

What are you doing?

Effective Computing

Intro

- This class is designed to help with the more complex computing problems that will arise in your research. By “complex” I don’t mean a difficult or large calculation, but instead all the other problems that arise around the edges of your work. These are things like: working across your laptop and a remote machine or designing code projects that don’t break every time you change something.
- Assignment, due at the next class: write answers to the questions on the following slides and send them to me. This will give me a better understanding of what topics we should focus on in order to solve your problems.
- For each question, I fill in my own answers as an example.

1. Describe what computers you use

- I mainly use a MacBook Pro (2017, OS X Catalina, RAM: 16 GB, Storage: 1 TB Flash which is 60% full).
- I also have 3 linux machines on campus (fjord, boiler and perigee) that I use for analysis and storage. All told, these have about 450 TB of RAID storage and dozens of cores. These have their own room in OSB and are maintained by my system administrator David Darr. I pay about half of his annual salary from grants, sharing his services with LuAnne Thompson and other oceanographers.
- I have part ownership of a linux cluster called gaggle, with about 200 cores connected by InfiniBand to do parallel calculations.
- I own 260 cores on the UW supercomputer platform called mox (part of hyak), also for large parallel jobs.

2. What software do you use most for research? (everything except the programming languages – that is the next page)

- Sophos virus protection: first thing to install on any laptop
- TextMate for editing text files on my mac, vi if I am working remotely.
- GitHub Desktop for keeping code organized across multiple machines, and for sharing
- Office 365 for writing, with EndNote X9 as a citation manager
- Dreamweaver (part of Adobe Creative Suite) for webpages
- Evernote to make notes – a log of what you did on any given day for a project can be valuable when trying to figure out where a bug came from
- Transmit for file transfer (also scp from the command line)
- ROMS for ocean modeling

3. What programming language do you use when writing your own code?

- For the past few years I write most new code in python (python 3, Anaconda distribution). I work from a terminal window, and with a text editor (TextMate), instead of using an IDE like Spyder. “IDE” stands for Integrated Development Environment – like what you see when you start MATLAB.
- Within python my most common modules are: matplotlib, numpy, pandas, datetime, sys, os, pickle, argparse, and requests
- For decades I did everything in MATLAB, but when I started to develop more complex systems of code for a daily forecast model python was the logical choice.

4. What do you wish you could do?

- I wish I could have my programs and data float magically in front of me so that I could manipulate them with my hands like in all the science fiction shows.
- I wish I never, ever had to deal with getting data from one format to another.
- I wish I knew how to program in Javascript and could make interactive 3D data exploration tools for people to use on my website.

5. Describe a coding problem you had recently, and how you tried to fix it.

- I was trying to resample a pandas DataFrame, making monthly averages from daily values for all the columns, but when I did this a few of the columns would mysteriously disappear.
- First, I try Googling the problem. Many problems are solved in Stack Overflow.
- When I am really stuck, my strategy is to make the smallest possible piece of test code that reproduces the error. The faster you can iterate the better. Then I add print statements. Then I examine suspicious variables in the iPython environment (typing varname? gives a lot of info).
- In this case I found that the resample method drops columns if for some reason pandas thinks they have non-numeric values. To fix this you have to use the `.to_numeric()` method on each column.
- I add comment lines around the fix, so that I can remember what I did later.
- If I solve a problem and it seems like it will come up again, I add an entry about it in my Evernote page python/pandas.