Tufte

Andrej Ferko

Comenius University, Bratislava, ferko@fmph.uniba.sk

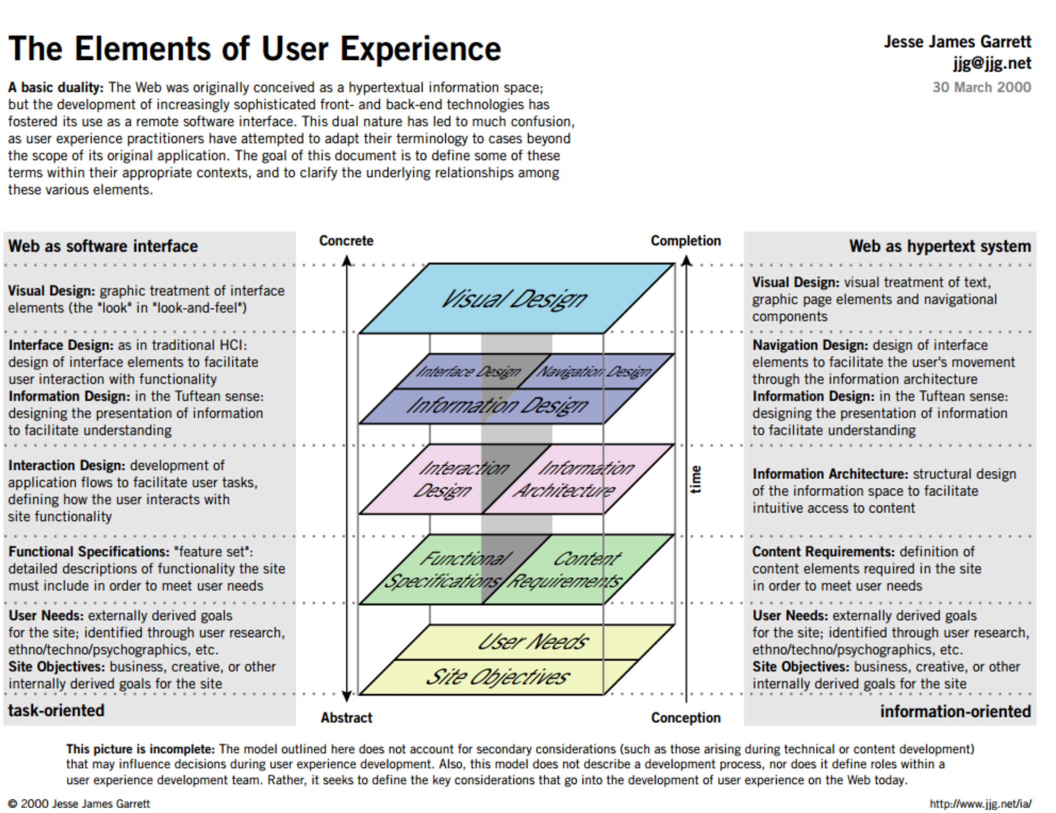
www.sccg.sk/wega

Communication / Semiotics

- Original <---> Recipient
- Original … Two recipients
- Incomplete original
- Two parts of original, two recipients
- No original => Model, representation
- No model => Darstellung, Ostension
- Knowledge direct or indirect
- Metacommunication, semiotics

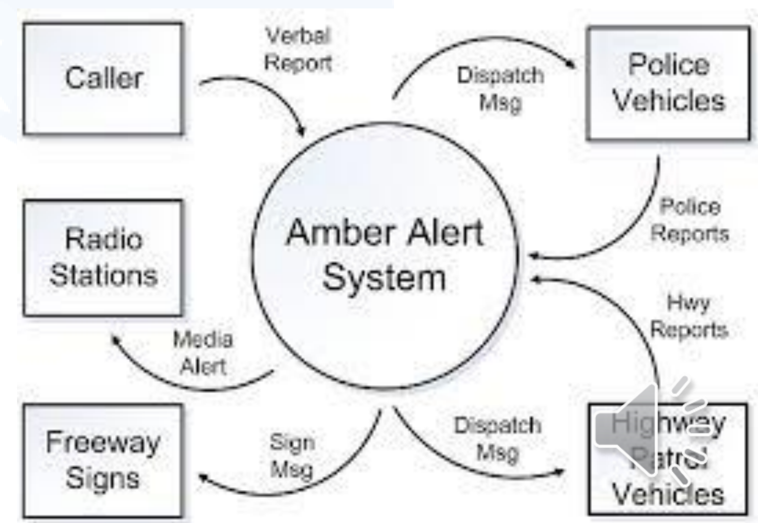
Jesse James GARRETT

- Prehľad odporúčanej metodiky rozmýšľania o špecifikácii Vášho diela, 5 rovín, 10 častí, **19 pojmov + logo**
- *http://www.jjg.net/elements/pdf/elements.pdf*



- **Kontextový diagram**

Zber požiadaviek, Wiegers



Key Source for Today

- **Graphics Design Based on Edward Tufte's Principles**

- https://chnm.gmu.edu/digitalhistory/links/pdf/chapter4/4.13b%20and%204.21.pdf
- https://chnm.gmu.edu/digitalhistory/links/cached/chapter4/4_13b_tufte.htm

- **Easy access -> WEGA page, AF page, Chapter 4 in Cohen et al. book**

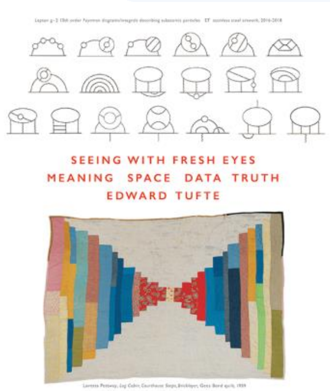
Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web

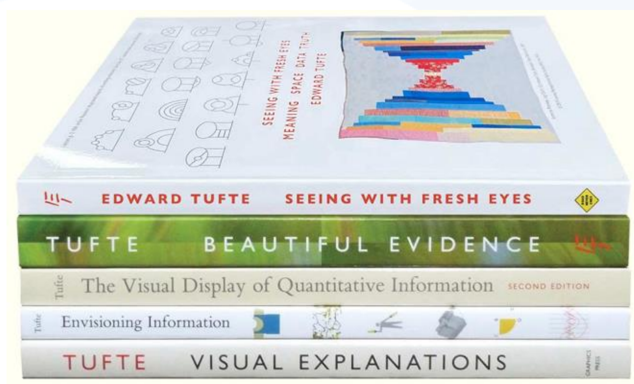
This book provides a plainspoken and thorough introduction to the web for historians—teachers and students, archivists and museum curators, professors a
work, or to build upon and improve the projects they have already started in this important new medium. It begins with an overview of the different genres of history websites, surveying a
beginning of the web. The book then takes the reader step-by-step through planning a project, understanding the technologies involved and how to choose the appropriate ones, designing
way that makes them web-friendly while preserving their historical integrity, and how to reach and respond to an intended audience effectively. It also explores the repercussions of copyri
cutting-edge web techniques involving interactivity, such as sites that use the medium to solicit and collect historical artifacts. Finally, the book provides basic guidance on insuring that th

Design Equation vs. Tufte

"Edward Tufte is a statistician/visualizer/artist, taught data analyis and policy making at Princeton and Yale 32 years, and also taught his one-day course on Presenting Data and Information to 328,000 students 1994-2020. He wrote, designed, and self-published **5 books on data visualization**. *The New York Times* described ET as the 'Leonardo da Vinci of data,' and *Bloomberg* as the 'Galileo of graphics.' "

https://www.edwardtufte.com/tufte/





Printing from WWW

- Landscape AND Portrait PDF
- Safe area in both orientations, responsive web design
- Both versions, e.g. Semantic Web milestone paper by Tim Berners-Lee from Scientific American page

- Typography – Donald KNUTH – TEX, LaTEX...
- One page typo ~ 100++ SKK

  - another „solved“ field Planar Graph Drawing

Edward Tufte Tips
on Powerpoint and Presentation Design

- Three simple suggestions from Edward Tufte:

- **Show up early** Something good is bound to happen—if there's no need to fix a mechanical problem or resolve a room conflict, you can always mingle with the audience.
- **How to start** –Clearly tell the audience: What the problem is, who cares, and what your solution is…
- **Always provide a handout** –Text on paper can provide more information than verbal communication (e.g. it takes 22 minutes to read the top half of the New York Times aloud). – This allows them to become engaged. – Assures that each point is covered (even if you forget something).

Tufte

Tufte: Aims

- Don't unneccessarily segregate text & graphics– don't turn them into silos.
- Aim for the truth. Truth wins.
- Aim for simplicity. Don't dumb, down however.
- Avoid distracting animations and clip art.
- Of course practice, practice, practice.

Tufte

Tufte: What to Avoid

Death by Powerpoint:

- It is used to guide and **to reassure a presenter**, rather than to enlighten the audience;
- It has **unhelpfully simplistic** tables and charts, resulting from the low resolution of early computer displays;
- The outliner causes ideas to be arranged in an unnecessarily deep hierarchy, itself subverted by the need **to restate the hierarchy** on each slide;
- **Enforcement of the audience's linear progression** through that hierarchy (whereas with handouts, readers could browse and relate items at their leisure);
- **Poor typography and chart layout**, from presenters who are poor designers and who use poorly designed templates and default settings (in particular, difficulty in using scientific notation);
- **Simplistic thinking**, from ideas being squashed into bulleted lists, and stories with beginning, middle, and end being turned into a collection of disparate, loosely disguised points. This may present an image of objectivity and neutrality that people associate with science, technology, and —bullet points.

Graph Drawing

- 1736 Euler

Kruja, E. et al. 2001. A Short Note on the History of Graph Drawing.
GD 2001: pp 272-286. [online]
http://www.merl.com/publications/docs/TR2001-49.pdf

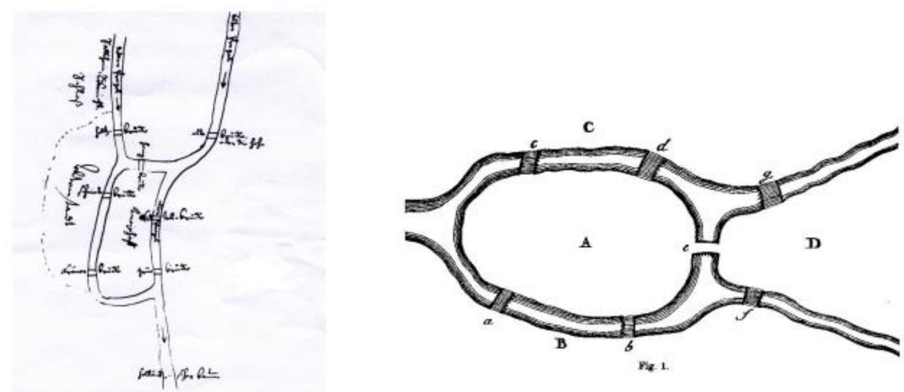


Fig. 9. Ehler's sketched map of Königsberg, 1736 (left), and Euler's more polished version [12]. Euler included one more sketched map (a variant of the first with more bridges included) in his paper, but no abstract graph drawing of the problem. Reproduced with permission.

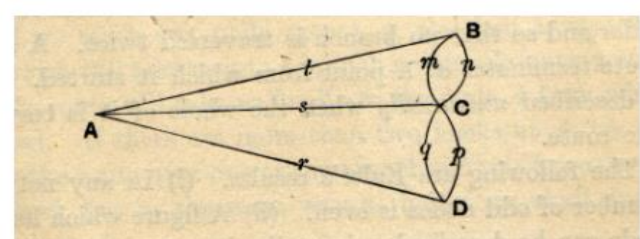
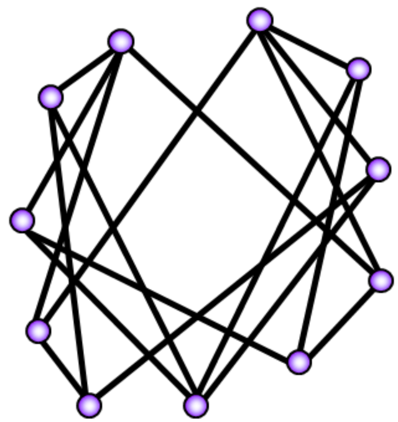


Fig. 10. Ball's 1892 graph-drawing abstraction of the bridges of Königsberg. The nodes represent the land areas and the edges represent the bridges connecting them.

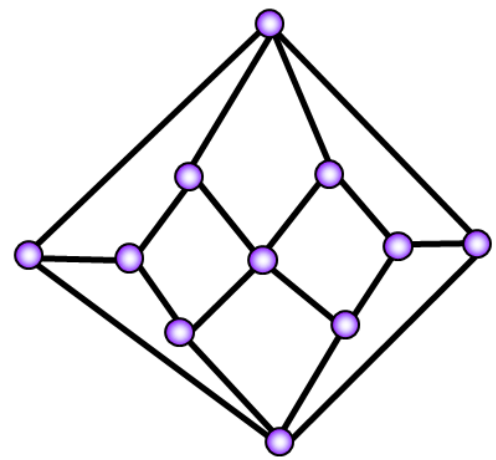
Tufte

Planar Graph Drawing

- Nishizeki, T. & Rahman, S. 2004. Planar Graph Drawing. World Scientific 2004.



structure of the graph is difficult to understand



structure of the graph is easy to understand

Tufte

Graph Drawing Styles

- Nishizeki, T. & Rahman, S. 2004. Planar Graph Drawing. World Scientific 2004.

- **Planar**
- **Polyline**
- **Straight Line**
- **Convex**
- **Orthogonal**
- **Box-Orthogonal**
- **Rectangular**
- **Box-Rectangular**
- **Grid**

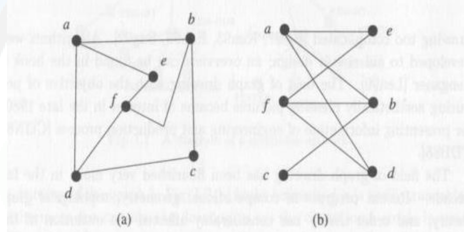
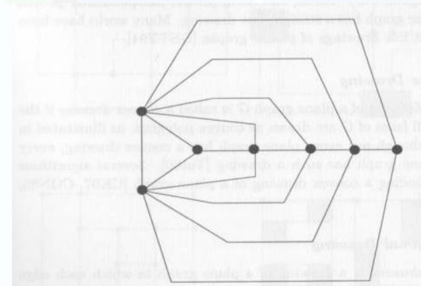


Fig. 1.3 (a) A planar drawing, and (b) a non-planar drawing of the same graph.



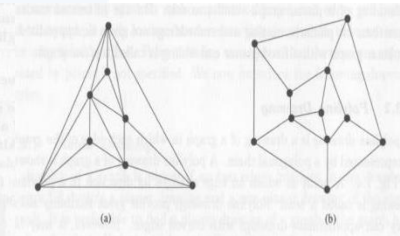


Fig. 1.5 (a) A straight line drawing, and (b) a convex drawing.

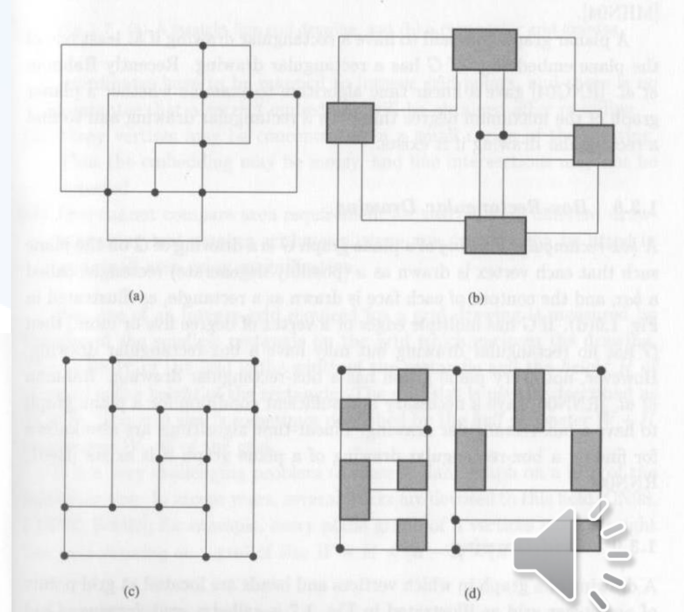


Fig. 1.6 (a) An orthogonal drawing, (b) a box-orthogonal drawing, (c) a rectangular drawing, and (d) a box-rectangular drawing.

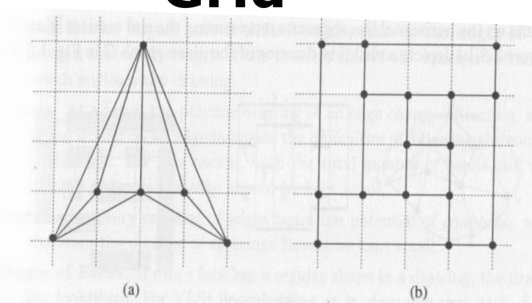


Fig. 1.7 (a) A straight line grid drawing, and (b) a rectangular grid drawing.

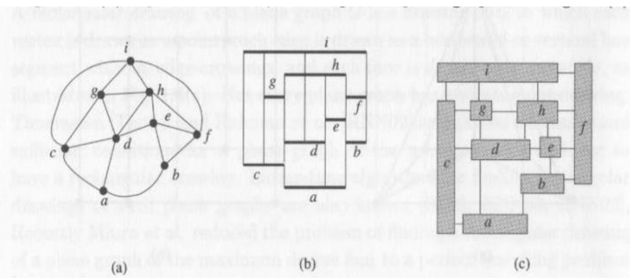
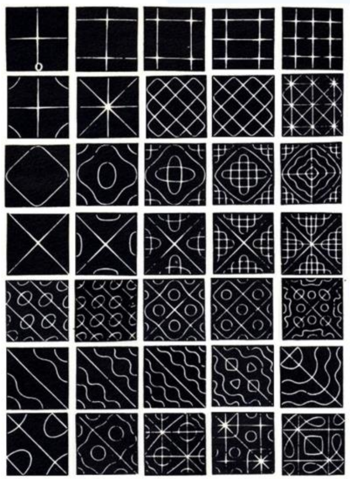
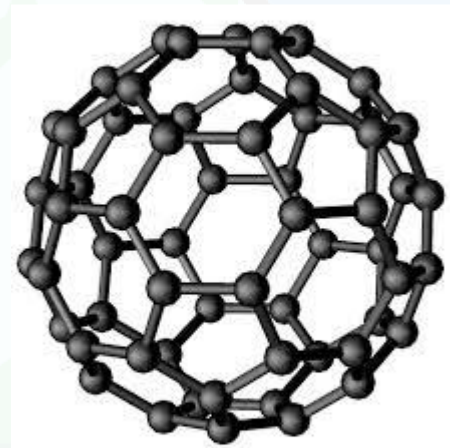
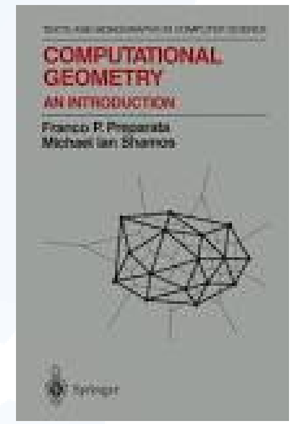


Fig. 1.8 (a) A plane graph G, (b) a visibility drawing of G, and (c) a 2-visibility drawing of G.

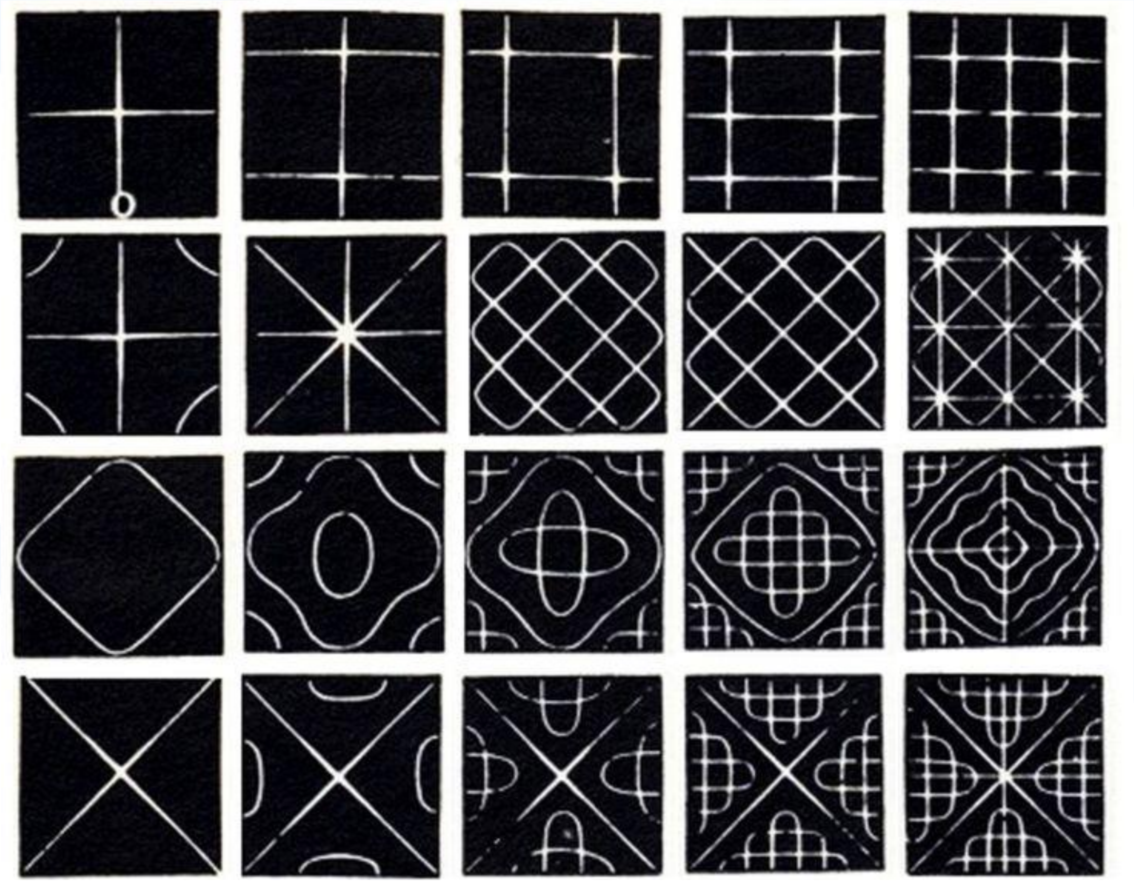
Observations

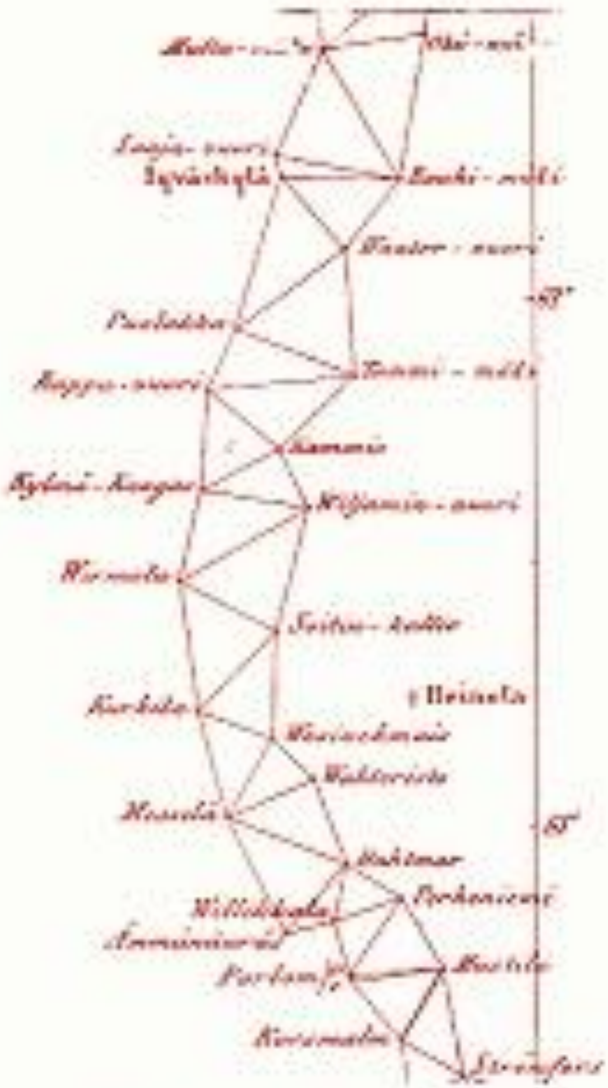


* Mezo: e.g. Chladni patterns, GPS, Struve arc 258
* Micro: C60 fulleren
* Macro: Graph drawing for star constellations 87/89

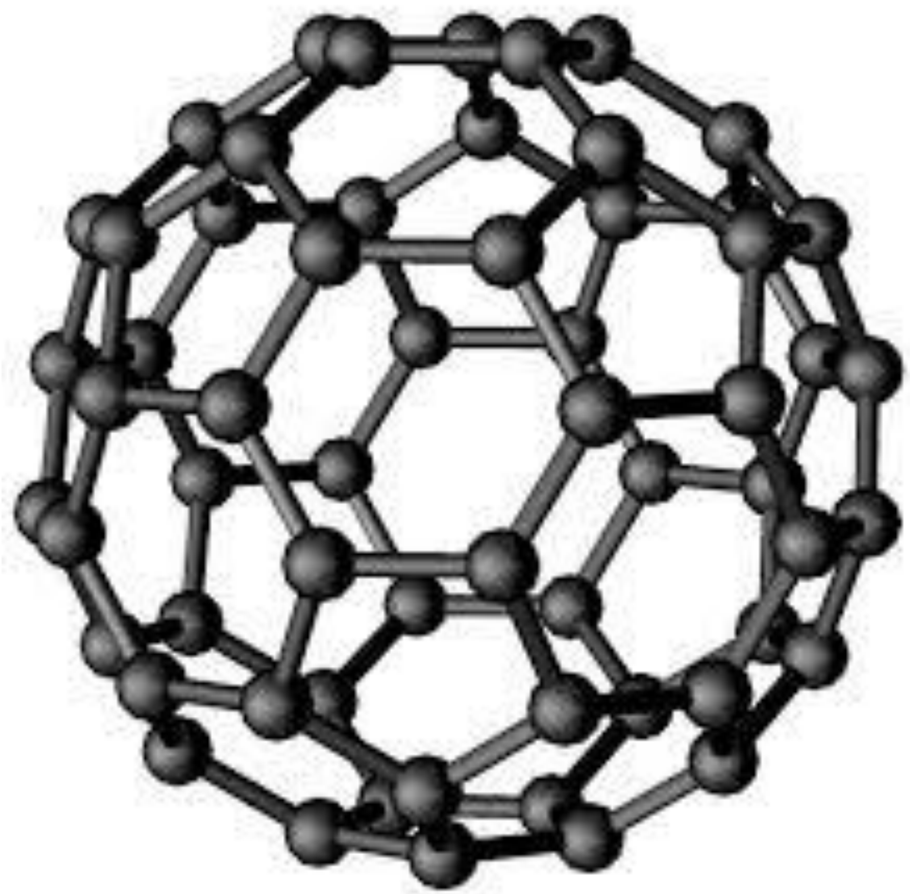
Chladni patterns

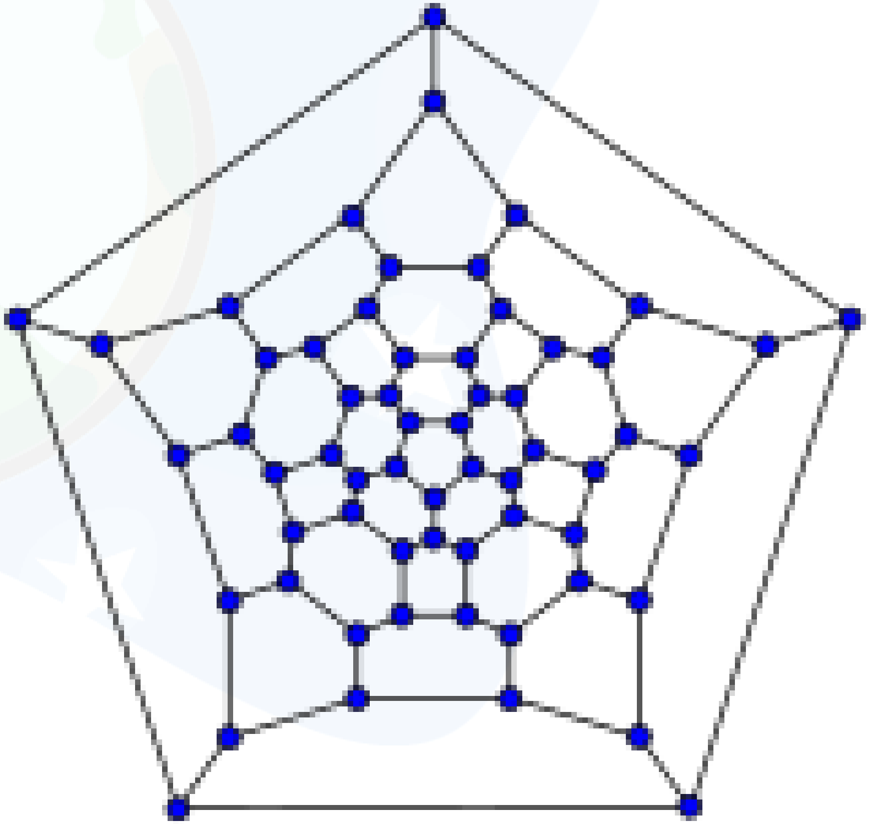




The Struve Geodetic Arc

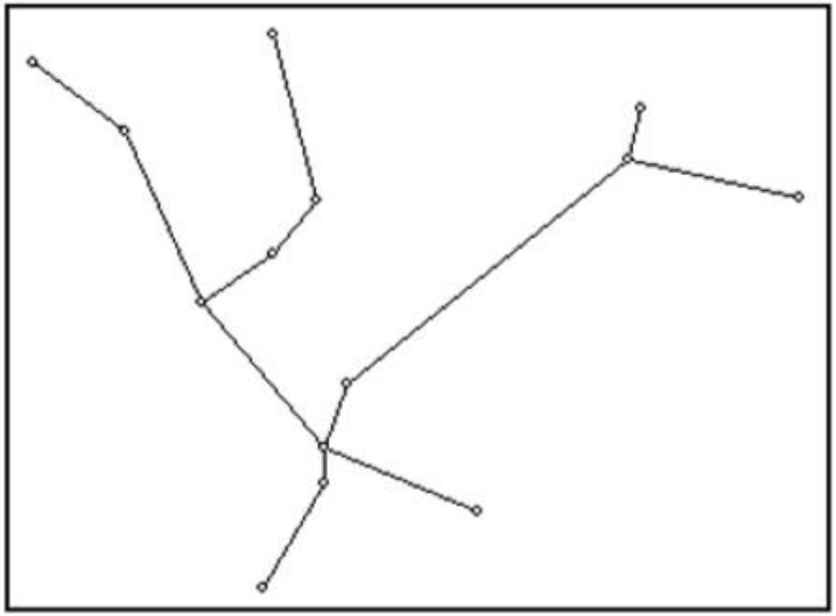
Fulleren C60





Andromeda

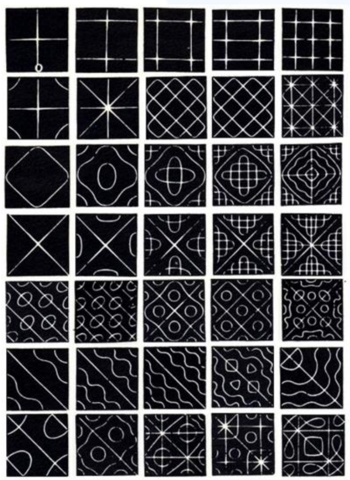
0.793642	0.625906
1.000000	0.507201
0.777619	0.558920
0.410155	0.264520
0.316204	0.720905
0.372658	0.505411
0.316201	0.436635
0.000000	0.686255
0.121954	0.596426
0.220482	0.371511
0.301413	0.000000
0.380619	0.134792
0.381947	0.181937
0.582601	0.098013

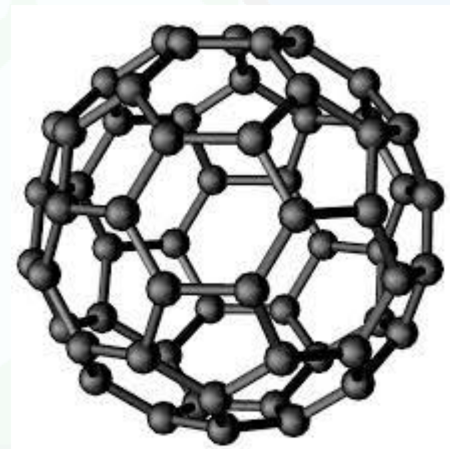
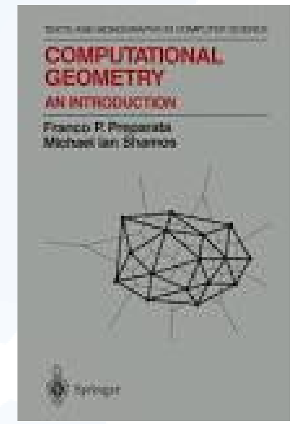


AND – Andromeda

Obr. 1. Příklad grafu souhvězdí Andromeda s přepočítanými planárními souřadnicemi poloh hvězd

Delaunay Triangulation (DT)



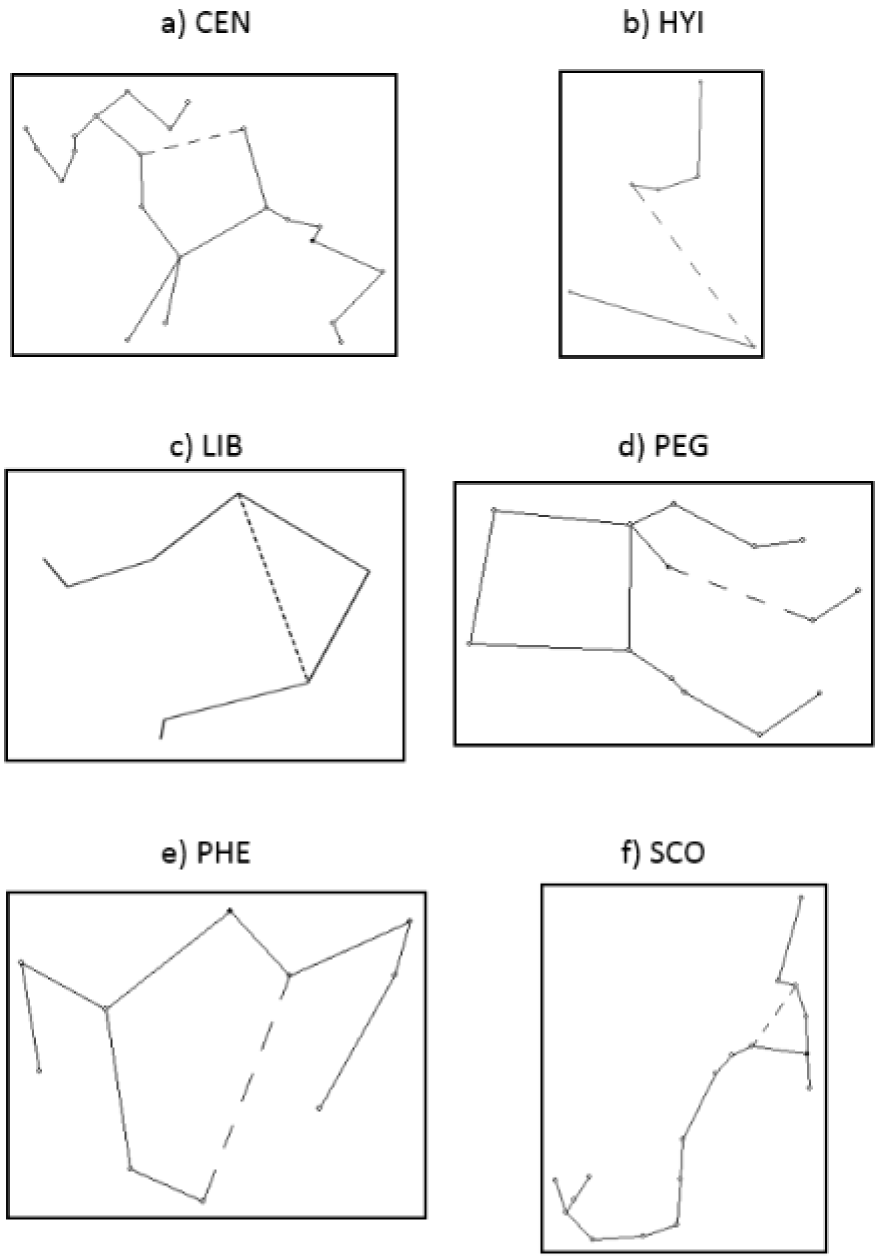
Mezo: e.g. Chladni patterns, GPS, Struve arc

Micro: C60 fulleren

Macro: Graph drawing for star constellations 87/89

Non DT

empty circle?



Obr. 2. Souhvězdí, která DT zcela nepokryje (nepokryté hrany jsou čárkovaně)

Graphics Design: ET's Principles

- 1 Introduction 2 History of Plots 3 The Explanatory Power of Graphics
- 4 Basic Philosophy of Approach 5 Graphical Integrity 6 Data Densities
- 7 Data Compression 8 Multifunctioning Graphical Elements
- 9 Maximize data-ink; minimize non-data ink
- 10 Small Multiples 11 Chartjunk 12 Colors
- 13 General Philosophy for Increasing Data Comprehension
- 14 Techniques for Increasing Data Comprehension
- 15 When NOT to Use Graphics 16 Aesthetics

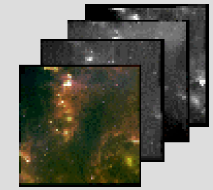
Tufte's works address the following issues:

E.T.

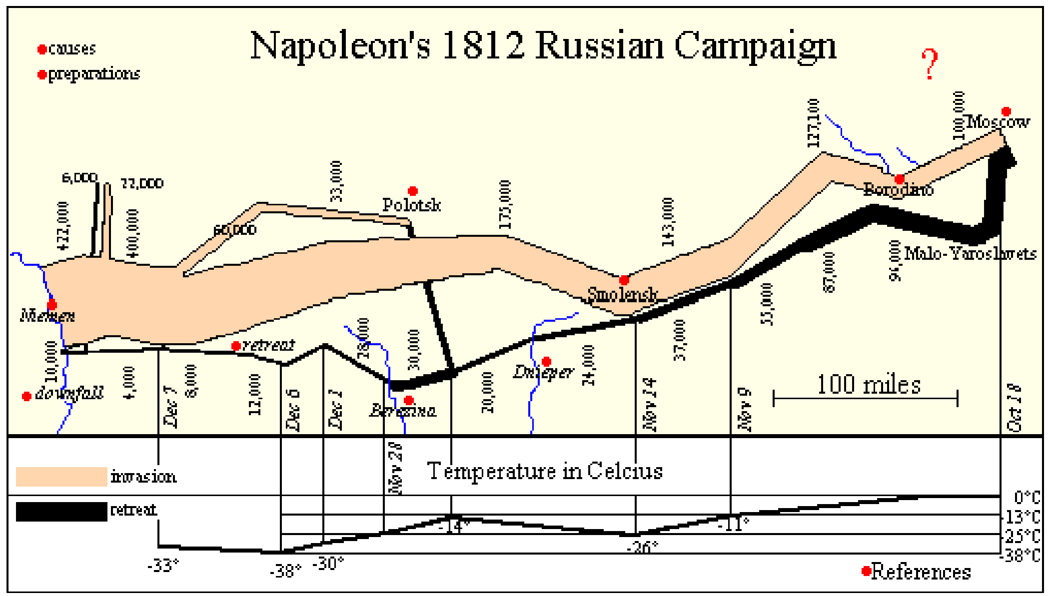
- The Problem: The problem is that of presenting large amounts of information in a way that is compact, accurate, adequate for the purpose, and easy to understand. Specifically, to show cause and effect, to insure that the proper comparisons are made, and to achieve the (valid) goals that are desired.
- Its Importance: Printed and graphical information is now the driving force behind all of our lives. It no longer is confined to specialized workers in selected fields but impacts nearly all people through the widespread use of computing and the Internet. Rapid and accurate transfers of information can be a life and death matter for many people (an example being the Challenger disaster). The extent to which symbols and graphics affect our lives can be seen by the dramatic increase in IQ scores in all cultures which have acquired information technology: in the United States there has been an average increase of 3 IQ points per decade over the last 60 years, for a total of an 18 IQ point increase. There is no known biological explanation for this increase and the most likely cause is widespread exposure to text, symbols, and graphics that accompany modern life. As mentioned above, this increase has been seen in all cultures exposed to information technology.
- Its Application: Some of the information relates to the displays of statistical information, but much applies to any type of display, even plain text.

Explanatory Power of Graphics

Explanatory Power of Graphics



Examples of Visualization



This graphic is an adaptation of M. Charles Joseph Minard´s „March of the Napoleon Army" by Sunny McClendon, as part of an Information Design Class at the University of Texas at Austin.

Tufte

Approach

Basic Philosophy of Approach

Important rules and themes to use when presenting graphics:

- Assume that the audience is intelligent (a paraphrase from E.B. White). Even publications, such as NY Times, assume that people are intelligent enough to read complex prose, but too stupid to read complex graphics.
- Don't limit people by "dumbing" the data -- allow people to use their abilities to get the most out of it.
- To clarify -- add detail (don't omit important detail; e.g., serif fonts are more "detailed" than san serif fonts but are actually easier to read). And Einstein once said that "an explanation should be as simple as possible, but no simpler".
- Above all else, show the data. Graphics is "intelligence made visible"
- Data rich plots can show huge amounts of information from many different perspectives: cause & effect, relationships, parallels, etc. (VD-31: train schedule, VD-17: Chloroplethic map, VD-41: Napoleon's campaign, EI-49: space junk)
- Plots need annotation to show data, data limitations, authentication, and exceptions (VE-32: text of exceptions)
- Don't use graphics to decorate a few numbers

Tufte

Approach

Basic Philosophy of Approach

Important rules and themes to use when presenting graphics:

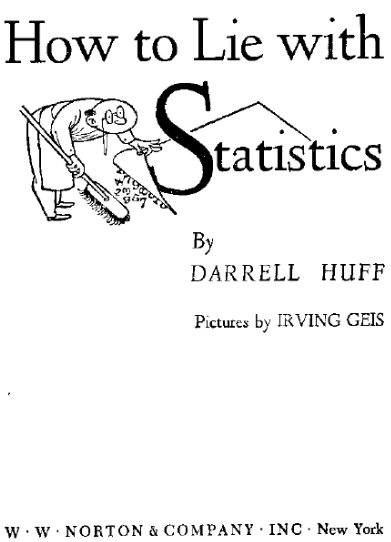
- Assume that the audience is intelligent (a paraphrase from E.B. White). Even publications, such as NY Times, assume that people are intelligent enough to read complex prose, but too stupid to read complex graphics.
- Don't limit people by "dumbing" the data -- allow people to use their abilities to get the most out of it.
- To clarify -- add detail (don't omit important detail; e.g., serif fonts are more "detailed" than san serif fonts but are actually easier to read). And Einstein once said that "an explanation should be as simple as possible, but no simpler".
- Above all else, show the data. Graphics is "intelligence made visible"
- Data rich plots can show huge amounts of information from many different perspectives: cause & effect, relationships, parallels, etc. (VD-31: train schedule, VD-17: Chloroplethic map, VD-41: Napoleon's campaign, EI-49: space junk)
- Plots need annotation to show data, data limitations, authentication, and exceptions (VE-32: text of exceptions)
- Don't use graphics to decorate a few numbers

Tufte

Lies

In addition to "lies, damn lies, and statistics", graphics can also be used to deceive. For example, deceptive graphics may:

- Compare full time periods with smaller time periods (VD-60: Nobel prizes, which compares 10 year time periods with one 5 year period)
- Use a "lie factor" [= (size of graphic)/(size of data)] to exaggerate differences or similarities
- Use area or volume representations instead of linear scales to exaggerate differences. See VD-69: "Shrinking family doctor" as an example of how to confuse people using 1 versus 2- and 3- dimensional size comparisons. Area and volume representations fool people with the square/cube law: an increase in linear size leads to a square of the increase for areas and a cube of the increase for volumes.
- Fail to adjust for population growth or inflation in financial graphs
- Make use of design variation to obscure or exaggerate data variation (VD-61: exaggeration of OPEC prices)
- Exaggerate the vertical scale
- Show only a part of a cycle so that data from other parts of the cycle cannot be used for proper comparison

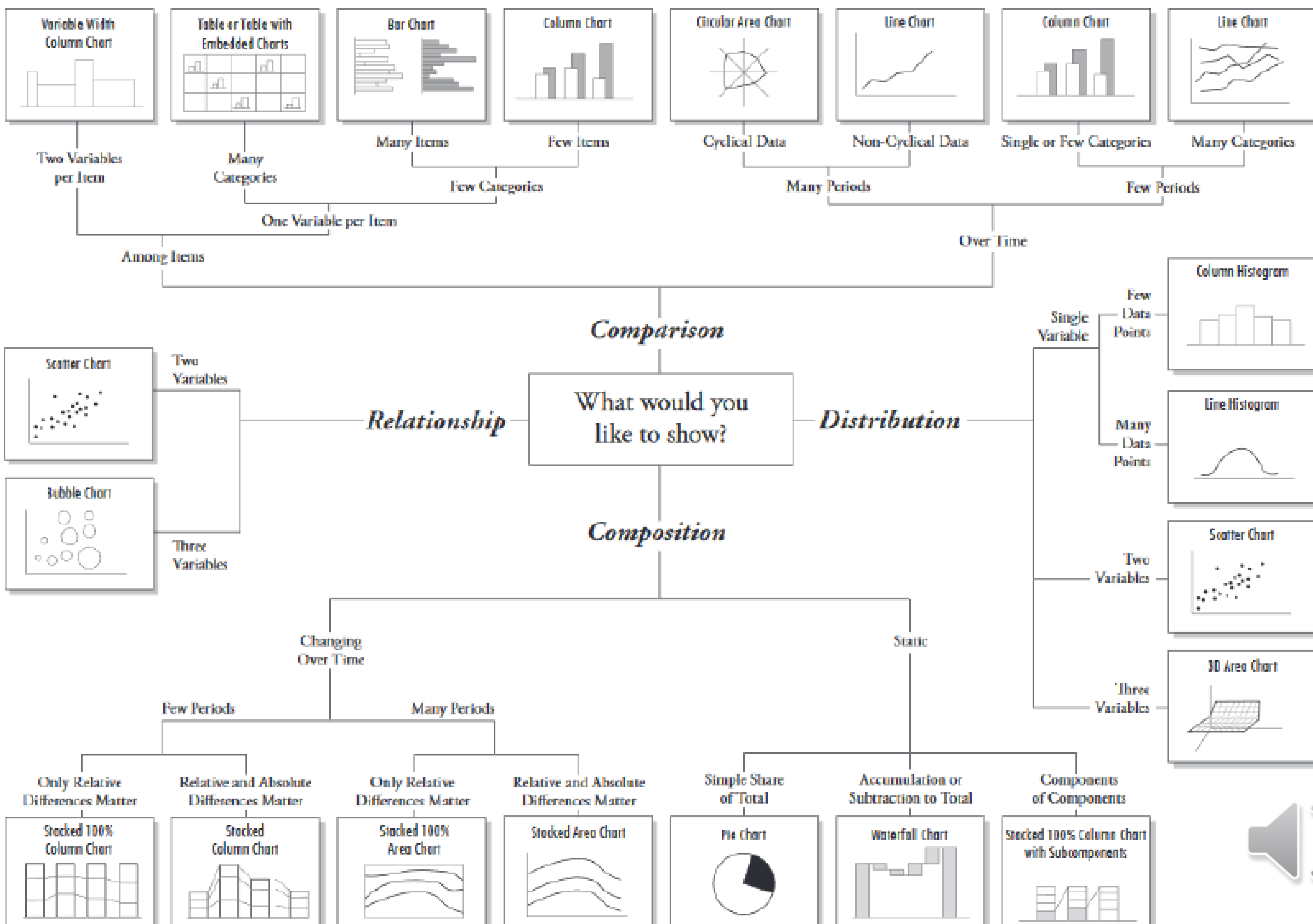


Questions (Polya 38, Alsthuller 40)

- Ask the right questions:
  1. Does the display tell the truth
  2. Is the representation accurate
  3. Are the data documented
  4. Do the display methods tell the truth
  5. Are appropriate comparisons, contrasts, and contexts shown

SCHMIDT, J. 2018. *Visualisation in Data Science*. TU Wien.

Chart Suggestions—A Thought-Starter



Density

Data Densities

Graphics are at their best when they represents very dense and rich datasets. Tufte defines data density as follows:

Data density = (no. of entries in data matrix)/(area of graphic)

Note that low data densities on computer displays force us to view information sequentially, rather than spatially, which is bad for comprehension. Good quality graphics are:

- Comparative
- Multivariate
- High density
- Able to reveal interactions, comparisons, etc
- And where nearly all of the ink is actual data ink

Example data densities include:

- 110,000 numbers/sq-inch for an astronomical graph. This is the maximum known density for a graph. For most scientific journals we get about 50-200 numbers/sq-inch

- 150 Mbits = human eye
  8 Mbits = typical computer screen
  25 Mbits = color slide

Tufte

Compress

Data Compression

- Use data compression to reveal (not hide) data . For example, EI-22: "Sun Spot cycles" displays sunspots as thin vertical lines in the y-axis direction only in order to present many such spots over a period of time on a single graph
- Use compression to show lots of information in a single graph, such as a plot that shows x-axis, y-axis, and x/y interactions. (VD-134: Pulsar signals; VE-111)
- Exclude bi-lateral symmetry when it is redundant (e.g., charnoff faces) or extend it when it aids comprehension (50% more view of the world on a world map provides a wrap-around context that aids understanding). Studies show that we often concentrate on one side of a symmetrical figure and only glance at the other side.

Ink

Maximize Data-ink; Minimize non-Data Ink

Tufte defines the data ink ratio as:

Data Ink Ratio = (data-ink)/(total ink in the plot)

The goal is to make this as large as is reasonable. To do this you:

- Avoid heavy grids
- Replace box plots with interrupted lines (VD-125: reduced box plot)
- Replace enclosing box with an x/y grid
- Use white space to indicate grid lines in bar charts (VD-128: white spaces)
- Use tics (w/o line) to show actual locations of x and y data
- Prune graphics by: replacing bars with single lines, erasing non-data ink; eliminating lines from axes; starting x/y axes at the data values [range frames])
- Avoid over busy grids, excess ticks, redundant representation of simple data, boxes, shadows, pointers, legends. Concentrate on the data and NOT the data containers.
- Always provide as much scale information (but in muted form) as is needed

Alias, moire…

Chartjunk

Chartjunk consists of decorative elements that provide no data and cause confusion.

- Tufte discusses the rule of 1+1=3 (or more): 2 elements in close proximity cause a visible interaction. Such interactions can be very fatiguing (e.g., moiré patterns, optical vibration) and can show information that is not really there (EI-60: data that is not there, VD-111: chart junk)
- In major science publications we see 2% to 20% moiré vibration. For example, in recent statistical and computer publications chartjunk ranges from 12% to 68%
- Techniques to avoid chartjunk include replacing crosshatching with (pastel) solids or gray, using direct labeling as opposed to legends, and avoiding heavy data containers

Colors

- Layering with colors is often effective
- Color grids are a form of layer which provides context but which should be unobtrusive and muted
- Pure bright colors should be reserved for small highlight areas and almost never used as backgrounds.
- Use color as the main identifier on computer screens as different objects are often considered the same if they have the same color regardless of their shape, size , or purpose
- Contour lines that change color based on the background standout without producing the 1+1=3 effects
- Colors can be used as labels, as measures, and to imitate reality (e.g., blue lakes in maps).
- Don't place bright colors mixed with White next to each other.
- Color spots against a light gray are effective
- Colors can convey multi-dimensional values
- Scroll bars should be solid pastel colors
- Note that surrounding colors can make two different colors look alike, and two similar colors look very different (EI-92/93: effects of context on colors).
- Subtle shades of color or gray scale are best if they are delimited with fine contour lines (EI-94: shades with contours)
- Be aware that 5-10% of people are color blind to some degree (red-green is the most common type followed by blue-yellow, which usually includes blue-green)

No Graphics

When NOT to Use Graphics

- Often text tables can replace graphs for simple data; you can also use 2D text tables, where row and column summaries are useful. Non-comparative data sets usually belong in tables, not charts
- Poster designs are meant just to capture attention, as opposed to conveying information -- generally they are not good designs for graphs.
- If a picture is not worth a 1000 words, to hell with it (quote from Ad Reinhardt -- note this is from the original Chinese quote that "a picture is worth 10,000 words).

Aesthetics

Aesthetics

Graphical excellence consists of simplicity of design and complexity and truth of data. To achieve this

- Use words, numbers, drawings in close proximity
- Display an accessible complexity of data
- Let the graphics tell the story
- Avoid context-free decoration
- Use lines of different weights as an attractive and compact way to display data (VD-185: Mondrian)
- Make use of symmetry to add beauty (although someone once said that "all true beauty requires some degree of asymmetry")

Once upon a time, there was the end

- Story >> another story

Graphics and Web Design Based on Edward Tufte's Principles

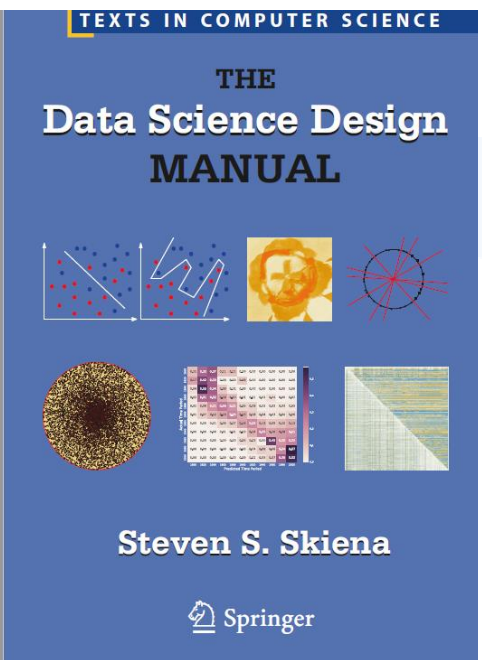
Other Topics: **Other References**

Chapter 6

Visualizing Data

At their best, graphics are instruments for reasoning.

– Edward Tufte



Andrej Ferko
Comenius University, Bratislava, ferko@fmph.uniba.sk
www.sccg.sk/wega

Oct 29, 2020