VISUALIZATION IN

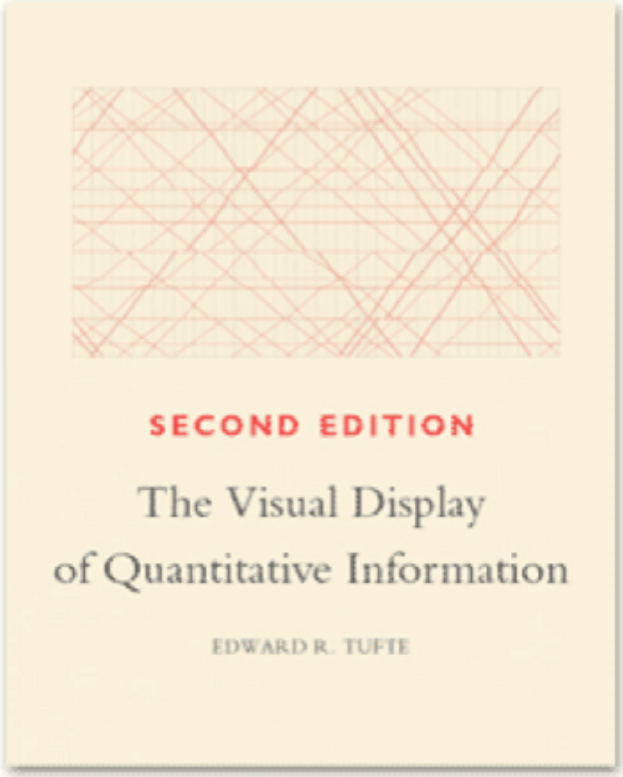
Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics
Radcliffe Insitute for Advanced Study
@AlyssaAGoodman

VISUALIZATION IN 20 MIN

THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION

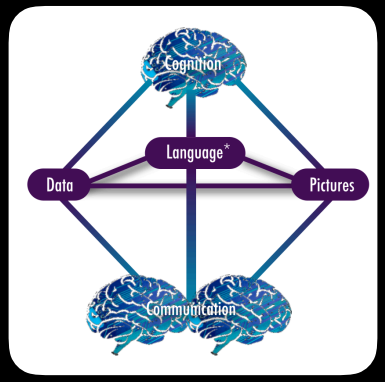
The classic book on statistical graphics, charts, tables. Theory and practice in the design of data graphics, 250 illustrations of the best (and a few of the worst) statistical graphics, with detailed analysis of how to display data for precise, effective, quick analysis. Design of the high-resolution displays, small multiples. Editing and improving graphics. The data-ink ratio. Time-series, relational graphics, data maps, multivariate designs. Detection of graphical deception: design variation vs. data variation. Sources of deception. Aesthetics and data graphical displays.

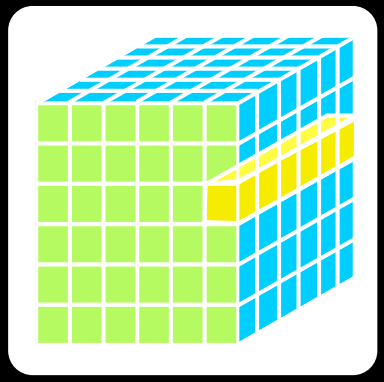
This is the second edition of *The Visual Display of Quantitative Information.* Recently published, this new edition provides excellent color reproductions of the many graphics of William Playfair, adds color to other images, and includes all the changes and corrections accumulated during 17 printings of the first edition.

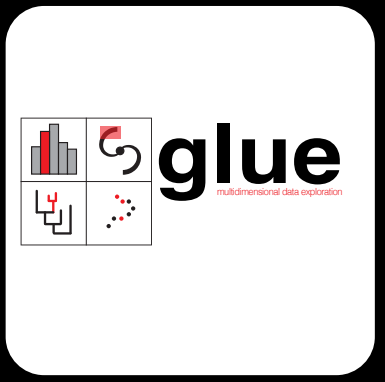


(READ CANON LAW)

VISUALIZATION IN 20 MIN







(LIVING BY THE CANON IN 2019)

your handout

≡ MENU

TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION

The 10 Questions

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. **Categories** |Do you want to show or explore pre-existing, known, human-interpretable, categories?
4. **Patterns** |Do you want to identify new, previously unknown or undefined patterns?
5. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
6. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
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?

Now, visit the 10QViz conversation! There's so much more to talk about.

Arzu Çöltekin
10QViz co-founder
(with Alyssa Goodman)

Curious about the **origins** of 10QViz? Try the About page.
Want to learn **how best to use** and **participate** in 10QViz? Try the How to page.
Want to read about the **scholarship** behind 10QViz.org's questions? Try Coltekin & Goodman 2018.

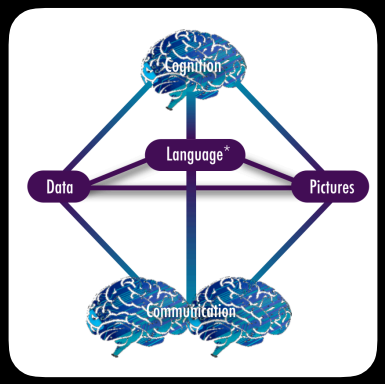
(VISUALIZATION IN >>20 MIN)

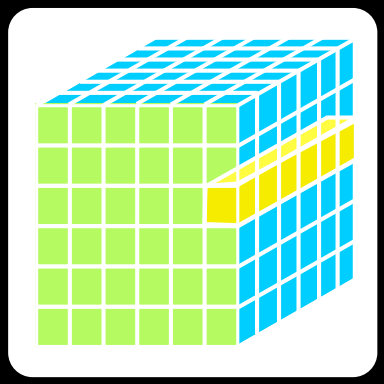
The 10 Questions

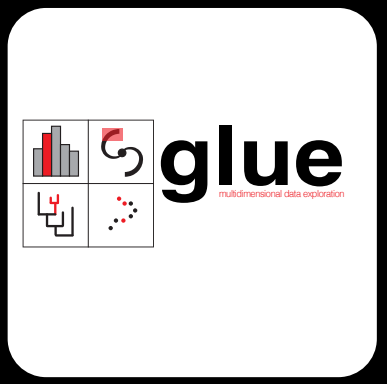
1. **Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
2. **Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
3. **Categories** | Do you want to show or explore pre-existing, known, human-interpretable, categories?
4. **Patterns** | Do you want to identify new, previously unknown or undefined patterns?
5. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
6. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
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?

10QVIZ.ORG

VISUALIZATION IN 20 MIN







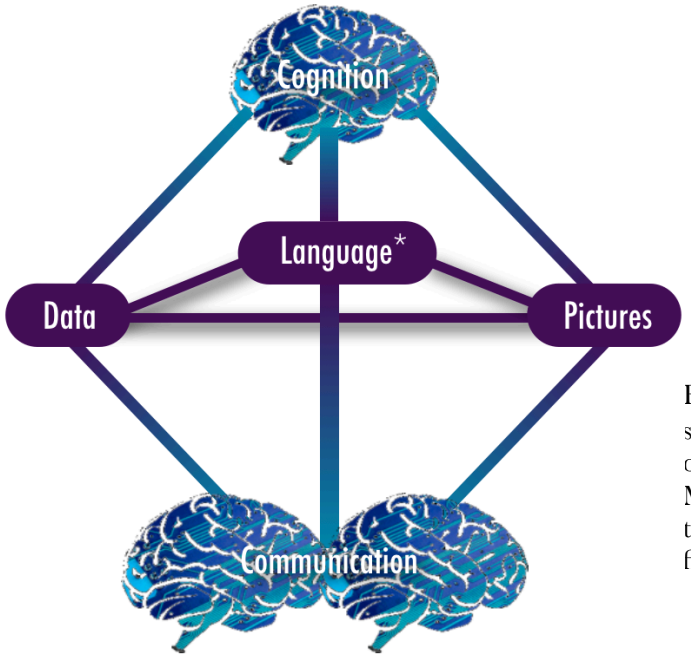
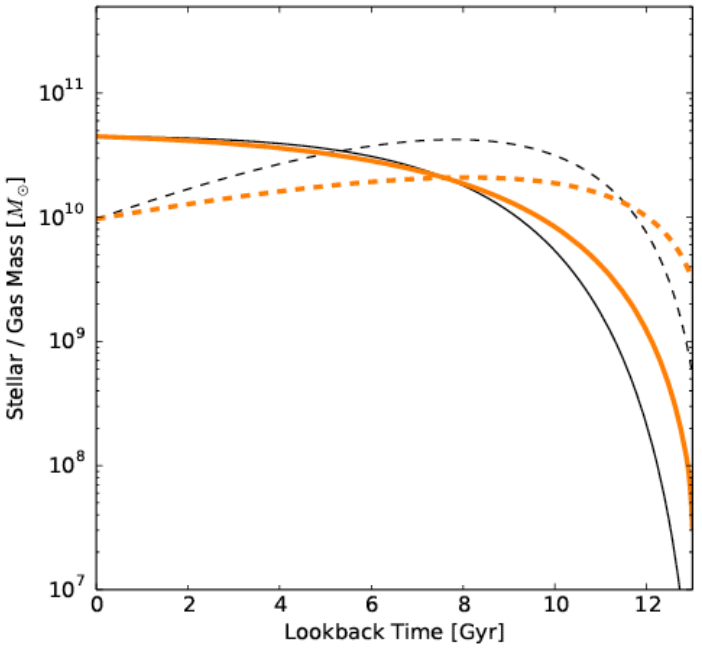
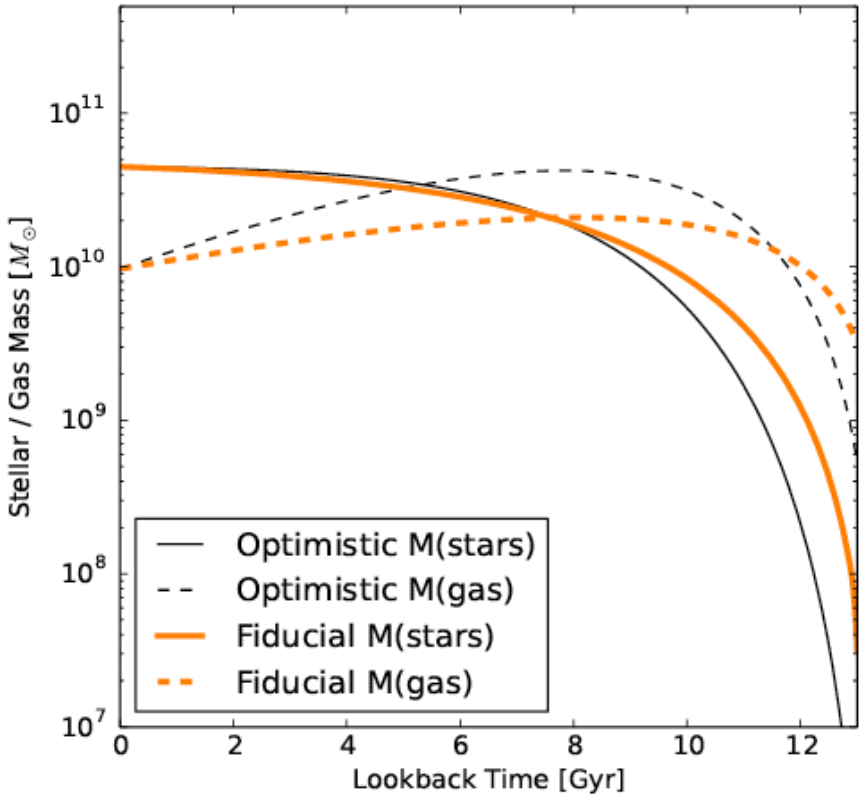


Figure Caption: The solid/ solid black line shows the optimistic case for M(stars)/ M(gas). The orange lines show the same quantities, for the fiducial case.





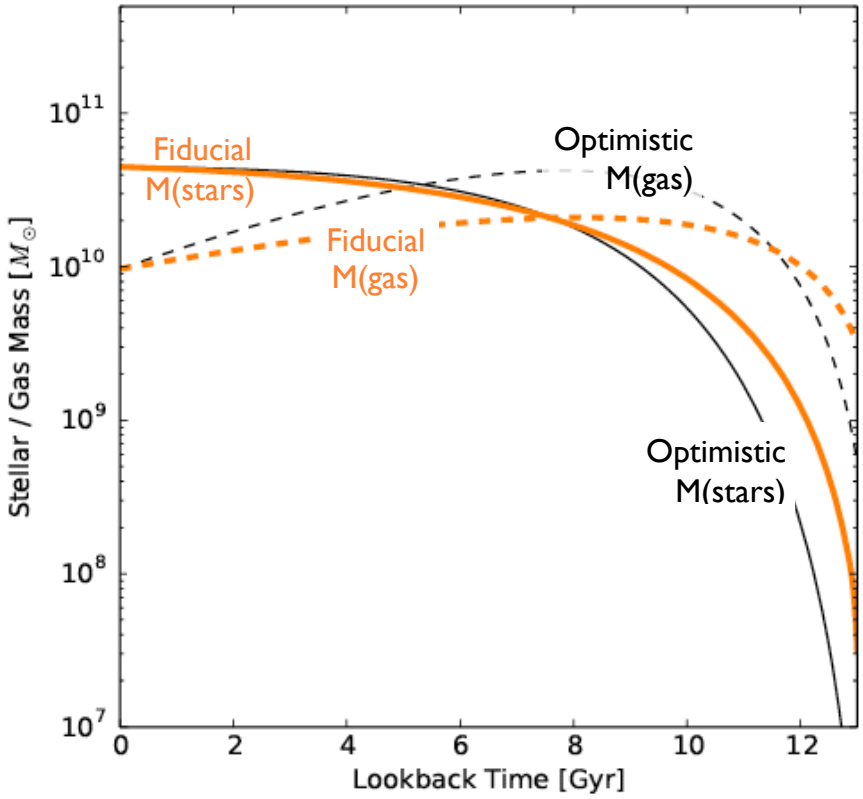
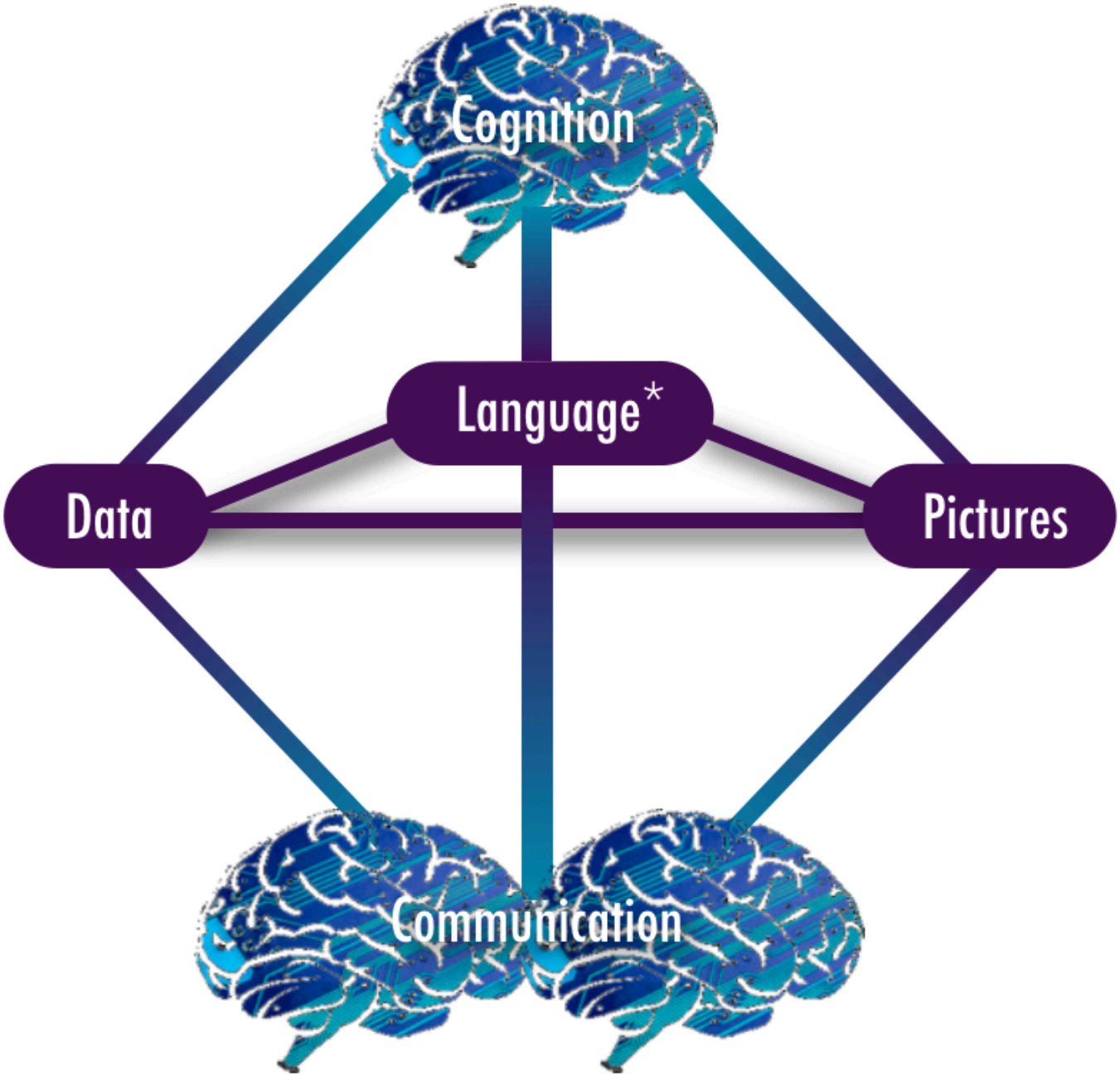


figure from draft of Ting, Conroy & Goodman 2015


Cognition
Language*
Data
Pictures
Communication

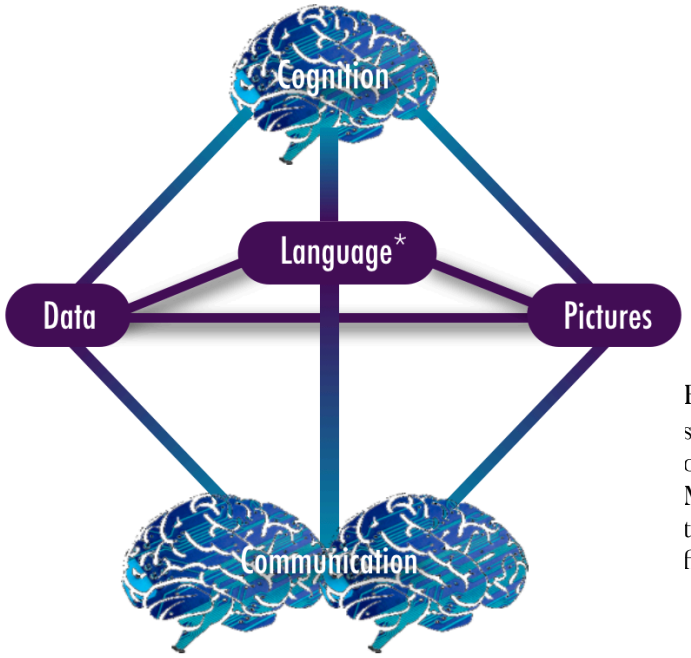
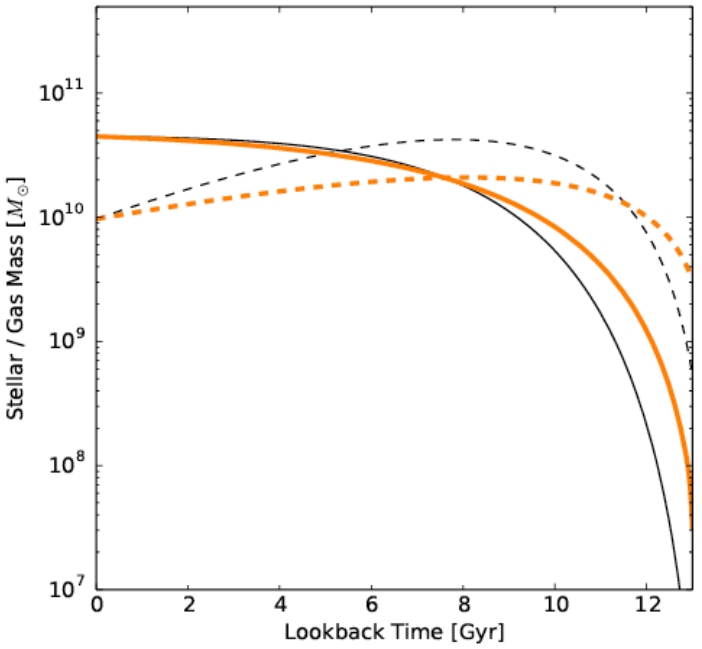
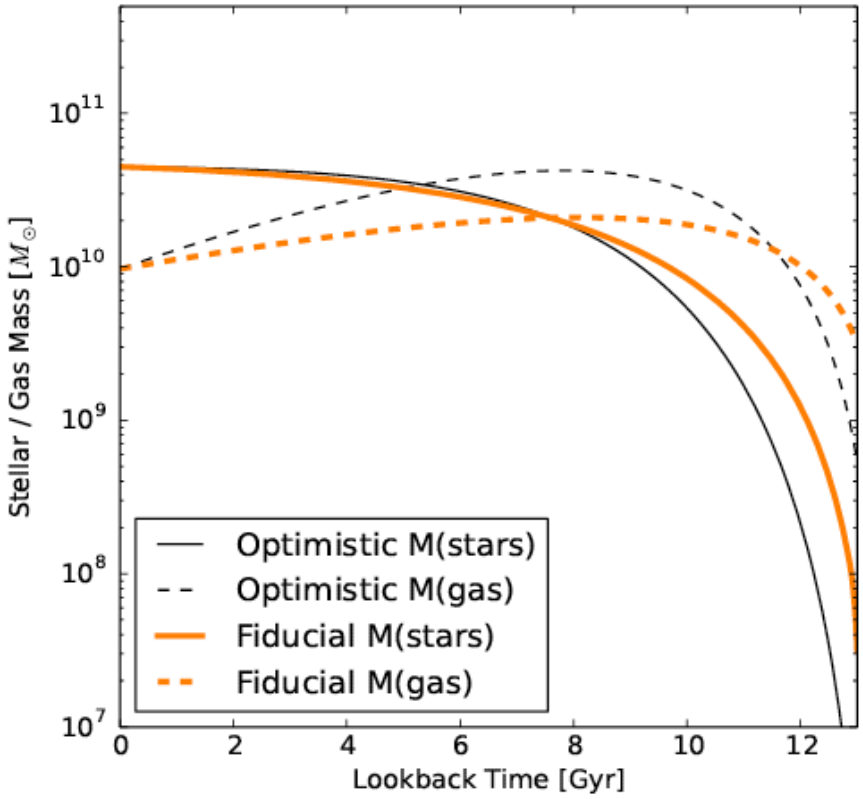


Figure Caption: The solid/ solid black line shows the optimistic case for M(stars)/ M(gas). The orange lines show the same quantities, for the fiducial case.





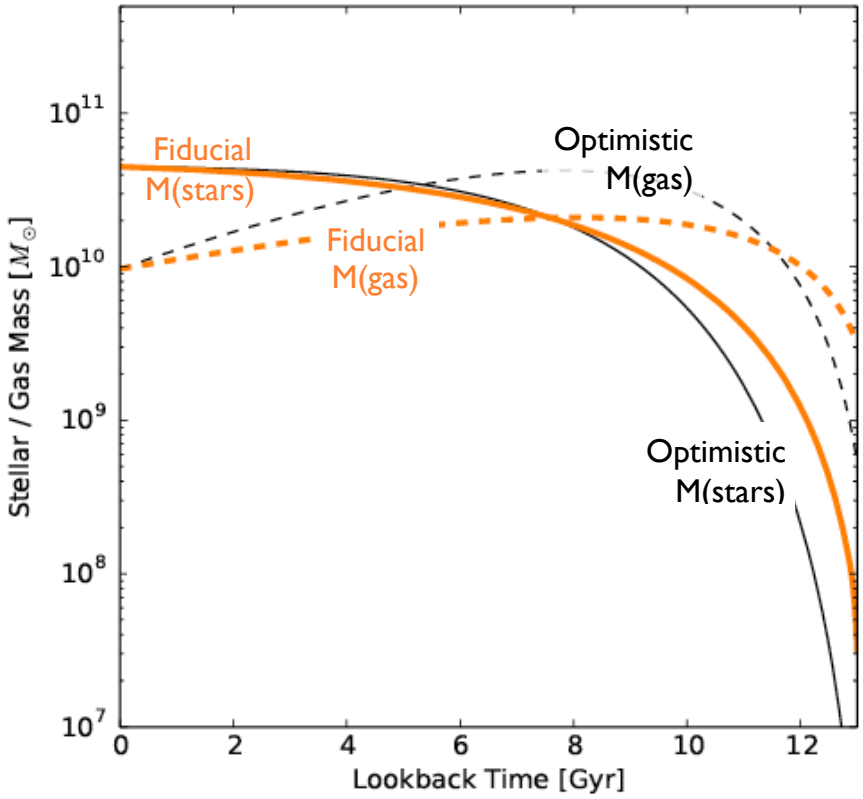
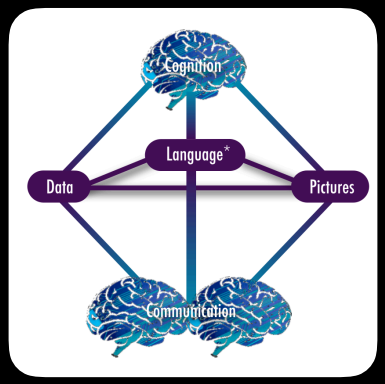
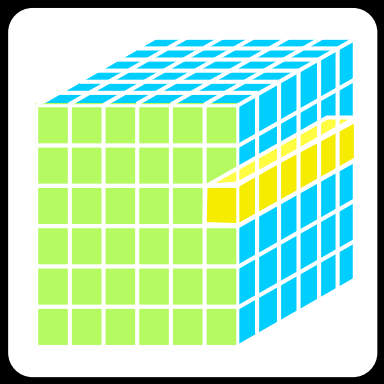
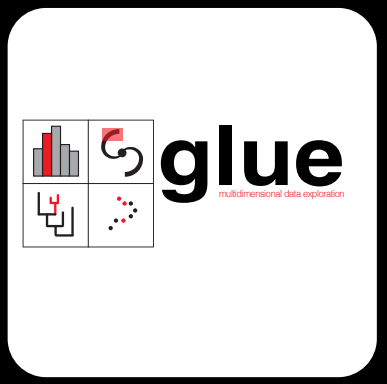


figure from draft of Ting, Conroy & Goodman 2015

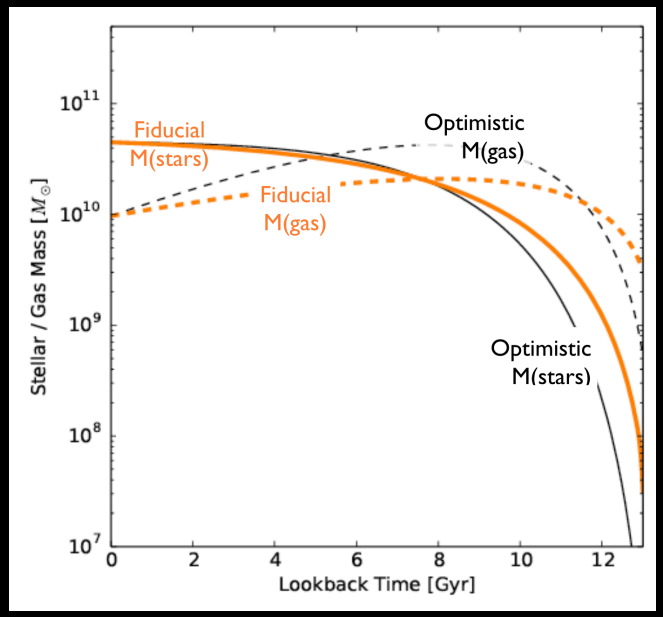
VISUALIZATION IN 20 MIN

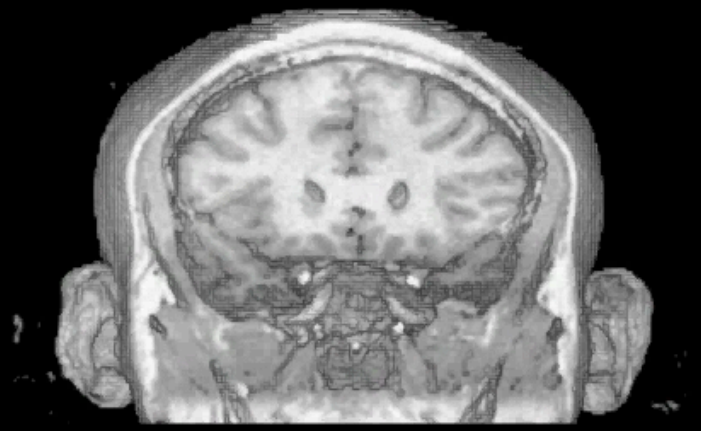


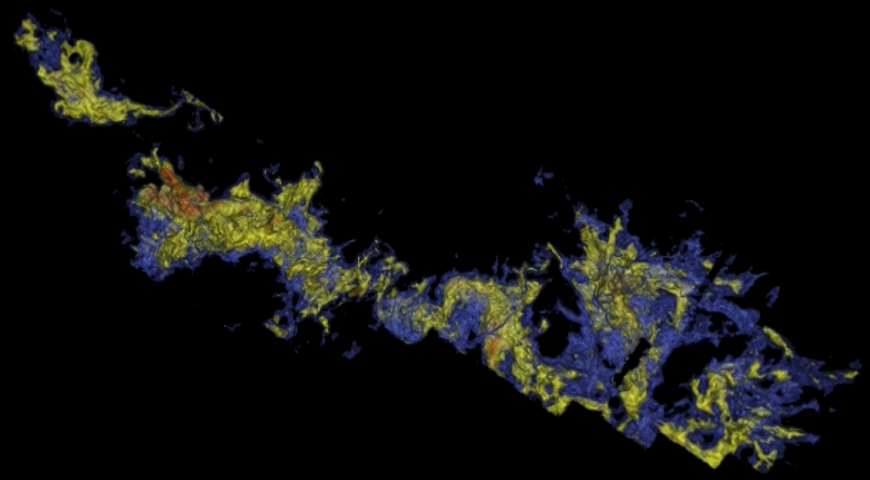


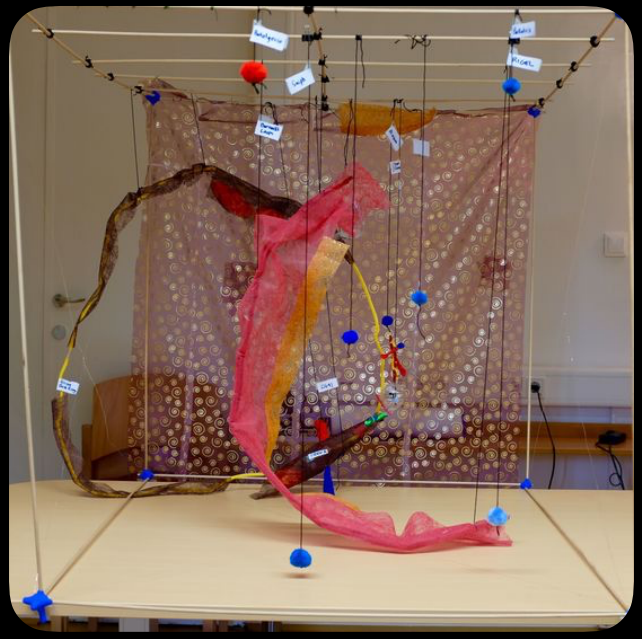


“DATA, DIMENSIONS, DISPLAY”

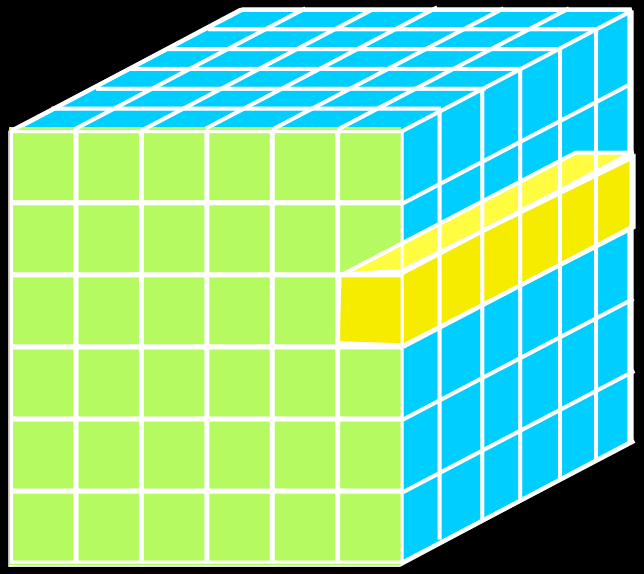






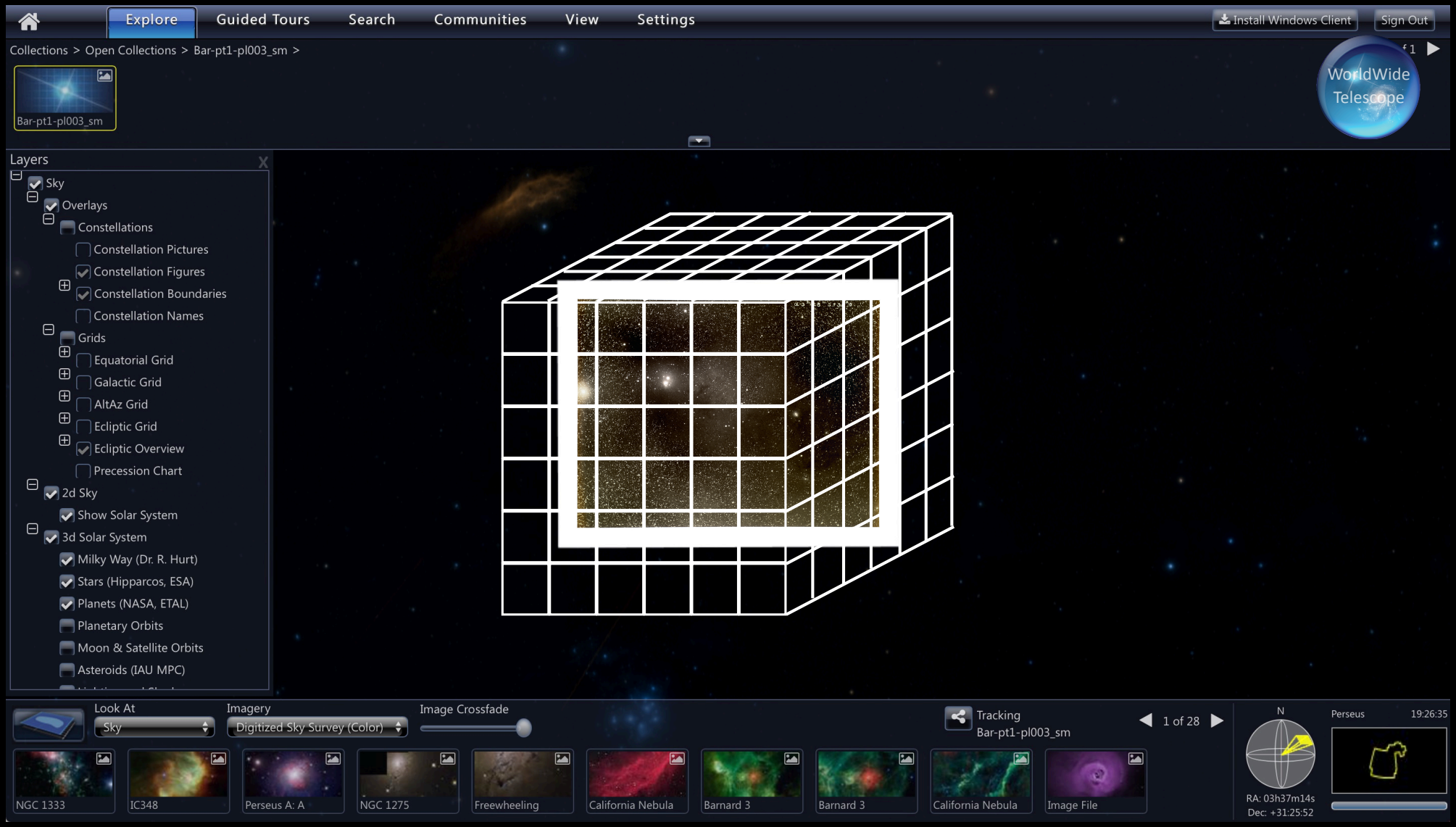


"DATA, DIMENSIONS, DISPLAY"

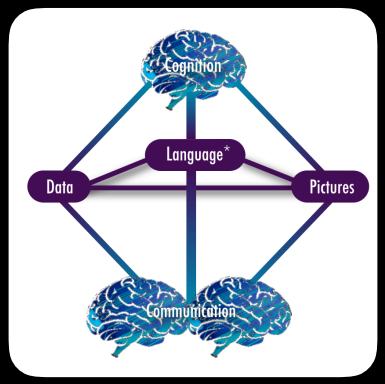


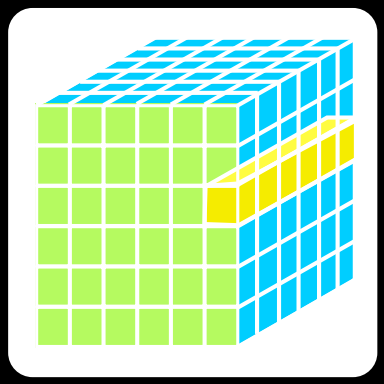
1D: Columns = "Graphs"
2D: Faces or Slices = "Images"
3D: Volumes = "3D Renderings", "2D Movies"
4D: Time Series of Volumes = "3D Movies"

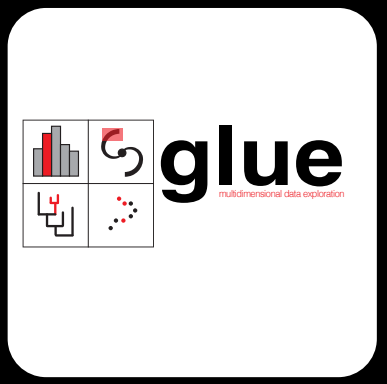
“DATA, DIMENSIONS, DISPLAY”



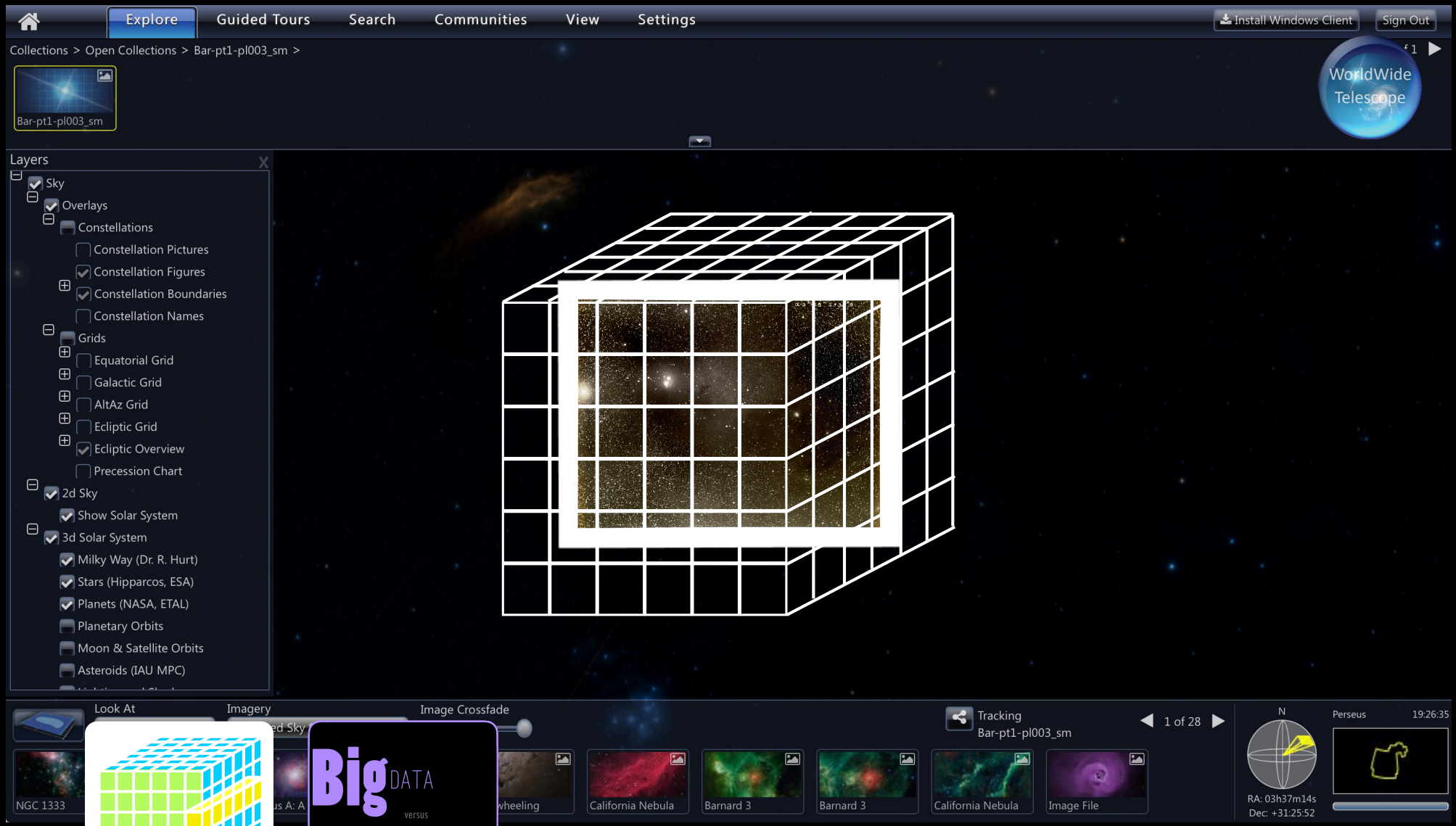
VISUALIZATION IN 20 MIN



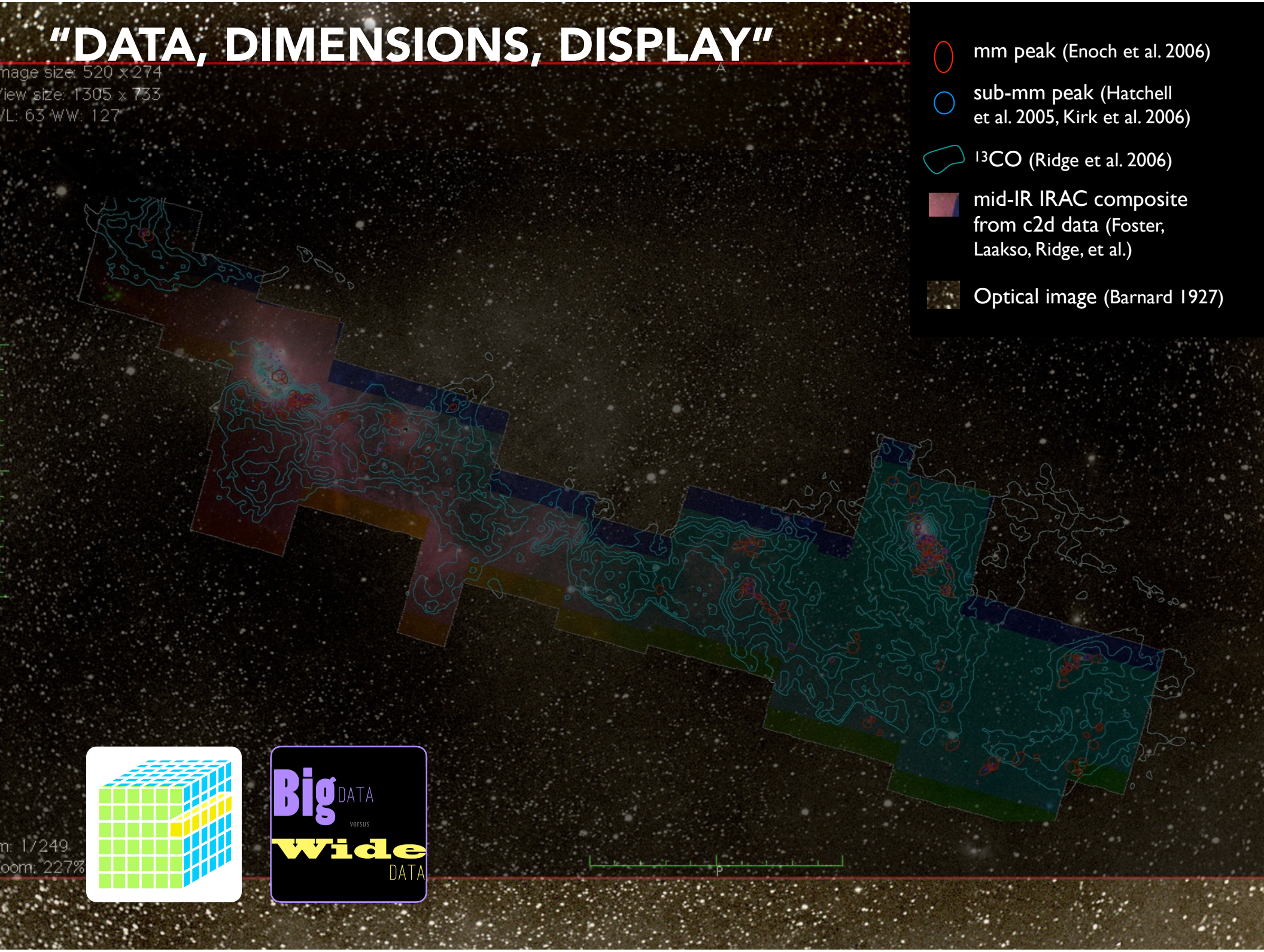




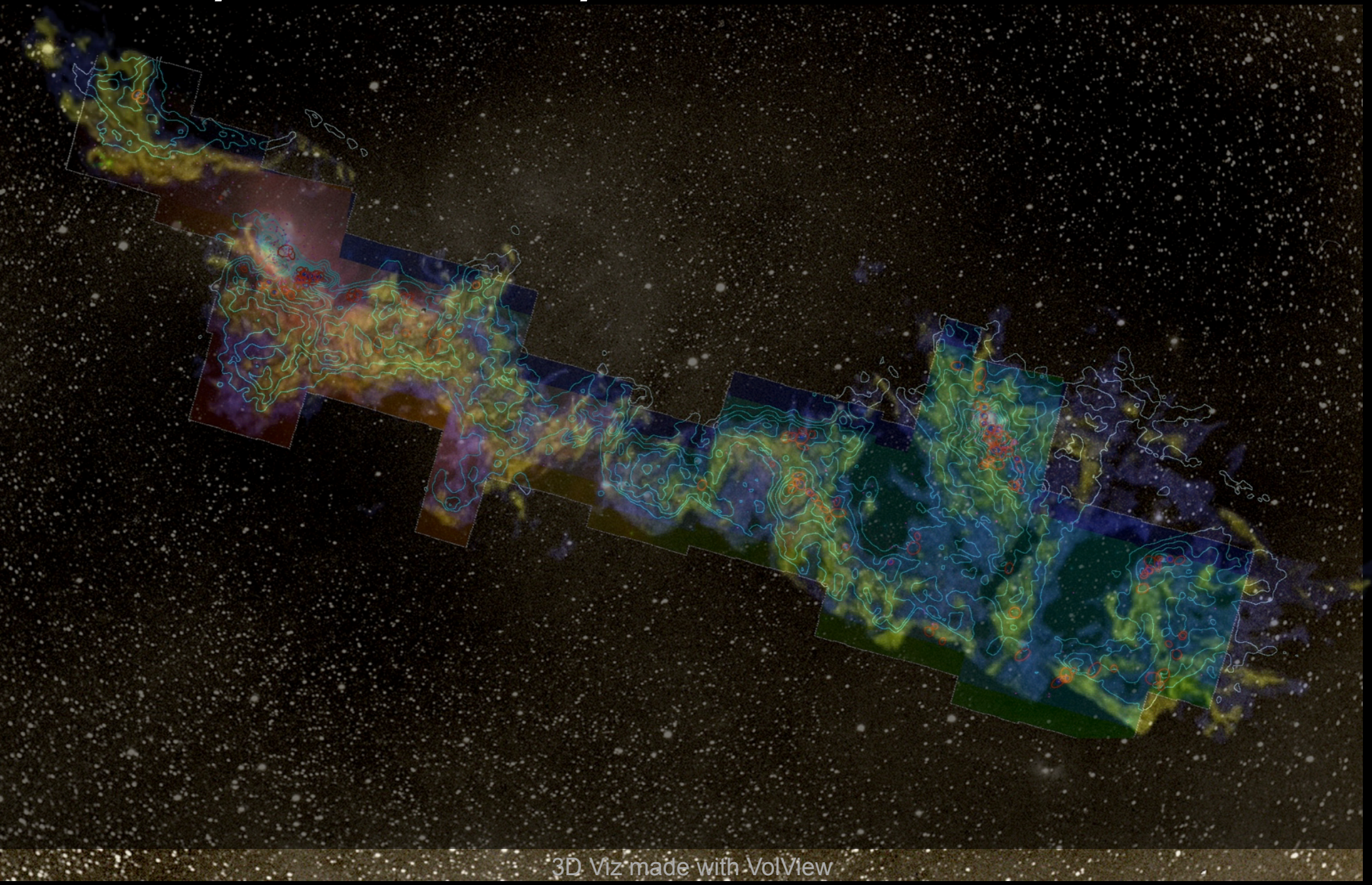
“DATA, DIMENSIONS, DISPLAY”



"DATA, DIMENSIONS, DISPLAY"



“DATA, DIMENSIONS, DISPLAY”



3D Viz made with VolView

CLOSER TO HOME (GLUE & MY FBI FILE)

Glue

Terminal | Open Session | Save Session | Add/edit data components | Selection Mode: | Preferences | View Error Console

Tab 1

3D Scatter

2D Scatter

2D Image

2D Scatter

Climbing
Landing
A Day in the Life of Logan

Plot Layers - 2D Scatter

- A Day in the Life of Logan (airplane_positions)
- Landing (airplane_positions)
- Climbing (airplane_positions)
- Descending (airplane_positions)
- nearglound (airplane_positions)

Color | Points | Line | Errors | Vectors

color: Fixed

opacity

Plot Options - 2D Scatter

General | Limits | Axes

x label: Heading (degrees)

y label: Ground Speed

	x	y
axis label size	10	10
axis label weight	medium	medium
tick label size	10	10

Apply to all plots

altitude

Ground Speed

Heading (degrees)

Latitude

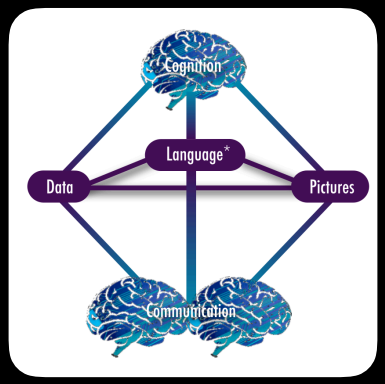
Longitude

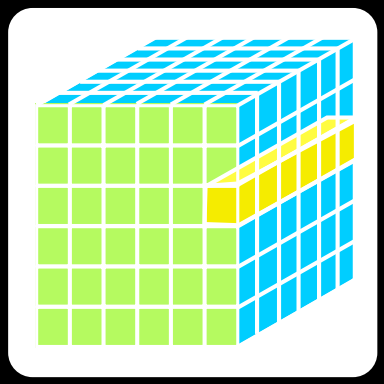
Vertical Rate

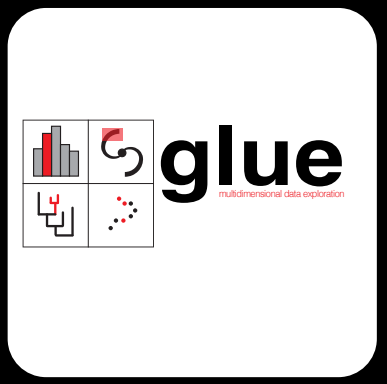
Time (seconds)

1e9

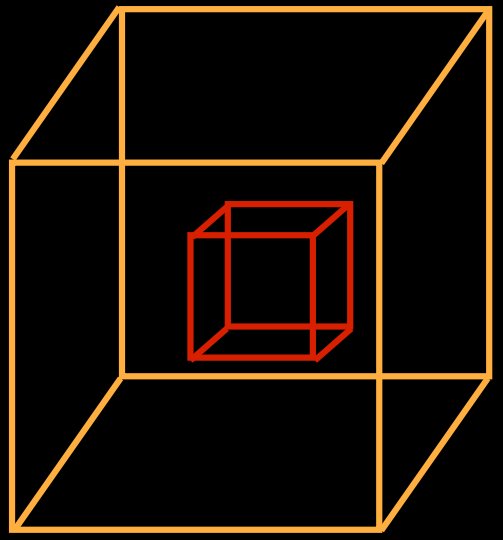
VISUALIZATION IN 20 MIN



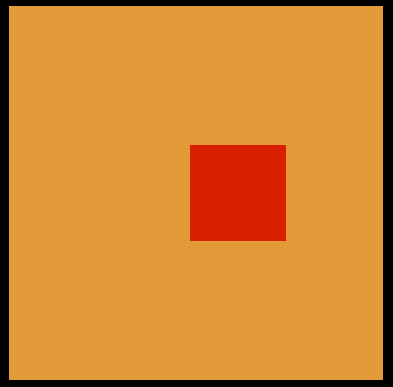




LINKED VIEWS OF HIGH-DIMENSIONAL DATA

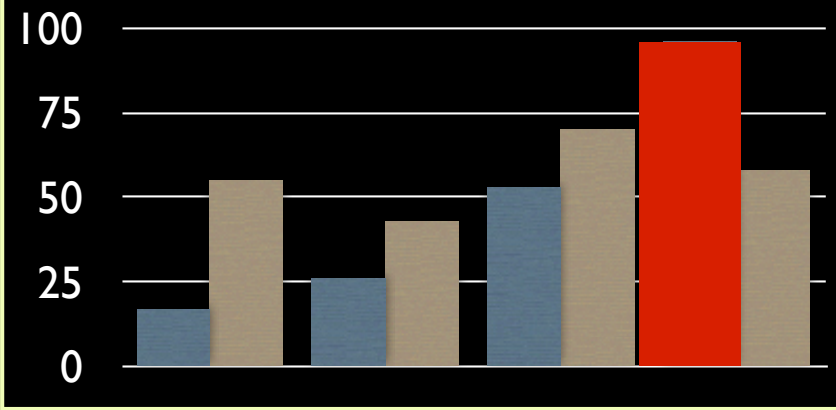


3D

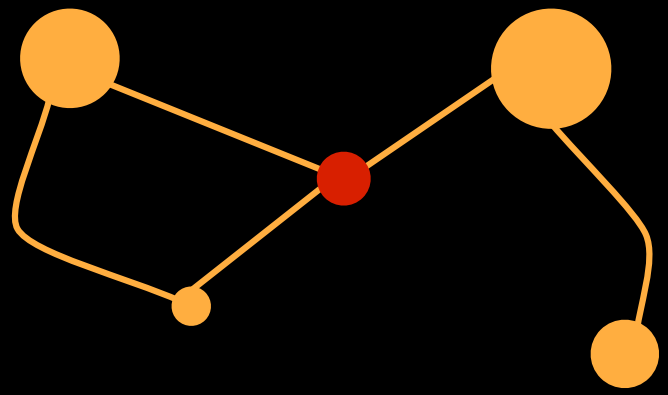


2D

Statistics



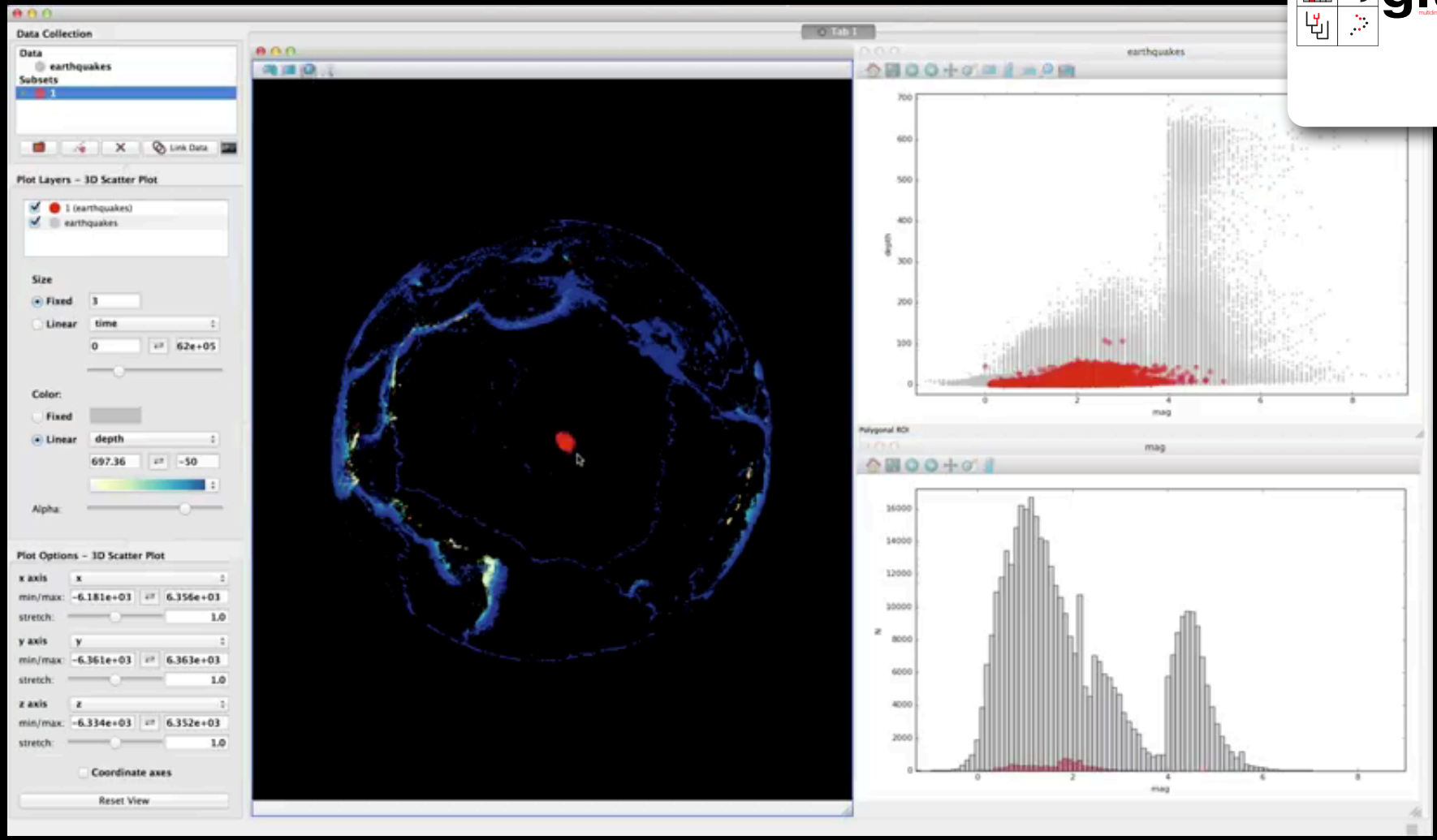
Data Abstraction



glue

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

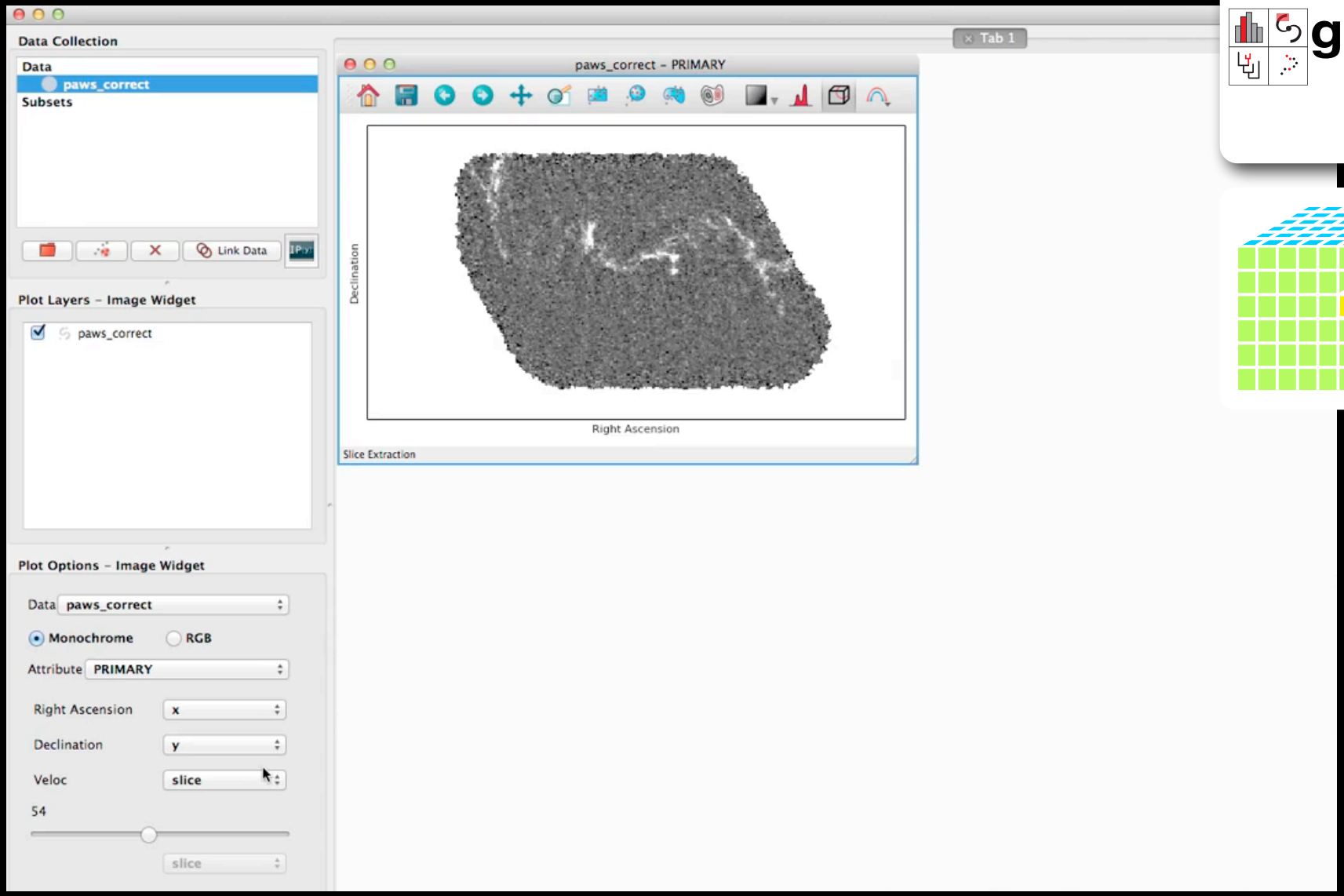
LINKED VIEWS OF HIGH-DIMENSIONAL DATA, IN PYTHON



video by Tom Robitaille, lead glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

LINKED VIEWS OF HIGH-DIMENSIONAL DATA, IN PYTHON



video by Chris Beaumont, glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

your handout

glue

multidimensional data exploration

built-in

standard data loaders

Data File 1

Data File 2

Data File N

define new variables, import/export insights, interactive plots for the web, save state, all from GUI

standard 1D, 2D & 3D plots

link data files' attributes

highlight live or algorithmic selections with Boolean logic

plug-in

custom data loaders

custom buttons, features

custom plots

?

+options

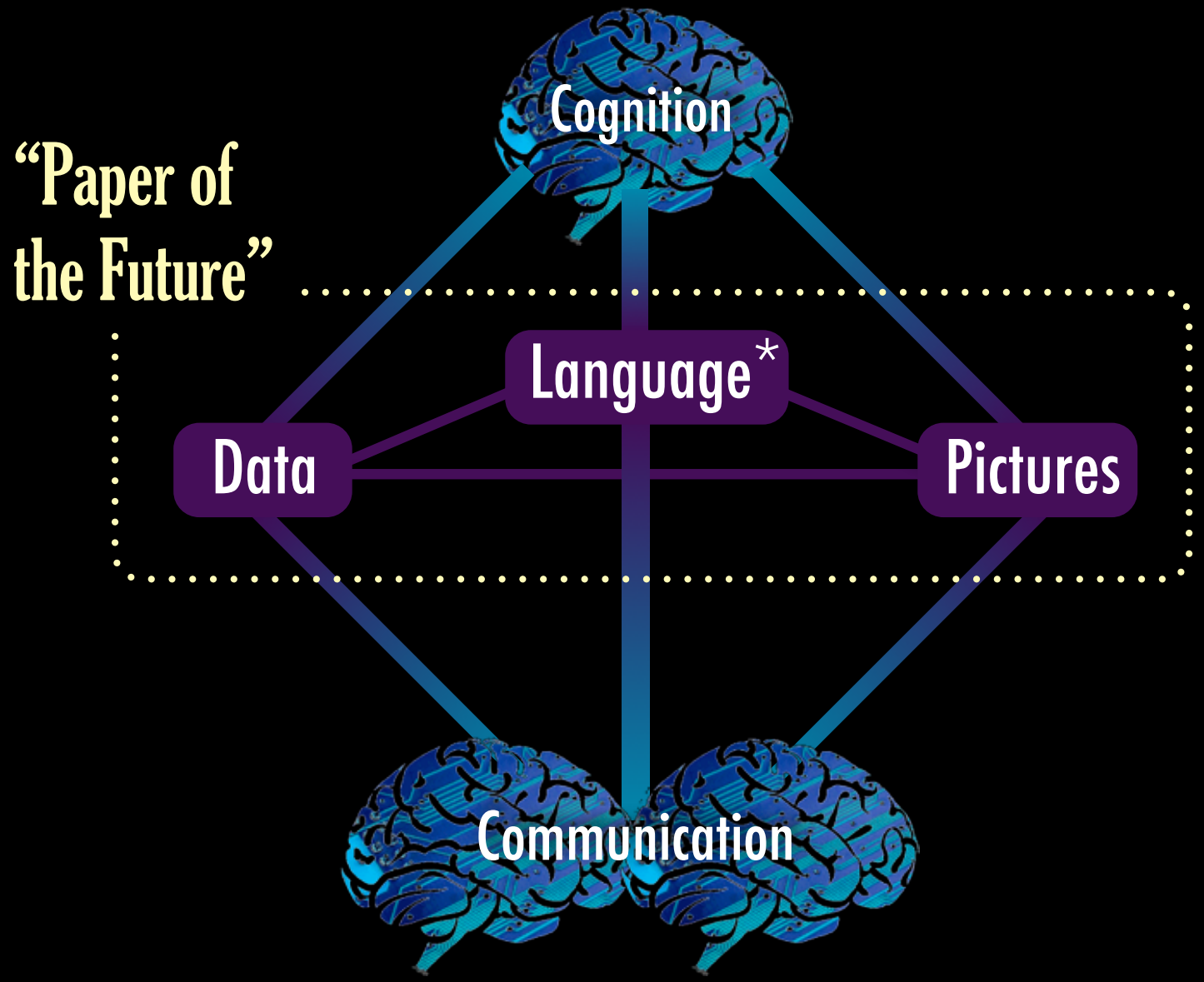
user config.py file (loaders, colors, plot types, +)

access to all matplotlib functions through built-in IPython terminal

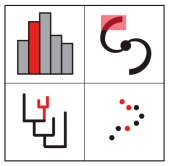
run & interact with glue from Jupyter notebook & other tools

glueviz.org

"Paper of the Future"



*"Language" includes words & math



glue

multidimensional data exploration

enabled by d3.js (javascript) outputs

plotly

d3po

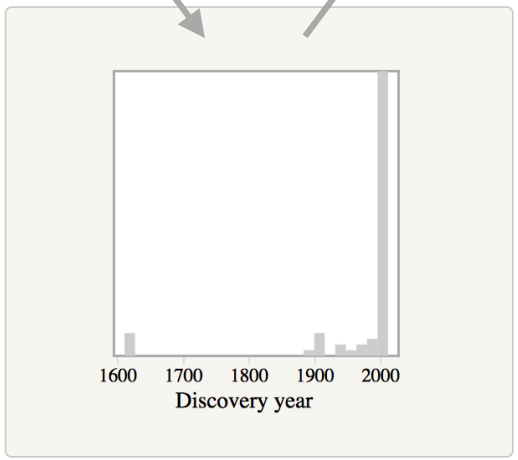
d3po is a project designed to allow an astronomer (or anyone), with no special data visualization skills, to make an interactive, publication-quality figure that has staged builds and linked brushing through scatter plots. Our current version can be previewed at d3po.org, and represents a figure from upcoming work by graduate student Elisabeth Newton. The figure describes how metallicity affects color in cool stars, and represents a nice use case for d3po. Try clicking and dragging in the scatter plots to understand the power of linked brushing in published figures.

Right now we are in search of alpha testers, who have figures that could be made interactive and who are willing to get their hands a little dirty (No javascript skills needed). In future versions, we plan to link to glue to allow the creation of d3po figures interactively. We are also exploring implementation of d3po within presentations and within authorea. Full 1.0 version expected in January 2014.

Installing your own d3po server

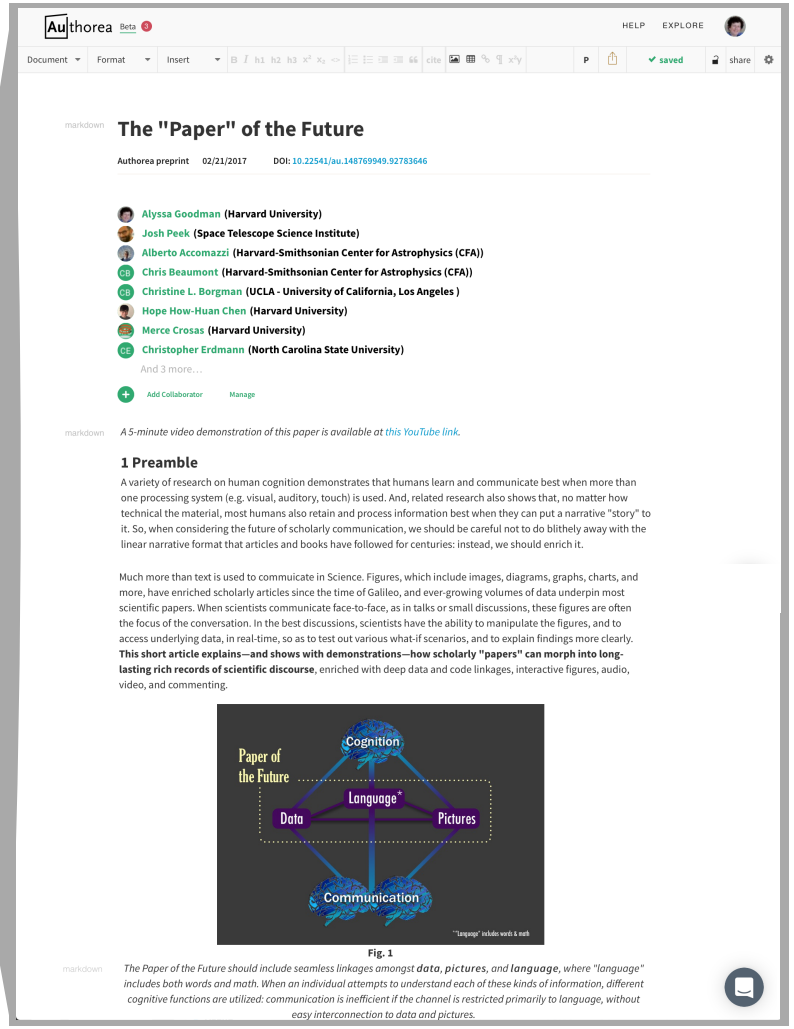
```
git clone git@github.com:adrn/d3po.git
cd d3po
virtualenv --no-site-packages venv
source venv/bin/activate
pip install -r pip-requirements.txt
python run.py
```

Authorea



Four Centuries of Discovery | A Chasm in Mass | Little Siblings | Close Cousins | The Strangers

After Galileo discovered the first four moons of Jupiter, it took nearly three hundred years to discover the next one.

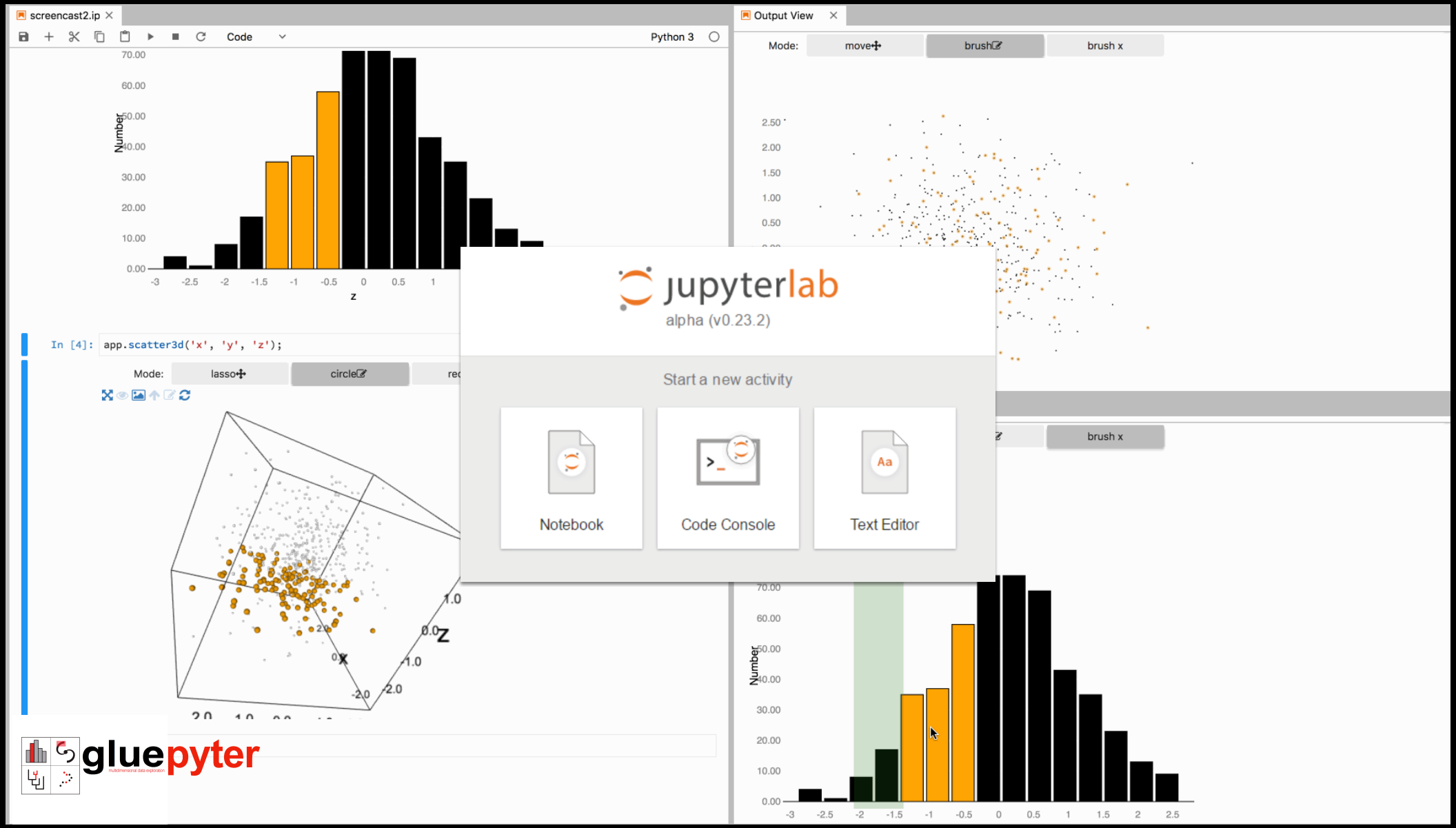


[demo]

Many thanks to Alberto Pepe, Josh Peek, Chris Beaumont, Tom Robitaille, Adrian Price-Whelan, Elizabeth Newton, Michelle Borkin & Matteo Cantiello for making this posible.

SNAP CIRCUITS JR.
Have fun learning all about electronics
Custom Parts
Organizer Box
Included!
U1
U2
U3

COMING NEXT: GLUE IN THE BROWSER

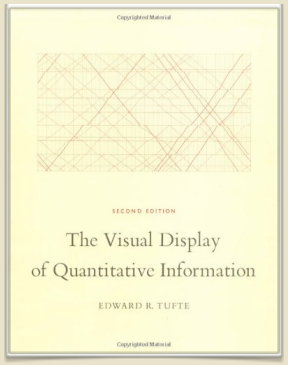


Video courtesy of Maarten Breddels, consulting developer

glueviz.org

HOW Best to Visualize, with examples

READ “The Visual Display of Quantitative Information,” Tufte 1983



Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

Data-ink Ratio

Multi-functioning
graphical elements

Data density

Small multiples

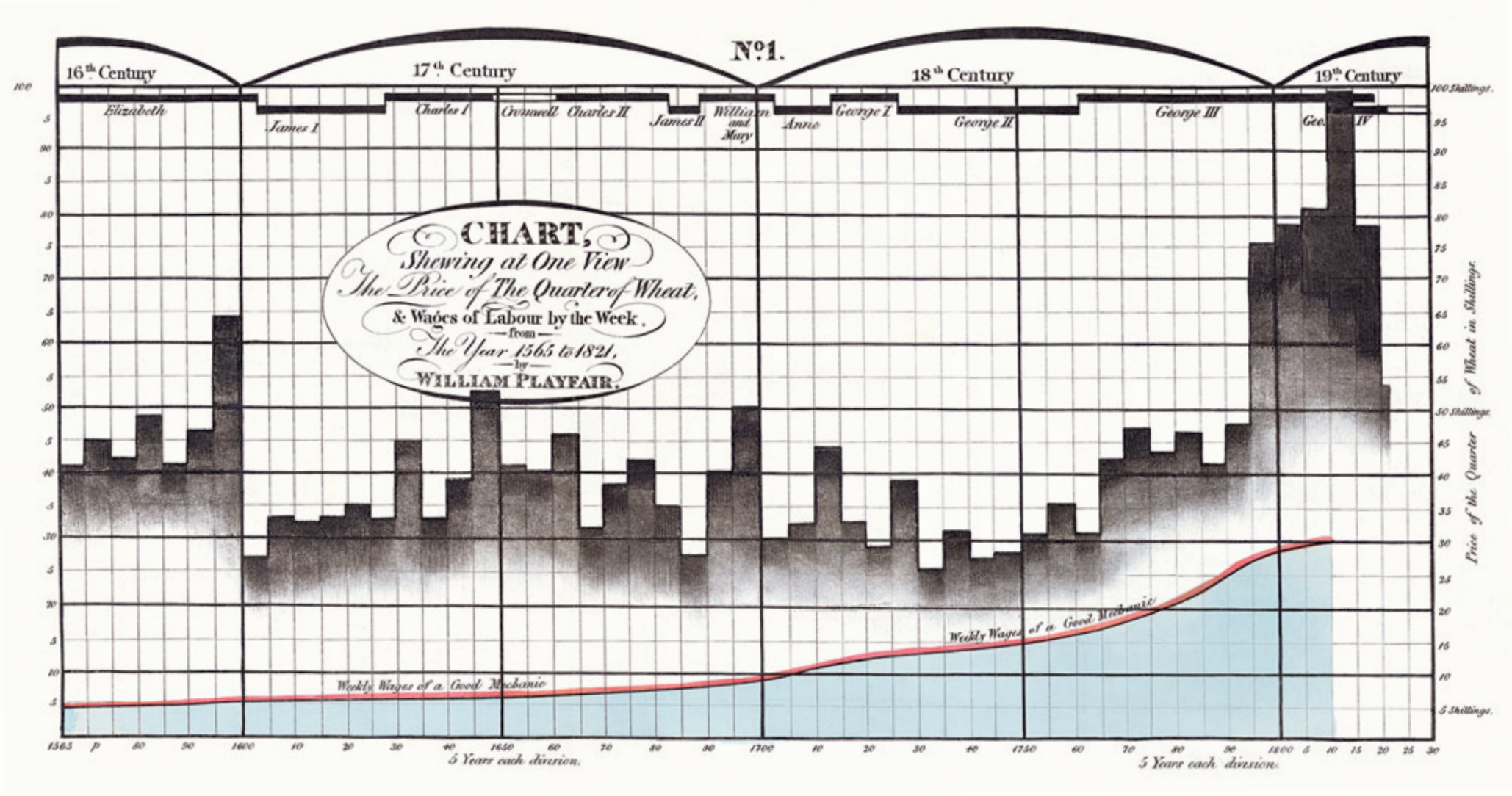
EDWARD TUFTE'S IDEAS

(from "The Visual Display of Quantitative Information," Tufte 1983.)

Graphical Excellence – displays should...

- **show** the **data**
- induce the viewer to **think** about the **substance** rather than about methodology, graphic design, the technology of graphic production, or something else
- **avoid distorting** what the data have to say
- present **many numbers** in a small space
- make large data sets **coherent**
- encourage the eye to **compare** different pieces of data
- reveal the data at **several levels of detail**, from a broad overview to the fine structure
- serve a reasonably clear **purpose**: description, exploration, tabulation, or decoration
- be closely **integrated** with the statistical and **verbal descriptions** of a data set

Graphical Excellence



William Playfair (1759-1823)

EDWARD TUFTE'S IDEAS

(from "The Visual Display of Quantitative Information," Tufte 1983.)

Graphical Excellence

Graphical Integrity
& "The Lie Factor"

Chartjunk
& Subtraction

Data-ink Ratio

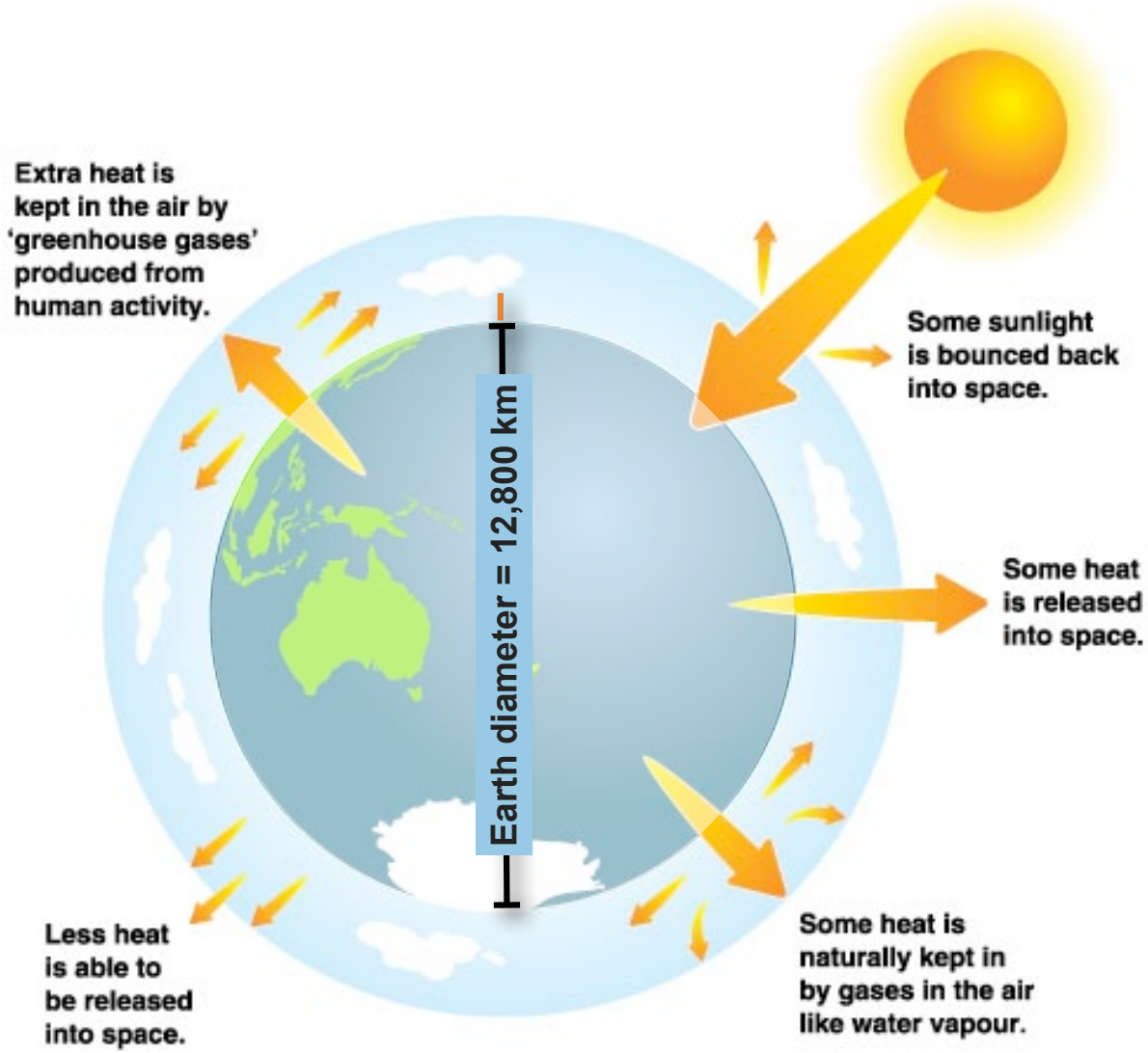
Multi-functioning
graphical elements

Data density

Small multiples

Graphical Integrity & "The Lie Factor"

http://www.aeronomie.be/en/topics/earthsystem/atmosphere-layers.htm



http://www.climatechange.govt.nz/science/what-is-climate-change.html

EDWARD TUFTE'S IDEAS

(from “The Visual Display of Quantitative Information,” Tufte 1983.)

Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

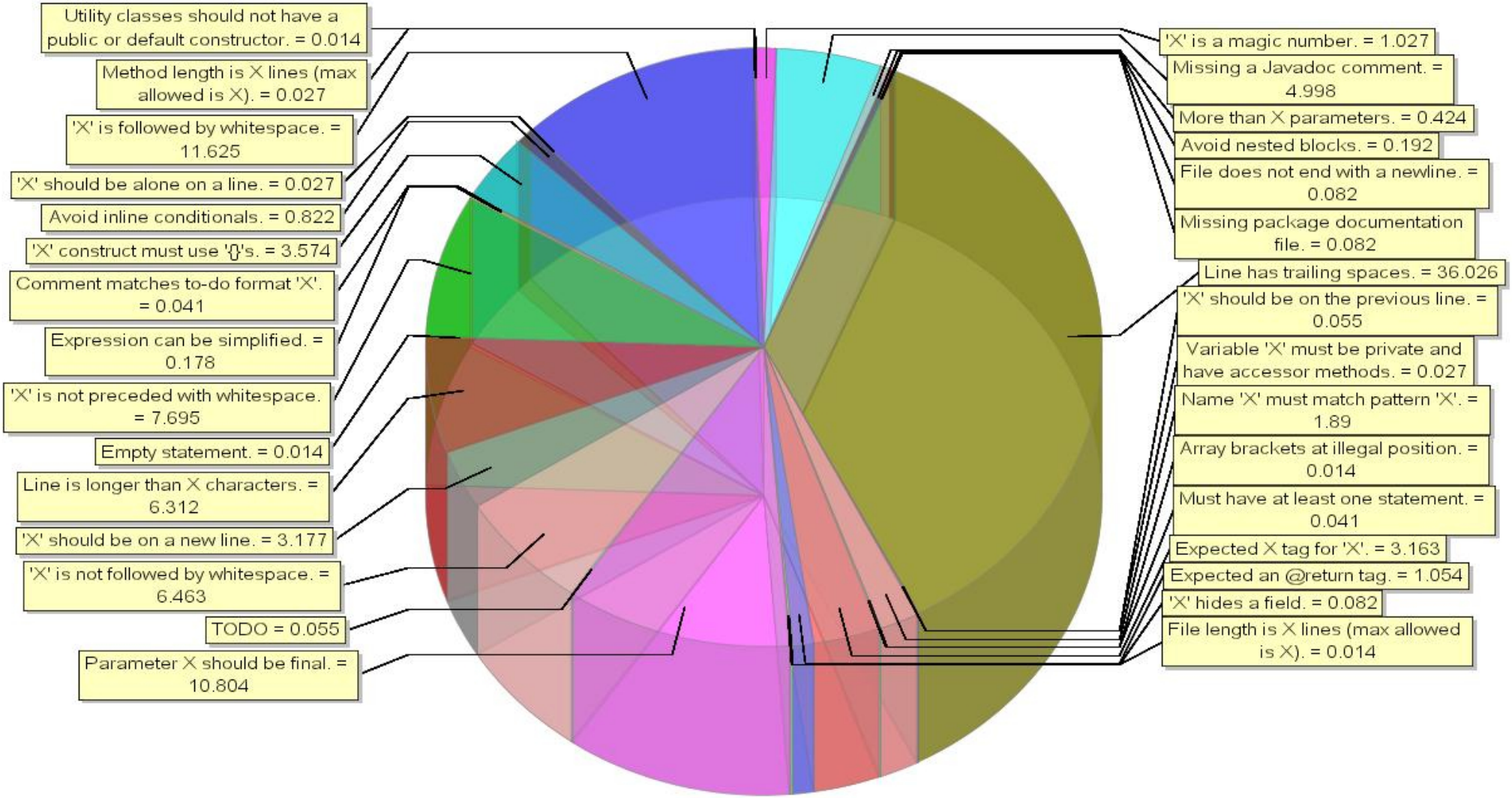
Data-ink Ratio

Multi-functioning
graphical elements

Data density

Small multiples

Chartjunk & Subtraction



http://www.theusrus.de/blog/yet-another-pie-chart/

EDWARD TUFTE'S IDEAS

(from “The Visual Display of Quantitative Information,” Tufte 1983.)

Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

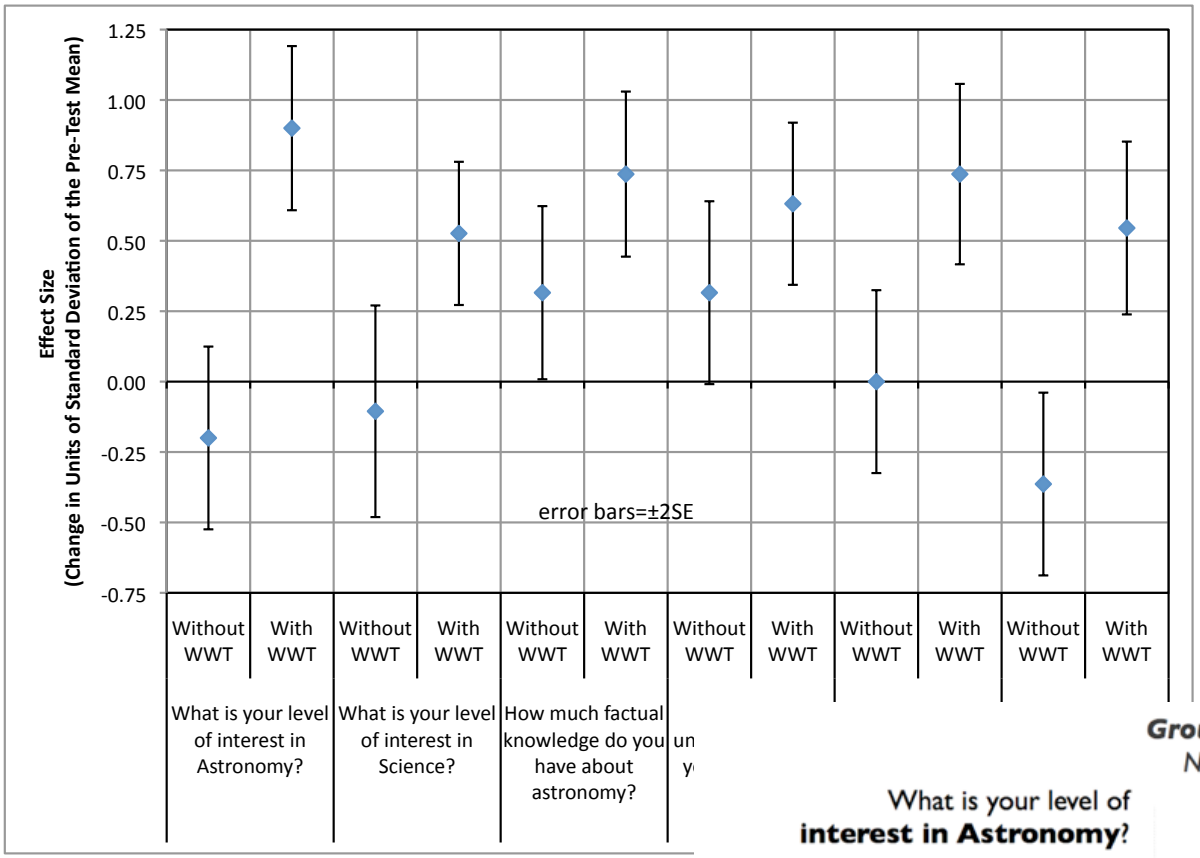
Data-ink Ratio

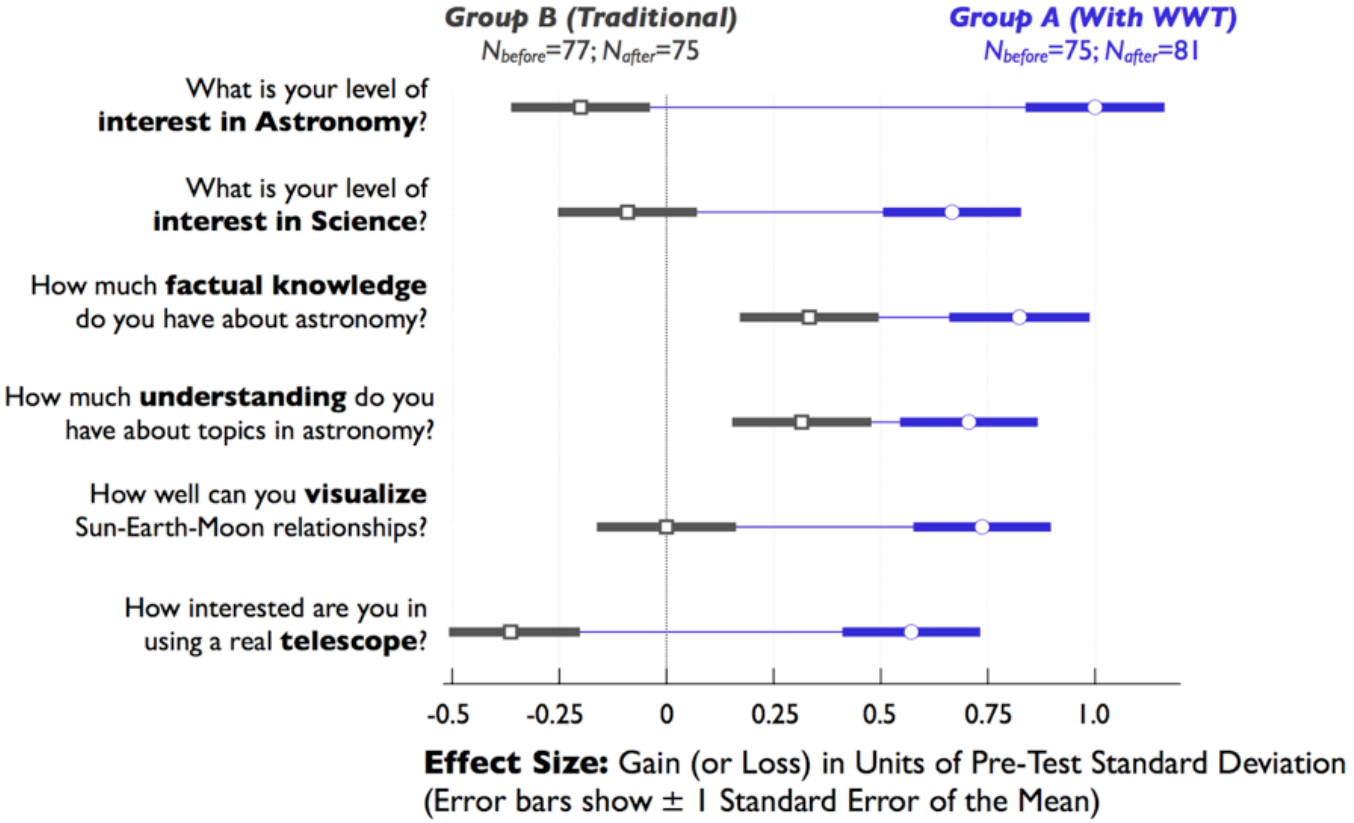
Multi-functioning
graphical elements

Data density

Small multiples

Data-ink Ratio





from AG example used in VISUAL STRATEGIES

EDWARD TUFTE'S IDEAS

(from “The Visual Display of Quantitative Information,” Tufte 1983.)

Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

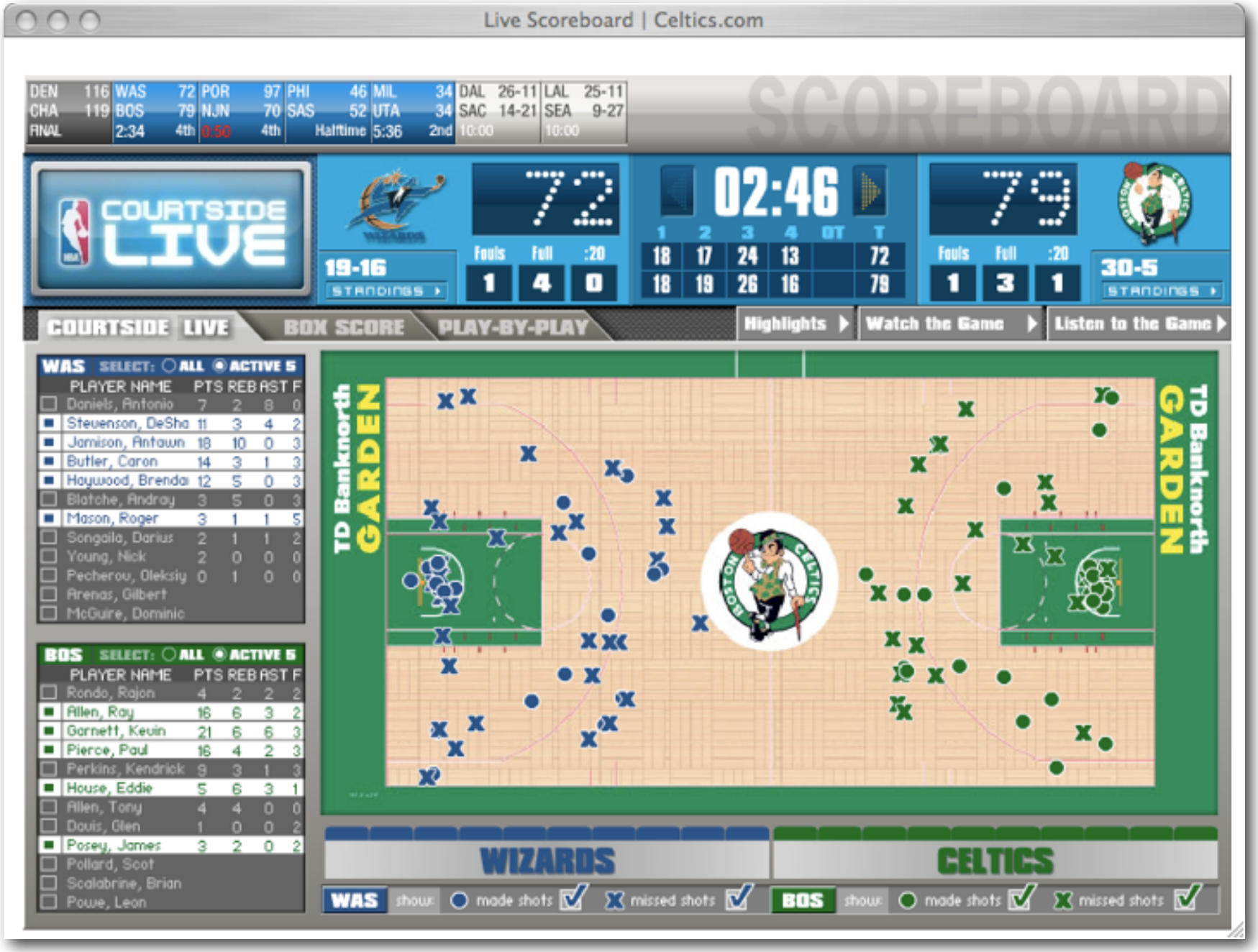
Data-ink Ratio

Multi-functioning
graphical elements

Data density

Small multiples

Multi-functioning graphical elements



EDWARD TUFTE'S IDEAS

(from “The Visual Display of Quantitative Information,” Tufte 1983.)

Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

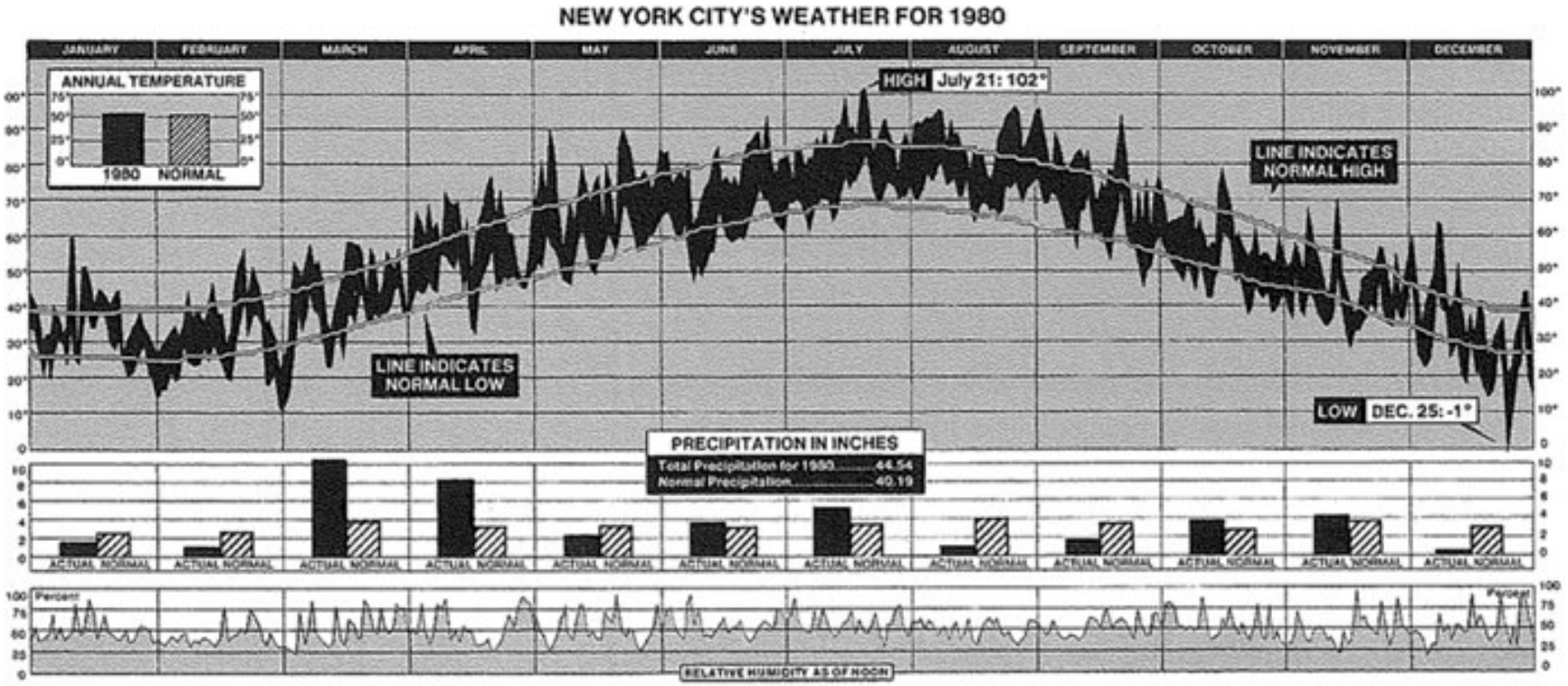
Data-ink Ratio

Multi-functioning
graphical elements

Data density

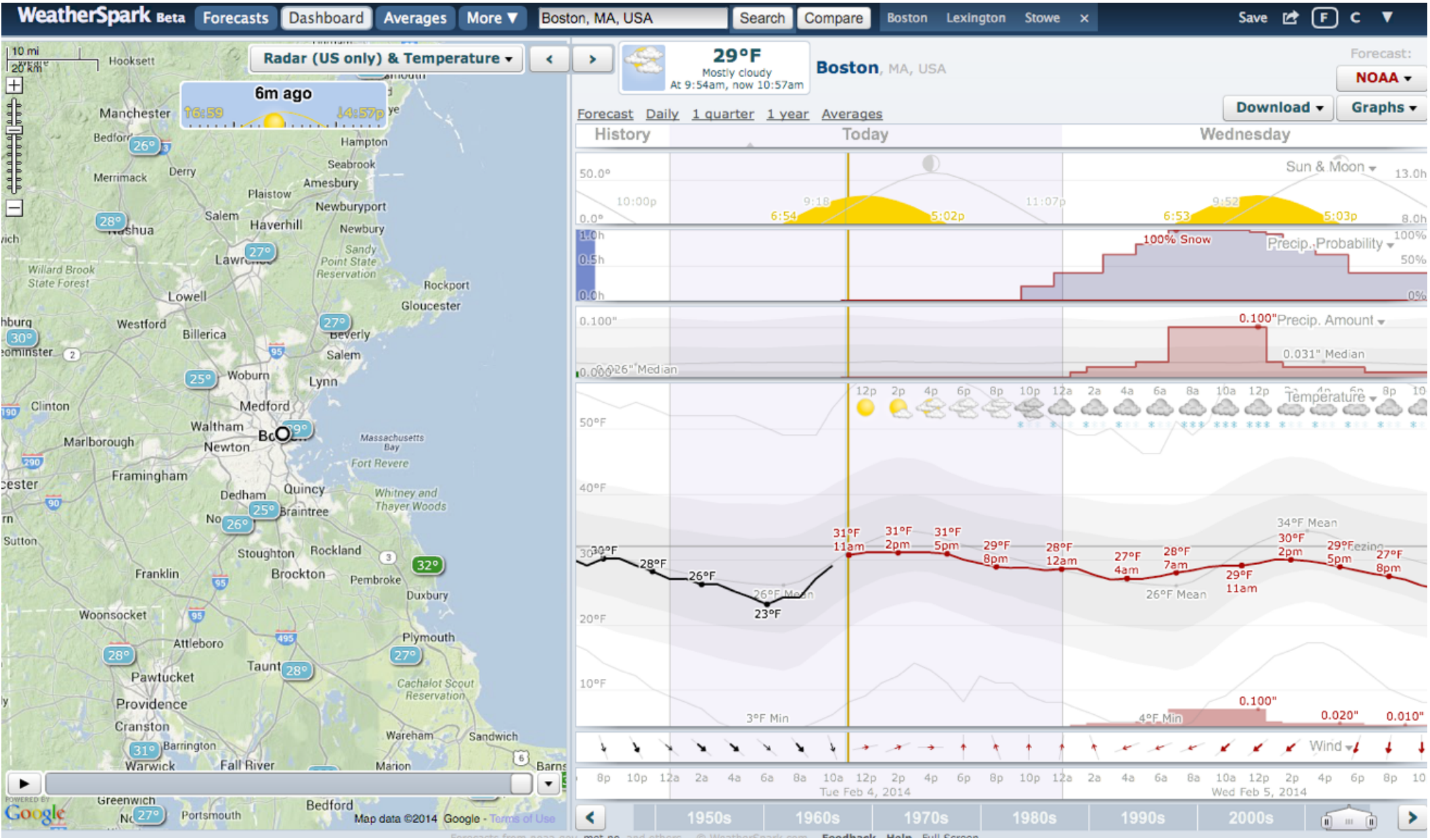
Small multiples

Data density



New York Times, January 11, 1981, p. 32.

Data (over?) density



http://weatherspark.com/#!dashboard;a=USA/MA/Boston

EDWARD TUFTE'S IDEAS

(from “The Visual Display of Quantitative Information,” Tufte 1983.)

Graphical Excellence

Graphical Integrity
& “The Lie Factor”

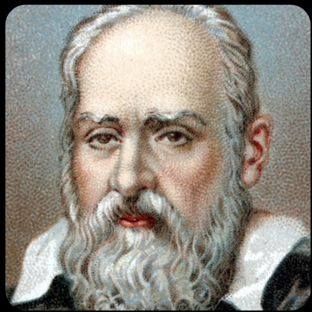
Chartjunk
& Subtraction

Data-ink Ratio

Multi-functioning
graphical elements

Data density

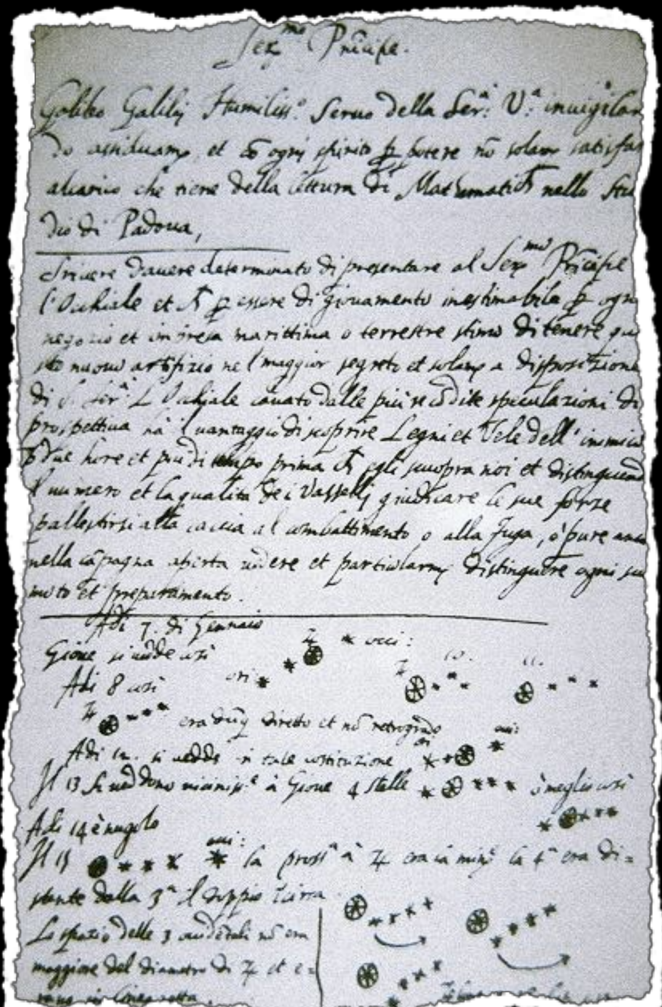
Small multiples

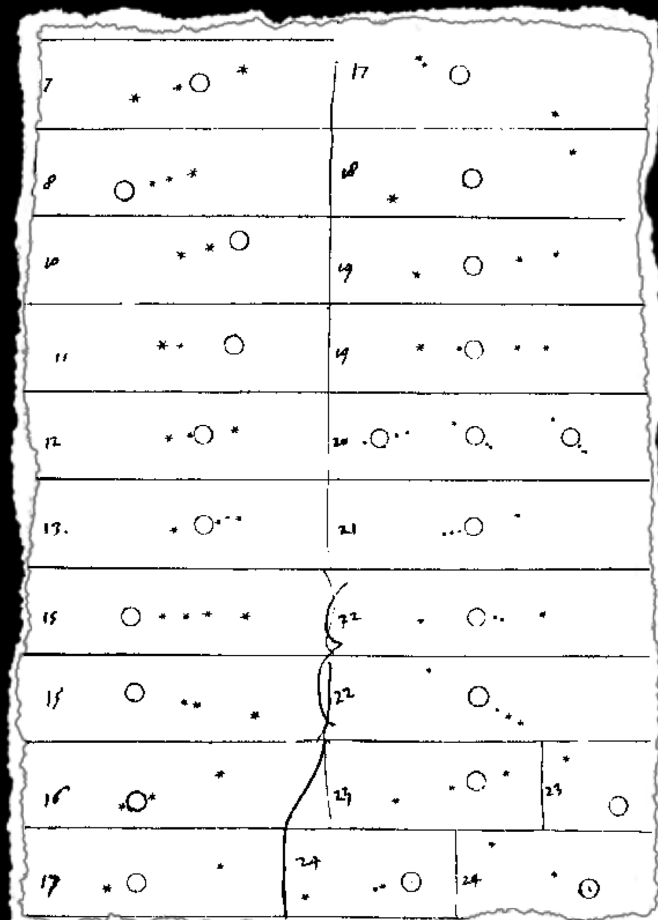


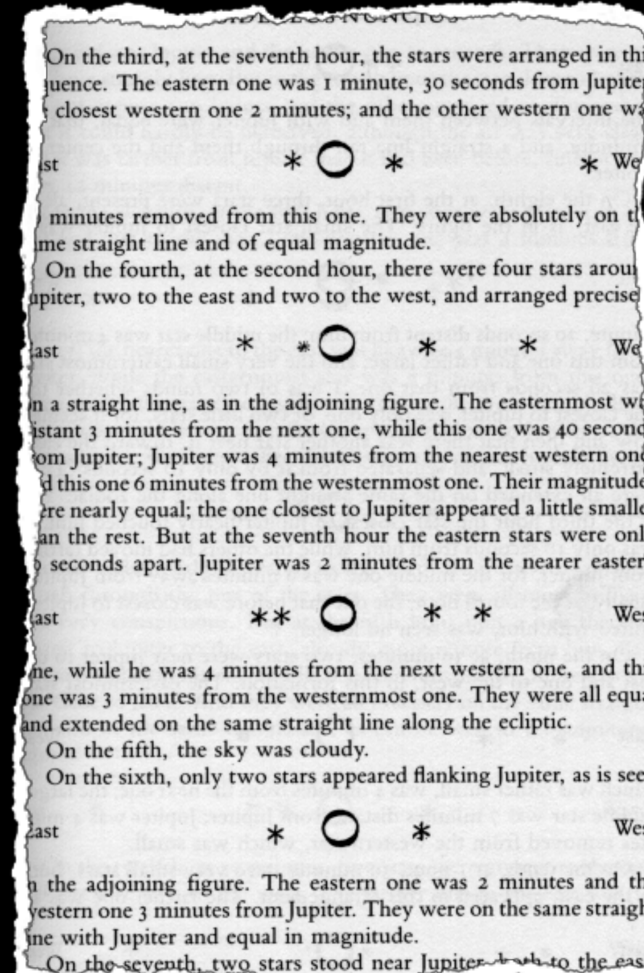
GALILEO GALILEI

(1564-1642)

Small multiples







On the third, at the seventh hour, the stars were arranged in thi
quence. The eastern one was 1 minute, 30 seconds from Jupite
closest western one 2 minutes; and the other western one w

ast * ○ * * We

minutes removed from this one. They were absolutely on t
me straight line and of equal magnitude.

On the fourth, at the second hour, there were four stars arou
upiter, two to the east and two to the west, and arranged precise

ast * *○ * * We

n a straight line, as in the adjoining figure. The easternmost w
istant 3 minutes from the next one, while this one was 40 second
om Jupiter; Jupiter was 4 minutes from the nearest western on
d this one 6 minutes from the westernmost one. Their magnitude
ere nearly equal; the one closest to Jupiter appeared a little smalle
an the rest. But at the seventh hour the eastern stars were onl
seconds apart. Jupiter was 2 minutes from the nearer easter

ast ** ○ * * Wes

ne, while he was 4 minutes from the next western one, and thi
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

ast * ○ * Wes

n the adjoining figure. The eastern one was 2 minutes and th
western one 3 minutes from Jupiter. They were on the same straigh
ine with Jupiter and equal in magnitude.

On the seventh, two stars stood near Jupiter

Notes for & re-productions of Siderius Nuncius

COMPOSITE TABLE*

*how AG met ET, 1995

Table 1: Summary of VLA NH_3 Observations in Low-Mass Dense Cores as of 1994
(double-boxed sources to be observed in this proposal)

Source	Central Source	Outflow	Year of Obsv'n.	Relationship of VLA NH_3 to Central Source	Comments (all observations are by the current proposers, unless otherwise indicated)
L1172	★		1984		
L1551	★		1989		Torrelles et al. 1985.
L1152	★	*	1988		improve sensitivity and extend map with proposed observations
TMC-1A (≡ L1534)	★		1988		improve sensitivity with proposed observations
B335	★		1990		Keene, Menten & Masson, in prep.
L1689N	NO	NO	1985		Wootten and Loren 1987. Star shown at left (w/outflow) not within primary beam of observation.
AQ13	NO	?	1988		noise higher in on-line channels, suggestive of signal
B217	NO	NO	1983		relatively poor sensitivity
L1582	★	?	1984		
L1498	NO	NO	1984		
B5	★		1985		relatively poor sensitivity
B1	★		1988		relatively poor sensitivity

New Sources to be Observed in this Proposal

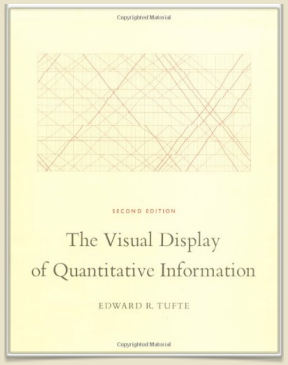
L1527	★		1995		infall candidate
L483	★		1995		infall candidate
B133	NO	NO	1995		strong extended sub-mm continuum, no known point source

*determination is unclear from existing data

Goodman et al. NRAO/VLA Proposal, 1994

HOW Best to Visualize, with examples

READ “The Visual Display of Quantitative Information,” Tufte 1983



Graphical Excellence

Graphical Integrity
& “The Lie Factor”

Chartjunk
& Subtraction

Data-ink Ratio

Multi-functioning
graphical elements

Data density

Small multiples