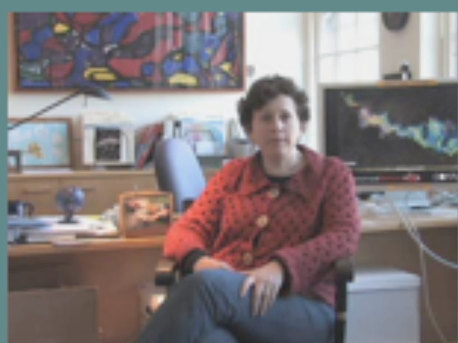
WorldWide Telescope

Microsoft Research

The Art of Numbers



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

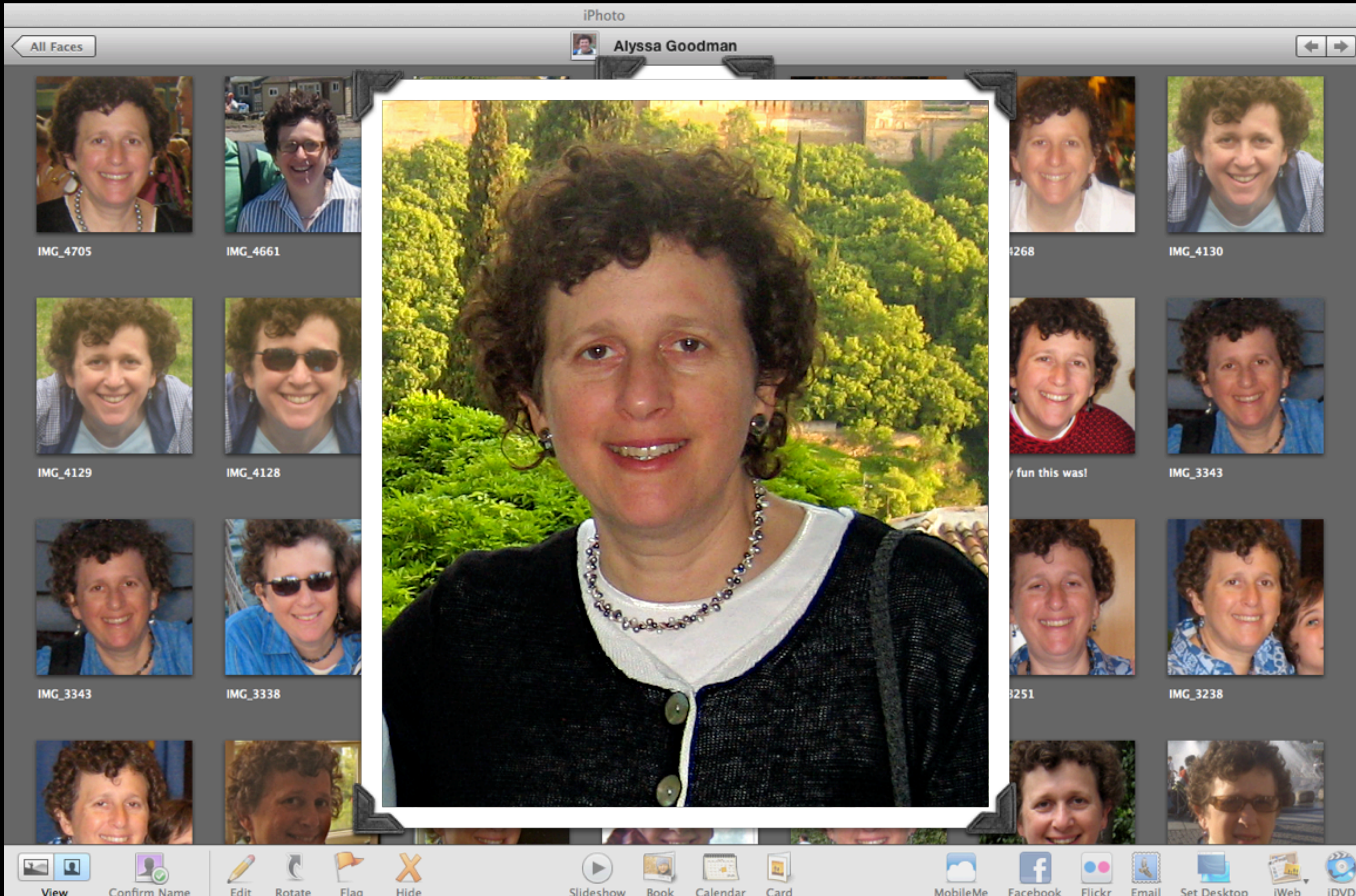
Duration: **05:30**

What kind of credentials are those??

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







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)

The Art of Numbers

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?

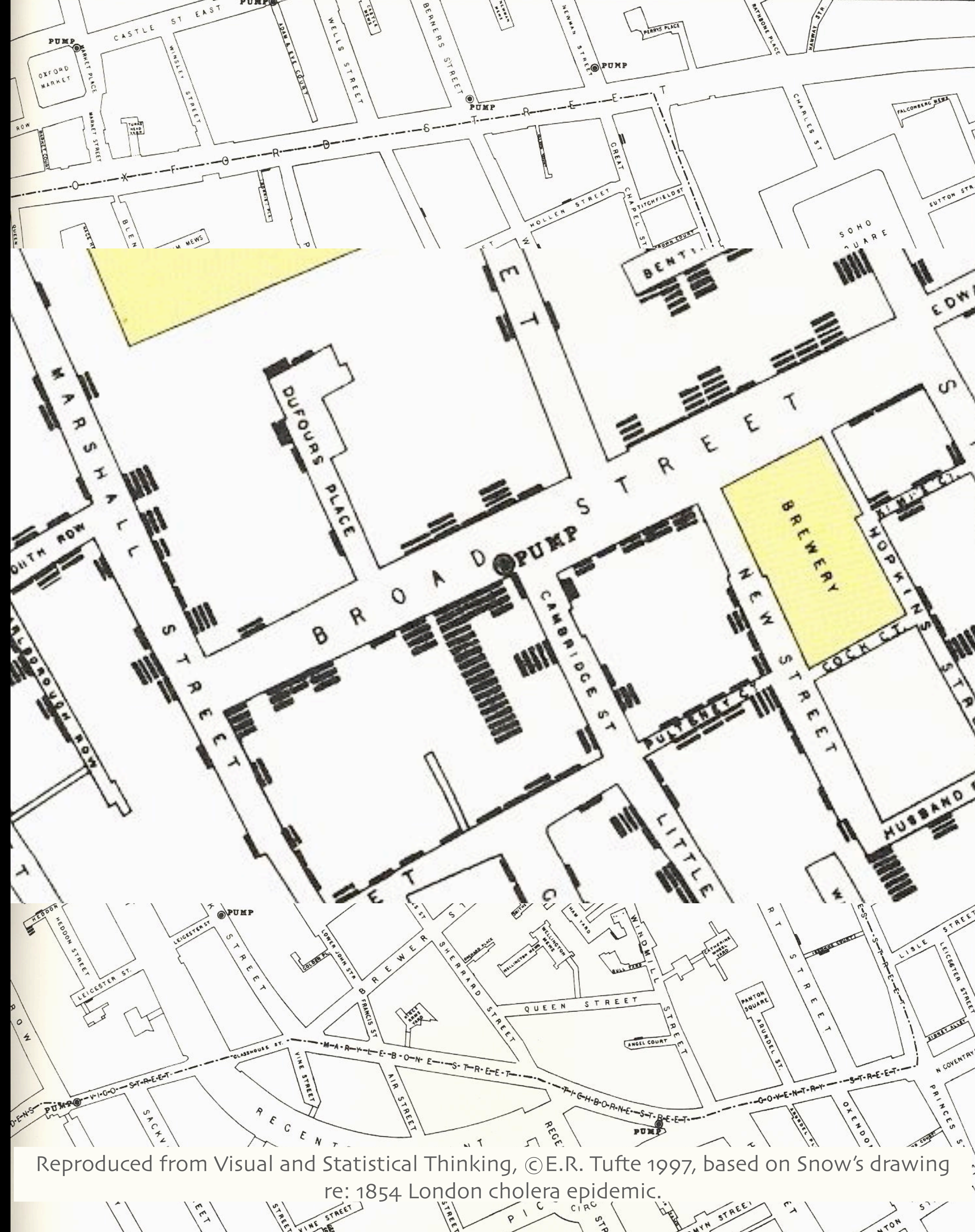
modular craftsmanship, linked views

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

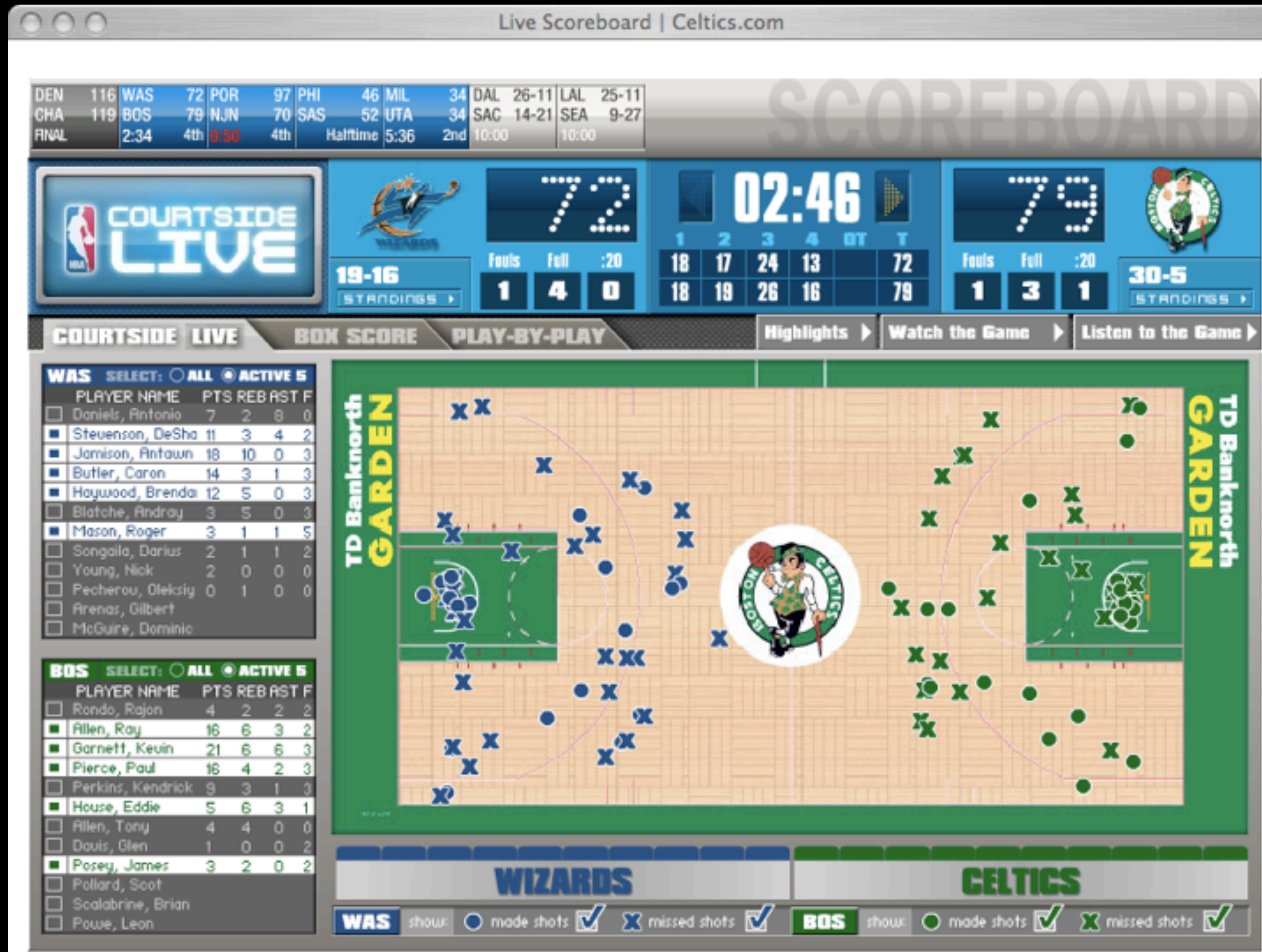
✓Tables

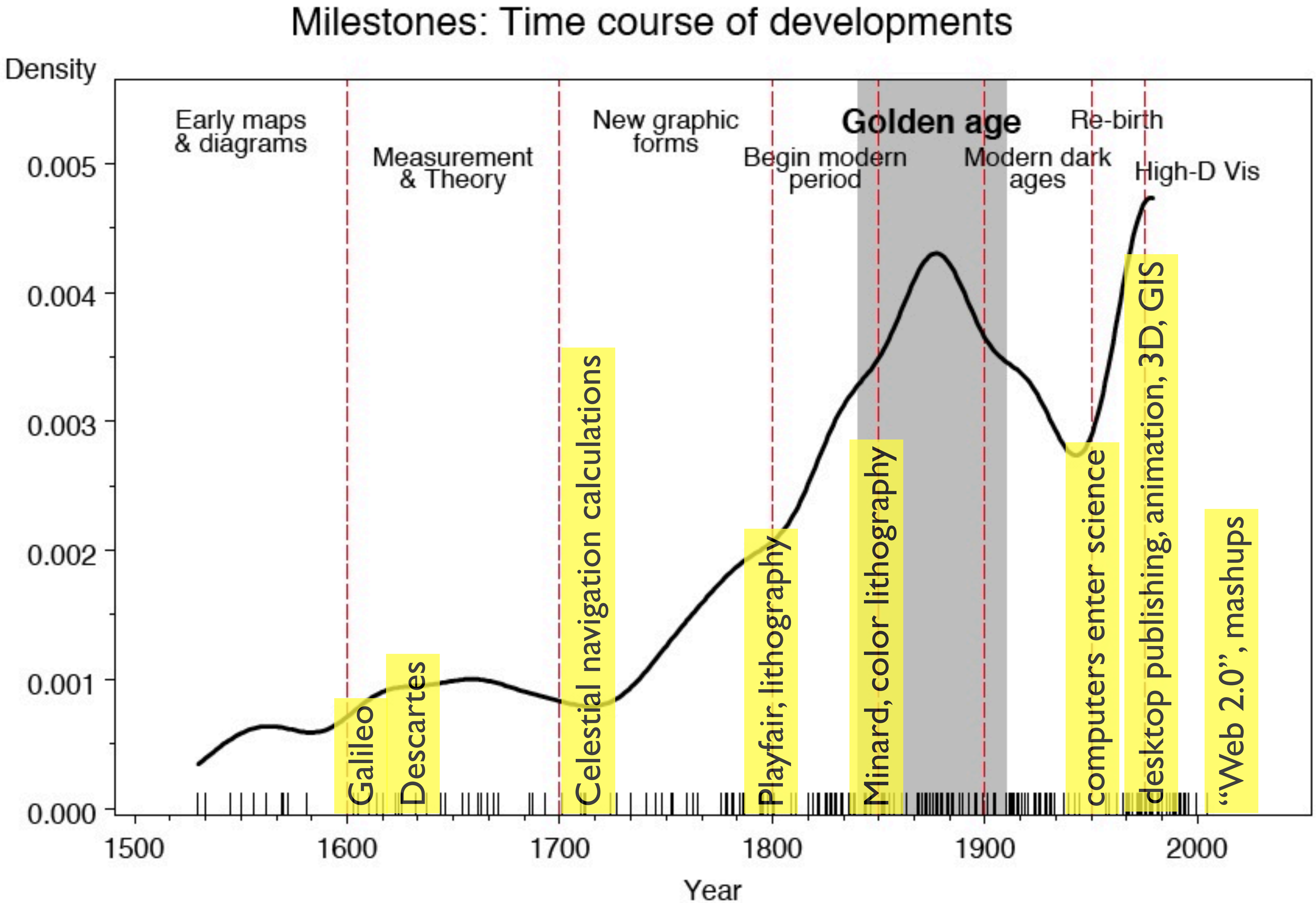
✖Graphs

✓Charts

✓Illustrations

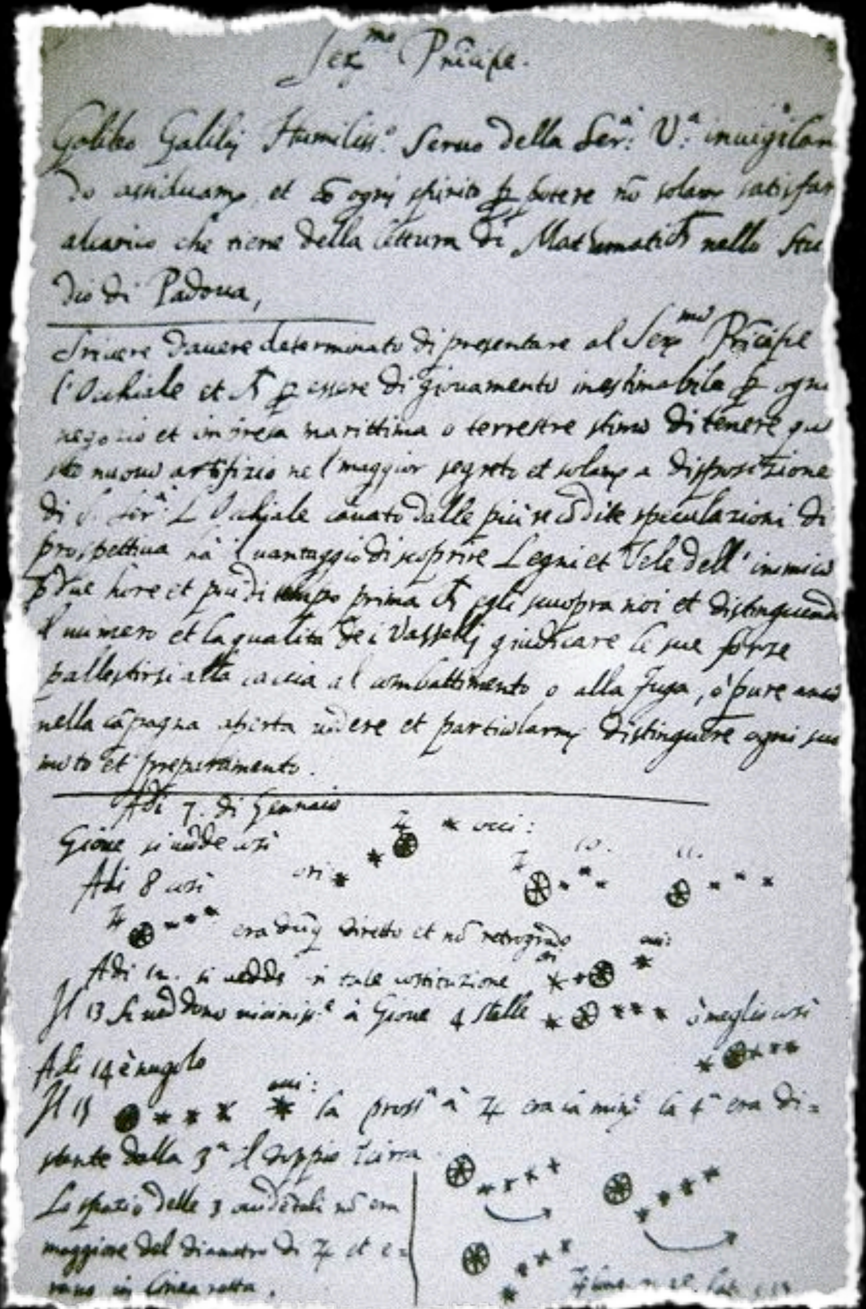
✓Combinations

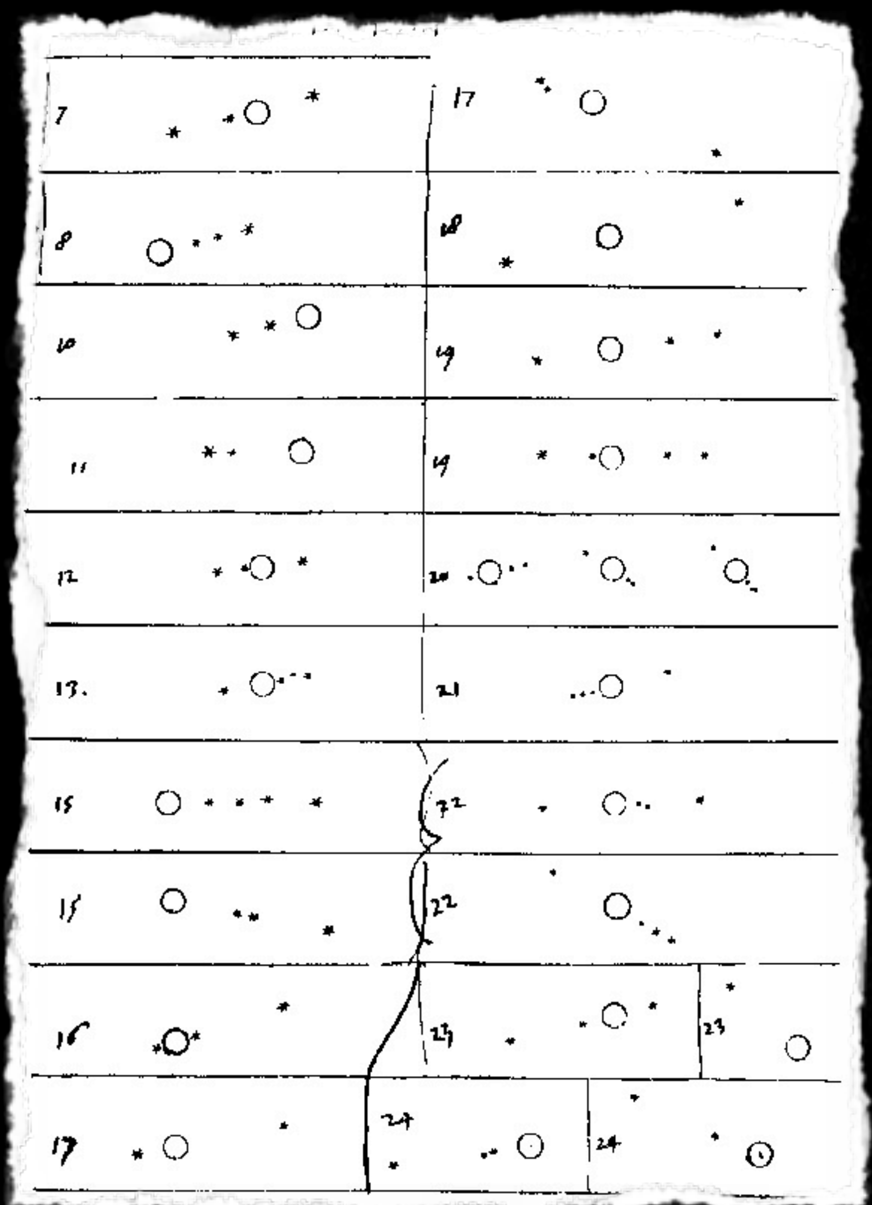


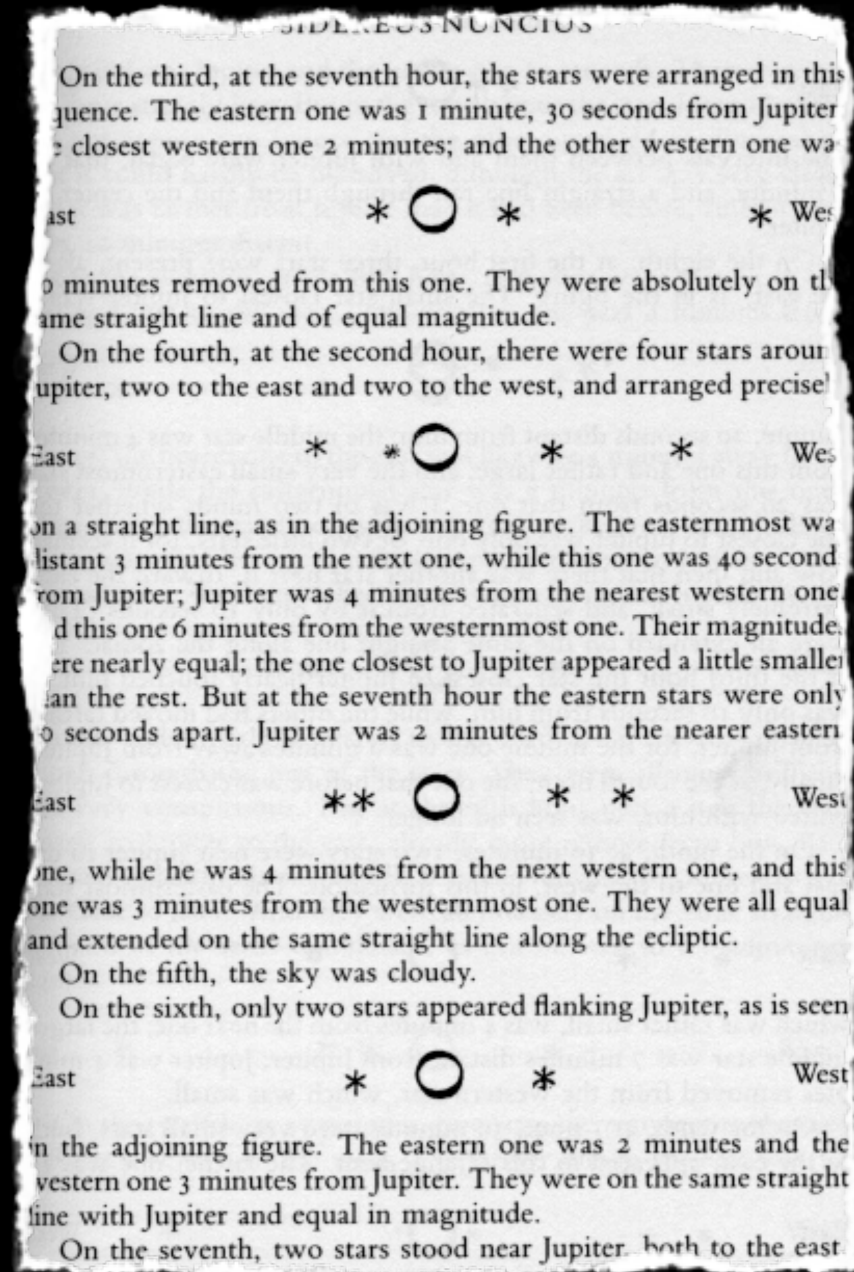


adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, 2009

Galileo 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

East * O * * West

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 aroun
upiter, two to the east and two to the west, and arranged precise

East * * O * * West

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 eastern

East ** O * * West

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

East * O * West

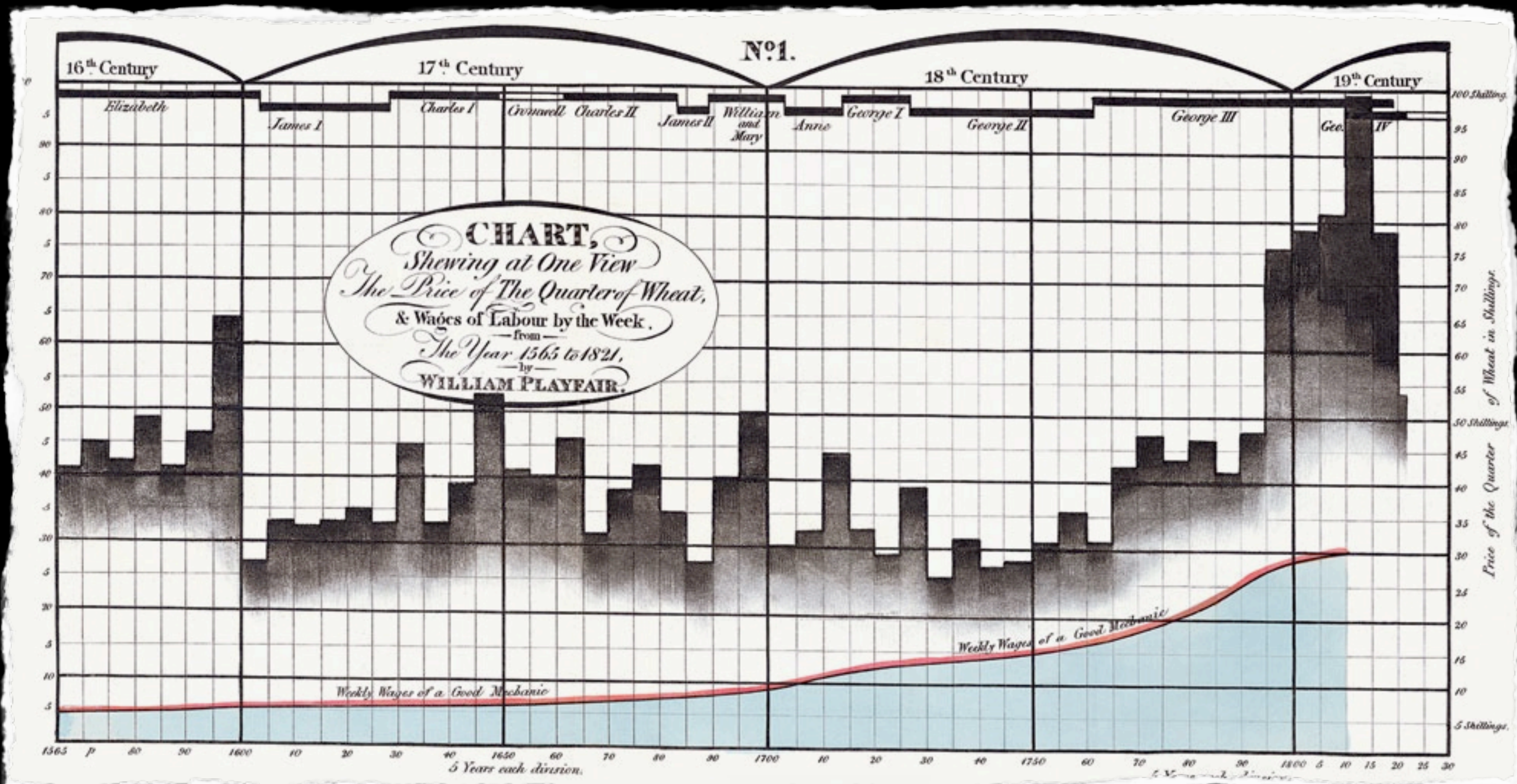
in the adjoining figure. The eastern one was 2 minutes and the
western one 3 minutes from Jupiter. They were on the same straight
line with Jupiter and equal in magnitude.

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

Notes for & re-productions of Siderius Nuncius

Playfair, lithography

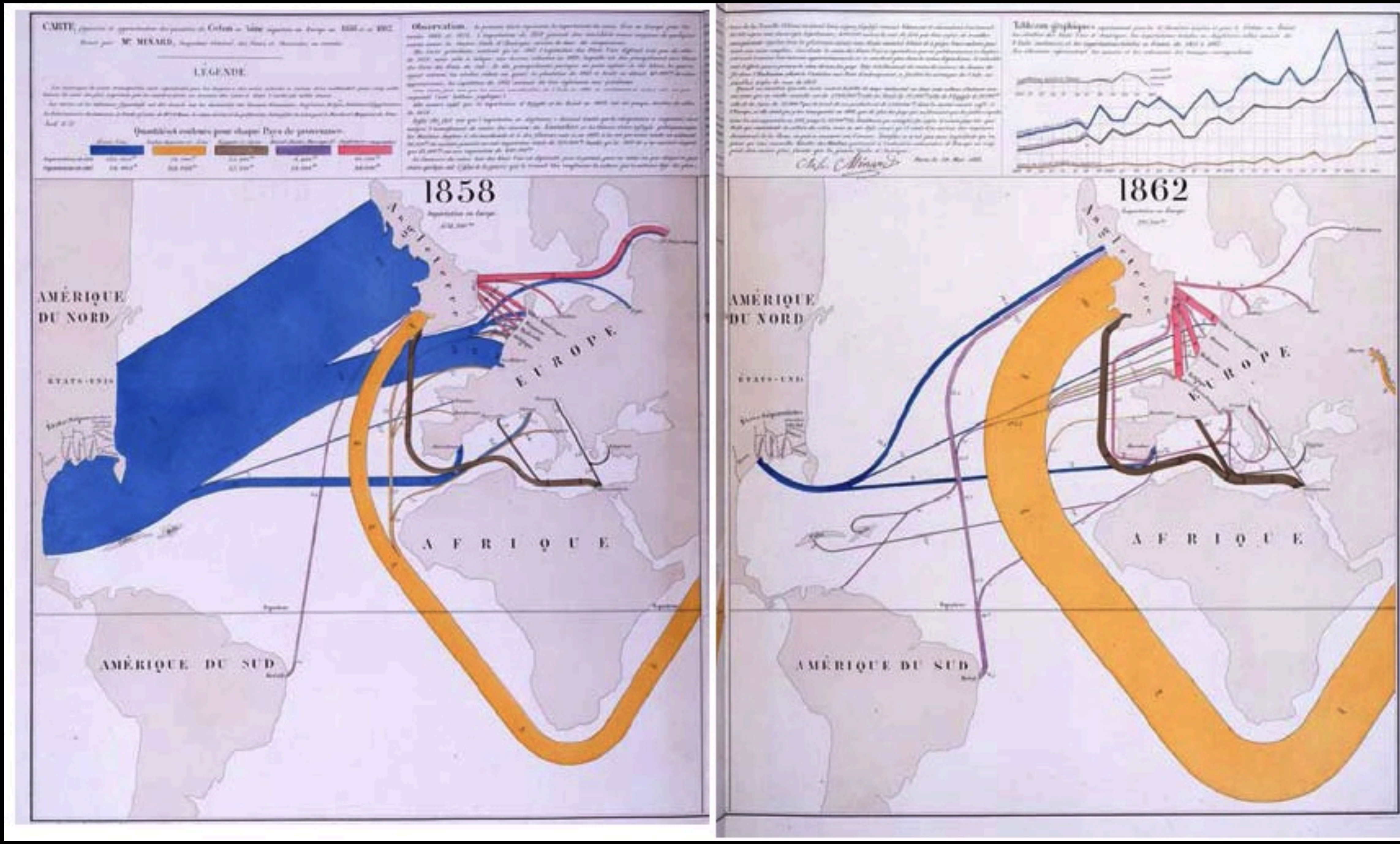
William Playfair
(1759-1823)

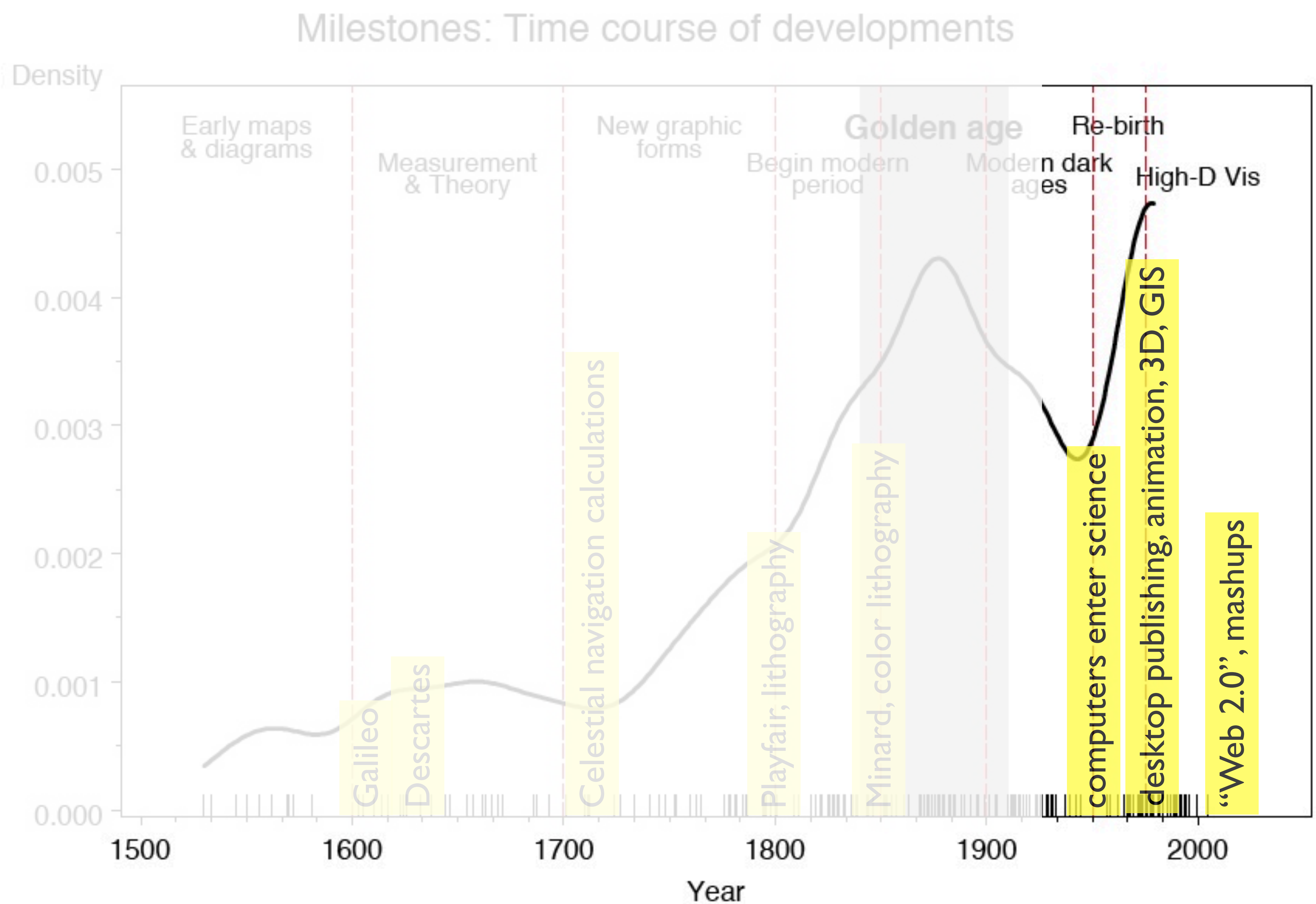


Minard, color lithography

Charles Joseph Minard, in color

(1781-1870)

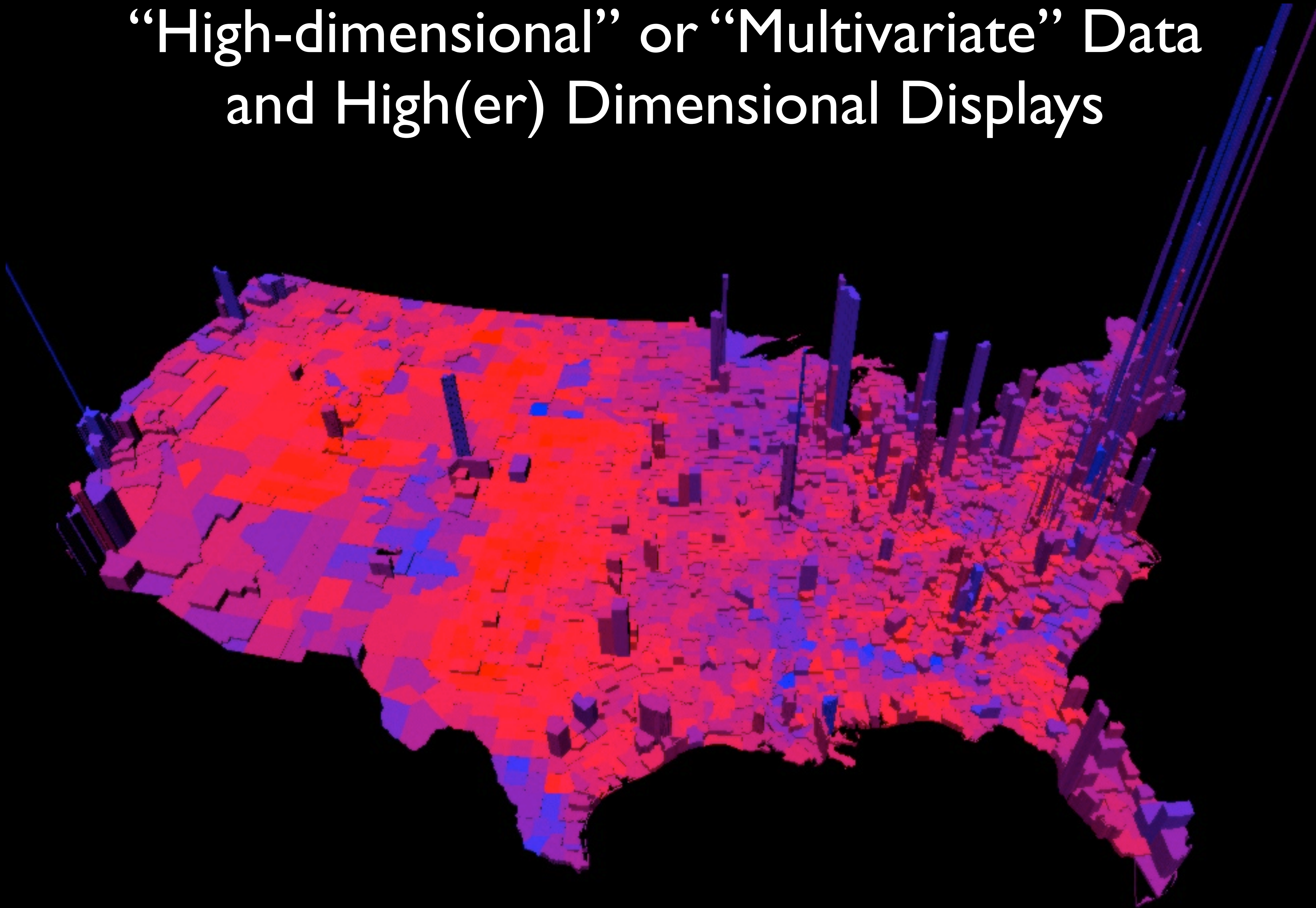




adapted from Friendly, "The Golden Age of Statistical Graphics," *Statistical Science*, in press (2008)

Data • Dimensions • Display

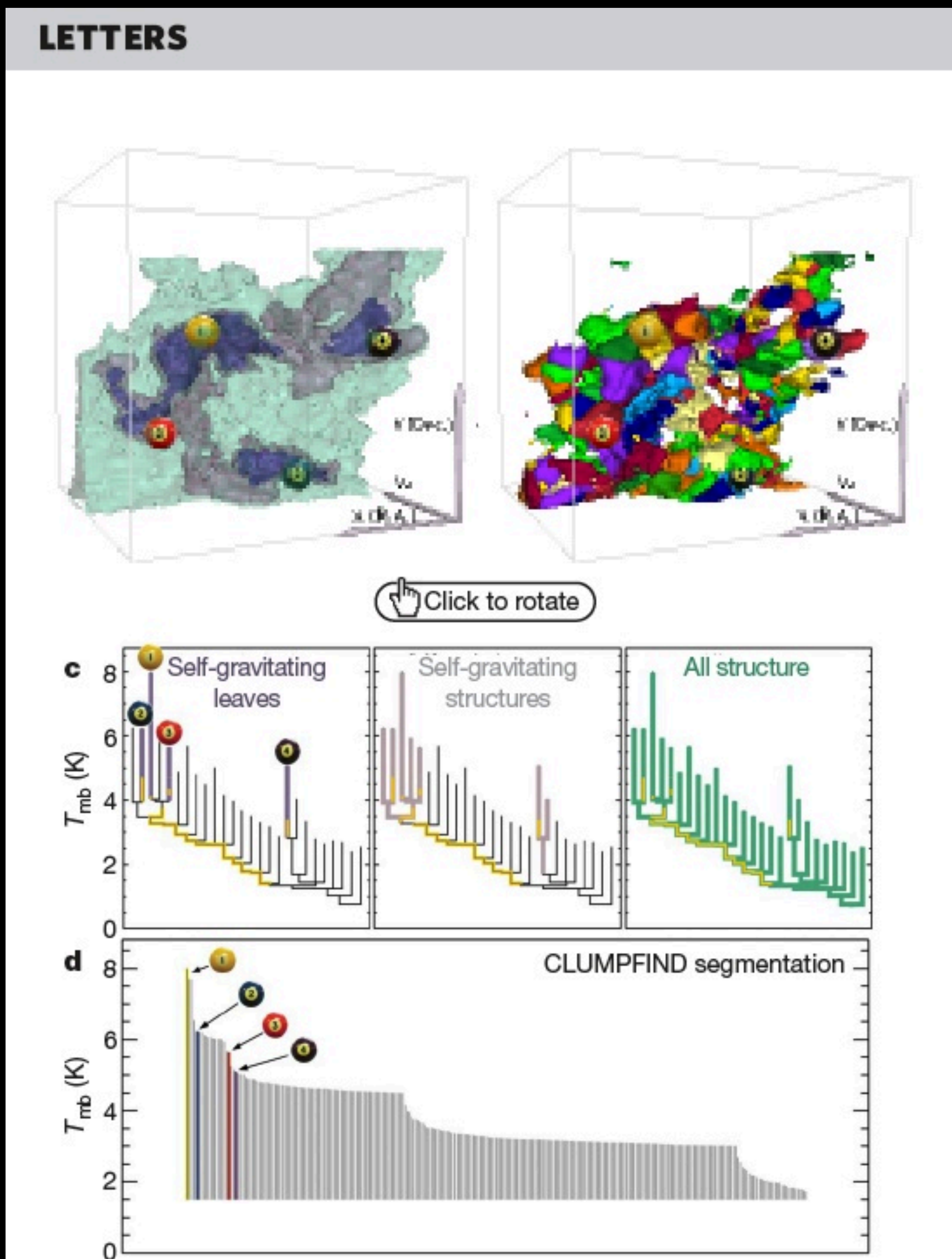
"High-dimensional" or "Multivariate" Data and High(er) Dimensional Displays



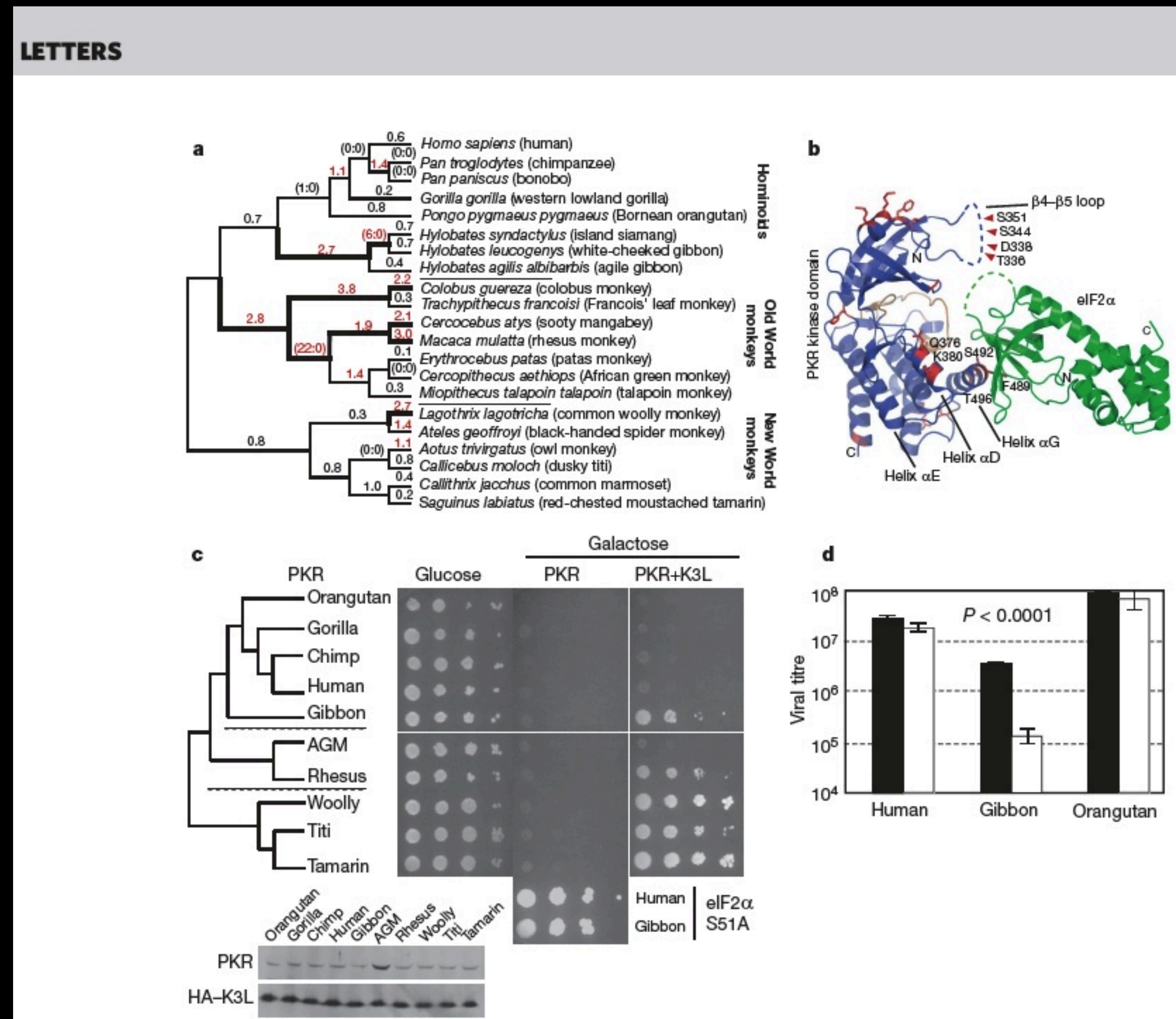
This map ***displays*** *2 quantities as a function of 2 spatial dimensions.*
...Is that 4 dimensions?

“High-dimensional” or “Multivariate” Data

(Astronomy=Biology)

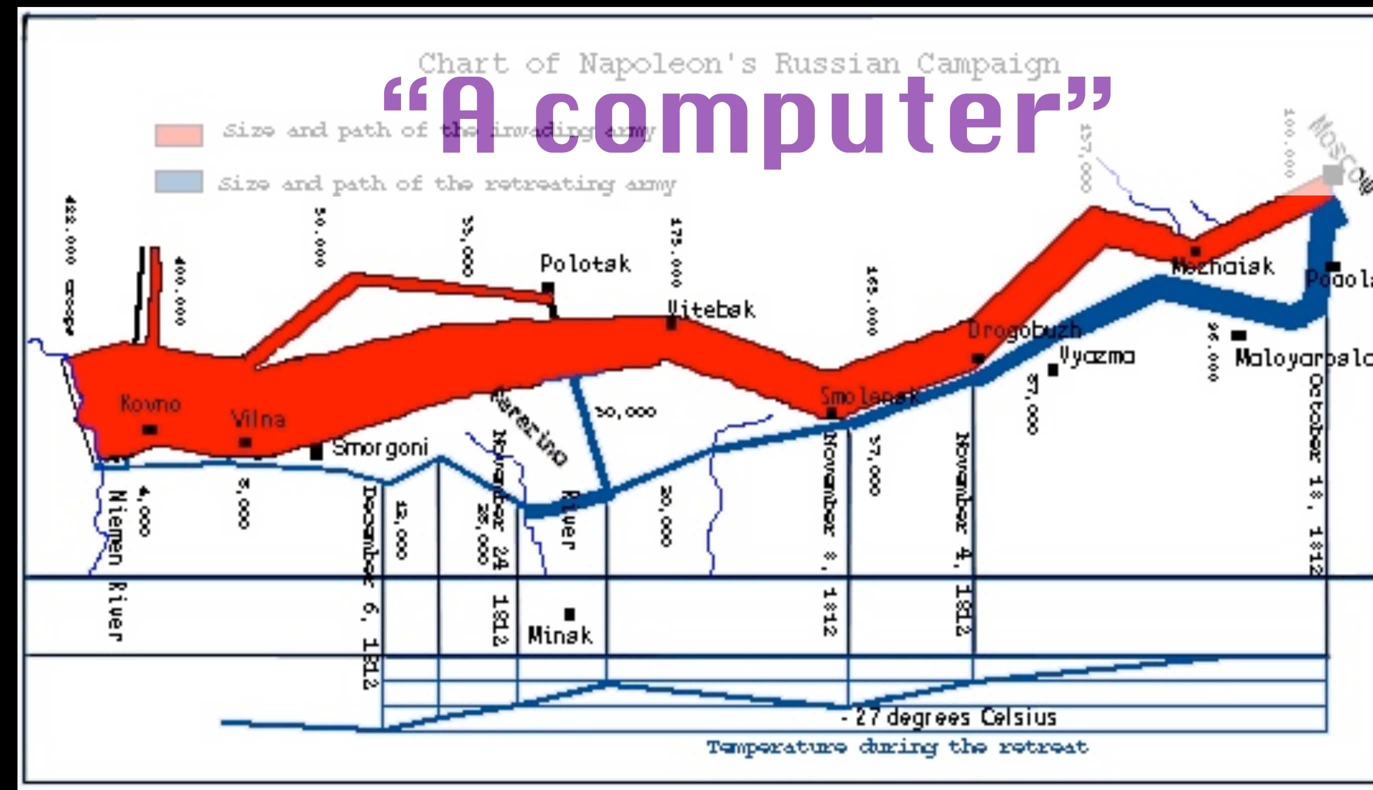


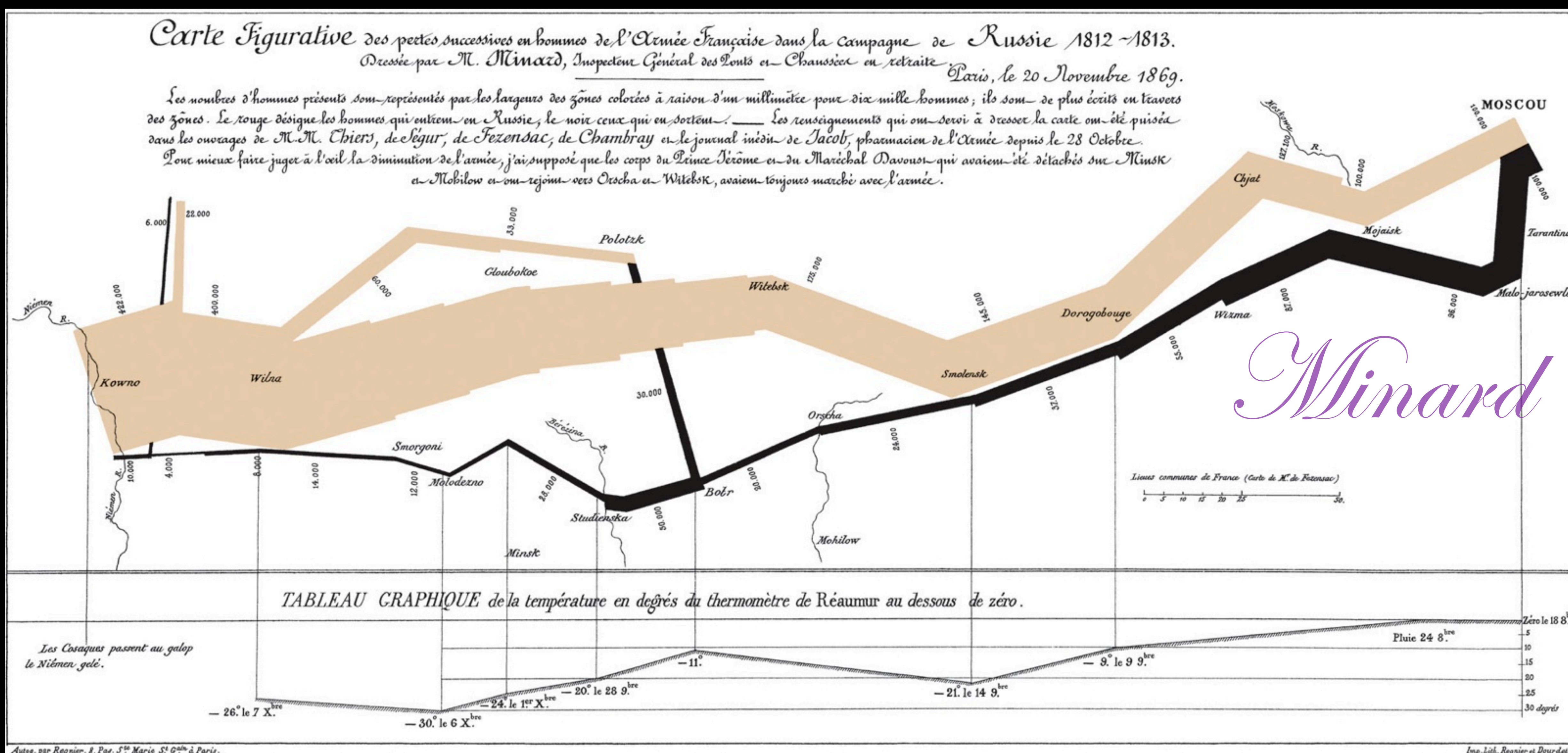
Goodman et al. *Nature*, 2009

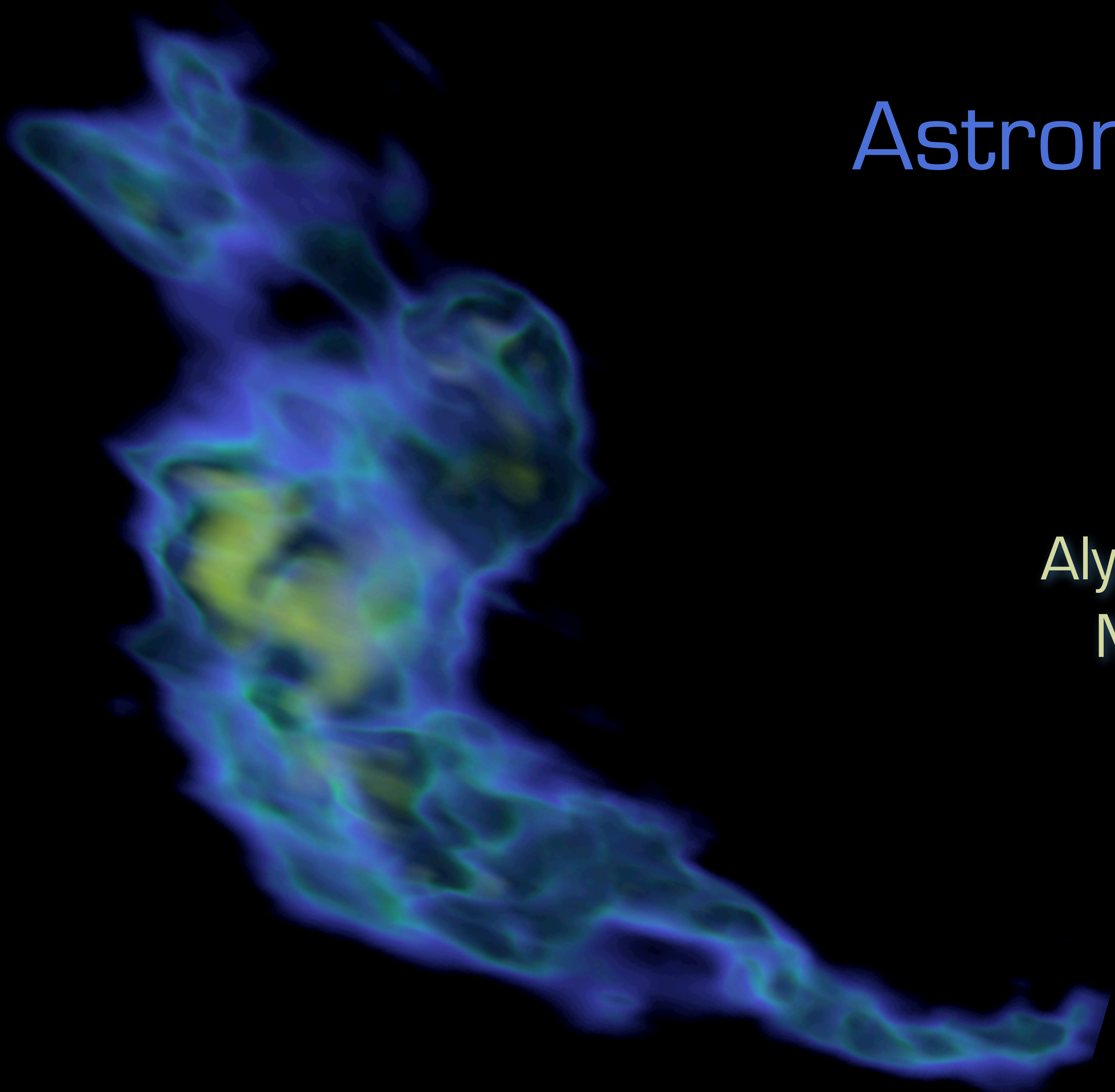


Elde et al. *Nature*, 2008

How much are we held back today by digital tools?





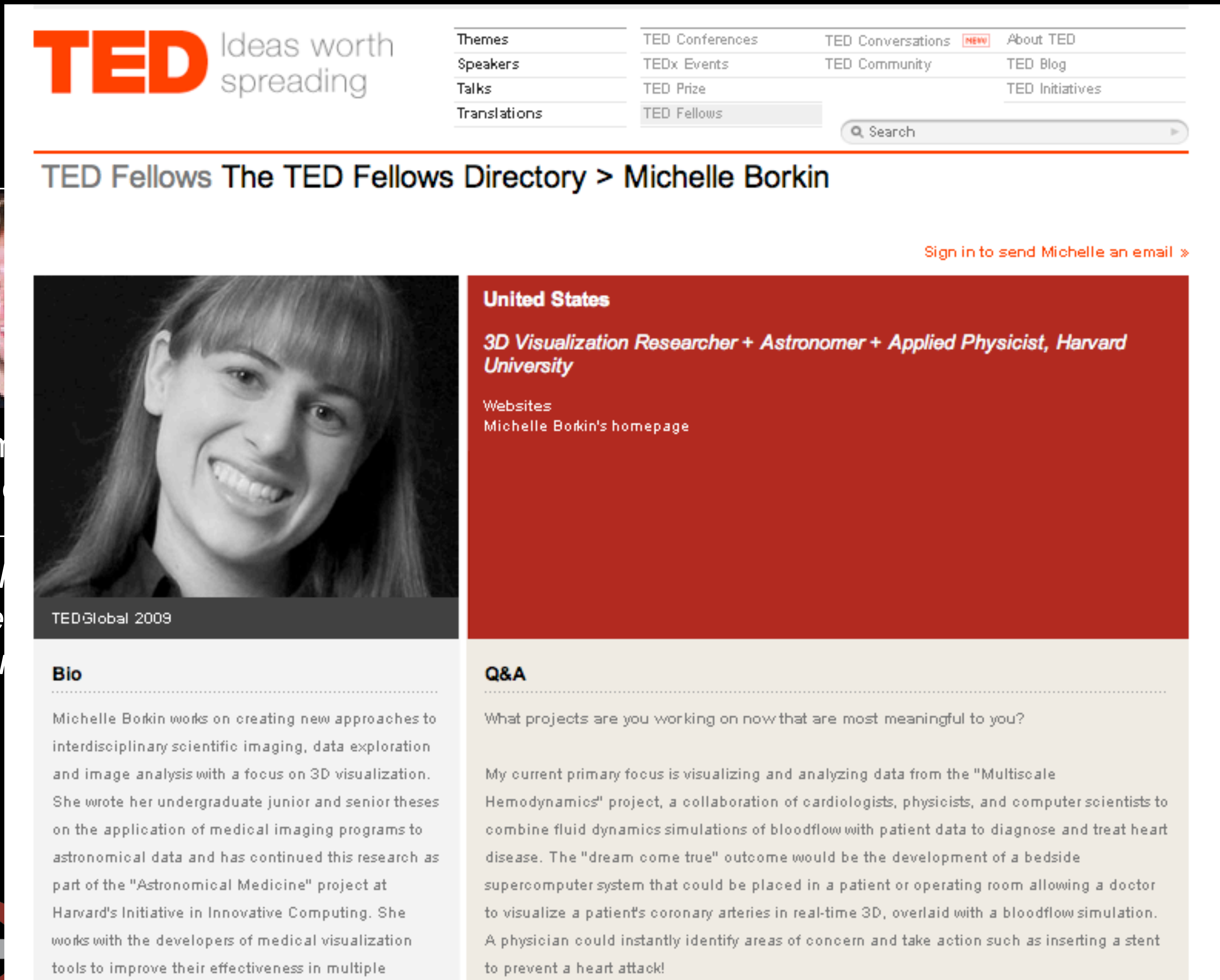


Astronomical Medicine

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



TED Ideas worth spreading

Themes | Speakers | Talks | Translations

TED Conferences | TEDx Events | TED Prize | TED Fellows

TED Conversations NEW | TED Community

About TED | TED Blog | TED Initiatives

Search

TED Fellows The TED Fellows Directory > Michelle Borkin

Sign in to send Michelle an email »

TEDGlobal 2009

United States

3D Visualization Researcher + Astronomer + Applied Physicist, Harvard University

Websites
Michelle Borkin's homepage

Bio

Michelle Borkin works on creating new approaches to interdisciplinary scientific imaging, data exploration and image analysis with a focus on 3D visualization. She wrote her undergraduate junior and senior theses on the application of medical imaging programs to astronomical data and has continued this research as part of the "Astronomical Medicine" project at Harvard's Initiative in Innovative Computing. She works with the developers of medical visualization tools to improve their effectiveness in multiple

Q&A

What projects are you working on now that are most meaningful to you?

My current primary focus is visualizing and analyzing data from the "Multiscale Hemodynamics" project, a collaboration of cardiologists, physicists, and computer scientists to combine fluid dynamics simulations of bloodflow with patient data to diagnose and treat heart disease. The "dream come true" outcome would be the development of a bedside supercomputer system that could be placed in a patient or operating room allowing a doctor to visualize a patient's coronary arteries in real-time 3D, overlaid with a bloodflow simulation. A physician could instantly identify areas of concern and take action such as inserting a stent to prevent a heart attack!

COMPLETE Perseus

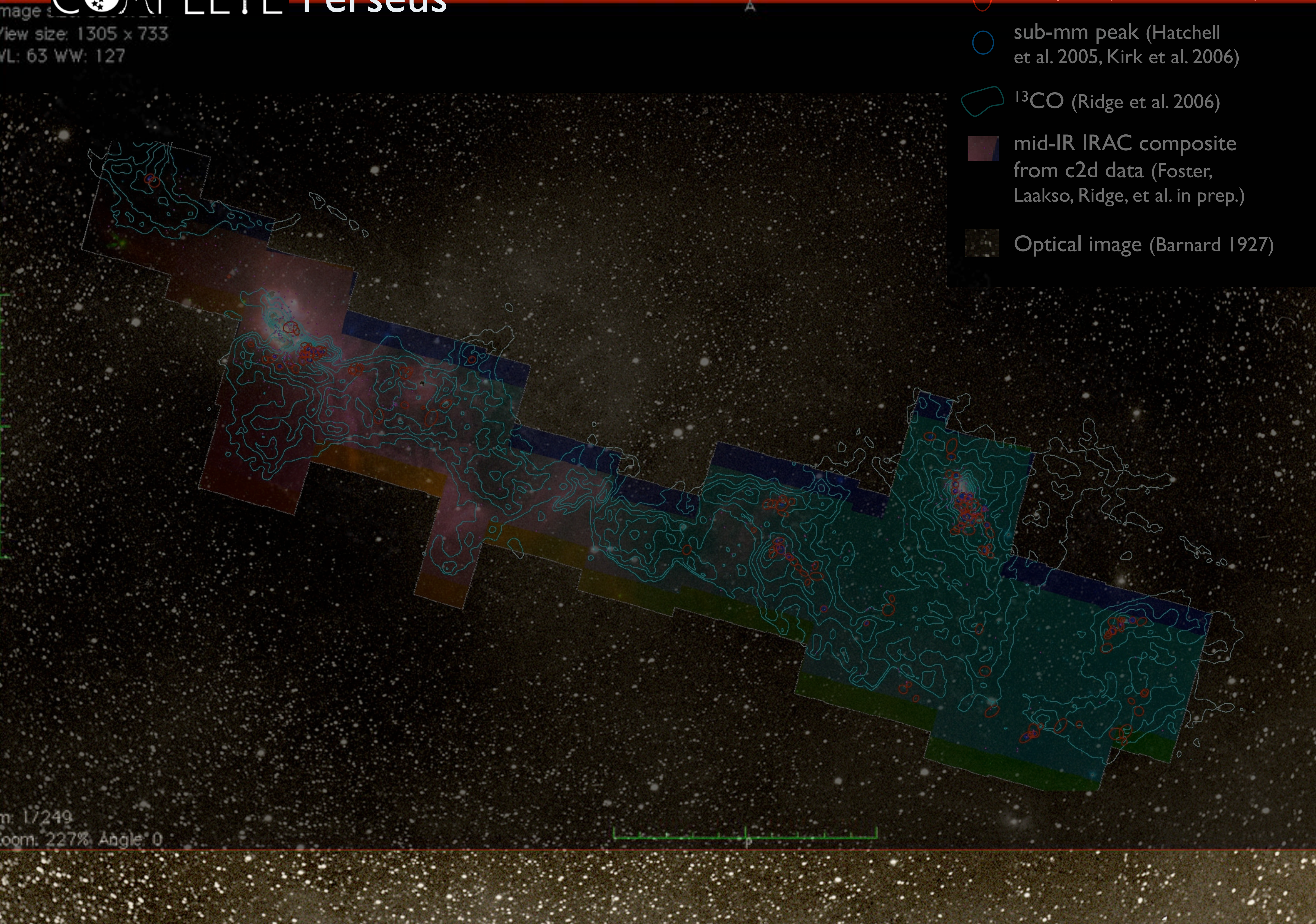
mm peak (Enoch et al. 2006)

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

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

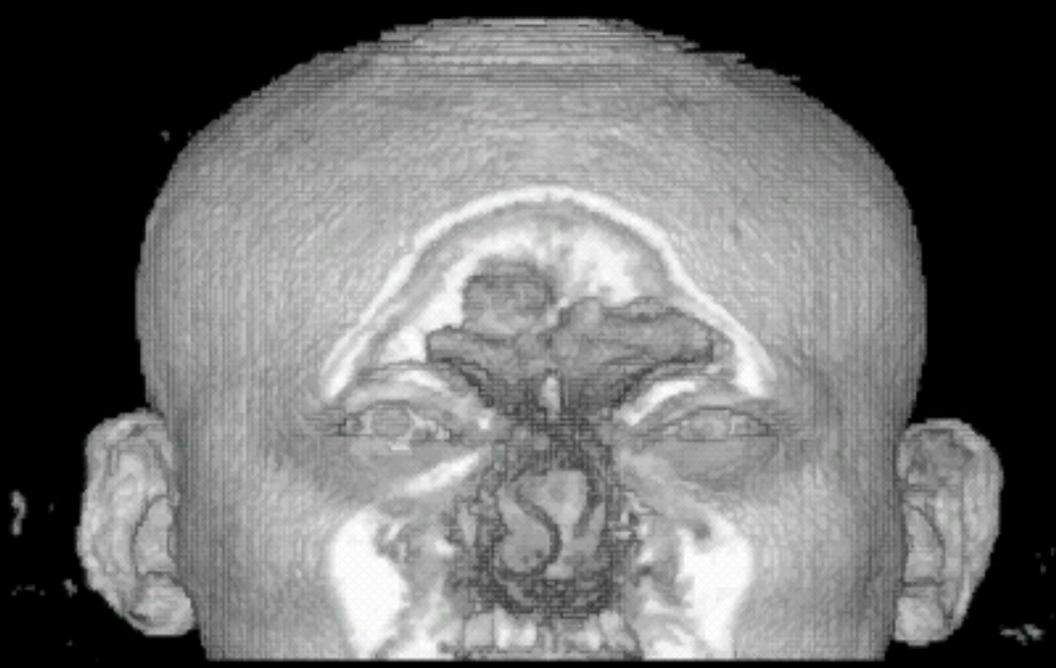
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al. in prep.)

Optical image (Barnard 1927)

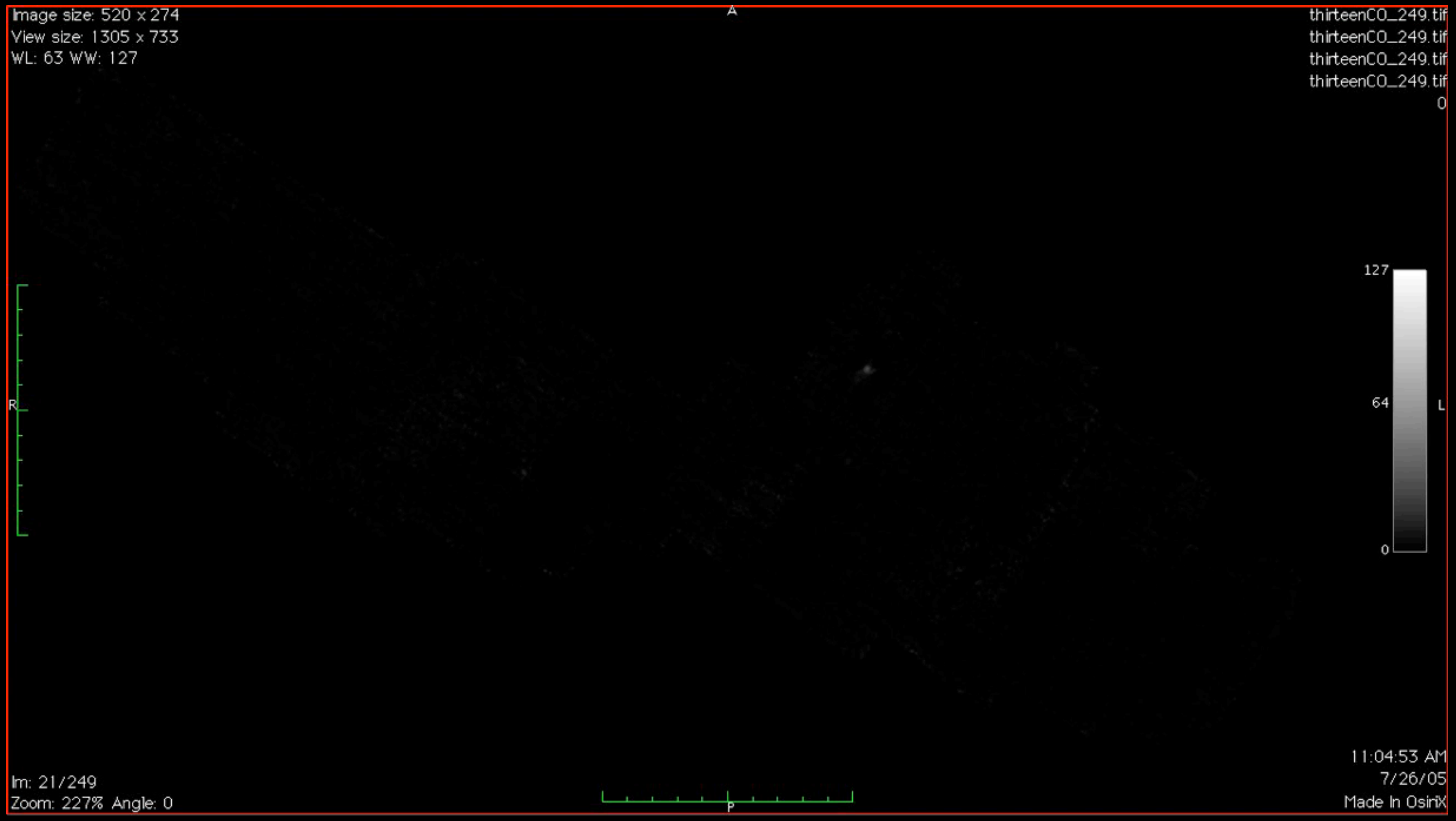


“Astronomical Medicine”

“KEITH”



“PERSEUS”

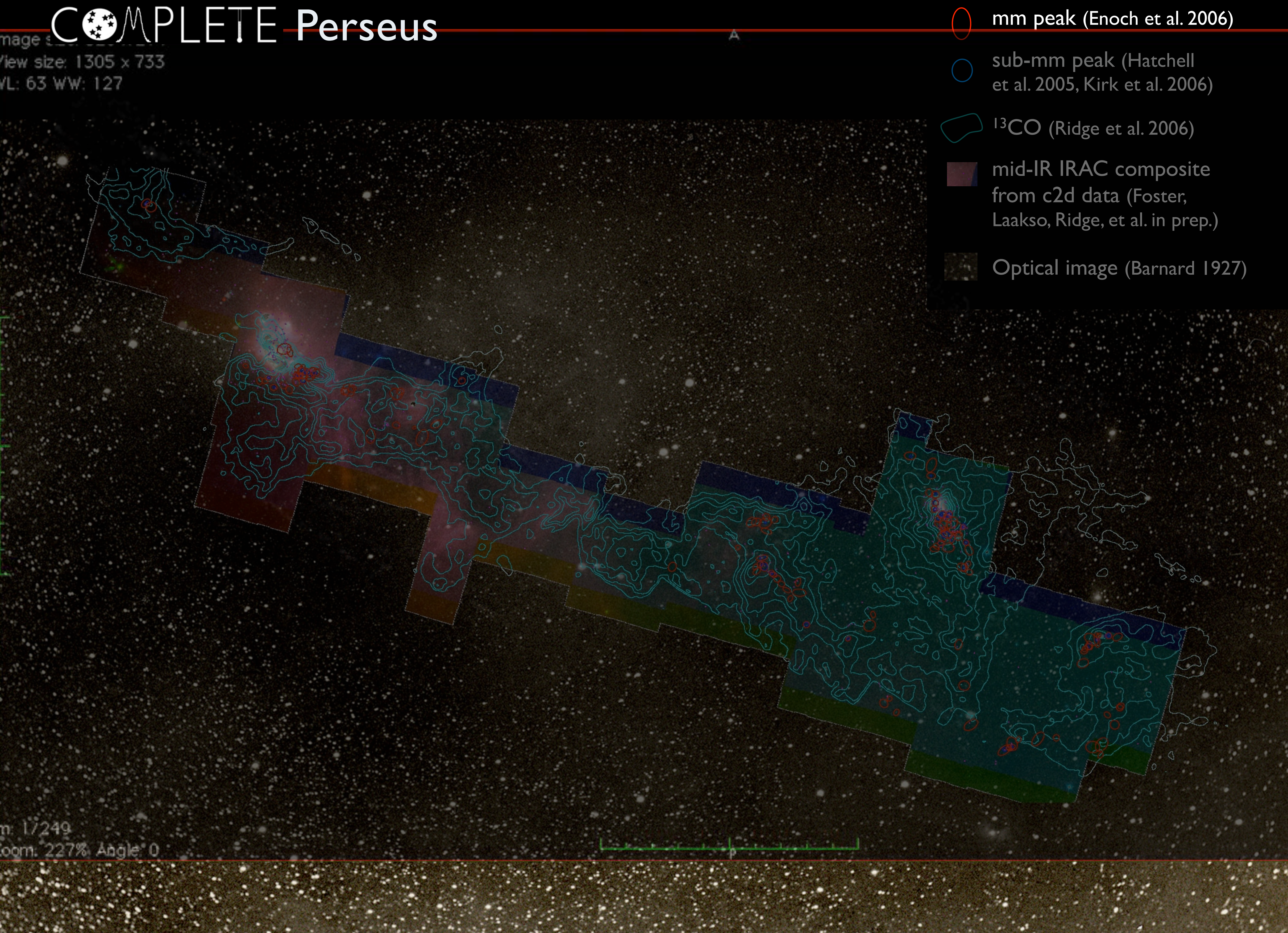


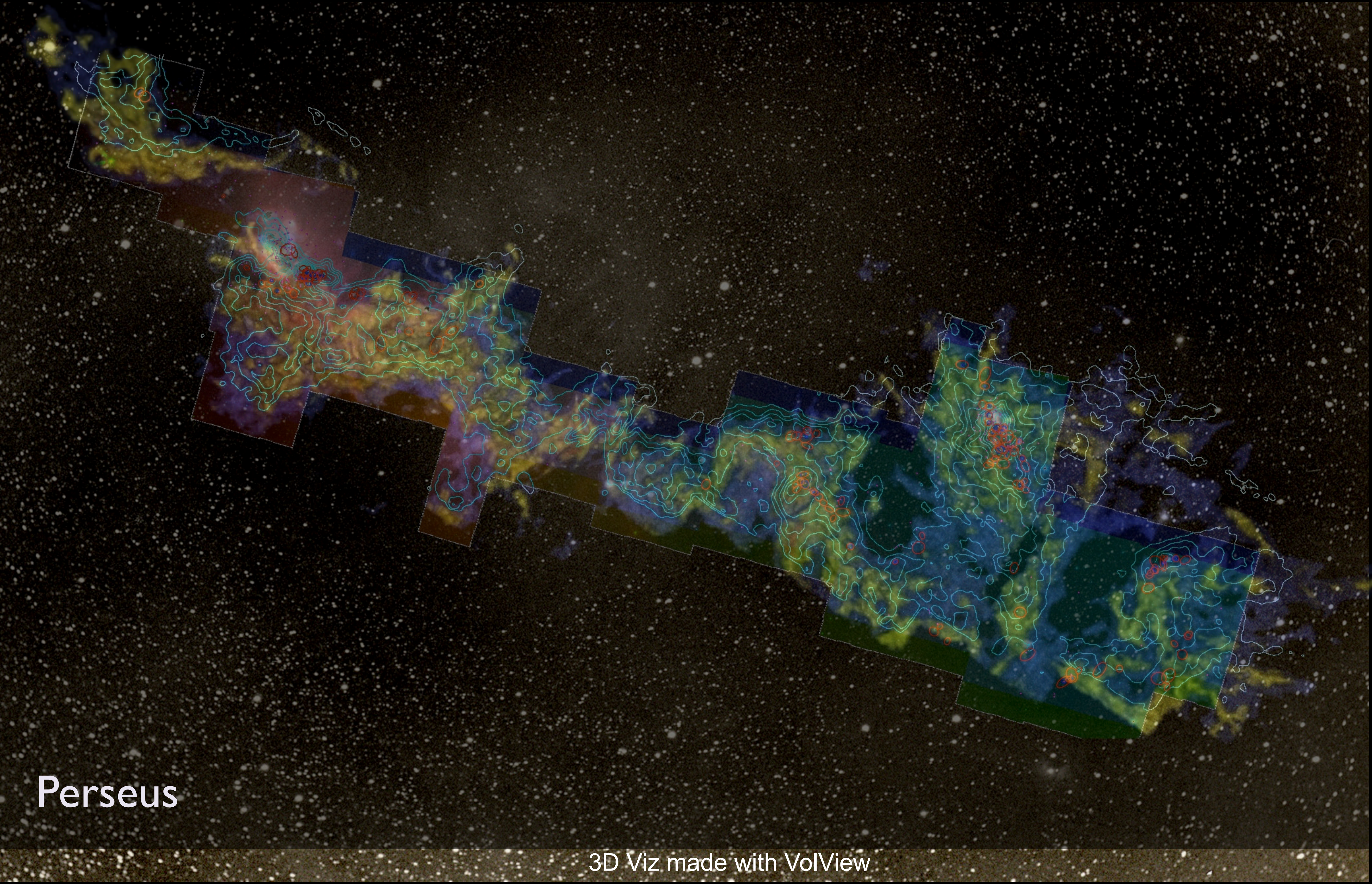
“z” is depth into head

“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




Perseus
3D Viz made with VolView

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: Modular Craftsmanship

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

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

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

live, interactive linked views (DataDesk, GapMinder, WWT, Dendroviz)

Unsolved Questions...

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

improved graphical representation of uncertainty display modes

3D PDF

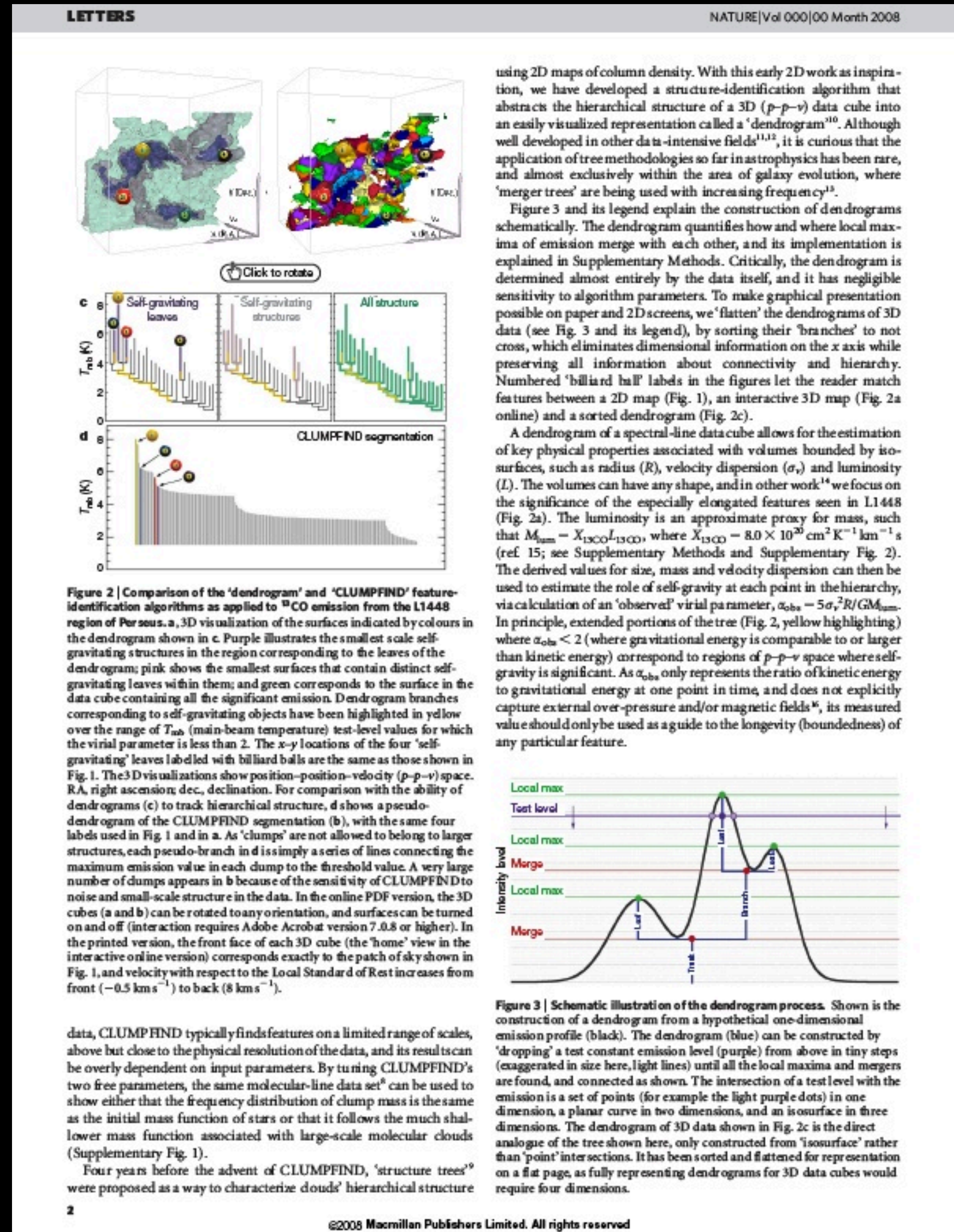


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating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

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

IIC Member login

iic

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

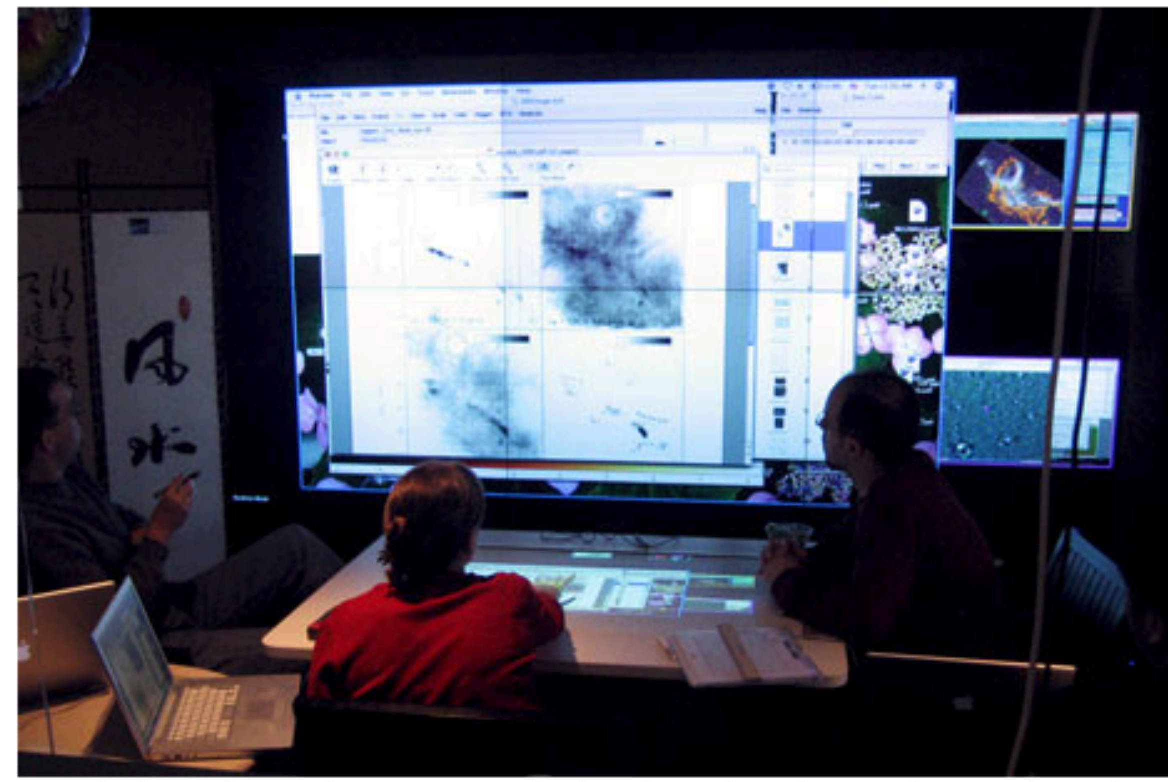


PHOTO: HAO JIANG, DANIEL WIGDOR, CLIFTON FORLINES, AND CHIA SHEN

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

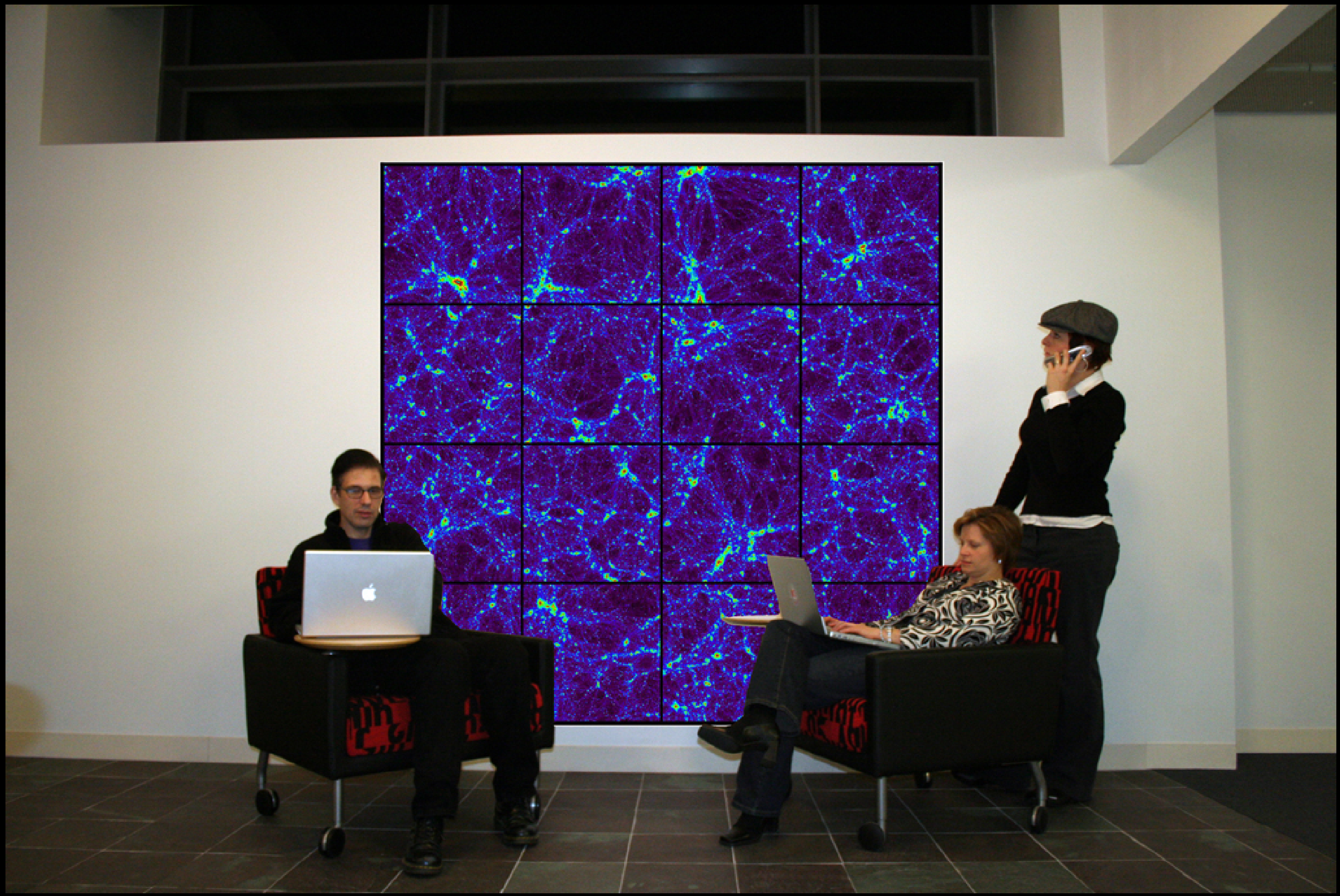
Touch Interfaces

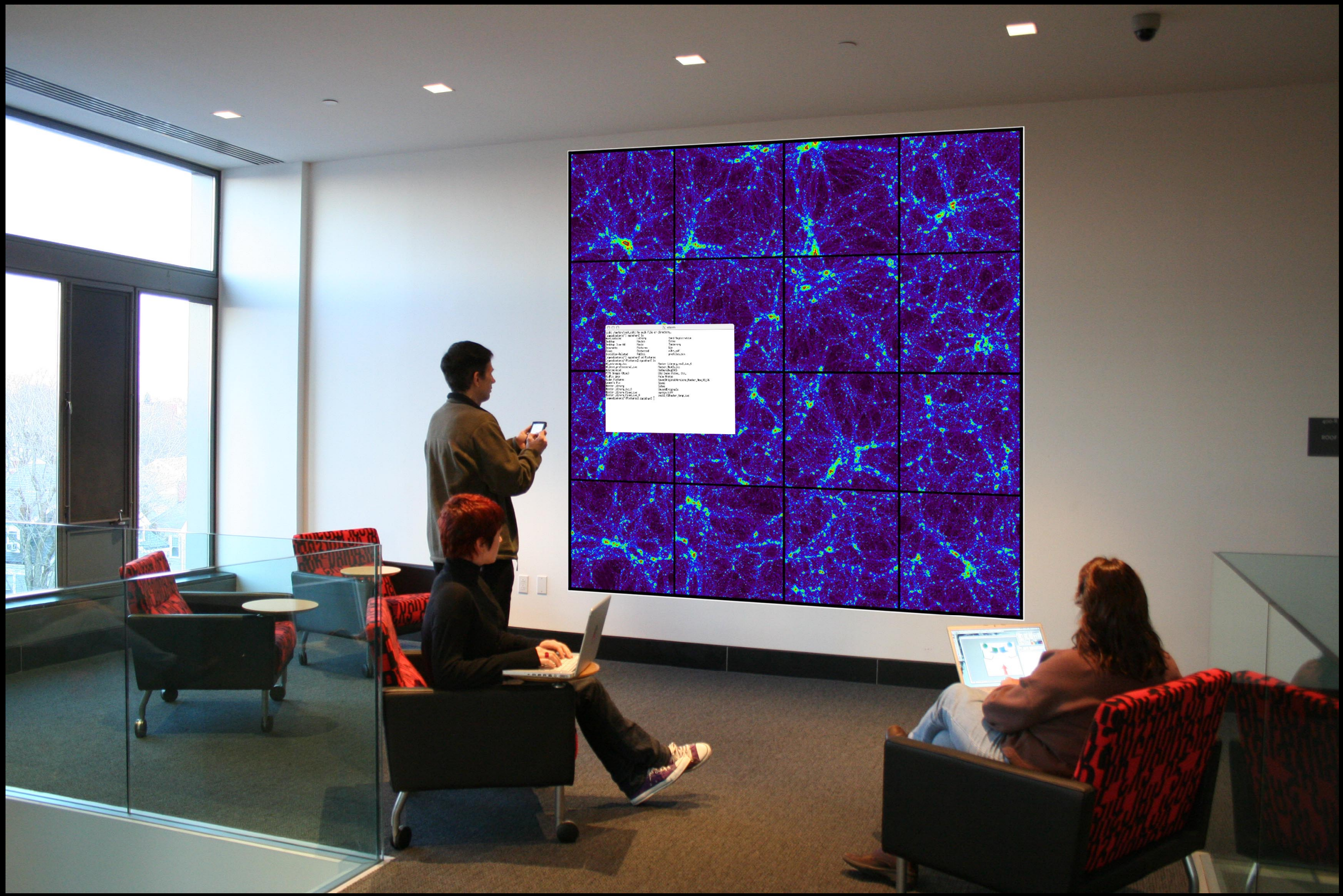
movie courtesy Daniel Wigdor, taken at MERL, Kendall Square, Cambridge

display modes

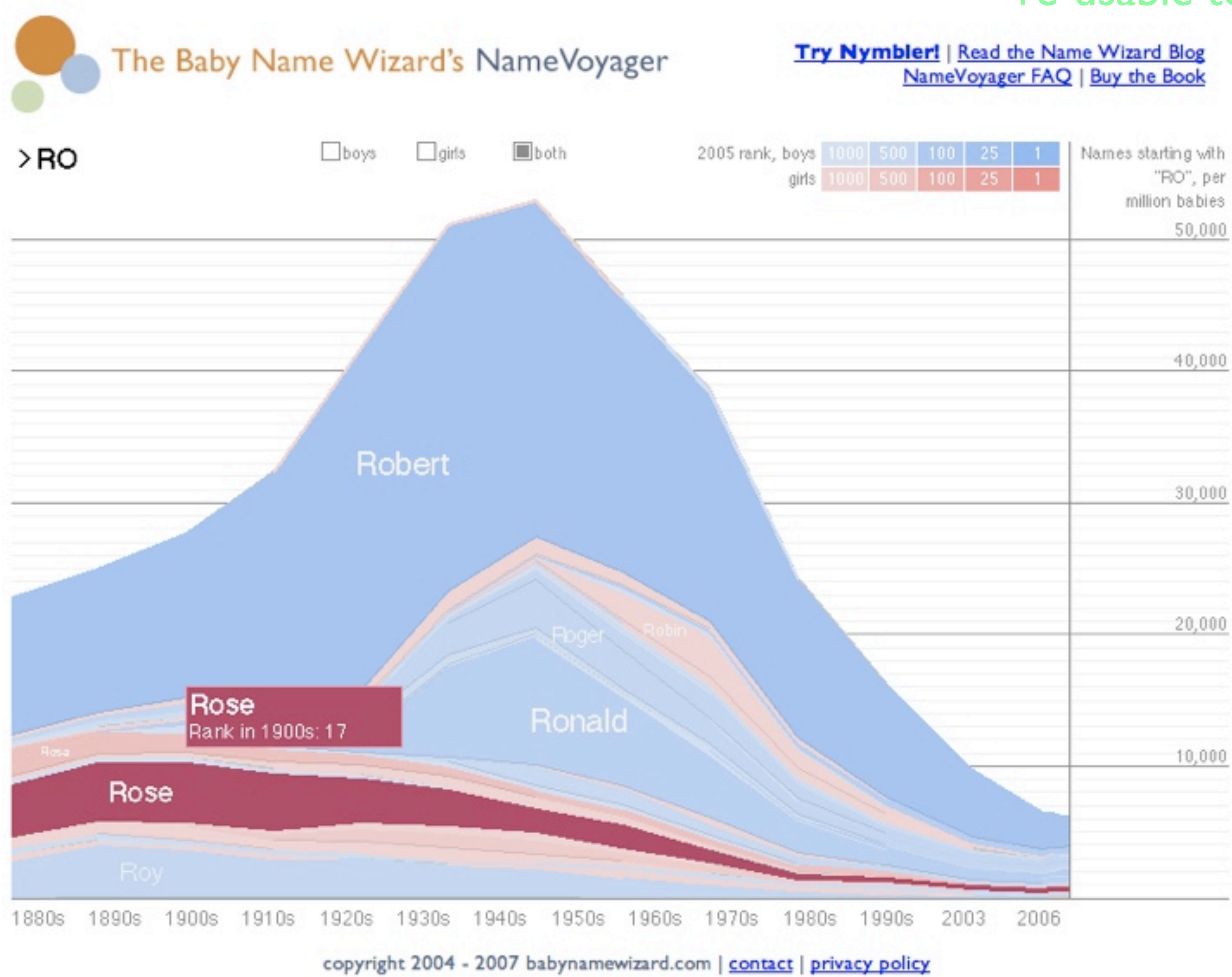
...why we must explain that...

"This is not art."





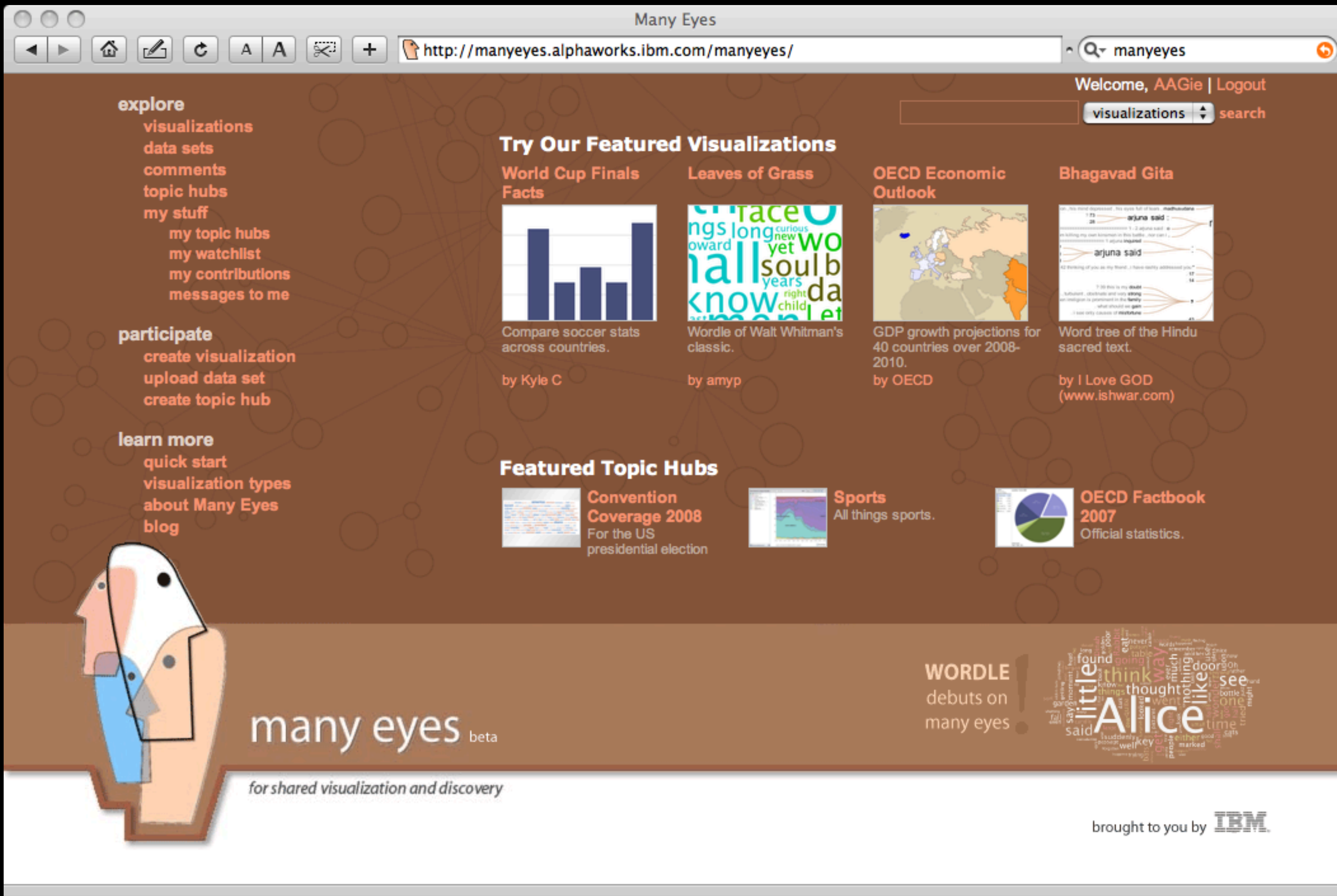
re-usable tools



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

Many Eyes:

Martin Wattenberg & Fernanda Viegas (now Google, formerly IBM)



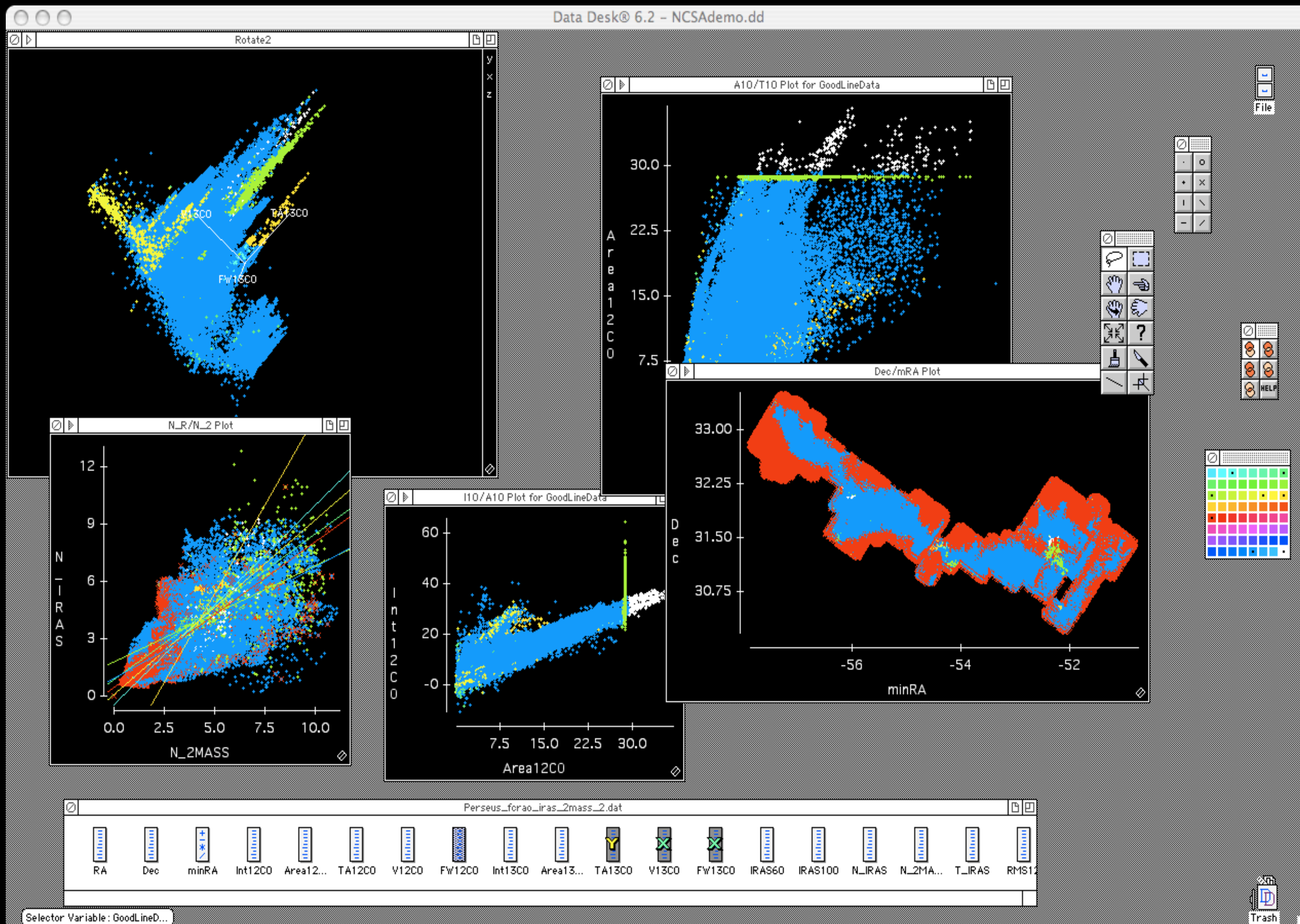
re-usable tools + crowdsourcing

Machine Learning

What riles up the ISM?

live linked views

Data Desk, c. 1986(!)



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

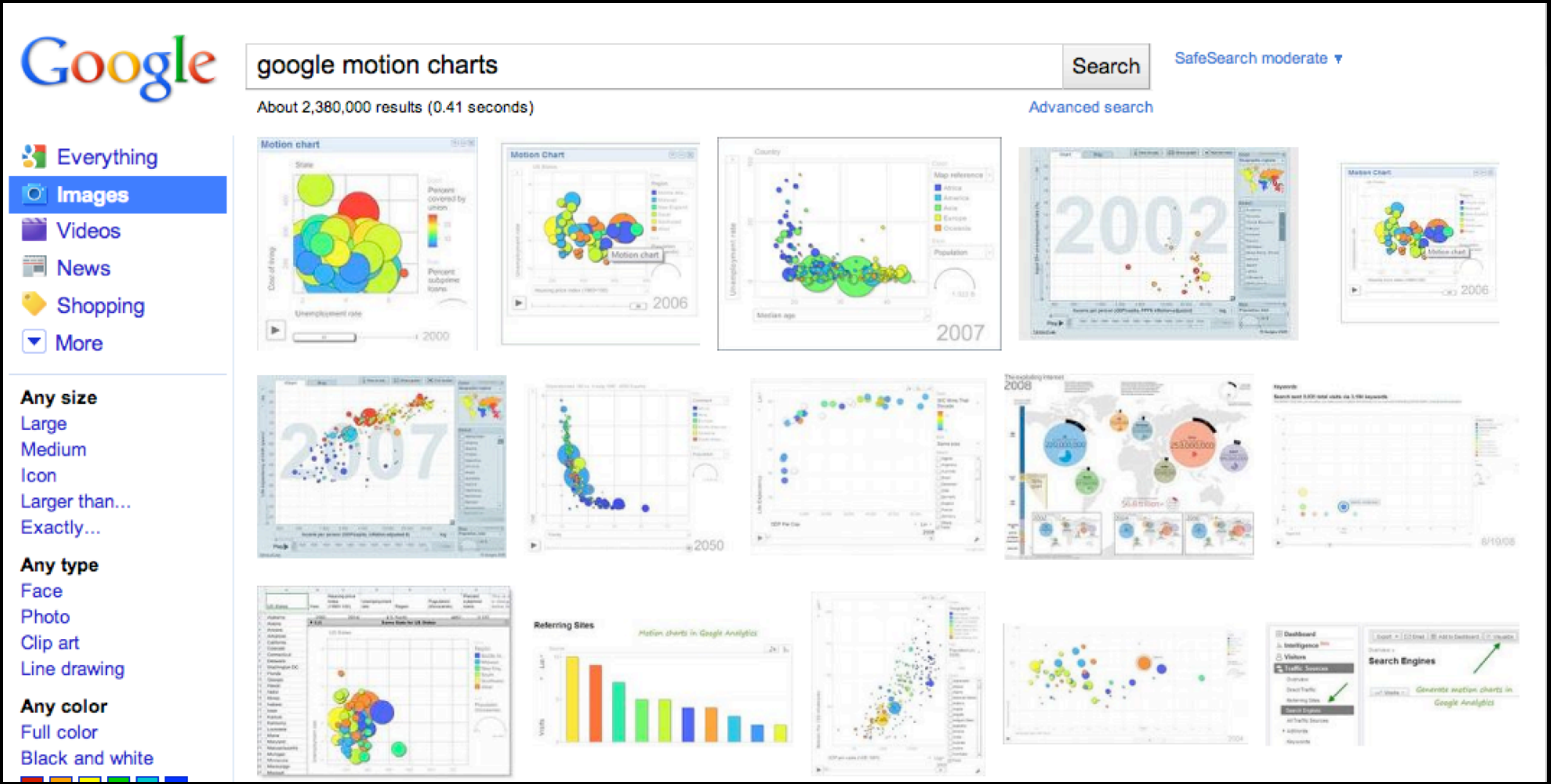
re-usable, live linked views

Gapminder (today)

http://gapminder.org

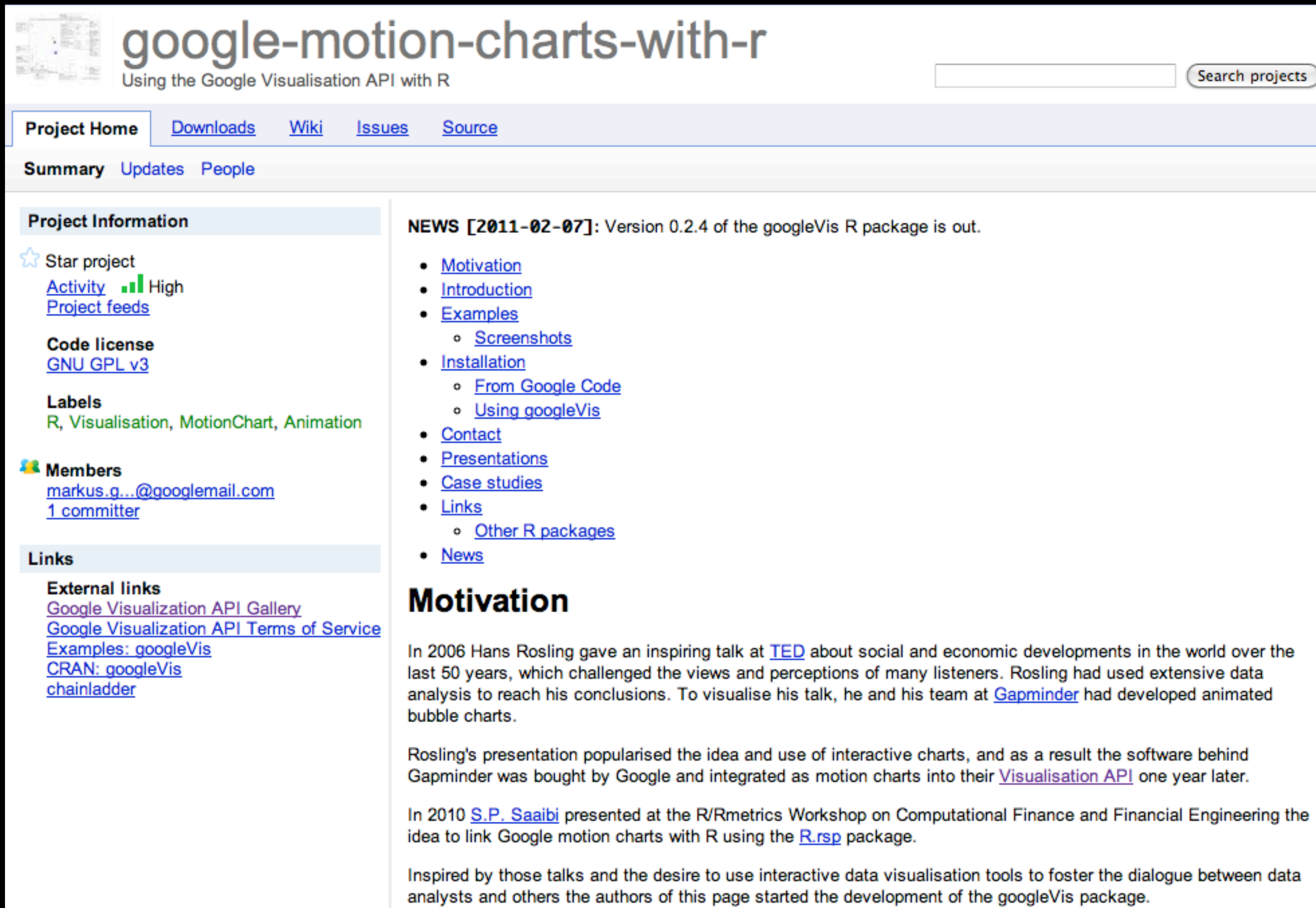
re-usable, live linked views

Gapminder for All = Google Motion Charts



re-usable, live linked views

“Modular Craftsmanship”=
Google Motion Charts with R



google-motion-charts-with-r

Using the Google Visualisation API with R

Search projects

Project Home | Downloads | Wiki | Issues | Source

Summary | Updates | People

Project Information

Star project

Activity High

Project feeds

Code license

GNU GPL v3

Labels

R, Visualisation, MotionChart, Animation

Members

markus.g...@googlemail.com

1 committer

Links

External links

Google Visualization API Gallery

Google Visualization API Terms of Service

Examples: googleVis

CRAN: googleVis

chainladder

NEWS [2011-02-07]: Version 0.2.4 of the googleVis R package is out.

- Motivation
- Introduction
- Examples
  - Screenshots
- Installation
  - From Google Code
  - Using googleVis
- Contact
- Presentations
- Case studies
- Links
  - Other R packages
- News

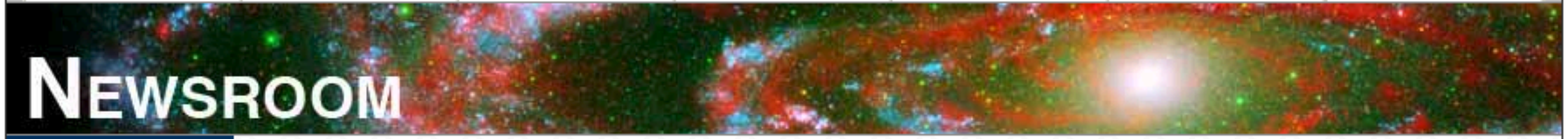
Motivation

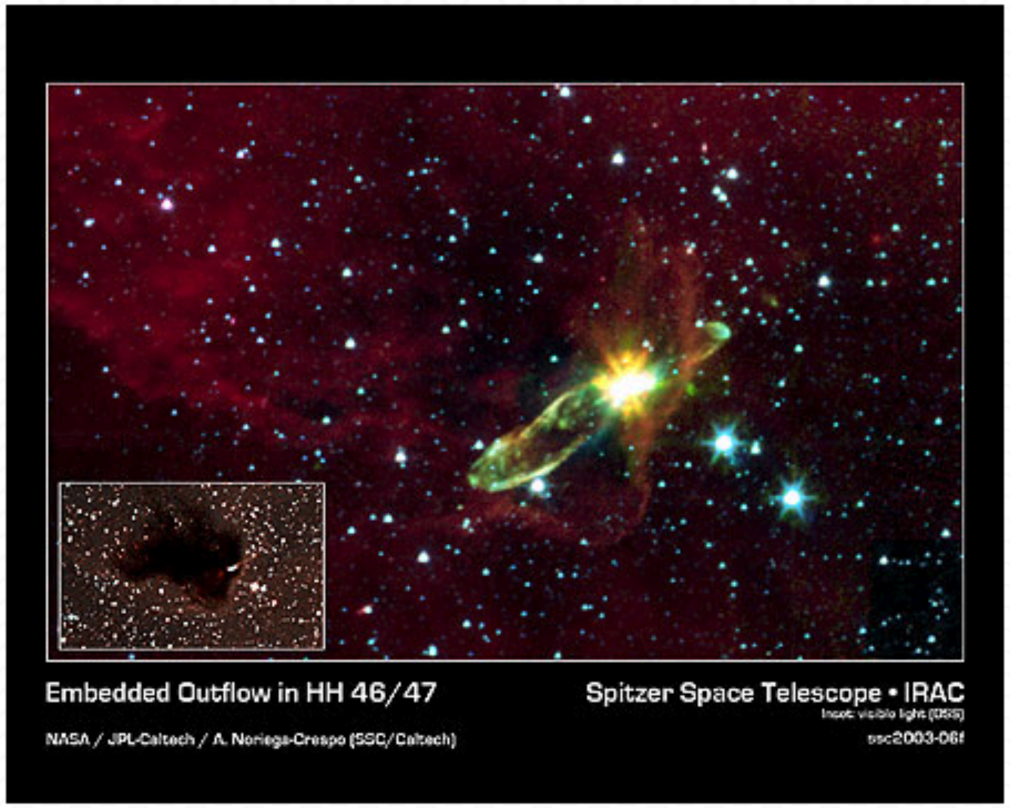
In 2006 Hans Rosling gave an inspiring talk at TED about social and economic developments in the world over the last 50 years, which challenged the views and perceptions of many listeners. Rosling had used extensive data analysis to reach his conclusions. To visualise his talk, he and his team at Gapminder had developed animated bubble charts.

Rosling's presentation popularised the idea and use of interactive charts, and as a result the software behind Gapminder was bought by Google and integrated as motion charts into their Visualisation API one year later.

In 2010 S.P. Saaibi presented at the R/Rmetrics Workshop on Computational Finance and Financial Engineering the idea to link Google motion charts with R using the R.rsp package.

Inspired by those talks and the desire to use interactive data visualisation tools to foster the dialogue between data analysts and others the authors of this page started the development of the googleVis R package.





Credit: NASA/JPL-Caltech/A. Noriega-Crespo (SSC/Caltech), Digital Sky Survey

HH46/47

This image from NASA's Spitzer Space Telescope transforms a dark cloud into a silky translucent veil, revealing the molecular outflow from an otherwise hidden newborn star. Using near-infrared light, Spitzer pierces through the dark cloud to detect the embedded outflow in an object called HH 46/47. Herbig-Haro (HH) objects are bright, nebulous regions of gas and dust that are usually buried within dark clouds. They are formed when supersonic gas ejected from a forming protostar, or embryonic star, interacts with the surrounding interstellar medium. These young stars are often detected only in the infrared.

The Spitzer image was obtained with the infrared array camera. Emission at 3.6 microns is shown as blue, emission from 4.5 and 5.8 microns has been combined as green, and 8.0 micron emission is depicted as red.

HH 46/47 is a striking example of a low mass protostar ejecting a jet and creating a bipolar, or two-sided, outflow. The central http://www.flickr.com/photos/66496709@N00/3175344278/

flickr

Signed in as Alyssa_Goodman (1 new) Help Sign Out

Home You Organize Contacts Groups Explore Search

HH4647

ADD NOTE | SEND TO GROUP | ADD TO SET | BLOG THIS | ALL SIZES | ORDER PRINTS | ROTATE | EDIT PHOTO | DELETE

Share This

Uploaded on January 6, 2009
by Alyssa_Goodman

Alyssa_Goodman's photostream


16 uploads

browse

This photo also belongs to:

astrometry (Pool)

Tags

Astrometrydotnet:version=10145
Astrometrydotnet:id=alpha-200901-20629873
Astrometrydotnet:status=solved

Add a tag

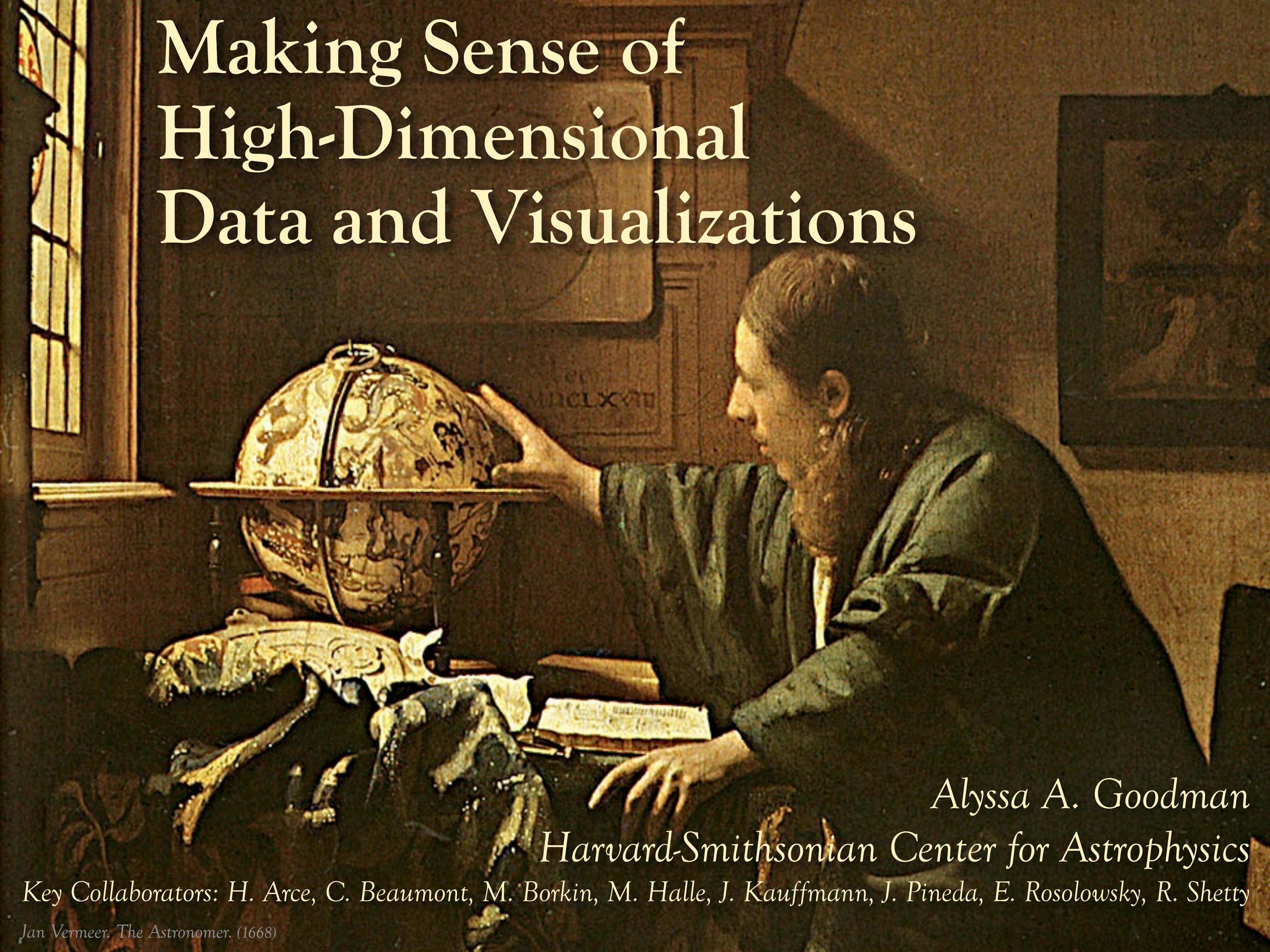
Additional Information

Anyone can see this photo (edit)

- Add to your map
- Taken on December 12, 2003 (edit)
- Photo stats
- Viewed 7 times (Not including you)
- Edit title, description, and tags

Flag your photo

Making Sense of High-Dimensional Data and Visualizations



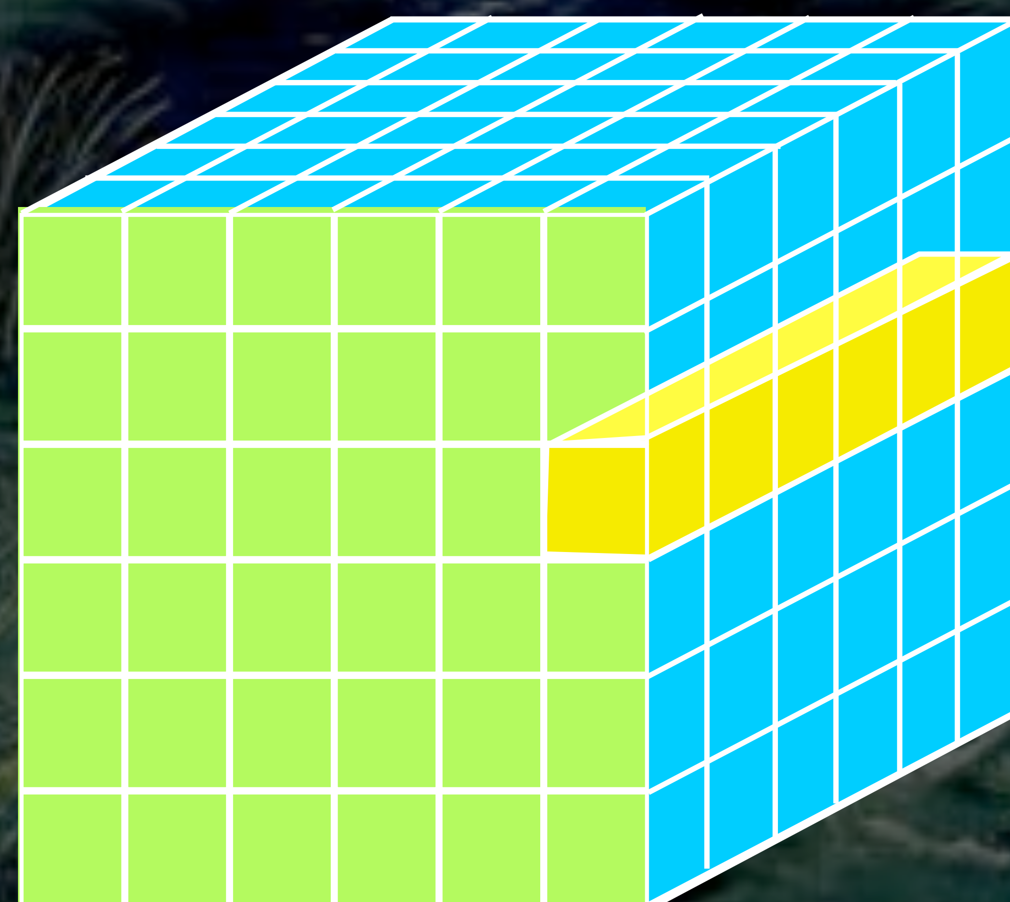
Alyssa A. Goodman

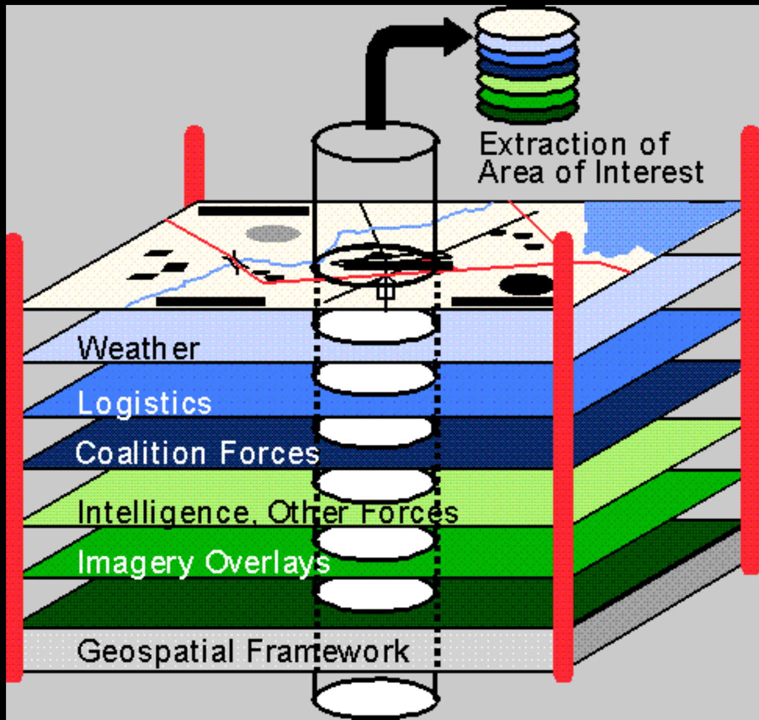
Harvard-Smithsonian Center for Astrophysics

Key Collaborators: H. Arce, C. Beaumont, M. Borkin, M. Halle, J. Kauffmann, J. Pineda, E. Rosolowsky, R. Shetty

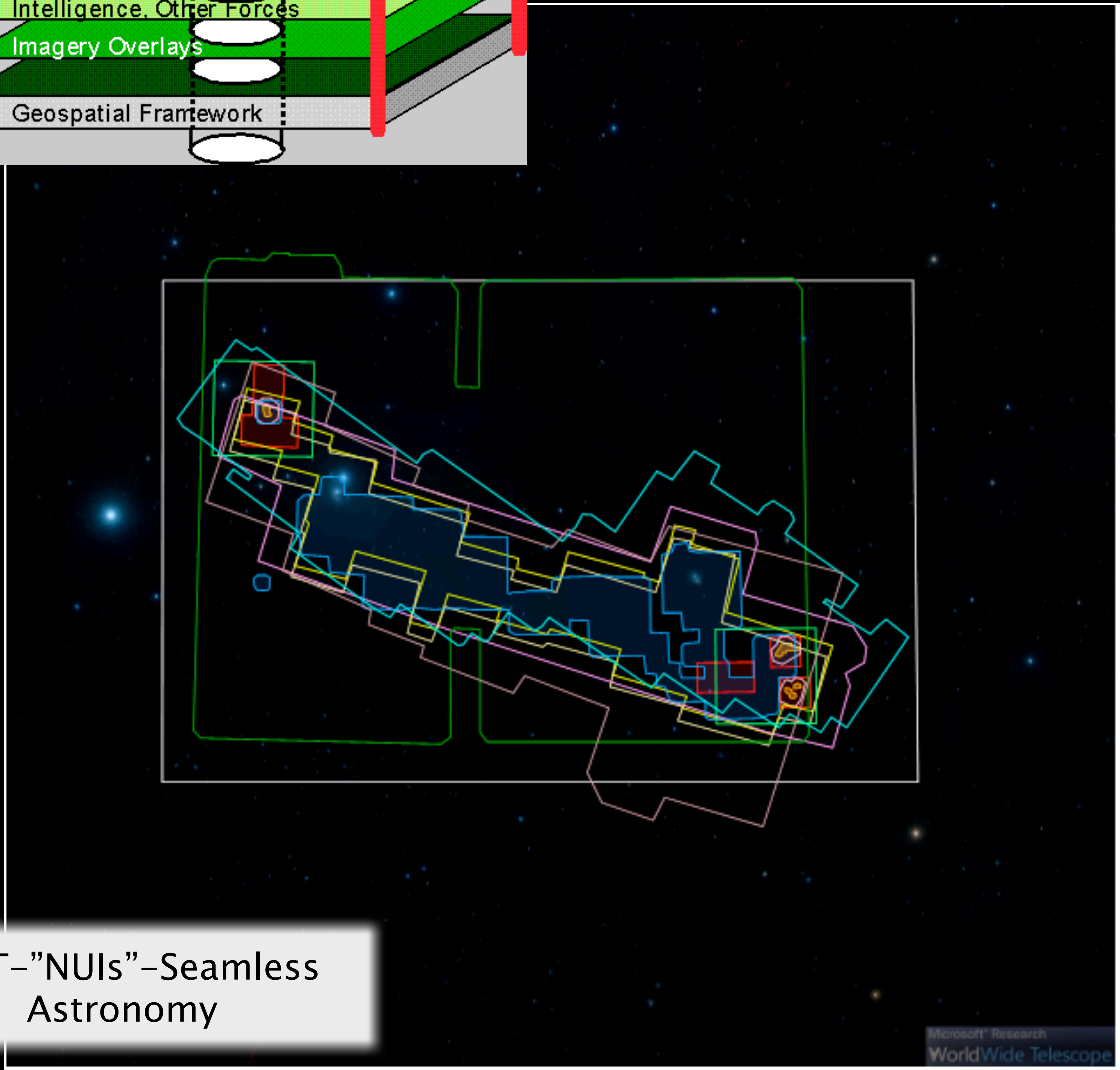
Jan Vermeer. The Astronomer. (1668)

The dream scenario...





COMPLETE



WWT-"NUIs"-Seamless Astronomy

COMPLETE Data Available

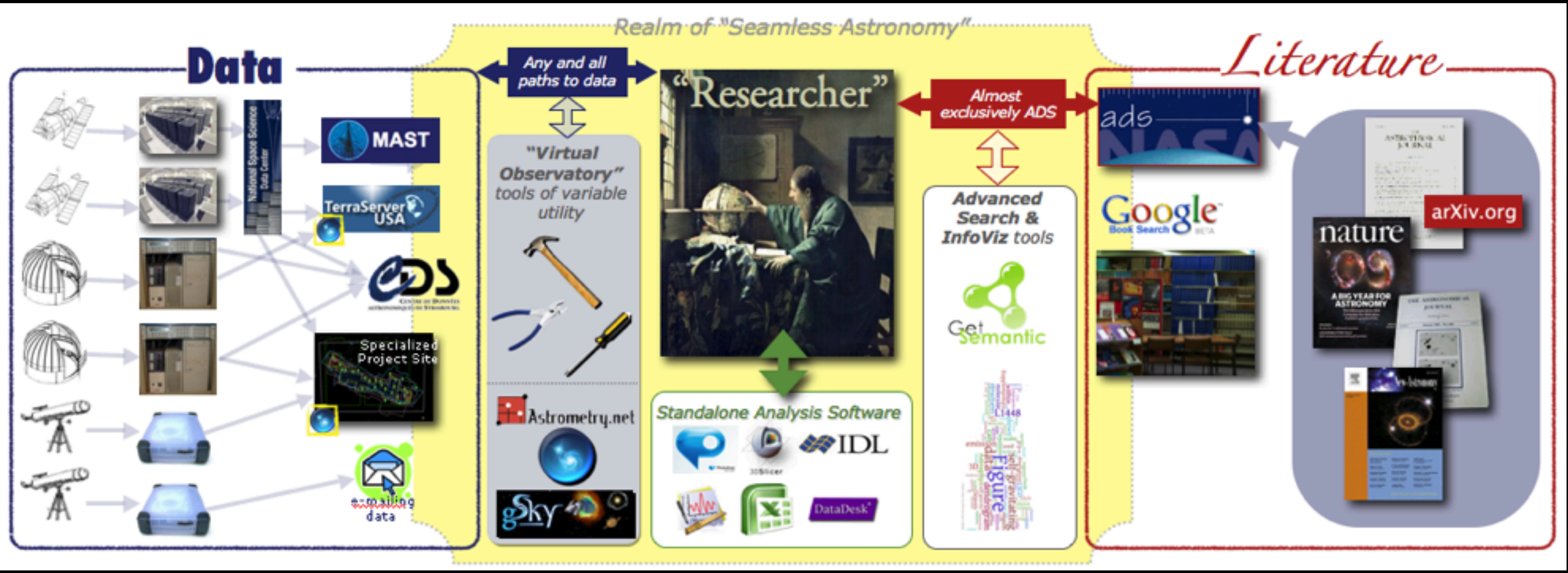
Center on Perseus | Center on Ophichus | Center on Serpens

Full-Cloud Data (Phase I, All Data Available)

Dataset	Show	Perseus	Ophiuchus	Serpens	Link
GBT: HI Data Cube	☑	√	√	∅	Data
IRAS: Av/Temp Maps	☑	√	√	√	Data
FCRAO: 12CO	☑	√	√	√	Data
FCRAO: 13CO		√	√	√	Data
JCMT: 850 microns	☑	√	√	∅	Data
Spitzer c2d: IRAC 1,3 (3.6,5.8 μm)	☑	√	√	√	Data
Spitzer c2d: IRAC 2,4 (4.5,8 μm)	☑	√	√	√	Data
CSO/Bolocam: 1.2-mm	☑	√	∅	∅	Data
Spitzer MIPS: Derived Dust Map	☑	√	∅	∅	Data
Targeted Regions (Phase II, Some Data Not Yet Available)					
CTIO/Calar Alto: NIR (J,H,Ks)	☑	√	√	∅	Data
IRAM 30-m: N2H+ and C18O	☑	√	∅	∅	Data
IRAM 30-m: 1.1-mm continuum	☑	√	∅	∅	Data
Megacam/MMT: r,i,z images	☑	√	∅	∅	Data
Catalogs & Pointed Surveys					
NH3 Pointed Survey	☐	√	∅	∅	Data
YSO Candidate list (c2d)	☐	√	√	√	Data

Seamless Astronomy

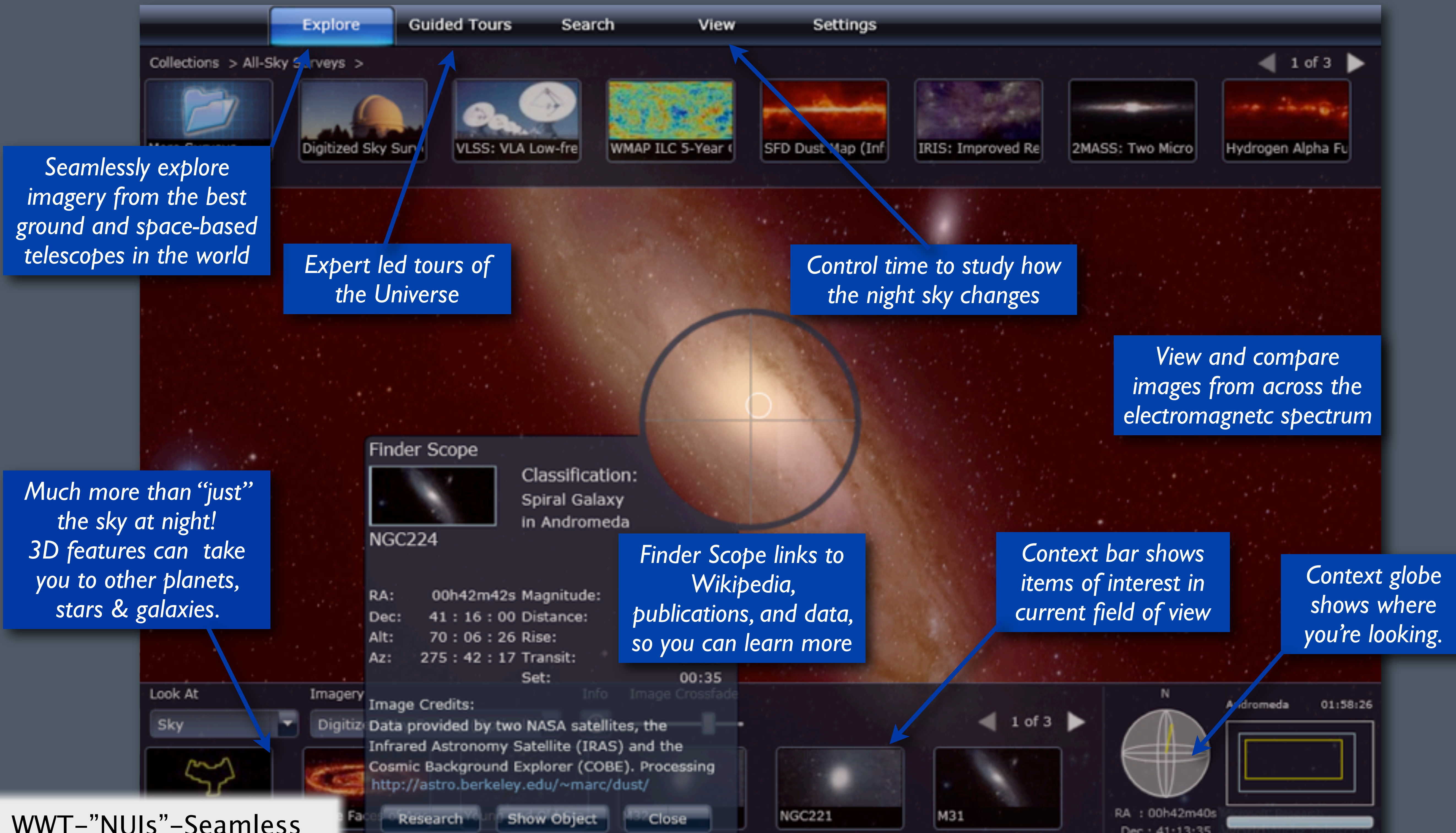
Alberto Accomazzi, Doug Burke, Alberto Conti, Carol Christian, Mercé Crosas, Raffaele D'Abrusco, Rahul Davé, Christopher Erdmann, Jonathan Fay, Jay Luker, Alyssa Goodman, Michael Kurtz, Gus Muench, Alberto Pepe, Curtis Wong



WWT-"NUIs"-Seamless Astronomy

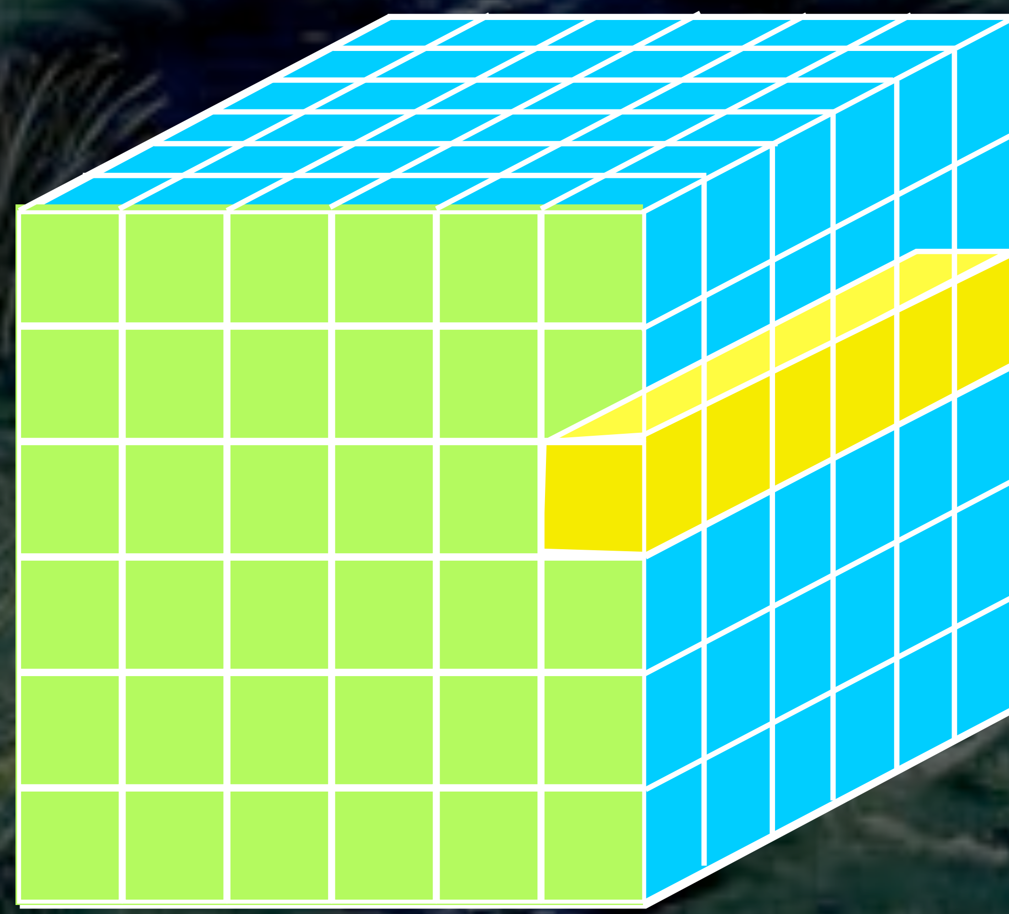
Microsoft® Research
WorldWide Telescope

Experience WWT at **worldwidetelescope.org**



WWT–"NUIs"–Seamless Astronomy

The dream scenario...



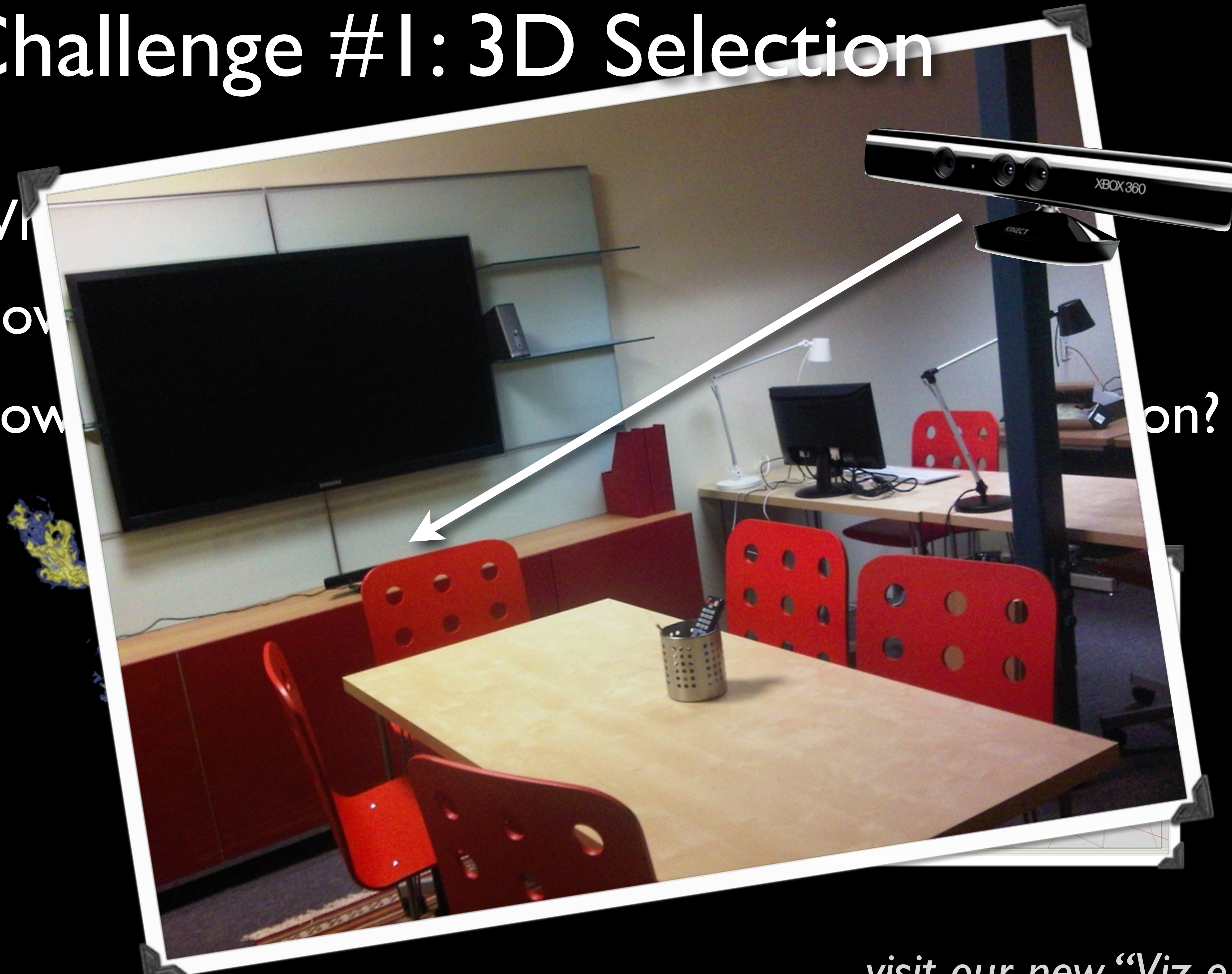
WWT-"NUIs"-Seamless Astronomy

Challenge #1: 3D Selection

Wh

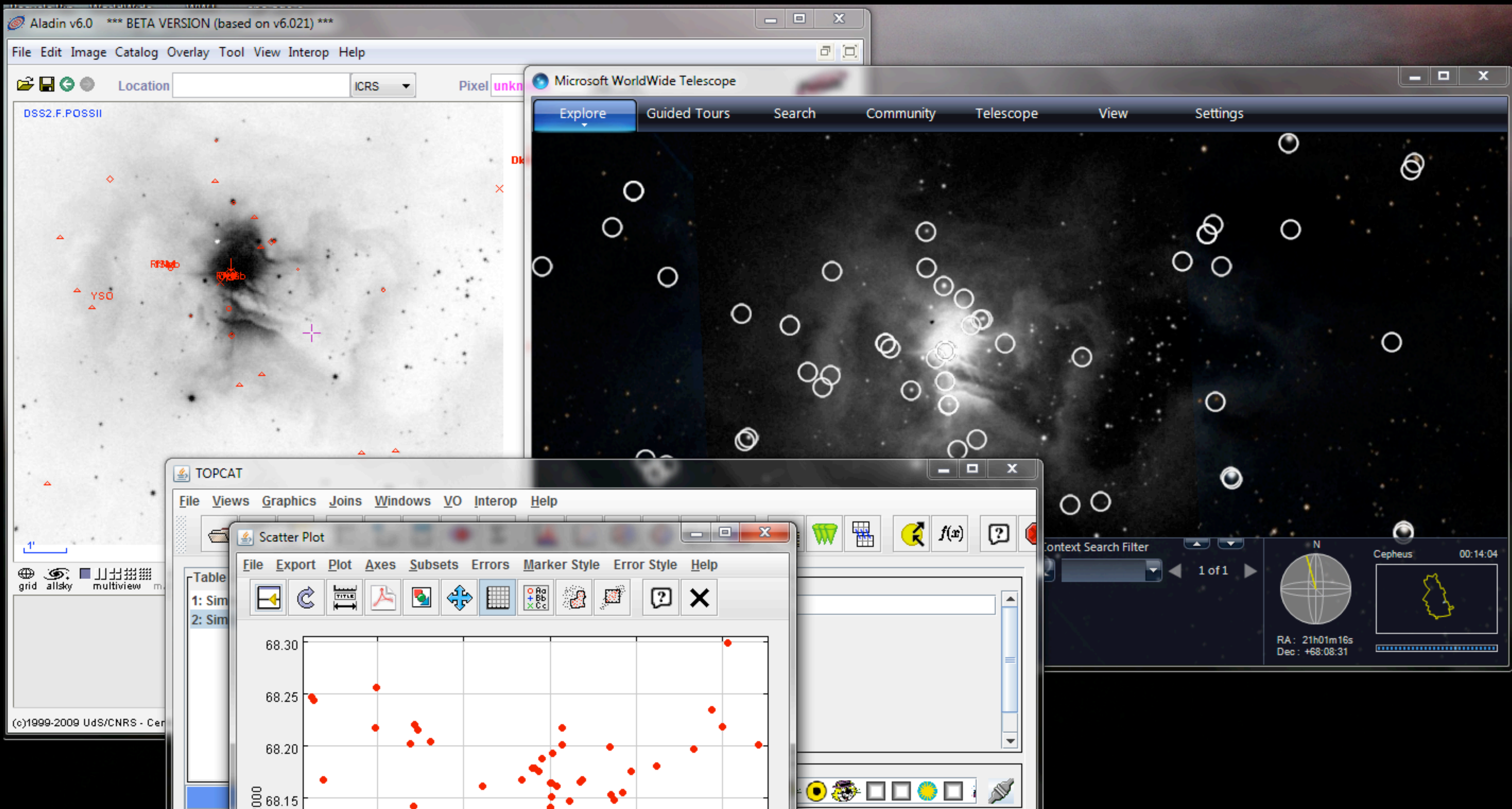
How

How ... on?



visit our new "Viz-e-Lab"!

Challenge #2: Too many windows...



Challenge #3:

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

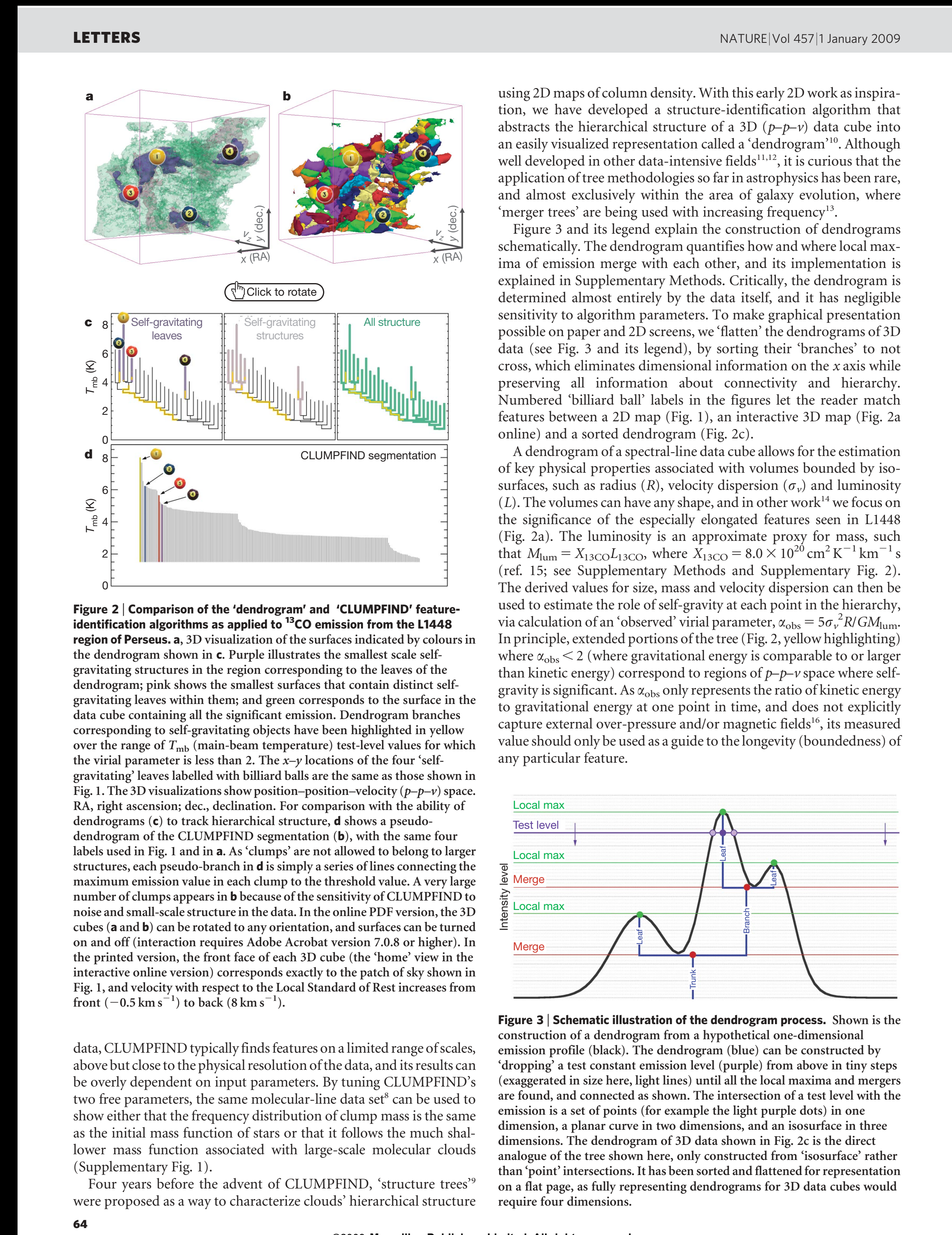


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating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

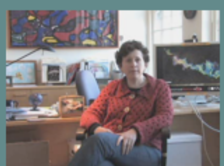
Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, 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.

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*

The Art of Numbers



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**

Alyssa A. Goodman

Harvard University (HCO+IIC)

Smithsonian Astrophysical Observatory

Scholar-in-Residence, WGBH