MENU

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:

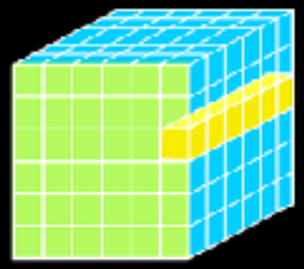
1. **Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
2. **Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
3. **Feature recognition** | Is feature and/or pattern recognition, a goal?
4. **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

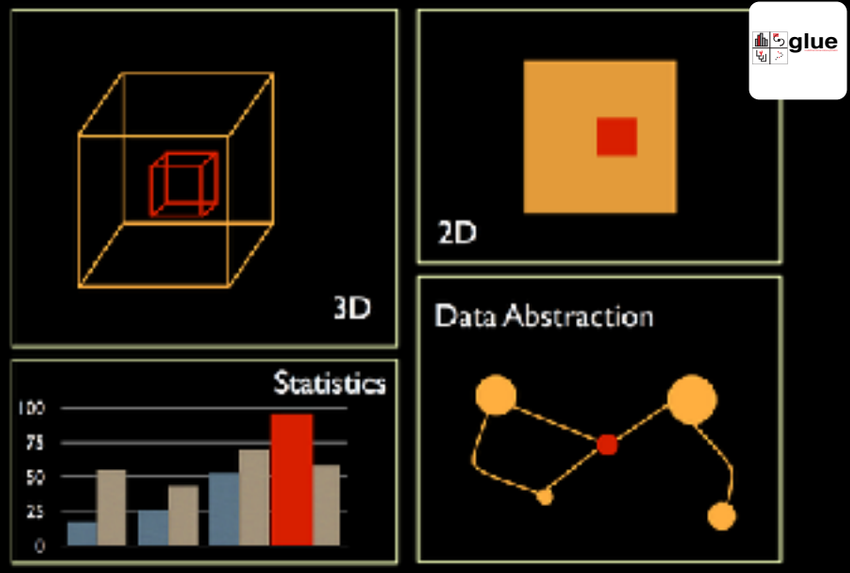
*with Pizz



"DATA. DIMENSIONS. DISPLAY"

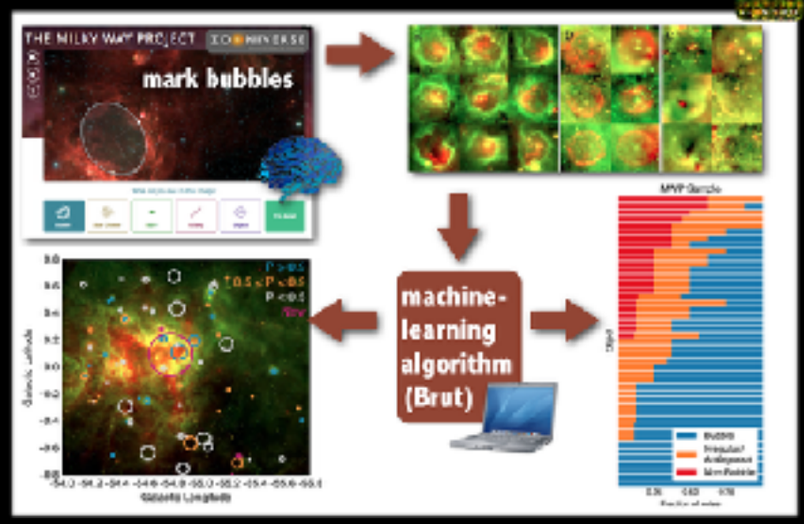
1D: Columns = "Spectra", "SEDs" or "Time Series"
2D: Faces or Slices = "Images"
3D: Volumes = "3D Renderings", "2D Movies"
4D: Time Series of Volumes = "3D Movies"

LINKED VIEWS OF HIGH-DIMENSIONAL DATA

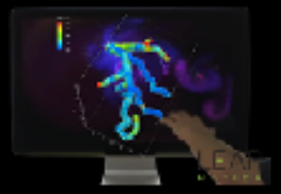


Big DATA versus Wide DATA

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"

COMMUNICATION

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 2016

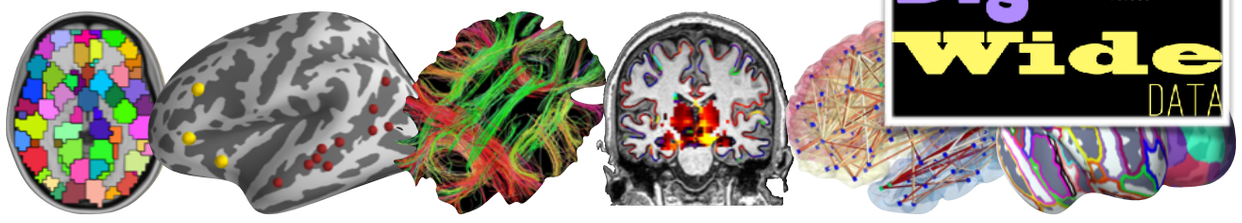
"Leveraging Human Pattern-Recognition Abilities with Machine Learning to Save Lives"

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 al
clinicians to better understand high-dimensional images. Specifically, we will develop an unp
for "smart selection" of features in volumetric data, using new human-computer interaction d
the gaming industry to train and adjust machine-learning algorithms.

BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in
critical in both research and clinical work. This process, called "segmentation," is essential in as
and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def
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.

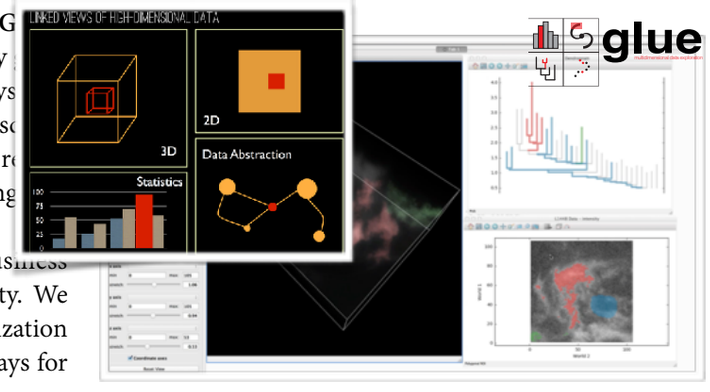
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft
ever-widening variety of neuroimaging modalities have grown tremendously o
brain images below all use colored highlighting to indicate salient "features."



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 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 "G**
The human visual perceptual system is especially
at detecting change. So data visualization sys
offering so-called "linked views," where one on-s
map, chart or graph data set updates *live* to r
selections made in another serve as insight eng
Linked-views of tabular and mapping data
become popular recently, especially in busi
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

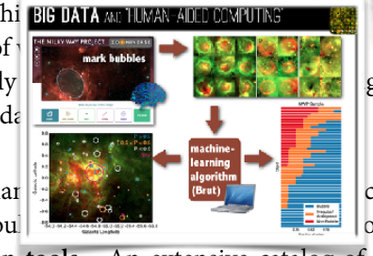


[1]See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

Glue," is supported by NASA because it offers
dimensional data that comes from so-called
Astronomical Medicine, Glue is just as useful
RI and CT scans, as it is for astronomy. As
e in one data view, live, across all views,
ng 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 device
eract with virtual environments in more and more i
nputers get more powerful and as profits grow. T
portunity on the horizon is the potential of devices lil
loLens to blend the real world, such as a person's real han
jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b
s 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. Thi
"selection" (outlining the object) and "segmentation" (the decision of
In both 2D and 3D, the segmentation problem is solved for relatively
For fuzzy features, though, in both astronomical and medical da
algorithms alone at defining boundaries of salient features.



PROPOSED SOLUTION Our recent work[2] used the output of mar
of nebulae to train a machine learning algorithm to find similar nebu
color images
o 3D, 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a
Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) project that
project. **Goodman** and **Rosen** both have extensive leadership experience, and the
one of their labs to host a part-time postdoctoral fellow and/or graduate stude
and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste
unprecedented nature of what we propose, a smart 3D selection without HoloL
3D selection tool, would be successes on their own, with immediate applicability

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

Pattern Recognition
Creativity

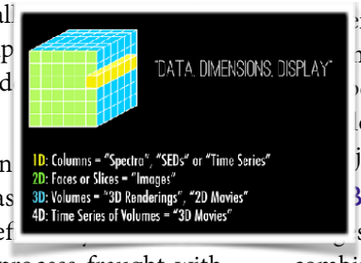
Calclulations
Endurance

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 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*

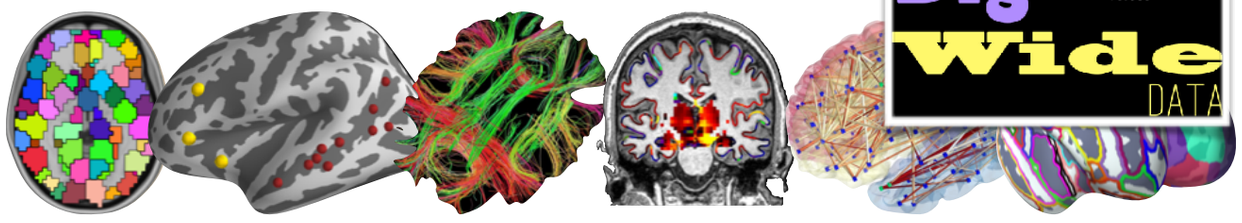
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

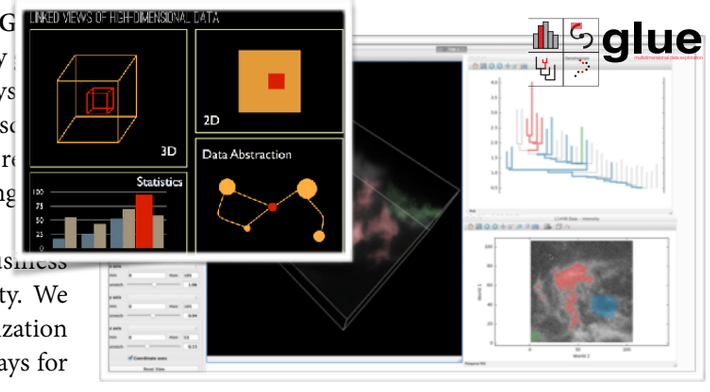
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

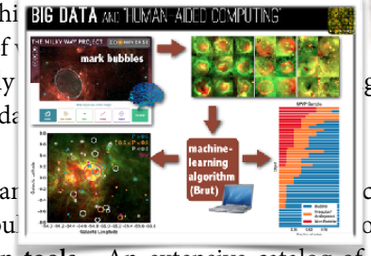
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.

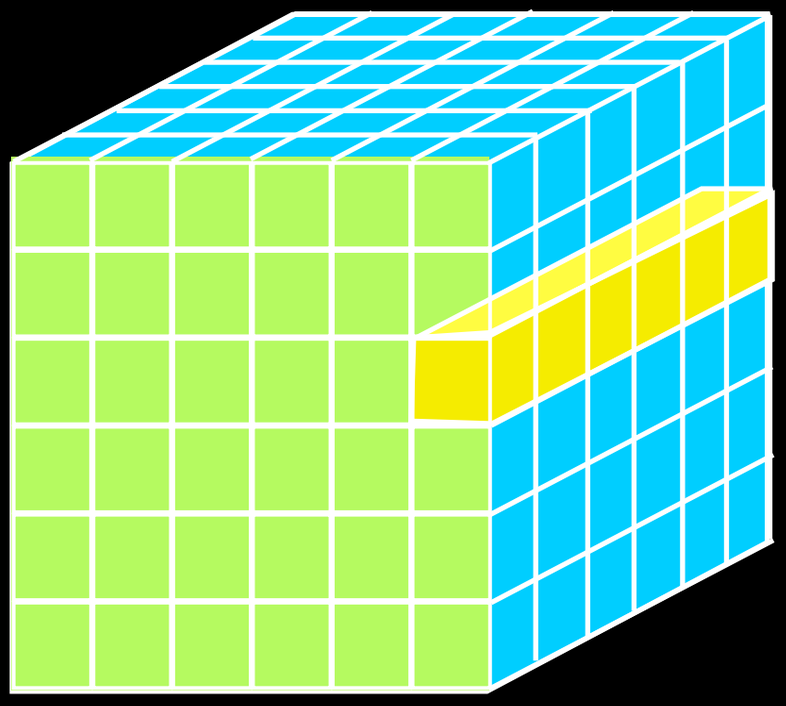


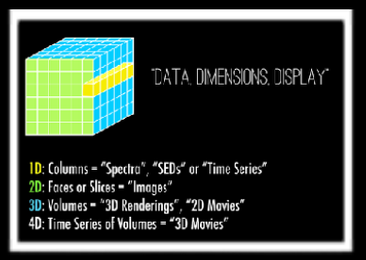
PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, we can **use experts' segmentations of brain imaging to train 3D selection tools**. An extensive catalog of such expert segmentation has already been assembled 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that project. **Goodman** and **Rosen** both have extensive leadership experience, and the one of their labs to host a part-time postdoctoral fellow and/or graduate stude and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste unprecedented nature of what we propose, a smart 3D selection without Holo 3D selection tool, would be successes on their own, with immediate applicability

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

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





"Data, Dimensions, Display"

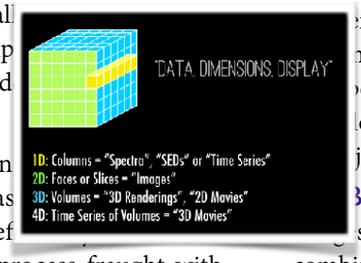
1D: Columns = "Spectra", "SEDs" or "Time Series"
2D: Faces or Slices = "Images"
3D: Volumes = "3D Renderings", "2D Movies"
4D: Time Series of Volumes = "3D Movies"

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 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*

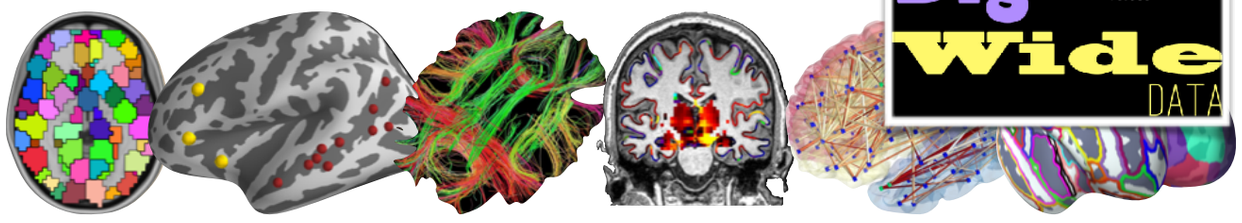
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

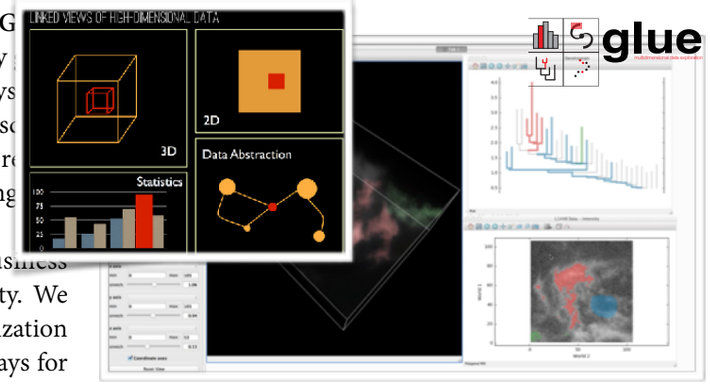
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

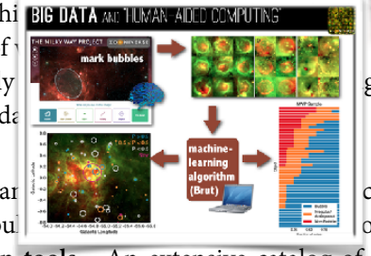
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.



PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that project. **Goodman** and **Rosen** both have extensive leadership experience, and the one of their labs to host a part-time postdoctoral fellow and/or graduate stude and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste unprecedented nature of what we propose, a smart 3D selection without Holo 3D selection tool, would be successes on their own, with immediate applicability

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

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

Big DATA
versus
Wide DATA

Explore | Guided Tours | Search | View | Settings | Install Windows Client

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

ZOONIVERSE
REAL SCIENCE ONLINE

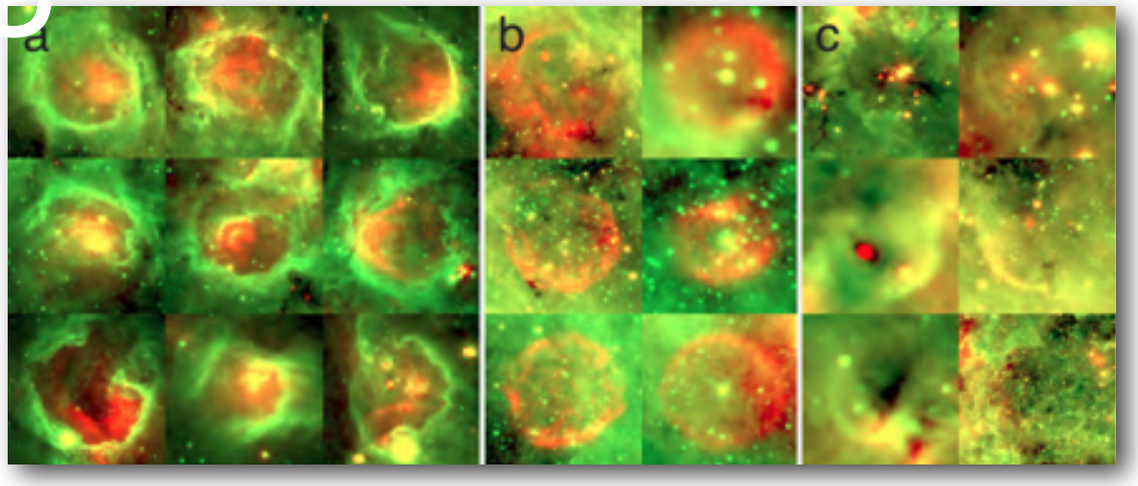
Look At: Sky
Imagery: Digitized Sky Survey (Color)
Image Crossfade

Tracking: GLIMPSE/MIPSGAL
1 of 3

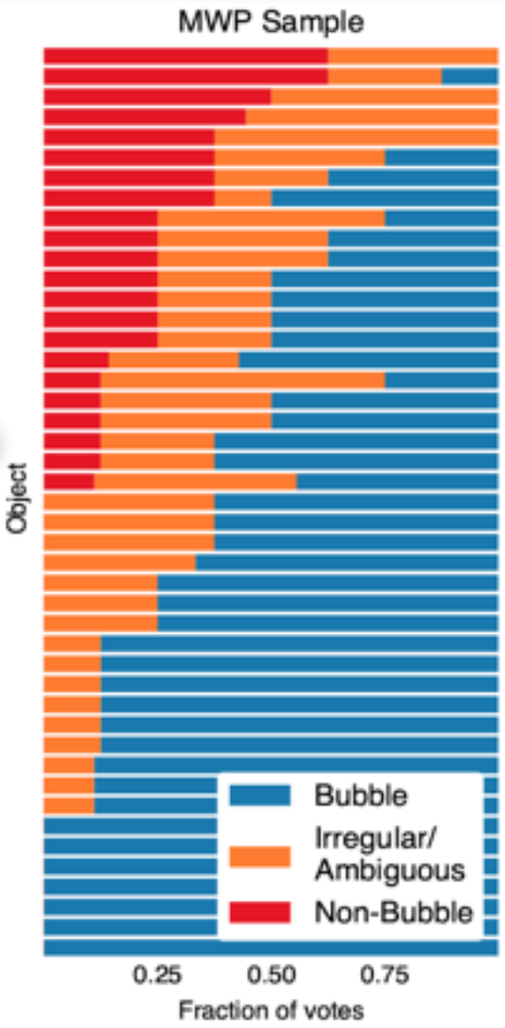
N
Scorpius
RA: 17h28m14s

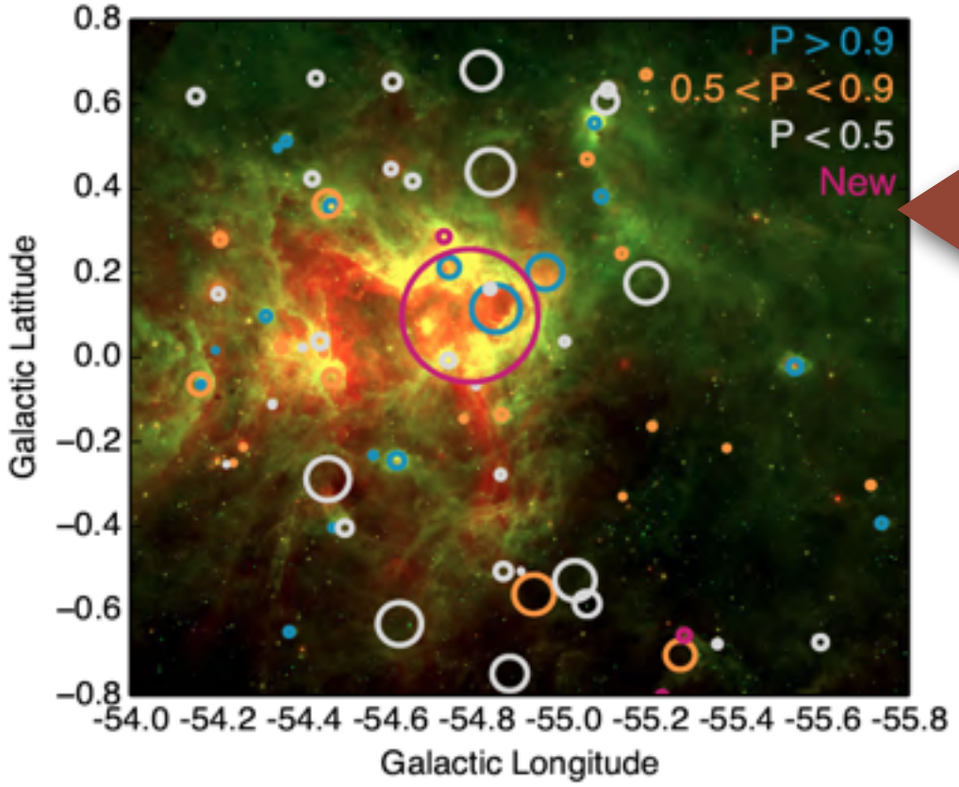
Pismis 24 and | NGC6334 | NGC6357 | NGC6374 | NGC6383 | NGC6396 | NGC6404 | Lesath | Shaula | HR6397 | HR6405

BIG DATA and "Human-Aided Computing"



machine-learning algorith





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*

Brut Based on "Random Forest*" (Supervised Learning)

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

The Astrophysical Journal Supplement Series, 214:3 (18pp), 2014 September doi:10.1088/0067-0049/214/1/3
© 2014. The American Astronomical Society. All rights reserved. Printed in the U.S.A.

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

Christopher N. Beaumont[1,2], Alyssa A. Goodman[2], Sarah Kendrew[3], Jonathan P. Williams[1], and Robert Simpson[3]
[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

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.

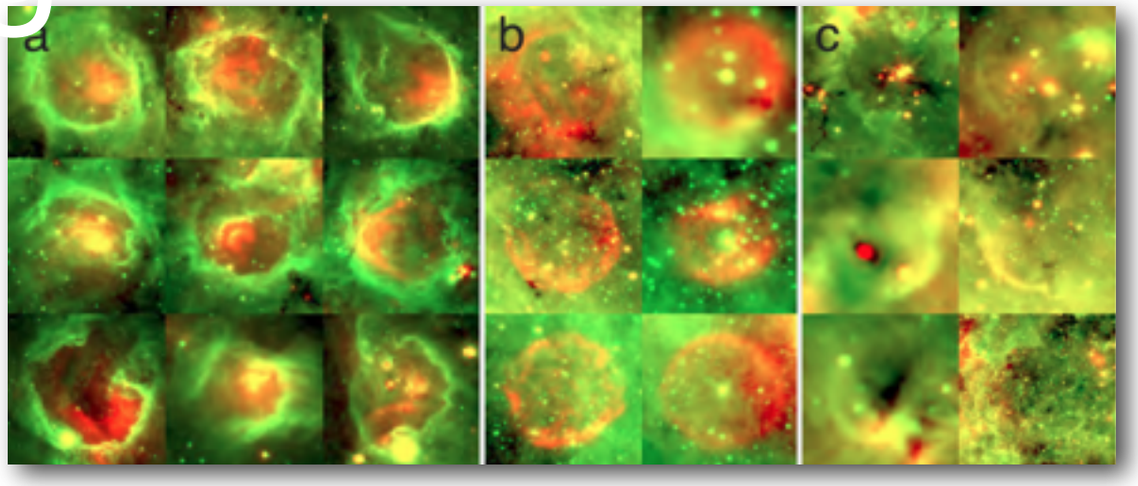
Website: https://github.com/ChrisBeaumont/brut
Appears in: http://adsabs.harvard.edu/abs/2014arXiv1406.2692B

Bibcode: 2014ascl.soft07016B

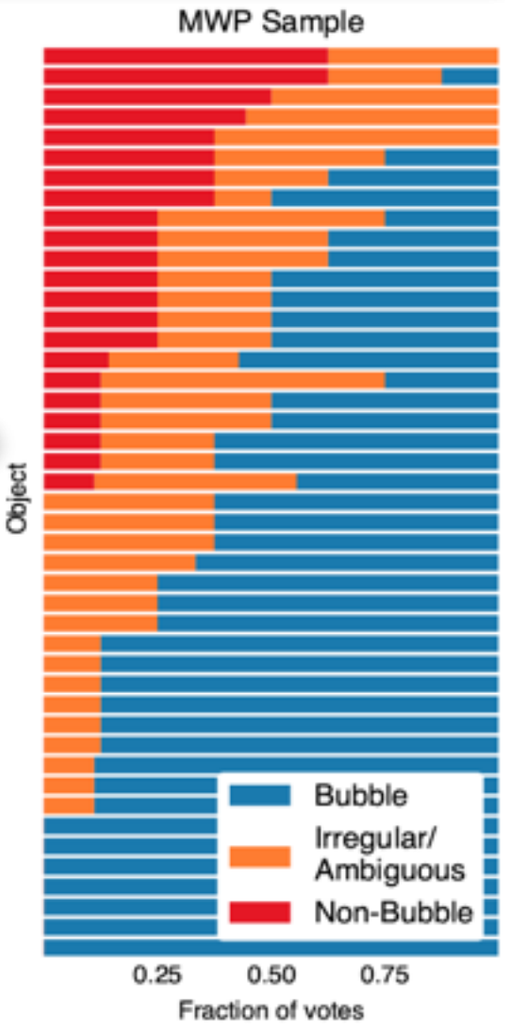
Explain these fields?

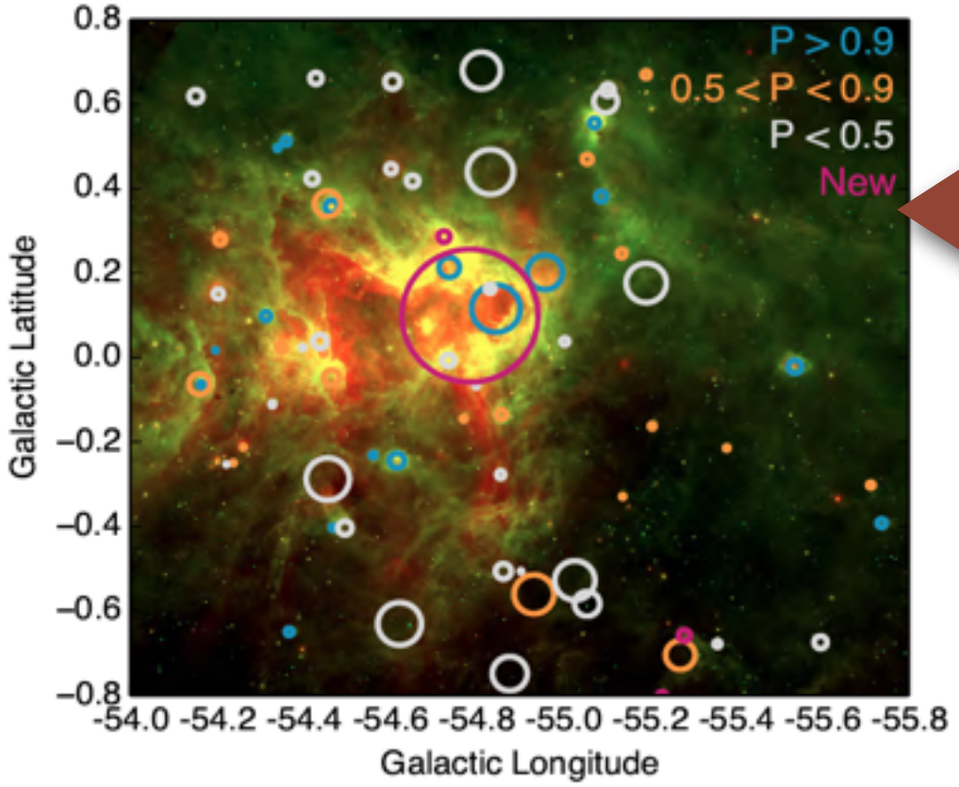
Discuss

BIG DATA and "Human-Aided Computing"



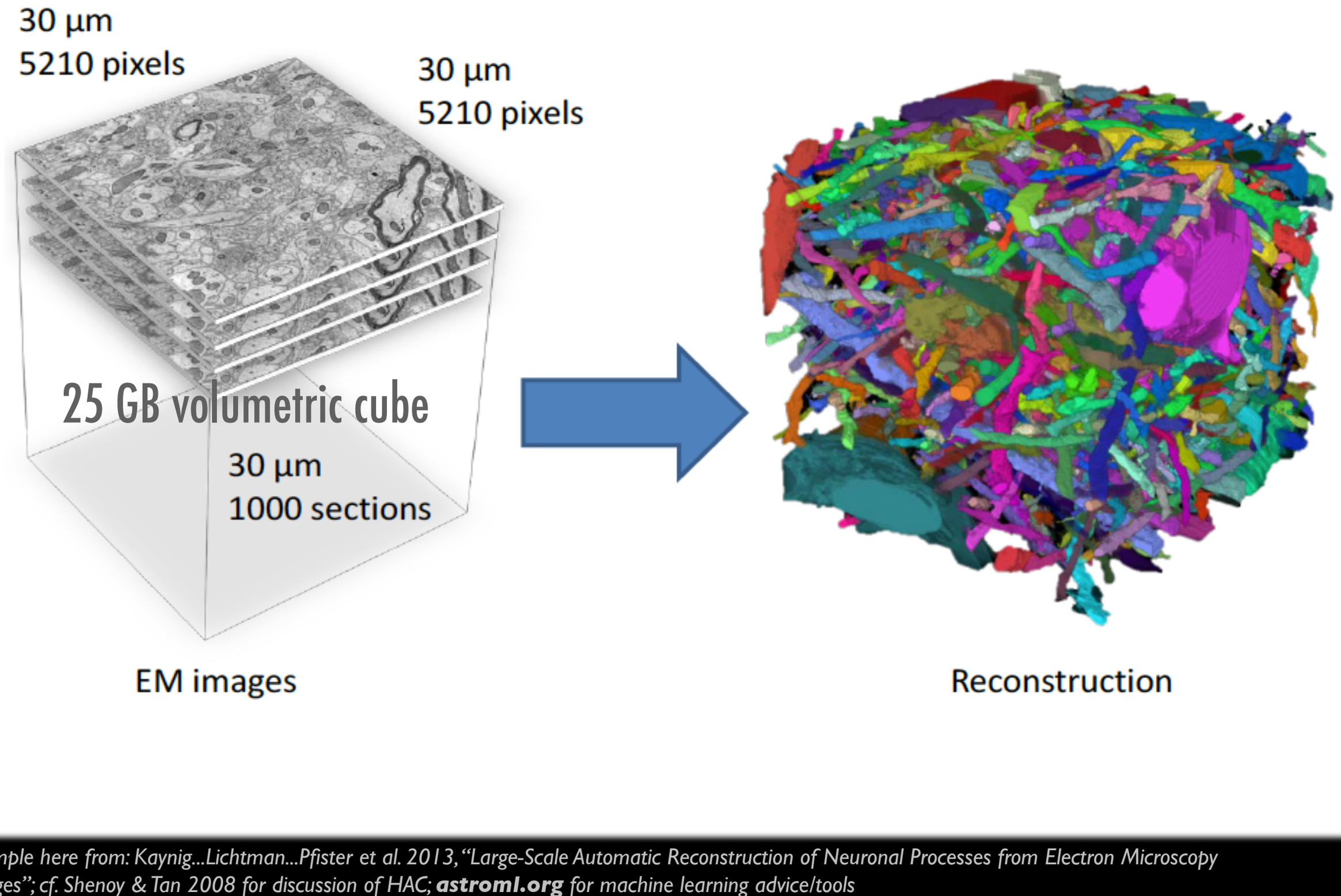
machine-learning algorith





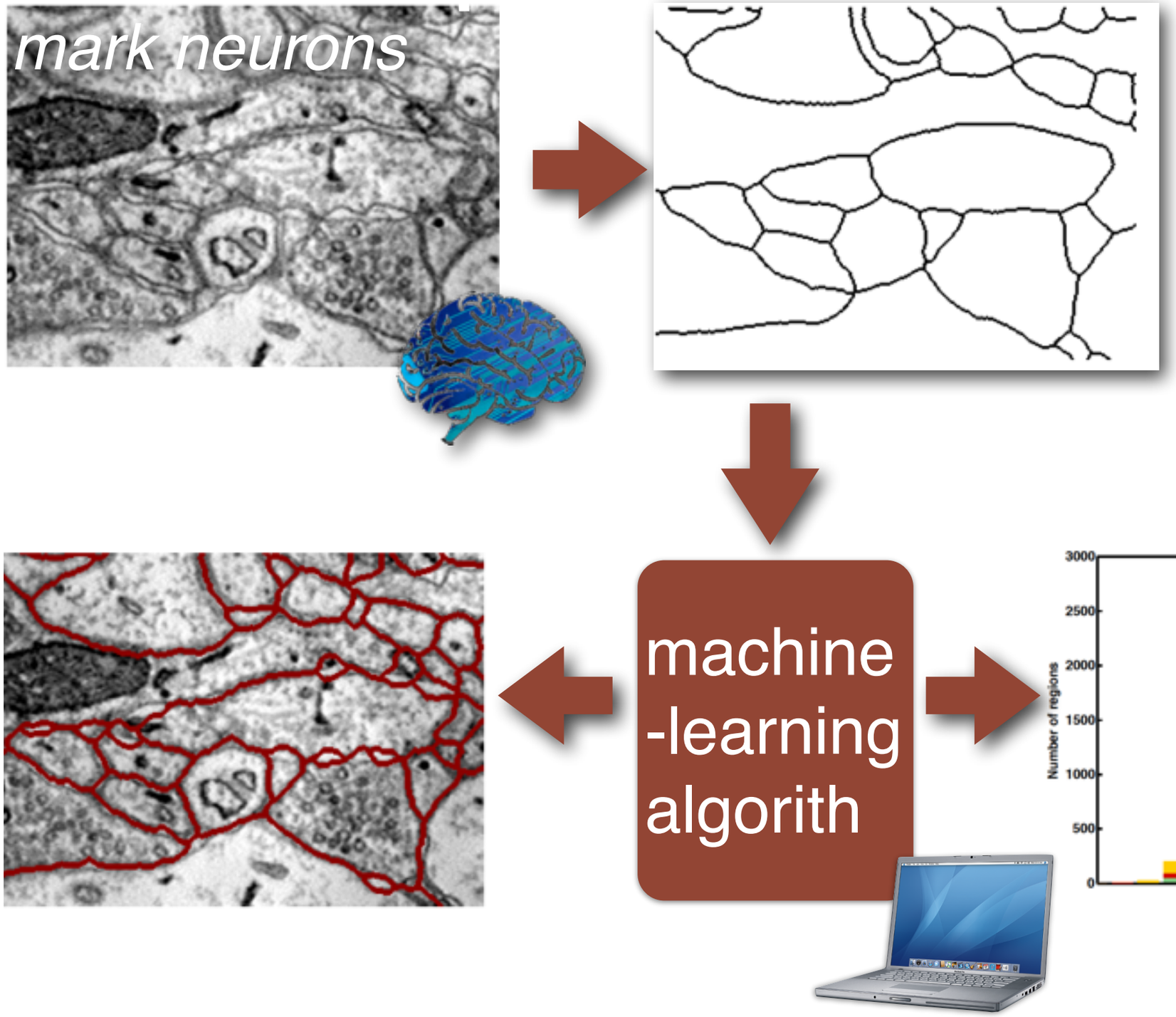
example here from: ***Beaumont****, Goodman, Kendrew, Williams & Simpson 2014; based on* ***Milky Way Project*** *catalog (Simpson et al. 2013), which came from* ***Spitzer/GLIMPSE*** *(Churchwell et al. 2009, Benjamin et al. 2003), cf. Shenoy & Tan 2008 for discussion of HAC;* ***astroml.org*** *for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

BIG DATA and "Human-Aided Computing"

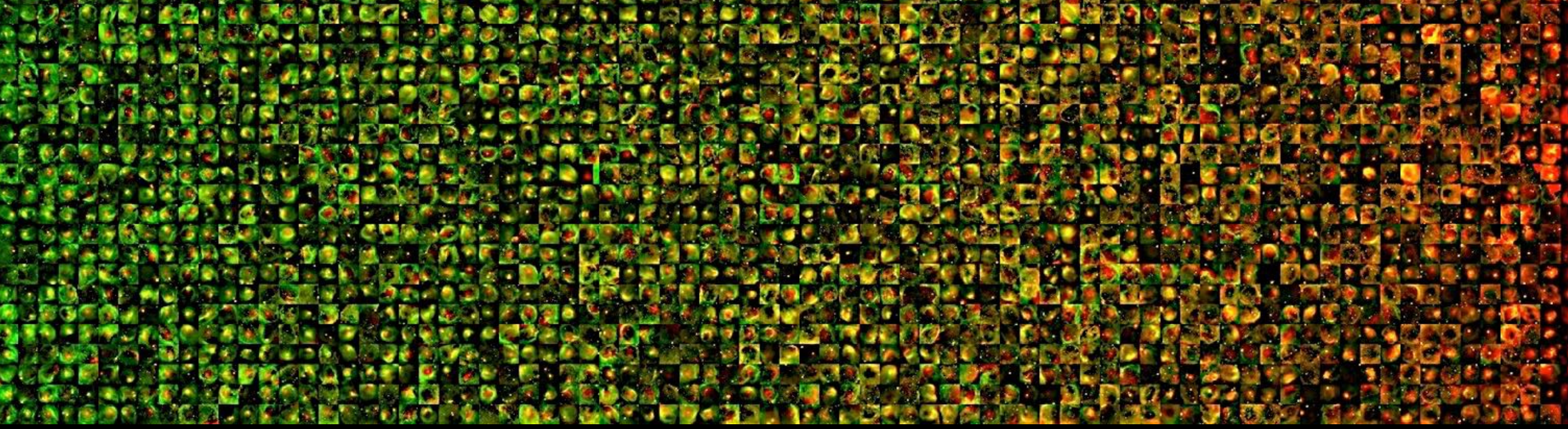


example here from: Kaynig...Lichtman...Pfister et al. 2013, "Large-Scale Automatic Reconstruction of Neuronal Processes from Electron Microscopy Images"; cf. Shenoy & Tan 2008 for discussion of HAC; ***astroml.org*** *for machine learning advice/tools*

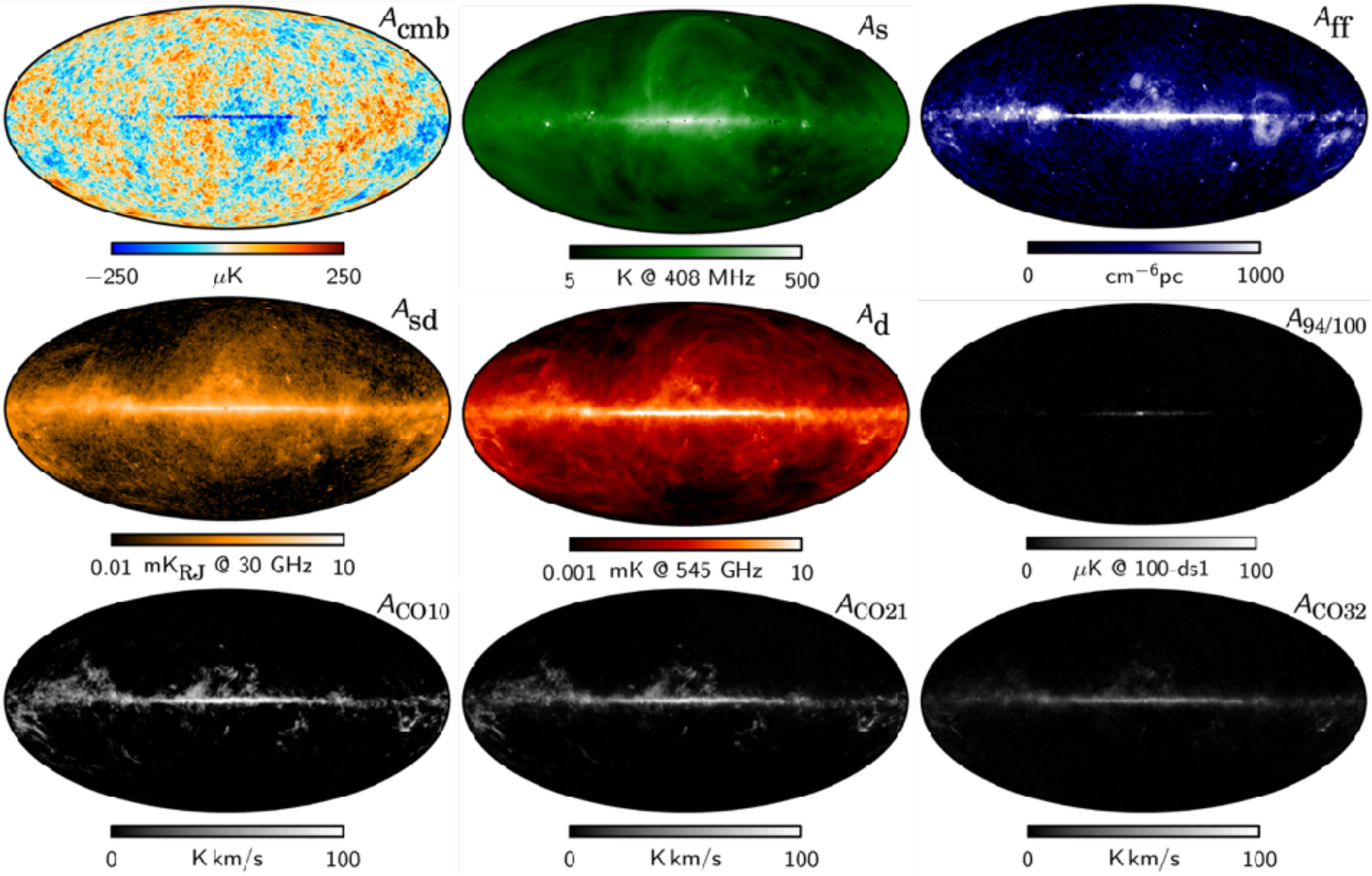
BIG DATA and "Human-Aided Computing"



example here from: Kaynig...Lichtman...Pfister et al. 2013, "Large-Scale Automatic Reconstruction of Neuronal Processes from Electron Microscopy Images"; cf. Shenoy & Tan 2008 for discussion of HAC; ***astroml.org*** *for machine learning advice/tools (Note: RF=Random Forest; CRF=Conditional Random Fields.)*



Wide Data



Temperature Foreground amplitudes from Commander, Planck Data [Feb 2015]

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”

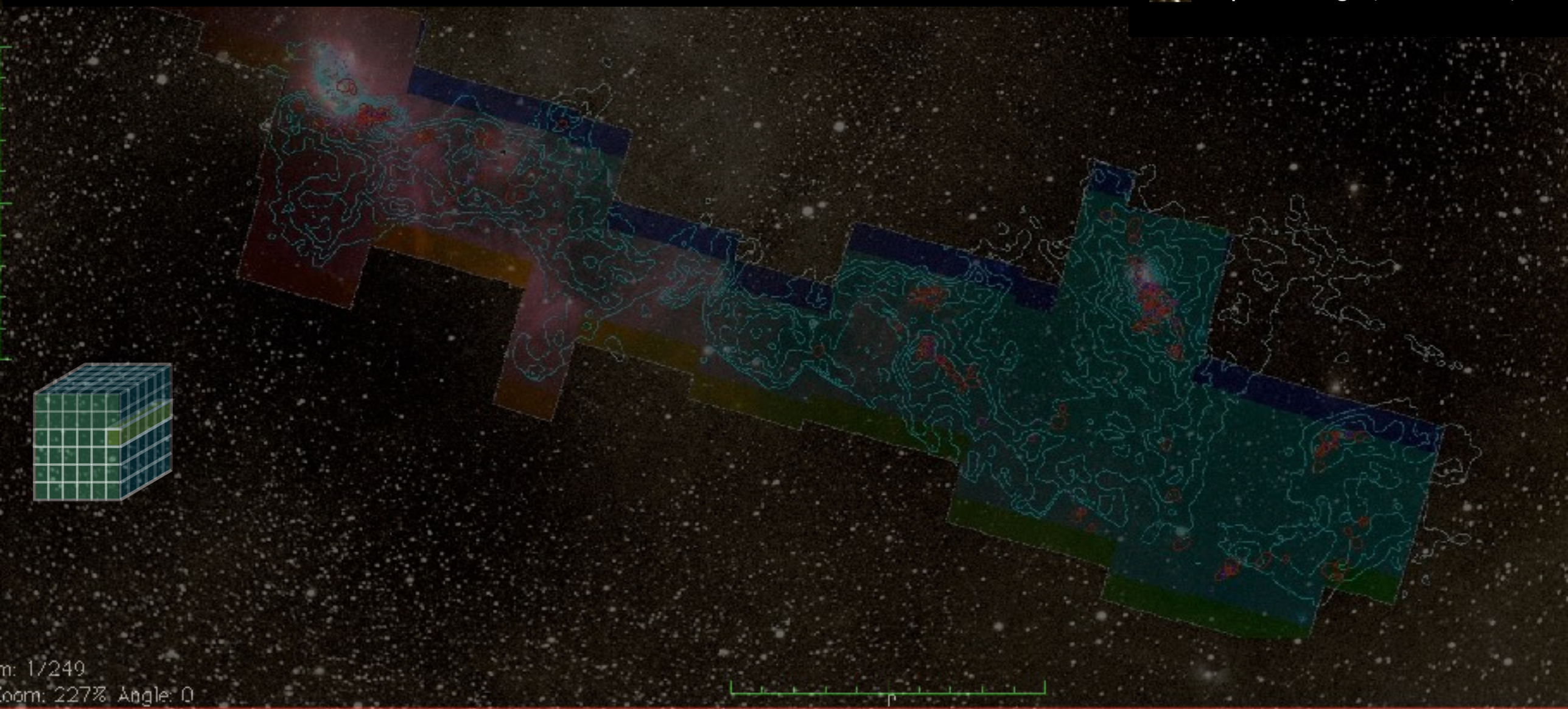
mm peak (Enoch et al. 2006)

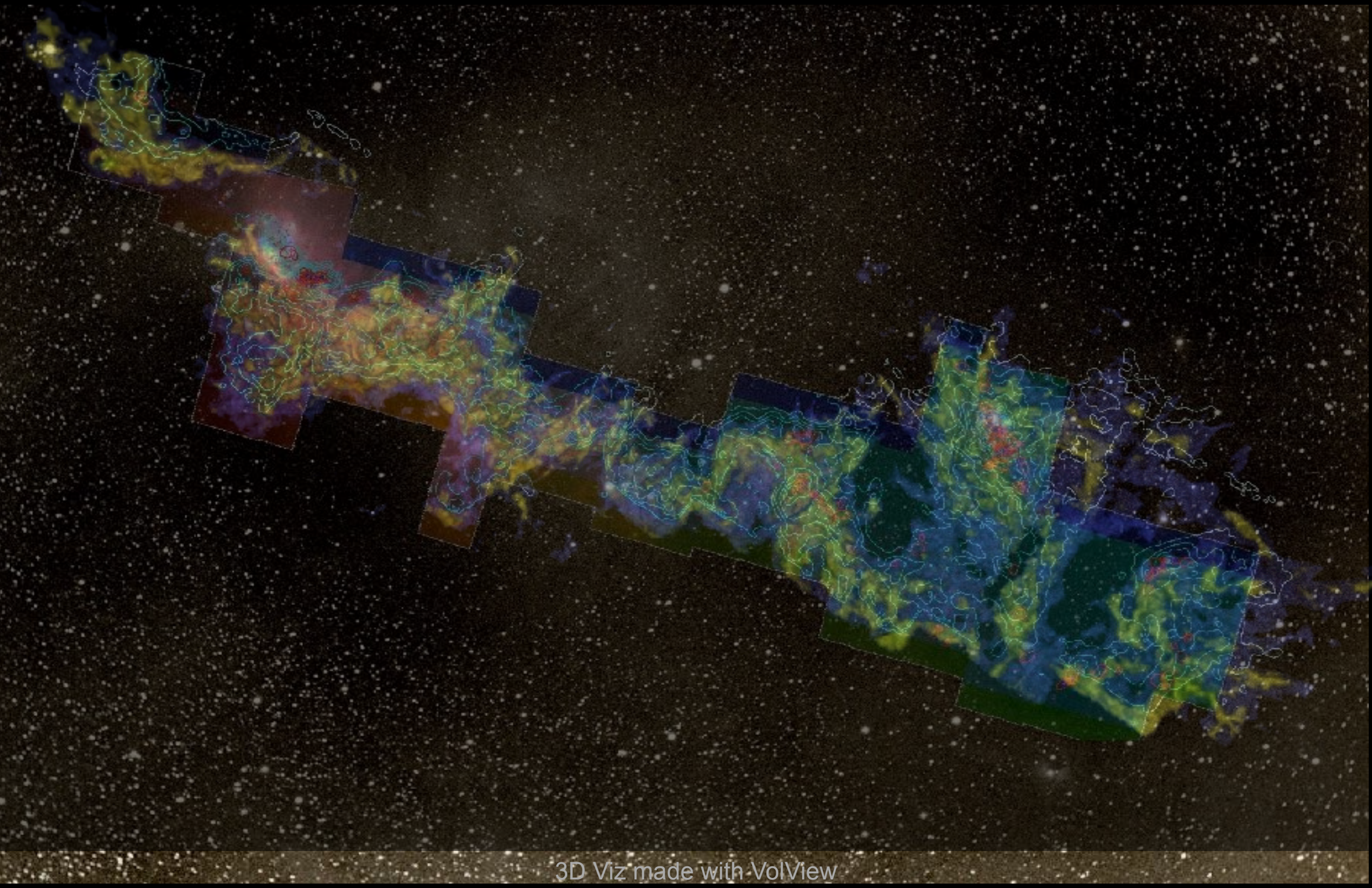
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

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

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

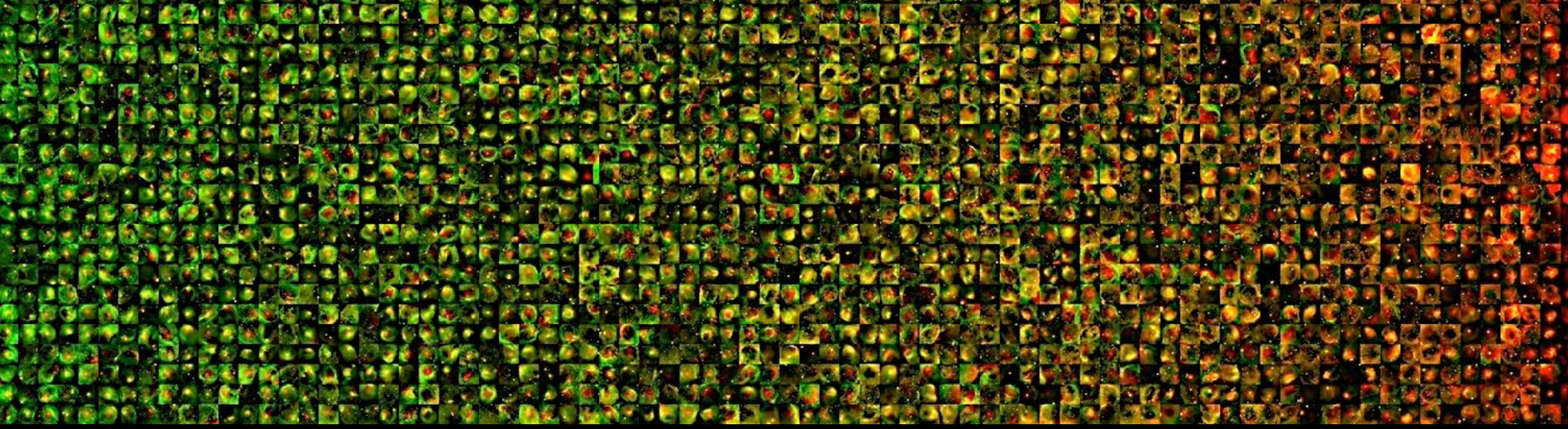
Optical image (Barnard 1927)

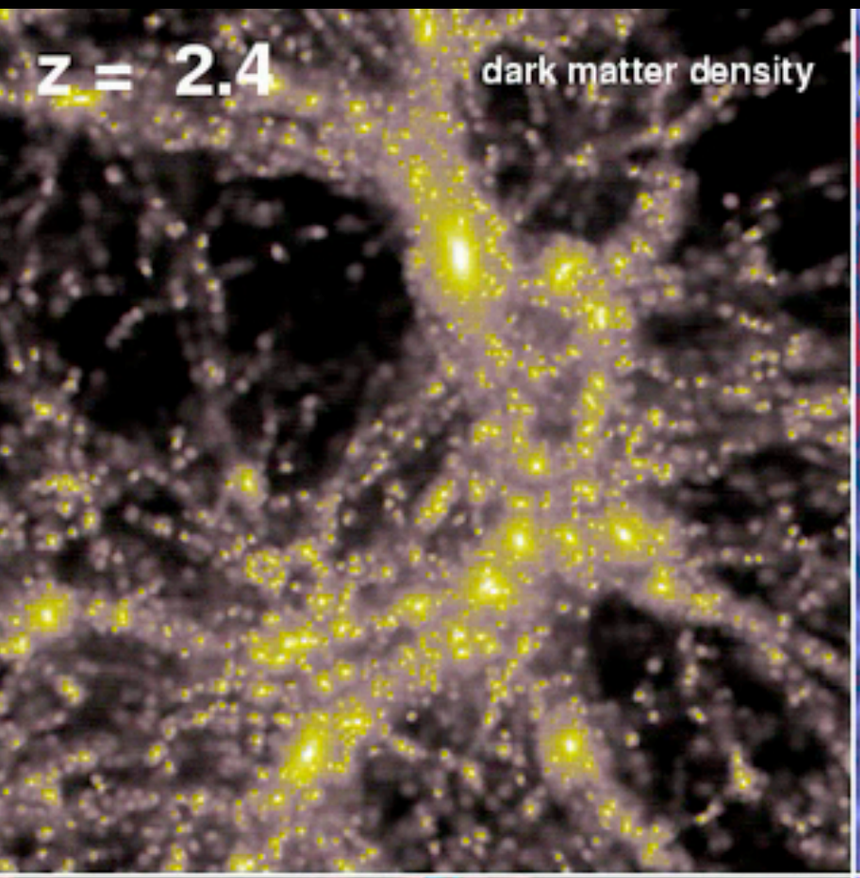


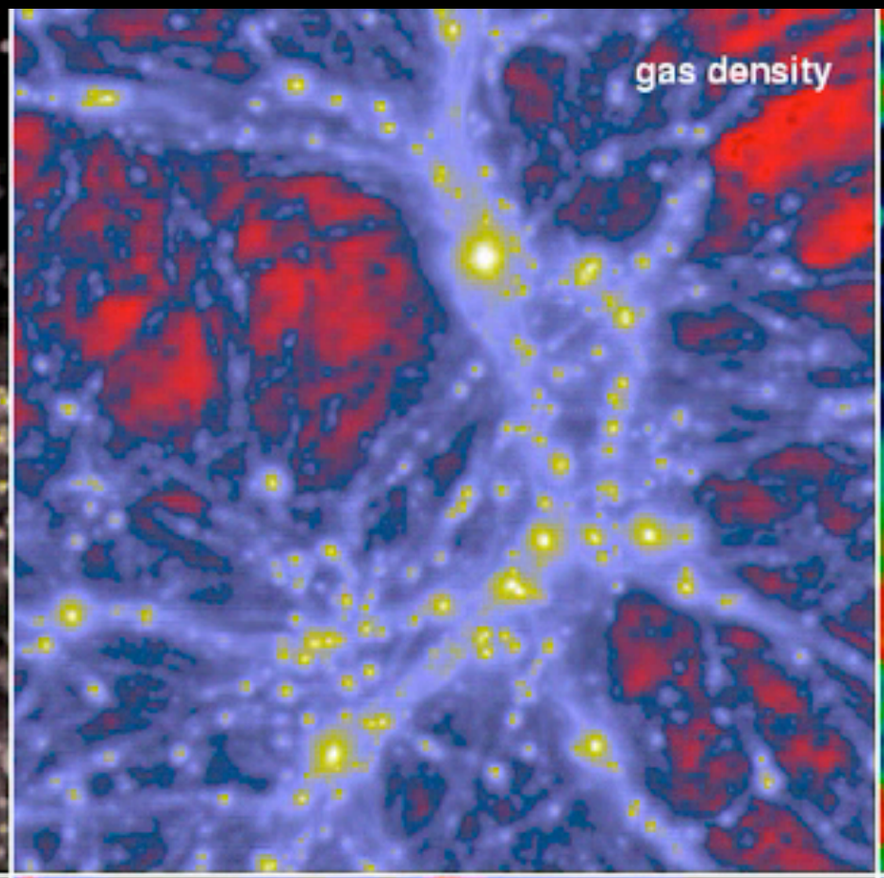


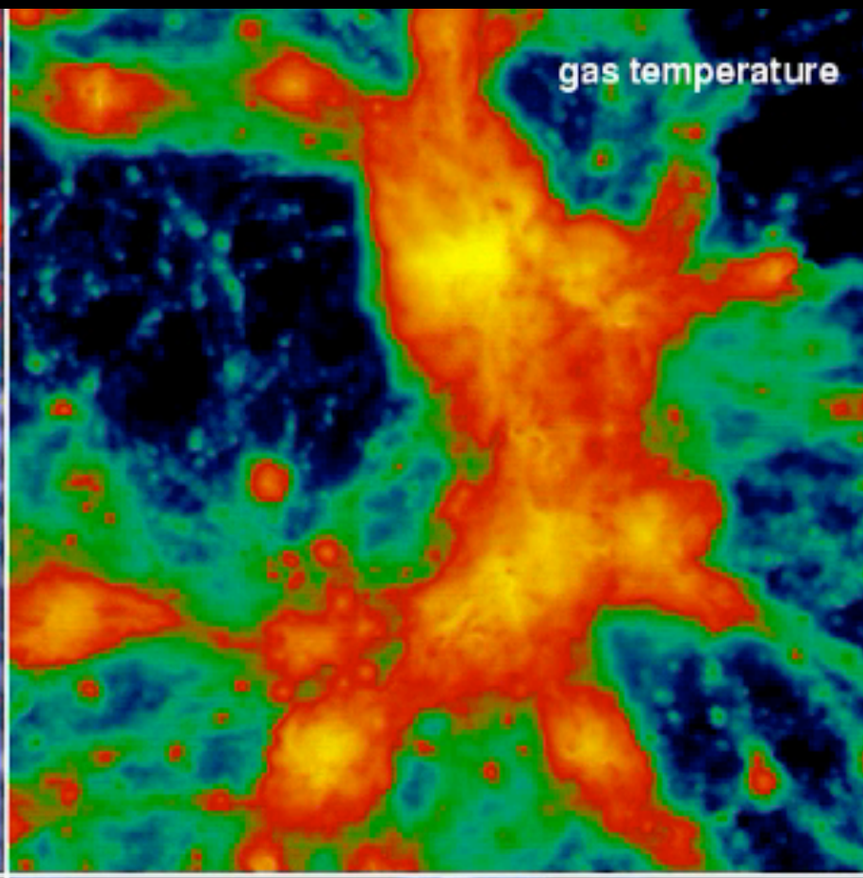
AstronomicalMedicine

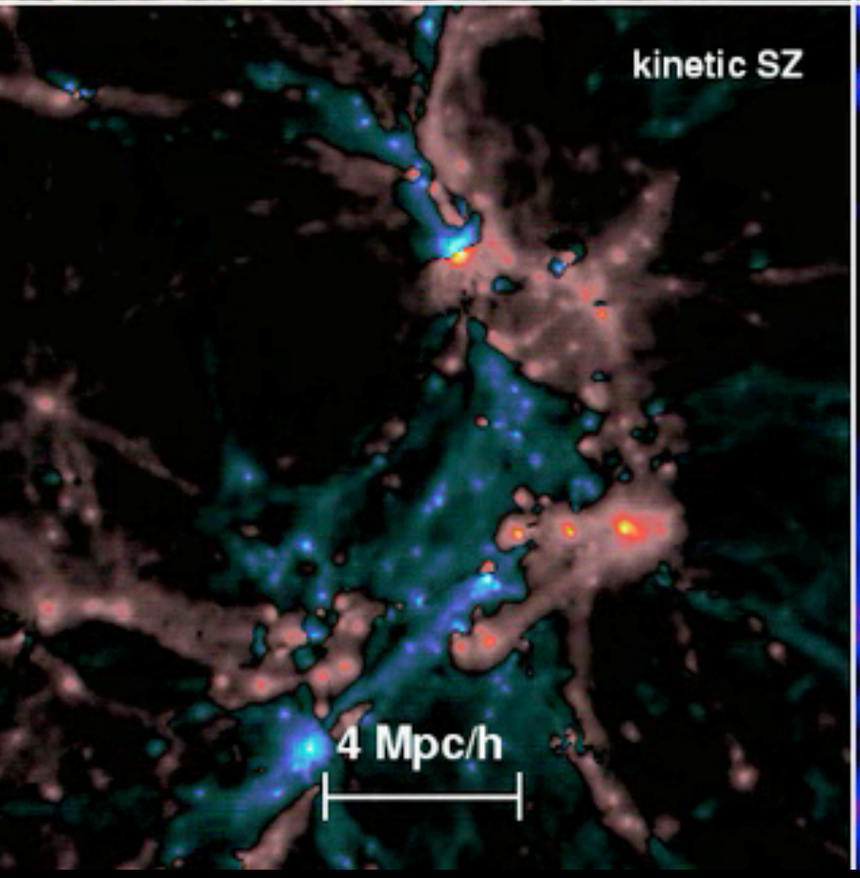
COMPLETE

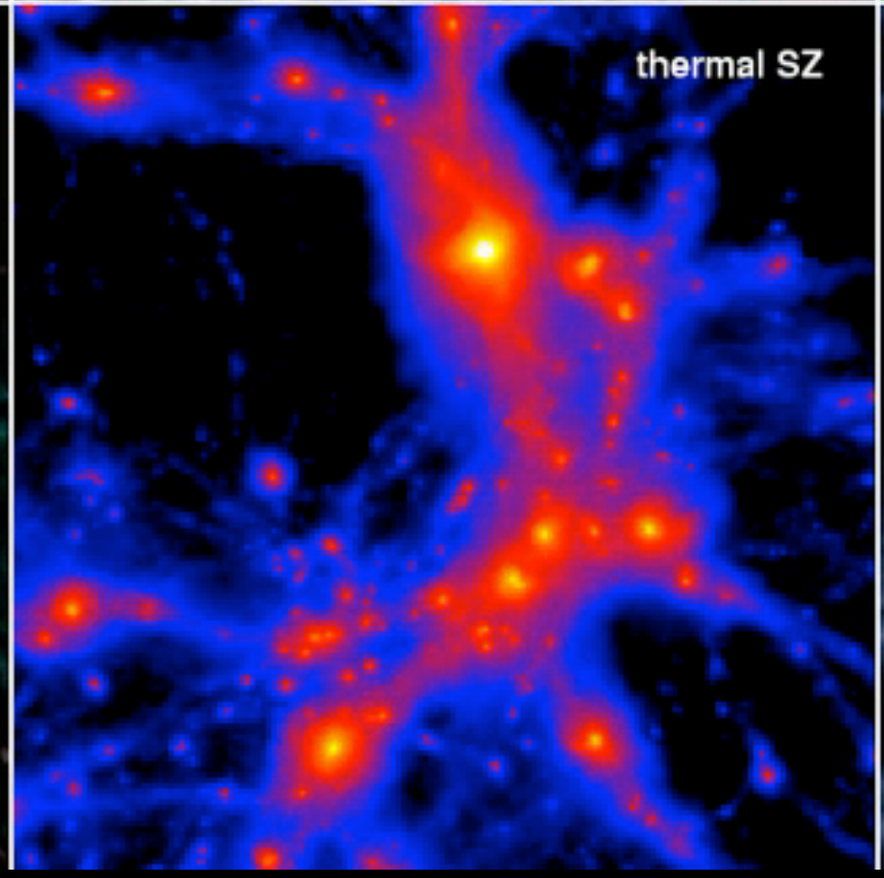


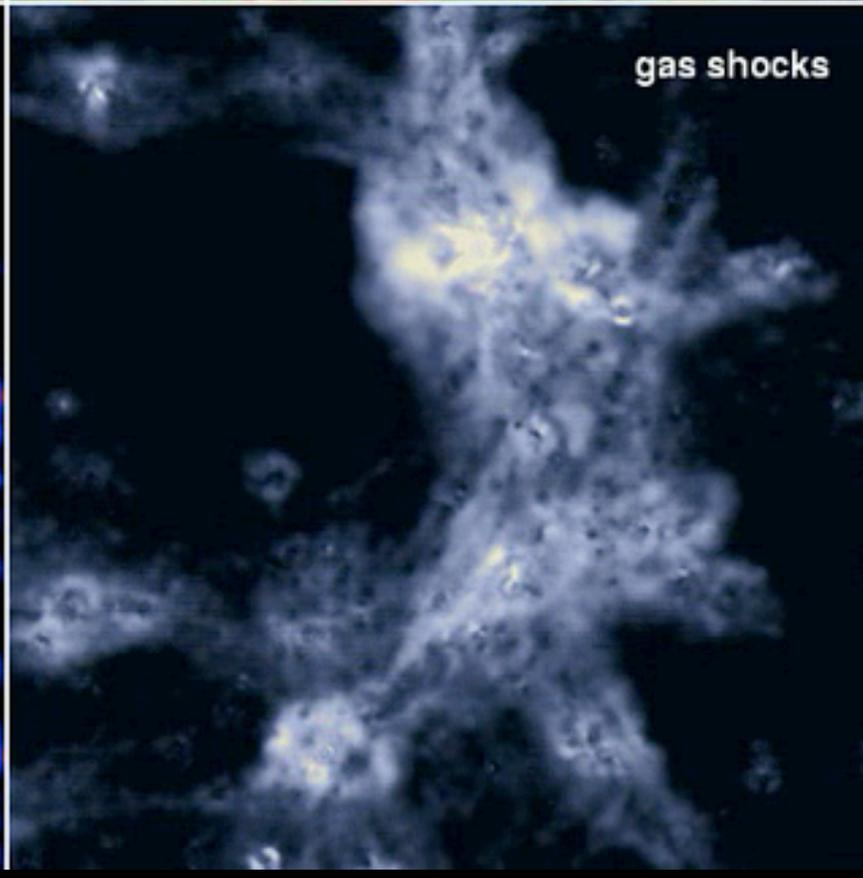












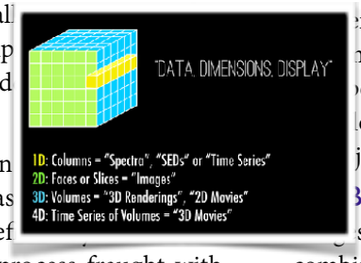
Movie: Volker Springel, formation of a cluster of galaxies. Millenium Simulation requires 25TB for output.

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 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*

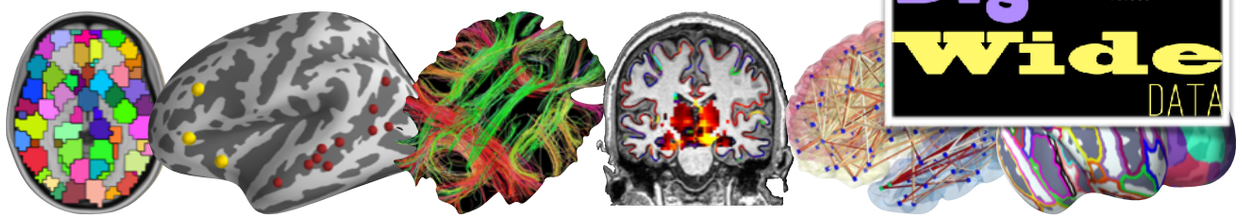
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

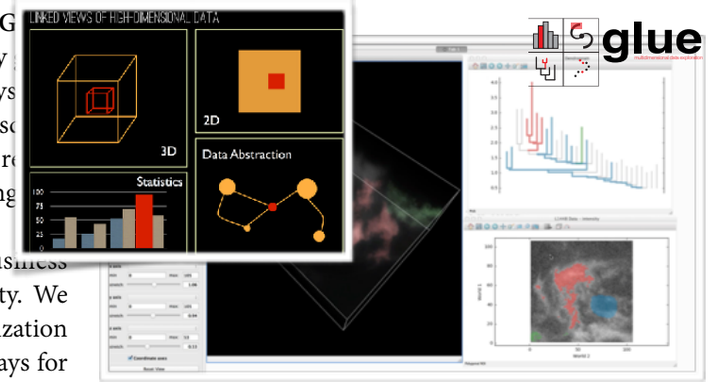
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

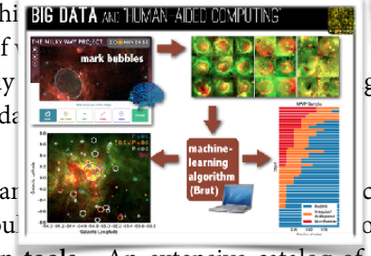
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.



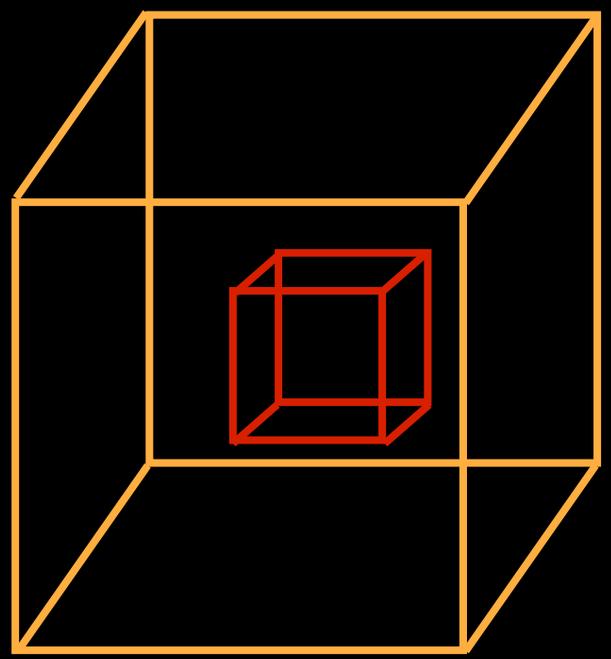
PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, we can **use experts' segmentations of brain imaging to train 3D selection tools**. An extensive catalog of such expert segmentation has already been assembled 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that project. **Goodman** and **Rosen** both have extensive leadership experience, and the one of their labs to host a part-time postdoctoral fellow and/or graduate stude and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste unprecedented nature of what we propose, a smart 3D selection without Holo 3D selection tool, would be successes on their own, with immediate applicability

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

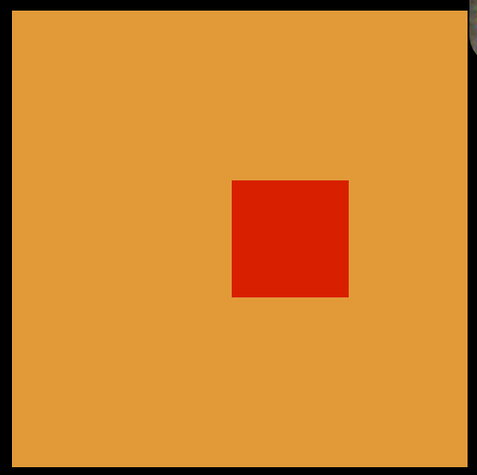
[2]Beaumont, Goodman, et al. *Brut: Automatic bubble classifier,* http://adsabs.harvard.edu/abs/2014ascl.soft07016B

Linked Views of High-dimensional Data



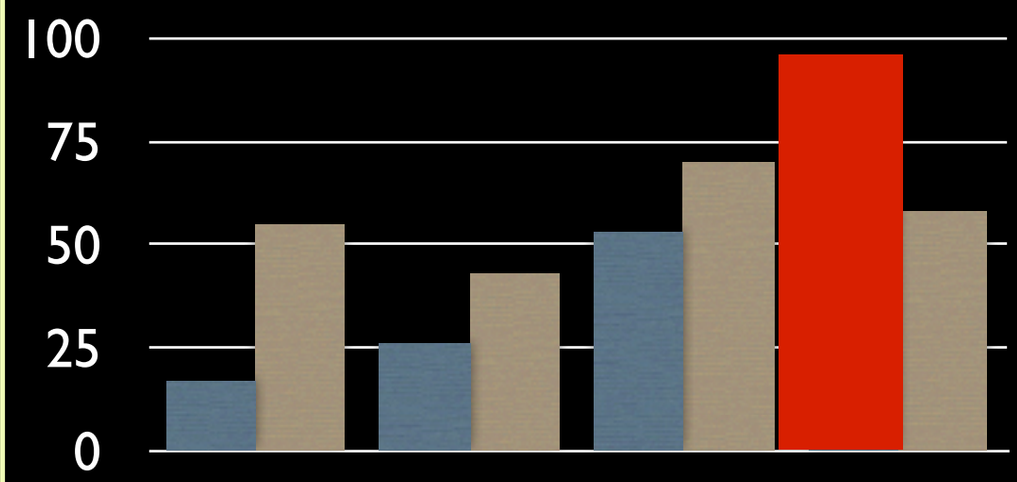
3D

John Tukey

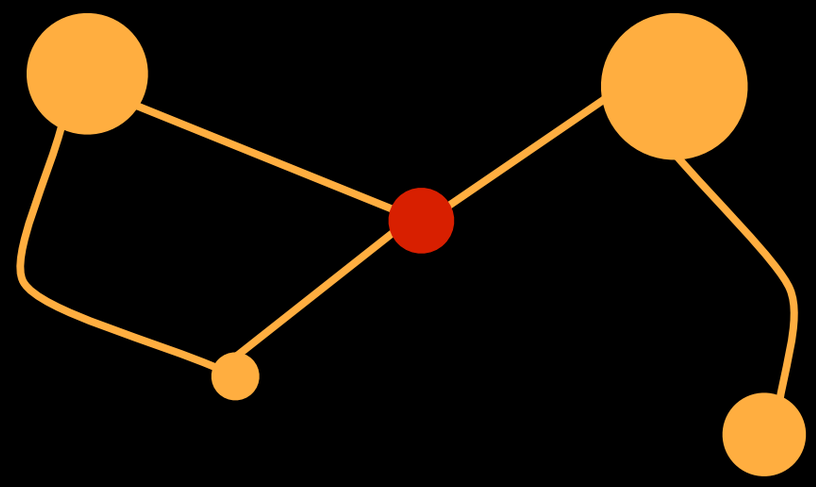


2D

Statistics

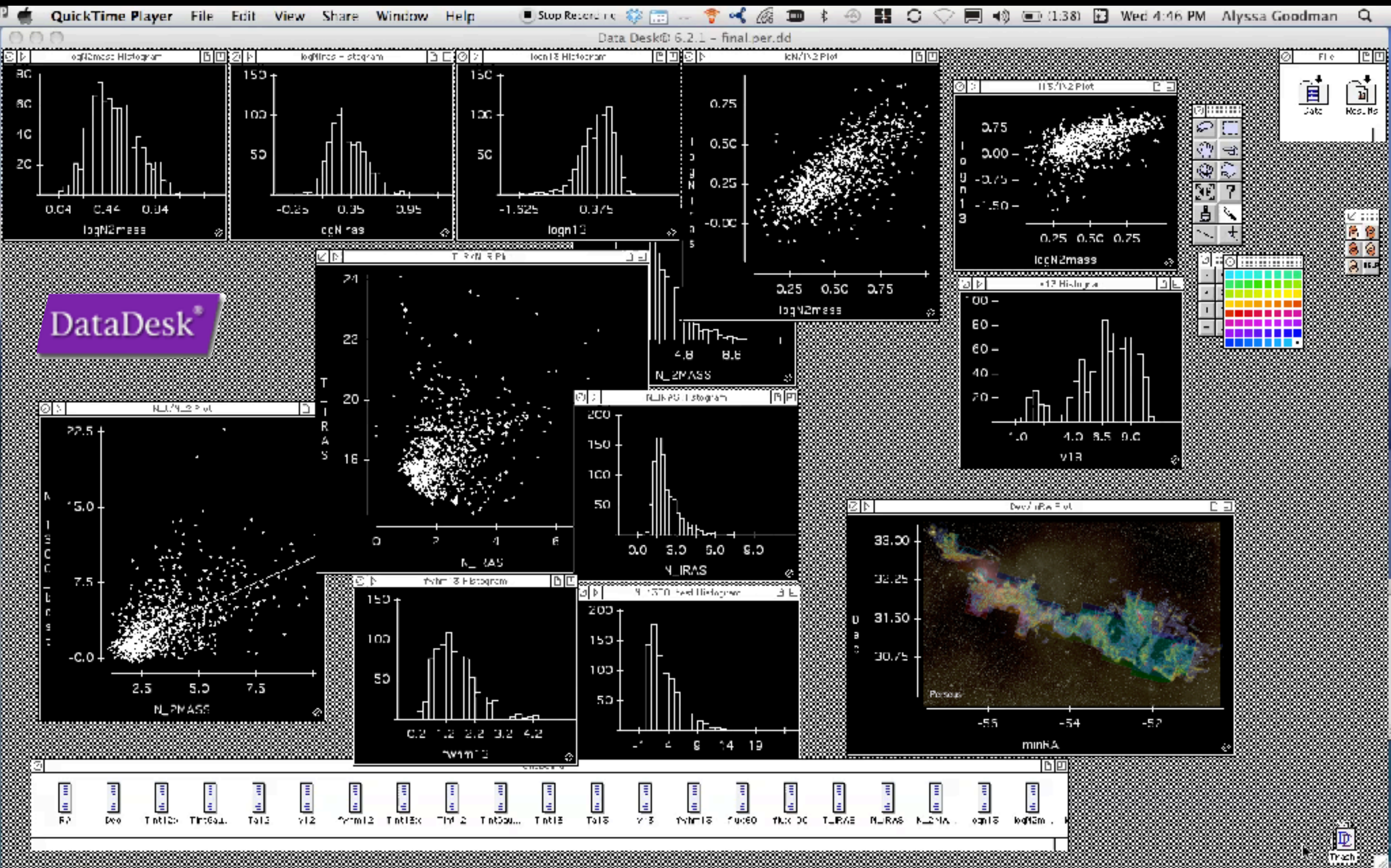


Data Abstraction

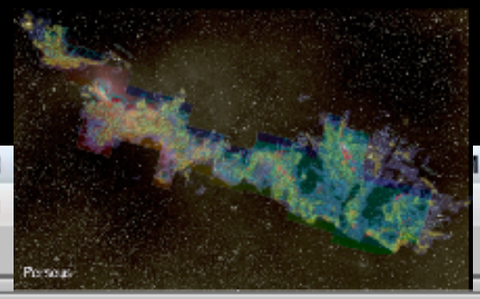


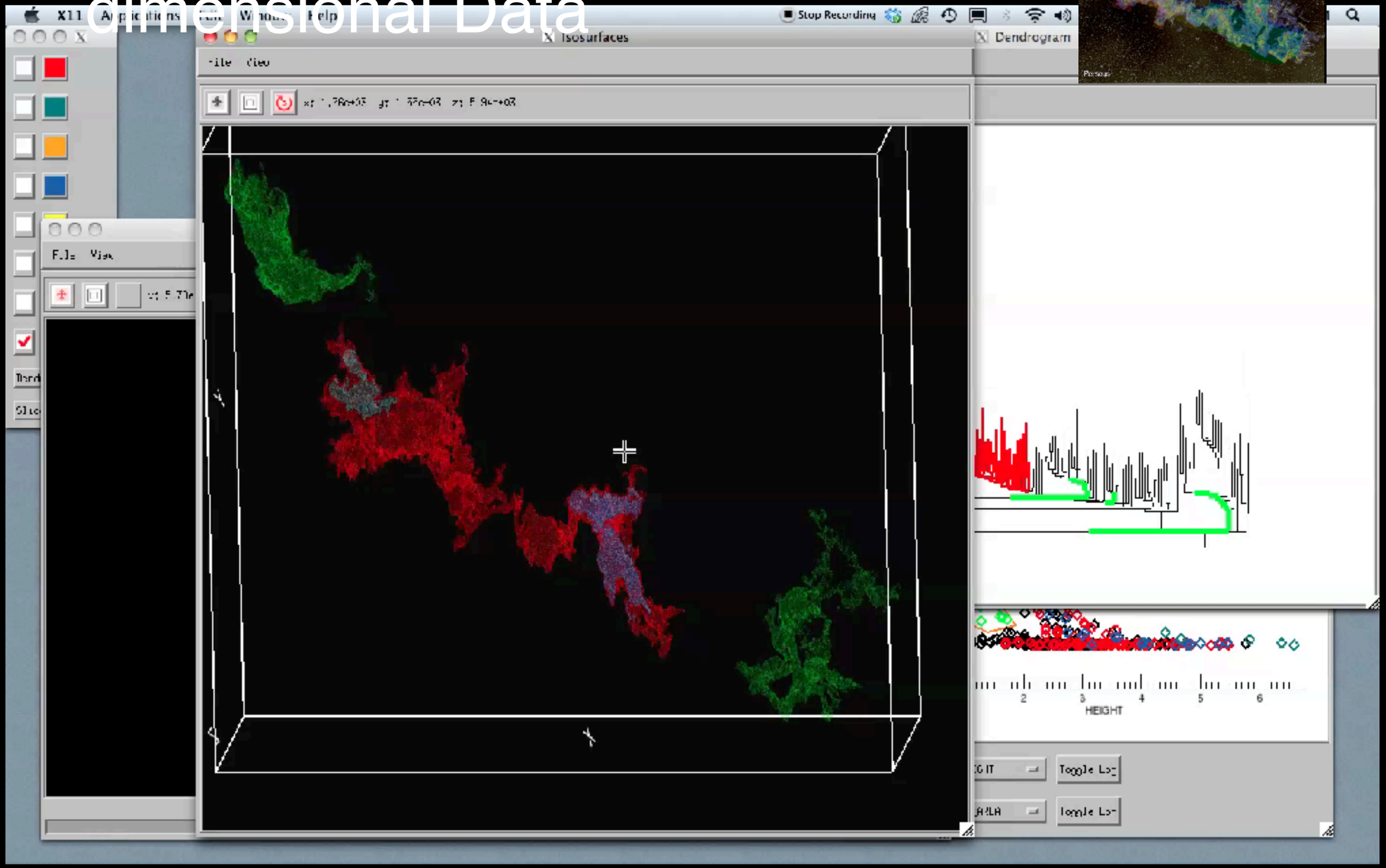
figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

DataDesk (est. 1986)



Linked Views of High-dimensional Data





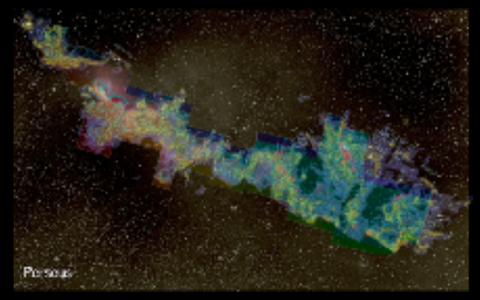
Video & implementation: Christopher Beaumont, Harvard→Counsyl;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

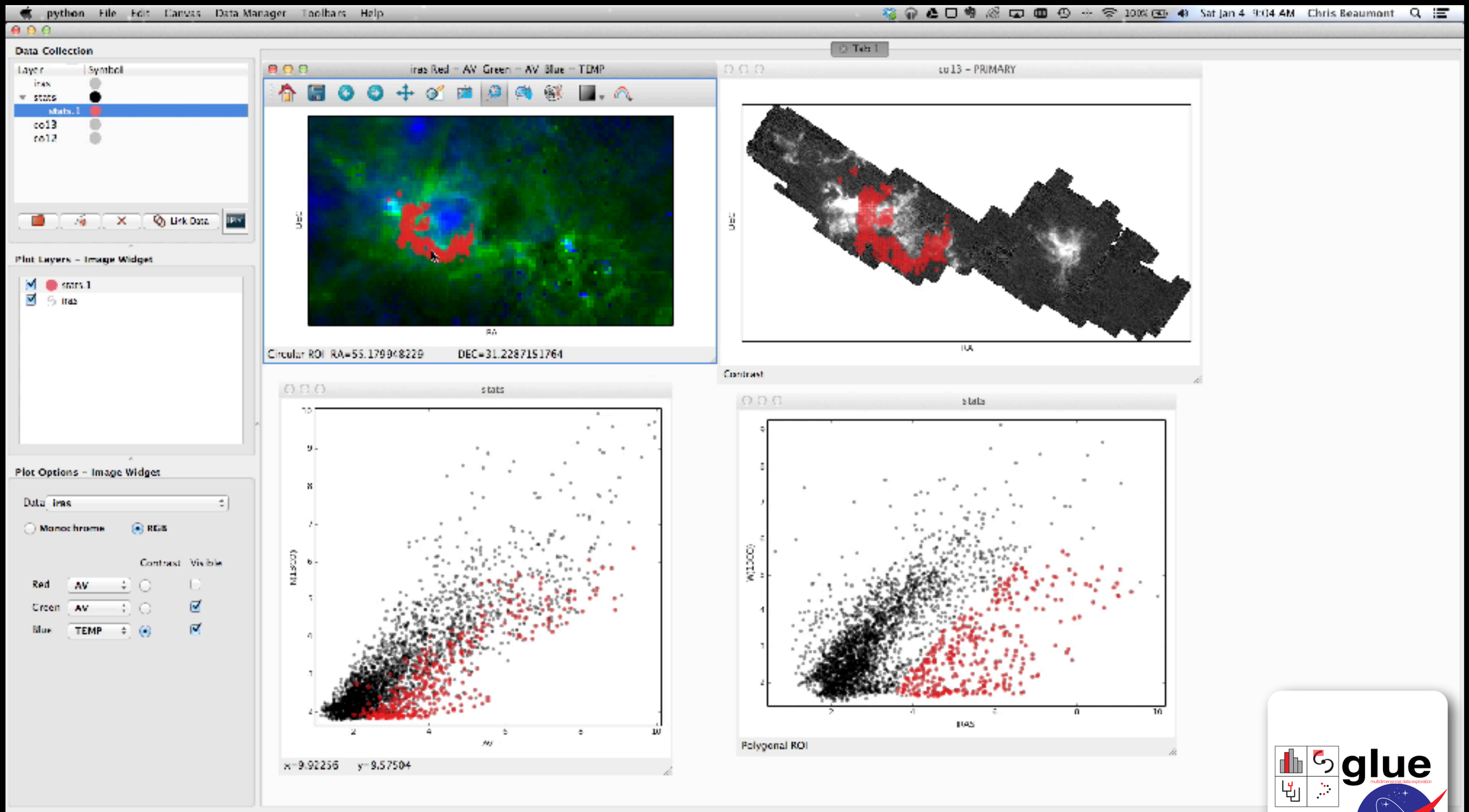
great. but that was all from one data file

(And it was in software that costs $1000.)

Linked Views of High-dimensional Data (in Python)

Glue

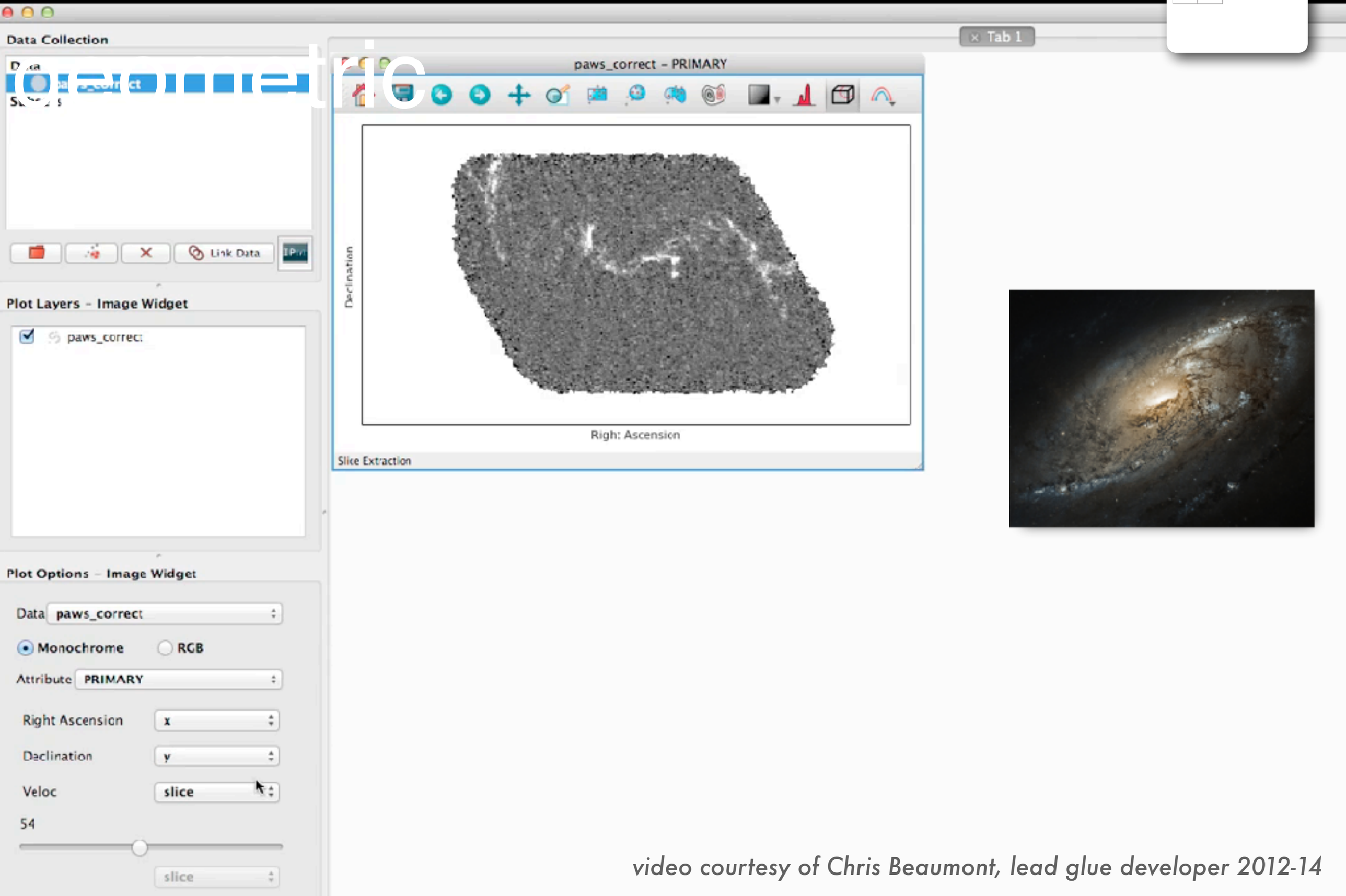




Christopher Beaumont, w/A. Goodman, T. Robitaille & M. Borkin

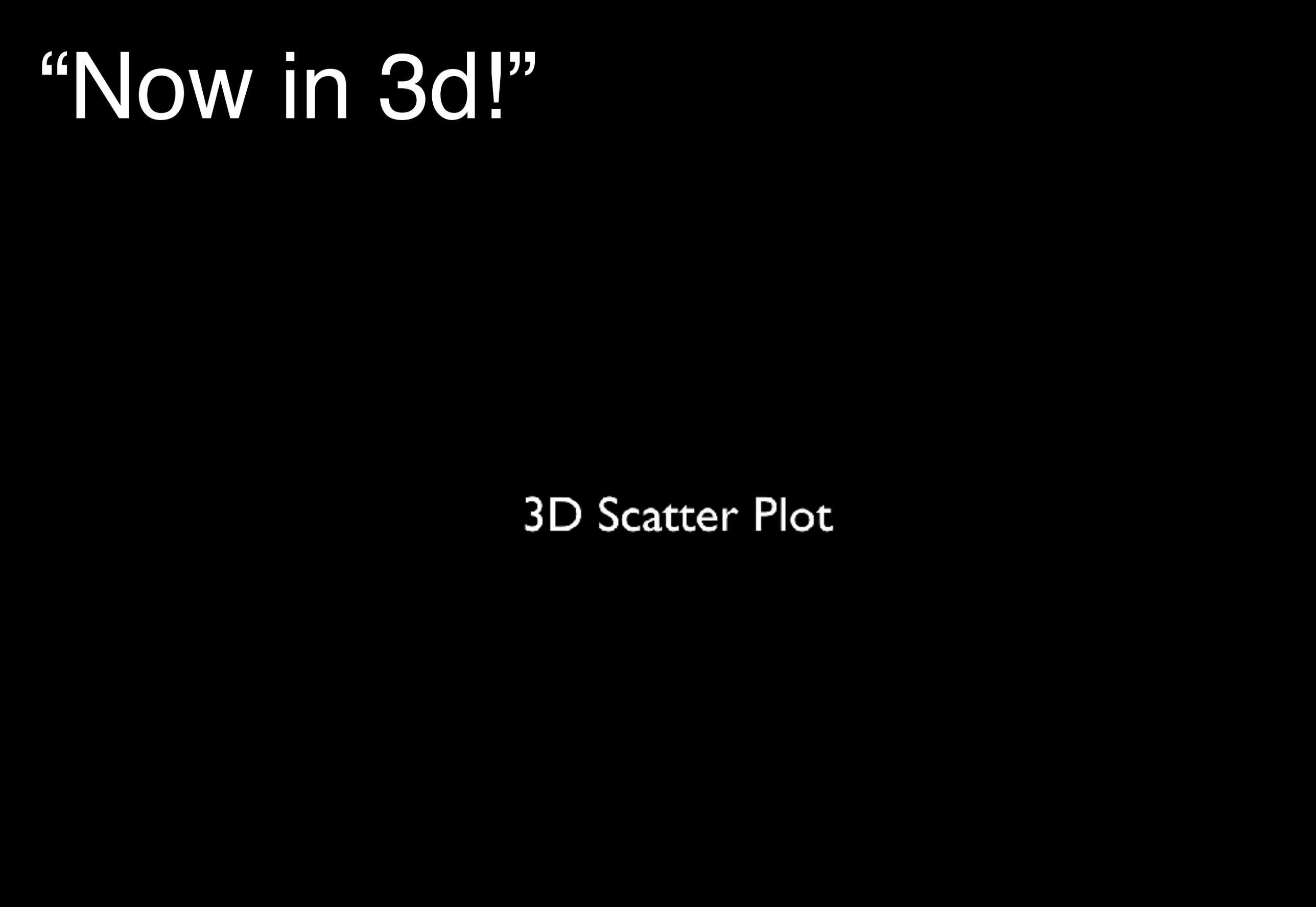
glue
multidimensional data exploration
NASA

Dimensions need not be geometric



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

“Now in 3d!”



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

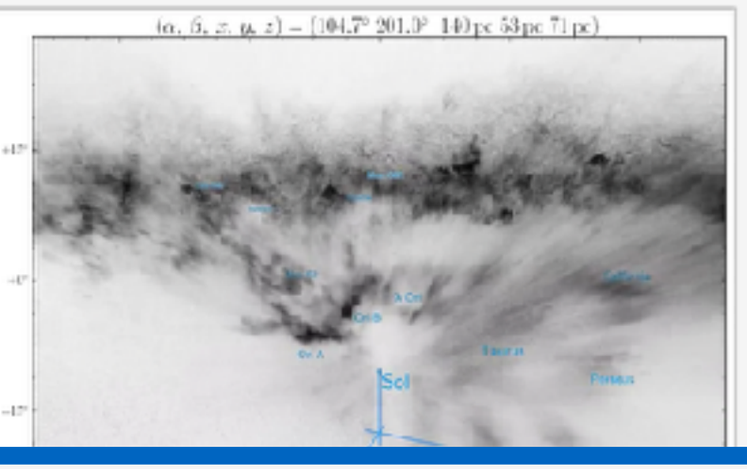
We present a 3D map of interstellar dust reddening. Using 800 million stars with Pan-STARRS 1 photometry, and matched 2MASS photometry for 200 million stars, the map traces dust reddening in three dimensions across three-quarters of the sky.

Volume Renderings

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

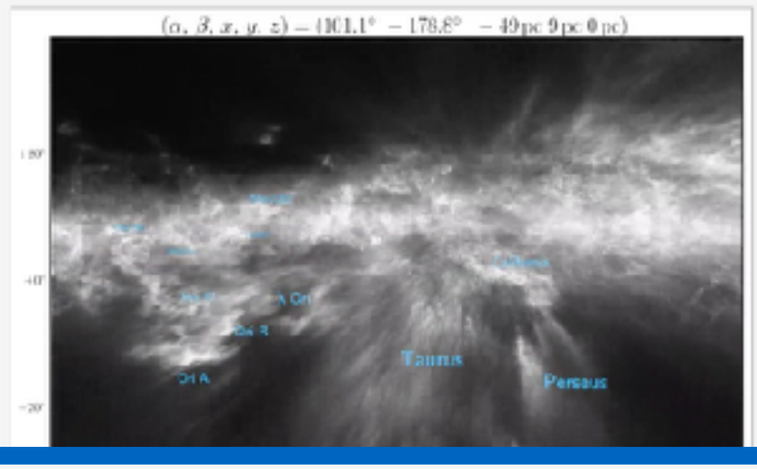
Local Dust

Looking towards the Galactic anticenter.



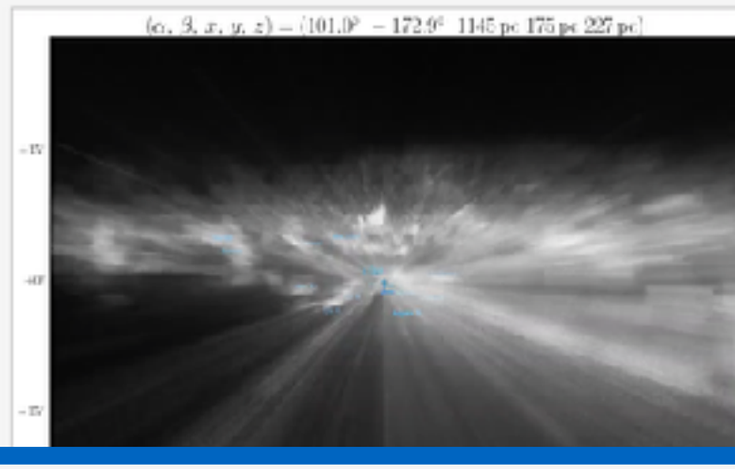
Orbiting the Sun

A 50-pc loop around the Sun.



A Tour of the Galaxy

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



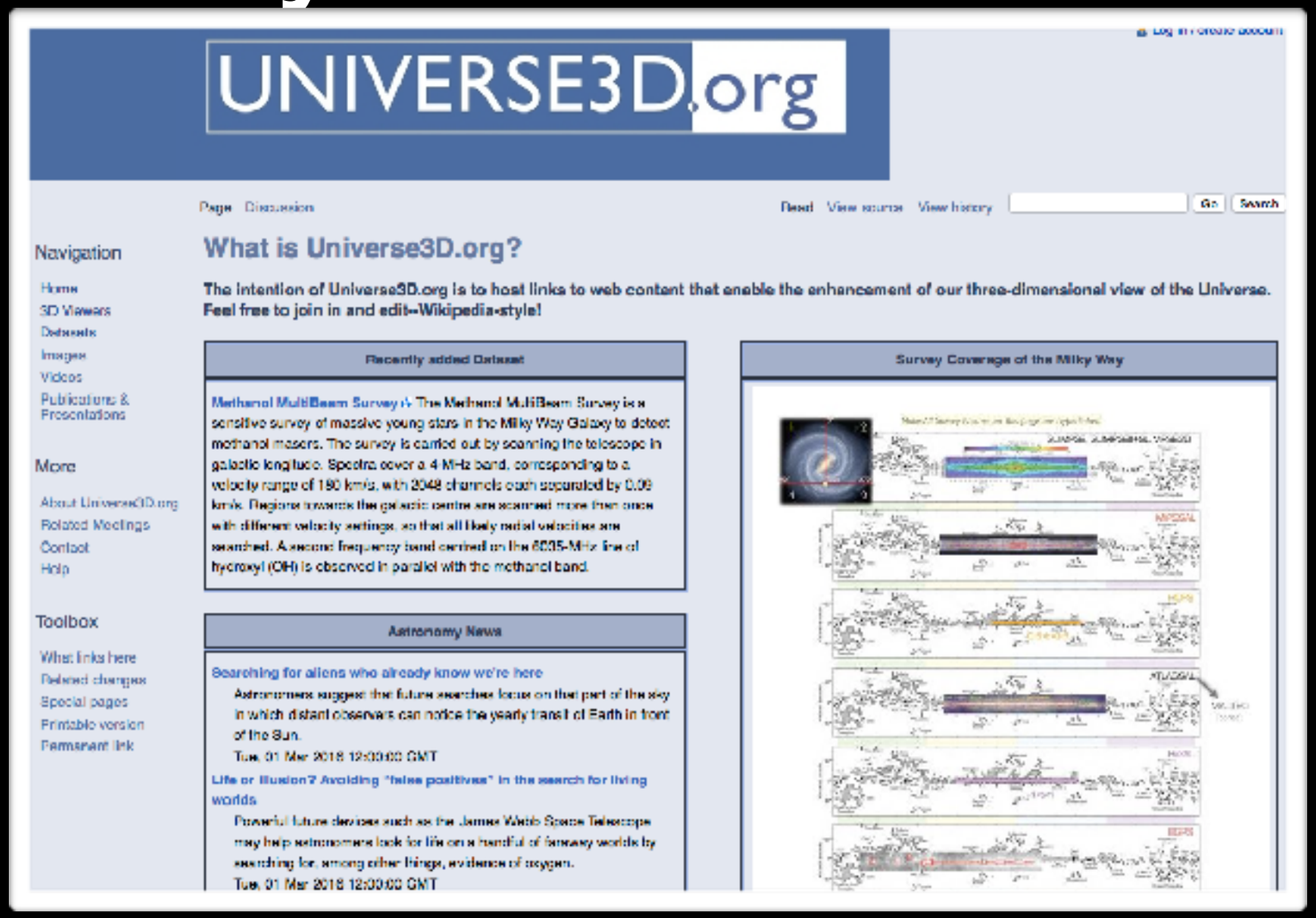
Gregory Green

Eddie Schlafly

Douglas Finkbeiner

Coming Soon: A “CfA Milky Way”

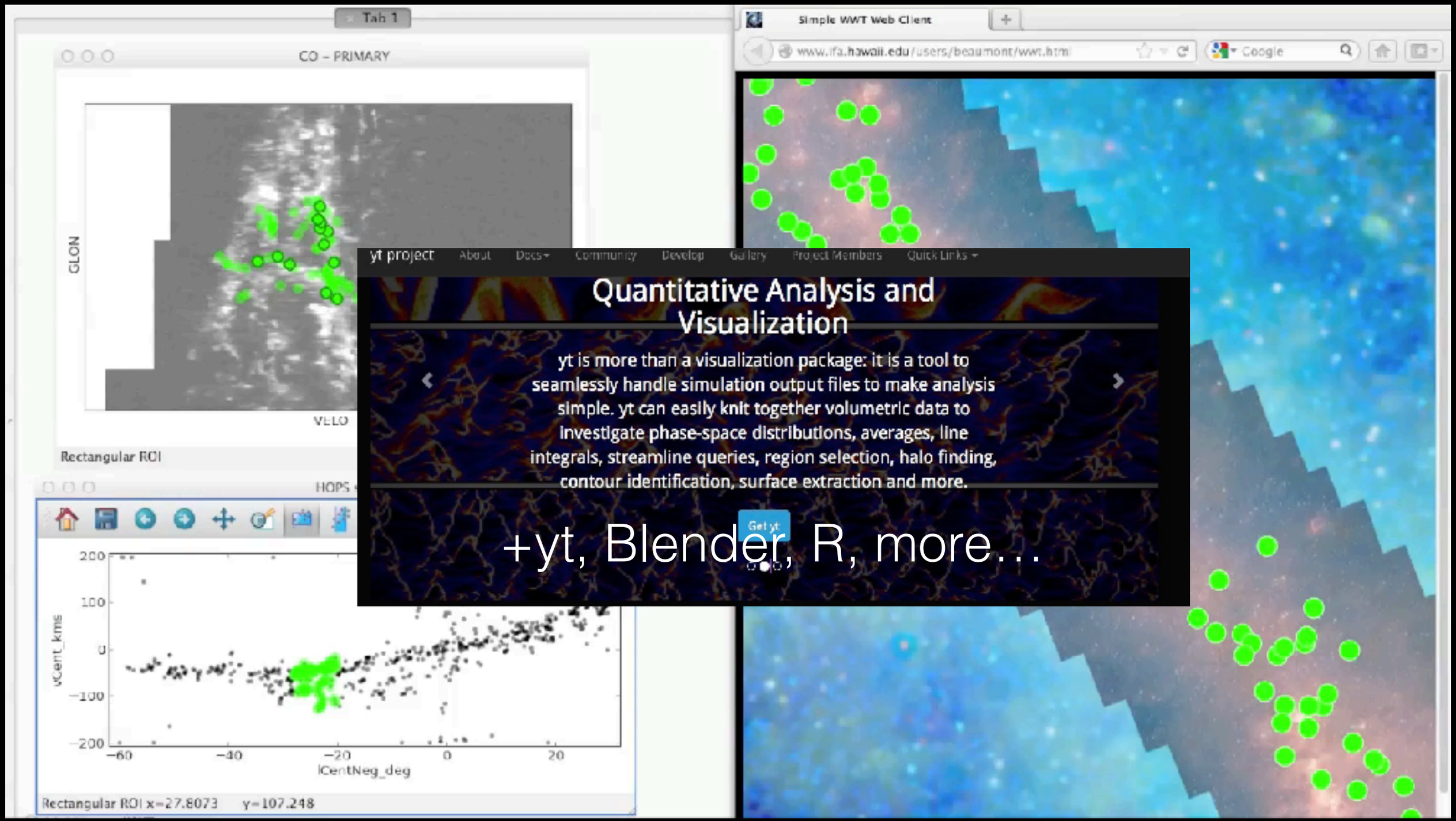
Battersby, Dame, Finkbeiner, Goodman, Green, Qian, Reid, Schlafly, Zucker, et al.



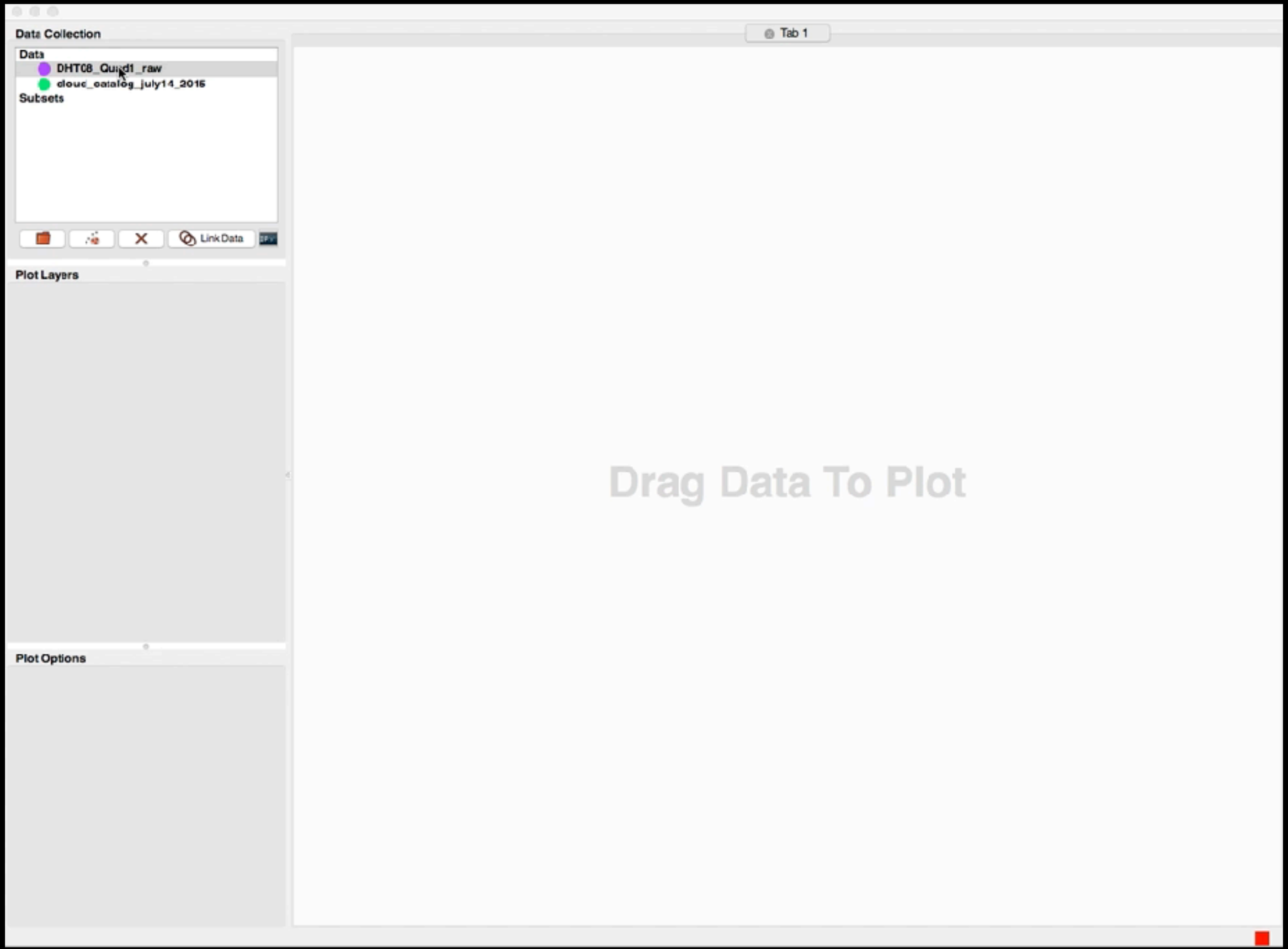
Video from WorldWide Telescope, based on work of Harvard Senior Matt Pasquini with Harvard Grad Student Catherine Zucker, using Rice et al. 2016 and Schlafly et al. 2014

glue

WorldWide Telescope



Video courtesy of Chris Beaumont



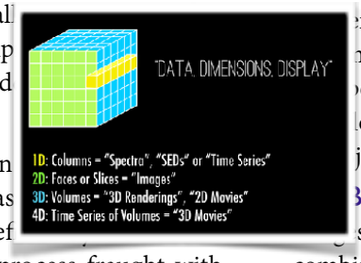
Video courtesy of Penny Qian (SAO predoc working on glue/Milky Way), Data from Dame et al. 2001; Rice et al. 2016

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 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*

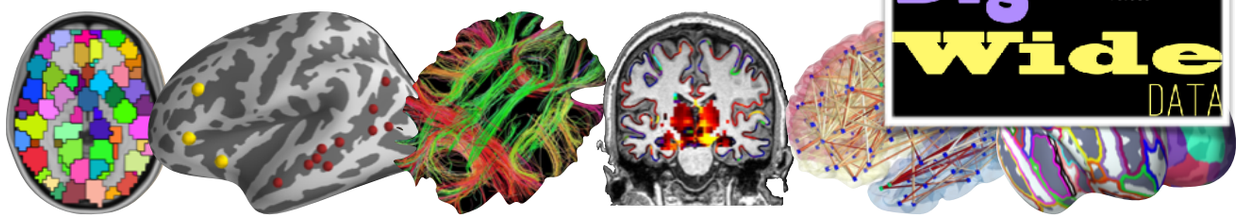
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

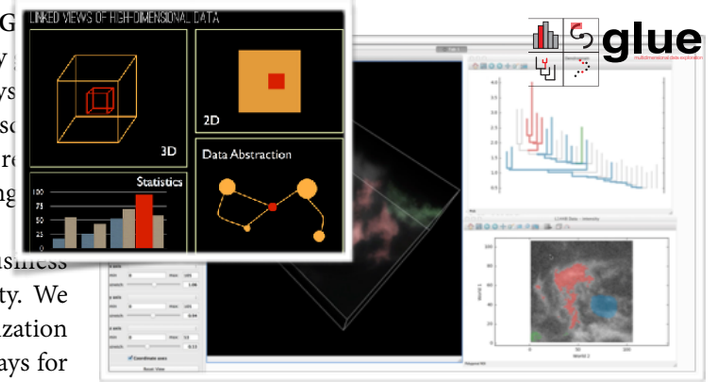
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

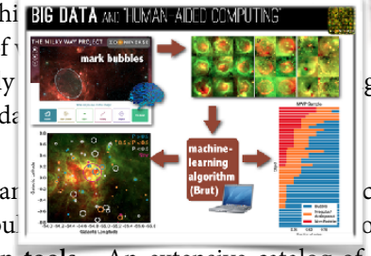
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.



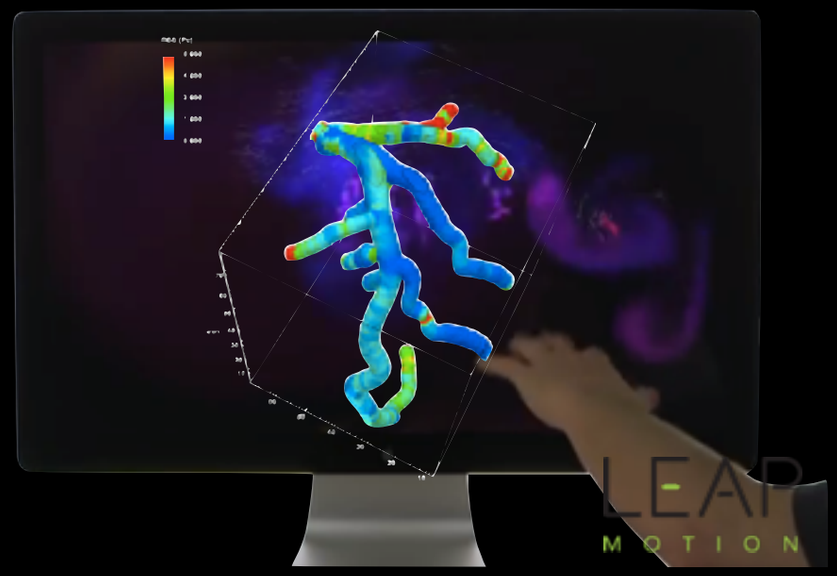
PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, we can **use experts' segmentations of brain imaging to train 3D selection tools**. An extensive catalog of such expert segmentation has already been assembled 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that project. **Goodman** and **Rosen** both have extensive leadership experience, and the one of their labs to host a part-time postdoctoral fellow and/or graduate stude and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste unprecedented nature of what we propose, a smart 3D selection without Holo 3D selection tool, would be successes on their own, with immediate applicability

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

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

Selection in 3D is an unsolved proble



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

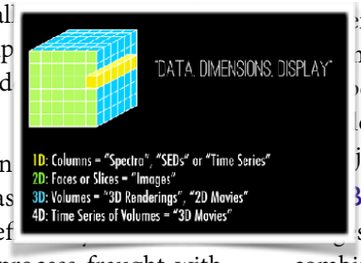
Microsoft
HoloLens

STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 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*

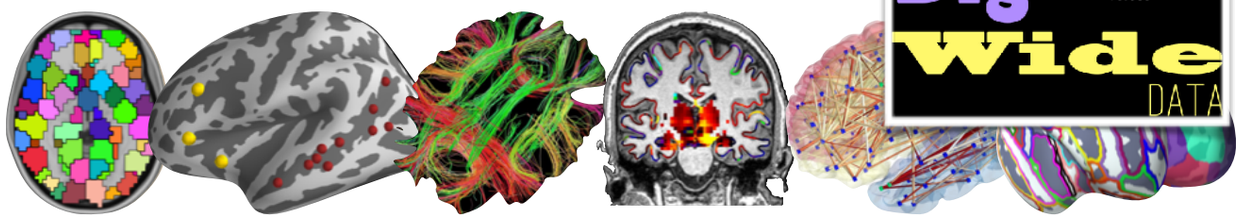
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

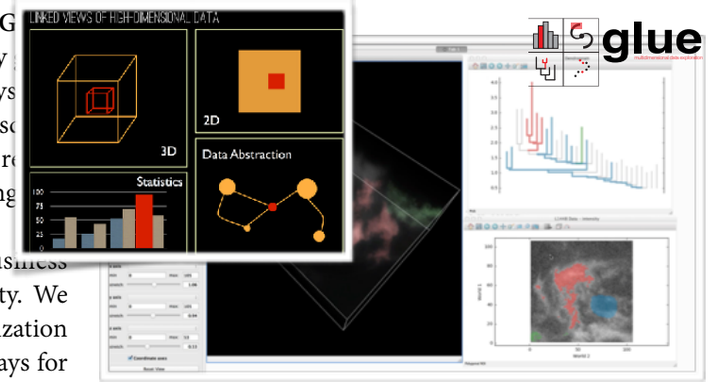
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

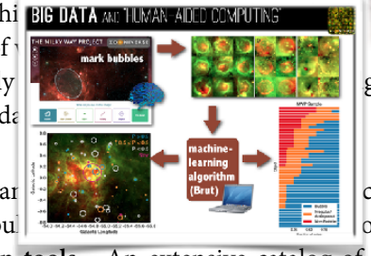
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.

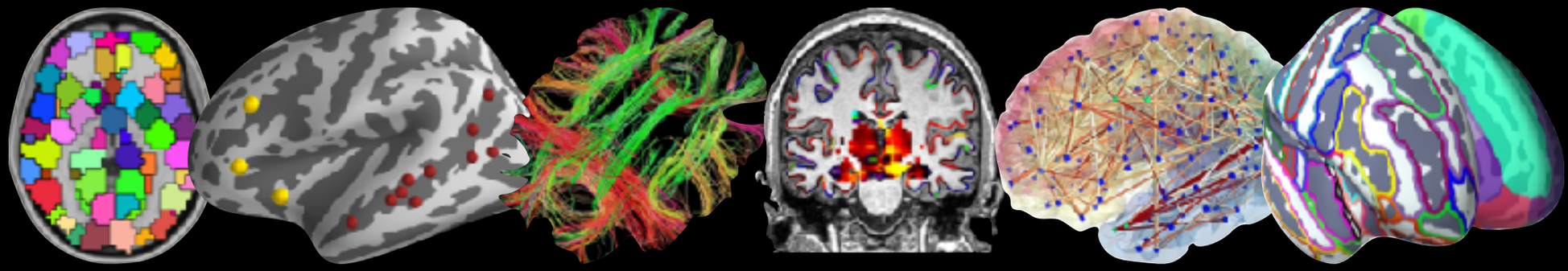


PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, we can **use experts' segmentations of brain imaging to train 3D selection tools**. An extensive catalog of such expert segmentation has already been assembled 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 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.

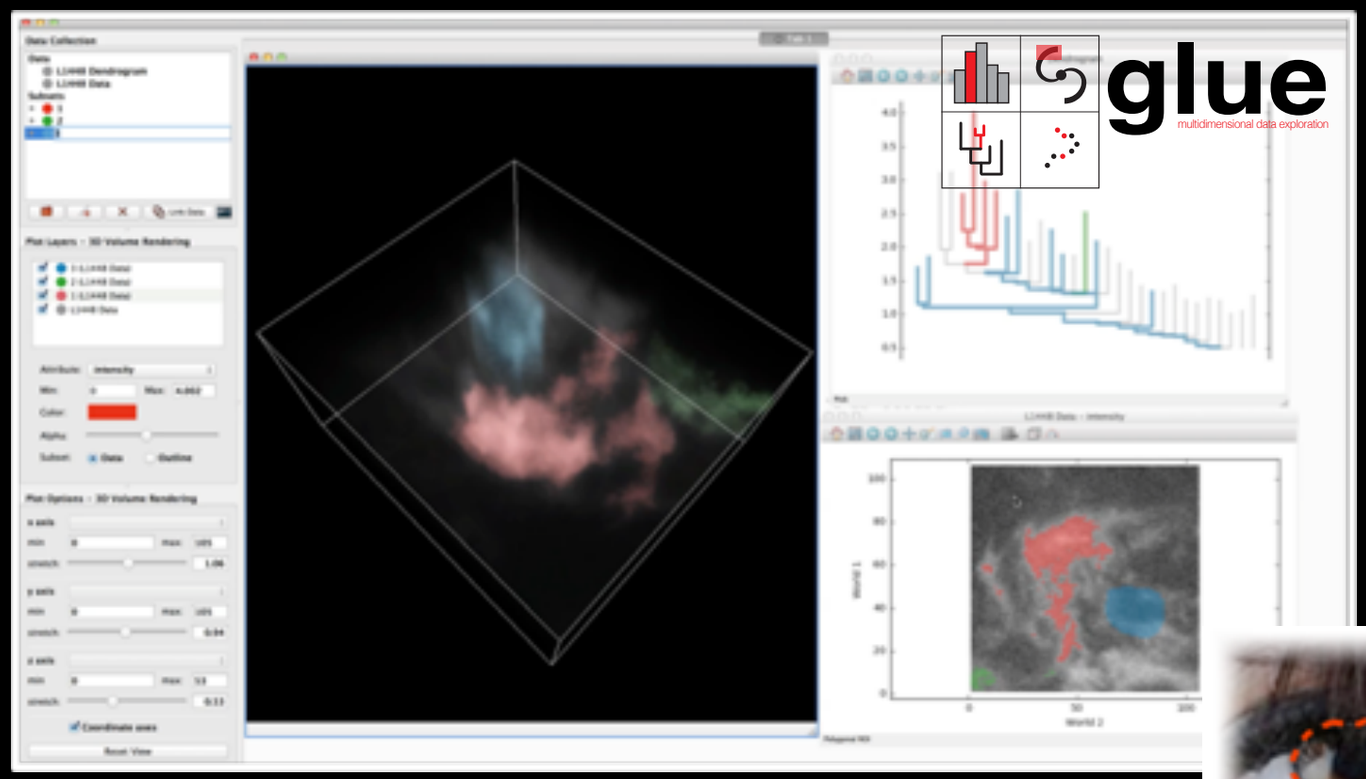
STAFFING, BUDGET, SUCCESS **Goodman** leads the Seamless Astronomy group at FAS, which develops Glue. She and Borkin founded the *Astronomical Medicine* project, of which this work is an outgrowth. **Borkin** is an expert on human-computer interaction. **Rosen**, trained as an MD-PhD physicist, leads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

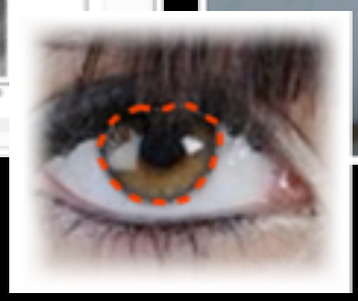
[1]See nitrc.org for a comprehensive searchable listing of neuroimaging projects and software.

[2]Beaumont, 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”



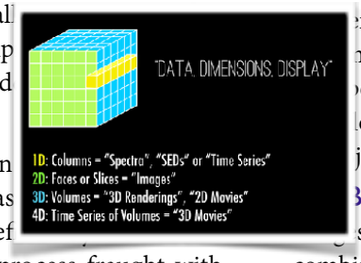


STAR FAMILY CHALLENGE FOR PROMISING SCIENTIFIC RESEARCH - SPRING 2016

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*

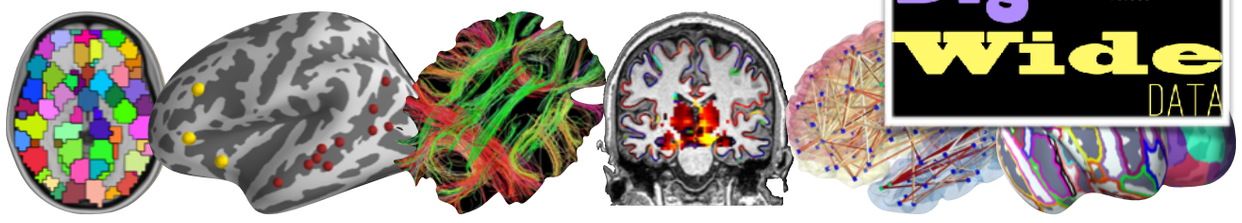
In spite of tremendous advances in computing, humans are still better than machines at finding patterns in visual data. We seek to leverage human pattern-recognition abilities with cutting-edge visualization, statistical and interaction techniques in order to save lives. Our groups' ongoing NASA- and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



BACKGROUND Identifying and quantitatively describing "regions of interest" (ROIs) in critical in both research and clinical work. This process, called "segmentation," is essential in as and in planning radiation therapy and robotic surgery. At present, 3D ROIs are manually def 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.

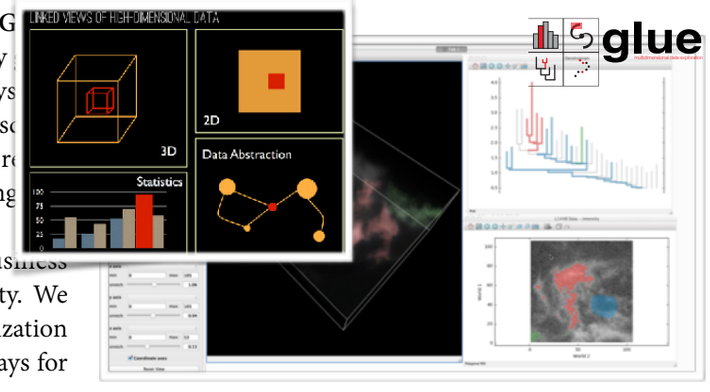
Three recent developments make the proposed research feasible now.

1. **Advances in Brain Imaging and Associated Software.** The capabilities of soft[1] ever-widening variety of neuroimaging modalities have grown tremendously o brain images below all use colored highlighting to indicate salient "features."



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 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).

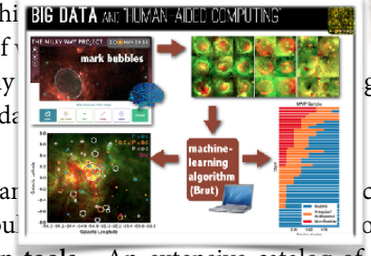
2. **Linked-view Visualization and the Advent of "G** The human visual perceptual system is especially at detecting change. So data visualization sys offering so-called "linked views," where one on-s map, chart or graph data set updates *live* to r selections made in another serve as insight eng Linked-views of tabular and mapping data become popular recently, especially in busi 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 device eract with virtual environments in more and more i nputers get more powerful and as profits grow. T portunity on the horizon is the potential of devices lil loLens to blend the real world, such as a person's real han jections, like the image of a human heart shown in the im

BLEM TO BE SOLVED Glue allows salient regions to b views (e.g. within s 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. Thi "selection" (outlining the object) and "segmentation" (the decision of In both 2D and 3D, the segmentation problem is solved for relatively For fuzzy features, though, in both astronomical and medical da algorithms alone at defining boundaries of salient features.



PROPOSED SOLUTION Our recent work[2] used the output of mar color images of nebulae to train a machine learning algorithm to find similar nebu o 3D, 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 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 expert in neuroimaging. **Kalpathy-Cramer** is an expert on quantitative image a *Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* project that project. **Goodman** and **Rosen** both have extensive leadership experience, and the one of their labs to host a part-time postdoctoral fellow and/or graduate stude and to purchase a HoloLens. Our **goal** is a working smart 3D selection syste unprecedented nature of what we propose, a smart 3D selection without Holo 3D selection tool, would be successes on their own, with immediate applicability

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

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

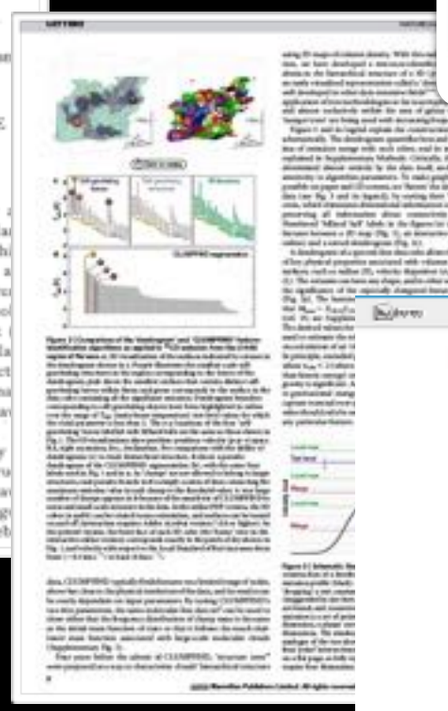
Communication

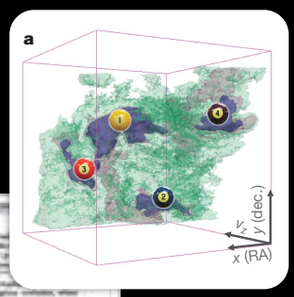
1610

1665

1895

2009





2015

WorldWide Telescope

glue
multidimensional data exploration

d3po

Au
Authorea

Au
Authorea

FEATURED ARTICLES ABOUT PLANS BLOG FEEDBACK HELP ALYSSA GOODMAN

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

The "Paper" of the Future

3

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

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

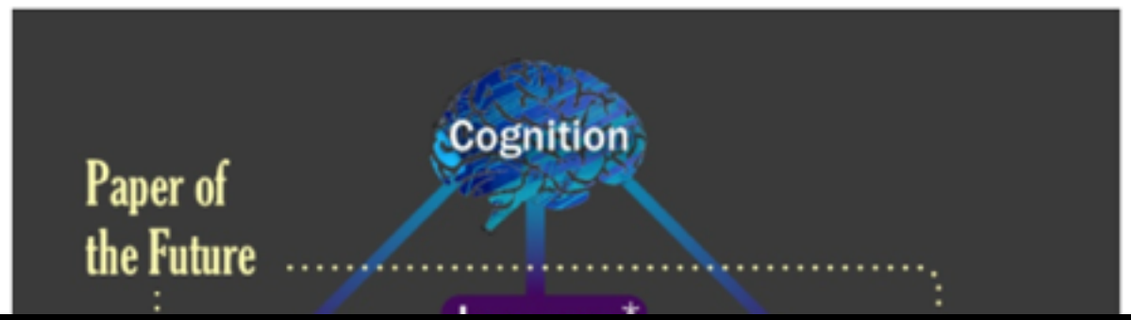
2

1 Preamble

A variety of research on human cognition demonstrates that humans learn and communicate best when more than one processing system (e.g. visual, auditory, touch) is used. And, related research also shows that, no matter how technical the material, most humans also retain and process information best when they can put a narrative "story" to it. So, when considering the future of scholarly communication, we should be careful not to do blithely away with the linear narrative format that articles and books have followed for centuries: instead, we should enrich it.

Much more than text is used to commuicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks or small discussions, these figures are often the focus of the conversation. In the best discussions, scientists have the ability to manipulate the figures, and to access underlying data, in real-time, so as to test out various what-if scenarios, and to explain findings more clearly. **This short article explains—and shows with demonstrations—how scholarly "papers" can morph into long-lasting rich records of scientific discourse**, enriched with deep data and code linkages, interactive figures, audio, video, and commenting.

0



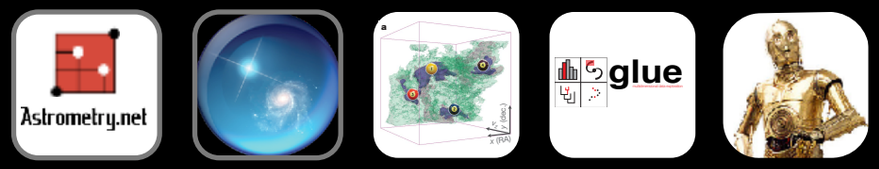
Konrad Hinsen 3 days ago · Public
Many good suggestions, but if the goal is "long-lasting rich records of scientific discourse", a more careful and critical attitude towards electronic artifacts is appropriate. I do see it concerning videos, but not a word on the much more critical situation in software. Archiving source code is not sufficient: all the dependencies, plus the complete build environment, would have to be conserved as well to make things work a few years from now. An "executable figure" in the form of an IPython notebook wil...
more

Merce Crosas 3 days ago · Public
Konrad, good points; this has been a concern for the community working on reproducibility. Regarding data repositories, Dataverse handles long-term preservation and access of data files in the following way: 1) for some data files that the repository recognizes (such as R Data, SPSS, STATA), which depend on a statistical package, the system converts them into a preservation format (such as a tab/CSV format). Even though the original format is also saved and can be accessed, the new preservation format gua...
more

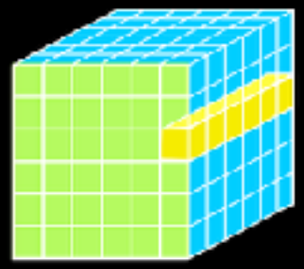
Konrad Hinsen 1 day ago · Public
That sounds good. I hope more repositories will follow the example of Dataverse. Figshare in particular has a very different attitude, encouraging researchers to deposit as much as possible. That's perhaps a good strategy to change habits, but in the long run it could well backfire when people find out in a few years that 90% of those deposits have become useless.

Christine L. Borgman 4 months ago · Private
"publications"

d3po/Authorea: Peek, Price-Whelan, Pepe, Beaumont, Borkin, Newton; PoF: Goodman, Peek; WWT: Wong, Fay et al.; Astrometry.net: Hogg, Lang, Roweis et al.



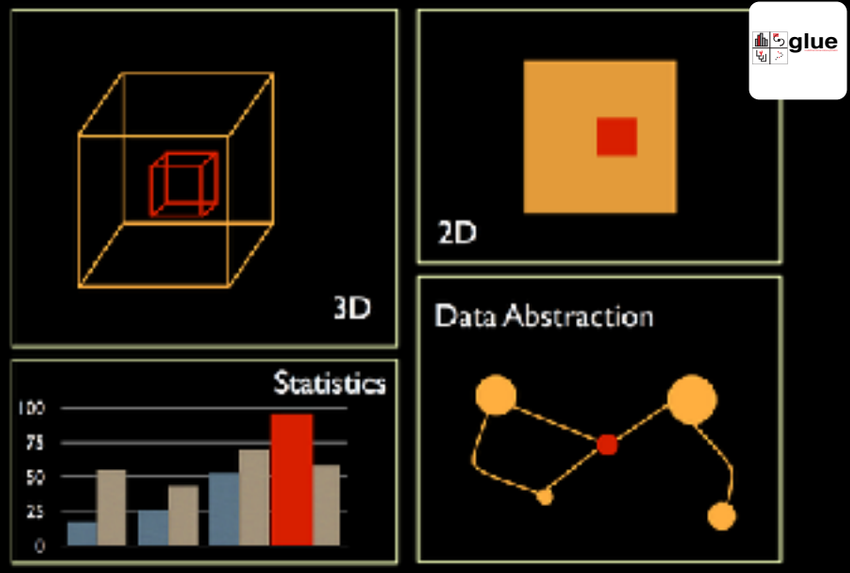
[video, demos]



"DATA. DIMENSIONS. DISPLAY"

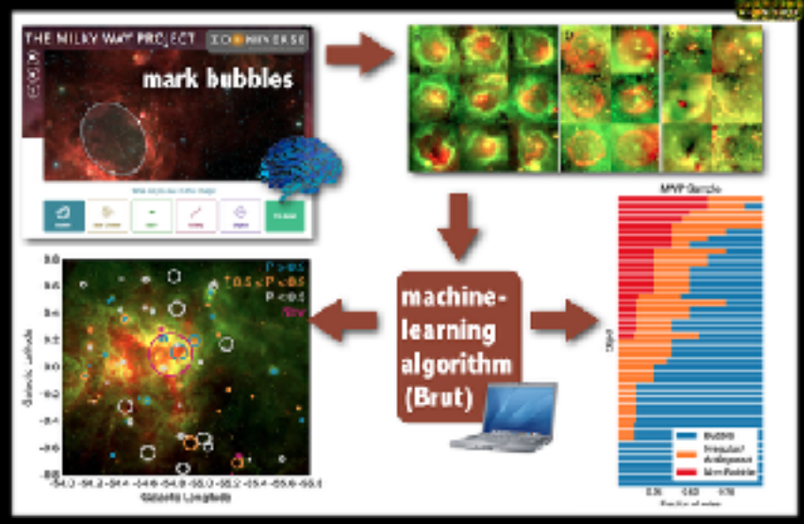
1D: Columns = "Spectra", "SEDs" or "Time Series"
2D: Faces or Slices = "Images"
3D: Volumes = "3D Renderings", "2D Movies"
4D: Time Series of Volumes = "3D Movies"

LINKED VIEWS OF HIGH-DIMENSIONAL DATA

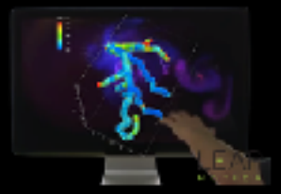


Big DATA versus Wide DATA

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"

COMMUNICATION

≡ MENU

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)