Notes for Class 2: What is Visualization?

Acknowledgements: Tamara Munzner & Torsten Möller, whose notes I will borrow and change throughout the course.

Objectives
By the end of the class, you will be able to…
• Define visualization
• Describe purposes for using visualization
• Explain some basic principles of good design
• Explain several ways in which visualization can be misleading

Agenda for Class
• Go over notes based on the readings. Show some additional examples in class.

Reading Assignment
Tufte Chapter 1 (Escaping Flatland)
Ware Chapter 1 (Foundation for a Science of Data Visualization)

Notes:

1. What is Visualization?
   A picture says more than a thousand numbers (variation on a Chinese proverb).
   Definitions:
   • A graphical representation of data or concepts. An external artifact supporting decision making. (Ware)
   • The use of computer-supported, interactive, visual representations of data to amplify cognition. (Card, Mackinlay, and Shneiderman)

2. What are some advantages of visualization?
• Data analysis
  o External representation – reduces load on working memory (e.g., grocery list)
  o View a lot of data in one place (e.g. Napoleon’s Russia campaign of 1812 and train schedule) Napoleon’s march plots several variables: army size, 2D location, temperature, direction of travel, date, river crossings (from Tufte)
Identify patterns (e.g. 1864 cholera map of London by John Snow, also Galileo’s sunspots)

Identify artifacts or errors in data
Understand both large scale and small scale features of the data. Put data in context of other data, spatial layout, etc.
Interact with and manipulate data to explore hypotheses

• Communication
  Present and explain information to others
  Provide a common picture to aid discussion

• Education

3. How can visualization be misleading?
Visual representations of data can be either intentionally or unintentionally misleading. For example:

• Visual Illusions
• Graphic exaggeration of effect sizes can be done by using different scales for the graphic representation or by not showing zero values

Lie Factor = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}

Here the ratio of line lengths is different from the ratio of data values

Height of the barrels may accurately reflect the data values but volume of the barrels does not and is visually prominent.

• Quoting data out of context (bold segment looks like a major decrease until put in context).

4. Some Principles of good design
• Show the context (e.g. Tufte - A Study of a Numerically Modeled Severe Storm)
• Avoid “chart junk” and visual clutter. Emphasize the data rather than the graphic design, method of production, glitziness, or something else (e.g., viz-o-matic)
• Avoid distorting the data
• Reveal the data at several levels of detail (e.g. micro / macro readings):
  o Provide an overview of a large set of data
  o Allow the viewer to access precise details – sometimes this is best done through interaction (“details-on-demand”, e.g. interactive flight schedule)
• Encourage the eye to compare different pieces of data (e.g. small multiples)
• Use clear and precise labeling
• Use symbols that do not need to be learned, when possible (sensory rather than arbitrary representations). If you need to use arbitrary symbols, use standard ones when possible.

5. Data Types
  o Entities vs relationships
  o Attributes have
    ▪ Quality (nominal, ordinal, interval, ratio)
    ▪ Dimensions (1D, 2D, 3D, …)
  o These properties influence how data should be encoded to be most easily understood.

**Key Concepts for Today:**
1. Visualization can be a valuable tool for data analysis and communication.
2. Visualization techniques must be carefully designed to be effective. Avoid misleading the user!
3. Principles of human perception and graphic design can help us to develop effective visualizations.