CS855: Data Visualization

Jaya Sreevalsan Nair

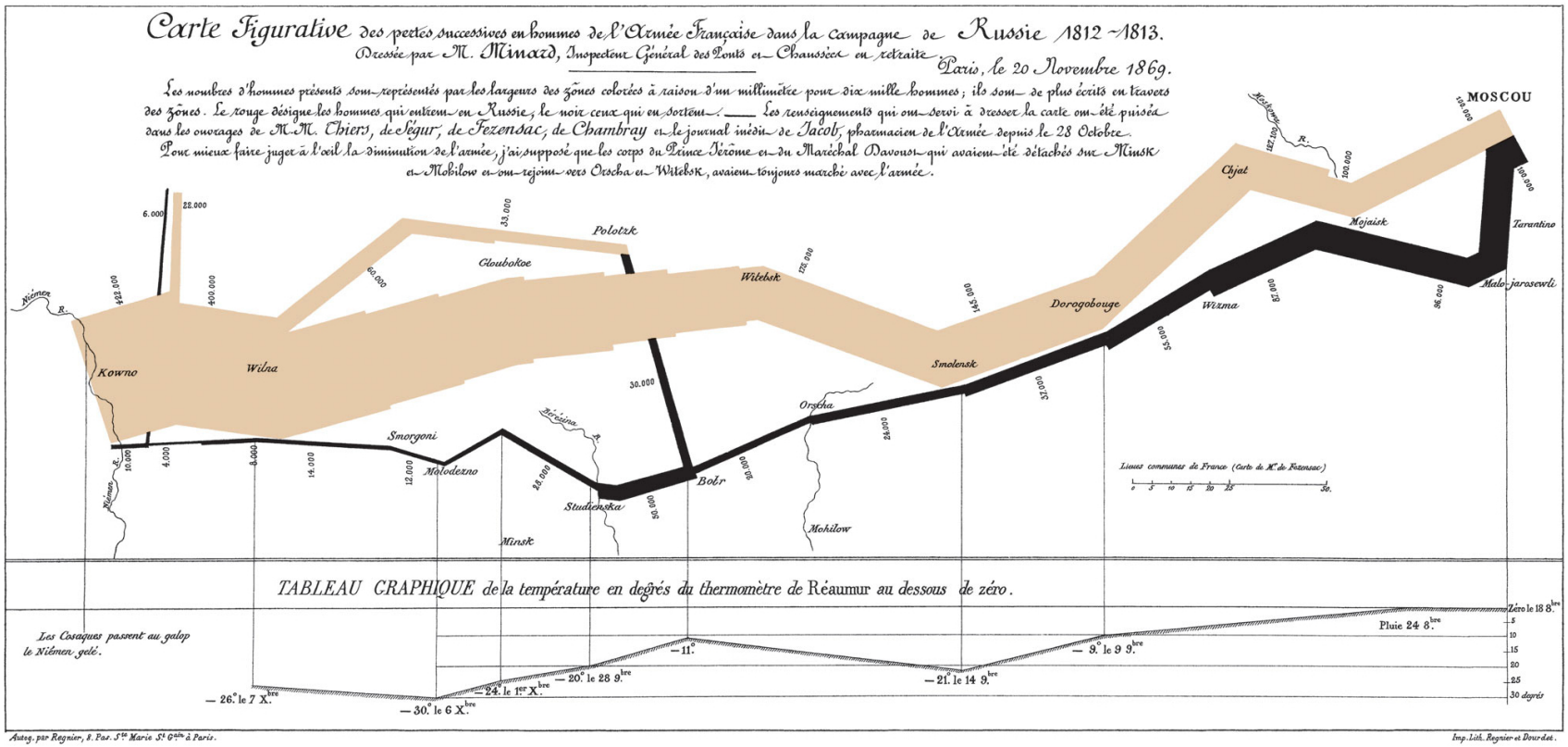
International Institute of Information Technology, Bangalore

Term I: 2014-15: Lecture 01

Agenda

- Introduction to Visualization
  - Need/motivation and purpose
  - Data representation

Classic example:



Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813, a flow map by Charles Minard, published in 1869 on the subject of Napoleon's disastrous Russian campaign of 1812.

Courtesy: Edward R. Tufte. 1986. The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT, USA.

Recommended reading of second edition of the book in 2007.

Graphical elegance is often found in simplicity of design and complexity of data.

...

Epilogue: ..

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.

From Tamara Munzner's textbook:

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods. The design space of possible vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations. Vis design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so validating the effectiveness of a design is both necessary and difficult. Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays. Vis usage can be analyzed in terms of why the user needs it, what data is shown, and how the idiom is designed.

From Tamara Munzner's textbook: She uses "vis" for visualization, in short.

Computer-based visualization systems provide **visual representations of datasets** *designed to help people carry out tasks more effectively.*

Visualization is suitable when there is a need to **augment human capabilities** *rather than replace people with computational decision-making methods. The* **design space of possible vis idioms** *is huge, and includes the considerations of both how to create and how to interact with visual representations. Vis design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so validating* **the effectiveness of a design** *is both necessary and difficult. Vis designers must take into account three very different kinds of* **resource limitations***: those of computers, of humans, and of displays. Vis usage can be analyzed in terms of* **why the user needs it, what data is shown, and how the idiom is designed.**

In Other Words ...

> *Visualization is a "cognitive process performed by humans in forming a mental image of a domain space. In computer and information science it is, more specifically, the visual representation of a domain space using graphics, images, animated sequences, and sound augmentation to present the data, structure, and dynamic behavior of large, complex data sets that represent systems, events, processes, objects, and concepts."*

- Williams, J. G., K. M. Sochats, and E. Morse, "Visualization," *Annual Review of Information Science and Technology (ARIST) 30* (1995), 161-207.

Visualization is a "cognitive process performed by humans in forming a mental image of a domain space. In computer and information science it is, more specifically, the visual representation of a domain space using graphics, images, animated sequences, and sound augmentation to present the data, structure, and dynamic behavior of large, complex data sets that represent systems, events, processes, objects, and concepts."

- Williams, J. G., K. M. Sochats, and E. Morse, "Visualization," *Annual Review of Information Science and Technology (ARIST) 30* (1995), 161-207.

The purpose of computing is **insight**, *not numbers.*

- Richard Hamming, 1962, *Numerical Methods for Scientists and Engineers*.

- Insight into data ? It has two meanings
  - Summary: questions leading to further questions $\Rightarrow$ deeper understanding.
  - Exploration: finding out novel facts and unexpected correlations.

- Insight into data ? It has two meanings
  - Summary: questions leading to further questions $\Rightarrow$ deeper understanding.
  - Exploration: finding out novel facts and unexpected correlations.
- Insight into data ? It has two complementary scenarios:
  - Precise to general questions answered in a bottom-up exploratory, open-ended manner.
  - General to precise questions answered in a top-down "overview, zoom, and details-on-demand"* manner.

    * attributed to [Shneiderman 1996] and is popularly known as the Visual Information Seeking Mantra.

[Shneiderman 1996] Ben Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages, pages 336-343, Washington. IEEE Computer Society Press, 1996.

Expectations from a Visualization

- What type of insight to expect ?
- What not to expect ?
- How much insight gained ?
- How to construct a visualization that provides the best insight ?

Expectations from a Visualization

- What type of insight to expect ?
- What not to expect ?
- How much insight gained ?
- How to construct a visualization that provides the best insight ?

Visualization largely needs a human-in-the-loop; for qualitative outcomes of the visualization applications.

- Quantative analysis maybe obtained automatically.

What exactly is visualization ?

- A graphical representation of data.

What exactly is visualization ?

- A graphical representation of data.

Is it anything more ? Give it a thought ....

What exactly is visualization ?

- A graphical representation of data.

Is it anything more ? Give it a thought ....

- Inherently a dimensionality reduction method – projection methods could be linear or nonlinear – finally a 2-d image.
- Data modeling needed to enable to the dimensionality reduction.
- Data modeling has to be precise, appropriate, intuitive, efficient.

What exactly is visualization ?

- A graphical representation of data.

Is it anything more ? Give it a thought ....

- Inherently a dimensionality reduction method – projection methods could be linear or nonlinear – finally a 2-d image.
- Data modeling needed to enable to the dimensionality reduction.
- Data modeling has to be precise, appropriate, intuitive, efficient.

Data representation using modeling is key.
Hence modeling becomes critical.

Data types:

- Continuous/sampled vs. discrete data
- Gridded vs. scattered data
  - For gridded data: various cell types giving uniform, rectilinear, regular/structured, vs. irregular/unstructured data.

Data types:

- Continuous/sampled vs. discrete data
- Gridded vs. scattered data
  - For gridded data: various cell types giving uniform, rectilinear, regular/structured, vs. irregular/unstructured data.

Processes:

- Aquisition, modeling, reconstruction.

- Multidimensional vs multivariate data: used interchangeably.
  - Right usage should be multidimensional multivariate data (mdmv data) where dimensions are linearly independent variables and attributes are dependent variables.

    Reference: Pak Chung Wong and R. Daniel Bergeron. 1994. 30 Years of Multidimensional Multivariate Visualization. In Scientific Visualization, Overviews, Methodologies, and Techniques, Gregory M. Nielson, Hans Hagen, and Heinrich Müller (Eds.). IEEE Computer Society, Washington, DC, USA, 3-33.
- Projecting from higher to lower dimensions may be achieved mathematically, e.g. using subspace projection.

- Multidimensional vs multivariate data: used interchangeably.
  - Right usage should be multidimensional multivariate data (mdmv data) where dimensions are linearly independent variables and attributes are dependent variables.

    Reference: Pak Chung Wong and R. Daniel Bergeron. 1994. 30 Years of Multidimensional Multivariate Visualization. In Scientific Visualization, Overviews, Methodologies, and Techniques, Gregory M. Nielson, Hans Hagen, and Heinrich Müller (Eds.). IEEE Computer Society, Washington, DC, USA, 3-33.
- Projecting from higher to lower dimensions may be achieved mathematically, e.g. using subspace projection.
- However, extending concepts from lower to higher dimensions is non-trivial and non-intuitive.

4-Circle Paradox

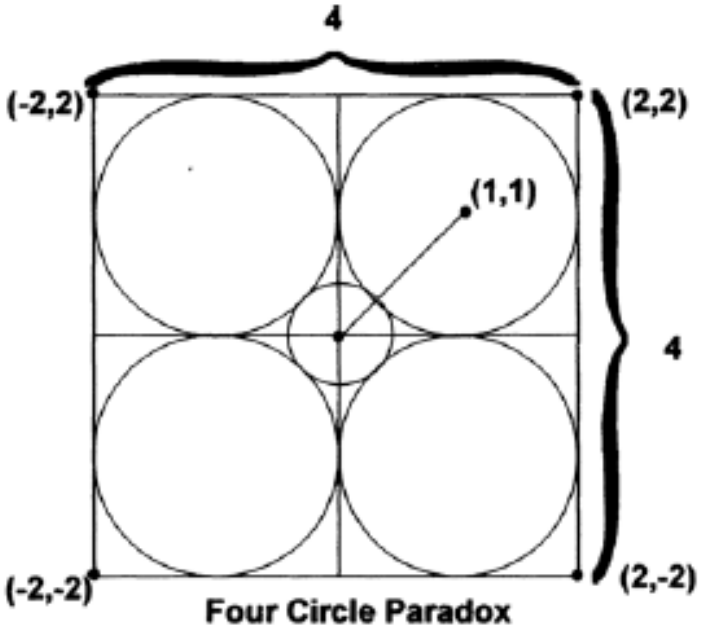


Image courtesy: Hamming, Richard (1997). "The Art of Doing Science and Engineering: Learning to Learn", Taylor & Francis e-Library, 2005.

In 2-d, radius of inner circle, $r_2 = \sqrt{2} - 1$, similarly in 3-d, it is $n_3 = \sqrt{3} - 1$; extending to n-dimensions, radius $r_n = \sqrt{n} - 1$.

4-Circle Paradox

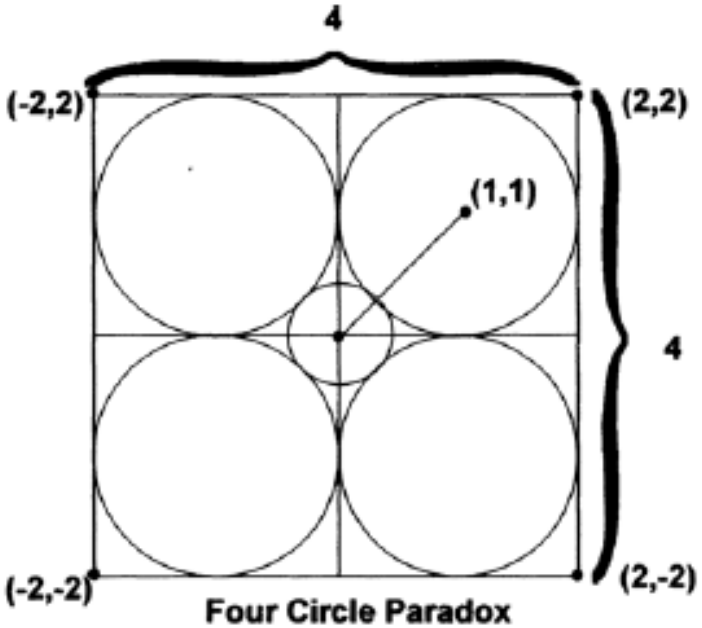


Image courtesy: Hamming, Richard (1997). "The Art of Doing Science and Engineering: Learning to Learn", Taylor & Francis e-Library, 2005.

In 2-d, radius of inner circle, $r_2 = \sqrt{2} - 1$, similarly in 3-d, it is $n_3 = \sqrt{3} - 1$; extending to n-dimensions, radius $r_n = \sqrt{n} - 1$.

However, for $n \geq 10$, inner radius $r_n > 2 \Rightarrow$ **violates the boundary condition of half length of bounding cube in any dimension being 2.**