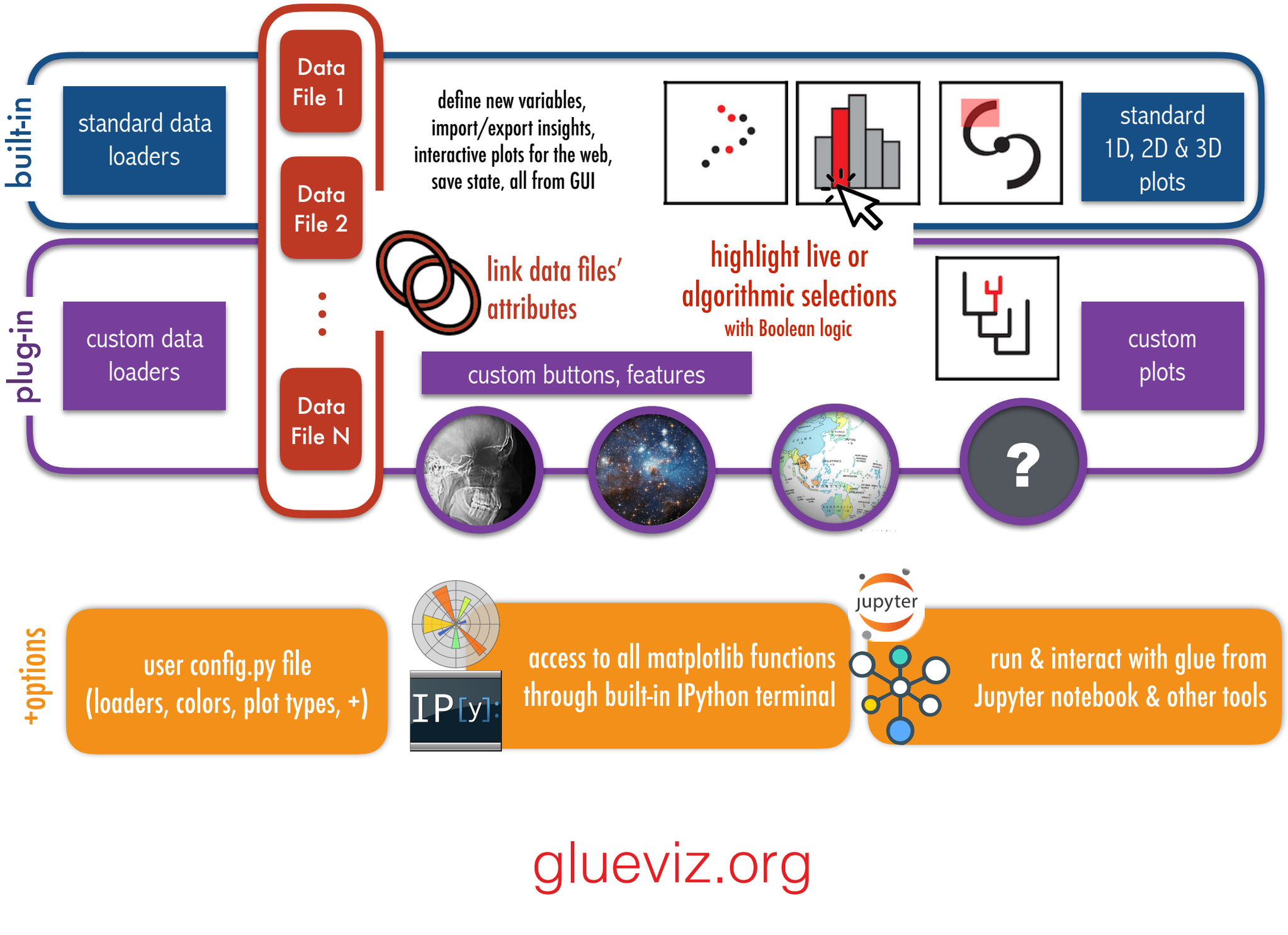
Thoughts on High-Dimensional Data Visualization, with glue

multidimensional data exploration

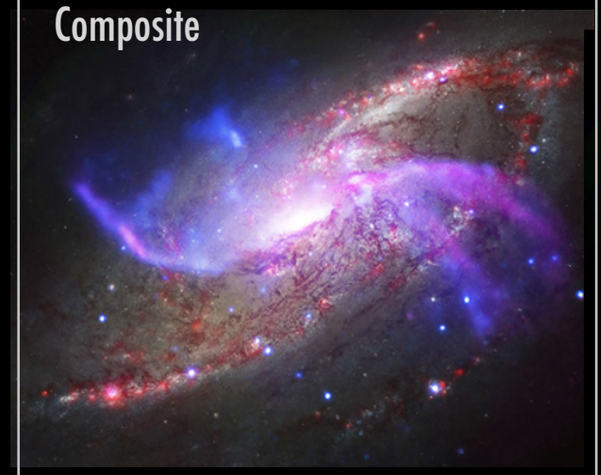
a discussion at Steward Observatory with Alyssa Goodman, March 23, 2018



glueviz.org

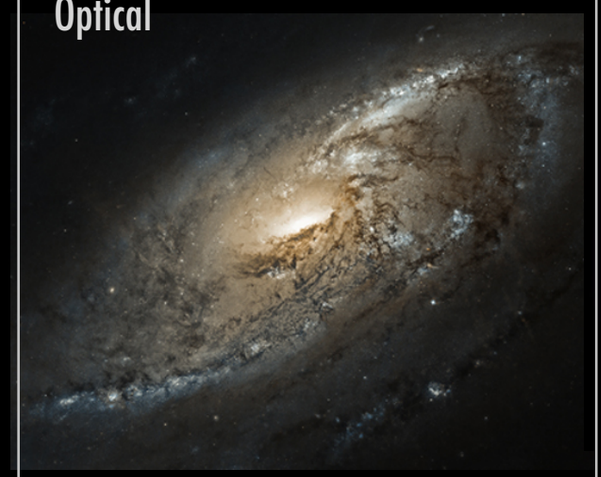
"ASTRONOMICAL MEDICINE"

Composite

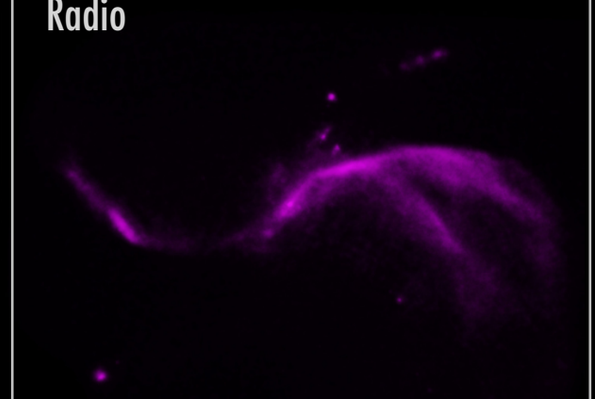


X-ray

Optical

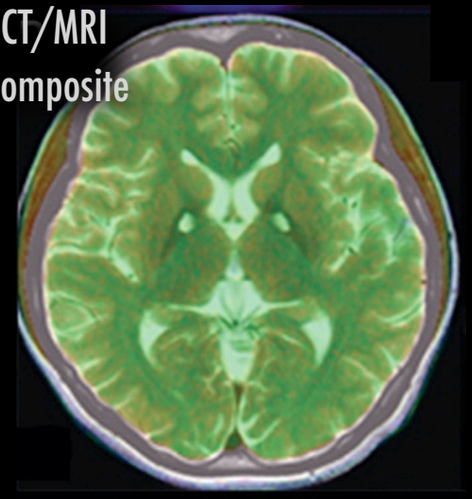


Radio

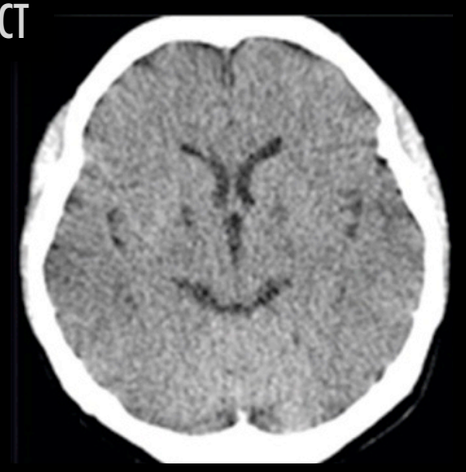


chandra.harvard.edu/photo/2014/m106/

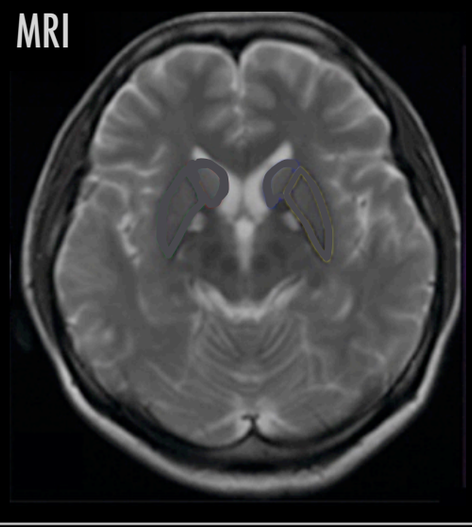
CT/MRI composite



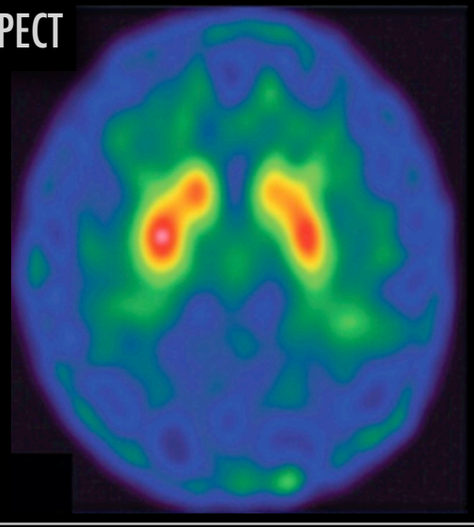
CT



MRI



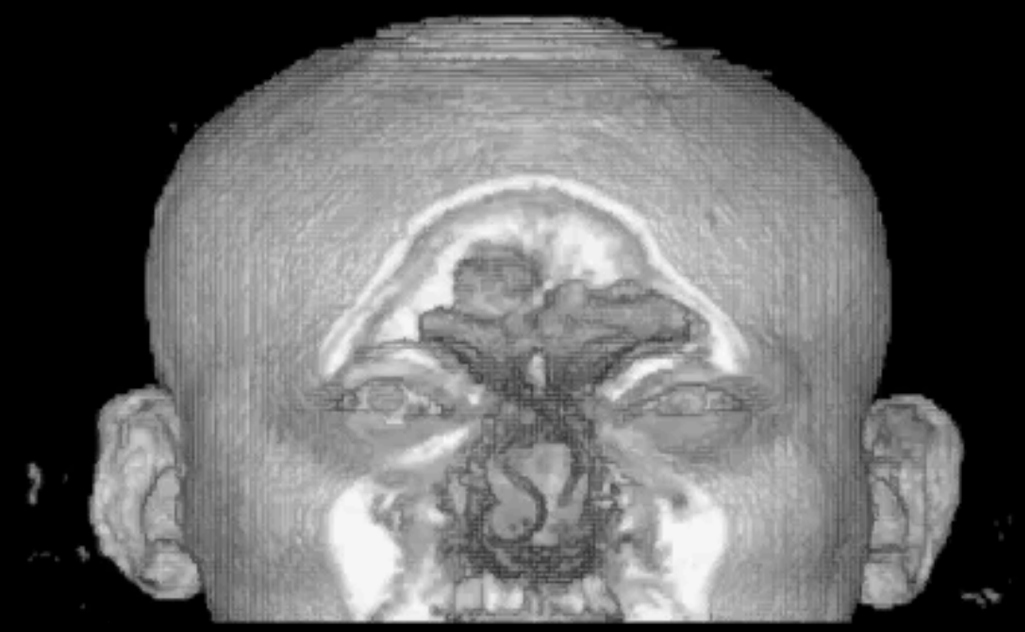
SPECT



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

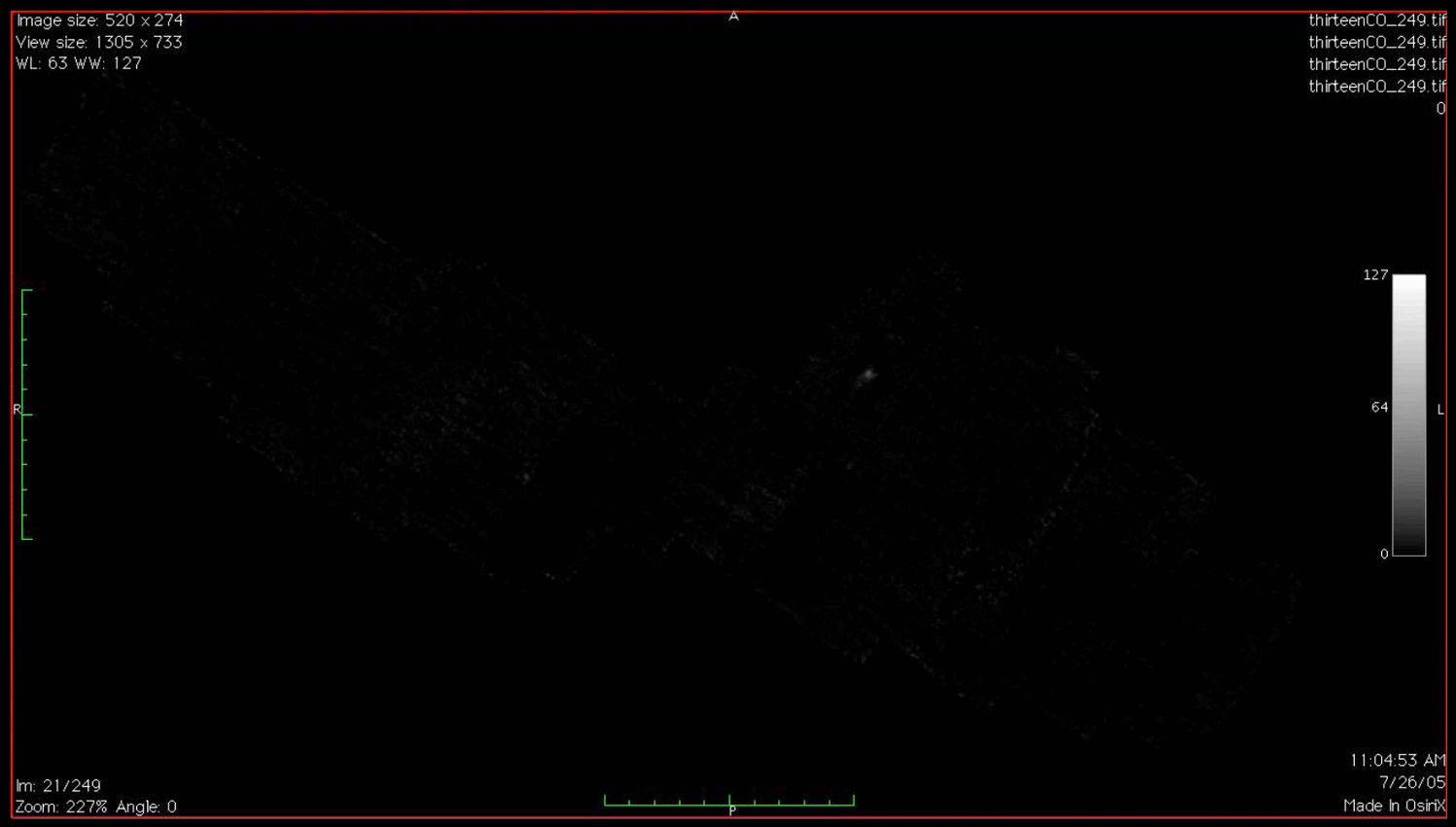
ASTRONOMICAL MEDICINE

“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)

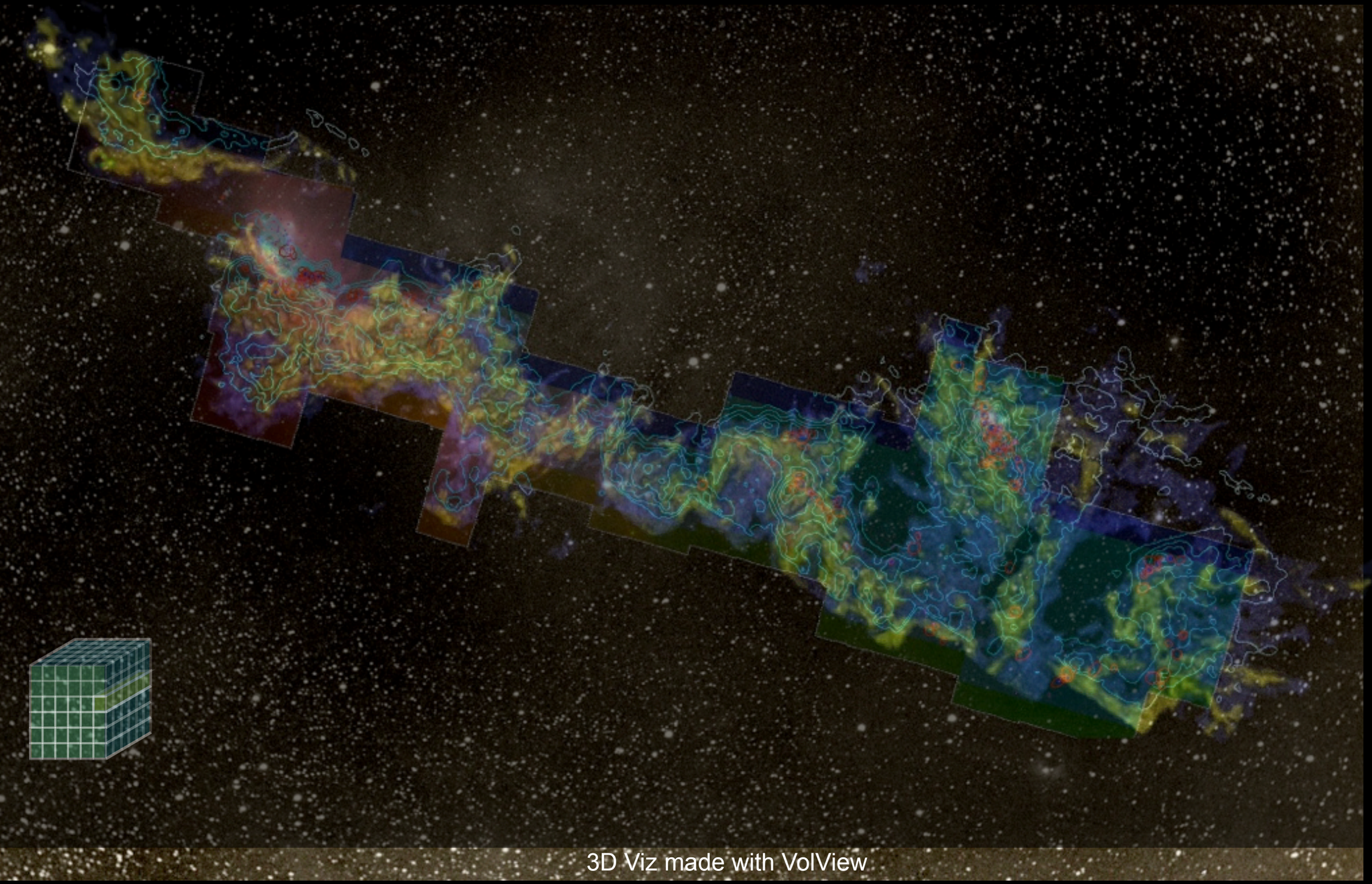
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

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

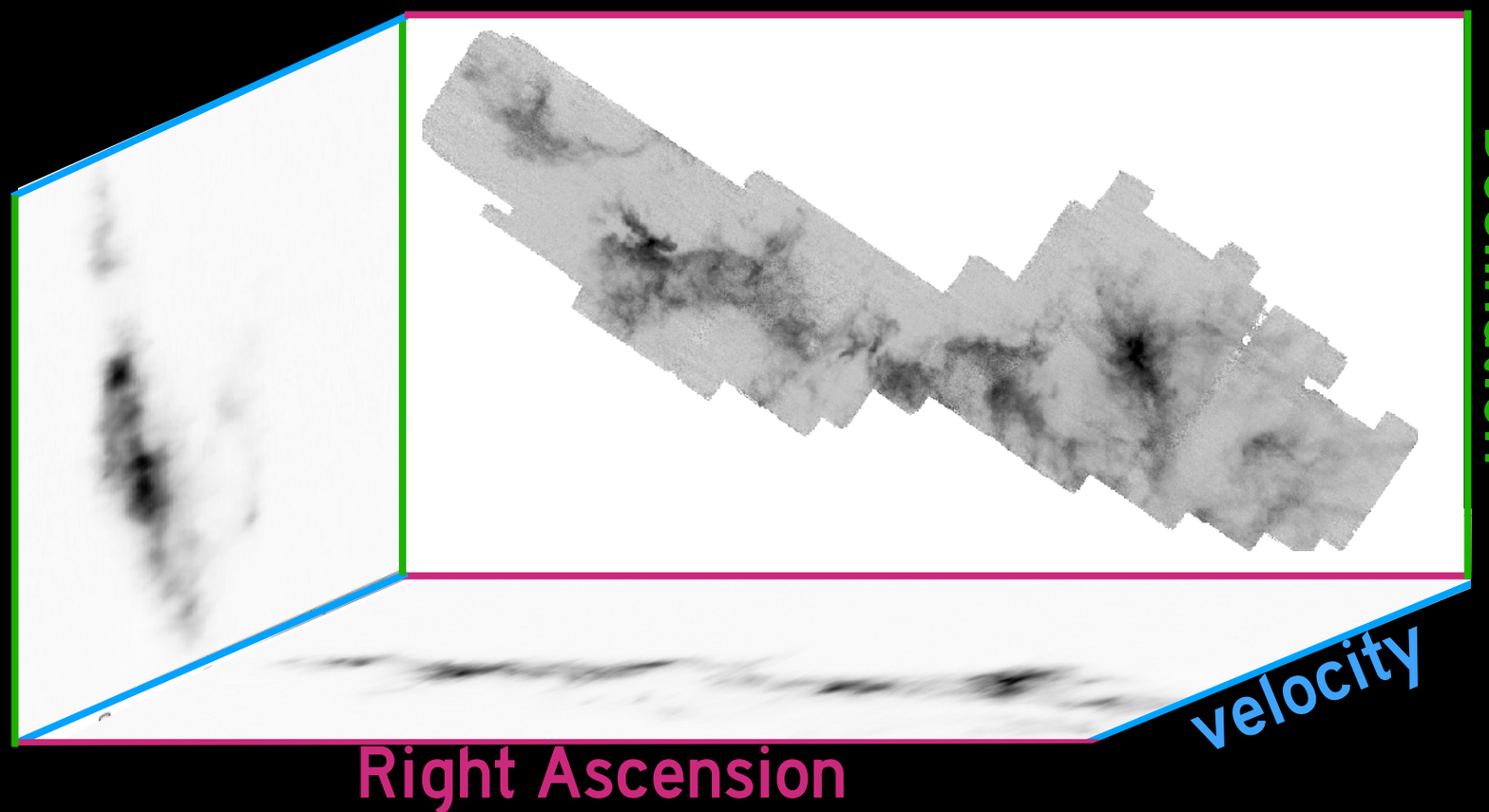
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)

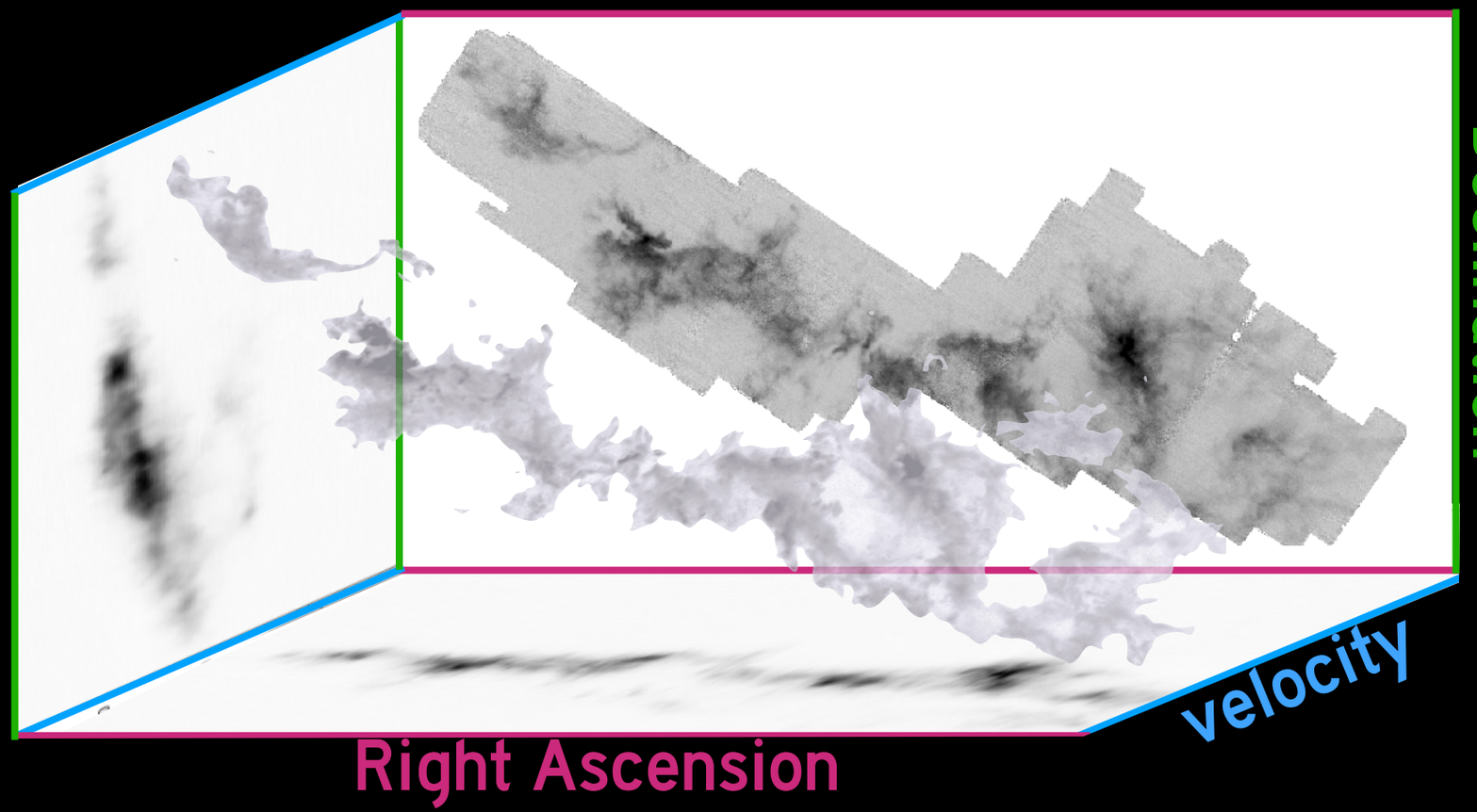
Optical image (Barnard 1927)

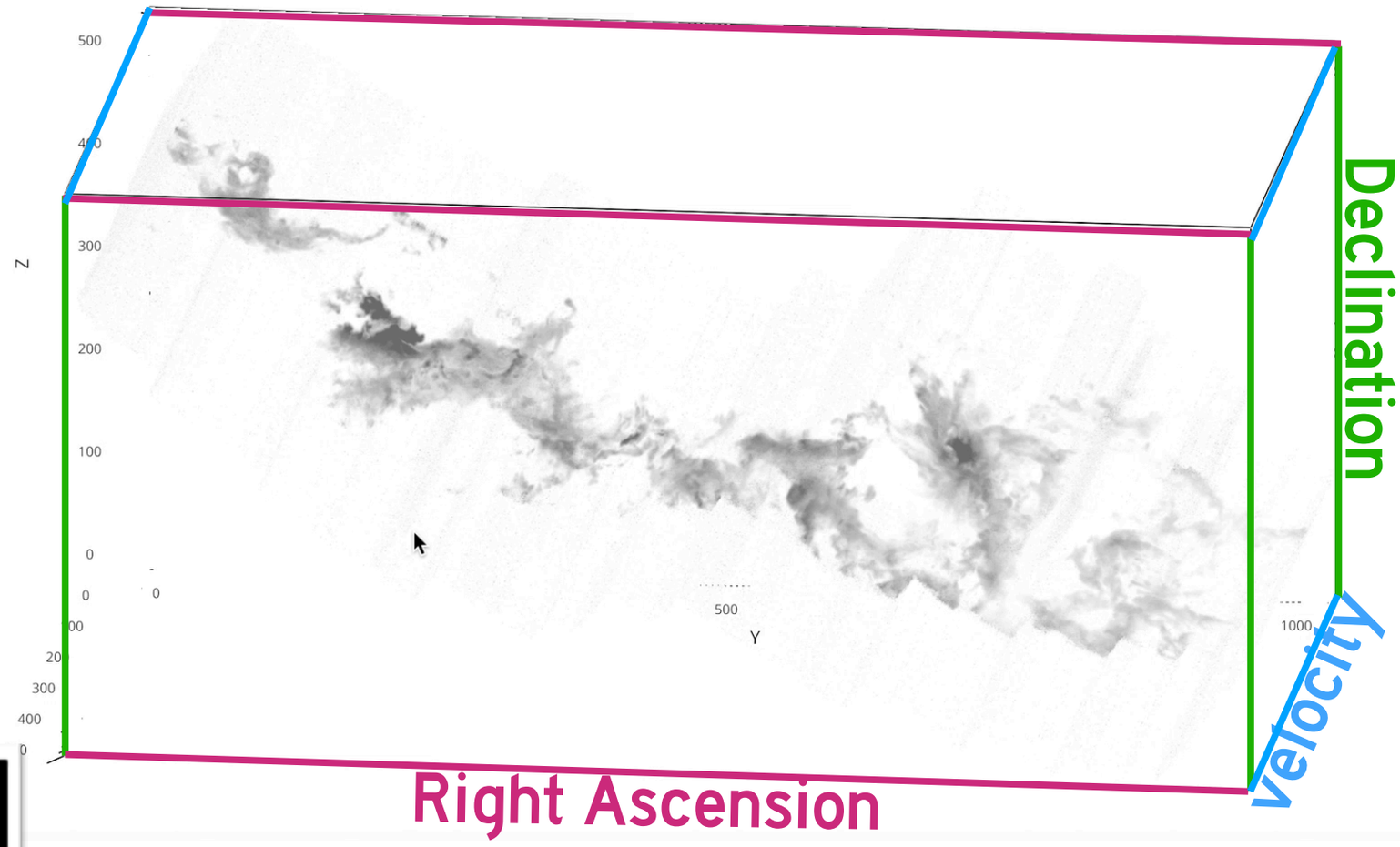
n: 1/249
oom: 227% Angle: 0

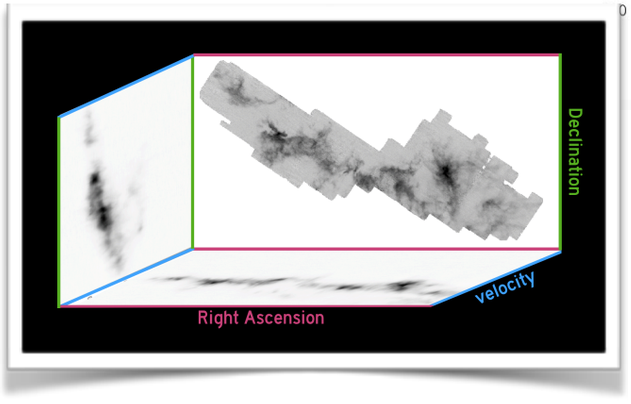


COMPLETE


Declination
Right Ascension
velocity


Declination
velocity
Right Ascension


Declination
velocity
Right Ascension
Z
Y
X
500
400
300
200
100
0
1000


Declination
velocity
Right Ascension

python File Edit View Canvas Data Manager Plugins Help 93% Sat Mar 10 1:29 PM Alyssa Goodman

Glue

Open Data | Save Data | Link Data | IPython Terminal | Open Session | Save Session | Add/edit data components | Selection Mode: | Preferences | View Error Console

Data Collection

Data
PerA_13coFCRAO_F_vxy[PRIMARY]
Subsets
IC348

Plot Layers - 2D Image
IC348 (PerA_13coFCRAO_F_vxy[PRIMARY])
PerA_13coFCRAO_F_vxy[PRIMARY]

attribute: PRIMARY
limits: Custom | Linear
4 | 0
lor/opacity | Sync
ntrast/bias | Reset

Options - 2D Image
General | Limits | Axes
mode: Colormaps
aspect: Square Pixels
reference: PerA_13coFCRAO_F_vxy[PRIM
x axis: Right Ascension
y axis: Velocity
Declination | Show real coordinates
241

Tab 1

Velocity (z)

58° 56°

Position Along (Red) Slice (user-defined)

Declination (y)

200 400 600 800 1000
Position Along Slice

3D Volume Ren...

python File Edit View Canvas Data Manager Plugins Help 93% Sat Mar 10 1:29 PM Alyssa Goodman

Glue

Open Data | Save Data | Link Data | IPython Terminal | Open Session | Save Session | Add/edit data components | Selection Mode: | Preferences | View Error Console

Data Collection

Data
PerA_13coFCRAO_F_vxy[PRIMARY]
Subsets
IC348

Plot Layers - 2D Image
IC348 (PerA_13coFCRAO_F_vxy[PRIMARY])
PerA_13coFCRAO_F_vxy[PRIMARY]

attribute: PRIMARY
limits: Custom | Linear
4 | 0
lor/opacity | Sync
ntrast/bias | Reset

Options - 2D Image
General | Limits | Axes
mode: Colormaps
aspect: Square Pixels
reference: PerA_13coFCRAO_F_vxy[PRIM
x axis: Right Ascension
y axis: Velocity
Declination | Show real coordinates
241

Tab 1

Velocity (z)

58° 56°

Position Along (Red) Slice (user-defined)

Declination (y)

200 400 600 800 1000
Position Along Slice

3D Volume Ren...

python File Edit View Canvas Data Manager Plugins Help 89% Sat Mar 10 1:21 PM Alyssa Goodman

Glue

Open Data | Save Data | Link Data | IPython Terminal | Open Session | Save Session | Add/edit data components | Selection Mode: | Preferences | View Error Console

Data Collection

Data
PerA_13coFCRAO_F_vxy[PRIMARY]
Subsets
IC348

Plot Layers - 2D Image
IC348 (PerA_13coFCRAO_F_vxy[PRIMARY])
PerA_13coFCRAO_F_vxy[PRIMARY]

attribute: PRIMARY
limits: Custom | Linear
4 | 0
lor/opacity | Sync
ntrast/bias | Reset

Options - 2D Image
General | Limits | Axes
mode: Colormaps
aspect: Square Pixels
reference: PerA_13coFCRAO_F_vxy[PRIM
x axis: Right Ascension
y axis: Velocity
Declination | Show real coordinates
241

Tab 1

Velocity (z)

58° 56°

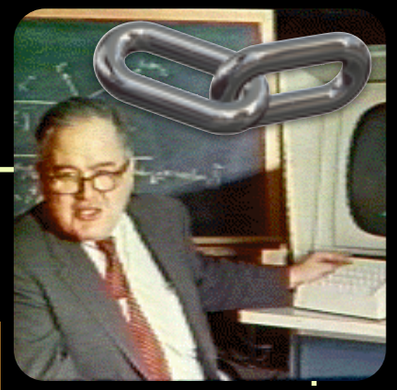
Position Along (Red) Slice (user-defined)

Declination (y)

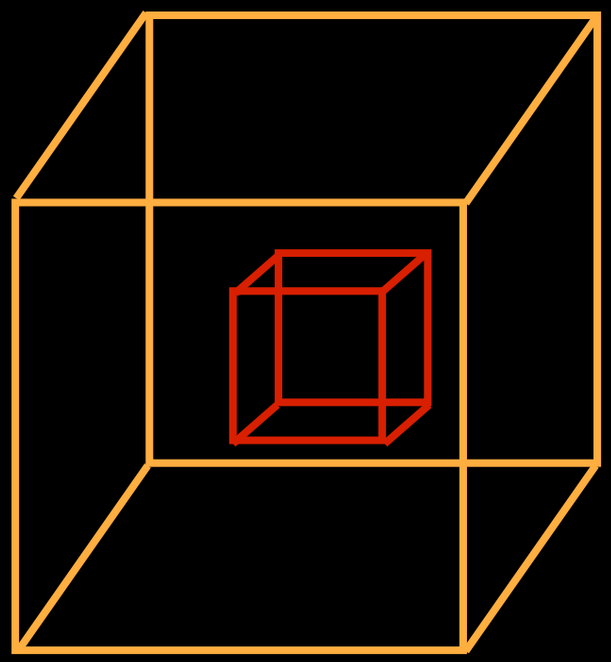
200 400 600 800 1000
Position Along Slice

3D Volume Ren...

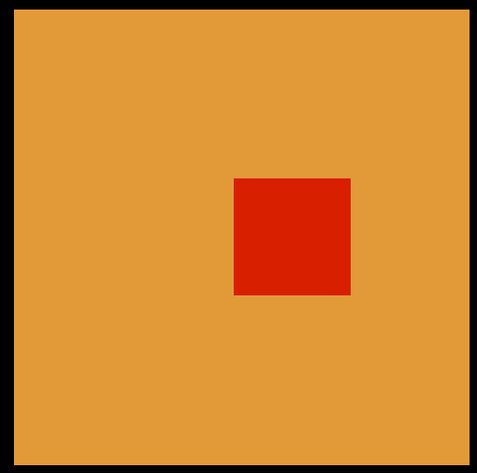
LINKED VIEWS OF HIGH-DIMENSIONAL DATA



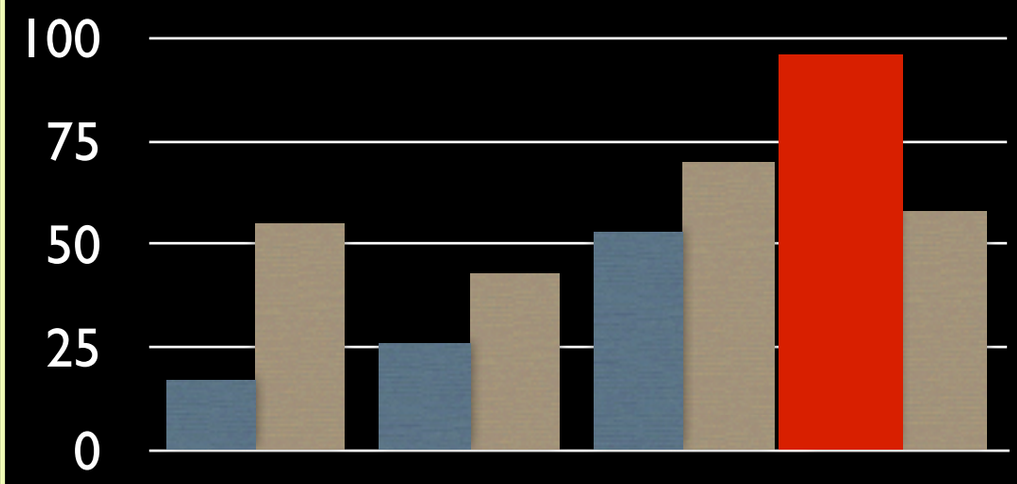
John Tukey



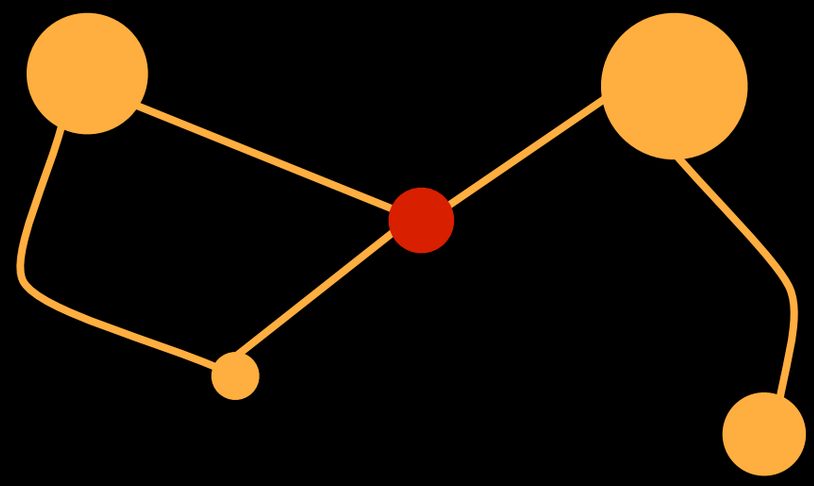
3D



2D

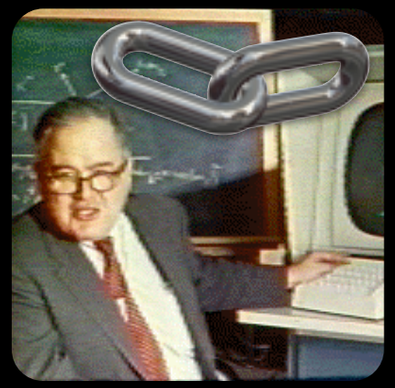


Data Abstraction



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

RGGobi

glue
multidimensional data exploration

Microsoft Power BI

Spotfire®

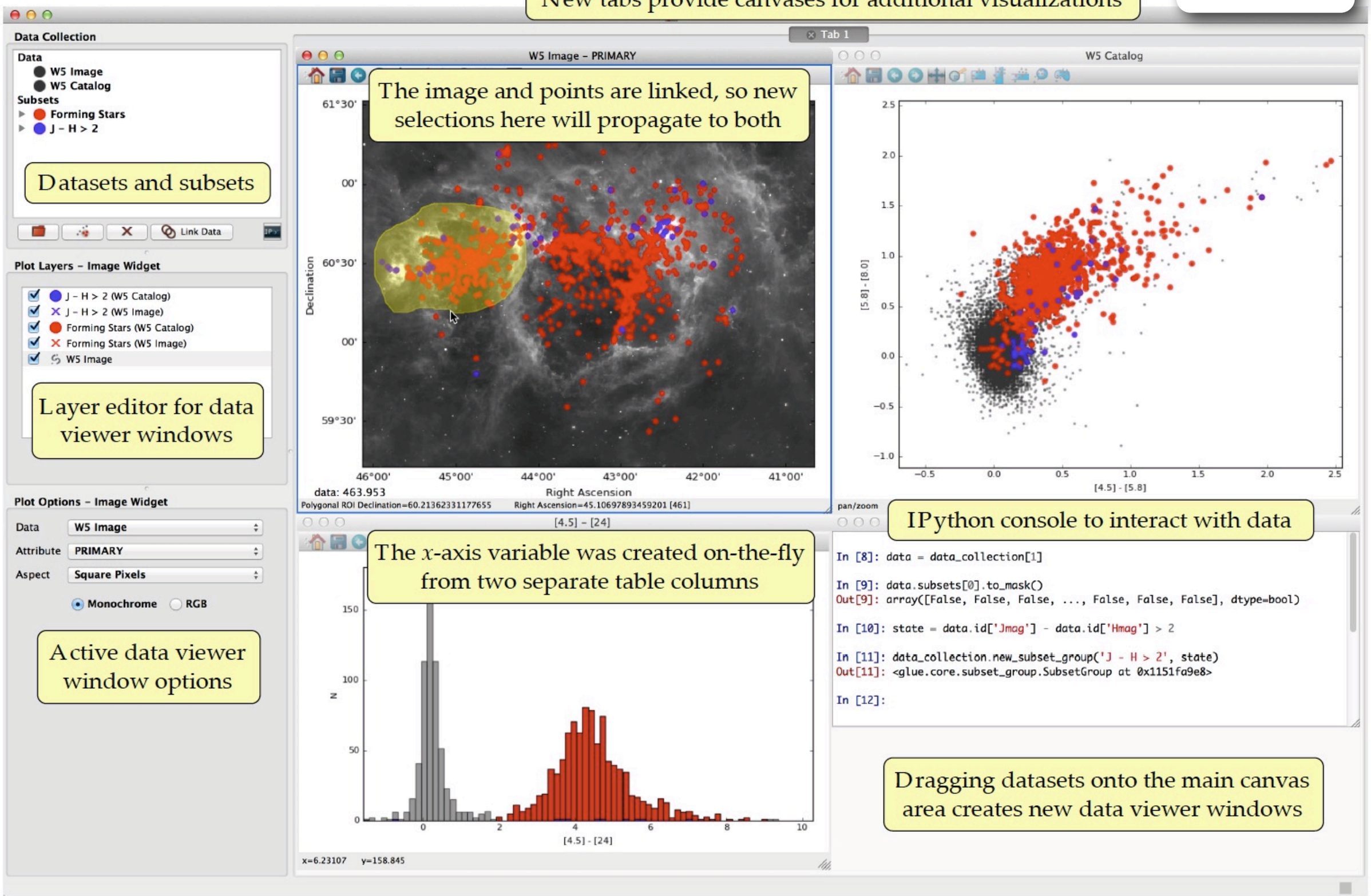
Polaris

tableau SOFTWARE

1970 1980 1990 2000 2010

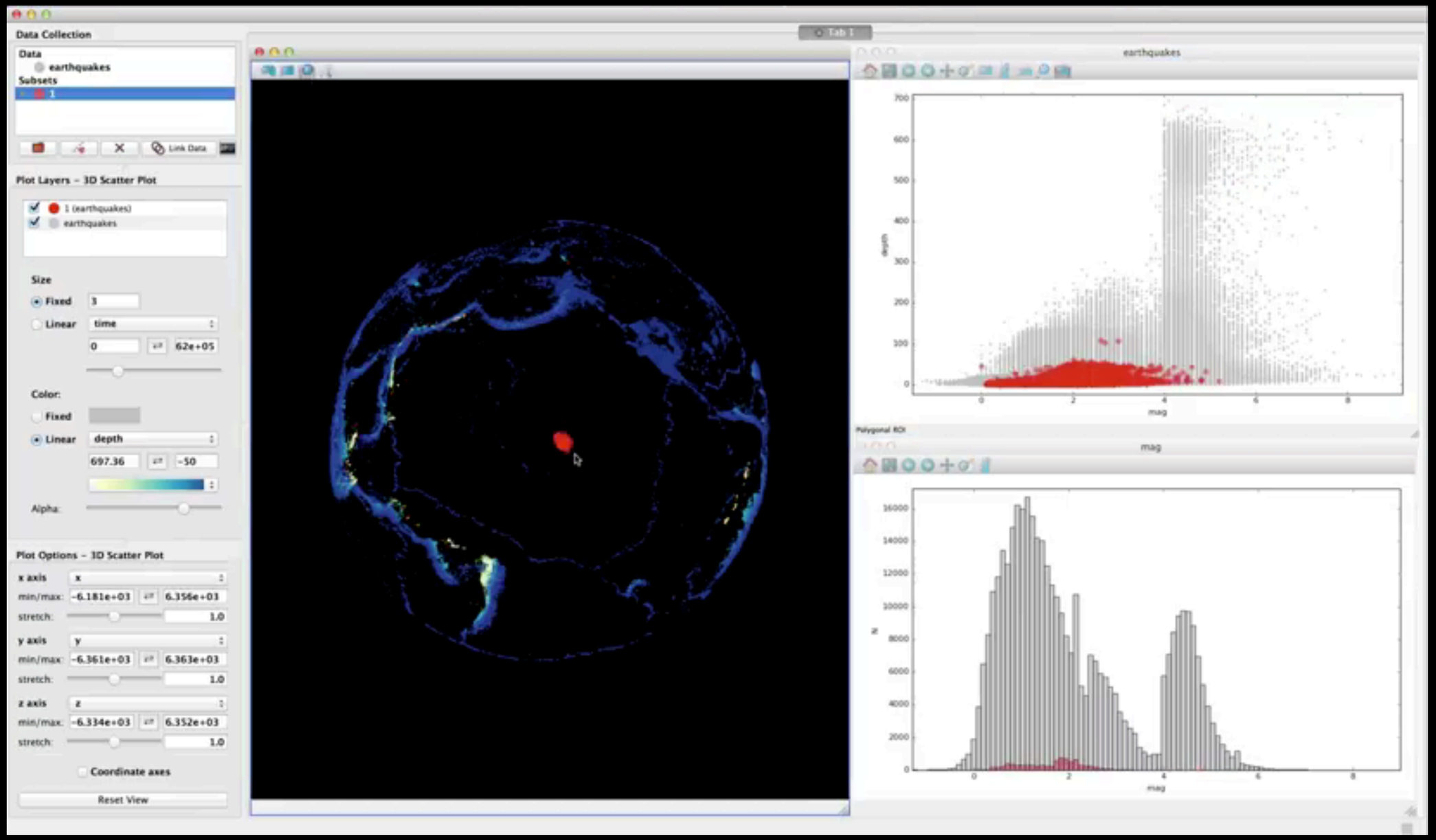
LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE

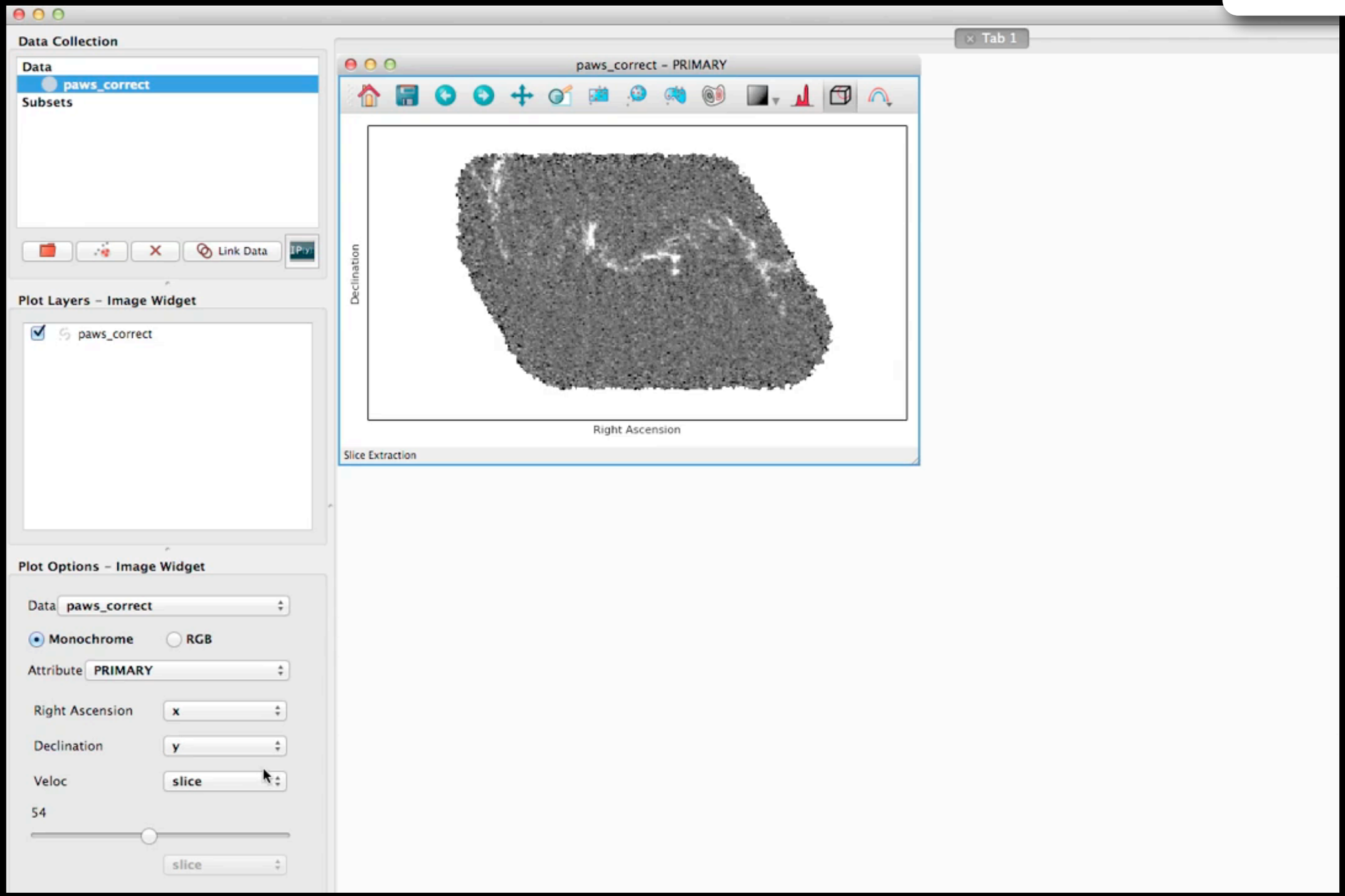


video by Tom Robitaille, lead glue developer

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

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



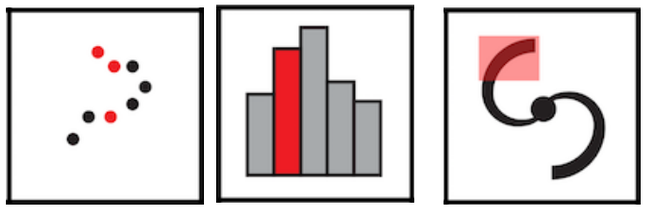
video by Chris Beaumont, glue developer

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

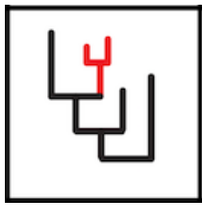

glue
multidimensional data exploration

glue
multidimensional data exploration

Linked Views



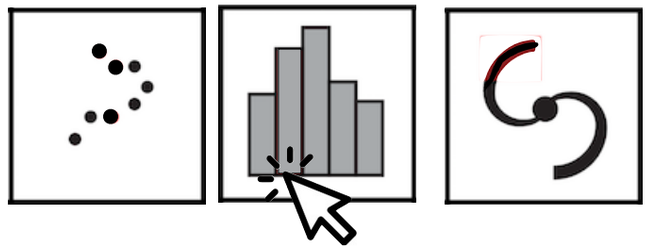
standard
1D, 2D & 3D
plots



custom
plots

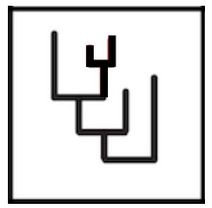
glue
multidimensional data exploration

Linked Views



standard
1D, 2D & 3D
plots

highlight live or
algorithmic selections
with Boolean logic

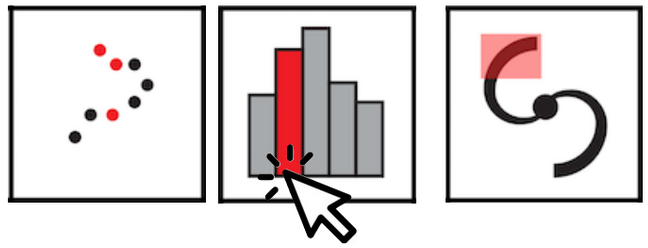


custom
plots

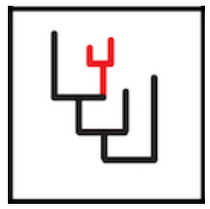
Multiple Data Sets at Once

link data files' attributes

+Linked Views

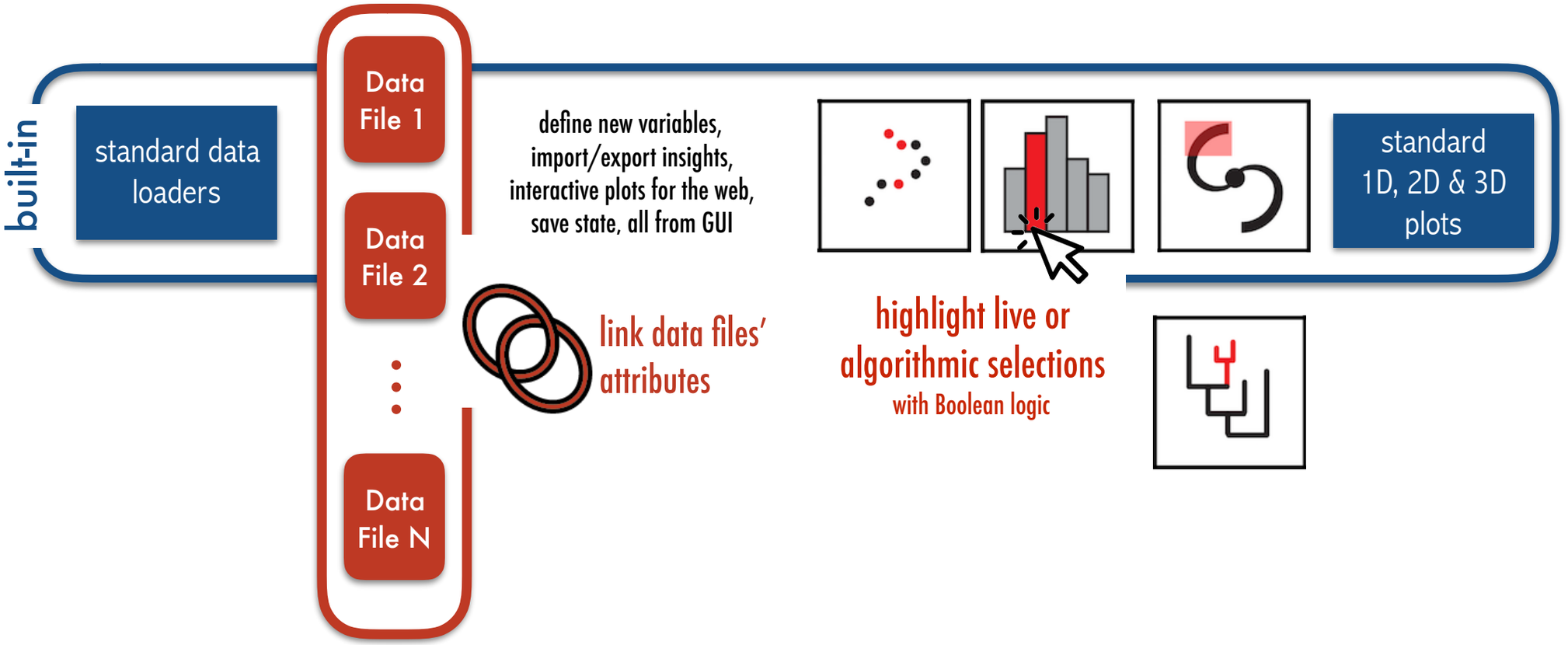


highlight live or algorithmic selections
with Boolean logic



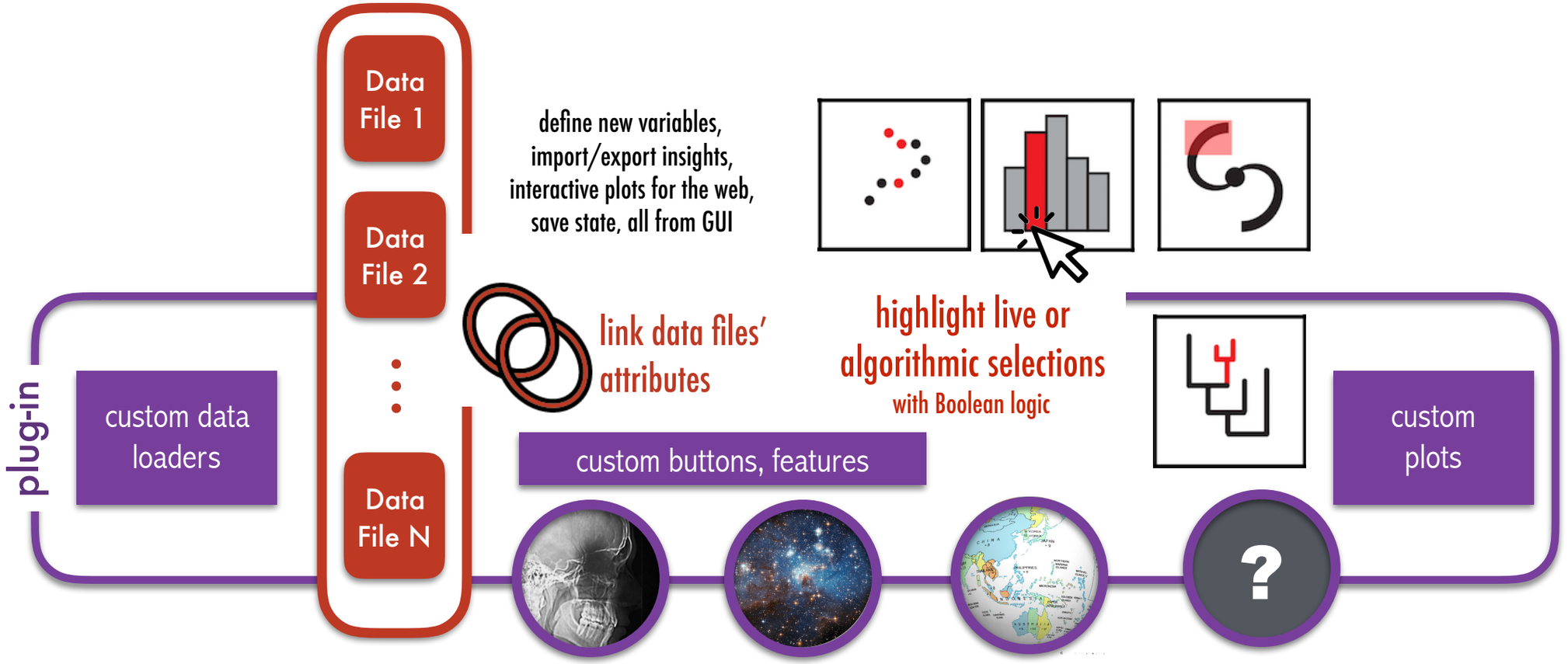
glue

multidimensional data exploration



glue

multidimensional data exploration



glue

multidimensional data exploration

built-in

standard data loaders

user setup "input formats"

plug-in

custom 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

link data files' attributes

highlight live or algorithmic selections with Boolean logic

custom buttons, features

?

standard 1D, 2D & 3D plots

user setup "output formats"

custom plots

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

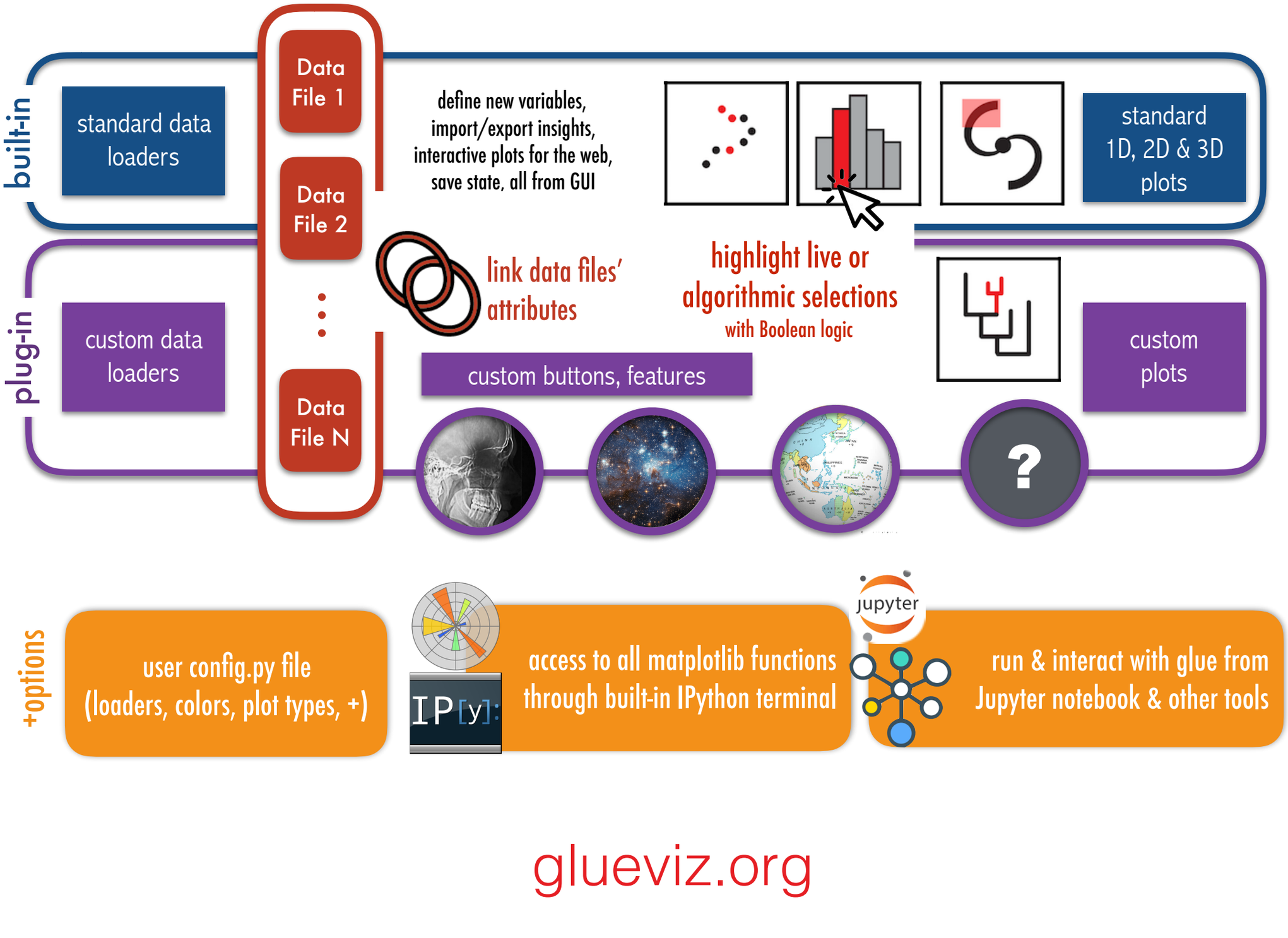
user config.py file (loaders, colors, plot types, +)

access to all matplotlib functions through built-in IPython terminal

run & interact with glue from Jupyter notebook & other tools

glue

multidimensional data exploration



glueviz.org

glue

multidimensional data exploration

your handout

built-in

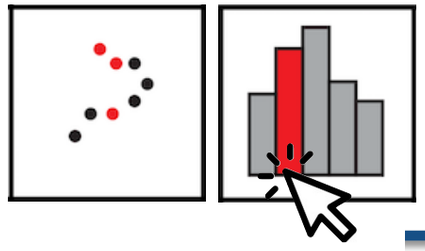
standard data loaders

Data File 1

Data File 2

Data File N

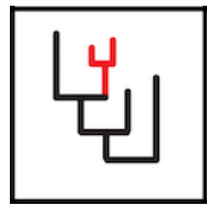
define new variables, import/export insights, interactive plots for the web, save state, all from GUI



standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic

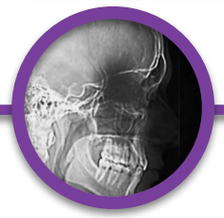


custom plots

plug-in

custom data loaders

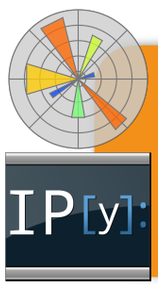
custom buttons, features



?

+options

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



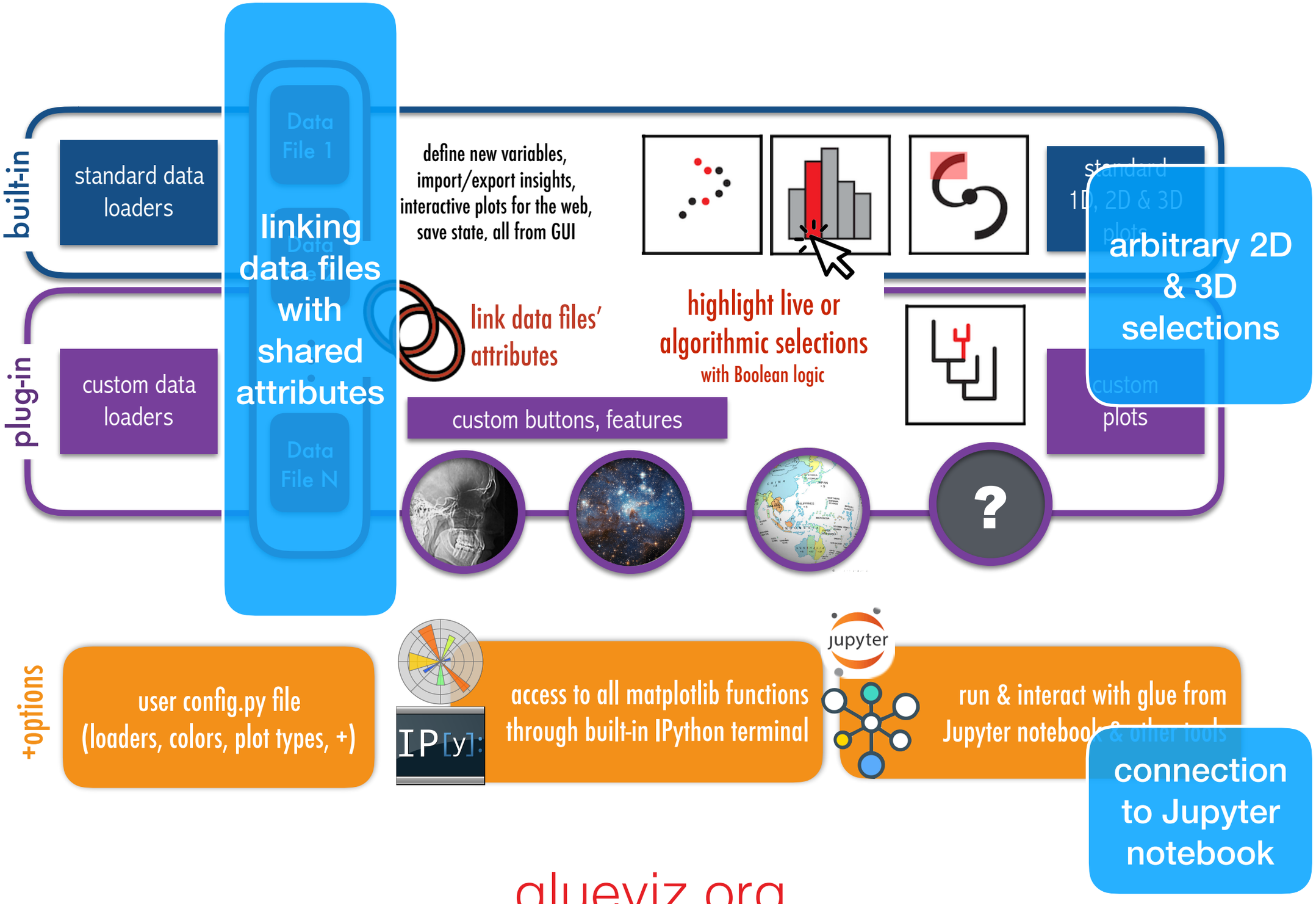
access to all matplotlib functions through built-in IPython terminal

jupyter

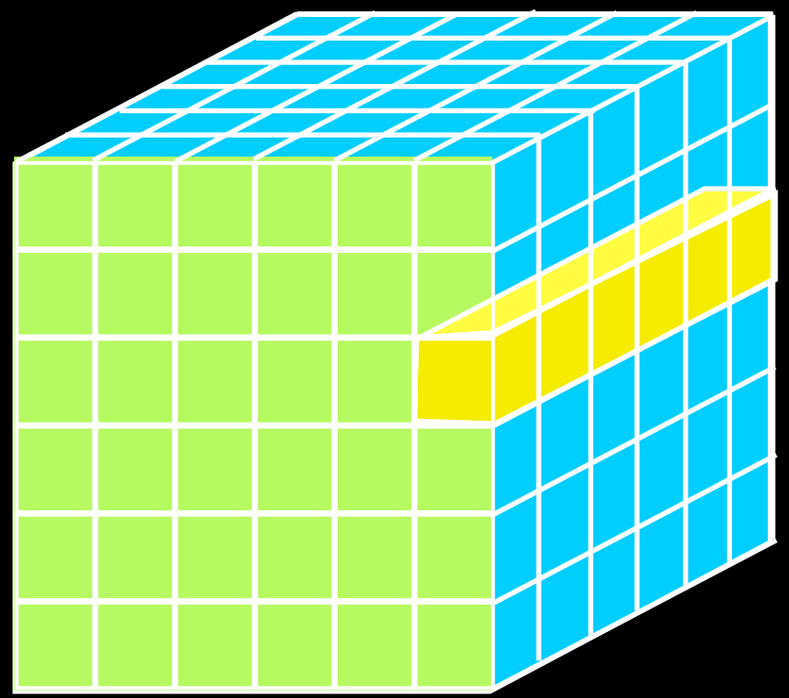
run & interact with glue from Jupyter notebook & other tools

glueviz.org

Which Parts are Novel?



glueviz.org



DATA-DIMENSIONS-DISPLA

explaining to astronomers

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”

Draft version December 29, 2017
Typeset using LaTeX **preprint** style in AASTeX61

PHYSICAL PROPERTIES OF LARGE-SCALE GALACTIC FILAMENTS

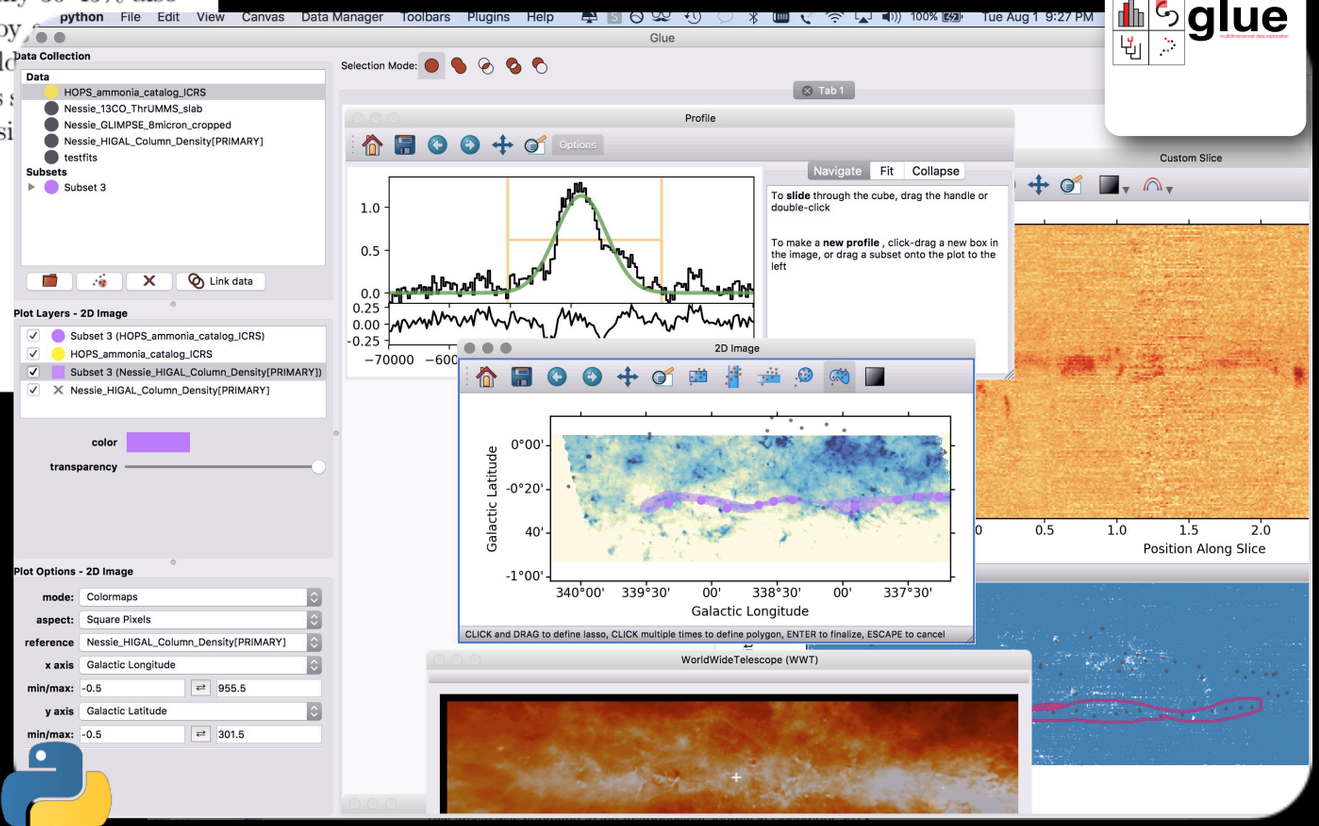
Catherine Zucker,[1] Cara Battersby,[1,2] and Alyssa Goodman[1]

[1] *Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138*
[2] *Department of Physics, University of Connecticut, Storrs, CT 06269, USA*

Abstract

The characterization of our Galaxy's longest filamentary gas features has been the subject of several studies in recent years, producing not only a sizeable sample of large-scale filaments, but also confusion as to whether all these features (e.g. "Bones", "Giant Molecular Filaments") are essentially the same. They are not. We undertake the first standardized analysis of the physical properties (densities, temperatures, morphologies, radial profiles) and kinematics of large-scale filaments in the literature. We expand and improve upon prior analyses by using the same data sets, techniques, and spiral arm models to disentangle the filaments' inherent properties from selection criteria and methodology. Our results suggest that the myriad filament finding techniques are uncovering different physical structures, with length (11-269 pc), width (1-40 pc), mass (3×10^3 $M_\odot$ $- 1.1 \times 10^6$ $M_\odot$), aspect ratio (3:1 - 117:1), and dense gas fraction (0.2-100%) varying by at least an order of magnitude across the sample of 45 filaments. As part of this analysis, we develop a radial profile fitting code, *RadFil*, which is publicly available. We also perform a *position-position-velocity* (*p-p-v*) analysis on a subset of the filaments and find that while 60%-70% lie in the plane of the Galaxy, only 30-45% also exhibit kinematic proximity to purported spiral arms. In a parameter space defined by s
temperature, and density, we broadly distinguish three filament categories, which could
of different formation mechanisms or histories. Highly elongated "Bone-like" filaments s
potential for tracing gross spiral structure (e.g. arms), while other categories could si
concentrations of molecular gas (GMCs, core complexes).

2017
"The Bone Wars"
(& glue)



python

WWT INTEGRATION

Glue

Data Collection

Data
- HOPS_ammonia_catalog_ICRS
- Nessie_13CO_ThrUMMS_slab
- Nessie_GLIMPSE_8micron_cropped
- Nessie_HIGAL_Column_Density[PRIMA...

Subsets
- Nessie
  - Nessie (HOPS_ammonia_catalog_I...
  - Nessie (Nessie_13CO_ThrUMMS_sl...
  - Nessie (Nessie_GLIMPSE_8micron_...
  - Nessie (Nessie_HIGAL_Column_De...

Link data

Plot Layers - WorldWideTelescope (WWT)
- Nessie (HOPS_ammonia_catalog_ICRS)
- HOPS_ammonia_catalog_ICRS

Color:
Size: 3
Opacity:
RA: ra
Dec: dec
Center view on layer

Plot Options - WorldWideTelescope (WWT)
Foreground: IRIS: Improved Reprocessing
Opacity:
Background: Digitized Sky Survey (Color)
Galactic Plane mode

Selection Mode:

View Console

Tab 1

2D Image

Galactic Latitude
0°00'
-0°15'
30'
45'
340°00' 339°30' 00' 338°30' 00' 337°30'
Galactic Longitude

0°00'
-1°00'

Pixel Axis 1 [y]
300
200
100
0
500
Pixel Axis 2 [x]

Custom Slice
-10000
-20000

WorldWideTelescope (WWT)

Profile
Options
0.0

glue

WWT INTEGRATION

Glue

Data Collection

Data

Nessie_HIGAL_Column_Density[PRIMARY]

HOPS_dense_gas_catalog

Subsets

Nessie_Boundary

Link data

Plot Layers

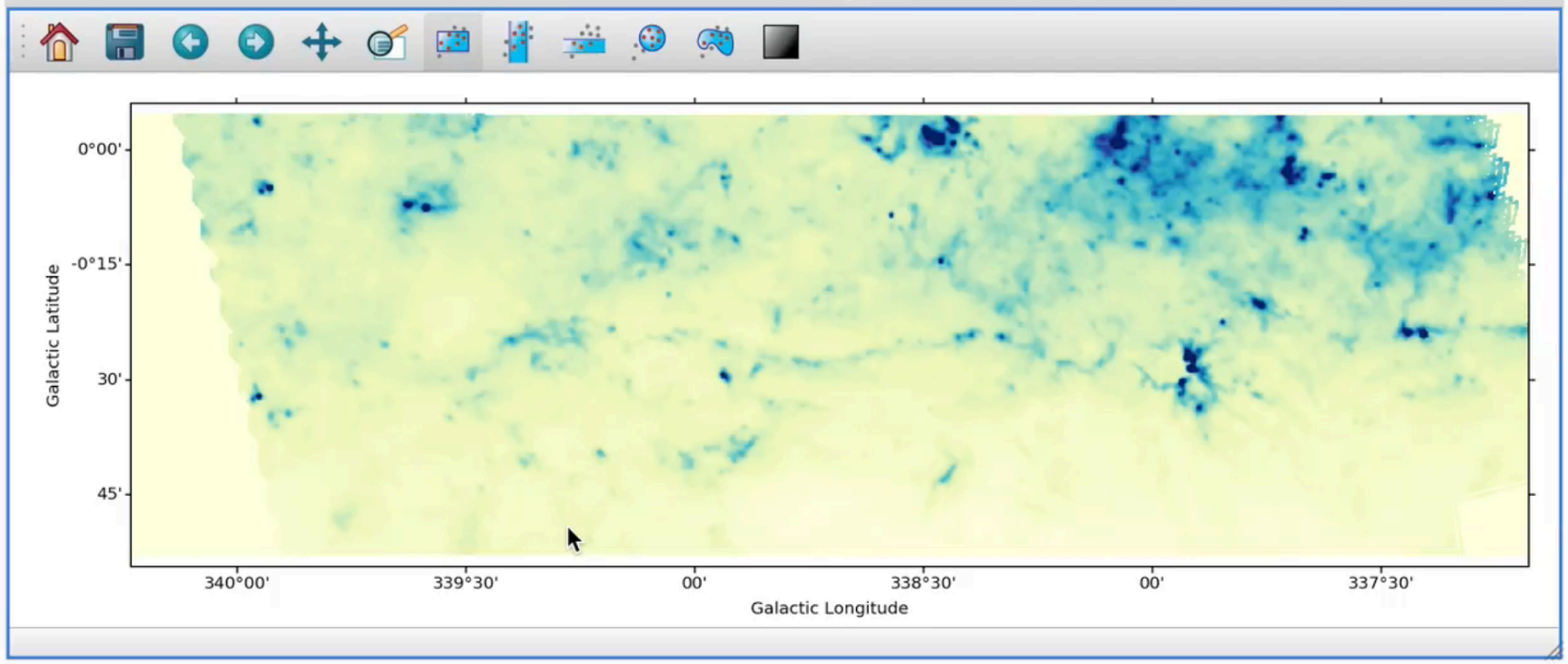
Plot Options

Selection Mode:

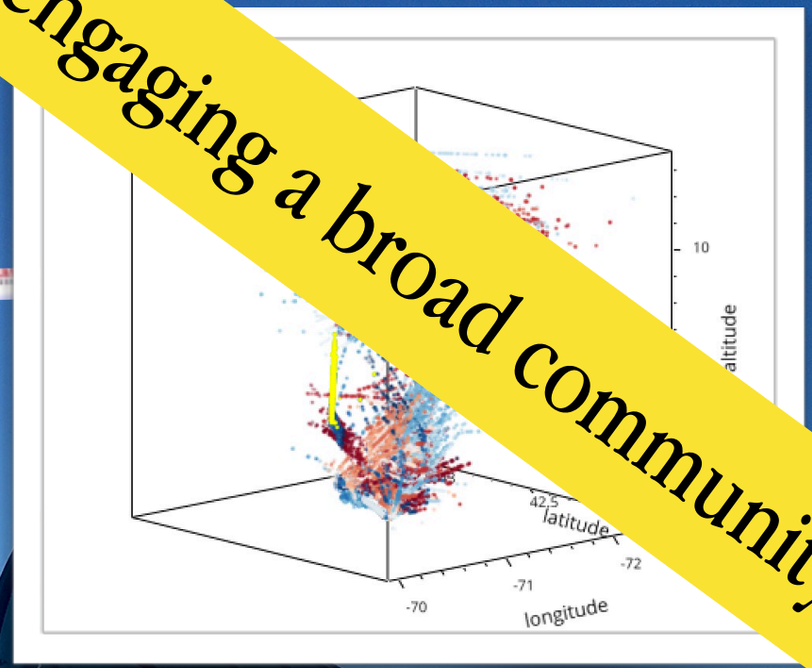
View Console

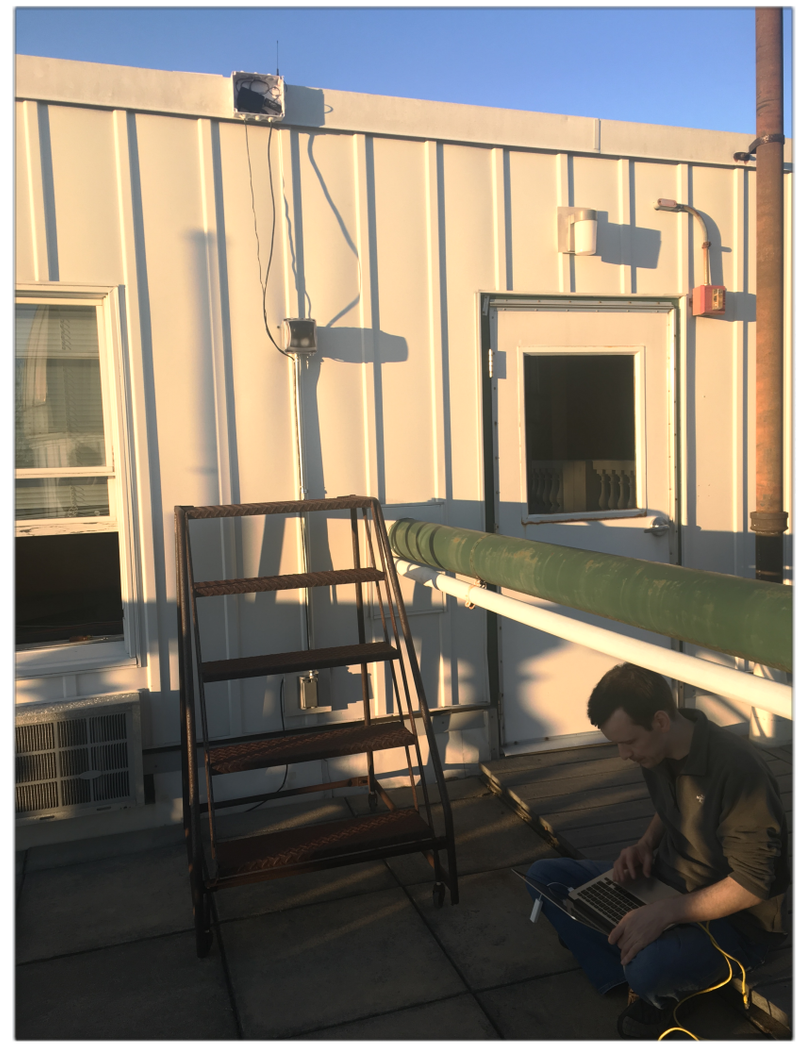
Tab 1

2D Image

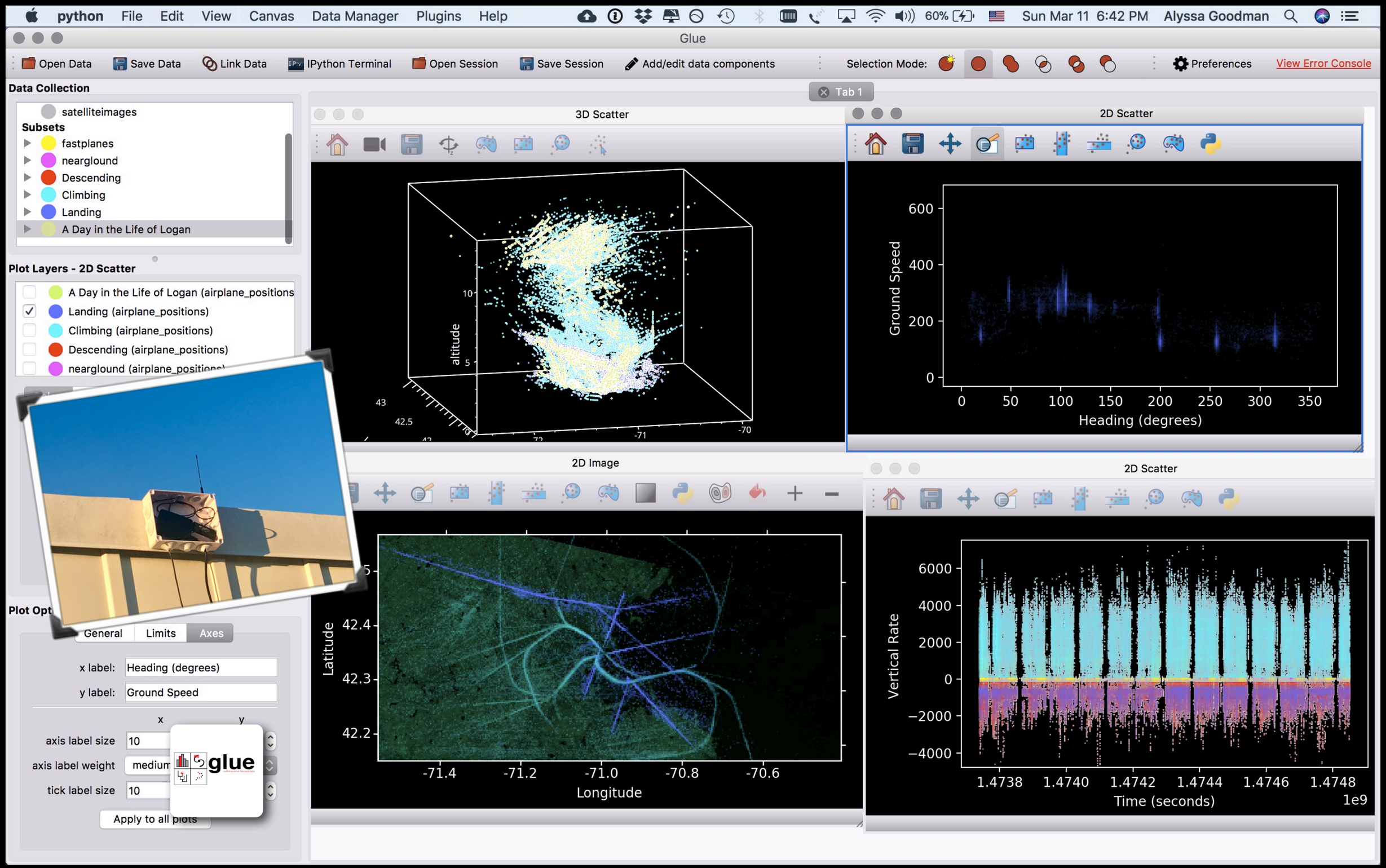


engaging a broad community



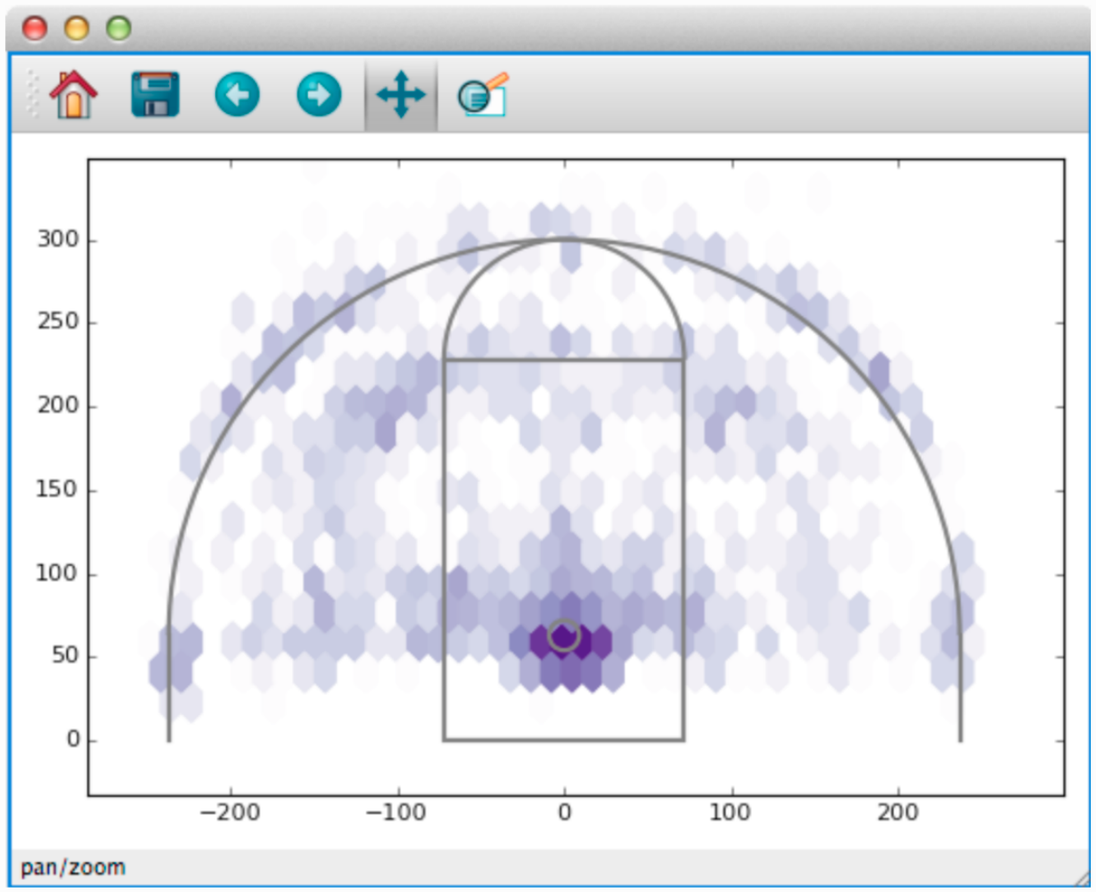


TRACKING PLANES (GIS TOOLS)



CUSTOMIZATION FOR BREADTH

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

JUPYTER LABS

glue "on the web"

Pre-Alpha Jupyter Lab Dem

127.0.0.1:8888/lab

IPython

File Notebook Editor Terminal Console Help

Files Commands

CONSOLE

Clear Cells
Execute Cell Shift-Enter
Interrupt Kernel
New Julia 0.4.5 console
New Python 2 console
New Python 3 console
New R console
Switch Kernel

EDITOR

Close all files
Line Numbers
Line Wrap
Match Brackets
Save File
Vim Mode
Vim Mode Off

FILE OPERATIONS

Close All Ctrl-Shift-Q
Close Document Ctrl-Q
New Notebook Ctrl-Shift-N
New Text File Ctrl-O
Revert Document
Save Document Cmd-S

HELP

About JupyterLab
FAQ
IPython Reference
JupyterLab Launcher
Markdown Reference
Matplotlib Reference
Notebook Tutorial
Numpy Reference
Pandas Reference
Python Reference
Scipy Lecture Notes
Scipy Reference
SymPy Reference

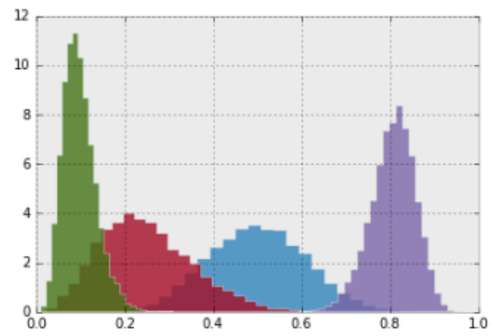
IMAGE WIDGET

Reset Zoom
Zoom In
Zoom Out

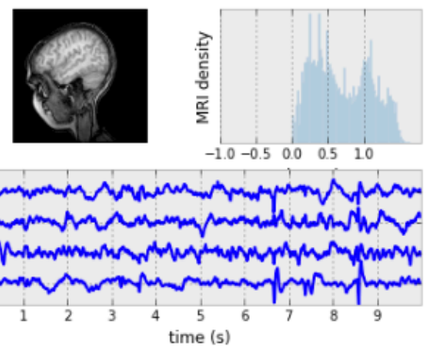
Untitled.ipynb | Terminal 1 | Python 3 (1)

```
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra
details.
```

```
In [1]: %matplotlib inline
        from numpy.random import beta
        import matplotlib.pyplot as plt
        plt.style.use('bmh')

        def plot_beta_hist(a, b):
            plt.hist(beta(a, b, size=10000), histtype="stepfilled",
                     bins=25, alpha=0.8, normed=True)
            return

        plot_beta_hist(10, 10)
        plot_beta_hist(4, 12)
        plot_beta_hist(50, 12)
        plot_beta_hist(6, 55)
```



```
In [2]: %run ~/Downloads/mri_with_eeg.py

  loading eeg /Users/fperez/usr/conda/lib/python3.5/site-packages/mat
  plotlib/mpl-data/sample_data/eeg.dat
```




```
In [ ]:
```

Launcher | mri_with_eeg.py

Running

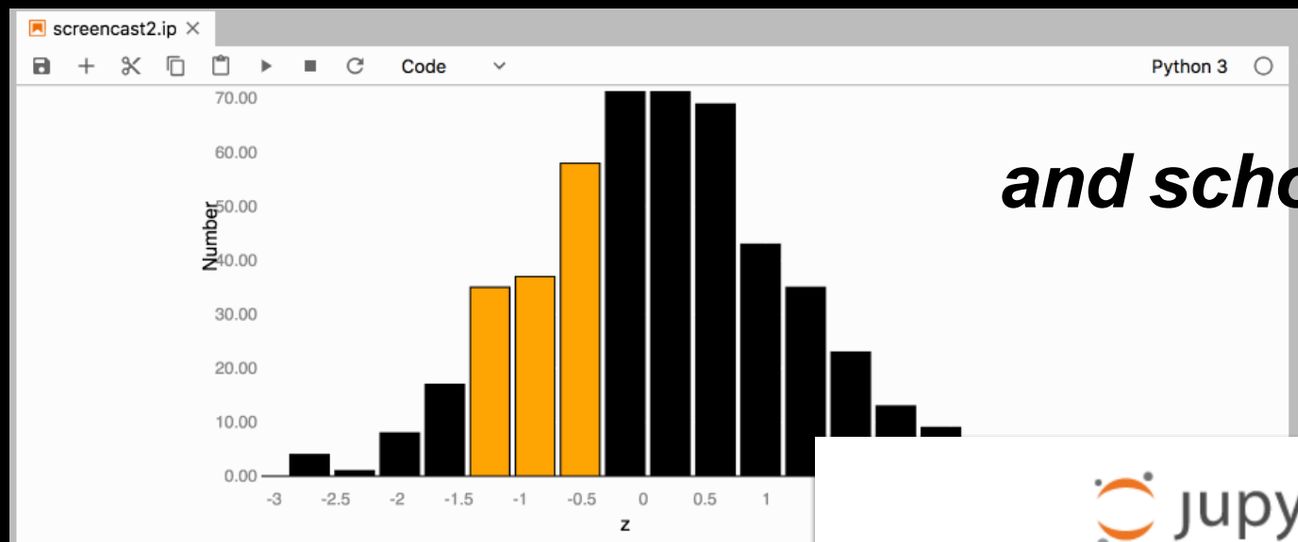
```
#!/usr/bin/env python

"""
This now uses the imshow command instead of pc
faster*
"""
from __future__ import division, print_function

import numpy as np

from matplotlib.pyplot import *
from matplotlib.collections import LineCollection
import matplotlib.cbook as cbook
# I use if 1 to break up the different regions of code visually

if 1:   # load the data
    # data are 256x256 16 bit integers
    dfile = cbook.get_sample_data('s1045.ima.gz')
    im = np.fromstring(dfile.read(), np.uint16).astype(float)
    im.shape = 256, 256

if 1:  # plot the MRI in pcolor
    subplot(221)
    imshow(im, cmap=cm.gray)
    axis('off')

if 1:  # plot the histogram of MRI intensity
    subplot(222)
    im = np.ravel(im)
    im = im[np.nonzero(im)]  # ignore the background
    im = im/(2.0**15)  # normalize
    hist(im, 100)
    xticks([-1, -.5, 0, .5, 1])
    yticks([])
    xlabel('intensity')
    ylabel('MRI density')

if 1:   # plot the EEG
    # load the data

    numSamples, numRows = 800, 4
    eegfile = cbook.get_sample_data('eeg.dat', asfileobj=False)
    print('loading eeg %s' % eegfile)
    data = np.fromstring(open(eegfile, 'rb').read(), float)
    data.shape = numSamples, numRows
    t = 10.0 * np.arange(numSamples, dtype=float)/numSamples
    ticklocs = []
    ax = subplot(212)
    xlim(0, 10)
    xticks(np.arange(10))
    dmin = data.min()
    dmax = data.max()
    dr = (dmax - dmin)*0.7  # Crowd them a bit.
    y0 = dmin
    y1 = (numRows - 1) * dr + dmax
    ylim(y0, y1)

    segs = []
    for i in range(numRows):
```

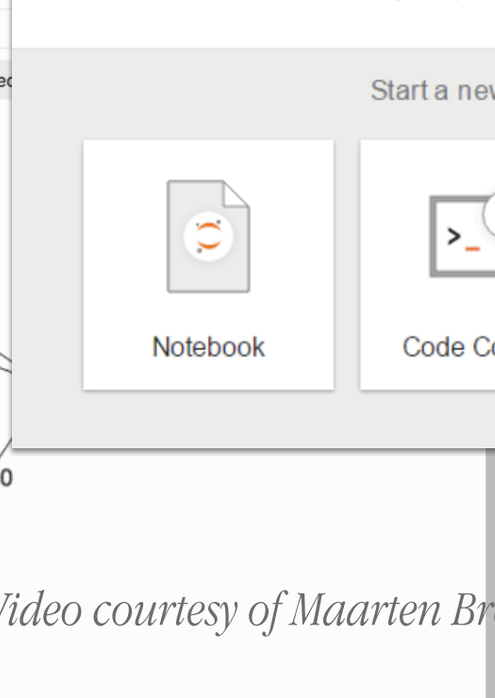
JUPYTER LABS

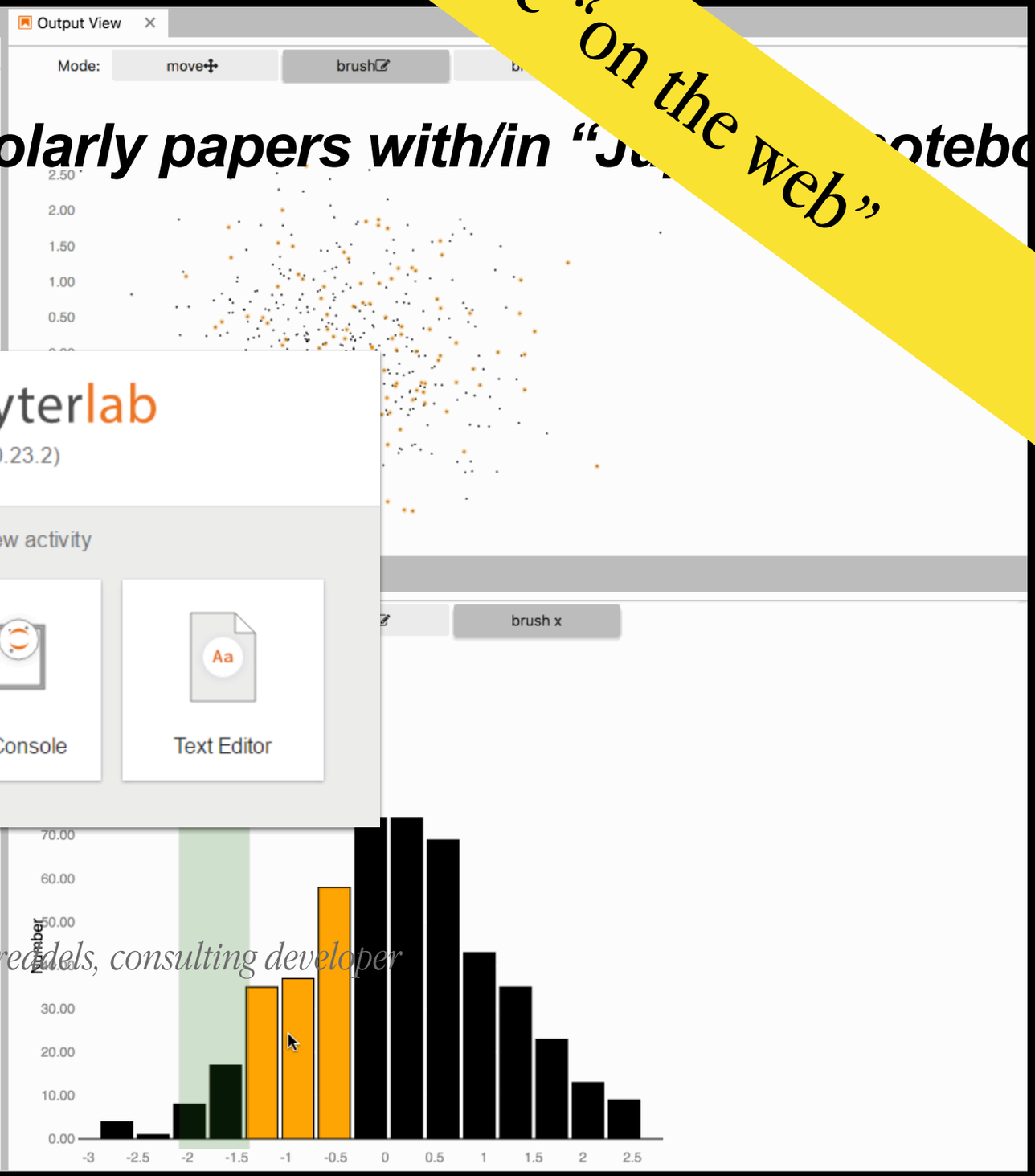
glue “on the web”

and scholarly papers with/in “Ju… notebo…

2018







Video courtesy of Maarten Breddels, consulting developer

glue *in the browser*

CODAP

web & outreach

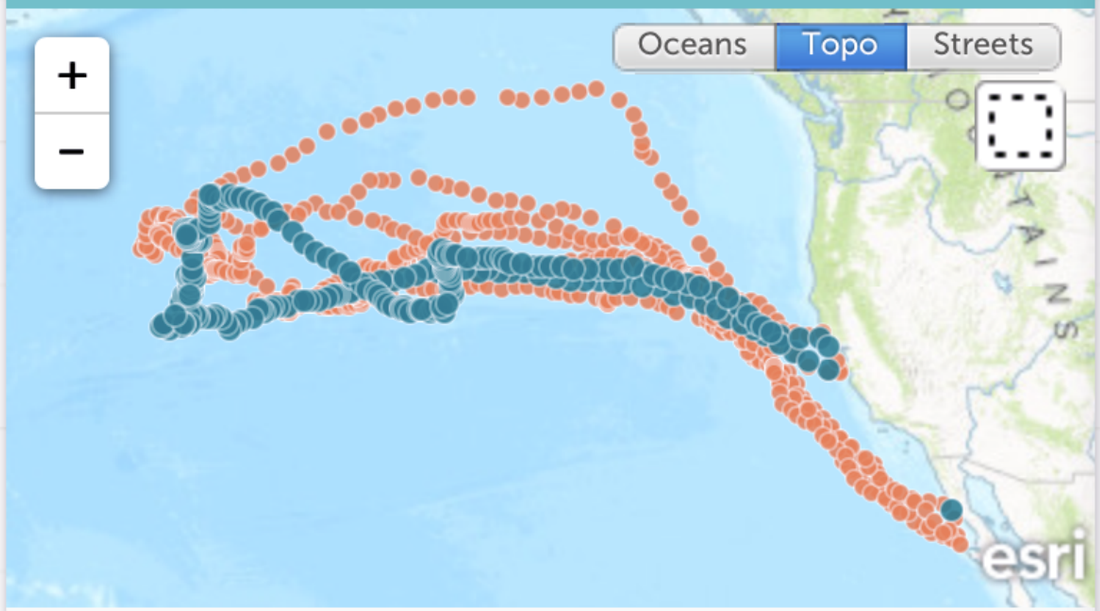
Four Seals UNSAVED Version 2.0 (0395)

Tables Graph Map Slider Calc Text Undo Redo Tiles

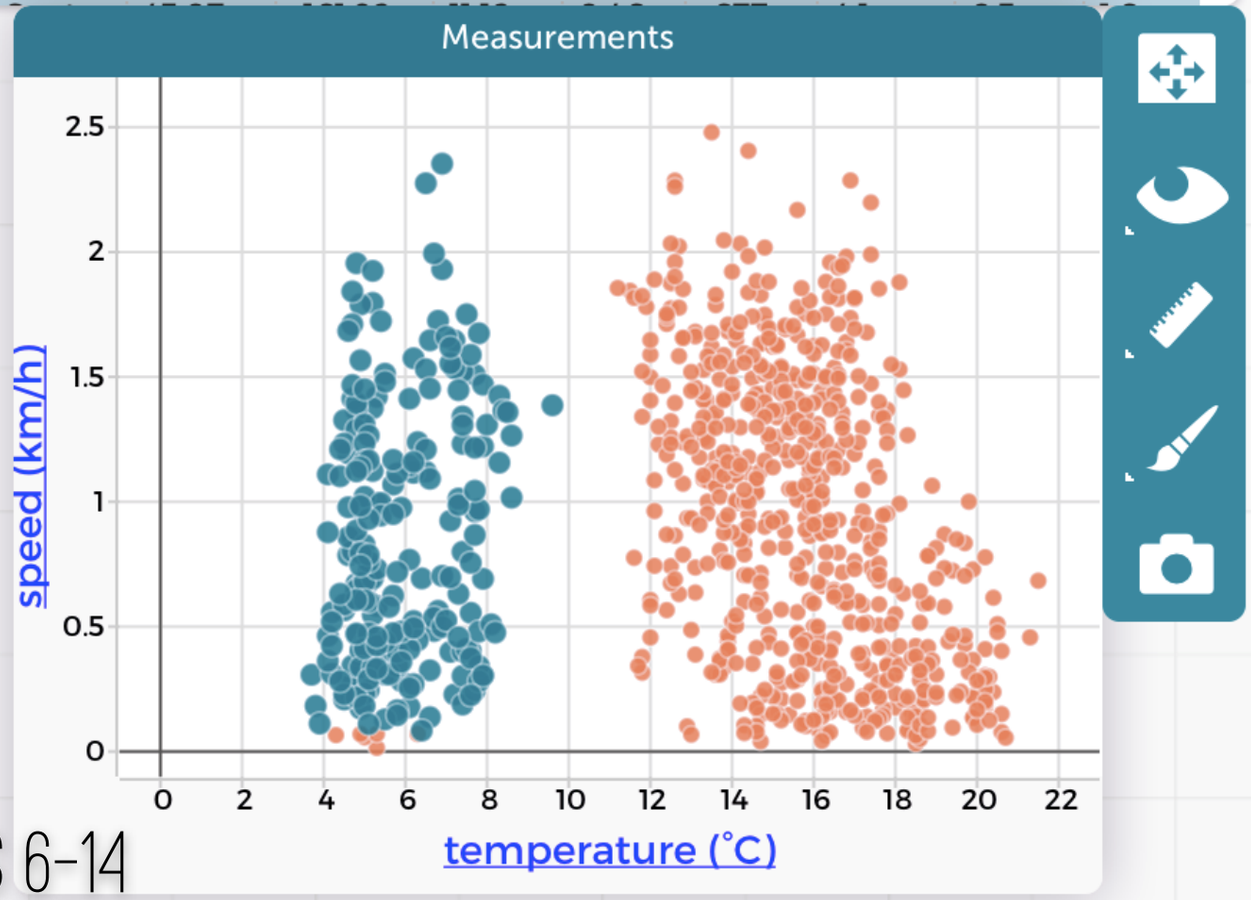
index	animal_	species
1	546	Elephant Seal
2	541	Elephant Seal
3	536	Elephant Seal
4	528	Elephant Seal

index	day	date	month	latitude	longitude	dist… e (km	spe… km/h	depth (m)	tem… ure (°C	chl… ophyl…	
100	99	8/30/20…	Augu…	41.8	-162.27	15.54	0.65	-608	4.8	0.1	1.5
101	100	8/31/20…	Augu…	42.27	-162.16	26.78	1.12	-371	6.2	0.2	1.5
102	101	9/1/2005	Sept…	42.92	-161.88	37.8	1.57	-332	6.2	0.3	1.5
103	102	9/2/2005	Sept…	43.08	-161.83	9.32	0.39	-368	5.5	0.3	1.5
104	103	9/3/2005	Sept…	43.06	-161.64	7.81	0.33	-291	6.6	0.2	1.5

Measurements



Measurements



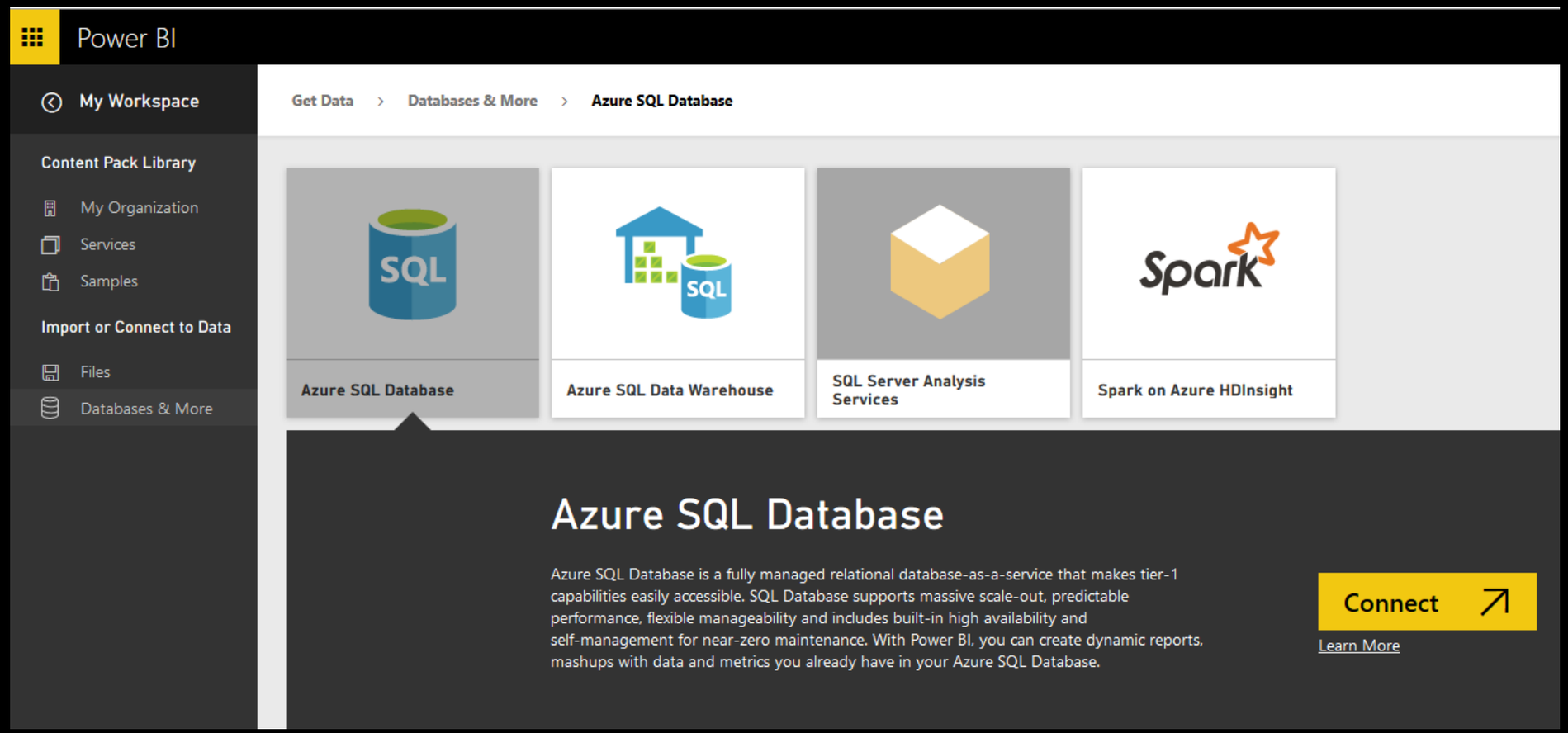
LINKED VIEWS & DATA SCIENCE FOR GRADES 6-14

Common Online Data Analysis Platform (CODAP), from Concord Consortium

REMOTE DATA ACCESS++

"BIG" Data

+data abstraction layer
+replace matplotlib with OpenGL-backed 3D viewer
+data shaders



THE CHALLENGE OF 3D SELECTION