Seeing Science
Data Visualization in Modern Research (and teaching!)

The Art of Numbers
Professor Alyssa A. Goodman (Astronomy)
Course website
Duration: 05:30

What kind of credentials are those??

Alyssa A. Goodman
Harvard University (HCO+IIC)
Smithsonian Astrophysical Observatory Scholar-in-Residence, WGBH
19 out of 22?
Relative Strengths

Pattern Recognition
Creativity

Calculations
“Interocularity”
(see work of John Tukey)

“Image and Meaning”
(see work of Felice Frankel, and imageandmeaning.org)
“Image and Meaning”
What...

...is easier now than before?

fast computation, animation, 3D

...was easier before than now?

craftsmanship

...should be easier in the future?

modular craftsmanship, linked views
“Easier”
Craftsmanship
(in 1854)
Data • Dimensions • Display
Craftsmanship (in 1854)

Displaying “high-dimensional” data with “multi-functioning graphical elements”

What Computers *Can* Let us Craft

**Elements...**

✓ Maps
✓ Tables
✗ Graphs
✓ Charts
✓ Illustrations
✓ Combinations
Milestones: Time course of developments

Galileo Galilei
(1564-1642)

Notes for & re-productions of Siderius Nuncius
William Playfair
(1759-1823)
Charles Joseph Minard, in color
(1781-1870)
Data • Dimensions • Display
“High-dimensional” or “Multivariate” Data and High(er) Dimensional Displays

This map displays 2 quantities as a function of 2 spatial dimensions. ...Is that 4 dimensions?
“High-dimensional” or “Multivariate” Data
(Astronomy=Biology)


How much are we held back today by digital tools?
How can we advance the **digital** tools for scientists?
Astronomical Medicine

am.iic.harvard.edu

Alyssa Goodman (IIC/CfA/FAS)
Michael Halle (IIC/SPL/HMS)
Ron Kikinis (SPL/HMS)
Douglas Alan (IIC)
Michelle Borkin (IIC)
Jens Kauffmann (CfA/IIC)
Erik Rosolowsky (CfA)
Nick Holliman (U. Durham)
Michelle Borkin is now a SEAS PhD Student, advised by Profs. Alyssa Goodman (Astronomy) and Hanspeter Pfister (SEAS), and IIC +AstroMed became the bases for the Viz-e-Lab.
Perseus

- 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”)
Perseus

- 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

3D Viz made with VolView

Astronomical Medicine @ IICT

COMPLETE
Astronomers are not alone...

*high-D data sets are everywhere in science*

**Brain Science**

---

**Astrophysics**

---

Buckner et al. 2005

---

Goodman et al. 2007

---

Borkin et al. 2011
What...

...is easier now than before?

*fast computation, animation, 3D*

...was easier before than now?

*craftsmanship*

...should be easier in the future?

*modular craftsmanship, linked views*
The “Easier” Future, for Everyone
Modular Craftsmanship & Linked Views

The Future we can see from “now”...

more display modes available (3D PDF, touch interfaces, stereo+)
re-usable tools/mashups (Many Eyes, Google Maps+, crowdsourcing)
live, interactive linked views (DataDesk, Tableau, GapMinder, WWT...)

 Unsolved Questions...

(feasibility of) templates/language (e.g. Grammar of Graphics)
improved graphical representation of uncertainty
Figure 2 | Comparison of the 'denogram' and 'CLUMPFIND' feature-identification algorithms as applied to N-dimension from the L1544 region of Perseus. a, 3D visualization of the surface indicated by colours in the denogram shown in c. b, Shows the smallest scale self-gravitating structures in the region corresponding to the leaves of the denogram, which shows the smallest surface that contains distinct self-gravitating leaves within them and (c) Corresponds to the surface in the data cube containing all of the significant emission. Denogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of $T_{MB}$ (main beam temperature) test-values for which the virial parameter is less than 2. The x-y locations of the four 'self-gravitating' leaves labeled with billiard balls are the same as those shown in Fig. 1. The x, y, z, u, v, w coordinates show position–position–velocity (p–p–v) space. R, A, D, position, declination, declination. For comparison with the ability of denograms (c) to track hierarchical structures, d shows a pseudodendrogram of the CLUMPFIND segmentation (b), with the same four labels used in Fig. 1 and in a. As a 'clump' is not allowed to belong to larger structures, each pseudo-branch in d is simply a set of lines connecting the maximum emission value in each dump to the threshold value. A very large number of clumps appears in b because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cube (a and b) can be rotated for arbitrary orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.6.8 or higher). In the printed version, the front face of each 3D cube (the 'here' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the local standard of rest increases from front (−0.5 km s$^{-1}$) to back (8 km s$^{-1}$).

Figure 3 | Schematic illustration of the denogram formation process. Shown is the construction of a denogram from a hypothetical one-dimensional emission profile (shaded). The denogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and minima are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example, the light purple dots) in one dimension, a plane curve in two dimensions, and an isosurface in three dimensions. The denogram of 3D data shown in Fig. 3c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' interactions. It has been sorted and flattened for representation on a flat page, as fully representing denograms for 3D data cubes would require four dimensions.
“Off the Desktop”

UBITABLE: Users can interact with surface computers through auxiliary devices, such as laptops, phones, and PDAs. The display on the auxiliary device can convey private or sensitive content to a single user, while group-appropriate content can appear on the tabletop display. Chia Shen and her colleagues at Mitsubishi Electric Research Laboratories, in Cambridge, Mass., have explored auxiliary interactions with surface computers in their UbiTable project, in which two people with laptops collaborate over a tabletop display. Recently, Shen expanded the UbiTable into an interactive room called the WeSpace. People can share data on their laptops with other people in the room, using both a table and a large display wall. Here, three Harvard University astrophysicists discuss radio and IR spectrum images using the WeSpace.

movie courtesy Daniel Wigdor, taken at MERL, Kendall Square, Cambridge, 2007
How can we advance the digital tools for scientists?
“Seamless Astronomy” and “Linked Views”

Contextual, 
High-Dimensional 
View

Flat, 
Text-Based 
View

Link
DataDesk (est. 1986)
John Tukey’s “Four Essentials” (c.1972)

Picturing  Rotation  Isolation  Masking  Selection

and these “need to work together” in a “dynamic display”

Results...
1. for immediate insight
2. as visual source of ideas for statistical algorithms (...relation to SVM)

Warning
“details of control can make or break such a system”

Watch the PRIM-9 video at: http://stat-graphics.org/movies/prim9.html
Exemplar: **Linked Dendrogram Views in IDL**

Video & implementation: Christopher Beaumont, CfA/UHawaii;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky
How can we advance the digital tools for scientists?
Seamlessly explore imagery from the best ground and space-based telescopes in the world.

Expert led tours of the Universe

Control time to study how the night sky changes.

View and compare images from across the electromagnetic spectrum.

Finder Scope links to Wikipedia, publications, and data, so you can learn more.

Context bar shows items of interest in current field of view.

Context globe shows where you’re looking.

Microsoft® Research
WorldWide Telescope

Experience WWT at worldwidetelescope.org
“Seamless Astronomy”...
astrometry.net + flickr + WWT

[published 1927]

ask me about ADSASS...
WWT
Ambassadors: WorldWide Telescope For Interactive Learning

Alyssa Goodman
Harvard University Professor of Astronomy, Microsoft Academic Partner
Pat Udomprasert
WWT Program Coordinator
Curtis Wong
Microsoft Research, WWT Creator

Stephen Strom
NOAO, WWTA Tucson Site Advisor
Sarah Block
Web site development
About the WWT Telescope Ambassadors Program

WorldWide Telescope (WWT) is a rich visualization environment that functions as a virtual telescope, allowing anyone to make use of professional astronomical data to explore and understand the universe. As of early 2010, the new WWT Ambassadors Program is recruiting astronomically-literate volunteers, including retired scientists engineers—all of whom will be trained to be experts in using WWT as a teaching tool. Ambassadors will give volunteer presentations at public libraries, community centers, museums, and schools, demonstrating WWT's power to help laypeople visualize and understand our universe.

Read more

John Huchra's Universe

Submitted by patudom on Jan. 11


John's colleagues at the Harvard-Smithsonian Center for Astrophysics, in collaboration with the creators of WorldWide Telescope at Microsoft Research, have created a new, interactive, WWT Tour to honor John and his career. The Tour primarily focuses on John's quest to map the Universe in three dimensions. It is 12.5 minutes long.

The Tour is best experienced inside the WorldWide Telescope program itself. (Note: You must have the version of WWT released on 1/13/2011 to view all of this Tour's content. You can download it from here.) As viewed within the WWT program, the Tour content is interactive, allowing users to pause and explore the parts of the Universe featured in the tour, explore web hyperlinks, and more. For those who do not have the desktop client, the Tour has been posted as a video as well.

Video (Interactive WWT features will be disabled)
Gains in Student Interest and Understanding
(“Traditional Way” vs “WWT Way”)

“Cooler than ‘Call of Duty’”
Seeing Science

Data Visualization in Modern Research (and teaching!)

The Art of Numbers

Professor Alyssa A. Goodman (Astronomy)
Course website
Duration: 05:30

Alyssa A. Goodman
Harvard University (HCO+IIC)
Smithsonian Astrophysical Observatory
Scholar-in-Residence, WGBH
The next challenges...
Challenge #1: 3D Selection

What is the 3D “magnetic lasso”?

How do you use it with a mouse?

How can a human “steer” computer-aided selection?
Challenge #2: Too many windows...
Challenge #3:

What does “Publication-Quality” Graphics Mean in an Interactive 3D World?

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and in results can be overly dependent on input parameters. By turning 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 functions associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, ‘structure trees’ to track hierarchical structure, a pseudo-branch in the dendrogram represented the minimum emission in each clump to the threshold value. A very large number of clumps appear in order because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (a and b) can be rotated in any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the ‘home’ view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front (8 km s$^{-1}$) to back (8 km s$^{-1}$).

Although well developed in other data-intensive fields, 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.

In principle, extended portions of the tree (Fig. 2, yellow highlighting) that ‘merge’ with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we ‘flatten’ the dendrogram of 3D data (see Fig. 3 and its legend), by setting their ‘branches’ to not cross, which eliminates dimensional information on the z 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 surgical 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-

"Self-gravitating" leaves within them; and green corresponds to the surface in the dendrogram shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front (8 km s$^{-1}$) to back (8 km s$^{-1}$). In the online PDF version, the 3D cubes (a and b) can be rotated in any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the ‘home’ view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front (8 km s$^{-1}$) to back (8 km s$^{-1}$).
