



The Visual Display of Quantitative Information

Edward R. Tufte

Julia Grippo, Judy Dai, and Mihir Keskar



Graphical Practice



What are
data graphics?



“Data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color.”

“Graphics *reveal* data.”

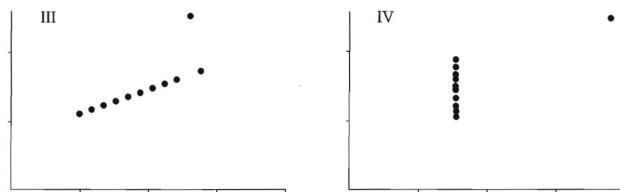
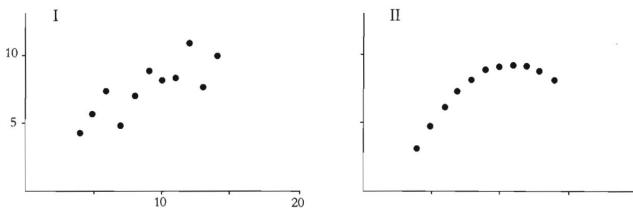
“Graphics were preferable to tables because graphics showed the shape of the data in a comparative perspective...any variable quantity could be placed in relationship to any other variable quantity, measured for the same unit of observation” - Playfair

Statistics

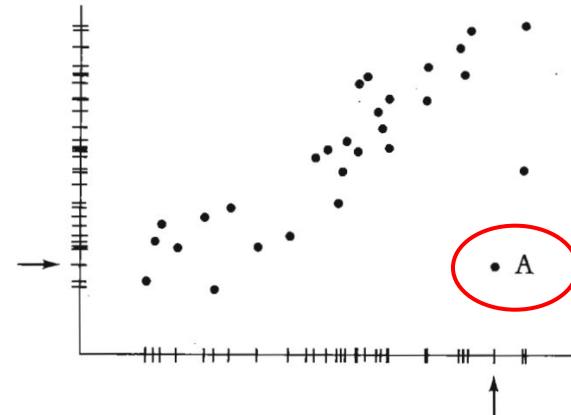
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
 $t = 4.24$
sum of squares $X - \bar{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
 $r^2 = .67$

Data Graphics



Benefit



One of the First Data Maps

1686: Data Map by Edmond Halley

A data map by Halley depicting trade winds and monsoons on a world map.

This map introduces the first use of the retinal variable, orientation, to establish where the wind comes from.



1546: Cosmographia by Petrus Apianus

At this time, they had yet to decide to mark mappings with quantitative values. They also have yet to create naming conventions and the establishment of the axes

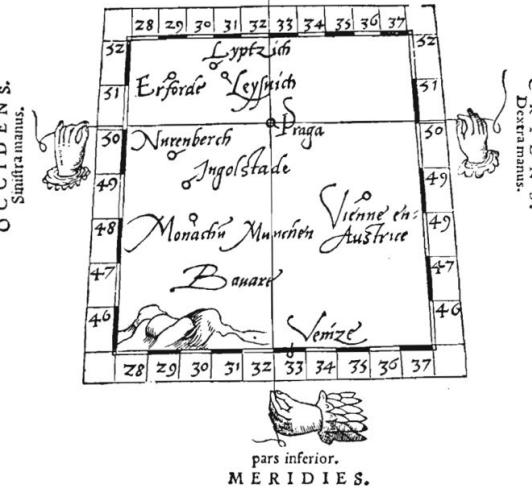
Ecce formulam, vsum, atque

structuram Tabularum Ptolomai, cum quibusdam locis, in
quibus studiosus Geographia se satis exercere potest.

S E P T E M B R I O .
pars superior.



O C C I D E N T I S .



- The data map was created in the 17th century once cartographic and statistical skills came together
- The first economic time-series was plotted in 1786



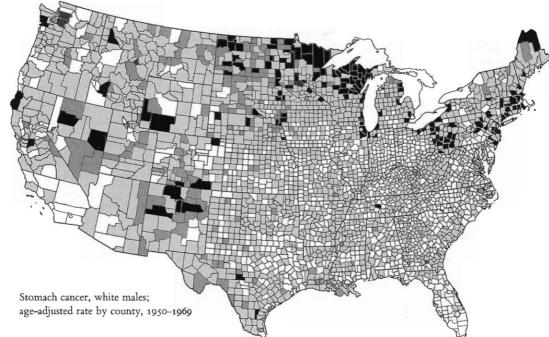
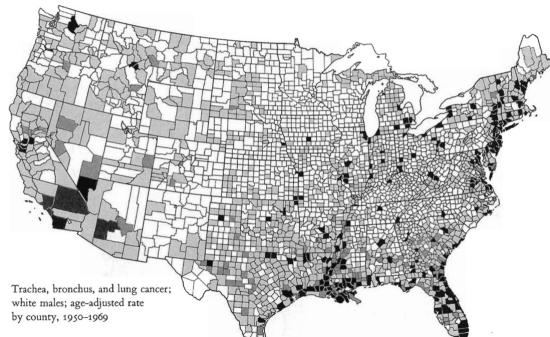
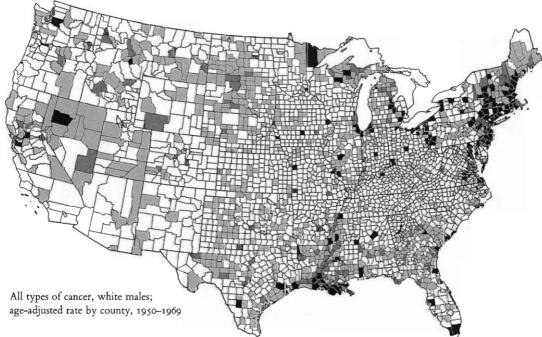
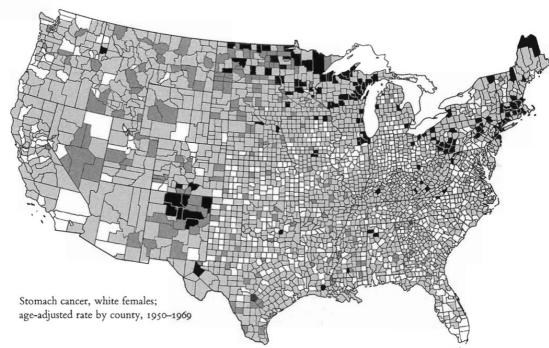
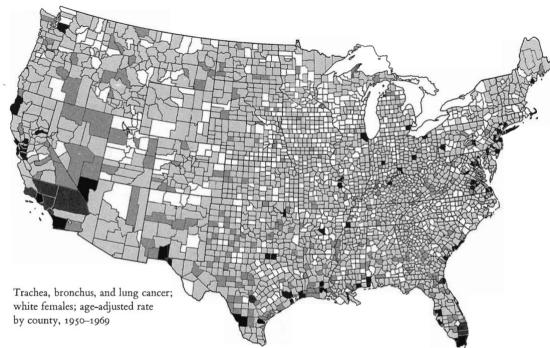
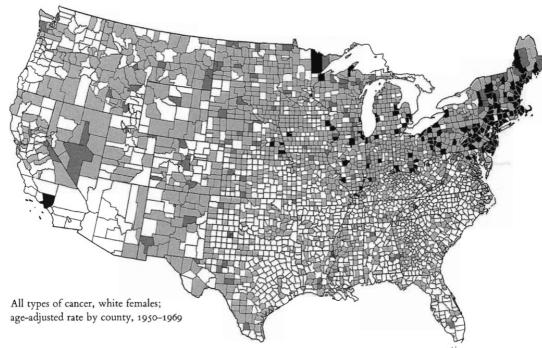
What should graphical displays do?



Graphical Excellence Should:

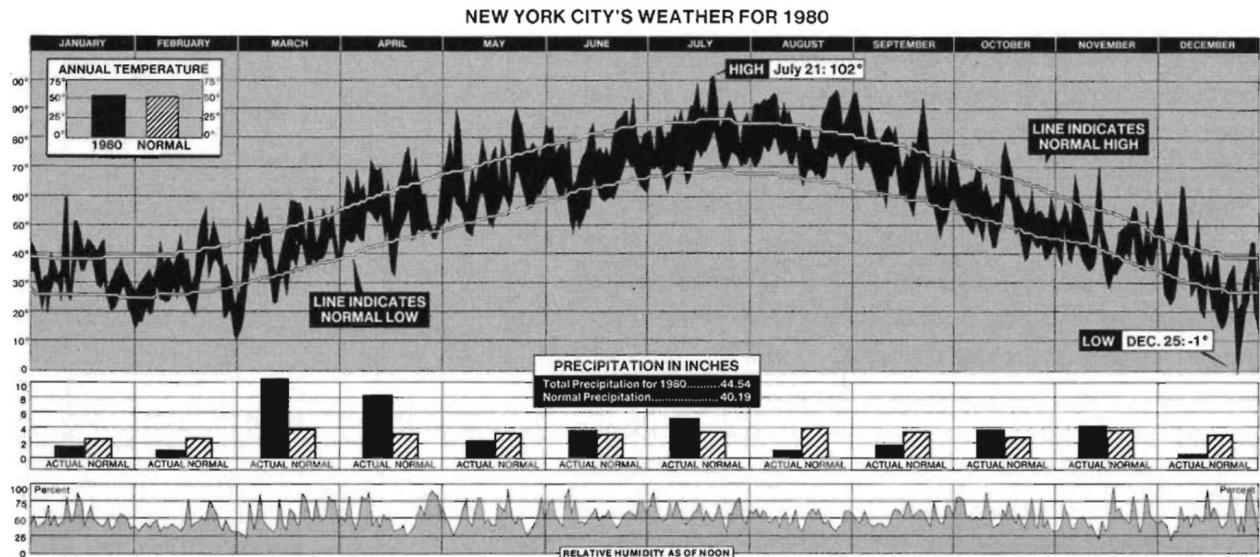
- Show the data while avoiding distortion
- Make large data sets coherent
- Encourage the eye to visually compare pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure

“Excellence in statistical graphics consists of complex ideas communicated with **clarity, precision, and efficiency**”



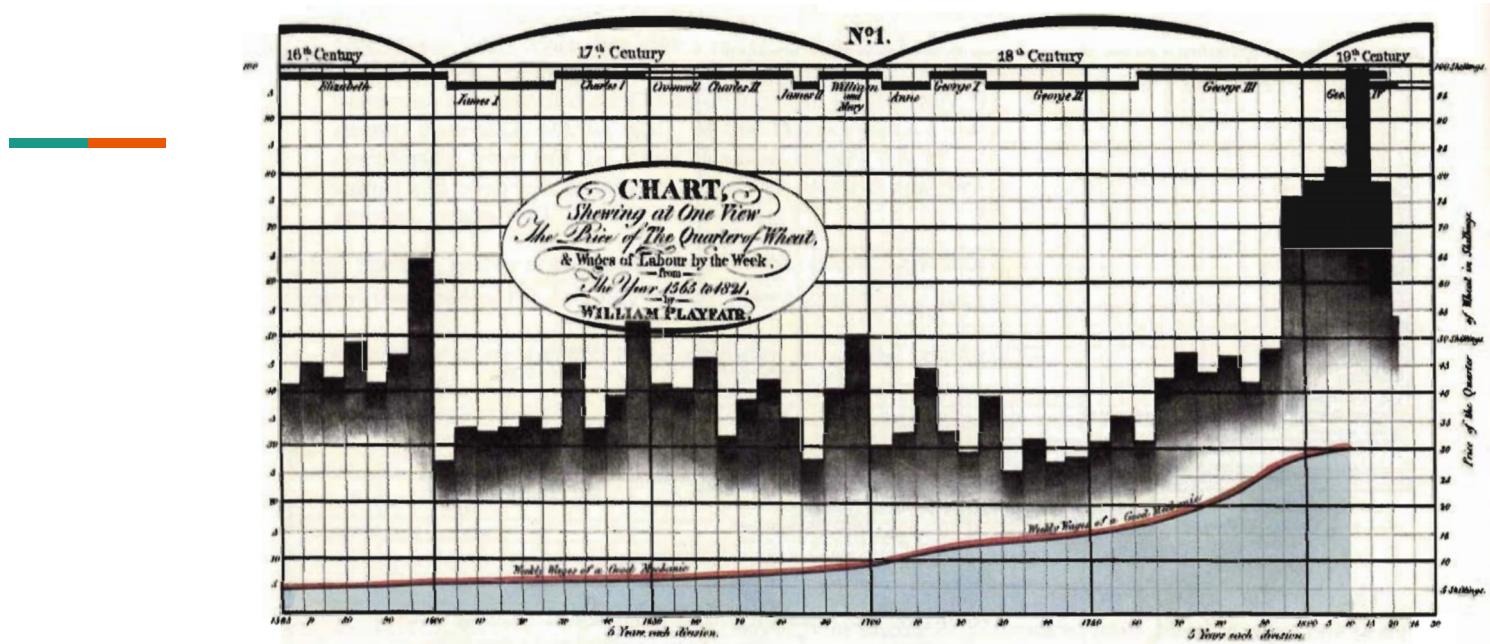
Time Series:

- Most frequently used form of graphic design
- Best for big data sets with real variability
- Successfully organizes a large collection of numbers, makes comparisons between different parts of the data, and tells a story



New York Times, January 11, 1981, p. 32.

*NOTE: The problem with time-series is that the simple passage of time is not a good explanatory variable

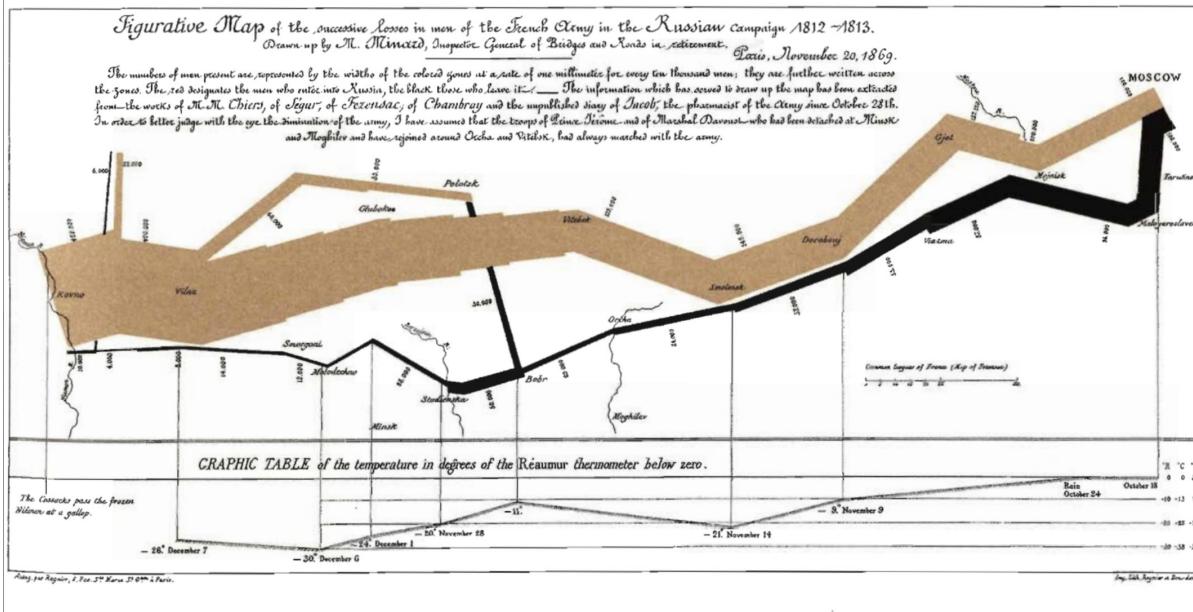


Bar Chart:

- Created by Playfair because year-to-year data were missing and he needed a design to portray the one-year data that were available
- Believed that small, noncomparative, highly labeled data sets belonged in tables
- Introduces labeling of x and y axes

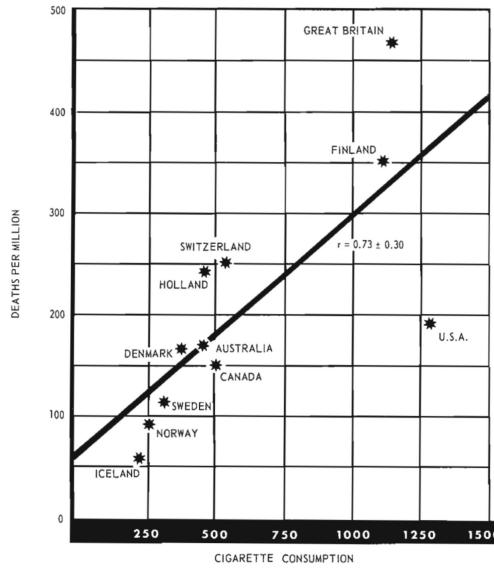
Space-Time Graphs:

- Adds spacial dimensions to the design so that the data is moving over time
- Below example is combination of data map and time series
- Tells a coherent story with multivariate data
- Depicts 6 plotted variables





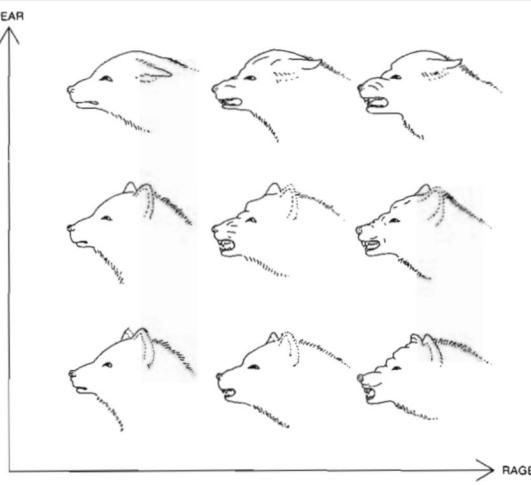
CRUDE MALE DEATH RATE FOR LUNG CANCER
IN 1950 AND PER CAPITA CONSUMPTION OF
CIGARETTES IN 1930 IN VARIOUS COUNTRIES.



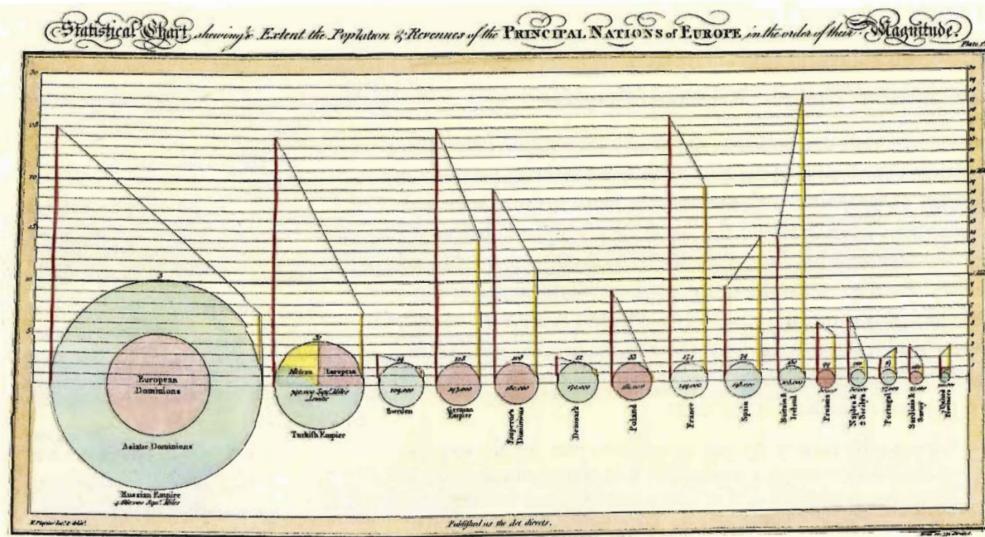
Scatterplot:

- Links at least two variables
- Depicts relationship between X and Y variables

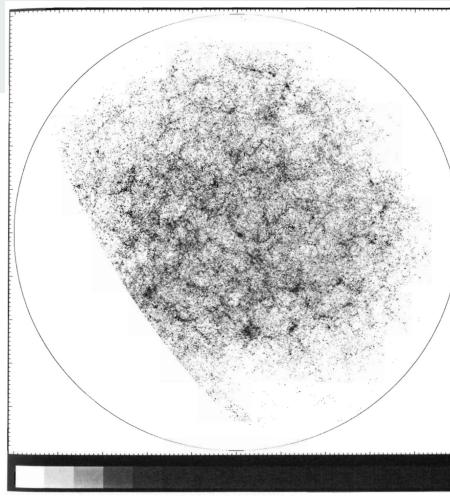
Shape



Area/Size



Value



Position





Principles of Graphical Excellence

- Well-designed presentation of interesting data
- Consists of complex ideas communicated with clarity, precision, and efficiency
- Gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space

Graphic excellence is **nearly always multivariate**
It also **requires the telling of truth** about the data



Graphical Integrity



“Graphical excellence begins with
telling the truth about the data.”

Deception results from the incorrect extrapolation of visual expectations generated at one place on the graph to other places



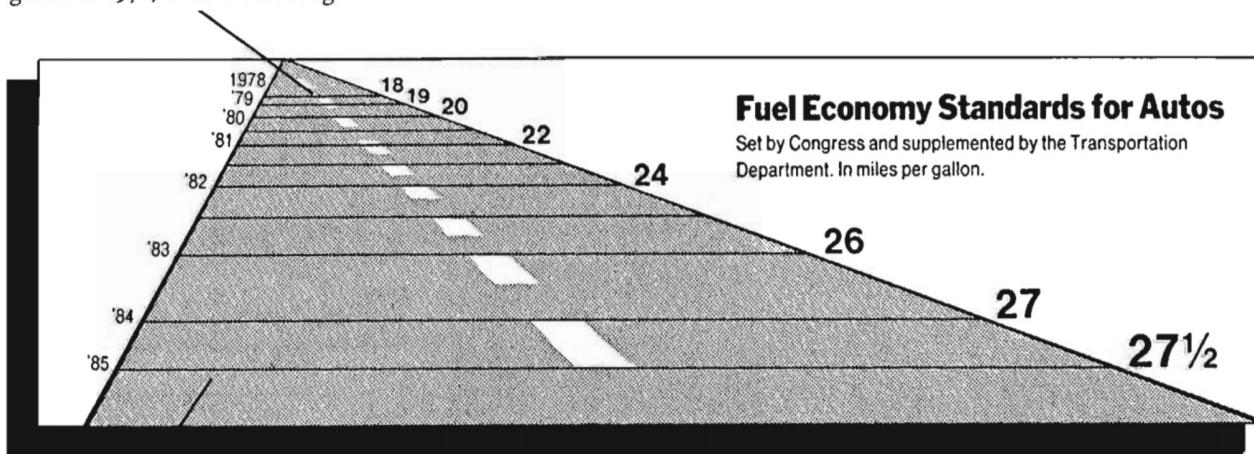
How to Achieve Graphical Integrity

- The representation of numbers should be directly proportional to the numerical quantities represented
- Create a clear, detailed graph through use of labeling and keys to defeat ambiguity and graphical distortion
- Show data variation, not design variation
- In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units
- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data
- Graphics must not quote data out of context

Lie Factor

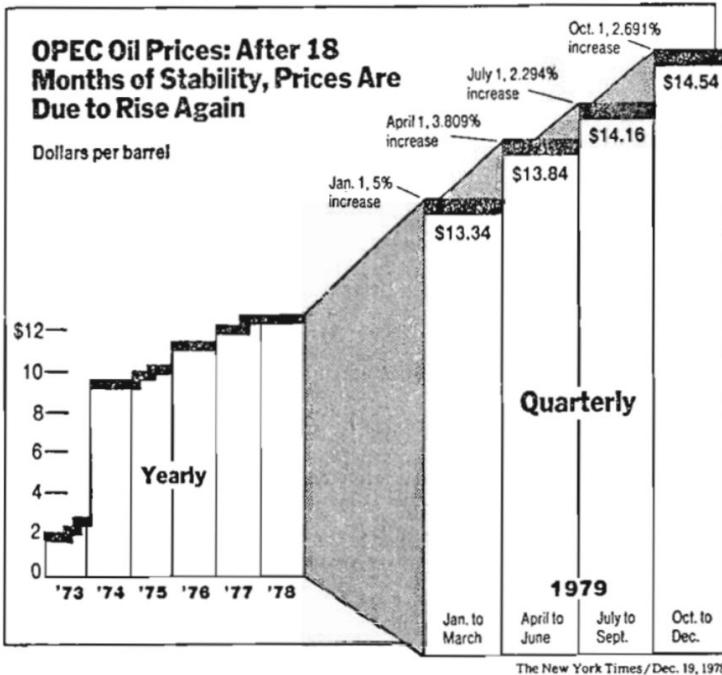
$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Design Variation



Five different vertical scales show the price:

During this time

1973–1978
January–March 1979
April–June 1979
July–September 1979
October–December 1979

one vertical inch equals

\$8.00
\$4.73
\$4.37
\$4.16
\$3.92

And two different horizontal scales show the passage of time:

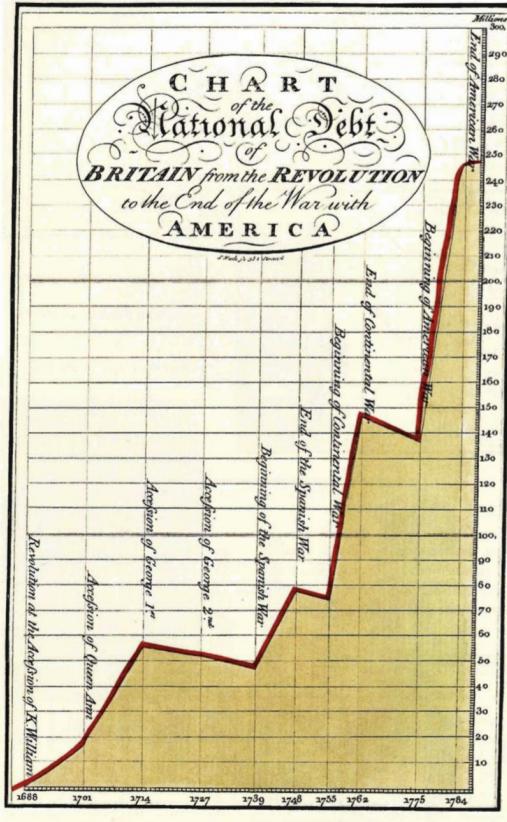
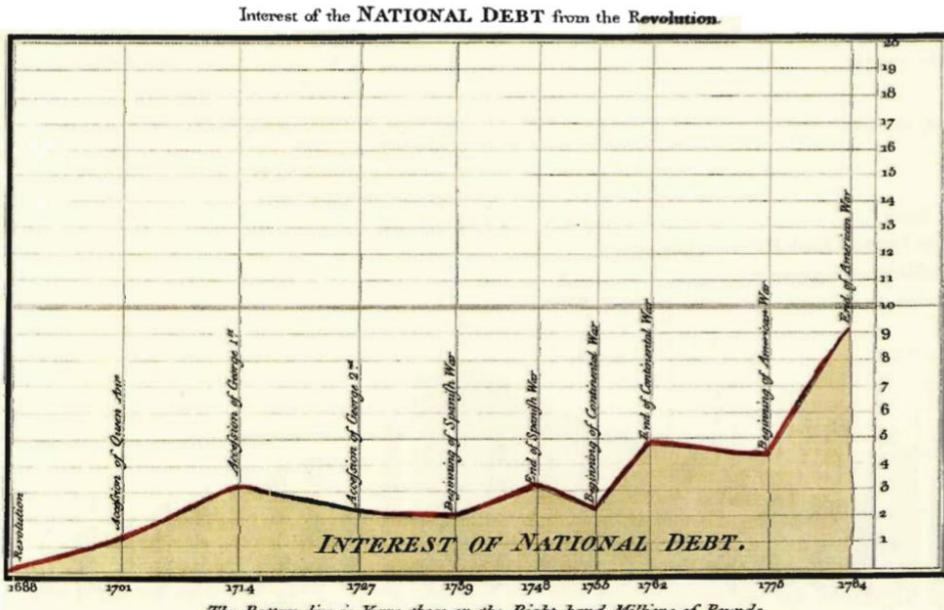
During this time

1973–1978
1979

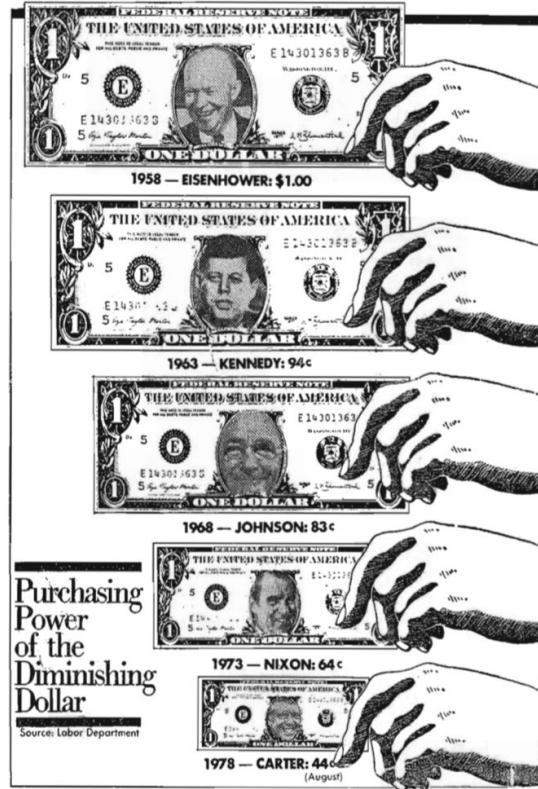
one horizontal inch equals

3.8 years
0.57 years

Graph Format



Visual Area



If the area of the dollar is accurately to reflect its purchasing power, then the 1978 dollar should be about twice as big as that shown.

“To be truthful and revealing, data graphics must bear on the question at the heart of quantitative thinking, “**Compared to what?**”



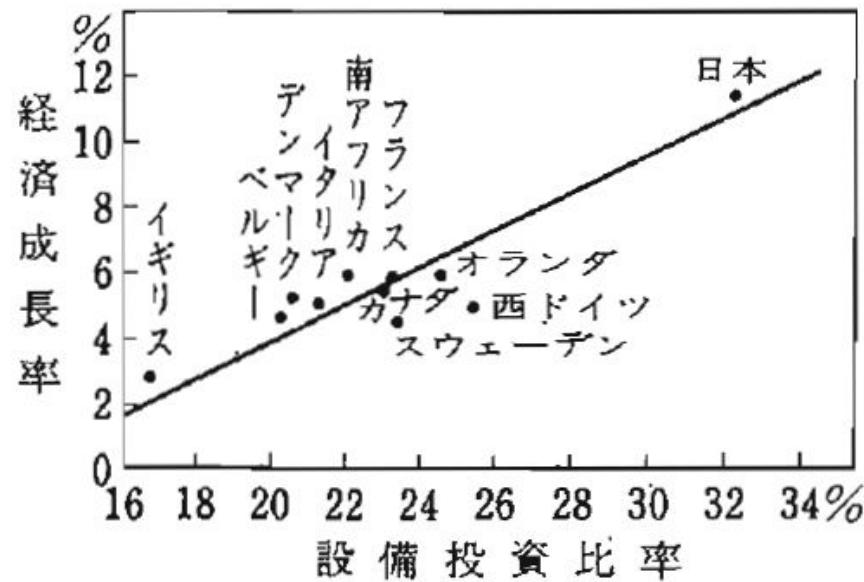
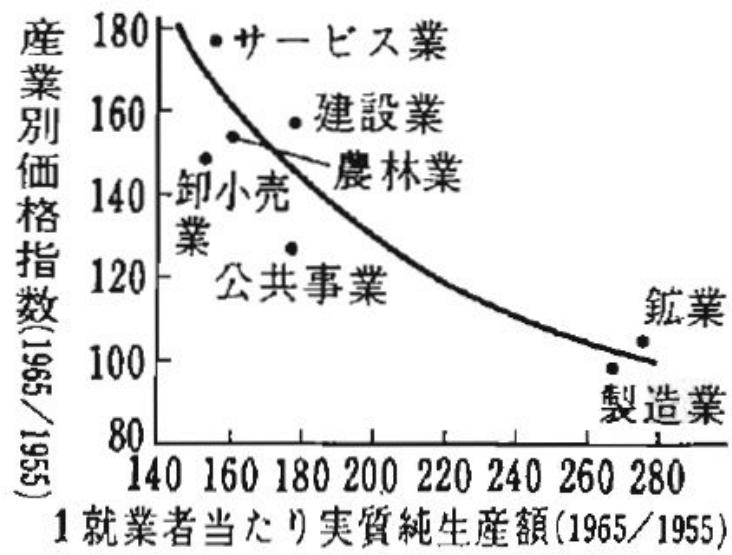


Graphic Mediocrity

The dislike of quantitative data

Engender graphics that:

- lies
- only the simplest design
- miss real news actually in the data



Good examples from Japan.

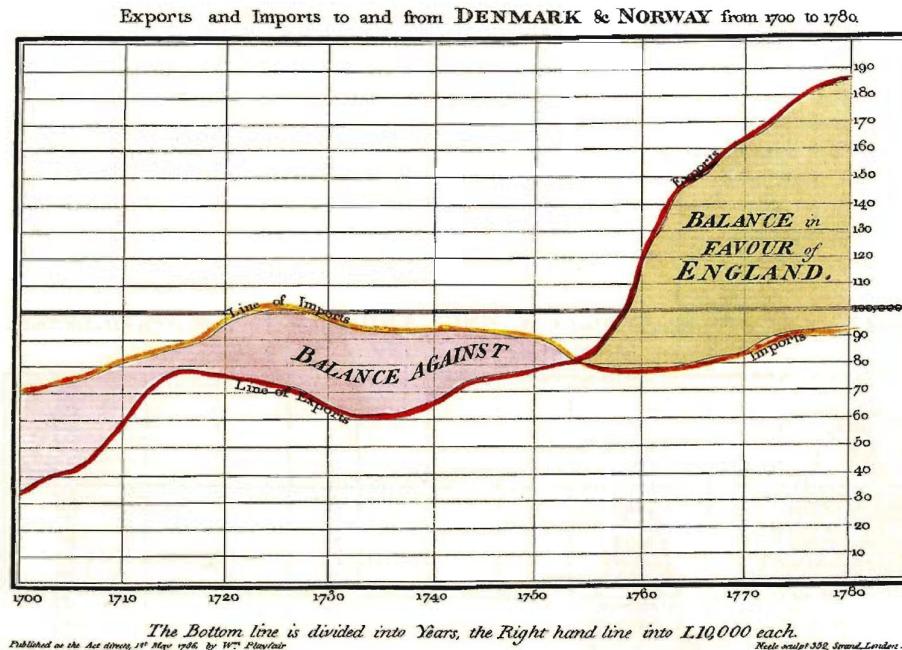
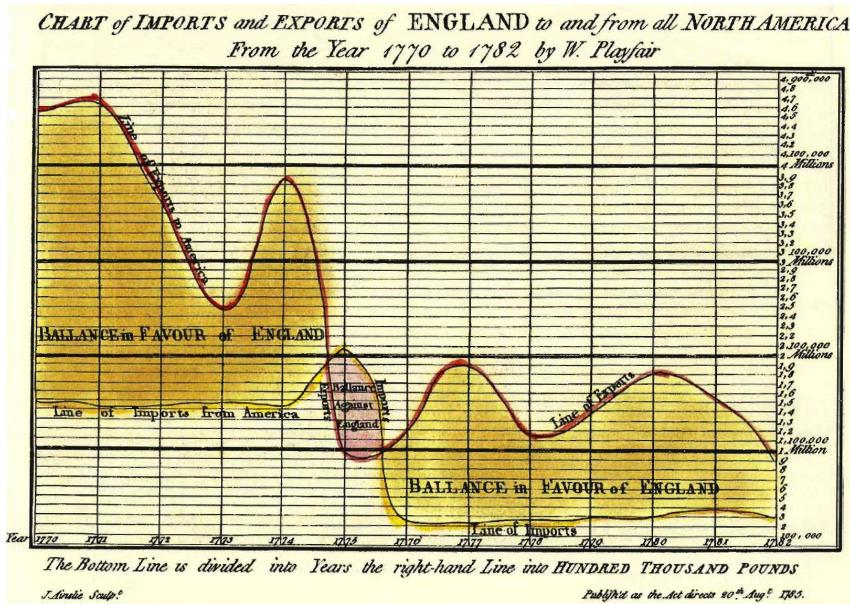


Theory of Data Graphics



“Above all else show the data”

The Devotion of Ink



The Commercial and Political Atlas, by W. Playfair.

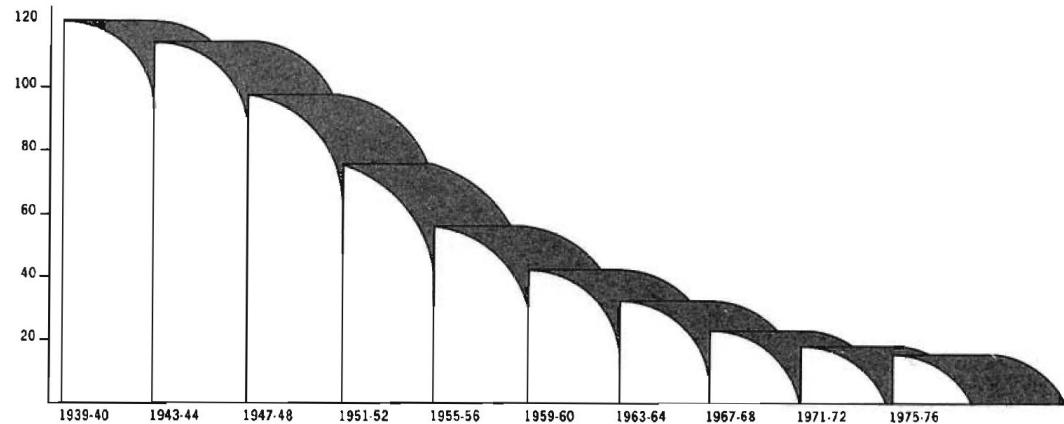
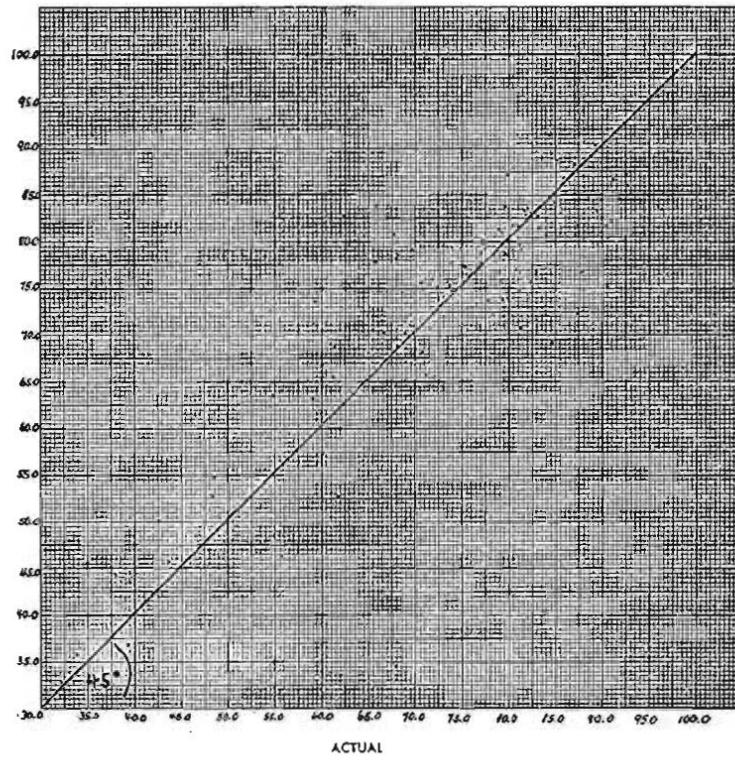

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

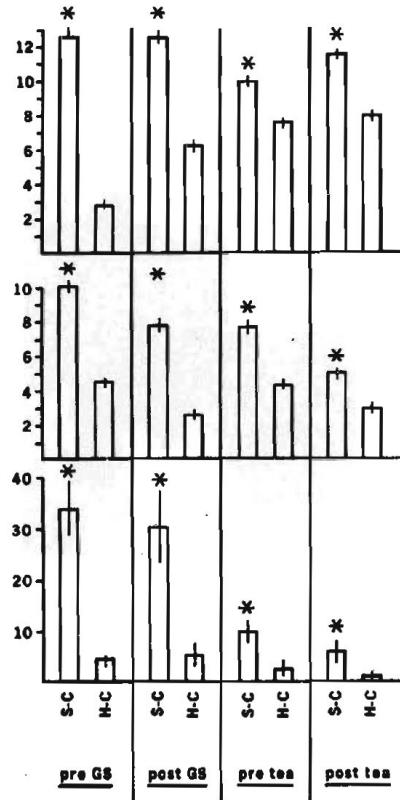
= 1.0 – proportion of a graphic that can be erased without loss of data-information.

- 
1. Above all else show the data
 2. Maximizing the Share of Data-Ink
 - Maximize the data-ink ratio
 3. Two Erasing Principles
 - Erase non-data-ink.
 - Erase redundant data-ink
 4. Revise and edit

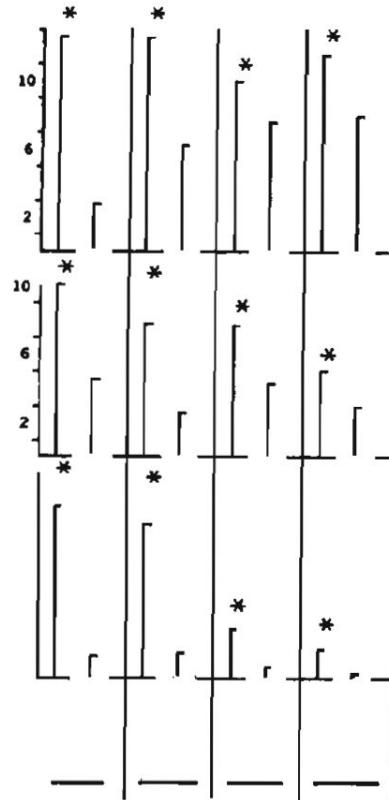
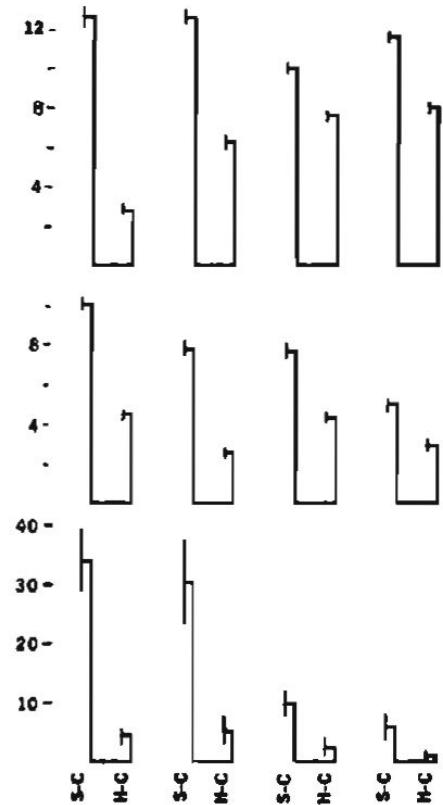
Relationship of Actual Rates of Registration to Predicted Rates
(104 cities 1960).



Unnecessary Data Ink



Improvements + Redesign

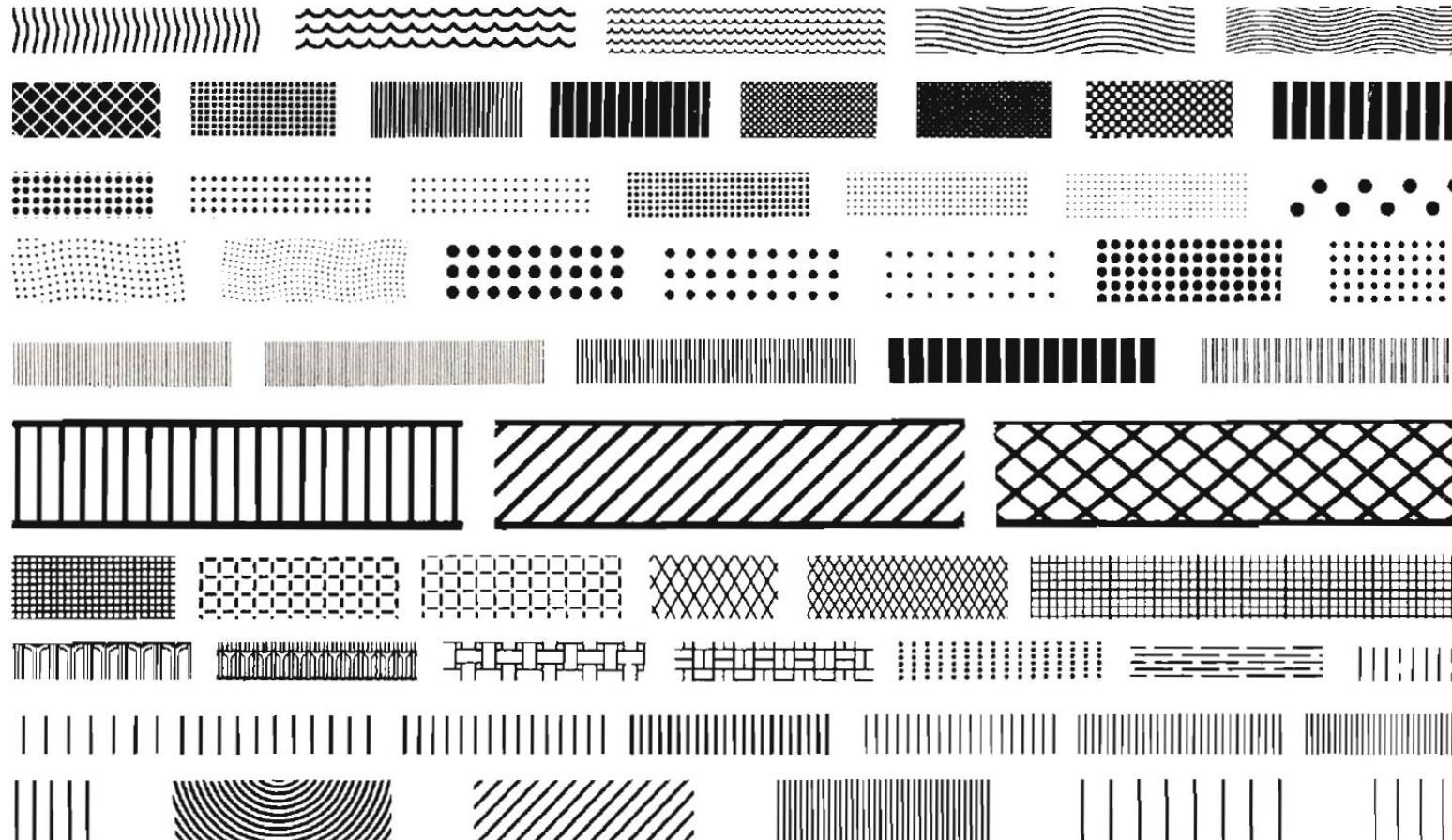


65% erased

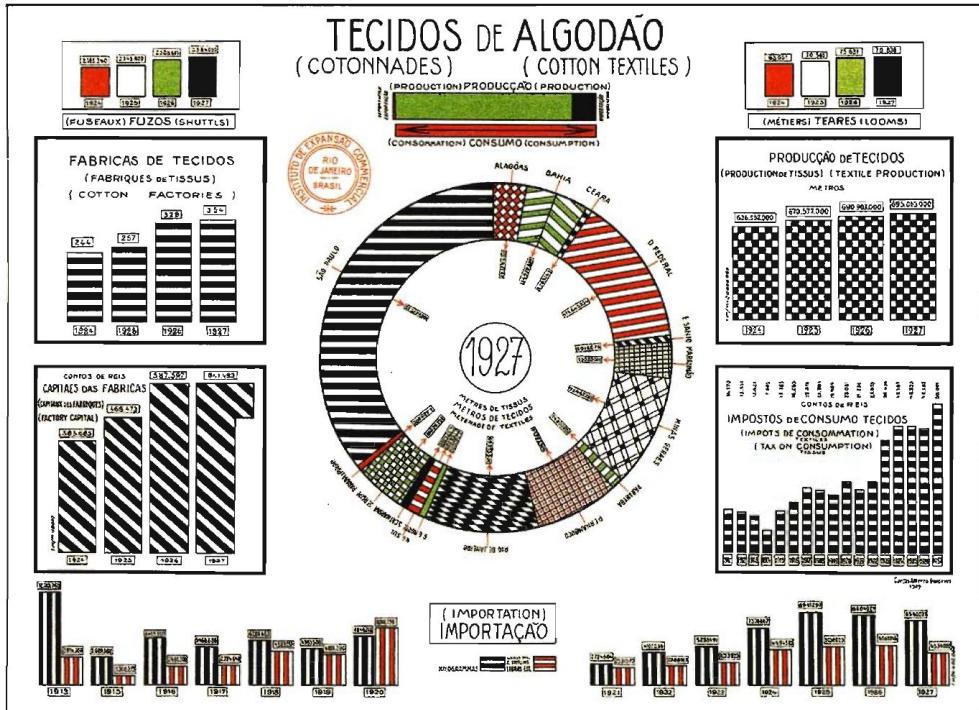


Chartjunk

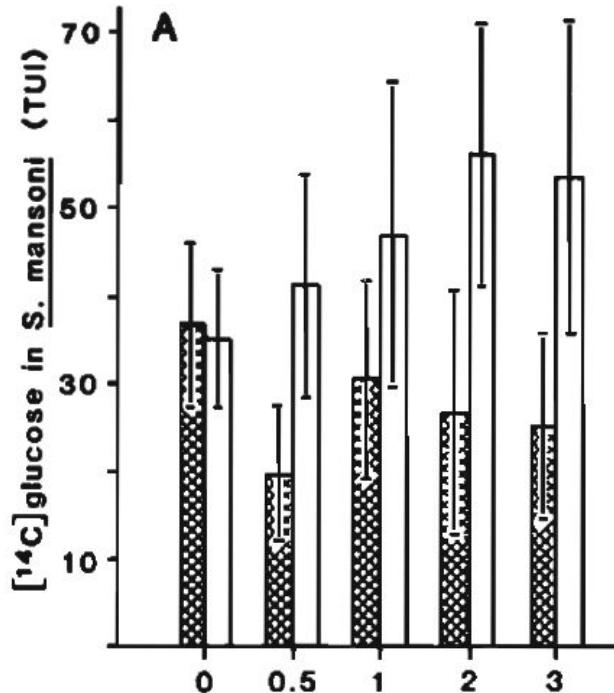
Overwhelming



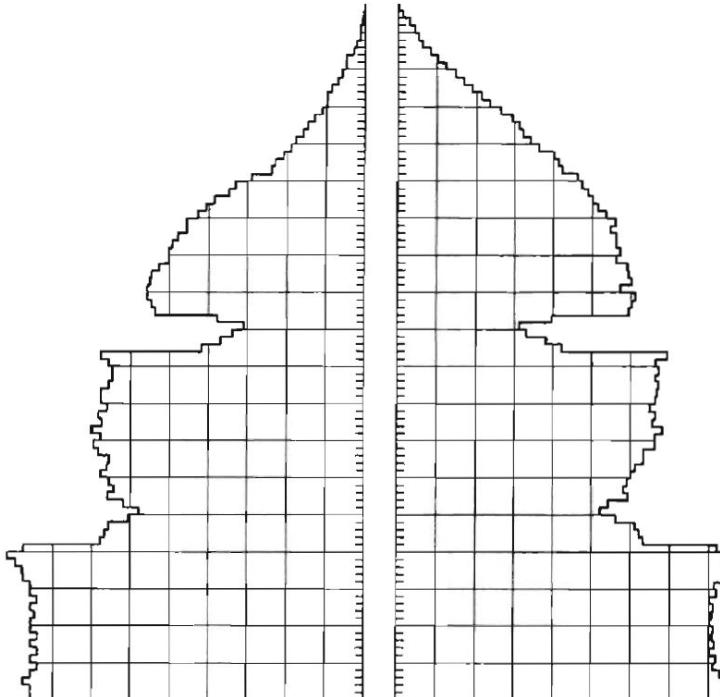
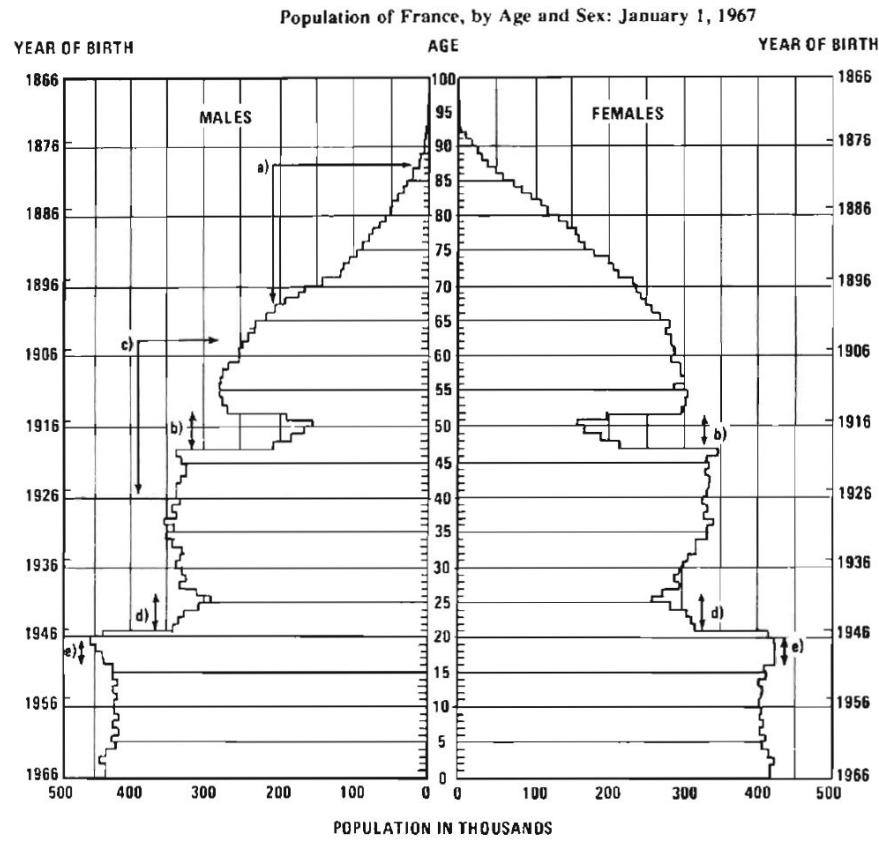
Optical Illusion



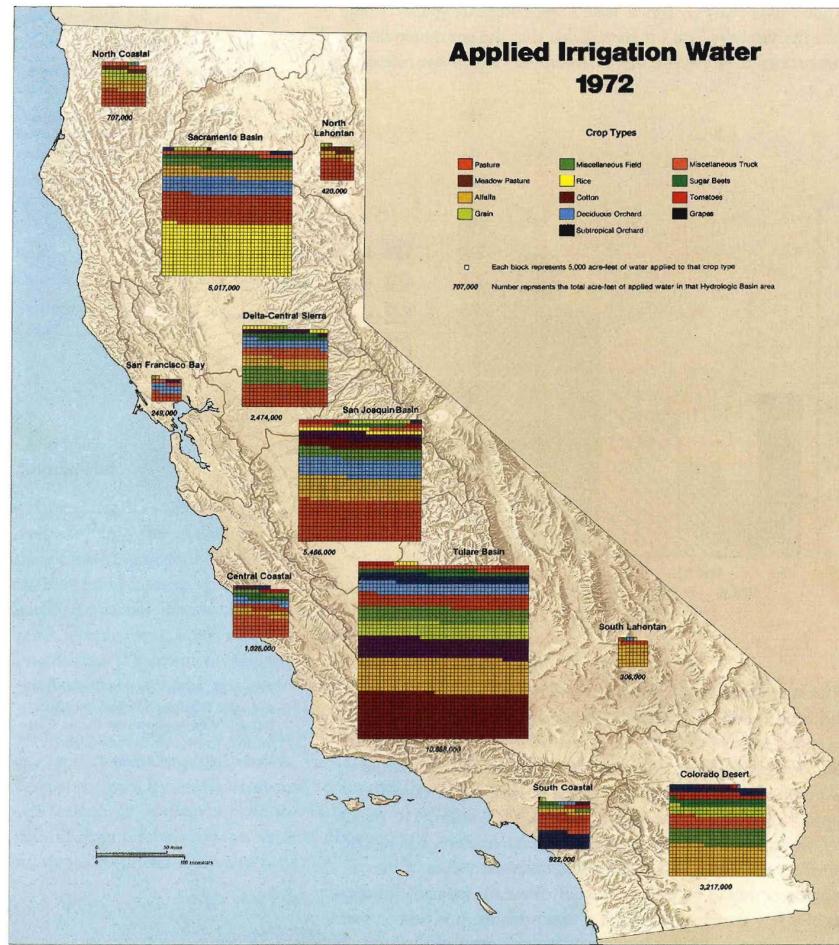
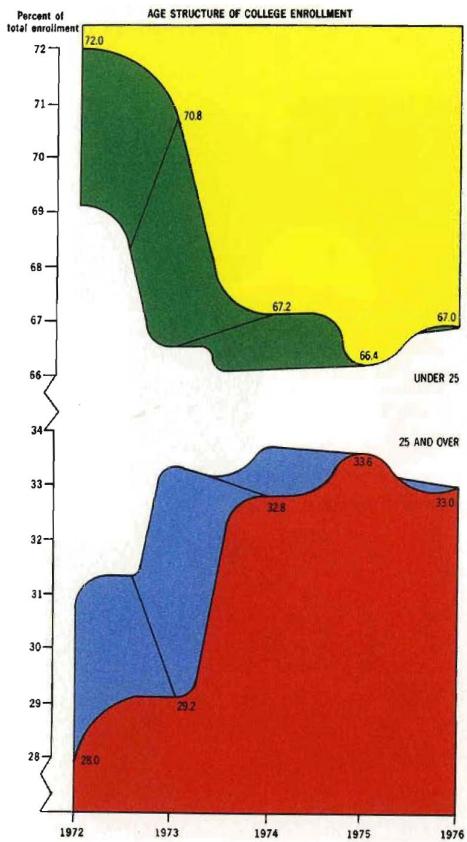
Moire Vibration



Grid

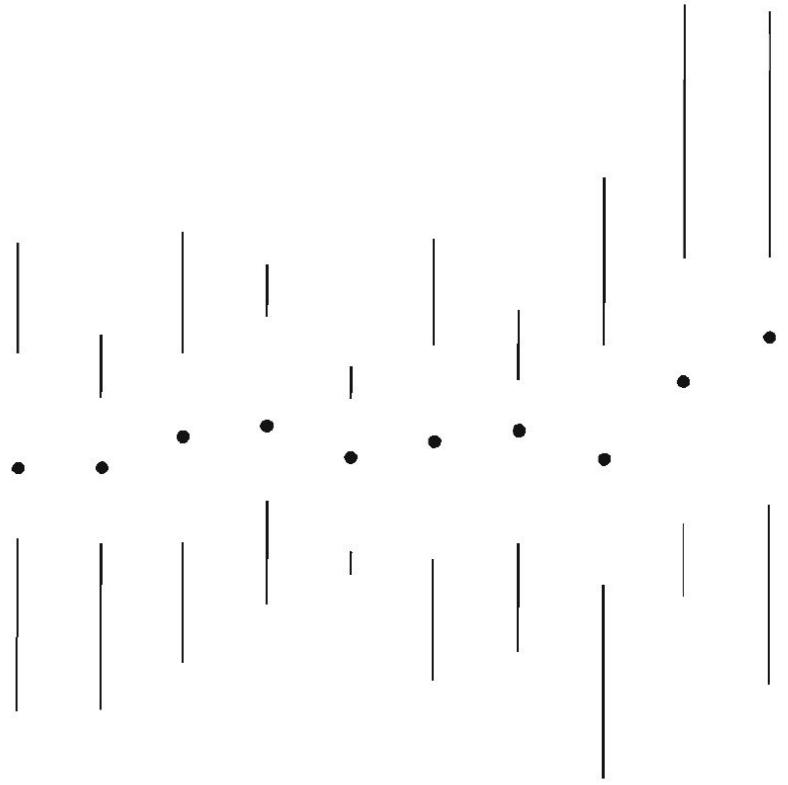
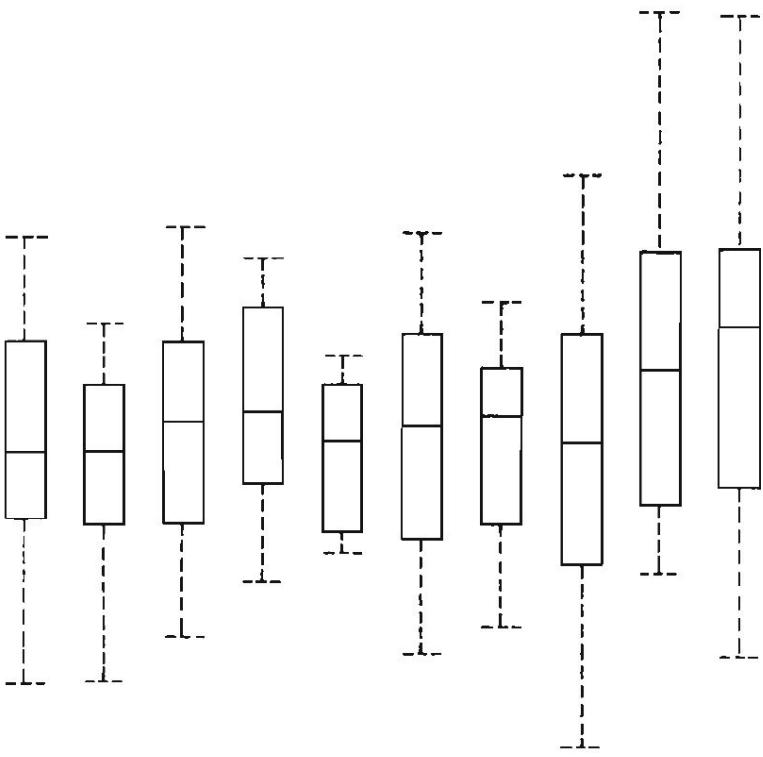


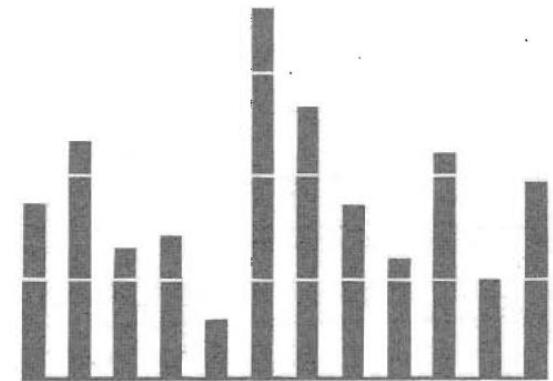
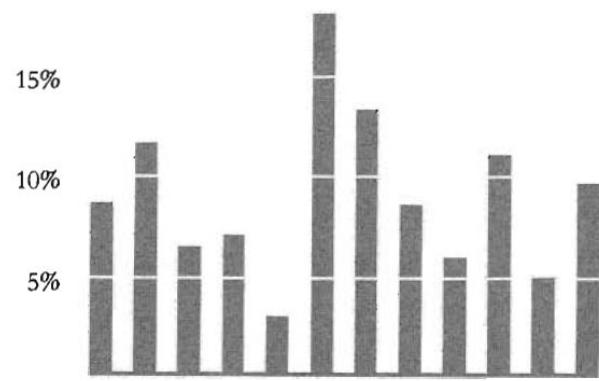
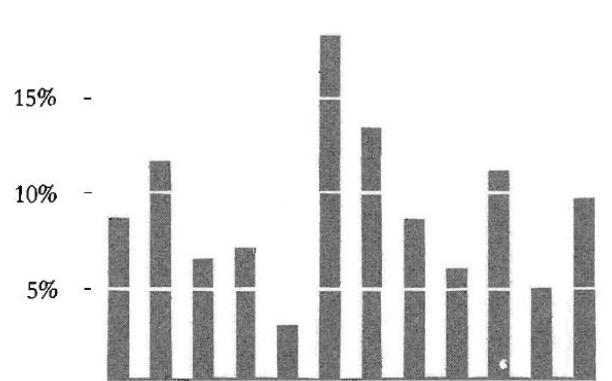
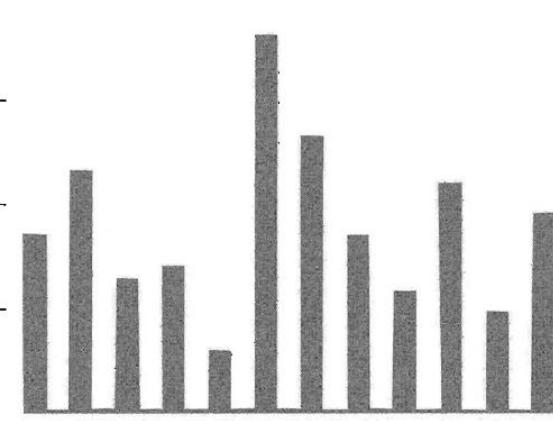
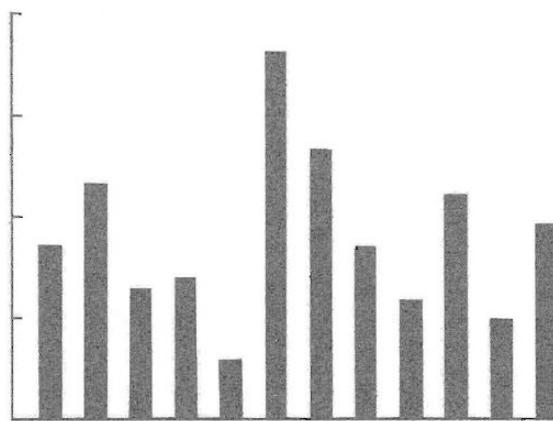
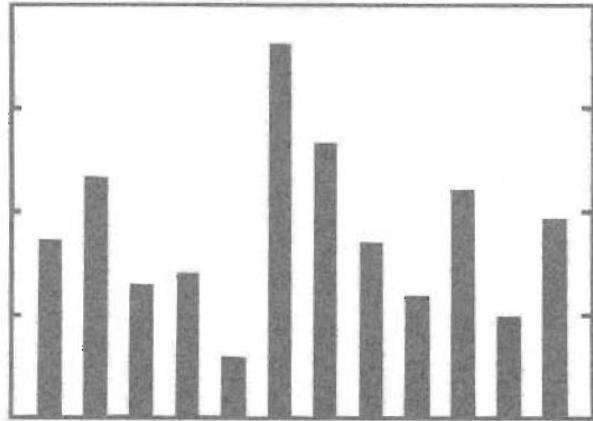
Duck





Graphical Design





**Efficiency, Complexity, Structure,
Density, Beauty.**

Multifunctioning Graphical Elements

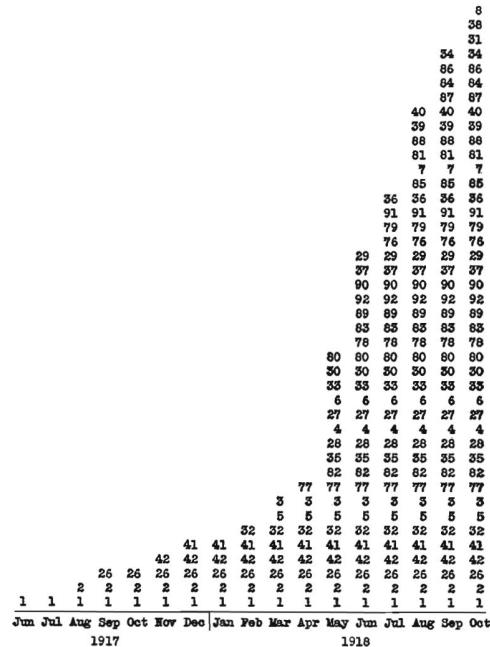
0 | 9 = 900 feet

0	98766562
1	97719630
2	69987766544422211009850
3	876655412099551426
4	9998844331929433361107
5	97666666554422210097731
6	898665441077761065
7	98855431100652108073
8	653322122937
9	377655421000493
10	0984433165212
11	4963201631
12	45421164
13	47830
14	00
15	676
16	52
17	92
18	5
19	39730

Stem-and-leaf displays:

heights of 218 volcanoes, unit 100 feet.

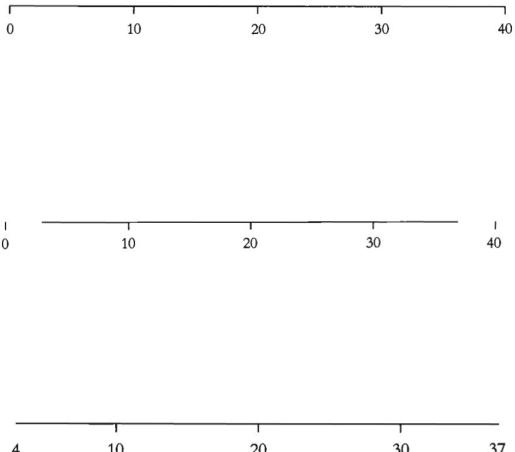
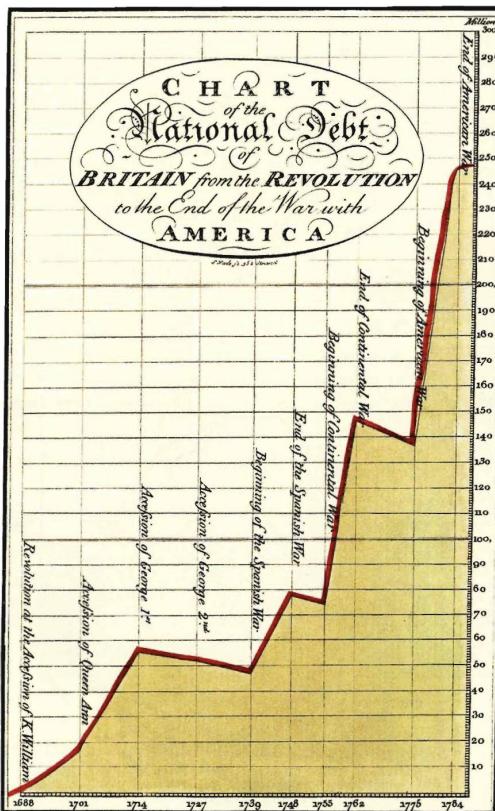
19 | 3 = 19,300 feet



Mobilize every graphical element, perhaps several times over, to show the data

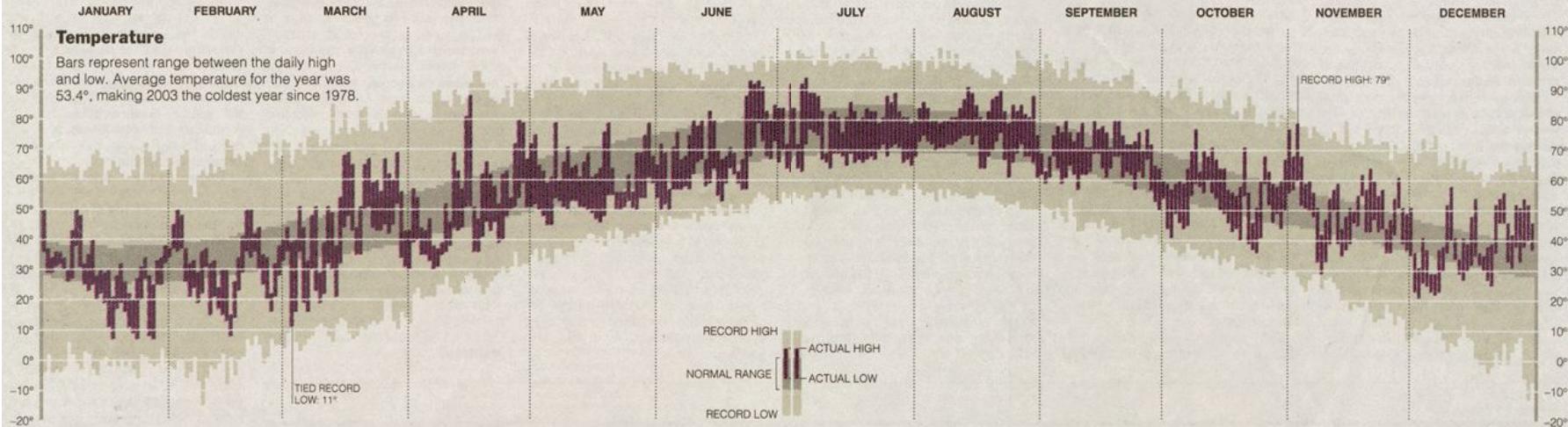


Central to maintaining clarity in the face of the complex are graphical methods that organize and order the flow of graphical information presented to the eye.

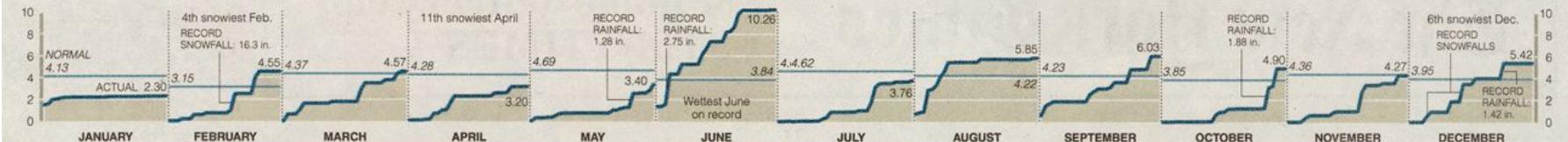


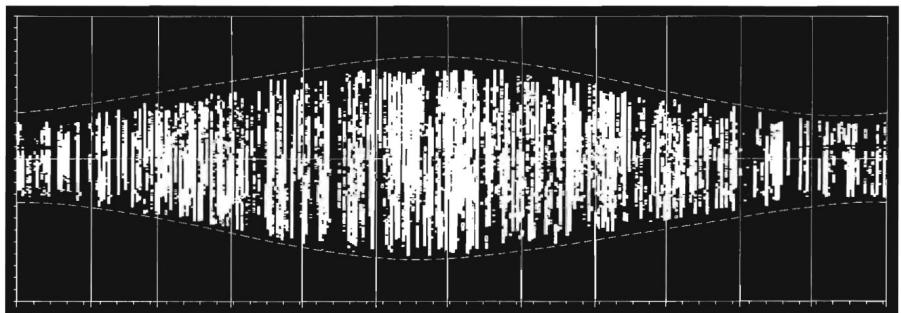
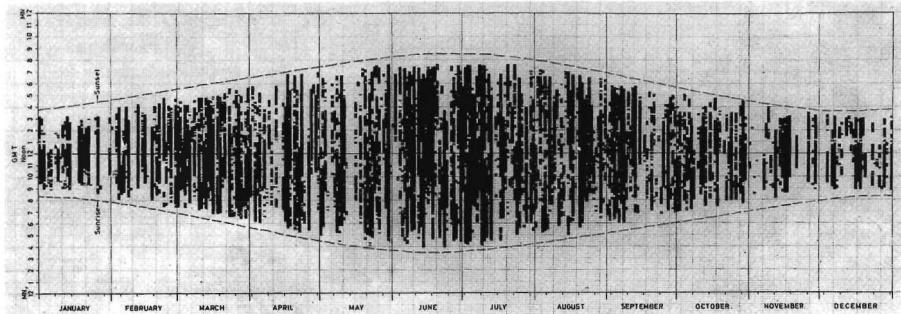
High Resolution Data Graphics

New York City's Weather in 2003

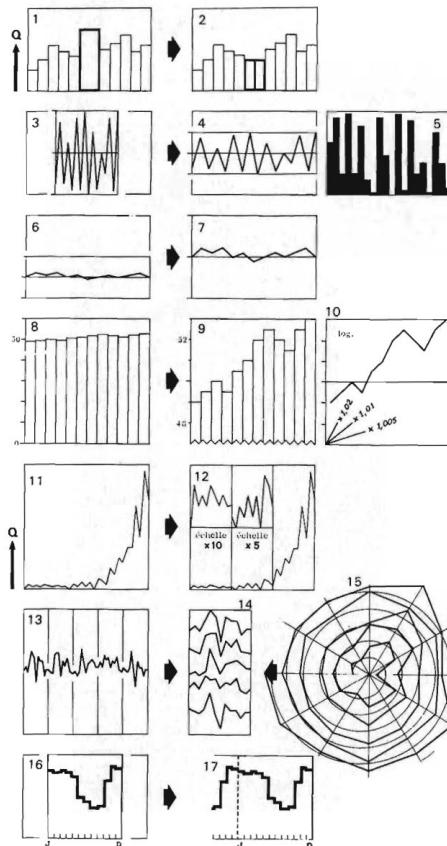


Precipitation Cumulative monthly precipitation in inches compared with normal monthly precipitation. Total precipitation in 2003 was 58.51 inches, 8.82 inches more than normal, which makes the year the sixth wettest on record.





Maximize data density and the size of the data matrix within reason while exploiting the maximum resolution of the available data-display technology.



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.

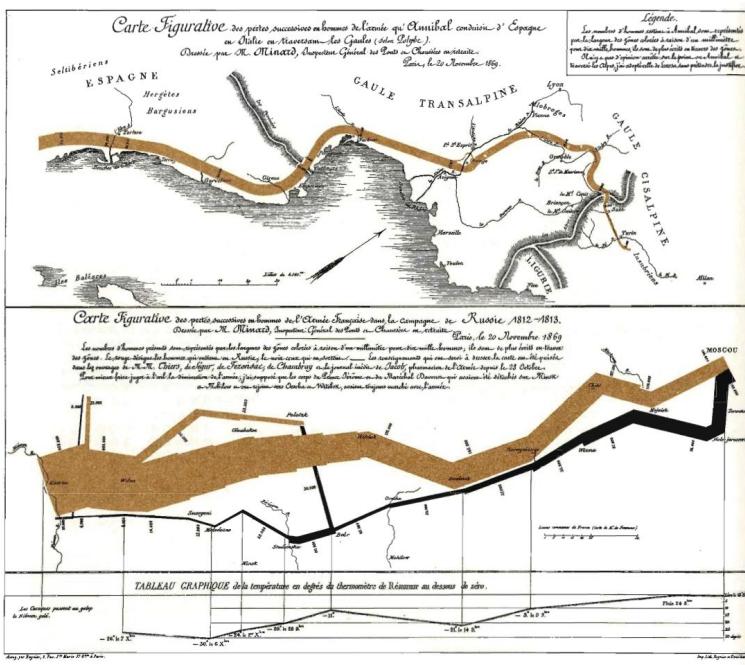
Courbes annuelles de pluie ou de température. Un cycle possède deux phases (17), pourra-t-il en offrir qu'une à la perception du spectateur ? (16).



For non-data-ink, less is more

For data-ink, less is a bore.

Aesthetics and Technique in Data Graphical Design



Good design has two key elements: Graphical elegance is often found in simplicity of design and complexity of data.

How Different Groups Voted for President

Based on 12,700 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%)	68	26	6	77-22
Conservatives (23%)	30	54	12	43-44
Moderates (30%)	11	34	4	3-54
Liberals (17%)	57	27	11	70-26
Moderates (46%)	42	48	6	51-48
Conservatives (26%)	23	71	4	20-70
Liberals (20%)	70	14	13	86-12
Moderate Democrats (12%)	68	28	6	77-22
Conservative Democrats (8%)	53	41	4	64-35
Political independents (13%)	72	19	6	—
Democrats favoring Kennedy in primaries (13%)	66	24	8	—
Liberal Republicans (4%)	60	29	15	84-29
Moderate Independents (12%)	21	53	13	45-33
Conservative Independents (7%)	22	69	6	26-72
Liberal Republicans (2%)	25	66	9	17-82
Moderate Republicans (13%)	13	81	5	18-80
Conservative Republicans (12%)	6	91	2	6-93
Politically active Republicans (2%)	5	89	6	—
East (22%)	43	47	8	51-47
South (27%)	43	51	3	54-45
Midwest (20%)	41	51	6	48-50
West (31%)	35	55	10	51-49
Blacks (10%)	62	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 (54%)	54	32	11	—
Female, opposes equal rights amendment (45%)	29	66	4	—
Catholic (25%)	40	51	7	54-44
Jewish (4%)	49	39	14	54-44
Protestant (46%)	37	58	6	44-55
Born again white Protestant (17%)	34	61	4	—
18-21 years old (15%)	44	43	11	48-44
22-39 years old (77%)	43	43	11	41-46
30-44 years old (31%)	37	54	7	49-49
45-59 years old (32%)	39	55	6	42-52
60 years or older (18%)	40	54	4	47-52
Family income				
Less than \$10,000 (15%)	50	41	9	58-40
\$10,000-\$19,999 (14%)	47	42	8	54-43
\$15,000-\$24,999 (30%)	38	53	7	48-50
\$25,000-\$50,000 (24%)	32	58	6	36-62
Over \$50,000 (5%)	25	55	8	—
Professional or manager (40%)	33	56	9	41-57
Clerical, sales or other white-collar workers (27%)	42	48	6	46-53
Blue-collar workers (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 (35%)	35	55	8	49-50
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 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
Farmers and agricultural party				
Democrats, better off than a year ago (7%)	77	16	6	69-31
Democrats, same (7%)	47	39	10	94-6
Independents, better off (1%)	46	36	12	—
Independents, same (1%)	21	65	11	—
Republicans, better off (4%)	18	77	5	3-97
Republicans, same (11%)	6	89	4	24-76
More important problems				
Unemployment (38%)	51	40	7	75-25
Inflation (44%)	30	60	9	35-65
Feel that U.S. should be more involved in dealing with world. Even if it would increase the risk of war (54%)	28	64	6	—
Disagree (31%)	56	32	10	—
Pay equal pay amendment (46%)	49	38	11	—
Oppose equal pay amendment (35%)	28	68	4	—
When decided about choice				
Know all along (41%)	47	50	2	44-45
During primaries (13%)	30	69	8	42-52
During conventions (8%)	36	55	7	51-48
Since Labor Day (8%)	30	54	13	49-49
In week before election (23%)	28	46	13	49-47

Some Winners and Losers in the Forecasting Game

**Council of Economic
Advisors:** +4.7%

**Data
Resources:** +4.5%

**Nat. Assoc. of Business
Economists:** +4.5%

**Wharton Econometric
Forecasting:** +4.5%

**Congressional Budget
Office:** +4.4%

**Conference
Board:** +4.2%

**I.B.M. Economics
Department:** +4.1%

**Real G.N.P.
Growth:** +3.8%
**Chase
Econometrics:** +2.8%

About a year ago, eight forecasters were asked for
their predictions on some key economic indicators.
Here's how the forecasts stack up against the
probable 1978 results (shown in the black panel).

**Industrial Production
Growth:** +5.8%
**Conference
Board:** +5.5%

**Change in Consumer
Prices:** +7.7%
**I.B.M. Economics
Department:** +6.8%

**Corporate Profits
Growth:** +13.3%
**Data
Resources:** +10.5%

**Wharton Econometric
Forecasting:** +21%
**Nat. Assoc. of Business
Economists:** +6.5%

**Unemployment
Rate:** 6%
**I.B.M. Economics
Department:** +10.4%

**Chase
Econometrics:** +4.8%
**Conference
Board:** +6.2%

**Chase
Econometrics:** +1.9%
**Data
Resources:** +6.2%

**Chase
Econometrics:** +5.9%
**Chase
Econometrics:** +6.5%

**Council of Economic
Advisors:** +5.9%
**Wharton Econometric
Forecasting:** +5.4%

**Chase
Econometrics:** 7.4%

**Wharton Econometric
Forecasting:** 6.8%

**Conference
Board:** 6.7%

**Nat. Assoc. of Business
Economists:** 6.7%

**I.B.M. Economics
Department:** 6.6%

**Data
Resources:** 6.5%

**Congressional Budget
Office:** 6.3%

**Chase
Econometrics:** 6.3%

Forecasters are not listed
in categories for which they
did not make a prediction.

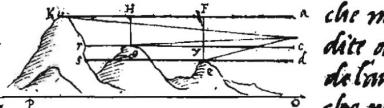
*After taxes

234.

cherai le cose vedute eforse tanto misteriose che no[n]che
le membra ma il tutto quasi rifara impossibile a po-
tre figurare Come sull'occhio fasse, o, c'la busa dun
guarro di braccio eguale alla tua trauola dipinta
sia, a., b., discosta . m. dal occhio mezo
braccio allorav tu  uedrai per esofar
cio tutte le cose che . 6 a. ueder si potessi dim-
tro alla lungheza . o. d'uno orizonte di
cento miglia intanua confusa diminutio[n]e che no[n]
che figurav di quelle alcuna parte c'habbia figura
ma apena potrai porre si piccolo punto di penello che
non sia maggiore c'hogm gran casamento posto in
diua miglia di distan[za].

*Perche li monti in langa distan[za]
si dimostrano piu scuri nella cima*

che nella basa -

Lavia c'acquista gradi di grossezza in ogni grado de
la sua basetta e della sua distan[za] e causa che le
cime de monti che piu s'inalzano piu mostrano la
sua natura per
sono im-
grossesse
nella cima 

Friendly

words are spelled out, mysterious and elaborate encoding avoided

words run from left to right, the usual direction for reading occidental languages

little messages help explain data

elaborately encoded shadings, cross-hatching, and colors are avoided; instead, labels are placed on the graphic itself; no legend is required

graphic attracts viewer, provokes curiosity

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)

type is clear, precise, modest; lettering may be done by hand

type is upper-and-lower case, with serifs

Unfriendly

abbreviations abound, requiring the viewer to sort through text to decode abbreviations

words run vertically, particularly along the Y-axis; words run in several different directions

graphic is cryptic, requires repeated references to scattered text

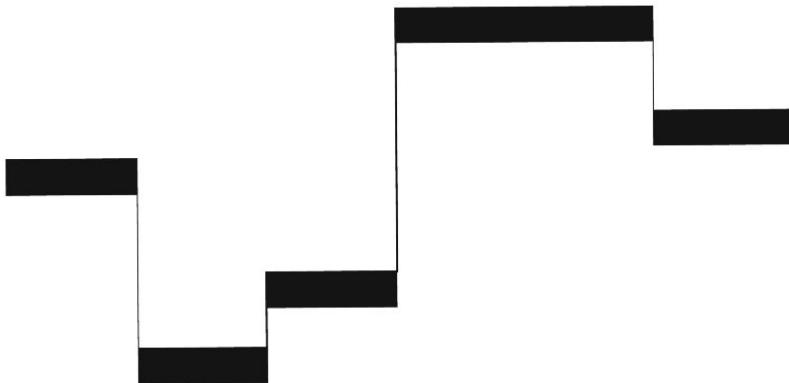
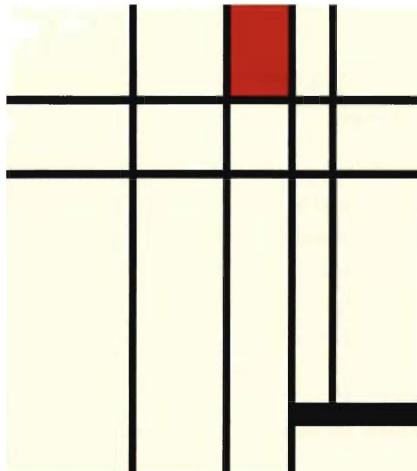
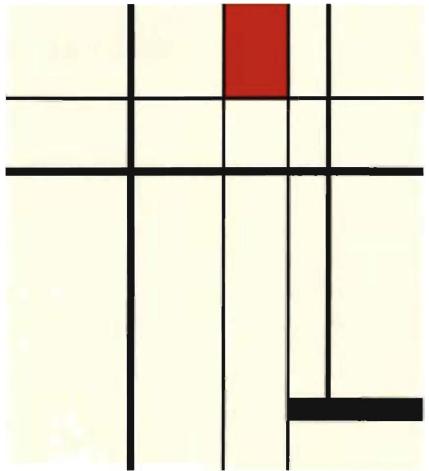
obscure codings require going back and forth between legend and graphic

graphic is repellent, filled with chartjunk

design insensitive to color-deficient viewers; red and green used for essential contrasts

type is clotted, overbearing

type is all capitals, sans serif



lesser height

greater length

some labels

some labels

instead of

some other labels

some other labels

effect

cause



Thank You!



Questions?