Info 424

Introduction to Tufte's Principles
The Visual Display of Quantitative Information (ch 1 & 2)
Envisioning Information (ch 1)

Today's lecture

Graphical Excellence & Graphical Integrity (Tufte)
- Overview with examples
- Visualization Critique example

Escaping Flatland
- What does it mean?
- Examples and discussion

Lab this week

Graphical Excellence

In brief
- Interesting data (complex ideas, multivariate)
- Clear, precise, concise presentation (data-ink ratio)
- Accurate communication (lie factor)

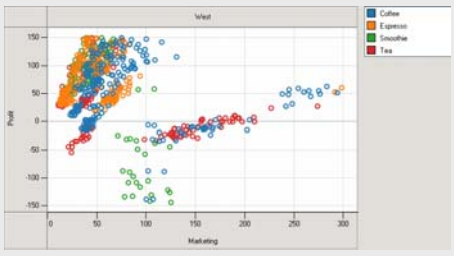
Tufte
- Well-designed presentation of interesting data—a matter of substance, of statistics, and of design
- Complex ideas communicated with clarity, precision and efficiency
- Gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- Nearly always multivariate
- Requires telling the truth

The data

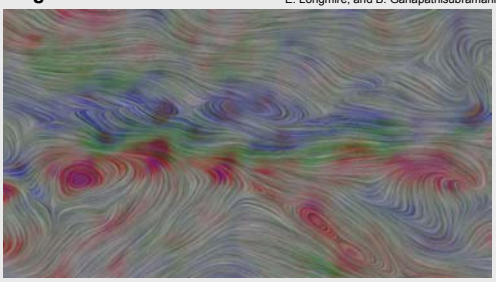
A visualization can be no better than its data
—Quantity, quality, schema, model

Multivariate

Multivariate: multiple variables
Relational: Relationship between variables

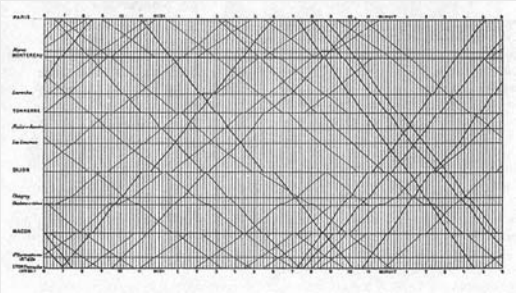


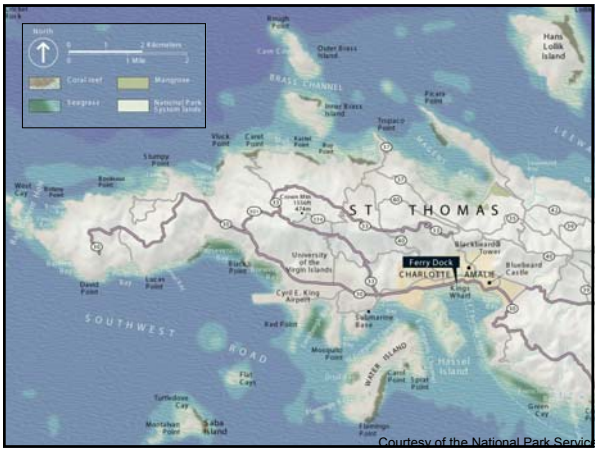
Effectively Visualizing Multi-Valued Flow Data Using Color and Texture
T. Urness, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani



Positive vorticity (red), Negative vorticity (blue)
Strongly negative Reynolds shear stress (green)
High swirl strength (orange or magenta, depending on direction) link

Time series (train schedule)





Show the data

Maximize the "Data-Ink Ratio"

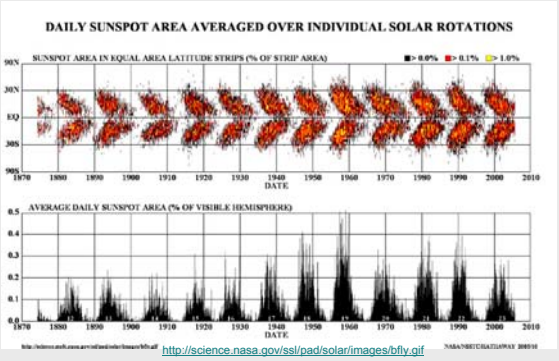
- Data-ink = the ink used to show data
- Data-ink ratio = data-ink/total ink used

Process

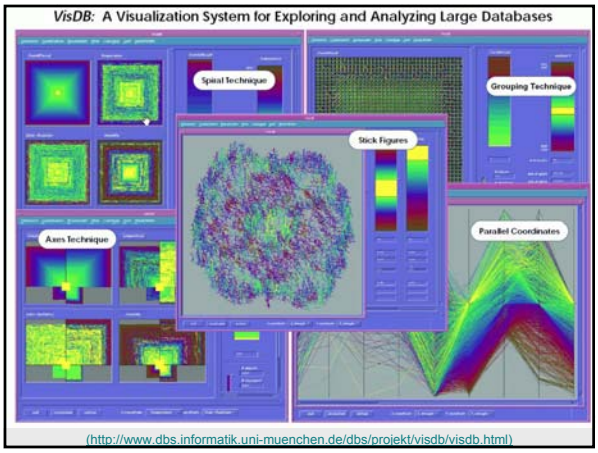
- Erase non-data ink
- Erase redundant data ink
- Revise and edit

http://www.tbray.org/ongoing/data-ink/di1

Data Density



http://science.nasa.gov/ssl/pad/solar/images/bfly.gif



(http://www.dbs.informatik.uni-muenchen.de/dbs/projekt/visdb/visdb.html)

Graphical Integrity

The representation of numbers as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented

Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.

Show data variation, not design variation.

In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

Graphics must not quote data out of context.

In brief

Present value relationships accurately
- Presentation size precisely matches data
- Avoid area and volume encodings
- Adjust currency values for inflation or other correlated changes (such as population changes)

Label carefully and clearly

Present data in context

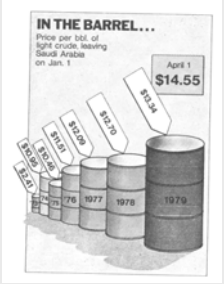
Lie Factor

Lie Factor
- (Size of effect in graphic)/(size of effect in data)
- Truth = 0.95-1.05 (5% error)
- 2-5 not uncommon

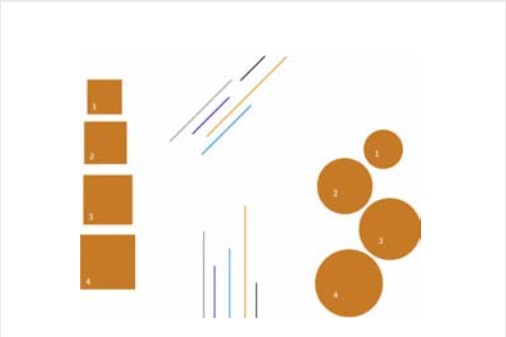
Lie factor by area = 9.4

Lie factor by volume = 59.4

NB: not adjusted for inflation



Estimating Relationships



Design Distortions

Gratuitous 3D and perspective

Non-uniform intervals and scales

Design Distortion
- Yearly vs. Quarterly
- Nonlinear scale
- Perspective distortion

Annual increase is actually similar to many previous years

Include detailed labels

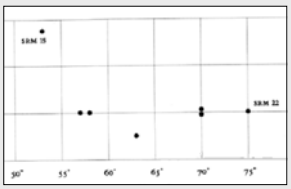
Clear, detailed, complete
- Provides precise values
- Indicates important events
- Defines units, context
- Corrects for presentation distortion

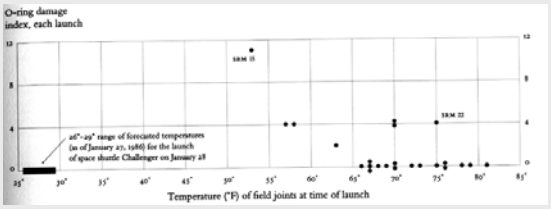
Label directly on the graphic
- Makes graphic self-explanatory
- Maintains focus (compared to legends)

Show context

Challenger data

Only complete context shows trend





Summary

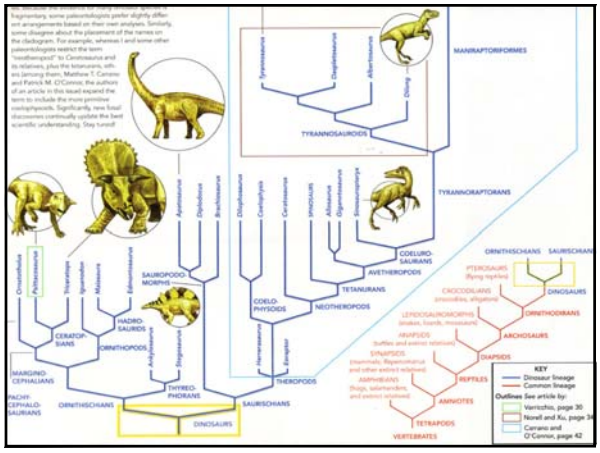
Graphical Excellence
- Interesting data (complex ideas, multivariate)
- Clear, precise, concise presentation (data-ink ratio)
- Accurate communication (lie factor)

Graphical Integrity
- Present value relationships accurately
  - Size matches data
  - Avoid area and volume encodings
  - Adjust currency values for inflation, etc.
- Label carefully and clearly
- Present data in context

Visualization Critique

One good and one bad example
Static, from a non-web source
Due Thursday, Oct 19
- Put on Wiki by 8 am

Requirements
- Image
- Explanation (what it shows, where it came from)
- Deconstruction (data and models)
- Critique (does it work?)



Natural History Magazine

Explanation
- Cladogram (evolutionary family tree) of dinosaurs
- Youth audience

Deconstruction
- Large, archeological datasets
- Connecting lines show hierarchical structure
- Color coding to link to overview tree, articles, etc.

Critique
- Good example
- Non-trivial data, well designed
- Dense data showing complex relationships
- Context and cross-linking provided
- Clearly labeled

Dell Computer Ad

Explanation
- Advertisement for Dell Pentium D computers
- Small and Medium business audience

Deconstruction
- Dataset is all possible processors and options
- Includes cost, availability, and marketing goals

Critique
- While the ad appears to be giving a detailed comparison, it is data thin and confusing.
- Many items are identical, wasting space
- Difficult to pull out critical differences
- Pictures lie, as monitor is not included in costs
- Actual set of configurations is much larger

Questions?

Envisioning Information

Escaping Flatland

Discussion Questions

What does "escaping flatland" mean?
What examples did you find confusing?
What examples did you especially like?

Friday's lab

Design exercise: redesign a bad example
Stephen Few examples