Visualization Basics with ParaView

UNIVERSITY OF LEEDS

Joanna Leng (j.leng@leeds.ac.uk)

EPSRC funded RSE Fellow

Working on Research Computing and Imaging

Ethos behind the course

- This is a fast overview of many topics.
- The key aims are to give you:
- Experience of using ParaView i.e., the exercises in an environment where you can get help.
- Teaches you enough key words and understanding of visualization so you can trouble shoot using ParaView and visualization topics beyond this course.
- Provide you with reference materials.

Contents

UNIVERSITY OF LEEDS

- A brief history of visualization
- The visualization pipeline
- Visualization is distributed computing
- Overview of ParaView GUI
- Exercises and break
- Colour
  - Digital image formats
  - Colour models
  - Human Vision
- Visualization Data Structures
- Visualization Design
- Visualization pipeline optimization

What is an RSE?

UNIVERSITY OF LEEDS

Stands for:

Research Software Engineer

New page on Wikipedia

- https://en.wikipedia.org/wiki/Research_software_engineering
- Research software engineering is the use of software engineering practices in research applications. The term started to be used in United Kingdom in 2012[1][2], when it was needed to define the type of software development needed in research. This focuses on reproducibility, reusability, and accuracy of data analysis and applications created for research.

What about you?

UNIVERSITY OF LEEDS

What is the oldest visualization?

The History of Visualization

What is the oldest visualization?

- The oldest map is from 6200 BC

Are Computers Necessary?

UNIVERSITY OF LEEDS

- Visualization predated the existence of computers by at least 8 thousand years and probably started before text existed as a means to communicate or store information.
- Its development has been connected to developments in art and textual communications.
- *Milestones in the history of thematic cartography, statistical graphics, and data visualization.*
- http://www.datavis.ca/milestones/.

Questions

UNIVERSITY OF LEEDS

- Are cave paintings art?
- Are cave paintings visualization?
- Is art visualization?
- Is visualization art?

- The answers are points of view and open to discussion.

I think:

UNIVERSITY OF LEEDS

- Visualization is a way of exploring, storing and understanding information.
- It comes from a need to create, learn, store and communicate information
- It pre-dates all civilized forms of communication and the quality of a visualization reflects the technology of the time in which it was created and the cultural practices in using the data/information

Visualization Cultural Influences

- Edward Tufte has studied how particular visualization techniques have developed – the "scientific" need and the benefits that resulted. (The scientific is in quotes because visualization pre-dates science.)
- He views visualization as the appropriate analysis and appropriate presentation of information

Improving User Experiences

- Colin Ware a failed artist and visualizer considers visualization to be an area of perceptual phycology.
- Perceptual phycology techniques are used to make the presentation of data suit the functions of our brains BUT does that improve the adoption of visualization AND do we know enough about perception and the brain to make this approach robust.

Academic Disciplines

UNIVERSITY OF LEEDS

- Academic disciples are a relatively new concept that have grown with ways of analysing and presenting their data
- Culturally visualization as a full academic discipline is new (about 30 years) and has similar difficulties to statistics – most disciplines use it but few do it well and doing it well is not visualization research

Visualization Topics an Overview

UNIVERSITY OF LEEDS

Hardware

Visualization Systems

e.g. large data,
Interfaces with VE
collaborative
Computational steering

Generic Visualization Techniques

e.g. isosurfaces

Applied Visualization

e.g. radiologist workbench,
Solar physics toolkit

Users + Data

Parallel programming
meta computing
Computational Science

Computing shape
Statistics
Image processing

Mathematics
Usability
Virtual Environments

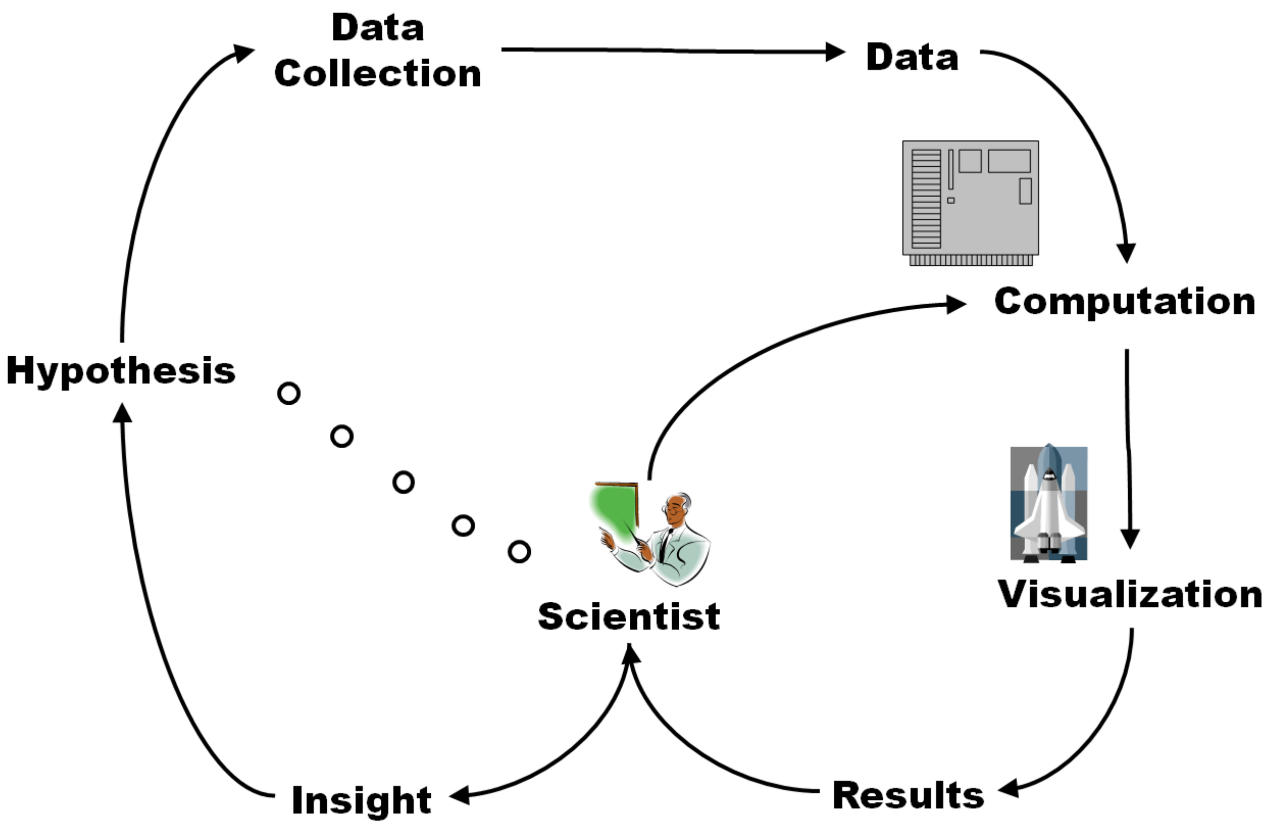
Art
Graphic Design
Domain Knowledge
Psychology

Scientific Vs Information Visualization

- Early in the era of computerized visualization the data structures that held the data dominated the field.
- Scientific Visualization was the first computerized visualization paradigm and related to the visualization of 2D or 3D simulation data or 3D medical scans
- Then Information Visualization arrived and was the visualization of databases and spreadsheets.
- It is more complex now!

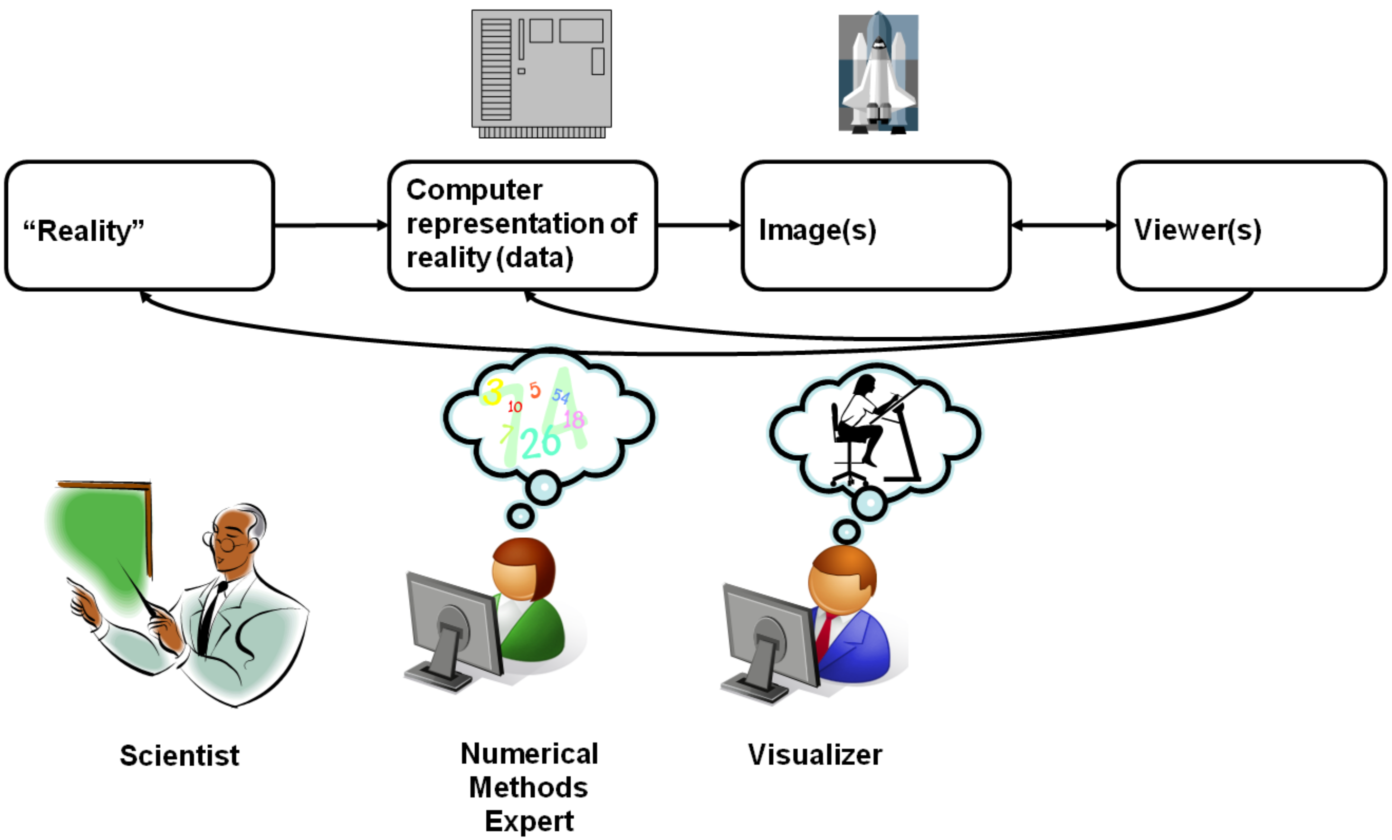
Scientific Visualization

UNIVERSITY OF LEEDS



Watson's model of scientific investigation was used by the visualization community to design software frameworks. It can be adapted to cover data science.

The Visualization Pipeline



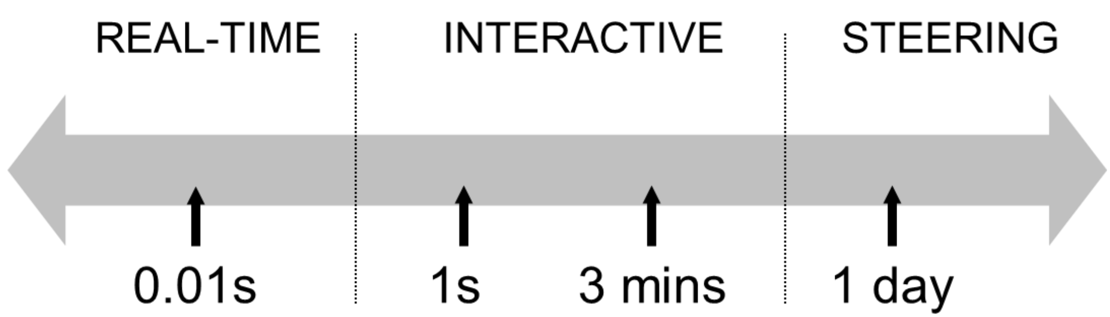
Hardware and Processing Key Words in the Visualization Pipeline

HARDWARE	CPU (nodes and cores)	Graphics card	Screen
"Reality" →	**Computer representation of reality** →	**Images(s)** →	**Viewer(s)**
PROCESSES	simulation or analysis	rendering	

Making Simulations Interactive – *In Situ* Visualization – Catalyst in ParaView

UNIVERSITY OF LEEDS

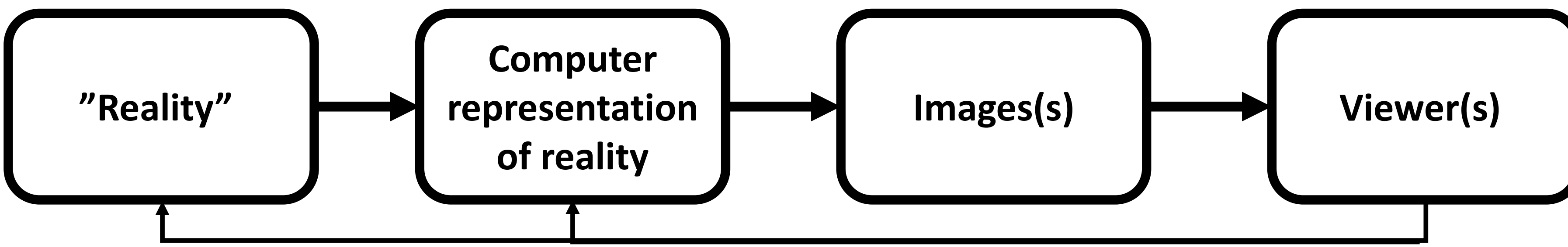
- If a visualization is too slow.
  - Reduce the size of the data.
  - Reduce the computational cost of the visualization techniques.
  - Use high performance graphics.
- Complex simulations execute slowly.
  - Use a remote supercomputers and **monitor** it to check they are not wasting compute time.
  - If interactive time on a supercomputer is possible monitor and alter simulation parameters during execution to **steer** the simulation.



Visualization is Distributed Computing 1

High performance, parallel or super computers “speed up” slow simulations.

Desk-top computers are passed the data from the simulation, visualize and render them on a screen.



Check with reality:

Compare with other data “talk” with peers for validation

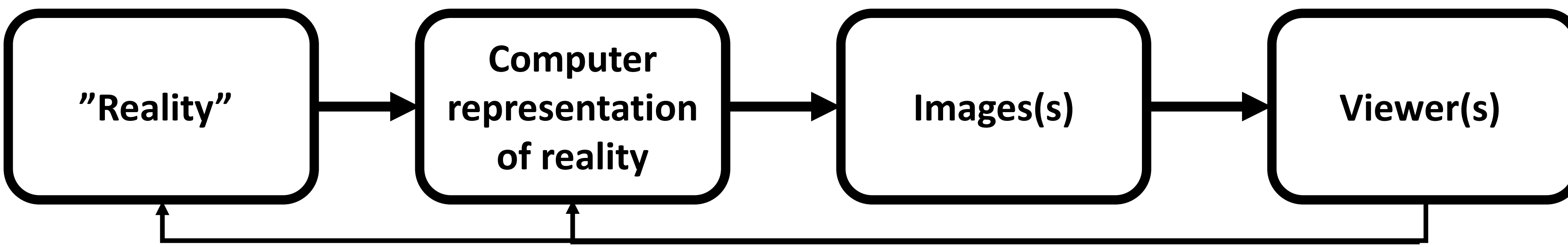
Computational steering allows scientists to interact with a simulation during its run time.

Visualization is Distributed Computing 2

UNIVERSITY OF LEEDS

High performance, parallel or super computers "speed up" slow simulations. The visualization results are rendered in software and passed to the user or sometimes a visualization technique is computed in the simulation e.g., iso-surfaces

Desk-top computers are passed OpenGL or a stream of compressed images and render them on a screen.



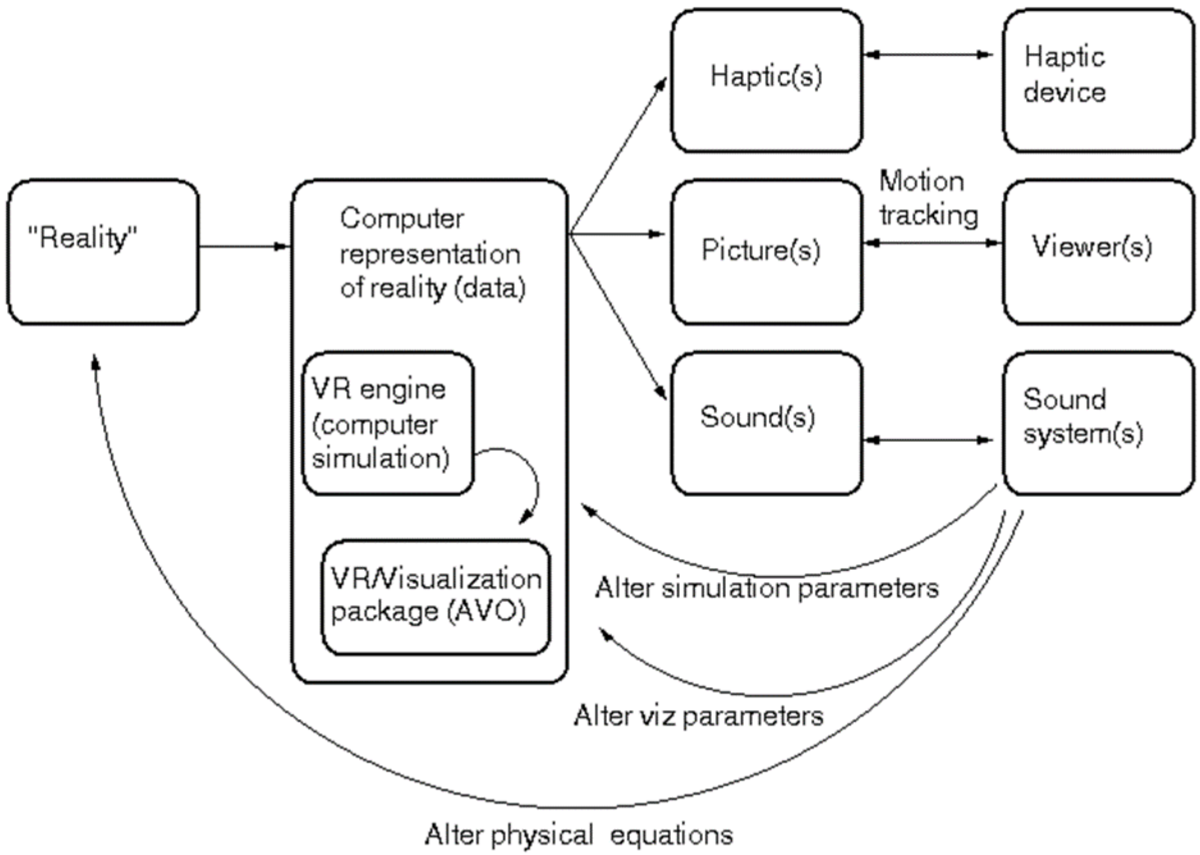
Check with reality:

Compare with other data "talk" with peers for validation

Computational steering allows scientists to interact with a simulation during its run time.

Visualization is Distributed Computing 3

UNIVERSITY OF LEEDS



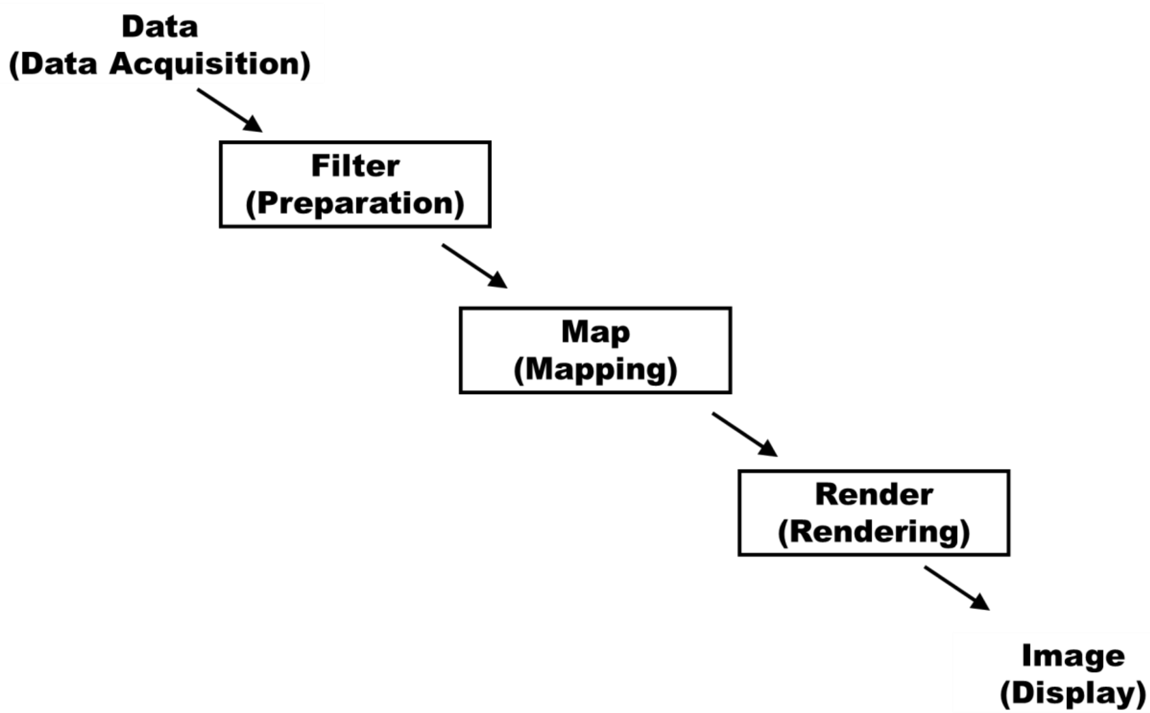
Overview of the ParaView GUI

UNIVERSITY OF LEEDS

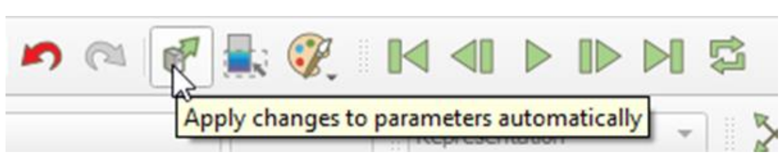
- Please look at the crib sheet:
  - It gives an overview of ParaView.
  - It tries to explain all the quirks in one place.
  - It gives quick key short cuts for interacting with data in the RenderView.
  - It gives many key terms that makes trouble shooting easier.
  - It gives links to many to documentation and materials that also make trouble shooting easier.

ParaView Visualization Pipeline Terminology

UNIVERSITY OF LEEDS



ParaView is designed for large data so you **PULL** rather than **PUSH** the data and is why ParaView has this button.



Modular Visualization Environments

UNIVERSITY OF LEEDS

Modular

1. Building blocks of code
2. Users can extend the functionality

Visualization

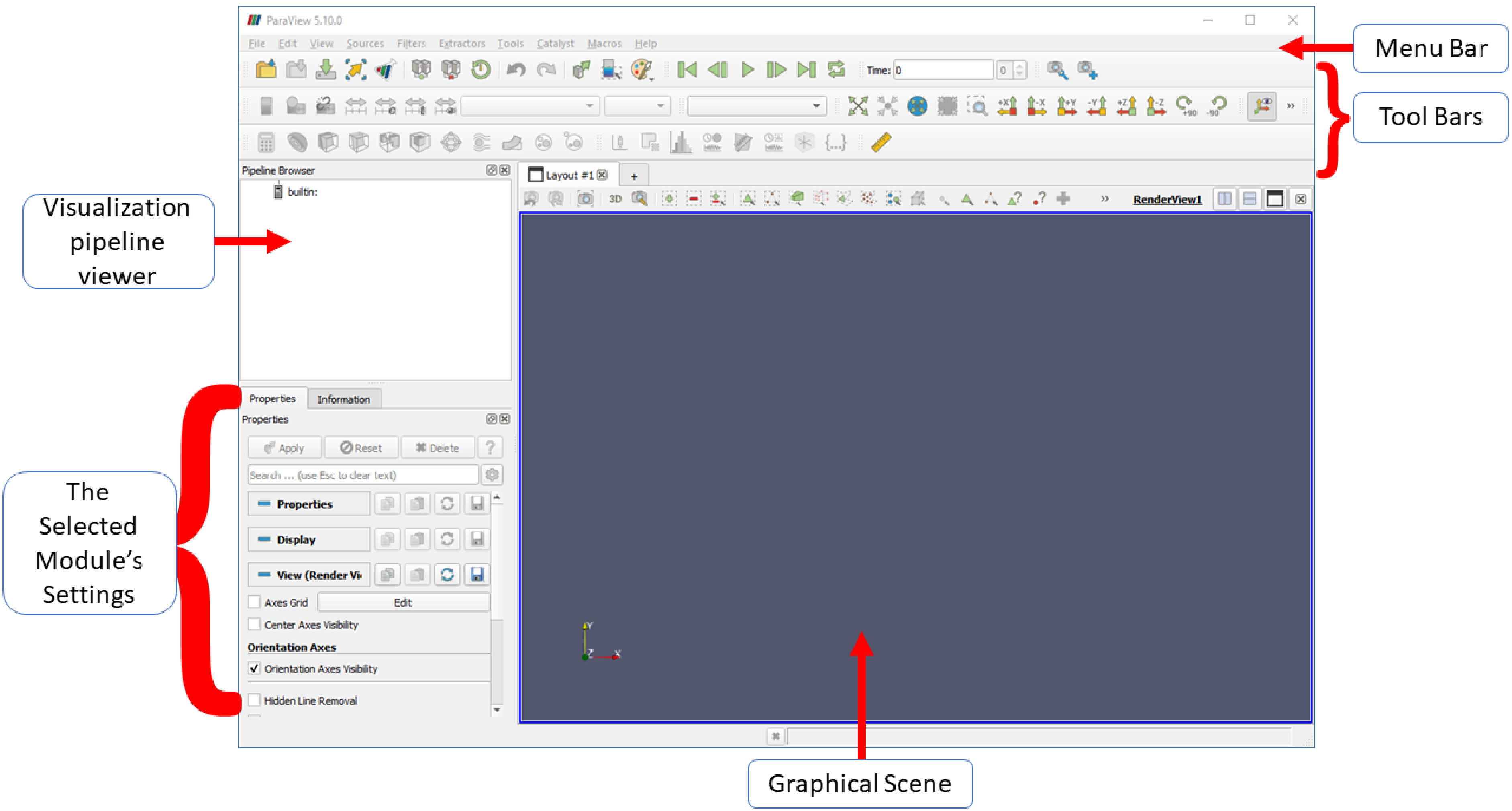
1. Analyses data and presents it in a visually appealing form that adds meaning.

Environment

1. A library of modules that perform particular types of analysis on the data.
2. A visual programming area where the modules are joined together into a pipeline; data is thought to flow through the pipeline from the top where it is read in to the bottom where it is displayed.
3. A graphics scene where the results of the pipeline are rendered and where the user can manipulate the objects on display e.g, displayed objects can be rotated.

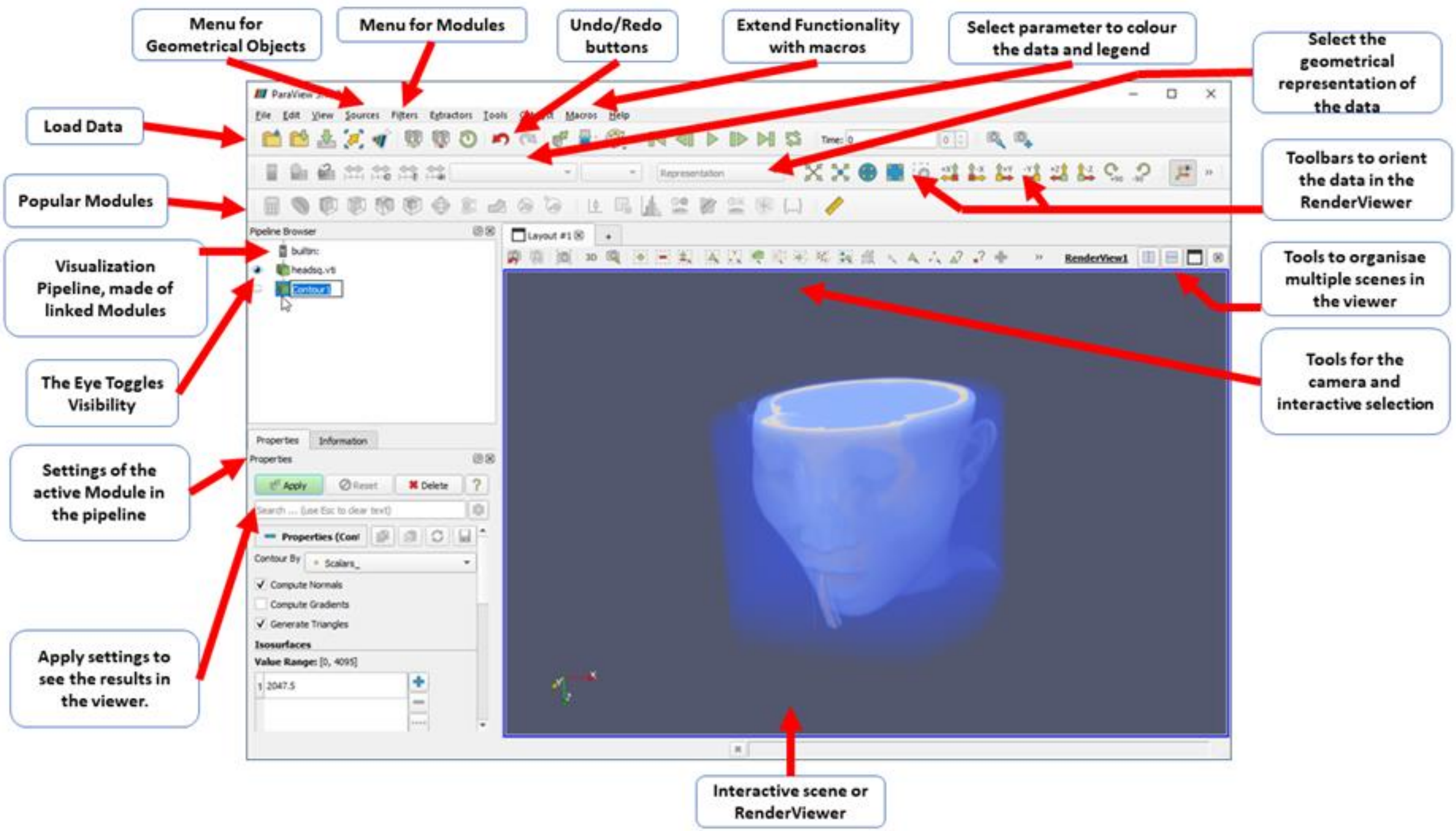
Overview of the Graphical User Interface

UNIVERSITY OF LEEDS



Common Functions

UNIVERSITY OF LEEDS



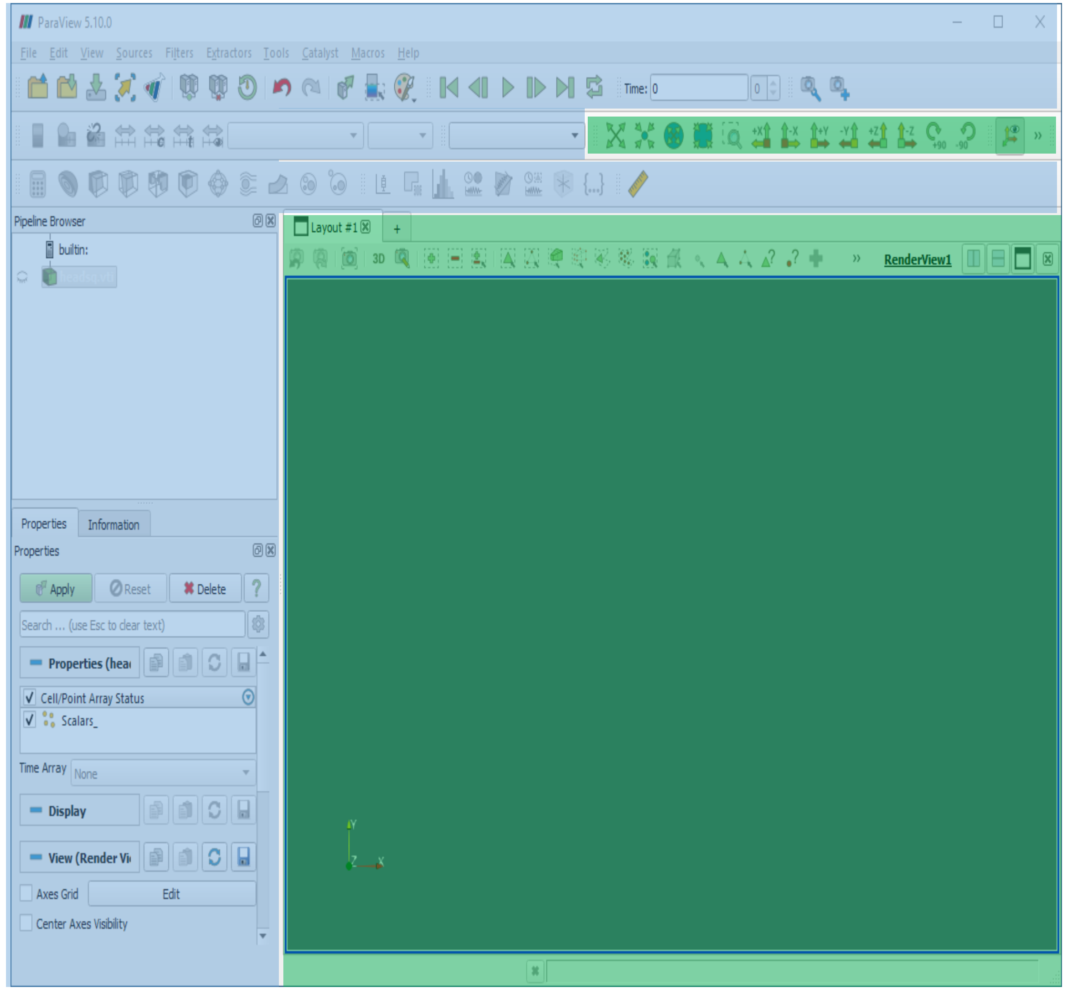
The ParaView GUI and Distributed Computing

Key:

GUI for the CPU

GUI for the graphics card

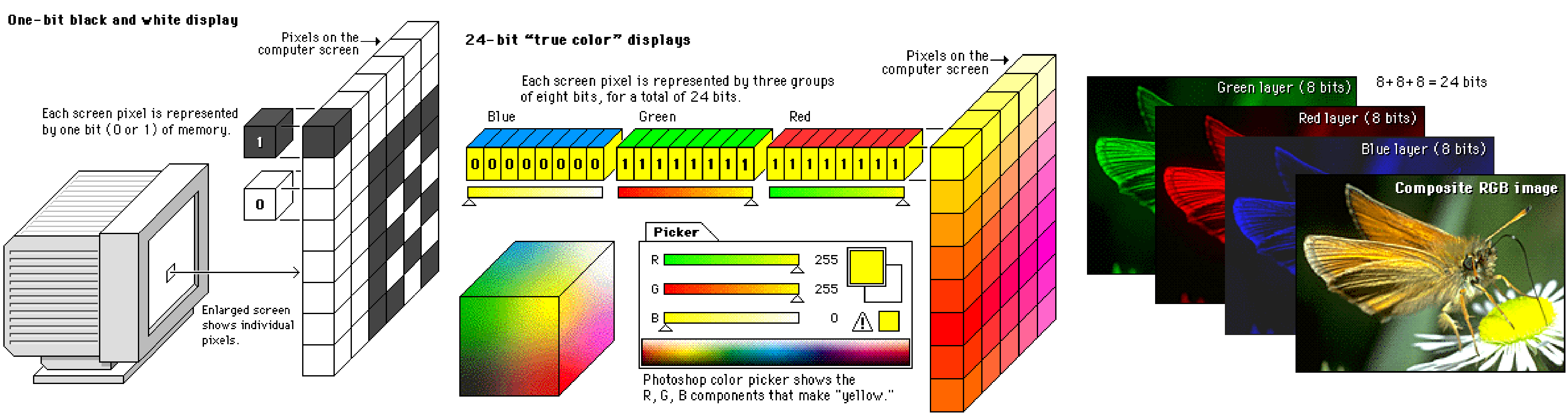
NB The redo, undo and apply changes functions apply only to CPU functions not the graphics card functions.



Digital Image Format Matches The Digital Screen Format

UNIVERSITY OF LEEDS

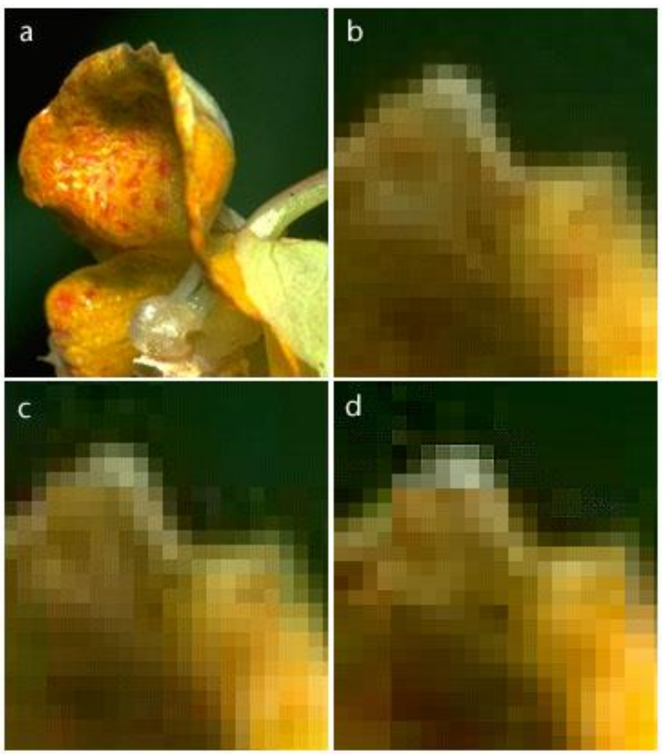
- You need to know about digital image formats because:
- the output images and animations in ParaView are in this format
- And because you can input digital images singularly and in stacks into ParaView.
- A screen is made up of pixels and these display colour as 4 parameters; Red, Green, Blue and Transparency. Digital images are based on this.



Digital Images Compression Artefacts

UNIVERSITY OF LEEDS

- Jpeg images have a losey compression. This means if you save an image in “low quality” the image irreversibly looses information.



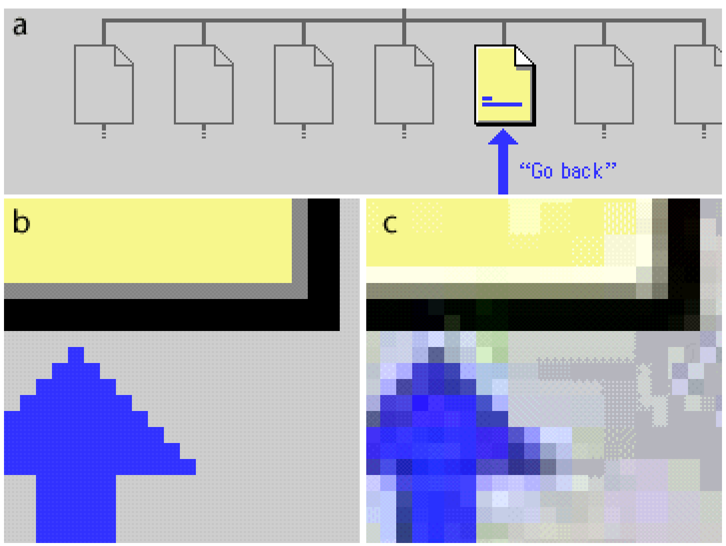
A,) original photograph and three detail views at different levels of JPEG compression:
B,) "excellent" quality jpeg compression
C,) "good" quality jpeg compression
D,) "poor" quality jpeg compression, notice the boxy quality of the image in.

Source: https://webstyleguide.com/wsg1/graphics/jpeg_graphics.html

Digital Image Compression Artefacts

UNIVERSITY OF LEEDS

Jpeg compression was designed for natural images with complex boundaries on objects. It does not handle well diagrams or images with harsh contrast and sharp edges, eg some visualizations.



A,) A diagram with harsh edges and sharp contrast

B,) A small detail with "excellent" quality jpeg compression

C,) A small detail with "poor" quality jpeg compression, shows a noise pattern on the transition area.

Source: https://webstyleguide.com/wsg1/graphics/jpeg_graphics.html

The Physics of Colour

UNIVERSITY OF LEEDS

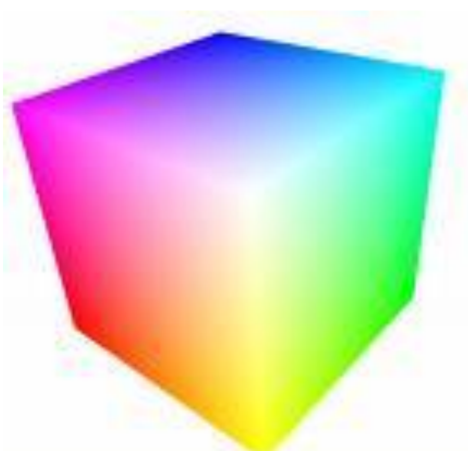
- The physics of colour is different to the digital image format. Computers can represent colours in other ways as well.
- I use Physics based colour models if I want to alter colours in a way that is more intuitive to how I perceive colour ie increasing the saturation of a colour to make an object stand out.

Colour Models – for Screens and Printing

UNIVERSITY OF LEEDS

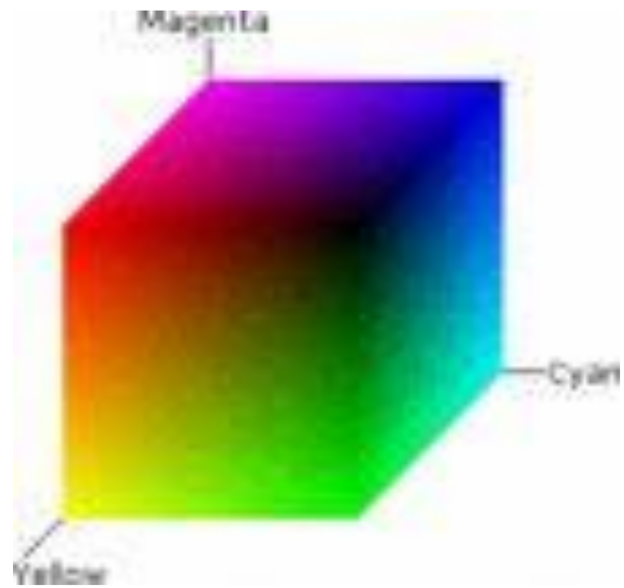
Red-Green-Blue (RGB)

- Used for Screens
- Red: 0% - 100%
- Green: 0% - 100%
- Blue: 0% - 100%



Cyan-Magenta-Yellow (CMYK)

- Used for printing

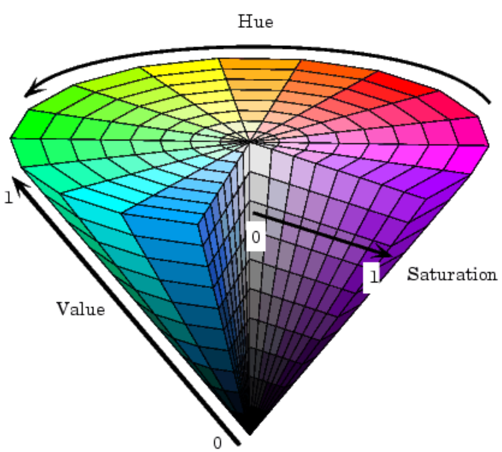


Physics Based Colour Models

UNIVERSITY OF LEEDS

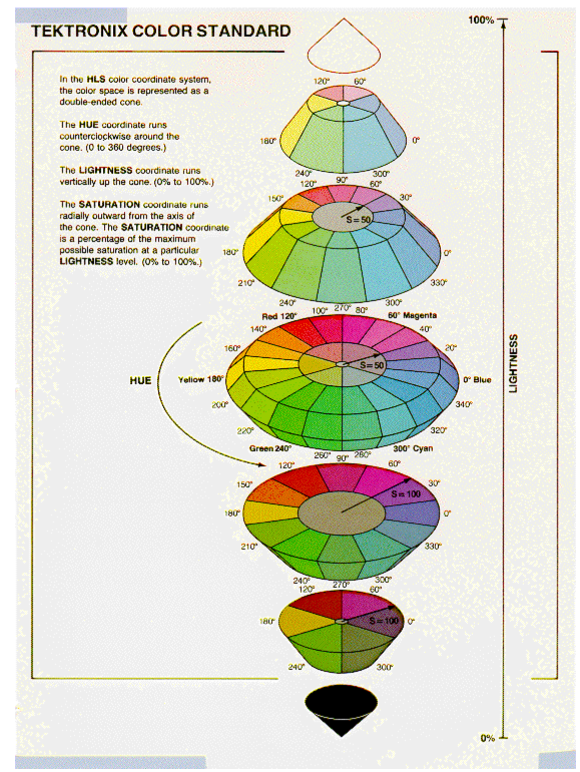
Hue-Saturation-Value (HSV)

- Hue is wavelength of colour
- Saturation is amount of pure colour
- 0% = gray, 100% = pure
- Value is brightness
- 0% = dark, 100% = bright



Hue-Lightness-Saturation (HLS)

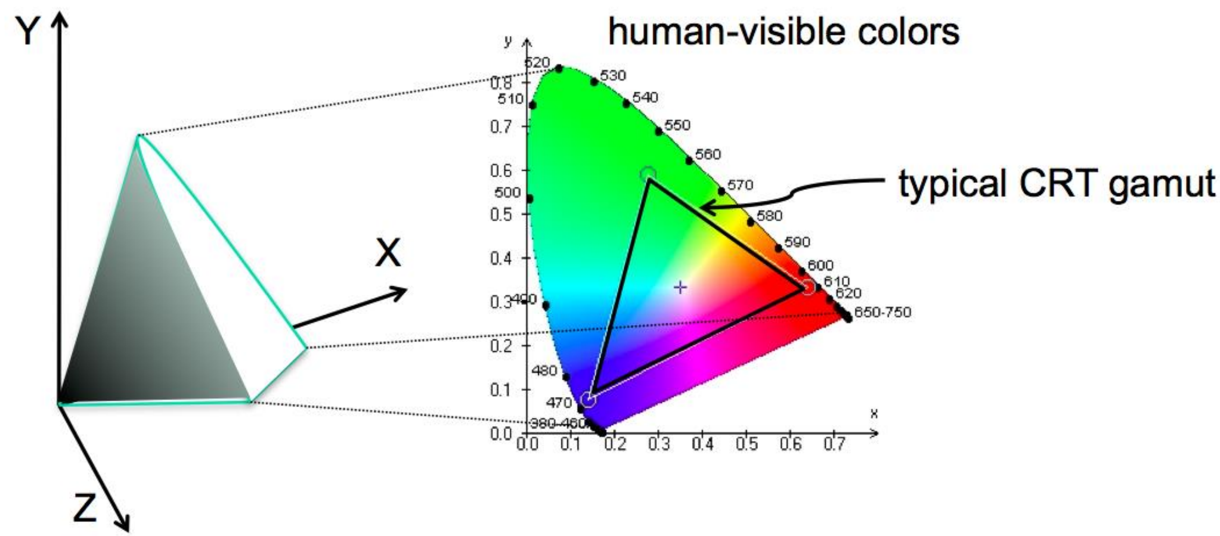
- White has lightness 1.0
- Pure colors have lightness 0.5



Obtaining Physically Accurate Colour

Colourometry: based on 3D space (xyz) for the primary colours bounded by what we can perceive.

Gamut: the colour pallet of a display device. You must calibrate display systems if the colours are to look the same.



Human Vision

UNIVERSITY OF LEEDS

Focusing the light

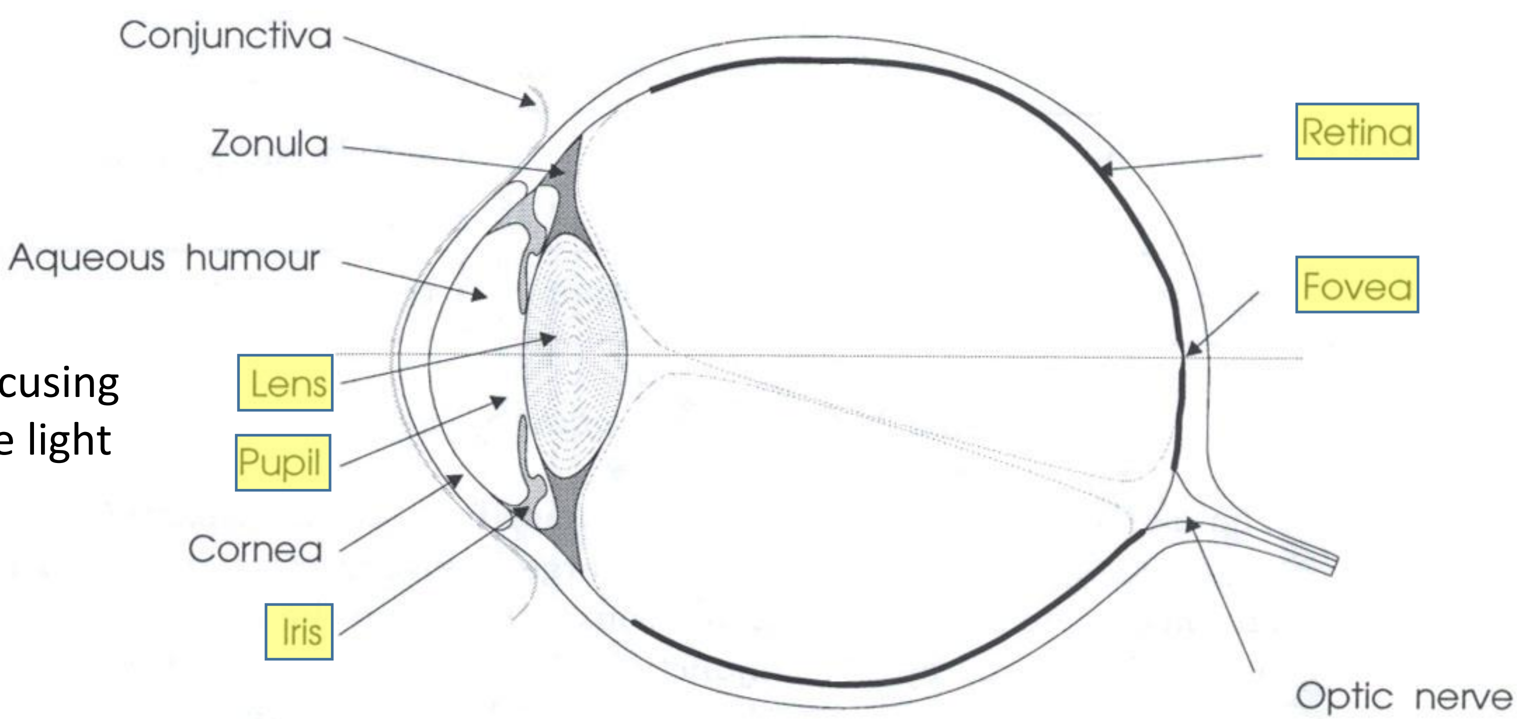


Figure 3: The human eye.

Where the light is focused so perception of tone and colour.

Source: Lilley, Lin, Hewitt, & Howard, "Colour in Computer Graphics", University of Manchester and http://web.mit.edu/6.813/www/sp16/classes/16-color/ for this and the next 14 slides.

Photoreceptors

Rods (grayscale vision across the retina)

- Only one kind (peak response in green wavelengths)
- Sensitive to low light (“scotopic vision”)
    - Multiple nearby rods aggregated into a single nerve signal
- Saturated at moderate light intensity (“photopic vision”)
    - Cones do most of the vision under photopic conditions

Cones (colour vision in the fovea)

- Operate in brighter light
- Three kinds: S(hort), M(edium), L(ong)
- S cones are very weak, centered in blue wavelengths
- M and L cones are more powerful, overlapping
- M centered in green, L in yellow (but called “red”)

Signals from Photoreceptors

UNIVERSITY OF LEEDS

Brightness

- M + L + rods

Red-green difference

- L - M

Blue-yellow difference

- weighted sum of S, M, L

Colour Differences underlie the Theory of opponent colours, good colours to pair are:

- Red and Green
- Black and white
- Blue and yellow

Key:
S – short wavelength (violet blue)
M – medium wavelength
L – long wavelength (red)

Information on the electromagnetic spectrum:
https://science.nasa.gov/ems/09_visiblelight

Chromatic Aberration

UNIVERSITY OF LEEDS

Different wavelengths focus differently

- Highly separated wavelengths (red & blue) can't be focused simultaneously

Guideline: don't use red-on-blue text

- It looks fuzzy and hurts to read

It hurts!

It hurts!

Blue Details Are Hard to Resolve

UNIVERSITY OF LEEDS

Fovea has few S cones

- Can't resolve small blue features (unless they have high contrast with background)

Lens and aqueous humor turn yellow with age

- Blue wavelengths are filtered out

Lens weakens with age

- Blue is harder to focus

Guideline: don't use blue against dark backgrounds where small details matter (text!)

Optical Illusions – Visualizers tend to be Obsessed with them!

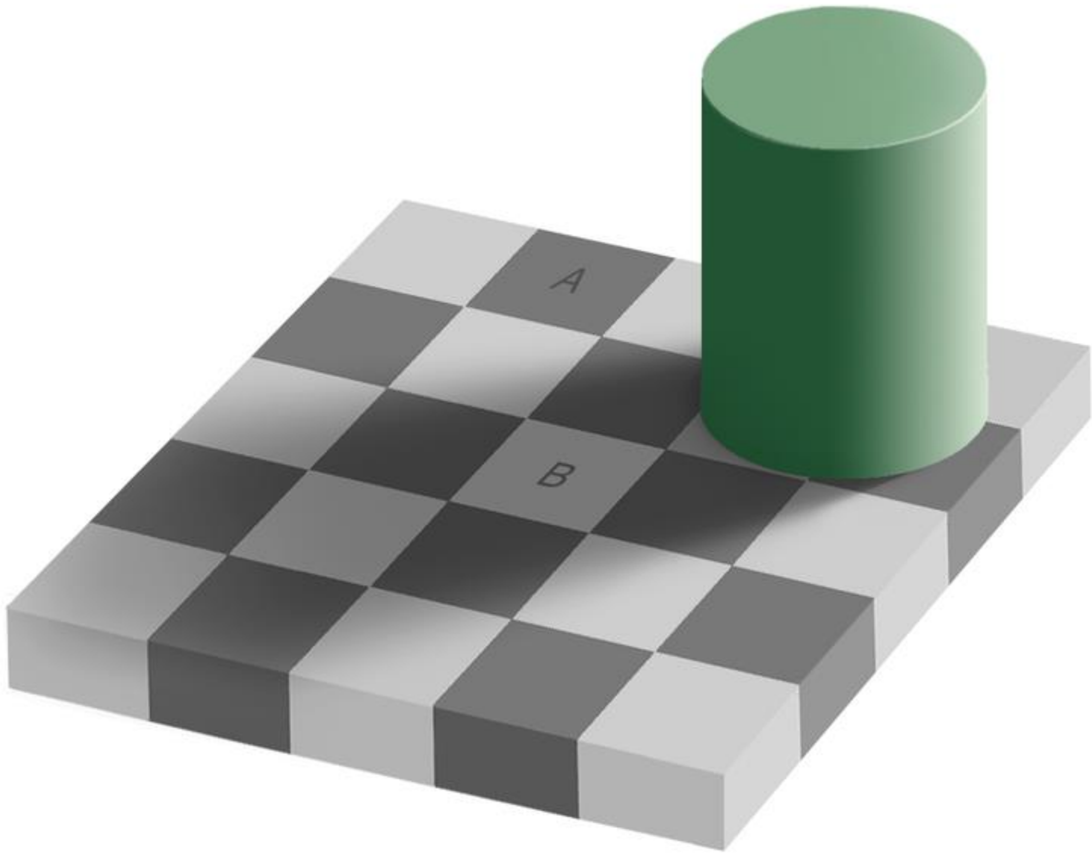
UNIVERSITY OF LEEDS

These four squares are all actually moving at the same speed. Your eye has more trouble distinguishing the blue from the black, so the blue squares appear to be moving in a jerky fashion.

Source: https://gizmodo.com/believe-it-or-not-these-four-squares-move-at-the-same-1559629417/+caseychan

Colour Constancy 1

UNIVERSITY OF LEEDS



Be careful when interpreting colour across an image or across a number of images.
NB use a legend and normalise your data ranges.

Colour Constancy 2

Be careful when interpreting colour across an image or across a number of images.
NB use a legend and normalise your data ranges.

Colour Blindness

UNIVERSITY OF LEEDS

Red-green colour blindness (protanopia & deuteranopia)

- 8% of males
- 0.4% of females

Blue-yellow colour blindness (tritanopia)

- Far more rare (~50 people in a million)

Guideline: don’t depend solely on colour distinctions

- use redundant signals: brightness, location, shape

A colour blindness simulator

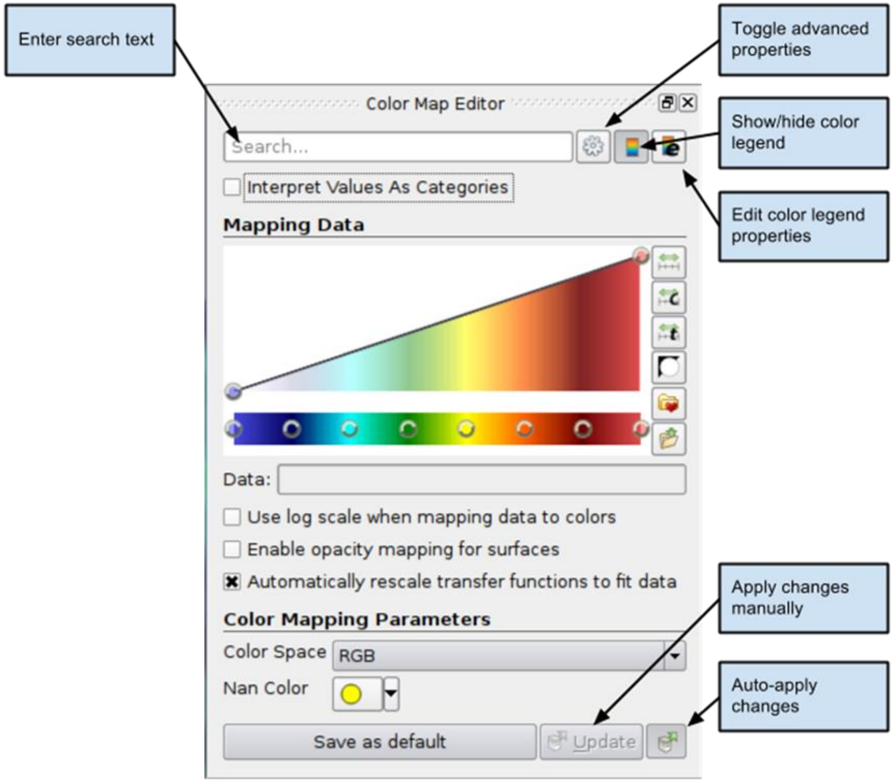
- http://www.color-blindness.com/coblis-color-blindness-simulator/

Colour and Pseudocolor

UNIVERSITY OF LEEDS

- Pseudocolor is the use of colour to represent data values in a visualization.
    - A transfer function maps the data to colours. This is commonly called the color or colour map.

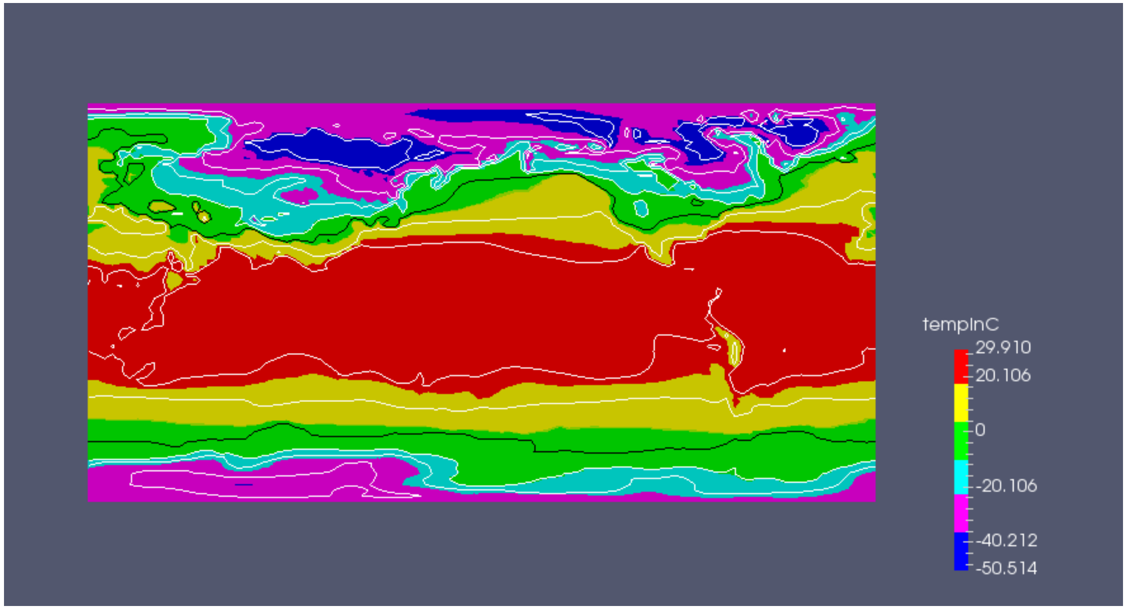
- The ParaView color map is a dockable panel
- Must press Update
    - Or auto-update



https://blog.kitware.com/using-the-color-map-editor-in-paraview-the-basics/

Colormap Examples

UNIVERSITY OF LEEDS

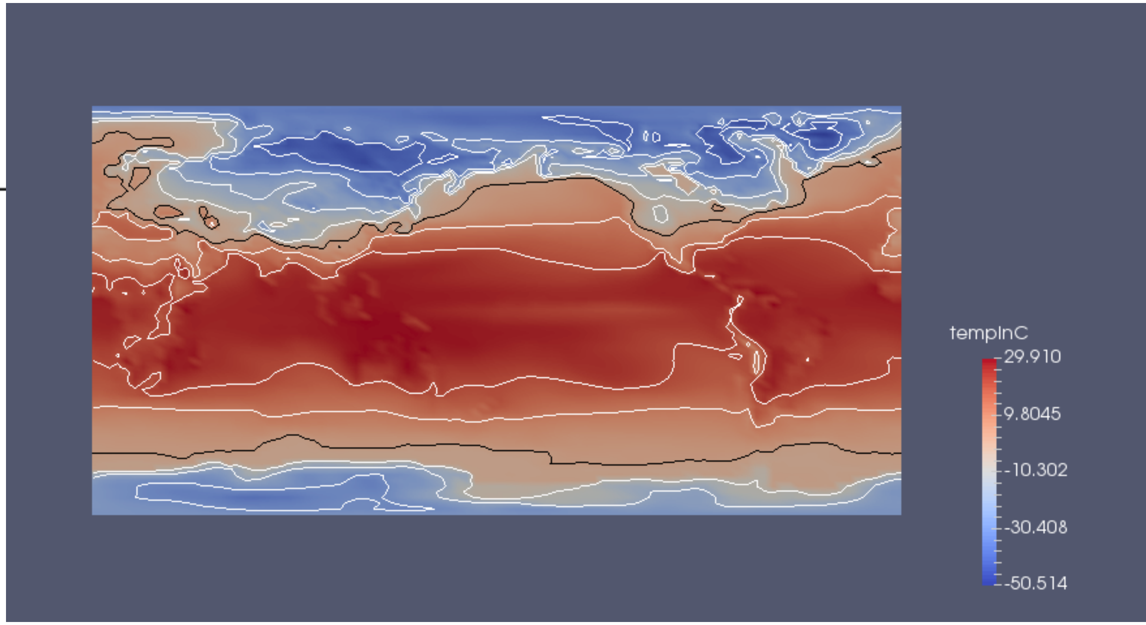


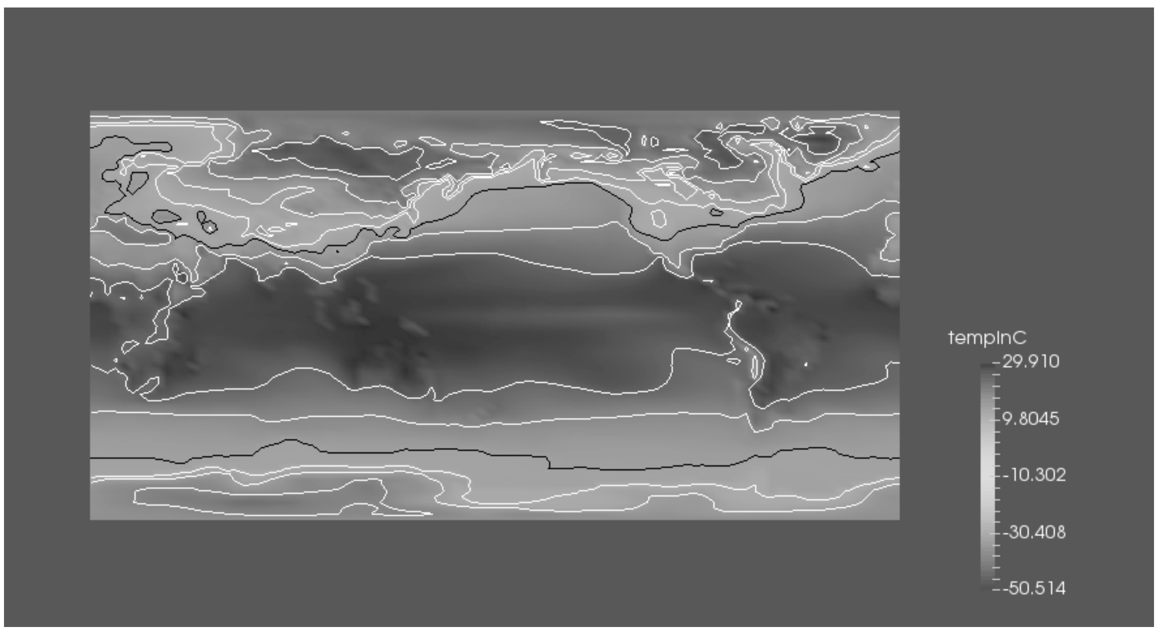
- The Blot colormap is similar to contours as the boundaries between colors are iso-contours and follow a contour line how ever the lines are not at set values and images with different max and min may look the same.

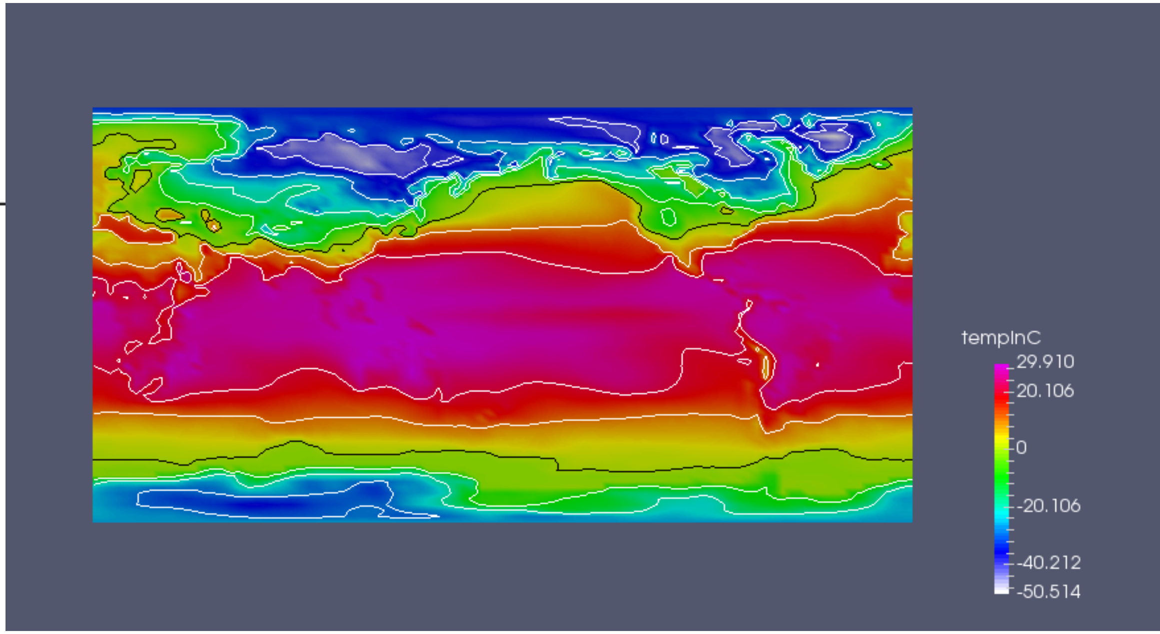
Image Intensity and Data Intensity

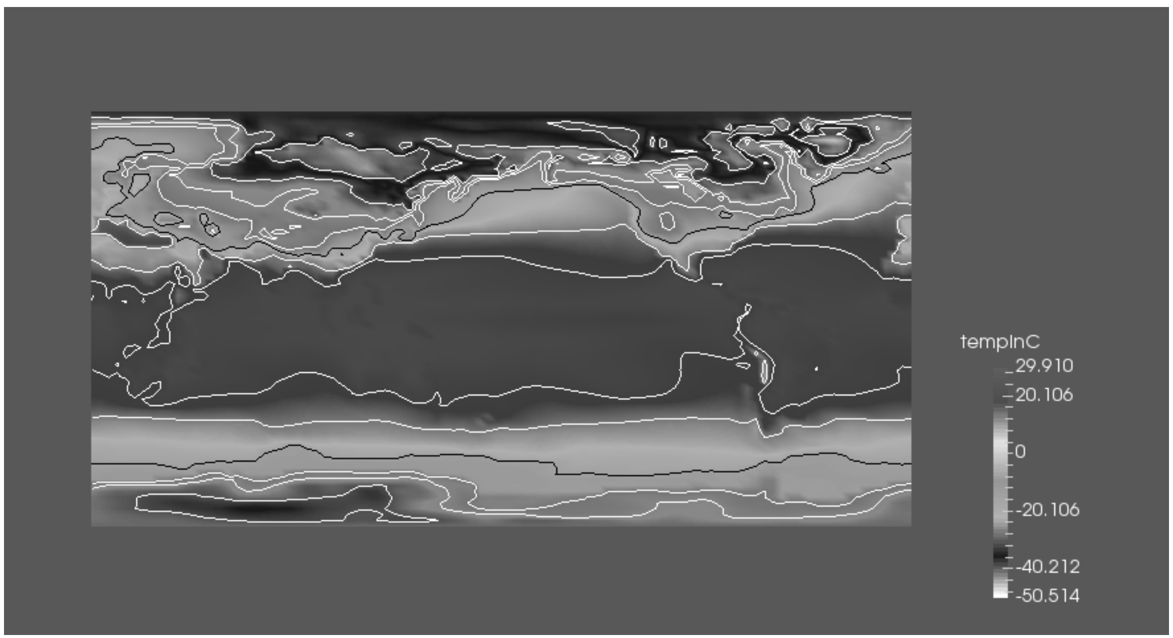
UNIVERSITY OF LEEDS

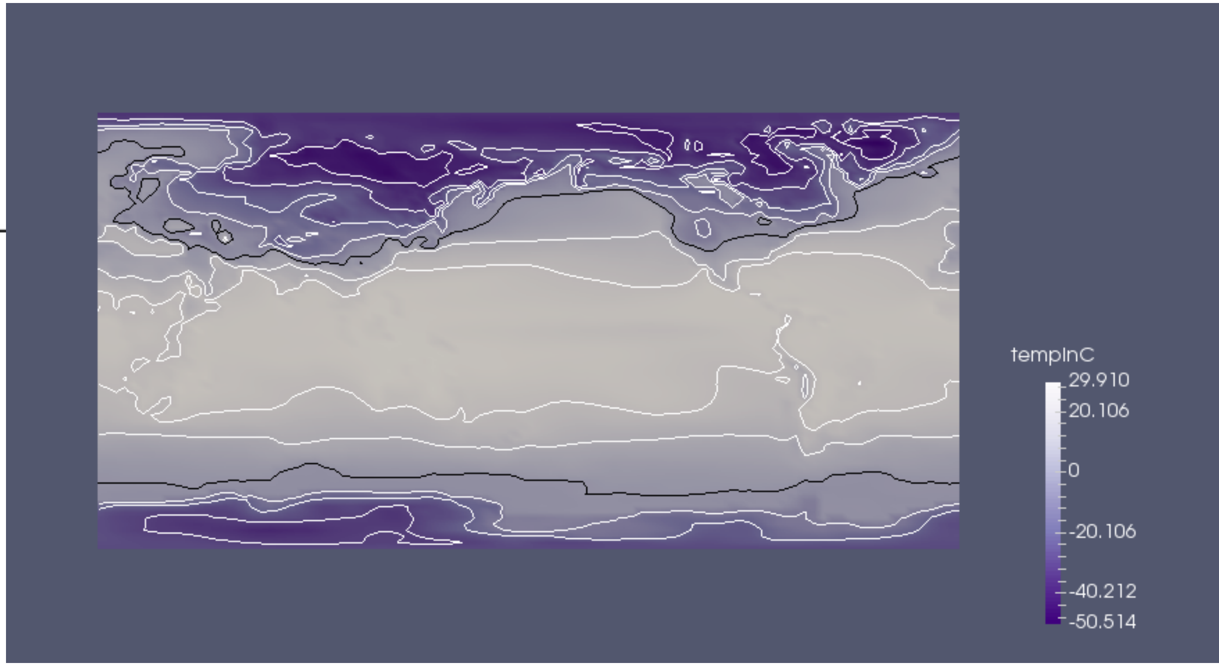
- We have 3 visual systems one in greyscale (brightness) which dominates in low light conditions and the others in colour which dominate in well lit conditions.
- If there is enough light our visual system prefers colour however the image intensity (the underlying grey scale) can confuse our perception.
- Good colourmaps increase (evenly) in brightness across the range of a parameter’s values – easiest to spot this in the legend when viewed in greyscale.

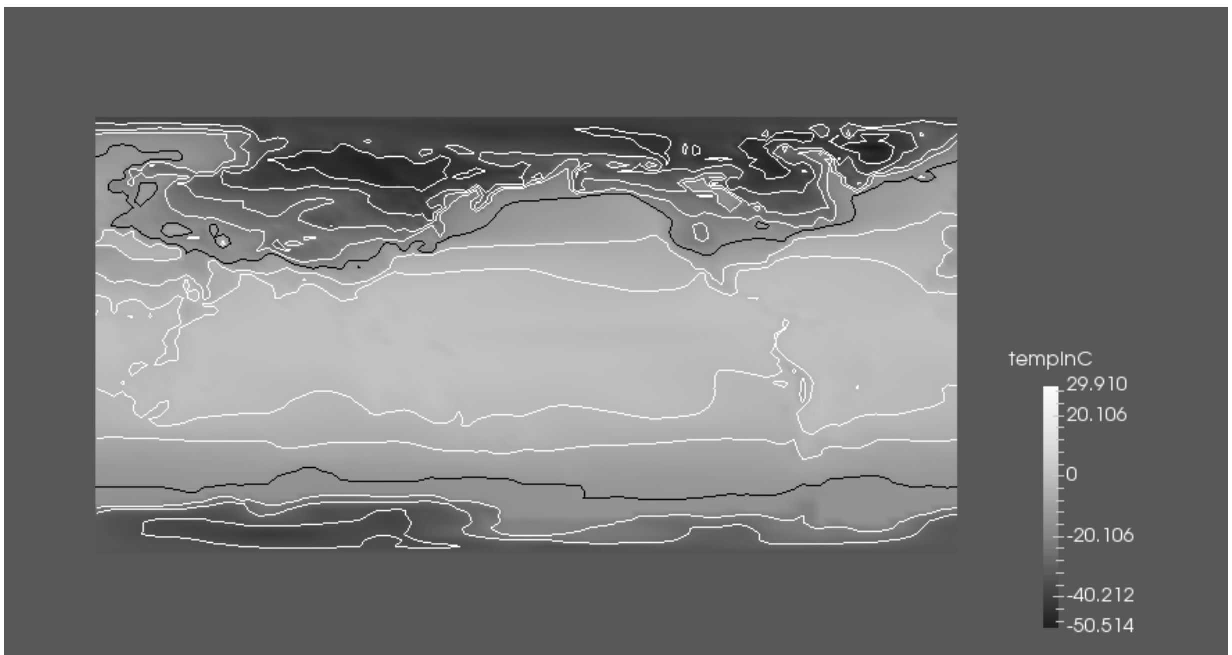

tempInC
29.910
9.8045
-10.302
-30.408
-50.514


tempInC
29.910
9.8045
-10.302
-30.408
-50.514

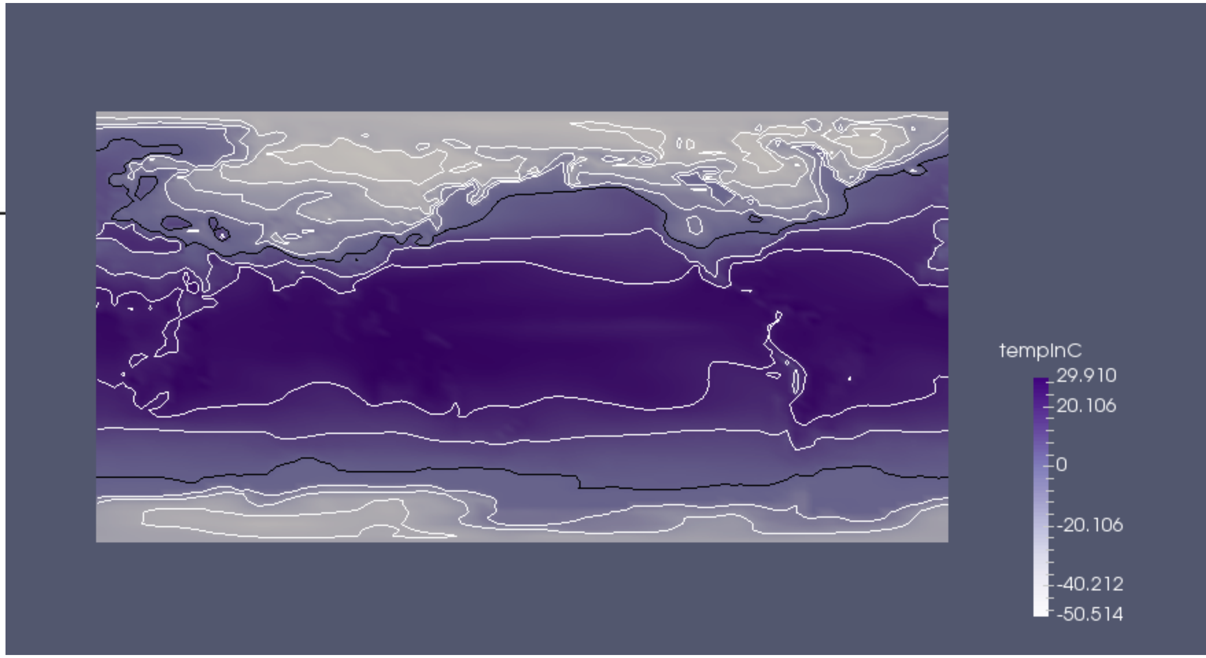

tempInC
29.910
20.106
0
-20.106
-40.212
-50.514

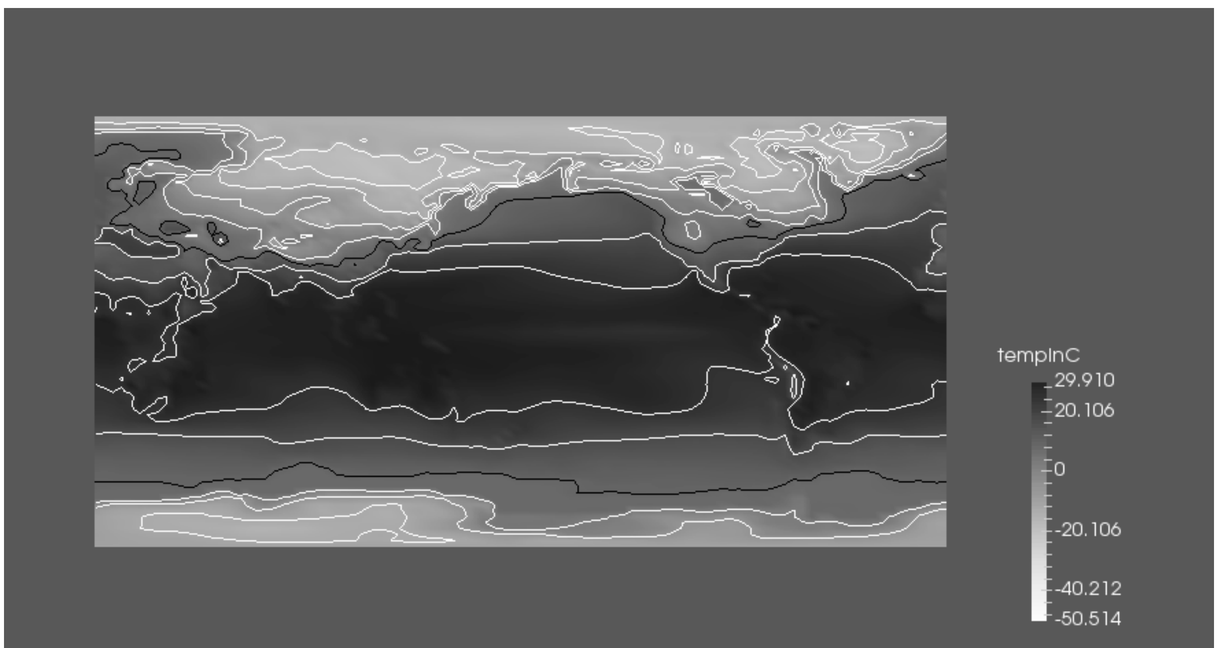

tempInC
29.910
20.106
0
-20.106
-40.212
-50.514





UNIVERSITY OF LEEDS


tempInC
29.910
20.106
0
-20.106
-40.212
-50.514


tempInC
29.910
20.106
0
-20.106
-40.212
-50.514

Colour My Guide Lines

UNIVERSITY OF LEEDS

- Be consistent - Always use the colormap that everybody uses in your area of science. It is good to be culturally accepted!
- Avoid saturated colours.
- Use different pseudo colour pallets for different parameters in the same scene.
- Have a light or white background when printing.
- Have good annotation and good contrast between that and the objects in your scene
- Do not use losey compression when saving images

Visualization Design

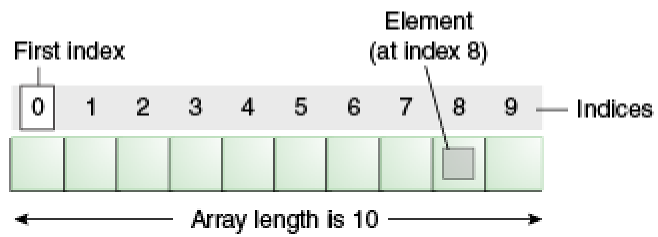
- How to create a visualization or set of visualizations that presents information to the viewer or viewers that is useful and accurate.
- Most design models focus on the combination of the data types and the visual information presented.
  - The design model used here is taken from the paper “Rethinking Visualization: A High-Level Taxonomy” by Tory and Moller.
- It is also important to consider the computational efficiency i.e., optimising the visualization pipeline.
  - This is a goal of *in situ* visualization

Storing Data in a Computer Program

- In a computer data is stored:
  - as binary (0s and 1s)
  - the length of a unit of binary data is determined by the hardware (32 bit or 64 bit)
  - Binary can be translated into integers (e.g., 1, 2), floating point numbers (e.g., 3.14) or alpha-numeric characters that makes up text or strings
  - An array is a list of values (integers, floating point numbers or strings) some programming languages have special types of lists that can be used for special purposes.

Visualization Data

- Data is held in arrays which are computer readable lists where the place in the list is given by an index.



- These data arrays may have a number of data values for each index e.g., temp and pressure.
- The location of the each data value in the array is called the mesh as is implicitly or explicitly given a Cartesian co-ordinate location (x, y, z).

Paraview Data Structures

UNIVERSITY OF LEEDS

A data array + a dims array, the location of the data is implied

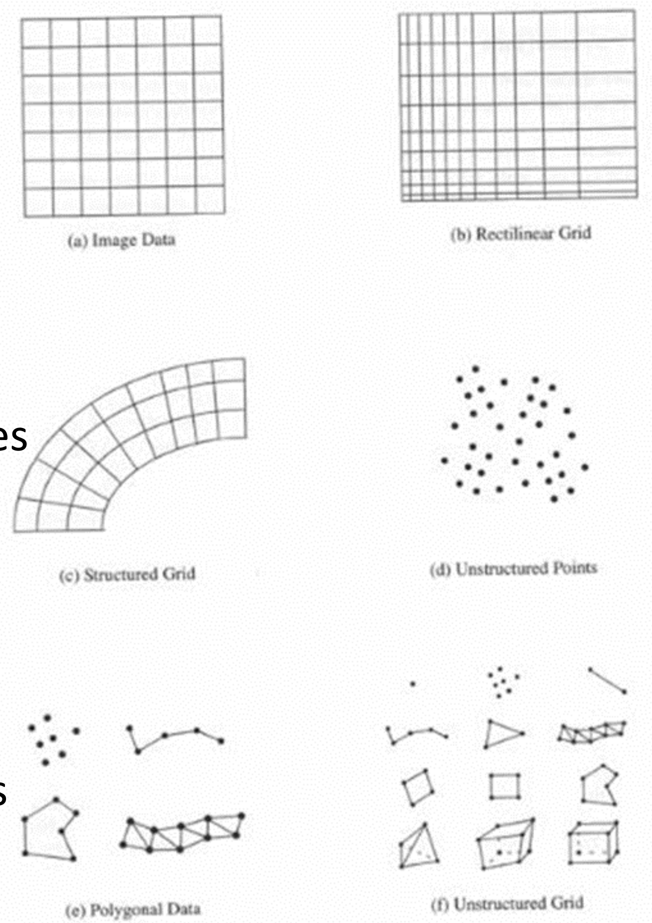
A data array + arrays for the relative location of data values along each principal axis

A data array + a coordinates array for each data value

A data array
OR
A data array + a coordinates array

A data array + a coordinates array + a connectivity array

For each type of cell:
a data array + a coordinates array for the point on the corner of each cell + connectivity array

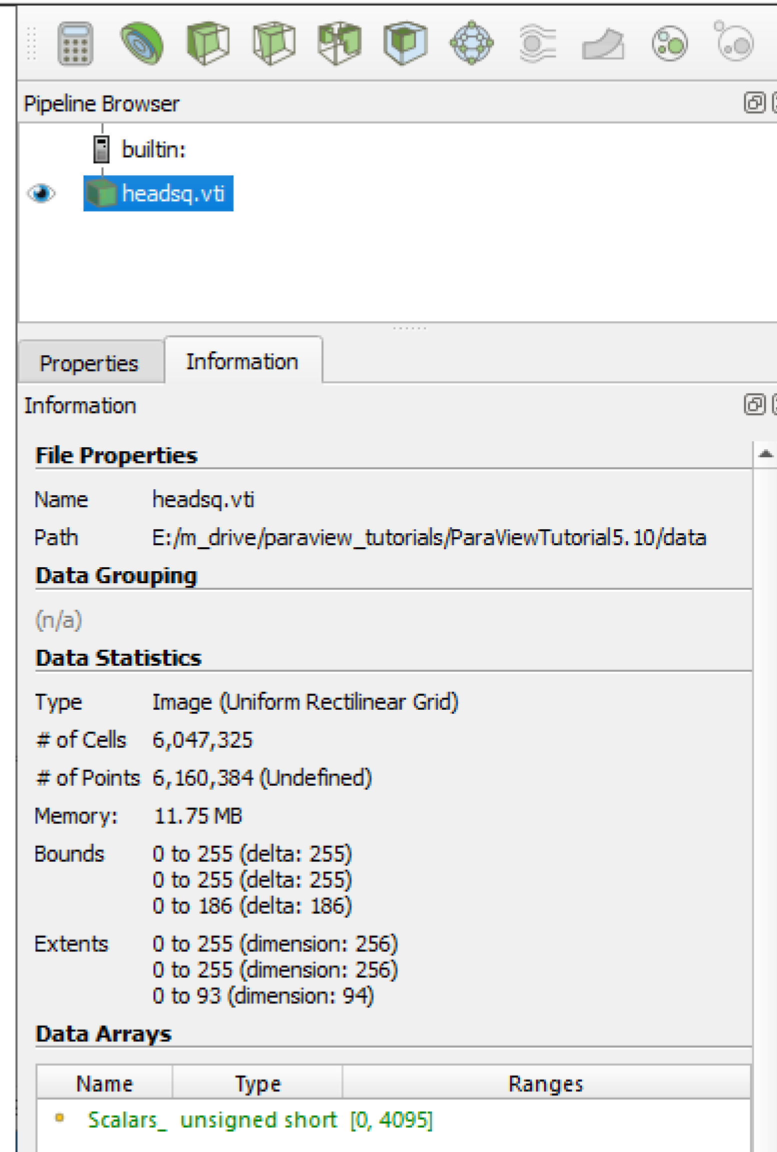


http://www.bu.edu/tech/support/research/training-consulting/online-tutorials/paraview/
Page 32, chap 3, Understanding data in The ParaView Guide

Inputs to Modules in ParaView 1

UNIVERSITY OF LEEDS

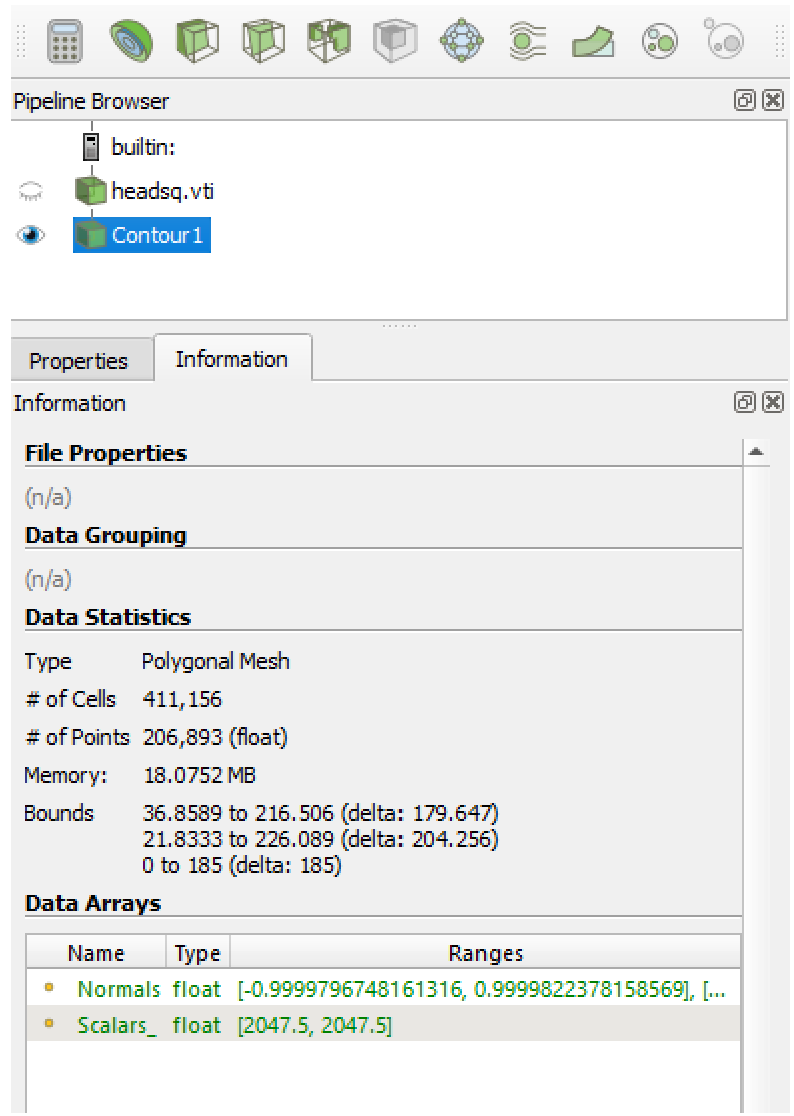
- Contours module as an example of data structure relevance.
- Read in head data and highlight it.
- Information panel shows it is a uniform grid of dims x=256, y=265, z=94 so has 6,160,384 points.
- Common menu has grayed out all modules not for uniform rectilinear grid data.



Inputs to Modules in ParaView 2

UNIVERSITY OF LEEDS

- Add the contour module.
- Now in the Information Panel we can see the data is a Polygonal Mesh.
- Different modules in the Common Menu are grayed out.
- NB some modules can read in many data types.



The Function of Visualization Data Structures

UNIVERSITY OF LEEDS

- Data in visualization systems is structured because:
  - It is easy for the visualization software to handle.
  - It has to be in a format that the graphics hardware and openGL understands.
- It is best if the data produced by simulation or analysis software is in a structure that is easy for the visualization software to handle otherwise the data will need further processing.
  - Use visualization data structures when designing a new simulation.

Design a Visualization

UNIVERSITY OF LEEDS

- How to create a visualization or set of visualizations that **presents information** to the viewer or viewers that is useful and accurate.
- What **data** do you have? Information and meta-data is related to data but is not data. Information is what is conveyed from data.
  1. What data can you collect?
  2. What visualization technique will you use to convey information from the data?
- Who will **view or be interested in your research?** – the visualization should answer a research question and **present it in a form that is tailored to your audience.**

Discrete or Continuous Design Model?

- Discrete / Continuous Classification
- The design model not the data is classified as continuous or discrete.
- Continuous models assume that the data can be interpolated.
- Discrete models assume that the data cannot be interpolated.
- Interval and ratio data can be visualized as either discrete or continuous
- Nominal or named (data with shared arbitrary characteristics) and ordinal or categorised (data with a small number of finite values) data can often only be visualized as discrete data as it cannot be interpolated

Converting between discrete and continuous

- Continuous model to discrete model: leave the data points as discrete entities
- Discrete to continuous: parameterise the model or embed it into a continuous space.

High-Level Visualization Taxonomy

UNIVERSITY OF LEEDS

		Display Constraints		
Design Model		Given	Display Attributes Constrained	Chosen
	Continuous	Images (e.g., medical) Fluid/gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc. Globe – distribution data (e.g., elevation levels)	Distortions of given / continuous ideas (e.g., flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analysis
	Discrete	Classified data / images (e.g., segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (e.g., city locations)	Distortions of given /discrete ideas (e.g., 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (e.g., file structure) Arbitrary multi-dimensional data (e.g., employment statistics)

Display Constraints

- Traditionally the way to display data is identified by whether the spatial location is given (is part of the data) or is chosen (is the choice of the visualizer).
- But in some cases the visualizer can choose the location e.g., selecting the projection for geographical data it could be displayed on a globe or a flat map
- Other attributes that can be chosen are colour, transparency and time

Low-Level Taxonomy of Continuous Models

UNIVERSITY OF LEEDS

# Independent Variables	Data Structure				
		Scalar	**Vector**	**Tensor**	**Multi-variate**
	1D	• Line graph			Combine scalar, vector, & tensor methods
	2D	• Colour map • Isolines	• LIC • Particle traces • Glyphs		
	3D	• Volume Rendering • Isosurfaces		• Tensor ellipsoids	
	nD	Multiple 1D, 2D or 3D views			

Low-Level Taxonomy of Discrete Models

UNIVERSITY OF LEEDS

Structure

Graph and Tree Visualization:

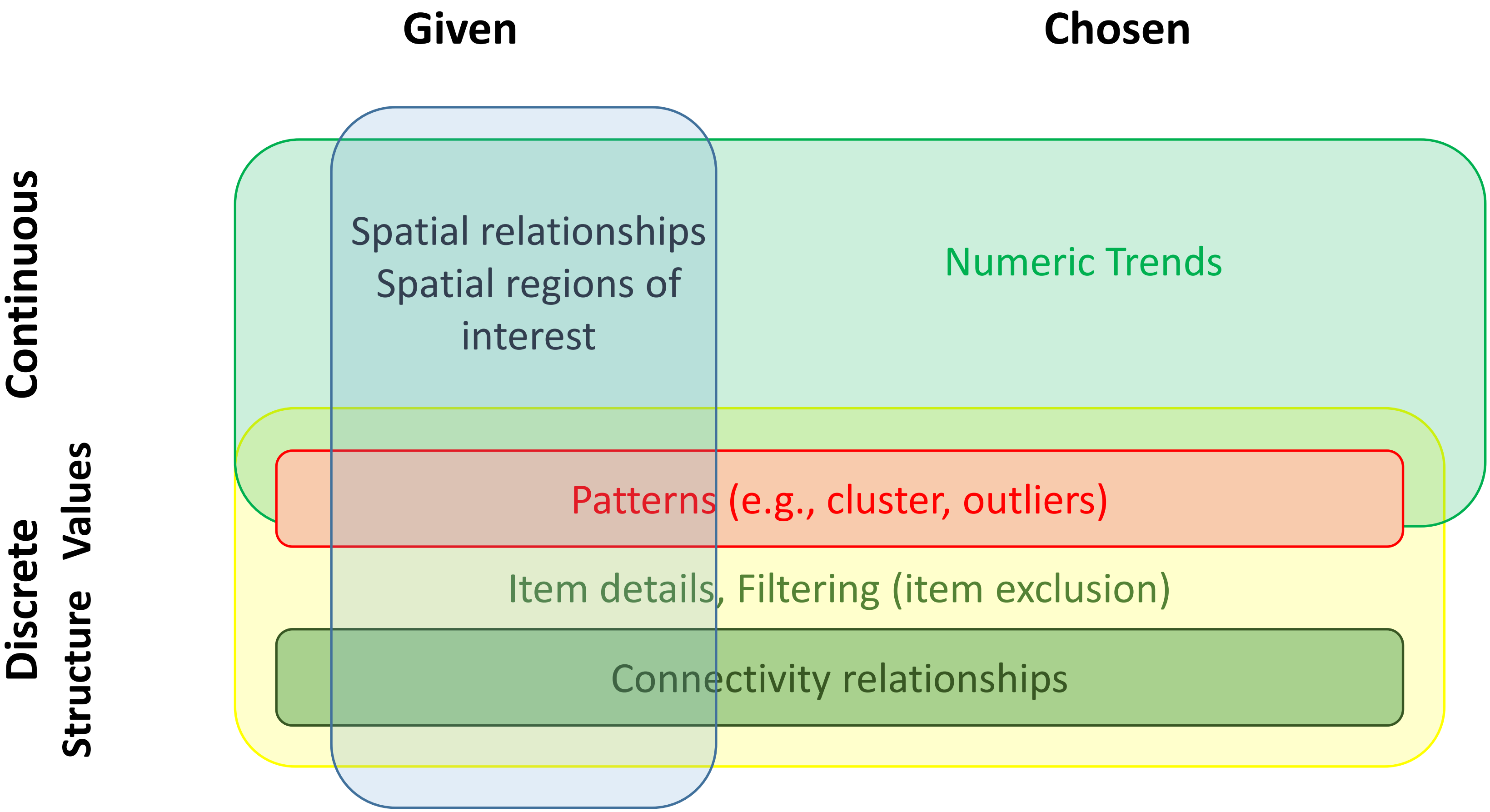
- Node – link diagrams (2D and 3D)
- Hierarchical graphs
- Space-filling mosaics

Values

		Variable Types	Example Techniques
Number of Variables	**2D**	1 dep + 1 indep variable	• Scatter plot • Bar chart
	3D	1 dep + 1 indep variable or vica versa	• 3D scatter plot • 3D bar chart
	nD	Any number of dep and indep variables	• Charts + colour • Multiple views • Glyphs • Parallel coordinates

Fitting Visualization Tasks to the Design Model

UNIVERSITY OF LEEDS



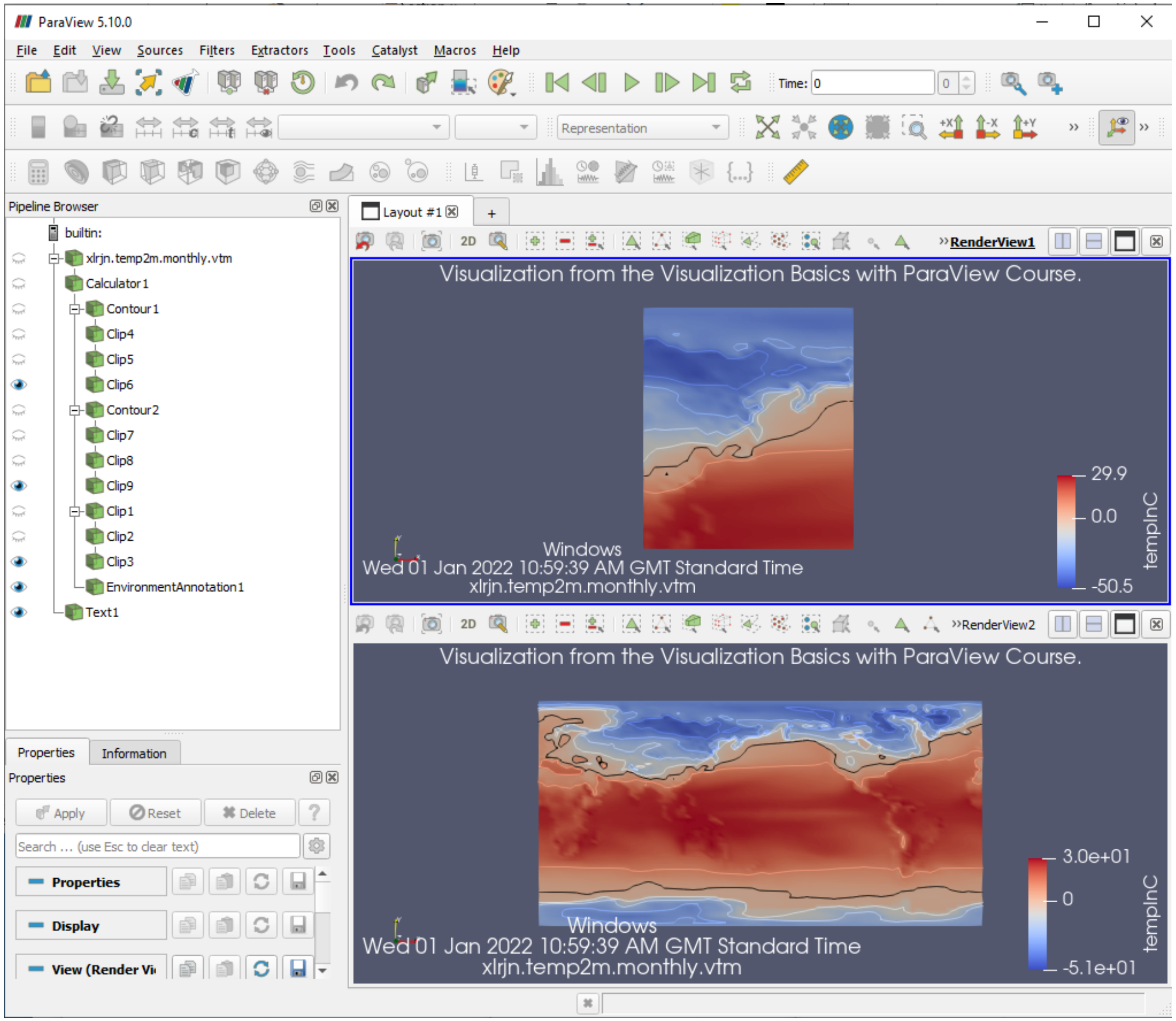
Optimizing Your Visualization Pipeline

- The Visualization Design helps you select appropriate visualization techniques.
- Once you have a good visualization with its associated pipeline you are ready to optimize the pipeline before you use it in production mode.
- Look for modules alter the size of the data e.g., clip reduces the data but contour if it creates a complex isosurface may increase the data
- Look for computationally expensive modules which the user may wish to alter input parameters triggering it to re-execute.

Example from the exercise sheet

UNIVERSITY OF LEEDS

- If you move the clipping to between the reader and calculator
- Reduce the number of clip modules from 9 to 3
- Increase the contour modules to 2 to 4
- This is good if the contouring is computationally cheap and the contour is not complex.



Thanks to:

N8 CIR:

https://n8cir.org.uk/

The School of Computing, University of Leeds, UK

EPSRC for funding me (EP/R025819/1)