

# Chapter 8

# Graphics

*KISS—Keep It Simple, but Scientific*  
*Emanuel Parzen*

What is the dimension of the information you will illustrate? Do you need to illustrate repeated information for several groups? Is a graphical illustration the best vehicle for communicating information to the reader? How do you select from a list of competing choices? How do you know whether the graphic you produce is effectively communicating the desired information?

**GRAPHICS SHOULD EMPHASIZE AND HIGHLIGHT SALIENT FEATURES. THEY** should reveal data properties and make large quantities of information coherent. While graphics provide the reader a break from dense prose, authors must not forget that their illustrations should be scientifically informative as well as decorative. In this chapter, we outline mistakes in selection, creation, and execution of graphics and discuss improvements for each of these three areas.

Graphical illustrations should be simple and pleasing to the eye, but the presentation must remain scientific. In other words, we want to avoid those graphical features that are purely decorative while keeping a critical eye open for opportunities to enhance the scientific inference we expect from the reader. A good graphical design should maximize the proportion of the ink used for communicating scientific information in the overall display.

## THE SOCCER DATA

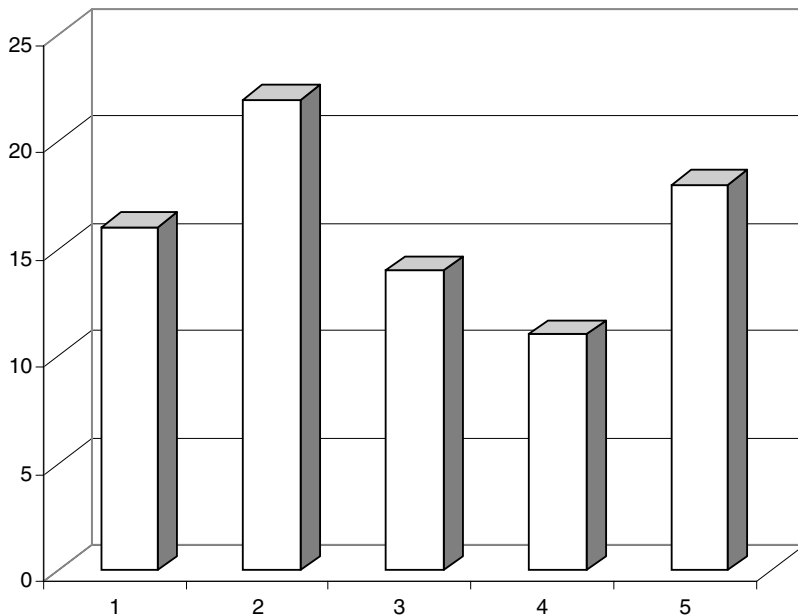
Dr. Hardin coaches youth soccer (players of age 5) and has collected the total number of goals for the top five teams during the eight-game spring 2001 season. The total number of goals scored per team was 16 (team 1), 22 (team 2), 14 (team 3), 11 (team 4), and 18 (team 5). There are many

ways we can describe this set of outcomes to the reader. In text above, we simply communicated the results in text.

A more effective presentation would be to write that the total number of goals scored by teams 1 through 5 was 16, 22, 14, 11, and 18, respectively. The College Station Soccer Club labeled the five teams as Team 1, Team 2, and so on. These labels show the remarkable lack of imagination that we encounter in many data collection efforts. Improving on this textual presentation, we could also say that the total number of goals with the team number identified by subscript was  $22_2$ ,  $18_5$ ,  $16_1$ ,  $14_3$ , and  $11_4$ . This presentation better communicates with the reader by ordering the outcomes because the reader will naturally want to know the order in this case.

## FIVE RULES FOR AVOIDING BAD GRAPHICS

There are a number of choices in presenting the soccer outcomes in graphical form. Many of these are poor choices; they hide information, make it difficult to discern actual values, or inefficiently use the space within the graph. Open almost any newspaper and you will see a bar chart graphic similar to Figure 8.1 illustrating the soccer data. In this section,



**FIGURE 8.1** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* The false third dimension makes it difficult to discern values. The reader must focus on the top of the obscured back face to accurately interpret the values plotted.

we illustrate five important rules for generating correct graphics. Subsequent sections will augment this list with other specific examples.

Figure 8.1 includes a false third dimension; a depth dimension that does not correspond to any information in the data. Furthermore, the resulting figure makes it difficult to discern the actual values presented. Can you tell by looking at Figure 8.1 that Team 3 scored 14 goals, or does it appear that they scored 13 goals? The reader must focus on the top back corner of the three-dimensional rectangle since that part of the three-dimensional bar is (almost) at the same level as the grid lines on the plot; actually, the reader must first focus on the floor of the plot to initially discern the vertical distance of the back right corner of the rectangular bar from the corresponding grid line at the back (these are at the same height). The viewer must then mentally transfer this difference to the top of the rectangular bars in order to accurately infer the correct value. The reality is that most people focus on the front face of the rectangle and will subsequently misinterpret this data representation.

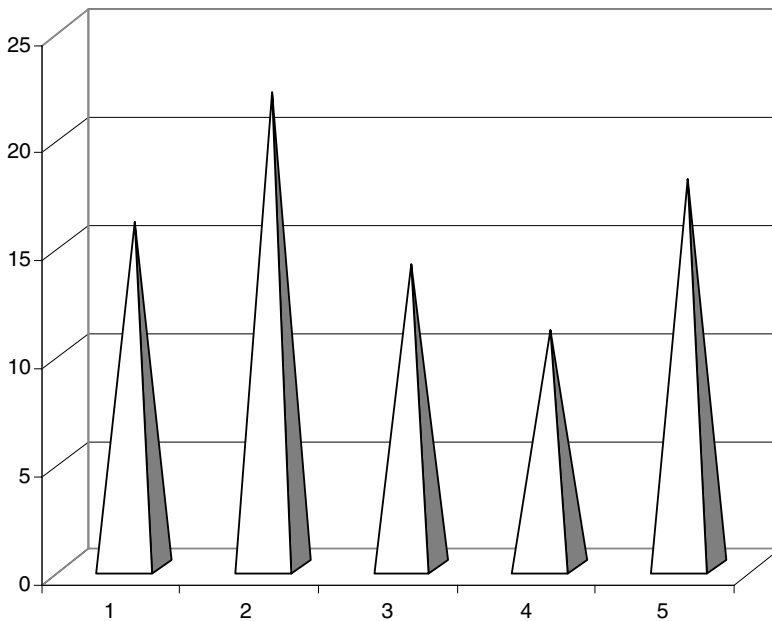
Figure 8.2 also includes a false third dimension. As before, the resulting illustration makes it difficult to discern the actual values presented. This illusion is further complicated by the fact that the depth dimension has been eliminated at the top of the three-dimensional pyramids so that it's nearly impossible to correctly ascertain the plotted values. Focus on the result of Team 4, compare it to the illustration in Figure 8.1, and judge whether you think the plots are using the same data (they are). Other types of plots that confuse the audience with false third dimensions include point plots with shadows and line plots where the data are connected with a three dimensional line or ribbon.

The lesson from these first two graphics is that we must avoid illustrations that utilize more dimensions than exist in the data. Clearly, a better presentation would indicate only two dimensions where one dimension identifies the teams and the other dimension identifies the number of goals scored.

*Rule 1: Don't produce graphics illustrating more dimensions than exist in the data.*

Figure 8.3 is an improvement over three-dimensional displays. It is easier to discern the outcomes for the teams, but the axis label obscures the outcome of Team 4. Axes should be moved outside of the plotting area with enough labels so that the reader can quickly scan the illustration and identify values.

*Rule 2: Don't superimpose labeling information on the graphical elements of interest. Labels can add information to the plot, but*

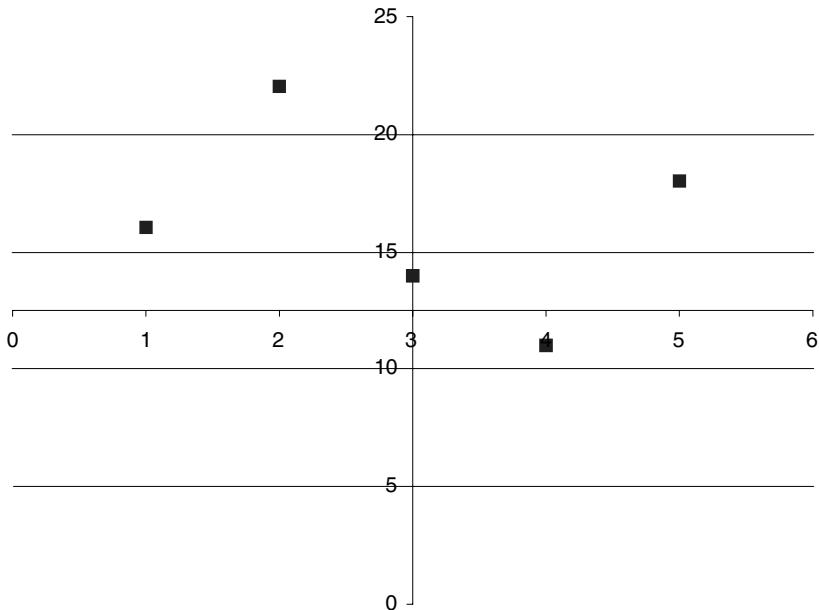


**FIGURE 8.2** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* The false third dimension makes it difficult to discern the values in the plot. Since the back face is the most important for interpreting the values, the fact that the decorative object comes to a point makes it impossible to correctly read values from the plot.

*should be placed in (otherwise) unused portions of the plotting region.*

Figure 8.4 is a much better display of the information of interest. The problem illustrated is that there is too much empty space in the graphic. Choosing to begin the vertical axis at zero means that about 40% of the plotting region is empty. Unless there is a scientific reason compelling you to include a specific baseline in the graph, the presentation should be limited to the range of the information at hand. There are several instances where axis range can exceed the information at hand, and we will illustrate those in a presentation.

**Rule 3:** *Don't allow the range of the axes labels to significantly decrease the area devoted to data presentation. Choose axis limits wisely and do not automatically accept default values for the axes that are far outside of the range of data.*

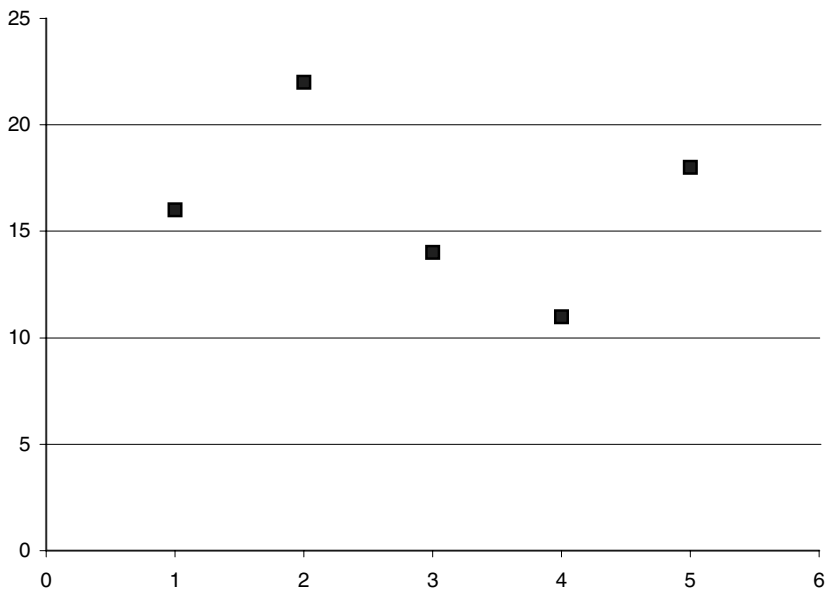


**FIGURE 8.3** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* Placing the axes inside of the plotting area effectively occludes data information. This violates the simplicity goal of graphics; the reader should be able to easily see all of the numeric labels in the axes and plot region.

Figure 8.5 eliminates the extra space included in Figure 8.4 where the vertical axis is allowed to more closely match the range of the outcomes. The presentation is fine, but could be made better. The data of interest in this case involve a continuous and a categorical variable. This presentation treats the categorical variable as numeric for the purposes of organizing the display, but this is not necessary.

*Rule 4: Carefully consider the nature of the information underlying the axes. Numeric axis labels imply a continuous range of values that can be confusing when the labels actually represent discrete values of an underlying categorical variable.*

Figures 8.5 and 8.6 are further improvements of the presentation. The graph region, area of the illustration devoted to the data, is illustrated with axes that more closely match the range of the data. Figure 8.6 connects the point information with a line that may help visualize the difference between the values, but also indicates a nonexistent relationship; the

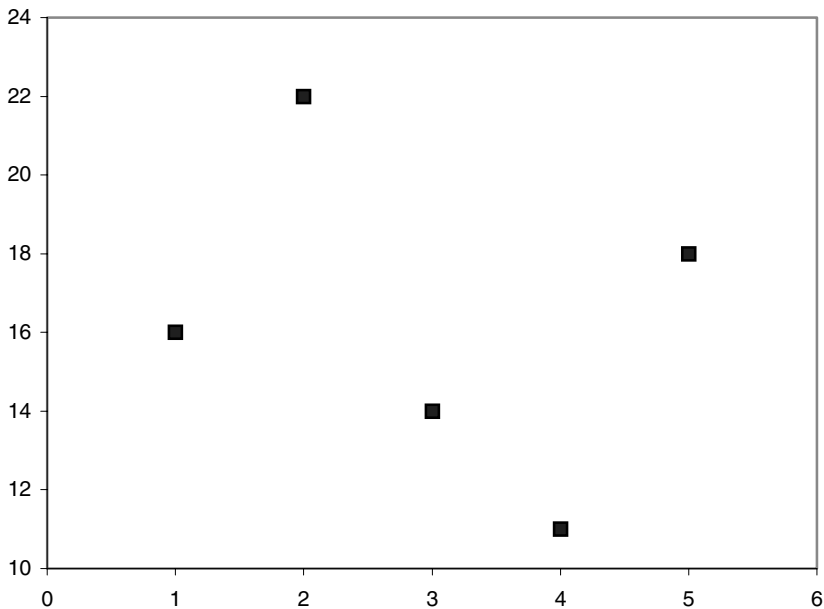


**FIGURE 8.4** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* By allowing the  $y$  axis to range from zero, the presentation reduces the proportion of the plotting area in which we are interested. Less than half of the vertical area of the plotting region is used to communicate data.

horizontal axis is discrete rather than continuous. Even though these presentations vastly improve the illustration of the desired information, we are still using a two-dimensional presentation. In fact, our data are not really two-dimensional and the final illustration more accurately reflects the true nature of the information.

*Rule 5: Do not connect discrete points unless there is either (a) a scientific meaning to the implied interpolation or (b) a collection of profiles for group level outcomes.*

Rules 4 and 5 are aimed at the practice of substituting numbers for labels and then treating those numeric labels as if they were in fact numeric. Had we included the word “Team” in front of the labels, there would be no confusion as to the nature of the labels. Even when nominative labels are used on an axis, we must consider the meaning of values between the labels. If the labels are truly discrete, data outcomes should not be connected or they may be misinterpreted as implying a continuous rather than discrete collection of values.



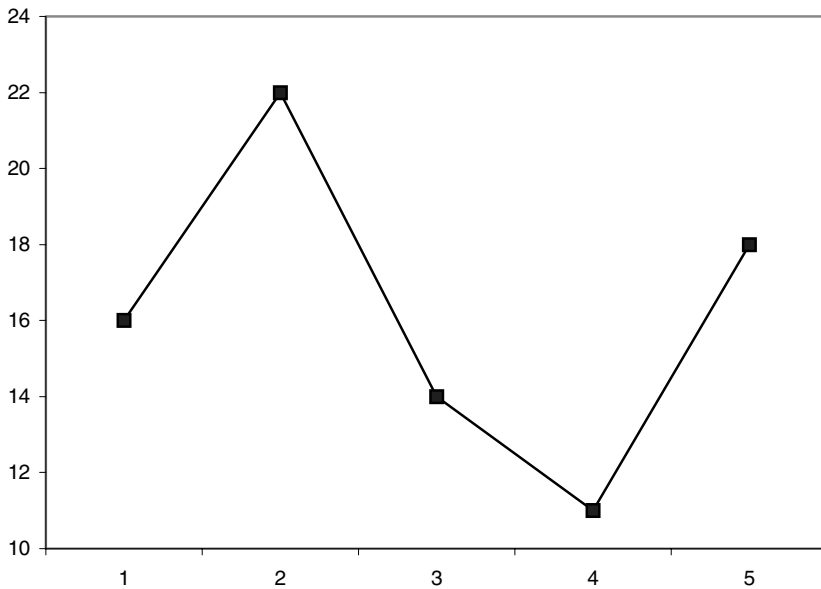
**FIGURE 8.5** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* This graph correctly scales the  $y$  axis, but still uses a categorical variable denoting the team on the  $x$  axis. Labels 0 and 6 do not correspond to a team number and the presentation appears as if the  $x$  axis is a continuous range of values when in fact it is merely a collection of labels. While a reasonable approach to communicating the desired information, we can still improve on this presentation by changing the numeric labels on the  $x$  axis to String labels corresponding to the actual team names.

Figure 8.7 is the best illustration of the soccer data. There are no false dimensions, the range of the graphic is close to the range of the data, there is no difficulty interpreting the values indicated by the plotting symbols, and the legend fully explains the material. Alternatively, we can produce a simple table.

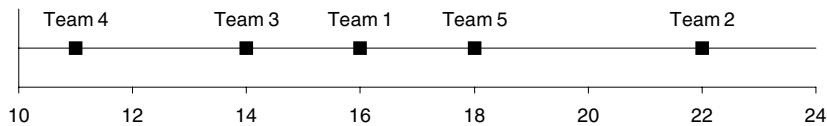
Table 8.1 succinctly presents the relevant information. Tables and figures have the advantage over in-text descriptions that the information is more easily found while scanning through the containing document. If the information is summary in nature, we should make that information easy to find for the reader and place it in a figure or table. If the information is ancillary to the discussion, it can be left in text.

### Choosing Between Tabular and Graphical Presentations

In choosing between tabular and graphical presentations, there are two issues to consider: the size (density) of the resulting graphic and the scale



**FIGURE 8.6** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates the team number, and the  $y$  axis indicates the number of goals scored by the respective team. *Problem:* The inclusion of a polyline connecting the five outcomes helps the reader to visualize changes in scores. However, the categorical values are not ordinal, and the polyline indicates an interpolation of values that does not exist across the categorical variable denoting the team number. In other words, there is no reason that Team 5 is to the right of Team 3 other than we ordered them that way, and there is no Team 3.5 as the presentation seems to suggest.



**FIGURE 8.7** Total Number of Goals Scored by Teams 1 through 5. The  $x$  axis indicates with a square the number of goals scored by the respective team. The associated team name is indicated above the square. Labeling the outcomes addresses the science of the KISS specification given at the beginning of the chapter.

of the information. If the required number of rows for a tabular presentation would require more than one page, the graphical representation is preferred. Usually, if the amount of information is small, the table is preferred. If the scale of the information makes it difficult to discern otherwise significant differences, a graphical presentation is better.



**TABLE 8.1 Total Number of Goals Scored by Teams 1 through 5 Ordered by Lowest Total to Highest Total<sup>a</sup>**

Team 4	Team 3	Team 1	Team 5	Team 2
11	14	16	18	22

<sup>a</sup> These totals are for the Spring 2001 season. The organization of the table correctly sorts on the numeric variable. That the team labels are not sorted is far less important since these labels are merely nominal; were it not for the fact that we labeled with integers, the team names would have no natural ordering.

## ONE RULE FOR CORRECT USAGE OF THREE-DIMENSIONAL GRAPHICS

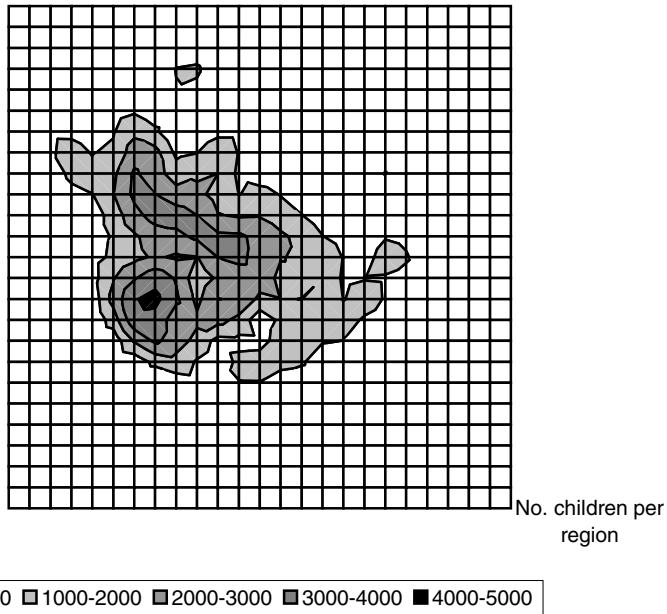
As illustrated in the previous section, the introduction of superfluous dimensions in graphics should be avoided. The prevalence of turnkey solutions in software that implement these decorative presentations is alarming. At one time, these graphics were limited to business-oriented software and presentations, but this is no longer true. Misleading illustrations are starting to appear in scientific talks. This is partly due to the introduction of business-oriented software in university service courses (demanded by the served departments). Errors abound when increased license costs for scientific- and business-oriented software lead departments to eliminate the more scientifically oriented software packages.

The reader should not necessarily interpret these statements as a mandate to avoid business-oriented software. Many of these maligned packages are perfectly capable of producing scientific plots. Our warning is that we must educate ourselves in the correct software specifications.

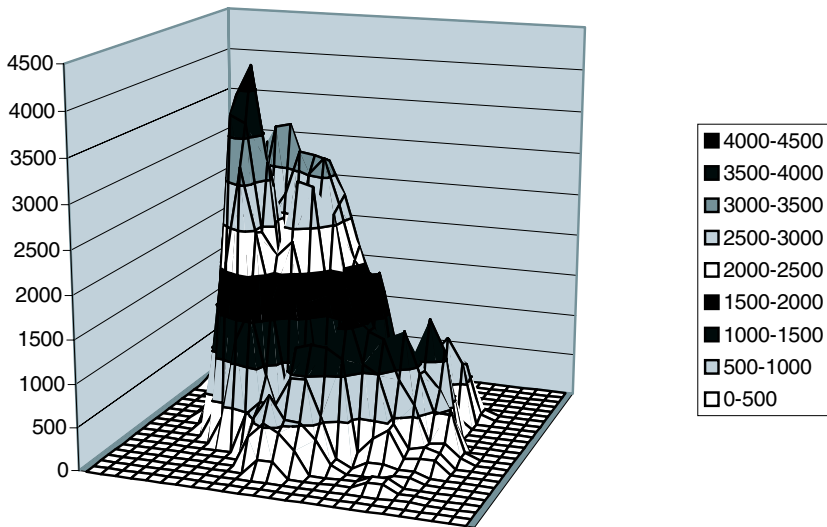
Three-dimensional perspective plots are very effective, but require specification of a viewpoint. Experiment with various viewpoints to highlight the properties of interest. Mathematical functions lend themselves to three-dimensional plots, but raw data are typically better illustrated with contour plots. This is especially true for map data, such as surface temperatures, or surface wind (where arrows can denote direction and the length of the arrow can denote the strength).

In Figures 8.8 and 8.9, we illustrate population density of children for Harris County, Texas. Illustration of the data on a map is a natural approach, and a contour plot reveals the pockets of dense and sparse populations.

While the contour plot in Figure 8.8 lends itself to comparison of maps, the perspective plot in Figure 8.9 is more difficult to interpret. The surface is more clearly illustrated, but the surface itself prevents viewing all of the data.



**FIGURE 8.8** Distribution of Child Population in Harris County, Texas. The  $x$  axis is the longitude ( $-96.04$  to  $-94.78$  degrees), and the  $y$  axis is the latitude ( $29.46$  to  $30.26$  degrees).



**FIGURE 8.9** Population Density of the Number of Children in Harris County, Texas. The  $x$  axis is the longitude ( $-96.04$  to  $-94.78$  degrees), and the  $y$  axis is the latitude ( $29.46$  to  $30.26$  degrees). The  $x$ - $y$  axis is rotated 35 degrees from Figure 8.10.

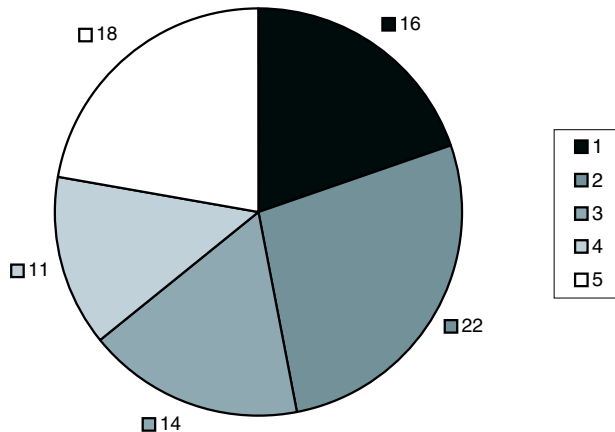
*Rule 6: Use a contour plot over a perspective plot if a good view-point is not available. Always use a contour plot over the perspective plot when the axes denote map coordinates.*

Though the contour plot is generally a better representation of mapped data, a desire to improve Figure 8.8 would lead us to suggest that the grid lines should be drawn in a lighter font so that they have less emphasis than lines for the data surface. Another improvement to data illustrated according to real-world maps is to overlay the contour plot where certain known places or geopolitical distinctions may be marked. The graphic designer must weigh the addition of such decorative items with the improvement in inference that they bring.

### ONE RULE FOR THE MISUNDERSTOOD PIE CHART

The pie chart is undoubtedly the graphical illustration with the worst reputation. Wilkinson (1999) points out that the pie chart is simply a bar chart that has been converted to polar coordinates.

Focusing on Wilkinson's point makes it easier to understand that the conversion of the bar height to an angle on the pie chart is most effective when the bar height represents a proportion. If the bars do not have values where the sum of all bars is meaningful, the pie chart is a poor choice for presenting the information (cf. Figure 8.10).



**FIGURE 8.10** Total Number of Goals Scored by Teams 1 through 5. The legend indicates the team number and associated slice color for the number of goals scored by the respective team. The actual number of goals is also included. *Problem:* The sum of the individual values is not of interest so that the treatment of the individuals as proportions of a total is not correct.

*Rule 7: Do not use pie charts unless the sum of the entries is scientifically meaningful and of interest to the reader.*

On the other hand, the pie chart is an effective display for illustrating proportions. This is especially true when we want to focus on a particular slice of the graphic that is near 25% or 50% of the data since we humans are adept at judging these size portions. Including the actual value as a text element decorating the associated pie slice effectively allows us to communicate both the raw number along with the visual clue of the proportion of the total that the category represents. A pie chart intended to display information on all sections where some sections are very small is very difficult to interpret. In these cases, a table or bar chart is to be preferred.

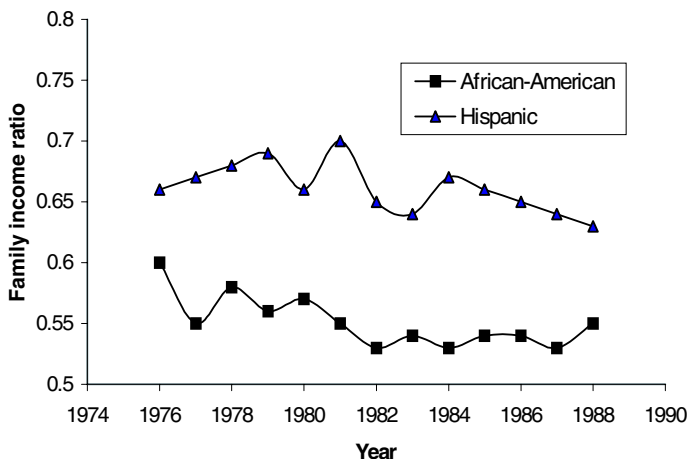
Additional research has addressed whether the information should be ordered before placement in the pie chart display. There are no general rules to follow other than to repeat that humans are fairly good at identifying pie shapes that are one-half or one-quarter of the total display. As such, a good ordering of outcomes that included such values would strive to place the leading edge of 25% and 50% pie slices along one of the major north–south or east–west axes. Reordering the set of values may lead to confusion if all other illustrations of the data used a different ordering, so the graphic designer may ultimately feel compelled to reproduce other illustrations.

### **THREE RULES FOR EFFECTIVE DISPLAY OF SUBGROUP INFORMATION**

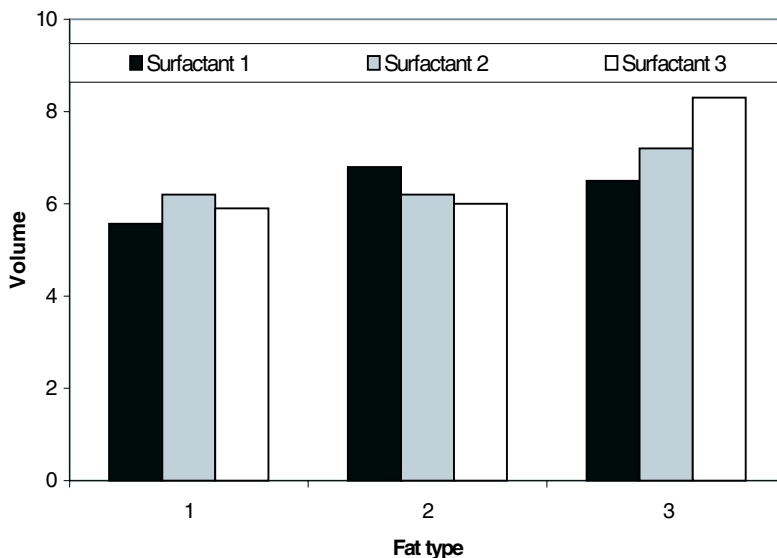
Graphical displays are very effective for communication of subgroup information—for example, when we wish to compare changes in median family income over time of African-Americans and Hispanics. With a moderate number of subgroups, a graphical presentation can be much more effective than a similar tabular display. Labels, stacked bar displays, or a tabular arrangement of graphics can effectively display subgroup information. Each of these approaches has its limits, as we will see in the following sections.

In Figure 8.11, separate connected polylines easily separate the subgroup information. Each line is further distinguished with a different plotting symbol. Note how easy it is to confuse the information due to the inverted legend. To avoid this type of confusion, ensure that the order of entries (top to bottom) matches that of the graphic.

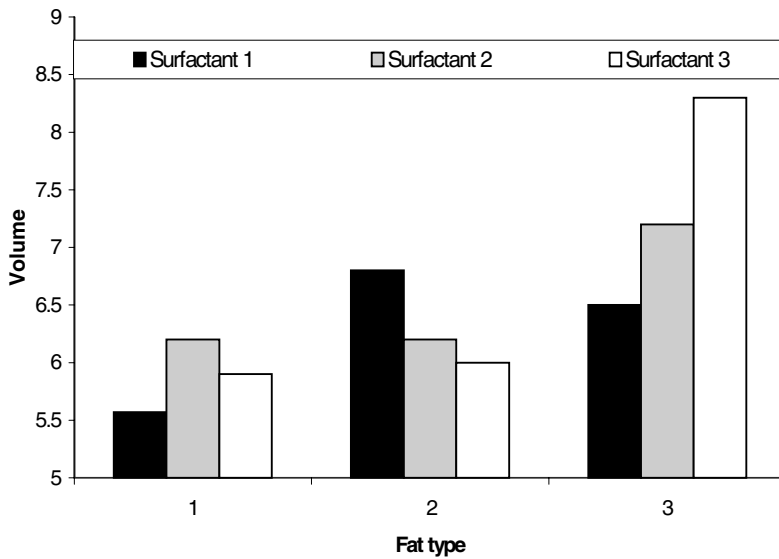
*Rule 8: Put the legend items in the same order they appear in the graphic whenever possible.*



**FIGURE 8.11** Median Family Income of African-Americans and Hispanics Divided by the Median Family Income for Anglo-American Families for Years 1976–1988. *Problem:* The legend identifies the two ethnic groups in the reverse order that they appear in the plot. It is easy to confuse the polylines due to the discrepancy in organizing the identifiers. The rule is that if the data follow a natural ordering in the plotting region, the legend should honor that order.



**FIGURE 8.12** Volume of a Mixture Based on the Included fat and Surfactant Types. *Problem:* As with a scatterplot, the arbitrary decision to include zero on the  $y$  axis in a bar plot detracts from the focus on the values plotted.



**FIGURE 8.13** Volume of a Mixture Based on the Included fat and Surfactant Types. Drawing the bar plot with a more reasonable scale clearly distinguishes the values for the reader.

Clearly, there are other illustrations that would work even better for this particular data. When one subgroup is always greater than the other subgroup, we can use vertical bars between each measurement instead of two separate polylines. Such a display not only points out the discrepancies in the data, but also allows easier inference as to whether the discrepancy is static or changes over time.

The construction of a table such as Table 8.2 effectively reduces the number of dimensions from two to one. This presentation makes it more difficult for the reader to discern the subgroup information that the analysis emphasizes. While this organization matches the input to most statistical packages for correct analysis, it is not the best presentation for humans to discern the groups.

Keep in mind that tables are simply text-based graphics. All of the rules presented for graphical displays apply equally to textual displays.

The proper organization of the table in two dimensions clarifies the subgroup analysis. Tables may be augmented with decorative elements just as we augment graphics. Effective additions to the table are judged on their ability to focus attention on the science; otherwise these additions serve as distracters. Specific additions to tables include horizontal and vertical lines to differentiate subgroups, and font/color changes to distinguish headings from data entries.

**TABLE 8.2 Volume of a Mixture Based on the Included Fat and Surfactant Types<sup>a</sup>**

Fat	Surfactant	Volume
1	1	5.57
1	2	6.20
1	3	5.90
2	1	6.80
2	2	6.20
2	3	6.00
3	1	6.50
3	2	7.20
3	3	8.30

<sup>a</sup> **Problem:** The two categorical variables are equally of interest, but the table uses only one direction for displaying the values of the categories. This demonstrates that table generation is similar to graphics generation, and we should apply the same graphical rules honoring dimensions to tables.

**TABLE 8.3 Volume of a Mixture Based on the Included Fat and Surfactant Types<sup>a</sup>**

Fat	Surfactant		
	1	2	3
1	5.57	6.20	5.90
2	6.80	6.20	6.00
3	6.50	7.20	8.30

<sup>a</sup> The two categorical variables are equally of interest. With two categorical variables, the correct approach is to allow one to vary over rows and the other to vary over columns. This presentation is much better than the presentation of Table 8.2 and probably easier to interpret than any graphical representation.

Specifying a *y* axis that starts at zero obscures the differences of the results and violates Rule 3 seen previously. If we focus on the actual values of the subgroups, we can more readily see the differences.

## TWO RULES FOR TEXT ELEMENTS IN GRAPHICS

If a picture were worth a thousand words, then the graphics we produce would considerably shorten our written reports. While attributing “a thousand words” for each graphic is an exaggeration, it remains true that the graphic is often much more efficient at communicating numeric information than equivalent prose. This efficiency is in terms of the amount of

information successfully communicated and not necessarily any space savings.

If the graphic is a summary of numeric information, then the caption is a summary of the graphic. This textual element should be considered part of the graphic design and should be carefully constructed rather than placed as an afterthought. Readers, for their own use, often copy graphics and tables that appear in articles and reports. Failure on the part of the graphic designer to completely document the graphic in the caption can result in gross misrepresentation in these cases. It is not the presenter who copied the graph who suffers, but the original author who generated the graphic. Tufte [1983] advises that graphics “should be closely integrated with the statistical and verbal descriptions of the data set” and that the caption of the graphic clearly provides the best avenue for ensuring this integration.

*Rule 9: Captions for your graphical presentations must be complete. Do not skimp on your descriptions.*

The most effective method for writing a caption is to show the graphic to a third party. Allow them to question the meaning and information presented. Finally, take your explanations and write them all down as a series of simple sentences for the caption. Readers rarely, if ever, complain that the caption is too long. If they do complain that the caption is too long, it is a clear indication that the graphic design is poor. Were the graphic more effective, the associated caption would be of a reasonable length.

Depending on the purpose of your report, editors may challenge the duplication of information within the caption and within the text. While we may not win every skirmish with those that want to abbreviate our reports, we are reminded that it is common for others to reproduce only tables and graphics from our reports for other purposes. Detailed captions help alleviate misrepresentations and other out-of-context references we certainly want to avoid, so we endeavor to win as many of these battles with editors as possible.

Other text elements that are important in graphical design are the axes labels, title, and symbols that can be replaced by textual identifiers. Recognizing that the plot region of the graph presents numerical data, the axis must declare associated units of measure. If the axis is transformed (log or otherwise), the associated label must present this information as well. The title should be short and serves as the title for the graphic and associated caption. By itself, the title usually does not contain enough information to fully interpret the graphic in isolation.

When symbols are used to denote points from the data that can be identified by meaningful labels, there are a few choices to consider for improving the information content of the graphic. First, we can replace all



symbols with associated labels if such replacement results in a readable (nonoverlapping) presentation. If our focus highlights a few key points, we can substitute labels for only those values.

When replacing (or decorating) symbols with labels results in an overlapping indecipherable display, a legend is an effective tool provided that there are not too many legend entries. Producing a graphical legend with 100 entries is not an effective design. It is an easy task to design these elements when we stop to consider the purpose of the graphic. It is wise to consider two separate graphics when the amount of information overwhelms our ability to document elements in legends and the caption.

Too many line styles or plotting points can be visually confusing and prevent inference on the part of the reader. You are better off splitting the single graphic into multiple presentations when there are too many subgroups. An ad hoc rule of thumb is to limit the number of colors or symbols to less than eight.

*Rule 10: Keep line styles, colors, and symbols to a minimum.*

## **MULTIDIMENSIONAL DISPLAYS**

Representing several distinct measures for a collection of points is problematic in both text and graphics. The construction of tables for this display is difficult due to the necessity of effectively communicating the array of subtabular information. The same is true in graphical displays, but the distinction of the various quantities is somewhat easier.

## **CHOOSING EFFECTIVE DISPLAY ELEMENTS**

As Cleveland and McGill (1988) emphasize, graphics involve both encoding of information by the graphic designer and decoding of the information by the reader. Various psychological properties affect the decoding of the information in terms of the reader's graphical perception. For example, when two or more elements are presented, the reader will also envision byproducts such as implied texture and shading. These byproducts can be distracting and even misleading.

Graphical displays represent a choice on the part of the designer in terms of the quantitative information that is highlighted. These decisions are based on the desire to assist the analyst and reader in discerning performance and properties of the data and associated models fitted to the data. While many of the decisions in graphical construction simply follow convention, the designer is still free to choose geometric shapes to represent points, color or style for lines, and shading or textures to represent areas. The referenced authors included a helpful study in which various graphical styles were presented to readers. The ability to discern the underlying infor-

**TABLE 8.4 Rank-Ordered List of Elementary Design Choices for Conveying Numeric Information<sup>a</sup>**

Rank	Graphical Element <sup>b</sup>
1	Positions along a common scale
2	Positions along identical, nonaligned scales
3	Lengths
4	Angles
4–10	Slopes
6	Areas
7	Volumes
8	Densities
9	Color saturations
10	Color hues

<sup>a</sup> Slopes are given a wide range of ranks since they can be very poor choices when the aspect ratio of the plot does not allow distinction of slopes. Areas and volumes introduce false dimensions to the display that prevent readers from effective interpretation of the underlying information.

<sup>b</sup> Graphical elements are ordered from most (1) to least (10) effective.

mation was measured for each style, and an ordered list of effective elementary design choices was inferred. The ordered list for illustrating numeric information is presented in Table 8.4. The goal of the list is to allow the reader to effectively differentiate among several values.

## CHOOSING GRAPHICAL DISPLAYS

When relying completely on the ability of software to produce scientific displays, many authors are limited by their mastery of the software. Most software packages will allow users to either (a) specify in advance the desired properties of the graph or (b) edit the graph to change individual items in the graph. Our ability to follow the guidelines outlined in this chapter is directly related to the time we spend learning to use the more advanced graphics features of software.

## SUMMARY

- Examine the data and results to determine the number of dimensions in the information to be illustrated. Limit your graphic to that many dimensions.
- Limit the axes to exactly (or closely) match the range of data in the presentation.
- Do not connect points in a scatterplot unless there is an underlying interpolation that makes scientific sense.

- Recognize that readers of your reports will copy tables and figures for their own use. Ensure that you are not misquoted by completely describing your graphics and tables in the associated legends. Do not skim on these descriptions or you force readers to scan the entire document for needed explanations.
- If readers are to accurately compare two different graphics for values (instead of shapes or predominant placement of outcomes), use the same axis ranges on the two plots.
- Use pie charts only when there are a small number of categories and the sum of the categorical values has scientific meaning.
- Tables are text-based graphics. Therefore, the rules governing organization and scientific presentation of graphics should be honored for the tables that we present. Headings should be differentiated from data entries by font weight or color change. Refrain from introducing multiple fonts in the tables and instead use one font where differences are denoted in weight (boldness), style (slanted), and size.
- Numeric entries in tables should be in the same number of significant digits. Furthermore, they should be right justified so that they line up and allow easy interpretation while scanning columns of numbers.
- Many of the charts could benefit from the addition of grid lines. Bar charts especially can benefit from horizontal grid lines from the  $y$ -axis labels. This is especially true of wider displays, but grid lines should be drawn in a lighter shade than the lines used to draw the major features of the graphic.
- Criticize your graphics and tables after production by isolating them with their associated caption. Determine if the salient information is obvious by asking a colleague to interpret the display. If we are serious about producing efficient communicative graphics, we must take the time ensure that our graphics are interpretable.

## TO LEARN MORE

Wilkinson (1999) presents a formal grammar for describing graphics, but more importantly (for our purposes), the author lists graphical element hierarchies from best to worst. Cleveland (1985) focuses on the elements of common illustrations where he explores the effectiveness of each element in communicating numeric information. A classic text is Tukey (1977), where the author lists both graphical and text-based graphical summaries of data. More recently, Tufte (1983, 1990) organized much of the previous work and combined that work with modern developments. For specific illustrations, subject-specific texts can be consulted for particular displays in context; for example, Hardin and Hilbe (2003, pp. 143–167) illustrate the use of graphics for assessing model accuracy.