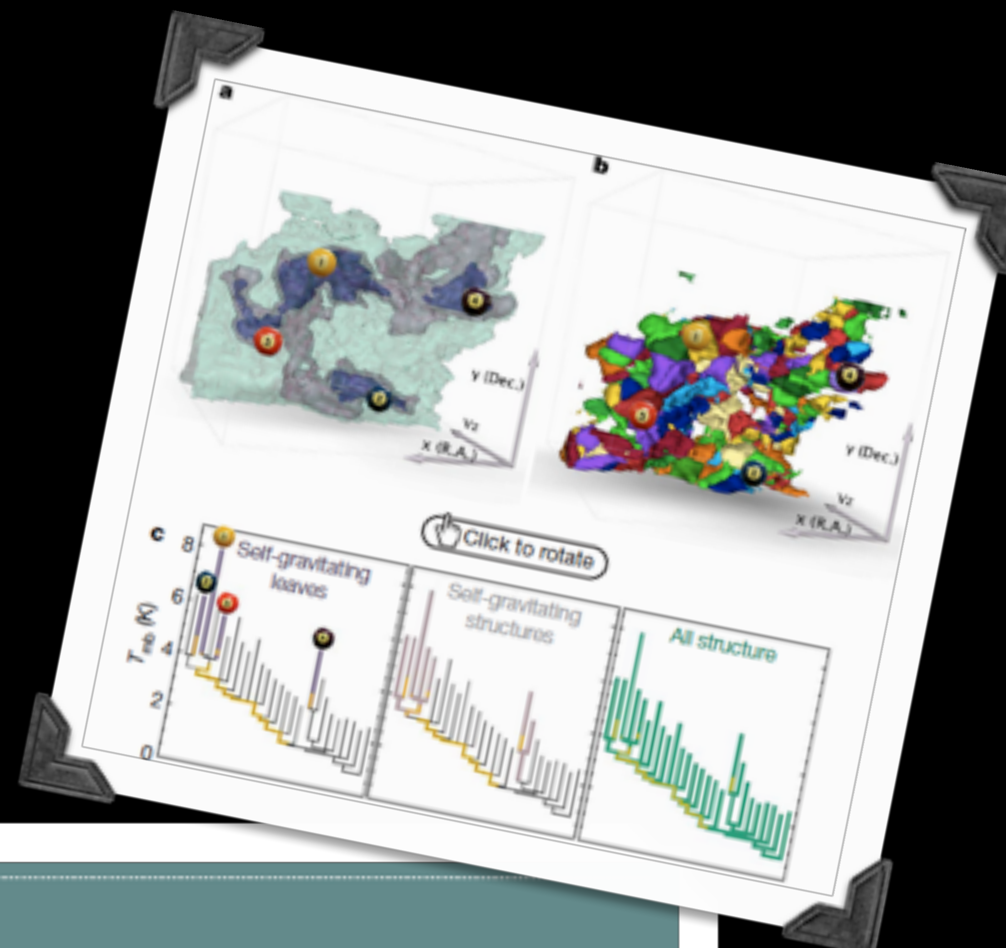
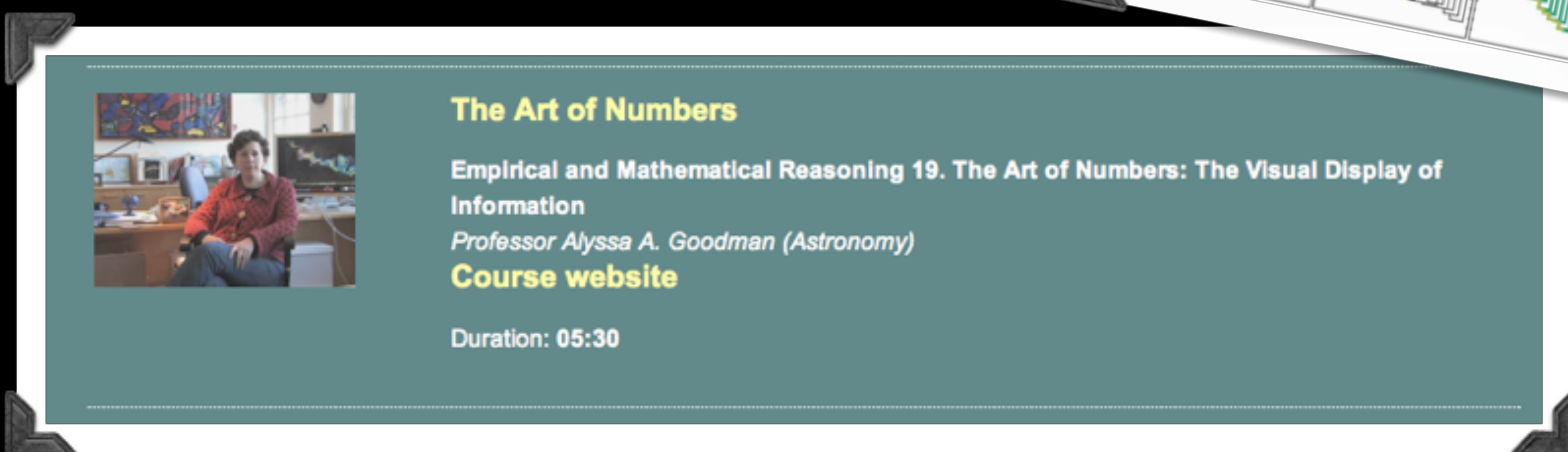
WorldWide Telescope

Microsoft Research

Seeing Science

Data Visualization in Modern Research (and teaching!)




The Art of Numbers

Empirical and Mathematical Reasoning 19. The Art of Numbers: The Visual Display of Information

Professor Alyssa A. Goodman (Astronomy)

Course website

Duration: 05:30

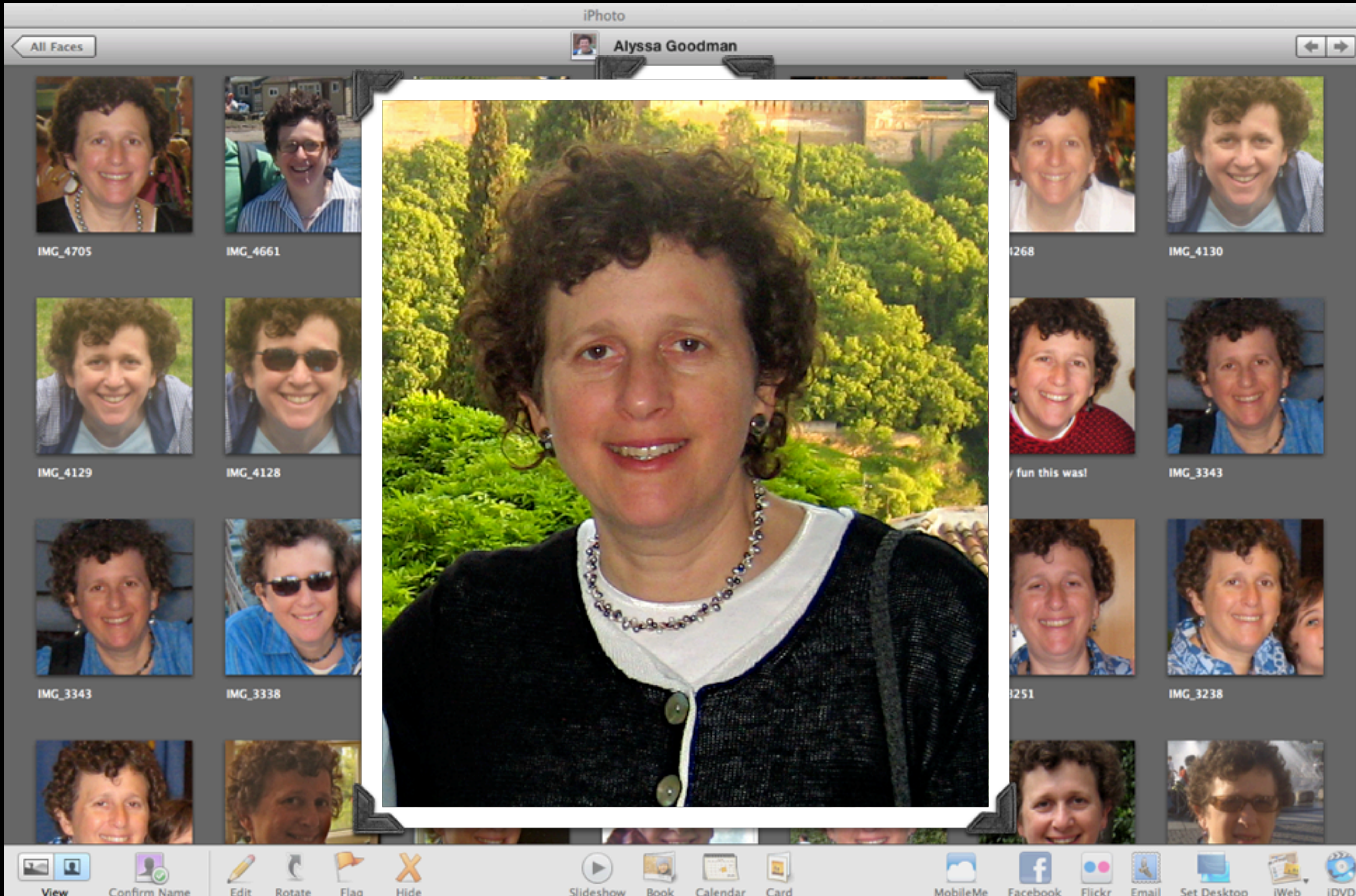
What kind of credentials are those??

Alyssa A. Goodman
Harvard University (HCO+IIC)
~~Smithsonian Astrophysical Observatory~~
~~Scholar-in-Residence, WGBH~~

Viz-e-Lab







iPhoto
All Faces
Alyssa Goodman
IMG_4705
IMG_4661
IMG_4130
IMG_4129
IMG_4128
fun this was!
IMG_3343
IMG_3343
IMG_3338
IMG_3238
View
Confirm Name
Edit
Rotate
Flag
Hide
Slideshow
Book
Calendar
Card
MobileMe
Facebook
Flickr
Email
Set Desktop
iWeb
iDVD

19 out of 22?

Relative Strengths

Pattern Recognition
Creativity

Calculations

“Interocularity”

(see work of John Tukey)

“Image and Meaning”

(see work of Felice Frankel,
and imageandmeaning.org)

“Image and Meaning”

What...

...is easier now than before?

fast computation, animation, 3D

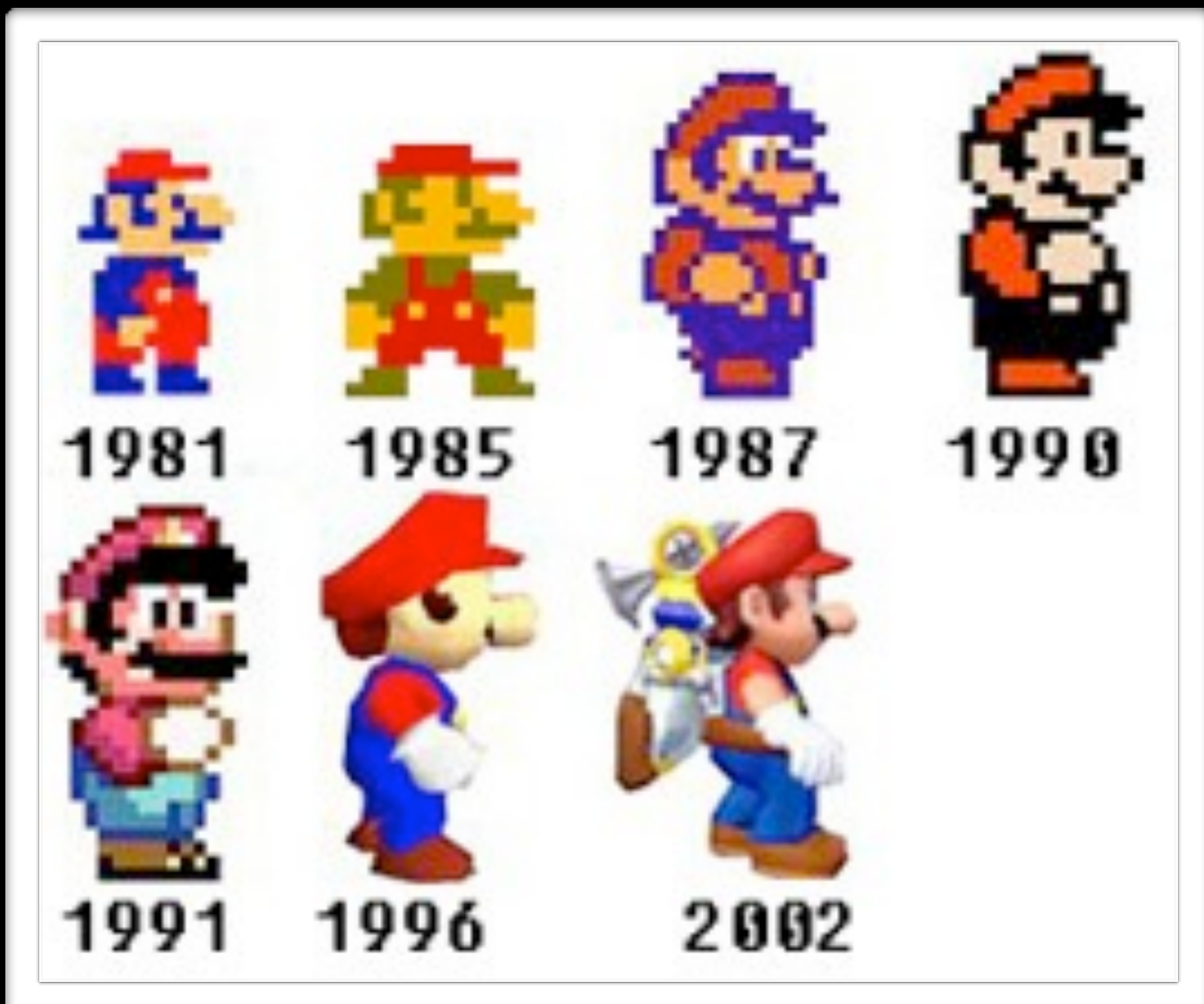
...was easier before than now?

craftsmanship

...should be easier in the future?

modular craftsmanship, linked views

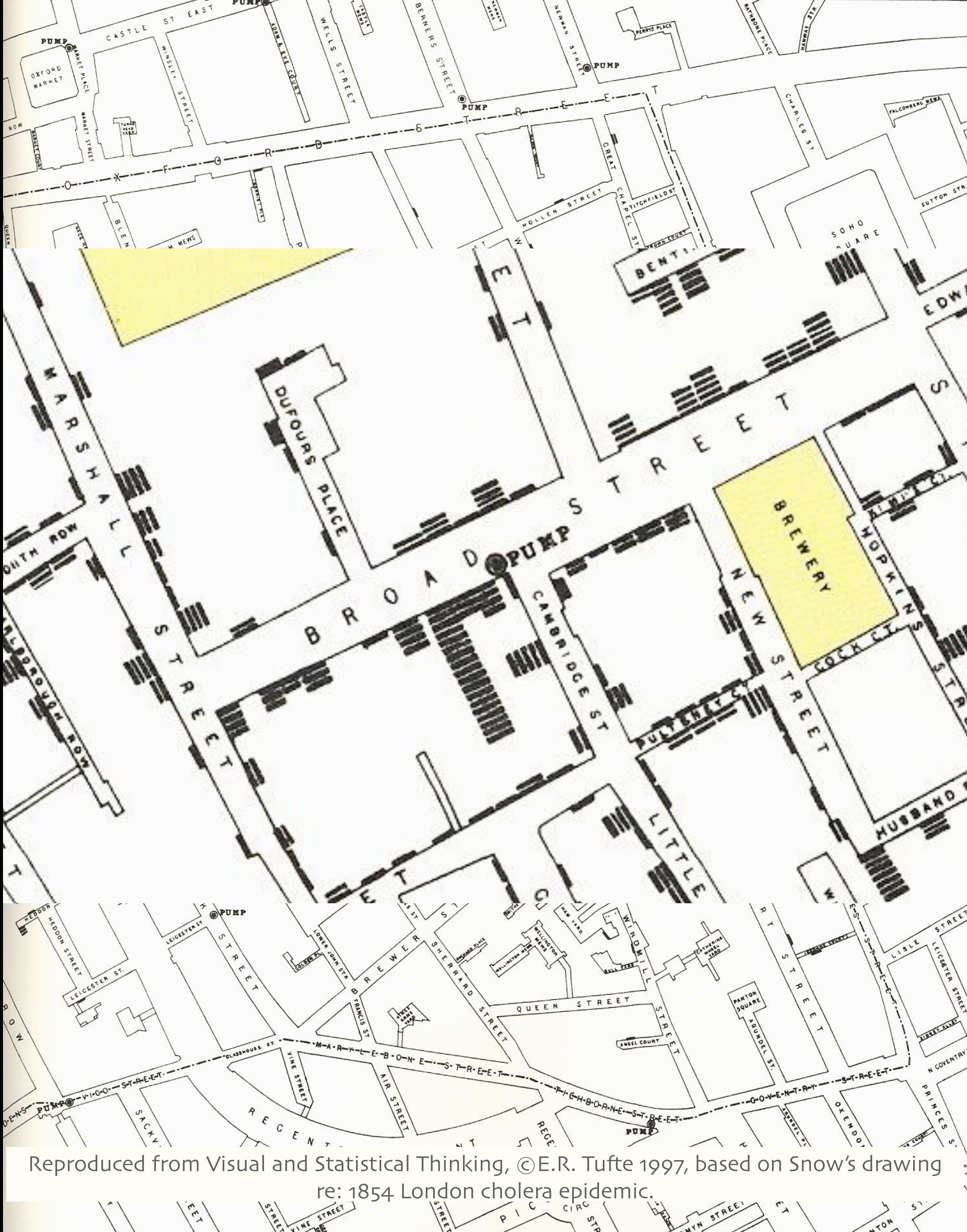
“Easier”



2011

Craftsmanship

(in 1854)



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

Data • Dimensions • Display

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

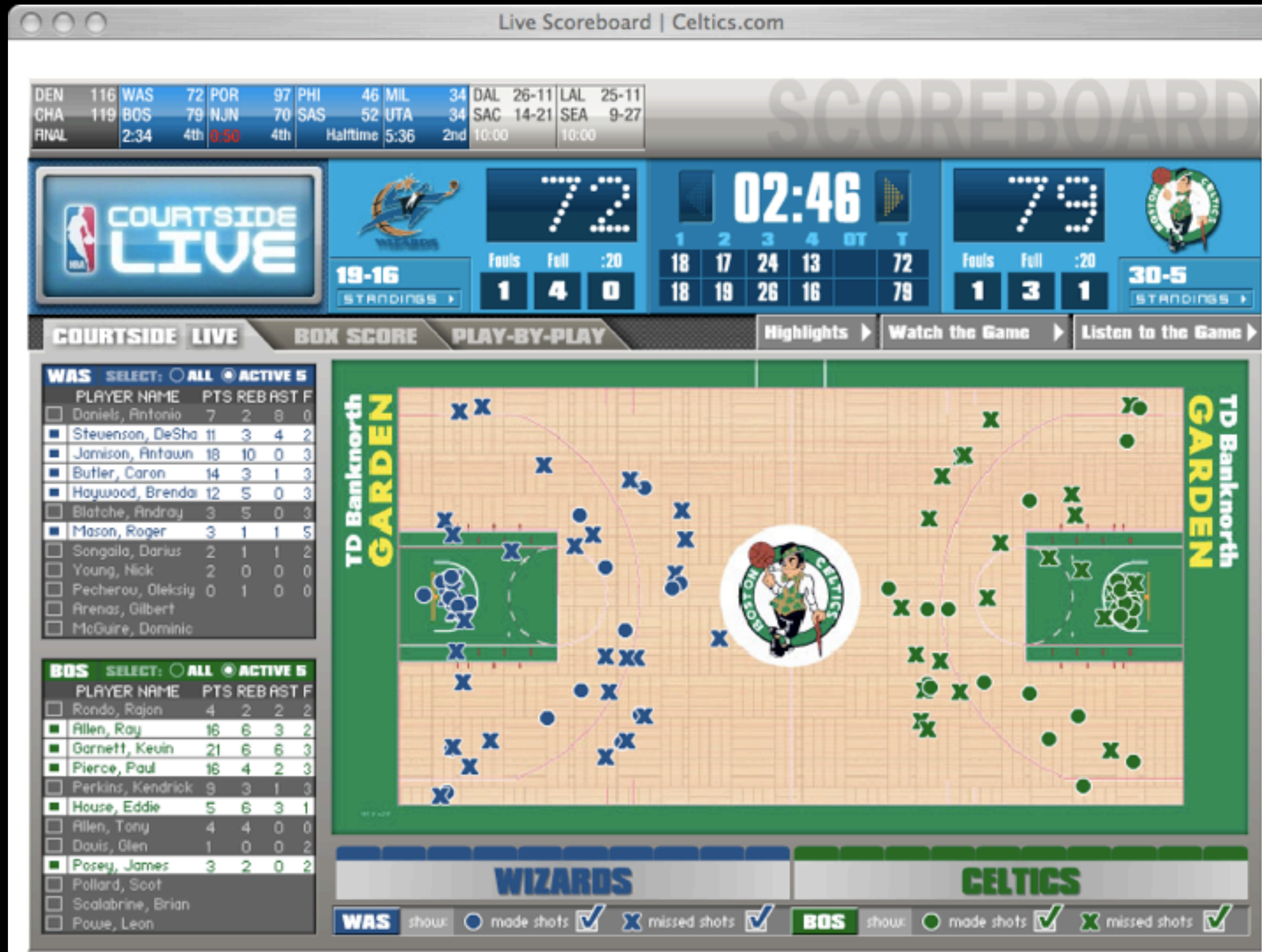
✓Tables

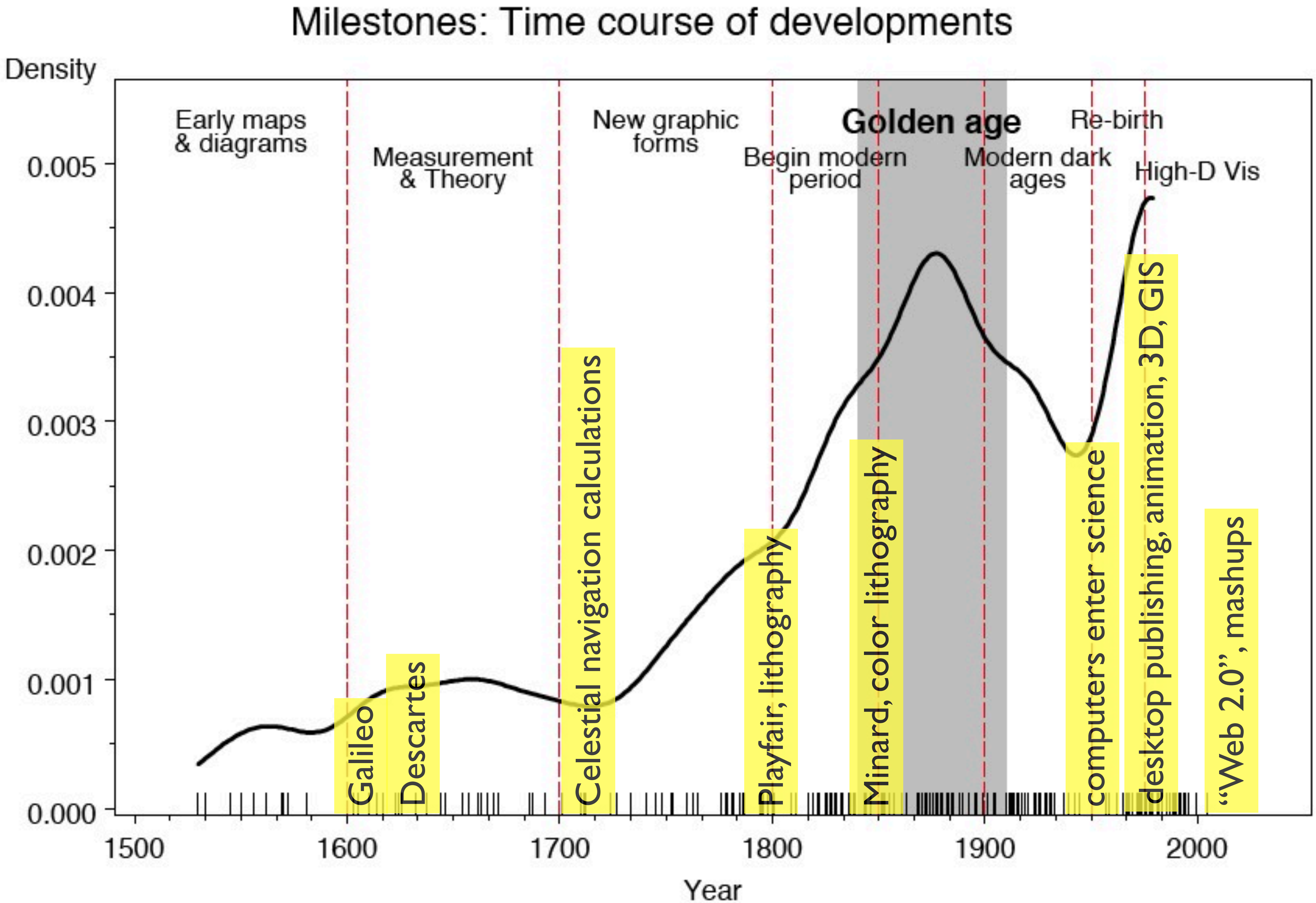
✖Graphs

✓Charts

✓Illustrations

✓Combinations

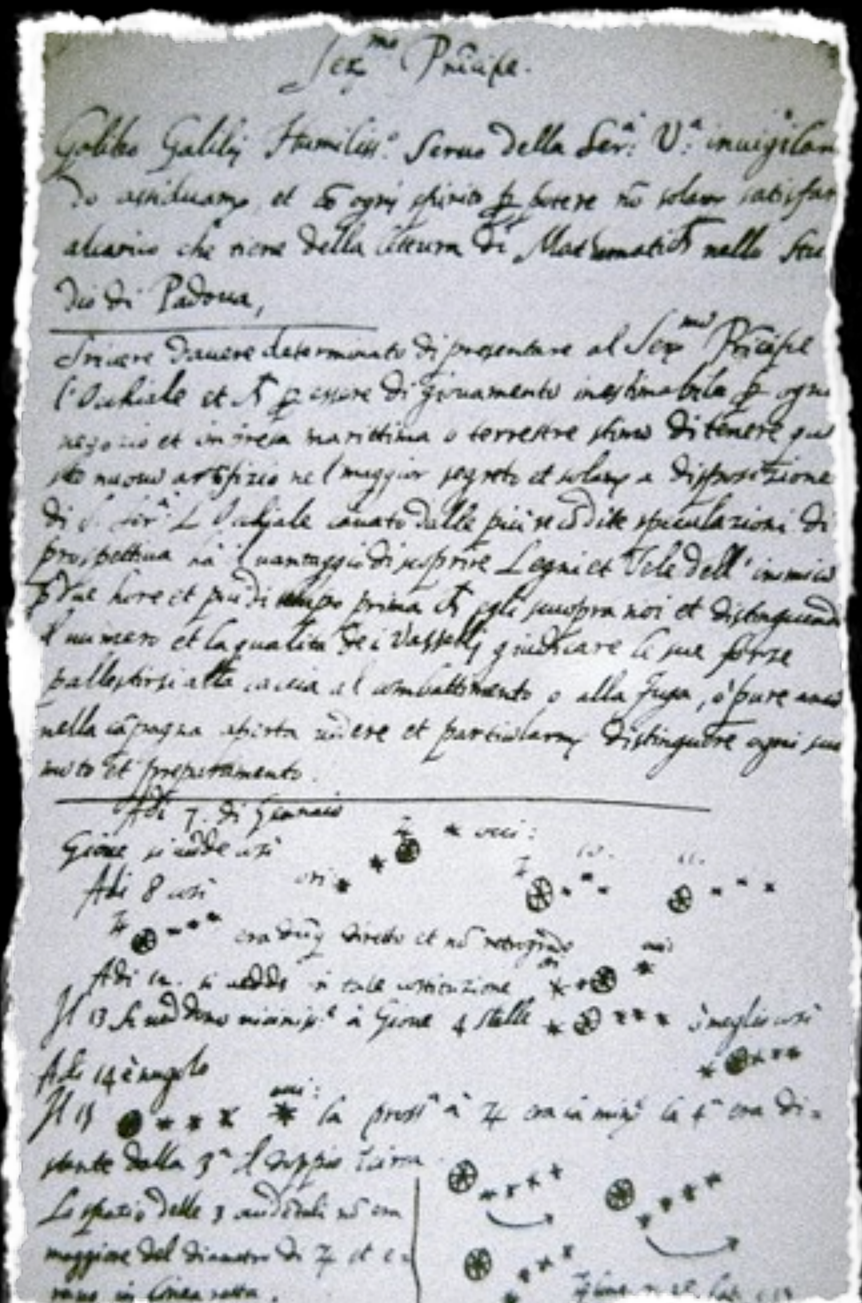


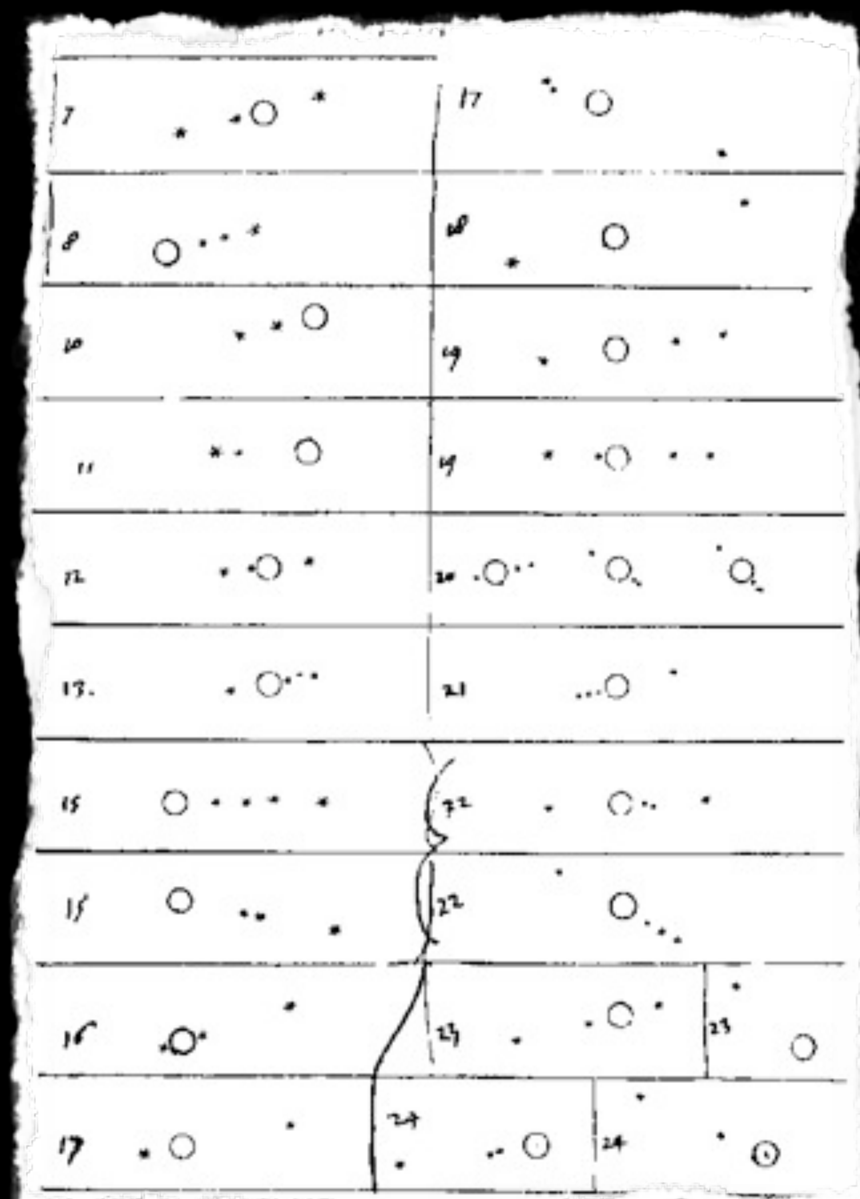


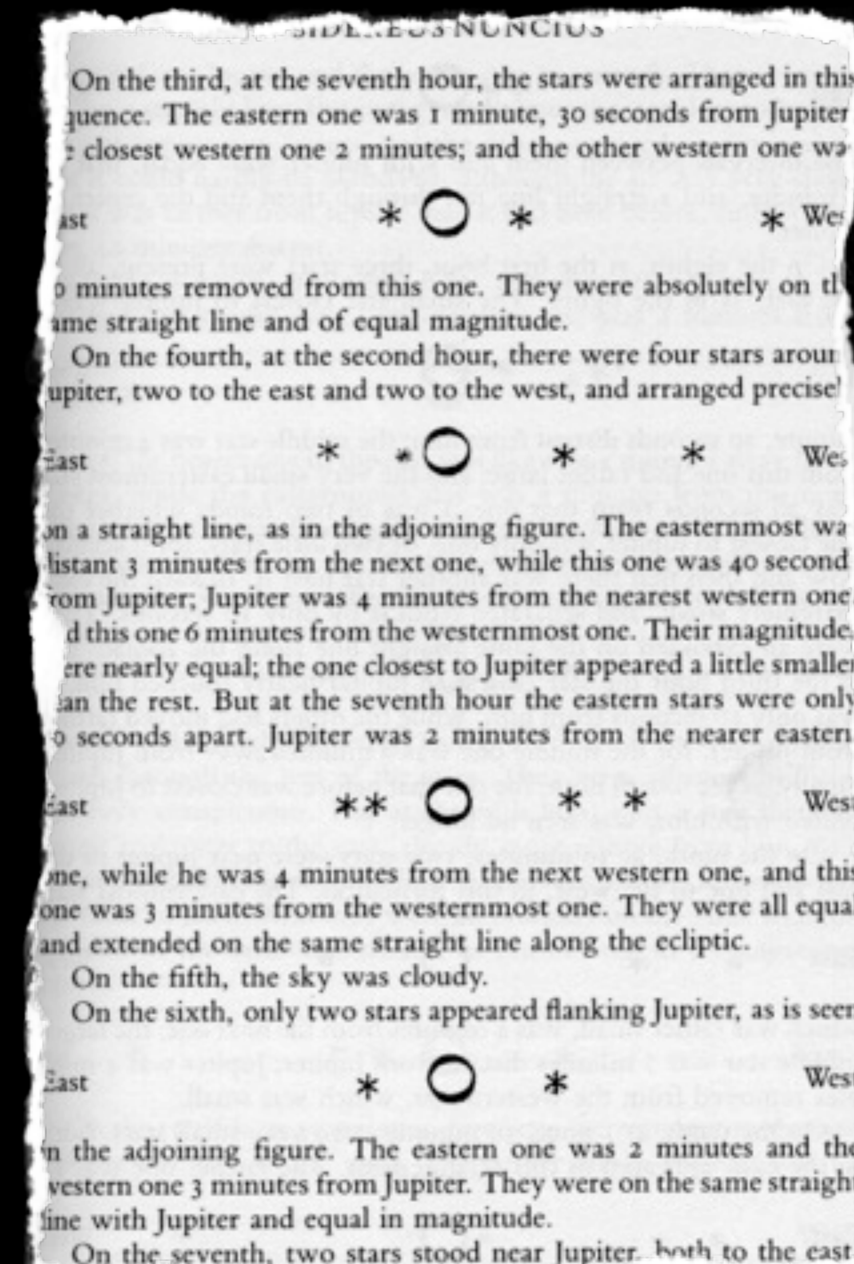
adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, 2009

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

ast * O * * Wes

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

East * *O * * Wes

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

East ** O * * West

ne, 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

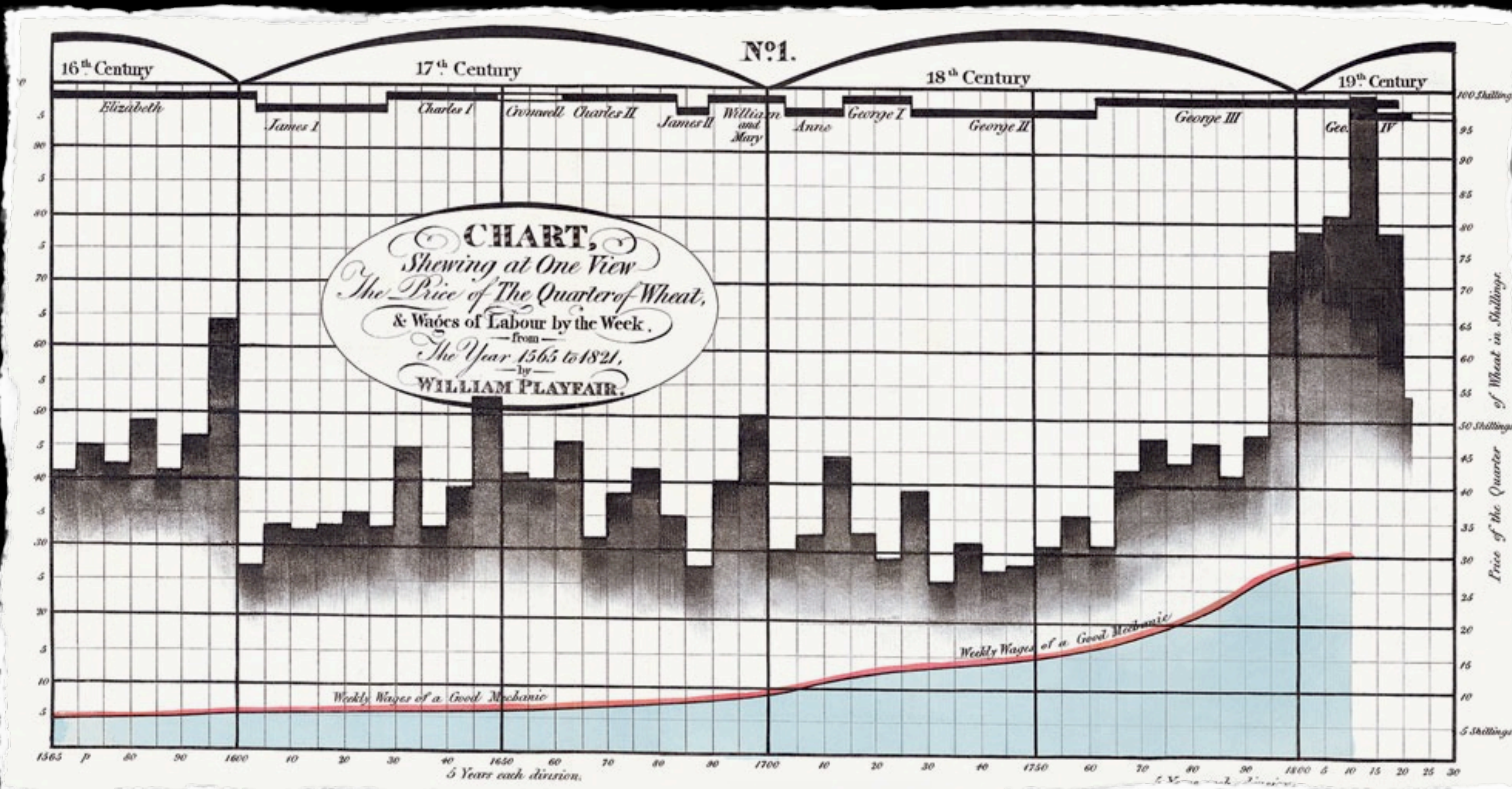
East * O * West

n the adjoining figure. The eastern one was 2 minutes and the
vestern 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

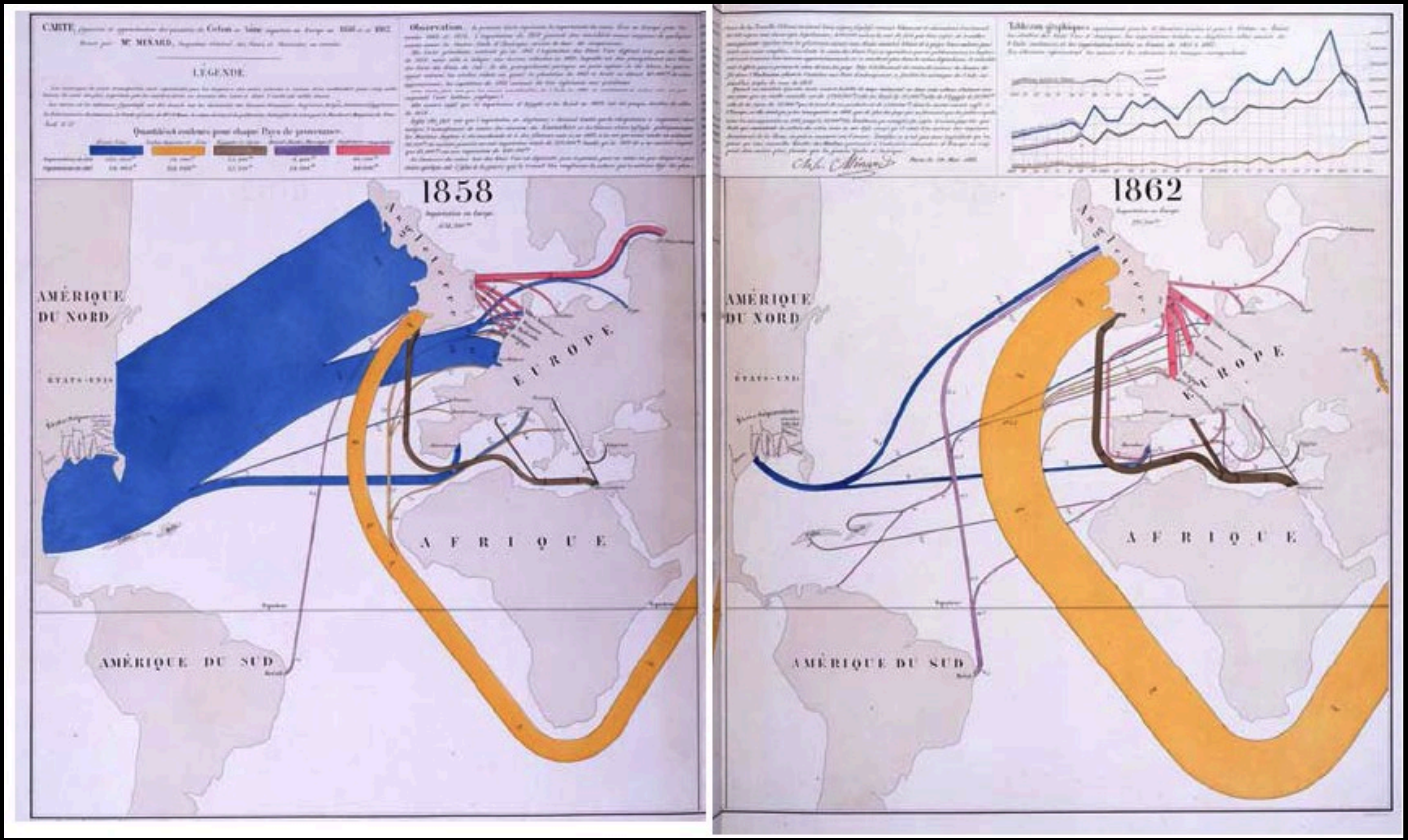
Playfair, lithography

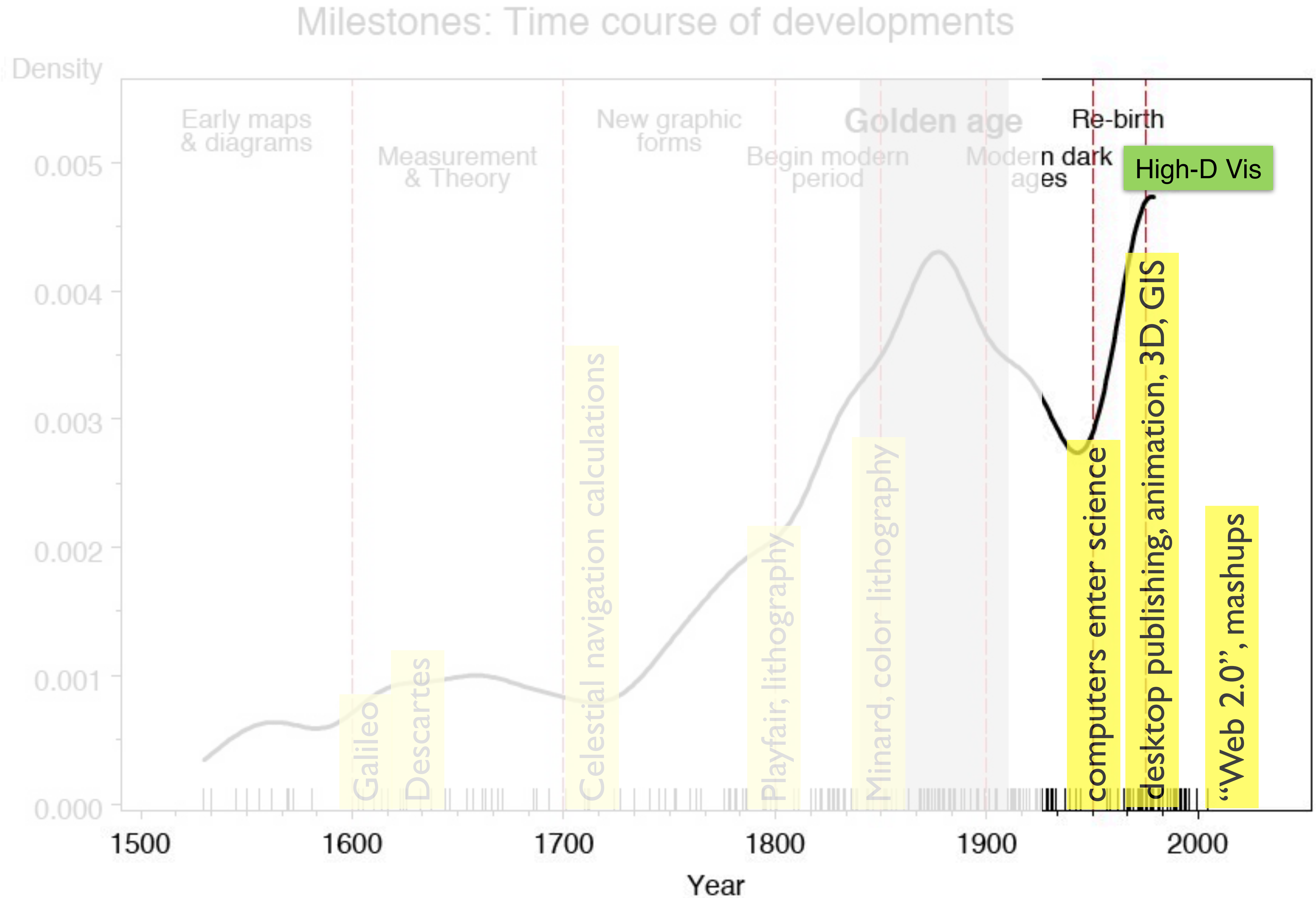
William Playfair
(1759-1823)



Minard, color lithography

Charles Joseph Minard, in color
(1781-1870)



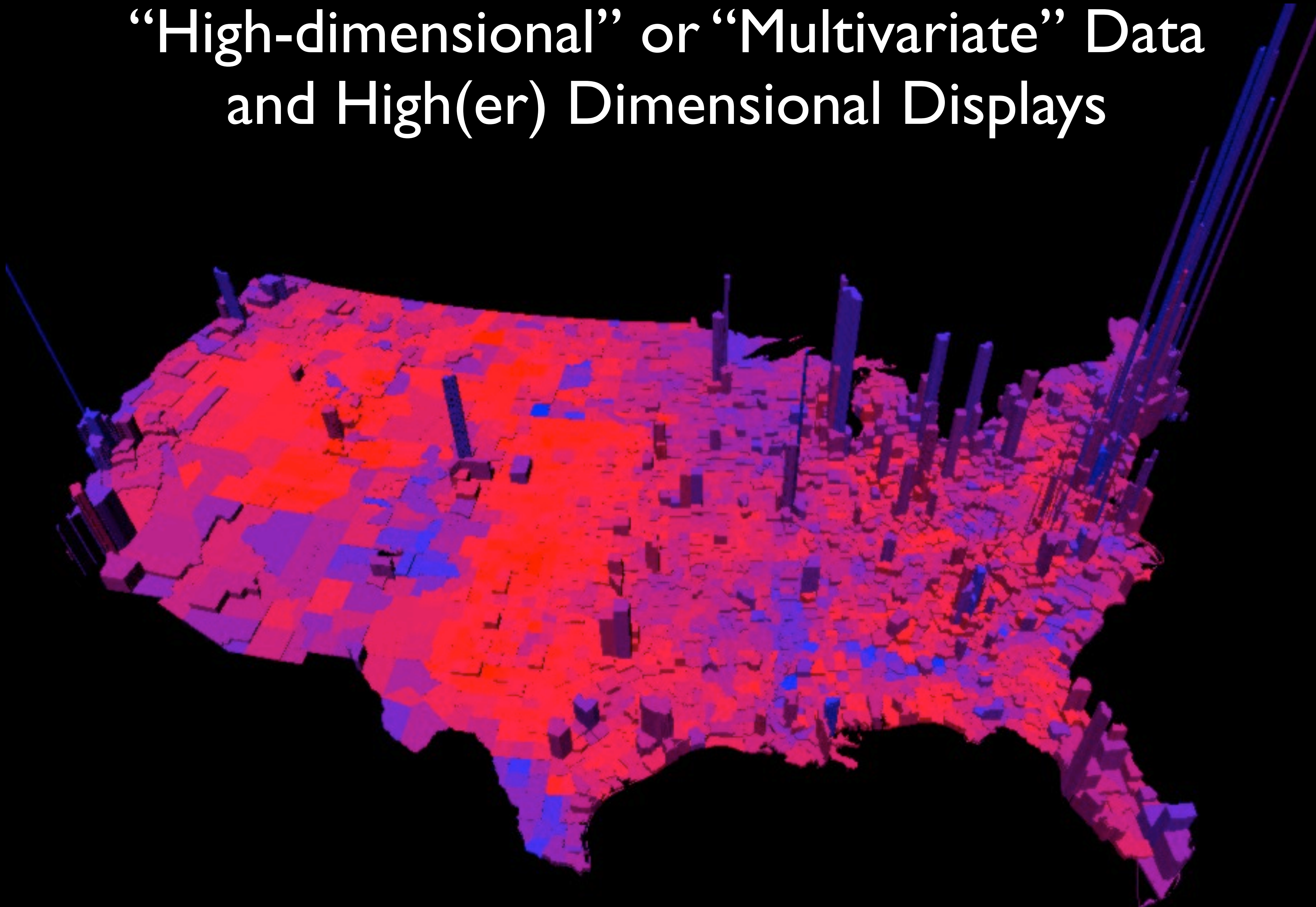


adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

High-D Vis

Data • Dimensions • Display

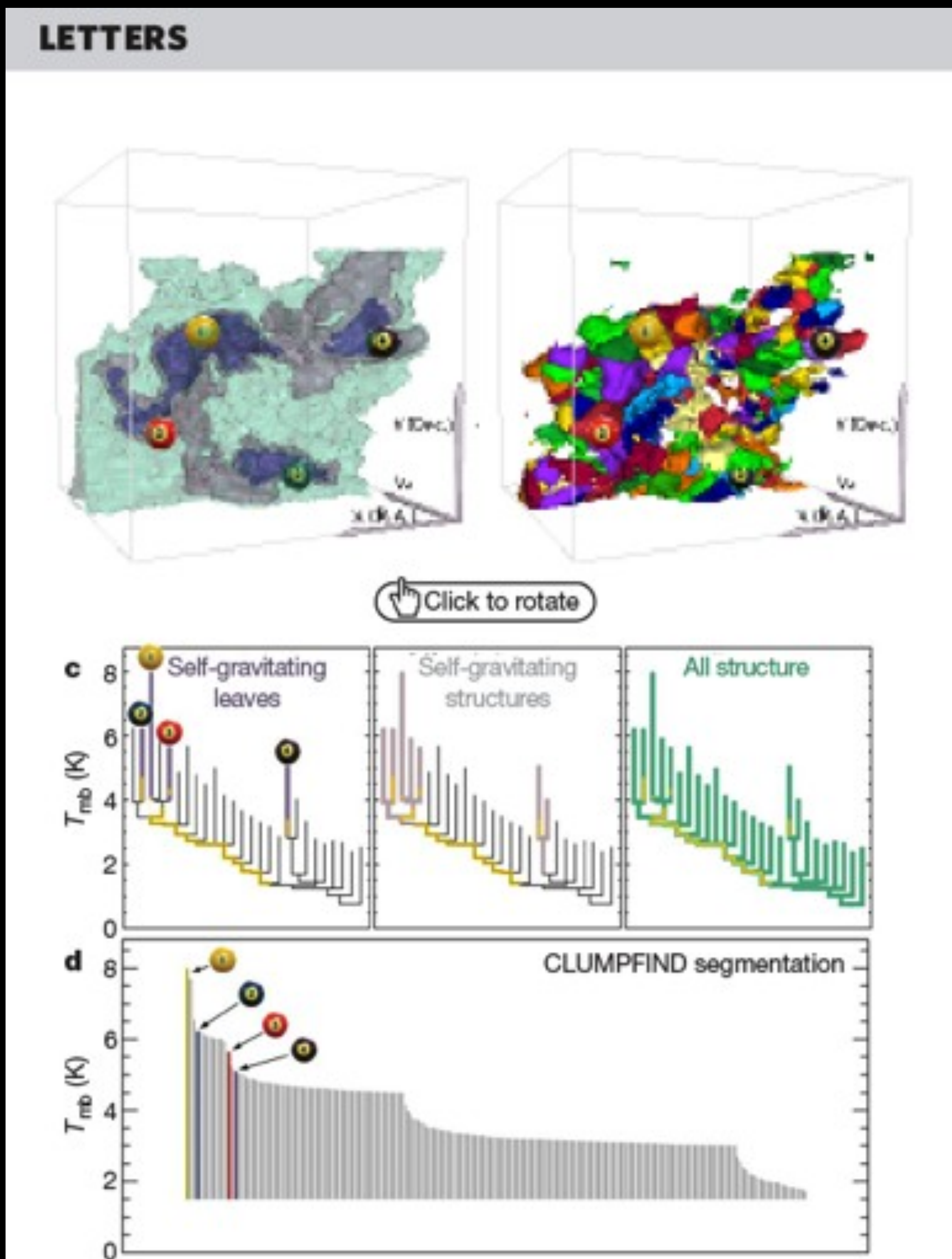
“High-dimensional” or “Multivariate” Data and High(er) Dimensional Displays



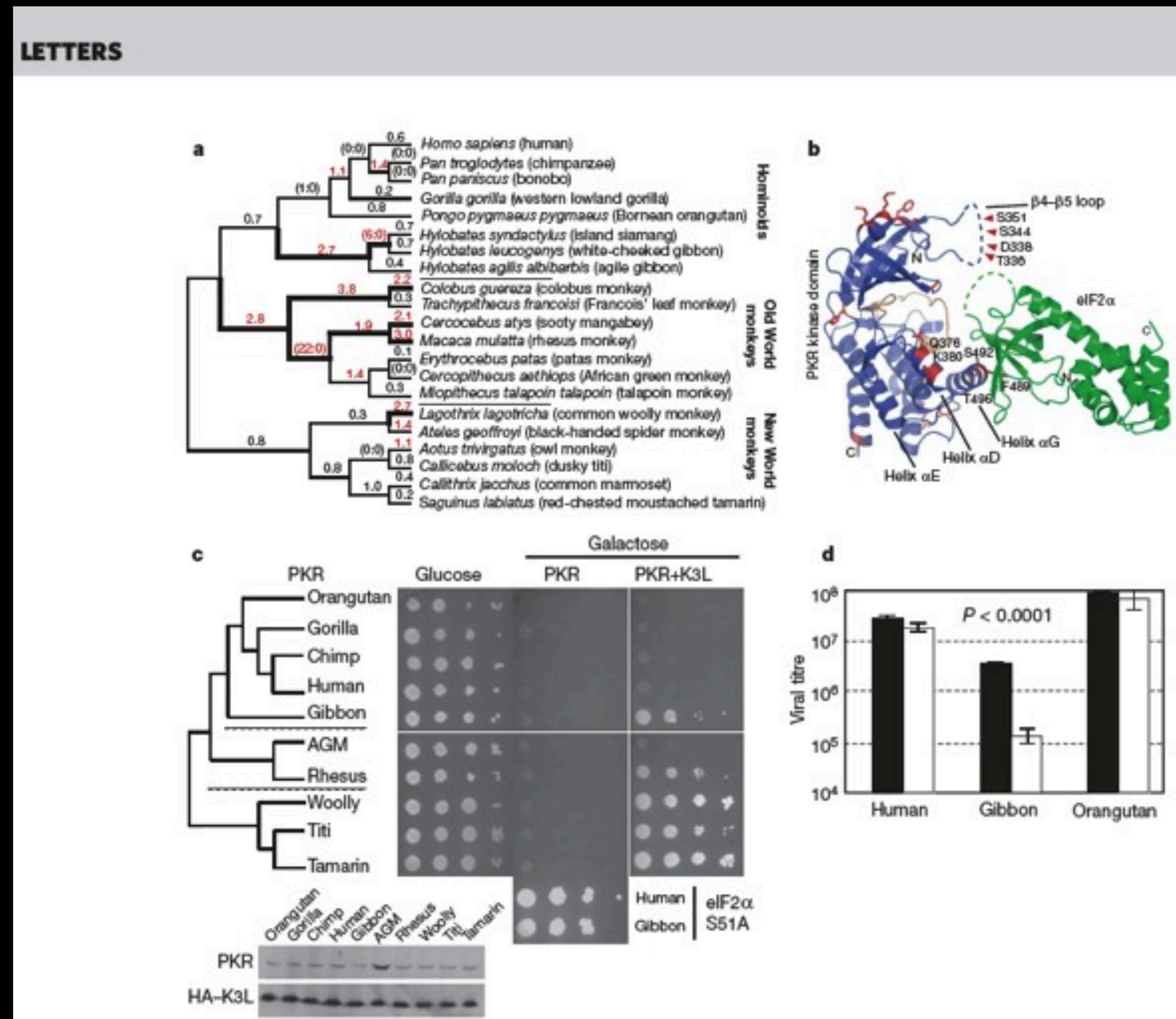
*This map **displays** 2 quantities as a function of 2 spatial dimensions.*
...Is that 4 dimensions?

“High-dimensional” or “Multivariate” Data

(Astronomy=Biology)

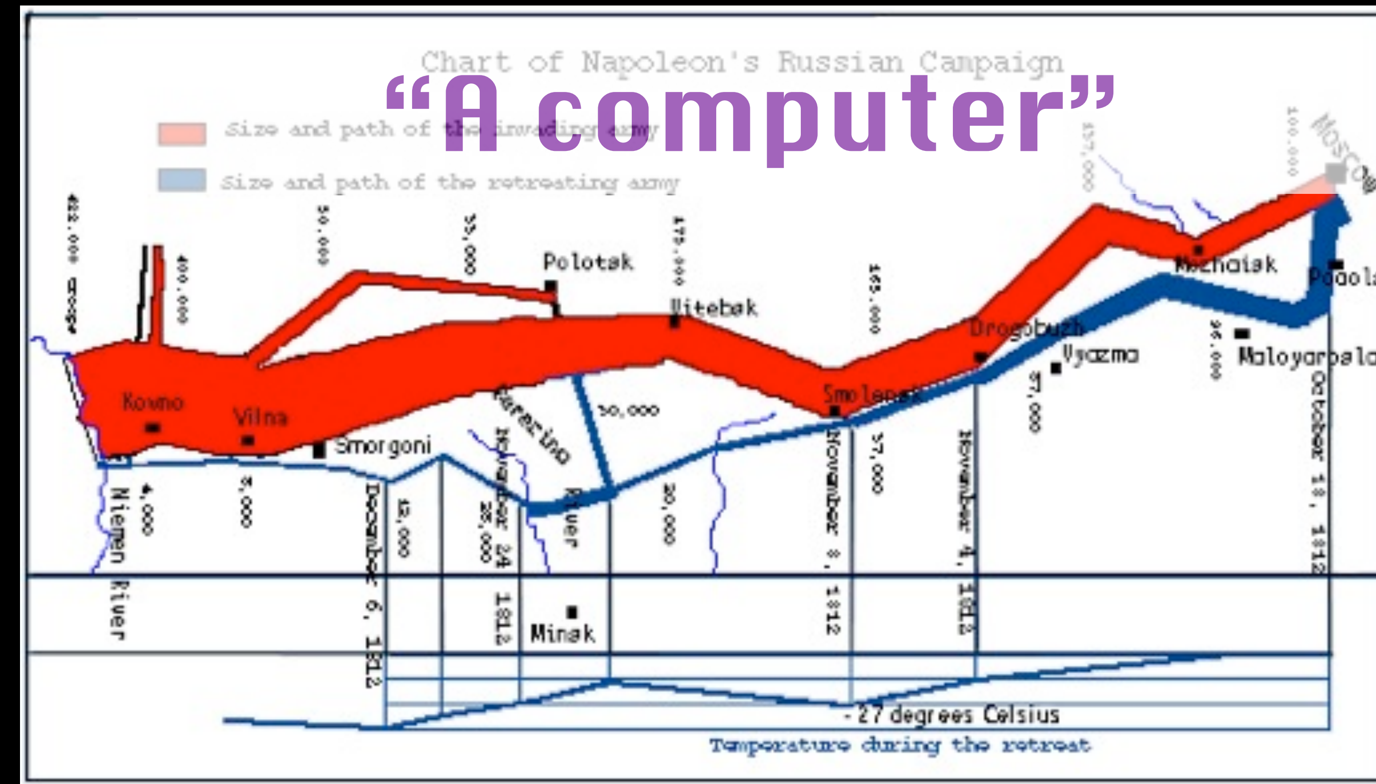


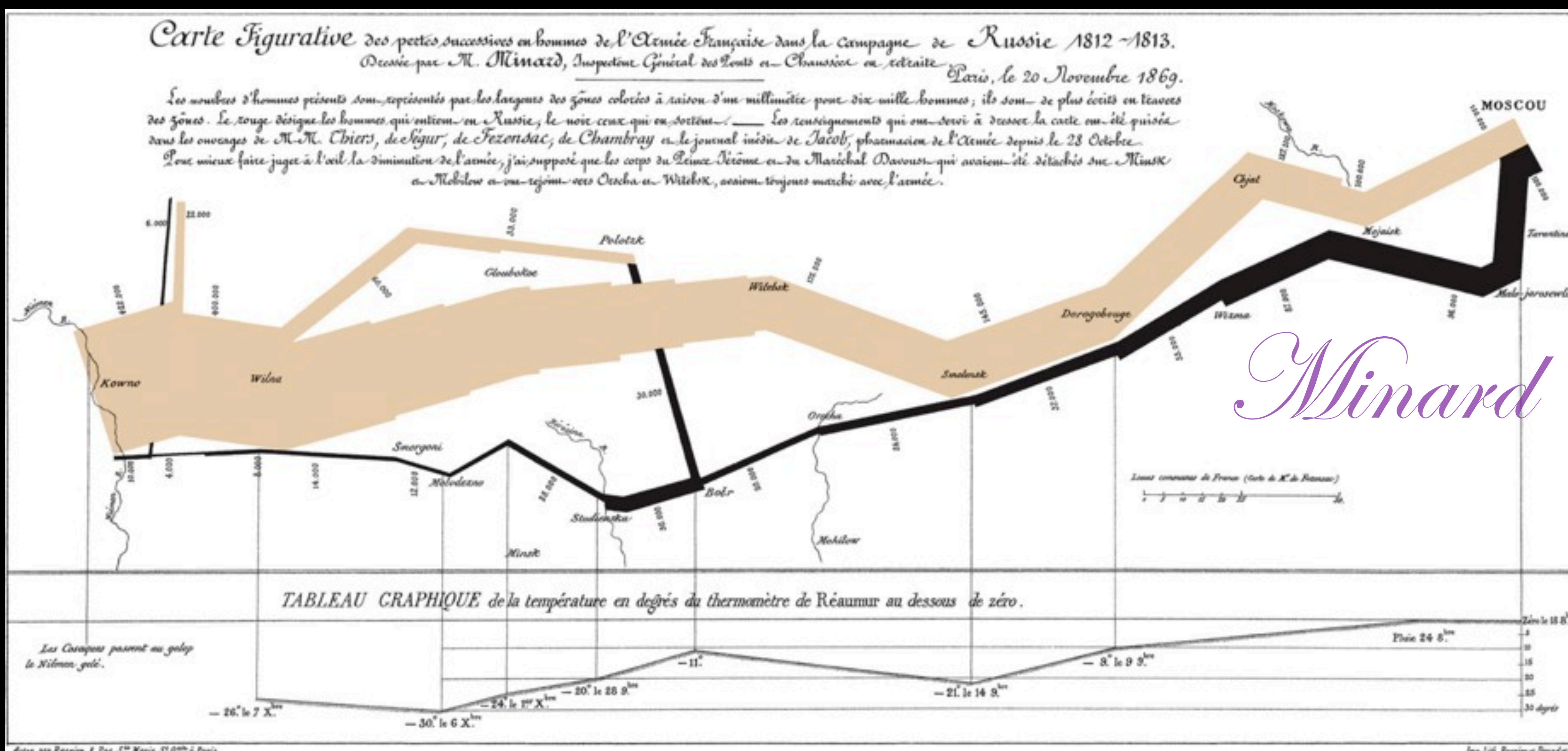
Goodman et al. *Nature*, 2009



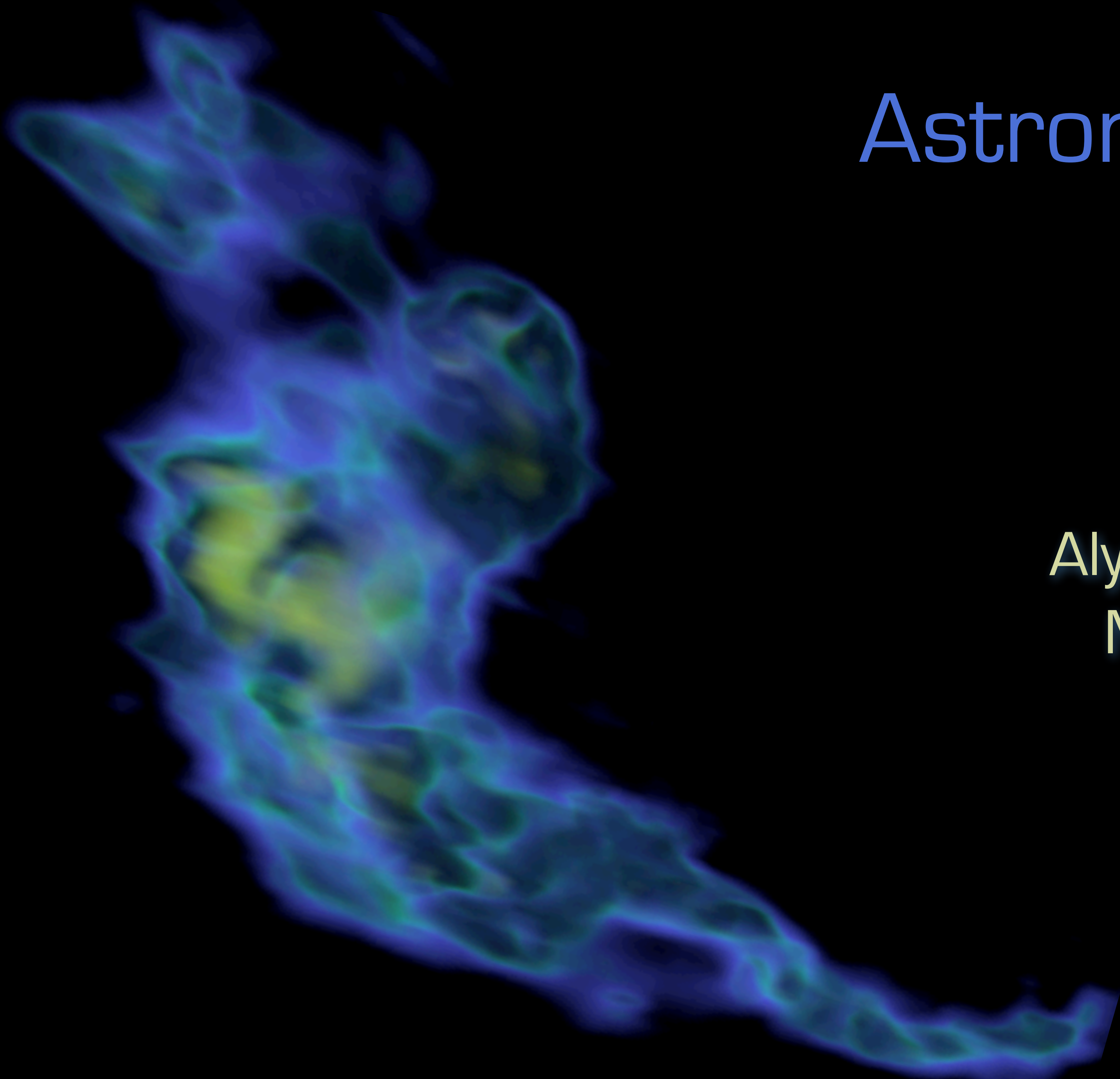
Elde et al. *Nature*, 2008

How much are we held back today by ***digital*** *tools?*





*How can we advance the **digital** tools for scientists?*



Astronomical Medicine

am.iic.harvard.edu

Alyssa Goodman (IIC/CfA/FAS)

Michael Halle (IIC/SPL/HMS)

Ron Kikinis (SPL/HMS)

Douglas Alan (IIC)

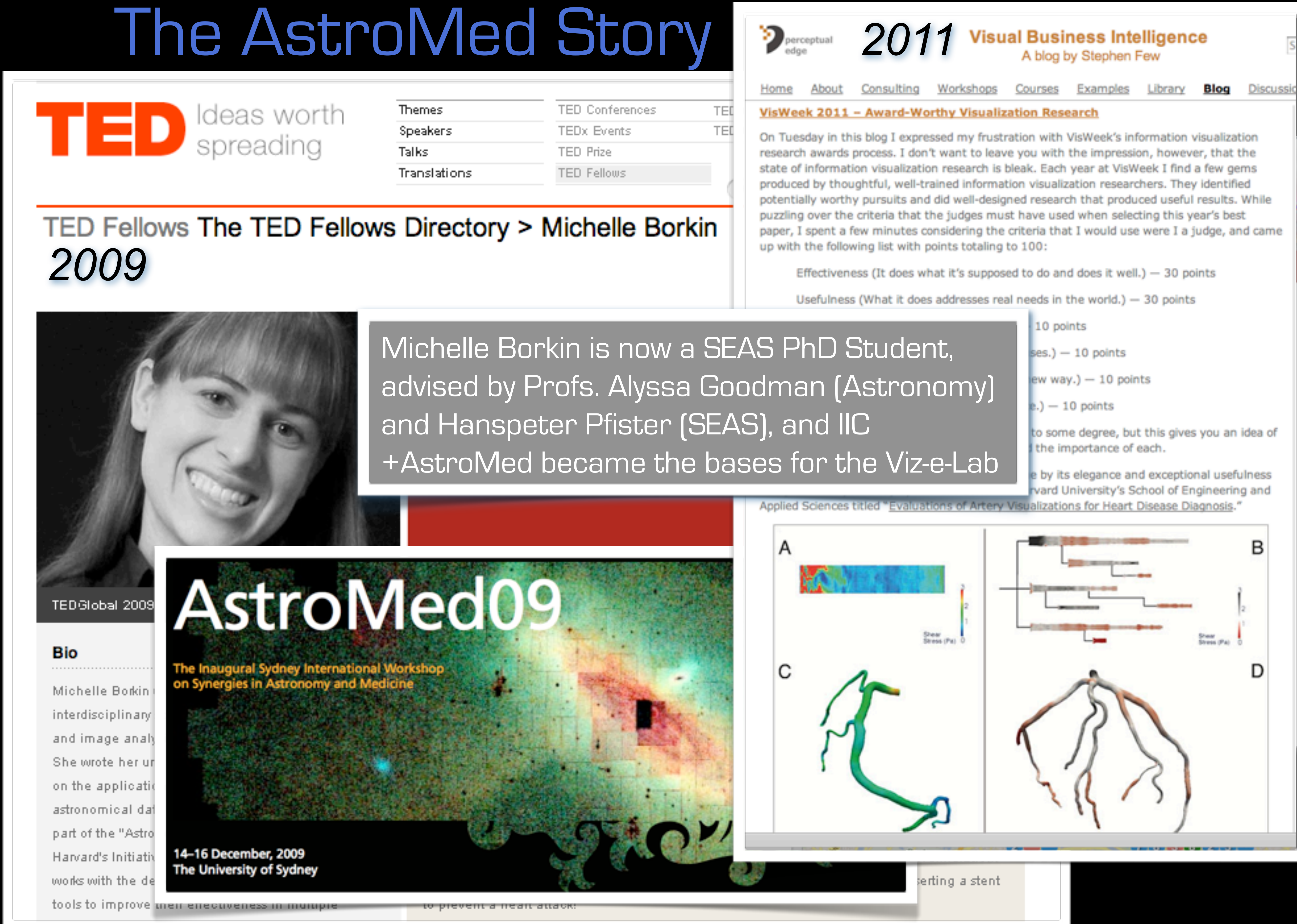
Michelle Borkin (IIC)

Jens Kauffmann (CfA/IIC)

Erik Rosolowsky (CfA)

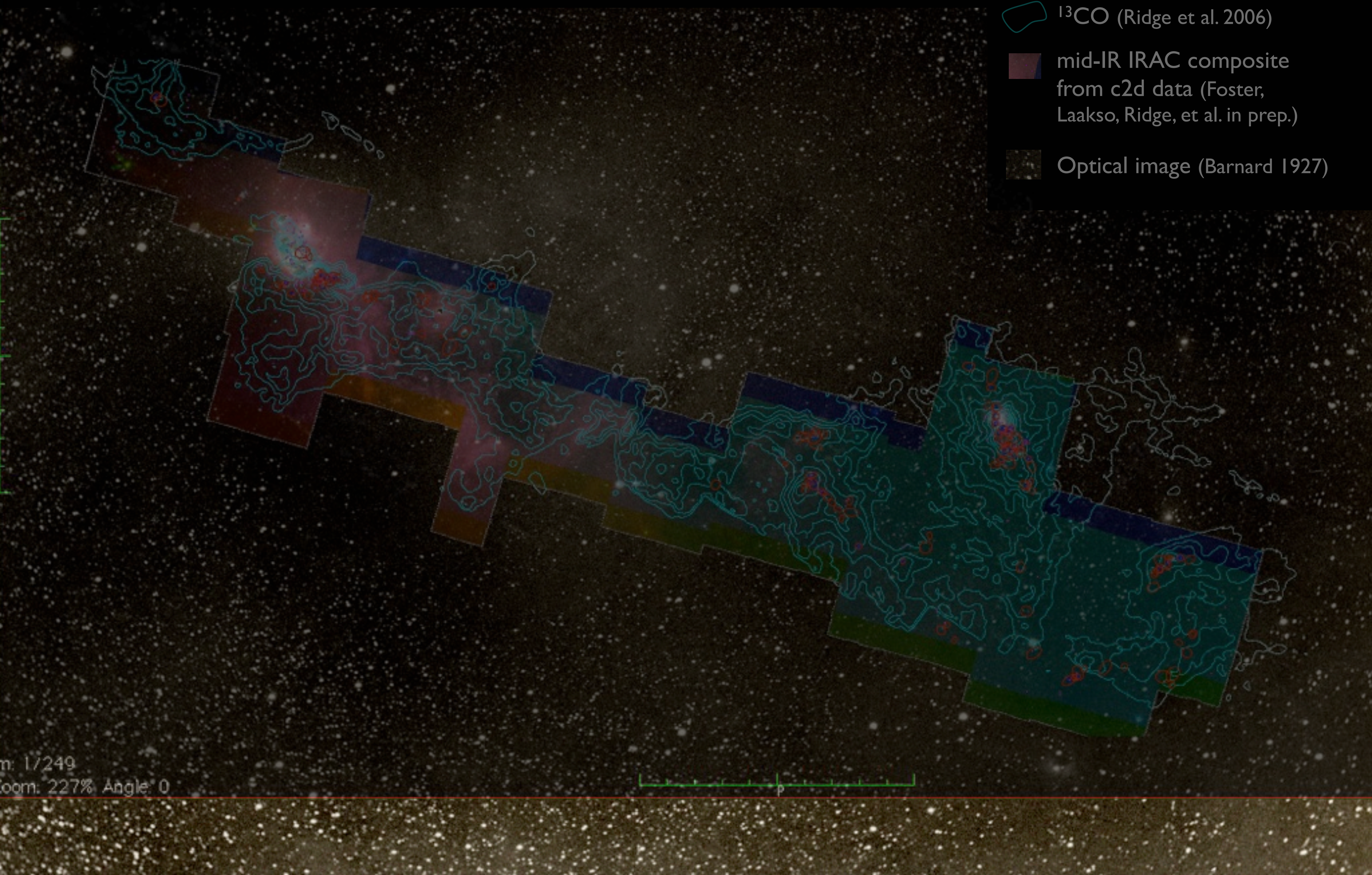
Nick Holliman (U. Durham)

The AstroMed Story



Michelle Borkin is now a SEAS PhD Student, advised by Profs. Alyssa Goodman (Astronomy) and Hanspeter Pfister (SEAS), and IIC +AstroMed became the bases for the Viz-e-Lab

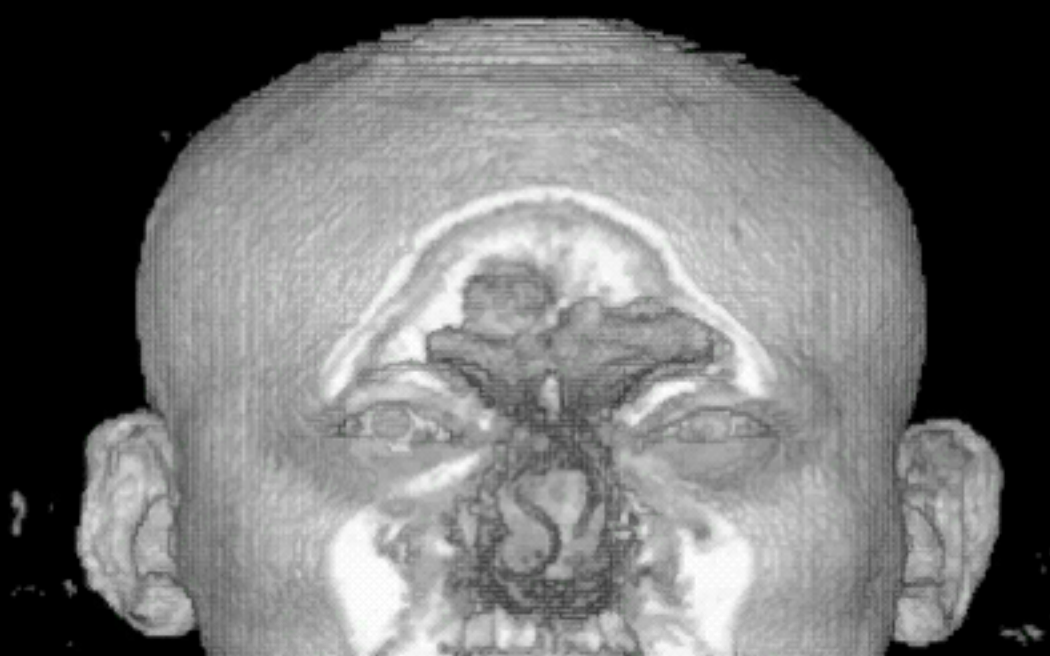
COMPLETE Perseus



- mm peak (Enoch et al. 2006)
- sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
- ^{13}CO (Ridge et al. 2006)
- mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al. in prep.)
- Optical image (Barnard 1927)

“Astronomical Medicine”

“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”)

COMPLETE Perseus

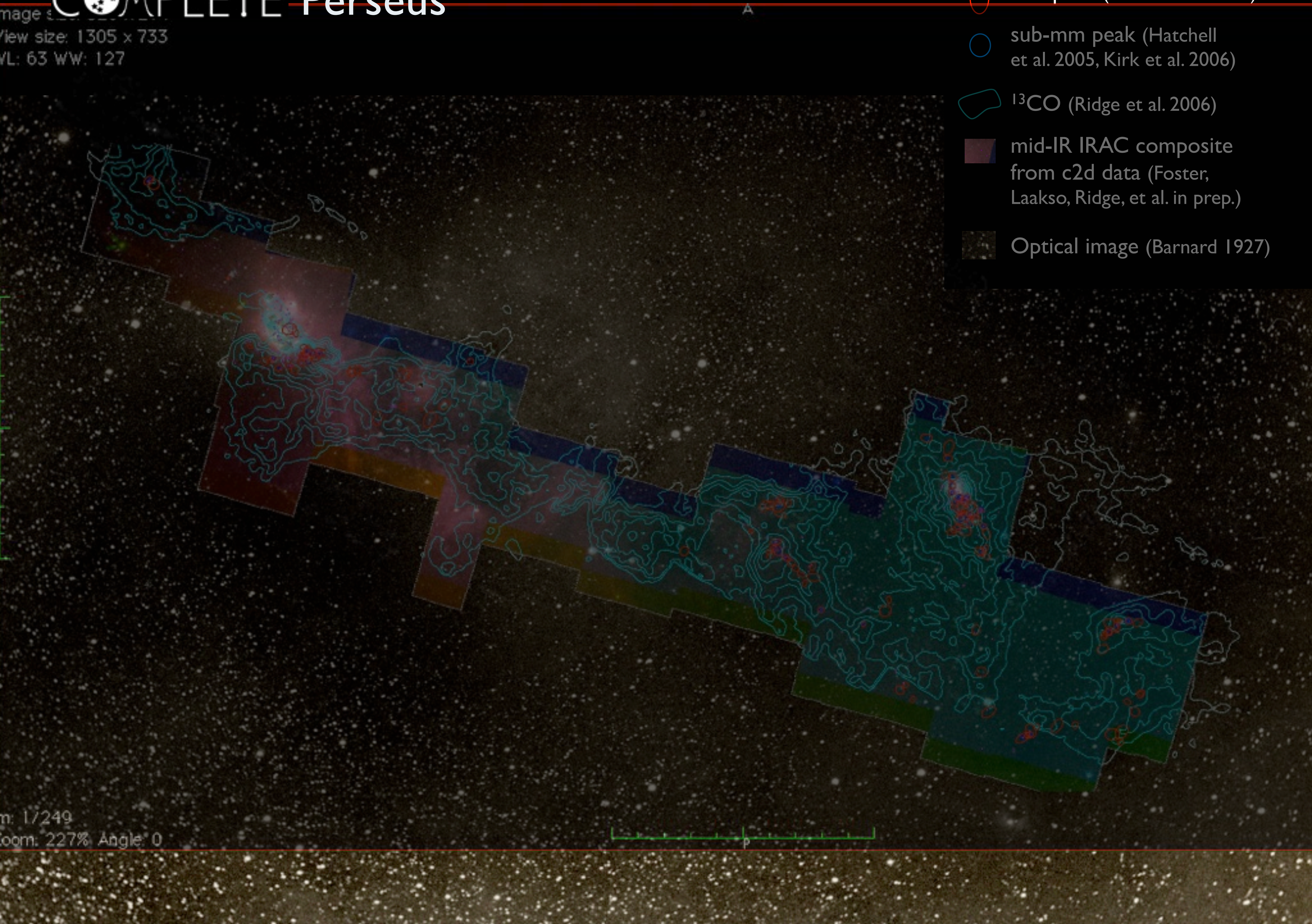
mm peak (Enoch et al. 2006)

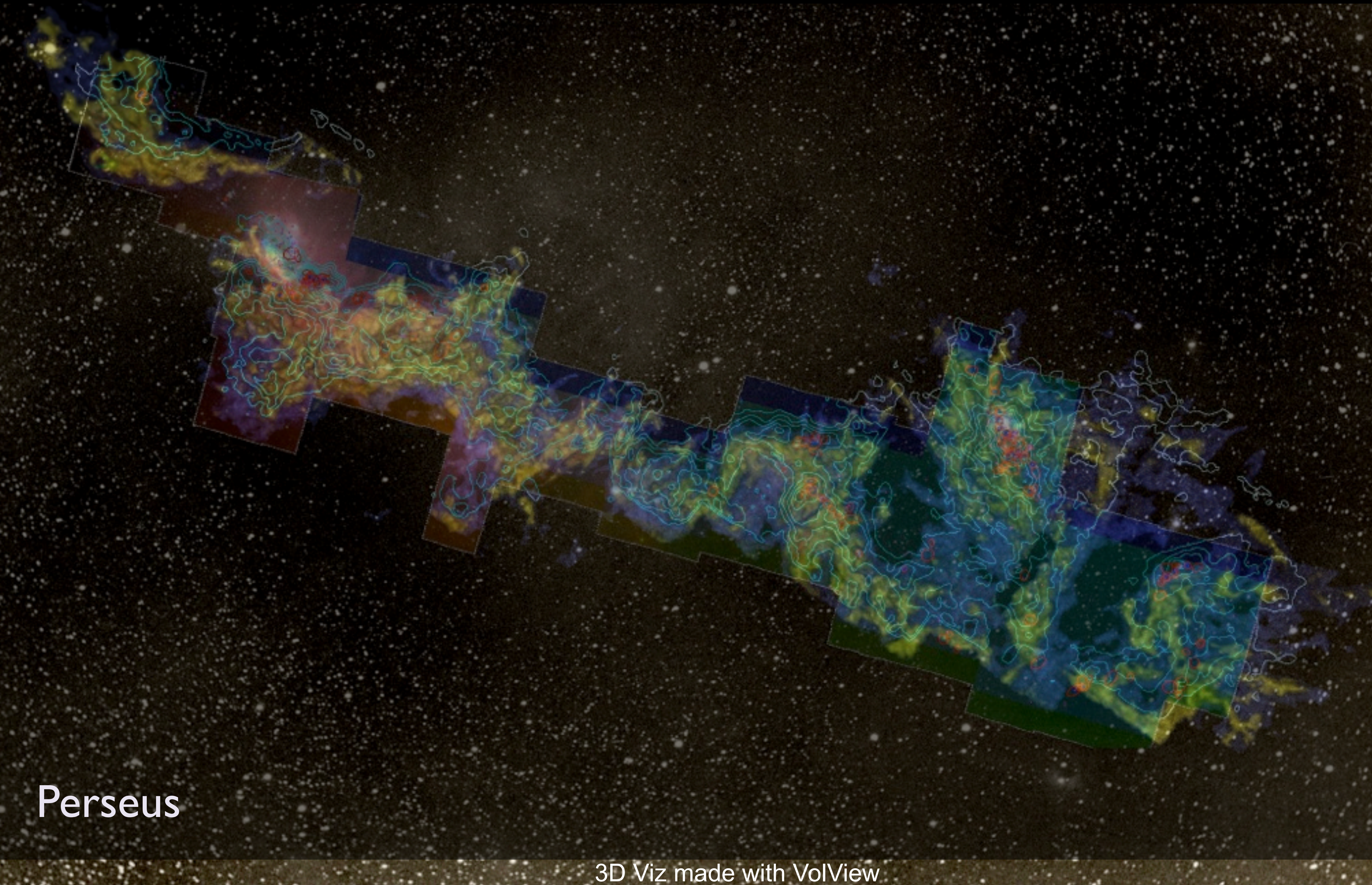
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

^{13}CO (Ridge et al. 2006)

mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al. in prep.)

Optical image (Barnard 1927)

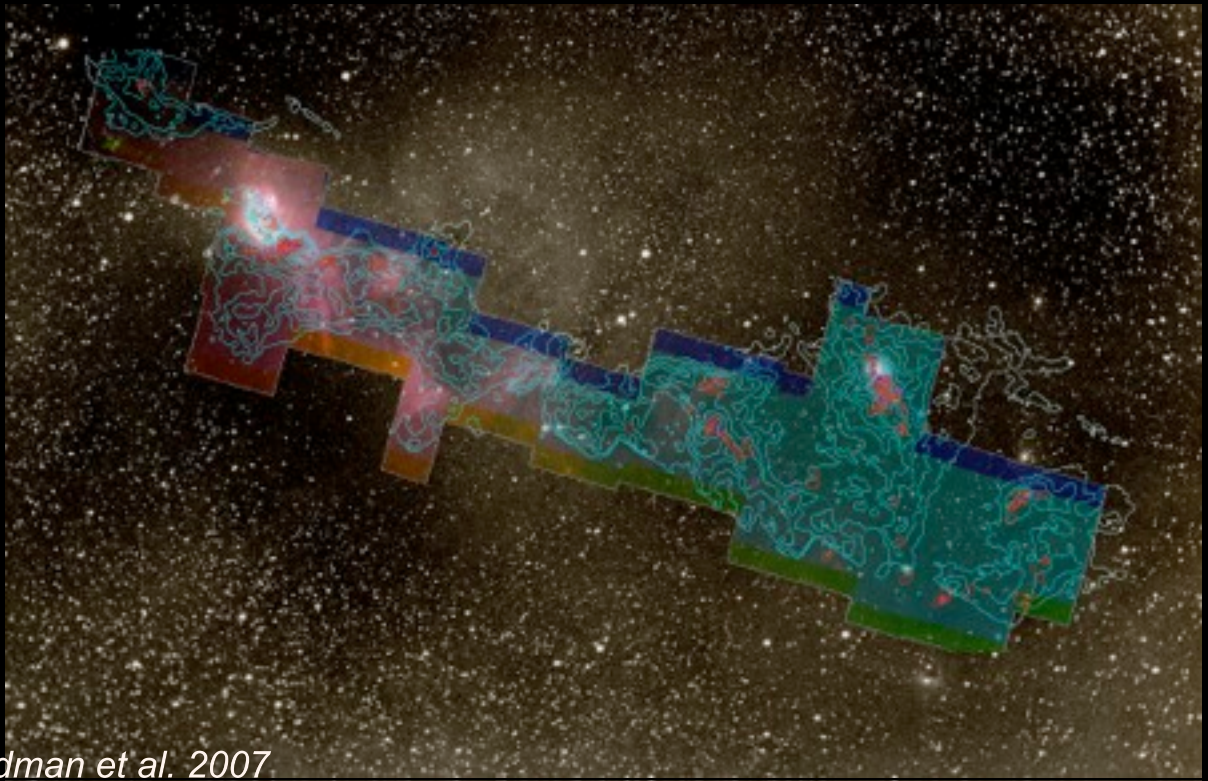




AstronomicalMedicine@IIC

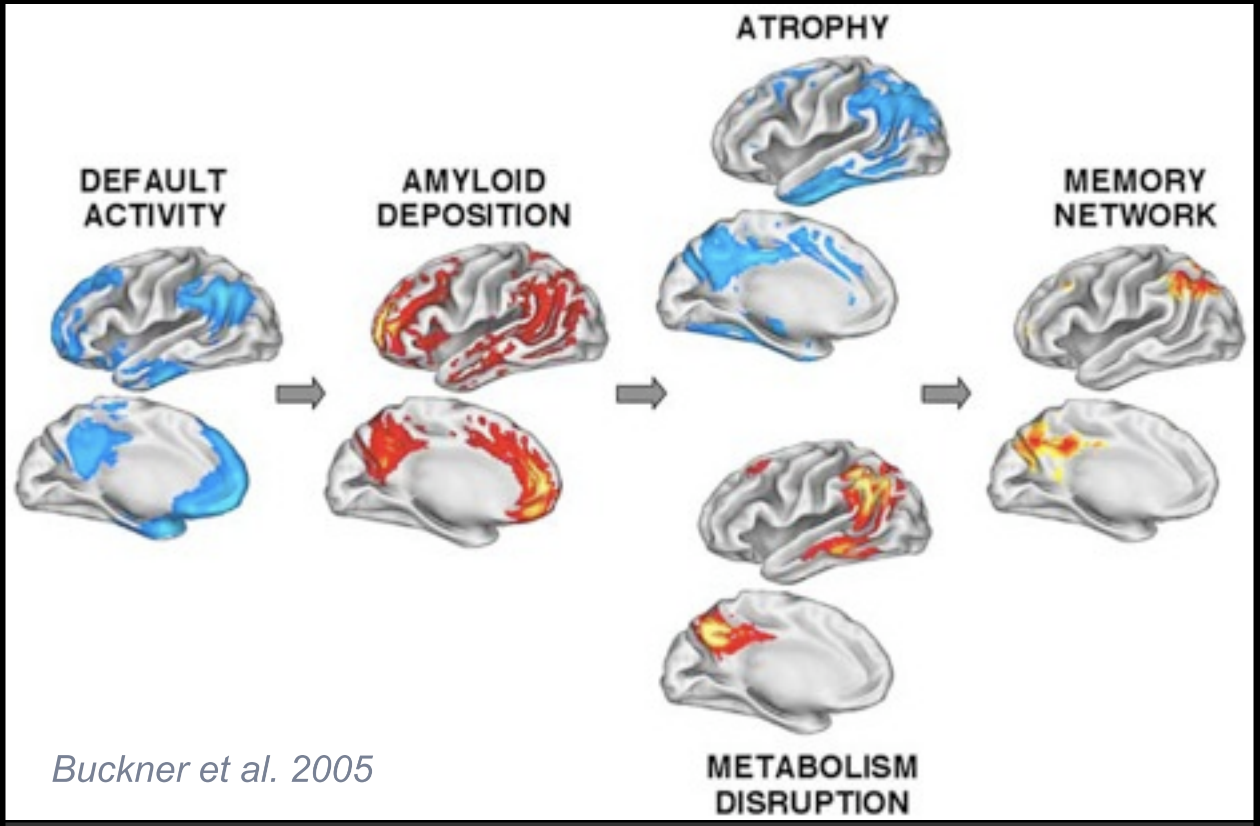
COMPLETE

Astronomers are not alone...
high-D data sets are everywhere in science



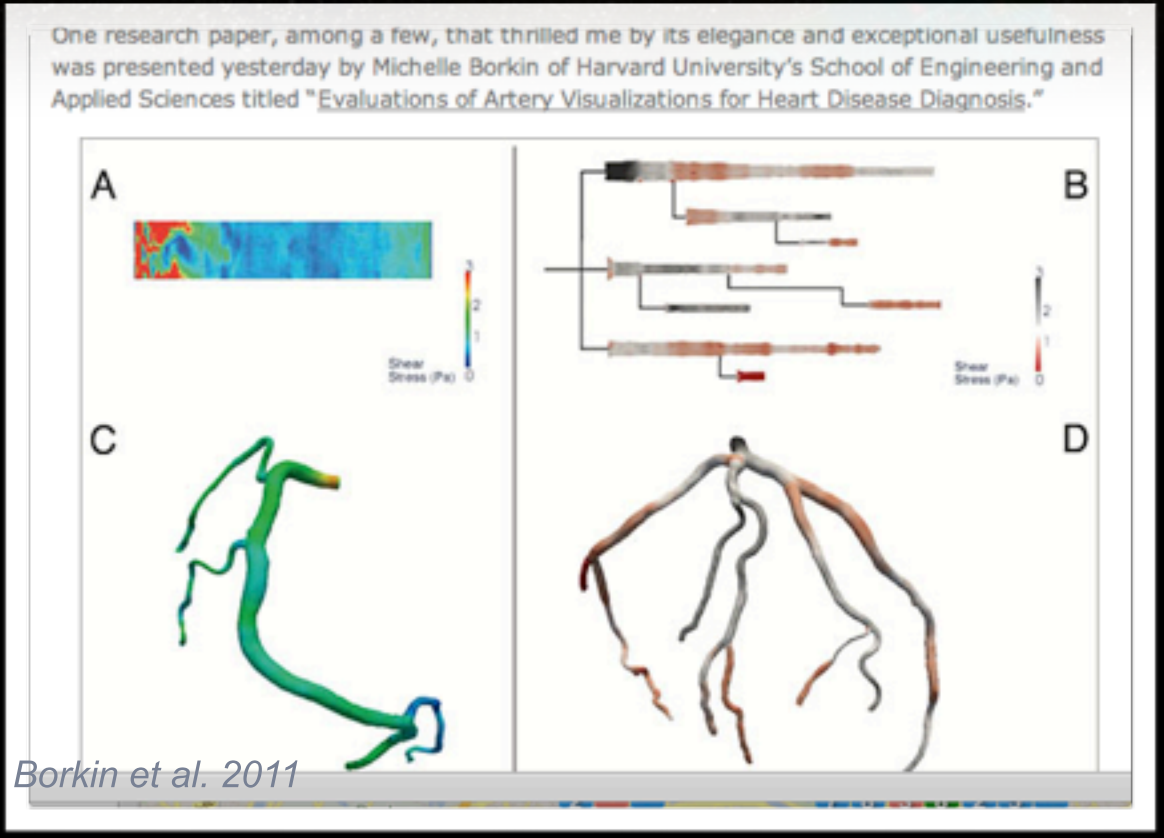
Goodman et al. 2007

Astrophysics



Buckner et al. 2005

Brain Science



Borkin et al. 2011

Blood Flow

What...

...is easier now than before?

fast computation, animation, 3D

...was easier before than now?

craftsmanship

...should be easier in the future?

modular craftsmanship, linked views

The "Easier" Future, for Everyone
Modular Craftsmanship & Linked Views

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

more display modes available (**3D PDF**, **touch** interfaces, stereo+)

re-usable tools/mashups (Many Eyes, Google Maps+, crowdsourcing)

live, interactive **linked views** (DataDesk, Tableau, GapMinder, ***WWT***...)

Unsolved Questions...

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

improved graphical representation of uncertainty display modes

3D PDF

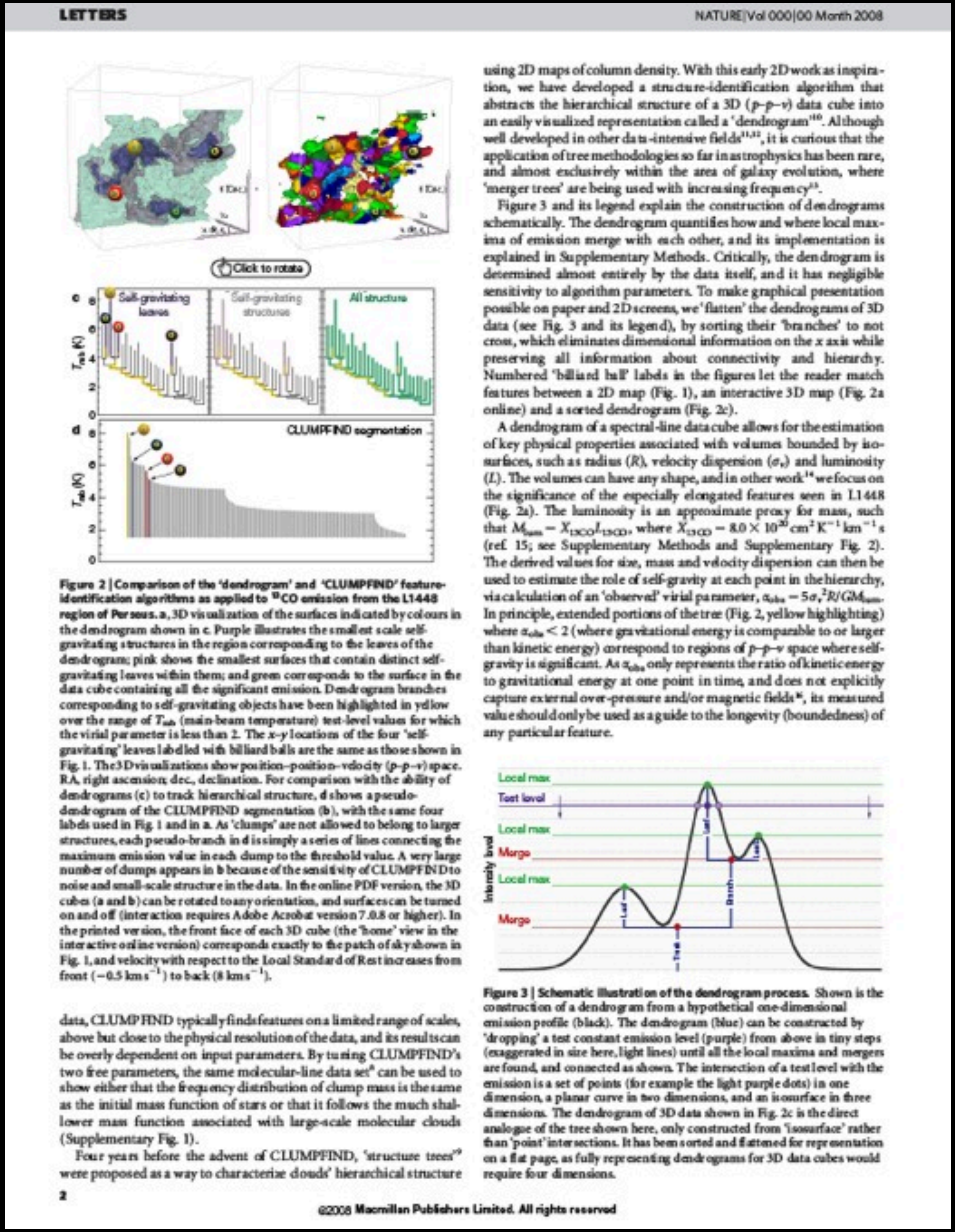


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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).

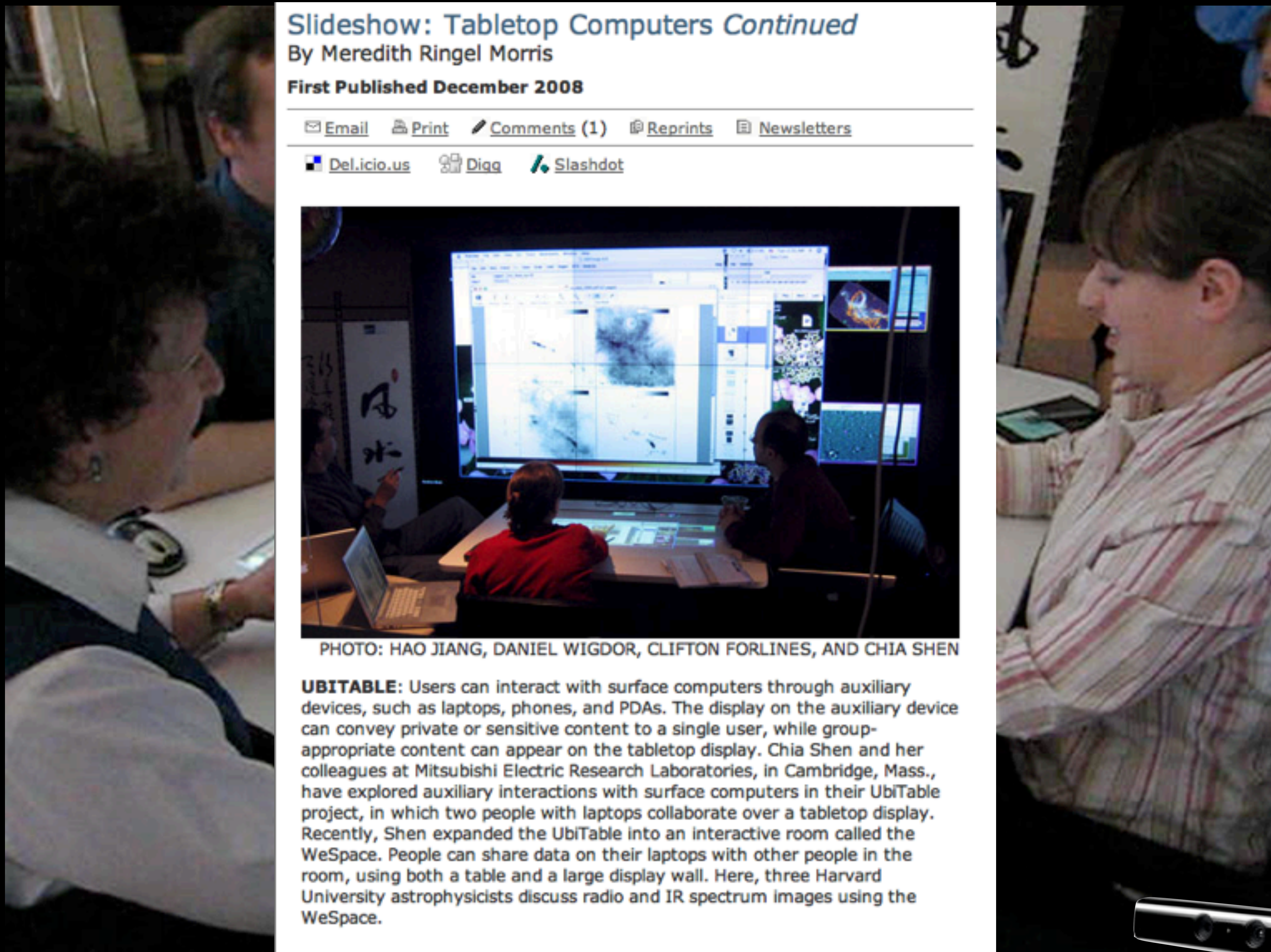
Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure 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 $X_{13CO} = 8.0 \times 10^{20}\ \mathrm{cm^2\,K^{-1}\,km^{-1}\,s}$ (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, $\alpha_{obs} = 5\sigma_v^2 R/GM_{lum}$. 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.

"Off the Desktop"



Slideshow: Tabletop Computers *Continued*

By Meredith Ringel Morris

First Published December 2008

Email Print Comments (1) Reprints Newsletters

Del.icio.us Digg Slashdot

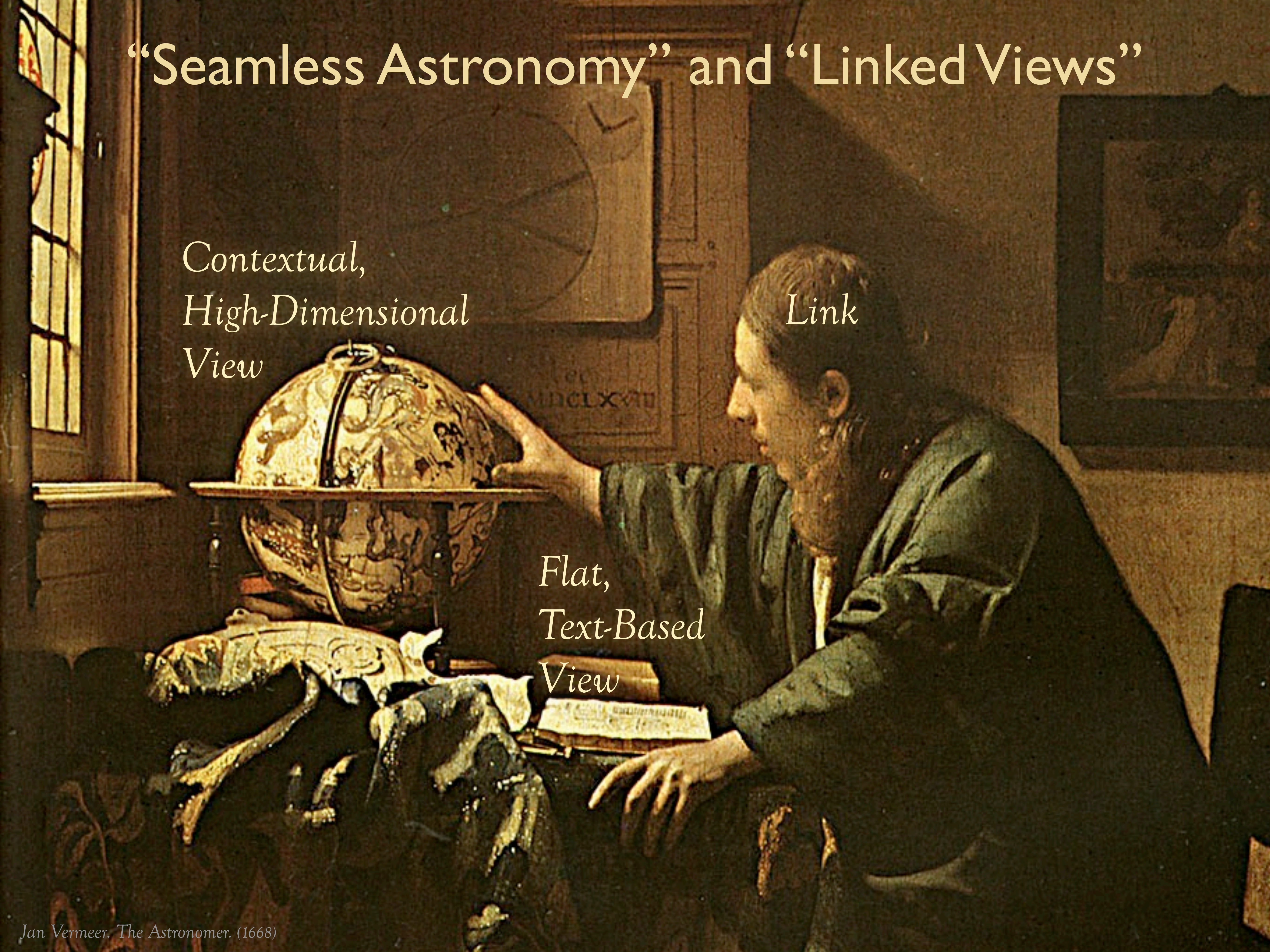
PHOTO: HAO JIANG, DANIEL WIGDOR, CLIFTON FORLINES, AND CHIA SHEN

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.

movie courtesy Daniel Wigdor, taken at MERL, Kendall Square, Cambridge, 2007

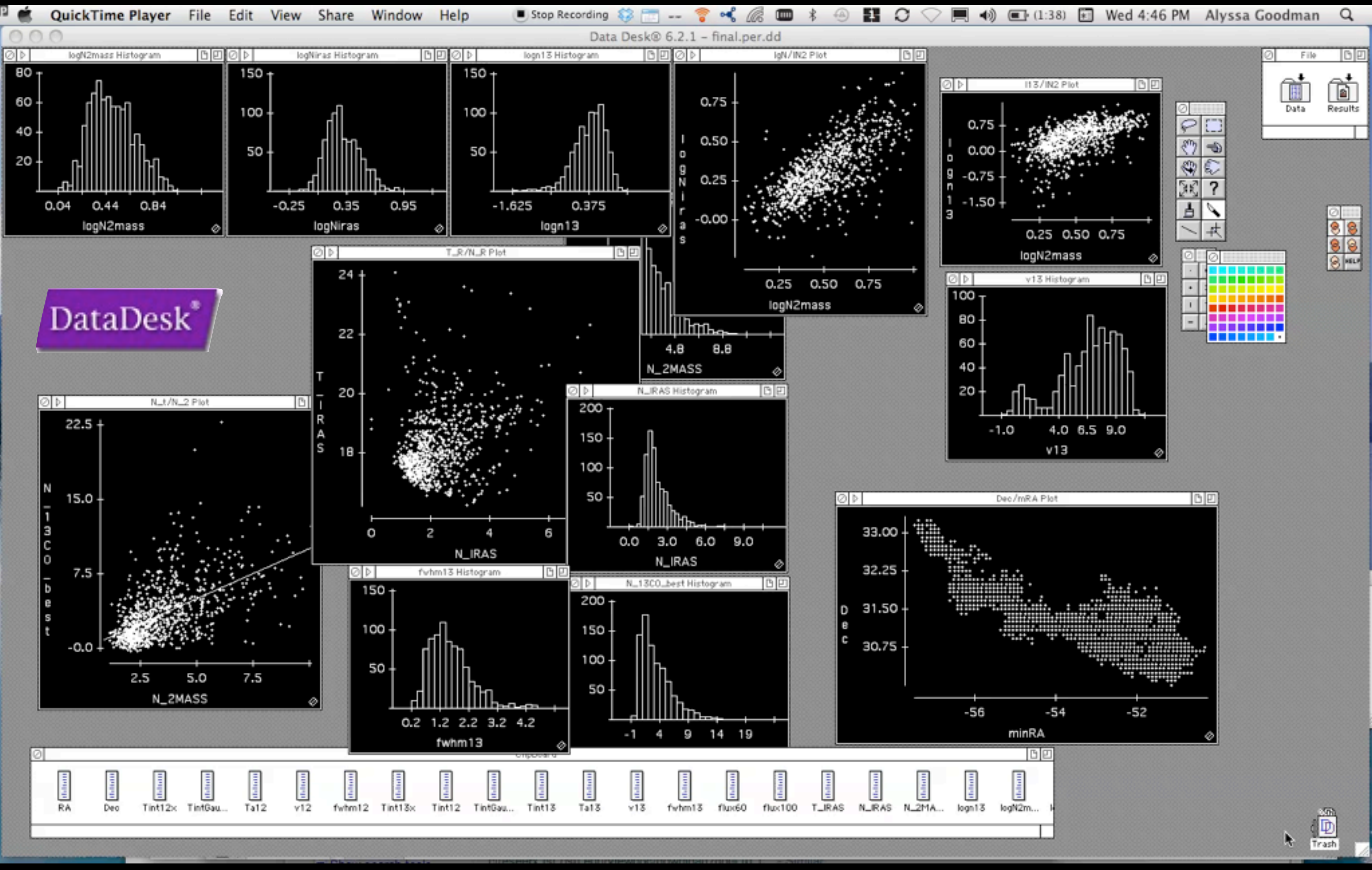
*How can we advance the **digital** tools for scientists?*

“Seamless Astronomy” and “Linked Views”



Jan Vermeer. The Astronomer. (1668)

DataDesk (est. 1986)



John Tukey's "Four Essentials" (c.1972)

Picturing

Rotation

Isolation

Masking

Selection

and these *"need to work together"*
in a *"dynamic display"*

Brushing

Linking

Results...

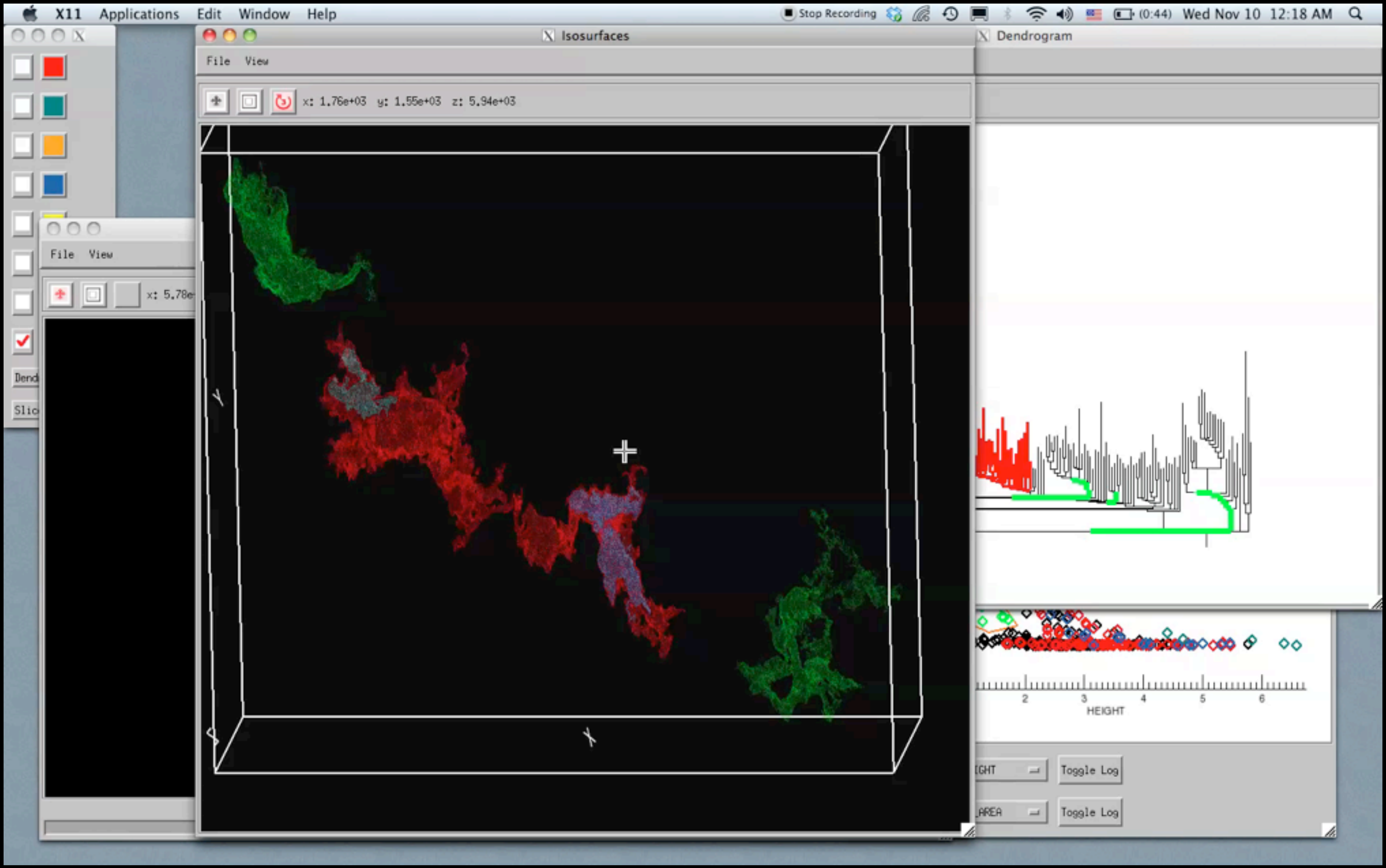
1. for immediate insight
2. as visual source of ideas for statistical algorithms (...relation to SVM)

Warning

"details of control can make or break such a system"

Watch the PRIM-9 video at: http://stat-graphics.org/movies/prim9.html

Exemplar: Linked Dendrogram Views in IDL

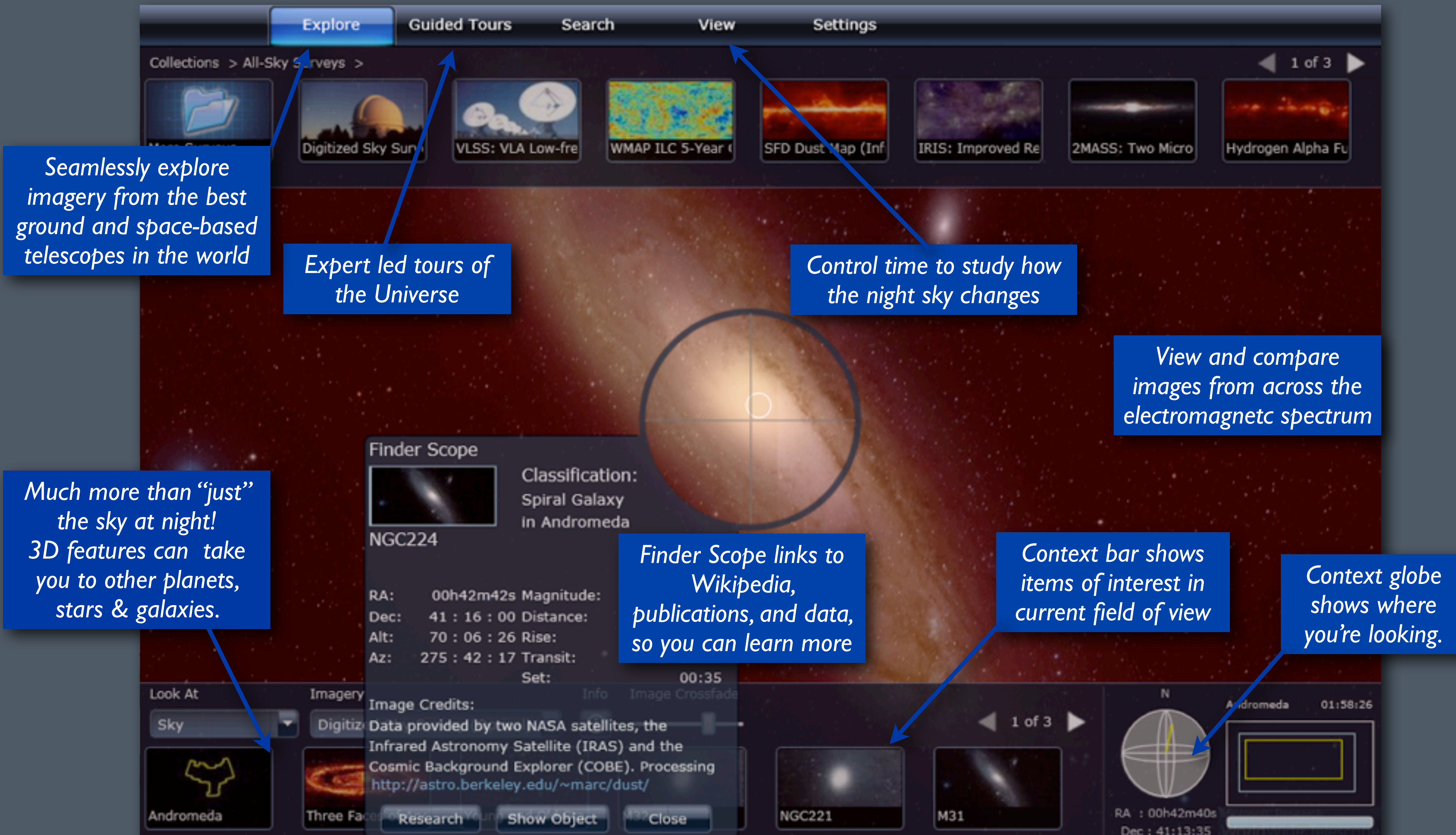


Video & implementation: Christopher Beaumont, CfA/UHawaii;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

*How can we advance the **digital** tools for scientists?*

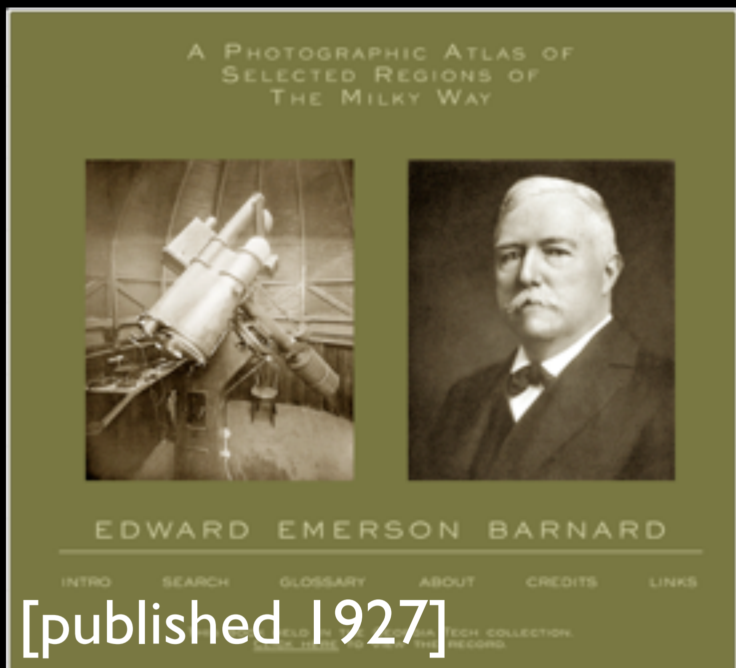
Microsoft® Research WorldWide Telescope

Experience WWT at **worldwidetelescope.org**

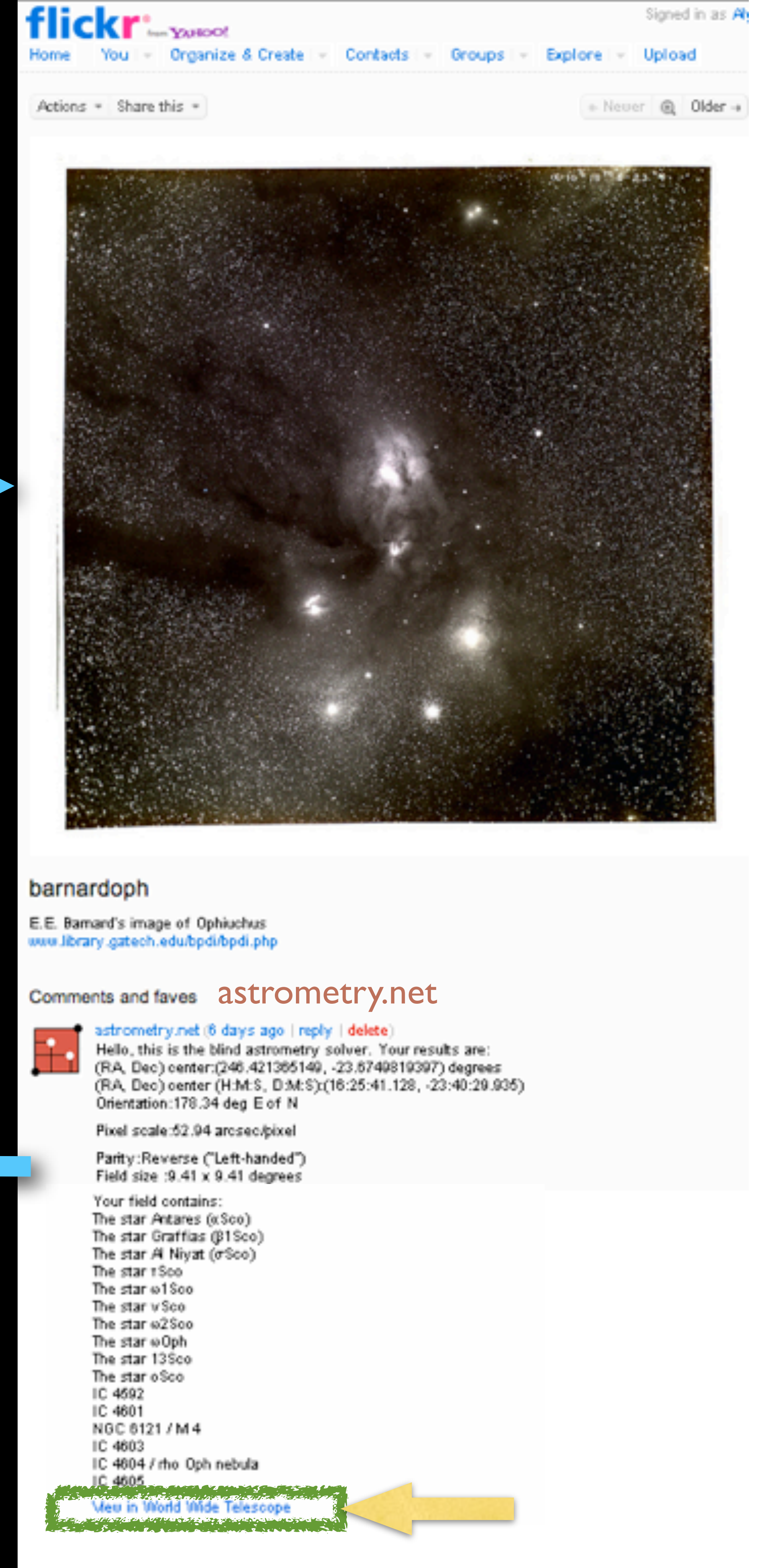


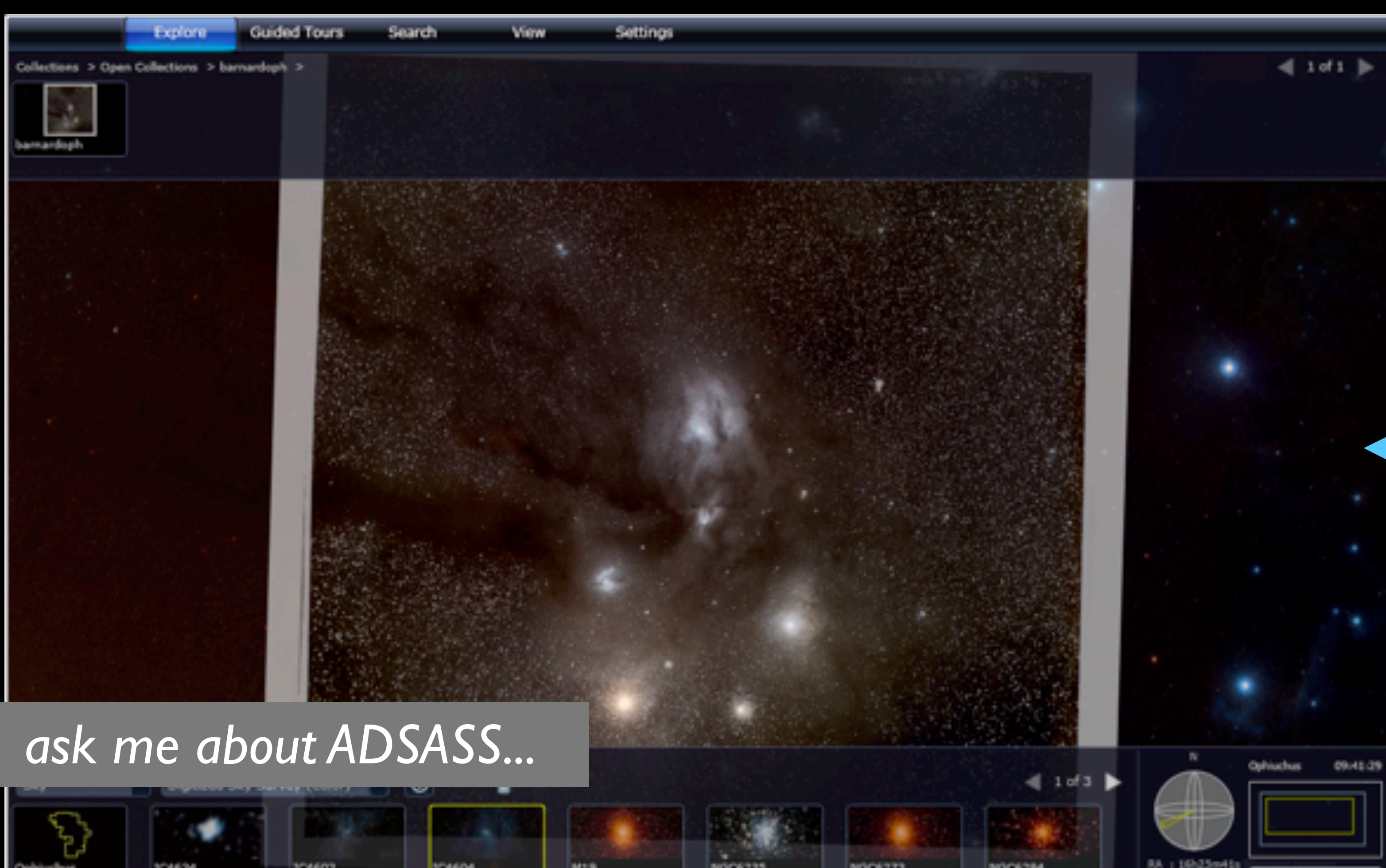
“Seamless Astronomy”...

astrometry.net + flickr + WWT



[published 1927]





ask me about ADSASS...

WWT Ambassadors: WorldWide Telescope For Interactive Learning

Alyssa Goodman
Harvard University Professor of Astronomy, Microsoft Academic Partner

Pat Udomprasert
WWT Program Coordinator

Curtis Wong
Microsoft Research, WWT Creator

Stephen Strom
NOAO, WWTA Tucson Site Advisor

Sarah Block
Web site development

wwtambassadors.org

Microsoft® Research
WorldWide Telescope

Ambassadors Program

Home | About | My Account | Find Tours | Resources | Log Out | Search

About the WWT Telescope Ambassadors Program

WorldWide Telescope (WWT) is a rich visualization environment that functions as a virtual telescope, allowing anyone to make use of professional astronomical data to explore and understand the universe. As of early 2010, the new WWT Ambassadors Program is recruiting astronomically-literate volunteers, including retired scientists engineers—all of whom will be trained to be experts in using WWT as a teaching tool. Ambassadors will give volunteer presentations at public libraries, community centers, museums, and schools, demonstrating WWT's power to help laypeople visualize and understand our universe.

Read more

John Huchra's Universe

Submitted by patudom on Jan. 11

John Huchra, former president of the **American Astronomical Society**, passed away on October 8, 2010.

John's colleagues at the Harvard-Smithsonian Center for Astrophysics, in collaboration with the creators of WorldWide Telescope at Microsoft Research, have created a new, interactive, WWT Tour to honor John and his career. The Tour primarily focuses on John's quest to map the Universe in three dimensions. It is 12.5 minutes long.

The Tour is best experienced inside the WorldWide Telescope program itself. (**Note: You must have the version of WWT released on 1/13/2011 to view all of this Tour's content. You can download it from here.**) As viewed within the WWT program, the Tour content is interactive, allowing users to pause and explore the parts of the Universe featured in the tour, explore web hyperlinks, and more. For those who do not have the desktop client, the Tour has been posted as a video as well.

Video (Interactive WWT features will be disabled)

John Huchra's Universe

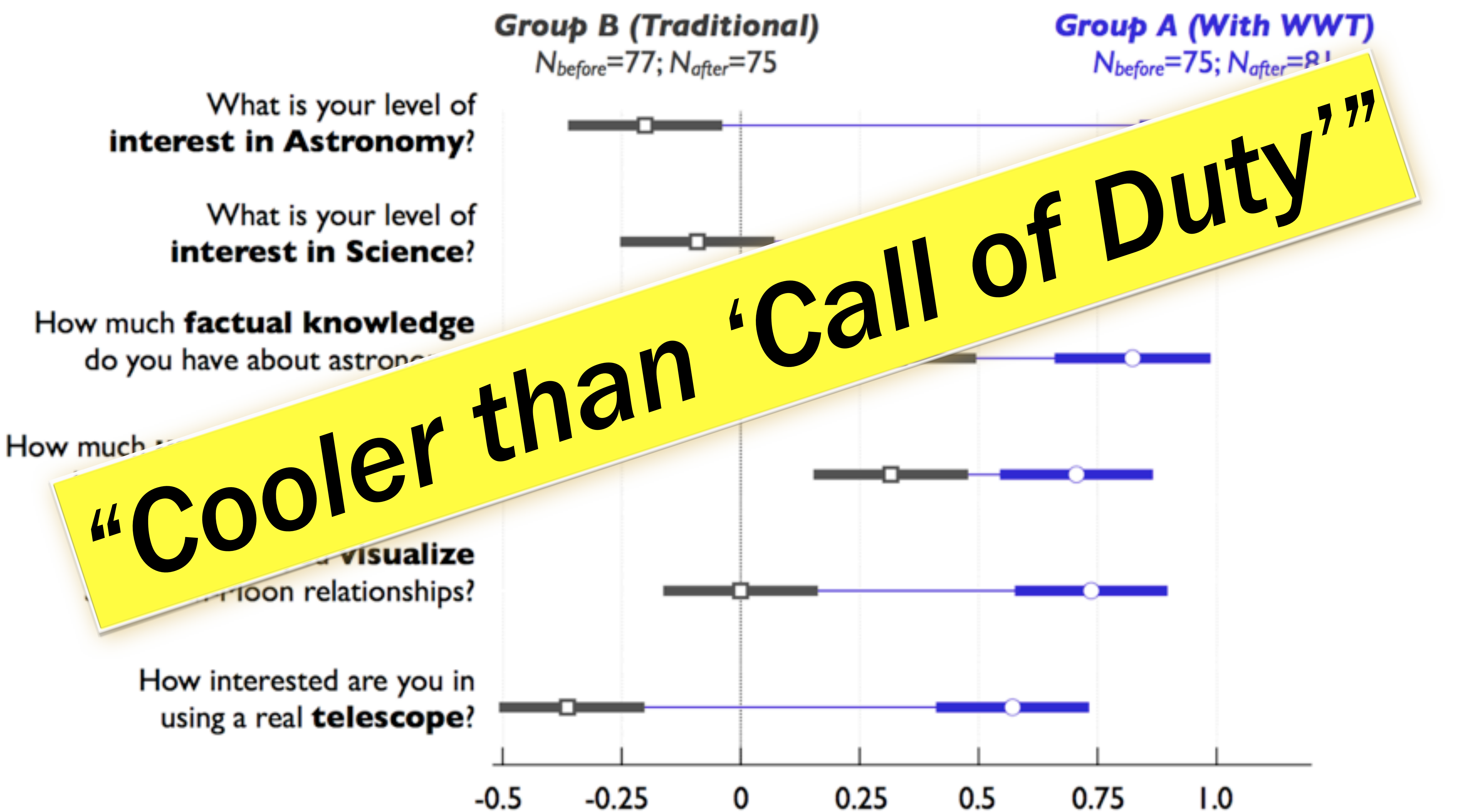
Friends of John Huchra have released a new WWT Tour to honor John and his work. The Tour primarily focuses on John's quest to map the Universe in three dimensions. You can view the Tour here.

Upcoming

- Cyberlearning Tools for STEM Education Conference
  Mar. 8 - Mar. 9
- Cambridge Science Festival
  Apr. 30 - May. 10

Gains in Student Interest and Understanding

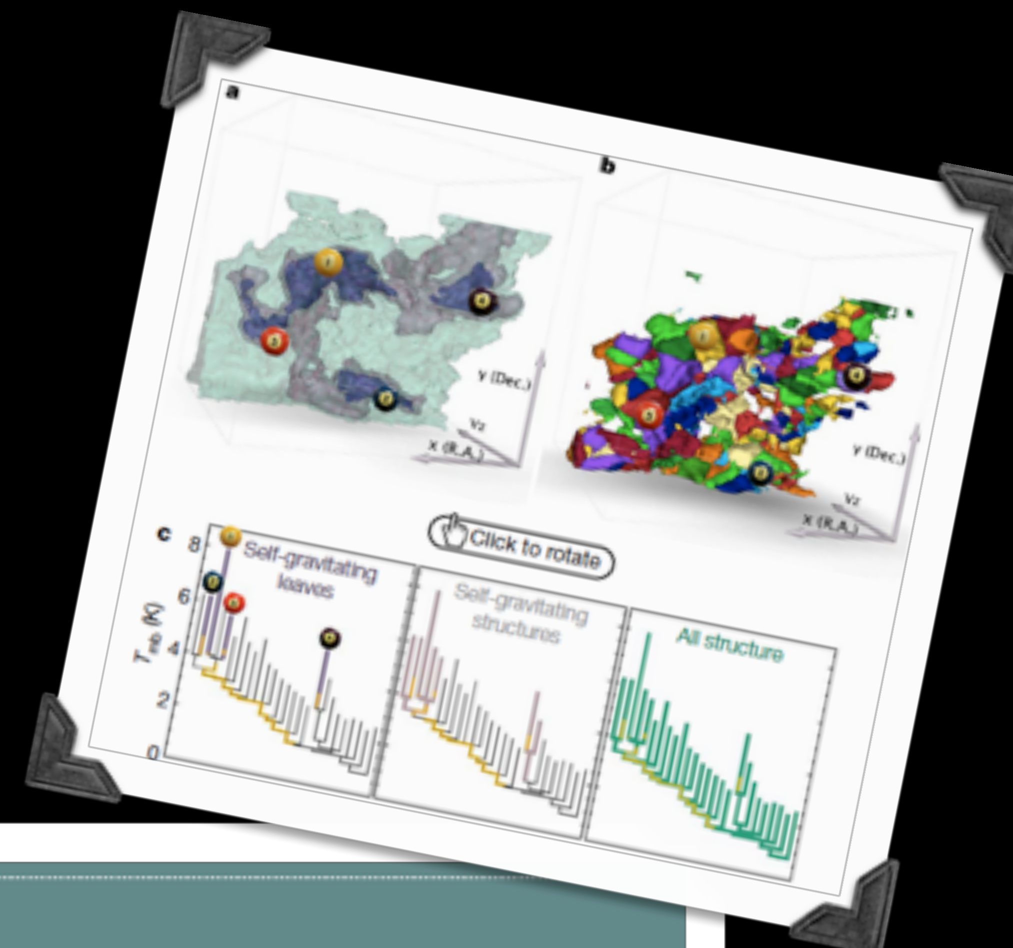
("Traditional Way" vs "WWT Way")

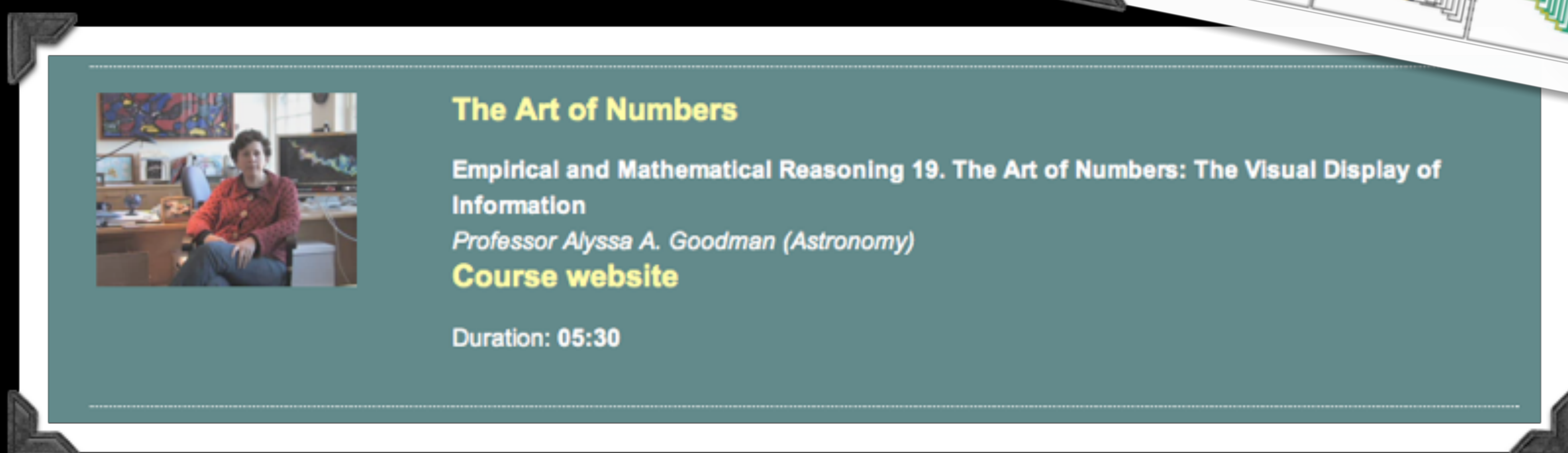


Effect Size: Gain (or Loss) in Units of Pre-Test Standard Deviation
(Error bars show ± 1 Standard Error of the Mean)

Seeing Science

Data Visualization in Modern Research (and teaching!)

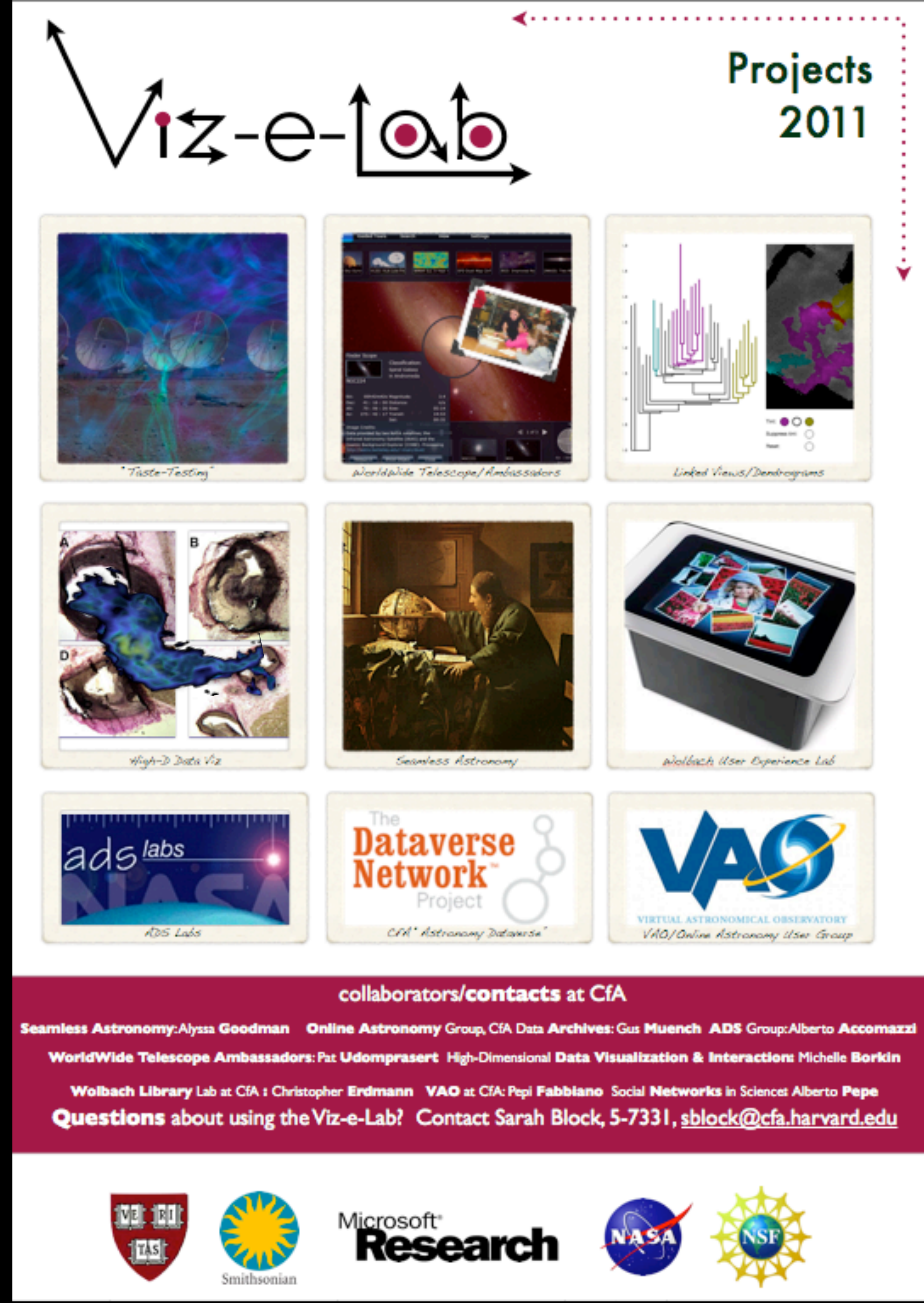




Alyssa A. Goodman
Harvard University (HCO+IIC)
Smithsonian Astrophysical Observatory
Scholar-in-Residence, WGBH

Viz-e-Lab

The next challenges…



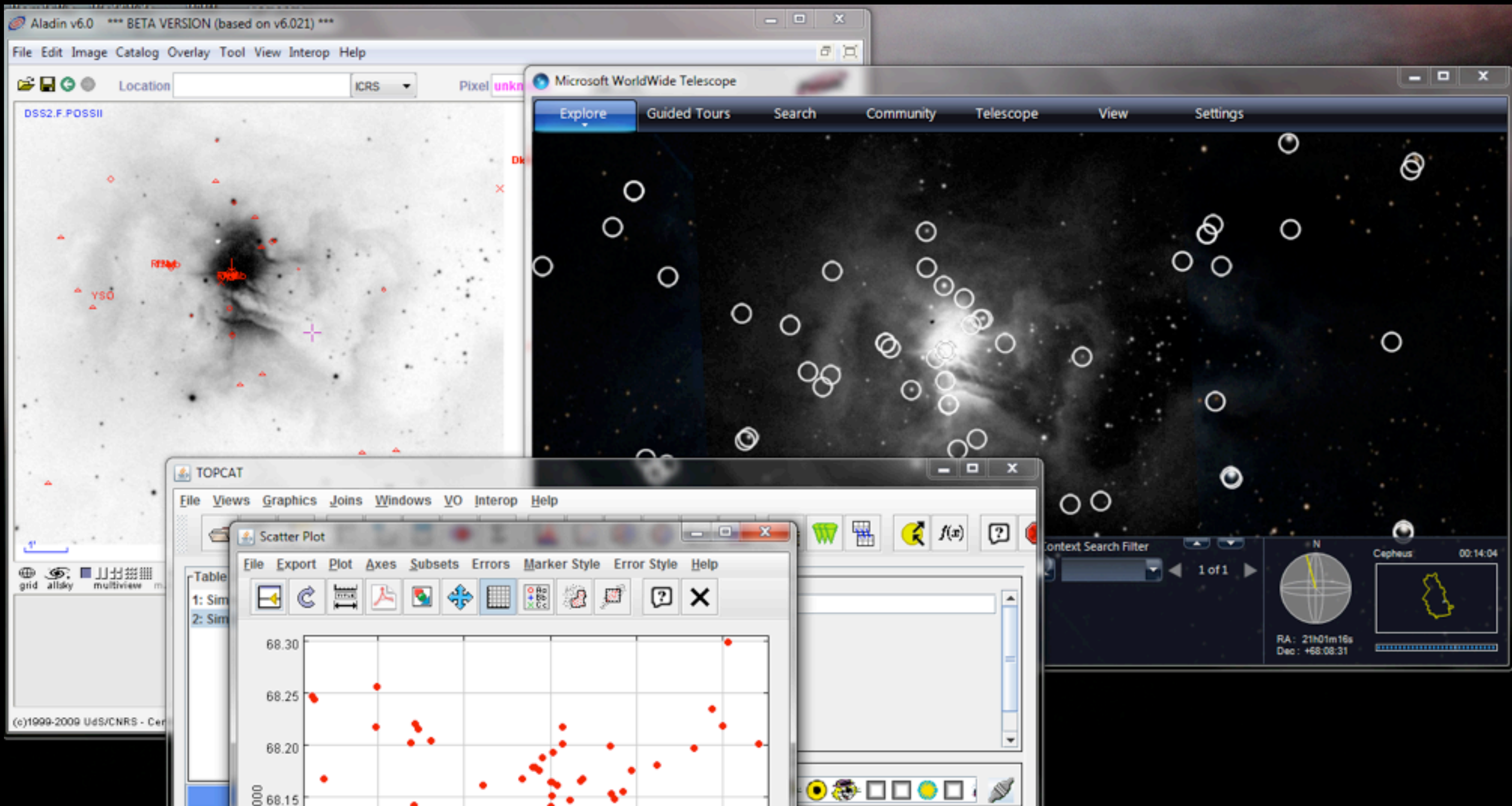
Challenge #1: 3D Selection

Wh

How

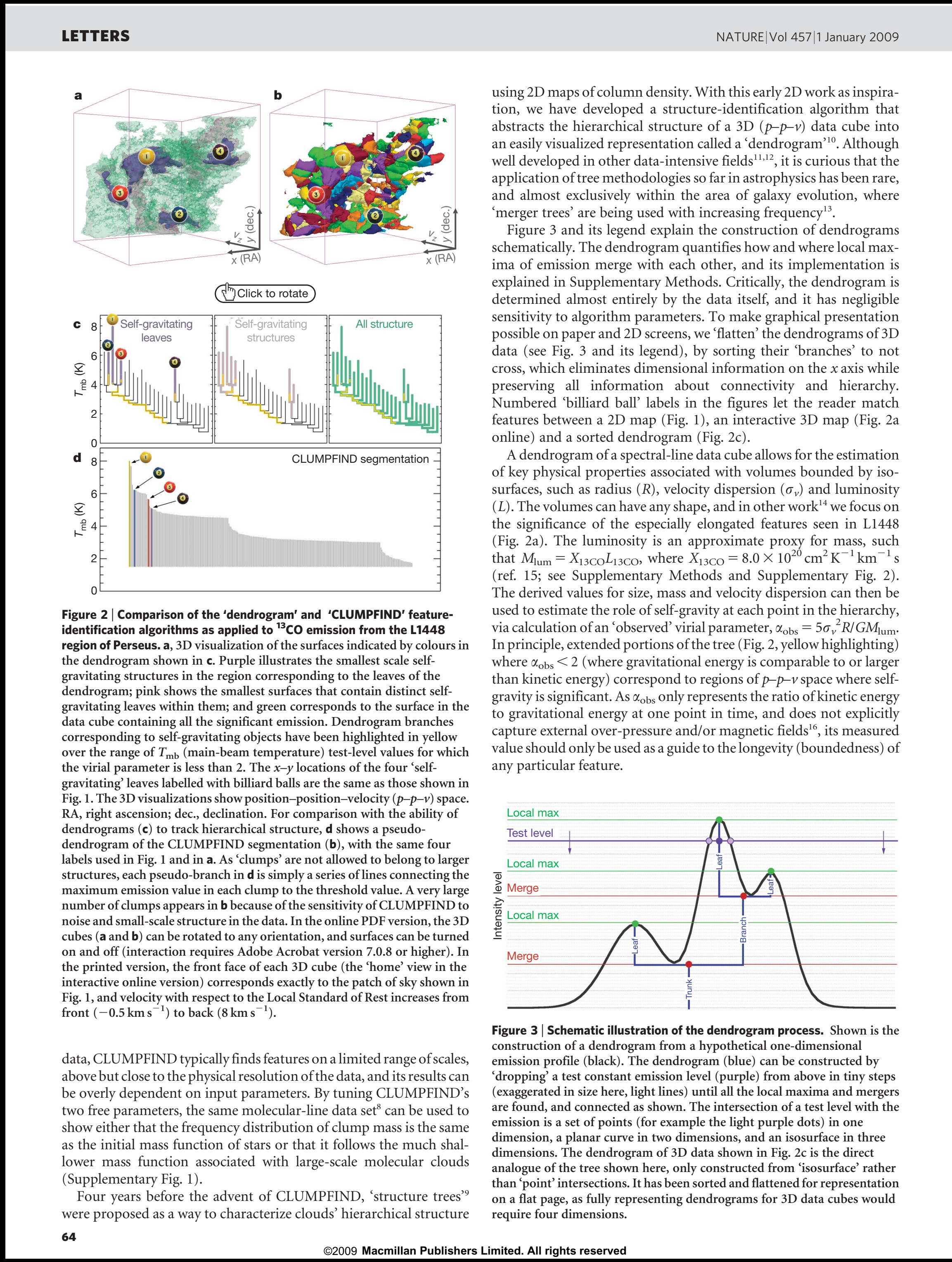
How on?

Challenge #2: Too many windows...



Challenge #3:

What does "Publication-Quality" Graphics Mean in an Interactive 3D World?



a b

y (dec.) x (RA)

Click to rotate

c Self-gravitating leaves Self-gravitating structures All structure

d 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, $\alpha_{obs} = 5\sigma_v^2 R/GM_{lum}$. 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.

Local max
Test level
Local max
Merge
Local max
Merge
Intensity level
Leaf
Leaf
Leaf
Branch
Trunk

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.

*Goodman, Rosolowsky, Borkin, Foster, Halle, Kauffmann & Pineda, **Nature**, 2009*