While you are waiting...here’s an in-the-works [iPhone] App on “Uncertainty” being made in conjunction with my course & a project on “Prediction” we’re doing at WGBH...

**main screen**

- **Temperature Distribution**: The graph shows a sample “distribution” of how far off temperature predictions have been in the past. Perfect predictions give zero as a difference value. The shaded bars just summarize the graph: the darker the shaded bar, the more predictions fell in that difference zone.

**info screen**

- **Number of Occurrences**: The graph below shows the distribution of temperature predictions. Perfect predictions give zero as a difference value. The shaded bars summarize the graph: the darker the shaded bar, the more predictions fell in that difference zone.

- **To learn more about this App and about how predictions are made, visit [prediction.wgbh.org](http://prediction.wgbh.org).**
From Baby Pictures to Baby Stars: What Scientists Can See

What kind of credentials are those??

Alyssa A. Goodman
Harvard University (HCO+IIC)
Smithsonian Astrophysical Observatory
Scholar-in-Residence, WGBH
19 out of 22?
You & I can see what I will show thanks to...

**Astronomical Medicine & 3D PDF:** Mike Halle, Michelle Borkin, Jens Kauffmann, Doug Alan, Ron Kikinis, Erik Rosolowsky, Nick Holliman, Jonathan Foster, Jaime Pineda, Héctor Arce, Dave Kennedy, Mark Thomas, Timo Hannay & Phil Campbell

**WorldWide Telescope:** Curtis Wong, Jonathan Fay & Gus Muench + MSR supporters

**Touch:** Chia Shen & Hanspeter Pfister

**Image & Meaning:** Felice Frankel & Ros Reid
These limit what we “can” see.
My Fuller Meaning of “Tools”

(hardware)

(nerdware)

(Nature?!)
Can this process ever be generalized in a “theory of data graphics”?

Shall we discuss that later??
Baby Stars: What Scientists “Can” See
What Astronomers “Can” See
as distinguished from what they “Cannot” vs. “Could” See

**Can**
- 2D (our sky)
- static graph(ic)s
- their own data
  - *baby pix*

**Cannot**
- aliens’ views
- N/A
- proprietary data
  - *matches to adult pix*

**Could**
- 3D proxies, 4D
  - Astronomical Medicine
- interactive graph(ic)s
  - 3D PDF, “3D” DataDesk w/selector
- “free” data
  - WWT, Google Sky
- with practice?...
Star (and Planet, and Moon) Formation 101

AG’s focus...

Demo

Number of Stars of each Mass

Stellar Mass

Galaxy

Molecular Cloud Complex

Star-Forming "Globule"

Circumstellar Disk

Extrasolar System

Star Cluster
What Astronomers “Can” See
as distinguished from what they “Cannot” vs. “Could” See

**Can**
- 2D (our sky)
- static graph(ic)s
- their own data
- baby pix

**Cannot**
- aliens’ views
- N/A
- proprietary data
- matches to adult pix

With practice?...

Goodman et al., Astrophysical Journal, 2009
Are creativity and discovery held back from "could" by confining technological tools?
Astronomical Medicine

Alyssa Goodman (IIC/CfA/FAS)
Michael Halle (IIC/SPL/HMS)
Ron Kikinis (SPL/HMS)
Douglas Alan (IIC)
Michelle Borkin (FAS/IIC)
Jens Kauffmann (CfA/IIC)
Erik Rosolowsky (CfA/UBC Okanagan)
Nick Holliman (U. Durham)
The Astronomical Medicine Story

“Viz has failed the scientific community…”

Computer Scientist

Astronomer

Unsuspecting Undergrad

+Nick Holliman (CS, 3D expert)
+Doug Alan (S/W Engineer)
+Jens Kauffmann (postdoc)
+Erik Rosolowsky (postdoc) + ...
COMPLETE = COordinated Molecular Probe Line Exinction Thermal Emission

- 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. in prep.)
- Optical image (Barnard 1927)
"Astronomical Medicine"

"KEITH"

"PERSEUS"

"z" is depth into head

"z" is line-of-sight velocity

(This kind of “series of 2D slices view” is known in the Viz as “the grand tour”)
What can we observe?

\[ I(E, s, \vec{x}, t) \]

Intensity

Energy (wavelength)

"State" (e.g. polarization)

Spatial Position \((x, y, z)\)

Time

...and the science is in the interpretation of these measurements into physical quantities & processes.
What can we observe?

$$I(\vec{x})$$

Intensity

Spatial Position

$$(x, y)$$

Optical Single-Band Image of NGC1333
What can we observe?

\[ I(\vec{x}) \]

Intensity

Spatial Position
\( (x, y) \)

X-Ray of Human Skull, c. 1920
What can we observe?

\[ I(\vec{x}, t) \]

- Intensity
- Spatial Position \((x, y, z)\)
- Time

“Nova Stella” of Tycho, 1572
What can we observe?

\[ I(\vec{x}, t) \]

- **Intensity**
- **Spatial Position** \((x, y, z)\)
- **Time**

Cardiac Motion
What can we observe?

\[ I(E, \vec{x}) \]

- Intensity
- Energy (wavelength)
- Spatial Position \((x, y)\)

Optical (B,V,R) image of NGC1333
What can we observe?

$I(E, \vec{x})$

- Intensity
- Energy (wavelength)
- Spatial Position $(x, y)$

Human Ear, Thermal Infrared
What can we observe?

\[ I(s, \vec{x}) \]

Intensity

Spatial Position \((x, y)\)

“State” (e.g. polarization)

Hall 1949

Fig. 4. Observational evidence that there is no one preferential orientation of the plane of polarization. Stars showing no polarization are represented by circles.
What can we observe?

\[ I(s, \vec{x}) \]

Intensity

Spatial Position \((x, y, z)\)

“State” (~diffusivity)
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. in prep.)

Optical image (Barnard 1927)

Perseus
Some of What We’ve Discovered...

Cores nest in cocoons (Kauffmann et al. 2009)

Gravity Matters (Goodman et al. 2009)

Tripled Outflows (Borkin et al. 2008, Arce et al. 2009a)

Shells Rule (Arce et al. 2009b)
“Seeing” the Role Self-Gravity in Star Formation

Figure 2 | Comparison of the ‘dendrogram’ and ‘CLUMPFIND’ feature-identification algorithms as applied to 13CO emission from the L1448 region of Perseus. A, 3D visualization of the surface illuminated by clumpy clouds. B, a selected plane of the dendrogram shown in A. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the dendrogram. The volumes can have any shape, and in other work focus on the significance of the especially elongated features seen in L1448 (ref. 13; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an ‘observed’ virial parameter, \( \frac{M_{\text{obs}} \sigma^2}{R_{\text{lum}}} \). In principle, extended portions of the tree (Fig. 2, yellow highlighting) where \( \frac{M_{\text{obs}} \sigma^2}{R_{\text{lum}}} \) is less than 2, where gravitational energy is comparable to or larger than kinetic energy, correspond to regions of \( p-p' \)-space where self-gravity is significant. As \( \frac{M_{\text{obs}} \sigma^2}{R_{\text{lum}}} \) only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields, its measured value should only be used as a guide to the elongated/boundaries of any particular feature.

Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by dropping a one-dimensional emission profile (purple) from above in tiny steps (exaggerated in size here, top left) and then merging the local maxima and mergers are found and connected as shown. The intersection of a test level with the merged region of maxima and local maxima is the dendrogram. In the final dendrogram, a plane curve is two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 3 is the 3D analogue of the tree shown here, only constructed from ‘nonleaf’ rather than ‘point’ intersections. It has been cut and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

Notes:
- CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND’s two free parameters, the same molecular line data set can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).
- Four years before the advent of CLUMPFIND, ‘structure trees’ were proposed as a way to characterize clouds’ hierarchical structure.
Interactivity for the Future
(from "Could" to "Can")
“Data Desk”

If only DataDesk were >2D...??
3D selection tools (& interaction) are challenging
The (secret) inherent in column density mapping.

Real Slide from Real Astronomy Conference in 2006

Goodman, Pineda & Schnee, 2007; see also Pineda, Caselli & Goodman 2007.
Goodman, Pineda & Schnee, 2009; see also Pineda, Caselli & Goodman 2008.
Mirage (Bell Labs)

cf. Avizo (Mercury Systems); some aspects of GenePattern; Taverna...
What we **can** (and others **could**)
see off the desktop...

http://iic.harvard.edu/research/scientists-discovery-room-lab-sdr-lab

http://spectrum.ieee.org/dec08/6999/9
The Scientists’ Discovery Room
(Shen & Pfister)

movie courtesy Daniel Wigdor, equipment now in Chia Shen’s SDR lab at Harvard SEAS
The Modular, Personalizable, Approach we “Can”, “Could” (& Should!?) Take to Interactions
“Made with Processing” (see Reas & Fry 2006)

DendroStar (Douglas Alan)
HH46/47

This image from NASA's Spitzer Space Telescope transforms a dark cloud into a silky translucent veil, revealing the molecular outflow from an otherwise hidden newborn star. Using near-infrared light, Spitzer pierces through the dark cloud to detect the embedded outflow in an object called HH 46/47. Herbig-Haro (HH) objects are bright, nebulous regions of gas and dust that are usually buried within dark clouds. They are formed when supersonic gas ejected from a forming protostar, or embryonic star, interacts with the surrounding interstellar medium. These young stars are often detected only in the infrared.

The Spitzer image was obtained with the infrared array camera. Emission at 3.6 microns is shown as blue, emission from 4.5 and 5.8 microns has been combined as green, and 8.0 micron emission is depicted as red.

HH 46/47 is a striking example of a low-mass protostar ejecting a jet and creating a bipolar, or two-sided, outflow. The central young star is positioned on the right side of the jet, and the outflow is seen as a series of shock waves emanating from the star.
From Baby Pictures to Baby Stars: What Scientists Can See

Shall we generalize?...
Discussion: Can this process ever be generalized in a "theory of data graphics"?
Objectivity

Generality

e.g. baseball box scores

“cartesian” plots in 2D, 3D, e.g. an EKG

“automated” (theory-based) graphics systems

“happiness” indicator in economics

“aesthetics”
Another approach: “A Virtual Graphical Collaborative” a-la-Frankel

(graphics courtesy David Curry, participant at IM2 “Group A” at Apple in 2007)
IM2.3 Prototype Visualization Rubric 1.0 February 2007

Tool Schematic

IM2 Visualization Rubric
Considerations
Click here to explore the "Interactivity" parameter, since these values have not been defined for this rubric.

Low
High

Overall Success

General Strategies

A: Practitioner Selects "Interactivity"

Executional Strategies

B: "Interactivity" Provides Overview, Best Practice Examples

Constraints/Contextual Issues

C: Practitioner drills-down on best practice example

Click on the images or use the nav above right to explore these pages further.

graphics courtesy David Curry, participant at IM2 "Group A" at Apple in 2007.
extra slides
data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By using 2D maps of column density with this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p-v) data cube into an easily visualized representation called a 'dendrogram' \(^5\). Although well developed in other data-intensive fields\(^6\), it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where ‘merger trees’ are being used with increasing frequency\(^7\).

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we ‘flatten’ the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their ‘branches’ to not cross, which eliminates dimensional information on the x axis while preserving all information about connectivity and hierarchy. Numbered ‘billiard ball’ labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius (\(R\)), velocity dispersion (\(v\)) and luminosity (\(L\)). The volumes can have any shape, and in other work\(^8\) we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that \(M_{\text{lum}} = X_{\text{obs}}X_{\text{lum}}\) where \(X_{\text{lum}} = 8.0 \times 10^{-5}\) cm\(^3\) K\(^{-1}\) km\(^{-1}\) s\(^{-1}\) (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an ‘observed’ virial parameter, \(x_\text{vir} = 5\sqrt{8}\, GM_{\text{obs}}\).

In principle, extended portions of the tree (Fig. 2, yellow highlighting) where \(x_\text{vir} < 2\) (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of \(p\)-\(v\) space where self-gravity is significant. As \(x_\text{vir} < 2\) only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields\(^9\), its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.
Galileo’s Moons + Technology

Notes for & re-productions of Siderius Nuncius (1610)