

Biểu đồ có mật độ dữ liệu cao

Nguyễn Thị Bích Thủy

Email: nbthuy2001@gmail.com

Tel: 0981 365 780

Nội dung:

1. Mật độ dữ liệu trên biểu đồ.
2. Biểu đồ có mật độ dữ liệu cao.
3. Dạng thiết kế biểu đồ có mật độ dữ liệu cao.

1. Mật độ dữ liệu trên biểu đồ:

Mật độ dữ liệu:

$$\text{Mật độ dữ liệu} = \frac{\text{kích thước ma trận dữ liệu}}{\text{diện tích biểu đồ.}}$$

Trong đó:

- ✓ ***Kích thước ma trận dữ liệu:*** số phần tử trong ma trận dữ liệu với 1 chiều là số độ chia để đo dữ liệu và 1 chiều là các biến.
- ✓ ***Diện tích biểu đồ:*** diện tích của khu vực dành để vẽ biểu đồ.

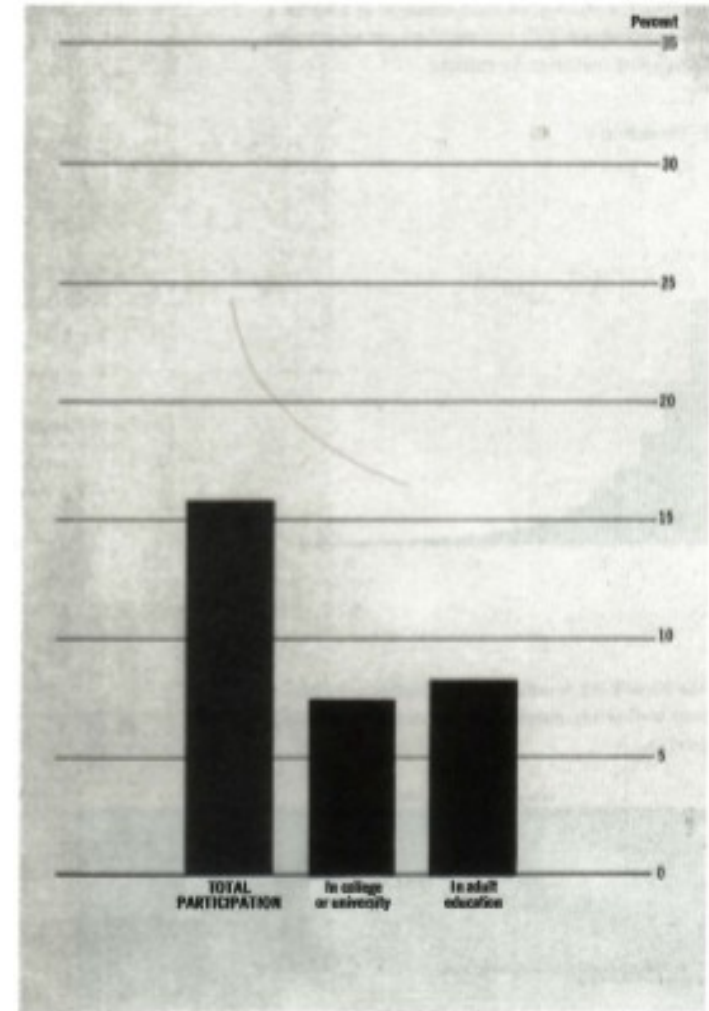
Ví dụ 1:

Mật độ dữ liệu quá nhỏ, biểu đồ chưa hiệu quả.

- ✓ Diện tích của biểu đồ = 171 cm^2 .
- ✓ Kích thước ma trận dữ liệu :
 $4 = 2 \text{ biến} * 2 \text{ mức chia}$.

Thanh bên trái không có ý nghĩa vì là tổng số liệu của 2 thanh bên phải.

- Mật độ dữ liệu trên hình bằng $0.02 \text{ dữ liệu}/1 \text{ cm}^2$.



? Cơ sở để xây dựng các biểu đồ có mật độ dữ liệu cao.

➤ *Khả năng phân biệt của mắt:*

Có thể nhìn thấy các ô vuông rất nhỏ trong 1 ô vuông diện tích 1cm^2 và được chia thành 900 ô vuông bằng nhau.

➤ *Thông tin dữ liệu lớn cần thể hiện trong biểu đồ.*

Ví dụ về khả năng phân giải của mắt:

Bảng tóm tắt thống kê về dân số Hoa Kỳ với lượng dữ liệu lớn:

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	3,406	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	1,860	2,129	2,094
Other domestic commercial equip.....	1,939	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	7,911	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	406	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.

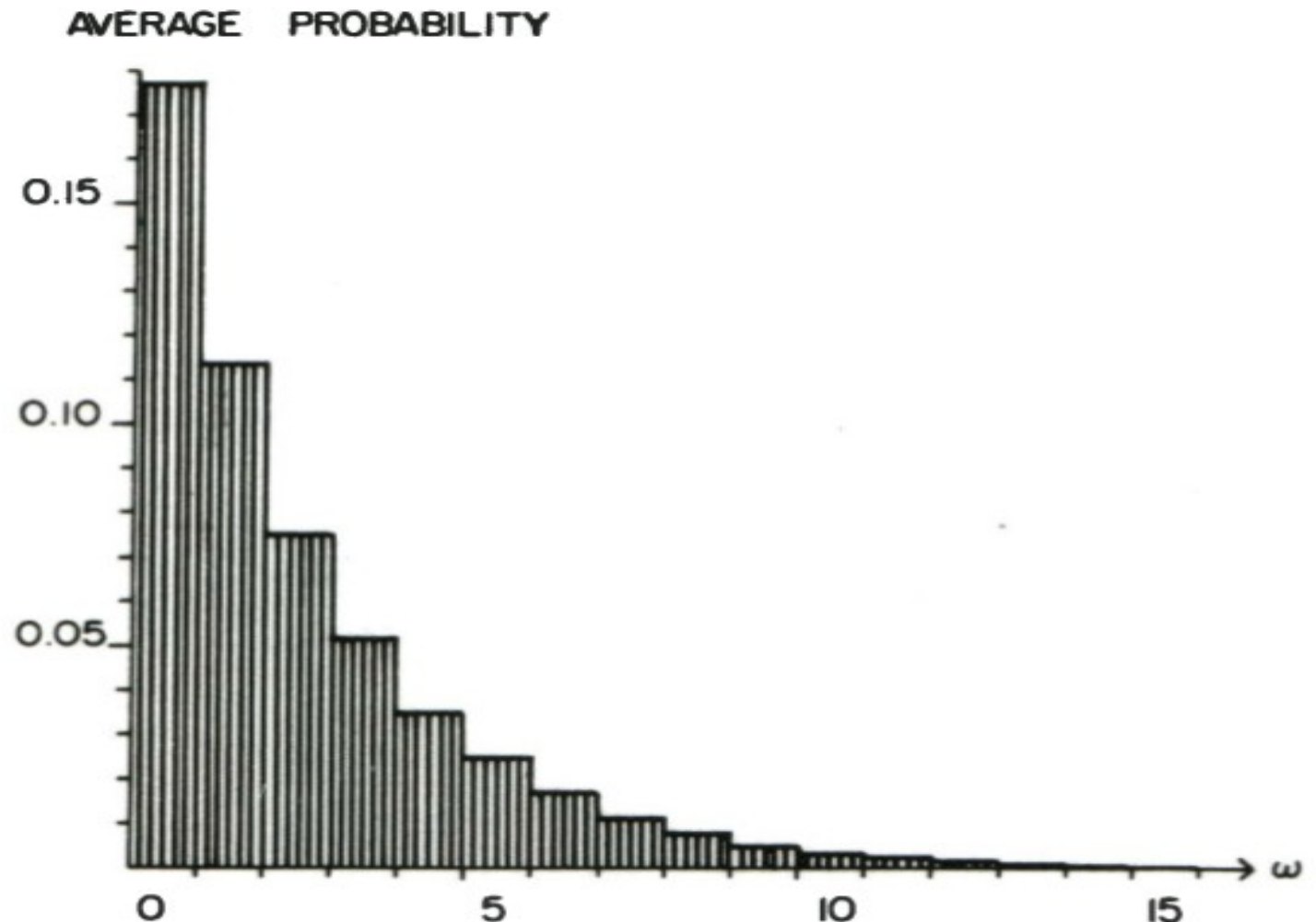
³ Includes railways, rapid transit systems, railroad rails, trackwork, and equipment.

2. Biểu đồ có mật độ dữ liệu cao

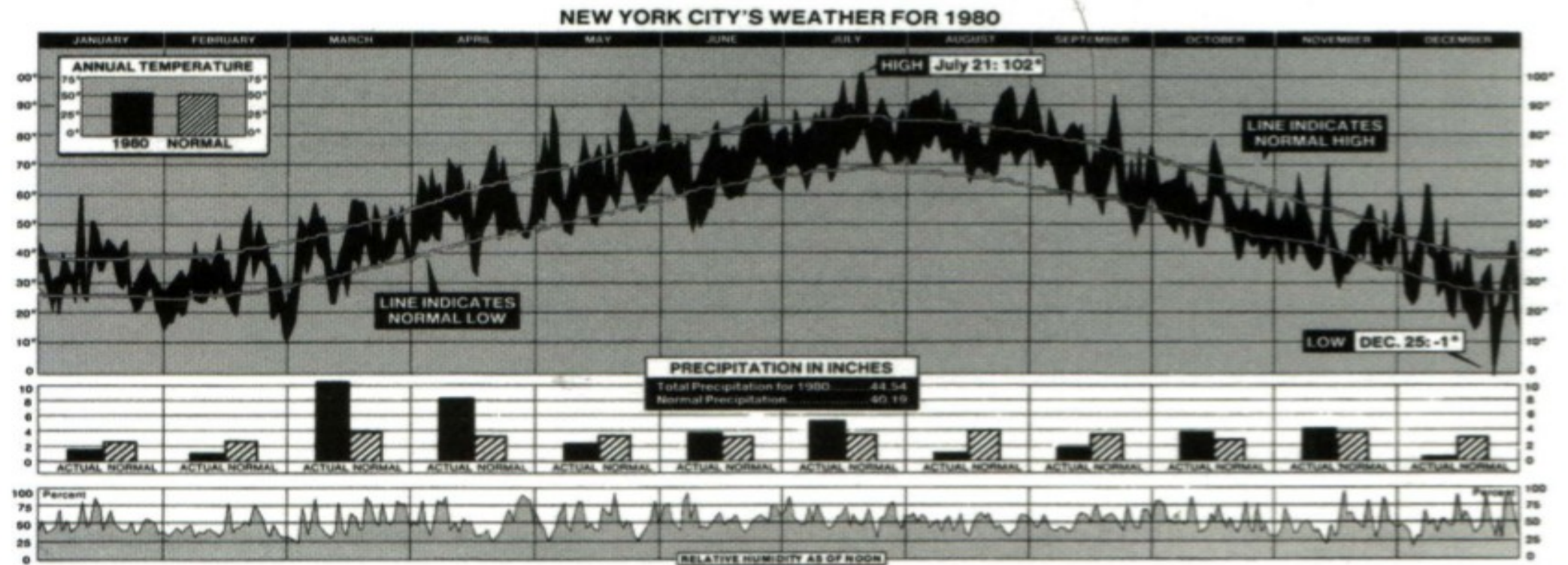
a. Một số ví dụ về biểu đồ có mật độ dữ liệu cao

Ví dụ 1:

- Mật độ dữ liệu trên hình bằng 0.6 dữ liệu/ 1cm^2 .

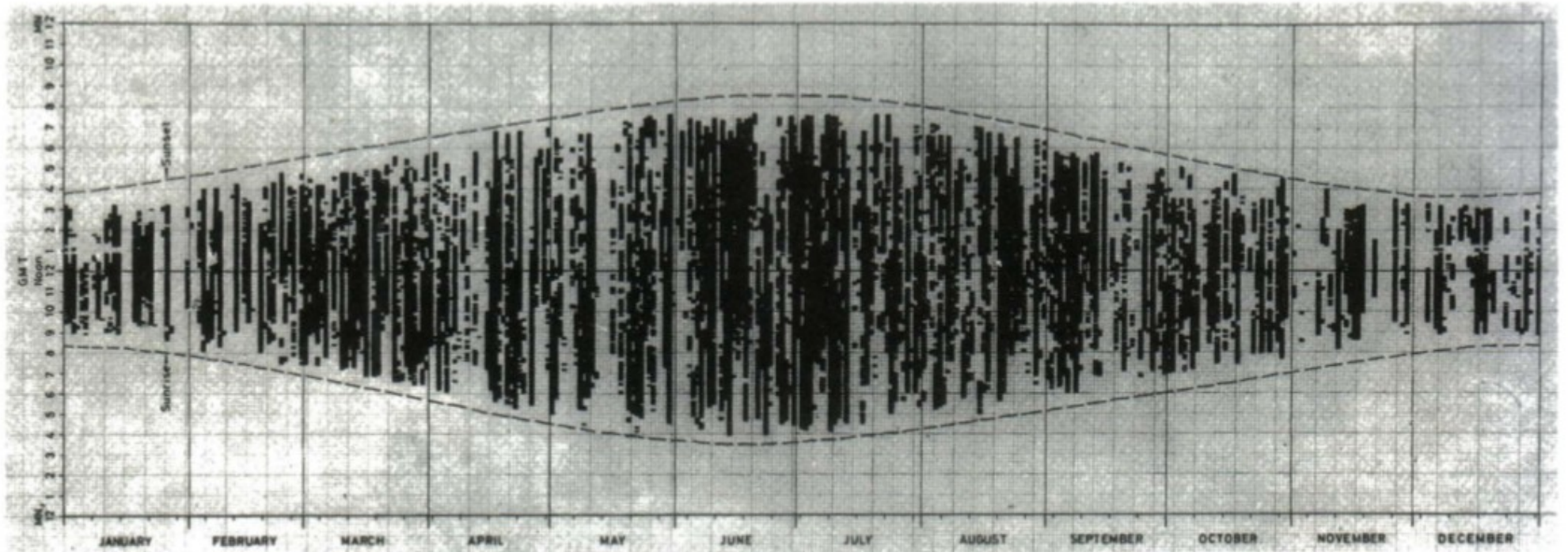


Ví dụ 2: ➤ Mật độ dữ liệu trên hình bằng 28 dữ liệu/1cm².
Biểu đồ hiệu quả.



Biểu đồ: chỉ số thời tiết của New York năm 1980

Ví dụ 3: ➤ Mật độ dữ liệu trên hình bằng 160 dữ liệu/1cm².
Biểu đồ hiệu quả.

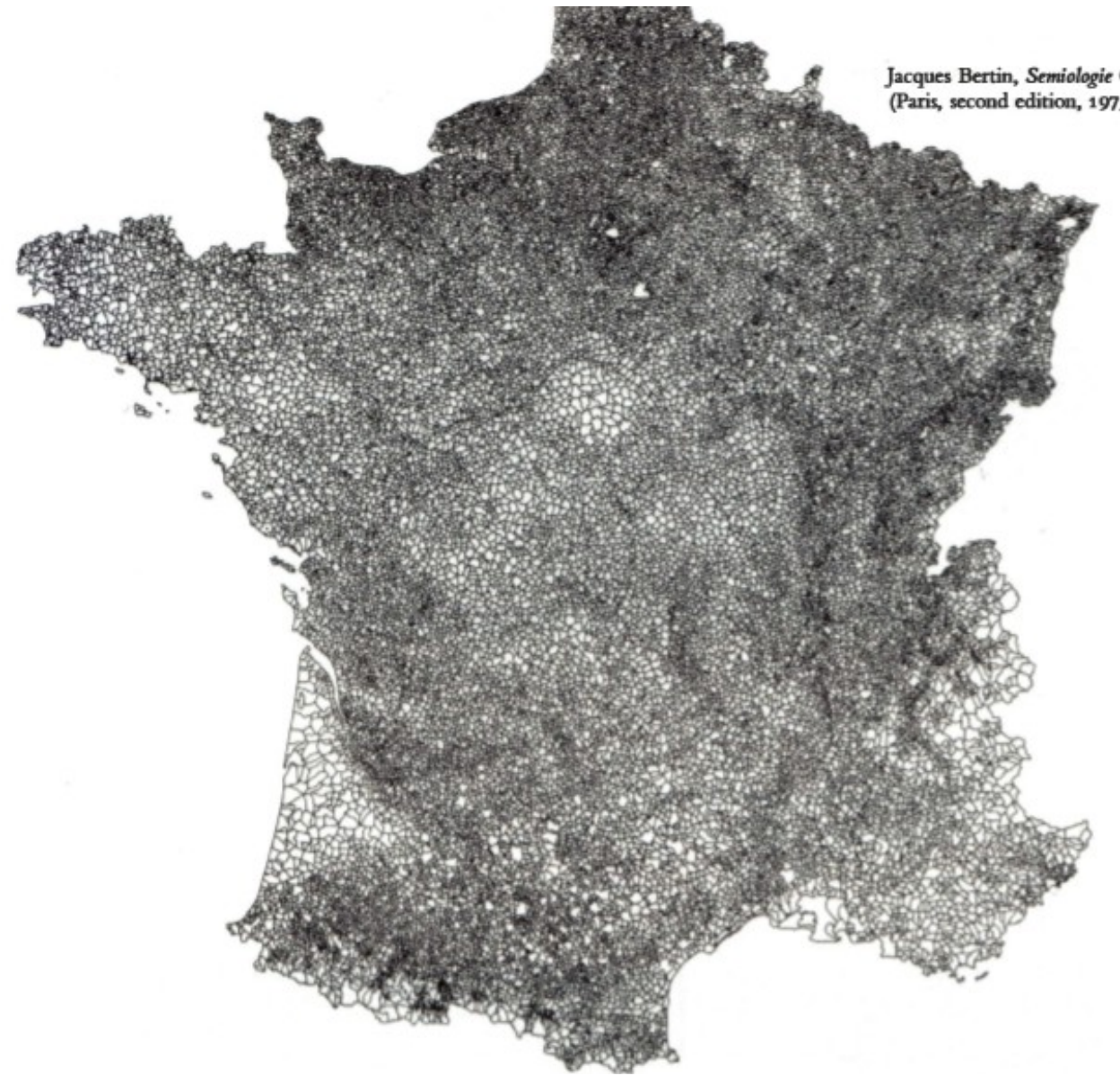


Biểu đồ: Lượng ánh nắng trung bình hàng năm

Ví dụ 4:

➤ Mật độ dữ liệu trên hình bằng 1400 dữ liệu/ 1cm^2 .

- Cần 30.000 chỉ số kinh độ, 30.000 chỉ số vĩ độ, 6 biến khác để mô tả và xác định hình dáng của ngôi làng. Ma trận dữ liệu có khoảng 240.000 phần tử.
- Diện tích của bản đồ: 175cm^2 .



*Bản đồ vị trí và ranh giới của
30.000 ngôi làng ở Pháp*

Bảng thống kê:

Bảng thống kê mật độ dữ liệu và kích thước ma trận dữ liệu của các biểu đồ trên 21 tạp chí lớn:

**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
<i>Nature</i>	48	3	362	177	15	3780
<i>Journal of the Royal Statistical Society, B</i>	27	4	115	200	10	1460
<i>Science</i>	21	5	44	109	26	316
<i>Wall Street Journal</i>	19	3	154	135	28	788
<i>Fortune</i>	18	5	31	96	42	156
<i>The Times (London)</i>	18	2	122	50	14	440
<i>Journal of the American Statistical Association</i>	17	4	167	150	46	1600
<i>Asahi</i>	13	2	113	29	15	472
<i>New England Journal of Medicine</i>	12	3	923	84	8	3600

<i>The Economist</i>	9	1	51	36	3	192
<i>Le Monde</i>	8	1	17	66	11	312
<i>Psychological Bulletin</i>	8	1	74	46	8	420
<i>Journal of the American Medical Association</i>	7	1	39	53	14	735
<i>New York Times</i>	7	1	13	35	6	580
<i>Business Week</i>	6	2	12	32	14	96
<i>Newsweek</i>	6	1	13	23	2	96
<i>Annuaire Statistique de la France</i>	6	1	25	96	12	540
<i>Scientific American</i>	5	1	69	46	14	652
<i>Statistical Abstract of the United States</i>	5	2	23	38	8	164
<i>American Political Science Review</i>	2	1	10	16	9	40
<i>Pravda</i>	0.2	0.1	1	5	4	20

Nhận xét về bảng thống kê

- Nhìn chung, một số biểu đồ có mật độ dữ liệu cao. Đa số có mật độ dữ liệu thấp. Những tạp chí như Wall Street Journal, The Times, Asahi, thường dùng các biểu đồ có mật độ dữ liệu cao. Trong khi đó, hầu hết các tạp chí ở Mỹ dùng các biểu đồ có mật độ dữ liệu ít hơn so với nơi khác.
- Rất ít biểu đồ có tỉ lệ hiển thị dữ liệu như bản đồ. Có lẽ cần hy vọng điều đó ở tương lai.

b. Biểu đồ có mật độ dữ liệu cao

- Dùng khi:
 - Muốn biểu diễn các dữ liệu có số lượng lớn, tác giả muốn truyền tải tối đa lượng thông tin của dữ liệu
 - Giúp người đọc có những so sánh, cái nhìn chính xác về thông tin mà biểu đồ truyền tải
- Nhược điểm: Cần nhiều công sức và chi phí để vẽ biểu đồ. Khi biểu đồ biểu diễn lượng dữ liệu lớn, dữ liệu bị co lại. Ví dụ, cần dùng những chấm nhỏ hơn để biểu diễn biểu đồ phân tán, hay nhiều đường rất mảnh để vẽ biểu đồ thanh.

Biểu đồ có mật độ dữ liệu cao

Một số lưu ý khi thiết kế biểu đồ

- ✓ Các tập dữ liệu đơn giản: dùng bảng biểu, đôi khi dùng từ ngữ, để thể hiện.
- ✓ Các tập dữ liệu lớn và phức tạp: dùng biểu đồ chứa nhiều dữ liệu sẽ hiệu quả hơn.
- ✓ Nếu 1 biểu đồ cần chứa nhiều dữ liệu, trước khi vẽ cần áp dụng một số kỹ thuật để giảm bớt lượng dữ liệu, ví dụ: tính trung bình dữ liệu, phân cụm dữ liệu, làm mịn dữ liệu.

Biểu đồ có mật độ dữ liệu cao

- Cách thiết kế: Để tăng mật độ dữ liệu có 2 cách:
 - Giảm diện tích của biểu đồ.
 - Tăng ma trận dữ liệu.
- Nhận xét:
 - Có thể giảm một nửa diện tích biểu đồ nếu dùng các nét vẽ mảnh hơn.
 - Áp dụng nhiều lần nguyên tắc co biểu đồ (the Shrink Principle), có thể tạo ra những biểu đồ có sức mạnh lớn hơn.

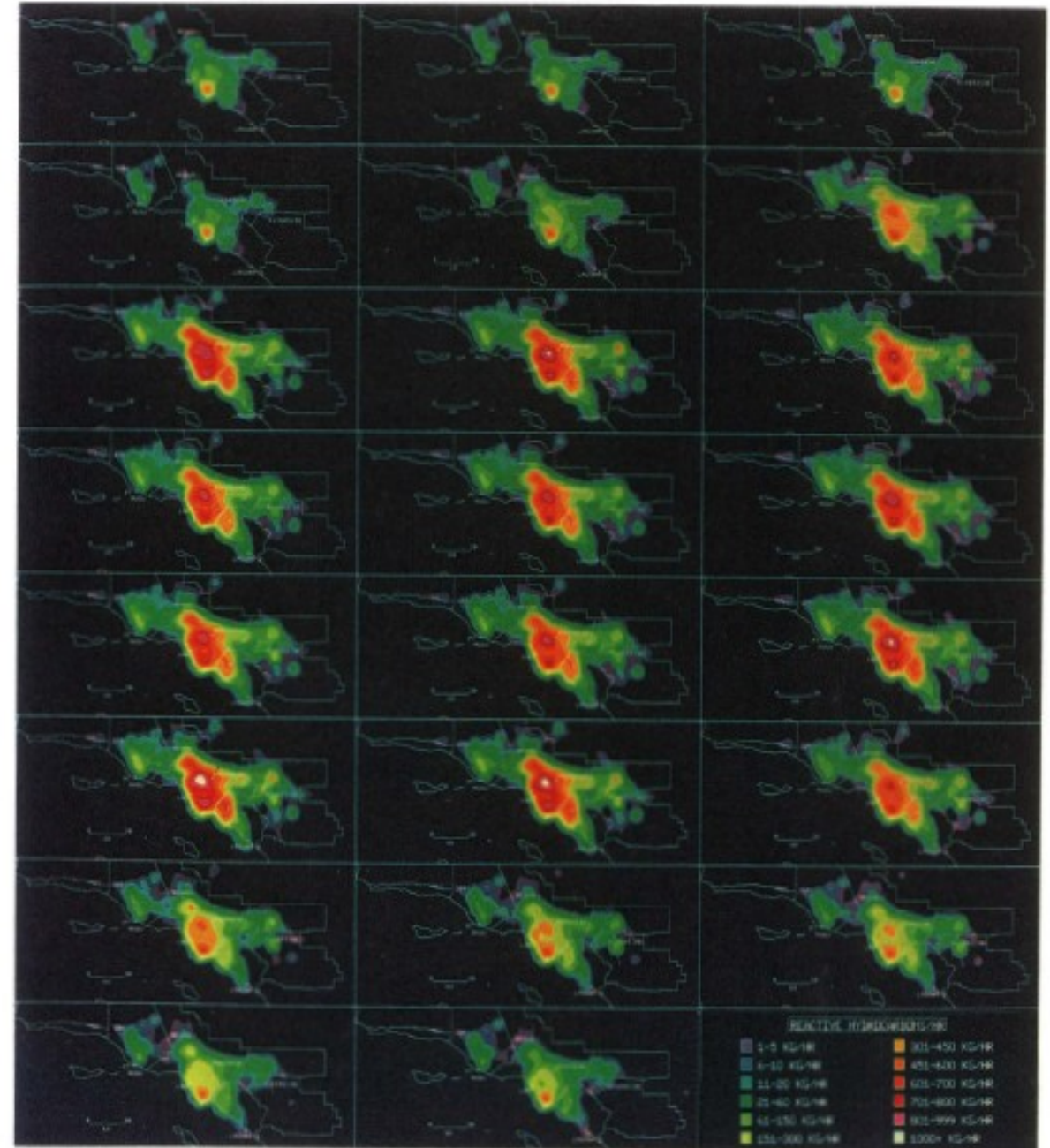
3. Dạng thiết kế biểu đồ có mật độ dữ liệu cao

- ✓ Biểu đồ con giống nhau:

Bằng cách sử dụng chuỗi các biểu đồ nhỏ giống hệt nhau trong một khung hình nhằm thể hiện sự dịch chuyển thông tin (ví dụ như theo thời gian, theo không gian,...)

Ví dụ:

*Sự thay đổi ô nhiễm không khí tại
Los Angeles trong 23 giờ liên tiếp
(biểu đồ nhiều cửa sổ):*



Ví dụ:

Biểu đồ về phân bố các tế bào ung thư ở nam và nữ

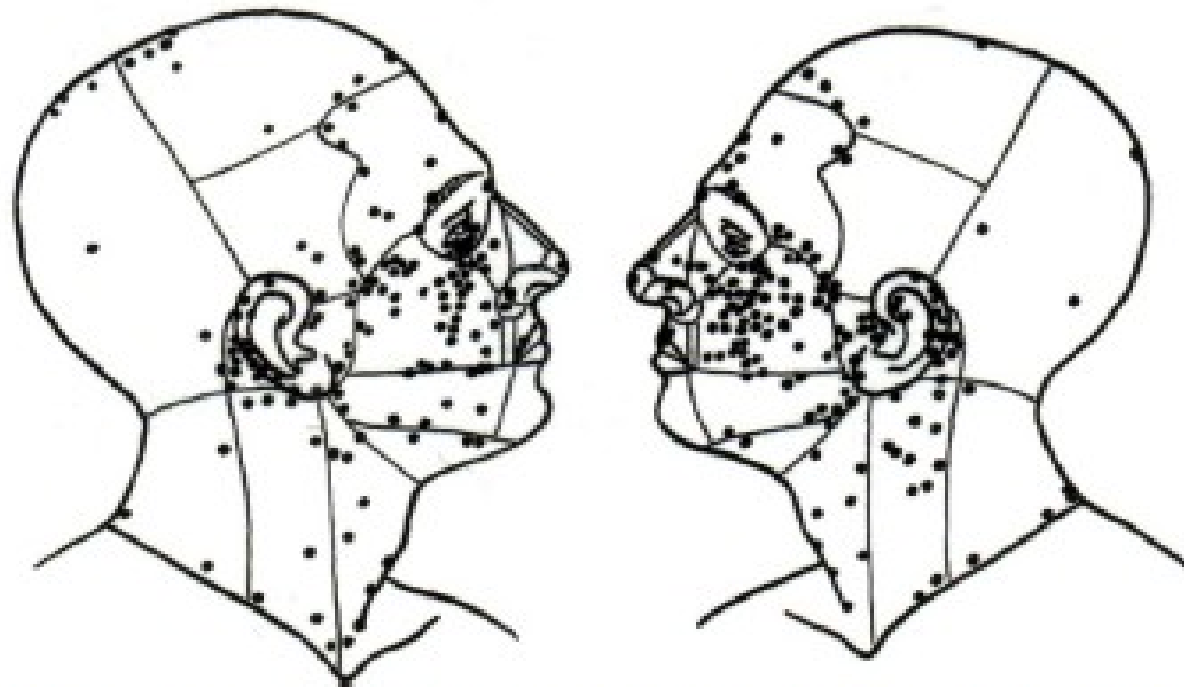


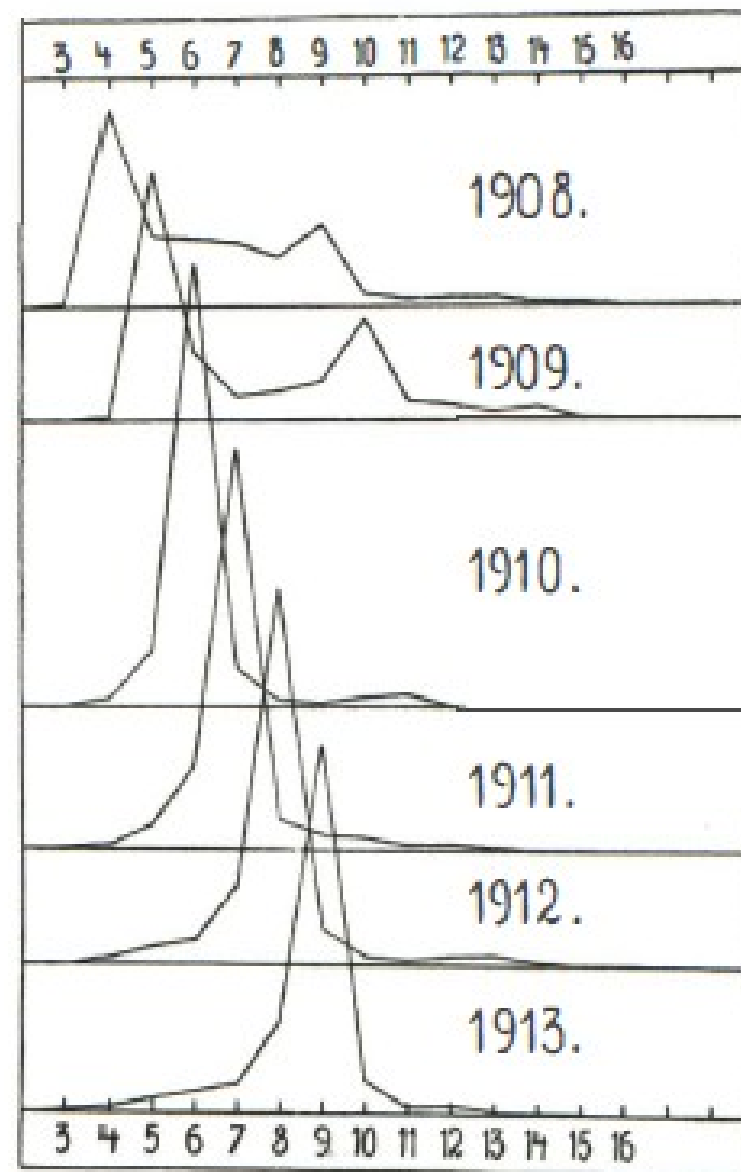
Abb. 1. Verteilung von 269 primären Melanomen auf Kopf und Hals

Ví dụ:

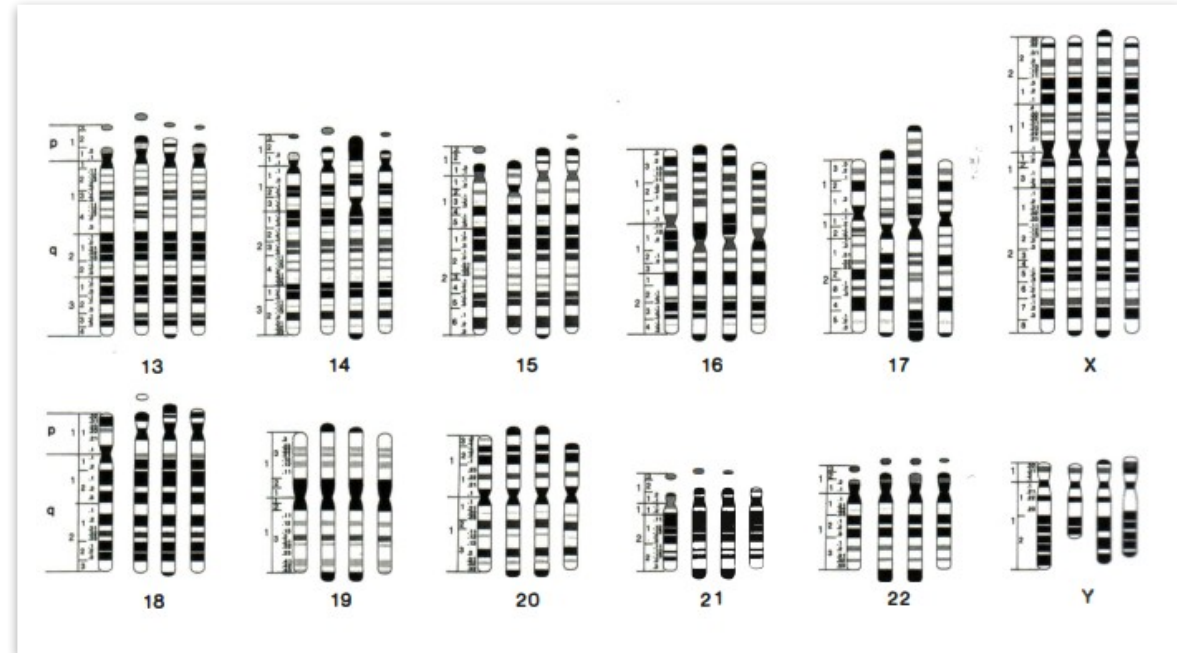
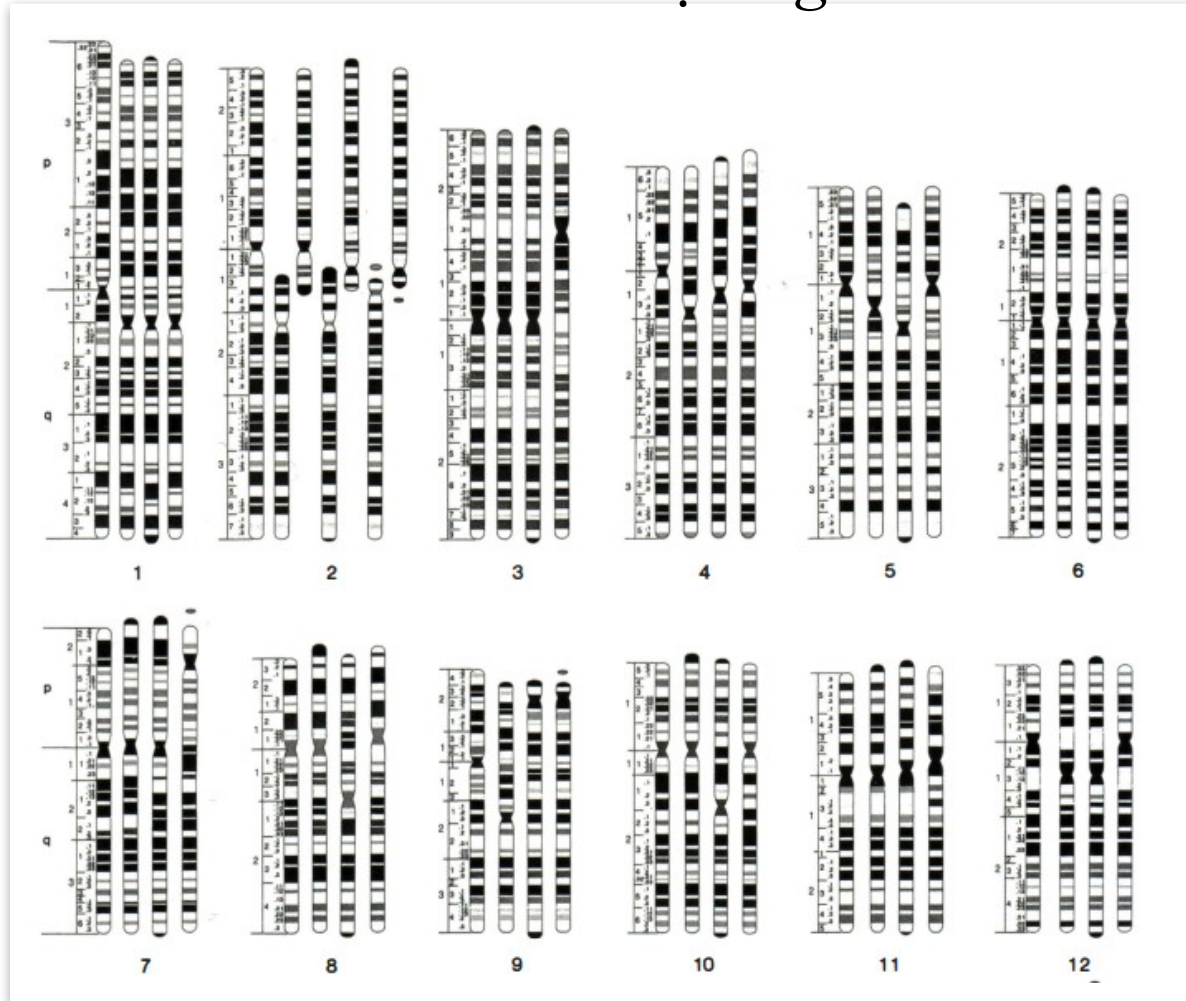
Biểu đồ về phân phối tuổi của cá trích được bắt từ năm 1908 đến 1913

Nhận xét:

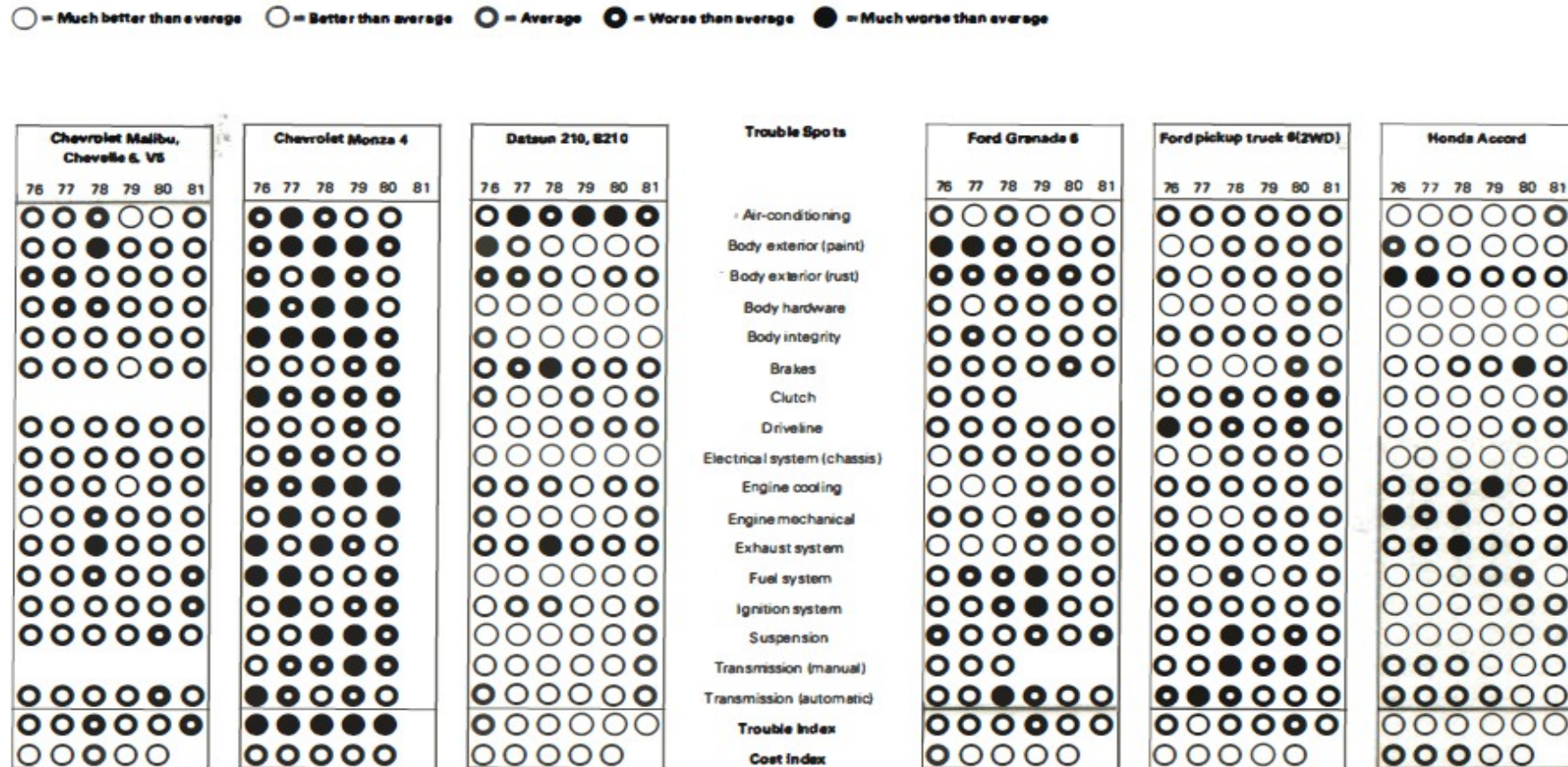
Cá trích được đẻ năm 1904 là nhiều nhất (1908 đạt 4 tuổi).



Ví dụ: *Biểu đồ so sánh nhiễm sắc thể của người, tinh tinh, vượn người và đười ươi.*



Ví dụ: *Biểu đồ so sánh việc bảo trì của các loại xe từ năm 1976 đến 1981*



Các thông tin được thể hiện: nhà sản xuất, bộ phận bị hỏng, mức độ, loại xe và năm sửa chữa.

Kết luận:

Một biểu đồ gồm các biểu đồ con giống nhau nếu được thiết kế tốt sẽ mang lại những lợi ích sau:

- Dễ dàng so sánh thông tin
 - Thể hiện được mối liên quan của nhiều biến.
 - Mật độ dữ liệu cao
 - Ma trận dữ liệu lớn
 - Hiệu quả trong giải thích thông tin.
- * Cách thiết kế: Để tăng mật độ dữ liệu có 2 cách:
- Giảm diện tích của biểu đồ.
 - Tăng ma trận dữ liệu.



**THANK
YOU**
for
**LISTENING TO
MY PRESENTATION**