



CDS6324 DATA VISUALIZATION

LECTURE 4: GRAPHICAL INTEGRITY, VISUALIZATION DESIGN AND TOOLS



Graphical Excellence & Graphical Integrity



Graphical Excellence

Graphical Excellence



- The well-designed presentations of interesting data are a matter of
 - substance,
 - statistics, and
 - design.

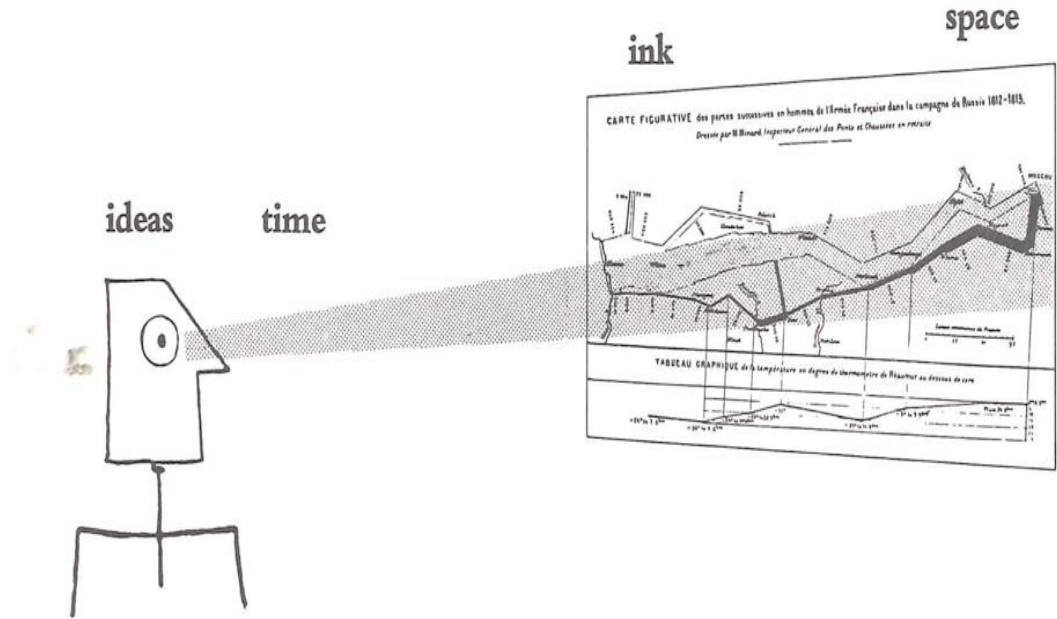
Graphical Excellence



- consists of complex ideas communicated with
 - clarity
 - precision
 - efficiency

Graphical Excellence

- is that which gives the viewer
 - the greatest number of ideas
 - in the shortest time
 - with the least ink



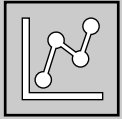
Graphical Excellence



Can be achieved by:

1. telling the truth about the data (**graphical integrity principles**)
2. visualizing with clarity and precision (**design principles**)

Graphical Integrity



“Not lying with statistics”



Tell the truth about data

Graphical Integrity

What do you notice here?

Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.

\$147

\$172

\$198

\$213

\$214



South Mountain



Pittsburgh



Harrisburg



Norristown



Warren

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

Graphical Integrity

What do you notice here?

Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.

\$147

\$172

\$198

\$213

\$214



South Mountain



Pittsburgh



Harrisburg



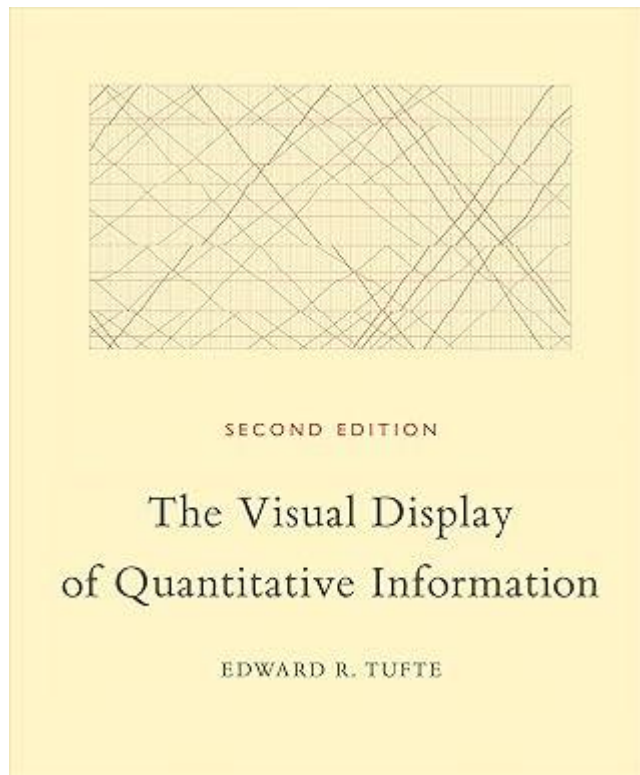
Norristown



Warren

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

Principles of Graphical Integrity



Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.

Principles of Graphical Integrity



1

The representation of numbers, as measured on the surface of a graph, should be directly proportional to the numerical quantities represented.

Principles of Graphical Integrity

1

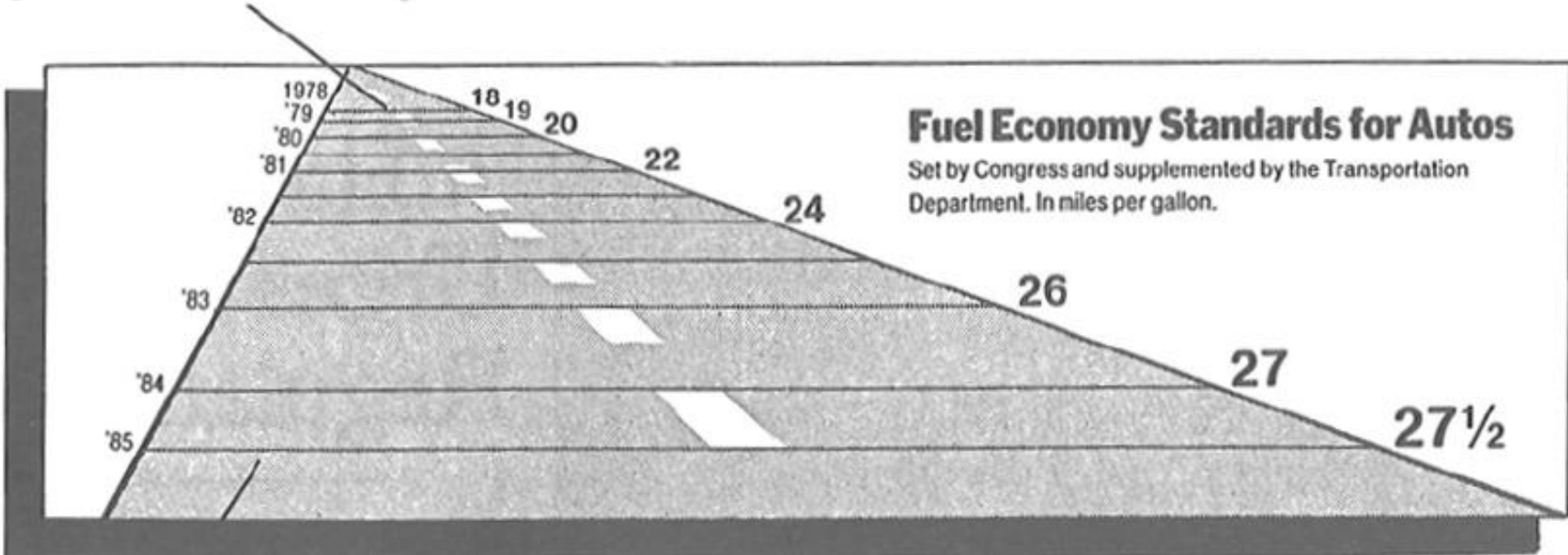
The representation of numbers, as measured on the surface of a graph, should be directly proportional to the numerical quantities represented.

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

Lie factor = 1 (accurate representation)

Principles of Graphical Integrity

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.

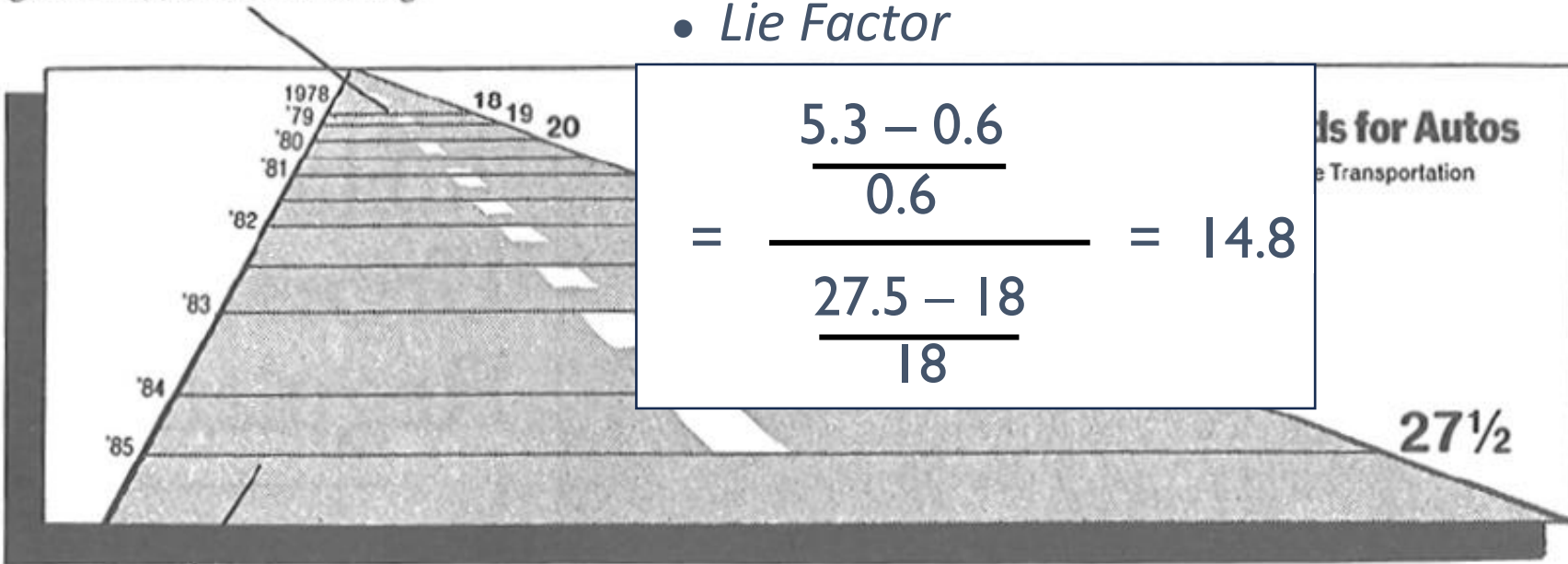
Principles of Graphical Integrity

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

- *Lie Factor*

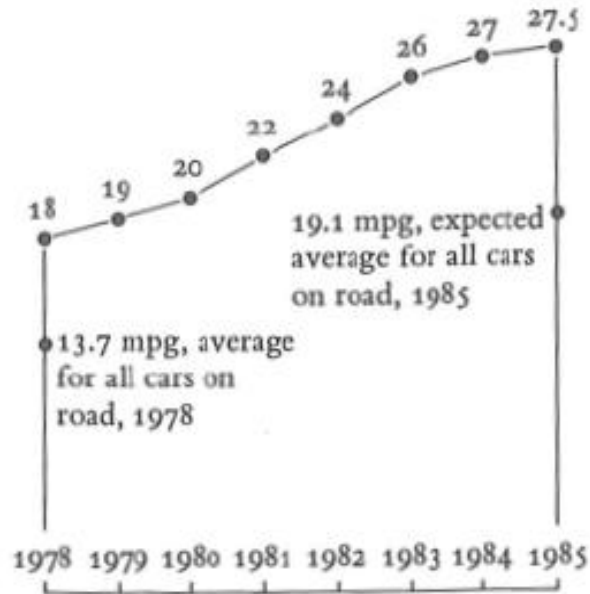
$$= \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$

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



Principles of Graphical Integrity

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985



Principles of Graphical Integrity



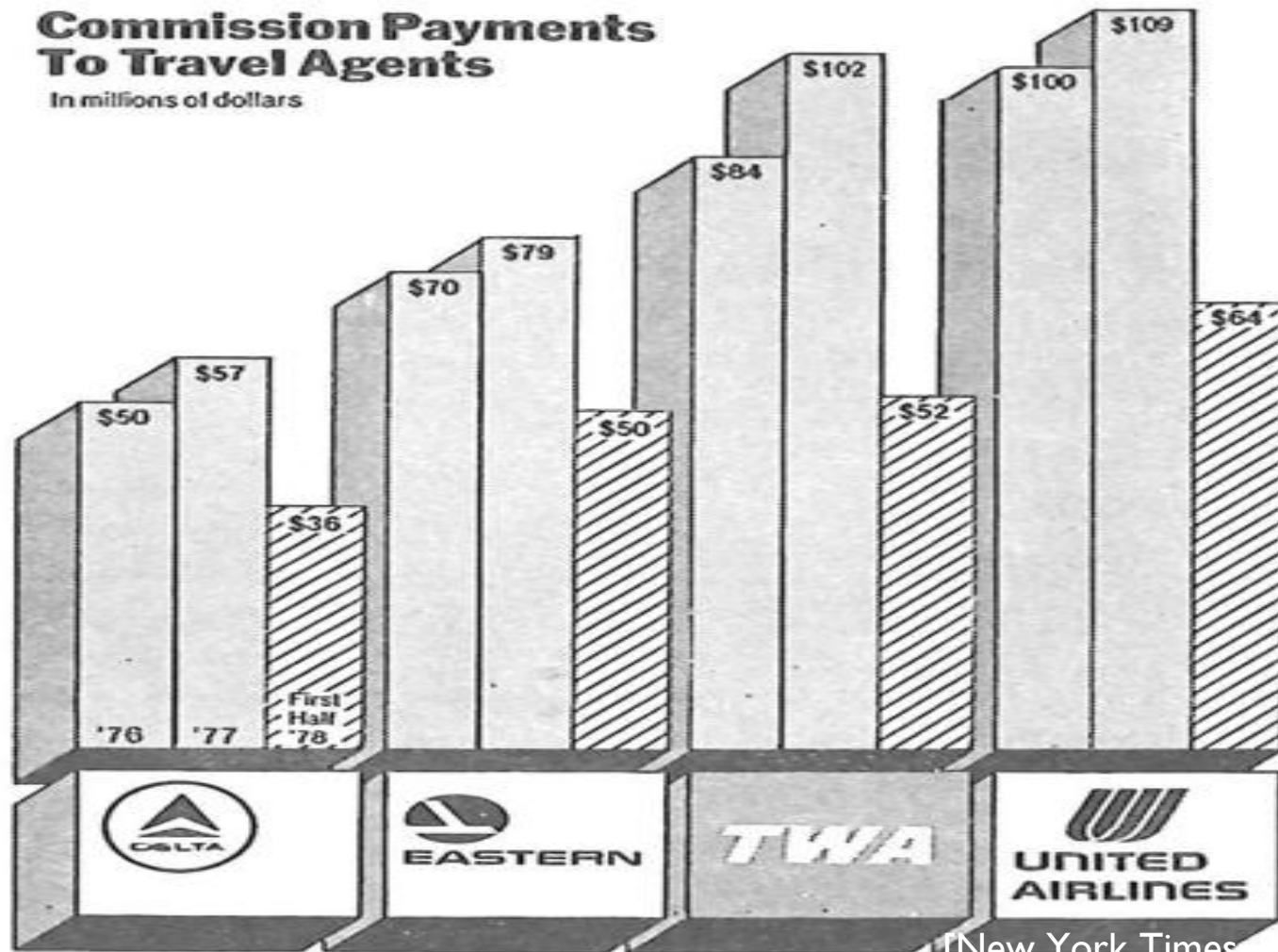
2

Clear, detailed, and thorough labeling should be used to prevent **graphical distortion and ambiguity**.

Write out explanations of the data on the graphic itself. Label important events in the data.

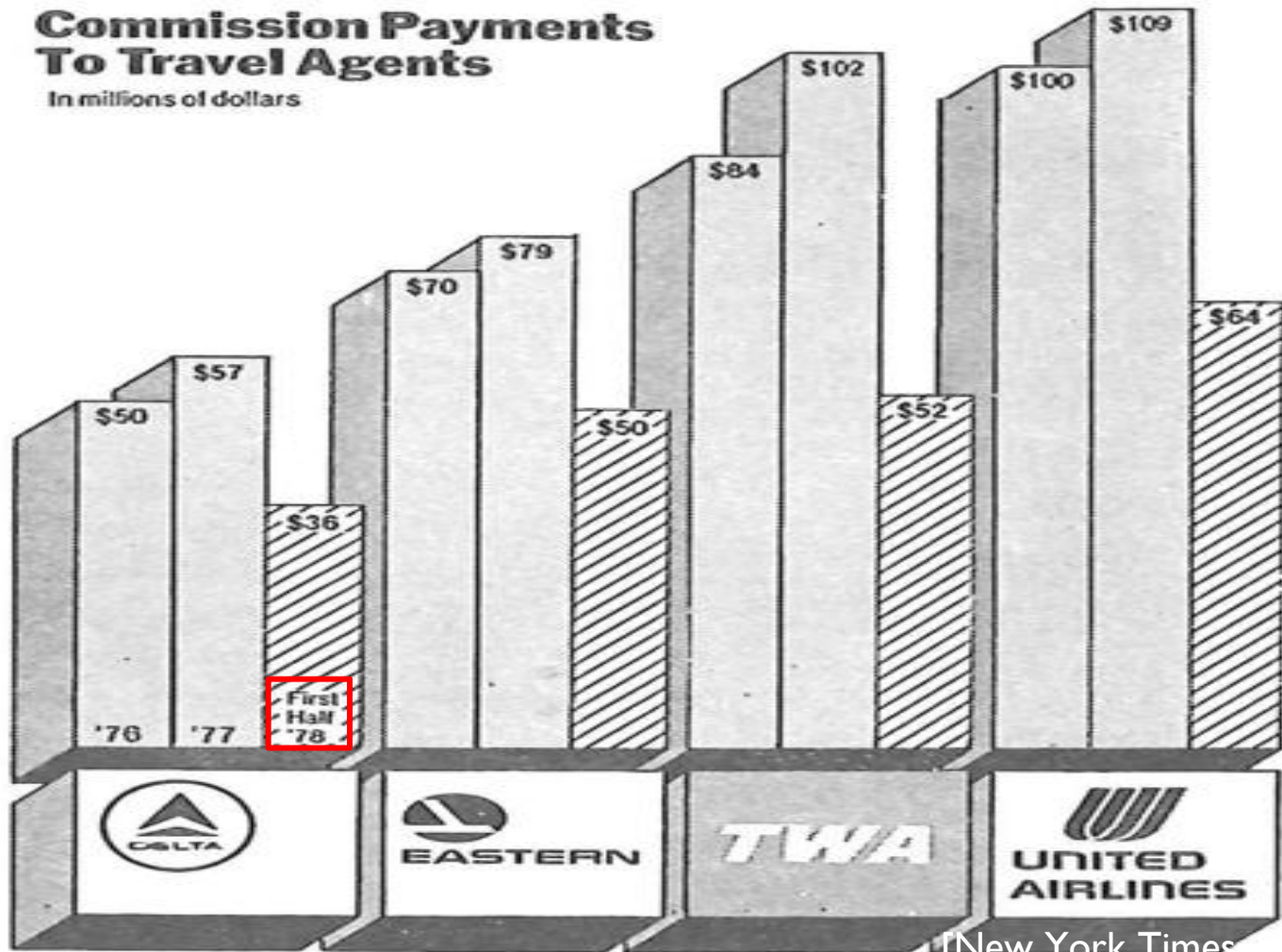
Commission Payments To Travel Agents

In millions of dollars



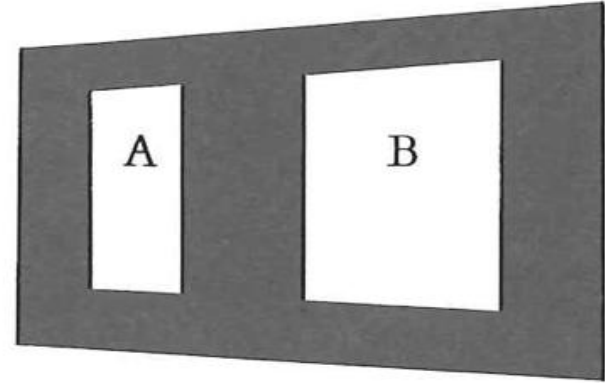
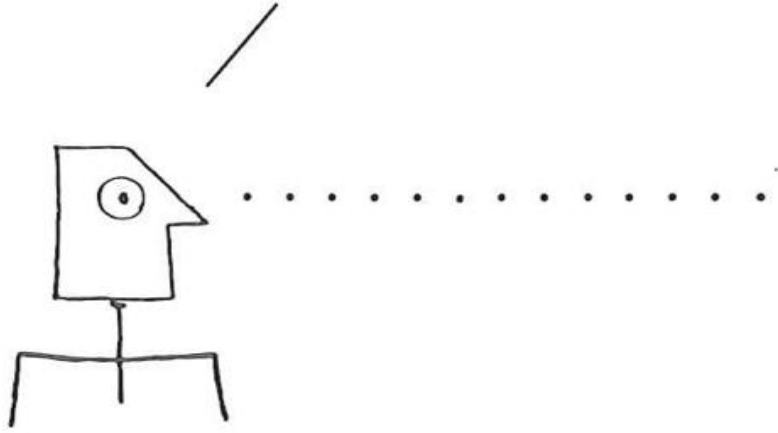
Commission Payments To Travel Agents

In millions of dollars



Distortion on Infographics

I think I see that area B is 3.14 times bigger than area A. Is that correct?



Distortion on Infographics



- Perceived Area
 - grows more slowly than measured area
 - varies between people
 - changes with experience
 - changes with context

Principles of Graphical Integrity



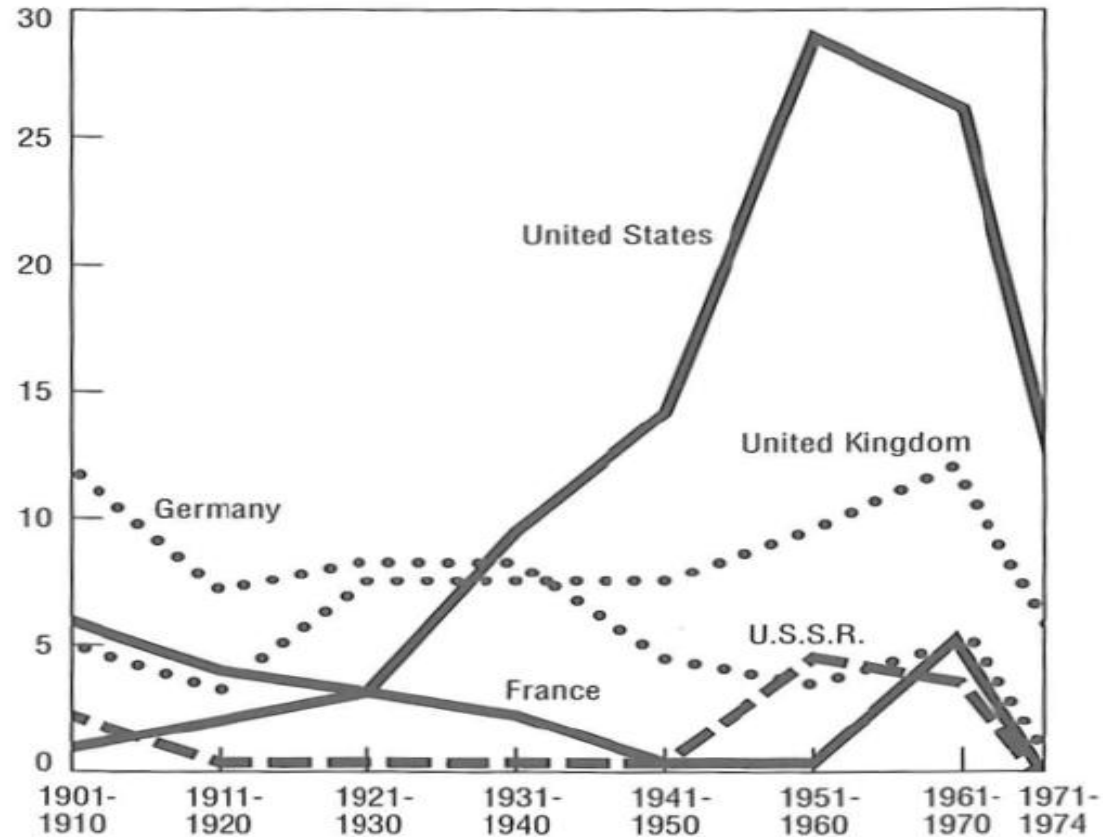
3

Show data variation, not design variation

- Design vs Data Variation
 - We expect that patterns will continue
 - Don't confuse design variation and data variation

Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974

(Number of Prizes)

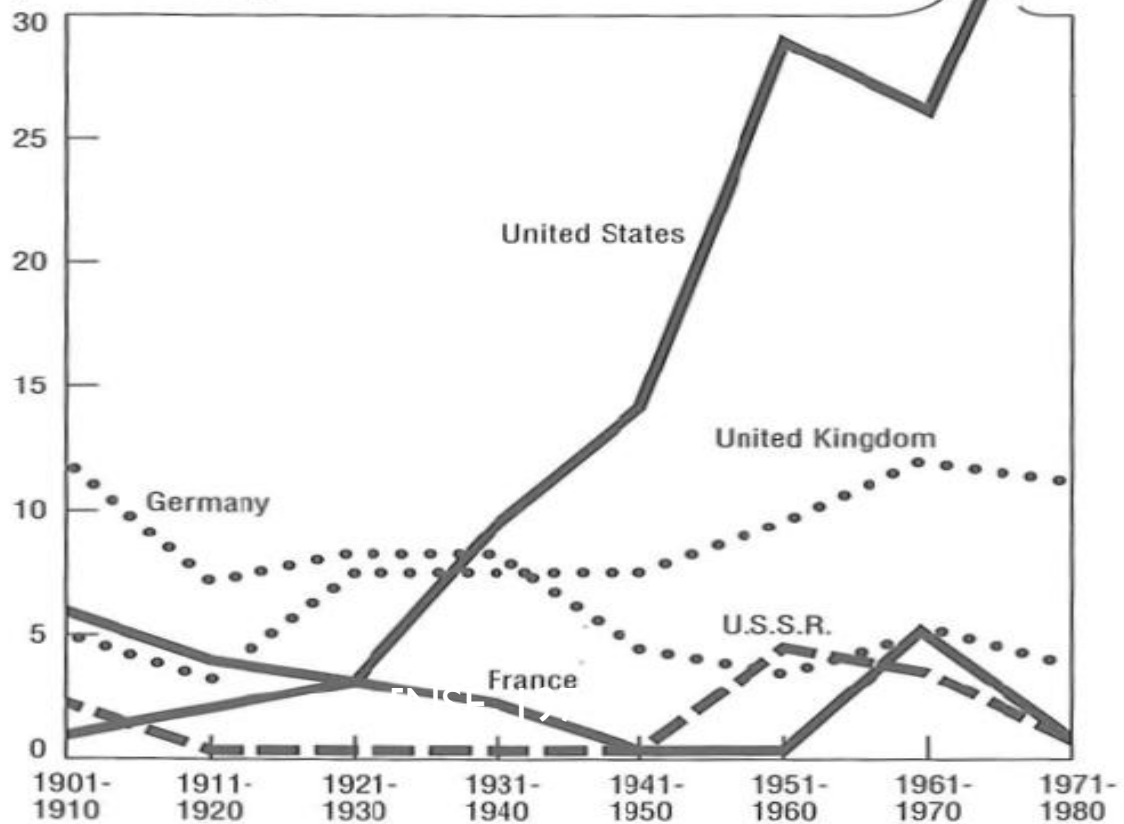


What is wrong
here?



Nobel Prizes Awarded in Science, for Selected Countries, 1901-1980

(Number of Prizes)

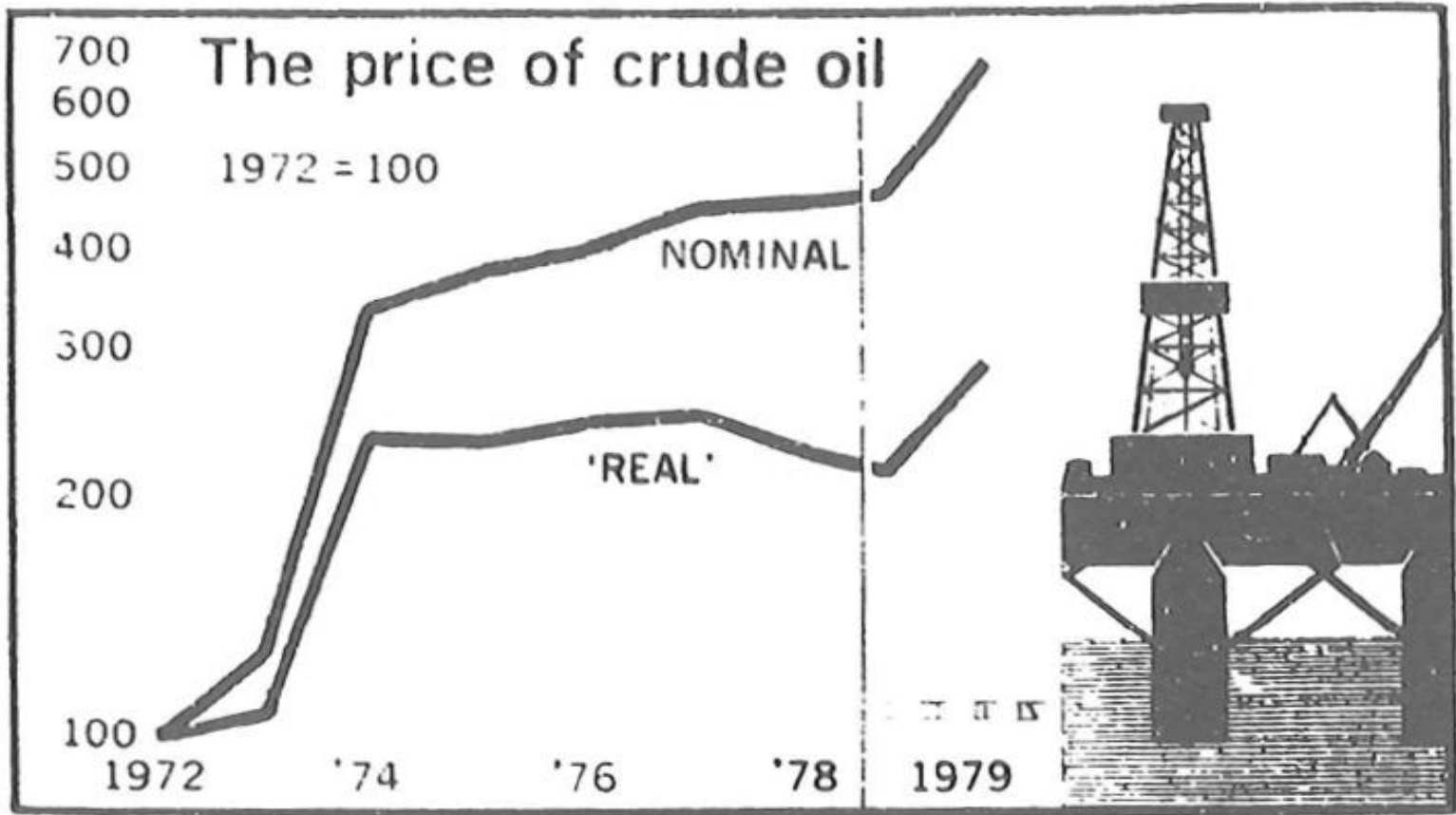


Principles of Graphical Integrity



4

In time-series displays of money, deflated and standardized units of monetary measurements are nearly always better than nominal units.



Example: Adjusted dollar amounts for inflation

Principles of Graphical Integrity



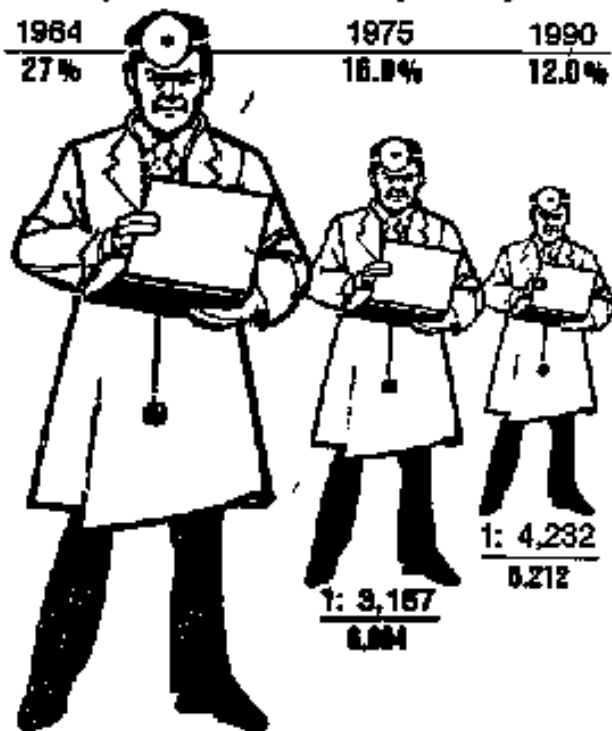
5

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

THE SHRINKING FAMILY DOCTOR in California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.8%	12.0%



1: 2,247 RATIO TO POPULATION
8,023 Doctors

1: 3,157
6,004

1: 4,232
0,212

New York State Total Budget Expenditures and Aid to Localities

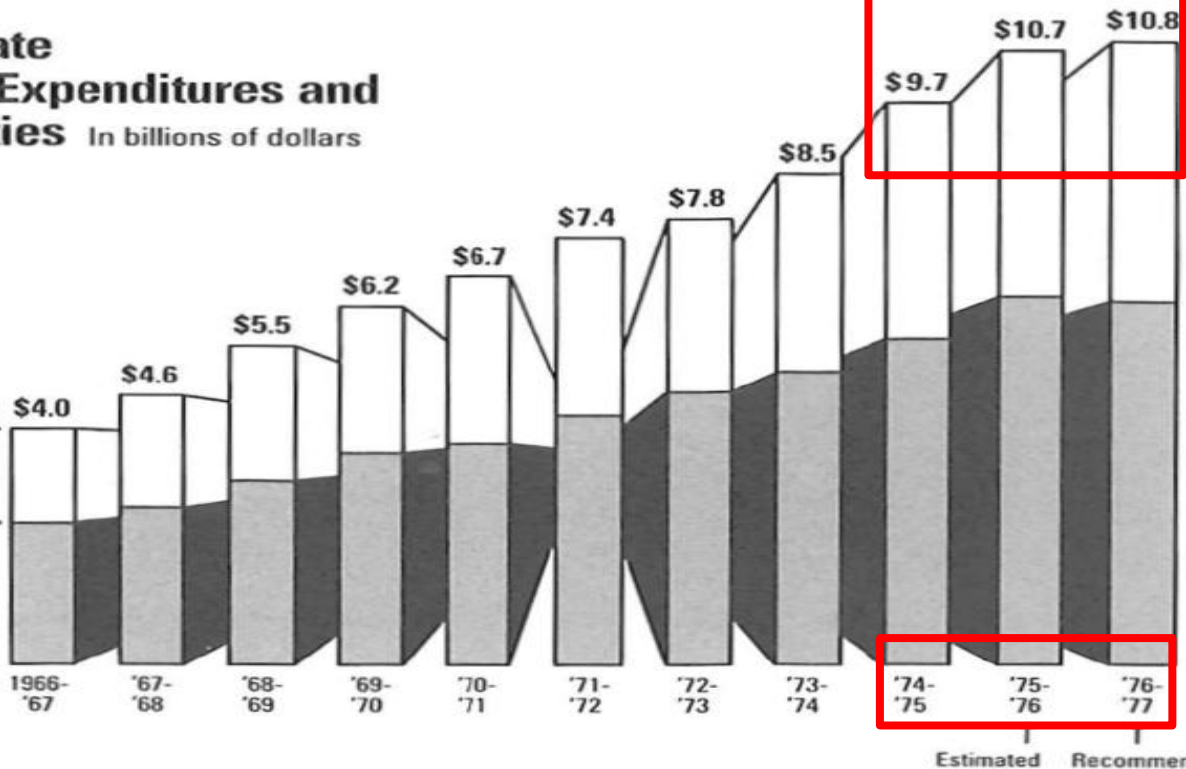
In billions of dollars

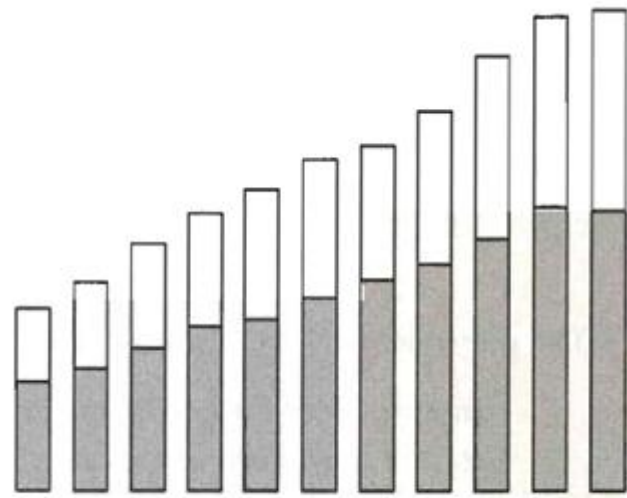
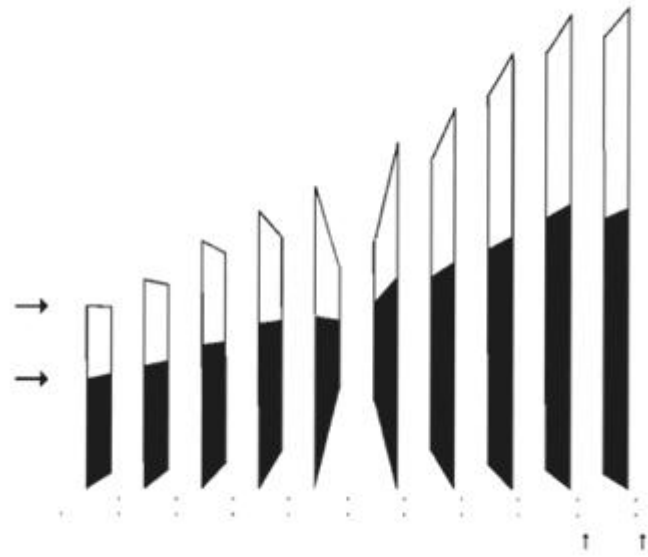
Fiscal 1966-1976

Total Budget

Total Aid to
Localities*

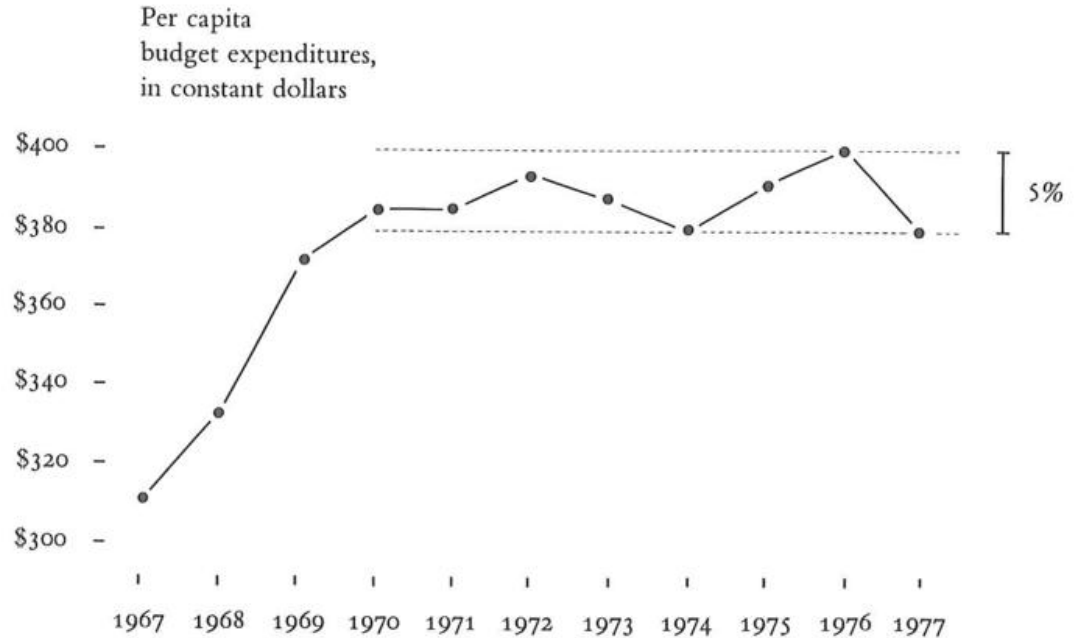
*Varying from a low
of 56.7 percent of
the total in 1970-71
to a high of 60.7
percent in 1972-73





Graphical Integrity

- Done well
- Adjusted for inflation
- Adjusted for population
- No chart junk

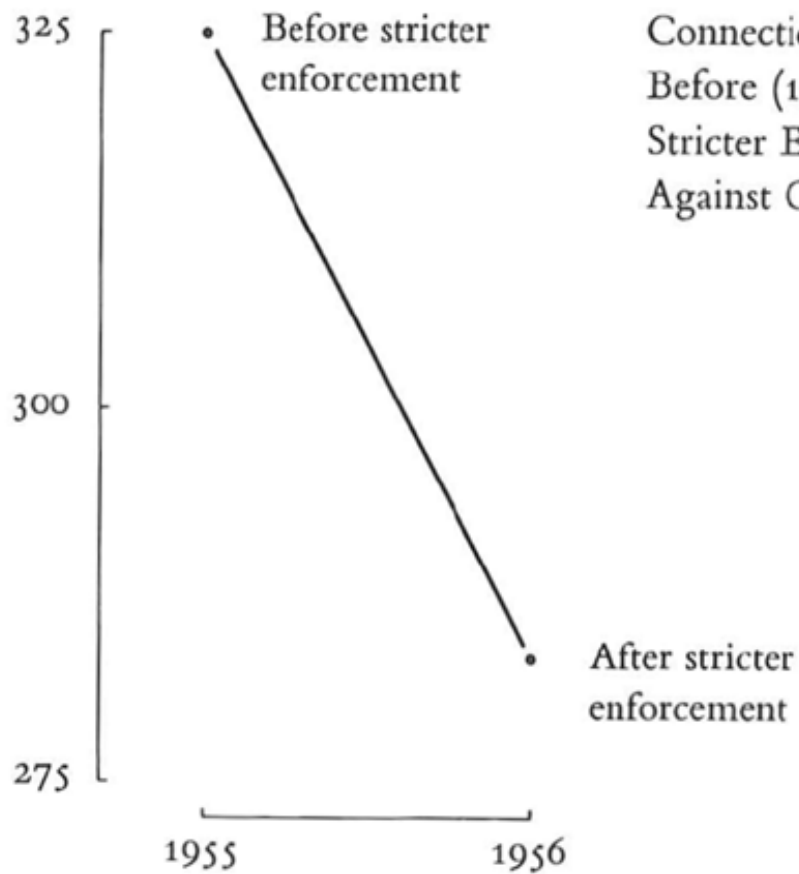


Principles of Graphical Integrity

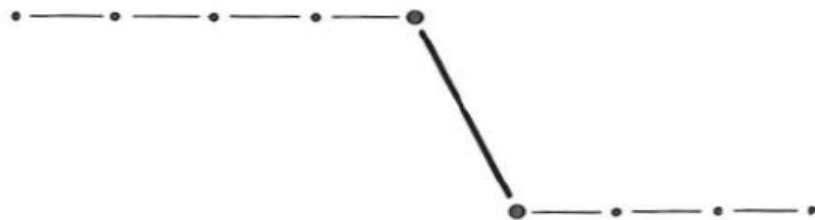


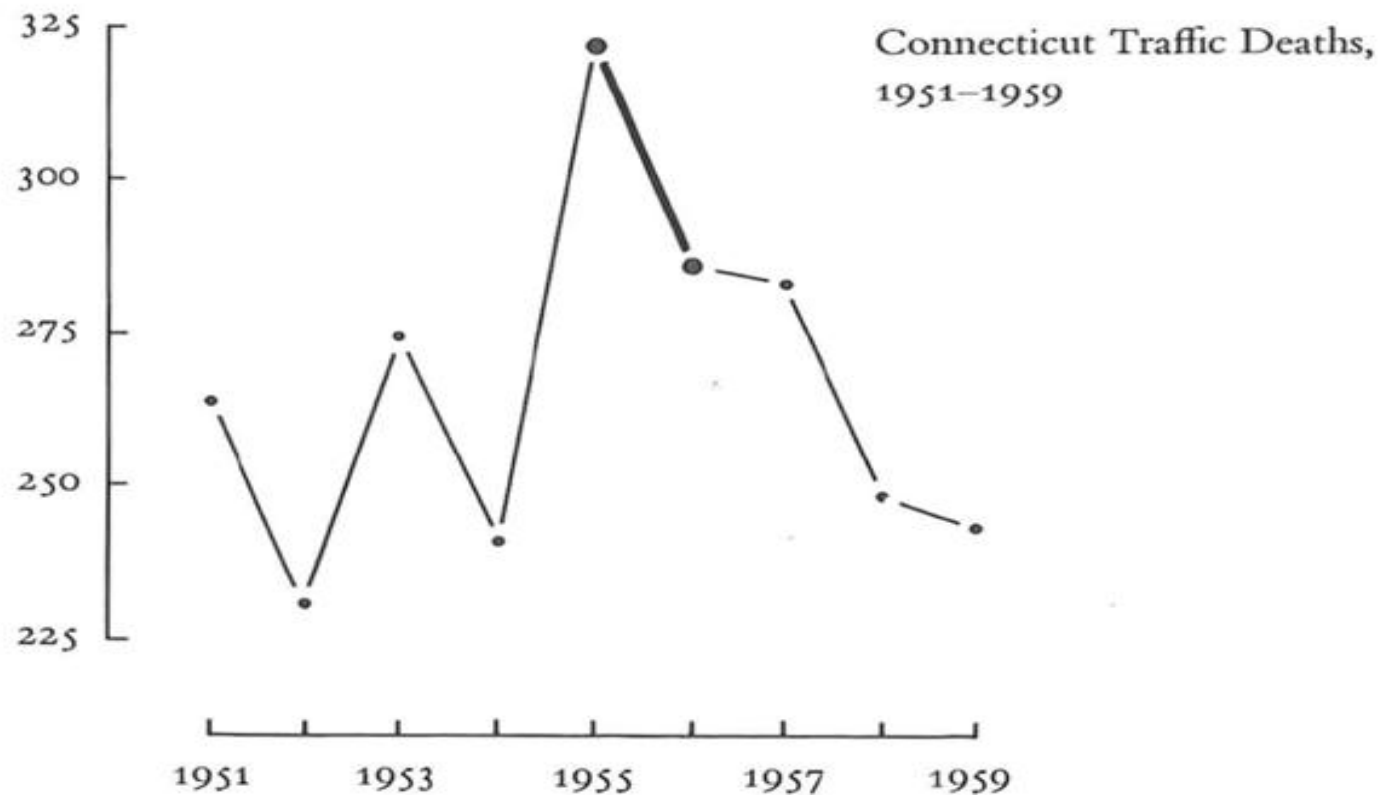
6

Graphics must not quote data out of context



Connecticut Traffic Deaths,
Before (1955) and After (1956)
Stricter Enforcement by the Police
Against Cars Exceeding Speed limit





Tufte's Principle for Graphical Integrity



1. The representation of numbers, as measured on the surface of a graph, should be directly proportional to the numerical quantities represented
2. Clear, detailed, and thorough labeling should be used to prevent graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
3. Show data variation, not design variation
4. In time-series displays of money, deflated and standardized units of monetary measurements are nearly always better than nominal units.
5. The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
6. Graphics must not quote data out of context



Visualization Design Principles

Graphics Excellence



Can be achieved by:

1. telling the truth about the data (**graphical integrity principles**)
2. visualizing with clarity and precision (**design principles**)

Tufte's Design Principles



- ❑ Data-Ink Maximization and Graphical Redesign
- ❑ Remove Chartjunk
- ❑ Multi-functioning Graphical Elements
- ❑ Data Density
- ❑ Small Multiples and Parallel Sequencing

Data Ink and Graphic Redesign



Tufte's Five Laws of Data-Ink:

Above all else show the data.

Maximize the data-ink ratio

Erase non-data-ink.

Erase redundant data-ink.

Revise and edit.

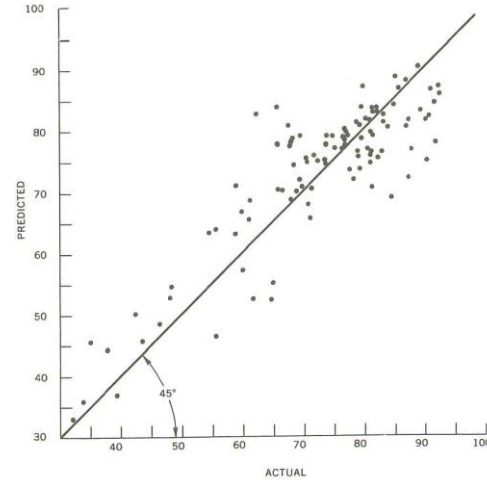
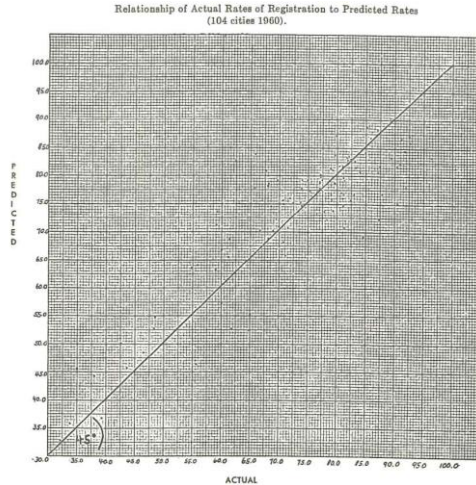
Data Ink and Graphic Redesign

- Maximize data-ink ratio

$$\text{Data ink ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

Data Ink and Graphic Redesign

□ Example

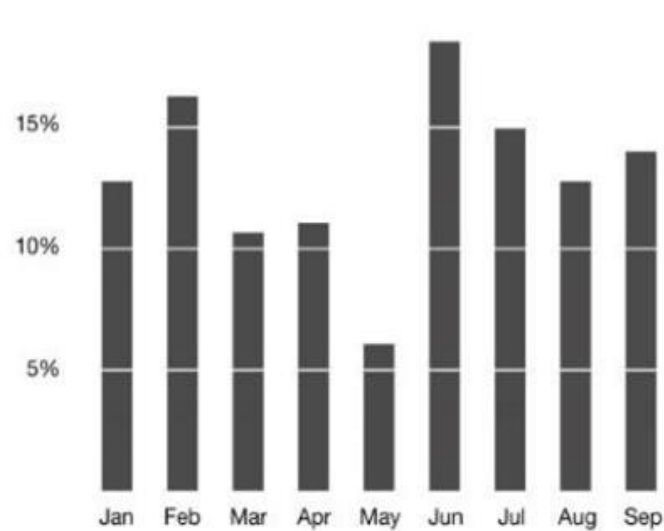
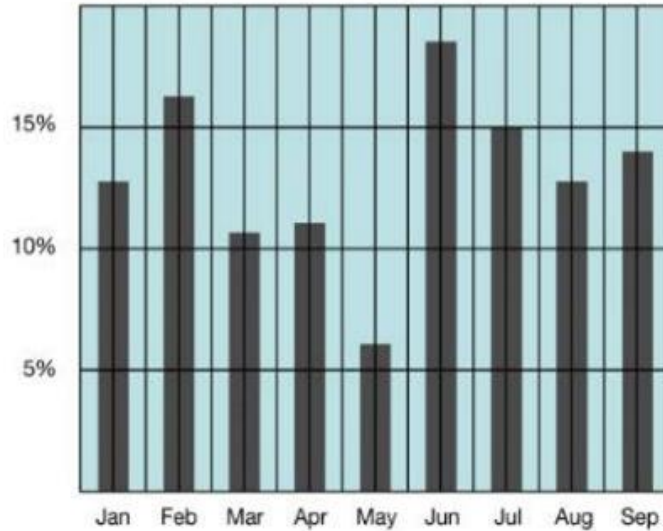


Erase non-data-ink (within reason)

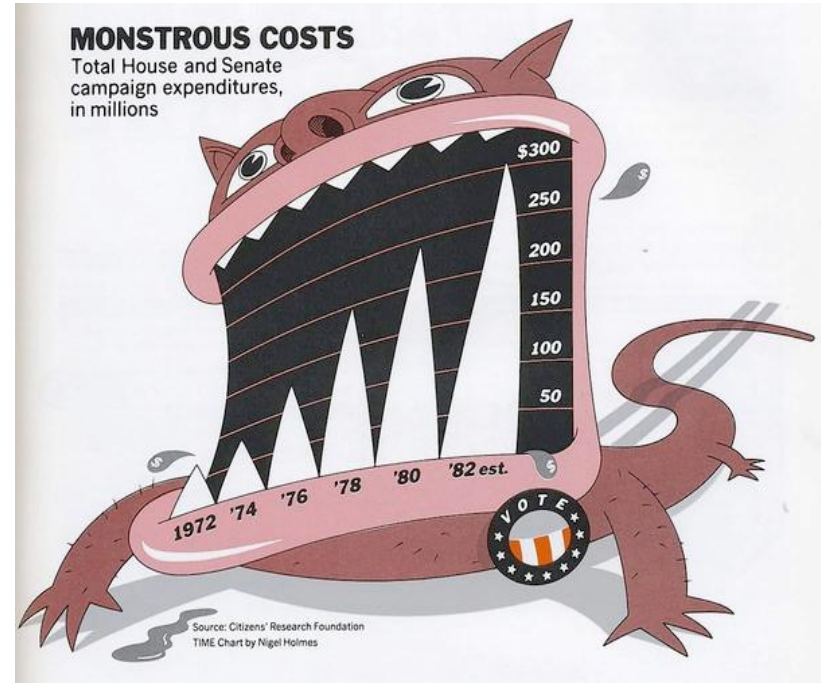
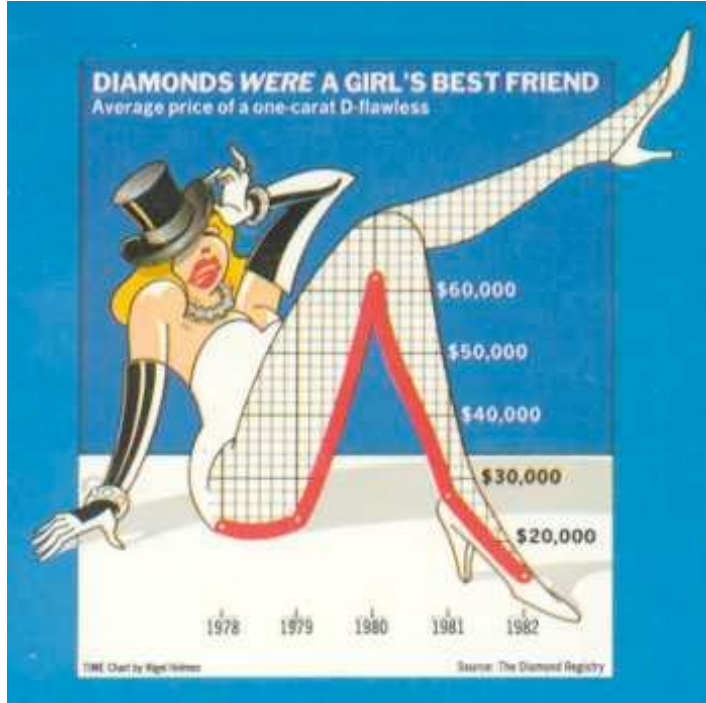
Erase redundant data-ink

Avoid Chartjunk

- ❑ Extraneous visual elements that detract from message:



Avoid Chartjunk



Multi-functioning Elements

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

- In other words, try to make all present graphical elements data encoding elements.

Multi-functioning Elements

- ❑ Stem and Leaf Plots
 - ❑ construct the distribution of a variable with numbers themselves

0 | 9 = 900 feet

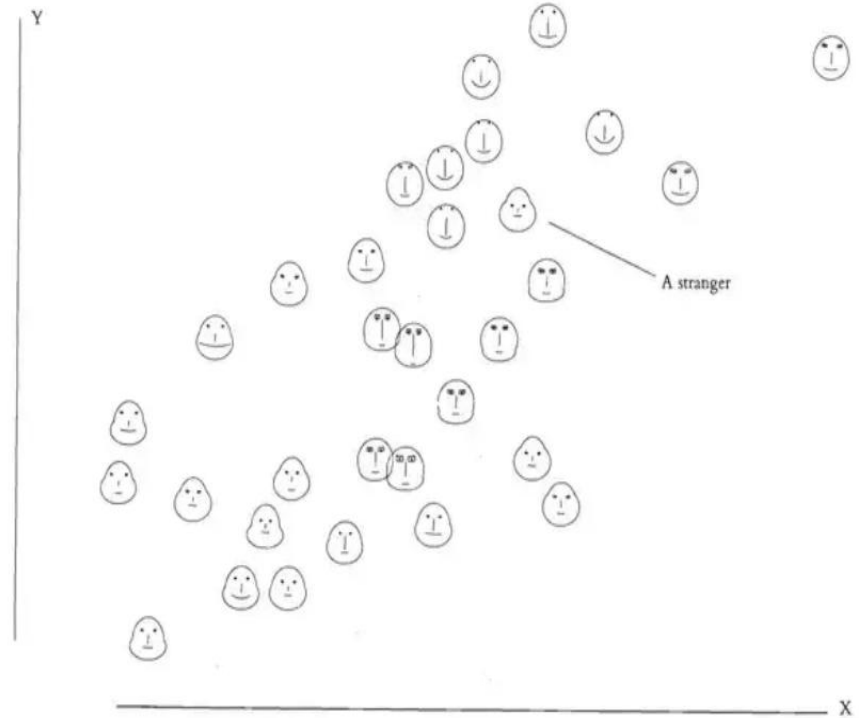
Stem-and-leaf displays:
heights of 218 volcanoes, unit 100 feet.

19 | 3 = 19,300 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

Multi-functioning Elements

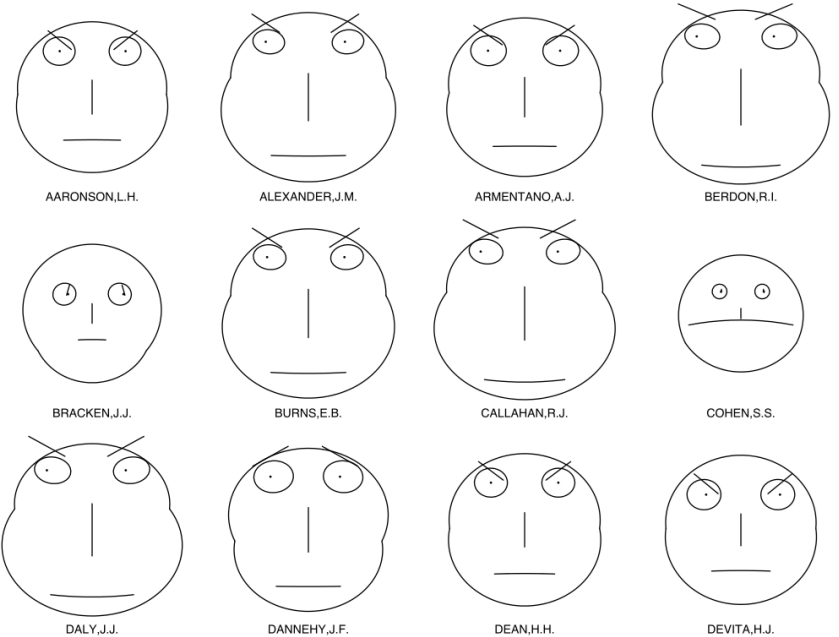
- ❑ Chernoff Faces
 - ❑ Show a bunch of variables at once via facial features like lips, eyes, and nose size.



Multi-functioning Elements

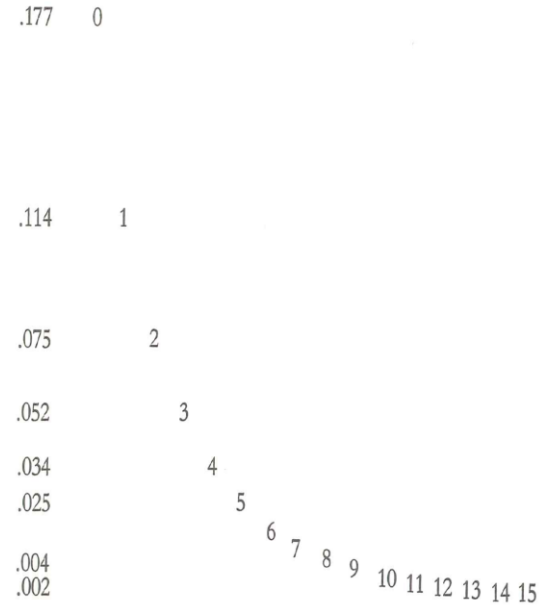
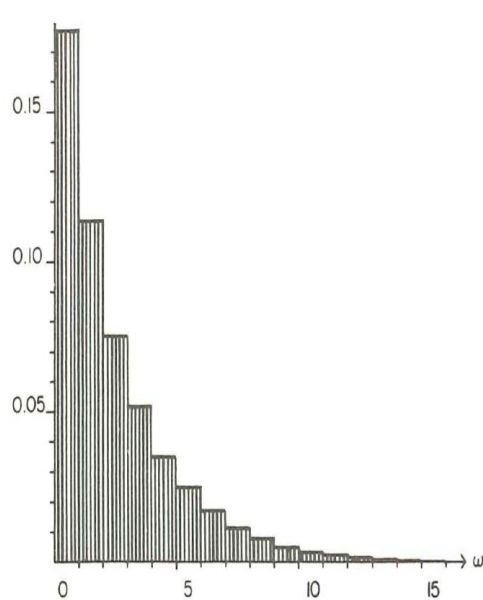
❑ Chernoff Faces

- ❑ Chernoff faces for lawyers' ratings of twelve judges



Multi-functioning Elements

- Coordinate labels as Marks

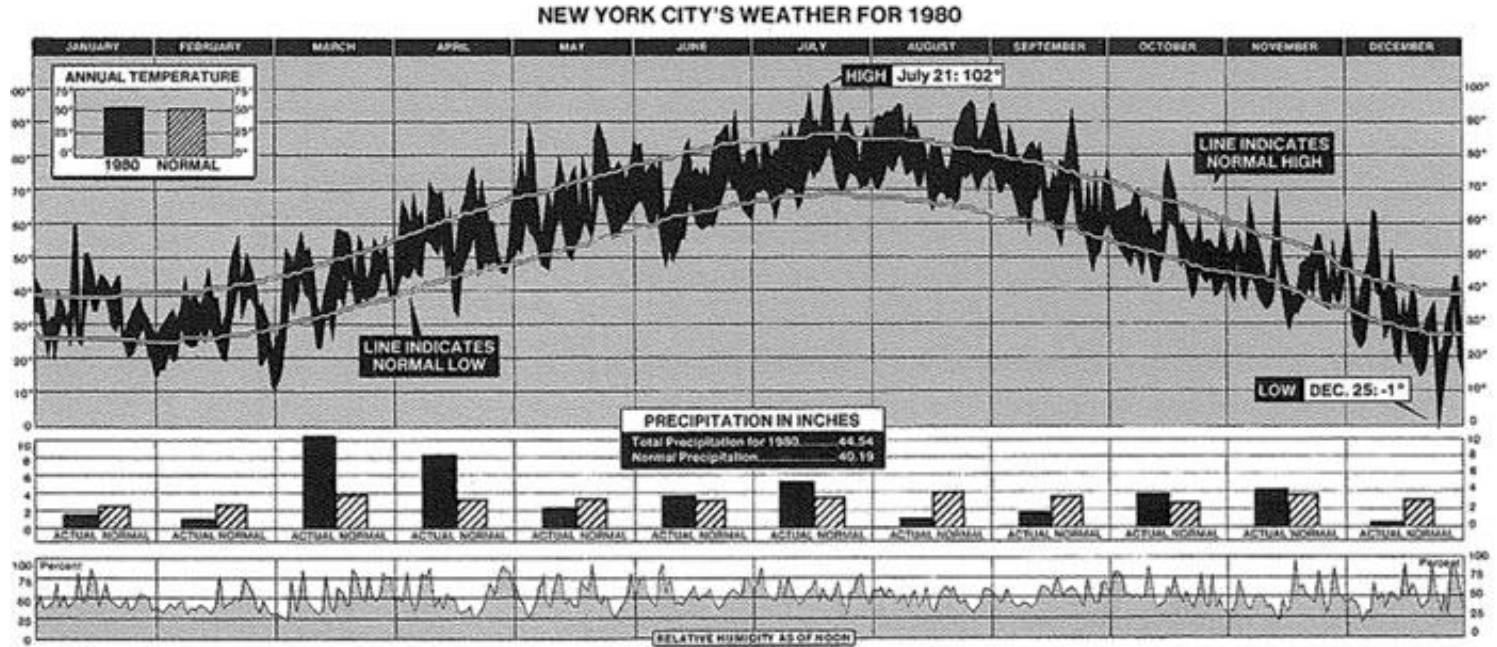


Maximize Data Density

- Maximize data density and the size of the data matrix within reason

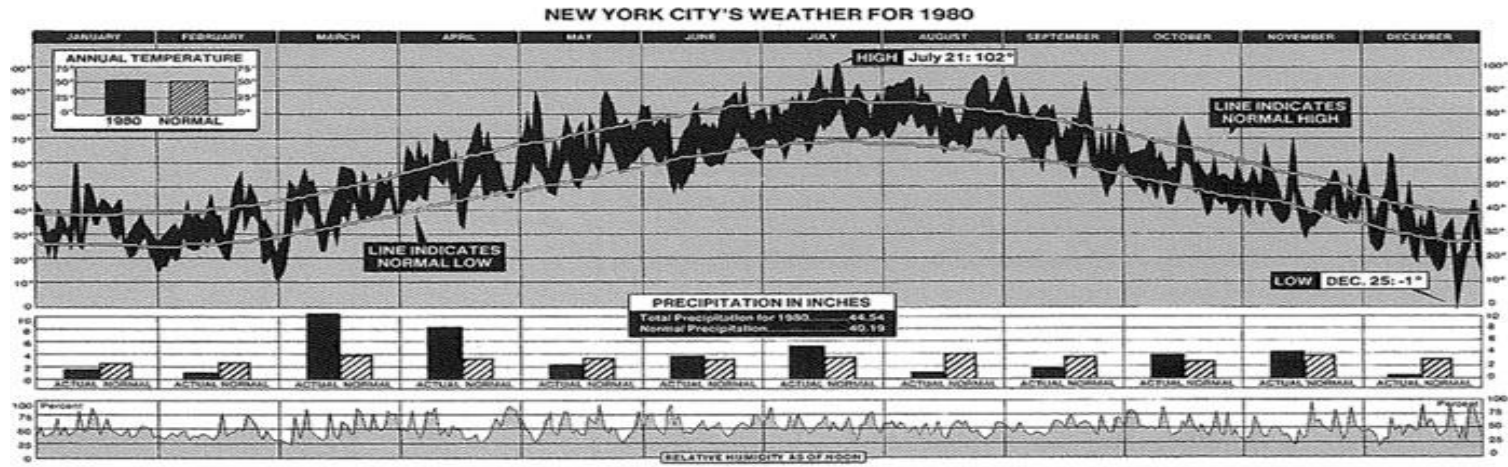
$$\text{Data density} = \frac{\# \text{ entries in data matrix}}{\text{area of data graphic}}$$

Maximize Data Density



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

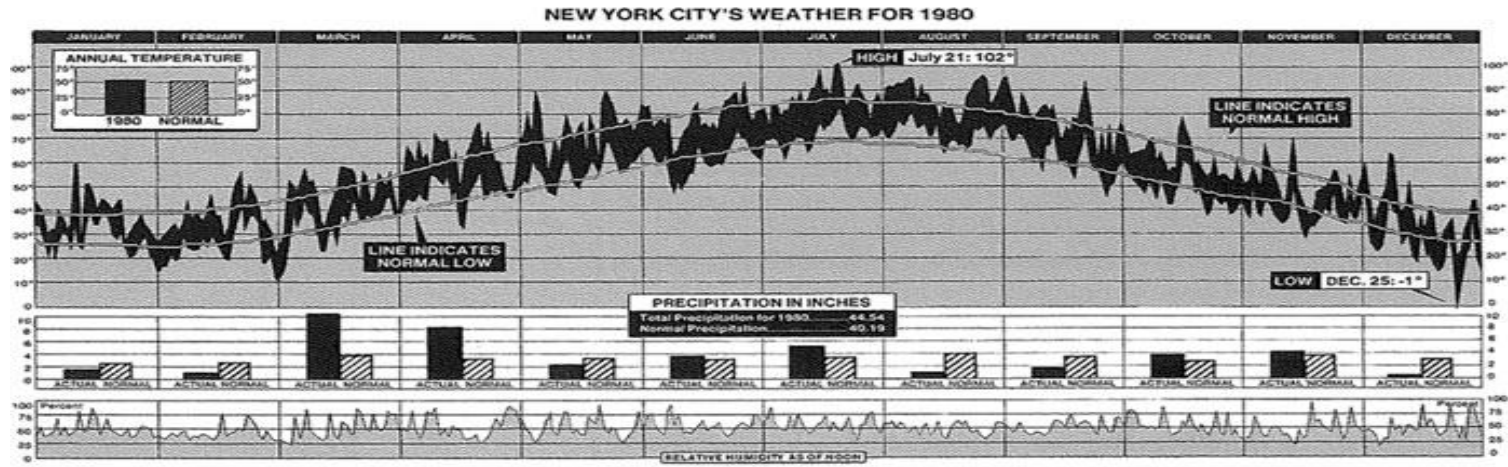
Maximize Data Density



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

- ❑ This graph shows *both valiantly, and successfully* 2200 numbers which summarize the trends and patterns in weather in New York City in 1980.
- ❑ The three aligned charts show *temperature, precipitation, and relative humidity*.
- ❑ In the graph of temperature, the area is filled between the daily low and daily high.

Maximize Data Density



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

- ❑ Segregates visually three different time series
- ❑ Aligns the axes on different plots to assist in making comparisons across the time series
- ❑ Uses markers to make to clearly denote when extremes occurred
- ❑ Contains clear textual labels

Small Multiples and Parallel Sequencing

- ❑ Previous examples improved data density by using a LARGE data matrix.

$$\text{Data density} = \frac{\# \text{ entries in data matrix}}{\text{area of data graphic}}$$

- ❑ Alternatively, we can reduce the size of the graphic = the **Shrink Principle**
- ❑ Repeated application of this principle leads to a **Small Multiples design**

Small Multiples

NITROGEN OXIDES

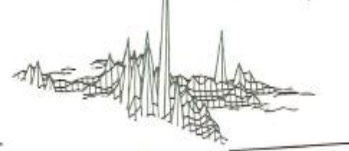
midnight to 3 am



6 am to 9 am



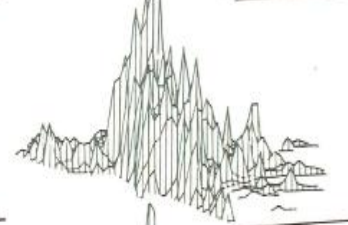
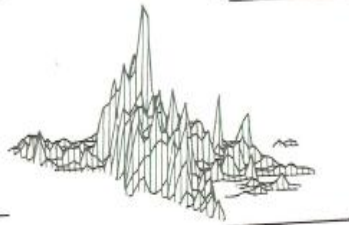
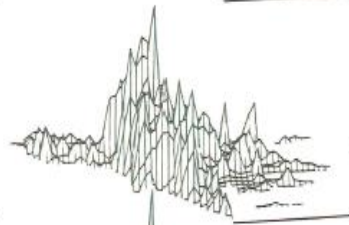
noon to 3 pm



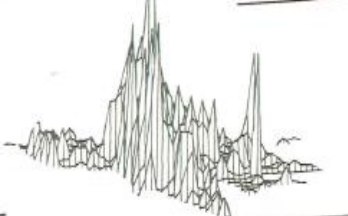
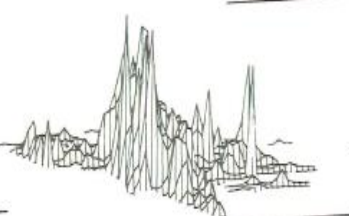
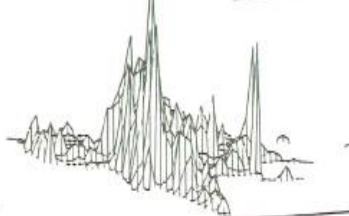
3 pm to 6 pm



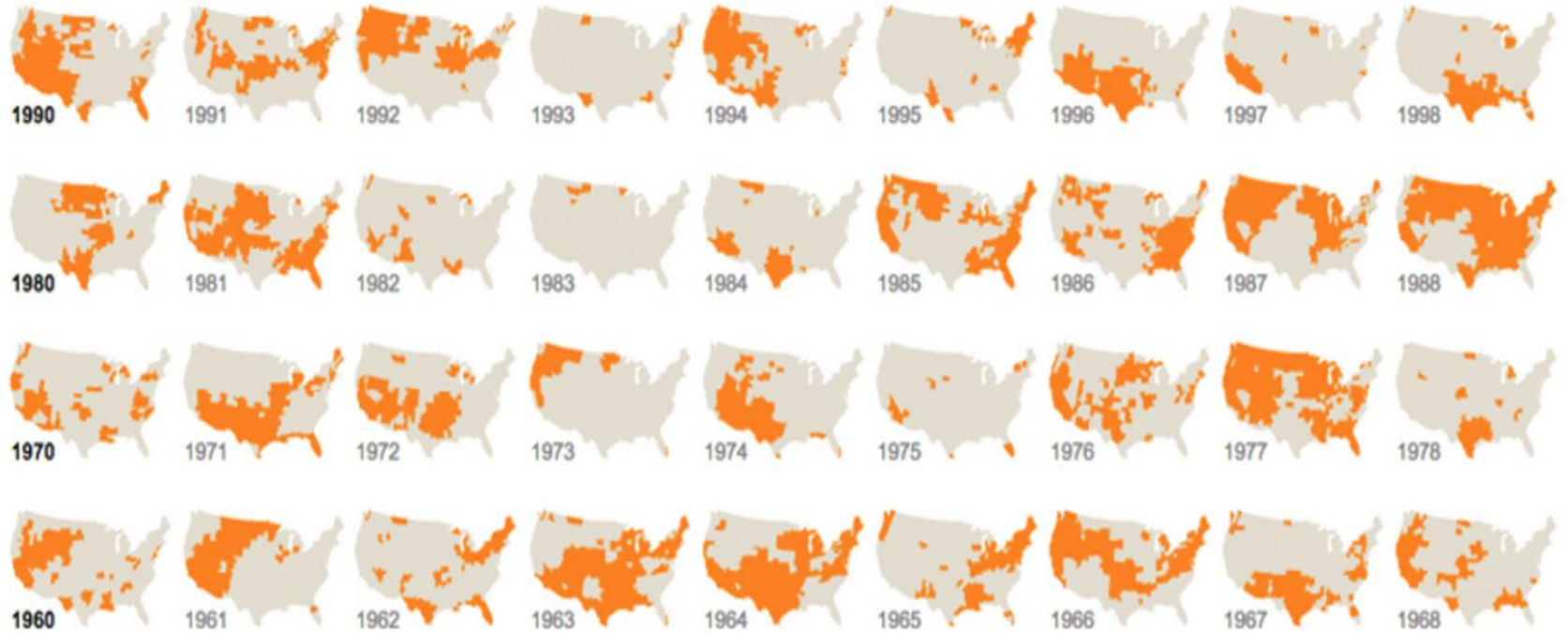
CARBON MONOXIDE



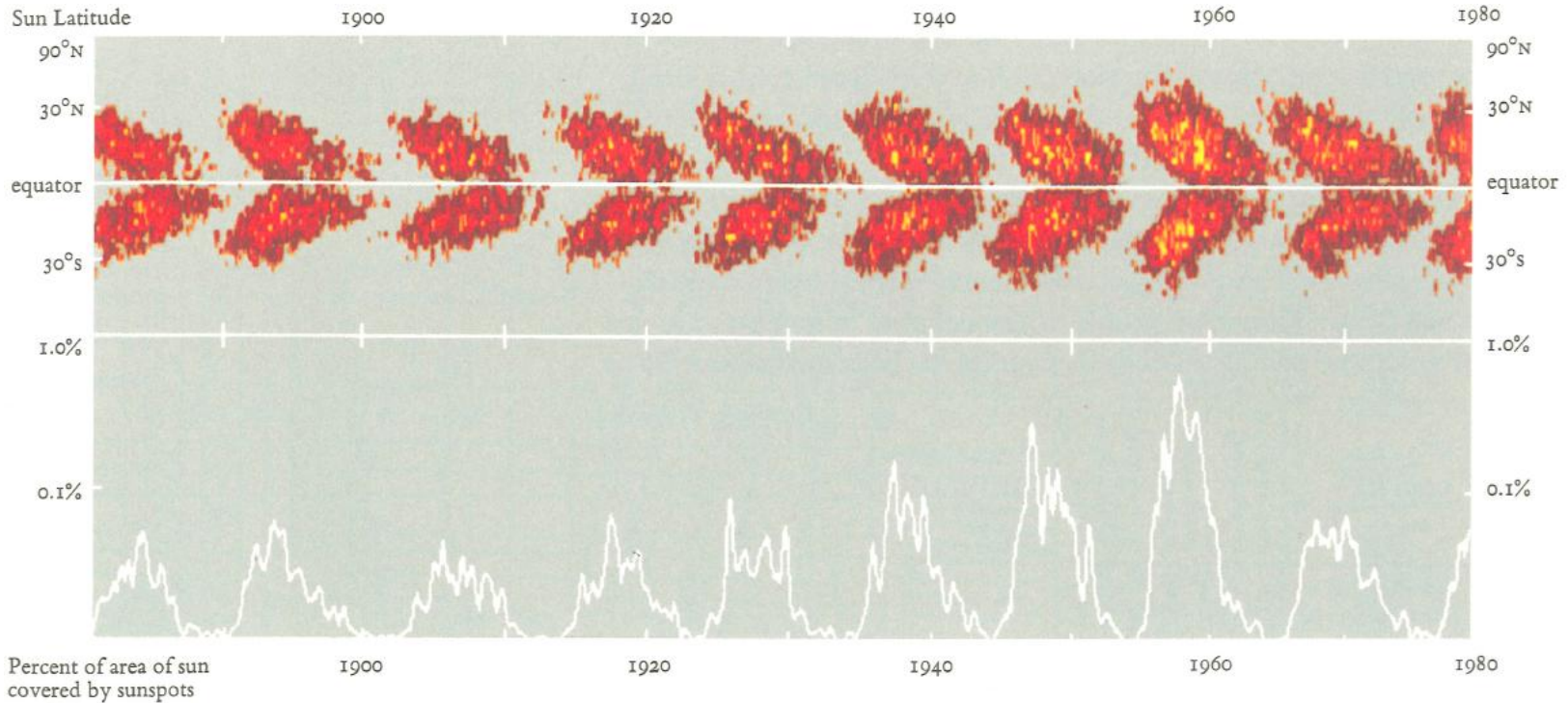
REACTIVE HYDROCARBONS



Small Multiples



Parallel Sequencing



Tufte's Design Principles



- ❑ Data-Ink Maximization and Graphical Redesign
- ❑ Remove Chartjunk
- ❑ Multi-functioning Graphical Elements
- ❑ Data Density
- ❑ Small Multiples and Parallel Sequencing

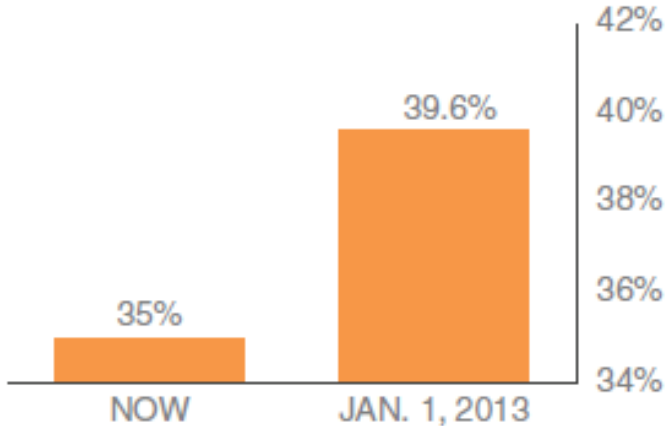


Deceptive Charts

Bars Must Always Start at Zero

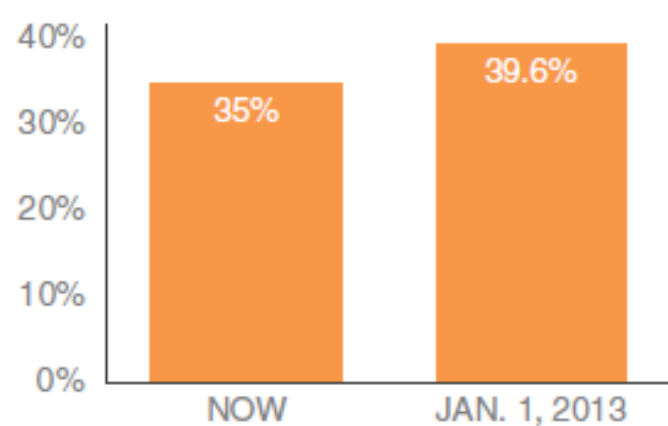
Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE



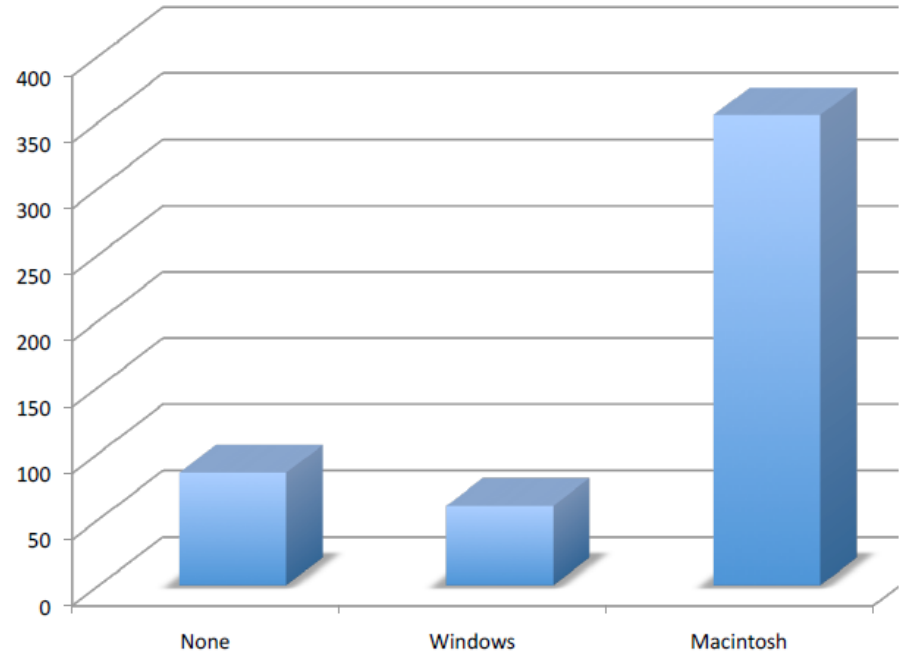
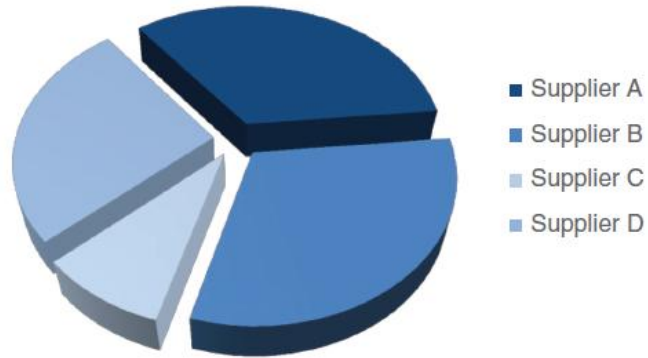
Zero baseline: as it should be graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE



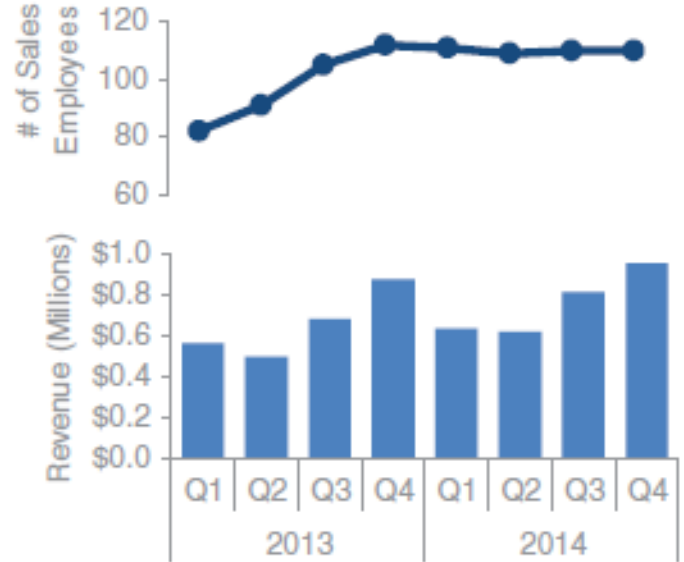
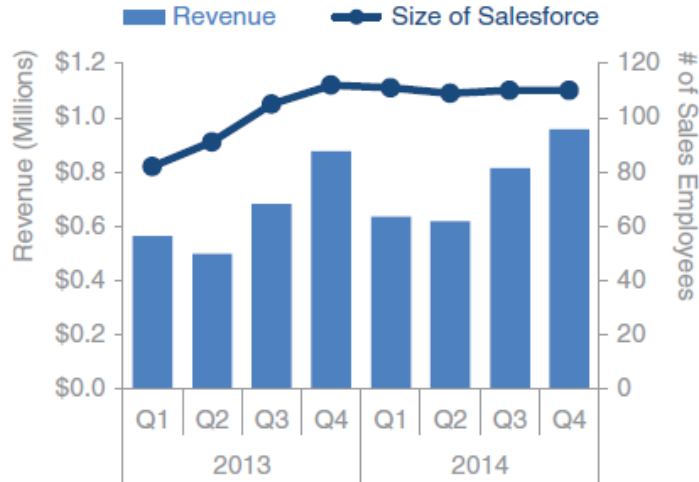
Avoid Pie Charts and 3-D

Supplier Market Share

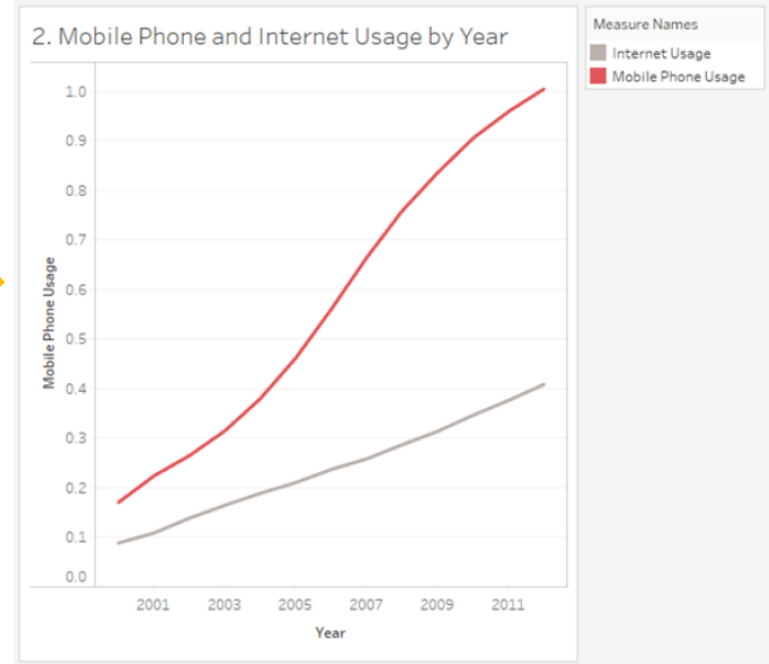
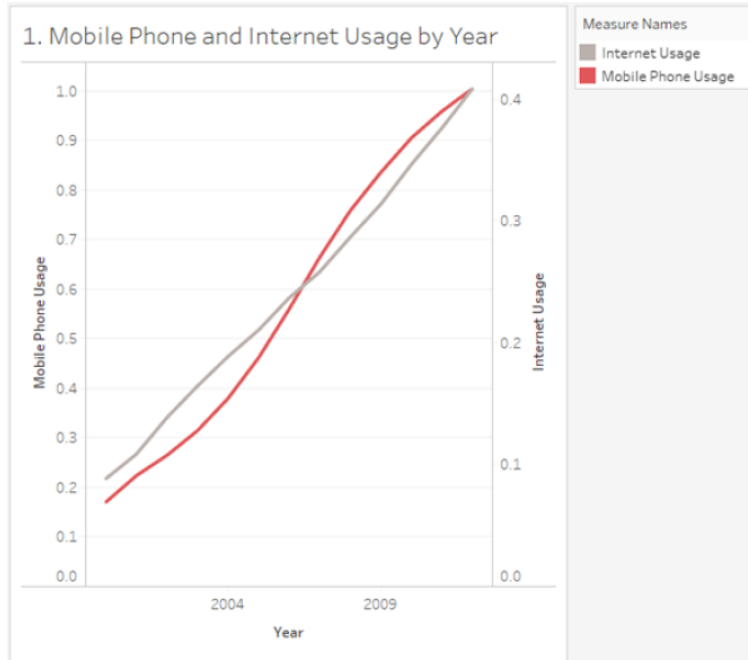


Be Careful with Dual Axes

Secondary y-axis



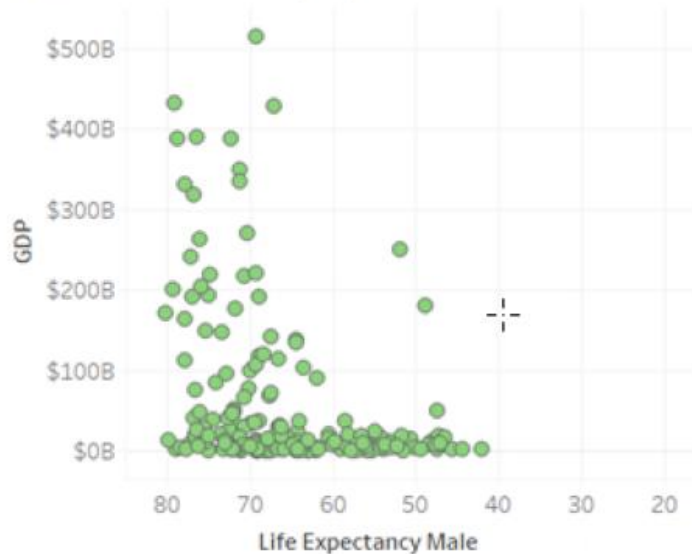
Synchronize Axes



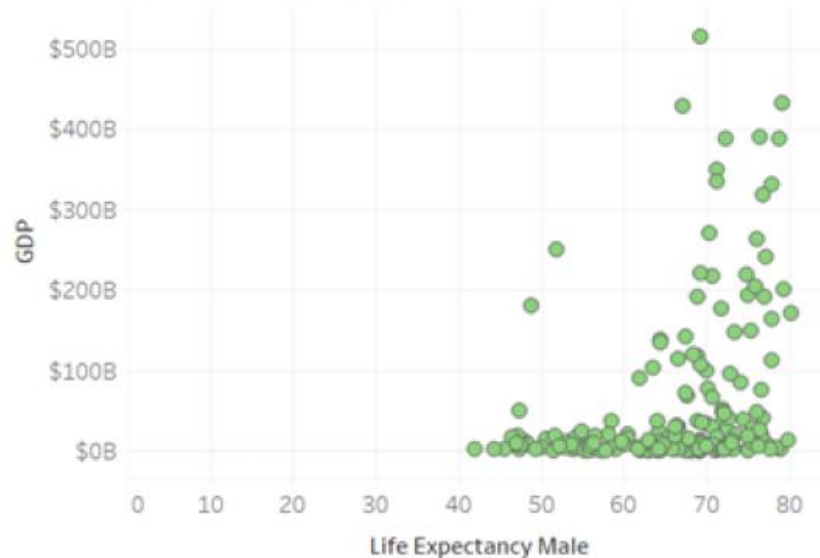
Make sure scale is in standard direction



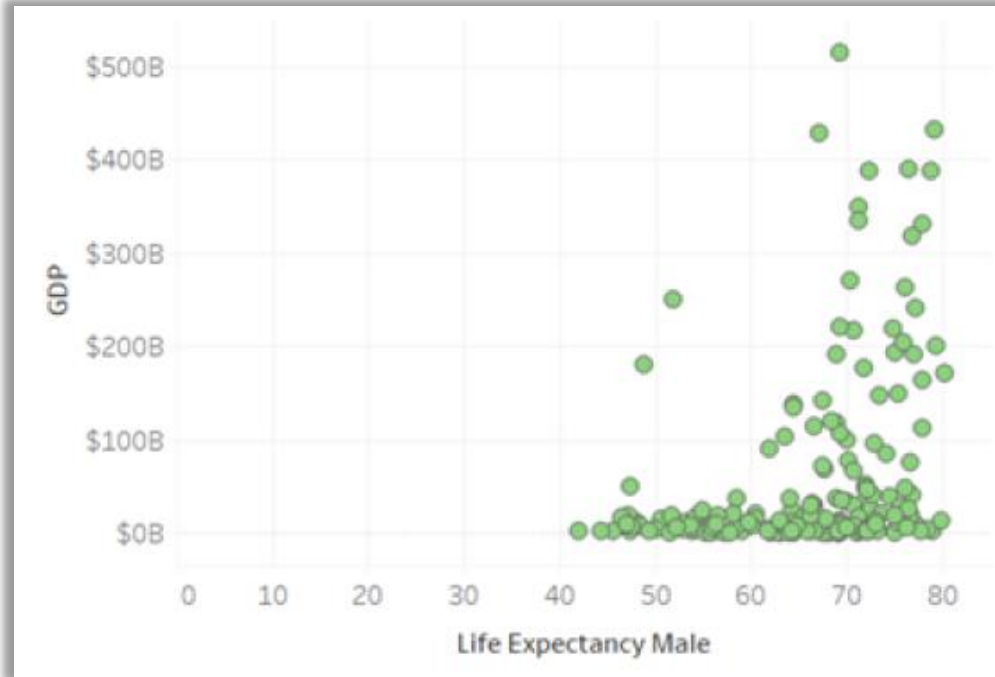
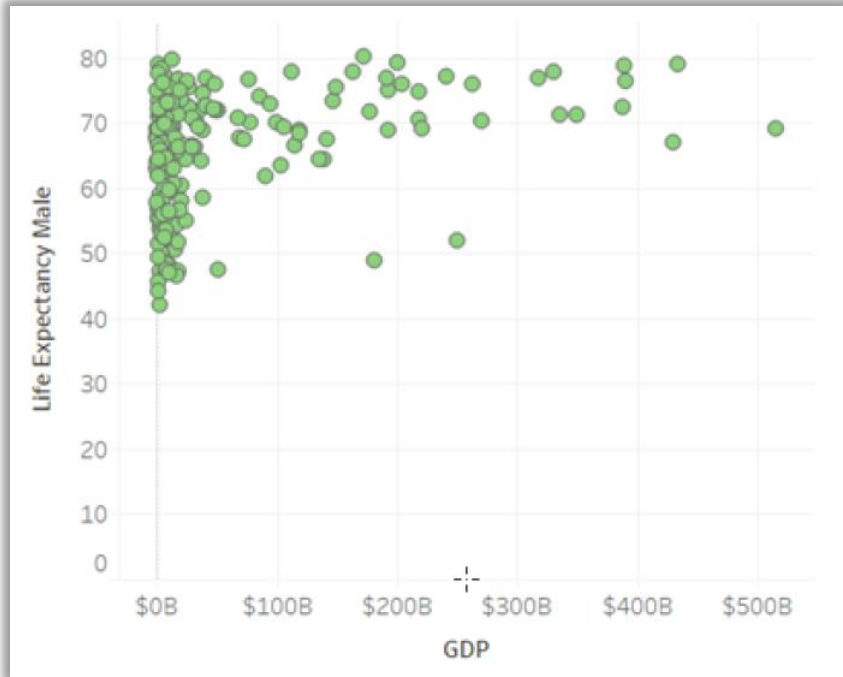
1. Life Expectancy by GDP



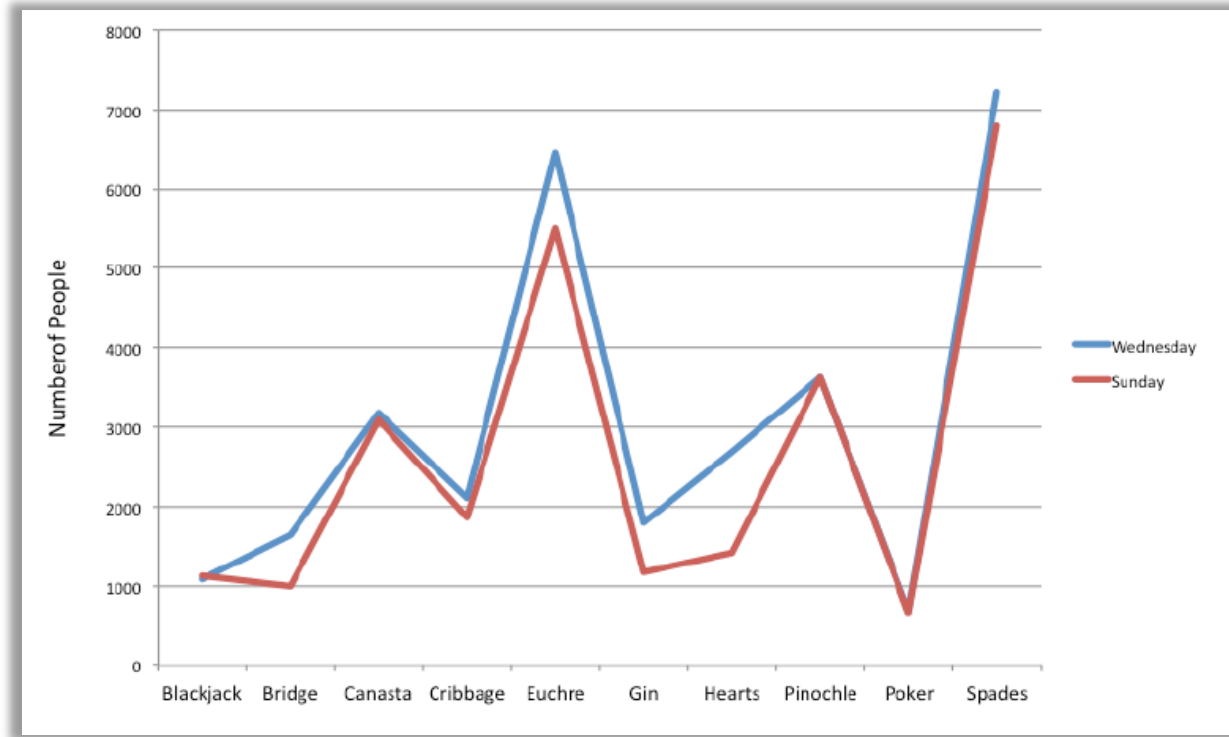
2. Life Expectancy by GDP



Independent variable on x-axis



What is Wrong With This Graph?





Visualization Tools

Visualization Tool Stack

Charting Tools

Excel, Many Eyes, Google Charts

Interactive Data Exploration

Tableau, Qlik, Lyra, Polestar, Voyager

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics / Visualization APIs

OpenGL, Java2D, Python, R

Charting Tools /
Graphical Interfaces

Declarative
Language

Programming
Toolkits

Visualization Tool Stack

Charting Tools

Excel, Many Eyes, Google Charts

Interactive Data Exploration

Tableau, Qlik, Lyra, Polestar, Voyager

Charting Tools /
Graphical Interfaces

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Declarative
Language

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics / Visualization APIs

OpenGL, Java2D, Python, R

Programming
Toolkits

Visualization Tool Stack

Charting Tools

Excel, Many Eyes, Google Charts

Interactive Data Exploration

Tableau, Qlik, Lyra, Polestar, Voyager

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics / Visualization APIs

OpenGL, Java2D, Python, R

Visualization Tool Stack

Charting Tools

Excel, Many Eyes, Google Charts

Interactive Data Exploration

Tableau, Qlik, Lyra, Polestar, Voyager

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics / Visualization APIs

OpenGL, Java2D, Python, R

How about
Power BI?

Visualization Tool Stack

Charting Tools

Excel, Many Eyes, Google Charts

Interactive Data Exploration

Tableau, Qlik, PowerBI, Lyra, Voyager

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics / Visualization APIs

OpenGL, Java2D, Python, R

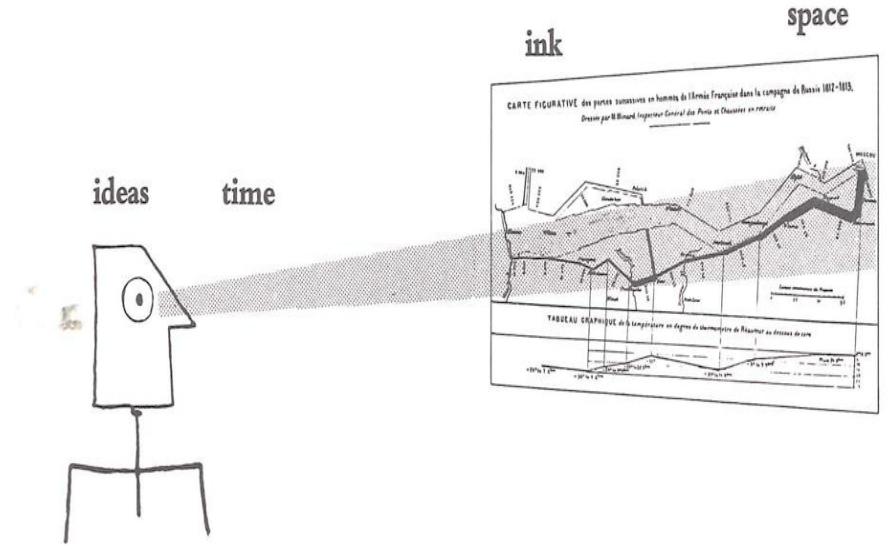


Recap

Graphical Excellence

Tufte's Principles:

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**

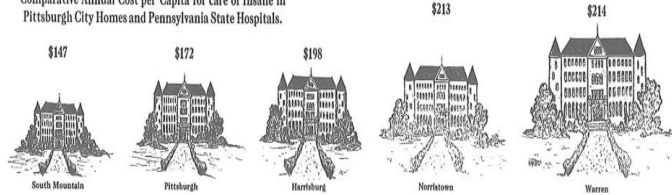


Graphical Integrity

Graphical Integrity Principles: Avoid distortion and ambiguity

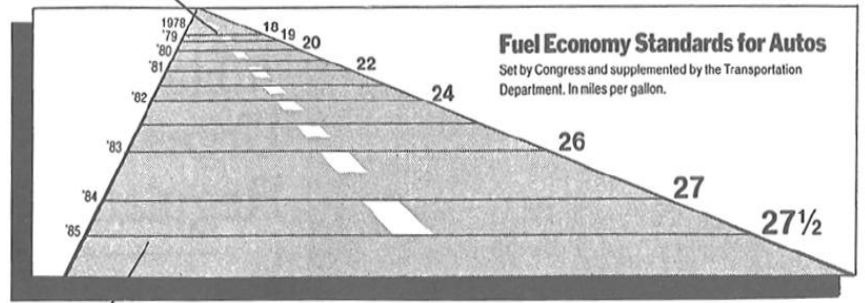
Lie Factor = 1.0

Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.



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

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

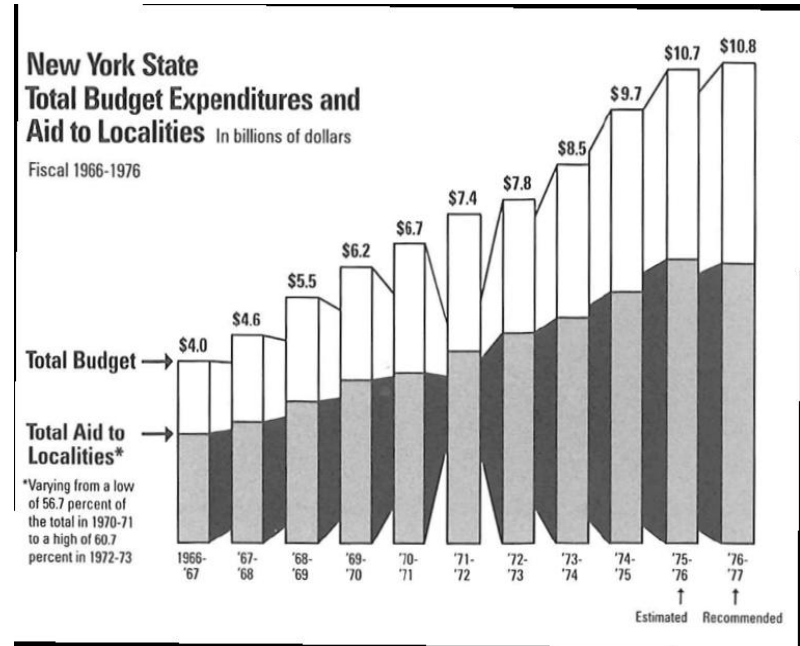


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Graphical Integrity

Graphical Integrity Principles

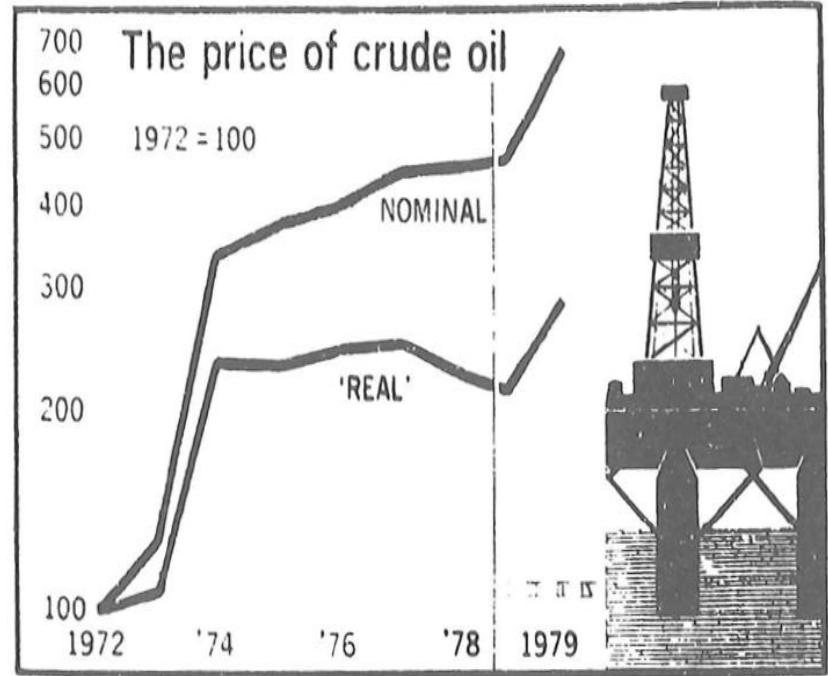
1. **Avoid distortion** and ambiguity
2. **Show data variation**, not design variation (avoid fancy chart junk / tricks!)
3. Account for **Inflation**



Graphical Integrity

Graphical Integrity Principles

1. **Avoid distortion** and ambiguity
2. **Show data variation**, not design variation (avoid fancy chart junk / tricks!)
3. Account for **Inflation**



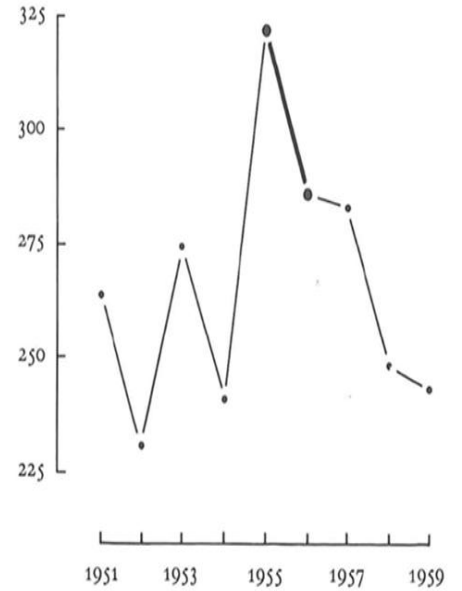
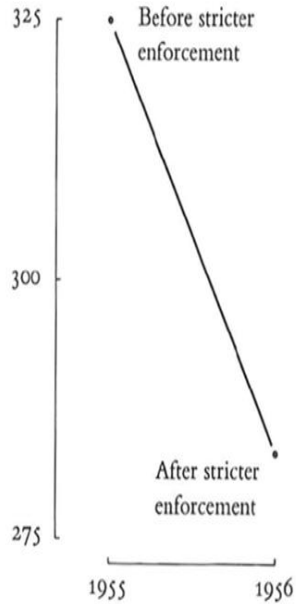
Tufte's Design Principles



- ❑ Data-Ink Maximization and Graphical Redesign
- ❑ Remove Chartjunk
- ❑ Multi-functioning Graphical Elements
- ❑ Data Density
- ❑ Small Multiples and Parallel Sequencing

Context

Graphics should be put in
context.



The KISS Principle



Keep it Simple, Stupid!

*“It seems that **perfection is reached** not when there is nothing to add, but when there is **nothing left to take away.**”*

- Antoine de Saint Exupéry