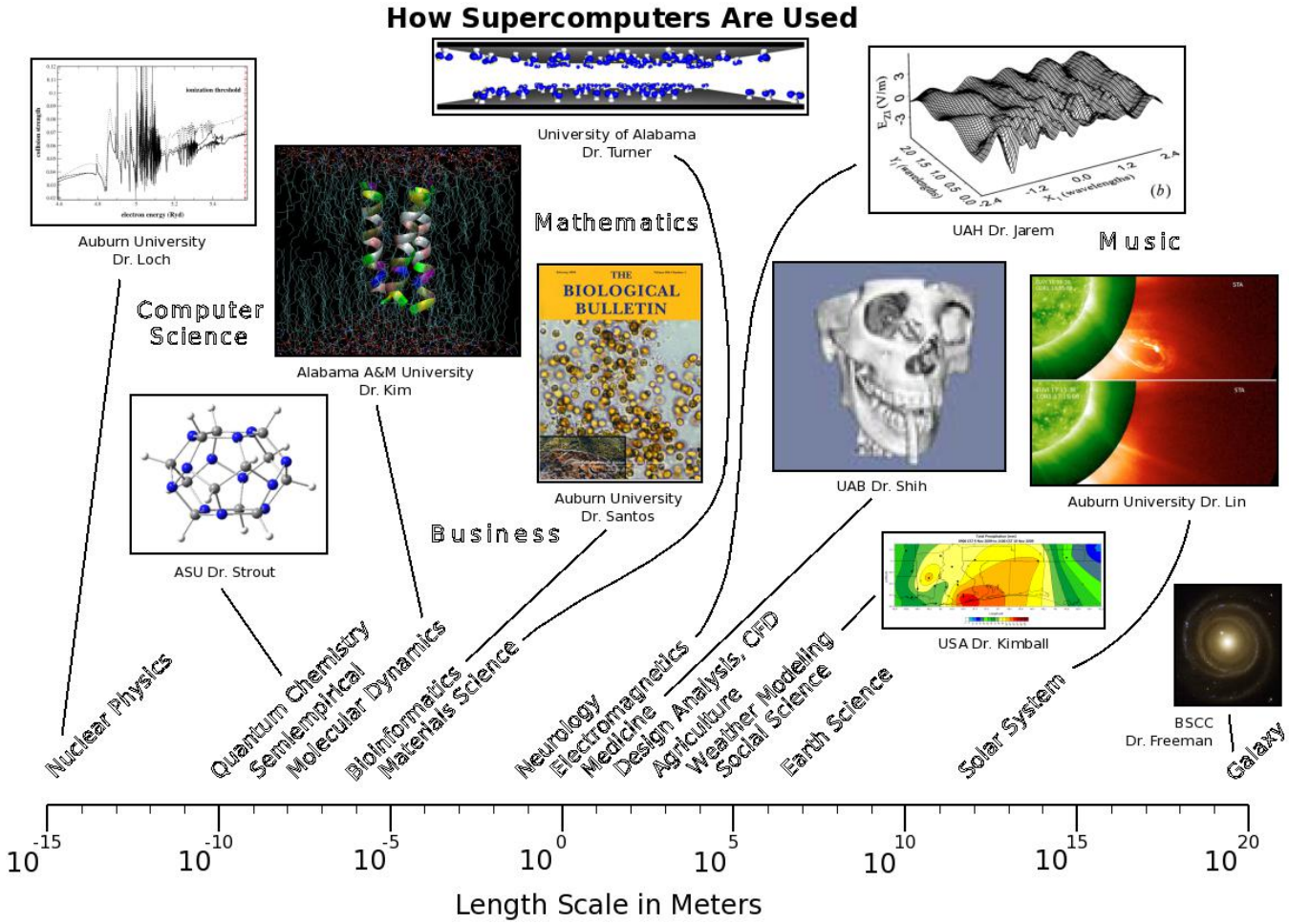
Getting Started with Visualization

About this white paper: This paper was written by David C. Young, an employee of CSC. It was written as supplemental documentation for use by the HPC account holders at the Alabama Supercomputer Center. This was originally written in 2011, and updated in 2013.

Visualization is the graphical display of data, concepts, metaphors, information, strategies, or processes. This can be as simple as a bar chart or can utilize complex 3D animated rendering of multiple types of data in a single scene. Visualization is valuable because it gives a more intuitive and immediate understanding of data than a table of numbers.

As an example of visualization, below is an illustration (created with Dia) of the type of work done on the high performance computing (HPC) systems at the Alabama Supercomputer Center. The organization of this image by length scale gives students learning about high performance computing some perspective on the types of problems that can be better understood by a computer simulation.



The first step in a visualization project is to have some ideas as to how the data might be displayed. Here are some sources that give some interesting visualization examples.

http://www.visual-literacy.org/periodic_table/periodic_table.html
http://www.visualcomplexity.com/vc/
http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches/
http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html
http://visualizing.org/
Recent years issues of National Geographic and the Harvard Business Review

Sometimes visualization capabilities are included in the graphic interface for a specific piece of software. For example, some chemistry software programs now come with a graphic interface to display graphical representations of molecules, electron density, and other chemically interesting properties. Sometimes data must be imported into a dedicated visualization program.

It can take some investigation to find software that will display your information most effectively. Regardless of the software that is used, it will probably take some effort to learn to use the software, and quite a bit of tinkering with the settings to find the best way to display the data. However, it can be well worth the effort to make a strong impact on your audience.

The following are some open source or free visualization programs worth noting.
NOTE: Not all of these are presently installed on the HPC systems at the Alabama Supercomputer Center.

Graphviz - visualization of trees, flow charts, connections between items.
http://www.graphviz.org/

OpenDX - visualization of 2D & 3D data sets.
http://www.opendx.org/

VisIT - for visualization of very large data sets
https://wci.llnl.gov/codes/visit/

GNUPlot - a graphing program with a command line interface
http://www.gnuplot.info/

JMOL - for chemical structure data
http://jmol.sourceforge.net/

VMD - for molecular dynamics data & protein structures
http://www.ks.uiuc.edu/Research/vmd/

Rasmol - scriptable chemical structure visualization
http://www.umass.edu/microbio/rasmol/

Gephi - for visualizing very complex graphs
https://gephi.org/features/

Codeswarm - for visualizing software development projects
http://code.google.com/p/codeswarm/

Gourse - software version control visualization
http://code.google.com/p/gource/

Wordle - create a word cloud of words found most frequently in a document
http://www.wordle.net/

Cytoscape - for complex network visualization
http://www.cytoscape.org/

ParaView - popular for CFD data
http://www.paraview.org/

Dia Diagram Editor - an open source alternative to Visio.
http://dia-installer.de/

Inkscape - an open source alternative to Adobe Illustrator.
http://inkscape.org/

Circos - popular for gene sequence data
http://circos.ca/

A number of general purpose mathematical programming environments also include various levels of graphical display capability. These include;

Mathematica – a symbolic manipulation and numeric mathematics package
http://www.wolfram.com/

R – a programming language for statistical modeling
http://www.r-project.org/

Matlab – a powerful mathematical tool, originally built around matrix algebra
http://www.mathworks.com/products/matlab/

In recent years, a plethora of open source visualization packages have become popular. However, there are also some important commercial software packages to be noted. These include;

Excel – the graphing capabilities of the world's most popular spreadsheet are sometimes overlooked
http://office.microsoft.com/en-us/excel/

Origin – a full featured data analysis and graphing tool
http://www.originlab.com/

AVS – a powerful and customizable general purpose visualization tool
http://www.avs.com/

Visio – one of the most popular tools for creating conceptual diagrams
http://office.microsoft.com/en-us/visio/

Iris Explorer – a powerful and customizable visualization tool
http://www.nag.com/Welcome_iec.asp

Adobe Illustrator – the Adobe tools are a favorite for many graphic artists. Illustrator is the ultimate computer drawing program.
http://www.adobe.com/products/illustrator.html

Photoshop – an incredible range of functionality built around the manipulation of images
http://www.adobe.com/products/photoshop.html

Tecplot – a visualization tool built around CFD and engineering data
http://www.tecplot.com/

There are many more visualization tools than can be mentioned specifically in this small introduction. The reader is encouraged to do some investigation of tools appropriate for the type of data they wish to display.