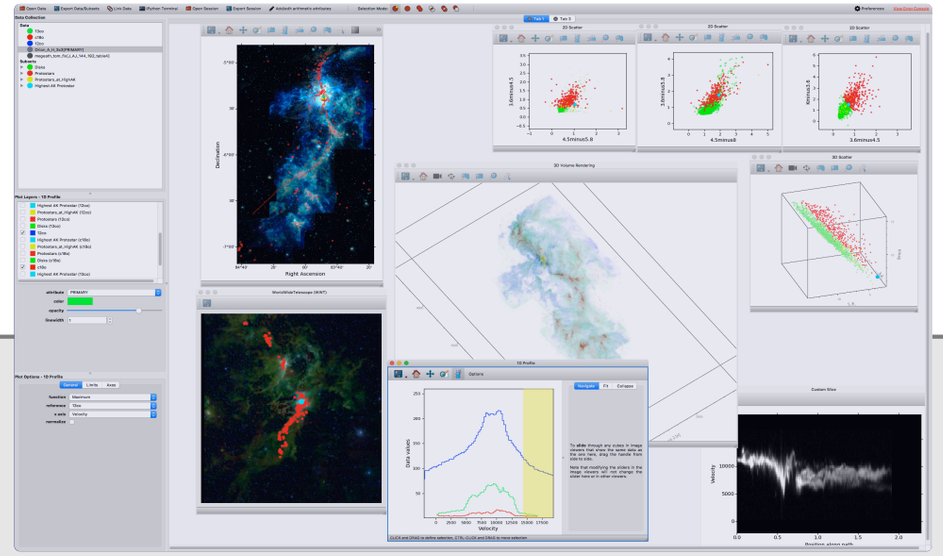


Preview

New Thinking on, and with, Data Visualization

Alyssa A. Goodman, Michelle A. Borkin, Thomas P. Robitaille

As the complexity and volume of datasets have increased along with the capabilities of modular, open-source, easy-to-implement, visualization tools, scientists' need for, and appreciation of, data visualization has risen too. Until recently, scientists thought of the "explanatory" graphics created at a research project's conclusion as "pretty pictures" needed only for journal publication or public outreach. The plots and displays produced during a research project – often intended only for experts – were thought of as a separate category, what we here call "exploratory" visualization. In this view, discovery comes from exploratory visualization, and explanatory visualization is just for communication. Our aim in this paper is to spark conversation amongst scientists, computer scientists, outreach professionals, educators, and graphics and perception experts about how to foster flexible data visualization practices that can facilitate discovery and communication at the same time. We present an example of a new finding made using the glue visualization environment to demonstrate how the border between explanatory and exploratory visualization is easily traversed. The linked-view principles as well as the actual code in glue are easily adapted to astronomy, medicine, and geographical information science – all fields where combining, visualizing, and analyzing several high-dimensional datasets yields insight. Whether or not scientists can use such a flexible "undisciplined" environment to its fullest potential without special training remains to be seen. We conclude with suggestions for improving the training of scientists in visualization practices, and of computer scientists in the iterative, non-workflow-like, ways in which modern science is carried out.

Comments: Submitted as an invited "Perspectives" Paper for PNAS, in conjunction with the 2018 Sackler Colloquium
License: http://arxiv.org/licenses/nonexclusive-distrib/1.0/

Categories

Primary: Instrumentation and Methods for Astrophysics (astro-ph.IM)

Cross lists: Astrophysics | Instrumentation and Methods for Astrophysics | Add

This article is currently submitted.

View Article

Goodman, Borkin & Robitaille 2018, available on arXiv

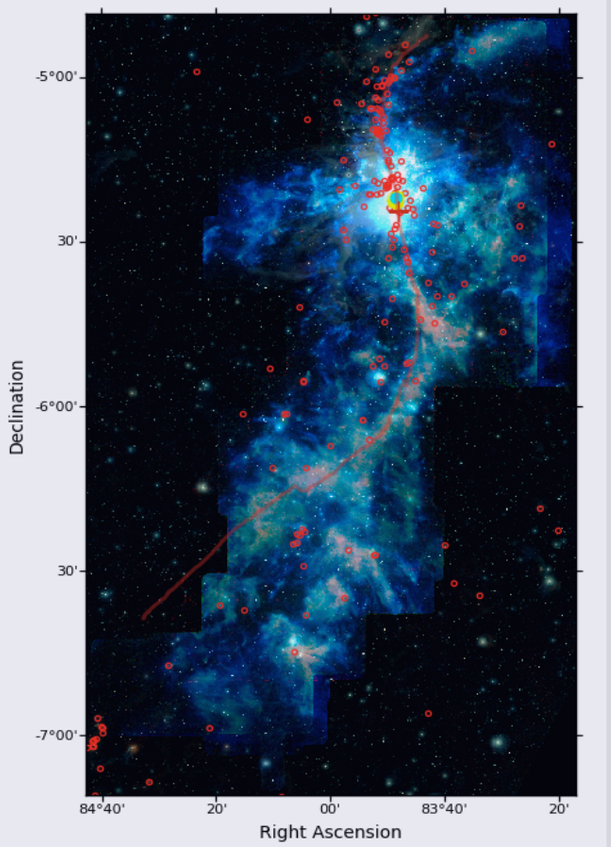
Visualization & the Universe

How and why astronomers, doctors, and you need to work together to understand the world around us...

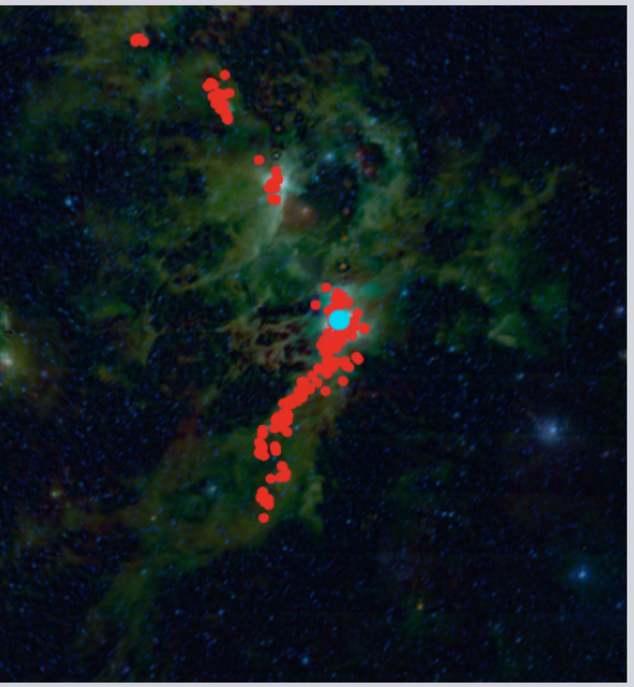
Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics + Radcliffe Insitute for Advanced Study
@AlyssaAGoodman

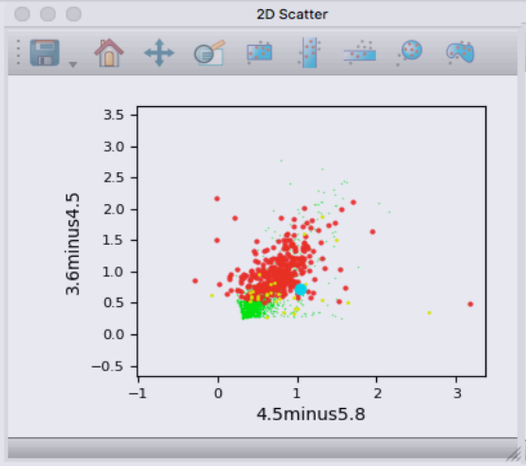
Export Session | Add/edit arithmetic attributes | Selection Mode: | Preferences | View Error Con

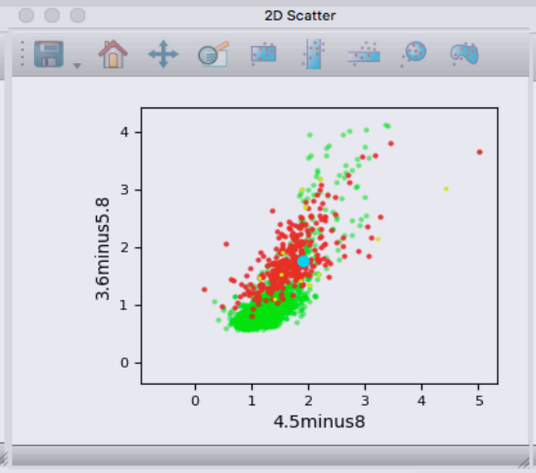
Tab 1 | Tab 3

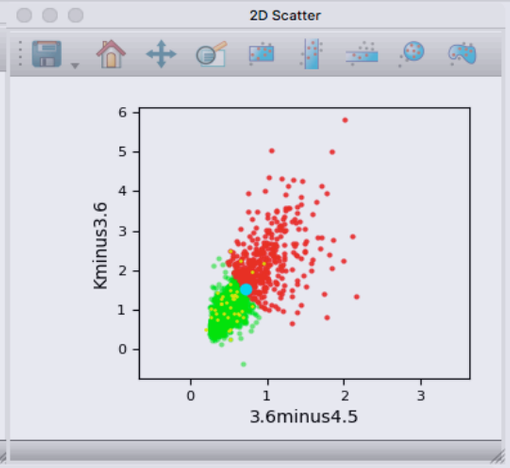


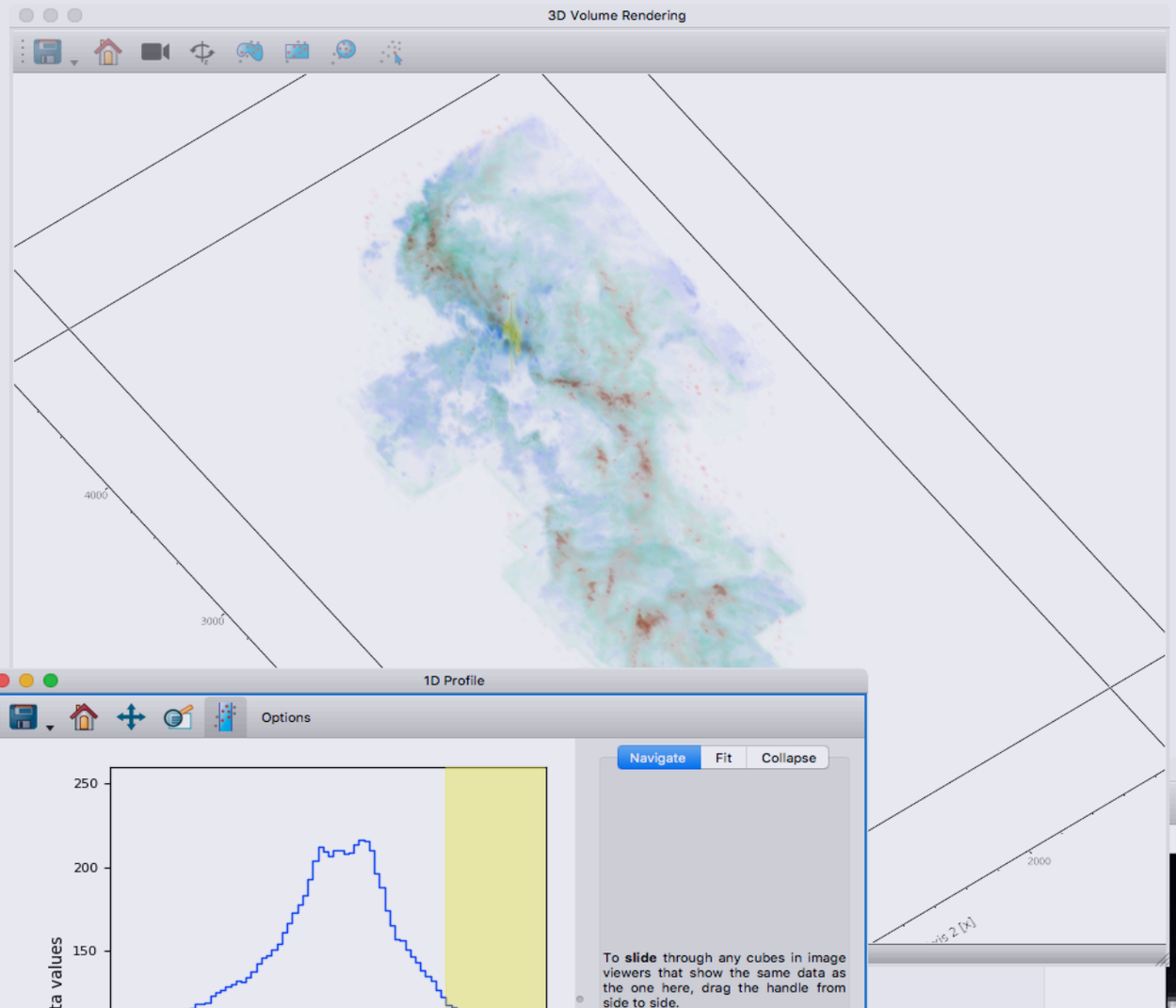
WorldWideTelescope (WWT)

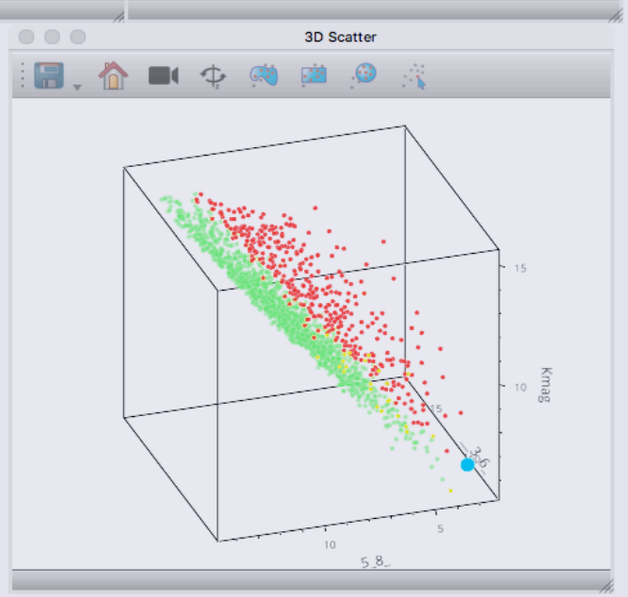


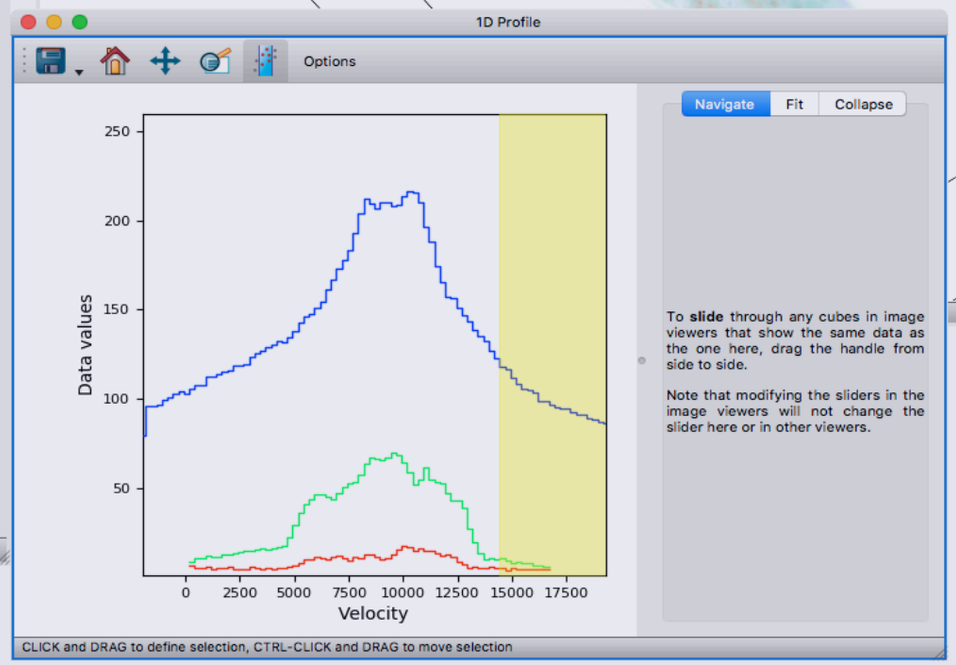




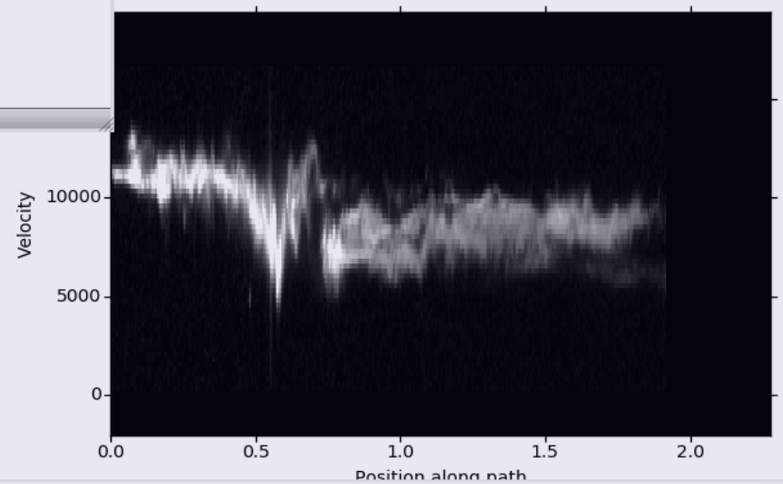






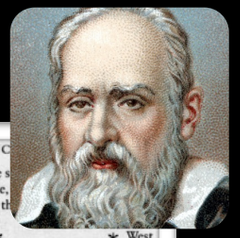


Custom Slice



4 Centuries from Galileo to Galileo

1610



SIDEREUS NUNC

On the third, at the seventh hour, the s
sequence. The eastern one was 1 minute,
the closest western one 2 minutes; and th

East * ○ * * West

10 minutes removed from this one. They were absolutely on the same straight line and of equal magnitude.

On the fourth, at the second hour, there were four stars around Jupiter, two to the east and two to the west, and arranged precisely

East * * ○ * * West

on a straight line, as in the adjoining figure. The easternmost was distant 3 minutes from the next one, while this one was 40 seconds from Jupiter; Jupiter was 4 minutes from the nearest western one, and this one 6 minutes from the westernmost one. Their magnitudes were nearly equal; the one closest to Jupiter appeared a little smaller than the rest. But at the seventh hour the eastern s
30 seconds apart. Jupiter was 2 minutes from the

East ** ○ * *

one, while he was 4 minutes from the next western
one was 3 minutes from the westernmost one. They
and extended on the same straight line along the ecl

On the fifth, the sky was cloudy.

On the sixth, only two stars appeared flanking Ju

East * ○ *

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

On the seventh, two stars stood near Jupiter, bo
arranged in this manner.

1665

PHILOSOPHICAL TRANSACTIONS

1895

ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

VOLUME I JANUARY 1895 NUMBER

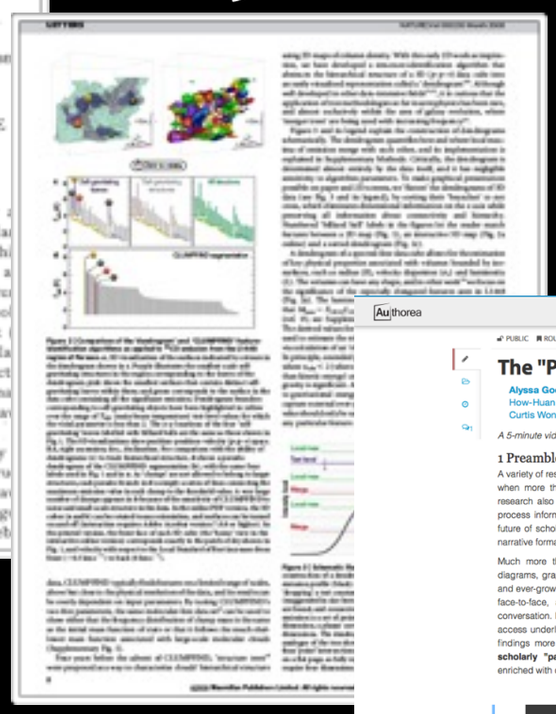
ON THE CONDITIONS WHICH AFFECT THE SPECTRO-PHOTOGRAPHY OF THE SUN.

By Albert A. Michelson.

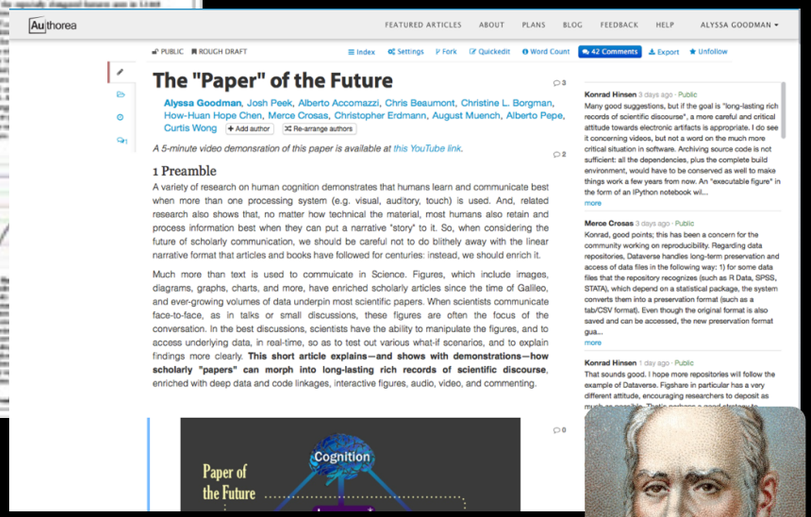
The recent developments in solar spectro-photography
in great measure due to the device originally suggested by Ja
sen and perfected by Hale and Deslandres, by means of wh
a photograph of the Sun's prominences may be obtained at a
time as readily as it is during an eclipse. The essential featu
of this device are the simultaneous movements of the co
mator-slit across the Sun's image, with that of a second slit
the focus of the photographic lens) over a photographic pla
If these relative motions are so adjusted that the same spect
line always falls on the second slit, then a photographic ima
of the Sun will be reproduced by light of this particular wa
length.

Evidently the process is not limited to the photography
the prominences, but extends to all other peculiarities of stru
ure which emit radiations of approximately constant wa
length; and the efficiency of the method depends very larg
upon the *contrast* which can be obtained by the greater endeeb

2009



2015



Authorea

FEATURED ARTICLES ABOUT PLANS BLOG FEEDBACK HELP ALYSSA GOODMAN

PUBLIC ROUGH DRAFT Index Settings Fork Quickedit Word Count 42 Comments Export Unfollow

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 + Add author Re-arrange authors

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.

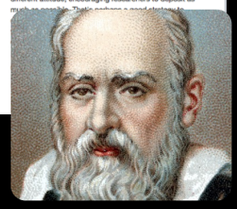
Much more than text is used to commuicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks 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.

Paper of the Future Cognition

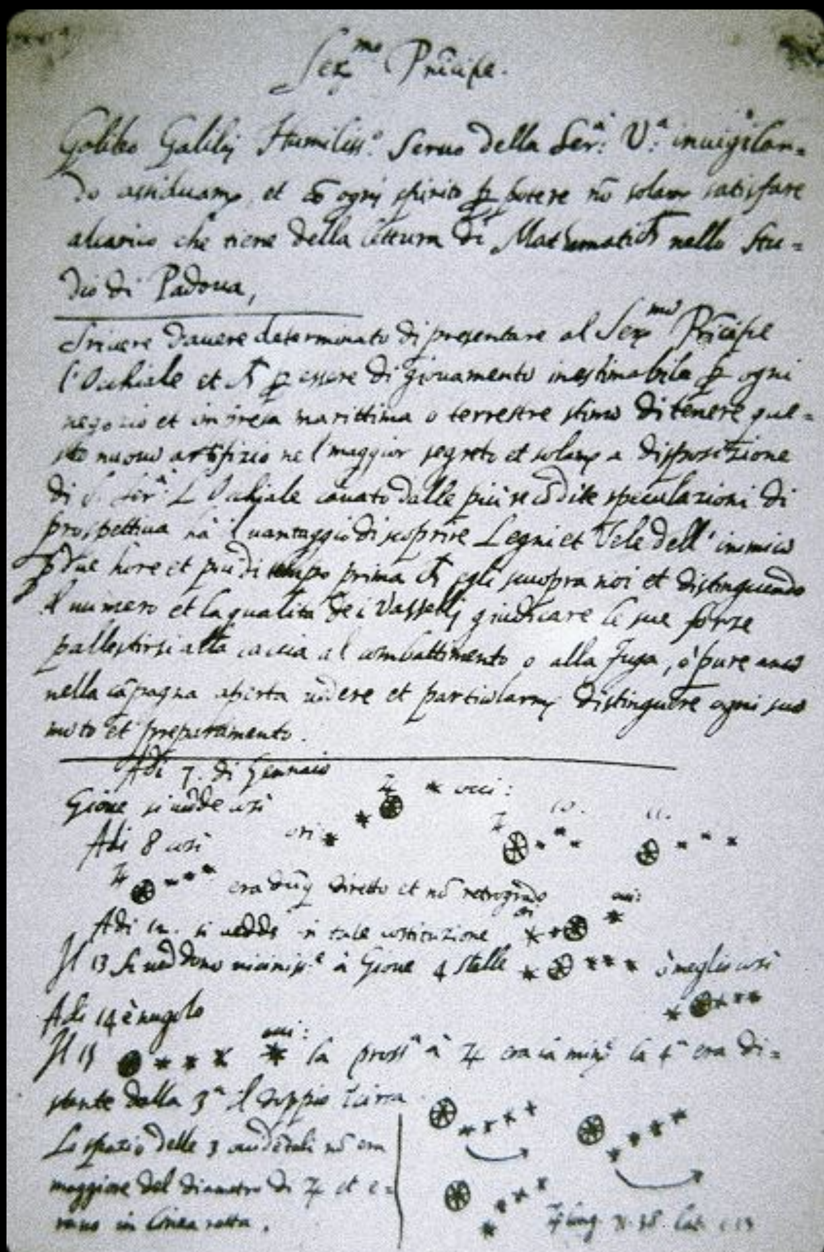
Konrad Hinsen 3 days ago · Public
Many good suggestions, but if the goal is "long-lasting rich records of scientific discourse", a more careful and critical attitude towards electronic artifacts is appropriate. I do see it concerning videos, but not a word on the much more critical situation in software. 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

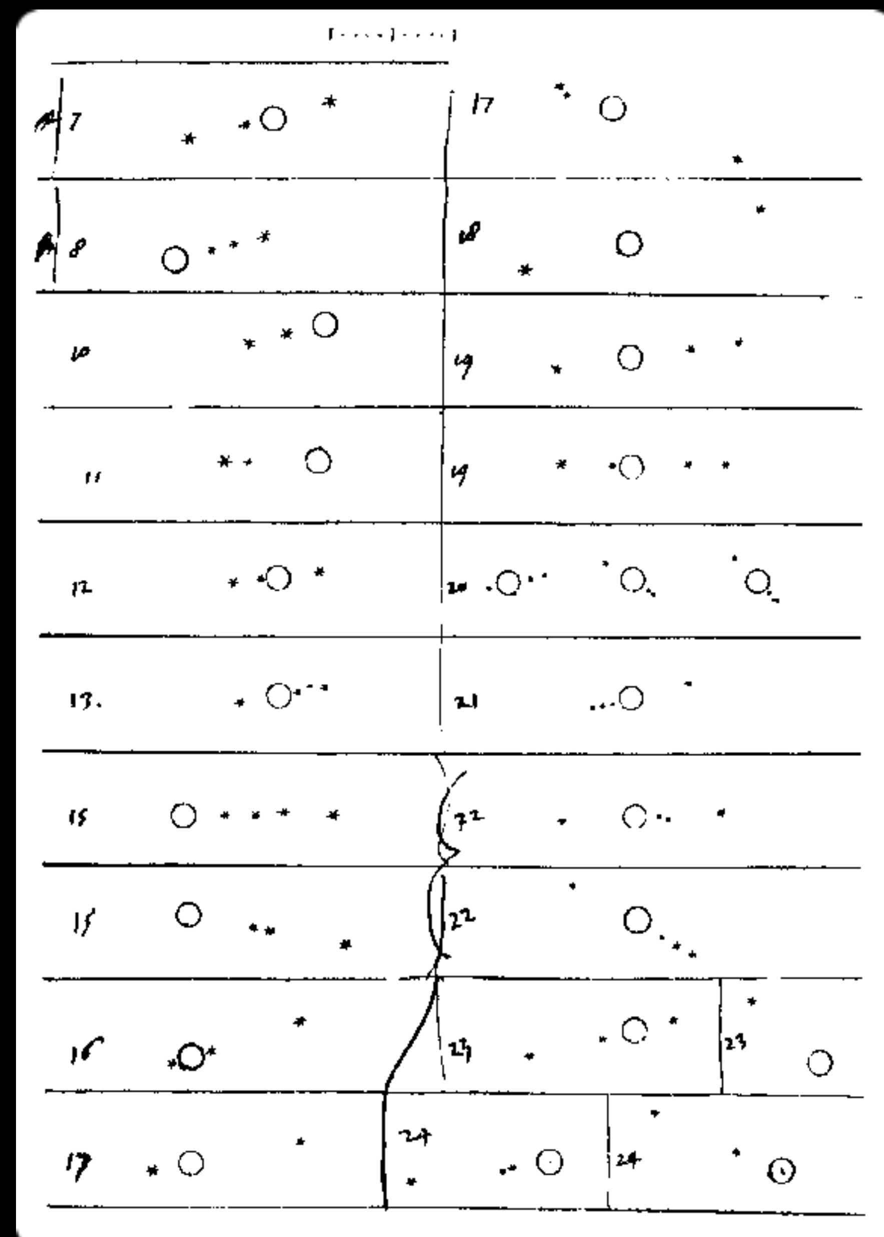


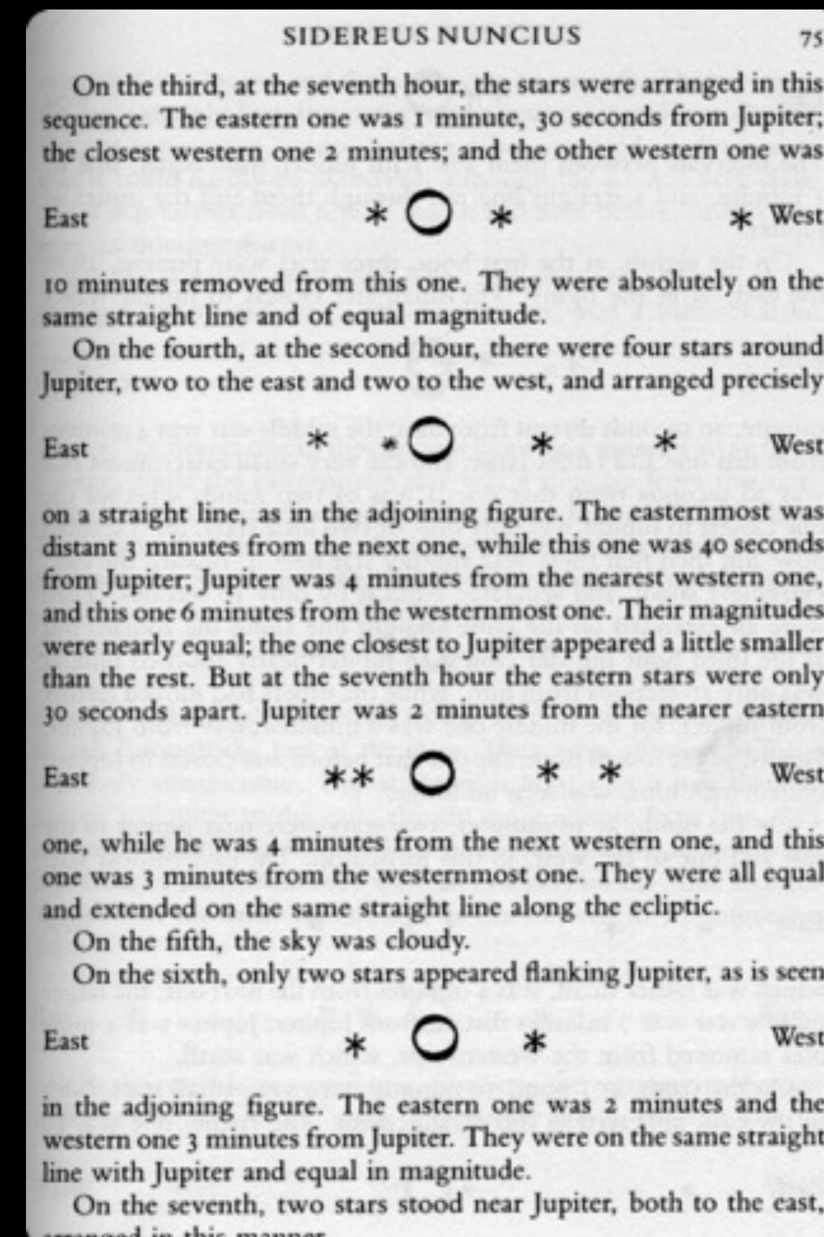
1610: Galileo Discovers Jupiter's Moons


Ser.mo Principe.

Galileo Galilei Humiliss.o Servo della Ser.tà V.a invigilan=do assiduam.e et con ogni spirito per potere non solam.e satisfare al carico che tiene della lettura di Matematiche nello Stu=dio di Padova,

Hora havere determinato di presentare al Ser.mo Principe l'Occhiale et di poterne di giovamento inestimabile per ogni negozio et impresa marittima o terrestre stimo di tenere questo nuovo artifizio nel maggior segreto et solam.e a disposizione di V. Ser.tà L'Occhiale cavato dalle più recondite speculazioni di prospettiva ha 'l vantaggio di scoprire Legni et Vele dell'inimico due hore et più di tempo prima che egli scuopra noi et distinguendo il numero et la qualità de i Vasselli giudicare le sue forze allestirsi alla caccia al combattimento o alla fuga, o pure anco nella campagna aperta videre et particolarm.e distinguere ogni suo moto et preparamento.





On the third, at the seventh hour, the stars were arranged in this sequence. The eastern one was 1 minute, 30 seconds from Jupiter; the closest western one 2 minutes; and the other western one was

East * O * * West

10 minutes removed from this one. They were absolutely on the same straight line and of equal magnitude.

On the fourth, at the second hour, there were four stars around Jupiter, two to the east and two to the west, and arranged precisely

East * *O * * West

on a straight line, as in the adjoining figure. The easternmost was distant 3 minutes from the next one, while this one was 40 seconds from Jupiter; Jupiter was 4 minutes from the nearest western one, and this one 6 minutes from the westernmost one. Their magnitudes were nearly equal; the one closest to Jupiter appeared a little smaller than the rest. But at the seventh hour the eastern stars were only 30 seconds apart. Jupiter was 2 minutes from the nearer eastern

East ** O * * West

one, while he was 4 minutes from the next western one, and this one was 3 minutes from the westernmost one. They were all equal and extended on the same straight line along the ecliptic.

On the fifth, the sky was cloudy.

On the sixth, only two stars appeared flanking Jupiter, as is seen

East * O * West

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

On the seventh, two stars stood near Jupiter, both to the east, arranged in this manner.

Notes for & re-productions of Siderius Nuncius

WorldWide Telescope: Explaining Galileo's Discovery

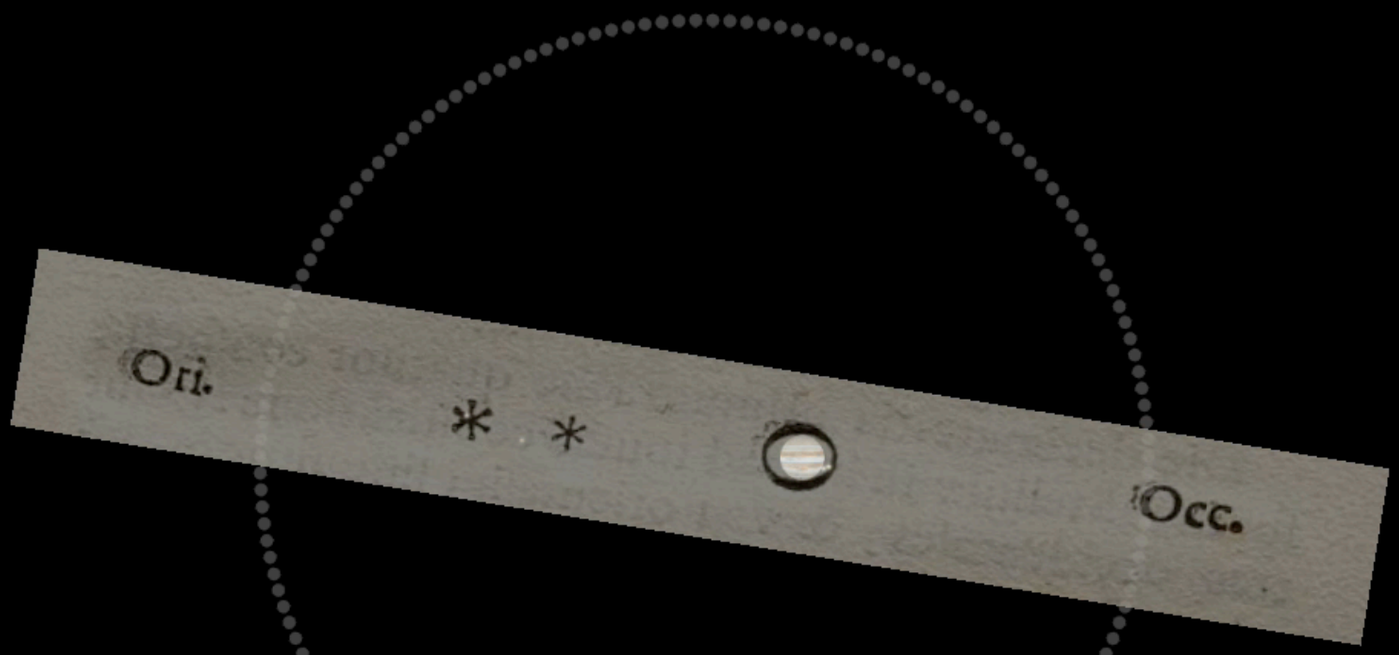
GALILEO'S "NEW ORDER"

Created by Alyssa Goodman, Curtis Wong and Pat Udomprasert,
with advice from Owen Gingerich and David Malin

Galileo's New Order, A WorldWide Telescope Tour by Goodman, Wong & Udomprasert 2010
WWT Software Wong (inventor, MS Research), Fay (architect, MS Research), et al., now open source, hosted by AAS
see wwtambassadors.org for more on WWT Outreach

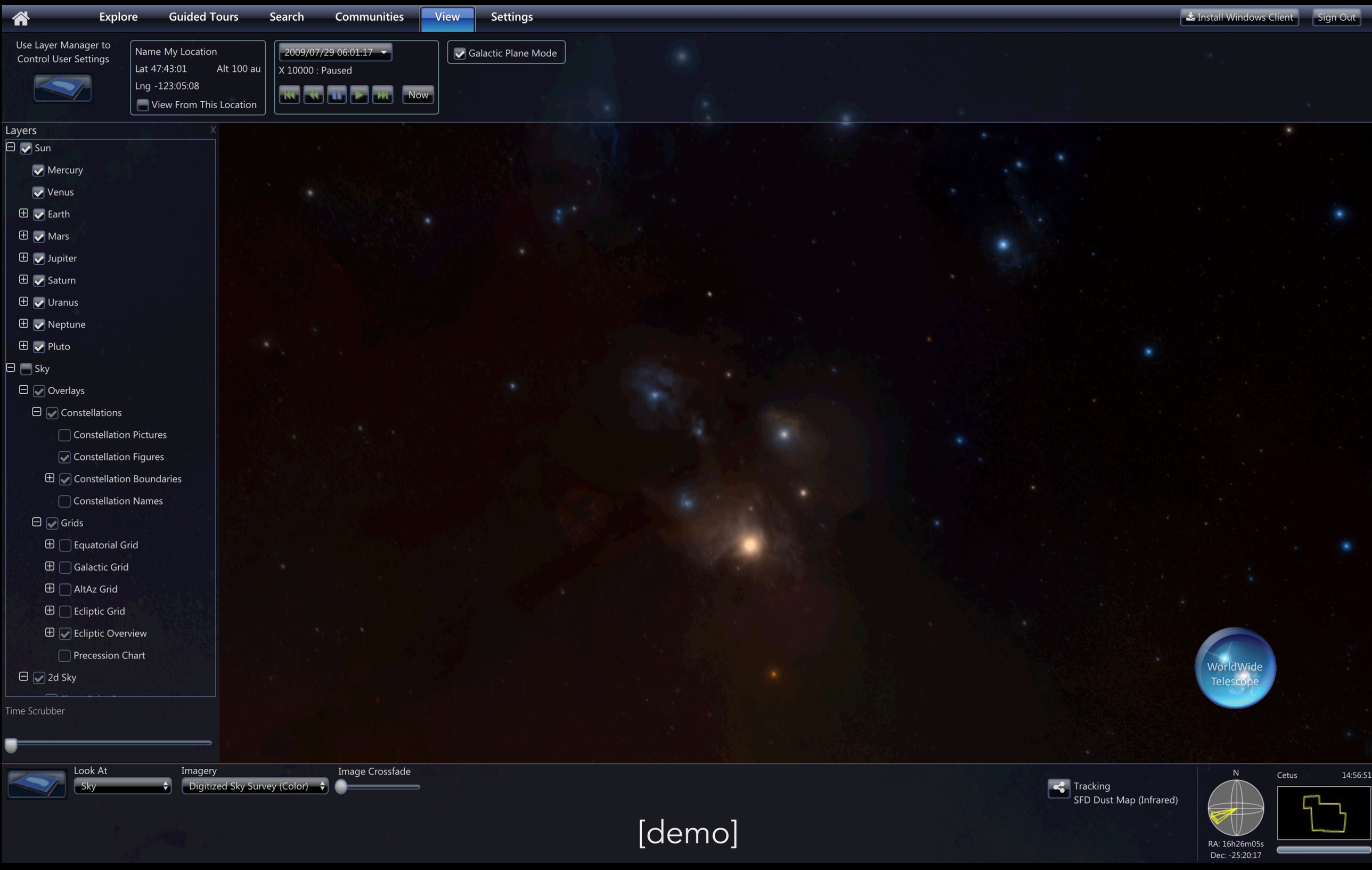
WorldWide Telescope: Explaining Galileo's Discovery

January 11, 1610

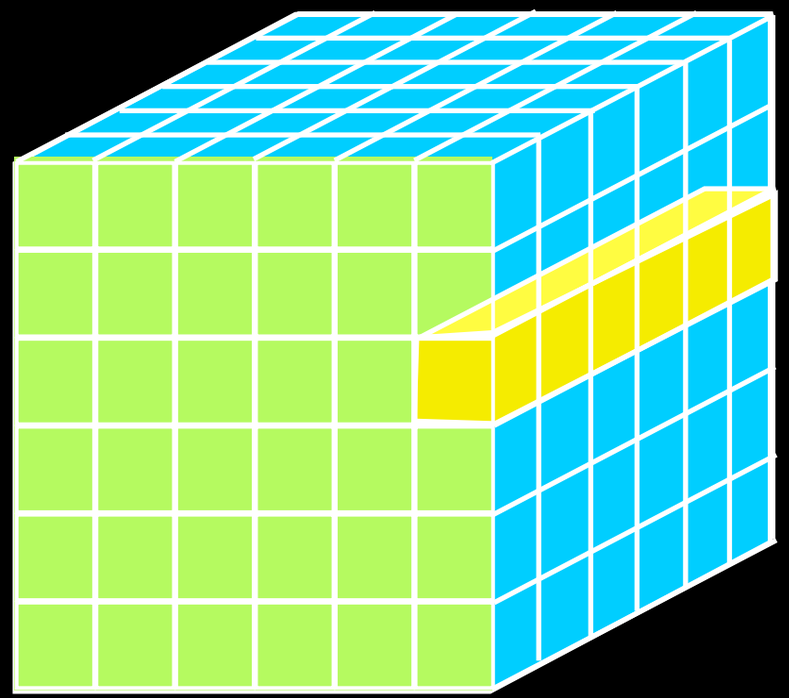


Galileo's New Order, A WorldWide Telescope Tour by Goodman, Wong & Udomprasert 2010
WWT Software Wong (inventor, MS Research), Fay (architect, MS Research), et al., now open source, hosted by AAS
see wwtambassadors.org for more on WWT Outreach

WorldWide Telescope: Exploring the Universe



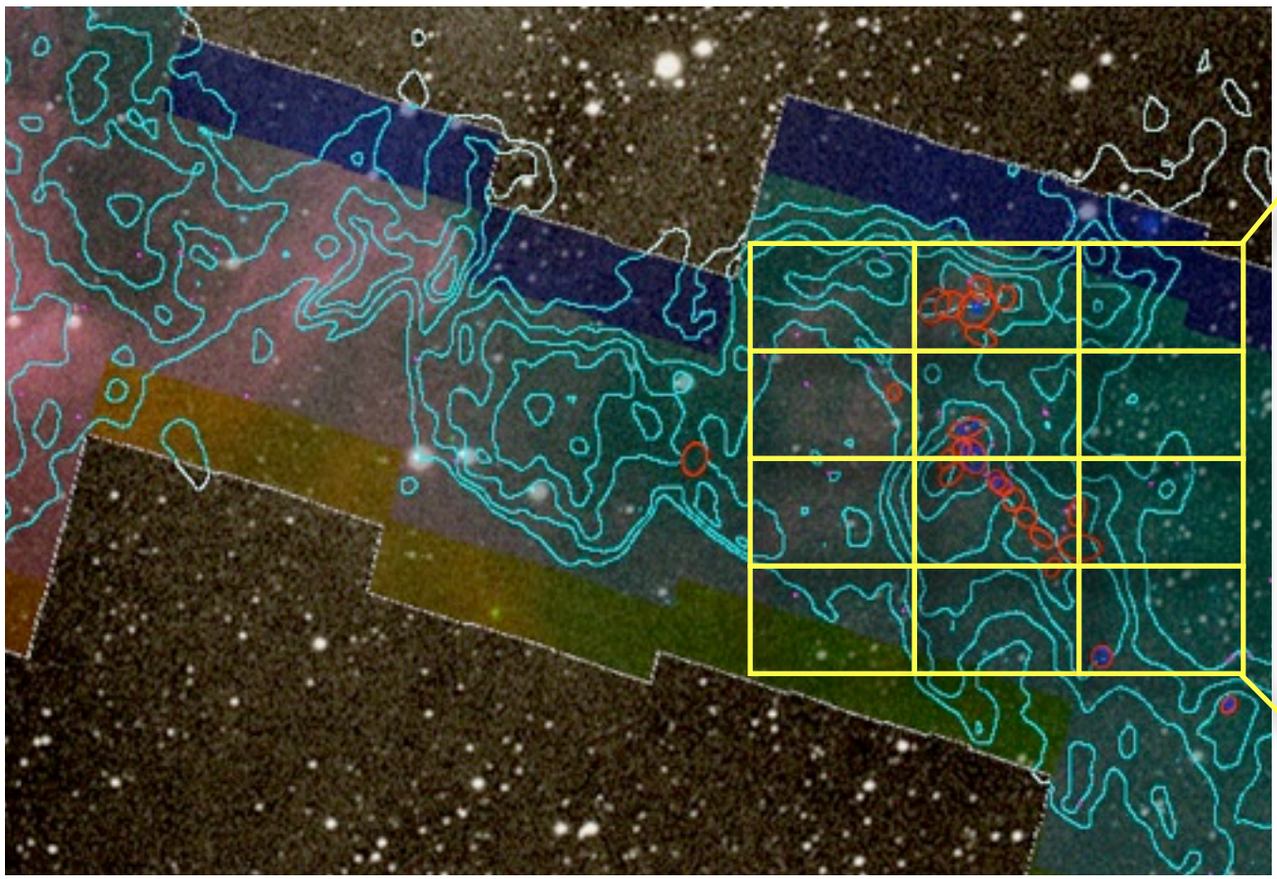
[demo]

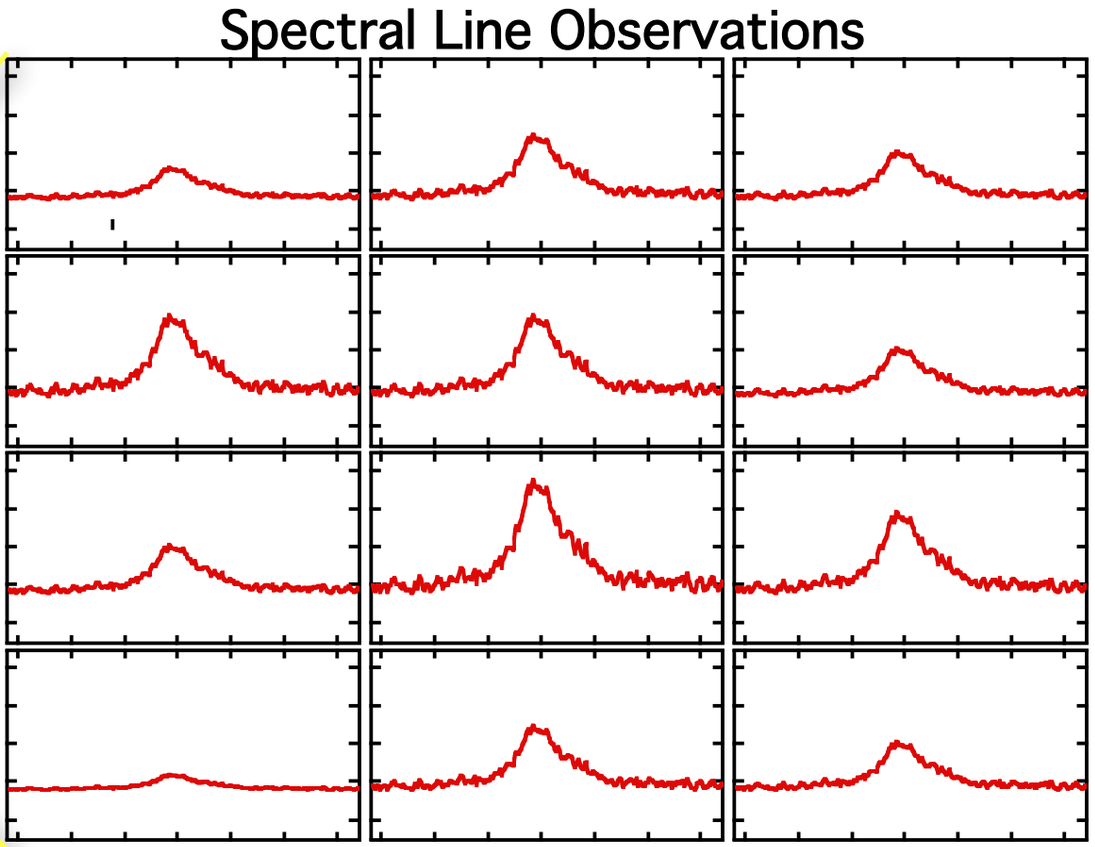


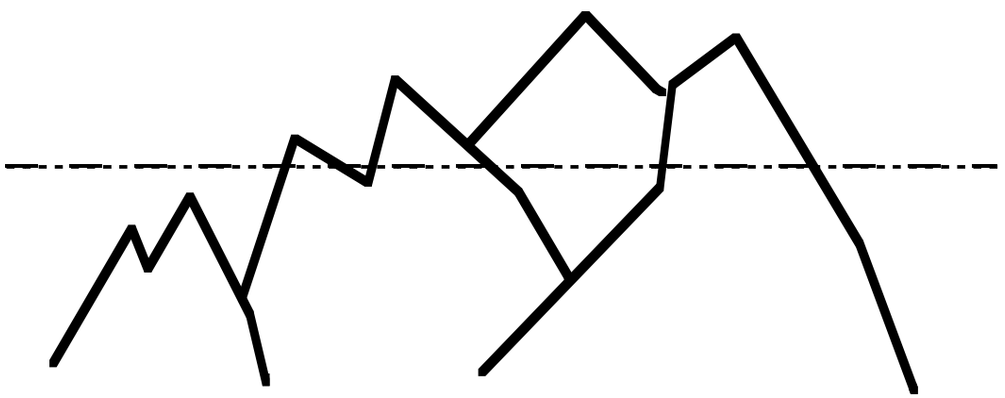
Data, Dimensions, Display

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”

Data, Dimensions, Display





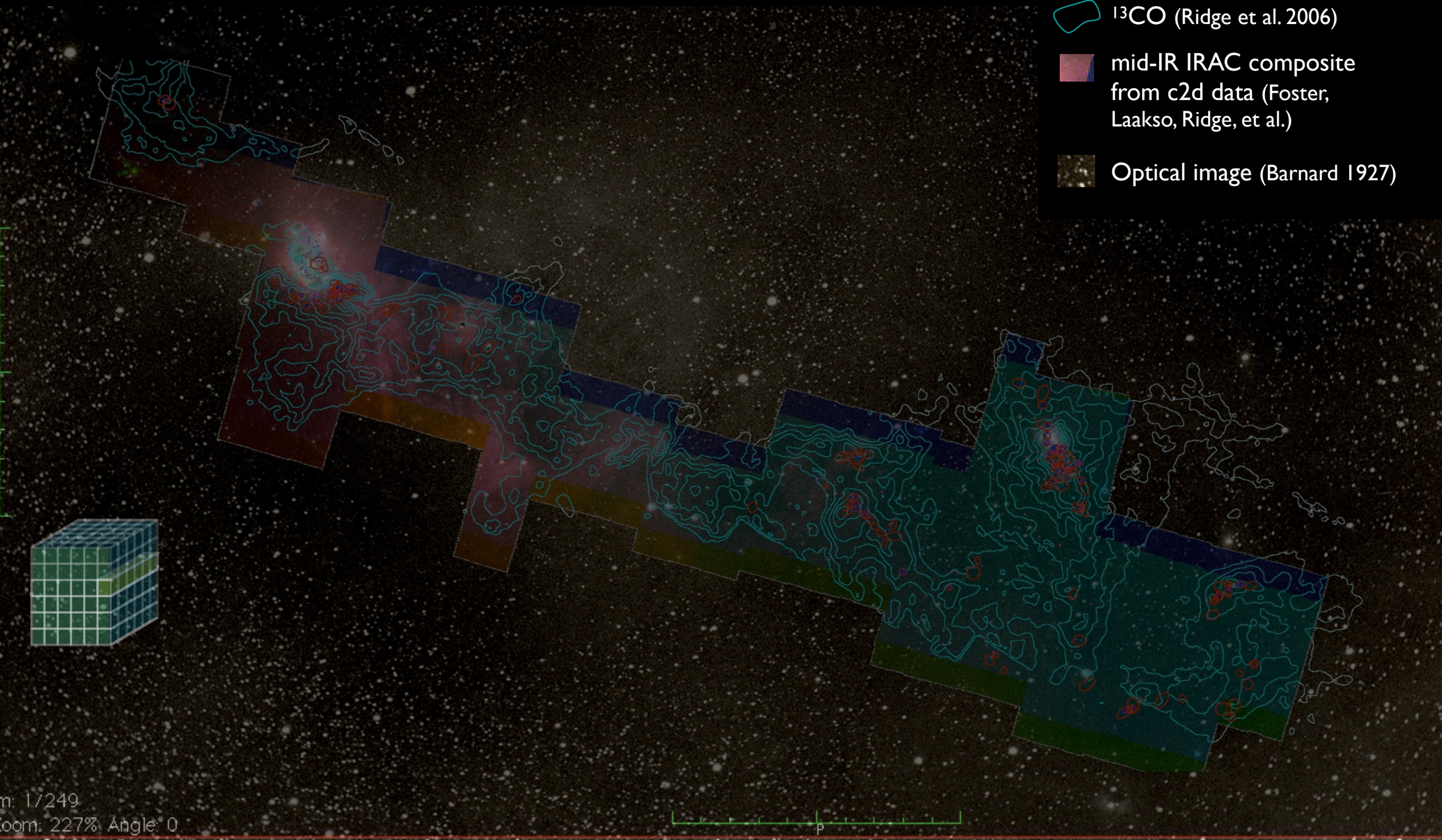


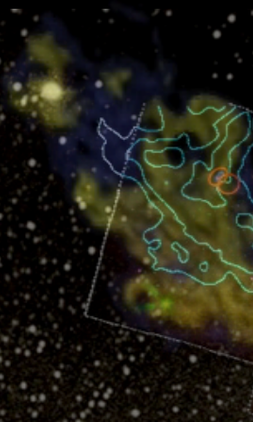
Mountain Range

No loss of information

Loss of 1 dimension

Data, Dimensions, Display

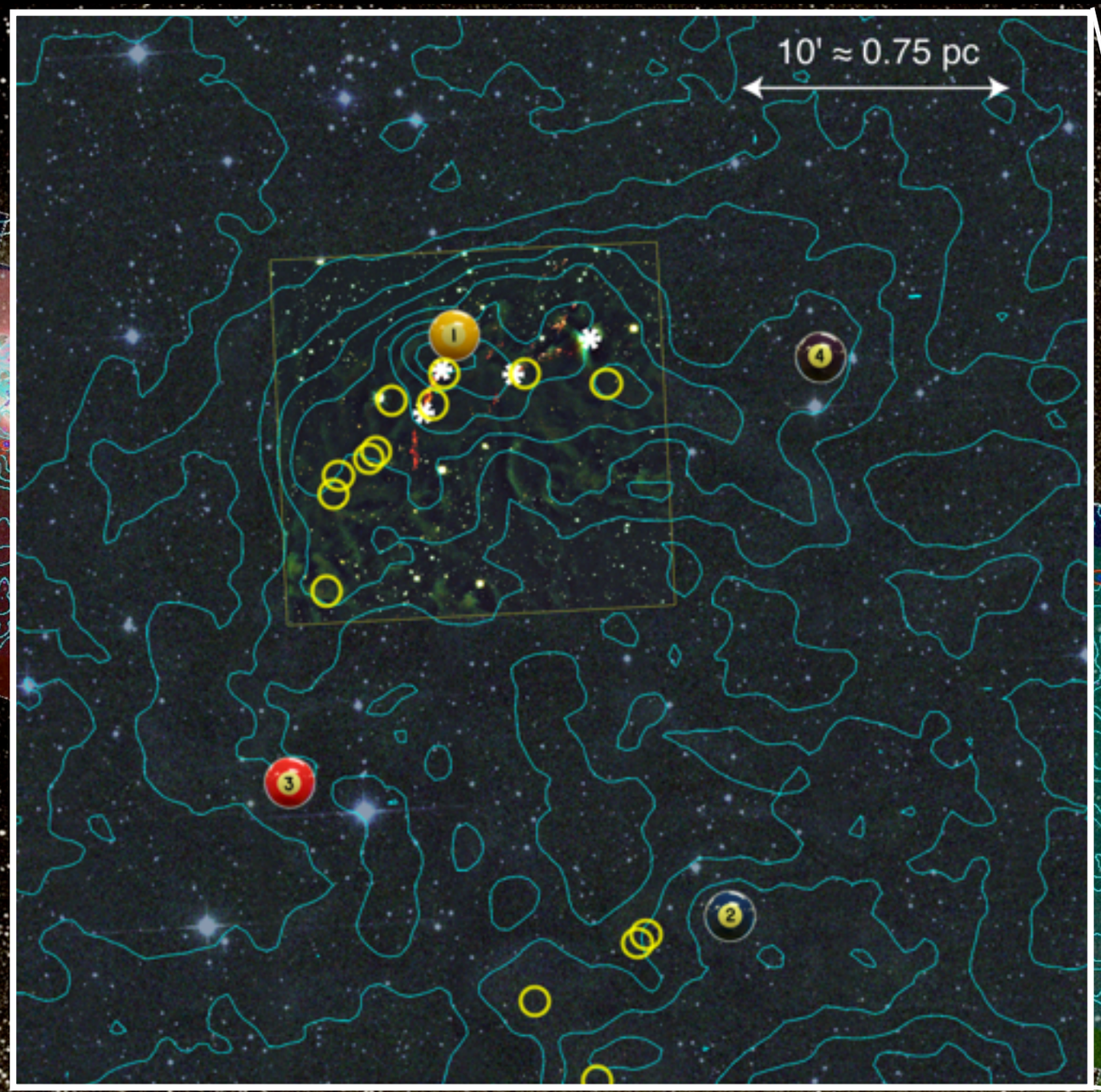


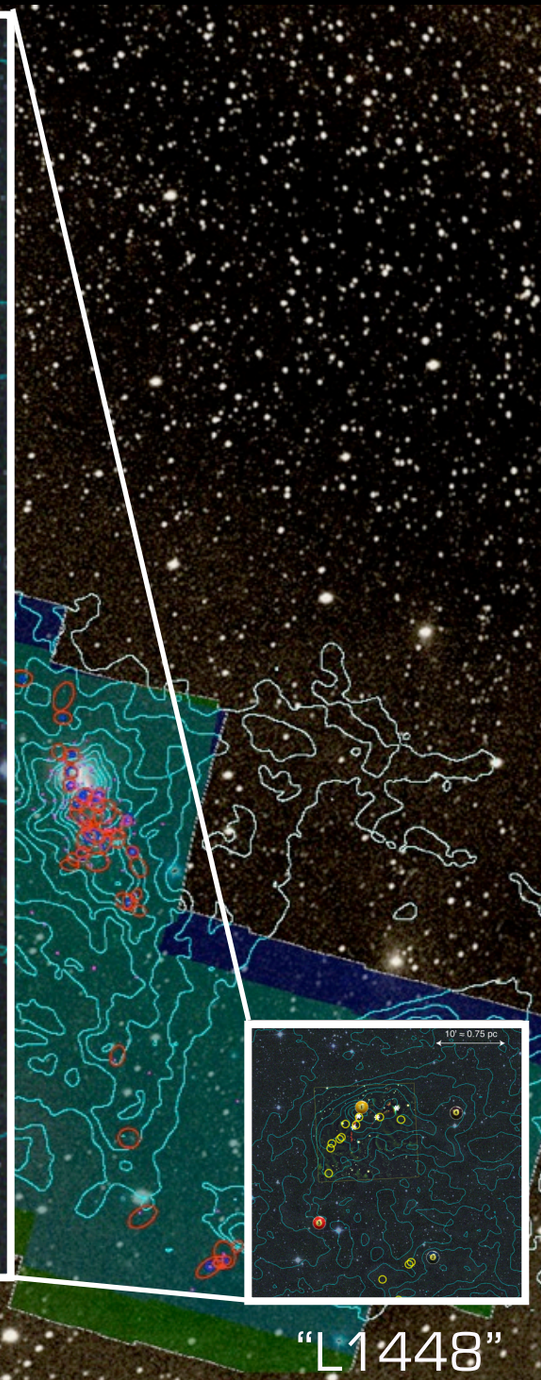


3D Viz made with VolView

AstronomicalMedicine@IIG

COMPLETE





"L1448"

Perseus

COMPLETE

"3D PDF" (Nature, 2009)

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 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

2009

3D PDF

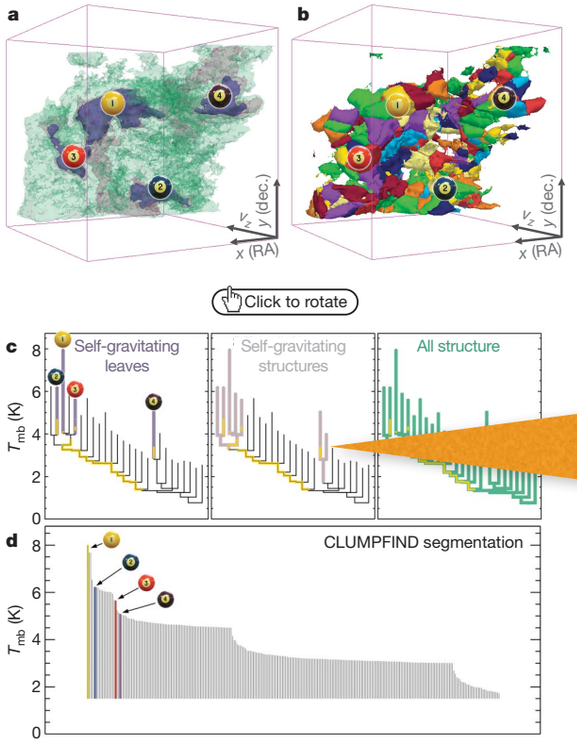


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The x–y locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (p–p–v) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\,\mathrm{km\,s^{-1}}$) to back ($8\,\mathrm{km\,s^{-1}}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'[9] ...roposed as a way to characterize clouds' hierarchical structure

using 2D maps of column density. With th...rly 2D work as inspiration, we have developed a structure-id...ithm 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 ar... 'merger trees' are being used with in...

Figure 3 and its legend explain th... schematically. The dendrogram qua...

...mass, stu... $\mathrm{K^{-1}\,km^{-1}\,s}$ ...ary Fig. 2). ... 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.

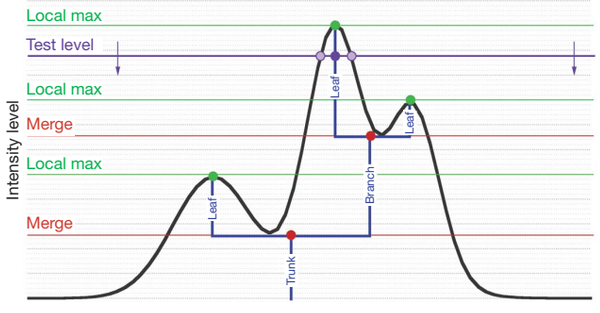
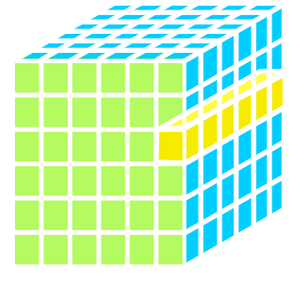
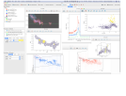


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in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

©2009 Macmillan Publishers Limited. All rights reserved

These are "dead" panels! That's not good enough.



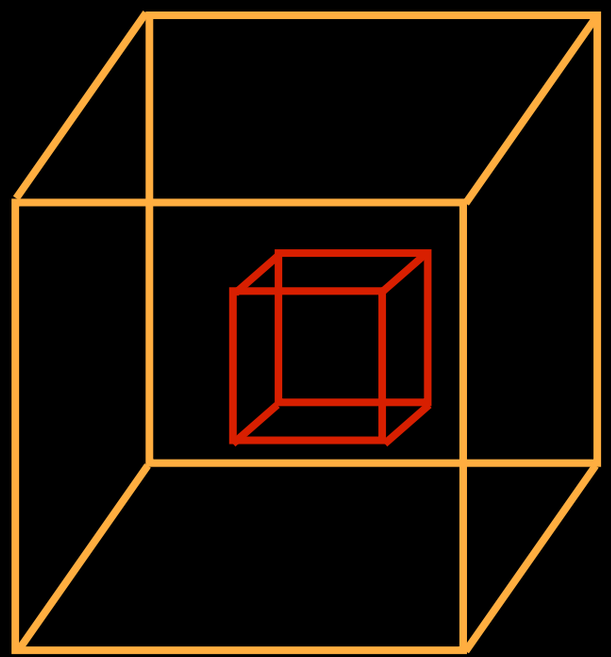


VAST

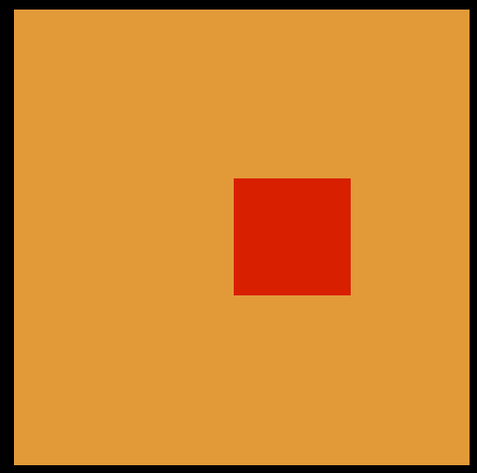
Goodman et al. 2009, Nature,
cf: Fluke et al. 2009

Linked Views of High-dimensional Data

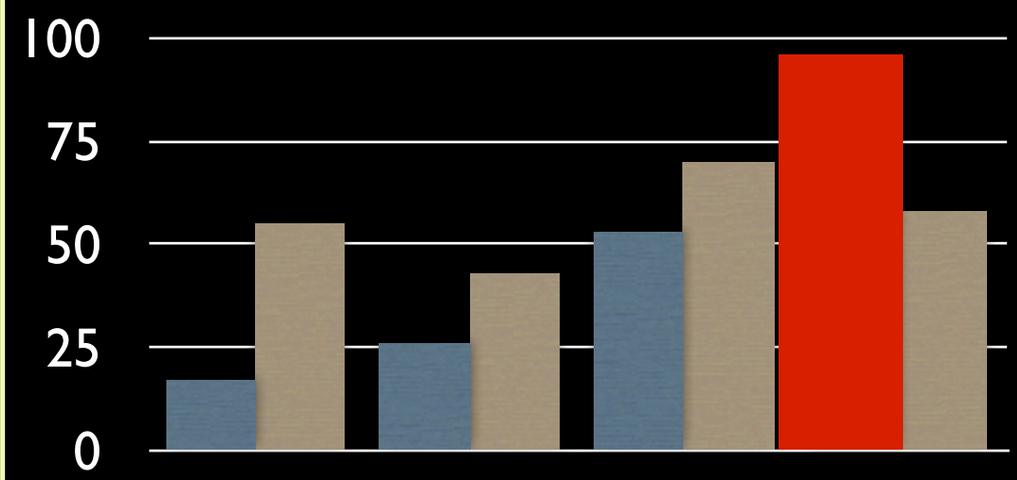
John Tukey



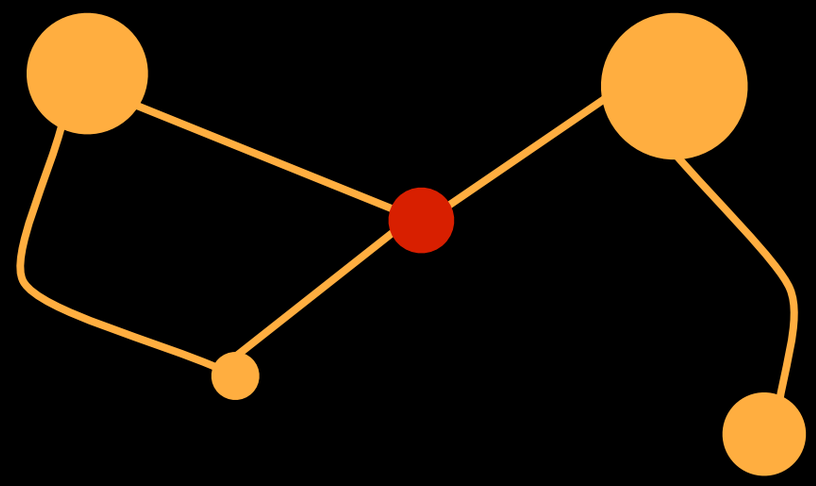
3D



2D



Data Abstraction



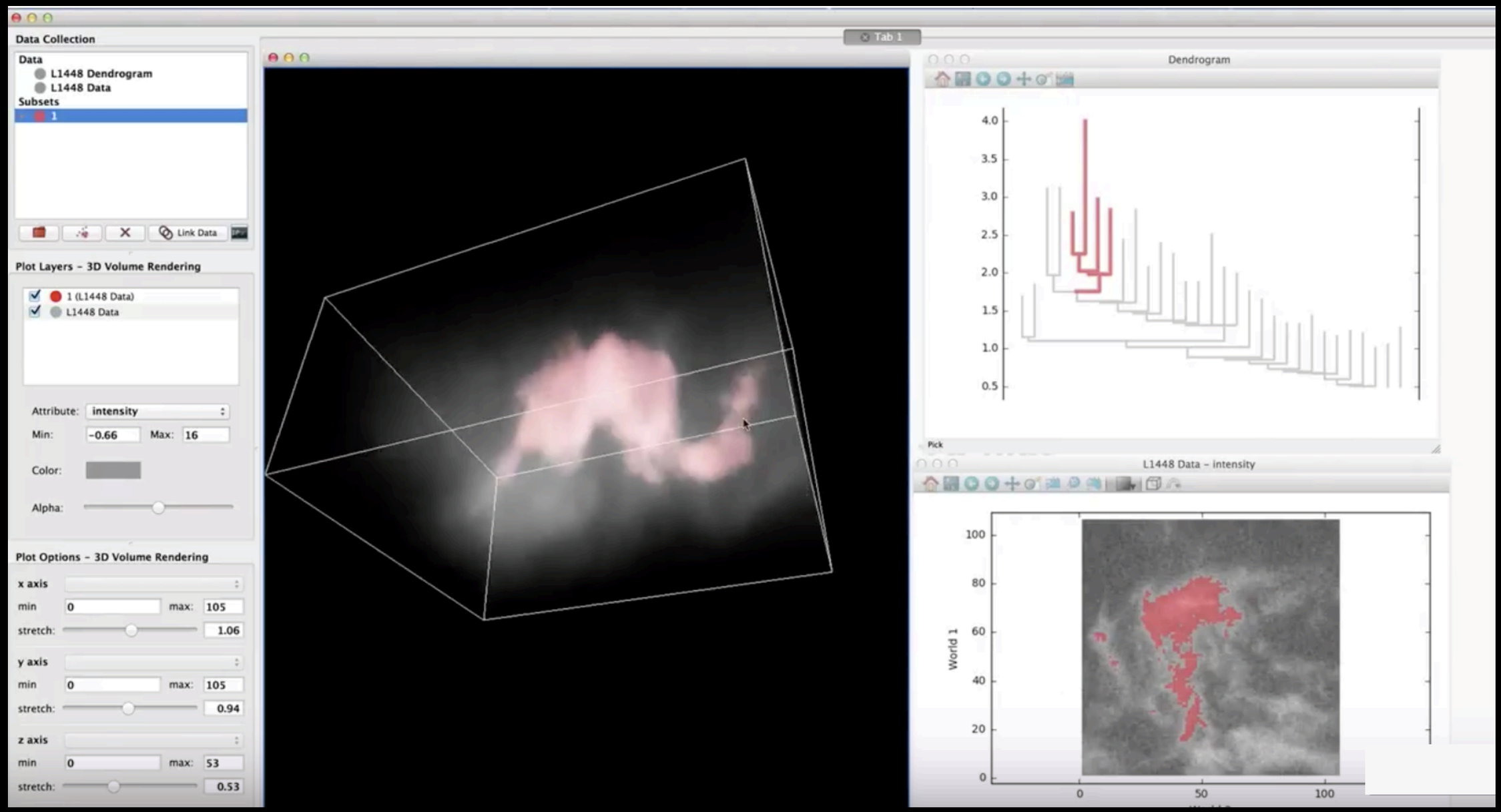
figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

VAST • INFOVIS • SCIVIS

VIS 2018

21–26 October 2018

BERLIN, GERMANY



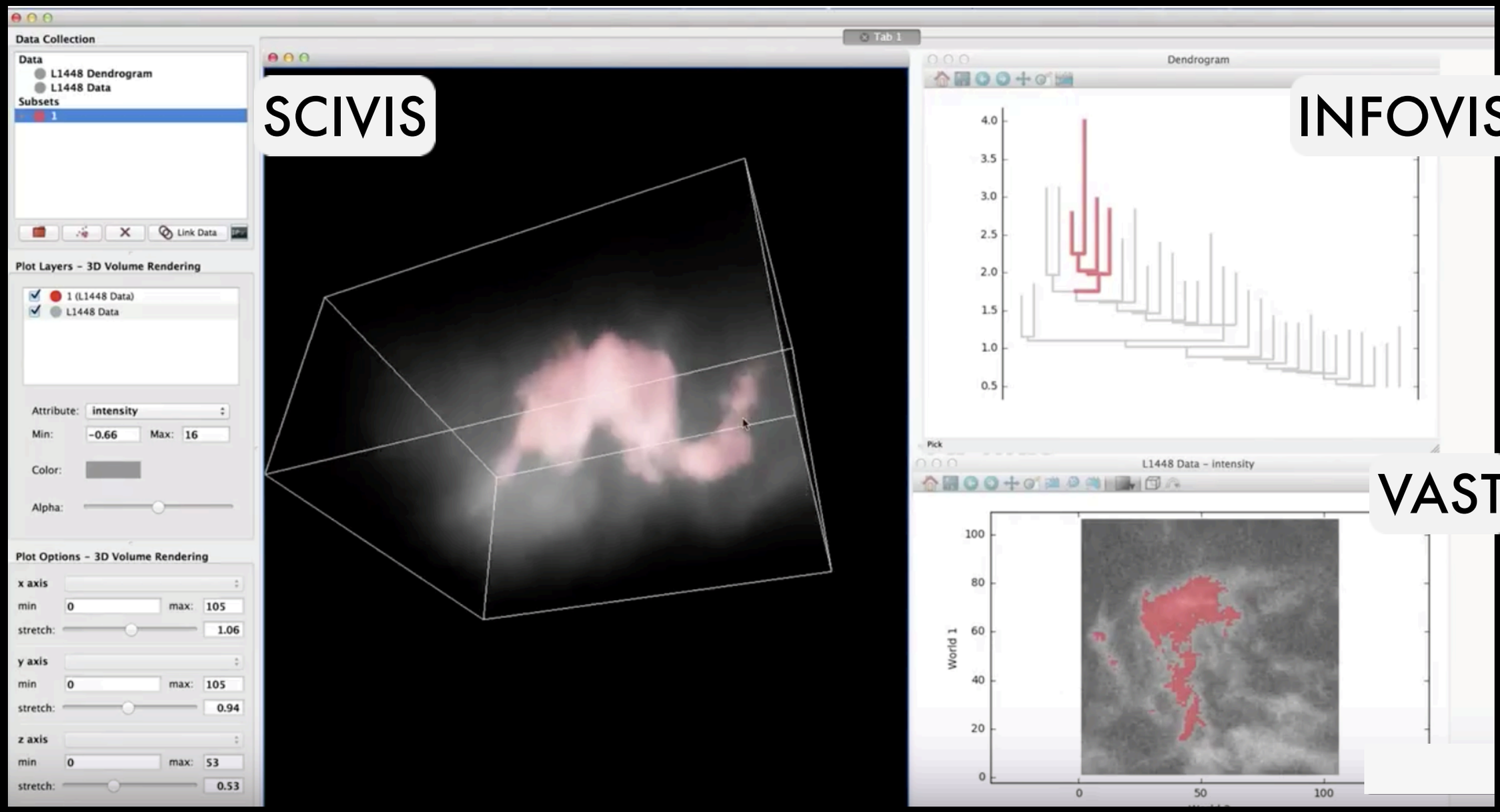
VAST • INFOVIS • SCIVIS

VIS 2018

21–26 October 2018

BERLIN, GERMANY

IEEE vgtc · IEEE · IEEE computer society

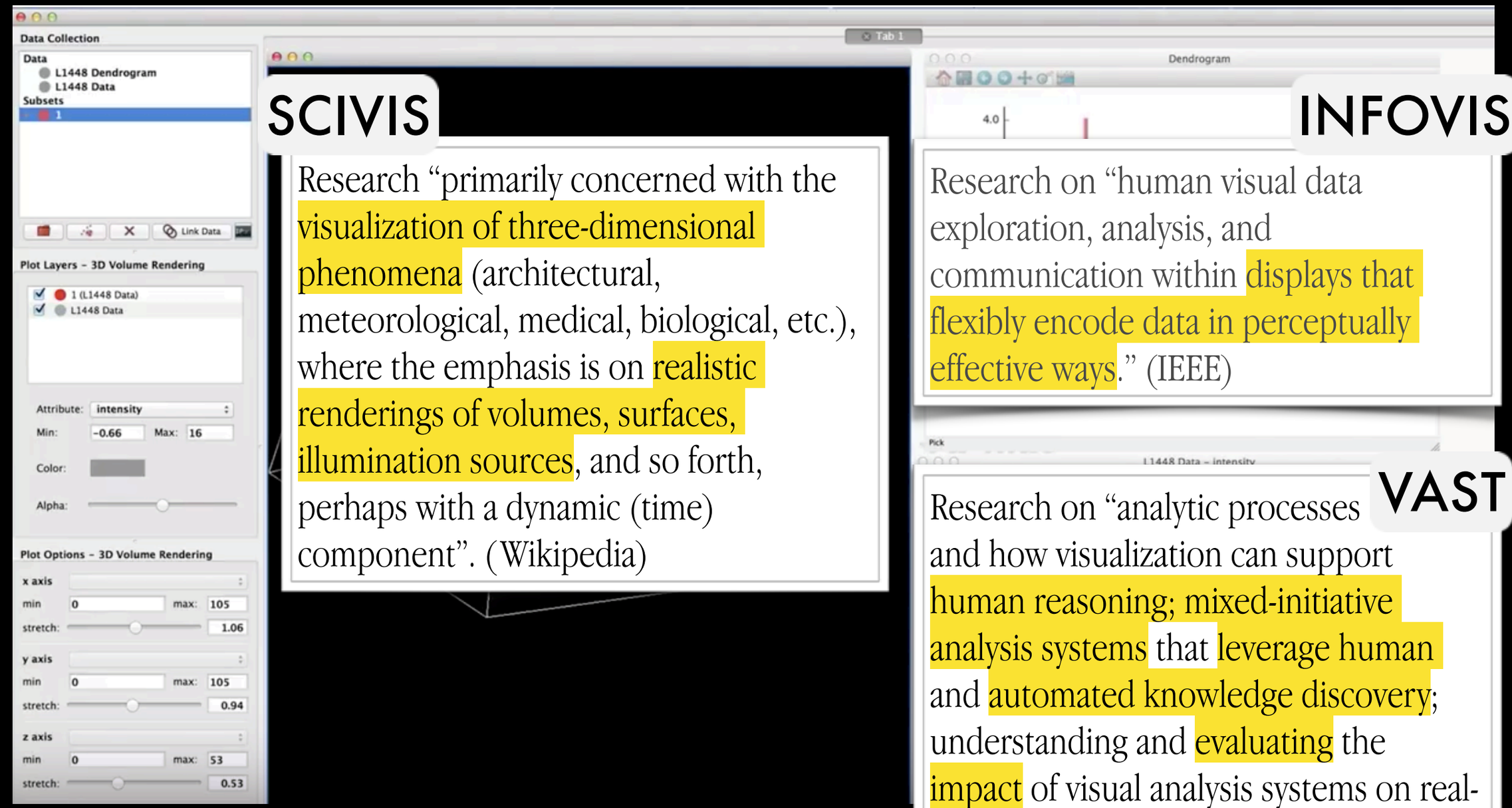


VAST • INFOVIS • SCIVIS

VIS 2018

21–26 October 2018

BERLIN, GERMANY



SCIVIS

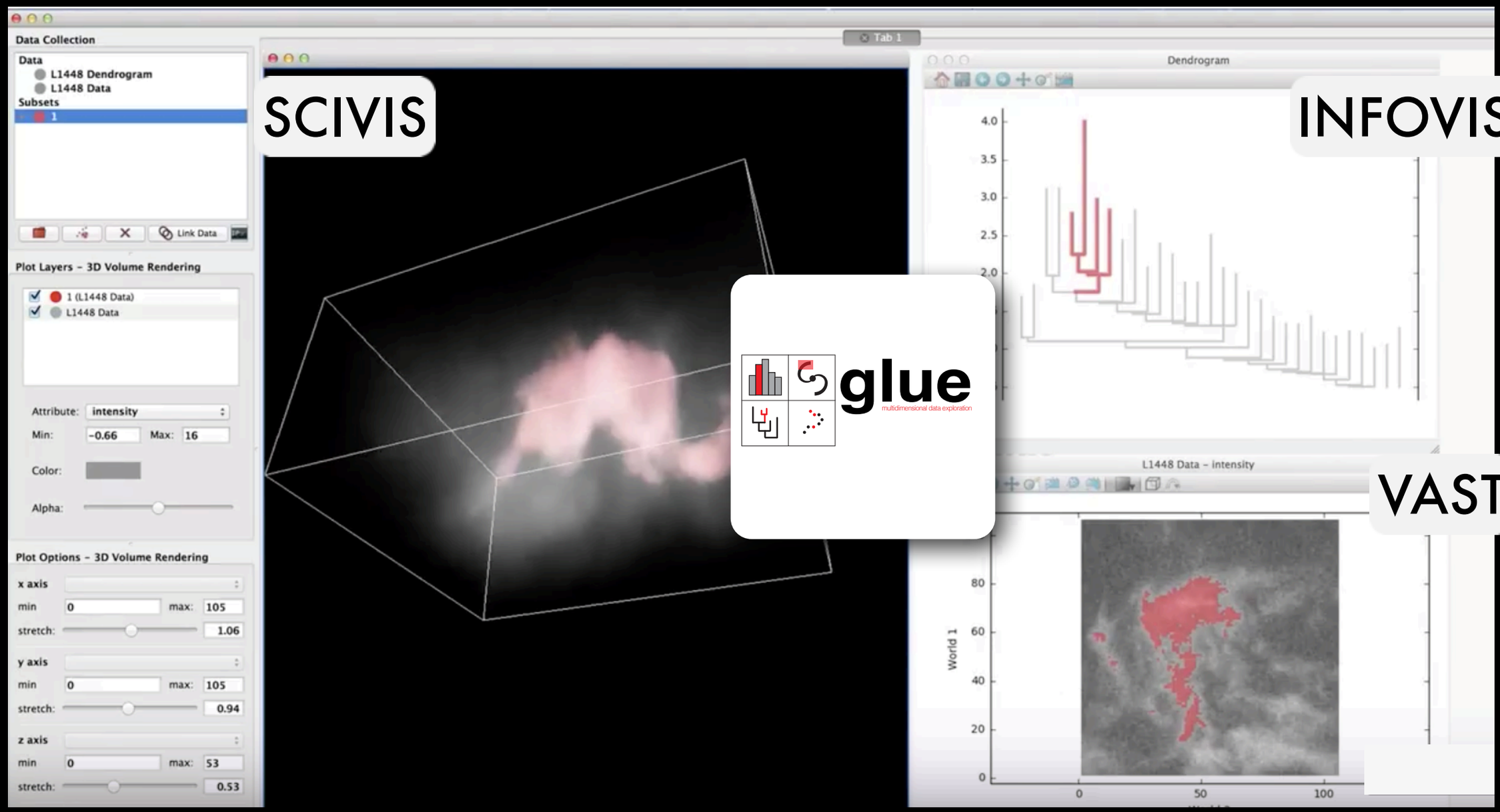
Research "primarily concerned with the visualization of three-dimensional phenomena (architectural, meteorological, medical, biological, etc.), where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component". (Wikipedia)

INFOVIS

Research on "human visual data exploration, analysis, and communication within displays that flexibly encode data in perceptually effective ways." (IEEE)

VAST

Research on "analytic processes and how visualization can support human reasoning; mixed-initiative analysis systems that leverage human and automated knowledge discovery; understanding and evaluating the impact of visual analysis systems on real-world applications." (IEEE)



core glue contributors: C. Beaumont, M. Borkin, M. Breddels, A. Goodman, T. Robitaille

SCIVIS

INFOVIS

2009

3D PDF

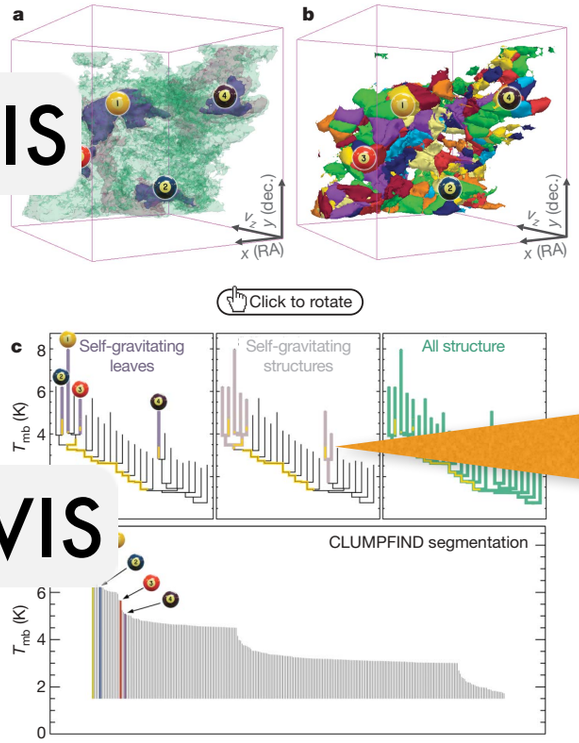


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The x–y locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (p–p–v) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ($-0.5\,\mathrm{km\,s^{-1}}$) to back ($8\,\mathrm{km\,s^{-1}}$).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'[9] ...roposed as a way to characterize clouds' hierarchical structure

using 2D maps of column density. With th...rly 2D work as inspiration, we have developed a structure-id...ithm 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 ar... 'merger trees' are being used with in...

Figure 3 and its legend explain th... schematically. The dendrogram qua...

...mass, stu... $\mathrm{K^{-1}\,km^{-1}\,s}$ ...ary Fig. 2). ...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.

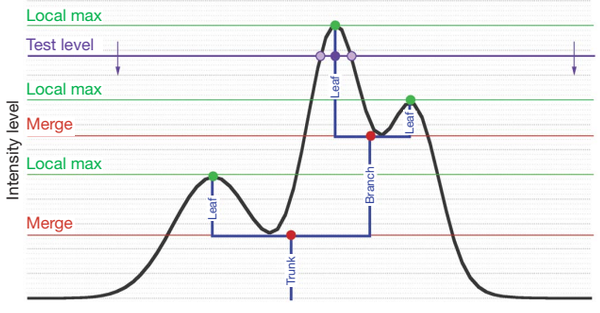
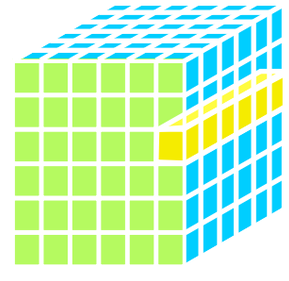
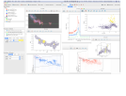


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in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

©2009 Macmillan Publishers Limited. All rights reserved

These are "dead" panels! That's not good enough.



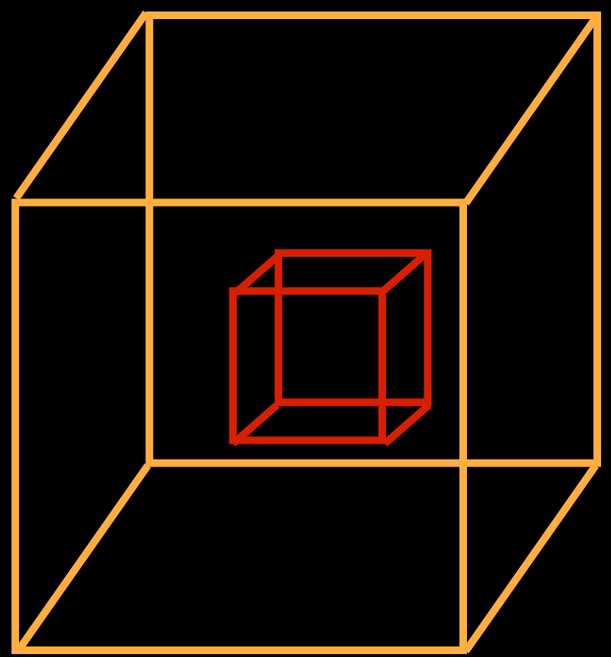


VAST

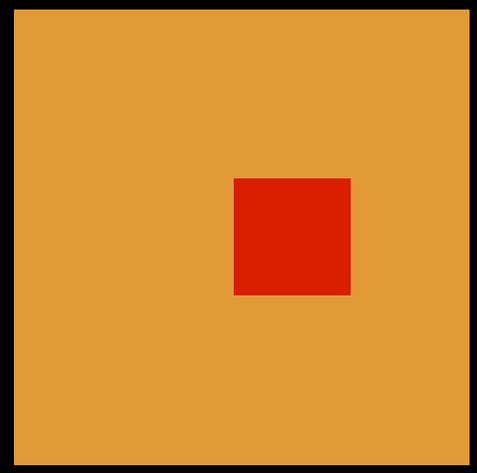
Goodman et al. 2009, Nature,
cf: Fluke et al. 2009

Linked Views of High-dimensional Data

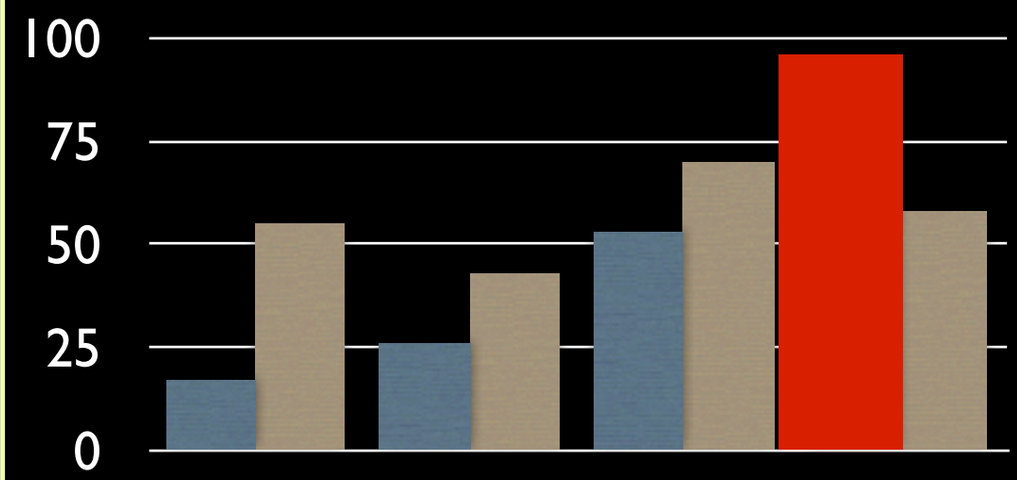
John Tukey



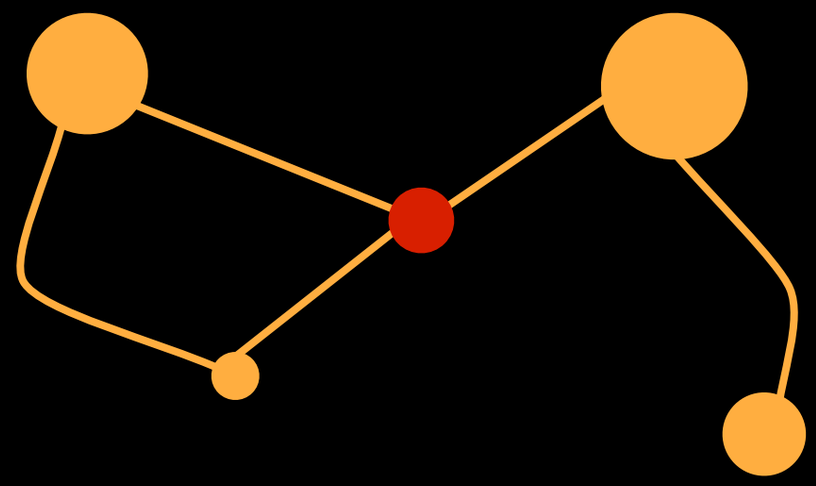
3D



2D

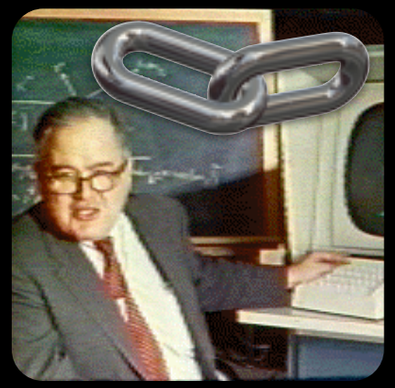


Data Abstraction



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®

XGobi

GGobi

glue

RGGobi

Spotfire®

Polaris

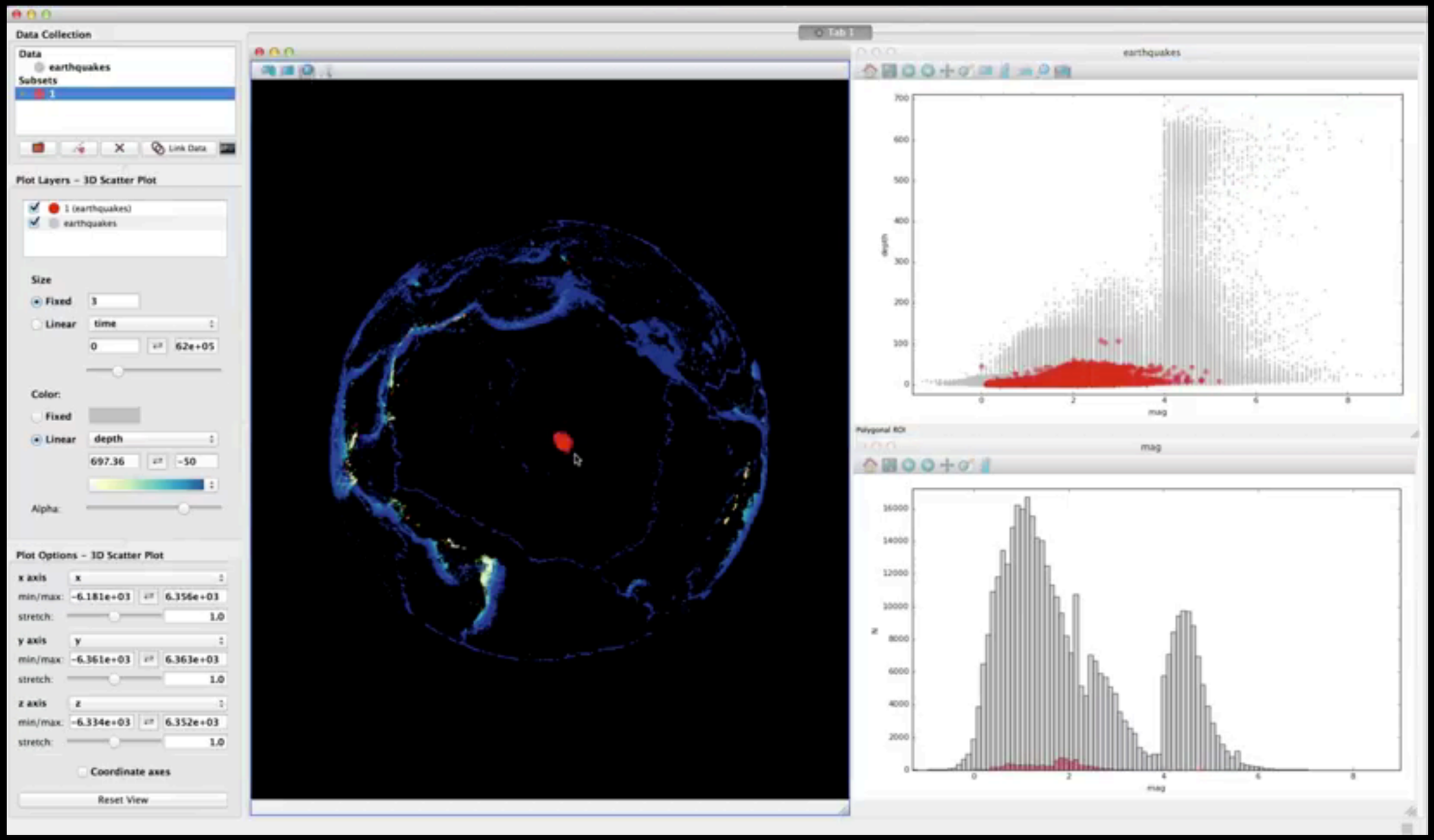
+ableau SOFTWARE

IEEE VIS established

1970 1980 1990 2000 2010

Linked Views of High-dimensional Data (in Python)

glue

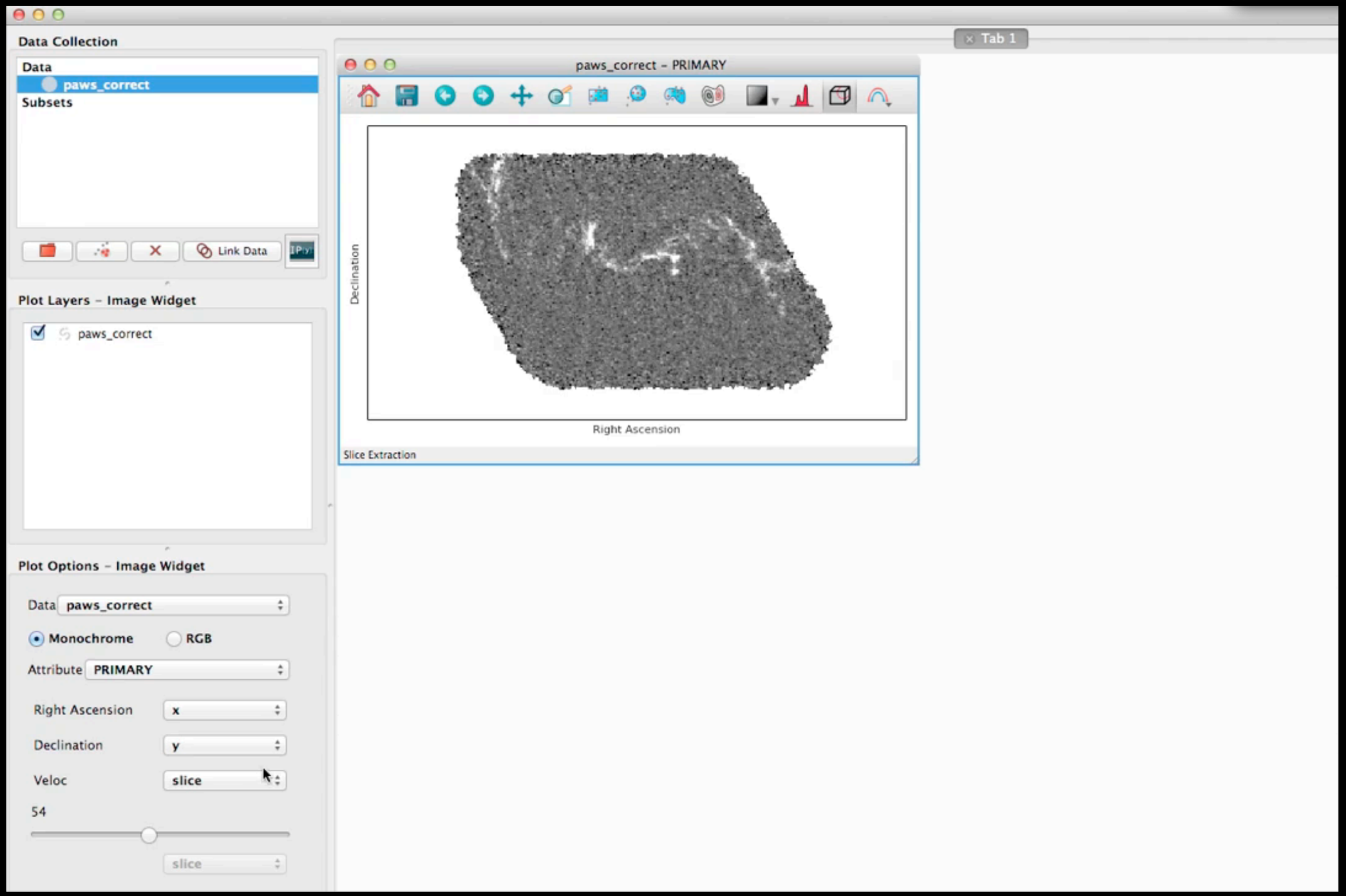


video by Tom Robitaille, lead glue developer

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

Linked Views of High-dimensional Data (in Python)

glue



video by Chris Beaumont, glue developer

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

glue

multidimensional data exploration

built-in

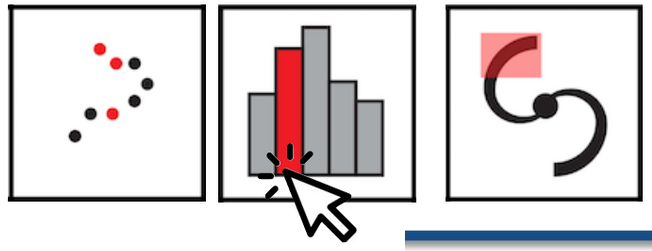
standard data loaders

Data File 1

Data File 2

Data File N

define new variables, import/export insights, interactive plots for the web, save state, all from GUI



standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic

plug-in

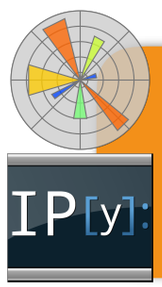
custom data loaders

custom buttons, features

custom plots

+options

user config.py file (loaders, colors, plot types, +)



access to all matplotlib functions through built-in IPython terminal

jupyter

run & interact with glue from Jupyter notebook & other tools

glueviz.org

[demo]
[ALMA]
[publishing]

python File Edit View Canvas Data Manager Plugins Help 21% Mon Oct 22 7:15 AM Alyssa A Goodman

Glue

Open Session | Export Session | Import Data | Export Data/Subsets | Link Data | Arithmetic attributes | Active Subset: Subset 3 | Mode: | Terminal | Preferences | Error Console

Data Collection

Data
- PerA_12coFCRAO_F_xyv
- PerA_13coFCRAO_F_xyv
- Perseus_Av_NICER
- PerA_AvTemMIPS_F_Av
- PerA_AvTemMIPS_F_T
- Perseus_fcrao_iras_2mass

Subsets
- spec_probe
- hot_highv
- Subset 3

Plot Layers - 2D Scatter
- Subset 3 (Perseus_fcrao_iras_2mass)
- hot_highv (Perseus_fcrao_iras_2mass)
- spec_probe (Perseus_fcrao_iras_2mass)
- Perseus_fcrao_iras_2mass

Color | Points | Line | Errors | Vectors

color: Fixed

opacity

Plot Options - 2D Scatter

General | Limits | Axes

x axis: V13CO log

y axis: TIRAS log

Tab 1 | Tab 2

2D Image

1D Profile

2D Scatter

Link Editor

Click on two datasets to set up links or click on an existing connection to edit links. Selected datasets are shown in green. When one dataset is selected, the colors show directly and indirectly linked (blue) and inaccessible (red) datasets.

PerA_AvTemMIPS_F_T — PerA_13coFCRAO_F_xyv

PerA_AvTemMIPS_F_Av — PerA_12coFCRAO_F_xyv

Perseus_Av_NICER — Perseus_fcrao_iras_2mass

Data File 1

Data File 2

Data File N

Datasets: PerA_AvTemMIPS_F_Av

Components

Coordinate components
el Axis 0 [y]
el Axis 1 [x]
clination
ht Ascension
n components
MARY

Glue

Datasets: PerA_12coFCRAO_F_xyv

Components

Coordinate components
Pixel Axis 0 [z]
Pixel Axis 1 [y]
Pixel Axis 2 [x]
Velocity
Declination
Right Ascension
Main components
PRIMARY

Current Links

Function	Component 1	Component 2
identity	Right Ascension	Right Ascension
identity	Declination	Declination

Unglue

Advanced linking | Cancel | OK

V13CO

3D Scatter

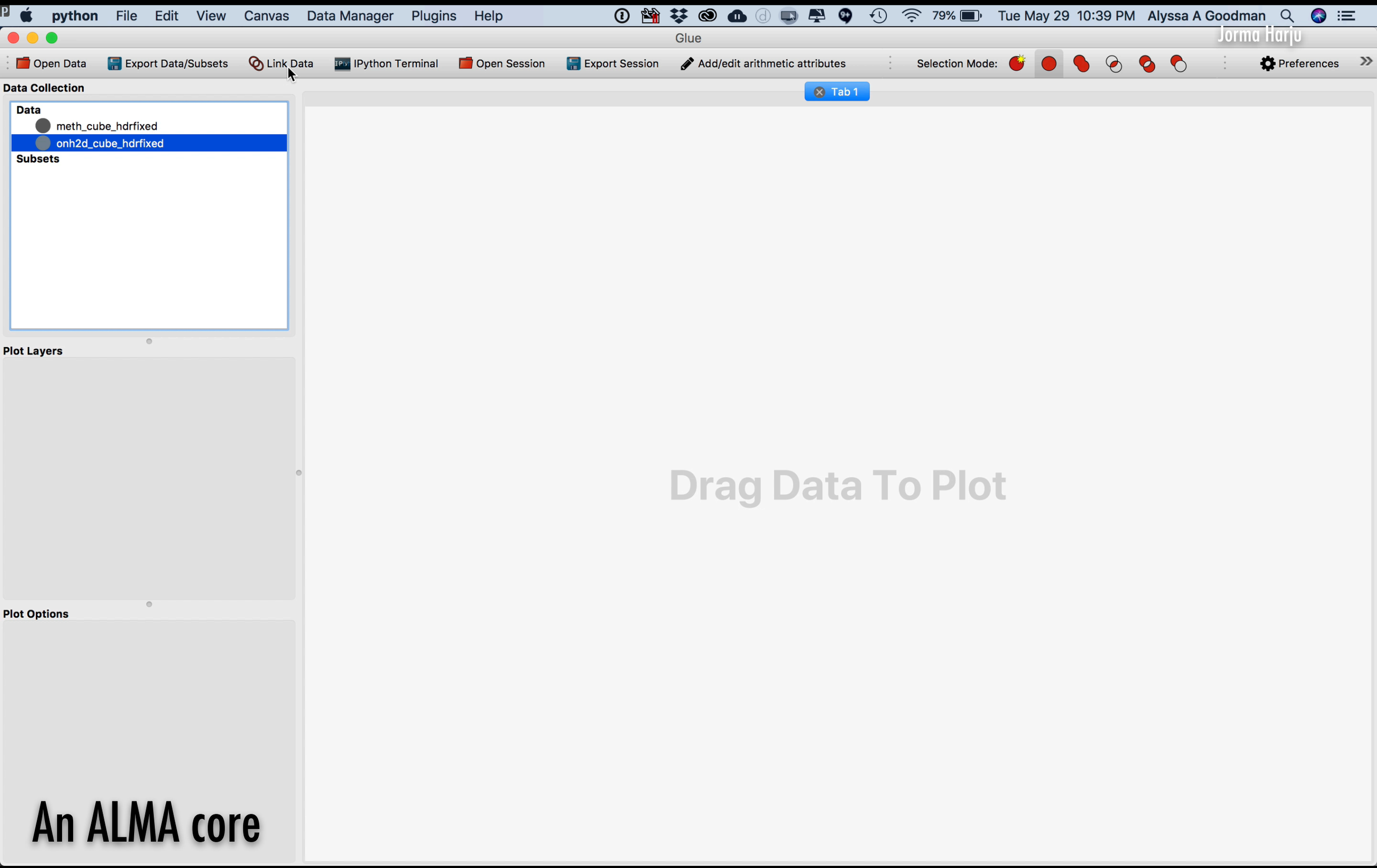
NIRAS

RA

Pixel Axis 0 [z]

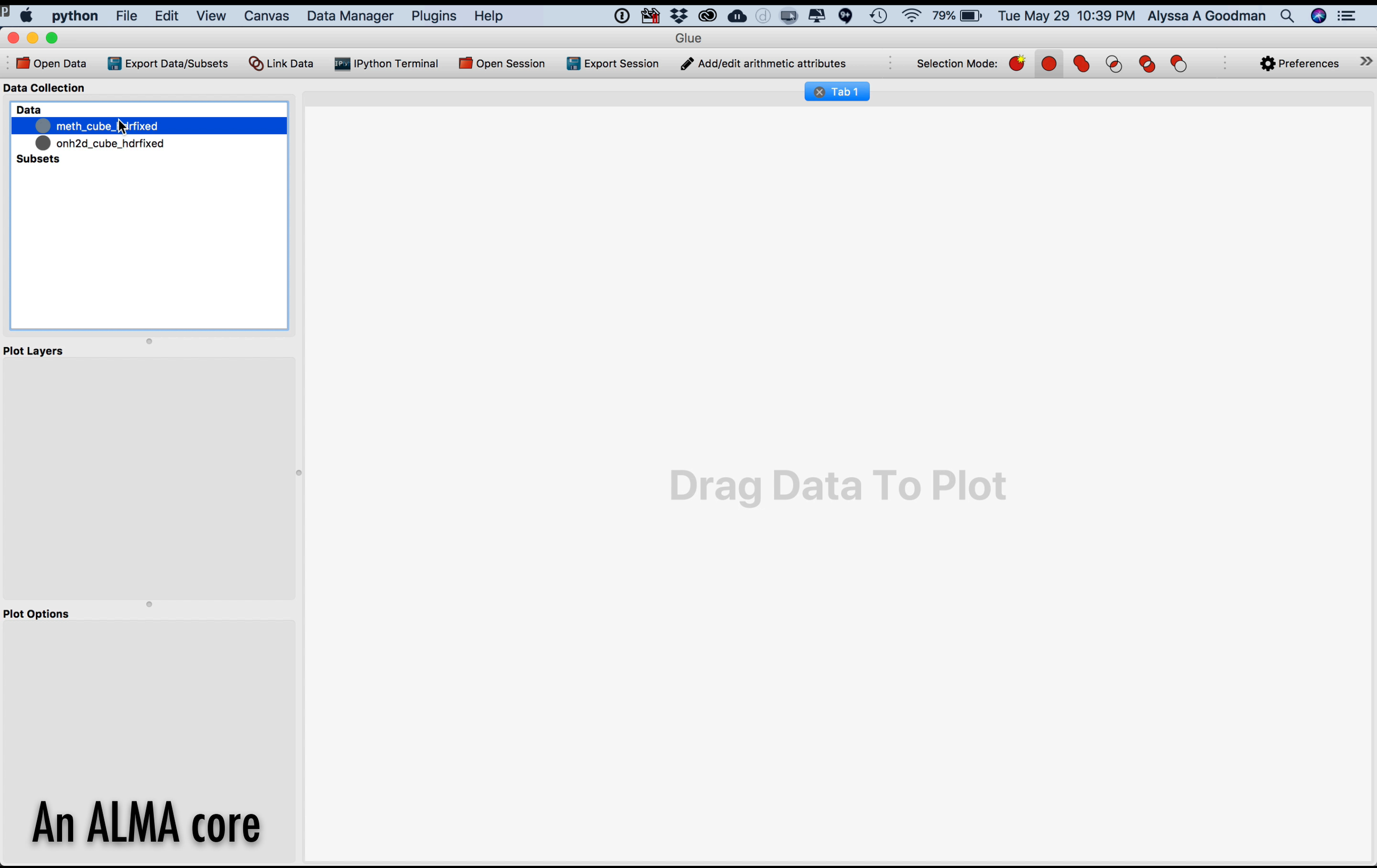
Pixel Axis 1 [y]

No merging of data sets—just glue them.



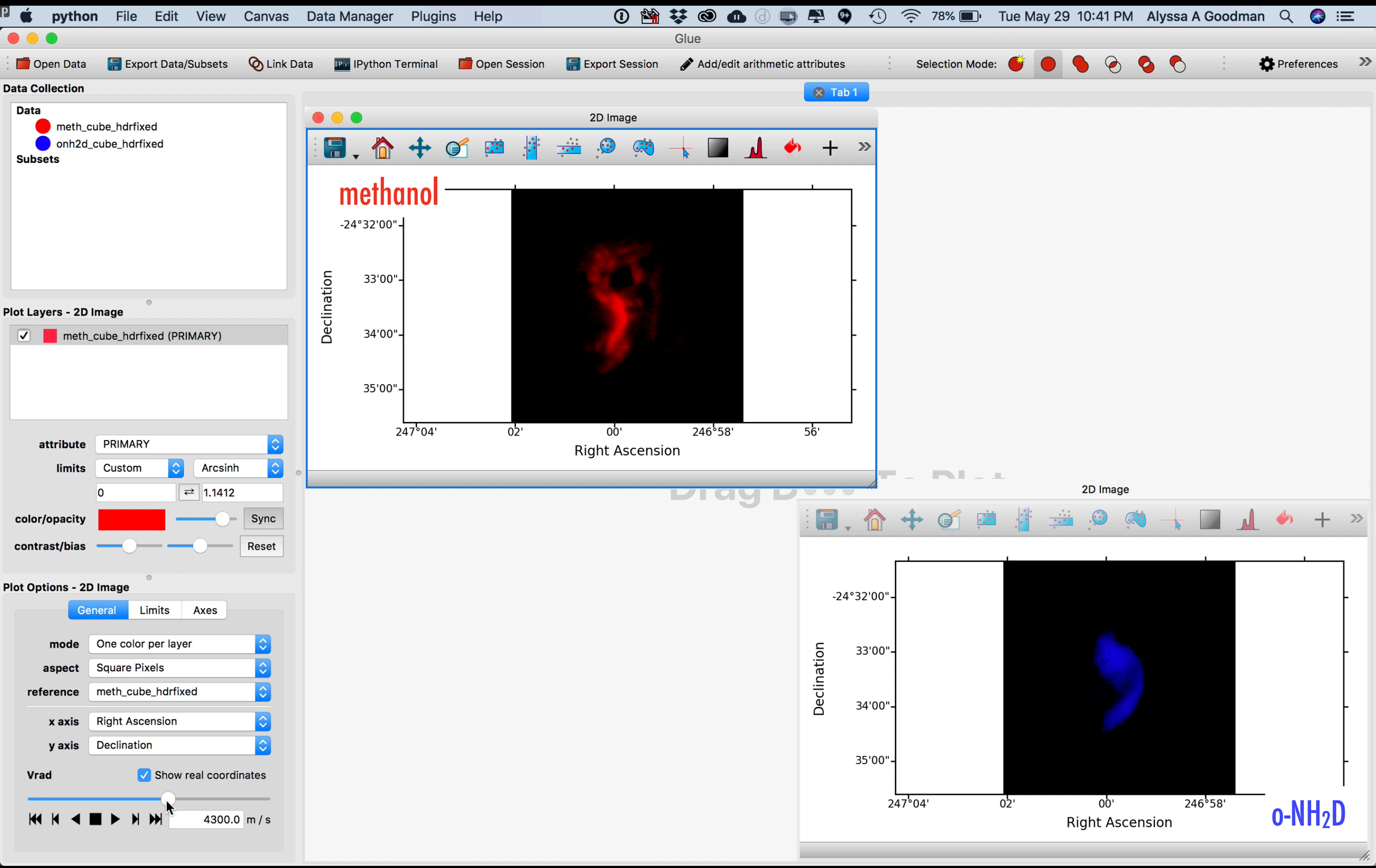
An ALMA core

Just drag to visualize, e.g. series of 2D "channel maps."

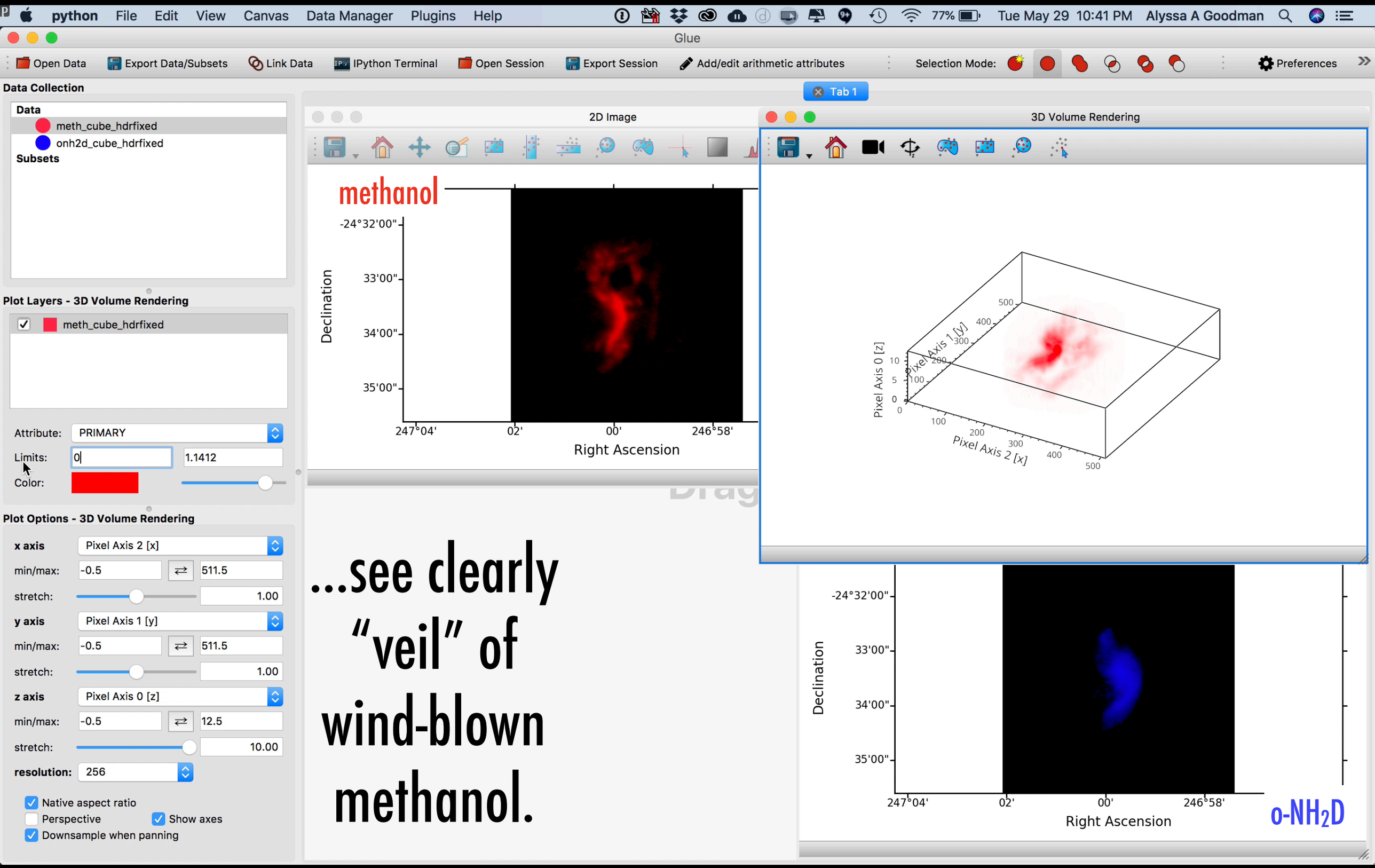


An ALMA core

Adjust so each tracer is a different color.

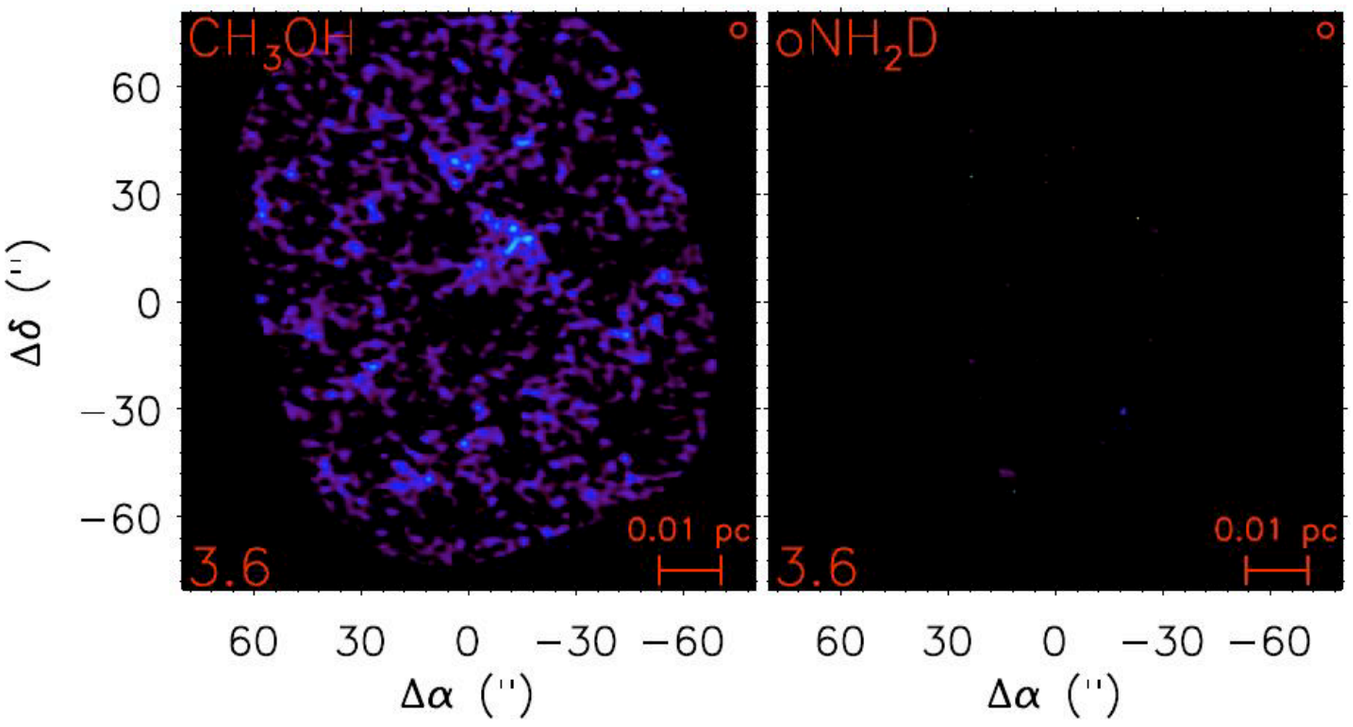


Create 3D views...

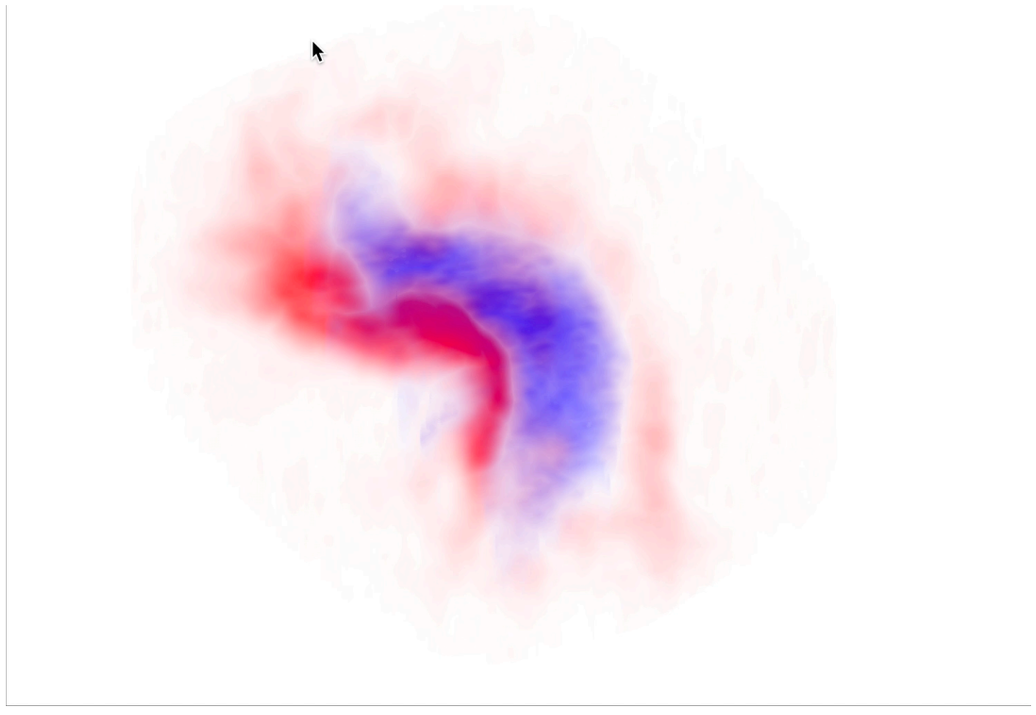


...see clearly "veil" of wind-blown methanol.

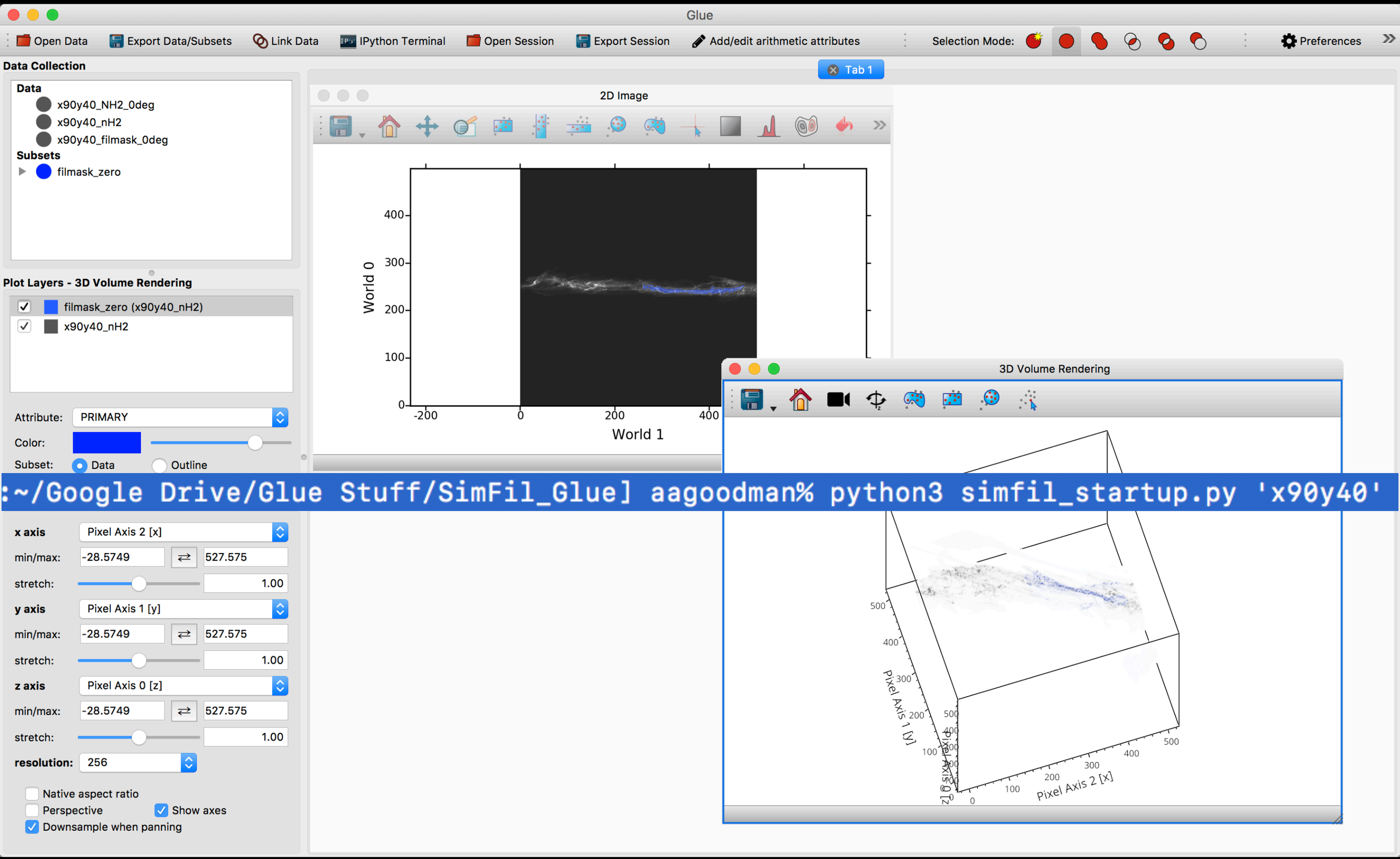
Traditional Rainbow Channel maps



glue



Scripting


Glue
Open Data
Export Data/Subsets
Link Data
IPython Terminal
Open Session
Export Session
Add/edit arithmetic attributes
Selection Mode:
Preferences
Data Collection
Data
x90y40_NH2_0deg
x90y40_nH2
x90y40_filmask_0deg
Subsets
filmask_zero
Tab 1
2D Image
World 0
World 1
Plot Layers - 3D Volume Rendering
filmask_zero (x90y40_nH2)
x90y40_nH2
Attribute:
PRIMARY
Color:
Subset:
Data
Outline
3D Volume Rendering
Pixel Axis 1 [y]
Pixel Axis 2 [x]
Pixel Axis 0 [z]
:~/Google Drive/Glue Stuff/SimFil_Glue] aagoodman% python3 simfil_startup.py 'x90y40'
x axis
Pixel Axis 2 [x]
min/max:
-28.5749
527.575
stretch:
1.00
y axis
Pixel Axis 1 [y]
z axis
Pixel Axis 0 [z]
resolution:
256
Native aspect ratio
Perspective
Show axes
Downsample when panning

2015: The "Paper" of the Future

Authorea

glue

Authorea FEATURED ARTICLES ABOUT PLANS BLOG FEEDBACK HELP ALYSSA GOODMAN

PUBLIC ROUGH DRAFT Index Settings Fork Quickedit Word Count 42 Comments Export Unfollow

The "Paper" of the Future

3

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 + Add author Re-arrange authors

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

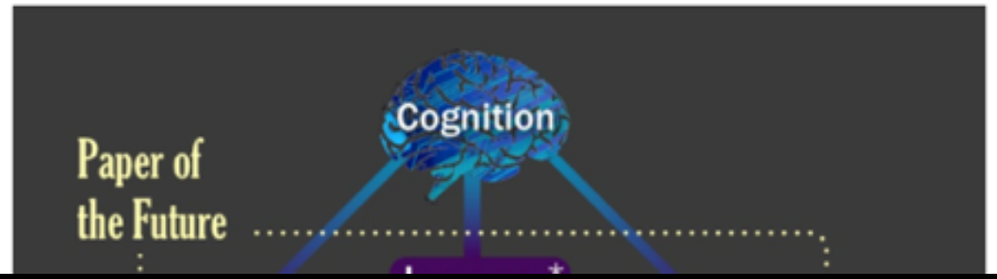
2

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.

0



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.

[demo]

[skip arch]

glue
multidimensional data exploration

enabled by d3.js (javascript) outputs

plotly

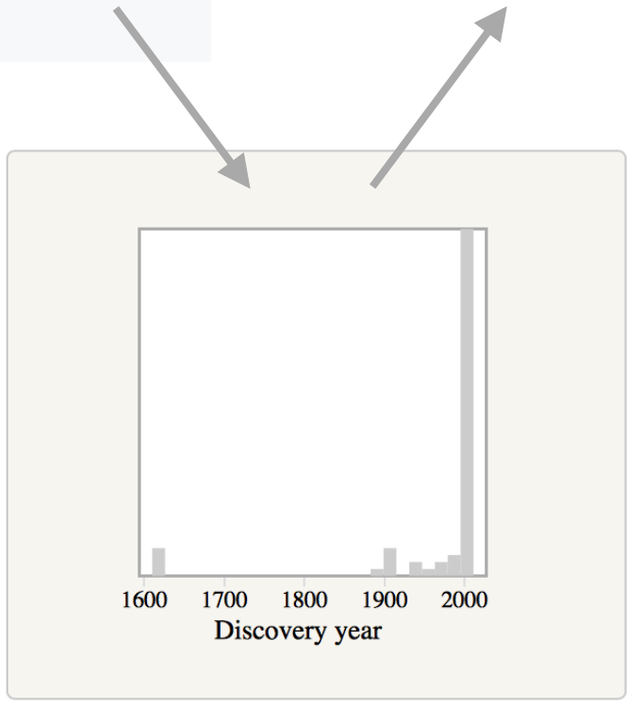
d3po

d3po is a project designed to allow an astronomer (or anyone), with no special data visualization skills, to make an interactive, publication-quality figure that has staged builds and linked brushing through scatter plots. Our current version can be previewed at d3po.org, and represents a figure from upcoming work by graduate student Elisabeth Newton. The figure describes how metalicity affects color in cool stars, and represents a nice use case for d3po. Try clicking and dragging in the scatter plots to understand the power of linked brushing in published figures.

Right now we are in search of alpha testers, who have figures that could be made interactive and who are willing to get their hands a little dirty (No javascript skills needed). In future versions, we plan to link to glue to allow the creation of d3po figures interactively. We are also exploring implementation of d3po within presentations and within authorea. Full 1.0 version expected in January 2014.

Installing your own d3po server

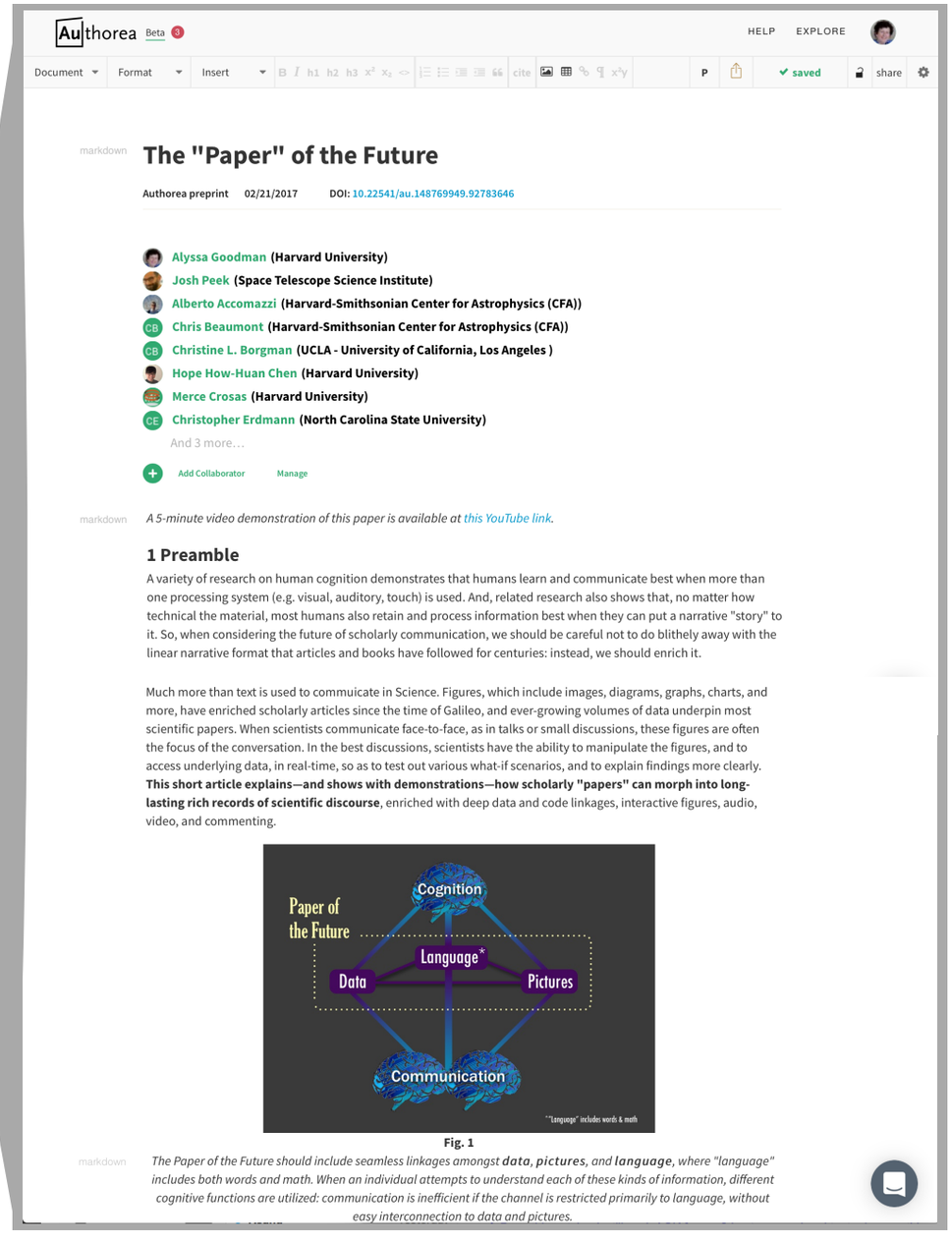
```
git clone git@github.com:adrn/d3po.git
cd d3po
virtualenv --no-site-packages venv
source venv/bin/activate
pip install -r pip-requirements.txt
python run.py
```



Four Centuries of Discovery | A Chasm in Mass | Little Siblings | Close Cousins | The Strangers

After Galileo discovered the first four moons of Jupiter, it took nearly three hundred years to discover the next one.

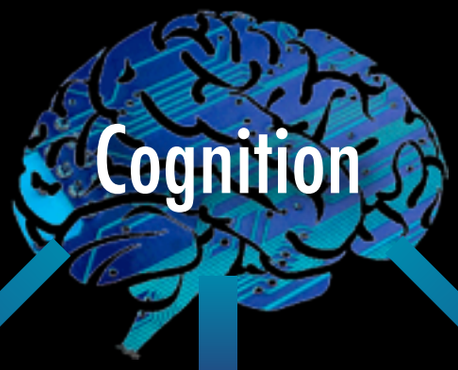
Authorea



[demo]

Many thanks to Alberto Pepe, Josh Peek, Chris Beaumont, Tom Robitaille, Adrian Price-Whelan, Elizabeth Newton, Michelle Borkin & Matteo Cantiello for making this posible.

“Paper of the Future”

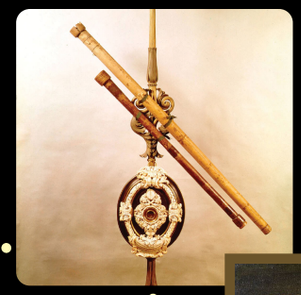


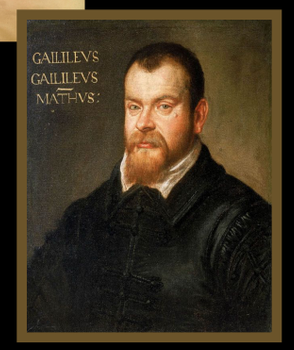
Cognition

Language*

Data

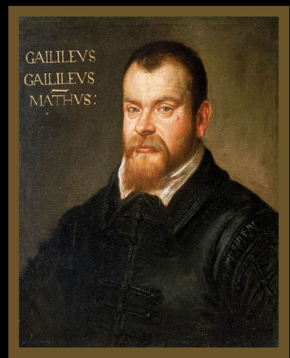
Pictures





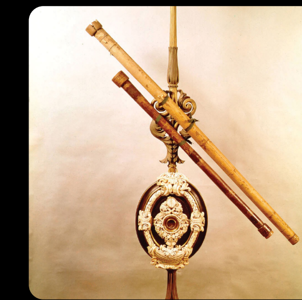
Communication

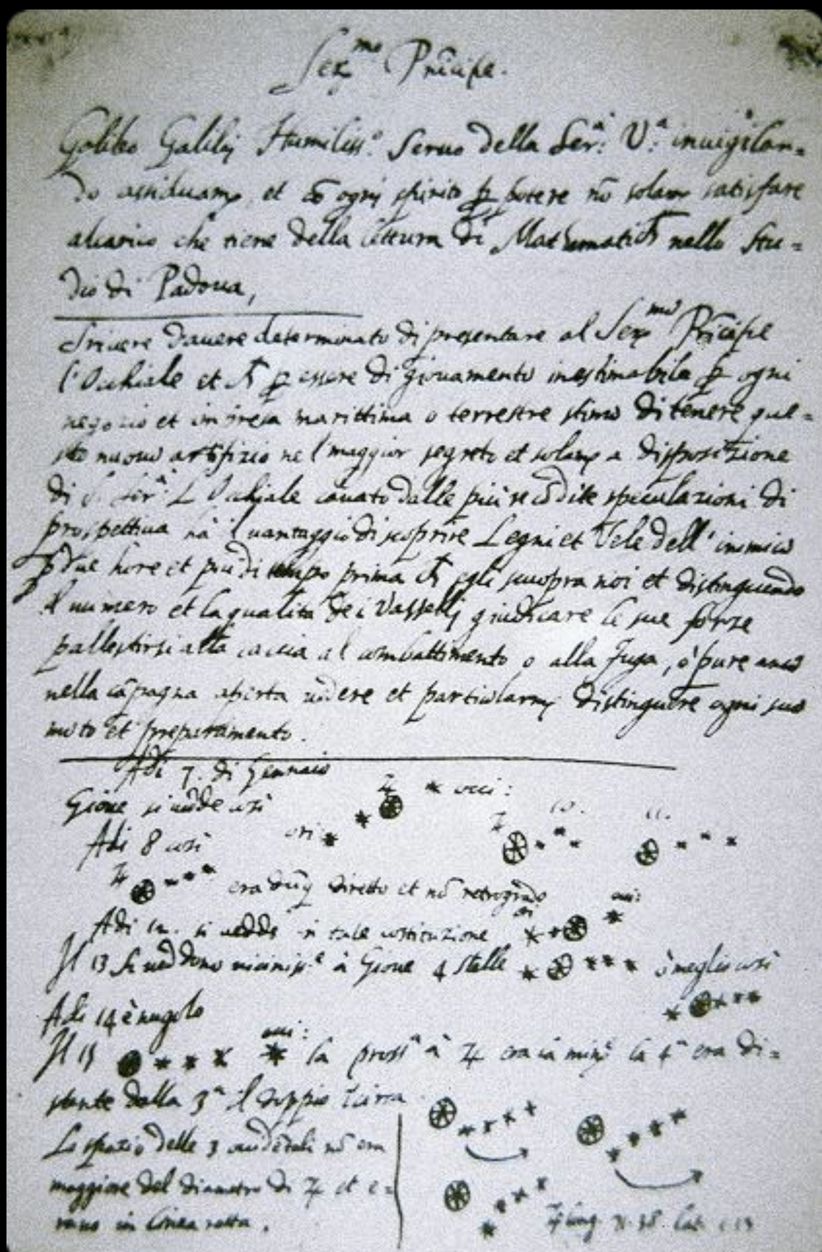
*“Language” includes words & math



1610:Why Galileo is My Hero

(Explore-Explain-Explore)

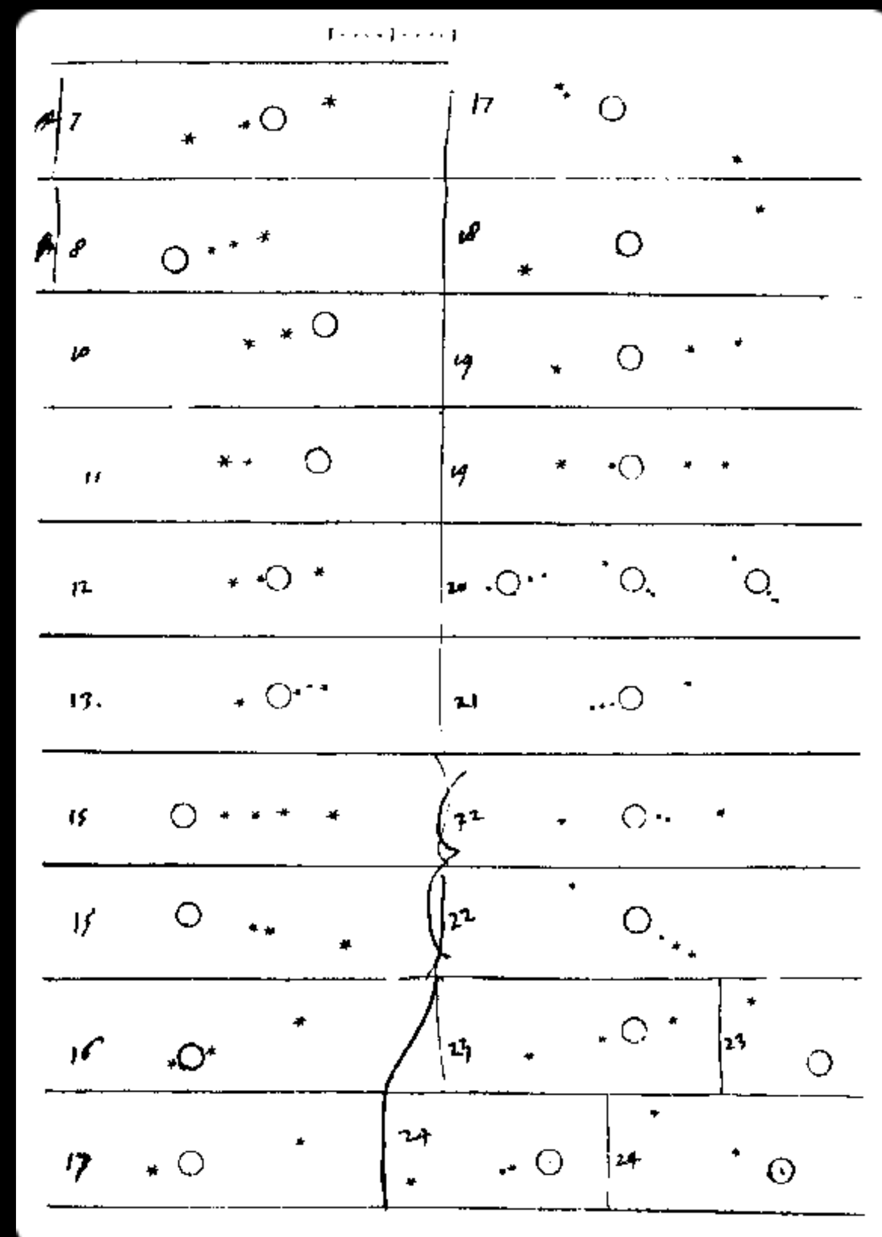


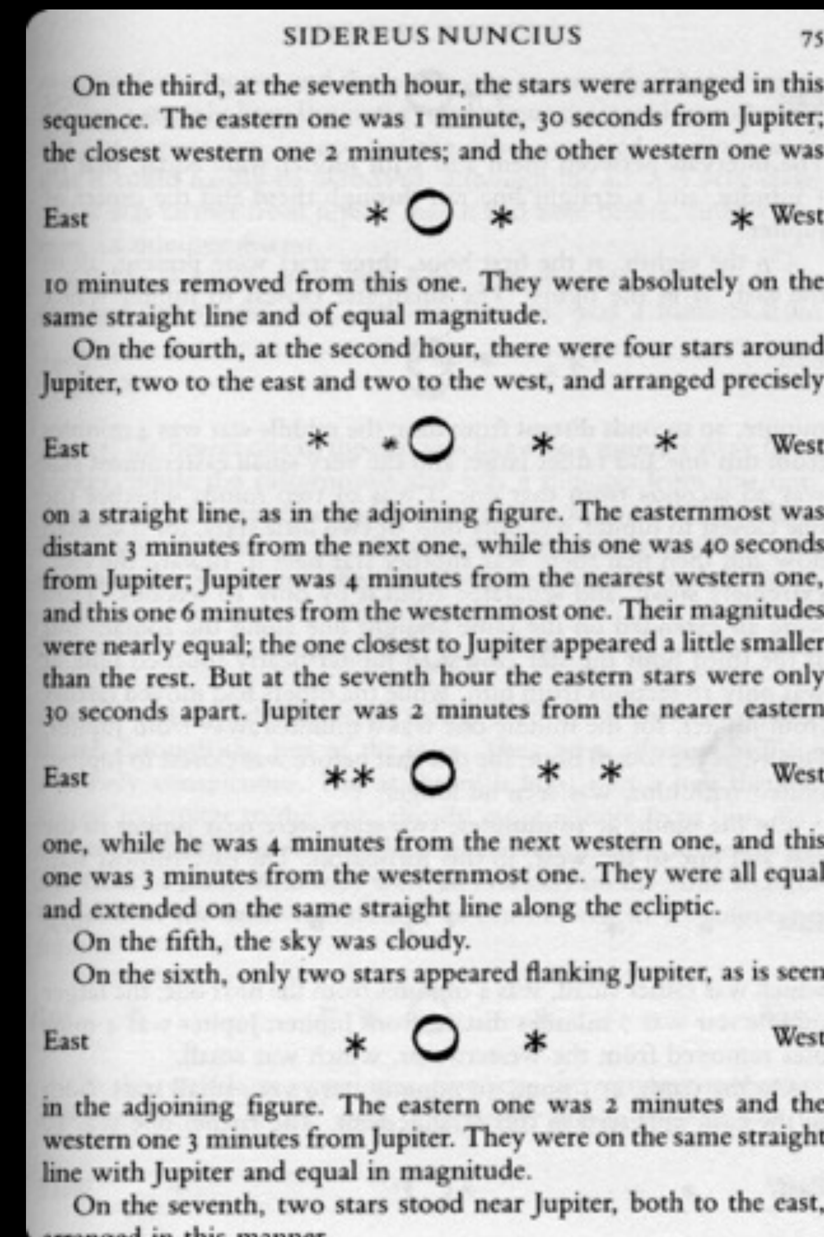


Ser.mo Principe.

Galileo Galilei Humiliss.o Servo della Ser.tà V.a invigilando assiduamente, et con ogni spirito per potere non solamente satisfare al carico che tiene della lettura di Matematiche nello Studio di Padova,

Invece havere determinato di presentare al Ser.mo Principe l'Occhiale et di essere di giovamento inestimabile per ogni negozio et impresa marittima o terrestre stimo di tenere questo nuovo artifizio nel maggior segreto et solamente a disposizione di V. Ser.tà L'Occhiale cavato dalle più recondite speculazioni di prospettiva ha il vantaggio di scoprire Legni et Vele dell'inimico due hore et più di tempo prima che egli scuopra noi et distinguendo il numero et la qualità de i Vasselli giudicare le sue forze allestirsi alla caccia al combattimento o alla fuga, o pure ancora nella campagna aperta vedere et particolarmente distinguere ogni suo moto et preparamento.





SIDEREUS NUNCIUS 75

On the third, at the seventh hour, the stars were arranged in this sequence. The eastern one was 1 minute, 30 seconds from Jupiter; the closest western one 2 minutes; and the other western one was

East * O * * West

10 minutes removed from this one. They were absolutely on the same straight line and of equal magnitude.

On the fourth, at the second hour, there were four stars around Jupiter, two to the east and two to the west, and arranged precisely

East * *O * * West

on a straight line, as in the adjoining figure. The easternmost was distant 3 minutes from the next one, while this one was 40 seconds from Jupiter; Jupiter was 4 minutes from the nearest western one, and this one 6 minutes from the westernmost one. Their magnitudes were nearly equal; the one closest to Jupiter appeared a little smaller than the rest. But at the seventh hour the eastern stars were only 30 seconds apart. Jupiter was 2 minutes from the nearer eastern

East ** O * * West

one, while he was 4 minutes from the next western one, and this one was 3 minutes from the westernmost one. They were all equal and extended on the same straight line along the ecliptic.

On the fifth, the sky was cloudy.

On the sixth, only two stars appeared flanking Jupiter, as is seen

East * O * West

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

On the seventh, two stars stood near Jupiter, both to the east, arranged in this manner.

Notes for & re-productions of Siderius Nuncius

[More later?]

2018: 10QViz.org

≡ MENU

TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION

The 10 Questions

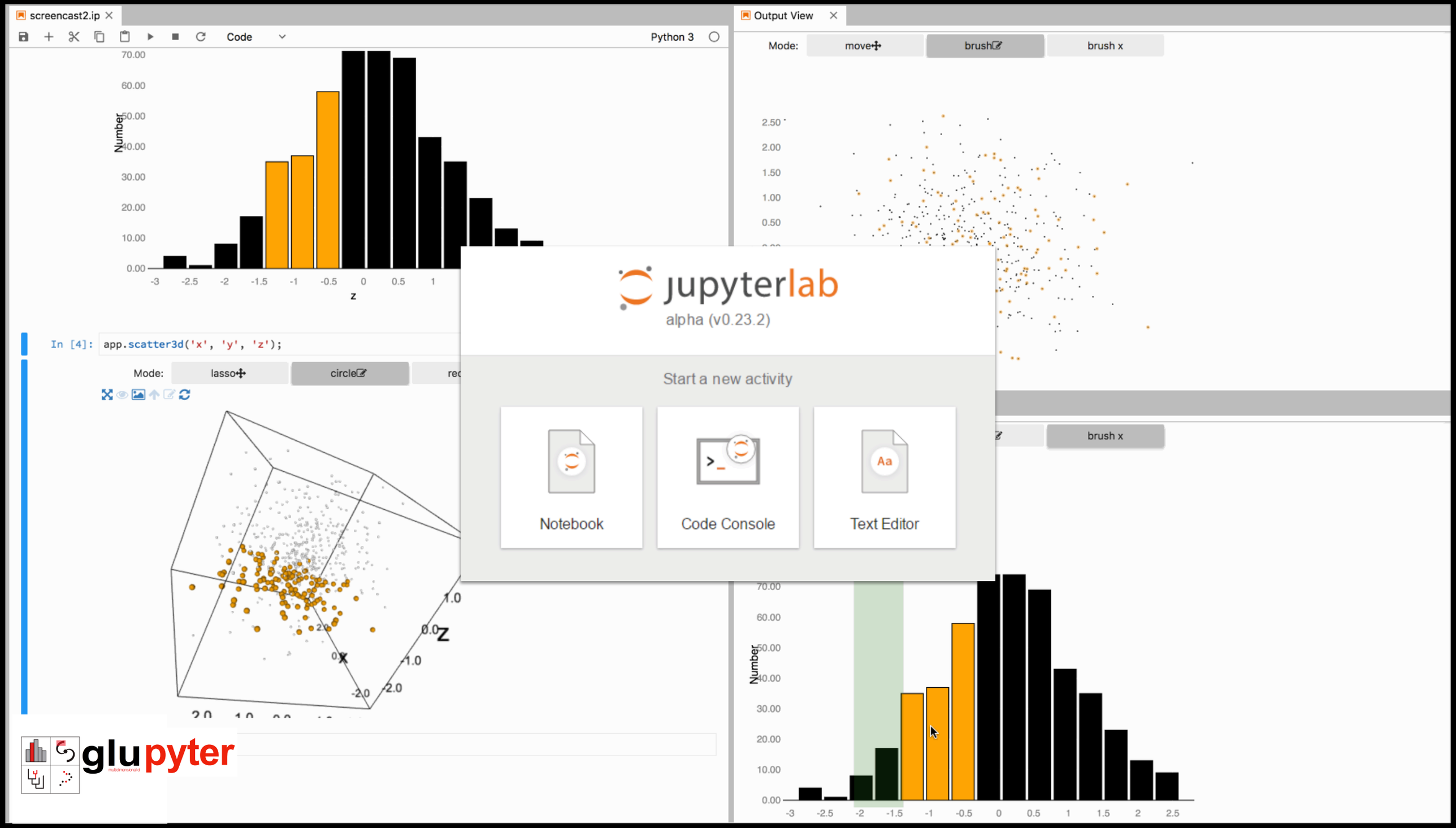
1. **Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
2. **Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
3. **Categories** |Do you want to show or explore pre-existing, known, human-interpretable, categories?
4. **Patterns** |Do you want to identify new, previously unknown or undefined patterns?
5. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
6. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
7. **Abstraction & Accuracy** | Do you need to show all the data, or is summary or abstraction OK?
8. **Context & Scale** | Can you, and do you want to, put the data into a standard frame of reference, coordinate system, or show scale(s)?
9. **Metadata** | Do you need to display or link to non-quantitative metadata? (including captions, labels, etc.)
10. **Display Modes** | What display modes might be used in experiencing your display?

Join the 10QViz Conversation!

*To learn more about this site, please visit the **About** page.*

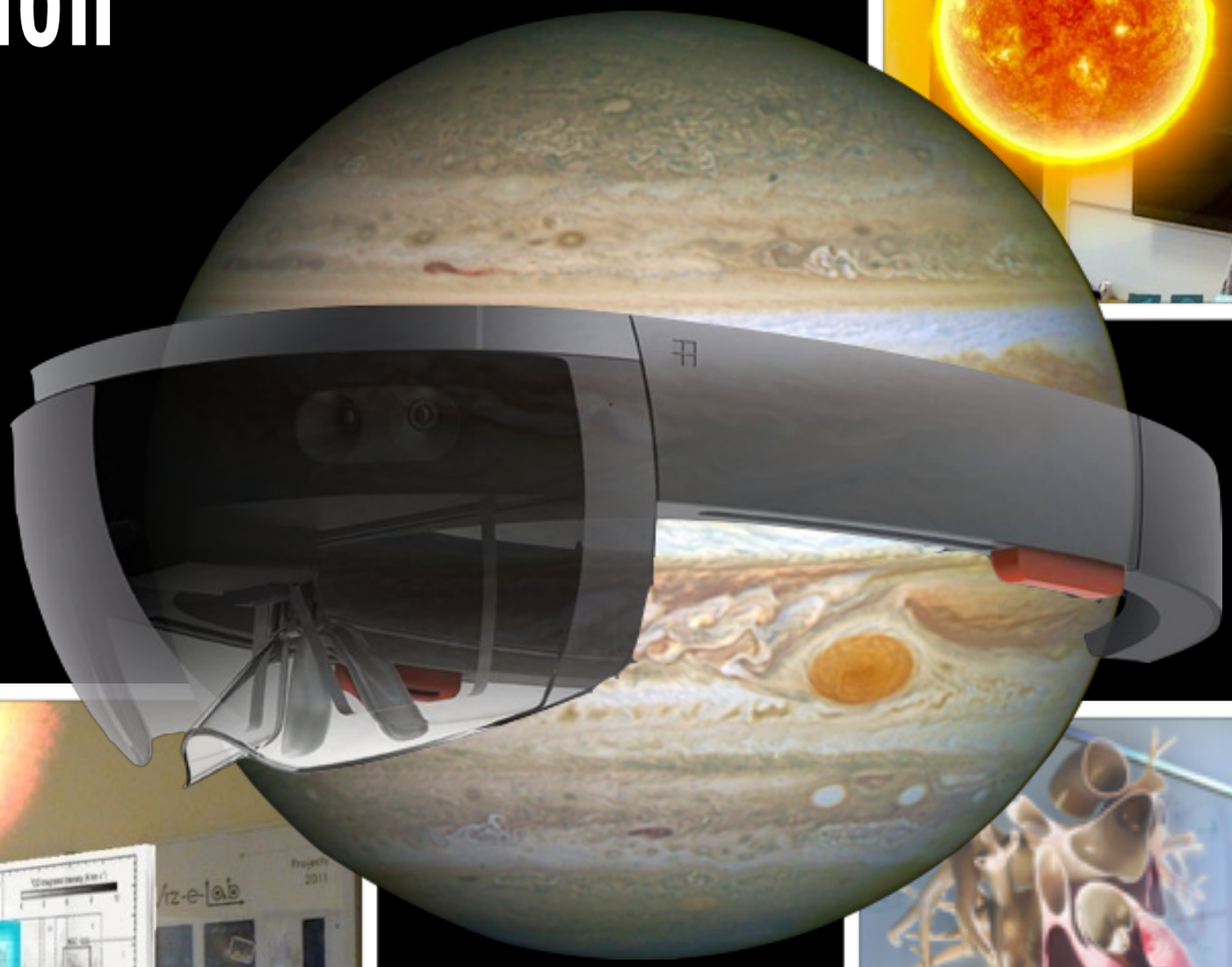
To read an in-process manuscript giving the scholarship behind the recommendations on this site, see Coltekin & Goodman 2018.

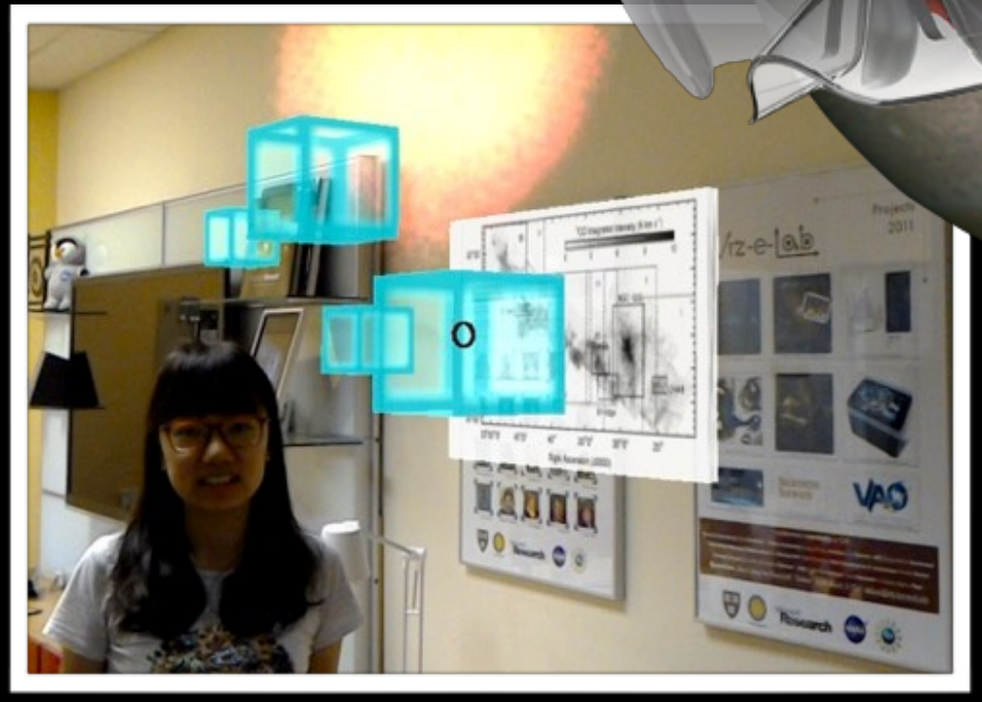
2019: glue in the browser



Video courtesy of Maarten Breddels, consulting developer

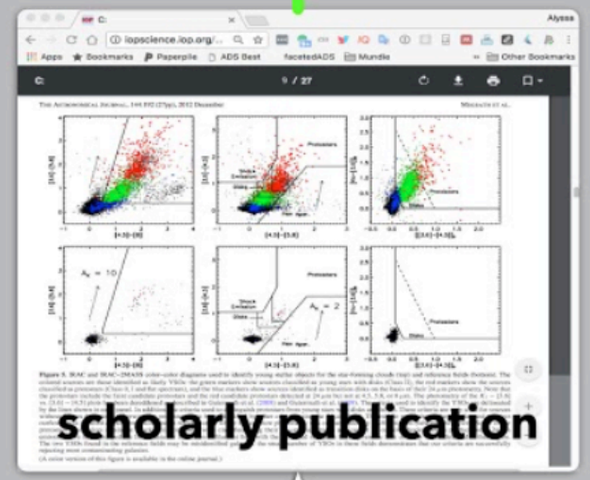
The challenge of 3D Selection

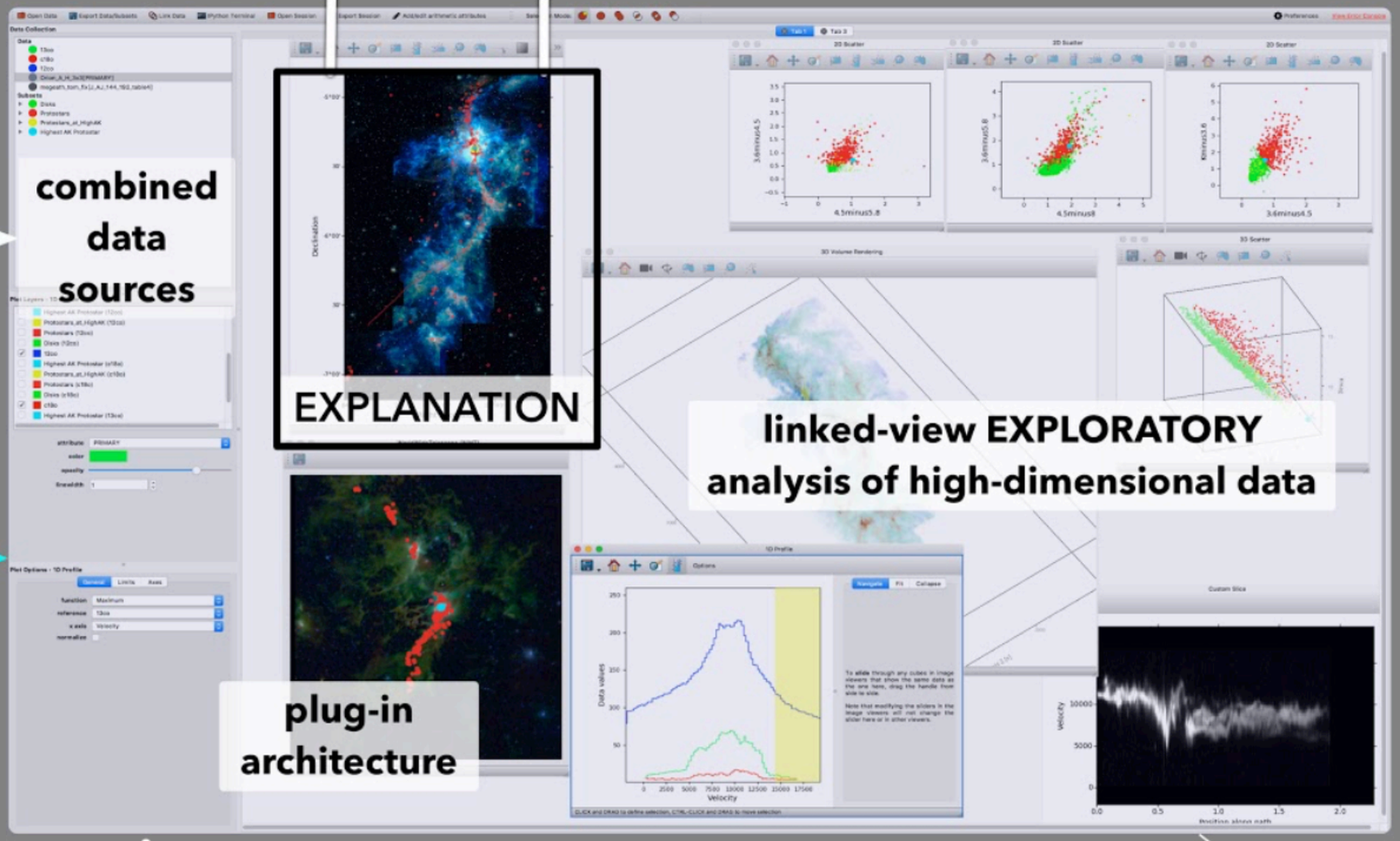




DATA,
CODE,
COLLABORATION

DATA-DRIVEN STORYTELLING





EXPLORATORATION ⇄ EXPLANATION

glue

multidimensional data exploration

glueviz.org github.com/glue-viz glueviz.slack.com

Linked-View Exploratory Visualization of High-Dimensional Data, for Everyone

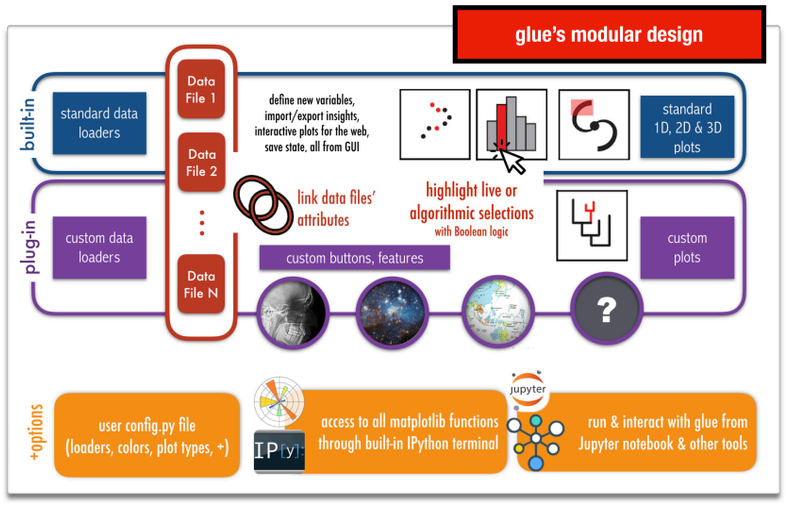
Alyssa **Goodman** (PI, Harvard)
Michellle **Borkin** (PI, Northeastern)
Thomas **Robitaille** (Lead Architect)

The **glue** project was founded in 2012, with funding from **NASA**'s **James Webb Space Telescope** (JWST) project. NASA contracts continue to support development of JWST-related (Astronomy) functionality.

Beginning in 2017, glue has also been funded by the **National Science Foundation**, under SI2-SSE 1739657/1740229: *Collaborative Research: A sustainable future for the glue multi-dimensional linked data visualization package.* The goal of the NSF SSE funding is to expand glue's functionality into domains beyond its traditional strengths in Astronomy and Medicine, by broadening both its user and developer communities.

All glue code is Open Source, at github.com/glue-viz

glue's modular design



Linked Visualizations

With Glue, users can create scatter plots, histograms and images (2D and 3D) of their data. Glue is focused on the brushing and linking paradigm, where selections in any graph propagate to all others.

Flexible linking across data

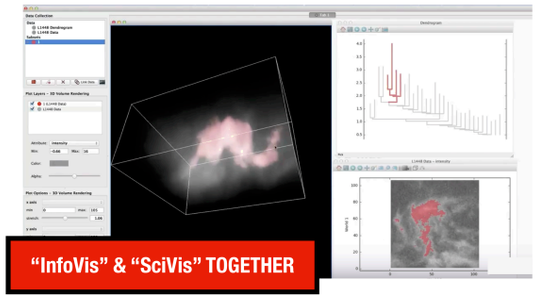
Glue uses the logical links that exist between different data sets to overlay visualizations of different data, and to propagate selections across data sets. These links are specified by the user, and are arbitrarily flexible

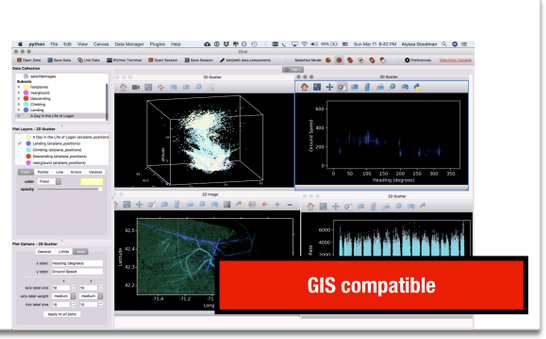
Full scripting capability

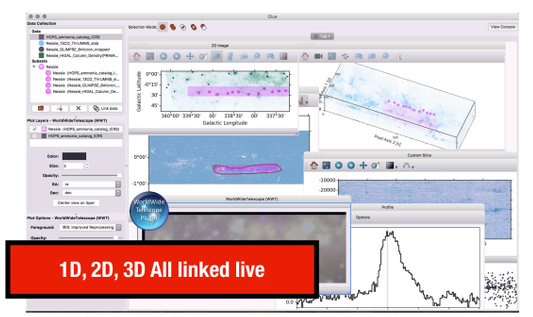
Glue is written in Python, and built on top of its standard scientific libraries (i.e., Numpy, Matplotlib, Scipy). Users can easily integrate their own python code for data input, cleaning, and analysis.

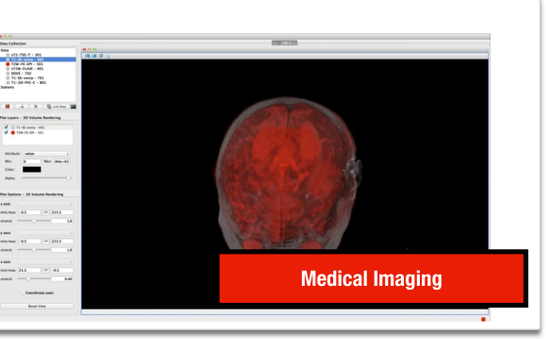
Want to plug-in your project or tool? Consider joining us for glue-con, right after JupyterCon, August 27-29, 2018, at Harvard.

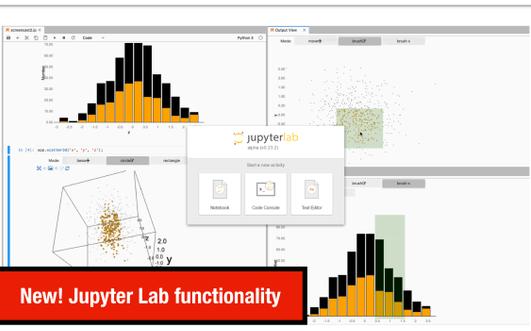
glue-con
multidimensional data exploration
2018, CAMBRIDGE, MA
projects.iq.harvard.edu/gluecon

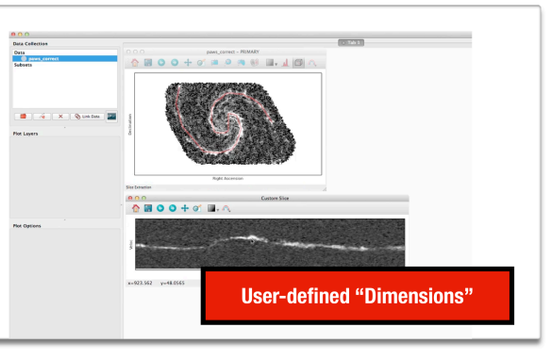








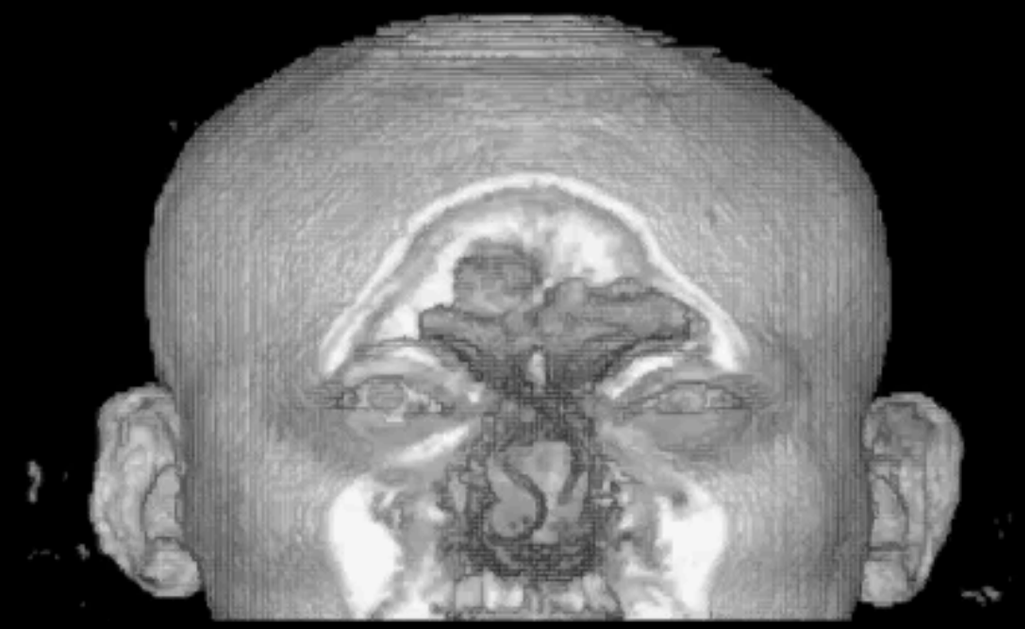




extra slides

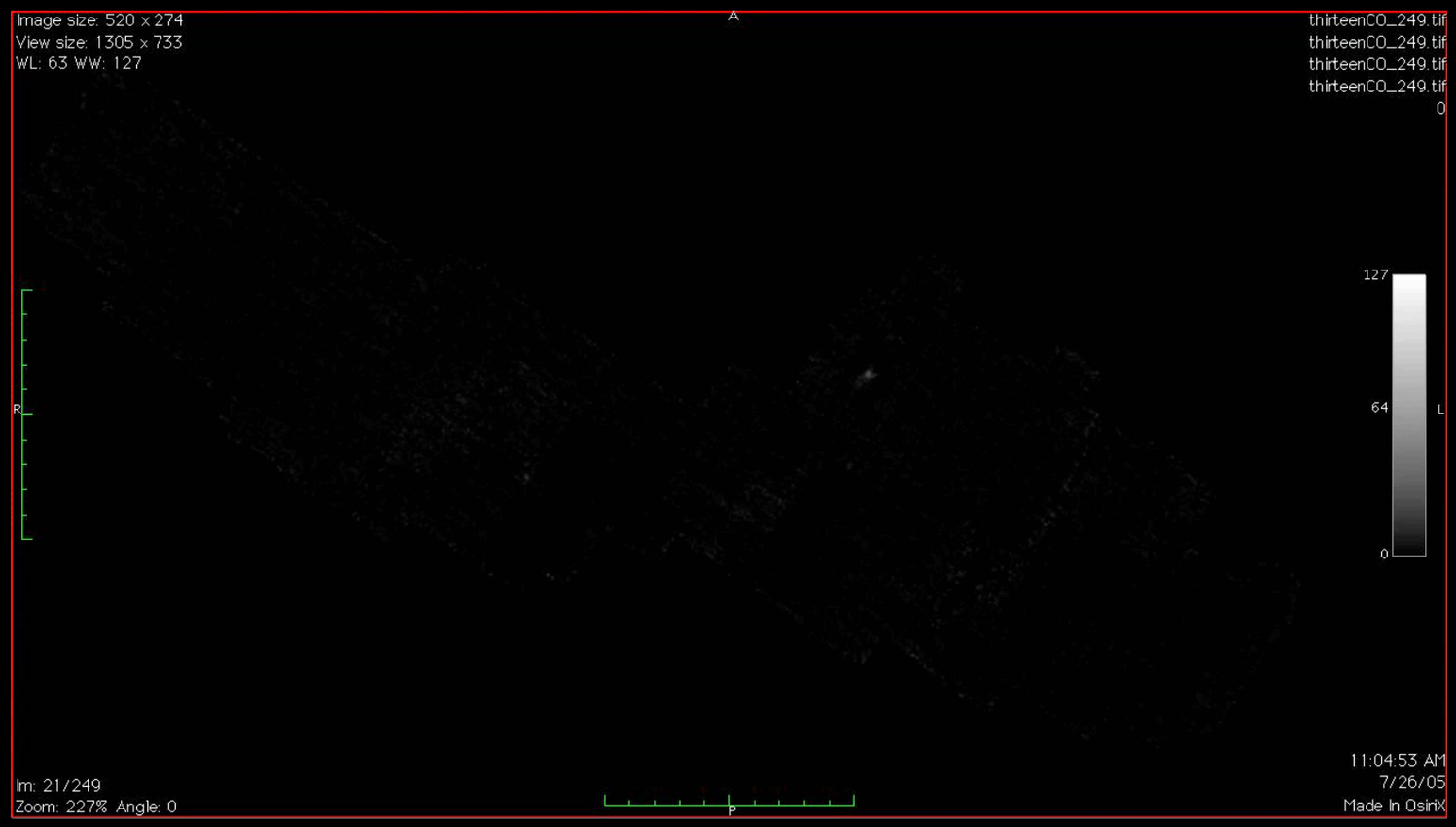
ASTRONOMICAL MEDICINE

“KEITH”



“z” is depth into head

“PERSEUS”



“z” is line-of-sight velocity

ASTRONOMICAL MEDICINE

mage size: 520 x 274
iew size: 1305 x 733
L: 63 WW: 127

mm peak (Enoch et al. 2006)

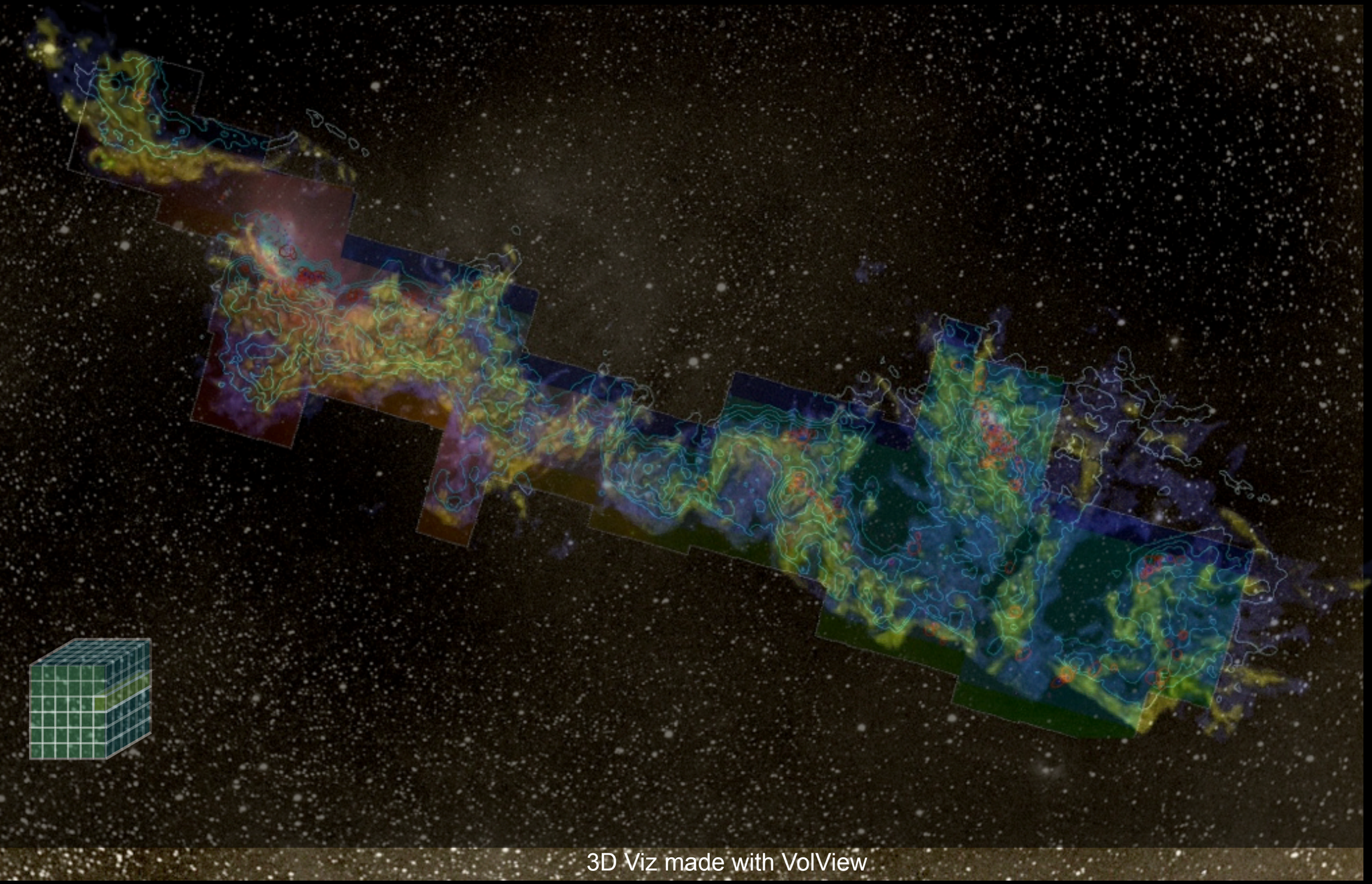
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

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

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

Optical image (Barnard 1927)

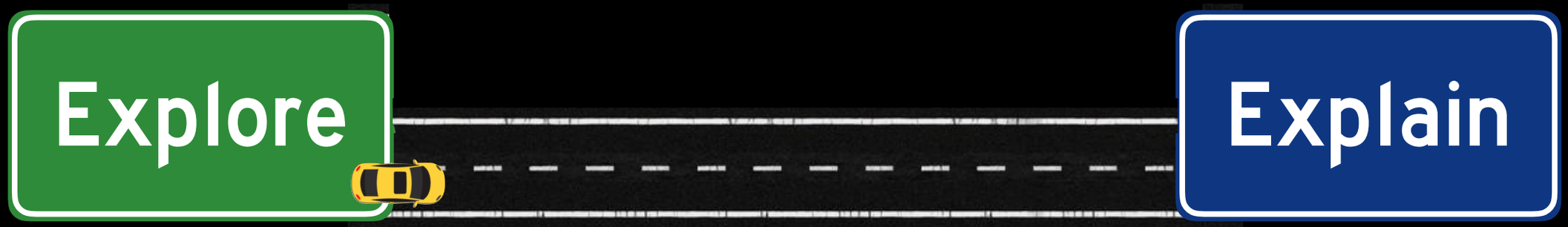
n: 1/249
oom: 227% Angle: 0

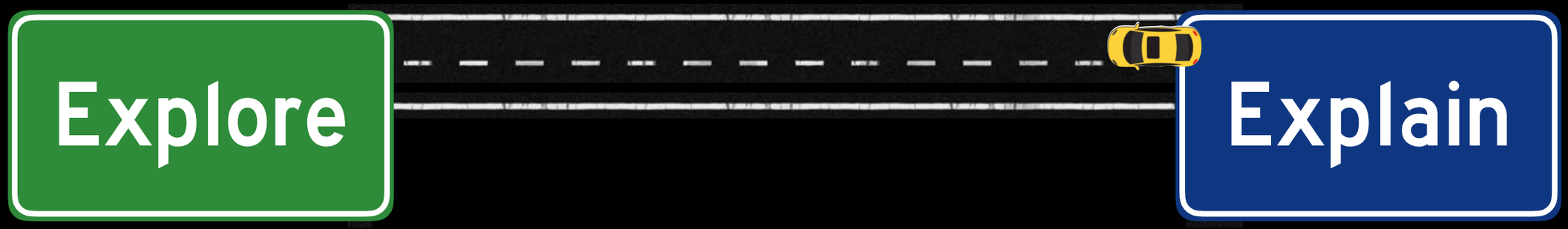


3D Viz made with VolView

COMPLETE

Traditionally, travel from exploration to explanation is called
"Scholarly Publishing" if its *dry*, and "Public Outreach," if it's *beautiful*.





It’s much harder to go the other way.

Explore
And, the *best* roads are two-way.
Explain