

Presenting Data: Tabular and graphic display of social indicators

(under construction)

[Gary Klass](#)

[Illinois State University](#)

© 2002

Home

Numbers?

■ Good Tables

Good Charts

References

Course Page

Chart of the week

Constructing Good Tables for the Display of Social Indicators.

- [Introduction](#)
- [Presenting Meaningful Data](#)
- [Presenting Unambiguous Data](#)
- [Presenting Data Efficiently](#)
- [Formatting tables with Microsoft Excel](#)

Introduction

For any writing based on the analysis of statistical data, there are likely to be two types of readers: those who tend to scan through the text ignoring most of the charts and tables and those who scan through the charts and tables and ignore most of the text. A well designed analysis will accommodate both sets of readers; both should be able to quickly discern the main points of your analysis.

A statistic is a summary measure of the size or magnitude of some real world phenomenon. Although statistics are often presented in a manner to complicate, confuse and obscure essential facts about the world, when they are used well their purpose is to *simplify*.

According to the 2000 US Census, in 1999 the 12.4 percent of Americans lived in families below the government defined standard of poverty. That 12.4 statistic is one of many statistics derived from a compilation of several tens of millions of other numbers provided by one out of every six American households who were asked questions about their earnings on the long form questionnaire used by the Census Bureau. The purpose of the statistic is to provide a summary measure of the economic well being of Americans in 1999. Other summary measures based on the same numbers measure the mean and median income of American households, the distribution of incomes across economic classes, the median incomes of black, white and Hispanic families of different family sizes and structures, and much more. Had the Census Bureau reported the numbers it collected, the reported income of some 50 million Americans, things would not be so simple.

The essential purpose of a tabular display is to simplify the presentation of statistical data that otherwise would have to be presented in a textual discussion. A table serves little or no purpose if all the information it contains could just as easily be summarized in a sentence. When properly constructed, data display tables generally provide a more efficient means of presenting data that would otherwise be presented in the form of text. But this depends on

whether or not the table is well constructed.

There are three general characteristics of a good tabular display. The table should present **meaningful** data. The data should **unambiguous**. The table should convey ideas about the data **efficiently**.

Whether or not the data are meaningful has to do with how closely the data relate to the main points that you are trying to make in your analysis or report. The data contained in a table constitute the premises, or evidence, offered to support a conclusion; ideally this should be an important conclusion, an essential part of the analysis you are making. Authors sometimes include tabular data in their writings merely to convey a feeling of scientific rigor offering data that is irrelevant or tangential to the main issues under discussion.

Whether or not the information presented in a table is unambiguous depends largely on the text contained in the table. The information presented in a table: the data, the title, headings, footnotes and citation of sources, should be completely interpretable on its own without assistance from the textual discussion. The titles and headings should precisely define what each number in the table represents. A citation of the data source should give the reader some indication of the reliability and validity of the data.

An efficient tabular display will allow a reader to draw a variety of interesting conclusions from a large amount of information. How quickly a reader can digest the information presented, discern the critical relationship among the data, and draw meaningful conclusions depends on how well the table is formatted.

Charts are graphical representations of an underlying statistical table. As a general rule, a chart or a graph will be a more effective way of presenting larger amounts of data, although with less accuracy, than a table. With large amounts of data, and especially in the case of time series charts or scattergrams, charts often are more effective at revealing trends and relationships among the data.

A note on terminology:

- "**Variable**" refers to the concept being measured, for example, murder rate, per capita income, Gross domestic Product. In most tables, the variables define the columns of the table.
- "**Cases**" refers to the geographic units, groups, time points or units of analysis. The cases usually define the rows of a table (but there are [exceptions](#)). The cases are in fact a variable.
- A **datum** or **data point** is the value of a variable for a particular case.

A statistical table generally consists of the a) data points for two or more variables and two or more cases and b) textual material explaining what the data represent. In the following table, each of the teams is a case. The team name, number wins, loses, the percentage of games won and the number of games behind first place are the variables.

Table 1:

American League Central Standings

Team	W	L	Pct.	GB
Minnesota	46	30	0.605	-
Cleveland	45	30	0.600	0.5
Chicago	36	38	0.486	9
Detroit	32	42	0.432	13
Kansas City	30	46	0.395	16

(Thursday June 28, 2001)

Presenting Meaningful Data

Report the most meaningful data, data that measure something important about the cases you are analyzing. As a general rule statistics based on rates, ratios, and per capita measures are more meaningful than rather than aggregate totals. Beware of measures that merely reflect the population size of the cases (as in Table 2- 10.1, below).

Table 2: 

Table 10.1: Murders* in Ten Largest US Cities, 1998		Table 10.2: Murder Rates* in Ten Largest US Cities, 1998	
Chicago	703	Detroit	43.0
New York	633	Chicago	25.6
Detroit	430	Philadelphia	23.3
Los Angeles	426	Dallas	23.1
Philadelphia	338	Phoenix	15.1
Houston	254	Houston	14.1
Dallas	252	Los Angeles	11.8
Phoenix	185	New York	8.6
San Antonio	89	San Antonio	8.1
San Diego	42	San Diego	3.5

*Murder and non-negligent manslaughter

*Murder and non-negligent manslaughter per 100,000 population

source: Statistical Abstract 2000 , CD-Rom, table 332.

Time. Two time points are often better than one. Including more than one year's measure of a variable in columns of data often provides for substantially more information and analytical comparisons.

Table 2a:

Table 10.3:
Murder Rates in Ten Largest US Cities, 1995-98

	1995	1998	Net Change
Detroit	47.6	43.0	- 4.6
Chicago	30.0	25.6	- 4.4
Philadelphia	28.2	23.3	- 4.9
Dallas	26.5	23.1	- 3.3
Los Angeles	24.5	11.8	-12.7
Phoenix	19.7	15.1	- 4.6
Houston	18.2	14.1	- 4.1
New York	16.1	8.6	- 7.5
San Antonio	14.2	8.1	- 6.1
San Diego	7.9	3.5	- 4.4

*Murder and non-negligent manslaughter per 100,000 population
sources: Statistical Abstract 2000, CD-Rom, table 332; and
Bureau of Justice Statistics:
<http://www.ojp.usdoj.gov/bjs/data/cities92.wk1>

Unless you are analyzing the before and after effect of something that happened at a single point in time, a one year difference in data often does not provide for meaningful comparisons. One-year changes in the city murder rates are subject to random fluctuations: 3, 5, and 10 year change intervals provide for a more reliable analysis. An exception to this rule is budgetary data, where the analysis of annual changes and comparisons the past, current and forthcoming fiscal years is often crucial for agencies living on annual appropriations.

Particularly with data that generally show substantial year-to-year variation (for example, murder rates in relatively small cities), calculating average score over three or five year intervals before and after a significant policy change (say, before and after the imposition of the death penalty) will prevent making too much out of random data fluctuations.

One should beware, however, of seemingly arbitrary selections of a base year. If table 2a. ([above](#)), for example, were presented to make a case for the effectiveness of the Los Angeles police force, one would have to check to see whether the 1995 data were the result of an unusually high jump in the city's murder rate.

Integrate tables with the text. Tabular displays of data should not be used merely to make an essay look more sophisticated. There are plenty of authors who provide displays of statistical data accompanied by a textual discussion that says nothing about the data or, worse, a discussion that contradicts the data presented. As a general rule, a single table ought to be accompanied by a half page of double spaced text directly related to the content of the table.

Presenting Unambiguous Data

The table titles, column headings and footnotes should precisely define what each data point in the table means. When rates or ratios are reported, both the numerator and denominator should be clearly defined. Pay particular attention to whether the statistics are reported in hundreds, thousands or millions. Students sometimes confuse statistics reported in thousands with rates per thousand of population. The following are three entirely different statistics:

- Highway fatalities (thousands)
- Highway fatalities per 100,000 population
- Highway fatalities per 100 million vehicle miles traveled.

Percentages. Percentages can usually be calculated in at least two different ways and are often a source of confusion. Consider the difference between tables 3a and 3b. In 3a, we see that 14% of poor families are headed by a householder under 24 years of age; in 3b 31% of families headed by a householder under 24 years old are poor. Although the table title clearly defines the difference, showing the 100% total and the "all families" rate. In general "composition" or "distribution" percentages are less meaningful than the rate statistics. This is especially true when the categories are arbitrary, as in table 3a: the 18 to 24 category has a six year range while others have a ten year range.

Table 3: 

Table 3a:		Table 3b:	
Distribution of Poverty Families by age of householder, 1999		Poverty Rate of Families by age of householder, 1999	
18 to 24 years old	14%	18 to 24 years old	31%
25 to 34 years old	29	25 to 34 years old	16
35 to 44 years old	27	35 to 44 years old	10
45 to 54 years old	12	45 to 54 years old	6
55 to 64 years old	9	55 to 64 years old	6
65 years old and over	10	65 years old and over	6
total	100	all families	9

source: US Statistical Abstract 2001, CD-Rom. Table 686

Change in Percentages. Calculating changes in percentages, rates and ratios is also a source of confusion both in tabular display and in textual summaries of data. The following is an example of a poorly defined data:

**Table 4: Change in Teenage Birth Rates:
1987-1998**

White	6.7%
Black	-4.9
Asian	-1.8

Hispanic

3.7

Source: Statistical Abstract 2000, Table 85

Consider the ambiguity of these data by trying to answer the following questions:

- Is the Teenage Birth Rate the percentage of all babies who were born to teenage mothers or the percentage of teenage mothers who gave birth?
- Would an increase in the teenage birth rate from 20 to 26.7 be a 6.7% change or a 3.35% change?

In table 2a (above), change is reported as a net change. It would have been possible to show the change as a percentage change. The change in San Diego's murder rate from 7.9 to 3.5 (murders per 100,000 population) can be reported either as a 4.4 net decline in the murder rate or as a 56% drop in the rate.

Although in the San Diego example most people (especially the San Diego chief of police) would be more comfortable with the measure of the percentage change in the rate, percentage changes in rates and ratios can be misleading. Consider this: if San Diego's murder rate is 7.9, its non-murder rate (the number of people not murdered per 100,000 population) is 92.1 and rose to 96.5. The net change in this "safety rate", is still the same 4.4, but the percentage change is less than 5%. While the probability of being murdered in San Diego fell 56%, the probability of not being murdered rose only 4.8%.

Perhaps unemployment rates would be a better example. If the *unemployment rate* fell from 10 to 5 percent, this would be a 50% drop in the rate. On the other hand, its inverse, the employment rate went from 90 to 95 percent: only a 5.6% increase. Calculating the percentage change in a percent is problematic because the use of a given measure or its inverse is entirely arbitrary. With labor statistics, unemployment rates rather than employment rates are the more commonly used indicator, but in the case of labor force participation measures, the percent of the population in the labor force is more commonly used than the percent of the population not in the labor force.

Presenting Data Efficiently.

Most statistical tables included in reference sources, such as the Statistical Abstract, contain multi-purpose tables and are not designed to illustrate particular points or to make particular comparisons. Typically these tables are designed as "look up" tables. An FBI table of crime statistics, for example, will provide quick access to a specific statistic such the murder rate in Illinois in 1999 (7.7 non-negligent homicides per 100,000 population). Generally, reference-source statistical tables will include much more information than would be included in a table included in an analytic paper.

State	1996	1999
Alabama	17.2	18.7
Alaska	11.2	10.0
Arizona	13.7	10.1
Arkansas	17.4	16.0
California	10.0	10.0
Colorado	10.0	10.0
Connecticut	10.0	10.0
Delaware	10.0	10.0
Florida	10.0	10.0
Georgia	10.0	10.0
Hawaii	10.0	10.0
Idaho	10.0	10.0
Illinois	10.0	7.7
Indiana	10.0	10.0
Iowa	10.0	10.0
Kansas	10.0	10.0
Kentucky	10.0	10.0
Louisiana	10.0	10.0
Maine	10.0	10.0
Maryland	10.0	10.0
Massachusetts	10.0	10.0
Michigan	10.0	10.0
Minnesota	10.0	10.0
Mississippi	10.0	10.0
Missouri	10.0	10.0
Montana	10.0	10.0
Nebraska	10.0	10.0
Nevada	10.0	10.0
New Hampshire	10.0	10.0
New Jersey	10.0	10.0
New Mexico	10.0	10.0
New York	10.0	10.0
North Carolina	10.0	10.0
North Dakota	10.0	10.0
Ohio	10.0	10.0
Oklahoma	10.0	10.0
Oregon	10.0	10.0
Pennsylvania	10.0	10.0
Rhode Island	10.0	10.0
South Carolina	10.0	10.0
South Dakota	10.0	10.0
Tennessee	10.0	10.0
Texas	10.0	10.0
Utah	10.0	10.0
Vermont	10.0	10.0
Virginia	10.0	10.0
Washington	10.0	10.0
West Virginia	10.0	10.0
Wisconsin	10.0	10.0
Wyoming	10.0	10.0
U.S. Average	10.0	5.7

Unlike the statistical data included in tables presented in a reference source, the information in a data display table consists of evidence (or premises) presented in support of a conclusion. Meaningful conclusions require more than citing a specific statistic, most often they require some kind of comparison: Although Illinois' 1999 murder rate is well below its 10.0 rate for 1996, it is still higher than that of all but seven states, and well above the national average of 5.7.

A properly formatted table (click on image to the left), allows the reader to quickly draw the right conclusion.

Sorting. Sort data by the most meaningful variable. Most reference sources will list data for geographic units (countries, states, or cities) alphabetically. The alphabet is almost never the most meaningful variable.

[Table 6](#) 

Not sorted		Sorted	
Youth Television Watching		Youth Television Watching	
	Percent of 9-year-olds who watch more than 5 hours of television per weekday		Percent of 9-year-olds who watch more than 5 hours of television per weekday
Canada	14.9	United States	21.5
Denmark	6.0	Spain	17.5
Finland	6.1	Canada	14.9
France	5.5	Netherlands	12.6
Germany	4.4	Ireland	11.8
Ireland	11.8	Italy	9.2
Italy	9.2	Finland	6.1
Netherlands	12.6	Denmark	6.0
Spain	17.5	France	5.5
Sweden	4.7	Sweden	4.7
United States	21.5	Germany	4.4

Source: Uri Bronfenburger, et. al. *The State of Americans* (New York: The Free Press, 1996)
from: William Bennett, *The Index of Leading Cultural Indicators* (New York: Broadway Books, 1999), p. 230

Source: Uri Bronfenburger, et. al. *The State of Americans* (New York: The Free Press, 1996)
from: William Bennett, *The Index of Leading Cultural Indicators* (New York: Broadway Books, 1999), p. 230

To fully appreciate the advantages of sorting, consider table 7.

Table 7: American League Central Standings

	W	L	Pct.	GB
Chicago	36	38	0.486	9
Cleveland	45	30	0.600	0.5
Detroit	32	42	0.432	13
Kansas City	30	46	0.395	16
Minnesota	46	30	0.605	-

(Thursday June 28, 2001) [notes on table format](#)

Decimal places and rounding. For most purposes, limit the number of decimal places to what are needed to display the data to two or three significant digits. It is usually not necessary to include dollar signs or percentage signs next to the numbers in a table, although this is sometimes done for the first number in a column.

[Table 8](#) 

Income, Poverty and Education, by Race and Hispanic Origin: 1998			
	Median Family Income	Poverty Rate	% High School Graduates
White	49,023	10.1	84.3
Asian*	52,826	12.5	84.7
Hispanic**	29,608	25.6	56.1
Black	29,404	26.1	77.0

*Asian and Pacific Islander

**Hispanics many be of any race

Source: Statistical Abstract, 2000. Tables 737, 755, 250

Howard Wainer, a leading authority on data presentation, insists that there is no reason to display more than two significant digits in most tabular displays. He would, therefore, eliminate the decimal points in the above table and round off the family income data to 49,000, 53,000 and 30,000 and 29,000. Presumably, he would have Major League Baseball record the Cubs' winning percentage as 49 percent rather than the .486 proportion. Hank Aaron's record of 755 home runs could be rounded to 760.

I think Wainer goes to far. It's true that readers will look at the income data in table 8 and, in their minds, round off to thousands. And the income data are based on estimates that make any conclusion based on differences of less than a hundred dollars practically meaningless. Percentages are usually fine without decimal points. But there are exceptions. In recent years, the US poverty rate has ranged from 16 to 14 percent. Reporting these rates without decimals might obscure many significant changes. Reporting Major League batting averages with just 2 digits would result in many ties and fail to distinguish important differences.

As a general rule, similar data ought to be presented in the same column. Mixing data of different types in the same column is disorienting, as we see in table 8a and in the Welcome to Farmer City sign, below.

[Table 8a](#) 

Income, Poverty and Education, by Race and Hispanic Origin: 1998				
	White	Asian	Hispanic	Black
Median Family Income	49,023	52,826	29,608	29,404
Poverty Rate	10.1	12.5	25.6	26.1
% High School Graduates	84.3	84.7	56.1	77.0

Welcome to Farmer City!	
Founded	1837
Population	2114
Elevation	965 ft.
Total	4916

Time: In tables where the time points define the columns, display years in adjacent columns, from left to right. Where the time points are in a column, sort so that the most recent year is at the bottom.

Time series trend data of more than five time points is generally better displayed in a time series chart than in a table. Times series charts convey trends more efficiently than tables, but with some loss of accuracy. The advantages and disadvantages of each may be seen in these examples. Compare table 9, below, to the charts and tables in displaying [annual data, 1960-1999](#) . (click on the table itself to view the original Excel file).

	Army	Air Force	Navy	Marines	Total
1960	871	814	618	171	2,474
1970	1,320	791	693	260	3,063
1980	773	558	527	188	2,046
1990	746	535	605	197	2,083
1998	492	363	381	173	1,409

from: Bennet, Index of Leading Cultural Indicators, pp. 202-203

Consistency: When a paper or report contains more than one table, the formatting ought to be consistent across tables: same fonts, same heading style, and same borders. If the four branches of the armed services are displayed as they are in table nine, they ought to be sorted in the same order (despite the sorting rule, above) if the same categories are used in another table. (Note: to some extent, this rule is not followed in this report in order to show alternative formats).

Other Considerations. While cramming too much data and too many different kinds of data into a single table should be avoided, you should also look for opportunities to combine several tables into one.



Table 10, on the left, is derived from Christina Hoff Sommers' *War Against Boys* nicely summarizes in a single table what could have been presented in six. The basic format used here is ideal for presenting

crosstabular data when a single variable is crosstabulated against several other. Sommers uses these data to make two points. The first is that teachers favor girls over boys. The second point is more subtle: that the American Association of University Women who conducted the original survey (and who sponsored a report arguing that girls are ignored by teachers) suppressed the release of these data. A less argumentative title for the table might have been, "Boys and Girls Perceptions of Teachers' Gender Partiality."

This California Field Poll tabulation illustrates the same approach, but with the demographic variables defining the rows. These segmented tabulations provide for a very efficient tabulation. Additional demographic categories do not complicate the data display, while adding a variety of interesting comparisons that would not be as easy to make if one used several tables. See, for example, that age is not as strong a determinant of candidate choice as education, although age, surprisingly, is a better predictor of whether or not a voter is undecided.

Borders: A common and simple table format is used in most of the tables on these pages. It includes a thin straight border under the title and heading cells and under the main body of data. Occasionally a thicker border is added at the top of a table. There is usually no need for vertical borders. The Title is in bold. Putting the headings in bold is advised only if they are very short headings, and not if it is inconsistent with the format for other tables in the report.

APA style recommends that table titles be aligned to the left with the text underlined and that the Table number be placed (again aligned to the left) above the title. These style recommendations, however, are for papers that are not in final form, i.e., manuscripts that will later be formatted by a publisher. Few APA-style publications underline table titles. Not centering titles is a carry over from the days of the typewriter, when centering multi-line text was a complicated chore.

Citations and footnotes: the source(s) of the data and any footnotes should be listed at the bottom of the table.

Formatting tables with Microsoft Excel:

[practice exercise with instructions.](#) 

Tips on using Excel.

- Let Excel align your data.
 - Merge and word wrap cells for table title
*Format | Cells | alignment * wrap text * merge cells*
 - Center column headings, center data under headings.
 - To center data on the decimal point, use a custom format.
*Format | Cells | Number * custom*
examples:

- in table 9 ([above](#)) the "Army" column is custom formatted to: ?,000
- in table 6 ([above](#)) the custom format is ?0.0

(To center and align data in an html table, use a courier font and spaces.)

- By default, Excel aligns the contents of cells on the bottom of the cell. Changing the alignment so that the data is centered vertically is better, particularly with cells that have borders.
use: **Format | Cells | alignment *vertical *center**
- Avoid fancy fonts: I recommend that a consistent 10pt font for the table, bold for the titles and headings, perhaps an 8pt font for the source. For tables that are to be printed in a paper, do everything in black and white.
- For tables that are used in PowerPoint or overhead presentations, use simpler tables with large bold fonts. PowerPoint tables should generally have fewer than 24 data points.
- Web page tables: Most of the tables displayed here were constructed using MS Excel, using **shift-Edit | Copy picture** and pasted into a single-cell table in MS Front Page. Tables 1, 4 and 7 are standard HTML tables. Note the difficulties with HTML tables:
 - the font size is set by the browser that is viewing the page, headings, in particular, may display differently with different browser and different users.
 - there is very little control over the use of borders in an HTML table, and using horizontal lines to display borders doesn't look good.

[References](#)

also: ([Field poll](#))

