Data Visualization

Spring 2017

Matthew Turk

Basics

10AM-1PM Fridays, LIS-46

Matthew Turk - mjturk@illinois.edu

Office Hours: Wednesday 2-4, LIS 222

TA Jarai Carter - carter31@illinois.edu

http://github.com/UIUC-iSchool-DataViz/

Schedule

Week 1 (Jan 20): Introduction, syllabus, and examples of visualization
Week 2 (Jan 27): History of visualization
Week 3 (Feb 3): Simple plotting: quantitative plots, breakdown of plot components
Week 4 (Feb 10): Histograms, binning, and distributions
Week 5 (Feb 17): Images: color theory, colormaps
Week 6 (Feb 23): Quantitative and time series
Week 7 (Mar 2): Geospatial visualizations
Week 8 (Mar 10): Synthesizing multiple datasets
Week 9 (Mar 17): Software ecosystem around visualization
Week 10 (Mar 31): Network visualization
Week 11 (Apr 7): Statistical visualization
Week 12 (Apr 14): Interactive visualizations
Week 13 (Apr 21): Advanced topics
Week 14 (Apr 27): Group presentations

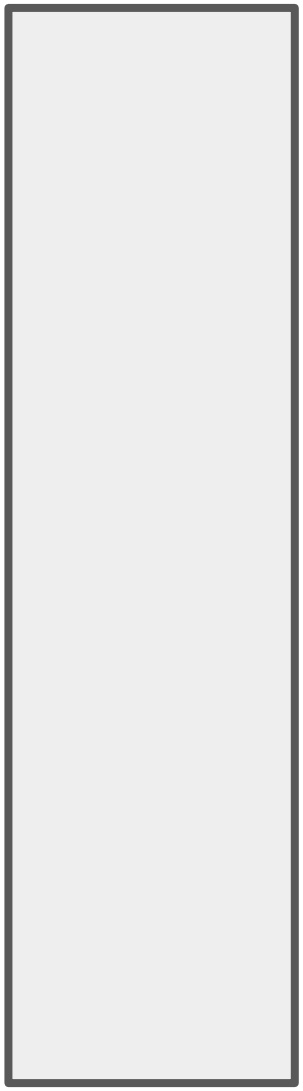
Overview - Themes

1. What are the components of an effective visualization of quantitative data?
2. What tools and ecosystems are available for visualizing data?
3. What systems can be put in place to generate visualizations rapidly and with high-fidelity representation?

Overview - Goals

- Students will be able to communicate information and data through visual representation
- Students will be able to examine a visualization and understand how it can be improved upon
- Students will have facility with the commonplace tools used for visualization, and a deeper understanding of where those tools have shortcomings

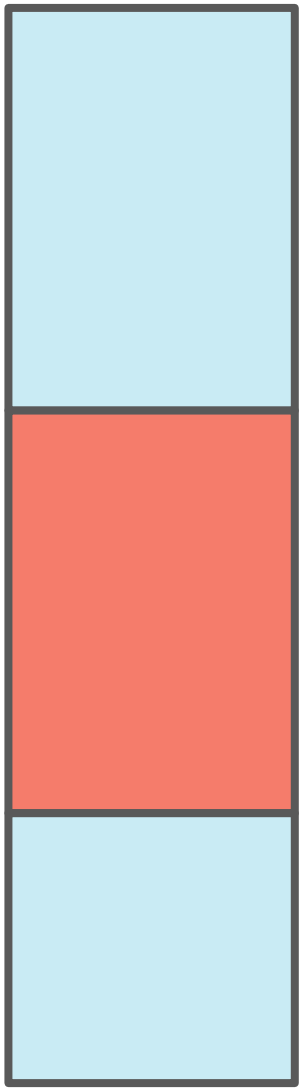
Overview - Structure



10:00 AM

12:50 PM

Overview - Structure



Topic introduction and lecture

Hands-on examples and discussion

Wrap-up

Overview - Grading

70%: Weekly Assignments in prose or code form
30%: Final project

Overview - Assignments

- Weekly, assigned in class, collected following class
- Prose assignments: deconstruction or analysis of a visualization or a dataset.
- Coding assignments: Jupyter notebooks following step by step through collection and processing of data and the visualization of that data

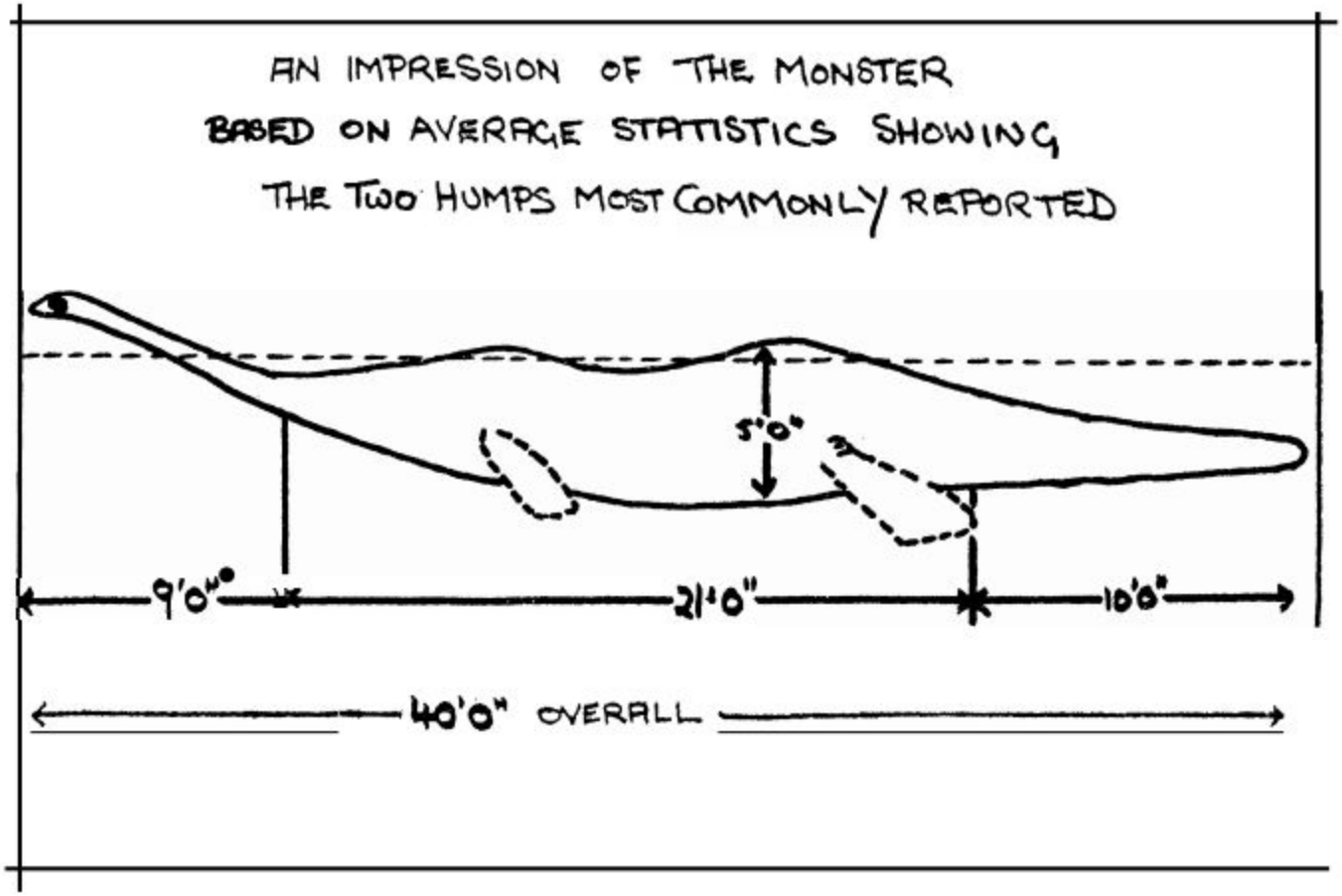
Plagiarism

Plagiarism is about copying ideas.

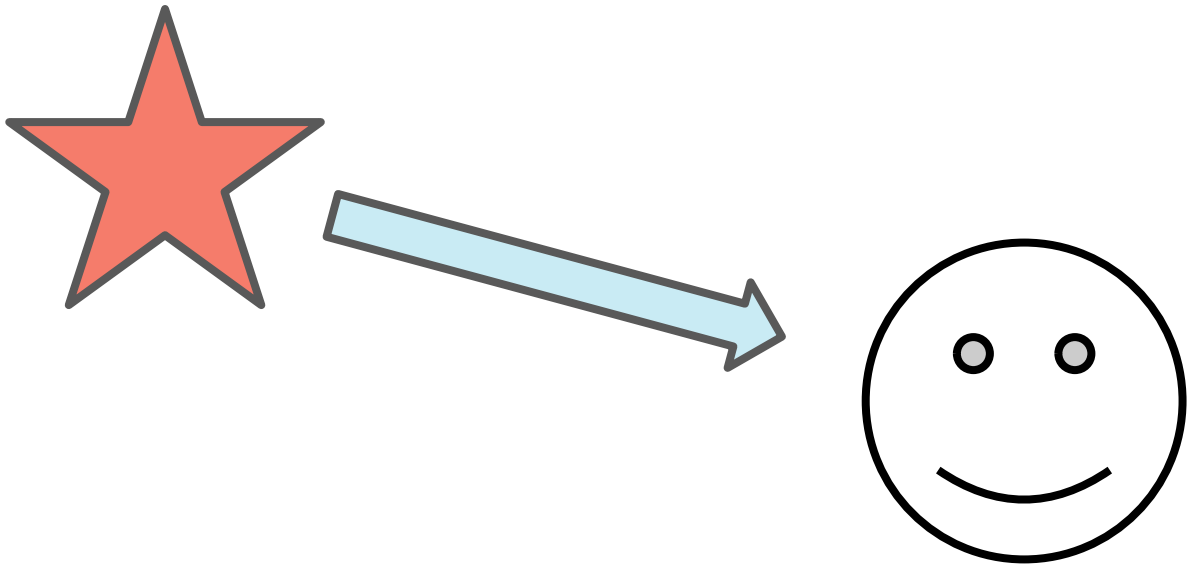
Cite all code you utilize from elsewhere.

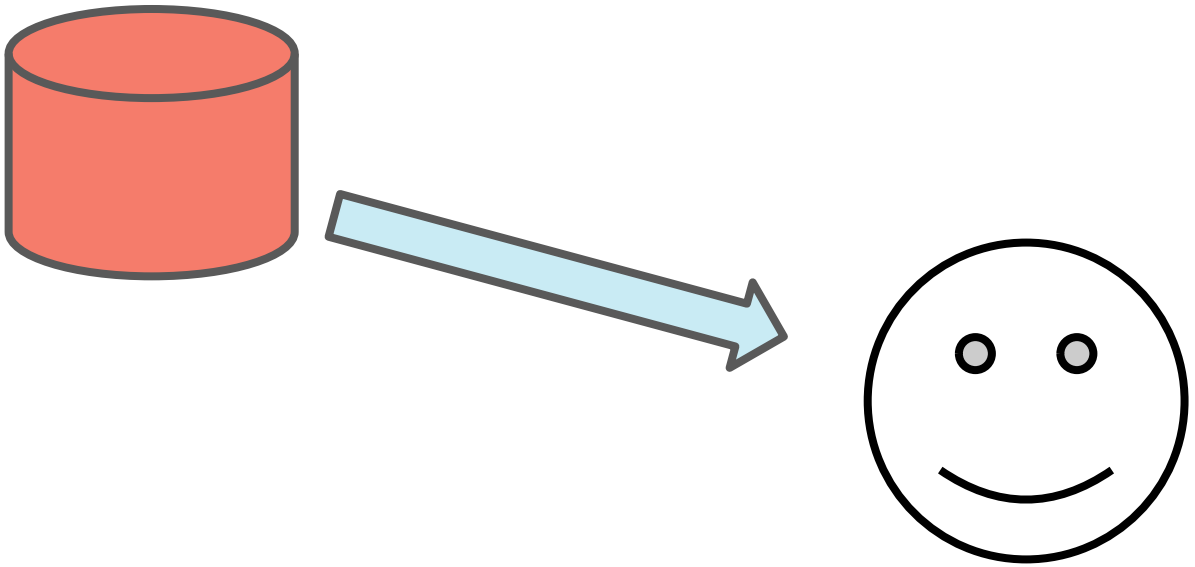
goo.gl/WzIDCa

Visualization


AN IMPRESSION OF THE MONSTER
BASED ON AVERAGE STATISTICS SHOWING
THE TWO HUMPS MOST COMMONLY REPORTED
5'0"
9'0"
21'0"
10'0"
40'0" OVERALL

Gowler, 1971

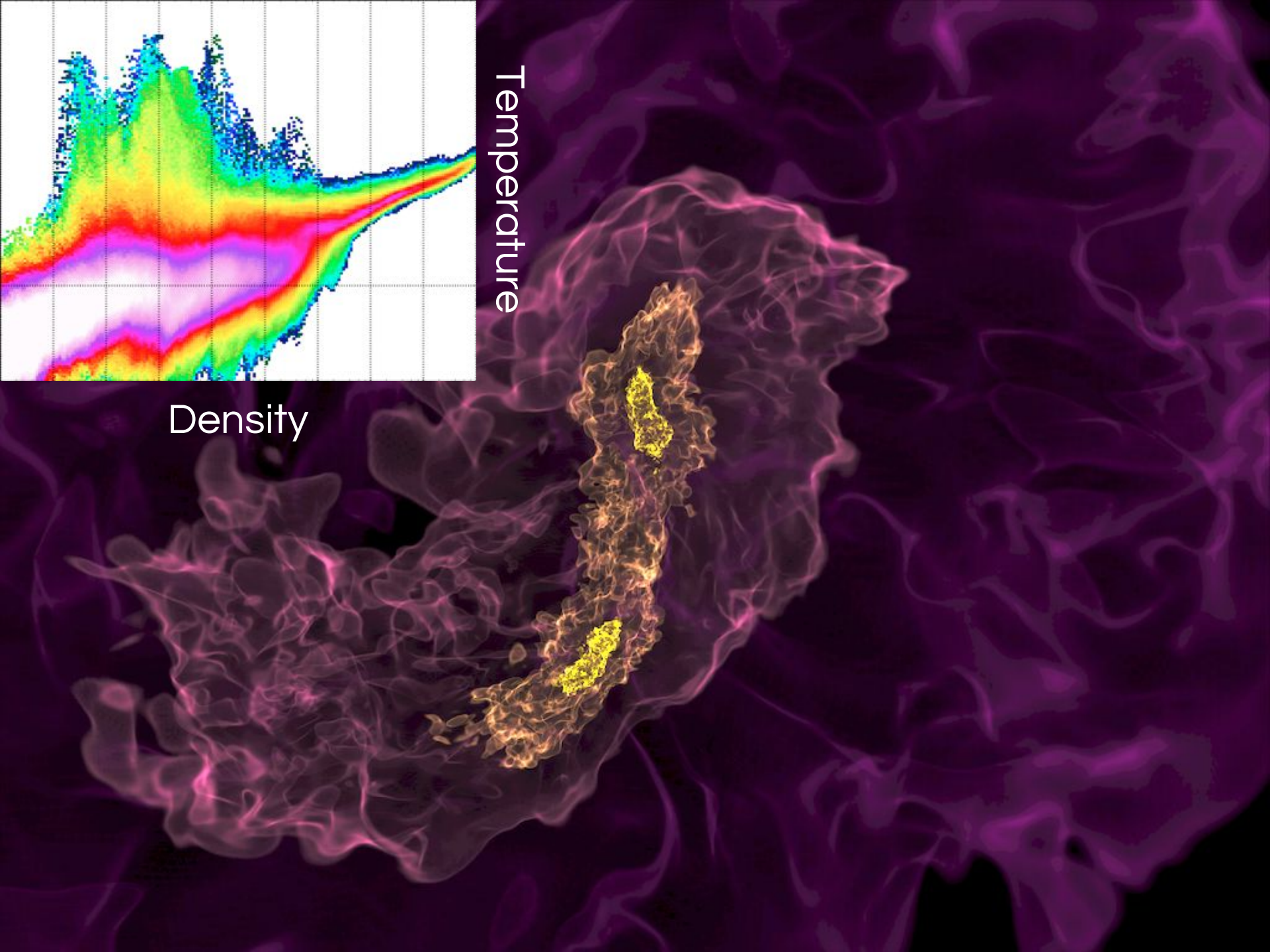


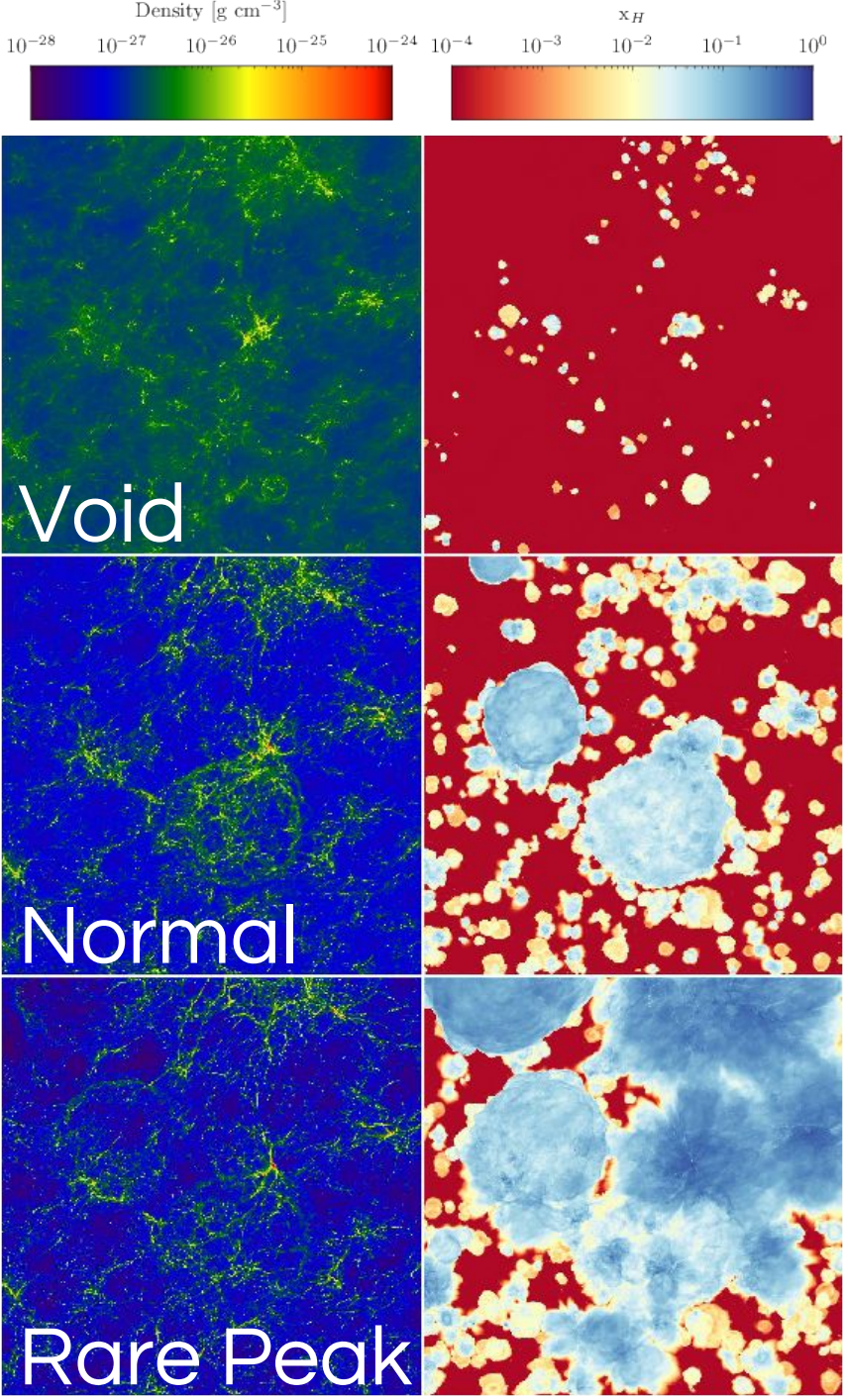


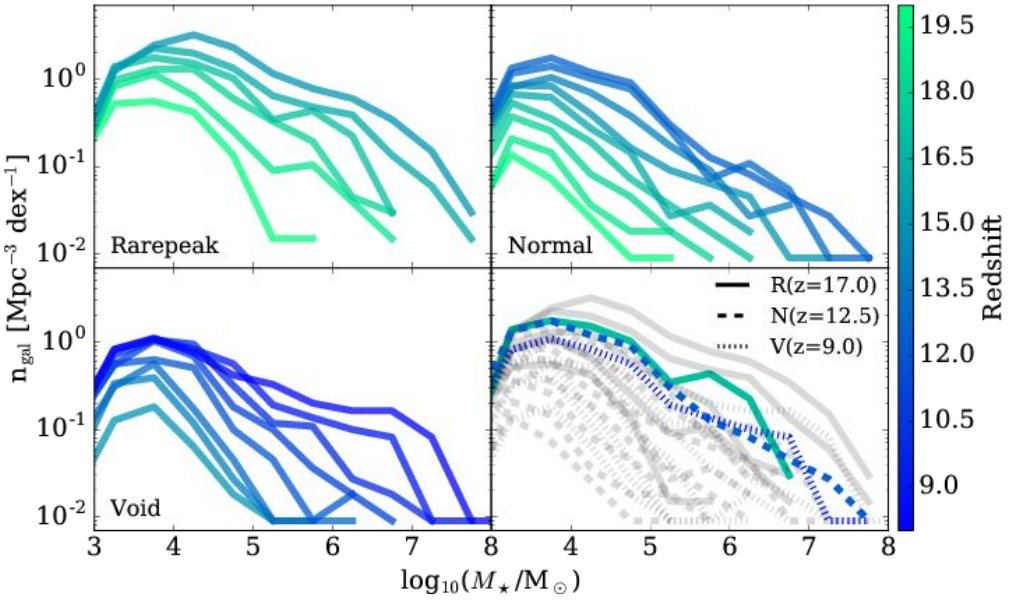
Visualization for self

Visualization for peers

Visualization for others


Temperature
Density


Density [g cm^{-3}]
10^{-28} 10^{-27} 10^{-26} 10^{-25} 10^{-24}
x_H
10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^{0}
Void
Normal
Rare Peak

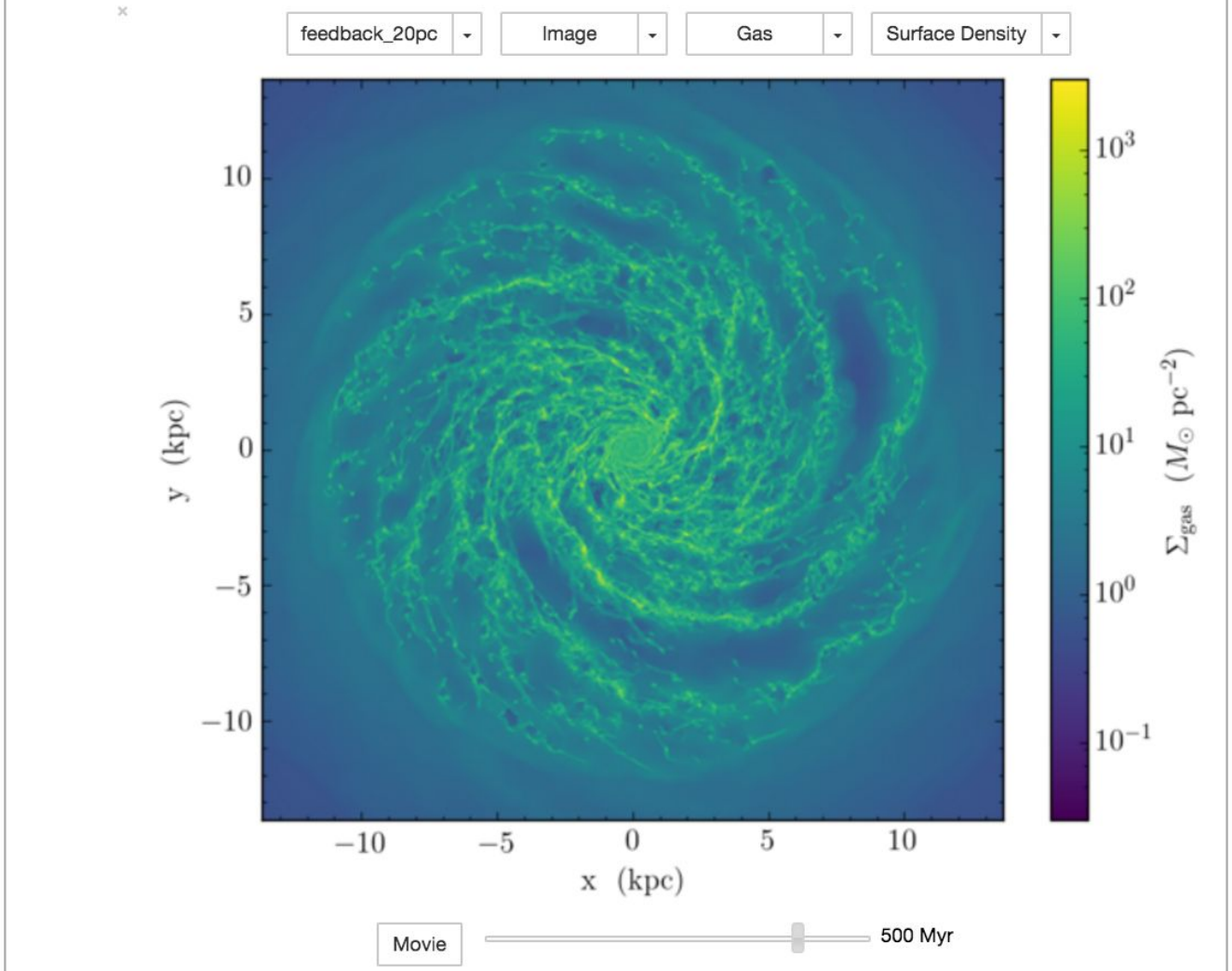

Rarepeak
Normal
Void
R(z=17.0)
N(z=12.5)
V(z=9.0)
n_{gal} [Mpc^{-3} dex^{-1}]
$\log_{10}(M_\star/M_\odot)$
Redshift
19.5
18.0
16.5
15.0
13.5
12.0
10.5
9.0

Galaxy Visualization Widget

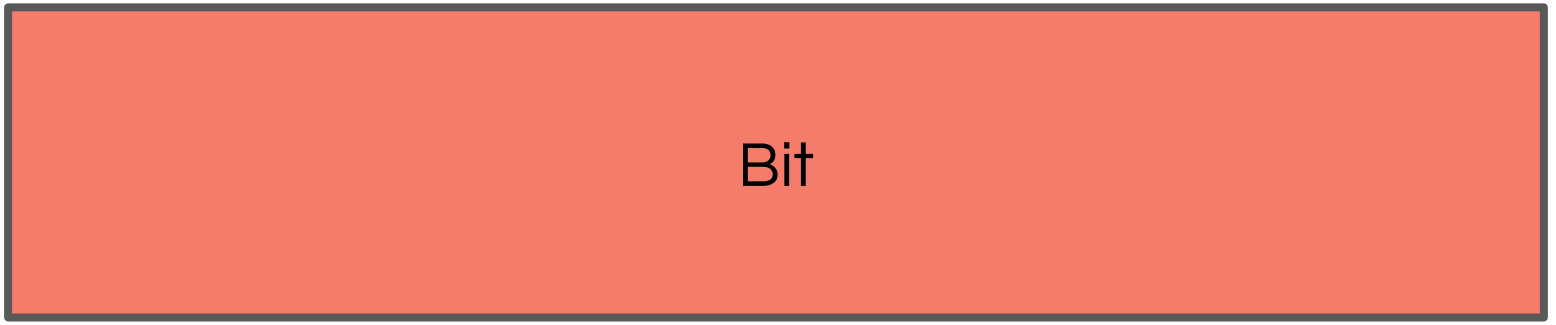
This notebook defines a simple widget used to visualize various fields derived from a set of 3D AMR hydrodynamic simulation of an isolated disk galaxy. The full simulation dataset weighs in at around 15 TB, but these sorts of interactive visualizations make it easy to quickly and easily visualize the physically important parts of the data.

The widget should appear after the final notebook cell.

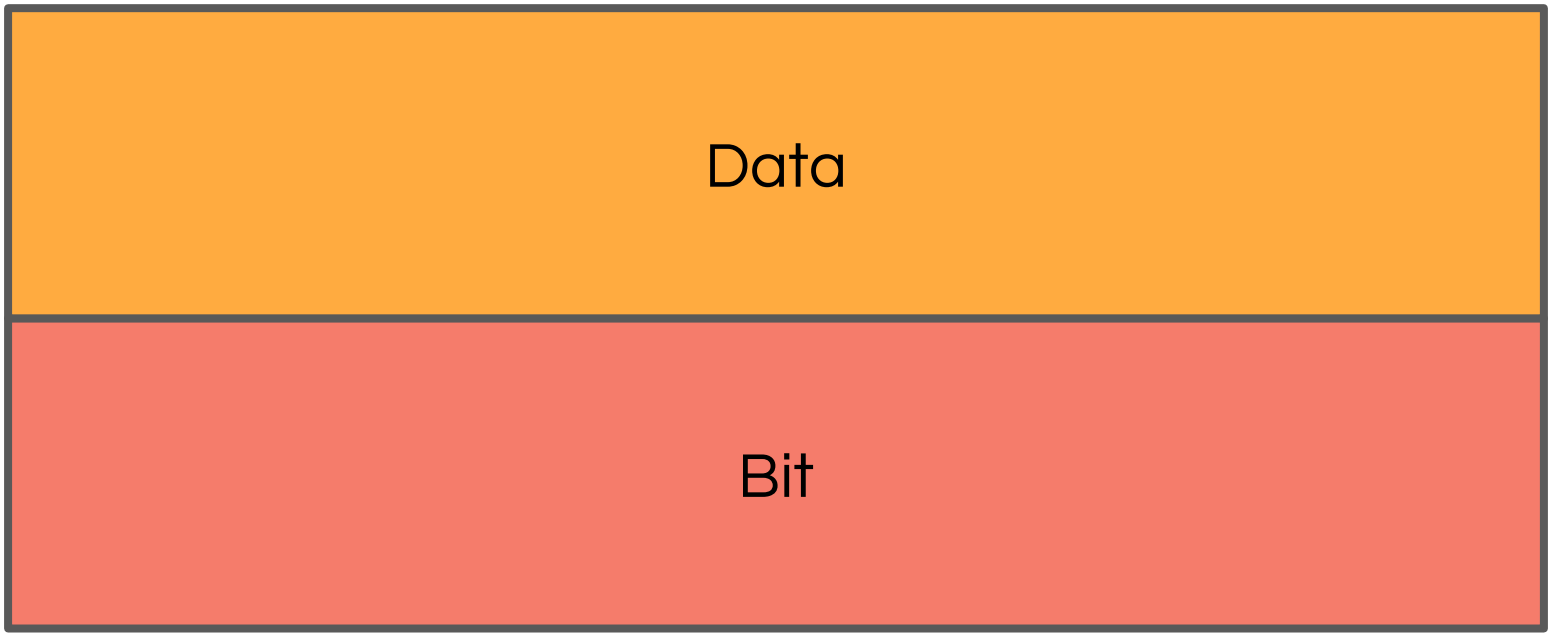
In [1]:
```
from galanyl.widget import galaxy_widget
from IPython.display import display

display(galaxy_widget)
```



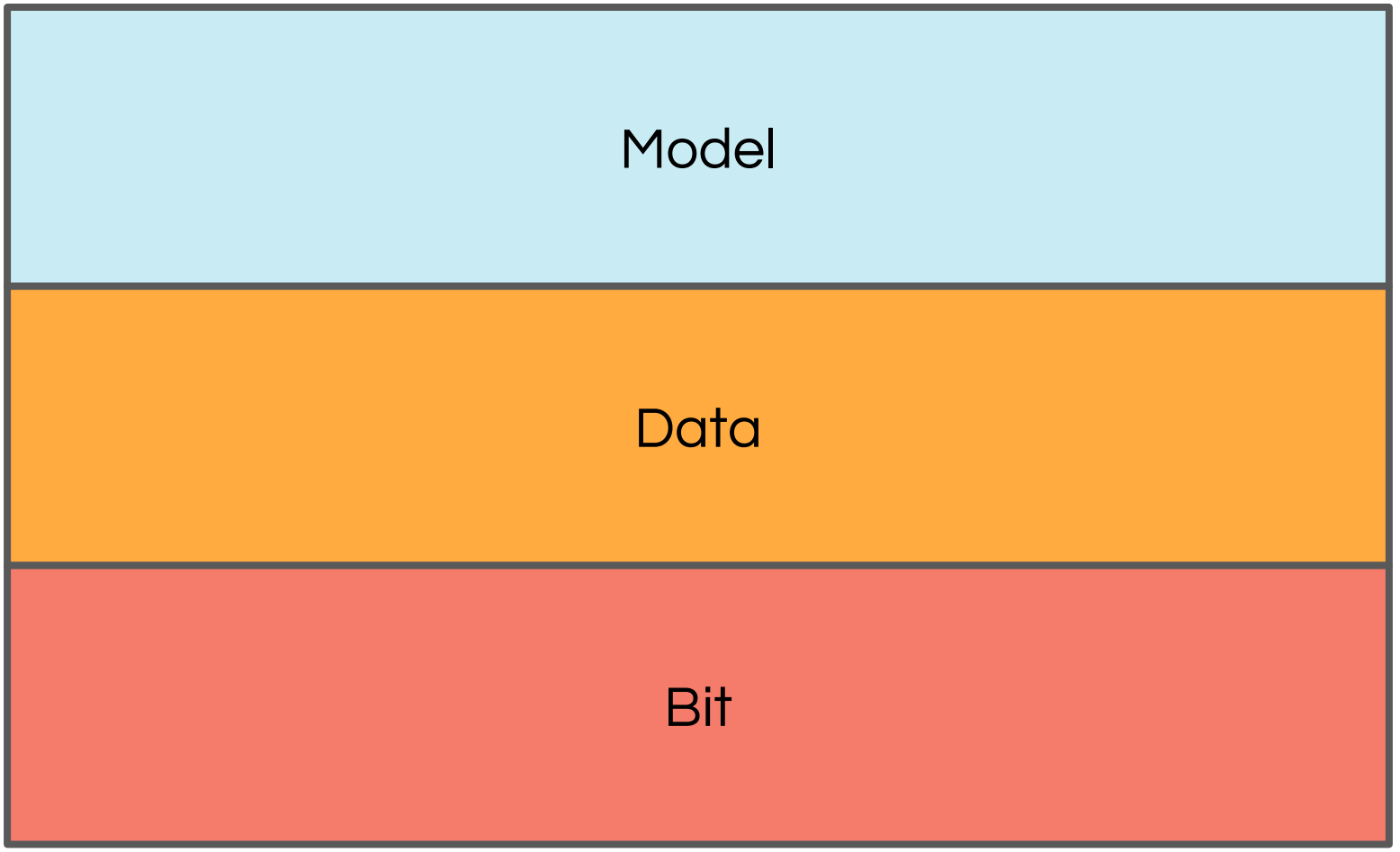
Layers of Representation



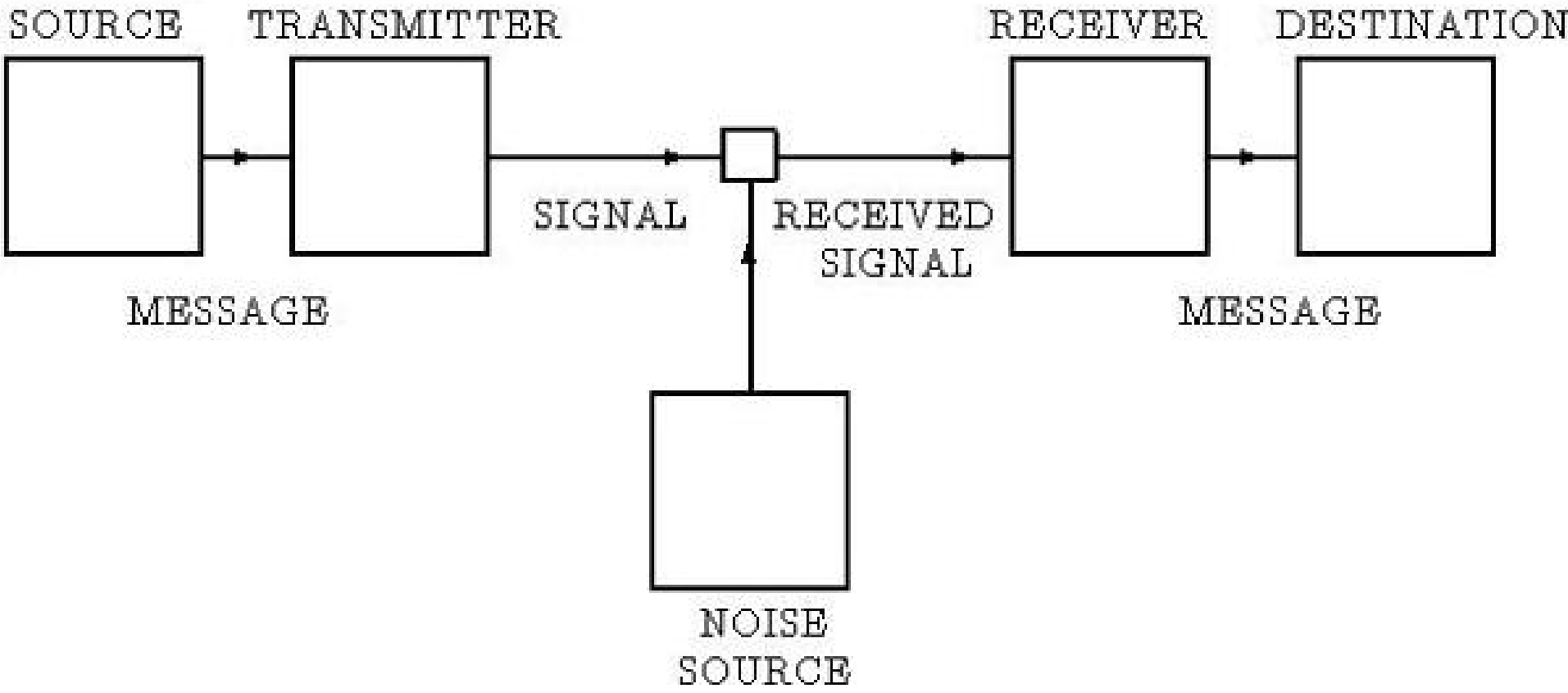
Layers of Representation



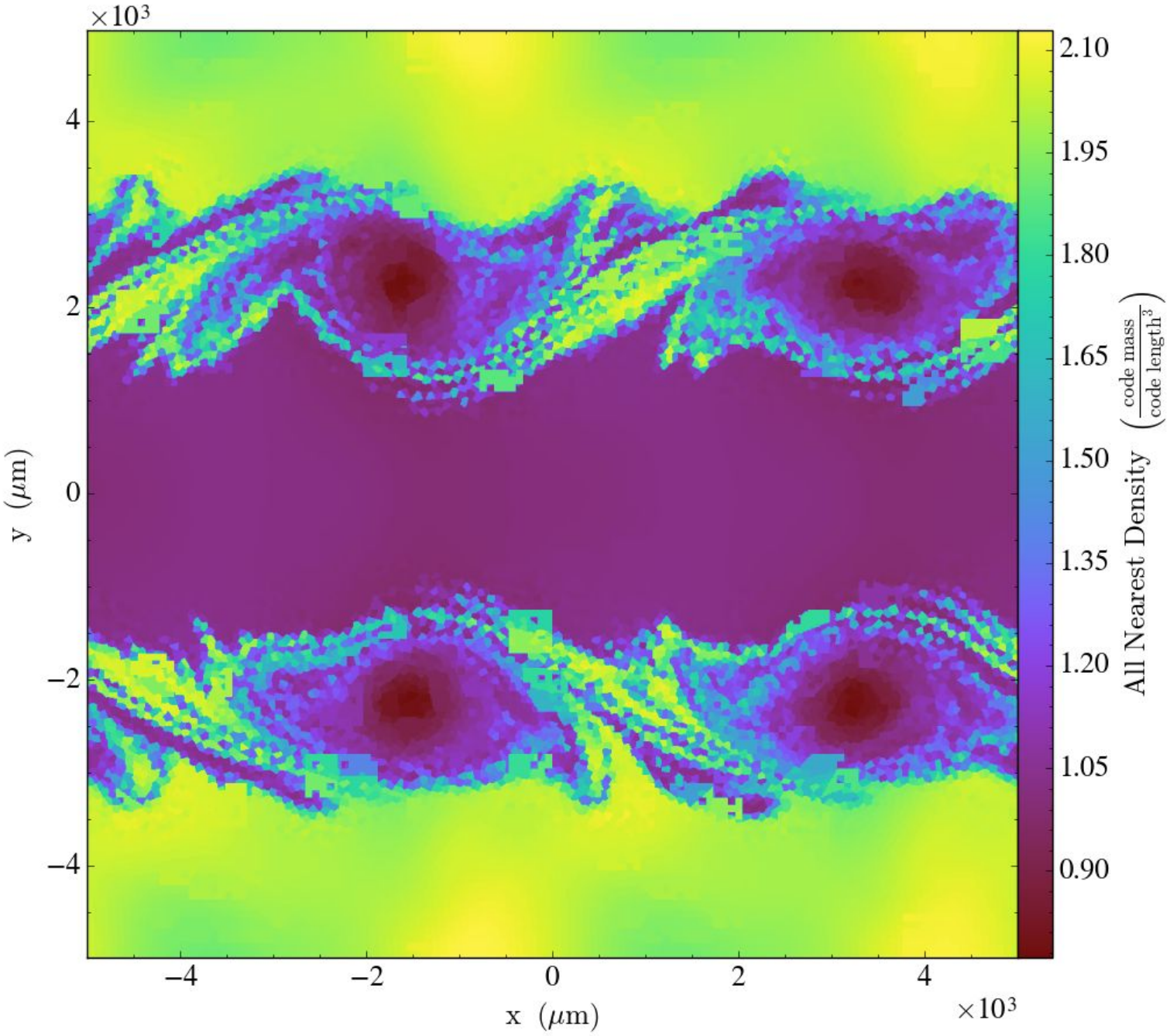
Layers of Representation

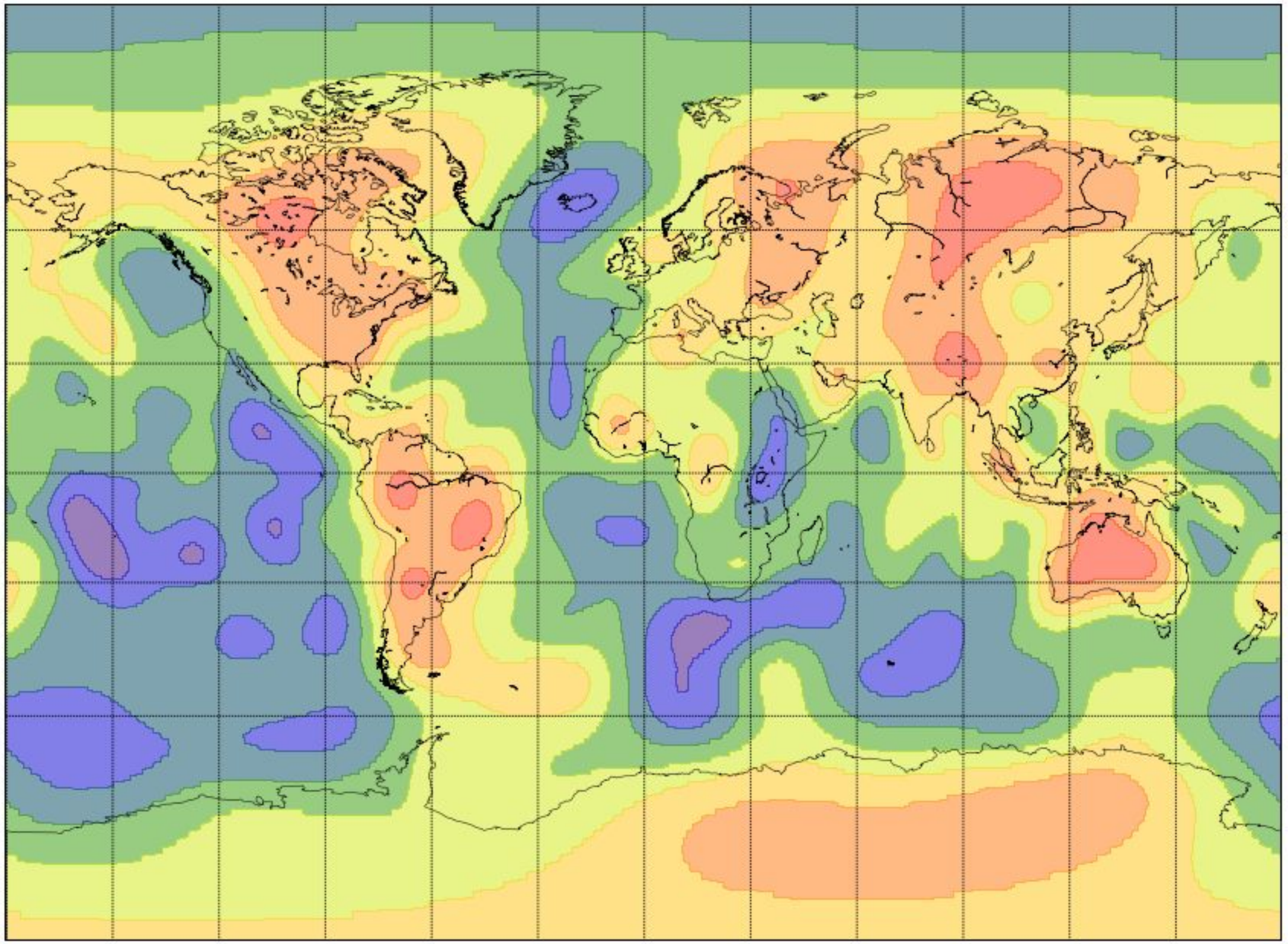


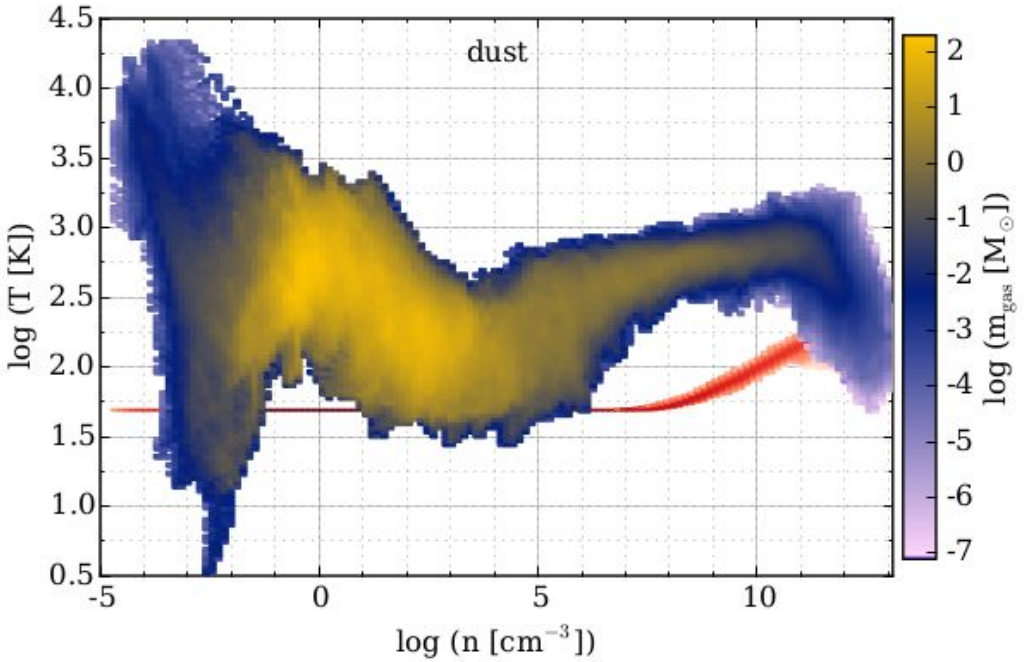
We tell lies to visualize.

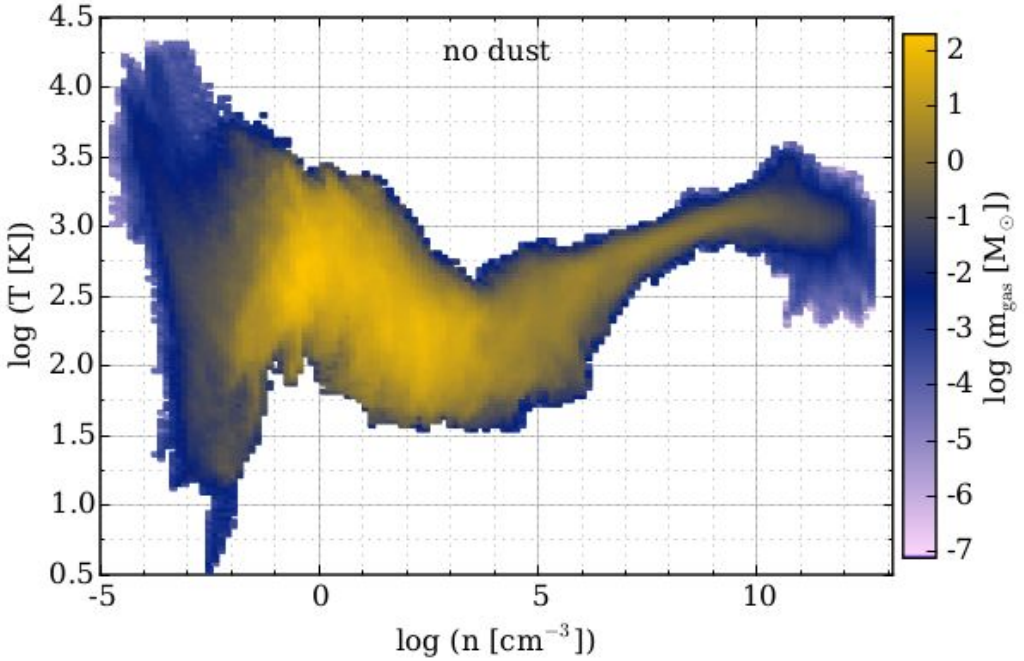

SOURCE
TRANSMITTER
RECEIVER
DESTINATION
SIGNAL
RECEIVED
SIGNAL
MESSAGE
MESSAGE
NOISE
SOURCE

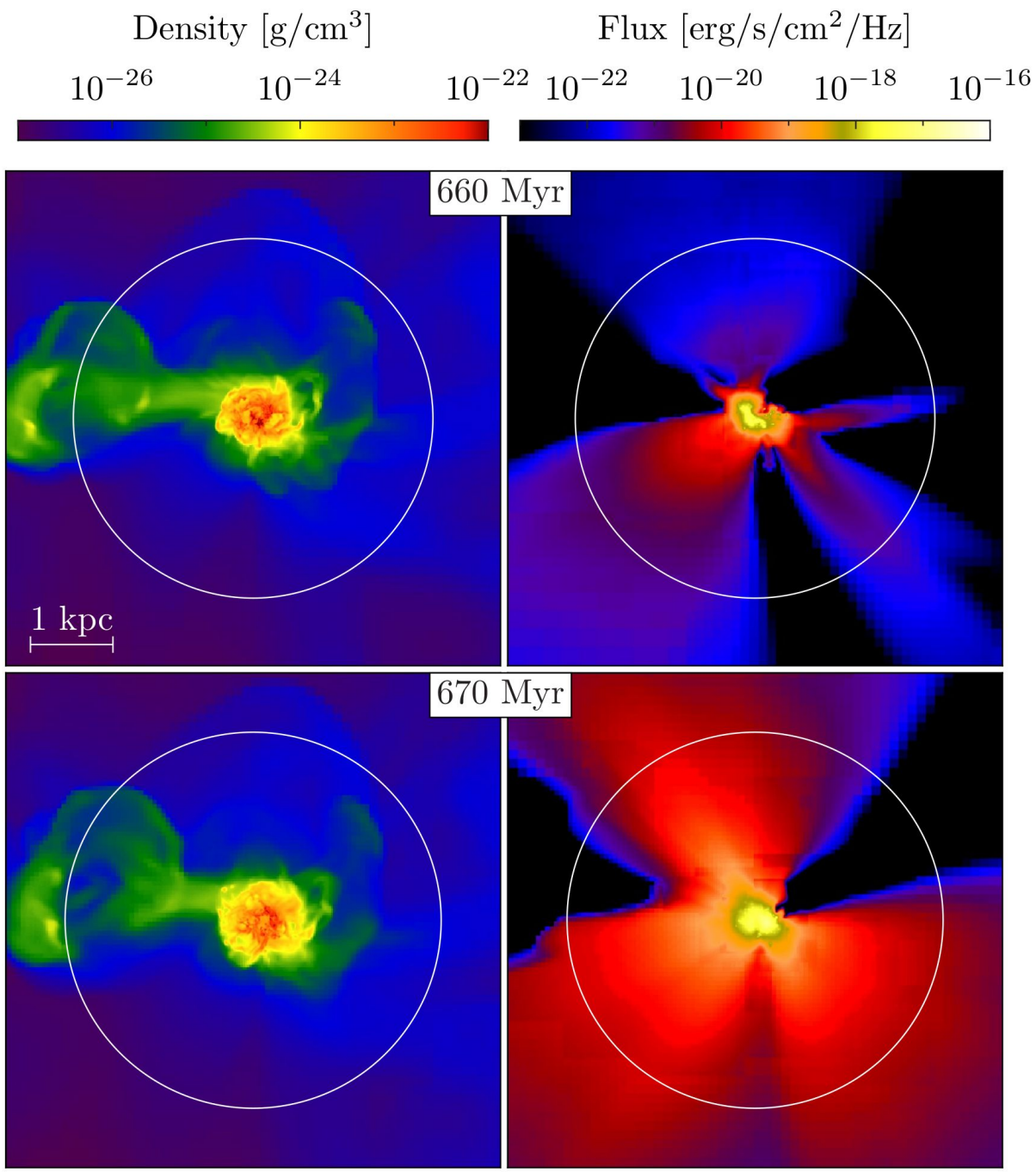
- The collection of the data
- The organization of that data
- Components of representation of that data
- Component generation


×10³
4
2
0
−2
−4
y (μm)
−4
−2
0
2
4
x (μm)
×10³
2.10
1.95
1.80
1.65
1.50
1.35
1.20
1.05
0.90
All Nearest Density $\left(\frac{\text{code mass}}{\text{code length}^3}\right)$




dust
log (T [K])
log (n [cm^{-3}])
log (m_{gas} [$M_\odot$])
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-5
0
5
10
2
1
0
-1
-2
-3
-4
-5
-6
-7


no dust
log (T [K])
log (n [cm^{-3}])
log (m_{gas} [$M_\odot$])
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-5
0
5
10
2
1
0
-1
-2
-3
-4
-5
-6
-7


Density [g/cm^3]
10^{-26}
10^{-24}
10^{-22}
Flux [erg/s/cm^2/Hz]
10^{-22}
10^{-20}
10^{-18}
10^{-16}
660 Myr
1 kpc
670 Myr

Line of sight velocity profiles in turbulent molecular clouds

This is a snapshot from a three dimensional simulation of about 5,700 Solar masses of gas in a 10 pc periodic box. The gas began with fully developed turbulence at Mach 8; this is after 1.27 Myr of self-gravitating evolution. Click around the image to explore the line-of-sight velocity structure.

*pick a column density image (current choice is **darker**)*

total column density (log scale)

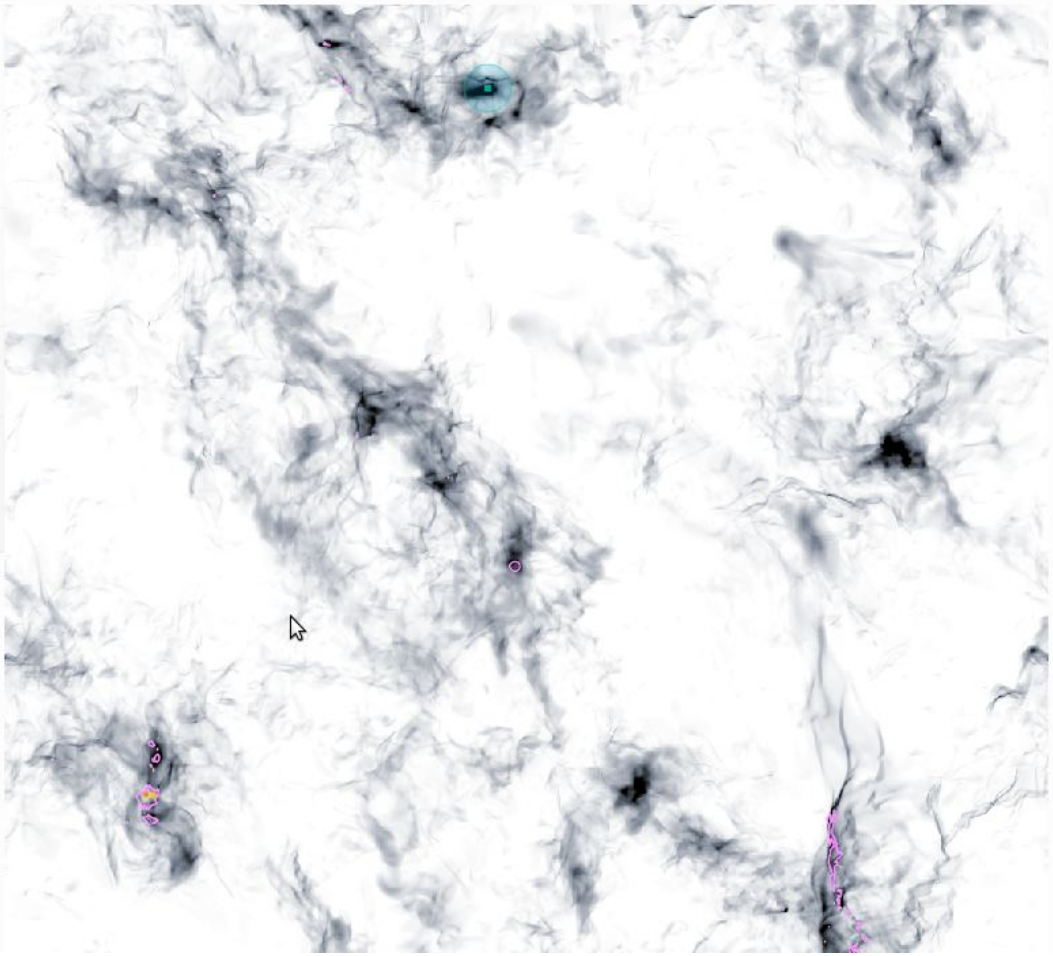
10^3 < n cm^{-3} < $10^{4.5}$ (linear scale; approximates C^{18}O)

toggle high density contours and sink particles

$10^{4.5}$ < n cm^{-3} (approximates N_2H^+) **sink particles**

click the image to inspect a point

tracer :	total	C^{18}O	N_2H^+
log N / cm^{-2} :	22.26	22.18	0



Assignment 1

- Gaia satellite: http://sci.esa.int/gaia/
- Examine subset of data
- Draw out *something* from this data
- Describe in detail what you did with the data, why you did what you did, and the context for these choices.

goo.gl/mSALXN

Overview - Installation

- Conda / Anaconda

  https://www.continuum.io/downloads
- We will utilize numpy, matplotlib, pandas, and seaborn
- Potential additional topics include altair, bokeh, datashader and yt