9 *Aesthetics and Technique in Data Graphical Design*

Along with the amazing graphic of the French losses in the Russian invasion, Minard includes a second "Carte Figurative." It portrays Hannibal's fading elephant campaign in Spain, Gaul, and Northern Italy. Minard uses a light transparent color for flow-lines, allowing the underlying type to show through. This refined use of color to depict more information contrasts with the garish tones too often seen in modern graphics.

Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845–1869*, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

What makes for such graphical elegance? What accounts for the quality of Minard's graphics, of those of Playfair and Marey, and of some recent work, such as the new view of the galaxies? Good design has two key elements:

> Graphical elegance is often found in simplicity of design and complexity of data.

Visually attractive graphics also gather power from content and interpretations beyond the immediate display of some numbers. The best graphics are about the useful and important, about life and death, about the universe. Beautiful graphics do not traffic with the trivial.

On rare occasions graphical architecture combines with the data content to yield a uniquely spectacular graphic. Such performances can be described and admired but there are no easy compositional principles on how to create that one wonderful graphic in millions. As Barnett Newman once said, "Aesthetics is for the artist like ornithology is for the birds."

What can be suggested, though, are some guides for enhancing the visual quality of routine, workaday designs. Attractive displays of statistical information

- have a properly chosen format and design
- use words, numbers, and drawing together
- reflect a balance, a proportion, a sense of relevant scale
- display an accessible complexity of detail
- often have a narrative quality, a story to tell about the data
- are drawn in a professional manner, with the technical details of production done with care
- avoid content-free decoration, including chartjunk.

The Choice of Design: Sentences, Text-Tables, Tables, Semi-Graphics, and Graphics

The substantive content, extensiveness of labels, and volume and ordering of data all help determine the choice of method for the display of quantitative materials. The basic structures for showing data are the sentence, the table, and the graphic. Often two or three of these devices should be combined.

The conventional sentence is a poor way to show more than two numbers because it prevents comparisons within the data. The linearly organized flow of words, folded over at arbitrary points (decided not by content but by the happenstance of column width), offers less than one effective dimension for organizing the data. Instead of:

> Nearly 53 percent of the type A group did something or other compared to 46 percent of B and slightly more than 57 percent of C.

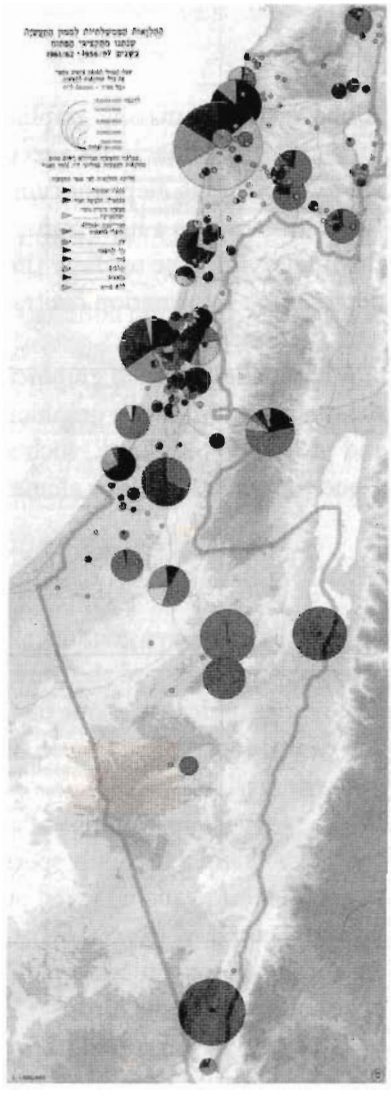
Arrange the type to facilitate comparisons, as in this *text-table*:

> The three groups differed in how they did something or other:
>
> | | |
> |---|---|
> | Group A | 53% |
> | Group B | 46% |
> | Group C | 57% |

There are nearly always better sequences than alphabetical—for example, ordering by content or by data values:

Group B	46%
Group A	53%
Group C	57%

Tables are clearly the best way to show exact numerical values, although the entries can also be arranged in semi-graphical form. Tables are preferable to graphics for many small data sets.[1] A table is nearly always better than a dumb pie chart; the only worse design than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between pies, as in this heavily encoded example from an atlas. Given their low data-density and failure to order numbers along a visual dimension, pie charts should never be used.[2]



Department of Surveys, Ministry of Labour, *Atlas of Israel* (Jerusalem, 1956–), vol. 8, p. 8.

[1] On the design of tables, see A.S.C. Ehrenberg, "Rudiments of Numeracy," *Journal of the Royal Statistical Society*, A, 140 (1977), 277–297.

[2] This point is made decisively in Jacques Bertin, *Graphics and Graphic Information Processing* (Berlin, 1981). Bertin describes multiple pie charts as "completely useless" (p. 111).

Tables also work well when the data presentation requires many localized comparisons. In this 410-number table that I designed for the *New York Times* to show how different people voted in presidential elections in the United States, comparisons between the elections of 1980 and 1976 are read across each line; within-election analysis is conducted by reading downward in the clusters of three to seven lines. The horizontal rules divide the data into topical paragraphs; the rows are ordered so as to tell an ordered story about the elections. This type of elaborate table, a *supertable*, is likely to attract and intrigue readers through its organized, sequential detail and reference-like quality. One supertable is far better than a hundred little bar charts.

New York Times, November 9, 1980, p. A–28.

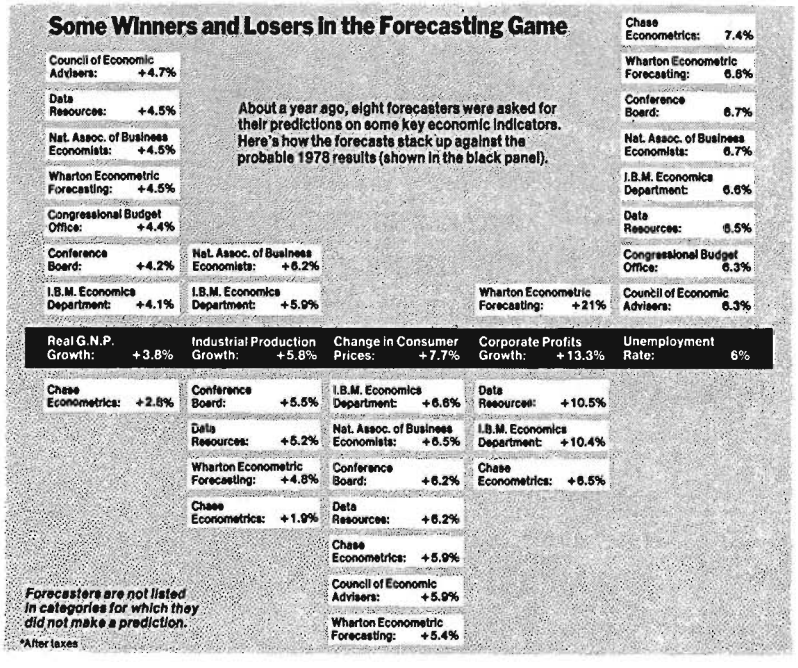
How Different Groups Voted for President

Based on 12,782 interviews with voters at their polling places. Shown is how each group divided its vote for President and, in parentheses, the percentage of the electorate belonging to each group.

	CARTER	REAGAN	ANDERSON	CARTER-FORD in 1976
Democrats (43%)	66	26	6	77 - 22
Independents (23%)	30	54	12	43 - 54
Republicans (28%)	11	84	4	9 - 90
Liberals (17%)	57	27	11	70 - 26
Moderates (46%)	42	48	8	51 - 48
Conservatives (28%)	23	71	4	29 - 70
Liberal Democrats (9%)	70	14	13	86 - 12
Moderate Democrats (22%)	66	28	6	77 - 22
Conservative Democrats (8%)	53	41	4	64 - 35
Politically active Democrats (3%)	72	19	8	—
Democrats favoring Kennedy in primaries (13%)	66	24	8	—
Liberal Independents (4%)	50	29	15	64 - 29
Moderate Independents (12%)	31	53	13	45 - 53
Conservative Independents (7%)	22	69	6	26 - 72
Liberal Republicans (2%)	25	66	9	17 - 82
Moderate Republicans (11%)	13	81	5	11 - 88
Conservative Republicans (12%)	6	91	2	6 - 93
Politically active Republicans (2%)	5	89	6	—
East (32%)	43	47	8	51 - 47
South (27%)	44	51	3	54 - 45
Midwest (20%)	41	51	6	48 - 50
West (11%)	35	52	10	46 - 51
Blacks (10%)	82	14	3	82 - 16
Hispanics (2%)	54	36	7	75 - 24
Whites(88%)	36	55	8	47 - 52
Female (49%)	45	46	7	50 - 48
Male (51%)	37	54	7	50 - 48
Female, favors equal rights amendment (22%)	54	32	11	—
Female, opposes equal rights amendment (15%)	29	66	4	—
Catholic (25%)	40	51	7	54 - 44
Jewish (5%)	45	39	14	64 - 34
Protestant (46%)	37	56	6	44 - 55
Born-again white Protestant (17%)	34	61	4	—
18 - 21 years old (6%)	44	43	11	48 - 50
22 - 29 years old (17%)	43	43	11	51 - 46
30 - 44 years old (31%)	37	54	7	49 - 49
45 - 59 years old (23%)	39	55	6	47 - 52
60 years or older (18%)	40	54	4	47 - 52
Family income				
Less than $10,000 (13%)	50	41	6	58 - 40
$10,000 - $14,999 (14%)	47	42	8	55 - 43
$15,000 - $24,999 (30%)	38	53	7	48 - 50
$25,000 - $50,000 (24%)	32	58	8	36 - 62
Over $50,000 (5%)	25	65	8	—
Professional or manager (40%)	33	56	9	41 - 57
Clerical, sales or other white-collar (11%)	42	48	8	46 - 53
Blue-collar worker (17%)	46	47	5	57 - 41
Agriculture (3%)	29	66	3	—
Looking for work (3%)	55	35	7	65 - 34
Education				
High school or less (39%)	46	48	4	57 - 43
Some college (28%)	35	55	8	51 - 49
College graduate (27%)	35	51	11	45 - 55
Labor union household (26%)	47	44	7	59 - 39
No member of household in union (62%)	35	55	8	43 - 55
Family finances				
Better off than a year ago (16%)	53	37	8	30 - 70
Same (40%)	46	46	7	51 - 49
Worse off than a year ago (34%)	25	64	8	77 - 23
Family finances and political party				
Democrats, better off than a year ago (7%)	77	16	6	69 - 31
Democrats, worse off than a year ago (13%)	47	39	10	94 - 6
Independents, better off (3%)	45	36	12	—
Independents, worse off (9%)	21	65	11	—
Republicans, better off (4%)	18	77	5	3 - 97
Republicans, worse off (11%)	6	89	4	24 - 76
More important problem				
Unemployment (39%)	51	40	7	75 - 25
Inflation (44%)	30	60	9	35 - 65
Feel that U.S. should be more forceful in dealing with Soviet Union even if it would increase the risk of war (54%)	28	64	6	—
Disagree (31%)	56	32	10	—
Favor equal rights amendment (46%)	49	38	11	—
Oppose equal rights amendment (35%)	26	68	4	—
When decided about choice				
Knew all along (41%)	47	50	2	44 - 55
During the primaries (13%)	30	60	8	57 - 42
During conventions (8%)	36	55	7	51 - 48
Since Labor Day (8%)	30	54	13	49 - 49
In week before election (23%)	38	46	13	49 - 47

Source: 1976 and 1980 election day surveys by The New York Times/CBS News Poll and 1976 election day survey by NBC News.

For sets of highly labeled numbers, a wordy data graphic—coming close to straight text—works well. This table of numbers is nicely organized into a graphic:



New York Times, January 2, 1979, p. D–3.

Making Complexity Accessible: Combining Words, Numbers, and Pictures

Explanations that give access to the richness of the data make graphics more attractive to the viewer. Words and pictures are sometimes jurisdictional enemies, as artists feud with writers for scarce space. An unfortunate legacy of these craft-union differences is the artificial separation of words and pictures; a few style sheets even forbid printing on graphics. What has gone wrong is that the techniques of production instead of the information conveyed have been given precedence.

Words and pictures belong together. Viewers need the help that words can provide. Words on graphics are data-ink, making effective use of the space freed up by erasing redundant and non-data-ink. It is nearly always helpful to write little messages on the plotting field to explain the data, to label outliers and interesting data points, to write equations and sometimes tables on the graphic itself, and to integrate the caption and legend into the design so that the eye is not required to dart back and forth between textual material and the graphic. (The size of type on and around graphics

can be quite small, since the phrases and sentences are usually not too long—and therefore the small type will not fatigue viewers the way it does in lengthy texts.)

The principle of *data/text integration* is

> Data graphics are paragraphs about data and should be treated as such.

Words, graphics, and tables are different mechanisms with but a single purpose—the presentation of information. Why should the flow of information be broken up into different places on the page because the information is packaged one way or another? Sometimes it may be useful to have multiple story-lines or multiple levels of presentation, but that should be a deliberate design judgment, not something decided by conventional production requirements. Imagine if graphics were replaced by paragraphs of words and those paragraphs scattered over the pages out of sequence with the rest of the text—that is how graphical and tabular information is now treated in the layout of many published pages, particularly in scientific journals and professional books.

Tables and graphics should be run into the text whenever possible, avoiding the clumsy and diverting segregation of "See Fig. 2," (figures all too often located on the back of the adjacent page).[3] If a display is discussed in various parts of the text, it might well be printed afresh near each reference to it, perhaps in reduced size in later showings. The principle of text/graphic/table integration also suggests that the same typeface be used for text and graphic and, further, that ruled lines separating different types of information be avoided. Albert Biderman notes that illustrations were once well-integrated with text in scientific manuscripts, such as those of Newton and Leonardo da Vinci, but that statistical graphics became segregated from text and table as printing technology developed:

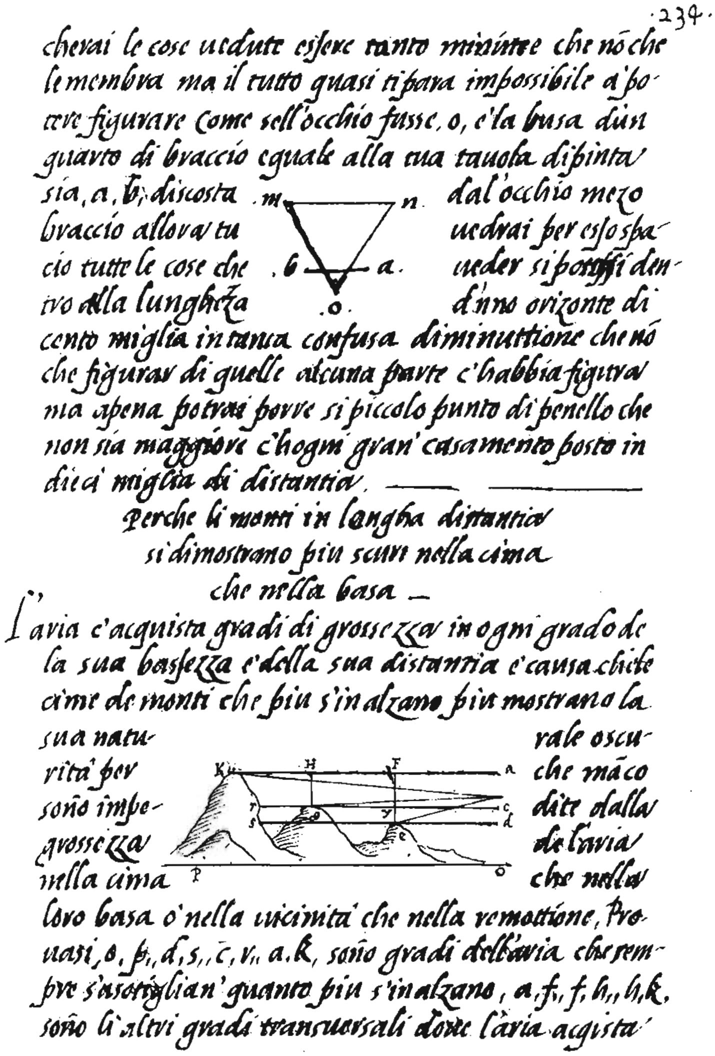
> The evolution of graphic methods as an element of the scientific enterprise has been handicapped by their adjunctive, segregated, and marginal position. The exigencies of typography that moved graphics to a segregated position in the printed work have in the past contributed to their intellectual segregation and marginality as well. There was a corresponding organizational segregation, with decisions on graphics often passing out of the hands of the original analyst and communicator into those of graphic specialists—the commercial artists and designers of graphic departments and audio-visual aids shops, for example, whose predilections and skills are usually more those of cosmeticians and merchandisers than of scientific analysts and communicators.[4]

[3] "Fig.," often used to refer to graphics, is an ugly abbreviation and is not worth the two spaces saved.

[4] Albert D. Biderman, "The Graph as a Victim of Adverse Discrimination and Segregation," *Information Design Journal*, 1 (1980), 238.

Page after page of Leonardo's manuscripts have a gentle but thorough integration of text and figure, a quality rarely seen in modern work:

Leonardo da Vinci, *Treatise on Painting* [*Codex Urbinas Latinus 1270*], vol. 2, facsimile (Princeton, 1956), p. 234, paragraph 827.


·234·
chevai le cose uedute essere tanto minime che nō che
le membra ma il tutto quasi ti parra impossibile a po
tere figurare come sell'occhio fusse, o, e la busa d'un
quarto di braccio equale alla sua tauola dipinta
sia, a, b, discosta dal occhio mezo
braccio allora tu uedrai per esso spa
cio tutte le cose che ueder si possino den
tro alla lungheza d'uno orizonte di
cento miglia in tanta confusa diminuttione che nō
che figurar di quelle alcuna parte c'habbia figura
ma a pena potrai porre si piccolo punto di penello che
non sia maggiore c'hogni gran casamento posto in
dieci miglia di distantia.
Perche li monti in longha distantia
si dimostrano piu scuri nella cima
che nella basa.
L'aria c'acquista gradi di grossezza in ogni grado de
la sua bassezza e della sua distantia e causa che le
cime de monti che piu s'inalzano piu mostrano la
sua natu- rale oscu-
rità per che mā co
sono impe- dite dalla
grossezza de l'aria
nella cima che nella
loro basa o nella uicinità che nella remottione, Pro-
uasi, o, p, d, s, c, r, a, k, sono gradi dell'aria che sem-
pre s'assottiglian' quanto piu s'inalzano, a, f, f, h, h, k,
sono li altri gradi transuersali dove l'aria acquista

Finally, a caveat: the use of words and pictures together requires a special sensitivity to the purpose of the design—in particular, whether the graphic is primarily for communication and illustration of a settled finding or, in contrast, for the exploration of a data set. Words on and around graphics are highly effective—sometimes all too effective—in telling viewers how to allocate their attention to the various parts of the data display.[5] Thus, for graphics in exploratory data analysis, words should tell the viewer *how* to read the design (if it is a technically complex arrangement) and not *what* to read in terms of content.

[5] Experiments in visual perception indicate that word instructions substantially determine eye movements in viewing pictures. See John D. Gould, "Looking at Pictures," in Richard A. Monty and John W. Senders, eds., *Eye Movements and Psychological Processes* (Hillsdale, N.J., 1976), 323–343.

Accessible Complexity: The Friendly Data Graphic

An occasional data graphic displays such care in design that it is particularly accessible and open to the eye, as if the designer had the viewer in mind at every turn while constructing the graphic. This is the *friendly data graphic.*

There are many specific differences between friendly and unfriendly graphics:

Friendly	Unfriendly
words are spelled out, mysterious and elaborate encoding avoided	abbreviations abound, requiring the viewer to sort through text to decode abbreviations
words run from left to right, the usual direction for reading occidental languages	words run vertically, particularly along the Y-axis; words run in several different directions
little messages help explain data	graphic is cryptic, requires repeated references to scattered text
elaborately encoded shadings, cross-hatching, and colors are avoided; instead, labels are placed on the graphic itself; no legend is required	obscure codings require going back and forth between legend and graphic
graphic attracts viewer, provokes curiosity	graphic is repellent, filled with chartjunk
colors, if used, are chosen so that the color-deficient and color-blind (5 to 10 percent of viewers) can make sense of the graphic (blue can be distinguished from other colors by most color-deficient people)	design insensitive to color-deficient viewers; red and green used for essential contrasts
type is clear, precise, modest; lettering may be done by hand	type is clotted, overbearing
type is upper-and-lower case, with serifs	type is all capitals, sans serif

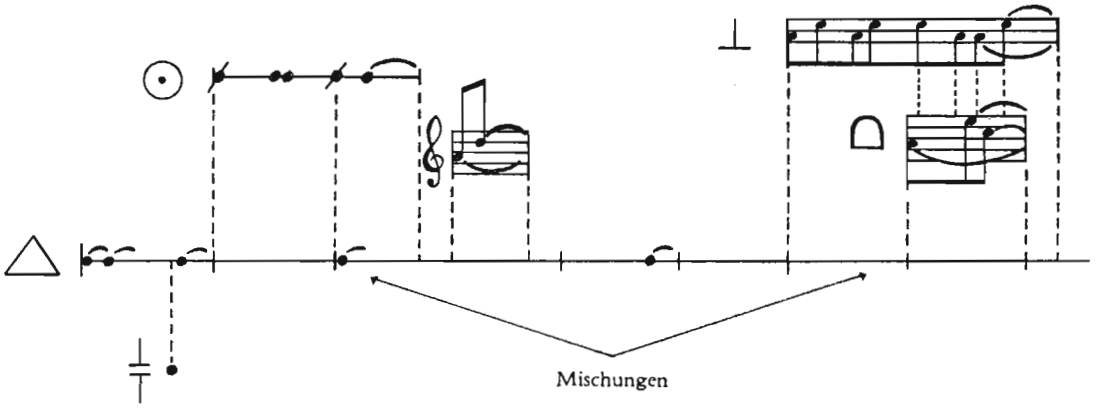
With regard to typography, Josef Albers writes:

> The concept that "the simpler the form of a letter the simpler its reading" was an obsession of beginning constructivism. It became something like a dogma, and is still followed by "modernistic" typographers. . . . Ophthalmology has disclosed that the more the letters are differentiated from each other, the easier is the reading. Without going into comparisons and details, it should be realized that words consisting of only capital letters present the most difficult reading—because of their equal height, equal volume, and, with most, their equal width. When comparing serif letters with sans-serif, the latter provide an uneasy reading. The fashionable preference for sans-serif in text shows neither historical nor practical competence.[6]

[6]Josef Albers, *Interaction of Color* (New Haven, 1963, revised edition 1975), p. 4.

Proportion and Scale: Line Weight and Lettering

Graphical elements look better together when their relative proportions are in balance. An integrated quality, an appropriate visual linkage between the various elements, results. This musical score of Karlheinz Stockhausen exhibits such a visual balance:



Karlheinz Stockhausen, *Texte*, vol. 2 (Cologne, 1964), p. 82, from the score of "Zyklus für einen Schlagzeuger."

In contrast, this next design is heavy handed, with nearly every element out of balance: the clotted ink, the poor style of lettering, the puffed-up display of a small data set, the coarse texture of the entire graphic, and the mismatch between drawing and surrounding text:

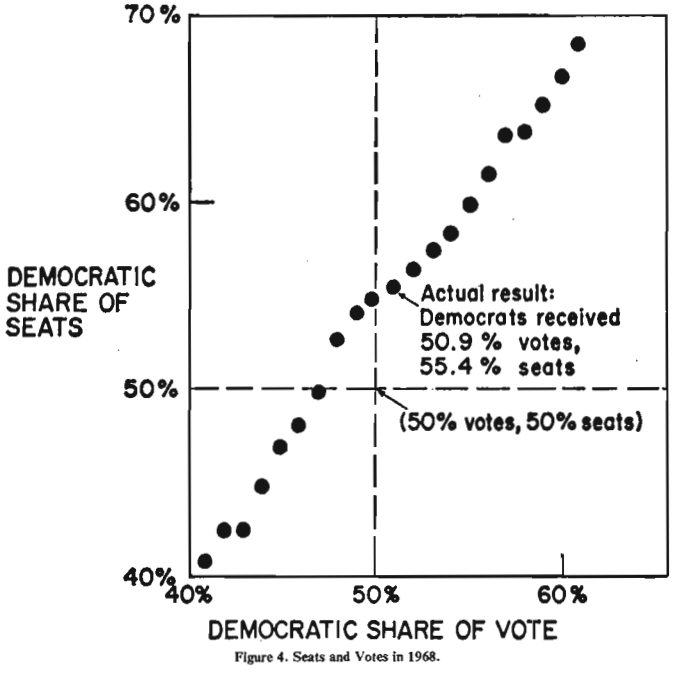
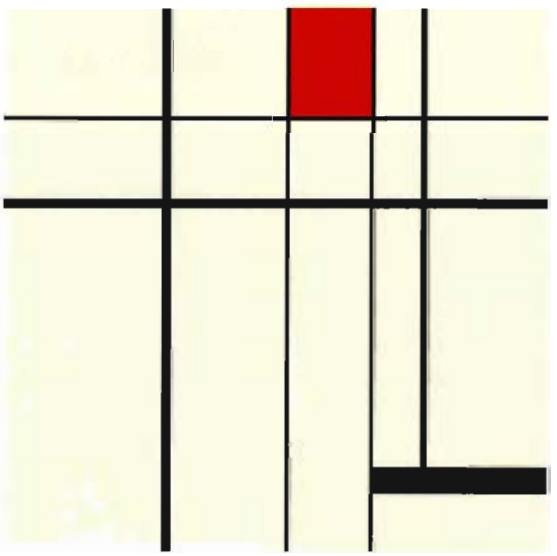


Figure 4. Seats and Votes in 1968.

Edward R. Tufte, "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review*, 67 (June 1973), 551.

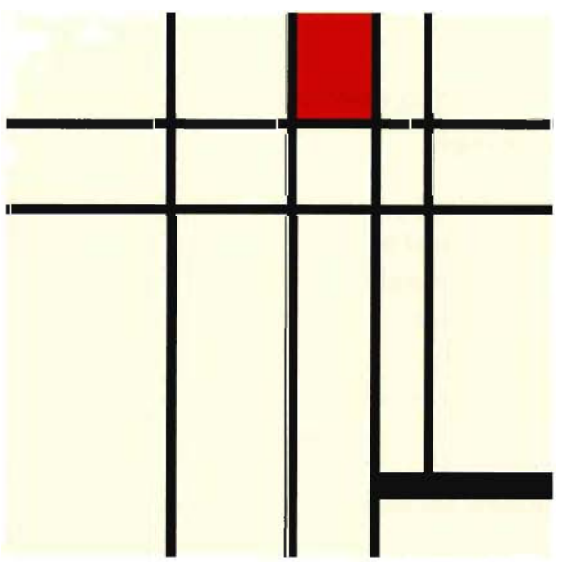
Lines in data graphics should be thin. One reason eighteenth- and nineteenth-century graphics look so good is that they were engraved on copper plates, with a characteristic hair-thin line. The drafting pens of twentieth-century mechanical drawing thickened linework, making it clumsy and unattractive.

An effective aesthetic device is the orthogonal intersection of lines of different weights:

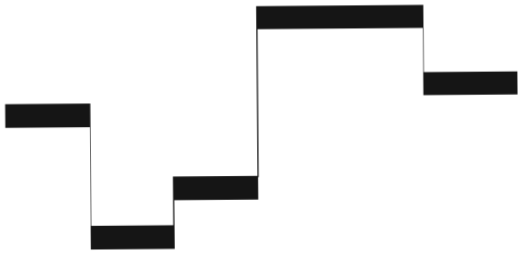


Poster for the exhibition "Mondrian and Neo-Plasticism in America," Yale University Art Gallery, October 18 to December 2, 1979. The original painting was done in 1941 by Diller; see Nancy J. Troy, *Mondrian and Neo-Plasticism in America* (New Haven, 1979), p. 28.

Nearly every intersection of the lines in this design (based on a painting by Burgoyne Diller) involves lines of differing weights, and it makes a difference, for the painting's character is diluted with lines of constant width:



Likewise, data graphics can be enhanced by the perpendicular intersections of lines of differing weights. The heavier line should be a data measure. In a time-series, for example:



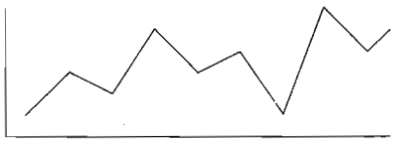
The contrast in line weight represents contrast in meaning. The greater meaning is given to the greater line weight; thus the data line should receive greater weight than the connecting verticals. The logic here is a restatement, in different language, of the principle of data-ink maximization.

Proportion and Scale: The Shape of Graphics

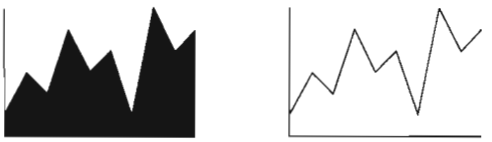
Graphics should tend toward the horizontal, greater in length than height:

Several lines of reasoning favor horizontal over vertical displays.

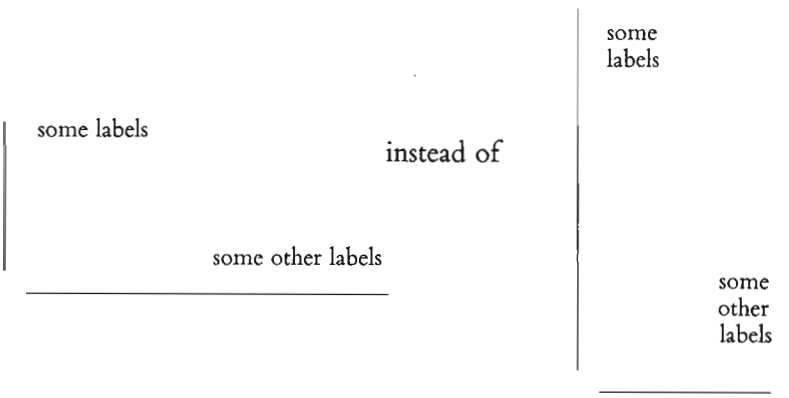
First, analogy to the horizon. Our eye is naturally practiced in detecting deviations from the horizon, and graphic design should take advantage of this fact. Horizontally stretched time-series are more accessible to the eye:



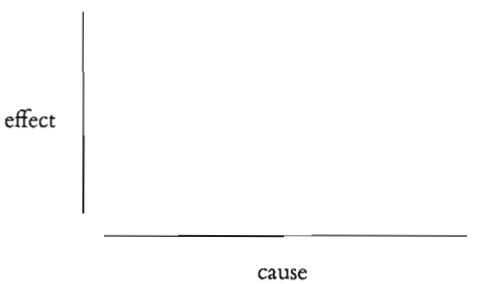
The analogy to the horizon also suggests that a shaded, high contrast display might occasionally be better than the floating snake. The shading should be calm, without moiré effects.



Second, ease of labeling. It is easier to write and to read words that read from left to right on a horizontally stretched plotting-field:



Third, emphasis on causal influence. Many graphics plot, in essence,



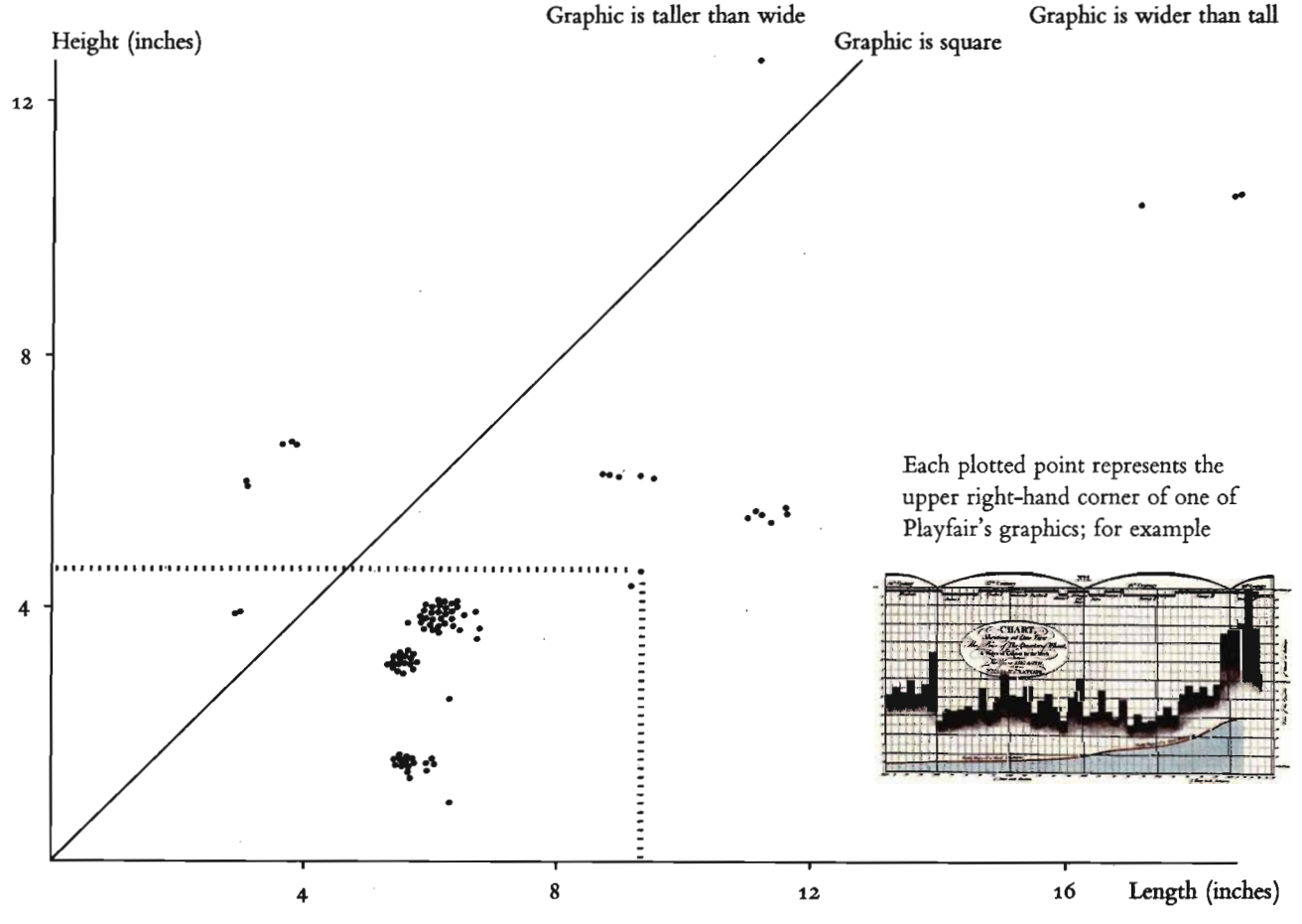
and a longer horizontal helps to elaborate the workings of the causal variable in more detail.

Fourth, Tukey's counsel.

> Most diagnostic plots involve either a more or less definite dependence that bobbles around a lot, or a point spatter. Such plots are rather more often better made *wider* than tall. Wider-than-tall shapes usually make it easier for the eye to follow from left to right.
>
> Perhaps the most general guidance we can offer is that smoothly-changing curves can stand being taller than wide, but a wiggly curve needs to be wider than tall. . . .[7]

[7]John W. Tukey, *Exploratory Data Analysis* (Reading, Mass., 1977), p. 129.

And, finally, Playfair's example. Of the 89 graphics in six different books by William Playfair, most (92 percent) are wider than tall. Several of the exceptions are his skyrocketing government debt displays. This plot shows the dimensions of each of those 89 graphics:

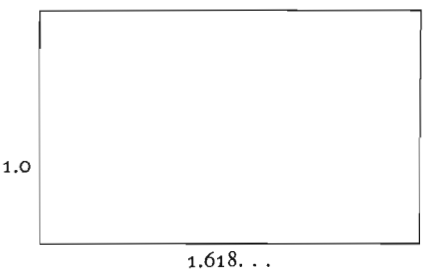


If graphics should tend toward the horizontal rather than the vertical, then how much so? A venerable (fifth-century B.C.) but dubious rule of aesthetic proportion is the Golden Section, a "divine division" of a line.[8] A length is divided such that the smaller is to the greater part as the greater is to the whole:

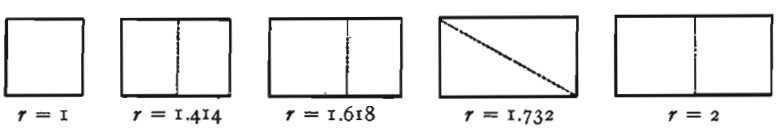
$$\frac{a}{b} = \frac{b}{a+b}$$

Solving the quadratic when $a = 1$ yields $b = \frac{\sqrt{5}+1}{2} = 1.618\ldots$.

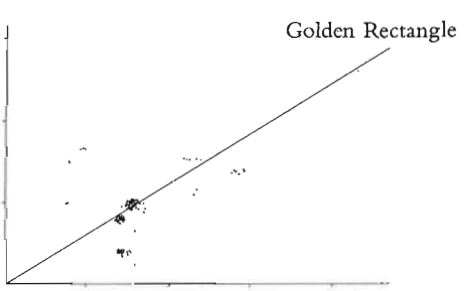
In turn the Golden Rectangle is



The nice geometry of the Golden Rectangle is not unique; Birkhoff points out that at least five other rectangles (including the square) have one simple mathematical property or another for which aesthetic claims might be made:[9]



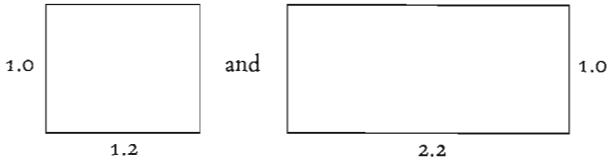
Playfair favored proportions between 1.4 and 1.8 in about two-thirds of his published graphics, with most of the exceptions moving more toward the horizontal than the golden prescription:



[8] The combination of geometry and mysticism surrounding the Golden Rectangle can be seen in Miloutine Borissavlièvitch, *The Golden Number and the Scientific Aesthetics of Architecture* (New York, 1958) and Tons Brunés, *The Secrets of Ancient Geometry* (Copenhagen, 1967), vols. 1 and 2.

[9] George D. Birkhoff, *Aesthetic Measure* (Cambridge, 1933), pp. 27–30.

Visual preferences for rectangular proportions have been studied by psychologists since 1860, but, even given the implausible assumption that such studies are relevant to graphic design, the findings are hardly decisive. A mild preference for proportions near to the Golden Rectangle is found among those taking part in the experiments, but the preferred height/length ratios also vary a great deal, ranging between

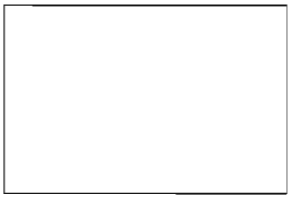


And, as is nearly always the case in experiments in graphical perception, viewer responses were found to be highly context-dependent.[10]

[10] I have relied on Leonard Zusne, *Visual Perception of Form* (New York, 1970), ch. 10, for a summary of the immense literature.

The conclusions:

- If the nature of the data suggests the shape of the graphic, follow that suggestion.
- Otherwise, move toward horizontal graphics about 50 percent wider than tall:



Epilogue: Designs for the Display of Information

Design is choice. The theory of the visual display of quantitative information consists of principles that generate design options and that guide choices among options. The principles should not be applied rigidly or in a peevish spirit; they are not logically or mathematically certain; and it is better to violate any principle than to place graceless or inelegant marks on paper. Most principles of design should be greeted with some skepticism, for word authority can dominate our vision, and we may come to see only through the lenses of word authority rather than with our own eyes.

What is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and the difficult—that is,

the revelation of the complex.