How to make graphs misleading

1. **Clutter.** Make it hard to find points that should be compared. Add decoration to distract the eye from the part showing the data.

2. **Cut axes.** Exaggerate (or shrink) differences by starting scales somewhere other than zero, or by cutting out a segment of an axis.

3. **Change scale.** Alter the scale in mid-graph, so that things that <u>look</u> comparable aren't.

4. **Use pictures.** Don't use boring bars to make a bar graph. Use a picture of something.

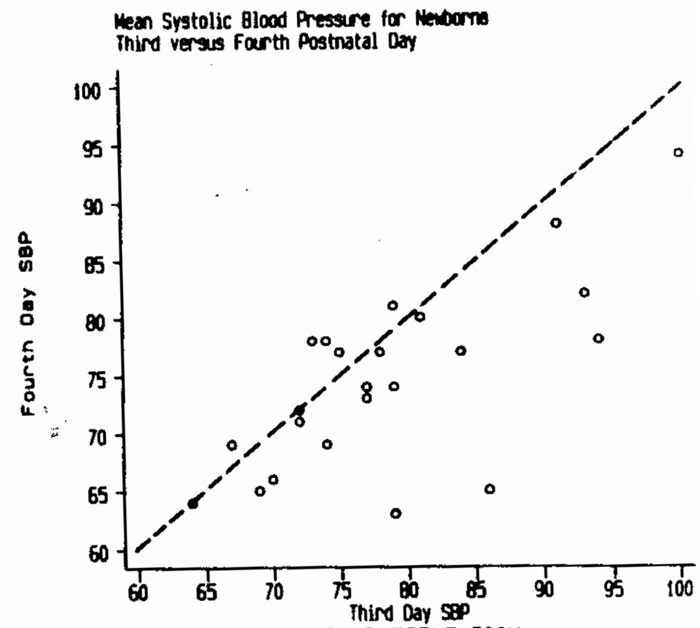
5. **Make blunders.** Label axes incorrectly, interchange variables, etc.

Blood Pressures of 23 Newborns Taken During Non-Rapid Eye Movement Sleep

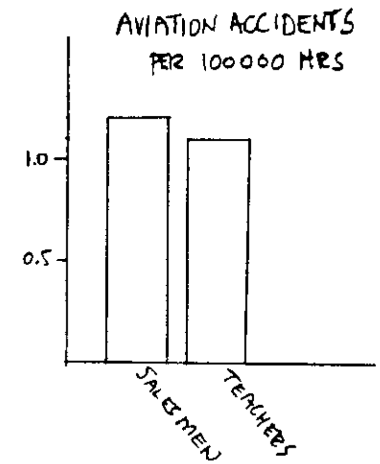
Mean Systolic Blood Pressures (mm Hg)

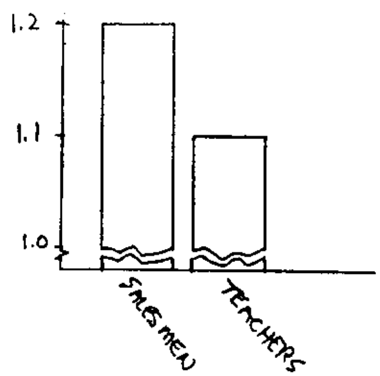
Individual	Day 3	Day 4
1	65	69
2	71	72
3	78	73
4	80	81
5	63	79
6	69	67
7	69	74
8	78	74
9	74	79
10	66	70
11	77	75
12	64	64
13	77	84
14	73	77
15	77	78
16	74	77
17	94	100
18	78	94
19	82	93
20	65	86
21	81	79
22	72	72
23	88	91
Average	74.6	78.0

Ref: Pediatrics 58(2), Aug. 1976

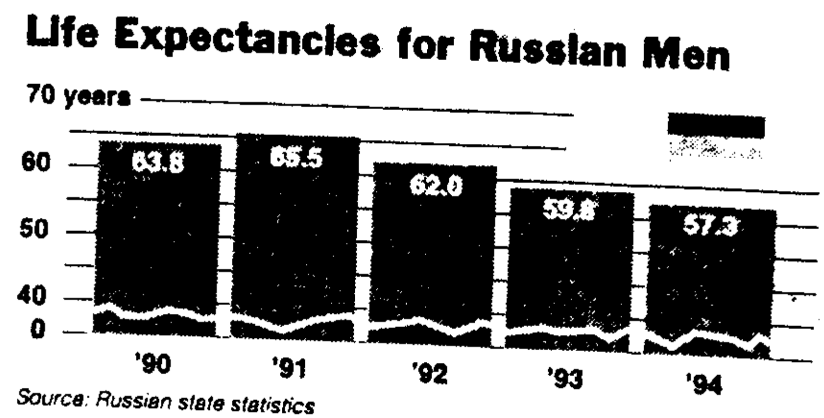


SCATTERPLOT PRESENTATION





Example: New York Times (August 2, 1995) Within masses of data on the people of Russia (ages, births, deaths), public health workers discovered some shocking patterns:



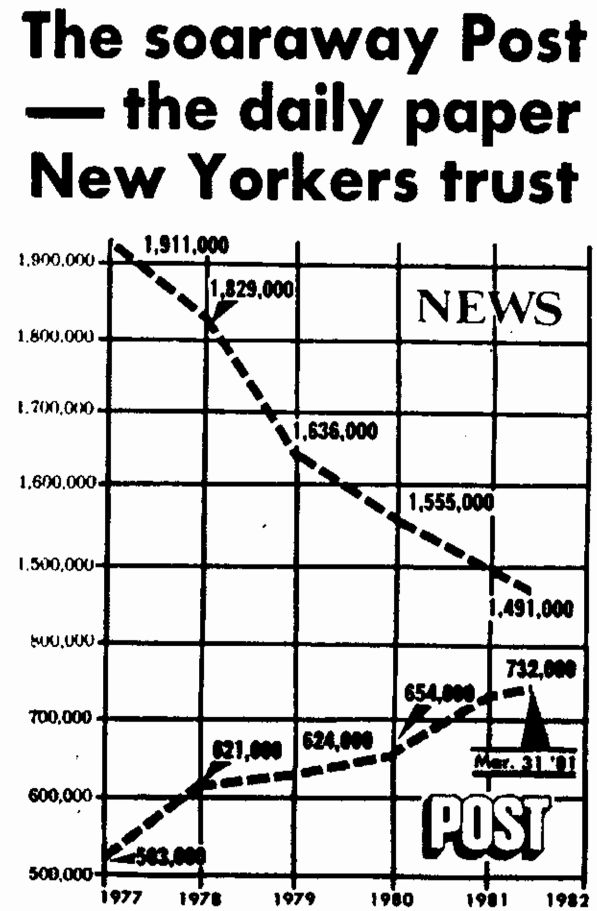
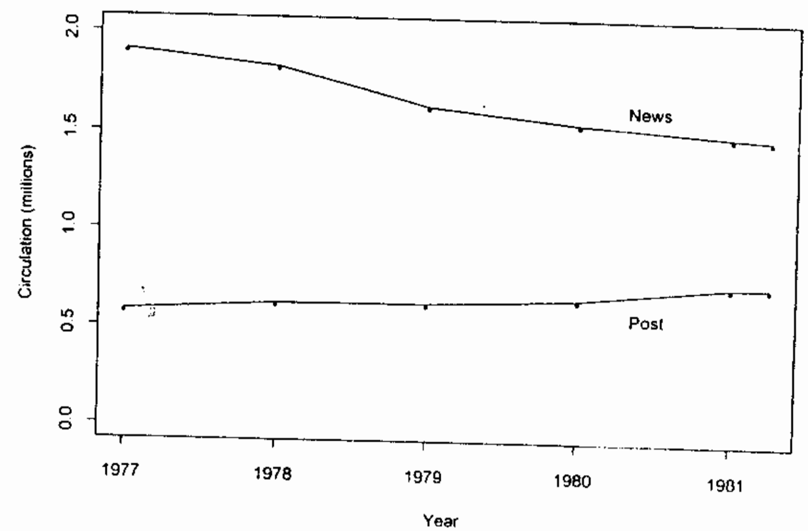
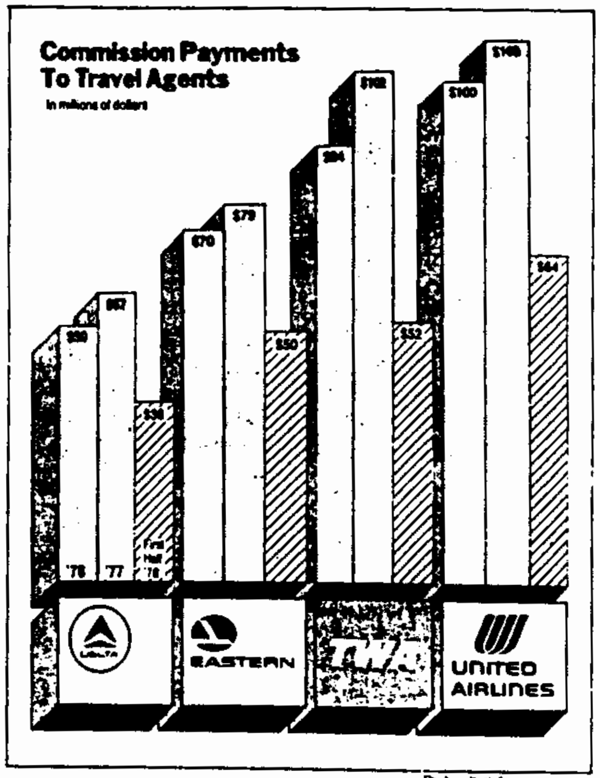


Figure 12. Changing scale in mid-axis to make large differences small (© 1981, New York Post).





Complex web of discount fares and airlines' telephone delays are raising travel agents' overhead, offsetting revenue gains from higher volume.

Figure 21. Mixing a changed metaphor with a tiny label reverses the meaning of the data (© 1978, The New York Times).

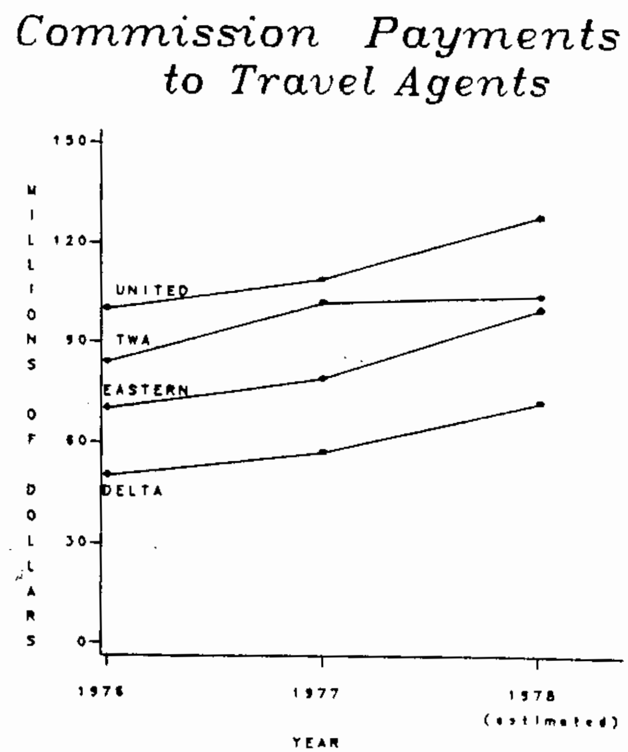
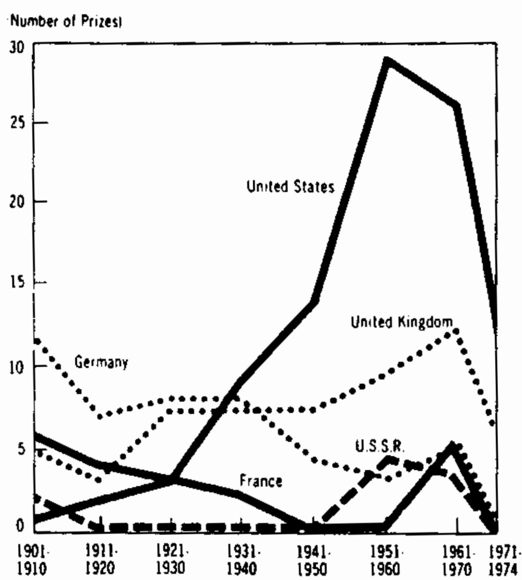
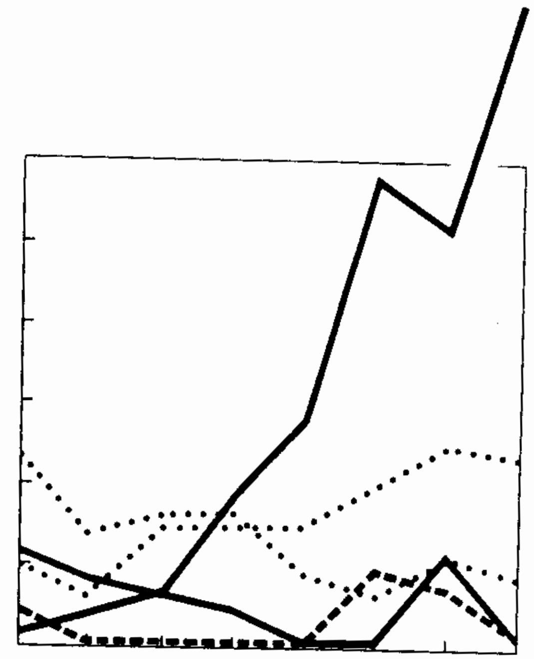


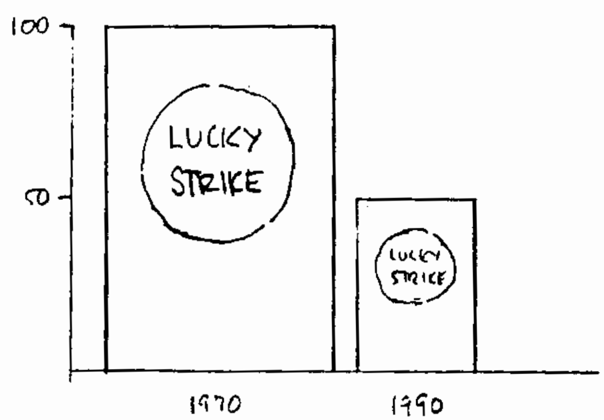
Figure 22. Figure 21 redrawn with 1978 data placed on a comparable basis (from Wainer 1980).

Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974



National Science Foundation, *Science Indicators, 1974* (Washington, D.C., 1976), p. 15.





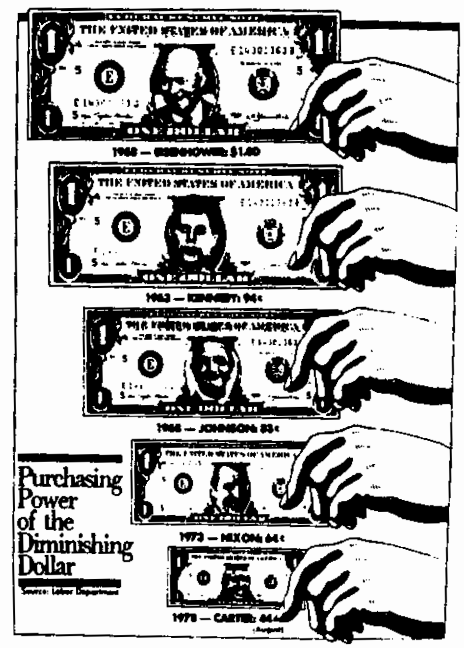


Figure 9. An example of how to goose up the effect by squaring the eyeball (© 1978, The Washington Post).

Table 2.5 Absolute frequencies of serum cholesterol levels for 1,067 U.S. males, aged 25 to 34 years, 1976–1980

Cholesterol Level (mg/100 ml)	Number of Men
80–119	13
120–159	150
160–199	442
200–239	299
240–279	115
280–319	34
320–359	9
360–399	5
Total	1,067

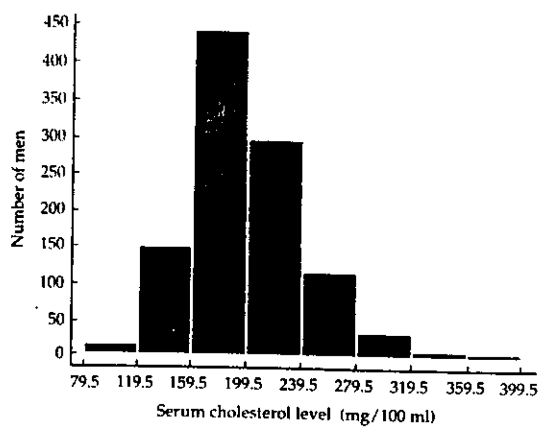


Figure 2.2. Histogram: Absolute frequencies of serum cholesterol levels for 1,067 U.S. males, aged 25 to 34 years, 1976–1980

Table 2.5A Absolute frequencies of serum cholesterol levels for 1,067 U.S. males, aged 25 to 34 years, 1976–1980

Cholesterol Level (mg/100 ml)	Number of Men
80–119	13
120–159	150
160–199	442
200–319	448
320–359	9
360–399	5
Total	1,067

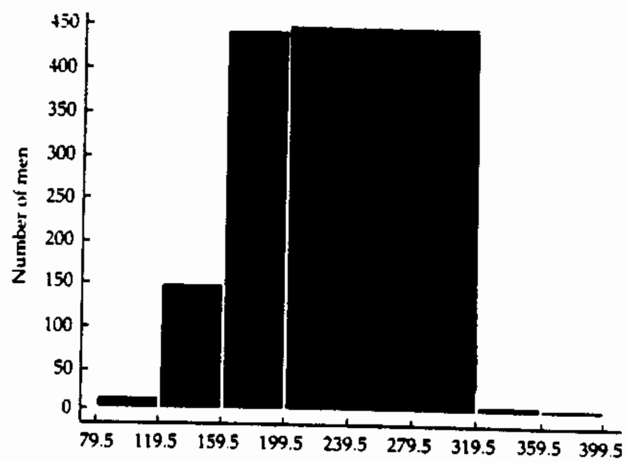


Figure 2.2.A Histogram: Absolute frequencies of serum cholesterol levels for 1,067 U.S. males, aged 25 to 34 years, 1976–1980

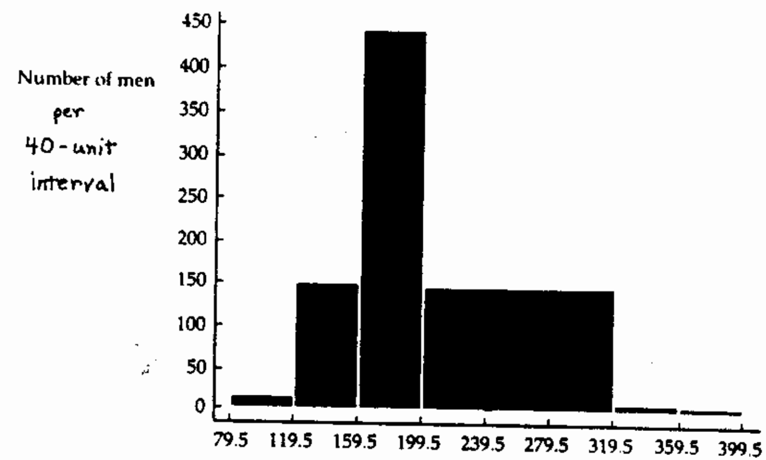


Figure 2.2.B Histogram: Absolute frequencies of serum cholesterol levels for 1,067 U.S. males, aged 25 to 34 years, 1976–1980

Table 3.3 Road accident casualties in the London Borough of Harrow in 1985 (excluding 65 with unknown age)

Age	Frequency
0– 4	28
5– 9	46
10–15	58
16	20
17	31
18–19	64
20–24	149
25–59	316
60+	103
Total	815

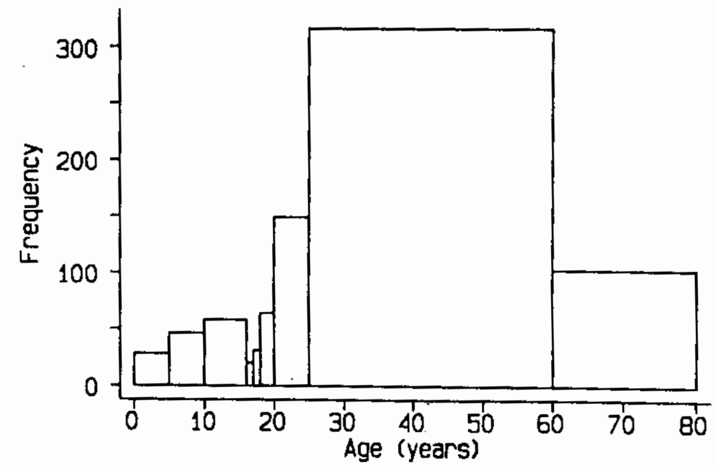


Figure 3.5 Incorrect histogram of road accident data of Table 3.3.

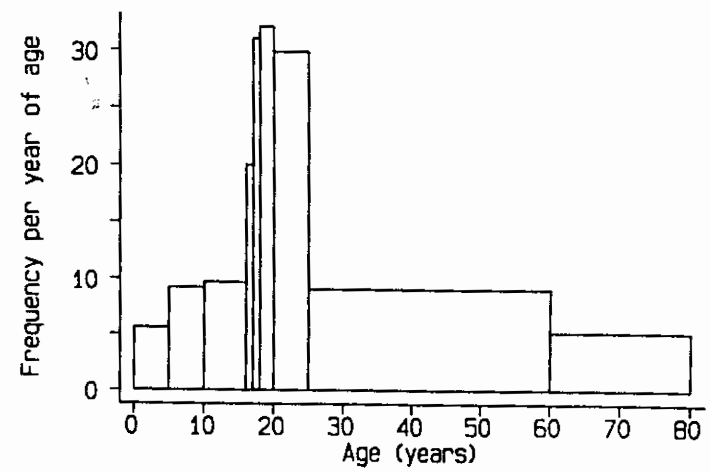


Figure 3.6 Correct histogram of road accident data.

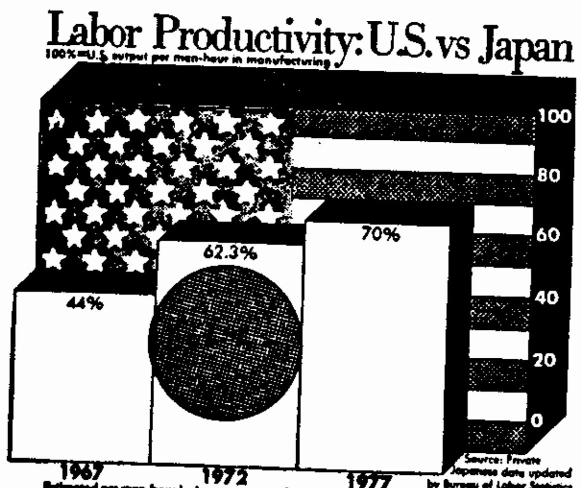


Figure 3. A low density graph (© 1978, The Washington Post) with chart-junk to fill in the space (ddi = .2).

Graphical presentation

In 1976 a Government publication[1] gave examples of some past successes in preventive medicine. One of these examples concerned the introduction in the 1930s of mass immunisation against diphtheria. Figure 1(a) shows their presentation of childhood mortality from diphtheria from 1871 to 1971. This appears to show that the introduction of immunisation resulted in a rapid decline in mortality. In their figure, however, mortality is plotted on a logarithmic scale and shows proportional changes. When the data are plotted on a linear scale,[3] as in fig 1(b), the visual effect is quite different, as is the interpretation. From this figure we can see that over the period in question mortality from diphtheria had been dropping very quickly, and this specific preventive measure was adopted relatively late in the day. This is not to say that the introduction of immunisation was not effective, but that the degree of its effectiveness that one accepts depends considerably on which way the data are presented.

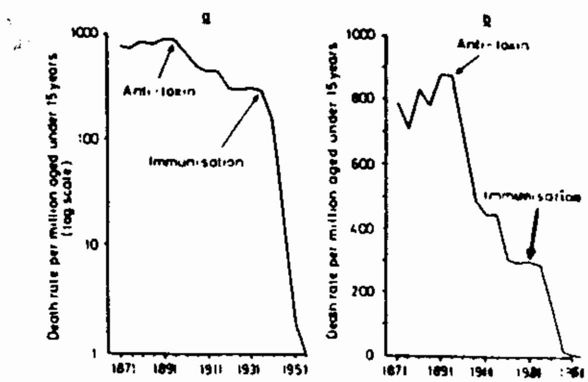


FIG 1—Childhood mortality from diphtheria (a) on a log scale[2] (b) on a linear scale.[3]

Source: Brit Med Journal 281, 1980

211

Guidelines for Graphs

Simplicity vs. Clutter: Clarity is the goal, and it is usually achieved through simplicity. A cluttered graph can result from trying to pack too much information into it (trying to do too much at once). It can also result from adding meaningless decoration, which might simply reveal poor judgement and bad taste, but which might also result from a deliberate effort to confuse or mislead.

Read title; Inspect axes; Ask questions

What am I looking at?
What does each axis represent? ...counts? ...rates? ...logarithms of rates? (When you make a graph be sure to label the axes, so your reader can answer this question)
Where is zero?
Are there breaks in the axes? ... shifts in the scales? (e.g. interval that represents one year suddenly represents only 6 months)
If it looks like a histogram, <u>is</u> it a histogram? (Are <u>areas</u> meaningful?)

Guidelines for Tables

1. Round numbers
2. Include totals, averages, other summaries
3. Make comparisons across columns
4. Order
5. Experiment with spacing, layout

example: Consider the following table which presents the number of deaths due to four specific causes in Baltimore City from 1950 to 1980.

Cause	1950	1960	1970	1980
Tuberculosis	535	163	94	12
Heart Disease	4,399	4,796	4,457	3,689
Cancer	1,582	1,856	2,018	2,054
Homocide	101	104	234	215

As presented it is not easy to detect trends in the table or other patterns in the data. We now consider the effect of some simple guidelines on the presentation of this data.

Guideline 1: Round

In order to understand a table of numbers it is almost always easier to do so if the numbers do not contain too many significant figures. The table above has the following appearance after rounding to two significant figures:

Cause	1950	1960	1970	1980
Tuberculosis	540	160	94	12
Heart Disease	4,400	4,800	4,500	3,700
Cancer	1,600	1,900	2,000	2,100
Homocide	100	100	230	220

Trends now are easier to see and there is much less "clutter" in the table.

Note: "A significant figure is any "digit" in a number that contributes to specification of its magnitude, apart from 0's which solely determine the position of the decimal point." (Johnson and Kotz: Encyclopedea of Statistics)
examples:

- 101.1 has four significant figures
- 101100 has four significant figures
- .001011 has four significant figures
- .0010110 has five significant figures

Note: To round a number leave the digit unchanged if the following digits are between 0 and 4.99..., increase the digit by 1 if the following digits are are greater than 5 and round the digit to the nearest even number if the following digits are exactly equal to 5.
examples:

- 40.14 rounded to three significant figures is 40.1
- 40.16 rounded to three significant figures is 40.2
- 40.15 rounded to three significant figures is 40.2
- 40.25 rounded to three significant figures is 40.2

Guideline 2: Add Averages or Totals

Adding appropriate row and/or column averages or totals to a table often provide a useful focus for establishing trends or patterns. In this table we add yearly totals as follows:

Cause	1950	1960	1970	1980
Tuberculosis	540	160	94	12
Heart Disease	4,400	4,800	4,500	3,700
Cancer	1,600	1,900	2,000	2,100
Homicide	100	100	230	220
Total Deaths	6,600	7,000	6,800	6,000

Guideline 3: Numbers are Easier to Compare in Columns

In the example we rearrange the table as follows:

Year	Tuberculosis	Heart Disease	Cancer	Homicide	Total
1950	540	4,400	1,600	100	6,600
1960	160	4,800	1,900	100	7,000
1970	94	4,500	2,000	230	6,800
1980	12	3,700	2,100	220	6,000

Guideline 4: Order by Size

A more effective presentation usually is based on rearranging so that the largest (and presumably most important numbers) appear first:

Year	Total	Heart Disease	Cancer	Homicide	Tuberculosis
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12

Guideline 5: Spacing and Layout

It is useful to present tables in single space format and not have a lot of "empty space" to detract the reader from concentrating on the numbers in the table.

Year	Total (4 causes)	Heart Disease	Cancer	Homicide	TB
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12
1990	5,100	2,600	2,300	280	9

Remove unnecessary lines from the table.

Year	Total (4 causes)	Heart Disease	Cancer	Homicide	TB
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12
1990	5,100	2,600	2,300	280	9

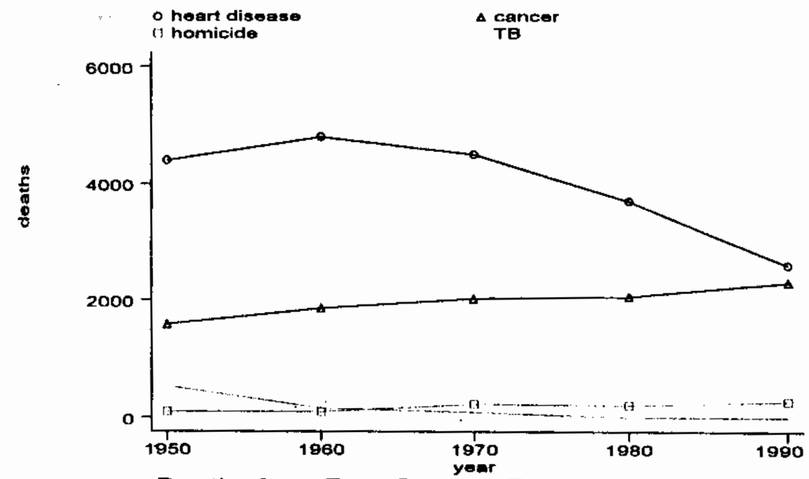
Year	Total (4 causes)	Heart Disease	Cancer	Homicide	TB
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12
1990	5,100	2,600	2,300	280	9

Year	Total (4 causes)	Heart Disease	Cancer	Homicide	TB
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12
1990	5,100	2,600	2,300	280	9

Maybe we've gone too far.

Baltimore City : Deaths from Four Causes

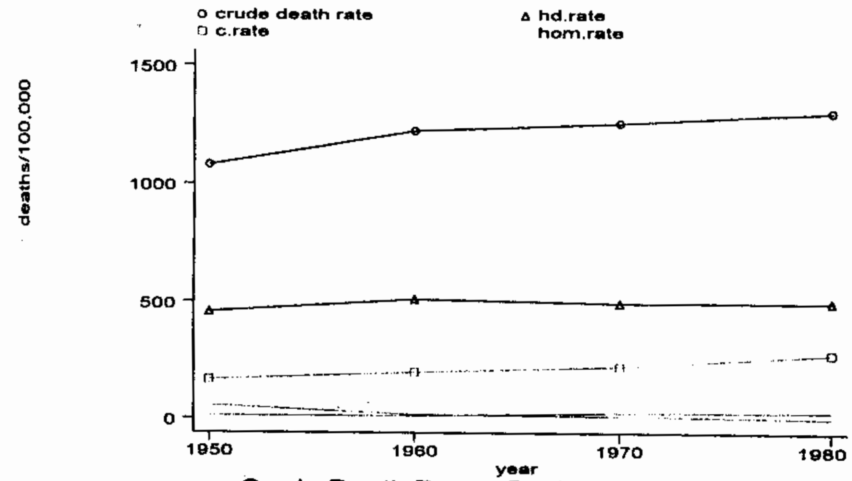
Year	Total (4 causes)	Heart Disease	Cancer	Homicide	TB
1950	6,600	4,400	1,600	100	540
1960	7,000	4,800	1,900	100	160
1970	6,800	4,500	2,000	230	94
1980	6,000	3,700	2,100	220	12
1990	5,100	2,600	2,300	280	9



Deaths from Four Causes: Baltimore City

Baltimore City : Crude Death Rates per 100,000

Year	All Causes	Heart Disease	Cancer	Homicide	TB
1950	1,120	460	170	11	56
1960	1,200	510	200	11	17
1970	1,300	490	220	26	10
1980	1,300	500	280	29	2



Crude Death Rates: Baltimore City