Graphical Excellence—Edward Tufte

Graphical Excellence—Edward Tufte

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency.

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency.

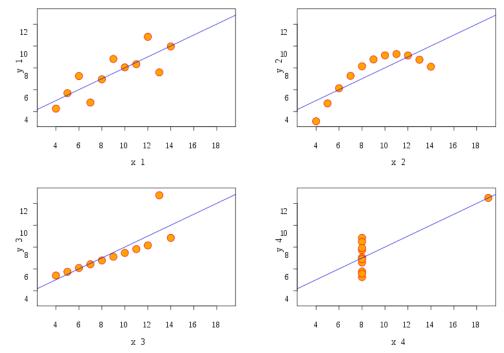
- Graphical displays should:
- show the data
- induce the viewer to think about the substance
- avoid distorting what the data says
- present many numbers in small space
- make large data sets coherent
- encourage comparison between data
- reveal the data at several levels of detail
- clear purpose: description, exploration, tabulation or decoration

Graphics reveal data.

Graphics reveal data.

I		II				IV	
x	у	x	У	x	у	x	у
10,00	8,04	10,00	9,14	10,00	7,46	8,00	6,58
8,00	6,95	8,00	8,14	8,00	6,77	8,00	5,76
13,00	7,58	13,00	8,74	13,00	12,74	8,00	7,71
9,00	8,81	9,00	8,77	9,00	7,11	8,00	8,84
11,00	8,33	11,00	9,26	11,00	7,81	8,00	8,47
14,00	9,96	14,00	8,10	14,00	8,84	8,00	7,04
6,00	7,24	6,00	6,13	6,00	6,08	8,00	5,25
4,00	4,26	4,00	3,10	4,00	5,39	19,00	12,50
12,00	10,84	12,00	9,13	12,00	8,15	8,00	5,56
7,00	4,82	7,00	7,26	7,00	6,42	8,00	7,91
5,00	5,68	5,00	4,74	5,00	5,73	8,00	6,89

Anscombe's quartet



They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing.

All four sets are identical when examined using simple summary statistics, but vary considerably when graphed.

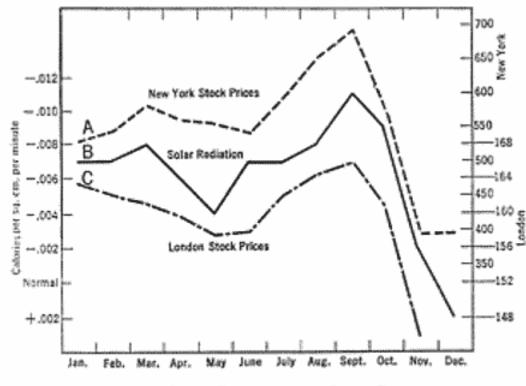
Statistical graphics are only as good as what goes into them.

Statistical graphics are only as good as what goes into them.

A- New York Stock prices

B- Solar Radiation inverted C- London Stock prices

An ill-specified model cannot be rescued by a graphic. No matter how clever or fancy they are.



SOLAR RADIATION AND STOCK PRICES

A New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner). Graphical Excellence =

efficient communication of complex quantitative ideas

- Fundamental Graphical Designs:
 - Data Maps
 - Time-Series
 - Space-time narrative designs
 - Relational Graphics

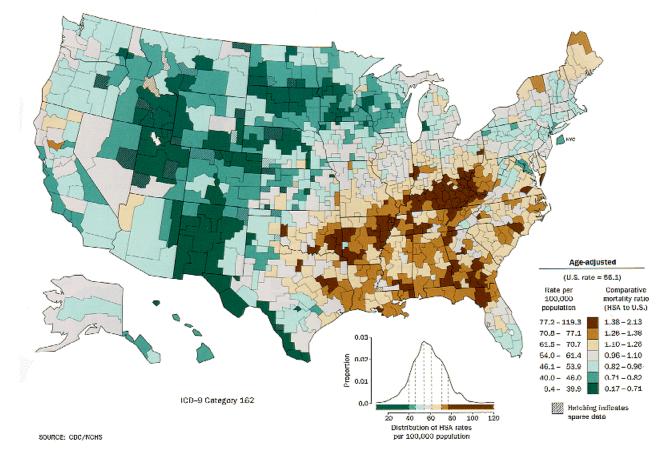
• Fundamental Graphical Designs:

Data Maps

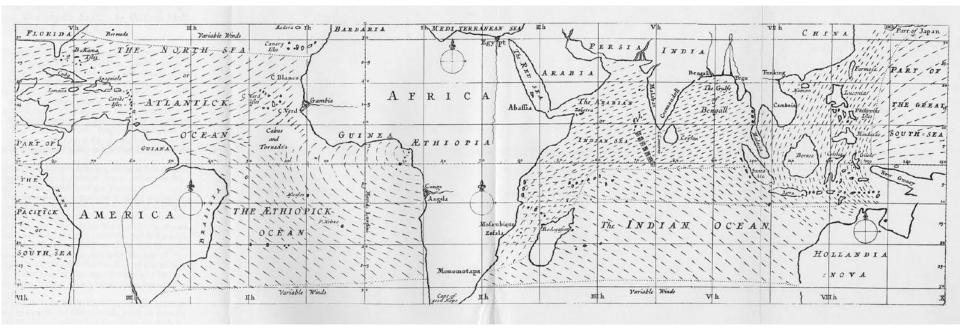
- Time-Series
- Space-time narrative designs
- Relational Graphics

- Picture- a thousand words
- Only a picture can carry such a volume of data in such a small space

- Picture- a thousand words
- Only a picture can carry such a volume of data in such a small space



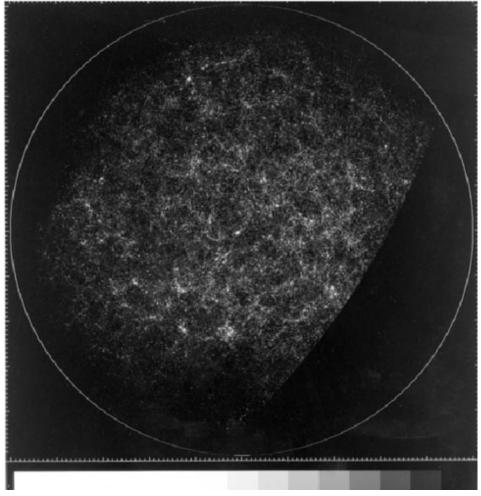
• Edmond Halley, 1686



• Dr. John Shaw, 1854

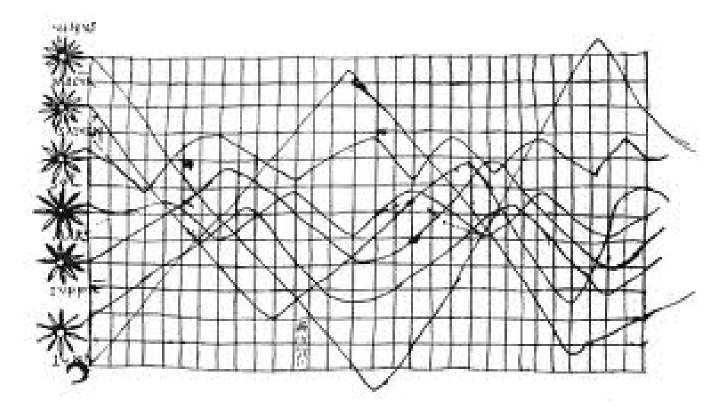


• "New Reduction of the Lick Catalog of Galaxies" - Seldner, Siebers, Groth and Peebles

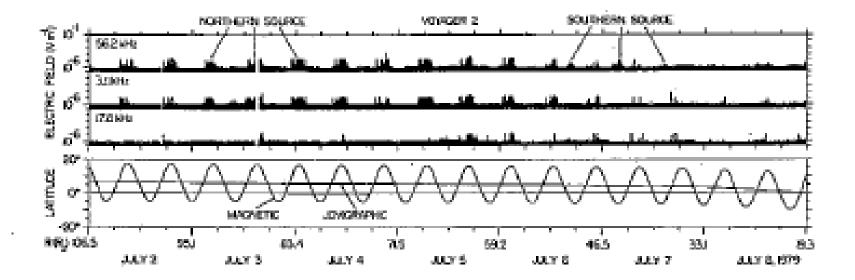


- Fundamental Graphical Designs:
 - Data Maps
 - Time-Series
 - Space-time narrative designs
 - Relational Graphics

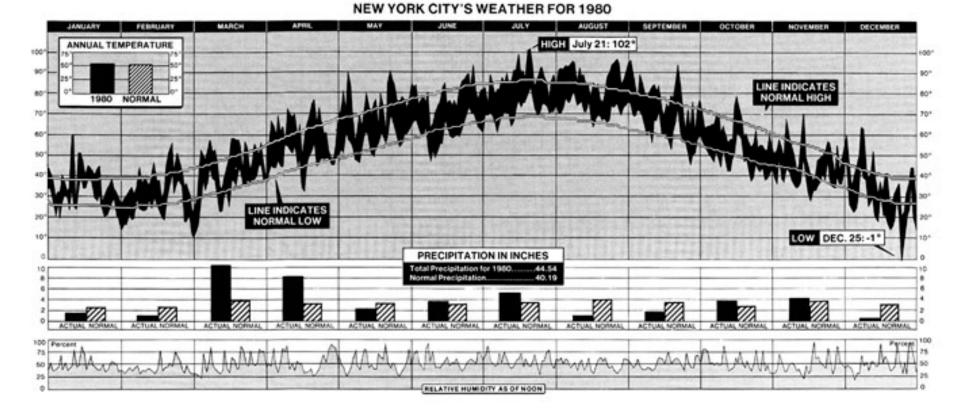
• "A Note on a Tenth Century Graph" - H. Gray Funkhouser

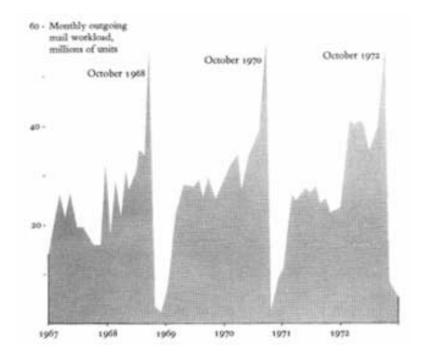


• "Plasma Wave Observation Near Jupiter"- Gurnett, Kurth and Scarf



• New York Weather 1980





FRANKED MAIL TIE TO VOTING SHOWN

sator John G. Towar, Rev.

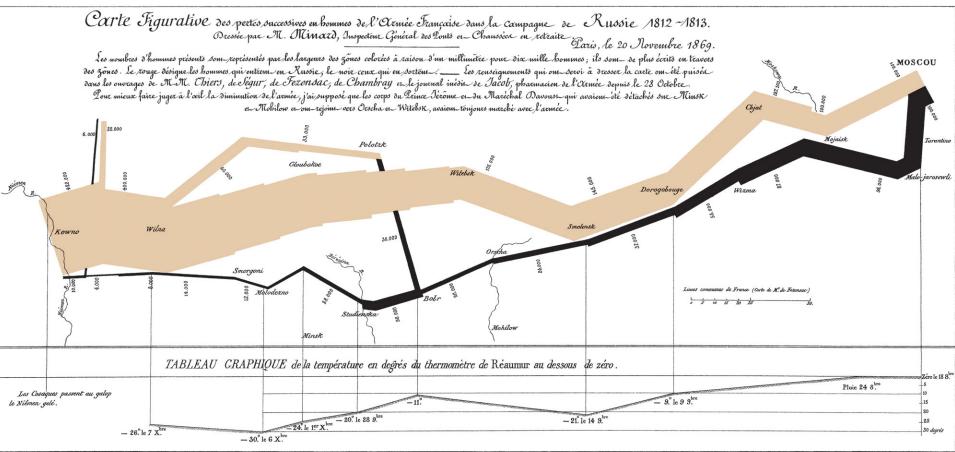
Provine Descents Testingung Finds the Volume and of Sa (202 modeline) forently, foreign page. Switch has der Soldan has the printed Even Before Dettines "Start and method company and toth of finding mail has well documented as to record and antiplied. A new federal to a minimum and decomment from all and what what aphotofice is a federal four for frames in request. And a fair what and after it a restore that the train which what and after the billy great which which

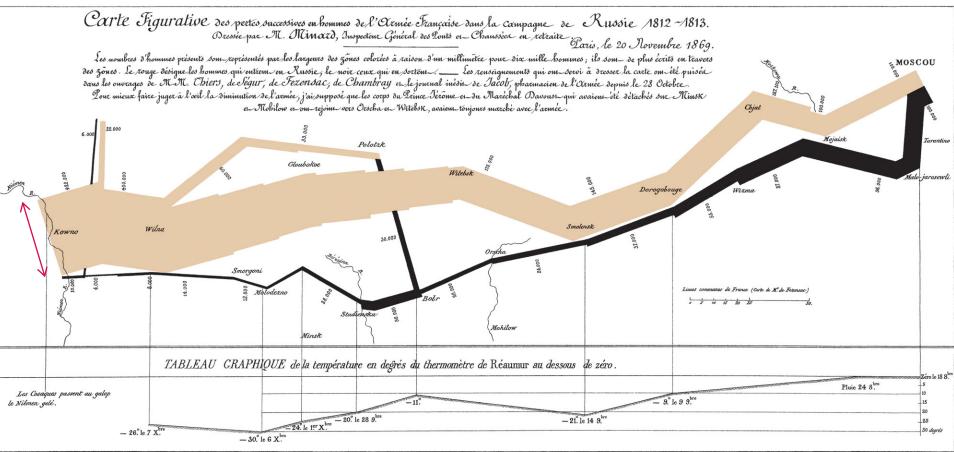
ine services from an an integral includes of the executive and include an optimize in the full links of finance Proof in first includes and in the service and the change of the free particular, which he change a complements of publical wat to require the old post. Admitted products made spectral arrandom int five W. Blackbroger, wonie i

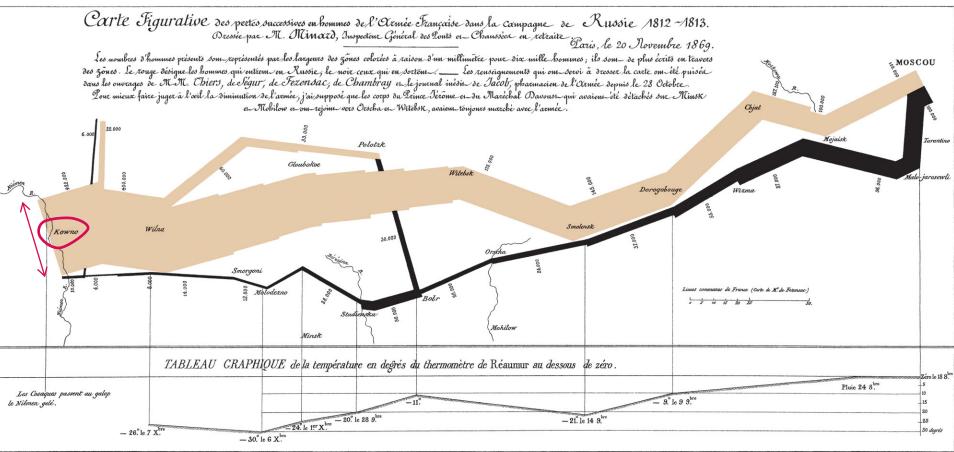
You Lorffler, a high-muthing a proposal for the use of range and, arone to a new dwated mail to be chief, serve organizes dated (b). 21 linest Danette, int 20076. "Darts, is 2015. (1712, that fairing the tanganget involved and program and by special interval house brinds

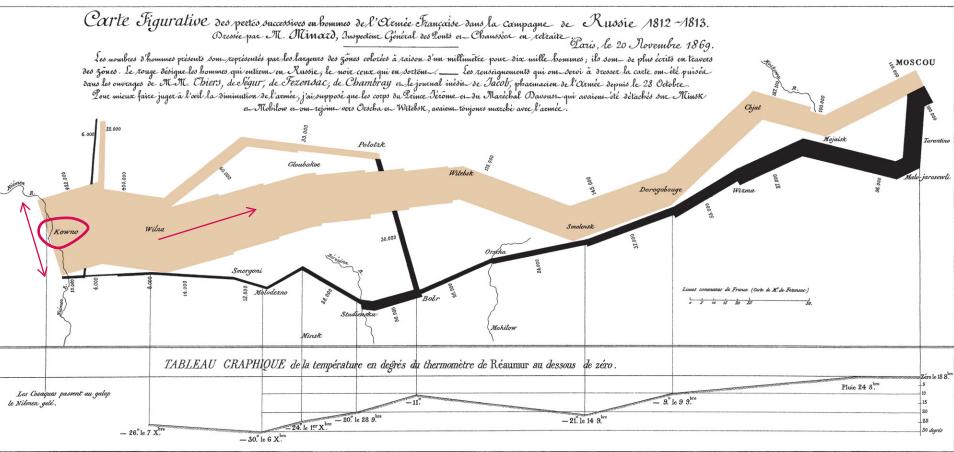
Alternative and how and shares a state a first state and state a state and state a state

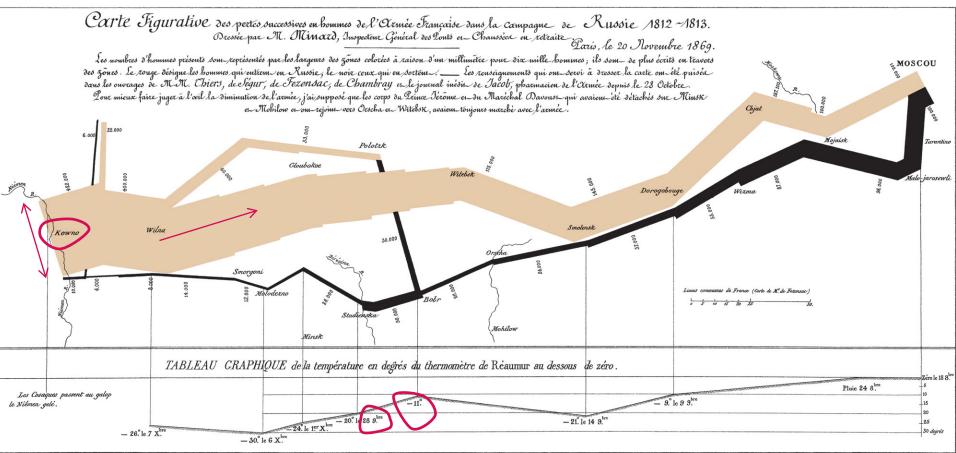
- Fundamental Graphical Designs:
 - Data Maps
 - Time-Series
 - Space-time narrative designs
 - Relational Graphics



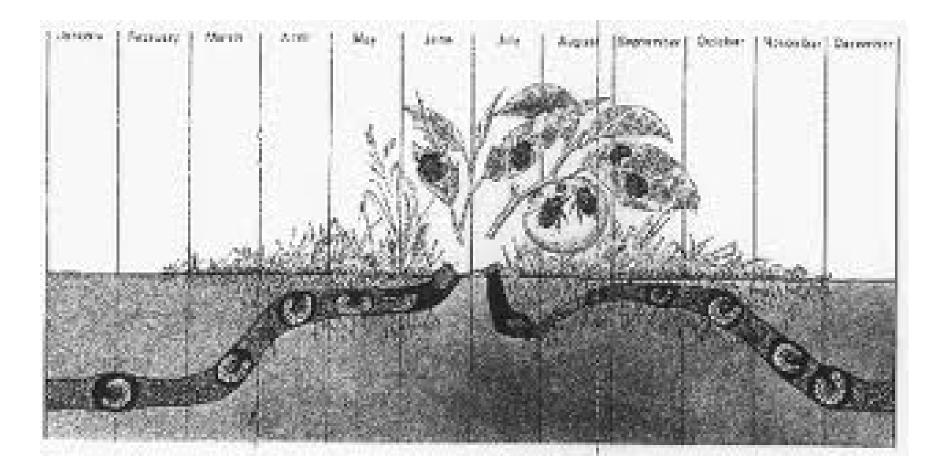






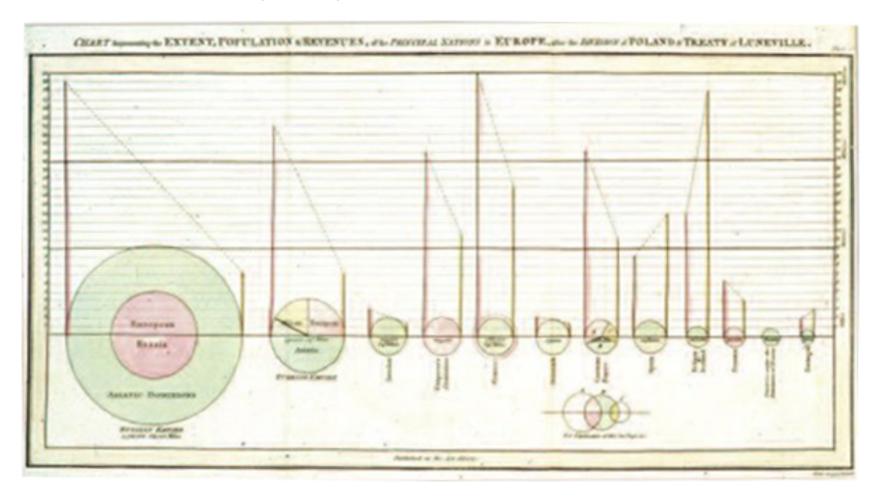


"Man and Insects"- I. Hugh Newman

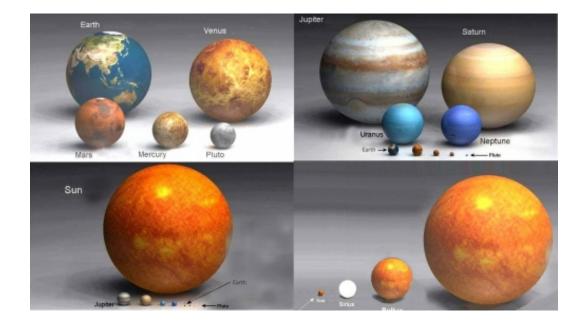


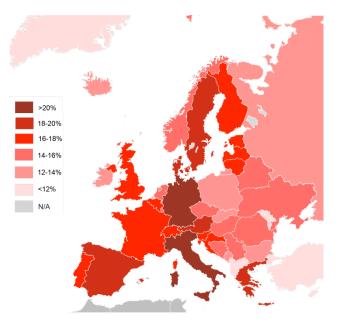
- Fundamental Graphical Designs:
 - Data Maps
 - Time-Series
 - Space-time narrative designs
 - Relational Graphics

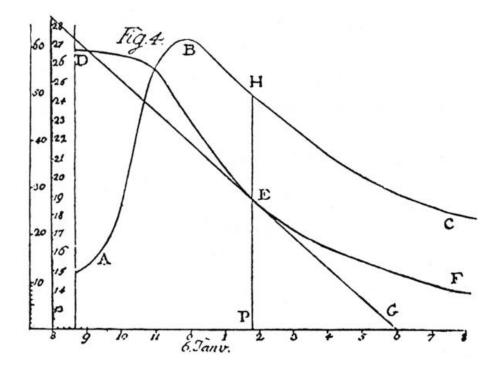
"The Statistical Breviary" - Playfair

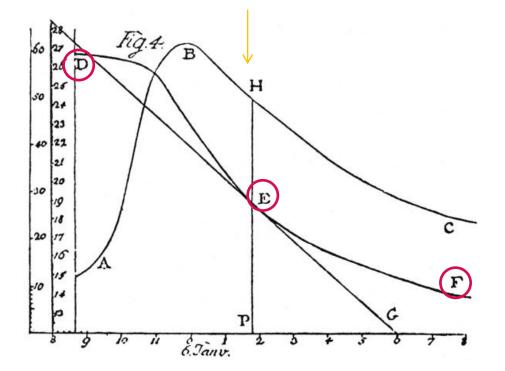


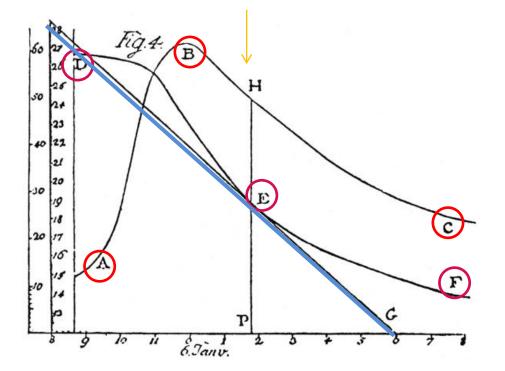
Comparing sizes represented with similar form.

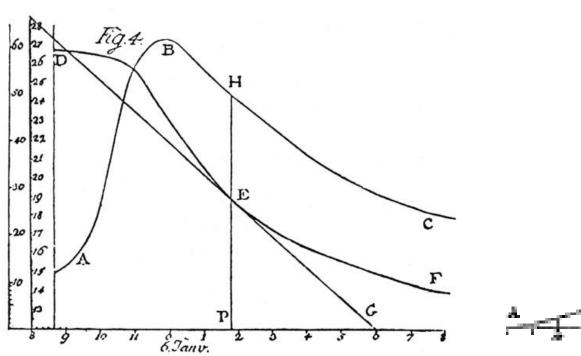


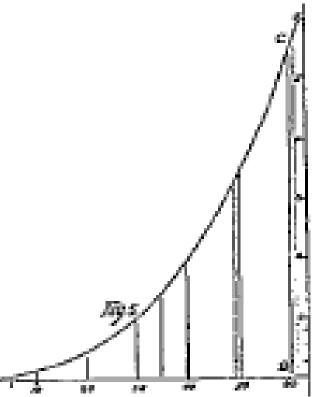












Graphical Excellence is the well-designed presentation of interesting data- a matter of *substance*, *statistics* and *design*.

Graphical Excellence is the well-designed presentation of interesting data- a matter of *substance*, *statistics* and *design*.

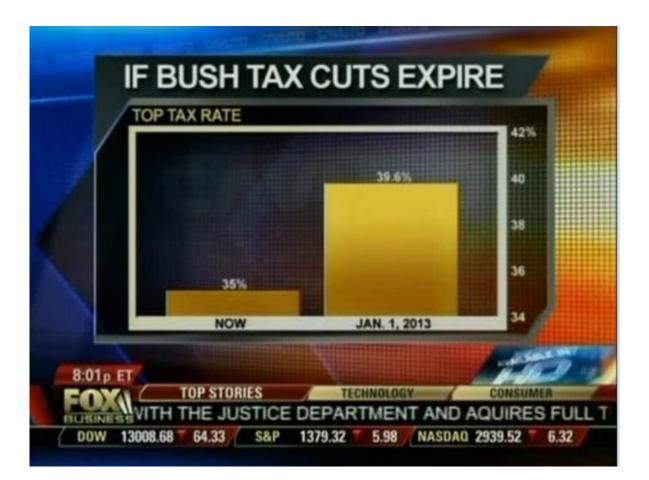
Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.

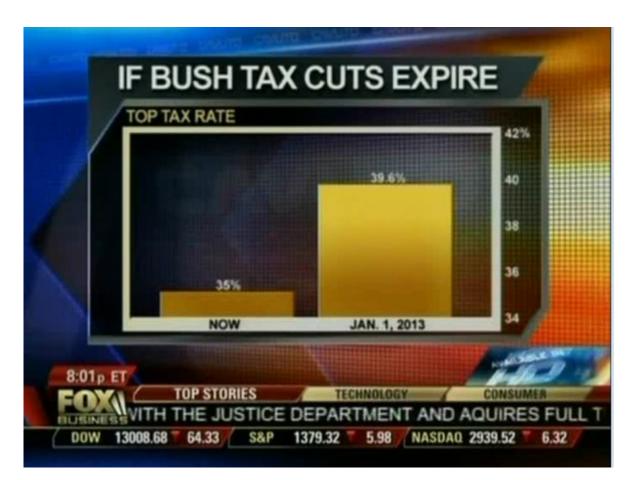
Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.

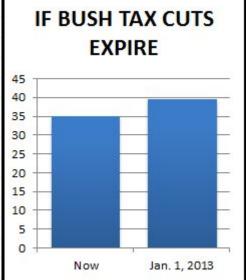
Graphical excellence requires telling the truth about data.

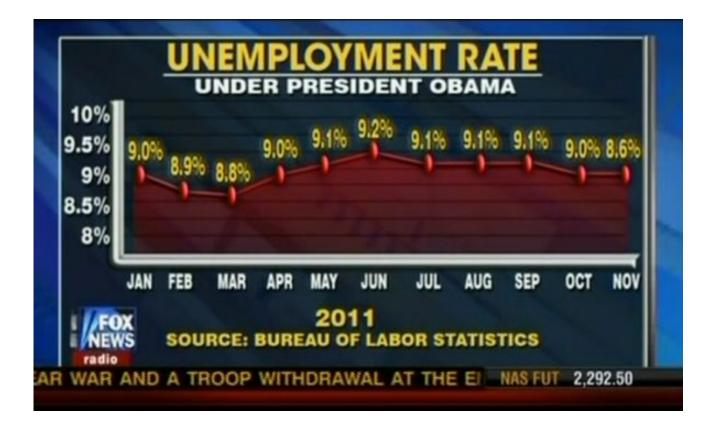
Graphical excellence requires telling the truth about data.

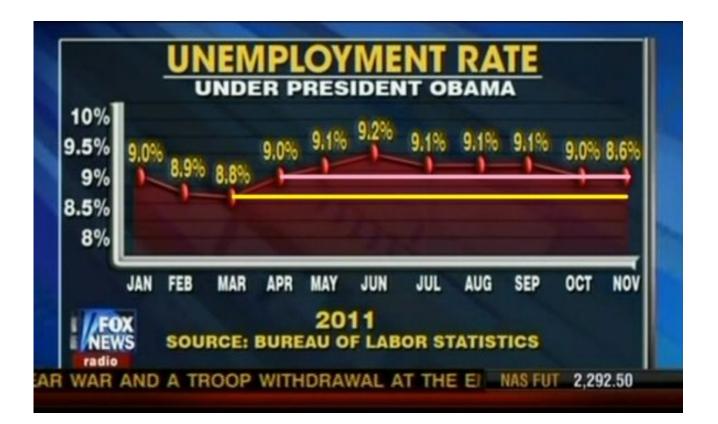
Graphical excellence requires telling the truth about data.









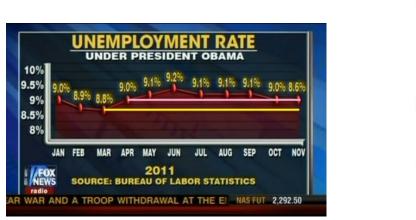


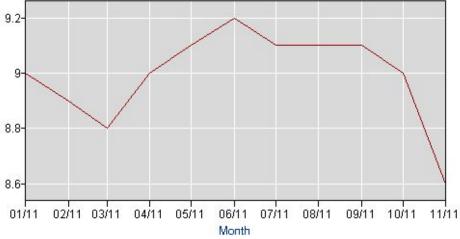
Misleading graphics

Data extracted on: December 12, 2011 (9:50:59 AM)

Labor Force Statistics from the Current Population Survey

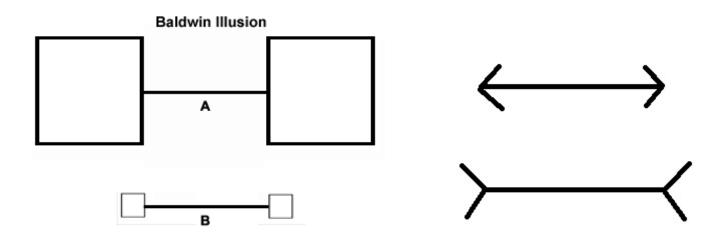
Series Id:	LNS14000000		
Seasonally Adjusted			
Series title:	(Seas) Unemployment Rate		
Labor force status:	Unemployment rate		
Type of data:	Percent or rate		
Age:	16 years and over		





A graphic does not distort if the visual representation of data is consistent with the numerical representation.

Visual representation of data \neq numerical representation



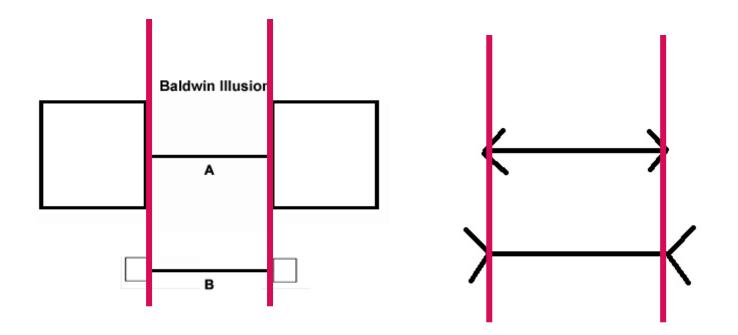


Table vs. Graph

Before: Table

Trust	Deprivation	Borough	Amount	Numbe
enk 1	index rank 3	Tower Hamlets	epproved £9,692,642	 2
2	1	Hackney	£7,809,808	
<u>-</u> }	12	Southwark	£7,266,118	
, 1	14	Camden	£8,140,419	
• 5	4	Islington	£5,424,137	
1	8	Lambeth	25,257,941	-
, 7	2	Newham	\$5,217,075	
3	13	Hammersmith and Fulham	£4,085,708	
))	29	Merton	£3,656,112	
10	20	Croydon	€3,629,066	-
11	9	Lewisham	£3,537,049	
12	17	Westminster	£3,357,911	
13	15	Ealing	£3,057,709	
14	30	Bromley	£3,038,621	1
15	19	Kensington and Chelsea	£2,979,468	
16	11	Brent	£2,898,224	
17	10	Greenwich	£2,837,858	
18	24	Barnet	£2,796,587	
19	21	Wandsworth	£2,592,453	
20	5	Waltham Forest	£2,505,730	100
21	28	Sutton	£2,468,511	
22	18	Hounslow	£2,383,393	
23	7	Haringey	£2,380,290	
24	22	Redbridge	£2,285,173	
25	33	Richmond upon Thames	£2,249,983	-
28	23	Hillingdon	£2,181,566	
27	16	Enfield	£2,145,800	
28	6	Barking and Dagenham	£1,943,597	
29	25	Havering	£1,934,424	
30	28	Bexley	£1,631,415	
31	27	Harrow	£1,516,193	
12	31	Kingston upon Thames	£1,353,125	
13	32	City of London	£402,060	
		Several Inner Boroughs	£18,704,877	
		Several Outer Boroughs	£8,392,100	1
		Other combination	£28,568,830	5
		London-wide	£86,583,750	1,3

After: Grable

Total grant spend by London Borough September 1995 - March 2011

		Index	Number	
Borough	rank	rank	of grants	Amount approved (£)
Tower Hamlets	1	3	269	9.692.642
Hackney	2	2	225	7,809.608
Southwark	3	12	232	7.266.118
Camden	4	14	136	6,140,419
Islington	5	4	156	5.424.137
Lambeth	6	8	156	5.257.941
Newham	7	2	154	5.217.075
Hammersmith and Fulham	8	13	109	4 085 708
Merton	9	29	113	3.656.112
Croydon	10	20	127	3.629.066
Lewisham	11	9	144	3 537.049
Westminster	12	17	100	3.357.911
Ealing	13	15	84	3.057.709
Bromley	14	30	131	3.038.621
Kensington and Chelsea	15	19	74	2.979.468
Brent	16	11	85	2 898 224
Greenwich	17	10	87	2 837 658
Barnet	18	24	99	2,796,587
Wandsworth	19	21	89	2 592 453
Waltham Forest	20	5	131	2 505 730
Sutton	21	28	87	2 468 511
Hounslow	22	18	75	2 383 393
Haringey	23	7	101	2 360 290
Redbridge	24	22	75	2 285 173
Rechmond upon Thames	25	33	133	2 249 983
Hullingdon	26	23	103	2 181 566
Enfield	27	16	86	2 145 800
Barking and Dagenham	28	6	68	1.943.597
Havering	29	25	95	1.934.424
Bexley	30	26	103	1.631.415
Harrow	31	27	62	1.516 193
Kingston upon Thames	32	31	55	1.353 125
City of London	33	32	11	02,060
Several Additional Inner Bouroughs			481	18,704,677
Several Additional Outer Bor			164	6.392.100
Other			566	28,566,830
London-wide			1,214	86,583,750
Total			6,180	252,883,123

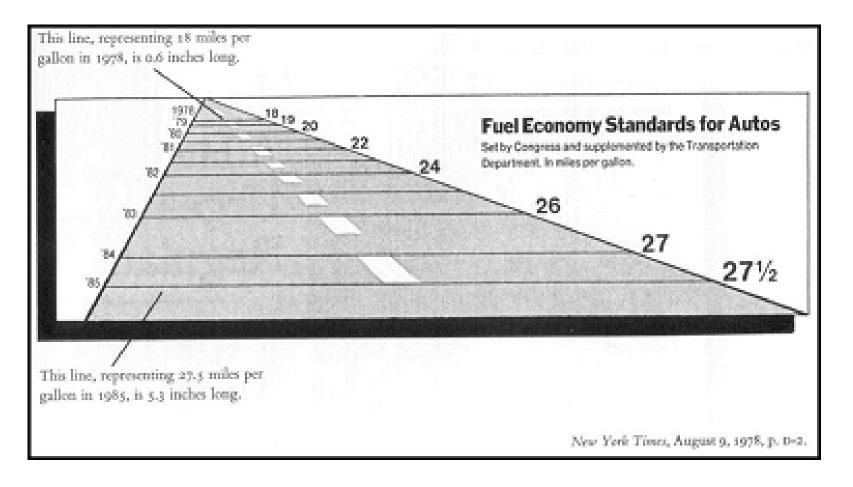
The power of graphics comes in the display of large data sets.

Two Principles:

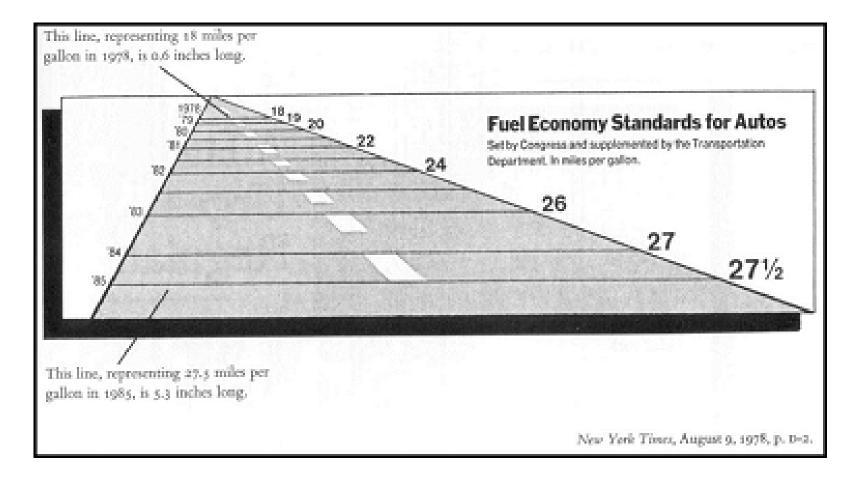
- Representation of numbers: the numbers should be directly proportional
- Clear, detailed and thorough labeling: write out explanations and label important events

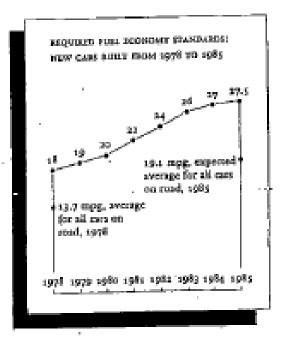
 $Lie \ Factor = \frac{size \ of \ effect \ shown \ in \ graphic}{size \ of \ effect \ in \ data}$

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

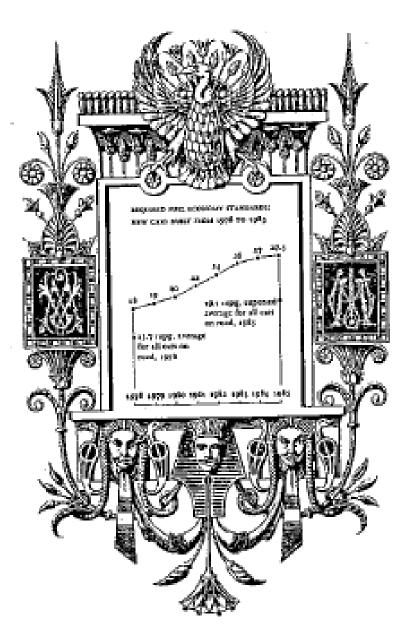


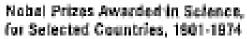
Lie Factor
$$= \frac{783}{53} = 14,8\%$$

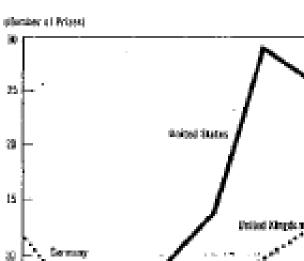




d.







Frank e

2243 -2628

1980-1940 1853-1970-1970 1974

[33]-1553

3

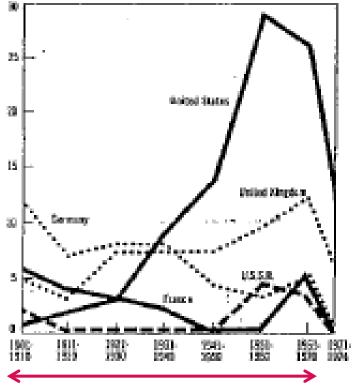
Q.,

 $\frac{1900}{1210}$

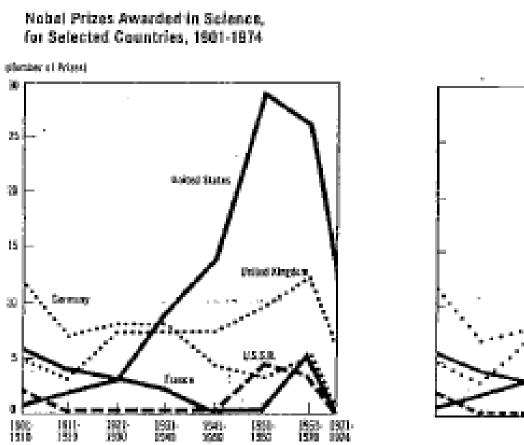
1811-1529 1923-1960

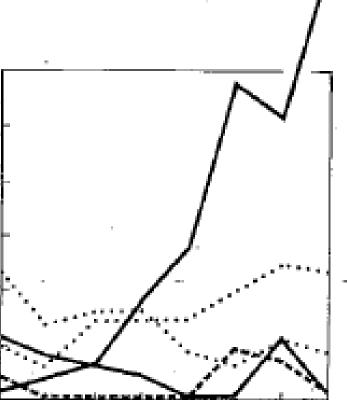


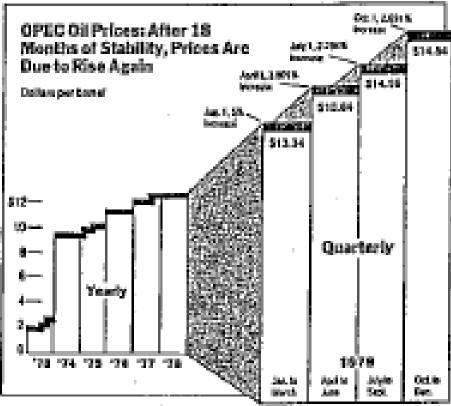




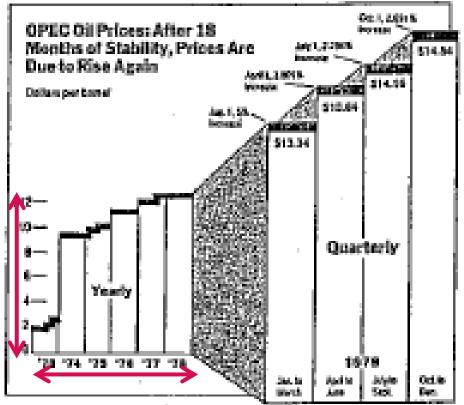
show data variation, not design variation





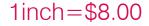


The Hern York Tarway Dec. 11, 1941

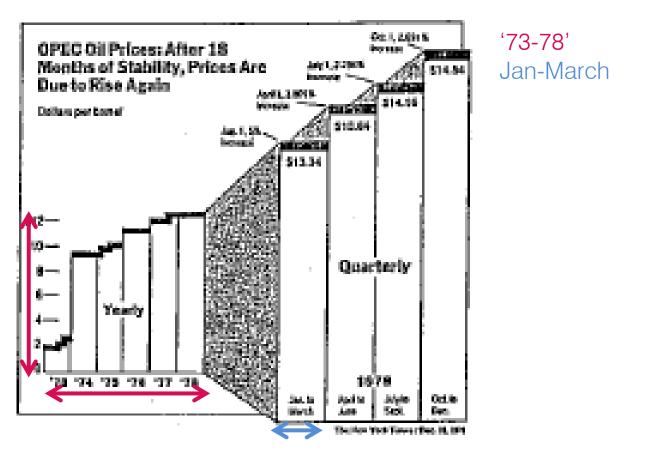


Prices:

'73-78'

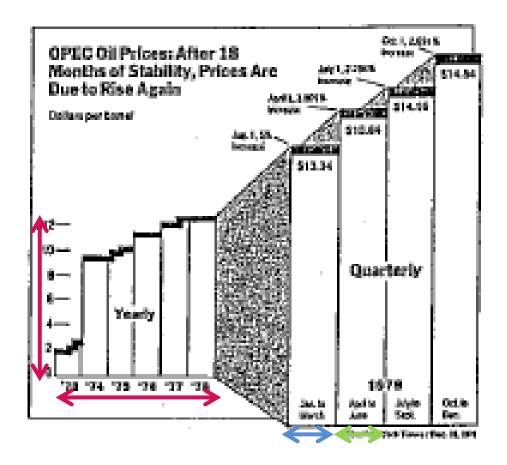






Prices:

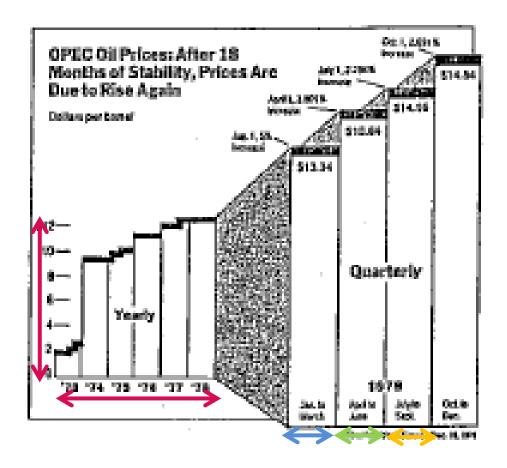
1inch=\$8.00 1inch=\$4.73



Prices:

'73-78' Jan-March Apr-June

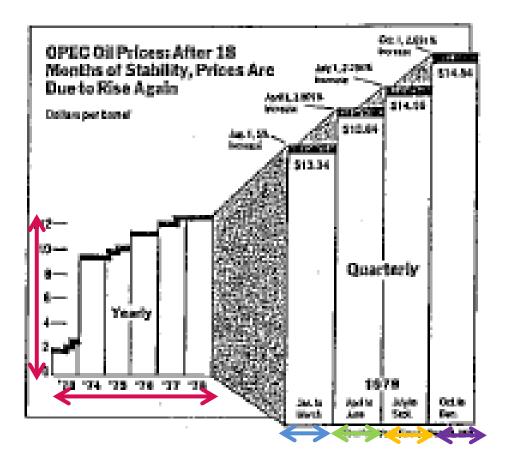
1inch=\$8.00 1inch=\$4.73 1inch=\$4.37





'73-78' Jan-March Apr-June

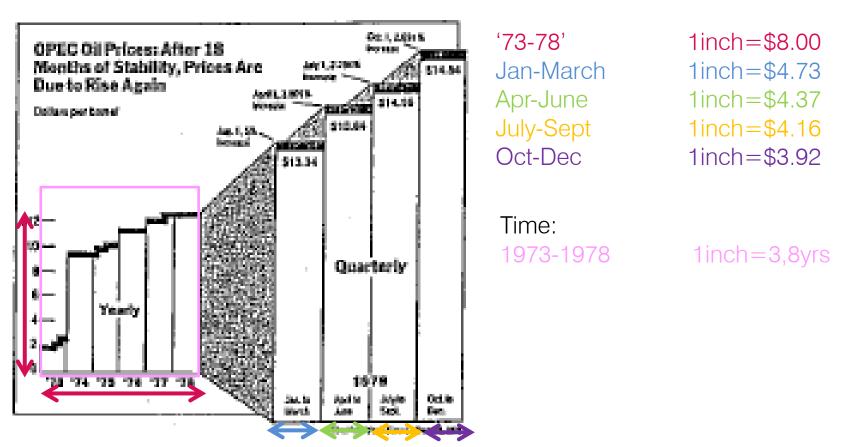
1inch=\$8.00 1inch=\$4.73 1inch=\$4.37 July-Sept 1inch=\$4.16



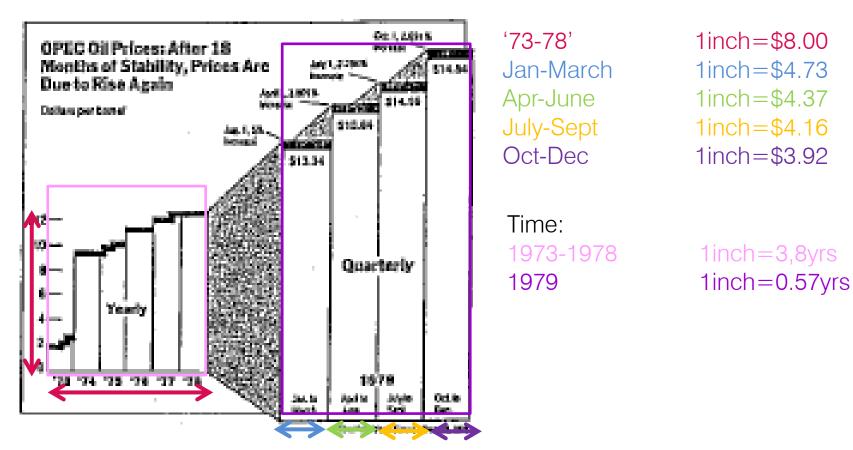


'73-78' Jan-March Apr-June July-Sept Oct-Dec

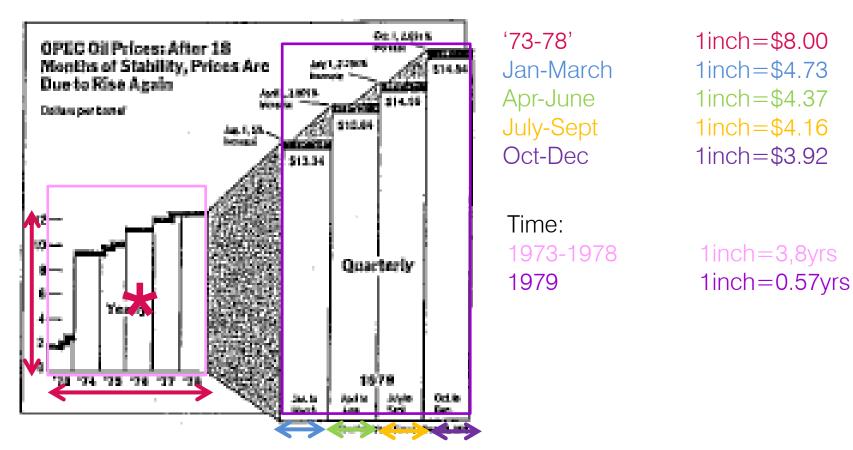
1inch=\$8.00 1inch=\$4.73 1inch=\$4.37 1inch=\$4.16 1inch=\$3.92



Prices:

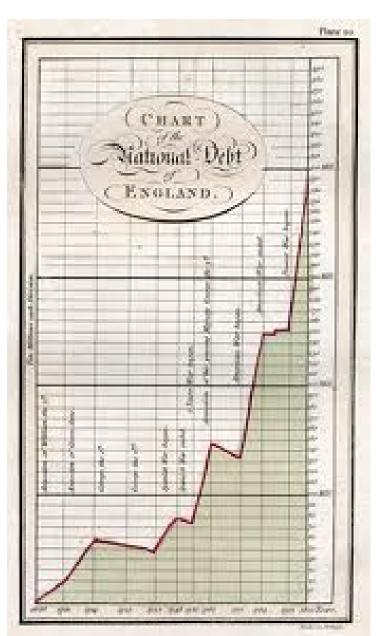


Prices:

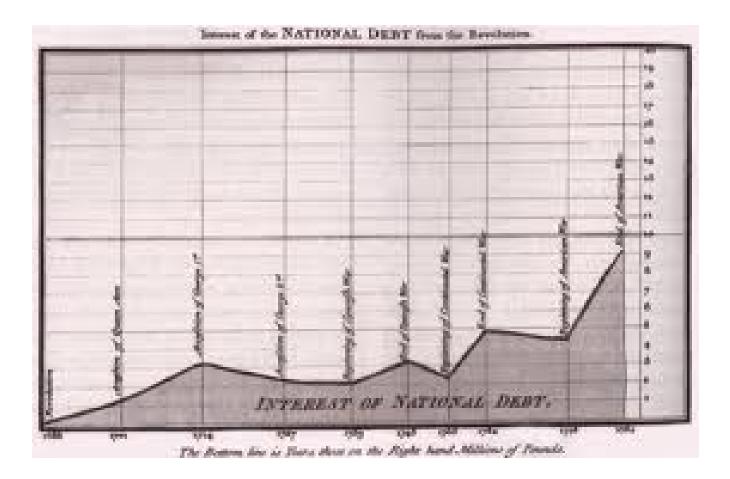


Prices:

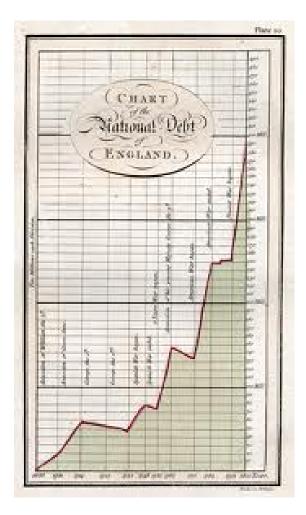
-Playfair, 1786

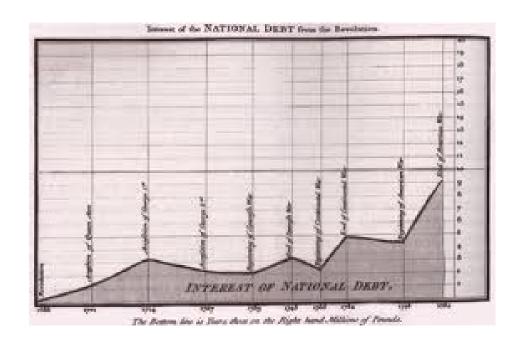


-Playfair, 1786



-Playfair, 1786

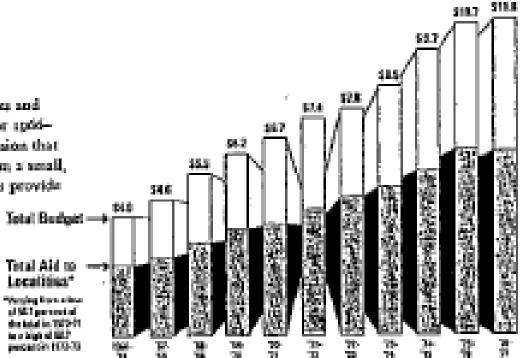




-New York Times, Feb 1, 1976

This closure of type emphasizes and structure out the low value for spid-1967, encouraging the impression that recent years have shot up from 2 small, stable base. Horizonesi acrows provide similar emphasis.

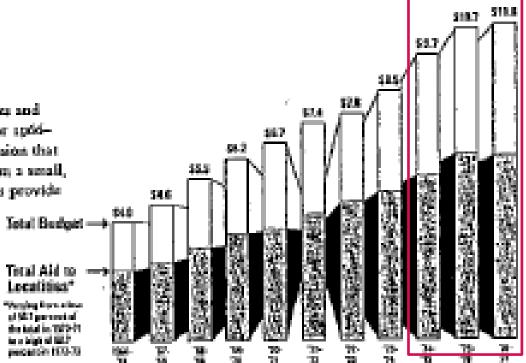
10



-New York Times, Feb 1, 1976

This closure of type emphasizes and structure out the low value for spid-1967, encouraging the impression that recent years have shot up from 2 small, stable base. Horizonesi acrows provide similar emphasis.

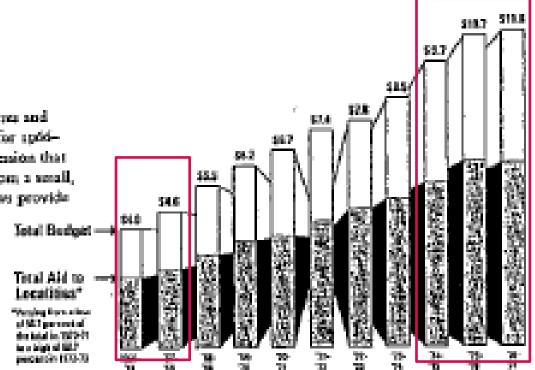
10



-New York Times, Feb 1, 1976

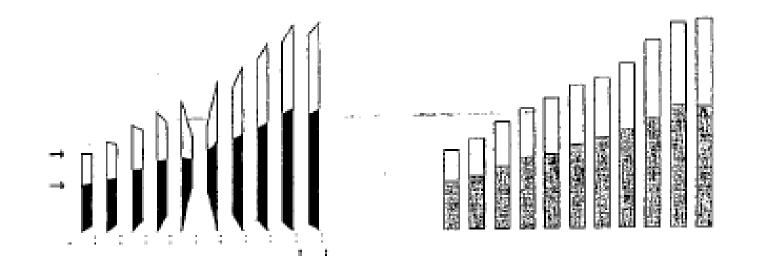
This closer of type emphasizes and structure out the low value for spid-1967, encouraging the impression that recent years have shot up from 2 small, stable base. Horizonesi acrows provide similar emphasis.

10



The Case of Skyrocketing Government Spending:

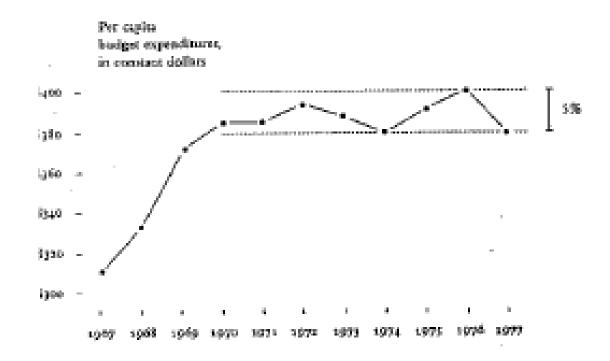
-New York Times, Feb 1, 1976



The Case of Skyrocketing Government Spending:

-New York Times, Feb 1, 1976

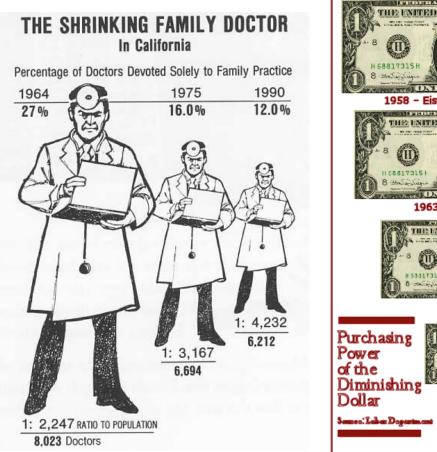
The Principle: In time-series of money, deflated and standardized units of monetary measurement are better than nominal units.



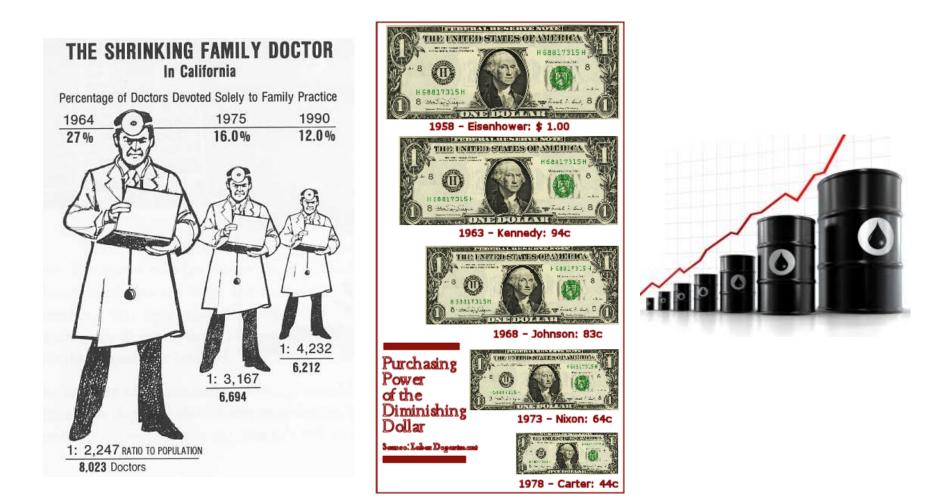
THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice



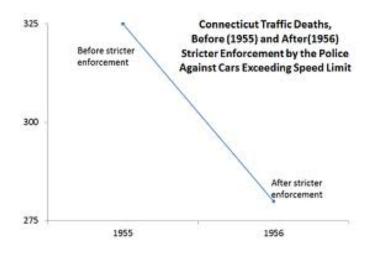


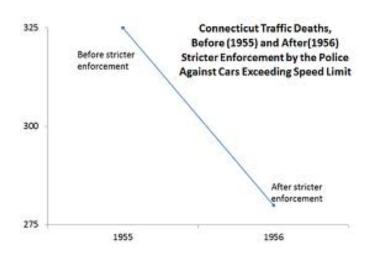


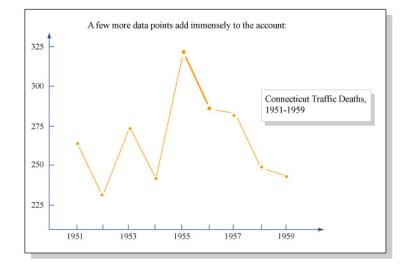


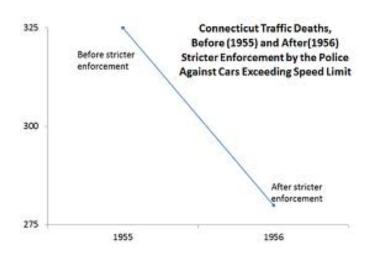
Conclusion: The use of two or three varying dimensions to show onedimensional data is a weak and inefficient technique.

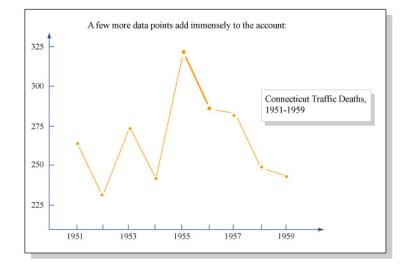
The number of information-carrying dimensions **should not exceed** the number of dimensions in the data.

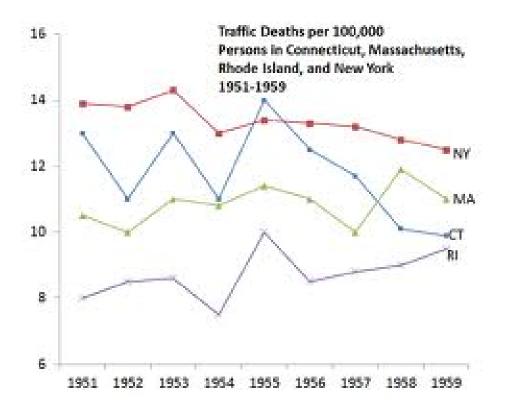












Lying graphics cheapen the graphical art everywhere.

Lying graphics cheapen the graphical art everywhere.

Six Principles of Graphical Integrity :

1. The representation of numbers should be directly proportional to the numerical quantities represented

Lying graphics cheapen the graphical art everywhere.

- 1. The representation of numbers should be directly proportional to the numerical quantities represented
- 2. Clear, detailed and thorough labeling. Write out explanations and label important events.

Lying graphics cheapen the graphical art everywhere.

- 1. The representation of numbers should be directly proportional to the numerical quantities represented
- 2. Clear, detailed and thorough labeling. Write out explanations and label important events.
- 3. Show data variation, not design variation.

Lying graphics cheapen the graphical art everywhere.

- 1. The representation of numbers should be directly proportional to the numerical quantities represented
- 2. Clear, detailed and thorough labeling. Write out explanations and label important events.
- 3. Show data variation, not design variation.
- 4. In time-series displays of money, use deflated and standardized units of monetary measurement.

Lying graphics cheapen the graphical art everywhere.

- 1. The representation of numbers should be directly proportional to the numerical quantities represented
- 2. Clear, detailed and thorough labeling. Write out explanations and label important events.
- 3. Show data variation, not design variation.
- 4. In time-series displays of money, use deflated and standardized units of monetary measurement.
- 5. The number of information-carrying dimensions, should not exceed the number of dimensions in the data.

Lying graphics cheapen the graphical art everywhere.

- 1. The representation of numbers should be directly proportional to the numerical quantities represented
- 2. Clear, detailed and thorough labeling. Write out explanations and label important events.
- 3. Show data variation, not design variation.
- 4. In time-series displays of money, use deflated and standardized units of monetary measurement.
- 5. The number of information-carrying dimensions, should not exceed the number of dimensions in the data.
- 6. Graphics must not quote data out of context.

Thank you for the attention.