Graphs

STAT8801
Statistical Consulting

School of Statistics
University of Minnesota

March 22, 2010

Good Graphics

Fundamental Principal of Statistical Graphics

Above all else show the data.

Ed Tufte

Graphics can be

- . . . all that is read in an article
- . . . efficiently summarize a problem
- . . . very aesthetic
- . . . misleading or otherwise awful

We must use them well, or else who will?

The Aesthetics of Graphics

Ed Tufte is at the top of the pantheon of statistical graphics gods.

Tufte has three **extremely** influential books on graphics.

Not everyone agrees with Tufte, but no one can ignore him.
Other important sources:

- Lee Wilkenson (**The Grammar of Graphics**)
- Bill Cleveland (**The Elements of Graphing Data**)
- Howard Wainer (lots of articles)

Technique

- Graphs may be the only part of an article that is read.
- Good format and design
  Aesthetics, elegance, and style difficult to prescribe.
  Construct, revise, edit, try again
- Words/numbers/graphics together
- Data graphics are paragraphs about numbers (Tufte, p 181).
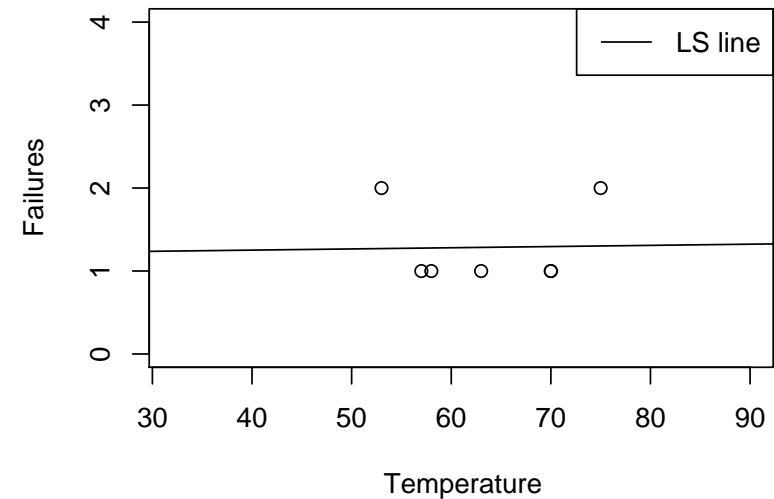- Graphics and tables must always reinforce message and text.

Show the right data

Show the <u>right</u> data
- Show enough data
- Don't hide important data
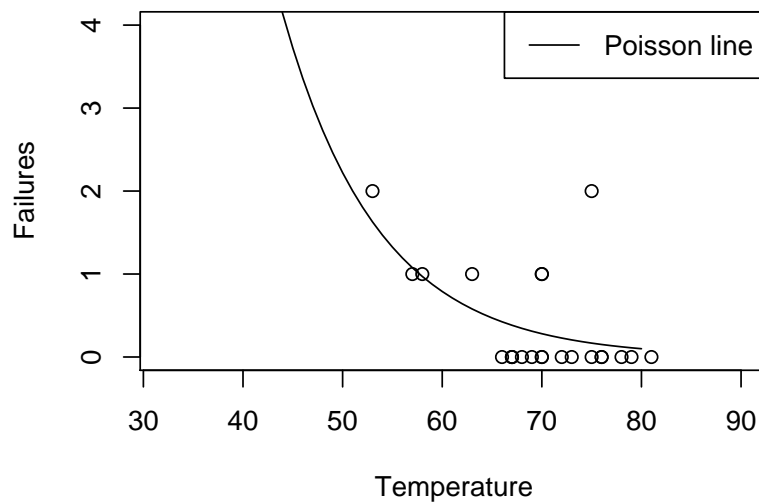- Keep data in context
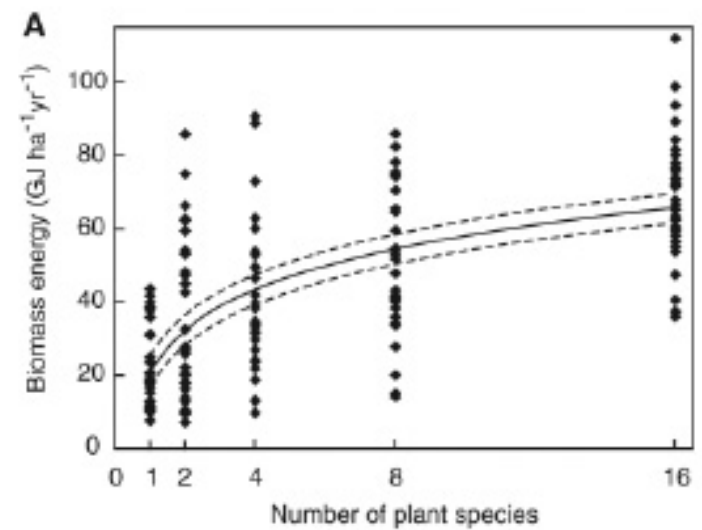
The Worst Graph Ever
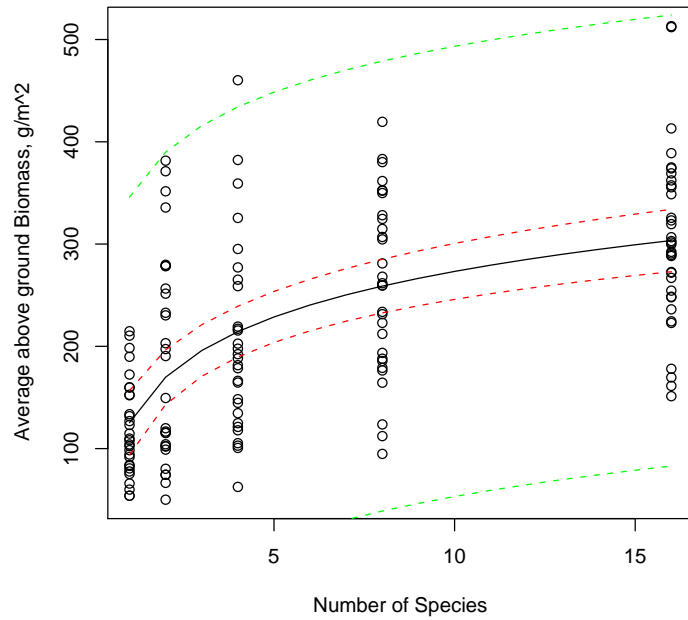

Challenger data

What they should have done

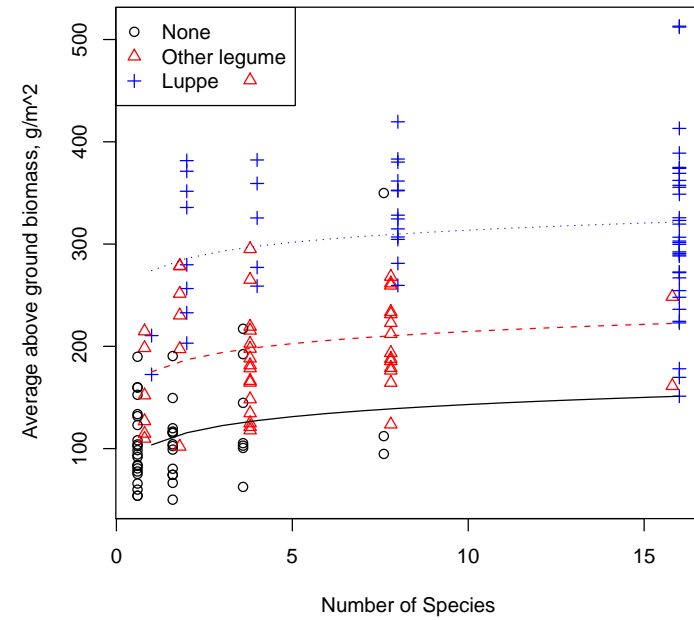**Challenger data**

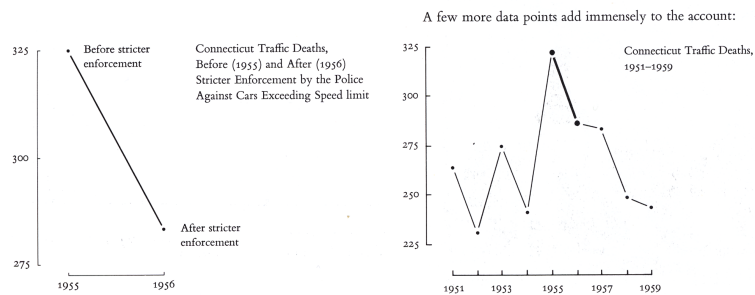## From Tilman, Hill and Lehman (2006) <u>Science</u>, p. 1598

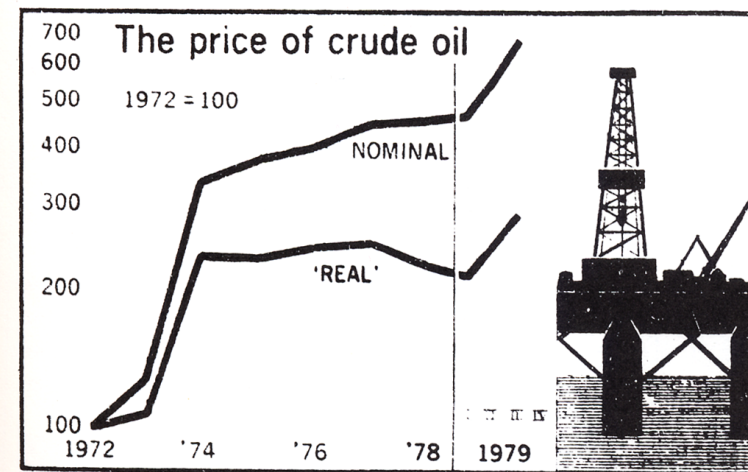## . . . adding prediction intervals

. . . adding species indicator

Connecticut Traffic Deaths

Correct for inflation
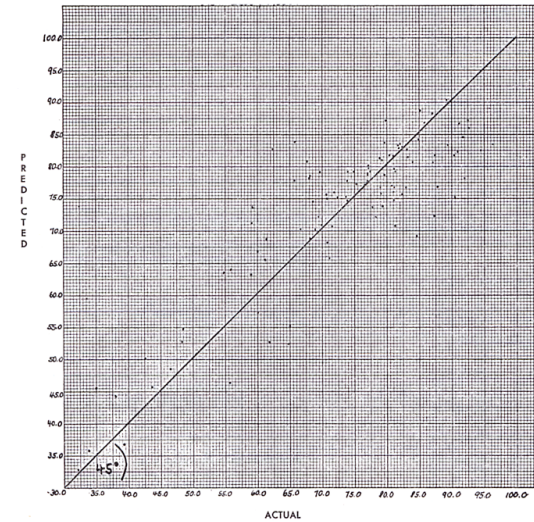
Show only the data

Definition (Data ink)

Data ink is the "ink" that displays non-redundant data information.
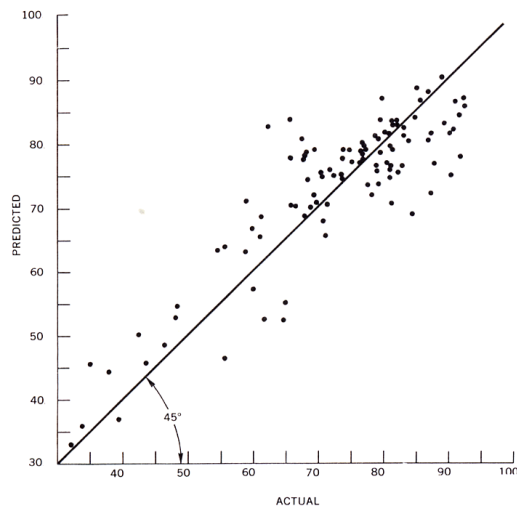
Definition (Data ink ratio)

Proportion of a graphic's ink devoted to the non-redundant display of data information.

1. Maximize data ink ratio, within reason
2. Erase non data ink, within reason
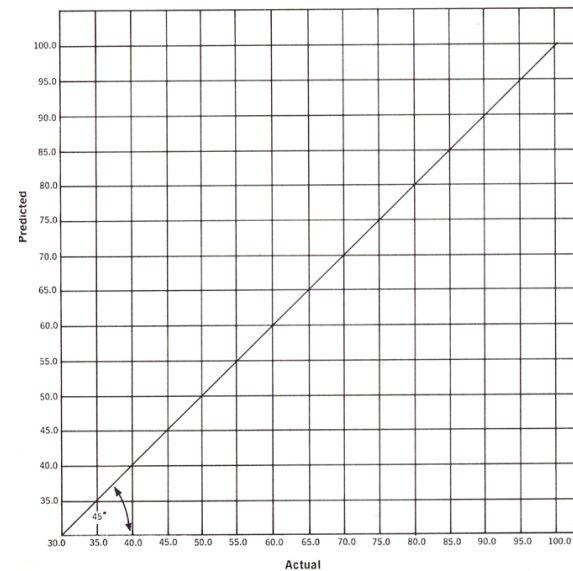3. Erase redundant data ink, within reason

Bad data-ink ratio
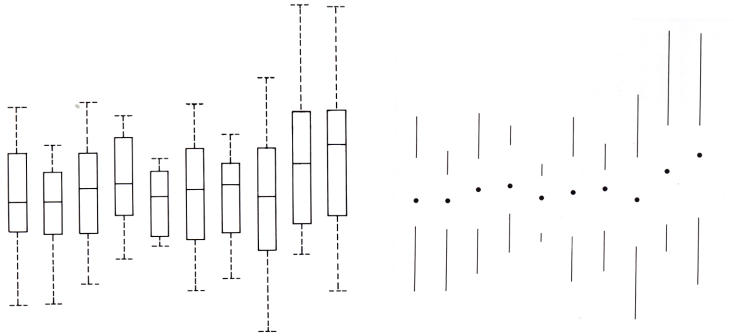
Good data-ink ratio

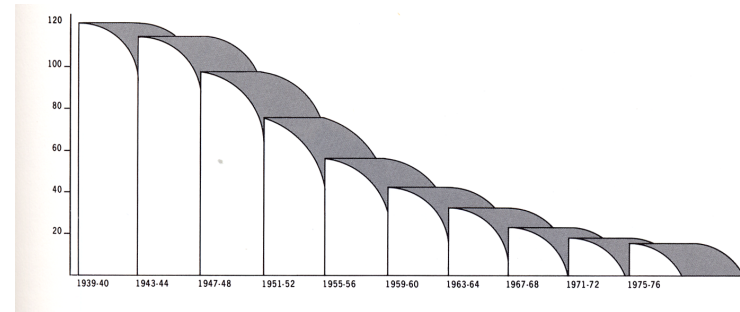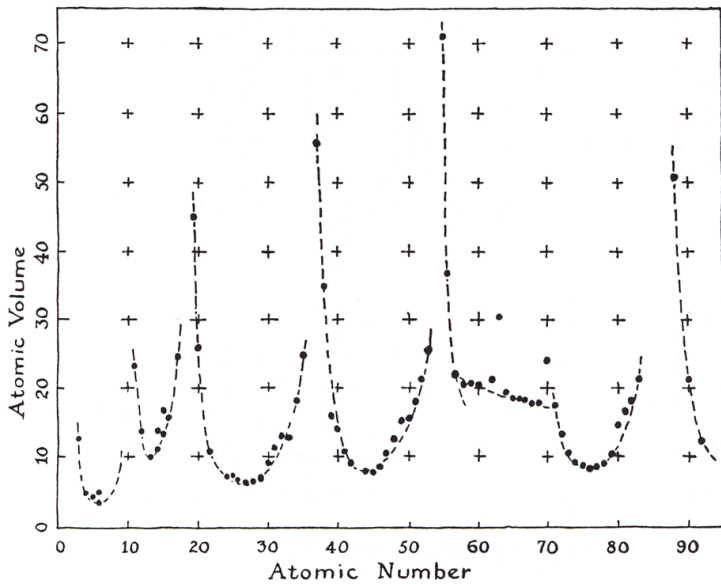## Zero data-ink ratio

Maximizing data-ink ratio
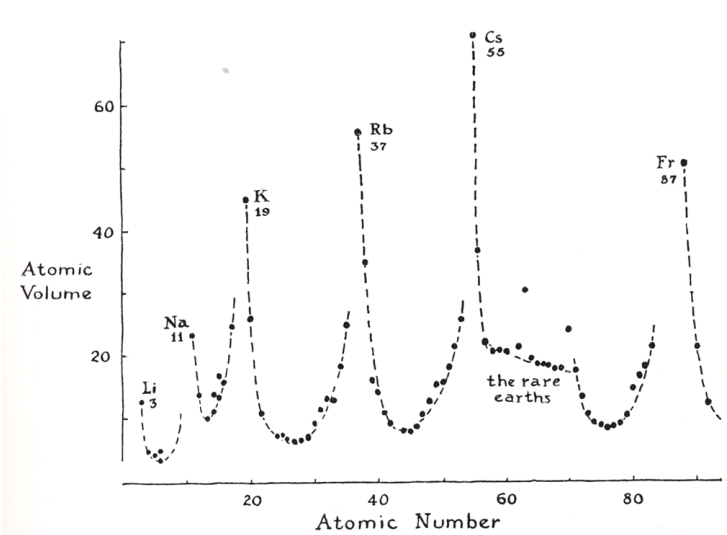
p125 top and p125 bottom of Tufte

Erasable non-data ink

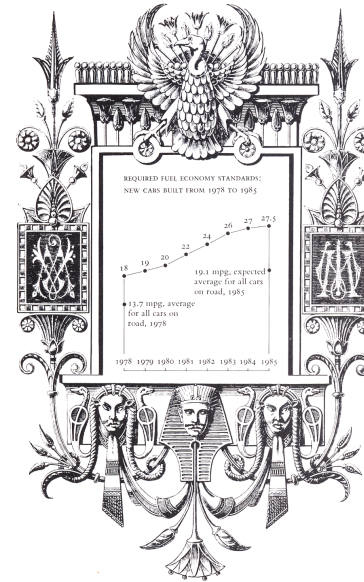# Erasable non-data ink

Improved non-data ink
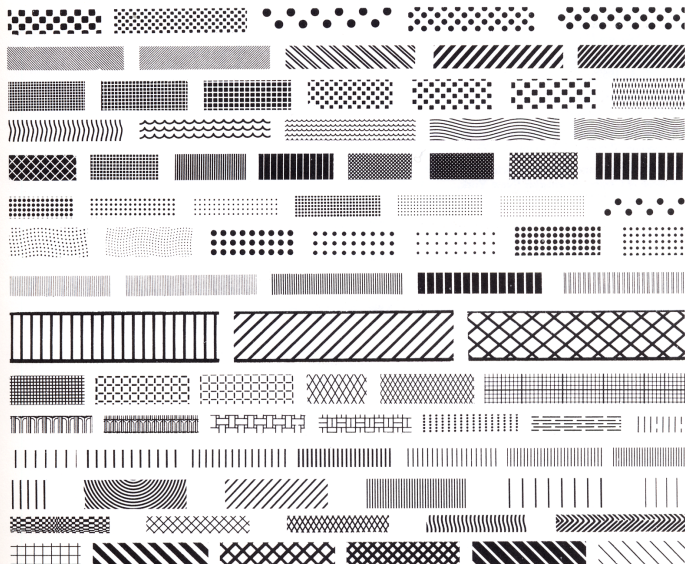
Show only the data

Non-data ink can be chartjunk.

Could be shading, hatching, grid, etc.
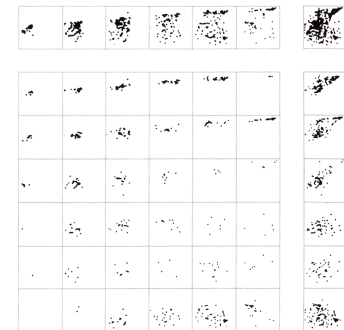
Get rid of it!

Content-free decoration

Moiré patterns
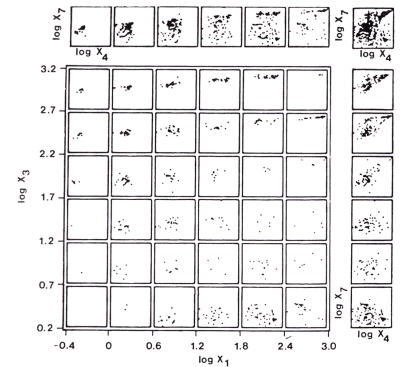
Show data, not frames

Almost no information!
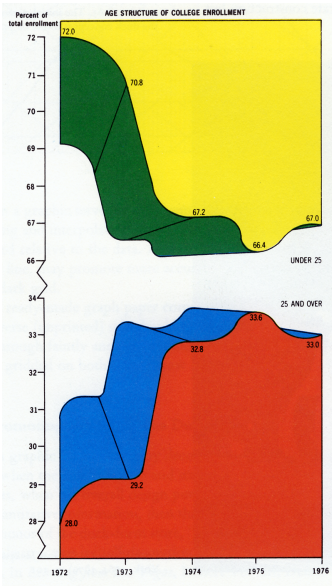
Don't lie with graphics

Lies, damned lies, and statistics

could also be

Lies, damned lies, and graphics.

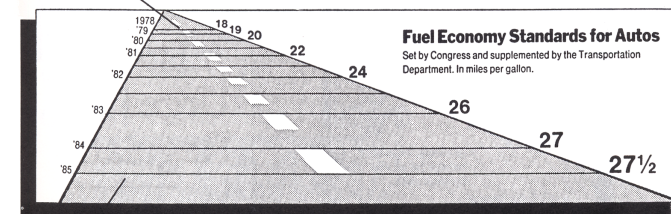What can we do to avoid misleading?

Data, area and dimension

The size of the representation of a number should be proportional to the number

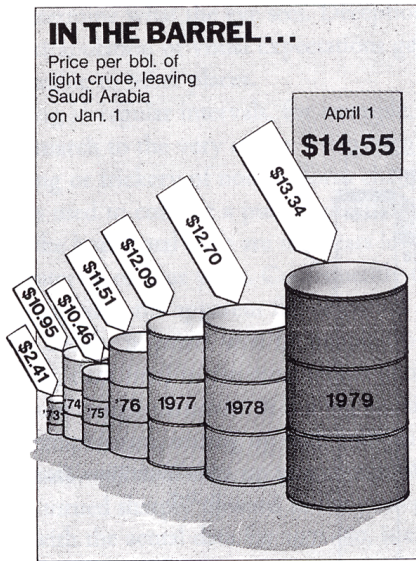The number of information carrying dimensions should not exceed the dimension of the data.

Backward in time?

Perspective

Change in axis meaning
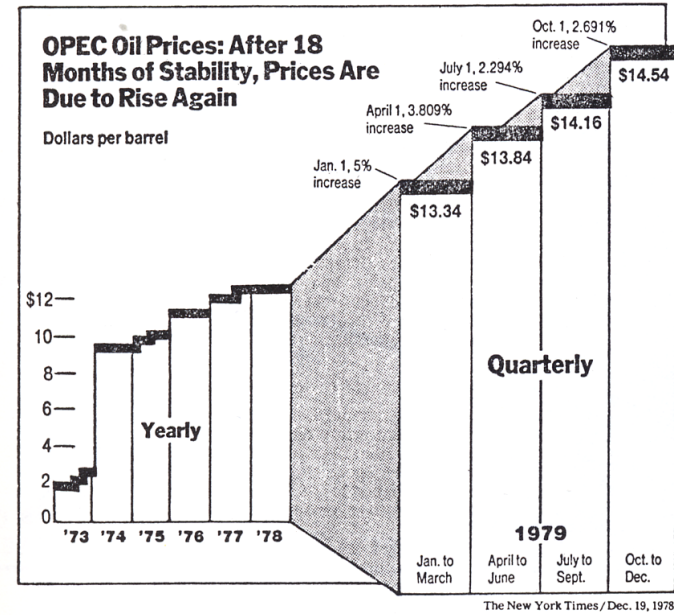
Paper usage, New York Times, Feb. 10, 2008
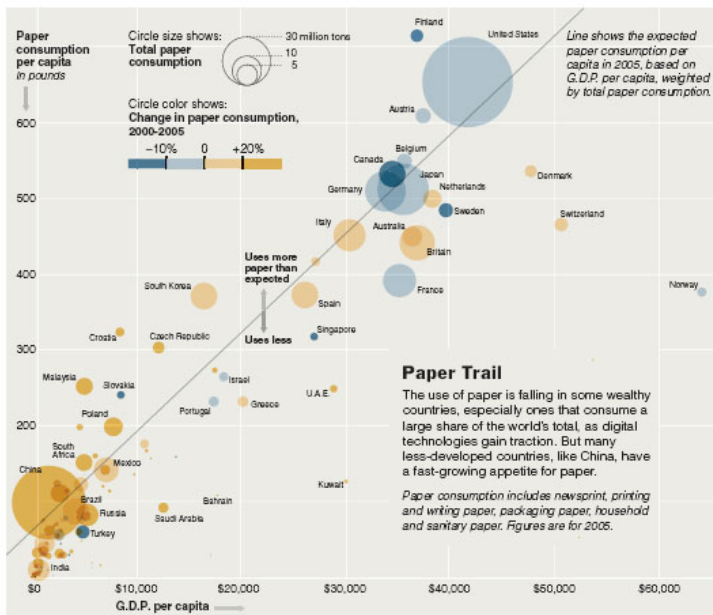
How to Display Data Badly (Wainer)

1. Show as few data as possible.
2. Hide what data you do show.
3. Ignore the visual metaphor.
4. Only order matters.
5. Graph data out of context.
6. Change scales in mid-axis.
7. Emphasize the trivial, not the important.
8. Jiggle the baseline.
9. Austria first.
10. Label illegibly, incompletely, inaccurately, and ambiguously.

Don't. . .

1. . . . Mislead
2. . . . Use mysterious abbreviations
3. . . . Include too much clutter (forest for the trees)
4. . . . Misuse placement of origin
5. . . . Include graphs without explanation
6. . . . Use gratuitous color/line variation
7. . . . SHOUT (use all capital letters)
8. . . . use chart junk
9. . . . use pie charts

Do. . .

1. . . . use accessible friendly graphic
2. . . . include axis labels, titles and legends
3. . . . use sensible tick marks
4. . . . facilitate comparisons between graphs by using common scales.
5. . . . avoid unclear abbreviations.

Summary

Many, many ways to do things badly.

- Show the data.
- Do not distort.
- Cause no pain.