Design World

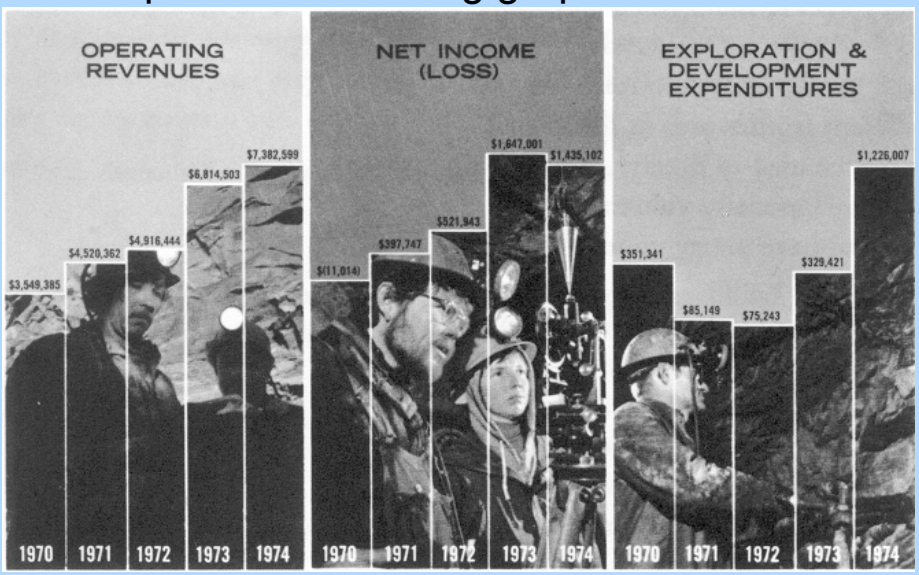
Graphical Integrity

largely from Edward Tufte,
The Visual Display of Quantitative Information, Graphics Press, 1983.

Graphical integrity

- Graphics can be a powerful communication tool
- Lies and falsehoods are possible
- Much focus on this ‘how to lie with maps’ or ‘statistics’

Examples of misleading graphics



Where is the bottom line? What is happening in 1970?

Day Mines, Inc. 1974 Annual Report, p1 (Tufte, 1983, p54)

Sheelagh Carpendale

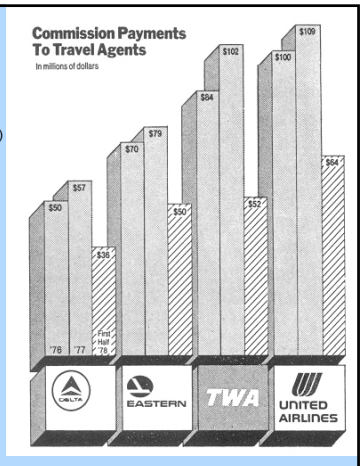
Misleading graphics

New York Times, August 8, 1978, p.D-1 (Tufte, 1983, p54)

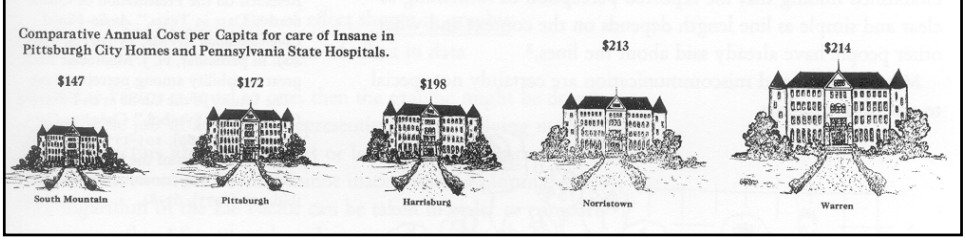
What is the first impression of the airlines relative success in 1978?

Order of numbers?

Magnitude of numbers? Impression?



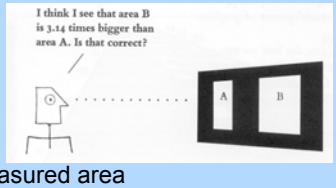
Pittsburgh Civic Commission, Report on Expenditures of the Department of Charities (Pittsburgh, 1911), p.7 (Tufte, 1983, p54)



Achieving graphical Integrity

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

- Is the magnitude of 'visual representations' as physically measured on the graphic?
- Or the perceived magnitude?



Approach

Conduct a study of visual perception of the graphics.

Circles – perceived area grows more slowly than measured area

reported perceived area = (actual area)^{x}, where x = 0.8+/-0.3

Lines -

Lie Factors

Given perceptual difficulties – strive for uniformity (predictability) in graphics (p56)

- 'the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.'
- 'Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.'

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

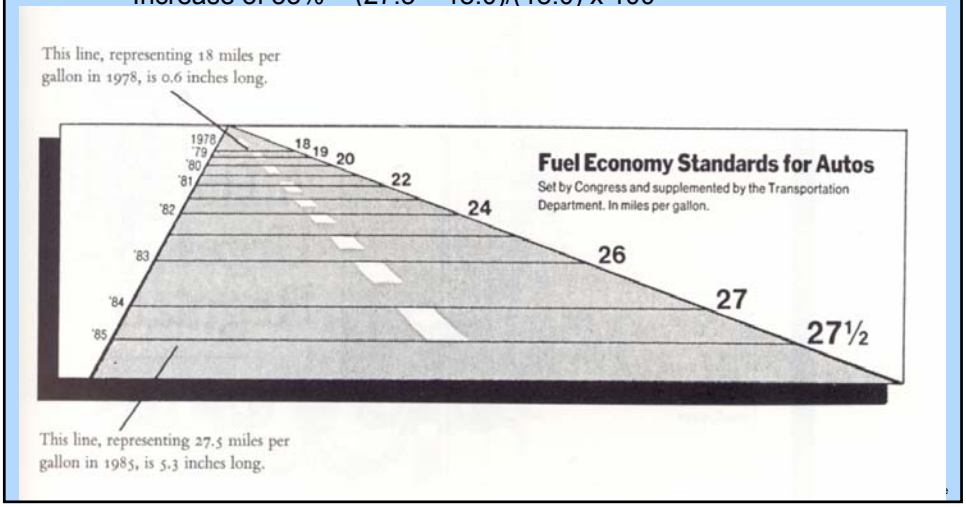
Lie factor of 1 – is desirable – lie factors > 1.05 or < 0.95 go beyond plotting errors

Extreme example

Fuel economy standards for automobiles

18 miles/gallon in 1978 to 27.5 miles/gallon in 1985

Increase of 53% = (27.5 – 18.0)/(18.0) x 100

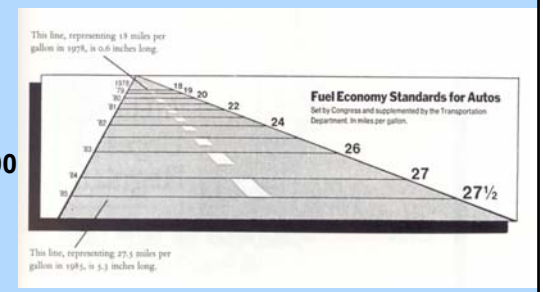


Extreme example

Graphic increase

783% = (5.3 – 0.6)/(0.6) x 100

Lie Factor = 783/53 = 14.8



Additional confounding factors

Usually the future is in front of us

Dates remain same size and fuel factors increase

Includes perspective distortion – how to read change in perspective

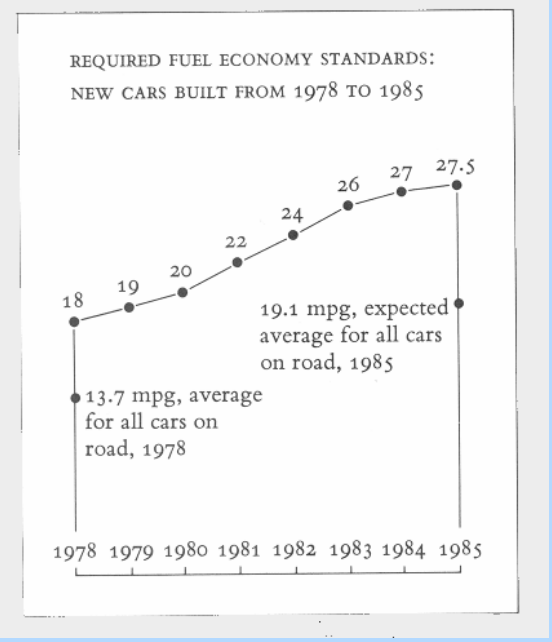
Extreme example

Same data

Simple graphic

Note

- Gradual improvement for 1st 2 years
- Increased improvement for next 3 years
- Tapering off for last 2 years



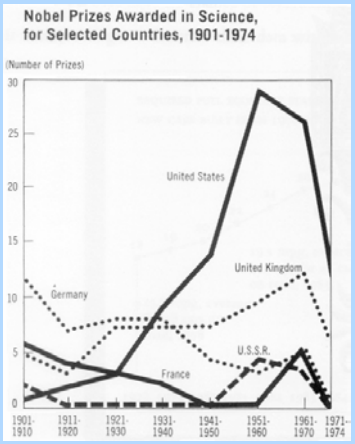
Extrapolation

a graphic generates visual expectations – deception can result from incorrect extrapolation of visual expectations

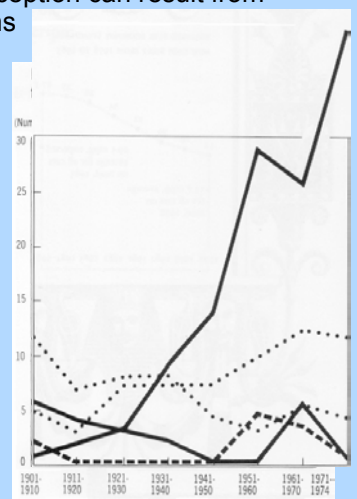
1st seven intervals are 10 years

The last interval is 4 years

Gives a false sense of decline



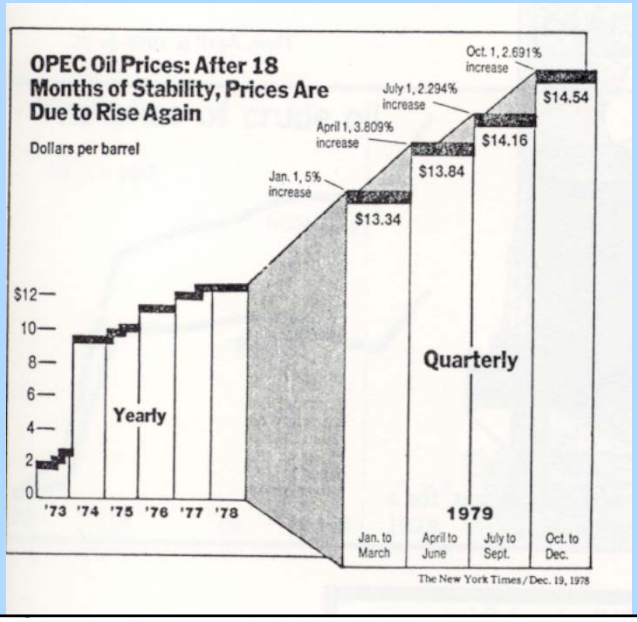
National Science Foundation, Science Indicators, 1974 (Washington D.C., 1976), p.15, (Tufte, 1983, p60)



Accurate data for the next 10 years

Design Variation vs Data Variation

New York Times, Dec. 19, 1978, p.D-7 (Tufte, 1983, p61)



Sheelagh Carpendale

Design Variation vs Data Variation

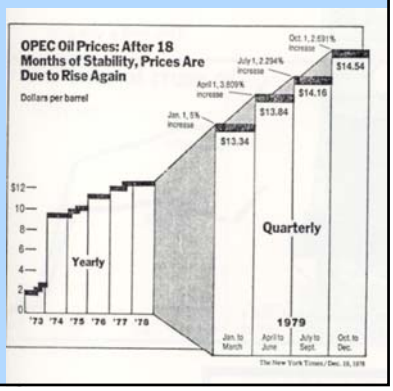
5 different vertical scales show price

1973 -1978	$8.00
Jan. – Mar. 1979	$4.73
Apr. – June 1979	$4.37
Jul. – Sept. 1979	$4.16
Oct. – Dec. 1979	$3.92

2 different horizontal scales show passage of time

1973-1978	3.8 years
1979	0.57 years

With both scales shifting the distortion is multiplicative

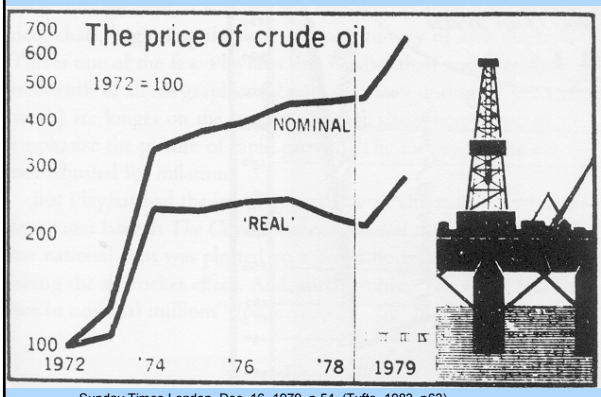


National Science Foundation, Science Indicators, 1974 (Washington D.C., 1976), p.15, (Tufte, 1983, p60)

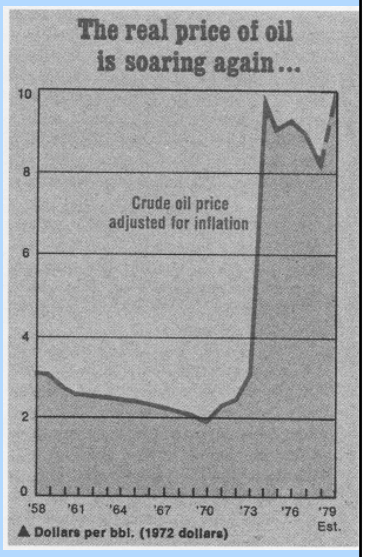
Design Variation vs Data Variation

Graphics that actually represent the data and take into consideration inflation adjusted money.

Business week includes more context



Sunday Times London, Dec. 16, 1979, p.54, (Tufte, 1983, p63)



Business Week, Apr. 9, 1979, p.99, (Tufte, 1983, p63)

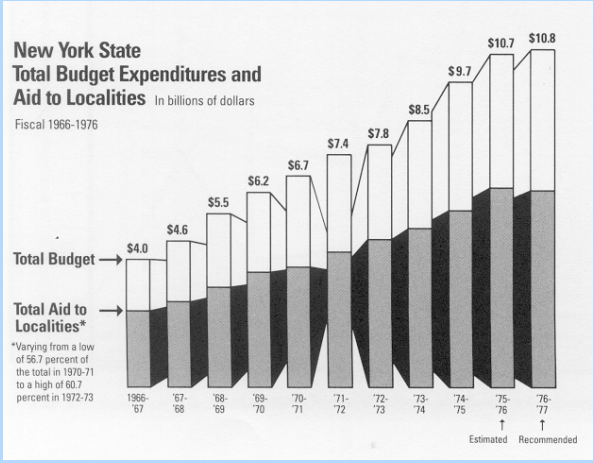
Sheelagh Carpendale

Chartjunk

These 3 parallelepipeds have been placed in front of the other 8 – giving the impression that they tower over the others

Clustering and horizontal arrows provide an impression of small stable base

Squeezed down block of type contributes to impression of small 'squeezed down budgets in the sixties



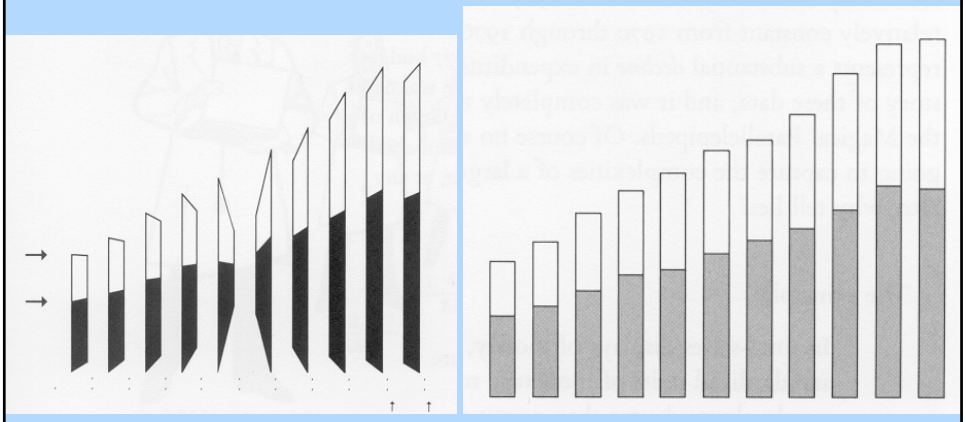
Up-arrows emphasize recent growth

New York Times, Feb. 1, 1976, p.IV-6, (Tufte, 1983, p66)

Sheelagh Carpendale

Chartjunk

Removing chart junk creates a calmer view



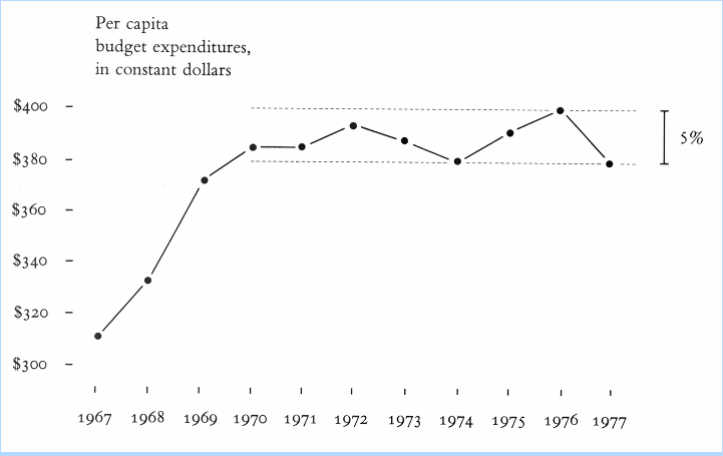
However, also statistical bias, introduced by ignoring increase in populations and inflation, is still present.

Tufte, 1983, p67

Sheelagh Carpendale

Chartjunk

Comparisons need to be made with comparable data



Removing chartjunk, and statistical bias tells the real story – a 20% increase 67 to 70, relative stability 70 to 76 and a decrease in spending per capita in 77

Principle: deflated and standardized monetary units are nearly always better the nominal units

Tufte, 1983, p68

Sheelagh Carpendale

Visual Area and Numerical measure

Use of area to portray 1D data can be confusing

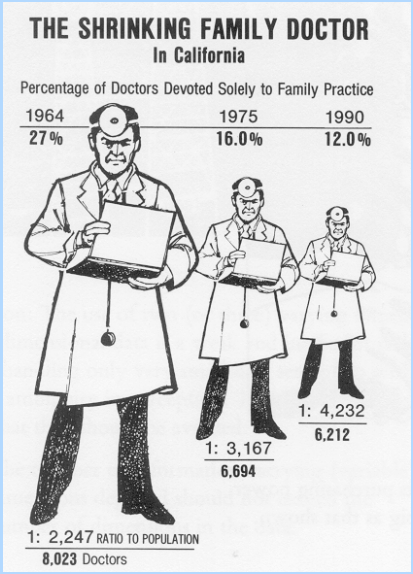
- Area has 2 dimensions

The ‘incredible’ shrinking family doctor

Lie factor of 2.8

Plus perspective distortion

Plus incorrect horizontal spacing



Los Angeles Times, August 5, 1979 p.3, (Tufte, 1983, p69)

Sheelagh Carpendale

Visual Area and Numerical measure

Use of area to portray 1D data can be confusing

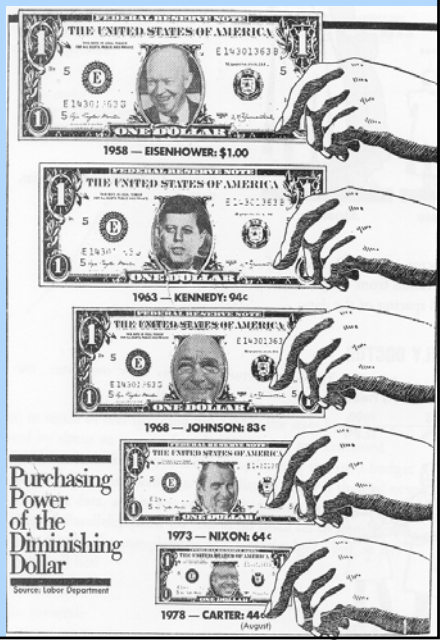
- Area has 2 dimensions

The ‘incredible’ shrinking dollar

The size of the dollar is adjusted in both height and width

The affect on the area in multiplicative

A more accurate representation of the 1978 dollar would be about twice the size of the one in this chart



Washington Post, Oct. 25, 1978 p.1, (Tufte, 1983, p70)

Visual Area and Numerical measure

Use of area to portray 1D data can be confusing

-Area has 2 dimensions

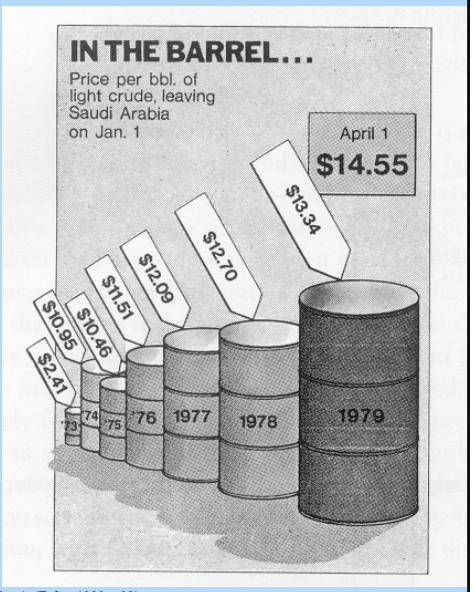
-There are considerable ambiguities of how people perceive area (2D) and then convert that to 1D data

-Even more so with volumes (3D)

By surface area lie factor is 9.4

By volume – lie factor is 59.4 – probably a record

The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.



Time, Apr. 9, 1979 p.1, (Tufte, 1983, p62)

Sheelagh Carpendale

Visual Area and Numerical measure

Multivariate history of Italian post office uses 2 dimensions according to this principle

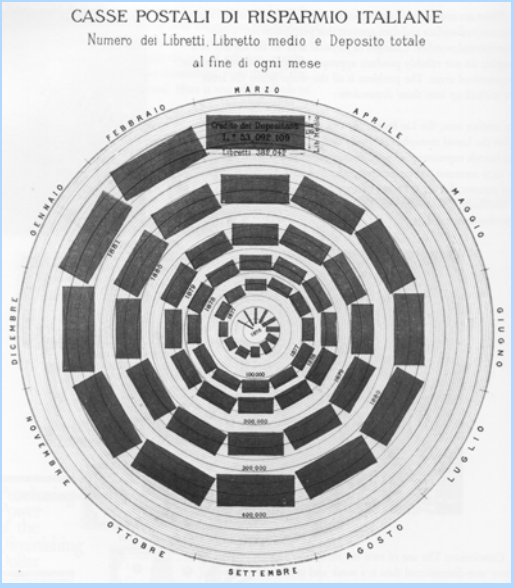
Each month

- width of the rectangle represents the number of postal savings books issued, and

- height represents the average size of the deposits

- therefore area represents the total deposits for each month

The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.



Antonio Gabaglio, Teoria Generale della Statistica (milan 2nd edition 1888), (Tufte, 1983, p72)

Sheelagh Carpendale

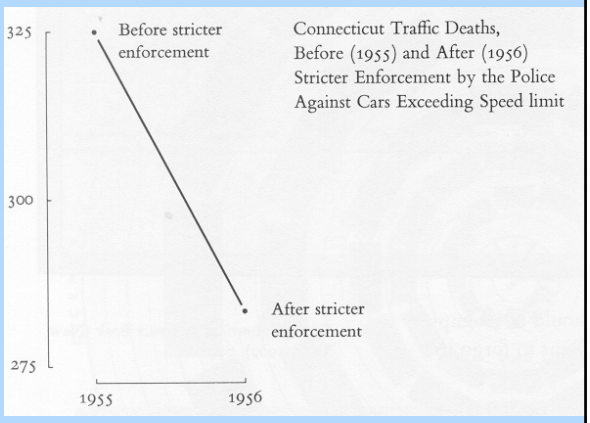
Context is Essential

Graphics must not quote data out of context

Data sparse graphics should provoke suspicion

Graphics often lie by omission

Nearly all important questions are left unanswered by this graphic

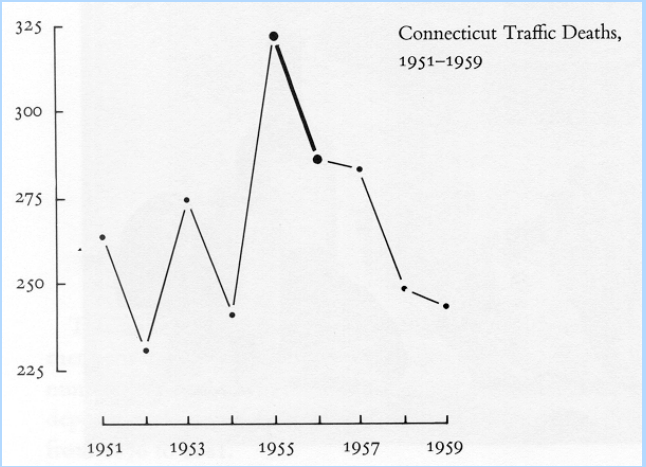


Sheelagh Carpendale

Context is Essential

Graphics must not quote data out of context

A few more data points tell a more complete story

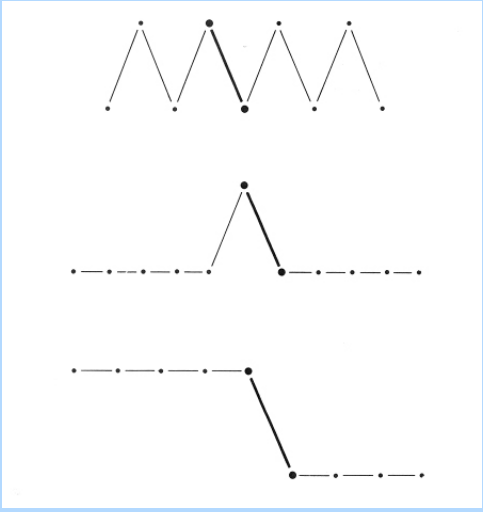


Sheelagh Carpendale

Context is Essential

Graphics must not quote data out of context

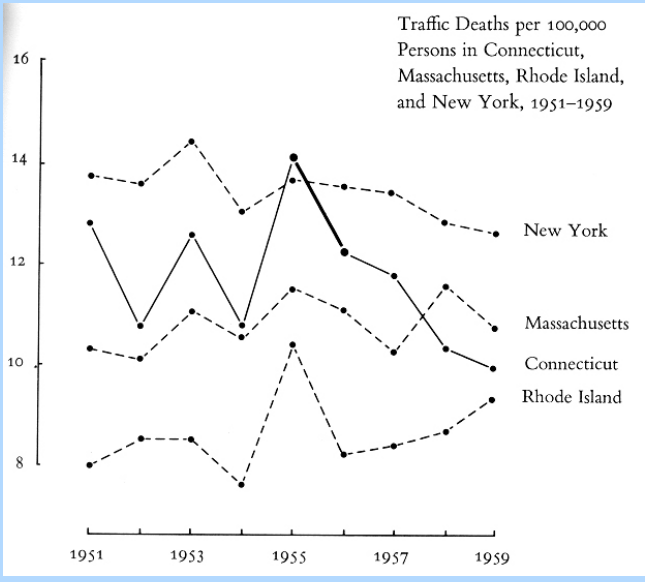
Different data points would tell a different stories



Context is Essential

Graphics must not quote data out of context

Comparisons with adjacent states give more context



Donald T. Campbell and H. Laurence Ross, "The Connecticut Crackdown on Speeding: Time Series Data in Quasi-Experimental Analysis," in Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems* (Reading, Mass., 1970), 110–125.

Graphical Integrity - Summary

- ‘The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the graphic itself. Label important events.
- 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.’

(Tufte, 1983, p77)