Better than Free:
Data Explorations with Public Data and Software Tools

Introduction to Free Tools

Mary Haley

SEA 2016 Tutorial
April 6, 2016

Quick survey (2/2)

What software tools do you currently use?

- Fortran, C, C++, Java
- Scripting languages:

  IDL, MATLAB, NCL, Python, R, Ruby, etc.
- GUIs:

  ncview, panoply

Categories of tools

- Command line tools
- GUIs
- Scripting languages
  - Python packages

Command line tools

- Quick way to look at data
- Easy to perform multiple operations on file(s)
- Somewhat limited functionality
- Data files may have to meet requirements
- Can be slow

Command line tools for examining data

- ncdump – NetCDF
- ncl_filedump – NetCDF, HDF, HDF-EOS, GRIB
- h5dump – HDF5
- wgrib / wgrib2 – GRIB1 / GRIB2

NCO – NetCDF Operators
(works on HDF5 too)

netCDF Operators (NCO) Software Stack

CDO – Climate Data Operators

NetCDF and GRIB 1/2

wgrib / wgrib2

- wgrib - slice and dice GRIB1 files
- wgrib2 – “four drawers of kitchen utensils as well as the microwave and blender”

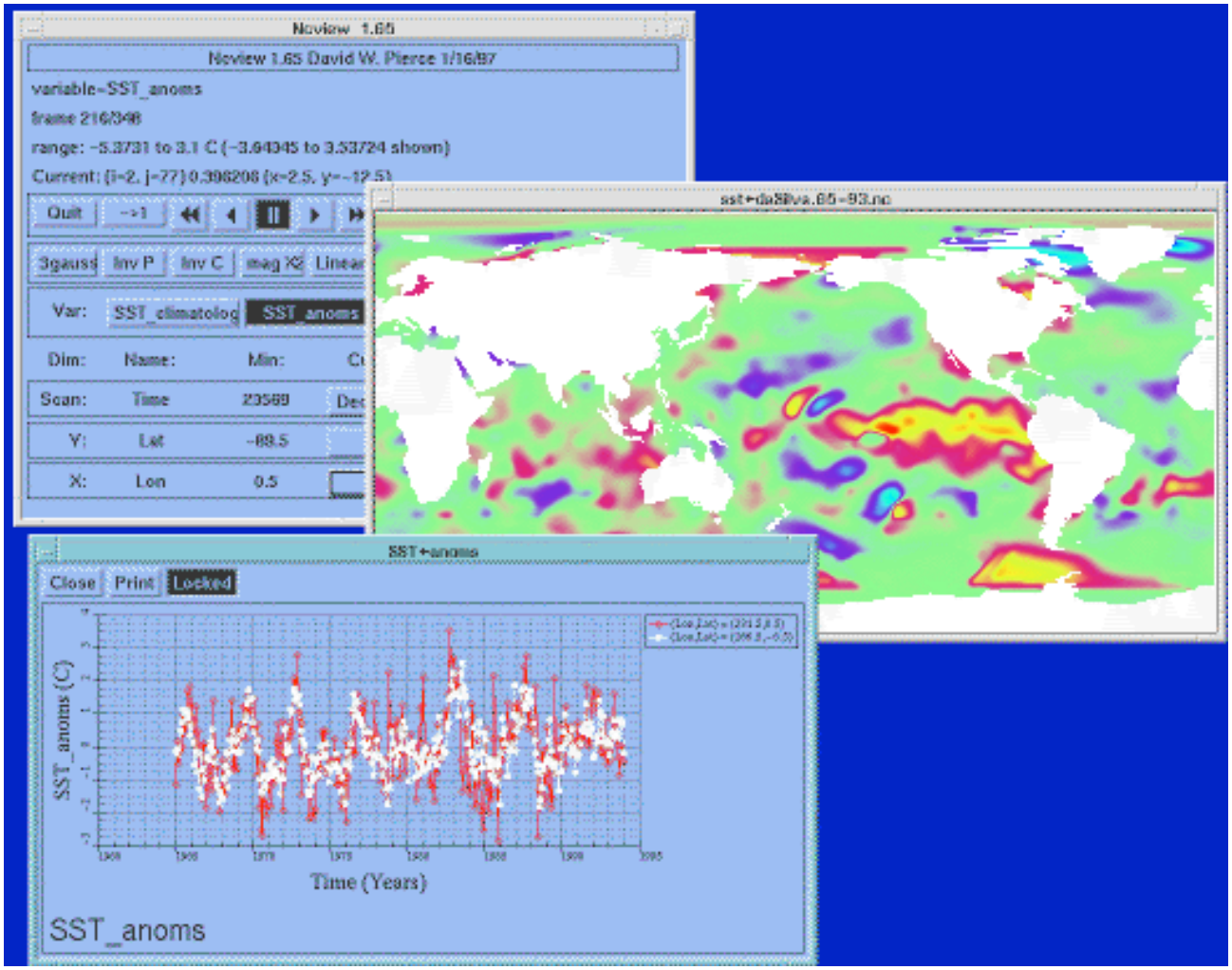
DEMO

Examining NetCDF and GRIB files

GUI-based tools

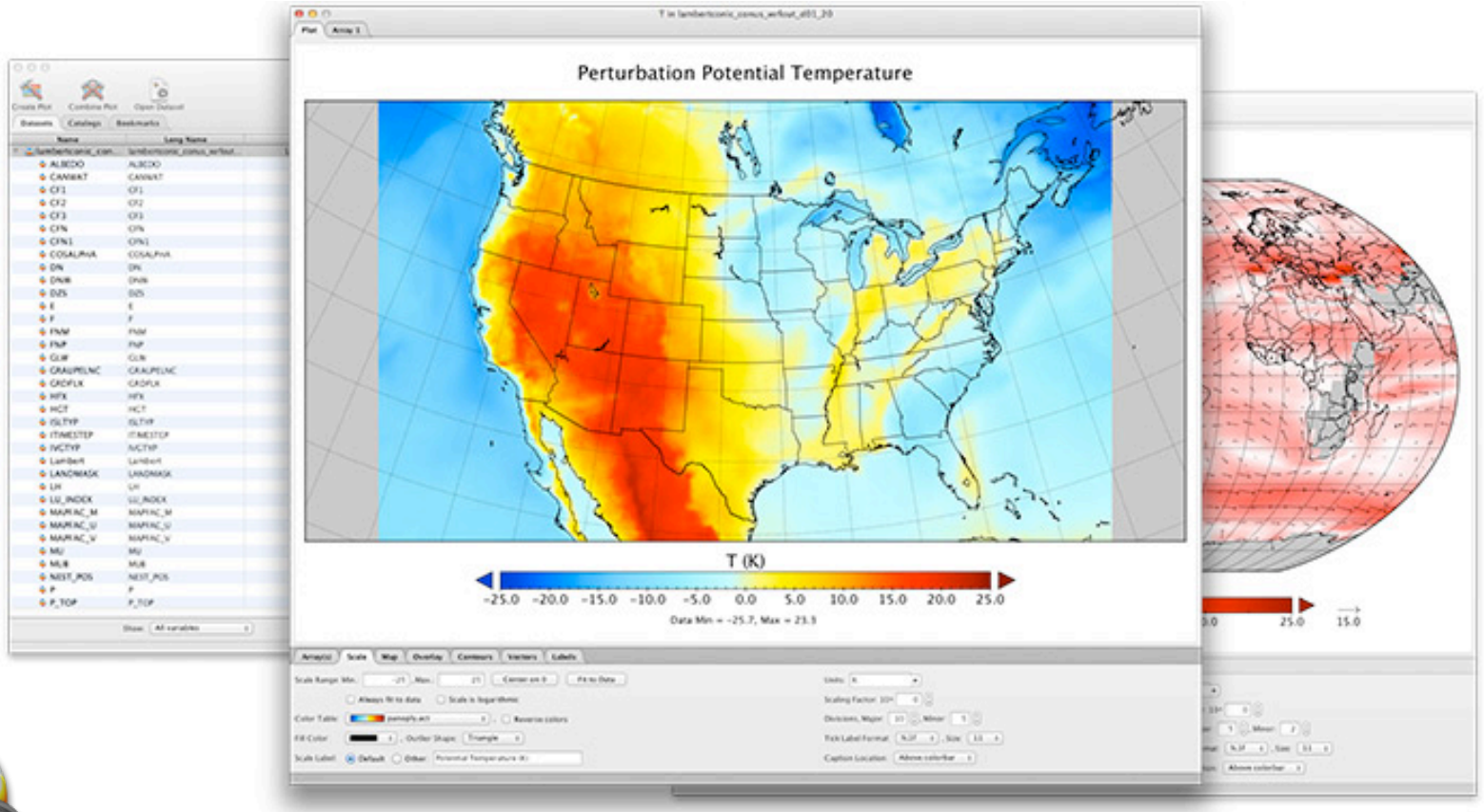
- Quick exploration of file contents
- Quick visualization of data
- Can be intuitive to use
- May have limited functionality

ncview – netCDF visual browser



Panoply – netcdf, hdf, grib data viewer

(defn: impressive collection of things)



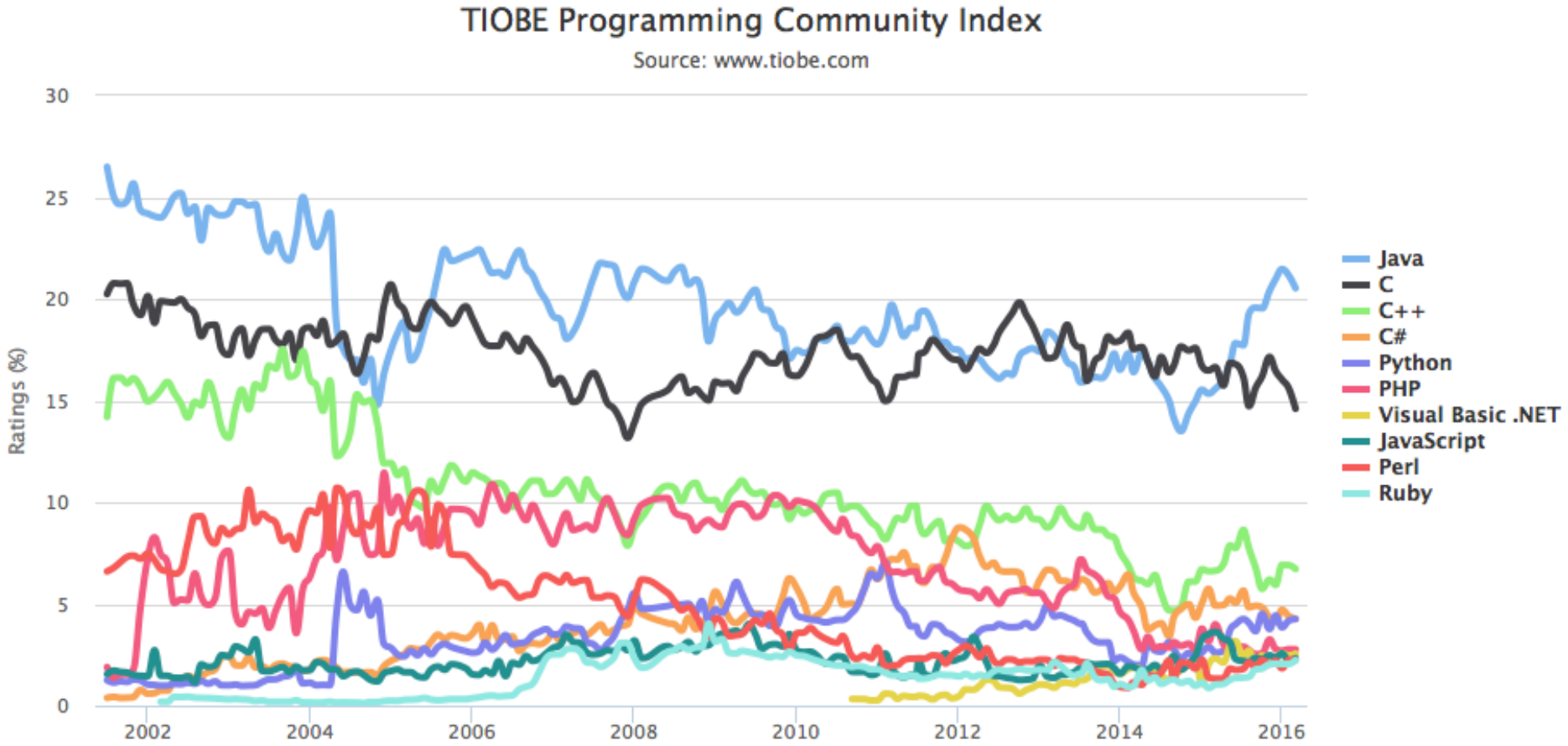
DEMO

panoply - SKIP

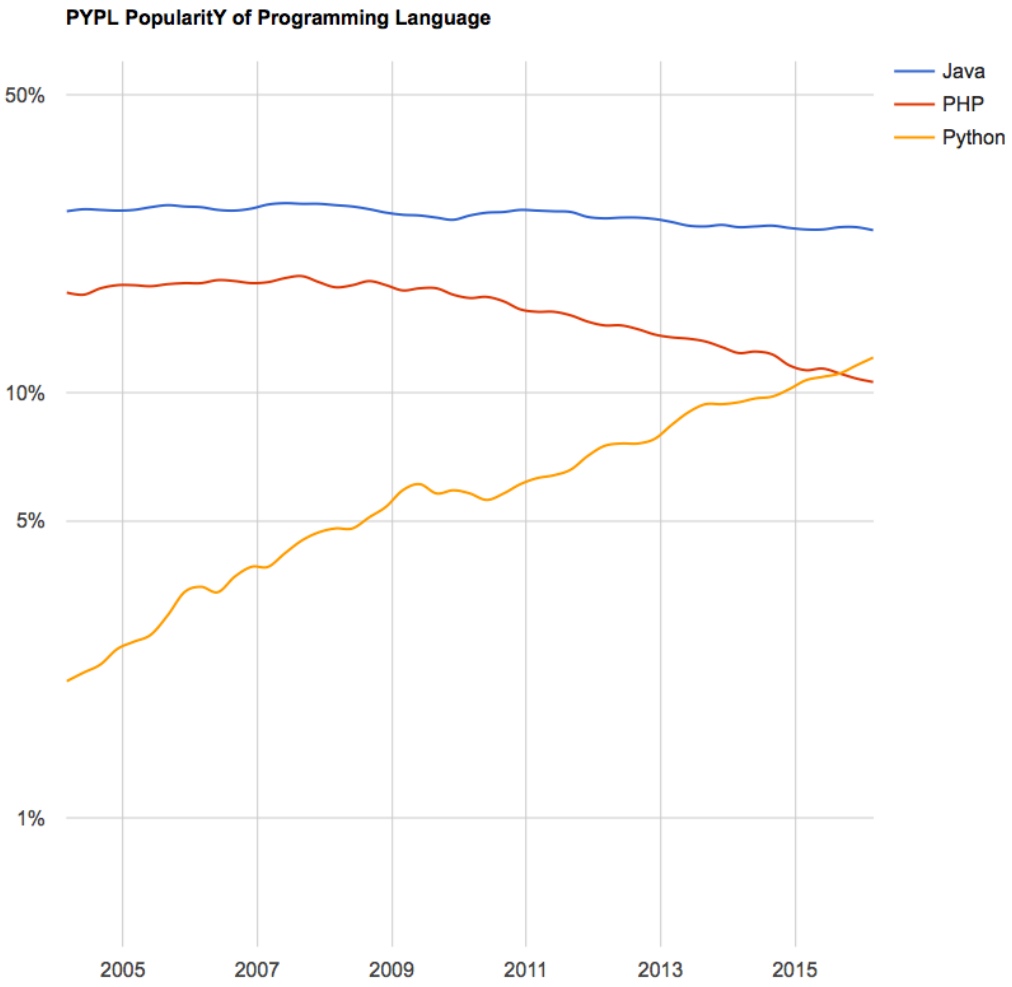
Scripting languages

- Powerful – loaded with functionality
- Have to write code; easier than C/Fortran
- Can be slow
- Can be difficult to get support
- Looks good on resume!

"Python is powerful... and fast;
plays well with others;
runs everywhere;
is friendly & easy to learn;
is Open. "



www.tiobe.com



http://pypl.github.io/PYPL.html

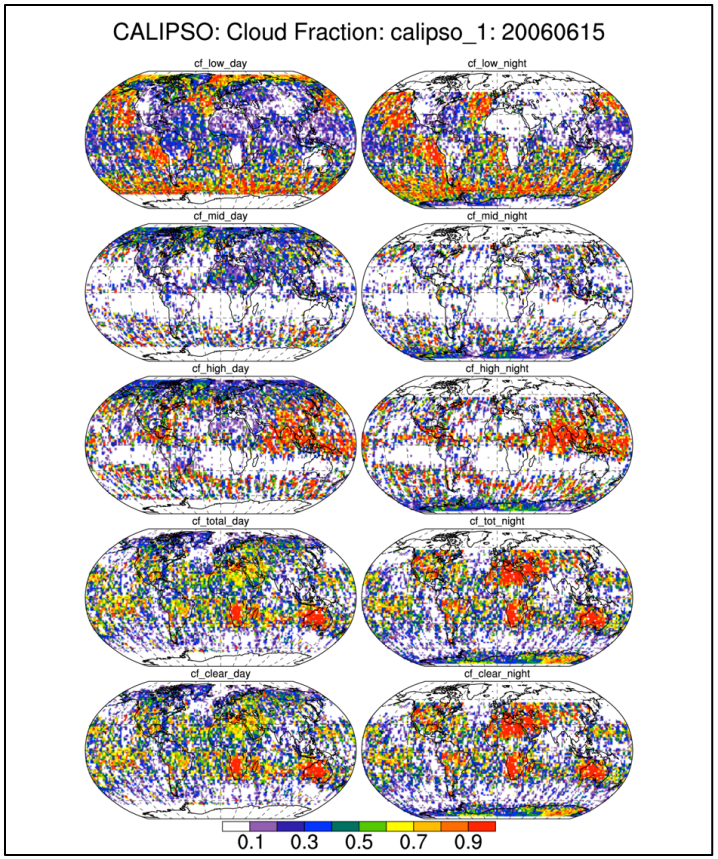
Open data science platform that empowers the scientific Python community

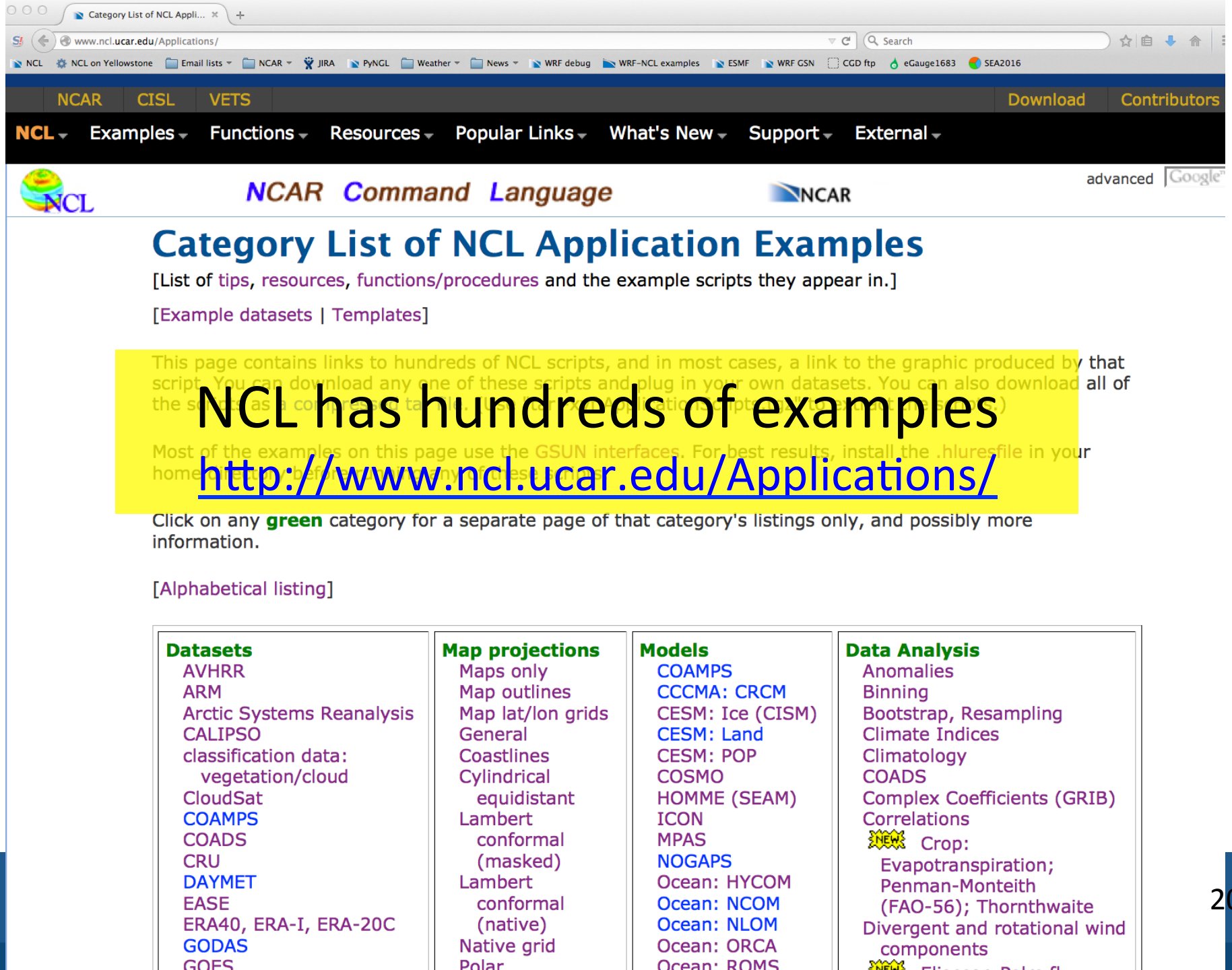
www.continuum.io

NCAR Command Language

Developed at NCAR, tailored for the analysis and visualization of geoscientific data

1. Simple, robust file input and output
2. Hundreds of analysis functions
3. Publication-quality visualizations
4. Well-supported, 1000+ examples
5. Extensive website, training
6. Now under “conda”





NCL has hundreds of examples

http://www.ncl.ucar.edu/Applications/

Python packages

- There's LOTS of them (pro and con)
  - More than one way to do something
- Wide range of functionality
- Easy installation via "pip" and "conda"
- Lack of unified interfaces and support

Tip

If writing your own public Python package:

Style Guide for Python Code

https://www.python.org/dev/peps/pep-0008/

Python packages and tools for AOS

- cf-python
- ESMPy
- H5py / PyTables
- Iris
- netcdf4-python
- ESMPy
- Pygrib
- PyNIO
- pandas / xarray
- Siphon
- Cfplot
- matplotlib / basemap
- cartopy
- PyGRaDS
- PyFerret
- PyNGL
- PyAverager / PyReshaper
- seaborn
- RPy

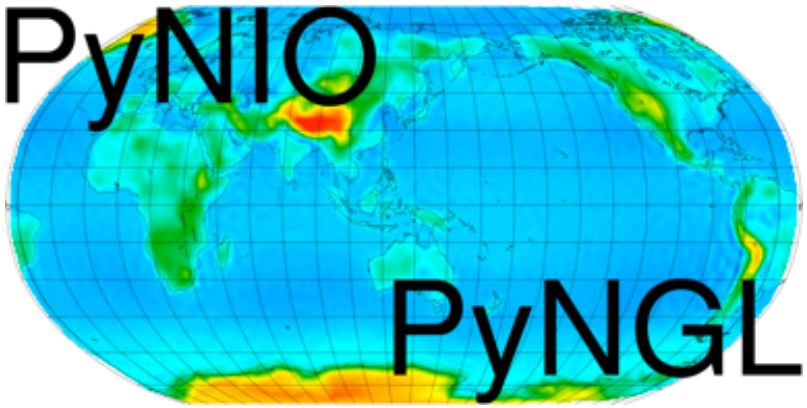
http://pyaos.johnny-lin.com/

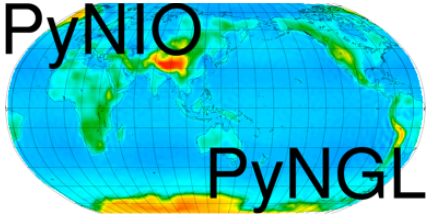
PyNIO and PyNGL

Python modules based on a subset of NCL's capabilities

PyNIO - Python interface to **N**CL's file **I/O** library

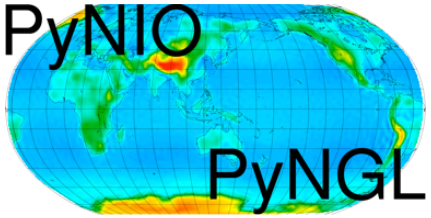
PyNGL - Python interface to **N**CL's **g**raphics **l**ibrary



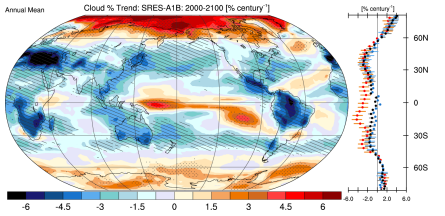


PyNIO

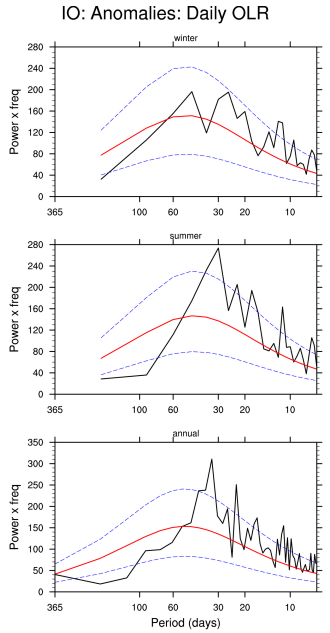
- Provides unified interface to these formats:
  - NetCDF-3 / NetCDF-4
  - HDF4 / HDF5
  - HDF-EOS2 / HDF-EOS5
  - GRIB1 / GRIB2
  - Shapefiles

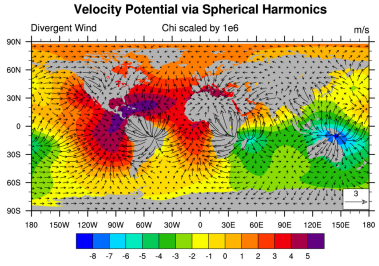


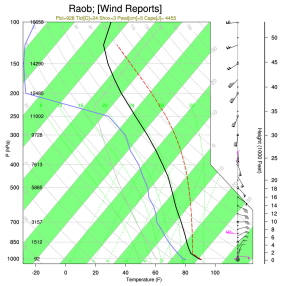
PyNGL

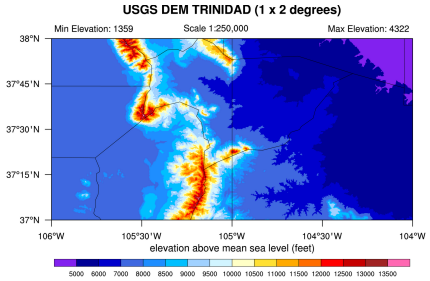


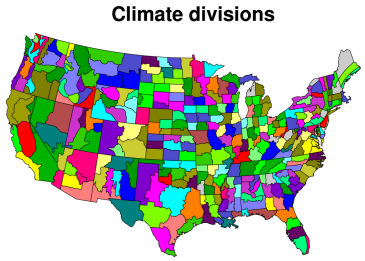
- Python 2D plotting library (based on NCL)
- Highly customizable
- Handles data on regular and irregular grids
- Specialized scripts for MJO CLIVAR, skewT, wind roses, bar charts, overlays





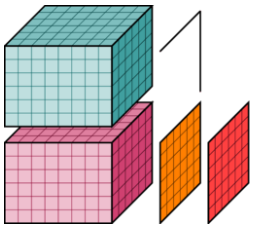






matplotlib

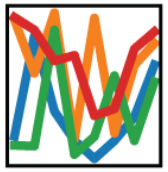
- Python 2D plotting library
- Basemap provides mapping utility
- Generates graphics similar to NCL
- More mainstream usage than NCL



xarray

pandas

$$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$$



pandas provides high-performance, easy-to-use data structures and data analysis.

xarray brings the labeled data power of pandas to the physical sciences, by providing N-dimensional variants of the core pandas data

Good starting point. . .

PyAOS website

Python resources for atmospheric and oceanic sciences

http://pyaos.johnny-lin.com/

All this software!

What should I use?!?

Depends. . .

- Tools can be domain specific
- How motivated you are – do you like to write code?
- Do you need quick look or heavy lifting?
- The one your advisor/boss told you to

Help available from NCAR and beyond

NCL email list

http://mailman.ucar.edu/mailman/listinfo/ncl-talk

PyAOS email list

http://lists.johnny-lin.com/listinfo.cgi/pyaos-johnny-lin.com

For the hands-on lab

- Focus will be downloading data from the RDA
- Use tools like NCL, panoply, Python (PyNIO, xarray, netcdf4python) to examine files
- Take advantage of the expertise in the room!

http://www.ncl.ucar.edu/Training/Tutorials/SEA2016/