Statistical Science
1990, Vol. 5, No. 3, 327–339

Data-Based Graphics: Visual Display in the Decades to Come

John W. Tukey

Abstract. Visual display based on data deserves careful attention to a long list of ideas and questions (19 are discussed below). While classical views of graphical display need to be re-examined and selectively used, the computer—mainly as display maker, but significantly as number cruncher—has so greatly enhanced our potentialities that we have much to explore and many important steps to take. In particular, we need to pay serious and continuing attention to securing: (a) immediate and strong impact, (b) easy flow of attention across parallel elements, (c) planning to show phenomena, not numbers, (d) attention to both prospecting for what the data might show and transfer (to others) of what we have learned from it, (e) partnership with computation, and (f) putting disproportionate response to work. The next decade or two should see major advances.

Key words and phrases: Impact, easy flow, showing phenomena, prospecting versus transfer.

INTRODUCTION

Attention to *graphical display* originating with data, at least some aspects of which are usually numbers, has grown greatly in the last two decades. There is every reason for this growth to continue—even to accelerate. Computer systems are reaching the flexibilities of production and control that will make almost any choice easy to implement on paper, not just on the screen. People have begun to ask below-surface questions and develop deep enough schisms to generate competitive pressures. Millions of computer systems with some graphic capabilities are already out there, and software houses will need new enhancements to sell. Data analysis has been taken seriously by enough people to build an increasing base on which more effective displays can stand. Desktop publishing will soon make a square-inch of picture no more expensive than a square-inch of text.

Both professionals, concerned with growth and innovation, and amateurs, concerned with learning to do, will gain much from (i) keeping their minds open to new ideas and attitudes, (ii) absorbing enough of most of these ideas to make tentative judgments of their relevance to a specific situation, and (iii) acting on these judgments, while keeping them tentative and under revision.

The judgments put forth in this paper are themselves tentative, and will, I am sure, by revised (as some have in the last few months), but I am acting on them, both in my writing and in my research.

What ideas and questions are driving progress today? While order is a matter of taste, I believe strongly that all of the following should be on the list:

(1) the importance of impact (enforcement of attention);
(2) the importance of easy flows;
(3) the relative importance of phenomena and numbers;
(4) the distinction between prospecting and transfer;
(5) the importance of recognizing the purpose to be served;
(6) the need for computation before display;
(7) discovery of the principles that govern busyness (visual clutter);
(8) learning from the classic era of statistical graphics;
(9) understanding classical choices anew;
(10) incorporating the techniques and principles of the graphic designer;
(11) moving toward graphic-making systems;
(12) evaluating competing graphics;

John W. Tukey is Senior Research Statistician at Princeton University. His mailing address is Fine Hall, Princeton University, Washington Road, Princeton, New Jersey 08544-1000.

(13) finding good uses for dangerous ideas ("data ink ratio" and "one picture only");
(14) learning to use different kinds of disproportionate response;
(15) learning to be effective in reducing unnecessary distraction;
(16) understanding the constituents of high impact;
(17) learning what color can and cannot do for us;
(18) making dynamic graphics powerful and more widely used;
(19) recognizing the place of planned, artificial irregularity of emphasis.

It is my responsibility then to make some aspects of each of these ideas or questions clear, and, as appropriate, to express judgments about their consequences.

In doing this, I intend to treat making visual displays as something done by many people who want to communicate—often, on the one hand, to communicate identified phenomena to others, and often, on the other, to communicate unidentified phenomena to themselves. This broad clientele needs a "consumer product," not an art course. To focus on a broad array of users is not to deny the existence of artists of visual communication, only to recognize how few they are and how small a share in the total volume of communication they can contribute. For such artists, very many statements that follow deserve escape clauses or caveats.

1. IMPACT, NOT ARCHAEOLOGY

The greatest possibilities of visual display lie in vividness and inescapability of the intended message. A visual display can stop your mental flow in its tracks and make you think. A visual display can force you to notice what you never expected to see. ("Why, that scatter diagram has a hole in the middle!") On the other hand, if one has to work almost as hard to drag something out of a visual display as one would to drag it out of a table of numbers, the visual display is a poor second to the table, which can easily provide so much more precision. (Here, as elsewhere, artists may deserve an escape clause.)

Another important aspect of impact is *immediacy.* One should see the intended at once; one should not even have to wait for it to gradually appear. If a visual display lacks immediacy in thrusting before us one of the phenomena for whose presentation it had been assigned responsibility, we ought to ask why and use the answer to modify the display so its impact will be more immediate. (As an aside, notice that psychological research seems to have shown that those text styles, like caps and lower case instead of all lower case, that are read more slowly also transfer more information to the reader. Perhaps we have erred in thinking that textual account best that looks smoothest and reads most smoothly.)

With a decade's more background from innovation in visual display and insight into why (at least at a practical level) some choices in visual display are usually better than others, we may then be ready to return to thinking about running text and asking how our classical aims and choices about running text should be modified, especially in view of new technological possibilities.

But today our concern is with the generation and management of impact in visual display. One example may suffice to point out how thinking about impact can clarify matters. Edward Tufte's tour de force, *The Visual Display of Quantitative Information,* puts forward many new ideas (with about half of which I fully agree). In one particular instance, it applies the minimization of ink to a form (schematic plot) of box plot in a way that *strips* it of almost all its *impact—its ability to enforce attention.*

The basic elements displayed in a schematic plot—a particular form of box plot (EDA, 1977, page 48)—are: median, hinges (near quartiles), adjacent values (the observations nearest to but still inside the fences) and the individual observations that fall outside the inner fences (which are defined to fall at "2.5 × one hinge MINUS 1.5 × the other hinge").

In personal communication, Tufte points out an unfortunate concomitant of a string of similarly placed schematic plots of the style I prefer—where, as he says, the boxy version "generates active white stripes between the boxes." This phenomenon can occur and needs to be thought about carefully. I wait with great interest (a) to see what my own thoughts eventually lead to, and (b) to see Tufte's next book, which promises attention to such questions.

Figure 1 shows (i) the original schematic plot, put together solely by "instinct"; (ii) the Tufte stripped plot, a subject for graphic archaeology (probably with a hand lens); and (iii) the modification of (i) I would now favor, as responding to ideas of instrinsic and situational emphasis and a desire for balanced emphasis.

On the page facing his almost "stripped" plot, Tufte shows an alternative of the schematic plot which begins to respond to the ideas of impact and balanced emphasis. (Why did not this later style destroy the style on page 124 before publication?) Figure 2 shows (i) the style from Tufte's page 124, (ii) the style from Tufte's page 125, and (iii) two modifications of the style of Tufte's page 125 that bring the total emphasis on the hinges up closer to the level of the total emphasis on the median and the total emphasis on each adjacent value. The final style of Figure 2, while not nearly as effective in conveying a diversity of aspects

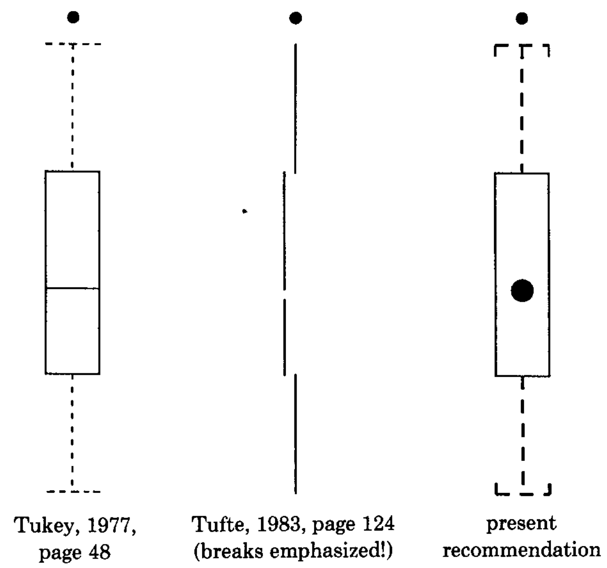


FIG. 1. *Three stages of evolution (backward as well as forward) of the schematic-plot style.*

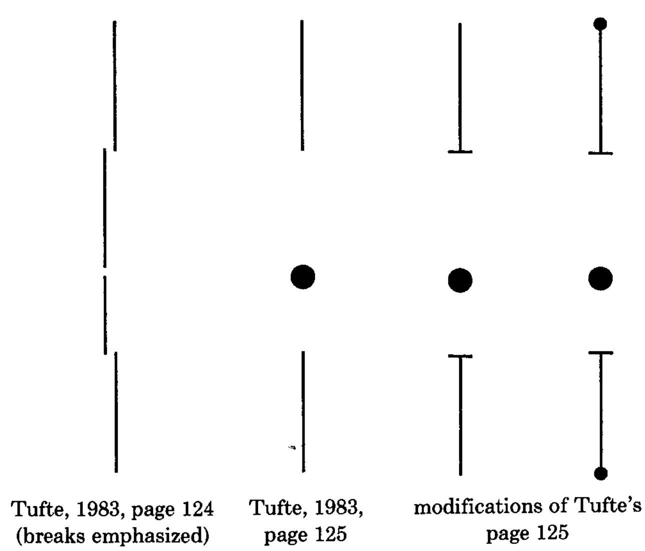


FIG. 2. *Same as Figure 1, but omitting outside values.*

as that of the final style of Figure 1, might be a possible compromise between Tufte's ideas and my own.

In thinking about the relation of the compromise to the "boxing" version, we need to give some attention to emphasis on a higher level, to the emphasis on parts of an array of linkages. One reason for the box, in the classical style, was to emphasize the middle half of the distribution; to emphasize "central clumping." The compromise style emphasizes this only through white space, the *space between* the two ends-emphasized bars. Is this strong enough? How should we make such judgments?

Because the elements in a schematic plot are linked together in an order, the *situational* impact of the elements at and near either end of the order will be greater than the situational impact of elements near the center. If we want balanced impact, as we have good reason to (see below), we need to choose ways of showing the elements that will give greater intrinsic impact to the center elements. On the right of Figure 1 this is just what we have done by using the four elements

● ——— - - - - - and •

whose intrinsic impact, as individual isolated elements, is strongly and obviously ordered.

2. EASY FLOW, USUALLY ONE PATH AMONG SEVERAL

Almost everything we do with data involves comparison—most often between two or more values derived from the data, sometimes between one value derived from the data and some mental reference or standard. The dedication of Richard Hamming's book on numerical analysis reads "The purpose of computation is insight, not numbers." We need a book on visual display that at least implies "The purpose of display is comparison (recognition of phenomena), not numbers."

Thus it is fortunate that most of what visual displays are effective in conveying involves comparison. Comparison means a shift of attention, often back and forth (perhaps repetitively so) between two or more targets. If we think of flow of attention along a path that runs across a sequence of targets and try to make this easy and effective, we will do about as well as we can to prepare for the more complex shifts of attention/perception that actually go on.

If we have only one sequence of targets, each described by one number—possibly means or medians under different circumstances or at different dates—our task is relatively easy; we have little to do but to make each target vivid enough and to avoid undue busyness.

But we usually have alternative sequences of targets and want to do well by each. In a schematic plot, for example, we have alternative sequences of location targets (the first kind of target sequences), involving: (1) the median, (2) a chosen hinge, and (3) a chosen adjacent value. Sometimes other alternatives, such as outside values or the midhinge, may also provide location sequences.

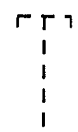
We also have separation targets, involving: (4) distance from median to chosen hinge, (5) distance between hinges, (6) distance from chosen adjacent value to the nearest hinge, (7) distance between adjacent values, and (8) overall impression of spread.

Beyond this, there are pattern targets, including: (9) placement of median (central, low, high) between the hinges, and (10) chosen-adjacent-value-to-hinge separation in comparison to hinge-to-hinge separation. If we are designing a schematic plot for general use, we need to seek for easy-sweep behavior for as many of these target sequences as we can.

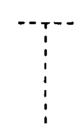
Making it easy to do any chosen location sweep calls for making it easy to see each location element, something that seems to be best done by making the impacts of all kinds of the location elements about the same, so no one will be pushed down toward unnoticeability by the others. Hence the desire for balanced impact above (Section 1).

But this is not enough; we also need the elements to be so distinct from one another that there is no chance of skipping from one to another as we sweep our attention across the sequence. This we have done quite effectively in the revised schematic plot, where what the elements point out are: (a) the center of the large dot, (b) the two ends of the solid box, (c) the two ends of the dashed whiskers, and (d) a small dot for each outside value; each is qualitatively different from all the others.

In the last section, we showed the adjacent values as



although we discussed them as if they were shown as



We can now see some reason for the difference. The turned-down ends draw some attention to the relation of the adjacent values to either the near hinge or the other adjacent value. The corresponding differences are involved in separation targets (6) and (7); thus turning down the ends slightly makes sweeping across a sequence of such separations somewhat easier and more effective. These "turn downs" also contribute to tying each schematic plot together visually, thus contributing to easy sweep for separation target (8). Not all readers will agree, however, some will prefer the simpler version.

Tuning the schematic plot for its location targets seems to have done fairly well for its separation and pattern targets, but I wish I understood much better and more generally how to make sweeping these two types of targets more effective. (Might gray lines help?)

Finally, we want to have an additional kind of distinctiveness. Other kinds of plots also focus on three or five values—for example, overlap plots to facilitate perception in a multiple-comparison situation. To the degree reasonably feasible, we want the whole appearance of different standard styles of plot to be so different that there is never doubt about what we are looking at. Figure 3 contrasts one choice for (double) overlap plots with our final choice for a schematic plot. (The same five numbers are involved in both.) Clearly we have been able to distinguish the two kinds of plots, and we have also been able to provide reasonably balanced impact for the outer and inner notches of the overlap plot. (The central hole deserves visibility, but not equal impact, because the overall message of this display emphasizes expressed uncertainty, not unreal precision.)

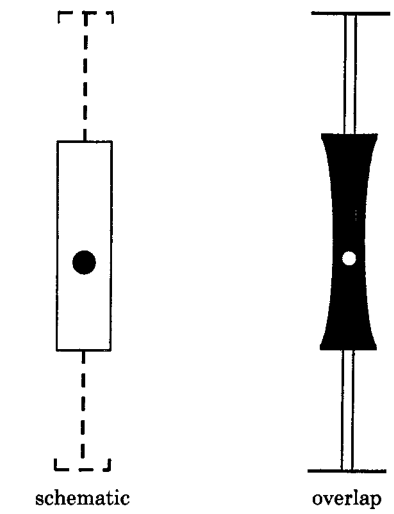


FIG. 3. *Schematic-plot and overlap-plot symbols sharing the same five numbers (almost certainly* not *for the same data set). The schematic-plot symbol is here, though not in general, symmetric so it can share the same five numbers with the overlap plot. For the same reason, its adjacent values are surprisingly close to its hinges. If the heavy central notches of two overlap-plot symbols overlap, the corresponding determinations are not significantly different at the chosen* individual *error rate. If two whole symbols overlap, the determinations are not different at the chosen* simultaneous error rate. *Failure to overlap shows significant difference in the indicated direction.*

3. PHENOMENA VERSUS NUMBERS

Much of what we want to know about the world is naturally expressed as *phenomena*, as potentially interesting things that can be described in *non* numerical words. That an economic growth rate has been declining steadily throughout President X's administration, for example, is a phenomenon, while the fact that the GNP has a given value is a number. With exceptions like "I owe him 27 dollars!" numbers are, when we look deeply enough, mainly of interest because they can be assembled, often only through analysis, to describe phenomena. To me phenomena are the main actors, numbers are the supporting cast. Clearly we most need help with the main actors.

The relative importance—as what we are to extract from visual displays—of phenomena and numbers is clearly reflected by a deep schism, one we can hope will either lessen, or become very well understood. The title of Ed Tufte's book, *The Visual Display of Quantitative Information*, stands in deep contrast to the title of one of my research proposals, *The Visual Display of Quantitative Phenomena*. It would be even better to go further, and say *The Visual Display of Phenomena With Quantitative Roots*.

If you really want numbers, presumably for later assembly into a phenomenon, a table is likely to serve you best. The graphic map of Napoleon's incursion into Russia that so stirs Tufte's imagination and admiration does quite well in showing the relevant phenomena, in giving the answers to "About where?," "About when?" and "With roughly what fraction of the original army left?" It serves certain phenomena well. But if we want numbers, we can do better either by reading the digits that may be attached to the graphic—a simple but often effective form of table—or by going to a conventional table.

The questions that visual display (in some graphic mode) answers best are phenomenological (in the sense of the first sentence of this section). For instance:

- Is the value small, medium or large?
- Is the difference, or change, up, down or neutral?
- Is the difference, or change, small, medium or large?
- Do the successive changes grow, shrink or stay roughly constant?
- What about change in ratio terms, perhaps thought of as percent of previous?
- Does the vertical scatter change, as we move from left to right?
- Is the scatter pattern doughnut-shaped?

One way that we will enhance the usefulness of visual display is to find new *phenomena* of potential interest and then learn how to make displays that will be likely to reveal them, when they are present.

The absence of a positive phenomenon is itself a phenomenon! Such absences as:

- the values are all about the same
- there does not seem to be any definite curvature
- the vertical scatter does not seem to change, as we go from left to right!

are certainly *potentially* interesting. (We can all find instances where they *are* interesting.) Thus they are, honestly, phenomena in themselves. We need to be able to view apparent absence of specific phenomena effectively as well as noticing them when they are present! This is one of the reasons why fitting scatter plots with summarizing devices like middle traces (Tukey, 1977a, page 279 ff.) can be important.

4. PROSPECTING VERSUS TRANSFER OF RECOGNITION

We all need to be clear that visual display can be very effective in serving two quite different functions, but only if used in correspondingly different ways. On the one hand, it can be essential in helping—or, even, in permitting—us to search in some data for phenomena, just as a prospector searches for gold or uranium. Our task differs from the usual prospector's task, in that we are concerned *both* with phenomena that do occur *and* with those that might occur but do not. On the other hand, visual display can be very helpful in *transferring* (to reader, viewer or listener) a recognition of the appearances that indicate the phenomena that deserve report. Indeed, when sufficient precomputation drives appropriately specialized displays, visual display can even also convey the statistical significances or non-significances of these appearances.

There is no reason why a good strategy for prospecting will also be a good strategy for transfer. We can expect some aspects of prospecting strategy (and most techniques) to carry over, but other aspects may not. (We say prospecting because we are optimistic that we may in due course have good lists of possible phenomena. If we do not know what we seek, we are "exploring" not "prospecting.")

One major difference is in prospecting's freedom to use multiple pictures. If it takes 5 or 10 kinds of pictures to adequately explore one narrow aspect of the data, the only question is: Will 5 or 10 pictures be needed, or can be condense this to 3 or 4 pictures, *without appreciable loss*? If "yes" we condense; if "no" we stick to the 5 or 10.

If it takes 500 to 1000 (quite different) pictures, however, our choice can only be *between* finding a relatively general way to do relatively well with many fewer pictures *and* asking the computer to sort out some number, perhaps 10 or 20 pictures of "greatest interest."

In doing transfer, once we have one set of pictures to do what is needed, economy of paper (or plastic) and time (to mention or to read) push even harder toward "no more pictures than needed." But, even here, we must be very careful *not to insist*, as a necessity not a desideratum, that a single picture can do it *all*. If it takes two pictures to convey the message effectively, we must use *two*.

For prospecting, we will want a bundle of pictures, probably of quite different kinds, so chosen that someone will reveal the presence of any one of possible phenomena, of potentially interesting behaviors, which will often have to be inadequately diverse. Developing, and improving, bundles of pictures for selected combinations of kinds of aspects and kinds of situations will be a continuing task for a long time.

For transfer, we will need a few good styles to transfer each phenomenon of possible importance, so that, when more than one phenomenon deserves transfer, we can choose compatible styles and try to transfer two, or even all, of these phenomena in a single picture. (We can look for the opportunity to do this, but, when we cannot find it, we will use two, or more, pictures as necessary.)

For prospecting, we look at *long* lists of what *might* occur—and expect to use many pictures. For transfer, we select *short* lists of what *must* be made plain—and use as few pictures as will serve us well.

5. THE IMPORTANCE OF PURPOSE

Most of what we have so far said emphasizes the importance of purpose. Especially in transfer, we badly need a detailed understanding of purpose. Do we want the reader or viewer to sweep the location of the upper hinge or to sweep the separation of lower hinge and lower adjacent point. It may be wise to readjust impact within, say, a schematic plot, to encourage such emphasis. (We might, for instance, extend one box end somewhat outside each side of the box or increase that end's line weight, or we might double or raise the line weight of the dashed line between one adjacent value and its near hinge.)

Treating visual display as a tabula rasa which will automatically and unbiasedly analyze the data, without need of computation, is to give up most of its value by asking it to do only what it does relatively poorly. A scatter plot, for instance, can serve many purposes, but not all!) There is a tabula-rasa fallacy for display, as well as one for the application of scientific structure to the results of experience or experiment.

Prospecting needs to be based on an (undoubtedly incompletely formulated) long list of potential purposes. Its results provide a shorter list of potential phenomena from among which we pick some for transfer. The whole shape of each of our transfer displays is crafted to the purposes of the specific instance at hand.

6. THE NEED FOR PRECOMPUTATION

There can be no substitute for computation as a support for display, computation that goes as deep as may be needed, computation that involves more than one layer of recursion, computation that embodies almost everything we have learned or intuited about the possibilities of data analysis.

If our data consists of several observations in each cell of a row-by-column display, for example, our analysis has only begun when we have gone through a conventional analysis-of-variance calculation, including: (i) a breakdown of the observations into additively-combinable overlays, for instance common term, row effects, column effects, interactions, residuals, and, often, subdivisions of these; (ii) summarization of these overlays by mean squares; and (iii) assessment of error terms, quite possibly both for model I and for model II. (Especially if the pattern is at all complex, we will need to use the mean squares, or some subtle analysis, to guide aggregation of overlays.)

We need to go much further, for instance, converting the error terms into sizing information for overlap plots, perhaps, at each single level of break-out, one overlap plot for each of the three main kinds of quantity: combined fits, residuals and (long-run) cell values. With these under our eyes, we can begin to make pictures relevant to purposes that are more realistic or deeper, or both. Since we have not as yet analyzed purposes deeply enough in this area, we do not yet know *either* a list of all the pictures we are likely to need *or* a list of all the computations that will be needed to support making the pictures.

This is but one example of many. If we are going to get full value from visual display, we are going to need to do much to enhance, and to explain, the underlying data analysis.

7. AVOIDING BUSYNESS (VISUAL CLUTTER)

Most of us recognize that being unnecessarily "busy" is a serious defect of any display. Most of us can judge relative busyness between two displays. Few, if any, of us understand how busyness comes about.

We need to find, understand and put to use some principles relative to busyness. What these will eventually be is still unclear, but some very tentative suggestions may help to clarify what such principles might be like and, if we are sufficiently fortunate, may even stimulate someone to provide better candidates.

When dealing with the special case of display elements that are linked together (bars, simple or divided; schematic plots; overlap plots; etc.), it seems likely that we can reduce busyness by following these maxims:

- Avoid slanted lines within the elements, except possibly for pointed substructures that weaken emphasis and direct attention (cp. pencil-point plots in Hoaglin and Tukey, 1985n) or central x's.
- Use circular forms for dot-like subelements, unless there is serious need for more distinctions than size and fill of circles can provide.
- Avoid noncircular curved elements, unless, as in overlap-plot symbols, they are needed to strengthen a particular feeling (here a feeling of "a notch").
- Be ready to avoid busyness by splitting one picture into two or more.

- Avoid open forms where closed forms will serve as well. (See Figure 4 for instance.)

It seems to me that Tufte's concern with "data-ink ratio" was essentially intended to help avoid busyness (*vide* his concern with "chart junk"). Whether this is indeed so is a matter for Tufte's long-term introspection, but if true it would be the best excuse for what, in detail, I see as a mistaken emphasis. What would, for instance, an unremitting emphasis on "data-ink ratio" leave of the famous Napoleon-in-and-out-of-Russia chart?

We do need to reduce busyness, but we must retain impact. The tension between busyness and the dropping of unhelpful elements on the one hand and impact and the keeping of helpful elements on the other can be the basis of well-tuned visual displays.

8. BUILDING FROM THE CLASSICAL SKILLS

Visual display in the form of "statistical graphics" has a long and honorable history. Much progress was made by those people who could, and did, evaluate carefully which of two modes of presentation did better at showing what they wanted to show. We will need to take advantage of what was done, in what we may label as the pre-Bertin era, building upon it where direct use is feasible, learning to understand it in new ways and new terms wherever transmogrification is needed before it will contribute its proper share to the concepts and attitudes of the future.

Why is it that not all of the classic results and attitudes can be taken over directly as fundamentals? Mainly, I think, because, for too many workers, statistical graphics was thought of as a *substitute* for analysis by computation, rather than as a *partner*. The truth was thought to be present in the data as it was received, so that all that was needed was to display the data graphically (in any suitable way), after that one could trust that "the truth will make you free." The needs for the concepts and mental structures of science—changeable though they may be, often in the small but very occasionally in the very large—have by now erased the idea that scientific observations can be always seen on a tabula rasa. Now the needs for greater computational presupport, for analysis of the numbers in ways of ever increasing diversity before the results are plotted, will have to erase the idea that good visual display can *always* be a display of *raw data*. This is so even though the resulting preanalyses, which will have to be increasingly diversified, may also have to be, fortunately to a lesser degree, of ever-increasing complexity.

As a result our insights must go deeper, our principles must be broader, our illustrations must be more diverse, in order that what visual display and data analysis can do as partners will become more and more useful.

A graph or chart should not be just another form of table, in which we can look up the facts. If it is to do its part effectively, its focus—or so I believe—will have to be one or more phenomena, and it will often be unable to play its role without extensive computational support.

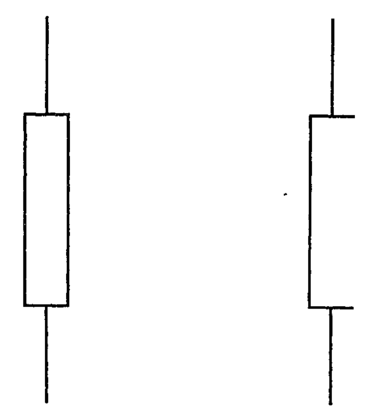


FIG. 4. *Leaving out a symbol closure can increase busyness and distraction (in a box and whiskers symbol). The openness of the right-hand symbol draws attention, both to what is not there and to why it is not there. It also directs our attention to the right, for no useful purpose.*

9. A NEED FOR NEW UNDERSTANDING OF OLD CHOICES

To use some of the insights of the pre-Bertin era, to let them contribute an appropriate share to the new synthesis, we will have to restudy them and reanalyze them in terms of new concepts and new principles. This needs to be thought of as saving the real essence of the old, not as destroying it!

Take simple bar charts as an example, which I would define much more generally than many classical authors. Why have they survived? Not because they are geometrically true, and not because they lead to good numerical estimates by the viewer! In my thoughts, their virtue lies in the fact that we can all compare two bars, perhaps only roughly, in two quite different ways, both "About how much difference?" and "About what ratio?" The latter, of course, is often translated into "About how much percent change?" (Going on to three or more successive bars, we can see globally whether the changes in amount are nearly the same, but asking the same question about ratios—rather than differences—requires either tedious assessment of ratios between adjacent bars for one adjacent pair after another, or going over to some other display, one necessarily based on precomputations.) If I am right about this, we will come to think quite differently about the use of bars, but we will be able to make use, probably in new ways, of much of the wisdom of our predecessors.

10. LEARNING FROM GRAPHIC DESIGNERS

The art of statistical graphics was for a long time a pen-and-pencil cottage industry, with the top professionals skilled with the drafting or mapping pen. In the meantime, graphic designers, especially for books, have had access to different sorts of techniques (the techniques of graphic communication), such as grays of different screen weights, against which, for instance, both white and black lines (and curves) are effective. They also have a set of principles shared in part with the Fine Arts (some written down by Leonardo da Vinci). I do not understand all this well enough to try to tell you about "visual centers" and how attention usually moves when one looks at a picture, but I do know enough to find this area important—and to tell you that more of us need to learn a lot more about it. (It will not wholly change our lives; book illustrations by good graphic designers rarely shock us. But it should help us both to ask new questions and to gain effectiveness by modifying our displays in previously unthought-of ways.) In particular, we can all try to learn from the works of Jacques Bertin.

11. STEPS TOWARD GRAPHIC SYSTEMS

Before long, most visual displays will be piloted—and more and more often finished—using some combination of computer, monitor and printer. This will happen because this will, by then, be the easy way. We will do easily not only what we are used to doing "by hand" but also use those detailed things that graphic designers are accustomed to use, like appropriate line weights (different for lines with different purposes), appropriate gray screen densities, flexibly sized (and eventually flexibly shaped) characters and the like. (Tufte's concern, in conjunction with "data-ink ratio" reduction, about the number of pen strokes required for a symbol will decay toward emptiness.)

We will need to avoid too rapid a freezing of what such a system would provide, but we all should do what we can to identify both capabilities that would help and simple ways to control their functioning.

We might for example keep the display under development on the monitor screen in as near to wysiwyg ("what you see is what you get") style as possible, and take off a printer plot for careful evaluation whenever we think we are doing significantly better. In such a *context* we might, for instance, be able to change the density of a gray-screened area with a mouse. We would hope to gradually converge to what is needed, both for a system and in individual instances.

We need to think of things to explore that have never been tried, sometimes for such obvious reasons as lack of availability. The question of what "mesh" a screen should have is a very different question when we can have dense sets of dots that are either locally irregular or locally regular. (The computer, and techniques of restricted randomization, can give us irregular-seeming patterns that are still quite uniform on the average. What they will prove to be good for is an interesting question, even an exciting one.)

As we move toward a good system for making visual displays in a computer-aided—nay computer-dependent—way, we will always have to keep in mind the partnership with computation and schemes of analysis and be sure, by careful trial and examination, that what comes out the far side has made good use of all the partner's abilities.

12. EVALUATING COMPETING GRAPHICS

We have begun to see evaluations of alternative forms of visual display, using experimental groups of subjects drawn from different experience pools, not just evaluations by individual inventors. In the large, we have to say that this is an important innovation, long overdue. But we are all going to have to be very careful about how we do such studies. Their planning and conduct need to reflect what we think of as the most important broad purposes of visual display.

As one who *denies* "reading off numbers" as the prime purpose of visual display, I can only denounce evaluating displays in terms of how well (given careful study) people read numbers off. If such an approach were to guide us well, it would have to be a very unusual accident.

How could we study "shows phenomena" with a group of subjects? I would suggest the following approach as one that should have merit.

Phase One (*Preparation*). Explain to each subject what each of, say, seven possible phenomena is like, using, in a balanced way, examples of all the styles of presentation to be compared.

Phase Two (*Experiment*). Give a subject a short-time (tachistoscopic) view of a single style applied to a single set of data, and then ask whether one specific phenomenon was present. Balance and randomize both presence and nonpresence and the different phenomena. Score correct or incorrect.

Phase Three (*Analysis*). Analyze the phase two data, perhaps to determine the time of display for 50% right, but more reasonably to assess the time for 90% right. Compare times across data set × style. Prefer the style in which 90% right is seen most rapidly.

Such experiments would be harder and would require greater involvement by trained psychologists (who might well tell us how to restructure the experiment) than those we have so far seen, but would test something much closer to the true purpose of visual display!

Even with moderately complex symbols like those of schematic plots, such an experiment can usefully be done on (i) each of several specific aspects (see Section 2) individually, and (ii) on small sets of specific aspects simultaneously. There will be good opportunities for all the work people may wish to do.

13. GAINING FROM "INK" AND SPACE RESTRICTIONS

Granted the presence of an important idea behind (or below) "data-ink ratio" and recognizing that its untempered use can be destructive, how can we reorient attitudes and excavate the important idea itself. It may well be that Tufte's underlying idea combines "avoid busyness" and "avoid distraction." Both of these can lead to trimming off elements but may not in fact, when carefully applied, favor removing all that could logically be removed without making reconstruction impossible. Turn back to Figure 4, where one side is taken out of a box (in a box-and-whiskers plot). This reduces "data-ink ratio" but, to my eye, increases both busyness and distraction.

We need to become keenly and broadly sensitive, both to busyness and to distraction. For the moment, the only suggestions about how to learn this I can offer is "practice, introspect and experiment." Perhaps others can do (or have done) better. Let us hope so.

Once we have these sensitivities, and an equally important sensitivity to impact, we are ready to put a variety of displays between the rock of busyness and distraction and the hard place of lost impact and see which ones seem to dominate the others.

It will not be sufficient to try a few styles, or just several styles, even if both macro- and micro-differences are represented. (My experience of nearly two decades ago with the development of a dynamic graphic display, PRIM-9, has left me convinced that 30 to 50 alternatives—perhaps realized as 30 to 50 steps, each hopefully forward—is not too many to reach the position we need to reach.)

A similar rock and a hard place arises when we try to squeeze down space requirements. Here busyness and distraction join with impact to form the hard place, while the need for adequate presentation, with its resistance to squeezing everything into a single one-inch square picture, forms the rock. The pressure of space saving will have its effect, but its effect must be a limited one.

14. PUTTING DISPROPORTIONATE RESPONSE TO WORK

Calculation can give us disproportionate responses to larger or smaller inputs as well as disproportionate responses to faster-changing or slower-changing inputs. (The latter is much of what classical smoothing is about!) Aptly planned nonlinearity of preparation (by computation) can enhance display in many ways. All our techniques of robustness, like most kinds of data editing, exploit disproportionality. We illustrate only two other instances.

Daniel's Expanded Residuals

In areas where both precise measurement and well-fitting empirical formulas are common, and there are such areas, Cuthbert Daniel's idea of expanding residuals compared to the fit, and using the result as a basis for display can be very helpful. Instead of displaying

$$y \equiv \hat{y}(x) + [y - \hat{y}(x)]$$

against x, where $\hat{y}(x)$ is some fit as a function of x (possibly even a straight-line fit) we can display, for instance

$$y^* = \hat{y}(x) + 7[y - \hat{y}(x)]$$

against x, where 7 stands for a well-chosen value depending on the instance, a value usually between 3 and 100. If $\hat{y}(x)$ is at all complex, we would have begun by displaying both data, y, and fit, $\hat{y}(x)$, on a single picture. Daniel's approach would lead us to go on to display both modified data, y^*, and fit, $\hat{y}(x)$, on a better single picture. Where some phenomenon we want to reveal involves a relationship of residual value to fitted value, this sort of approach can be very helpful.

In areas where residuals are so large as to almost drown out the fits, there may be a role for Leinad displays (Daniel spelled backward) where we plot, say

$$y_* = \hat{y}(x) + \frac{1}{7}[y - \hat{y}(x)]$$

against x, instead of y against x. In both Leinad and Daniel displays we are seeking to balance relative emphasis on fits and residuals.

Detrivialization

Smoothing is often thought of as "getting rid of the middle-sized wiggles but not bothering with the small ones." We can, in fact, often gain by doing the opposite: by *detrivializing.*

In a sense, significance testing—as opposed to confidence interval production, which is more flexible and helpful—is a classic example of detrivialization. In the form "forget everything not significant at 5% and attend only to the exact observed values of what was significant at 5%" it does a lot to avoid distraction and diffusion of attention. (Replacing "forget . . ." by "act as if everything not significant at 5% ... were zero" is of course a very bad deed—one that we ought to consider a crime! Conversely, replacing "attend only things significant at 5%" by stating a suitable confidence interval is an improvement!) The main point of significance analysis is disproportionate response (though usually not as part of a display process). Essentially, we are using an operator D that makes small things vanish and leaves larger things untouched.

Disproportion in smoothing is easily introduced by starting with a possibly naive, possibly sophisticated smoothing $y \to \check{y}$, applying some D to the rough, and adding back the result to the smooth to reach

$$y_{**} = \check{y} + D(y - \check{y})$$

in which $y_{**} = \check{y}$ for most values (all those for which $y - \check{y}$ was small), but $y_{**} = y$ for some values (those for which $y - \check{y}$ was large). There are some situations where this is very helpful and others where it is very bad. (It is, of course, an antithesis of the usual operations of robustness.) More sophisticated versions can be even more useful.

Close

The purpose of display is to make messages about phenomena clear. There is no place for a doctrinaire approach to "truth in geometry." We must be honest and say what we did, but this need not mean plotting raw data.

The point is that we have a choice, not only in (x, y)-plots, but more generally. Planned disproportionality needs to be a widely-available option, one that requires the partnership of computation and display.

15. REDUCING DISTRACTION/DIVERSION/ DEFOCUSING

We have just seen, in detrivialization, one example of reducing distraction. Another example arises when we take care to make our "smooths" smoother by polishing them. We do this (not at all because the polished *values* displayed will be made better, say, by being closer to what they "ought to be") but rather because we do not want the viewer, whether ourselves or a client, to be distracted by "little wiggles." What we are really polishing-immediately-before-display is the phenomenon. Most often, perhaps, we will do such polishing-immediately-before-display with proportionate smoothers, even though the initial smoothing, probably by being robust, may have been disproportionate. (Occasionally, of course, we may want to polish by detrivializing.) Again we stress the partnership of computation and display.

16. IMMEDIACY; INTEROCULARITY; INESCAPABILITY

We started with a desire for impact, mentioning both *inescapability* and *immediacy* as two of its constituents. Now we have discussed a variety of topics at least briefly, it is time to come back to impact in a little more detail.

A good picture of what makes up *moderate* impact does not seem to be available. If we think of a sequence of schematic plots with, say, five location targets, three separation or scale targets, and a couple of pattern targets, for each of which we want "adequate impact when appropriate," a clear analysis of impact for each of the specific targets is not easy.

But if we are dealing with fewer possible phenomena and large impact, we can do quite well by looking for these three elements:

- It hits you fast (immediacy).
- It hits you between the eyes (interocularity).
- You cannot avoid its message (inescapability or unavoidability).

When a display provides all three of these, it must "really hit you"—which means it has high impact. (Clearly we can only expect to attain such properties in full measure when either the phenomenon is very clearly present or clearly absent.)

Giving up a little of one or more of these three elements, in order to diversify the list of phenomena that our display will reveal if present, is often right. Whether there is any other good and sufficient reason for giving up something in the way of impact is far from certain. (It may be for some phenomena that there are no high-impact displays, however strongly the phenomena are present. In such instances, though, it is right to worry that we have perhaps merely not yet thought of the appropriate display.) Overall, though, we ought to push hard as we can for impact.

The Limits of Compromise

We have admitted that compromising *somewhat* on impact can be worthwhile if we increase the number of detectable phenomena considerably. We would allow some increase of distraction for a similar reason. But we ought not consider giving up any of: ease of use, a lusty dose of impact, or relatively slight distraction.

Above all, we should not accept displays that require archaeology, with or without hand lens, to draw from them what we need to learn.

17. WHAT CAN WE DO WITH COLOR?

We have concentrated, so far, on unchanging black-and-white displays. What about color? (And what about dynamic displays?)

On the whole, color is a disappointment. Adjusting emphasis by the use of color is certainly possible, but maintaining balance in the process is far from easy and often treacherous. Color is a moderately good labeling device, providing as many as three pure-color labels easily and as many as seven pure-color labels with some difficulty. Visual linking of similar things in different parts of the display through common color is probably feasible but far from easy. Transferring response-like information by color is at best very difficult. As an enhancement, color's great advantage is that it can be put on *paper*.

The rise of color copying will make the use of color in reports of all kinds, including statistical ones, more frequent, splashier, and perhaps more impactful. I do not yet see color, however, as a first-class opportunity to enhance the visual display of quantitative phenomena.

For qualitative phenomena at two or three levels we can use color effectively. The speedometer that turns red when you go too far above the speed limit—or the booklet of 2080 = 1770 + 300 + 10 scatter diagrams (involving 60 variables and 5 composites) which the 100 the computer thinks most interesting are printed in red—both illustrate how color can be an effective flagging device. But they do not call for new thoughts about visual display. (There are other ways to wave flags!)

The maps in certain atlases (including *The Oxford Atlas* and *The American Oxford Atlas*, 1951) use sophisticated sets of "layer tints" to show 10 zones of elevation and 3 of water depth. They show the phenomena (mountain ranges, marshes, etc.) quite well, but they are, for instance, of very limited utility for "reading-off" heights of individual locations. They may well be examples of the most we can do with color in difficult situations.

18. OPPORTUNITIES IN DYNAMIC GRAPHICS

Dynamic graphics is another matter. Here we are still quite seriously bound by technology, but we can see better times ahead. (Even today there are important possibilities.)

We can easily group three people about a monitor, or with a lot of difficulty make a video cassette. (Indeed, as my fifth cousin Paul Tukey and I have shown (Tukey and Tukey, 1990), there are real "lo-tech" possibilities for limited dynamics on overhead projectors. There may also be "middle-tech" possibilities as well.) Good impact on class-sized or meeting-sized audiences is difficult, but rewarding enough to make study and practice worthwhile. Dynamic displays for those who, like traditional readers of books, want to receive the message while individually alone and collectively scattered here and there are even more difficult to provide, but VCR tapes may well be the answer.

For the audience of three around the monitor, alternation of images is less demanding on the computer system than is static color. The same memory that stores three basic colors will store three basic images, and the same "color tables" can combine three black-and-white images with equal facility. There is a large variety of opportunities here, particularly with sequences, often cyclic, of 4 to 10 views. There will be principles about unfamiliar quantities, such as the ratio of the times two images appear on the screen, but the basic ideas—impact, easy comparison (now not necessarily by sweep), balanced emphasis, etc.—can and should be the same.

Depth cues, our most natural visual approach to a third dimension, are best sought by apparent rotation. Good things can be done with one form or another of rotation. (Other depth cues, like perspective and haziness, seem much less effective.) Fully simulated rotation can be demanding indeed of computer resources. But alternation among four to eight individual views can give a very good feeling of limited rotation.

The new ideas for dynamic display, in my opinion, are going to come from alternation rather than from rotation. Let us welcome as much dynamic character as each individual can afford to put in his or her graphics, whether for prospecting or for transfer. And let us try to learn to do it effectively. This can, I believe, be a major enhancement.

19. PLANNED IRREGULARITY: RIPPLING EMPHASIS

We are going to illustrate our closing point with a column display, where any display component that visually connects a lower point with an upper one is called a *column*, and any parts into which a column is visually divided is called a *bar*. (An elementary bar chart, which we shall not illustrate, is the case where the columns are rectangles—usually open or filled—and each column is a single bar.)

Figure 5 shows two nine-point eight-bar styles of column display (using single-line bars of somewhat

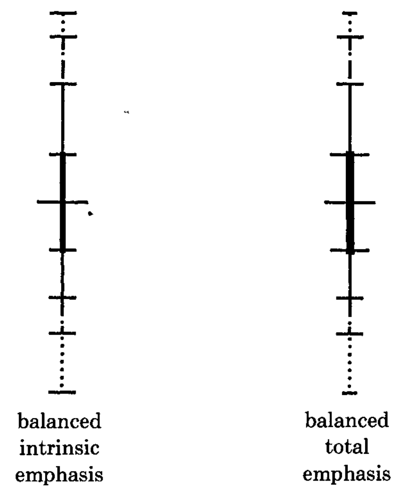


FIG. 5. *Two versions of a nine-point, eight-bar column; one with balanced intrinsic emphasis, the other with balanced total emphasis.*

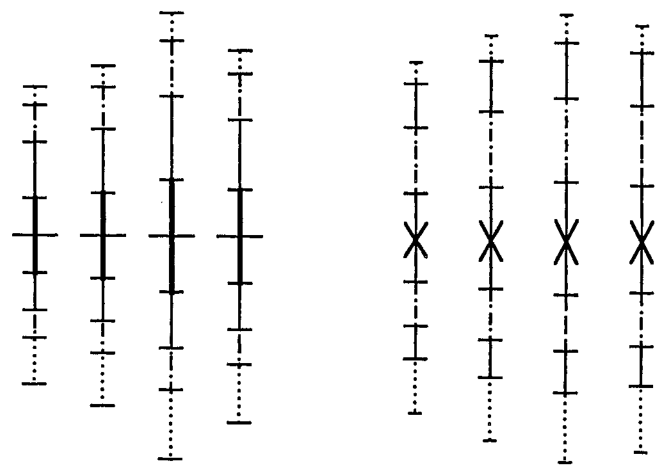


FIG. 6. *Four eight-bar columns shown in two ways: left as in Figure 5 (smoothly balanced emphasis), right with rippling emphasis (easier to sweep).*

balanced total emphasis), on the left with equal intrinsic emphasis on the nine points and on the right with intrinsic emphasis chosen to balance the situational emphasis quite smoothly. (Imaginary data!) On the right, one's eye is not appreciably more attracted to any one of the nine points rather than another. So long as we need to look at only one or two such nine-point eight-bar columns together this quietude of emphasis is probably a good thing.

When we need to look at a number of columns, however, perhaps with individual bars (rather than their separating points) as targets, we are too likely to slide from one point to another, or from one bar to another. As a result, sweeping is fearful, not easy. We need a different style.

Figure 6 shows an illustrative configuration of four columns (a) in the smooth-emphasis style just discussed and (b) a rippled-emphasis style obtained by rearranging the bar styles to alternate between more and less emphasis. Now each bar is much more clearly distinguished in feeling from those that abut on it.

Rippling emphasis—balanced in the large but alternating in the small—has restored easy sweep. The price could be thought of, at least by some, as increased distraction; after all, the alternation is quite noticeable and can be argued not to be "in the data"! But it *is* "in the data's use," where it plays an important role!

20. CLOSE

Visual display of quantitative things is in a major transition. Once it combined, on the one hand, a cottage industry stemming from cartography (and the mapping pen) and, on the other, a graphic art (as practiced by book designers). It seemed to emphasize plotting either the raw data or the results of the simplest analyses (like population per square mile). Tomorrow it can be quite different, with well understood (but not necessarily agreed-upon) principles, a facility of computer-driven implementations making diverse choices easy to a degree far beyond anything we have yet seen, and a partnership (two-way of course) of display with all forms of data analysis.

If we seek a landmark for the beginning of the transition, we cannot do better than turn to the work of Bertin (e.g., 1983, a translation of the second edition of a 1967 original). If we seek a marker for where we have almost reached today, I cannot do better than put down a short list of key ideas: phenomena, impact, busyness, distraction, prospecting, easy flow, partnership with analysis. If we seek a landmark for the end of the transition, we must wait for roughly a decade or two.

Let us go forward actively, trying to understand the roots of disagreements and the need for a diversity of concepts. In the next decades, the practice of visual display can be more conscious, both of what is being done in each individual display and of what principles are being recognized. Can be, should be—and will be!

ACKNOWLEDGMENTS

I owe major thanks to Nancy Clark, Ron Thisted, Ed Tufte, Pat Costigan-Eaves, Judith Tanur, Lyle V. Jones and Howard Wainer for helping me say more clearly what I meant. Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham), DAAL03-88-K-0045, and in part in connection with the Language of Data Project, sponsored by System Development Foundation.

Note: Letters used with years on John Tukey's publications correspond to bibliographies in all volumes of his collected papers.

REFERENCES

BERTIN, J. (1983). *Semiology of Graphics.* Univ. Wisconsin Press, Madison. (Translation of the *Semiologie Graphique*, 2nd ed., Mouton and Gauthier-Villars, Paris, 1973.)

HOAGLIN, D. C. and TUKEY, J. W. (1985n). Checking the shape of discrete distributions. In *Exploring Data Tables, Trends, and Shapes* (D. C. Hoaglin, F. Mosteller and J. W. Tukey, eds.) 345–416. Wiley, New York.

TUFTE, E. R. (1983). *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, Conn.

TUKEY, J. W. (1977a). *Exploratory Data Analysis.* Addison-Wesley, Reading, Mass.

TUKEY, J. W. and TUKEY, P. A. (1990). Dynamic graphical display of data can be "low tech"—and projected! Unpublished manuscript.