DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

CSC318S

THE DESIGN OF INTERACTIVE COMPUTATIONAL MEDIA

Lecture 19 — 30 March 1998

PRINCIPLES OF DATA DISPLAY AND VISUALIZATION

Ronald Baecker
Professor of Computer Science,
Electrical and Computer Engineering, and Management
University of Toronto

19.1 Nature, purpose of data display and visualization

What is computer science?
- The study of algorithms
- The study of data management and data processing
- The study of systems design and software engineering

We manage and process data by computers because:
- Data sets are large
- Data sets change over time
- Data sets embody complex interrelationships

But how do we comprehend the data?
- How do we know what is there and what is not?
- How do we know if there are errors in the data?
- How do we know what relationships exist in the data?
- How do we know what “the data means”?

Usually, we don't, because we can't see it and we can't visualize/perceive/understand it

The purpose of data display and visualization techniques is to help us see the data and understand the data (Fig. 19.1)

“We thrive in information-thick worlds because of our marvelous and everyday capacities to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, refine, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, sort, pick over, group, pigeonhole, integrate, blend, average, filter, lump, skip, smooth, chunk, inspect, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, list, glean, synopsize, winnow wheat from chaff, and separate the sheep from the goats.” (Tufte, 1990, p. 50)

Computer graphics has introduced a revolution in data display, but it is easier to produce visual garbage than it is to produce elegant visual representations that convey meaning

The goal of this lecture is to convey insight into the key issues

Fig. 19.1. Data sets displayed as numbers & as graphs (Tufte, 1983, pp. 13,14)

19.2 Excellence in data display

Tufte (1983, p. 13) suggests that graphical displays should:

"• Show the data
• Induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
• Avoid distorting what the data have to say
• Present many numbers in a small space
• Make large data sets coherent
• Encourage the eye to compare different pieces of data
• Reveal the data at several levels of detail, from a broad overview to the fine structure
• Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
• Be closely integrated with the statistical and verbal descriptions of a data set"

He goes on (1983, p. 51):

"Graphical excellence is the well-designed presentation of interesting data — a matter of *substance*, of *statistics*, and of *design*.
... consists of complex idea communication with clarity, precision, and efficiency.
... is that which gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
... is nearly always multivariate.
... requires telling the truth about the data."

And finally (1983, p. 191):

"What is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple, rather the task of the designer is to give visual access to the subtle and the difficult — that is, *the revelation of the complex*."

19.3 Techniques for data display

Tables (Fig. 19.2)
  Organizing and displaying lists, series of related numbers, categories, concepts, etc.
  Example: Comparing techniques for data display

Technique	**Major use**
Tables	Organized description of relationships among discrete elements
Graphs and charts	Pictures of relationships among quantitative data items
Maps	Displays of geographical or spatially-organized data
Diagrams	Portrayals of interrelationships among abstractions

Charts and graphs (e.g., relational graphics, time series)
  Distributions of one quantitative variable (Fig. 19.3)
  One quantitative variable with labels (Fig. 19.4)
  Two quantitative variables (Fig. 19.5)
  Two quantitative variables: Time series (Fig. 19.6, 19.7)
  Two quan. var.'s with labels (categories) (Slides 19.1, Cleveland, 1985, p. 207; 19.2, *ibid.*, opp. p. 213)

Maps (e.g., data maps)
  Data maps of discretized data (Fig. 19.8)
  Data maps of topographic, continuous data (Fig. 19.9)

Small multiples and table-graphics
  A small multiple (Fig. 19.10)
  A table graphic (Fig. 19.11)
  A small multiple of table graphics (Fig. 19.12)

Diagrams, e.g.,
  Block diagram, organization (tree) chart, pert chart

Fig. 19.2 A table of data (Tufte, 1990, p. 29)

Fig. 19.3 Graphing distributions of one quantitative variable: Point graph, histograms, percentile graphs (Cleveland, 1985, pp. 124, 126, 128)

Fig. 19.4 Graphing distributions of one quantitative variable with labels: Dot chart (Cleveland, 1985, p. 145)

Fig. 19.5 Graphing two quantitative variables (Tufte, 1983, p. 47)

Fig. 19.6 Graphing two quantitative variables with time series:
Symbol graph, connected symbol graph, connected graph, vertical line graph
(Cleveland, 1985, pp. 181, 180, 182, 183)

Fig. 19.7 Graphing two quantitative variables:
Time series (Tufte, 1983, pp. 30, 37)

Fig. 19.8 Data maps of discretized data (Tufte, 1983, p. 17)

Fig. 19.9 Data maps of topographic, continuous data
(Kerlow and Rosebush, 1986, pp. 244, 246, 246)

Fig. 19.10 A small multiple (Tufte, 1983, p. 42)

Fig. 19.11 A table graphic (Tufte, 1983, p. 158)

Fig. 19.12 A small multiple of table graphics (Tufte, 1983, p. 204)

19.4 Graphical integrity

Lying with data graphics
- The lie factor (Fig. 19.13)
- Confounding the lie factor: what is being compared?
  - Length, area, volume (Fig. 19.14)
- The importance of context (Fig. 19.15)

Unintentional lies due to phenomena of graphical perception
- An example (Fig. 19.16)

19.5 Pitfalls, flaws, and opportunities in data display

Data ink and the data ink ratio (Fig. 19.17, 19.18)

Chartjunk
- Vibrations (Fig. 19.19)
- Grids (Fig. 19.20)
- Ducks (Slides 19.3, Tufte, 1983, p. 118; 19.4, Tufte, 1990, p. 34)

The role of colour
- Data map with colour (Slide 19.5, *ibid.*, p. 40)
- Small multiples diagram with colour (Slide 19.6, *ibid.*, p. 63)
- Geometry proof with colour (Slide 19.7, *ibid.*, p. 85)
- Interfaces with colour

19.6 Applications of computer-aided visualization

Banking and finance
Industrial and manufacturing
Resources and exploration
Printing and publishing
Medical
Educational

Fig. 19.13 A lie factor of 14.8, a truthful graphic, and a needlessly decorated version (Tufte, 1983, pp. 57, 58, 59)

Fig. 19.14 Confounding the lie factor: length vs. area (Tufte, 1983, p. 70)

Fig. 19.15 The importance of context (Tufte, 1983, pp. 74, 74, 75, 75)

Fig. 19.16 Visual perception problem (Cleveland, 1985, p. 277)

Fig. 19.17 High and low data ink ratios (Tufte, 1983, p. 94)

Fig. 19.18 Data ink maximization (Tufte, 1983, p. 125)

Fig. 19.19 Chartjunk: Vibrations (Tufte, 1983, p. 108)

Fig. 19.20 Chartjunk: Grids (Tufte, 1983, p. 114)

Slides19.1, 19.2 Use of colour in a scatterplot (Cleveland, 1985, p. 207, 213)

Slide19.3 A preposterous use of colour (Tufte, 1983, p. 118)

Slide19.4 Another silly use of colour (Tufte, 1990, p. 34)

Slide19.5 Data map with colour (Tufte, 1990, p. 40)

Slide19.6 Small multiples diagram with colour(Tufte, 1990, p. 63)

Slide19.7 Geometry proof with colour(Tufte, 1990, p. 85)