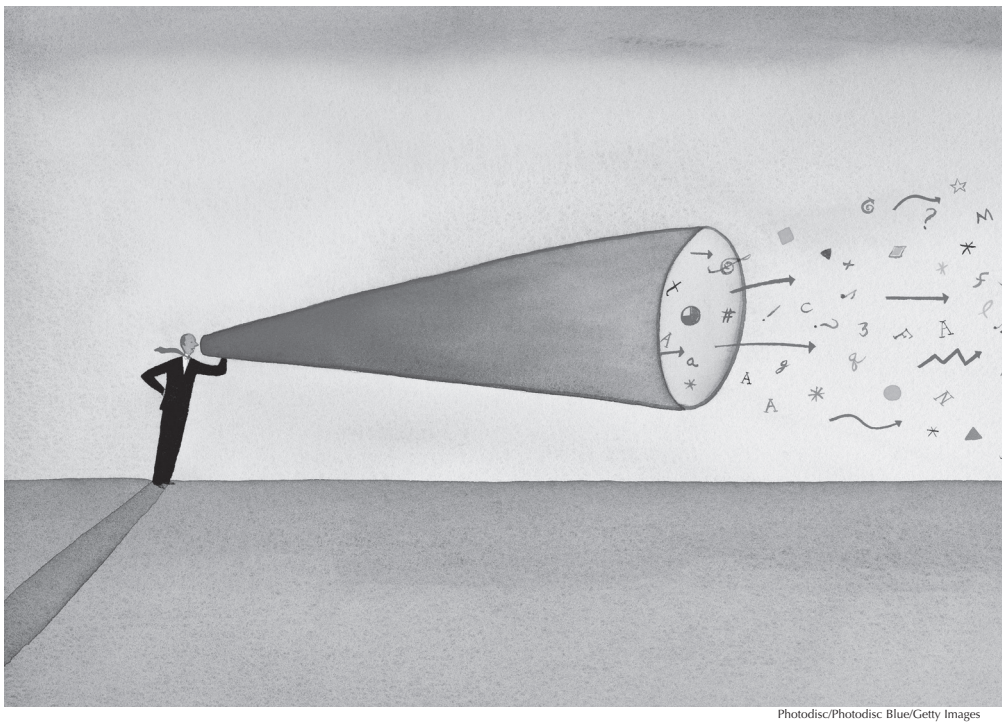
# How To Communicate Statistical Findings: an Expository Writing Approach

**Jane E. Miller**



Photodisc/Photodisc Blue/Getty Images

What do the hit television show "CSI: Crime Scene Investigation" and good quantitative writing have in common? Both involve using technical information as evidence in an inquiry, asking whether the scientific or numeric facts support or refute a hypothesis. On "CSI," the investigators use a range of laboratory tests to conduct a coherent scientific investigation. Tying together evidence from ballistics, DNA tests, and trace chemical analysis, they use established scientific approaches to build a convincing case. And they use technical language when speaking with one another, but must be able to translate their findings into plain English so that juries made up of nonscientists will understand.

Likewise, people who write about statistics should seek to conduct a coherent statistical inquiry, using statistics to test a hypothesis about the relationship between the concepts under study. When conducting the tests and talking with other statisticians, they too use a specialized vocabulary. However, when communicating to a general audience, statisticians must explain the answers in words that the audience can understand. Unfortunately, few courses teach how to write a clear narrative linking numeric evidence to substantive questions, or how to present statistics in words that nonstatisticians can comprehend easily. As a consequence, many people present statistical results in ways that leave general audiences struggling to understand what questions those numbers are intended to answer or what conclusions they support.

By introducing the question, describing facts in context, and relating the entire body of evidence back to the original question, statisticians can tell a clear story with numbers as evidence. Basic principles for describing numeric patterns include introducing the topic, reporting and interpreting data and statistics, specifying direction and magnitude of an association, summarizing patterns, writing a conclusion, and explaining a chart to a live audience. Using these

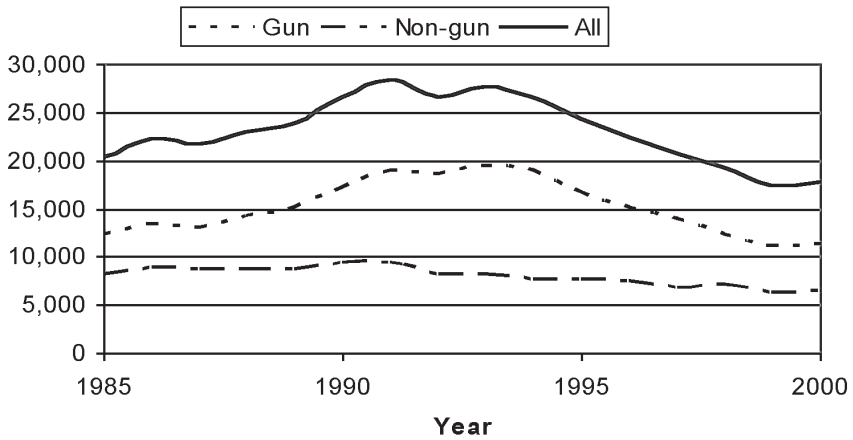



**Figure 1.** **Gun and non-gun homicides, United States, 1985–2000**
Source: Fox, J.A., and Zawitz, M.W. (2004). Homicide Trends in the United States. U.S. Department of Justice, Bureau of Justice Statistics: Washington, DC

guidelines, statisticians and other quantitative writers can present a coherent statistical inquiry, with statistics playing a fundamental but unobtrusive role.

## Introducing the Topic

As with other types of expository writing, start by introducing the topic of your work and the questions you seek to answer with the numbers that follow. If an investigator starts talking about a specific fingerprint or piece of DNA without having first outlined the basic facts of the crime and its context, a jury will have a hard time understanding where that evidence fits in the overall case. To help set the scene for your statistics, begin with a topic sentence that introduces the characters (variables, in a statistical paper) and the Ws (when, where, what).

> **Poor**: [No introductory sentence.] "In 2000, there were 11,280 gun-related homicides (Figure 1)."

This version jumps directly to presenting data without orienting the audience to the topic and objectives, so it will be difficult for them to see what those numbers mean.

> **Better**: "What factors explain the observed rise and fall in overall homicide rates in the United States in the 1990s (Figure 1)?"

This version uses a rhetorical question to introduce the context (where and when) and pattern to be investigated (the time trend in homicide rates). However, it does not specify which variables will be considered as possible explanatory factors.

> **Best**: "Was the substantial rise and subsequent fall in the number of homicides in the 1990s in the United States (Figure 1) observed across all age groups and types of weapons?"

This version sets the context and mentions the outcome (homicide) and hypothesized explanatory factors (age and type of weapon).

Notice there are no numbers yet, just a statement that establishes the purpose of the statistics to be described later in the paragraph. Such introductions are especially important when presenting a series of charts and tables, each of which addresses one part of an analysis. For instance, this investigation of homicide trends includes one chart comparing time trends in gun and non-gun homicides, another showing time trends in gun-related homicides by age of offender, and a third showing gun involvement by age of victim. As you move from one topic to another, introduce it before presenting the associated numeric evidence, using paragraphs to organize the material and transition sentences to guide the reader from one major point to another.

> **Poor**: "In 1985, there were 20,370 homicides, 12,263 of which were gun-related (Figure 1). The total number of homicides rose until the mid-1990s, and then declined until 2000. In 1985, 14–17 year olds committed 952 gun-related homicides, while persons 18–24 and 25+ committed 3,633 and 7,621 gun-related homicides, respectively (Figure 2). Three-quarters of teenage homicide victims were killed by guns, compared to only one-quarter of those aged 80 and older (Figure 3).

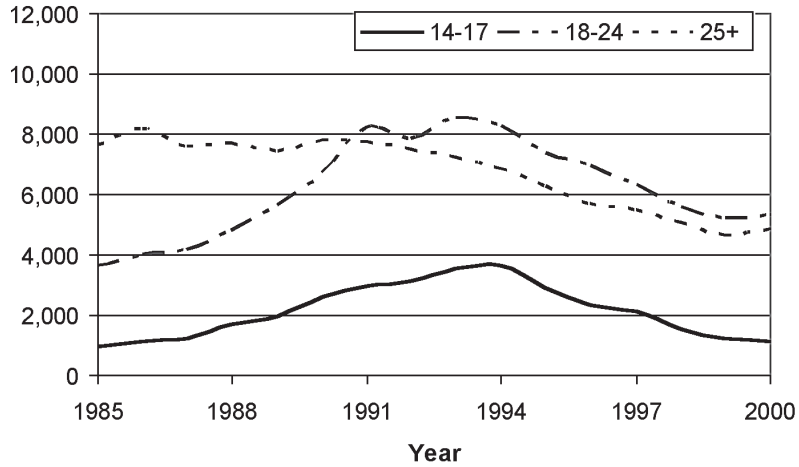



**Figure 2.** **Gun homicides by age of offender, United States, 1985–2000**
Source: Fox, J.A., and Zawitz, M.W. (2004). Homicide Trends in the United States. U.S. Department of Justice, Bureau of Justice Statistics: Washington, DC

This description simply lists statistics from each of three figures without explaining how they relate to one another or how the statistics address the initial research question.

> **Better**: [Transition sentence from a paragraph describing Figure 1 to a second paragraph describing Figure 2] "As shown in Figure 1, the increase and subsequent decrease in homicides were driven entirely by trends in gun-related homicides. Figure 2 examines whether gun-related homicides exhibited the same time trend among all age groups of offenders."

By starting a new paragraph and section to present evidence on homicide patterns by age of offender, this version signals a second step in the investigation. The first sentence summarizes the conclusions of the preceding section (on time trends in homicide by type of weapon). The second sentence introduces a new variable—age of offender—to be considered in a further dissection of those time trends. Subheadings such as "Homicide trends by type of weapon" and "Homicide trends by age of offender" could be used to provide further guidance through the different parts of the analysis.

## Reporting and Interpreting Numbers

As you present evidence to test your hypotheses, report the raw data and interpret the statistics. Reporting the numbers you work with is an important first step toward writing effective numeric descriptions. By including the numbers in the text, table, or chart, you give your audience the raw materials with which to perform additional calculations or to compare your data with information for other times, places, or groups. Likewise, presenting the DNA profile of material found at a crime scene would allow comparison with many other DNA samples.

However, if you stop there, you leave it to your readers to figure out how those data answer the question at hand. An isolated number or scientific fact that has not been introduced or explained leaves its explication entirely to your audience. Those who are not familiar with your topic are unlikely to know which comparisons to make or to have the information for those comparisons immediately at hand.

Consequently, after reporting the raw data, interpret it. To help readers grasp the meaning of the numbers you report, provide the relevant data and explain the comparisons. To make his or her point about whether a particular suspect appears guilty, an investigator should compare explicitly DNA evidence from the crime scene to DNA evidence collected from that suspect. Similarly, to help readers assess how his or her data support or refute a particular hypothesis, a statistician should present statistical or mathematical comparisons of those data.

> **Poor**: "In 1985, there were 12,263 gun-related homicides (Figure 1)."

From this sentence, readers can't assess whether the number of gun-related homicides was high or low, changing or stable. If they knew the number of homicides committed with other weapons or in other time periods, they could compare the figures, but you will make the point more directly if you do the calculation for them.

> **Better**: "Throughout the period shown, homicides were committed most often with guns. In 1985, for example, roughly 1.5 times as many homicides were committed with guns as with other types of weapons (12,263 versus 8,107; Figure 1) ."

The first sentence reports which type of weapon is used most commonly to commit homicide, while the second compares the number of homicides by type of weapon for one time point with reference to the associated chart. Subsequent sentences would be used to describe time trends in homicides by type of weapon and how the ratio of gun to non-gun homicides changed across time.

Although it is important to interpret quantitative information, it is also essential to report the original data upon which it is based. If you only present a ratio or percentage change (two measures of relative level), for example, you will have painted an incomplete picture. Suppose a report states that the number of gun-related homicides in the United States was 60% higher in 1994 than in 1985, but does not report the number of such homicides for either year. A 60% increase is consistent with many possible combinations: 10 and 16 homicides, or 1,000 and 1,600, or 100,000 and 160,000, for example. The first pair of numbers suggests a very low incidence of homicide; the last pair suggests an extremely high rate. Unless the absolute homicide levels are mentioned, one can't determine whether the nation has nearly eradicated homicide or faces a huge homicide problem. Furthermore, one can't compare homicide statistics from other times or places.

## Explaining Direction and Magnitude

Writing about numbers often involves portraying associations between two or more variables. To describe an association, explain both its shape and size, rather than simply stating whether the variables are correlated. In other words, which value is higher, and how much higher? Is a trend rising, falling, or level? For instance, to compare homicides by type of weapon, report which type of weapon was mostly common as well as how much more common. Well-chosen adjectives ("minuscule difference"), adverbs ("increased markedly"), and analogies to familiar shapes ("bell-shaped" or "J-shaped") can be effective in conveying the shape and size of a relationship between variables. For statistically oriented audiences, also report results of inferential statistical tests.

## Direction of Association

For ordinal, interval, or ratio variables—such as age group or year for which the values have an inherent numeric order—direction of association also can be described in terms of positive (direct) or negative (inverse) associations—whether the outcome variable increases or decreases as the predictor increases (see examples below).

For nominal variables—such as gender, race, geographic region, or type of weapon—that are classified into categories with no inherent order, describe direction of association by


**Figure 3. Percentage of homicides involving guns by age of victim, United States, 1976–2002**
Source: Fox, J.A., and Zawitz, M.W. (2004). Homicide Trends in the United States. U.S. Department of Justice, Bureau of Justice Statistics: Washington, DC

specifying which category has the highest or lowest value and how other categories compare.

> **Poor**: "Homicide is negatively associated with type of weapon."

Type of weapon has no directionality, so this sentence cannot be interpreted.

> **Better**: "Homicides are committed more often with guns than with other weapons."

This version identifies the most common type of weapon used in homicides and names the other category against which it is being compared.

## Size of Association

An association can be large (e.g., a given change in one variable is associated with a big change in the other variable) or small (e.g., a given change in one variable is associated with a small change in the other). If several factors are each associated with homicide risk, knowing which have the strongest association can help identify which attributes to target.

> **Poor**: "Age of victim is correlated with the percentage of homicides due to guns (Figure 3)."

This sentence doesn't say whether age of victim and homicide risk are positively or negatively related, or how much homicide risk differs by age.

> **Better**: "As the victim's age increases, the percentage of homicides involving guns decreases (Figure 3)."

Although this version specifies the direction of the association, the size of the age/gun homicide association is still unclear.

> **Best**: "Among victims aged 15 and older, the percentage of homicides involving guns is inversely related to age. For example, more than three-quarters of teenage homicide victims were killed by guns, compared to less than one-quarter of victims aged 80 and older (Figure 3)."

This version explains the direction and size of the age/gun-homicide pattern.

The size of a difference between two values can be calculated in any of several ways, including absolute difference (subtracting one value from the other), relative difference or ratio (dividing one value by the other), or percentage difference or change. To decide which of these alternatives to use as you write, read similar comparisons in the literature for your field.

## Summarizing Patterns from Tables or Charts

Often, answering a research question requires describing a pattern involving many numbers, such as trends in homicides for each of three age groups over several decades or rates of violent crime by type of offense and region in the United States. Typically, the data for such patterns are reported in tables and charts, which provide an efficient way to organize lots of numbers. Just as an investigator must help the jury see the overall pattern suggested by the body of evidence, statisticians should guide their readers through the patterns shown in their tables and charts. If you only provide a table or chart, you leave it to your audience to figure out for themselves what that evidence says. An important step in telling a clear, logical story with numbers as evidence is to summarize the patterns shown in your tables and charts and relate them back to the substantive question at hand. Use prose to summarize the patterns so your audience can see the general relationship in the table or chart—the forest, not the individual trees. If readers want to know the number of homicides for a particular date and age group, they can look it up in the associated chart.

When summarizing a table or chart, inexperienced writers often make one of two opposite mistakes: (1) they report every number from the table or chart in their description or (2) they pick a few arbitrary numbers to contrast in sentence form without considering whether those numbers represent an underlying general pattern. Neither approach adds much to the information being presented in the table or chart, and both approaches can confuse or mislead the audience. Paint the big picture, rather than reiterating all of the little details.

As you summarize, link the evidence back to the substantive topic—in this case, considering whether gun-related homicides showed similar time trends in all age groups. Describe broad patterns with a few simple statements, pointing out similarities and differences across groups.

> **Poor**: "In 1985, 14–17 year olds committed 952 gun-related homicides. In 1986 and 1987, persons in that age group committed 1,099 and 1,207 gun-related homicides, respectively (Figure 2). [Description continues by reporting annual statistics for each of the three age groups]."

Individual statistics on the number of homicides in each of 15 years for each of several age groups force readers to do the math to figure out whether homicides are rising, falling, or staying level, and whether the time trend is similar for the three age groups being compared. All those sentences reporting numbers also will obscure the general pattern and tire your readers.

> **Poor [version 2]**: "Between 1985 and 1986, the number of gun-related homicides committed by 14–17 year olds increased from 952 to 1,099. Between 1986 and 1987, it increased again, to 1,207 (Figure 2). [Description continues by reporting one-year changes in the number of homicides for each of the three age groups.]"

Although this version presents single-year changes in homicides committed by one age group, instead of merely reporting the value of each data point, it fails to paint the overall shape of the time trend across the period shown or to compare across age groups.

> **(Somewhat) better**: "Among offenders aged 14–17, gun-related homicides nearly quadrupled between 1985 and 1994 (from 952 to 3,617), and then declined to 1,079 in 2000. Among offenders aged 18–24, gun-related homicides more than doubled between 1985 and 1994 (from 3,633 to 8,253), then decreased through 2000. Among offenders aged 25 and older, gun-related homicides declined slightly throughout the period from 1985 to 2000 (Figure 2)."

Although this version describes the shape and size of the time trend for each age group, it doesn't compare time trends across age groups, requiring readers to figure out whether all three age groups followed the general time trends observed in Figure 1 or whether the patterns varied.

> **Best**: "As shown in Figure 2, in the two youngest groups of offenders, gun-related homicides increased substantially between 1985 and 1994, and then decreased steadily until 2000. In contrast, the number of gun-related homicides committed by older offenders decreased slowly throughout the time period shown."

This description points out that the time trend in gun-related homicides was similar for two of the three age groups, and then describes the general shape of the pattern. The phrase "in contrast" is used to emphasize that the time trend for the third age group was different from the other two before going on to describe the shape of that pattern.

## Generalization, Example, Exceptions

Here is a mantra I devised to guide you through the steps of writing an effective description of a pattern involving three or more numbers: generalization, example, exceptions, or GEE for short. The idea is to identify and describe the general shape of a pattern, give a representative numeric example to illustrate that pattern, and then explain and illustrate any exceptions. This approach also can be used to summarize findings of previous studies, identifying consensus and pointing out discrepancies—as when one or more facts from the crime scene contradict the pattern implied by other pieces of evidence.

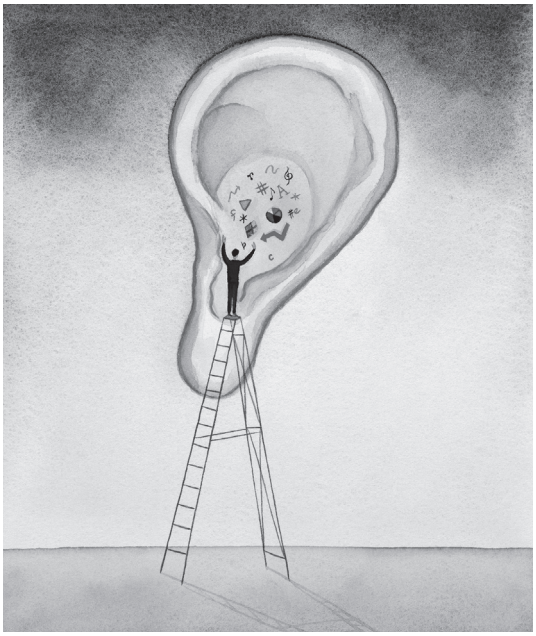
### *Generalization*

For a generalization, come up with a description that characterizes a relationship among most, if not all, of the numbers. In Figure 2, ask "Did gun-related homicides show similar time trends in all three age groups? Does one age group consistently commit the most gun-related homicides?" Start by describing one such pattern (e.g., time trends in gun-related homicides committed by persons aged 14–17), then consider whether that temporal pattern applies to each of the other two age groups as well. Or, determine which age group of offender committed the most homicides in 1985 and see whether it was also the worst age group in 1995 and 2000. If the pattern fits most of the groups most of the time, it is a generalization. The few situations that don't fit are your exceptions (see below).

Photodisc/Photodisc Blue/Getty Images

In Figure 2, there are two generalizations of interest: the relationship of each independent variable (year and age group of offender) with the dependent variable (number of gun-related homicides). So, we start with a verbal generalization about each of those patterns, which serve as the topic sentences for separate paragraphs—one about time trends in gun-related homicides, the other about age differences in homicides committed by gun.

> **Generalization #1**: "In the youngest two age groups of offenders, the number of gun-related homicides increased markedly between 1985 and 1994, then decreased until 2000 (Figure 2)."



Photodisc/Photodisc Blue/Getty Images

> **Generalization #2**: "From 1985 until 1990, persons aged 25 and older were most likely to commit gun-related homicides (Figure 2)."

Notice that although these sentences each convey direction of association, they don't include any numbers to assess size. That comes in the next step. Readers are referred to the accompanying chart, which depicts the relationships and allows readers to look up values for specific age groups and years.

*Example*

Having described a generalizable pattern in intuitive language, illustrate it with numbers from your table or chart. This step anchors your generalization to the specific numbers upon which it is based. It ties the prose and table or chart together. By reporting a few illustrative numbers, you implicitly show your audience where in the table or chart those numbers come from as well as the comparison involved. They can then test whether the pattern applies to other times, groups, or places using other data from the table or chart.

To illustrate the above generalizations, include sentences that incorporate examples from Figure 2. For the generalization about time trends, pick one of the age groups that showed this pattern and compare number of homicides across time for that group: "For example, among offenders aged 14–17, the number of gun-related homicides nearly quadrupled between 1985 and 1994, then declined to near-1985 levels."

For the generalization across age groups, choose one year and present the difference in number of gun-related homicides by age of offender for that date: "In 1985, for example, persons 25 and older accounted for more than twice as many gun-related homicides as 18–24 year-old offenders, and nearly eight times as many as those aged 14–17."

*Exceptions*

Sometimes you will be lucky enough that your generalizations capture all the relevant variation in your data. If you are working with real-world data, however, there will often be important exceptions to the general pattern you have sketched. Tiny blips can usually be ignored, but if some parts of a table or chart depart substantially from your generalization, describe those departures.

When portraying an exception, explain its overall shape and how it differs from the generalization you have described and illustrated in your preceding sentences. Is it higher or lower? By how much? If a trend, is it moving toward or away from the pattern you are contrasting it against? In other words, describe both direction and magnitude of the difference between the generalization and the exception. Use phrases such as "on the other hand" or "in contrast" to differentiate an exception from a generalization; "conversely" or "on the contrary" can be used to point out when one pattern is the opposite direction of another. Finally, provide numeric examples from the table or chart to illustrate the exception.

In the case of Figure 2, the above generalizations about age group of offender fail to capture the changing rank order in number of gun-related homicides from the early 1990s onward.

> [**To follow the generalization comparing age groups**]: "However, after 1990, 18–24 years olds became the largest group of gun-related homicide offenders. Throughout the period shown, offenders aged 14–17 committed the fewest gun-related homicides."

The first sentence describes how the original generalization changed in later time periods. The second sentence generalizes the relative position of the youngest age group.

This particular exception occurred in terms of direction: Homicides committed by one age group declined, while those for the other age groups rose. Exceptions also can occur in terms of magnitude, such as a rising trend in each of the groups, but at a notably slower rate in some groups than in others.

## Writing the Conclusion

Having presented the individual pieces of evidence, an investigator must summarize how that evidence, taken together, incriminates or eliminates a particular suspect of the criminal charges. Likewise, in the concluding section of a paper, statisticians should explain how the statistical evidence answers the question posed at the beginning of the paper, following standard expository writing guidelines to writing an analytic essay.

> **Poor**: "In the two youngest groups of offenders, gun-related homicides increased two-fold to four-fold between 1985 and 1994, then decreased until 2000. In contrast, the number of gun-related homicides committed by offenders aged 25 and older decreased slowly throughout the time period shown. Homicides committed with other weapons remained relatively steady over the entire period."

This version merely restates the statistics from the results section of the paper, adding nothing to what has been demonstrated already and failing to put the evidence back into the "big picture" of the main research question.

> **Better**: "The rise in the total number of homicides in the early 1990s was due to increases only in gun-related homicides—increases that were concentrated largely among perpetrators in their teens and early 20s."

This sentence brings the analysis full circle, relating the statistical evidence back to the original question about the roles of weapon type and age of offender to the overall time trend in homicides. The conclusion could then be fleshed out with possible explanations for this pattern, such as availability of guns, penalties associated with gun-related homicide, or factors that explain age patterns of violent criminal behavior.

## Explaining a Chart or Table to a Live Audience

Tables, charts, maps, and other diagrams offer real advantages for presenting evidence, whether results of forensic tests at a trial or results of statistical tests in a course lecture or conference presentation. Unfortunately, many speakers devote far too little time to describing such exhibits. They put up a slide with the table or chart, state "as you can see," and then describe the pattern in a few seconds before moving on to the next slide. As the slide disappears, many listeners are still trying to locate the numbers or pattern in question and have not had time to digest the meaning of the statistics.

Although it may appear to save time, failing to orient the audience to your tables or charts reduces the effectiveness of your talk. An investigator who works every day with certain types of diagrams and computer output from forensic tests knows exactly where to look and how to interpret the information shown. Likewise, if you designed a statistical chart and wrote the corresponding lecture, you know it well enough to home in quickly on the exact number, table cell, or trend line you wish to discuss. Give your viewers a similar advantage by showing them where to find the statistics you plan to discuss and what questions they address before you report and interpret patterns.

### *Introduce the Topic*

State the topic or purpose of the table, chart, or other diagram, just as you would in the introductory sentence of a written paragraph. Rather than reading the title from the slide, paraphrase it into a full sentence or rephrase it as a rhetorical question: "Figure 2 examines trends in gun homicides by age of offender. In other words, 'did all age groups of offenders show similar changes in the number of gun-related homicides between 1985 and 2000?'"

### *Explain the Layout*

Explain the layout of the table or chart. Don't discuss any numbers, patterns, or contrasts yet. Just give your audience a chance to digest what is where. For a table, name what is in the columns and rows. For a chart, identify the concepts and units on the different axes and in the legend, mentioning the color or shading of bars or line styles that correspond to each major group you will discuss. Also, explain the purpose of features such as reference lines or regions, colors, symbols, or other annotations. (If you don't have time to mention such features, omit them to avoid distracting or confusing your viewers.)

Use a "Vanna White" approach as you explain the layout: Literally point out the applicable portion of the table or chart as you mention it. Point with a laser pointer, pen, or finger—it doesn't matter what you use. The important thing is to lead your viewers' eyes across the key features of the exhibit before reporting or interpreting the information found there. At first, this may seem silly or awkward, but most audiences follow and retain the subsequent description much more easily than if you omit the guided tour. Below, I use bracketed comments to describe the Vanna White motions that accompany the surrounding script; they are there to guide you, not to be spoken as part of the presentation.

For Figure 2: "Year is shown on the x axis, ranging from 1985 to 2000 [wave left to right along the x axis]. On the vertical (y) axis is the number of homicides committed using a gun [gesture vertically along y axis]. Offenders aged 14–17 years are shown with the solid line, those aged 18-24 with the dashed line, and those 25 and older with the dotted line [Point to each line in turn. Note: For a color slide, mention the line colors instead of the line styles.].

### *Describe the Patterns*

Having introduced your audience to the purpose and layout of the table or chart, describe the patterns it embodies. Use the GEE approach, starting with a general descriptive sentence and following with specific numeric examples and exceptions (where pertinent), as in the above description of Figure 2. Again, gesture to show comparisons and point to identify specific values on the chart as you mention them. For example, when reporting numeric values to illustrate the trend generalization, coordinate the bracketed gestures with the associated script: "Among persons aged 25 and older, the number of gun-related homicides declined from about 8,000 in 1985 to about 5,500 in 2000" [wave along dotted line for ages 25+, pointing out y values corresponding to 1985 and 2000].

To show your audience where the numbers for the across-age group generalization come from, say: "In 1985, for example, persons 25 and older [pointed to dotted line in 1985] accounted for more than twice as many gun-related homicides as offenders aged 18–24 [point to dashed line for 1985] and nearly eight times as many as offenders aged 14–17 years [point to solid line for 1985]."

In summary, both crime scene investigators and people who work with statistics face the often challenging task of weaving together complex scientific information into a form that answers a substantive question in a straightforward and clear fashion. Technical forensic or statistical skills alone are not sufficient to convey results to an audience, whether a jury weighing forensic evidence or a general audience reading a paper involving statistics. Write about statistical findings so readers understand what they mean, crafting a logical narrative with a beginning, middle, and end. In the introduction, ask the question in plain English, mentioning the specific concepts under study. In the body of the results section, systematically review the statistical evidence using introductory and transition sentences to keep readers oriented to the logical flow of the analysis. Finally, close by answering the original question in everyday language. These approaches will help yield a clear story with numbers as evidence.

**Editor's Note:** This paper is adapted in part from material in T*he Chicago Guide to Writing about Numbers*. 