

Scholarship Skills

Andrew Black & David Maier

Text Elements

All material © 1996–2017

Andrew Black, Todd Leen,
David Maier, Tim
Sheard

Footnotes and Parenthetical Comments

- Key idea: Use them sparingly
- If the main text is unclear or incomplete without the footnote or comment, it should probably be rewritten.
- Don't use a footnote or parenthesis to restate something already stated in the main text.
- Regard footnotes as parenthetical remarks; that is, something the reader can skip the first time through without losing vital information.

A hash table (which is a frequent database access method) gives good performance on single-item lookups. Other common access methods are ...

Figures and Tables

- Use the same name format and font in text, caption and figure. (Core Rule: consistent lexical set.)

Buf1 Buf_1 *Buf₁*

- Pay attention is you prepare the figure with one tool, and the text with another

Captions

- Give them meaningful and informative captions

Table 3. Performance characteristics.

Table 3. Comparative performance of simulated annealing and Monte Carlo methods on join planning.

- Include a figure or table number.
- Captions should describe *all* frames in the figure

Figure 2: (a) Plot of the posterior mean for each point in the latent space. (b) Plot of the 12-dimensional data projected onto the 2-dimensional subspace that maximizes the negentropy. (c) Plot of 12-dimensional data projected onto 2-dimensional principal subspace.

Credits

- Include a credit for material from another source — copied or adapted.

Fig. 3 Univariate factor analysis of IQ test data. (Adapted from *The Mismeasure of Man*, Stephen J. Gould, W.W. Norton & Co., 1996. Used with permission.)

- Give the credit in the figure caption, not the text. (Puts credit line close to information.)
- Make sure that you do get permission before using such material!

Callouts

- “Call out **all** figures, tables, programs ...” explicitly in the main text. [Dupré, §62]

A *callout* is a citation like “in Figure 2 ...”

This tells the reader *when* to look at the figure or table

Figure 3(a) shows the results using a nonlinear GTM mapping under the conditions reported by Bishop et al.[4].

- Use single numbers (as above) for figures in articles. Decimal numbers such as “Figure 4.3(a) shows ...” are used in the fourth chapter of a book or thesis.
- The order of appearance of figures or tables should match the order in which they are called out in the text.

Callouts

- Capitalize the words “Figure” and “Table” when used with a number.

The results with this modified algorithm are shown in Figure 3.

- Don't use a “buried callout” the first time you refer to a figure or table.

After the modification, the results are improved (see Figure 3).

You can use this format for subsequent references to the figure or table.

Tables

- Number your tables, call them out in the text, *and* tell the reader what to pay attention to in the table.
- Give only the information that you're going to discuss. (Avoid clutter, be concise ...)

Re-set the table, rather than cutting from a source?

- Write a caption for each table.

Tables

- Put units in the column header.

Elapsed time / μ s

- Keep style consistent among tables.
- Use the minimal number of vertical and horizontal rule lines in the body of a table — they add visual clutter.

The “minimal number” is often zero vertical lines, and one horizontal below the header. Most tables look better with fewer lines; there is never a need for double lines.

Lines can help if tables have many rows. Try a light grey background on every other row instead.

- Align columns of numbers on the decimal point, and use leading zeros (e.g., 0.5).

Code

- Use a different font for typesetting a code block, and indent the code block. `\usepackage{listings}`

- Dupré (§ 94) suggests using a monospace font.

Monospace fonts use the same amount of horizontal space for each character, like a mechanical typewriter.

(Does anyone remember typewriters?)

- Black says: this is a bad idea. We don't use typewriters any more, and there is a reason

If you do choose a monospace font, don't use Courier.

Code

- Some folks claim that you should distinguish code blocks further by using a sans-serif font for them if your text is set in a serif font — or vice versa.

Serif Fonts — Times Roman, Garamond, Palatino

Sans Serif Fonts — Arial, Helvetica, Lucida

This means that you don't have to use quotes when mentioning an element from your code.

- Set large blocks of code as a numbered figure.
(Why?)

Method 2.2: *Initializing instances of LOCell*

```
1 initialize
2   super initialize.
3   self label: ".
4   self borderWidth: 2.
5   bounds := 0@0 corner: 16@16.
6   offColor := Color paleYellow.
7   onColor := Color paleBlue darker.
8   self useSquareCorners.
9   self turnOff
```

Note that the characters " on line 3 are two separate single quotes with nothing between them, not a double quote! " denotes the empty string.



Accept *this method definition*.

What does the above code do? We won't go into all of the details here (that's what the rest of the book is for!), but we will give you a quick preview. Let's take it line by line.

Graphs

- Keep graphs and charts uncluttered.

Work as hard to make figures clear as you do to make your text clear.

Include *only* the information that you're going to refer to in the text.

Core Rule: Be Concise, Be Simple

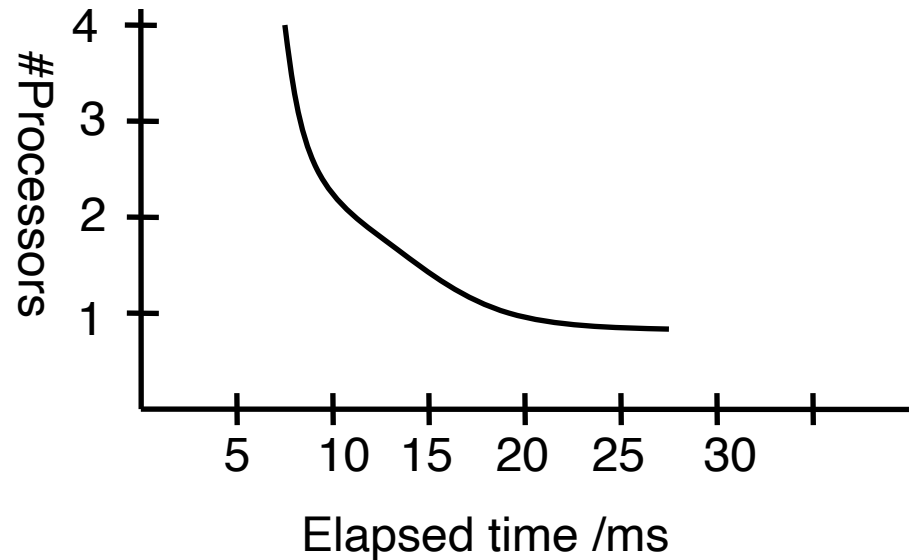
- Label the axes!

- Use REALLY BIG fonts for axis labels and tick mark numbers. Your figure will be reduced in size when typeset.

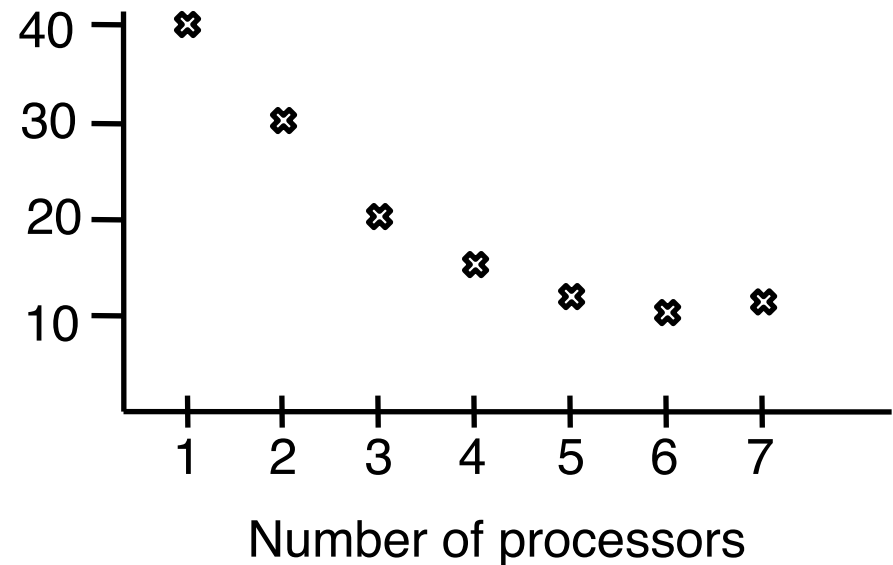
If you use MatLab to generate figures, use 16 point fonts or larger for the axes.

Graphs

- Independent axis generally goes along the bottom.

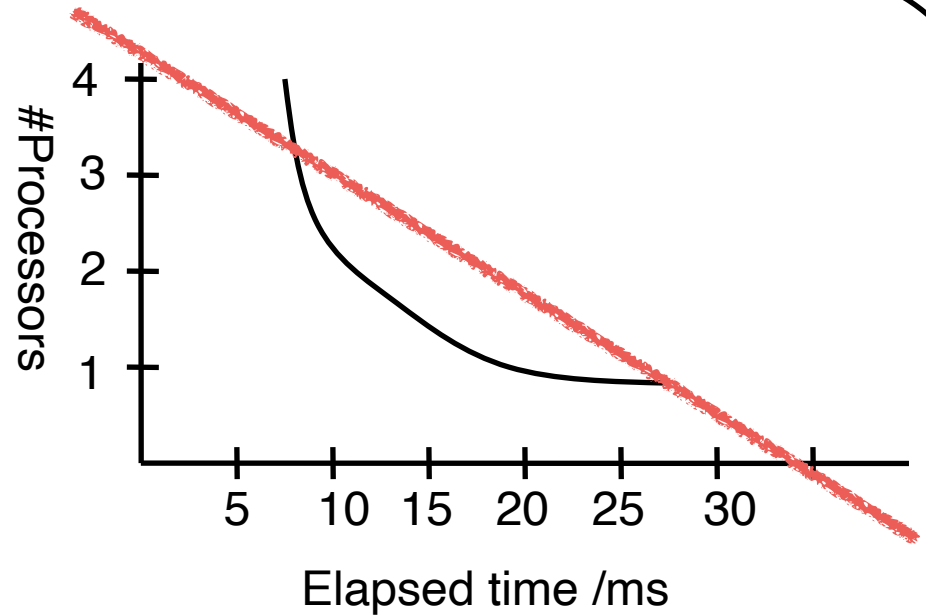


Time to completion /ms

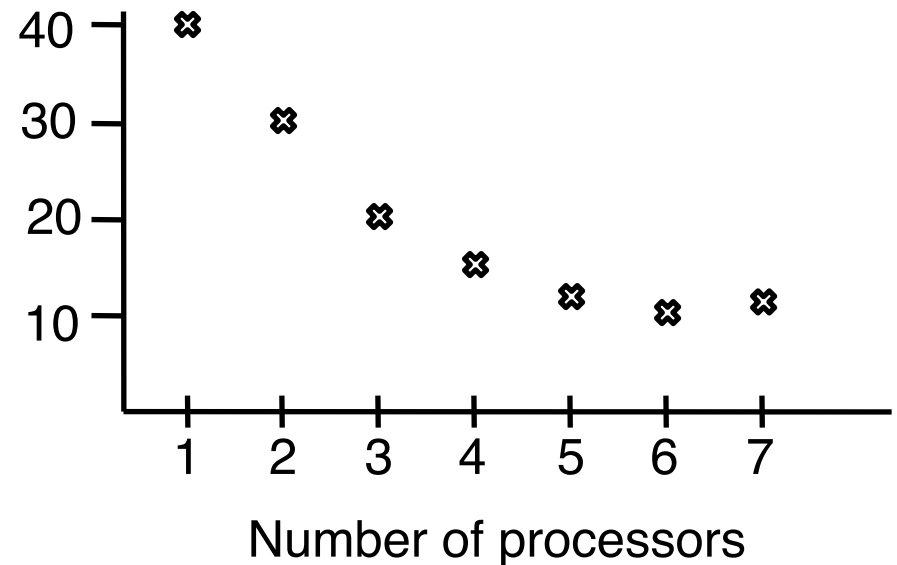


Graphs

- Independent axis generally goes along the bottom.

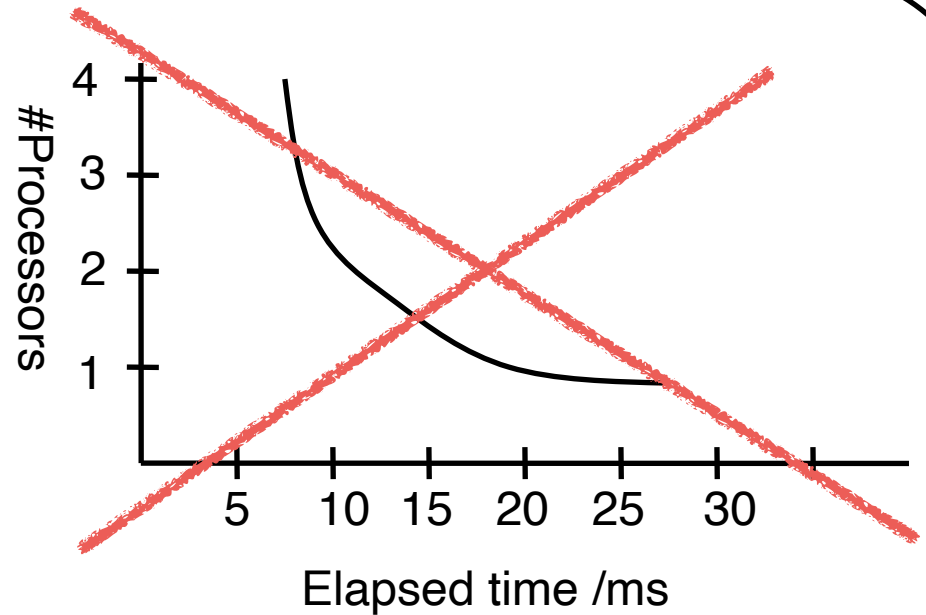


Time to completion /ms

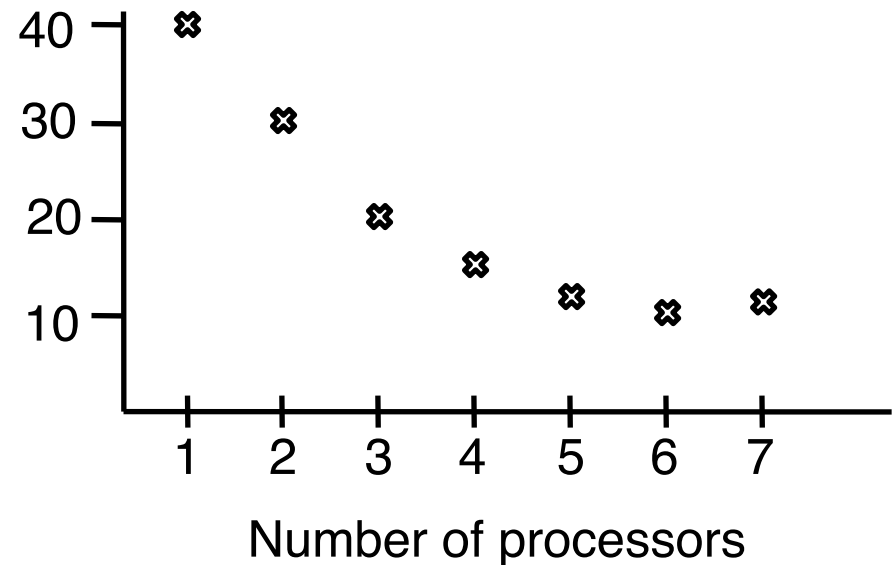


Graphs

- Independent axis generally goes along the bottom.



Time to completion /ms



Graphs

- Where several curves or classes of data points appear on the same graph, use different line types or marking symbols, and provide a legend.

Keep fonts in the legend large so they're legible.

If possible, label the lines rather than providing a legend

- Look at Edward Tufte's beautiful book *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

Graphs

- Where several curves or classes of data points appear on the same graph, use different line types or marking symbols, and provide a legend.

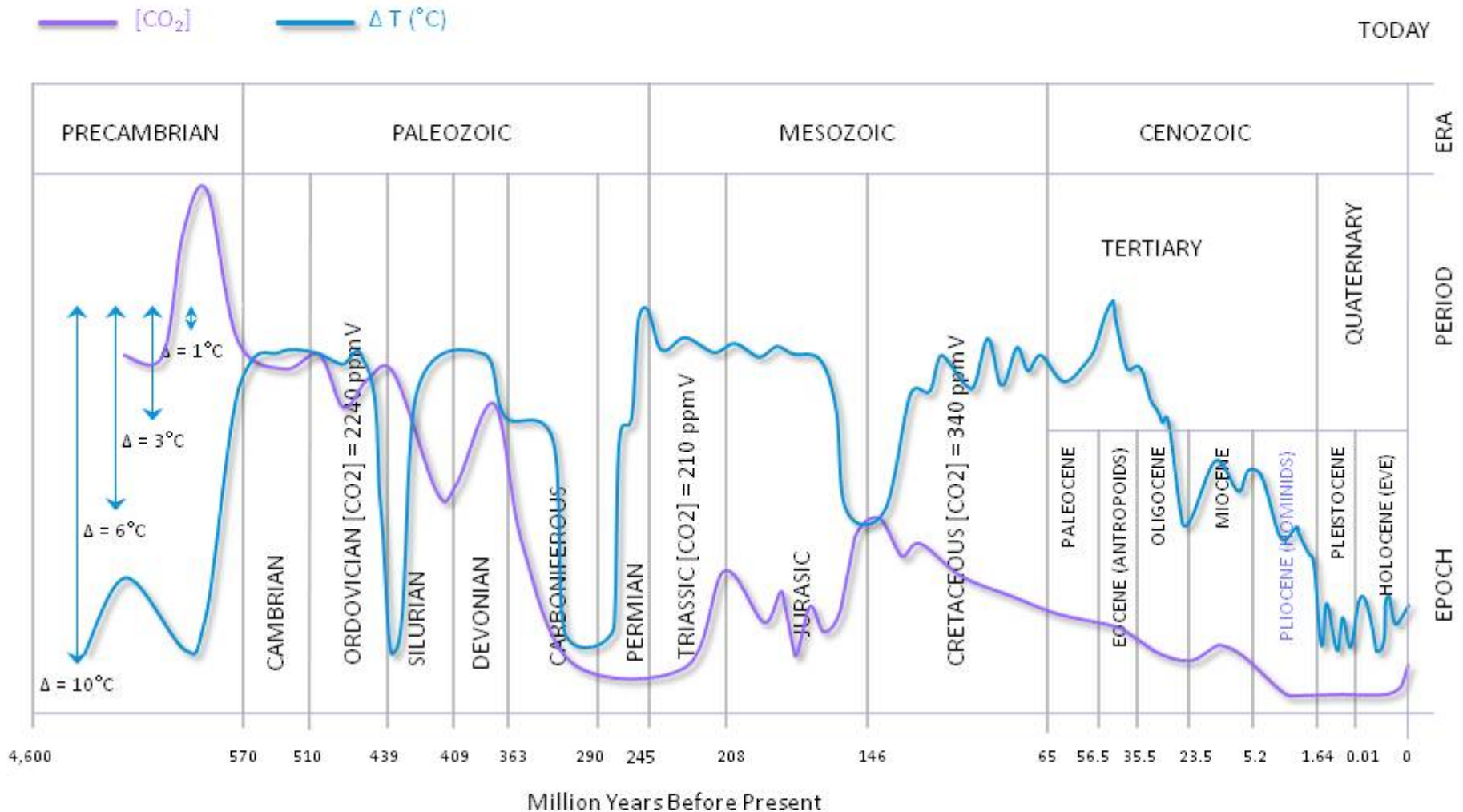
Keep fonts in the legend large so they're legible.

If possible, label the lines rather than providing a legend

Why?

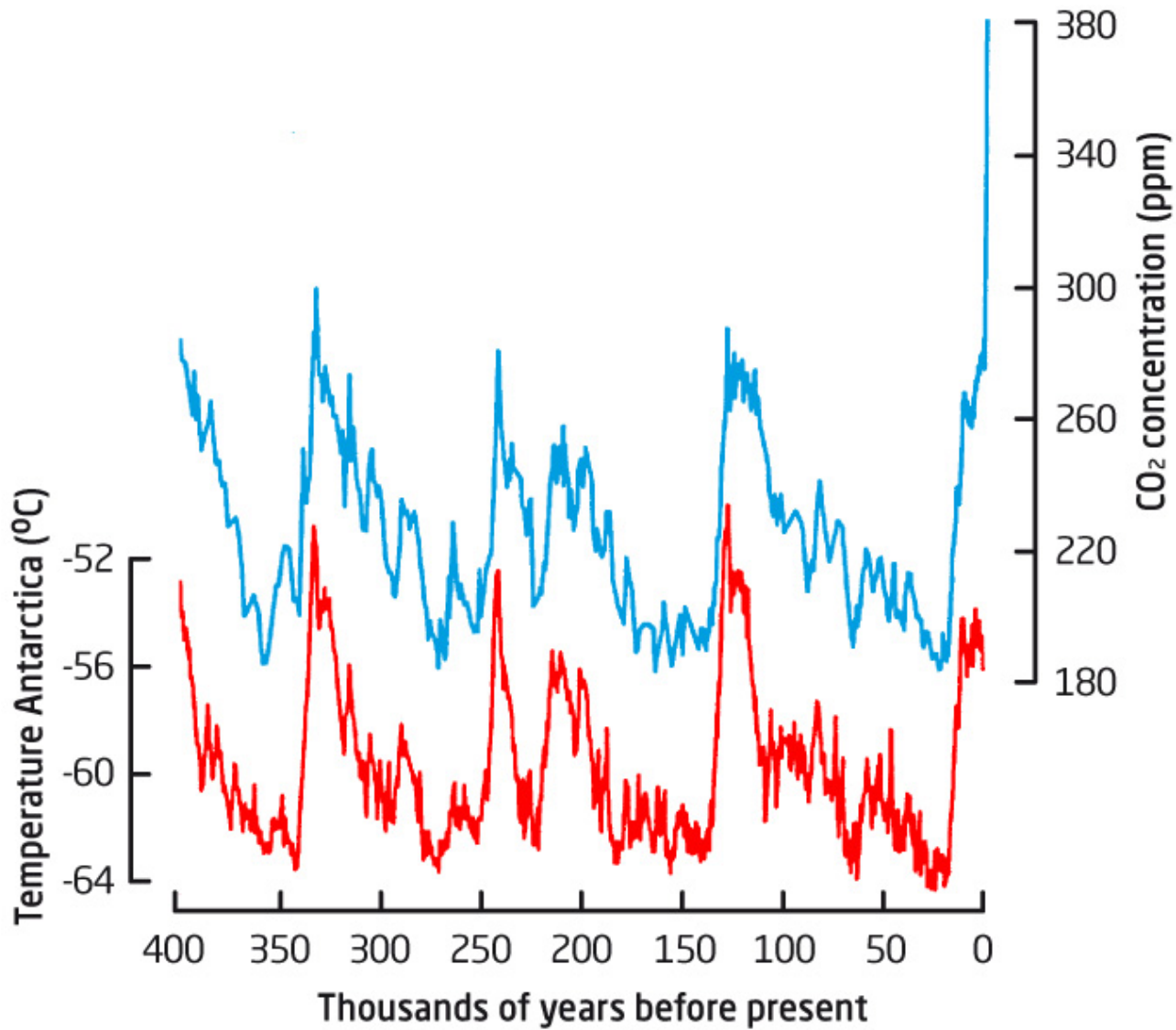
- Look at Edward Tufte's beautiful book *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

Geological Timescale: Concentration of CO₂ and Temperature fluctuations



1- Analysis of the Temperature Oscillations in Geological Eras by Dr. C. R. Scotese © 2002. 2- Ruddiman, W. F. 2001. *Earth's Climate: past and future*. W. H. Freeman & Sons. New York, NY. 3- Mark Pagani et al. *Marked Decline in Atmospheric Carbon Dioxide Concentrations During the Paleocene*. *Science*; Vol. 309, No. 5734; pp. 600-603. 22 July 2005. *Conclusion and Interpretation* by Nasif Nahle ©2005, 2007. *Corrected on 07 July 2008 (CO₂: Ordovician Period)*.

Better:

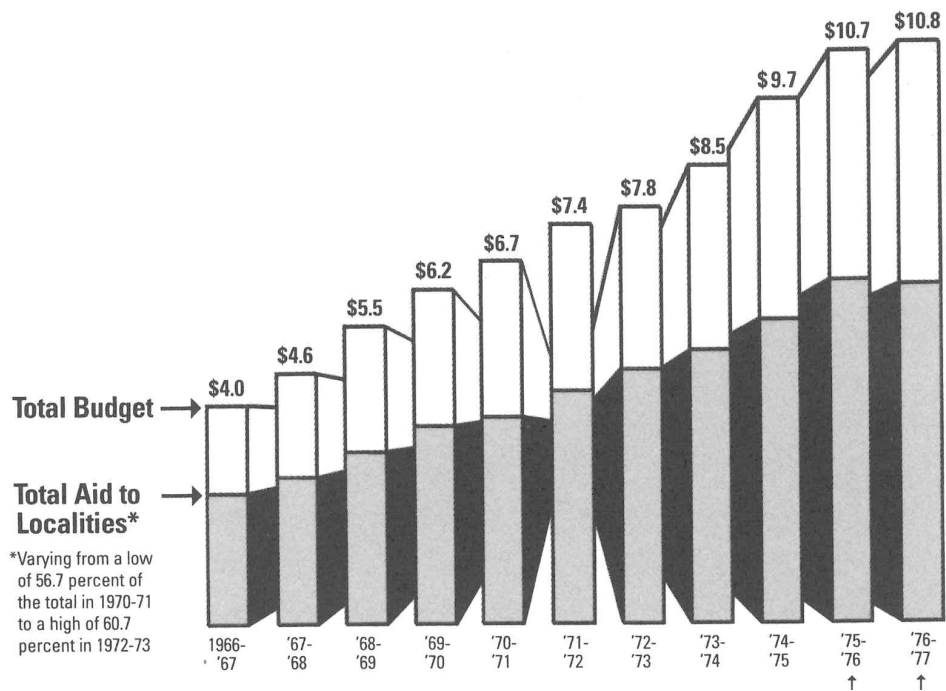


Deconstructing a Budget Chart

New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976



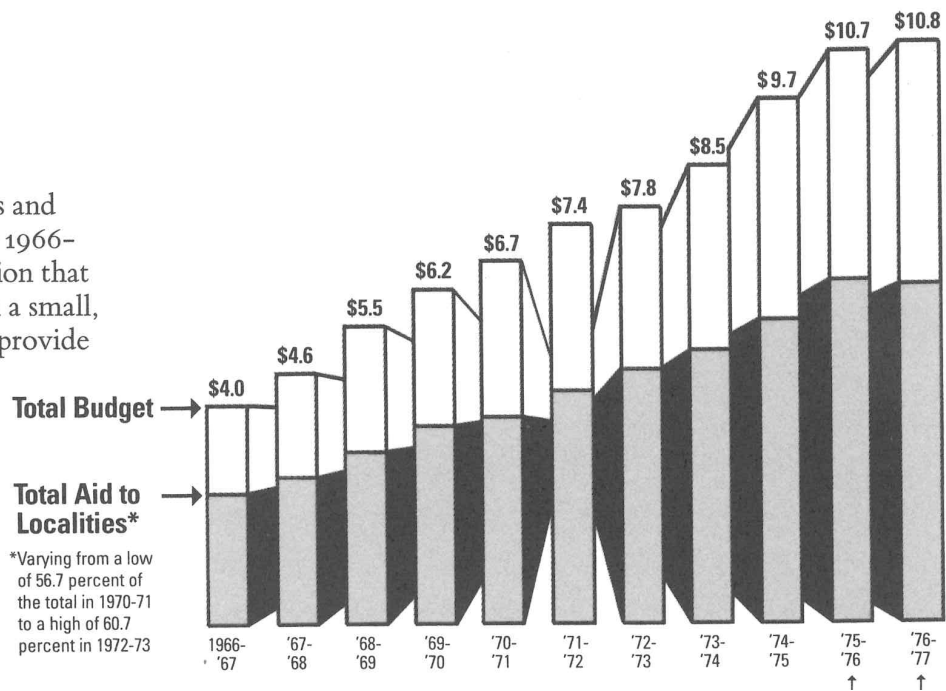
Deconstructing a Budget Chart

New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

This cluster of type emphasizes and stretches out the low value for 1966-1967, encouraging the impression that recent years have shot up from a small, stable base. Horizontal arrows provide similar emphasis.



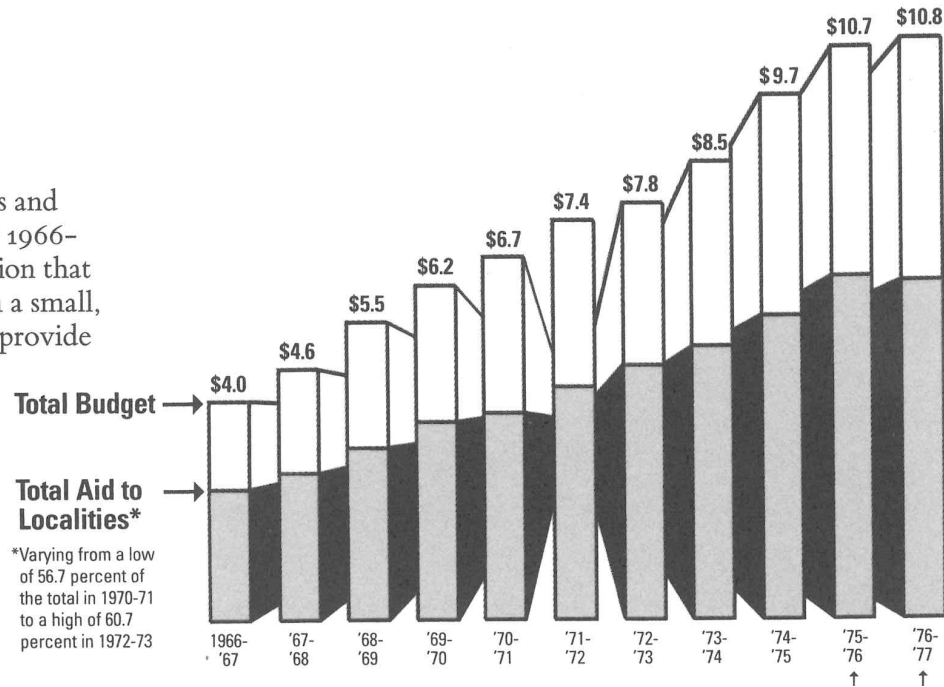
Deconstructing a Budget Chart

New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

This cluster of type emphasizes and stretches out the low value for 1966-1967, encouraging the impression that recent years have shot up from a small, stable base. Horizontal arrows provide similar emphasis.



This squeezed-down block of type contributes to an image of small, squeezed-down budgets back in the good old days.

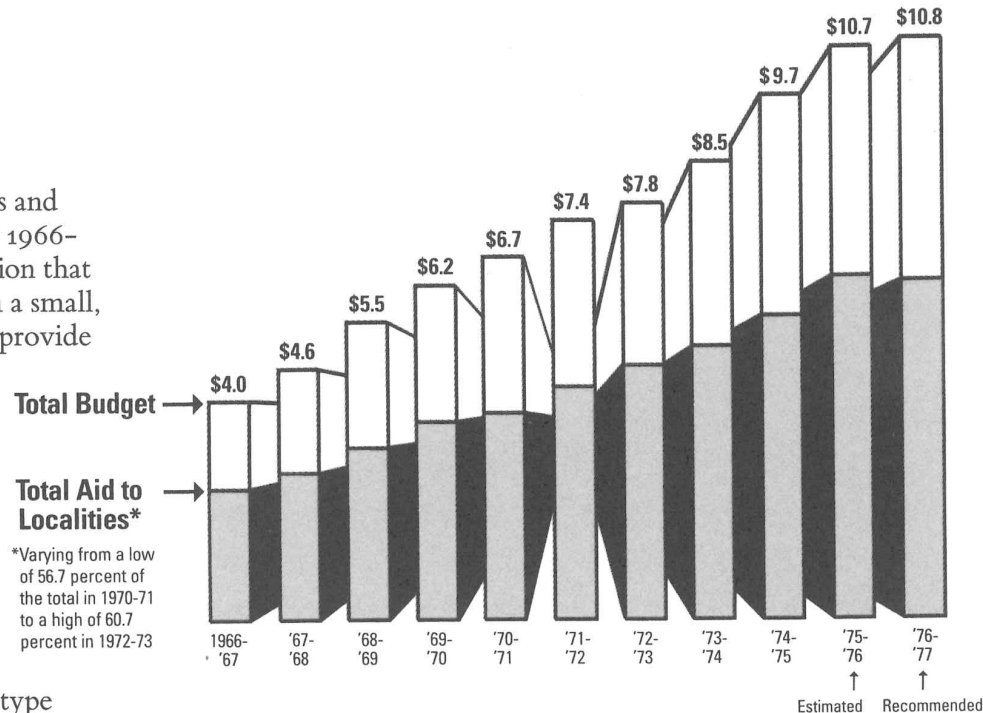
Deconstructing a Budget Chart

New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

This cluster of type emphasizes and stretches out the low value for 1966-1967, encouraging the impression that recent years have shot up from a small, stable base. Horizontal arrows provide similar emphasis.



This squeezed-down block of type contributes to an image of small, squeezed-down budgets back in the good old days.

Arrows pointing straight up emphasize recent growth. Compare with horizontal arrows at left.

Deconstructing a Budget Chart

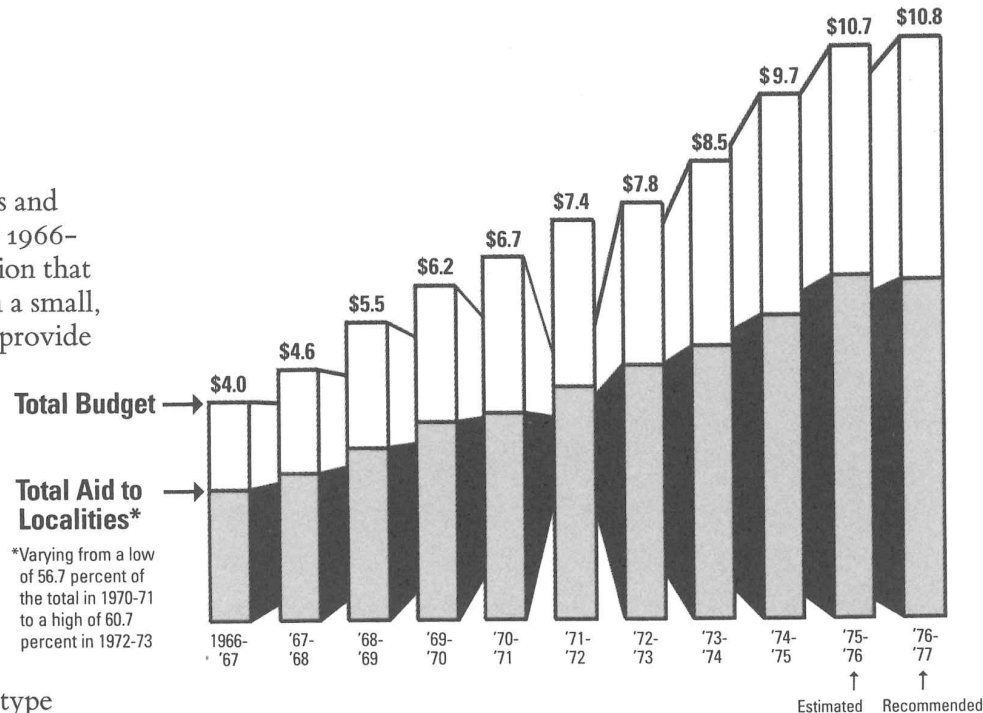
New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

These three parallelepipeds have been placed on an optical plane *in front* of the other eight, creating the image that the newer budgets tower over the older ones.

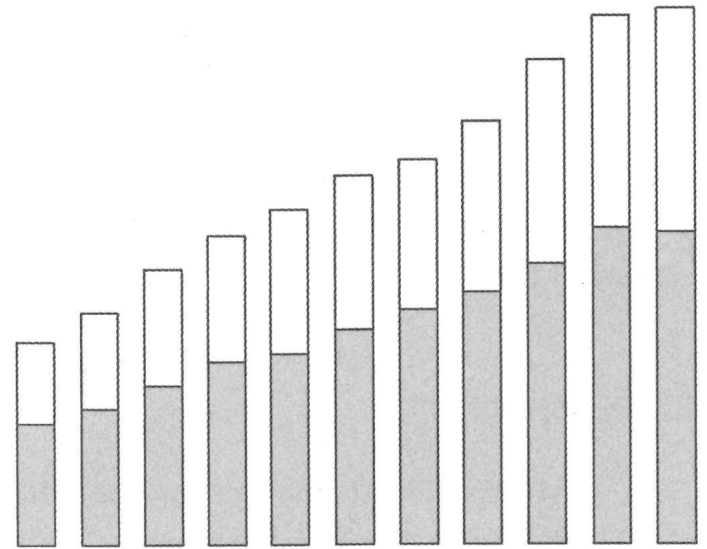
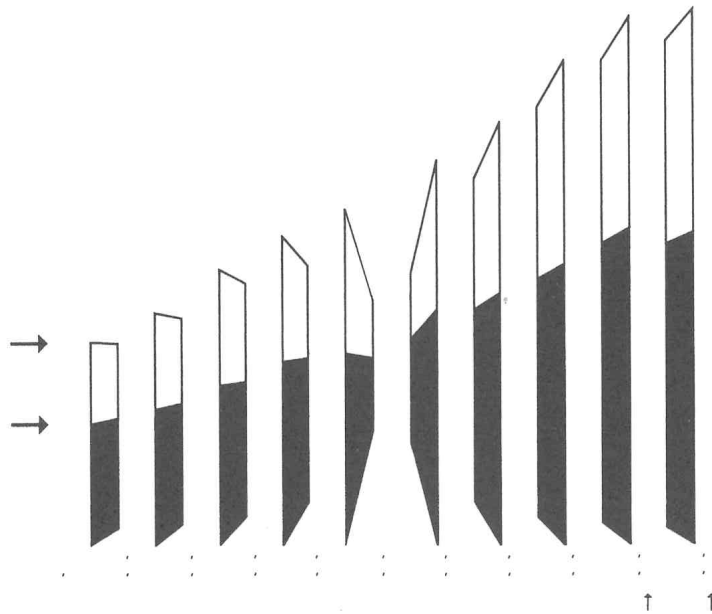
This cluster of type emphasizes and stretches out the low value for 1966-1967, encouraging the impression that recent years have shot up from a small, stable base. Horizontal arrows provide similar emphasis.



This squeezed-down block of type contributes to an image of small, squeezed-down budgets back in the good old days.

Arrows pointing straight up emphasize recent growth. Compare with horizontal arrows at left.

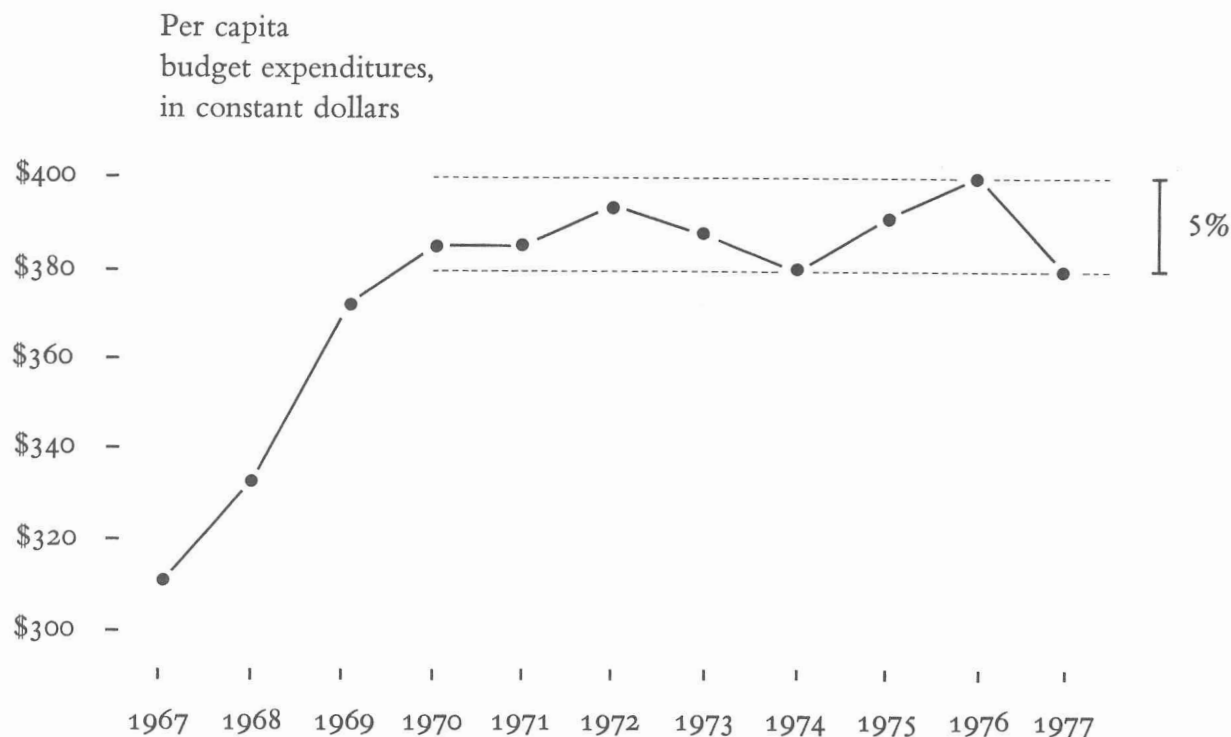
Leaving behind the distortion in the chartjunk heap at the left yields a calmer view:



■ But: chart ignores

change in population and
change in the value of money

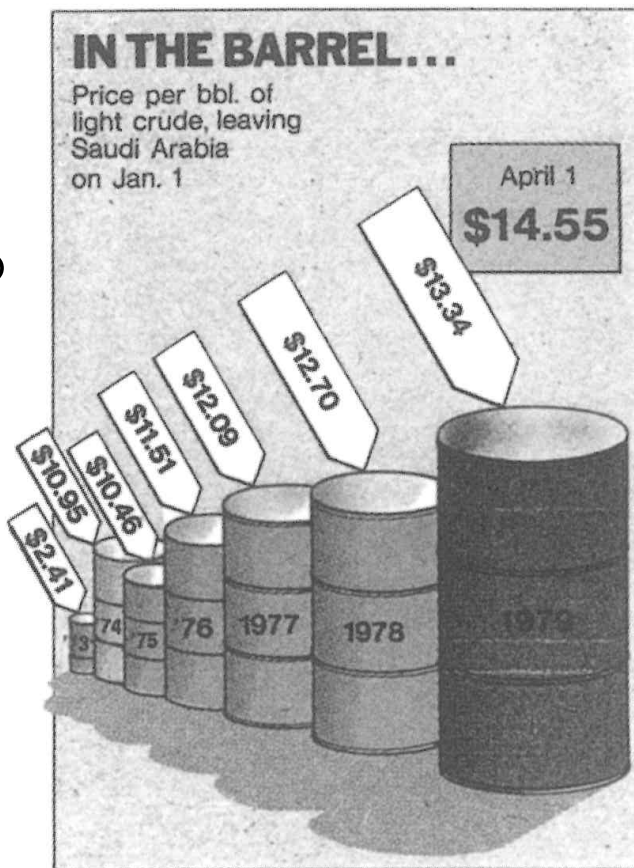
Application of arithmetic makes it possible to take population and inflation into account. Computing expenditures in *constant (real) dollars per capita* reveals a quite different—and far more accurate—picture:



Tufte's "Lie Factor"

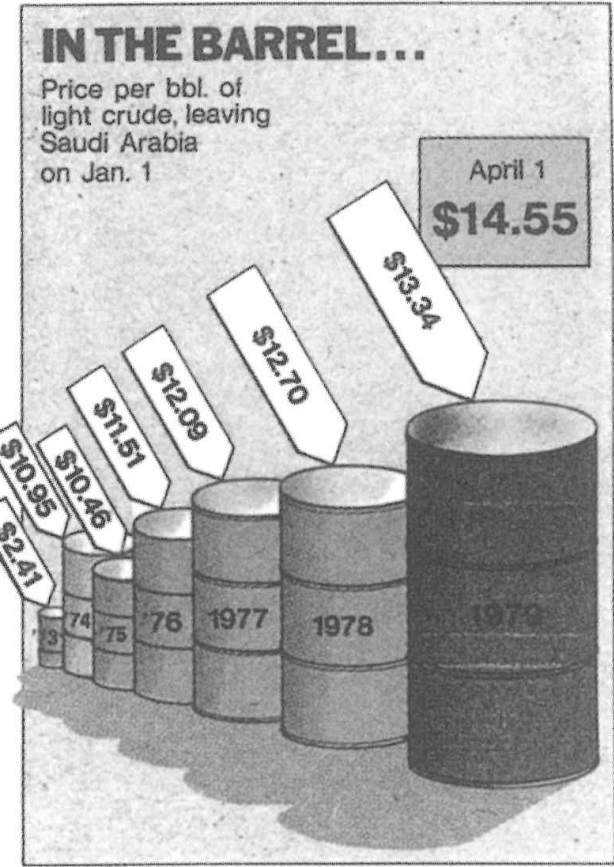
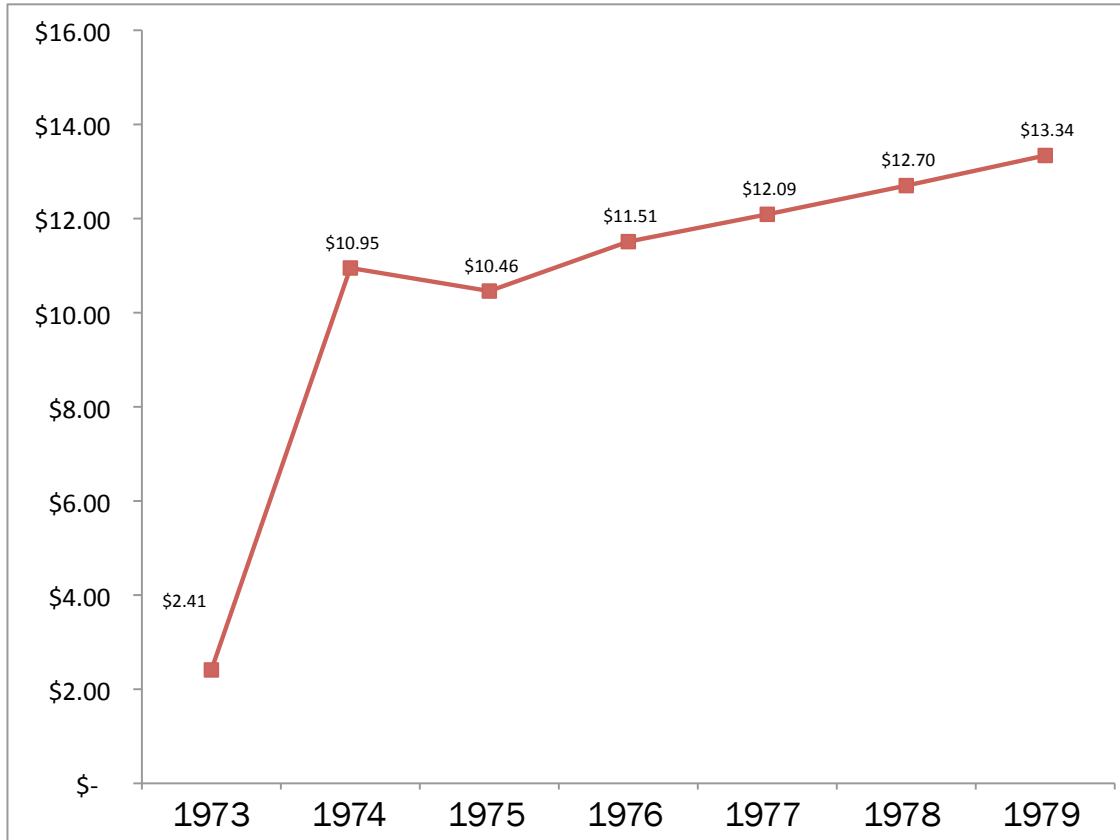
■ Ratio of actual values to depicted values

By surface area, the Lie Factor is 9.4. But, if one takes the barrel metaphor seriously and assumes that the volume of the barrels represents the price change, then the volume from 1973 to 1979 increases 27,000 percent, compared to a data increase of 454 percent, for a Lie Factor of 59.4, which is a record.



Tufte's "Lie Factor"

- Ratio of actual values to depicted values



Lists

- Keep lists and references to them in the same order

We cover three basic tree-traversal methods: in-order, pre-order, and post-order. With pre-order, the parent node is visited ...

- Consider setting off lists typographically
- Keep lists “pure” — readers assume that elements will be of like kind (Parallel construction)

Stony Brook Travel

- Cruises
- Airline Tickets
- Trains
- Golf Buffs

- Watch for list *splices*

With a cache size of M , P_1, P_2, \dots, P_n are accessed one-by-one.

- Don't use "etc." when the reader is unlikely to be able to fill in the rest of the list.

The stable marriage problem has been explored by Floyd [21], Sedgewick [34], etc.

- Watch for list *splices*

With a cache size of M , P_1, P_2, \dots, P_n are accessed one-by-one.

- Don't use "etc." when the reader is unlikely to be able to fill in the rest of the list.

The stable marriage problem has been explored by Floyd [21], Sedgewick [34], etc.

- Watch for list *splices*

With a cache size of M , P_1, P_2, \dots, P_n are accessed one-by-one.

- Don't use "etc." when the reader is unlikely to be able to fill in the rest of the list.

The stable marriage problem has been explored by Floyd [21], Sedgewick [34], ~~etc.~~

- Watch for list *splices*

With a cache size of M , P_1, P_2, \dots, P_n are accessed one-by-one.

- Don't use "etc." when the reader is unlikely to be able to fill in the rest of the list.

The stable marriage problem has been explored by Floyd [21], Sedgewick [34], ~~etc.~~

and many others

Intext Lists

- With intext lists — lists set in running text, rather than displayed — the collection of elements taken together forms a sentence. [Dupré, § 26]

Individual entries are *not* sentences. Entries should be separated by commas or semicolons.

We discuss three types of model: (1) parametric, (2) non-parametric, and (3) semi-parametric.

We discuss several types of model: parametric, non-parametric, and semi-parametric.

Please do the following: rewrite the exercise, paying close attention to the core rules for writing; read the assigned sections in the text; and complete the new exercise.

- Number the list elements only if necessary

When the number is part of the discussion

Trees can be traversed in three ways: (1) pre-order, (2) in-order, and (3) post-order.

or you will refer to the numbers later.

- Use a colon before an intext list *only* if it's logical and grammatically correct.

Programming languages contain: data structures, control structures, and abstraction mechanisms.

Programming languages collect features as they age: no one ever removes a data structure, a control structure or an abstraction mechanism from a deployed language.

- Number the list elements only if necessary

When the number is part of the discussion

Trees can be traversed in three ways: (1) pre-order, (2) in-order, and (3) post-order.

or you will refer to the numbers later.

- Use a colon before an intext list *only* if it's logical and grammatically correct.

Programming languages contain/ data structures, control structures, and abstraction mechanisms.

Programming languages collect features as they age: no one ever removes a data structure, a control structure or an abstraction mechanism from a deployed language.

- If at least one entry needs to be a complete sentence (with a period), then turn the list into running text.

Our marketing strategy will include several actions. First, we will organize a kickoff meeting. Second, we will write to all the magazines in the field, and ask them to review our product. Third, we will rent booths at the three major conferences. Finally, we will purchase two mailing lists, and send out brochures.

Displayed Lists

- You can apply the same rules to displayed lists

We tried several alternatives:

1. Unimodal density estimates,
2. Mixture model density estimates, and
3. Parzen window density estimates.

- Dupré [§26] says that displayed lists can be treated as if they were not text at all:

We might list the reasons for living in a city:

Convenient shopping

Great theaters

Good restaurants

Exciting nightlife

- This works for lists set as tables, but I (Black) don't like it when the list is part of the text