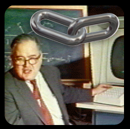
LINKED-VIEW VISUALIZATION OF HIGH-DIMENSIONAL DATA IN GLUE



Alyssa A. Goodman

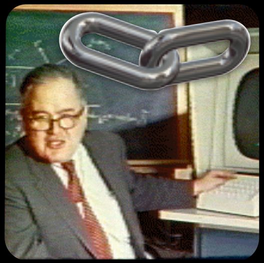
Harvard-Smithsonian Center for Astrophysics & Radcliffe Institute

with Chris Beaumont, Michelle Borkin, Penny Qian & Tom Robitaille

@aagie
@glueviz
@astrofrog

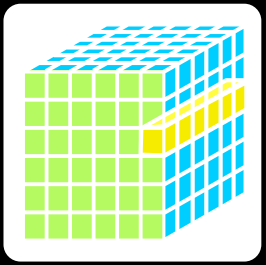
glueviz.org
github.com/glue-viz
Tom Robitaille, lead developer

James Webb
Space Telescope



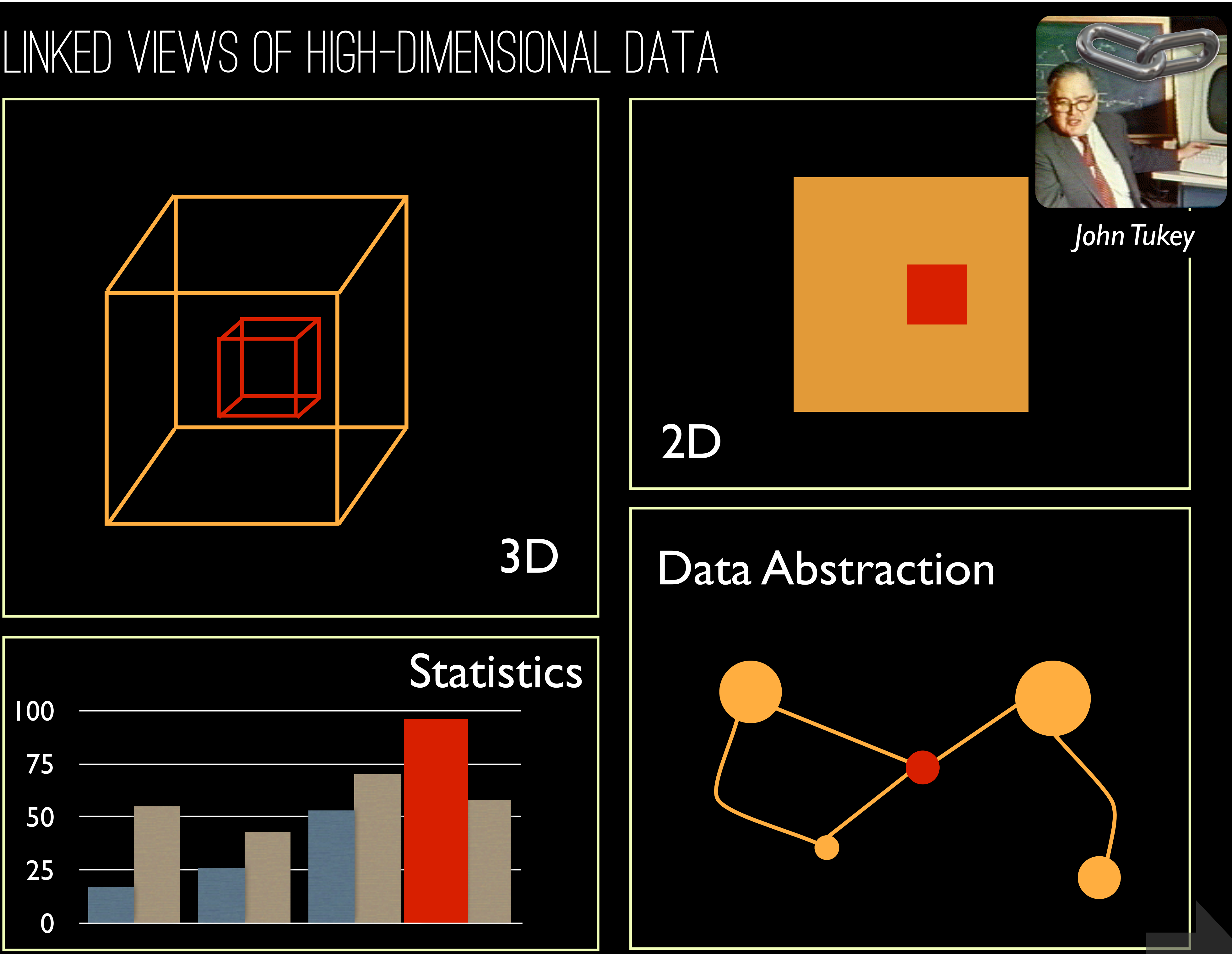
“Linked Views”

Open Source Python, on GitHub



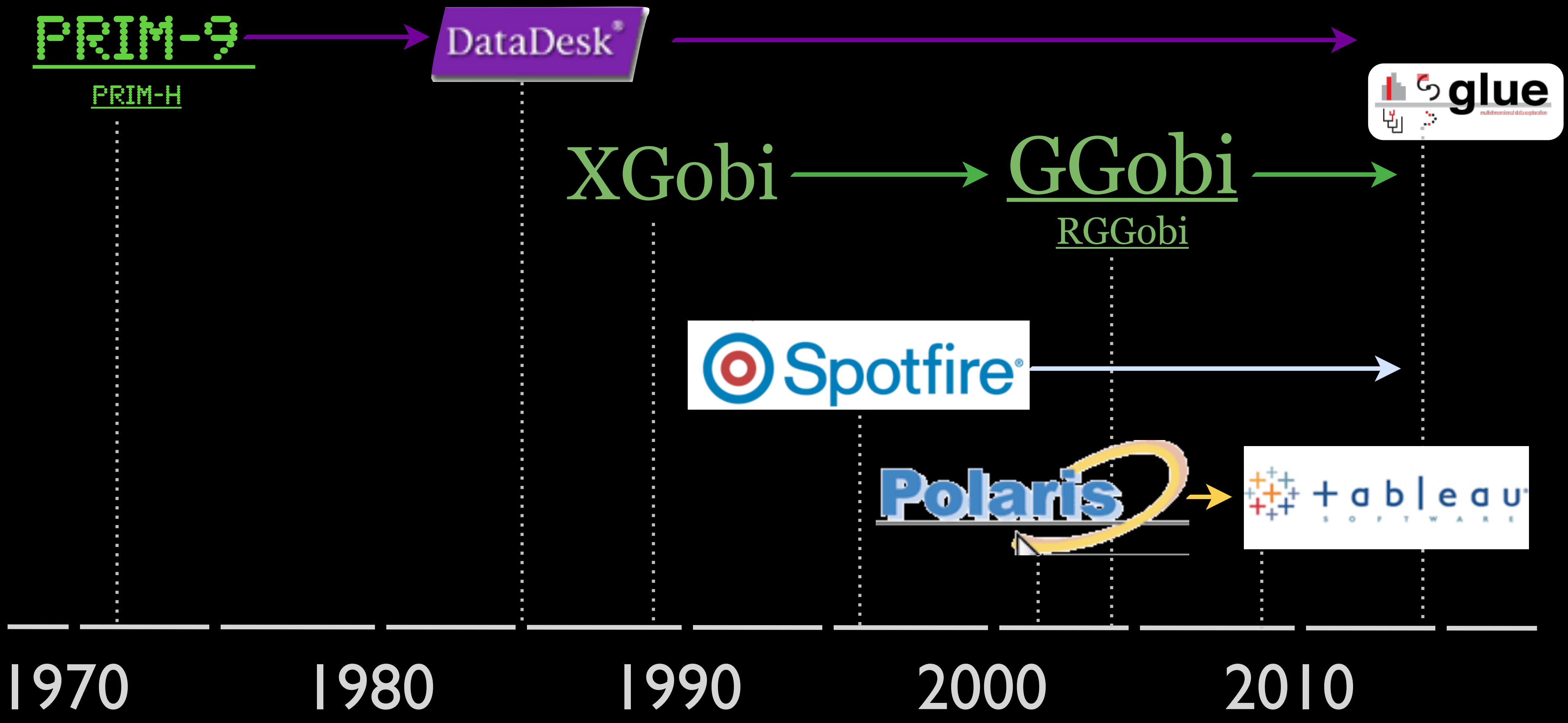
3D+

LINKED VIEWS OF HIGH-DIMENSIONAL DATA



figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

JOHN TUKEY'S LEGACY



PRIM-9

PRIM-H

DataDesk®

glue

multidimensional data exploration

XGobi

GGobi

RGGobi

Spotfire®

Polaris

+ a b l e a u

SOFTWARE

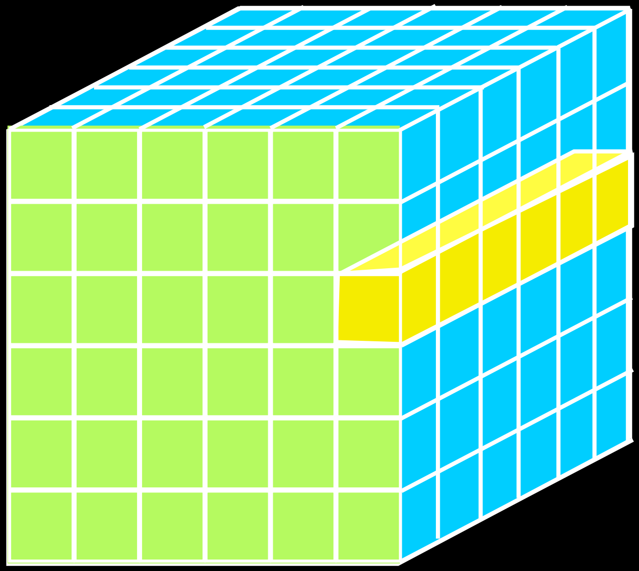
1970

1980

1990

2000

2010

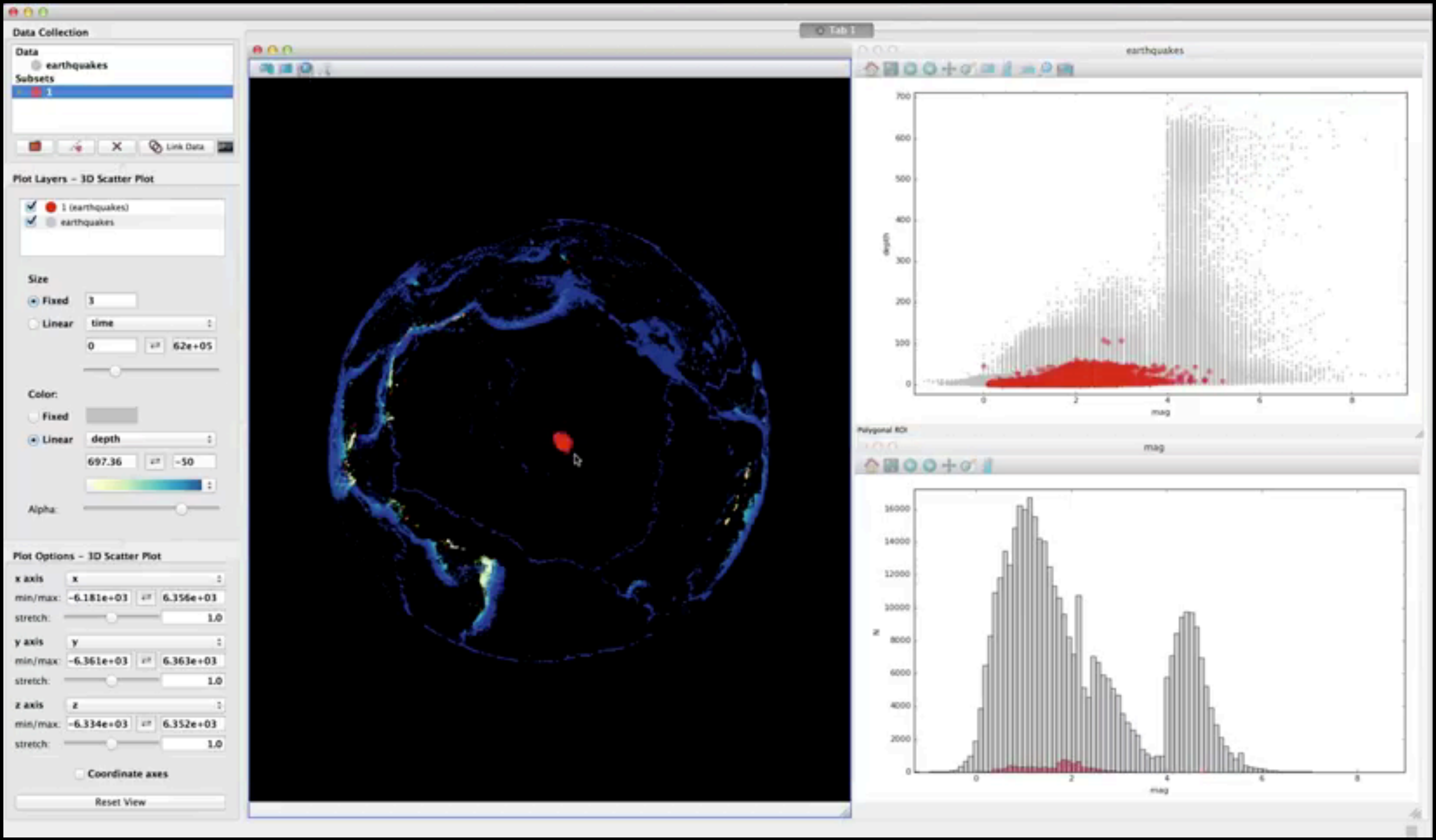


DATA-DIMENSIONS-DISPLAY

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

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



Data Collection

Data
- earthquakes

Subsets
- 1

Plot Layers – 3D Scatter Plot

- ☑ 1 (earthquakes)
- ☑ earthquakes

Size
- ● Fixed: 3
- ○ Linear: time, 0 – 62e+05

Color:
- ○ Fixed
- ● Linear: depth, 697.36 – -50

Alpha:

Plot Options – 3D Scatter Plot

x axis: x
min/max: -6.181e+03 – 6.356e+03
stretch: 1.0

y axis: y
min/max: -6.361e+03 – 6.363e+03
stretch: 1.0

z axis: z
min/max: -6.334e+03 – 6.352e+03
stretch: 1.0

□ Coordinate axes

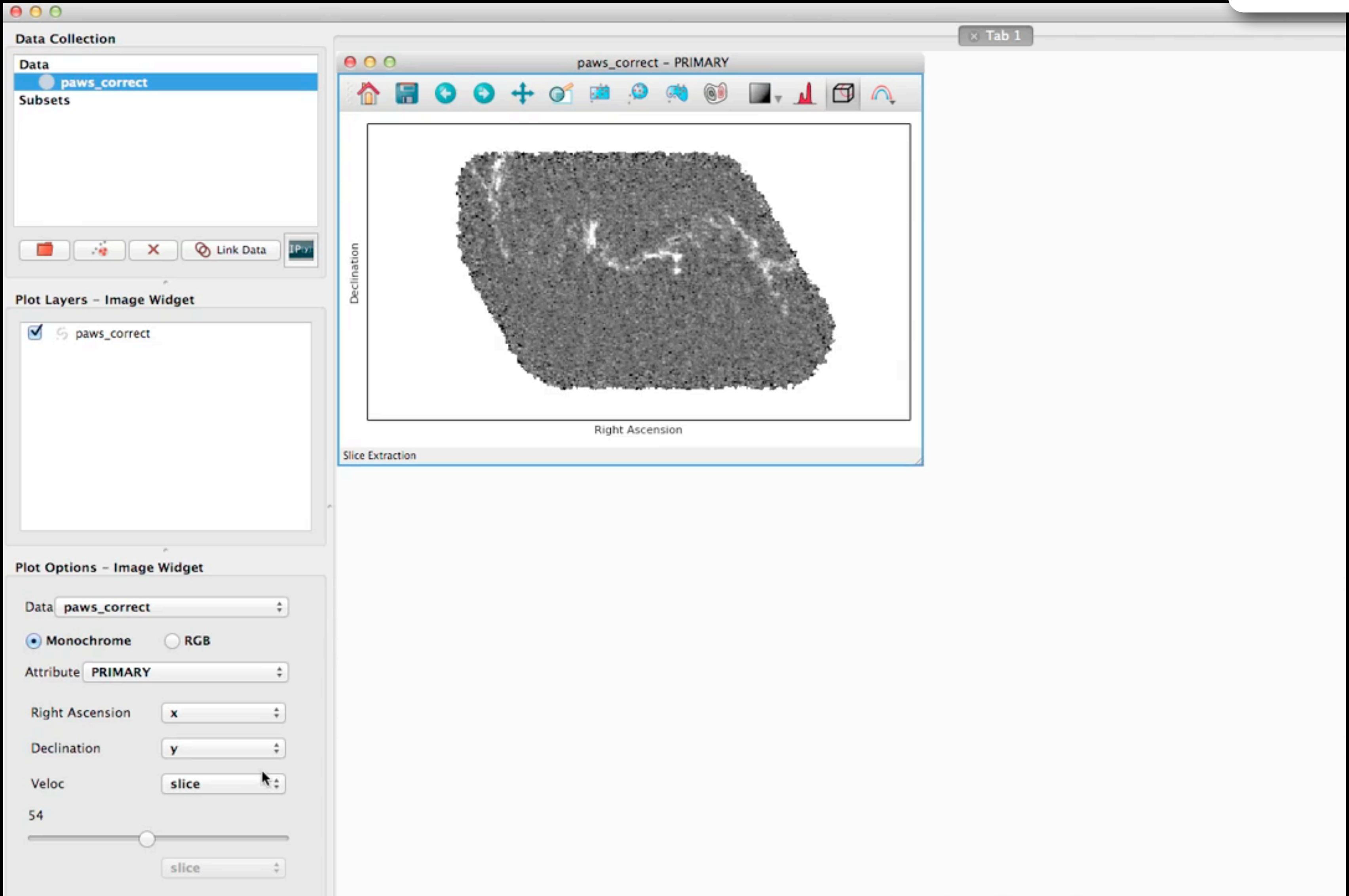
Reset View

video by Tom Robitaille, lead glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



Data: paws_correct

● Monochrome ○ RGB

Attribute: PRIMARY

Right Ascension: x

Declination: y

Veloc: slice

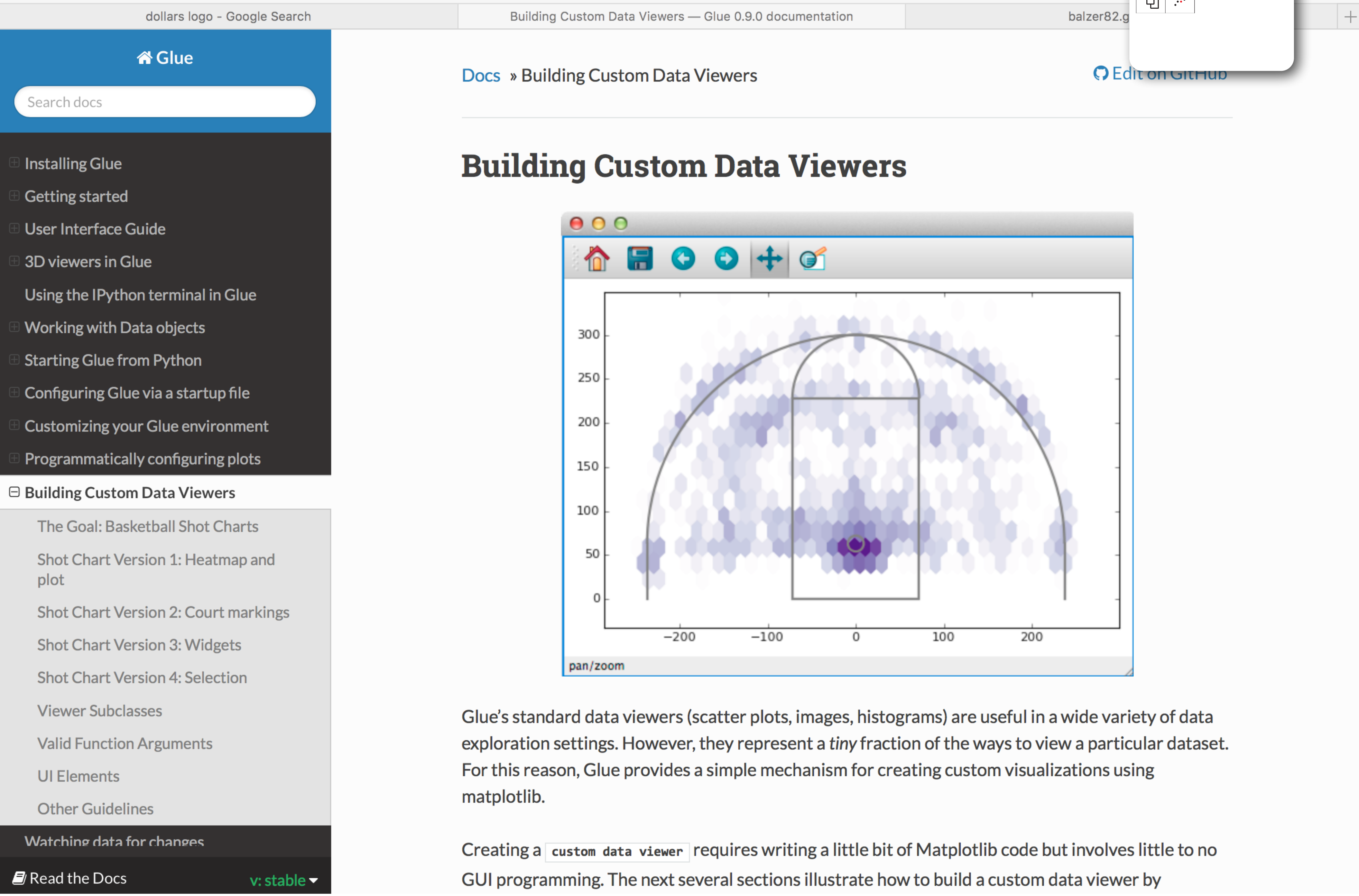
54

slice

video by Chris Beaumont, glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

"BUT WAIT, THERE'S MORE…"



dollars logo - Google Search | Building Custom Data Viewers — Glue 0.9.0 documentation | balzer82.g

Glue

Search docs

- Installing Glue
- Getting started
- User Interface Guide
- 3D viewers in Glue
- Using the IPython terminal in Glue
- Working with Data objects
- Starting Glue from Python
- Configuring Glue via a startup file
- Customizing your Glue environment
- Programmatically configuring plots
- Building Custom Data Viewers
  - The Goal: Basketball Shot Charts
  - Shot Chart Version 1: Heatmap and plot
  - Shot Chart Version 2: Court markings
  - Shot Chart Version 3: Widgets
  - Shot Chart Version 4: Selection
  - Viewer Subclasses
  - Valid Function Arguments
  - UI Elements
  - Other Guidelines
- Watching data for changes

Read the Docs v: stable

Docs » Building Custom Data Viewers

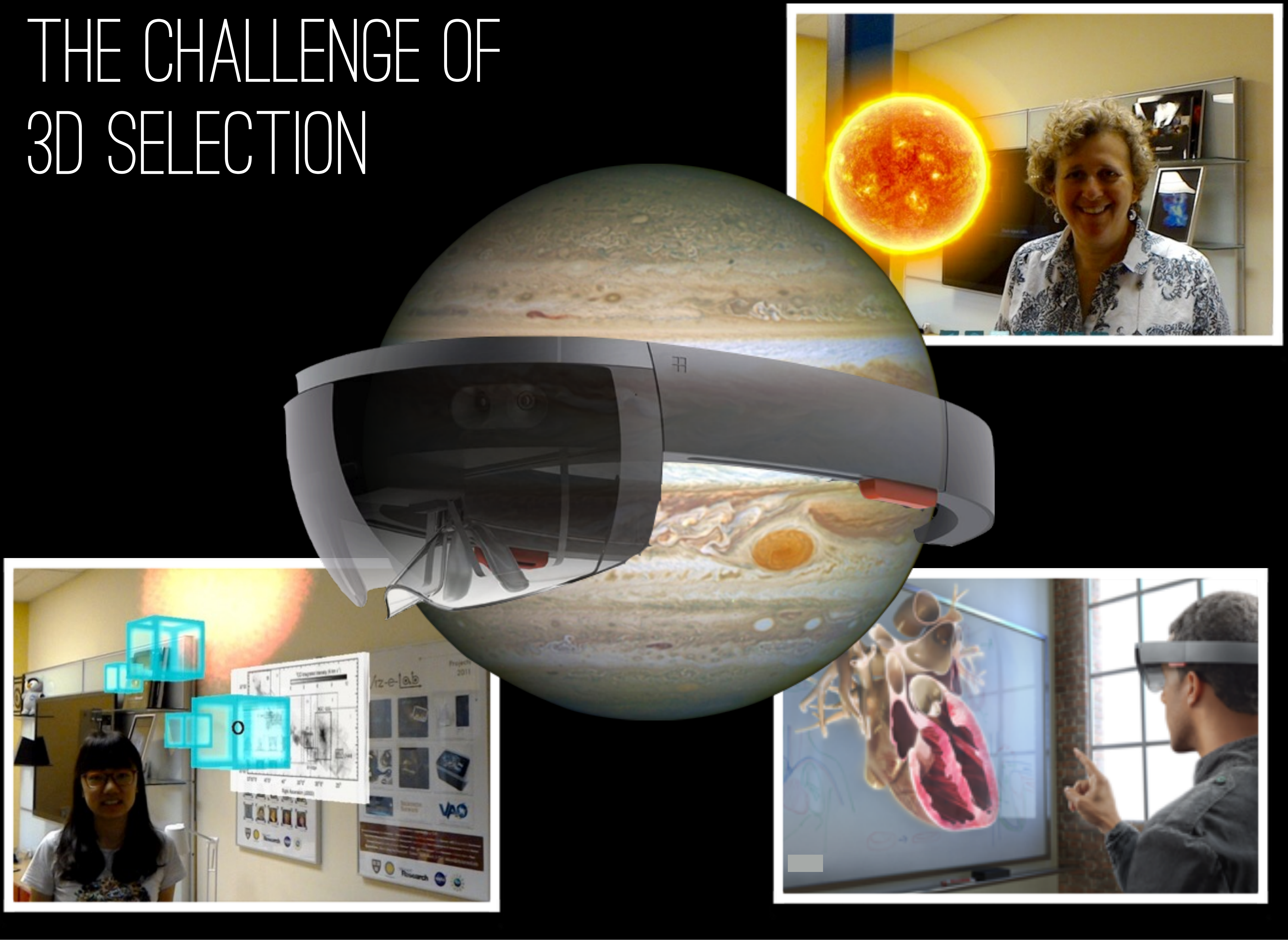
Edit on GitHub

Building Custom Data Viewers

Glue's standard data viewers (scatter plots, images, histograms) are useful in a wide variety of data exploration settings. However, they represent a *tiny* fraction of the ways to view a particular dataset. For this reason, Glue provides a simple mechanism for creating custom visualizations using matplotlib.

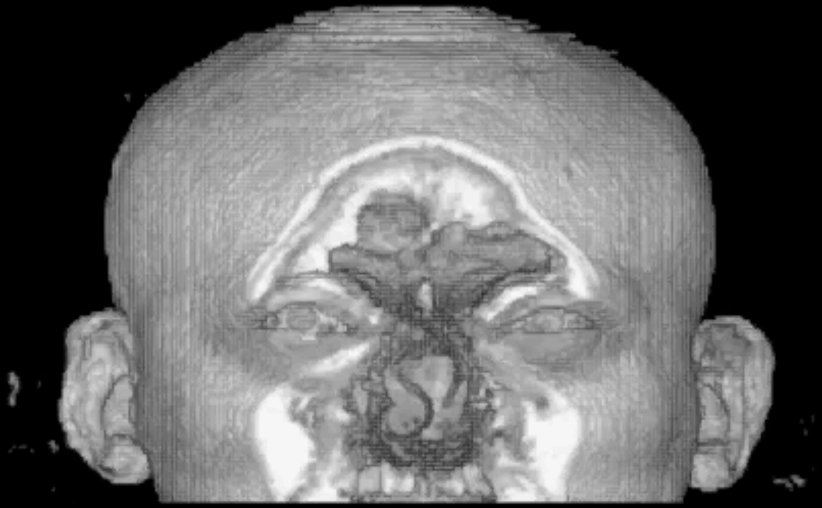
Creating a `custom data viewer` requires writing a little bit of Matplotlib code but involves little to no GUI programming. The next several sections illustrate how to build a custom data viewer by

THE CHALLENGE OF 3D SELECTION



ASTRONOMICAL MEDICINE

"KEITH"



"z" is depth into head

"PERSEUS"

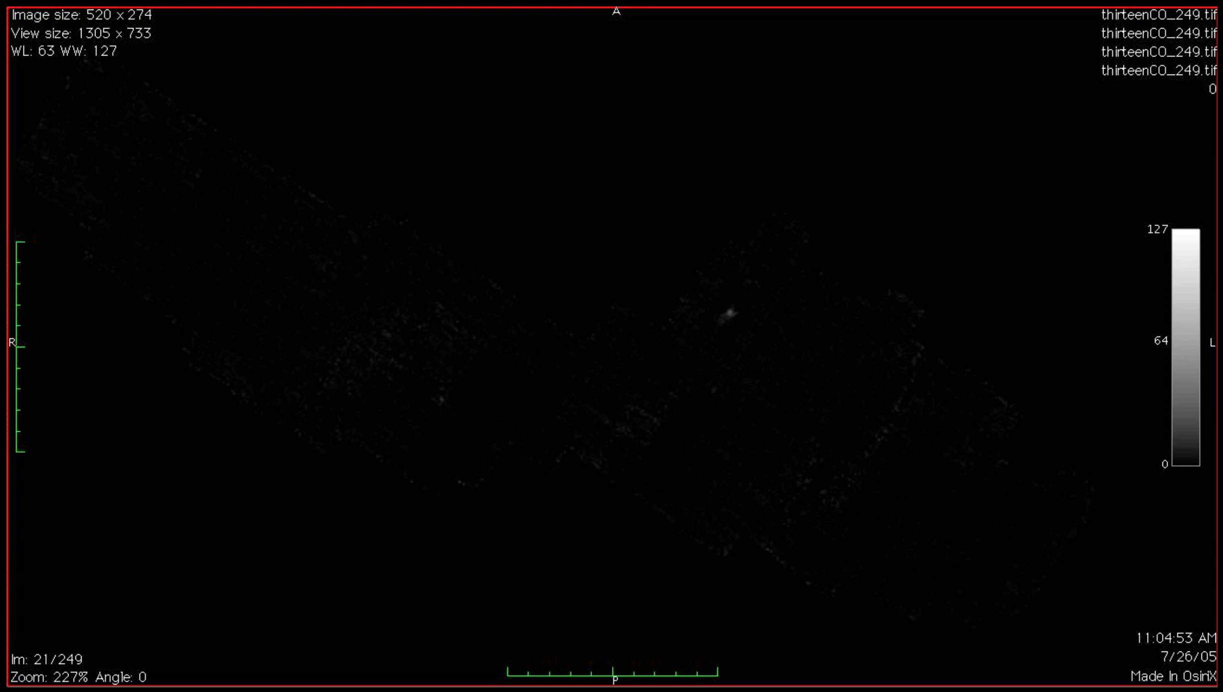


Image size: 520 x 274
View size: 1305 x 733
WL: 63 WW: 127

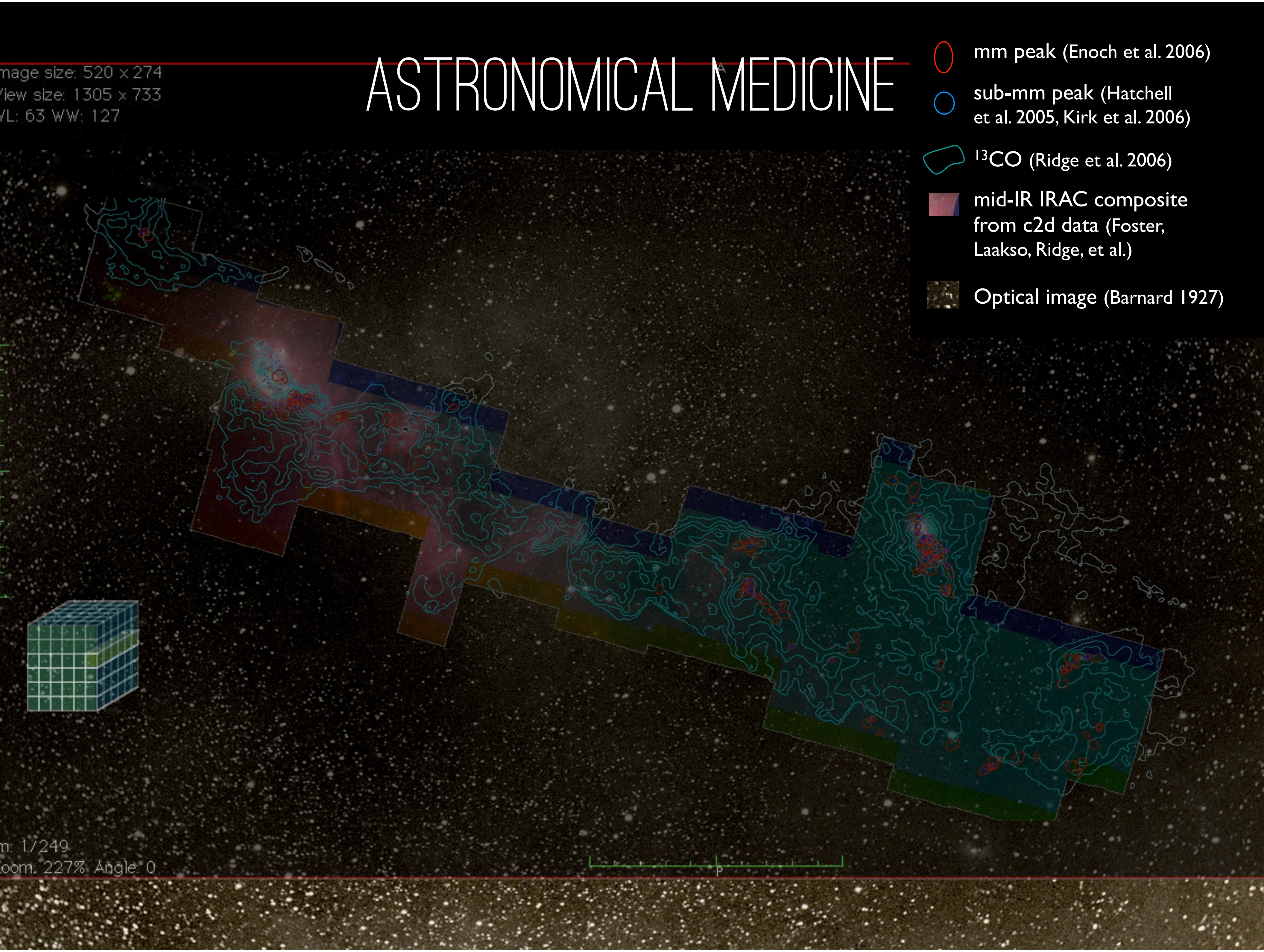
thirteenCO_249.tif
thirteenCO_249.tif
thirteenCO_249.tif
thirteenCO_249.tif

Im: 21/249
Zoom: 227% Angle: 0

11:04:53 AM
7/26/05
Made In OsiriX

"z" is line-of-sight velocity

ASTRONOMICAL MEDICINE



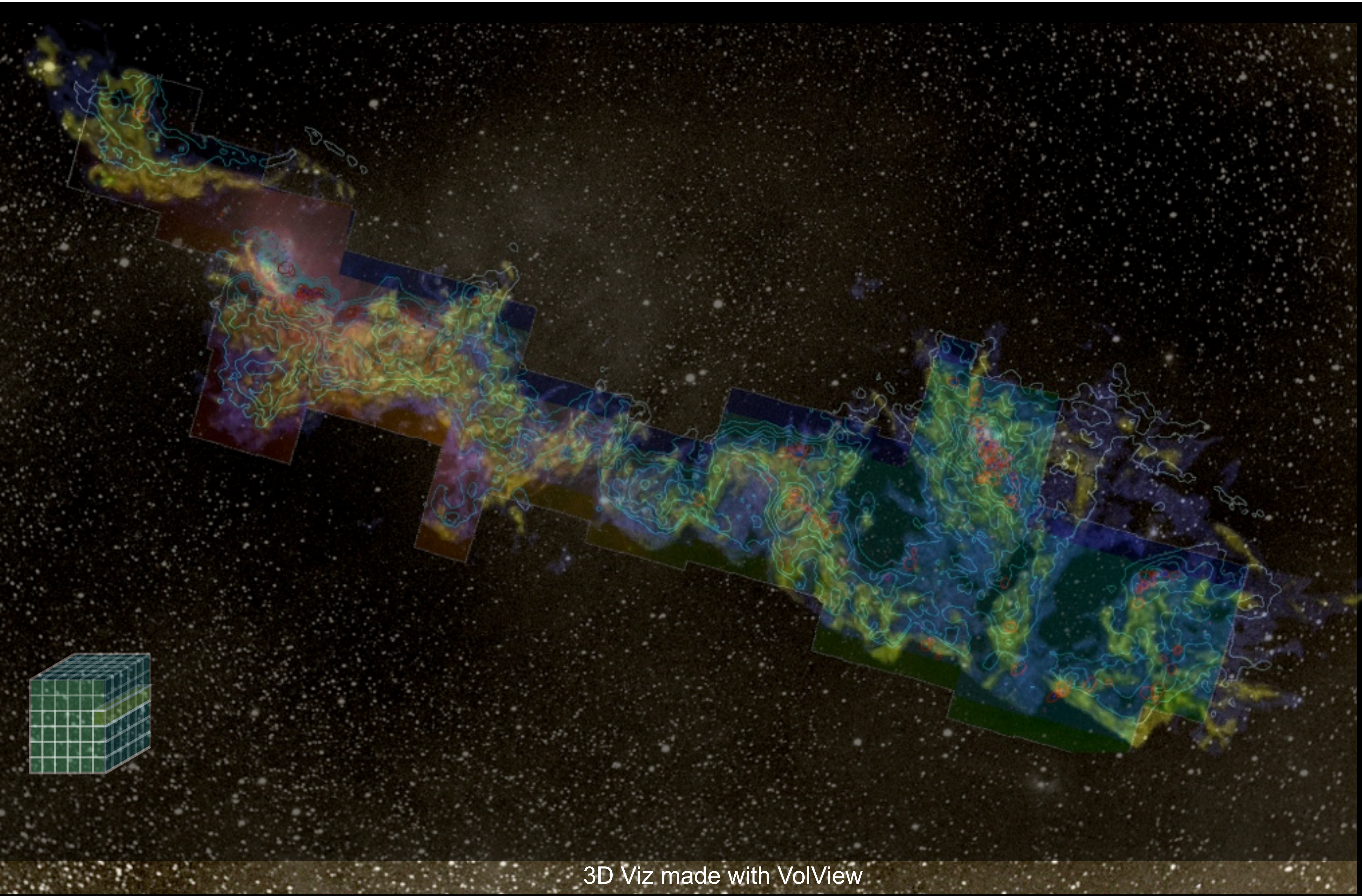
○ mm peak (Enoch et al. 2006)

○ 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)



3D Viz made with VolView

COMPLETE

altitude
10
5
0
latitude
42.5
42
longitude
-72
-71
-70

"BUT WAIT, THERE'S MORE..."

"cuts" along arbitrary paths

flood-fill selection (2D, 3D)

export to d3po, plotly

custom viewers (e.g. GIS, WorldWide Telescope, Super Mario)

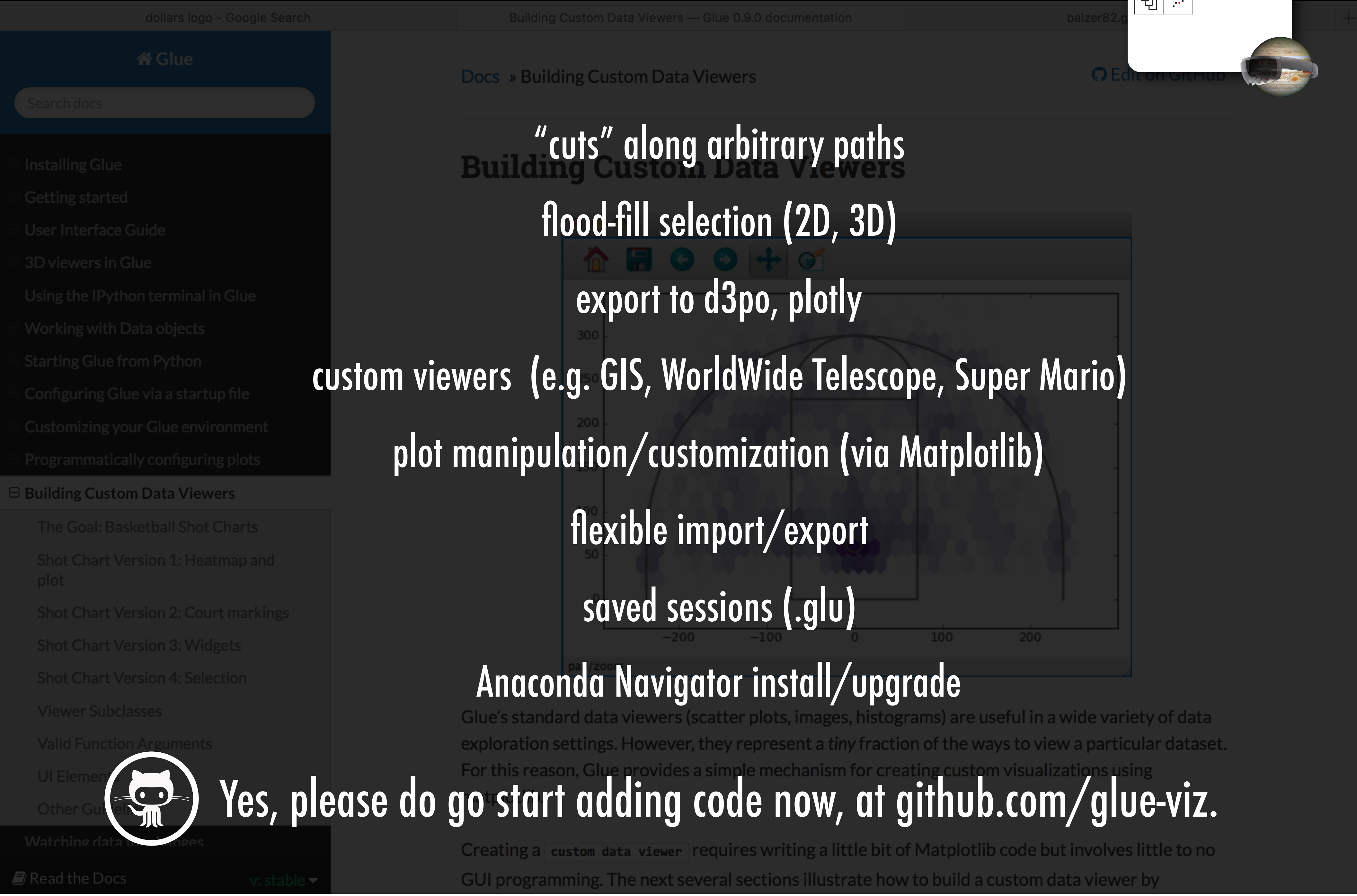
plot manipulation/customization (via Matplotlib)

flexible import/export

saved sessions (.glu)

Anaconda Navigator install/upgrade

Yes, please do go start adding code now, at github.com/glue-viz.

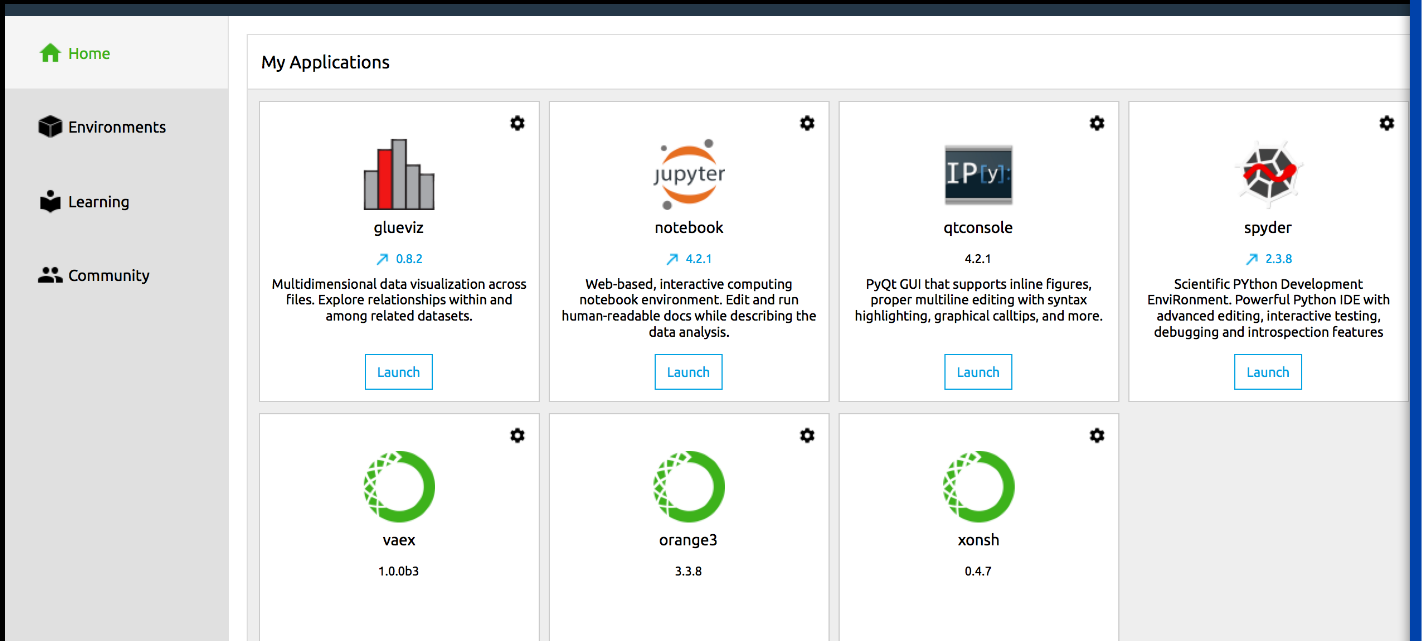

dollars logo - Google Search
Building Custom Data Viewers — Glue 0.9.0 documentation
balzer82.g
Glue
Search docs
Installing Glue
Getting started
User Interface Guide
3D viewers in Glue
Using the IPython terminal in Glue
Working with Data objects
Starting Glue from Python
Configuring Glue via a startup file
Customizing your Glue environment
Programmatically configuring plots
Building Custom Data Viewers
The Goal: Basketball Shot Charts
Shot Chart Version 1: Heatmap and plot
Shot Chart Version 2: Court markings
Shot Chart Version 3: Widgets
Shot Chart Version 4: Selection
Viewer Subclasses
Valid Function Arguments
UI Elemen
Other Gui
Watching data
Read the Docs
v: stable
Docs » Building Custom Data Viewers
Edit on GitHub
300
200
50
0
−200
−100
0
100
200
pan/zoom

Building Custom Data Viewers

Glue's standard data viewers (scatter plots, images, histograms) are useful in a wide variety of data exploration settings. However, they represent a *tiny* fraction of the ways to view a particular dataset. For this reason, Glue provides a simple mechanism for creating custom visualizations using

Creating a `custom data viewer` requires writing a little bit of Matplotlib code but involves little to no GUI programming. The next several sections illustrate how to build a custom data viewer by

INTEGRATION



Home

Environments

Learning

Community

Documentation

Developer Blog

Feedback

My Applications

glueviz

↗ 0.8.2

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Launch

notebook

↗ 4.2.1

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch

qtconsole

4.2.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch

spyder

↗ 2.3.8

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch

vaex

1.0.0b3

Launch

orange3

3.3.8

Install

xonsh

0.4.7

Install

python™

SciPy

jupyter

astropy

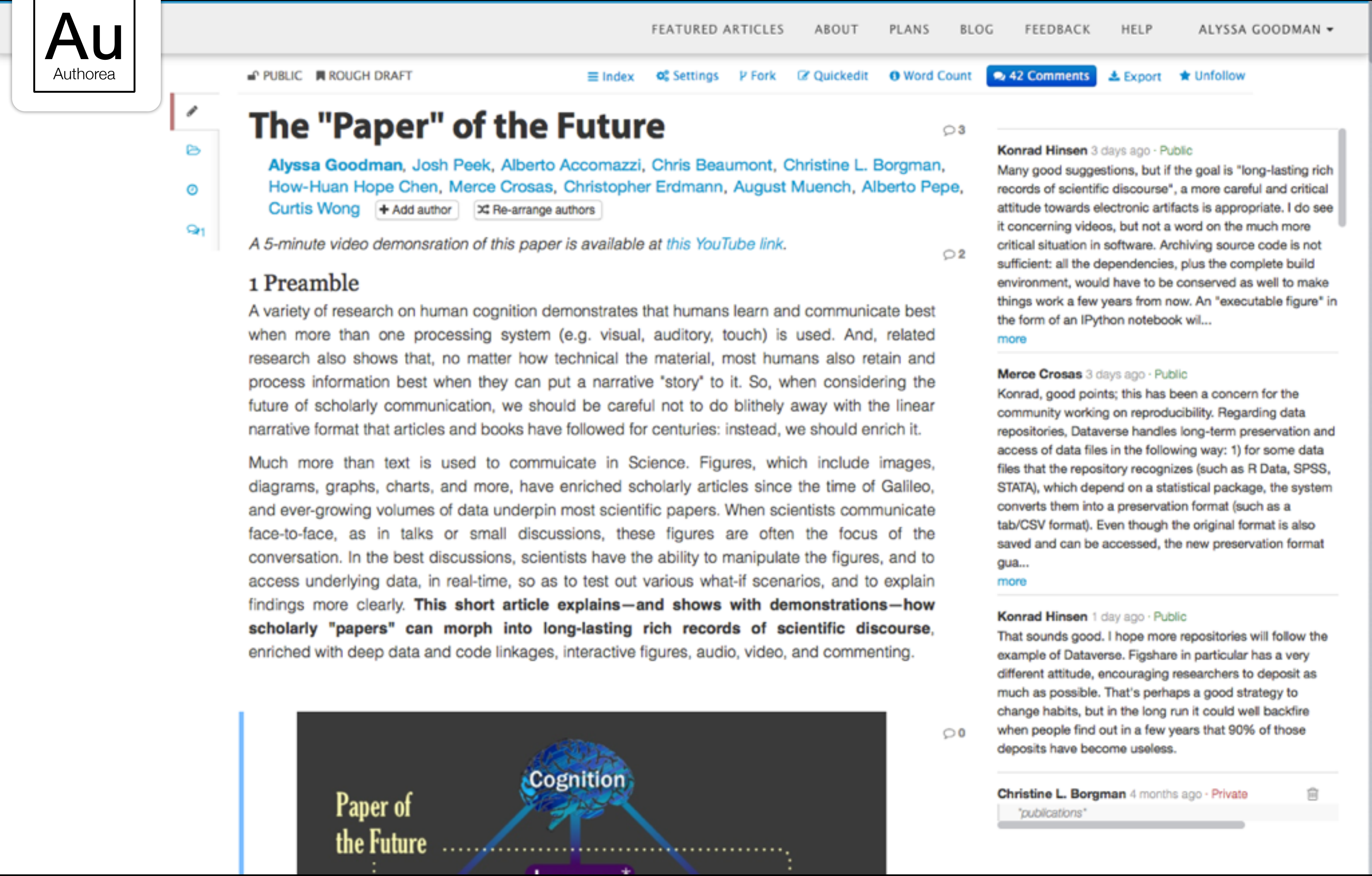
A Community Python Library for Astronomy

JavaScript

d3.js

plotly

HTML 5



The "Paper" of the Future

Alyssa Goodman, Josh Peek, Alberto Accomazzi, Chris Beaumont, Christine L. Borgman, How-Huan Hope Chen, Merce Crosas, Christopher Erdmann, August Muench, Alberto Pepe, Curtis Wong

A 5-minute video demonsration of this paper is available at this YouTube link.

1 Preamble

A variety of research on human cognition demonstrates that humans learn and communicate best when more than one processing system (e.g. visual, auditory, touch) is used. And, related research also shows that, no matter how technical the material, most humans also retain and process information best when they can put a narrative "story" to it. So, when considering the future of scholarly communication, we should be careful not to do blithely away with the linear narrative format that articles and books have followed for centuries: instead, we should enrich it.

Much more than text is used to commuicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks or small discussions, these figures are often the focus of the conversation. In the best discussions, scientists have the ability to manipulate the figures, and to access underlying data, in real-time, so as to test out various what-if scenarios, and to explain findings more clearly. **This short article explains—and shows with demonstrations—how scholarly "papers" can morph into long-lasting rich records of scientific discourse**, enriched with deep data and code linkages, interactive figures, audio, video, and commenting.

Konrad Hinsen 3 days ago · Public
Many good suggestions, but if the goal is "long-lasting rich records of scientific discourse", a more careful and critical attitude towards electronic artifacts is appropriate. I do see it concerning videos, but not a word on the much more critical situation in software. Archiving source code is not sufficient: all the dependencies, plus the complete build environment, would have to be conserved as well to make things work a few years from now. An "executable figure" in the form of an IPython notebook wil...
more

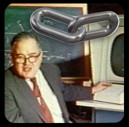
Merce Crosas 3 days ago · Public
Konrad, good points; this has been a concern for the community working on reproducibility. Regarding data repositories, Dataverse handles long-term preservation and access of data files in the following way: 1) for some data files that the repository recognizes (such as R Data, SPSS, STATA), which depend on a statistical package, the system converts them into a preservation format (such as a tab/CSV format). Even though the original format is also saved and can be accessed, the new preservation format gua...
more

Konrad Hinsen 1 day ago · Public
That sounds good. I hope more repositories will follow the example of Dataverse. Figshare in particular has a very different attitude, encouraging researchers to deposit as much as possible. That's perhaps a good strategy to change habits, but in the long run it could well backfire when people find out in a few years that 90% of those deposits have become useless.

Christine L. Borgman 4 months ago · Private
"publications"

d3po/Authorea: Peek, Price-Whelan, Pepe, Beaumont, Borkin, Newton; PoF: Goodman, Peek; WWT: Wong, Fay et al.; Astrometry.net: Hogg, Lang, Roweis et al.

[video, demos]

LINKED-VIEW  VISUALIZATION OF HIGH-DIMENSIONAL DATA IN GLUE

Alyssa A. Goodman

Harvard-Smithsonian Center for Astrophysics & Radcliffe Institute

with Chris Beaumont, Michelle Borkin, Penny Qian & Tom Robitaille

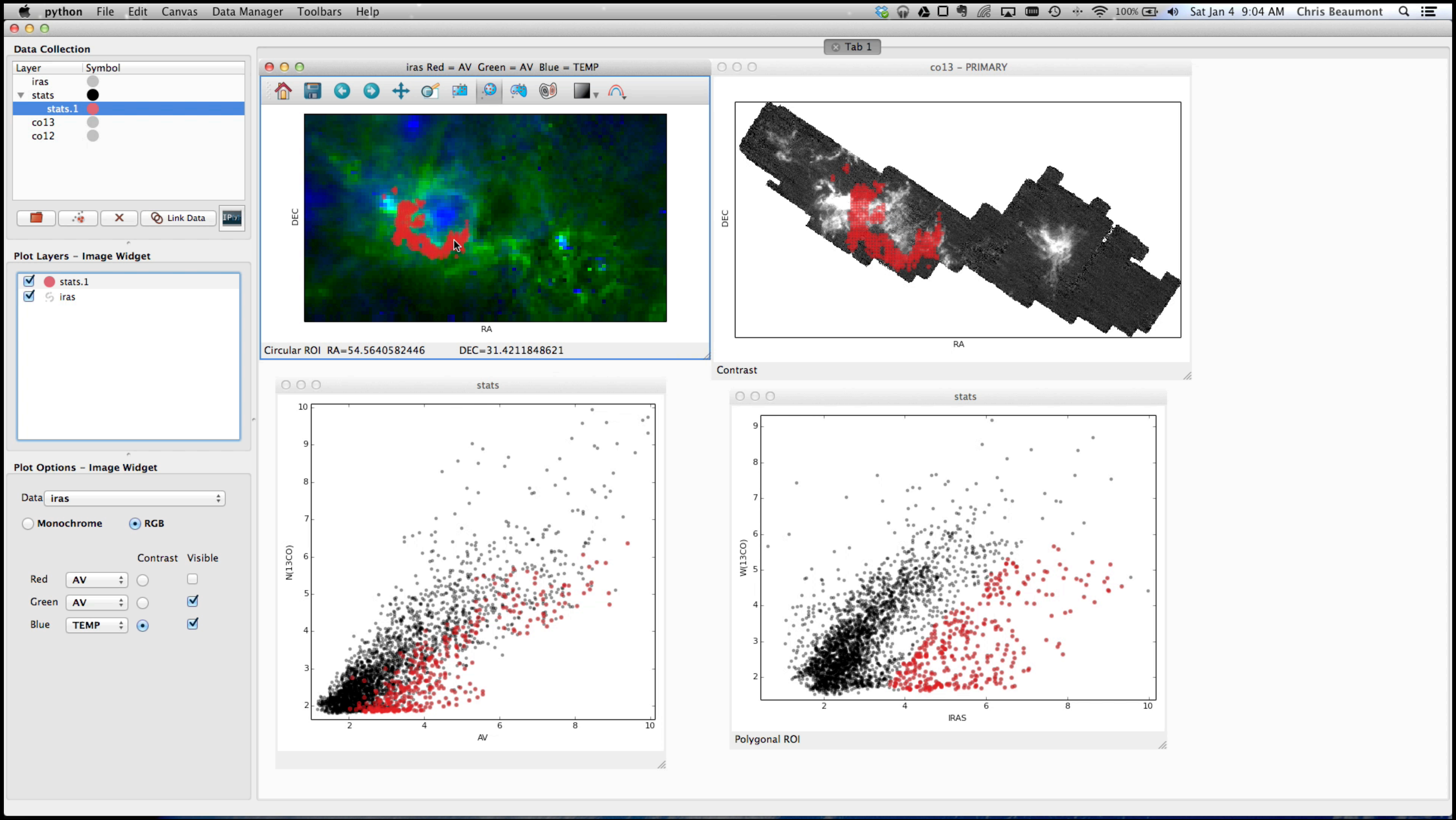
@aagie
@glueviz
@astrofrog

glueviz.org
github.com/glue-viz
Tom Robitaille, lead developer

James Webb
Space Telescope

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE


python File Edit Canvas Data Manager Toolbars Help
100% Sat Jan 4 9:04 AM Chris Beaumont
Tab 1
iras Red = AV Green = AV Blue = TEMP
DEC
RA
co13 – PRIMARY
DEC
RA
Contrast
stats
N(13CO)
AV
2
4
6
8
10
stats
W(13CO)
IRAS
Polygonal ROI
Link Data

Data Collection

Layer	Symbol
iras	●
▾ stats	●
stats.1	●
co13	●
co12	●

Plot Layers – Image Widget

☑ ● stats.1
☑ iras

Plot Options – Image Widget

Data iras

○ Monochrome ● RGB

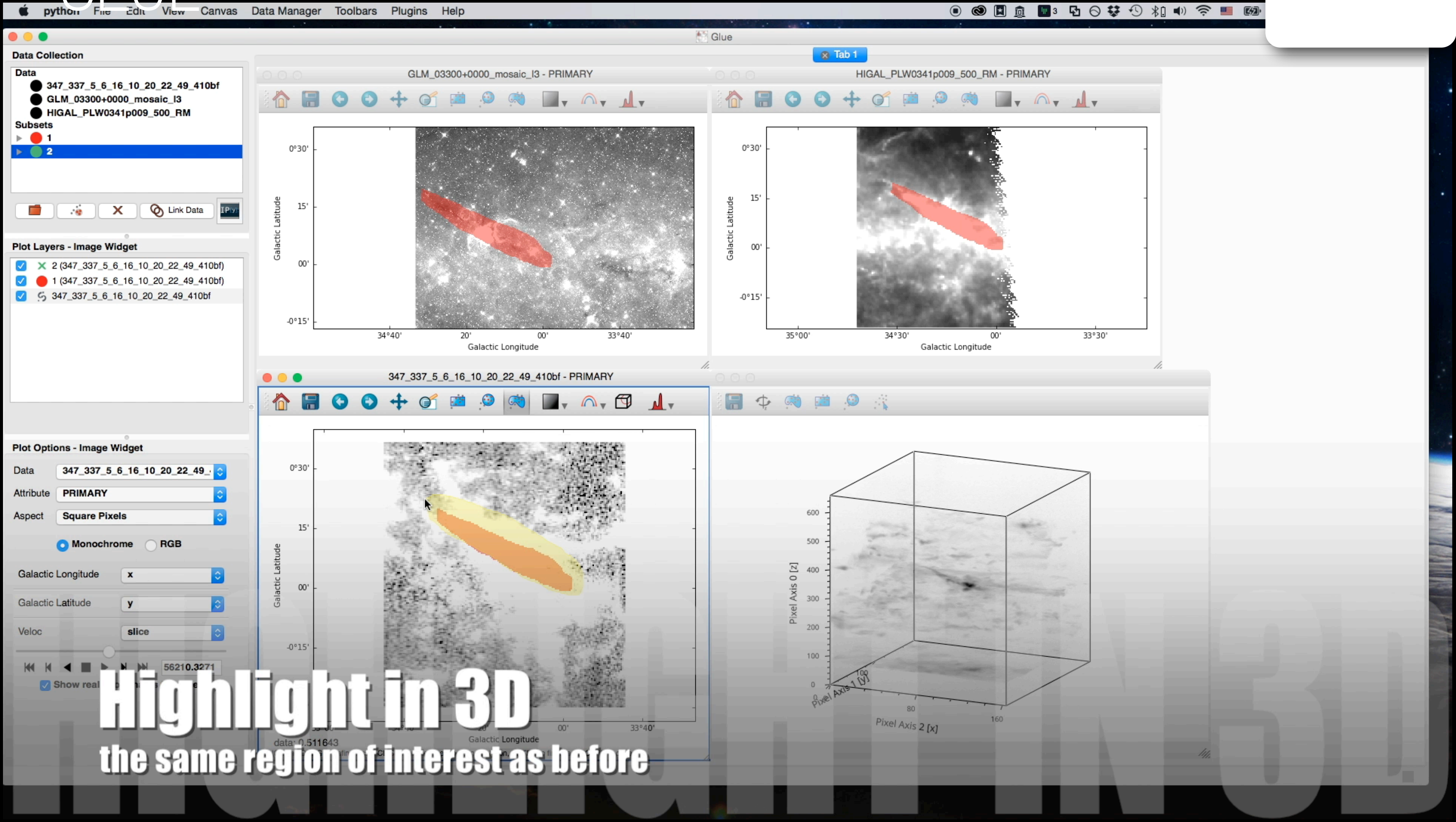
		Contrast	Visible
Red	AV	○	□
Green	AV	○	☑
Blue	TEMP	●	☑

Circular ROI RA=54.5640582446 DEC=31.4211848621

Christopher Beaumont, w/A. Goodman, T. Robitaille & M. Borkin

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



Data Collection

Data
- 347_337_5_6_16_10_20_22_49_410bf
- GLM_03300+0000_mosaic_I3
- HIGAL_PLW0341p009_500_RM

Subsets
- 🟥 1
- 🟩 2

Plot Layers - Image Widget

- ☑ 2 (347_337_5_6_16_10_20_22_49_410bf)
- ☑ 1 (347_337_5_6_16_10_20_22_49_410bf)
- ☑ 347_337_5_6_16_10_20_22_49_410bf

Plot Options - Image Widget

Data: 347_337_5_6_16_10_20_22_49_
Attribute: PRIMARY
Aspect: Square Pixels
● Monochrome ○ RGB
Galactic Longitude: x
Galactic Latitude: y
Veloc: slice
56210.3271
☑ Show real

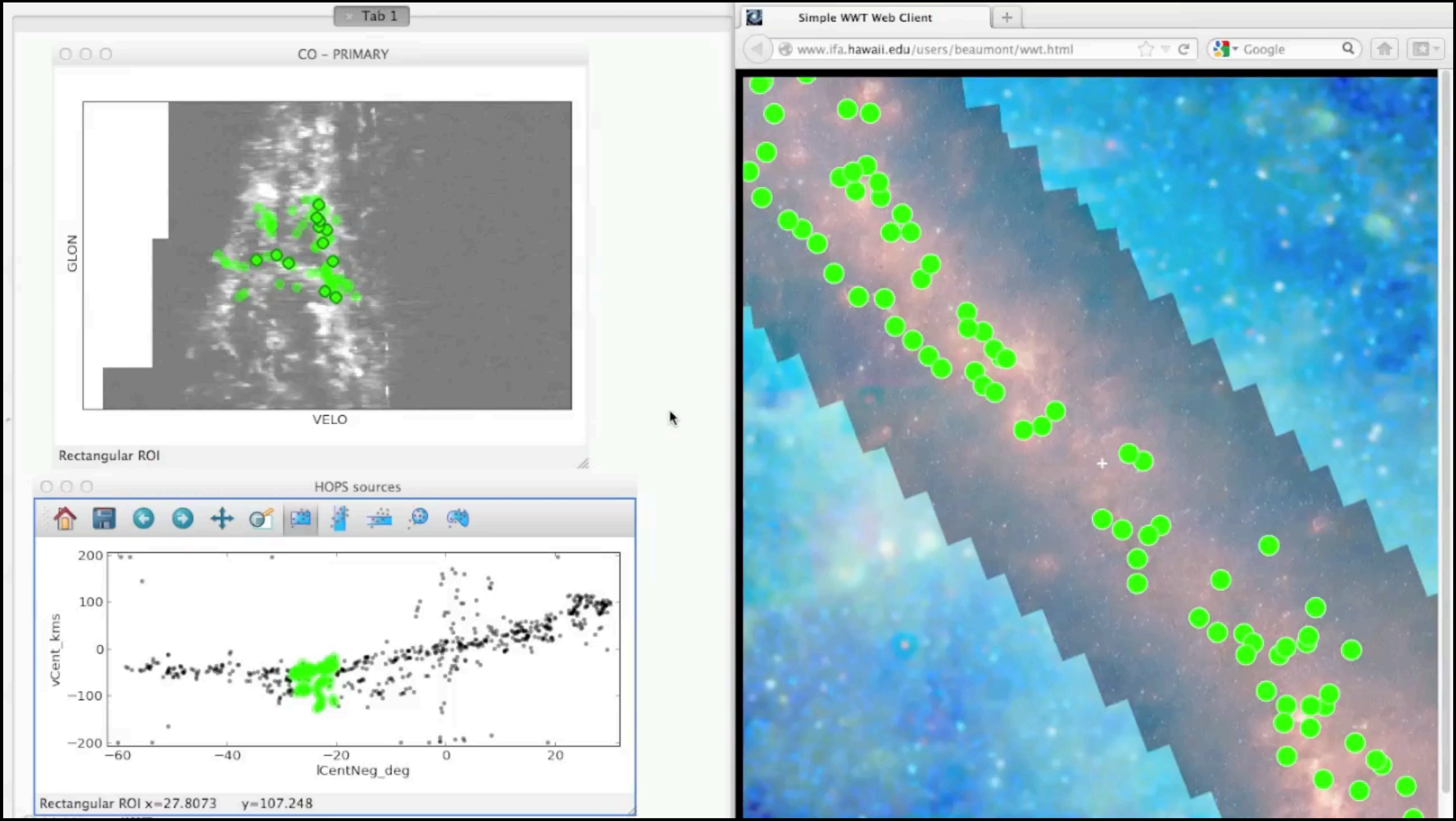
Highlight in 3D

the same region of interest as before

video by Penny Qian, wth Catherine Zucker, graduate students
glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

glue

WorldWide Telescope



Tab 1

CO – PRIMARY

Rectangular ROI

HOPS sources

Rectangular ROI x=27.8073 y=107.248

Simple WWT Web Client

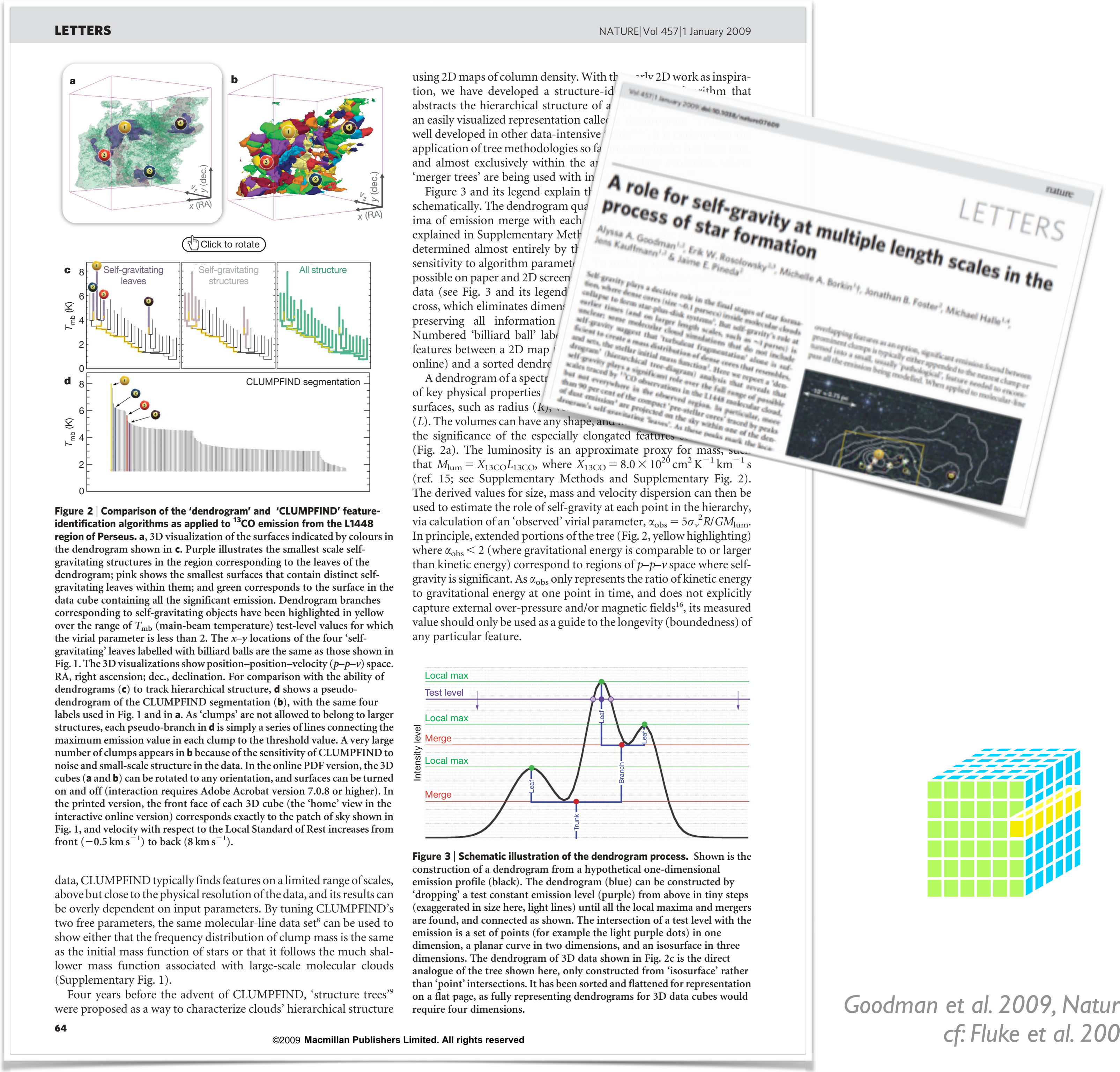
www.ifa.hawaii.edu/users/beaumont/wwt.html

Video courtesy of Chris Beaumont

2009

3D PDF

HIGH-DIMENSIONAL DATA IN A "PAPER"



LETTERS

NATURE|Vol 457|1 January 2009

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 th arly 2D work as inspira-
tion, we have developed a structure-id rithm that
abstracts the hierarchical structure of a
an easily visualized representation calle
well developed in other data-intensive
application of tree methodologies so fa
and almost exclusively within the a
'merger trees' are being used with in

Figure 3 and its legend explain th
schematically. The dendrogram qua
ima of emission merge with each
explained in Supplementary Meth
determined almost entirely by th
sensitivity to algorithm paramete
possible on paper and 2D screen
data (see Fig. 3 and its legend
cross, which eliminates dimens
preserving all information
Numbered 'billiard ball' lab
features between a 2D map (
online) and a sorted dendro

A dendrogram of a spectr
of key physical properties
surfaces, such as radius (*R*), v
(*L*). The volumes can have any shape, and i
the significance of the especially elongated features
(Fig. 2a). The luminosity is an approximate proxy for mass, suc
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.

64

Vol 457|1 January 2009|doi:10.1038/nature07609

nature

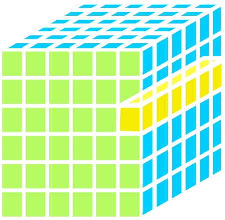
LETTERS

A role for self-gravity at multiple length scales in the process of star formation

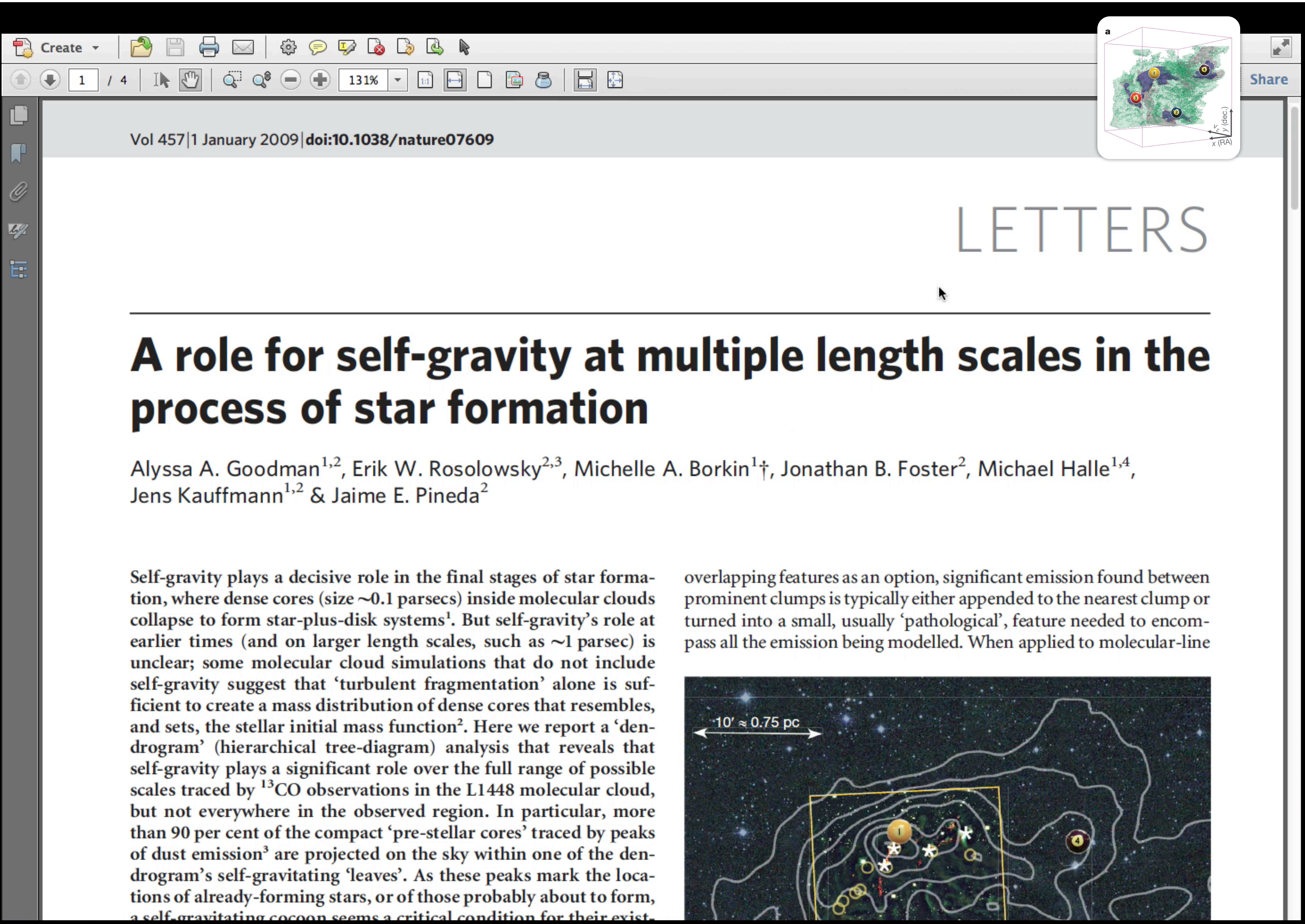
Alyssa A. Goodman[1,2], Erik W. Rosolowsky[2,3], Michelle A. Borkin[1]†, Jonathan B. Foster[2], Michael Halle[1,4], Jens Kauffmann[1,2] & Jaime E. Pineda[2]

Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size ~0.1 parsecs) inside molecular clouds collapse to form star-plus-disk systems[1]. But self-gravity's role at earlier times (and on larger length scales, such as ~1 parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function[2]. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by ^{13}CO observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission[3] are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the loca-

overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line



Goodman et al. 2009, Nature, cf: Fluke et al. 2009



Vol 457|1 January 2009|doi:10.1038/nature07609

LETTERS

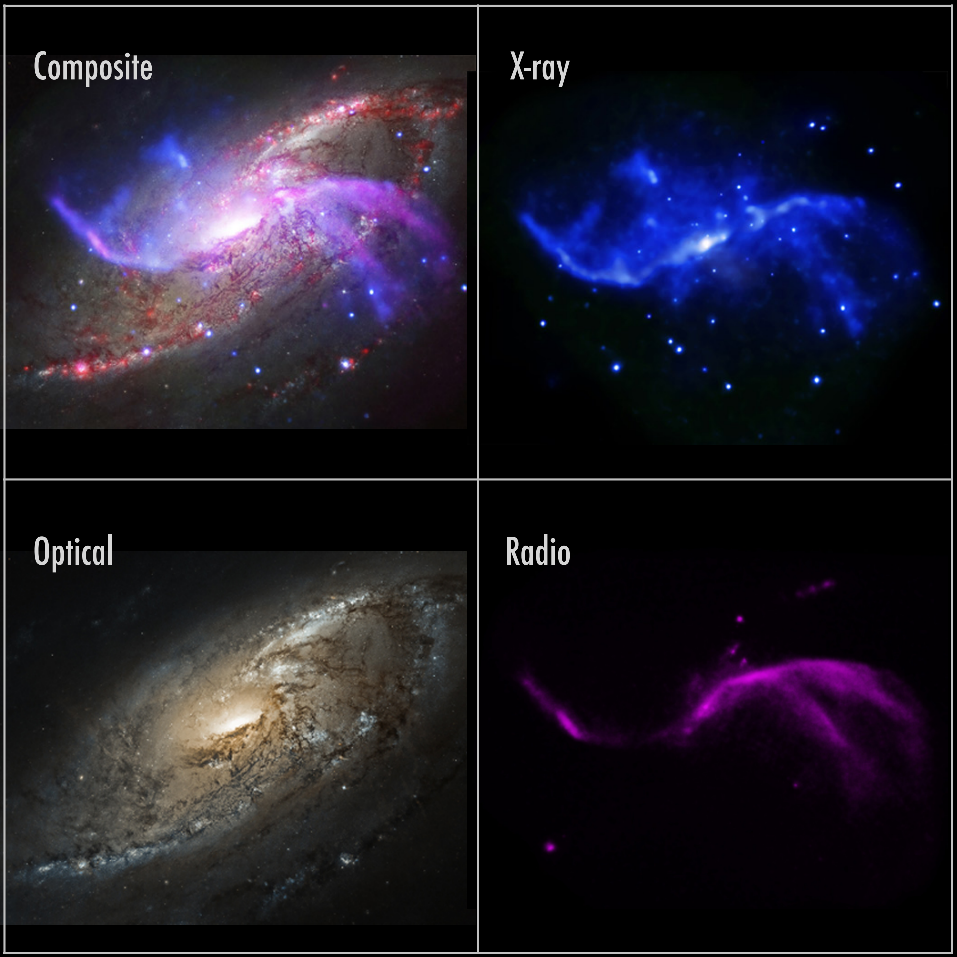
A role for self-gravity at multiple length scales in the process of star formation

Alyssa A. Goodman[1,2], Erik W. Rosolowsky[2,3], Michelle A. Borkin[1]†, Jonathan B. Foster[2], Michael Halle[1,4], Jens Kauffmann[1,2] & Jaime E. Pineda[2]

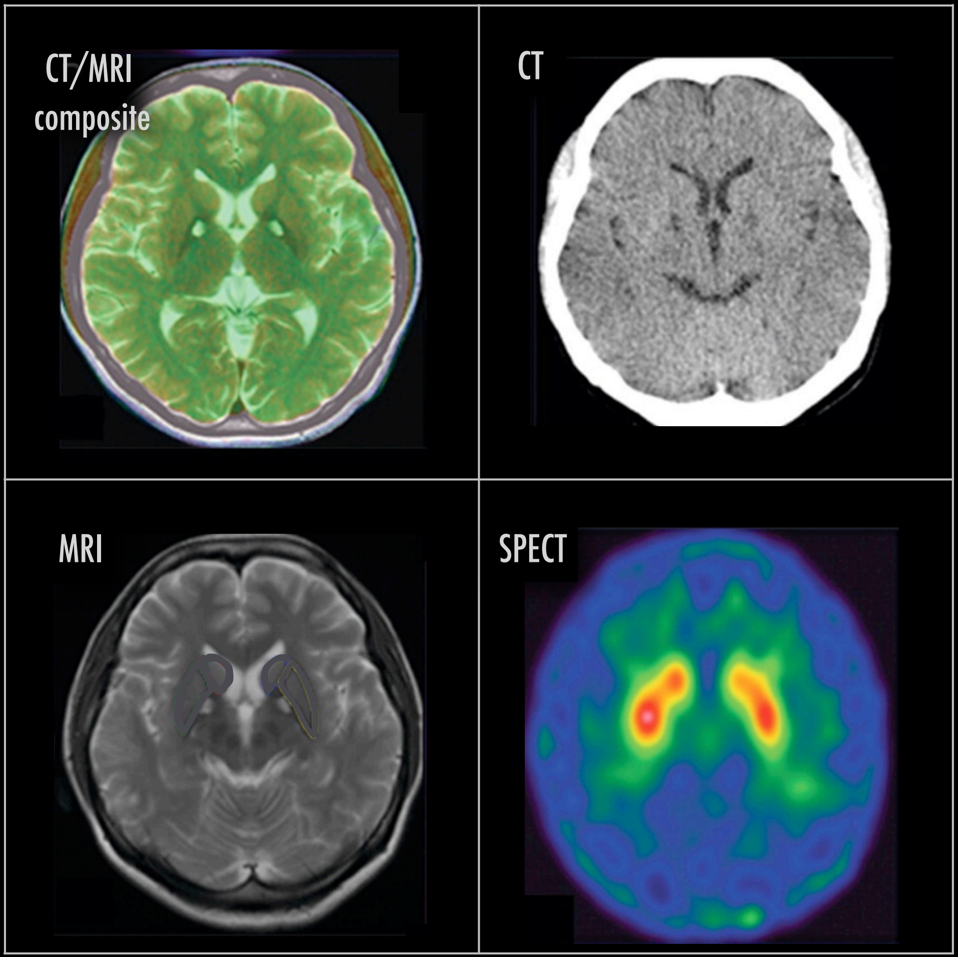
Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size ~0.1 parsecs) inside molecular clouds collapse to form star-plus-disk systems[1]. But self-gravity's role at earlier times (and on larger length scales, such as ~1 parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function[2]. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by ^{13}CO observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission[3] are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the locations of already-forming stars, or of those probably about to form, a self-gravitating cocoon seems a critical condition for their exist-

overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line

ASTRONOMICAL MEDICINE



chandra.harvard.edu/photo/2014/m106/



Chang, et al. 2011, brain.oxfordjournals.org/content/134/12/3632