

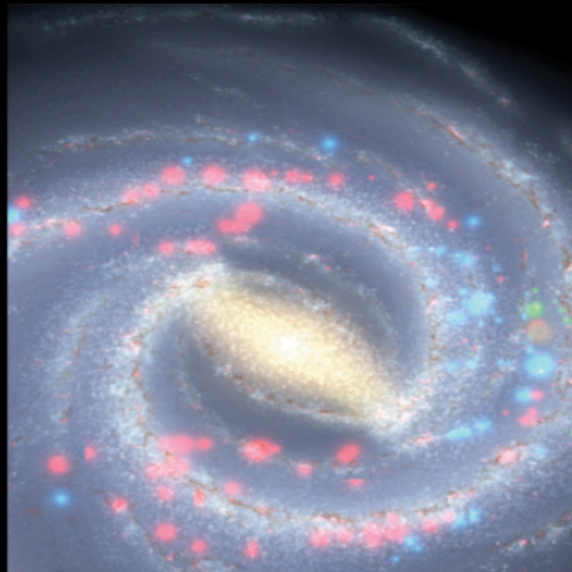
C I E R A

CENTER FOR
INTERDISCIPLINARY
EXPLORATION AND
RESEARCH IN
ASTROPHYSICS

Fall 2012 Interdisciplinary Colloquium

“THE CLOSE RELATIONSHIP BETWEEN HEART DISEASE AND SEA MONSTERS”

*(or, why good high-dimensional data visualization
can solve almost any problem!)*



This talk will explore the role of high-dimensional data visualization in a variety of research areas. In fields as diverse as Astronomy, Medicine, Geology, and Genomics, the progress of research is now limited more by the speed of data analysis and understanding than it is by data acquisition. Rich online resources offer not only ever-larger tables of numbers, but also images, movies, and other high-dimensional data streams. And, most results of previous research are available on-demand thanks to a growing wealth of digitized literature. The talk will offer real-world scenarios, focusing in particular on examples from Astronomy (including sea monsters in the sky) and Medicine (including how your doctor *should* look at your heart), where high-dimensional linked-view data visualization tools, such as the WorldWide Telescope and Glue software packages, have quickened the pace of insight.

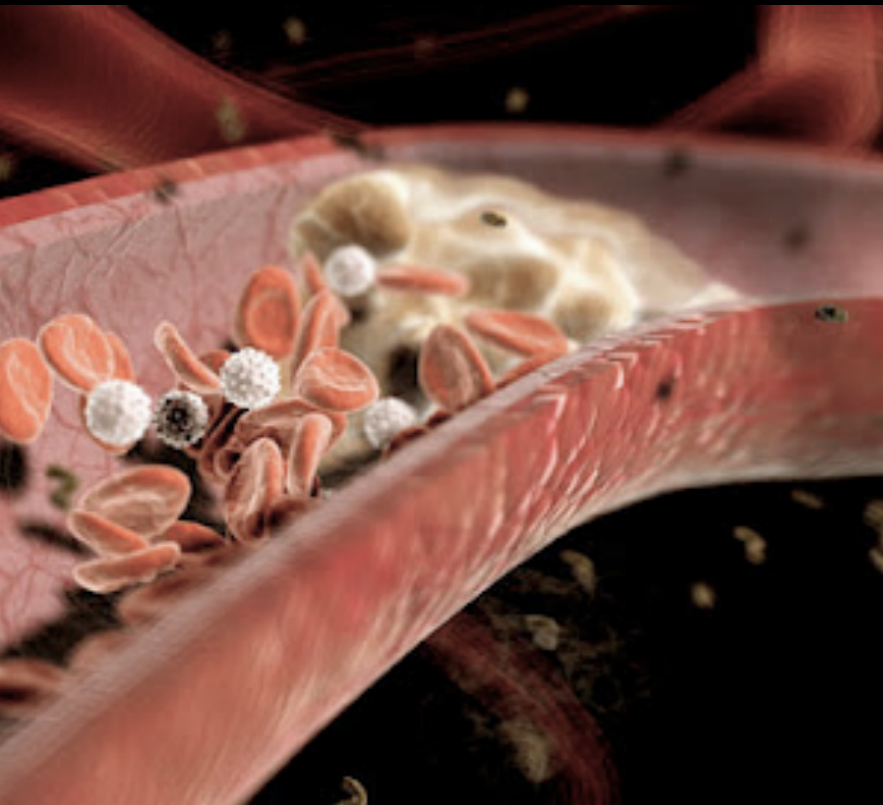
**Friday, November 30th
at 11:00am**
Technological Institute
2145 Sheridan Road, Room F160



Dr. Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics

The Close Relationship between Heart Disease and Sea Monsters*

**or, why good high-dimensional data visualization can solve almost any problem!*



Alyssa A. Goodman

Harvard-Smithsonian Center for Astrophysics

Relative Strengths



Pattern Recognition
Creativity



Calculations



"e-Science"

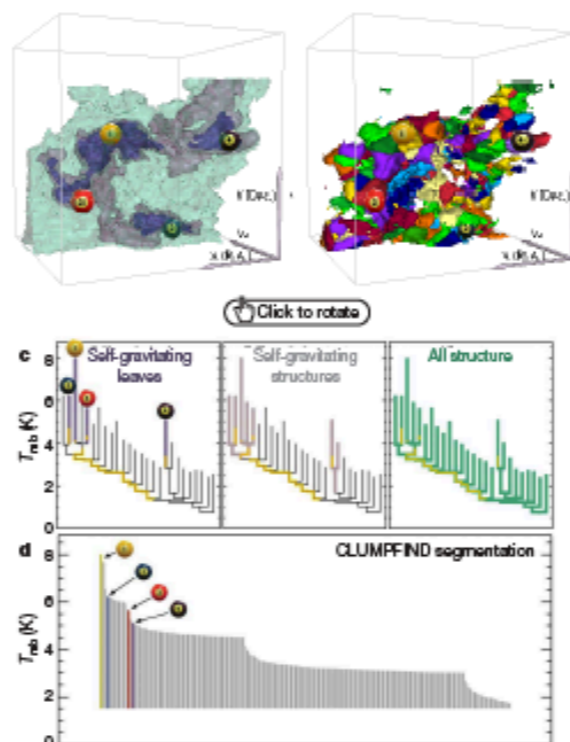
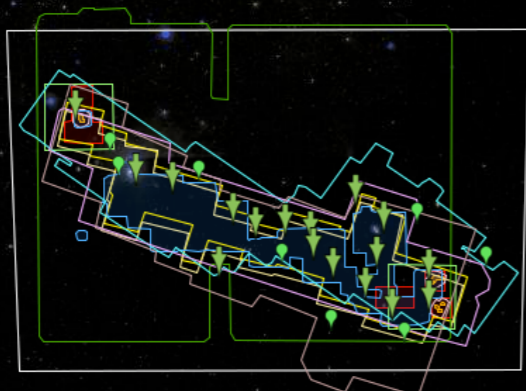


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ¹³CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in c. Purple illustrates the smallest scale self-gravitating structures 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 set⁸ 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 frequency¹³.

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 work¹⁴ 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 $M_{\text{gas}} = X_{13\text{CO}} L_{13\text{CO}}$, where $X_{13\text{CO}} = 8.0 \times 10^{20} \text{ cm}^2 \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ (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, $\alpha_{\text{obs}} = 5\sigma_v^2 R / GM_{\text{gas}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{\text{obs}} < 2$ (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 fields¹⁶, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

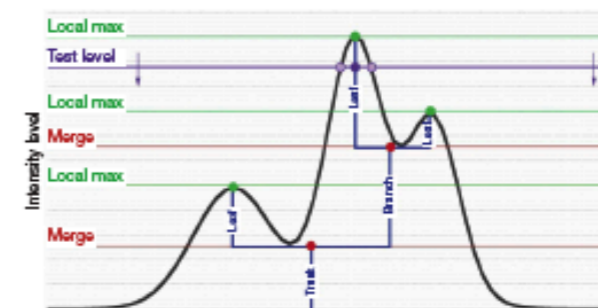


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in 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.

"High-dimensional" or "Multivariate" Data (Astronomy=Biology)

LETTERS

NATURE | Vol 000 | 00 Month 2008

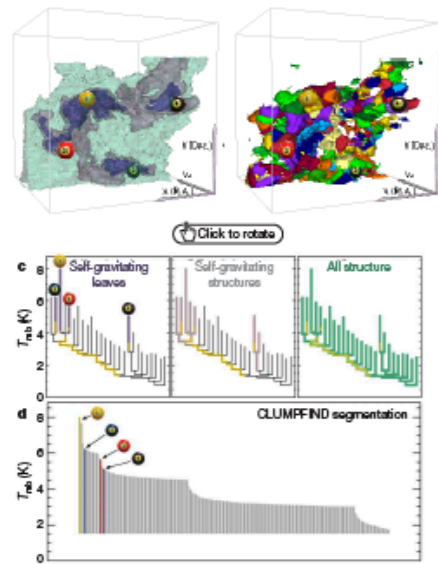


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ¹³CO emission from the L1448 region of Perseus. **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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 labeled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position-position-velocity ($p-p-v$) space. R.A., 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 set⁹ 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 frequency¹³.

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 work¹⁴ 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 $M_{\text{gas}} = X_{13\text{CO}} L_{13\text{CO}}$, where $X_{13\text{CO}} = 8.0 \times 10^{20} \text{ cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ (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, $\alpha_{\text{obs}} = 5\sigma_v^2 R / GM_{\text{gas}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{\text{obs}} < 2$ (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 fields¹⁶, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

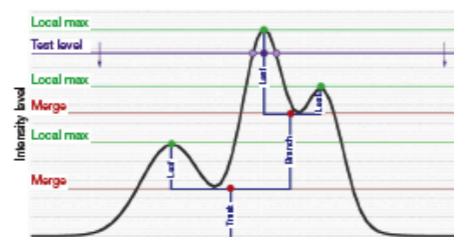
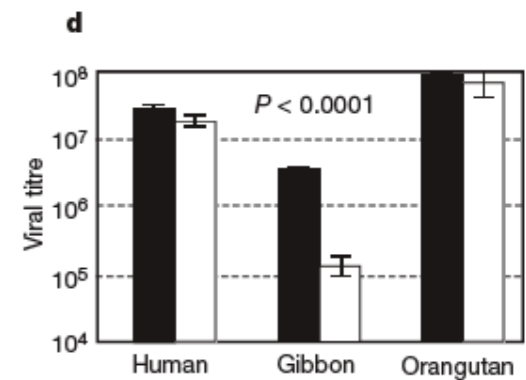
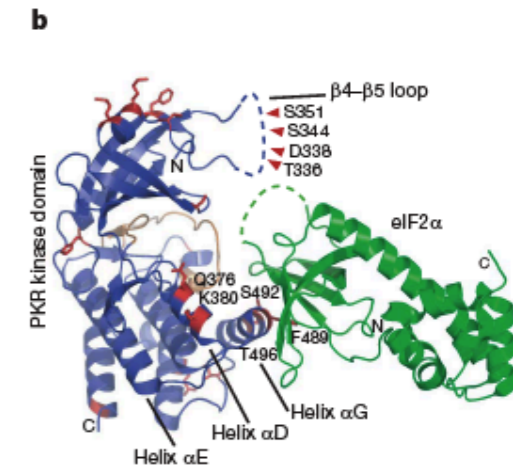
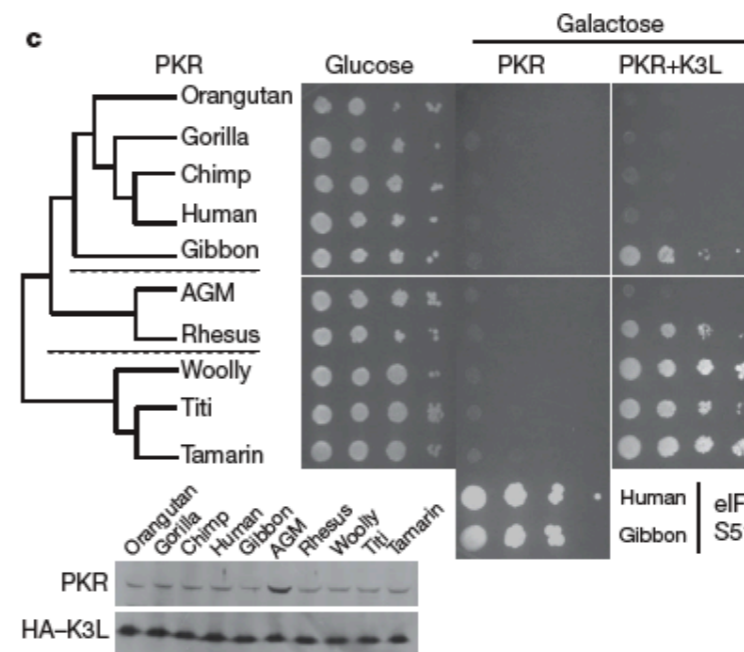
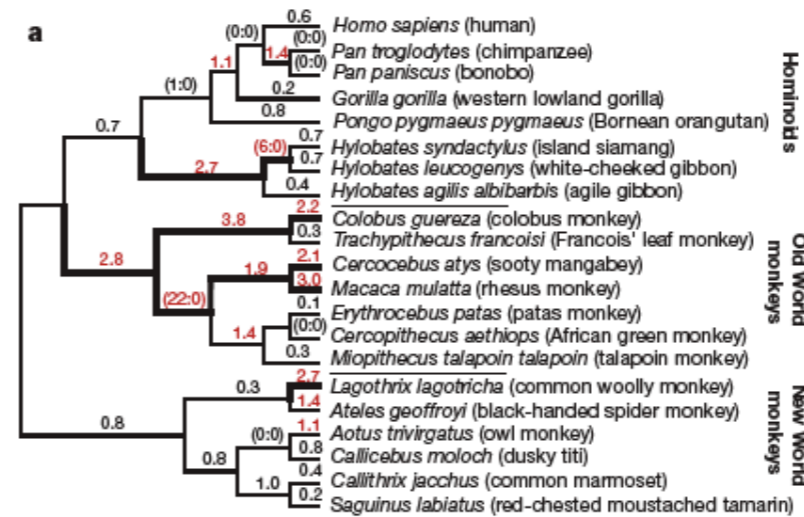


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and merges 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.

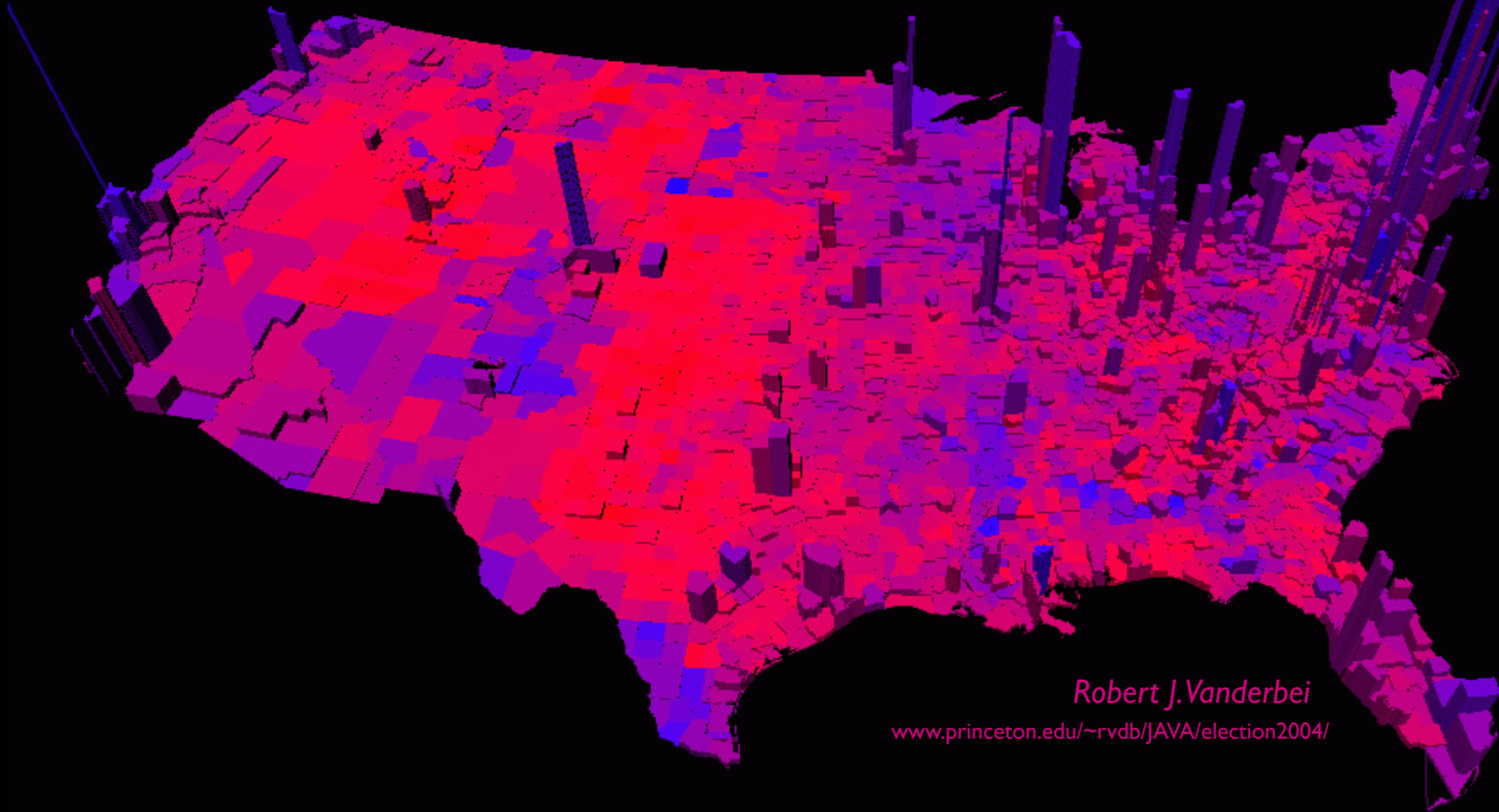
2

©2008 Macmillan Publishers Limited. All rights reserved

LETTERS



"High-dimensional" or "Multivariate" Data



Robert J. Vanderbei

www.princeton.edu/~rvdb/JAVA/election2004/

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

Data • Dimensions • Display

Astron. Nachr. / AN 333, No. 5/6, 505–514 (2012) / DOI 10.1002/asna.201211705

Principles of high-dimensional data visualization in astronomy

A.A. Goodman*

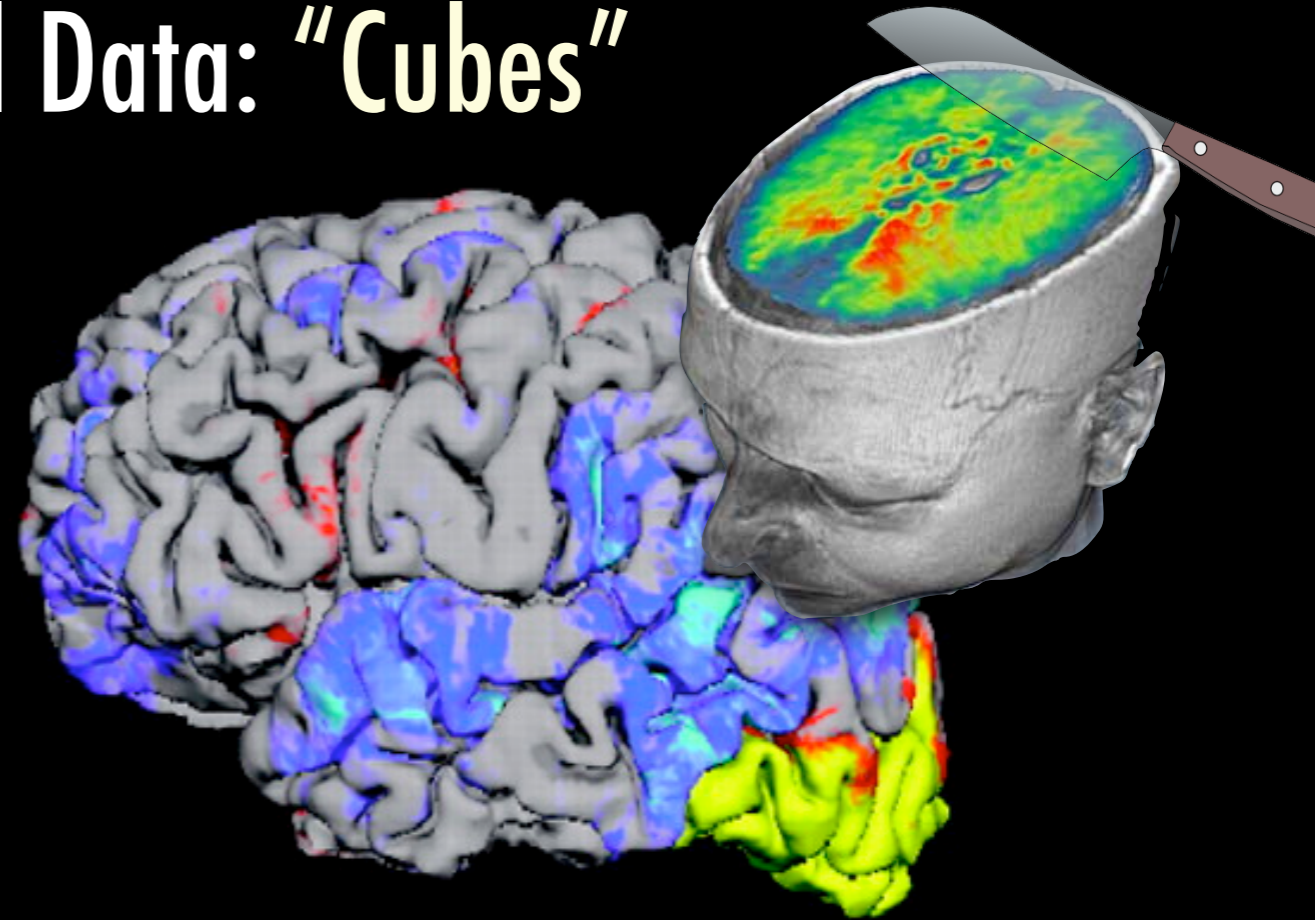
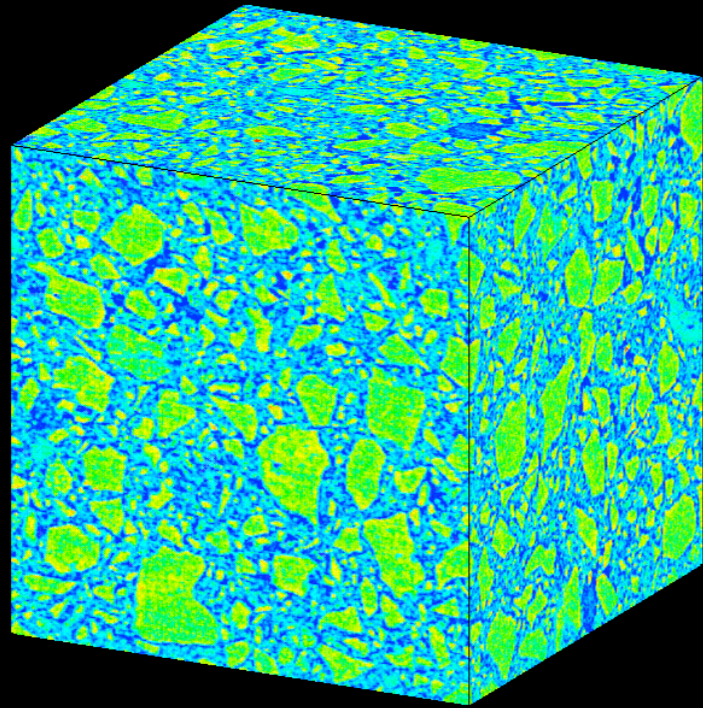
Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2012 May 3, accepted 2012 May 4
Published online 2012 Jun 15

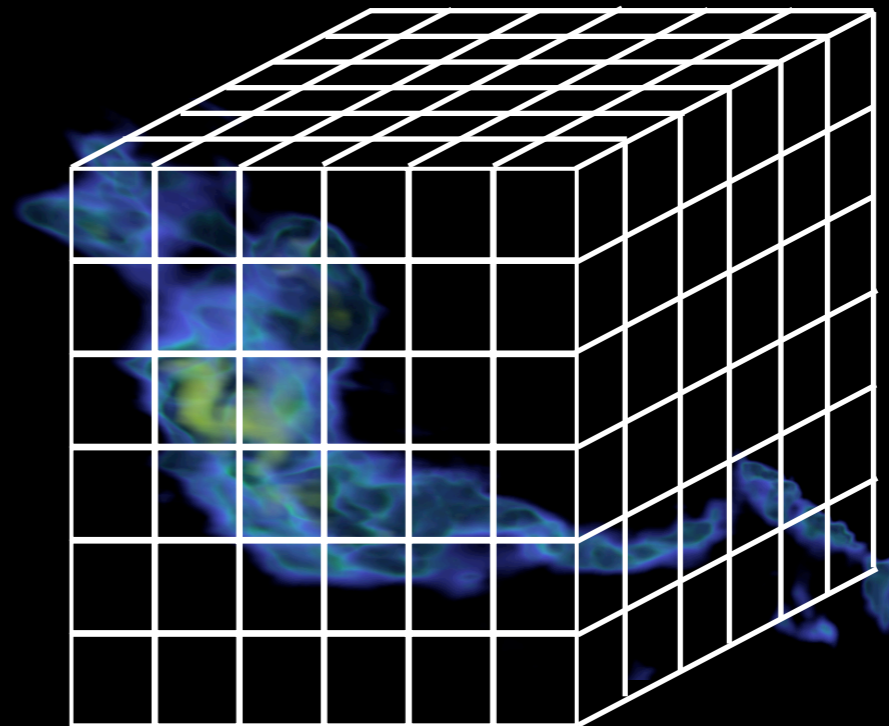
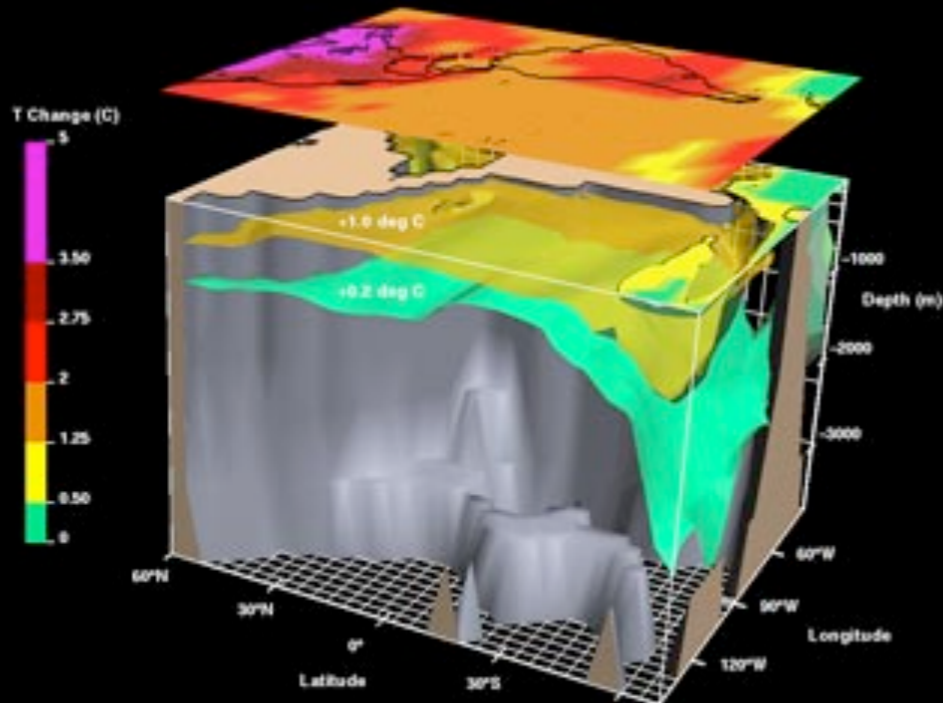
Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing
techniques: radial velocities

Astronomical researchers often think of analysis and visualization as a linear process, starting with data sets, though, interactive *exploratory data analysis* and statistical analysis are often used to explore the data.

High-Dimensional Data: "Cubes"



ATMOSPHERIC AND OCEANIC TEMPERATURE CHANGE

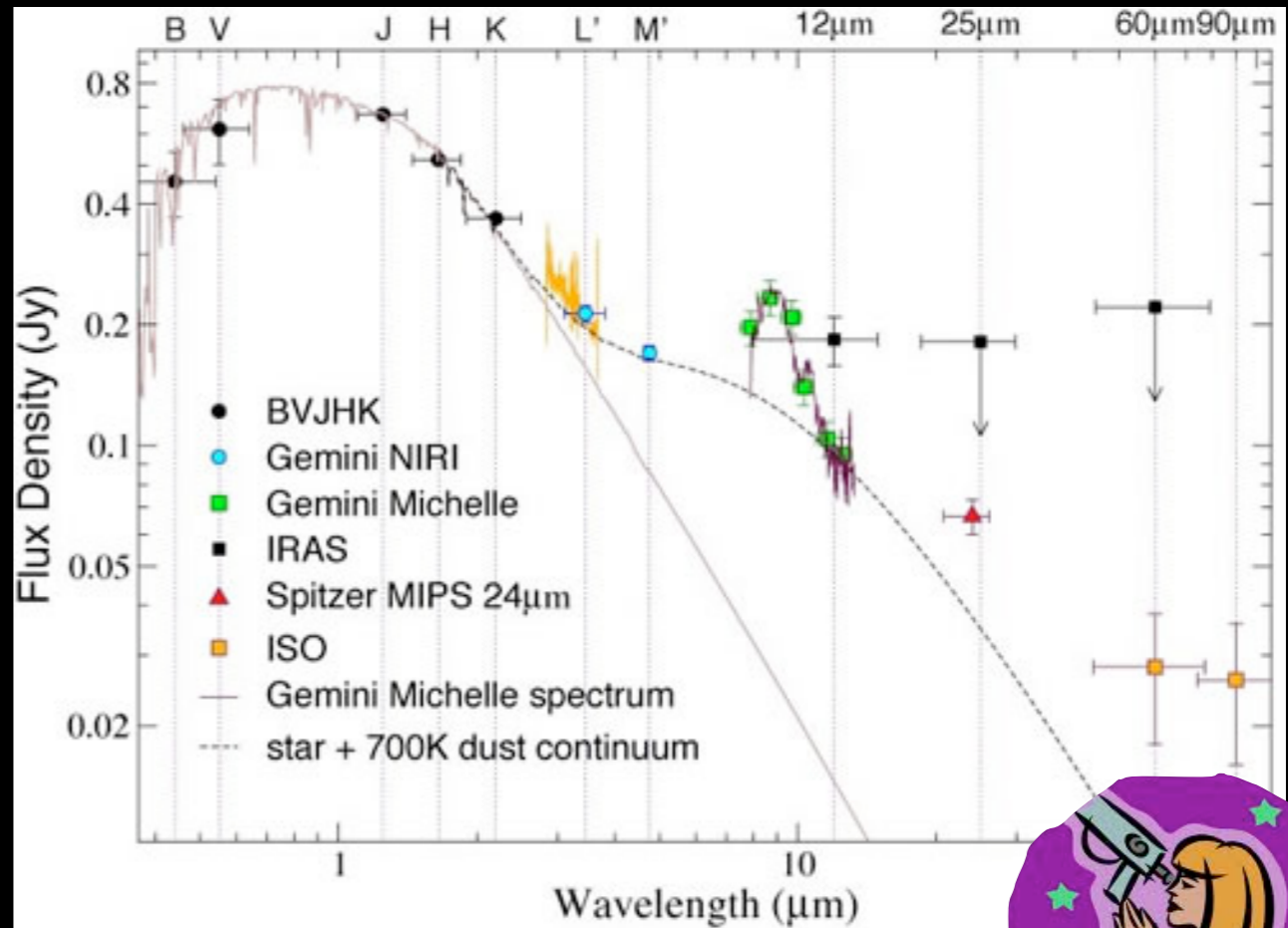
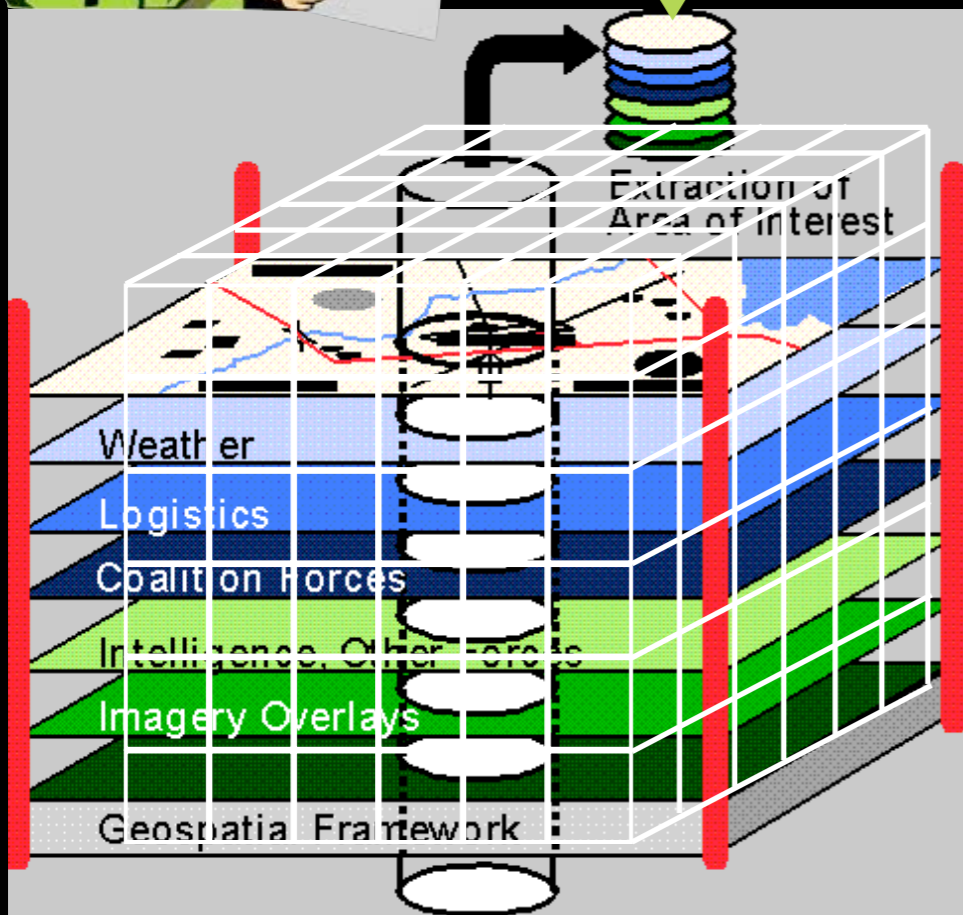


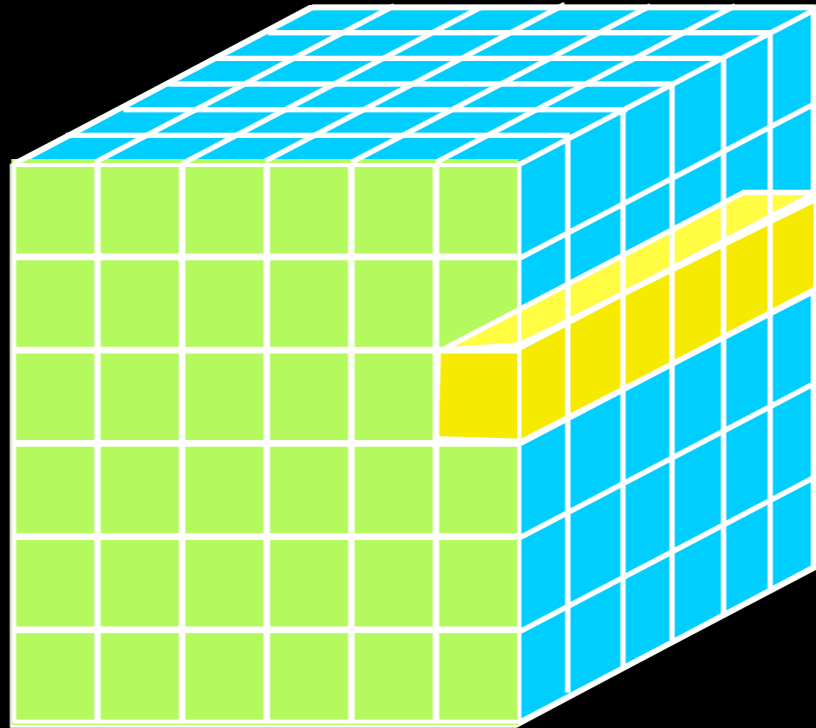


This I-D “Bore”



is a “spectral energy distribution”



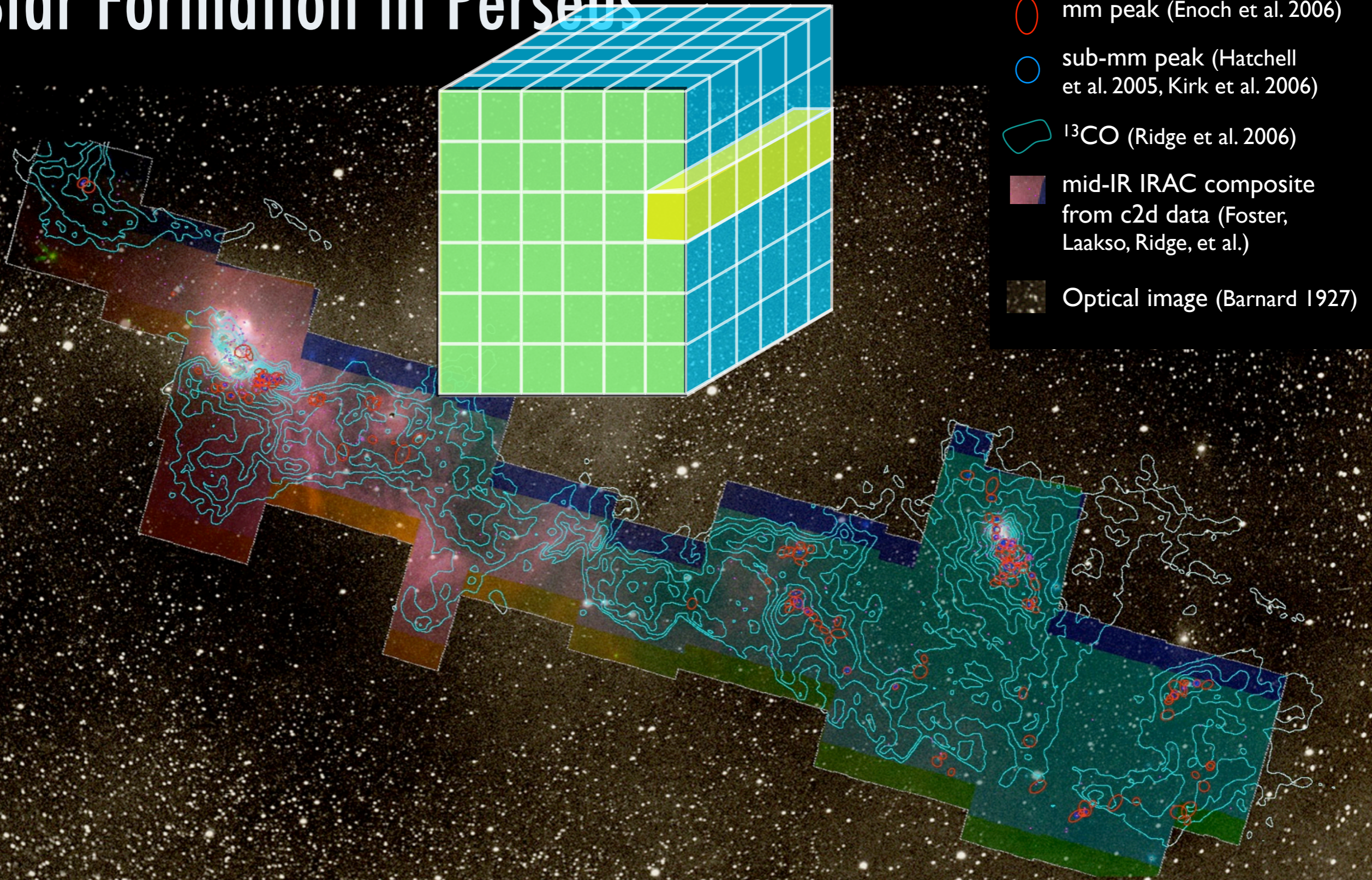


GENERALLY

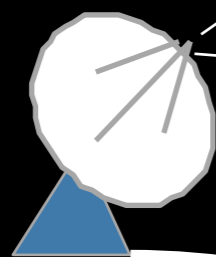
- 1D:** Columns = “Spectra”, “SEDs” or “Time Series”
- 2D:** Faces or Slices = “Images”
- 3D:** Volumes = “3D Renderings”, “2D Movies”
- 4D:** Time Series of Volumes = “3D Movies”

Star Formation in Perseus

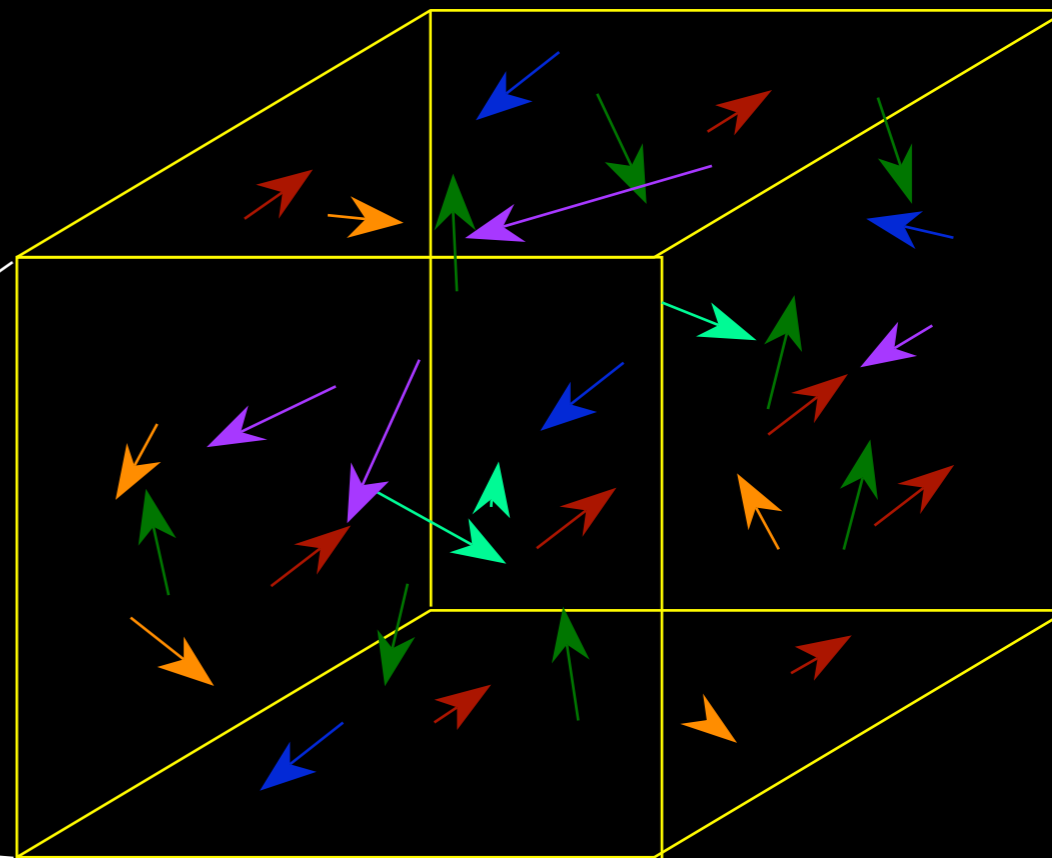
COMPLETE



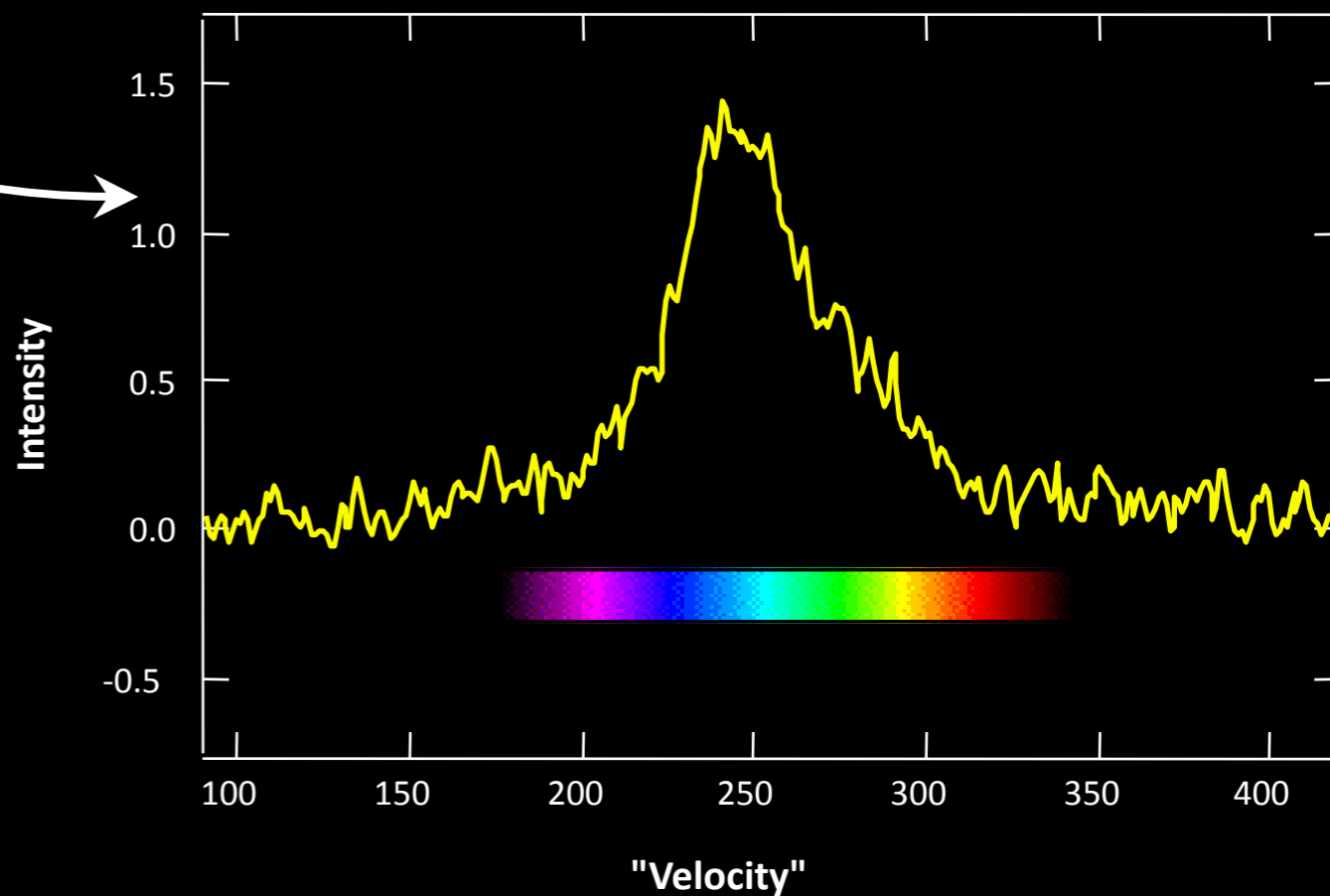
Velocity from Spectroscopy



Telescope +
Spectrometer



Observed Spectrum

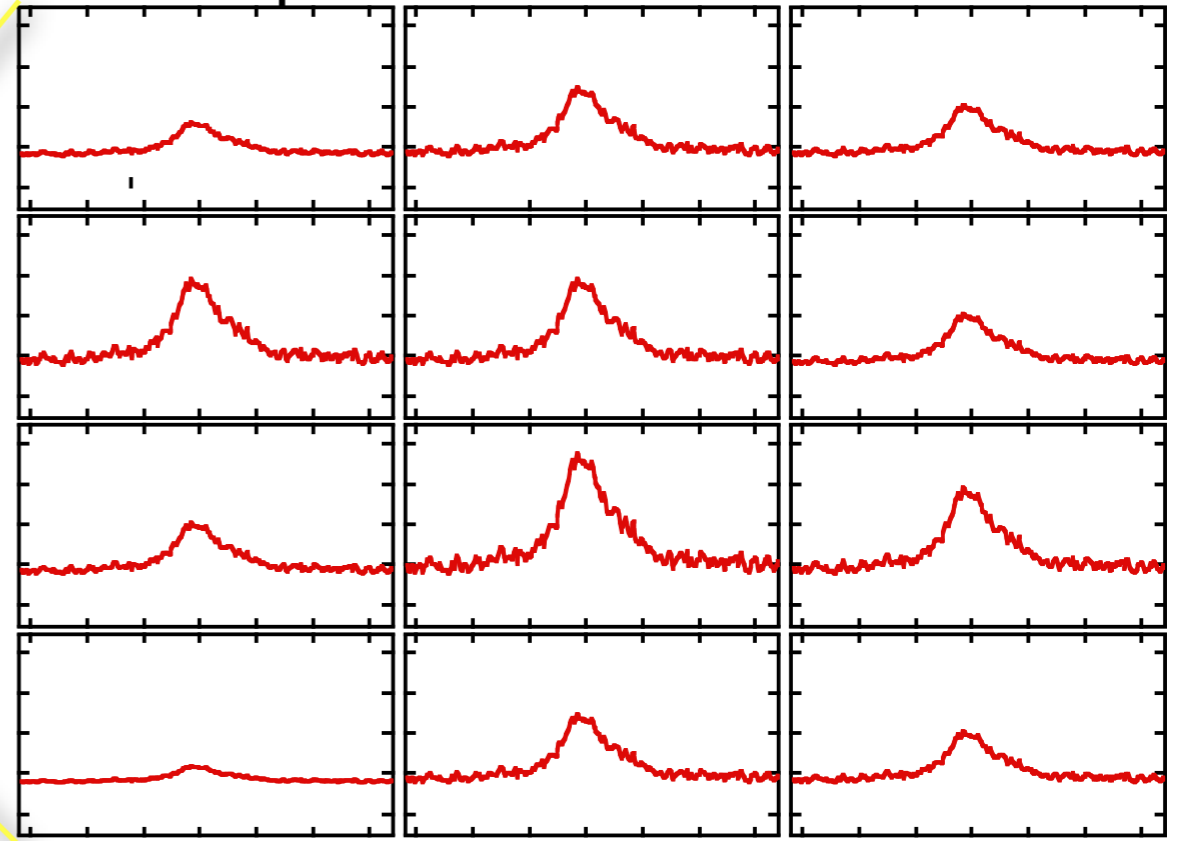


All thanks to Doppler

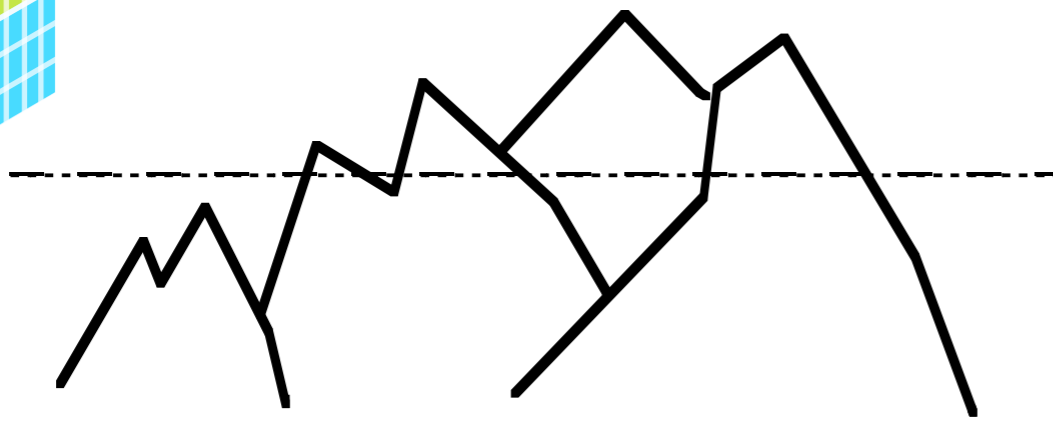
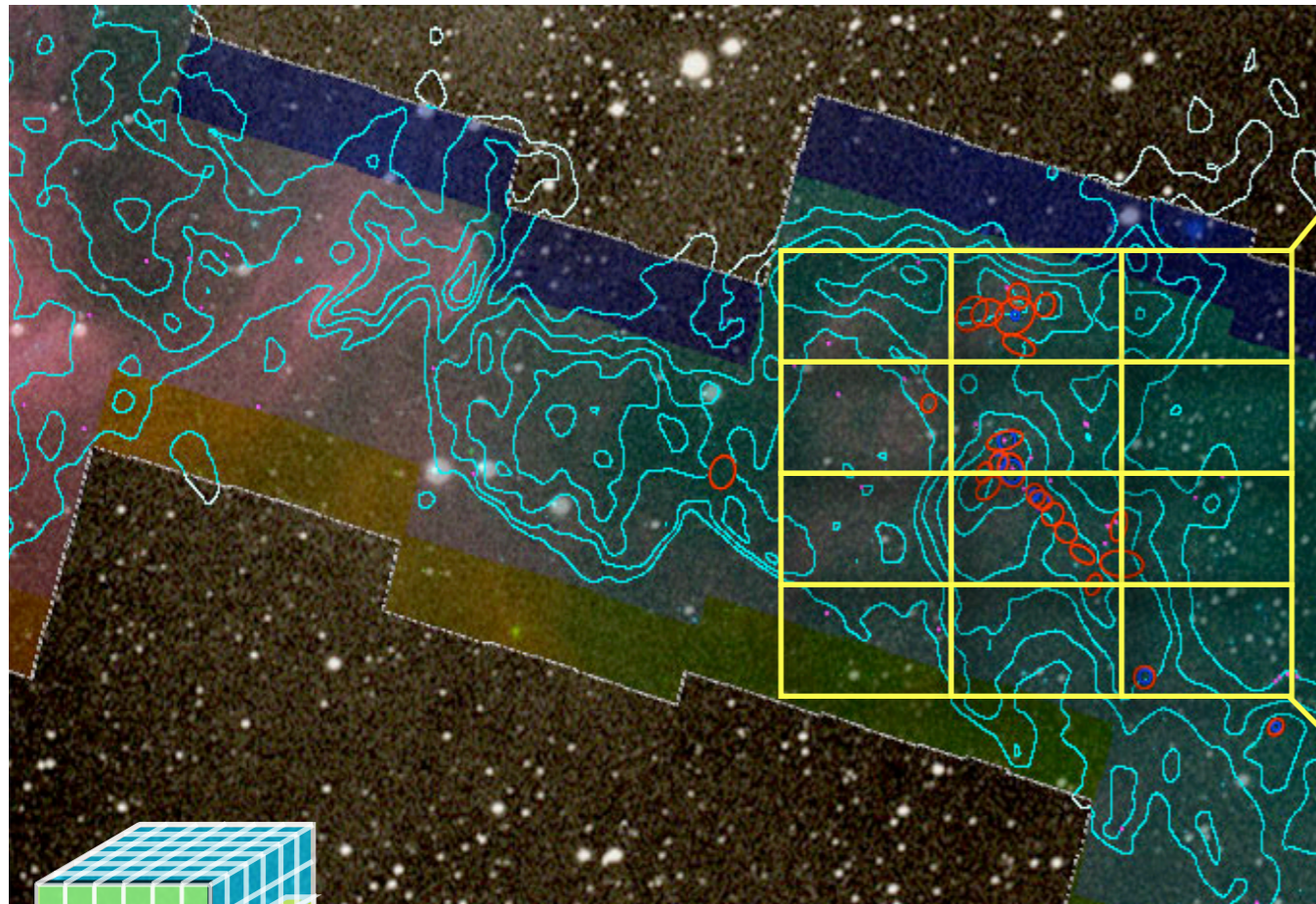
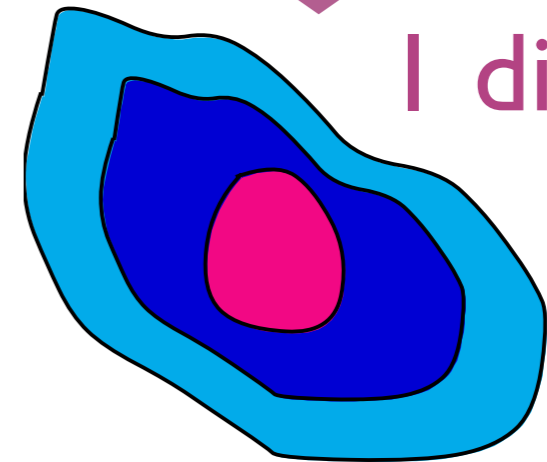


Spectral-Line Mapping

Spectral Line Observations



Loss of
1 dimension








Mountain Range

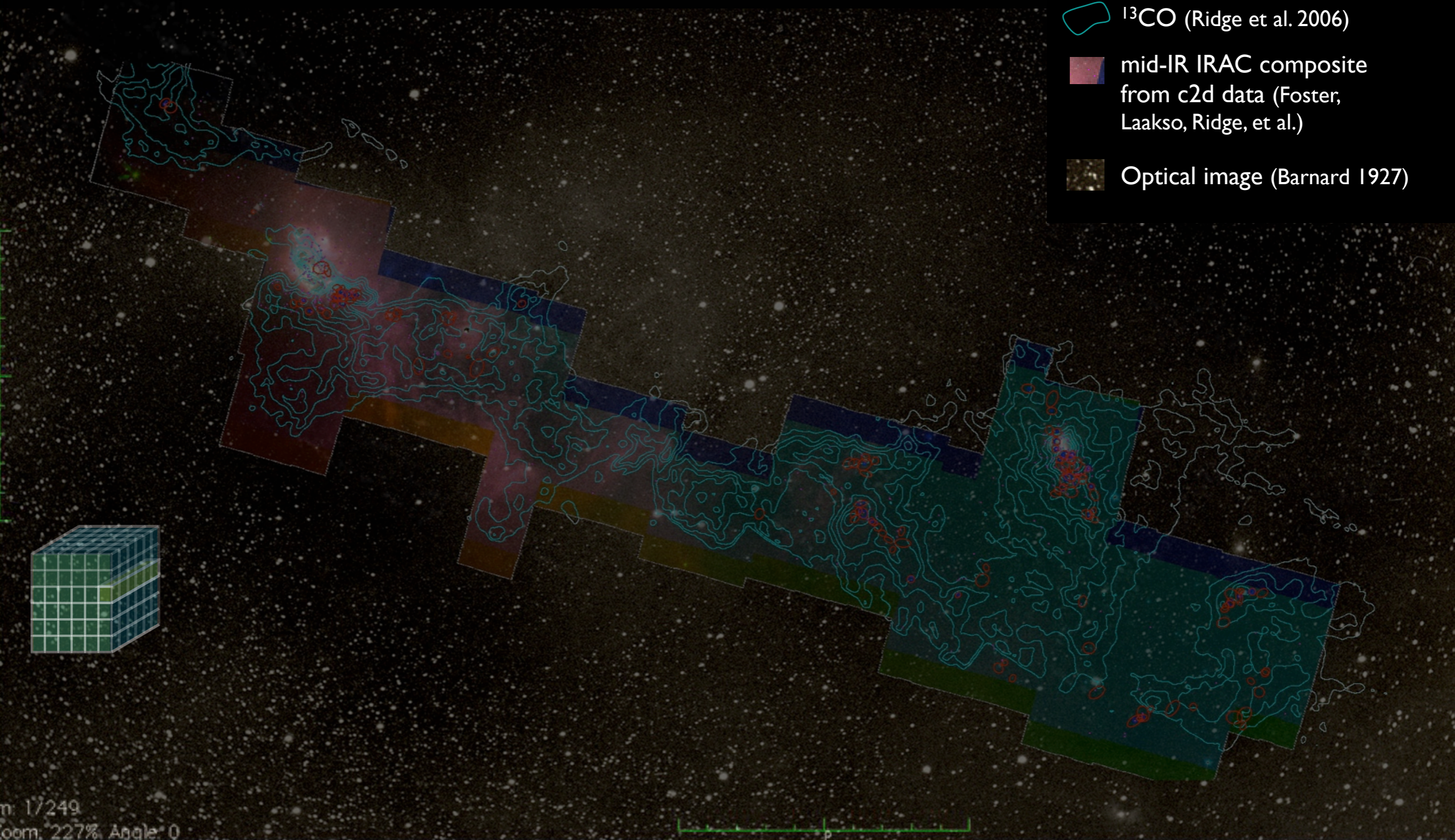
No loss of
information



COMPLETE Perseus

Image size: 1305 x 733
VL: 63 WW: 127

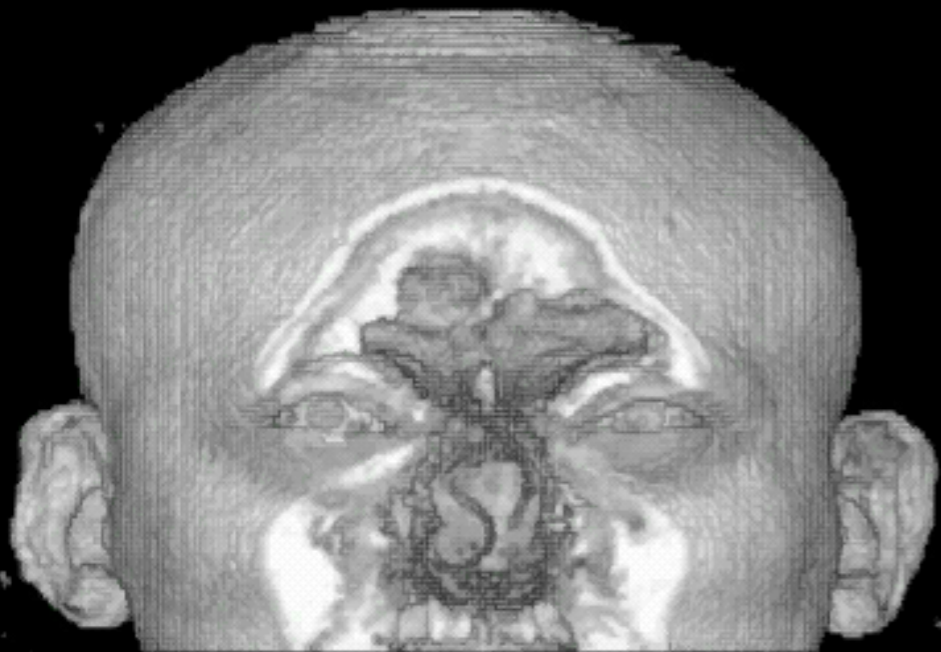
-  mm peak (Enoch et al. 2006)
-  sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
-  ^{13}CO (Ridge et al. 2006)
-  mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
-  Optical image (Barnard 1927)



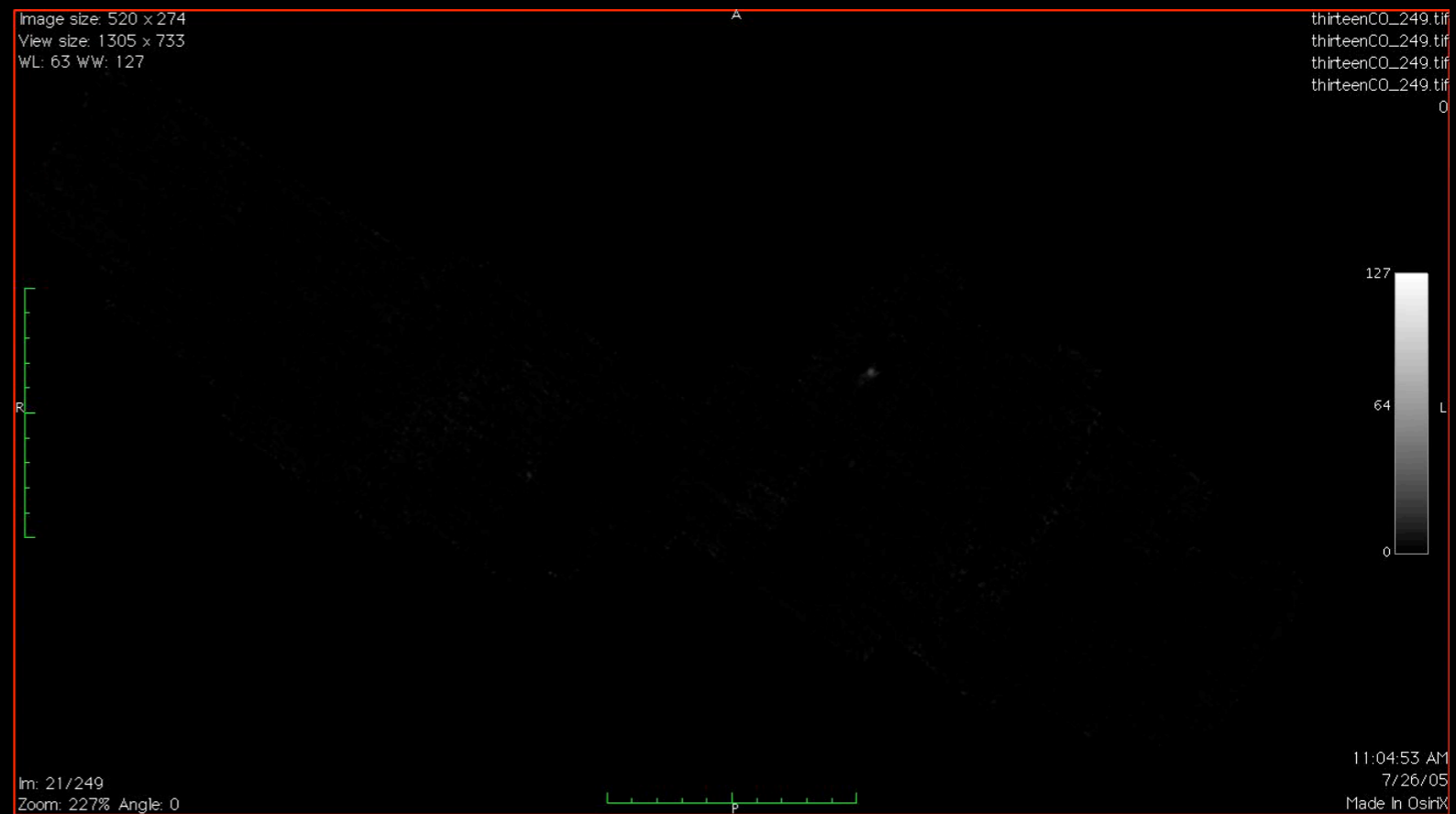
m: 17249
Zoom: 227% Angle: 0

"Astronomical Medicine"

"KEITH"



"PERSEUS"



"z" is depth into head

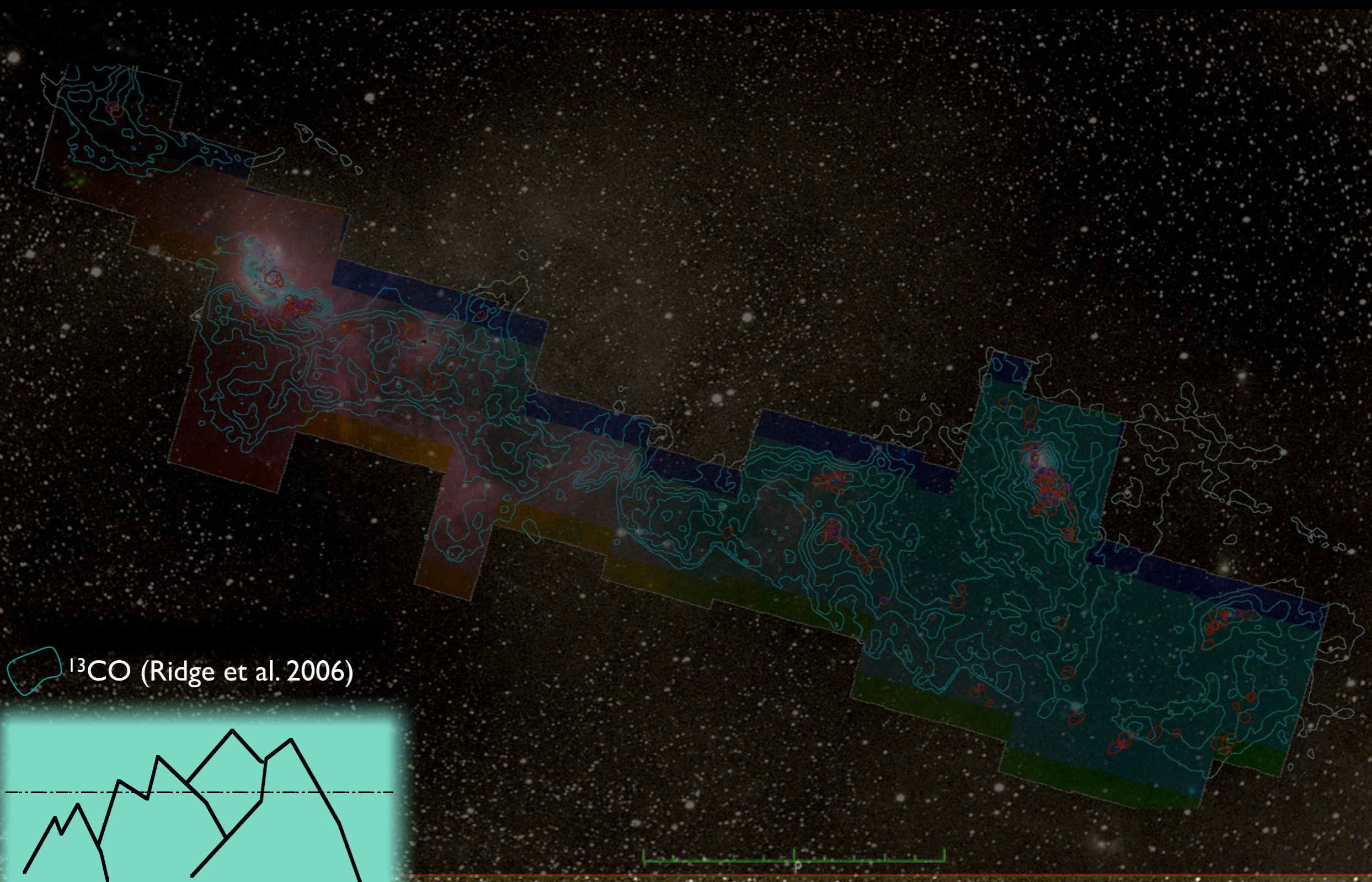
"z" is line-of-sight velocity

<http://am.iic.harvard.edu/>

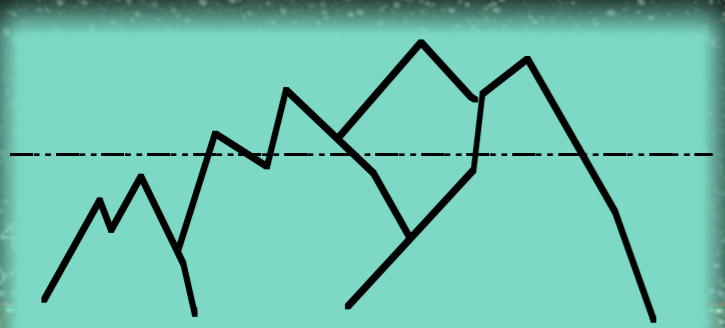


Image size: 520 x 274
View size: 1305 x 733
VL: 63 WW: 127

COMPLETE Perseus

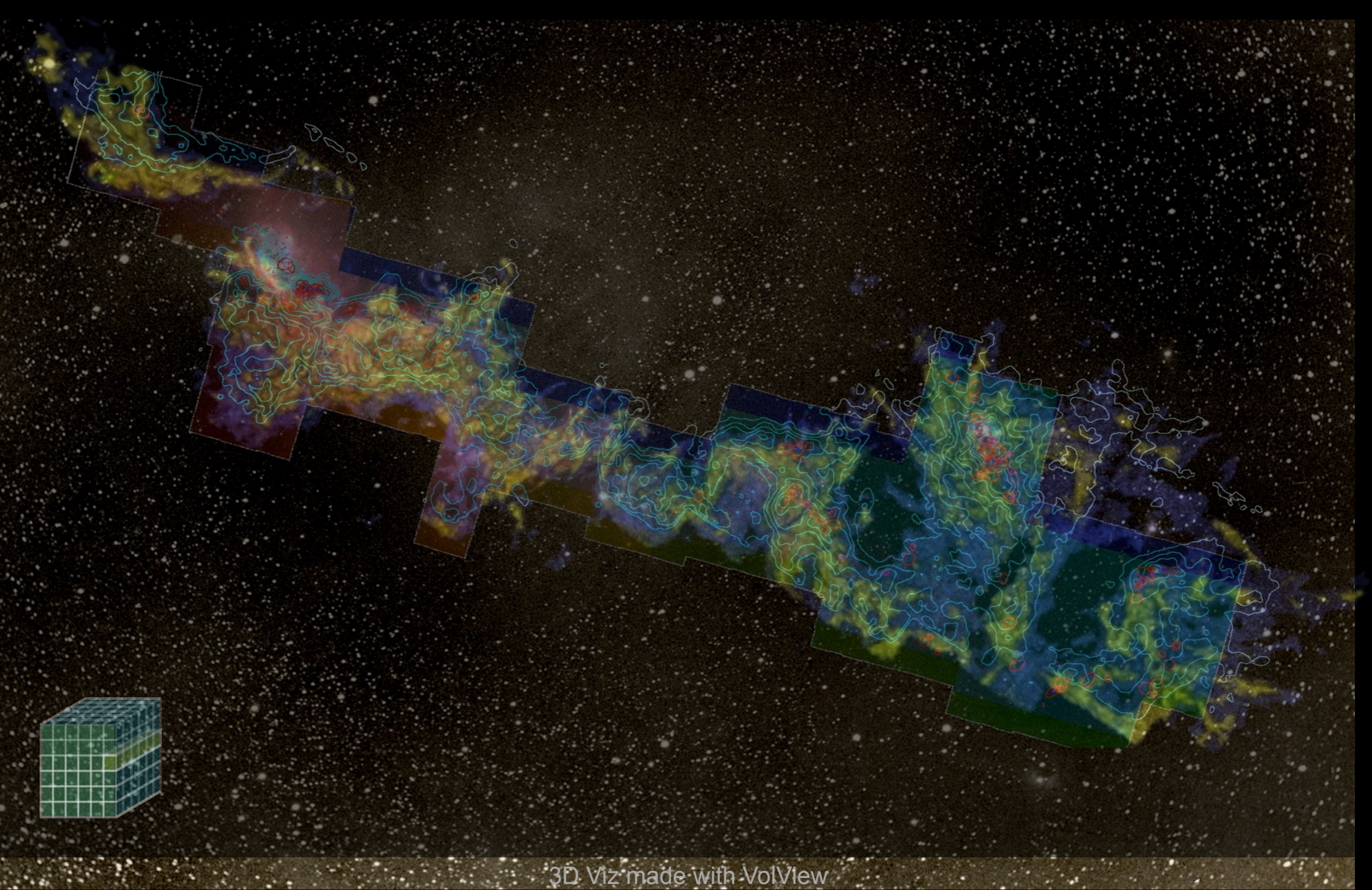


 ^{13}CO (Ridge et al. 2006)

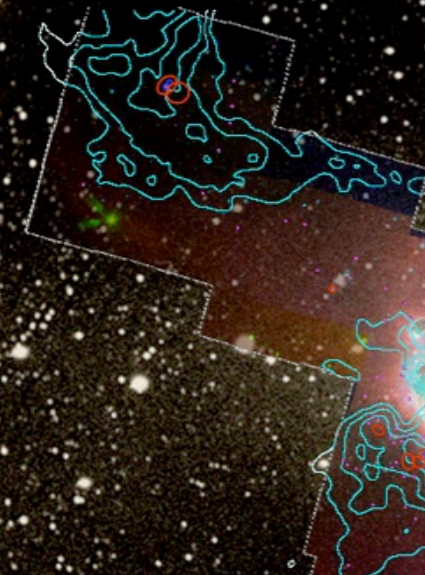


Mountain Range

m: 1724
oom: 2



3D Viz made with VolView



LETTERS

NATURE | Vol 000 | 00 Month 2008

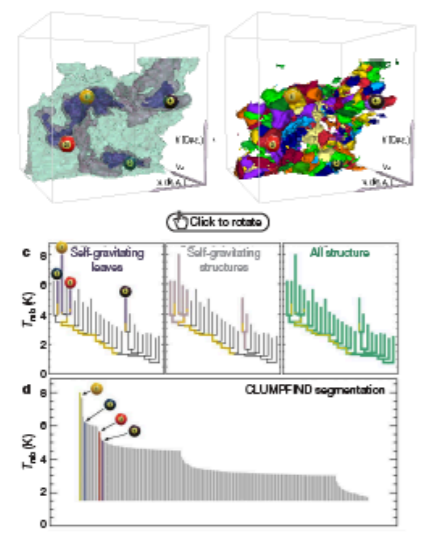


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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 labeled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position-position-velocity (p - p - v) space. R.A., 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 set⁴ 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 frequency¹³.

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 iso-surfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work¹⁴ 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 $M_{\text{gas}} = X_{13\text{CO}} L_{13\text{CO}}$, where $X_{13\text{CO}} = 8.0 \times 10^{20} \text{ cm}^{-2} \text{ K}^{-1} \text{ s}$ (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, $\alpha_{\text{obs}} = 5\sigma_v^2 R / GM_{\text{gas}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{\text{obs}} < 2$ (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 fields⁶, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

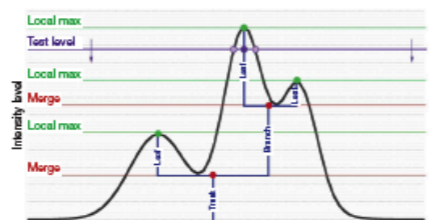
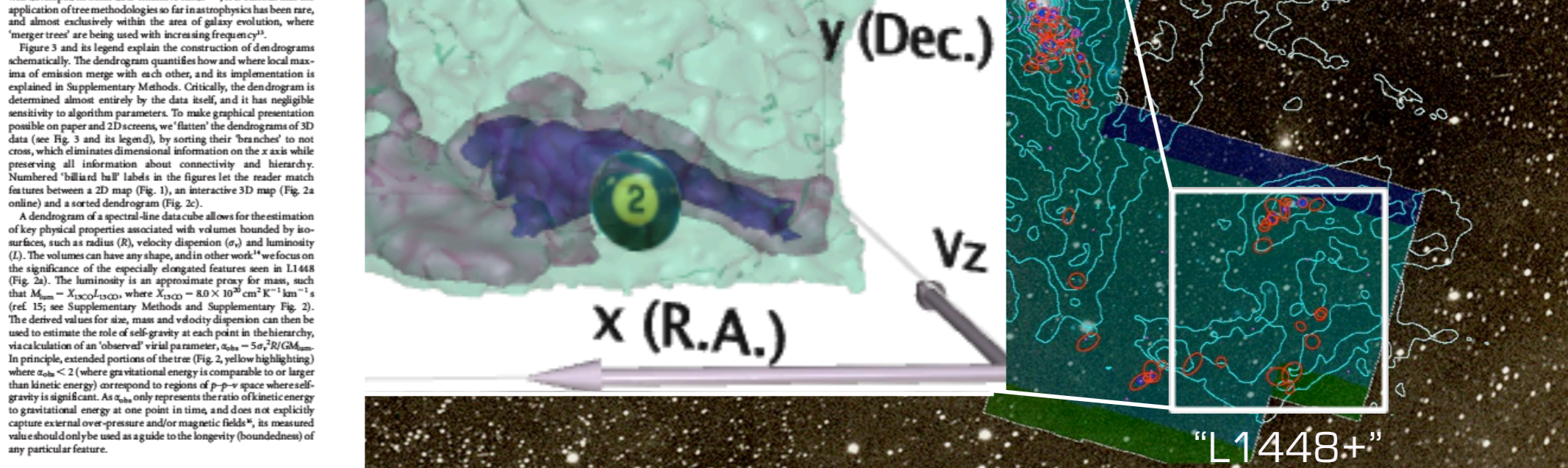


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in 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 iso-surface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'iso-surface' 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.



"Rich Online Resources"

offer not only ever-larger tables of numbers, but also images, movies, and other high-dimensional data streams.

And, most results of previous research are available on-demand thanks to a growing wealth of digitized literature.

SEAMLESS ASTRONOMY
Linking scientific data, publications, and communities

ABOUT PROJECTS PEOPLE RESOURCES DATAVERSE

SEAMLESS ASTRONOMY

About

The Seamless Astronomy Group at the Harvard-Smithsonian Center for Astrophysics brings together astronomers, computer scientists, information scientists, librarians and visualization experts involved in the development of tools and systems to study and enable the next generation of online astronomical research.

Current projects include research on the development of systems that seamlessly integrate scientific data and literature, the semantic interlinking and annotation of scientific resources, the study of the impact of social media and networking sites on scientific dissemination, and the analysis and visualization of astronomical research communities. Visit our project page to find out more.

COMPLETE Data Available

Center on Perseus Center on Ophiuchus Center on Serpens

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

Dataset	Show	Perseus	Ophiuchus	Serpens	Link
GBT: HI Data Cube	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Data
IRAS: Av/Temp Maps	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data
FCRAO: 12CO	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data
FCRAO: 13CO	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data
JCMT: 850 microns	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Data
Spitzer c2d: IRAC 1,3 (3.6,5.8 μm)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data
Spitzer c2d: IRAC 2,4 (4.5,8 μm)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data
CSO/Bolocam: 1.2-mm	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data
Spitzer MIPS: Derived Dust Map	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data

Targeted Regions (Phase II, Some Data Not Yet Available)

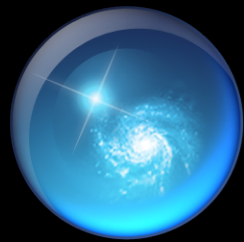
CTIO/Calar Alto: NIR (J,H,Ks)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Data
IRAM 30-m: N2H+ and C18O	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data
IRAM 30-m: 1.1-mm continuum	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data
Megacam/MMT: r ₁ z Images	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data

Catalogs & Pointed Surveys

NH3 Pointed Survey	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data
YSO Candidate list (c2d)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data

WorldWide Telescope

[NGC 1333/theastrodata demo]



Microsoft® Research WorldWide Telescope

Experience WWT at worldwidetelescope.org



Explore | Guided Tours | Search | View | Settings

Collections > All-Sky Surveys >

Digitized Sky Survey | VLSS: VLA Low-frequency Sky Survey | WMAP ILC 5-Year Cosmic Microwave Background | SFD Dust Map (Infrared) | IRIS: Improved Resolution | 2MASS: Two Micron All Sky Survey | Hydrogen Alpha Full Sky

1 of 3

Finder Scope

Classification: Spiral Galaxy In Andromeda

NGC224

RA: 00h42m42s Magnitude:
 Dec: 41 : 16 : 00 Distance:
 Alt: 70 : 06 : 26 Rise:
 Az: 275 : 42 : 17 Transit:
 Set: 00:35

Image Credits: Data provided by two NASA satellites, the Infrared Astronomy Satellite (IRAS) and the Cosmic Background Explorer (COBE). Processing <http://astro.berkeley.edu/~marc/dust/>

Research | Show Object | Close

Look At: Sky | Imagery: Digitized Sky Survey | Info | Image Crossfade

Andromeda | Three Faces of Andromeda | NGC221 | M31

1 of 3

Context bar shows items of interest in current field of view

Context globe shows where you're looking.

RA : 00h42m40s
Dec : 41:13:35

Seamlessly explore imagery from the best ground and space-based telescopes in the world

Expert led tours of the Universe

Control time to study how the night sky changes

View and compare images from across the electromagnetic spectrum

Much more than "just" the sky at night! 3D features can take you to other planets, stars & galaxies.

Finder Scope links to Wikipedia, publications, and data, so you can learn more

Context bar shows items of interest in current field of view

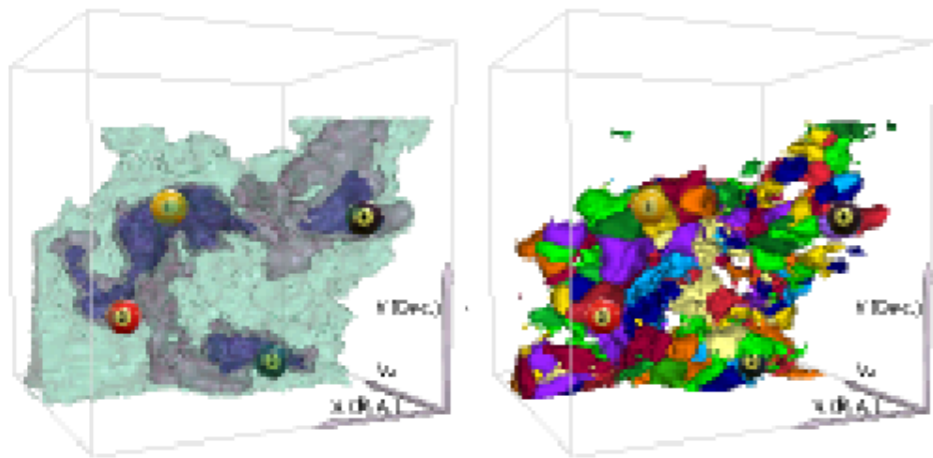
Context globe shows where you're looking.



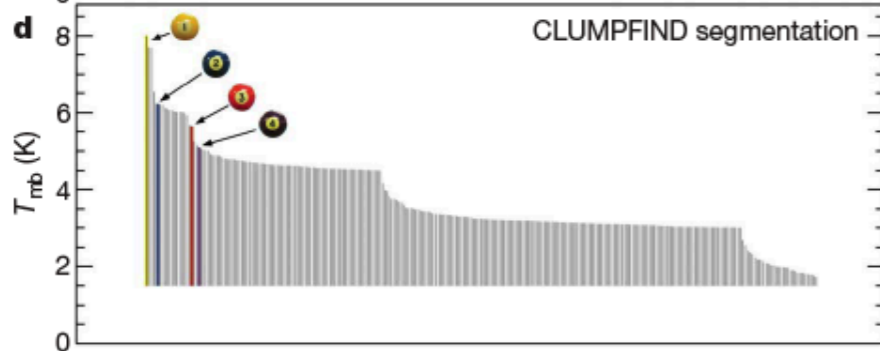
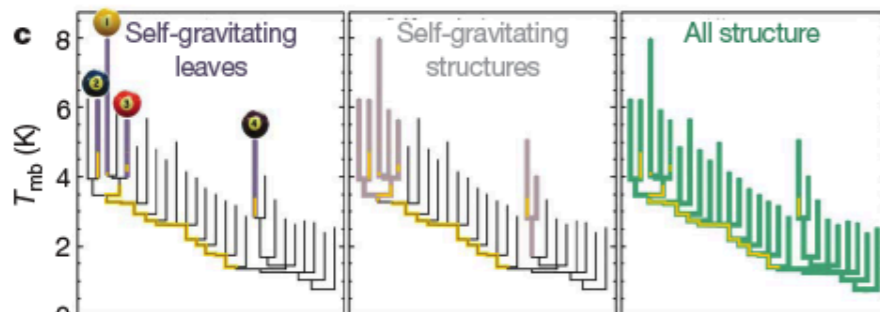
Heart Disease & Sea Monsters

"High-dimensional" or "Multivariate" Data (Astronomy=Biology)

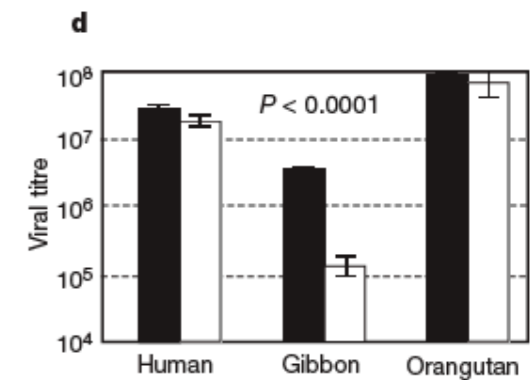
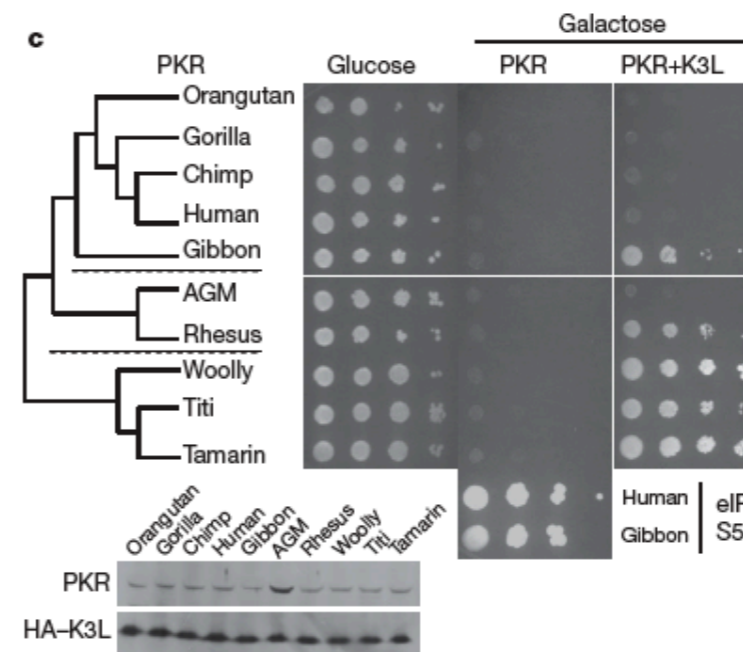
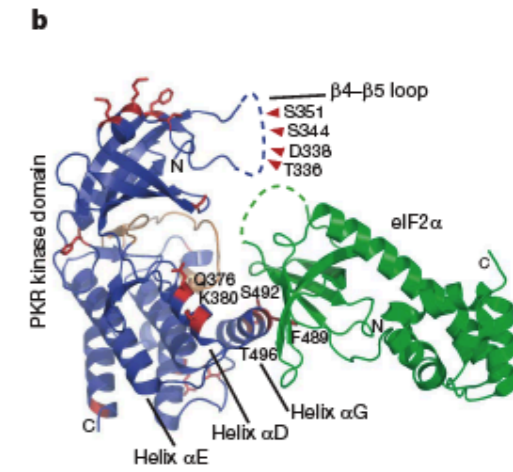
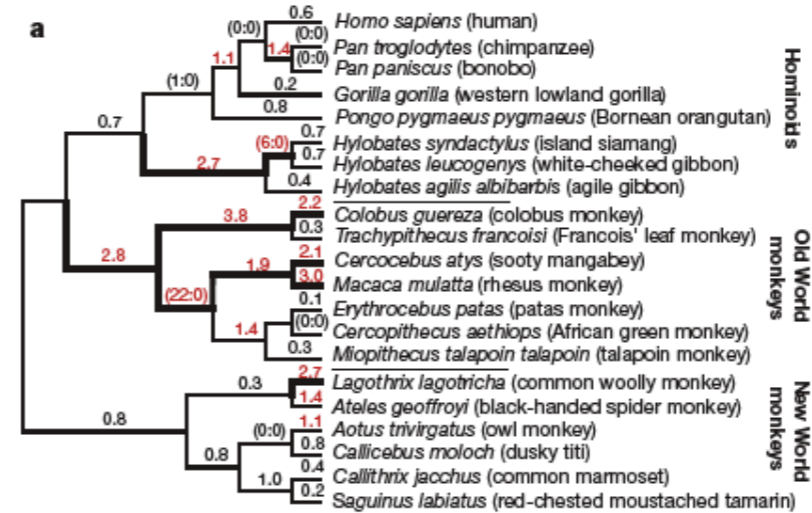
LETTERS



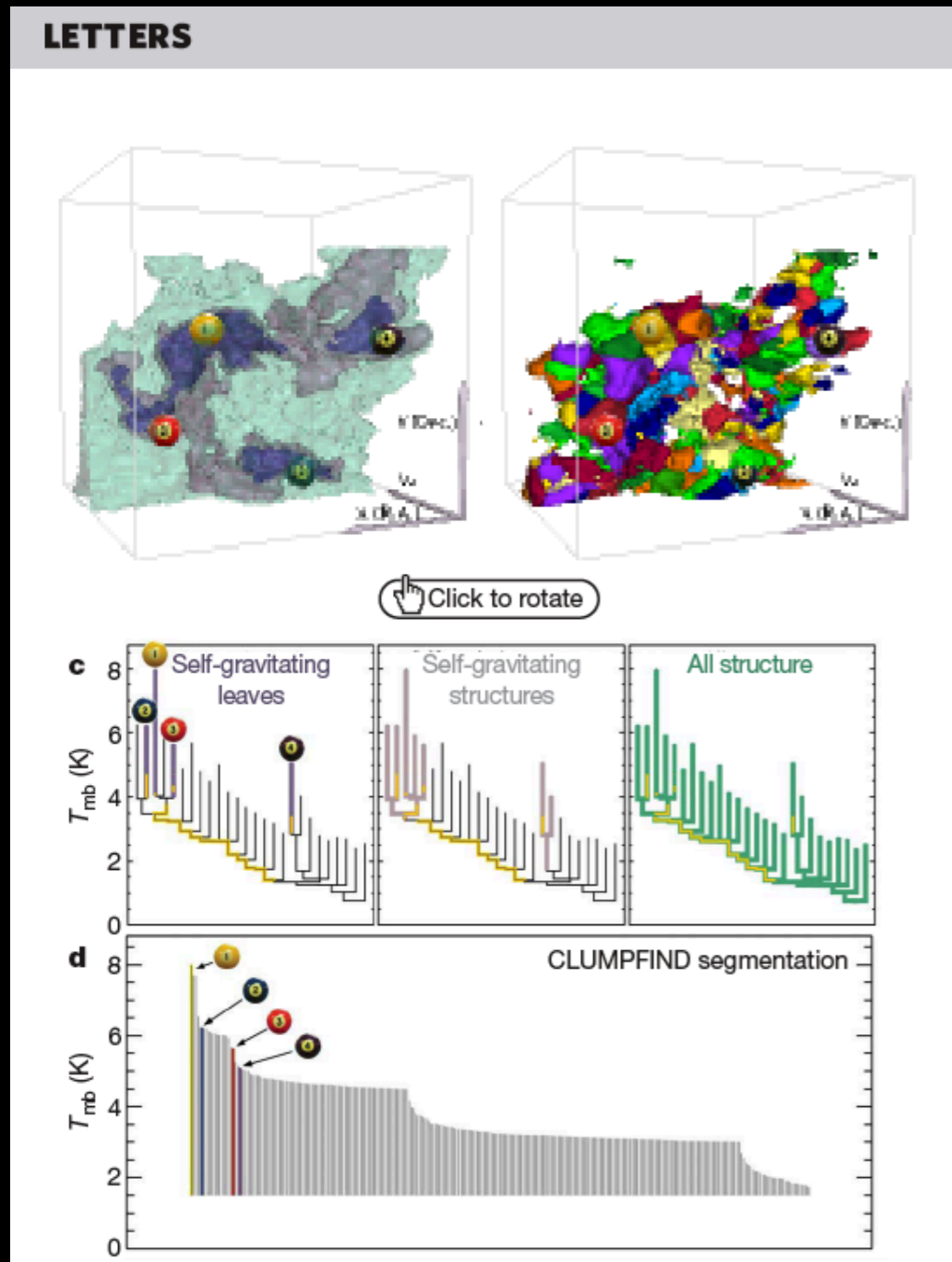
Click to rotate



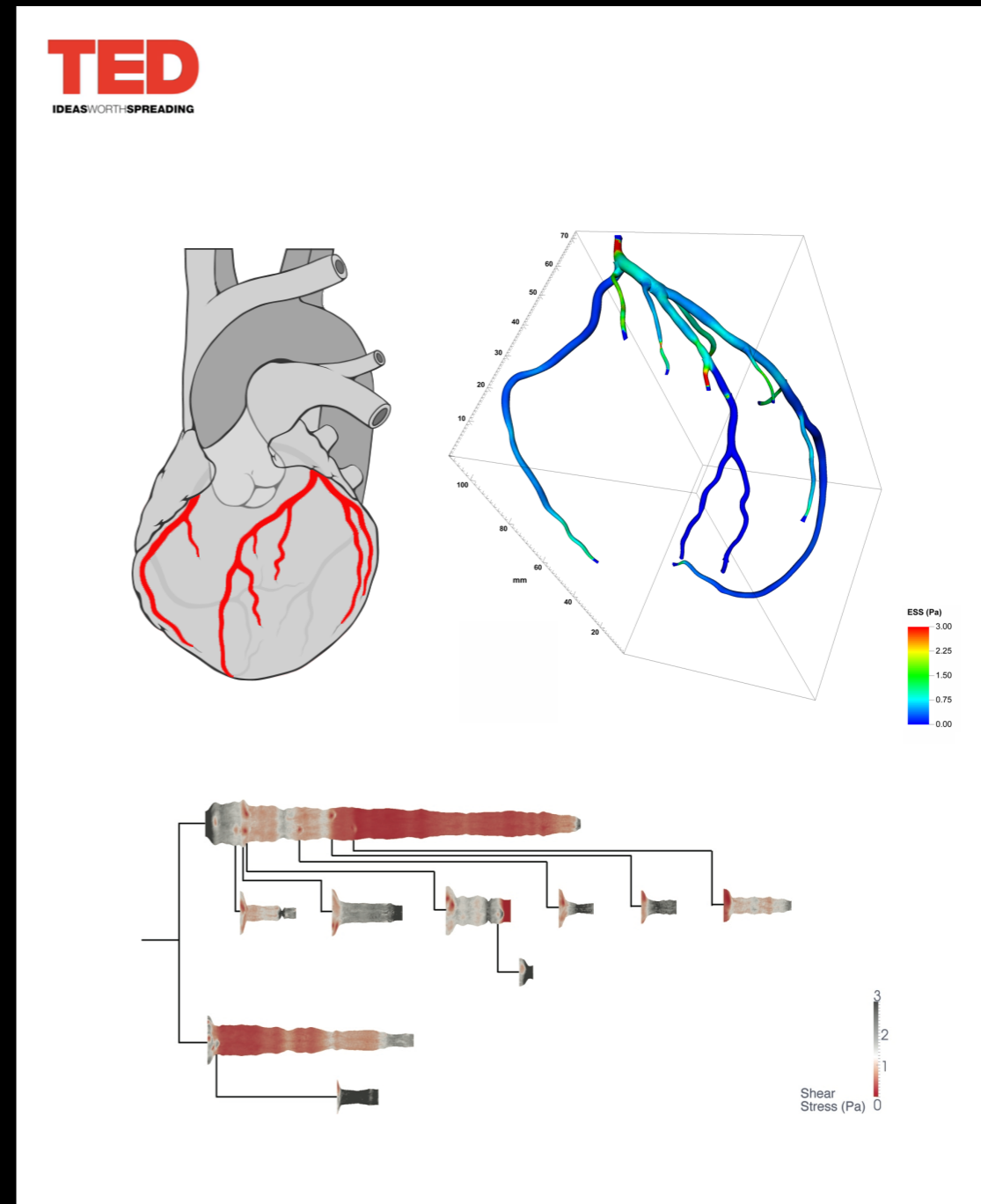
LETTERS



"High-dimensional" or "Multivariate" Data (Astronomy=Medicine)

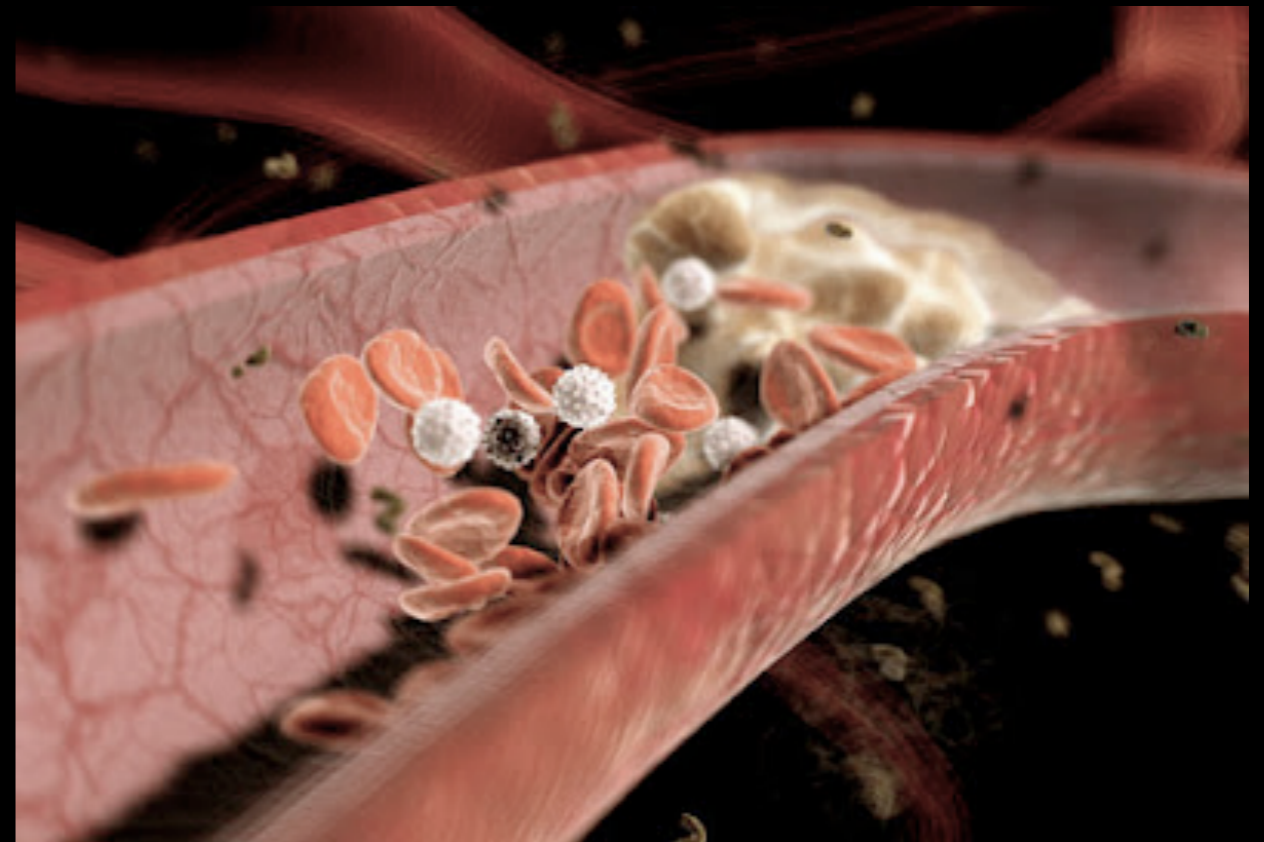


Goodman et al. *Nature*, 2009



Borkin et al. *IEEE Viz*, 2011

Saving Lives



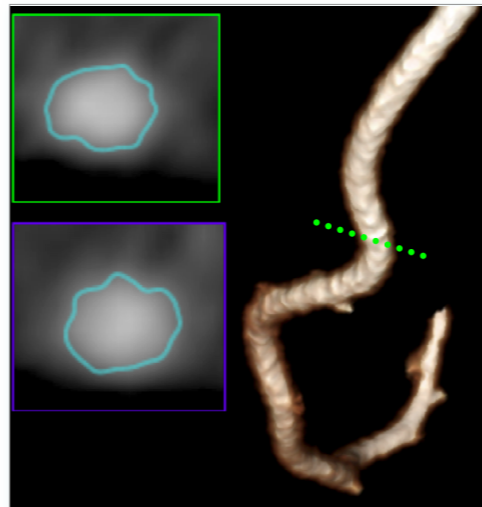
Michelle Borkin

*Harvard School of Engineering & Applied Science Ph.D. student,
supervised by Alyssa Goodman (Astronomer) & Hanspeter Pfister (Computer Scientist)*

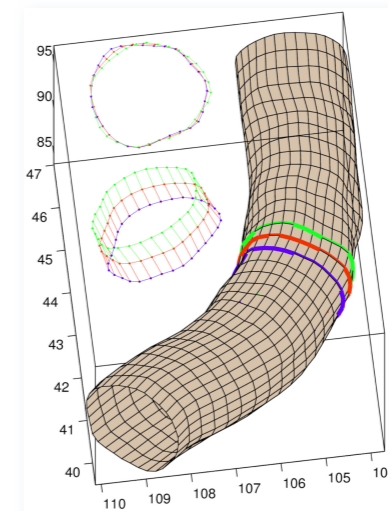
Patients Troubled Hearts, in 3D



Obtain patient CT data



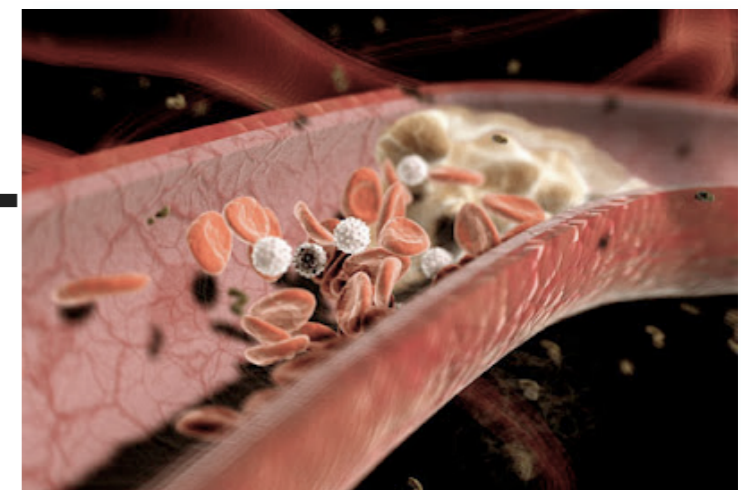
Segment arteries



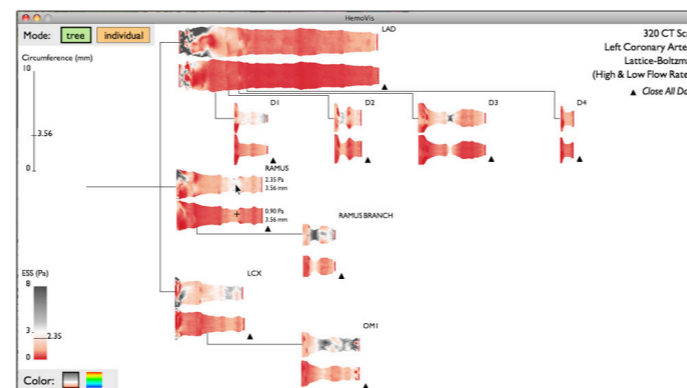
Generate patient geometries



Patient specific flow simulation



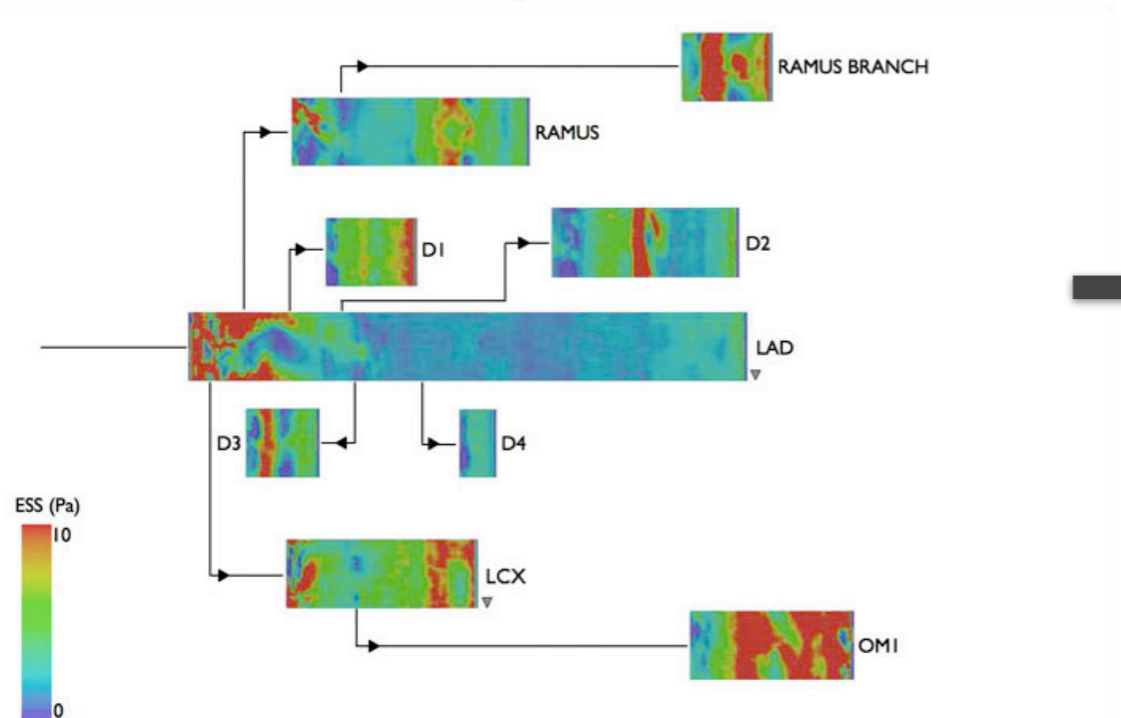
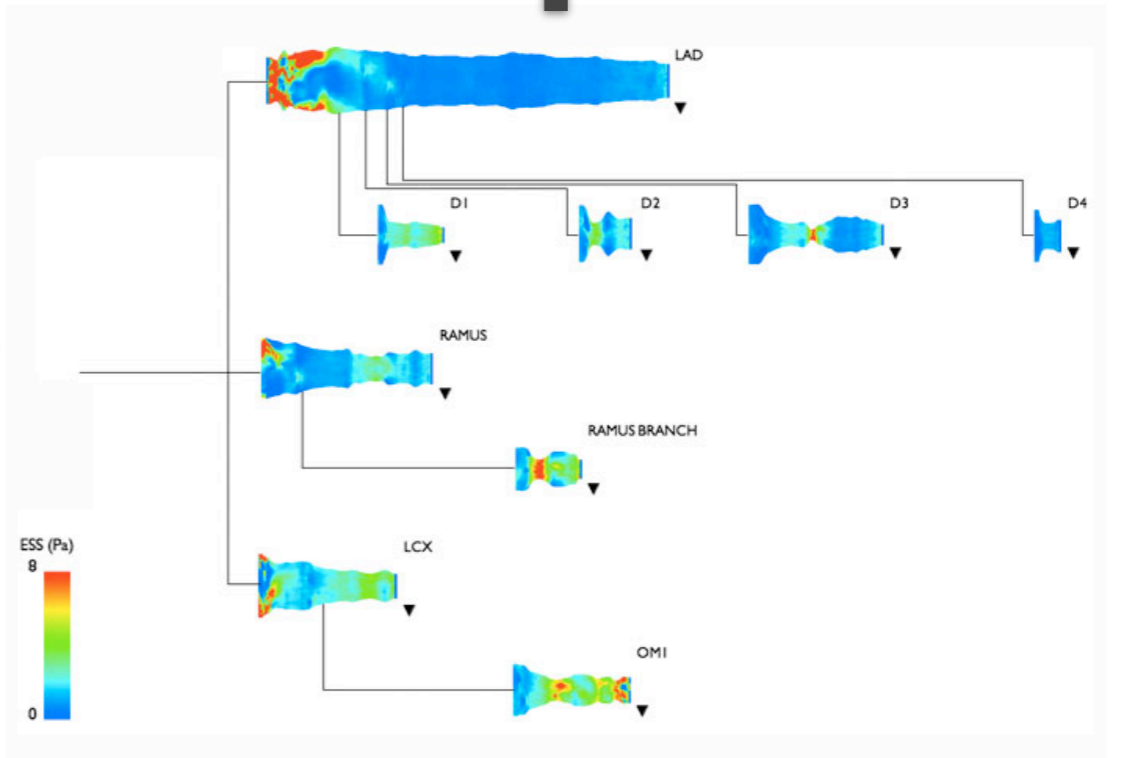
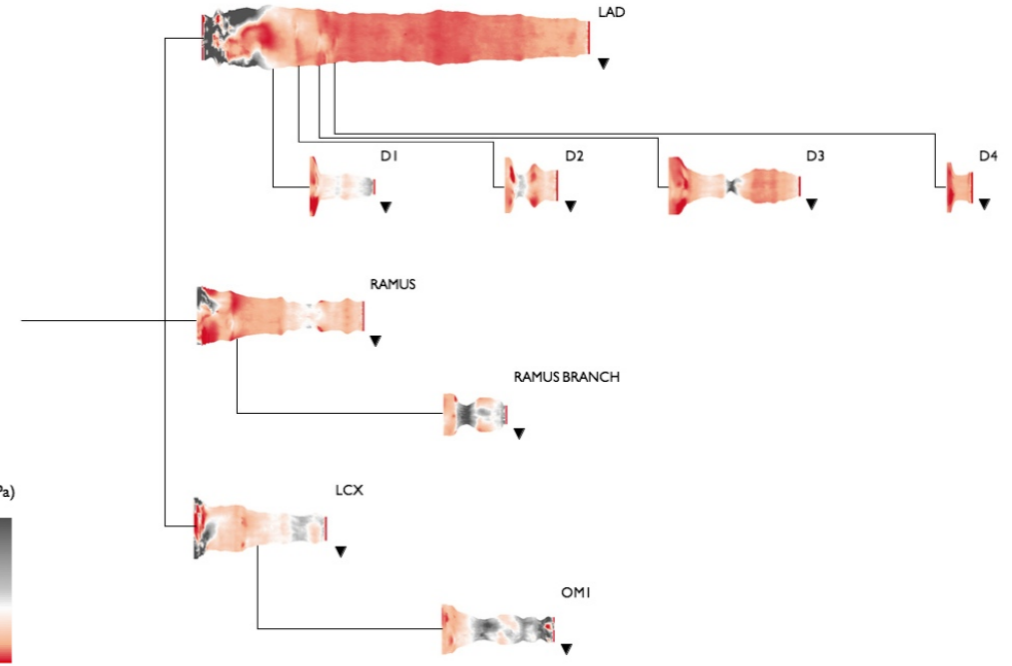
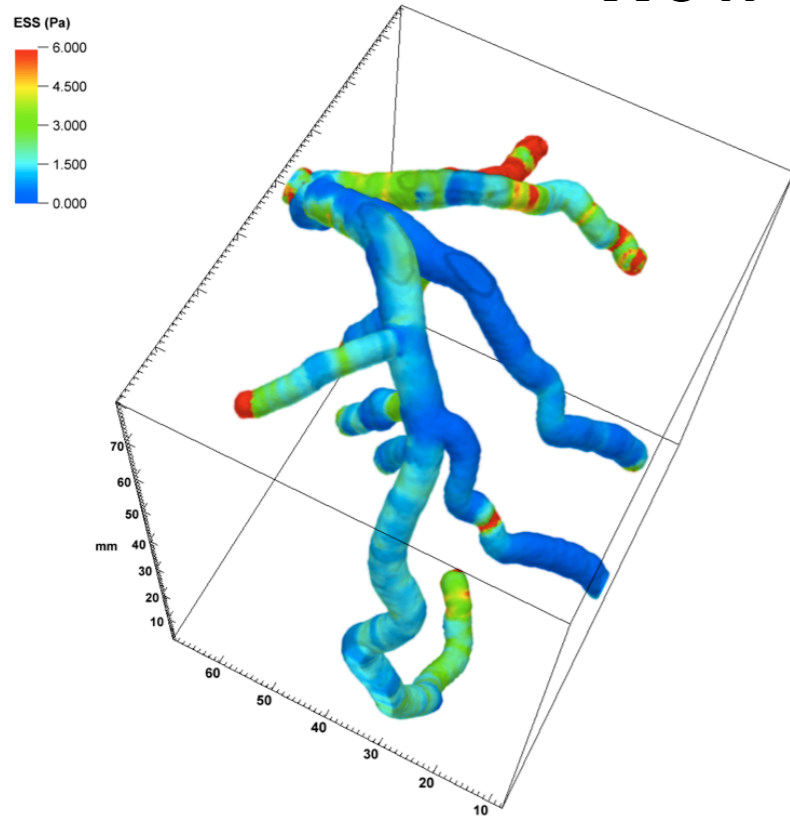
Visualize/analyze data



Clinical decision



How much does viz matter?



ACCURACY

Strong effect of **dimensionality** on accuracy

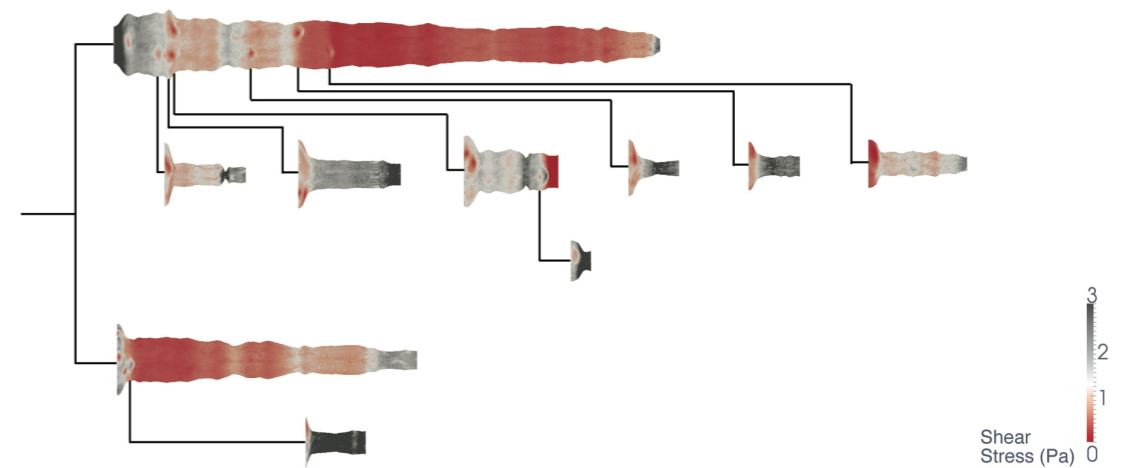
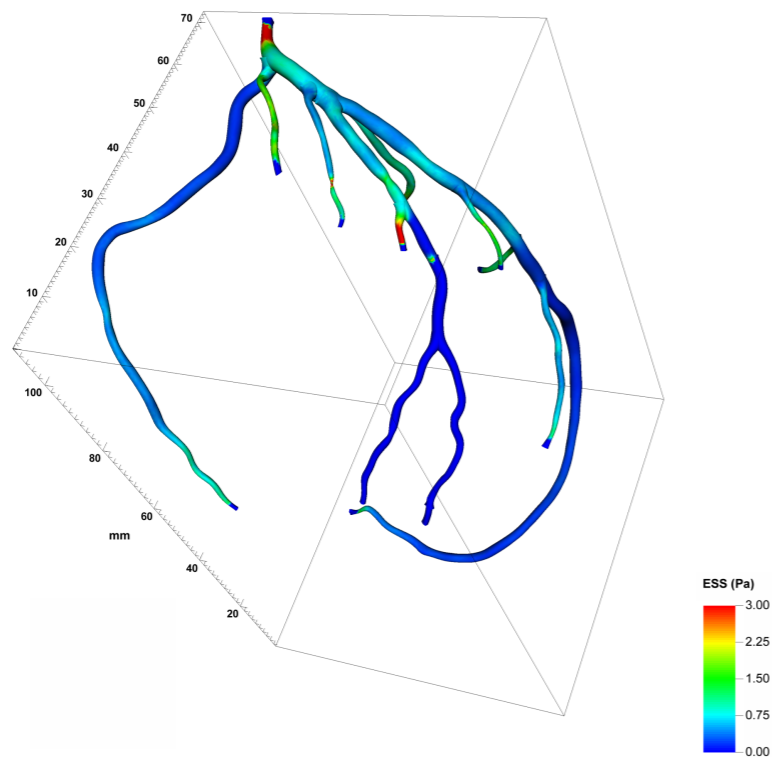
And strong effect of **color**...

39%

percent low ESS regions found

62%

91%



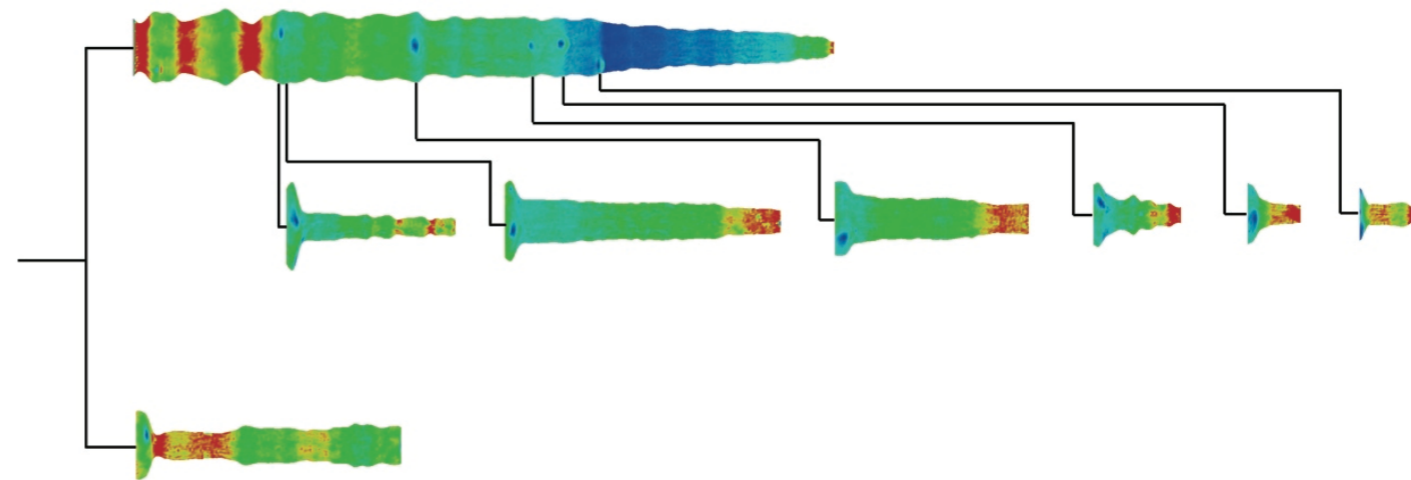
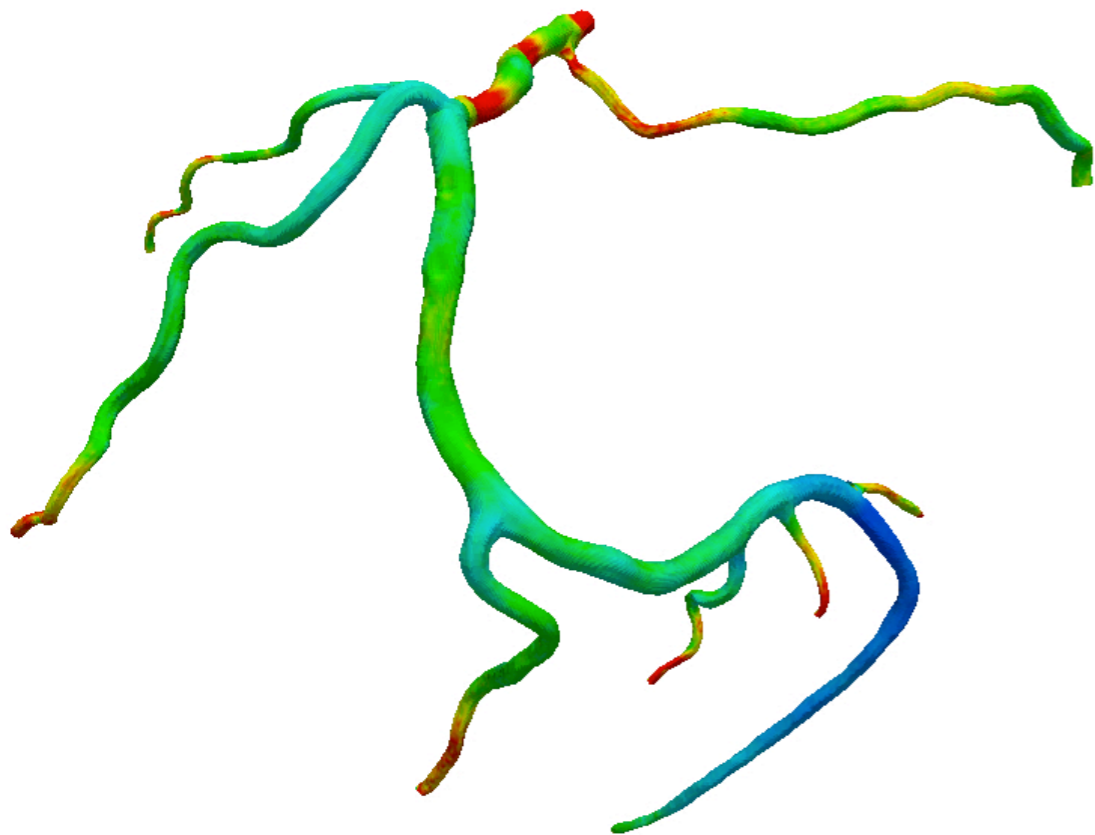
EFFICIENCY

Participants more **efficient** in **2D**.

Rainbow color map has greater detriment in 3D.

10.2 sec/region
5.6 sec/region

2.6 sec/region
2.4 sec/region

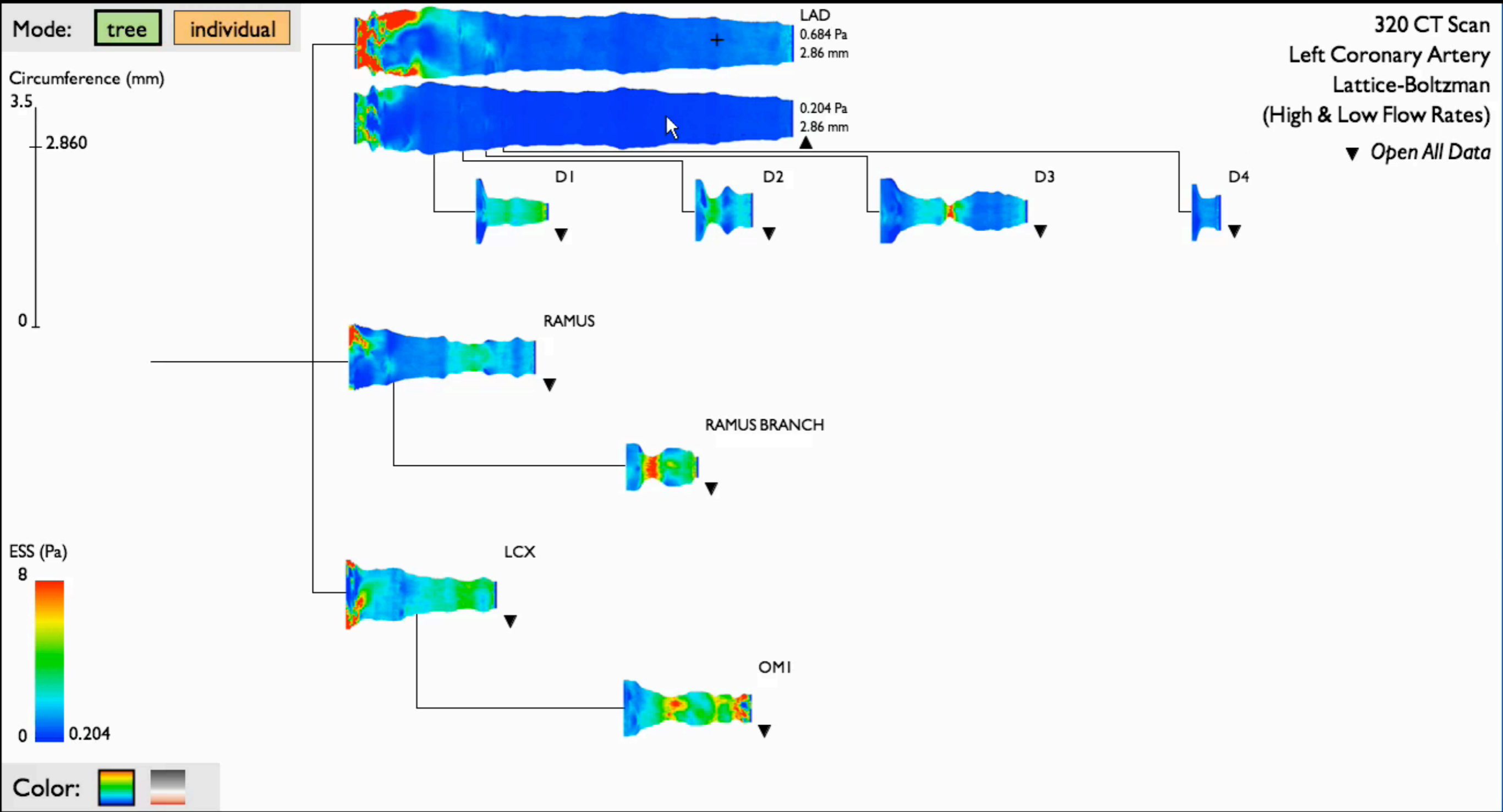


BUT—3D still essential for surgical planning.

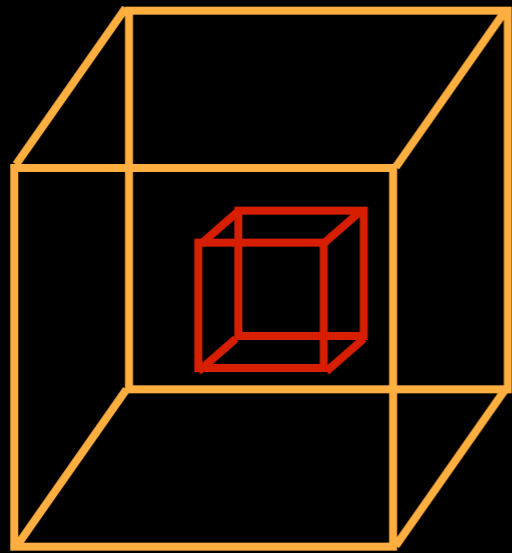
*Contextual,
High-Dimensional
View*

*Interactive
Link*

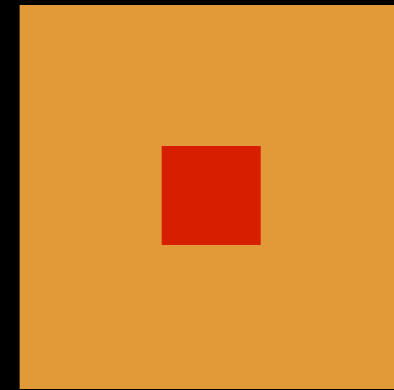
*Flat,
(Text-Based)
View*



"Linked Views"

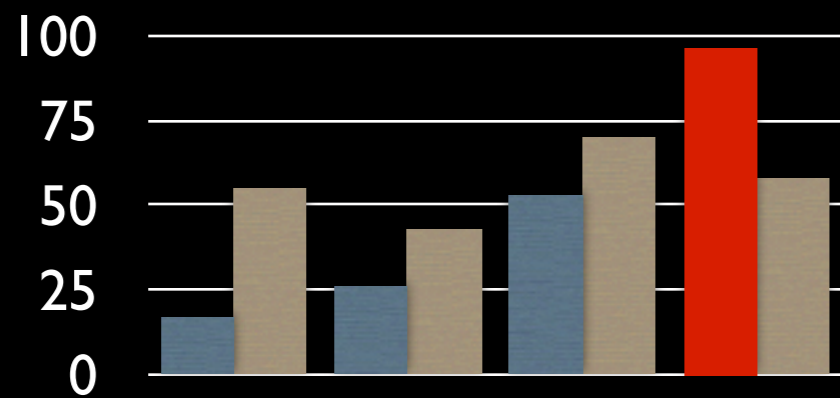


3D

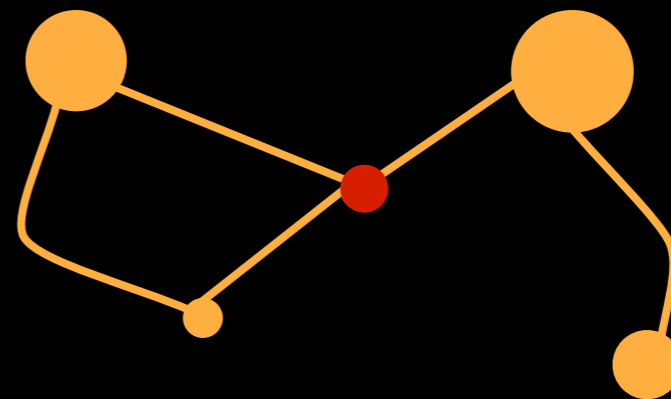


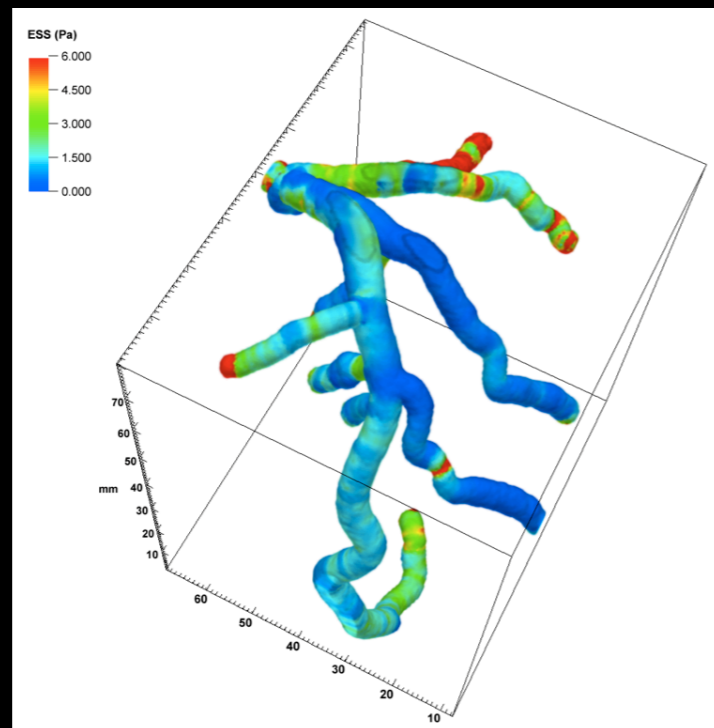
2D

Statistics

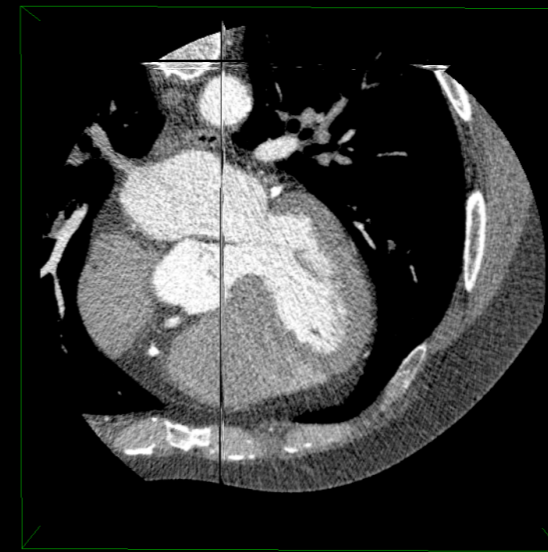


Data Abstraction





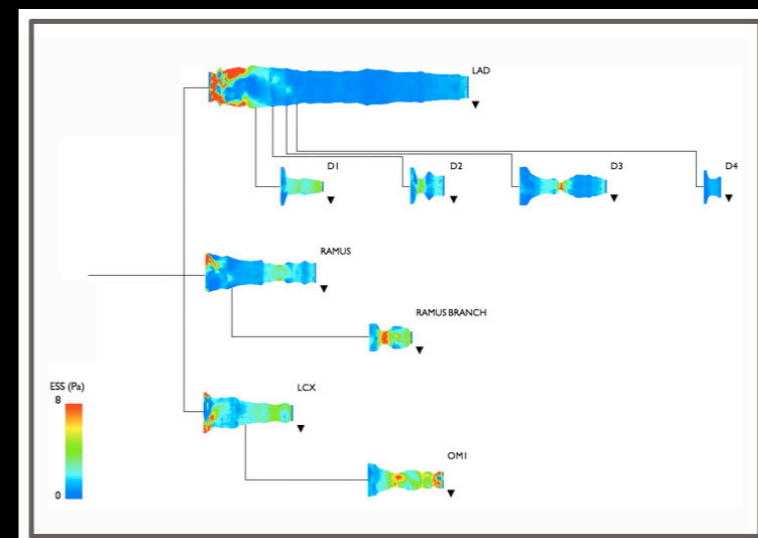
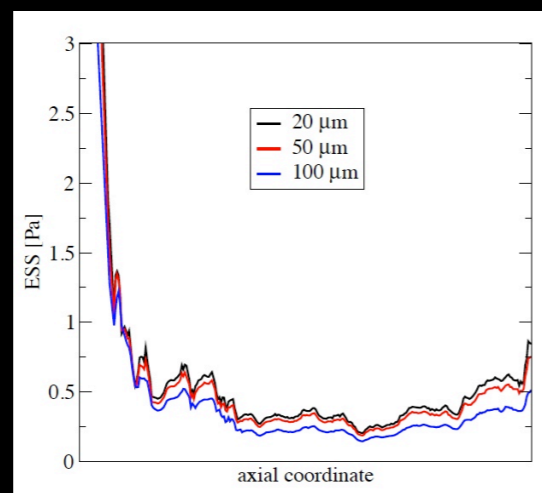
3D

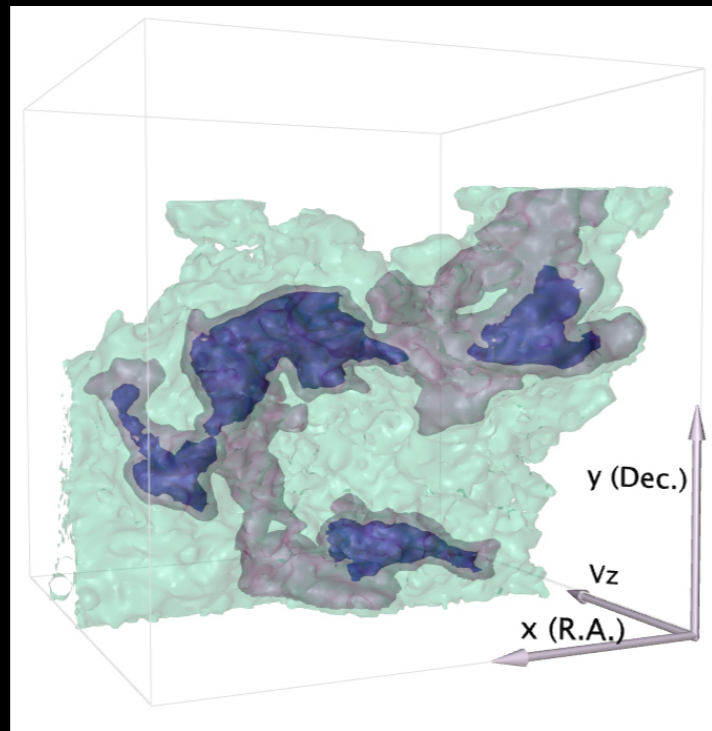


2D

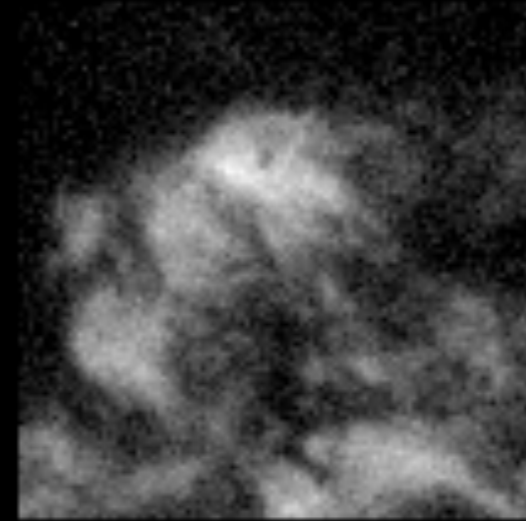
Data Abstraction

Statistics





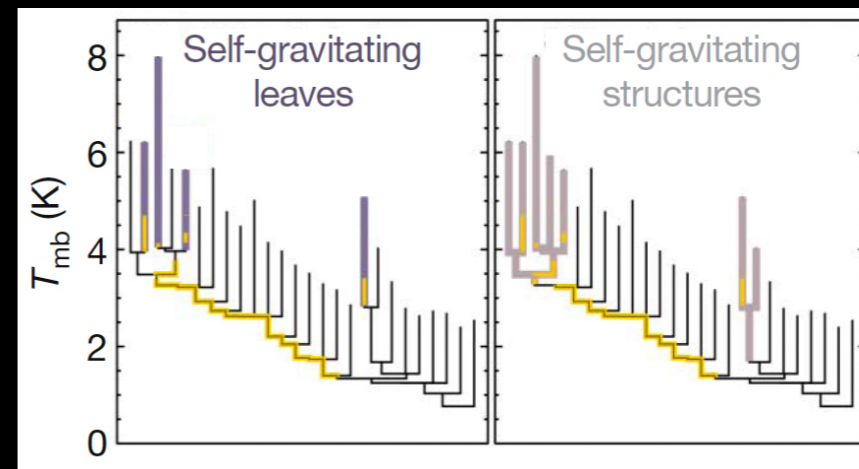
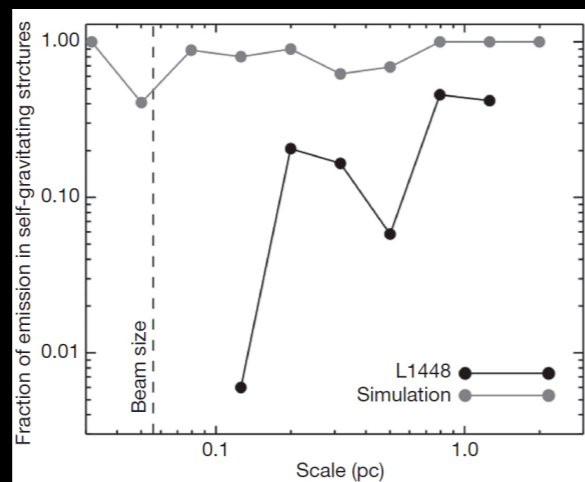
3D



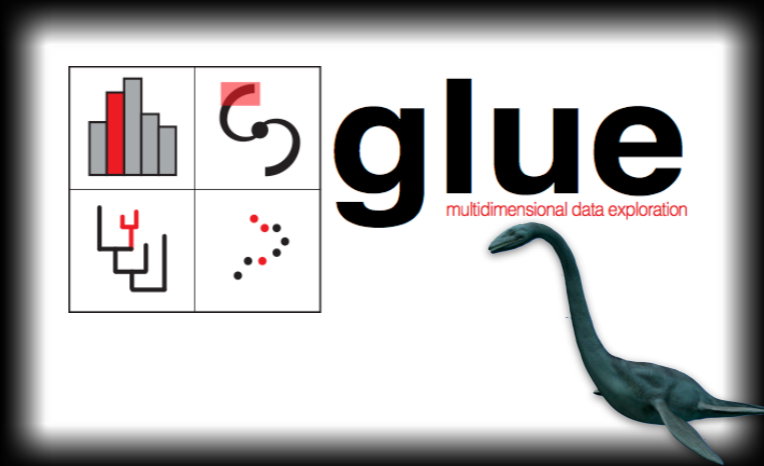
2D

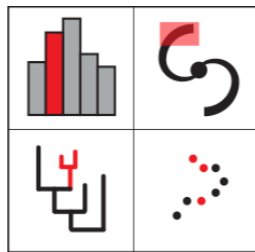
Data Abstraction

Statistics



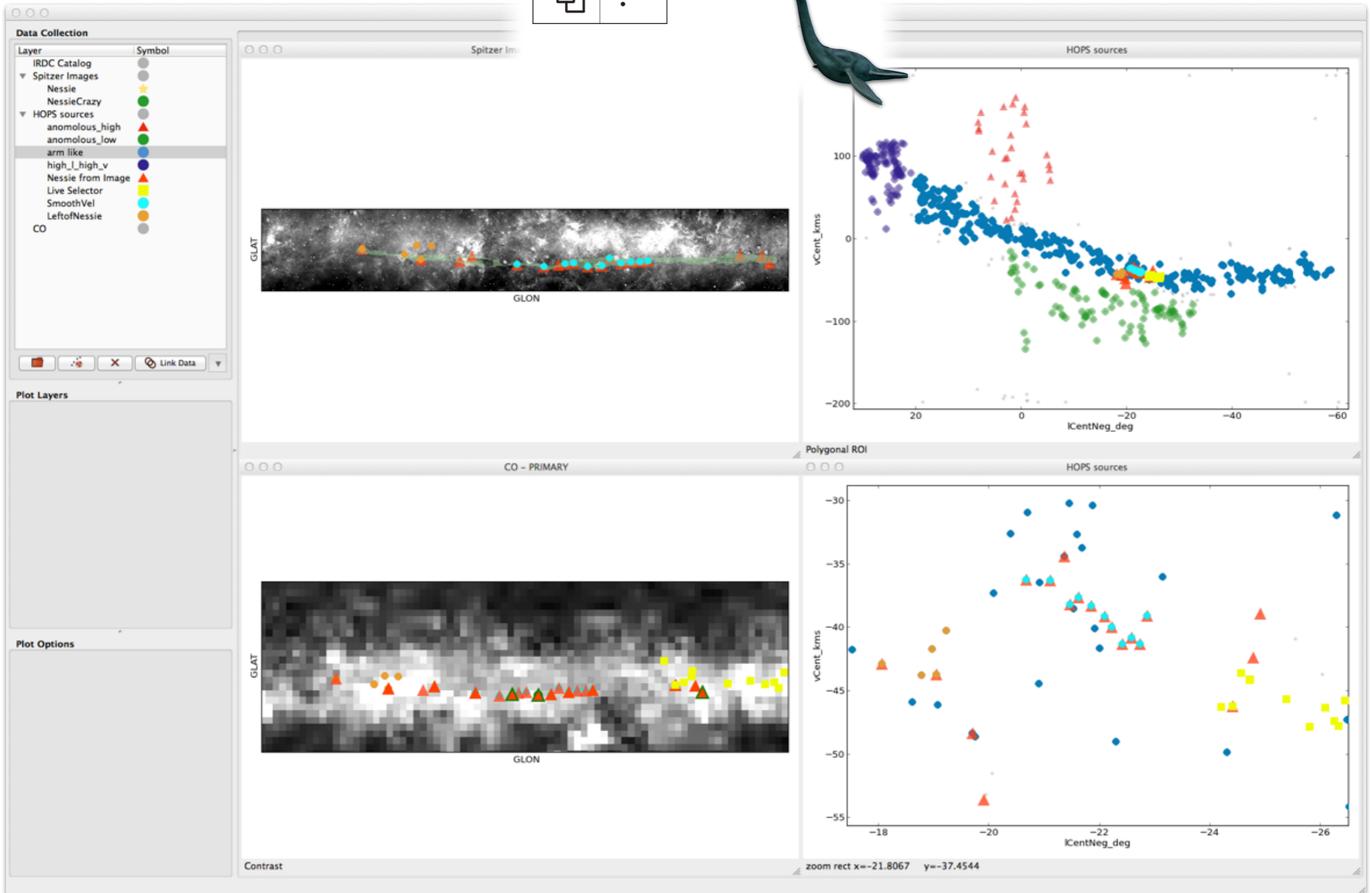
How?





glue

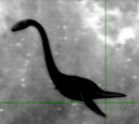
multidimensional data exploration



Glue collaboration: **Beaumont**, Borkin, Goodman, Pfister, Robitaille

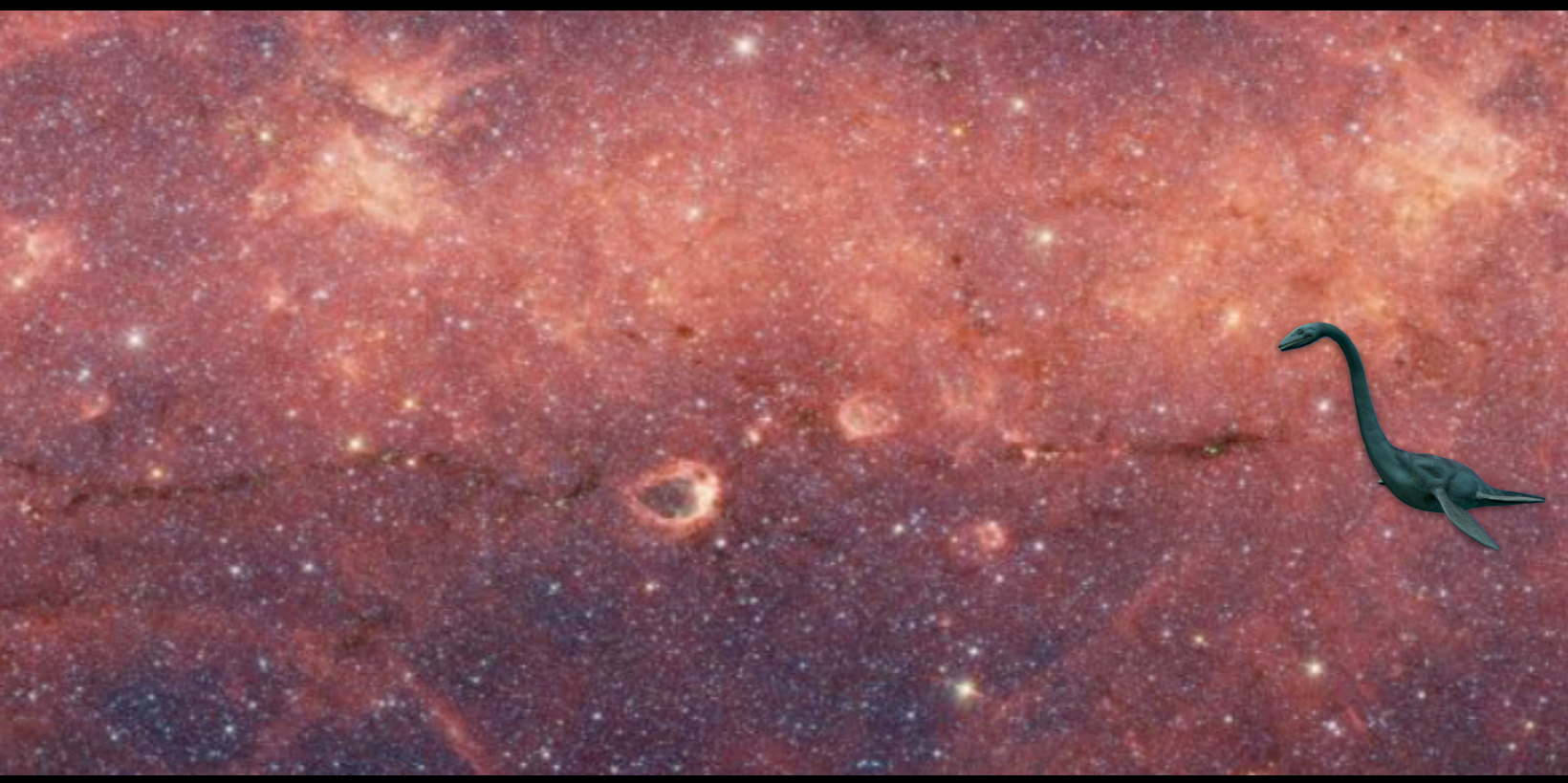


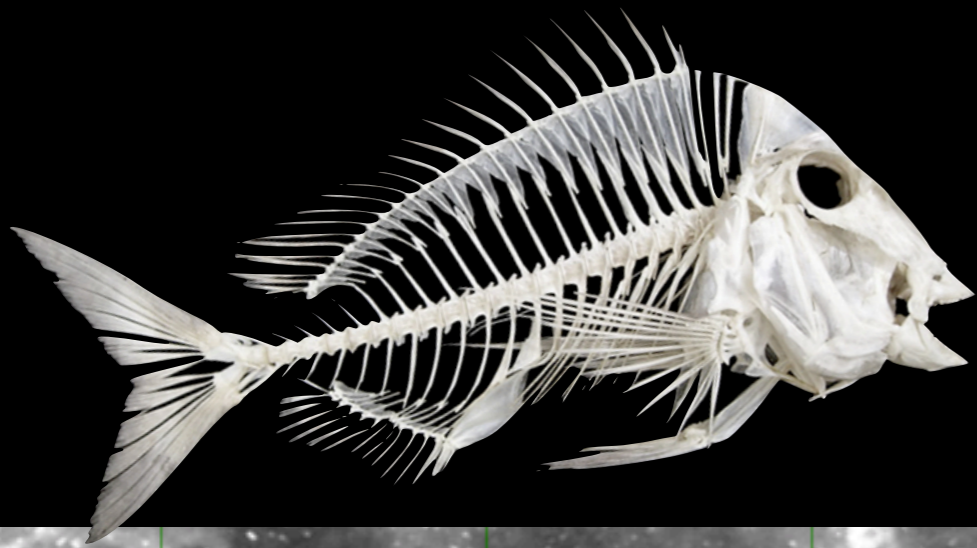
The Bones of the Milky Way



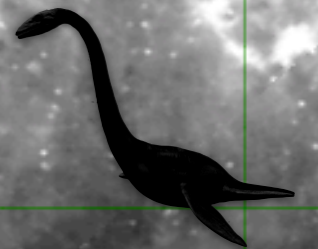
Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics
with Joao F. Alves, Christopher Beaumont, Robert A. Benjamin, Michelle A. Borkin,
Andreas Burkert, Thomas M. Dame, Jens Kauffmann & Thomas Robitaille

fish skeleton images from
www.helterskeletons.com





The Bones of the Milky Way



Alyssa A. Goodman

Harvard-Smithsonian Center for Astrophysics

with Joao F. Alves, Christopher Beaumont, Robert A. Benjamin, Michelle A. Borkin,
Andreas Burkert, Thomas M. Dame, Jens Kauffmann & Thomas Robitaille



IC342



image from Jarrett et al. 2012; WISE Enhanced Resolution Galaxy Atlas

Once upon a time in an enchanted castle by a lake, a sea monster...





...named “Nessie” ...

THE ASTROPHYSICAL JOURNAL LETTERS, 719:L185–L189, 2010 August 20

doi:10.1088/2041-8205/719/2/L185

© 2010. The American Astronomical Society. All rights reserved. Printed in the U.S.A.

THE “NESSIE” NEBULA: CLUSTER FORMATION IN A FILAMENTARY INFRARED DARK CLOUD

JAMES M. JACKSON¹, SUSANNA C. FINN¹, EDWARD T. CHAMBERS², JILL M. RATHBORNE³, AND ROBERT SIMON⁴

¹ Institute for Astrophysical Research, Boston University, Boston, MA 02215, USA; jackson@bu.edu, sfinn@bu.edu

² Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA; e-chambers@northwestern.edu

³ Australia Telescope National Facility and Universidad de Chile, Santiago, Chile; rathborn@das.uchile.cl

⁴ I. Physikalisches Institut, Universität zu Köln, 50937 Köln, Germany; simonr@ph1.uni-koeln.de

Received 2010 April 13; accepted 2010 July 21; published 2010 August 3

ABSTRACT

The “Nessie” Nebula is a filamentary infrared dark cloud (IRDC) with a large aspect ratio of over 150:1 ($1^{\circ}5 \times 0^{\circ}01$ or $80 \text{ pc} \times 0.5 \text{ pc}$ at a kinematic distance of 3.1 kpc). Maps of HNC (1–0) emission, a tracer of dense molecular gas, made with the Australia Telescope National Facility Mopra telescope, show an excellent morphological match to the mid-IR extinction. Moreover, because the molecular line emission from the entire nebula has the same radial velocity to within $\pm 3.4 \text{ km s}^{-1}$, the nebula is a single, coherent cloud and not the chance alignment of multiple unrelated clouds along the line of sight. The Nessie Nebula contains a number of compact, dense molecular cores which have a characteristic projected spacing of $\sim 4.5 \text{ pc}$ along the filament. The theory of gravitationally bound gaseous cylinders predicts the existence of such cores, which, due to the “sausage” or “varicose” fluid instability, fragment from the cylinder at a characteristic length scale. If turbulent pressure dominates over thermal pressure in Nessie, then the observed core spacing matches theoretical predictions. We speculate that the formation of high-mass stars and massive star clusters arises from the fragmentation of filamentary IRDCs caused by the “sausage” fluid instability that leads to the formation of massive, dense molecular cores. The filamentary molecular gas clouds often found near high-mass star-forming regions (e.g., Orion, NGC 6334, etc.) may represent a later stage of IRDC evolution.

Key words: ISM: clouds – stars: formation

Jackson et al. 2010

Ringberg Castle, Bavaria
“Early Phases of Star Formation”
July 2012



QUESTION *Andi Burkert*: Is Nessie “parallel to the Galactic Plane”?

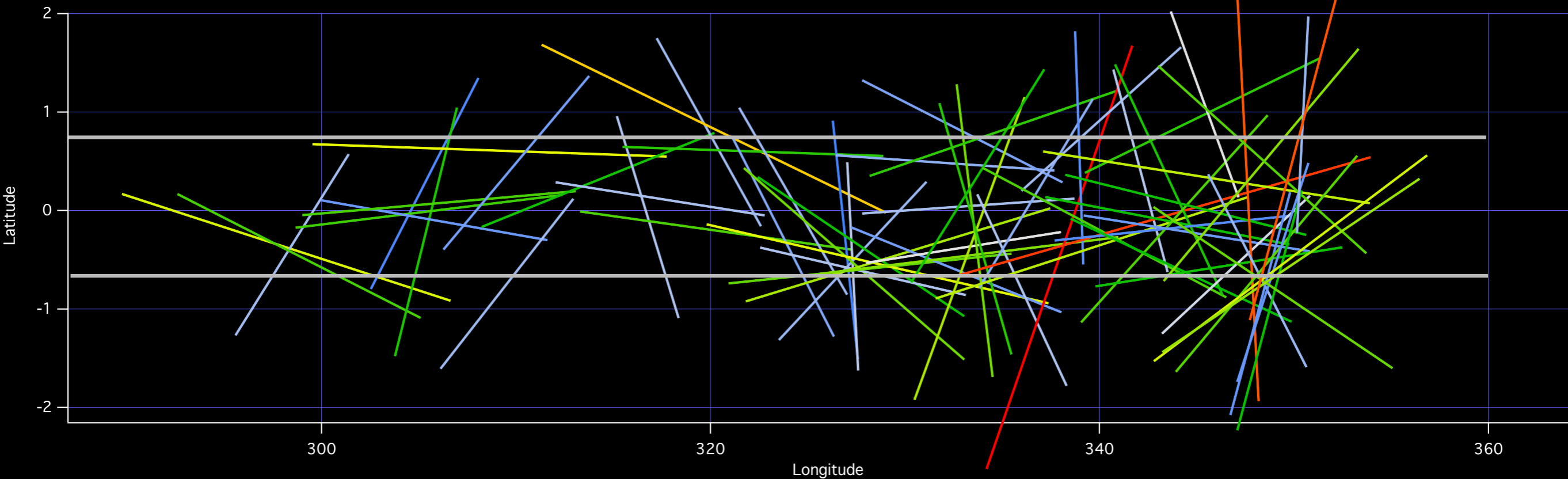
ANSWER *no one* immediately knew the answer!

AG decides to look into this and...

Quick Answer: Yes, looks like it is...

The image shows a screenshot of the Glimpse/MIPSGAL Viewer interface. At the top left, there is a logo and the text "GLIMPSE / MIPSGAL VIEWER". To the right of this are three buttons: "LINK TO CURRENT VIEW", "TOGGLE PINS", and "QUESTIONS?". Below these buttons is a horizontal bar with a green gradient and a white slider. The main area of the interface is a large, rounded rectangle containing a star field image. The stars are predominantly red and orange, with some blue and white stars. A blue dinosaur-like creature is overlaid on the right side of the star field. In the top right corner of the star field, there is a button labeled "COORDINATES". At the bottom of the interface, there is a control bar with a question mark icon, a slider labeled "IRAC" with a white marker, a label "IRAC/MIPS", another question mark icon, and a set of navigation icons including a plus/minus sign, left/right arrows, up/down arrows, and a refresh icon.

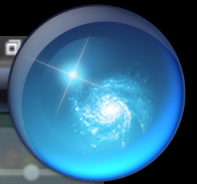
Are others too?



filtering out only “long” Peretto & Fuller clouds, and showing their orientation (color~length)

Huh? Let's look...

WWT



The screenshot displays the Microsoft WorldWide Telescope (WWT) interface. At the top, there is a menu bar with options: Explore, Guided Tours, Search, Community, Telescope, View, Settings, and a window titled 'hops test'. Below the menu bar is a toolbar with a play button, a 'Run Time' indicator (0:30), and three slide thumbnails labeled '3D Milky Way', 'With beer', and 'No beer'. The 'No beer' slide is currently selected. To the right of the toolbar are controls for 'Tour Properties', 'Save', 'Music', and 'Voiceover'. The main viewing area shows a vibrant, multi-colored star field. On the left side, there is a 'Layers' panel with a tree view containing 'Sun', 'Mercury', 'Venus', 'Earth', 'Mars', 'Jupiter', 'Saturn', 'Uranus', 'Neptune', 'Pluto', 'Sky', 'Overlays', '3d Solar System', 'perettoww/long.xml', 'peretto.xml', 'HOPS', and 'Dome'. Below the layers panel is a table with astronomical data:

Name	Value
Glon	340.2600
Glat	-0.2237
RAJ2000	252.22437
DEJ2000	-45.19194
Seq	10358
Name	340.260-0.223
RAJ2000	16.48.53.85
DEJ2000	-45.11.31.0
lmin	49.8
IMIR	99.2
IMIR	1.70

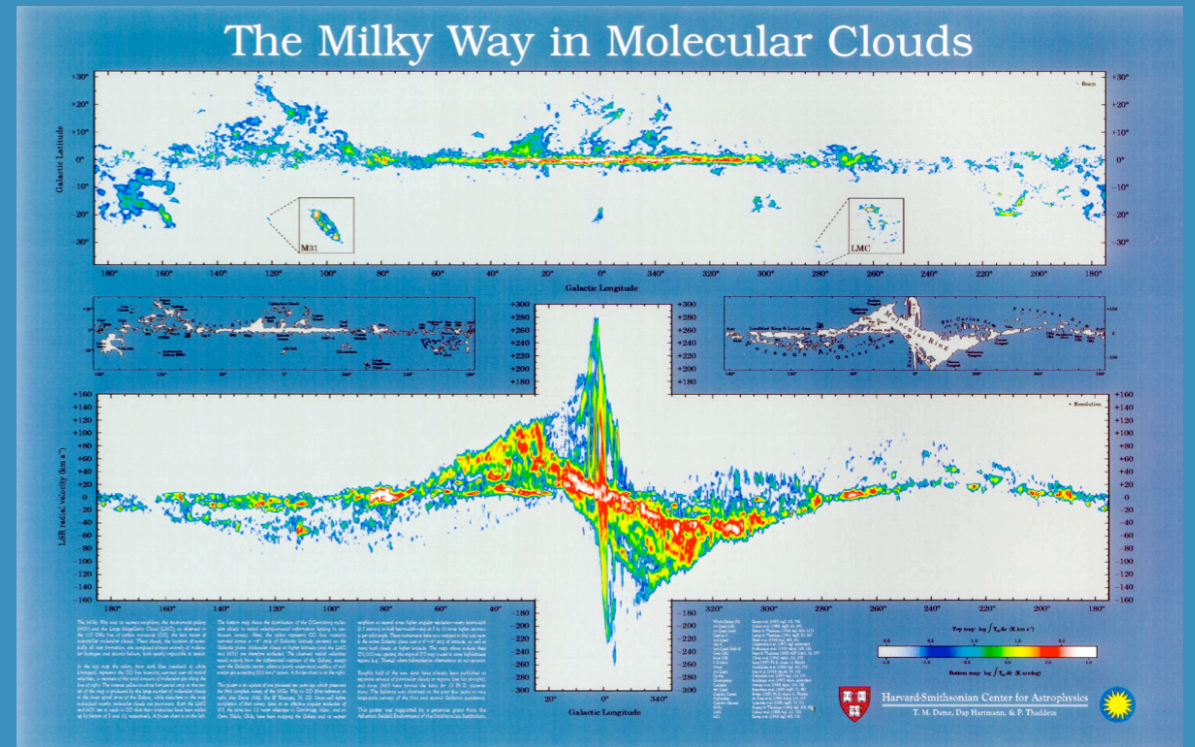
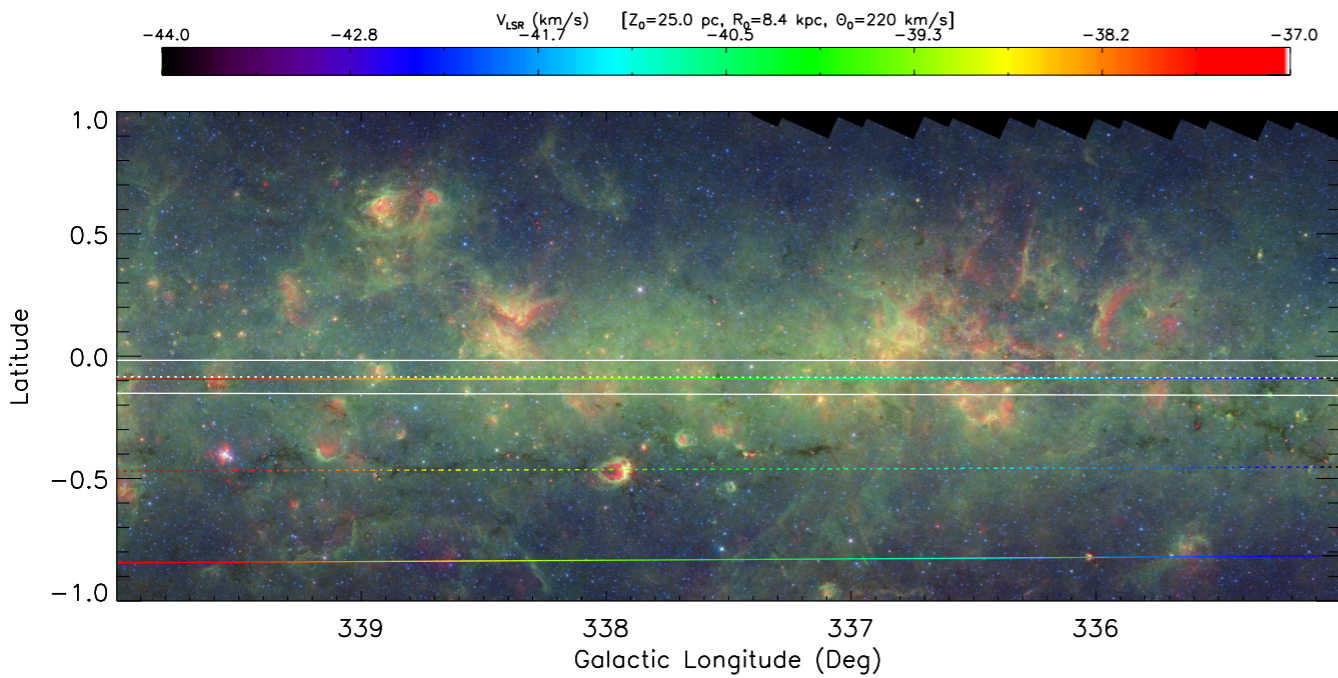
Below the table is a 'Time Scrubber' with a slider and a 'Time Series' section with 'Auto Loop' and 'Delete', 'Add', 'Paste', and 'Reset' buttons. At the bottom, there is a 'Look At' section with 'Sky' selected, an 'Imagery' dropdown set to 'Digitized Sky Survey (Color)', and an 'Image Crossfade' slider. On the right, there is a 'Context Search Filter' set to 'All', a '1 of 1' indicator, a compass, and a small map showing the current location. The bottom right corner displays coordinates: RA: 16h44m02s, Dec: -46.34.15, and a time indicator '01:13:01'.

[Demo: Much more to Nessie than meets the eye!]

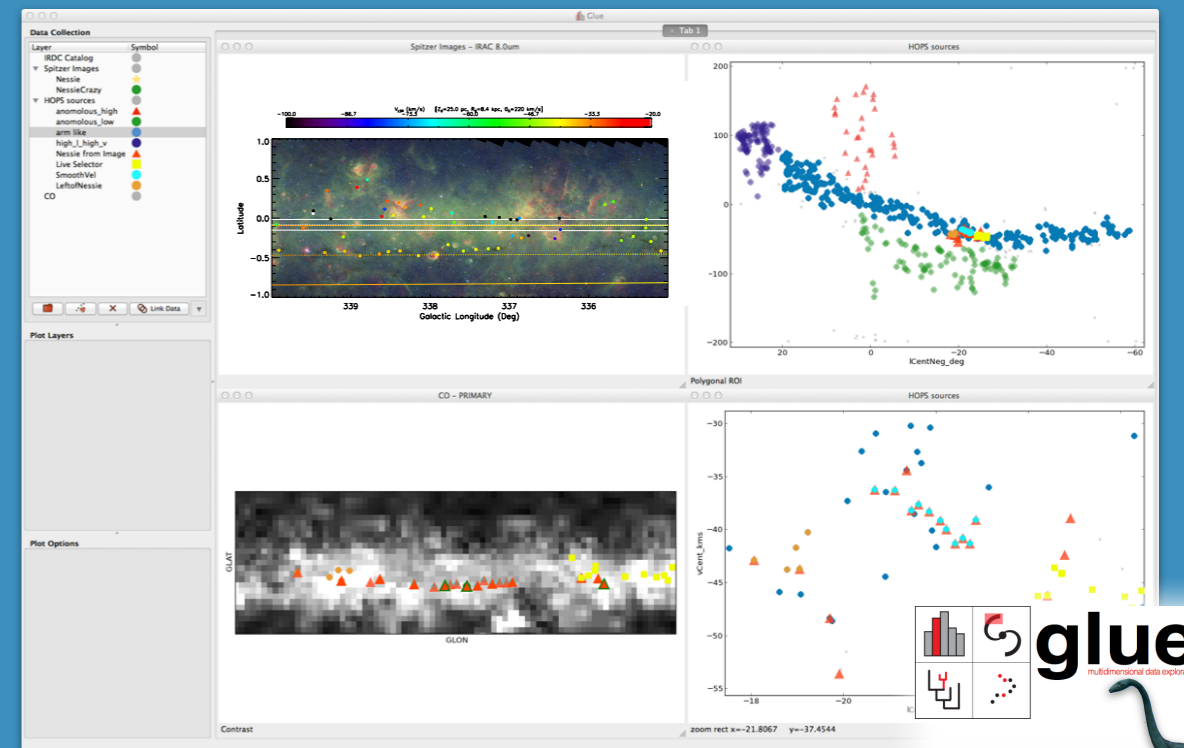
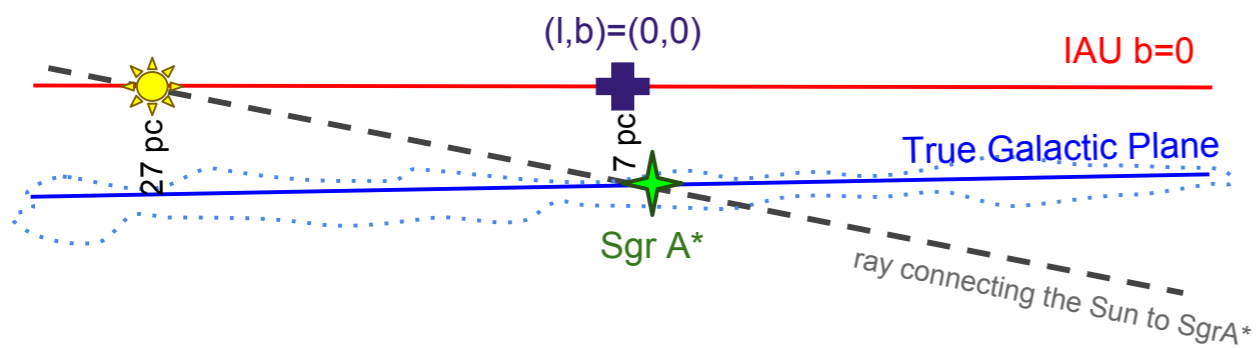
Where is "Nessie," in 3D?

How close to "in" the plane?

At what distance & inclination to l.o.s?



Drawing is schematic--NOT to scale

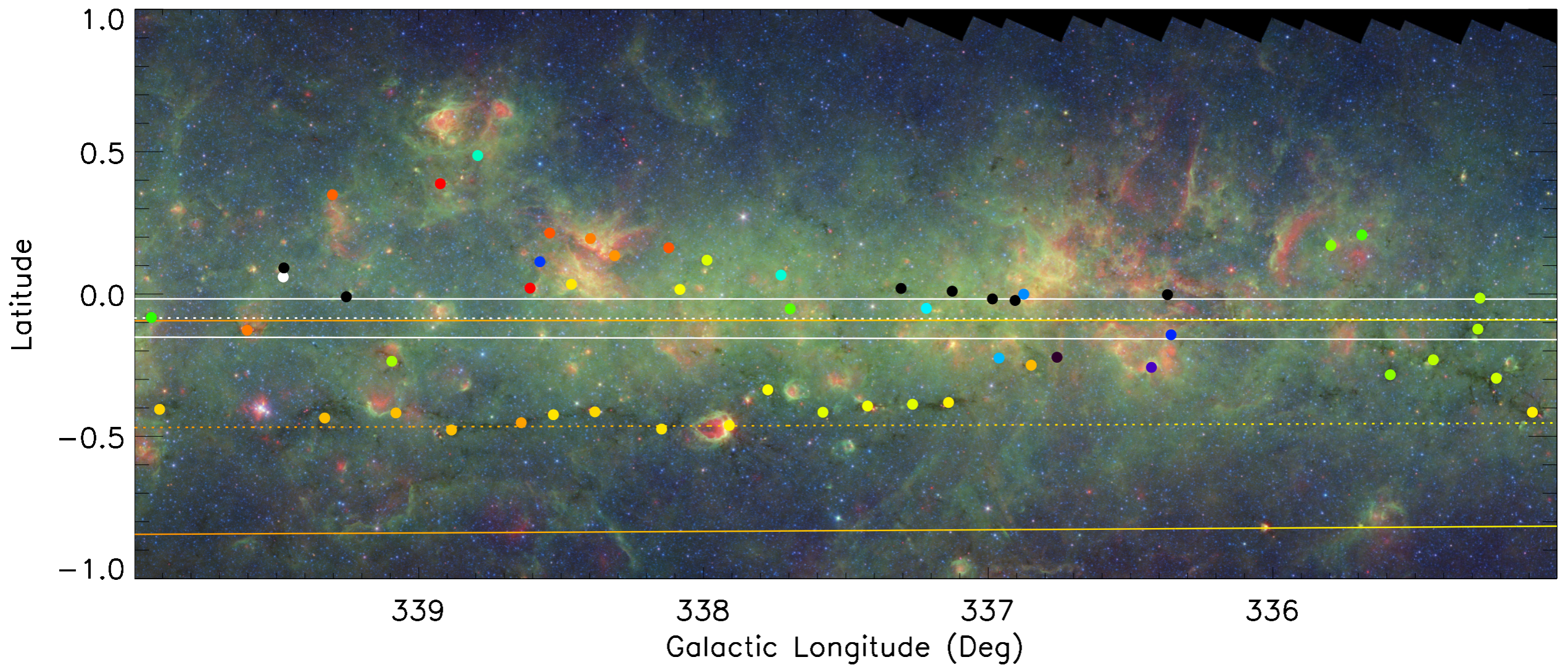
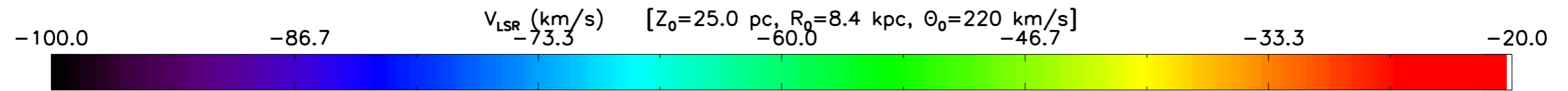


Notes:
 IAU b=0 set from HI, which is uncertain by ~0.1 degrees
 tilt of red w.r.t. blue would be $(20/8400) * 180/\pi = 0.13$ degrees



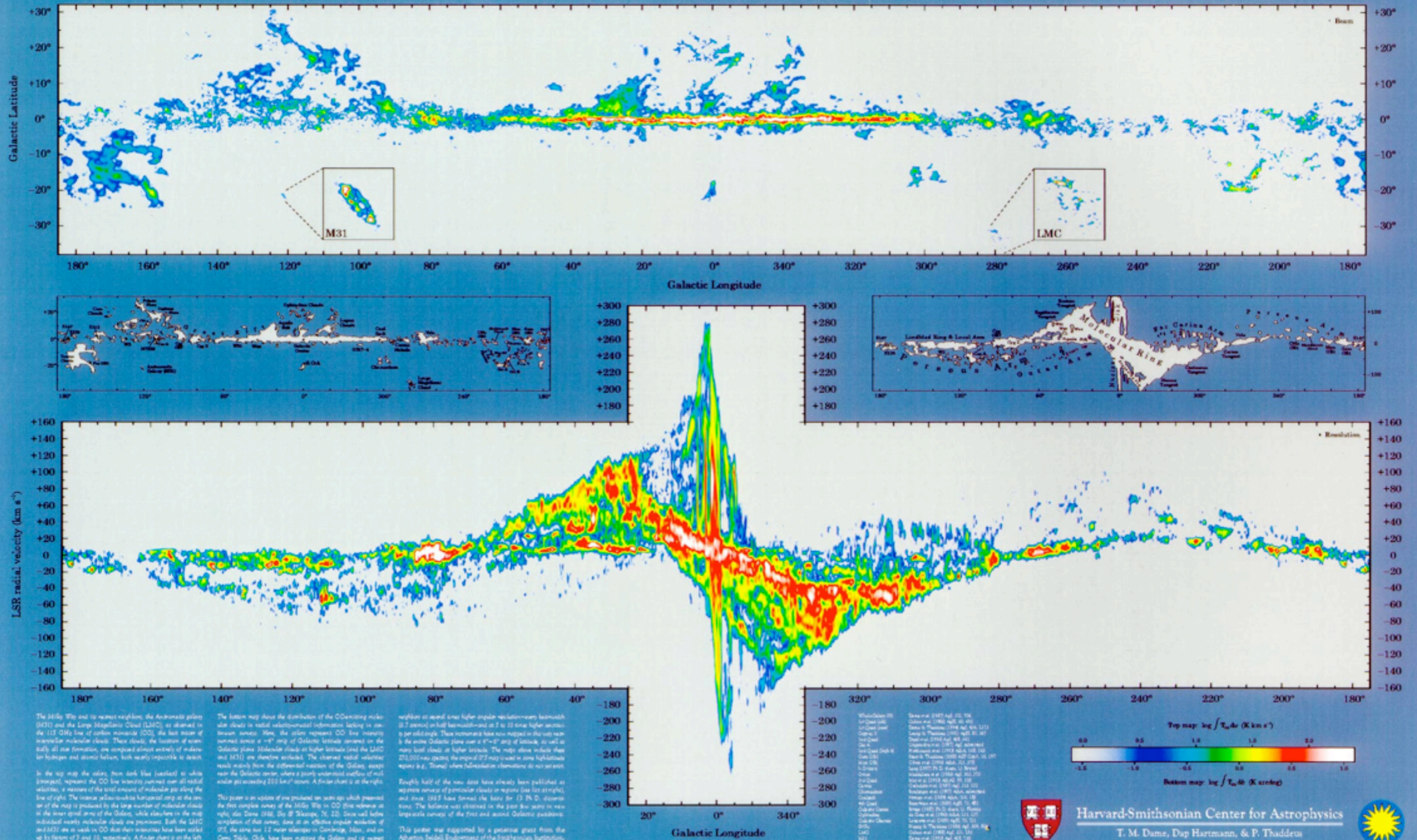
Where is "Nessie," in 3D?

Velocity constraints on distance & inclination to l.o.s



colored dots show HOPS data (dense gas velocities) from Purcell et al. 2012

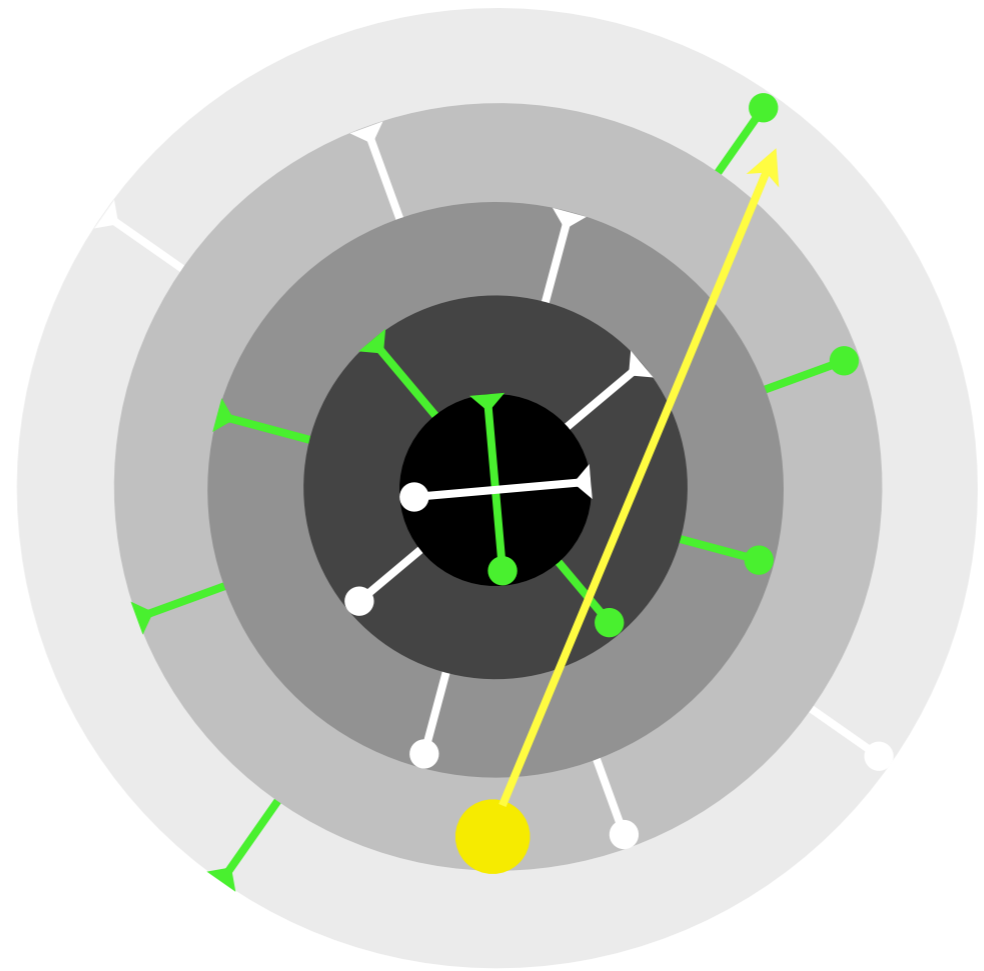
Milky Way Gas Kinematics 101

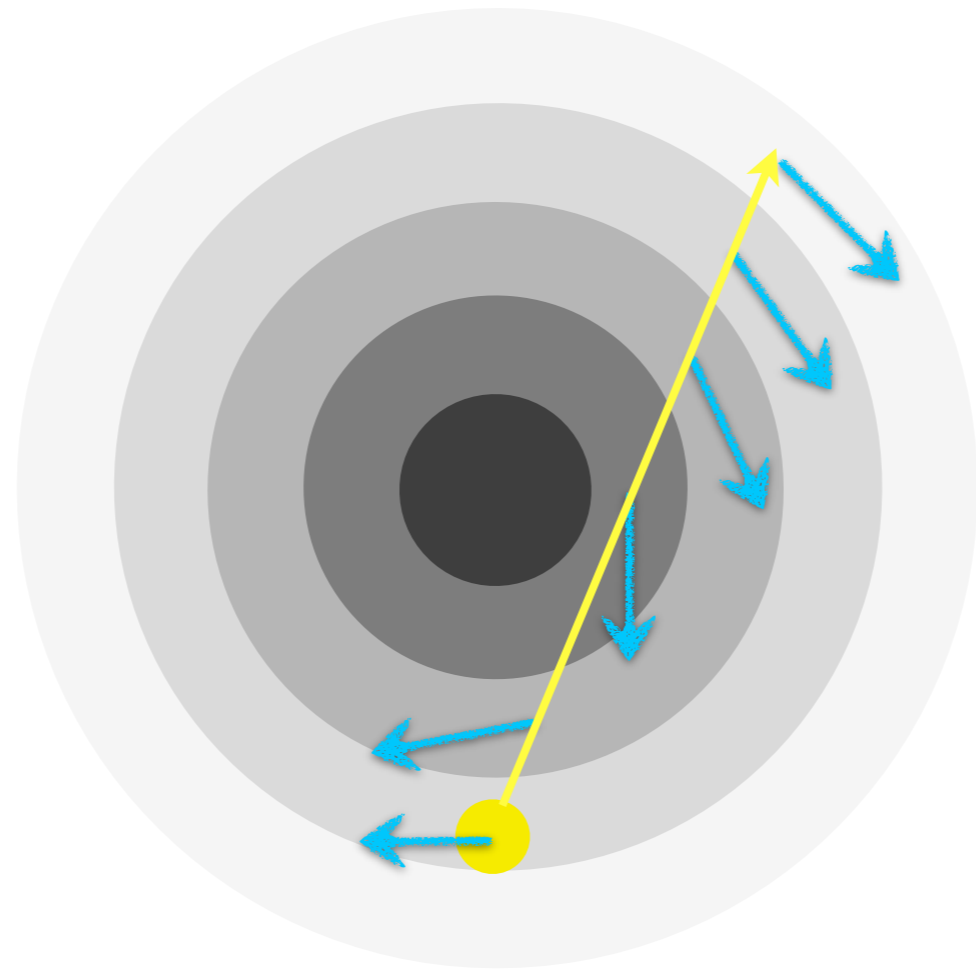


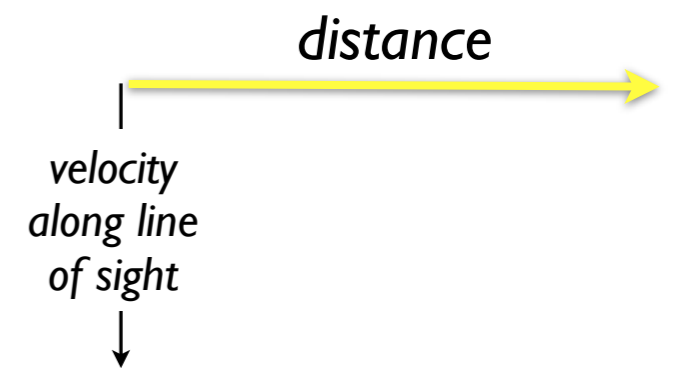
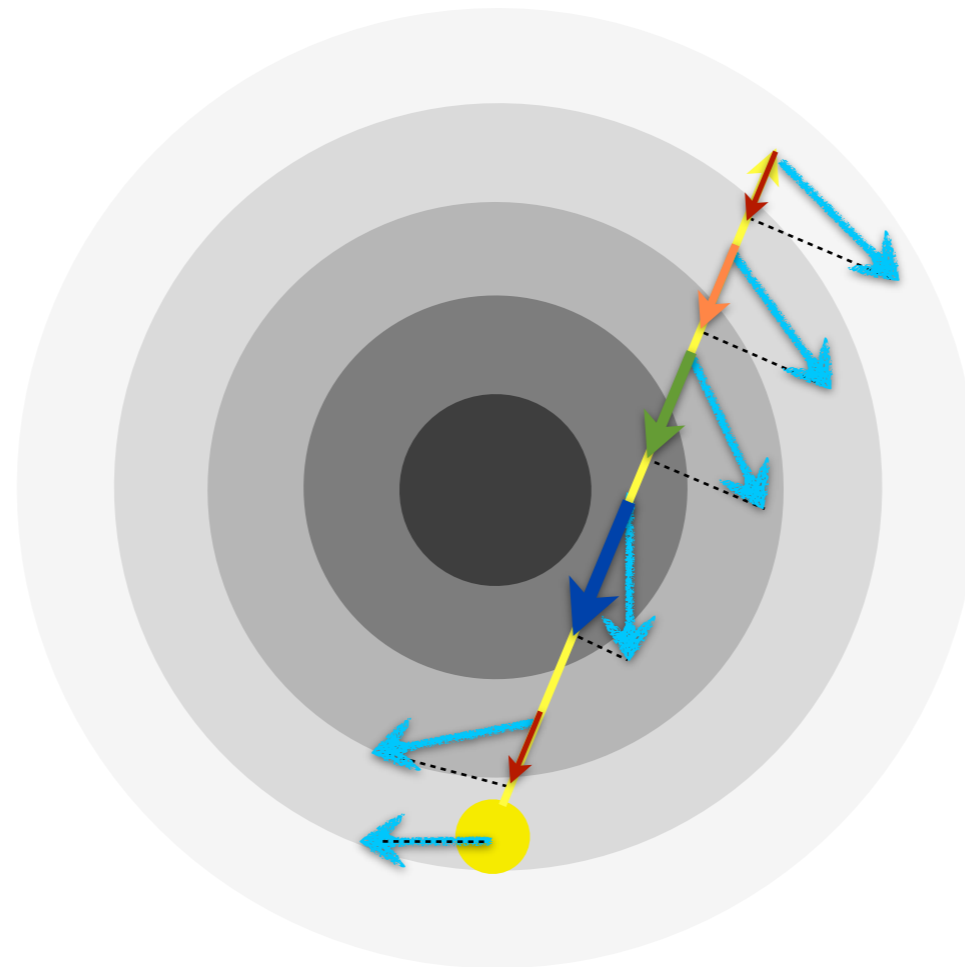
Dame et al. 2001

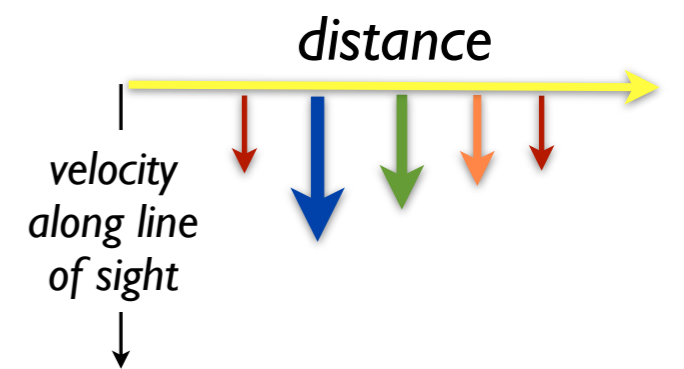
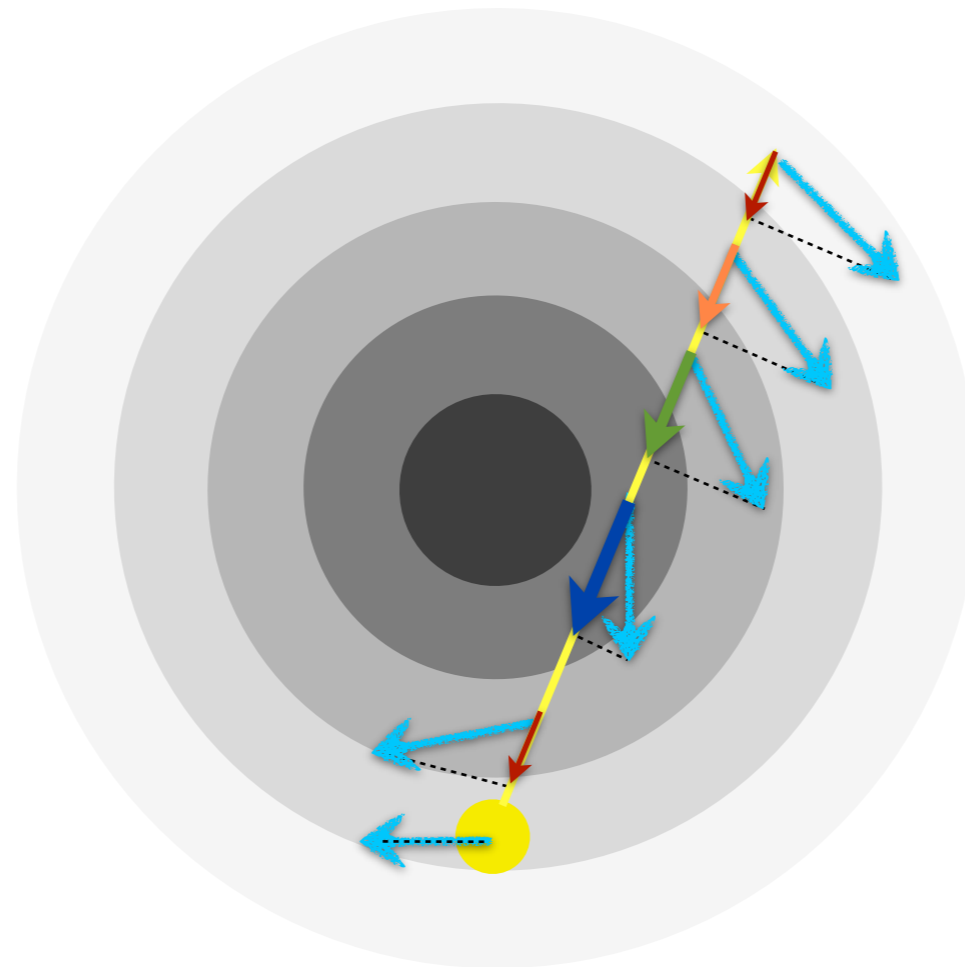
A Spiral Galaxy Observed from its Outskirts...



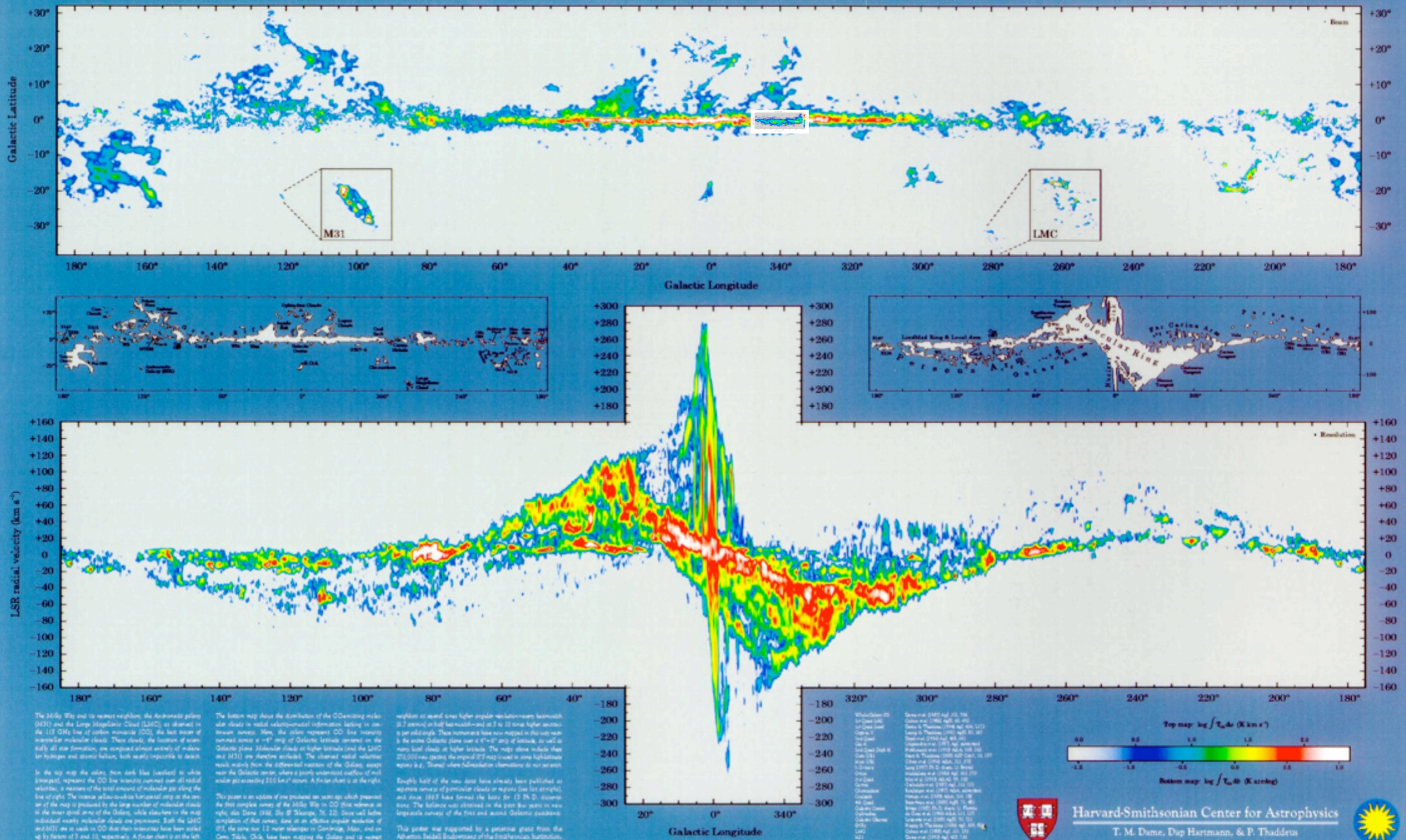








Milky Way Gas Kinematics 101



The Milky Way and its nearest neighbors, the Andromeda galaxy (M31) and the Large Magellanic Cloud (LMC), are shown in the top map. The color-coded map shows the distribution of CO-emitting molecular gas. The color bar at the bottom right indicates the intensity scale for both maps, ranging from -2.5 to 2.5 K km s⁻¹.

The bottom map shows the distribution of the CO-emitting molecular gas. The color bar at the bottom right indicates the intensity scale for both maps, ranging from -2.5 to 2.5 K km s⁻¹.

roughly half of the new data have already been published as separate surveys of particular clouds or regions (see list on right), and these lists have formed the basis for 13 Ph.D. dissertations. The balance was obtained in the past few years in new large-scale surveys of the first and second Galactic quadrants.

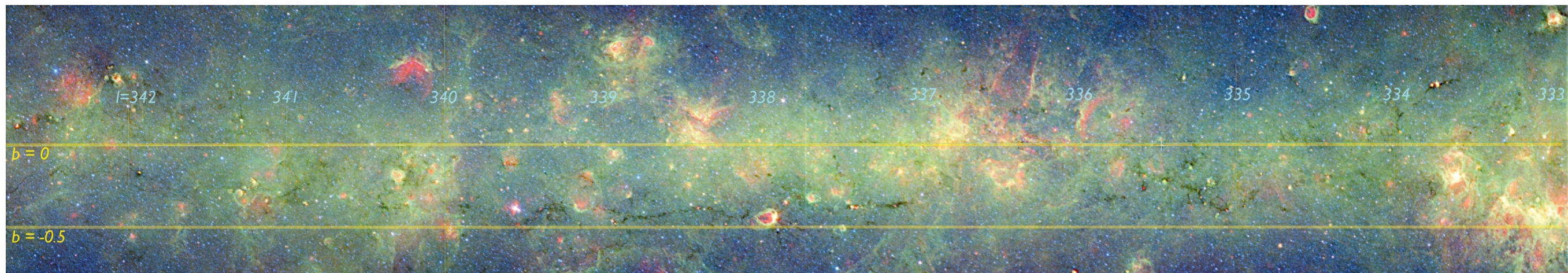
This paper was supported by a generous grant from the Arthur S. Holden Endowment of the Smithsonian Institution.

- Wolfe-Gordon 1975
- ... (list of references)

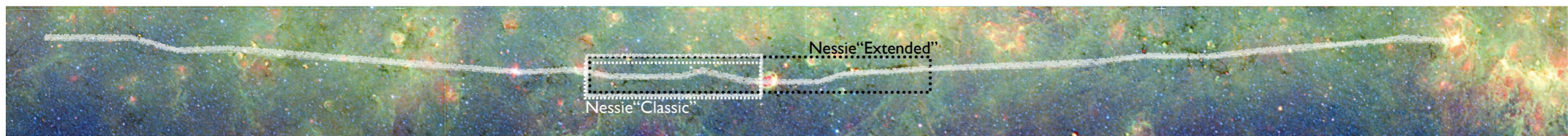
- ... (list of references)

Top map: $\log I_{CO} \text{ (K km s}^{-1}\text{)}$
 Bottom map: $\log I_{CO} \text{ (K km s}^{-1}\text{)}$

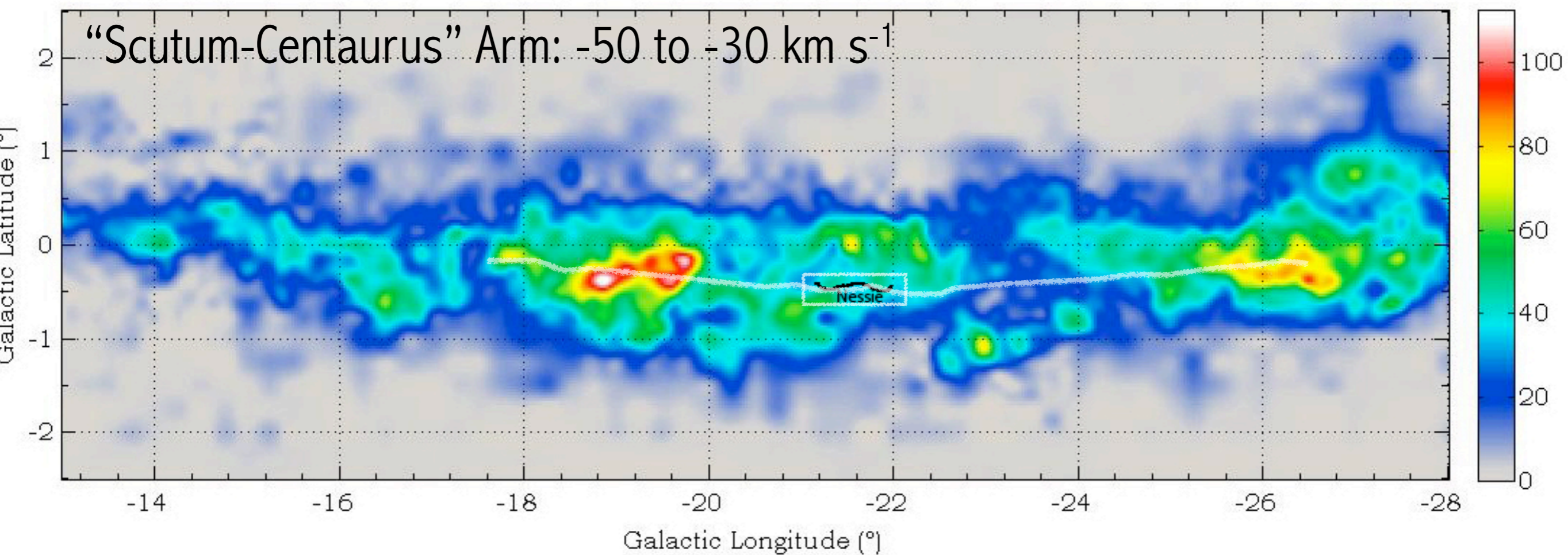




$1 \text{ degree} \sim 60 \text{ pc at } 3.5 \text{ kpc}$

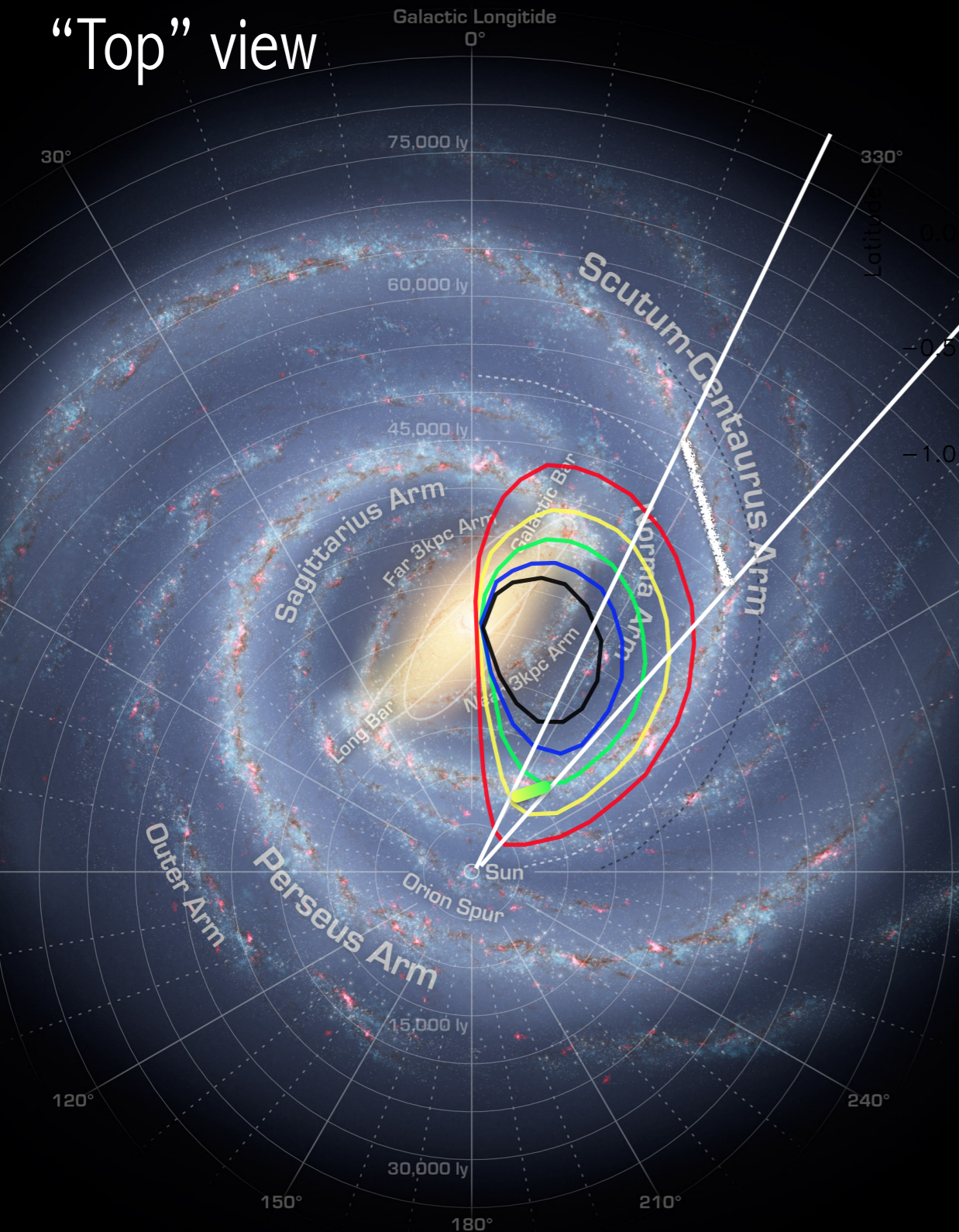


Wco m50 m30.fits

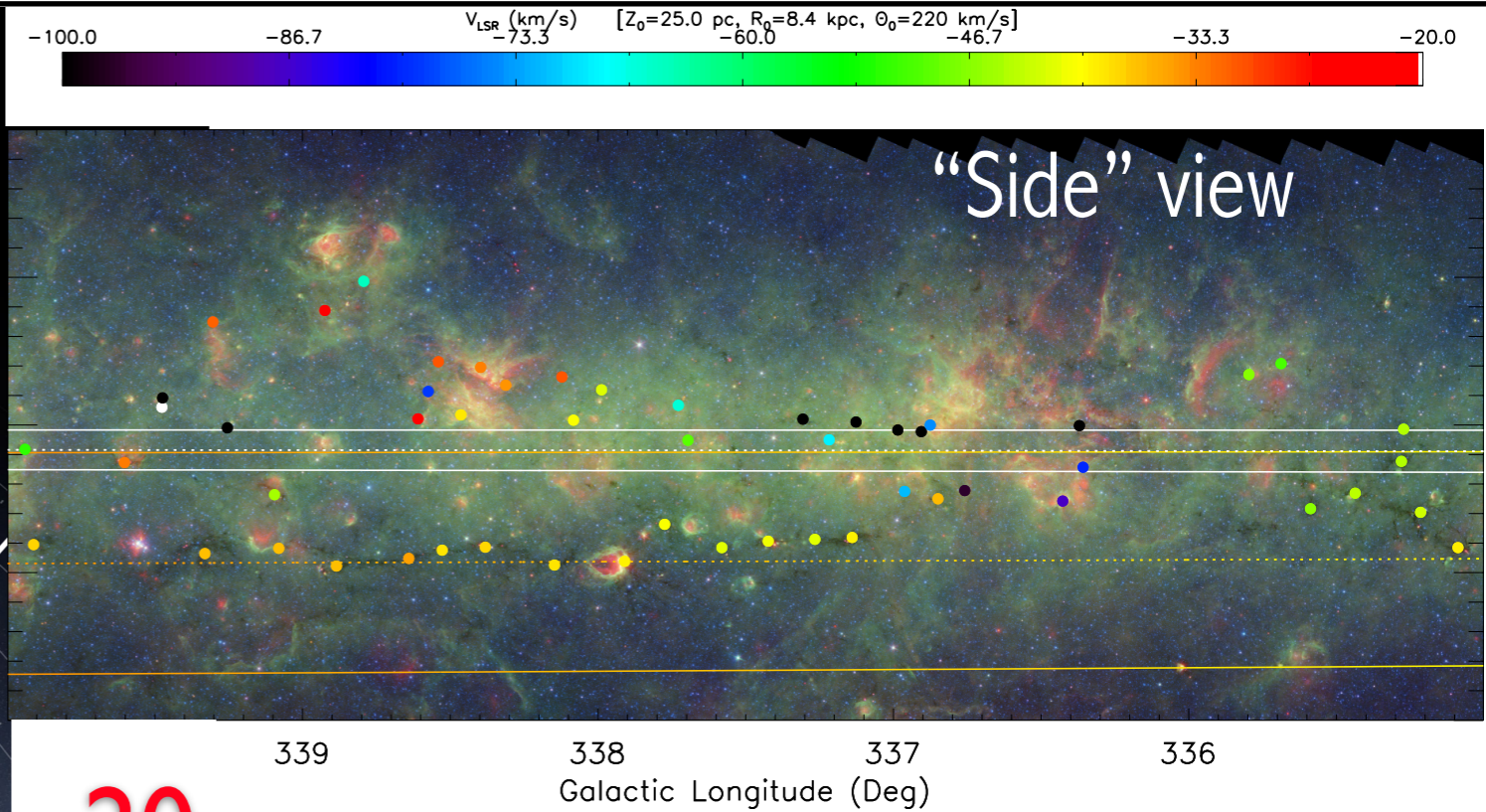


Using Velocity Constraints

“Top” view



“Side” view



-20
-40
-60
-80
-100

Left: Scientist-artist Robert Hurt’s view of the Milky Way (cartoon based on stars, CO, HI, masers & HII regions; Benjamin, Dame et al.)

IC 349

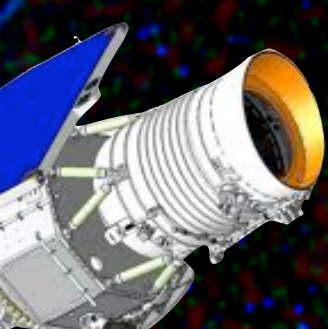
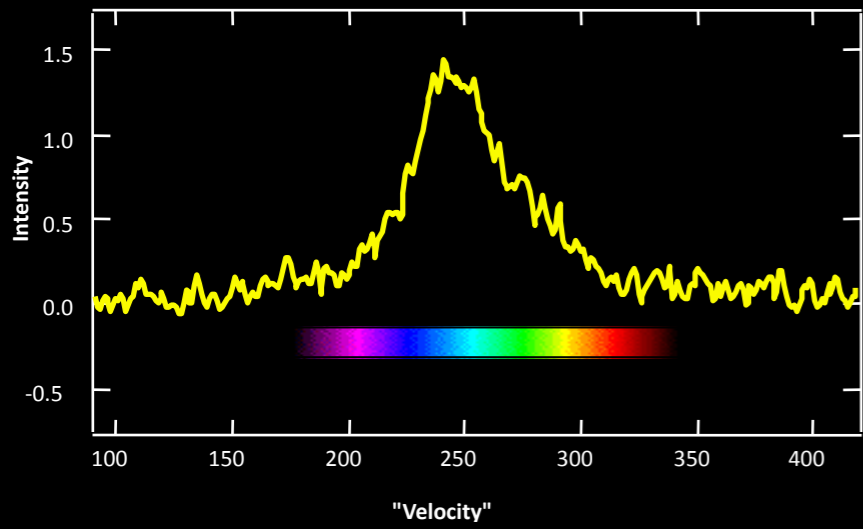
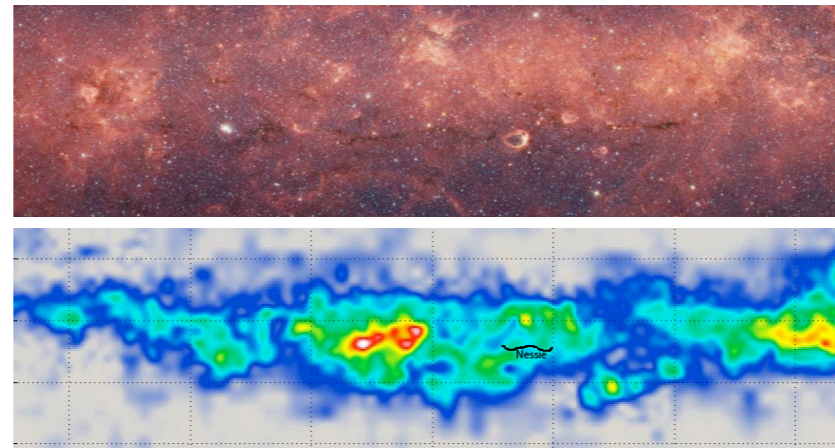
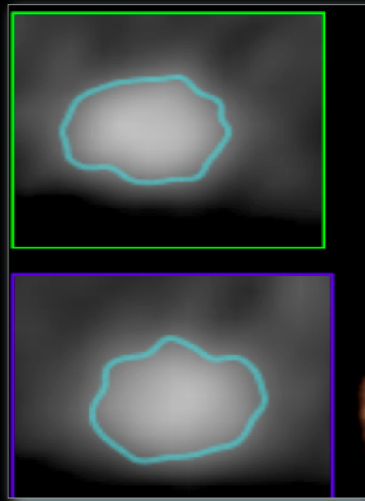


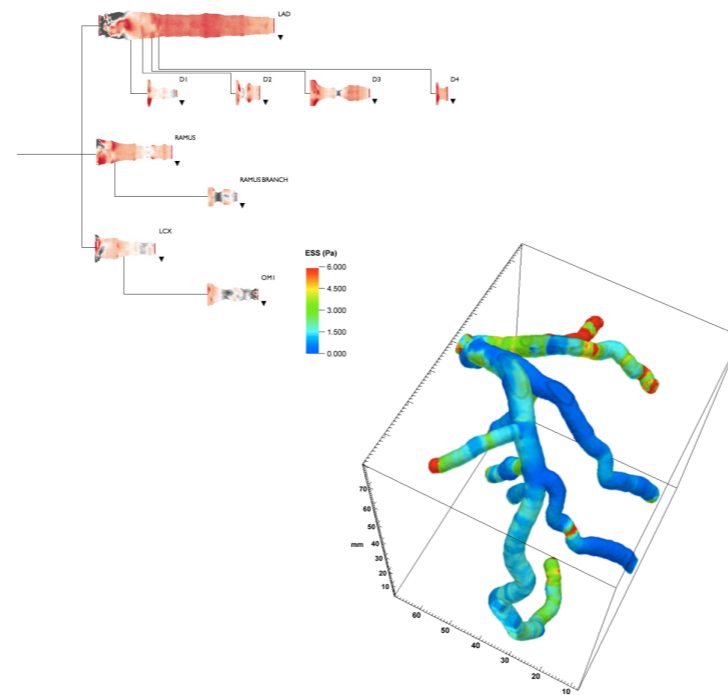
image from Jarrett et al. 2012; WISE Enhanced Resolution Galaxy Atlas



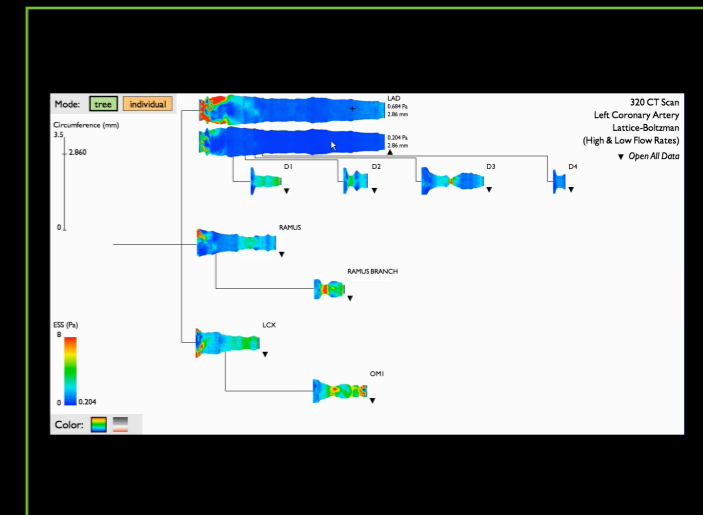
Data

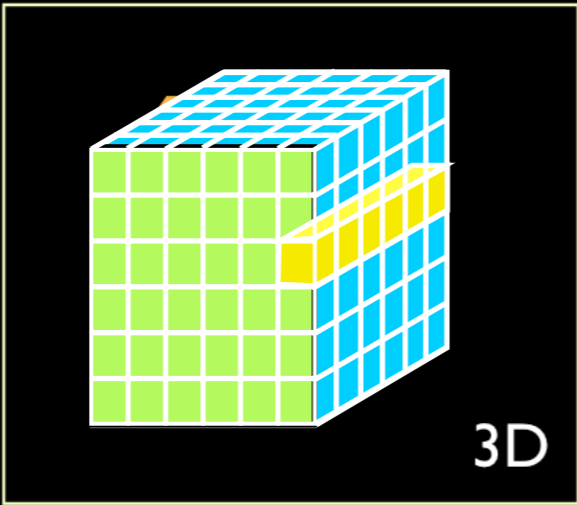


Dimensions

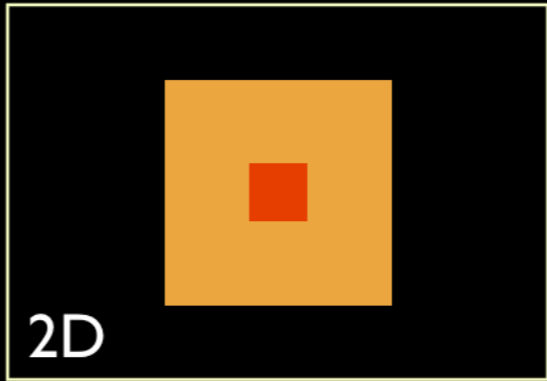


Display

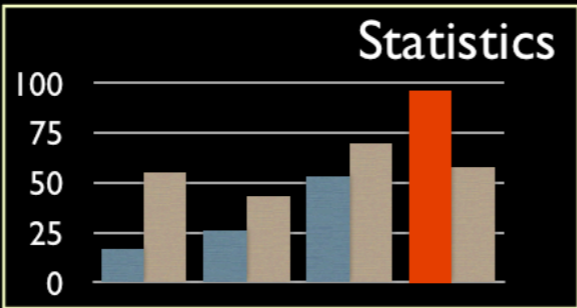




3D



2D

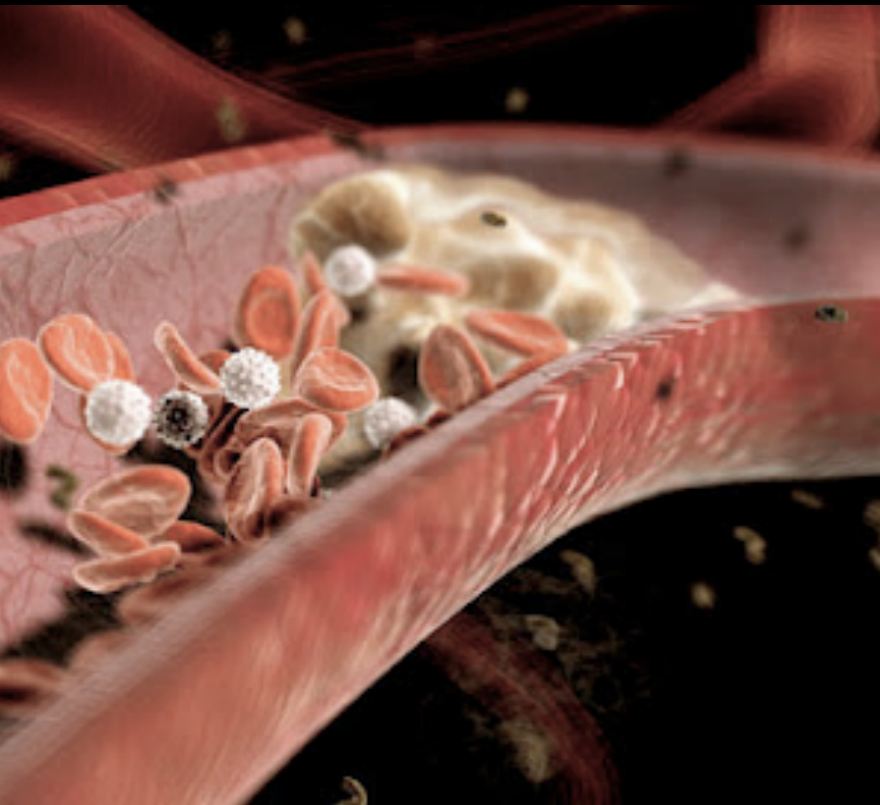


Data Abstraction



The Close Relationship between Heart Disease and Sea Monsters*

**or, why good high-dimensional data visualization can solve almost any problem!*



Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics



John Tukey

Principles of high-dimensional data visualization in astronomy

A.A. Goodman*

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2012 May 3, accepted 2012 May 4

Published online 2012 Jun 15

Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing – techniques: radial velocities

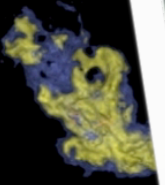
Astronomical researchers often think of analysis and visualization as separate tasks. In the case of high-dimensional data sets, though, interactive *exploratory data visualization* can give far more insight than an approach where data processing and statistical analysis are followed, rather than accompanied, by visualization. This paper attempts to chart a course toward “linked view” systems, where multiple views of high-dimensional data sets update live as a researcher selects, highlights, or otherwise manipulates, one of several open views. For example, imagine a researcher looking at a 3D volume visualization of simulated or observed data, and simultaneously viewing statistical displays of the data set’s properties (such as an x - y plot of temperature vs. velocity, or a histogram of vorticities). Then, imagine that when the researcher selects an interesting group of points in any one of these displays, that the same points become a highlighted subset in all other open displays. Selections can be graphical or algorithmic, and they can be combined, and saved. For tabular (ASCII) data, this kind of analysis has long been possible, even though it has been under-used in astronomy. The bigger issue for astronomy and other “high-dimensional” fields, though, is that no extant system allows for full integration of images and data cubes within a linked-view environment. The paper concludes its history and analysis of the present situation with suggestions that look toward cooperatively-developed open-source modular software as a way to create an evolving, flexible, high-dimensional, linked-view visualization environment useful in astrophysical research.

3D Selection

Why
How
How



on?



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

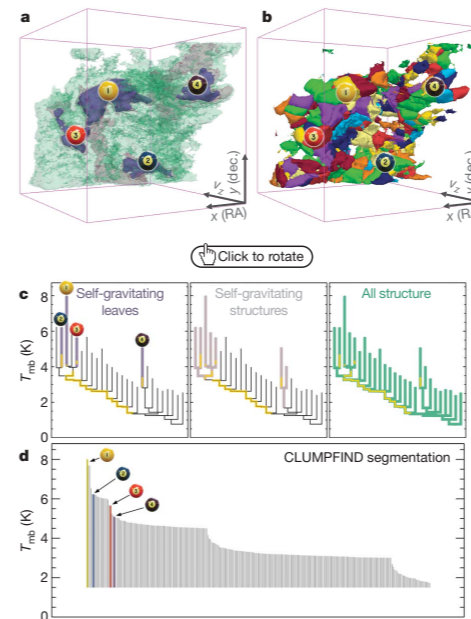


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ¹³CO emission from the L1448 region of Perseus. **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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 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 set⁸ 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 frequency¹³.

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 work¹⁴ 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 $X_{13CO} = 8.0 \times 10^{20} \text{ cm}^2 \text{ K}^{-1} \text{ s}$ (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 fields⁶, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

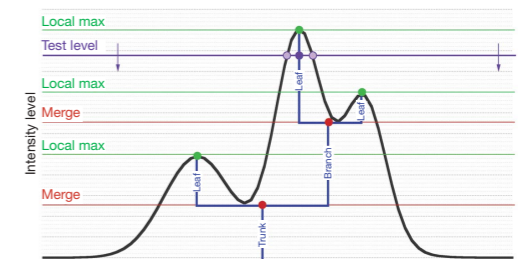


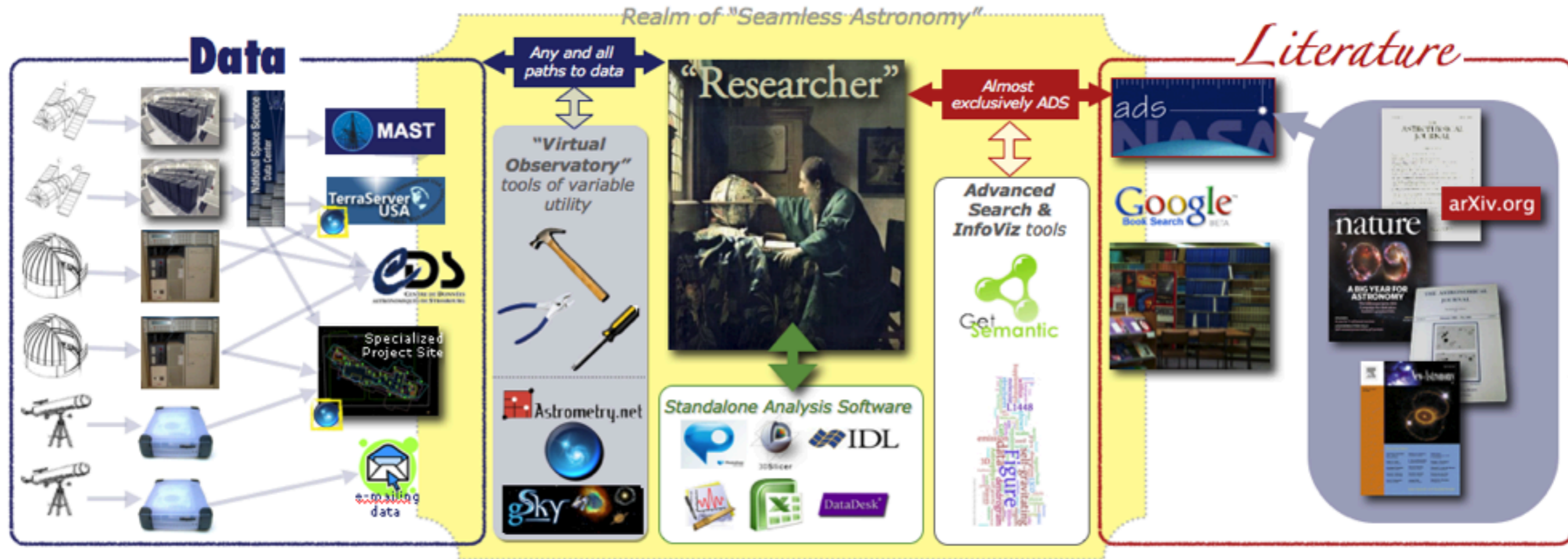
Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in 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

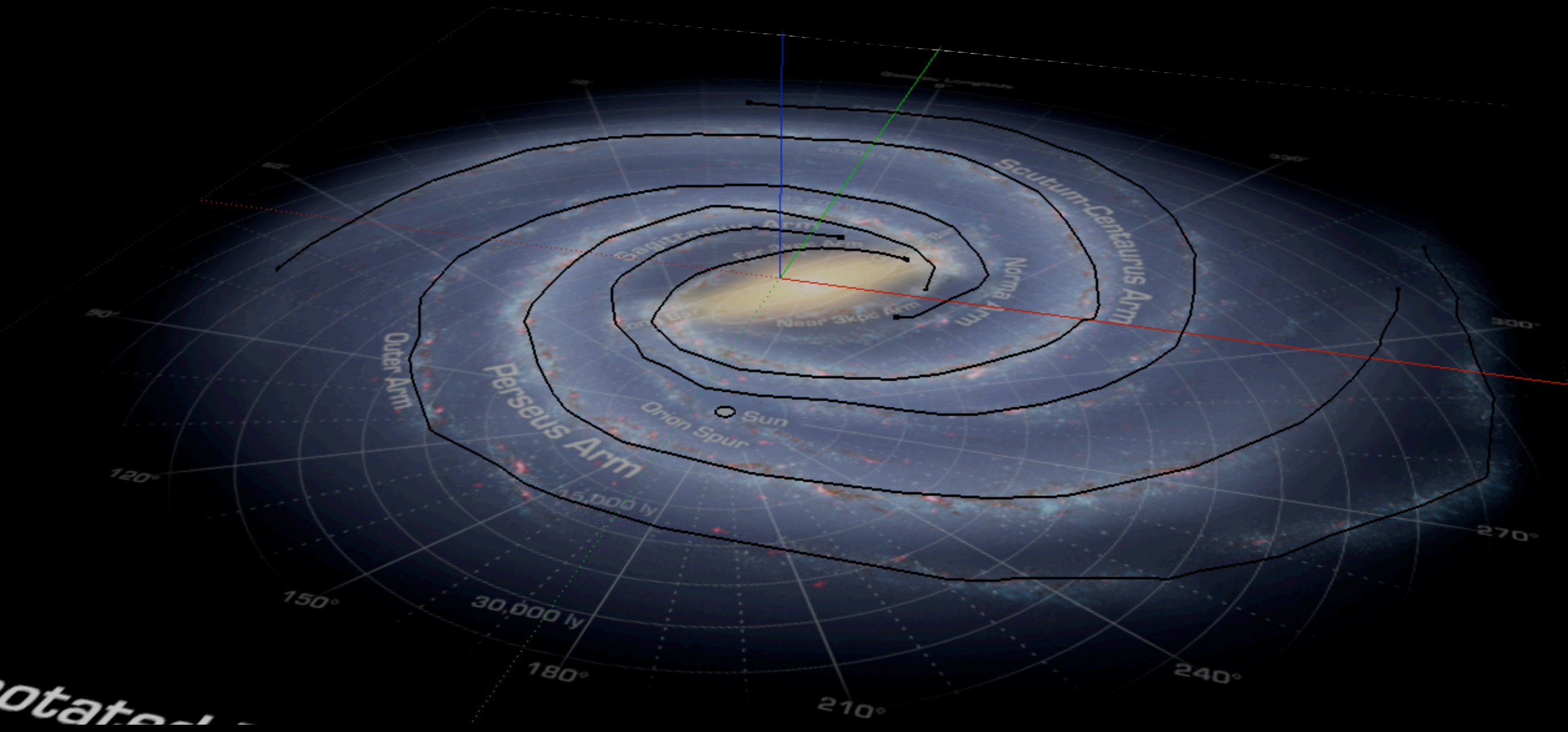


SEAMLESS ASTRONOMY

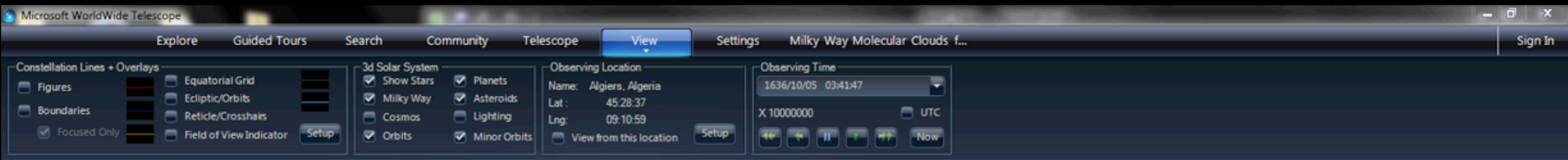
Linking scientific data, publications, and communities



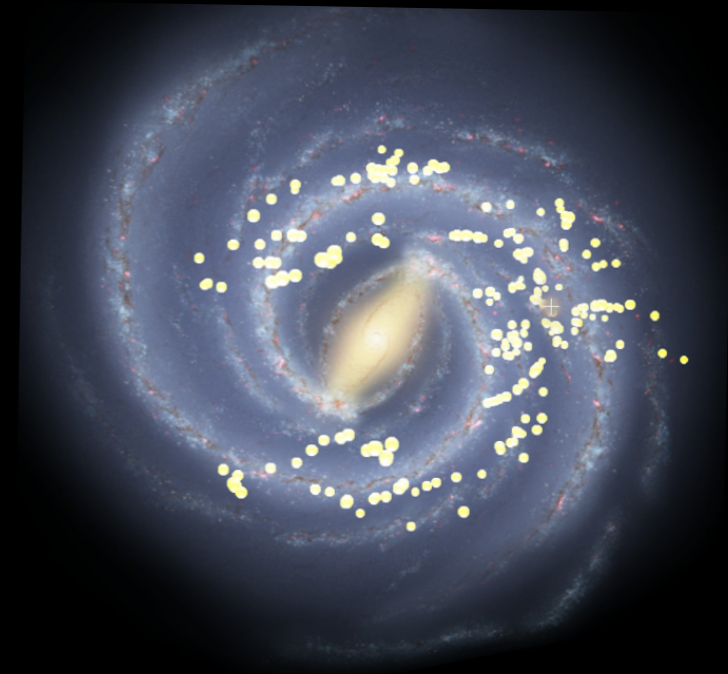
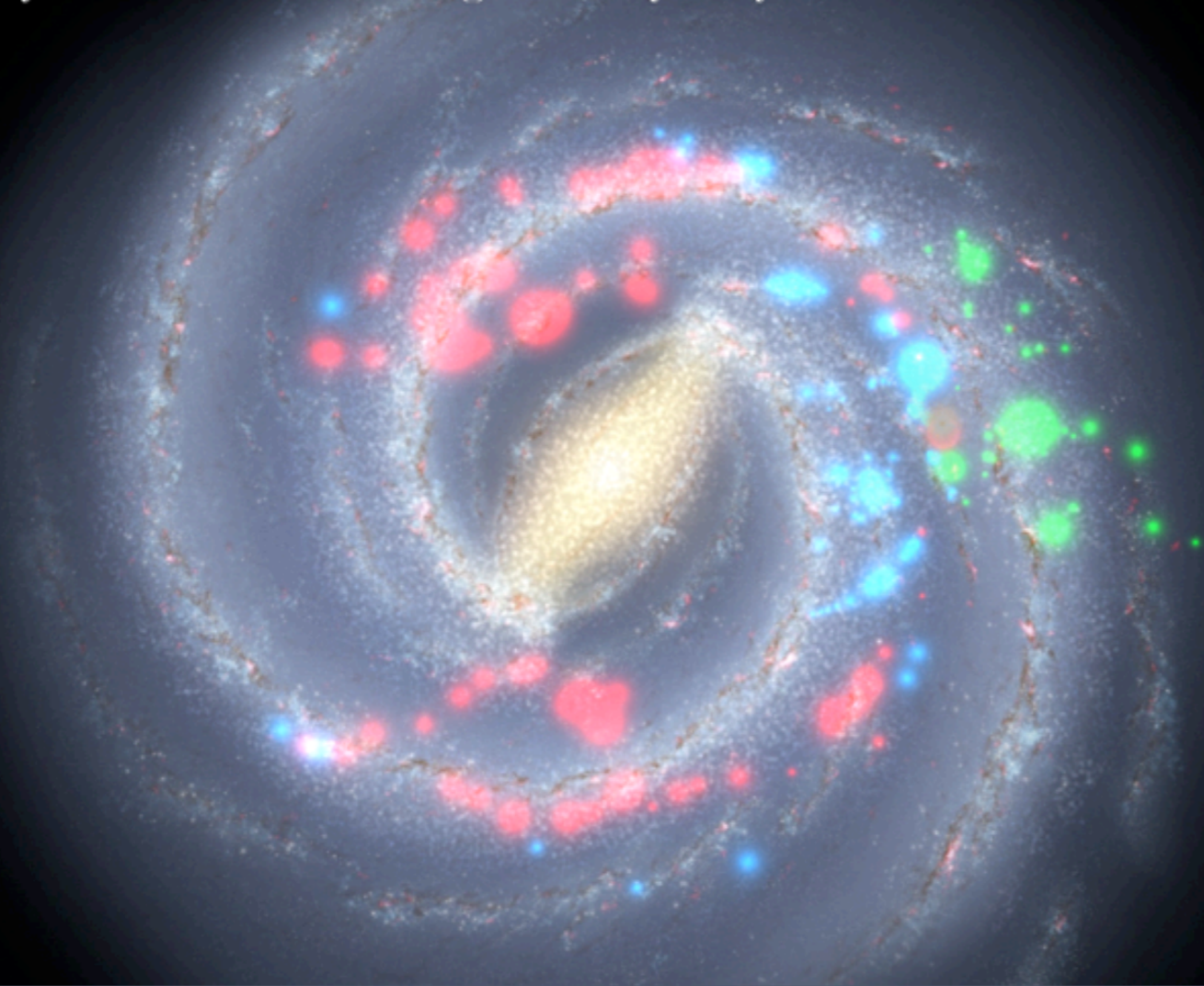
projects.iq.harvard.edu/seamlessastronomy/



Value of "Linked Views" from Harvard Undergrad Thesis* Work



Results from Tom Rice's Thesis:
Preliminary Hierarchical Catalog of Milky Way Plane Molecular Clouds



*thesis of Tom Rice, '12, awarded Hoopes Prize