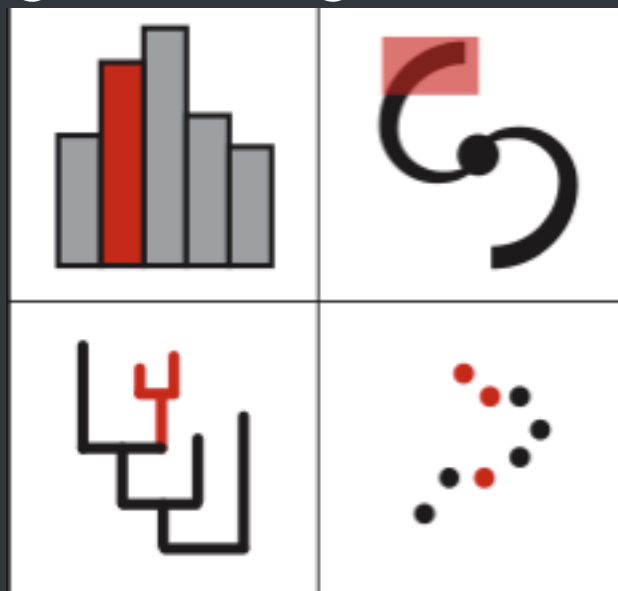


Data Exploration and Hackable User Interfaces with Glue

glueviz.org



Chris Beaumont

Harvard Center for Astrophysics
Boston DataCon | September 14

The screenshot shows a web browser window with the URL `www.glueviz.org/en/stable/`. The page has a blue header with the Glue logo and a search bar. A dark sidebar on the left contains a list of navigation links: "Installing Glue", "Getting started", "User Interface Guide", "Python Guide", "Developer Guide", "Frequently Asked Questions", "Demo Videos", and "Getting Help with Glue". The main content area is titled "Glue Documentation" and features a large graphic with the word "glue" in a bold, lowercase font, with the tagline "multidimensional data exploration" in red below it. To the left of the word "glue" is a 2x2 grid of icons: a bar chart, a circular arrow, a tree diagram, and a scatter plot. Below the graphic, a paragraph states: "Glue is a Python library to explore relationships within and among related datasets. Its main features include:". This is followed by a bulleted list of three features: "Linked Statistical Graphics", "Flexible linking across data", and "Full scripting capability". At the bottom of the page, there is a video player showing a thumbnail of a man with glasses and the text "What is Glue? from Chris Beaumont".

Glue Documentation

glue
multidimensional data exploration

Glue is a Python library to explore relationships within and among related datasets. Its main features include:

- **Linked Statistical Graphics.** 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.

What is Glue?
from Chris Beaumont

★ 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.

★ 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)

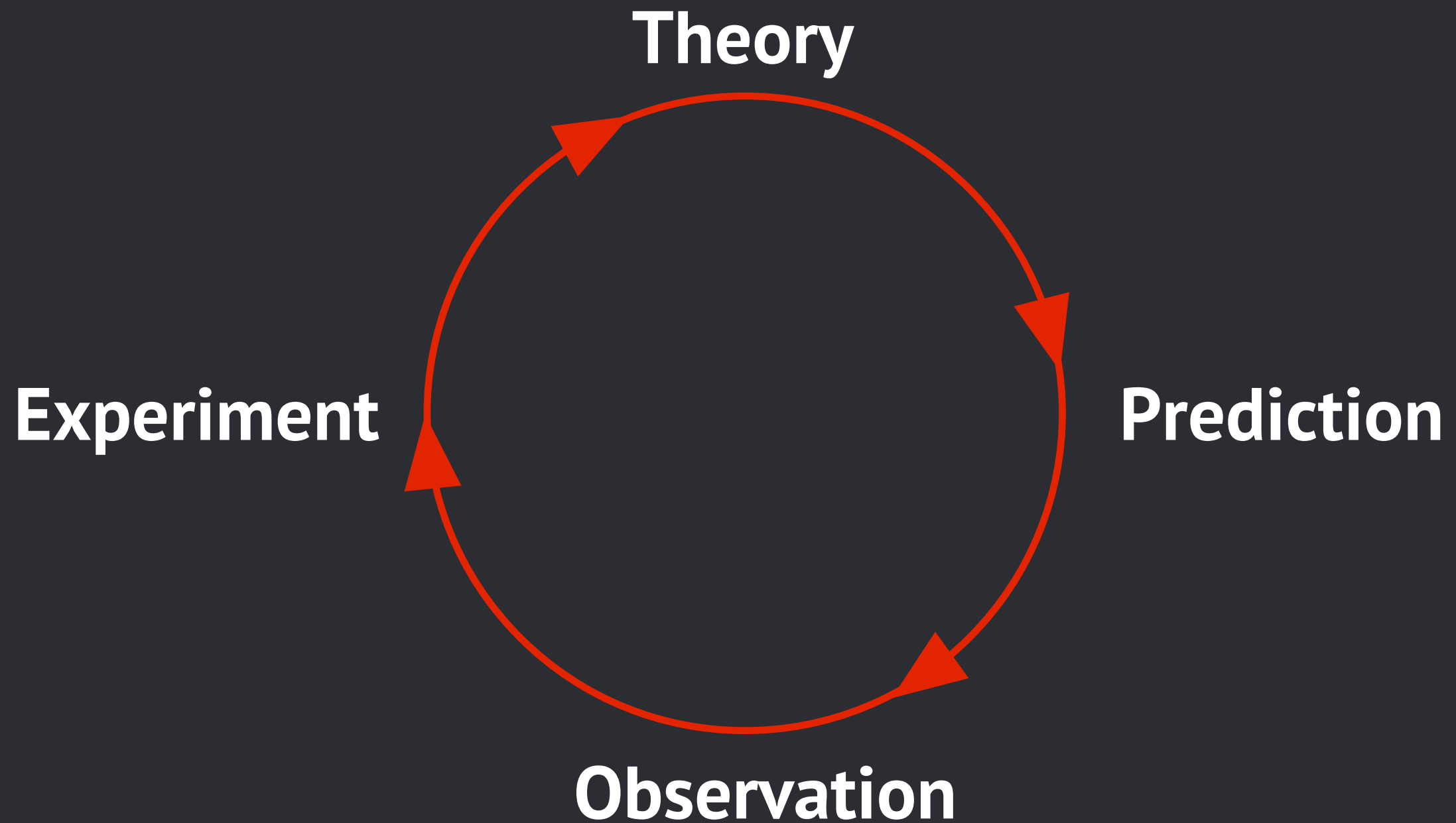
My guess, and this is just a guess of course, is that the big breakthroughs in data analysis will come neither from advances in statistical methods nor from ever-more-complex calculations on ever-larger data sets.

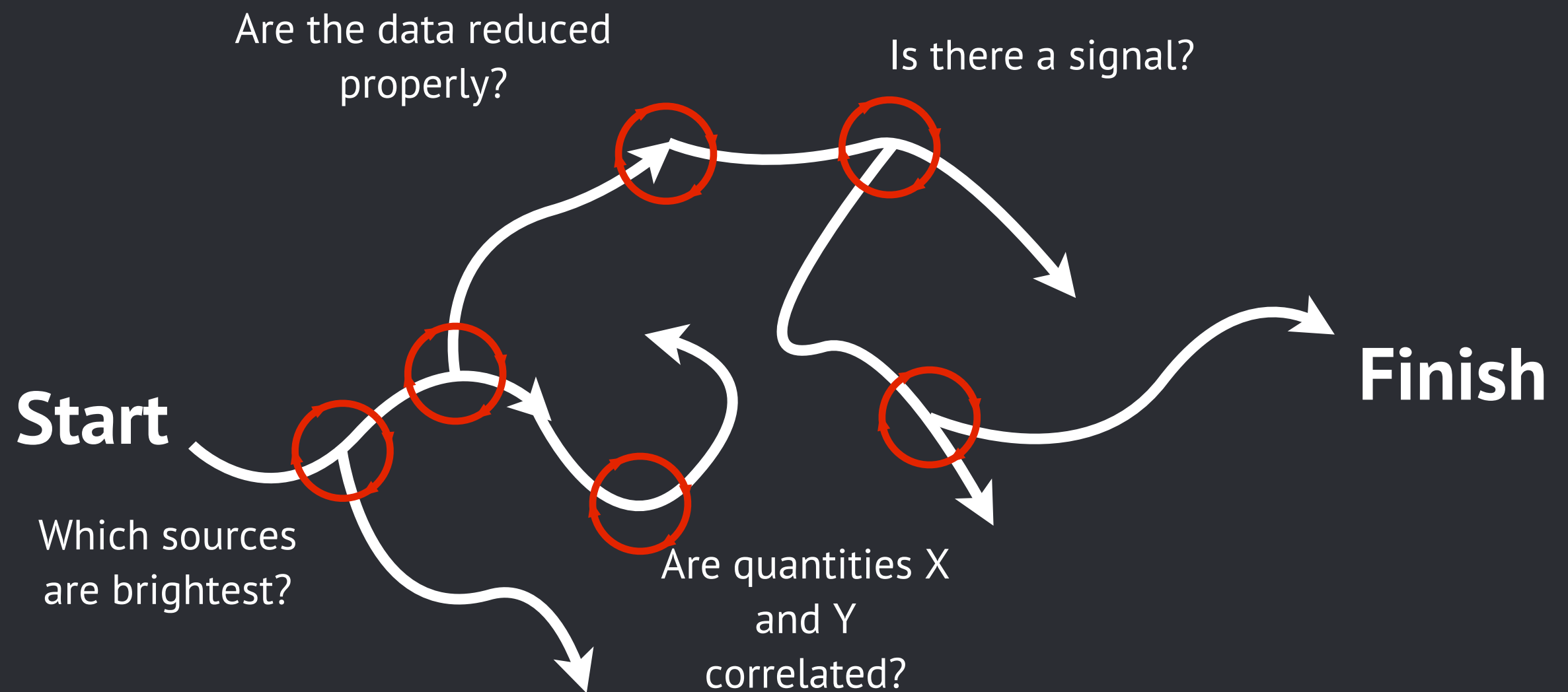
I think that the biggest gains will come from tools that will allow analysts and others to browse, query, summarize, filter, aggregate, disaggregate, and view their data interactively.

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





Data Interfaces

Code

GUI

Data Interfaces

Code

GUI

Expressive
Reproducible
Precise
Scalable

Slow
Hard

Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard

GUI

Intuitive
Nonsequential
Fast

Canned
Imprecise

Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard



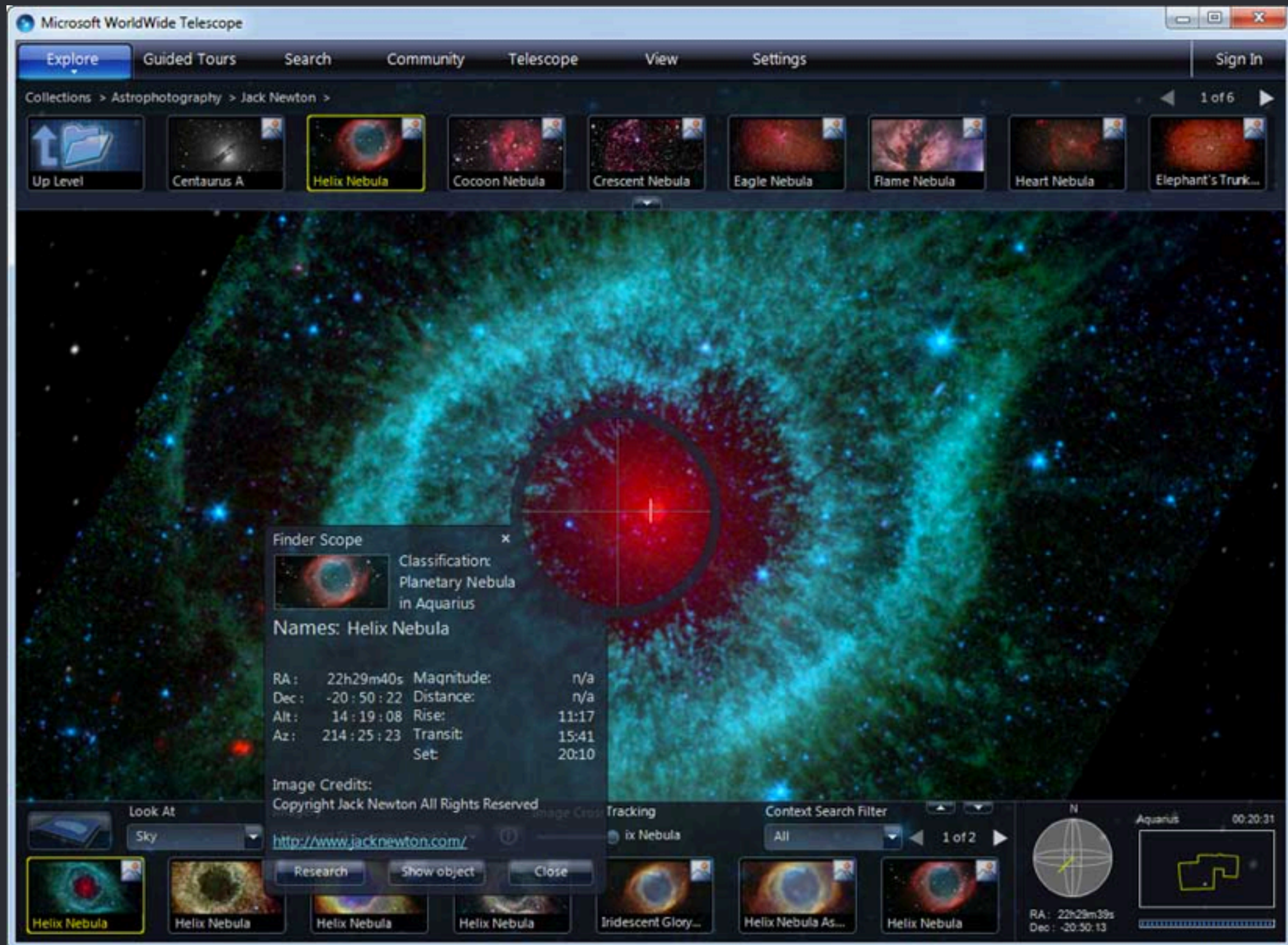
Valley of Sorrow

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

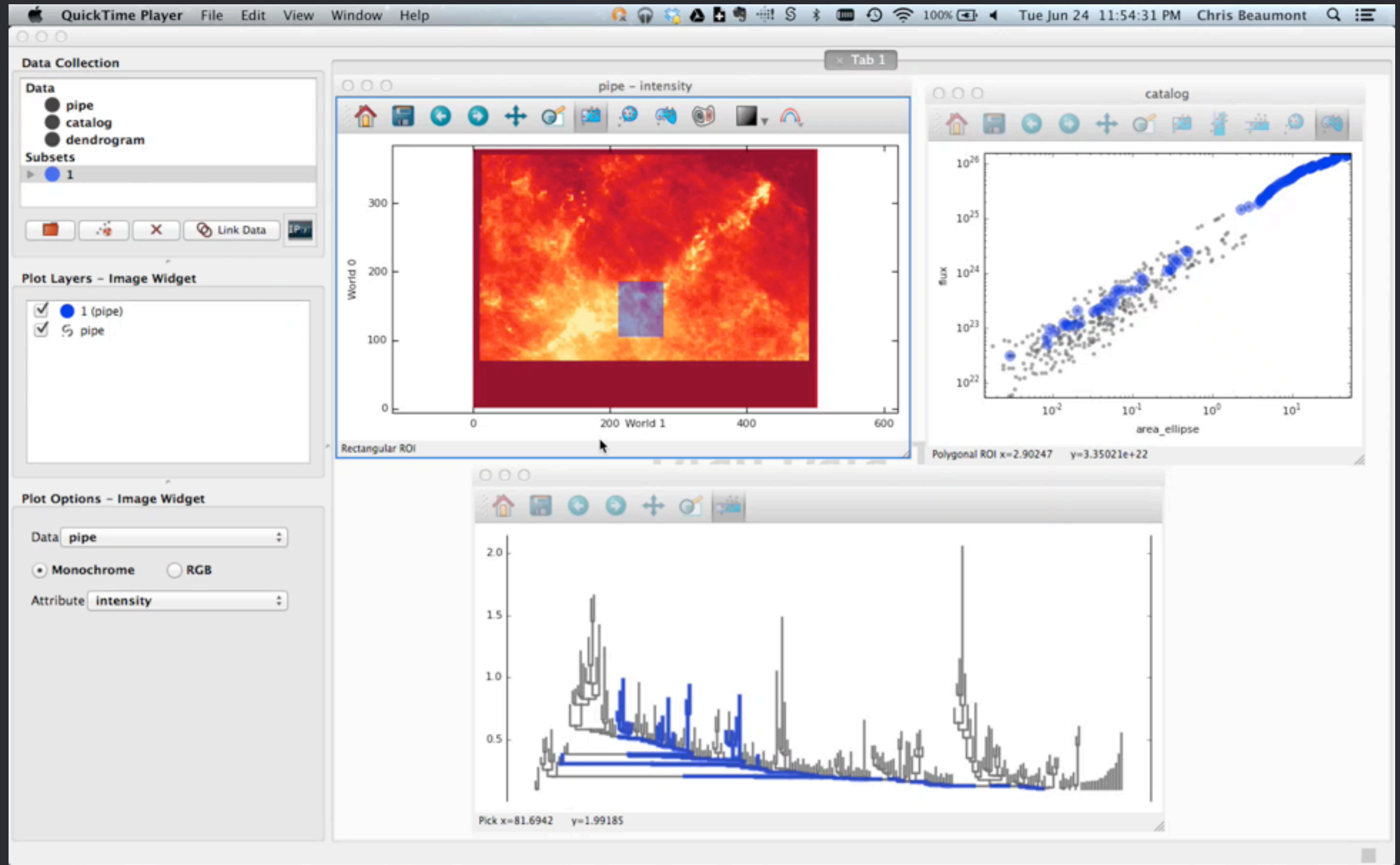
That understood images

That was hackable

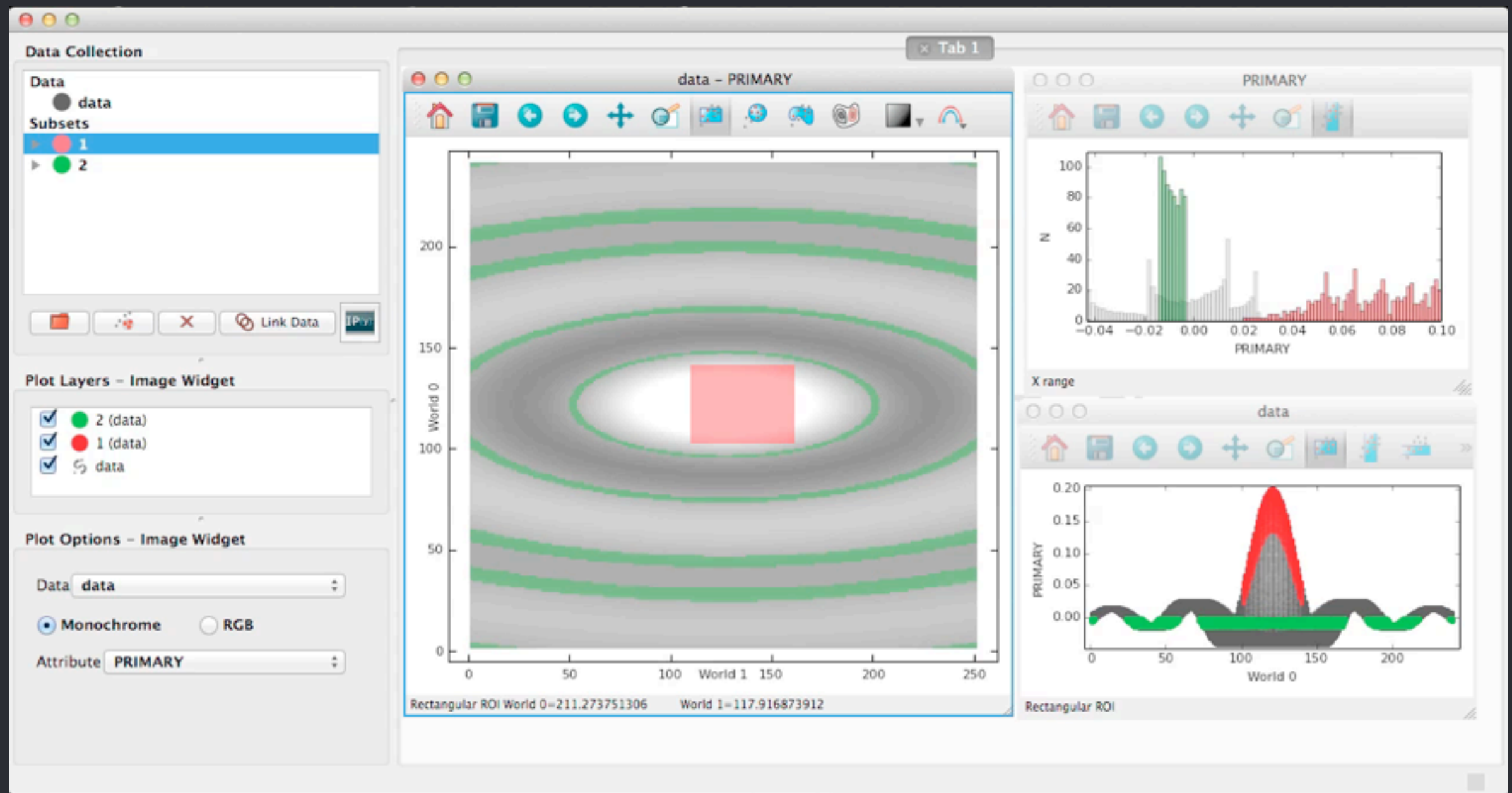


Live Demo

Nonspatial Linking



Automatic File Watching



IPython Integration

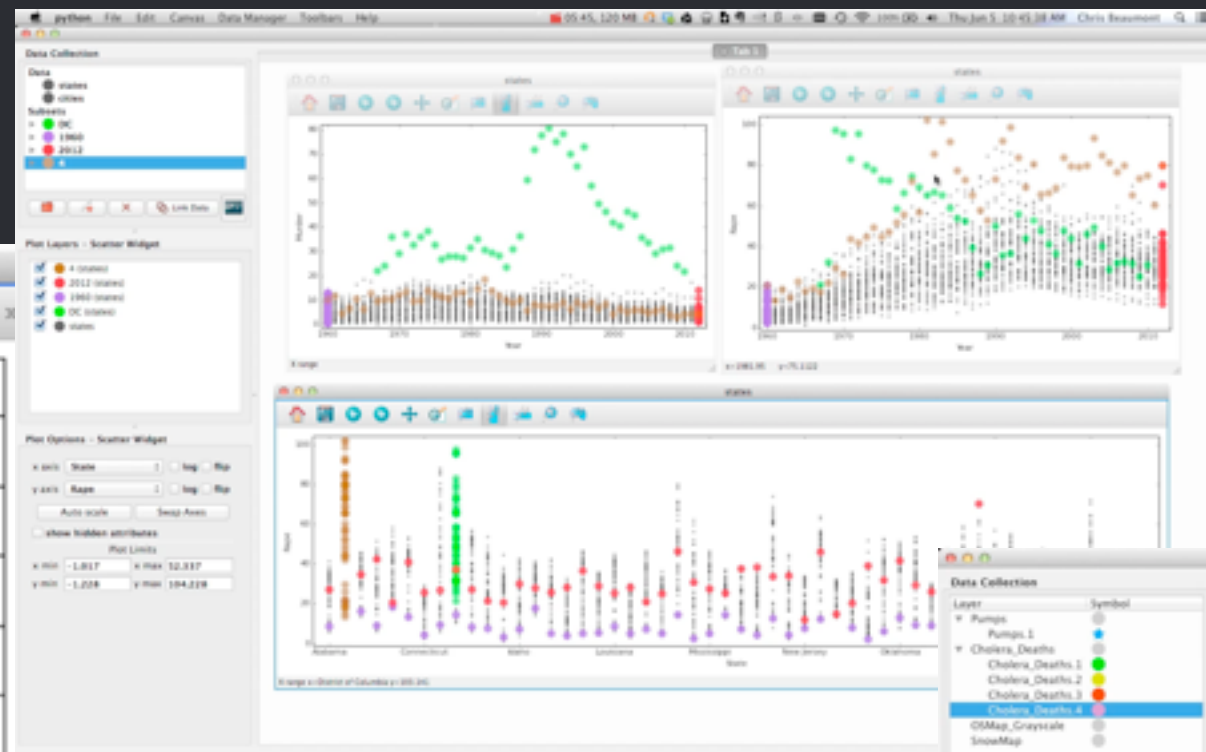
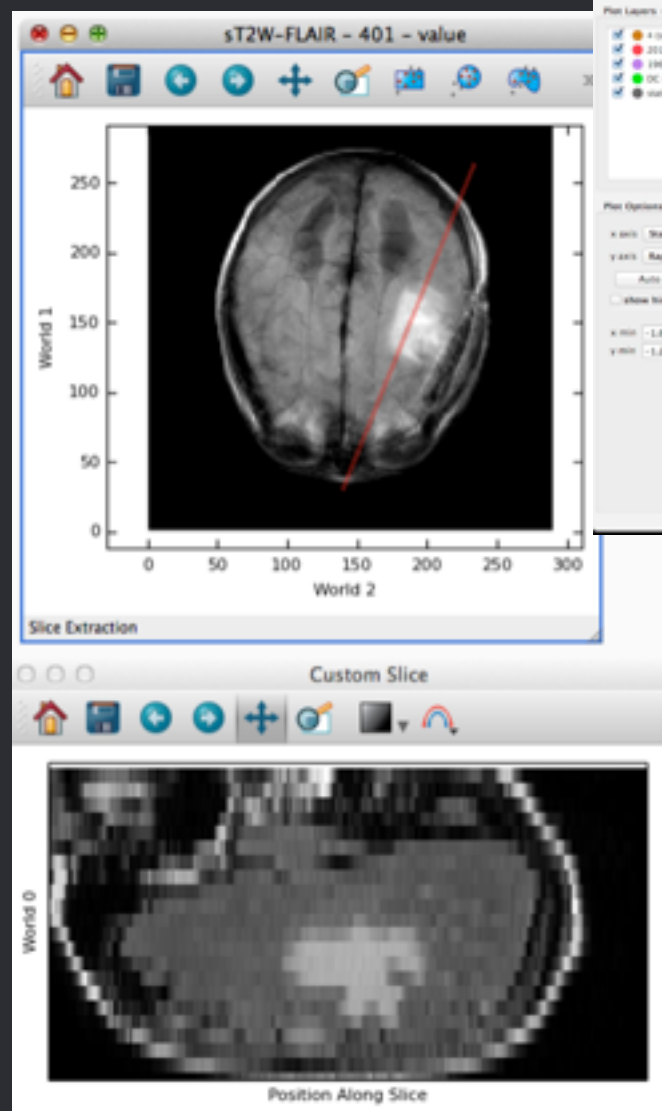
Lots of data formats

(and you can make your own loaders)

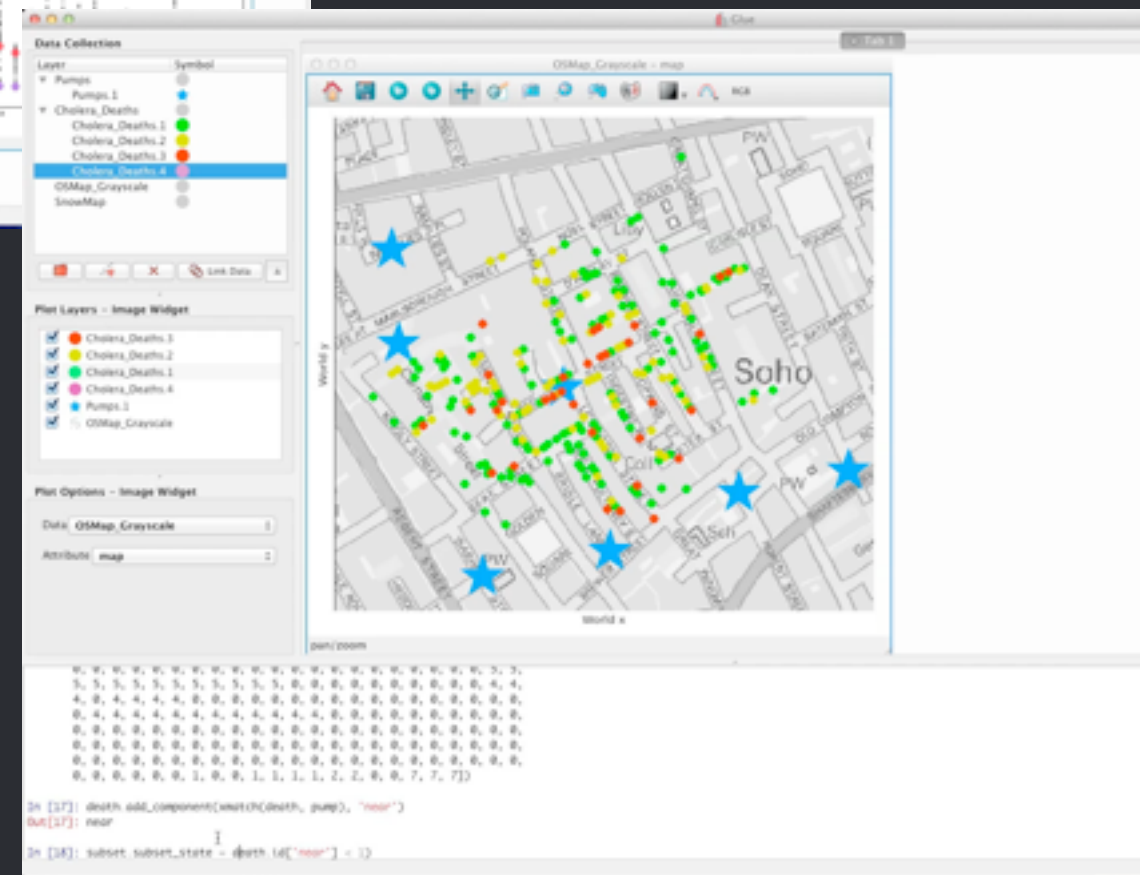
(or you can use python variables)

FBI Crime Statistics

MRI Cubes



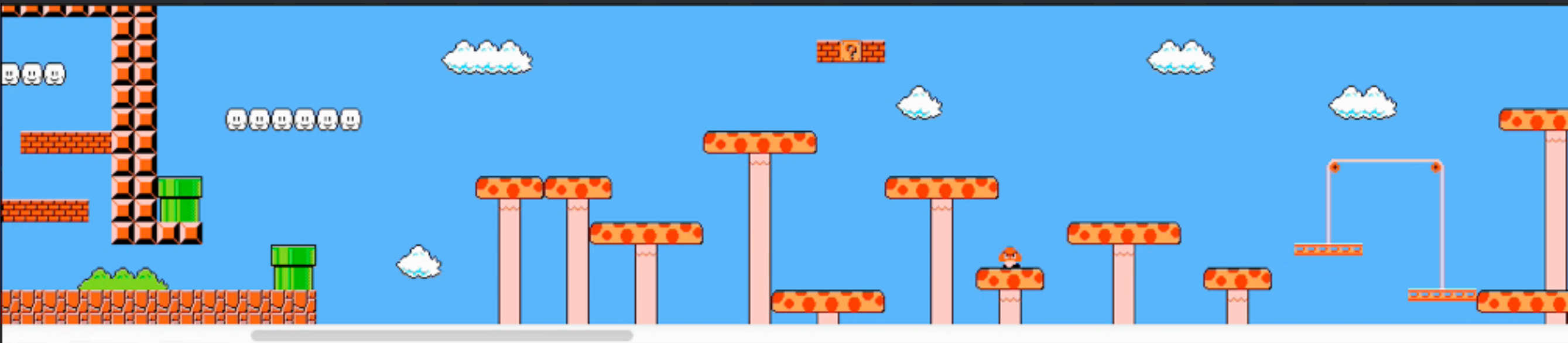
Map Data



User-Defined Viewers

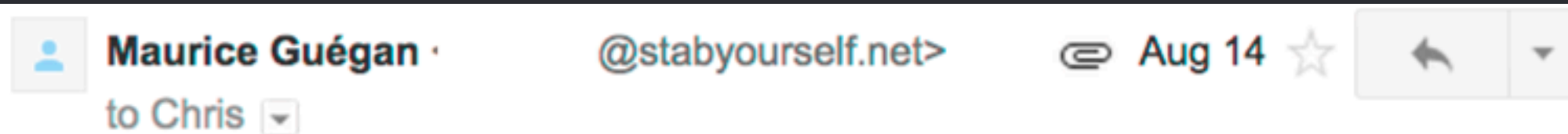
Maurice Guégan | <http://stabyourself.net/mari0/>

User-Defined Viewers



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

User-Defined Viewers



Heya,

sure! Here you go:

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

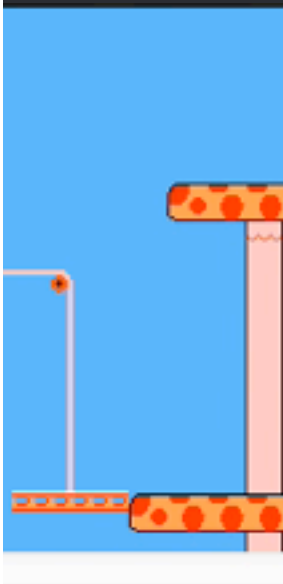
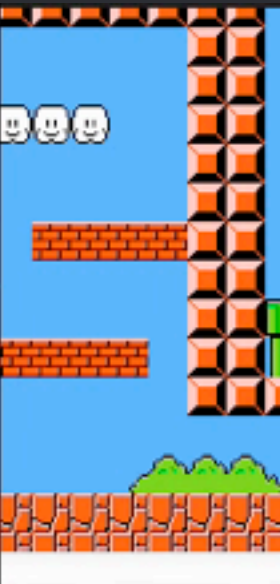
Some information:

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>)

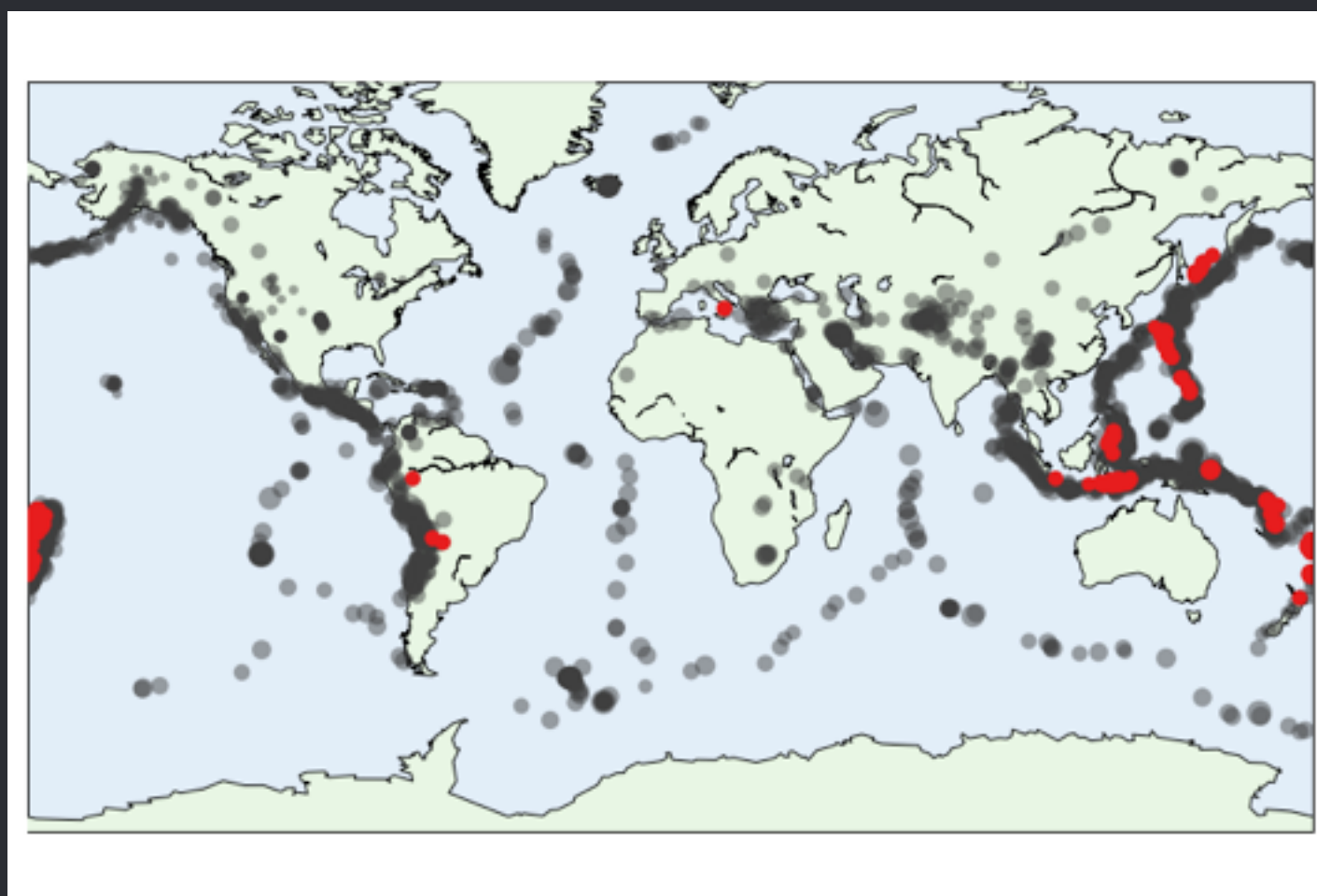
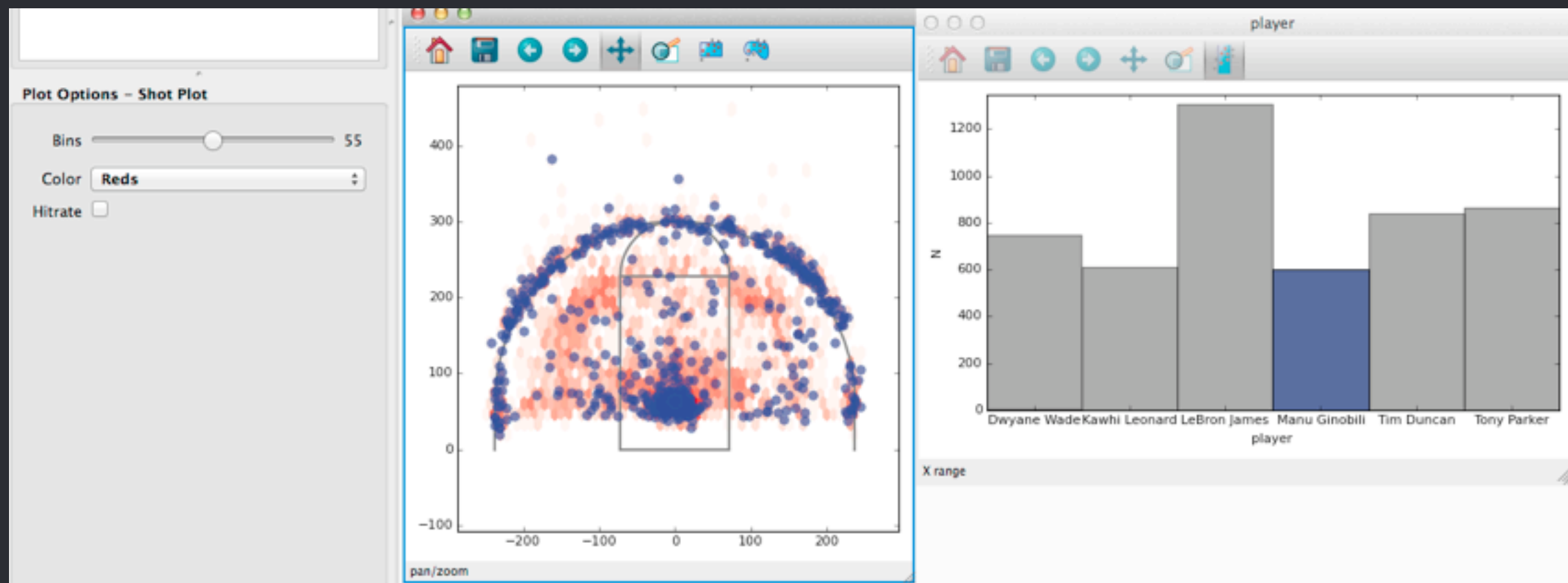
People could see 200 previous replays as grey ghost marios while they played

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>



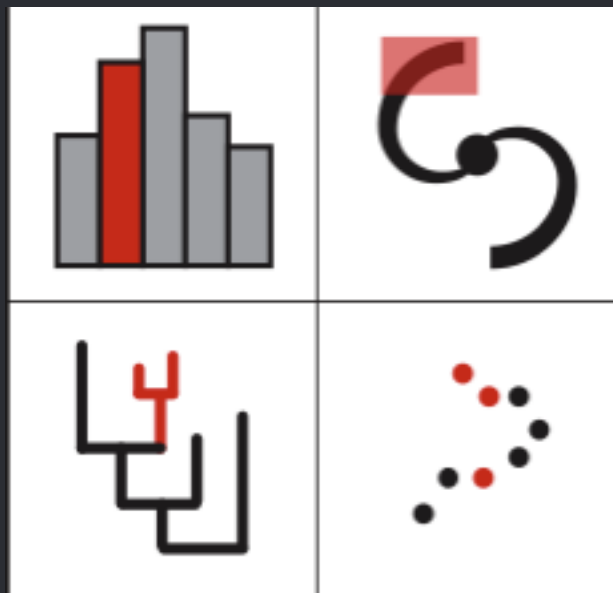
Maurice Guégan | <http://stabyourself.net/mari0/>



Data Interfaces

Code

GUI



Data Interfaces

Code

GUI

Expressive
Reproducible
Precise
Scalable

Slow
Hard

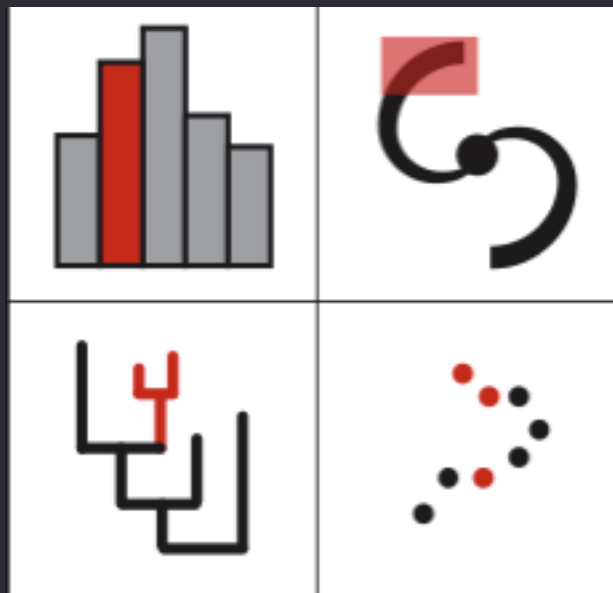


Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard



GUI

Intuitive
Nonsequential
Fast

Canned
Imprecise

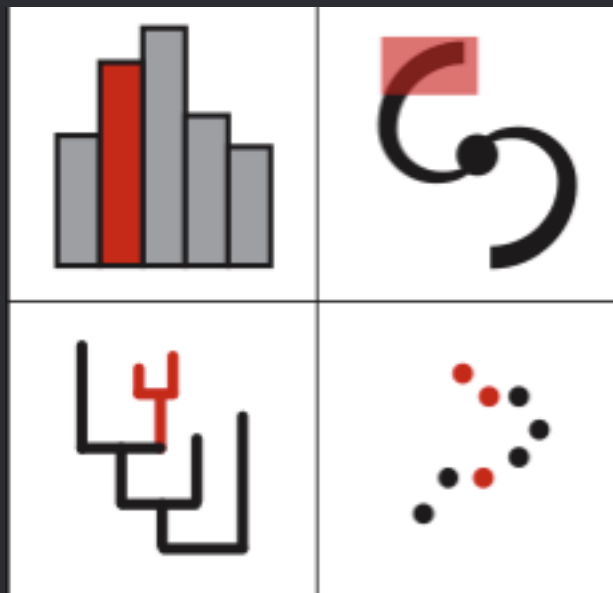
Data Interfaces

Code

Expressive
Reproducible
Precise
Scalable

Slow
Hard

Hackable
User Interfaces



GUI

Intuitive
Nonsequential
Fast

Canned
Imprecise

Thanks!

glueviz.org
@BeaumontChris

