Data Exploration

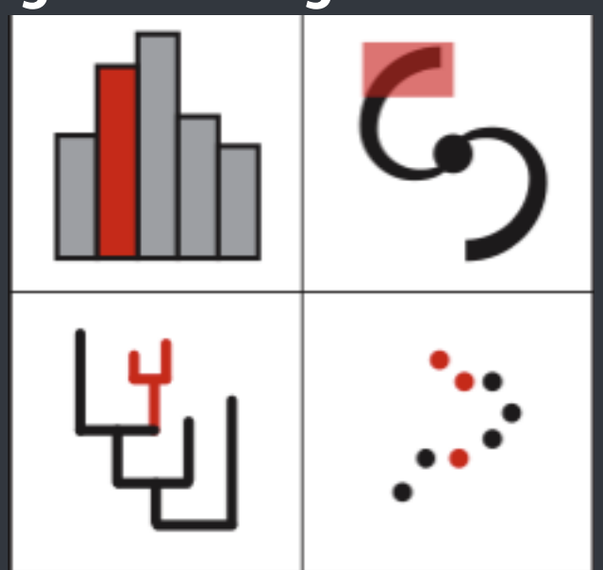
and

Hackable User Interfaces

with

Glue

glueviz.org



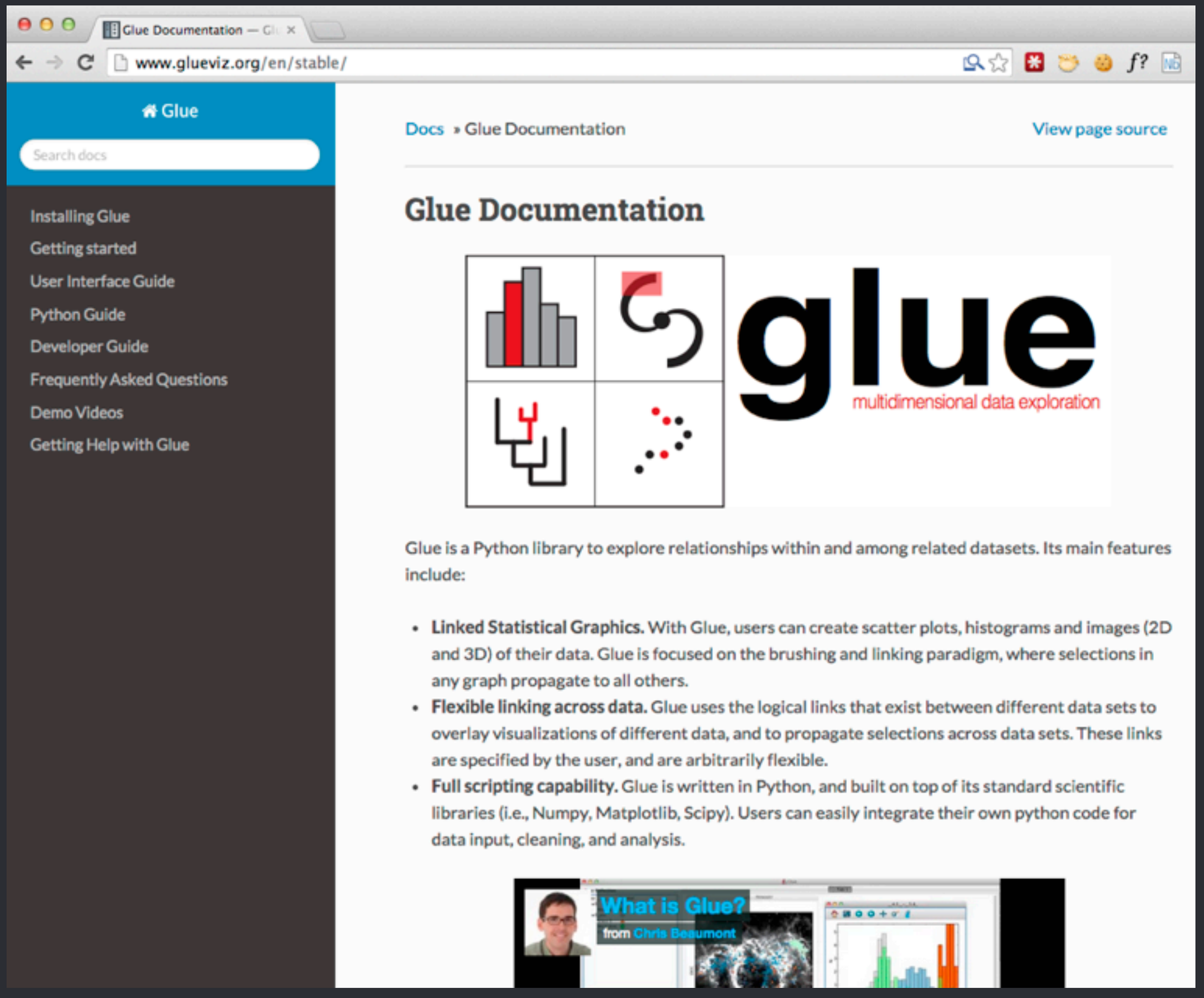
Chris Beaumont

Harvard Center for Astrophysics

Boston DataCon | September 14

@BeaumontChris

glueviz.org



Quora

Search

Home

Open

★ Statistics (academic discipline): What is the future of data analysis?

Follow Question 479 Comment 1 Share 5 •••

Michael Hochster, PhD in Statistics, Stanford; Director of Data Science, LinkedIn

176 upvotes by Justin Rising (PhD in statistics with a dissertation in probab... (more)), Don van der Drift (In PhD Physics program for 2.5 years at Technis... (more)), Alexander Blocker (Statistician at Google, PhD in statistics from ... (more)), Joe Blitzstein (Professor in the Harvard Statistics Department), Yair Livne (Econ PhD from Stanford, took 2 years of stats P... (more)), (more)

My guess, and this is just a guess of course, is that the big breakthroughs in data analysis will come neither from advances in statistical methods nor from ever-more-complex calculations on ever-larger data sets.

I think that the biggest gains will come from tools that will allow analysts and others to browse, query, summarize, filter, aggregate, disaggregate, and view their data interactively. The static or minimally interactive tables and charts that are the endpoint of most analyses I see seem very primitive, given the technology we have at our disposal today.

Much of the work I see in data visualization seems to me on the wrong path, focusing too much on making things look cool rather than on answering questions. I am with Andrew Gelman on this (see Infovis, infographics, and data visualization: Where I'm coming from, and where I'd like to go). But I still think this is the area with the greatest potential for changing the way real data analysis is done.

Quora

Search

Home

Open

★ Statistics (academic discipline): What is the future of data analysis?

Follow Question 479 Comment 1 Share 5 •••

Michael Hochster, PhD in Statistics, Stanford; Director of Data Science, LinkedIn

176 upvotes by Justin Rising (PhD in statistics with a dissertation in probab... (more)), Don van der Drift (In PhD Physics program for 2.5 years at Technis... (more)), Alexander Blocker (Statistician at Google, PhD in statistics from ... (more)), Joe Blitzstein (Professor in the Harvard Statistics Department), Yair Livne (Econ PhD from Stanford, took 2 years of stats P... (more)), (more)

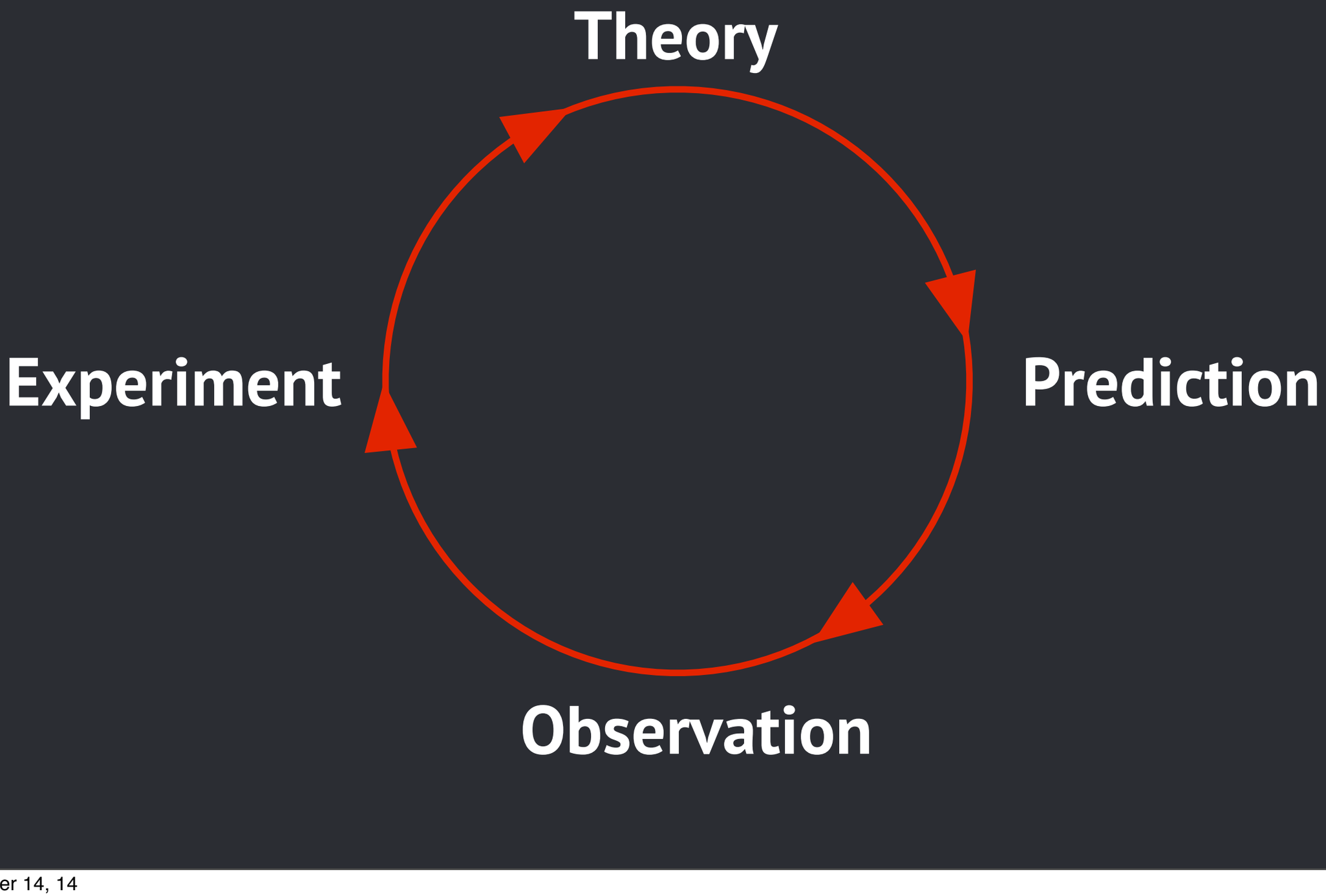
My guess, and this is just a guess of course, is that the big breakthroughs in data analysis will come neither from advances in statistical methods nor from ever-more-complex calculations on ever-larger data sets.

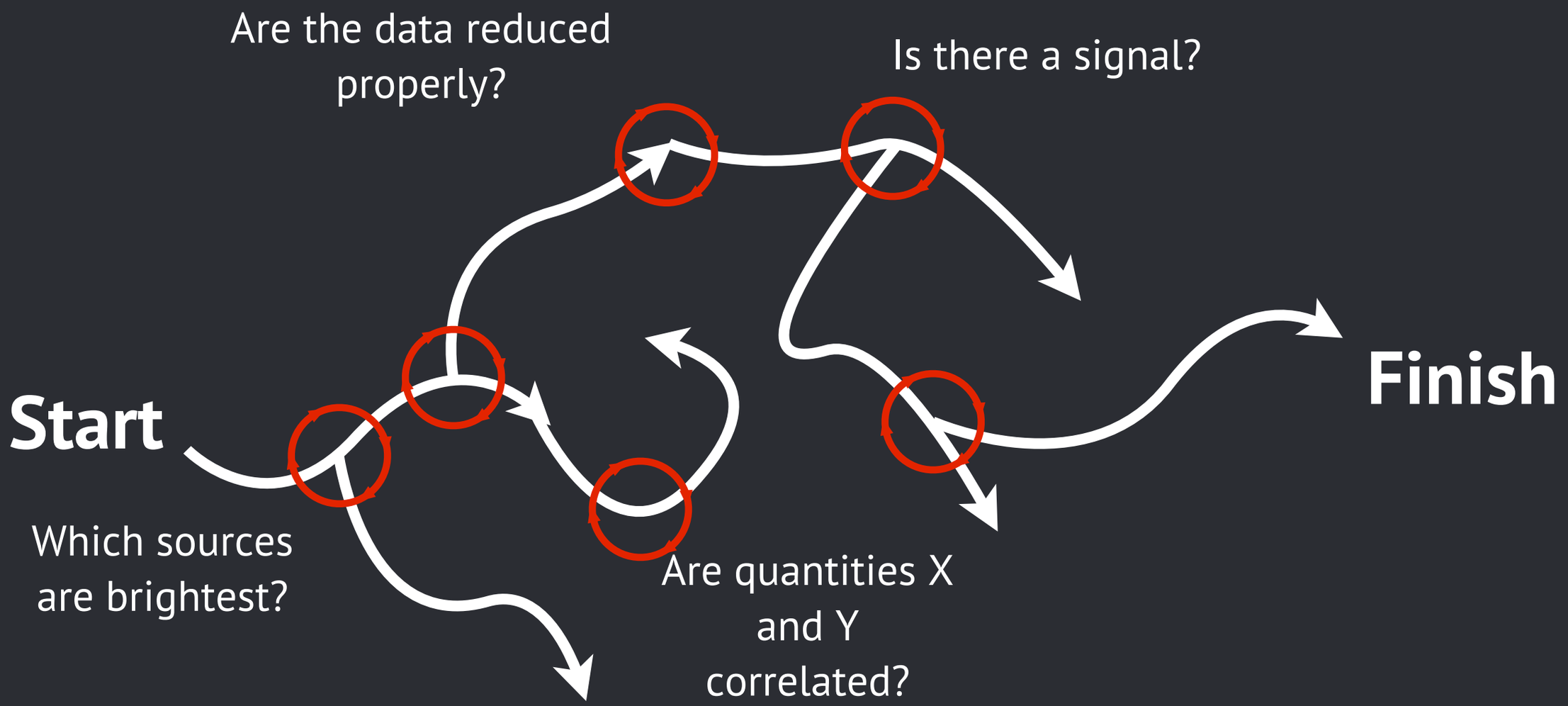
I think that the biggest gains will come from tools that will allow analysts and others to browse, query, summarize, filter, aggregate, disaggregate, and view their data interactively.

Much of the work I see in data visualization seems to me on the wrong path, focusing too much on making things look cool rather than on answering questions. I am with Andrew Gelman on this (see Infovis, infographics, and data visualization: Where I'm coming from, and where I'd like to go). But I still think this is the area with the greatest potential for changing the way real data analysis is done.

The Scientific Method*

*not true




Are the data reduced properly?
Is there a signal?
Start
Finish
Which sources are brightest?
Are quantities X and Y correlated?

Data Interfaces

Code

GUI

Data Interfaces

Code		GUI
Expressive		
Reproducible		
Precise		
Scalable		
Slow		
Hard		

Data Interfaces

Code	GUI
Expressive	Intuitive
Reproducible	Nonsequential
Precise	Fast
Scalable	
Slow	Canned
Hard	Imprecise

Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard

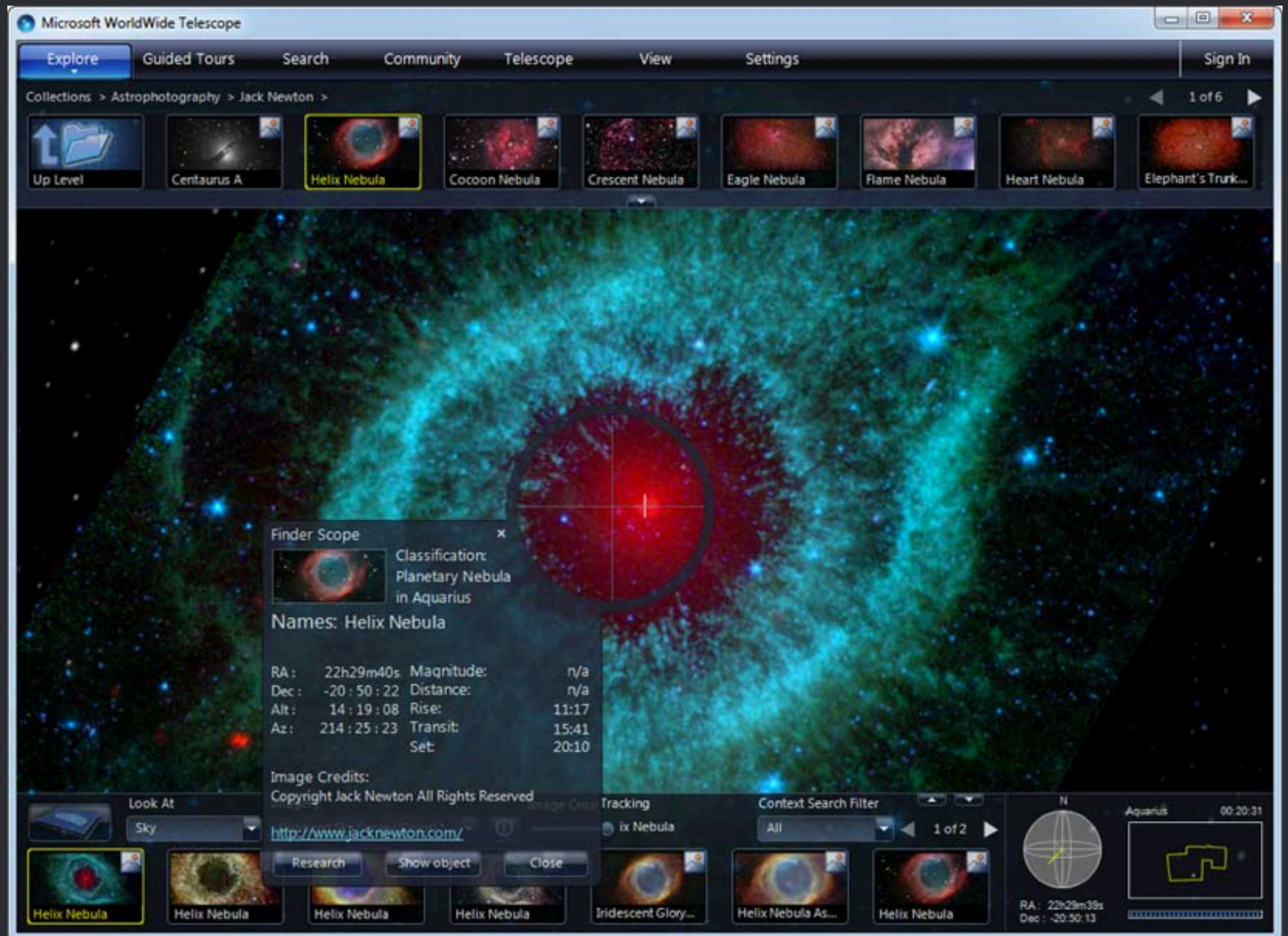
Valley of Sorrow

GUI

Intuitive
Nonsequential
Fast

Canned
Imprecise

Inter-file comparison



Stuff I Wanted

A Data Exploration GUI

That could build linked views

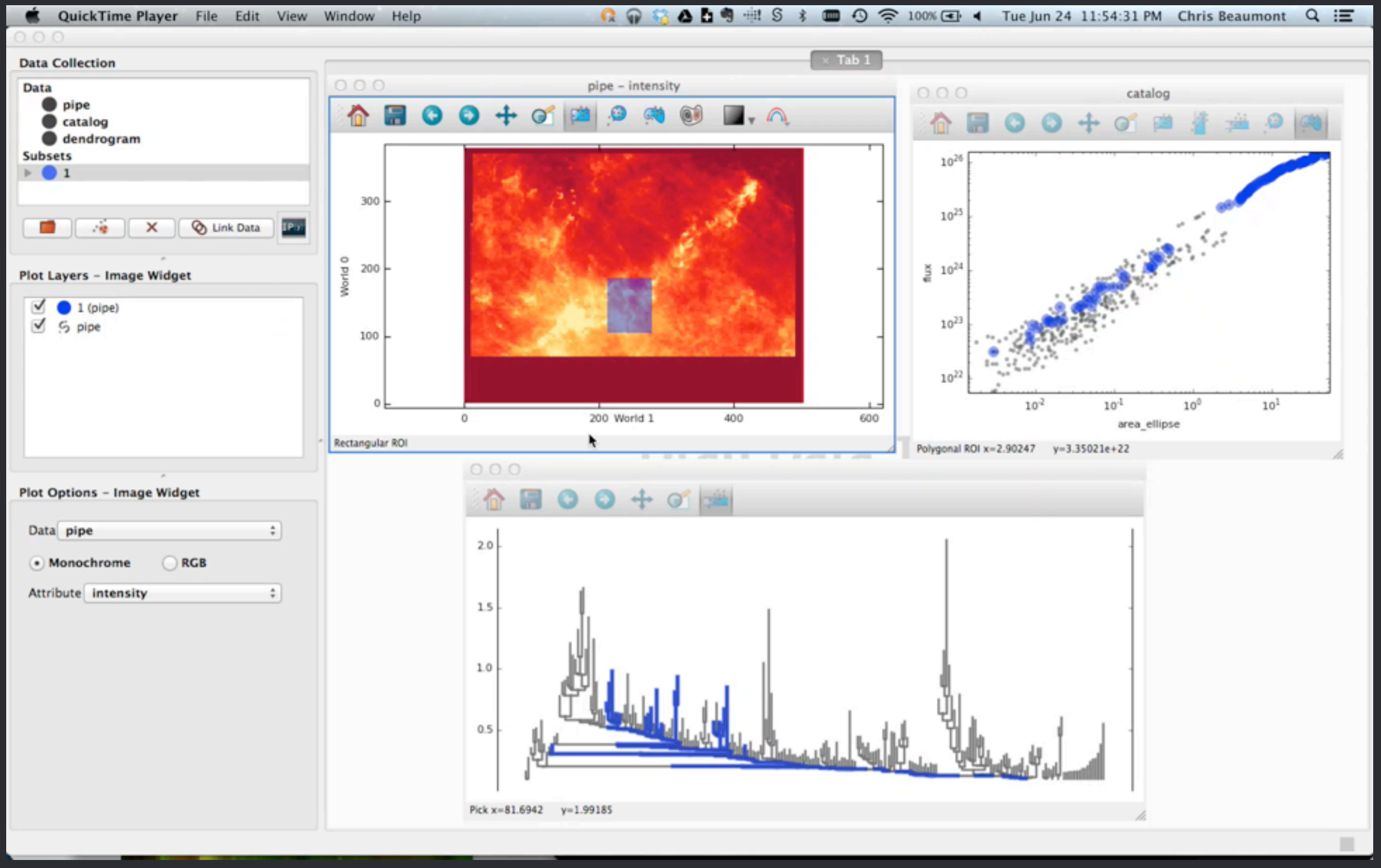
That could compare different datasets

That understood images

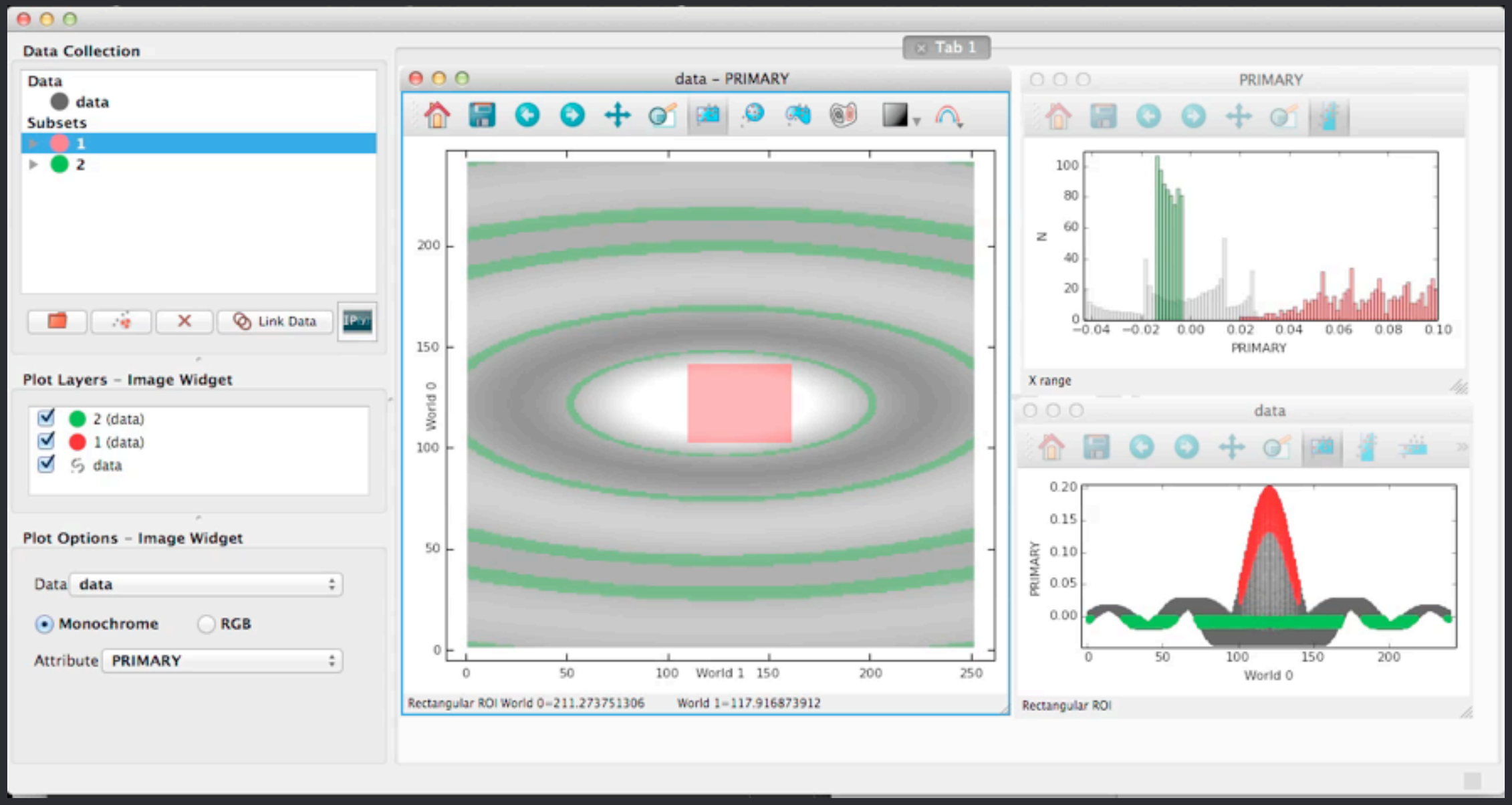
That was hackable

Live Demo

Nonspatial Linking



Automatic File Watching



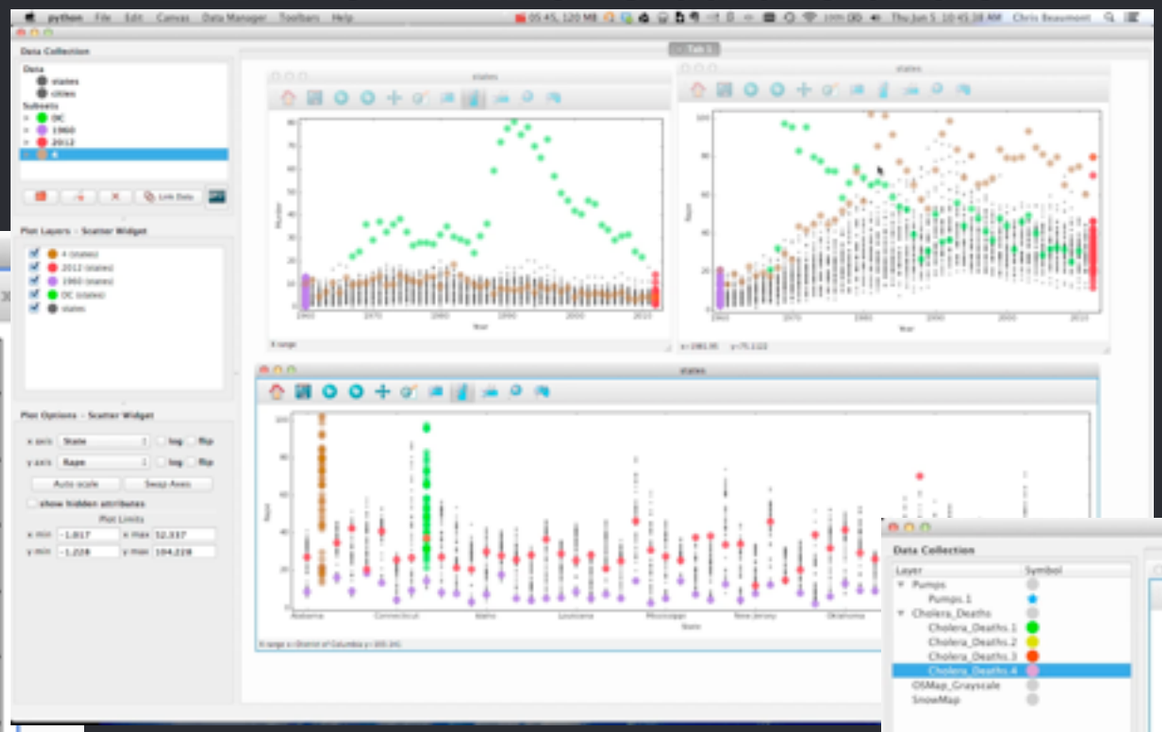
IPython Integration

Lots of data formats

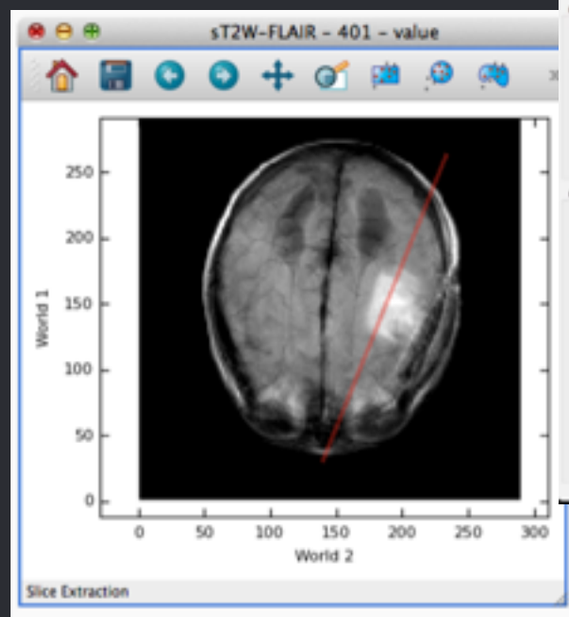
(and you can make your own loaders)

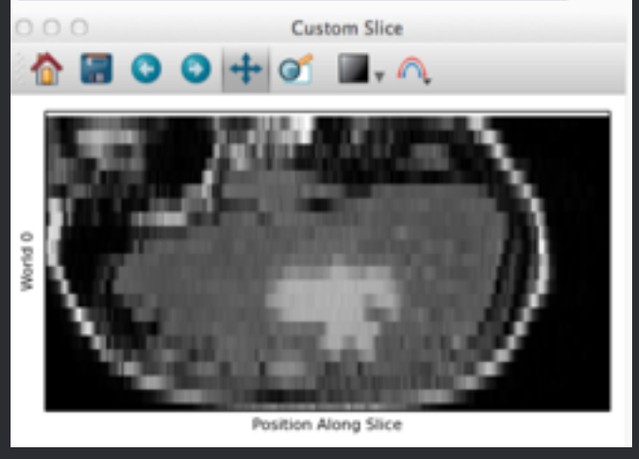
(or you can use python variables)

FBI Crime Statistics

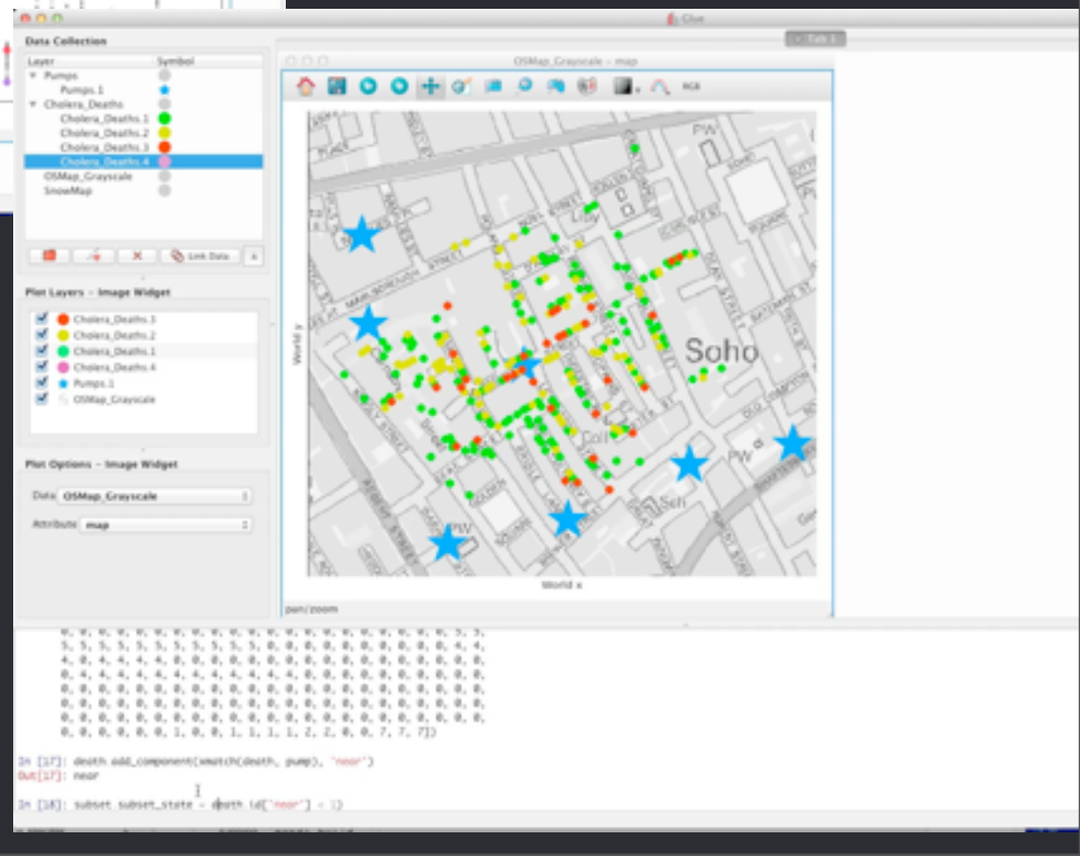


MRI Cubes





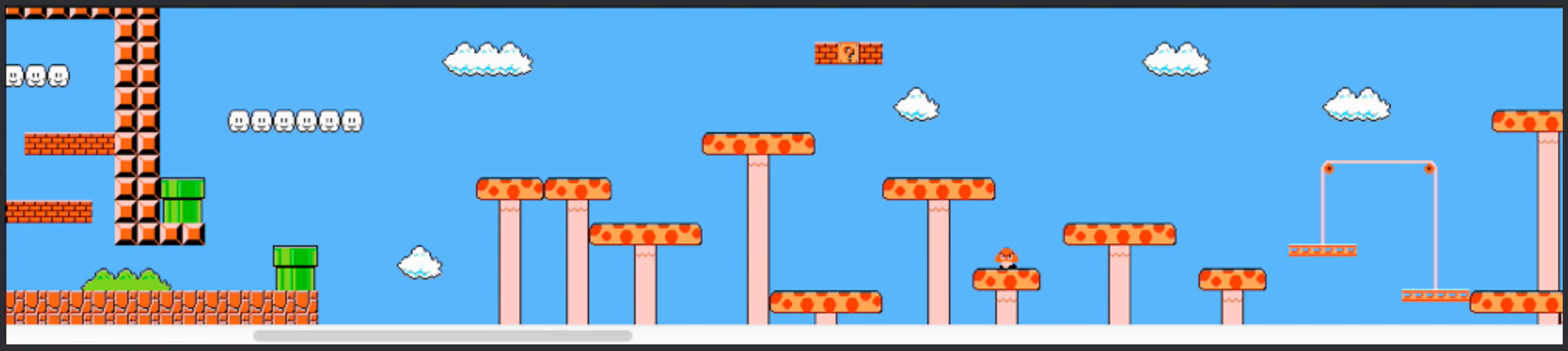
Map Data



User-Defined Viewers

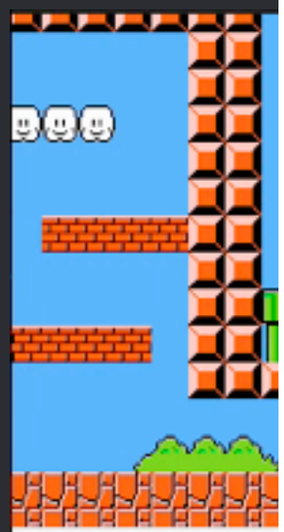
Maurice Guégan | http://stabyourself.net/mari0/

User-Defined Viewers



Maurice Guégan | http://stabyourself.net/mari0/

User-Defined Viewers



Maurice Guégan · @stabyourself.net> Aug 14

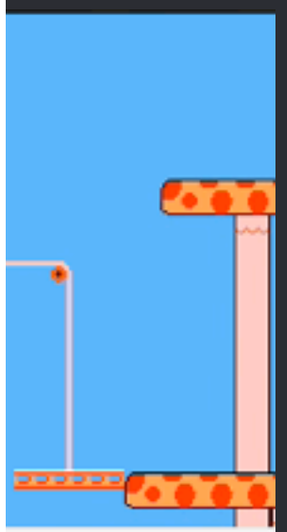
to Chris

Heya,

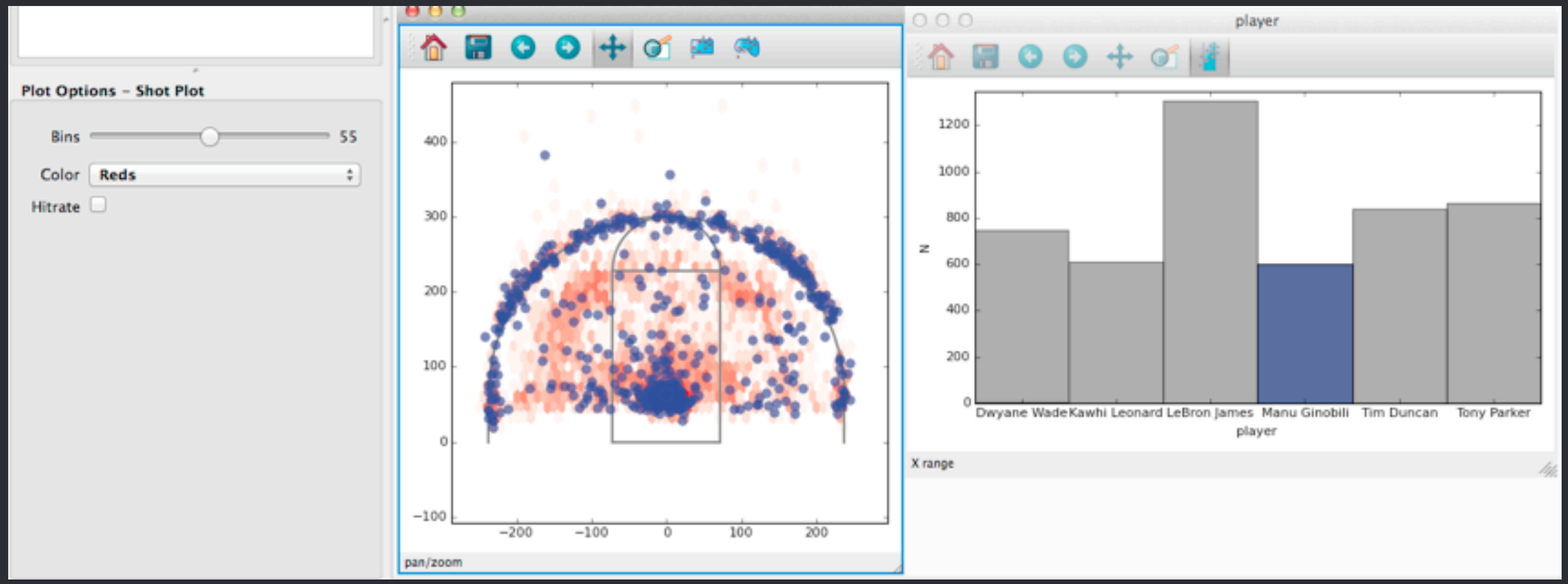
sure! Here you go:

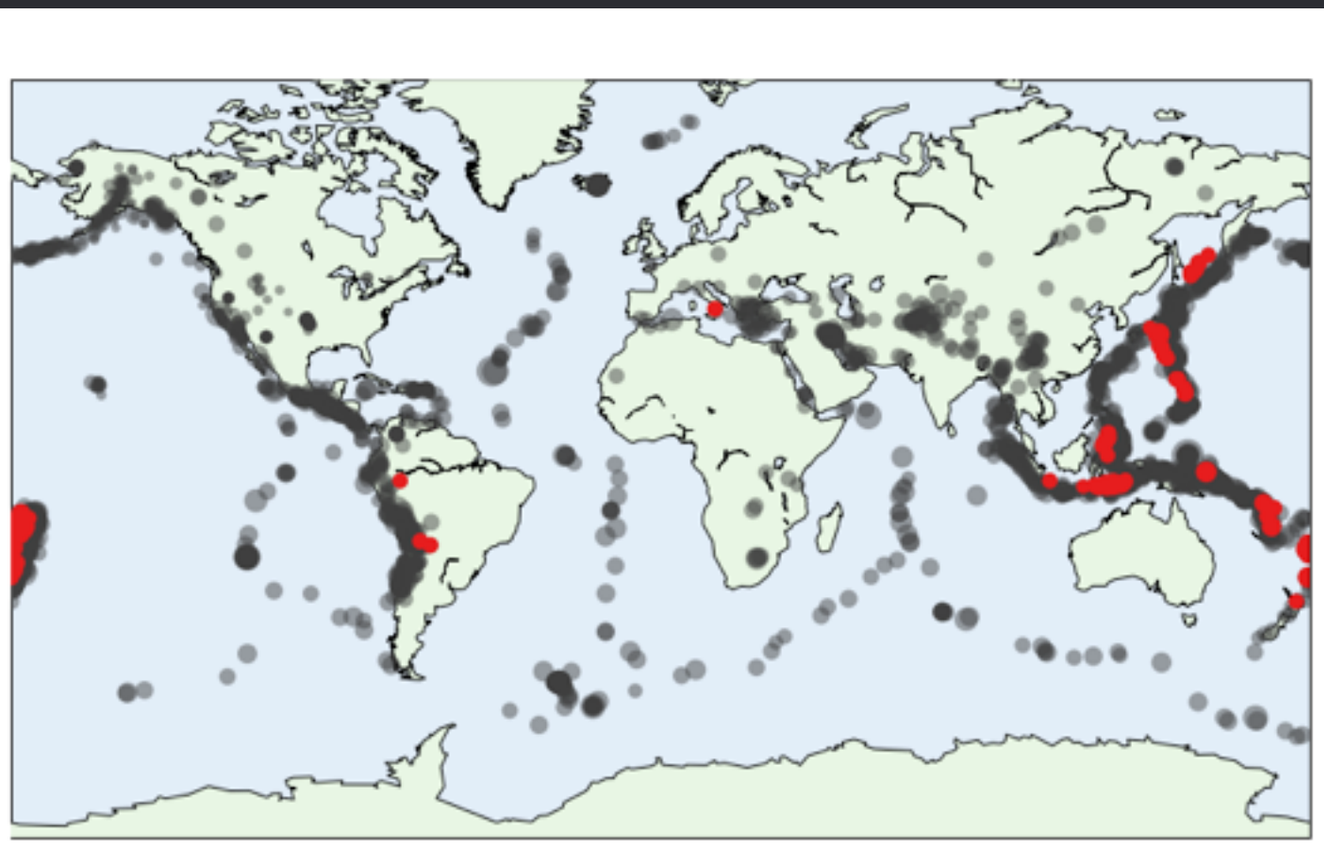
http://guegan.de/stuff/Mario_Time_Trials_Replays.zip

Some information:
This was free to play for anyone at Gamescom (This is what it looked like: http://www.flickr.com/photos/stabyourself/9594326986/in/set-72157635161214148)
People could see 200 previous replays as grey ghost marios while they played
There was a timer (counting up) at the bottom of the screen
The game showed this screen before starting http://i.imgur.com/bs9ZaHt.png



Maurice Guégan | http://stabyourself.net/mari0/

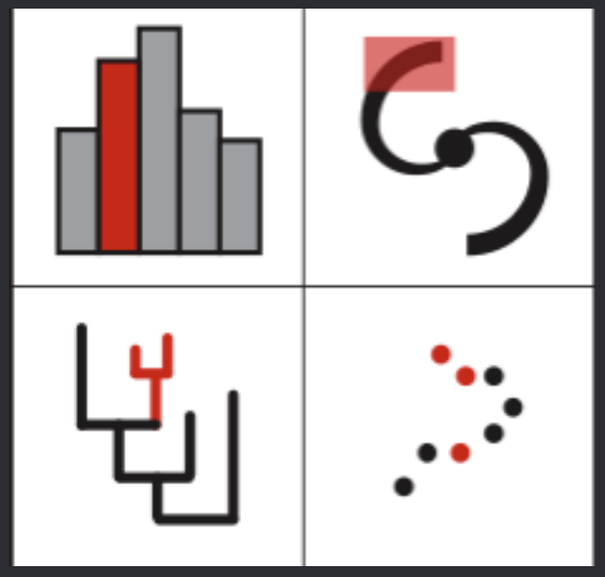

Plot Options – Shot Plot
Bins
55
Color
Reds
Hitrate
400
300
200
100
0
−100
−200
−100
0
100
200
pan/zoom
player
1200
1000
800
600
400
200
0
N
Dwyane Wade
Kawhi Leonard
LeBron James
Manu Ginobili
Tim Duncan
Tony Parker
player
X range



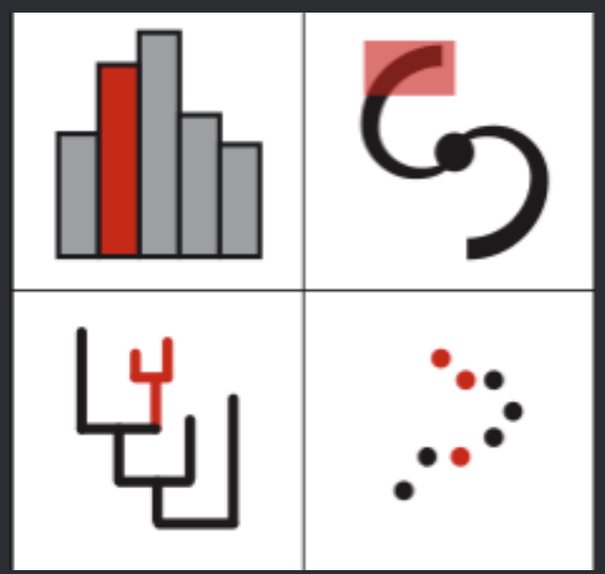
Data Interfaces

Code

GUI

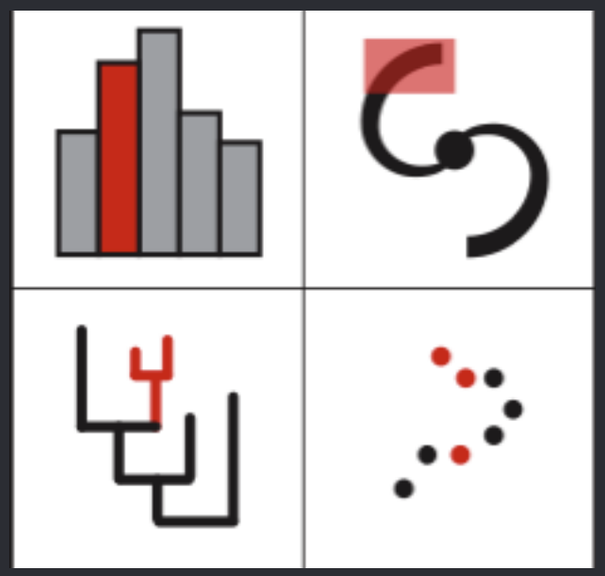


Data Interfaces

Code		GUI
Expressive Reproducible Precise Scalable Slow Hard		

Data Interfaces

Code	GUI
Expressive	Intuitive
Reproducible	Nonsequential
Precise	Fast
Scalable	
Slow	Canned
Hard	Imprecise



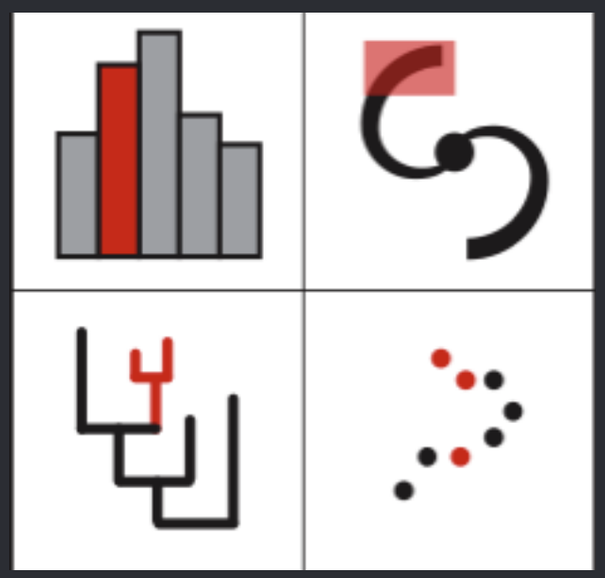
Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard

Hackable User Interfaces



GUI

Intuitive
Nonsequential
Fast

Canned
Imprecise

Thanks!

glueviz.org
@BeaumontChris

Sunday, September 14, 14