Kalina Borkiewicz
Advanced Visualization Lab
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

AJ Christensen
Advanced Visualization Lab
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

Helen-Nicole Kostis
Scientific Visualization Studio
USRA/GESTRA & NASA Goddard Space Flight Center

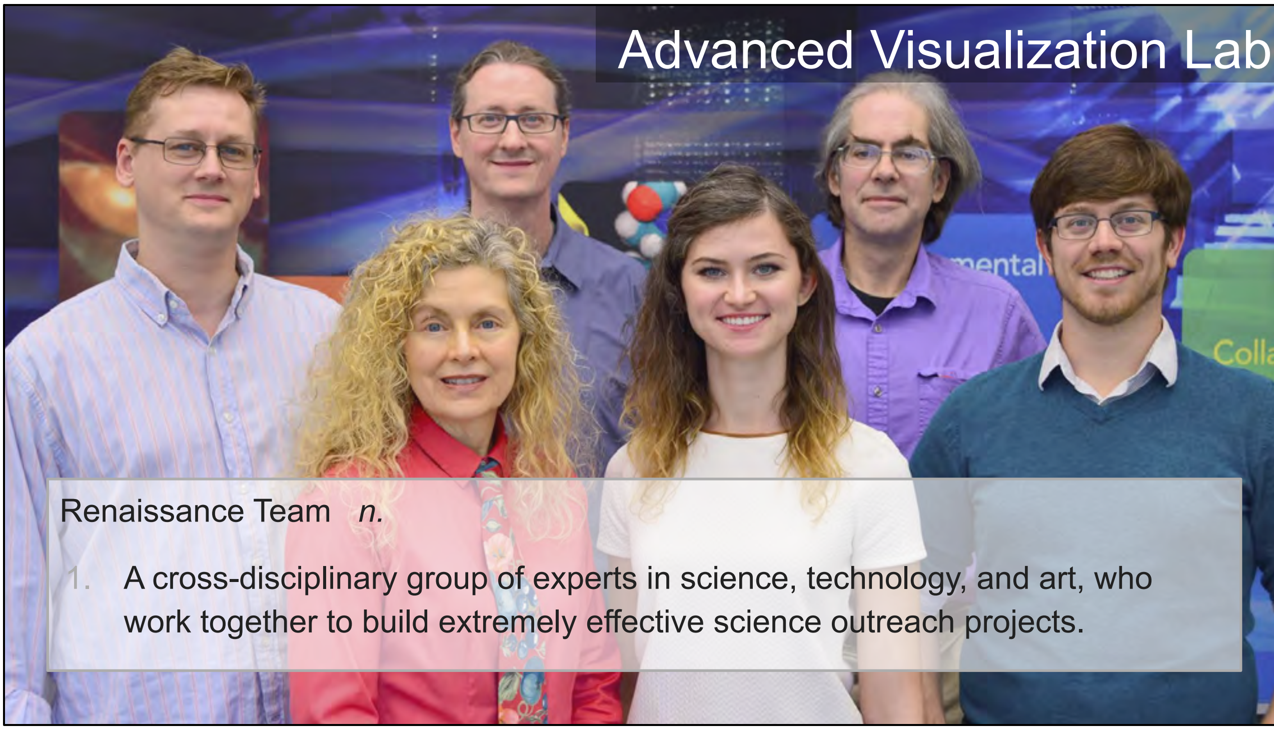
Greg Shirah
Scientific Visualization Studio
NASA Goddard Space Flight Center

Ryan Wyatt
Morrison Planetarium
California Academy of Sciences

SIGGRAPH '19 Courses, July 28 - August 01, 2019, Los Angeles, CA, USA
ACM 978-1-4503-6307-5/19/07.
10.1145/3305366.3328056

Table of Contents

Who We Are: Introduction



The Advanced Visualization Lab (AVL) is part of the the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. The AVL is led by Professor Donna Cox, who coined the term "Renaissance Team", with the belief that bringing together specialists of diverse backgrounds creates a team that is greater than the sum of its parts, and members of the AVL team reflect that in our interdisciplinarity. We specialize in creating high-quality cinematic scientific visualizations of supercomputer simulations for public outreach.

The AVL has contributed to dozens of television, digital fulldome, and IMAX science documentaries. This is a selection of a few recent films.

IMAX Films:
Hubble 3D (2010), narrated by Leonardo DiCaprio
A Beautiful Planet (2016), narrated by Jennifer Lawrence

Narrative Films:
The Tree of Life (2011), written and directed by Terrence Malick
Europa Report (2013), directed by Sebastián Cordero

Fulldome Shows:
Dynamic Earth (2012), narrated by Liam Neeson
Destination Solar System (2014), at the Chicago Adler Planetarium
Solar Superstorms (2015), narrated by Benedict Cumberbatch
Imagine the Moon (2019), at the Chicago Adler Planetarium
Birth of Planet Earth (2019), narrated by Richard Dormer

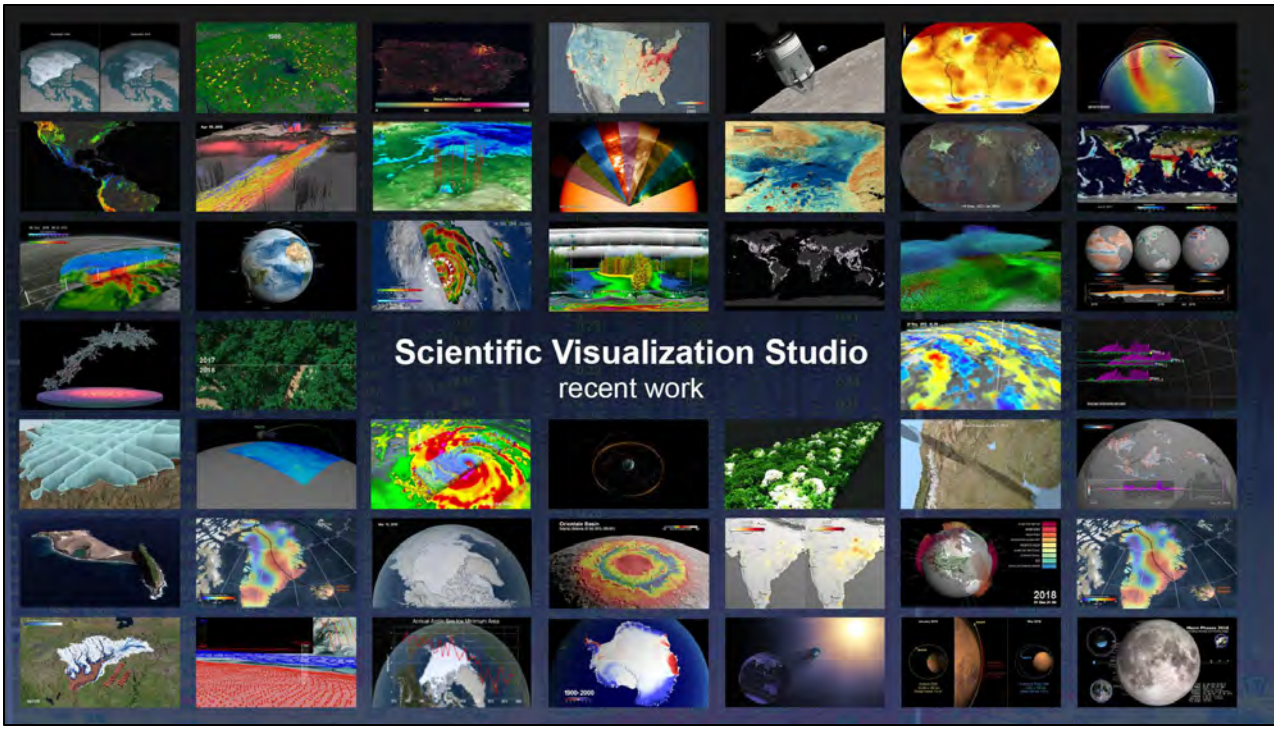
TV Documentaries:
Solar Superstorms: Journey to the Center of the Sun (2015)
SuperTornado: Anatomy of a MegaDisaster (2016)
Seeing the Beginning of Time (2017)
The Jupiter Enigma (2018)

Morrison Planetarium shows at the California Academy of Sciences are fueled by cutting-edge scientific data, resulting in stunning visualizations of the latest findings, discoveries, and theories about our Universe. Every star or galaxy a viewer encounters in the planetarium precisely mirrors a real-world counterpart, and when this virtual cosmos is projected onto Morrison's 75-foot-diameter screen, the dome itself seems to disappear, resulting in a uniquely immersive experience.

Ryan Wyatt has written and directed the Academy's six original, award-winning productions: *Fragile Planet* (2008), *Life: A Cosmic Story* (2010), *Earthquake: Evidence of a Restless Planet* (2012), *Habitat Earth* (2015), *Incoming!* (2016), and *Expedition Reef* (2018).

The Scientific Visualization Studio (SVS) works closely with scientists in the creation of data-driven visualizations, animations, and images in order to promote a greater understanding of Earth and Space Science research activities at NASA and within the academic research community supported by NASA. The SVS is an integral member of NASA's Science Storytelling team, along with conceptual animators, producers, science writers and social media experts aim to deliver science news and research findings to the general public. Visualized products are vetted by scientists and upon release become freely and openly available to the public through the SVS website: https://svs.gsfc.nasa.gov/

The Scientific Visualization Studio (SVS) is based at NASA's Goddard Space Flight Center and serves the entire agency.

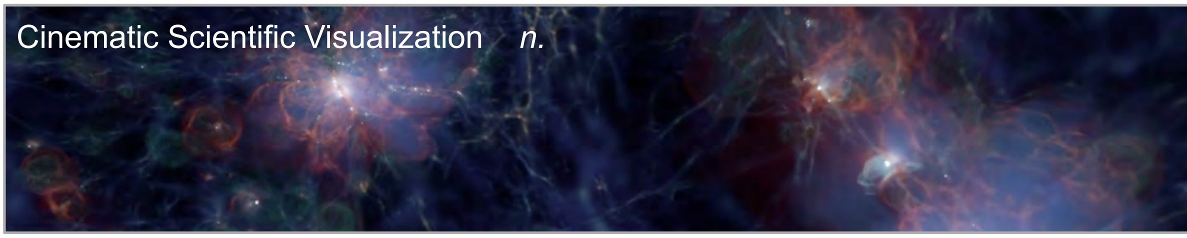


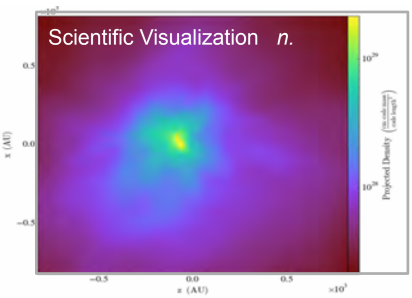
The Scientific Visualization Studio develops data-driven computer graphics (animations and still images) that cover all of NASA's science themes (Earth, Planetary, Heliophysics, Astrophysics), producing a wide range of products and different formats, including:

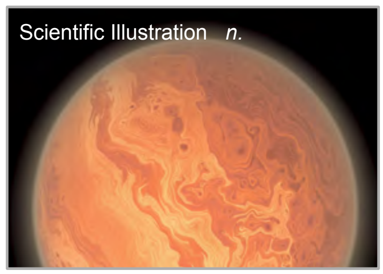
- HD
- UltraHD for tiled displays and hyperwall
- Dome shows
- Mobile
- 360 videos

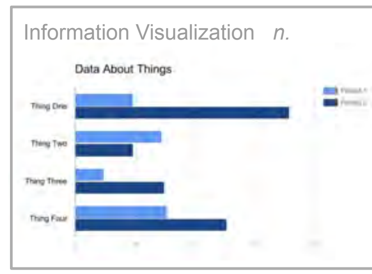
What is Visualization

Vocabulary







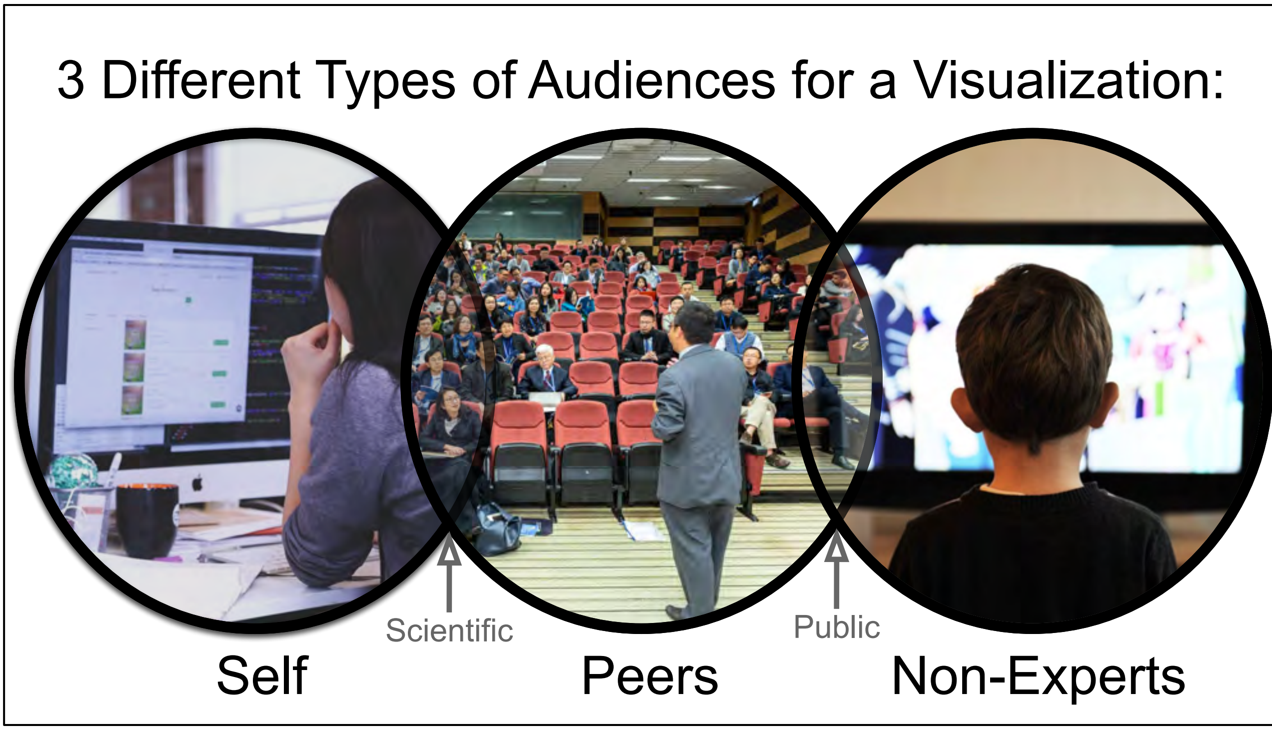


Visualization is a wide field, and in this course we are focusing specifically on "cinematic scientific visualization". Cinematic scientific visualization focuses on using Hollywood storytelling techniques to communicate spatial computational data with the public. This type of visualization is production-quality, data-driven imagery created with movie-making tools with good composition, camera direction, and artistic aesthetics suitable for distribution in immersive giant screen theaters.

This differs from traditional scientific visualization, which does not focus on cinematic quality but rather on data analysis functionality for communicating with experts.

The other primary category of visualization is "information visualization" which is used for relational or other non-spatial data, and is more frequently seen in 2-dimensional representations. Where scientific visualization deals with datasets which most often have an inherent structure or geometry, such as fires, hurricanes, topographic data, MRI data, information visualization deals with data that do not have spatial structure such as genetic information and stock market.

And it's worth noting that "visualization" is not "illustration", which is not data-driven. These are often images or animations created based on expert input but using predominantly artistic tools, with the goal to bring science, theory, design and concepts to life in a scientifically accurate and visually compelling way. It is an integral part of telling a story when data are not available. For example, NASA's Conceptual Image Laboratory creates animations and visual effects in collaboration with science teams to illustrate - what would a future NASA mission look like?



Scientific visualization can be done for different audiences, and for different purposes.
1. Scientists use visualization to validate, analyze, or error-check their own data. The purpose of such a visualization is to enhance data analysis capabilities and the resulting visualizations may not be shared with anyone other than the scientist themself.
2. A scientist may also create visualizations for publication in academic journals, or for presentations to their peers and other experts. These visualizations often assume some base level of knowledge, such as the context around the data, or any meanings assigned to the specific colorations used.
3. The third and final audience is the primary target of cinematic scientific visualizations: non-experts. The general public requires additional context and care to be taken to not misrepresent data, as the visualization creators cannot assume any background knowledge in the topic. These visualizations also benefit from being more eye-catching, and using advanced computer graphics techniques pioneered by Hollywood and not often employed for the other types of visualization.

The purpose of visualization for oneself can be thought of as validation; visualization for peers is visual communication; and visualization for non-experts is storytelling. These different audiences require vastly different treatments, experts, design, and consideration to be taken from the very beginning of the visualization process, through to the packaging and distribution of the created content. There is however overlap between these categories. Visualization for oneself and one's peers can both benefit from traditional scientific (non-cinematic) visualization. And visualization for peers as well as non-experts requires elements of communication.
The Advanced Visualization Lab and the Morrison Planetarium specialize in creating cinematic scientific visualizations for non-experts, while NASA's Scientific Visualization Studio creates visualizations that target both experts and non-experts.

“I think I’m more visual than symbolic. I try to get a feeling of what’s going on. Equations come later.”

- Claude Shannon, founder of information theory

The science community often likes to talk about how research is objective and discoveries are simply the outcome of robust methods. But this discounts the visceral and intuitive way subject matter experts understand their field. For example, Claude Shannon - the inventor of Information Theory - was frequently caught saying “Something LIKE THIS should be true…” and it would lead his research toward a real discovery.

Because “a picture is worth a thousand words”, cinematic scientific visualizations help us teach people the way science “feels”.

Cinematic visualization is a combination of many disciplines:

- Science
    - Domain science (e.g. astrophysics, biology, climate, geospatial…)
    - Mathematics
- Art
    - Art & Design
    - Filmmaking
- Technology
    - Data science
    - Computer science
        - Computer graphics
    - Computer hardware
- Humanities
    - Communication
    - Education
    - Psychology

Though we are not domain scientists, our skill lies in being able to *understand* the science and communicate to broad audiences in a visual way, using advanced computing techniques.

Vocabulary

Scientific simulation *n.*

1. A scientific simulation is a computational model that solves physical and statistical equations to simulate real-world phenomena. These simulations are often done on supercomputers.

Visual effects simulation *n.*

1. If it looks right, it is right.

Scientific simulations allow scientists to work in virtual laboratories where a physical experiment is either too costly, time-consuming, or impossible. A scientific simulation differs from the Hollywood definition of an “FX simulation”, which is calculated in graphics software to approximate physics at a flexible computational cost. In this course, when we use the word “simulation”, we are referring to a scientific simulation, which might take weeks or months to calculate on a powerful supercomputer.

Vocabulary

Scientific model *n.*

1. A mathematical description of a scientific phenomenon.

Visual effects model *n.*

1. A surface defined by geometric coordinates.

The word “model” is one that is used both by the scientific community as well as the VFX community, though the definitions are quite different in each of those contexts. Within the context of cinematic scientific visualization, the meaning of the word “model” can fall somewhere in the middle, along a continuum - for example, a model of an asteroid, or photogrammetry of a coral reef, may fit both of these definitions.

Vocabulary

Frame *n.*

1. One image out of a sequence of images that make up a video.

Timestep (Also: Datastep) *n.*

1. One data file out of a sequence of data files in a time-evolving dataset.

Finally, one interesting thing to note is that, while timing imagery and data in the VFX world is often thought of in terms of frames and FPS, there is a parallel timeline that visualization designers also have to work with - the timing of the data. In an ideal world, the number of frames per second and the number of data timesteps per second would be the same, but this is rarely the case. Data is expensive to create or collect, so a visualization may consist of a couple hundred timesteps, but a couple thousand frames. Interpolation, dissolving, or other techniques often need to be employed for a smooth viewing experience in such a case.

Data Visualization **Critique**

Data is subjective.

Data can be unintentionally biased.

Data can be maliciously incorrect.

People tend to think of “data” as true, unbiased fact, but that is not always the case.

First of all, data can be wrong or fabricated. The research questions being asked can be leading and biased, leading to incorrect results. You can’t even necessarily trust datasets of video, photos, or audio because they are becoming more and more easy to fake.

But even data that is true can be biased. It’s important to ask - who collected and sponsored the research, and what was the method? Things like low sample size or cherry-picked results can be an indication of biased data. The data may be a one-off result. If it is not reproduced or reproducible by other scientists, this may be a bad sign.

Visualization is subjective.

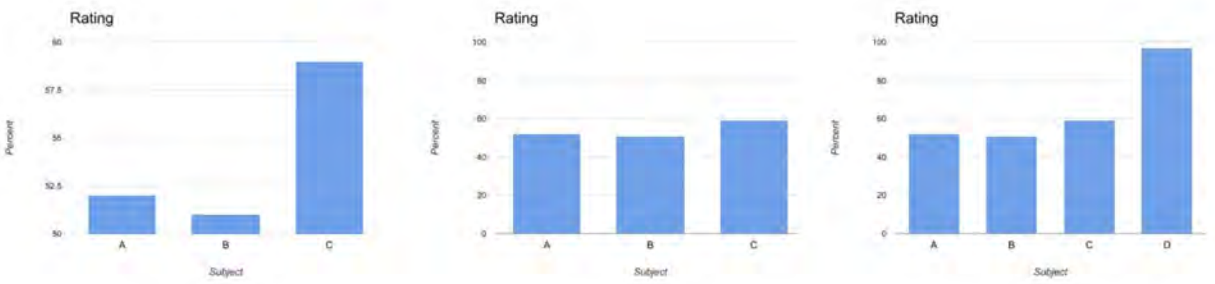
Visualization can be unintentionally biased.

Visualization can be maliciously incorrect.

People believe visualizations more than words. They are not necessarily right to do so, but this gives you, as a visualization designer, power that you have to yield responsibly. It is incredibly easy to lie about a dataset that is not set in context. Even visualizations based on true, unbiased data, can still convey a biased or incorrect message, whether intentionally or unintentionally.

The fact that both data and visualization are subjective but are not perceived that way gives you, a visualization designer, the responsibility to work with reputable data and to create accurate and vetted visualizations.

Graphs showing the same data with different y-axis labeling



C is much better than A or B

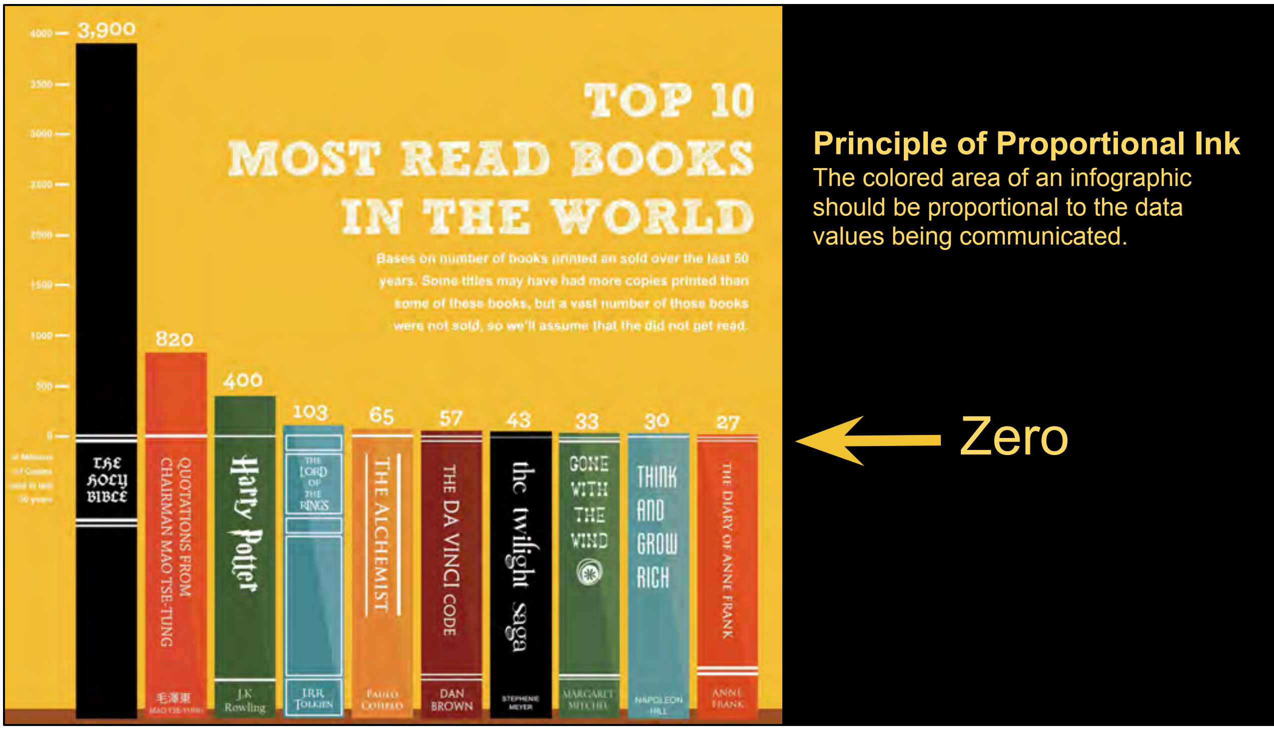
A, B, and C are about equal.
C is a little better than A or B.

D is much better than A, B, or C.

For a simple abstract example of how visualization can be misleading, let's look at these bar charts which illustrate how different conclusions can be drawn from context.

In the first graph, we can assume that C is much bigger than A and B. We may not notice that the y-axis doesn't start at 0, and that there's a very narrow range.
If we rescale the y-axis to go from 0-100%, A, B, and C are about equal.
But, perhaps we were just looking at a subset of the full dataset. The third graph introduces D, which has a much higher value than A, B, and C.

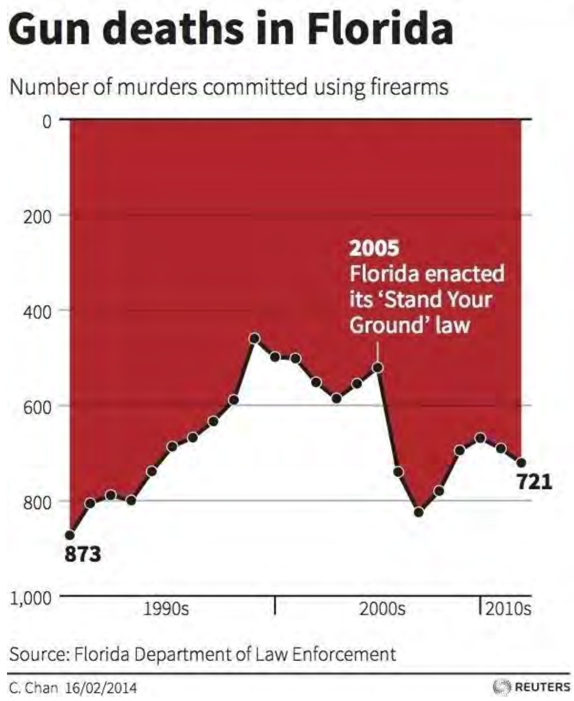
All of these graphs show the same data, in the same bar graph format, but they portray different ideas. Visualizations can be intentionally misleading, or accidentally misinterpreted. These graphs cannot be correctly interpreted if their context isn't properly encoded into the axes.



There is a website about communicating with data that you might enjoy called CallingBullshit.org.

One idea they discuss is called the "Priniciple of Proportional Ink" which describes why some graphs can be misleading. It suggests that the colored area of an infographic should be proportional to the data values being communicated. As you can see in the book example, there's a lot of extra ink below zero on the Y-axis.

Although this is a 2D information visualization example, and not 3D scientific visualization, it is important to learn from the field of information visualization because it has several well-established principles like this one. Scientific visualization is a newer, broader field that does not yet have it's own established rules, but some 2D principles can be applied to a 3D environment as well. The Principle of Proportional Ink, if applied to scientific visualization, suggests that an unimportant feature in the dataset should not take up a majority of the frame. This is a reasonable 2D principle to follow even in three dimensions.

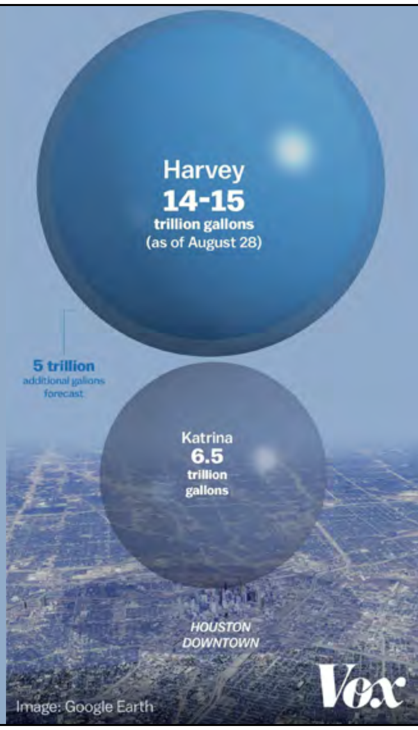


Intentionally Malicious or Just Poor Design?

(does it matter if the effect is the same?)

The y-axis is upside down, making an increase in gun deaths appear like a decrease.

The previous examples have been misleading but overall harmless. But there are many examples of images that have had a real world effect, misrepresenting data for political reasons. That's the case with this graph that was widely distributed around the time of a high-profile shooting that happened in Florida. You can see that the y-axis is upside down to make gun deaths seem to decrease instead of increase.



Humans are inherently bad at comparing certain perceptual attributes, like volume.

The sphere on top is supposed to represent 3x as much rainfall.

The human mind can misinterpret multi-dimensional objects. Although the giant water droplet on top is meant to represent more than three times as much water as the the water droplet on bottom, it intuitively feels like the bigger droplet is only bigger by a little bit. There is an indirect relationship between 3D quantities flattened on a 2D screen and our perception of them. Humans are generally bad at judging scale without many different clues interacting with each other and it’s important to reinforce and reiterate data where possible.

These graphs represent sort of a curiosity about how humans might perceive images differently than expected.

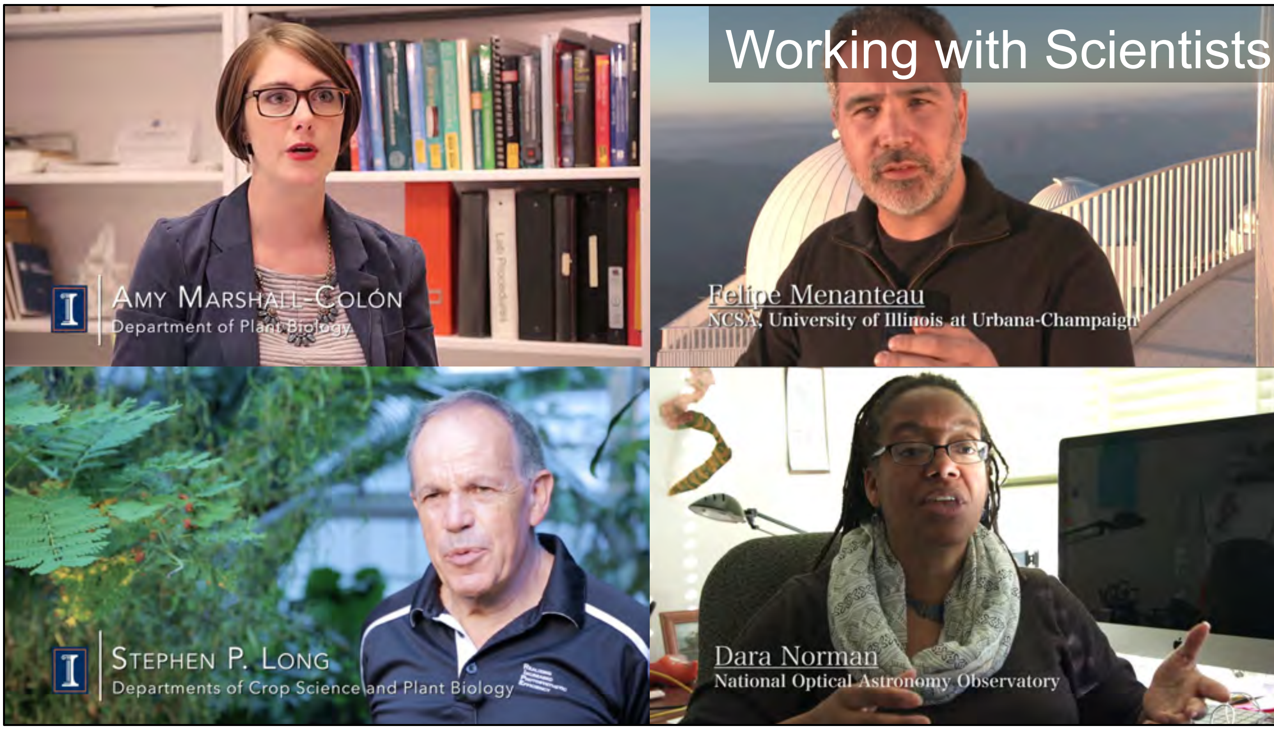
In this example, the data points are not at regular intervals along the horizontal axis, distorting the rate of change in the price of fuel. Many Americans votes are influenced by the price of fuel, so this sort of miscommunication can have a very real effect if distributed widely.

It is possible that the people who originally created these images made an honest mistake, but when the news room shared them with millions of people, the effects were the same. Especially in the age of social media, images are powerful and they can live forever. This is why it’s especially important for visualization designers who claim to be communicating with facts are careful that their unedited work does not reach the public prematurely.

Recent research by our colleagues at the University of Illinois [1,2] suggests that even if your data is accurate and unbiased, and your resulting visualization is accurate and unbiased, the title or narration given to the visualization can still influence the perceived message of the visualization, with different viewers deriving opposing messages from the same imagery. Subjects were shown a visualization with a title that contradicted what the imagery was showing. When asked to recall the visualization at a later point in time, the majority of people recalled that the data showed what the title had said, rather than what was actually happening.

1) Ha-Kyung Kong, Zhicheng Liu & Karrie Karahalios. "Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment," CHI 2019
2) Ha-Kyung Kong, Zhicheng Liu & Karrie Karahalios. "Frames and Slants in Titles of Visualizations on Controversial Topics," CHI 2018

How to **Validate** a Visualization



It is extremely important to have frequent contact and build a strong relationship with scientific domain experts so they can identify inaccuracies long before they make it into the visualization. Working with scientists is important because they are able to:

- Explain the purpose of the simulation and its context in the larger field
- Validate that the visualization conveys the science accurately (or as accurately as possible)
- Validate and edit the documentary script
- Re-run or fill gaps in their simulation

While many design decisions can be influenced by formal research, in most cases there is no formal research to justify making one choice versus another. In these situations, visualization designers must rely on informed intuition. This is why perhaps the most important way in which we evaluate our work is direct interaction with audiences through daily demonstrations in our lab, and through public presentations and Q&A sessions like this. We need to acknowledge that intuition isn't as structured as formal research, and so to avoid falling into the trap of creating our own "alternative facts", we must collect data on these interactions and document and analyze our discoveries periodically

For the following, please choose how much you agree with the statement:

	strongly disagree	disagree	neutral	agree	strongly agree
[8] I learned something new today.	1	2	3	4	5
[9] I now better understand some science concept(s).	1	2	3	4	5
[10] I'm going to try to learn more about what I saw today.	1	2	3	4	5
[11] I have more questions about what I saw.	1	2	3	4	5
[12] I felt confused by a visualization I saw.	1	2	3	4	5
[13] I would like to see more visualizations.	1	2	3	4	5
[14] I'd like to learn how to make visualizations.	1	2	3	4	5

[15] What visualization impressed you the most, and why?

[16] What is your race/ethnicity?

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Other (please specify) _______________

[17] Age:

[18] Gender: Female Male Other

[19] Occupation or Major:

We've been working with researchers in the fields of education and psychology to understand the value of audience interaction. What we've learned is that while our intuition as designers is improved by direct contact with the audience, we must collect data on these interactions and document and analyze our discoveries periodically. We are just starting to collect audience surveys in our lab demonstrations that our psychology research partner plans to include in her research.

We are taking demographic information to make sure that people of all ethnicities, ages, and genders understand our visualizations, and to ensure that we are not introducing any unconscious bias into our work.

So - how do we know our visualizations are having the effect we want them to have? Understanding the far-reaching impact of these visuals is notoriously difficult to quantify.

Part of the trouble of understanding the value of cinematic scientific visualization comes from the fact that it is not meant to work like a textbook. The National Science Foundation, an agency of the United States federal government, defines the space of cinematic scientific visualization as “informal science”, which is measured indirectly. The expectation is not for people to go home and remember facts that they can be tested on, but rather, to have an enjoyable experience, to feel intellectually empowered, and to be motivated to learn more.

Some of the more important educational outcomes are:

- How many questions have audience members generated by the end of the documentary?
- Are audience members more capable of making scientific explanations to others?
- Do audience members know how to further interrogate the material?

Audience Testing

Example Question

Where do these images come from?

A. This is real video, observed through a telescope
B. This is a visualization, based on a scientific computer simulation
C. This is an artistic drawing of what we think it looks like

The most straightforward way to quantify understanding is to work with educational researchers to interrogate audiences about how they are interpreting the visualizations. We have held advance screenings of our films to get feedback while there is still time to change the design. One of the communication challenges we focus on with these audiences is trying to convey that the images we make are scientific visualizations - they have not been captured in the lens of a camera, nor have they been captured from an artist's imagination.

Audience Testing

Example Question Answers

44% thought it was both real pictures and data visualizations

33% thought it was a combination of all three

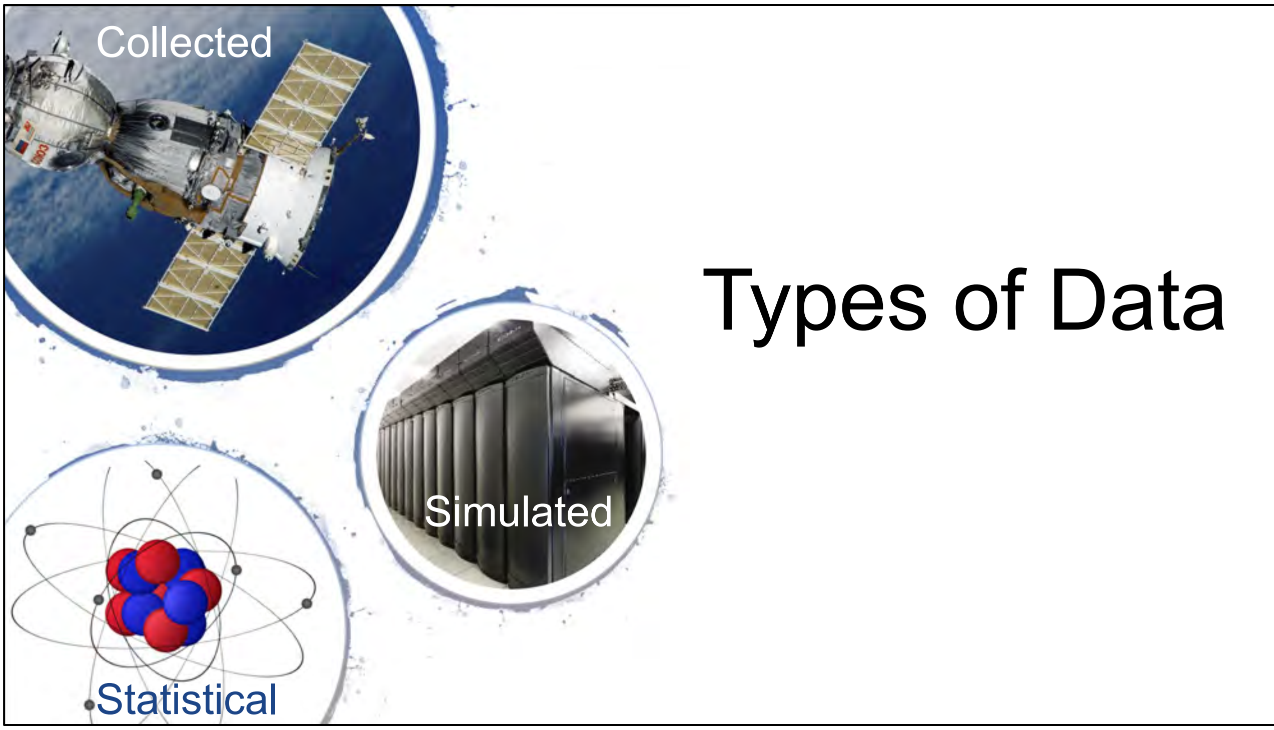
23% thought it was solely visualized data

All are correct or partially-correct!

In these survey responses, you can see that the 44% of the audience chose what we consider “the most correct” answer, and no one chose “the most incorrect” answer that everything was purely artistic.

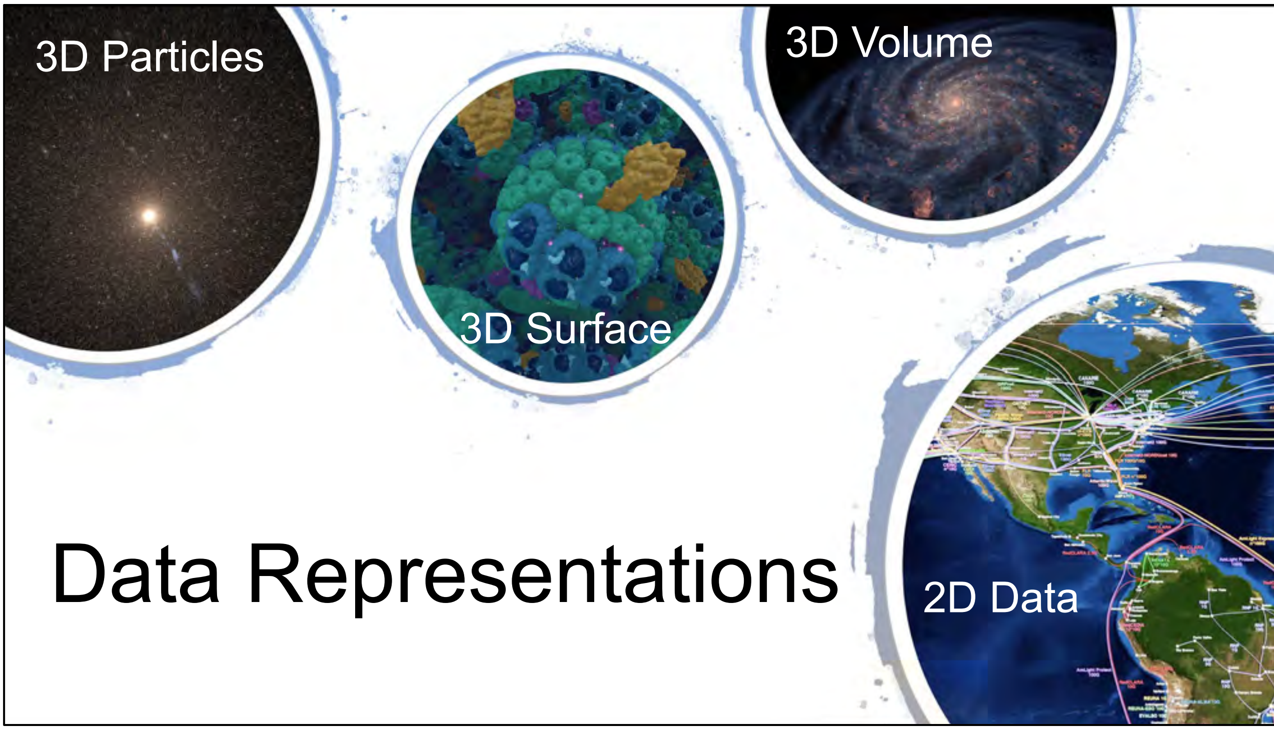
Working with Data

"Data" is a word that covers a lot ground; everyone here works with data of some kind. What do we mean by scientific data and what are some challenges of working with it?



There are three types of scientific data:

1. Data that is collected from the real world with an instrument like a satellite, telescope, microscope, 3D scanner, LIDAR, photogrammetry, etc. This type of data is also called observational data. Collected data is often thought of as the most “real” type of data, though the comparison is not so black-and-white. Working with observational data requires it to be processed, cleaned, transformed, translated, and it may even be run through a scientific model.
2. Data that is simulated on a computer based on a scientific model. These simulations encode laws of physics and are often run on extremely powerful supercomputers for extended periods of time, and are validated by comparing against experimental results and theorized results by theoretical scientists such as Albert Einstein. Though computational science is a more recent invention, it has become known as the “third pillar” of science, alongside experimental science and theoretical science.
3. Data that is represented mathematically with probabilities by a statistical model. Examples of this type of data include positions or densities of objects that are not possible to measure - e.g. the exact positions of the stars in the Milky Way galaxy; the positions of individual electrons in an atomic simulation; the density of mass in the core of the sun.



Regardless of how spatial data is collected or created, it can be represented in a number of different ways. There are many valid representations that can be chosen for a particular dataset. Some common choices include:

1. 3D particles that can be used to represent small, discrete items such as distant stars.
2. 3D surfaces that can be used to represent objects with solid surfaces, such as mountainous terrain.
3. 3D volumes that can be used to represent continuous fields and are often used to represent gasses.
4. 2D data thta can be thought of as images, although the pixel values do not necessarily have to map to RGB colors. A “pixel” can hold any data value, such as height, which can be used to seed a 3D representation.

REAL DATA

Collision that Formed the Moon

SIMULATION
STATS

DATA SIZE	30GB raw + 241GB derived
TIME STEPS	492 interpolated to 3676
RESOLUTION	1.2 million particles
SPATIAL SCALE	1 Angstrom (1e-10 m)
TIME SCALE	1 femtosecond (1e-15 s)
SCIENTIST	Robin Canup
INSTITUTION	Southwest Research Institute

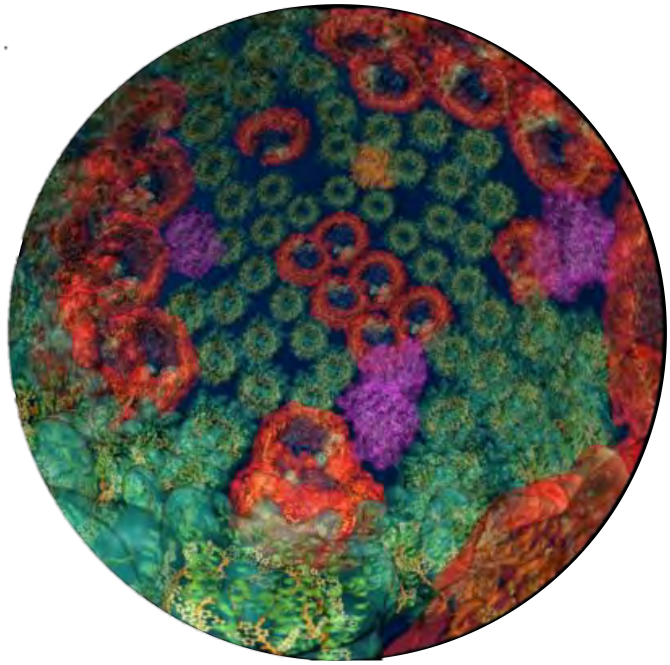
This dataset appears to start out as surface geometry and ends up looking volumetric - but in fact the underlying data is represented as particles. There are 1,172,150 particles, and each particle contains a multitude of variables such as density, energy, mass, temperature, pressure, and velocity. The data is represented in a hydrodynamical format called SPH. These discrete particles actually represent continuous bits of matter, where each particle has a “smoothing length” that defines how a function fills in the gaps between the particles. This is a relatively small dataset in terms of scientific data, coming it at about 0.25TB.

REAL DATA

Photosynthetic Organelle

SIMULATION STATS

DATA SIZE	60 TB
TIME STEPS	~100,000
RESOLUTION	~100,000,000 atoms
SPATIAL SCALE	1 Angstrom (1e-10 m)
TIME SCALE	1 femtosecond (1e-15 s)
SCIENTIST	Klaus Schulten
INSTITUTION	University of Illinois at U-C
SUPERCOMPUTER	Blue Waters and Titan



This visualization is based on a 60TB simulation of a photosynthetic organelle called a chromatophore - a predecessor to the chloroplasts we find today in plant cells. This simulation was developed from a full atom-by-atom simulation on some of the biggest supercomputers in the world. This model of how photosynthesis first began was the life work of the renowned late scientist from the University of Illinois, Klaus Schulten. This visualization is featured in this year’s Electronic Theater at SIGGRAPH and is from our most recent dome show, titled “Birth of Planet Earth”, which describes how the early universe led to the beginning of life on Earth. Rather than showing the individual atoms of the simulation, we are using a 3D surface to represent a statistical boundary that makes up a more recognizable structure.

If you would like to do data visualization but are not a data creator yourself, it can be intimidating to try to find a good, scientifically-validated, unbiased, complete, usable dataset to visualize.

Thankfully there are many datasets online that are free and easy to find. Government agencies like the NASA, the ESA, and others provide various types of free data for public use and are credible sources of information. Distributed Active Archive Centers, or DAACs, are a good place to start. Many cities like Chicago, London, and Singapore provide data through “Open Data Portals” that relates to local civic concerns. Certain university groups and individual researchers freely provide their data as well. You may be able to find data that is written about in academic papers - even if the data isn’t publicly available, you may be able to reach out to the scientist directly to request a copy. More and more, research journals are demanding that researchers make their data available as a primary research output.

If you use somebody else’s data for your visualization, it is important to credit the computational scientists, collaborators, and funding agencies, so make sure to get this information as well as written permission. This isn’t just about avoiding potential legal issues - remember that a scientist’s research data is their career, and may be their life’s work. We need to be respectful of that, and ensure that the scientists always retain the rights to their own data. There may be different rules for using and crediting data depending on whether it is being used for education or for profit, but you should always credit the scientists and software, regardless of whether or not this is explicitly required.

Challenge: Data is Messy

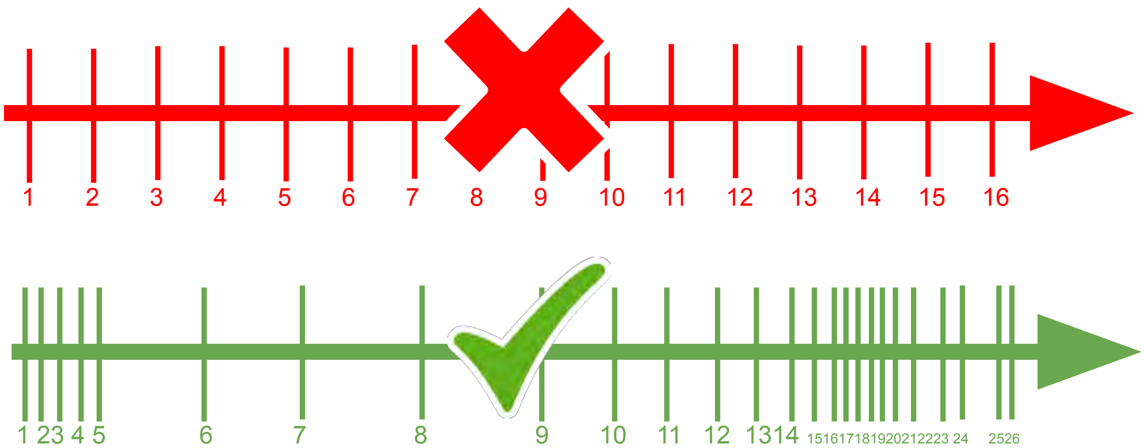
Any dataset that you find will never be trivial to visualize, for a variety of reasons. Data is messy and requires cleaning and processing - especially observational data. Sensors can fail or be spotty, with dropouts or partial coverage. In this example, the sensor is overwhelmed by the brightness of the solar flare, making perhaps the most interesting part of the data completely unintelligible.

Simulated data is not free from artifacts, either. These can be a result of errors in the scientific model or the data processing code, or they may be missing areas that are not being studied. For example, a scientist simulating the effect that a black hole has on its surrounding galaxy may model the gas and dust of the galaxy, but not the black hole itself - it may be represented as a dense point or sphere without any features of its own. This makes sense from the scientist's perspective, since they are studying what happens to the galaxy and the details of the black hole do not matter. But creating a visualization out of such a dataset for a non-expert audience would require creating some representation of a black hole that is missing from the data.

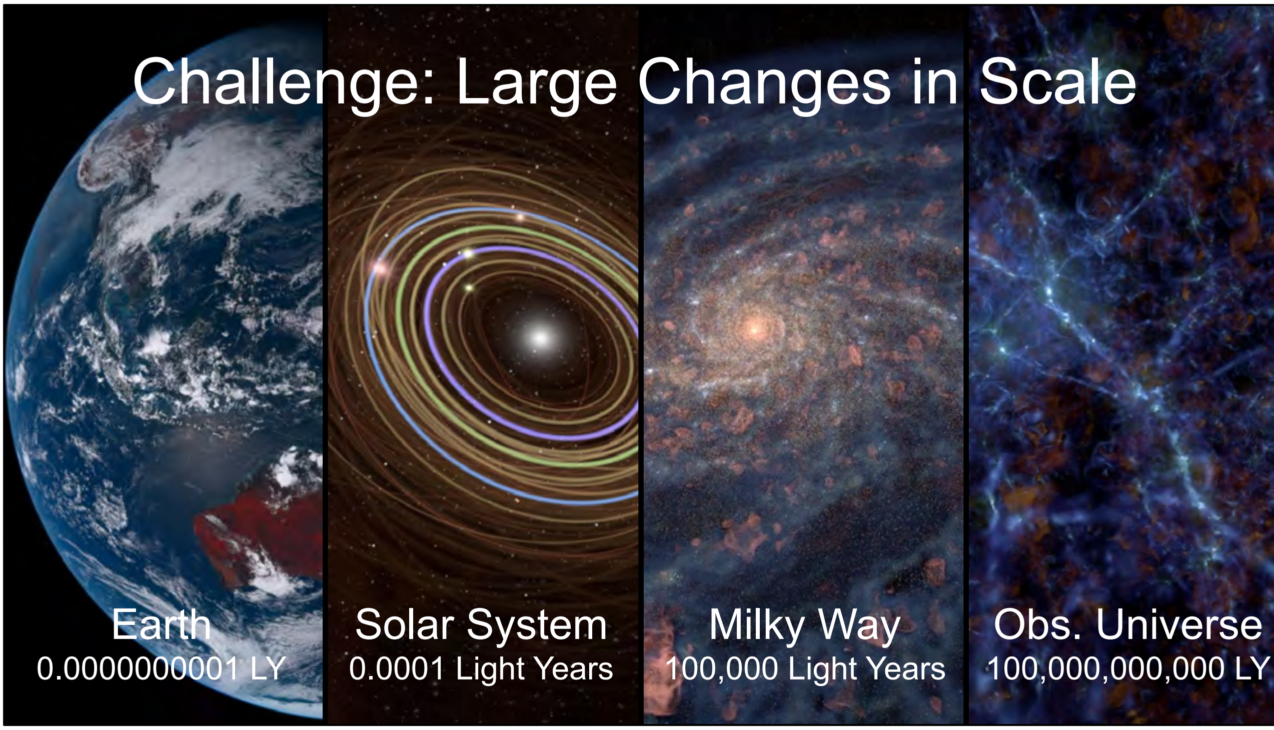
Another challenge in dealing with data are the many different coordinate systems that the data may be stored in - many of which may not be natively understood by the visualization software, or that may not be understandable to a non-expert audience. This image shows a selection of various map projections of the Earth. Each of these is valid, and some are especially useful for some single specific purpose. The same data can be represented in many different ways, and will often require translating between formats in order to get it into an appropriate format for the visualization.

Image credit: https://en.wikipedia.org/wiki/List_of_map_projections

Challenge: Non-Linear Time



Simulations can be very expensive to run, as they can take months of time on a supercomputer and may be terabytes or petabytes in size. Observational data is becoming extremely large as well - the Large Synoptic Survey Telescope, which will be completed in 2020, will produce 20TB of data per day! In such cases, it may not be practical for scientists to capture or create data at regular intervals - more data may be produced when interesting events are occurring, and less may be produced during uneventful periods of time. This results in datasets where time may not move linearly, which cannot produce smooth-looking visualizations without additional processing.



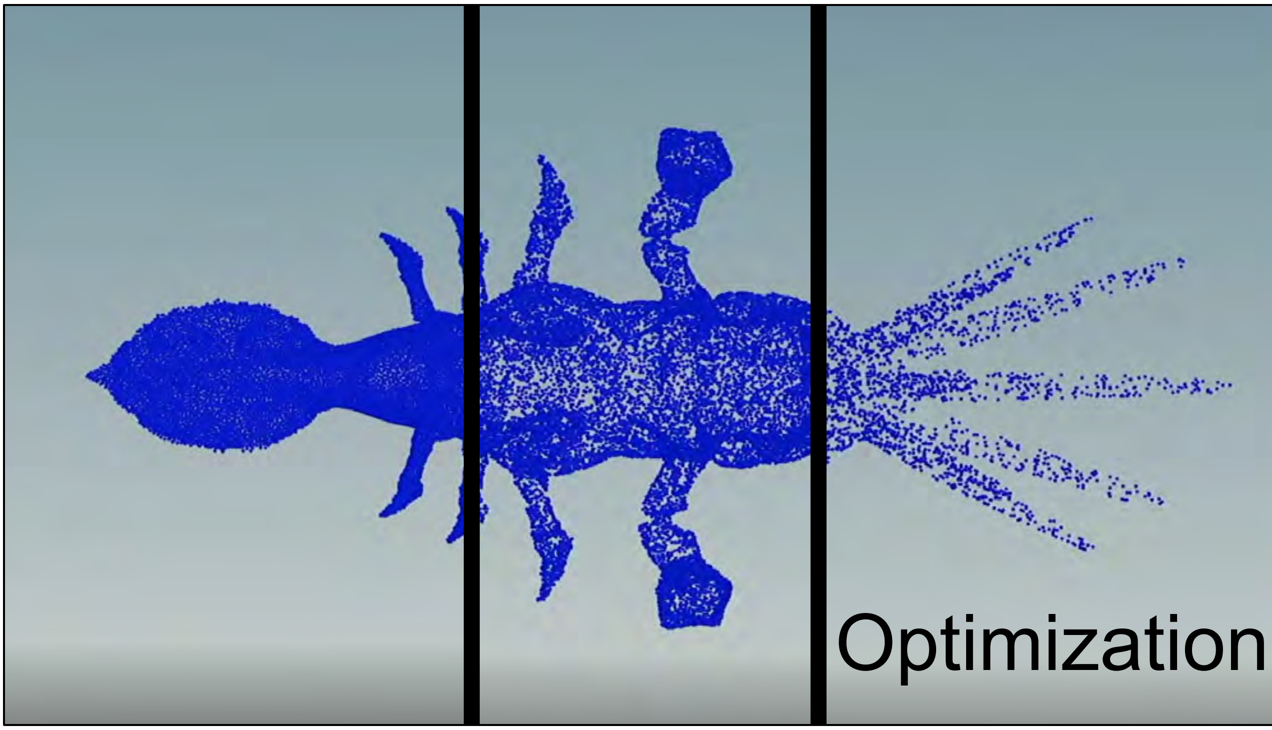
When dealing with data from physical sciences - astrophysics, molecular biology, etc - objects are often interacting at vastly different scales. This often leads to floating-point errors and creative methods need to be taken to be able to show both the large-scale and small-scale objects simultaneously.

Though a single dataset is not likely to have both resolution on the scale of the Earth as well as the observable universe, sometimes it is beneficial to link multiple such datasets together to tell a compelling story. The AVL recently completed such a project, where the camera path takes the audience on a seamless journey, starting with the Earth and ending with the large-scale structure of the universe.

More Data Challenges

- Different standards across scientific domains
- Complex data formats
- Software incompatibility
- Extremely large datasets
  - Many attributes per voxel/pixel/vertex/particle
- May need to create derivative data
- Data processing
  - Cleaning, translating, transforming, registering...

There are many other challenges when dealing with data, and the exact tasks that need to be taken to make a dataset ready for visualization will vary from dataset to dataset.



A common problem when working at a supercomputing center is acquiring data so large that it is impossible to work with within our standard pipeline. One solution to this problem is to create a proxy dataset, which is representative of the full dataset but only a fraction of its size. This proxy data is used for interactive viewing and designing camera choreography, and it is swapped out for the full dataset at render time. Particle data can easily be downsampled through subsampling, where some percentage of data is discarded evenly throughout the data. Thresholding can also be used to discard data that does not contribute much to the final image, e.g. extremely low-density data. In the past few years, AVL has been making more and more use of the Blue Waters supercomputer for data processing and rendering of these extremely large datasets.

Tools for Visualization

Software

VFX Software

Houdini
MAYA
blender
NUKE
Ae
RENDERMAN

Scientific Software

SPICE
Matlab
IDL
ArcGIS
STK
VMD
yt

Custom Software

- Partiview
- Virtual Director
- Ytini
- Plugins
- OSL shaders
- MEL scripts
- IDL code

Visual effects software and scientific software can both be used to create cinematic scientific visualizations - each category of tools has its pros and cons. We use a variety of both custom-built and off-the-shelf software. This is a non-comprehensive list of all the tools out there, just a selection of some of the tools that our teams use. There are many other options that you can choose to use that may work in similar ways.

Comparison of Visual Effects vs Scientific Software

	Houdini	Maya	Blender	yt	ParaView	VisIt	Vapor	IDL	Glue	AstroPy[a]
Data Analysis				X	X	X	X	X	X	X
Modular Coding Paradigm[b]	X			X	X	X	X	X	X	X
Supports Multiple Scientific Data Formats				X	X	X		X	X	X
Open Source			X	X	X	X	X		X	X
Graphical User Interface	X	X	X		X	X	X		X	
Renders Volumes	X	X	X	X	X	X	X	X	X	
Renders Surfaces	X	X	X		X	X	X	X		
Renders Multiple Geometry Types in Scene	X	X	X		X	X	X		X	
Renders Multiple Datasets in Scene	X	X	X		X	X	X		X	
Visual Transfer Function Editor	X	X	X		X	X	X	X		
Complex Lighting[c]	X	X	X							
Complex Shading, Material Effects	X	X	X							
Interactive Camera Controls[d]	X	X	X							
Interactive Design of Elements/Effects	X	X	X							

Borkiewicz, K., Naiman, J. P., & Lai, H. (2019). Cinematic Visualization of Multiresolution Data: Ytini for Adaptive Mesh Refinement in Houdini. *The Astronomical Journal*.

Visual effects software and scientific software have complementary strengths for creating cinematic scientific visualizations, but neither category of tools is ideal for doing the job alone.
Visual effects software excels at creating production-quality visuals, but can’t read scientific data formats.

- Pros:
    - Excellent animation controls
    - Excellent camera controls
    - Robust due to large user base and large commercial developer base
    - Excellent output quality (e.g., high resolution, motion-blur, etc.)
    - Render farm ready for fast turnaround
    - General purpose - visualizers can learn primary tool and apply to most/all projects
- Cons:
    - Can’t read scientific data sets
    - Scale and precision problems
    - Doesn’t know about date/time

Scientific software excels at quickly and efficiently reading and rendering complex scientific data formats, but its primary purpose is visualization for the purpose of analysis, not communication or storytelling, so it is difficult to control the aesthetics and design of the scene and rendered images.

- Pros:
    - Great at reading specific formats
    - Great for niche processing purposes but tend to be narrowly focused
    - Okay -to- great at rendering images
    - Often real-time
- Cons:
    - Poor camera controls
    - Poor animation controls
    - Not general purpose - need to use different software for different types of data
    - Sometimes buggy due to small user base and small developer base

Find the right ______ for your ______.

tool data

Depending on what you are the needs of your project are, it may be difficult to work with a particular tool, or a particular dataset. It takes varying amounts of and effort to develop features or work around roadblocks for software, so you may need to weigh your options.

You may need to choose the right data for your tool if:

- There are multiple similar datasets that could work for your needs.
- You are proficient in the use of a particular tool and are not confident you could produce work of similar quality in a new tool.

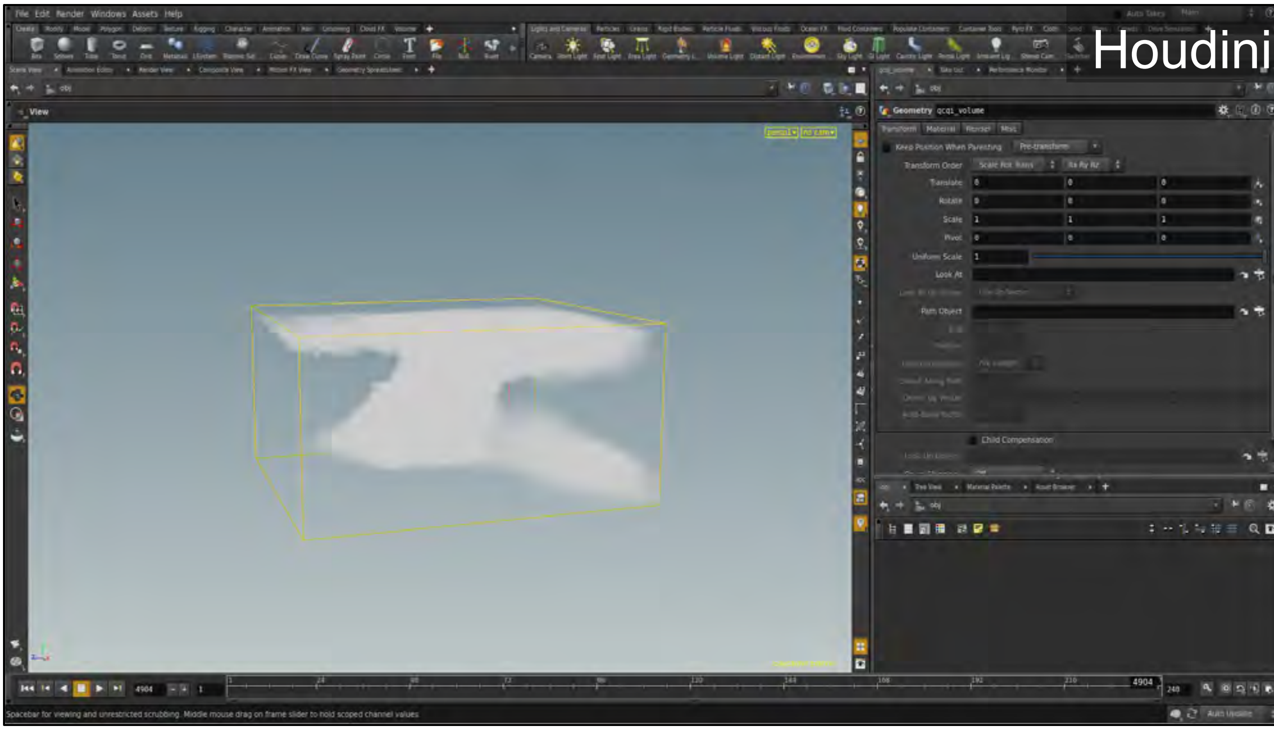
You may need to choose the right tool for your data if:

- This particular dataset is only one, or the best one, for your needs.
- The tool(s) you typically work with are unable to read this particular dataset or to be extended to read it.
- The data cannot be moved from its native environment so you must find a tool that works on that operating system or hardware.

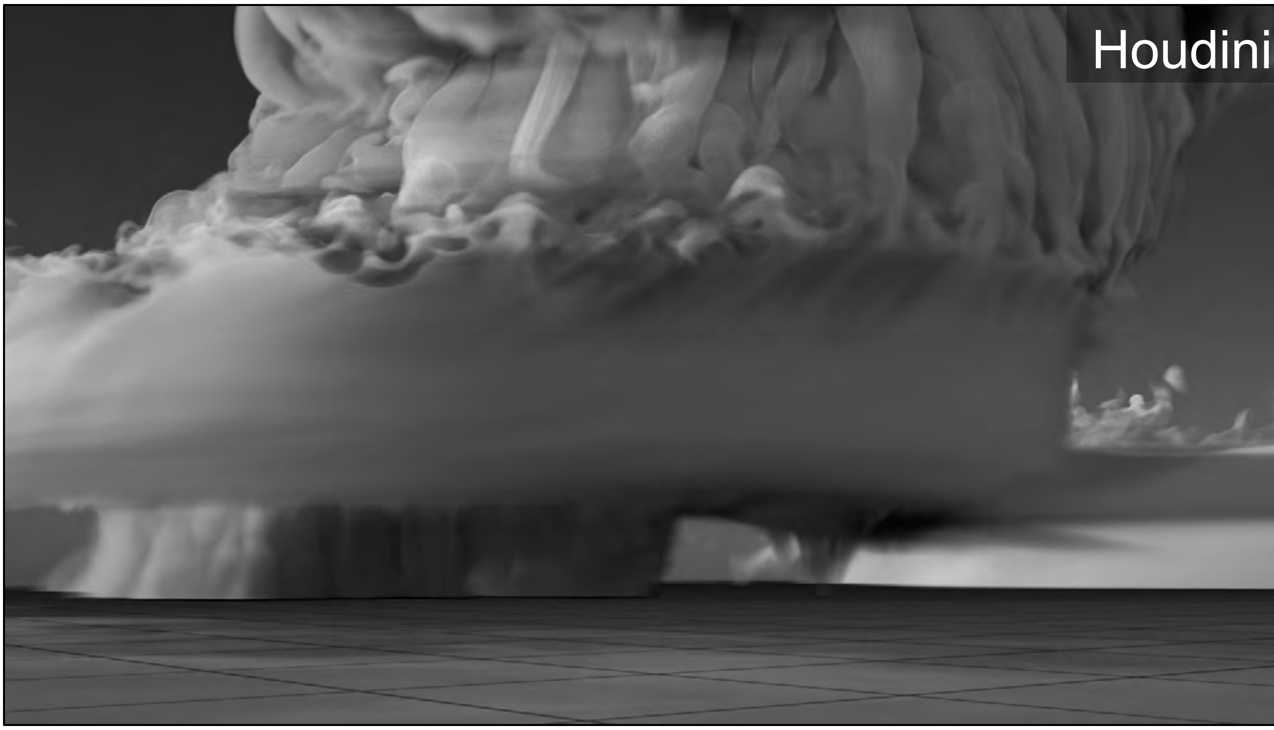
Tools for **Visual Effects**

Case Study: **Houdini**™

Let's take a closer look at an example of creating a cinematic scientific visualization in a visual effects tool, and later, a scientific tool. Houdini is the primary VFX tool that the AVL has been using in recent years.



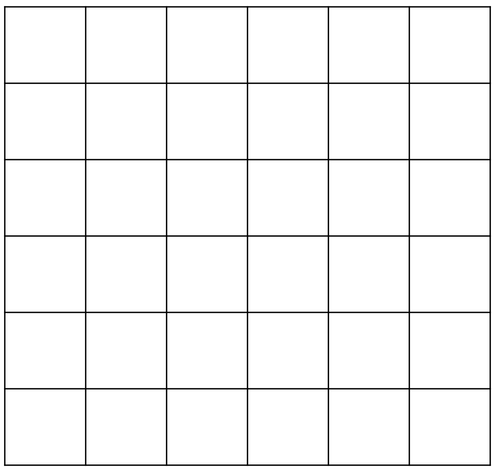
Houdini is not purpose-built for scientific visualization, but we continuously adapt it for our needs. AVL programmer Kalina Borkiewicz has written custom plugins and scripts to allow Houdini to work with scientific data. This includes tools for data loading and resampling, as well as artistic tools such as smooth falloff, isosurfaces, and spatial and temporal interpolation. The Houdini+Mantra combination works well in rendering multiple embedded high-resolution volumes efficiently and beautifully. This is a tornado simulation, represented as a stretched grid in NetCDF format,imported into Houdini with a custom C++ plugin.



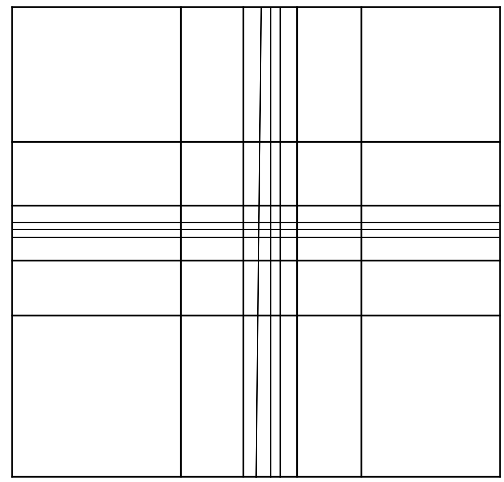
Here is a final rendering of the same tornado data, as a VDB sparse volume, done with Houdini and Mantra.

A Problem with SciVis in VFX Software

Houdini expects data to look like this:



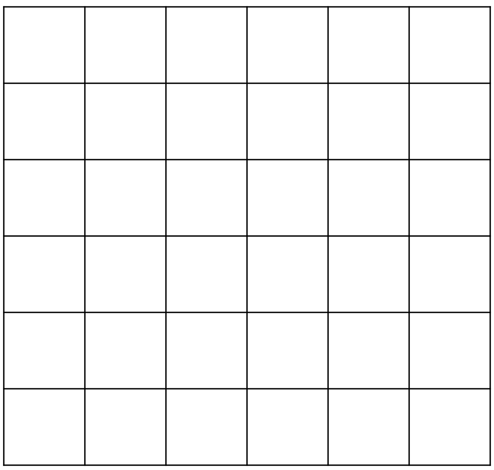
But some scientific data looks like this:



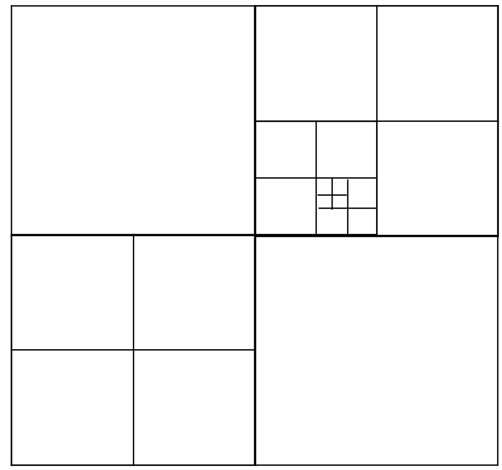
Houdini expects volumetric data to be in a uniform, Cartesian grid. However, this often not an efficient way to store scientific data. The tornado data from the previous slides came in a stretched grid, like the grid shown here on the right. This allowed for there to be more detail around the actual tornado, and less resolution around the less-interesting clouds.

A Problem with SciVis in VFX Software

Houdini expects data to look like this:



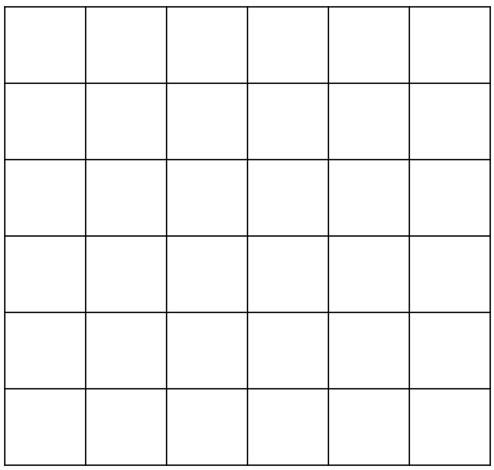
But some scientific data looks like this:



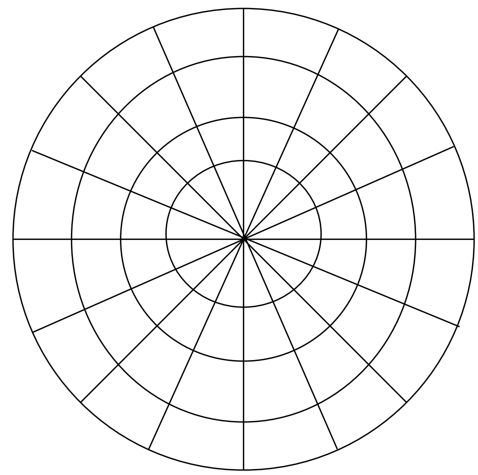
Another data format that has more resolution only where it’s needed is called Adaptive Mesh Refinement (AMR), which is made of nested, multi-resolution grids.

A Problem with SciVis in VFX Software

Houdini expects data to look like this:

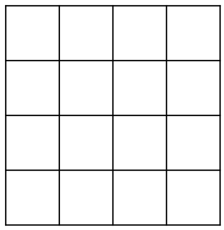


But some scientific data looks like this:

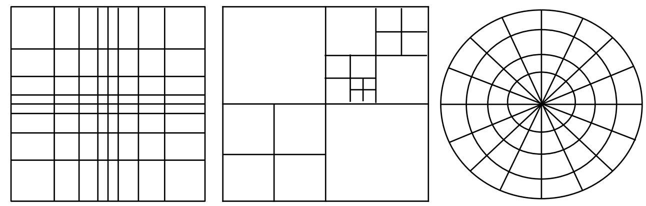


Data can also come in spherical or cylindrical grids, or many other complex data formats that don’t easily convert into a Houdini-friendly grid.

Houdini expects data to look like this:



But some scientific data looks like this:



Possible solutions:

A. Sample the data… How often?
   - Use the finest data resolution? ⟶ Wastes resources.
   - Use a convenient “middle” resolution? ⟶ Loses data.

B. Use different software? ⟶ Has pros and cons...

So, what are our options when dealing with complex data types?

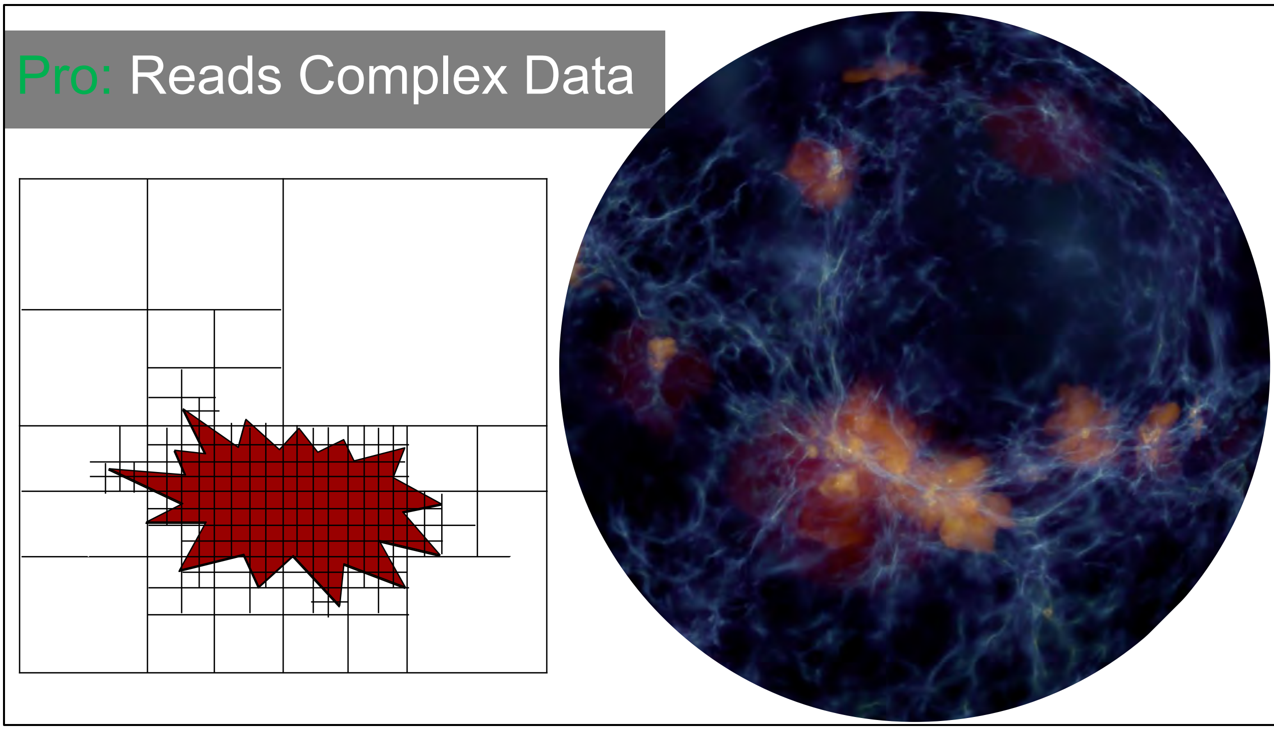
We can force the data to fit into the structure that the software requires, by sampling it. The data can be sampled in such a way that the full resolution is maintained, by having the sampling rate equal to the smallest cell size of the data. But, this oversamples areas where there is less data resolution, and this can easily become unmanageable due to memory and time constraints, if there is a large difference between the smallest and largest cell sizes. Alternatively, the data can be sampled at a more convenient rate, one that is not at the finest data resolution. This will lead to more manageable time and space requirements, but it loses data in the highest-resolution areas of the simulation, which is where the most interesting events are occurring. This is not ideal.

Which brings us to option B – stepping away from visual effects software, and using scientific software instead.

Tools for Science

Case Study: yt

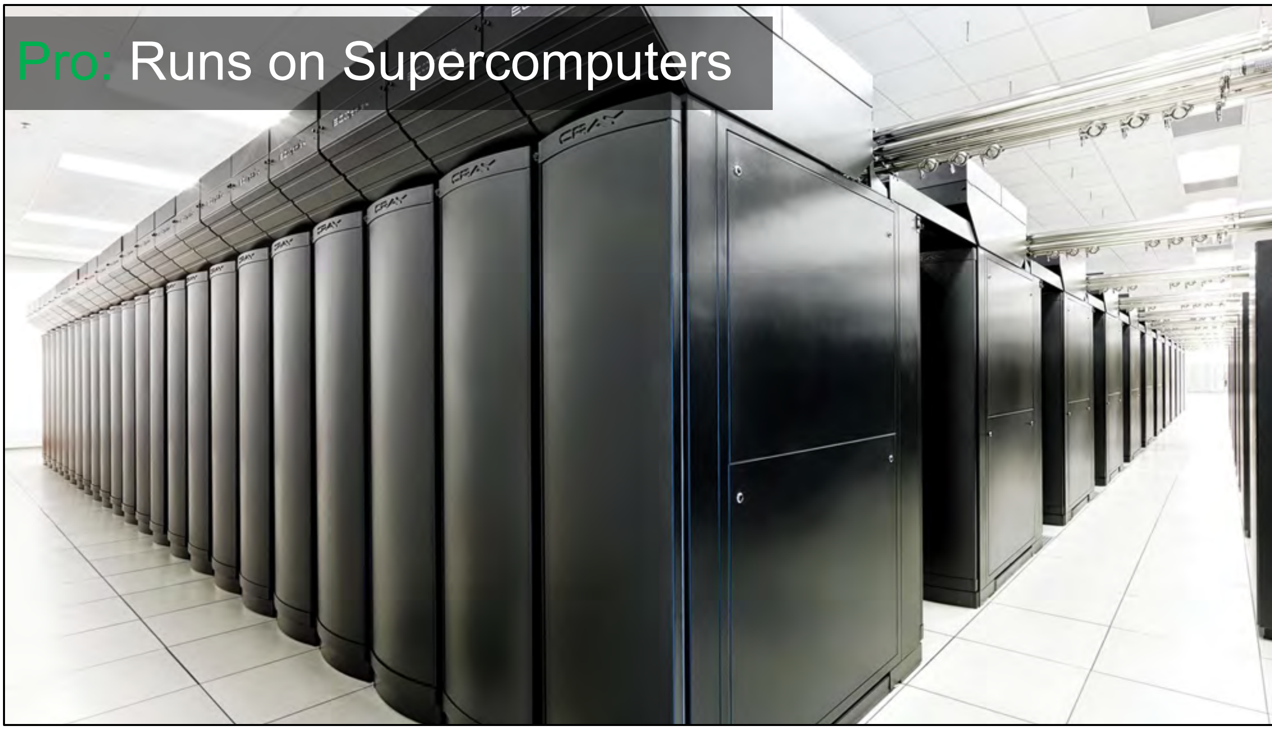
Like there are many options for VFX software (Houdini, Maya, Blender, etc), there are also many options for scientific visualization software (yt, VMD, VisIt, ParaView, etc). Here, we discuss yt, a scientific analysis and visualization Python package.



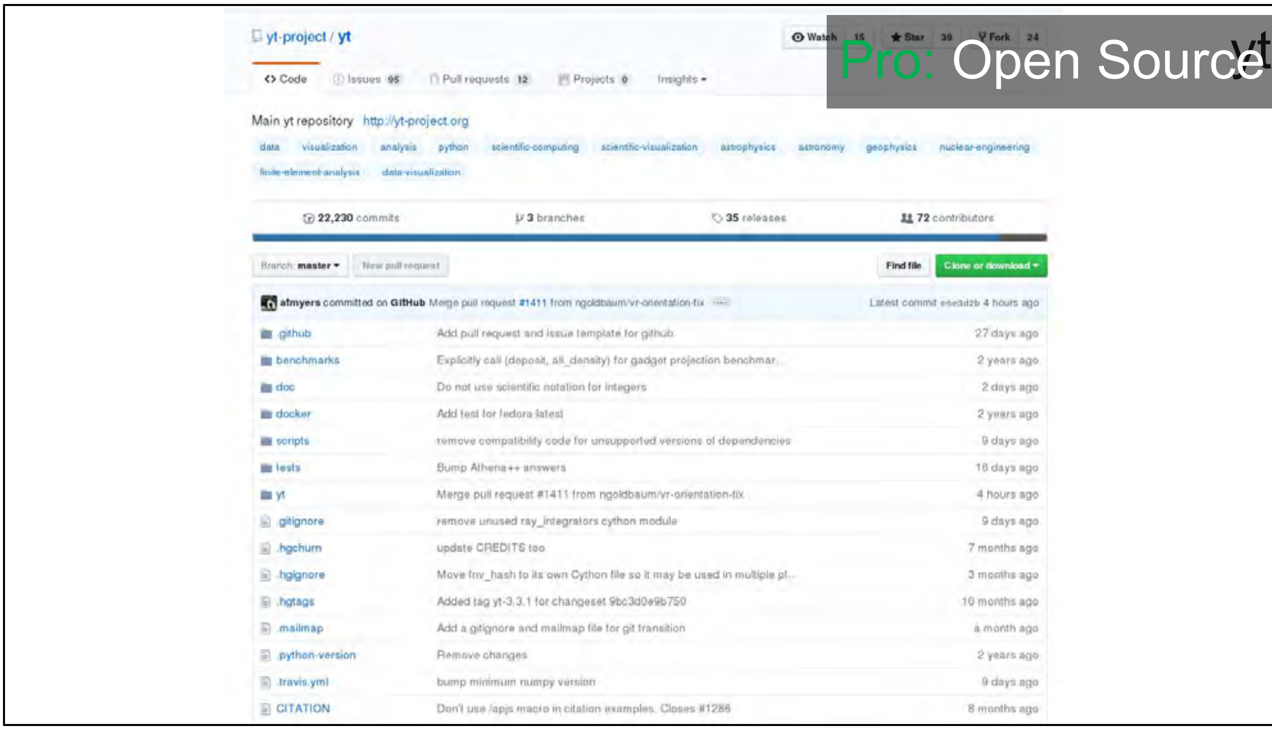
yt is the main scientific visualization tool we have used for visualization of complex astrophysical data formats, such as nested multi-resolution grids as described in previous slides. The image on the right is an AMR data set rendered with yt.

yt is free and open-source and available at http://yt-project.org/

yt is a tool aimed at scientists more so than at visualization experts. As such, it includes many features useful to scientists, such as creating derivative data like tracing streamlines through data. This is something that we often do in Houdini, but yt makes this simpler, just taking a few lines of code.



yt works in parallel and is installed on the Blue Waters supercomputer. We have taken advantage of being able to run yt across thousands of cores. We have used as many as 16,327 hours in one day of rendering on Blue Waters, which would come out to over 3 weeks on our 2 local clusters, or 2 years on a single machine.

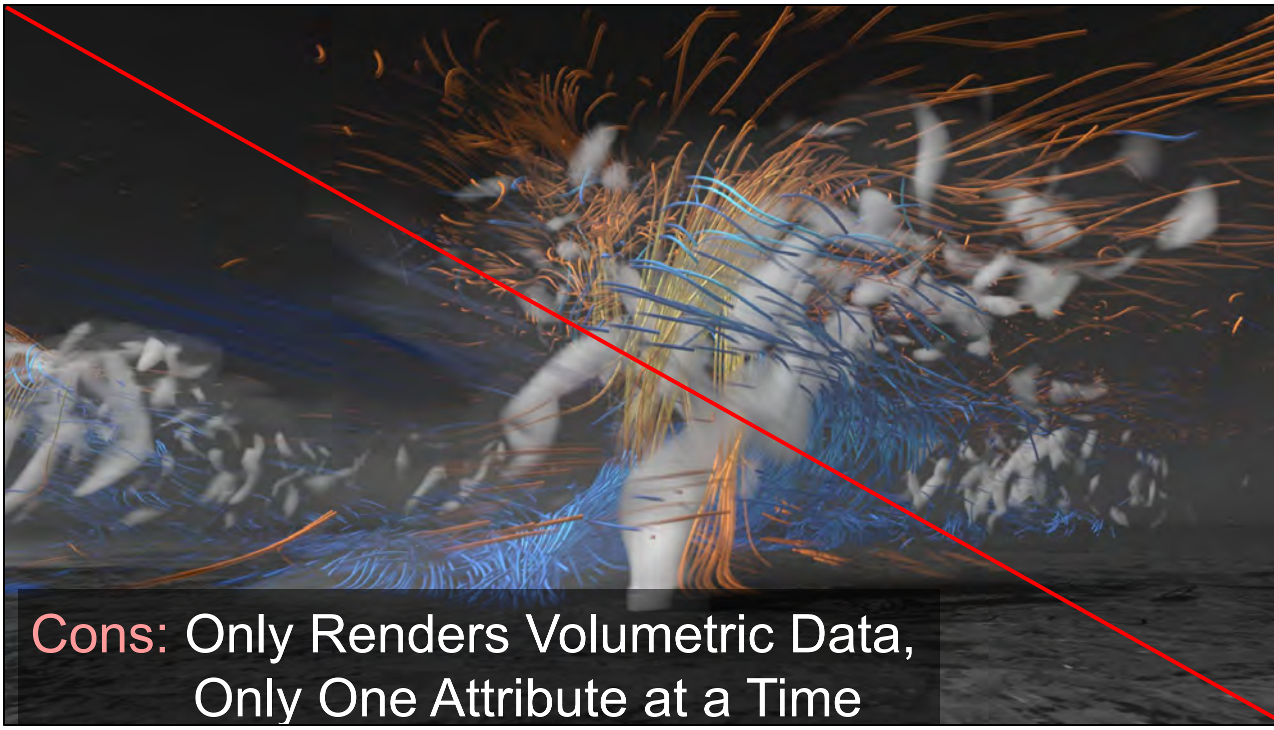


Off-the-shelf, more general-purpose tools can never do *everything* that you want them to. Conveniently, yt is open-source, and one of its primary creators, Dr. Matt Turk, is an NCSA colleague. We have worked with him and his team to develop the software so that it’s usable for cinematic rendering.

So why isn't yt our primary tool?

A problem that arises with scientific tools is that they often require domain experts to be able to use the software. In this case, yt is a Python tool, and does not have a user interface, so the required experts are programmers.

Typically, while working on a project, our pipeline starts with our programmers ingesting the data, and creating tools which are handed off to the artists. When using yt, the artists play less of a role, as most of the steps of the pipeline must be done programmatically. Color maps, camera paths, shaders, and all artistic elements need to be either created programmatically, or created in external software and then imported - both of which require more coding and more work than a more artist-friendly tool like Houdini.



More downsides: yt only renders volumetric data, and only one object at a time. yt also doesn't have many settings that artists may be used to, such as raytraced lighting and shadows, ambient occlusion, shading and texturing, etc. It makes shots like this impossible, where a volume is integrated with geometry, with lighting and shadows.

Customizing the Tools

	yt	Houdini
Data analysis	X	
Runs on supercomputers	X	
Can read complex scientific data	X	
Open source	X	
Artist-friendly		X
Can render geometry		X
Can render multiple objects in a scene		X
Lighting		X

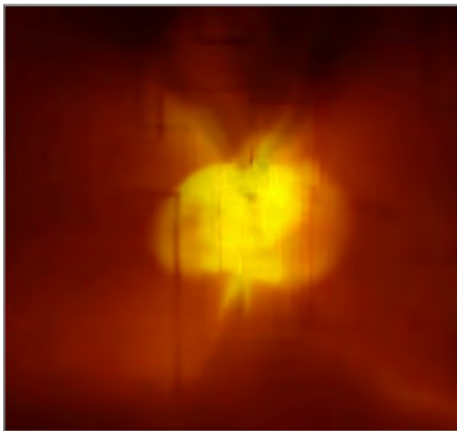
Scientific tools like yt and visual effects tools like Houdini have complementary strengths, but neither alone is ideal for cinematic scientific visualization. yt is a great tool for easily reading and analyzing many complex types of scientific data, but Houdini provides the artistic flexibility that we need.

	yt	Houdini	ytini
Data analysis	X		
Runs on supercomputers	X		
Can read complex scientific data	X		X
Open source	X		X
Artist-friendly		X	X
Can render geometry		X	X
Can render multiple objects in a scene		X	X
Lighting		X	X

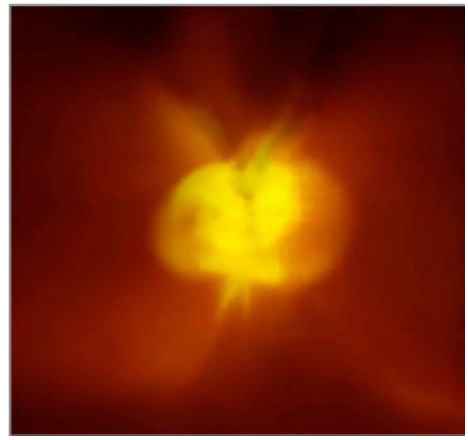
Custom middleware

To solve some of these problems, we created an open-source middleware between yt and Houdini which we call Ytini (rhymes with "martini"). This allows us to keep doing the bulk of our work in Houdini, but uses yt as a data reader for complex scientific data types.

AMR in Houdini with ytini



Rendering showing the underlying AMR structure



Final rendering with Ytini

This work is supported by NSF award for CADENS ACI-1445176, NSF grant AST-1402480, and the Fiddler endowment

In the simplest use case, we use yt to load the data, export it into the Houdini-compatible, sparse volumetric OpenVDB data format, and load the volume into Houdini.

Here is a rendering of AMR data done in Houdini. Our team has been dealing with AMR data for decades, and this is the first time we have been able to render it in visual effects software.

On the left is an early attempt, which has very visible lines between the different AMR levels, and on the right is the final, seamless rendering.

Programming and Scripting

* Import / Export
* Custom Software
* Plugins into Existing Software
* Middleware
* Data Wrangling
* Data Transformation
* Data Translation
* Derivative Data Creation
* Interpolation
* Proceduralism

Writing middleware or customizing software via plugins and scripts is often a necessary and crucial part of the pipeline, for a wide variety of tasks. Proceduralism is often important in just about every step of the pipeline. Scripts can be platform independent, allowing visualizers to not worry about recompiles, links, operating systems, versions, etc.

The full 5-minute, narrated, 360-degree movie is available at
http://svs.gsfc.nasa.gov/4685

Goals and Purpose

- Highlight capabilities of NASA's **GPM** mission using 2D and 3D data
- Explain a few science concepts about hurricanes
- View a hurricane from inside - at 'ground' level - surrounded by data
- Desire for cool content for social media to increase engagement

GPM = Global Precipitation Measurement mission
a spacecraft with microwave and radar instruments that measures rainfall globally

Project Team

Scientists (4)

Visualizers (4)

Media Professionals (2)

Systems Support (4)

Story Development: Science Points

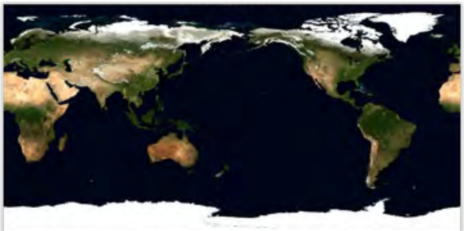
- Fly through hurricane precipitation rain bands
- Visit ‘hot tower’ on northeast side of storm
- Show liquid/frozen precipitation and the “melting layer”
- Visit hurricane eye
- Show 3d drop size data

Story Development: Visualization

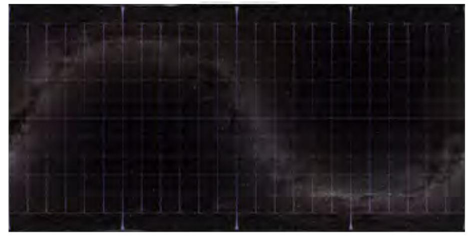
- Experimentation with 3d precipitation data sets
- Drafted
  - Camera path (locations, speed)
  - Data representation at different points in the story
- Slowly converged on (~112 iterations)
  - Camera path and speed
  - Data representation at different times
  - Narration script explaining what the view sees
  - Visuals drive Story ; Story drives Visuals

Context: Standard Static Data Sets

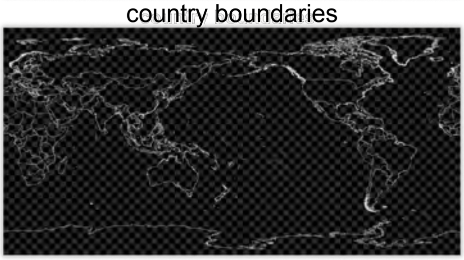
Blue Marble



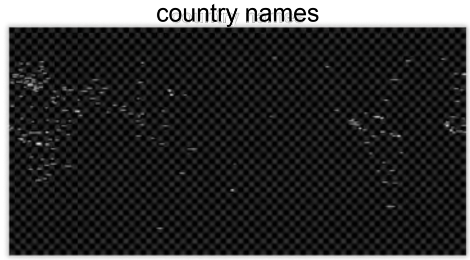
star field



country boundaries



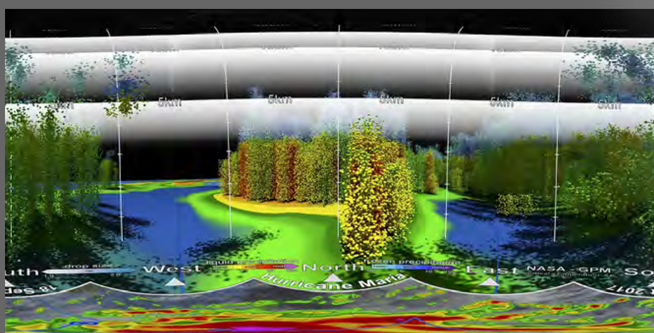
country names

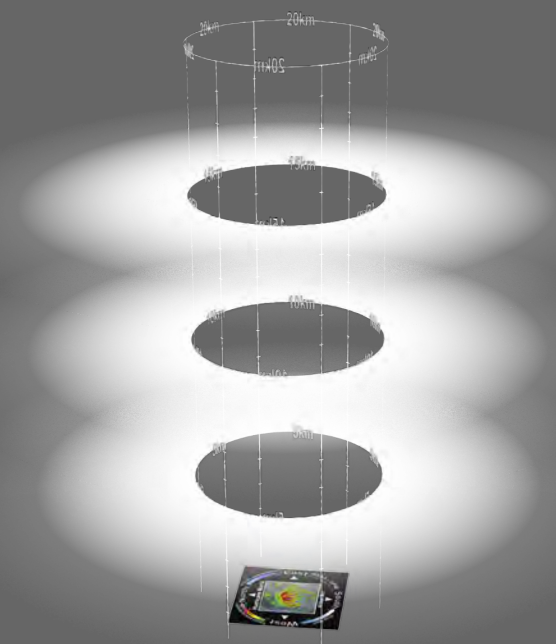


Context is a critical part of a visualization

- Standard, reusable, ready to render data sets
- Single time step
- Very high resolution mip-mapped

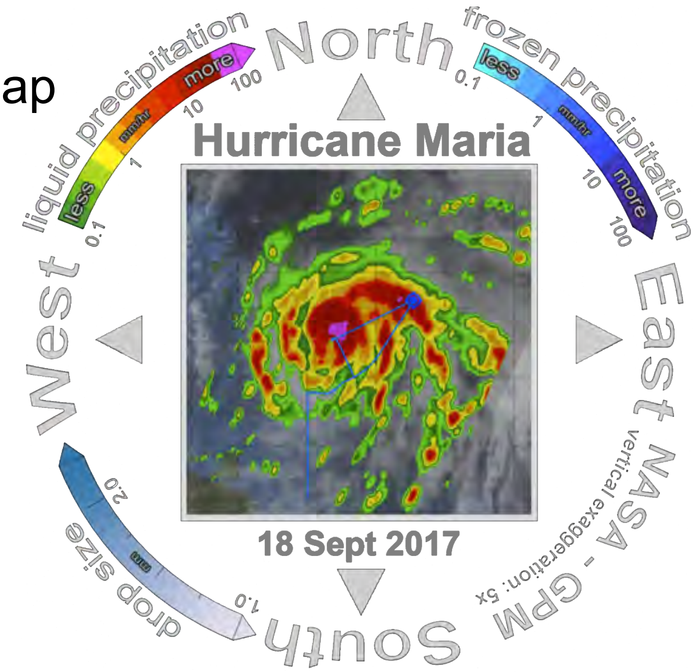
Context: Altitude





3d height cylinder that camera sits in - moving up and down through the center of the cylinder

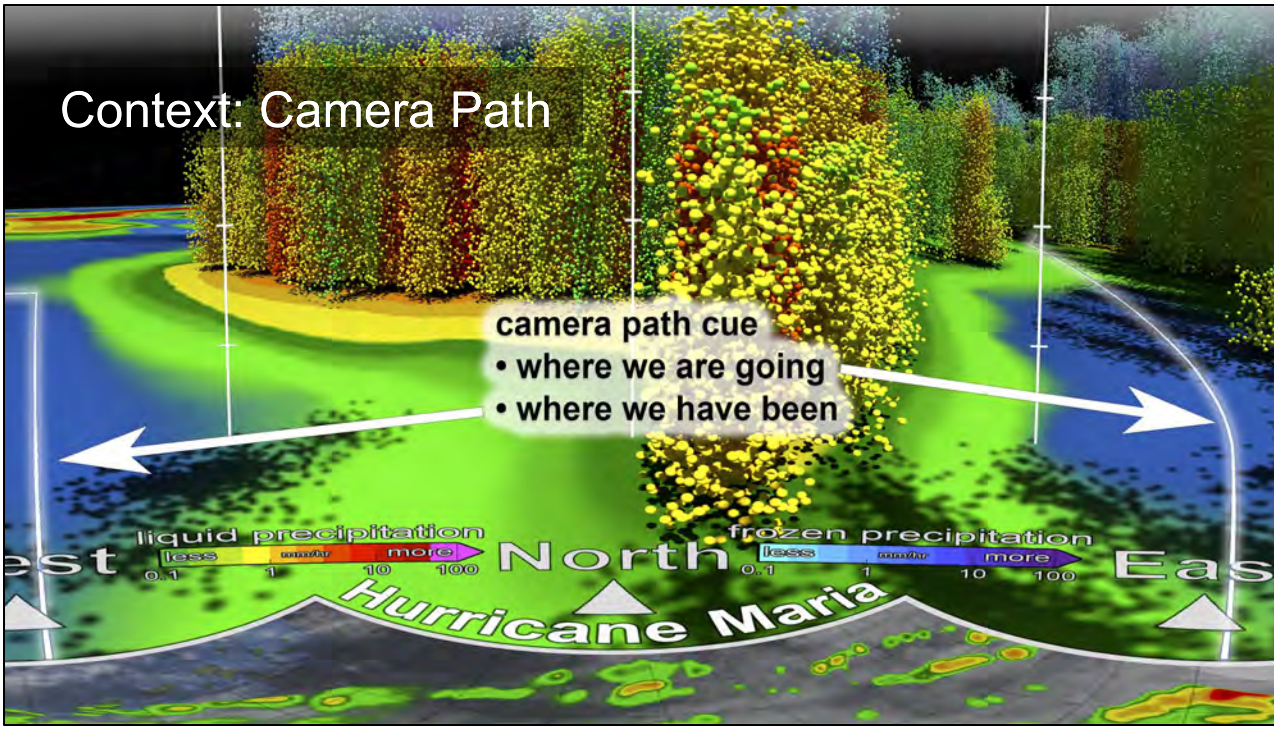
Context: Map



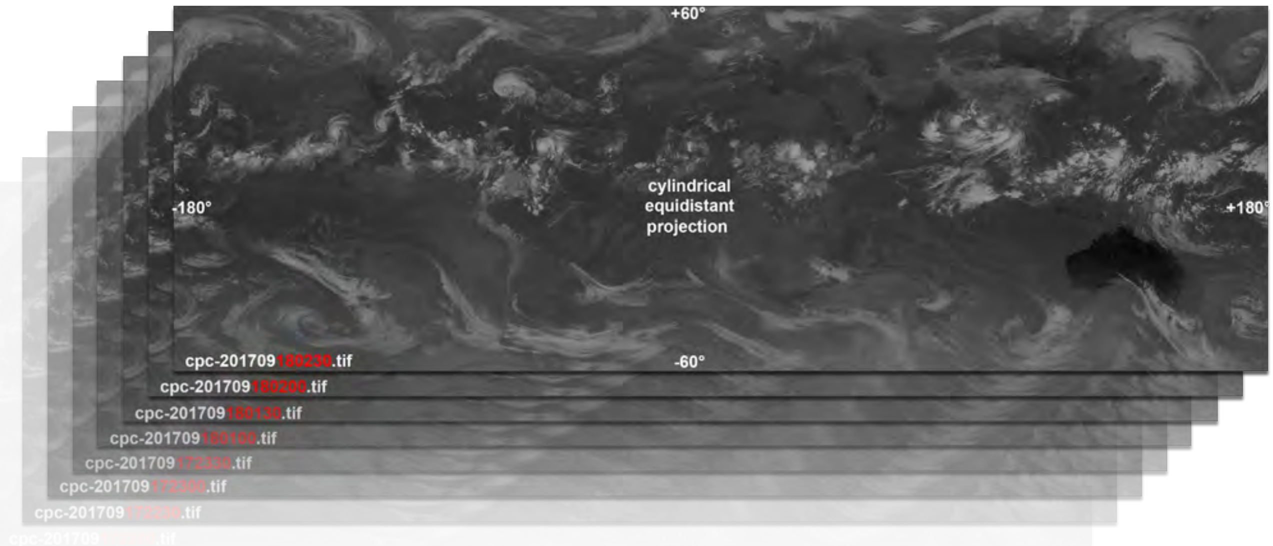
Map includes

- Color bars
- Compass directions
- Date
- Attribution to NASA and GPM
- Camera path and current location overlay in blue

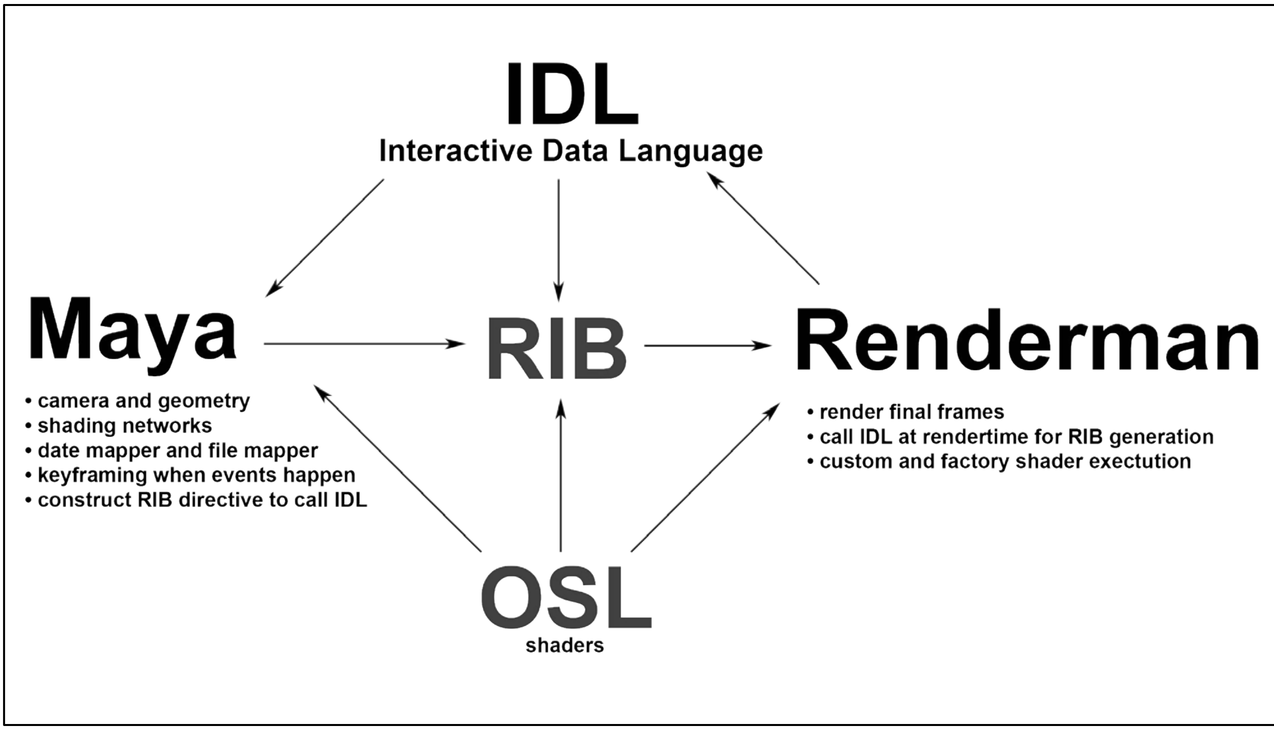
Context: Camera Path



Context: Cloud Data



- This is about a hurricane, we need some clouds for context
- Climate Prediction Center (CPC) infrared cloud mosaic
- Time varying - different data set every 30 minutes
- Infrared good to show clouds day or night
- Only +/-60 degrees
- Motivation to show date mapper and file mapper on next slide -- then following slide is another example of time-varying
    - Science data sets are long and messy - these are simple examples
    - We’ve learned from experience that keeping derivative files (like floating point textures) close to the names of the science data sets is MUCH better than renaming/linking to data.0001.tif, data.0002.tif - because you’ll quickly lose track things and have difficulty changing time rates


IDL
Interactive Data Language
Maya
RIB
Renderman
• camera and geometry
• shading networks
• date mapper and file mapper
• keyframing when events happen
• construct RIB directive to call IDL
• render final frames
• call IDL at rendertime for RIB generation
• custom and factory shader exectution
OSL
shaders

Software

Date and Time

- Scientific data visualizations often include data from multiple dates/times
- but...VFX software doesn't know about dates/times
- A good way to represent a date/time is using a Julian Date (JD)
  - double precision number used in Astronomy
  - number of days (and fractional days) since noon UT January 1, 4713 BC
- We often use a 'modified' JD
  - modified MJD = JD(current) - JD(Jan 1 2000)
  - this makes the number more manageable, but still not human readable
  - example: August 1 2019 11:30am
    JD = 2458696.979167
    MJD= 7152.479167

Date Mapper

- Maya container (set) of attributes: MJD, year, month, day, hour, min, sec
- MEL script sets MJD keyframes to desired date/time
- MJD can be keyframed since it’s a single number, controlling the flow of time
- MEL script runs every frame that converts the MJD to date/time
- Results stored in user attributes: year, month, day, hour, minutes, second
- Generic - works across all projects

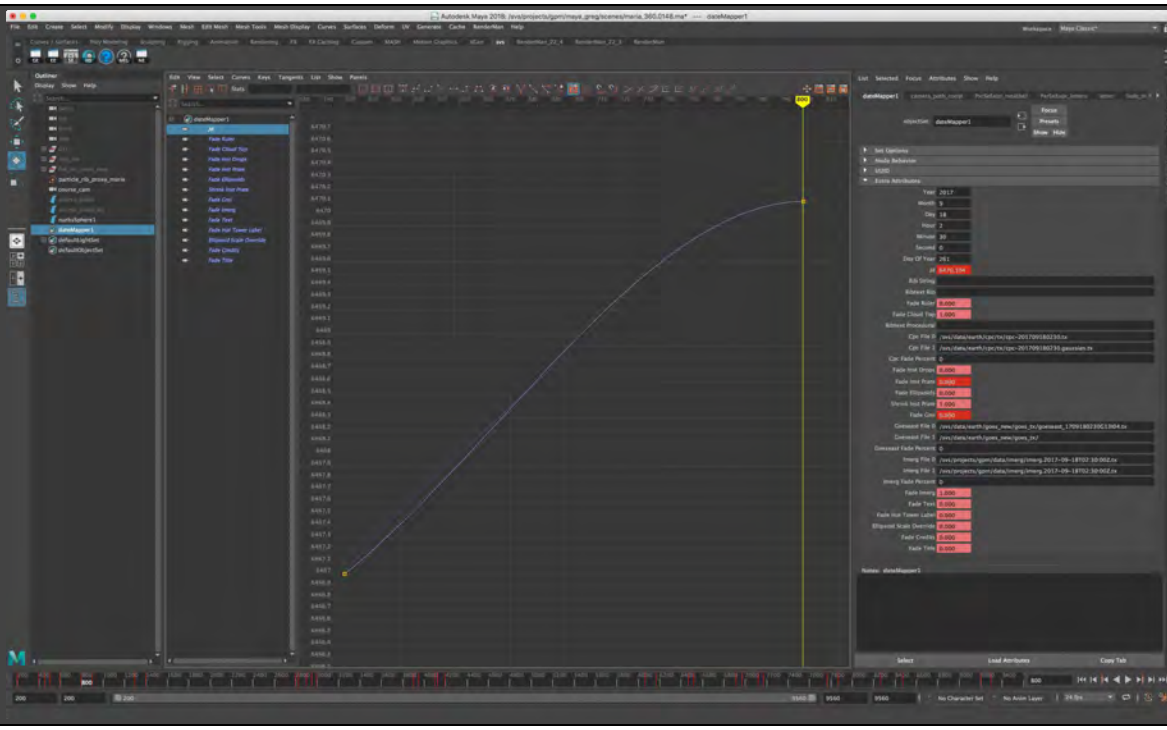
File Mappers

- custom MEL scripts that generate file names given current date
- File name is full path to the data file with science data name embedded
- File mapper scripts are executed every frame after the Date Mapper
- analogous to a shader: given certain inputs, return prescribed output
- keeping science data file names for processed files (like floating point tif files) is better that renaming/linking to data.0001.tif, data.0002.tif, etc.

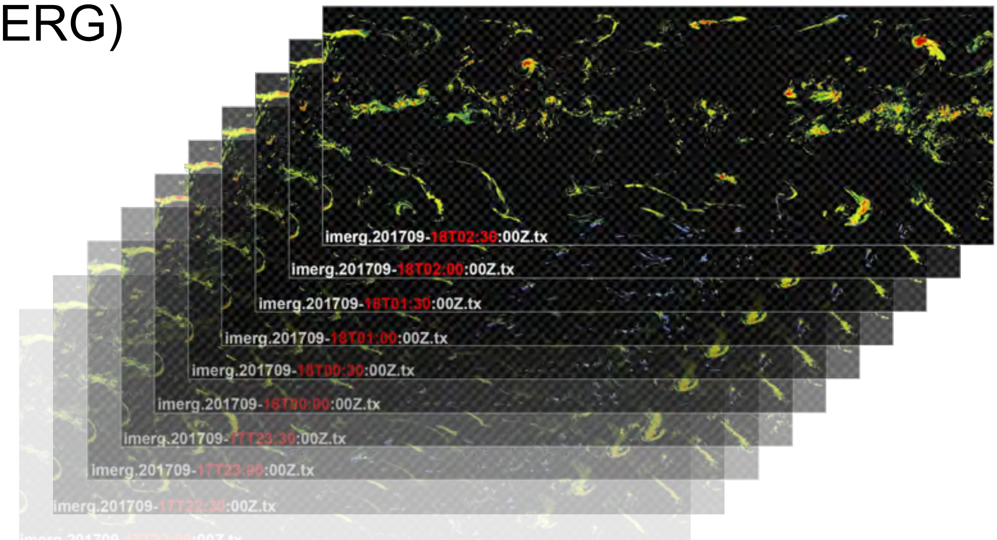
We've learned from experience that keeping derivative files (like floating point textures) close to the names of the science data sets is MUCH better than renaming/linking to data.0001.tif, data.0002.tif - because you'll quickly lose track things and have difficulty changing time rates

Makes interpolation between frames easy -- file0, file1, fade

Date Mapper Interface



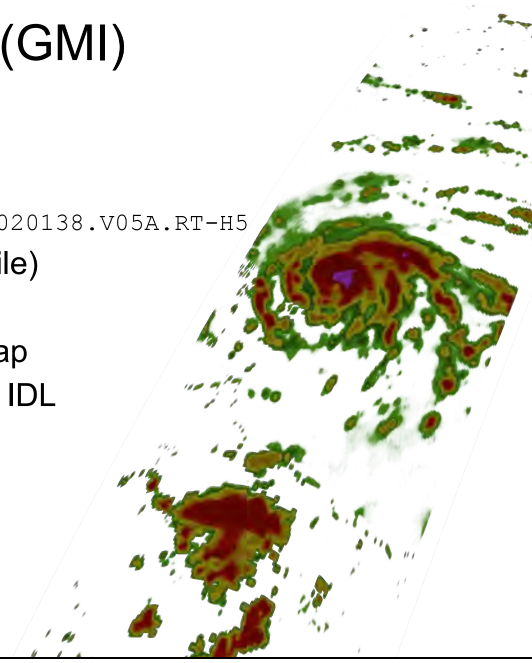
Data: Integrated Multi-satellite Retrievals for GPM (IMERG)



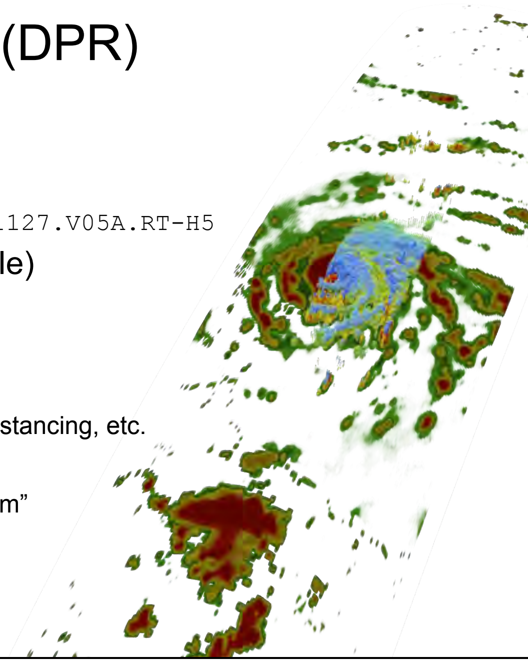
- time varying every 30 minutes
- range 0 mm/hr to over 50 mm/hr
- projection: cylindrical equidistant
- color table: imerg legacy

Data: GPM Microwave Imager (GMI)

- ground level precipitation rates
- swath width 904 km
- `2A.GPM.GMI.GPROF2017v1.20170918-S015640-E020138.V05A.RT-H5`
- HDF5 file format (lots of parameters in the file)
- liquid / frozen flag
- 2d data set used as the base level of the map
- parameters from this data set processed by IDL
- saved to floating point texture



Data: Dual Precipitation Radar (DPR)

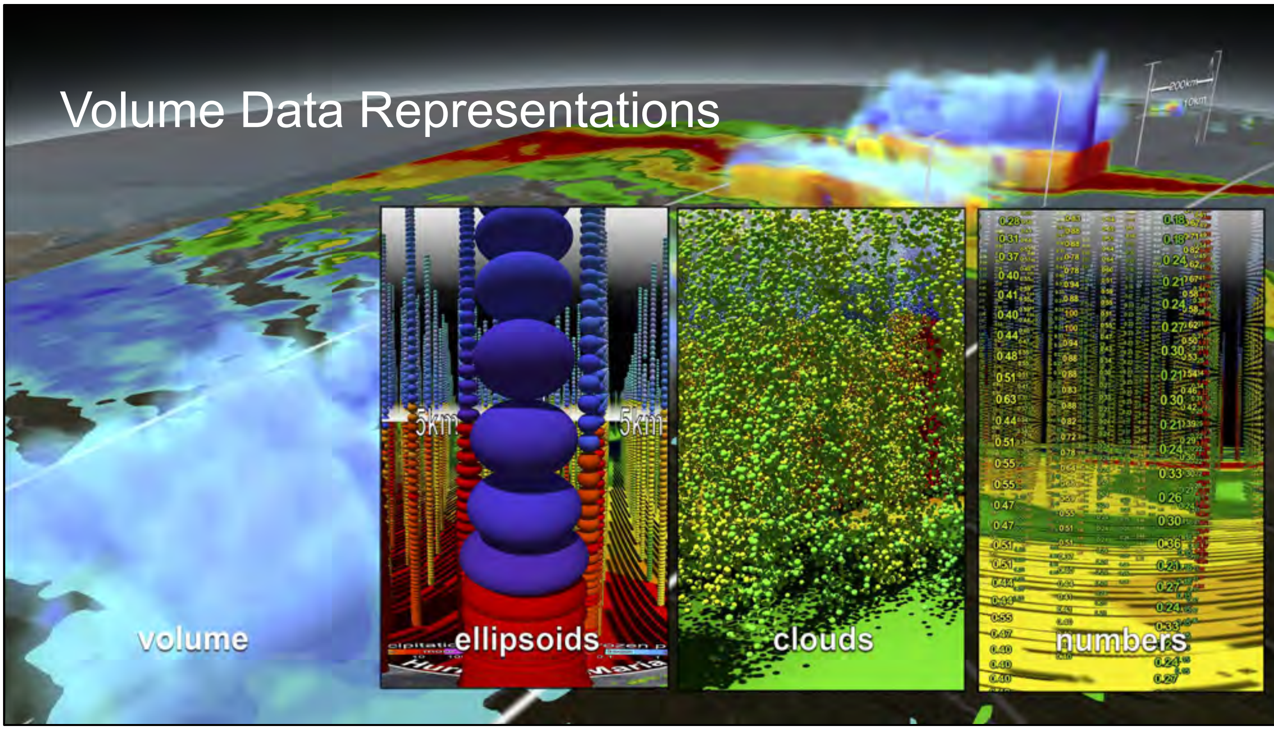


- volumetric precipitation rates
- swath width: 245 km
- `2A.GPM.Ku.V720170308.20170918-S014128-E021127.V05A.RT-H5`
- HDF5 file format (lots of parameters in the file)
- liquid / frozen flag for every voxel
- MEL script runs every frame
  - builds RiProcedural RIB directive
  - Collecting keyframed particle sizes, colors, type, instancing, etc.
- at render time
  - Renderman calls IDL via RiProcedural “RunProgram”
  - IDL generates RIB based on parameters passed in
  - RIB is sent pack to Renderman for rendering

RiProcedural call example

```
Procedural "RunProgram"  ["/usr/bin/python    /svs/share/lib/idl/particle_rib/dev/particle_rib.py"
"name_suffix=text file_type=2 frustum_nudge=0.01,0,0.01 object_rotate=0,0,0 scale_override=0,0,0
scale_max_override=0,0,0 diagnostic_hud=0 glow_gain=0.1 object_type=3
cbar_tx_file=/svs/projects/gpm/idl/colorbar_gpm_linear_GREG.tex cbar_min=-100 cbar_max=100 binary=1 align_flag=1
align_to=52.27851508,-82.4173314,22.54690322 align_up=0.5193868565,-0.8188158866,0.2444950659vals_min=-100
vals_max=100 size_min=0.04 size_max=0.04 percent_shown=1 exag=5 shader_lighting=1 shader_color_source=5 kd=1
density_power=1 inst_val_min=0 inst_val_max=0 inst_distribution_type=0
data_file=/svs/data/earth/gpm/hdf5/20170918_Maria/2A.GPM.Ku.V720170308.20170918-S014128-E021127.V05A.RT-H5
gpm_i0=1750 gpm_i1=2000 gpm_upscale_levels=1 gpm_upscale_across=1 gpm_upscale_along=1 inst_size_lon=0
inst_size_lat=0 inst_size_alt=0 projection=1 inst_max_count=0 abs_of_density_variable=1 camera_cull=0
camera_pixel_width=4096 camera_pixel_height=2048 camera_world=52.27851508,-82.4173314,22.54690322
camera_incident_vec=-0.1309622796,0.206462666,0.9696504777 camera_up_vec=0.5193868565,-0.8188158866,0.2444950659
camera_focal_length=35 camera_aperture_width=1.41732 camera_aperture_height=0.94488 cam_name=camera_main
cam_up_name=cam_up cam_in_name=cam_in film_fit=0 density_variable=vals
idl_pre_rib=pos=where((vals)ne(0.0))&x=x[pos]&y=y[pos]&z=z[pos]&opacs=0.0*opacs[pos]+1&vals=vals[pos]&particle_s
izes=particle_sizes[pos]&particle_sizes=(0.015+0.030*(clamp(abs(vals),0,150))/150.) "]
[-200 200    -200 200    -200 200]
```

MEL script executes every frame creating a string that is a RIB RiProcedural call
Red = RiProcedural
Blue = helper application that gets called (it's calling IDL)
Green = HDF file name that is passed to IDL to be read and processed at render time
Orange = alignment information for the numbers per frame

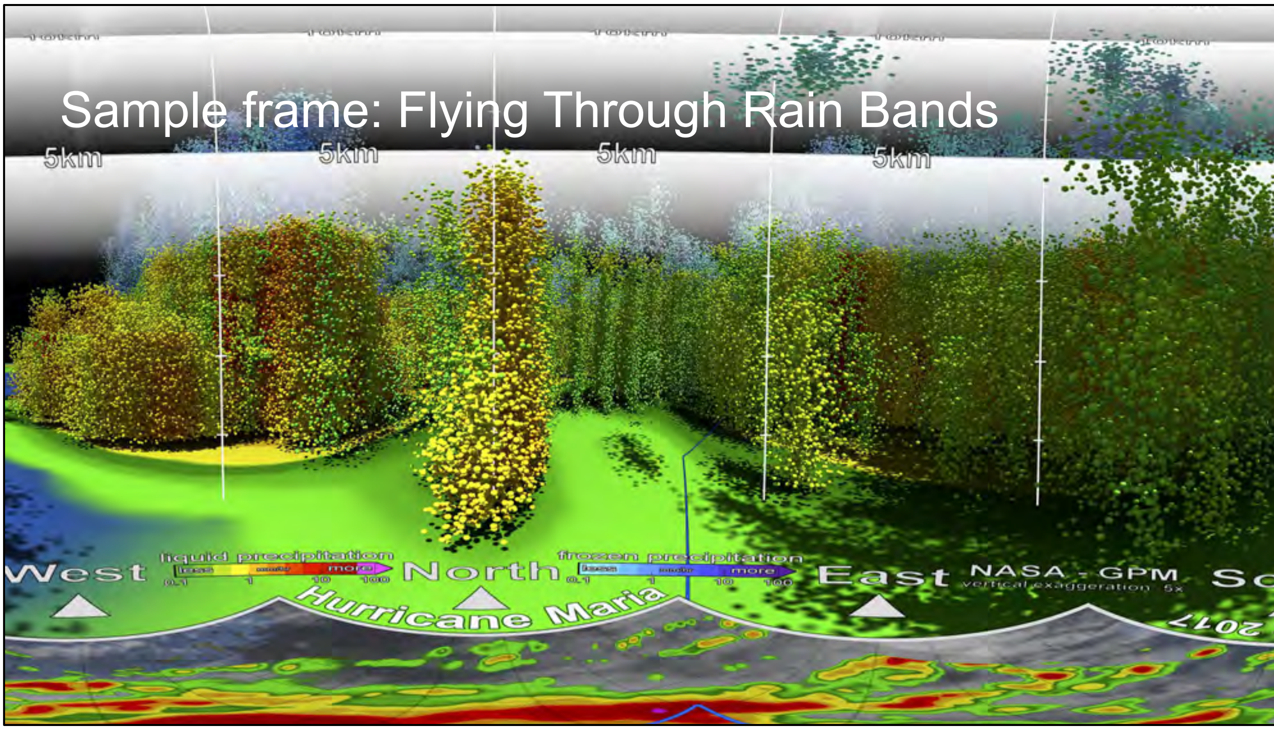


We chose particles to represent volumes instead of traditional volume rendering
Why not volume rendering?

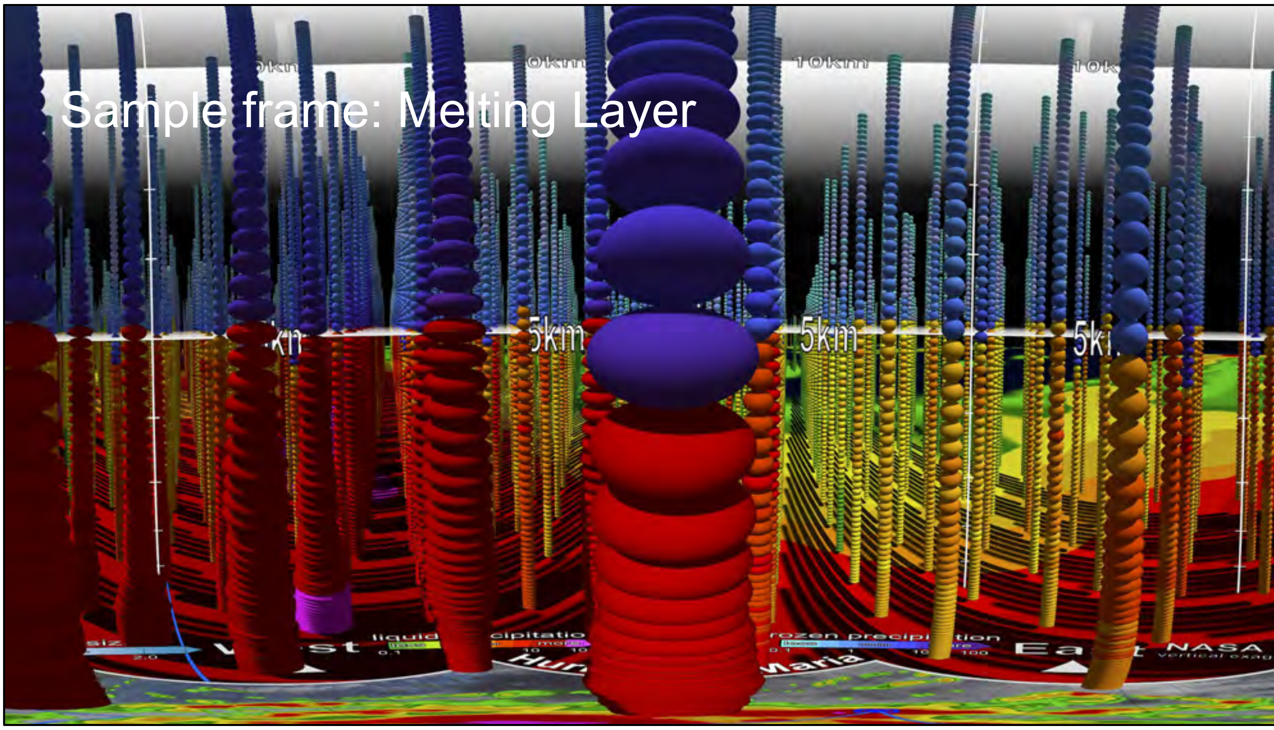
- Particle rendering much faster especially when close to volume
- Explicitly represent each individual voxel measurement revealing instrument’s data collection resolution and spacing
- volume rendering can blur data, particularly close up
- New look

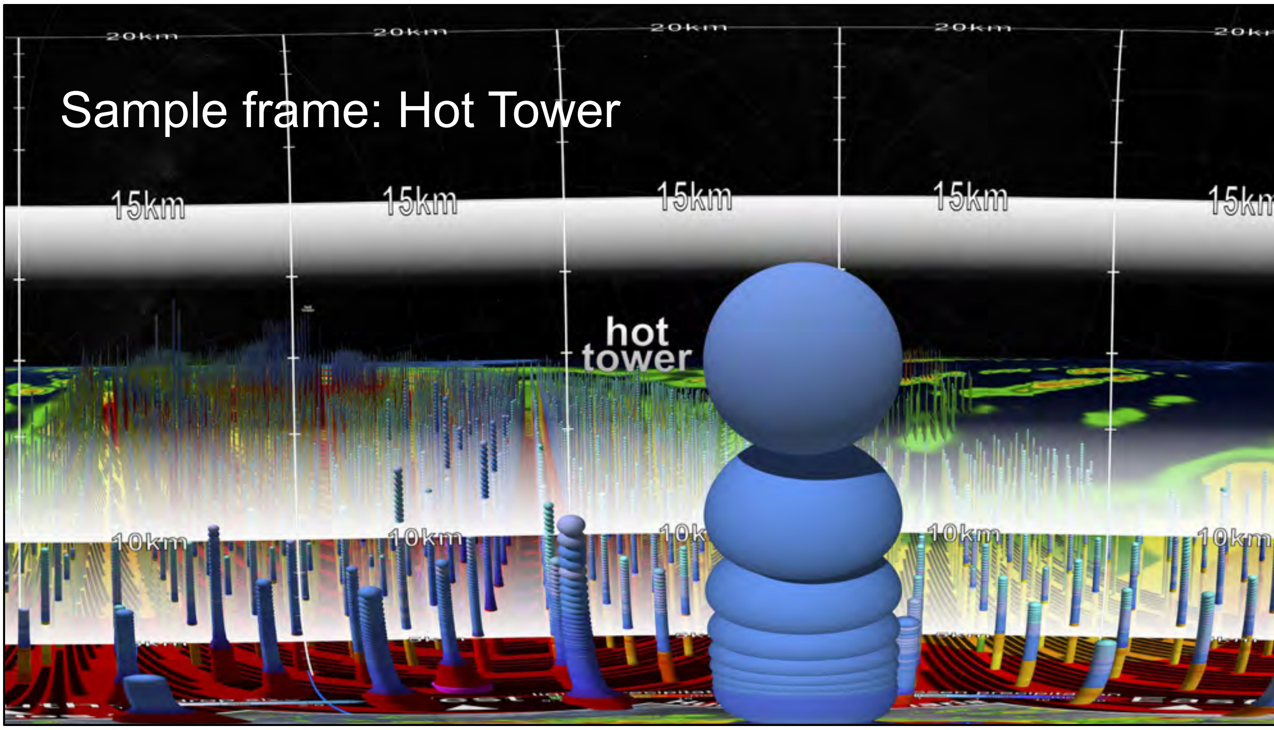
Vertical exaggeration - 5x - a way to deal with scale problems
Shadows

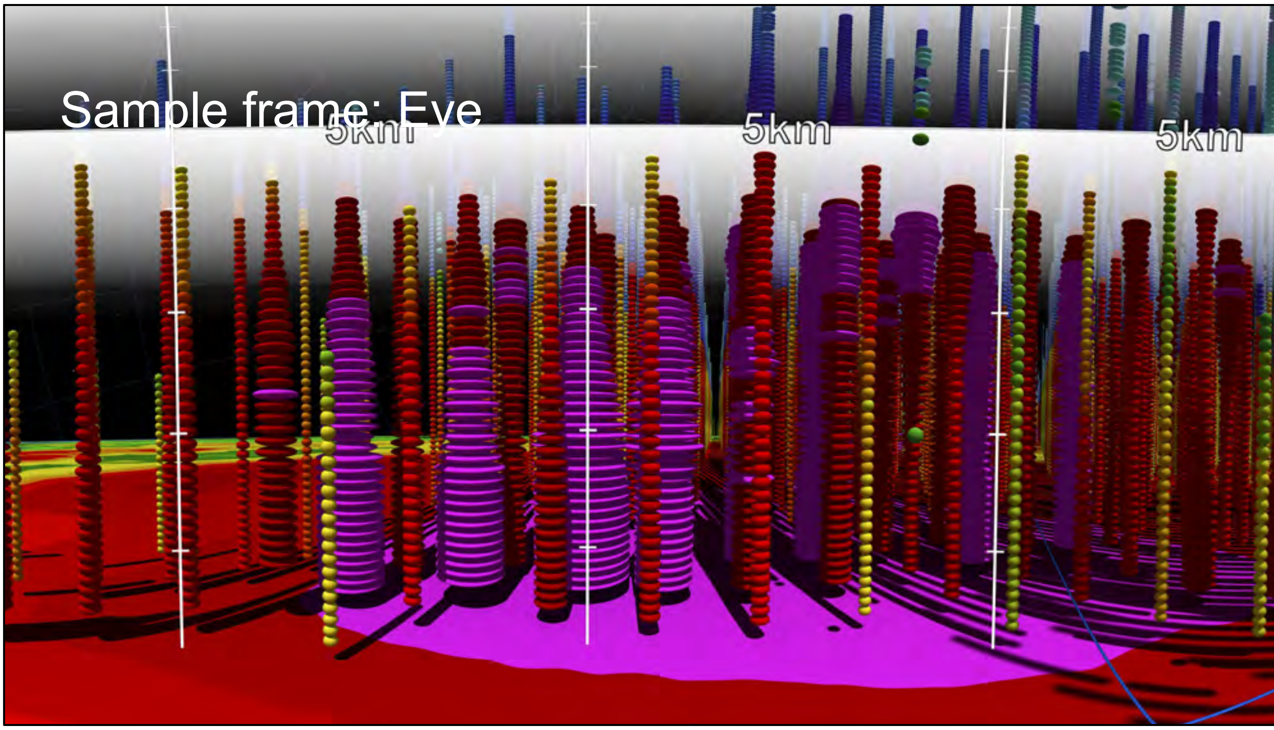
Sample frame: Flying Through Rain Bands

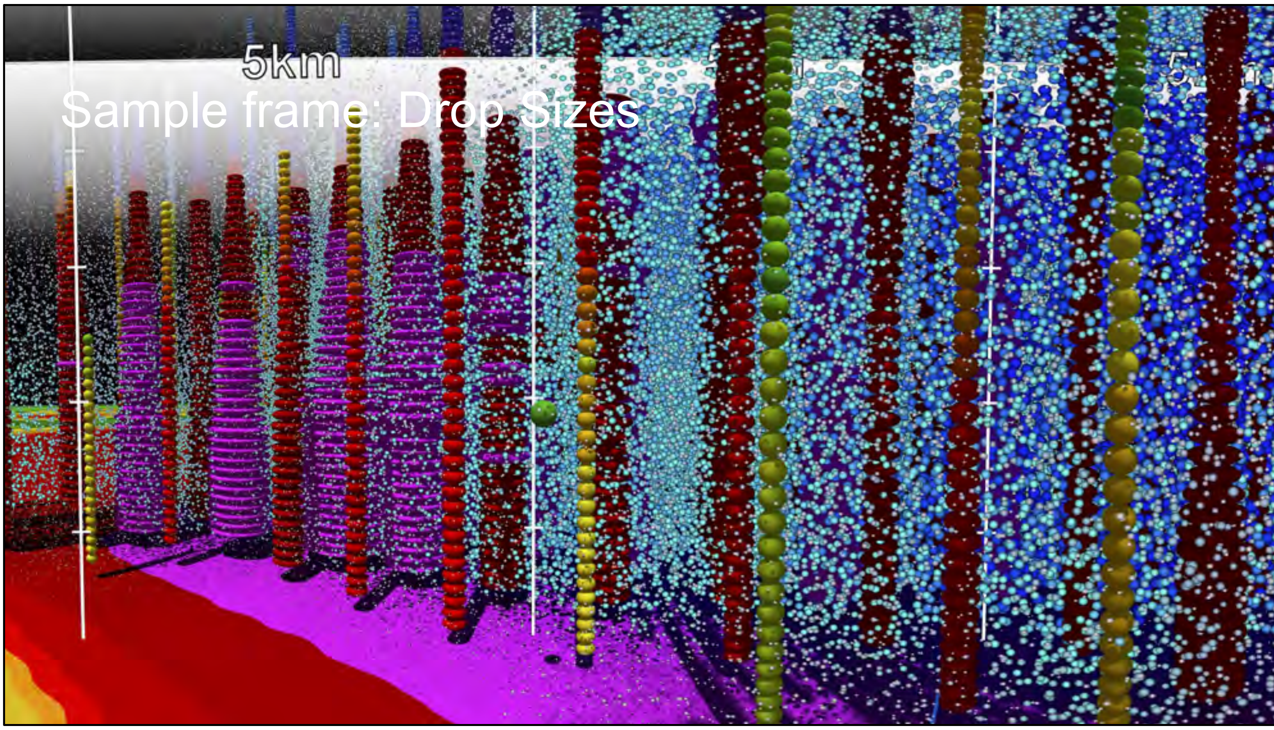


Sample frame: Melting Layer

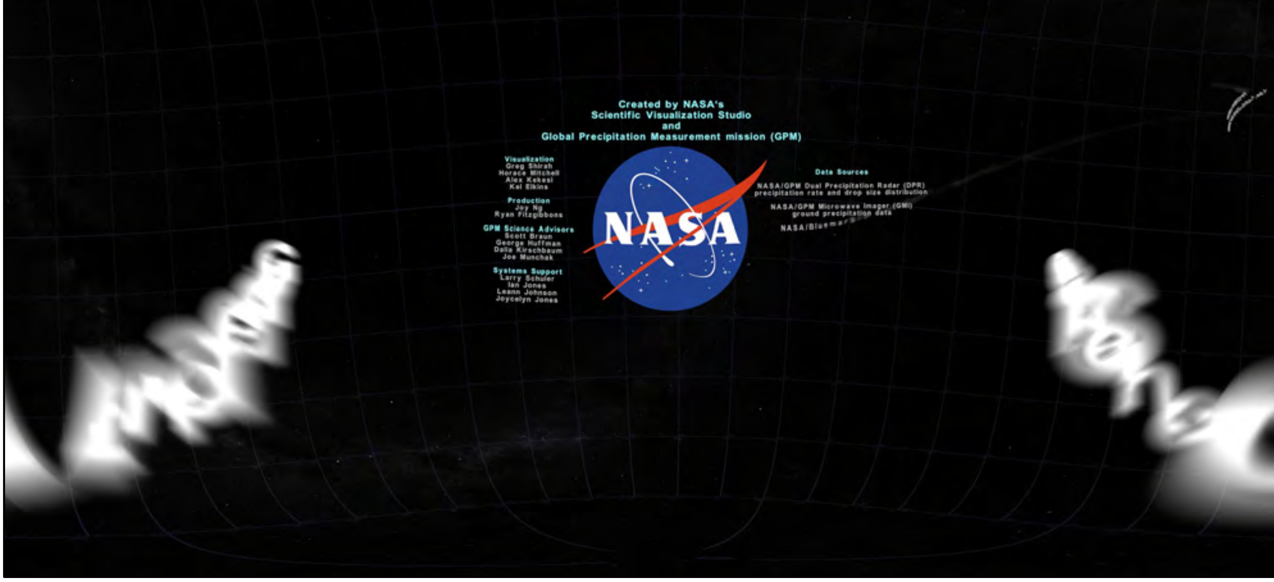



Sample frame: Hot Tower
20km
15km
hot
tower
10km


Sample frame: Eye
5km
5km
5km


5km
Sample frame: Drop Sizes

Sample frame: Fun With Credits



360 Production challenges

- camera aim - let viewer chose where to look
- Tests take longer
  - hard to judge without warping correctly
  - youtube uploads or specialized app like GoPro
  - 4k x 2k took longer to render
- narration script aware viewer might be looking away from point of interest
- numbers individually oriented towards camera every frame
- 8k
- artifacts

Particle procedural code orients planes, OSL shader that renders text

Results

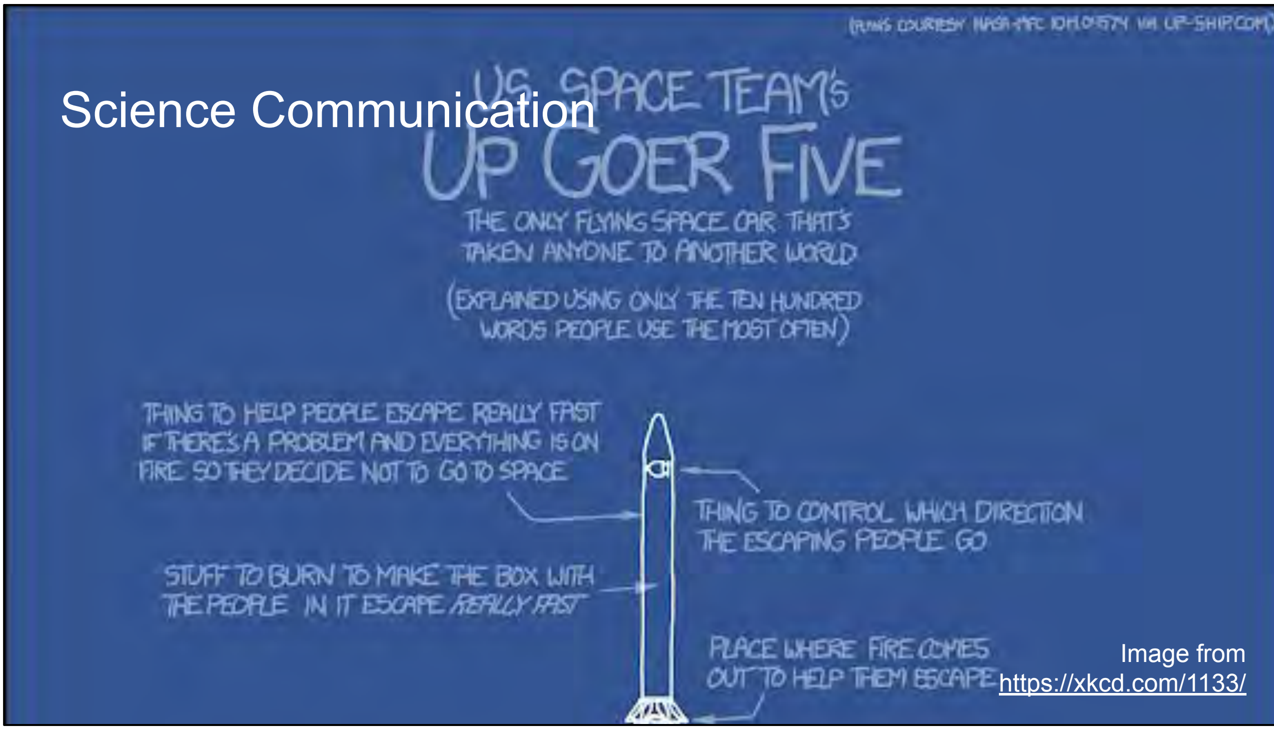
- released on October 4, 2018
  - www.nasa.gov
  - NASA, Goddard, and SVS youtube channels
- over 200,000 views on social media
- GPM's most watched video to date
- sudden assumption that SVS can turn any visualization into a 360 movie
- SIGGRAPH 2019 VR Theater

Making the Visualization **Meaningful**

Cinematic Scientific Visualization is a powerful tool for anyone looking to do fact-based outreach.
Effectively, we translate "film language" into "science language". “Film language” is something very familiar in our culture, and which learners of all ages can relate to, as kids are typically exposed to family films before they even understand spoken language.

Storytelling

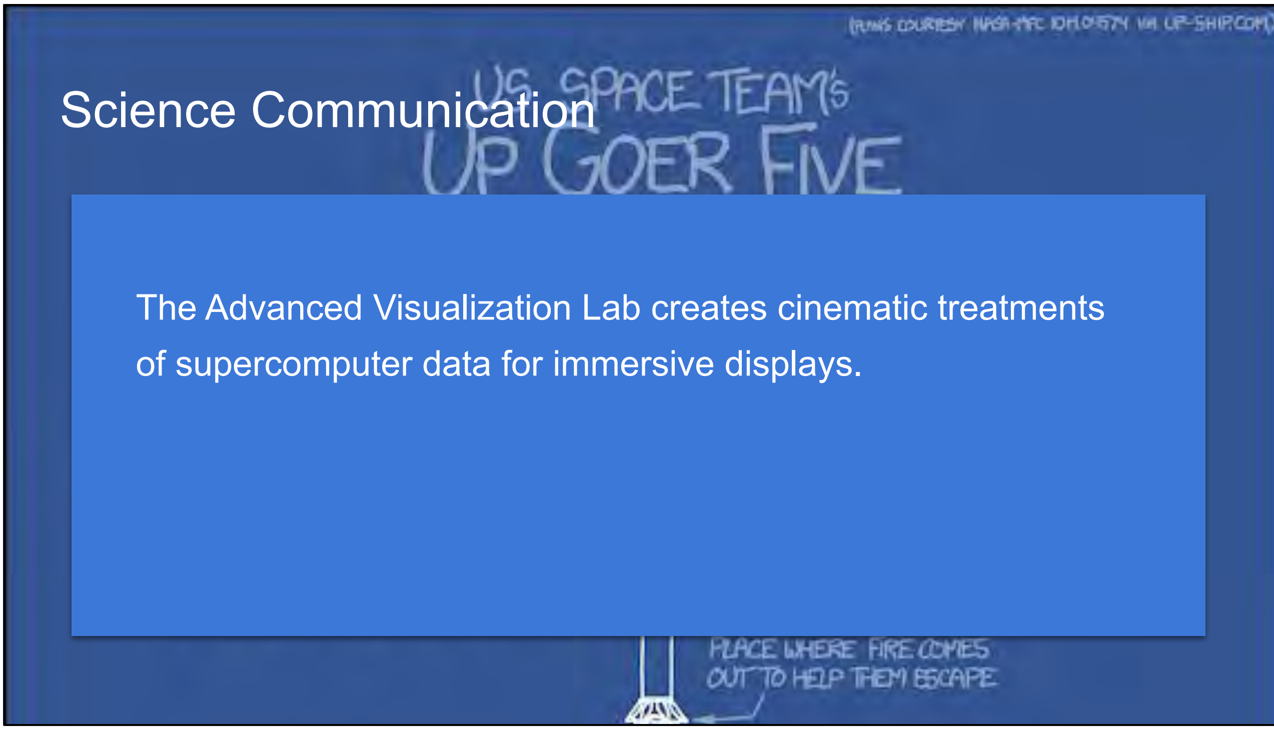
- Science communication
- Identify and tailor to your audience
- Narrative
  - Science Narrative
  - Documentary Narrative
- Describing your visualization vs describing the science



At its heart, scientific visualization is a form of science communication, much like Radiolab podcasts, dinosaur museum exhibits, and lectures by Bill Nye the Science Guy.

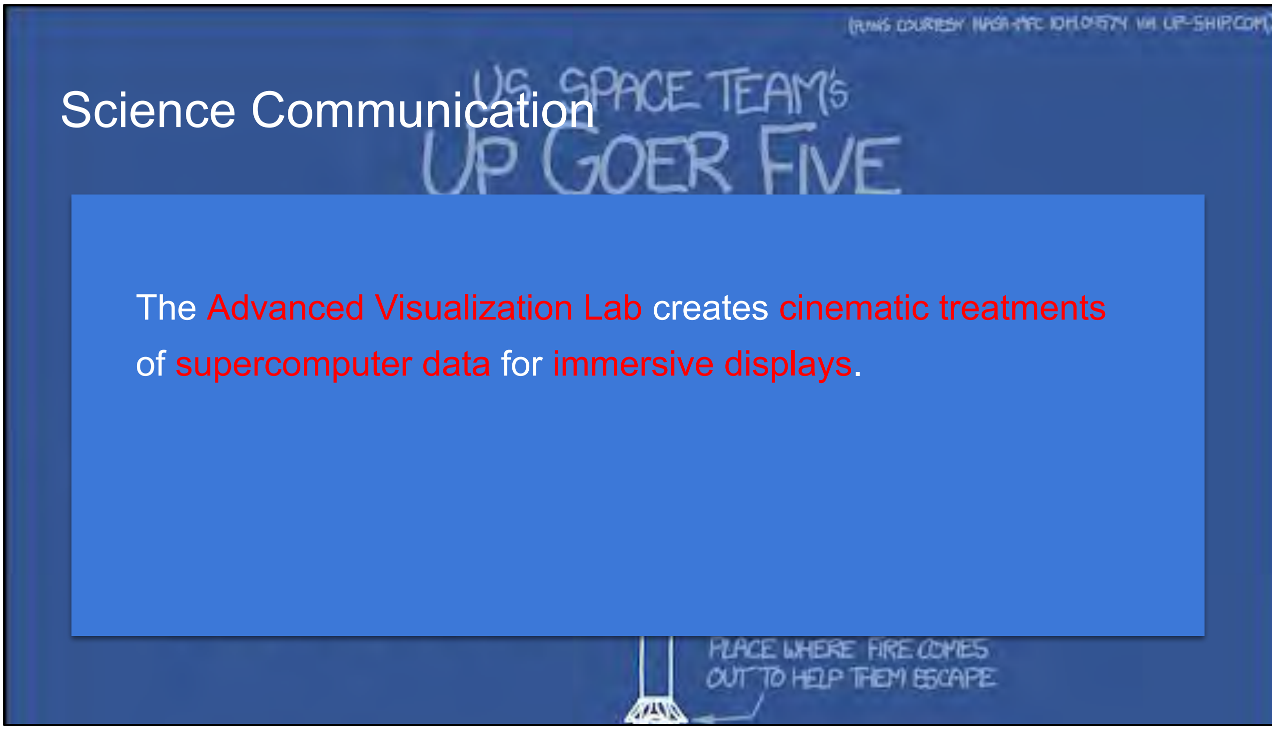
As with all science communication, the primary risk is of losing your audience with assumptions of their prior knowledge and use of science **jargon**.

This XKCD comic makes fun of the jargon of rocket science by dumbing down a diagram of the Saturn V rocket to descriptions using only the 1000 most common words in the English language.

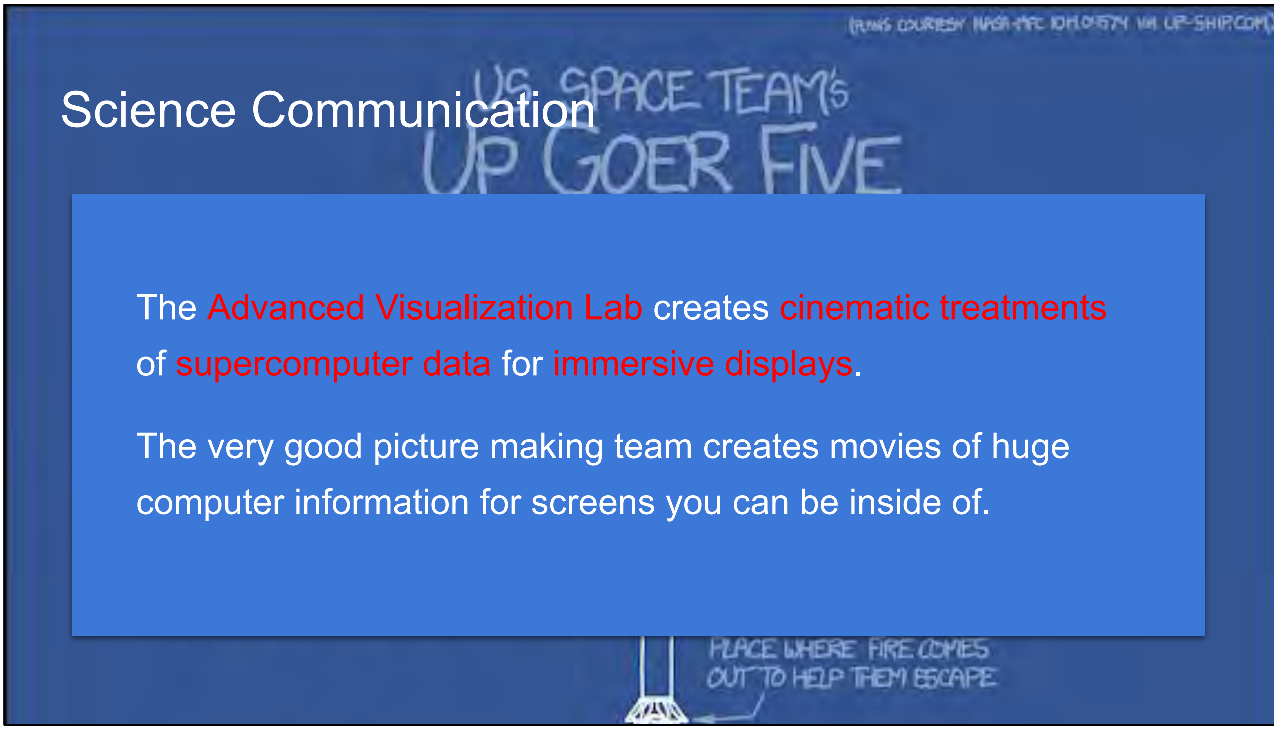


The reason other people do not understand our jargon is NOT because they are too stupid. It’s because their expertise lies elsewhere. As science communicators, we need to find a way to make a science message accessible to broad audiences without merely “dumbing it down”.

This is an example of the jargon I use to describe my own work.

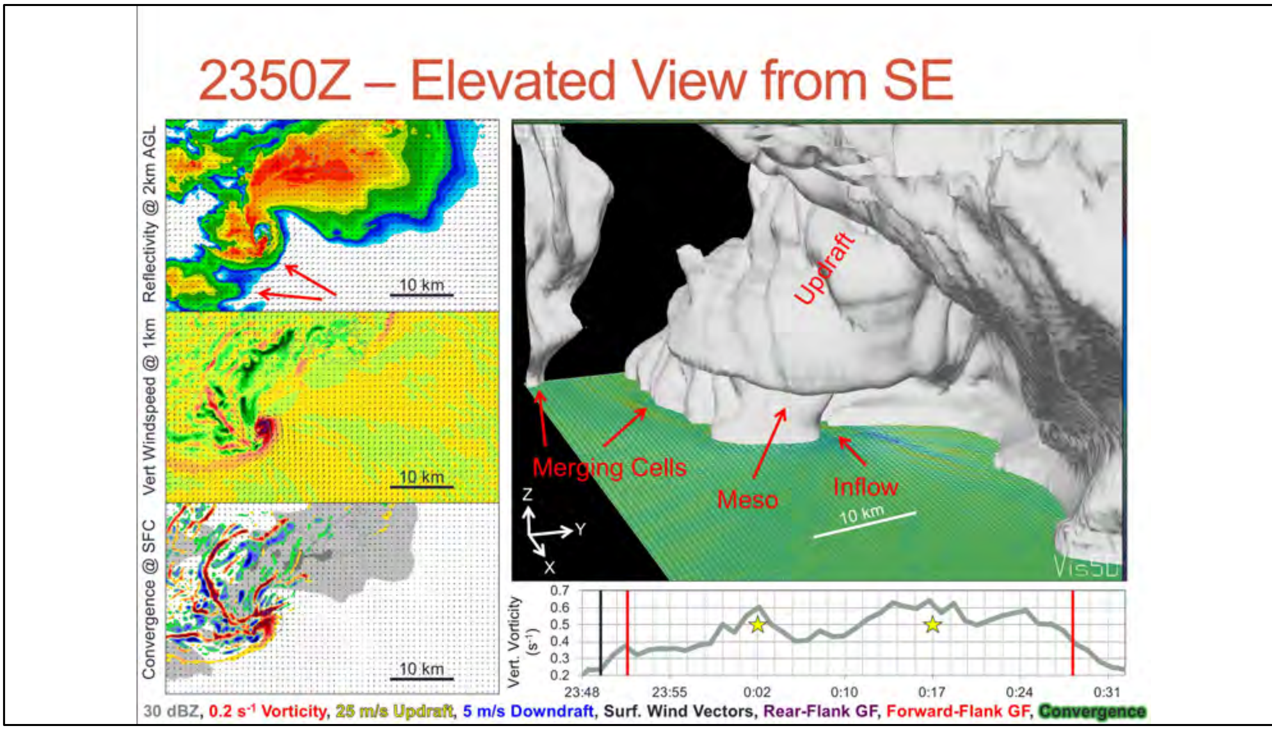


As you can see, most of these words are not one of the 1000 most common English words.



While this alternative seems almost laughably simple - it's undeniably a lot easier for people to understand.

There's no shame in using understandable language to help connect you with your audience!



In the same way, scientific imagery doesn't have to be wildly complex and covered in numbers and figures that your audience won't understand. This type of visualization is *just fine* for communicating with the right expert audience, but those of you in the audience that don't study tornadoes probably find this baffling.

Compare that to this visualization of the exact same data. Using cinematic techniques, we show the data in a more familiar way.

Chances are, decision-makers are more likely to be impressed if they understand the work that is being done.

- “I tried to put on my dress but Harper was in my way.”
- “We had to get in line and then we had to go on stage.”
- “I sang my song but I could see Robbie the whole time I was singing.”
- “I can’t believe Amber’s mom was backstage the whole time.”
- “Gordon wouldn’t go out when it was his turn because he was afraid.”
- “I think we should get ice cream. No cake. No ice cream AND cake.”
- “Harper is my best friend but she has to go home because her dad said it’s her bedtime.”
- “I put my dress on a hanger and Mrs. Chamberlain took it.”

So, let’s have a frank conversation about facts. Our minds are not wired to digest lists of facts. We don’t know how to process large amounts of information when none of it is more or less relevant or important.

Think about the way a young child tells you about their day. They tend to enumerate events seemingly at random. This type of exchange can be exhausting to follow.

- “I was worried about my dance performance tonight.”
- “At first my friend Harper was making me more nervous.”
- “But then the performance went great, and I deserve a reward!”

Our minds are wired to digest stories with conflict and resolution. It’s evolutionary - stories help us structure information about actions and consequences. They warn us of danger and help us learn from each other.

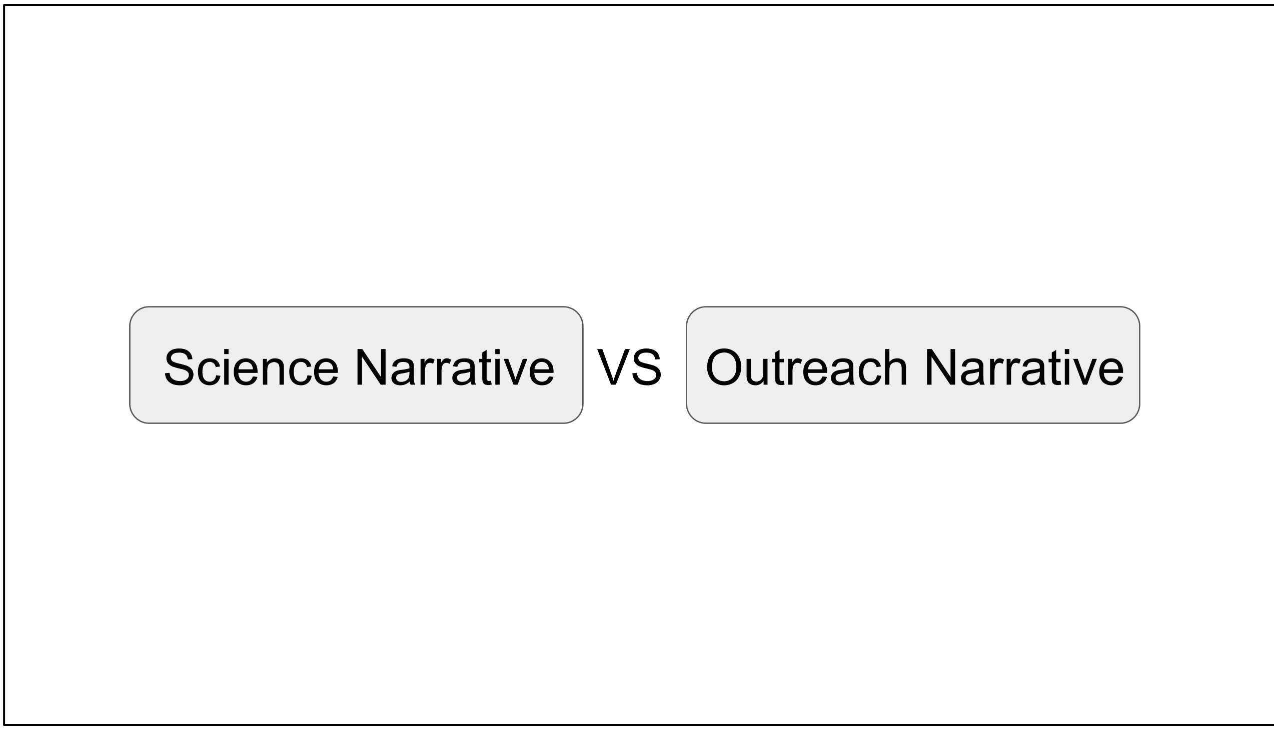
Moreover, stories provide context and generate empathy. Stories reach the emotional parts of our minds rather than just the logical parts.

Visual story

The imagery of a good visualization should tell a story even without text or narration.

These two photographs are both ostensibly of the US Capitol building, but one tells a story. Stories thrive on conflict and characters.

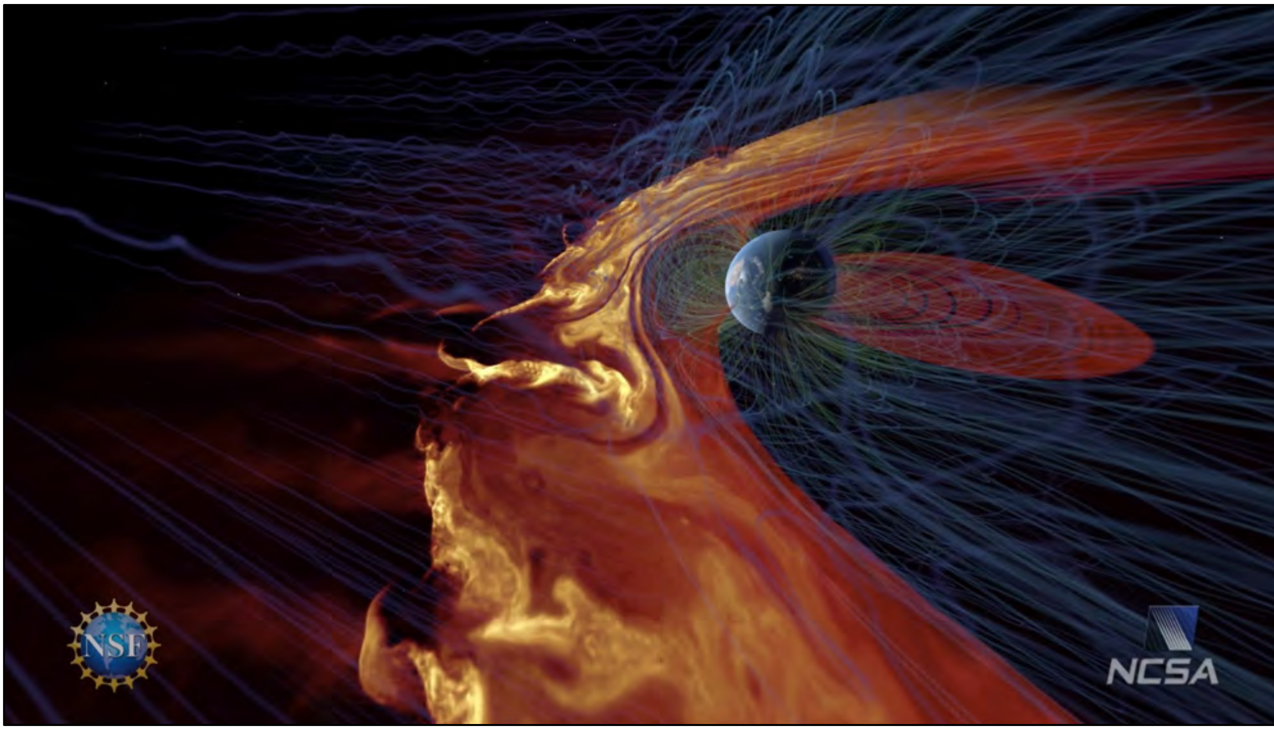
https://www.flickr.com/photos/mobili/32593123745
https://commons.wikimedia.org/wiki/File:United_States_Capitol_Building.jpg



Be aware that especially when working with someone else's data that you may face a dichotomy between the data's science narrative, and your own outreach narrative.

The **science narrative** is the story of the research that created the data - what question was the researcher asking and how does this data get us closer to answering it?

The **outreach narrative** is the story you are trying to tell to promote some broader understanding of science. This story might not even be focusing on the field of the researcher.

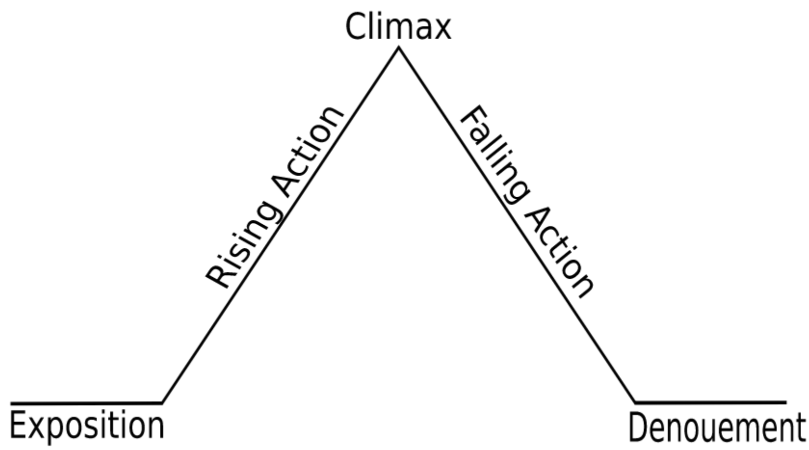


For instance, the science narrative of this visualization was about how solar plasma affects the Earth's magnetic field. It was focused on how the turbulence of the magnetic field causes it to break down. Originally, the data didn't even have the Earth in it, because it wasn't concerned about effects at the scale of the Earth.

But the outreach narrative was about how a burst of solar plasma could be hazardous to life on Earth. You can see in the visualization that all movement is centered around the Earth - which we added to the scene - and we kept the turbulent swirls of the magnetic field as a dynamic environment.

When designing a visualization that suits your outreach narrative, you always want to be honest to the science narrative as well. It's a bad idea to extrapolate from the data to suit your story's narrative - it comes across as dishonest and can risk the trust audiences place in the rest of your work. Communicate and share your progress with the scientist and other experts frequently.

Freytag's Pyramid



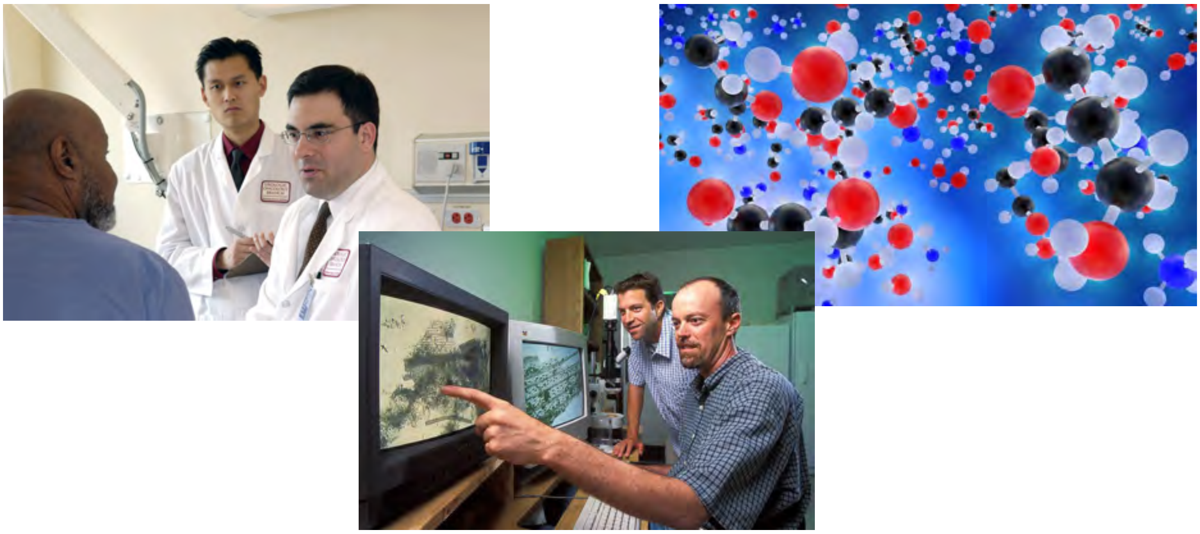
In order to tell a story well, it's helpful to understand the common traits in all stories. While finding new ways to use old storytelling techniques makes the stories feel original, we continue to use these same techniques that we have been using for millenia because they appeal to our common human psychology.

In short, we can transform a "list of events" into a "story" by introducing **characters** and **conflict**.

Freytag's pyramid is a common device used to illustrate the way that conflict rises and falls in all compelling stories.

https://commons.wikimedia.org/wiki/File:Freytags_pyramid.svg

Conflict in Science



Good stories always center around characters with relatable human emotions.
But how do we identify those emotions in a story about science?? Science is supposed to be objective, right?

1. The story is about direct recipients of the research - perhaps about political statistics or how a disease affects a population or how an animal population is being conserved
2. The story is about a scientist's process of discovery
3. The story anthropomorphizes abstract objects like galaxies and molecules

https://commons.wikimedia.org/wiki/File:Doctor_consults_with_patient_(4).jpg
http://www.freestockphotos.biz/stockphoto/10089

Education

- Learning theories
  - Behaviorism, Cognitivism, Constructivism
- Edu-tainment
- Informal Science Education
  - Agency - ability to form questions and find own answers

Learning Theories

1. Behaviorism

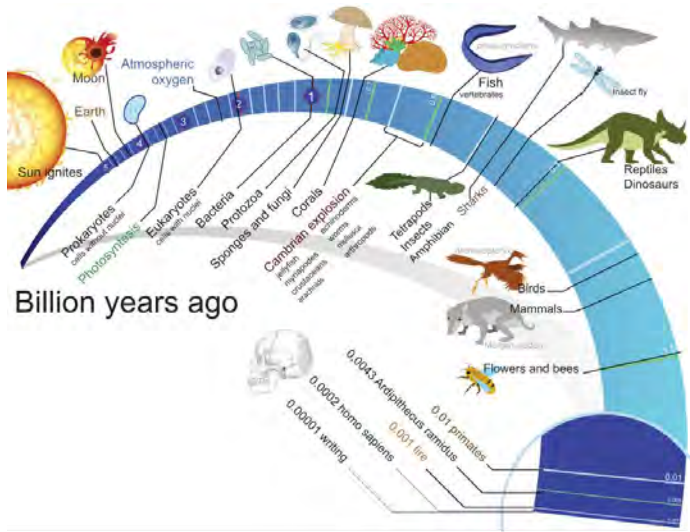
The early 20th century was dominated by a learning theory called “**behaviorism**”. The idea was that learning is a direct result of rewards - like good grades - and punishments - like detention.

While this works for simple memorization tasks, it’s not usually associated with higher-order critical thinking, so for audiences of visualization it’s largely ignored.

https://pixabay.com/illustrations/disobedience-mother-and-son-boy-1673196/

Learning Theories

1. Behaviorism
2. Cognitivism



Several decades later, "**cognitivism**" became a popular psychological model for how people process information. Learning theorists interpreted the cognitivism model to mean that learning is a result of organized, sequenced, and understandable presentation of facts.

This philosophy suggests that a sequence of visualizations that are sorted by size or chronologically might resonate well with people.

https://commons.wikimedia.org/wiki/File:Timeline_evolution_of_life.svg

Learning Theories

1. Behaviorism
2. Cognitivism
3. Constructivism

One of the learning theories that has become popular more recently is called "**constructivism**". Constructivism claims that learning is a result of building knowledge through interaction with the physical world..

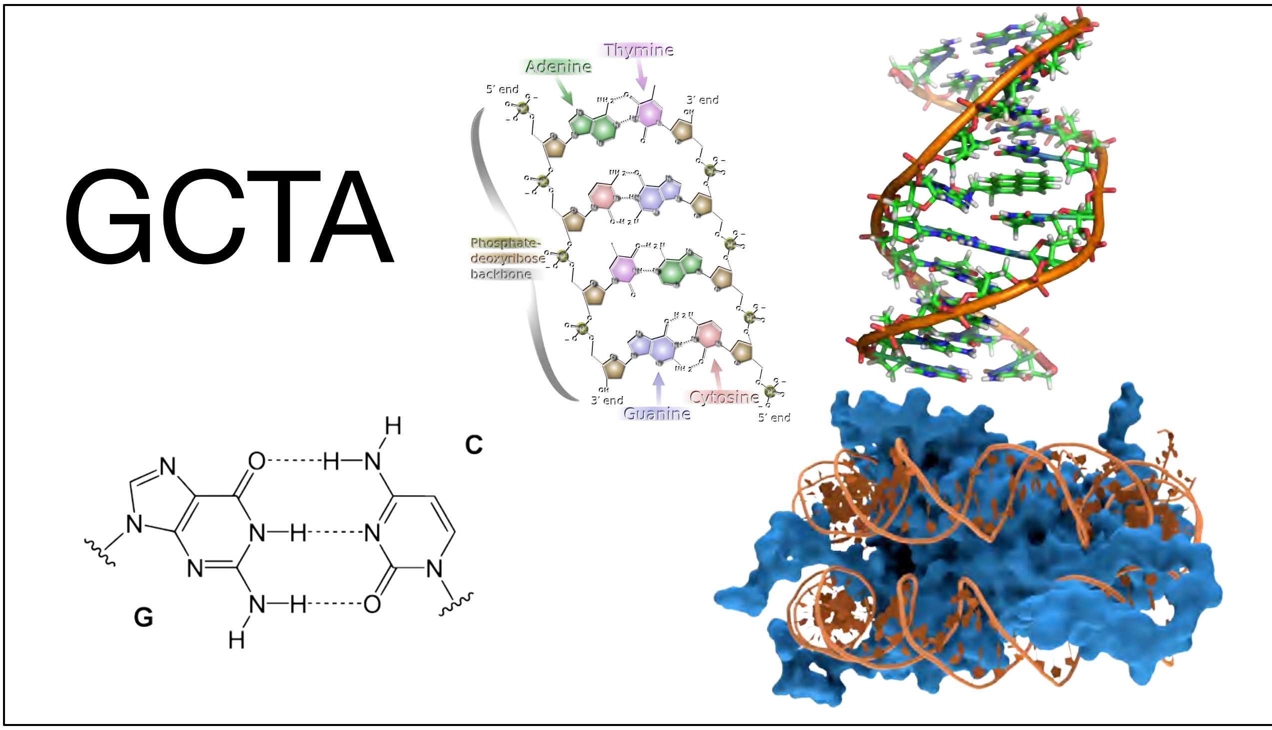
Constructivism can be applied to visual information through what experts call "metavisual capability" - that is, our ability to build a mental image from previous mental models.

You can see this theory at work in the modern school system teaching scientific concepts - like DNA.

https://pixabay.com/photos/legoland-building-blocks-legos-lego-172296/

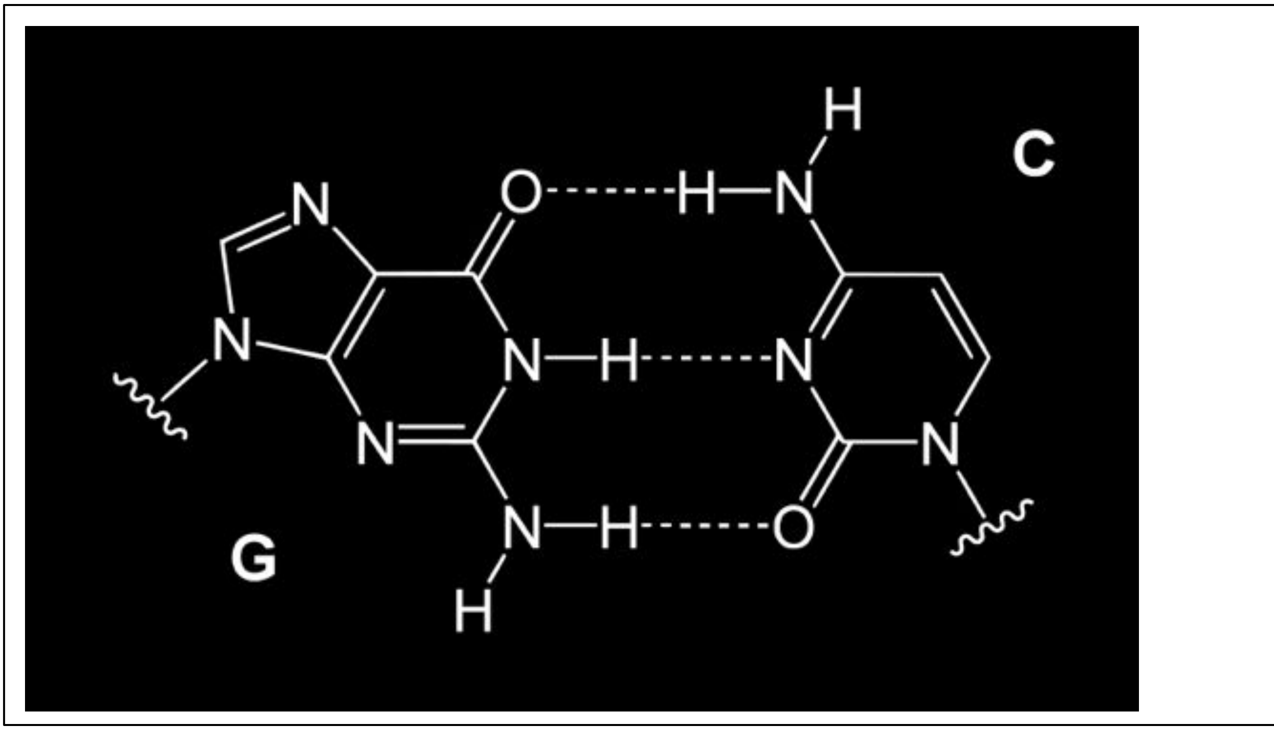
$$F = ma$$

Communicating science at all levels of education depends on abstraction. Abstractions come in many shapes and levels of simplicity. Some abstractions are strong because they rely on established language that is accumulated in students' minds through years of formal education. For instance, this equivalence of Newton's Second Law of Motion can be abstracted to symbols which in turn represent the English words "force", "mass", and "acceleration". Using symbols like these, physics students can begin to grasp many of the rules of the natural universe.

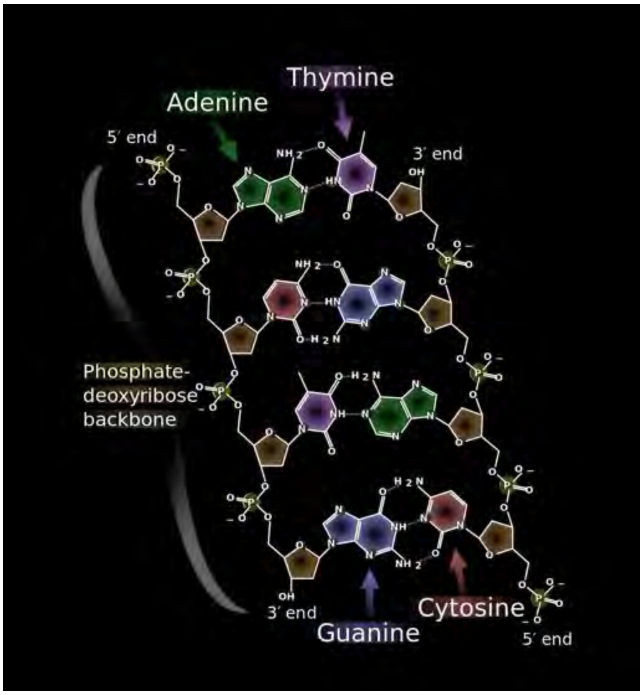


Educational literature refers to "metavisual capability" - the ability to understand a visual abstraction of science, and store that as a mental model that can be built upon and used to extrapolate to other domains.

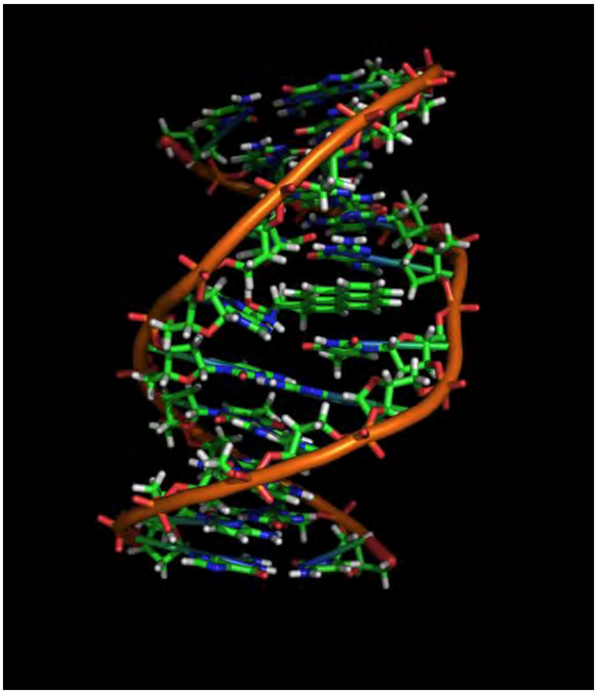
As young students, we start with an extreme abstraction of DNA that gives us a very limited, manageable amount of information: there are four DNA base pairs that are the building blocks of all life. We have no idea what shapes they are or that these shapes are called molecules. The base pairs start in our mental model of DNA as four of the letters we've known since we first saw a spelling toy or a book. The base pairs are four different objects in the same category, and that's about it.



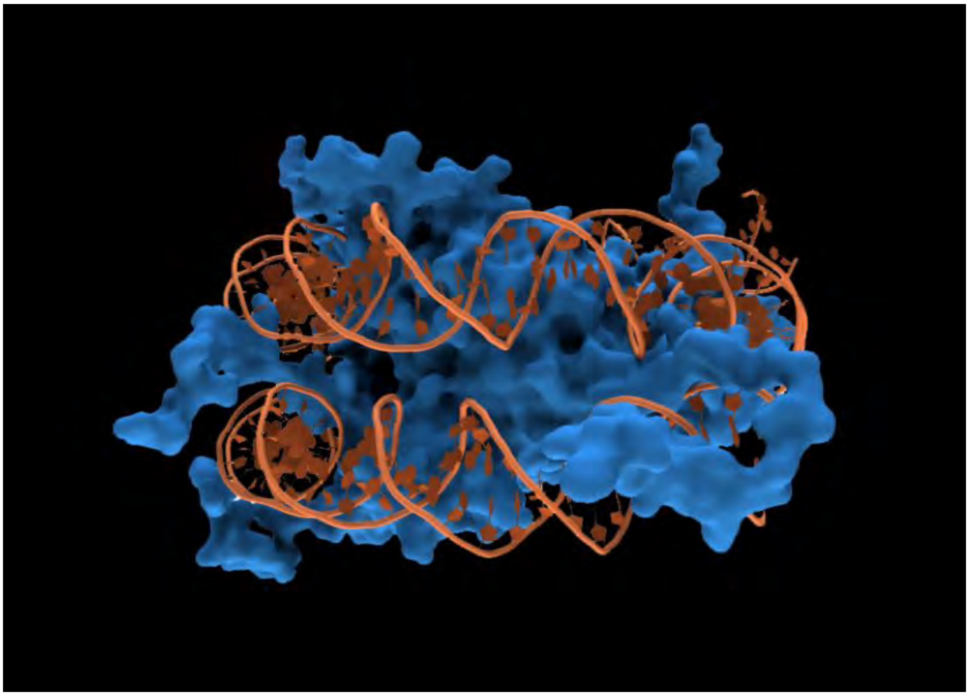
Through activities that allow us to manipulate those four letters, we are eventually introduced to the full molecules guanine, cytosine, thymine, and adenine. We draw them in neat straight-lined molecular diagrams. Our metavisual model of DNA grows from letters to line drawings.



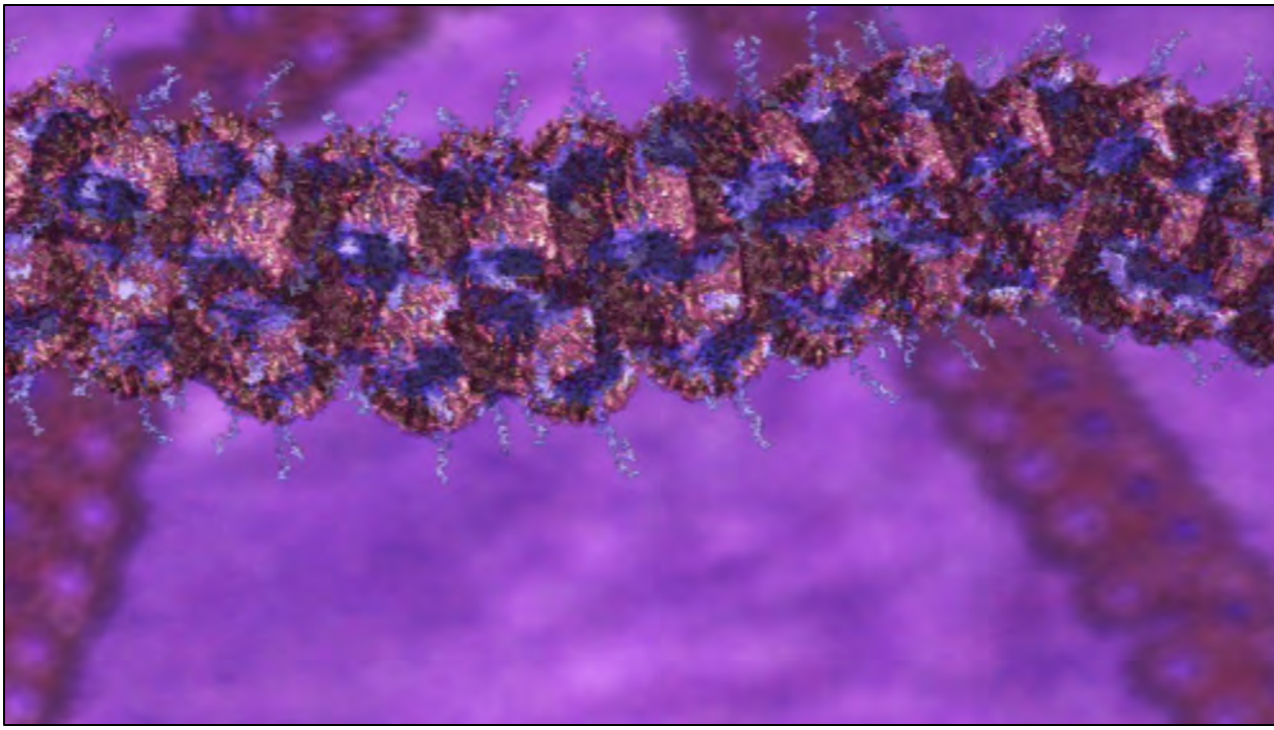
Again through interaction with imagery in books or on computers, we progress to connecting our base pairs into a ladder and start to get some idea of the context in which the base pairs reside. We learn that each of these base pairs actually relies on a phosphate backbone to hold the structure together.



Most students eventually construct the “double helix” model of DNA in their mind’s eye. They begin to understand the organic complexity of DNA in the real world, witnessing how base pairs twist around each other. As this visual description of DNA becomes more complex, letters are replaced by colors and more physically accurate shapes.



Most people don’t get the opportunity to further build their metavisual model of DNA in the classroom. But visualization can give them a chance to interact with new visual information. Here we can see an orange DNA double-helix wrapped around a protein in its natural environment. This image encourages audiences to reference multiple mental models - and not only our pre-existing images of the DNA double helix. As visualizations look more three-dimensional and tangible, it’s easier to equate them to real world experiences



And in this animated visualization, we actually see DNA in motion wrapping itself around those proteins, which brings the imagery even closer to our real world experience.

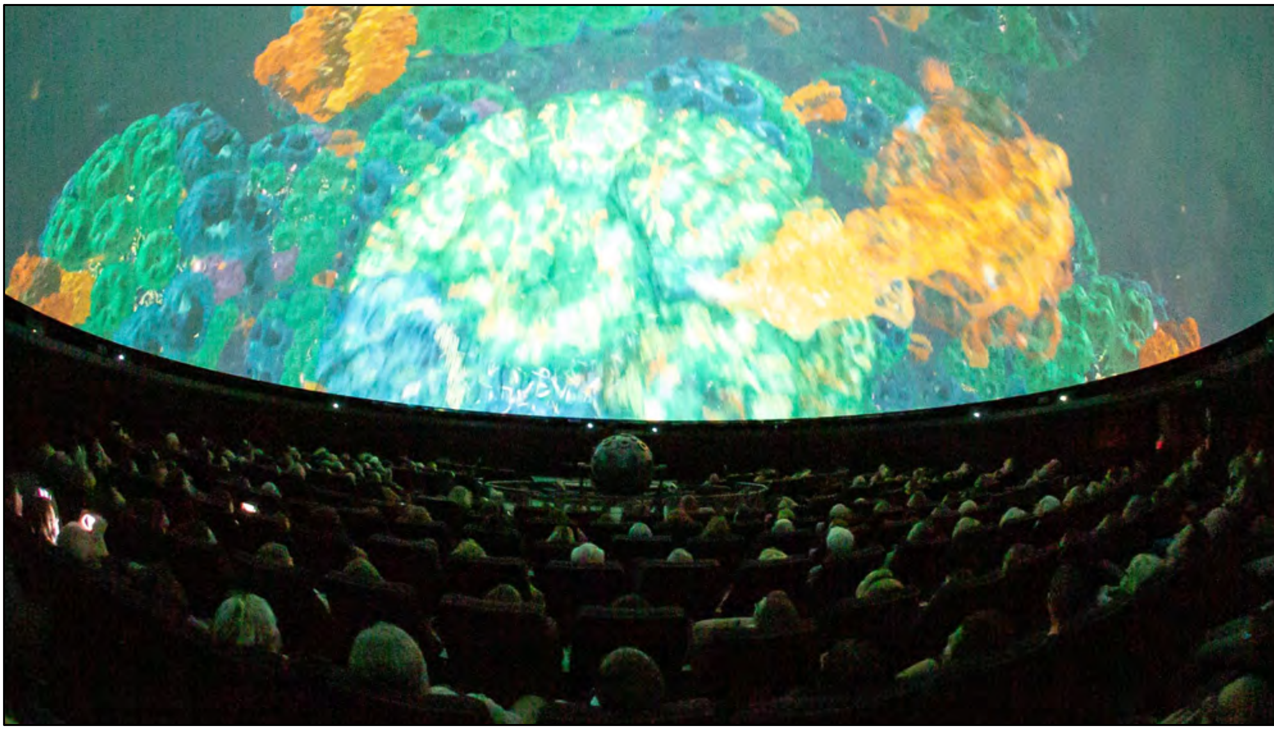
When we start talking about visualization in motion, there's another mental model we can easily relate to - the films of Hollywood. Even children with no formal schooling are familiar with movies and their visual language. Movies' use of framing, movement, and color are ingrained into most modern cultures.

And at the end, this visualization helps us make a connection to another familiar image - the butterfly shape of a chromosome. The more these mental models interact, the more we learn.

It bears mentioning that conveying science in the style of Hollywood films runs the risk of being too much fun and obscuring the scientific message. Years ago we collaborated on a show about an alien traveling the universe visiting astronomical events. The show received criticism for not properly separating the science from the science fiction. Audiences must be able to understand what the line between fact and fiction is.

Educational gaming researchers describe the phenomenon of "chocolate covered broccoli" - the idea that when you try to disguise an educational experience as a game, players will quickly discover they've been deceived. Because of this, they are more likely to dislike or abandon the game. In the world of EDU-TAINMENT, audiences actually tend to prefer fact over fiction.

[Note: I'm trying to get permission to use this image: https://vh26s5v0qdz9bkd1-zippykid.netdna-ssl.com/wp-content/uploads/2012/04/IMG_6101.jpg]



Ultimately, educational experiences should aim to develop a person's **agency.** Agency is someone's ability to form questions about a subject and find their own answers to those questions.

Teaching science through movies and video games is part of the process of building agency. It works differently than formal schooling in a classroom, and so we refer to it as **informal science education.** The primary goal of informal science education is not for students to memorize facts and figures - it is to build someone's interest in a subject, and help send them on their own journey of discovery.

Research actually shows that students learn more in their formal education if they are first primed with an informal educational experience.

Just because the audience isn’t memorizing facts, that doesn’t mean they aren’t inspired or influenced by the facts. It’s extremely important to present factually accurate information in a science education experience.

Verifying that a visualization is factually accurate can be a challenge, but it is well worth the effort. It establishes you as a reliable source of information and attracts a broader audience.

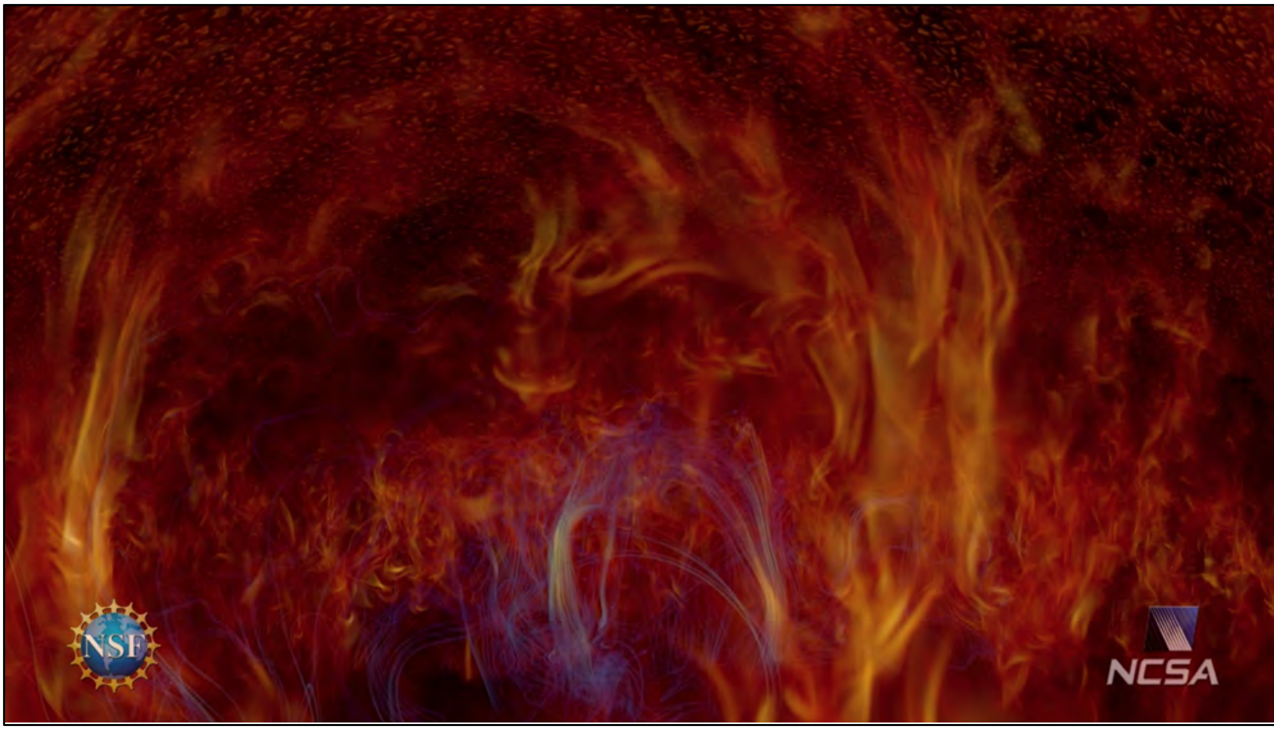
- Communicate with Domain experts
- Communicate with computer graphics experts
- Survey your audience
- Do side-by-side comparisons for different audiences

Color

- Psychology of Color
  - Different colors evoke different emotions
  - Colors have different meanings across different cultures (Himba color circle)
  - Physically real colors
  - Color can present bias (think skin tones)

In the same way that animation studios tightly control the color script of their films, color is extremely meaningful and controllable in visualization.

Colors don’t have objective meaning, but context can sway the feelings your audience gets from them. Color is cultural - it has different meanings in different times and places.



This visualization shows convective solar plasma just beneath the surface of our sun. Does the color design of this environment make it feel more romantic or hellish? Try to analyze what you feel when you watch it. When I watch this, I feel like I'm in an underground cave, surrounded by fire.

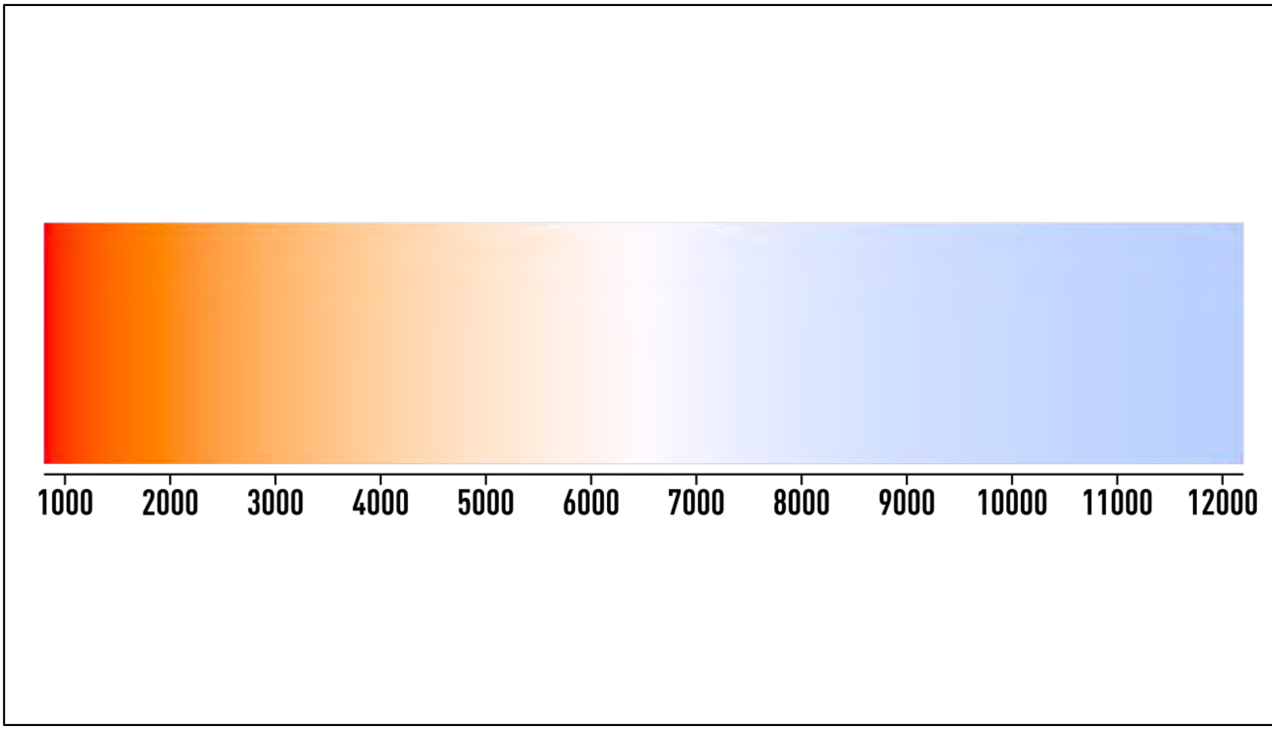
Any analysis of color in society will inevitably touch on politics and history.

For instance, why is it that we refer to skin colors as white and black when there are much more accurate words - like brown, tan, and peach. It is no mistake that colonial cultures equated fair skin with the color white, which is often associated with purity, and dark skin with the color black which is more often associated with fear. This use of color psychology has influenced racial attitudes for centuries.

Understanding the cultural implications of your color choices can help you to avoid confusing - and possibly even offending - your audience.

https://commons.wikimedia.org/wiki/File:Skin_color.png

Temperature is an example of cultural confusion you might encounter more frequently in the context of science. Most people encounter temperature indicators in their home everyday - on sink faucets or stovetops - and these label low temperatures with the color blue and high temperatures with the color red. Ice is blue, and fire is red, so this makes intuitive sense.

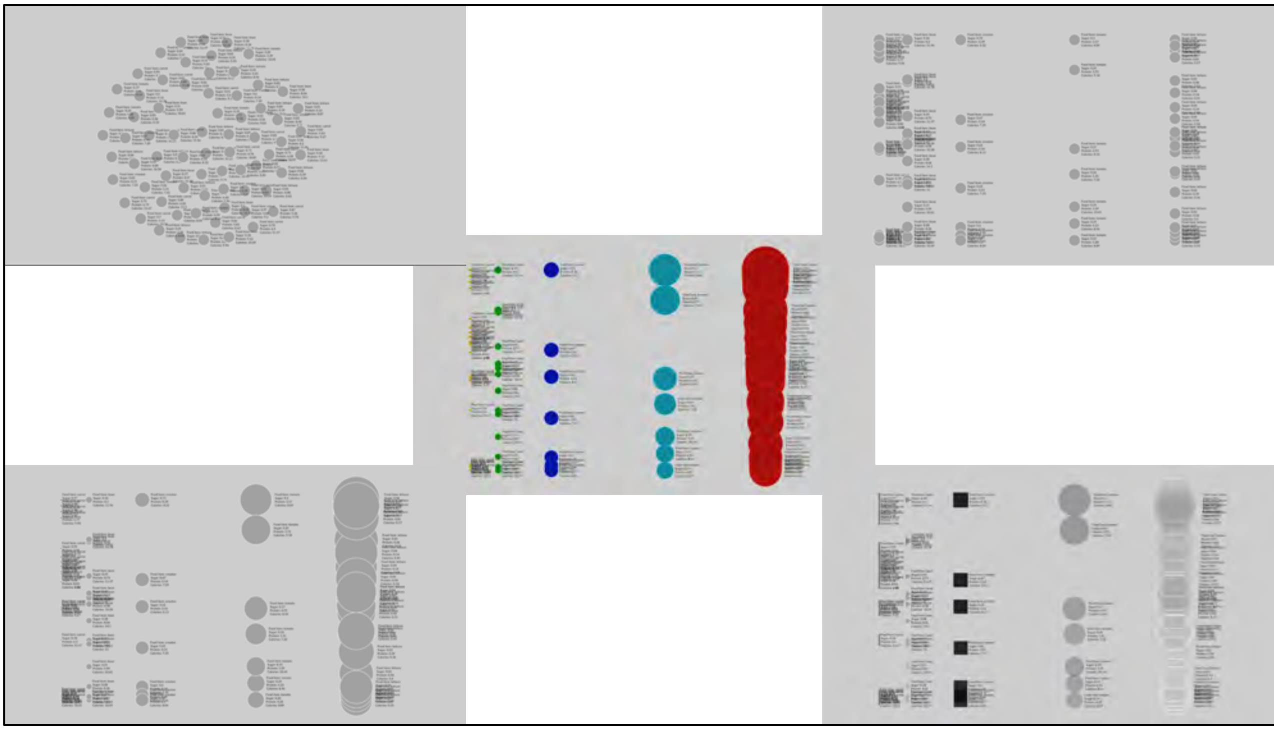


However, this common use of color temperature becomes problematic when we start mapping color to extremely hot astronomical objects.

Stars are one kind of black body object, which emits color according to this color chart - as you can see, the red and blue are reversed from your faucet at home. As black body objects get hotter, they become less red and more blue.

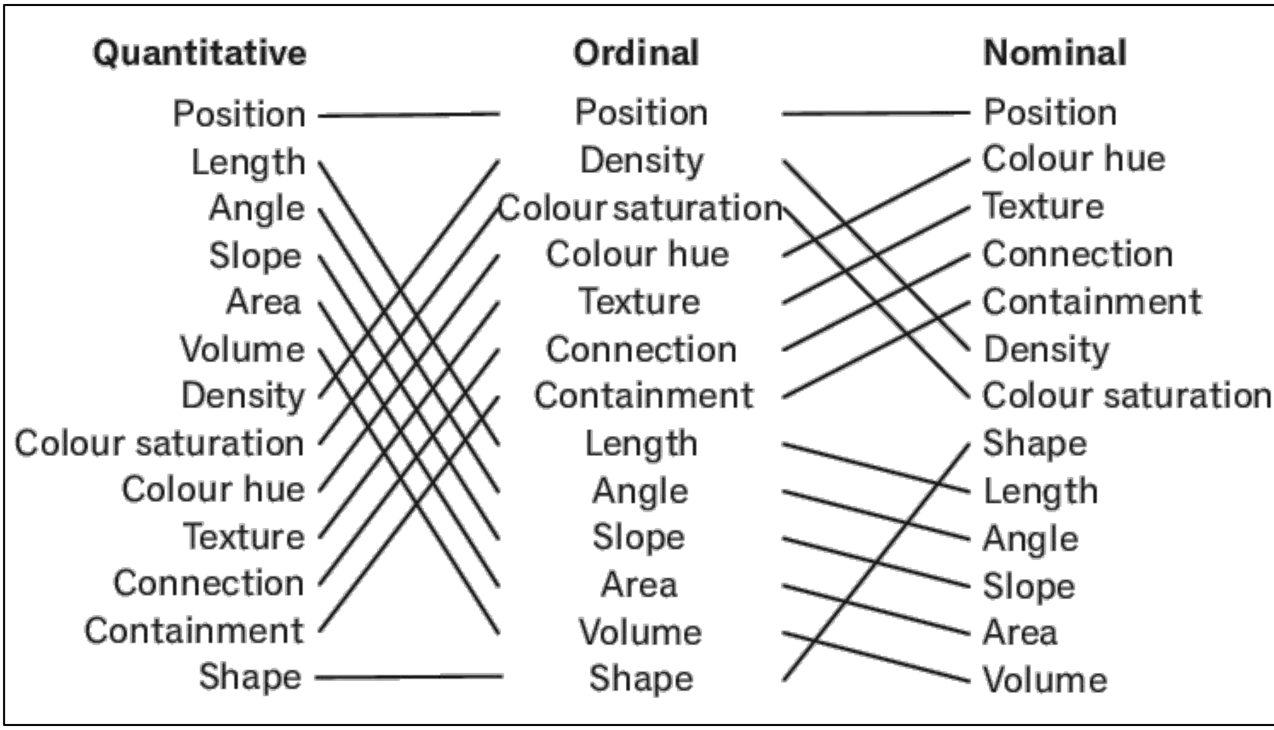
https://commons.wikimedia.org/wiki/File:Color_temperature_black_body_800-12200K.svg

Driving Visual Features With Data



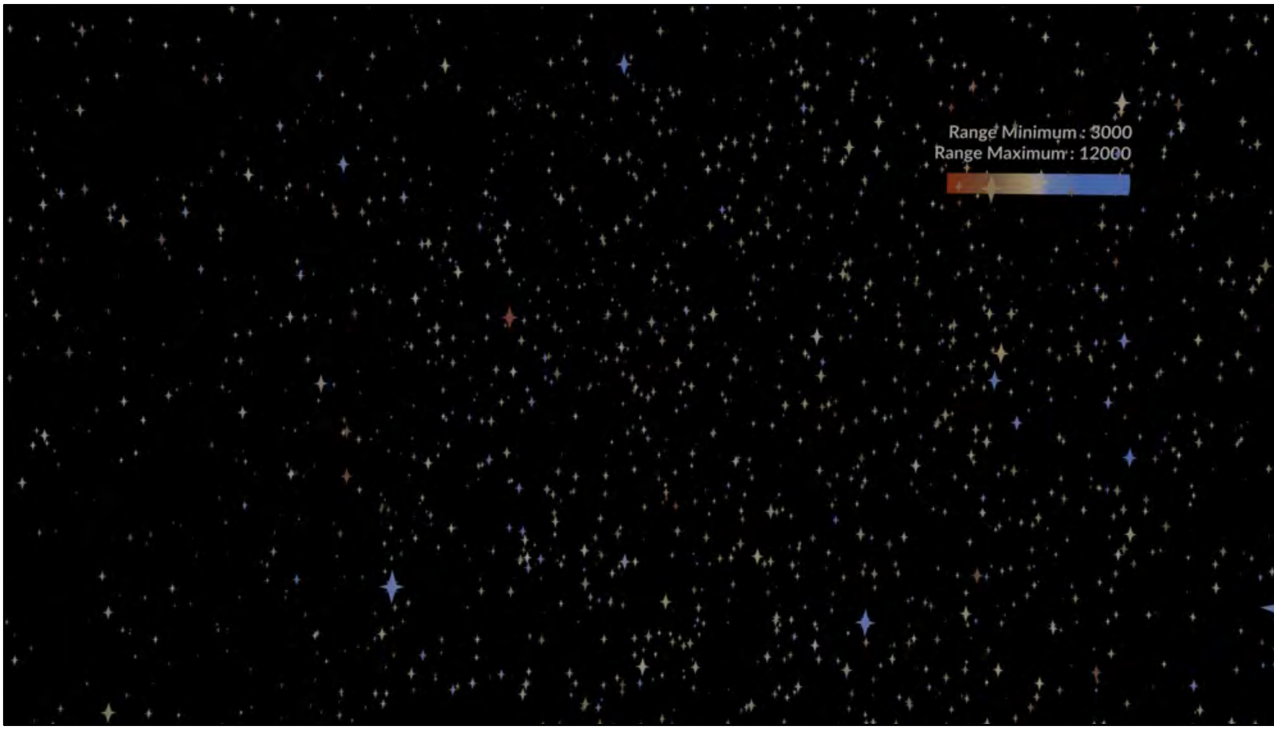
One of the neat things about using images to communicate data is that you can encode information in many ways other than labels and text. Visual objects can have lots of visual attributes, sometimes called “visual dimensions” that visualization designers get to play with.

We can change objects in the scene to have data-driven positions, or sizes, or colors, or shapes. In animated video we can drive the motion of objects with data as well.



Not all image dimensions are created equal. Position and size are directly comparable measurements - one value is objectively higher or larger than another.

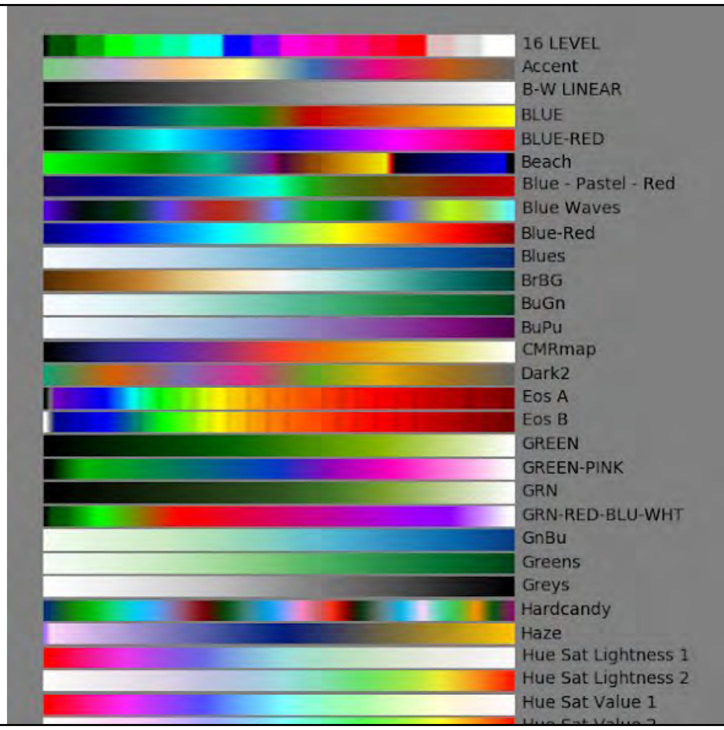
A dimension like color is more categorical though. You can recognize that one color is different from another - and one color might look similar to another - but different hues don't look bigger or smaller than each other.



A major part of scientific visualization is designing **transfer functions**. These are algorithms that translate your data into your visual dimensions. Transfer functions always include a **ramp -** which can either be a spline ramp for scalar dimensions, or a color map for vector dimensions. They also always include a **range**, which determines the minimum and maximum data values affected by the ramp.

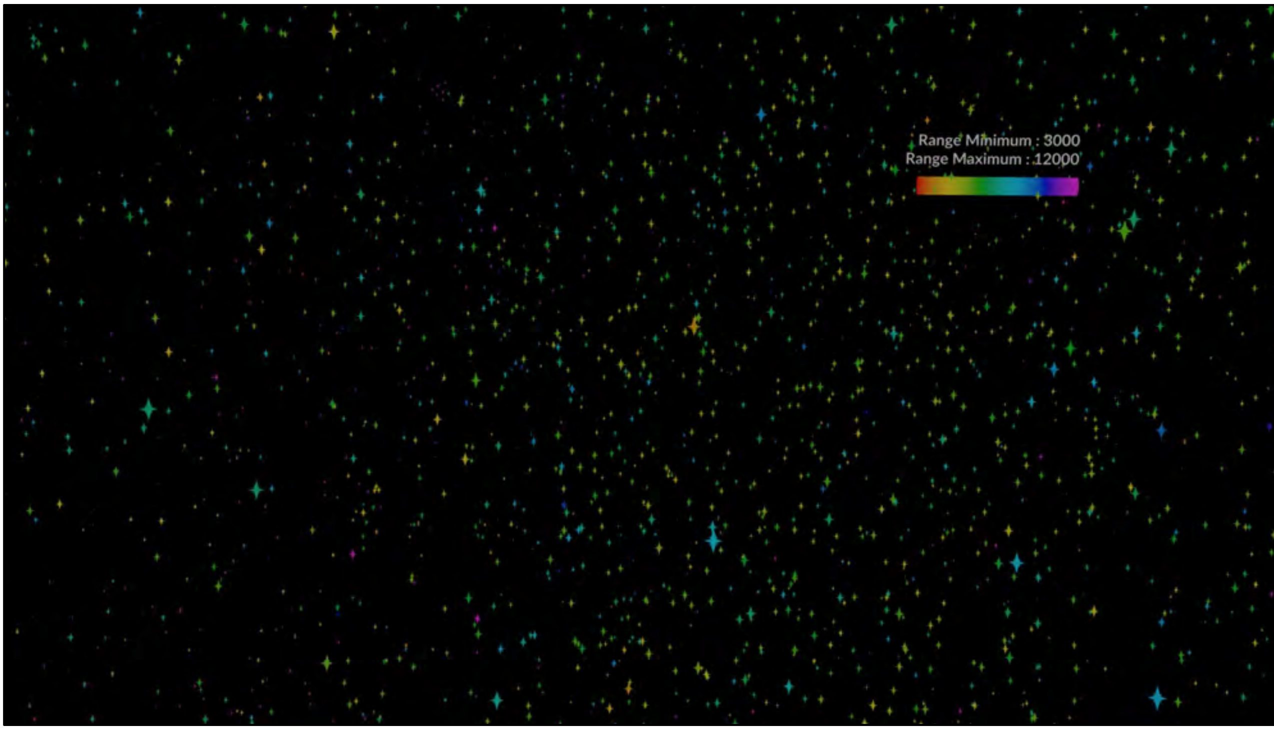
Transfer functions can include additional math, such as a logarithm operation to get better control of data that crosses many orders of scale.

The most common type of transfer function is a color transfer function, which drives the color of your computer graphics with a “field” or “attribute” in the data. This dataset of our local stellar neighborhood drives the color of the stars using their temperature in Kelvin using the blackbody color map I mentioned earlier.

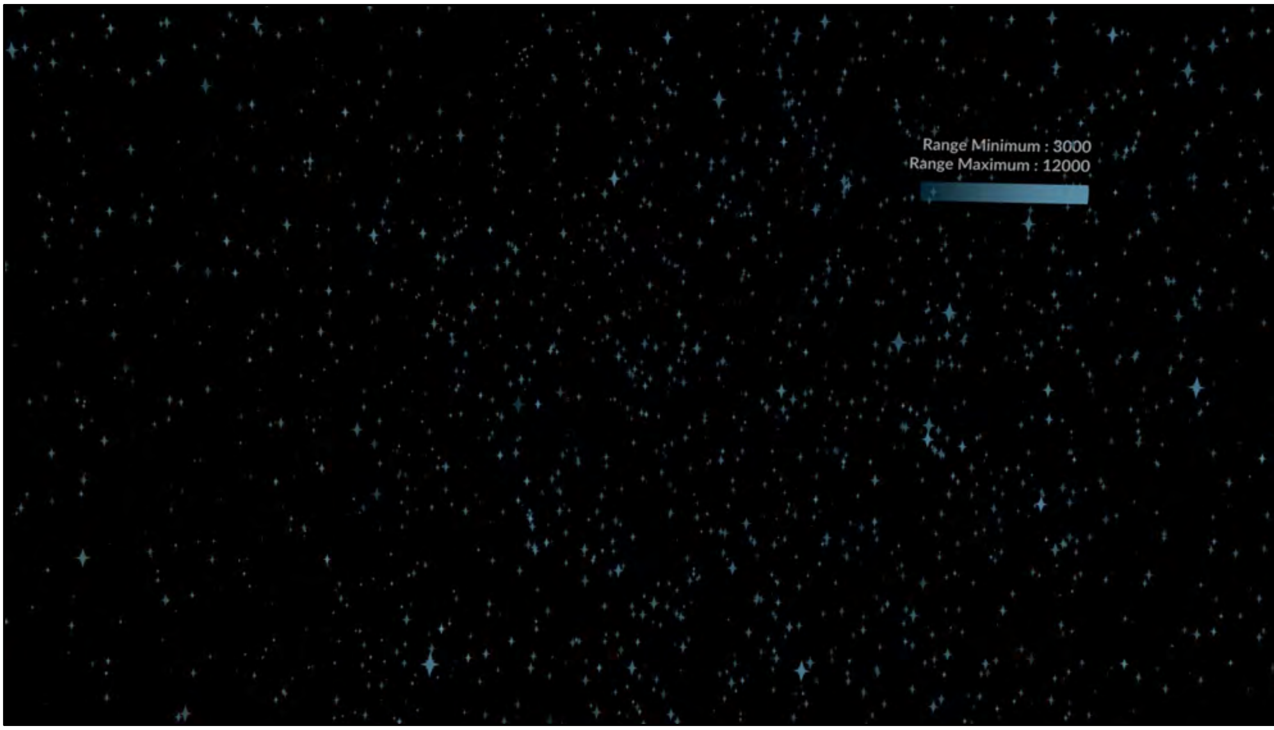


Many visualization tools come with color ramp presets with names like “cool”, “coolwarm”, “pastel” and “doom”. Some of these presets are better and worse for certain purposes.

http://yt-project.org/doc/_images/all_colormaps.png



A common piece of wisdom is that the rainbow color map is usually a poor choice - because the rainbow colors cycle, there’s no clear high or low end.



Because the human eye is more sensitive to brightness than hue, it’s usually recommended to have a consistent change in brightness from one end of the color ramp to the other as well.

Even the most photorealistic science imagery you're used to seeing is not showing REAL colors. Take for instance this Hubble image of the Orion nebula. Its beautiful pinks and blues work in perfect harmony. But actually this would all look like a wall of green to the human eye. This is because the human eye is less sensitive to visible color than the Hubble telescope.

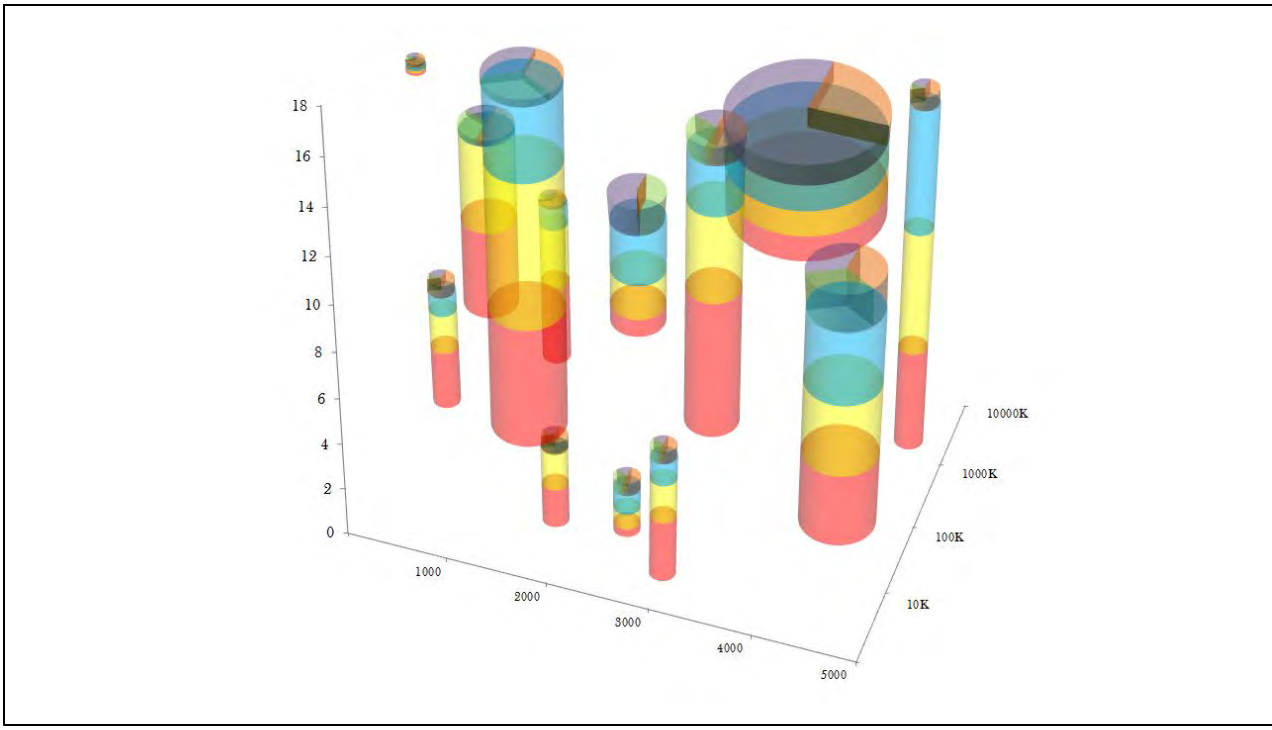
https://svs.gsfc.nasa.gov/30947

But actually this would all look like a wall of green to the human eye. This is because the human eye is less sensitive to visible color than the Hubble telescope.

http://wise.ssl.berkeley.edu/gallery_OrionNebula.html

The lovely pink and blue colors were chosen to emphasize features, to inspire audiences, and to show us things we wouldn't see with our own eyes. After all, why should we stick to what the human eye would see when we have the technology to show something more clearly than the human eye would see it?

https://svs.gsfc.nasa.gov/30947



One last thing to note about driving image dimensions with data: Trying to pack as much information as possible into every dimension of the image can make a visualization overwhelming, especially to an untrained audience.

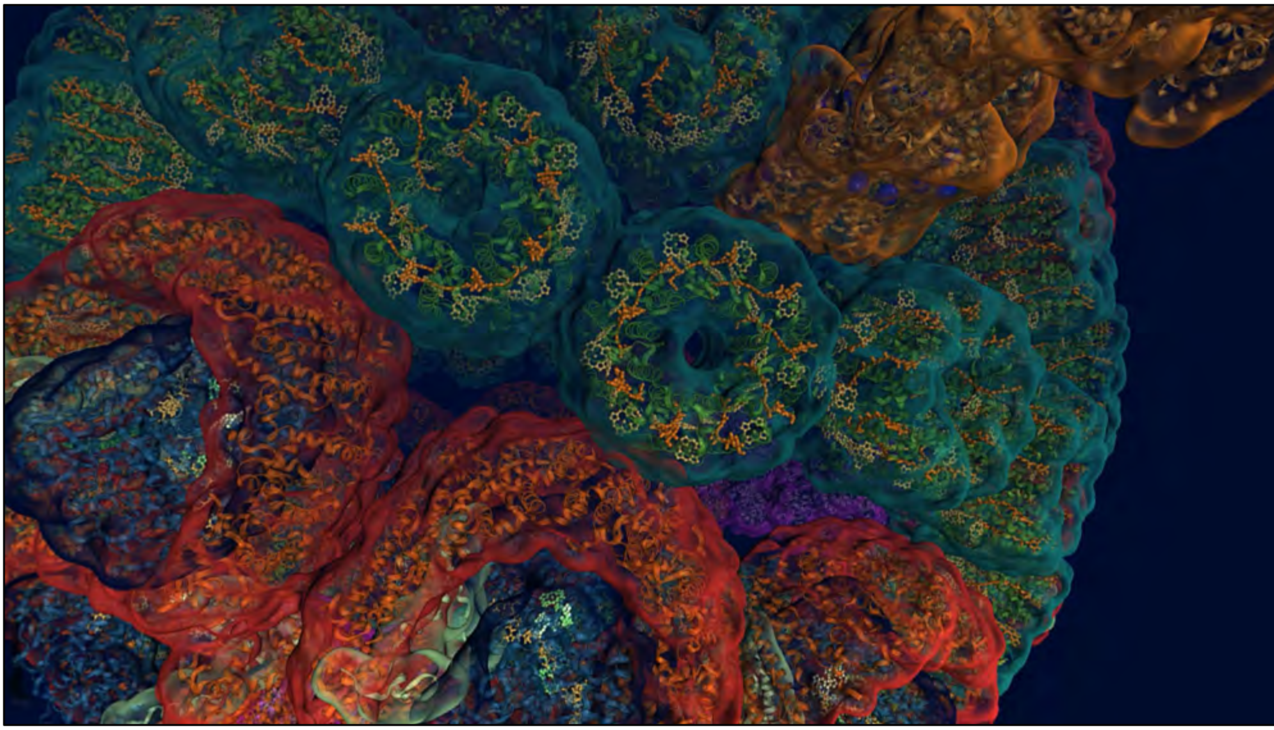
For instance, this visualization has so much going on that it's difficult to interpret. There's color, height, 3-dimensional position, shape, surface area, volume - All of these are making it very difficult to understand the point.

Especially for outreach visualization, repetition can help your audience. There's no shame in driving multiple image dimensions with the same data field.

https://www.reddit.com/r/dataisugly/comments/8msftx/the_marvel_that_is_3d_stacked_scatter_pie_columns/

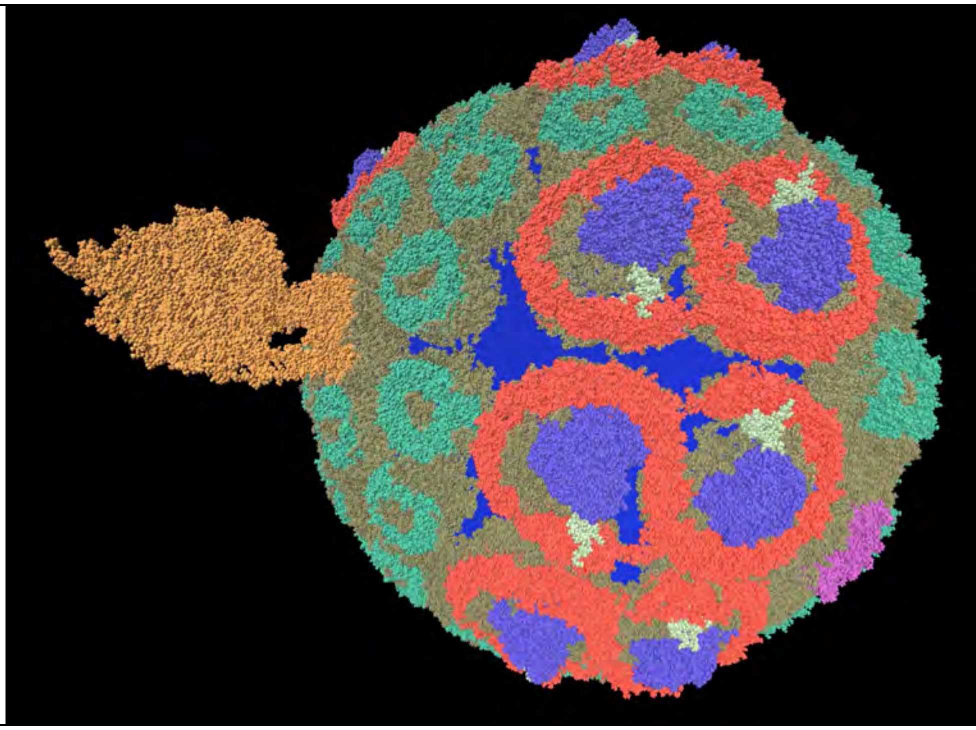
Abstraction
and Representation

So now I want to transition to talking about how we actually design our visualizations. As visualization designers, we need to make choices about what details to show, and how to show them. Choosing not to show everything all at once, to simplify, and to emphasize features is called "abstraction" and it is a key skill for any visualization designer.

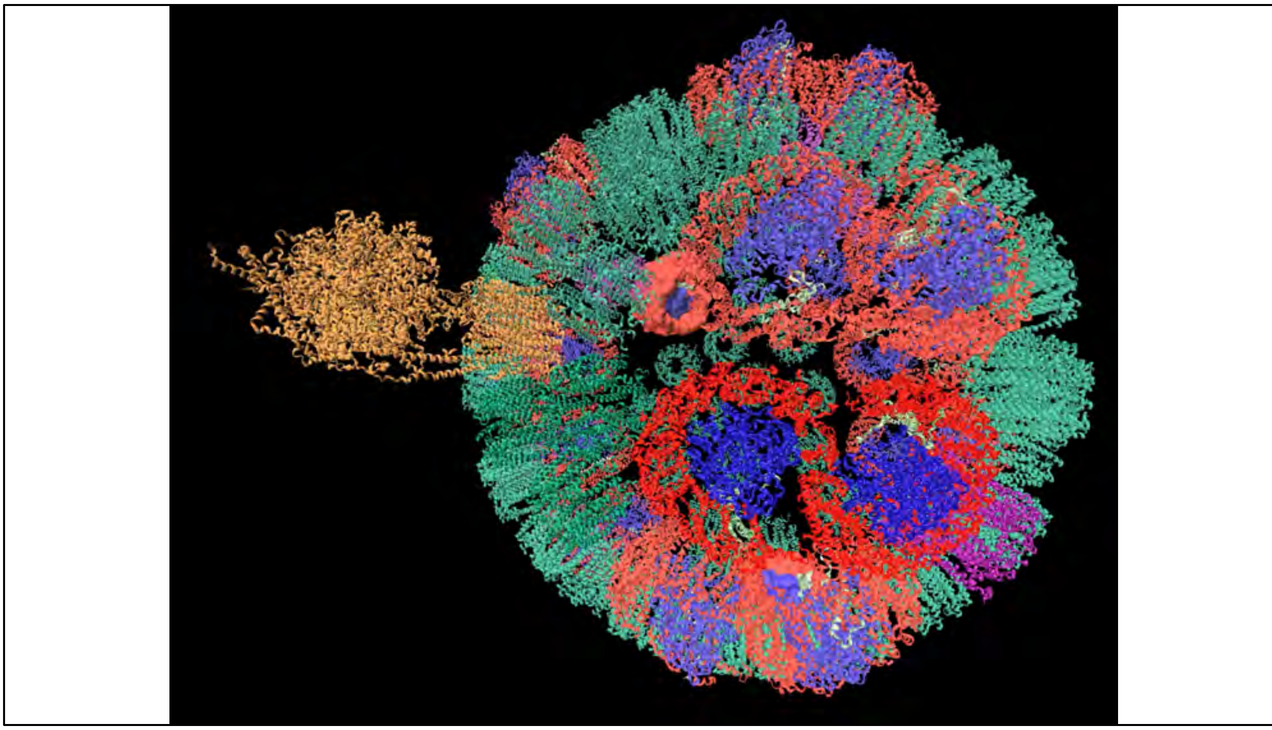


In visualization we have to lie to tell the truth. That is to say, realism isn't always the best way to communicate the complexity of your data.

In this biological visualization, we see hundreds of ribbons floating through the structure of the proteins in this organelle. These ribbons are not a realistic representation of shapes that actually appear - they are technically a lie - but they are a *meaningful* simplification of strands of DNA.

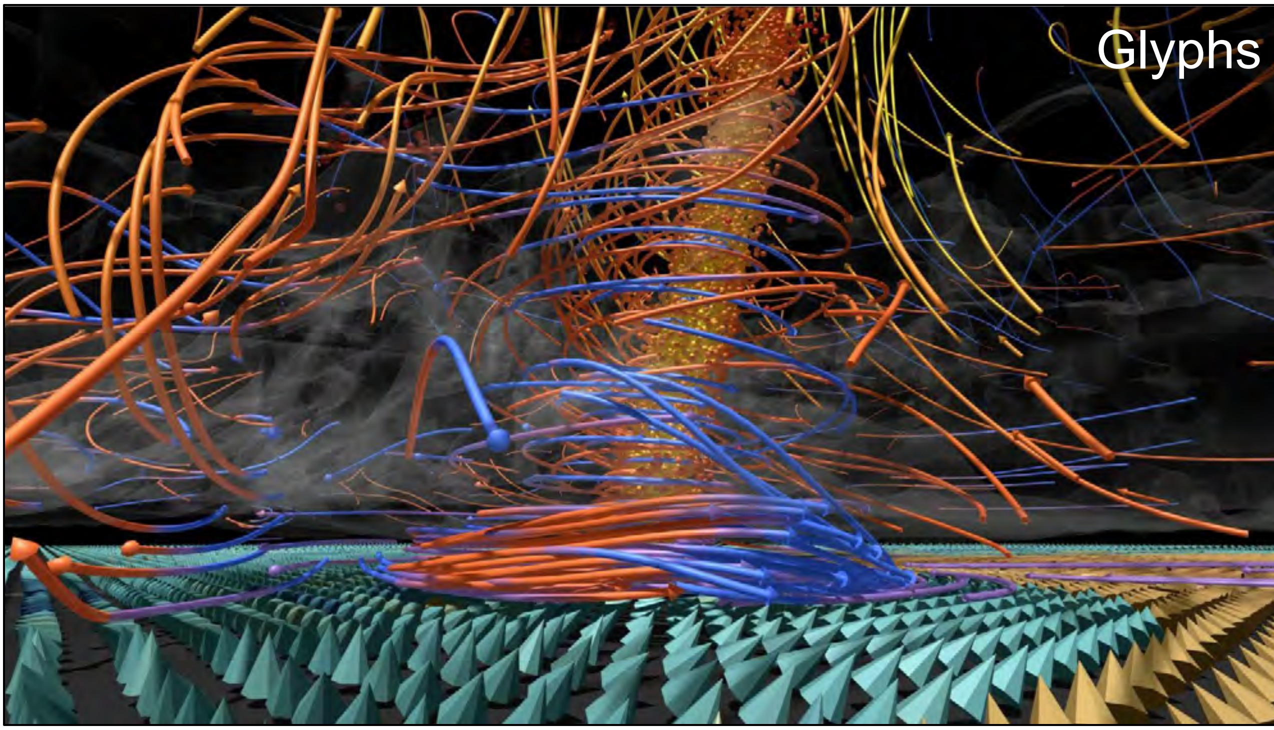


We used a visualization tool called VMD to analyze the hundreds of thousands of atoms in this simulation and identify which atoms were connected in patterns identifiable as DNA, then replaced those atoms with a ribbon geometry.



These ribbons are an **abstraction** of a more complex reality. In visualization we often use abstractions to take something visually or conceptually cluttered and turn it into something meaningful for our audience.

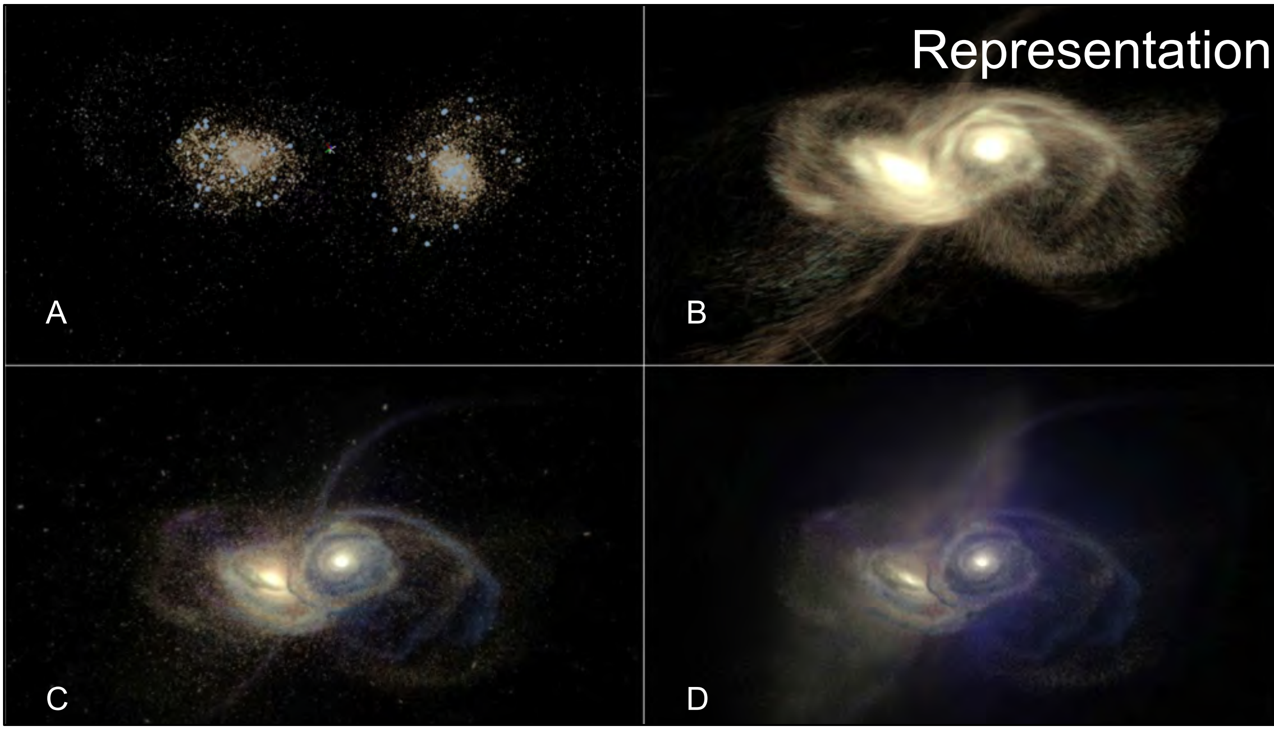
There are many ways to create abstractions. One of the most common abstractions is called a **glyph**. Glyphs are geometric metaphors - like the DNA ribbons.



This visualization of a tornado is filled with glyphs. Arrows are a very common glyph - this visualization uses arrows on the ground to show ground wind speed. But here they also draw an analogy to grass or trees blowing in the wind.

Streamtubes are another common glyph, and the orange-blue streamtubes here are tracing out the atmospheric wind direction. Balls near the center show areas of high pressure - what we associate as the actual funnel cloud of the tornado. There is also a faint gray shell around the storm, an abstraction of the storm cloud.

These abstractions show us scientific information much more clearly than recorded video of a tornado. A real world tornado is filled with dirt and fog, so it can be hard to see the inside of it, and we certainly can't tell by looking at it where changes in wind speed, temperature, and pressure occur.

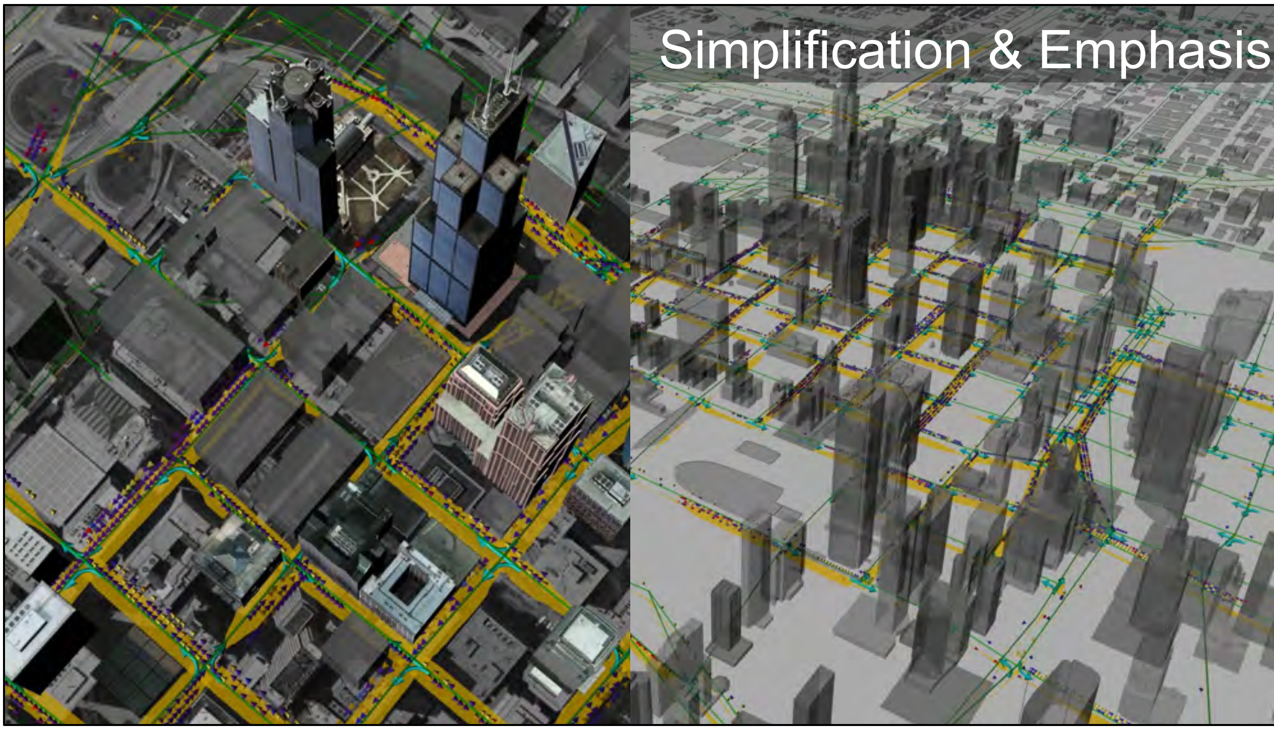


A **representation** is the appearance of a data object that is the result of many design choices.

These four images show the exact same colliding galaxy data at the exact same moment. But they have significant differences.

In the top left image, we see only stars, with colors emphasizing regions of rapid star birth.
In the top right image, we still see only stars, but the motion blur is exaggerated to indicate the directions in which the stars are moving.
In the bottom left, gas has been added, colored by the identity of which galaxy it originated in.
And in the bottom right, the gas has been colored by its destination.

Each of these representations tells a different story of the underlying data.



Representations can also help cut through the noise of a busy dataset.

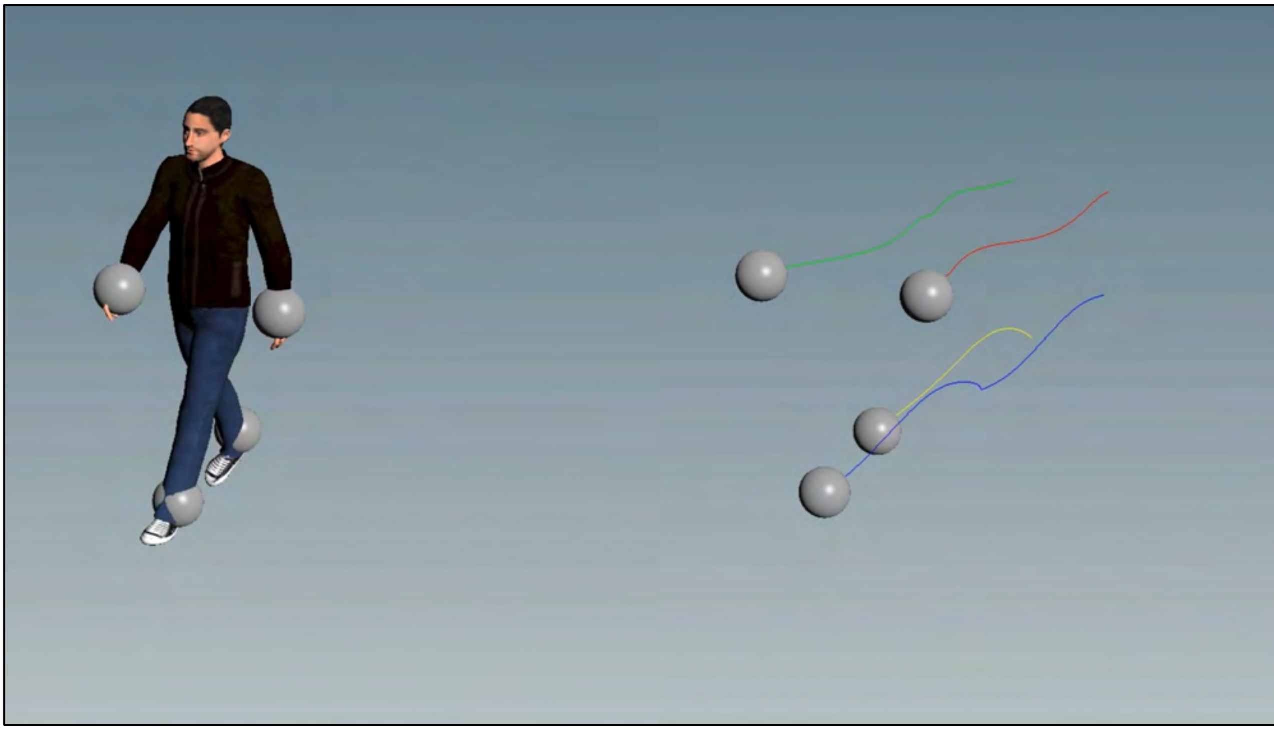
This visualization of traffic shows the city of Chicago outlined by recognizable buildings from the downtown area. The translucent gray representation isn't distracting - the viewer's attention is drawn to the brightly colored drivers which serves the data narrative.

In the earlier draft of this sequence, the buildings were shown more realistically. The visualization designers realized the shapes and locations of the buildings were familiar enough that they didn't need to include all the distracting detail of photographic textures.

Derivative Data

Types of Derivative Data

- Streamlines
- Isosurfaces
- Interpolation
  - Extrapolation

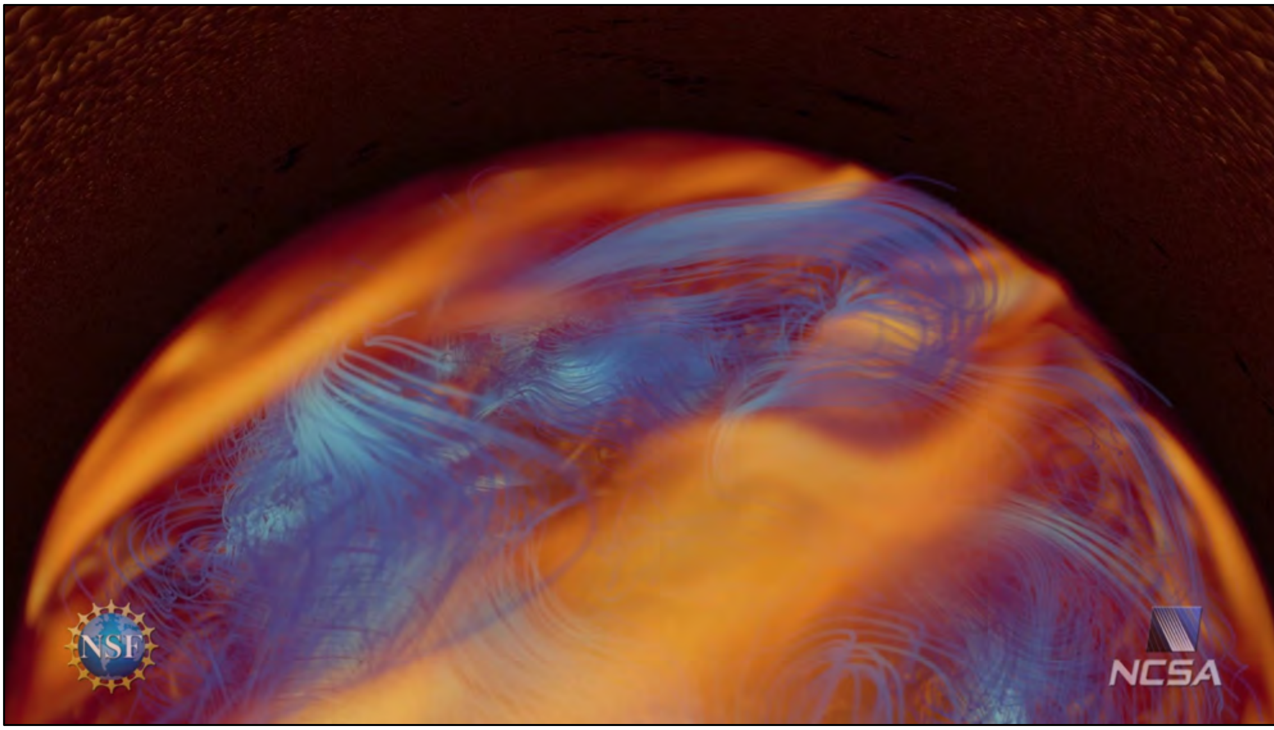


If you wanted to visualize the history of how a person moved as they walked through a room, you could attach motion trackers to a few points on their body, and track the positions of those points. Then, you could draw lines that show the movement of those points over time. These lines are a type of derivative data called **streamlines**.

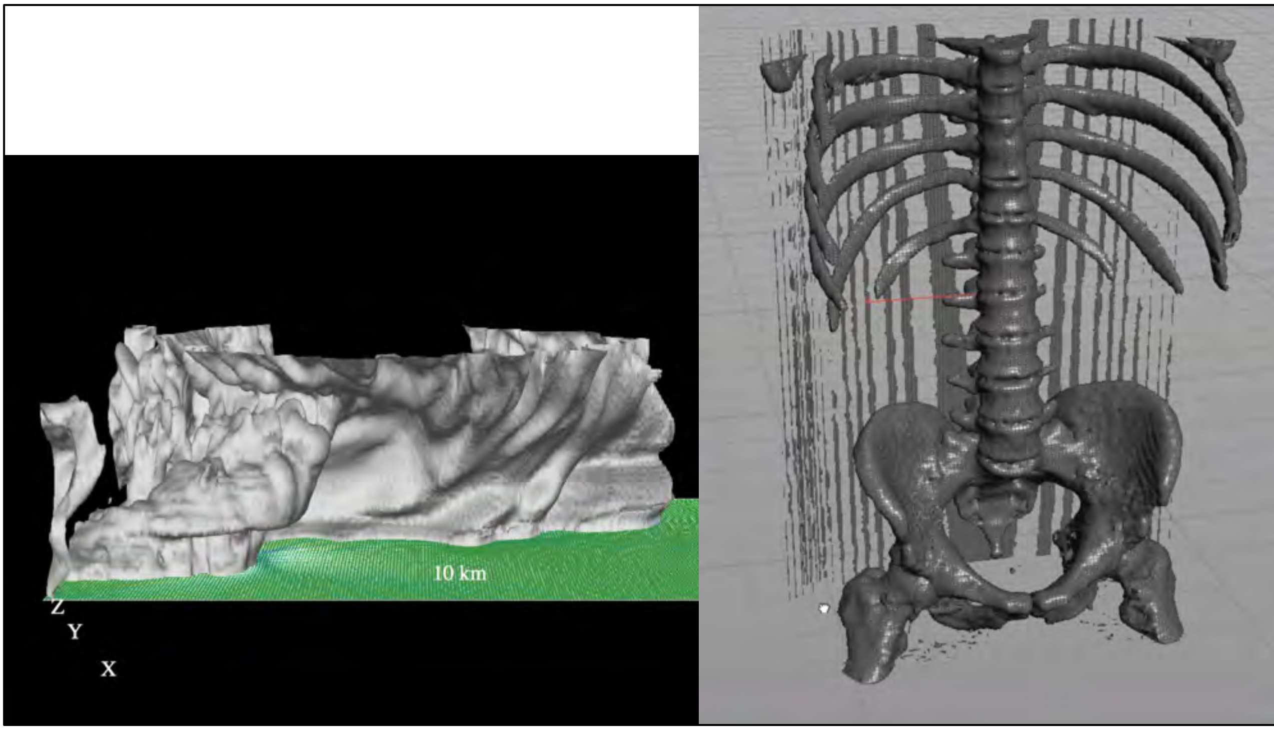
You can trace streamlines through any type of data that has motion or flow.

Imagine that you took a handful of leaves or particles and threw it into a tornado dataset. The wind attribute would carry those particles through the data.

If you traced the path of each particle with a streamline, you could see how the wind moved over time, which would give you a better understanding of the motion of the wind than just trying to visualize it as a volume, directly.

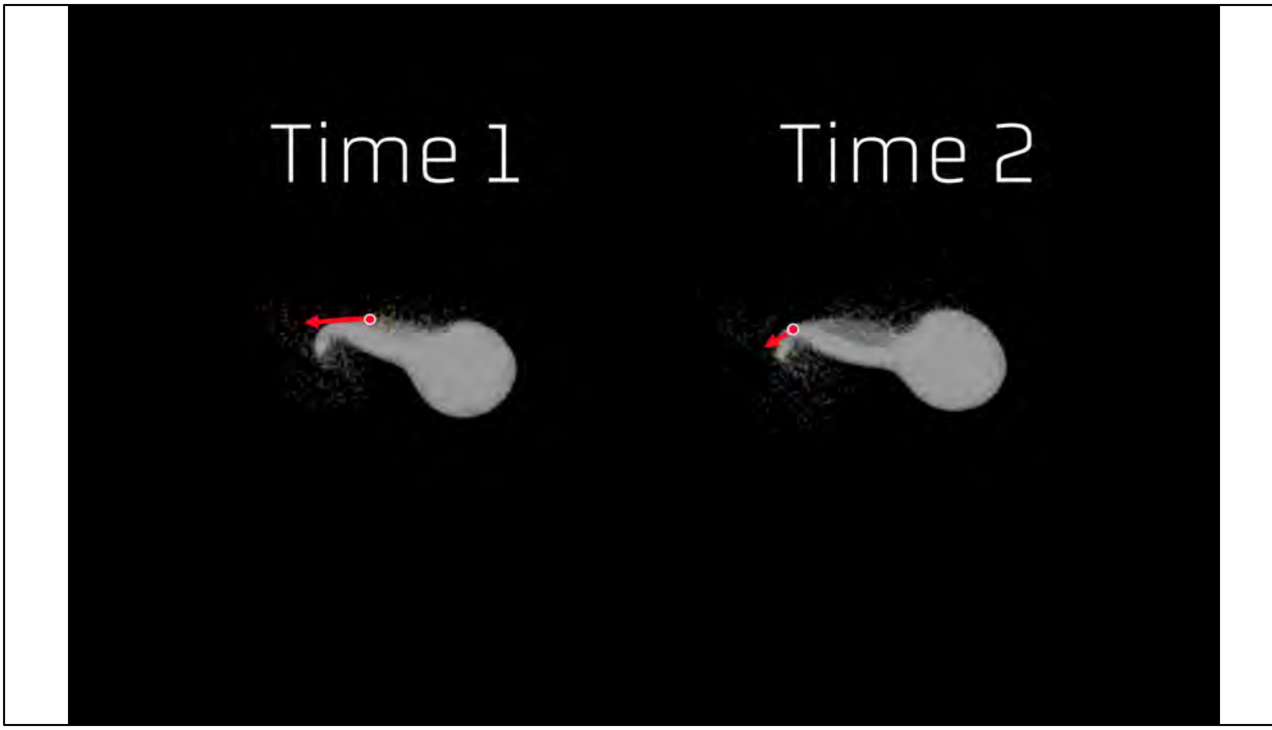


You can trace streamlines to show other types of flow like magnetic fields near the surface of the sun as well.



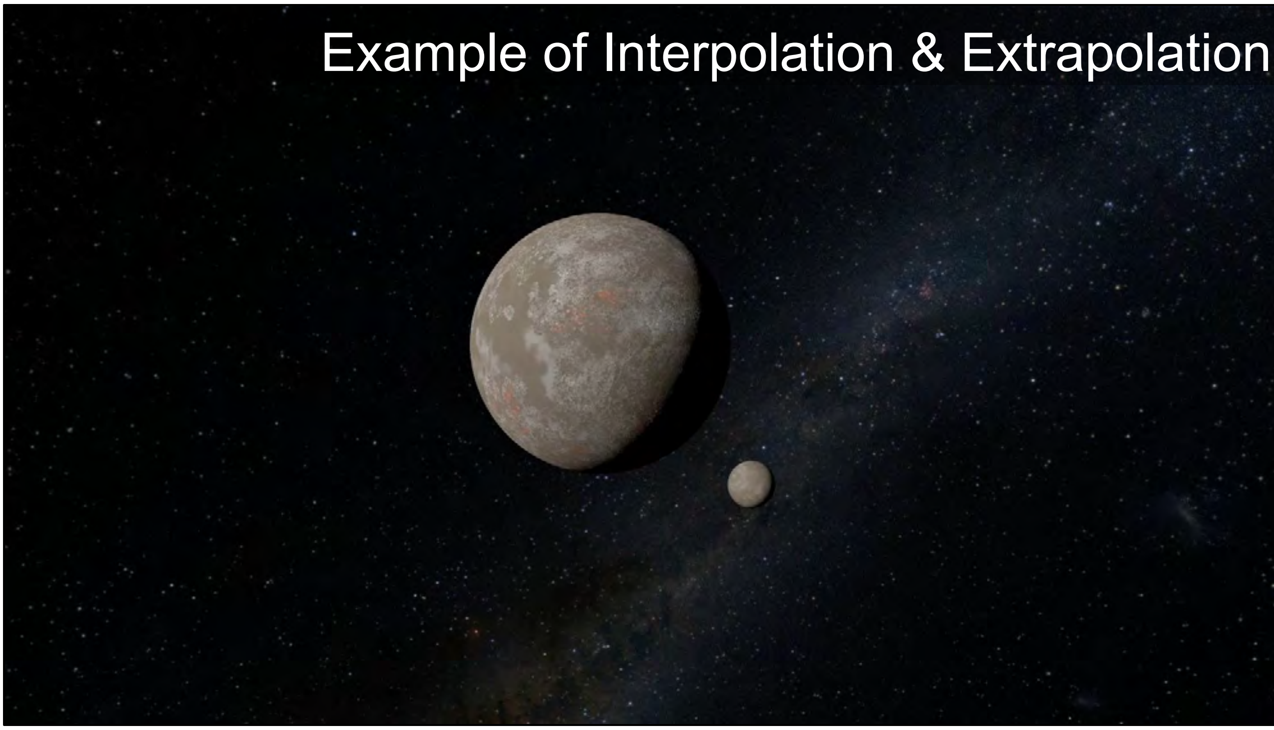
You can pull out specific features of ai dataset by using **isosurfaces**. You can think of this as drawing a wall around a part of the data where the value is some specific number.

This can be an effective way to simplify a complex 3-dimensional volume dataset into something more concrete.



A smoother way to play back your timesteps involves creating derivative data through **interpolation**, in which you calculate in-between timesteps. If you know where a point is at time 1, and where it is at time 2, you can estimate where it might be at time 1.5.

Interpolation is a heavily debated topic in the visualization community. Purists can argue that interpolation of data is lying or misleading. Scientific simulations use many incredibly complex physics formulas, and using a simple interpolation scheme to bypass those formulas can be inconsistent with the data. Furthermore, showing a simulation with more time resolution than was actually calculated can misrepresent realistic motion, for instance by showing the wiggling of atoms and molecules that all move faster than the frame rate of the movie, but which move at vastly different timescales from each other.

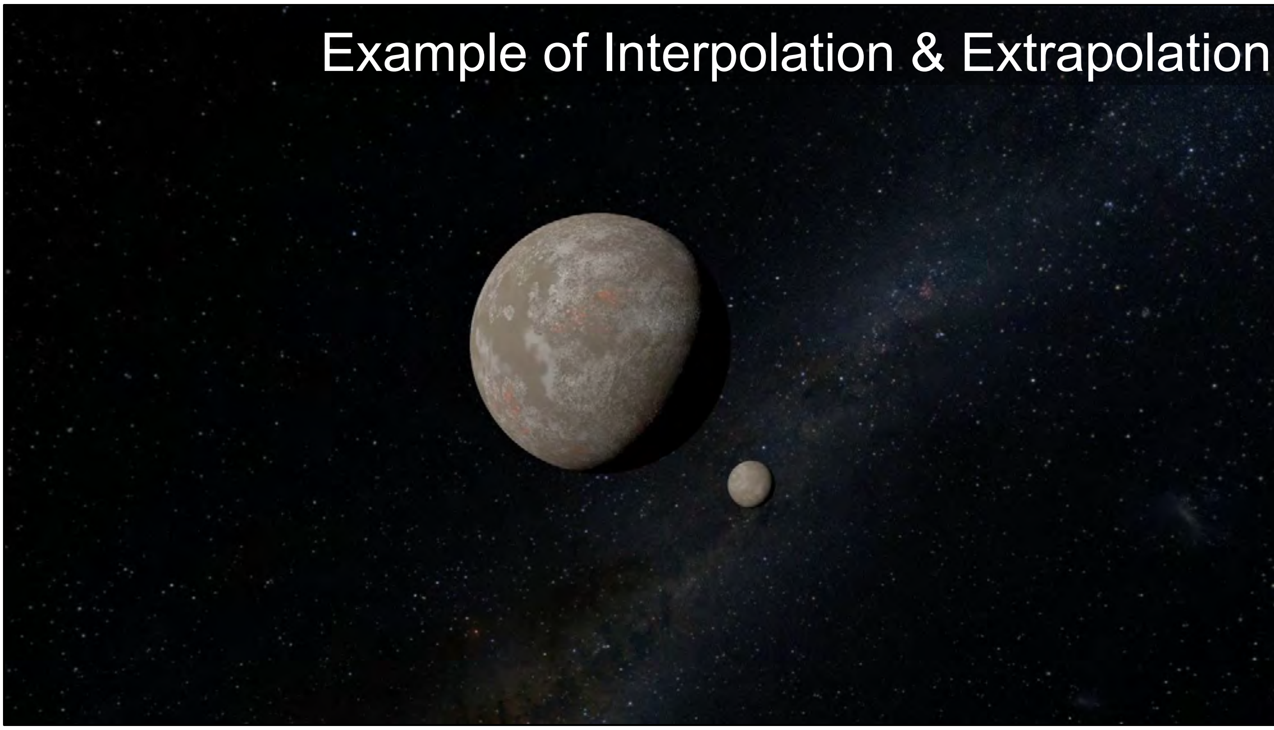


However, when a scientist runs a simulation, they may not write out very many timesteps. Each timestep might have many variables at every single point, so file sizes can get extremely large, and simulations are very expensive to run and can take weeks or months of computation time.

And, it's not necessarily scientifically useful to write out many timesteps. When there's less exciting stuff happening in a simulation, scientists may write out data less frequently, and write it out more densely where an event actually happens. So, the data we receive can be very sparse, and even non-linear in time.

We may get fewer than 100 timesteps, which we want to turn into a 2-minute shot at 30FPS. Doing so smoothly requires interpolating the data.

Here we are looking at a pre-visualization of a planetary collision. If we did not interpolate, the entire event would be over in a moment. While we *could* stay true to the data by stepping through it across multiple frames, like on the left…



… that doesn't produce a cinematic outcome that audiences can get emotional about. That choppiness is not only distracting, but makes it difficult for an audience to internalize the physical interactions captured in the simulation.

The scientist's goal is to create an accurate simulation while being efficient about using expensive supercomputing resources.
But our goal is to convey the science to the general public, and for our purposes, this type of interpolation is necessary.

Registration

People inherently trust visualizations as authoritative sources of information - after all, they are based on data from numerical calculations so how could they be wrong?

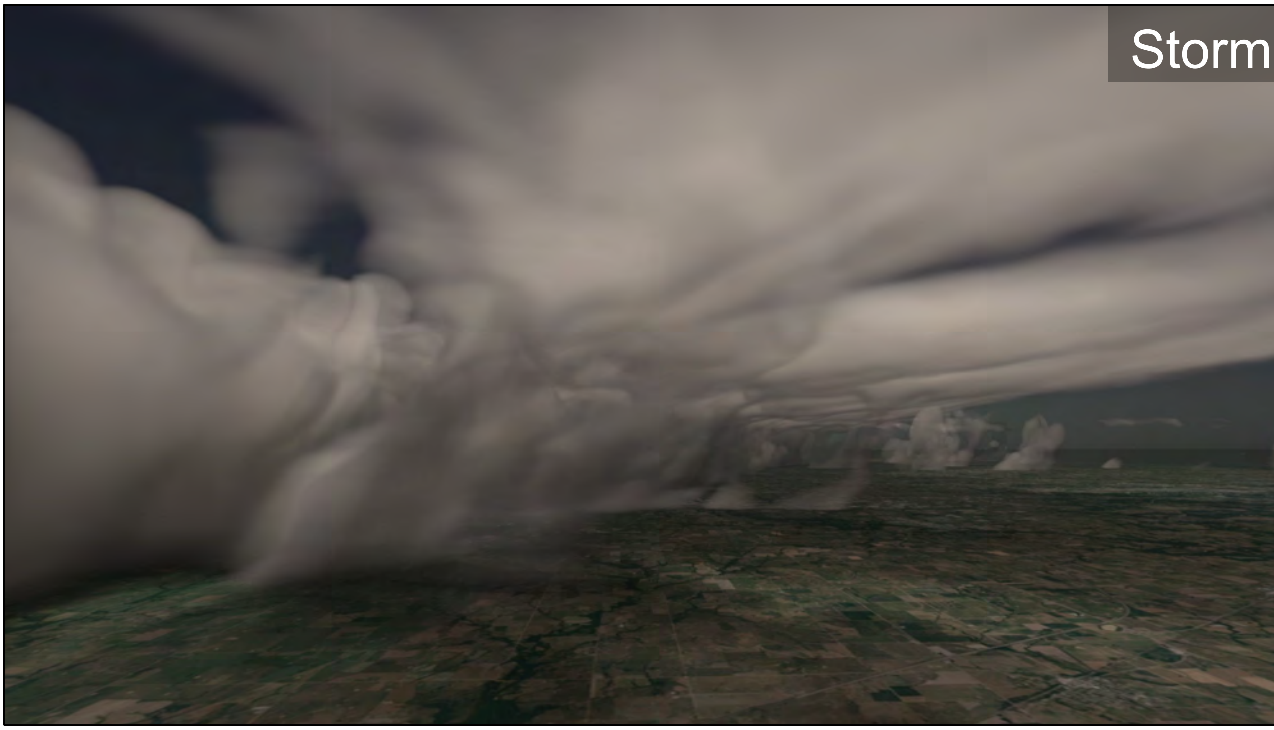
Especially because modern audiences are not as skeptical toward visualizations as they should be, good visualizations need clarity and context.

One of the best ways to provide context in a scientific visualization is to embed datasets inside each other.

Registration & Compositing

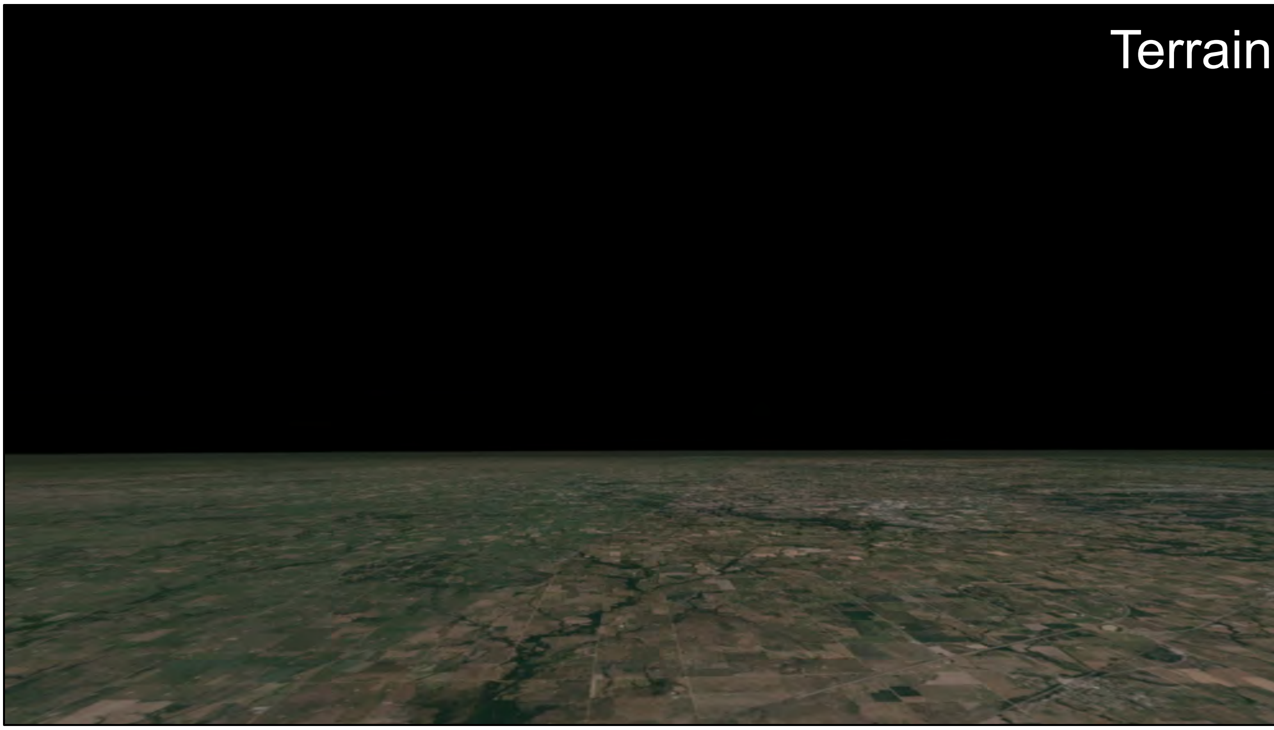
- Register data layers
  - May be rendered using different software
  - Combined in compositing

Embedding datasets is useful but it's not trivial. In addition to doubling the work you need to do reading and filtering data, you also need to find how the two datasets fit together. In the process of **registration**, you need to unify your datasets' units of time and scale, and synchronize the positions and start times of key features and events.



In this visualization of a tornado-forming storm that hit Joplin Missouri in 2011, an important element of the story was about the real city and human lives that were affected. It was important to ensure that the tornado was placed in the correct geographic position.

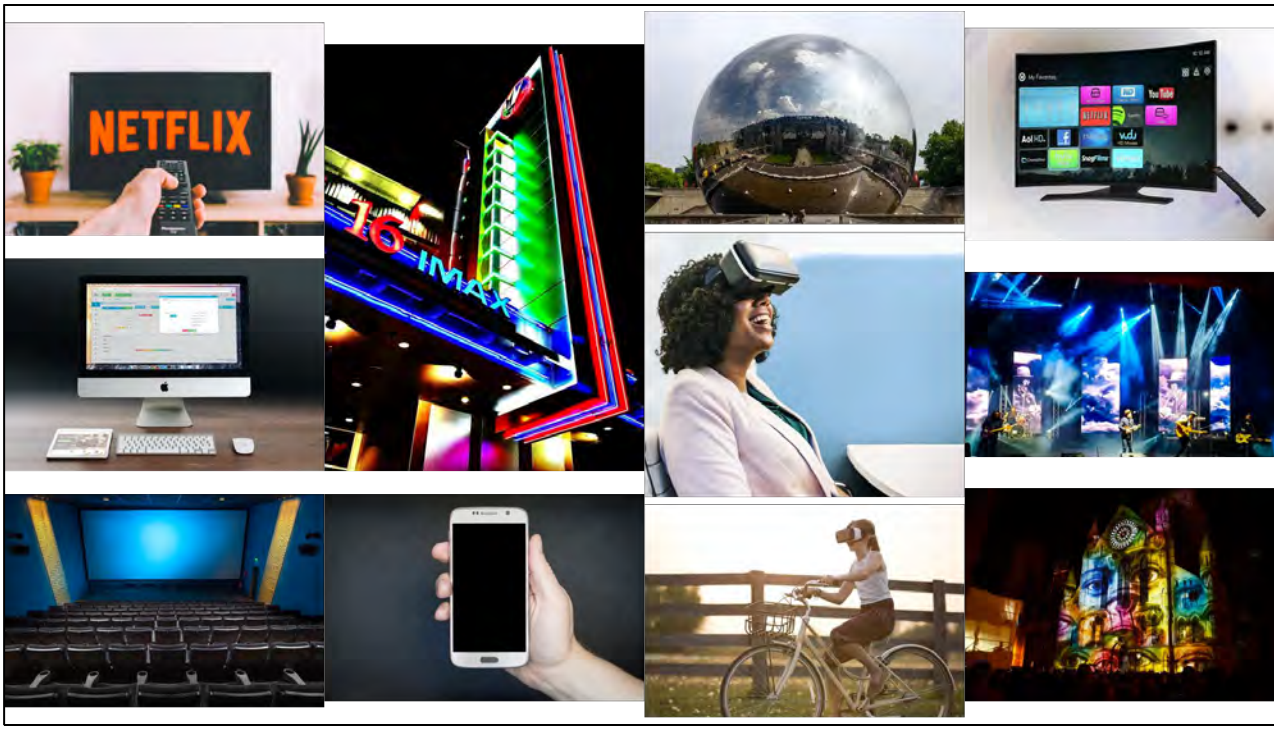
This is what the storm simulation looked like on a context-free black background.



We worked with a Geographical Information Systems expert to find the highest-resolution map of the terrain that was available, and then registered it to the coordinate system of the storm simulation.

Registration is taking multiple asynchronous data sources and transforming them into synchronized temporal and spatial units. In this visualization of a tornado-forming storm that hit Joplin Missouri in 2011, an important element of the story was about the real city and human lives that were affected. It was important to ensure that the tornado was placed in the correct geographic position.

Cinematic Presentation



A visualization can be presented to an audience in a multitude of display devices and resolutions—on a rectangular screen, in virtual reality, in a museum exhibit, in a theater, or in a planetarium dome. There are design challenges and opportunities for each of these display environments.

Modern digital planetariums can immerse audiences in new worlds, transporting people to the far reaches of the cosmos—or into the tiniest details of a microscopic realm. But crafting true immersion requires discipline. The feeling of immersion can be easily disrupted, so content must be designed with great care, respecting the cognitive load of audiences who are trying to make sense of complex visuals and sound.

Successful immersive storytelling takes people on a “narrative journey,” creating a sense of travel—an embodied experience unlike traditional cinema. In the last decade, the California Academy of Sciences has produced six award-winning shows that cover topics ranging from the origin of life in the Universe to the fate of Earth’s most precious ecosystems.

Many of the ideas that follow are rooted in the work of Ben Shedd, whose “Exploding the Frame” article has influenced many large-format and fulldome filmmakers. You can find some of Ben’s work here:

https://benshedd.com/exploding-the-frame-papers/

Guiding Viewer Attention


Filmmaking for Fulldome:

Best Practices and Guidelines for Immersive Cinema (Part I)


Filmmaking for Fulldome:

Best Practices and Guidelines for Immersive Cinema (Part II)

tinyurl.com/fulldome-filmmaking

My colleagues and I have published two articles outlining design principals for immersive storytelling in planetariums. They are both available at:

http://tinyurl.com/fulldome-filmmaking

Guiding Viewer Attention

- Object Motion
- Faces (human or animal)
- Brightness
- Growing in Size (collision course)
- Focus
- Framing of Space
- Color Variation
- Camera Motion

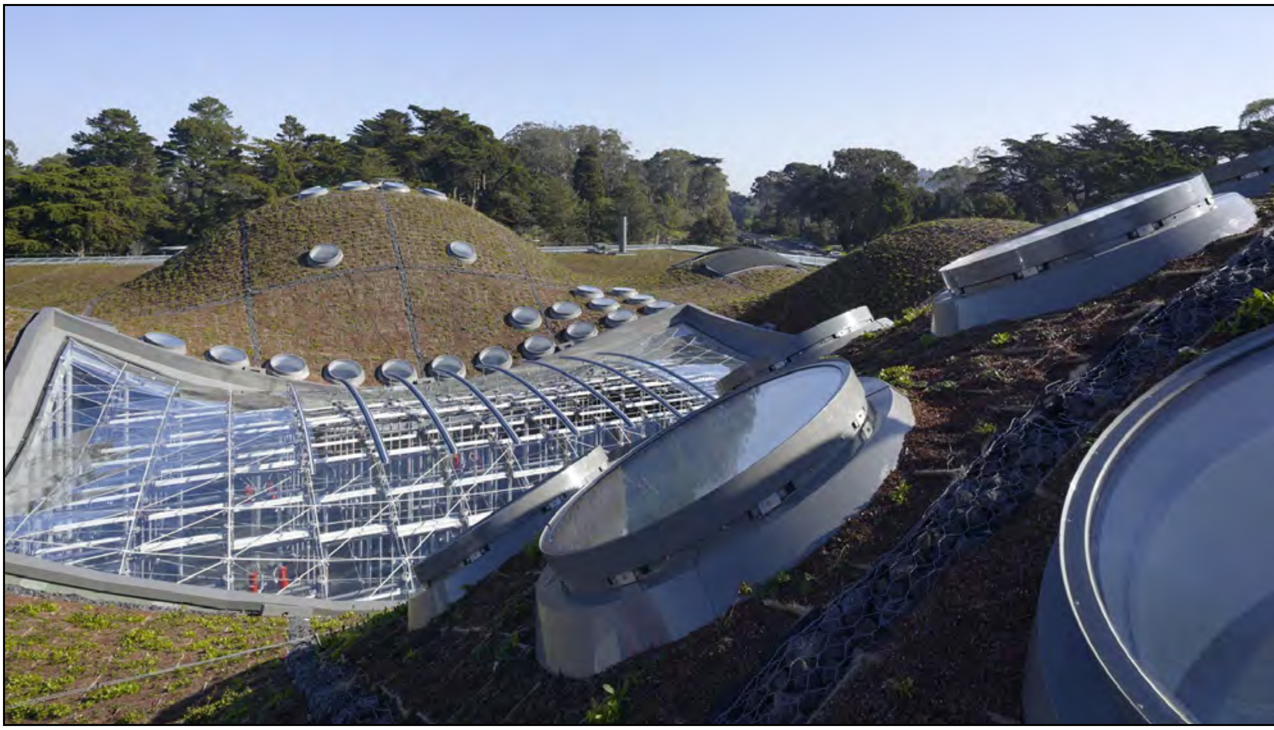
Among the ideas we address in the “Filmmaking for the Fulldome: Best Practices and Guidelines” articles is how to guide viewer attention. This is a brief list of visual cues in rough order of the importance for audience attentiveness.

Cinematic presentation is all about context. So for context, I would like to introduce you to the place I work, the California Academy of Sciences, in its five-year-old building in Golden Gate Park in San Francisco.

I want to call your attention to two things:

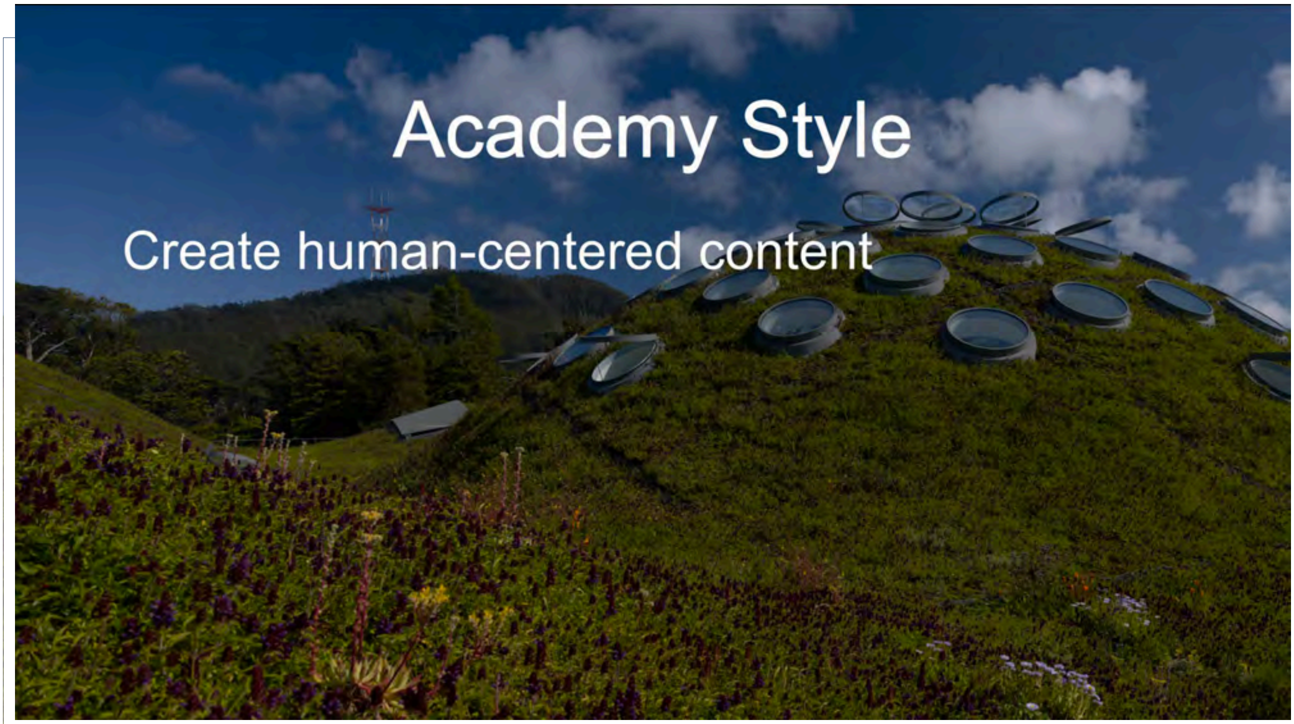
- First, we are a natural history museum—or as we like to say, a natural future museum—with active research in the life sciences as well as home an aquarium and planetarium.
- Second, we use our location in a park to help connect people to the natural world—our mission is to explore, explain, and sustain life on Earth

This is important context for the work we do in the visualization studio.



If you were to settle onto the roof of the Academy, you would see the roof's artificial hills (mirroring the hills of San Francisco) covering two of the institution's main attractions—the rainforest (in the foreground) and the planetarium (in the distance).

The environment in which we present our content in some sense drives the values with which we choose to communicate.



I think of us as having developed what I have nicknamed the “Academy Style,” a conservative but effective approach toward fulldome filmmaking.

Academy Style leverages four core values: creating human- and audience-centered content, starting shows at a human scale, maintaining camera motion and continuity, and using data (or at the very least, deeply-informed artwork) whenever possible. This basically adds up to treating the audience member with respect—respecting their comfort, their intellect, and their cognitive load.

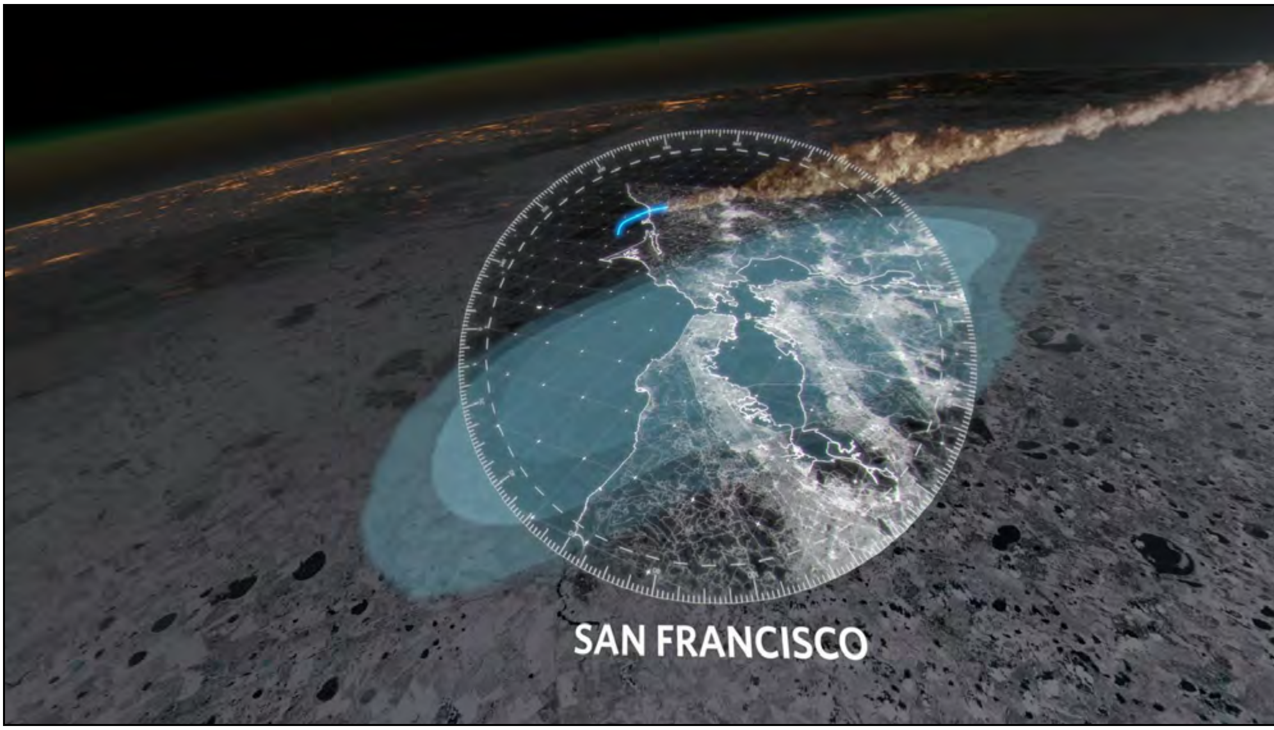
Elements of the “Academy Style” include:

- Create human- and audience-centered content
- Start at human scale
- Minimize cognitive load
- Maintain camera motion and continuity
- Keep visualizations “easy to read”

Re human- and audience-centered content: “I like to think that our shows bring the warm heart of poetry to the cold facts of science”

I’d like to start by addressing the focus on creating human- and audience-centered content.

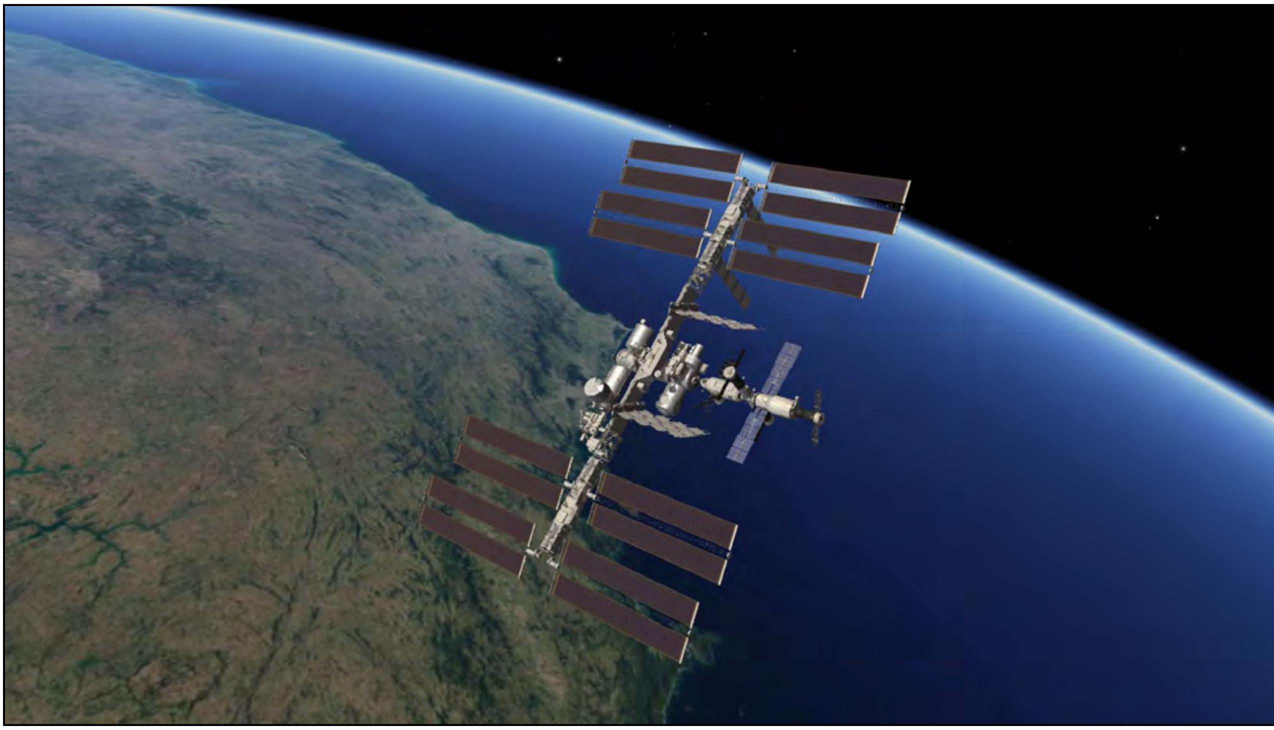
For example, when we set out to teach audiences about earthquakes, we knew that we needed to recreate the 1906 San Francisco earthquake. We chose to do this not least because, as an institution based in the city, we knew that our audiences expected it. (As it turns out, modern seismology was born out of research into the causes of the 1906 earthquake, so there are solid didactic reasons to mention the event, too.) But also because we could use the visceral and dramatic experience of the immersive recreation of the event to draw people into the scientific story.



When we reveal the magnitude of the destruction in the wake of the Chelyabinsk meteor impact in our show about asteroids and comets, we overlay maps of New York City, the San Francisco Bay Area, and the Tokyo metropolitan area. We chose potentially familiar areas with distinctive coastlines (turns out landlocked cities don’t provide good geographic reference at that scale) to help make the content more relatable.

When we wanted to highlight the role of sea otters as a keystone species in California’s kelp forests, we leveraged the audiences affection for mammals—and high-contrast mammalian faces in particular—to help create a connection to the species. And to provide an entry point into the complex topic of food webs, as I’ll describe later in the presentation.

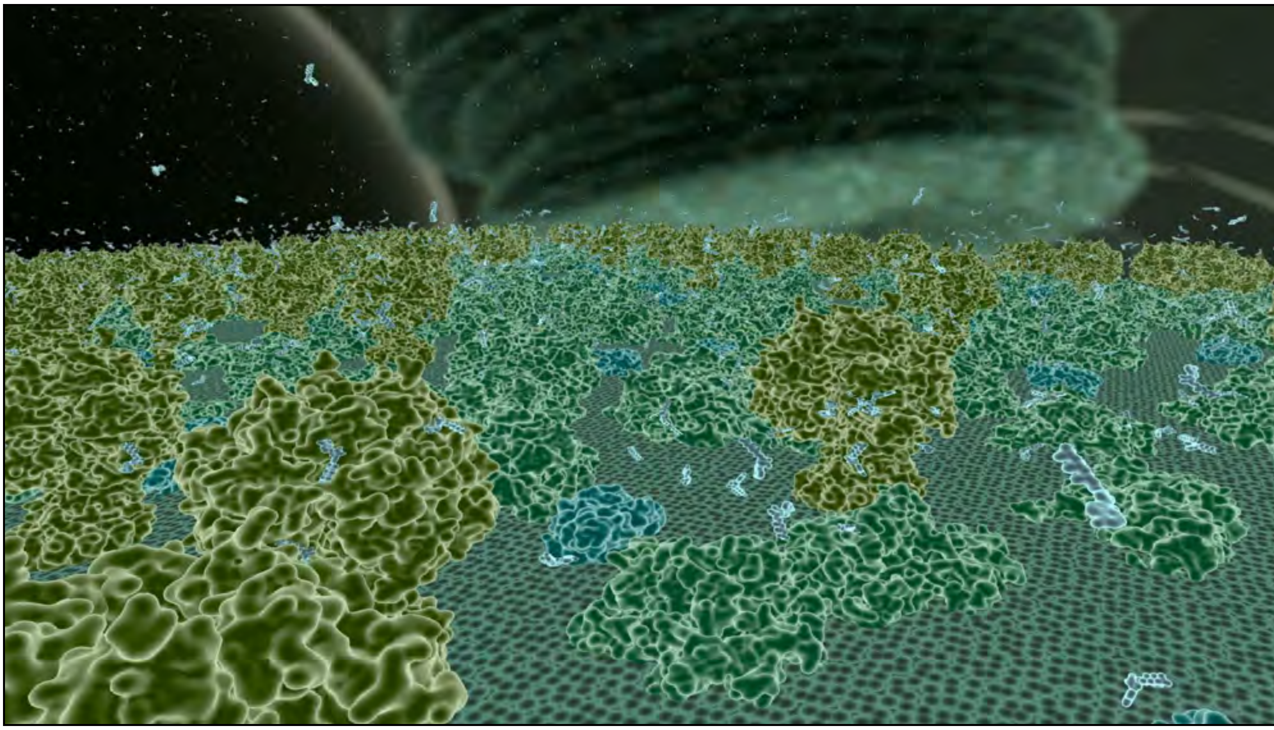
A corollary of creating human- and audience-centered content is that we start our shows at a human scale. The planetarium is a potentially disorienting space, especially for first-time visitors—and we want to use the immersive venue to explore wildly different scales in space and time—so we start by "grounding" our audiences in the familiar.



This snapshot actually comes from our real-time software, used to deliver a live “Tour of the Universe” program in the planetarium. We begin at the International Space Station, in large part because it is human-scaled—people reside onboard, and the entire structure is about the size of a football field.

[VIDEO]

For *Life: A Cosmic Story*, the opening shot is of a redwood forest, fading in over the sounds of birds, insects, and wind, pretty much the view you would have if you were standing in Muir Woods a few dozen kilometers north of San Francisco. We then shrink the perspective continuously to examine the interior of a leaf on one of those redwoods.

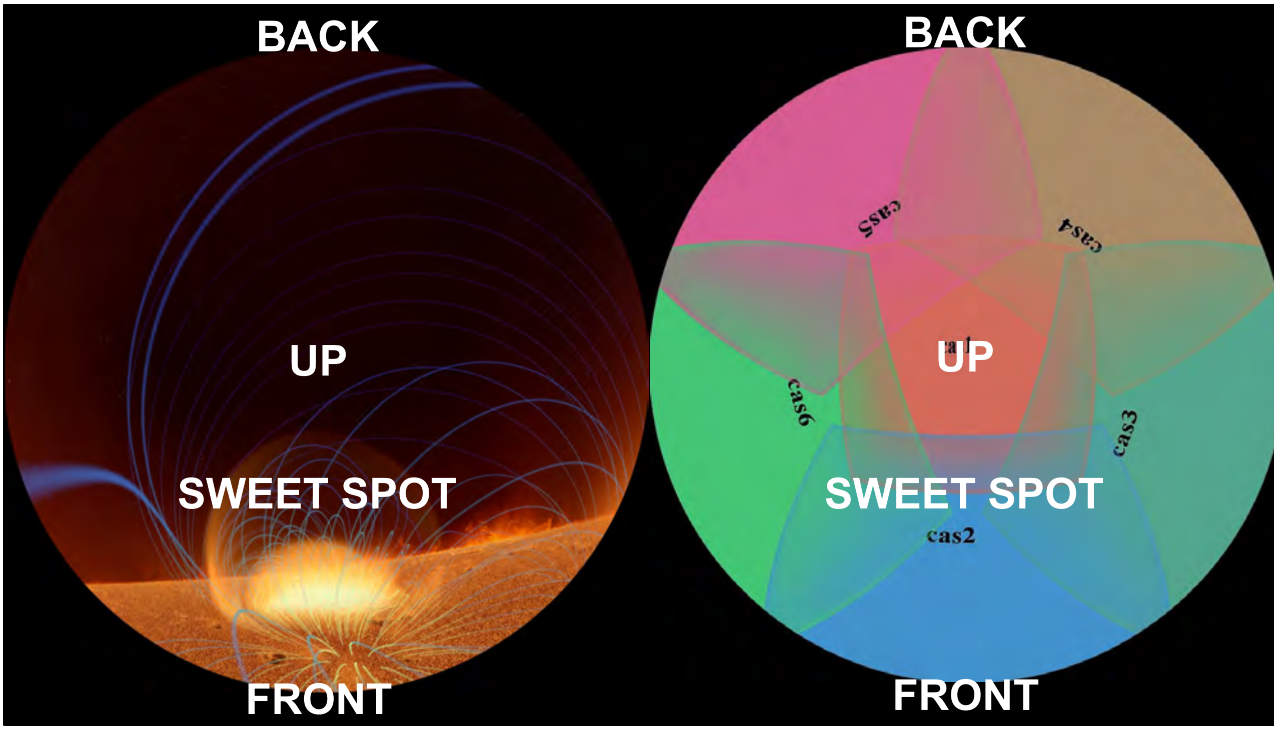


Ultimately, we end up inside a tree leaf, at the scale of the molecules involved in photosynthesis.

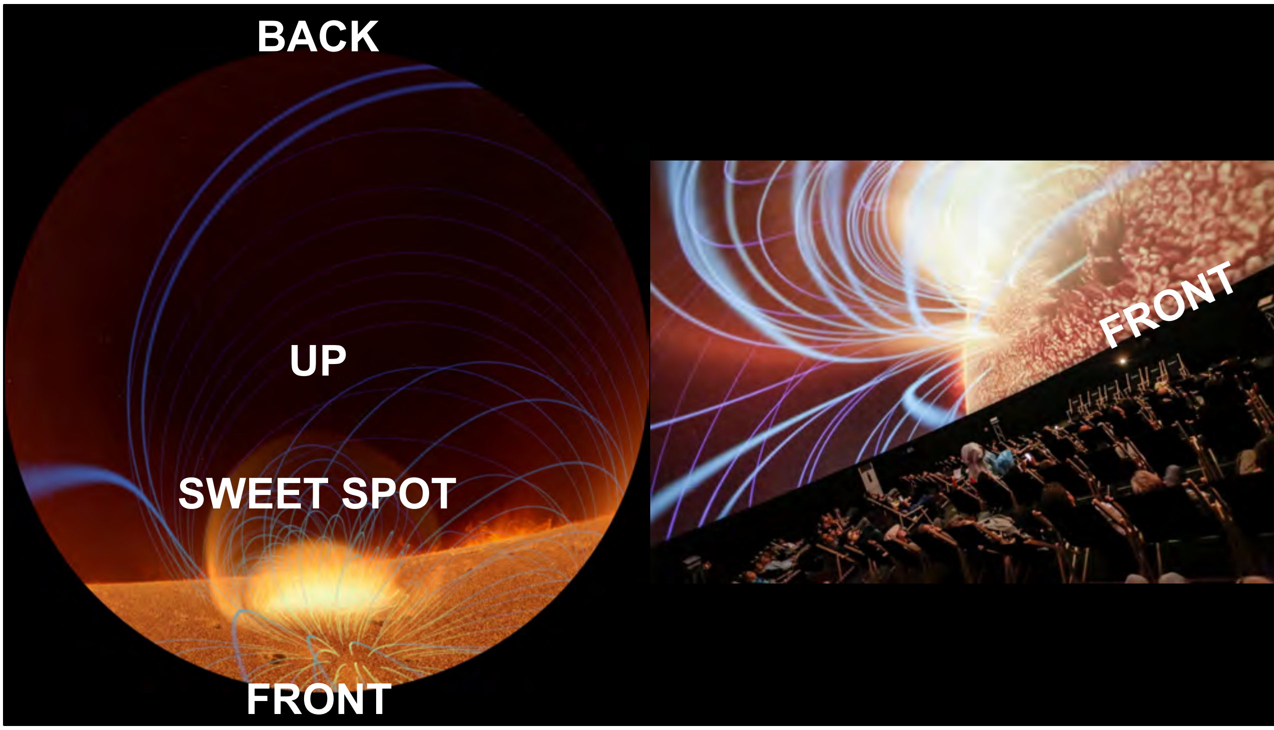
For *Earthquake: Evidence of a Restless Planet*, the show opens on McClures Beach, another site within easy driving distance of San Francisco but also a reasonable stand in for a generic beach. That location allows us the opportunity to use the ocean as a metaphor for the passage of time. More specifically, it allows us to examine the different time scales in the scene—from the lapping of waves over a period of seconds to the setting of the sun over several minutes, from the accumulation of sand over a period of thousands of years to the shifting of tectonic plates over the period of millions of years. In just a few moments, with an establishing shot of a familiar locale, we connect to a core theme of the show, namely “telescoping time” from personal to historical to geological time. Then we rocket to an aerial perspective to see the northern California coastline deformed by tectonic forces.

For *Incoming!*, we start with a lizard peeking over a rock in the Arizona desert (Fig. 2e), pulling back to see a hummingbird, cacti, and desert hares. We then follow a hawk flying over the nearby Barringer Crater, before launching to a satellite view of Earth, highlighting the locations of other impact craters in North America.

But I’ve been showing you HD crops of these images. It‘s important to note that they are just a subset of the imagery audiences are viewing.

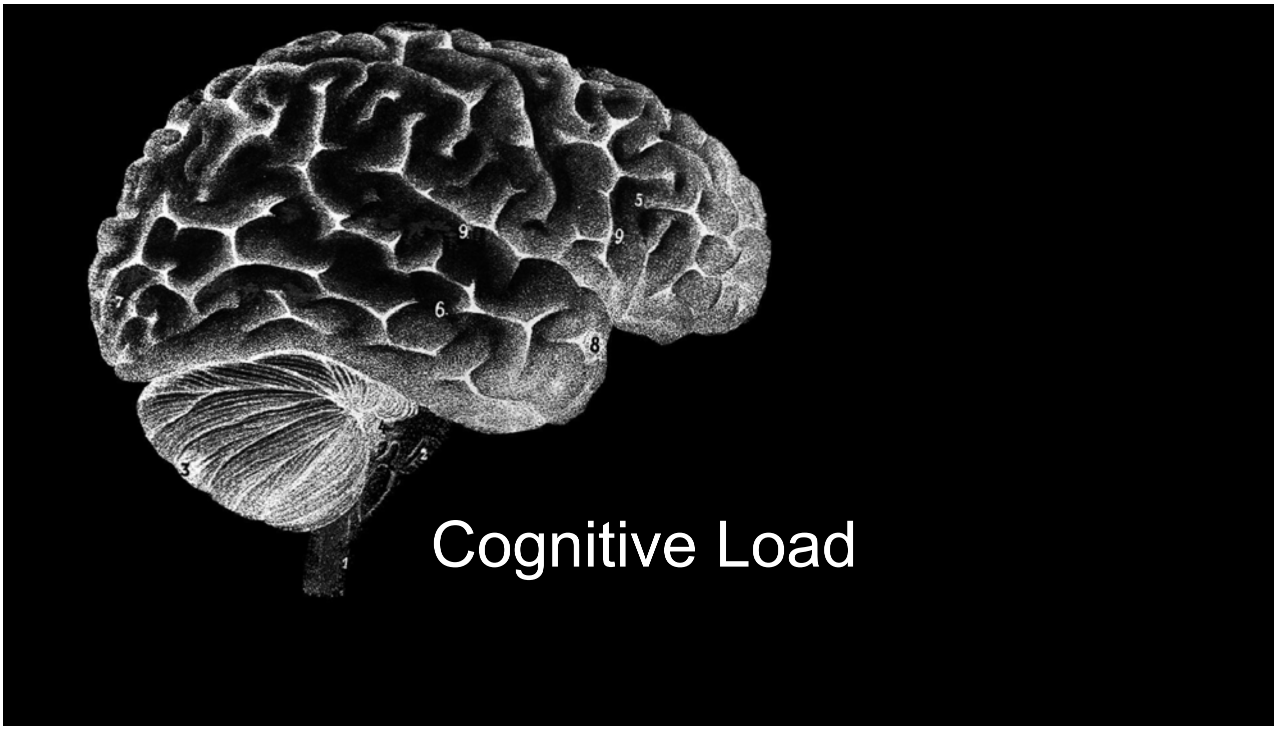


It's important to be aware of the type of screen for which you're designing a scene and a camera move and how audiences experience it.

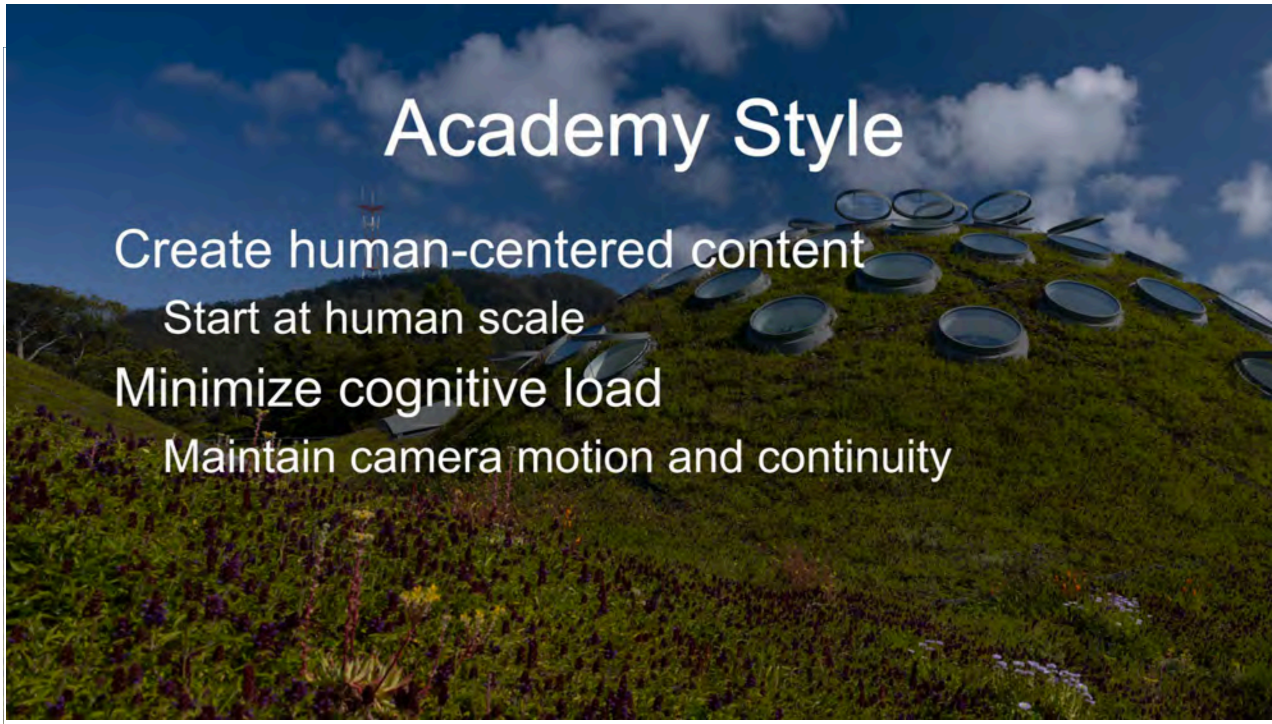


Fulldome theaters are 360 by 180 degrees, and often have ”sweet spots” – the area of the screen where the audience is predominantly facing. Dome theaters also have no upper frame edge, so to make something appear from off-screen, it must come from a far-off distance or from below the bottom frame edge.

So, designing a camera path for this type of theater is very different than creating one for a smaller screen, where the camera can move more rapidly and have more cuts. Learning the intuition to design camera paths for various environments takes research and practice. Camera choreographers are often surprised by the unintended effects of their camera moves when switching between different display environments.



All this information creates significant cognitive load for the audiences in the dome—they have immersive imagery they want to digest, while they are listening to a narration that may include a fair bit of factual information, along with a soundtrack that includes music and sound effects. That’s a lot to take in! How can we ensure that they aren’t overwhelmed?



So how do we minimize cognitive load on the audiences to help them follow the story and enjoy the narrative journey? First and foremost, we maintain camera motion and continuity in order to avoid disruptions in the viewer's sense of place.

First and foremost, maintaining camera motion increases the sense of immersion, reinforcing a sense of dimensionality by providing depth cues for foreground, midground, and background objects. Without relying on stereoscopic tomfoolery in the dome, we can make audiences feel situated in a space by giving them parallax cues for a three-dimensional environment. Because a dome also engages people's peripheral vision, this can be a powerful effect—for good or ill, in that individuals prone to motion sickness might feel a little *too* engaged by these techniques.

Avoiding cuts further supports immersion by allowing audiences to reside in a virtual environment for an extended period of time. When cutting from one immersive environment to another, the audience is effectively instantly transported from one place to another—potentially accompanied by a shift in scale—and it takes time for the human brain to make sense of such a dislocation. We can design the transitions in such a way to soften the blow; however, particularly for novice fulldome viewers, cuts can be very distracting and disorienting. So in our planetarium shows, we have avoided cuts as much as possible.

Average Shot Length

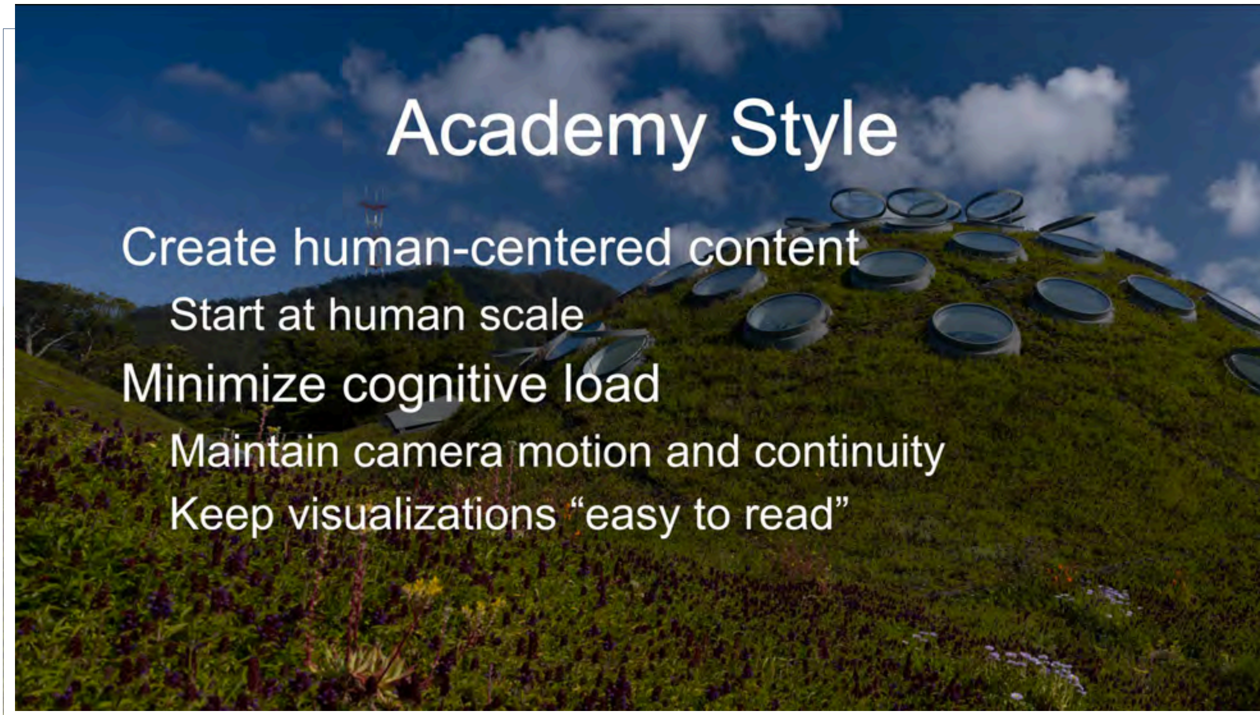
Fragile Planet (2008)	1545
Life: A Cosmic Story (2010)	225
Earthquake (2012)	192
Habitat Earth (2015)	183
Incoming! (2016)	766
Expedition Reef (2018)	160

Average shot length (in seconds) for shows created by the California Academy of Sciences.

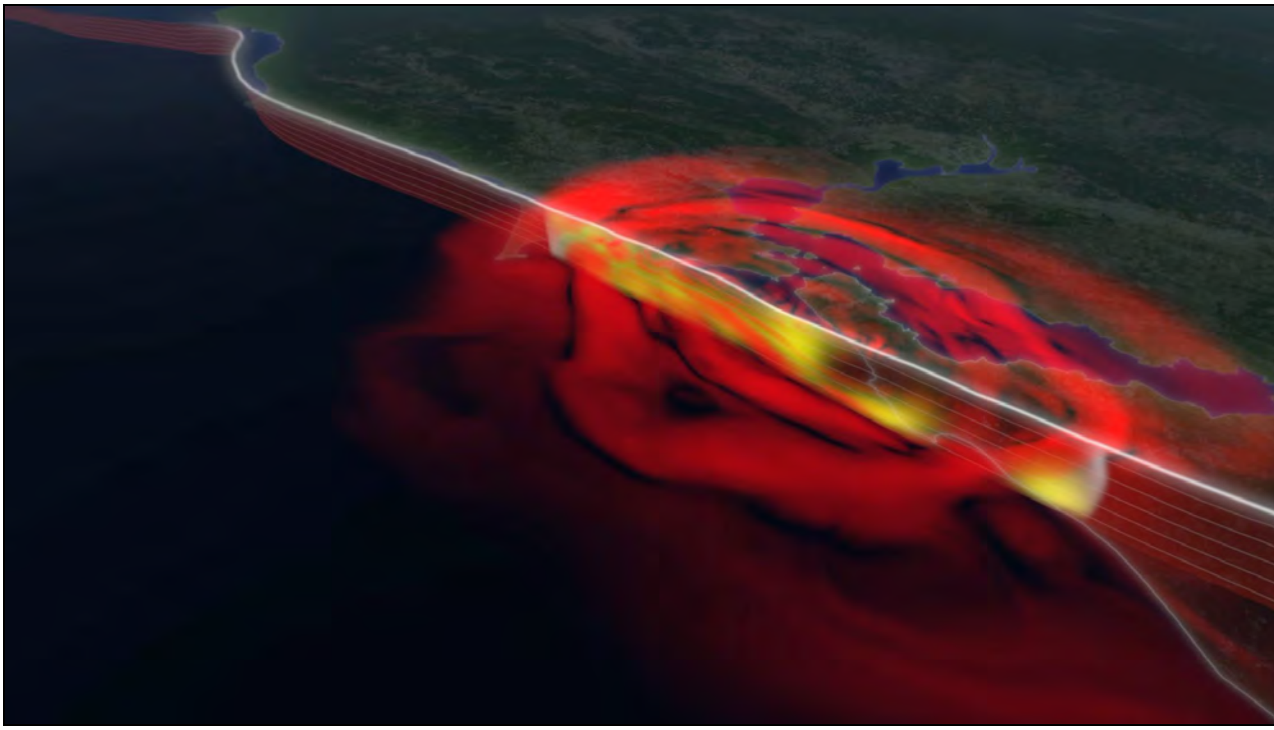
My first example comes from *Earthquake: Evidence of a Restless Planet* (2012).

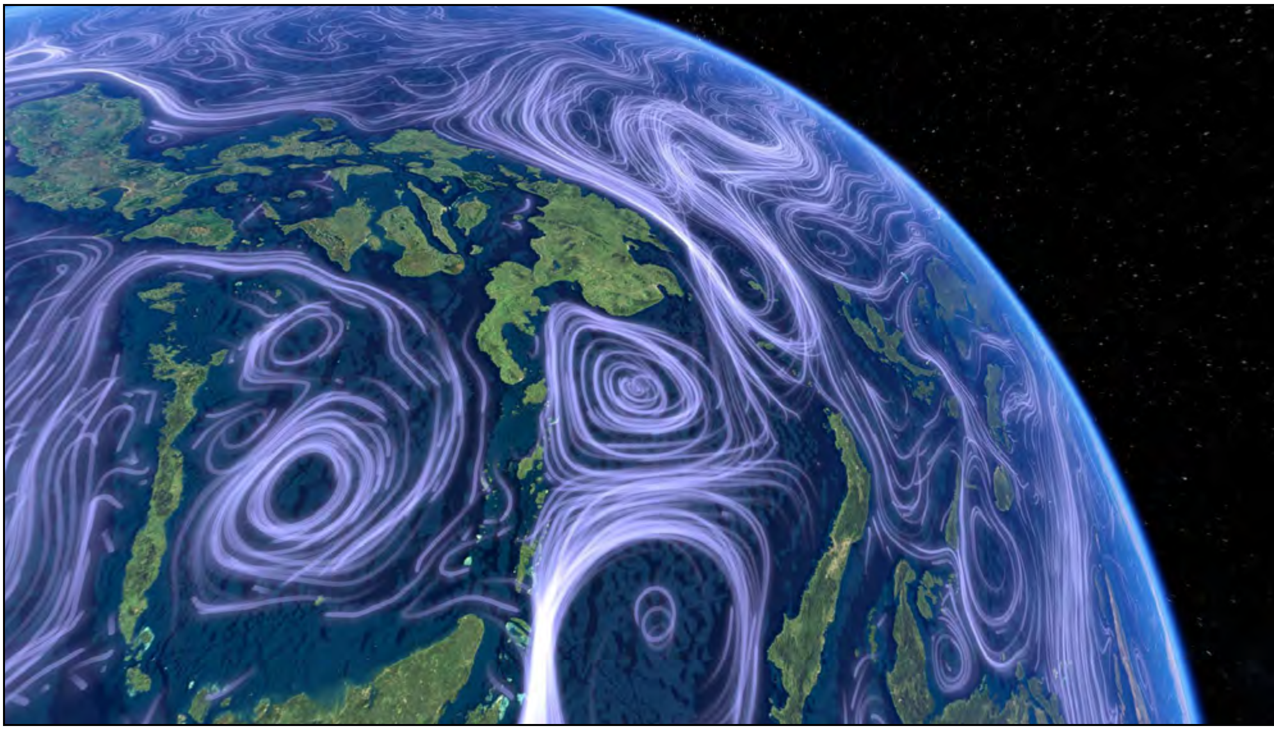
[VIDEO]

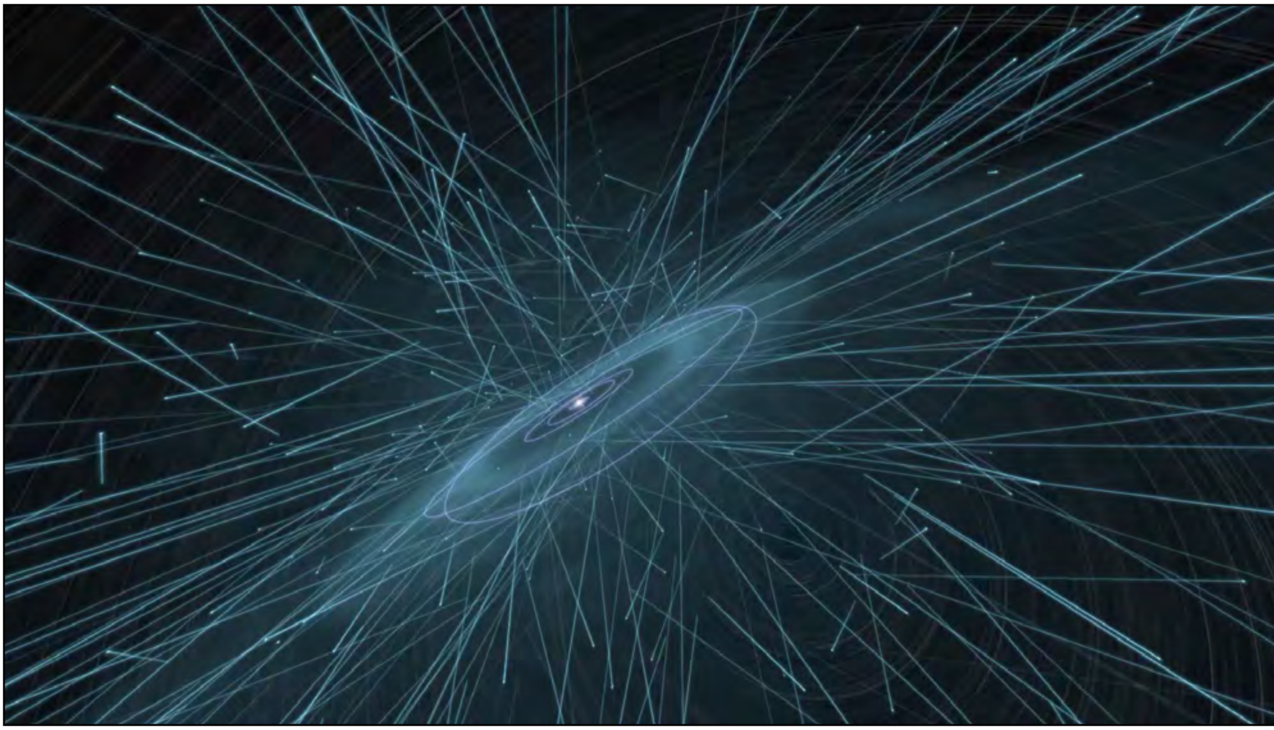
The shot in question begins with a street-level view of 1906 San Francisco, transitioning to planet-wide tectonic plate boundaries. This seamless transition starts with the visceral experience of an earthquake on the ground; as we pull out to a view of San Francisco, data overlays show the damage inflicted to the city by fire and the regional effects of surface shaking; broadening our perspective, we then follow the penetration of seismic energy underground where it reveals our global representation of Earth’s core, mantle, and crust. This sequence thus visually and conceptually situates an isolated seismic event in the context of the planetary-scale system that caused it.

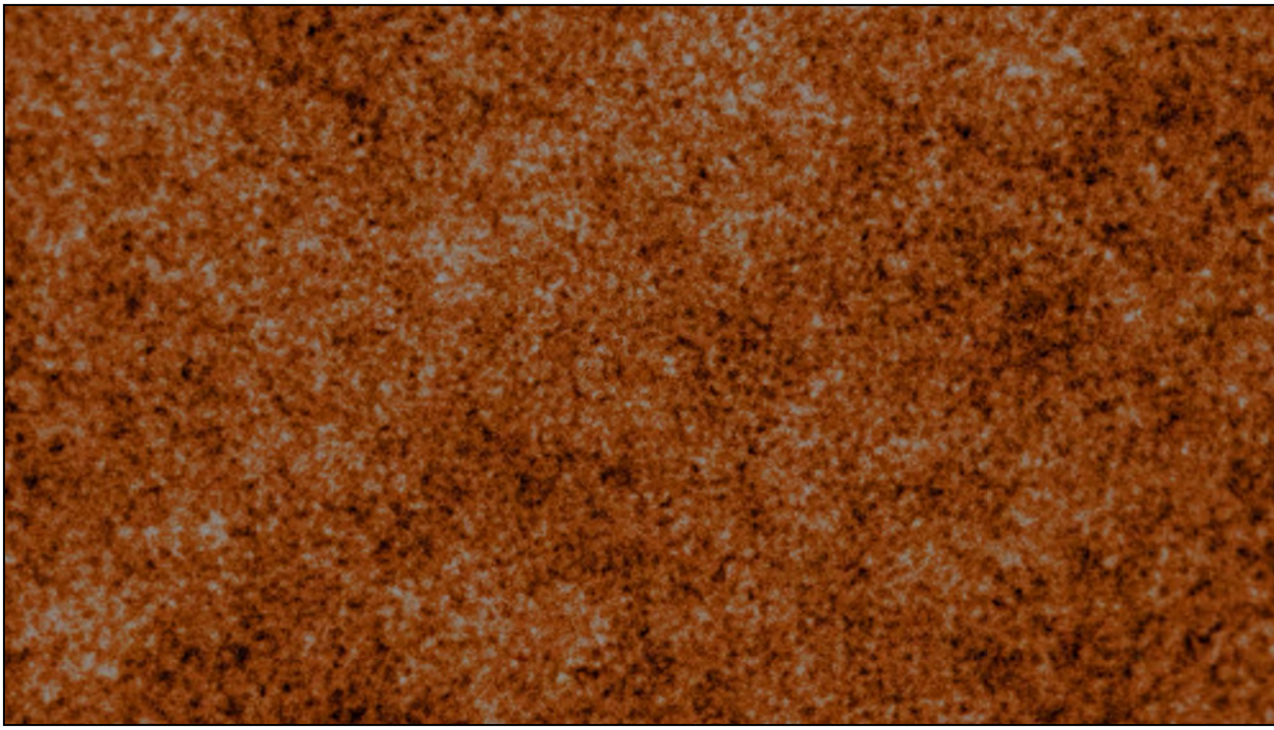


Finally, we try to keep visualizations “easy to read” in order for people to follow along with the story.









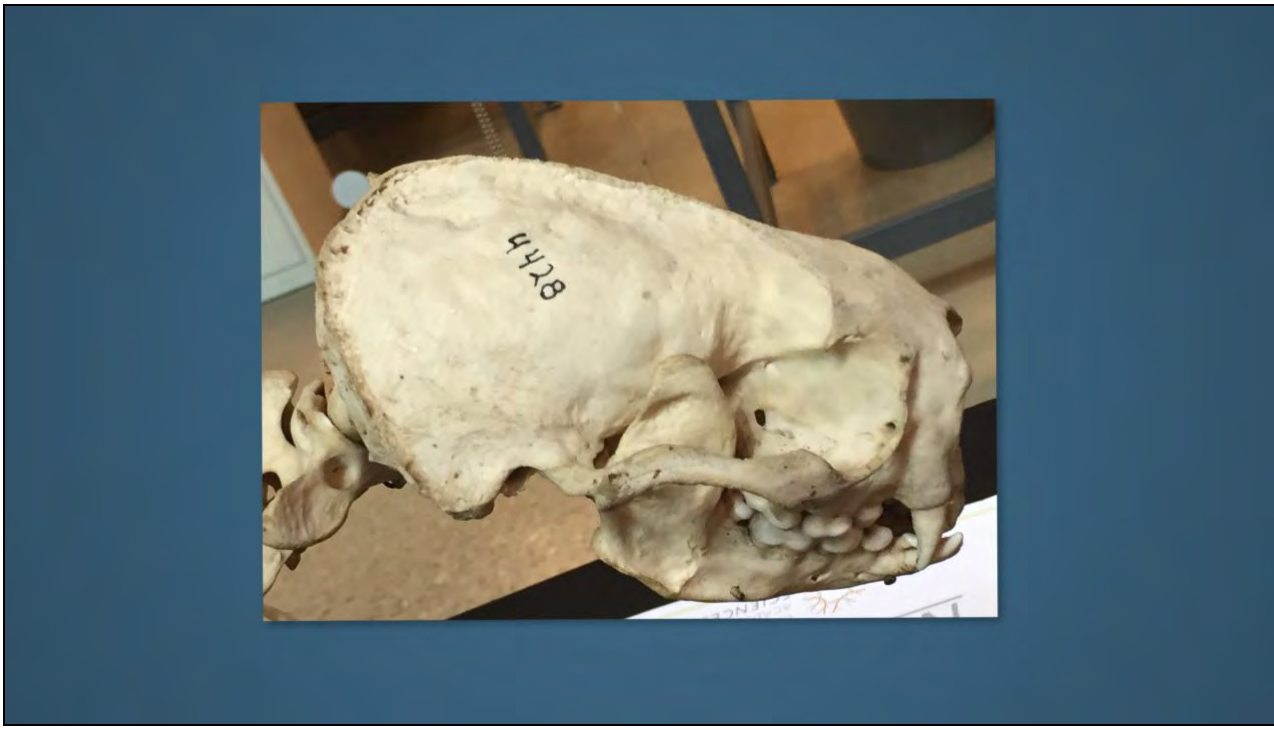
I have a few more examples, and these come from the Academy's 2015 production, "Habitat Earth."

We also wanted to highlight the role of sea otters as a keystone species in California's kelp forests.

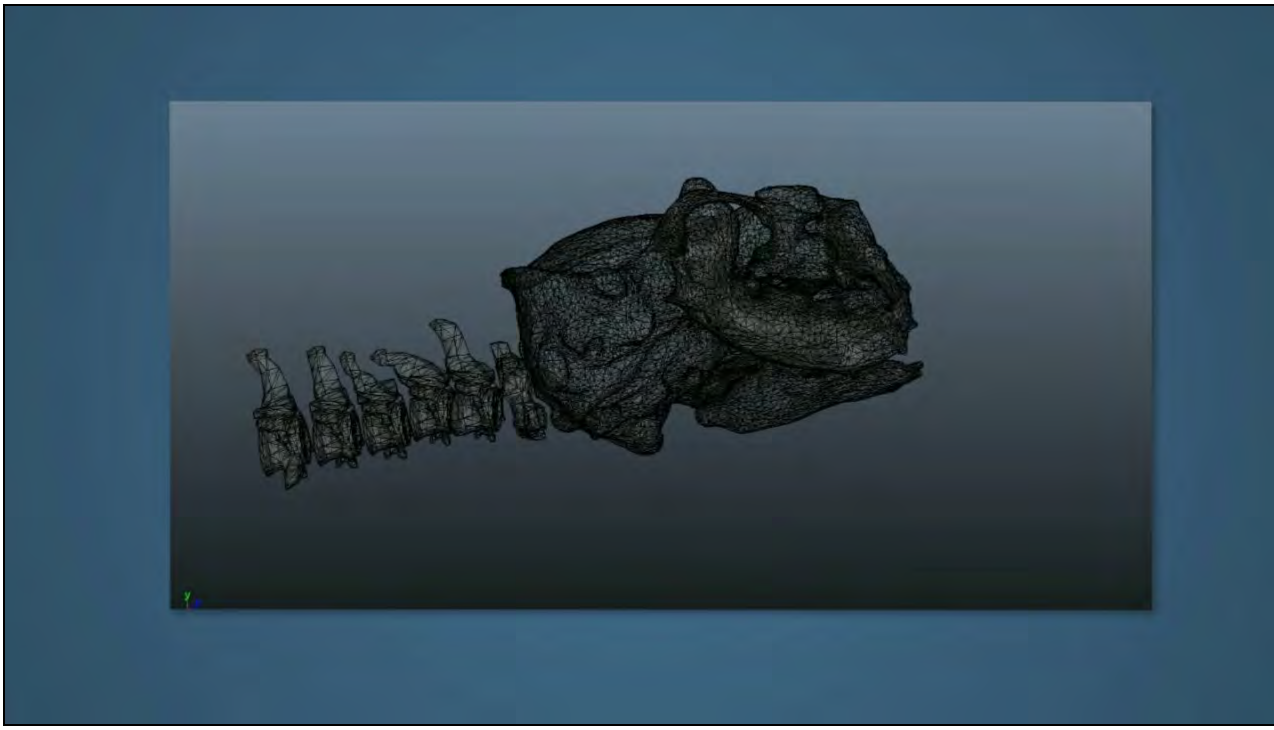
[VIDEO]

Animating the otters… These are truly Academy otters! Digital creations that relied on the Academy's collections to ensure their accuracy and authenticity.

We then went to the Academy's collections to reference otter skeletal structure…



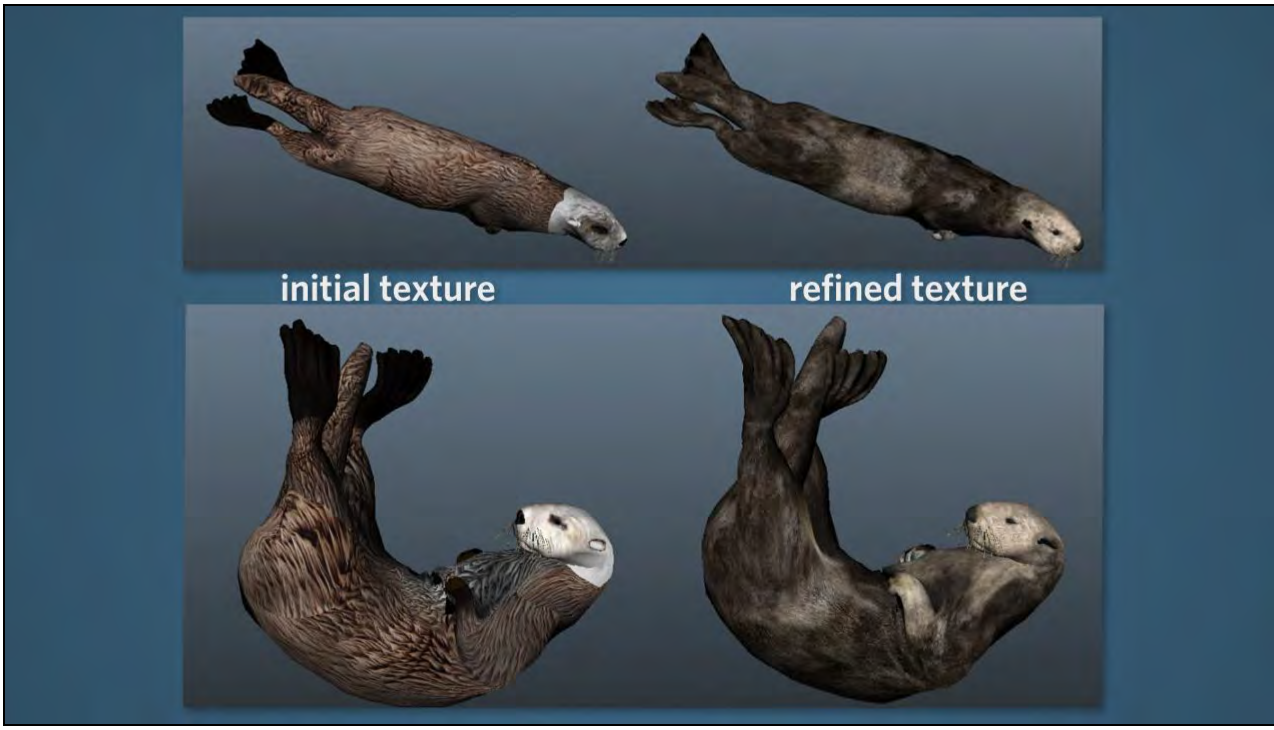
As well as the otters’ skull shape and articulation.



[VIDEO]

Photogrammetry of the otter skull allowed us to create a 3D digital model that was then articulated and used to refine the fleshed-out otter model.

We also accessed the Academy's collection of otter pelts to create a more realistic otter.



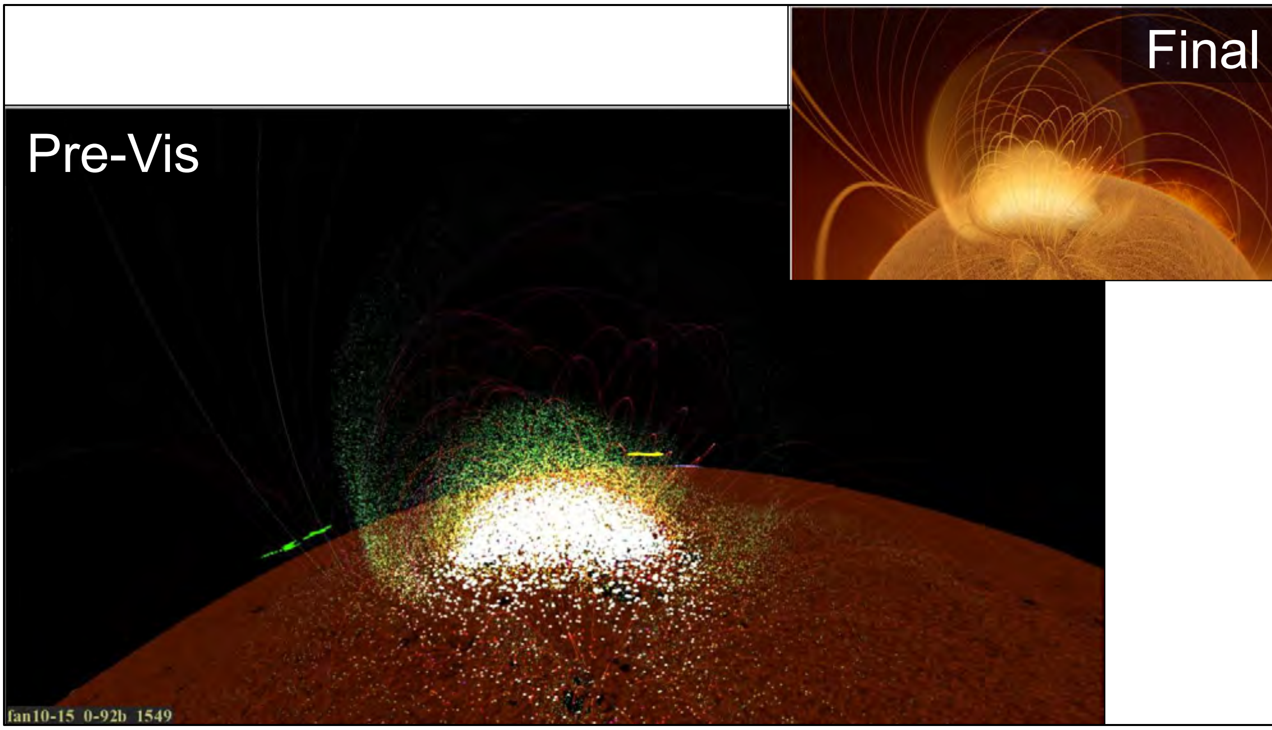
High-resolution photography of the otter pelts provided a refined “texture map” for the digital otter.

Visualization **Production**

Data Pre-Visualization

Data exploration is a critical step that lets you gain insight into your data.

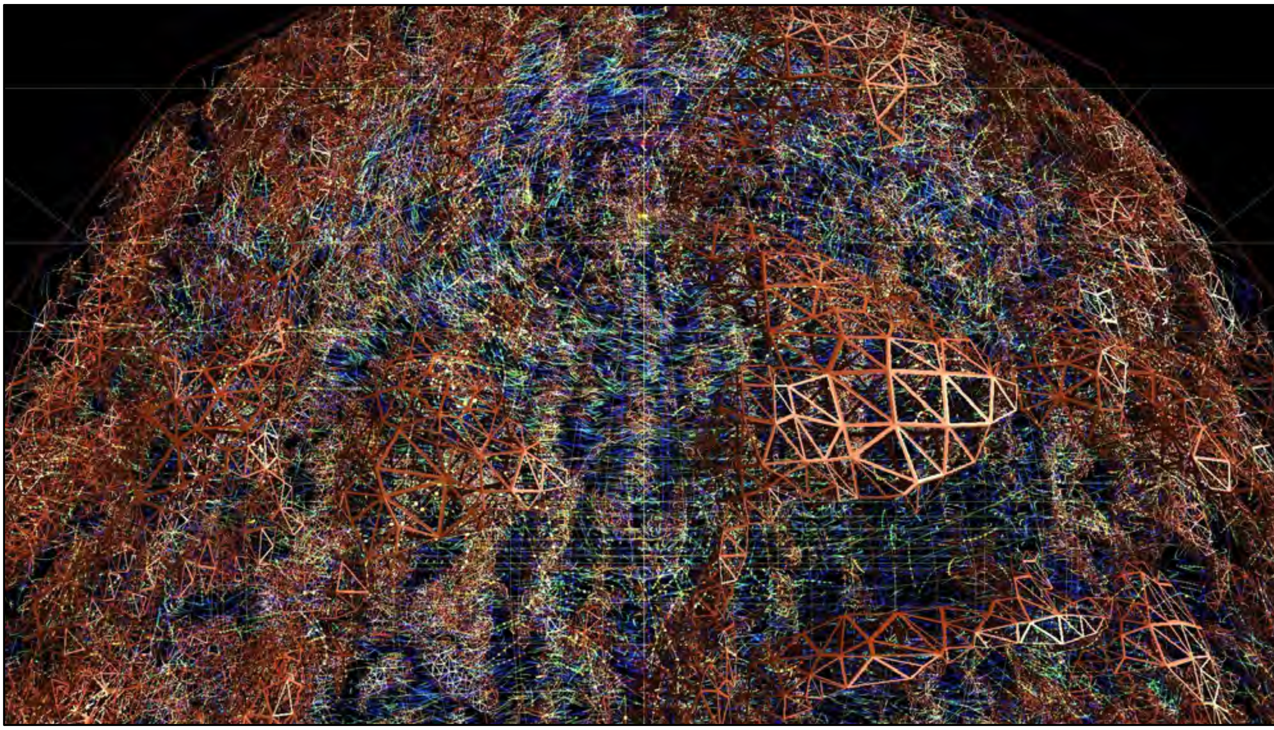
If your data files are large, it may take minutes to load in a single timestep, or an hour to render a single image. It’s not very easy to get a first look at your data or get a sense for how the data evolves over time at this rate.



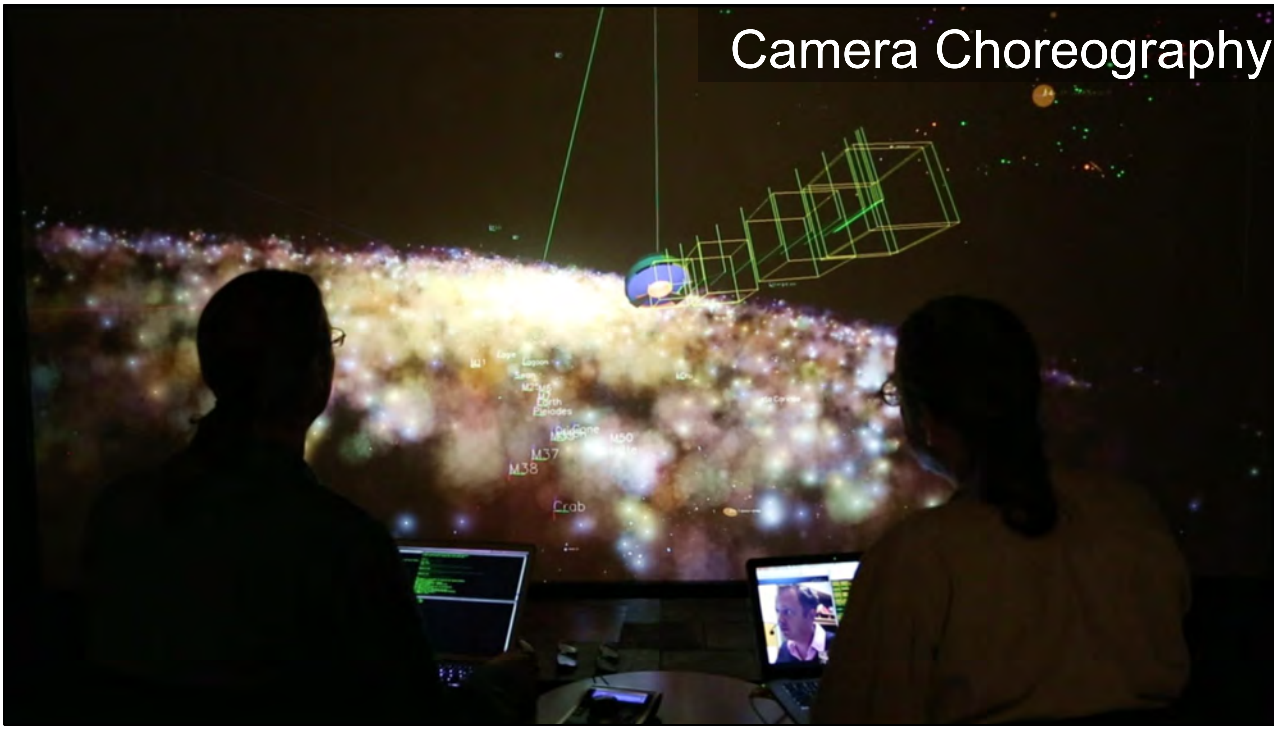
Pre-visualization is critical in data exploration, getting a first look at the data, and creating camera paths.

Since we are working with real data, and not just creating exactly what we want, we need a good way to see what that data is doing. But, the size of each data file is large and can take seconds or minutes to load. So, we sample down the data so that it fits into memory, and so that we can render it interactively. The goal is to have enough information to compose the frame and create camera moves, which we can then export and render with the full data-set. We also use pre-visualization to dial in the desired data rate, and set the length of the scene, so we can interpolate accordingly.

The tool that we use for this is an in-house, open-source tool called Partiview, created by one of our team members, Stuart Levy. Partiview can render star-like particles, lines, and triangles, so we subsample and convert enormous datasets into these geometry types.ndering.



In addition to subsampling the data, we can also create proxy geometry. This is an example of a triangulated polygonal isosurface derived from volumetric simulation data.



Previsualization is also an ideal use of immersive technologies like stereoscopic 3D and virtual reality because it improves your spatial awareness in a scene that can feel very abstract. We almost always explore and choreograph in an immersive environment.

We often use PartiView in conjunction with another piece of in-house software, called Virtual Director. Virtual Director is a tool for both data exploration, as well as camera choreography. We've been using it since …

AJ introduced the concept of Renaissance Teams. It’s important for us to be able to work with our multi-disciplinary collaborators, and Virtual Director helps us do that across distances, by being remote-collaborative.
Each person has their own view of the dataset, and you can see where the other person is flying around.

Technique: Camera Design

One of the major arguments for doing scientific visualization in a cinematic style is that we can take advantage of audiences' familiarity with filmmaking techniques - techniques that were pioneered with physical moviemaking equipment like film cameras, fresnel lights, and ALL the limitations of the physical world.

When we're trying to design a visualization that immerses our audiences and suspends their disbelief, we're the most successful when we follow the rules Hollywood filmmakers have established for themselves.

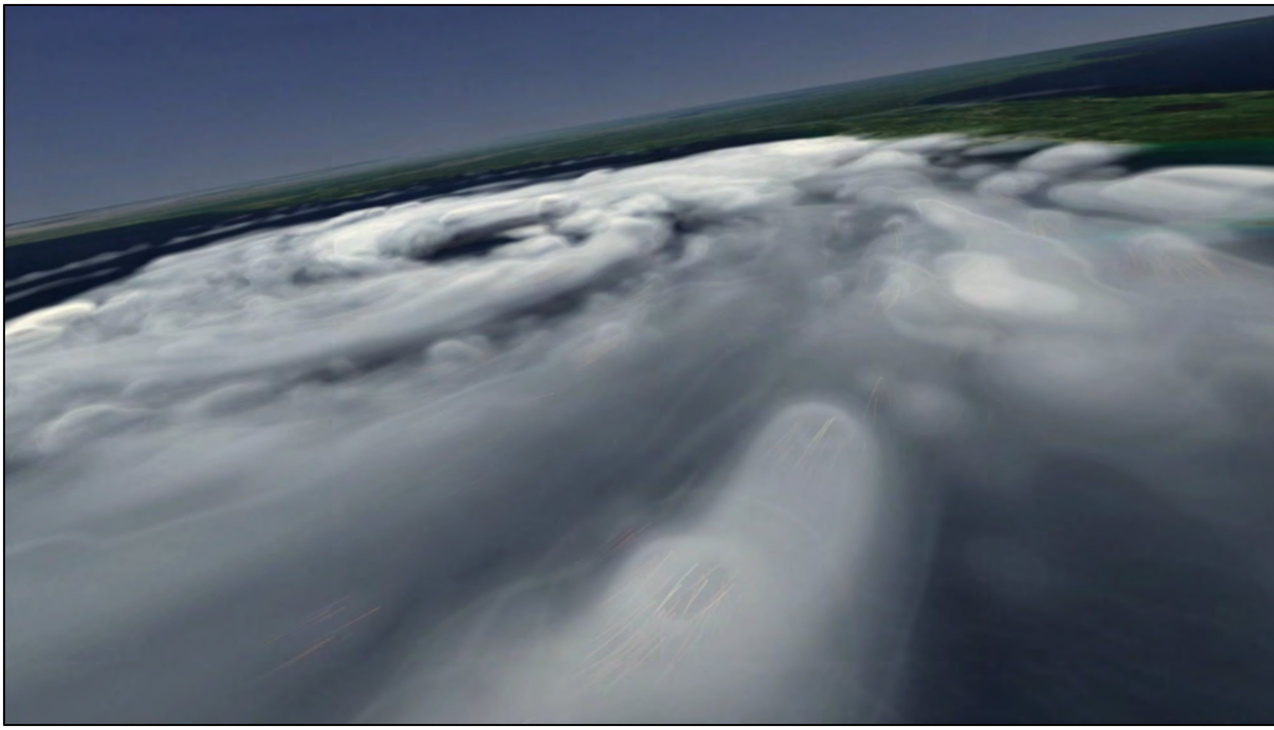
Or at least follow the same rules as much as we can, considering we're touring data environments that humans and cameras could never go.

Parallels to Live Action Cinematography

- Rule of thirds
- Camera at the height of a human
  - Look up to make object seem large, down to make it seem small
- Motion blur, depth of field, etc
- React to action, don't preempt it
- Speed of camera relative to speed of data
- Camera inside or outside of data boundaries
- Interocular distance

Many of the rules of photography and cinematography that you're used to hearing about at SIGGRAPH are relevant to scientific visualization as well. However, while most narrative animation is analogous to scripted drama and comedy, scientific visualization is actually more akin to filming a documentary.

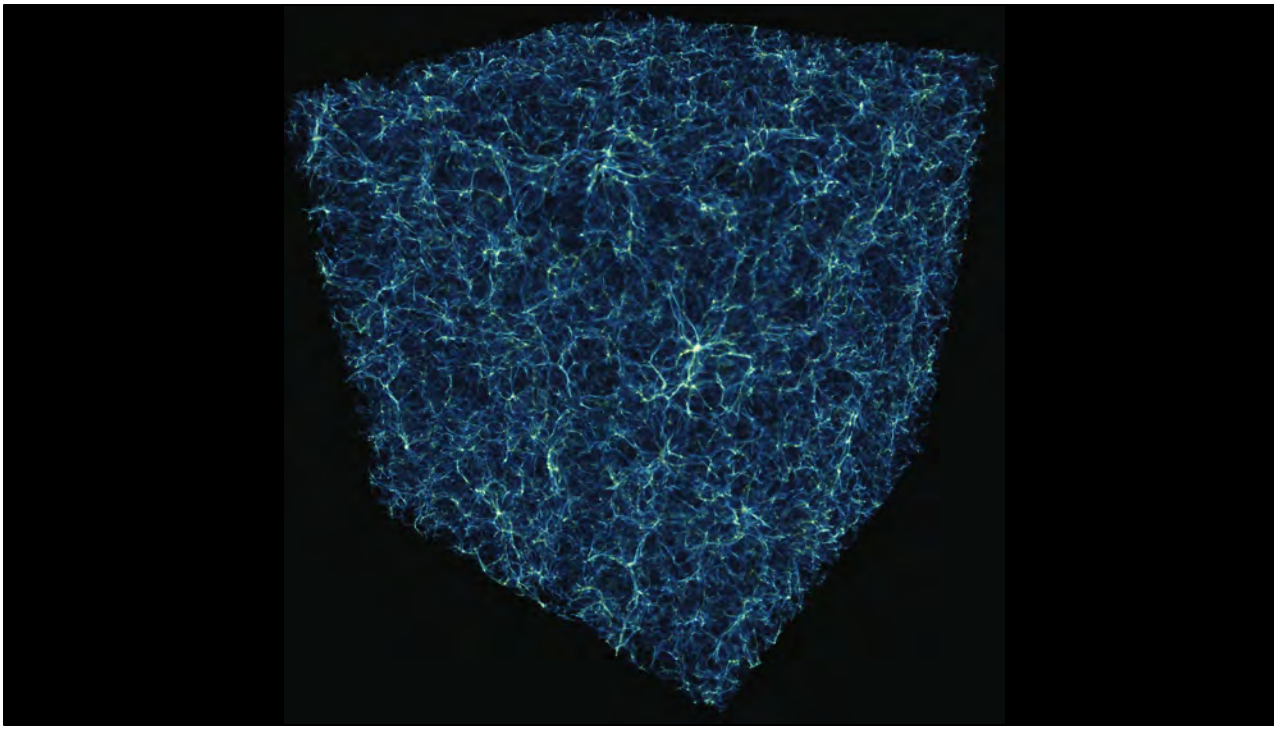
So for instance, you will still want to compose your shots following "the Rule of Thirds", respecting the 180-degree line, and carefully choosing closeups and wide shots, even if your characters are molecules instead of humans. You still absolutely want to incorporate motion blur and depth of field in your renders.



But you might also want to employ cinematographic techniques from documentary film. The motion of your camera should feel motivated. You shouldn't move your camera before an event happens in your data, but rather you should react to it, only moving the camera after you've noticed it happening. This can be a difficult effect to keyframe, which is why performative camera choreography can be valuable. This is why our team prefers to design choreography in Virtual Director.

It also helps to immerse the camera in the logic of the scene. You might imagine that your camera is mounted to an aircraft to give it a sense of physicality and momentum. Or you might have it get swept up by the winds of a hurricane or the gravitational pull of a galaxy. To get everything you can out of the data, you also want to make sure the camera isn't moving faster than the features in the data.

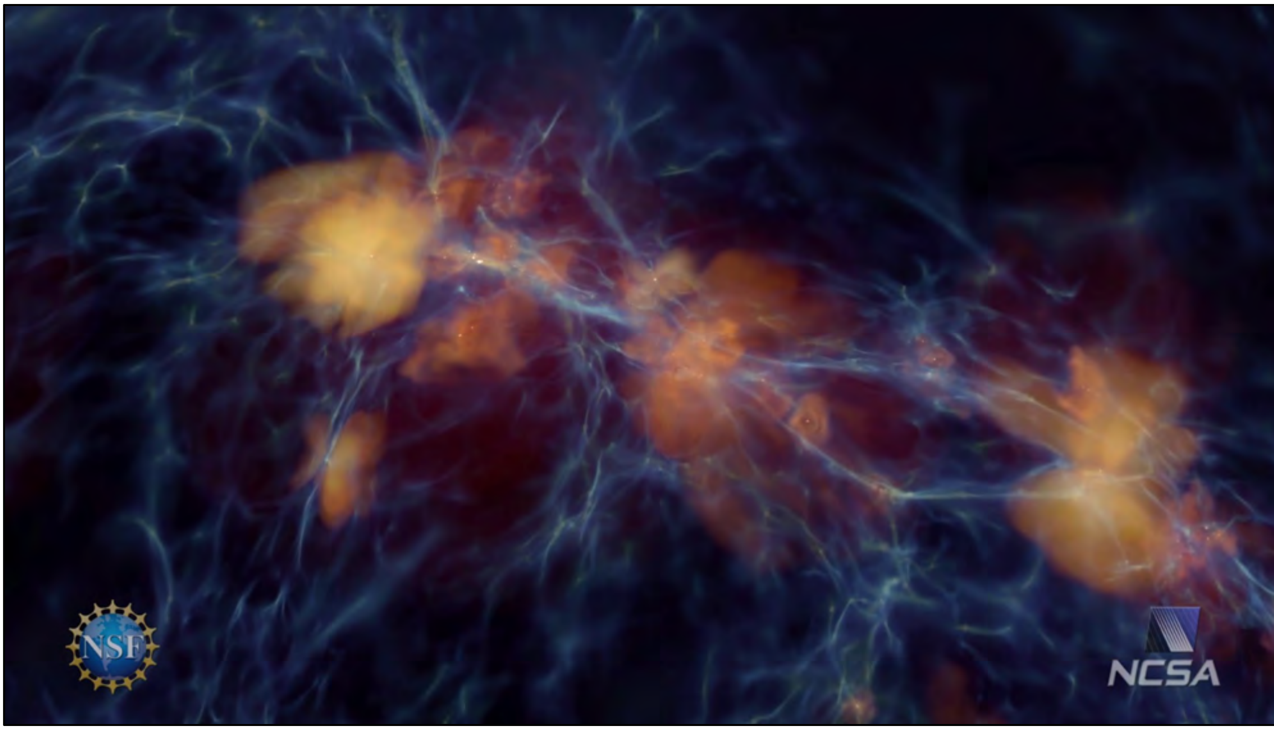
At some points in a data simulation, you will also find that not much dynamic is happening. An active camera can compensate for a lack of data motion to keep the scene consistent and exciting for an audience.



Another major factor in data cinematography is how close you get to it.

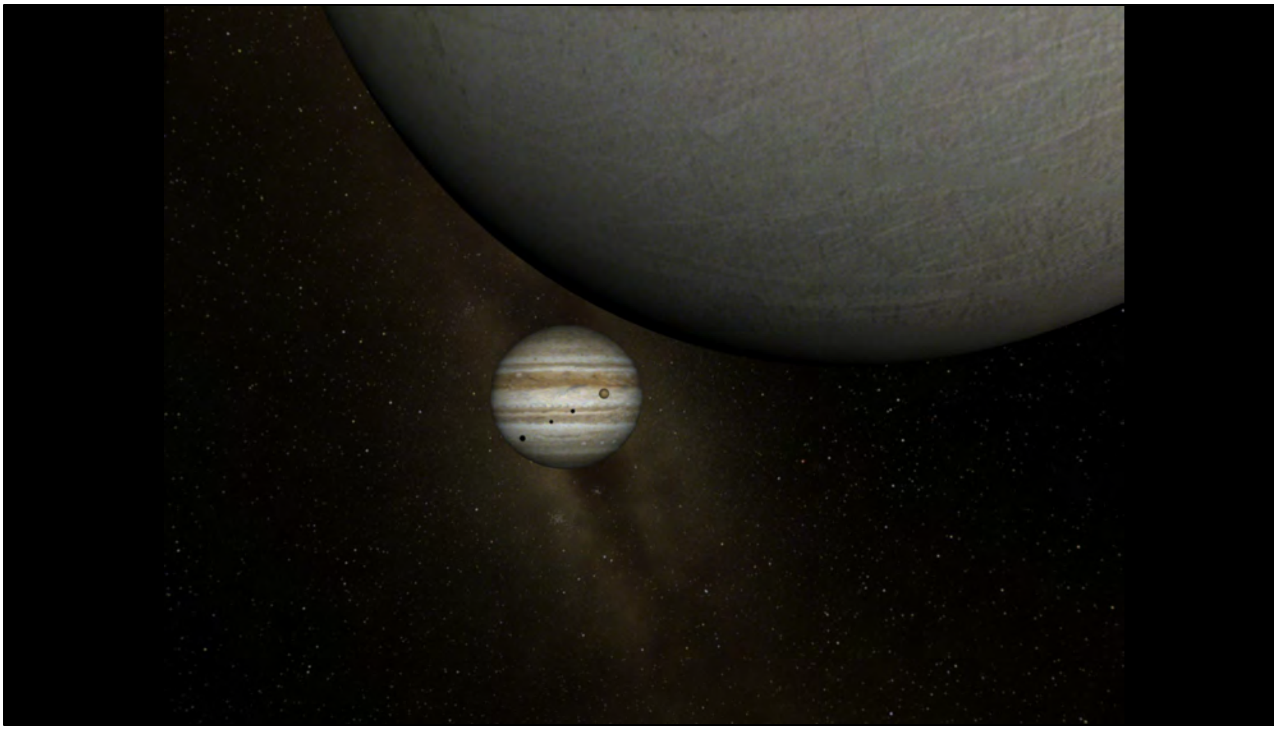
Many more objective visualizations will observe the data from a distance, outside "the box".

All datasets have borders or limitations, because you can't simulate or collect the whole universe! The edges of the box aren't artifacts, but you may or may not want to hide them, depending on your intended audience. If your primary audience is scientists, it can be *beneficial* to show the edges of the box, so that they can see exactly what is and isn't being studied. But if your audience is the general public, seeing the edges of a simulation that's supposed to represent the universe may be confusing and distracting from the narrative. Hiding the edges is especially a good idea if you are trying to make an immersive visualization.



But often our team likes to get right inside of it, getting as close as possible to feel more like a person in a spaceship exploring the data as an environment. Hiding data boundaries isn't always easy - you can fade them off, and sometimes they are periodic so you can instance them. Other times you might create an artistic element to make the background continuous.

There's also a concern of data resolution when getting close. Often if you get extremely close to a simulation of limited resolution, you can expose pixels, blurry surfaces, and volume hatching. While you can keep the camera further away, often it will be preferable to up-rez the data using out of the box techniques, or by incorporating extra artistic elements or algorithmic noise.



Another common issue in the cinematography of scientific visualization is that you don’t often have a great reference for the scale of objects. In Hollywood, you can put a human next to anything and immediately understand how large it is.

Playing with the height of the camera is a good way to convey the scale of your subject matter. By placing the viewer high above or down below the subject, you can make the subject seem huge or tiny.

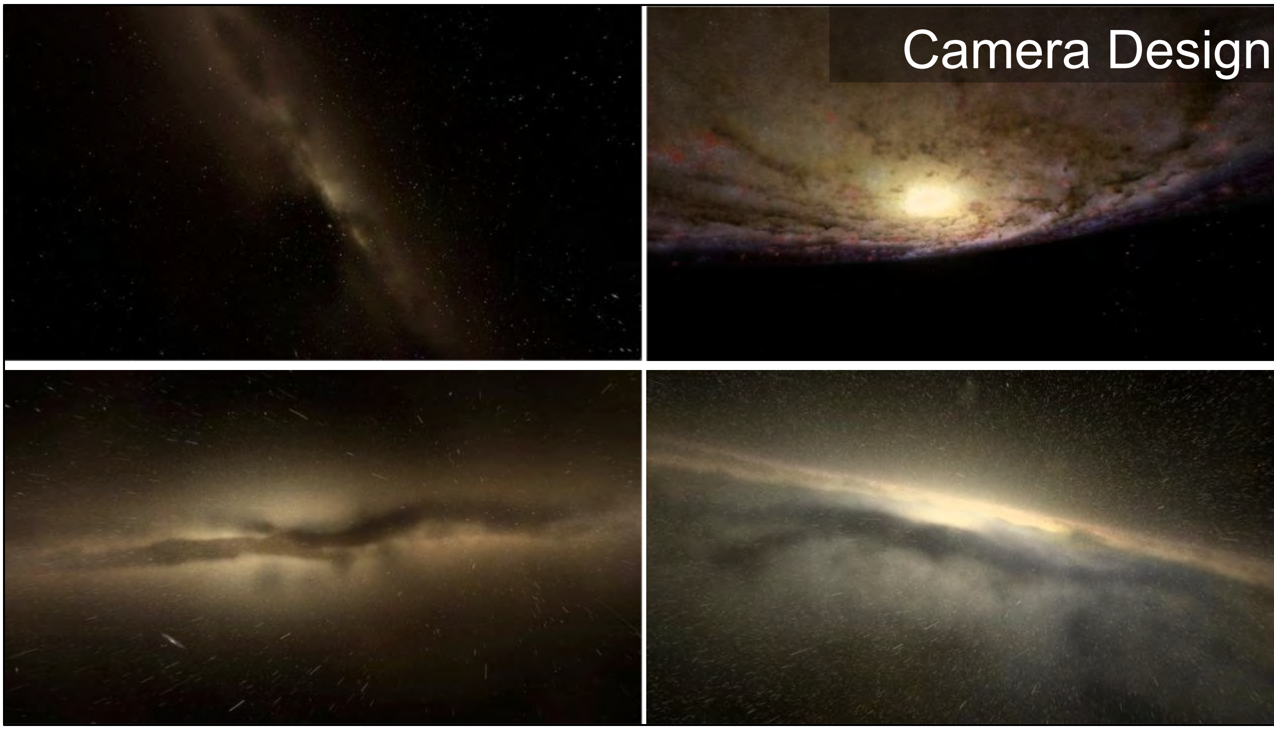
Another way to convey scale is reducing the speed of the scene. This means slowing down the rate of your data timesteps, which might require some more or less extreme interpolation. This also means taking a deliberately slow camera path, and focusing on movement that reveals motion parallax.

We can also make things appear small using a shallow depth of field.

Cinematic visualization scenes tend play out as unbroken seamless journeys to create a sense of immersion and also to help maintain context, a sense of scale and to communicate the dynamics or structure of the science. We try to create a sense of narrative with the camera move while capturing the most insightful and engaging views of the data geometry and timing. Giant screen immersive theaters are able to provide a more embodied experience and sometimes benefit from longer scenes that make you feel like you are there.

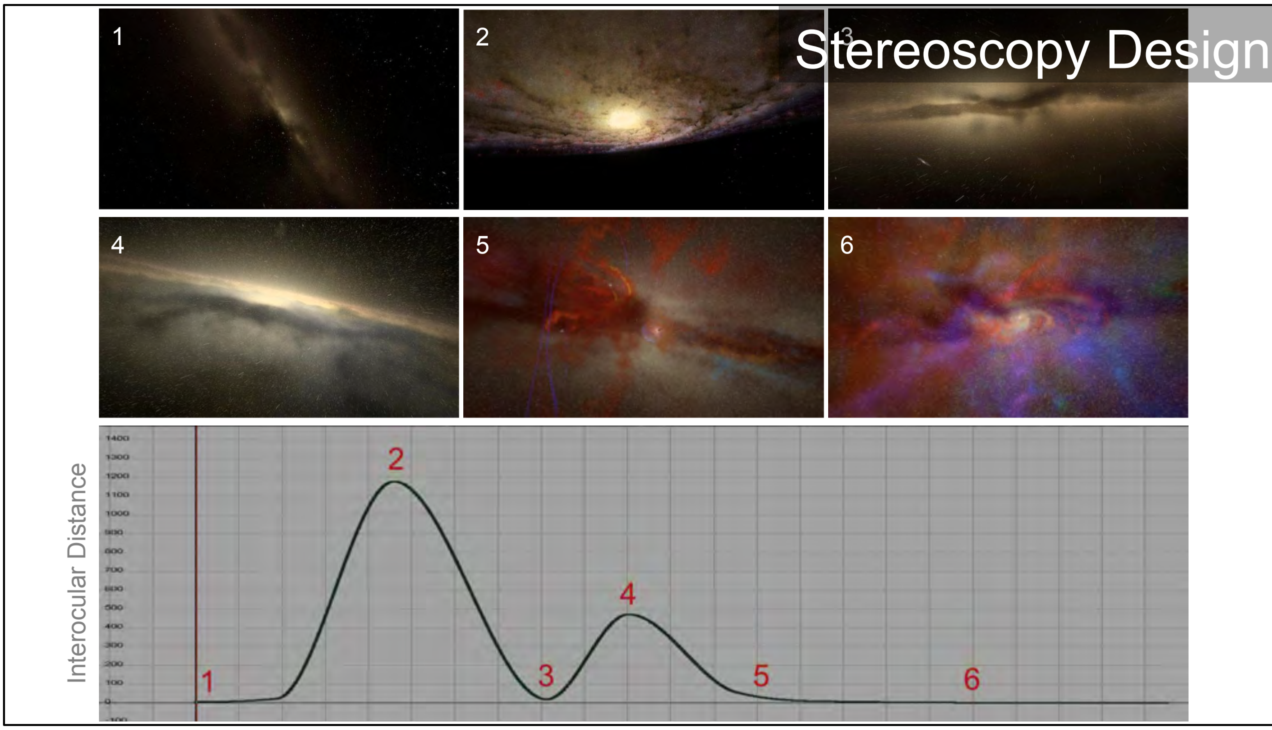
Because of this, large format screens benefit from long unbroken shots. These long shots have the added benefit of providing some breathing room between subsequent events in the simulation data. This breathing room is important to give the narrator enough time to describe the phenomenon, and to give the audience time to digest the information.

Due to their coverage of our peripheral vision, large format screens also make viewer more susceptible to motion sickness. Gradual and smooth camera paths can help avoid this, too.



The long shots have the added benefit of giving some breathing room between events in the data, giving the narrator enough time to describe what's going on, giving the audience time to digest the information, and adding a bit of drama.

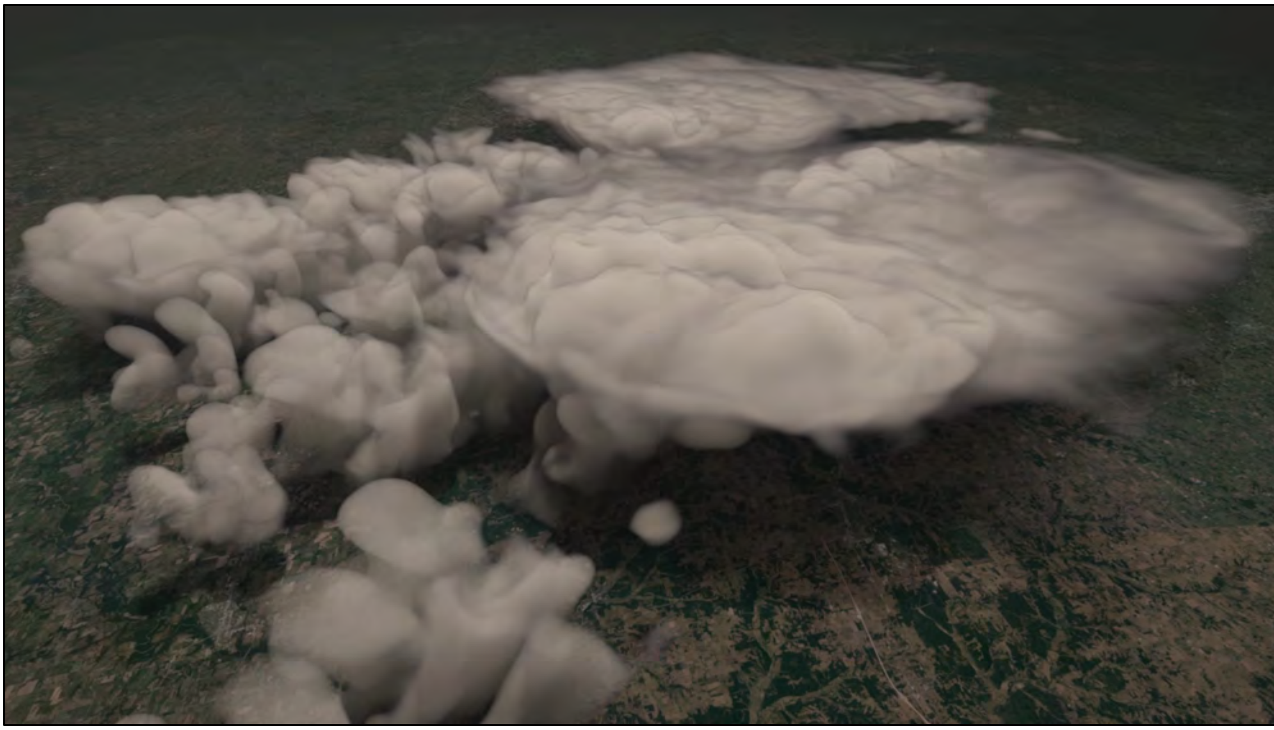
Cuts don't work well in scenes that don't have inherent directionality. Which way is "up" in space? How do we know which way we are facing, or even if we're looking at the same thing after a cut? Furthermore, frequent cuts also confuse our sense of depth in a 3D theater.



In addition to designing the camera path itself, we also consider stereoscopic design for 3D sequences. When dealing with data from physical sciences, be that astrophysics or molecular biology, objects are interacting across VASTLY different scales. When we create a camera move that takes us from looking at the entire Milky Way, to looking around the area of a single star, we have to adjust the stereoscopic depths accordingly, by changing the eye separation between the left and right-eye cameras.

Lighting

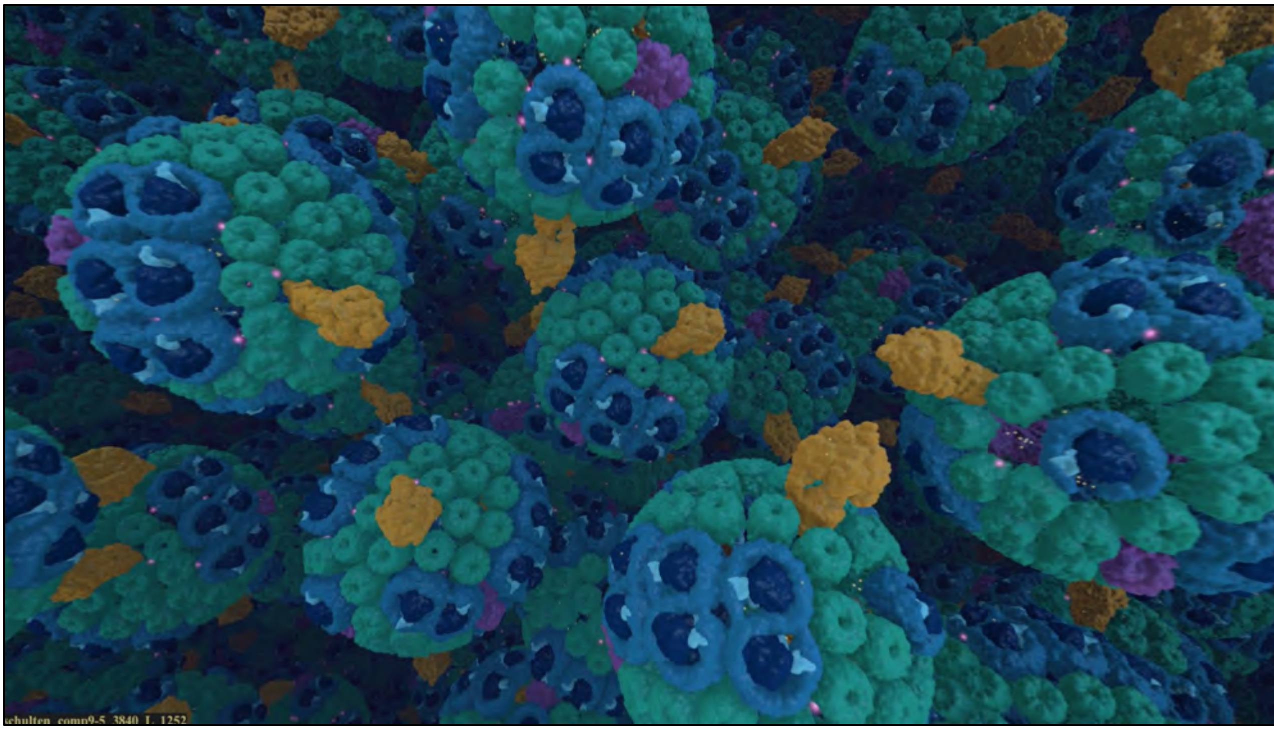
- Can add lighting for dramatic effect, or it can be data-driven / data-inspired
  - Set the tornado at the correct time of day
  - Use temperature and material to drive whether a particle is emissive
- Astrophysical data is often emissive



Lighting a visualization scene is again a combination of narrative animation techniques and documentary film techniques. Typically we rely on more naturalistic lighting than a controlled studio setup. Three-point-lighting occasionally comes into play, but usually when lighting a scene, the question we are trying to answer is: what motivates these light sources.

For a thunderstorm, it makes sense to create a directional sunlight and an environment light for diffuse shadow fill.

But for astrophysics we might find that all of the objects in the scene are self-illuminating, and we don’t want to have any external lights at all. In this visualization of the formation of the moon we had an extremely dynamic lighting setup. First the two diffuse planetoids are lit by a directional sunlight. Then as they collide, the molten surface becomes a morphing geometry light. Finally, as they’re swallowed in vaporized rock, the simulation becomes totally self-illuminating.



And in biophysics, when we're dealing with extremely small objects, lighting an environment is really more metaphorical, because these objects are smaller than wavelengths of light. In this visualization a photosynthetic organelle, we chose to have explicit external lighting because light is part of the story of photosynthesis, not because it was a realistic rendering of the microscopic universe.

Data Artifacts

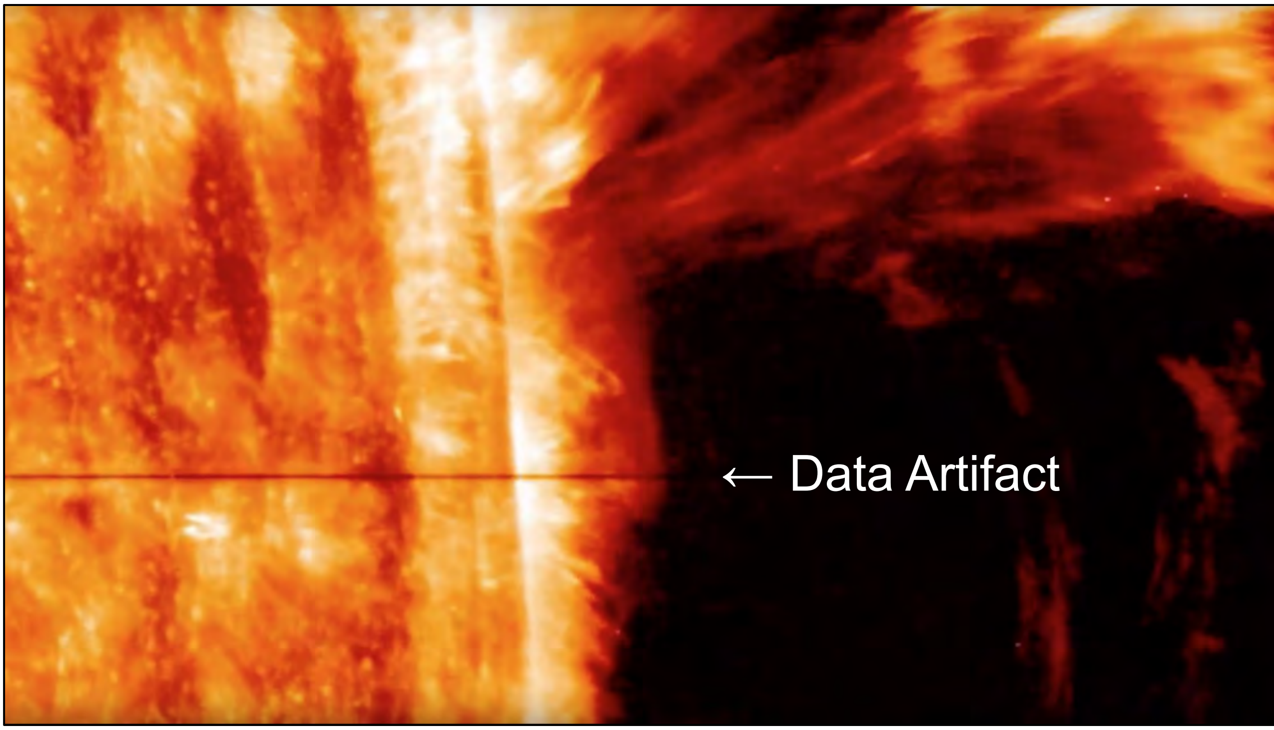
So now I want to transition to talking about how we actually design our visualizations. As visualization designers, we need to make choices about what details to show, and how to show them. Choosing not to show everything all at once, to simplify, and to emphasize features is called “abstraction” and it is a key skill for any visualization designer.

Data Artifacts

- Data collected with sensors is susceptible to noise
- Data simulated on computers is susceptible to bugs and errors
- Time may not be linear
- It can still be problematic even if it isn't an "error"
  - Pieces that are important to the documentary narrative, but not the scientific research, may be left out of the simulation/collection
  - Additional pieces may be present which are important to make the simulation run, but aren't part of the result
- Data-Driven Solutions
  - Fill missing gaps with other datasets
  - Use clever transfer function or camera placement to hide artifact
  - Fix in composite
  - Tile dataset if it's periodic

You may find missing data in some datasets, but in others, you may find additional data that seems like it doesn't belong. (Boley)

Regardless of whether or not the artifacts are errors or are intentional, they can often be distracting and can look strange in a visualization, so you may want to take steps toward hiding or minimizing them.



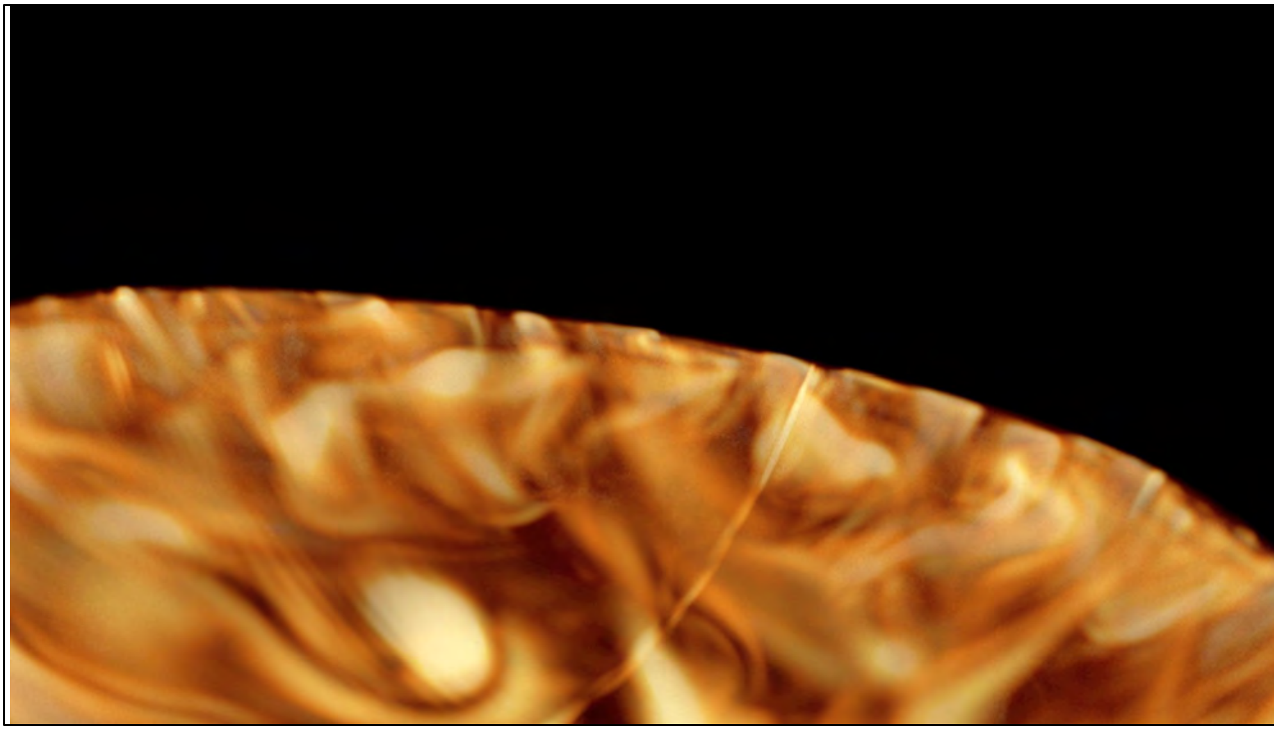
This is a video of the surface of the sun, taken a NASA telescope. But what’s that strange black line moving through the image?

The different sources of data we use for visualization can produce different kinds of data artifacts. Observational data collected with sensors, for instance, is susceptible to interference, noise, and other errors.

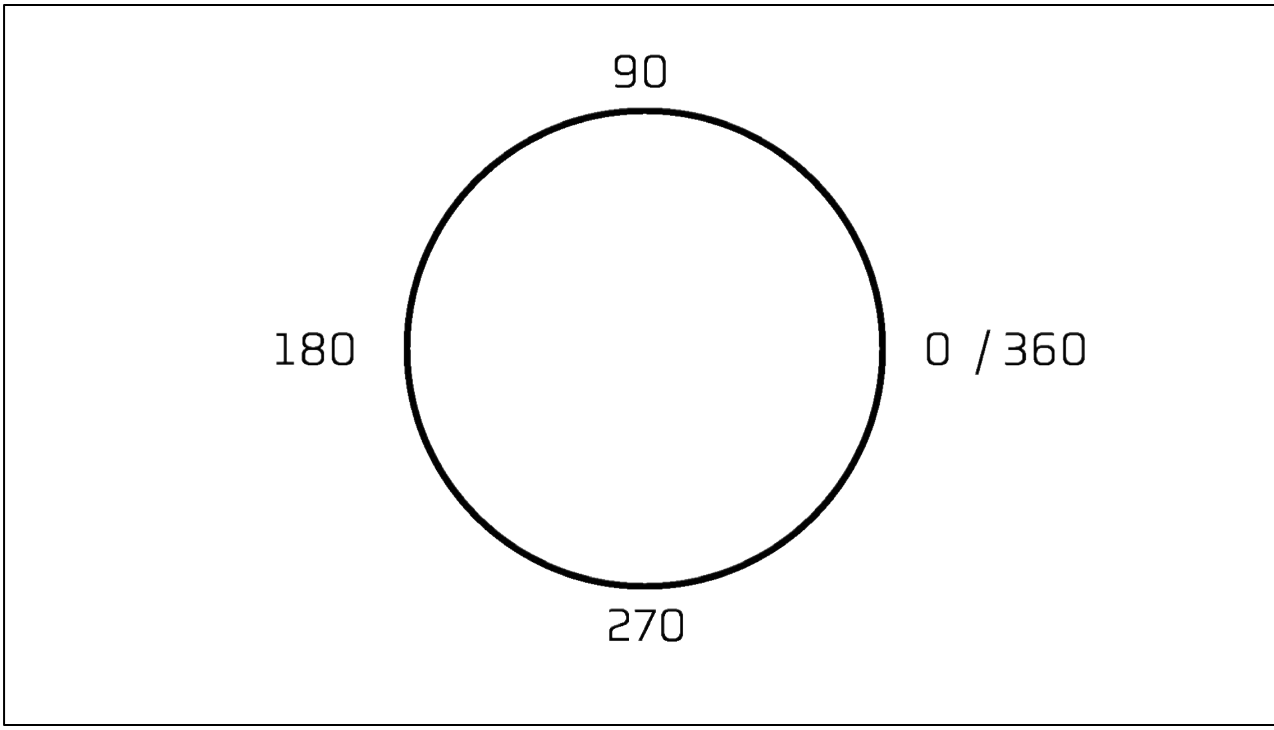
Data that is simulated and not collected isn't susceptible to these same types of collection errors, but can still produce data artifacts. Sometimes, these can be the result of an error in the scientific simulation, or a mistake made during the data processing step.

This volume hatching error is common in multi-resolution volumes when one level of resolution goes missing.

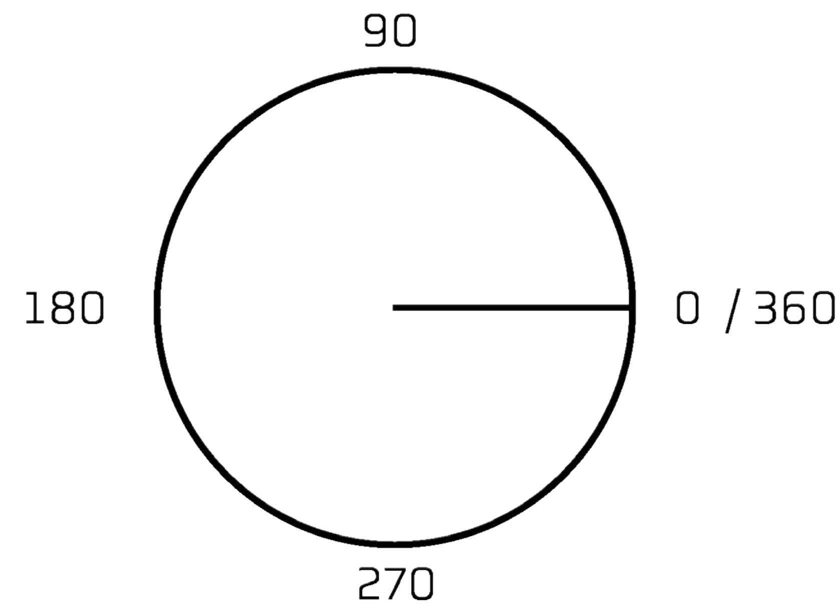
We were able to fix this error by replacing that one level of the grid with the same level from an adjacent timestep.



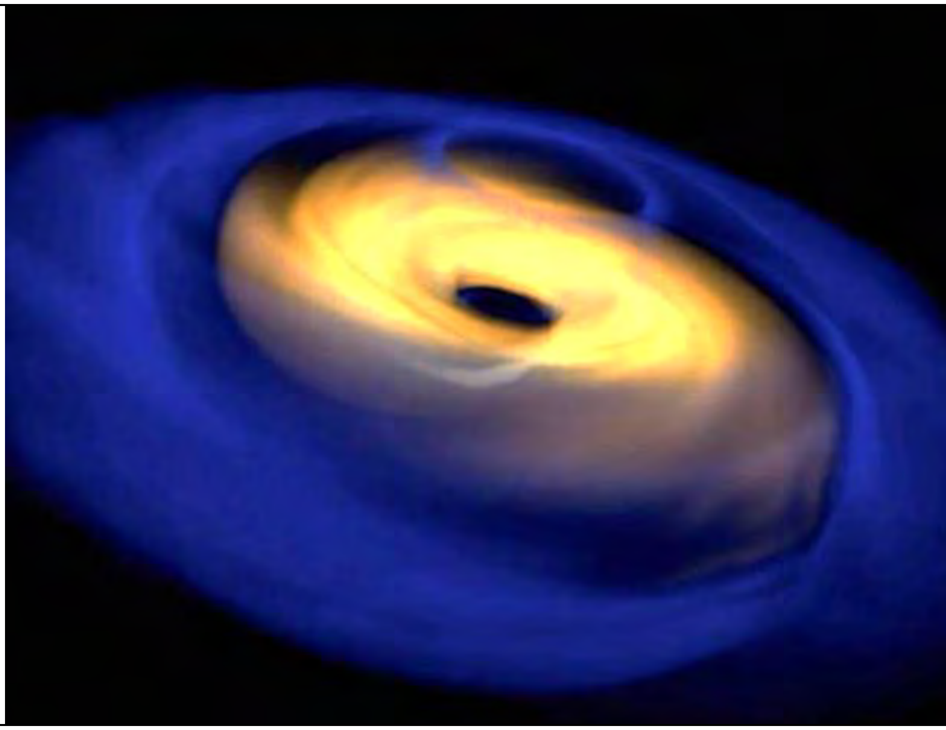
In this early draft of a visualization, we see a seam in this volume. This artifact is caused by an error in translating the data from spherical to cartesian coordinates.



On a circle, 0 degrees and 360 degrees refer to the same position.



This artifact was caused by an accidental overlap where this was not considered.



You may find missing data in some datasets, but in others, you may find additional data that seems like it doesn't belong.

The dataset of this visualization of a protoplanetary system actually has a second ring floating above it. This was needed to make the simulation run properly, but the scientist told us that we should ignore it because it wasn't physically accurate.

We were able to hide it by locating the specific density values that represented it.

We often encounter issues with inconsistent timing as well. This is often intentional on the part of the scientist - they write out datasteps for the parts of the simulation they care about, which are not necessarily evenly spaced.

We use many re-timing techniques in animation and compositing software to try to smooth out the playback rate.

Compositing

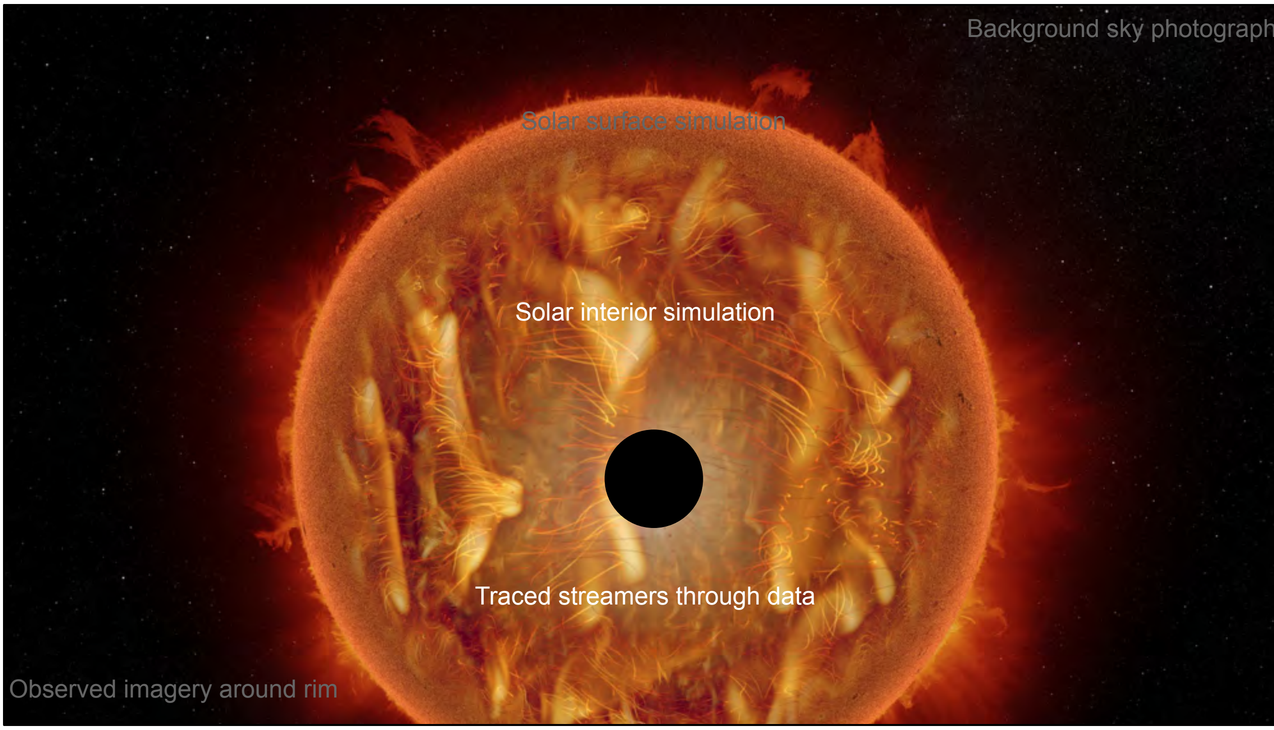
We often use compositing both for 2D image manipulation and combining datasets. This Nuke composite demonstrates combining datasets in a pleasing way.

This visualization has us looking through the surface of the sun, into the interior. This X-ray view required combining several different datasets from different scientists.

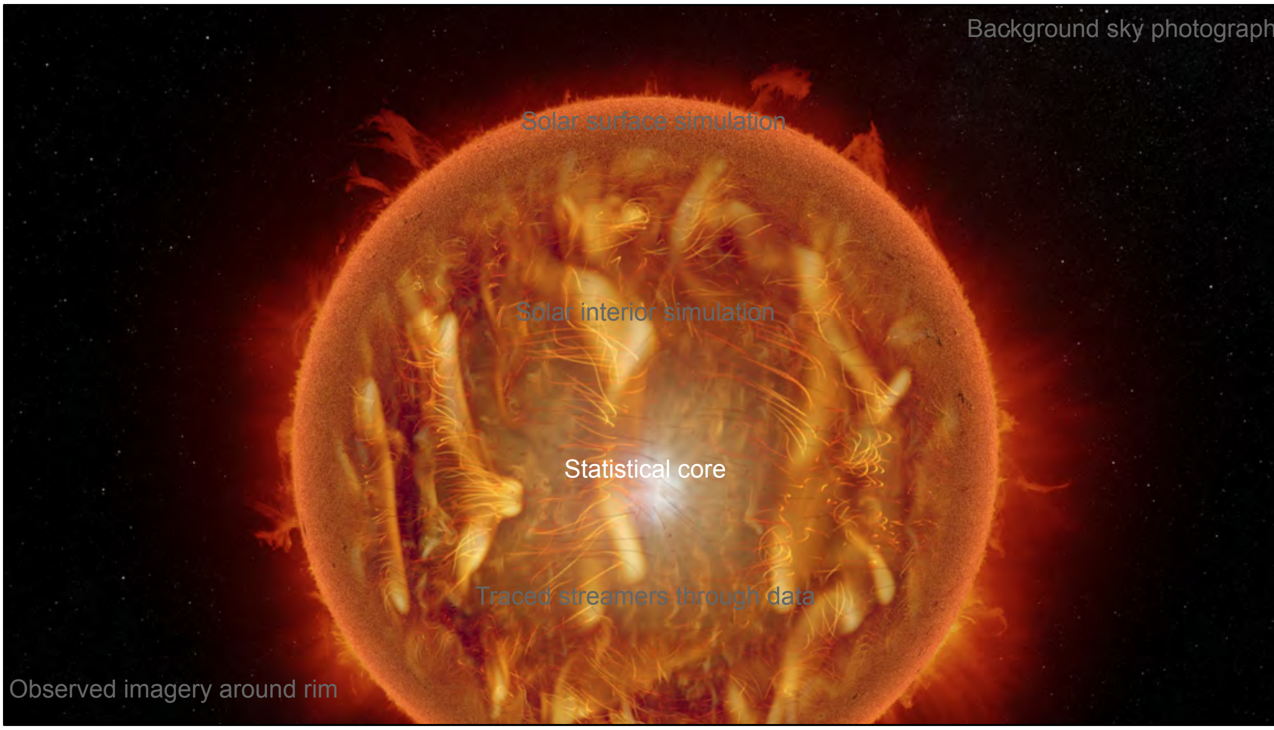
We began by putting the whole sun in the correct context - the starry background sky image is a photographic mosaic from Axel Mellinger, and it is registered to what we see when looking in the direction of the sun from the Earth.

Around the rim of the sun, we used photographic time lapse imagery from NASA's Solar Dynamics Observatory, a space-based telescope that has several different cameras looking at the sun.

The sun's surface was a supercomputer simulation of solar convection by Matthias Rempel. We mapped this data onto a sphere and then used compositing techniques to make the sphere more transparent where it was facing the camera.

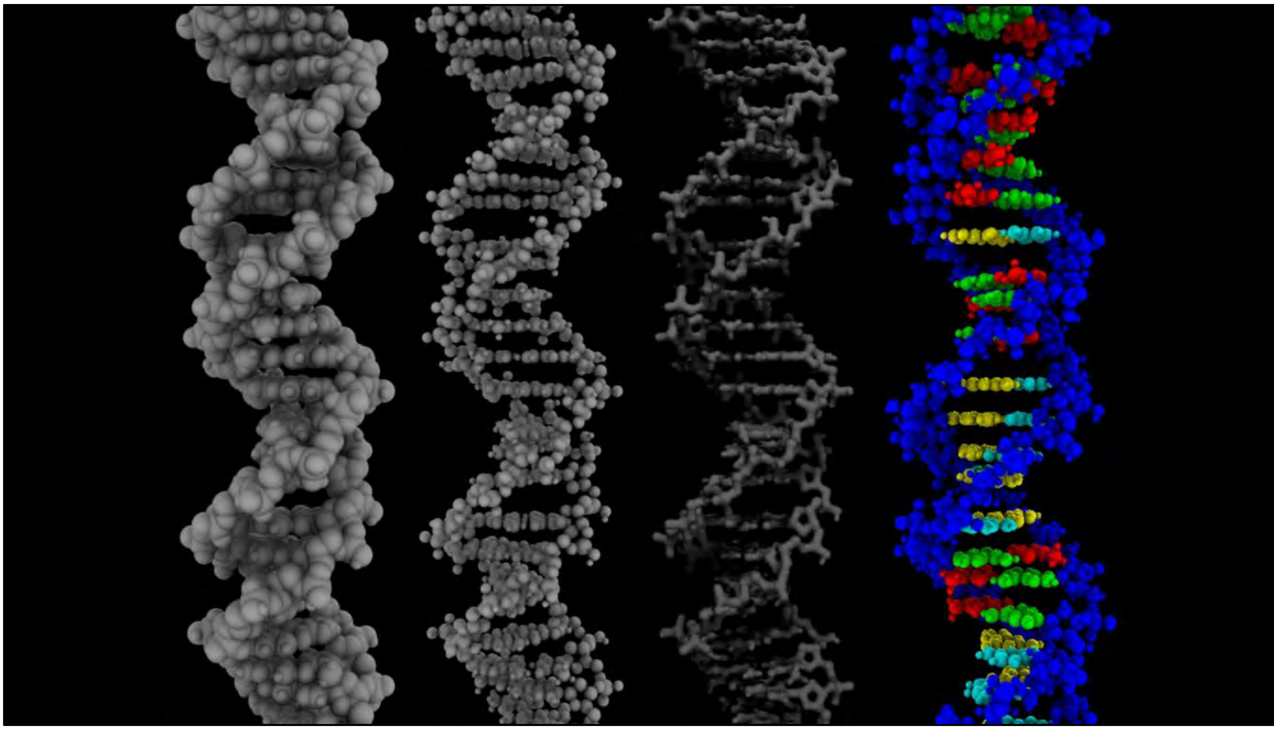


The interior simulation by Mark Miesch and Nick Nelson showed an important feature of the sun's interior, vortices around the equator that we nicknamed "bananas". We traced streamtubes through the vorticity field to wrap spirals around the bananas to show that they were rotating.



This interior simulation covered the outer third of the sun's volume. To fill in the very interior, we registered a fairly plain statistical representation of the sun's core inside of it.

Compositing is also a great way to integrate real-world lens effects, like the way that light interacts with the glass in front of a camera. Here we see the glare of a star wrapping in front of the planet in front of it.



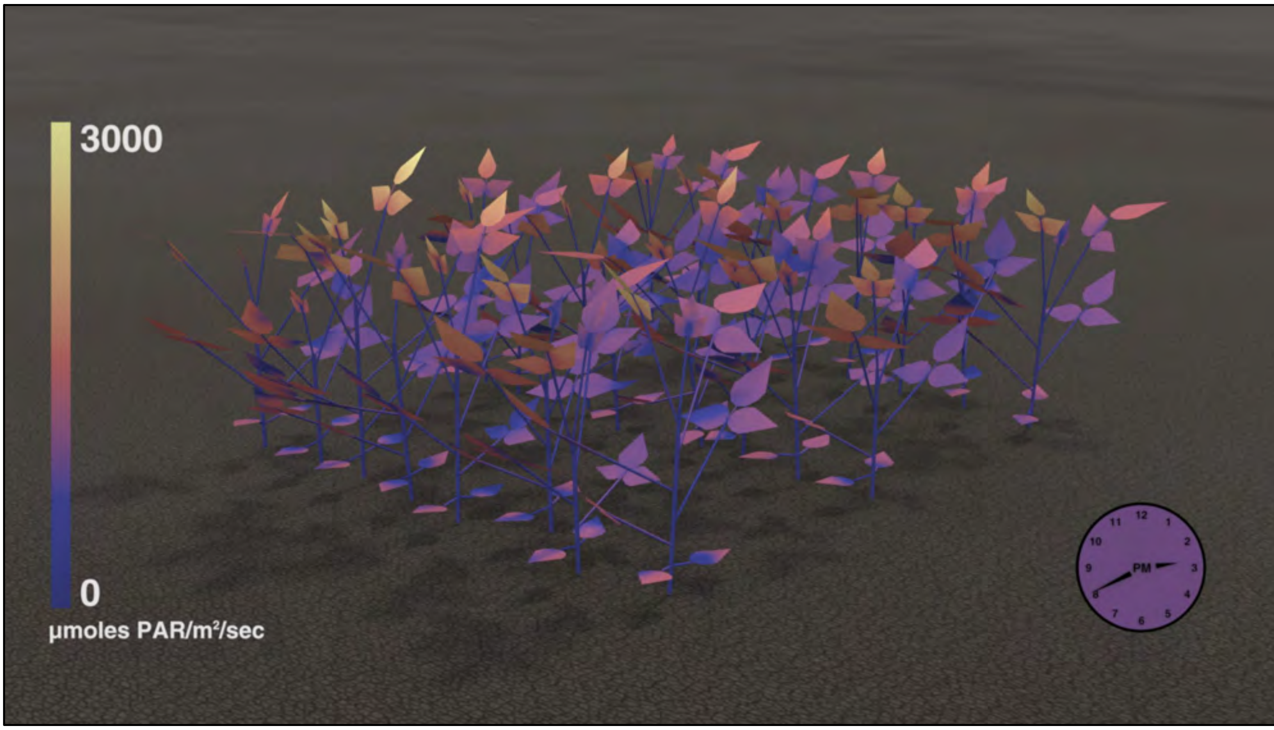
We can also use compositing trickery for more complex effects. By writing a depth attribute as an image layer, we can control depth of field and motion blur in the composite. We can re-color molecular data based on the atom id or the molecule id.

We can re-time a texture, or re-frame an immersive scene.

Packaging & Distribution

Packaging & Distribution

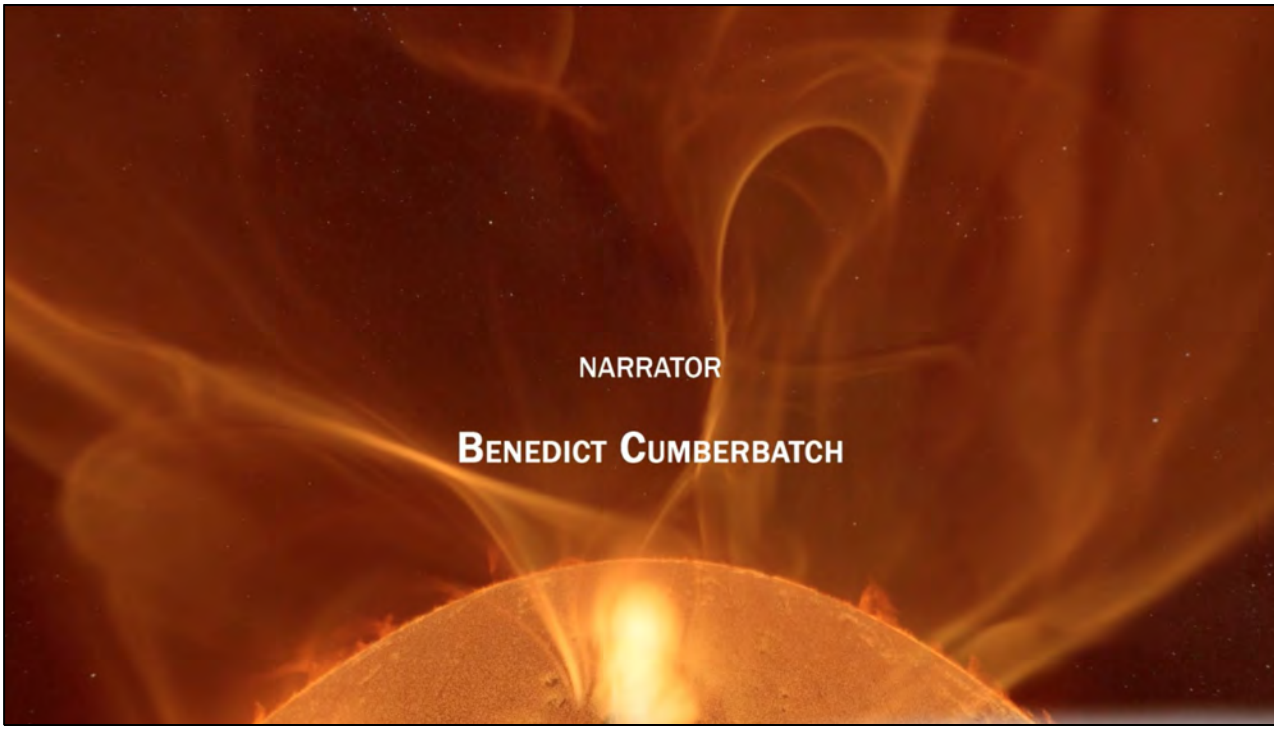
- Add color bars as needed, titles, credits, sound
- Acknowledge collaborators and funding agencies
- Small-scale distribution: YouTube, social media
- Large-scale distribution: museums, television
- Vis should be able to stand alone and speak for itself, it may end up in surprising places out of context



When producing a visualization for science domain experts, it’s a good idea to include an annotation of your transfer function that shows the color map and value range over the top of your visualization.

You may want to add labels that describe what the audience is looking at at different points throughout the video, or a time bar, or other titles and annotations.

In a cinematic setting you might choose to avoid these more didactic elements though.

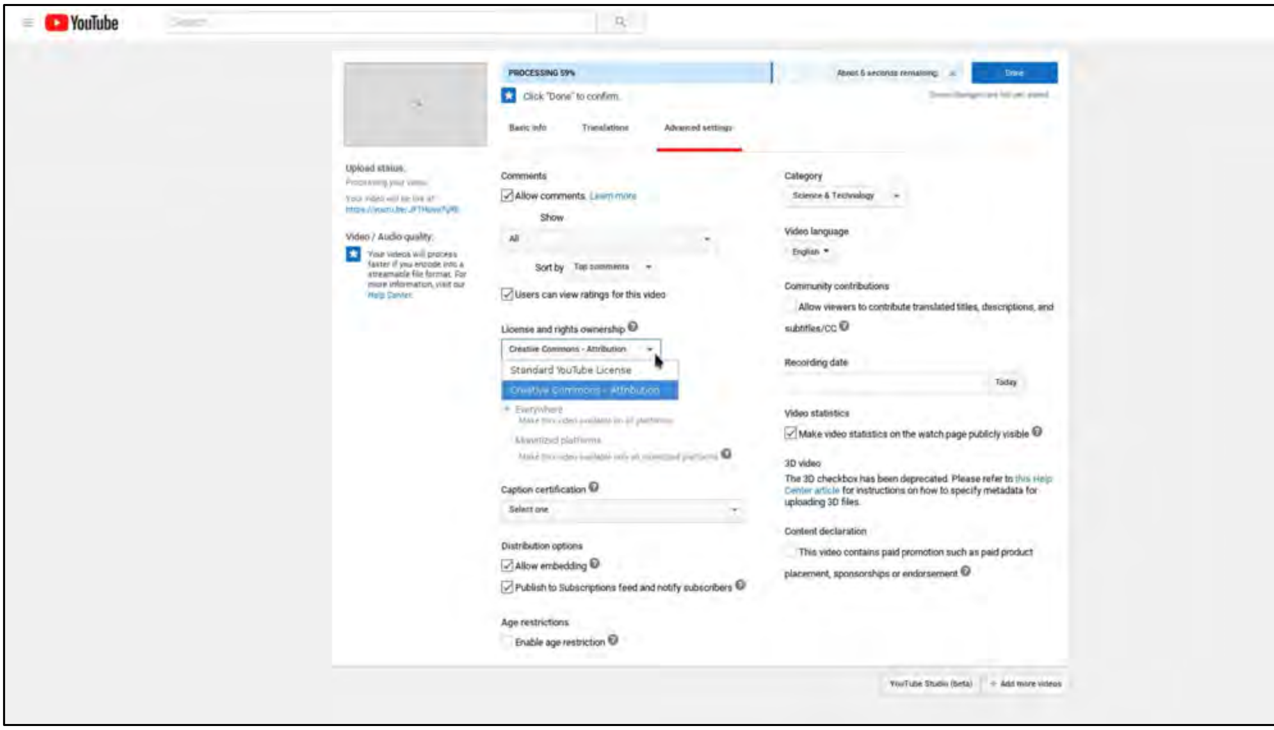


Visualizations are less compelling without sound. But bad sound can ruin an experience immediately, so it is important to spend time on a high-quality soundtrack.

The sound accompanying your visualization can have three main elements:

1. Narration
    a. Script: Should be vetted by scientists, and aimed at some target audience
    b. Timing: Timed to the action, and slow enough to allow viewer to absorb information
    c. Quality: Audio quality, as well as quality of voice acting (e.g. not monotonous)
    d. Celebrity narrators are a marketing tool that can pull audiences to see a documentary as well, but this is a double-edged sword. Some celebrities are voices for misinformation movements around scientific fields like vaccines, evolution, and climate, which is troubling because they are not experts in the subject. For our documentary films, we seek celebrities who are enthusiastic about the topics they narrate, and who respect their role as ambassadors for scientific research. Past narrators like Leonardo DiCaprio, Jodie Foster, and Morgan Freeman all have proven themselves to be science advocates before and after the films we've collaborated with them on.
2. Sound effects
    a. Used for emphasis and/or realism, can be data-driven as well
3. Score
    a. Can set the overall mood and raise the production quality

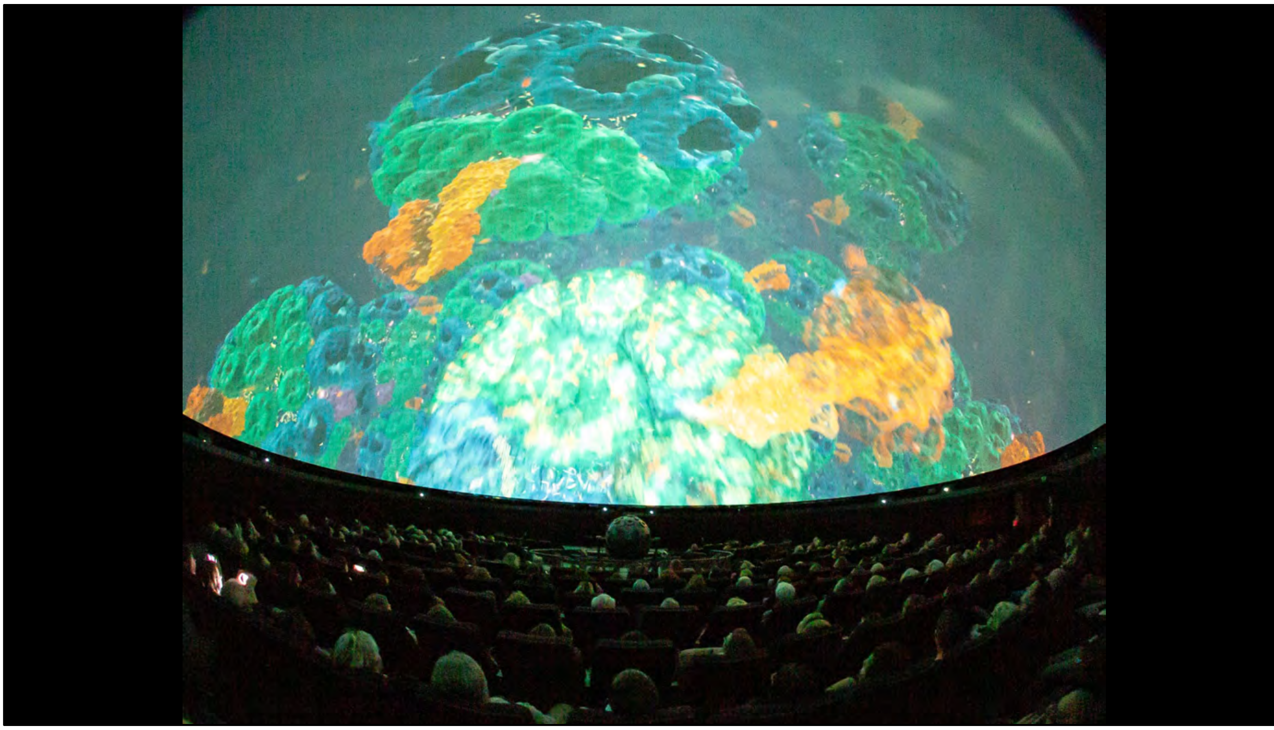
Of course, credits are absolutely necessary to acknowledge the scientists, visualization designers, and funding agencies. This kind of outreach can help scientists get funding so it's important to give them proper attribution.



Start thinking about how and where to share your video. You can share it yourself on a small scale, by uploading it to a video hosting service or social media, or publishing it as supplemental material for an academic paper. Some services like Vimeo allow you to sell a movie rather than just provide it for free.

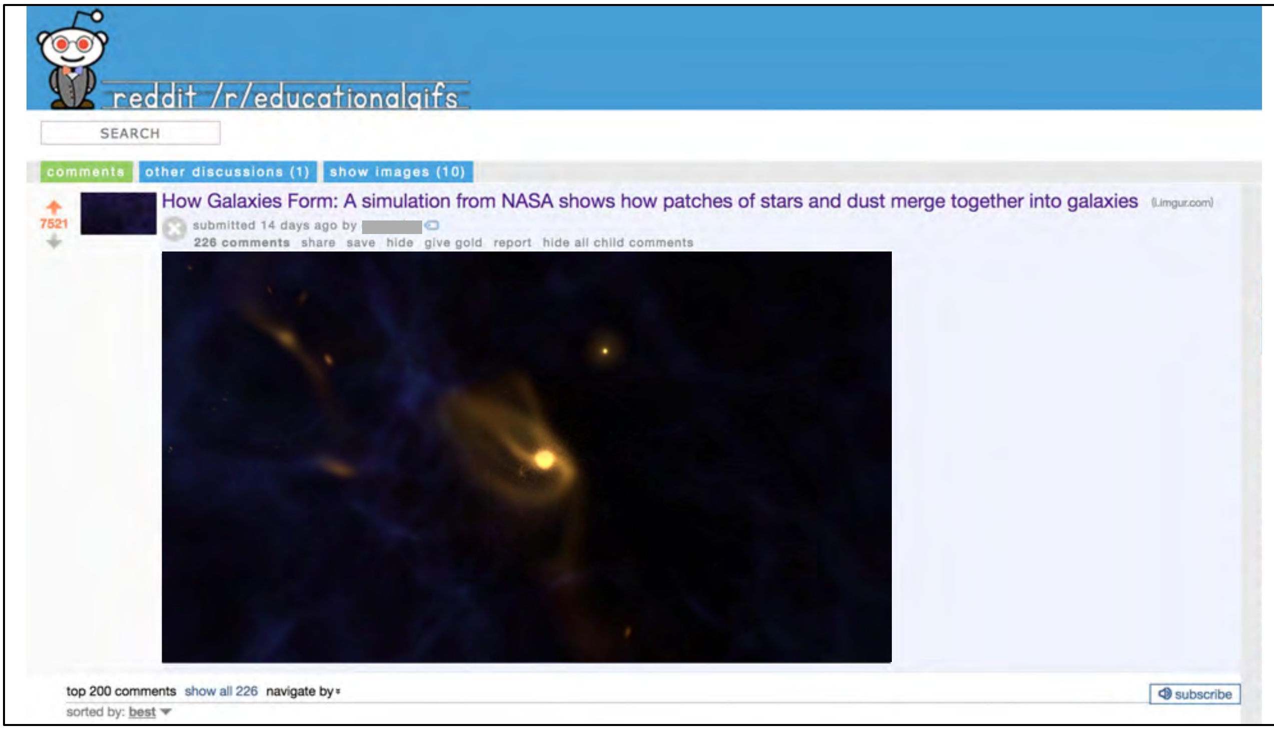
These services may have their own copyright and rules for what happens to the rights of the video once it's uploaded, so take a look at those as well. Some services claim they own content that you upload, and others will try to open up your license to the point where anyone can take and reuse your work for free.

This approach is accessible to wide audiences and is shareable through social media which can increase its virality.



If you want to distribute your visualization on a larger scale, say for TV or in a museum, you'll benefit from building a network and collaborating with a team that includes a professional distributor.

Making a whole film is a lot of work beyond the work of producing a visualization - 25 minute dome films can take a year or more to produce, and a million-plus-dollar budget. You should know what your work is worth, and budget for all the overhead that comes with producing a film.



A decade-old visualization of ours popped up on the front page of Reddit.com this summer. Our visualizations continue to resonate with audiences even outside of the contexts for which they have been created.
These discoveries are reinforcement of the impact that these images have on broad audiences.

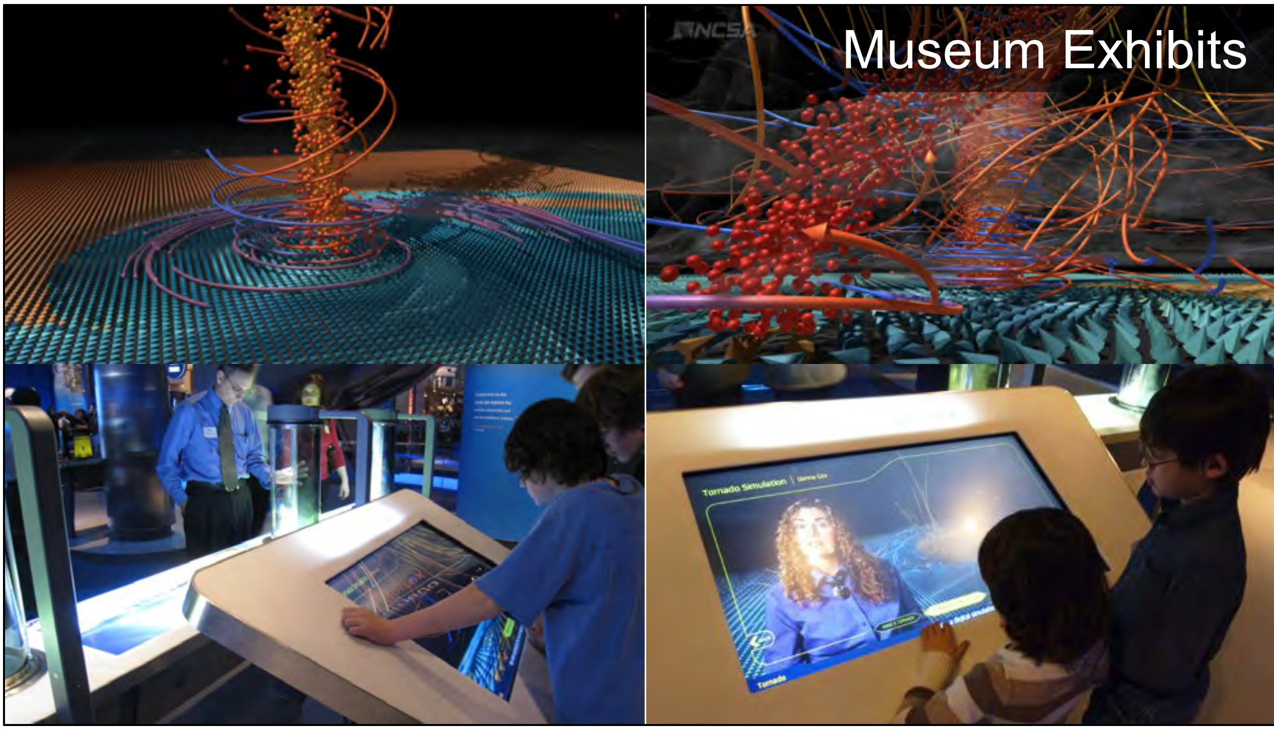
Captions, narrations, titles and credits are likely to get lost along the way in the world of social media. You, as a responsible media creator, should never remove this information from other videos, because you understand that context and credit is crucial. But, knowing this is something that *can* happen, you should always make sure that your visualization speaks for itself. You shouldn't distribute something that looks confusing and relies on explanation by a narrator or caption, because there's a possibility that someone may share the video without that explanation, or with an *incorrect* explanation. So, you should take precautions against spreading misinformation by showing annotations directly in the video where necessary, and always being truthful, accurate, and clear in your visualizations.

Re-Use for maximum Impact

Because visualizations can take a lot of time and energy to put together, it's a good idea to get as much use out of them as you can, and to distribute them in more than one way. A good rule of thumb is to aim to get 3 uses out of a visualization.

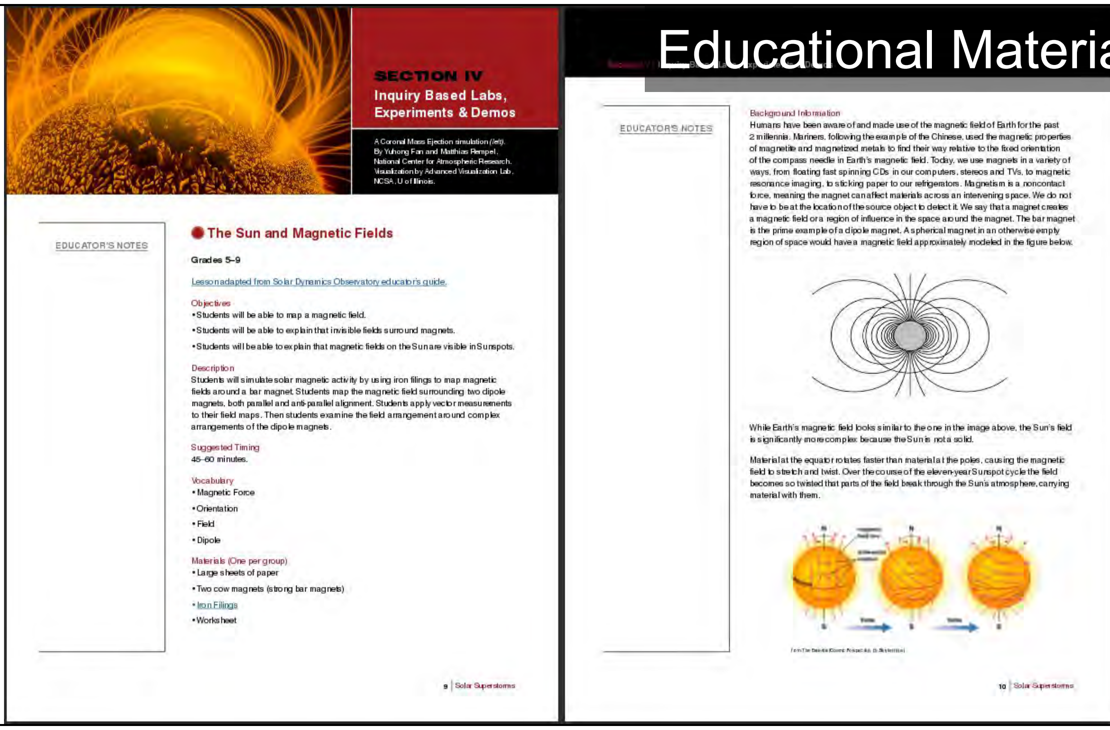
Our work gets re-used in various places and presented in new formats. Through repetition and re-contextualization, we help our audiences experience supercomputer data in different ways so they can build a more thorough understanding.

Our work has been recontextualized to be featured in news weather reports and Neil Degrasse Tyson's "Cosmos", the IMAX films "Space Junk" and "Voyage of Time", and the independent film "The Europa Report". It has also been written about by journalists and used in museum installations.



Educational literature suggests that re-contextualizing data is of enormous value for learners. Showing a data set in multiple representations can help people expand their mental models. For instance, a cinematic experience can be complemented by an interactive museum kiosk.

Educational Materials



SECTION IV

Inquiry Based Labs, Experiments & Demos

A Coronal Mass Ejection simulation *(left)*. By Yuhong Fan and Matthias Rempel, National Center for Atmospheric Research. Visualization by Advanced Visualization Lab, NCSA, U of Illinois.

EDUCATOR'S NOTES

The Sun and Magnetic Fields

Grades 5–9

Lesson adapted from Solar Dynamics Observatory educator's guide.

Objectives
- Students will be able to map a magnetic field.
- Students will be able to explain that invisible fields surround magnets.
- Students will be able to explain that magnetic fields on the Sun are visible in Sunspots.

Description

Students will simulate solar magnetic activity by using iron filings to map magnetic fields around a bar magnet. Students map the magnetic field surrounding two dipole magnets, both parallel and anti-parallel alignment. Students apply vector measurements to their field maps. Then students examine the field arrangement around complex arrangements of the dipole magnets.

Suggested Timing

45–60 minutes.

Vocabulary
- Magnetic Force
- Orientation
- Field
- Dipole

Materials (One per group)
- Large sheets of paper
- Two cow magnets (strong bar magnets)
- Iron Filings
- Worksheet

9 | Solar Superstorms

EDUCATOR'S NOTES

Background Information

Humans have been aware of and made use of the magnetic field of Earth for the past 2 millennia. Mariners, following the example of the Chinese, used the magnetic properties of magnetite and magnetized metals to find their way relative to the fixed orientation of the compass needle in Earth's magnetic field. Today, we use magnets in a variety of ways, from floating fast spinning CDs in our computers, stereos and TVs, to magnetic resonance imaging, to sticking paper to our refrigerators. Magnetism is a noncontact force, meaning the magnet can affect materials across an intervening space. We do not have to be at the location of the source object to detect it. We say that a magnet creates a magnetic field or a region of influence in the space around the magnet. The bar magnet is the prime example of a dipole magnet. A spherical magnet in an otherwise empty region of space would have a magnetic field approximately modeled in the figure below.

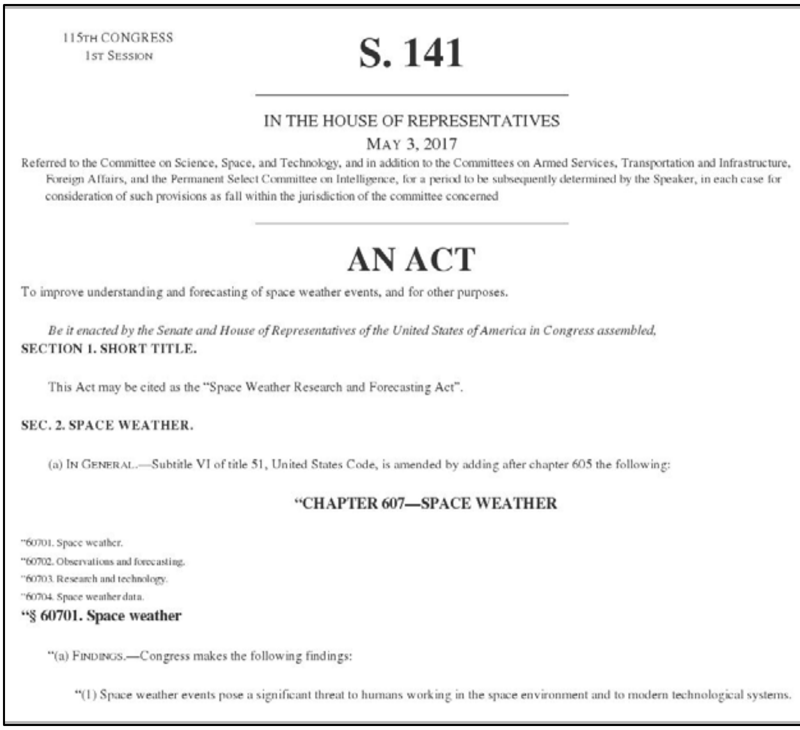
While Earth's magnetic field looks similar to the one in the image above, the Sun's field is significantly more complex because the Sun is not a solid.

Material at the equator rotates faster than material at the poles, causing the magnetic field to stretch and twist. Over the course of the eleven-year Sunspot cycle the field becomes so twisted that parts of the field break through the Sun's atmosphere, carrying material with them.

10 | Solar Superstorms

Our education research collaborators also create supplemental educational materials to go along with our documentaries, such as the booklet pictured here. This repackaging of the information can help students answer their more in-depth questions after they've seen the film.

Much of our government-funded work is free to download and use for educational purposes, and oftentimes we find our visualizations pop up in unexpected places. Our GLIF network map was a backdrop for a scene in the movie “Annie” (2014).



115TH CONGRESS
1ST SESSION

S. 141

IN THE HOUSE OF REPRESENTATIVES

MAY 3, 2017

Referred to the Committee on Science, Space, and Technology, and in addition to the Committees on Armed Services, Transportation and Infrastructure, Foreign Affairs, and the Permanent Select Committee on Intelligence, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned

AN ACT

To improve understanding and forecasting of space weather events, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Space Weather Research and Forecasting Act".

SEC. 2. SPACE WEATHER.

(a) IN GENERAL.—Subtitle VI of title 51, United States Code, is amended by adding after chapter 605 the following:

"CHAPTER 607—SPACE WEATHER

"60701. Space weather.
"60702. Observations and forecasting.
"60703. Research and technology.
"60704. Space weather data.

"§ 60701. Space weather

"(a) FINDINGS.—Congress makes the following findings:

"(1) Space weather events pose a significant threat to humans working in the space environment and to modern technological systems.

Note that policymakers are also part of a "general audience".

Our film "Solar Superstorms" was shown to Congress prior to a vote on de-funding space weather research.

(it got funded, with a unanimous vote)

Occasionally, visualizations even have very observable impacts on policy makers. In May 2016, AVL Director Donna Cox was invited by the American National Science Foundation to present "Solar Superstorms" to members of the Senate in Washington, D.C., after the Space Weather Research and Forecasting Act was introduced. In May 2017, the Act passed the Senate unanimously.

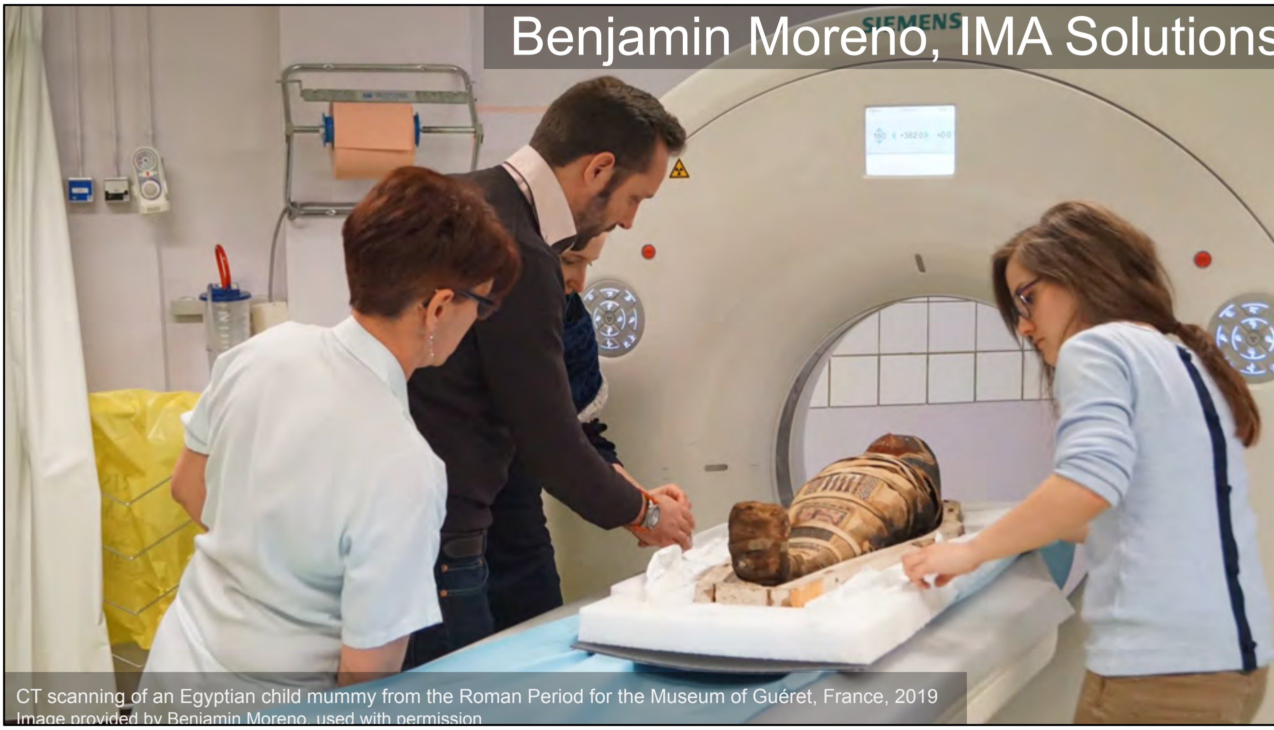
While we do keep in mind that that the target audience for our visualizations is the "general public", it is important to remember that experts and policy makers are among that public. This is why it is crucially important for our representations and explanations of science to be not only visually appealing, but also unambiguous and truthful.

State of the Art
within our Community

The purpose of this section is threefold:

1. Share with the ACM SIGGRAPH community exemplary efforts and latest developments from colleagues within our community.
2. Highlight different perspectives, directions, tools utilized and skillsets employed in our efforts to support science.
3. Demonstrate the need of platforms and programs within ACM SIGGRAPH to share latest trends and advancements within Cinematic Scientific Storytelling.

Please note that in our course notes we present only a small sample of projects and work from visualization researchers and practitioners. Providing an exhaustive list is outside the scope of this course. The list aims to be indicative of a variety of efforts and during our course presentation we will shed light to additional efforts.



CT scanning of an Egyptian child mummy from the Roman Period for the Museum of Guéret, France, 2019
Image provided by Benjamin Moreno, used with permission

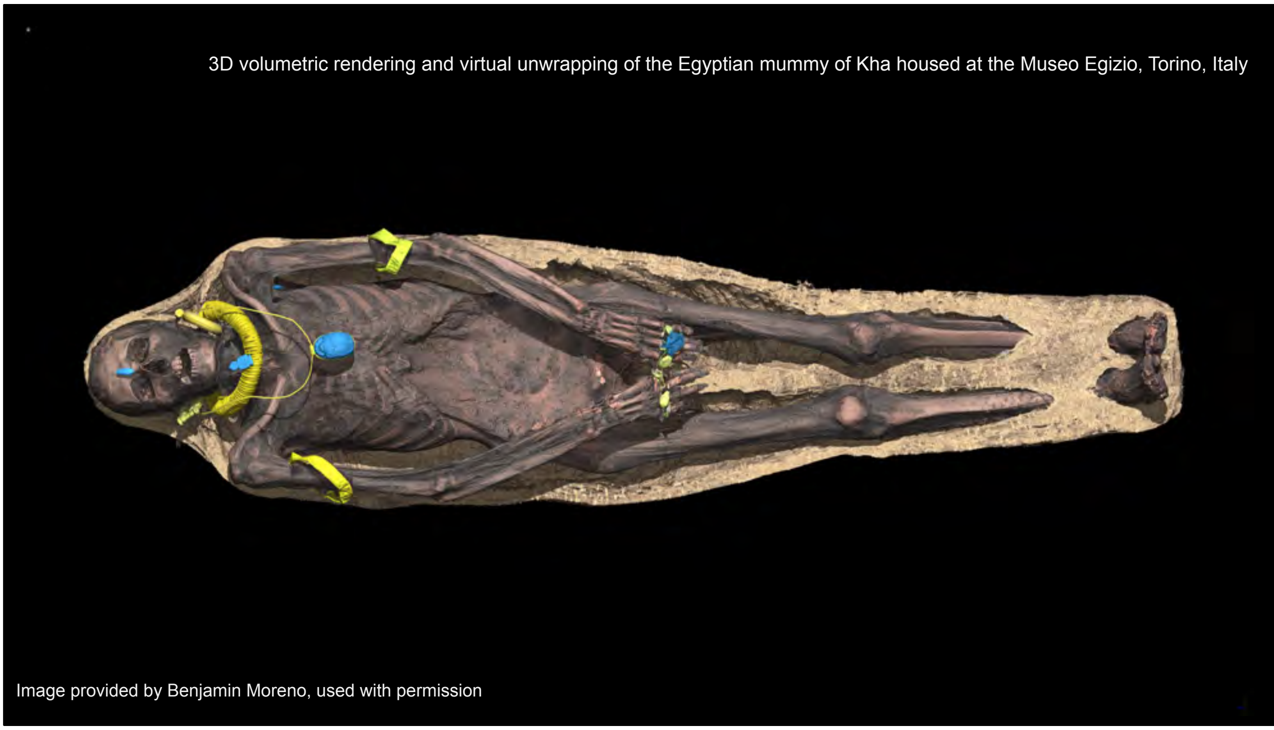
Benjamin Moreno, CEO of IMA Solutions; Toulouse, France

In the image above, Benjamin Moreno prepares an Egyptian child mummy from the Roman Period for a CT scan. The mummy is located at the Museum of Guéret, France.

Information about Benjamin Moreno and IMA Solutions:
Benjamin Moreno, after studying biology, founded IMA Solutions in 2008 to bring to Museum professionals (e.g.: Museums, Science Centers, Archeology Laboratories) scientific tools and technical skills to 3D digitize, analyze and present their collections in new ways.

Projects developed from IMA Solutions start with data acquisition that involve a wide variety of technologies (Xray, CT scanning, MRI, microCT, Synchrotron, 3d scanning, LIDAR to name few) to 3D digitize artworks, ancient human/animal remains or even complete archeological fields. IMA Solutions uses certified scanning systems to maintain the highest quality of data captured, as it aims to develop cinematic storytelling experiences that are scientifically accurately and approved by scientists. The data acquisition process is a critical one as it leads to new content/knowledge and of course a better understanding of the artwork.

The second step is the presentation of the discoveries to the audience and museum production. Depending on storytelling needs, data, and available budget, IMA Solutions scales production to use a wide range of systems and techniques, such as 3D volumetric animations, virtual reality, augmented reality, 3D pedagogic interactive applications or even 3D prints of virtually extracted objects (like 3D prints of the jewels hidden inside an Egyptian mummy). The goal is to always find and use the best way to transmit to the audience the discoveries and knowledge with a unique and engaging experience.

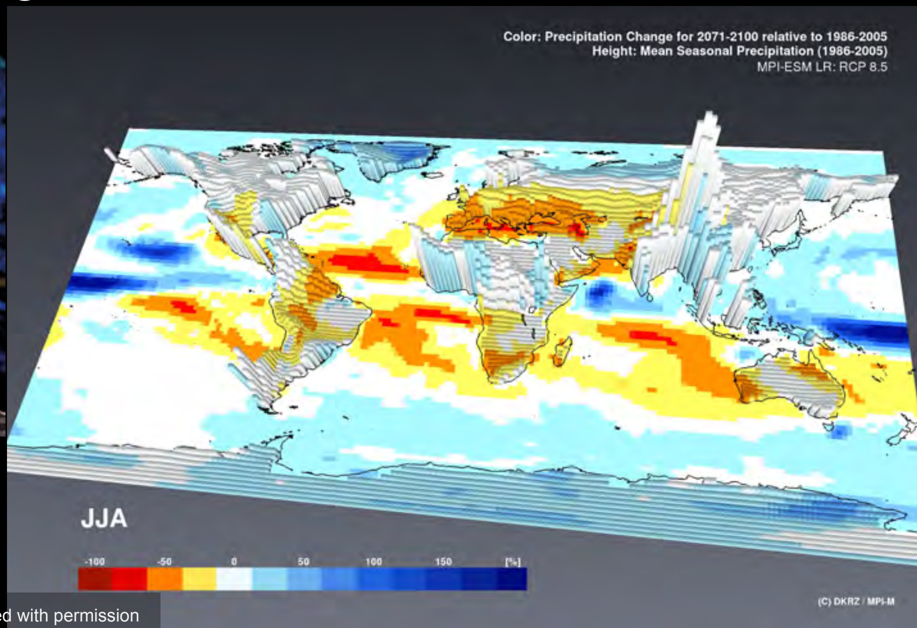


Sample work: Still image from the cinematic visualization of the Egyptian mummy of Kha housed at the Museo Egizio, Torino, Italy. This visualization work includes 3D volumetric renderings and virtual unwrappings of the mummy in the format of 3D animations from a CT scanner dataset. This work is currently shown at the exhibition "Archeologia Invisibile", Museo Egizio, Torino in Italy shown until January 2020.
For more Information about the exhibit please visit:
https://museoegizio.it/esplora/mostre/archeologia-invisibile/

Other recent projects:

- Exhibition *"Microbiota"* at Cité des Sciences et de l'Industrie, Paris, France until August 2019:
  http://www.cite-sciences.fr/en/explore/temporary-exhibitions/microbiota/the-exhibition/
  IMA Solutions is responsible for: 1) CT scanning of a living human to virtually explore the human digestive system, 2) Development of an interactive pedagogic multitouch application to explore the human digestive system using a 3D realtime voxel engine.
- Exhibition *"Easter Island"*, Museum of Natural History of Toulouse, France, until June 2019: https://www.iledepaquesexpo.fr/en
  IMA Solutions is responsible for 3D scanning of the Moai statue housed at the British Museum and 3D CNC machining of a size to size replica for video beaming projections for the exhibition.

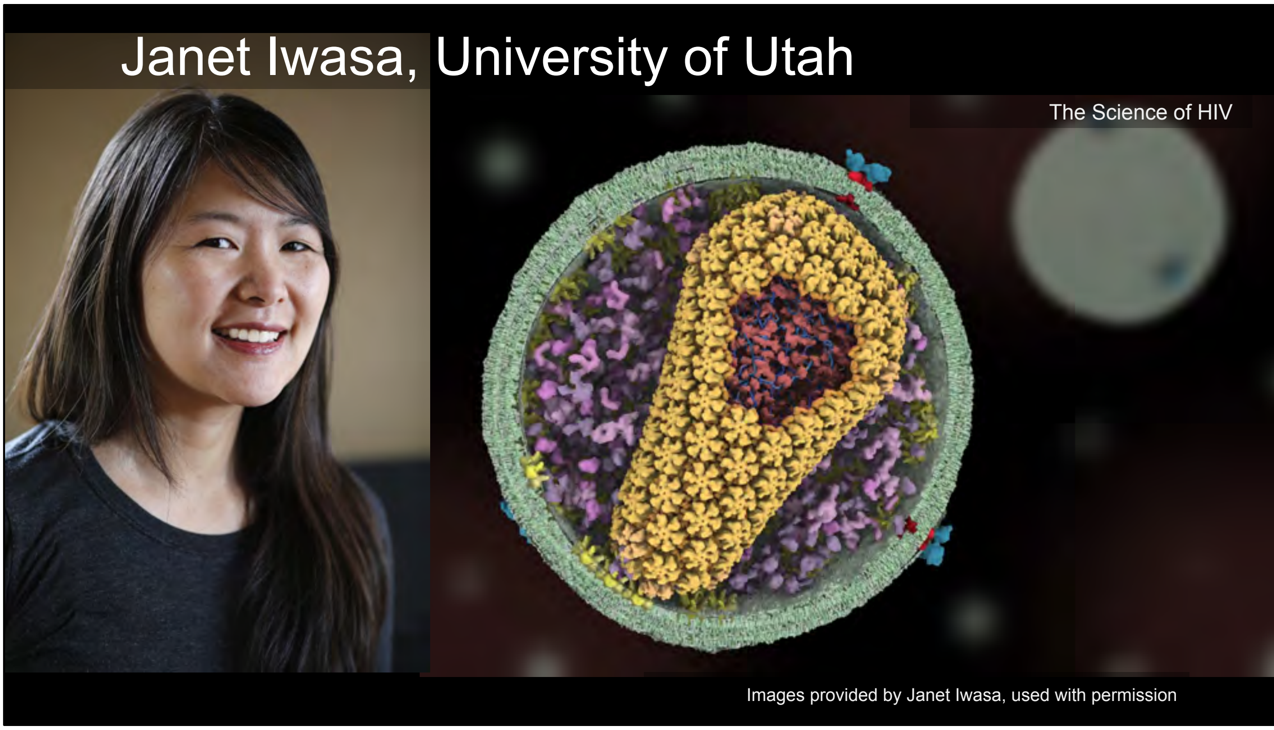
Michael Böttinger, DKRZ



Images provided by Michael Böttinger, used with permission

Michael Böttinger, Group Head Visualization & PR, German Climate Computing Center (DKRZ)
https://www.dkrz.de/about-en/staff/MichaelBoettinger?set_language=en&cl=en

Information about Michael Böttinger and DKRZ:
Michael Böttinger is a science visualizer at the DKRZ, the German Climate Computing Center, which is a domain specific service facility that provides high performance computing, data storage, archiving, and associated services to its scientific community.
Scientific visualization of climate research data for both data analysis and communication purposes has been one of the central services of DKRZ for almost 30 years, i.e. since the first IPCC report was published. These visualization services strive to cover the high-end part of the visualization work, i.e. projects that go beyond simple 2D techniques used by most domain scientists. Data visualizations developed at DKRZ (still images or animations) are used by the scientific community and to communicate to the general public and policy makers. Climate and climate change visualizations of DKRZ have been used for TV documentaries, newspapers, school books, in science centers, planetariums, climate conferences and public exhibitions such as several World EXPO exhibitions. Throughout the last 25 years DKRZ's visualization and public relations group consists of 2-5 members, which is an impressive effort.
The still image (right) above is of a bivariate visualization of precipitation change: The height of the bars on land indicate the mean summer precipitation for 1986-2005 as simulated with MPI-ESM. The colors show the projected changes in summer precipitation for 2071-2100 relative to 1986-2005 for the pessimistic greenhouse scenario RCP8.5. Blue colors indicate an increase, yellow to red colors indicate a decrease in precipitation for the summer season, and a bright color is used to show neutral areas with only little changes.
This work will appear in the book *Foundations of Data Visualization*, Edited by Gerik Scheuermann, Penny Rheingans, Min Chen. Publisher: Springer (upcoming fall 2019)



Janet Iwasa, Assistant Professor, Department of Biochemistry, University of Utah
https://animationlab.utah.edu/

Information about Janet Iwasa:
After receiving her Ph.D. from UCSF for work completed in Dyche Mullins' lab, she completed a postdoc with Jack Szostak (MGH/Harvard) and later worked on biological visualizations as a faculty member at Harvard Medical School.

The Science of HIV project (right - still image) aims to use animated visualizations and illustrations to captivate, inform, and educate diverse audiences. Molecular animations and illustrations are created in collaboration with HIV researchers around the globe, and reflect current hypotheses of how HIV infects the human body. For more information about the Science of HIV project please visit: http://scienceofhiv.org/

Ivan Viola, Associate Professor, Computer Science, KAUST
https://www.kaust.edu.sa/en/study/faculty/ivan-viola

Information about Ivan Viola:
The research interest of Professor Ivan Viola (pictured above) is scalable technology for interactive molecular visualization with the ultimate goal of constructing, visualizing, and modeling the entire complex biological cell at atomistic detail. This technology will allow people to interact, explore, study, and understand the life at nanoscale.

Over the years, Ivan Viola has collaborated with researchers and practitioners to develop technology and techniques at the frontier of data-driven scientific storytelling and has made contributions in the theory of visualization science. A sample of collaborative efforts presented in this course are:
1) Tobias Klein, Ivan Viola, Peter Mindek, “A Multi-Scale Animation Framework for Biological Fibrous Structures”, Proceedings of Eurovis 2019, June 3-7, 2019
2) Collaboration with Dr. Drew Berry, Biomedical Animator at The Walter and Eliza Hall Institute of Medical Research.

EUROVIS 2019/ J. Madeiras Pereira and R. Raidou *Poster*

A Multi-Scale Animation Framework for Biological Fibrous Structures

T. Klein[1,2], I. Viola[2,3], P. Mindek[1,2]

[1]TU Wien, Institute of Visual Computing & Human-Centered Technology
[2]Nanographics GmbH
[3]KAUST, Visual Computing Center

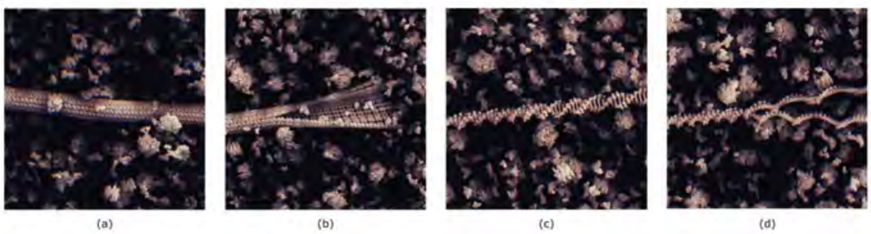
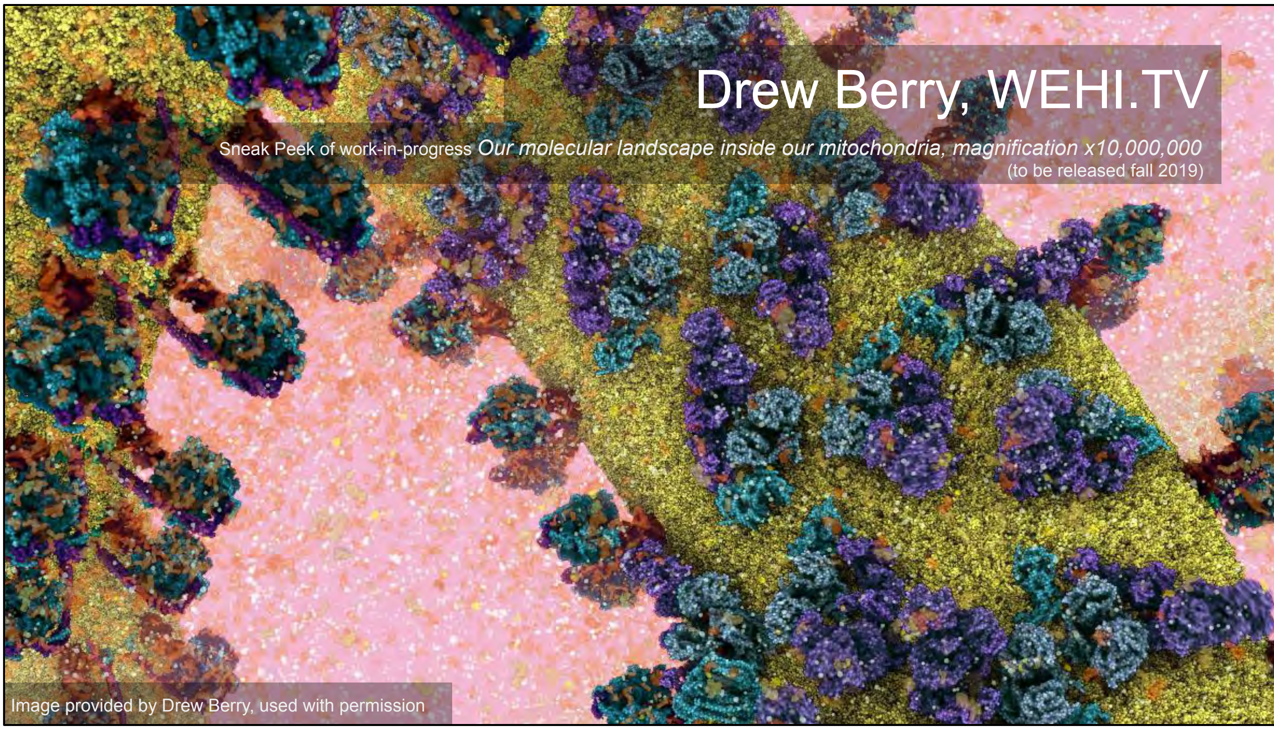


Figure 1: *Overview of the generation of multi-scale procedural animations of biological fibrous structures.*

Tobias Klein, Ivan Viola, Peter Mindek, "A Multi-Scale Animation Framework for Biological Fibrous Structures", Proceedings of Eurovis 2019, June 3-7, 2019

Relevant video:
Of Catastrophes and Rescues: Making the Invisible Visible, available here*:*
https://vimeo.com/330528972
By Peter Mindek, Tobias Klein, Ludovic Autin, Haichao Miao, Theresia Gschwandtner
Created for the Pacific Vis Storytelling Contest



Dr Drew Berry, Biomedical Animator, The Walter and Eliza Hall Institute of Medical Research
https://wehi.tv

Combining data from microscopy, 3D tomography, and X-Ray crystallography, the wehi.tv team is bringing to life the dynamic molecular engines that convert the food you eat and oxygen you breathe into chemical energy that powers your cells and tissues.

Exploiting the incredible power of video game GPU technology, vast molecular landscapes can be generated in ultra-high resolution detail, full 360 degree cinema for museum and science centre full-dome theatres, interactive exploration of data via AR on student mobile phones.

Open Challenges

In this part of the course we will discuss the open challenges relevant to our work. The next few slides shed light to a sample only of open challenges in our field. During the course we will illustrate them with examples and present even more.

Our practice and research is part of the story of science that is being told today, and is embedded in other stories with societal, educational and cultural dimensions.

We are science visualizers, storytellers, data wranglers, each one of us is called to represent the work of multiple departments within a typical computer graphics studio and we breathe in the sciences.

Collaboration with domain scientists, media professionals & educators **is key**

..and **shared language** is one of the prerequisites to foster collaboration

*“For if we do not have a vocabulary with which to discuss graphic concepts, we cannot discuss these concepts in an unambiguous manner.“ **

*Wainer, H. and C. M. Francolini (1980). "An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps." The American Statistician **34**(2): 81-93.

*Bujack, R., et al. (2018). "The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps." IEEE Transactions on Visualization and Computer Graphics **24**(1): 923-933

Exemplary efforts to define concepts, terminology and to formalize the theory of visualization, which in turn support collaboration.

Ivan Viola and Tobias Isenberg (2018) ***Pondering the Concept of Abstraction in (Illustrative) Visualization***. IEEE Transactions on Visualization and Computer Graphics, 24(9):2573–2588, September 2018.

Impact: embarked on a research agenda to understand and formally define the concept of abstraction for the visual representation and exploration of data. The paper provides a first formalization of the abstraction concept and focuses on open questions still waiting to be answered.

Bujack, R., et al. (2018). T***he Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps.*** IEEE Transactions on Visualization and Computer Graphics **24**(1): 923-933

Impact: developed a cross-discipline framework and language for the assessment of continuous colormaps through the discipline and universal power of mathematics.

visual literacy, **data literacy, computational thinking**, **visual thinking**

Impact of Art in shaping our world and crafting memorable, emotional experiences that transport us to new dimensions: *Art & The power of the Extrasensory*
Artworks have defined my language and thinking.

Need to insert computational thinking within the arts-based visualization programs and **visual thinking,** sonic and visualization literacy within the sciences. Not as an afterthought, for example taking few credit courses here and there, as electives, this requires the development of new integrated programs.

*Chen M., Kindlmann G., Hauser H., Rheingans P., Scheuermann G., editors , Foundations of Data Visualization. Springer. Forthcoming 2019

..and on the same time…"*communication of science is increasingly recognized as a responsibility of scientists, yet how do scientists learn these skills?*"*

*Brownell, S. E., Price, J. V., & Steinman, L. (2013). Science Communication to the General Public: "Why We Need to Teach Undergraduate and Graduate Students this Skill as Part of Their Formal Scientific Training". Journal of undergraduate neuroscience education : JUNE : a publication of FUN, Faculty for Undergraduate Neuroscience, 12(1), E6–E10.

And some more challenges..

- How can we keep up with radical updates in animation and VFX tools, especially since we build our own custom pipelines, shaders and plugins on top to address scientific and data needs?

- How can we evaluate visualizations without bias?
  Heather Krause, “Is Feminist Data Visualization Actually a Thing? (Yes, and How!)”, Evergreen Data Guest Post

- Should we visualize **uncertainty** for the general public?
  Or should we focus on what we know (rather on what we do not know), especially in works relevant to Climate Change. What are the best practices?

"Now more than ever we must stand up for science.

We need cross-pollination and engagement. We need experimentation on form, tone, content and distribution. We need to try to find a way to take the message to where people are, through digital promotion, distribution and social engagement."

***Dan Rather**, Scientific American*

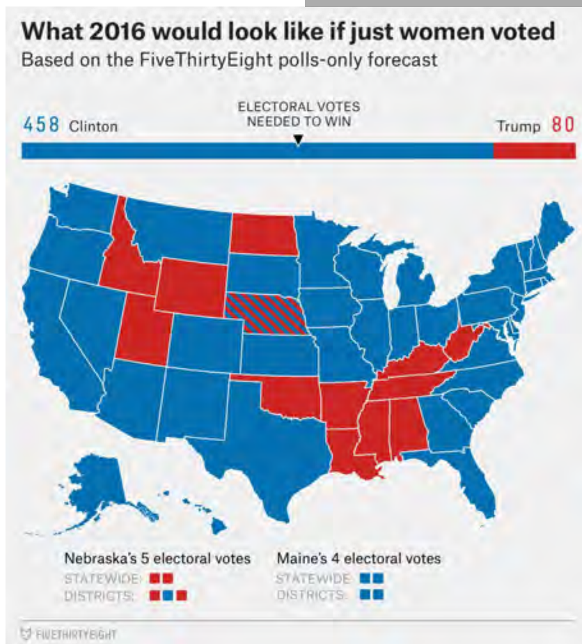
Amanda Montanez *(December 20, 2016)* How science visualization can Help Save the World

The pubic consumes also digital content on mobile devices (small screens).
How can we design, develop content that scales in small**er** screens, especially since our content is visually rich and detailed?

Emerging Directions

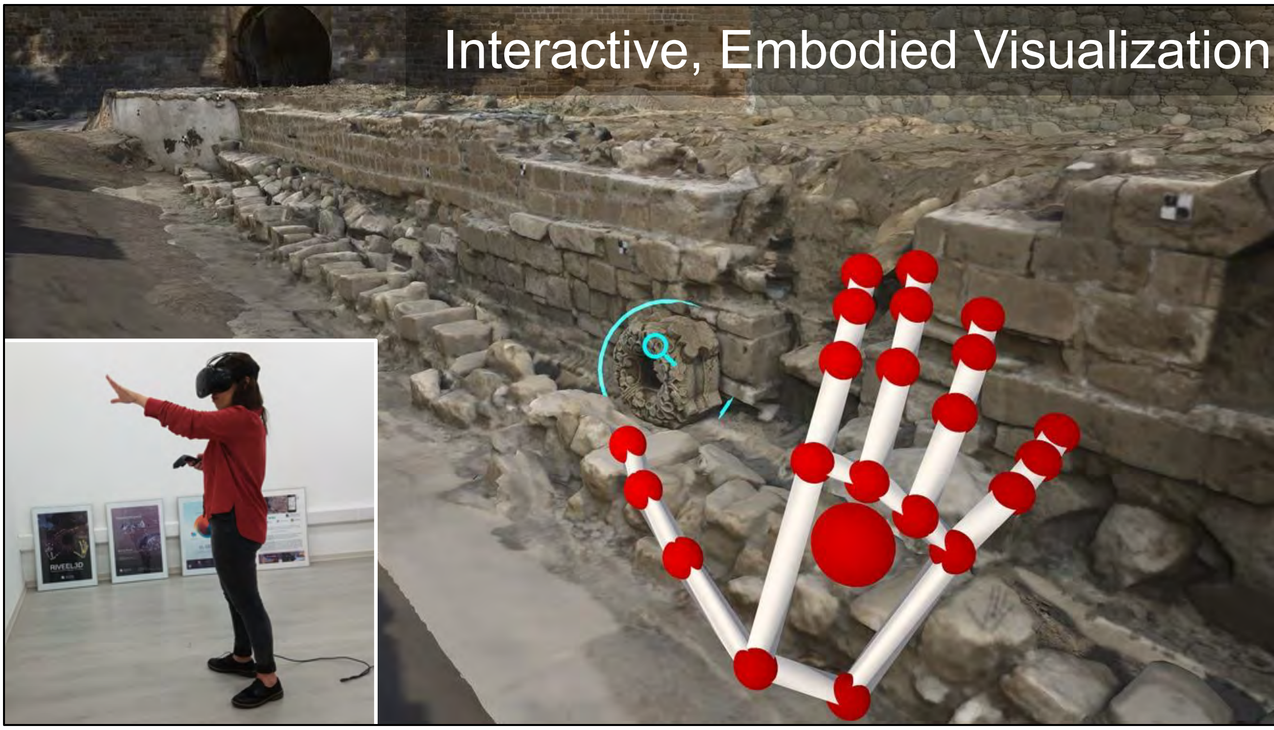
So where do we see visualization going?

How We Consume News

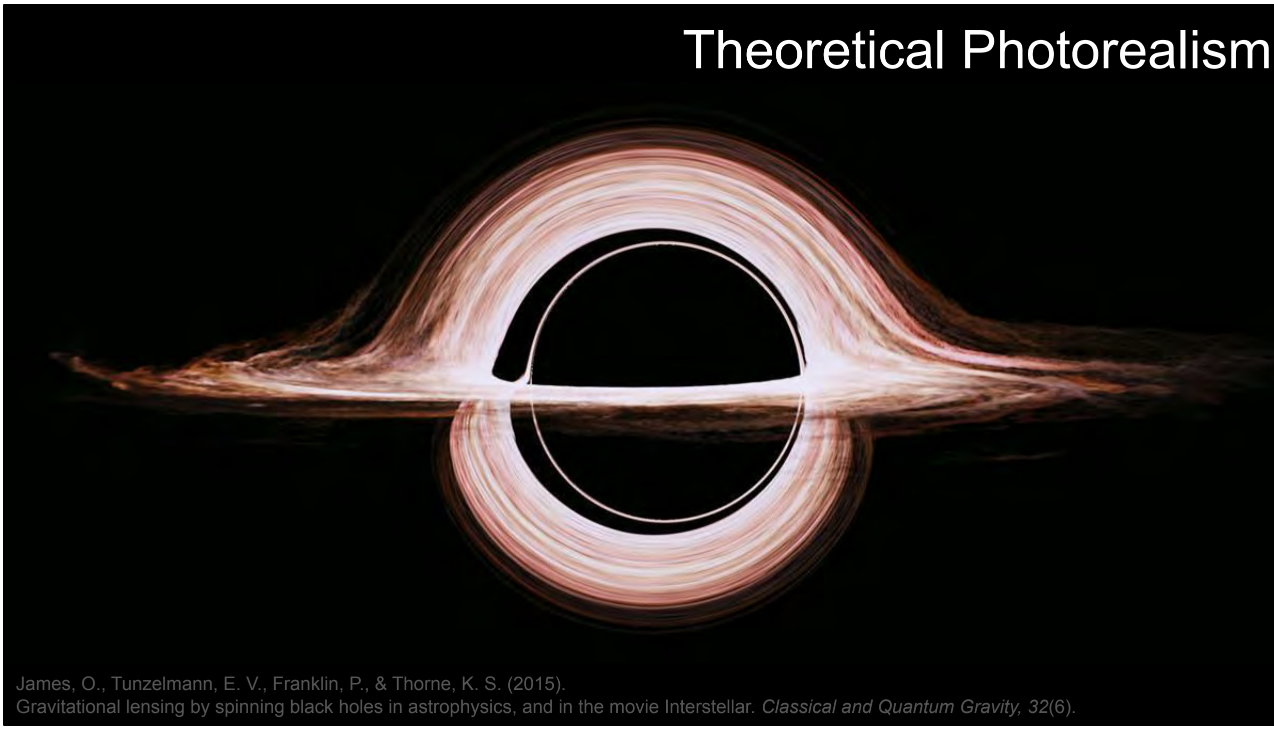


Our understanding of the world around us is more influenced by data everyday. A new type of journalist has emerged in the past few years, someone like Nate Silver, a journalist who started his career analyzing baseball stats for the New York Times but now runs a data journalism website called FIVE THIRTY EIGHT. Nate and his staff use data to tell stories about current events, but he and other data journalists understand that visualization is a key component of this storytelling.

We are bombarded by infographics, educational videos, and comic illustrations on social media. But as we are discovering the influence that international governments may have had over European and American election results in the past couple years, it's becoming clear that the way we share information is not well regulated.



Of course journalism is only one of many industries embracing visualization. As graphics processing becomes more affordable, interactivity becomes more common, and display devices for virtual reality and augmented reality advance, we see more industrial and research applications. Historians can digitally recreate ancient landscapes, auto manufacturers are wearing Google Glass for quality assurance, and surgeons are practicing their first surgeries in virtual space.



Perhaps one of the most exciting future directions of visualization is what Alan Warburton called in a recent video “Theoretical Photorealism”. He discussed the visual effects of the movie “Interstellar” which is noteworthy for its work with a well known astrophysicist named Kip Thorne. In the movie, the effects team created a real effect observed by telescopes known as “gravitational lensing” where light bends around the immense gravity of a black hole, turning a flat disk of gas and dust into the bizarre 3-dimensional waterfall you see here.

There are many phenomena described by theoretical science that are difficult to observe with cameras, but which computer software can describe. This is the space supercomputer scientists operate in, and in which our team strives to innovate.