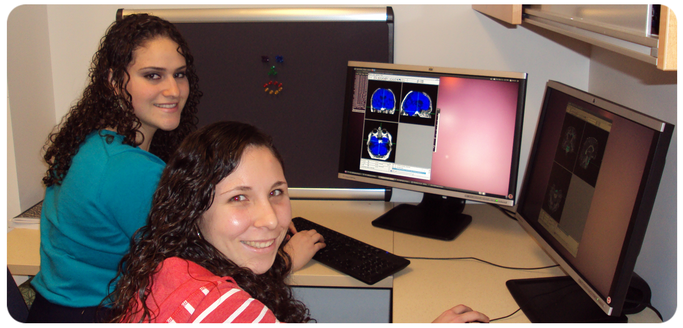
Using GNU Make to Teach Undergraduates Neuroimaging Workflow

Tara Madhyastha[1,3], Zoé Mestre[3], Elliot Collins[3], Thomas Grabowski[1,2,3]

[1]Department of Radiology, [2]Department of Neurology, [3]Integrated Brain Imaging Center at the University of Washington

Motivation

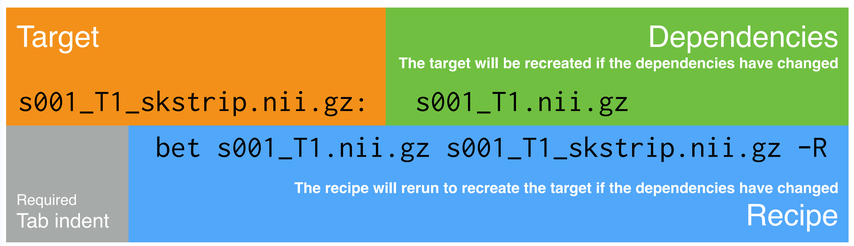
- Study of neuroscience using neuroimaging attracts undergraduates from diverse disciplines
- Goal is to teach neuroimaging workflows and their potential use in quantifying brain anatomy and function
- Most students lack familiarity with Linux, bash shell scripting, python, and parallelism



Zoé Mestre and former student research assistant appreciating the time saved using make to check images with fslview

Approach

- Design workflows using GNU Make, a freeware utility that allows one to specify how to create files from other files if they are out of date
- General form of a makefile:



Example 1. *A very simple makefile.*

```
T1files=$(wildcard s???_T1.nii.gz)
T1skullstrip=$(T1files:%_T1.nii.gz=%_T1_skstrip.nii.gz)


all: $(T1skullstrip)


%_T1_skstrip.nii.gz:   %_T1.nii.gz
     bet $< $@ -R
```

%	Matches a pattern
$@	Is the target
$<	Is the first dependency
$(variable)	Is a make variable

Example 2. *A more realistic example. This makefile uses wildcards to obtain the subject T1 images that the skull strips depends on, and pattern substitution to avoid having to type out all the subjects. Writing this out without pattern substitution is a useful exercise.*

Rationale

- Students can run complex pipelines using a common interface (e.g., “make freesurfer”) and see what commands are run, going from a high level of abstraction to a lower one
- Places emphasis on expressing logical workflow chains, less so on programming constructs
- Easy to modify existing workflows for specific purposes
- Naturally expresses dependencies that are essential for parallelization, making it possible for students to quickly write large scale workflows that will run on a multicore workstation or cluster
- Error recovery is automatic; the workflow will pick up where it left off when stopped or recreate what has been deleted
- Enforces good file naming and directory organization conventions

```
STD_BRAIN=/usr/share/fsl/5.0/data/standard/
MNI152_T1_2mm_brain.nii.gz

# epi to T1 registration
xfm_dir/epi2struct.mat: T1_skstrip.nii.gz MPRAGE.nii.gz
tsoc.nii.gz
     mkdir -p xfm_dir ;\
     epi_reg --epi=tsoc.nii.gz  –t1=MPRAGE.nii.gz
--t1brain=T1_skstrip.nii.gz --out=xfm_dir/epi2struct

# T1 to std registration
xfm_dir/struct2std.mat: T1_skstrip.nii.gz
     mkdir -p xfm_dir ;\
     flirt -in $< -ref $(STD_BRAIN) -omat $@
-out xfm_dir/struct2std.nii.gz

# std to T1 matrix transform
xfm_dir/std2struct.mat: xfm_dir/struct2std.mat
     mkdir -p xfm_dir ;\
     convert_xfm -omat $@ -inverse $<

# structural to epi registration
xfm_dir/struct2epi.mat: xfm_dir/epi2struct.mat
     mkdir -p xfm_dir ;\
     convert_xfm -omat $@ -inverse $<

# standard to epi
xfm_dir/std2epi.mat: xfm_dir/struct2epi.mat
xfm_dir/std2struct.mat
```

Example 3. *Registrations. Note consistent naming conventions, use of multiple registration tools*

Process of Writing and Running a Makefile Workflow

- Conceptualize targets and dependencies, normally within a subject directory
- Write and test rules on a single subject, in a text file called `Makefile`
- Create targets, one by one, fixing problems as you go
  `$ make target`
- Run job in parallel on multicore machine
  `$ make -j 8 target`
- Or on cluster
  `$ qmake -cwd -V -- -j 48 target`

Teaching Students Neuroimaging Workflow with Make

Basic Tasks

- Run preexisting workflows, developed for running FreeSurfer, dti analysis, tractography, white matter hyperintensity quantification, feat, melodic denoising
- Use make for interactive checking of images (e.g., editing white matter control points in FreeSurfer or melodic denoising)
- Use make in parallel for large jobs, some percent of which fail
- Run through workflow to redo problematic steps by hand
- Document problems and steps for solution

Intermediate Tasks

- Extend existing workflow with some simple target (document as appropriate)
- Port existing workflow to new dataset or timepoint

Advanced Tasks

- Pick a task to automate, script it using some combination of programming and makefiles
- Document new workflow

Links

- We provide a manual for teaching and using make for basic neuroimaging workflows:

  Using GNU Make for Neuroimaging Workflow:
  ibic.washington.edu/wiki

  The GNU Make manual is excellent:
  www.gnu.org/software/make/manual/

Acknowledgements

National Institutes of Health 1RC4NS073008-01

IBIC
INTEGRATED BRAIN IMAGING CENTER