Graphical Excellence—Edward Tufte

Graphical Excellence—Edward Tufte

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

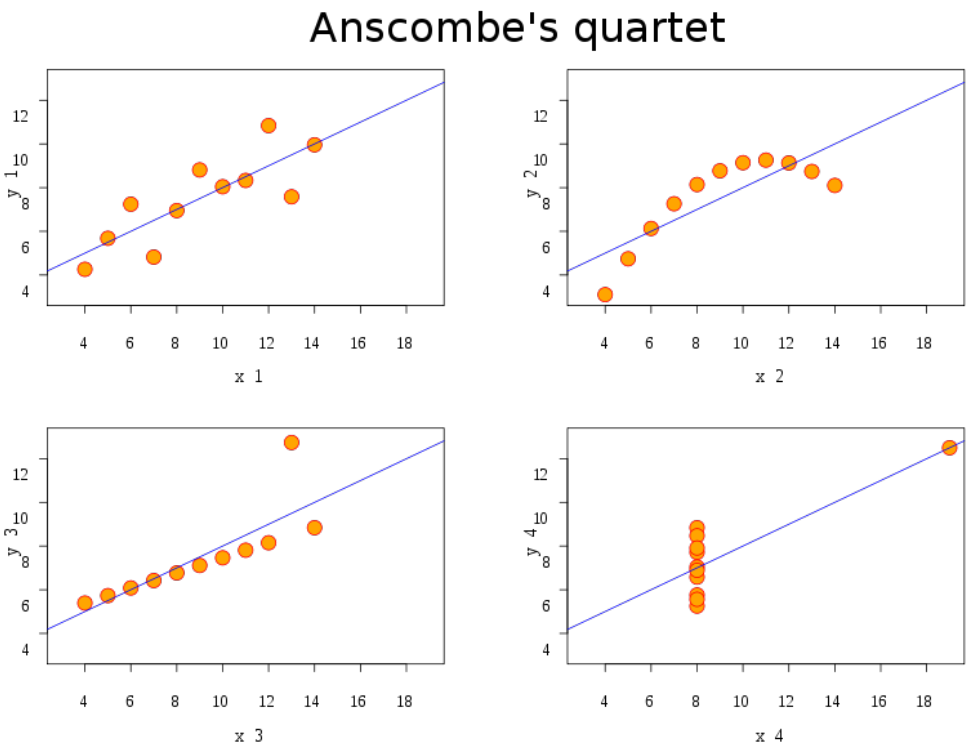
Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency.

- 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		III		IV	
x	y	x	y	x	y	x	y
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



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.

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

A- New York Stock prices
B- Solar Radiation inverted
C- London Stock prices

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

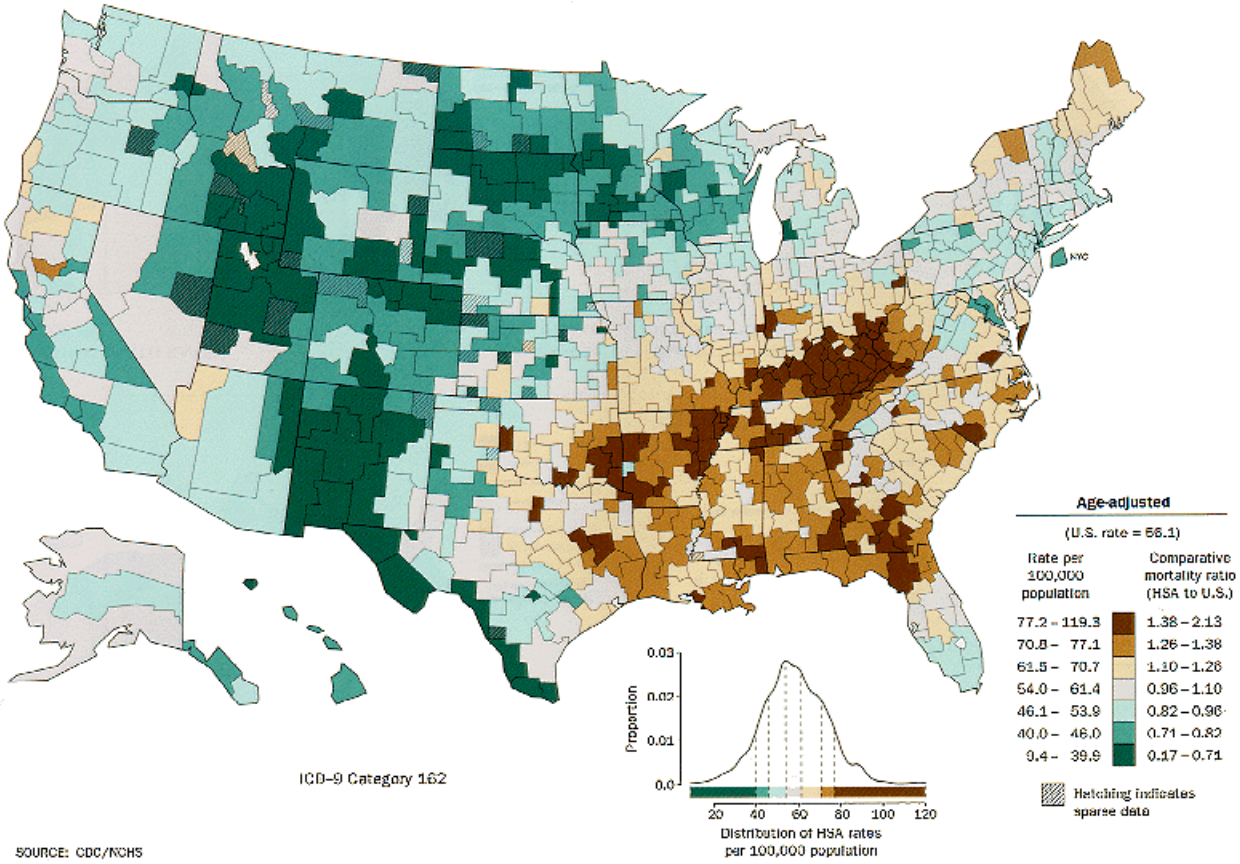
- Fundamental Graphical Designs:
  - Data Maps
  - Time-Series
  - Space-time narrative designs
  - Relational Graphics

Data Maps

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

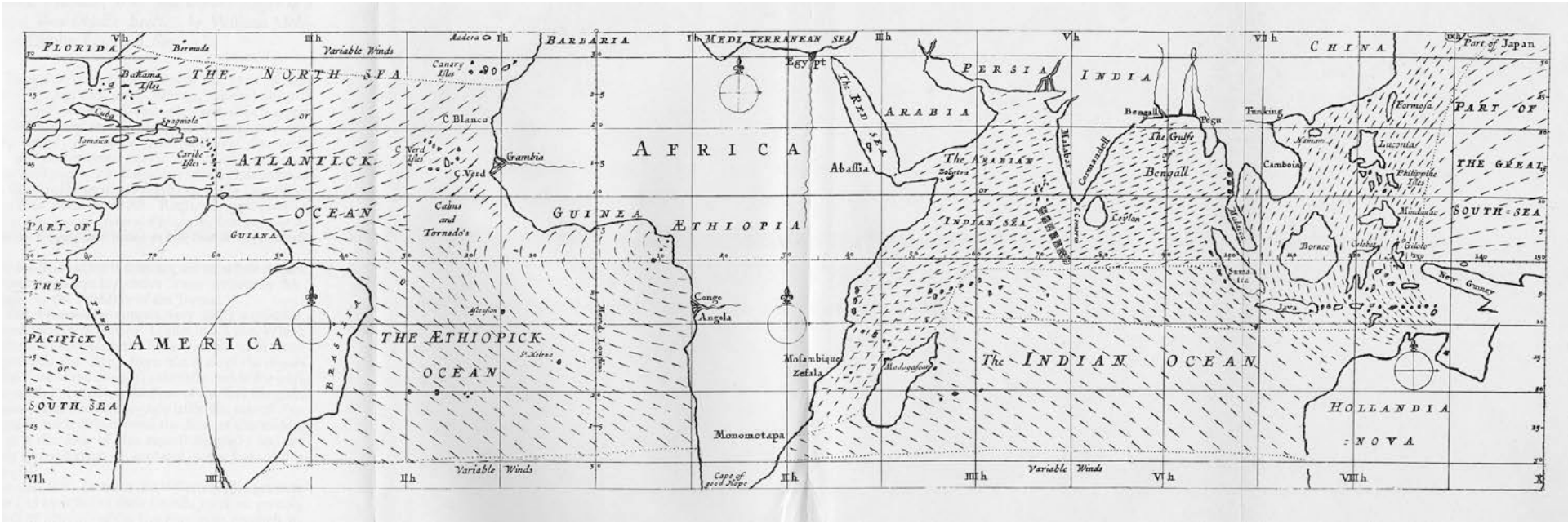
Data Maps

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



Data Maps

- Edmond Halley, 1686

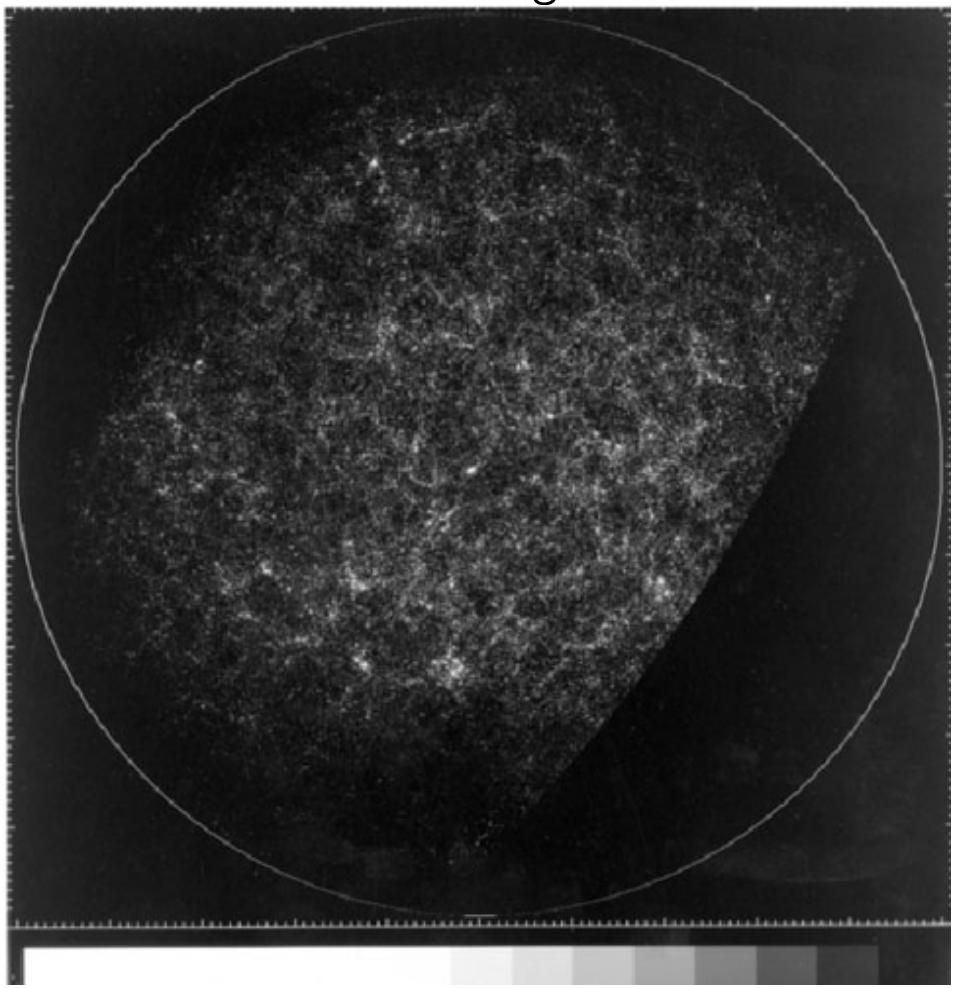


Data Maps

- Dr. John Shaw, 1854

Data Maps

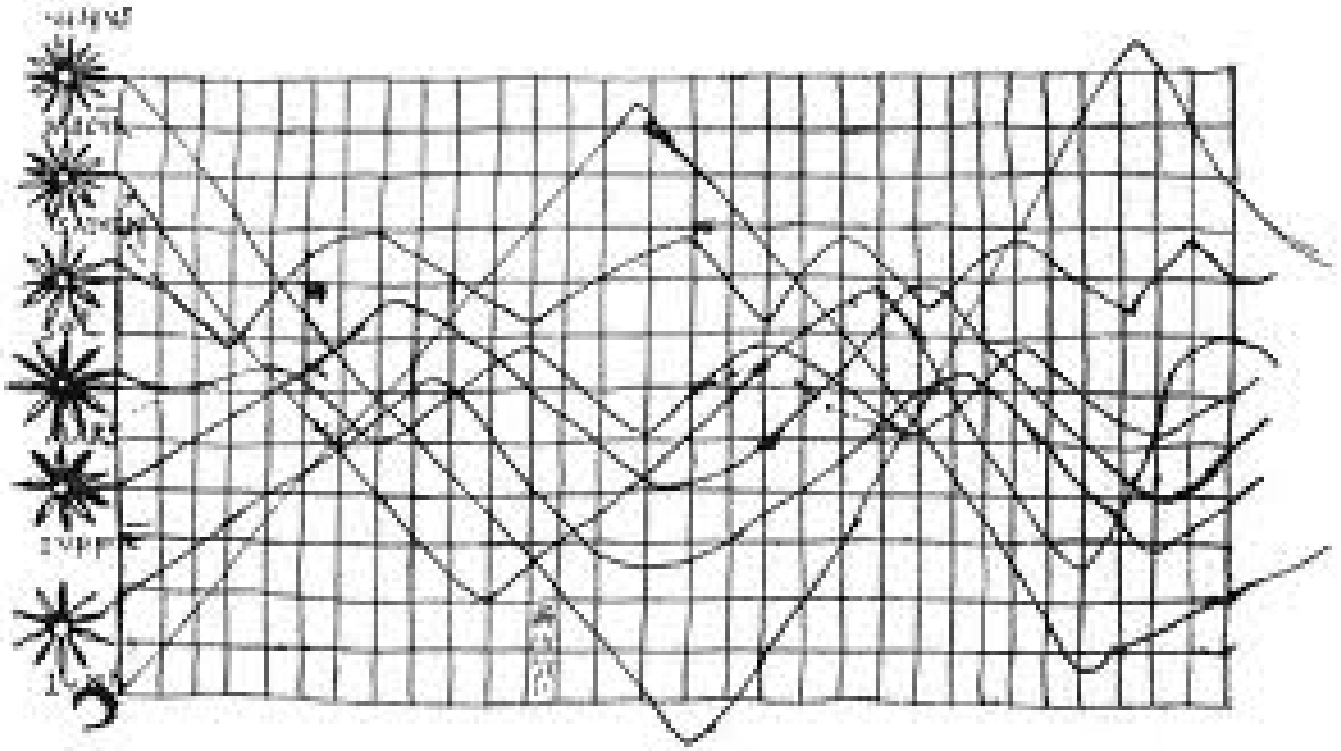
- “New Reduction of the Lick Catalog of Galaxies”- Seldner,Siebers,Groth and Peebles



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

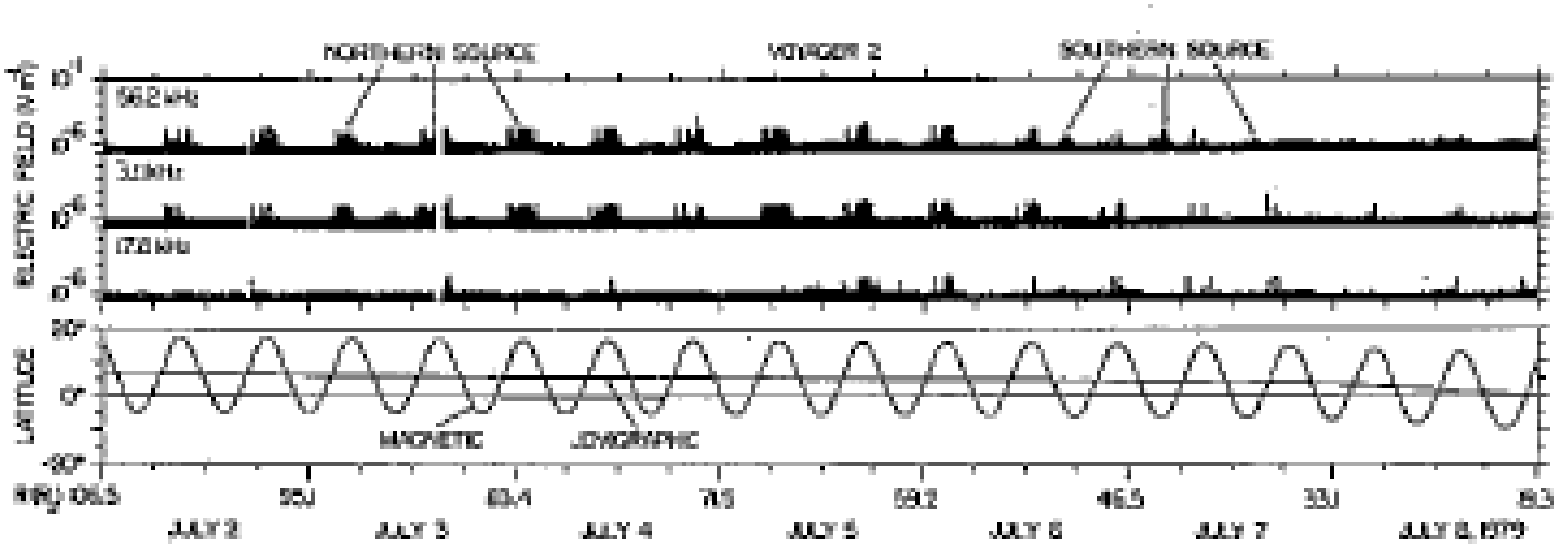
Time-Series

- “A Note on a Tenth Century Graph”- H. Gray Funkhouser



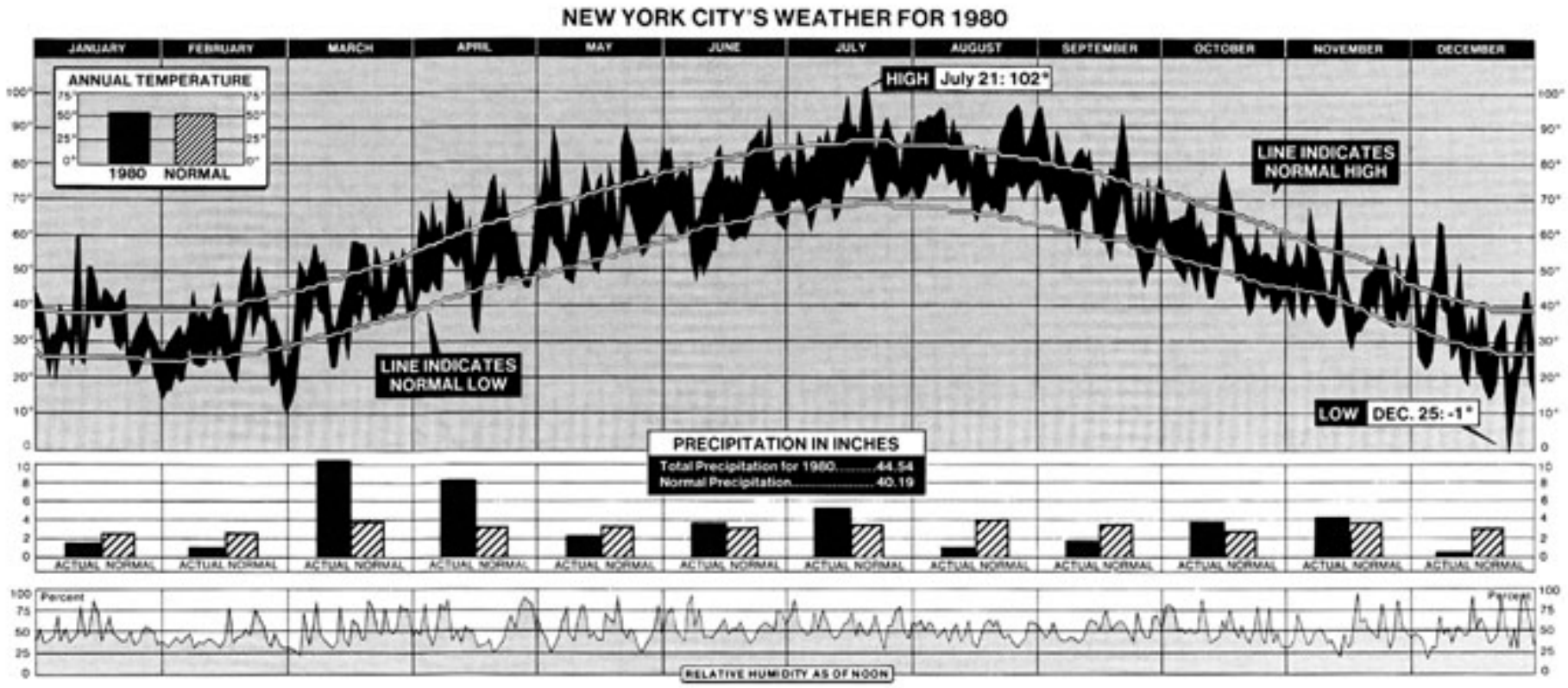
Time-Series

- “Plasma Wave Observation Near Jupiter”- Gurnett, Kurth and Scarf

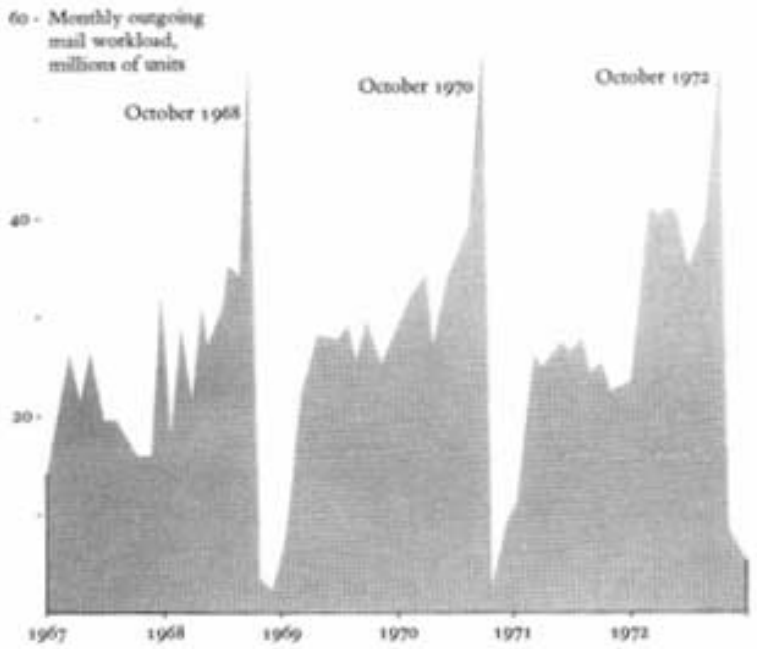


Time-Series

- New York Weather 1980



Time-Series



FRANKED MAIL TIE TO VOTING SHOWN

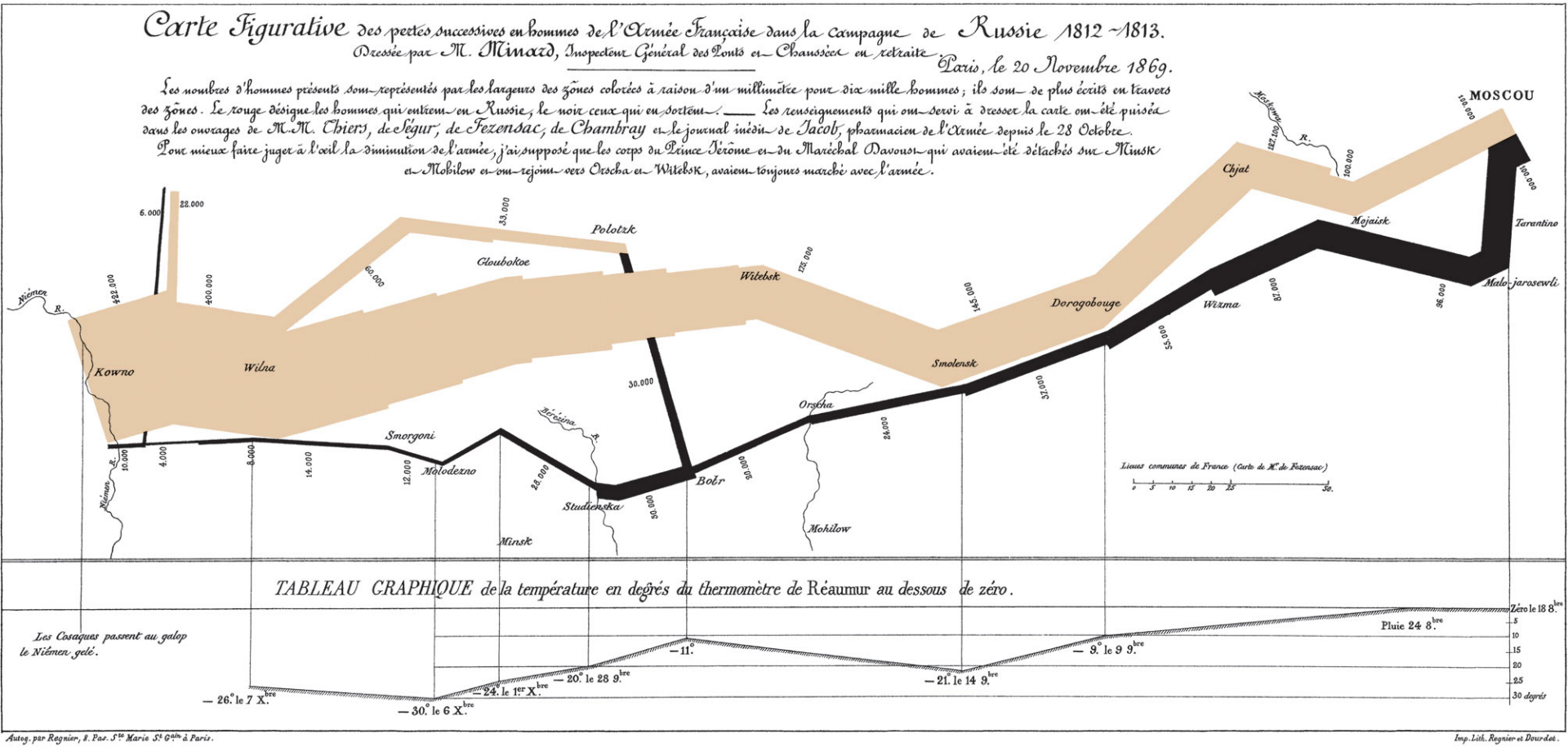
Testimony Finds the Volume Rises Before Elections

WASHINGTON, June 1 (AP)

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

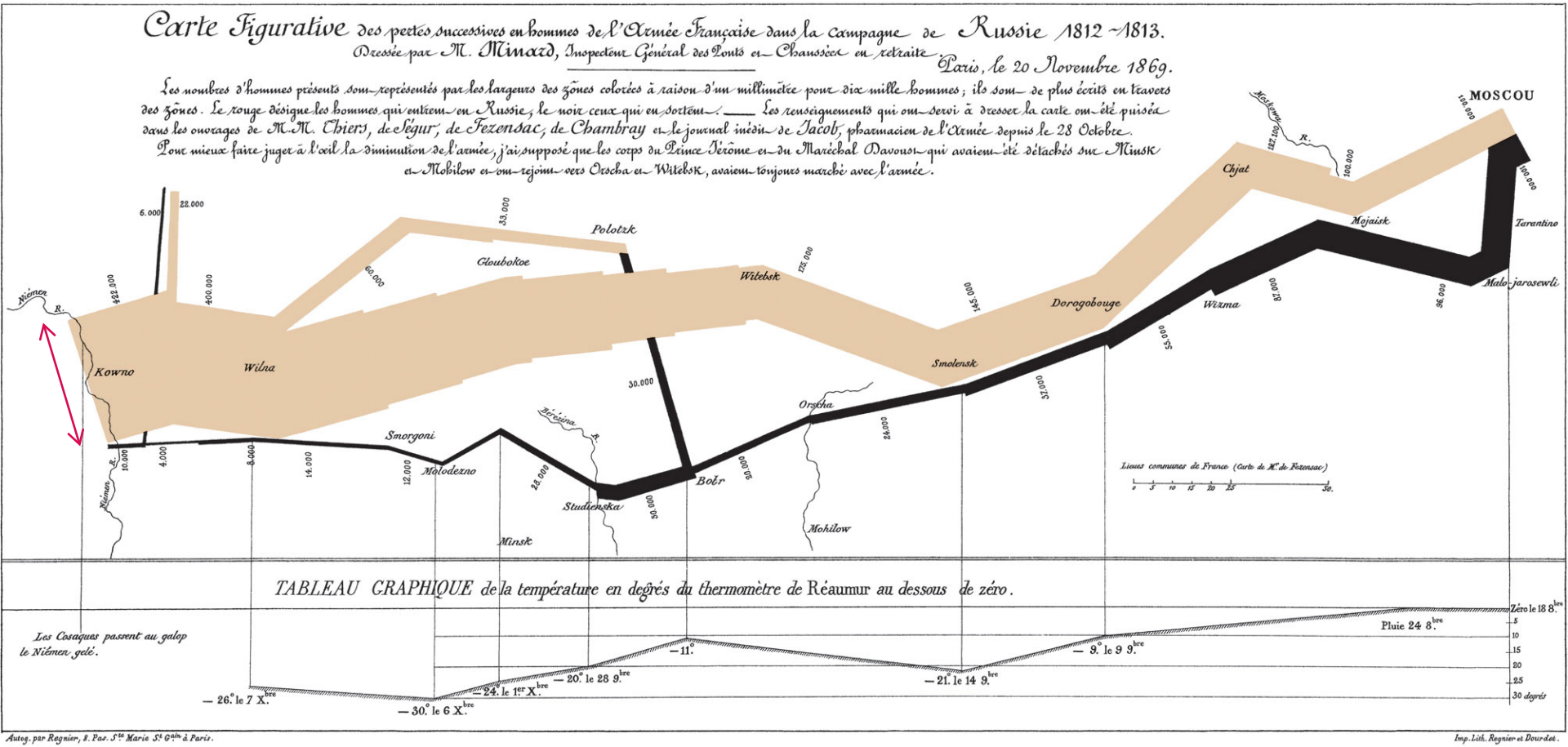
Narrative Graphics of Space and Time

Charles Joseph Minard



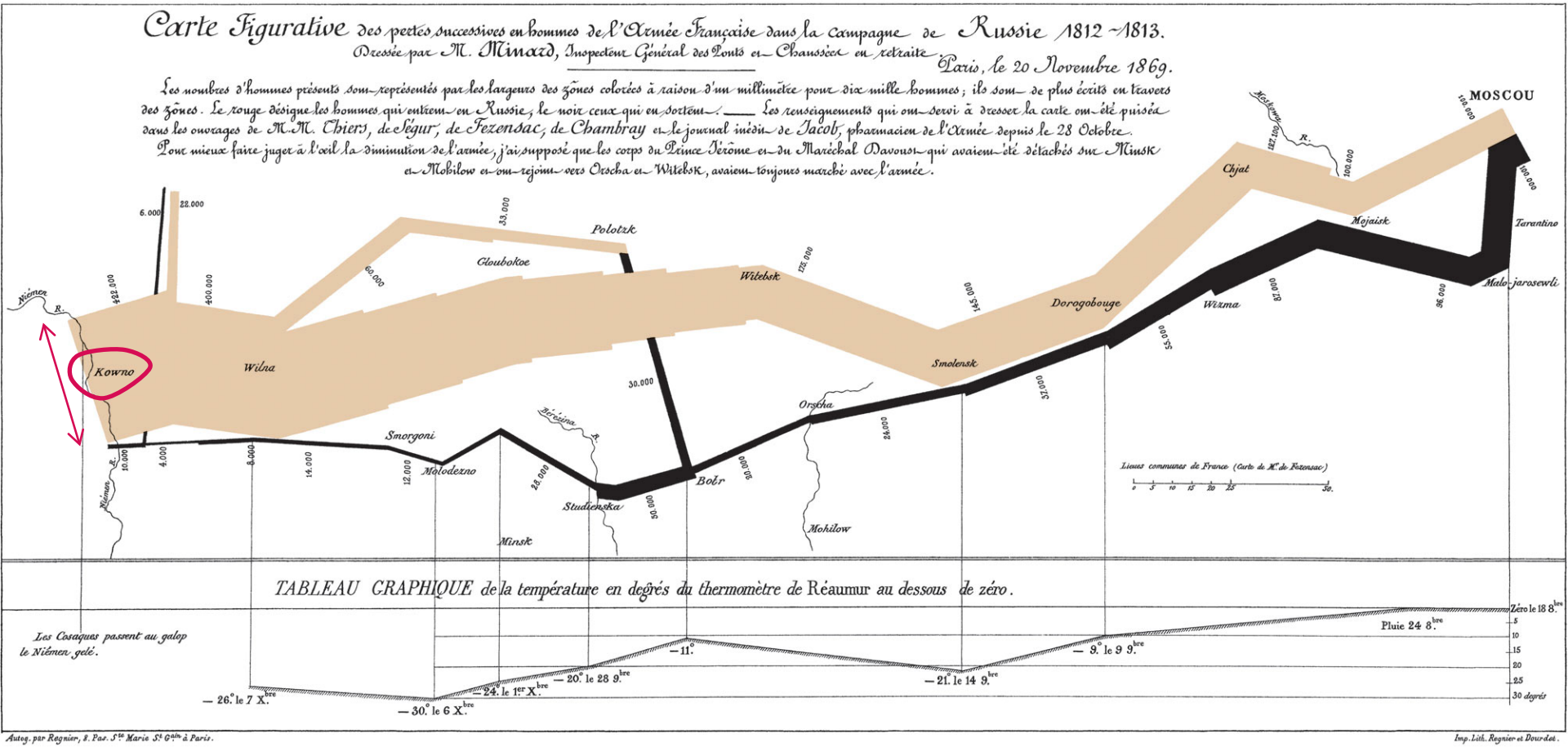
Narrative Graphics of Space and Time

Charles Joseph Minard



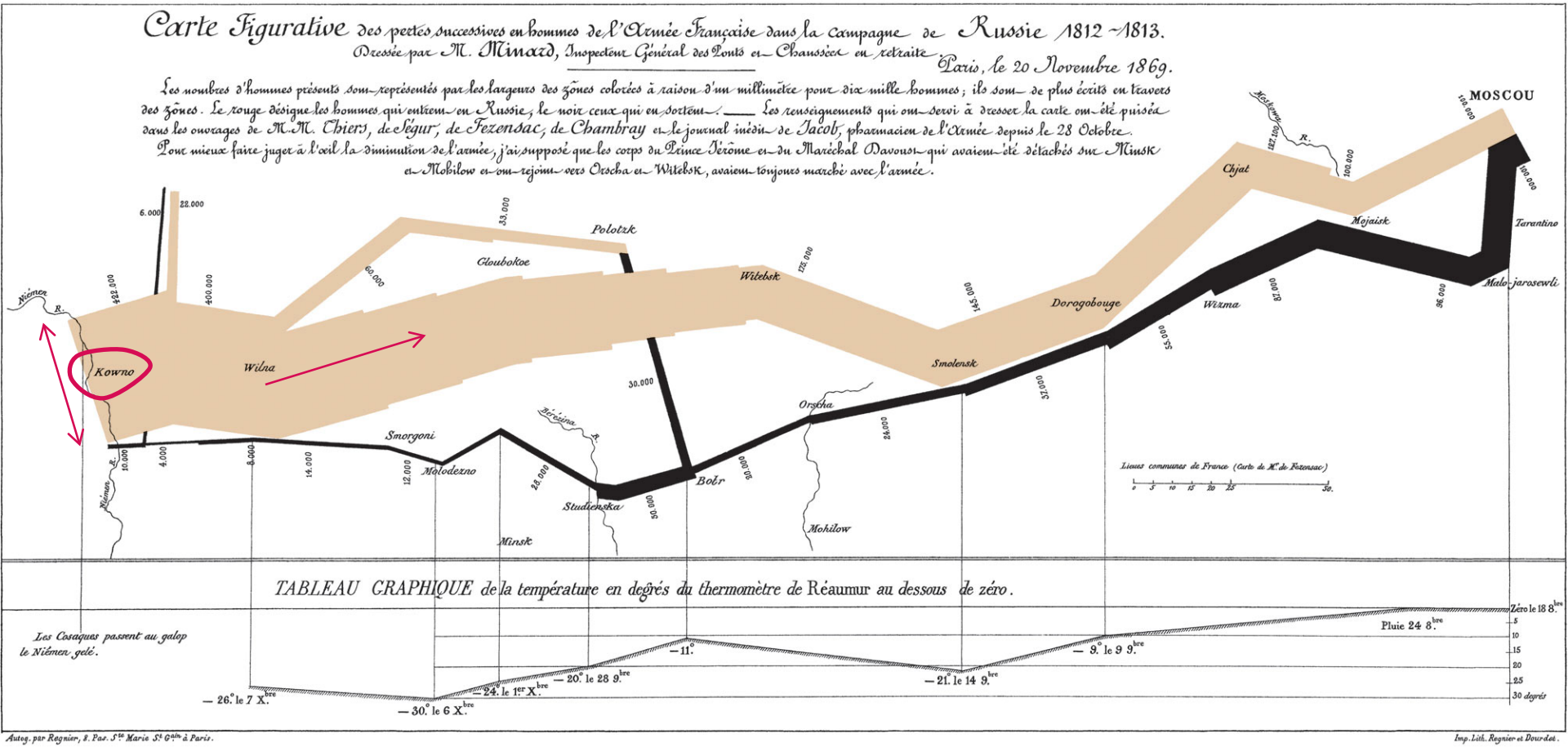
Narrative Graphics of Space and Time

Charles Joseph Minard



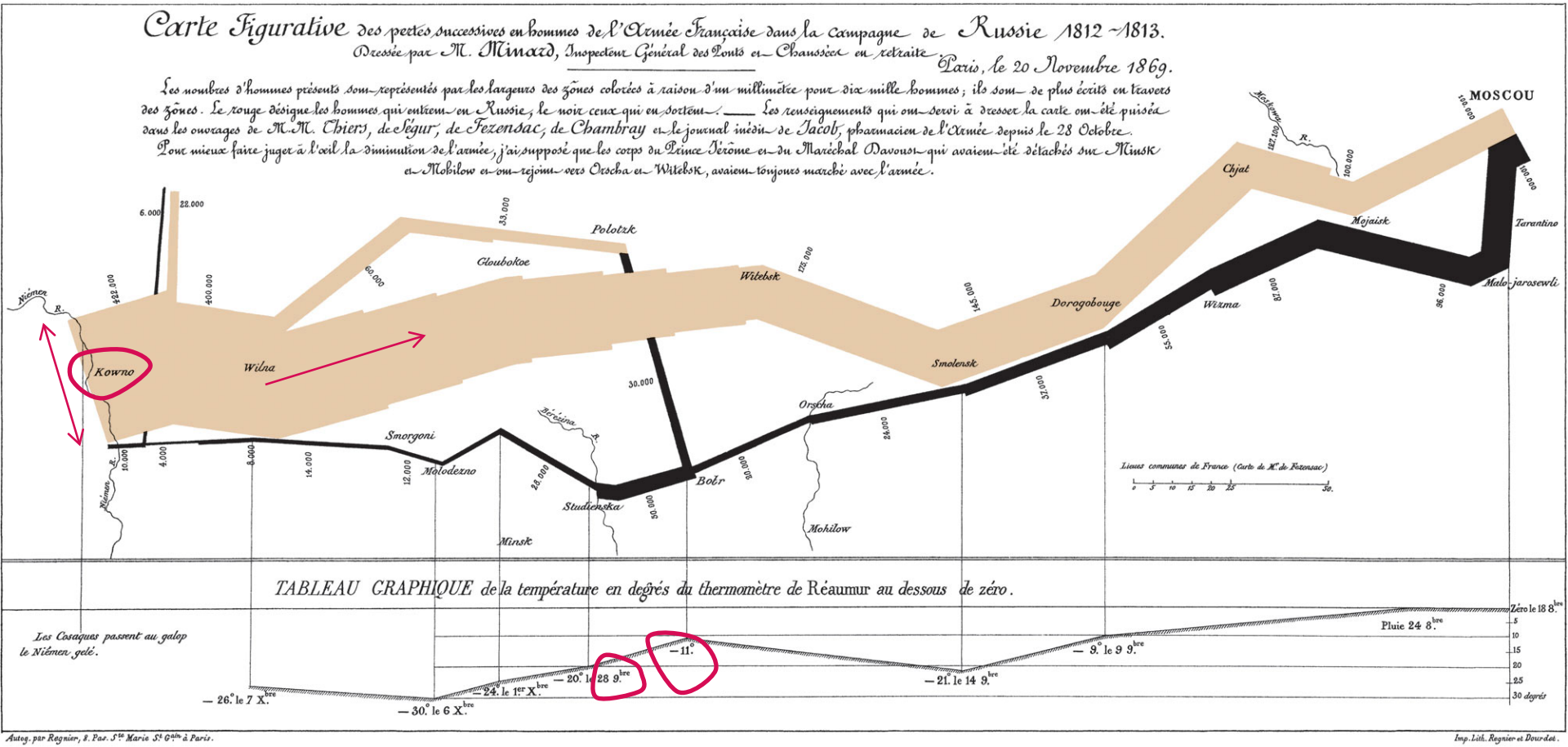
Narrative Graphics of Space and Time

Charles Joseph Minard



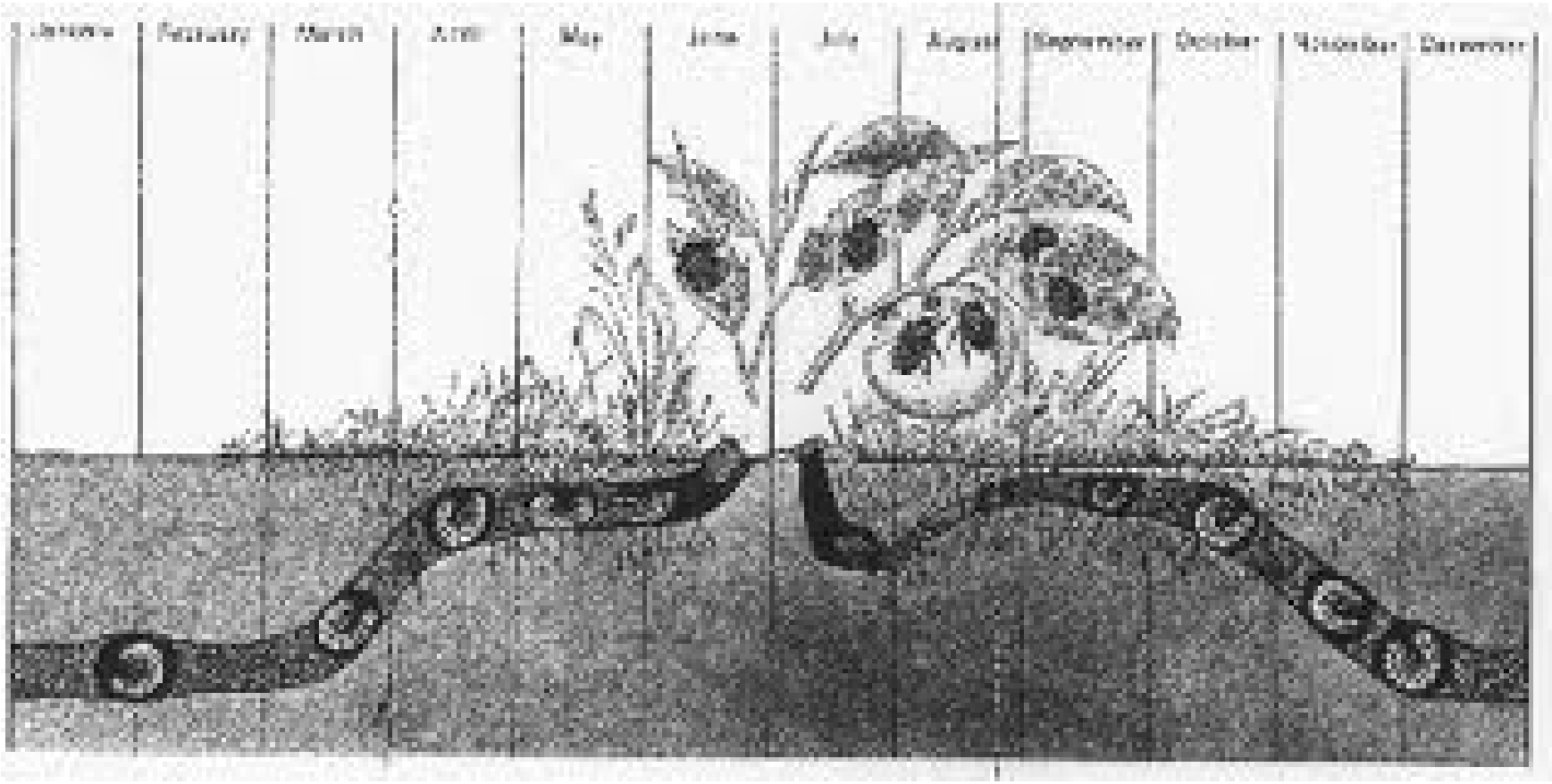
Narrative Graphics of Space and Time

Charles Joseph Minard



Narrative Graphics of Space and Time

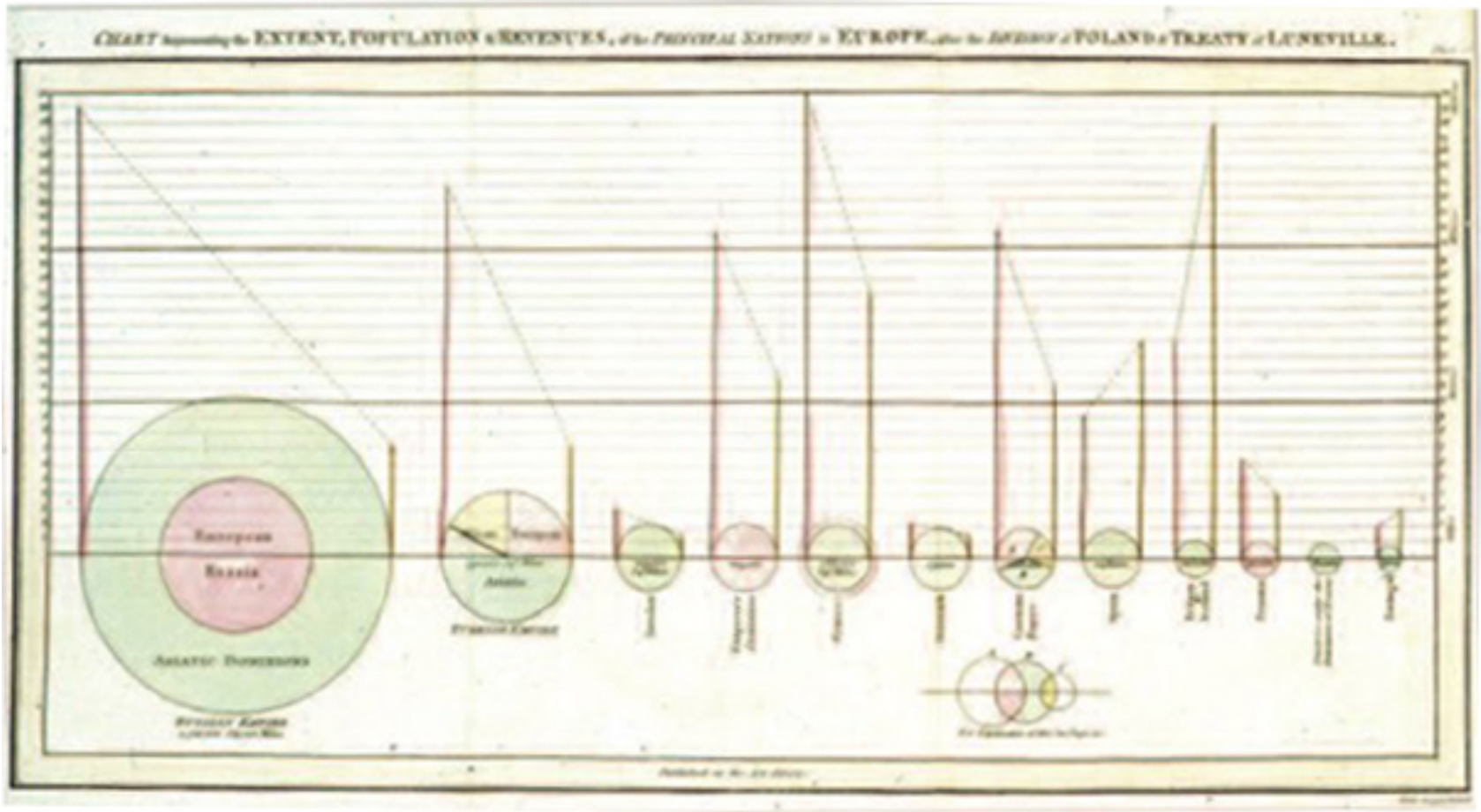
“Man and Insects”- I. Hugh Newman



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

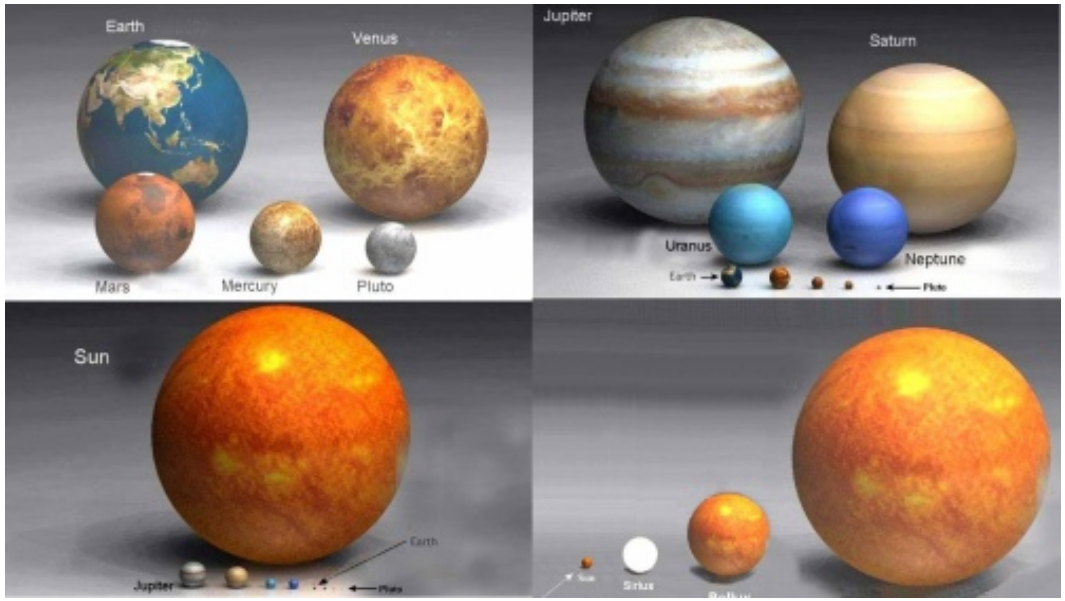
Relational Graphics

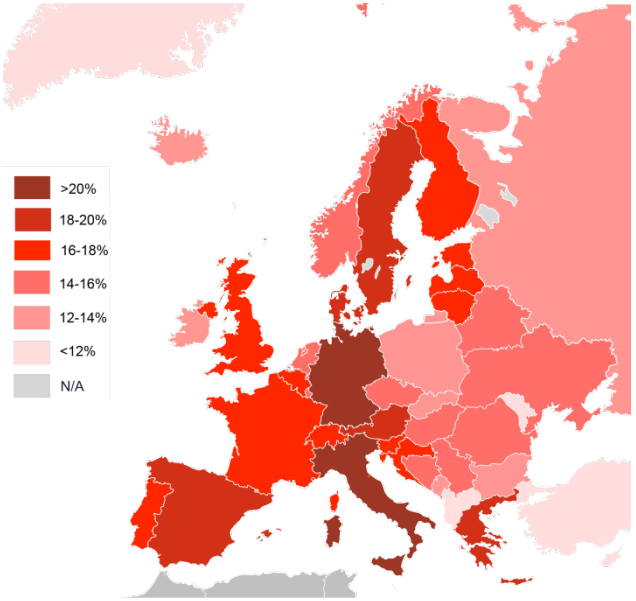
“The Statistical Breviary”- Playfair



Relational Graphics

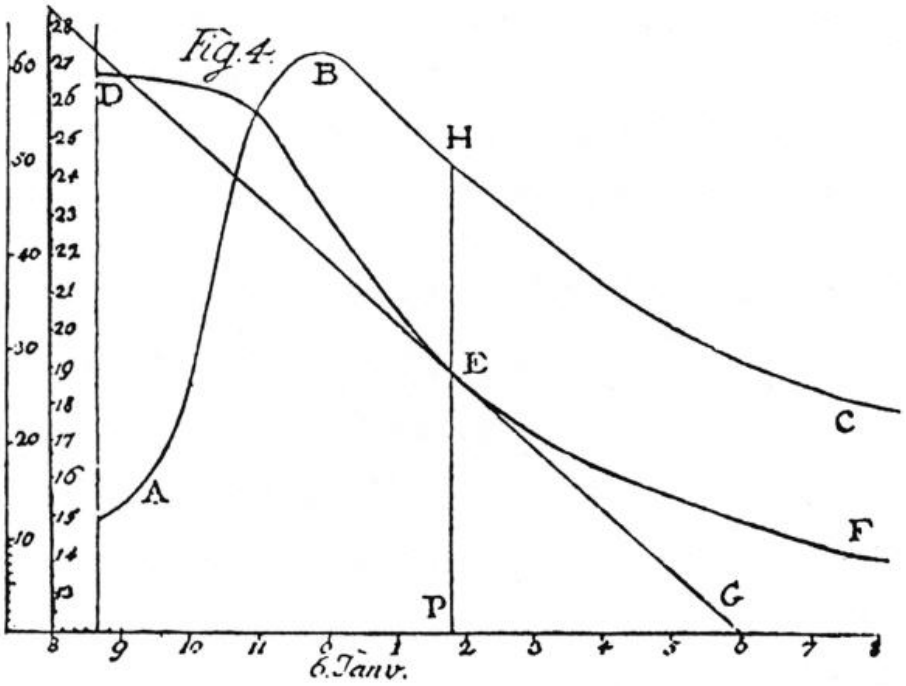
Comparing sizes represented with similar form.





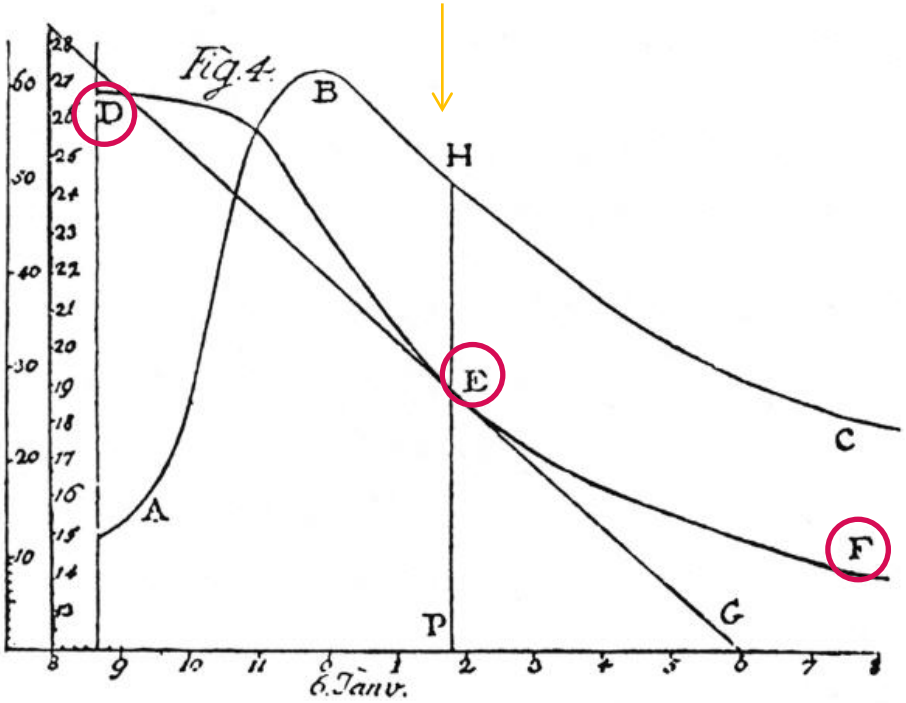
Relational Graphics

“Evaporation rate of water”- J.H. Lambert



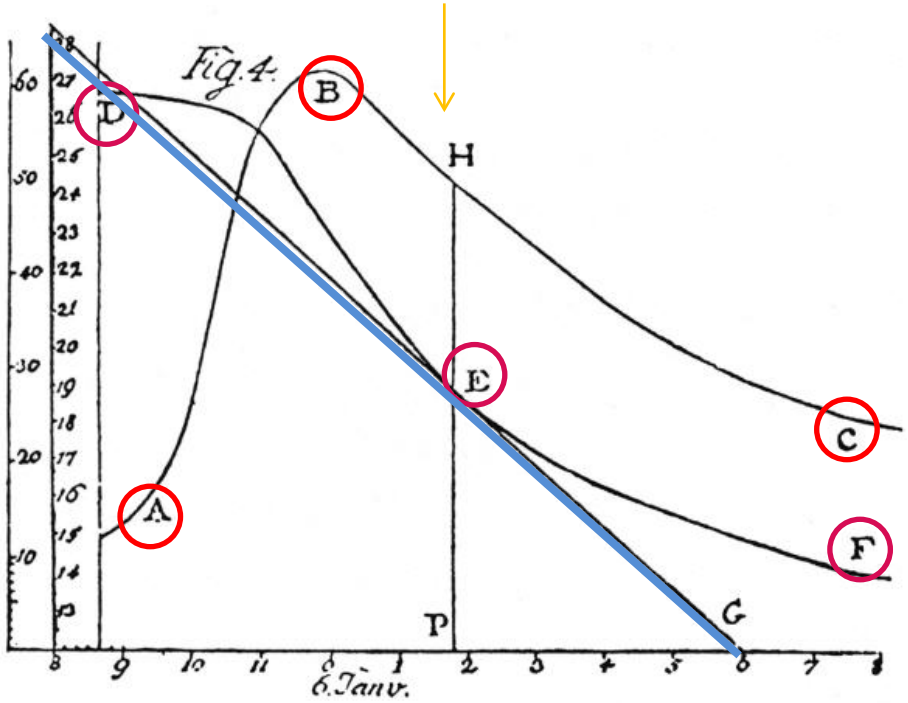
Relational Graphics

“Evaporation rate of water”- J.H. Lambert



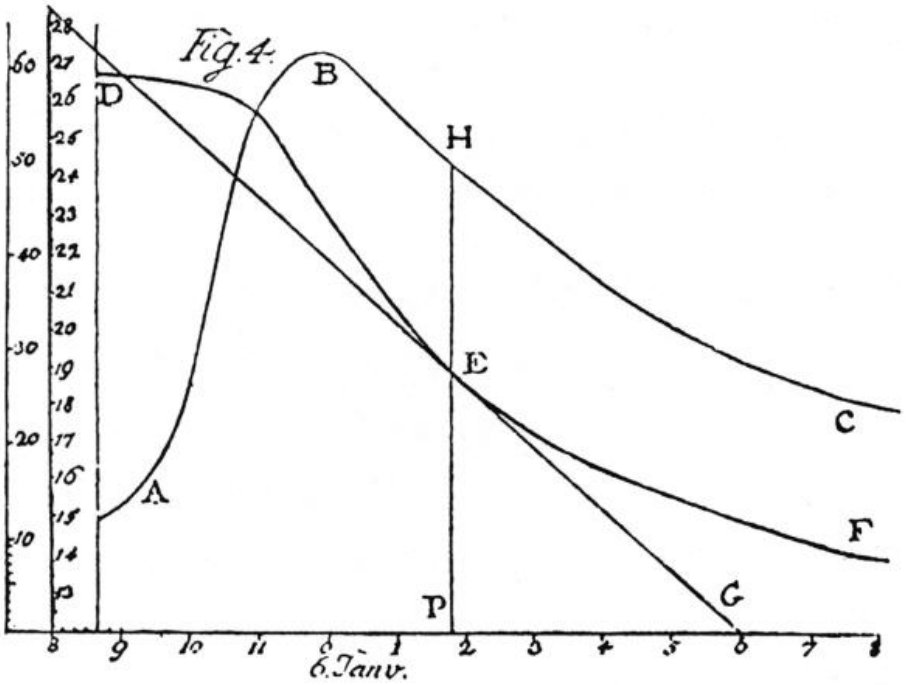
Relational Graphics

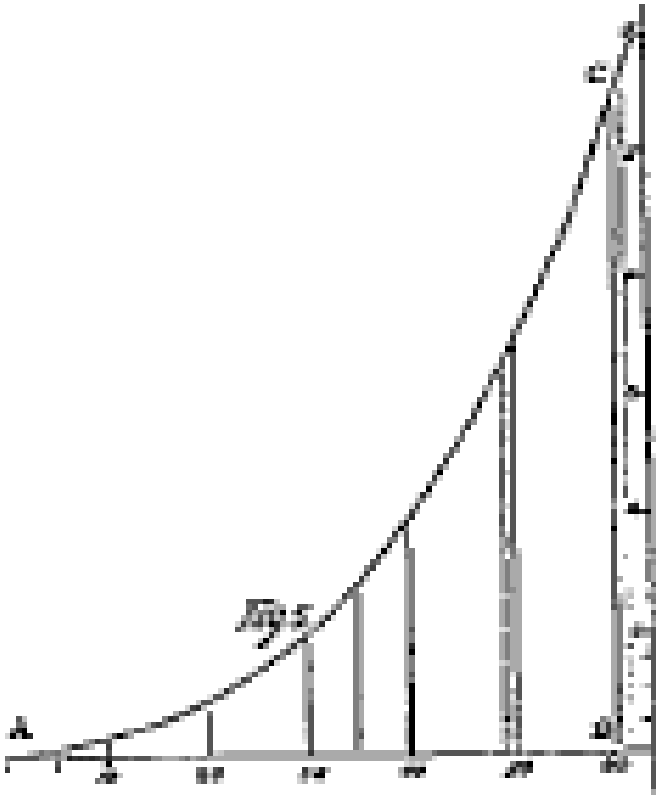
“Evaporation rate of water”- J.H. Lambert



Relational Graphics

"Evaporation rate of water"- J.H. Lambert





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.

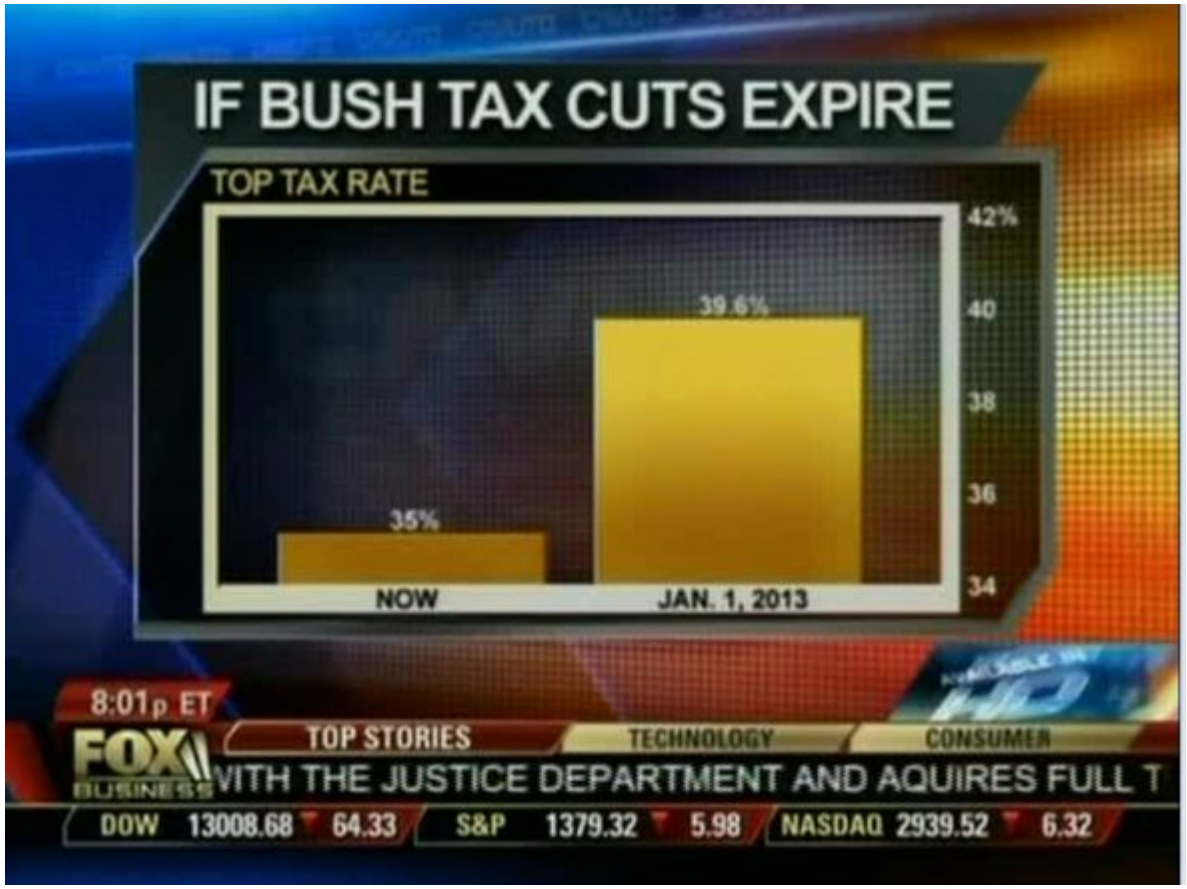
Graphical excellence requires telling the truth about data.

Graphical Integrity—Edward Tufte

Graphical excellence requires telling the truth about data.

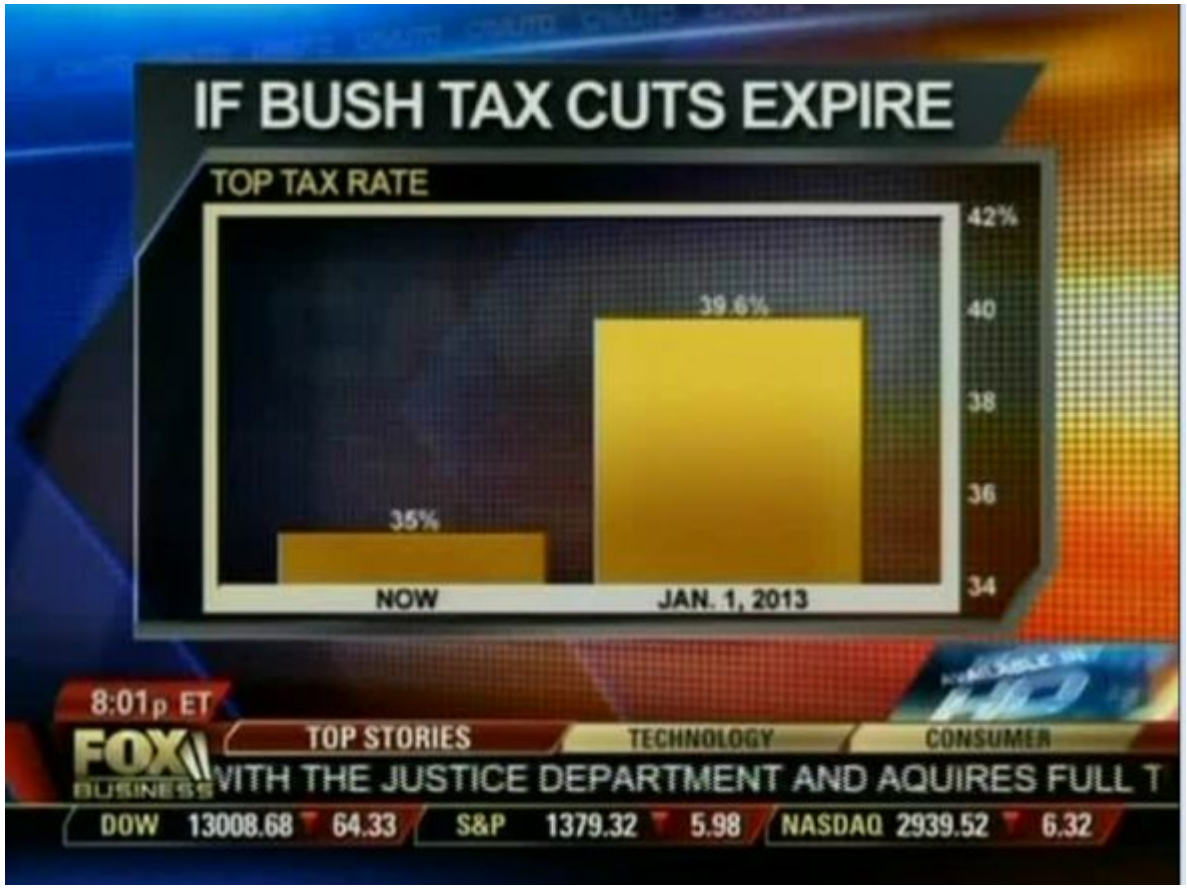
Graphical Integrity—Edward Tufte

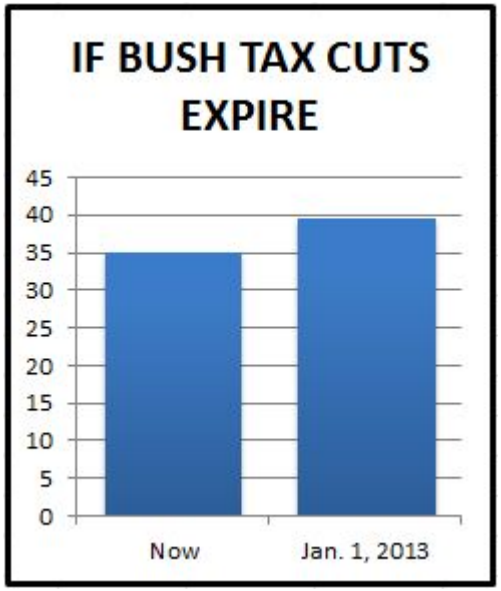
Misleading graphics



Graphical Integrity—Edward Tufte

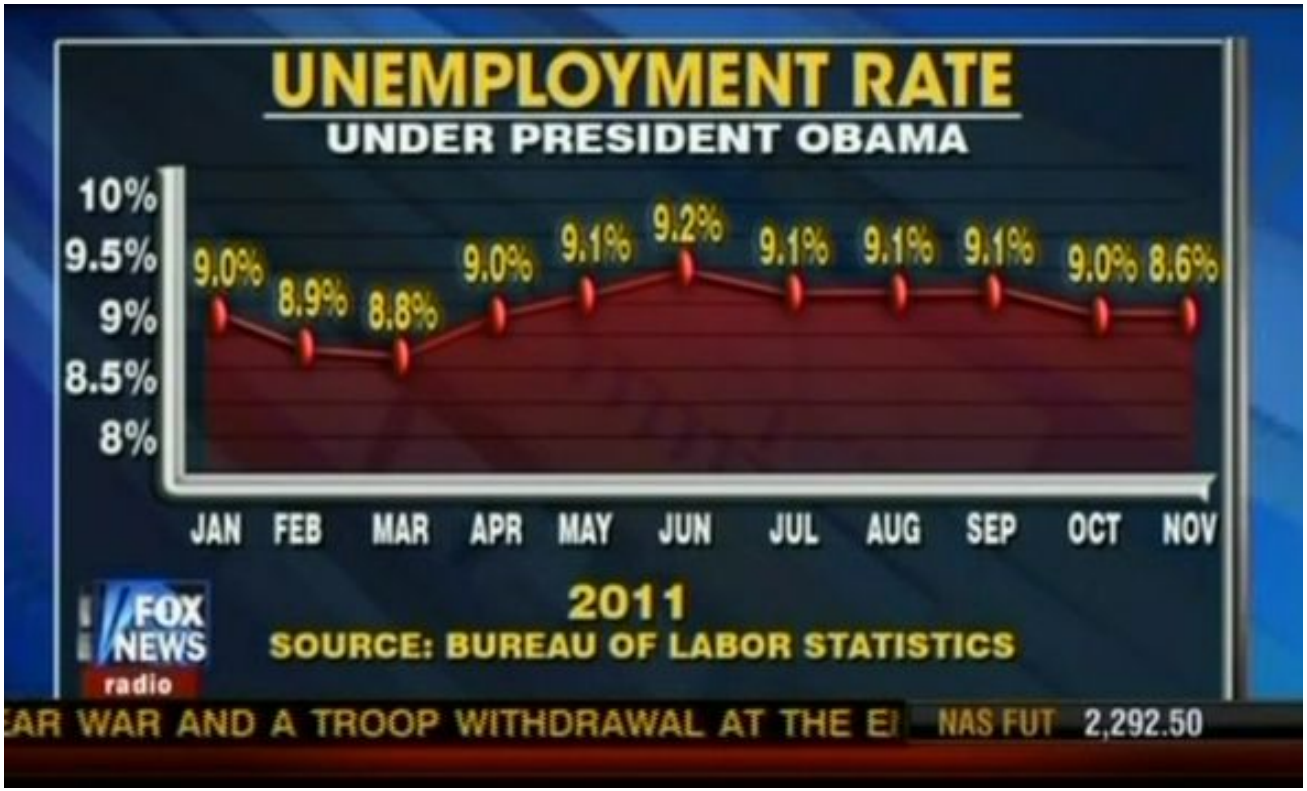
Misleading graphics





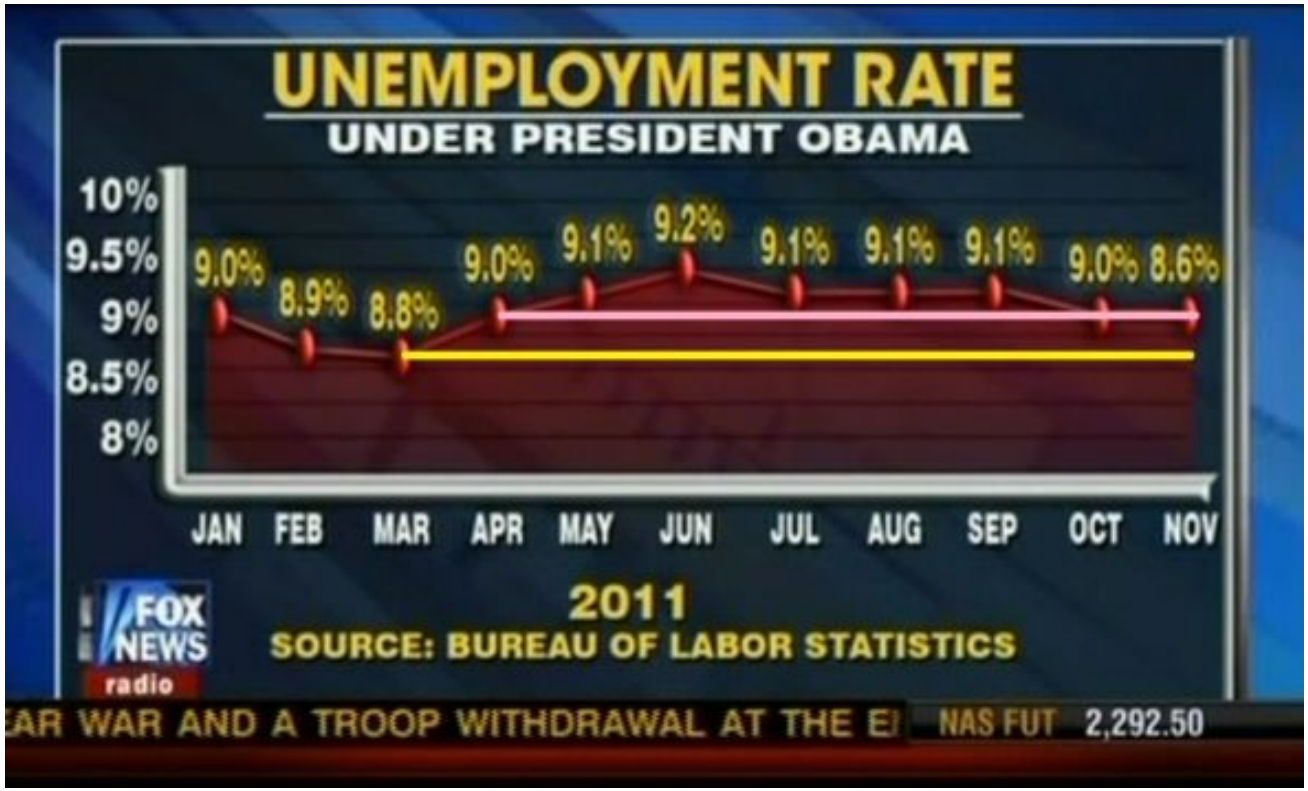
Graphical Integrity—Edward Tufte

Misleading graphics



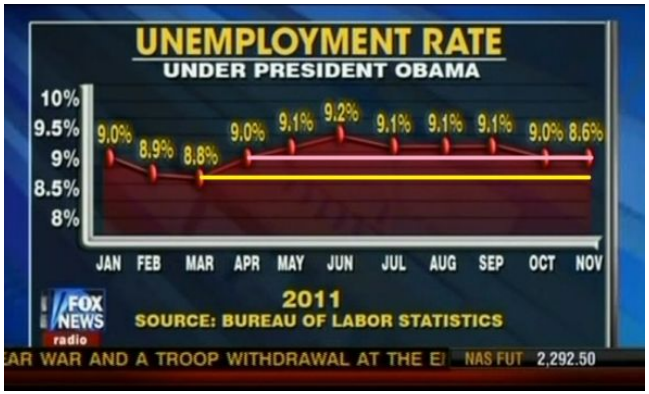
Graphical Integrity—Edward Tufte

Misleading graphics



Graphical Integrity—Edward Tufte

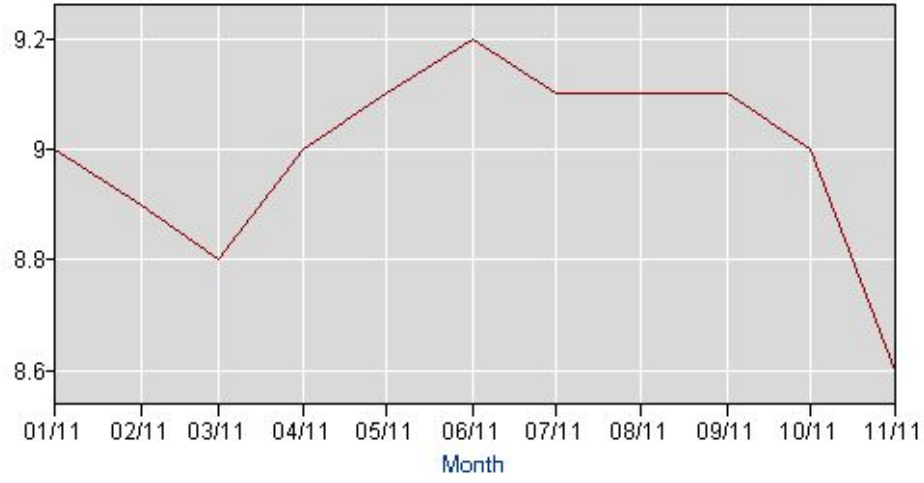
Misleading graphics



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



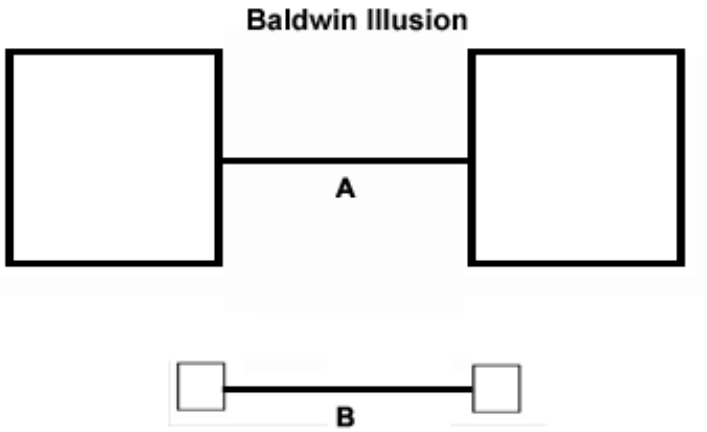
What is Distortion in a Data Graphic?

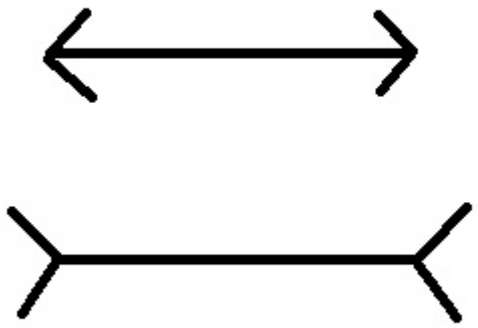
What is Distortion in a Data Graphic?

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

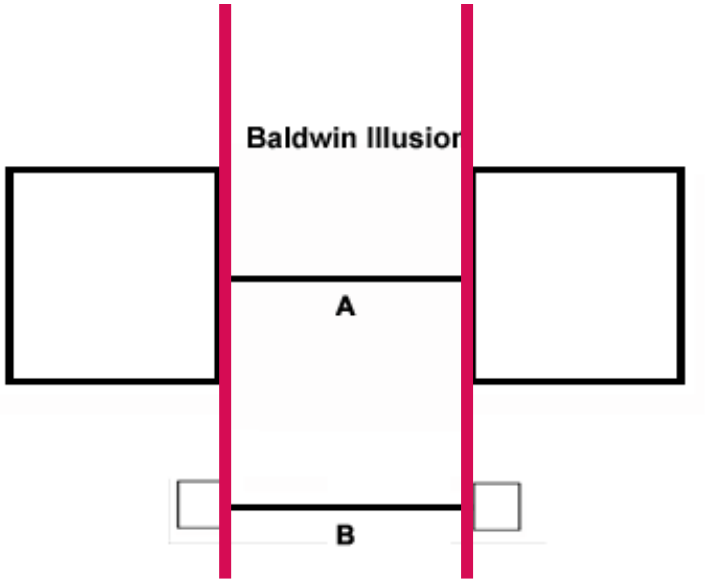
Visual representation of data $\neq$ numerical representation

What is Distortion in a Data Graphic?





What is Distortion in a Data Graphic?



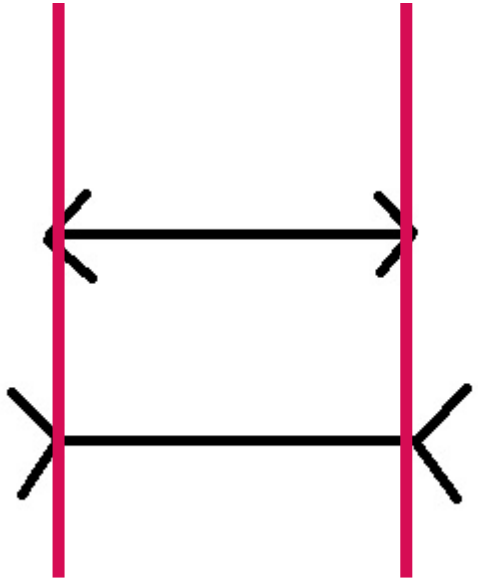


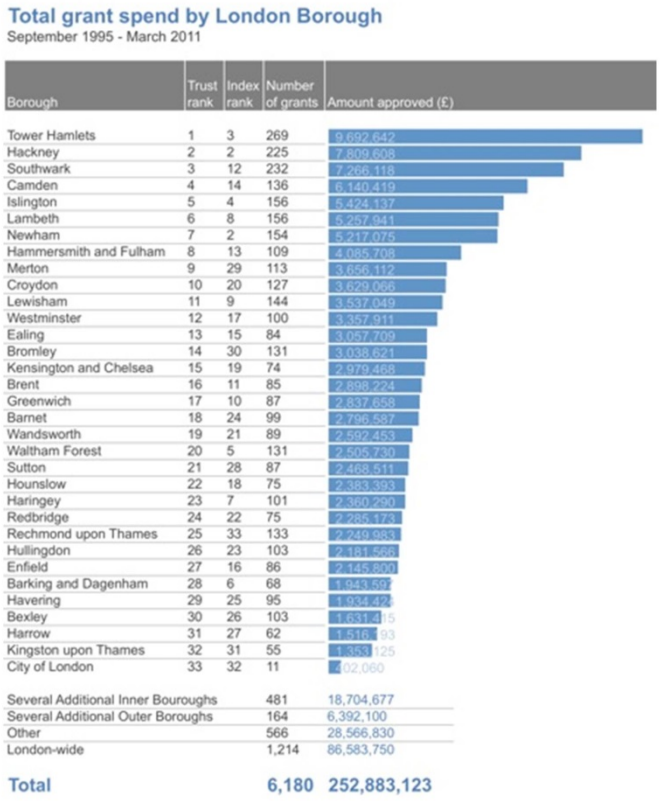
Table vs. Graph

Before: **Table**

Total grants spend by London Borough
September 1995 to March 2011

Trust rank	Deprivation index rank	Borough	Amount approved	Number of grants
1	3	Tower Hamlets	£9,692,642	269
2	1	Hackney	£7,809,608	225
3	12	Southwark	£7,266,118	232
4	14	Camden	£6,140,419	136
5	4	Islington	£5,424,137	156
6	8	Lambeth	£5,257,941	156
7	2	Newham	£5,217,075	154
8	13	Hammersmith and Fulham	£4,085,708	109
9	29	Merton	£3,856,112	113
10	20	Croydon	£3,829,066	127
11	9	Lewisham	£3,537,049	144
12	17	Westminster	£3,357,911	100
13	15	Ealing	£3,057,709	84
14	30	Bromley	£3,038,621	131
15	19	Kensington and Chelsea	£2,979,468	74
16	11	Brent	£2,898,224	85
17	10	Greenwich	£2,837,658	87
18	24	Barnet	£2,796,587	99
19	21	Wandsworth	£2,592,453	89
20	5	Waltham Forest	£2,505,730	131
21	28	Sutton	£2,468,511	87
22	18	Hounslow	£2,383,393	75
23	7	Haringey	£2,360,290	101
24	22	Redbridge	£2,285,173	75
25	33	Richmond upon Thames	£2,249,983	133
26	23	Hillingdon	£2,181,566	103
27	16	Enfield	£2,145,800	86
28	6	Barking and Dagenham	£1,943,597	68
29	25	Havering	£1,934,424	95
30	26	Bexley	£1,631,415	103
31	27	Harrow	£1,516,193	62
32	31	Kingston upon Thames	£1,353,125	55
33	32	City of London	£402,060	11
		Several Inner Boroughs	£18,704,677	481
		Several Outer Boroughs	£6,392,100	164
		Other combination	£28,566,830	566
		London-wide	£86,583,750	1,214
		Total	£252,883,123	6,180

After: **Grable**



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

Two Principles:

- Representation of numbers: the numbers should be directly proportional

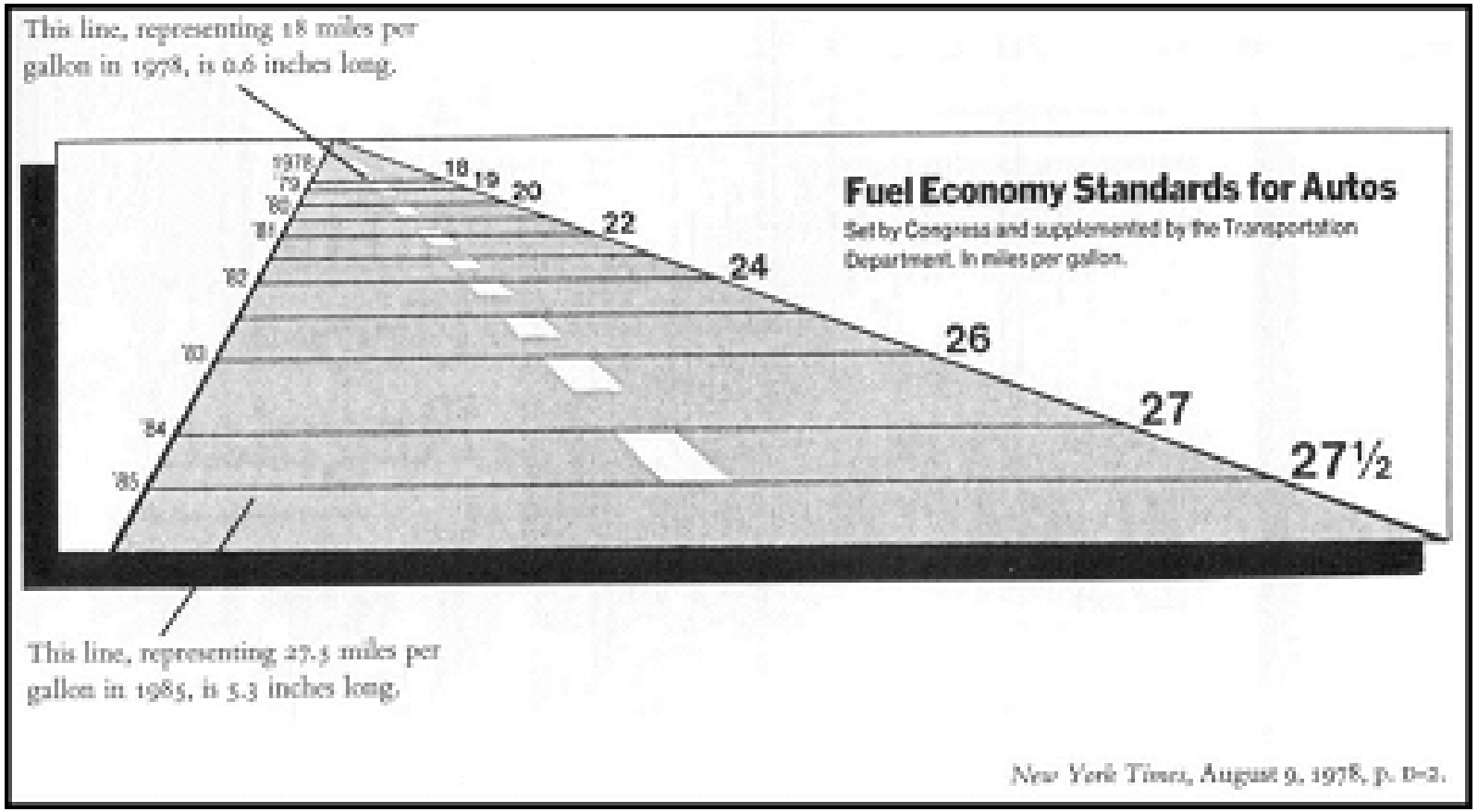
- Clear, detailed and thorough labeling: write out explanations and label important events

First Principle:

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

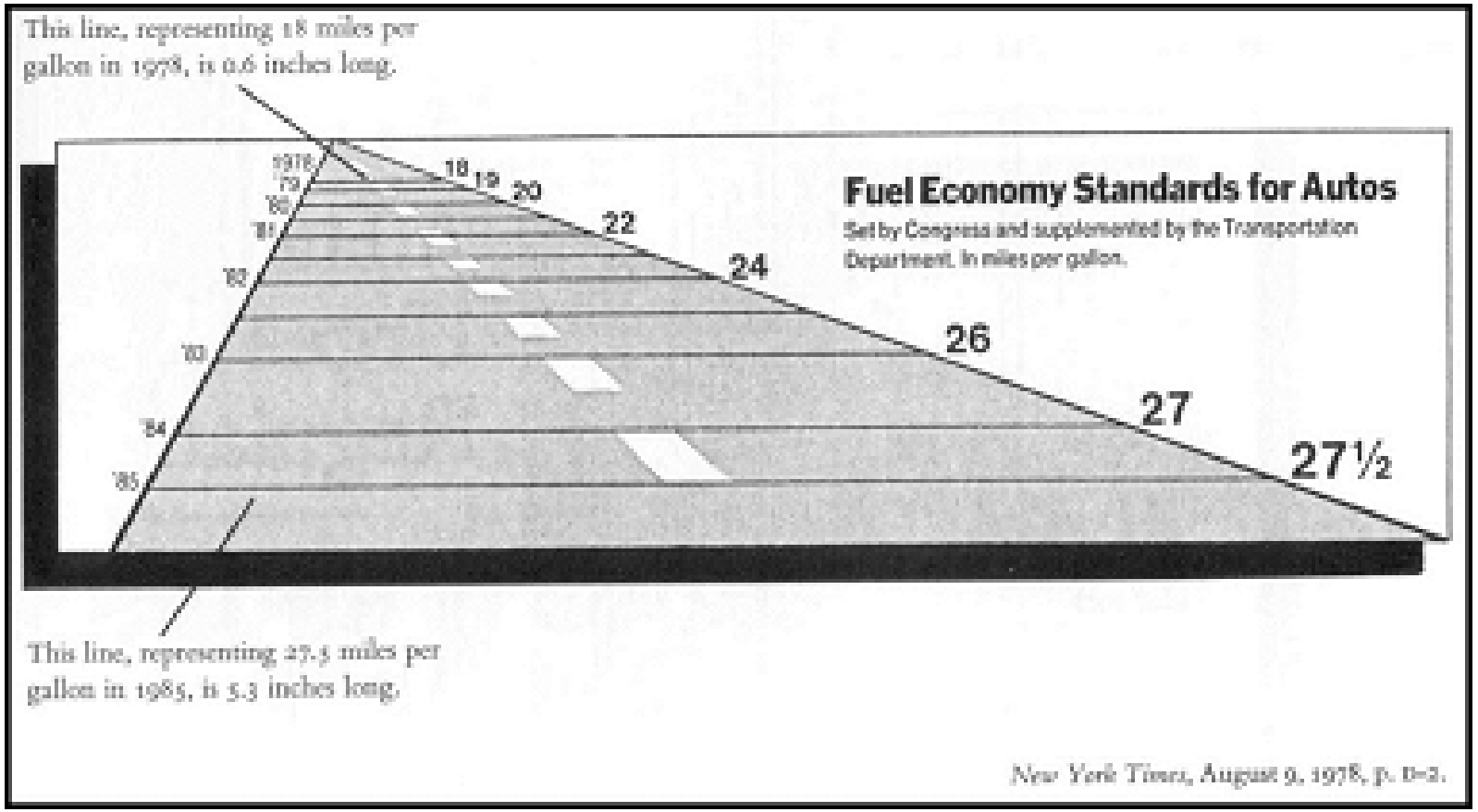
First Principle:

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

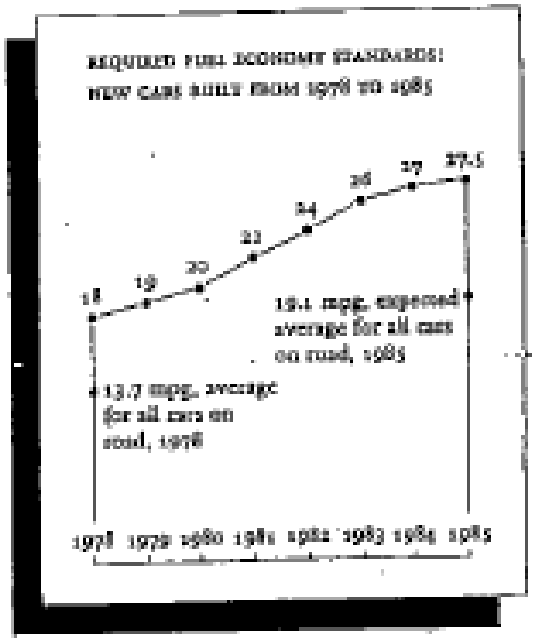


First Principle:

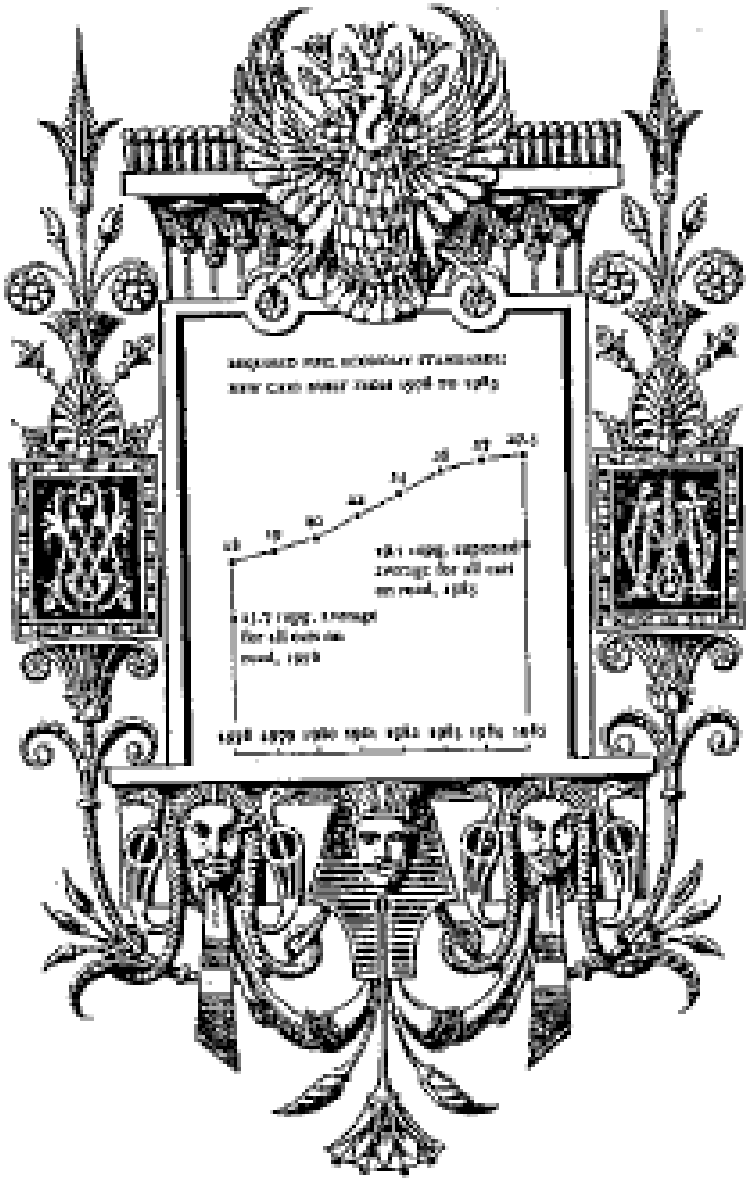
$$Lie\ Factor = \frac{783}{53} = 14{,}8\%$$



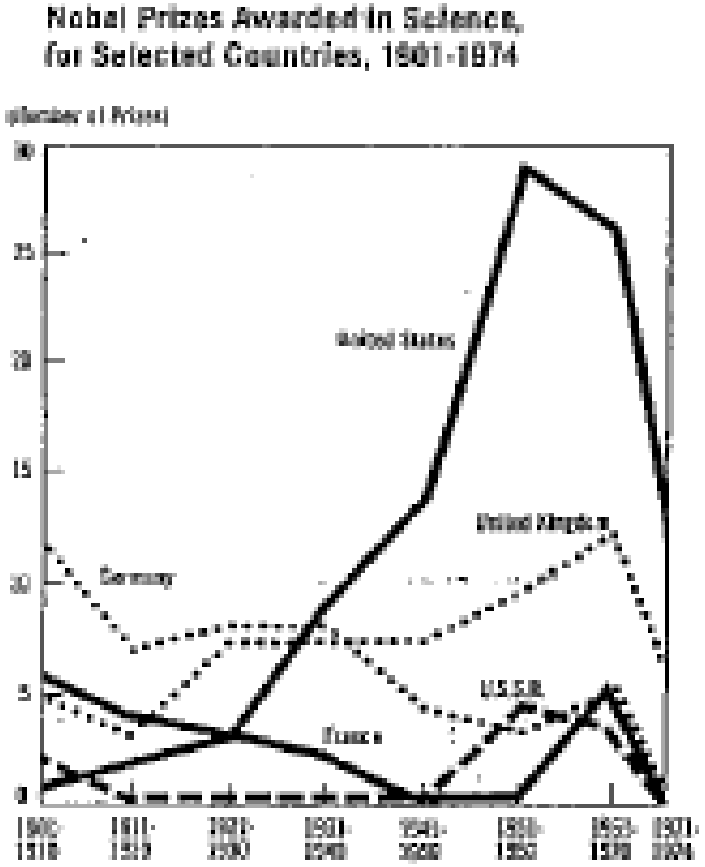
First Principle:



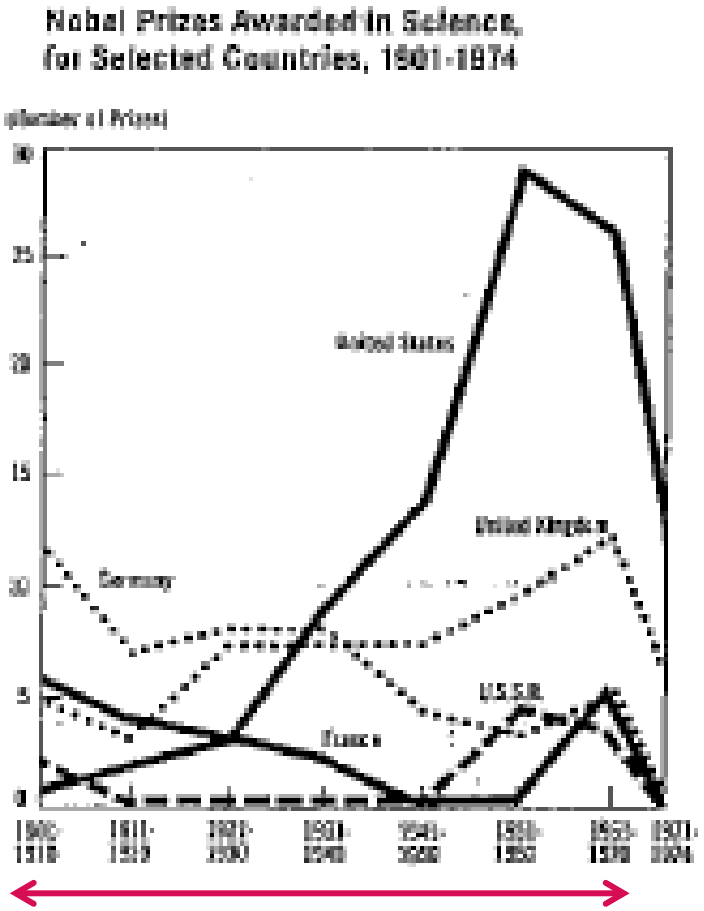
First Principle:



Design and Data Variation:

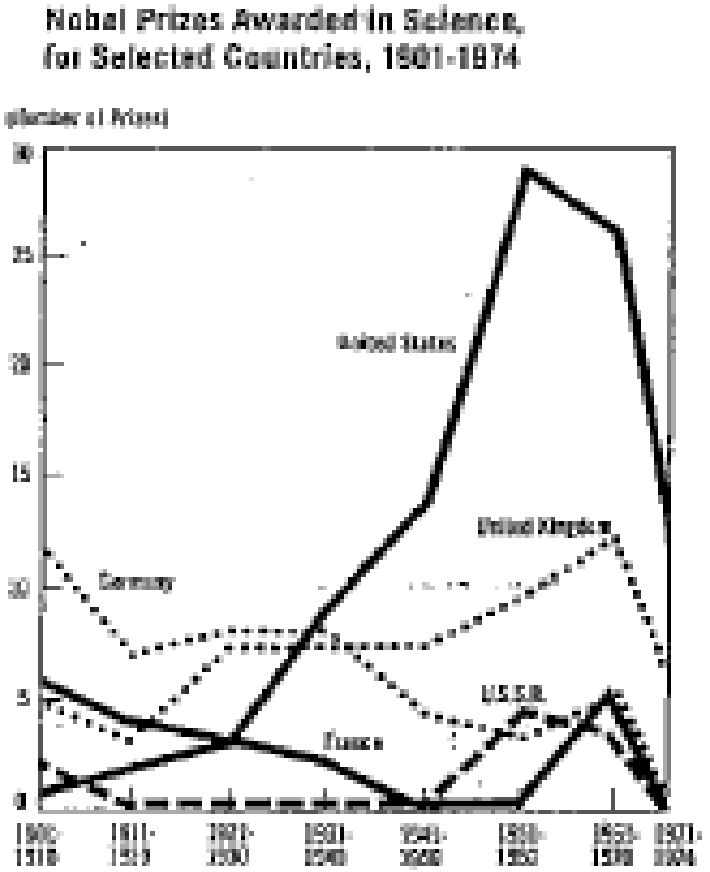


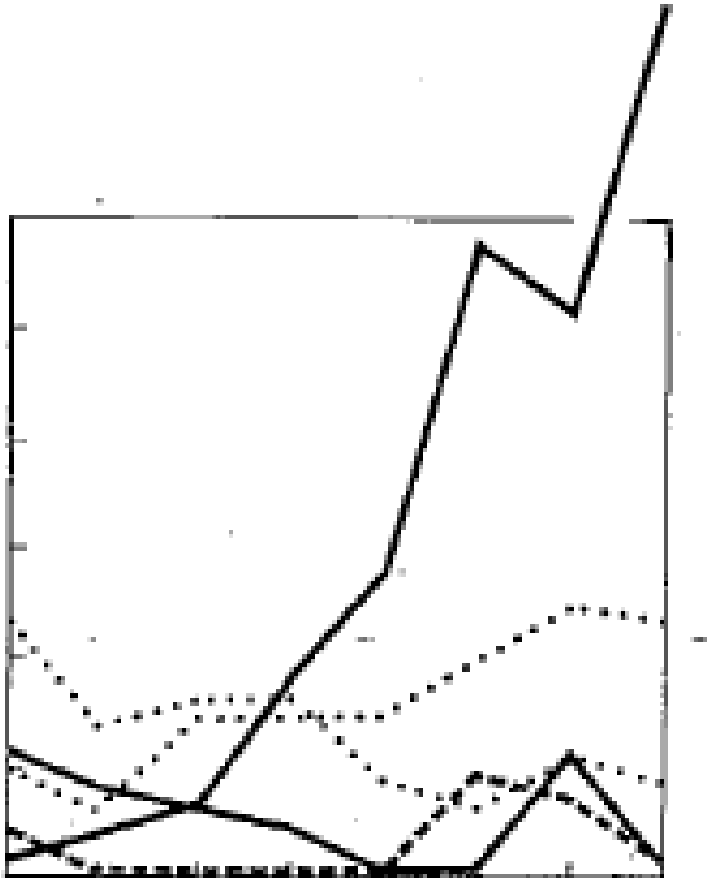
Design and Data Variation:



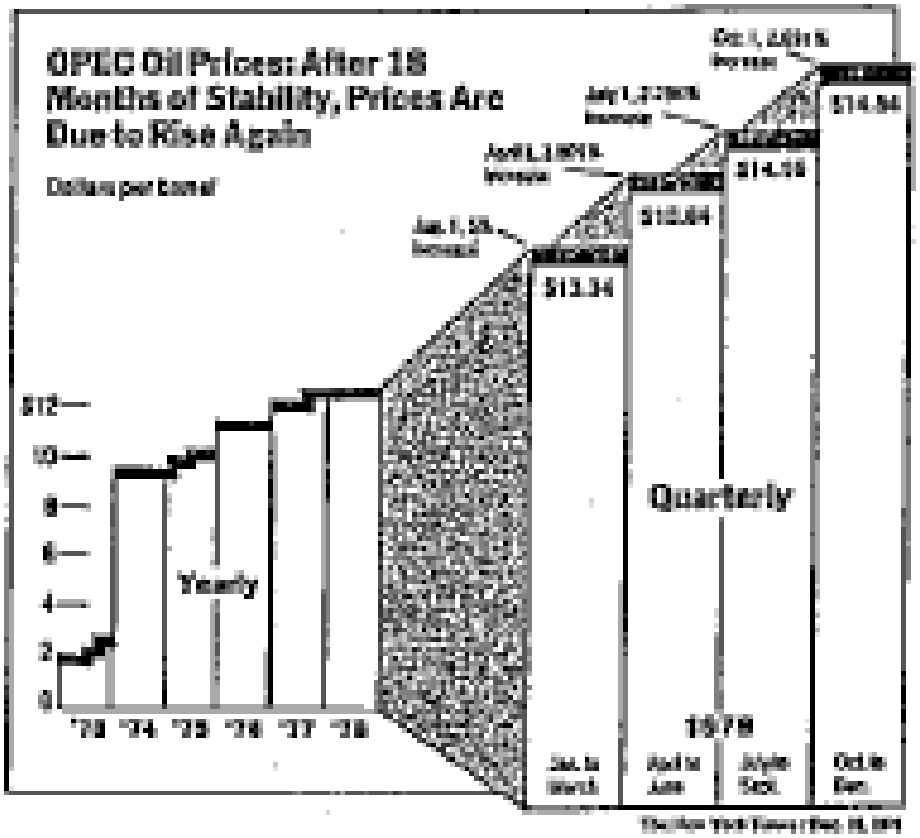
Design and Data Variation:

show data variation, not design variation

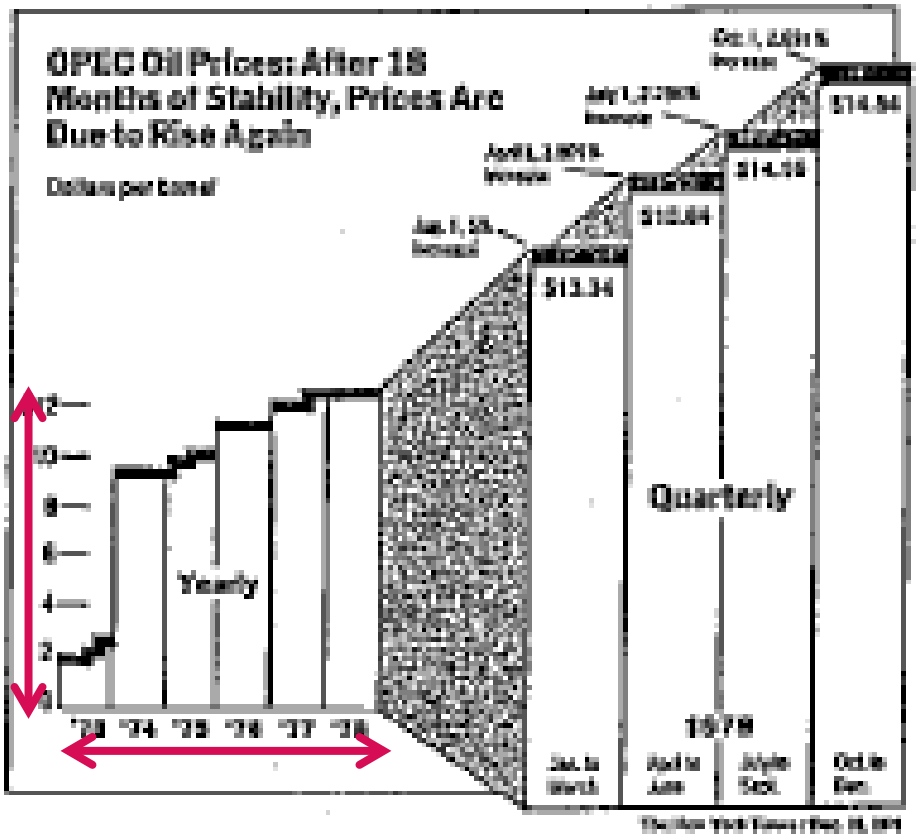




Design and Data Variation:



Design and Data Variation:

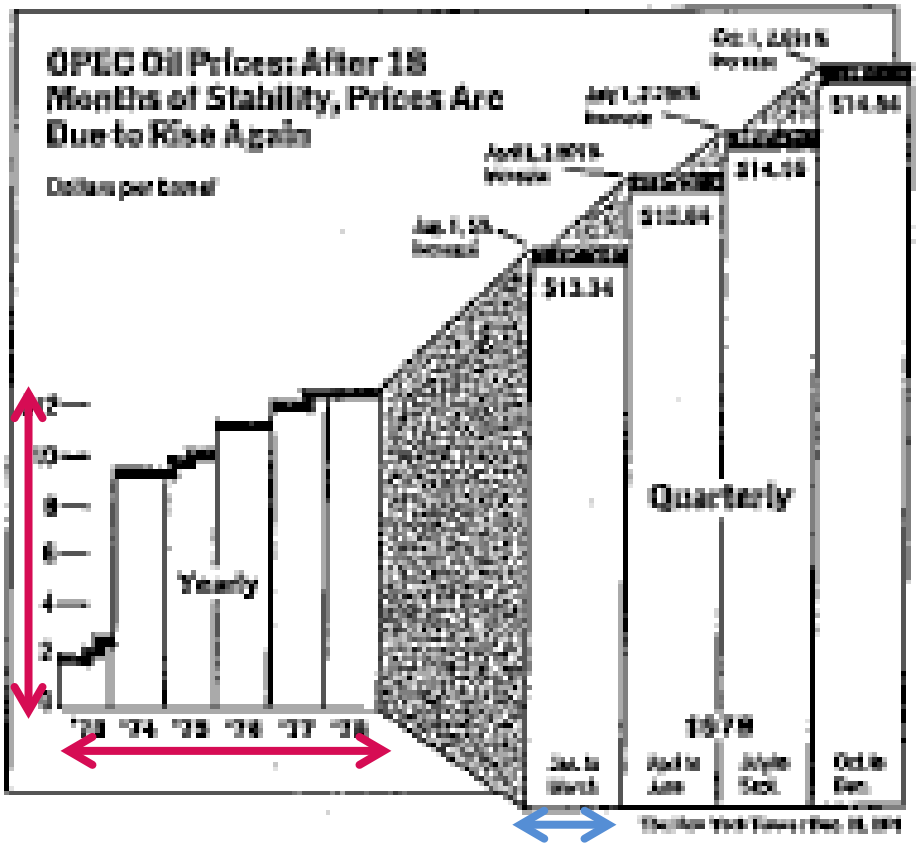


Prices:

‘73-78’

1inch=$8.00

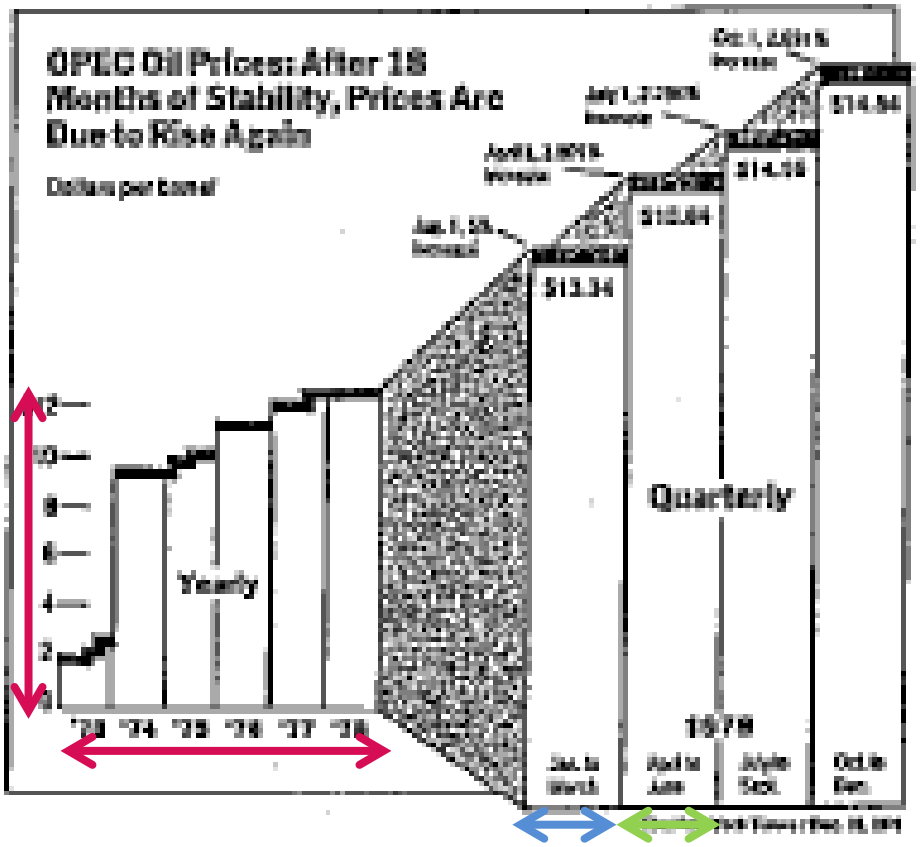
Design and Data Variation:



Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73

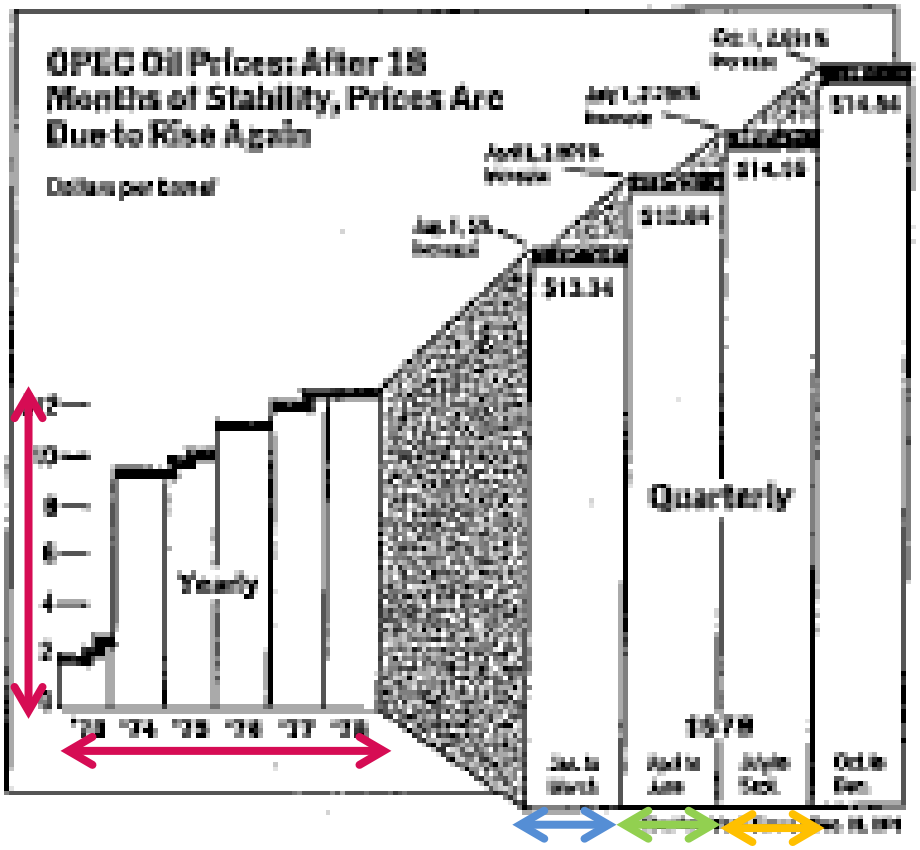
Design and Data Variation:



Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37

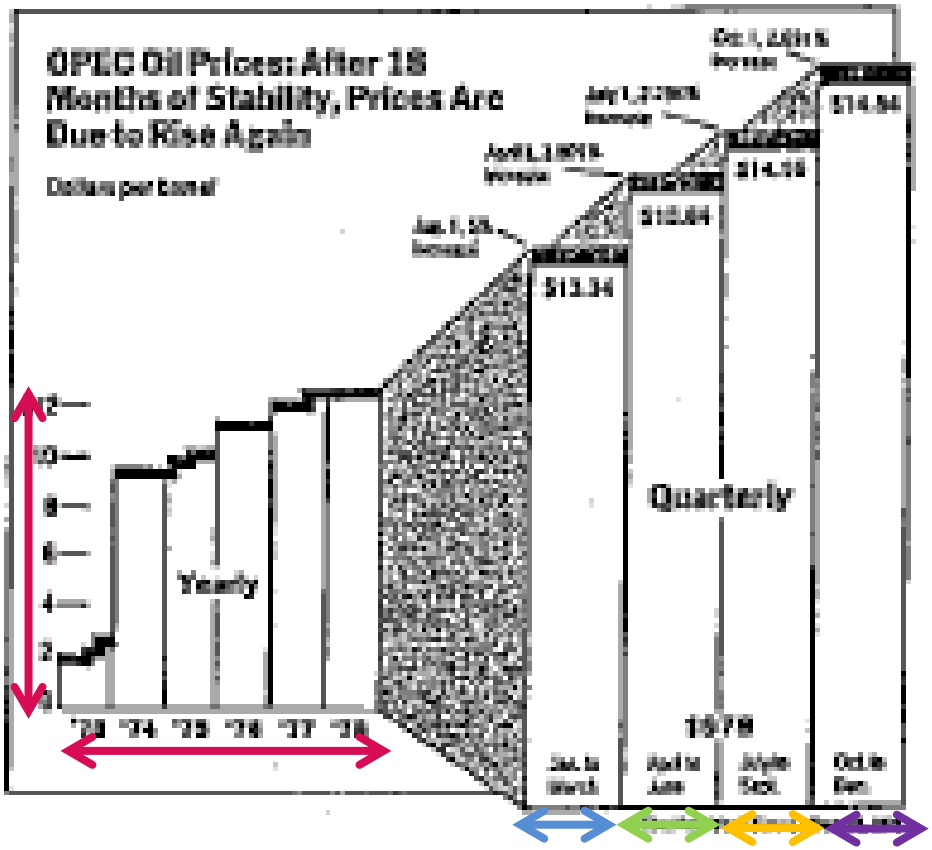
Design and Data Variation:



Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37
July-Sept	1inch=$4.16

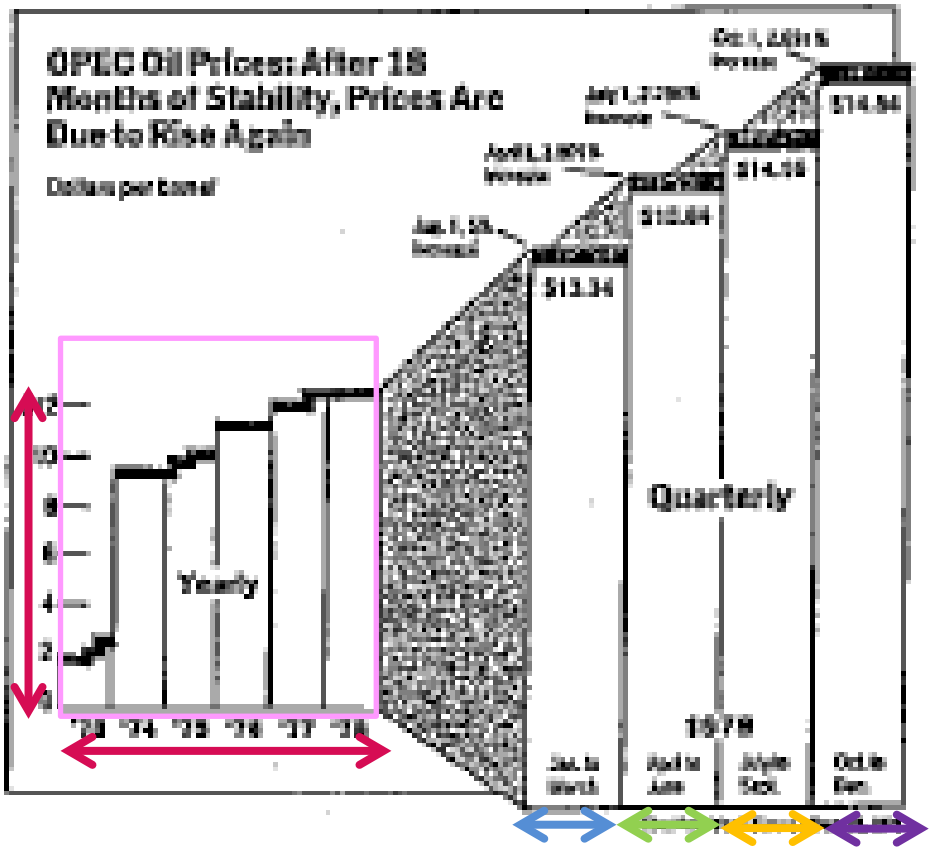
Design and Data Variation:



Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37
July-Sept	1inch=$4.16
Oct-Dec	1inch=$3.92

Design and Data Variation:



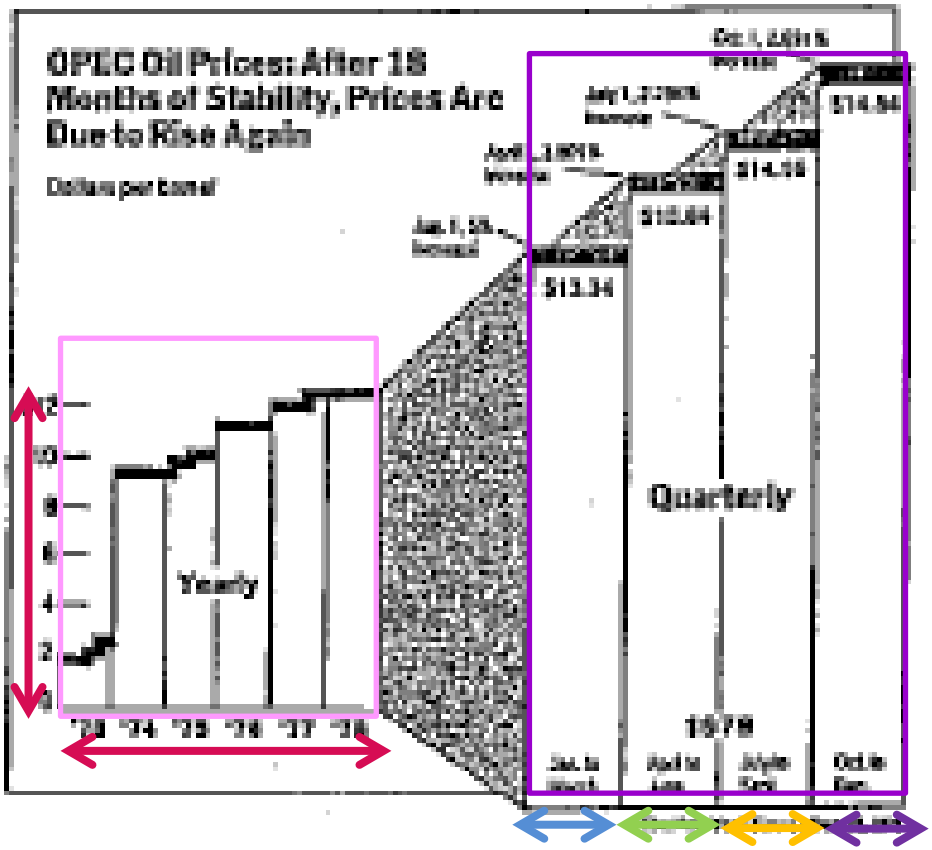
Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37
July-Sept	1inch=$4.16
Oct-Dec	1inch=$3.92

Time:

1973-1978	1inch=3,8yrs

Design and Data Variation:



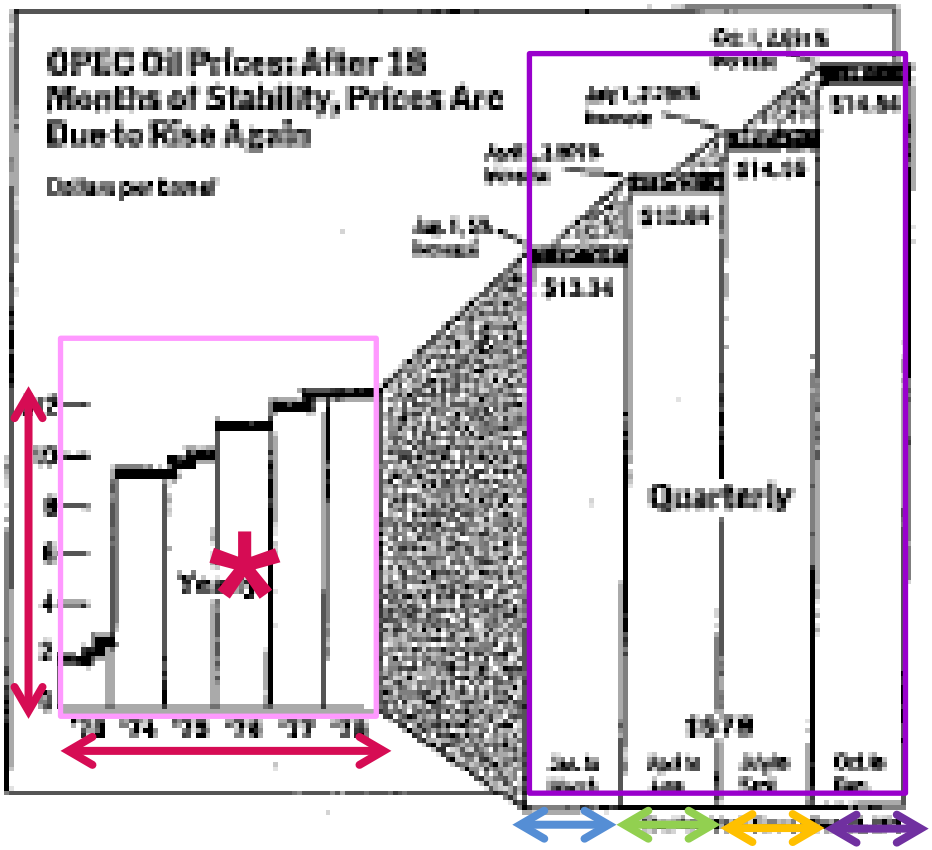
Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37
July-Sept	1inch=$4.16
Oct-Dec	1inch=$3.92

Time:

1973-1978	1inch=3,8yrs
1979	1inch=0.57yrs

Design and Data Variation:



Prices:

'73-78'	1inch=$8.00
Jan-March	1inch=$4.73
Apr-June	1inch=$4.37
July-Sept	1inch=$4.16
Oct-Dec	1inch=$3.92

Time:

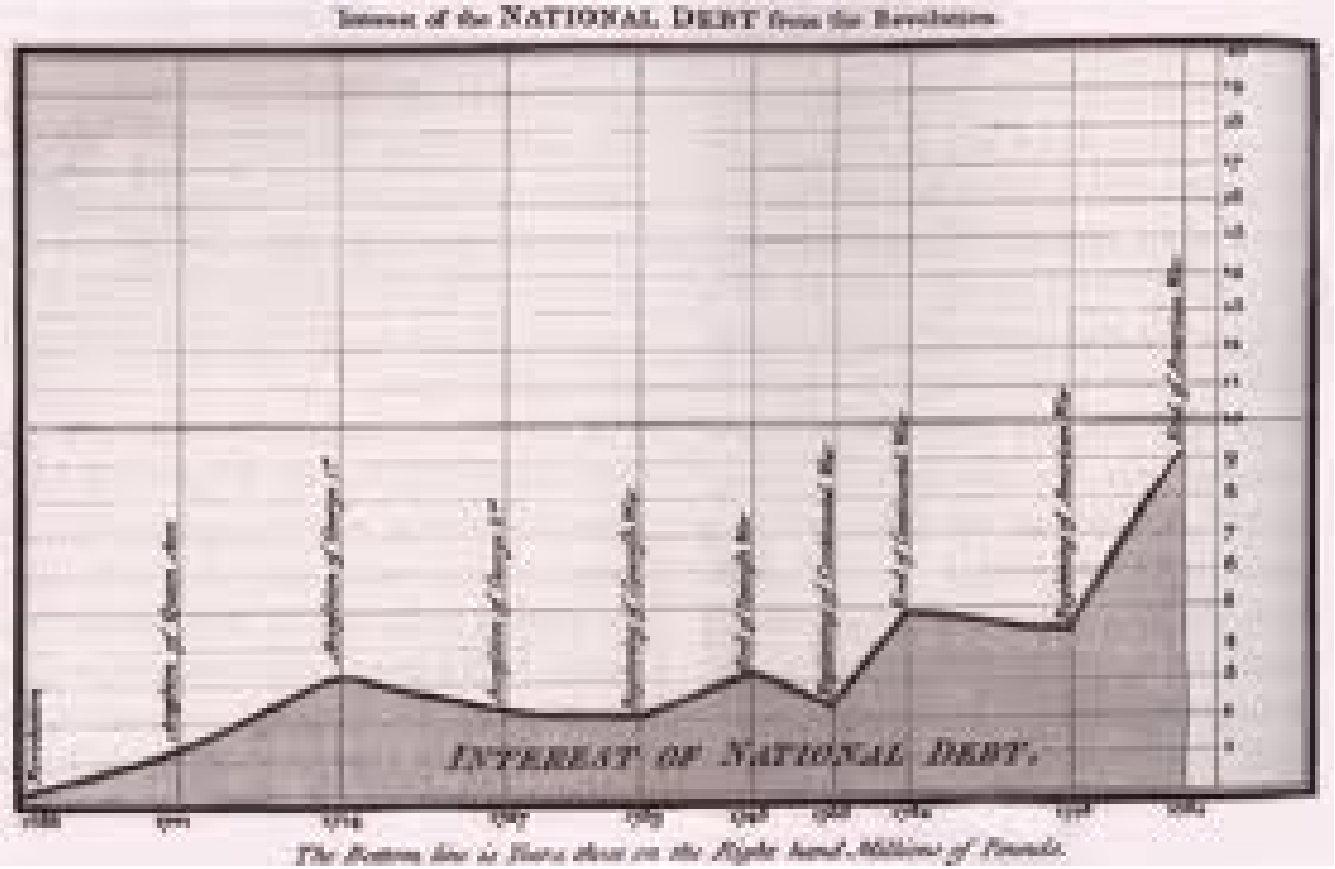
1973-1978	1inch=3,8yrs
1979	1inch=0.57yrs

The Case of Skyrocketing Government Spending:

-Playfair, 1786

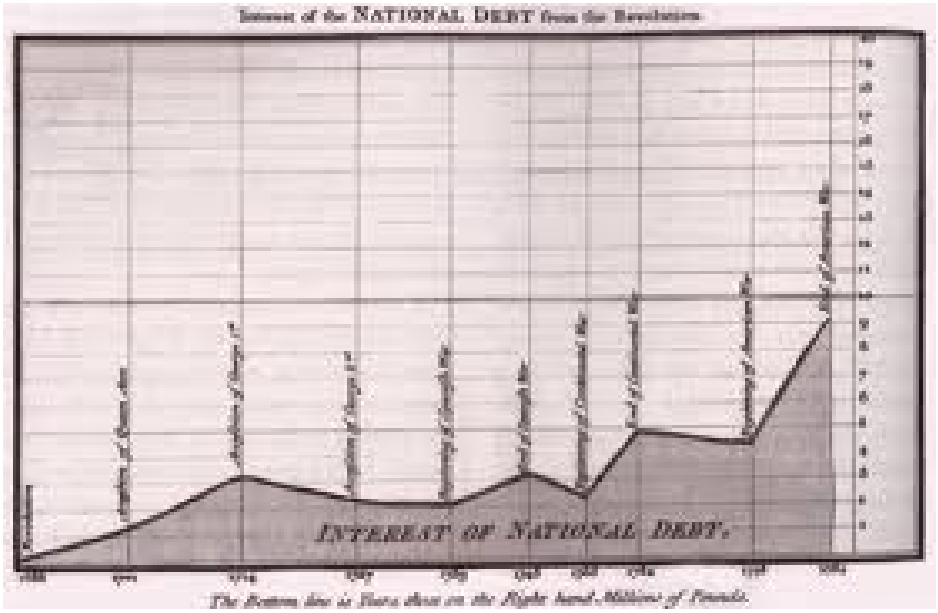
The Case of Skyrocketing Government Spending:

-Playfair, 1786



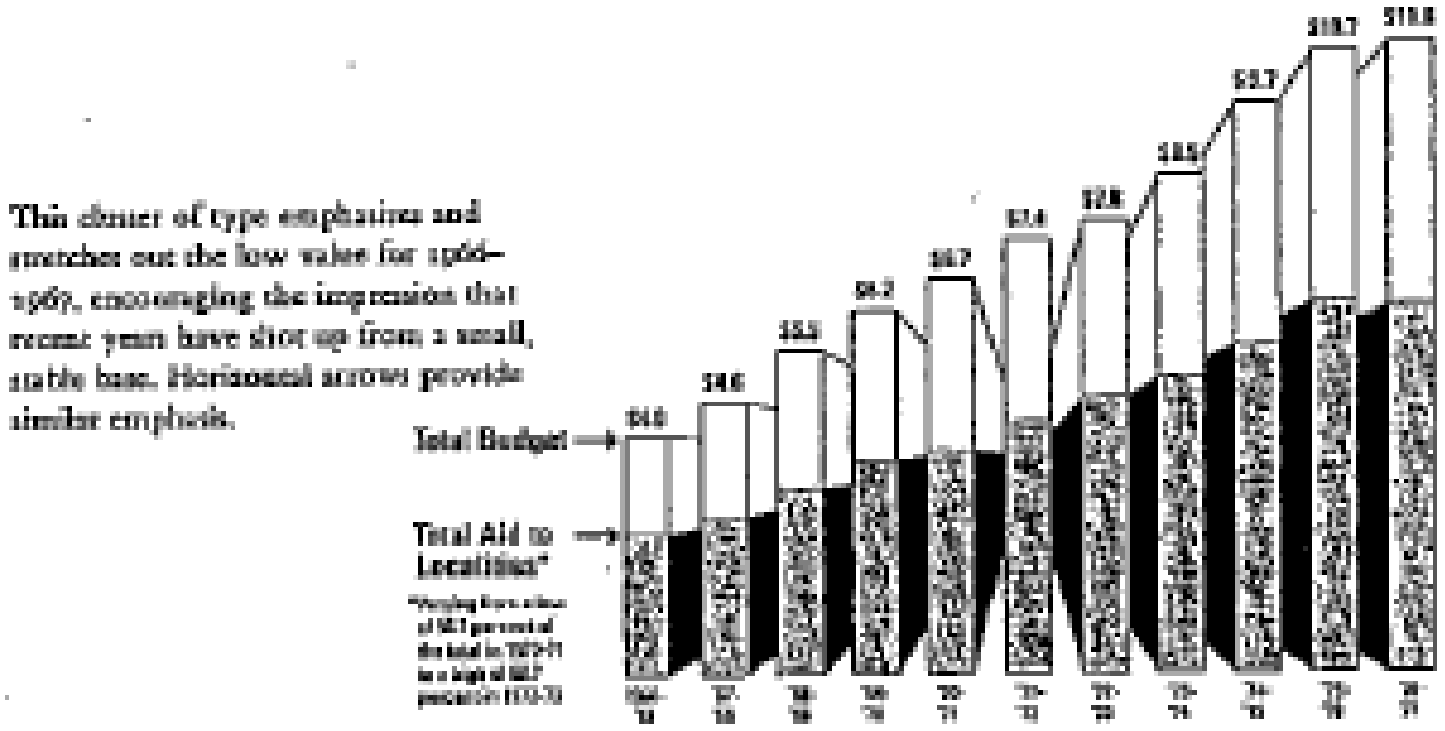
The Case of Skyrocketing Government Spending:

-Playfair, 1786



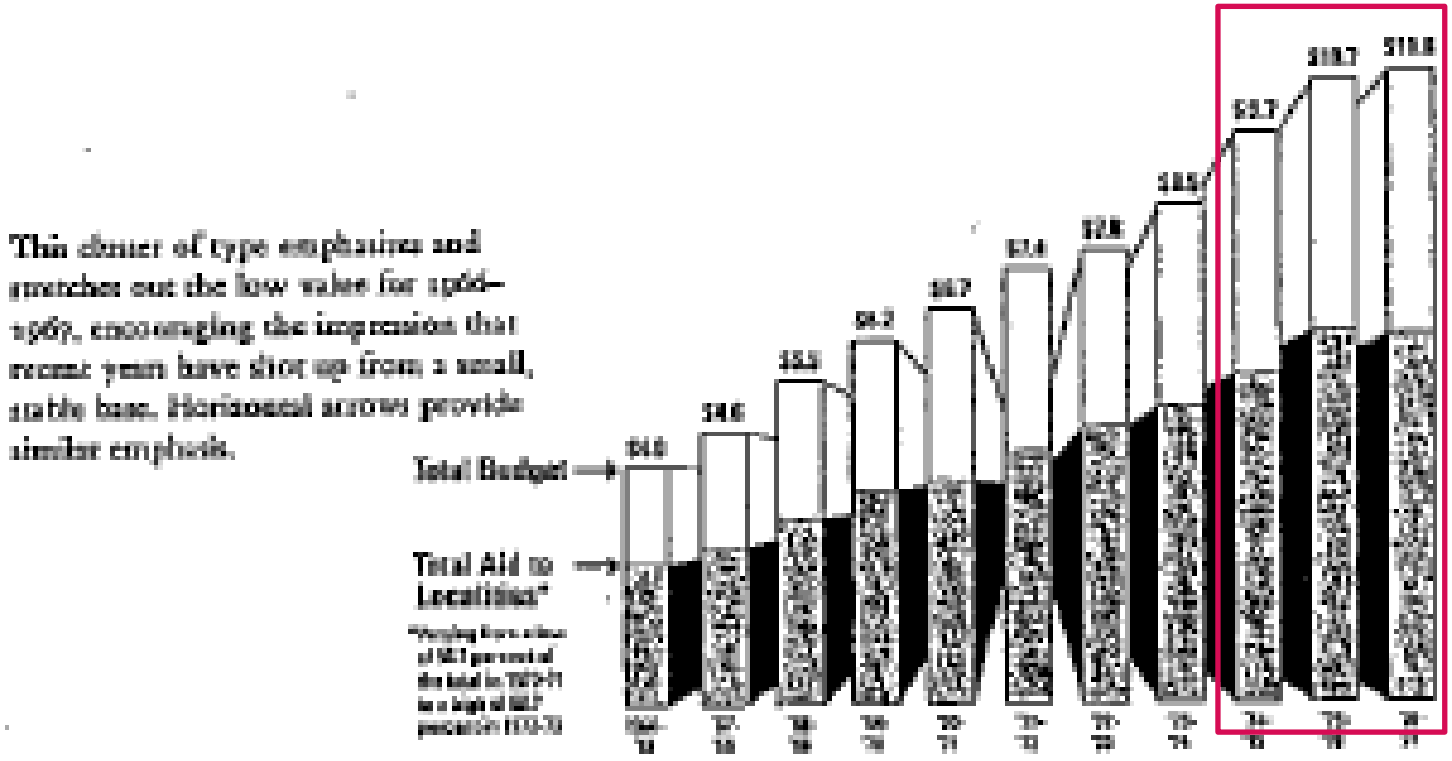
The Case of Skyrocketing Government Spending:

-New York Times, Feb 1, 1976



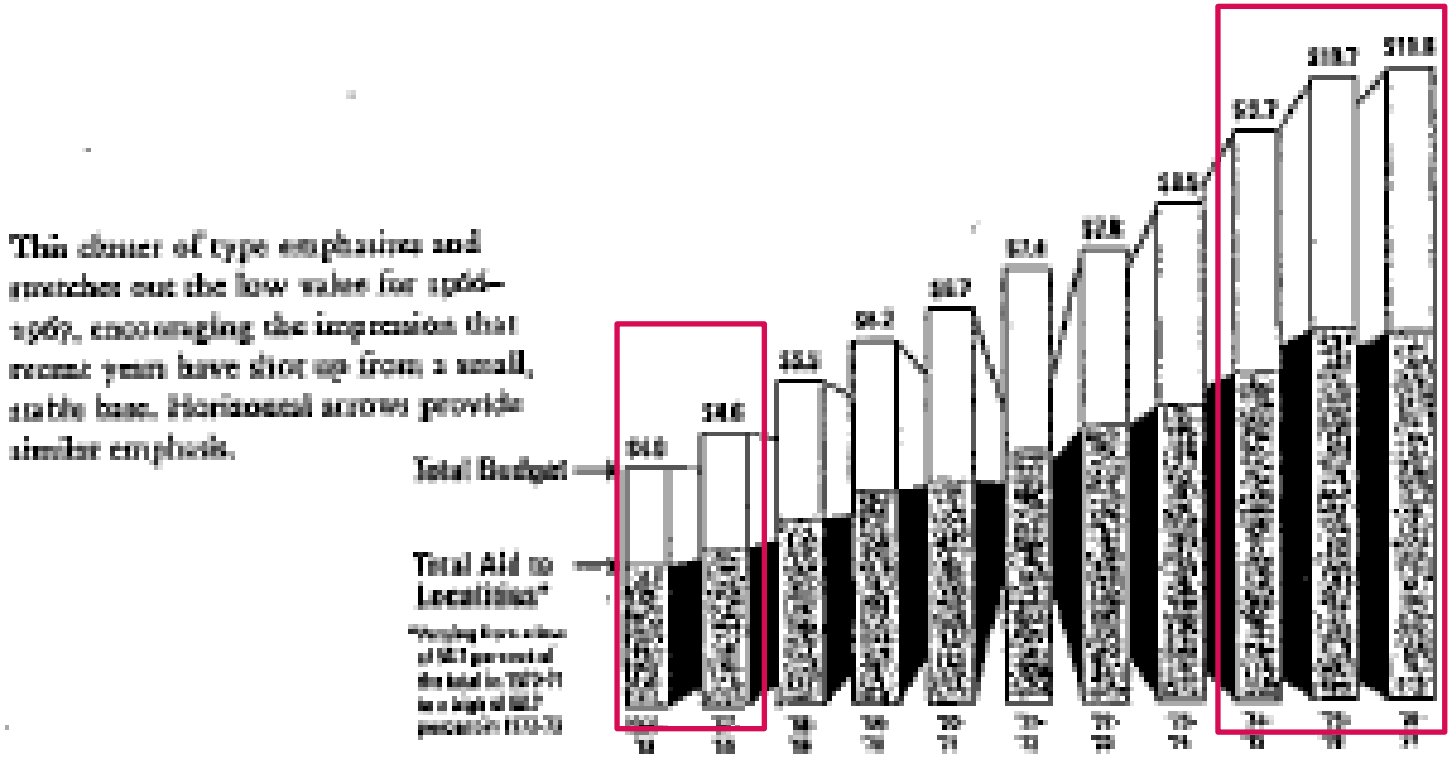
The Case of Skyrocketing Government Spending:

-New York Times, Feb 1, 1976



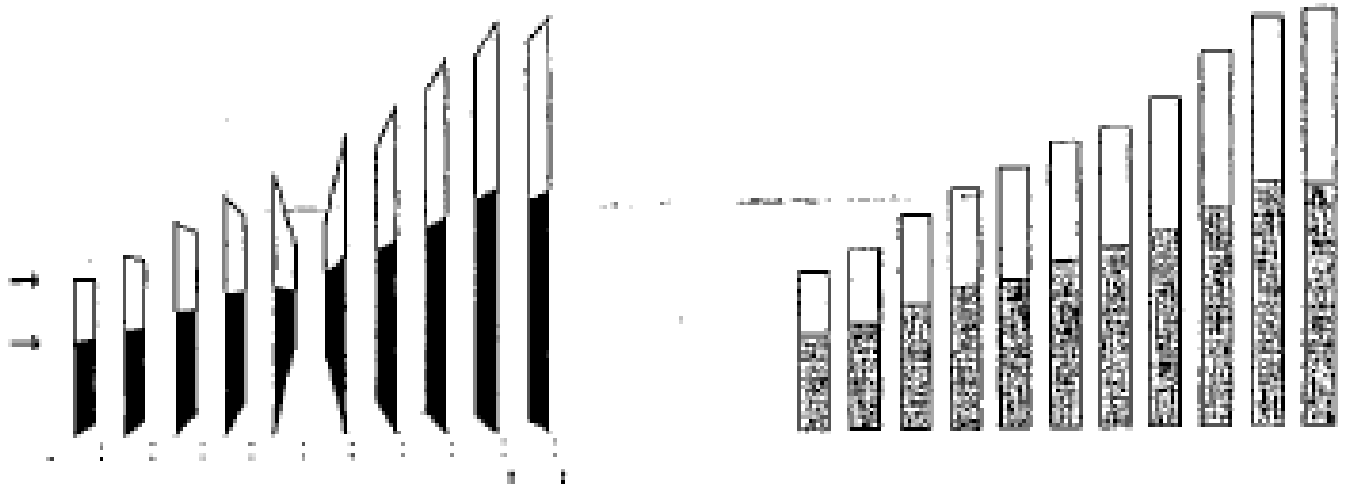
The Case of Skyrocketing Government Spending:

-New York Times, Feb 1, 1976



The Case of Skyrocketing Government Spending:

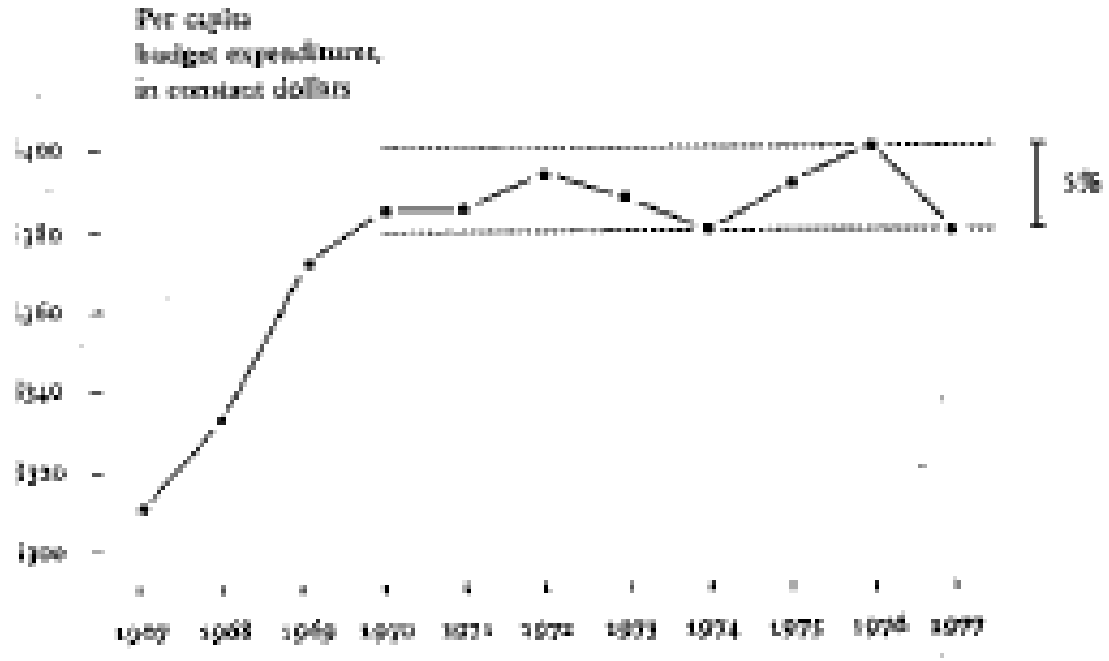
-New York Times, Feb 1, 1976



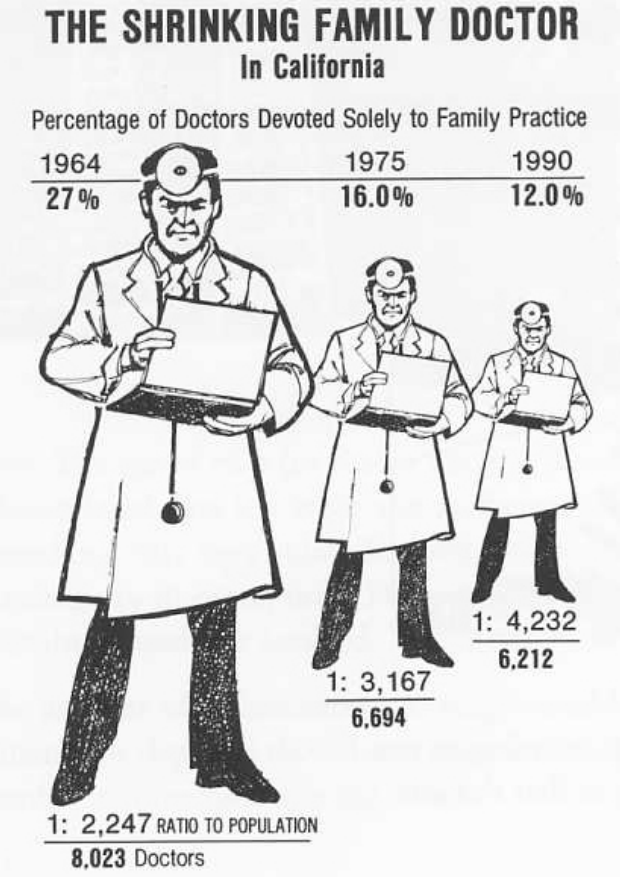
The Case of Skyrocketing Government Spending:

-New York Times, Feb 1, 1976

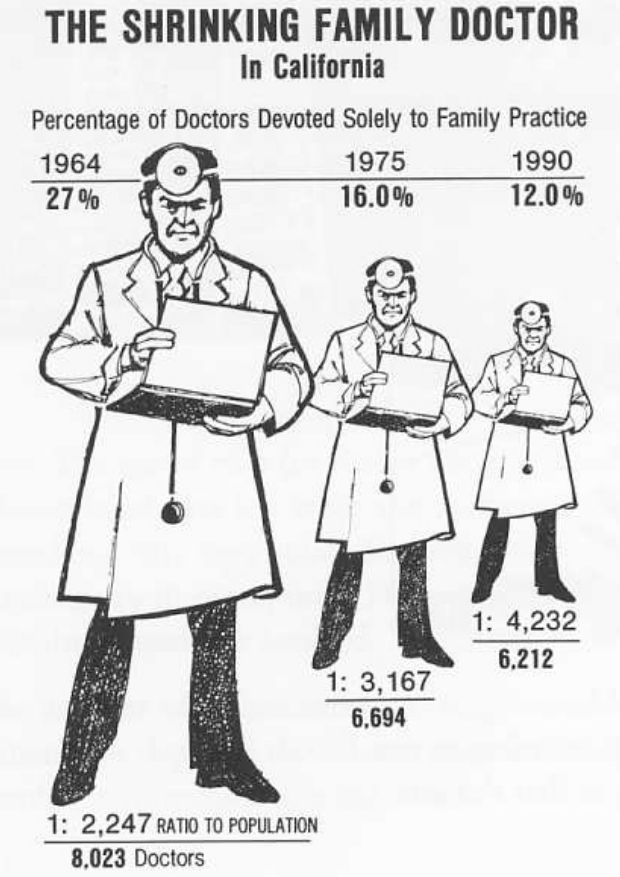
The Principle: In time-series of money, deflated and standardized units of monetary measurement are better than nominal units.



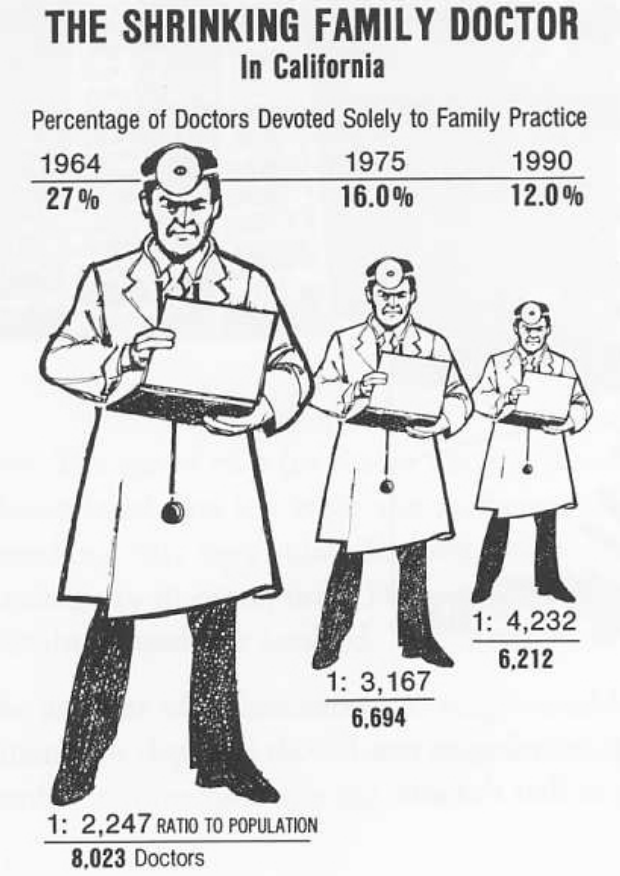
Visual Area and Numerical Measure:



Visual Area and Numerical Measure:



Visual Area and Numerical Measure:



Visual Area and Numerical Measure:

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

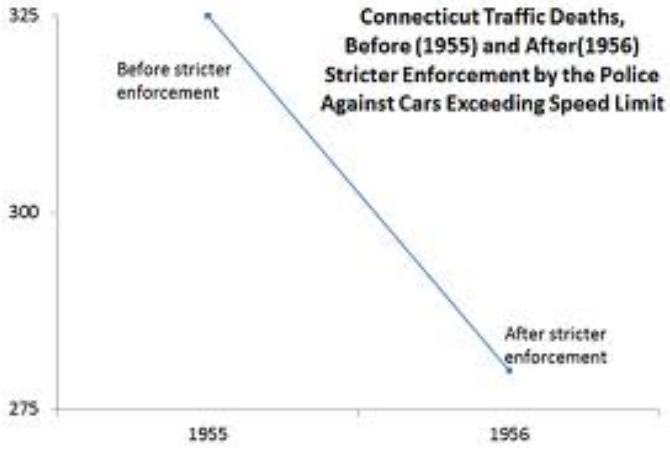
The number of information-carrying dimensions **should not exceed** the number of dimensions in the data.

Context is Essential for Graphical Integrity:

Graphics must not quote **data out of context.**

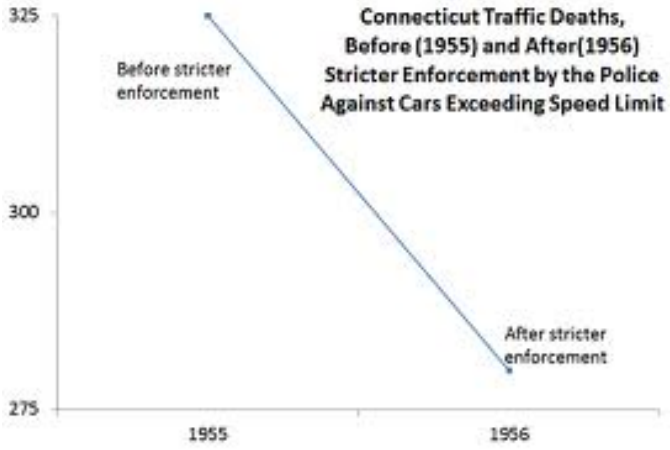
Context is Essential for Graphical Integrity:

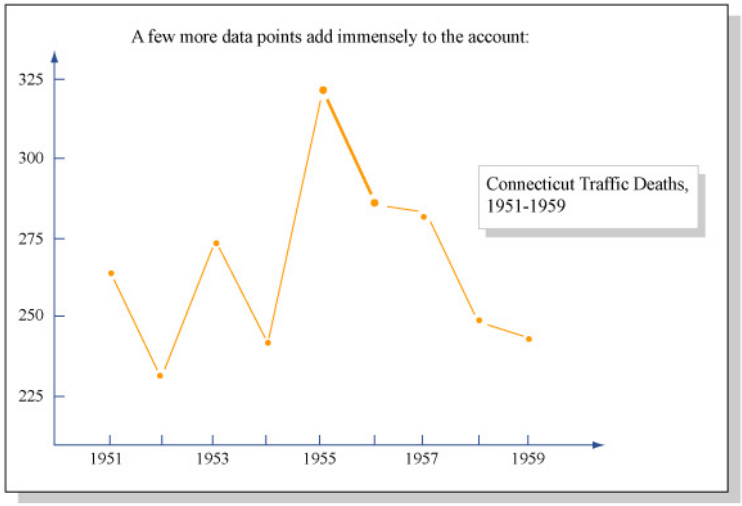
Graphics must not quote **data out of context.**



Context is Essential for Graphical Integrity:

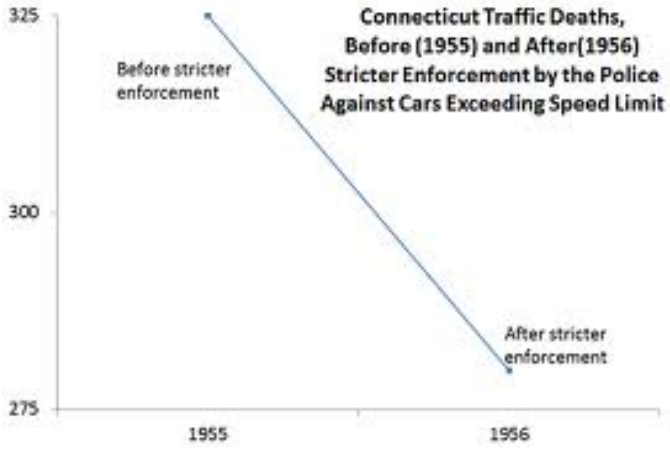
Graphics must not quote **data out of context.**

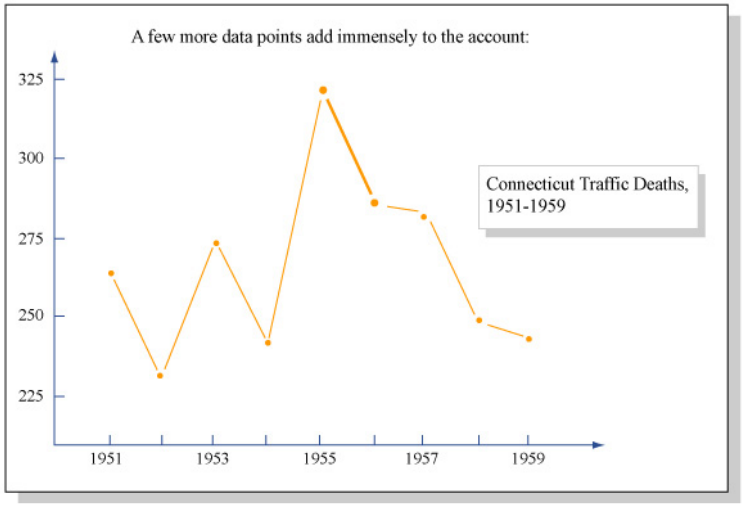




Context is Essential for Graphical Integrity:

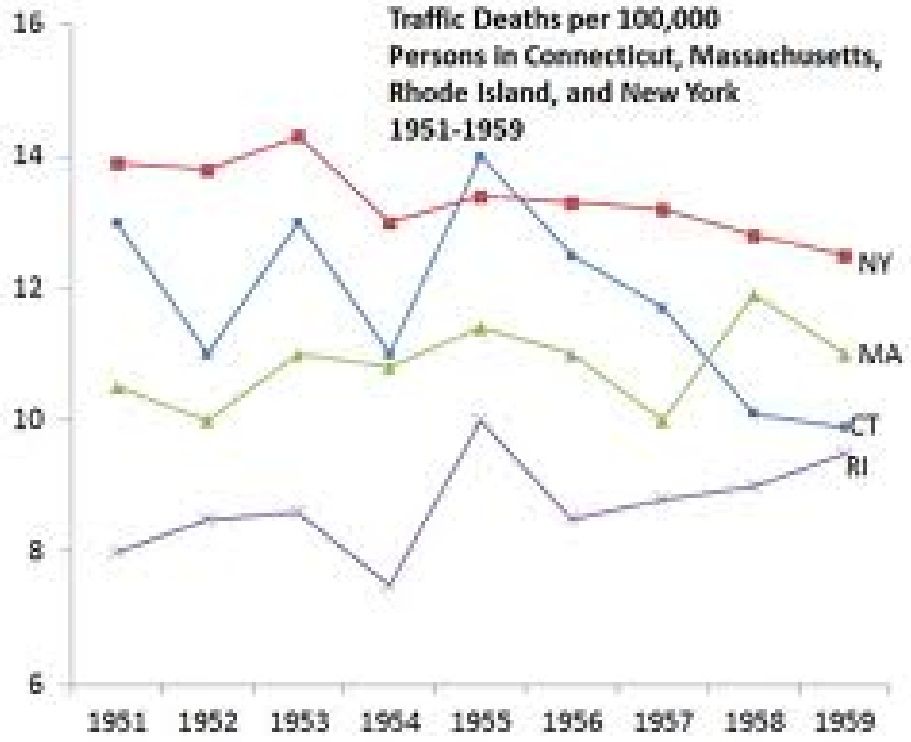
Graphics must not quote **data out of context.**





Context is Essential for Graphical Integrity:

Graphics must not quote data out of context.



Conclusion:

Lying graphics cheapen the graphical art everywhere.

Conclusion:

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

Conclusion:

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
2. Clear, detailed and thorough labeling. Write out explanations and label important events.

Conclusion:

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
2. Clear, detailed and thorough labeling. Write out explanations and label important events.
3. Show **data variation**, not design variation.

Conclusion:

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
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.

Conclusion:

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
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.

Conclusion:

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
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.