## SGENEE

## BEYONB



FLLATLAND:

Alyssa A. Goodman • Harvard-Smithsonian Center for Astrophysics • @aagie

flightaware.com/miserymap, by David Chouinard

BIG DATA WIDE DATA<br>DMENSIONALTY<br>LINKED VEWS<br>NTERACTION<br>COMMUNCATION EDUCATION



## GALIEO GALIEI (1564-1642)



On the third, at the seventh hour, the stars were arranged in this juence. The eastern one was I minute, 30 seconds from Jupiter closest western one 2 minutes; and the other western one wa
ast * * * Wer

0 minutes removed from this one. They were absolutely on th ame straight line and of equal magnitude.
On the fourth, at the second hour, there were four stars arout upiter, two to the east and two to the west, and arranged precisel
East $* * \bigcirc \quad * \quad * \quad$ Wes
in a straight line, as in the adjoining figure. The easternmost wa tistant 3 minutes from the next one, while this one was 40 second rom Jupiter: Jupiter was 4 minutes from the nearest western one d this one 6 minutes from the westernmost one. Their magnitude are nearly equal; the one closest to Jupiter appeared a little smaller an the rest. But at the seventh hour the eastern stars were only o seconds apart. Jupiter was 2 minutes from the nearer easteri
East ** * * Wes
me, while he was 4 minutes from the next western one, and this one was 3 minutes from the westernmost one. They were all equal and extended on the same straight line along the ecliptic.

On the fifth, the sky was cloudy.
On the sixth, only two stars appeared flanking Jupiter, as is seen

$$
\text { East } *<\quad * \quad \text { West }
$$

in the adjoining figure. The eastern one was 2 minutes and the vestern one 3 minutes from Jupiter. They were on the same straight fine with Jupiter and equal in magnitude.
On the seventh, two stars stood near Jupiter. beth to the east.

## RESOLUTON \& CONTEXT



## GALILEO'S "NEW OR <br> Created by Alyssa Goodman, Curtis Won with advice from Owen Gingerich and I



Galileo's New Order, A WorldWide Telescope Tour by Goodman, Wong \& Udomprasert 2010


## RESOLUTION \& CONTEXT + DIMENSIONALITY



January 11, 1610


## STAR \& PLANET FORMATON N 1 SLDE












## BIG DATA and HUMAN-ADED COMPUTNG


example here from: Beaumont, Goodman, Kendrew, Williams \& Simpson 2014; based on Milky Way Project catalog (Simpson et al. 20 I3), 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 HIMAN-ADED COMPUTNG



## BIG DATA aid HIMAN-ADED COMPUTNG


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.)


## $V \sqrt{V} \square A \rightarrow A$



[^0]
## WIDE DATA

## C ${ }^{(2) M P L E T E}$

 mm peak (Enoch et al. 2006)sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
${ }^{13} \mathrm{CO}$ (Ridge et al. 2006)
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)

Optical image (Barnard I927)

 +2. *2 x 3.3



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

## ADDNG A THIRD DMENSION



## VELOCITY FROM SPECTROSCOPY

All thanks to Doppler


## SPECTRAL-LINE MAPPNG GVES A "THRD" DMENSON

We wish we could measure..


But we con measure...


## THIRD DIMENSION OFTEN HIDDEN

Spectral Line Observations



## "DATA. DIMENSIONS. DISPLAY"

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

## WUE OATA NEDD

mm peak (Enoch et al. 2006)
sub-mm peak (Hatchell
et al. 2005, Kirk et al. 2006)
${ }^{13} \mathrm{CO}$ (Ridge et al. 2006)mid-IR IRAC composite
from c2d data (Foster,
Laakso, Ridge, et al.)
: Optical image (Barnard 1927)


## RESOLUTION. CONTEXT. DIMENSIONALITY. WIDE DATA



## COHERENT CORES <br> "ISLANDS OF CALM N TURBULENT SEAS"(?)



30-year story: Myers \& Benson 1983, Goodman et al. 1998, Pineda et al. 2010, 2011, 2014

## २०10: "HIGH" RESOLUTION $\rightarrow$ EVIDENCE FOR "COHERENCE" IN DENSE CORES



## DIMENSONALITY: COHERENT CORE BURED WITHIN B5


many thanks to Jaime Pineda \& Jens Kauffmann for this figure COMPLETE data: ${ }^{13} \mathrm{CO}$ from Ridge et al. 2006; $\mathrm{NH}_{3}$ from Pineda et al. 2010

## EVEN HIGHER RESOLUTION... UNEXPECTED SUB-STRUCTURE?!

The Astrophysical Journal Letters, 739:L2 (5pp), 2011 September 20


Figure 1. Left panel: integrated intensity map of B5 in $\mathrm{NH}_{3}(1,1)$ obtained with GBT. Gray contours show the 0.15 and $0.3 \mathrm{~K} \mathrm{~km} \mathrm{~s}^{-1}$ level in $\mathrm{NH}_{3}(1,1)$ integrated intensity. The orange contours show the region in the GBT data where the non-thermal velocity dispersion is subsonic. The young star, B5-IRS1, is shown by the star in both panels. The outflow direction is shown by the arrows. The blue contour shows the area observed with the EVLA and the red box shows the area shown in the right panel. Right panel: integrated intensity map of B5 in $\mathrm{NH}_{3}(1,1)$ obtained combining the EVLA and GBT data. Black contour shows the $50 \mathrm{mJy} \mathrm{beam}^{-1} \mathrm{~km} \mathrm{~s}^{-1}$ level in $\mathrm{NH}_{3}(1,1)$ integrated intensity. The yellow box shows the region used in Figure 4 . The northern starless condensation is shown by the dashed circle.

## BUT MAYBE IT'S DFFERENT?



## SHHH. .THS WILL APPEAR IN NATURE, TOMORROW

## What if flaments continue across "core" boundaries?!

blue =VLA ammonia (high-density gas); green=GBT ammonia (lower-res high-density gass; red-Herschel 250 micron continuum (dust)

## RESOLUTION. CONTEXT. WIDE DATA




## 2009 3D PDF INTERACTVITY N A " PAPER"



Figure $\mathbf{2}$ | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-
identification algorithms as applied to ${ }^{13}$ CO emssion from identification algorithms as applied to ${ }^{13} \mathrm{CO}$ emission from the L 1448
region of Perseus. a , 3D visualization of the surfaces indicated by colours region of Perseus. a, 3D visualization of the surfaces indicated by colou
the dendrogram shown in $c$. Purple illustrates the smallest scale selfgravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct selfgravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of $T_{\mathrm{mb}}$ (main-beam temperature) test-level values for whic
the virial parameter is less than 2. The $x-y$ locations of the four selfgravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position-position-velocity ( $p-p-v$ ) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (c) to track hierarchical structure, d shows a pseudodendrogram of the CLUMPFIND segmentation (b), with the same four
labels used in Fig. 1 and in a. As 'clumps' are not allowed to belong to larger labels used in Fig. 1 and in a. As clumps are not allowed to belong to larger
structures, each pseudo-branch in dis simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in $\boldsymbol{b}$ 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 $\mathbf{b}$ ) can be rotated to 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 $\left(-0.5 \mathrm{~km} \mathrm{~s}^{-1}\right)$ to back $\left(8 \mathrm{~km} \mathrm{~s}^{-1}\right)$.
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 tuning CLUMPFIND's two free parameters, the same molecular-line data set ${ }^{8}$ 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 shal (Supplementary Fig. 1)
Four years before the advent of CLUMPFIND, 'structure trees' were proposed as a way to characterize clouds hierarchical structure
using 2D maps of column density. With tt $\quad{ }^{-1 v}$ 2D work as inspiration, we have developed a structure-id abstracts the hierarchical structure of a an easily visualized representation calle well developed in other data-intensive application of tree methodologies so for and almost exclusively within the merger trees' are being used with is Figure 3 and schema cmission merge with que malained in Supple with determined almost entirely by tt sensitivity to algorithm paramet possible on paper and 2D screen data (see Fig. 3 and its legend cross, which eliminates dimen preserving all information Numbered 'billiard ball' lab features between a 2D map online) and a sorted dendri A dendrogram of a spectr of key physical properties surfaces, such as radius ( $k$ (L). The volumes can have any shape,
the significance of the especially elongated featurion mive mon (Fig. 2a). The luminosity is an approximate proxy for mass, that $M_{\text {lum }}=X_{13 C O} L_{\text {l3CO }}$, where $X_{13 C O}=8.0 \times 10^{20} \mathrm{~cm}^{2} \mathrm{~K}^{-1} \mathrm{~km}^{-1}$ (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be via calculation In principle extended portions of the tree ( Fig 2 yellowhighlighting) where $\alpha^{2}<2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of $p-p-v$ space where selfgravity is significant. As $\alpha_{\text {obs }}$ 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 ${ }^{16}$, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.


Figure $\mathbf{3}$ | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers 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 planar curve in wo dimensions, and an isosurface in thre dimensions. The dendrogram of 3 D data shown in Fig. 2 c is the direct than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

Goodman et al. 2009, Nature, cf: Fluke et al. 2009


Alyssa A. Goodman ${ }^{1,2}$, Erik W. Rosolowsky ${ }^{2,3}$, Michelle A. Borkin ${ }^{1} \dagger$, Jonathan B. Foster ${ }^{2}$, Michael Halle ${ }^{1,4}$, Jens Kauffmann ${ }^{1,2}$ \& Jaime E. Pineda ${ }^{2}$

Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size $\sim 0.1$ parsecs) inside molecular clouds collapse to form star-plus-disk systems ${ }^{1}$. But self-gravity's role at earlier times (and on larger length scales, such as $\sim 1$ parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function ${ }^{2}$. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by ${ }^{13} \mathrm{CO}$ observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission ${ }^{3}$ are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the locations of already-forming stars, or of those probably about to form, a celfagravitatino cocoon seemea critical condition for their exist.
overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line


DIMENSIONALITY (AND COLOR)


Borkin et al. 2011 cf. colorbrewer2.org


## LINKED VEWS OF HIGH-DIMENSONAL DATA



## TUKEYS "FOUR ESSENTIALS" (C1972)

## Picturing Rotation

## Isolation Masking

## Selection

## and these "need to work together" in a "dynamic display"

## Brushing

## Linking

## Results...

1. for immediate insight
2. as visual source of ideas for statistical algorihms

Warning
"details of control can make or break such a system"


## DATADESK (EST. 1986)




## LINKED VEWS OF HIGH-DIMENSIONAL DATA



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

GREAT. BUT THAT WAS ALL FROM ONE DATA FIE. AND IT WAS $\mathbb{N}$ SOFTWARE THAT COSTS \$1000.

## LINKED VEWS OF HGH-DMENSONAL DATA (N PYTHON) GLUE



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


# Once upon a time (2012), in an chchaniced casile (in Bavaria) 

## ...at a conierence albout

 "The Farily Phases of Star Foration"Andi Burkert asked'a question:
Is Nessie "paralle Ito the Gulactic Plane"?

No one knew.

The Milky Way
(Artist's Conception)

## "Is Nessie Parollel to the Goluctic Plane?"

Celestial
North

Yes but why not at Zero of Latitude $(\mathrm{b}=0)$ ?

## Where are we, really?

## "IAU Milky Way", est. 1959



## True Milky Way, modern

The equatorial plane of the new co-ordinate system must of necessity pass through the sun. It is a fortunate circumstance that, within the observational uncertainty, both the sun and Sagittarius A lie in the mean plane of the Galaxy as determined from the hydrogen observations. If the sun had not been so placed, points in the mean plane would not lie on the galactic equator.
[Blaauw et al. I 959]

Sun is
~ 75 light years
"above" the
IAU Milky Way Plane

Galactic

## Center is

~20 light years offset from the
IAU Milky Way
Center

The Galactic Plane is not quite where you'd think it is when you look at the sky

## In the plane! And at distance of spiral arm!




## A full 3D skeleton?



## New!

## 2014 Simulation



## New!

## 2014 Simulation



Smith et al. 2014, using AREPO

Data Collection

## Subsets

- Nessie on the Sky
- BigVrangeEast

BigVrangeEast (HOPS_NH3-11-D...
BigVrangeEast (peretto)
BigVrangeEast (glimpse_nessie_4)
BigVrangeEast (DHT36_Quad4_i...

- CentralNessie
$\square \therefore \dot{\square} \quad$ Q Link Data $I P$

Plot Layers - Scatter Plot
$\checkmark$ NessieWest (HOPS_NH3-11-DuchampCat)

- CentralNessie (HOPS_NH3-11-DuchampCa
$\checkmark$ BigVrangeEast (HOPS_NH3-11-DuchampC:
v Nessie on the Sky (HOPS_NH3-11-Ducham
ป HOPS_NH3-11-DuchampCat

Plot Options - Scatter Plot

| x axis | ICent_deg | * | $\square \log$ flip |
| :---: | :---: | :---: | :---: |
| y axis | vCent_kms | $\uparrow$ | $\square \log \square$ flip |
| Auto scale |  |  | wap Axes |
| show hidden attributes |  |  |  |
| Plot Limits |  |  |  |
| x min | 332.424 | x max | 345.468 |
| $y$ min | -69.056 | $y$ max | -6.445 |



## NTERACTON BEYOND FLATLAND IS AN UNSOLVED PROBLEM



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


\author{
RESOLUTION
CONTEXT
BIG DATA
WIDE DATA <br> DMENSIONALTTY <br> LINKED VEWS <br> 



## The "Paper" of the Future

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

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

## 1 Preamble

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

Much more than text is used to 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.


Konrad Hinsen 3 days ago - Puolic
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 possble. That's perhaps a good strategy to change habits, but in the long run it could well backire when people find out in a few years that $90 \%$ of those doposits have become useless.

Christine L. Borgman 4 months ago • Private

## COMMUNCATION: LITERATURE AS A FLTER FOR DATA

## nDS ALL SKY SURVEY

View in Aladin • View in WorldWide Telescope • Demo Videos

## [Demo]



## EDUCATION. 2015+

| Stephen |  |
| :---: | :---: |
| Yuan-Sen Ting |  |
| Interstellar <br> Absorption and the Lyman Alpha |  |
| Forest |  |
| ${ }_{\text {Jaxascript JavaScript }}^{\text {S. }}$ | https://www.cfa.harvard.edu/-yuan-sen.ting/yman_alpha.html |
| Janseripe JovaScript | htep://portillo.ca/nebula/ |

## online learning



## WWT Ambassadors



## THE FUTURE IS ABOUT INTEGRATION



Tab 1 C. Simple WWT Web Client CO - PRIMARY


Rectangular ROI


[^1]Video courtesy of Chris Beaumont, Lead Glue Architect


## wind map

## hint.fm/wind (Fernanda Viegas \& Martin Wattenberg)



Darker regions correspond to stronger polarised emission, and the striations indicate the direction of the magnetic field projected on the plane of the sky.


Darker regions correspond to stronger polarised emission, and the striations indicate the direction of the magnetic field projected on the plane of the (ESA, May 2014.)


Darker regions correspond to stronger polarised emission, and the striations indicate the direction of the magnetic field projected on the plane of the (ESA, May 2014.)


Alyssa A. Goodman • Harvard-Smithsonian Center for Astrophysics • @aagie


## Thanks for the Warmth!


takeasweater.com, and "TakeASweater" in the Apple App Store

projects.iq.harvard.edu/seamlessasstronomy/presentations

## Nessie to B5, the movie.

EXTRA SLIDES

## A Rotating (Spiral) Galaxy Observed from its Outskirts...






## Microsoft ${ }^{\circledR}$ Research

WorldWide Telescope


View and compare images from across the electromagnetc spectrum

Much more than "just" the sky at night!
3D features can take you to other planets, stars \& galaxies.

## Finder Scope



NGC224

RA: 00h42m42s Magnitude: Dec: 41 : 16 : 00 Distance: Alt: 70 : 06 : 26 Rise: Az: 275: 42:17 Transit:

Classification:
Spiral Galaxy in Andromeda

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


https://www.cfa.harvard.edu/~agoodman/seamless/

## Supported by "Research <br> $\square$ 0



Stellar Jet in the Carina Nebula Hubble Space Telescope • WFC3/UVIS/IR


Expertise


Expertise


등 ㄷ


Expertise


등 ㄷ


Expertise
Expertise




synchrotron polarization


screenshot of WEAVE from Gresh et al. 2000,
reproduced as shown in Goodman 2012


[^0]:    Temperature Foreground amplitudes from Commander, Planck Data [Feb 2015]

[^1]:    Rectangular ROI $x=27.8073 \quad y=107.248$

