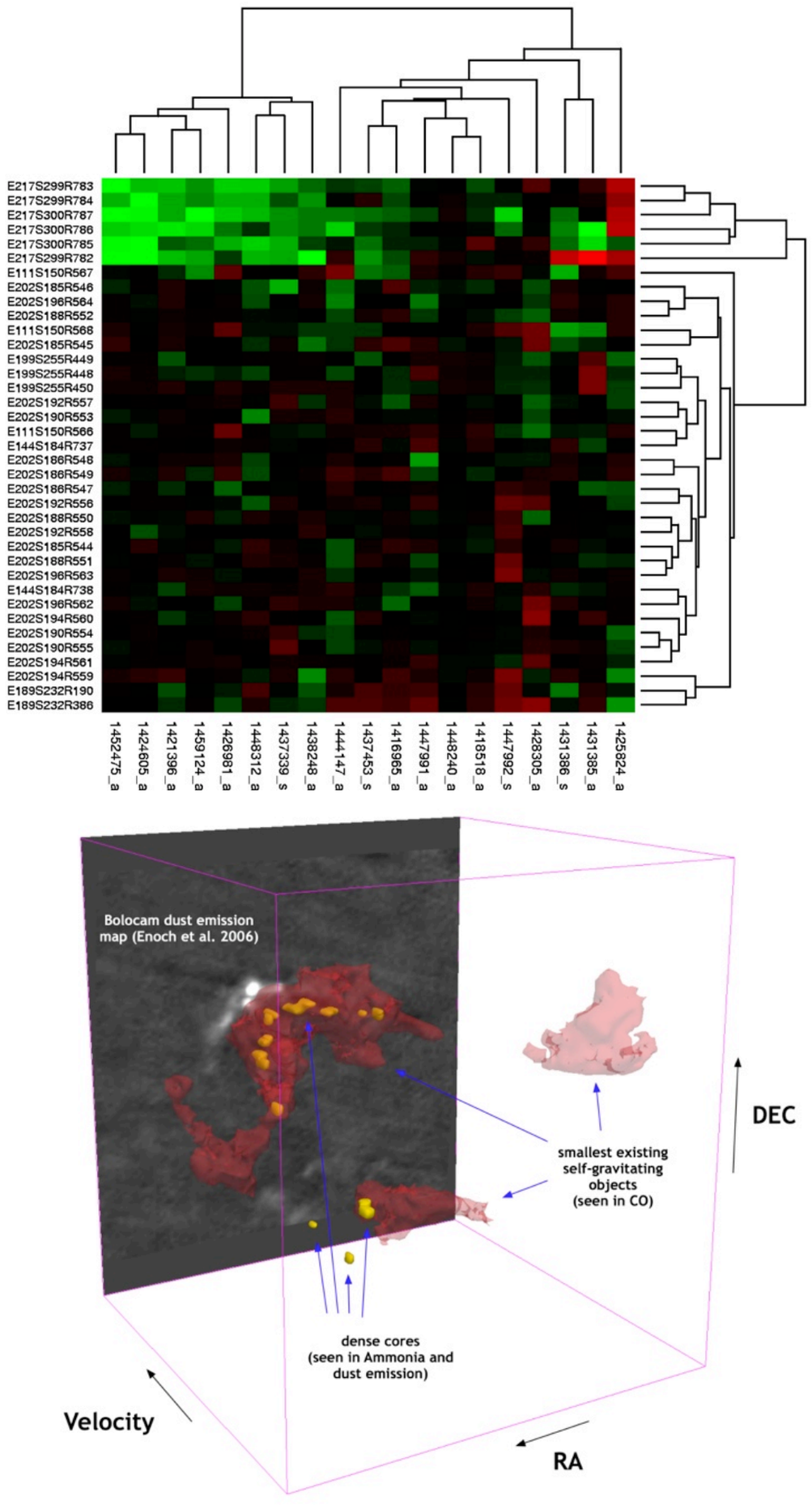
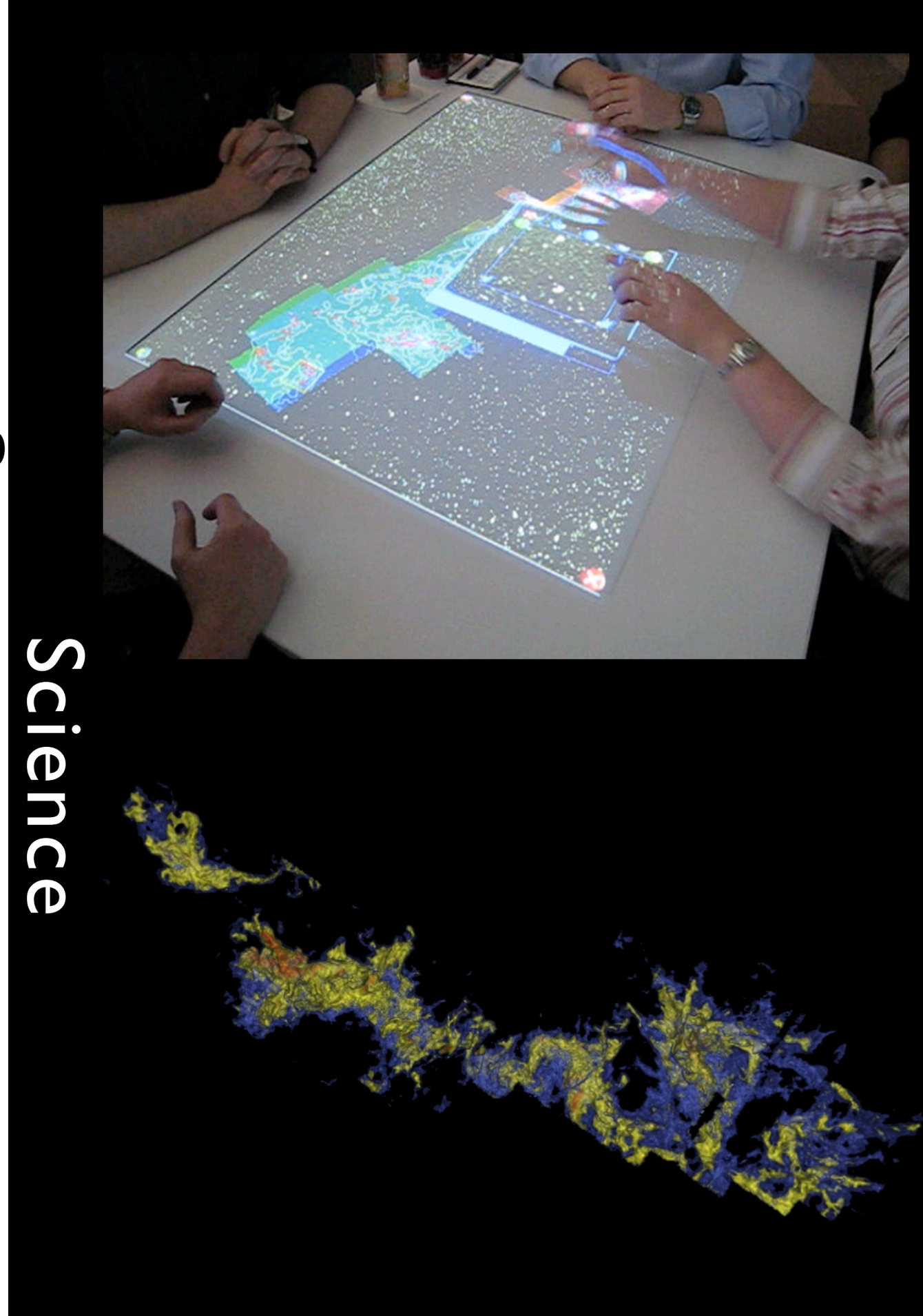
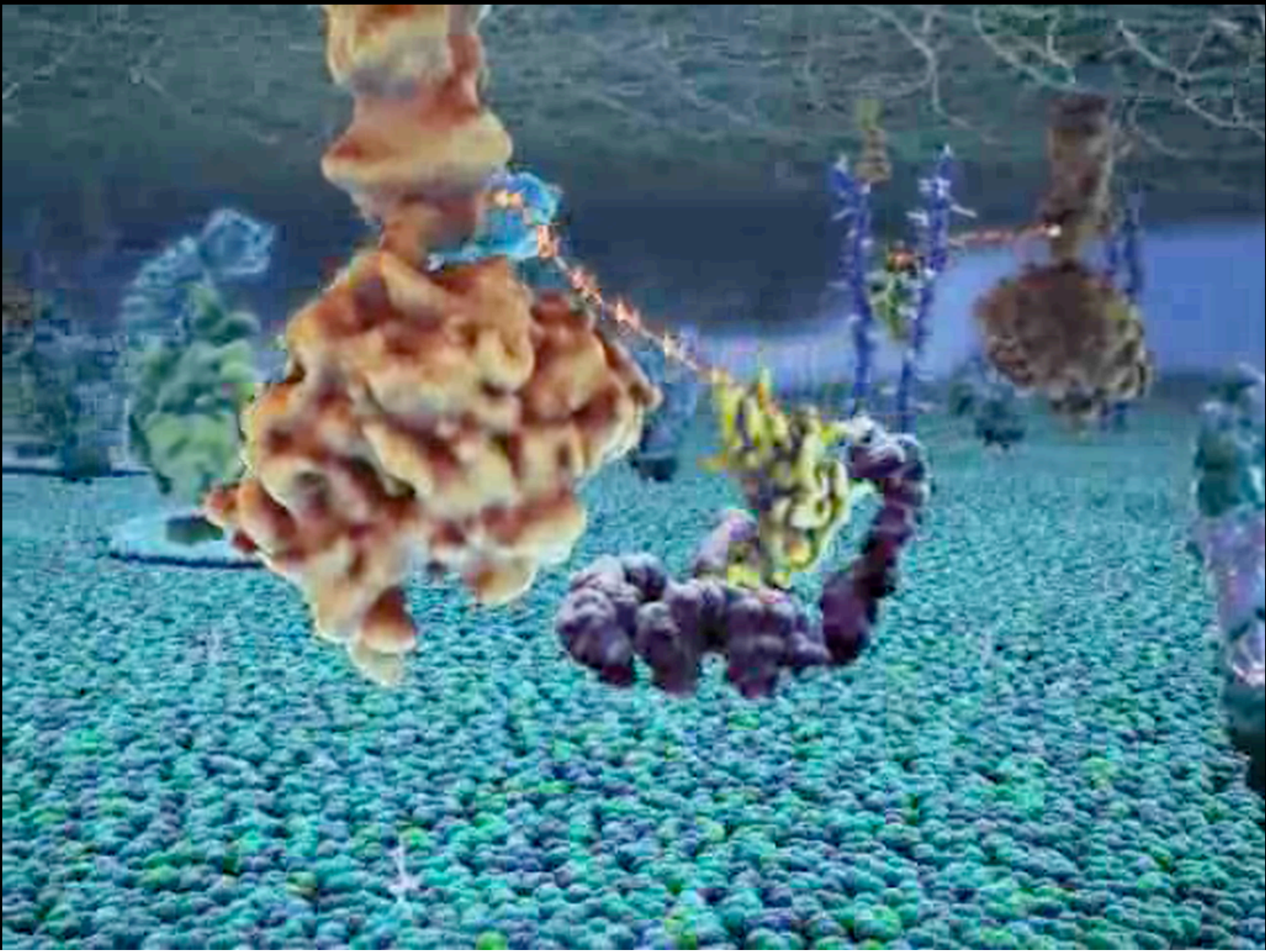


Seeing Science





Seeing Science

Alyssa A. Goodman

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

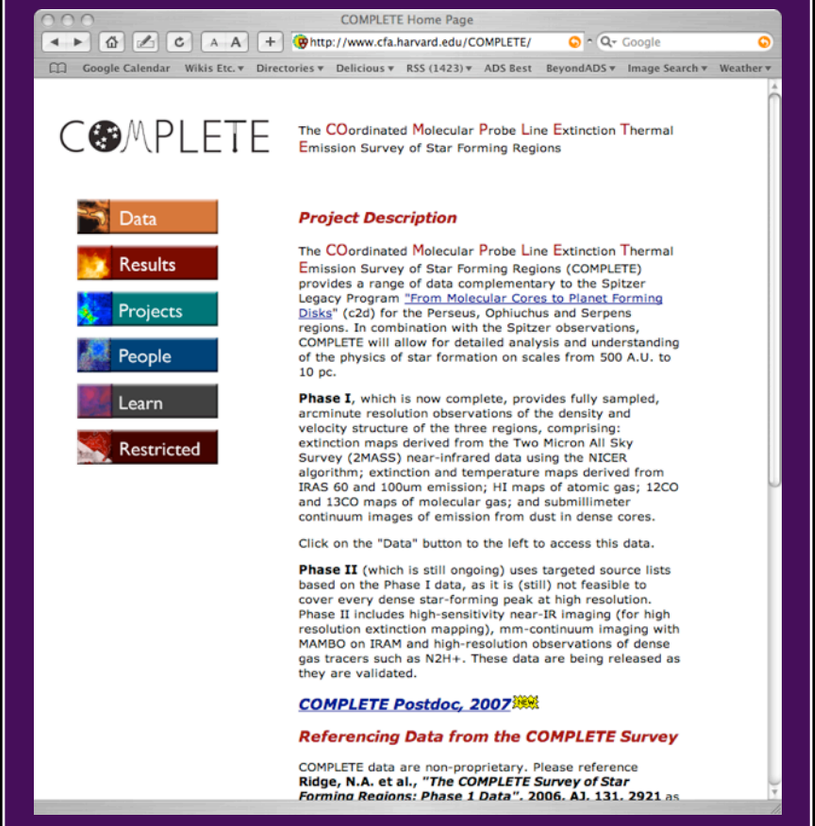
Scholar-in-Residence,
WGBH Boston

Who am I?

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

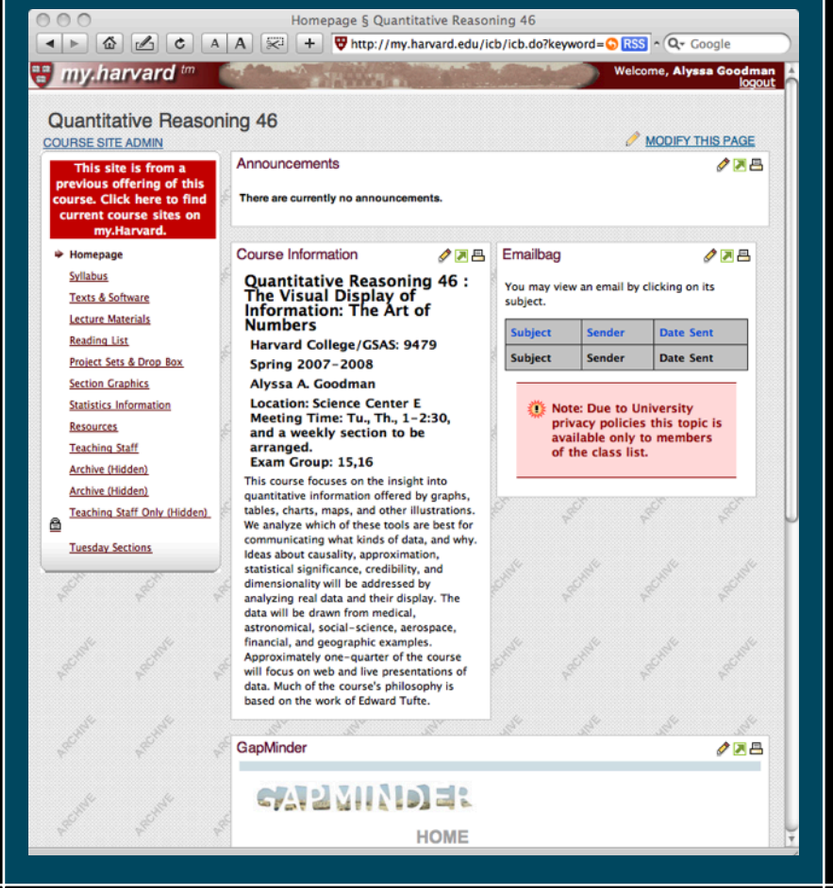
Art & Photography

Astronomy



Scientific Computing

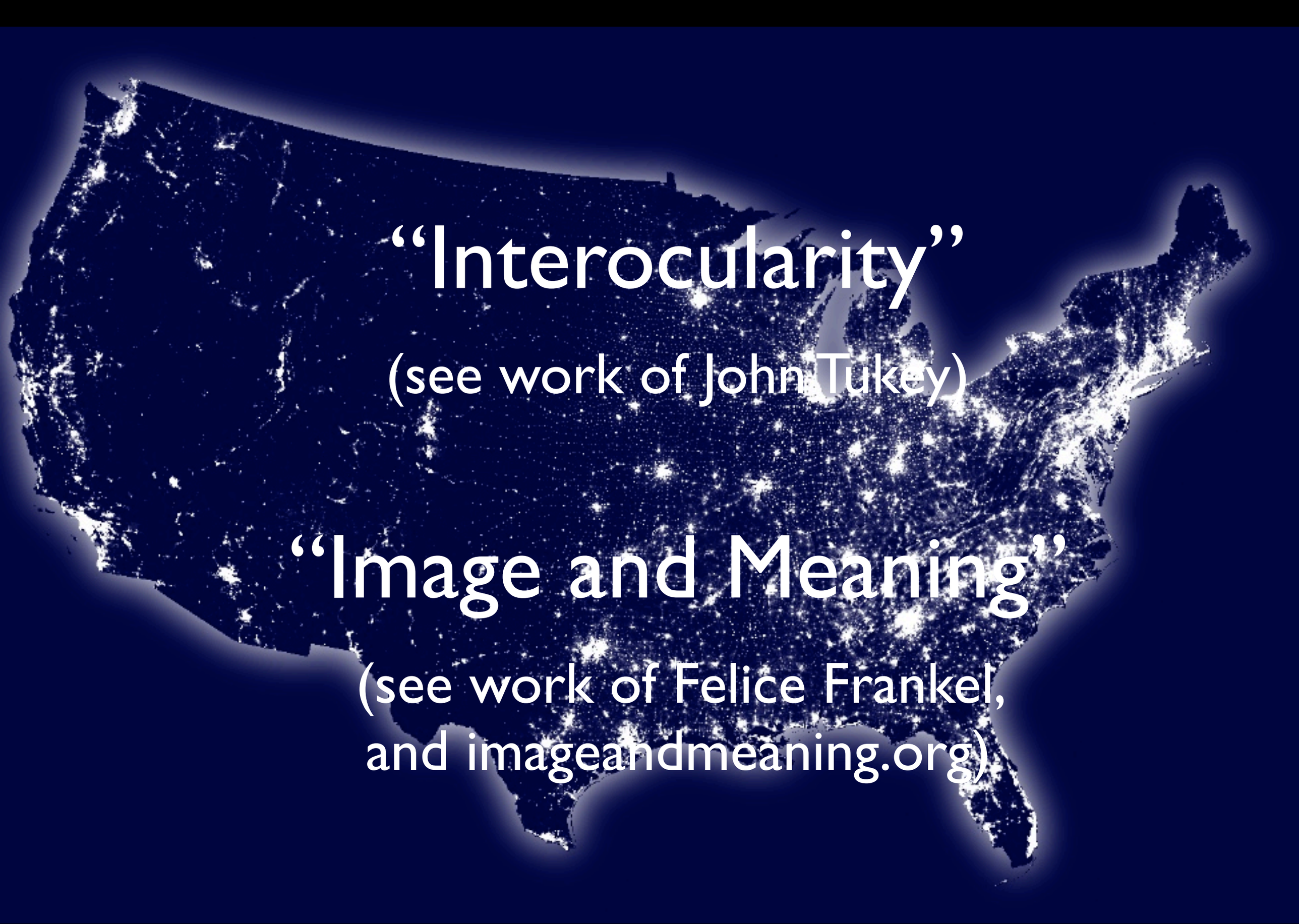
Teaching



Relative Strengths

Pattern Recognition
Creativity

Calculations


“Interocularity”
(see work of John Tukey)
“Image and Meaning”
(see work of Felice Frankel,
and imageandmeaning.org)

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?

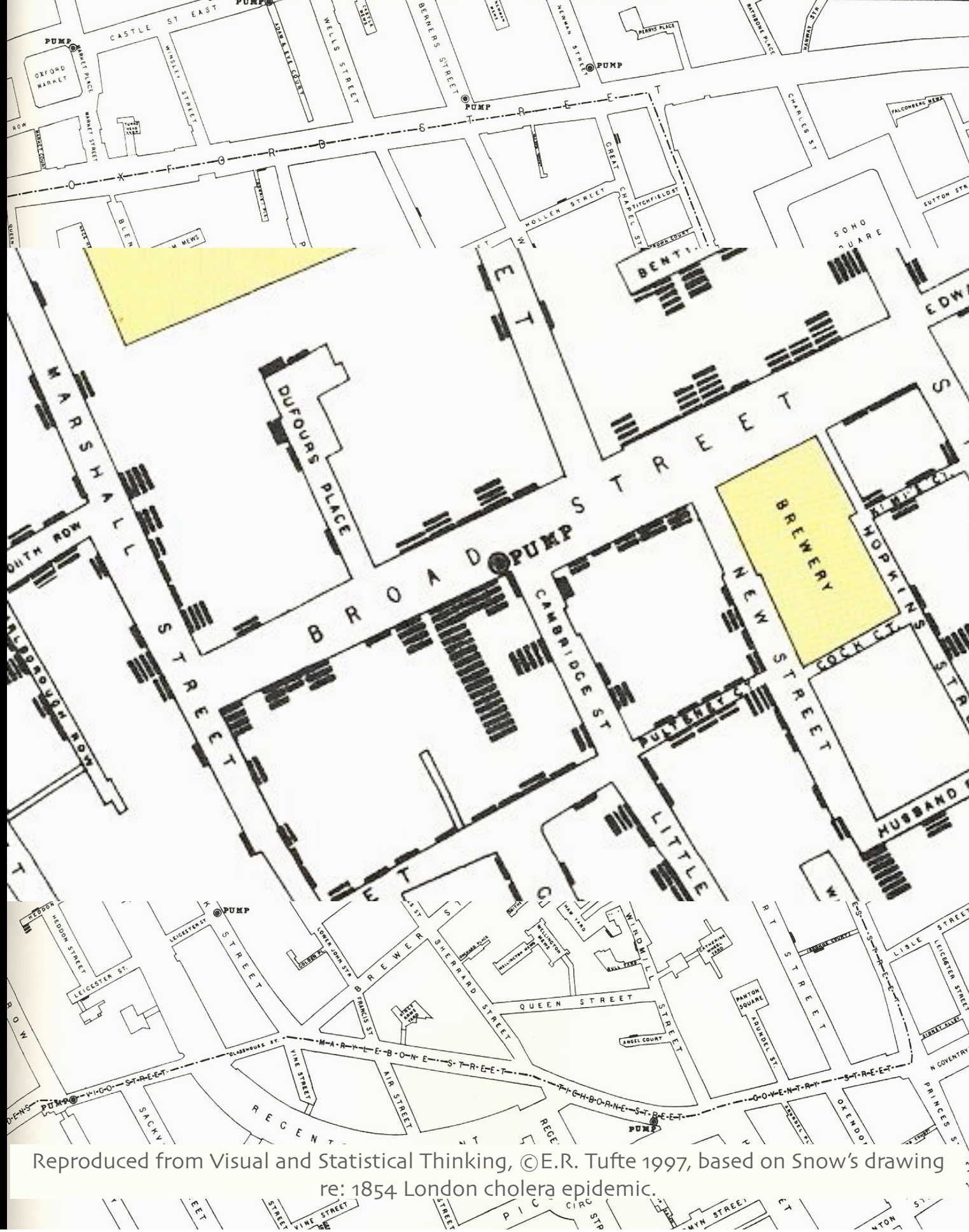
modular craftsmanship

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

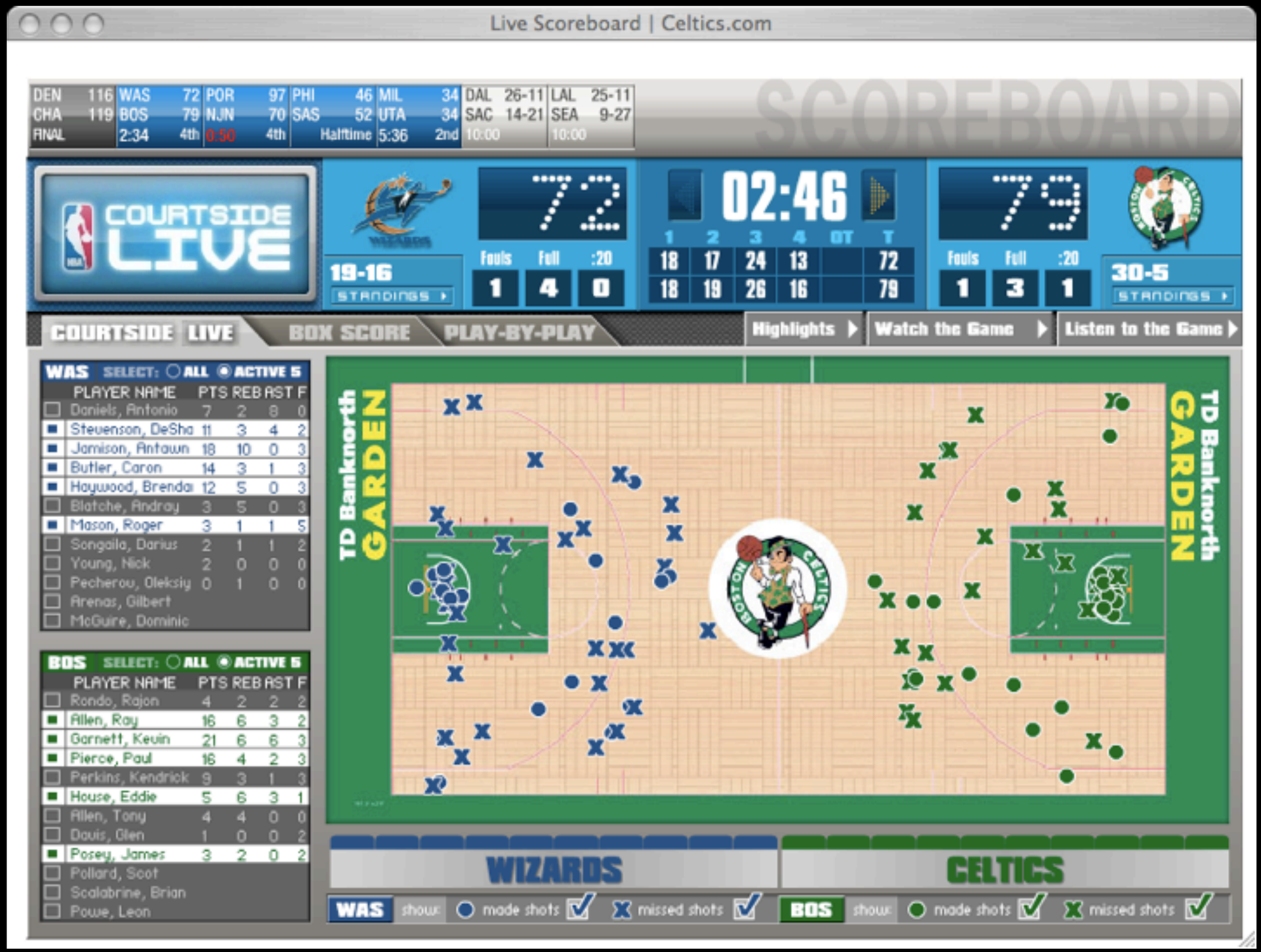
✓Tables

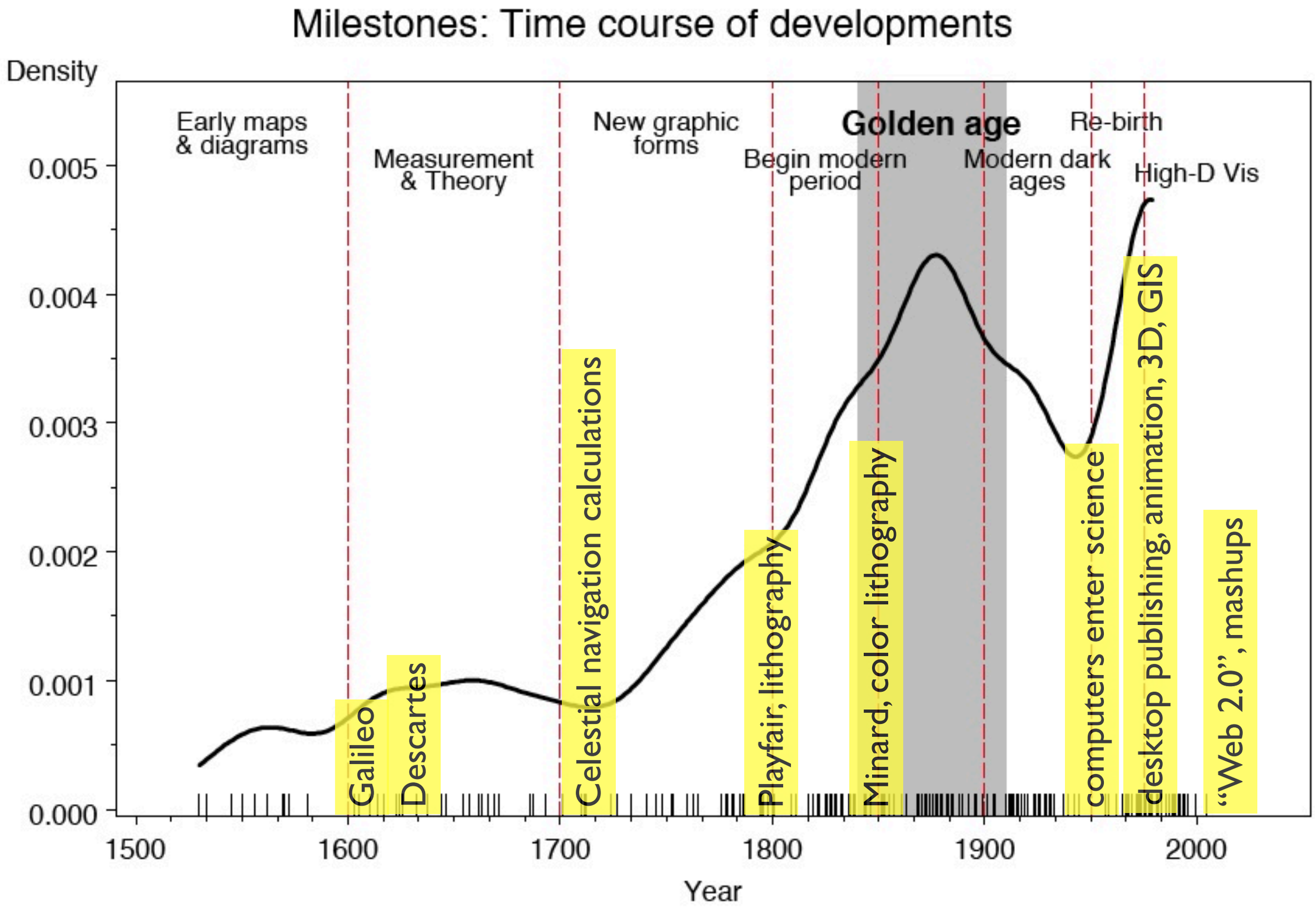
✖Graphs

✓Charts

✓Illustrations

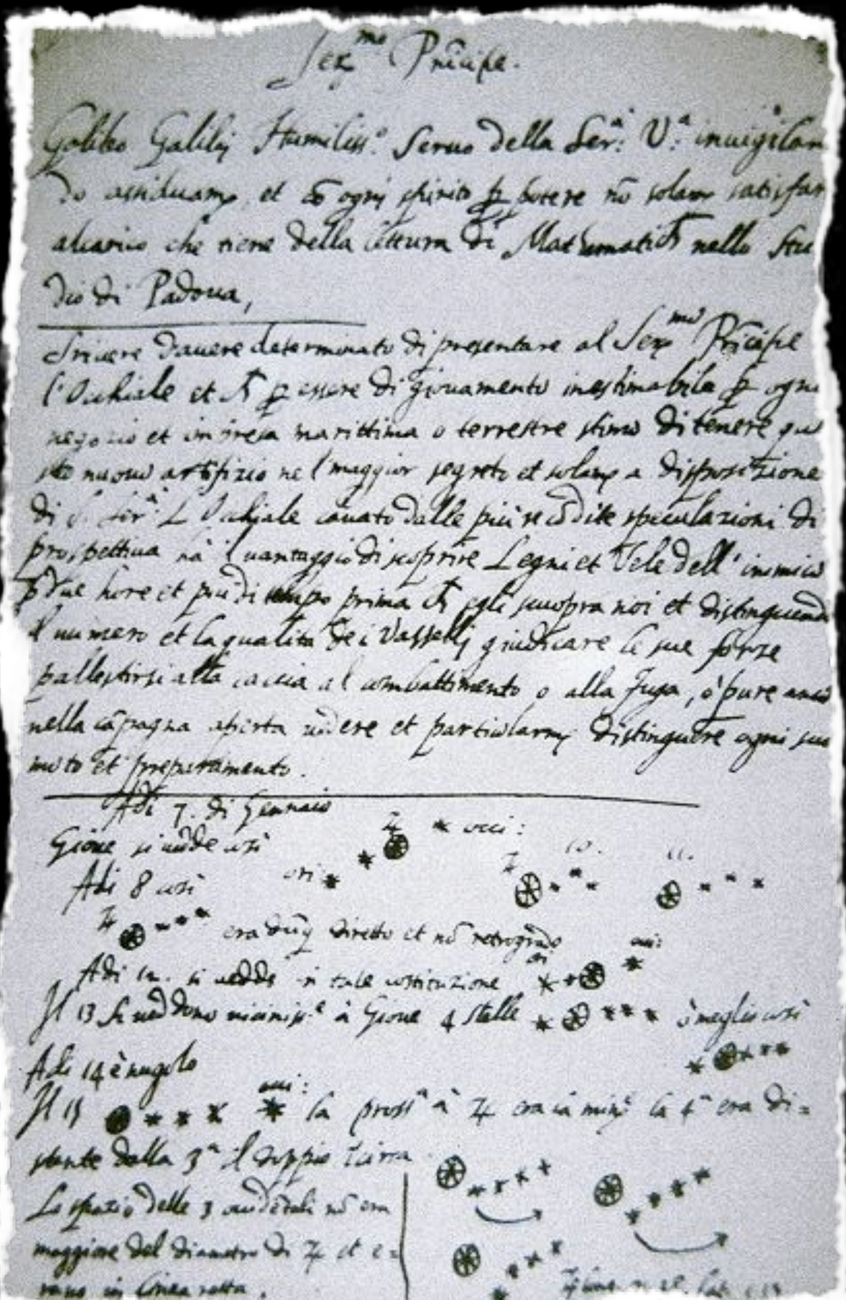
✓Combinations

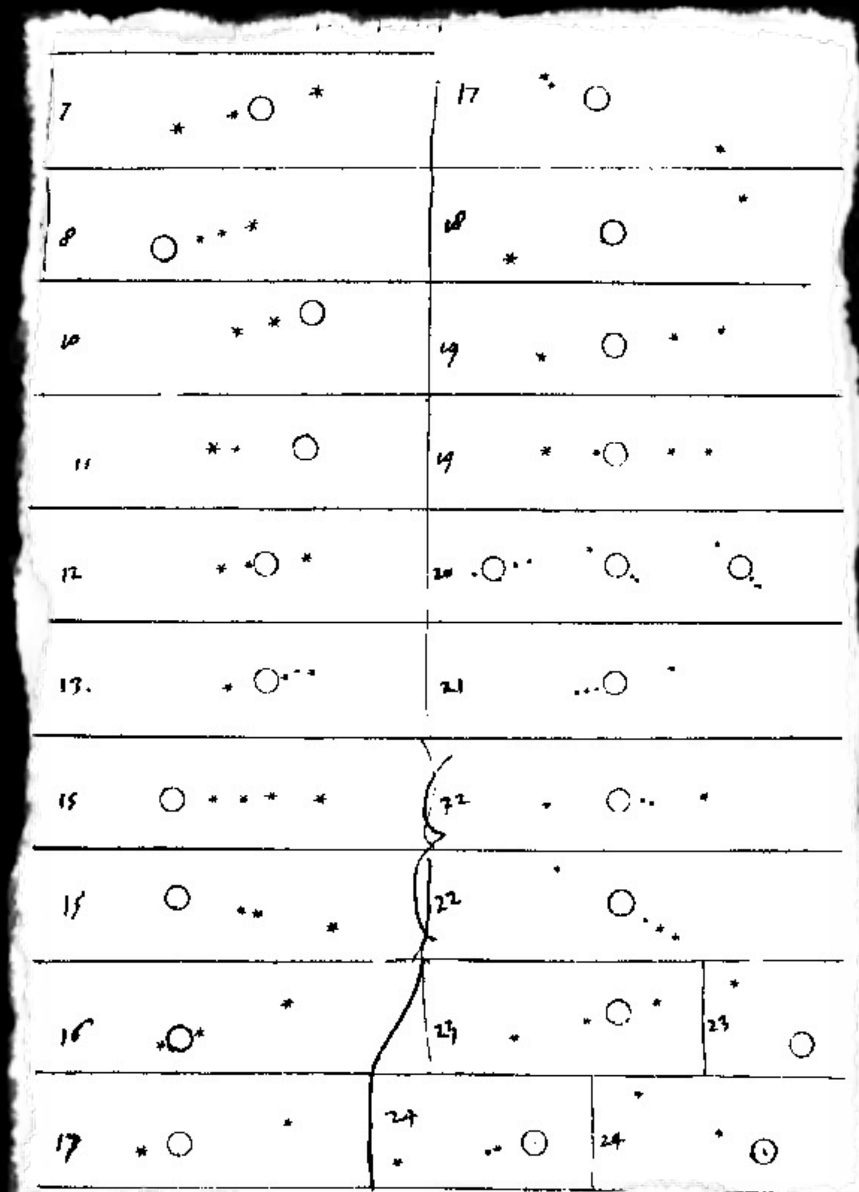


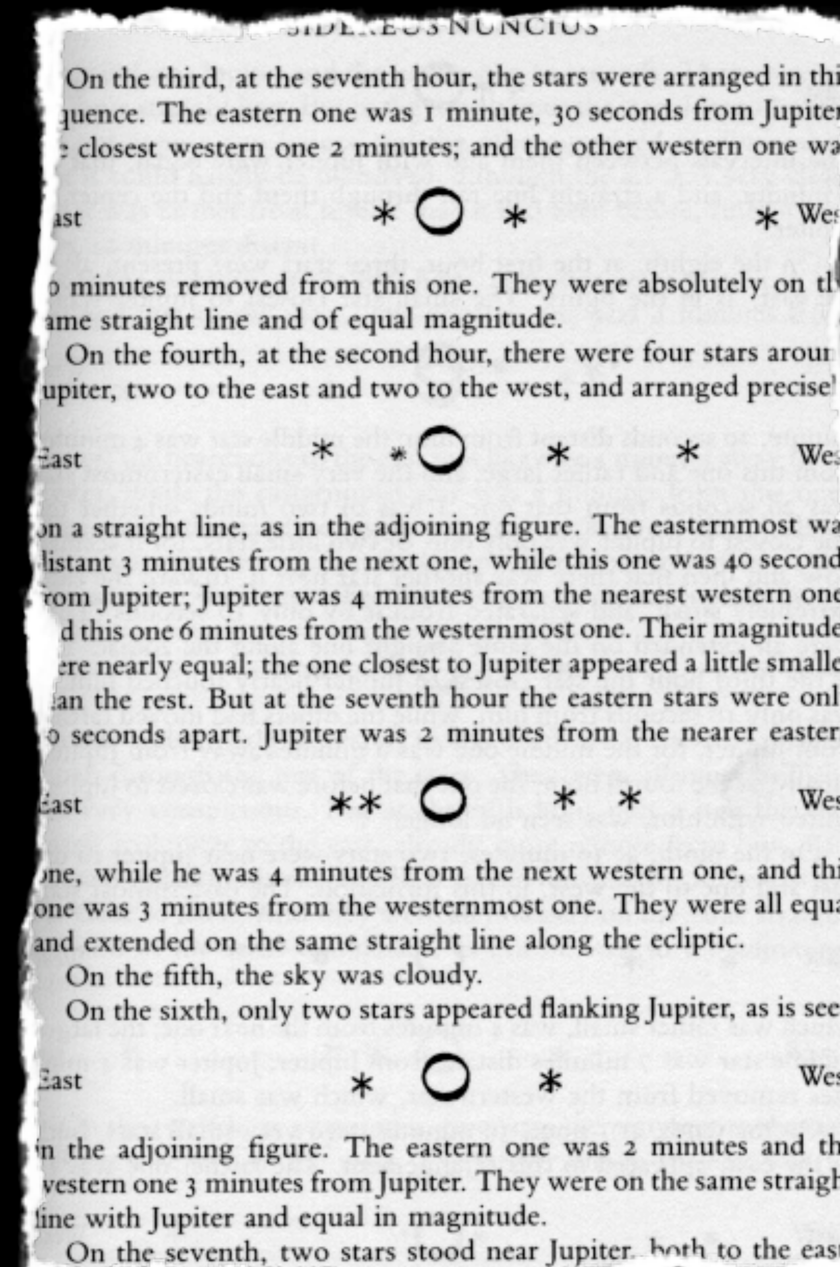


adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

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 smaller
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

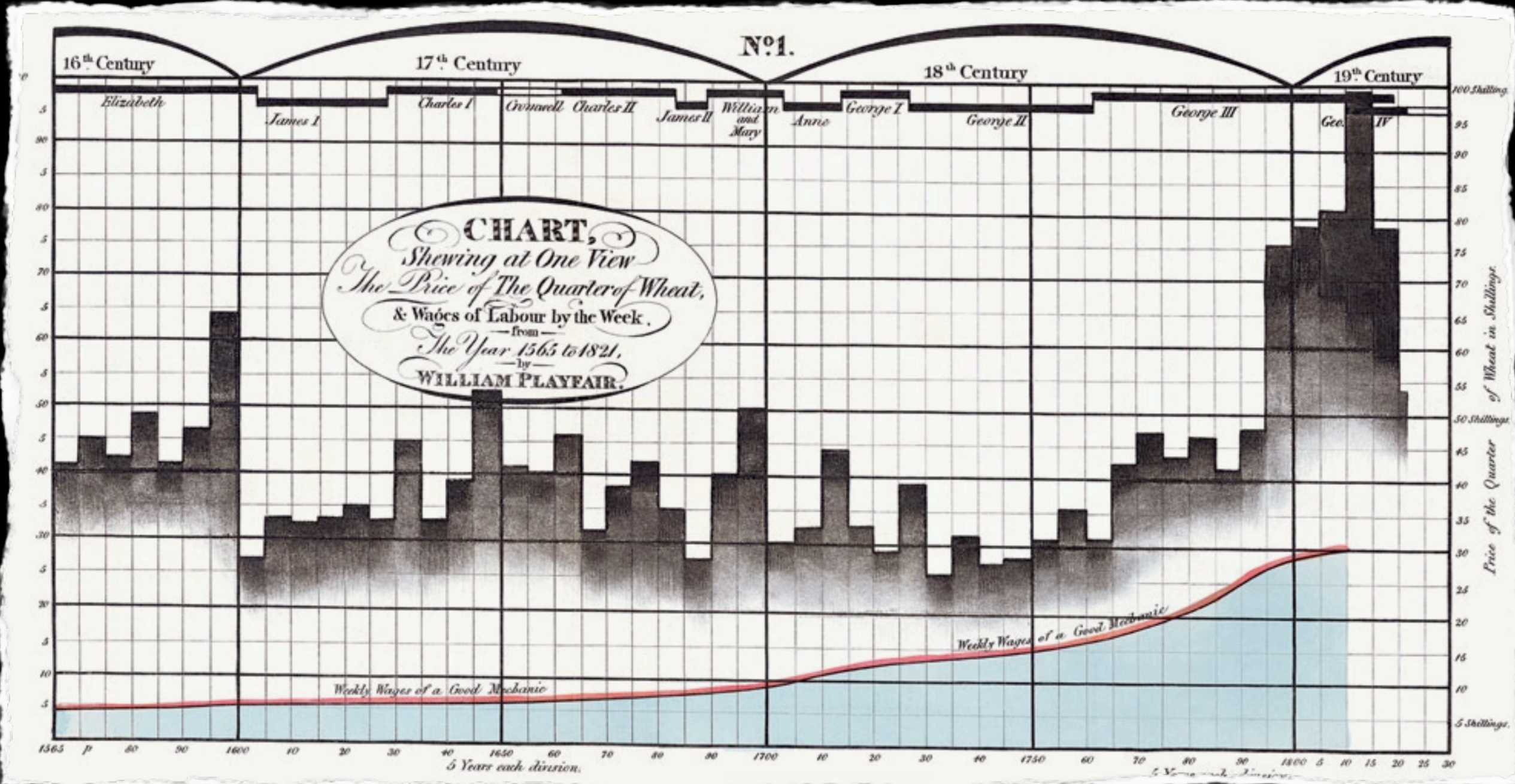
in the adjoining figure. The eastern one was 2 minutes and the
western one 3 minutes from Jupiter. They were on the same straight
line with Jupiter and equal in magnitude.

On the seventh, two stars stood near Jupiter, both to the east

Notes for & re-productions of Siderius Nuncius

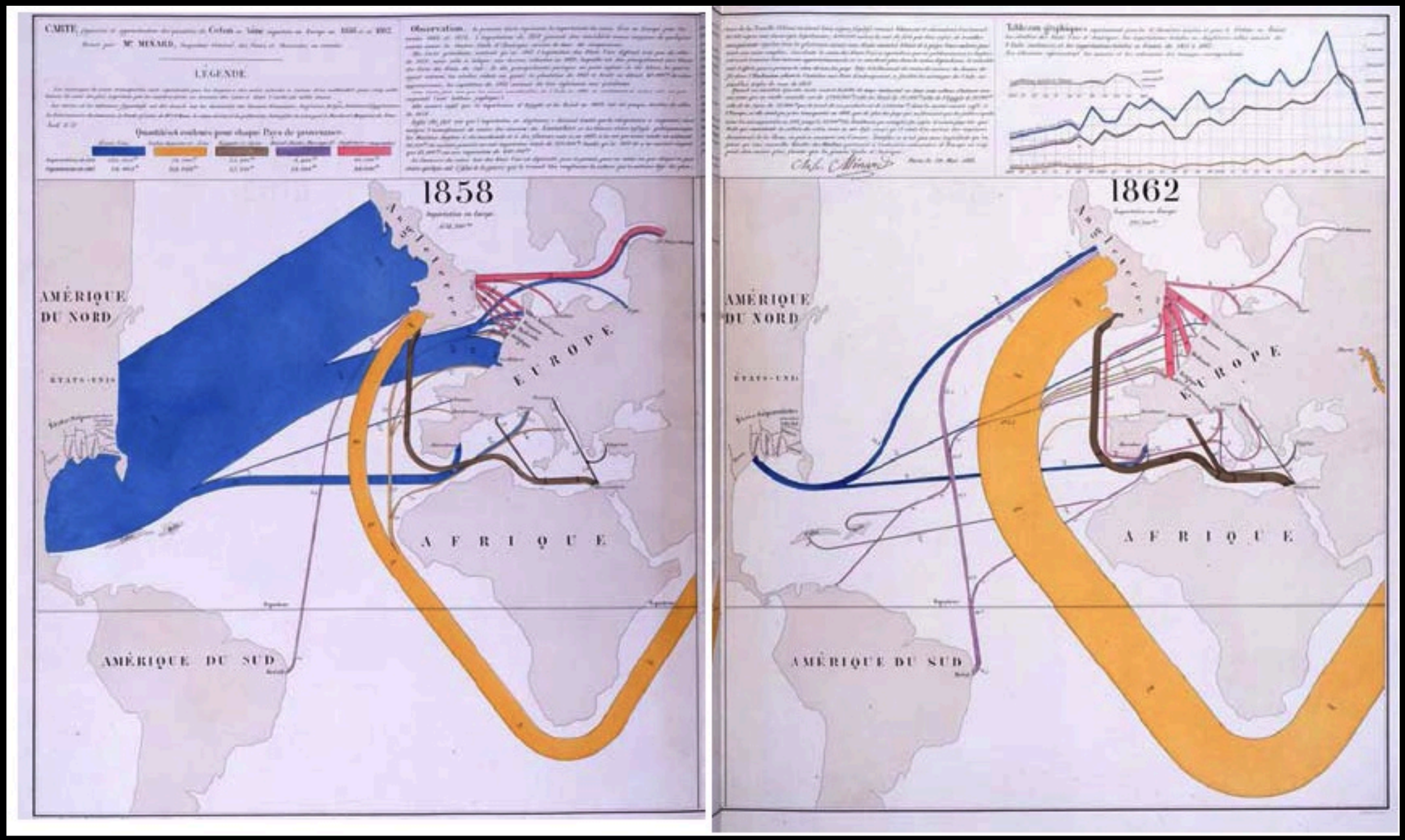
Playfair, lithography

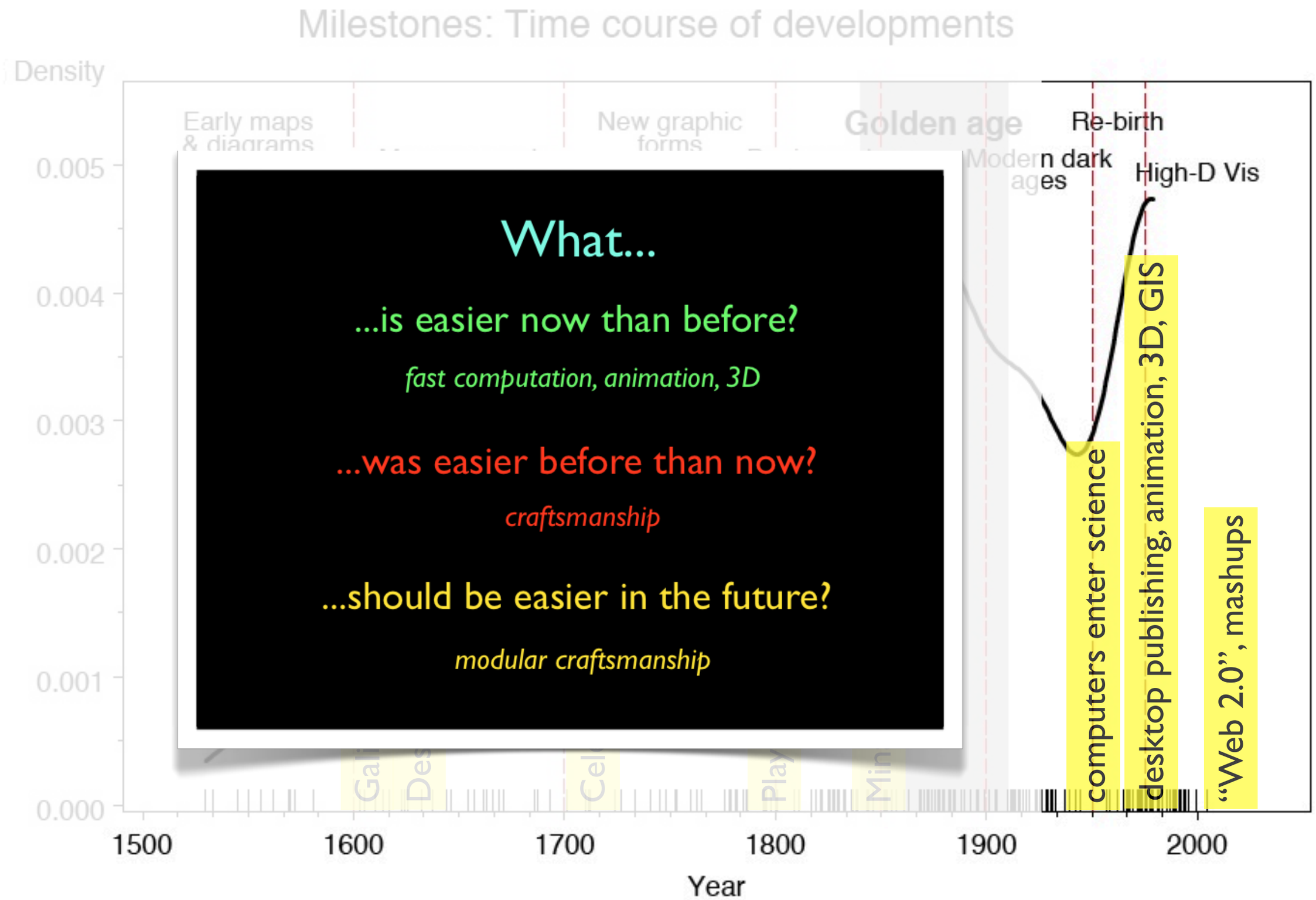
William Playfair
(1759-1823)



Minard, color lithography

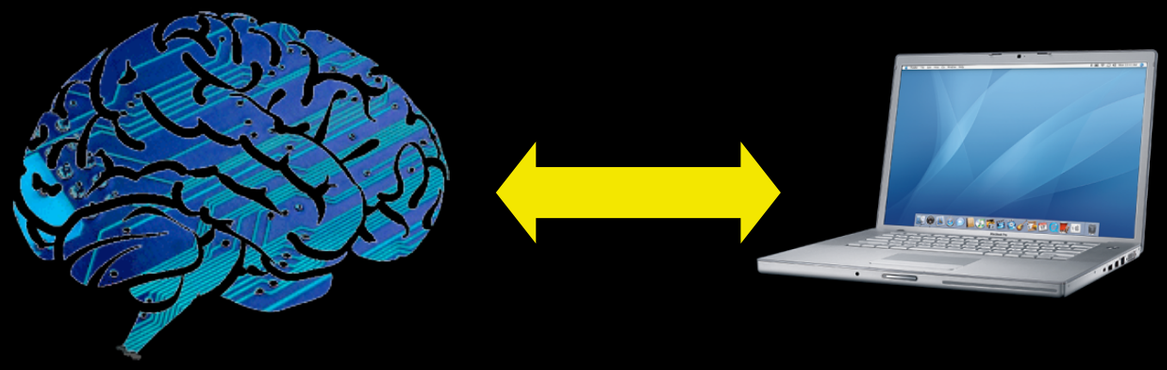
Charles Joseph Minard, in color
(1781-1870)





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)

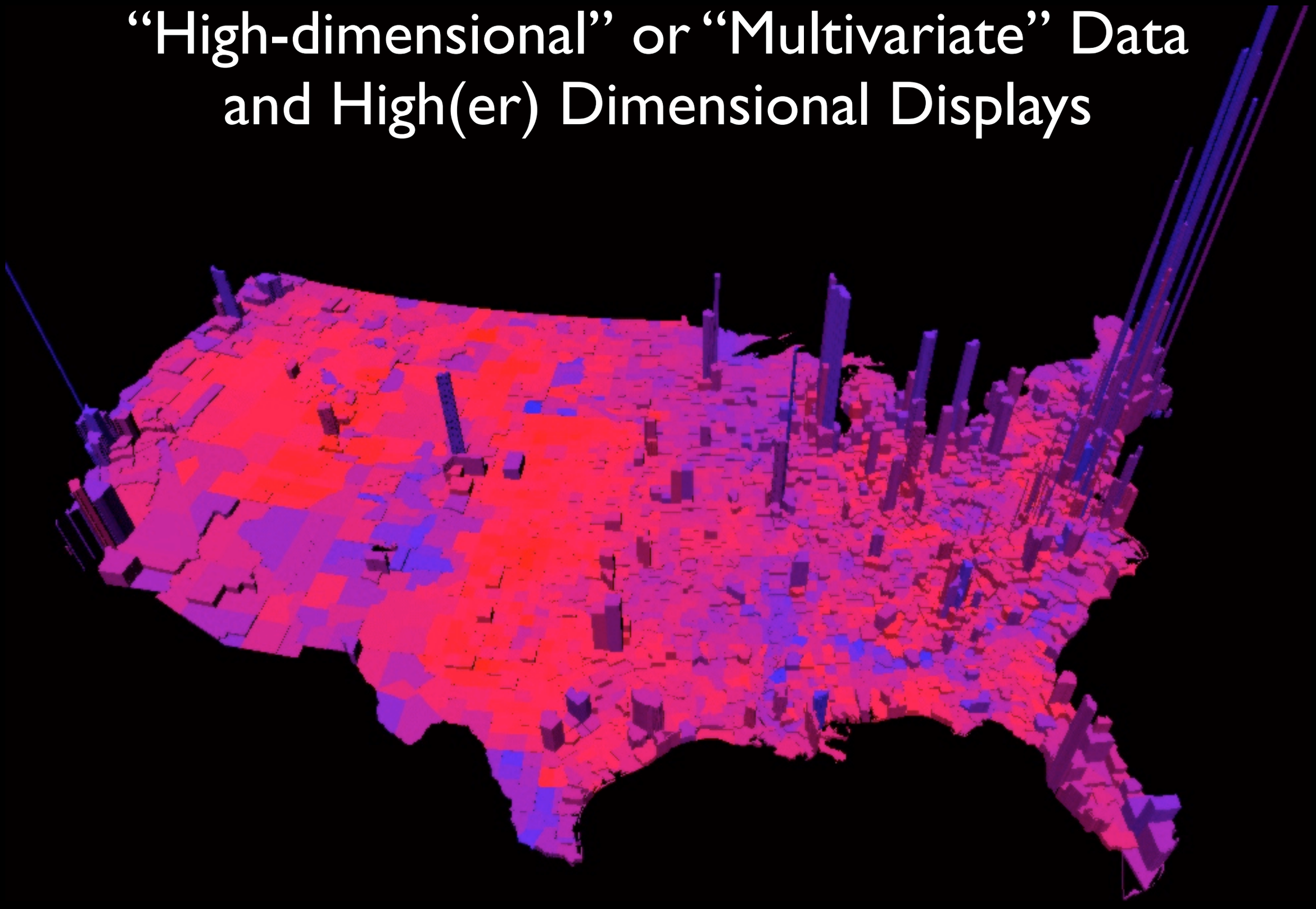
Simulation Design

Statistics Design

Data Exploration (Visualization)

Data • Dimensions • Display

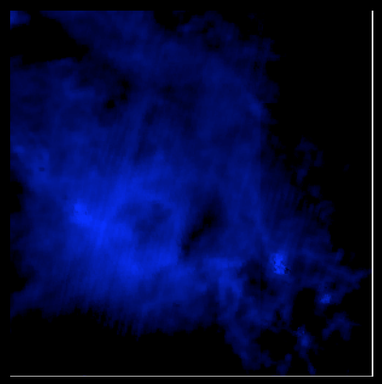
"High-dimensional" or "Multivariate" Data and High(er) Dimensional Displays



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

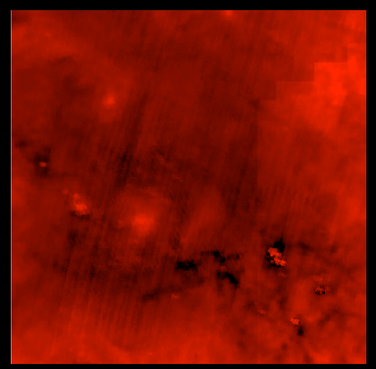
Note: Computers make this Easy

Column Density

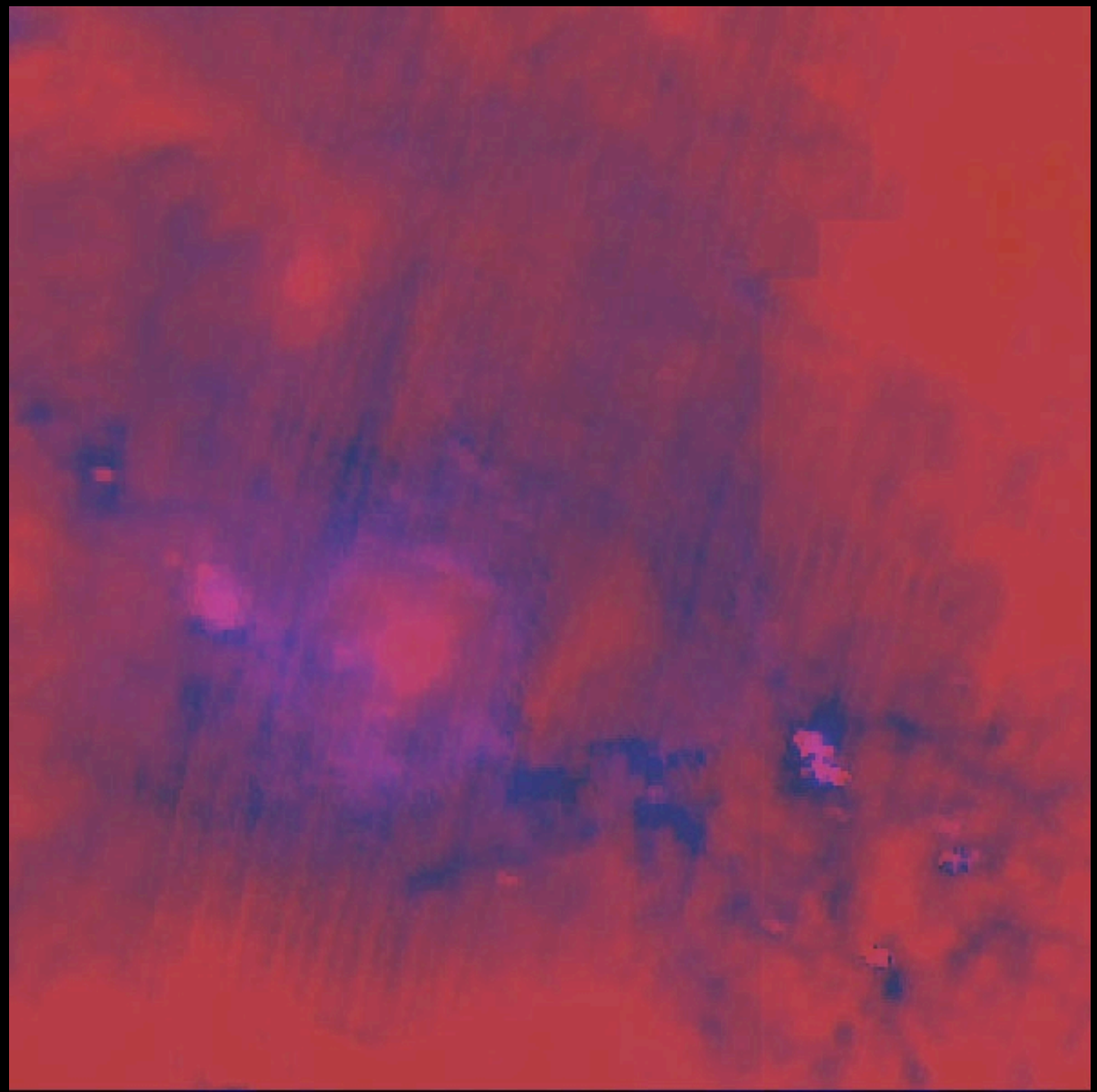


+

Temperature

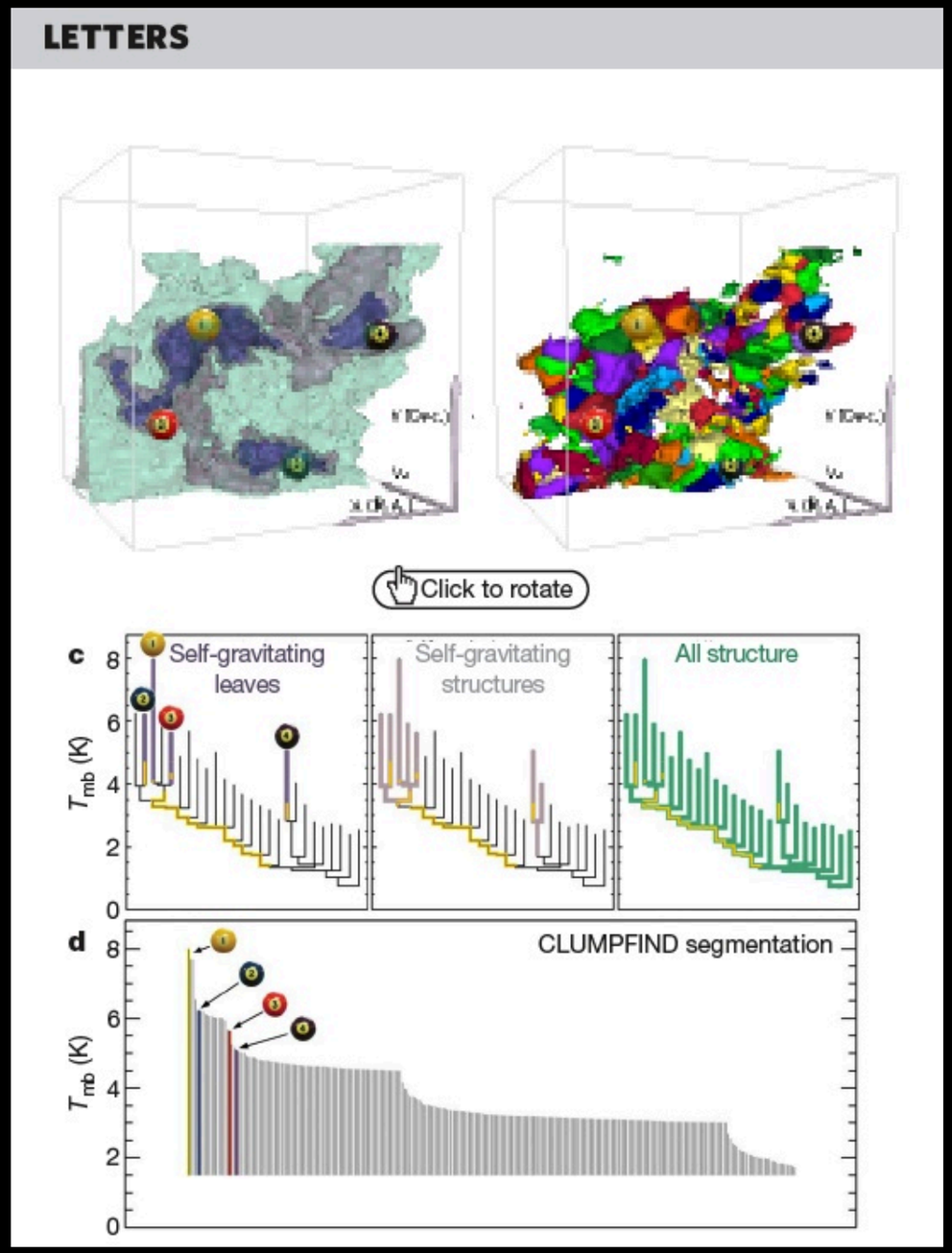


=

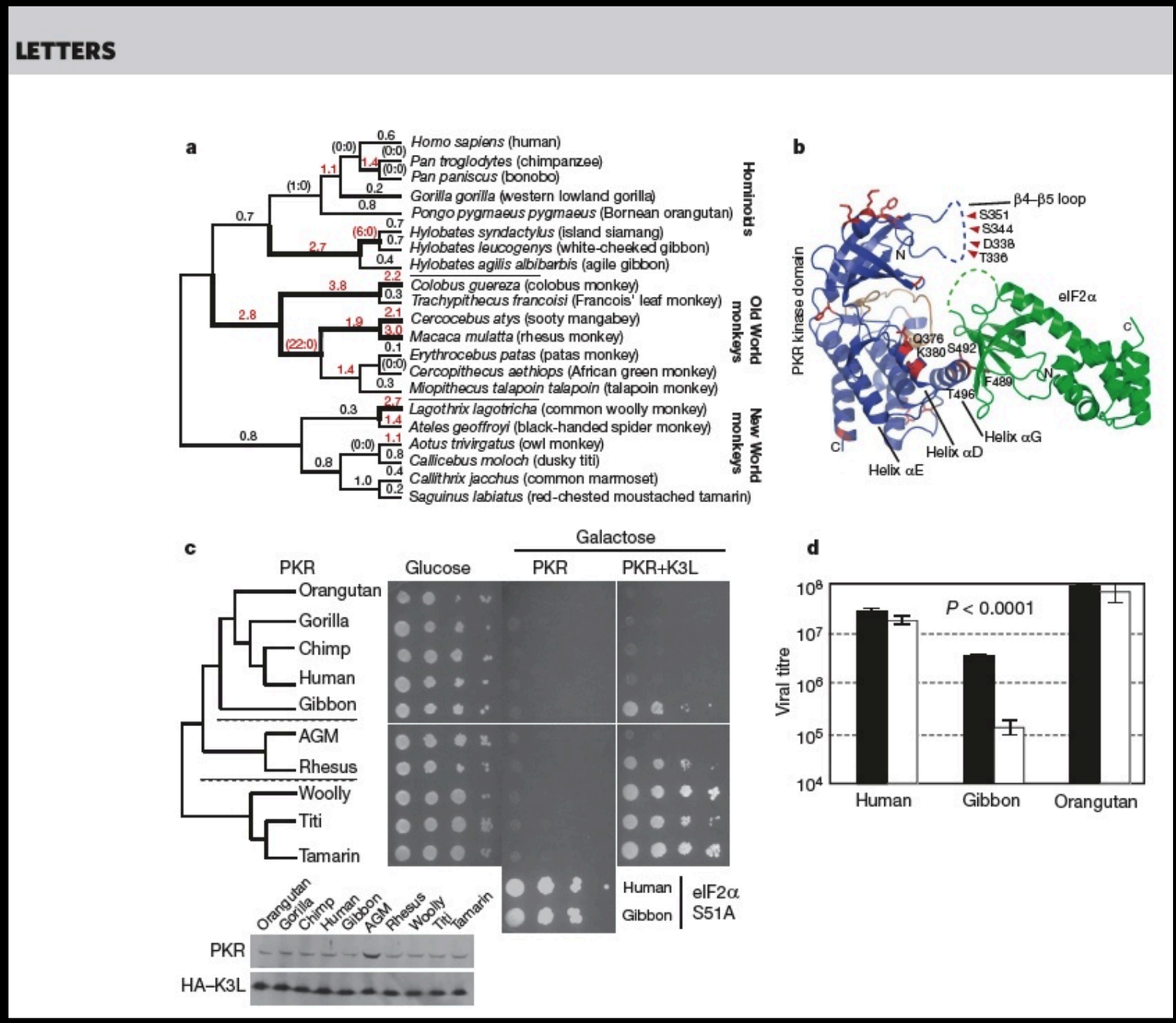


“High-dimensional” or “Multivariate” Data

(Astronomy=Biology)



Goodman et al. *Nature*, 2009



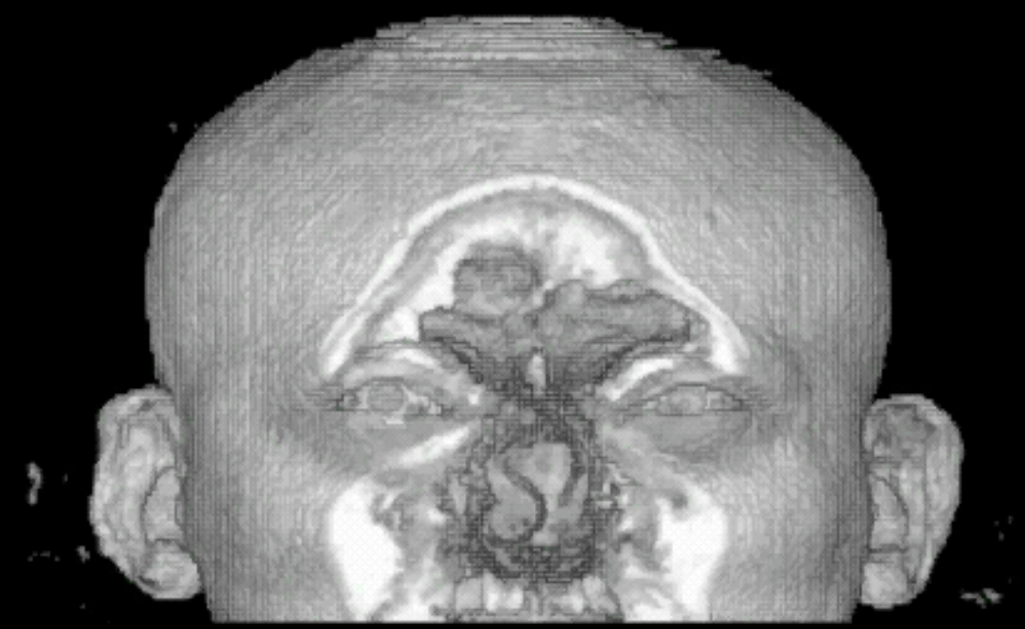
Elde et al. *Nature*, 2008

What about Animation?

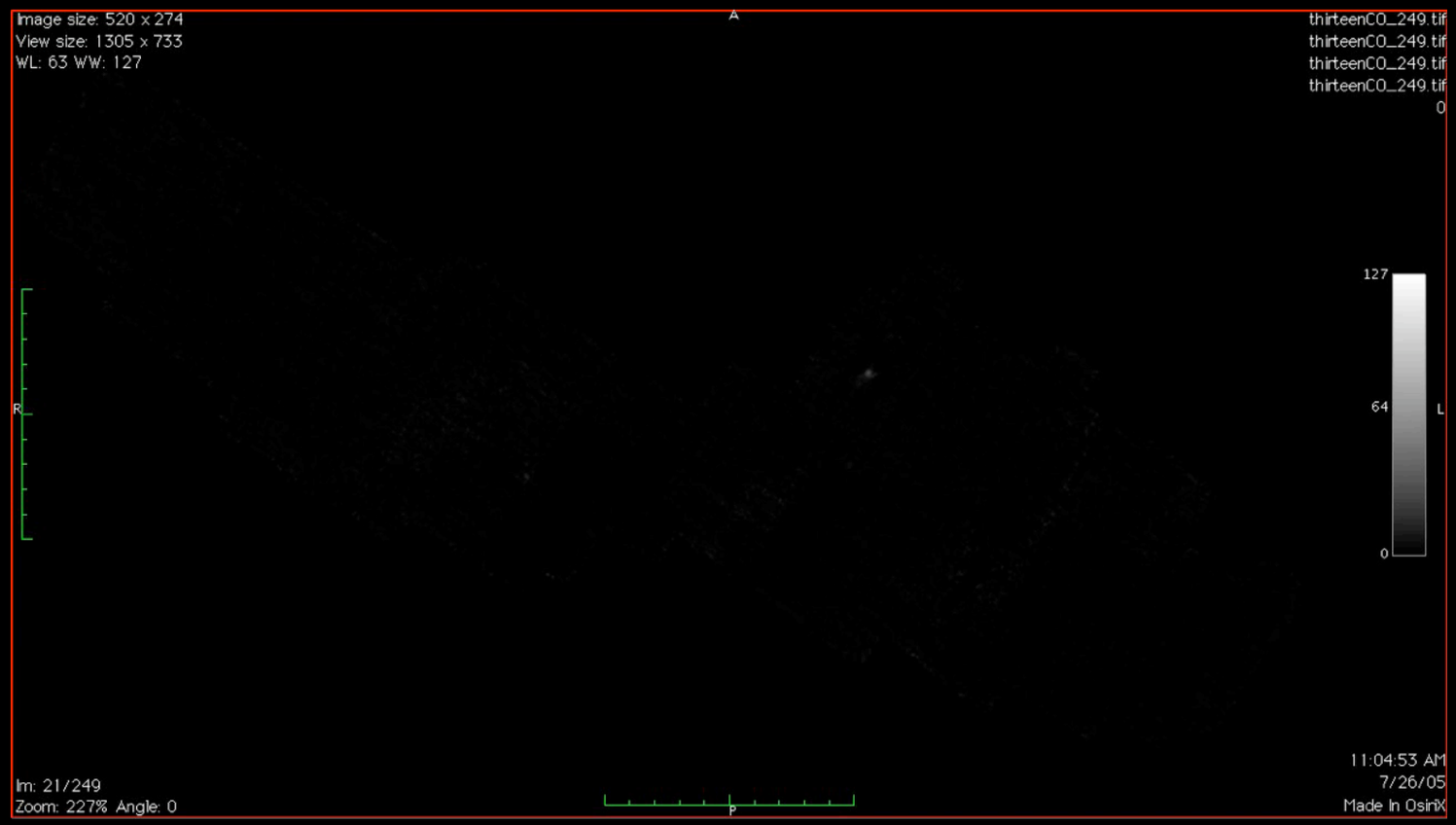
How many dimensions at once?

Can/should time (animation) substitute for dimensions?

"KEITH"



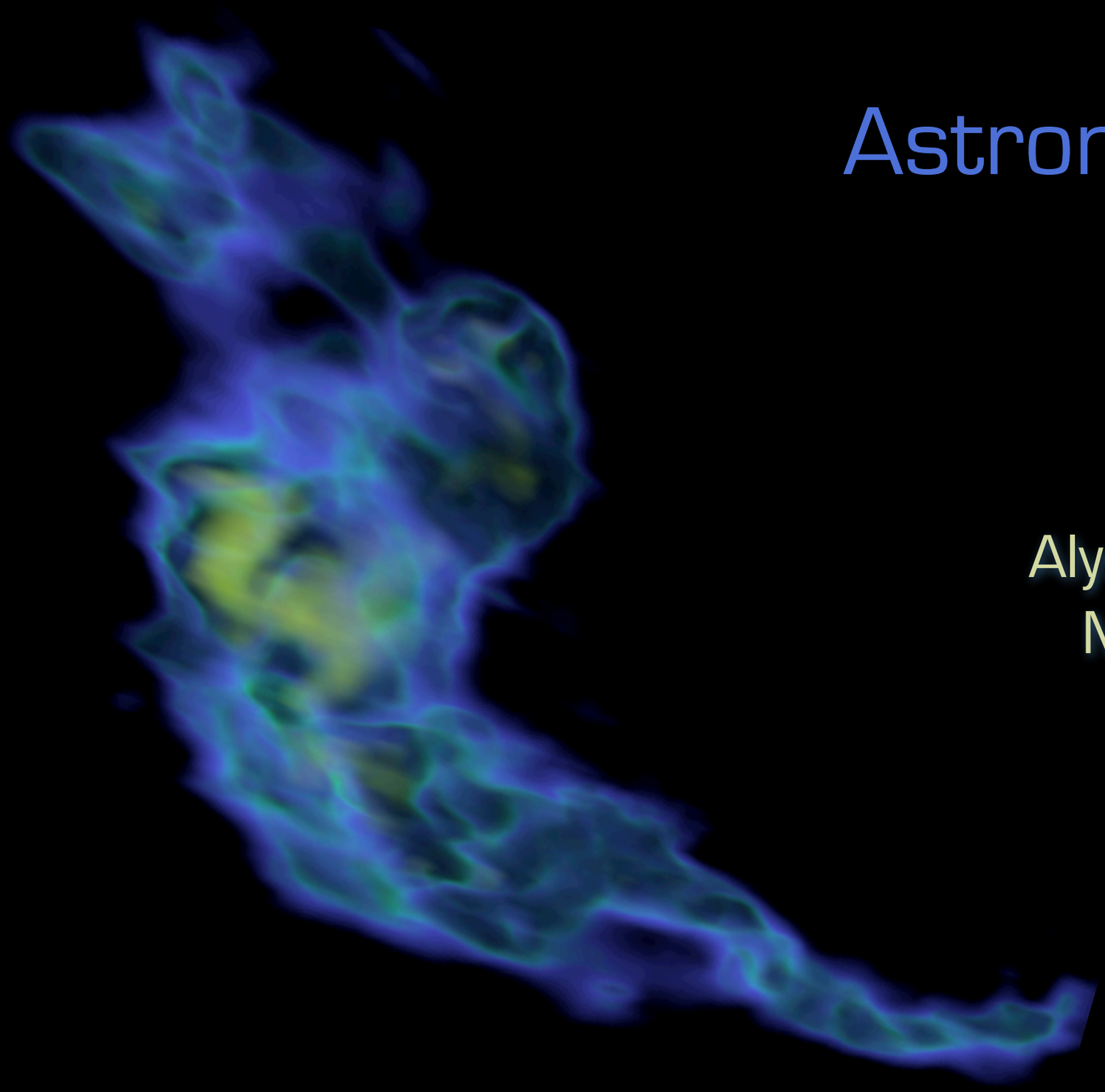
"PERSEUS"



"z" is depth into head

"z" is line-of-sight velocity

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



Astronomical Medicine

am.iic.harvard.edu

Alyssa Goodman (IIC/CfA/FAS)

Michael Halle (IIC/SPL/HMS)

Ron Kikinis (SPL/HMS)

Douglas Alan (IIC)

Michelle Borkin (IIC)

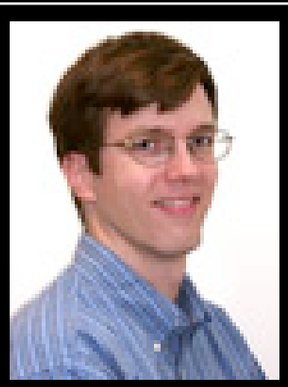
Jens Kauffmann (CfA/IIC)

Erik Rosolowsky (CfA)

Nick Holliman (U. Durham)

The AstroMed Story

Computer Scientist

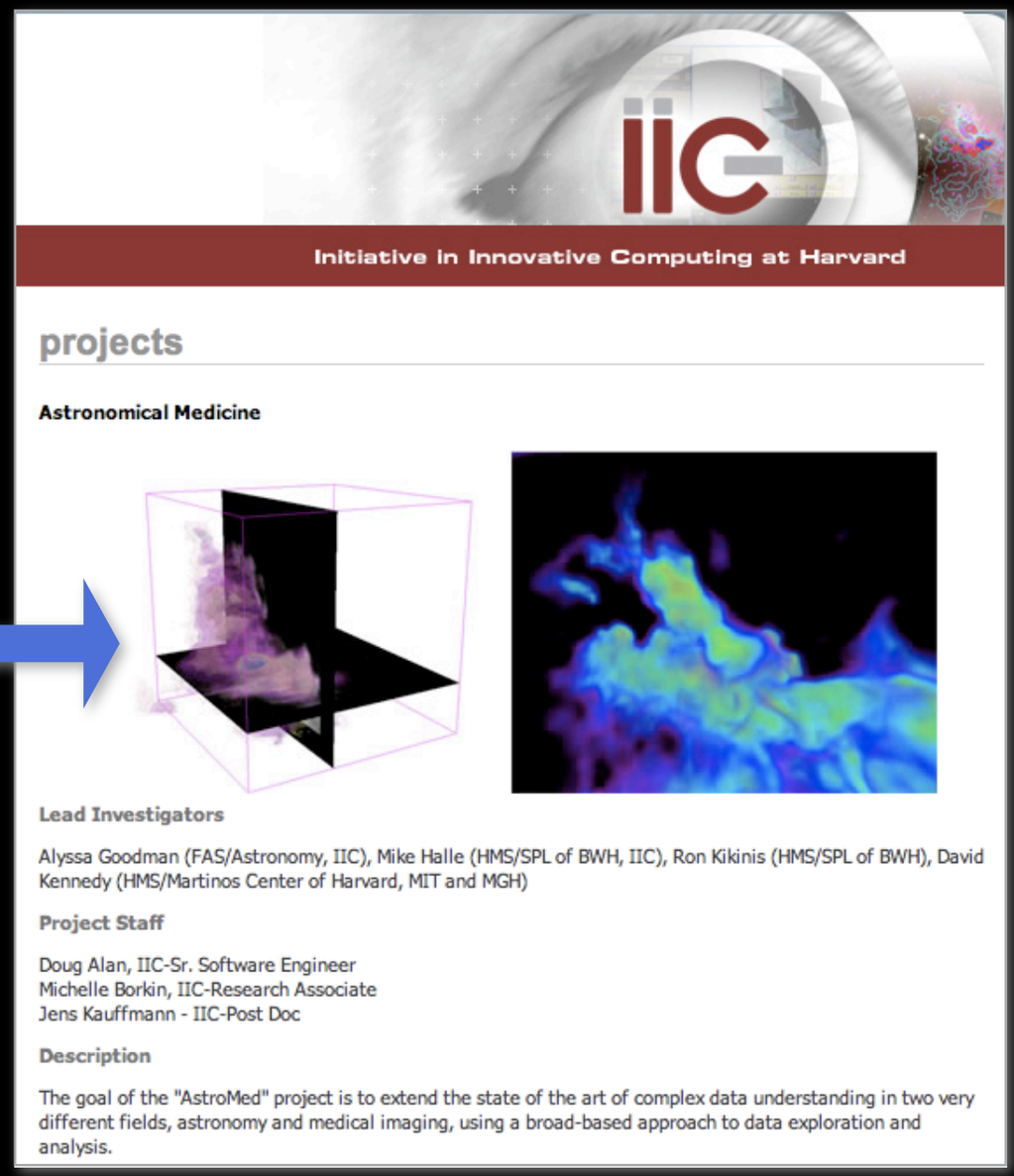


Computer Scientist

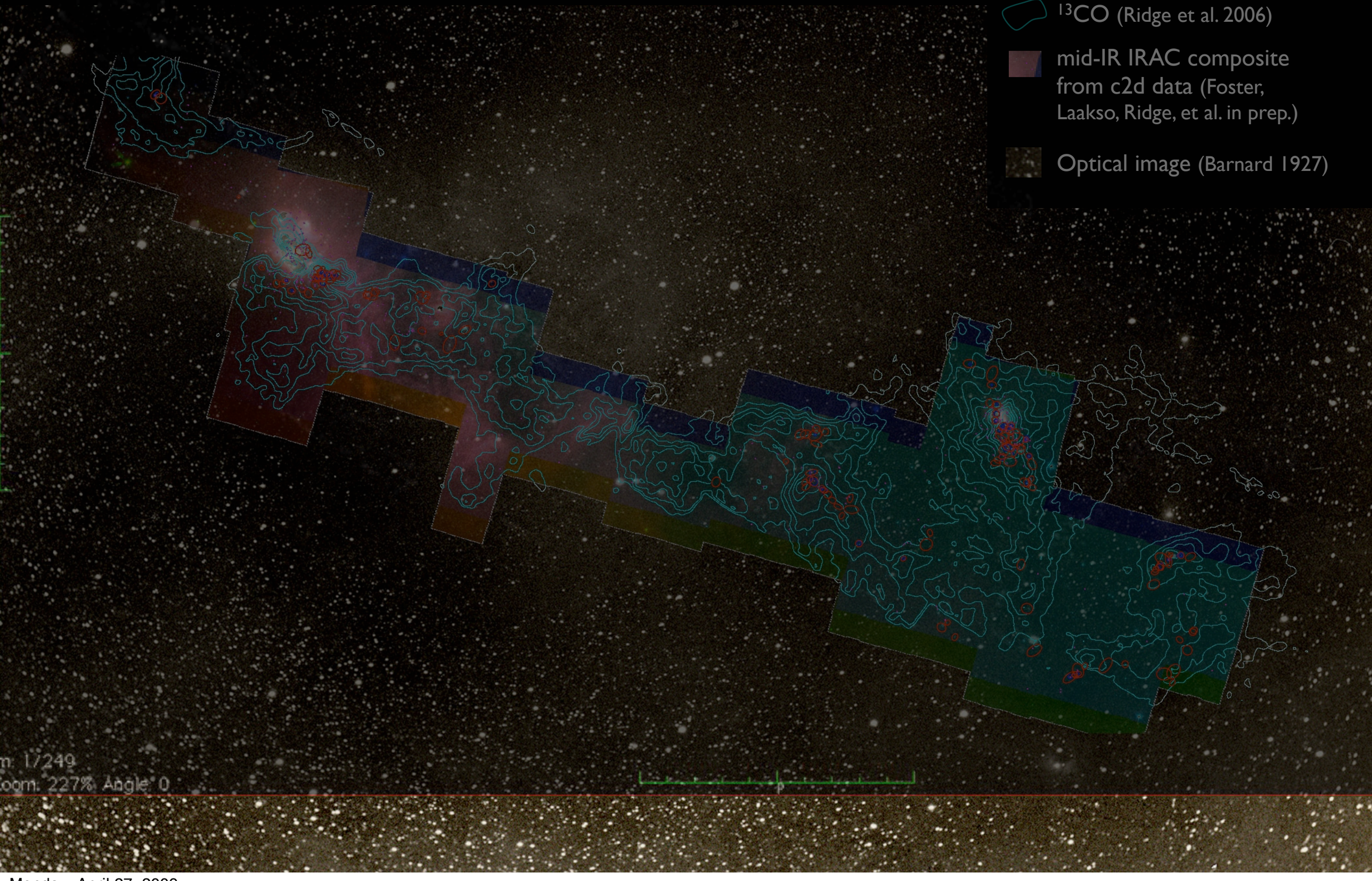
"Viz has failed the scientific community…"

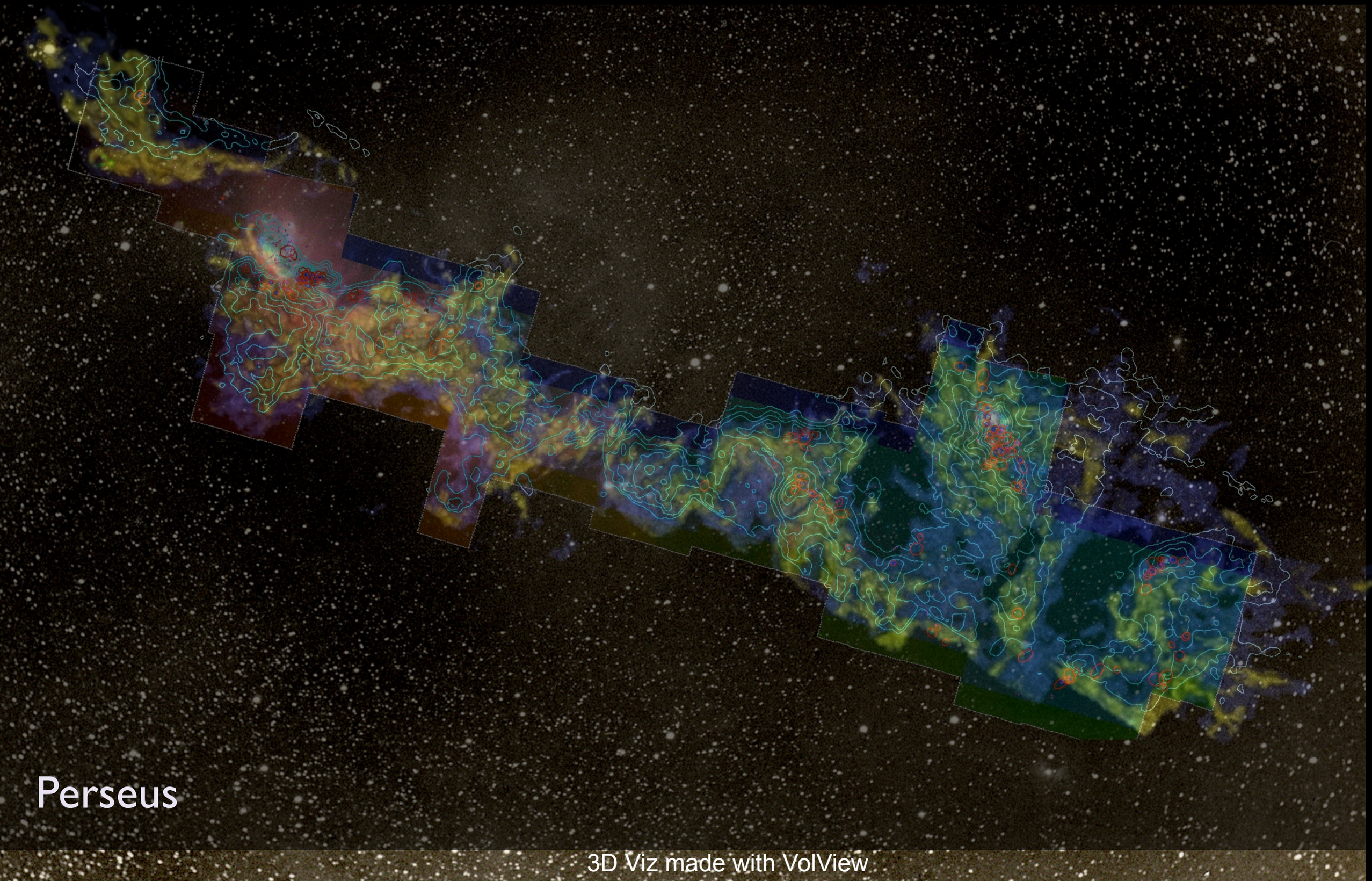
Astronomer

Unsuspecting Undergrad



COMPLETE Perseus





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)

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

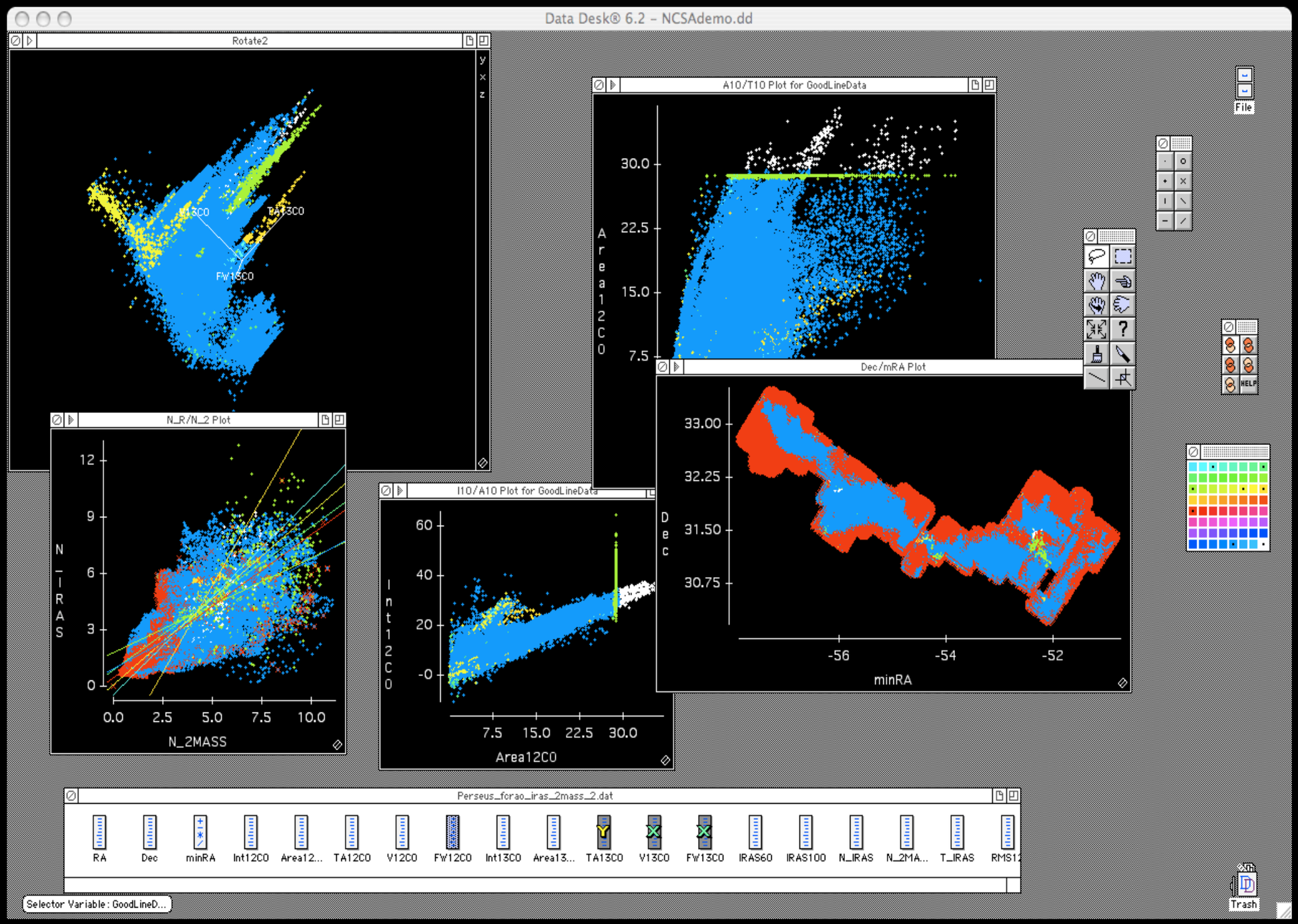
re-usable tools/mashups (Many Eyes, UFOmaps.com)

Unsolved Questions...

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

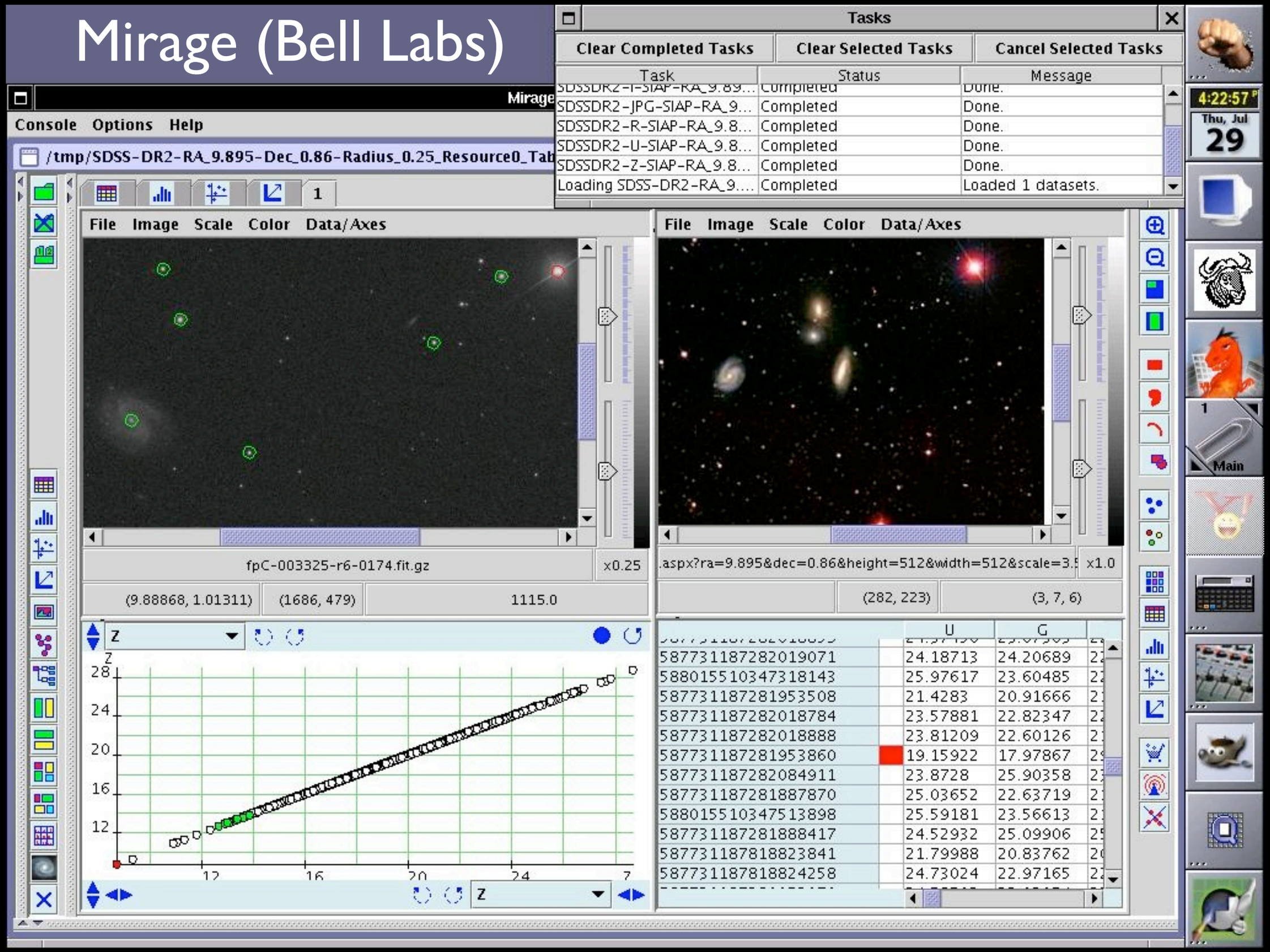
improved graphical representation of uncertainty

“Data Desk”



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

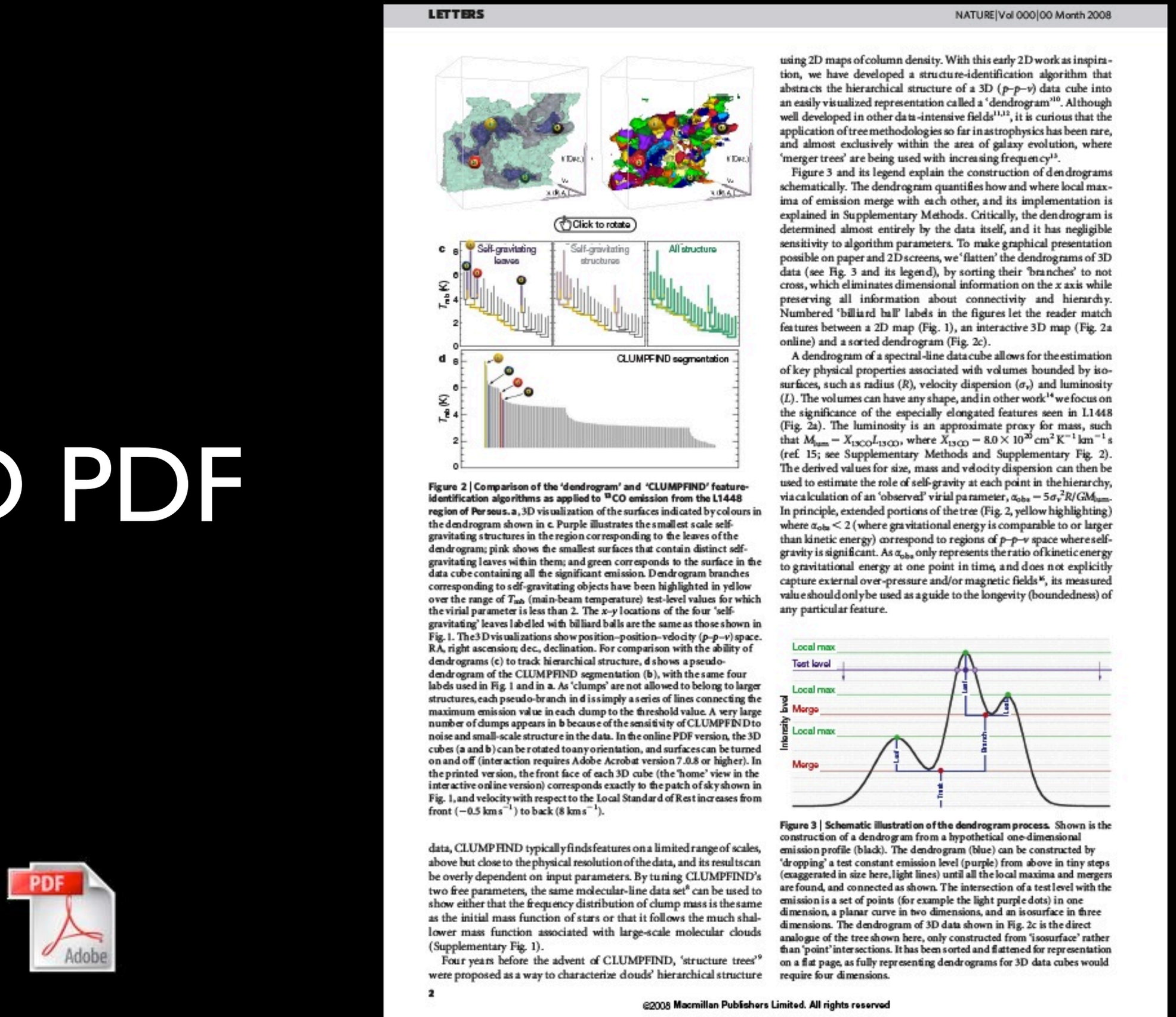
Mirage (Bell Labs)



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

display modes

3D PDF



LETTERS NATURE|Vol 000|00 Month 2008

Click to rotate

Self-gravitating leaves
Self-gravitating structures
All structure
CLUMPFIND segmentation
T_{mb} (K)

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 km s^{-1}) to back (8 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 years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure 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 $X_{13CO} = 8.0 \times 10^{20}$ 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.

2

Off the desktop

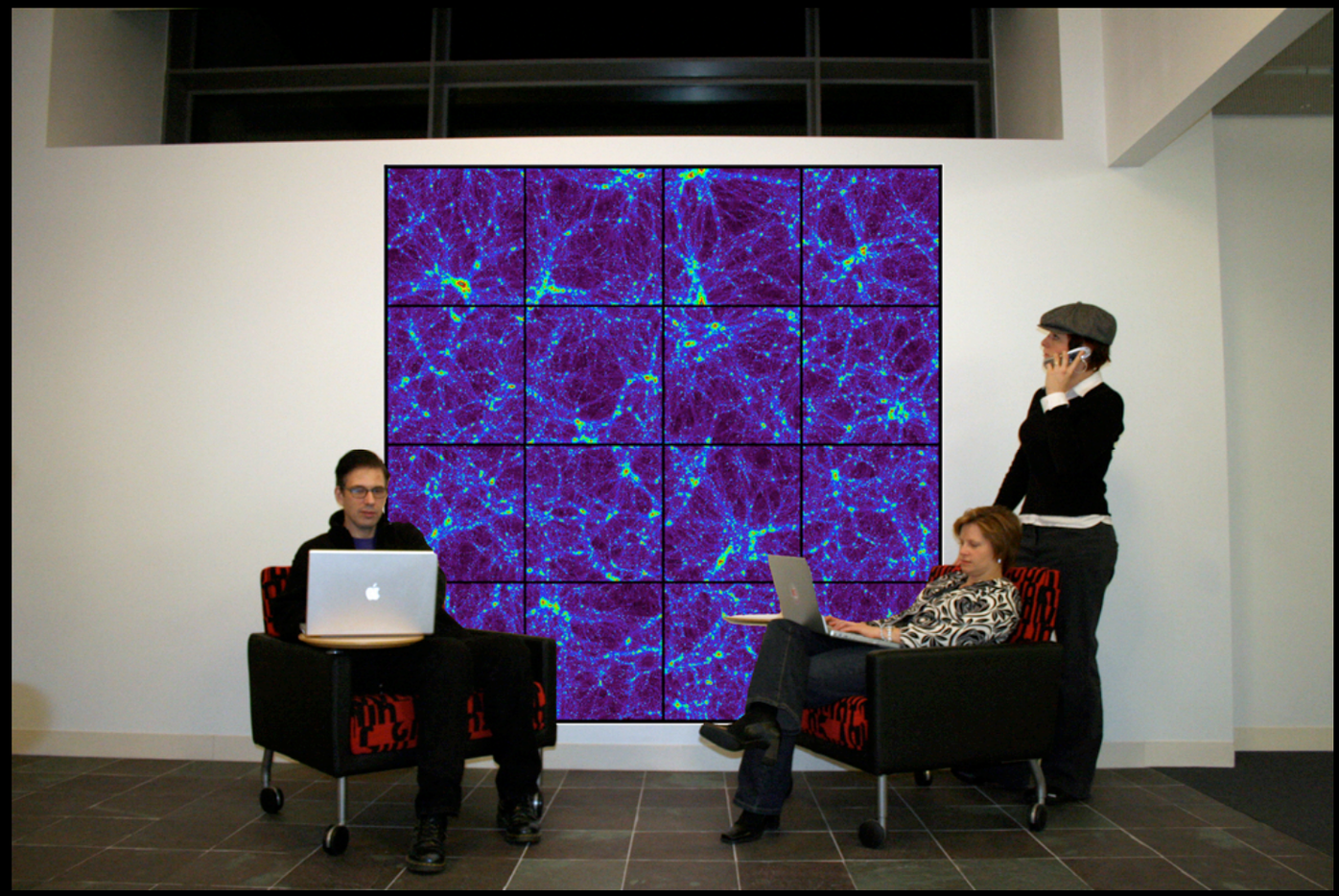
http://iic.harvard.edu/research/scientists-discovery-room-lab-sdr-lab

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

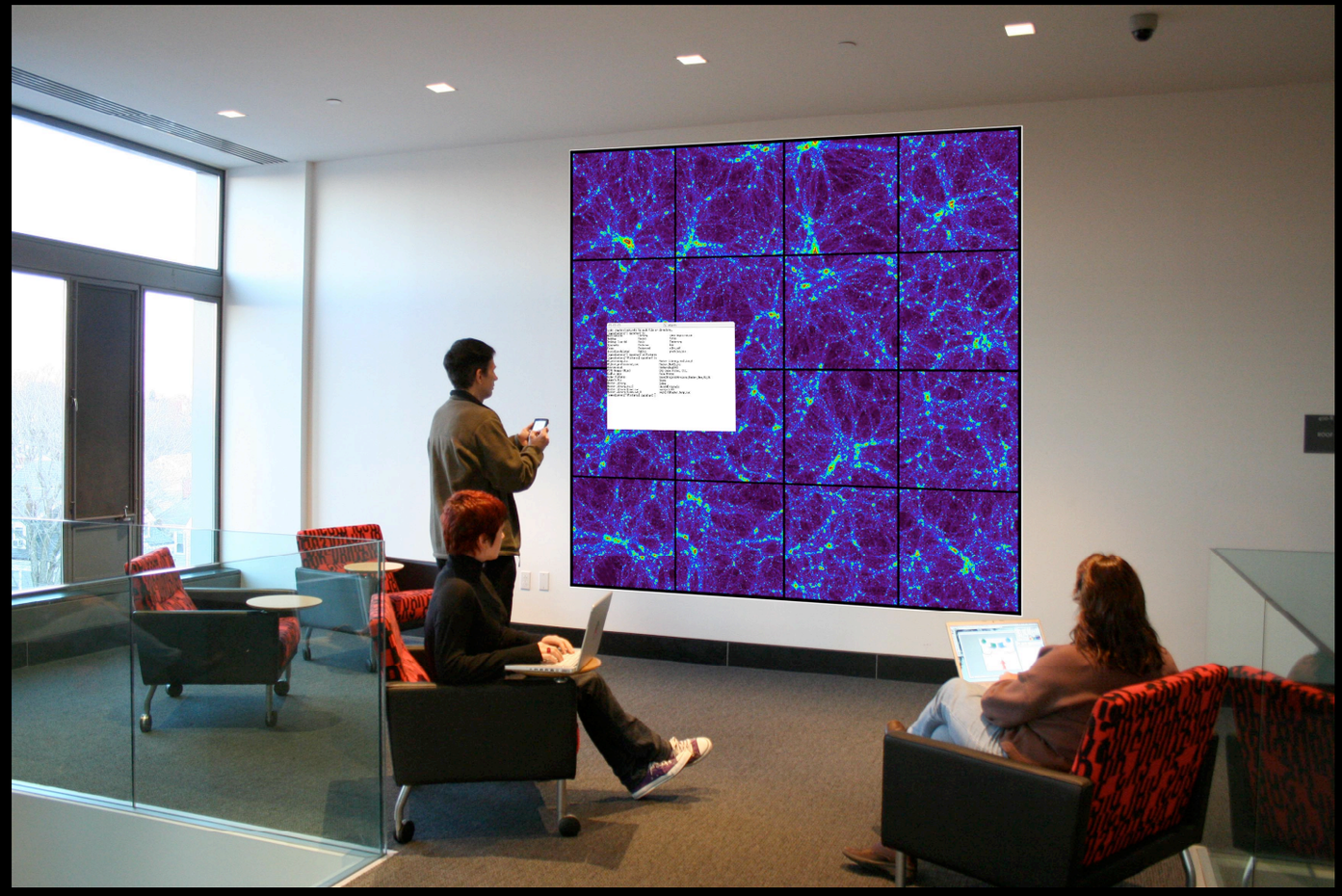
The Scientists' Discovery Room: Version 0.01

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

...why we must explain that...



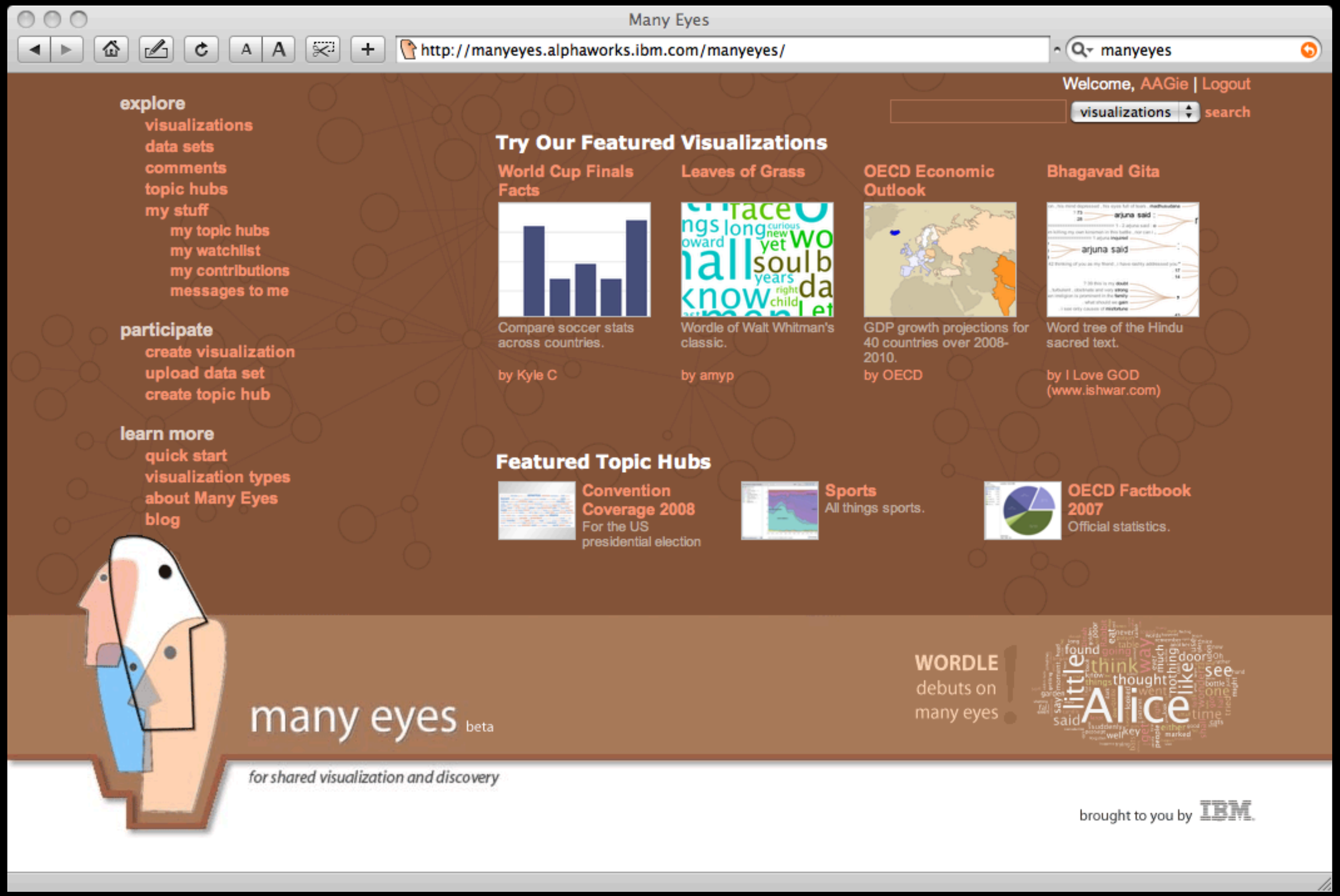
"This is not art."



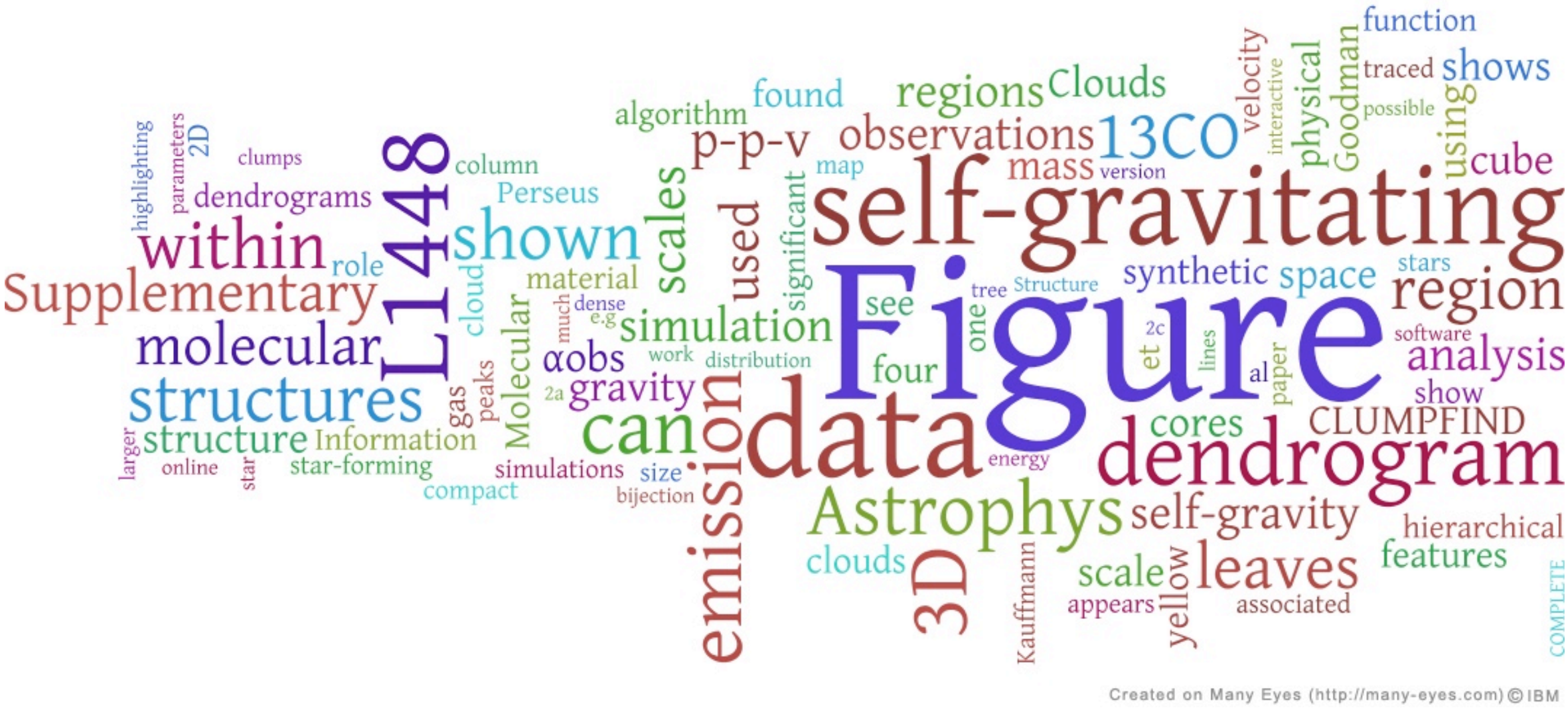
re-usable tools

Many Eyes:

Martin Wattenberg & Fernanda Viegas (IBM)

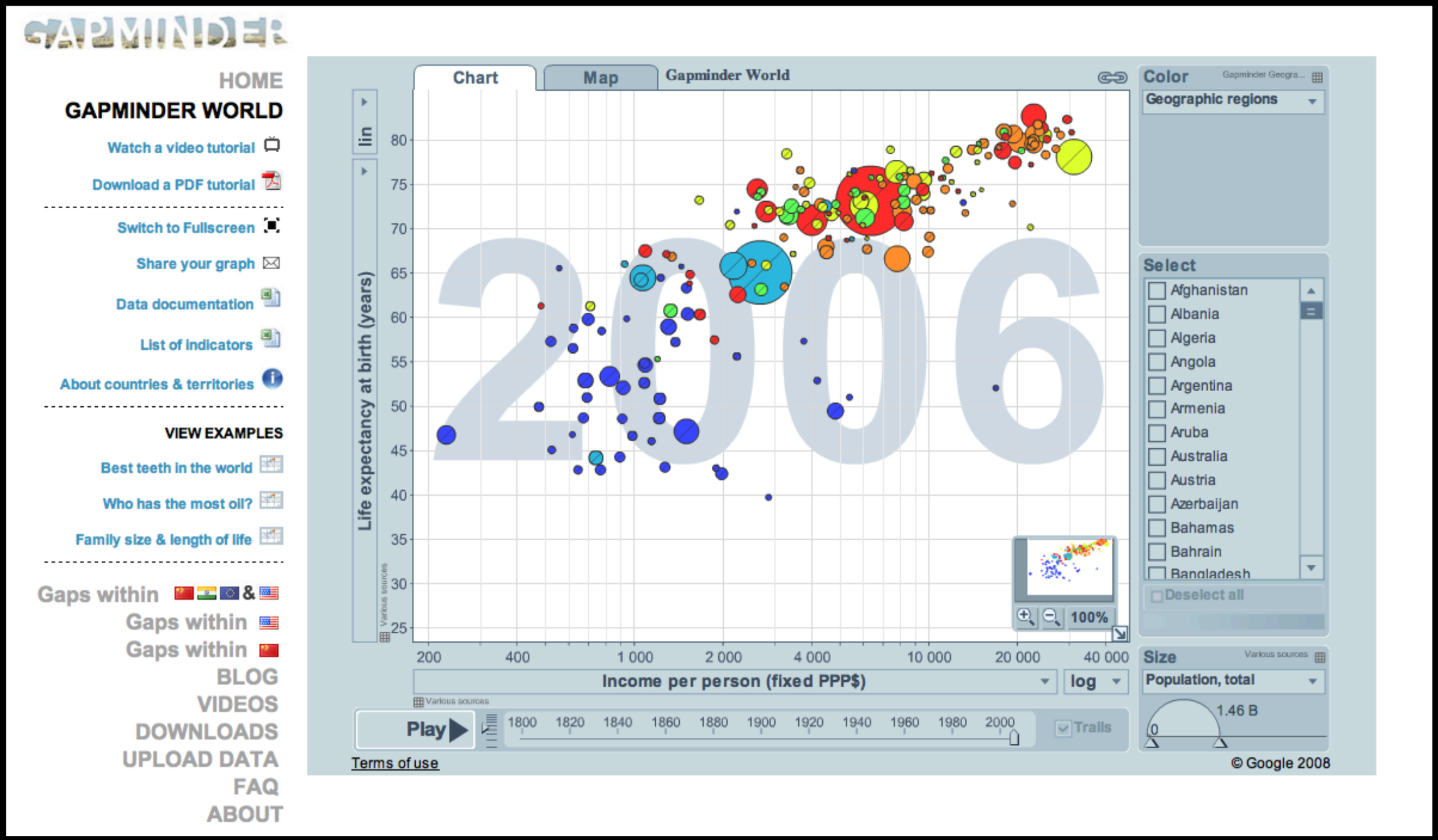


What Many Eyes thinks my "3D PDF" paper on self-gravity in hierarchical star-forming regions is about...



Gapminder:

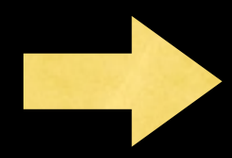
Re-usable, interactive graphical tools



http://gapminder.org

Facilitating re-usabilty...

Mashups, Resource Hubs, APIs

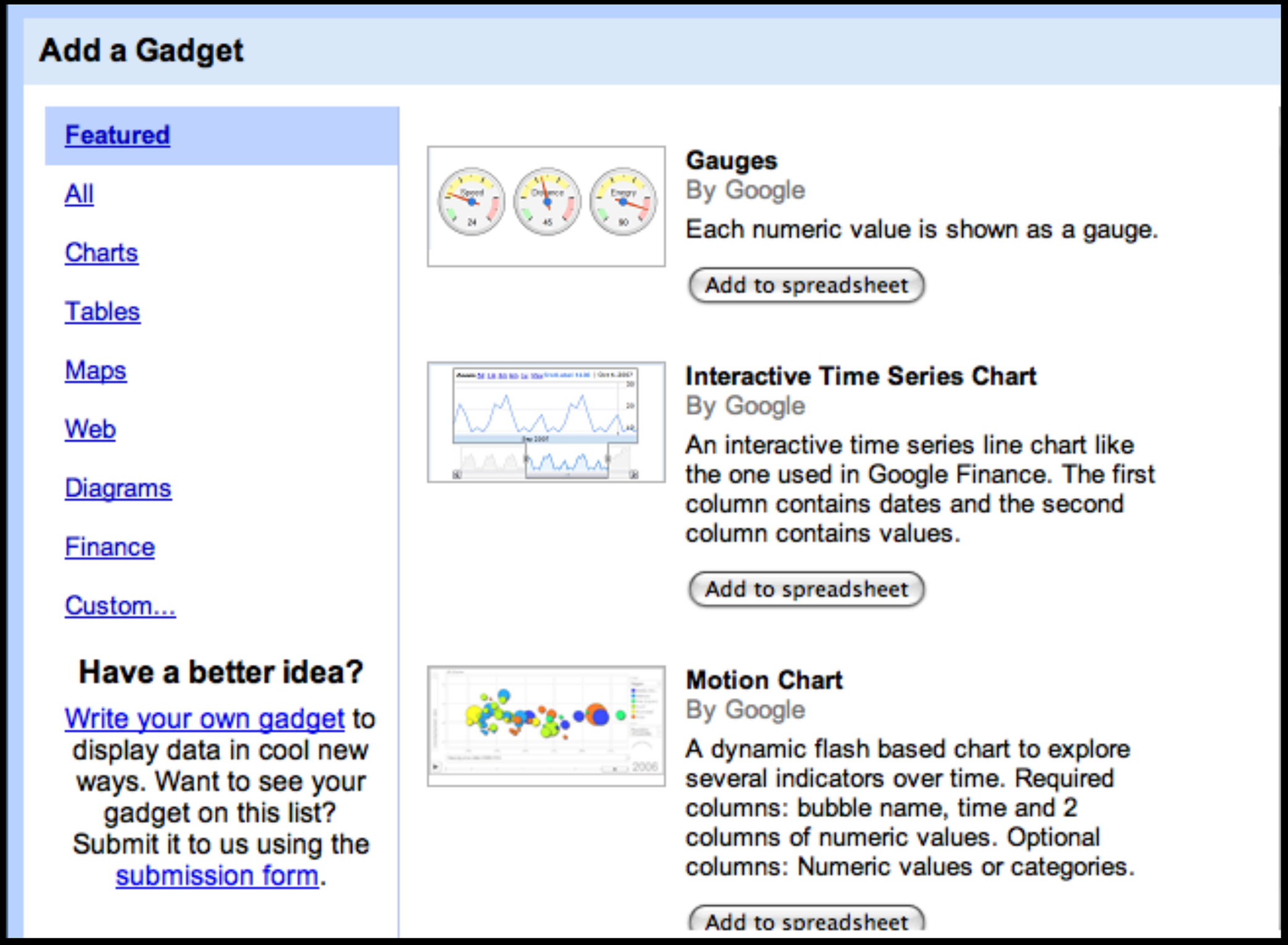


More on re-usabilty

Exemplar mashup c. 2006

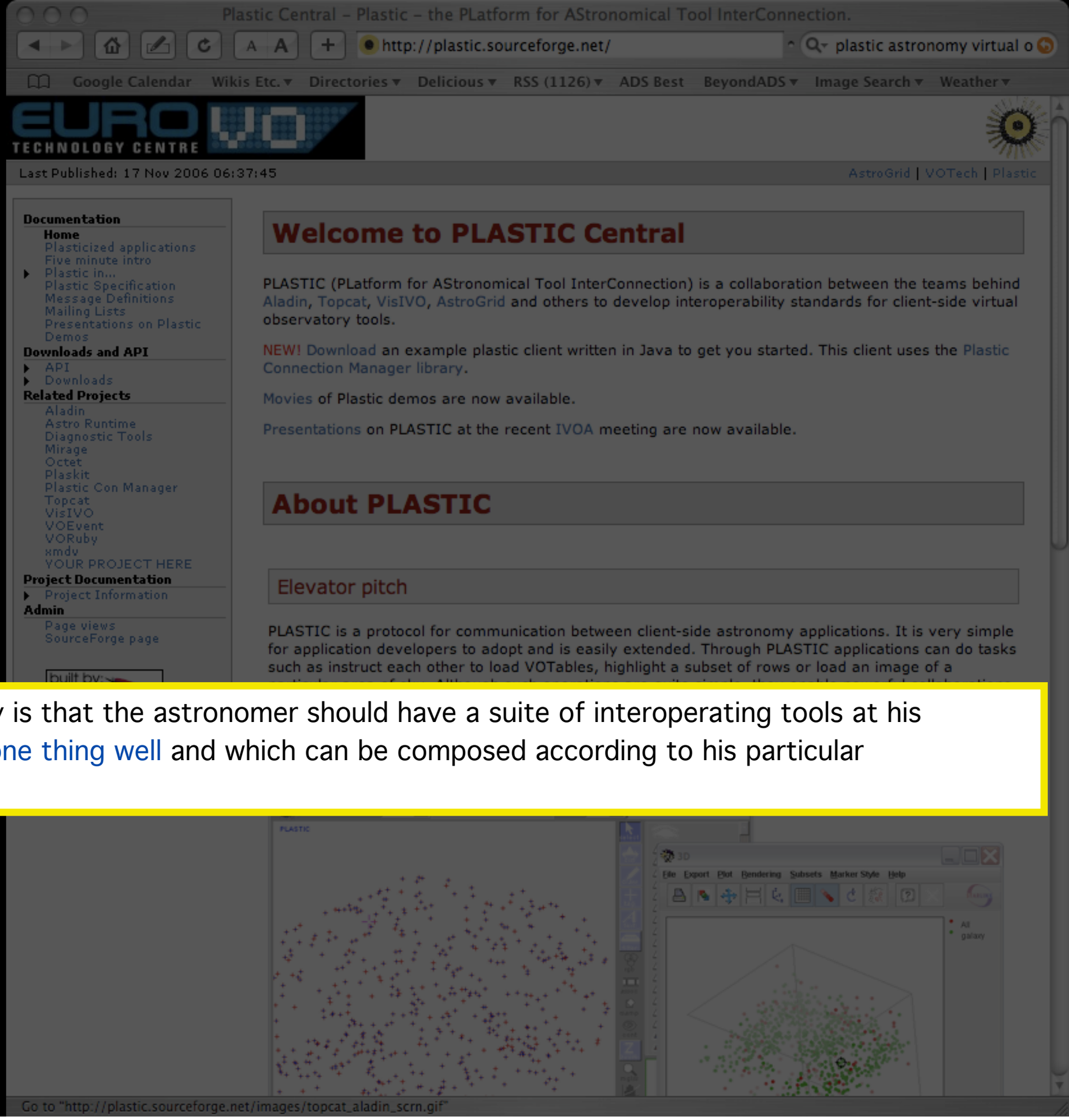
More on re-usabilty

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

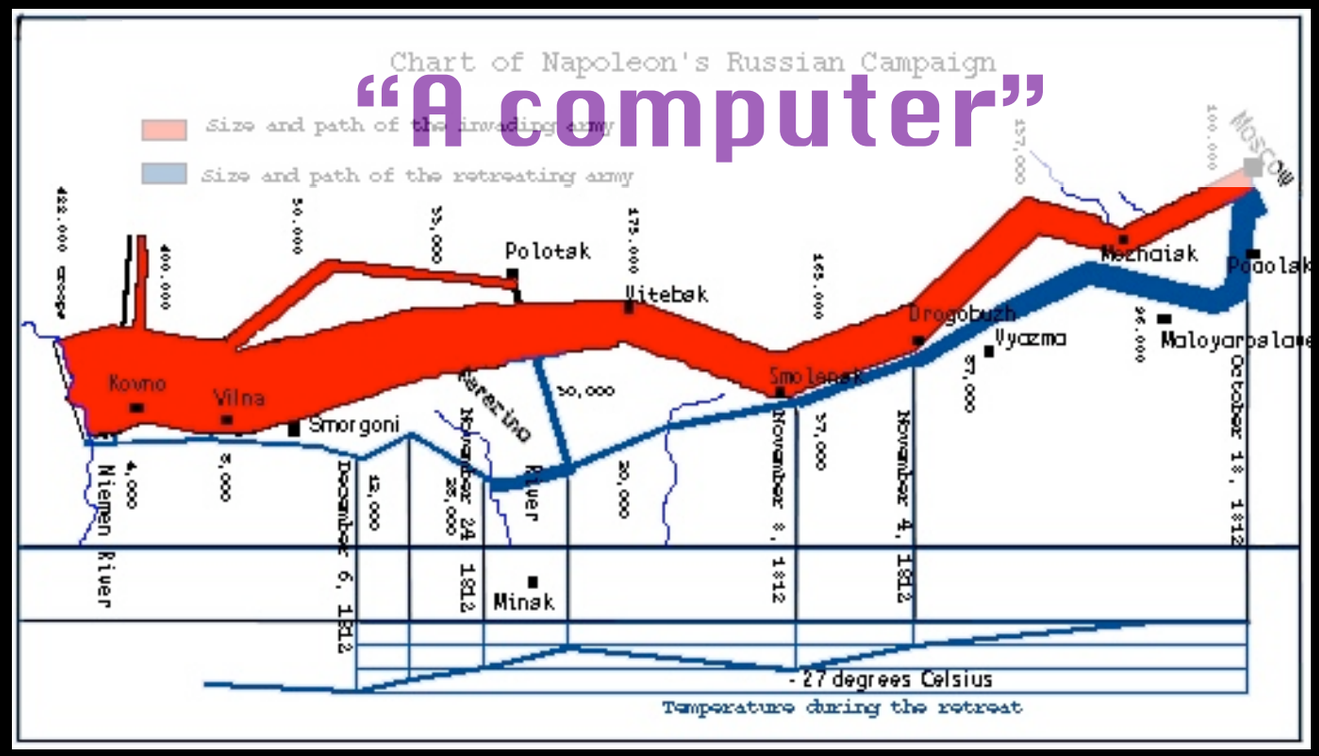
...was easier before than now?

craftsmanship

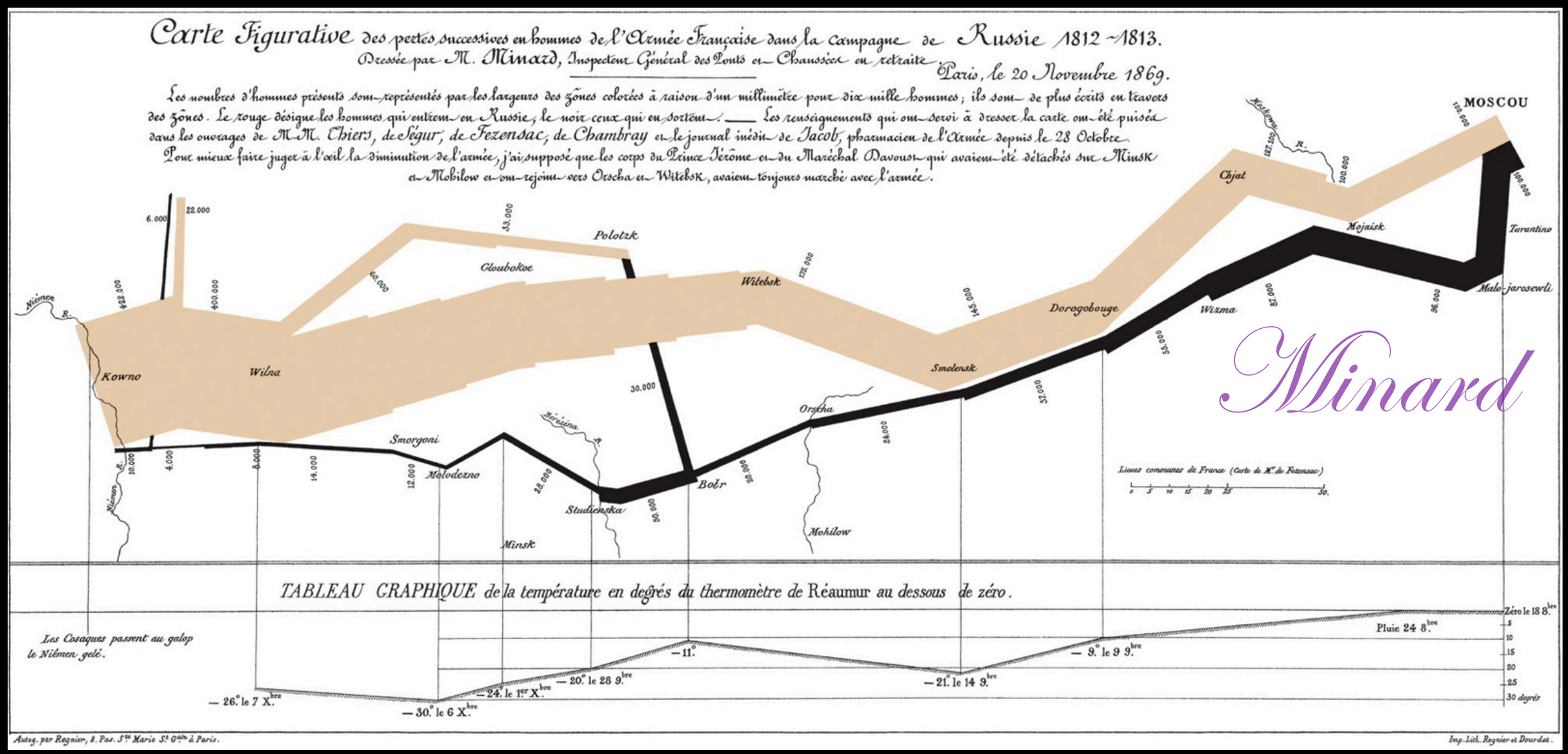
...should be easier in the future?

modular craftsmanship

Are we held back by confining tools?



"A computer"



Lattice – Multivariate Data Visualization with R – Figures and Code

http://lmdvr.r-forge.r-project.org/figures/figures.html

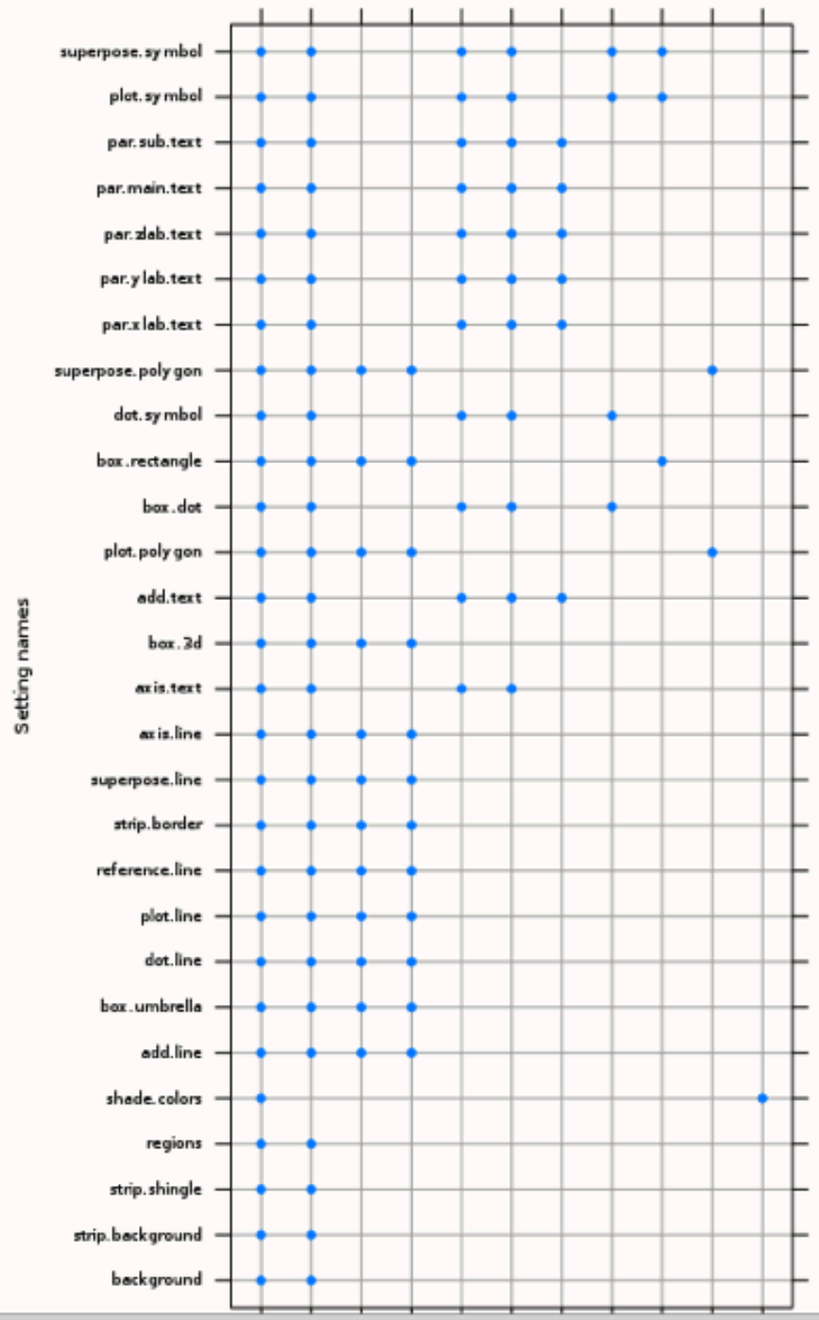
lattice data visualization

Lattice: Multivariate Data Visualization with R - Figures and Code

Black and White Theme | Default Color Theme | Classic Gray Theme

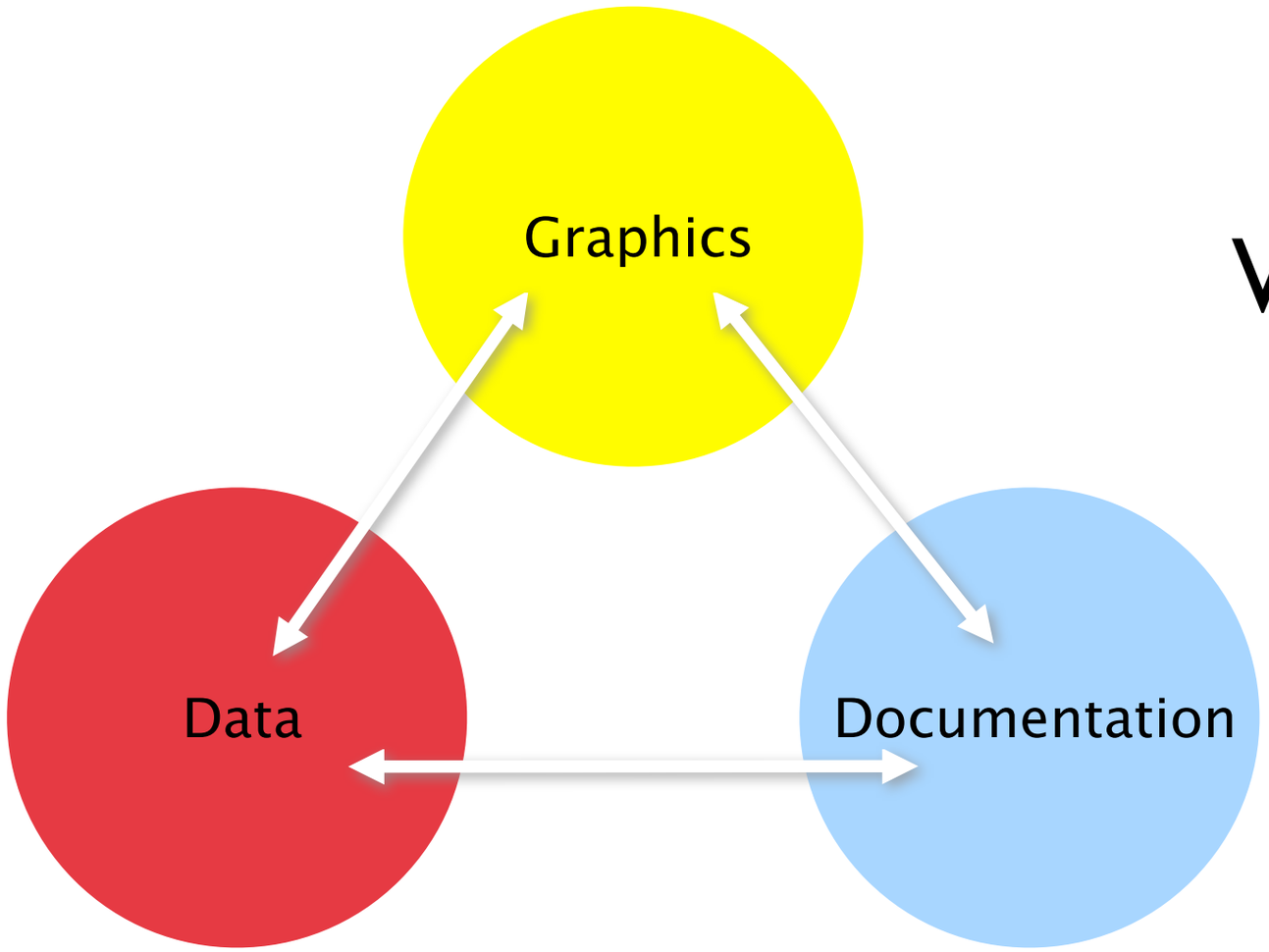
Reset Permalink Contact

Figure 7.3



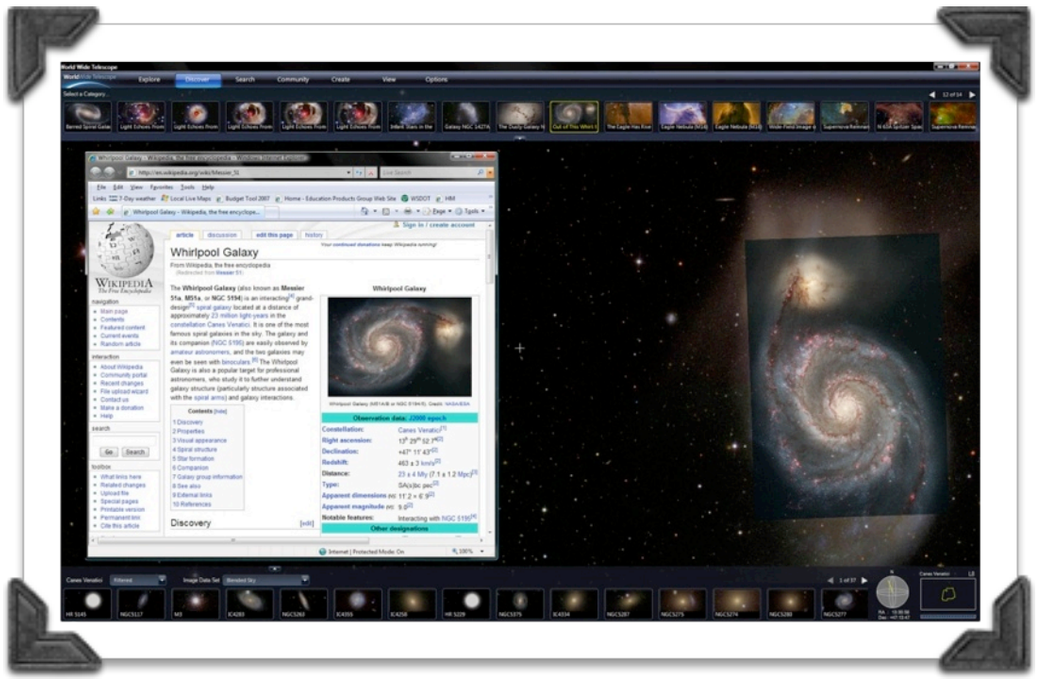

```
## Chapter 7                                                    witch code layout

vad.plot <-
    dotplot(reorder(Var2, Freq) ~ Freq | Var1,
            data = as.data.frame.table(VADeaths),
            origin = 0, type = c("p", "h"),
            main = "Death Rates in Virginia - 1940",
            xlab = "Number of deaths per 100")

## Figure 7.1
vad.plot

dot.line.settings <- trellis.par.get("dot.line")
str(dot.line.settings)
dot.line.settings$col <- "transparent"
trellis.par.set("dot.line", dot.line.settings)
plot.line.settings <- trellis.par.get("plot.line")
str(plot.line.settings)
plot.line.settings$lwd <- 2
trellis.par.set("plot.line", plot.line.settings)

## Figure 7.2
vad.plot

panel.dotline <-
    function(x, y,
             col = plot.symbol$col, pch = dot.symbol$pch,
             cex = dot.symbol$cex, alpha = dot.symbol$alpha,
             col.line = plot.line$col, lty = plot.line$lty,
             lwd = plot.line$lwd, alpha.line = plot.line$alpha,
             ...)
{
    dot.symbol <- trellis.par.get("dot.symbol")
    plot.line <- trellis.par.get("plot.line")
    panel.segments(0, y, x, y, col = col.line, lty = lty,
                   lwd = lwd, alpha = alpha.line)
    panel.points(x, y, col = col, pch = pch, cex = cex, alpha = alpha)
}
trellis.par.set(dot.line = dot.line.settings,
                plot.line = plot.line.settings)
trellis.par.set(dot.line = list(col = "transparent"),
                plot.line = list(lwd = 2))
trellis.par.set(list(dot.line = list(col = "transparent"),
                     plot.line = list(lwd = 2)))

## Figure 7.2 (alternative)
update(vad.plot,
       par.settings = list(dot.line = list(col = "transparent"),
                           plot.line = list(lwd = 2)))

tp <- trellis.par.get()
unusual <-
    c("grid.pars", "fontsize", "clip",
      "axis.components",
      "layout.heights", "layout.widths")
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...)



WorldWide Telescope

Microsoft Research