Tips and Tricks for Computing Work Flows

Benjamin Hicks (bhicks@princeton.edu)
Research Computing OIT

Why worry about workflow?

- Organized code and computer usage from the start
- Testability
- Saving yourself from problems later
- Getting a sense of problem scale

- Learn to use your own personal computers more efficiently (esp. Mac and Linux based)

- More easily juggle conflicting package requirements (i.e. Python virtual environments)

- Be more familiar with using the university compute clusters

Computing Resources at Princeton

- PICSciE
- Research Computing
- Lewis Library
- A point of contact: cses@princeton.edu

Services

- Weekly Help Sessions (M 10-11, Th 2-3, Lewis 347)
- Computing Clusters (Nobel and Adroit, plus other clusters with a sponsoring PI)
- cses@princeton.edu for reference with computational research questions

The Command Line

https://github.com/bwhicks/bash-notes

- .bash_profile?
- .profile?
- .bashrc?
- What’s the difference?

- Nano
- Vim
- Emacs
- On workstation: Sublime, Atom, Eclipse

Stupid Bash Tricks

Commands to Know

- man
- cp
- ls
- mv
- find & grep

Organizing Data and Code

(Why you should use some kind of CVS)

GIT

No really.

- Start early.
- Private Repo? (https://www.princeton.edu/researchcomputing/services/github-form-new/)
- Way to back up code for documenting results

Code and Documentation

- R packages for example - Weave and KnitR
- Python (Sphinx[uses ReStructured Text])
- Pandoc

Virtual Environments

Python

- virtualenv (https://virtualenv.pypa.io/en/stable/)
- conda (https://www.continuum.io/downloads) [also R and Scala!]
- Pros/Cons?

Ruby

- rbenv (https://github.com/rbenv/rbenv)

- RVM (Ruby Version Manager)

Meta Considerations

- Is this code that will require HPC resources?
- Can I make it parallel (implicit like algebraic libraries vs. parallel loops)?
- What sort of storage do I need?
- Does my PI sponsor accounts vs. Nobel/Adroit?

Working on the Cluster

- ssh (terminal is great, PuTTY on Win)
- scp
- sbatch and squeue (part of SLURM)

Survival Commands

- module load anaconda (or anaconda3)
- module avail
- rh/devtoolset/4 - updated gcc
- intel, openmpi, mkl math libraries