

CSE 332: Introduction to Visualization

Lecture 2: Principles of Excellence and Integrity in Visualization

(adopted from Edward R. Tufte's texts)

Klaus Mueller

Stony Brook University

Computer Science Department

Graphical Display

(taken from Edward R. Tufte “The Visual Display of Quantitative Information”)

- Can be more precise and revealing than numerical display
 - example: Anscombe’s quartet (pg. 13/14)
 - example: cholera map of central London, 1854, by Dr. John Snow (pg. 24)
- Can capture a large amount of information in a very small space (billions of bits on one page)
 - example: data maps for cancer incidence (pg. 17)
 - example: galaxy maps (pg. 27)
 - example: space debris (pg. 48, Tufte “Envisioning Information”)
- Can extend to time-series display
 - example: train schedule Paris-Lyon, 1880s (pg. 31)
- Can be narrative
 - example: Napoleon’s Russia campaign, 1812, plots 6 variables on a 2D graph (pg. 41)
- Can represent each datapoint by visual information (graphic, icon, image, color, pattern)
 - examples: fear-rage graph (pg. 50), Chernoff faces (pg. 97, 142)

Principles of Graphical Excellence

According to Tufte (pg. 51):

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

(Nevertheless, visualizations should be visually pleasing and may very well have an artistic touch)

Story Telling In Paintings

The corset symbolizes the restrictions that society, religion, etc. impose on women [The Artist is a woman]. The scissors are breaking up these ties and after that the path towards liberation is open. The dragonflies are insects that only come when a storm is up to happen. The storm in Jungian [a German author] symbolism symbolizes a time of change.

The eclipse symbolizes the union of primordial polarities represented generally by sun (male) and moon (female) (See Carl Gustav Jung: Man and his Symbols). This union of opposites represents a state of absolute unity. This sacred union of basic polarities forms the central motif of many religions and philosophies and for Jungian thought is the highest mark of psychological development. Jung sees the development of the Self, our true individuality, as dependent on the realization of all aspects of our nature, in particular the masculine and feminine archetypes. So the flying women have freed themselves from those things which kept them anchored in models of behavior imposed by somebody else.



Maria Alejandra Zanetta
“Solar Eclipse”

Graphical Integrity (What to Avoid in Visualization)

(taken from Edward R. Tufte “The Visual Display of Quantitative Information”)

- Is cosmetic decoration really needed to make data more interesting (may only distract):
 - example: diamond graph (adds a useless 3rd dimension)
- Misleading graphical representation
 - example: missing baseline in Day Mines, Inc. annual report (pg. 54)
 - example: non-uniform data spans in Commision Payments graph (pg. 54)
 - example: non-uniform scaling of icons in Pittsburgh Civic Commission report (pg. 55)
- The Lie Factor = $\frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$ (should be within [0.95, 1.05])
 - example: graph on fuel economy standards for autos (lie factor = 14.8) (pg. 57)
- Visualizing data bearing some dimension by means of objects of higher dimensions:
 - example: the *growing barrel* (lie factor: 9.4 (2D), 59.4 (3D)) (pg. 62)
 - example: the *growing oil pump* (lie factor: 9.5) (pg. 62)
 - example: the *shrinking dollar bill* (lie factor: ~6) (pg. 70)
 - example: the *incredibly shrinking family doctor* (pg. 69)

→ the number of information carrying dimensions should not exceed the data dimensions

Graphical Integrity II (What to Avoid in Visualization)

(Edward R. Tufte: “The Visual Display of Quantitative Information” and “Visual Explanations”)

- Quoting data out of context and/or too sparse (recall: graphics allows high data density)
 - example: Connecticut traffic deaths (pg. 74/75)

Principles that ensure graphical integrity:

- The representation of numbers should be directly proportional to the numerical quantities represented (see the growing barrels)
- Clear and detailed labeling should be used to defeat graphical distortion and ambiguity
- Show data variations and not design variations (see the fuel economy graph)
- In time-series displays of money, show deflated and standardized units
- The number of information carrying dimensions should not exceed the data dimensions (see the growing barrels, the shrinking doctor)
- Graphics must not quote data out of context (see the Connecticut traffic deaths)
- Convincing graphics must demonstrate cause and effect (see Challenger disaster)