Scholarship Skills

Andrew Black, PSU

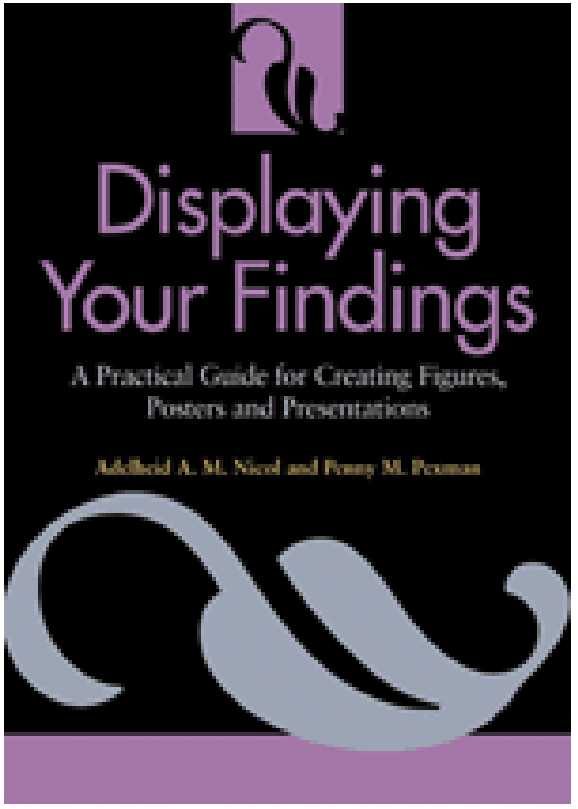
Figures, Tables & Graphics

Sources

Displaying Your Findings: A Practical Guide for Creating Figures, Posters, and Presentations
Adelheid A. M. Nicol and Penny M. Pexman. American Psychological Association. 2003

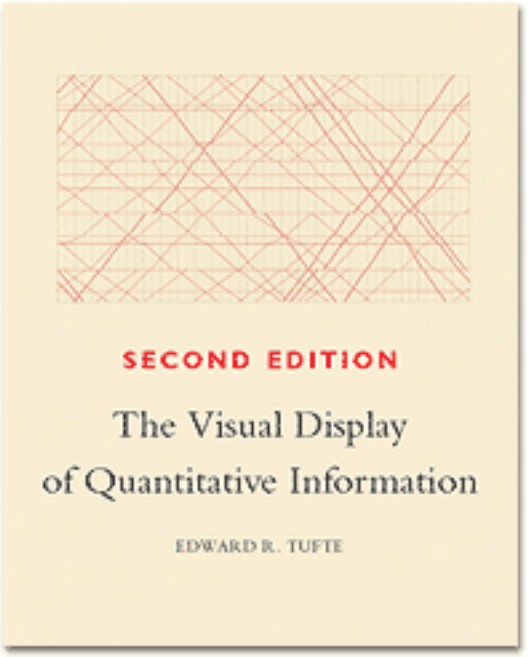
Used at Powell's and Amazon (<$10)

Practical, somewhat obvious, guidelines.



The *Visual Display of Quantitative Information* (2nd ed.). Edward Tufte. Graphics Press, 2001.

At Powell's and Amazon ($30)



Elevating, beautiful, even spiritual, one of the great books of the 20th century — but can be hard to put into practice

Visualizations Work

From Tufte:

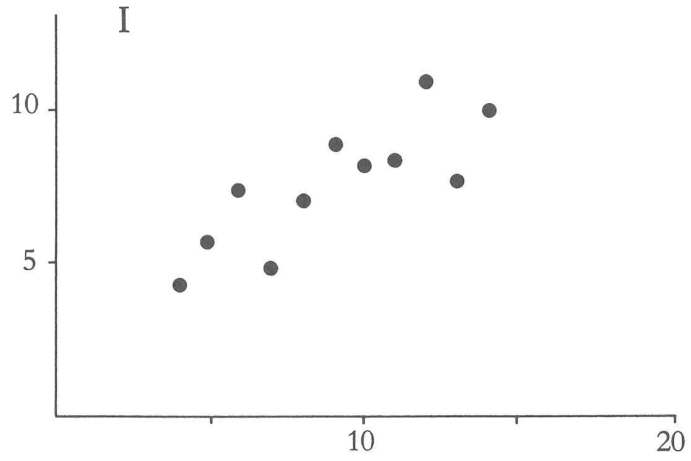
How do these measurements compare?

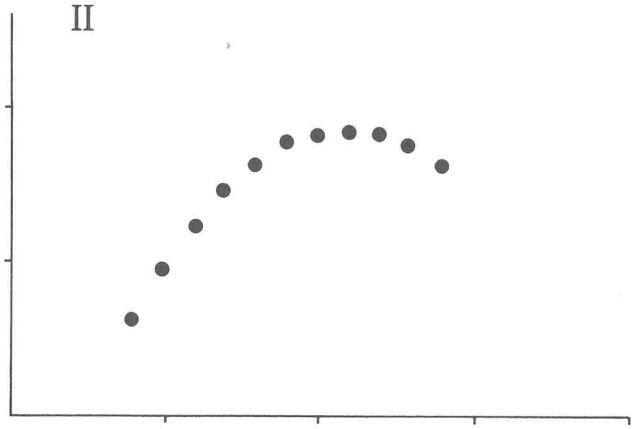
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0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

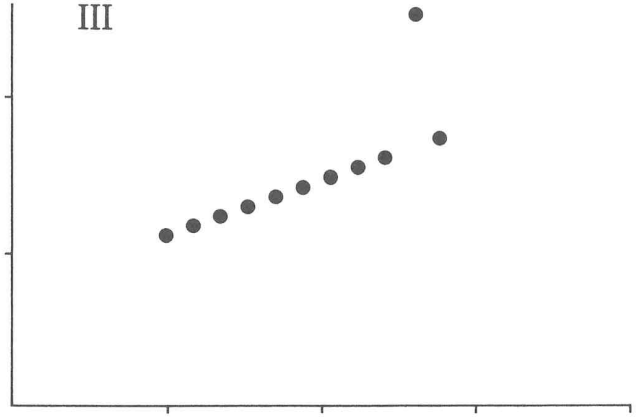
Lets' try some statistics:

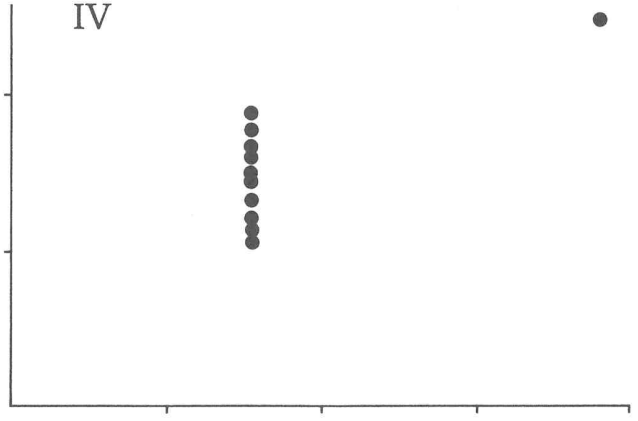
N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
t = 4.24
sum of squares $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$

Now let's try some graphs:

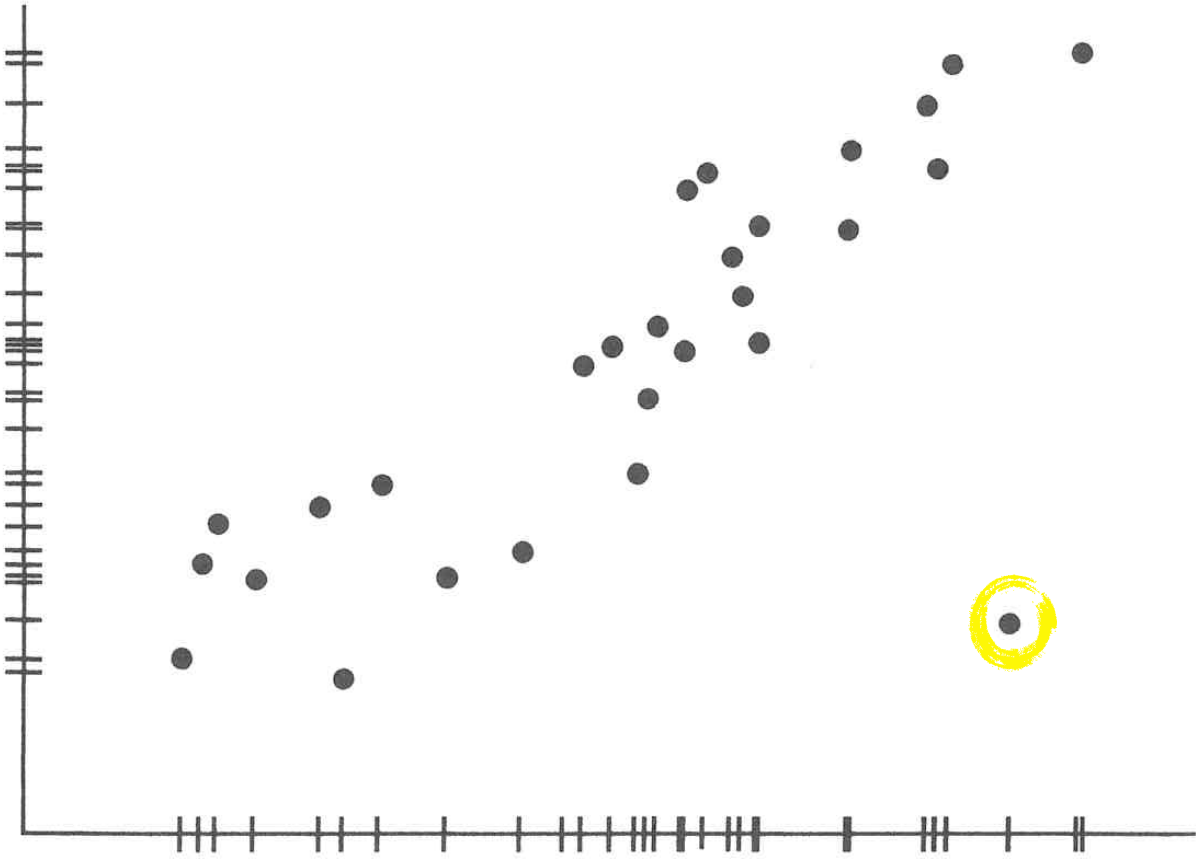




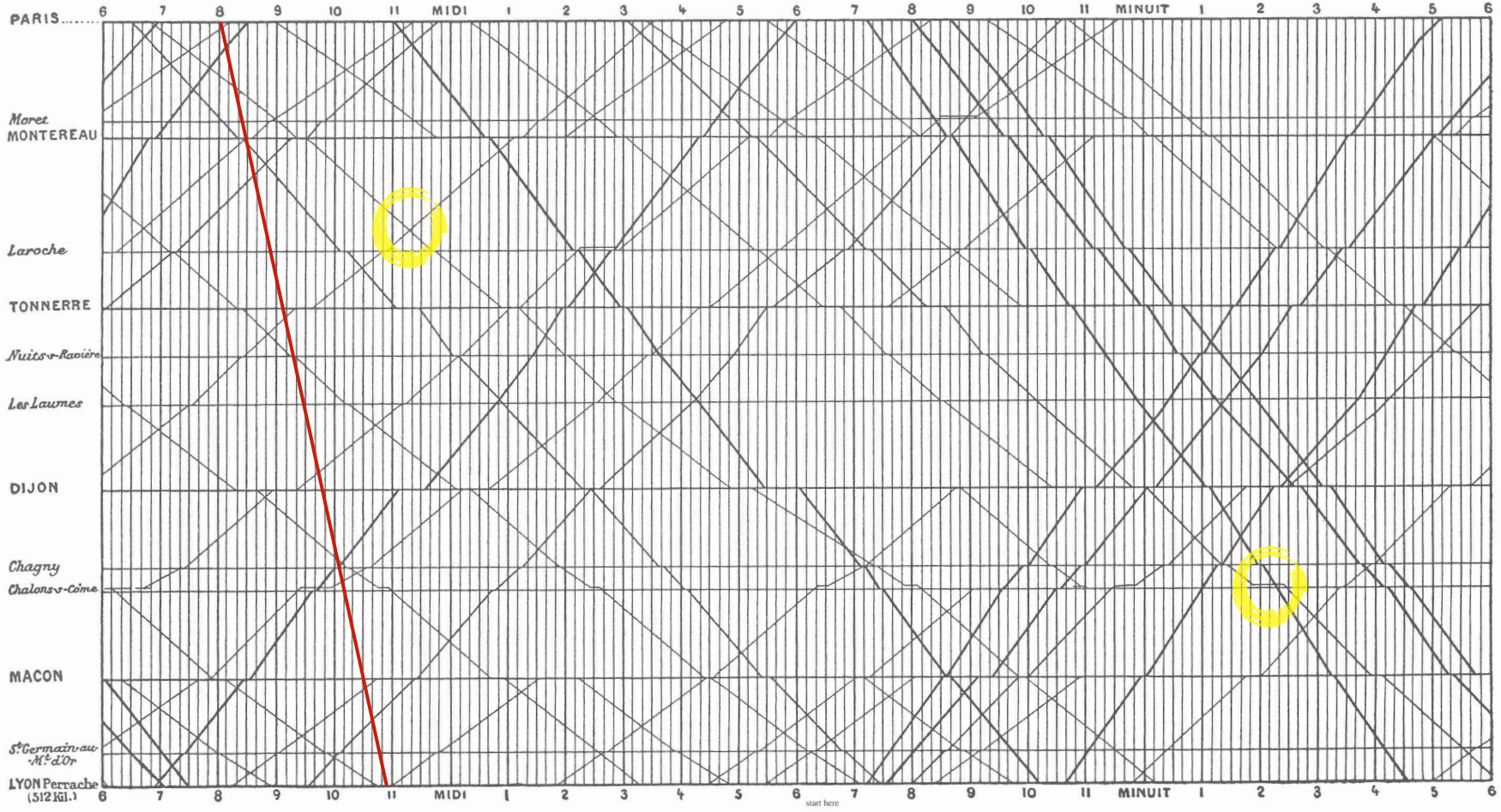




Detect outliers:



A Data-rich graphic:

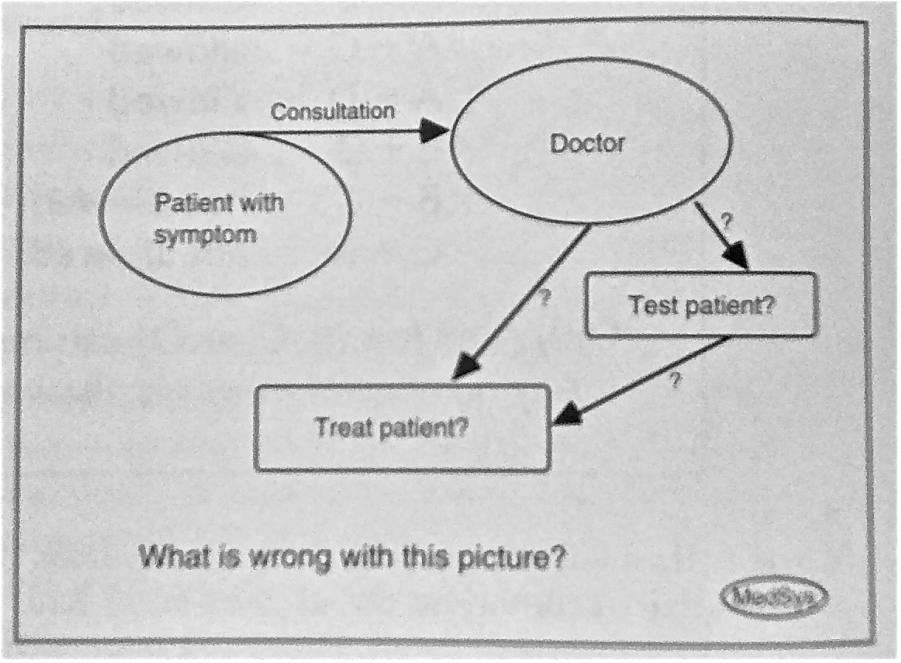


TGV: 1981

E. J. Marey, *La méthode graphique* (Paris, 1885), 20. The method is attributed to the French engineer, Ibry.

Dupré Segment 122

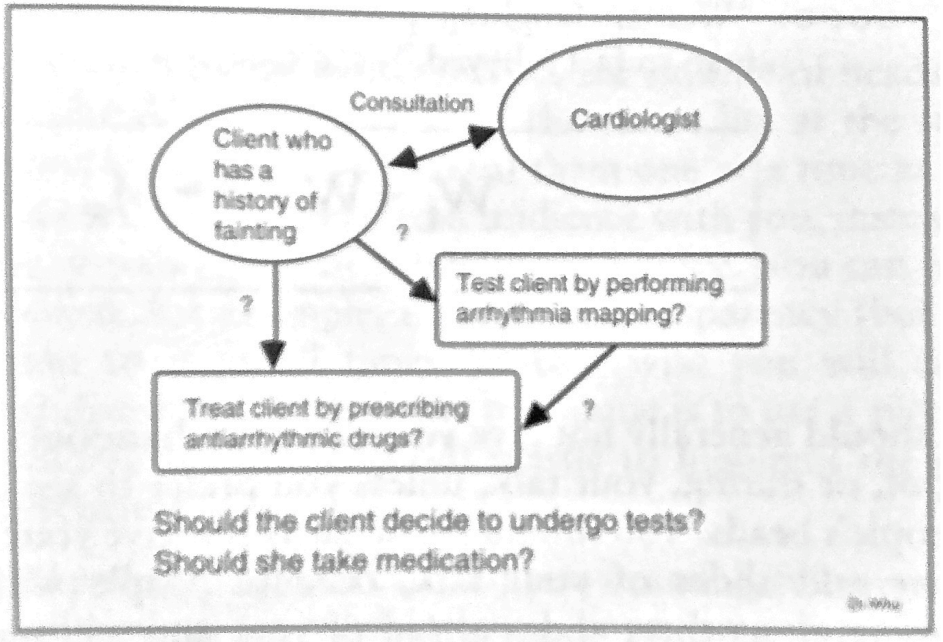
Explicit examples are more interesting and informative than vague examples



The patient presents with a symptom, and the doctor must decide whether to order diagnostic tests, and whether to prescribe treatment.

Dupré Segment 122

Explicit examples are more interesting and informative than vague examples



The patient presents with a history of fainting, and the physician must help the patient to decide whether to undergo arrhythmia mapping (an invasive procedure), and whether to take antiarrhythmic drugs (which have unpleasant side effects).

What's a Figure?

Any kind of graph, chart, plot, drawing, or photograph…

that is not a table, and is not running text.

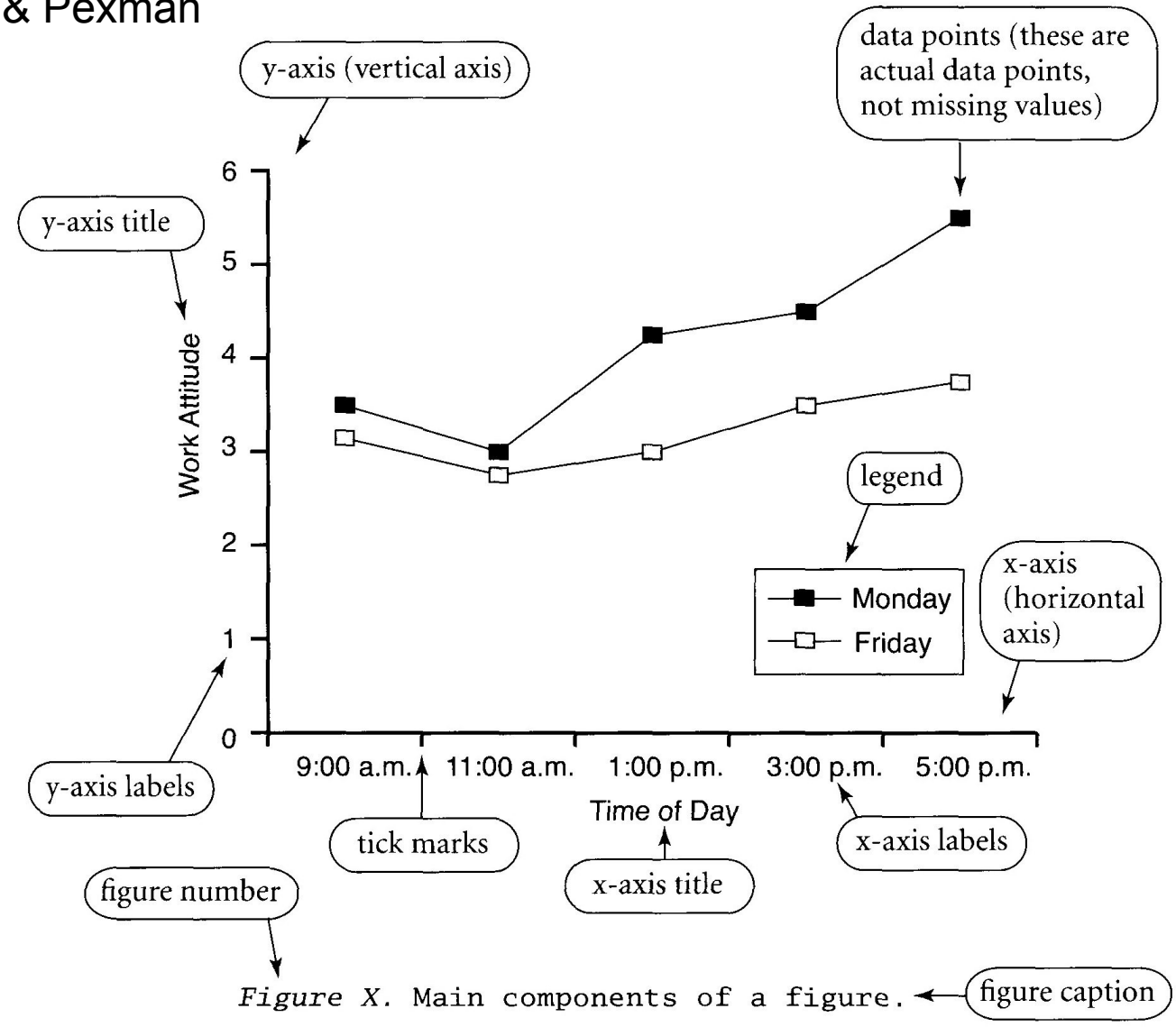
code segments can be figures

Three parts to a figure:

1. Graph or graphic
2. Legend
3. Caption

Anatomy of a Figure

from Nicol & Pexman



General Guidelines for Figures

Figure must be relevant to the paper

it should move the story along

Image should be as simple as possible

no chartjunk!

Labels should be concise

Fonts in all figures should be consistent

Avoid color in figures for journal articles, conference proceedings, etc.

Specify units *within* the figure

Figures should stand alone

all information necessary to interpret the figure should be included in the caption.

Points to Watch

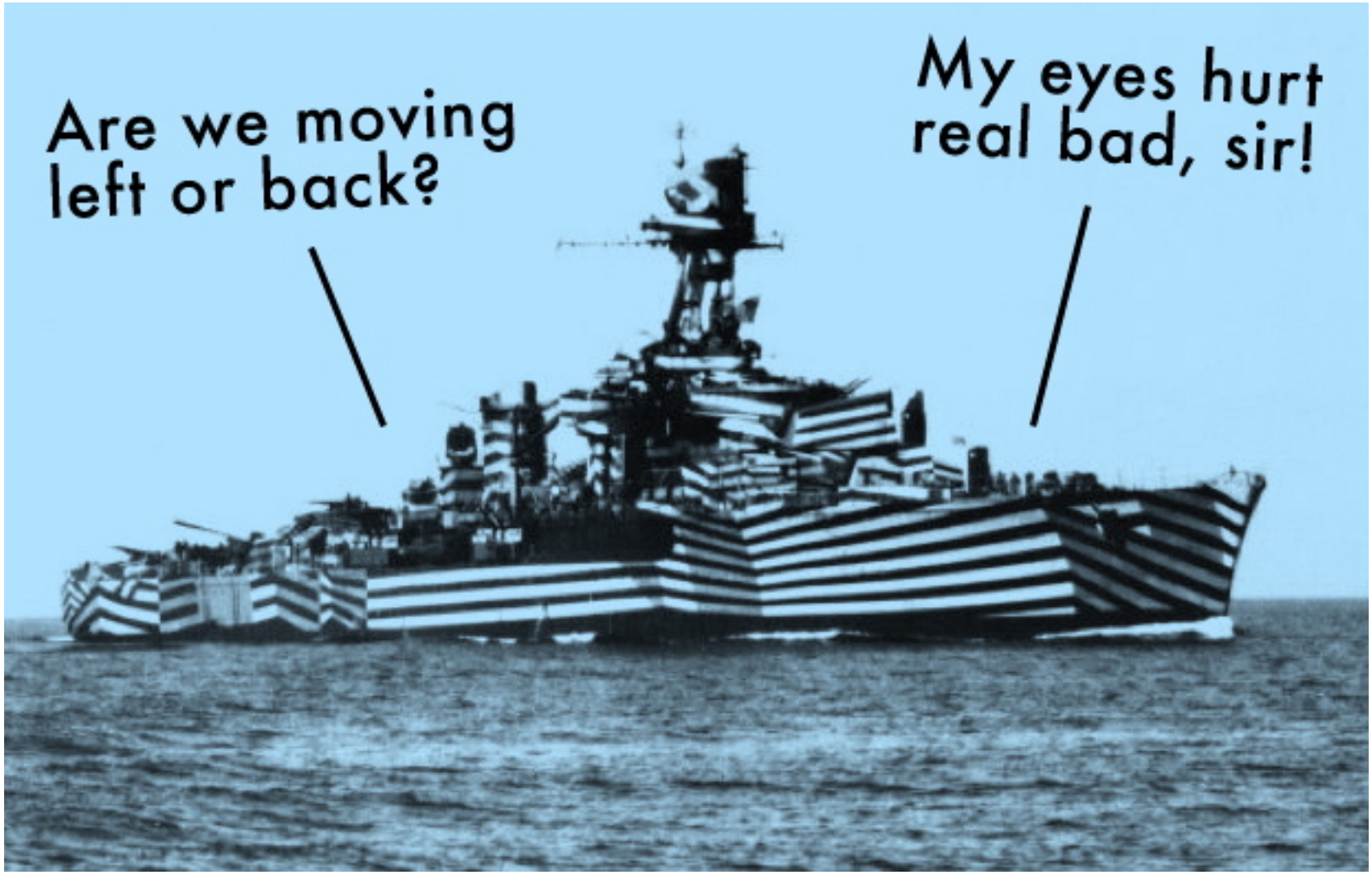
Figures may be reduced to fit the page

- Are they still legible? Is the type size still appropriate?
- Sans serif fonts scale better than serif fonts

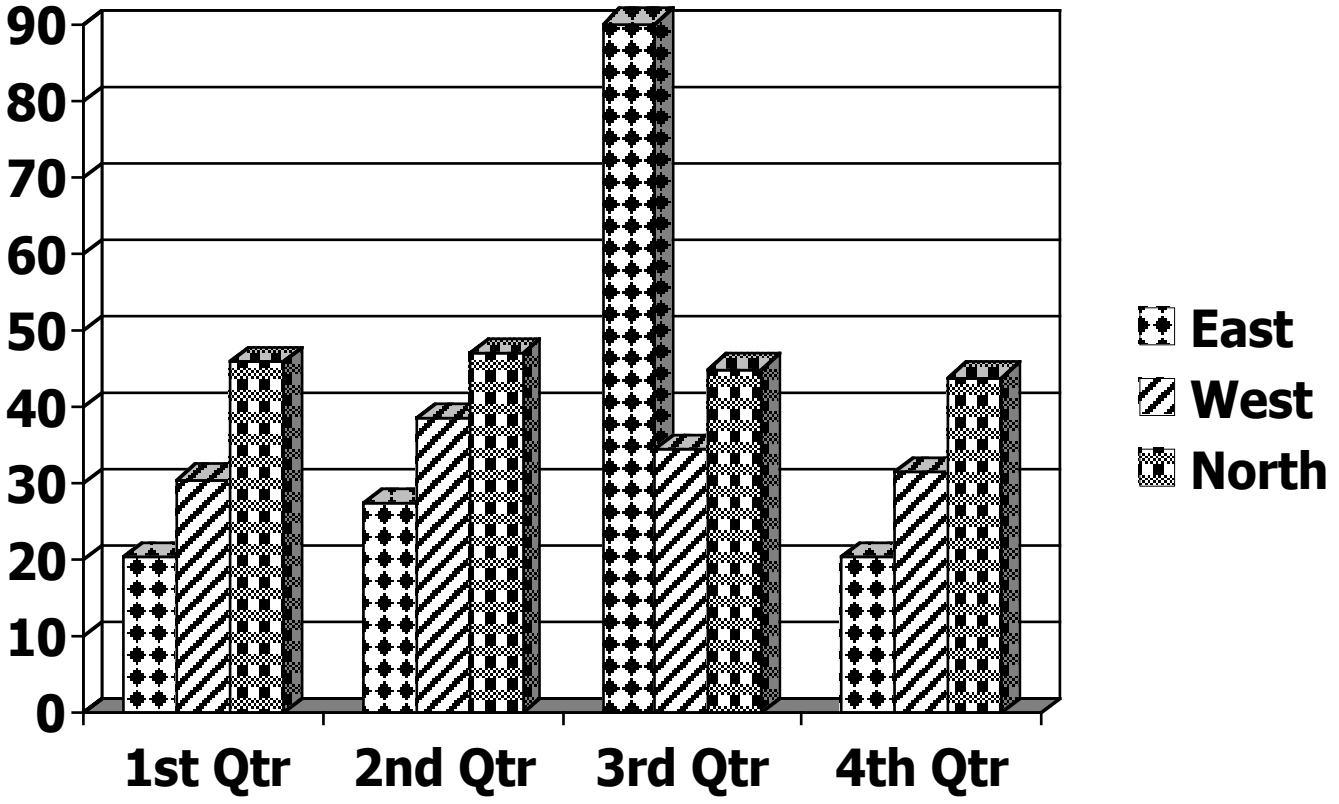
Similar figures (within same article) have similar style

- All text in same font style, and point sizes don't vary by more than 40%.

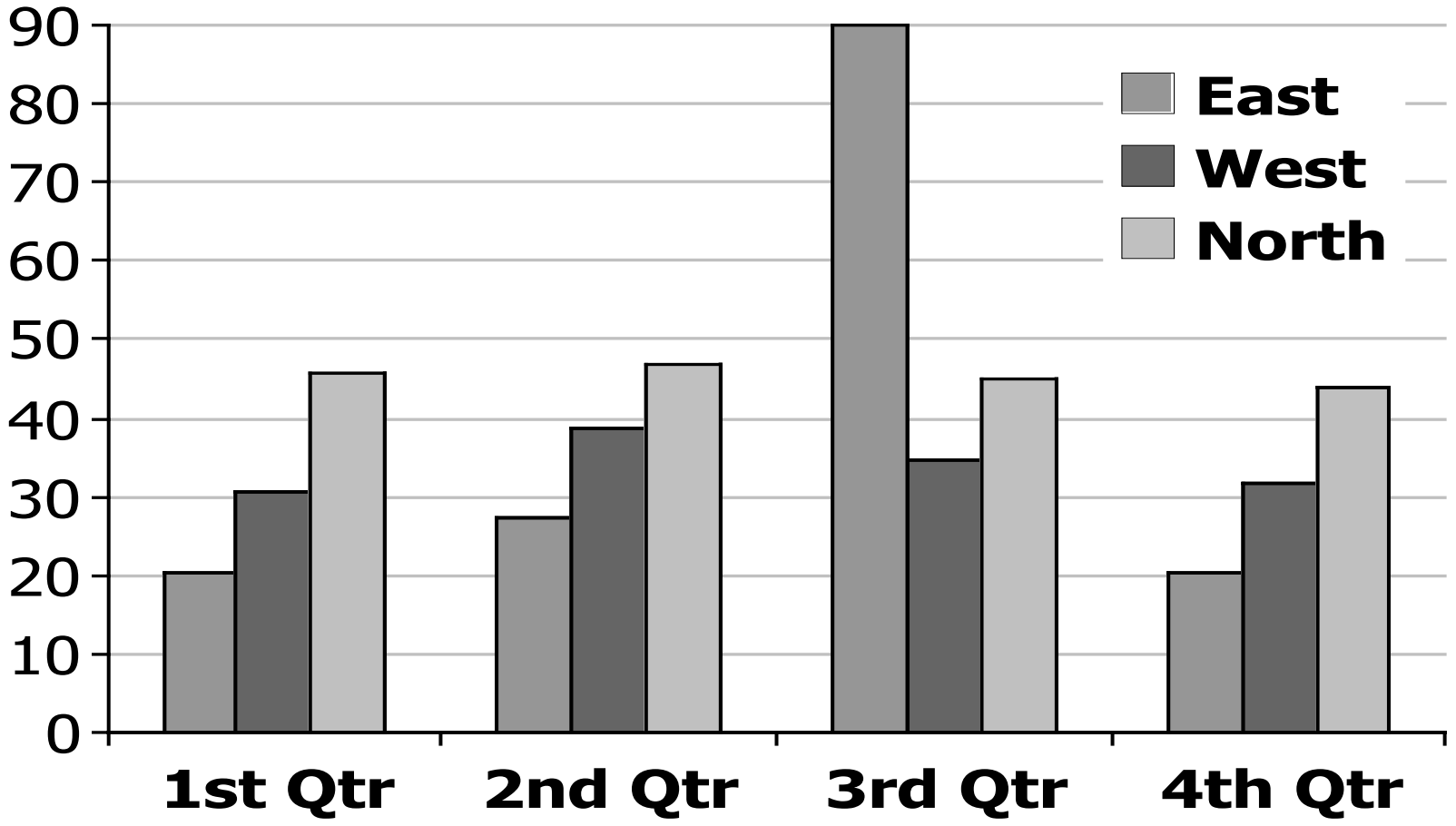
Dazzle Camouflage



Use shading carefully



Prefer Grayscale



Make the figure easy to read

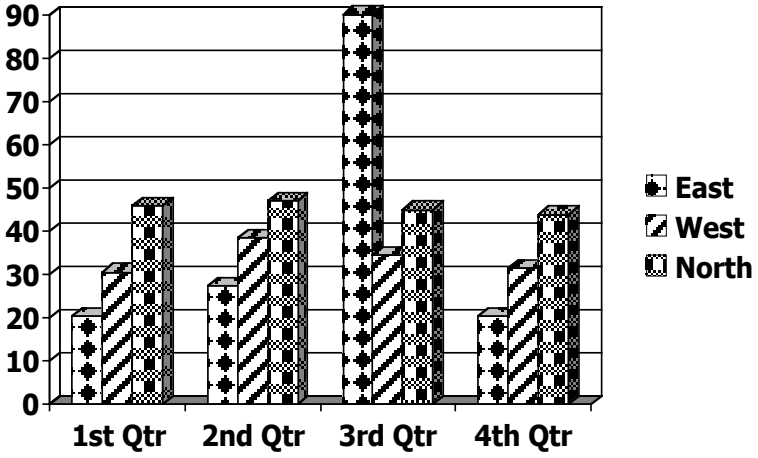
Avoid Bogus use of 3D

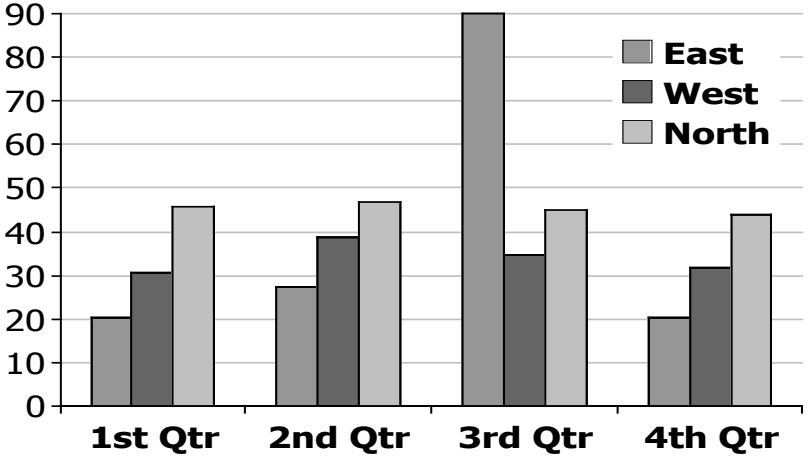
Put legend *in* the image, not next to the image

- better: avoid legend by labeling the graph directly

Omit gridlines, or use a pale grey

Omit “walls”

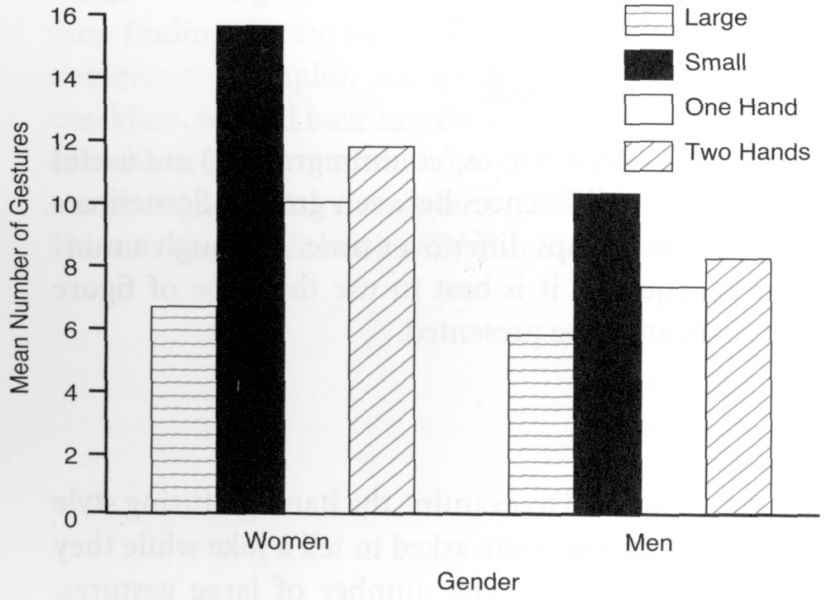




1: Bar Charts

The value of the dependent variable is most frequently placed on the y-axis (vertical axis).

The y-axis to x-axis length ratio should be appropriate (e.g., the y-axis should be from two thirds to three quarters the length of the x-axis).



The top of each bar indicates the value for a particular group (e.g., women presented an average of 6.5 large hand gestures).

The independent variable is most often placed on the x-axis (horizontal axis).

Figure X. Average number of times different types of hand gestures were used by women (n = 20) and men (n = 20) when telling a joke over a 5-min period.

Although differences between the patterns used to identify the bars may be observable on the computer monitor, when printed they may be indistinguishable. The appearance of the shading can best be verified by checking a laser printout.

Color can be used instead of shading for overheads, multimedia, or poster presentations to differentiate the bars. Most books and journals do not print color.

Another version

Adds numbers to the top of the columns — better than gridlines

In this version of the figure, the numerical values for the dependent variable are provided above the top of each bar.

The legend should be placed within the boundaries of the axes so that the figure requires as little reduction as possible.

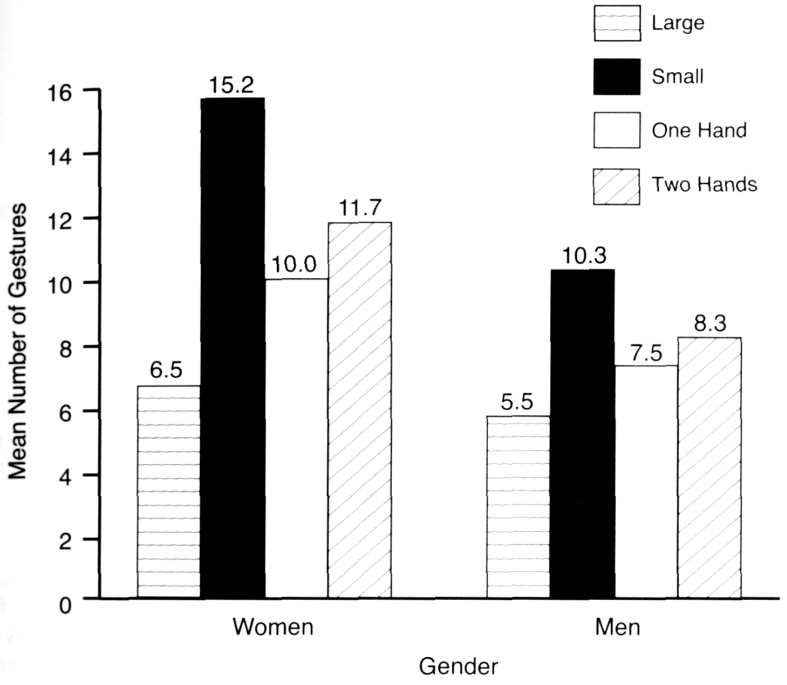


Figure X. Average number of times different types of hand gestures were used by women (n = 20) and men (n = 20) when telling a joke over a 5-min period.

The larger space between the *Women* and *Men* bars and the shorter space on either end of the x-axis help readers visually group the bars.

Little or no space should separate bars grouped within one level of the independent variable (e.g., the four bars within *Women*). Large spaces would make the graph not only larger but more difficult to read.

Adding graphics to a table

Category	JDT	CDT	RDT	HaRe
analysis problem	0	0	0	0
inaccurate analysis	35	4	2	2
incompatibility	5	1	0	0
compilation errors	27	1	3	0
internal error	24	5	0	36
inconsistent state	15	2	0	0
unsaved	4	1	0	0
deleted	4	0	0	0
misselection	0	0	0	0
selection not understood	30	26	19	33
improper quantity	0	5	0	2
misconfiguration	3	0	0	0
illegal name	6	7	1	15
unconventional name	11	0	0	0
clash	6	5	0	24
control clash	17	3	5	9
data clash	16	0	3	3
name clash	38	3	0	2
inheritance clash	9	0	0	0
inherent	0	0	0	0
context	38	0	7	4
own parent	4	0	0	0
structure	17	0	13	9
property	45	3	3	0
vague	37	1	0	22
unknown	6	1	2	21

TABLE 2
Our taxonomy of preconditions (column 1), with counts and bars indicating the number of error messages in each category for each refactoring tool (columns 2–5).

Box and Arrow Diagram

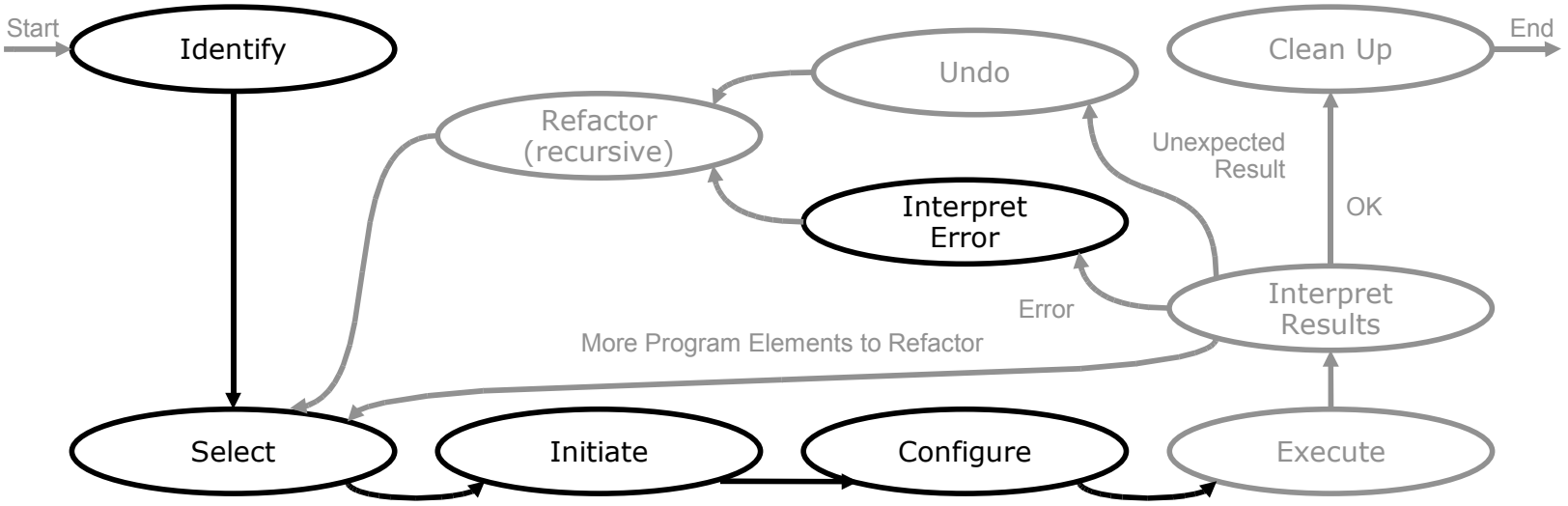
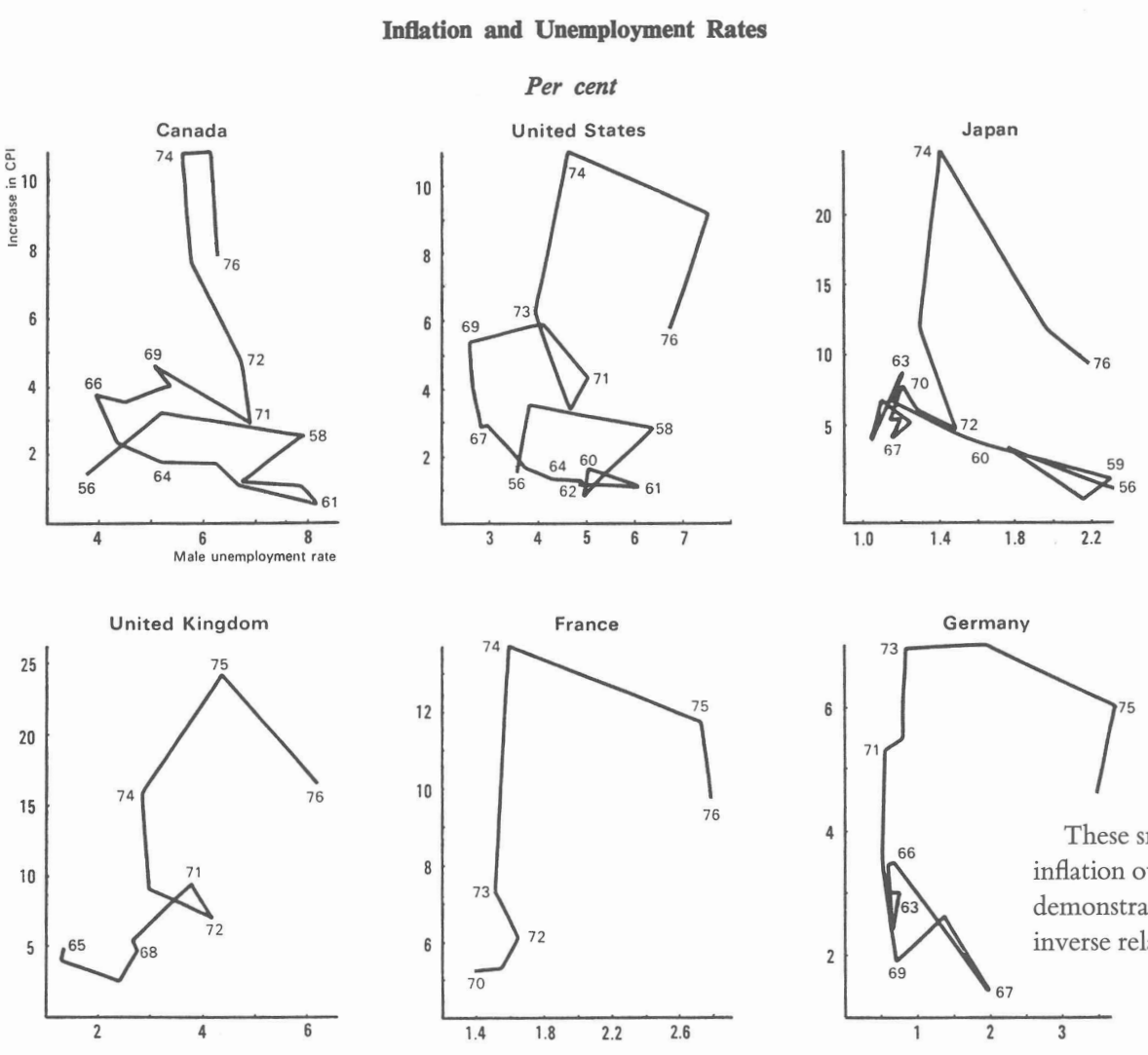


Figure 1.5: A model of how programmers use conventional refactoring tools. Steps outlined in black are the focus of this thesis.

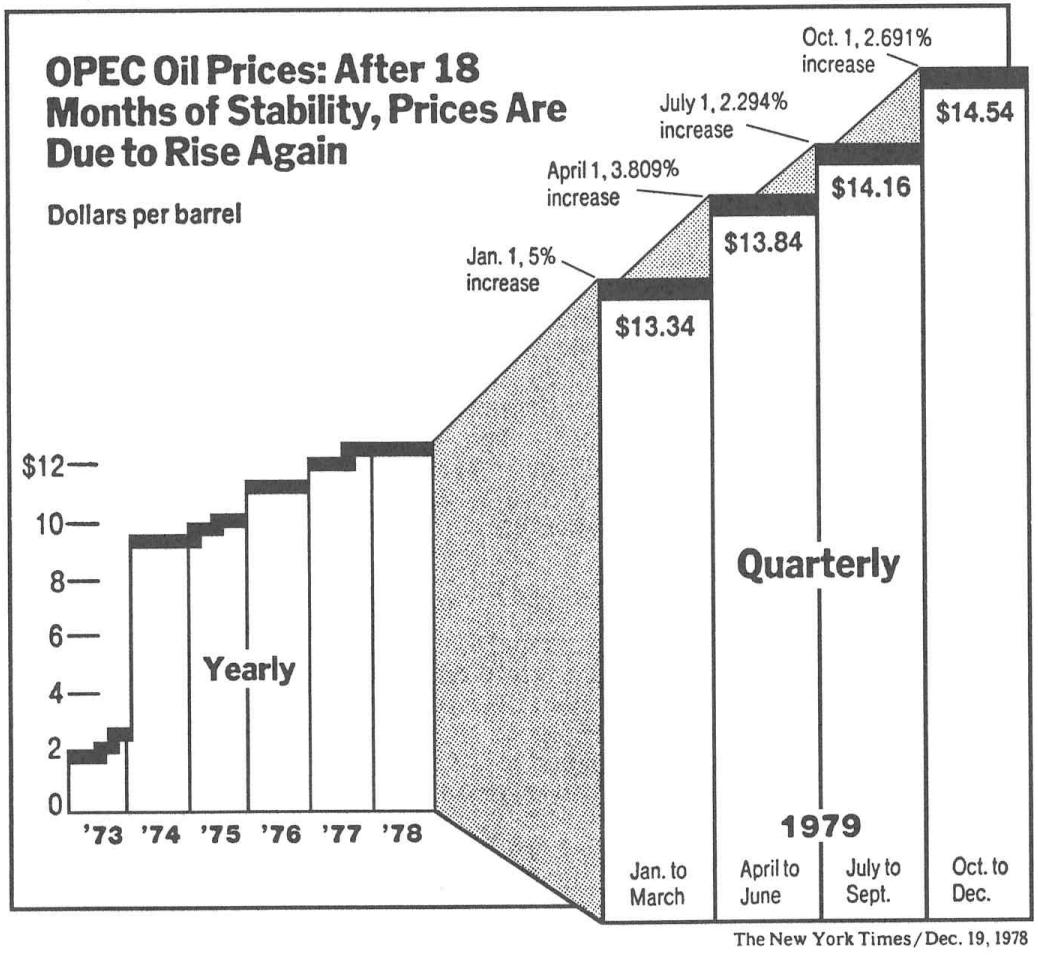
More than one way to show a time-series:



These small-multiple relational graphs show unemployment and inflation over time in "Phillips curve" plots for nine countries, demonstrating the collapse of what was once thought to be an inverse relationship between the variables.

Some Principles from Tufte

Show data variation, not design variation



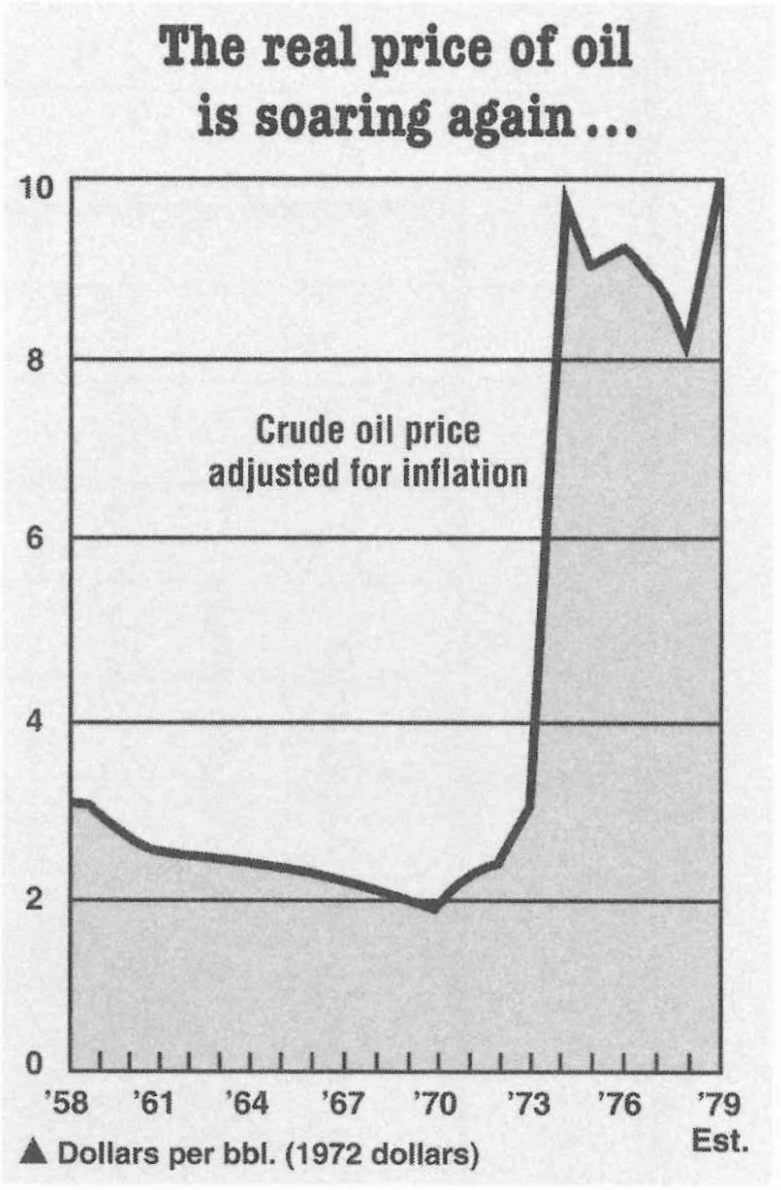
five different vertical scales,

two different horizontal scales

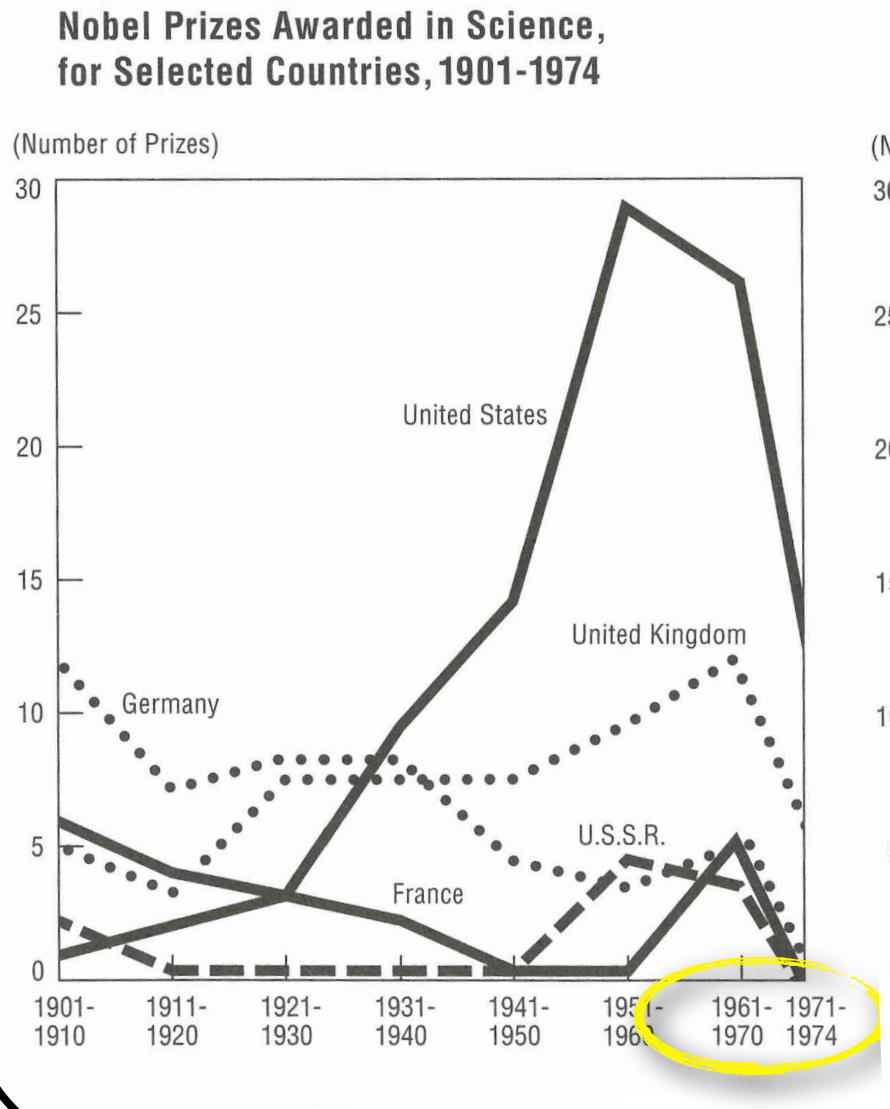
on the left, \$10 ~ 0.31 in^2
on the right, \$10 ~ 4.69 in^2

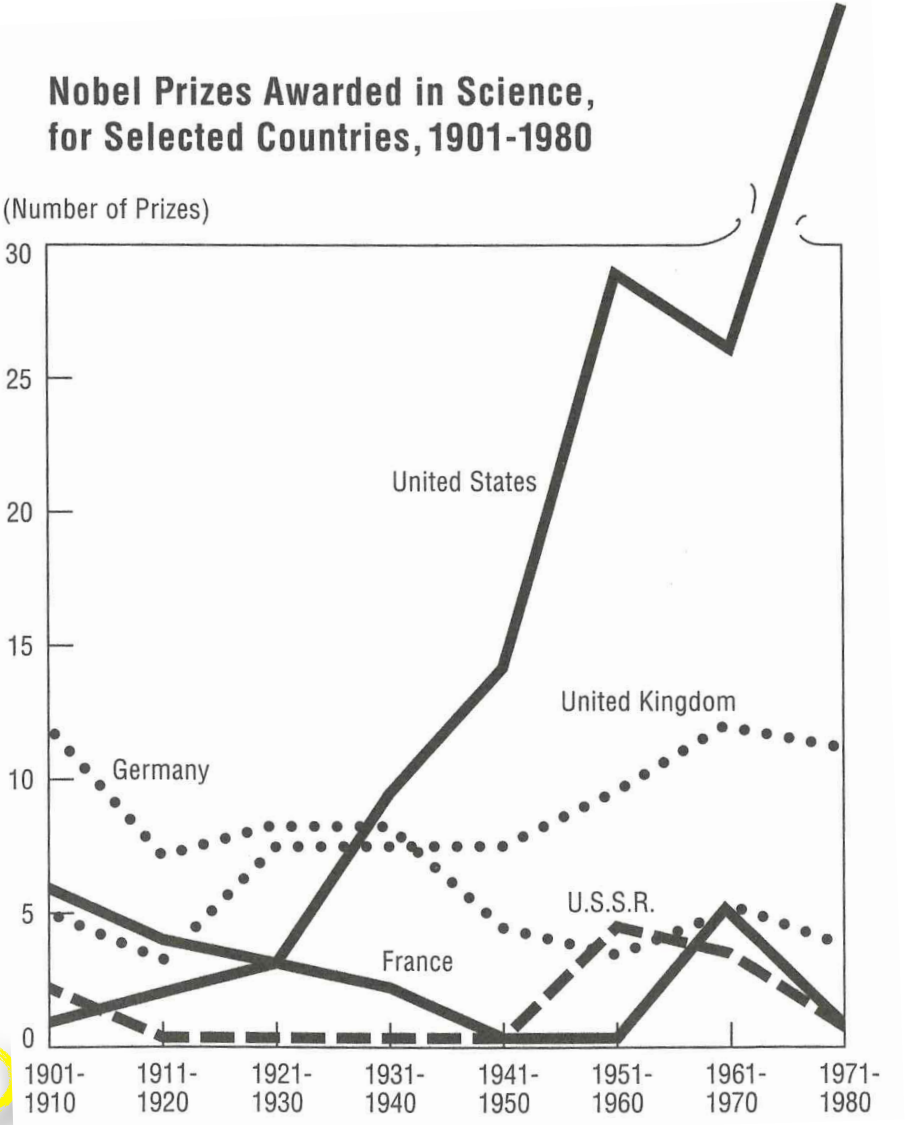
The real story:

Business Week,
April 9, 1979, p99



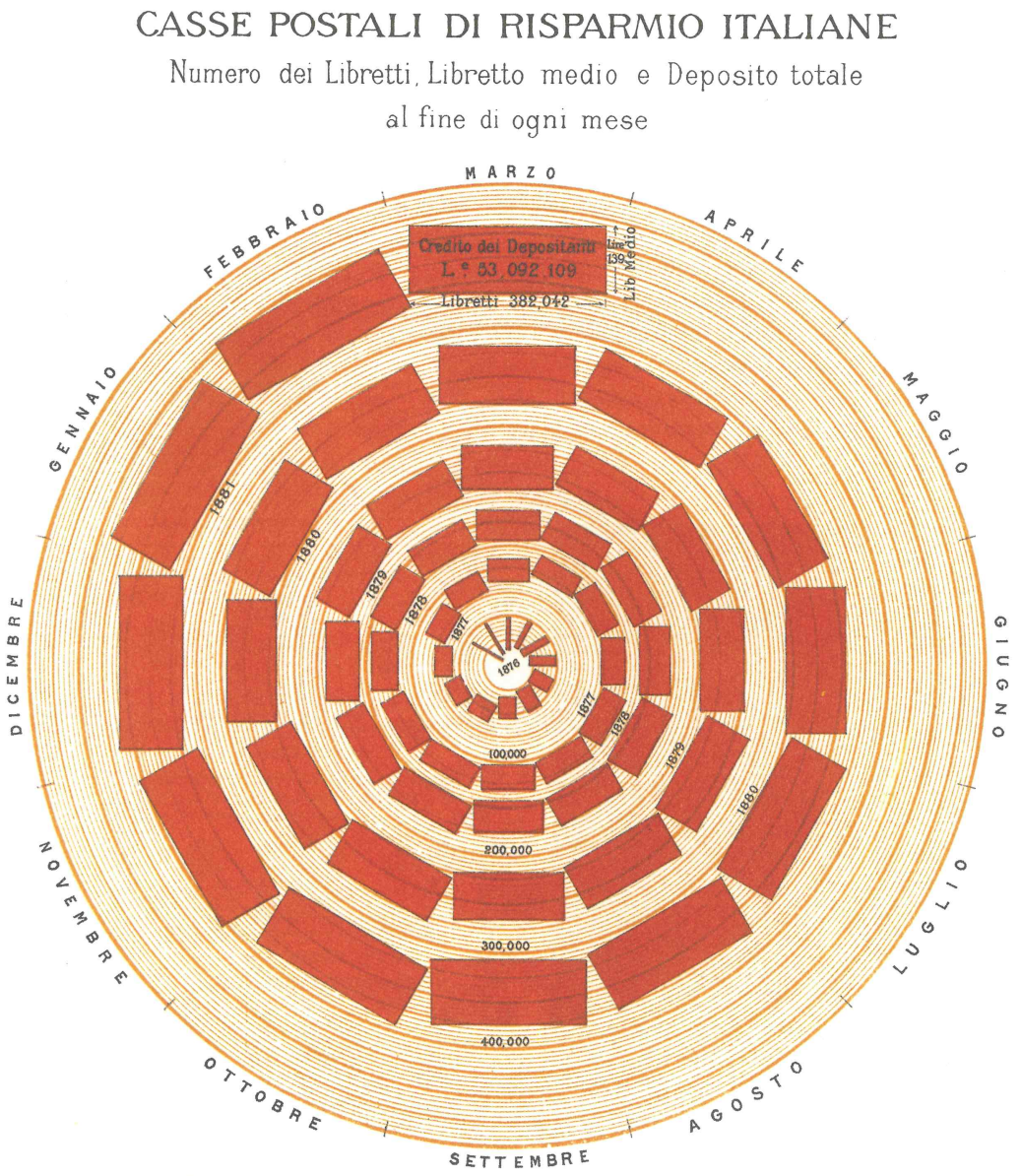
What's wrong with this?



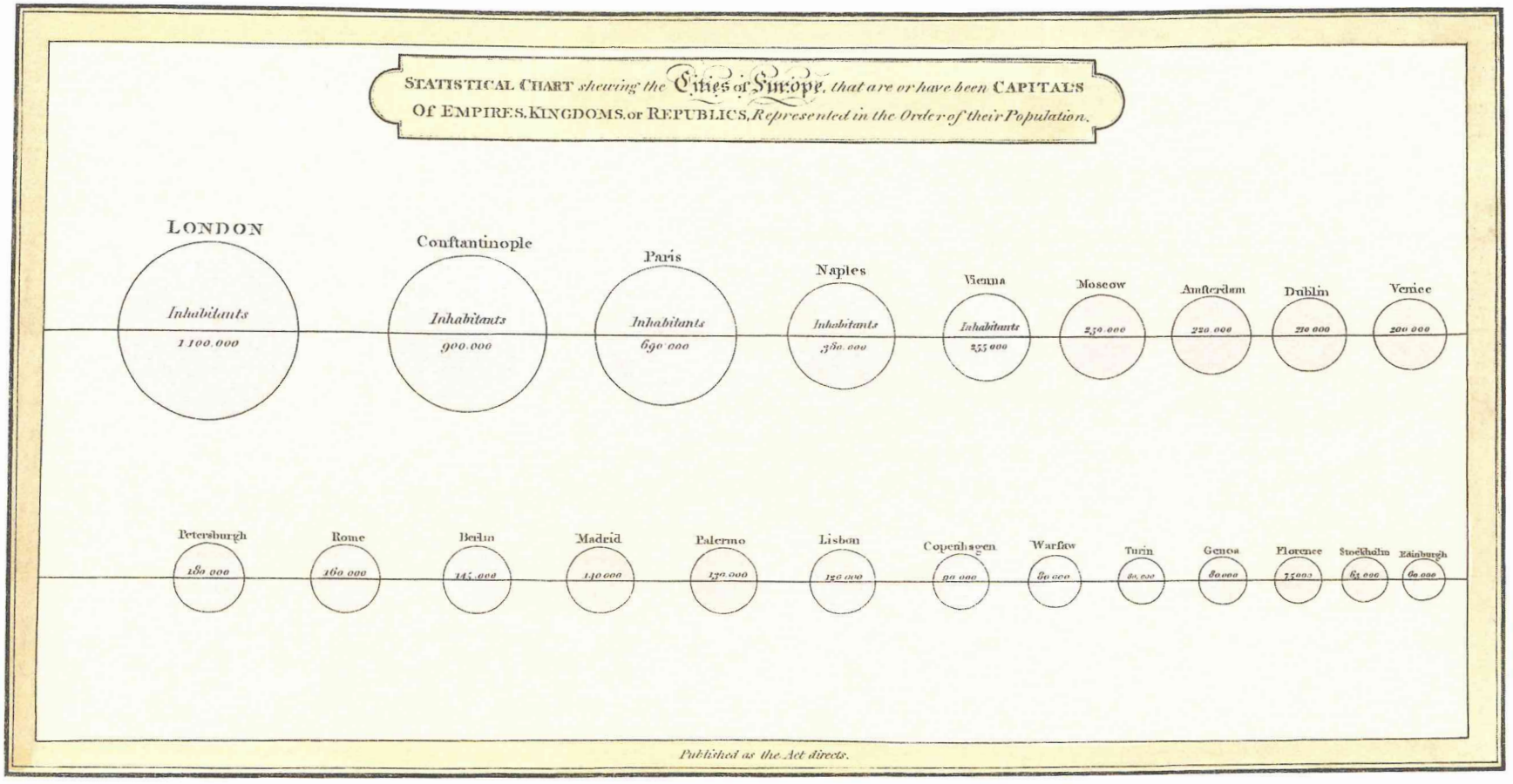


Number of dimensions in the graphic should not exceed the number of dimensions in the data

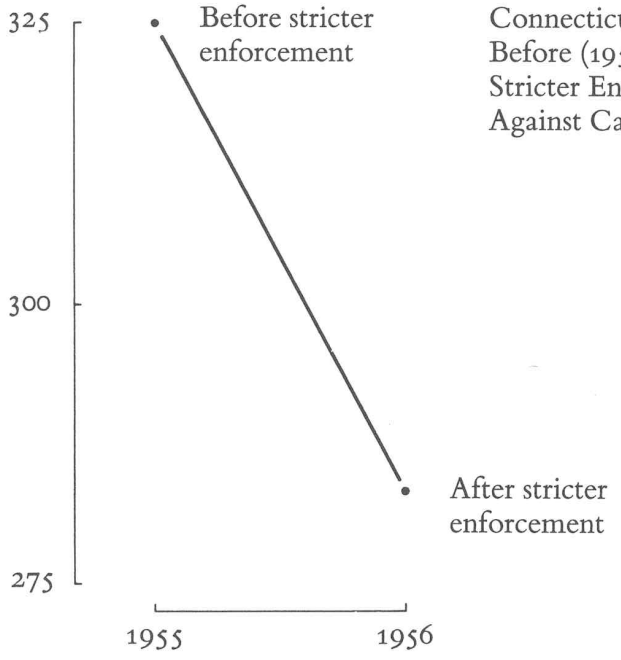
- Number of accounts, average Lira amount, and date



Counter-example (from Playfair): uses circles to show 1-dimensional population

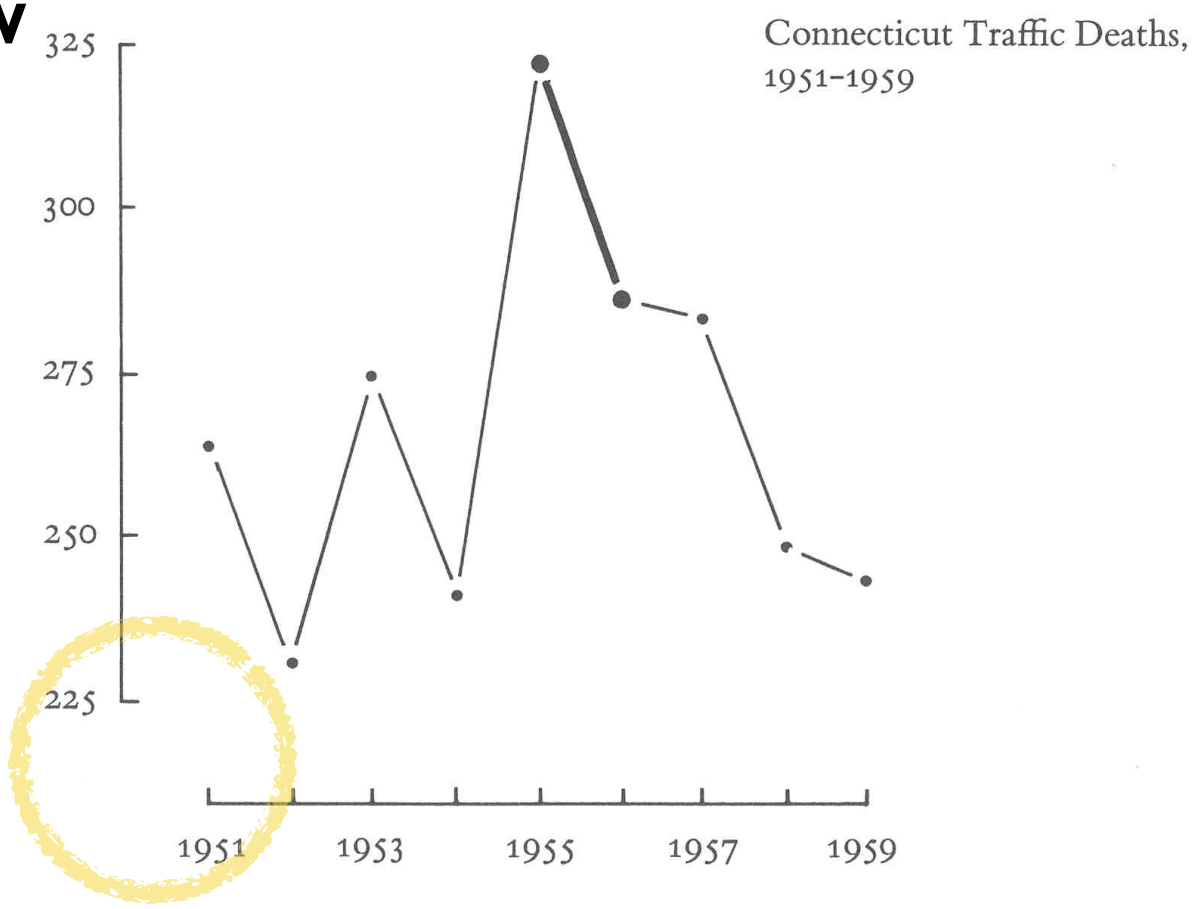


Don't show data out of context



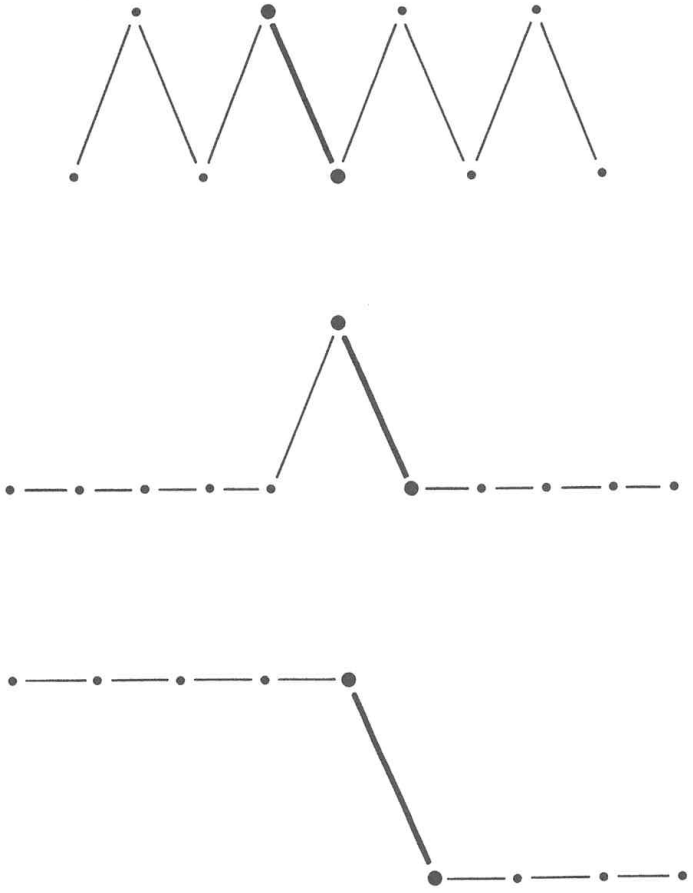
Connecticut Traffic Deaths, Before (1955) and After (1956) Stricter Enforcement by the Police Against Cars Exceeding Speed limit

Let's add some context! A few more data points add immensely to the story:



Context is everything!

Imagine the very different interpretations other possible time-paths surrounding the 1955–1956 change would have:



Tufte on Graphical Excellence

Graphical excellence is a matter of *substance*, *statistics* and *design*

An excellent graphic* communicates complex ideas with *clarity*, *precision* and *efficiency

***An excellent graphic* gives the viewer:**

- the greatest number of ideas,
- in the shortest time,
- with the least ink,
- in the smallest space

- **Excellent graphics are usually multivariate**
- **Excellent graphics tell the truth about the data**