WORKFLOW

MATAN GAVISH
STANFORD STATISTICS
5/2013
What is workflow?

- How we accomplish daily chores on computer
- Computer routine are different
- Tells apart professionals
- Tells apart professional groups
WHO CARES?

YOU.

- Be more productive
- Be less frustrated
- Produce better quality work
- Do impossible things before breakfast
THINGS WE DO

COMPUTATIONAL SCIENTIST’S WORKFLOW

- Write code
- Execute code
- Keep track
- Write papers & slides

ALONE. IN GROUP. IN COMMUNITY.
WORKFLOW ELEMENTS

BASICS

- Workstation
- Directory structure
- Shell
- Text editor
- Knowledge base
WORKFLOW ELEMENTS

SOFTWARE DEVELOPMENT

- Source control and collaboration
- Codebase, packages, dependencies
- Style guide and lint
- Tests and coverage
- Docs generator
- Code review
WORKFLOW ELEMENTS

PRODUCTION

- Production environment
- Job monitoring
- Result harvest
Keeping track

- Research Journal
- Lab Journal
- Result archive
WORKFLOW ELEMENTS

WRITING PAPERS & SLIDES

- Typesetting system
- Source control and collaboration
- Citation manager
To: advisor
From: Student
Re: Paper draft

attached draft_2_final_4_submitted_student_changes12.tex and draft_references_for_final_4_15.bib
MY WORKFLOW

BASICS

- **Workstation:** unix only, Mac ($$$), Ubuntu ($)
- **Directory structure:** ~/r/project-name/talks/asilomar
- **Shell:** ZShell + ohmyzsh goodness, .zshrc, .ssh/config
- **Text editor:** vim (filetypes{colors,templates,indentation, macros}, pathogen, fugitive, completion, keybinds) .vimrc
- **Knowledge base:** wiki
Source control and collaboration: Git! git! git! (branches, remotes, tags, submodules, hooks, github)

Codebase, packages, dependencies: Git+Phabricator, pip+virtualenv (python)

Style guide and lint: Google’s style guide, pylint, .pylintrc, mlint, R “lint”

Tests and coverage: nose (python)

Docs generator: sphinx-doc

Code review: Phabricator
Production Environment:
- qsub, starcluster, MRJob

Job Monitoring:
- qsub, MRJob

Result Harvest:
- rsync, scp, dropbox, MySQL, VCR
WORKFLOW ELEMENTS

KEEPING TRACK

- **Research Journal**: black notebook, tex notes in project repo
- **Lab Journal**: iPython Notebook, VCR
- **Important result archive**: VCR
WORKFLOW ELEMENTS

WRITING PAPERS & SLIDES

- **Typesetting system:** vim-latex, soft links (macros, bibtex, graphics), Skim (back+forward search), beamer

- **Source control and collaboration:** Git, github, Meld(diff)

- **Citation manager:** Mendeley + git + soft links
Ultimate Workflow?

- Effortless to collaborate **within group**
- Easy to collaborate **outside group**
- Easy to learn
- Easy to maintain
- Effortless to set up
- Portable across computers
Tutorial + wiki for each element

Regularly maintained

One or few installers
GET A WORKFLOW

START YOURS! (1 WEEK)
OPENSOURCE IT!

- http://software-carpentry.org
- http://verifiable-research.org
- https://github.com
- http://vimcasts.org
- http://www/~gavish/workflow.html