

TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION



Ten Questions to Ask When Creating a Scientific Visualization

Here are 10 good questions to ask when you create a scientific visualization:

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. **Feature recognition** | Is feature and/or pattern recognition, a goal?
4. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
5. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
6. **Categories & Clustering** | Are there natural, or imposed, categories within the data? Are you interested in clustering?
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?

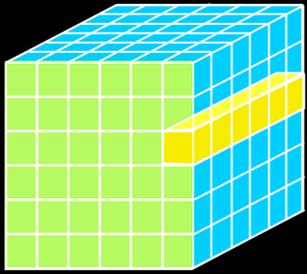
Astronomy 215hf (T,Th at 10 AM starting 3/22 for 2 weeks)

HUMANS-IN-THE-LOOP: VISUALIZATION + MACHINE LEARNING*



ALYSSA GOODMAN

*WITH PIZZA



"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"

LINKED VIEWS OF HIGH-DIMENSIONAL DATA

Big DATA versus wide DATA



BIG DATA AND "HUMAN-AIDED COMPUTING"

INTERACTION BEYOND FLATLAND IS AN UNSOLVED PROBLEM



John Tukey's warning: "details of control can make or break such a system"

Microsoft HoloLens



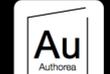
1665

1895

2009

2015

COMMUNICATION



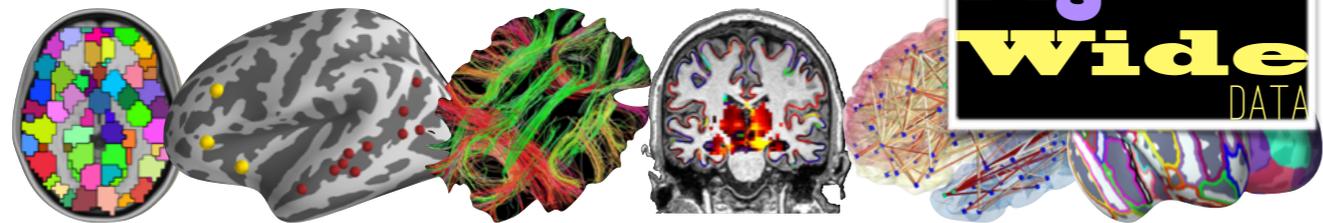
“Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives”

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups’ ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven approach for “smart selection” of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing “regions of interest” (ROIs) is critical in both research and clinical work. This process, called “segmentation,” is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D “slice” images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

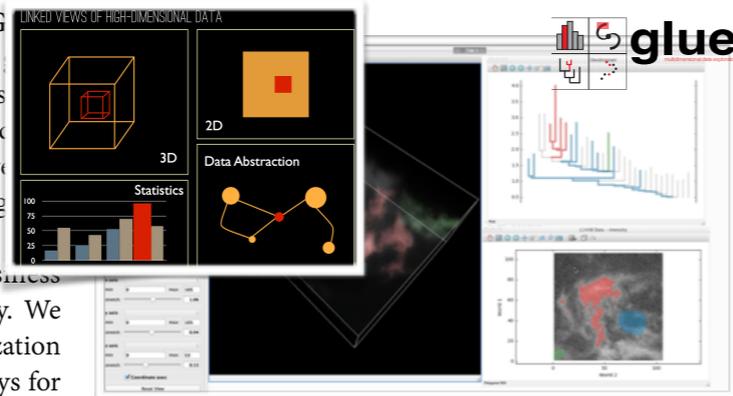
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. In brain images below all use colored highlighting to indicate salient “features.”



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow **multiple data sets** from diverse sources to be **analyzed in concert** (see #2). And, second, there is no way to draw **arbitrary surfaces** needed to make a “**selection**” in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of “Glue.” The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called “linked views,” where one on-screen map, chart or graph data set updates *live* to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence and data analytics, as data sets grow in size and diversity. We have recently developed the world’s first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices for interacting with virtual environments in more and more immersive ways are being developed as computers get more powerful and as profits grow. The opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person’s real hands and eyes, with virtual projections, like the image of a human heart shown in the image above.

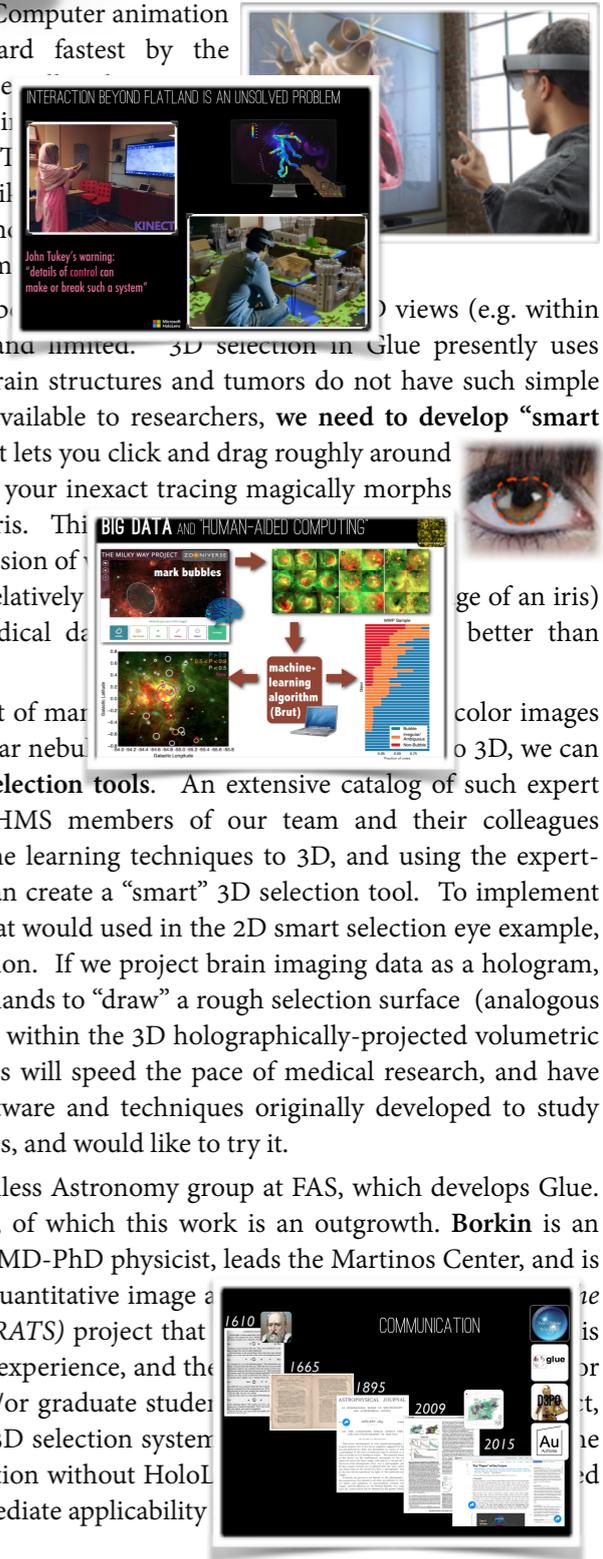
PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within slices or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop “**smart selection**” in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris’ edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This “selection” (outlining the object) and “segmentation” (the decision of whether a pixel belongs to the object) are solved for relatively simple shapes. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms to find similar nebulae to train a machine learning algorithm to find similar nebulae. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a “smart” 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to “draw” a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that is one of their labs to host a part-time postdoctoral fellow and/or graduate student and to purchase a HoloLens. Our goal is a working smart 3D selection system of unprecedented nature of what we propose, a smart 3D selection without HoloLens. A working smart 3D selection tool, would be successes on their own, with immediate applicability.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>

Glue,” is supported by NASA because it offers high-dimensional data that comes from so-called *Astronomical Medicine*, Glue is just as useful for MRI and CT scans, as it is for astronomy. As a result, in one data view, live, across all views, you can interact with the data. (Short demo videos: [2D](#), [3D](#))





PATTERN RECOGNITION
CREATIVITY



CALCULATIONS
ENDURANCE

Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

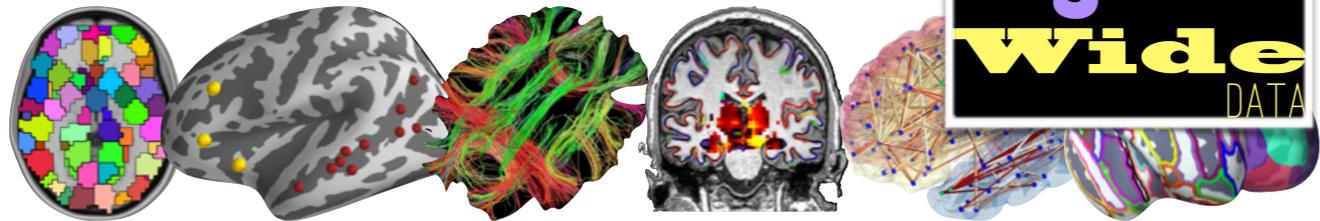
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

Three recent developments make the proposed research feasible now.

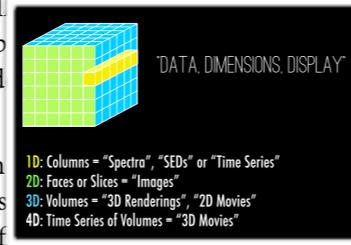
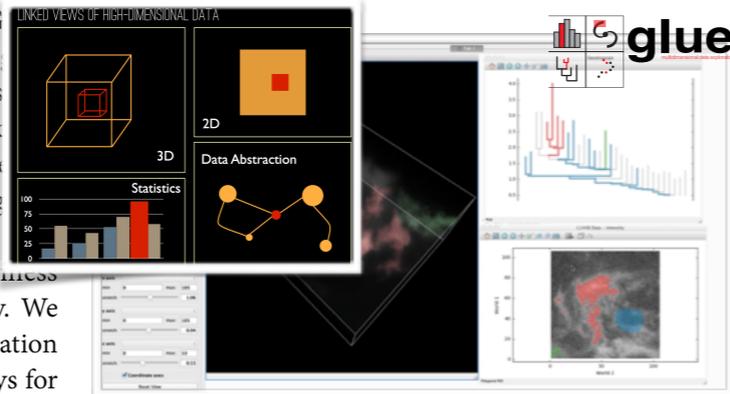
1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow multiple data sets from diverse sources to be analyzed in concert (see #2). And, second, there is no way to draw arbitrary surfaces needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"

The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates live to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence and analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



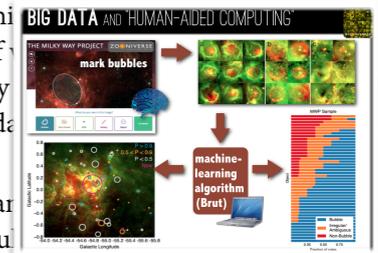
images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on *Astronomical Medicine*, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether a pixel belongs to the object) are solved for relatively quickly. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms to find similar nebulae. For example, we used the output of machine learning algorithms to find similar nebulae. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that has been used by one of their labs to host a part-time postdoctoral fellow and/or graduate student and to purchase a HoloLens. Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

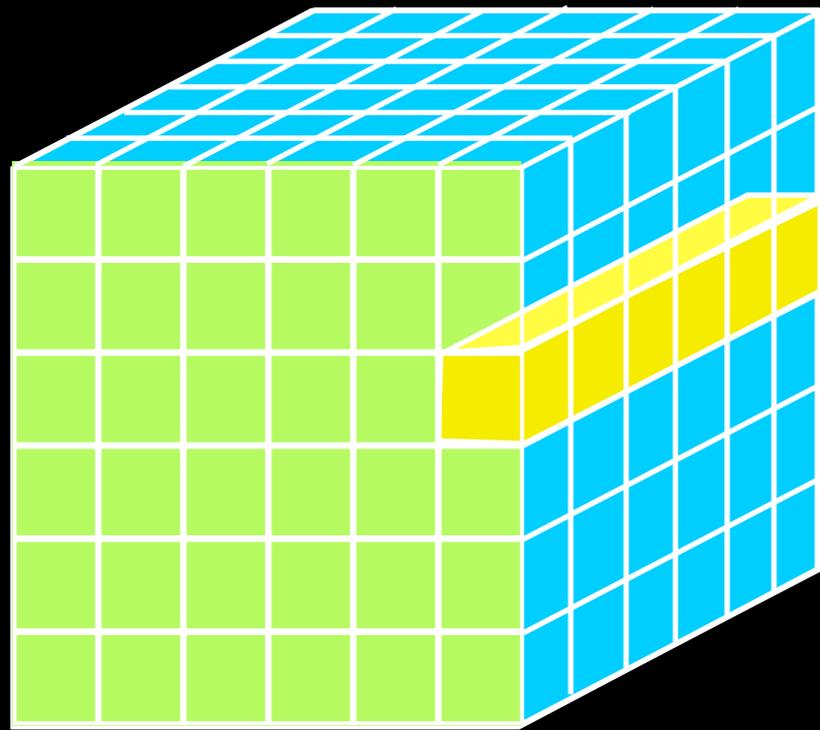
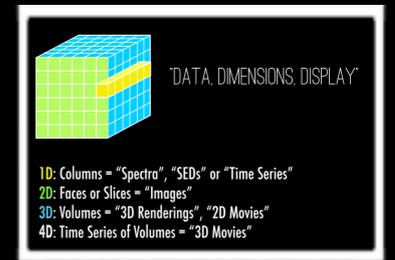


...ge of an iris) better than color images to 3D, we can



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>



"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"

Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

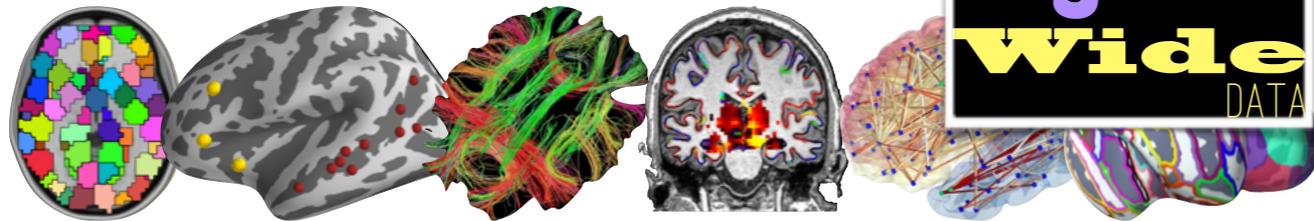
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

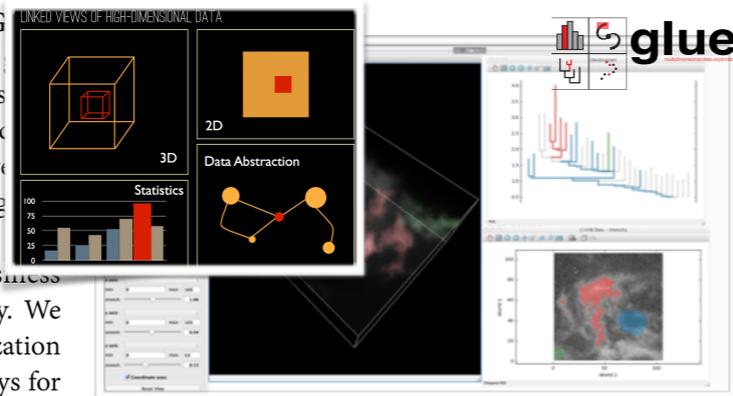
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow multiple data sets from diverse sources to be analyzed in concert (see #2). And, second, there is no way to draw arbitrary surfaces needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"
The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates live to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence and data analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on *Astronomical Medicine*, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

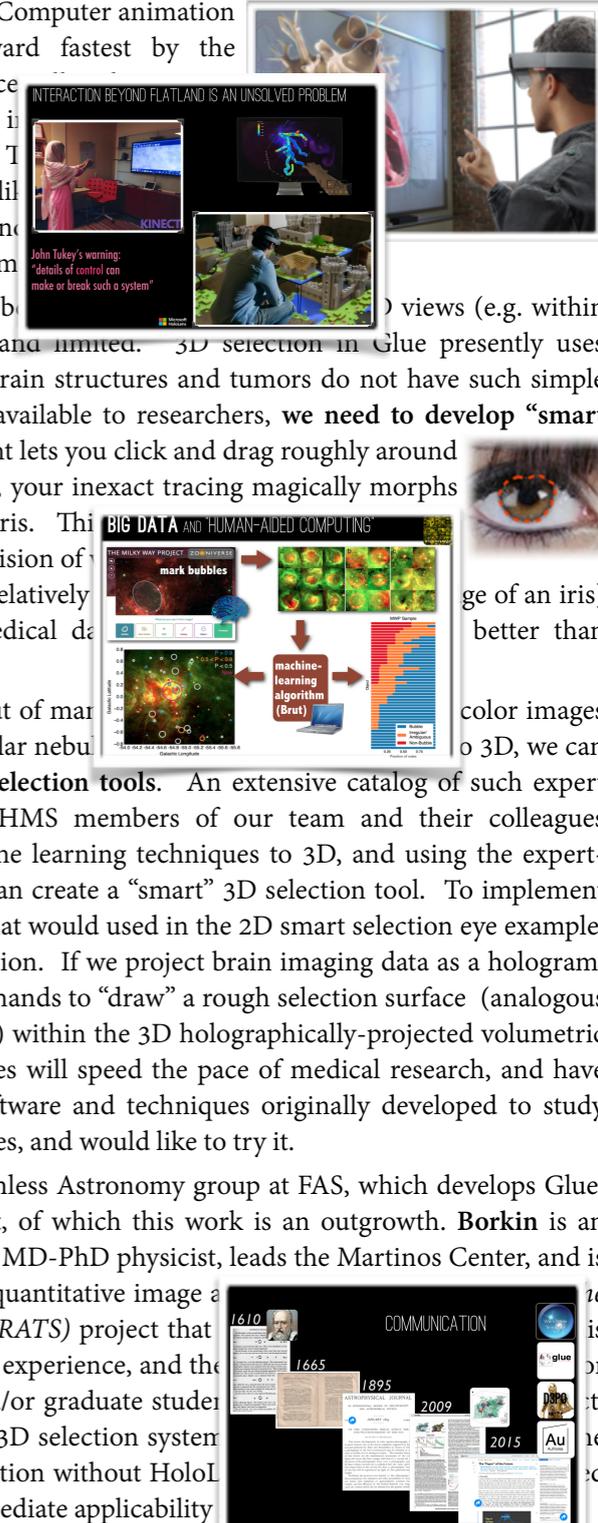
3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether a pixel belongs to the object) are solved for relatively quickly. In both 2D and 3D, the segmentation problem is solved for relatively simple features. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms of nebulae to train a machine learning algorithm to find similar nebulae. We can use experts' segmentations of brain imaging to train 3D selection tools. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that is one of their labs to host a part-time postdoctoral fellow and/or graduate student and to purchase a HoloLens. Our goal is a working smart 3D selection system. The unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>





BIG DATA, WIDE DATA

Use Layer Manager to Control User Settings



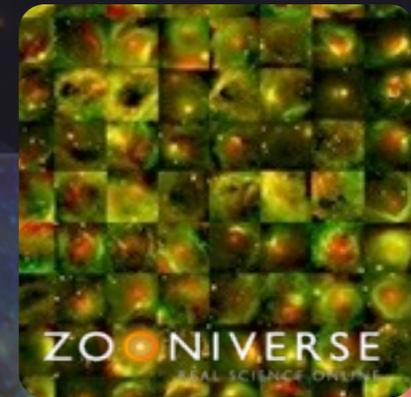
Name My Location
Lat 37:47:15 Alt 0 m
Lng -123:35:23
View From This Location

2015/02/11 04:40:33
Real Time
Now

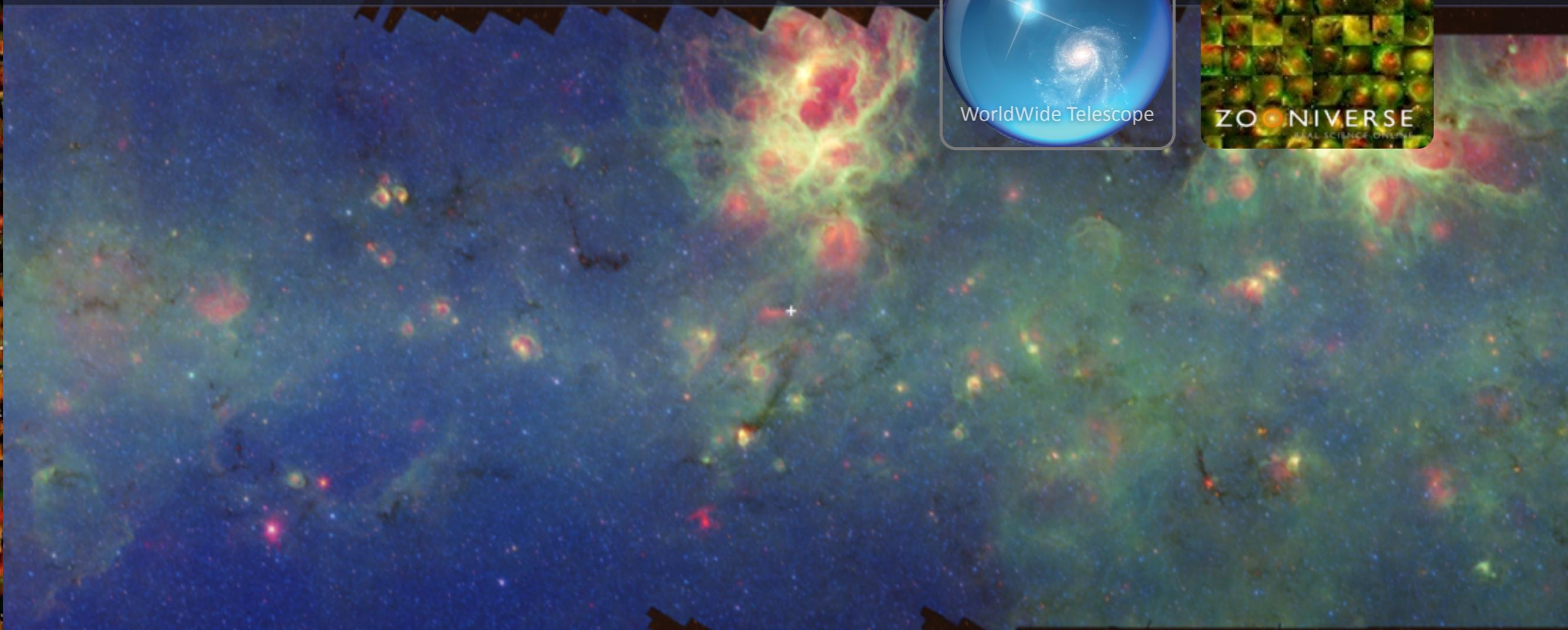
Galactic Plane Mode



WorldWide Telescope



ZOONIVERSE
REAL SCIENCE ONLINE



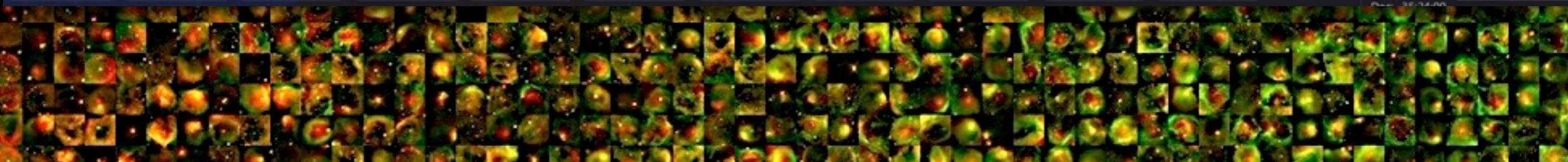
Look At: Sky Imagery: Digitized Sky Survey (Color) Image Crossfade: [Slider]

Tracking: GLIMPSE/MIPSGAL 1 of 3

Scorpius 03:10:14

RA: 17h28m14s

Pismis 24 and NGC6334	NGC6357	NGC6374	NGC6383	NGC6396	NGC6404	Lesath	Shaula	HR6397	HR6405



BIG DATA AND "HUMAN-AIDED COMPUTING"

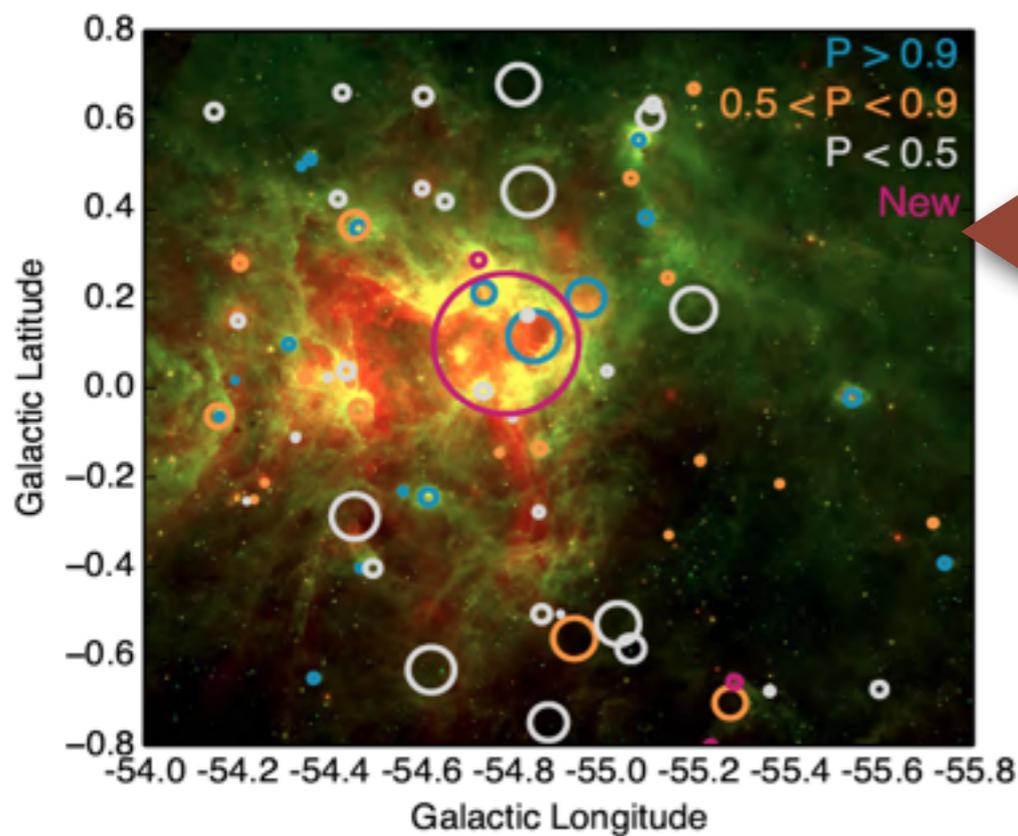
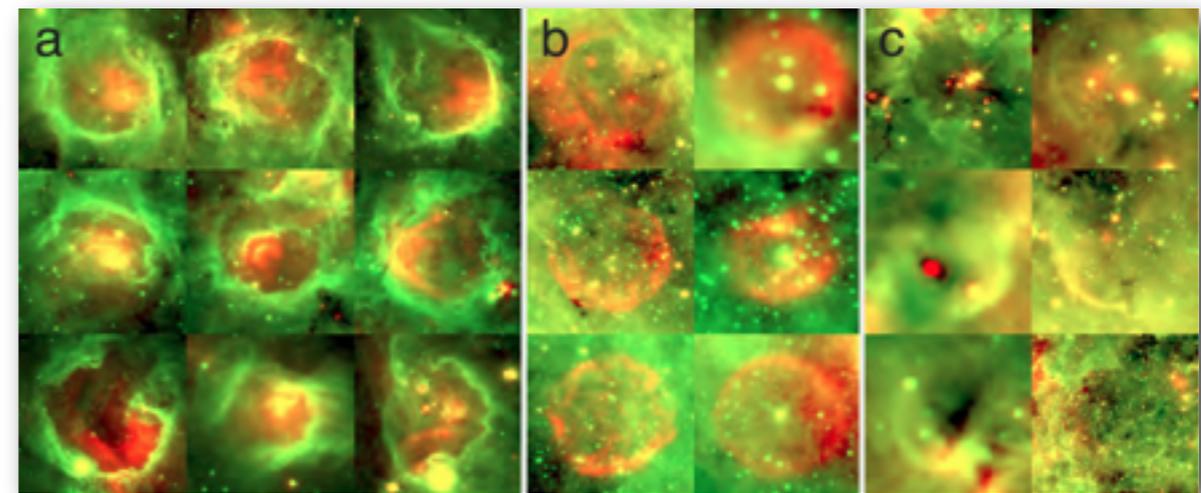


THE MILKY WAY PROJECT ZOO NIVERSE REAL SCIENCE ONLINE

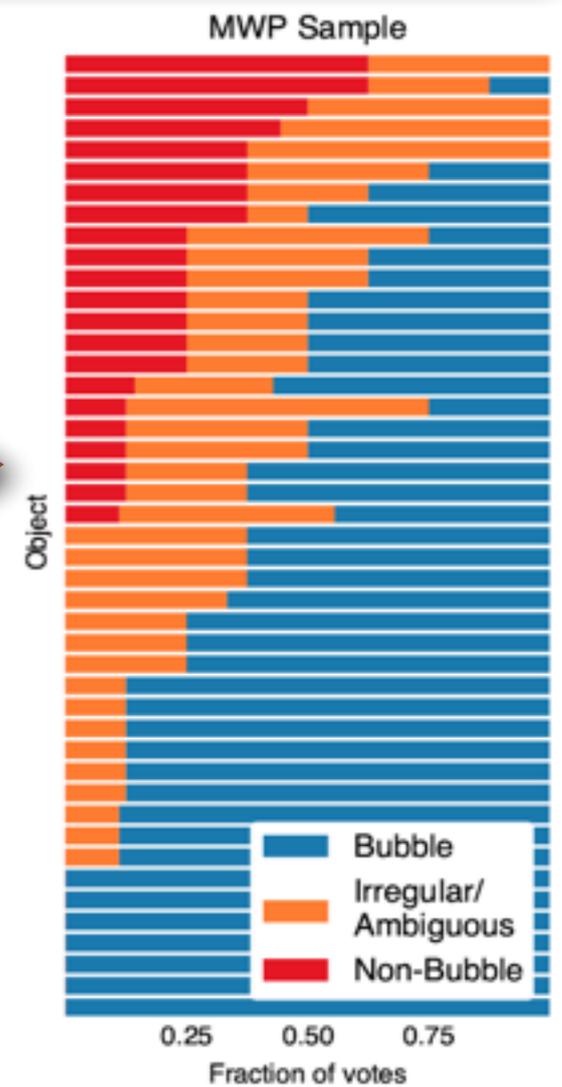
mark bubbles

What do you see in this image?

Bubble Star Cluster EGO Galaxy Object I'm done!



machine-learning algorithm (Brut)



example here from: Beaumont, Goodman, Kendrew, Williams & Simpson 2014; based on Milky Way Project catalog (Simpson et al. 2013), which came from Spitzer/GLIMPSE (Churchwell et al. 2009, Benjamin et al. 2003), cf. Shenoy & Tan 2008 for discussion of HAC; astroml.org for machine learning advice/tools

THE MILKY WAY PROJECT: LEVERAGING CITIZEN SCIENCE AND MACHINE LEARNING TO DETECT INTERSTELLAR BUBBLES

CHRISTOPHER N. BEAUMONT^{1,2}, ALYSSA A. GOODMAN², SARAH KENDREW³, JONATHAN P. WILLIAMS¹, AND ROBERT SIMPSON³

¹ Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA; beaumont@ifa.hawaii.edu

² Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

³ Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Received 2013 October 3; accepted 2014 June 7; published 2014 August 20

ABSTRACT

We present Brut, an algorithm to identify bubbles in infrared images of the Galactic midplane. Brut is based on the Random Forest algorithm, and uses bubbles identified by >35,000 citizen scientists from the Milky Way Project to discover the identifying characteristics of bubbles in images from the *Spitzer Space Telescope*. We demonstrate that Brut's ability to identify bubbles is comparable to expert astronomers. We use Brut to re-assess the bubbles in the Milky Way Project catalog, and find that 10%–30% of the objects in this catalog are non-bubble interlopers. Relative to these interlopers, high-reliability bubbles are more confined to the mid-plane, and display a stronger excess of young stellar objects along and within bubble rims. Furthermore, Brut is able to discover bubbles missed by previous searches—particularly bubbles near bright sources which have low contrast relative to their surroundings. Brut demonstrates the synergies that exist between citizen scientists, professional scientists, and machine learning techniques. In cases where “untrained” citizens can identify patterns that machines cannot detect without training, machine learning algorithms like Brut can use the *output* of citizen science projects as *input* training sets, offering tremendous opportunities to speed the pace of scientific discovery. A hybrid model of machine learning combined with crowdsourced training data from citizen scientists can not only classify large quantities of data, but also address the weakness of each approach if deployed alone.

Key words: H II regions – ISM: bubbles – methods: data analysis – stars: formation

BRUT BASED ON
"RANDOM FOREST"*
(SUPERVISED LEARNING)

*Breiman, L. 2001, *Mach. Learn.*, 45, 5

ASCL.net
Astrophysics Source Code Library

Making codes discoverable since 1999

Search Site

Search

Home About Resources Browse Submissions News Forum

ASCL Code Record

[[ascl:1407.016](#)] [Brut: Automatic bubble classifier](#)

[Beaumont, Christopher](#); [Goodman, Alyssa](#); [Williams, Jonathan](#); [Kendrew, Sarah](#); [Simpson, Robert](#)

Brut, written in Python, identifies bubbles in infrared images of the Galactic midplane; it uses a database of known bubbles from the Milky Way Project and Spitzer images to build an automatic bubble classifier. The classifier is based on the Random Forest algorithm, and uses the WiseRF implementation of this algorithm.

Website: <https://github.com/ChrisBeaumont/brut>

Appears in: <http://adsabs.harvard.edu/abs/2014arXiv1406.2692B>

Bibcode: [2014ascl.soft07016B](#)

[Explain these fields?](#)

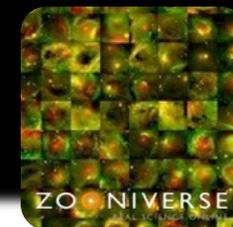
Discuss ↗



scikit-learn
Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

BIG DATA AND "HUMAN-AIDED COMPUTING"

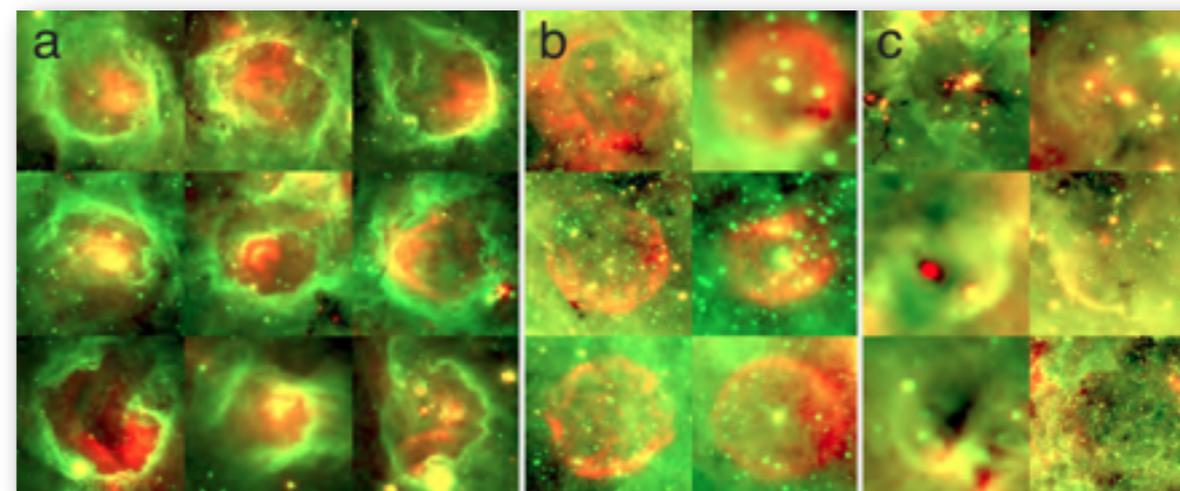
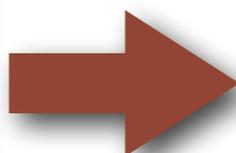


THE MILKY WAY PROJECT ZOONIVERSE REAL SCIENCE ONLINE

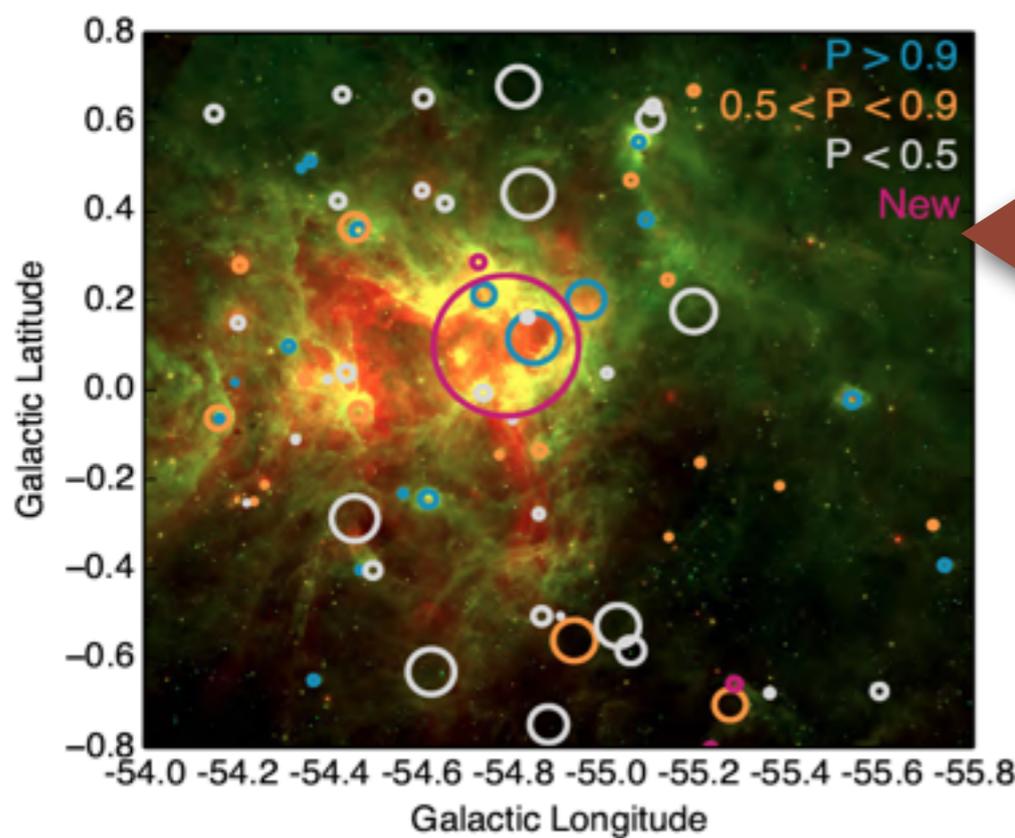
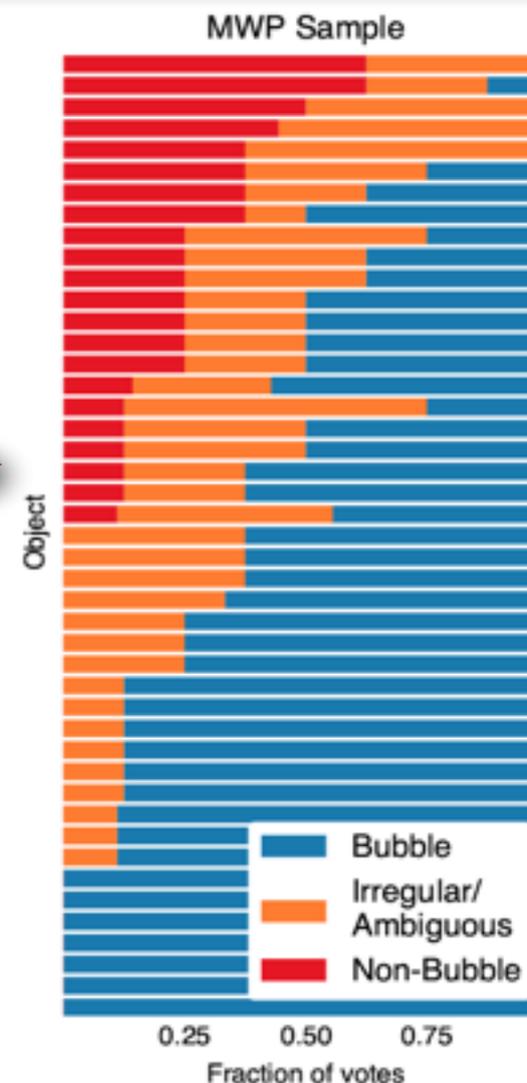
mark bubbles

What do you see in this image?

Bubble Star Cluster EGO Galaxy Object I'm done!

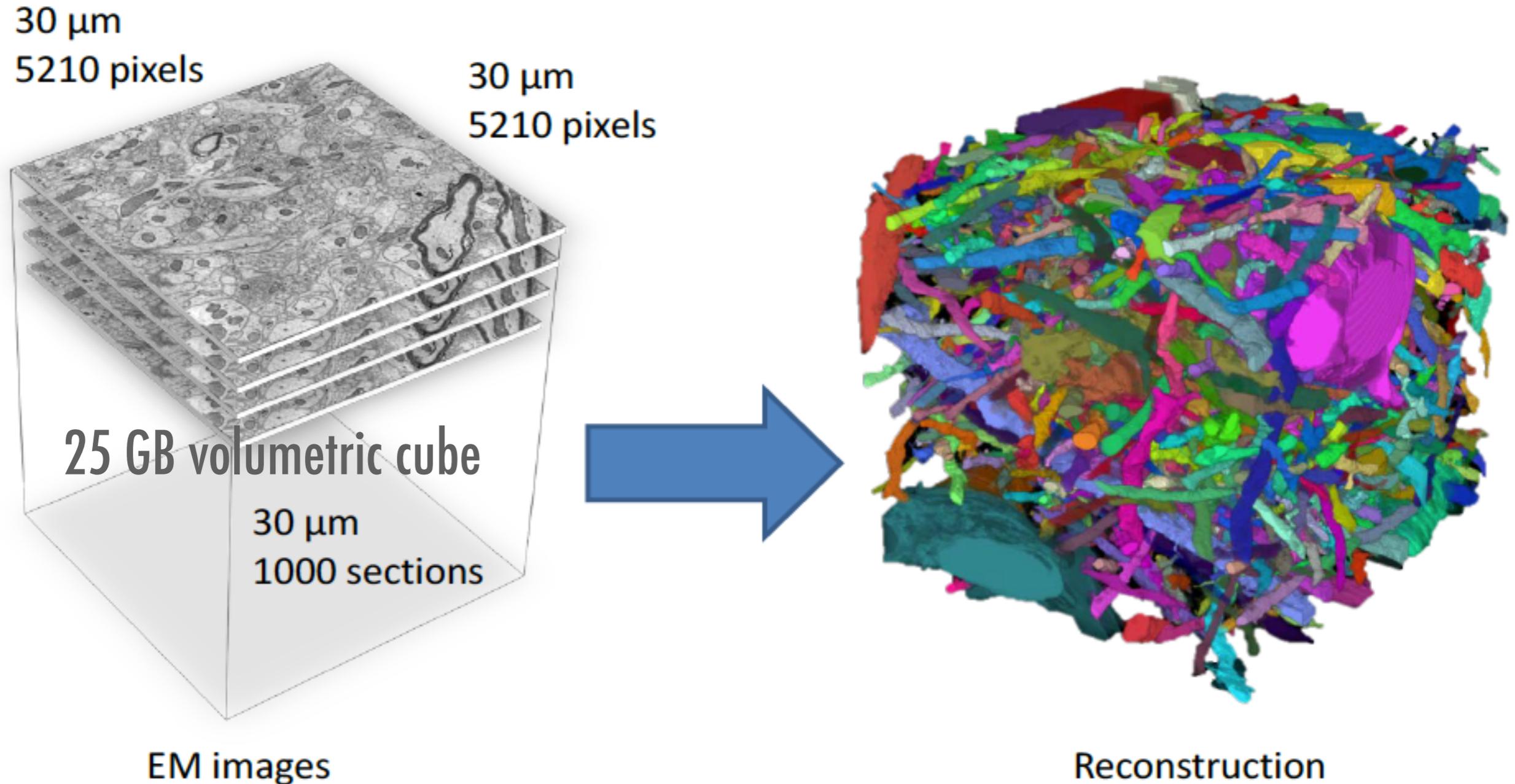


machine-learning algorithm (Brut)

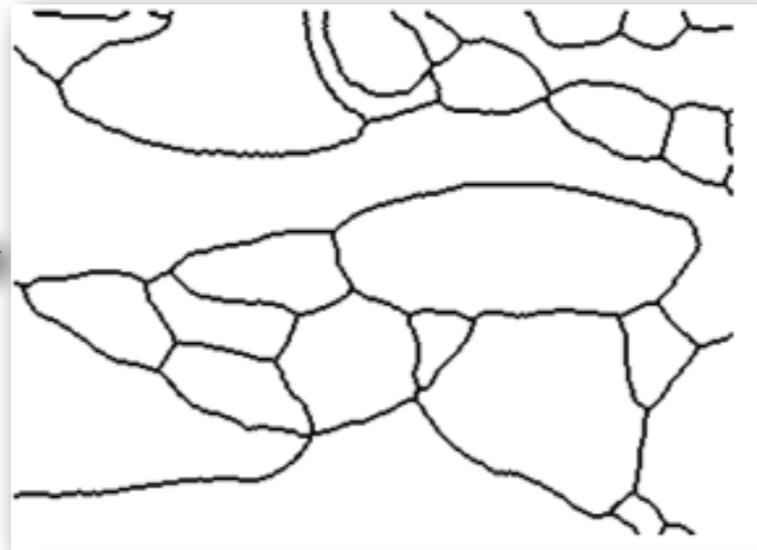
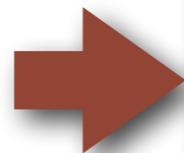
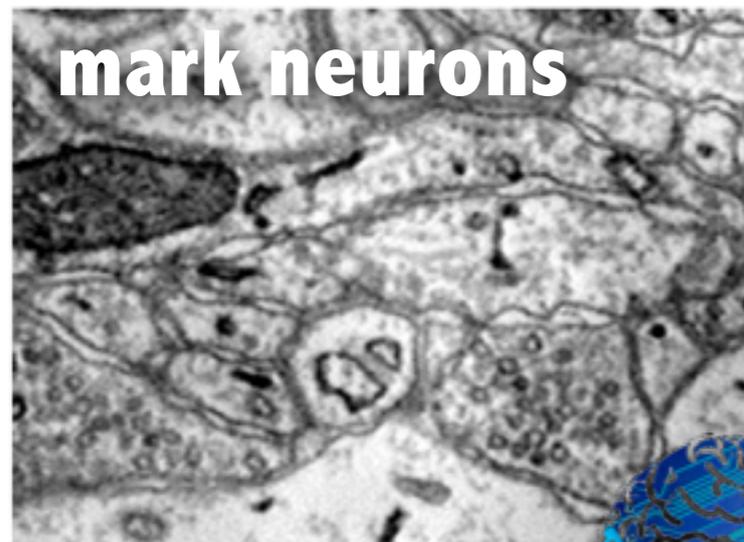


example here from: Beaumont, Goodman, Kendrew, Williams & Simpson 2014; based on Milky Way Project catalog (Simpson et al. 2013), which came from Spitzer/GLIMPSE (Churchwell et al. 2009, Benjamin et al. 2003), cf. Shenoy & Tan 2008 for discussion of HAC; astroml.org for machine learning advice/tools

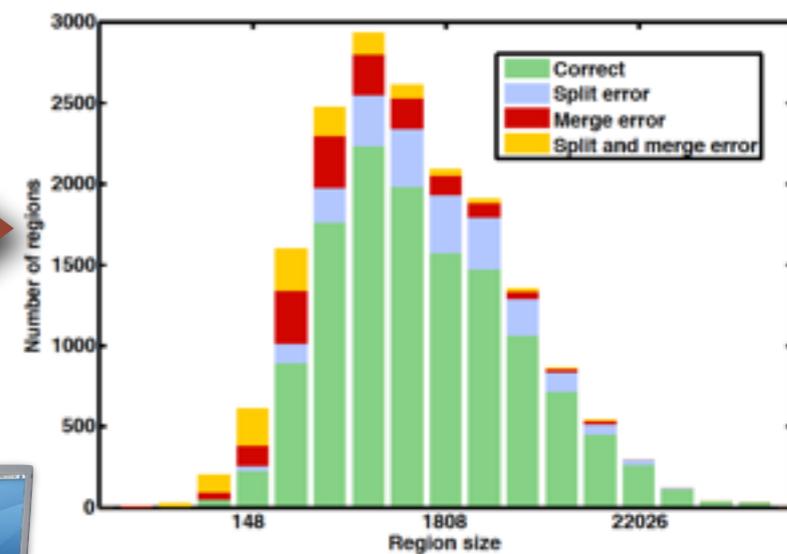
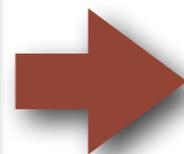
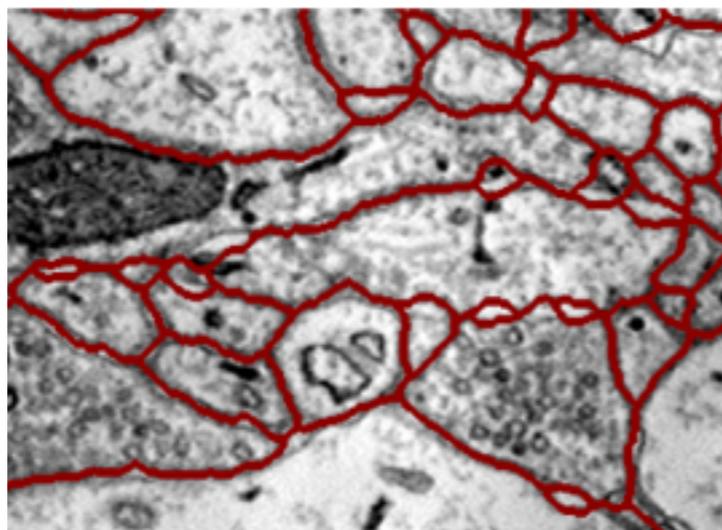
BIG DATA AND "HUMAN-AIDED COMPUTING"

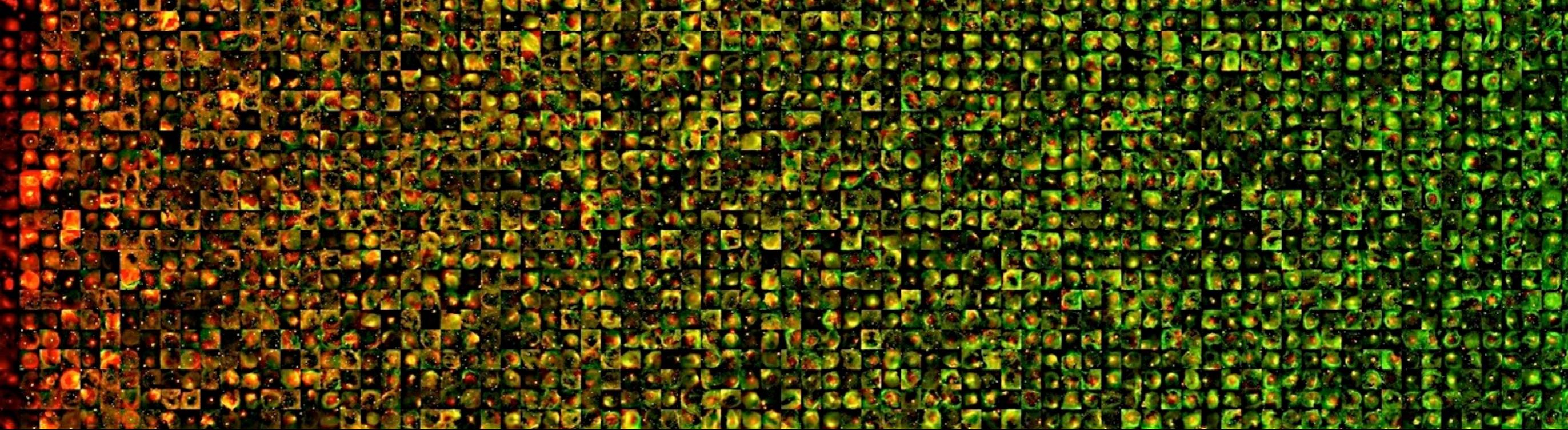


BIG DATA AND "HUMAN-AIDED COMPUTING"

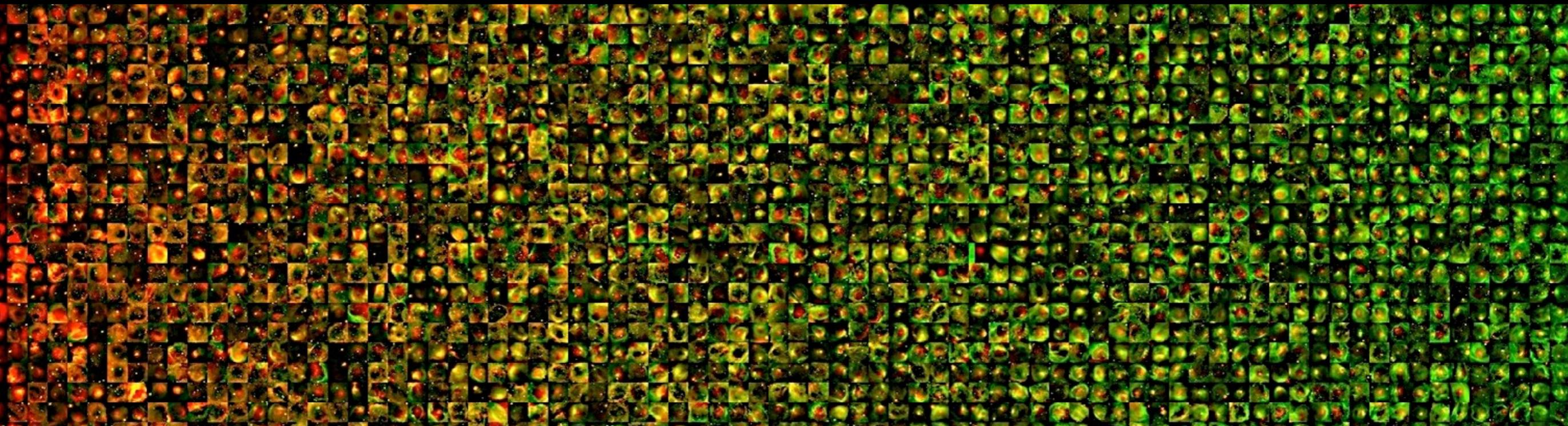


machine-learning algorithm (RF+CRF)

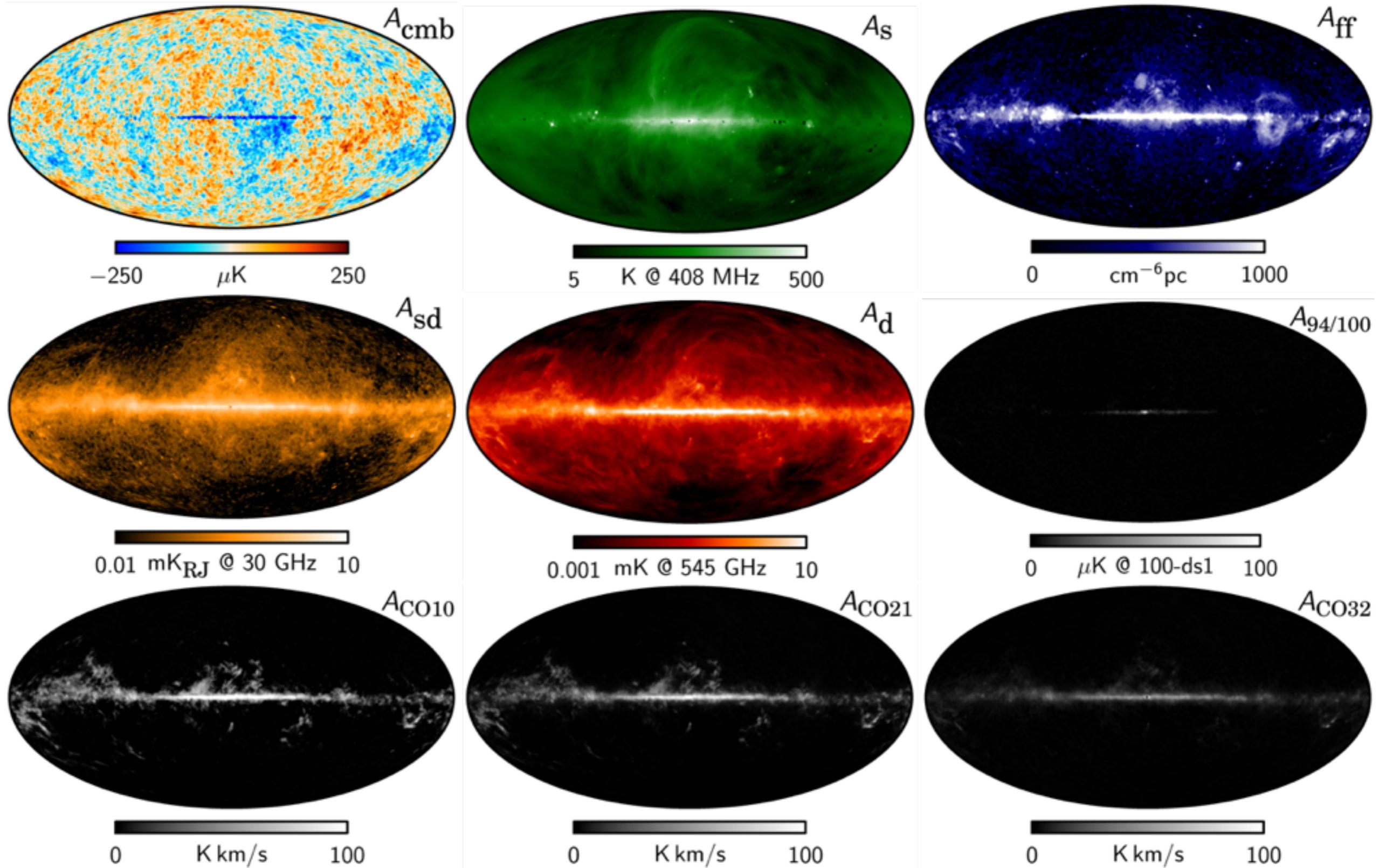




BIG DATA, WIDE DATA



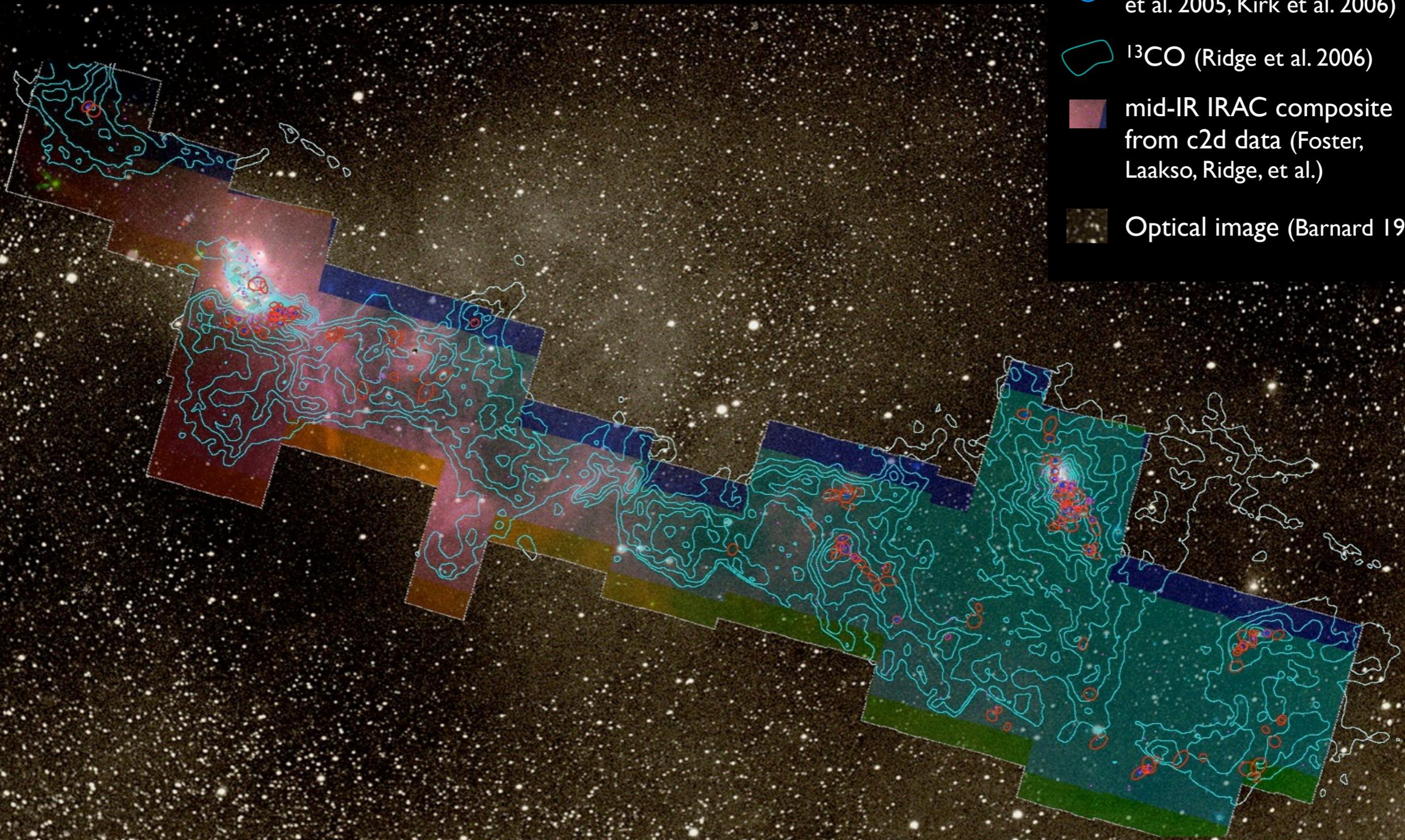
WIDE DATA



WIDE DATA

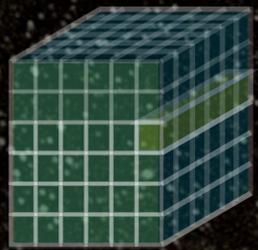
COMPLETE

-  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)



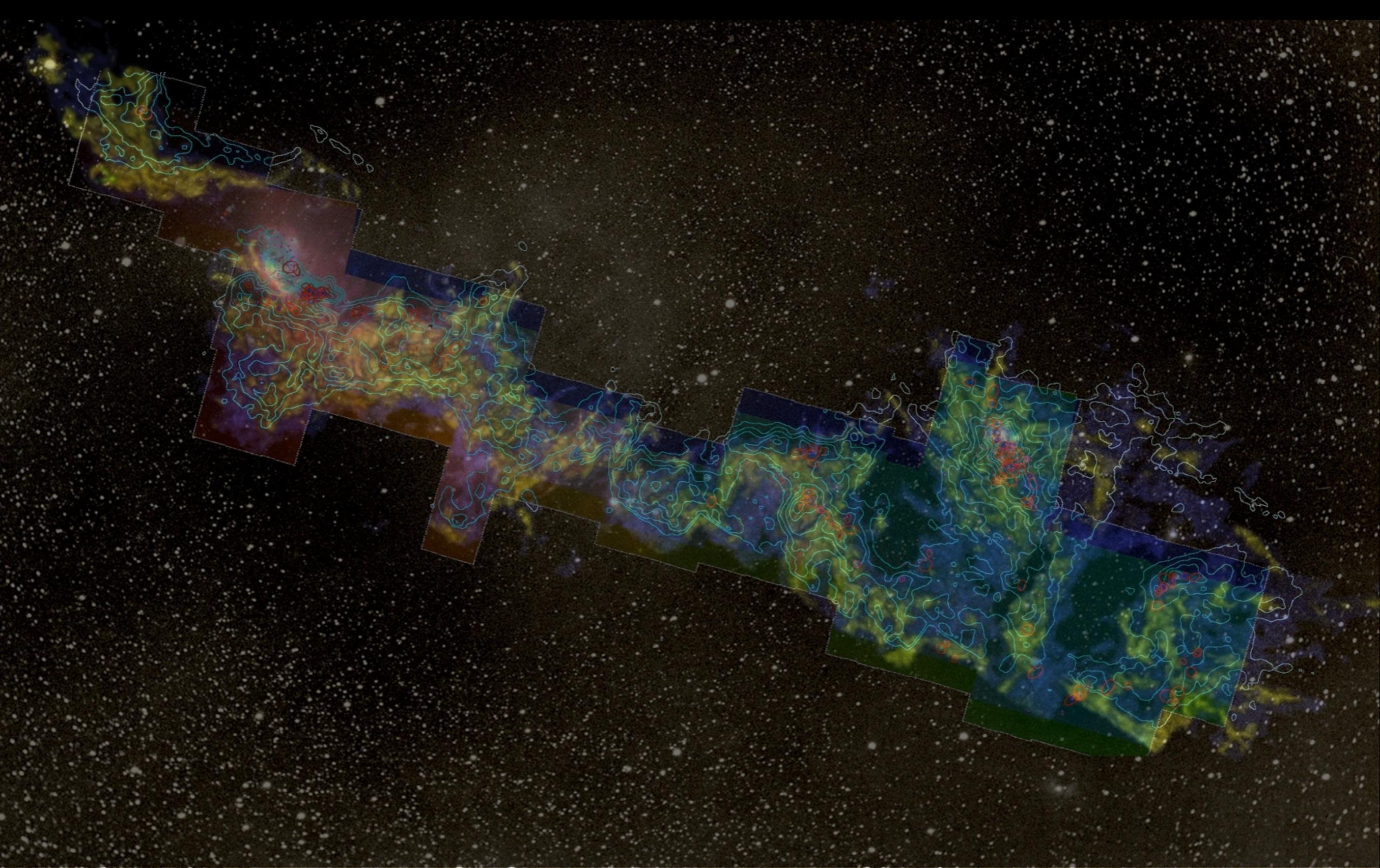
WIDE DATA, "IN 3D"

-  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)

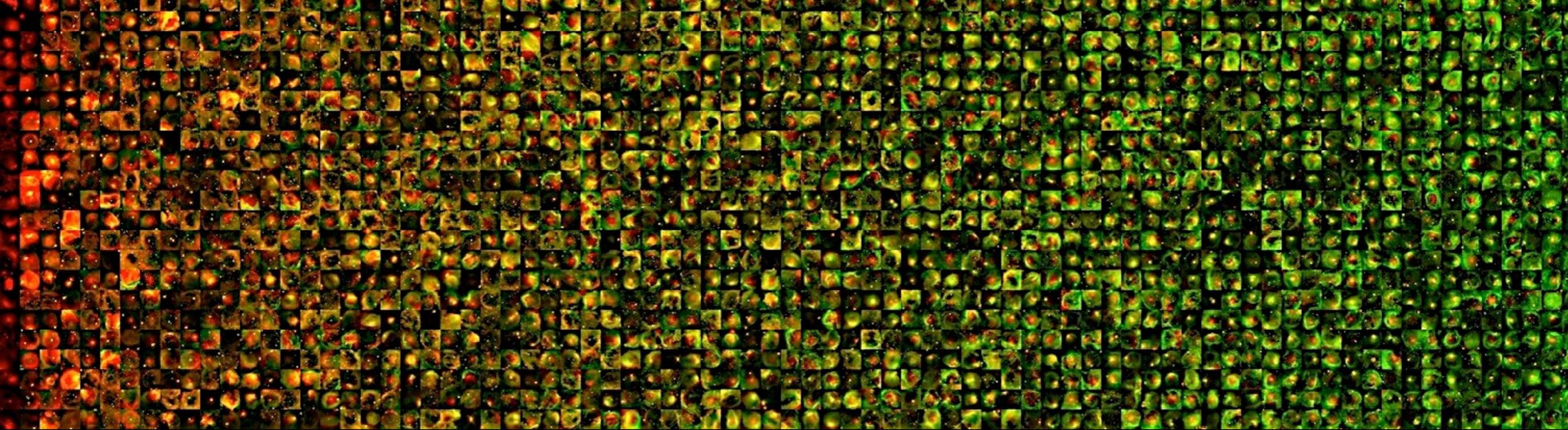


m: 17249
Zoom: 227% Angle: 0

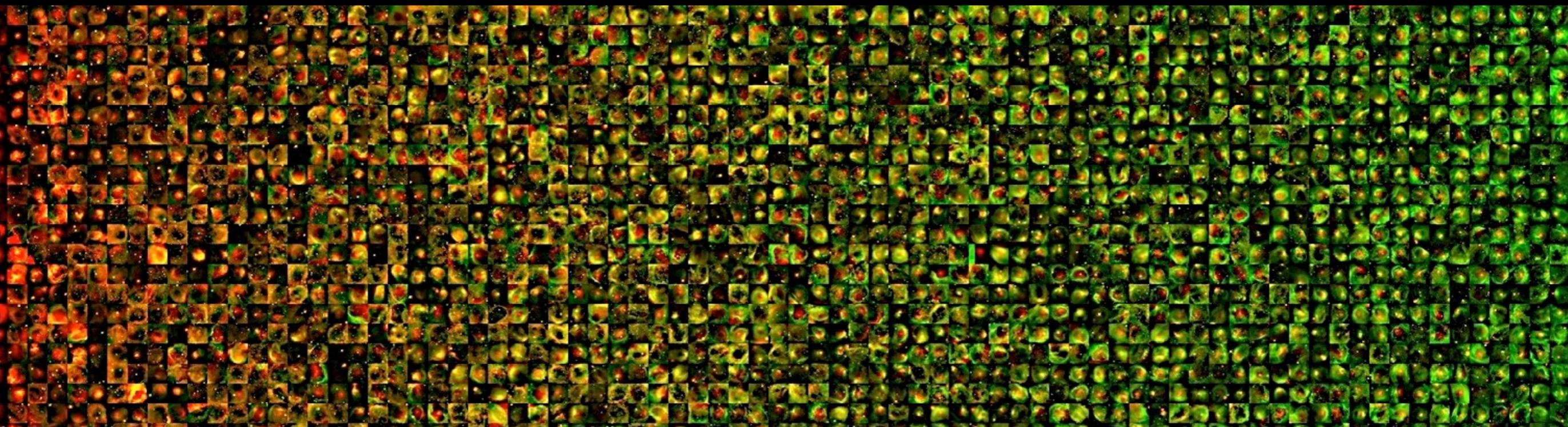


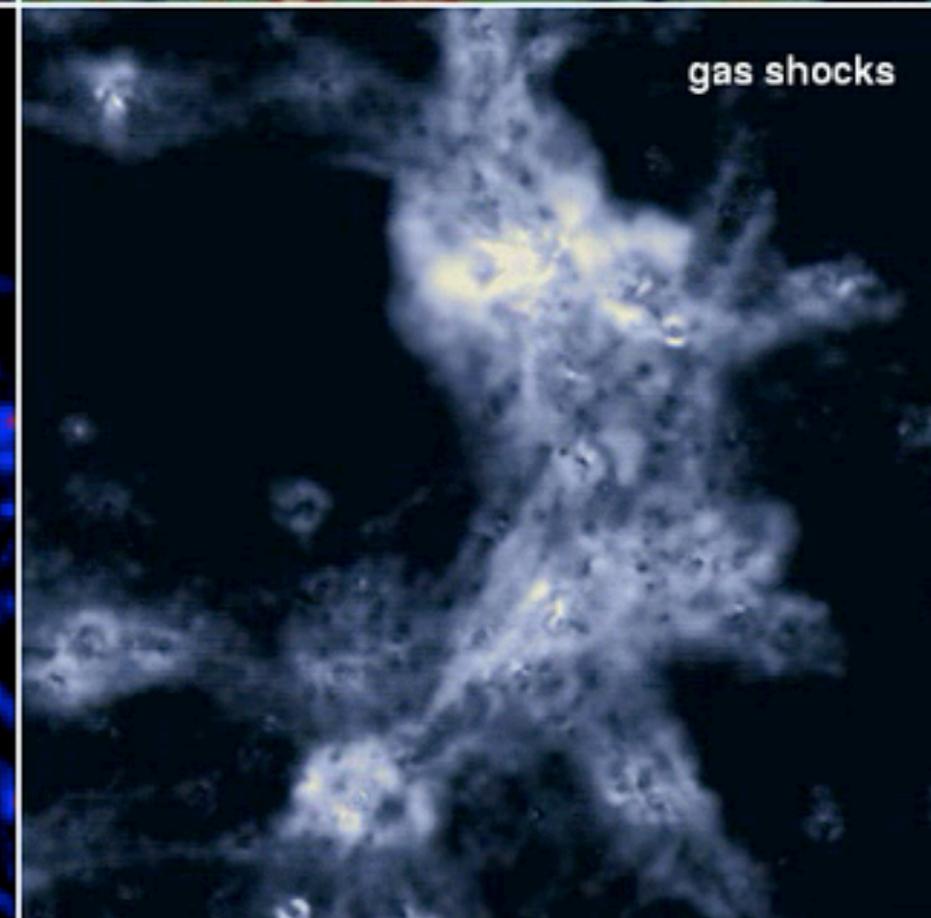
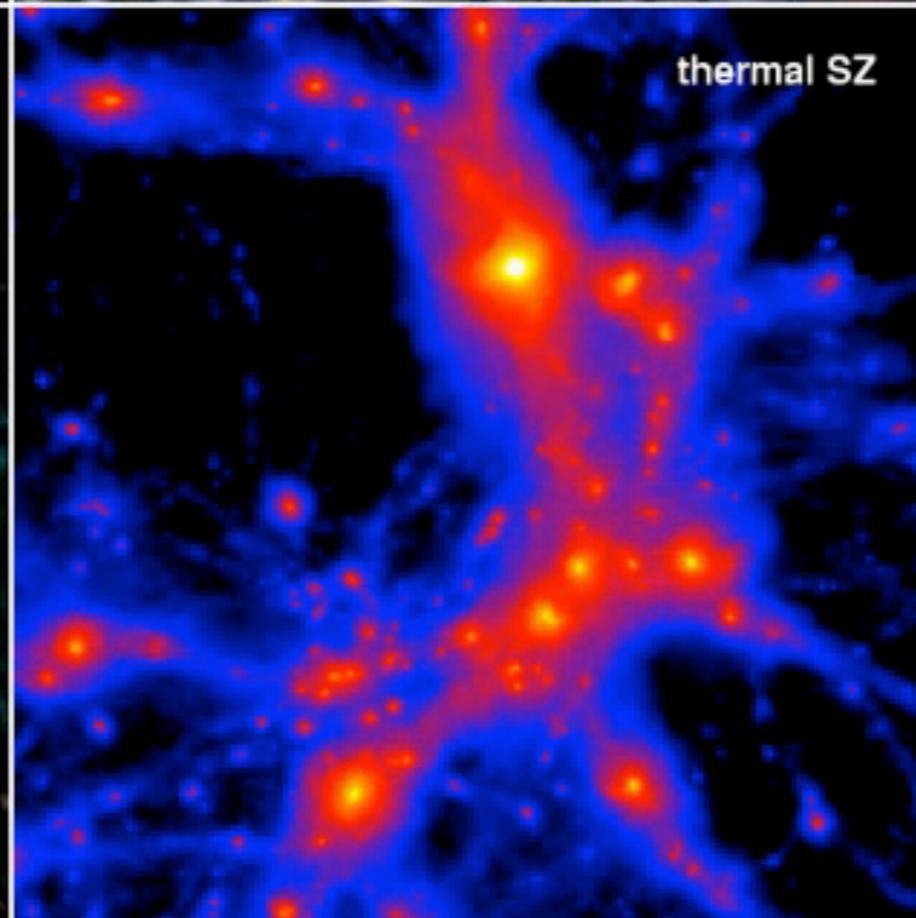
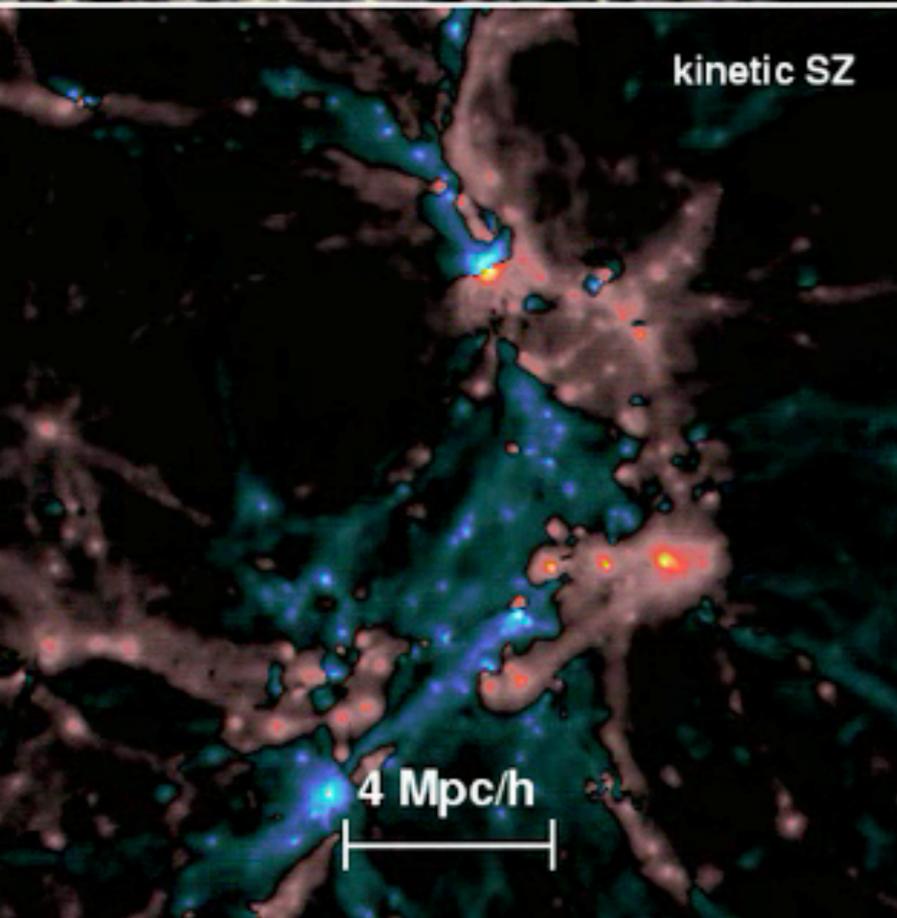
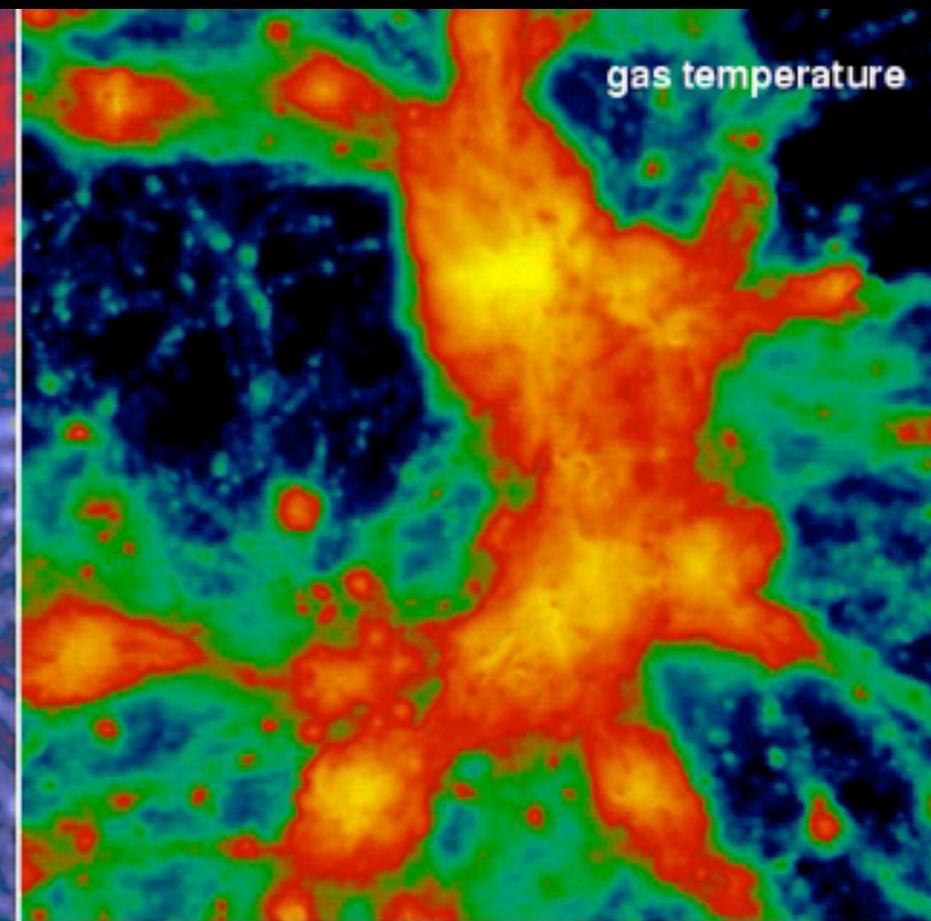
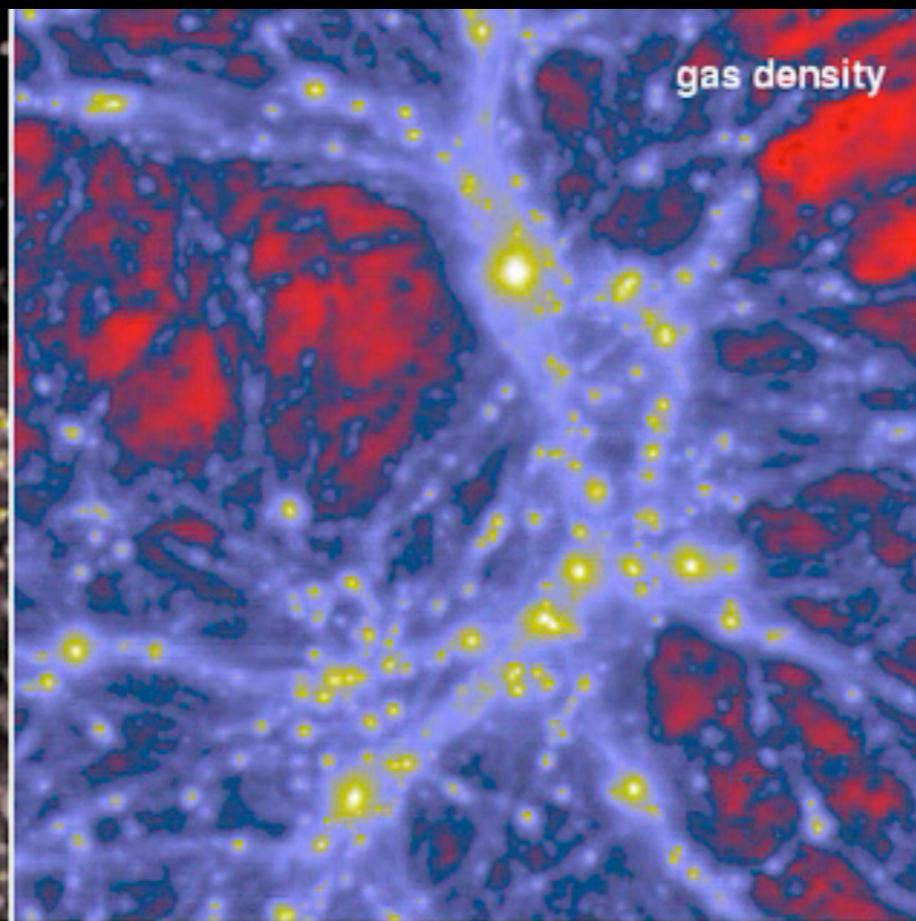
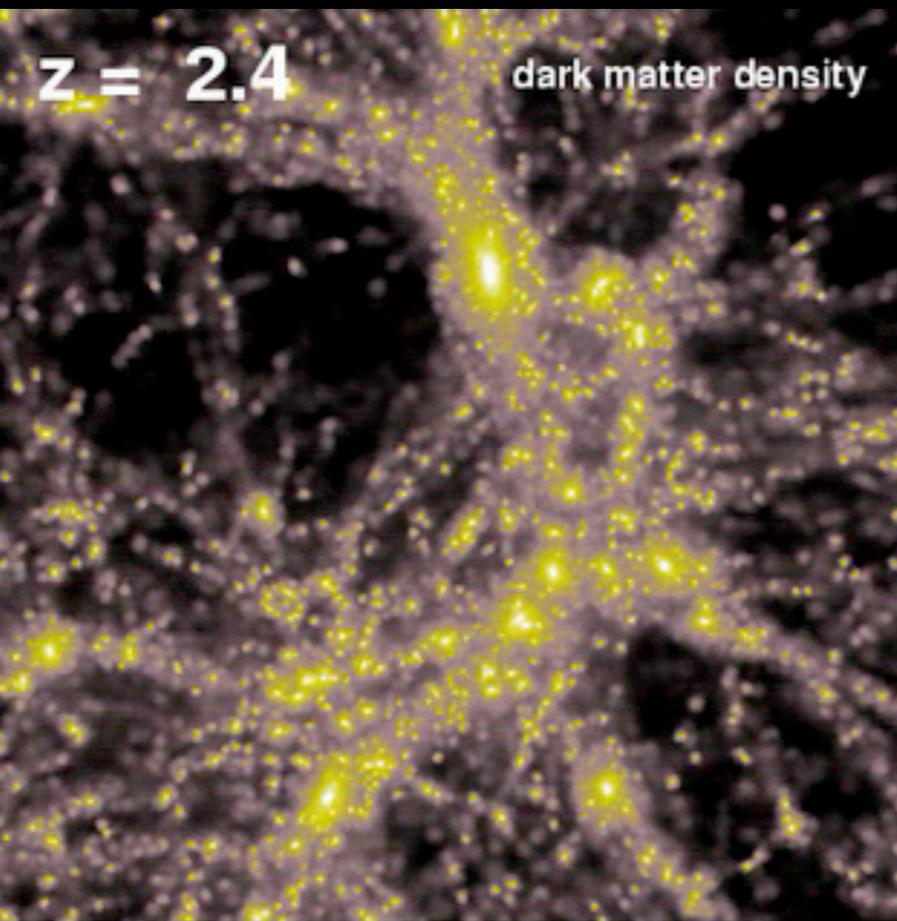


3D Viz made with VolView



BIG AND WIDE DATA





Movie: Volker Springel, formation of a cluster of galaxies. Millenium Simulation requires 25TB for output.

Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

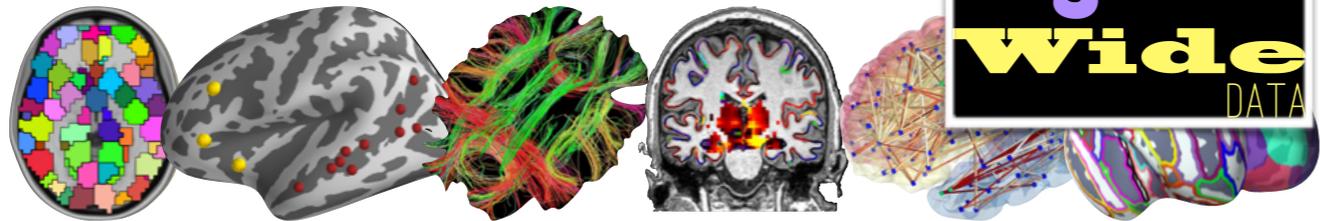
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

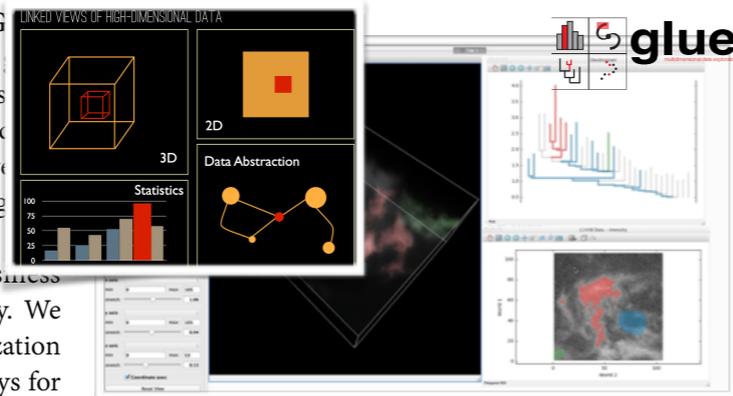
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow multiple data sets from diverse sources to be analyzed in concert (see #2). And, second, there is no way to draw arbitrary surfaces needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"
The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates live to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on *Astronomical Medicine*, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

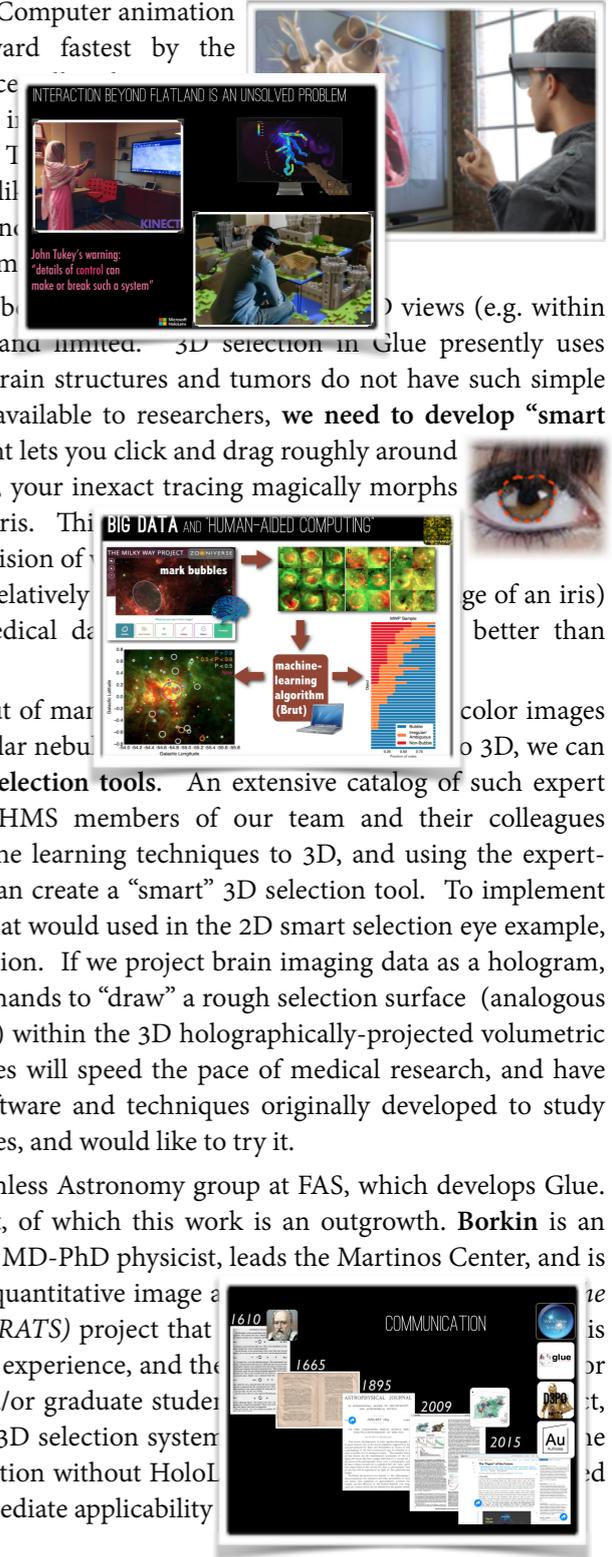
PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether a pixel belongs to the object) are solved for relatively quickly. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms to find similar nebulae. For example, we used the output of machine learning algorithms to find similar nebulae. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that has been used by one of their labs to host a part-time postdoctoral fellow and/or graduate student, and to purchase a HoloLens. Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

glue

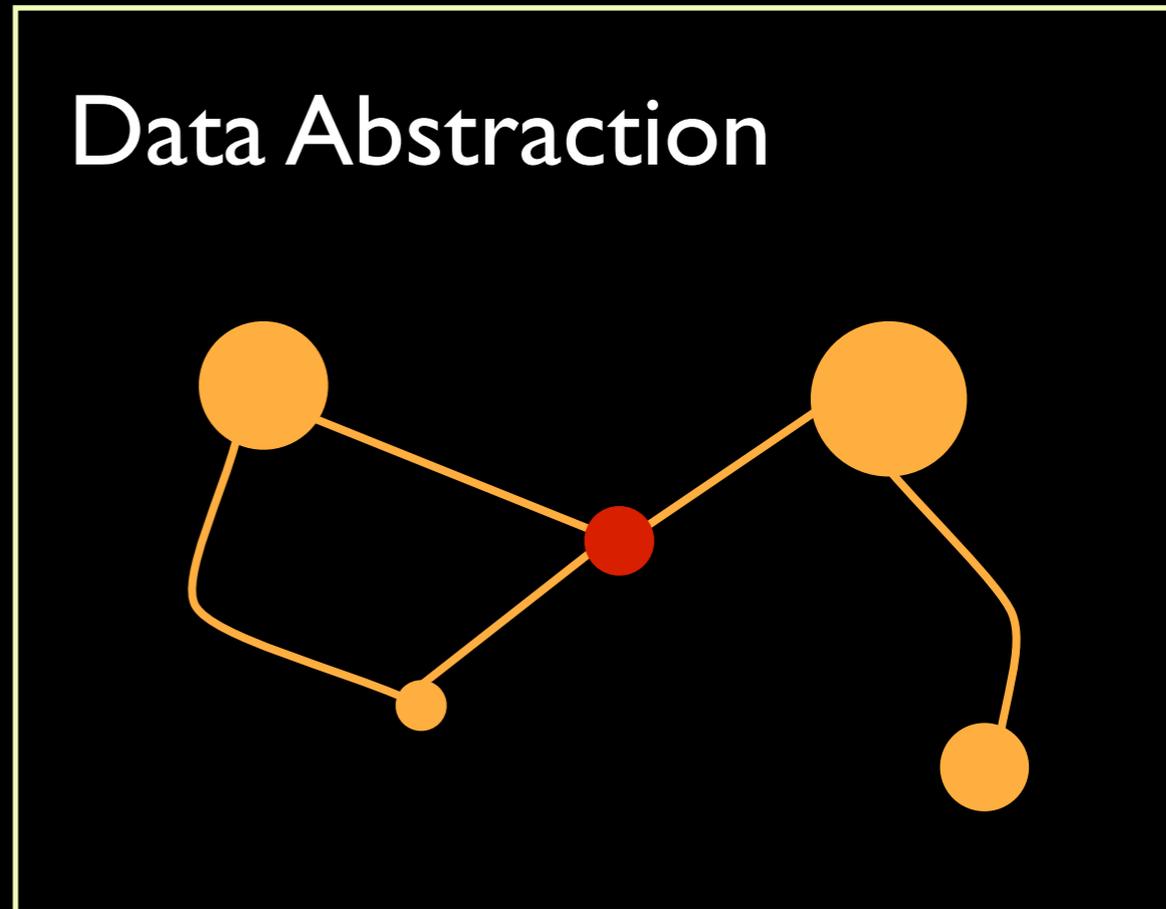
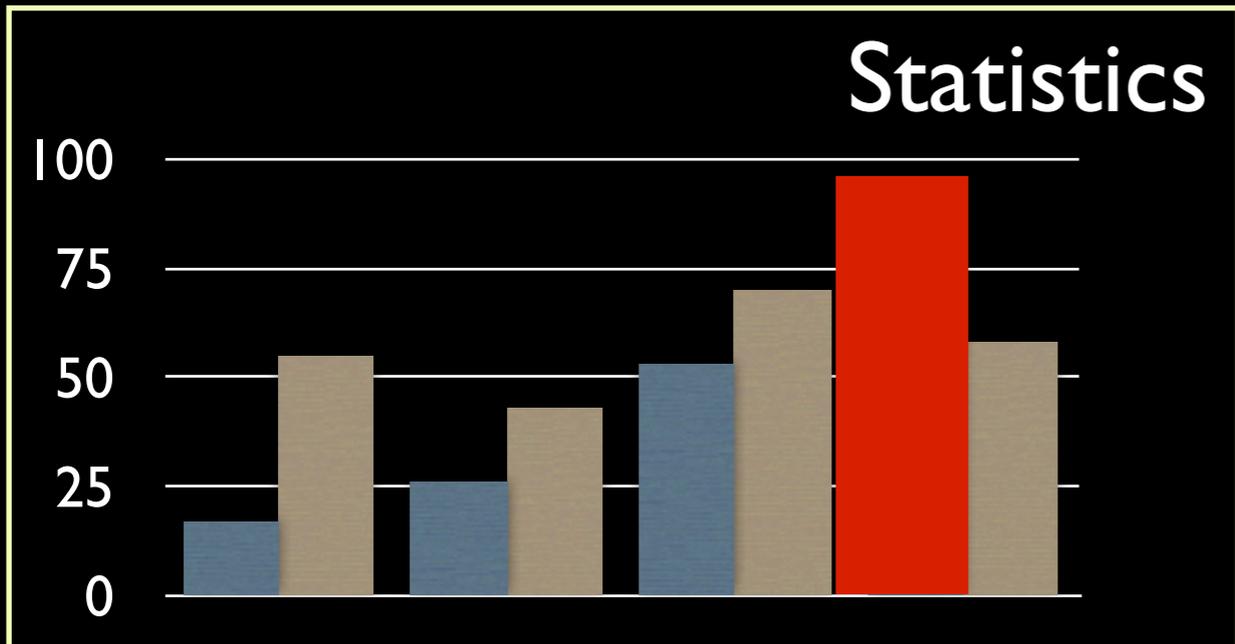
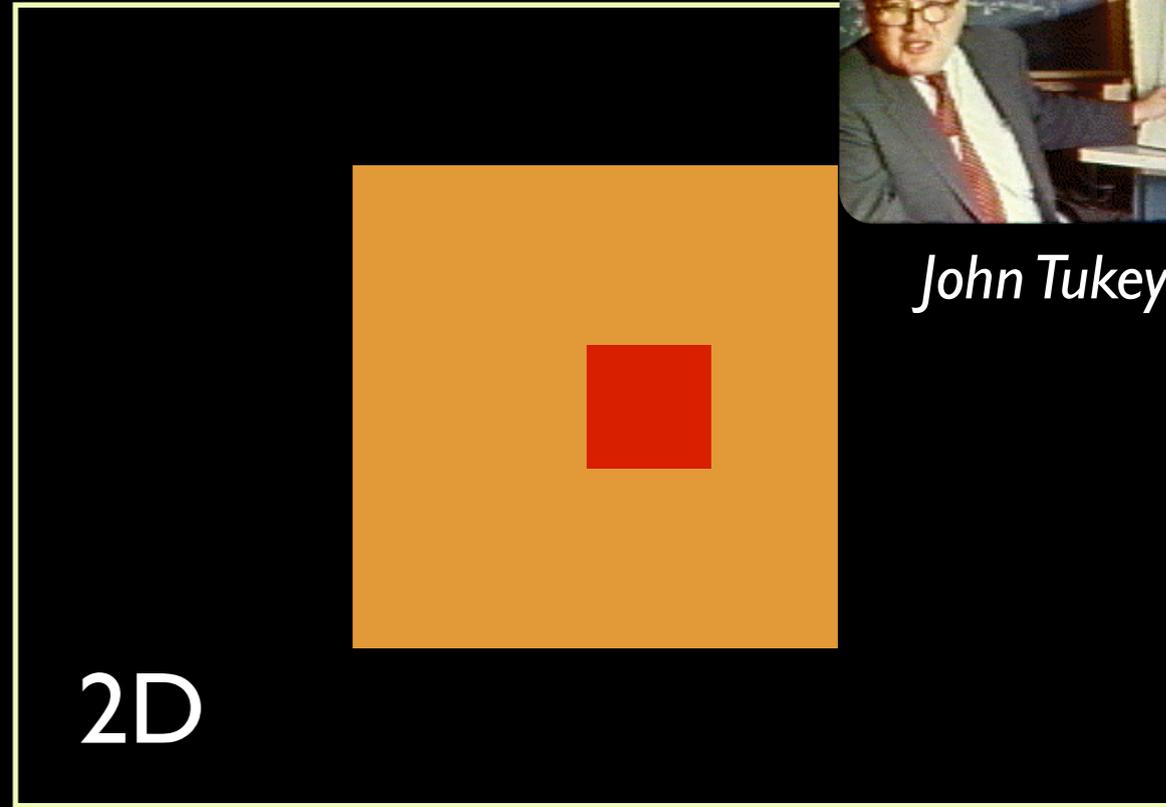
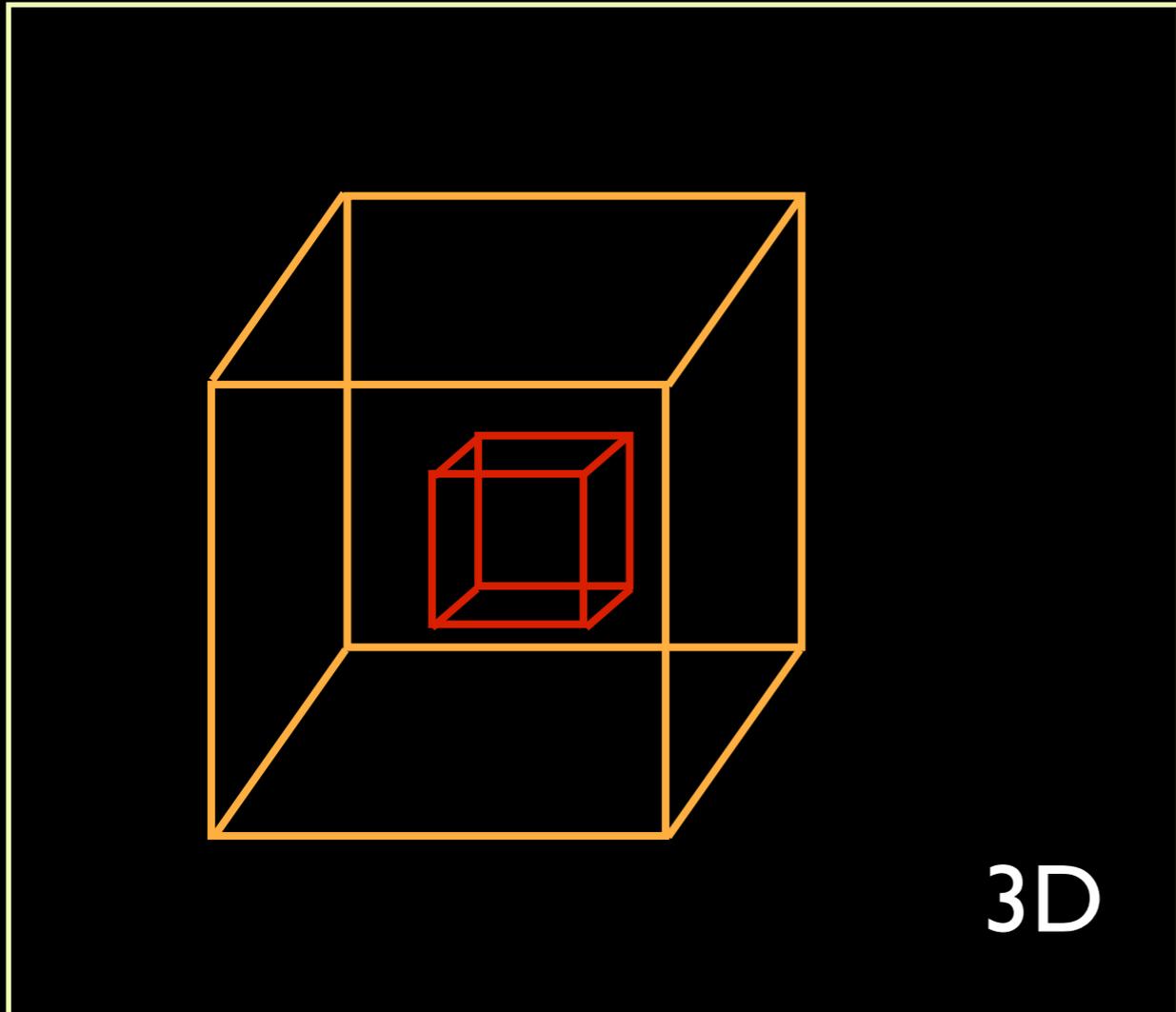
²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>



LINKED VIEWS OF HIGH-DIMENSIONAL DATA

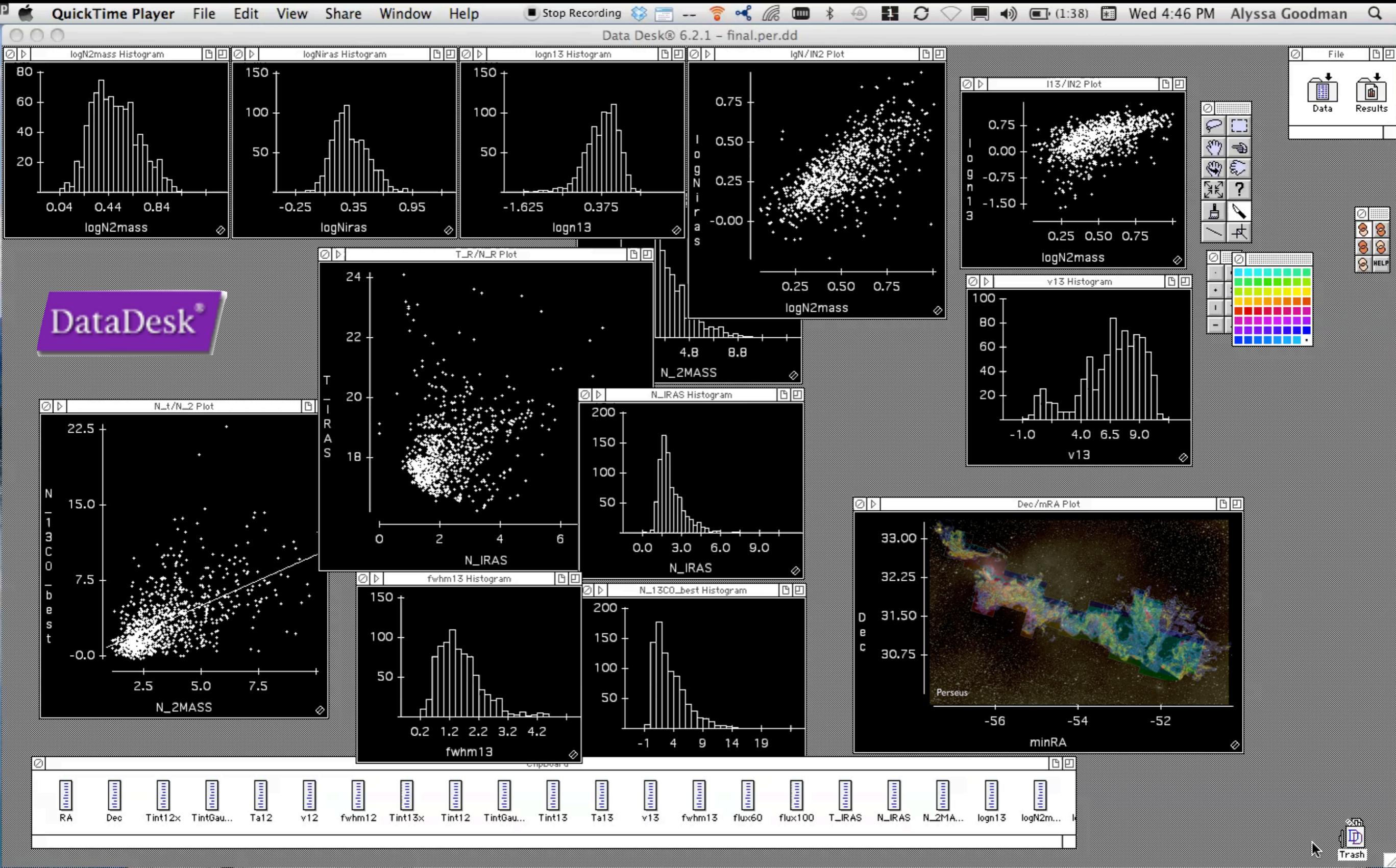


John Tukey

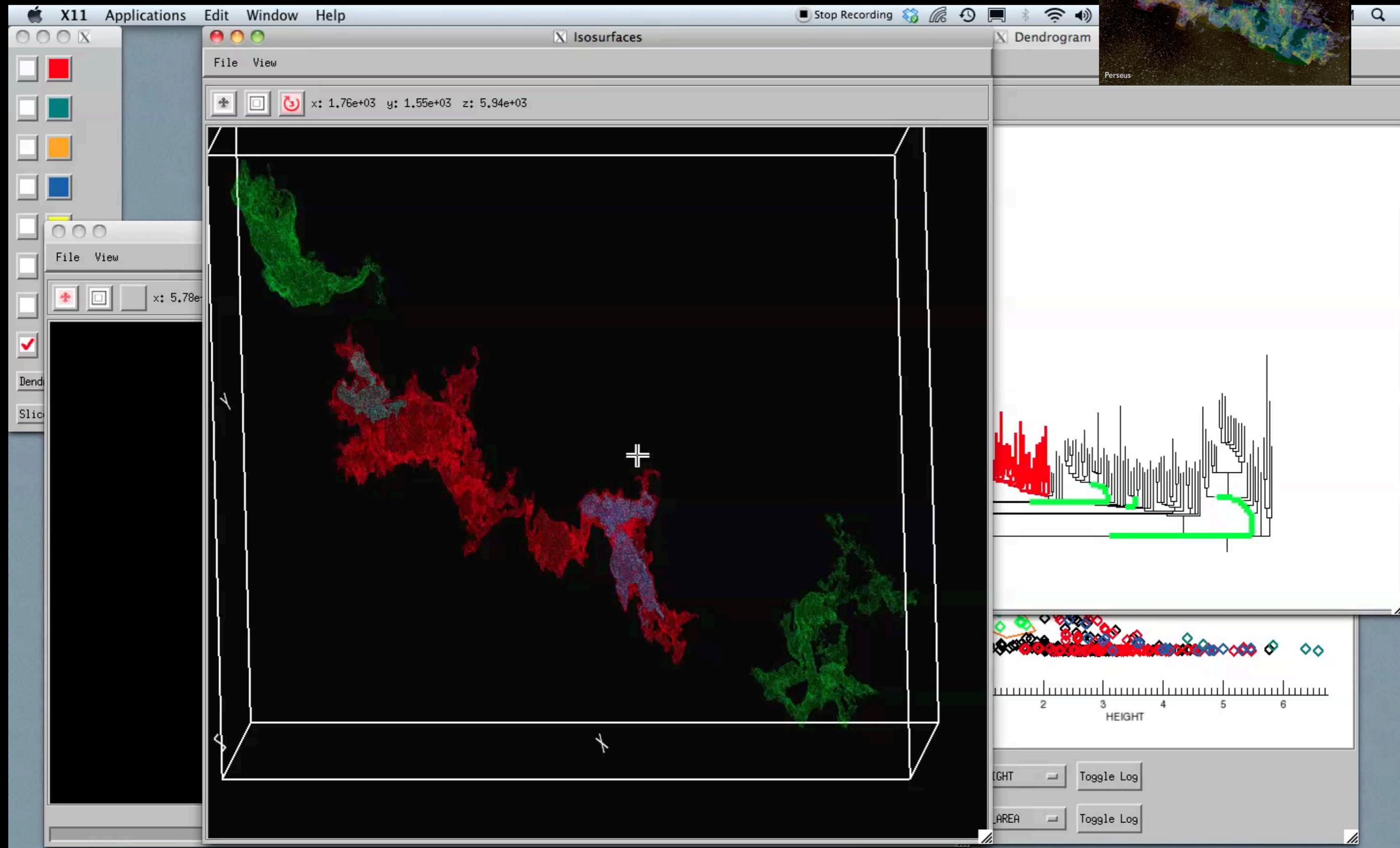
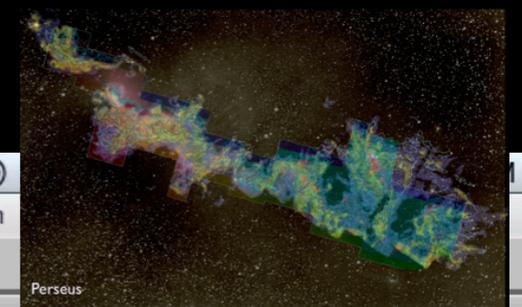


figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

DATADESK (EST. 1986)



LINKED VIEWS OF HIGH-DIMENSIONAL DATA



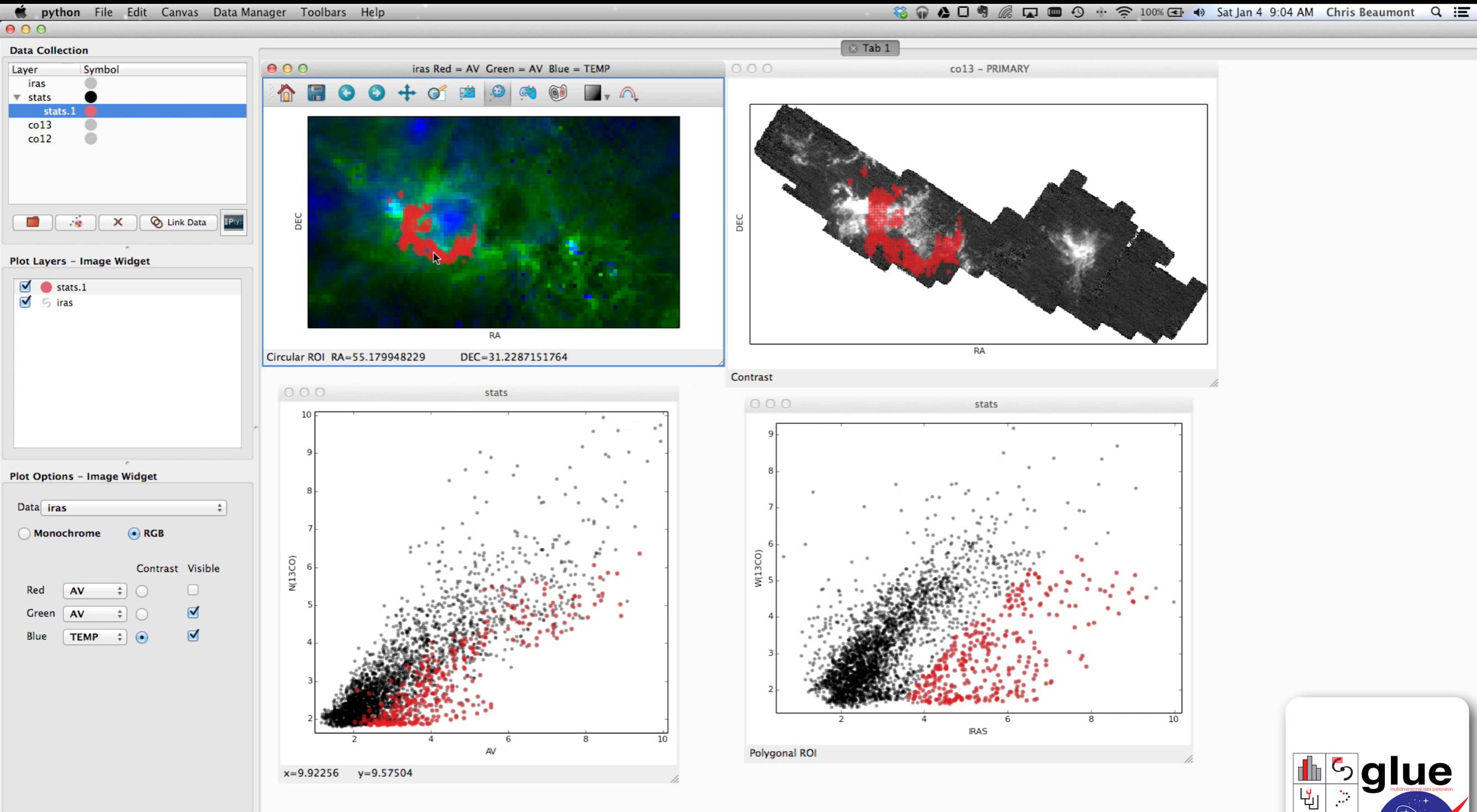
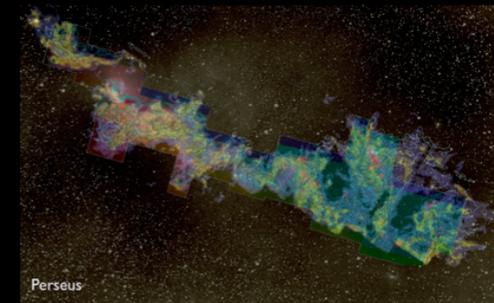
Video & implementation: Christopher Beaumont, Harvard→Counsyl;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

GREAT. BUT THAT WAS ALL FROM ONE DATA FILE.

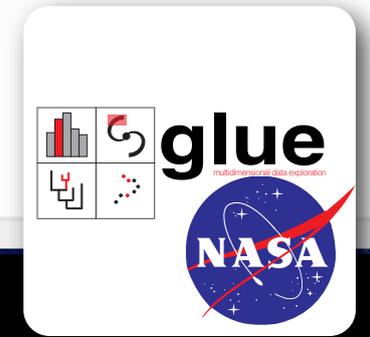
(AND IT WAS IN SOFTWARE THAT COSTS \$1000.)

LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)

GLUE



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



DIMENSIONS NEED NOT BE GEOMETRIC



Data Collection

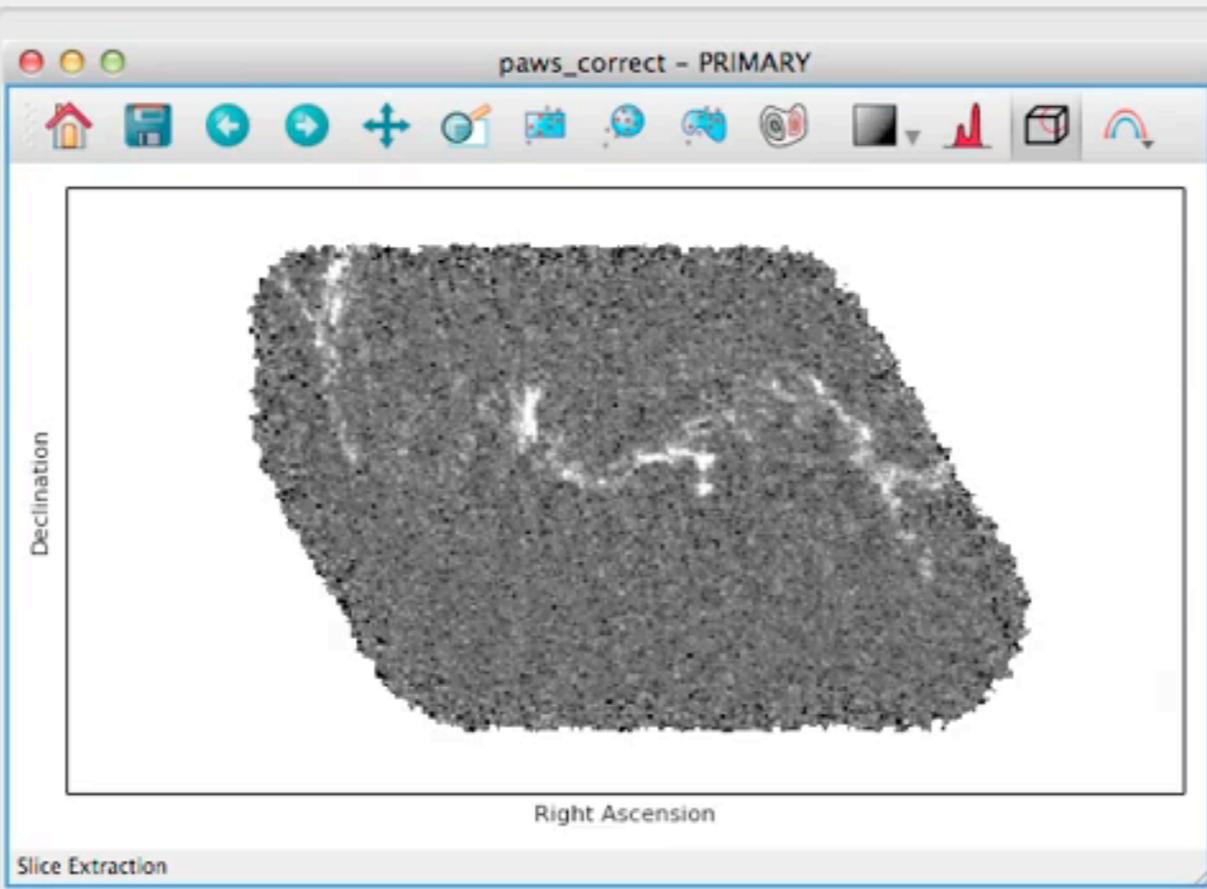
Data

- paws_correct

Subsets

Link Data

IPynb



Plot Layers - Image Widget

- paws_correct

Plot Options - Image Widget

Data: paws_correct

Monochrome RGB

Attribute: PRIMARY

Right Ascension: x

Declination: y

Veloc: slice

54

slice



video courtesy of Chris Beaumont, lead glue developer 2012-14

"NOW IN 3D!"

3D Scatter Plot

Video courtesy of lead glue developer, Tom Robitaille (showcasing work of glue team)

3D Dust Mapping

with Pan-STARRS 1

Query Map



Download & API



Read Paper



argonaut.skymaps.info

The Map

We present a 3D map of interstellar dust reddening. Using 800 million stars with Pan-STARRS 1 photometry, and matched 2MASS photometry for 200 million stars, the map traces dust reddening in three dimensions across three-quarters of the sky.

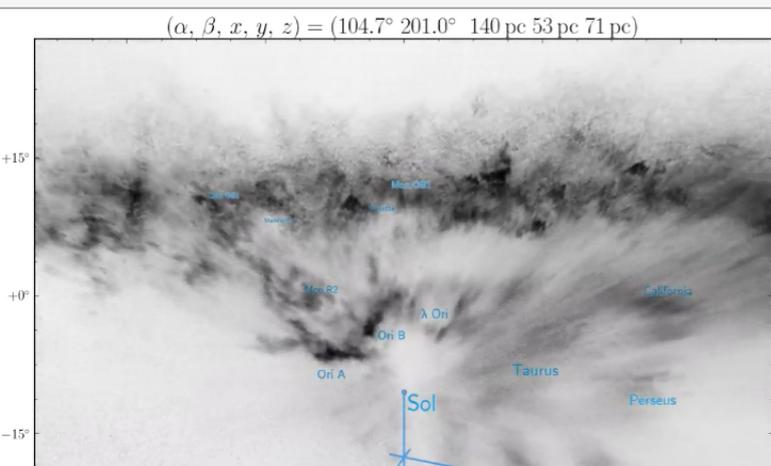
Volume Renderings

Volume renderings of the inferred dust opacity, moving the camera through the Galaxy on different trajectories.

Local Dust

Looking towards the Galactic anticenter.

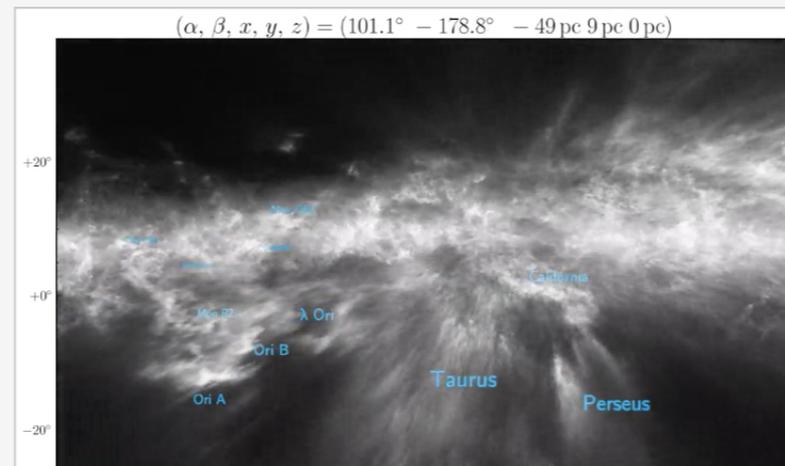
$(\alpha, \beta, x, y, z) = (104.7^\circ, 201.0^\circ, 140 \text{ pc}, 53 \text{ pc}, 71 \text{ pc})$



Orbiting the Sun

A 50-pc loop around the Sun.

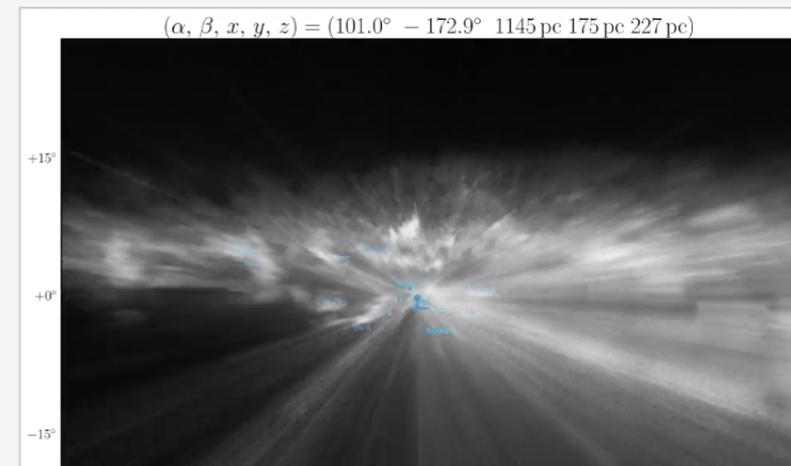
$(\alpha, \beta, x, y, z) = (101.1^\circ, -178.8^\circ, -49 \text{ pc}, 9 \text{ pc}, 0 \text{ pc})$



A Tour of the Galaxy

A several-kpc loop through and out of the Galactic plane.

$(\alpha, \beta, x, y, z) = (101.0^\circ, -172.9^\circ, 1145 \text{ pc}, 175 \text{ pc}, 227 \text{ pc})$



Gregory Green



Eddie Schlafly



Douglas Finkbeiner

COMING SOON: A "CFA MILKY WAY"

Battersby, Dame, Finkbeiner, Goodman, Green, Qian, Reid, Schlafly, Zucker, et al.



A screenshot of the Universe3D.org website. The page has a blue header with the site name "UNIVERSE3D.org" and a "Log in / create account" link. Below the header, there are navigation links for "Page", "Discussion", "Read", "View source", and "View history". A search bar is also present. The main content area is divided into several sections: "Navigation" with links like Home, 3D Viewers, Datasets, Images, Videos, Publications & Presentations; "More" with links like About Universe3D.org, Related Meetings, Contact, Help; "Toolbox" with links like What links here, Related changes, Special pages, Printable version, Permanent link; "Recently added Dataset" featuring the "Methanol MultiBeam Survey" with a description of its purpose and methodology; "Astronomy News" with two articles: "Searching for aliens who already know we're here" and "Life or illusion? Avoiding 'false positives' in the search for living worlds"; and "Survey Coverage of the Milky Way" which displays a series of maps showing the coverage of various surveys like GLIMPSE, MIPS GAL, HGPS, ATLAS GAL, HIGAL, and BGPS.

Video from WorldWide Telescope, based on work of Harvard Senior Matt Pasquini with Harvard Grad Student Catherine Zucker, using Rice et al. 2016 and Schlafly et al. 2014



yt project About Docs Community Develop Gallery Project Members Quick Links

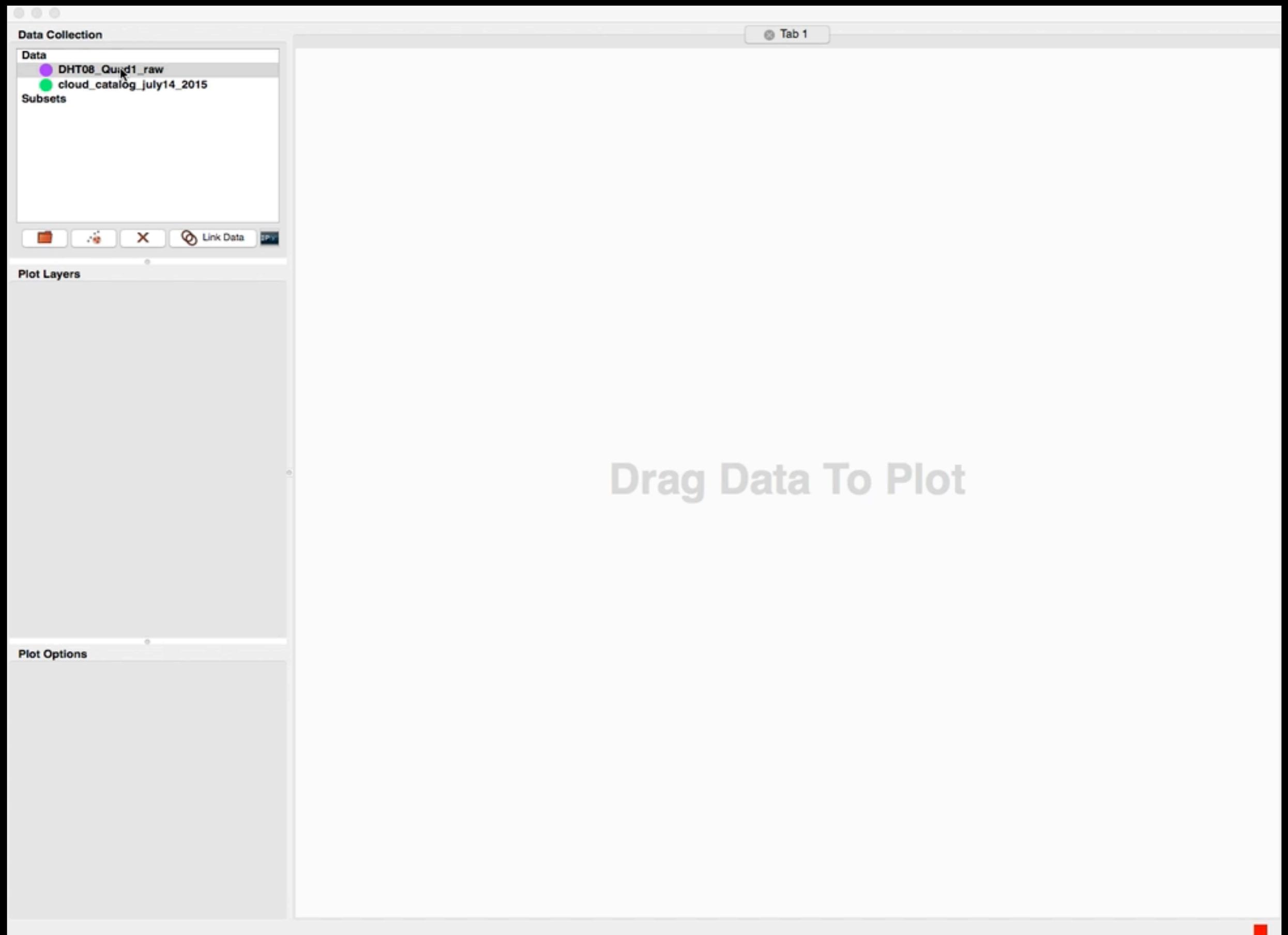
Quantitative Analysis and Visualization

yt is more than a visualization package: it is a tool to seamlessly handle simulation output files to make analysis simple. yt can easily knit together volumetric data to investigate phase-space distributions, averages, line integrals, streamline queries, region selection, halo finding, contour identification, surface extraction and more.

+yt, Blender, R, more...

Rectangular ROI x=27.8073 y=107.248

Video courtesy of Chris Beaumont



Video courtesy of Penny Qian (SAO predoc working on glue/Milky Way), Data from Dame et al. 2001; Rice et al. 2016

Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

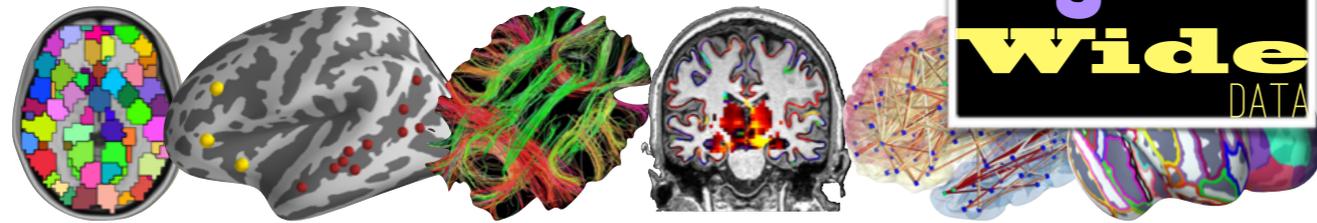
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

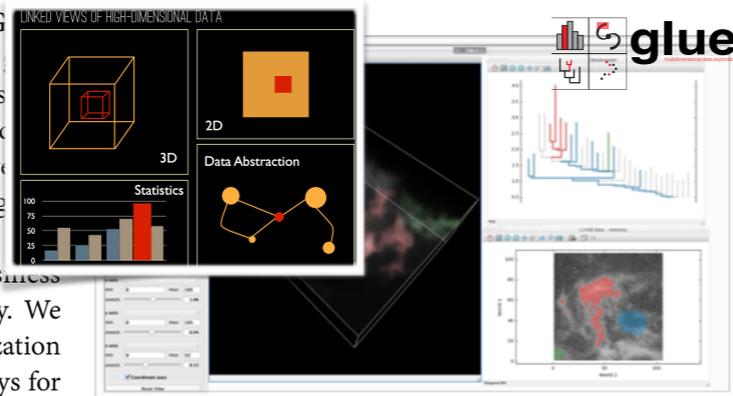
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow **multiple data sets** from diverse sources to be **analyzed in concert** (see #2). And, second, there is no way to draw **arbitrary surfaces** needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"
The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates *live* to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on **Astronomical Medicine**, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

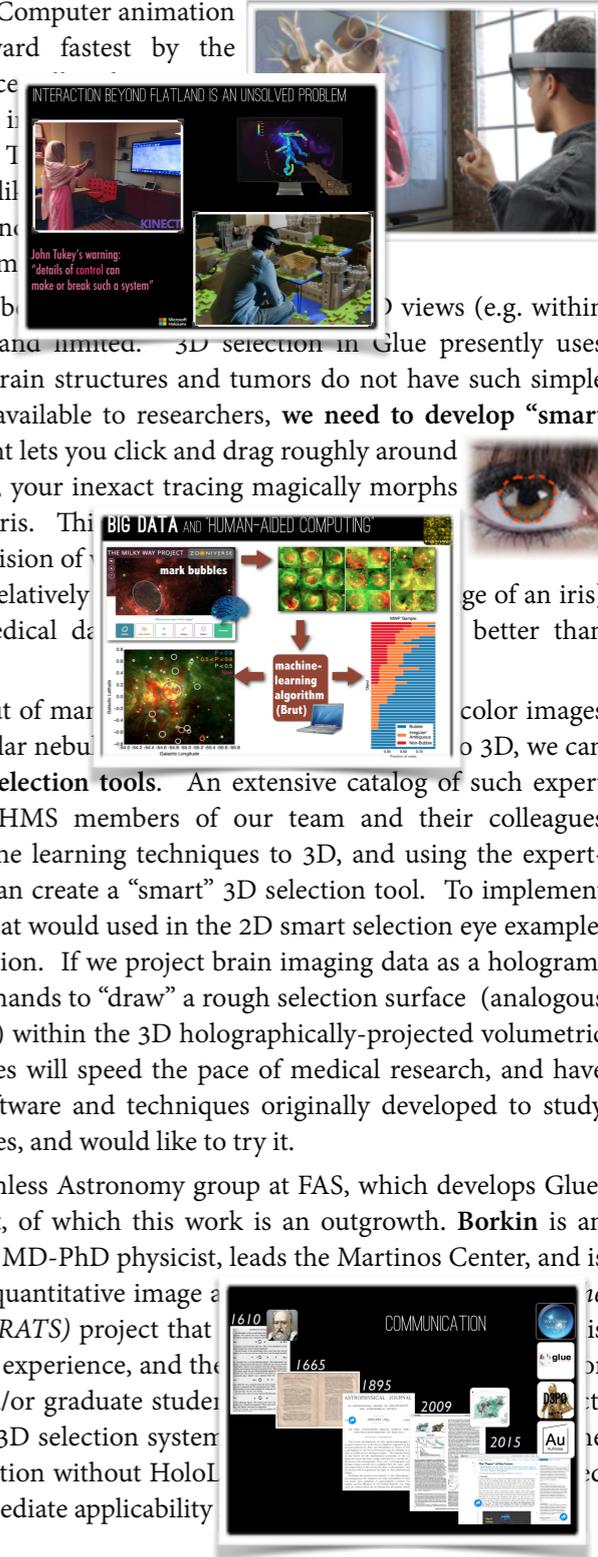
3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether a pixel belongs to the object) are solved for relatively quickly. In both 2D and 3D, the segmentation problem is solved for relatively simple features. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

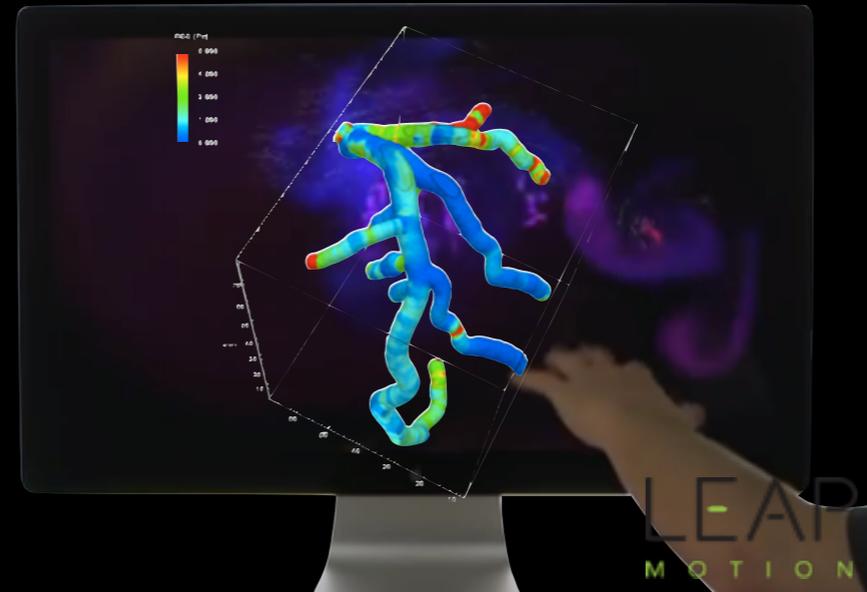
PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms of nebulae to train a machine learning algorithm to find similar nebulae. We can use experts' segmentations of brain imaging to train 3D selection tools. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that has been used by one of their labs to host a part-time postdoctoral fellow and/or graduate student, and to purchase a HoloLens. Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>



SELECTION IN 3D IS AN UNSOLVED PROBLEM



John Tukey's warning:
"details of control can
make or break such a system"



Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

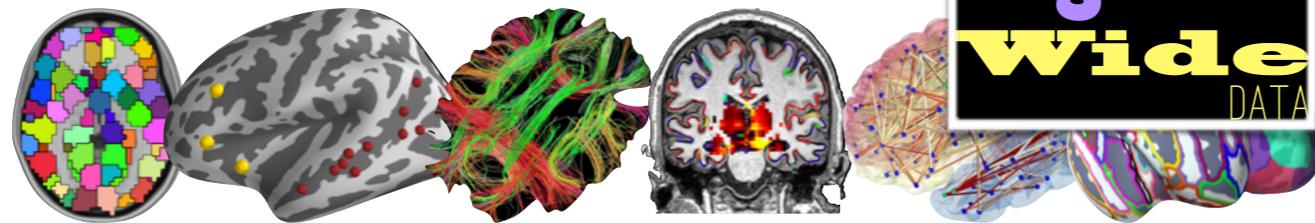
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

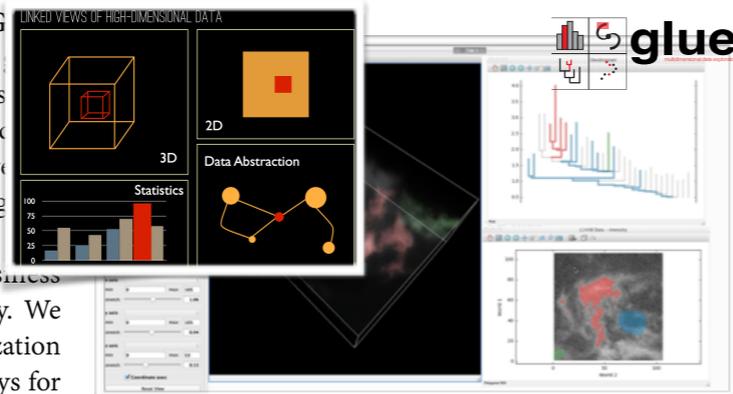
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow multiple data sets from diverse sources to be analyzed in concert (see #2). And, second, there is no way to draw arbitrary surfaces needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"
The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates live to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence and data analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on *Astronomical Medicine*, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

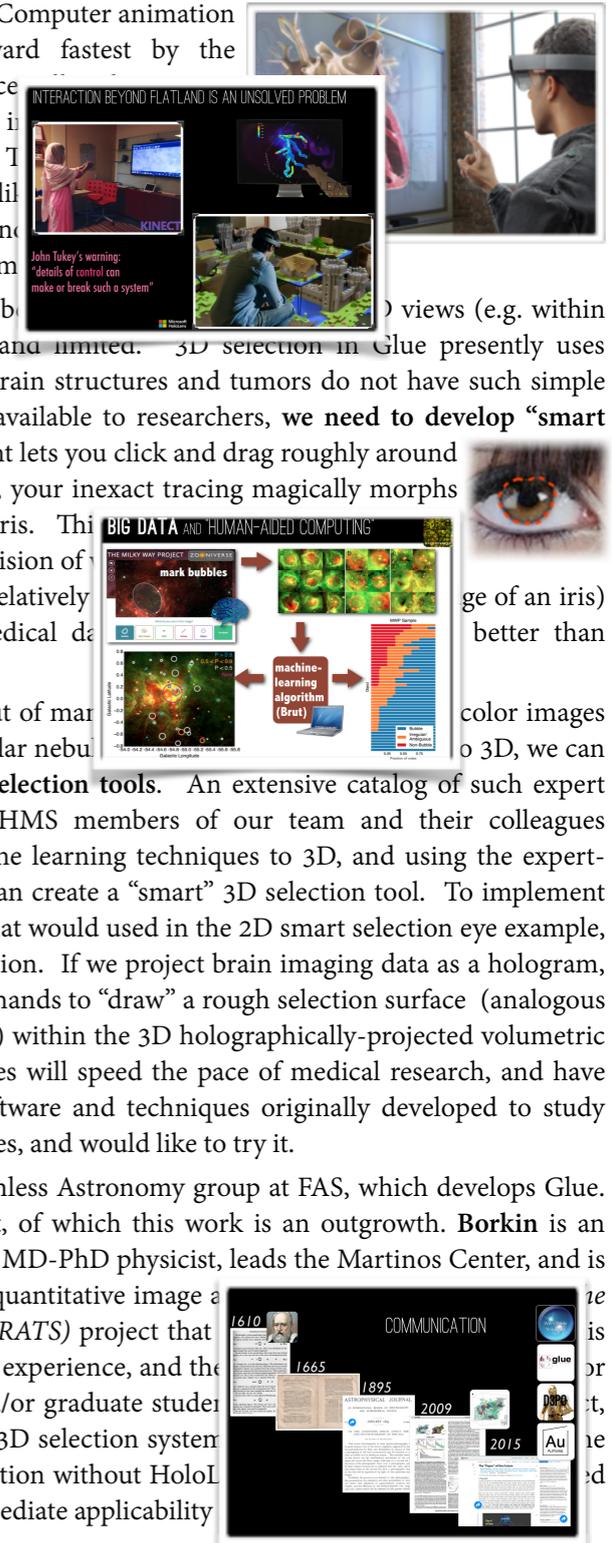
PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether a pixel belongs to the object) are solved for relatively quickly. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

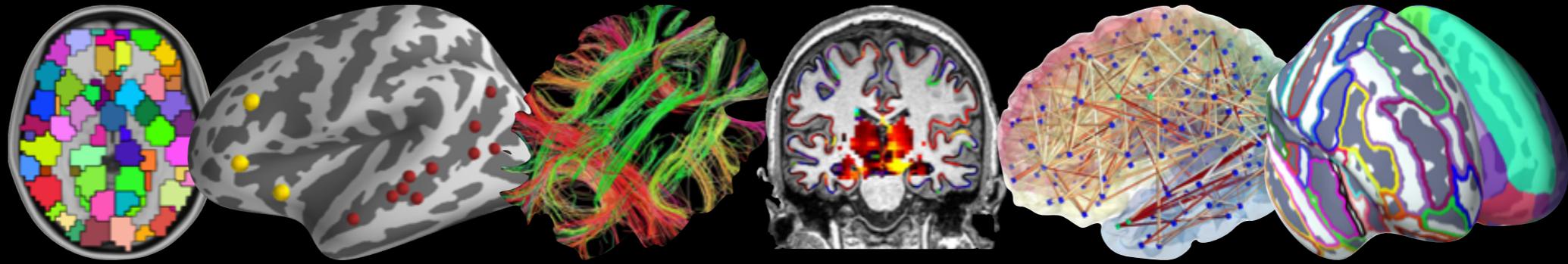
PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms to find similar nebulae. For example, we used the output of machine learning algorithms to find similar nebulae. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that has been used by one of their labs to host a part-time postdoctoral fellow and/or graduate student and to purchase a HoloLens. Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

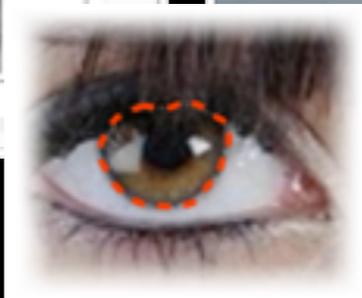
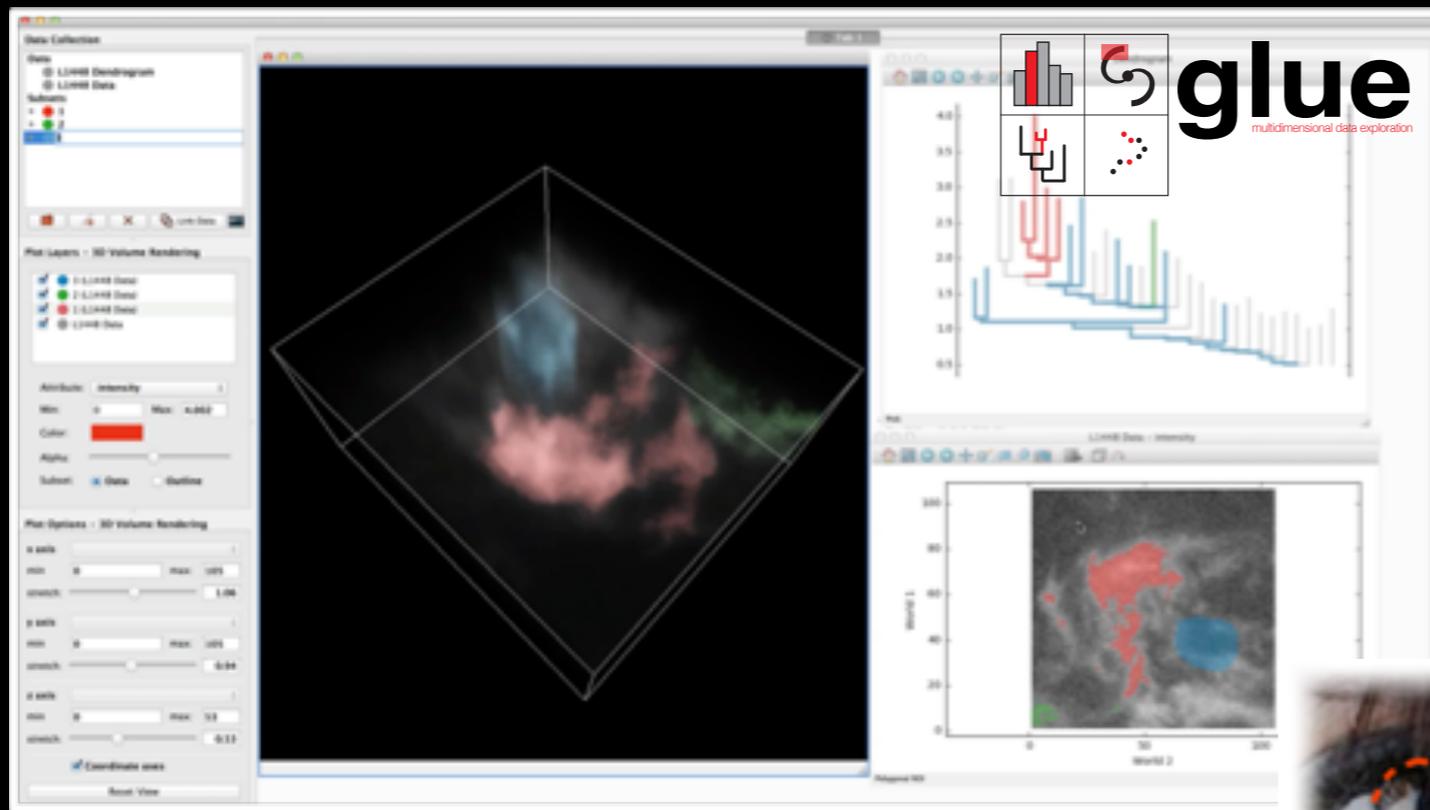
Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>





“Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives”



Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives

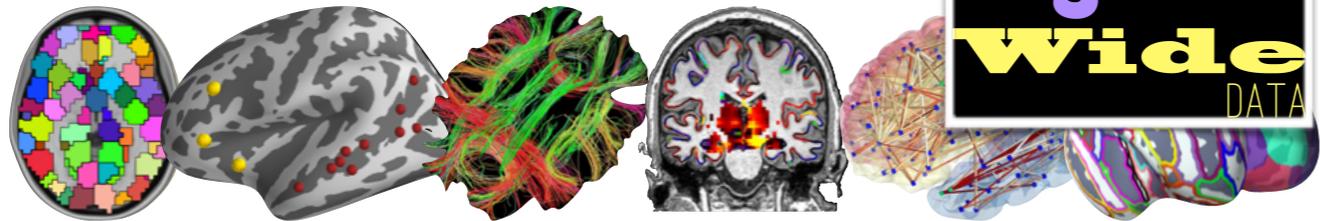
Alyssa A. Goodman Robert Wheeler Willson Professor of Applied Astronomy, FAS; Bruce Rosen Lawrence Lamson Robbins Professor of Radiology, HMS and Director of the Athinoula A. Martinos Center for Biomedical Imaging, MGH; Michelle A. Borkin Research Fellow in Surgery, HMS & Assistant Professor, College of Computer and Information Science, Northeastern University; Jayashree Kalpathy-Cramer Assistant Professor, HMS

In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and NIH-funded research on data visualization and brain imaging will be combined with new work to create a system that will allow clinicians to better understand high-dimensional images. Specifically, we will develop an unproven for "smart selection" of features in volumetric data, using new human-computer interaction techniques and the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) is critical in both research and clinical work. This process, called "segmentation," is essential in astronomy and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually defined by tracing out cross-sections of 3D image cubes on 2D "slice" images, a laborious and slow process fraught with inaccuracy. The new segmentation approach proposed will improve efficiency and reduce variability, saving lives.

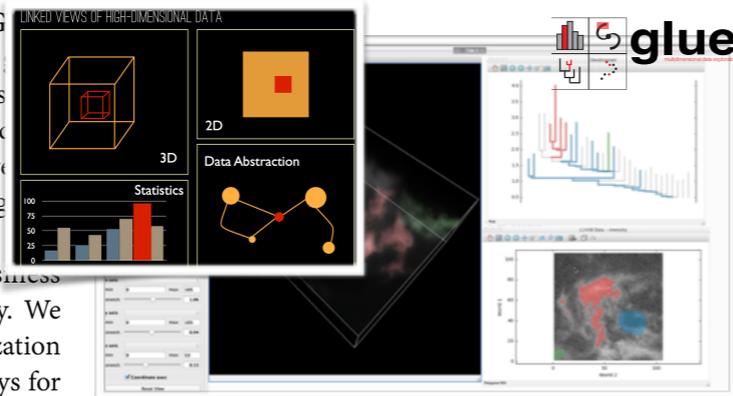
Three recent developments make the proposed research feasible now.

1. Advances in Brain Imaging and Associated Software. The capabilities of software packages for analyzing an ever-widening variety of neuroimaging modalities have grown tremendously over the past decade. Brain images below all use colored highlighting to indicate salient "features."



These features are identified algorithmically or manually, by computationally analyzing properties of the data themselves and/or by experts who markup and annotate the data. The software packages used to create these images typically suffer from two key limitations. First, they do not allow multiple data sets from diverse sources to be analyzed in concert (see #2). And, second, there is no way to draw arbitrary surfaces needed to make a "selection" in a 3D volumetric view. A standard computer mouse can be used to trace out a region of interest on a 2D image, but at present there is no analogous device to draw a surface to trace out a volume of interest in 3D (see #3).

2. Linked-view Visualization and the Advent of "Glue"
The human visual perceptual system is especially good at detecting change. So data visualization systems are offering so-called "linked views," where one on-screen map, chart or graph data set updates live to reflect selections made in another. These linked views serve as insight engines. Linked-views of tabular and mapping data have become popular recently, especially in business intelligence and data analytics, as data sets grow in size and diversity. We have recently developed the world's first visualization system that allows linking of data sets and displays for



¹See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

images and three-dimensional data formats. This system, called "Glue," is supported by NASA because it offers astronomers a way to analyze and make the best use of the three-dimensional data that comes from so-called spectroscopic data cubes. As in our earlier collaborative work on *Astronomical Medicine*, Glue is just as useful for analyzing three-dimensional medical image cubes, such as MRI and CT scans, as it is for astronomy. As illustrated in screenshot above, Glue propagates selections made in one data view, live, across all views, leveraging the human ability to see change, revealing hidden meaning in data. (Short demo videos: [2D](#), [3D](#))

3. Advances in Computer Gaming and Interaction Devices Computer animation and interaction techniques today are propelled forward fastest by the tremendous profitability of games. Innovative new devices allow users to interact with virtual environments in more and more immersive ways. As computers get more powerful and as profits grow, the opportunity on the horizon is the potential of devices like HoloLens to blend the real world, such as a person's real hands, with virtual projections, like the image of a human heart shown in the image above.

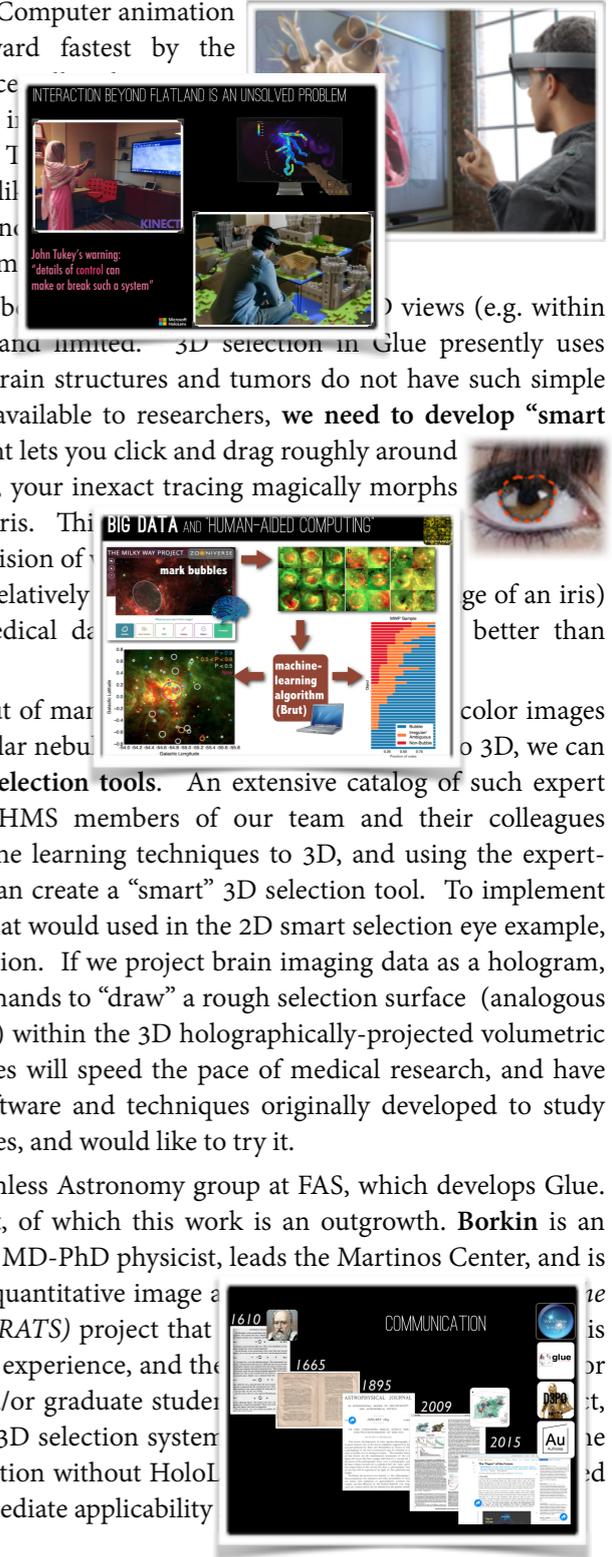
PROBLEM TO BE SOLVED Glue allows salient regions to be selected in 2D views (e.g. within maps or graphs), but selection in 3D views is onerous and limited. 3D selection in Glue presently uses combinations of simple shapes (e.g. spheres, prisms). But, brain structures and tumors do not have such simple shapes. To understand the tremendous wealth of 3D data available to researchers, we need to develop "smart selection" in 3D. Smart selection applied to the 2D eye at right lets you click and drag roughly around the iris' edge, and then, when you release the mouse button, your inexact tracing magically morphs into an excellent outline (red dashed line) of the iris. This "selection" (outlining the object) and "segmentation" (the decision of whether to include the object) are solved for relatively quickly. For fuzzy features, though, in both astronomical and medical data, machine learning algorithms alone at defining boundaries of salient features.

PROPOSED SOLUTION Our recent work² used the output of machine learning algorithms to find similar nebulae. For example, we used the output of machine learning algorithms to find similar nebulae. An extensive catalog of such expert segmentation has already been assembled by the MGH/HMS members of our team and their colleagues (braintumorsegmentation.org). By adapting our 2D machine learning techniques to 3D, and using the expert-created segmentations as training for those algorithms, we can create a "smart" 3D selection tool. To implement that tool in Glue, we also need the equivalent of the mouse that would be used in the 2D smart selection eye example, above. We believe that the HoloLens is likely to offer a solution. If we project brain imaging data as a hologram, using HoloLens, a researcher should be able to use their real hands to "draw" a rough selection surface (analogous to the dashed red line around the iris in the 2D image above) within the 3D holographically-projected volumetric data. The ability to make smart selections within 3D images will speed the pace of medical research, and have immediate clinical applications. We love the idea that software and techniques originally developed to study arcane questions about our Universe could ultimately save lives, and would like to try it.

STAFFING, BUDGET, SUCCESS Goodman leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. Borkin is an expert on human-computer interaction. Rosen, trained as an MD-PhD physicist, leads the Martinos Center, and is an expert in neuroimaging. Kalpathy-Cramer is an expert on quantitative image analysis and leads the *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that has been used by one of their labs to host a part-time postdoctoral fellow and/or graduate student and to purchase a HoloLens. Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

Our goal is a working smart 3D selection system. The success of the unprecedented nature of what we propose, a smart 3D selection without HoloLens, would be successes on their own, with immediate applicability.

²Beaumont, Goodman, et al. *Brut: Automatic bubble classifier*, <http://adsabs.harvard.edu/abs/2014ascl.soft07016B>



1610



SIDEREUS NUNCIUS

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

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 star was 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 ecliptic.

On the fifth, the sky was cloudy.

On the sixth, only two stars appeared flanking Jupiter

East * ○ *

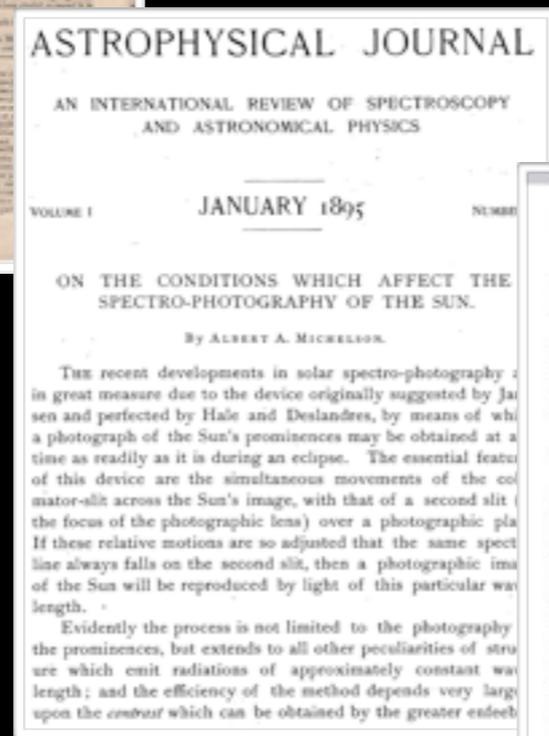
in the adjoining figure. The eastern one was 2 minutes from the next western one 3 minutes from Jupiter. They were on the same straight line with Jupiter and equal in magnitude.

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

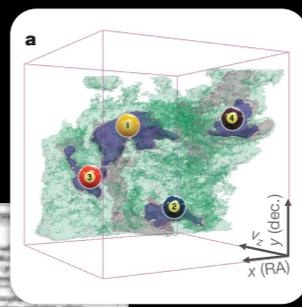
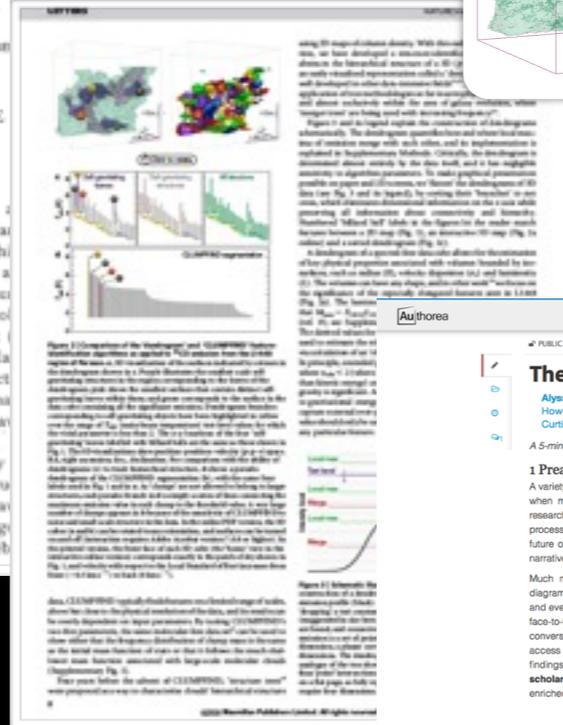
1665



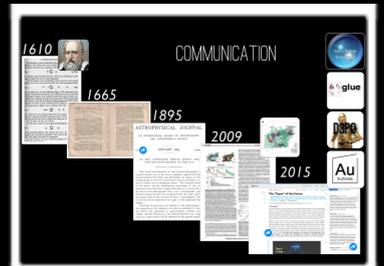
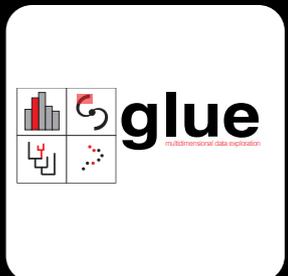
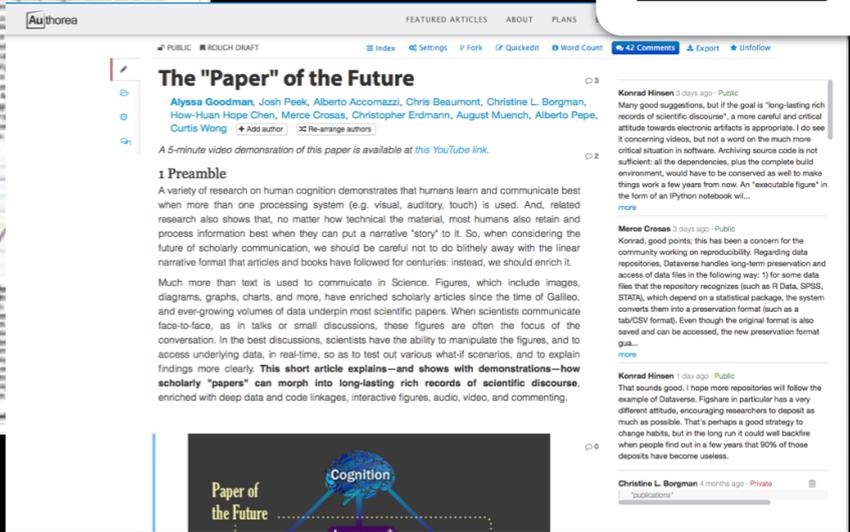
1895

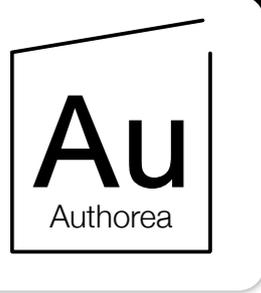


2009



2015





PUBLIC ROUGH DRAFT

Index Settings Fork Quickedit Word Count 42 Comments Export Unfollow

The "Paper" of the Future

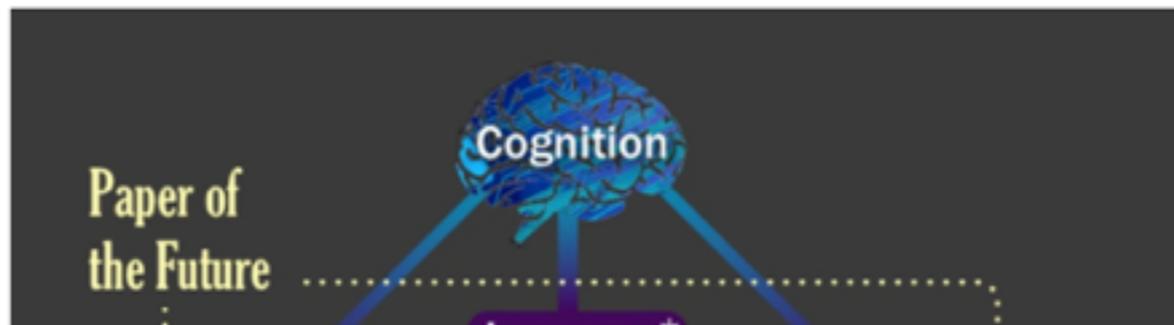
Alyssa Goodman, Josh Peek, Alberto Accomazzi, Chris Beaumont, Christine L. Borgman, How-Huan Hope Chen, Merce Crosas, Christopher Erdmann, August Muench, Alberto Pepe, Curtis Wong [+ Add author](#) [✕ Re-arrange authors](#)

A 5-minute video demonstration 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 communicate 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.**



3

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 will...

[more](#)

2

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](#)

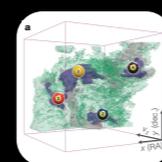
0

Konrad Hinsen 1 day ago · Public

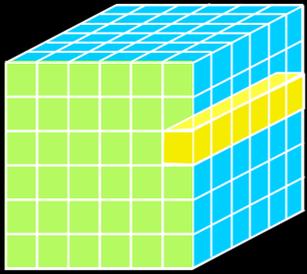
That sounds good. I hope more repositories will follow the example of Dataverse. Figshare in particular has a very different attitude, encouraging researchers to deposit as much as possible. That's perhaps a good strategy to change habits, but in the long run it could well backfire when people find out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

Christine L. Borgman 4 months ago · Private

"publications"



[video, demos]



"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"

LINKED VIEWS OF HIGH-DIMENSIONAL DATA

Big DATA versus wide DATA



BIG DATA AND "HUMAN-AIDED COMPUTING"

INTERACTION BEYOND FLATLAND IS AN UNSOLVED PROBLEM



John Tukey's warning: "details of control can make or break such a system"

Microsoft HoloLens



1665

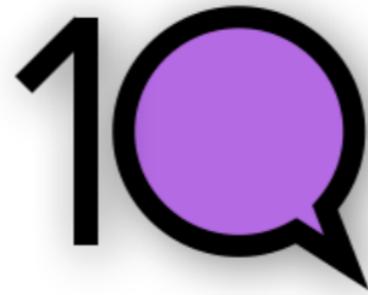
1895

2009

2015

COMMUNICATION





TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION



Ten Questions to Ask When Creating a Scientific Visualization

Here are 10 good questions to ask when you create a scientific visualization:

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. **Feature recognition** | Is feature and/or pattern recognition, a goal?
4. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
5. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
6. **Categories & Clustering** | Are there natural, or imposed, categories within the data? Are you interested in clustering?
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?

Astronomy 215hf (T,Th at 10 AM starting 3/22 for 2 weeks)