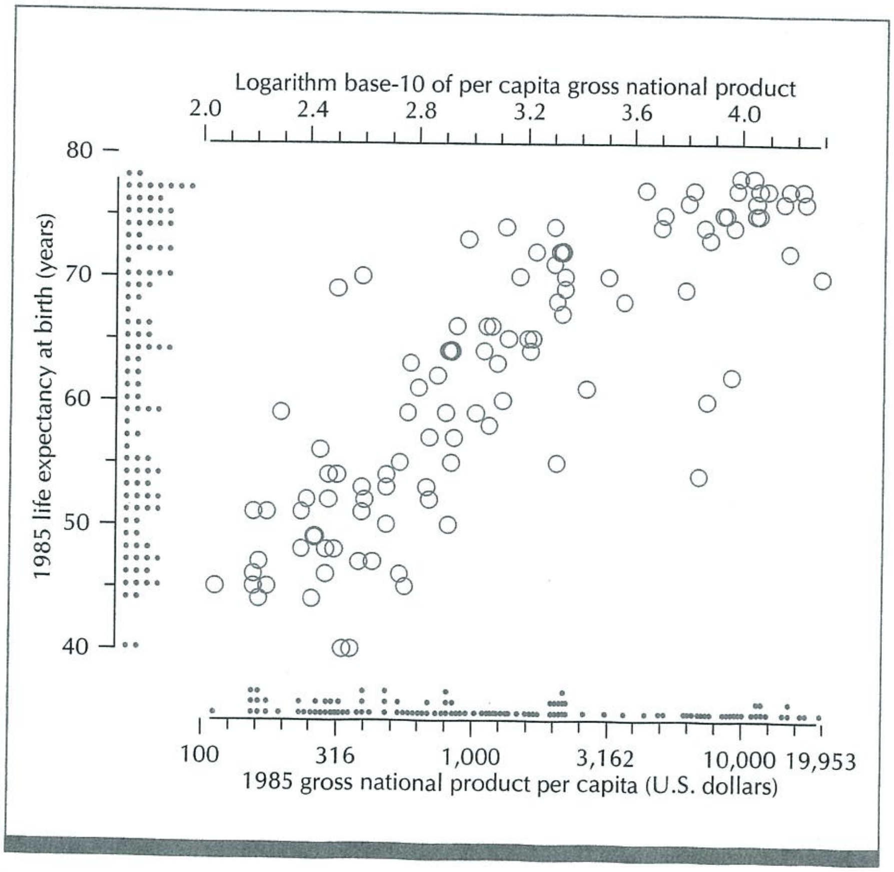

FIGURE 4-10 Scatterplot of 1985 life expectancy at birth and the logarithm of 1985 gross national product per capita for 109 countries. Nineteen countries are excluded because of missing values. The range of values for each variable is indicated by the line on the corresponding axis. A dot plot of the 109 life expectancies is on the left vertical axis. A dot plot of the 109 values of per capita gross national product is on the bottom axis.

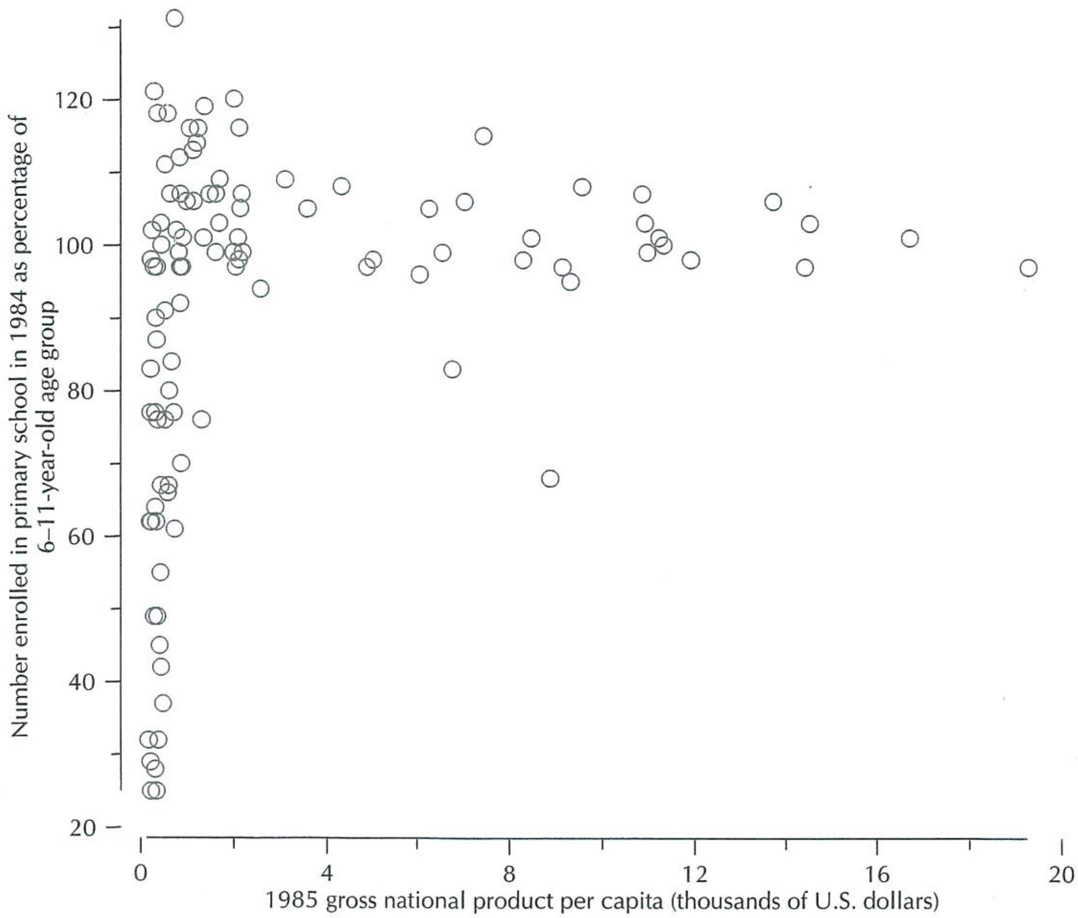

FIGURE 4-11 Scatterplot of number enrolled in primary school in 1984 as percentage of 6–11-year age group and 1985 gross national product per capita for 104 countries. Twenty-four countries are excluded because of missing values. The range of values for each variable is indicated by the line on the corresponding axis.

between two variables. Two quantitative variables can also have a negative or decreasing relationship. Figure 4-13 shows that birth rates decrease with increasing gross national product for the four high-income oil exporters.

Sometimes we see little or no relationship between two variables. Figure 4-14 reveals no meaningful relationship between calorie supply and gross national product for the four high-income oil exporters.

In this chapter we have considered a number of ways to study two variables at a time. In Chapter 5 we examine relationships among three or more variables.

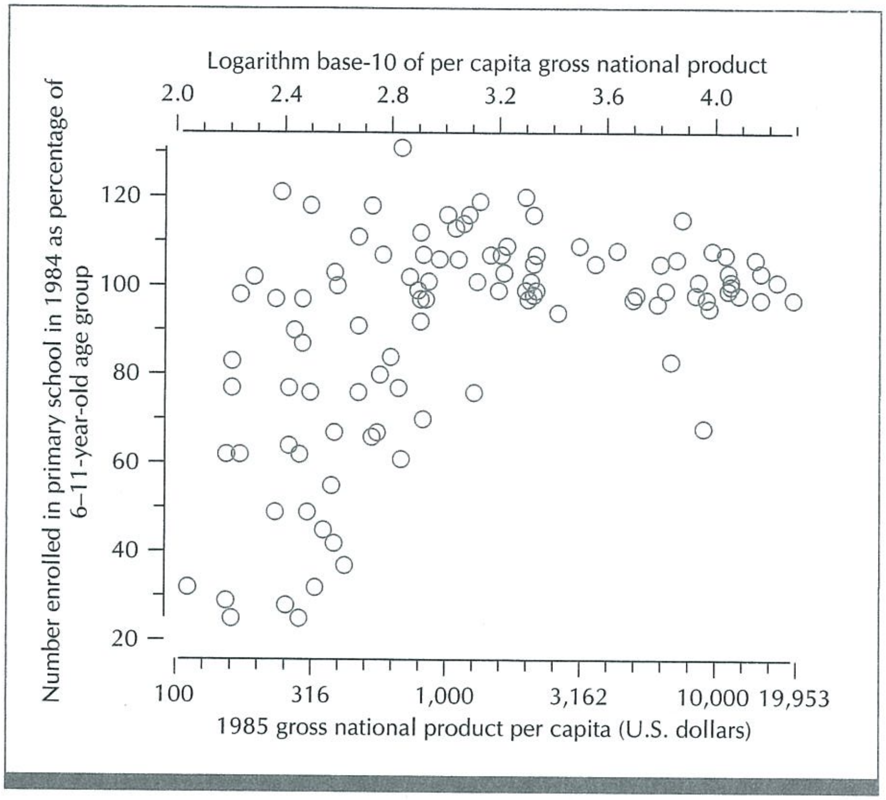

FIGURE 4-12 Scatterplot of number enrolled in primary school in 1984 as percentage of 6–11-year age group and the logarithm of 1985 gross national product per capita for 104 countries. Twenty-four countries are excluded because of missing values. The range of values for each variable is indicated by the line on the corresponding axis.

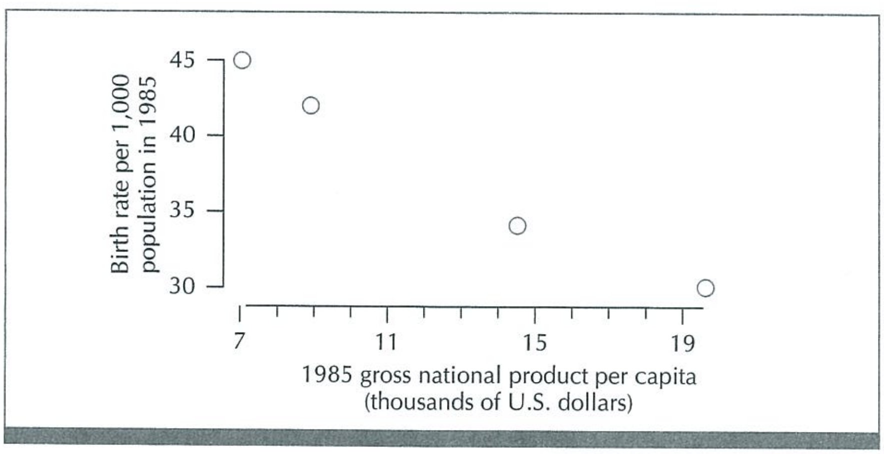

FIGURE 4-13 Scatterplot of birth rate in 1985 and 1985 gross national product per capita for four high-income oil exporting nations. The range of values for each variable is indicated by the line on the corresponding axis.

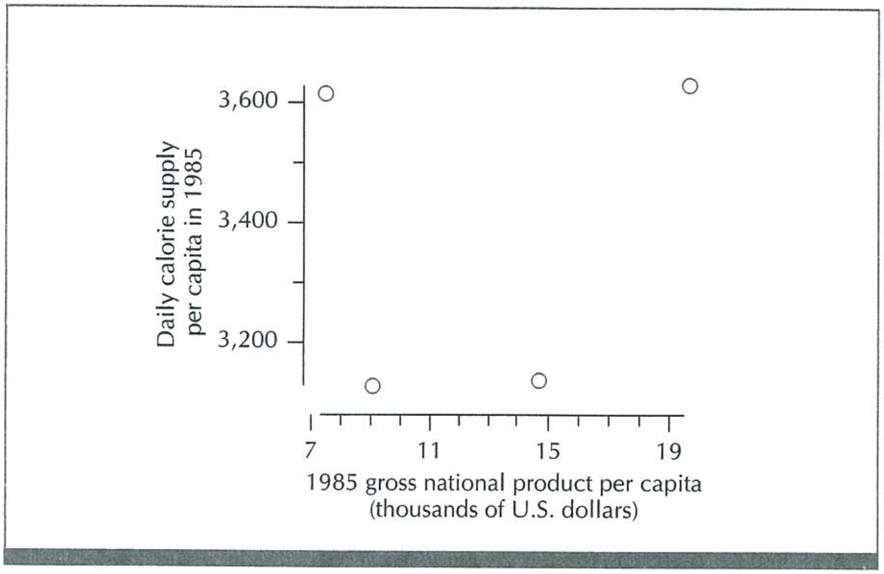

FIGURE 4-14 Scatterplot of daily calorie supply per capita and gross national product per capita in 1985 for four high-income oil exporting nations. The range of values for each variable is indicated by the line on the corresponding axis.

Summary of Chapter 4

Two-way frequency tables (or contingency tables) are tools for examining the relationship between two qualitative variables.

There are a number of ways we can study a quantitative variable at different levels of a qualitative variable. We can tabulate descriptive statistics for the quantitative variable across categories of the qualitative variable. Box plots, histograms, and dot plots allow us to compare the distribution of a quantitative variable across levels of a qualitative variable.

A dot chart displays totals (or percentages) of a quantitative variable for each level of a qualitative variable.

The scatterplot is a valuable graphical tool for looking at the relationship between two quantitative variables. With a range frame, we make efficient use of all the space in a scatterplot. Taking a transformation of a variable can lead to a scatterplot that is easier to look at, or shows a simpler relationship between two variables. Sometimes it is useful to transform both variables before plotting.

Minitab Appendix for Chapter 4

Creating a Two-Way Frequency Table

Recall that in the Minitab Appendix for Chapter 1, we worked with the data for Exercise 5-35. We created a coded variable for child mortality in column 6

```
             -
             -   *   *     *
         2.40+          3  *4* 4
             -                 * *  2
     logmort -
             -
             -
         1.80+
             -
             -
             -
             -
         1.20+
             -                                                  3 35*
             -                                                    **5*
             -
             -
               --------+---------+---------+---------+---------+--------life
                    40.0      48.0      56.0      64.0      72.0
```

FIGURE M4-8 Scatterplot of LOGMORT versus LIFE.

Transforming Data with the LET Command

We can use the LET command to transform data before plotting. The sequence of commands

```
MTB> let c29=logten(c2)
MTB> name c29 'logmort'
MTB> plot 'logmort' 'life'
```

will produce a scatterplot with the logarithm base-10 of child mortality on the vertical axis and life expectancy on the horizontal axis, as shown in Figure M4-8.

Minitab does not have a command for producing a dot chart.

Exercises for Chapter 4

In the exercises, answer the following questions: What would you need to know about the sample to be willing to use it to make inferences about a larger population? What is that larger population (if any)? What limitations do you see in the sample?

For each figure and table, include a legend that completely describes its contents. Note the number of cases included and the number excluded if there are missing values.

Use range frames when drawing scatterplots. When we say plot variable 1 versus variable 2, we mean plot variable 1 on the vertical axis and variable 2 on the horizontal axis.

EXERCISE 4-1

Consider the values of 1985 gross national product per capita and 1985 life expectancy at birth for four high-income oil exporters, shown in Table 4-5.

A scatterplot of life expectancy versus gross national product is displayed in Figure 4-7 for these four countries. In that scatterplot, we used a range frame. Each axis extends from the minimum to the maximum value for the corresponding variable. The axes do not meet. Compare this scatterplot with two other ways to draw a plot:

a. Plot life expectancy versus per capita gross national product for the four high-income oil exporters. In this plot, let each axis extend from 0 to the maximum value for the variable. Let the axes meet at 0.

b. Plot life expectancy versus per capita gross national product for the four high-income oil exporters. Let the horizontal axis go from $7,000 to $20,000. Let the vertical axis go from 55 to 75 years. Let the axes meet (at 55 for life expectancy and $7,000 for per capita gross national product).

c. Discuss the advantages of the scatterplot with range frame in Figure 4-7 over those you drew in parts (a) and (b).

EXERCISE 4-2

The Boston Red Sox spend an estimated $36,754 per minor-league player per year ("The Cost to Develop a Major Leaguer," *The Boston Globe,* June 7, 1989, page 47):

Category of expense	Cost ($)
Player contracts (including meals, equipment, expenses)	14,215
Signing bonuses	6,085
Scouts (salaries)	6,000
Scouts (expenses)	3,723
Managers, coaches, instructors (salaries)	2,885
Minor-league spring training	2,192
Instructors (travel)	554
Player insurance	508
Statistics	269
National association fees	208
Medical	115

Construct a dot chart to show the breakdown of yearly expenses per minor-league player. Show percentage of total as well as dollar costs.

EXERCISE 4-3

Researchers investigated the cancer risk associated with saccharin use in this study. For 2 years, the experimenters fed male rats diets with percent sodium saccharin ranging from 0 to 7.5. A response was development of a bladder tumor in a rat living past the time the first tumor appeared. Dose (percent sodium saccharin in the diet) and response information are shown below (Van Ryzin and Rai, 1987; from Carlborg, 1985).

Dose	Number of responses/ Number of animals tested	Proportion of responses
0	0/324	.000
1	5/658	.008
3	8/472	.017
4	12/189	.063
5	15/120	.125
6.25	20/120	.167
7.5	37/118	.314

Plot the proportion of responses versus dose. How would you describe the dose–response curve based on these plotted points?

EXERCISE 4-4

The accompanying frequency table summarizes the results of a study of 1,224 diabetic people, with each person classified by duration of diabetes and presence of the eye disease retinopathy (Knuiman and Speed, 1988).

Duration of patient's diabetes (years)	Patient has retinopathy			Proportion yes
	Yes	No	Total	
0–2	46	290	336	.14
3–5	52	211	263	.20
6–8	44	134	178	.25
9–11	54	91	145	.37
12–14	38	53	91	.42
15–17	39	42	81	.48
18–20	23	23	46	.50
21 +	52	32	84	.62

a. Plot proportion yes versus duration of diabetes. (You will have to pick a single value to represent a duration interval. You might reasonably use 1 for the interval 0–2, for instance.)

b. Describe the relationship between duration of diabetes and the proportion of diabetic patients with retinopathy.

EXERCISE 4-5

Times to failure (time units not given in the source) are listed below for tires manufactured with three different methods of production (Hsieh, 1986; from Bain, 1978).

Method 1:							
	10.03	10.47	10.58	11.48	11.60	12.41	13.03
	13.51	14.48	16.96	17.08	17.27	17.90	18.21
	19.30	20.10	21.51	21.78	21.79	25.34	

Method 2:	10.10	11.01	11.20	12.95	13.19	14.81	16.03
	17.01	18.96	24.10	24.15	24.52	26.05	26.44
	28.59	30.24	31.03	33.51	33.61	40.68	
Method 3:	19.07	19.51	19.62	20.47	20.78	21.37	22.08
	22.61	23.47	26.02	26.23	26.47	27.07	27.43
	28.28	29.10	29.66	30.67	30.81	34.36	

a. Construct a dot plot of failure times for each production method. Use the same scale for each plot.

b. Compare the three distributions of failure times. In particular, compare location and variation for the three sets of failure times.

EXERCISE 4-6 Field-goal ratio and point-after-touchdown ratio for the 1983 regular season are listed here for 29 National Football League kickers (Berry and Berry, 1985).

Kicker (Team)	Point-after-touchdown Conversions/ Attempts (%)	Field goals Goals/Attempts (%)
Allegre (Baltimore)	22/24 (92)	30/35 (86)
Anderson (Pittsburgh)	38/39 (97)	27/31 (87)
Karlis (Denver)	33/34 (97)	21/25 (84)
Lowery (Kansas City)	44/45 (98)	24/30 (80)
Wersching (San Francisco)	51/51 (100)	25/30 (83)
Haji-Sheikh (New York Giants)	22/23 (96)	35/42 (83)
M. Bahr (Cleveland)	38/40 (95)	21/24 (88)
Septien (Dallas)	57/59 (97)	22/27 (81)
Kempf (Houston)	33/34 (97)	17/21 (81)
Murray (Detroit)	38/38 (100)	25/32 (78)
Stenerud (Green Bay)	52/52 (100)	21/26 (81)
Luckhurst (Atlanta)	43/45 (96)	17/22 (77)
C. Bahr (Los Angeles Raiders)	51/53 (96)	21/27 (78)
M. Andersen (New Orleans)	37/38 (97)	18/23 (78)
Johnson (Seattle)	49/50 (98)	18/25 (72)
Ricardo (Minnesota)	33/34 (97)	25/33 (76)
Moseley (Washington)	62/63 (98)	33/47 (70)
von Schamann (Miami)	45/48 (94)	18/27 (67)
Leahy (New York Jets)	36/37 (97)	16/24 (67)
O'Donoghue (St. Louis)	45/47 (96)	15/28 (54)
Benirschke (San Diego)	43/45 (96)	15/24 (63)
Breech (Cincinnati)	39/41 (95)	16/23 (70)
Franklin (Philadelphia)	24/27 (89)	15/26 (58)
Danelo (Buffalo)	33/34 (97)	10/20 (50)
Thomas (Chicago)	35/38 (92)	14/25 (56)
Smith (New England)	12/15 (80)	3/6 (50)
Capece (Tampa Bay)	23/26 (88)	10/20 (50)
Steinfort (New England and Buffalo)	17/18 (94)	7/21 (33)
Nelson (Los Angeles Rams)	33/37 (89)	5/11 (45)

a. Construct two dot plots: one for percent successful point-after-touchdown attempts and one for percent successful field-goal attempts. Describe and compare the two distributions in terms of location, variation, peaks, and symmetry or skewness.

b. Plot percent successful field-goal attempts versus percent successful point-after-touchdown attempts. Describe the relationship between the two variables shown in the scatterplot.

EXERCISE 4-7

For the 1973 entering class of 15 American law schools, the class average scores on an examination known as the LSAT and the class average of undergraduate grades are shown below (Efron and Tibshirani, 1986).

LSAT	Grades	LSAT	Grades	LSAT	Grades
576	3.39	578	3.03	555	3.00
605	3.13	545	2.76	635	3.30
666	3.44	661	3.43	653	3.12
572	2.88	558	2.81	580	3.07
651	3.36	575	2.74	594	2.96

a. Construct a dot plot of class average LSAT scores and a dot plot of class average undergraduate grades. Describe the distribution of each set of values.

b. Construct a scatterplot of class average LSAT versus class average undergraduate grades. Discuss the information provided in the plot.

EXERCISE 4-8

Wire is wound around plastic spools used in electric motors. When current passes through the wire, the temperature of the spool rises. The accompanying table shows two measurements of temperature rise (°C) made on each of 12 such plastic spools (Nelson, 1986, page 21).

Spool:	1	2	3	4	5	6	7	8	9	10	11	12
First measurement:	45.0	45.1	45.4	45.9	45.9	46.0	46.2	46.5	46.5	46.8	47.0	50.6
Second measurement:	44.9	44.7	45.8	45.3	45.8	45.2	45.2	45.5	46.0	46.1	45.5	50.0

a. Construct a dot plot of the first set of temperature rise measurements. Using the same scale, construct a dot plot of the second set of temperature rise measurements. Describe the two distributions and compare them.

b. Construct a scatterplot of the second measurement versus the first measurement. Sketch the relationship you would expect if the two measurements were perfectly consistent. Do the two measurements appear to be consistent? Discuss the information about the two sets of measurements revealed in the scatterplot.

EXERCISE 4-9

A researcher rated the readability of unpublished reports and published articles written by engineers. A report with a low score is more easily understood. Results are shown below (Milton and Arnold, 1986, page 320; from "Engineers' English" by W. H. Emerson, *CME,* June 1983, pages 54–56).

Unpublished reports: 2.39 2.56 2.36 2.62 2.51 2.29 2.58 2.41 2.86 2.49 2.33 1.94 2.14

Published articles: 1.79 1.87 1.62 1.96 1.75 1.74 2.06 1.69 1.67 1.94 1.33 1.70 1.65

a. Construct a stem-and-leaf plot of scores for unpublished reports and another for published articles. Use the first digits as the stem and the last digit as the leaf.

b. Describe the distribution of each set of values. How do the two distributions compare?

EXERCISE 4-10

Researchers measured airborne bacteria (number of colonies per cubic foot) in eight carpeted hospital rooms and eight uncarpeted hospital rooms (Devore, 1982, page 315; from "Microbial Air Sampling in a Carpeted Hospital," *J. Environmental Health,* 1968, page 405).

Carpeted:	7.1	8.2	10.1	10.8	11.8	13.0	14.0	14.6
Uncarpeted:	3.8	7.2	8.3	10.1	11.1	12.0	12.1	13.7

a. Construct dot plots of the two sets of values, using the same scales for each.

b. Describe the distribution of bacterial counts for the carpeted rooms and for the uncarpeted rooms. Compare the two distributions.

EXERCISE 4-11

Researchers measured carbon monoxide concentration (parts per million) and benzo(a)pyrene concentration (μg per 1,000 cubic meters) in 16 different air samples from Herald Square in New York City (Devore, 1982, page 457; from "Carcinogenic Air Pollutants in Relation to Automobile Traffic in New York City," *Environmental Science and Technology,* 1971, pages 145–150).

Carbon monoxide	Benzo(a)-pyrene	Carbon monoxide	Benzo(a)-pyrene
2.8	.5	5.5	1.3
15.5	.1	12.0	5.7
19.0	.8	5.6	1.5
6.8	.9	19.5	6.0
5.5	1.0	11.0	7.3
5.6	1.1	12.8	8.1
9.6	3.9	5.5	2.2
13.3	4.0	10.5	9.5

a. Construct a dot plot for each of these two variables. Describe each distribution in terms of location, variation, concentrations of values, and symmetry or skewness.

b. Construct a scatterplot of carbon monoxide concentration versus benzo(a)pyrene concentration. Discuss the relationship between the two variables shown in the plot.

EXERCISE 4-12

Sodium content and potassium content (no units given) in perspiration of 10 healthy women are shown here (Oja and Nyblom, 1989; from Johnson and Wichern, 1982, page 182).

Woman:	1	2	3	4	5	6	7	8	9	10
Sodium:	48.5	65.1	47.2	53.2	55.5	36.1	24.8	33.1	47.4	54.1
Potassium:	9.3	8.0	10.9	12.2	9.7	7.9	14.0	7.6	8.5	11.3

a. Construct a dot plot for each of these two sets of values. Describe each distribution.

b. Draw a scatterplot of sodium content versus potassium content. What is the relationship between the two variables revealed in this plot?

EXERCISE 4-13

Maximal oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$) is a measure of lung function and capacity for work. Values of maximal oxygen uptake are listed below for male world-class athletes in their 20's. Several athletes were tested in each of five different sports (Wilmore, 1984).

Wrestling:	58.3	50.4	60.9	64.0	54.3	
Weightlifting:	40.1	42.6	49.5	50.7	46.3	41.5
Shot/discus:	49.5	42.8	42.6	47.5		
Ice hockey:	61.5	54.6	53.6			
Cross-country skiing:	63.9	73.9	78.3	73.0		

a. Using the same scales, construct a dot plot of maximal oxygen uptake for athletes in each of the five sports. Line up the plots under one another for easy comparisons. Discuss and compare these distributions.

b. Find the mean and the median of each of the five sets of values. Find the range, interquartile range, and standard deviation for each set of values. Construct a table showing measures of central tendency and variation, as well as sample size, for each sport. Discuss these measures within each sport and compare sports.

EXERCISE 4-14

Maximal oxygen uptake is a measure of lung function and capacity for work. Values of maximal oxygen uptake (in $ml \cdot kg^{-1} \cdot min^{-1}$) are listed below for female world-class athletes in their teens and early 20's. Several athletes were tested in each of four sports (Wilmore, 1984).

Basketball:	42.3	42.9	49.6
Swimming:	46.2	43.4	40.5

Distance running:	63.2	50.8	57.5	
Volleyball:	43.5	56.0	41.7	50.6

a. Using the same scales, construct a dot plot of maximal oxygen uptake for athletes in each of the four sports. Line up the plots under one another for easy comparisons. Discuss and compare these distributions.

b. Find the mean, median, range, interquartile range, and standard deviation for each of the four sets of values. Construct a table showing measures of location and variation, as well as sample size, for each sport. Discuss these descriptive statistics within each sport and compare sports.

EXERCISE 4-15

Is cigarette smoking associated with delayed conception? To answer this question, investigators studied 586 women who were pregnant with planned pregnancies and had gotten pregnant within 24 cycles of trying. Since oral contraceptives are associated with delayed conception, women whose most recent method of birth control had been the pill were not included. A woman was classified as a smoker if she smoked on average at least one cigarette a day during at least the first cycle she was trying to get pregnant. The accompanying frequency table classifies women by their smoking status and number of cycles to pregnancy (Weinberg and Gladen, 1986; Baird and Wilcox, 1985):

	Nonsmokers		Smokers	
Cycle	Number	(Percent)	Number	(Percent)
1	198	(40.7)	29	(29)
2	107	(22.0)	16	(16)
3	55	(11.3)	17	(17)
4	38	(7.8)	4	(4)
5	18	(3.7)	3	(3)
6	22	(4.5)	9	(9)
7	7	(1.4)	4	(4)
8	9	(1.9)	5	(5)
9	5	(1.0)	1	(1)
10	3	(.6)	1	(1)
11	6	(1.2)	1	(1)
12	6	(1.2)	3	(3)
More than 12	12	(2.5)	7	(7)
Total	486	(100)	100	(100)

a. Construct a frequency plot showing the number of nonsmokers in each cycle category. Using the same scale for frequencies, construct a frequency plot showing the number of smokers in each cycle category.

b. Construct a frequency plot showing the percentage of nonsmokers in each cycle category. Using the same scale for percentages, construct a frequency plot showing the percentage of smokers in each cycle category.

c. What information is provided in the two plots you constructed in part (a)? Compare this with the information provided in the two plots from part (b). What do these plots suggest about the relationship between smoking and delayed conception?

d. Suppose you are concerned that some women reporting pregnancy in the first cycle might have become pregnant by accident. Exclude the first cycle category from your plots. Do your interpretations change?

EXERCISE 4-16 Researchers are investigating an antibody known as 64K autoantibody as a possible early warning of Type I diabetes (*Science News,* volume 133, June 18, 1988, page 389). In one study, researchers found 64K autoantibodies in 18 of 20 patients newly diagnosed with Type I diabetes. They found the antibodies in none of 18 controls (people without diabetes). Display these results in a two-way frequency table. What do these results suggest?

EXERCISE 4-17 Will substances that cause cancer in people cause cancer in mice? Will substances that cause cancer in mice cause cancer in people? These questions are important because scientists routinely use animal studies to evaluate possible carcinogenicity of chemicals in humans (*Statistical Science,* volume 3, 1988, pages 3–56). An evaluation of 266 chemicals tested in rats and in mice by the National Cancer Institute and the National Toxicology Program yielded the following results (*Statistical Science,* volume 3, 1988, page 34; from "Species Correlation in Long-Term Carcinogenicity Studies," by J. K. Haseman and J. E. Huff, *Cancer Lett.,* volume 37, 1987, pages 125–132):

	Rats	
Mice	Carcino-genic	Not car-cinogenic
Carcinogenic	67	36
Not carcinogenic	32	131

a. For what percentage of the 266 chemicals do the rats and mice agree with respect to carcinogenicity (or lack of carcinogenicity)?

b. Of the chemicals carcinogenic for at least one species, what percentage was carcinogenic in both?

c. What do these results suggest about agreement between humans and mice (or rats) with respect to potential cancer-causing agents?

EXERCISE 4-18 In the United States in 1984–1985, bachelor's degrees were received by 12,402 Hispanic men, 13,472 Hispanic women, 23,018 black men, 34,455 black women, 405,085 white men, 421,021 white women, 13,554 Asian/Pacific Islander men, 11,841 Asian/Pacific Islander women, 1,998 American Indian men, 2,248 American Indian women, 20,091 nonresident alien men, and 9,126 nonresident alien women (American Council on Education, 1987, page 19).

a. Arrange this information in a two-way frequency table showing number of degrees by race and sex.

b. Construct two frequency plots (or dot charts) of bachelor's degrees by race—one for men and one for women. Use scales that allow the best comparisons of the two plots.

c. Display the data in any other way you find informative.

d. Discuss your findings.

EXERCISE 4-19

Researchers in Martinique found that 10 of 17 patients with tropical spastic paraparesis (a common paralytic disease in the tropics) had antibodies to HTLV-I (human T-cell lymphotropic virus I, a virus associated with some lymphomas and leukemias). Twelve of 303 controls (people without tropical spastic paraparesis) had antibodies to HTLV-I. (Numbers with antibodies were calculated from percentages reported in *Science,* volume 236, May 29, 1987, page 1059.)

a. Arrange these results in a two-way frequency table showing patient group and presence/absence of HTLV-I antibodies.

b. Discuss the possible implications of these findings.

EXERCISE 4-20

The snowberry fly has markings on its wings that make it resemble one of its predators, the zebra spider. Do these markings discourage zebra spiders from attacking snowberry flies? Investigators carried out an experiment to determine the reaction of zebra spiders to four types of potential prey: another spider, a housefly, a snowberry fly, and a snowberry fly with blackened wings (special markings obscured). The table shows the number of trials in which a predator spider pounced on the potential prey. (The numbers in the last column were calculated from percentages given in Mather and Roitberg, 1987.)

Prey	Number of trials	Number of trials in which predator spider pounced on prey
Spider	40	2
Housefly	40	24
Snowberry fly	76	15
Blackened wing snowberry fly	33	13

a. Arrange these results in a two-way frequency table.

b. Do these findings suggest that the markings of the snowberry fly are protective against attacks by zebra spiders?

EXERCISE 4-21 A study of Rhode Island 12th graders identified 59 females and 74 males who were academically prepared for science studies, with course work including calculus and physics. When interviewed, 11 of the 59 females and 47 of the 74 males expressed an interest in a career in engineering, science, or technology (*Science,* volume 236, May 8, 1987, page 660).

a. Display these results in a two-way frequency table.

b. Discuss the relationship between sex and science career interest in this group of students.

EXERCISE 4-22 A large group of domestic cats (in the United States and Canada) considered to be at high risk of infection with feline immunodeficiency virus were tested for presence of the virus. Of 663 female cats tested, 51 were positive for the virus. Of 855 male cats tested, 168 were positive (Yamamoto et al., 1989).

a. Arrange these results in a two-way frequency table.

b. Does there appear to be an association between sex and presence of the virus in this group of high-risk cats?

EXERCISE 4-23 Is obesity related to abnormal liver function tests? Researchers studied 39 people who were at least 11% above their ideal body weight and had abnormal results of liver function tests (*Science News,* volume 135, May 27, 1989, page 332). None of the volunteers had problems such as alcohol abuse that might contribute to liver problems. The researchers gave the volunteers a diet and exercise program to follow. At the end of the study period (about a year and a half), four volunteers had gained weight; all four still had abnormal liver function tests. Eighteen volunteers had lost less than 10% of their body weight; 11 of the 18 still had abnormal liver function tests. Seventeen volunteers had lost more than 10% of their body weight; 4 of the 17 still had abnormal liver function tests.

a. Arrange these results in a two-way frequency table to show the relationship between extent of weight loss and presence of abnormal liver function tests.

b. What do these results suggest about the relationship between obesity and evidence of abnormal liver function?

EXERCISE 4-24 Injuries treated at the University of Rochester Section of Sports Medicine from May 1975 to July 1983 are classified below by sex of the injured person and site of the injury (DeHaven and Lintner, 1986):

Site	Males	Females
Knee	1,157	401
Ankle	326	89
Shoulder/upper arm	224	15
Hand/finger	151	20
Hip/thigh	126	11
Elbow/forearm	107	28
Foot/toe	72	20

a. Tabulate and/or plot these results in any reasonable way.

b. Discuss the relationship between sex and site of injury in this group of patients.

EXERCISE 4-25 In a study of smokeless tobacco use among high school students in two Arkansas communities, researchers classified 901 students by use and by grade (Marty, McDermott, and Williams, 1986):

	Grade 10	Grade 11	Grade 12
Uses smokeless tobacco	45	68	58
Does not use smokeless tobacco	281	262	187

Of the 171 users of smokeless tobacco, 162 provided information on duration of use:

Duration of use (years)	<1	1–2	2–3	3–4	4–5	>5
Number of users	26	35	26	22	28	25

One hundred seventy provided information on frequency of use:

Frequency of use (days/week)	≤1	2–3	4–5	6–7
Number of users	31	20	20	99

One hundred seventy provided information on extent of daily use:

Number of dips or chews/day	1	2–3	4–5	6–7	8–9	≥10
Number of users	30	51	36	24	8	21

a. Tabulate and/or plot these results in any reasonable way.

b. Is there an association between grade in school and use of smokeless tobacco among this group of students?

c. Construct a frequency table and frequency plot of each of these variables: duration of use, frequency of use, and number of dips or chews per day. Discuss your findings.

EXERCISE 4-26 Use the information in the accompanying table to look for a possible association between the age of the mother and the likelihood of a baby with Down's syndrome. (Data contributed by P. A. P. Moran to the collection of problems in Andrews and Herzberg, 1985, pages 221–222. The data, for births in Australia from 1942 to 1952, originally appeared in "A Survey of Mongoloid Births in Victoria, Australia," by R. D. Collman and A. Stoller, *American Journal of Public Health,* volume 57, 1962, pages 813–829.)

Age of mother (years)	"Center" of age interval (years)	Number of births	Number of mothers of babies with Down's syndrome	Down's ratio
20 or younger	17.5	35,555	15	.00042
Over 20, under 25	22.5	207,931	128	.00062
At least 25, under 30	27.5	253,450	208	.00082
At least 30, under 35	32.5	170,970	194	.00113
At least 35, under 40	37.5	86,046	297	.00345
At least 40, under 45	42.5	24,498	240	.00980
45 or older	47.5	1,707	37	.02168

Let age of mother refer to the "center" of the age interval in column 2 of the table. Let Down's ratio refer to the number of mothers of babies with Down's syndrome divided by the total number of births, shown in the last column. You may wish to multiply each Down's ratio by 10,000 (or 100,000) before answering the following questions.

a. Construct a scatterplot of Down's ratio versus age of mother.

b. Construct a scatterplot of the logarithm of Down's ratio versus age of mother.

c. Construct a scatterplot of Down's ratio versus the logarithm of age of mother.

d. Construct a scatterplot of the logarithm of Down's ratio versus the logarithm of age of mother.

e. Discuss the apparent relationship between maternal age and Down's ratio, as shown in parts (a)–(d).

f. Does taking the logarithm of one or both variables help to illustrate the relationship between the two variables?

g. Discuss limitations of this data set and the scatterplots you constructed.

EXERCISE 4-27 In this experiment investigators studied the response of male beetles to an airborne sex pheromone (Nordheim, Tsiatis, and Shapas, 1983). The experimenters considered four dose rates (units not given), with 30 beetles exposed at each dose rate. The number of beetles responding within 60 seconds is shown below for each group:

Dose rate	Number of beetles responding within 60 seconds
10^{-6}	2
10^{-5}	10
10^{-4}	17
10^{-3}	25

Display these experimental results in a two-way frequency table. Plot the data in an informative way. Discuss your findings.

EXERCISE 4-28 In 1982 in Western Australia, 1,317 males and 854 females died of ischaemic heart disease, 1,119 males and 828 females died of cancer, 371 males and 460 females died of cerebral vascular disease, and 346 males and 147 females died from accidents (Hatton and Clarke-Hundley, 1984, page 44). Display these results in tabular and/or graphical form. Discuss your findings.