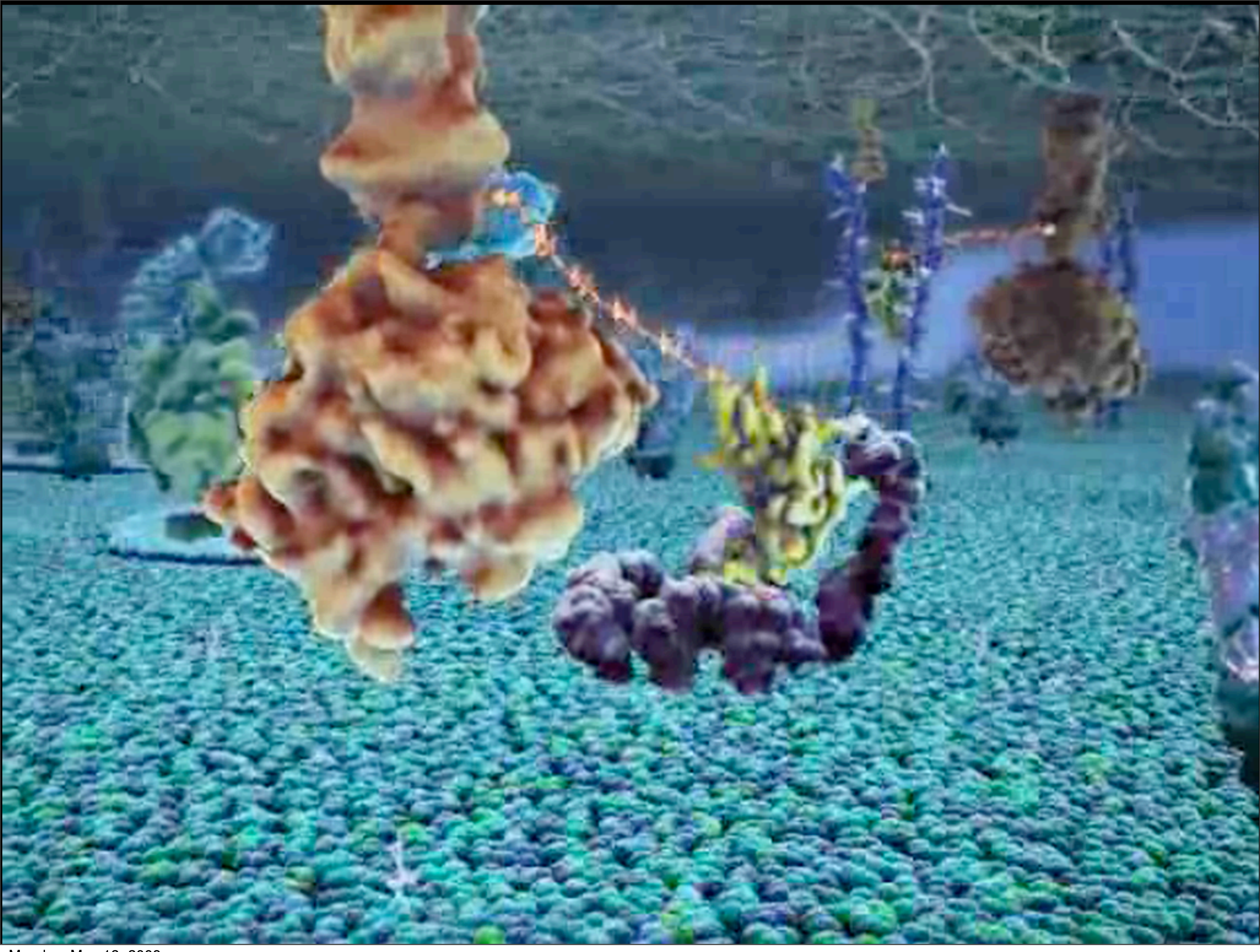
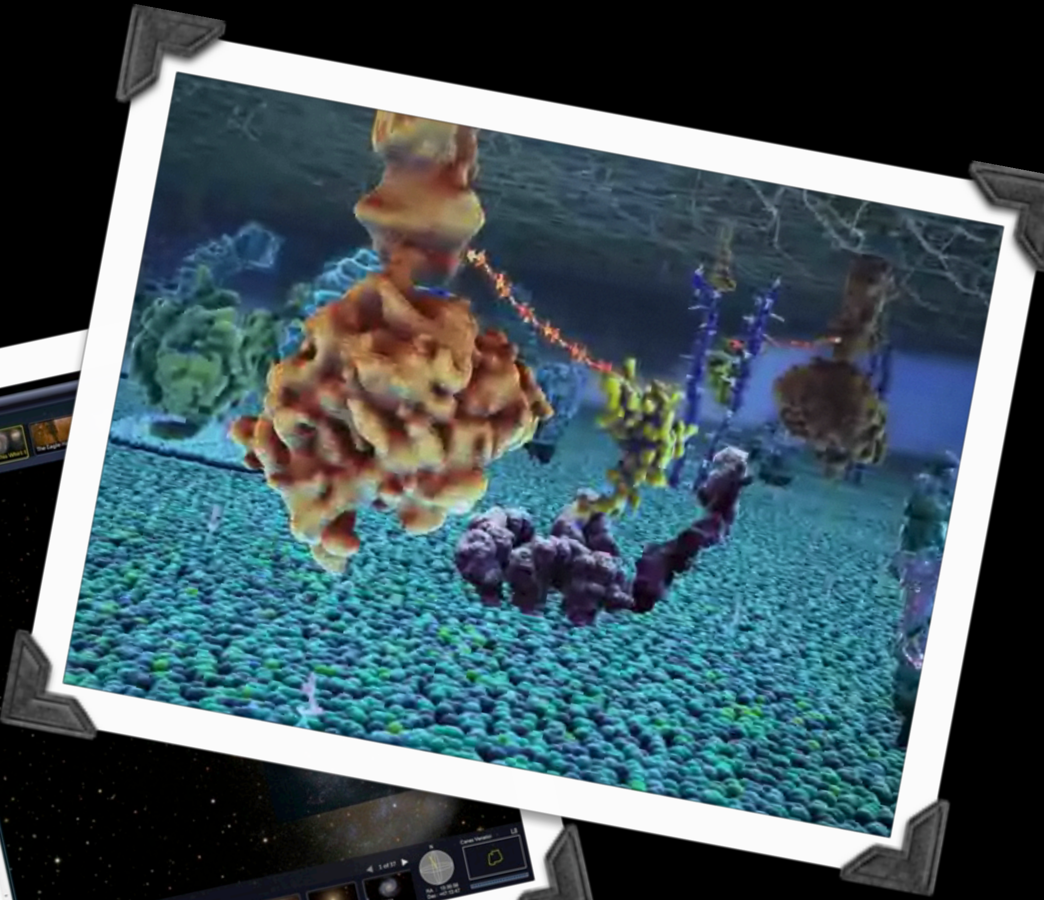


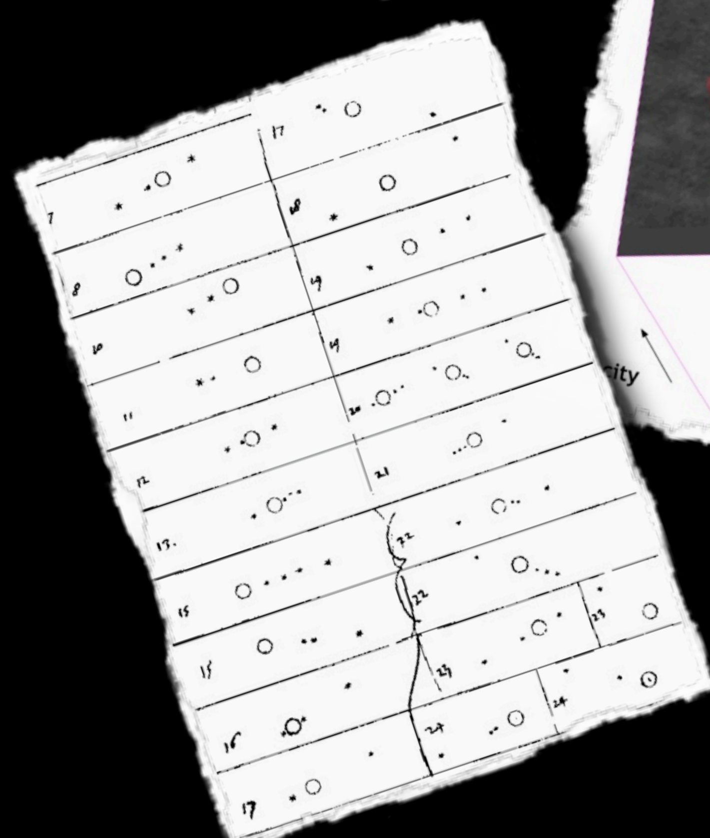
Seeing Science

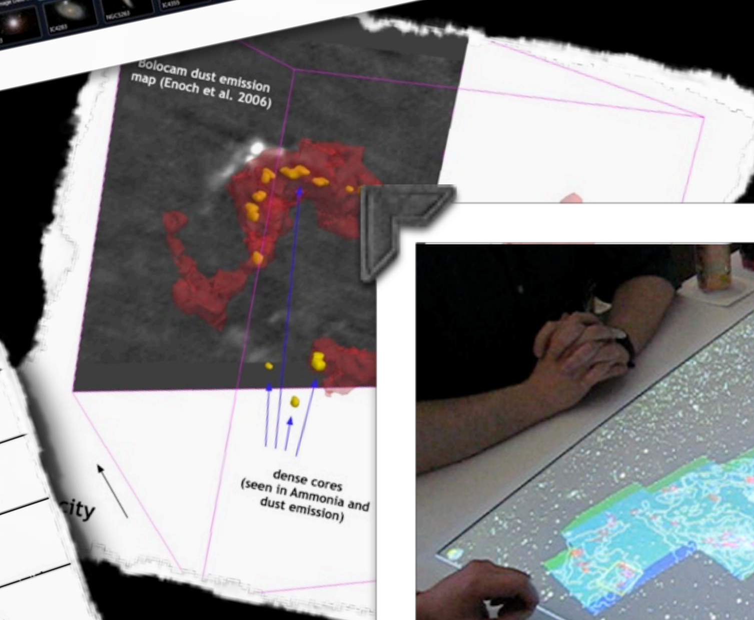
Alyssa A. Goodman

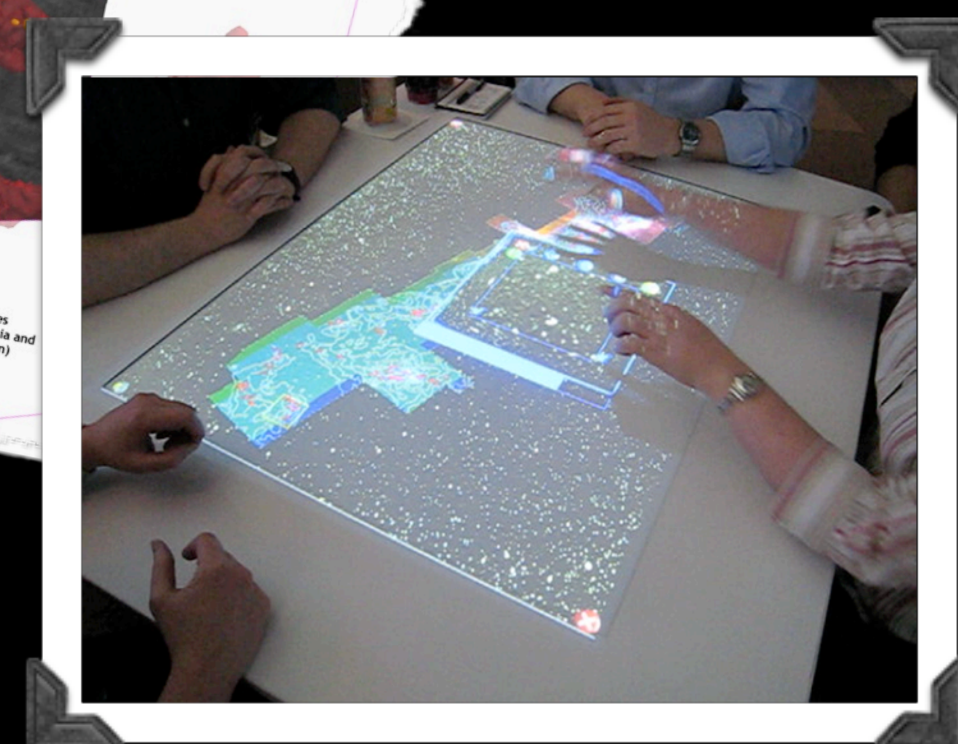
Professor of Astronomy
& Founding Director of the
Initiative in Innovative Computing,
Harvard University

Scholar-in-Residence,
WGBH Boston







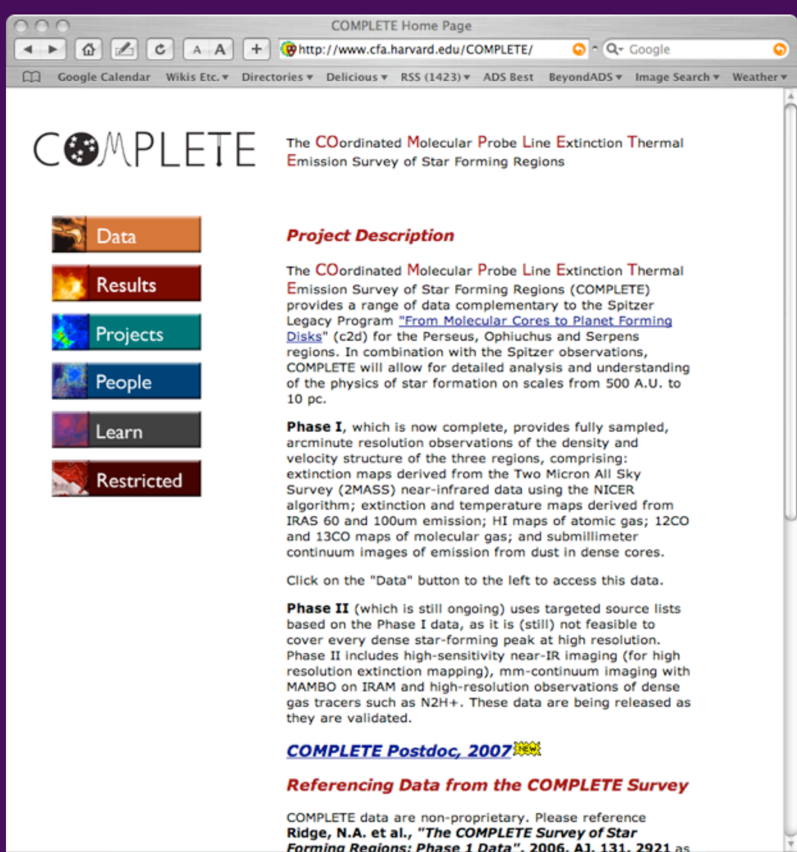


Who am I?

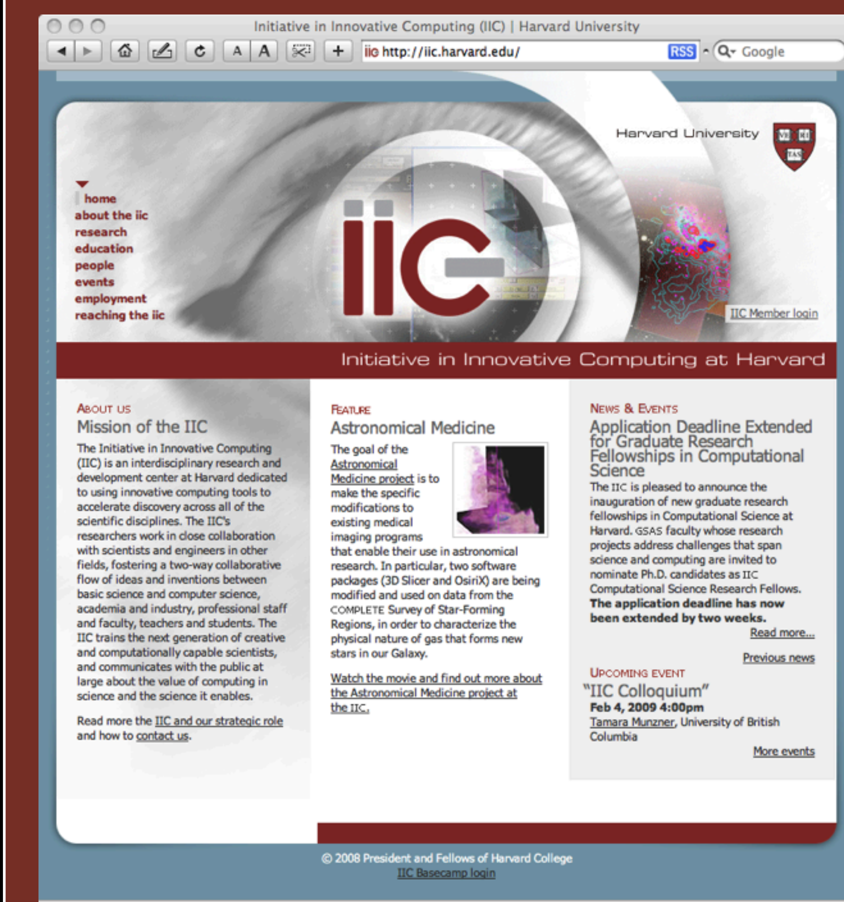
...my interest(s) in "Seeing Science"...

Art & Photography

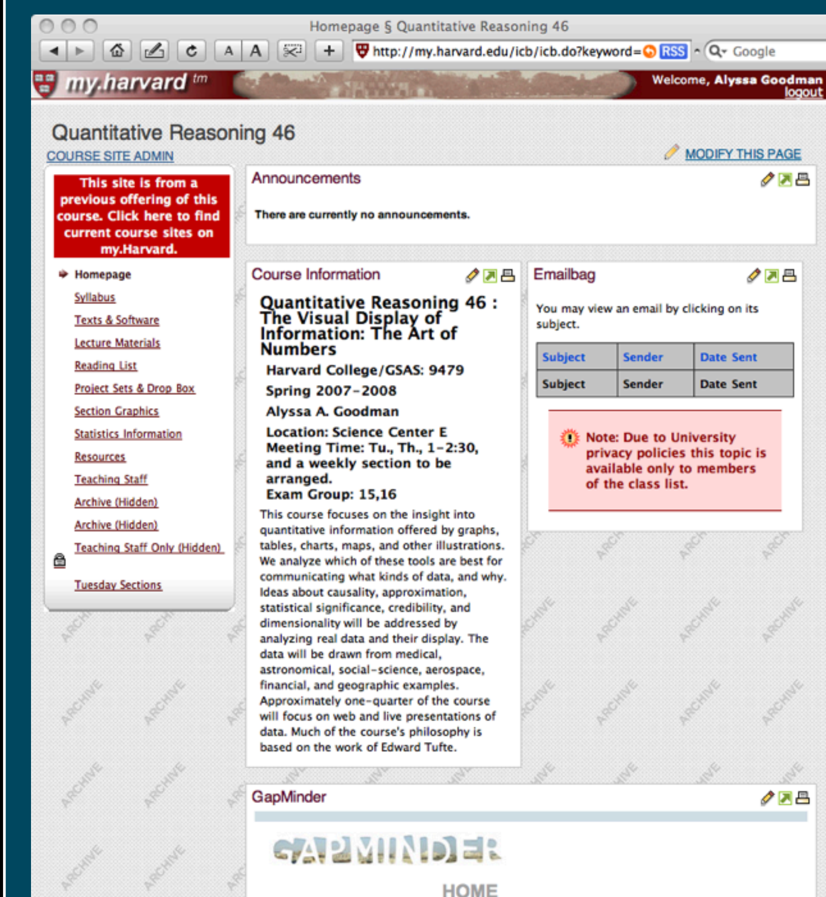
Astronomy



Scientific Computing



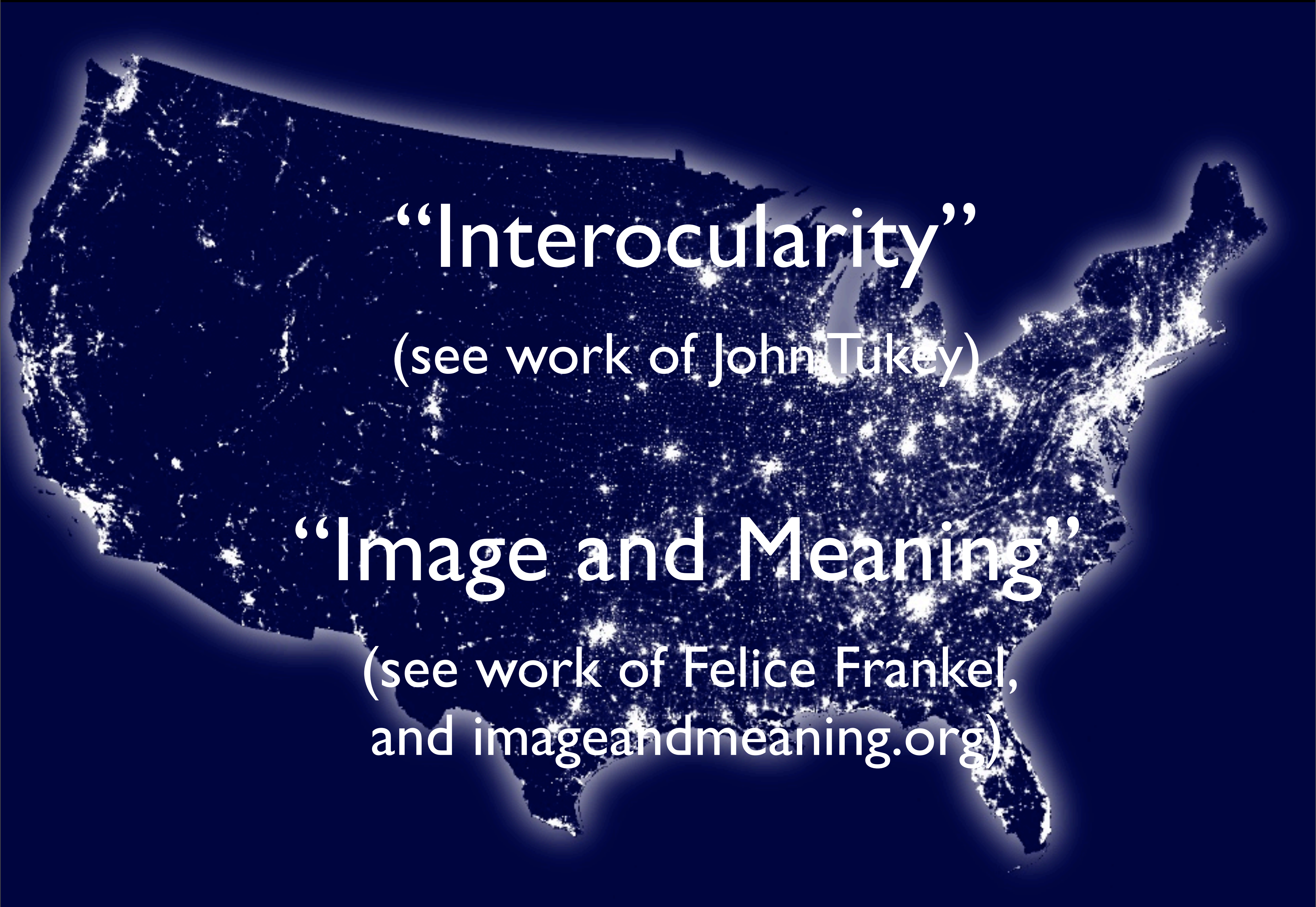
Teaching



Relative Strengths

Pattern Recognition
Creativity

Calculations



Monday, May 18, 2009

Seeing Science

Data • Dimensions • Display

What...

...is easier now than before?

fast computation, animation, 3D

...was easier before than now?

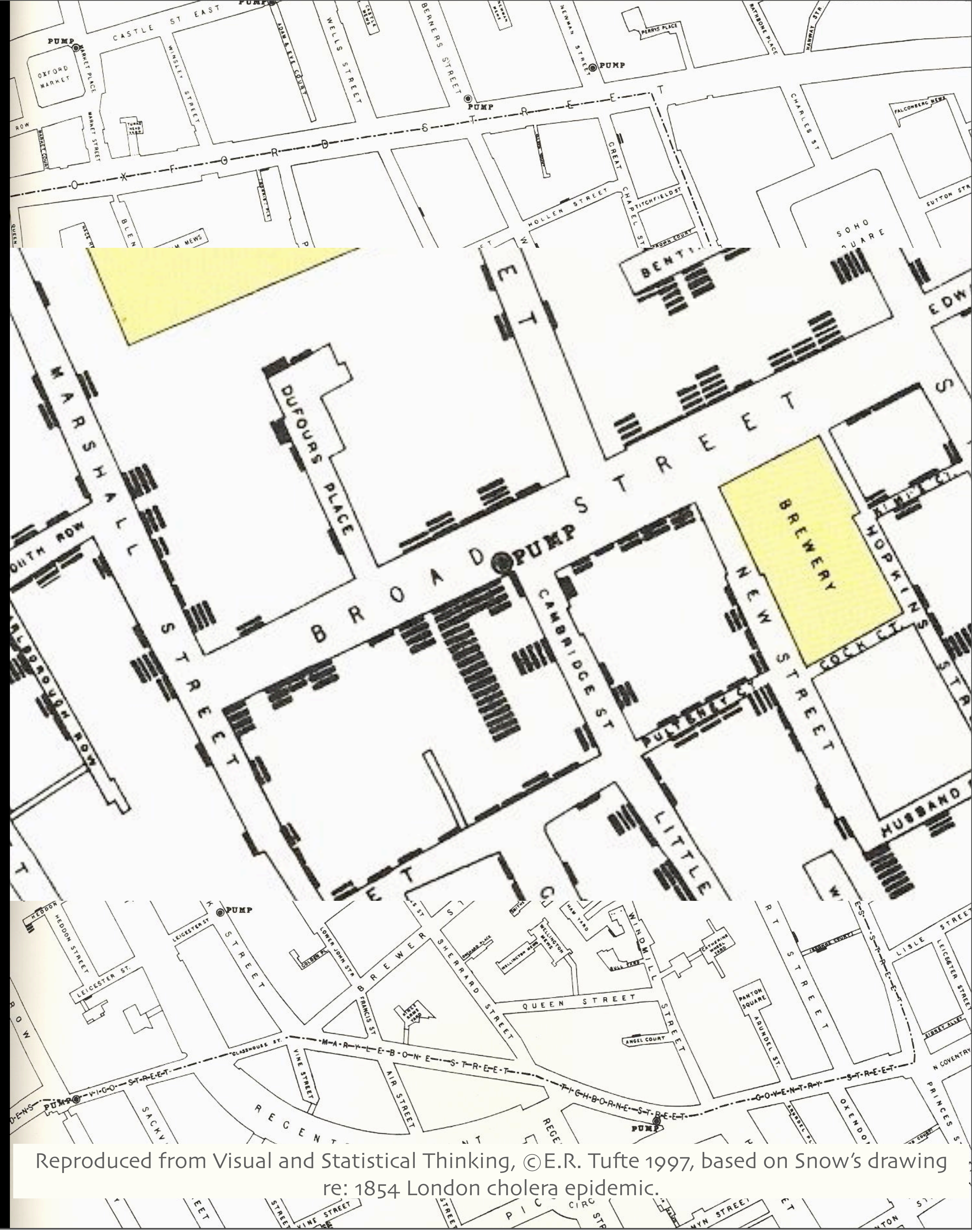
craftsmanship

...should be easier in the future?

modular craftsmanship

Craftsmanship
(in 1854)

Displaying
"high-dimensional" data
with
"multi-functioning
graphical elements"

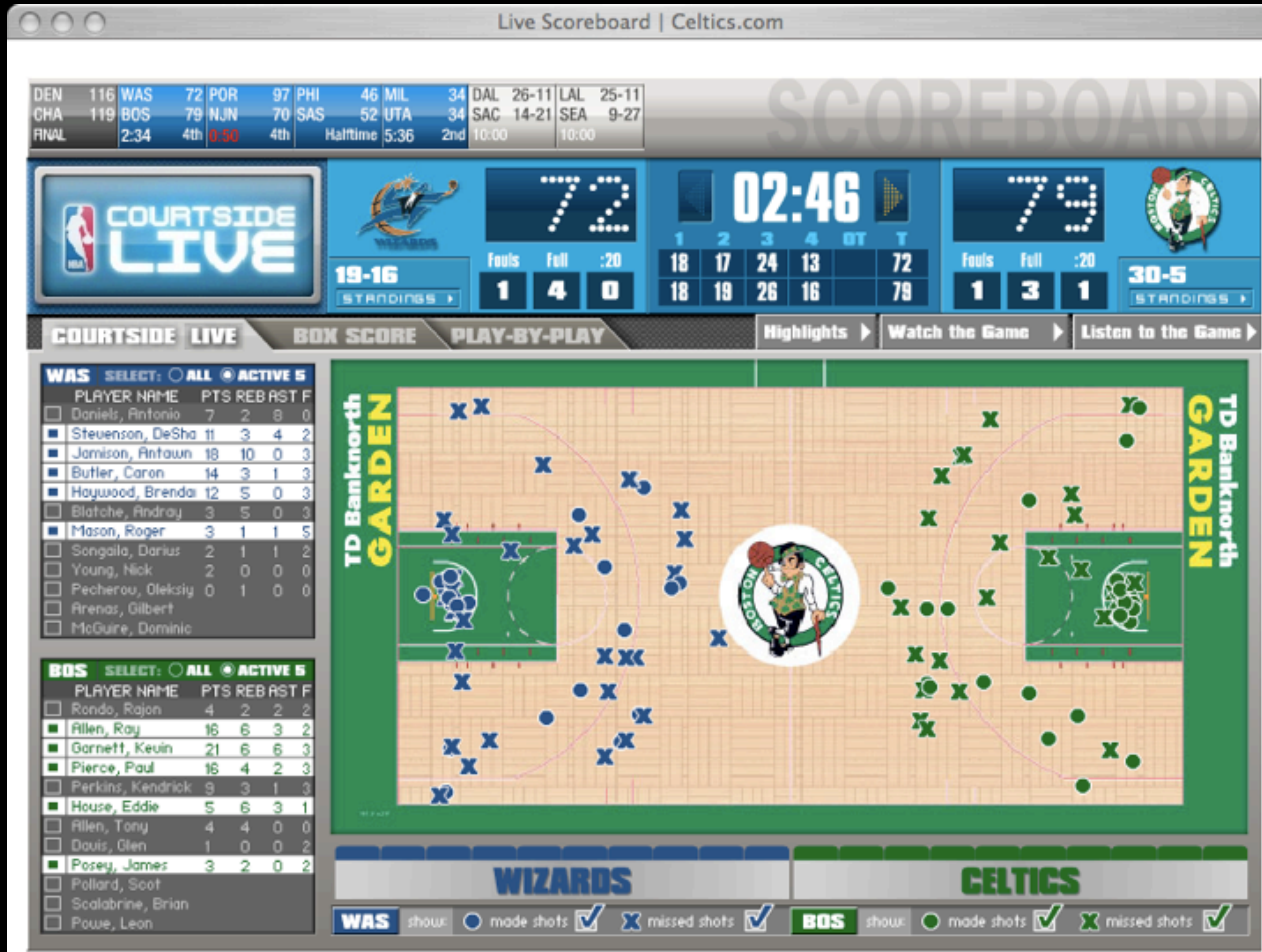


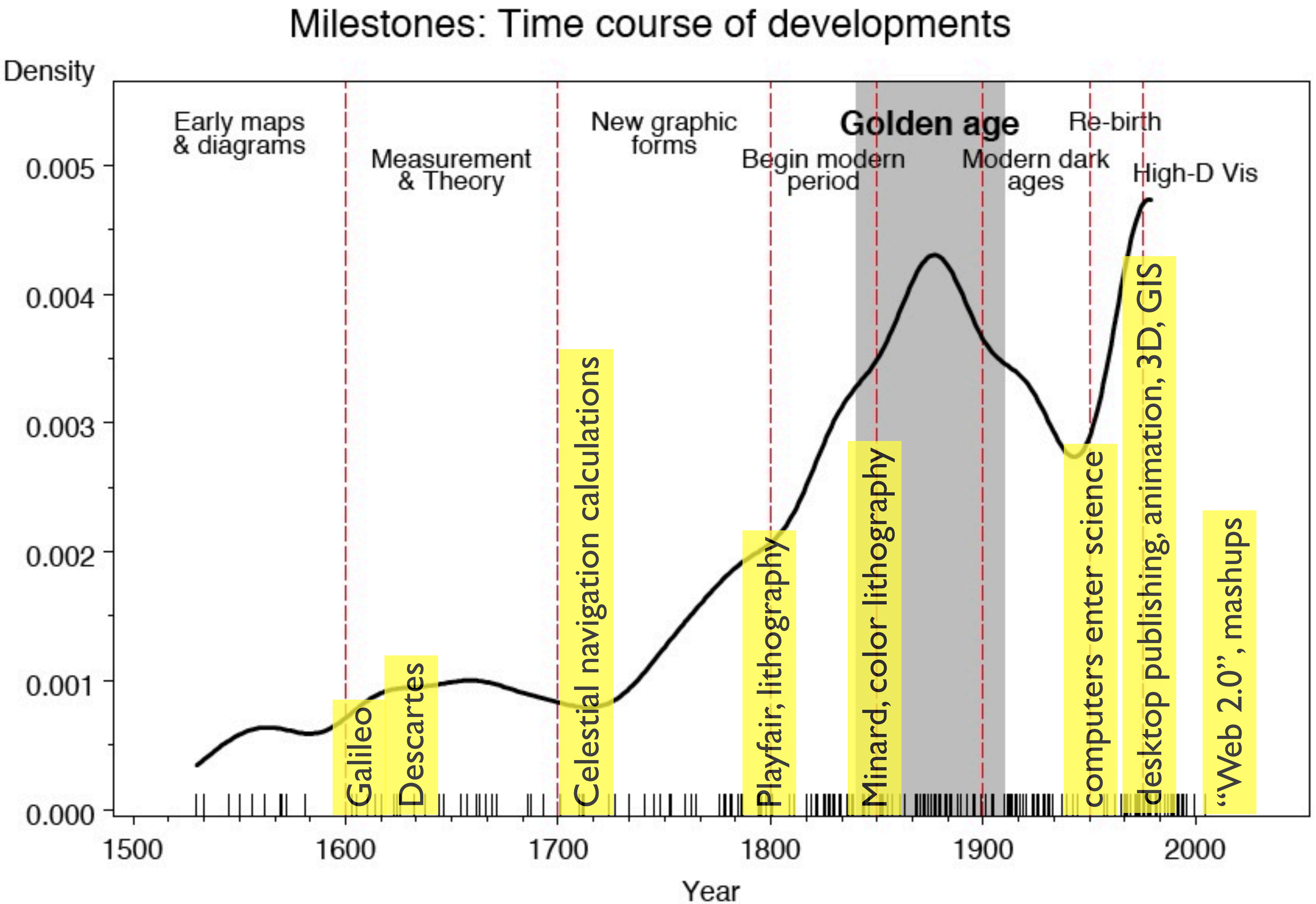
Reproduced from Visual and Statistical Thinking, ©E.R. Tufte 1997, based on Snow's drawing re: 1854 London cholera epidemic.

What Computers *Can* Let us Craft

Elements...

- ✓Maps
- ✓Tables
- ✖Graphs
- ✓Charts
- ✓Illustrations
- ✓Combinations





adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

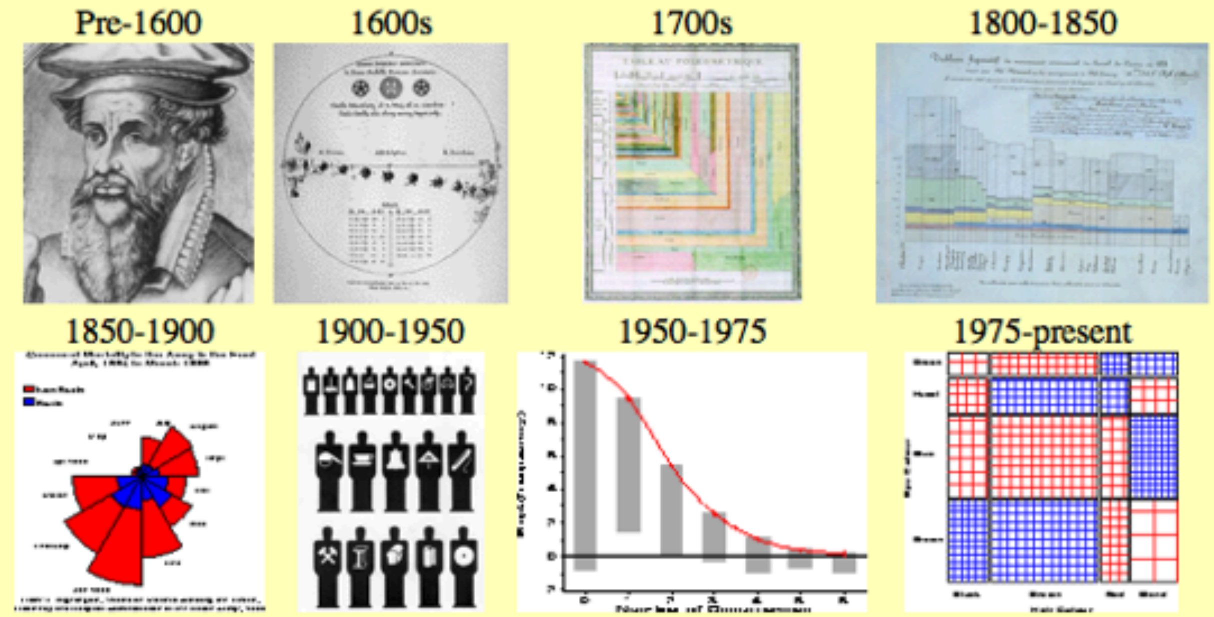
The Milestones Project (Friendly & Denis)

Milestones in the history of thematic cartography, statistical graphics, and data visualization

http://www.math.yorku.ca/SCS/Gallery/milestone/index.html

Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization

An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis



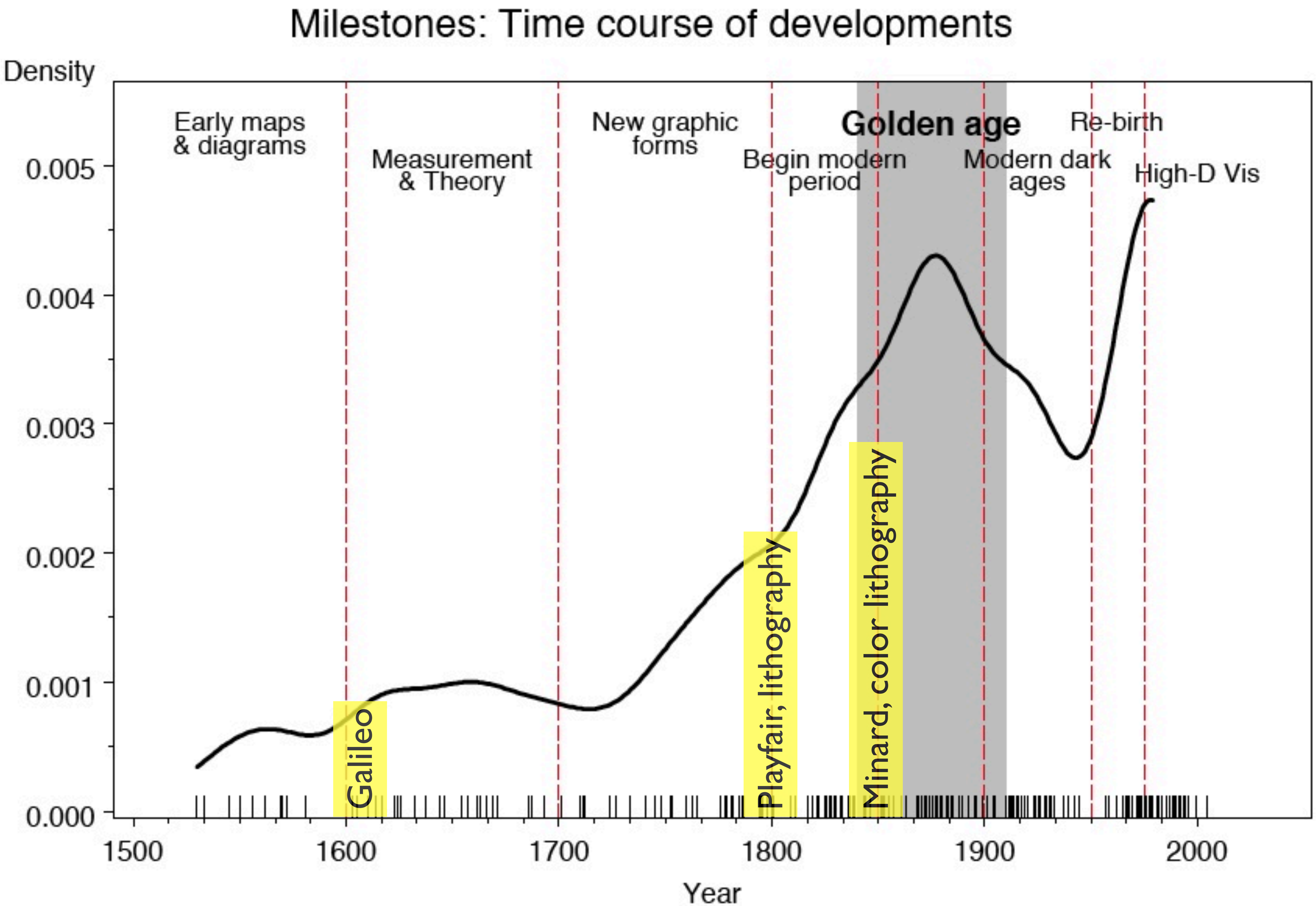
Up: Gallery | Introduction | Related | References | Term Index | Category XRef | Search

Pre-1600 | 1600s | 1700s | 1800+ | 1850+ | 1900+ | 1950+ | 1975+

Related information:

- **Citations**: References to information or images obtained from this web site should be cited as follows: Friendly, M. & Denis, D. J. (2001). *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. Web document, `http://www.math.yorku.ca/SCS/Gallery/milestone/`. Accessed: Thu, 16 Oct 2008 15:28:13 GMT
- **PDF version**: This document in PDF form, with active links.
- **Other publications**:
  - Chapter on the Milestone Project in C. Weihs and W. Gaul (eds.), *Classification-- The Ubiquitous Challenge*, *Handbook of Computational Statistics: Data Visualization*, Springer, 2005.
  - A Brief History of Data Visualization in C. Chen, W. Hardle and A. Unwin (eds.), *Handbook of Computational Statistics: Data Visualization*, Vol III, Springer, 2007.
- **Visual brochures** designed by Mario Kanno
  - Milestones in the History of Data Visualization
  - Marcos na História da Visualização de Dados(Brazillian Portuguese)
- **Images** from the JSM 2002 Technical Poster Session [Thanks to Andy Mauromoustakos!]:
  - Image1 (864 x 648; 123K);
  - Image2 (864 x 648; 124K).

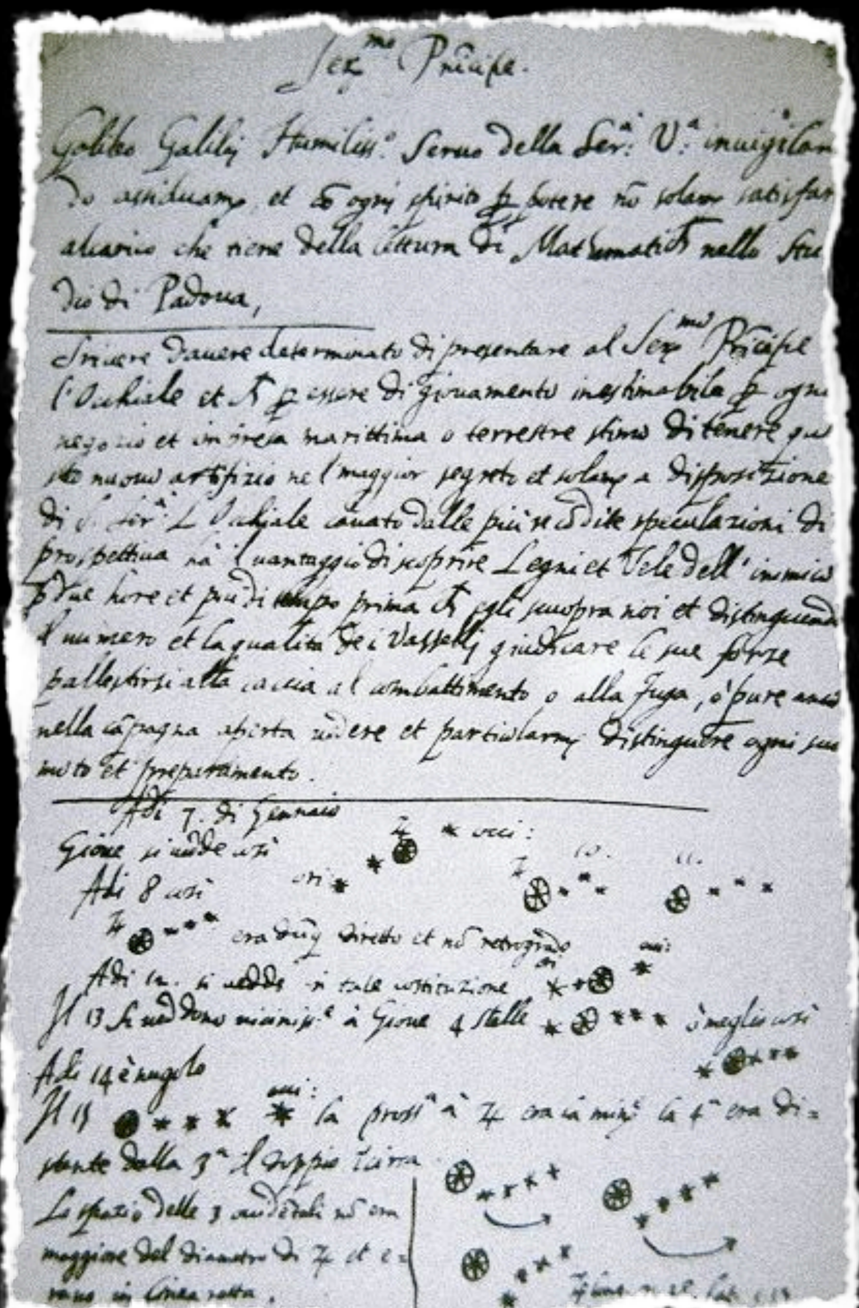
This web version is dedicated to Arthur H. Robinson (1915-2004), who inspired and encouraged our interest; to Antoine de Falguerolles, who initiated it, and to *les Chevaliers des Album de Statistique Graphique*, who supported it with interest, enthusiasm, and resources. In particular, Gilles Palsky, Antoine de Falguerolles, Antony Unwin and Ruddy Ostermann contributed important images and background information. This work is supported by the National Sciences and Engineering Research Council of Canada, Grant OGP0138748.

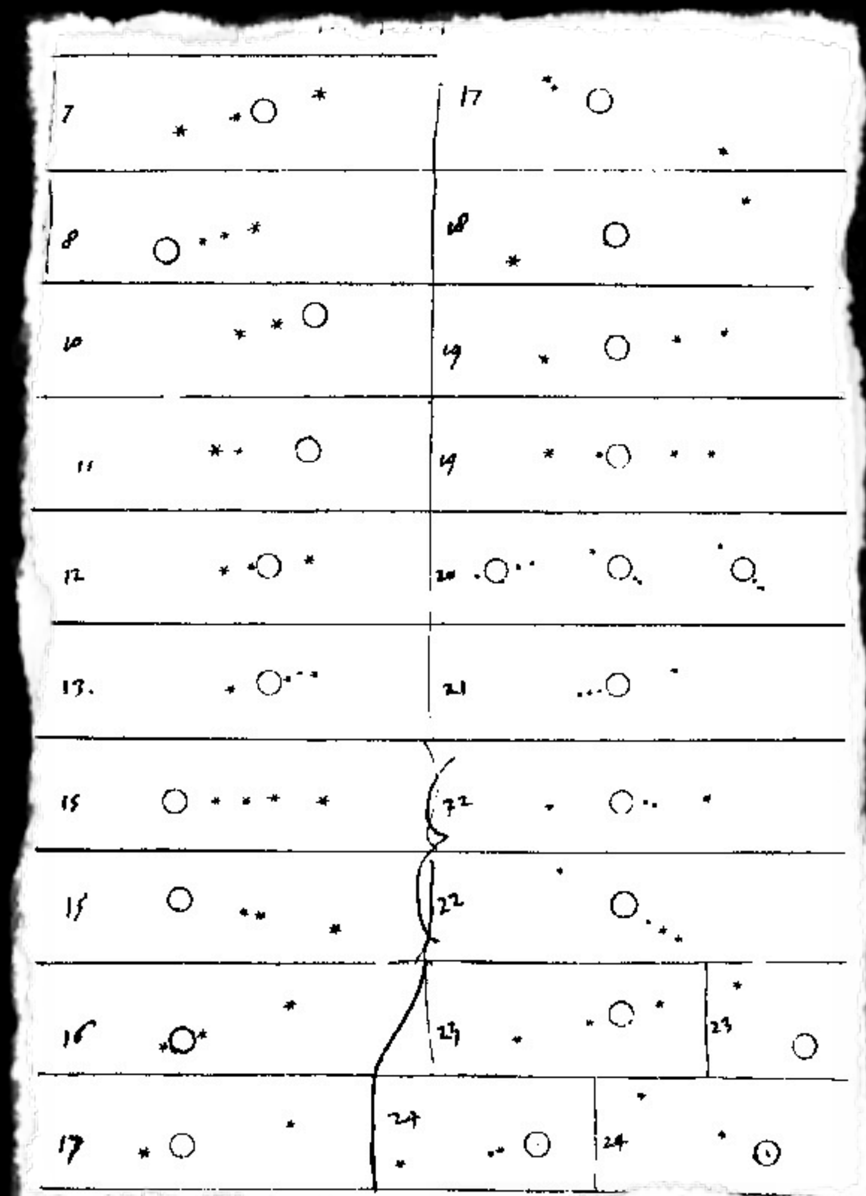


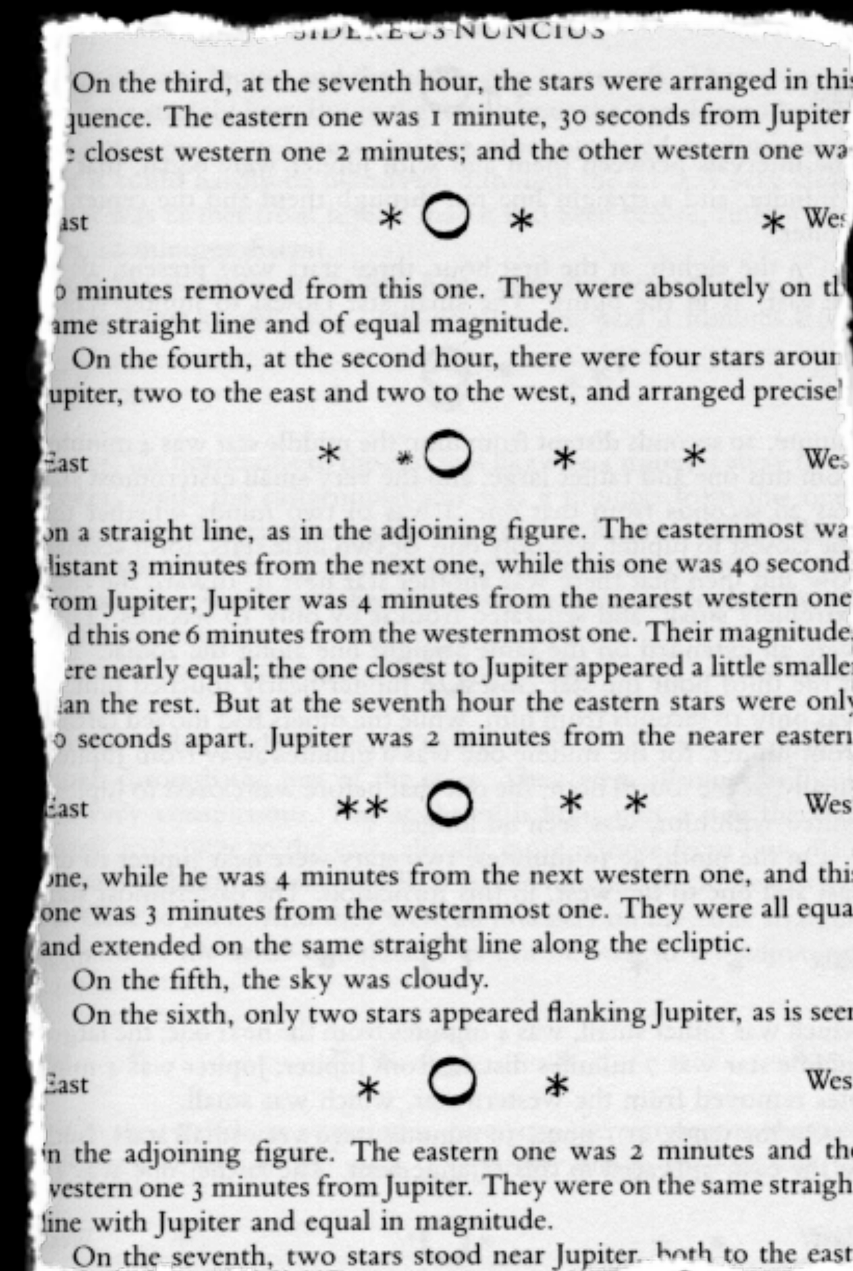
adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

Galileo

Galileo Galilei
(1564-1642)






SIDEREUS NUNCIUS

On the third, at the seventh hour, the stars were arranged in this
quence. The eastern one was 1 minute, 30 seconds from Jupiter
closest western one 2 minutes; and the other western one wa

o minutes removed from this one. They were absolutely on th
ame straight line and of equal magnitude.
On the fourth, at the second hour, there were four stars arour
upiter, two to the east and two to the west, and arranged precise

on a straight line, as in the adjoining figure. The easternmost wa
listant 3 minutes from the next one, while this one was 40 second
rom Jupiter; Jupiter was 4 minutes from the nearest western one
d this one 6 minutes from the westernmost one. Their magnitude
ere nearly equal; the one closest to Jupiter appeared a little smalle
an the rest. But at the seventh hour the eastern stars were only
o seconds apart. Jupiter was 2 minutes from the nearer easter

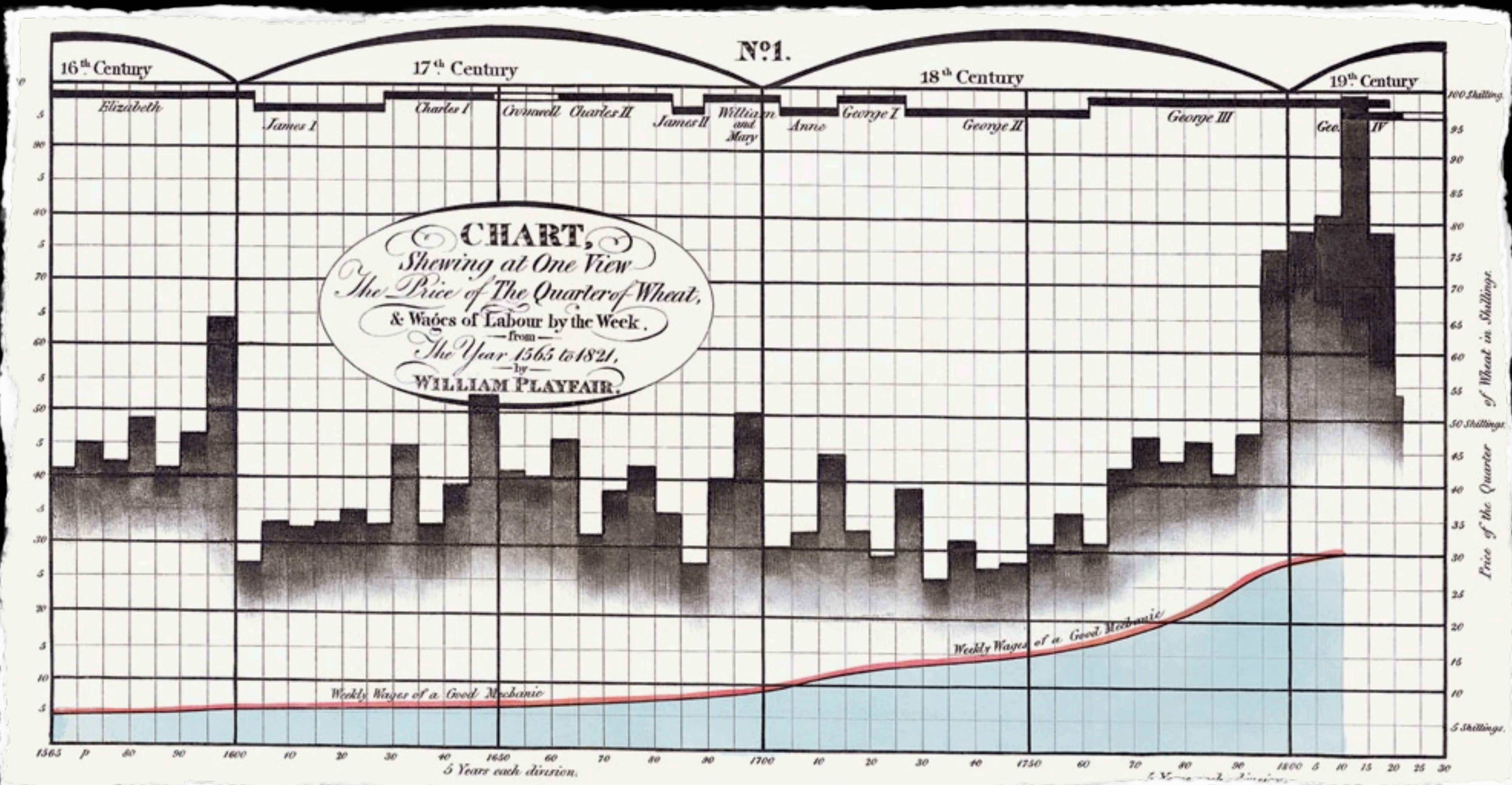
one, while he was 4 minutes from the next western one, and this
one was 3 minutes from the westernmost one. They were all equal
and extended on the same straight line along the ecliptic.
On the fifth, the sky was cloudy.
On the sixth, only two stars appeared flanking Jupiter, as is seen

in the adjoining figure. The eastern one was 2 minutes and the
western one 3 minutes from Jupiter. They were on the same straight
line with Jupiter and equal in magnitude.
On the seventh, two stars stood near Jupiter, both to the east

Notes for & re-productions of Siderius Nuncius

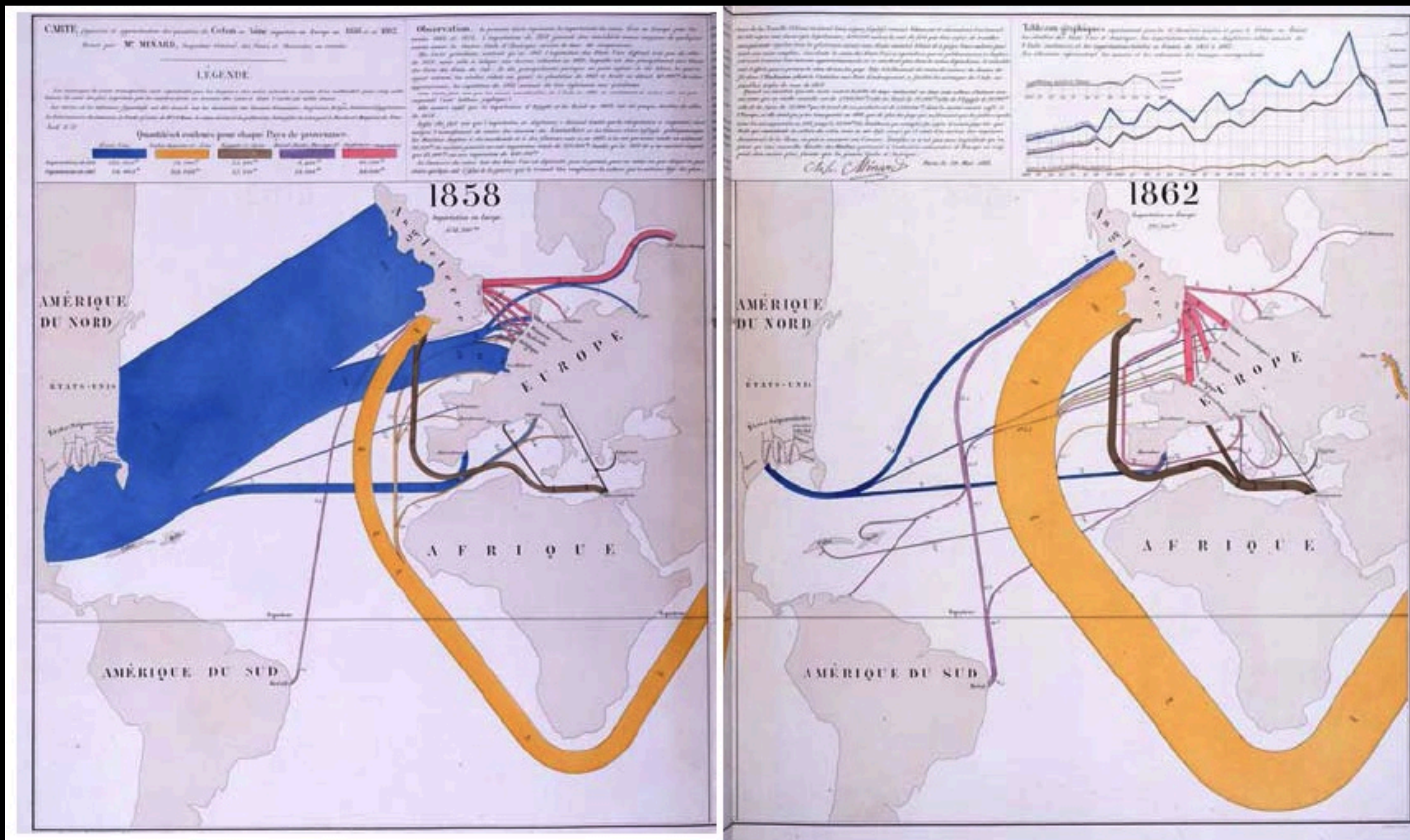
Playfair, lithography

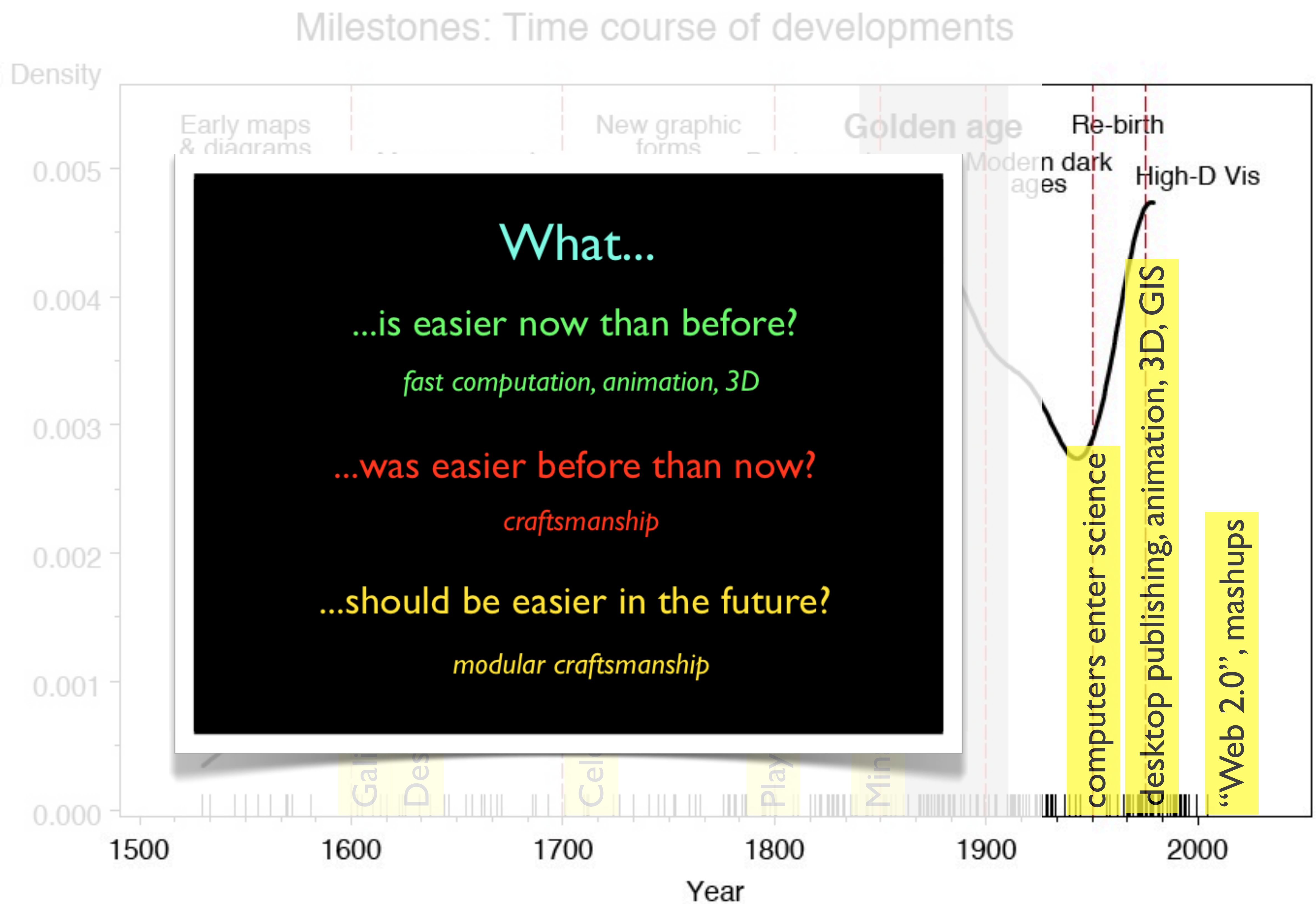
William Playfair
(1759-1823)



Minard, color lithography

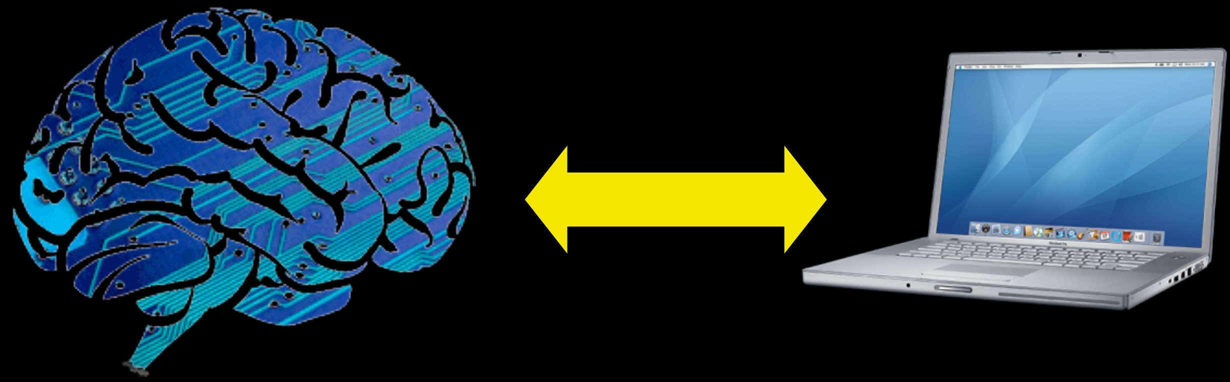
Charles Joseph Minard, in color
(1781-1870)





adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

computers enter science



Data Reduction

Data Display

Context (e.g. journals + online data)

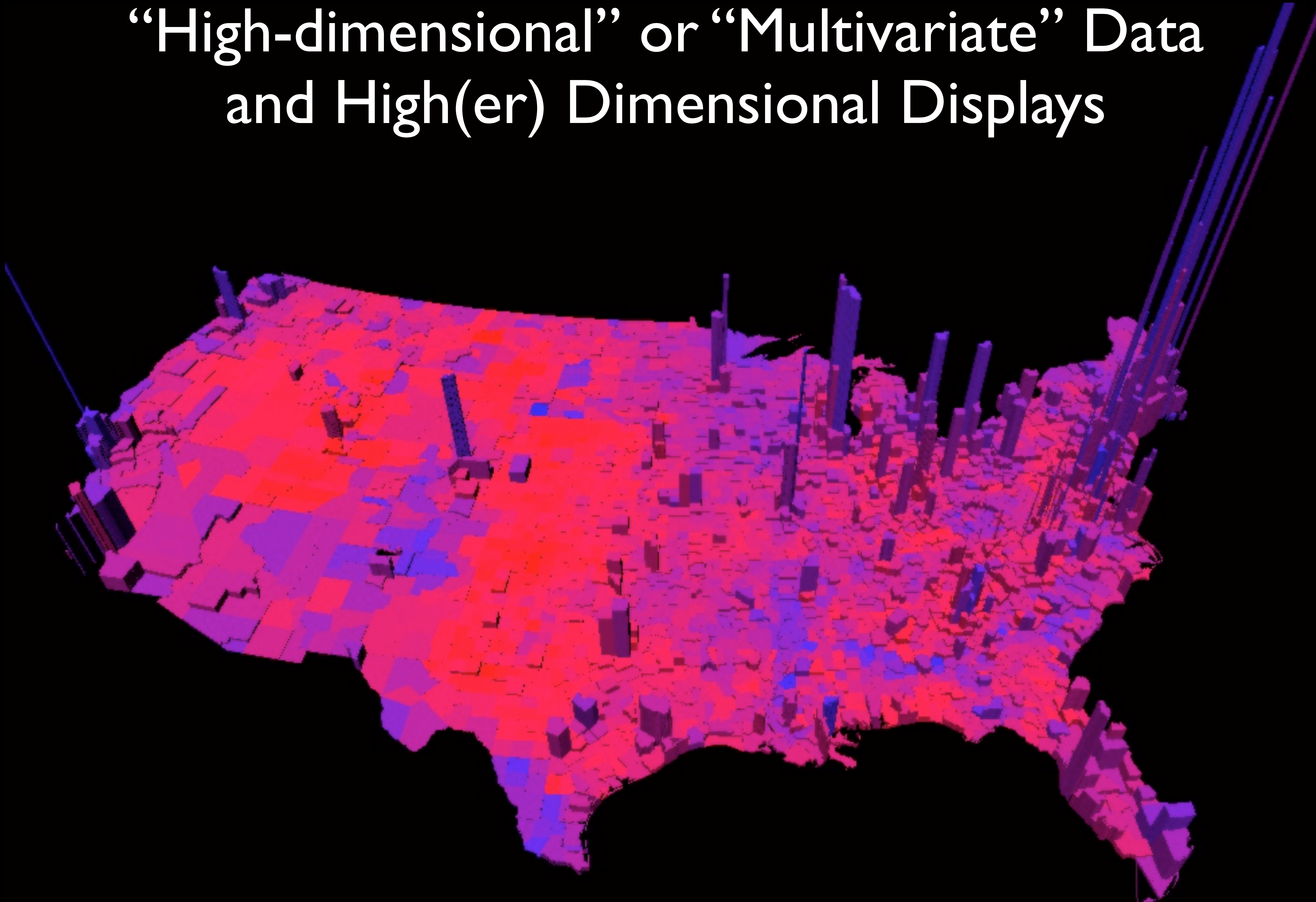
Simulation Design

Statistics Design

Data Exploration (Visualization)

Data • Dimensions • Display

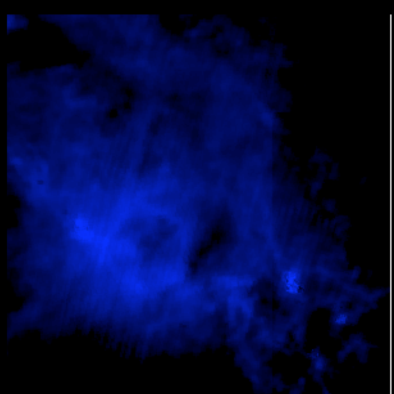
"High-dimensional" or "Multivariate" Data and High(er) Dimensional Displays



*This map **displays** 2 quantities as a function of 2 spatial dimensions.*
...Is that 4 dimensions?

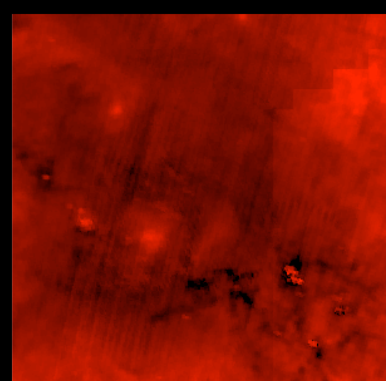
Note: Computers make this Easy

Column Density

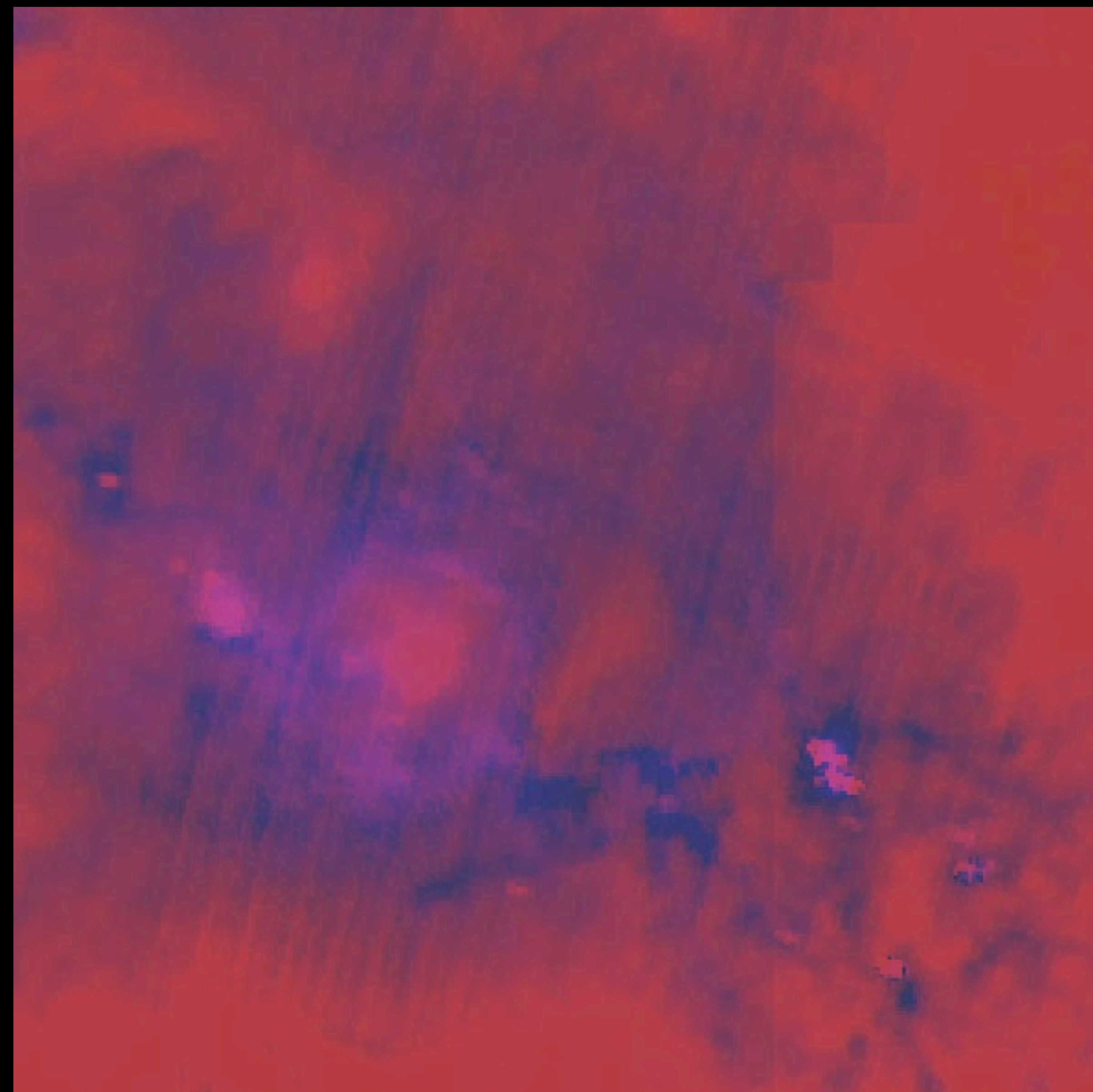


+

Temperature

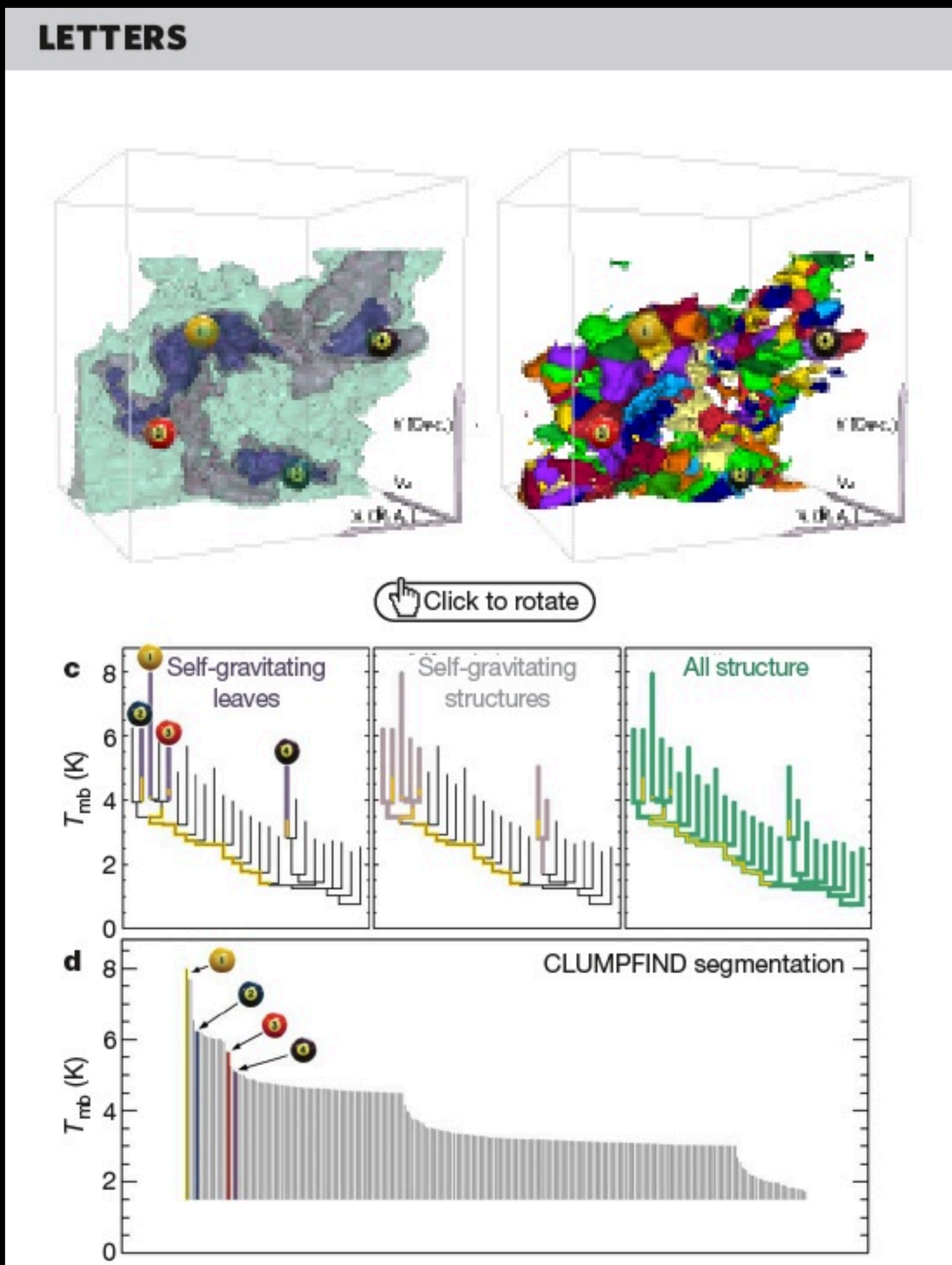


=

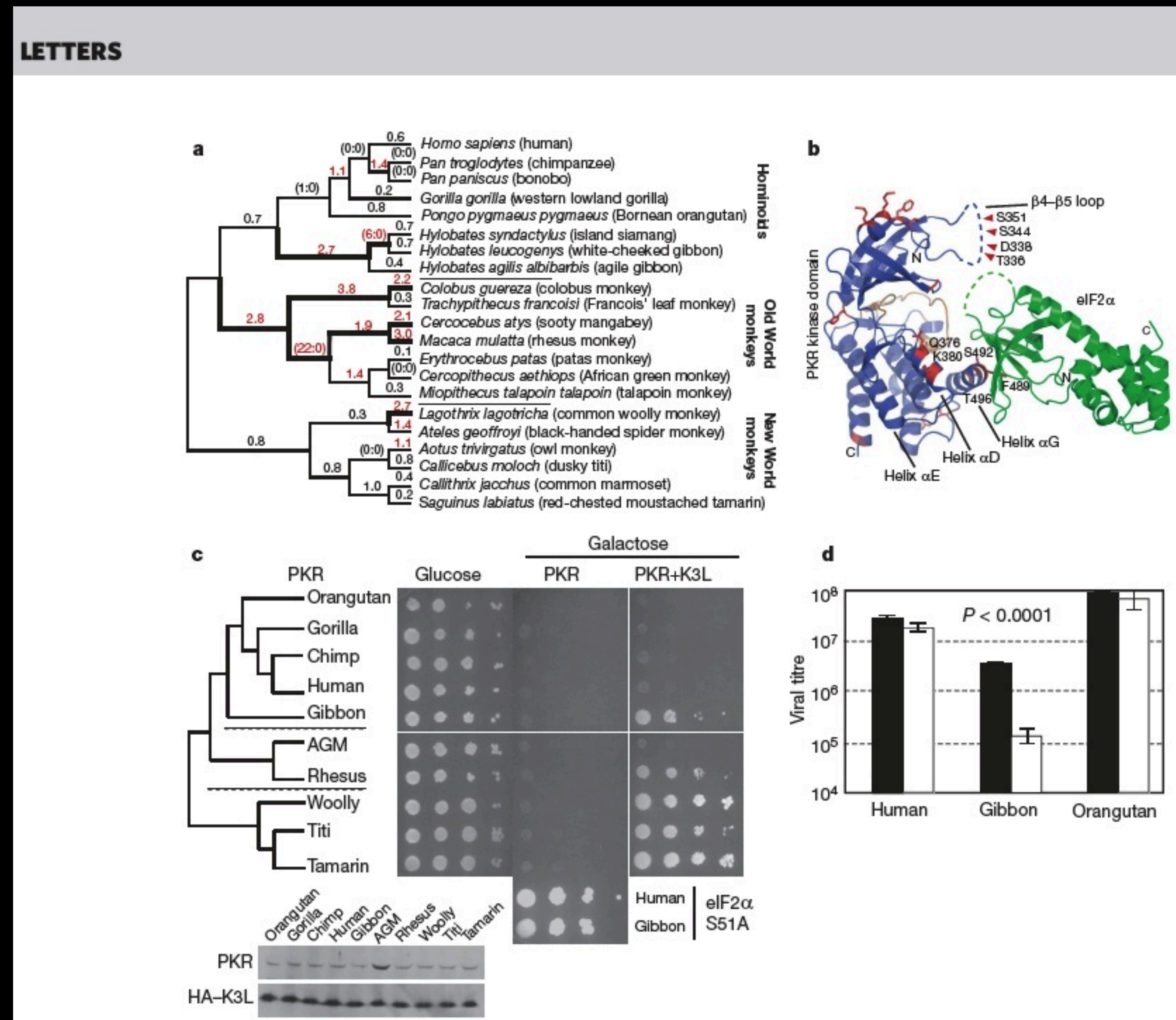


“High-dimensional” or “Multivariate” Data

(Astronomy=Biology)



Goodman et al. *Nature*, 2009



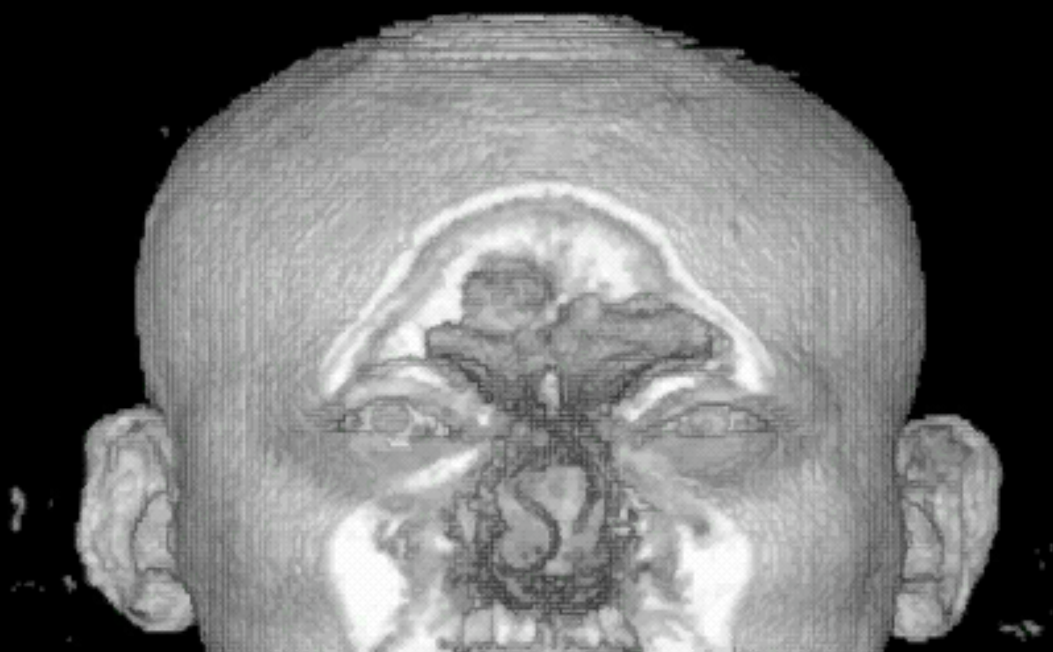
Elde et al. *Nature*, 2008

What about Animation?

How many dimensions at once?

Can/should time (animation) substitute for dimensions?

"KEITH"



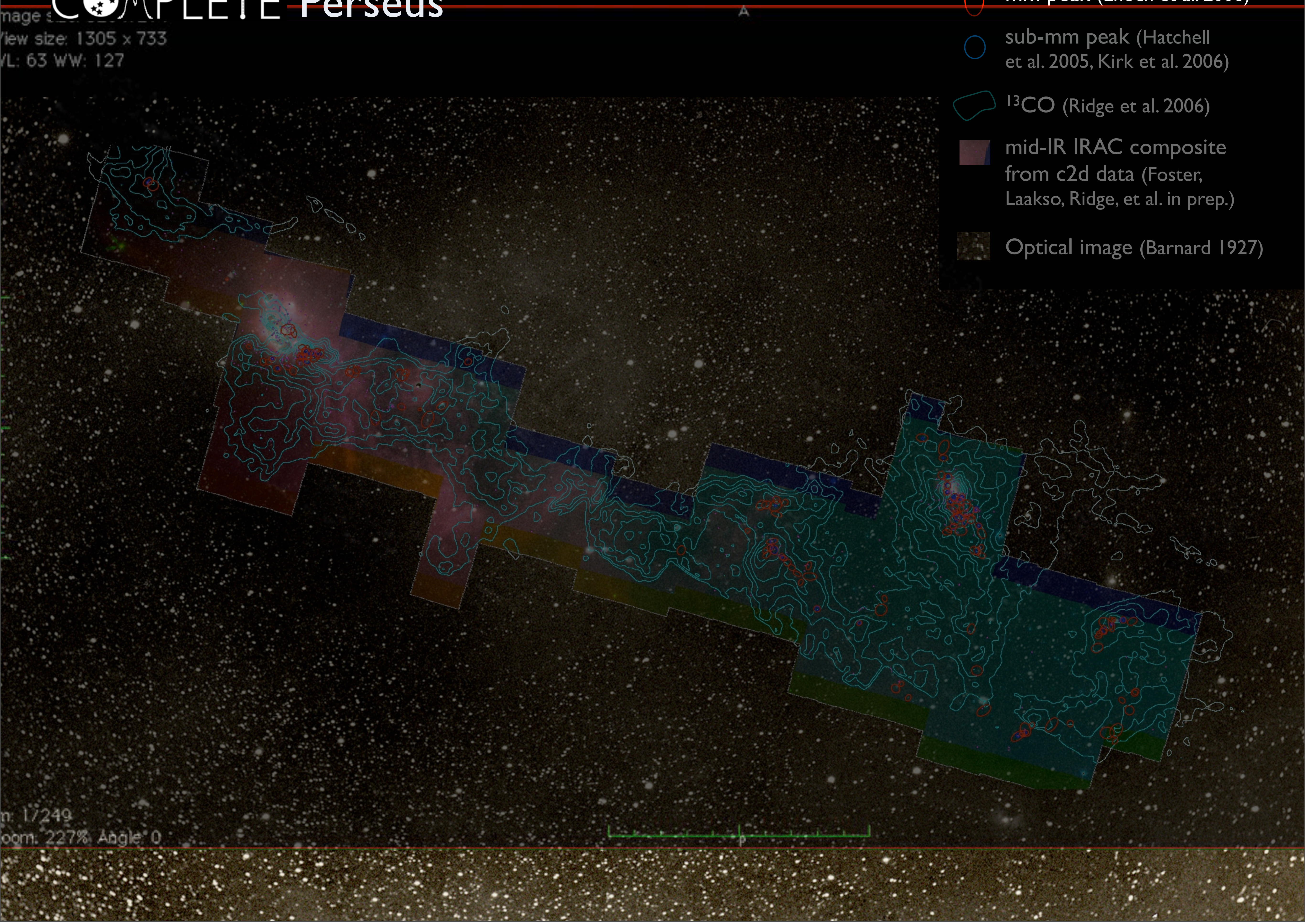
"z" is depth into head

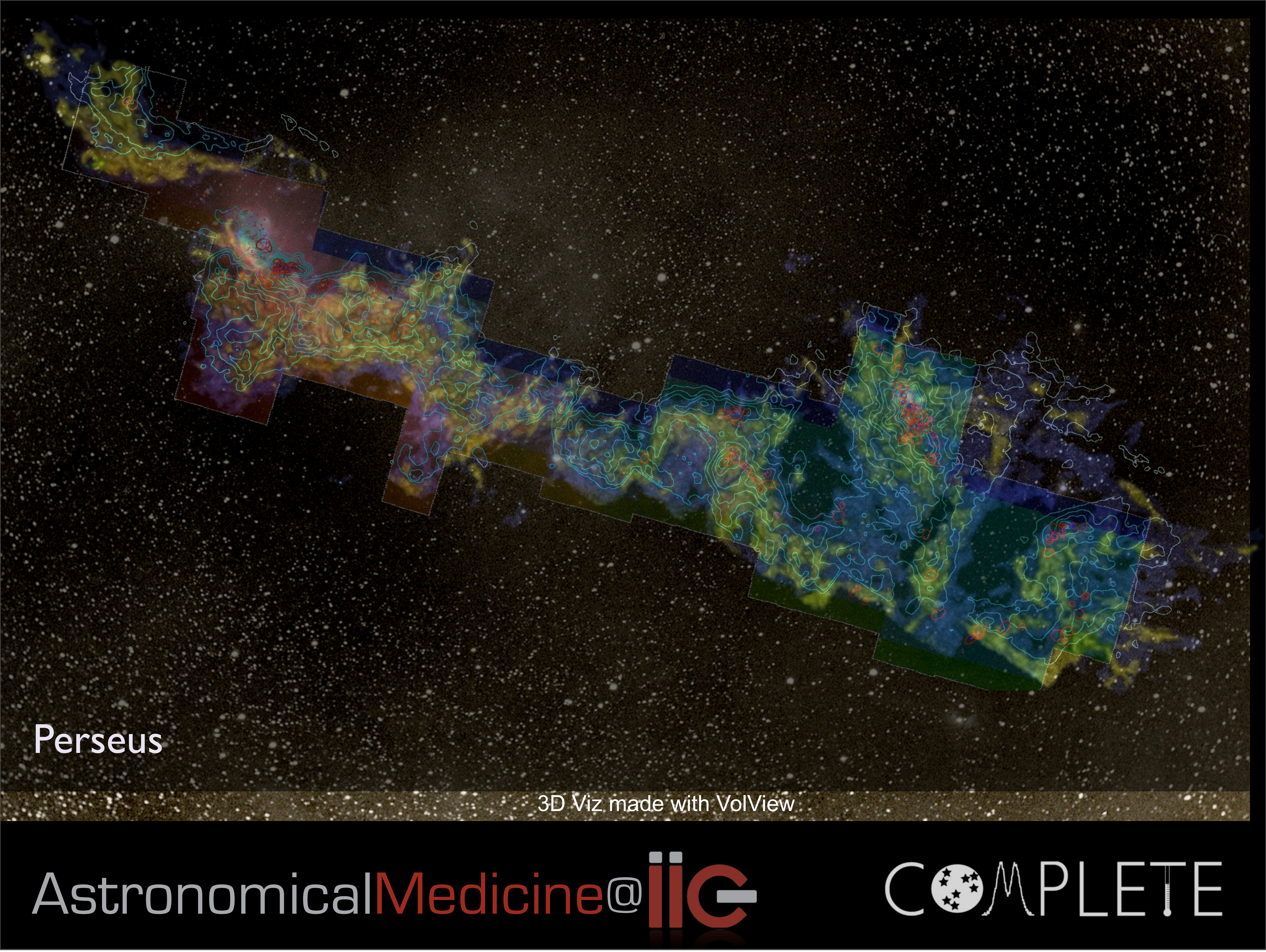
"PERSEUS"

"z" is line-of-sight velocity

(This kind of "series of 2D slices view" is known in the Viz as "the grand tour")

COMPLETE Perseus




Perseus
3D Viz made with VolView
AstronomicalMedicine@IIC
COMPLETE

What...

...is easier now than before?

fast computation, animation, 3D

...was easier before than now?

craftsmanship

...should be easier in the future?

modular craftsmanship

The "Easier" Future: Modular Craftsmanship

The Future we can see from "now"...

"live" interaction with data (DataDesk, WWT)
(scripts *only when useful*)

more display modes available (3D PDF, touch tables/walls, stereo+)

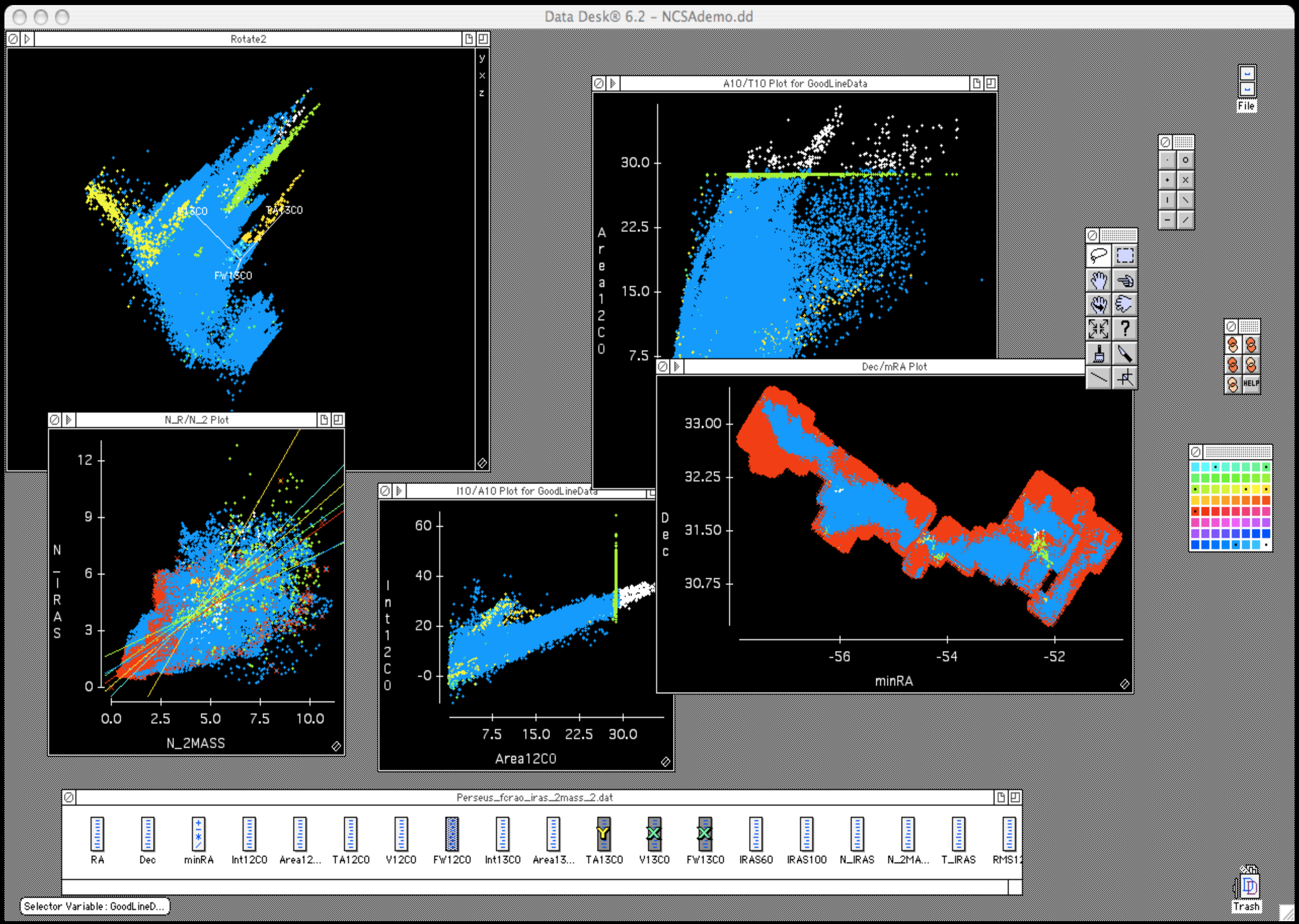
re-usable tools/mashups (Many Eyes, UFOmaps.com)

Unsolved Questions...

(feasibility of) templates/language (e.g. Grammar of Graphics)

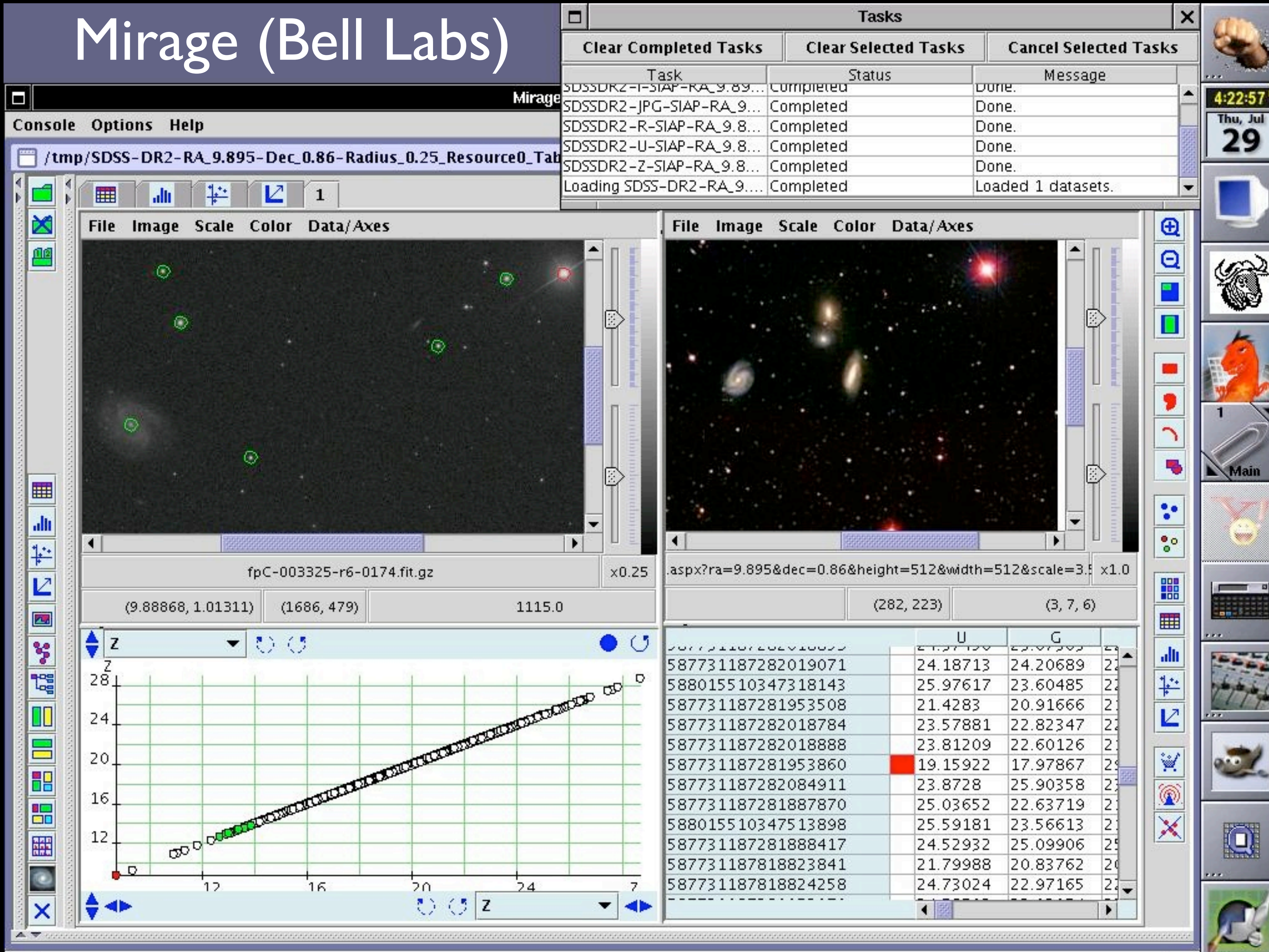
improved graphical representation of uncertainty

"Data Desk"



If only DataDesk were >2D...??

Mirage (Bell Labs)



cf. Avizo (Mercury Systems); some aspects of GenePattern; Taverna...

3D PDF

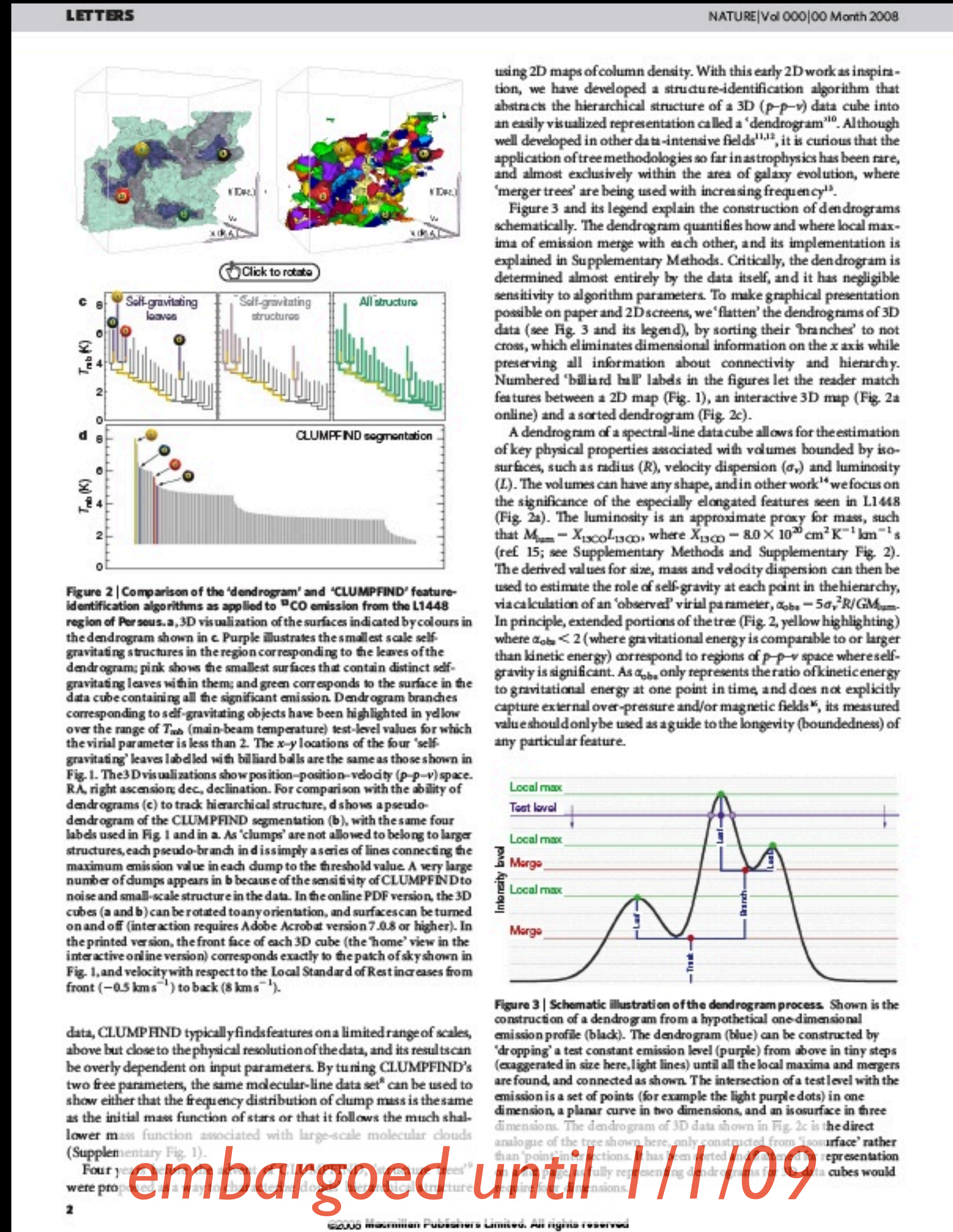


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to [13]CO emission from the L1448 region of Perseus. **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The *x*–*y* locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (*p*–*p*–*v*) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\ \mathrm{km\,s^{-1}}$) to back ($8\ \mathrm{km\,s^{-1}}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four y... were pro...

using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (*p*–*p*–*v*) data cube into an easily visualized representation called a 'dendrogram'[10]. Although well developed in other data-intensive fields[11,12], it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency[13].

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the *x* axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by isosurfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work[14] we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{lum} = X_{13CO} L_{13CO}$, where $X_{13CO} = 8.0 \times 10^{20}\ \mathrm{cm^2\,K^{-1}\,km^{-1}\,s}$ (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\alpha_{obs} = 5\sigma_v^2 R / G M_{lum}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{obs} < 2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of *p*–*p*–*v* space where self-gravity is significant. As α_{obs} only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields[16], its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

embargoed until 1/1/09

Off the desktop

IIC Member login

Initiative in Innovative Computing at Harvard

home › research

scientists' discovery room lab (sdr lab)

Lead investigators
Chia Shen (IIC), Hanspeter Pfister (SEAS/IIC) and Robert Lue (FAS/Molecular and Cellular Biology)

Project staff
Michael Horn, Hao Jiang and Meekal Bajaj

Description

The Scientists' Discovery Room (SDR) is a next-generation visual digital laboratory for science discovery, collaborative learning and education. Our research focuses on experimenting with new modalities of human-computer interaction and visualization, to create a new genre of navigation, exploration and detailed analyses in multi-dimensional information spaces. All projects in SDR are in close collaboration with domain scientists and educators.

CThru, currently a collaborative endeavor with Molecular and Cellular Biology faculty, aims to develop a self-guided educational environment. In CThru, we examine methods for constructing interactive video-based educational modules. Using the animation "The Inner Life of the Cell" as a testbed, CThru addresses research issues of embedding interactive visible objects, extensive multimedia information and manipulatable 3D models within a video flow for self-explanatory learning, replacing sequential video viewing with the experience of of exploring and manipulating in a multi-dimentional information space.

INVOLV is a generalizable multi-user interactive visualization framework for large hierarchical data sets. In this project, we address the visual layout of both the primary data representation and the overlay of alternate structures of the same data. Our first case study is the visualization of life on earth based on the Encyclopedia of Life (www.eol.org). We address the challenge of allowing free-form exploration of more than 1.2 million named species while communicating issues of biodiversity and phylogeny. The current visualization, designed for biodiversity science education settings, combines a Voronoi Treemap tessellation (see photo) with innovative human-computer interaction designs to support collaborative exploration and learning.

Slideshow: Tabletop Computers *Continued*

By Meredith Ringel Morris

First Published December 2008

Email | Print | Comments (1) | Reprints | Newsletters

Del.icio.us | Digg | Slashdot

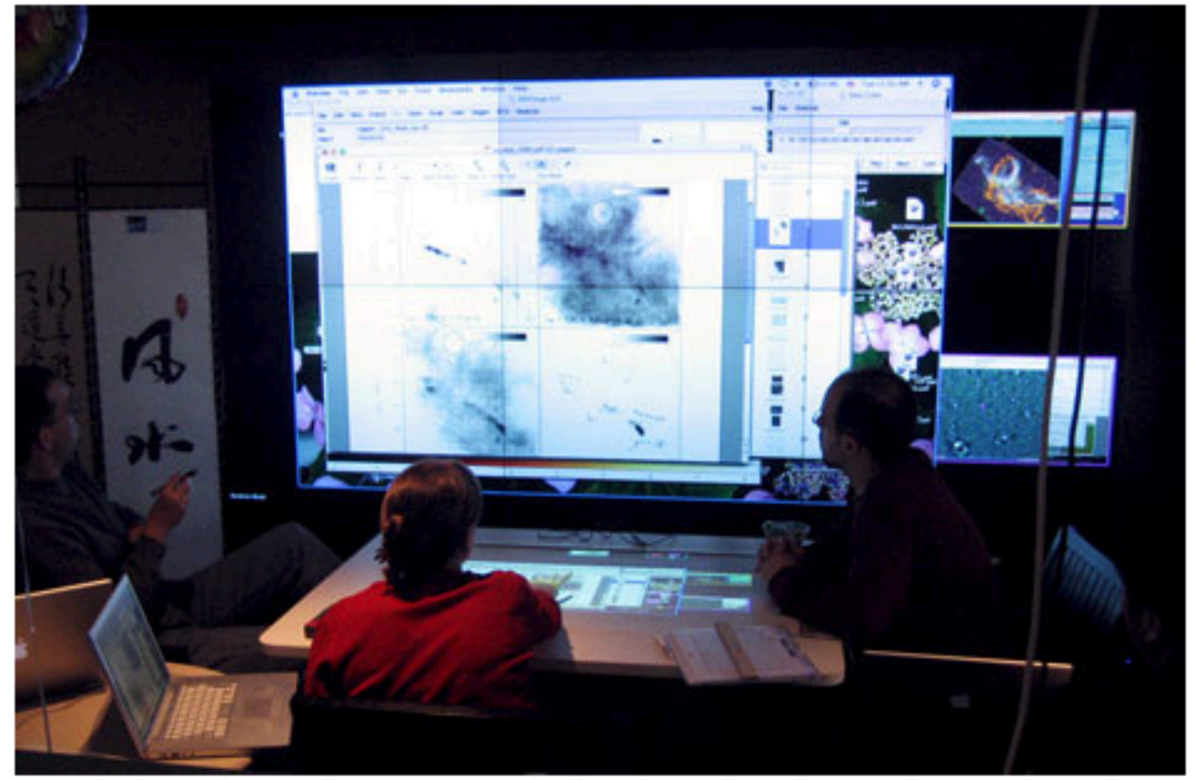


PHOTO: HAO JIANG, DANIEL WIGDOR, CLIFTON FORLINES, AND CHIA SHEN

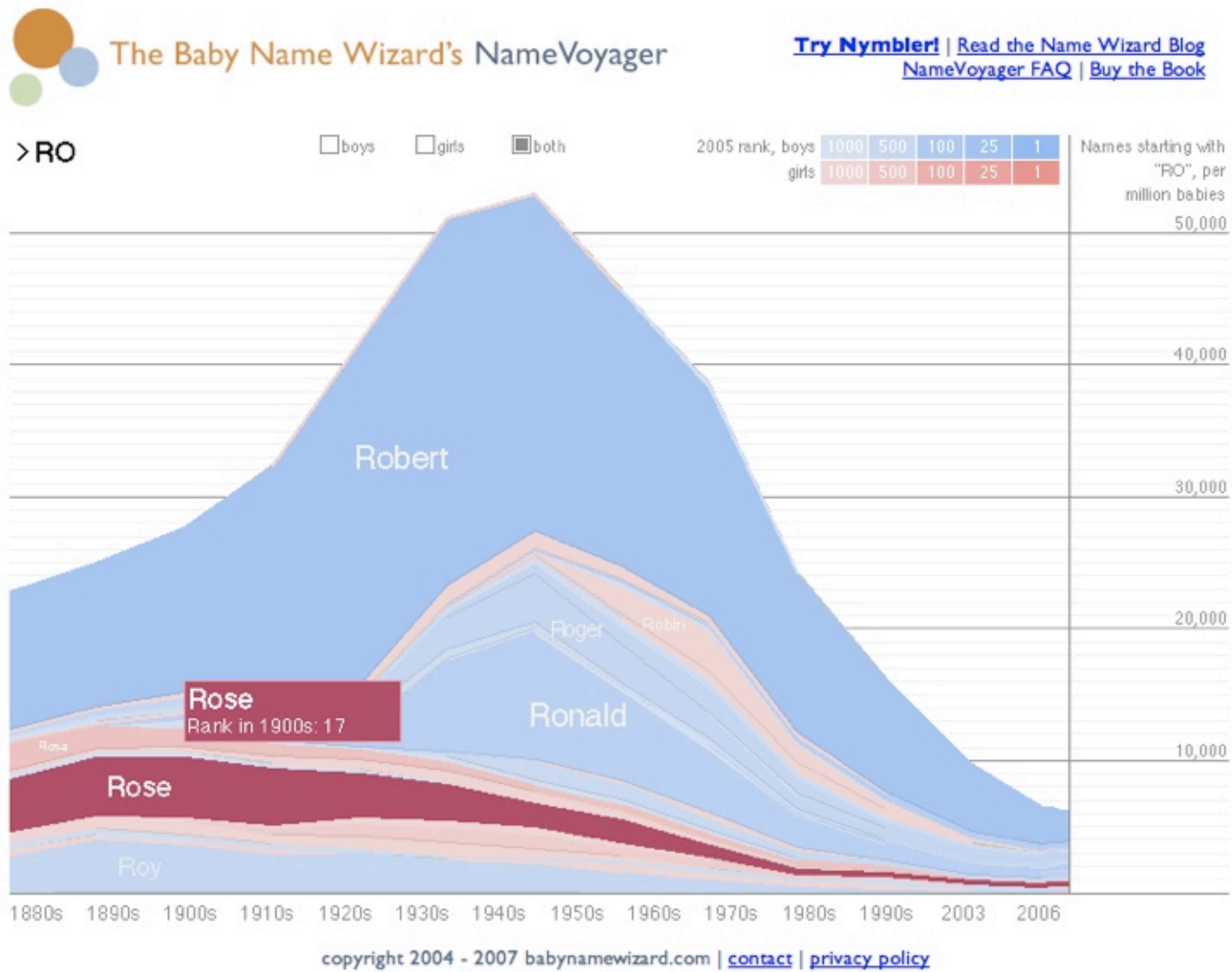
UBITABLE: Users can interact with surface computers through auxiliary devices, such as laptops, phones, and PDAs. The display on the auxiliary device can convey private or sensitive content to a single user, while group-appropriate content can appear on the tabletop display. Chia Shen and her colleagues at Mitsubishi Electric Research Laboratories, in Cambridge, Mass., have explored auxiliary interactions with surface computers in their UbiTable project, in which two people with laptops collaborate over a tabletop display. Recently, Shen expanded the UbiTable into an interactive room called the WeSpace. People can share data on their laptops with other people in the room, using both a table and a large display wall. Here, three Harvard University astrophysicists discuss radio and IR spectrum images using the WeSpace.

http://iic.harvard.edu/research/scientists-discovery-room-lab-sdr-lab

http://spectrum.ieee.org/dec08/6999/9

The Scientists' Discovery Room: Version 0.01

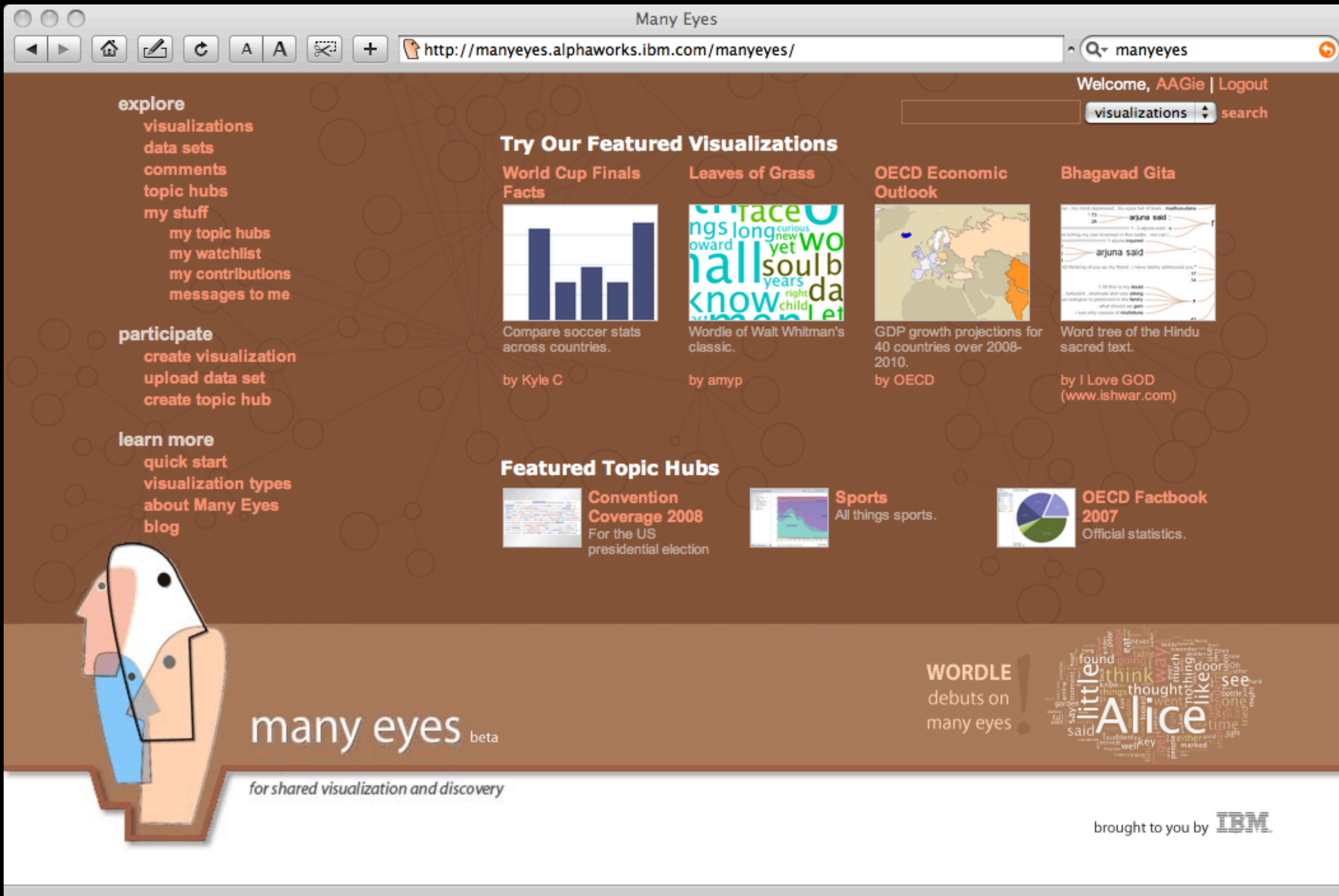
movie courtesy Daniel Wigdor, taken at MERL, Kendall Square, Cambridge



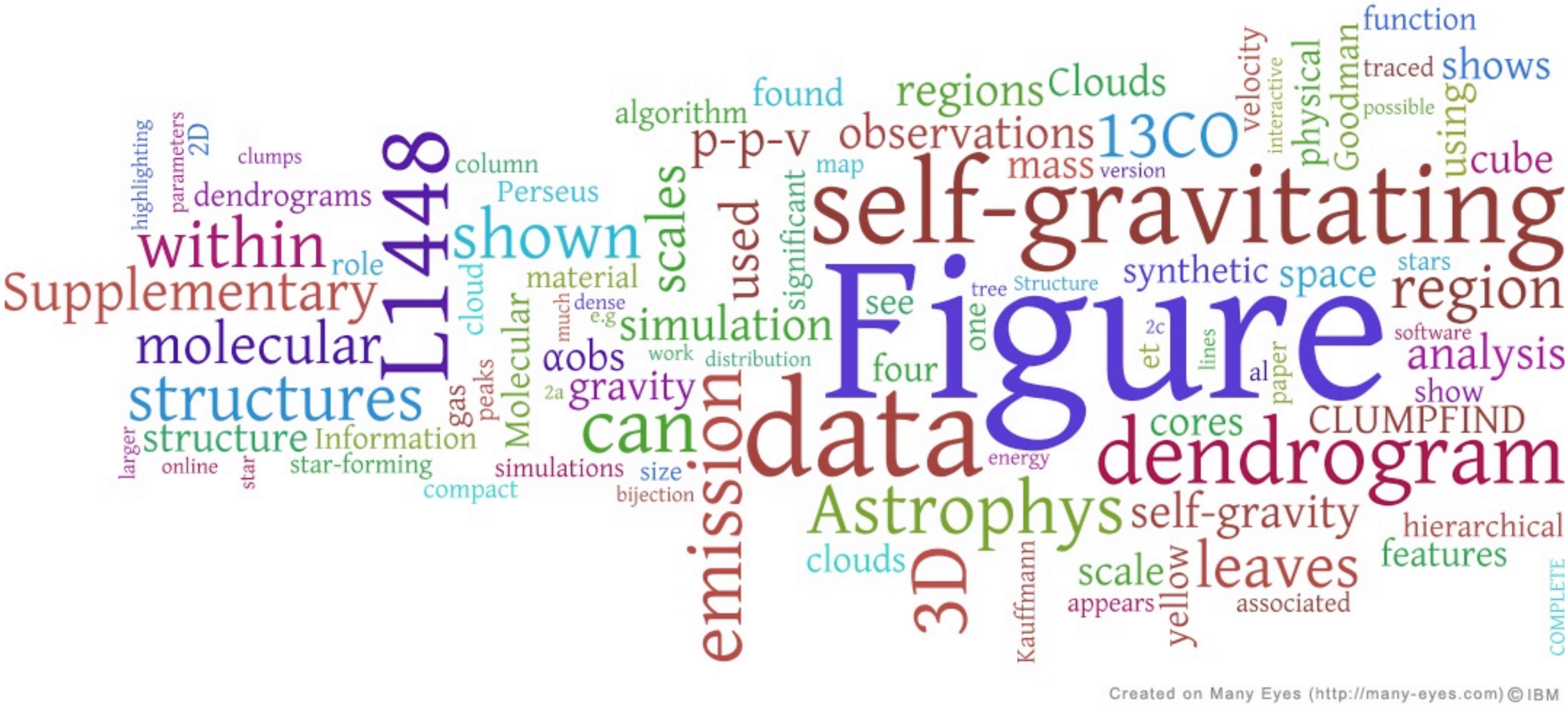
Data Viz at its Best: Baby Name Wizard's Name Voyager from Martin Wattenberg re-usable tools

Many Eyes:

Martin Wattenberg & Fernanda Viegas (IBM)

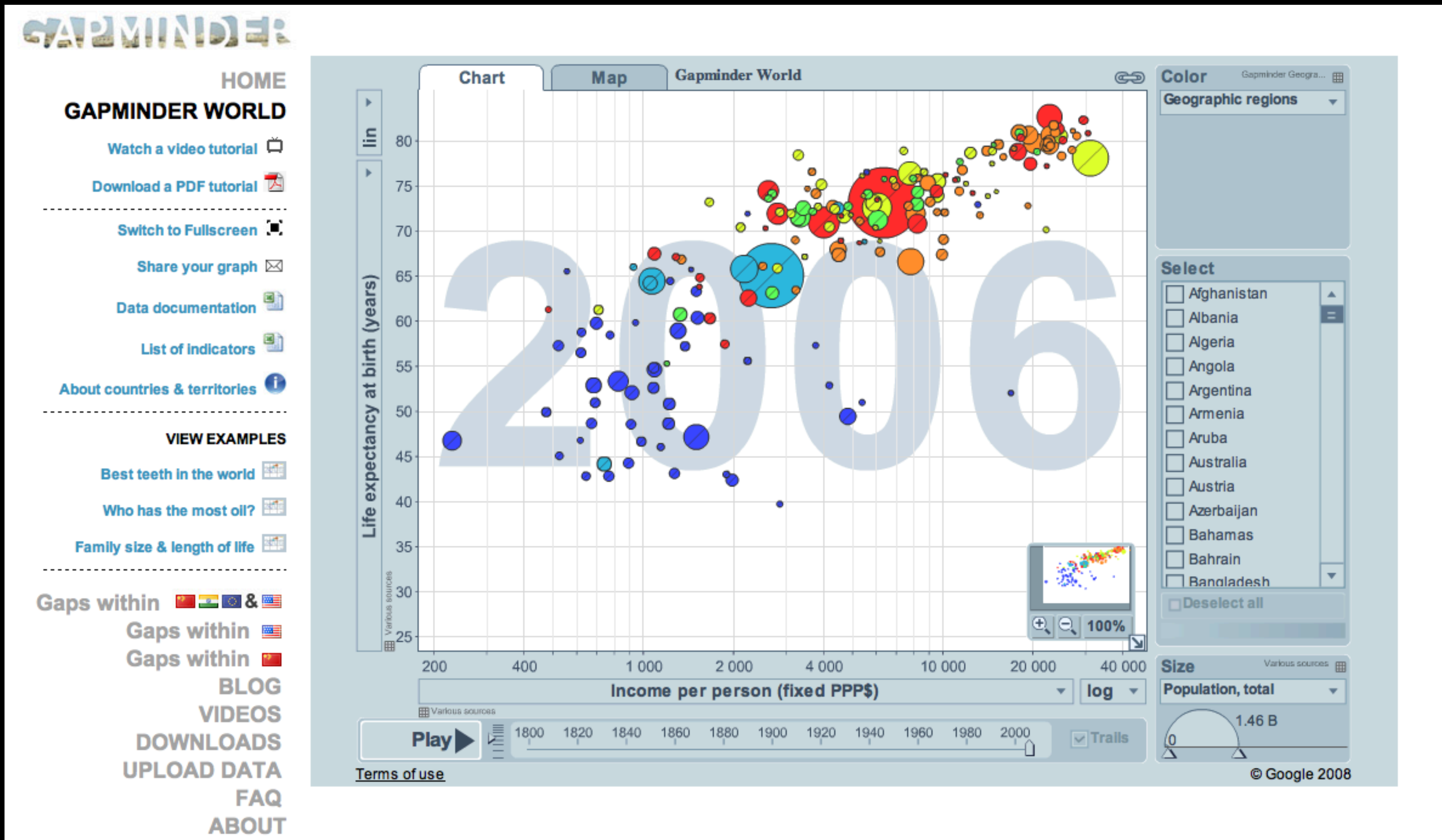


What Many Eyes thinks my "3D PDF" paper on self-gravity in hierarchical star-forming regions is about...



Gapminder:

Re-usable, interactive graphical tools



http://gapminder.org

re-usable tools

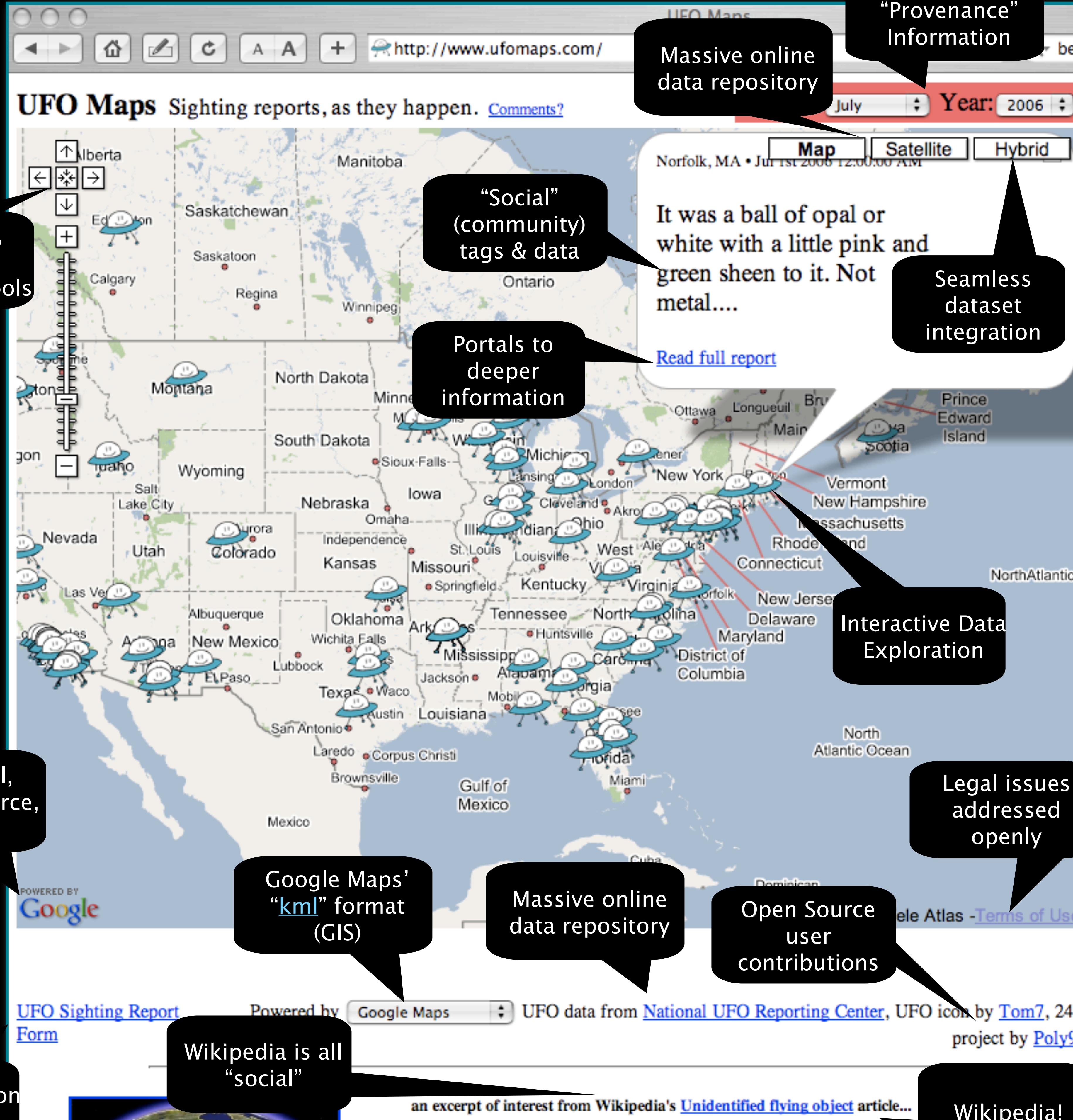
http://finance.google.com/finance?client=ob&q=INDEXDJX:DJI

Facilitating re-usabilty...

Mashups, Resource Hubs, APIs

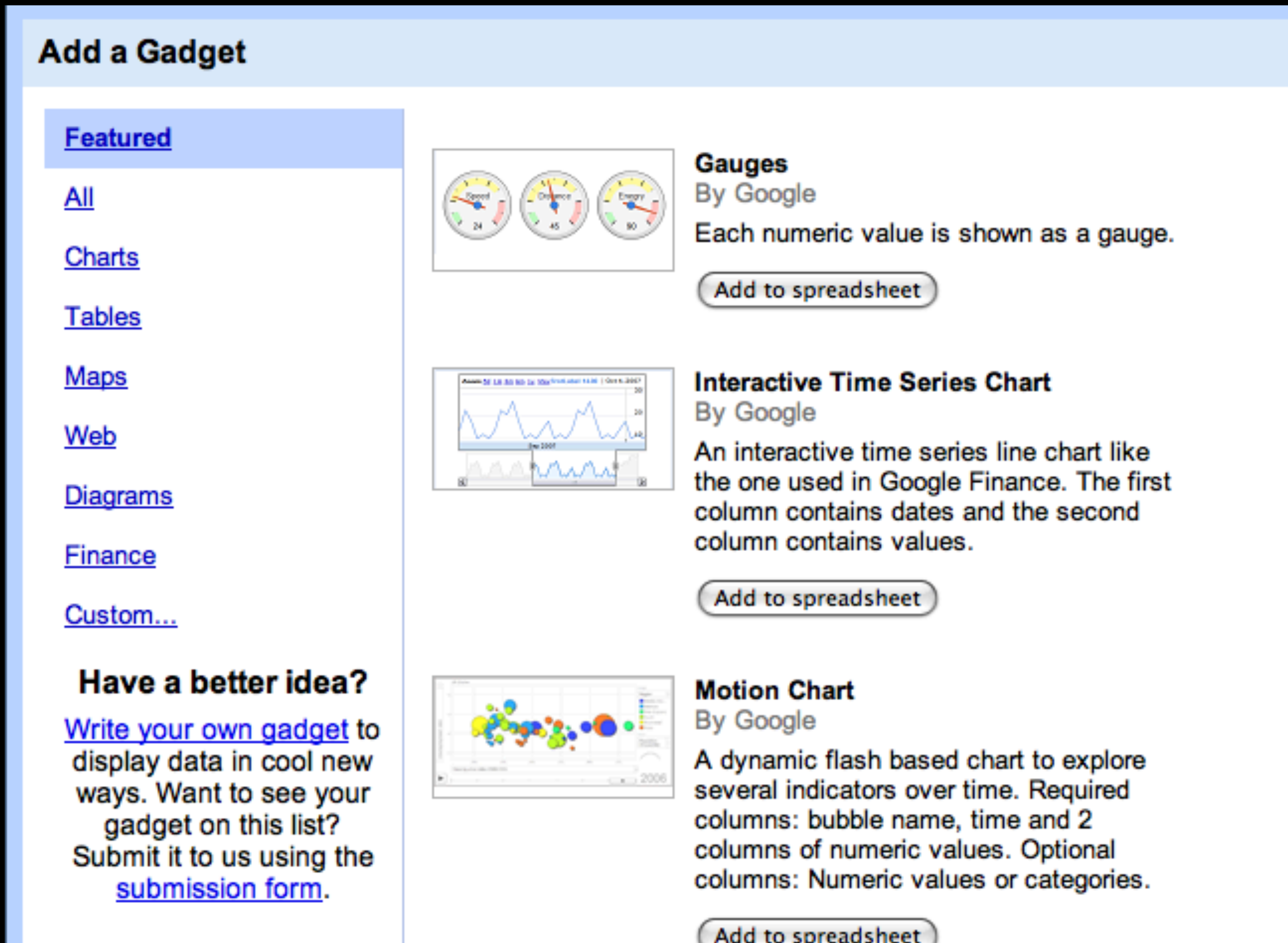
More on re-usabilty

Exemplar mashup c. 2006

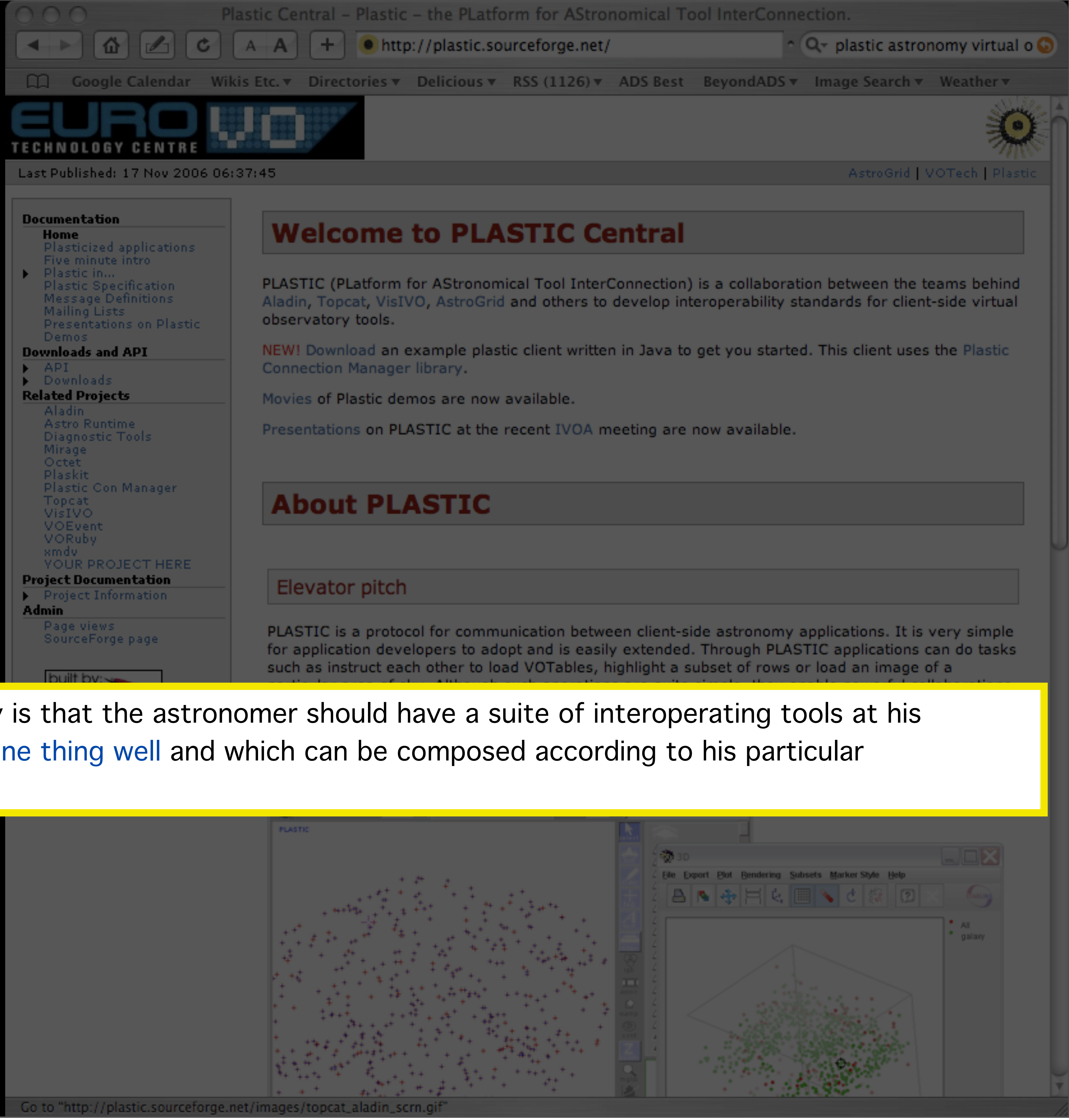


More on re-usabilty

Google Gadgets & APIs



More on re-usabilty



"PLASTIC"
(SAMP)

> between tools. The philosophy is that the astronomer should have a suite of interoperating tools at his disposal, each of which does one thing well and which can be composed according to his particular needs

What...

...is easier now than before?

fast computation, animation, 3D

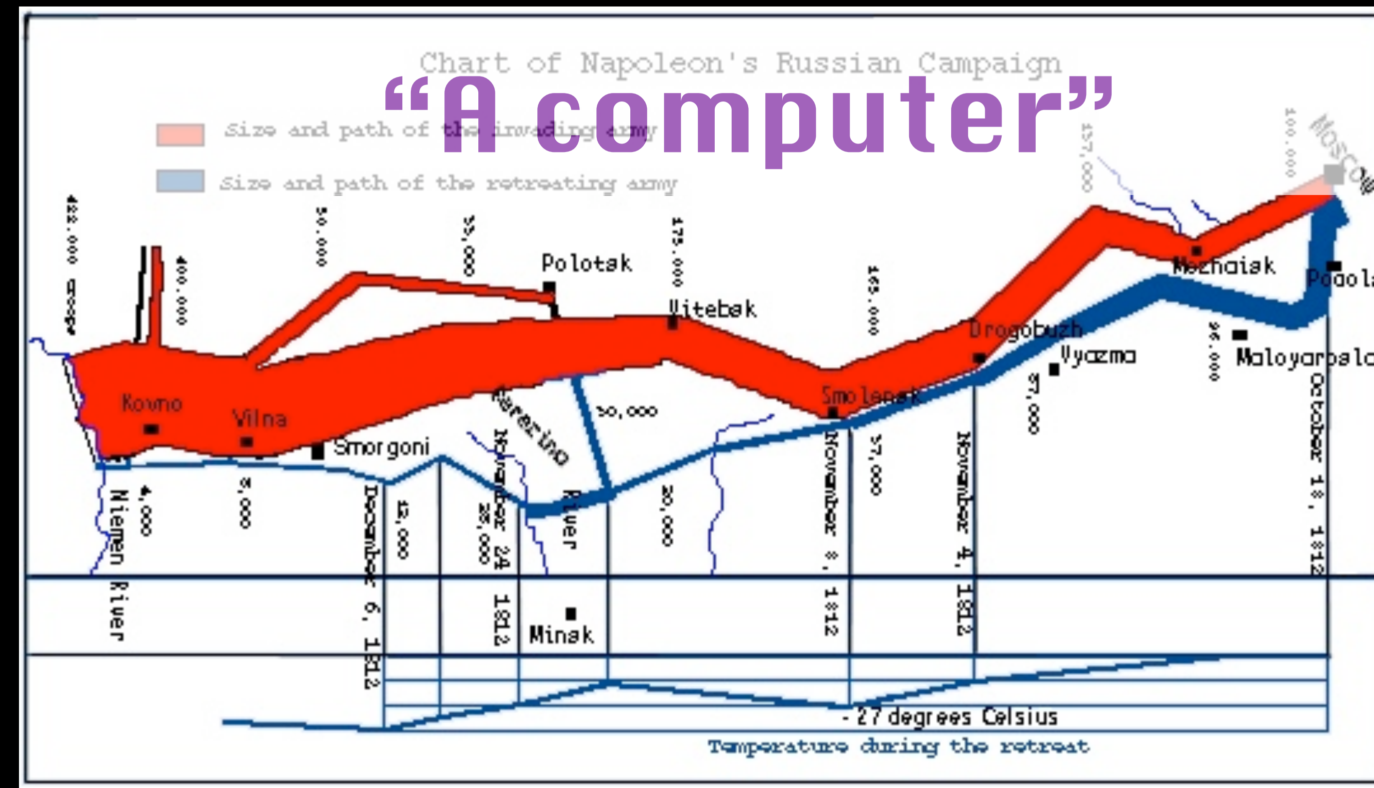
...was easier before than now?

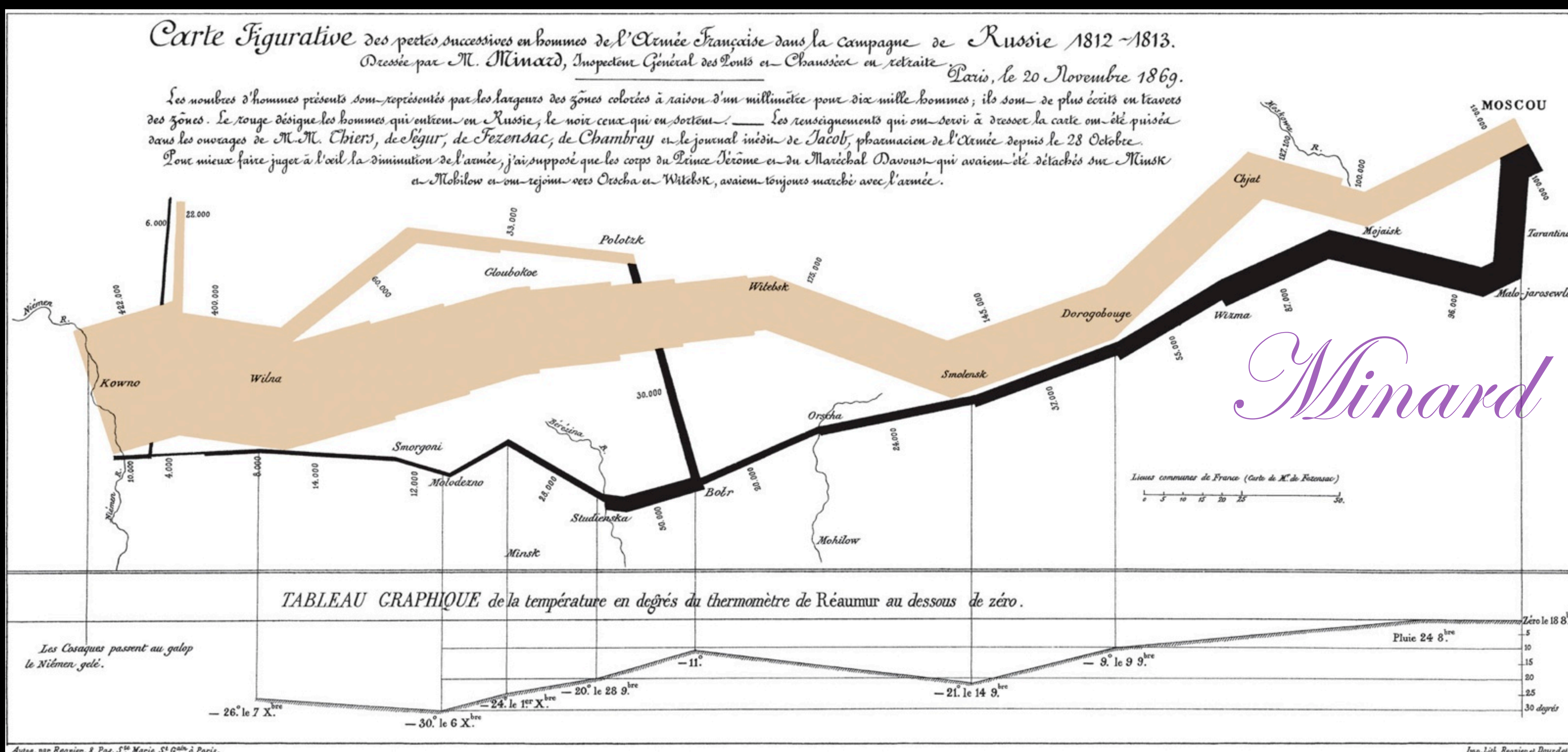
craftsmanship

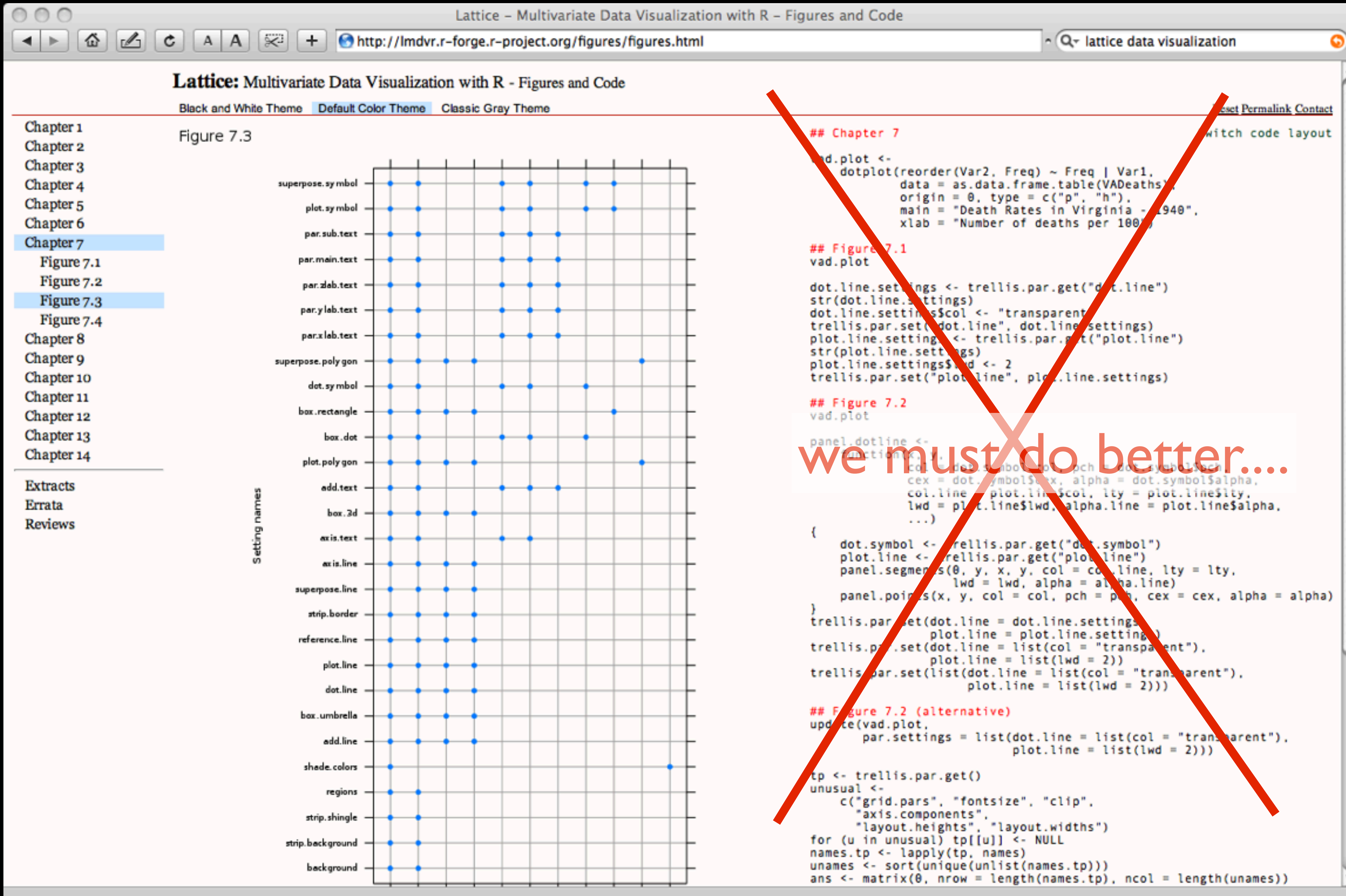
...should be easier in the future?

modular craftsmanship

Are we held back by confining tools?

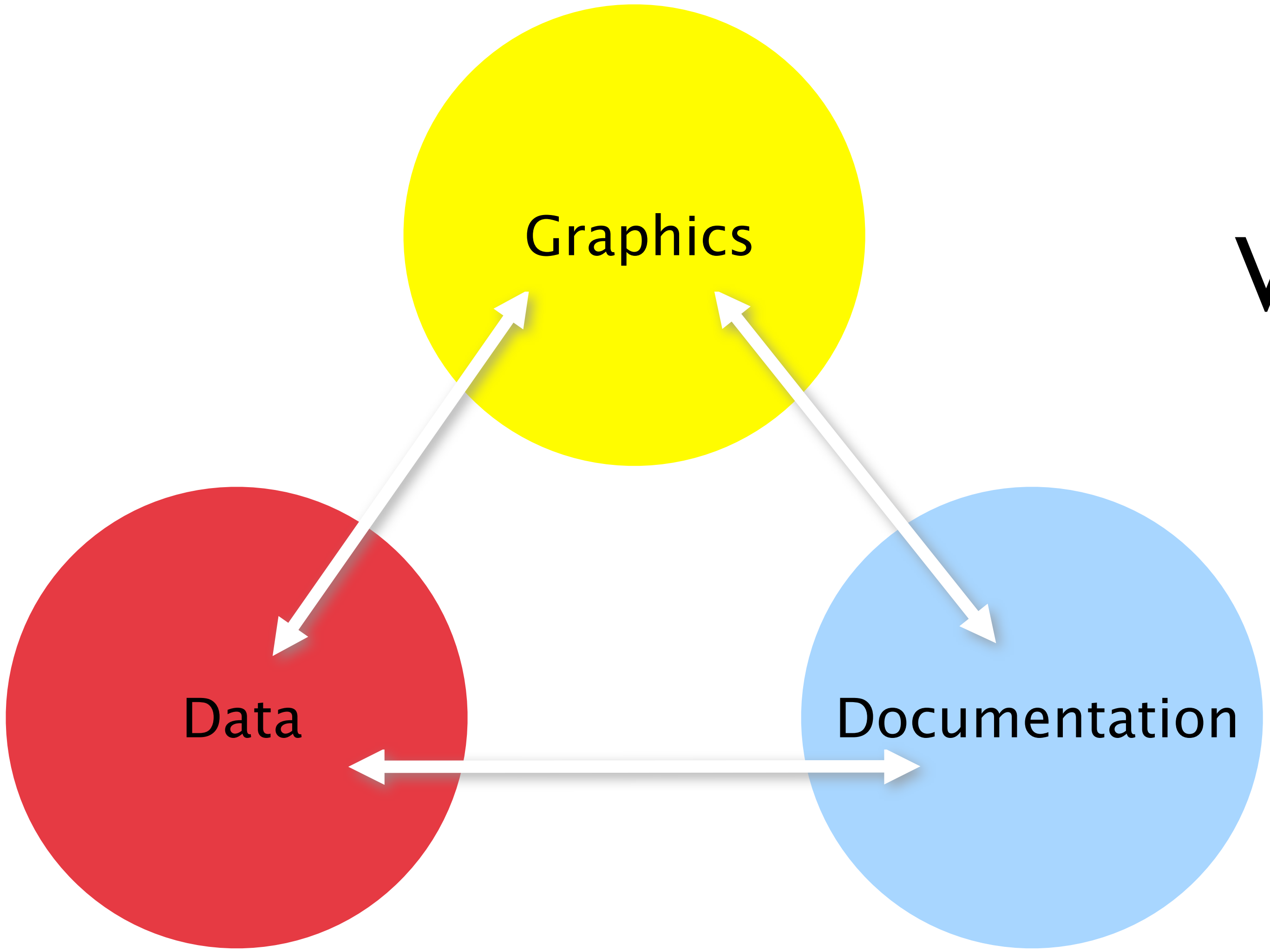





Lattice – Multivariate Data Visualization with R – Figures and Code
http://lmdvr.r-forge.r-project.org/figures/figures.html
lattice data visualization
Lattice: Multivariate Data Visualization with R - Figures and Code
Black and White Theme
Default Color Theme
Classic Gray Theme
Permalink Contact
Chapter 1
Chapter 2
Chapter 3
Chapter 4
Chapter 5
Chapter 6
Chapter 7
Figure 7.1
Figure 7.2
Figure 7.3
Figure 7.4
Chapter 8
Chapter 9
Chapter 10
Chapter 11
Chapter 12
Chapter 13
Chapter 14
Extracts
Errata
Reviews
Figure 7.3
Setting names
superpose.symbol
plot.symbol
par.sub.text
par.main.text
par.zlab.text
par.ylab.text
par.xlab.text
superpose.polygon
dot.symbol
box.rectangle
box.dot
plot.polygon
add.text
box.3d
axis.text
axis.line
superpose.line
strip.border
reference.line
plot.line
dot.line
box.umbrella
add.line
shade.colors
regions
strip.shingle
strip.background
background
Chapter 7
switch code layout
Figure 7.1
vad.plot
Figure 7.2
vad.plot
Figure 7.2 (alternative)
tp <- trellis.par.get()
for (u in unusual) tp[[u]] <- NULL
names.tp <- lapply(tp, names)
unames <- sort(unique(unlist(names.tp)))
ans <- matrix(0, nrow = length(names.tp), ncol = length(unames))
we must do better....

The Future, Now

World Wide Telescope
(ask for a demo...)



Seeing Science