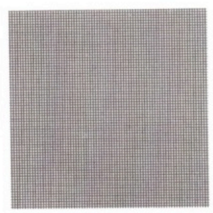
I want to reach that state of condensation of sensations which constitutes a picture.

Henri Matisse

8 *Data Density and Small Multiples*

Our eyes can make a remarkable number of distinctions within a small area. With the use of very light grid lines, it is easy to locate 625 points in one square inch or, equivalently, 100 points in one square centimeter.

Or consider how an 80 by 80 grid over a square inch—about 30 by 30 over a square centimeter—divides the space:[1]



25,281 distinctions

With the help of considerable redundancy and context, our eyes make fine distinctions of this sort all the time. Measurement instruments used in engineering, architectural, and machine work are engraved with scales of 20 increments to the centimeter and 50 to the inch. Or consider the reading of fine print. The type in the U.S. *Statistical Abstract* is set at 12 lines per vertical inch, with each line running at about 23 characters per inch for a maximum density of 276 characters per square inch. The actual density, given the white space, is in this case 185 characters per square inch or 28 per square centimeter.

[1] A square grid formed on each side by n parallel black and n–1 parallel white lines contains n^2 intersections of two black lines (corners of squares), $(n-1)^2$ intersections of two white lines (white squares), and $2n(n-1)$ intersections of a black and white line (sides of squares), for a total of $(2n-1)^2$ line intersections or distinct locations.

No. 1450. Steel Products—Net Shipments, by Market Classes: 1960 to 1978

[**In thousands of short tons.** Comprises carbon, alloy, and stainless steel. "N.e.c." means not elsewhere classified]

MARKET CLASS	1960	1965	1970	1973	1974	1975	1976	1977	1978
Total [1]	**71,149**	**92,666**	**90,798**	**111,430**	**109,472**	**79,957**	**89,447**	**91,147**	**97,935**
Steel for converting and processing	2,928	3,932	3,443	4,714	4,486	3,255	4,036	3,679	4,612
Independent forgers, n.e.c.	841	1,250	1,048	1,213	1,339	1,098	952	998	1,192
Industrial fasteners [2]	1,071	1,234	1,005	1,278	1,331	675	912	848	870
Steel service centers, distributors	11,125	14,813	16,025	20,383	20,400	12,700	14,615	15,346	17,333
Construction, incl. maintenance	9,664	11,836	8,913	10,731	11,360	8,119	7,508	7,553	9,612
Contractors' products	3,602	5,018	4,440	6,459	6,249	3,927	4,502	4,500	3,480
Automotive	14,610	20,123	14,475	23,217	18,928	15,214	21,351	21,490	21,253
Rail transportation	2,525	3,805	3,098	3,228	3,417	3,152	3,056	3,238	3,549
Freight cars, passenger cars, locomotives	1,763	2,875	2,005	1,997	2,097	1,794	1,428	1,709	2,188
Rails and all other [3]	762	930	1,093	1,231	1,320	1,358	1,628	1,529	1,361
Shipbuilding and marine equip.	622	1,051	859	1,019	1,339	1,413	969	869	845
Aircraft and aerospace	78	94	56	69	79	69	59	63	60
Oil and gas industries	1,759	1,936	3,550	3,405	4,210	4,171	2,653	3,650	4,140
Mining, quarrying, and lumbering	288	392	497	534	644	596	536	486	508
Agricultural, incl. machinery	1,003	1,483	1,126	1,772	1,859	1,429	1,784	1,743	1,805
Machinery, industrial equip., tools	3,958	5,873	5,169	6,351	6,440	5,173	5,180	5,566	5,992
Electrical equipment	2,078	2,985	2,694	3,348	3,242	2,173	2,671	2,639	2,811
Appliances, utensils, and cutlery	1,760	2,179	2,160	2,747	2,412	1,653	1,950	2,129	2,094
Other domestic commercial equip.	1,959	2,179	1,778	1,990	1,941	1,390	1,813	1,846	1,889
Containers, packaging, shipping	6,429	7,331	7,775	7,811	8,218	6,053	6,914	6,714	6,595
Cans and closures	4,976	5,867	6,239	6,070	6,349	4,859	5,290	5,173	4,950
Ordnance and other military	165	289	1,222	918	654	405	219	193	207
Exports (reporting companies only)	2,563	2,078	5,985	3,138	3,961	1,755	1,839	1,076	1,224

[1] Total includes nonclassified shipments, and, beginning 1970, data include estimates for a relatively small number of companies which report raw steel production but not shipments. [2] Bolts, nuts, rivets, and screws. [3] Includes railways, rapid transit systems, railroad rails, trackwork, and equipment.

U.S. Bureau of the Census, *Statistical Abstract of the United States: 1979* (Washington, D.C., 1979), p. 822.

Maps routinely present even finer detail. A cartographer writes that "the resolving power of the eye enables it to differentiate to 0.1 mm where provoked to do so. Clearly, therefore, conciseness is of the essence and high resolution graphics are a common denominator of cartography."[2] Distinctions at 0.1 mm mean 254 per inch.

[2]D. P. Bickmore, "The Relevance of Cartography," in John C. Davis and Michael J. McCullagh, eds., *Display and Analysis of Spatial Data* (London, 1975), p. 331.

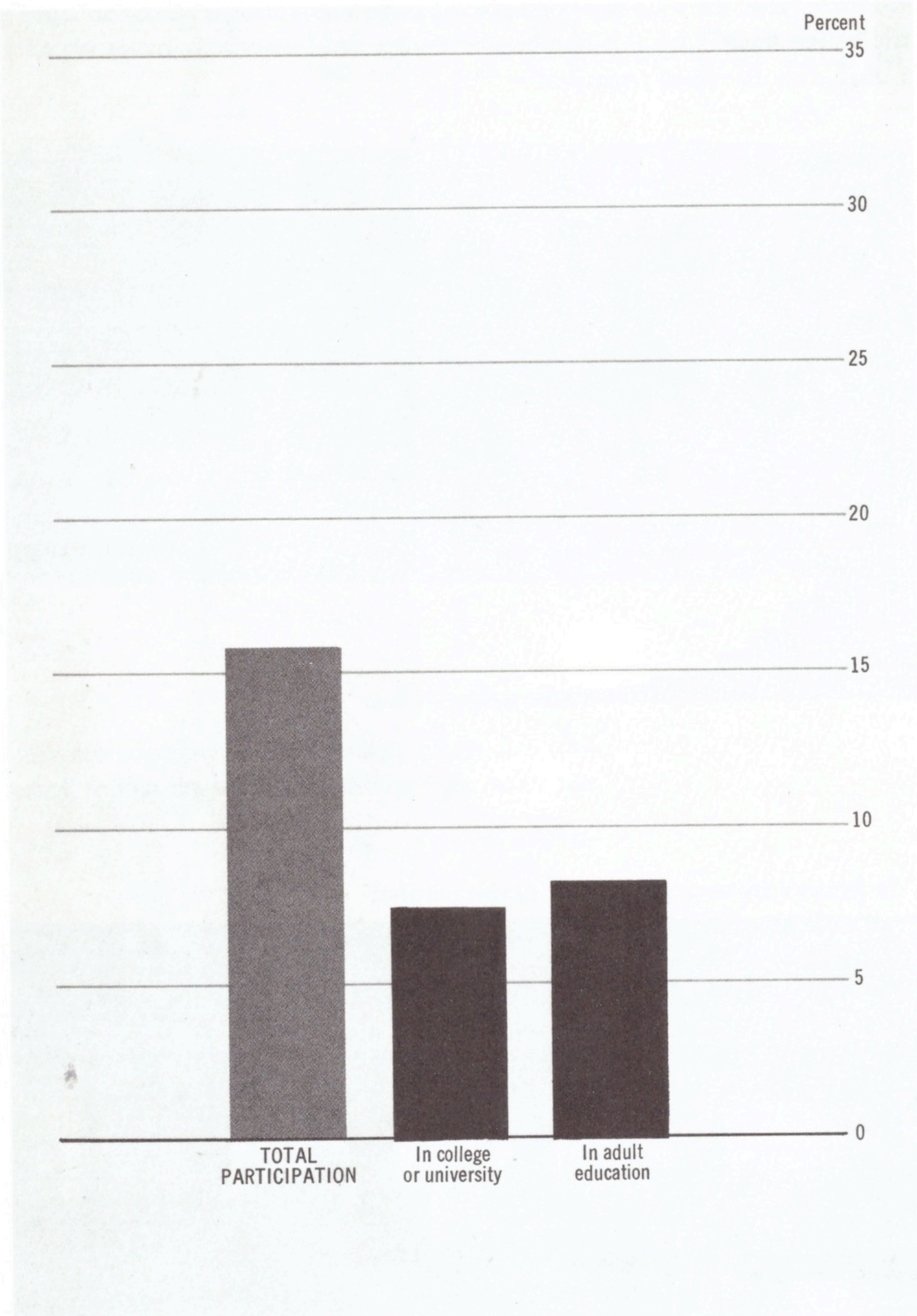
How many statistical graphics take advantage of the ability of the eye to detect large amounts of information in small spaces? And how much information should graphics show? Let us begin by considering an empirical measure of graphical performance, the data density.

Data Density in Graphical Practice

The numbers that go into a graphic can be organized into a data matrix of observations by variables. Taking into account the size of the graphic in relation to the amount of data displayed yields the *data density*:

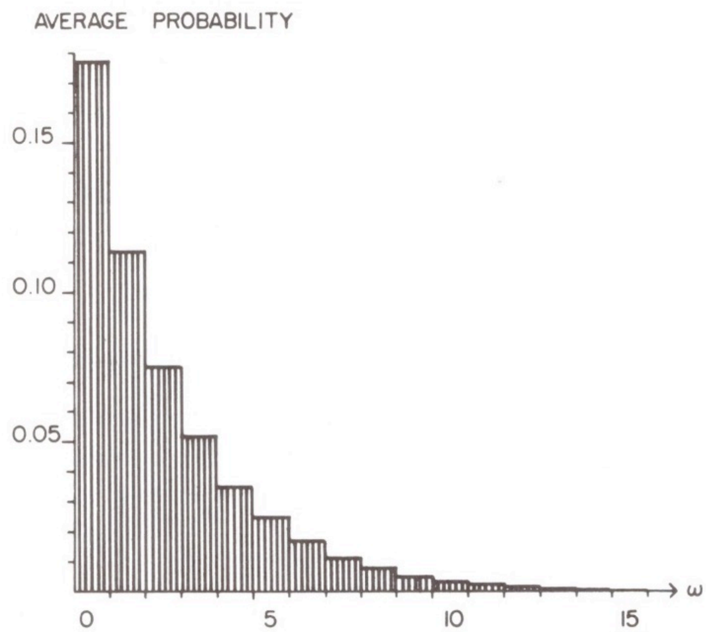
$$\text{data density of a graphic} = \frac{\text{number of entries in data matrix}}{\text{area of data graphic}}$$

Data matrices and data densities vary enormously in practice. At one extreme, this overwrought display (originally printed in five colors) presents a data matrix of four entries, the names and the numbers for the two bars on the right. The left bar is merely the total of the other two. The graph covers 26.5 square inches (171 square centimeters), resulting in a data density of .15 numbers per square inch (.02 numbers per square centimeter), which is thin indeed.

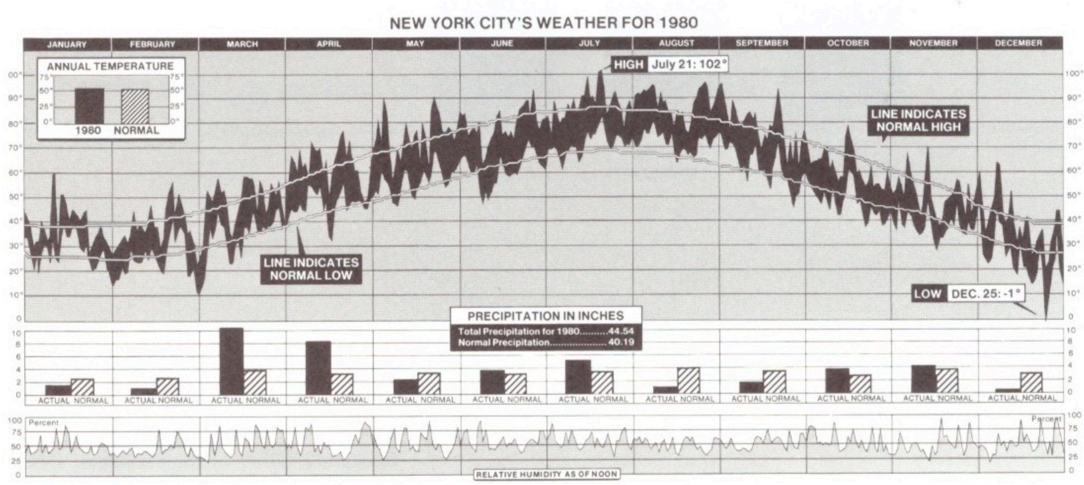


Executive Office of the President, Office of Management and Budget, *Social Indicators, 1973* (Washington, D.C., 1973), p. 86.

The exemplar from the JASA style sheet comes in at a lightweight 3.8 numbers per square inch (0.6 numbers per square centimeter) and a small data matrix of 32 entries:

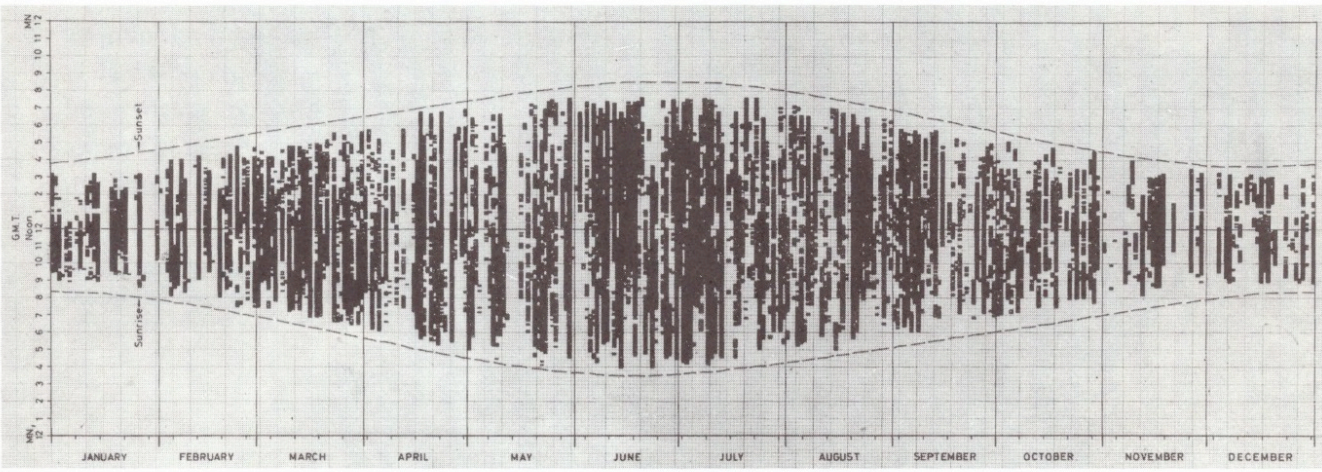


In contrast, the New York weather history, in this reduced version, does very well at 181 numbers per square inch (28 per square centimeter):

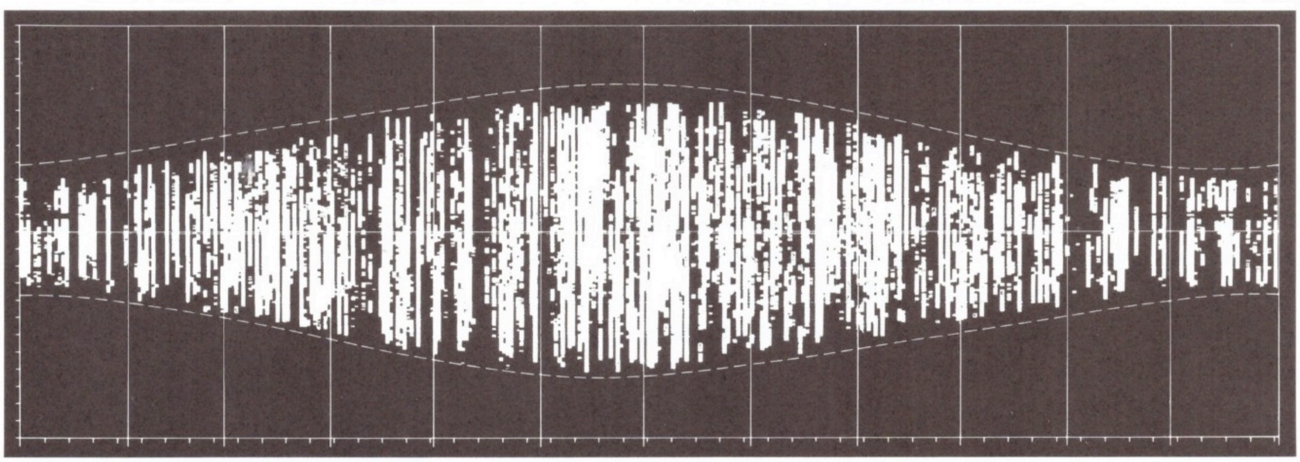


An annual sunshine record reports about 1,000 numbers per square inch (160 per square centimeter):

F. J. Monkhouse and H. R. Wilkinson, *Maps and Diagrams* (London, third edition, 1971), pp. 242–243.

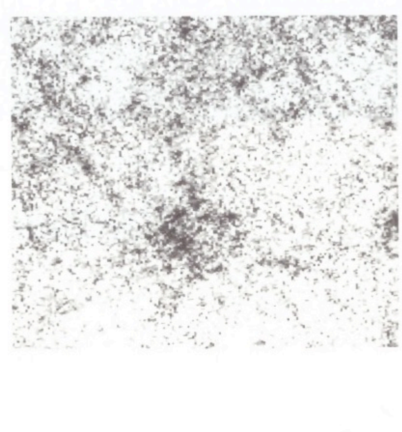


The visual metaphor corresponds appropriately to the data if the image is reversed, so that the light areas are the times when the sun shines:



Jacques Bertin, *Semiologie Graphique* (Paris, second edition, 1973), p. 152.

This map (27 square inches, 175 square centimeters) shows the location and boundaries of 30,000 communes of France. It would require at least 240,000 numbers to recreate the data of the map (30,000 latitudes, 30,000 longitudes, and perhaps six numbers describing the shape of each commune). Thus that data density is nearly 9,000 numbers per square inch, or 1,400 numbers per square centimeter.



The new map of the galaxies locates 2,275,328 encoded rectangles on a two-dimensional surface of 61 square inches (390 square centimeters). Each rectangle represents three numbers (two by its location, one by its shading), yielding a data density of 110,000 numbers per square inch or 17,000 numbers per square centimeter. That is the current record.

Data Density and the Size of the Data Matrix: Publication Practices

The table shows the data density and the size of the data matrix for graphics sampled from scientific and news publications. At least 20 graphics from each publication were examined.

The table records an enormous diversity of graphical performances both within and between publications. A few data-rich designs appear in nearly every publication. The opportunity is there but it is rarely exploited: the average published graphic is rather thin,

Data Density and Size of Data Matrix, Statistical Graphics in Selected Publications, Circa 1979–1980

	Data Density (Numbers per square inch)			Size of Data Matrix		
	median	minimum	maximum	median	minimum	maximum
Nature	48	3	362	177	15	3780
Journal of the Royal Statistical Society, B	27	4	115	200	10	1460
Science	21	5	44	109	26	316
Wall Street Journal	19	3	154	135	28	788
Fortune	18	5	31	96	42	156
The Times (London)	18	2	122	50	14	440
Journal of the American Statistical Association	17	4	167	150	46	1600
Asahi	13	2	113	29	15	472
New England Journal of Medicine	12	3	923	84	8	3600
The Economist	9	1	51	36	3	192
Le Monde	8	1	17	66	11	312
Psychological Bulletin	8	1	74	46	8	420
Journal of the American Medical Association	7	1	39	53	14	735
New York Times	7	1	13	35	6	580
Business Week	6	2	12	32	14	96
Newsweek	6	1	13	23	2	96
Annuaire Statistique de la France	6	1	25	96	12	540
Scientific American	5	1	69	46	14	652
Statistical Abstract of the United States	5	2	23	38	8	164
American Political Science Review	2	1	10	16	9	40
Pravda	0.2	0.1	1	5	4	20

based on about 50 numbers shown at the rate of 10 per square inch. Among the world's newspapers, the *Wall Street Journal*, *The Times* (London), and *Asahi* publish data-rich graphics, with data densities equal to those of the *Journal of the American Statistical Association*. Most of the American papers and magazines, along with *Pravda*, publish less data per graphic than the major papers of other industrialized countries.

Very few statistical graphics achieve the information display rates found in maps. Highly detailed maps portray 100,000 to 150,000 bits per square inch. For example, the average U.S. Geological Survey topographic quadrangle (measuring 17 by 23 inches) is estimated to contain over 100 million bits of information, or about 250,000 per square inch (40,000 per square centimeter).[3] Perhaps some day statistical graphics will perform as successfully as maps in carrying information.

[3] Morris M. Thompson, *Maps for America* (Washington, D.C., 1979), p. 187.

High-Information Graphics

Data graphics should often be based on large rather than small data matrices and have a high rather than low data density. More information is better than less information, especially when the marginal costs of handling and interpreting additional information are low, as they are for most graphics. The simple things belong in tables or in the text; graphics can give a sense of large and complex data sets that cannot be managed in any other way. If the graphic becomes overcrowded (although several thousand numbers represented may be just fine), a variety of data-reduction techniques—averaging, clustering, smoothing—can thin the numbers out before plotting.[4] Summary graphics can emerge from high-information displays, but there is nowhere to go if we begin with a low-information design.

Data-rich designs give a context and credibility to statistical evidence. Low-information designs are suspect: what is left out, what is hidden, why are we shown so little? High-density graphics help us to compare parts of the data by displaying much information within the view of the eye: we look at one page at a time and the more on the page, the more effective and comparative our eye can be.[5] The principle, then, is:

> Maximize data density and the size of the data matrix, within reason.

High-information graphics must be designed with special care. As the volume of data increases, data measures must shrink (smaller dots for scatters, thinner lines for busy time-series). The clutter of

[4] Paul A. Tukey and John W. Tukey, "Summarization; Smoothing; Supplemented Views," in Vic Barnett, ed., *Interpreting Multivariate Data* (Chichester, England, 1982), ch. 12; and William S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74 (1979), 829–836, are recent papers in the large literature.

[5] It is suggested in the analysis of x-ray films to "search a reduced image so that the whole display can be perceived on at least one occasion without large eye movement." Edward Llewellyn Thomas, "Advice to the Searcher or What Do We Tell Them?" in Richard A. Monty and John W. Senders, eds., *Eye Movements and Psychological Processes* (Hillsdale, N.J., 1976), p. 349.

chartjunk, non-data-ink, and redundant data-ink is even more costly than usual in data-rich designs.

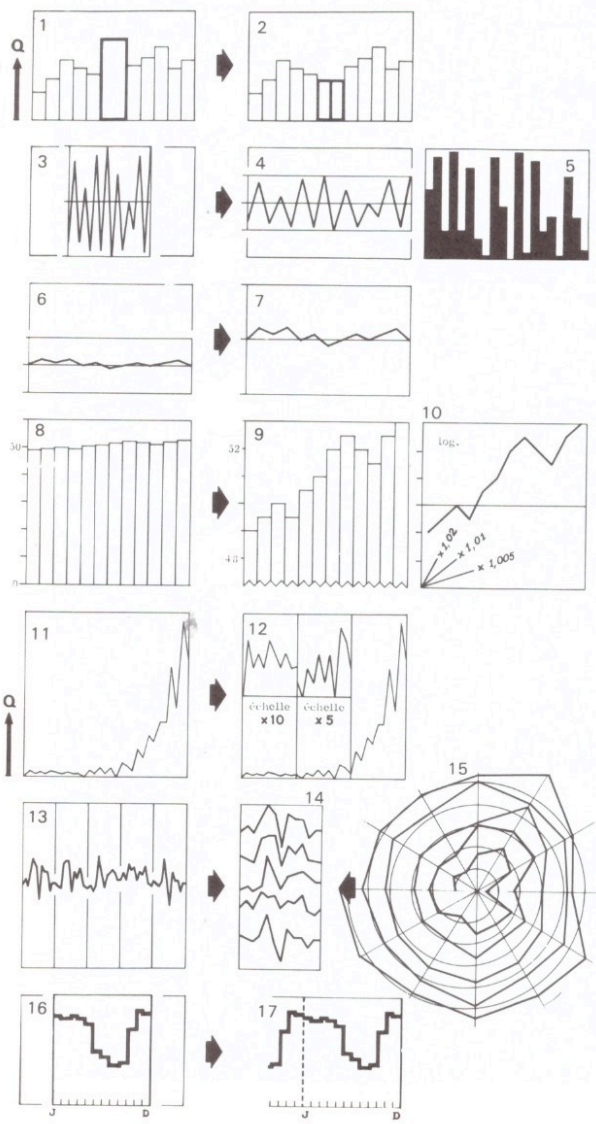
The way to increase data density other than by enlarging the data matrix is to reduce the area of a graphic. The Shrink Principle has wide application:

Graphics can be shrunk way down.

Many data graphics can be reduced in area to half their currently published size with virtually no loss in legibility and information. For example, Bertin's crisp and elegant line allows the display of 17 small-scale graphics on a single page along with extensive text. Repeated application of the Shrink Principle leads to a powerful and effective graphical design, the small multiple.

Jacques Bertin, *Semiologie Graphique* (Paris, second edition, 1973), p. 214.

PROBLEMES GRAPHIQUES POSES PAR LES CHRONIQUES



Un total sur deux cases (sur deux ans) doit être divisé par deux (1).
Un total pour six mois sera multiplié par deux dans des cases annuelles.

Courbes trop pointues, réduire l'échelle des Q; la sensibilité angulaire s'inscrit dans une zone moyenne autour de 70°.
Si la courbe n'est pas réductible (grandes et petites variations) employer les colonnes remplies (5).
Courbes trop plates : augmenter l'échelle des Q.

Variations très faibles par rapport au total.
Celui-ci perd de l'importance et le zéro peut être supprimé, à condition que le lecteur voit sa suppression (9). Le graphique peut être interprété comme une accélération si l'étude fine des variations est nécessaire (échelle logarithmique (10) (v. p. 240).

Très grande amplitude entre les valeurs extrêmes. Il faut admettre :
1°) Soit de ne pas percevoir les plus petites variations.
2°) Soit de ne s'intéresser qu'aux différences relatives (échelle logarithmique) sans connaître la quantité absolue.
3°) Soit admettre des périodes différentes dans la composante ordonnée et les traiter à des échelles différentes au-dessus de l'échelle commune (12).

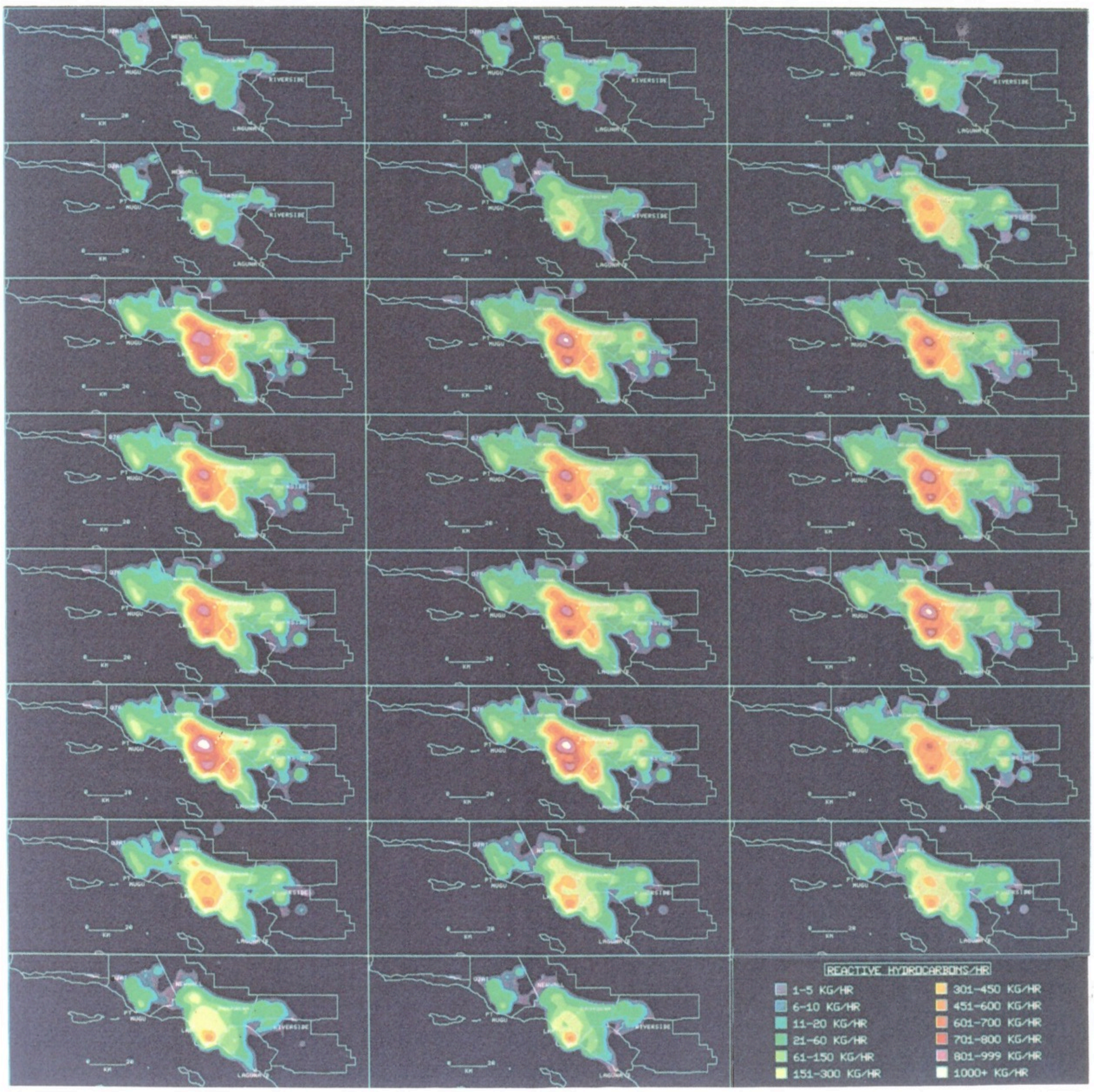
Cycles très marqués.
Si l'étude porte sur la comparaison des phases de chaque cycle, il est préférable de décomposer (13) de manière à superposer les cycles (14). La construction polaire peut être employée, de préférence dans une forme spirale (15) (ne pas commencer par un trop petit cercle); pour spectaculaire qu'elle soit, elle est moins efficace que la construction orthogonale.

Courbes annuelles de pluie ou de température. Un cycle possède deux phases (17), pourquoi n'en offrir qu'une à la perception du spectateur ? (16).

Small Multiples

Small multiples resemble the frames of a movie: a series of graphics, showing the same combination of variables, indexed by changes in another variable. Twenty-three hours of Los Angeles air pollution are organized into this display, based on a computer generated video tape. Shown is the hourly average distribution of reactive hydrocarbon emissions. The design remains constant through all the frames, so that attention is devoted entirely to shifts in the data:

From video tape by Gregory J. McRae, California Institute of Technology. The model is described in G. J. McRae, W. R. Goodin, and J. H. Seinfeld, "Development of a Second-Generation Mathematical Model for Urban Air Pollution. I. Model Formulation," *Atmospheric Environment*, 16 (1982), 679–696.



These grim small multiples show the distribution of occurrence of the cancer melanoma. The sites of 269 primary melanomas are recorded, along with the distribution between men and women. Note the data graphical arithmetic, similar to that of the multiwindow plot.

Arthur Wiskemann, "Zur Melanomentstehung durch chronische Lichteinwirkung," *Der Hautarzt*, 25 (1974), 21.

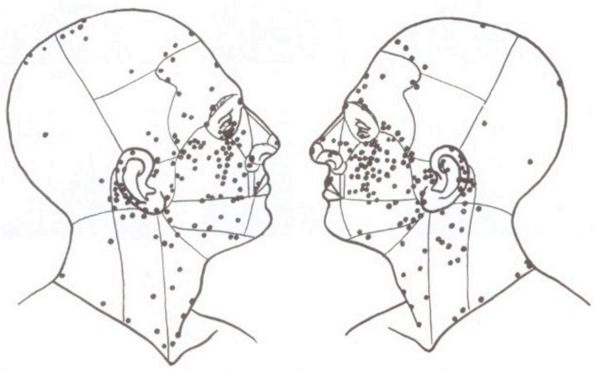


Abb. 1. Verteilung von 269 primären Melanomen auf Kopf und Hals

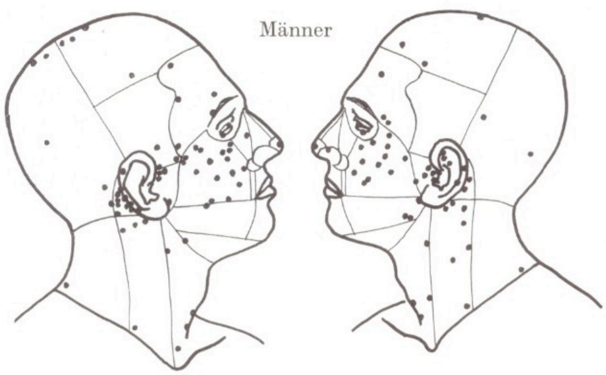


Abb. 2

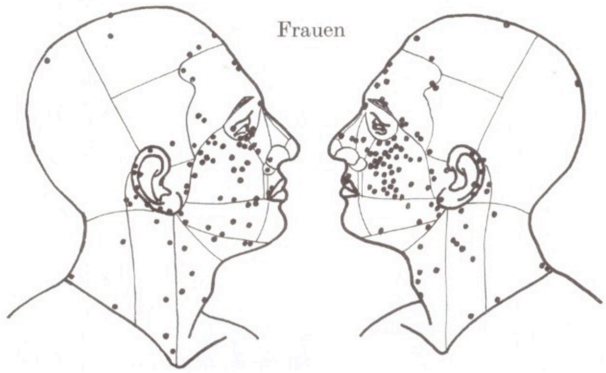
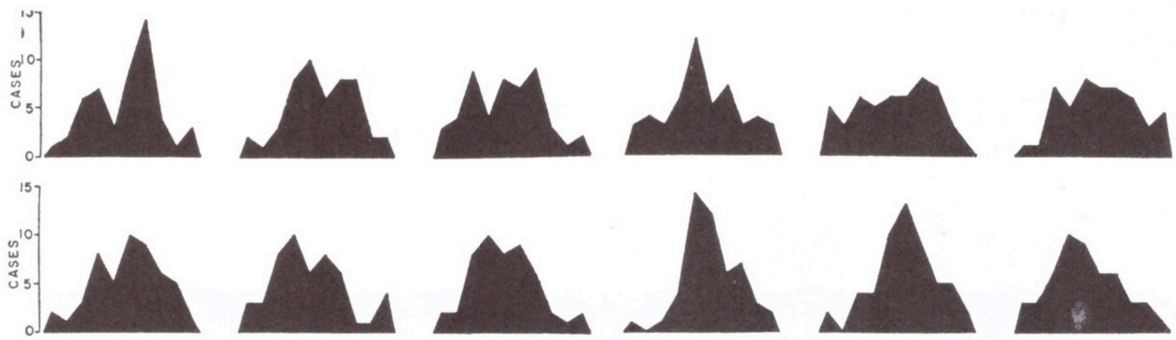


Abb. 3

Abb. 2 u. 3. Differenzierung der Melanomverteilung nach Geschlechtern

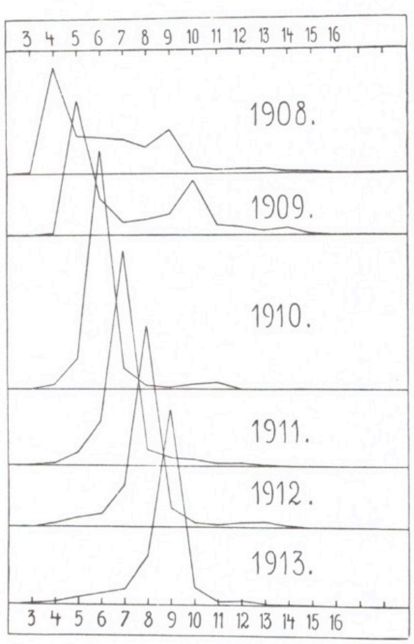
The effects of sampling errors are shown in these 12 distributions, each based on a sample of 50 random normal deviates:

Edmond A. Murphy, "One Cause? Many Causes? The Argument from the Bimodal Distribution," *Journal of Chronic Diseases*, 17 (1964), 309.



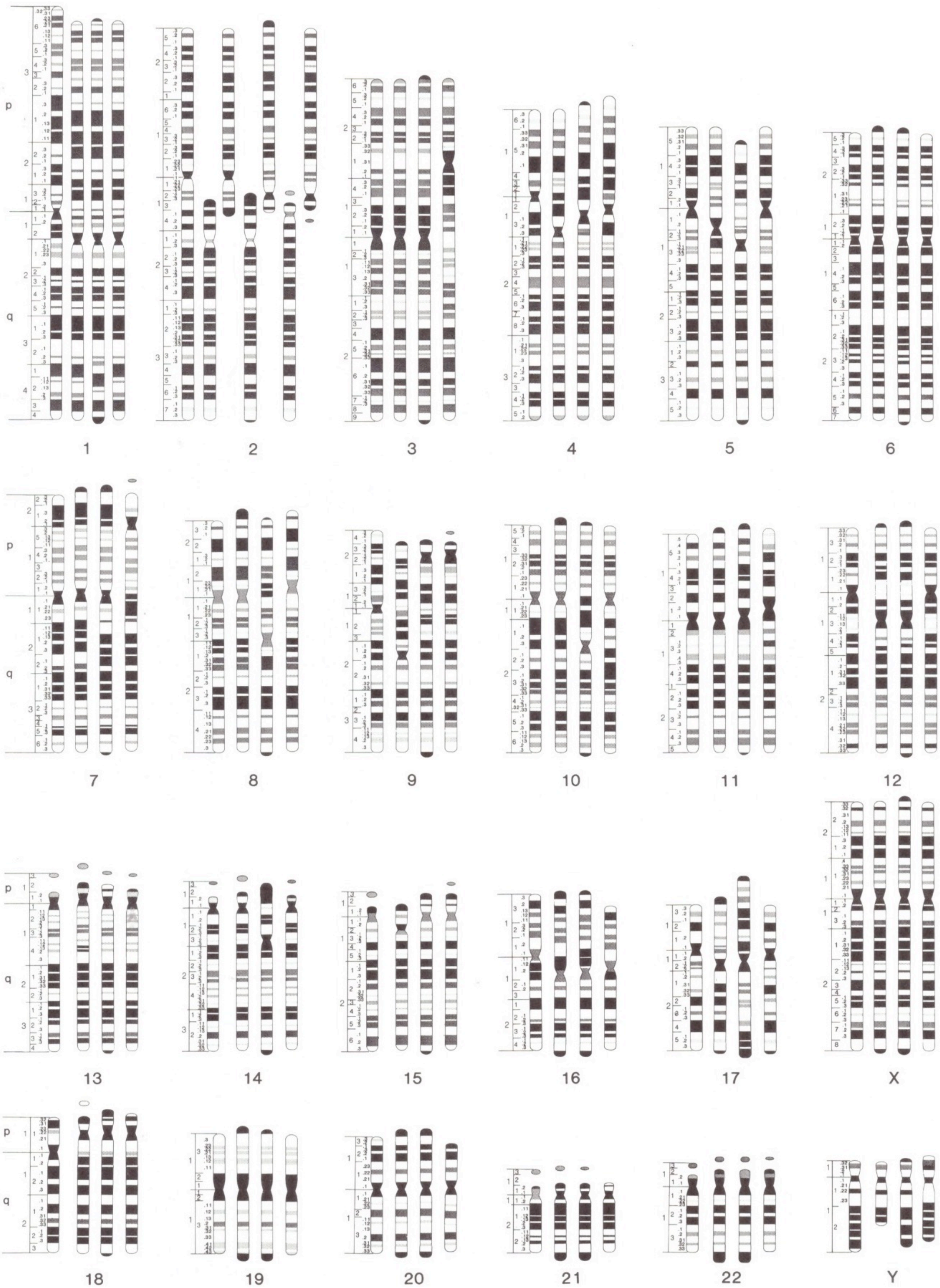
These six distributions show the age composition of herring catches each year from 1908 to 1913. A tremendous number of herring were spawned in 1904, and that class began to dominate the 1908 catch as four-year-olds, then the 1909 catch as five-year-olds, and so on:

Johan Hjort, "Fluctuations in the Great Fisheries of Northern Europe," *Rapports et Proces-Verbaux*, 20 (1914), in Susan Schlee, *The Edge of an Unfamiliar World* (New York, 1973), p. 226.



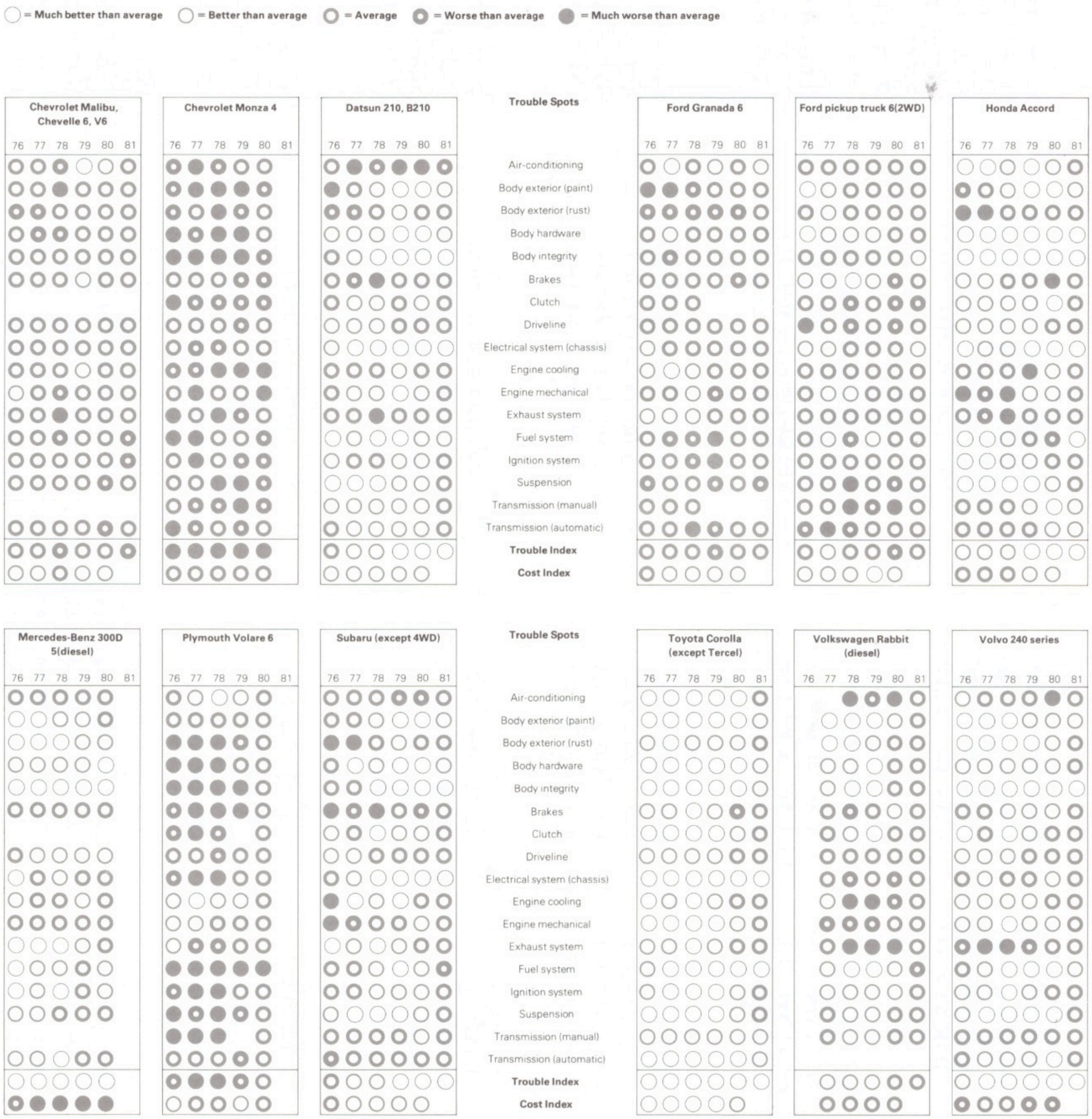
This next design compares a complex set of data: shown are the chromosomes of (from left to right) man, chimpanzee, gorilla, and orangutan. The similarities between humans and the great apes are to be noted.

Jorge J. Yunis and Om Prakash, "The Origin of Man: A Chromosomal Pictorial Legacy," *Science*, 215 (March 19, 1982), 1527.


p
q
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
X
18
19
20
21
22
Y

And, finally, a visually similar small multiple, the *Consumer Reports* frequency-of-repair records for automobiles built from 1976 to 1981. This is a particularly ingenious mix of table and graphic, portraying a complex set of comparisons between manufacturers, types of cars, year, and trouble spots.

Consumer Reports, 47 (April 1982), 199–207. Redrawn.



Conclusion

Well-designed small multiples are

- inevitably comparative
- deftly multivariate
- shrunken, high-density graphics
- usually based on a large data matrix
- drawn almost entirely with data-ink
- efficient in interpretation
- often narrative in content, showing shifts in the relationship between variables as the index variable changes (thereby revealing interaction or multiplicative effects).

Small multiples reflect much of the theory of data graphics:

For non-data-ink, less is more.

For data-ink, less is a bore.[6]

[6] The two aphorisms on the meaning of "less" are, respectively, credited to Ludwig Mies van der Rohe and to Robert Venturi, *Complexity and Contradiction in Architecture* (New York, second edition, 1977), p. 17.