DataFlow Using sysCS part I

or

super yet simple Computing for Science

James Ackman

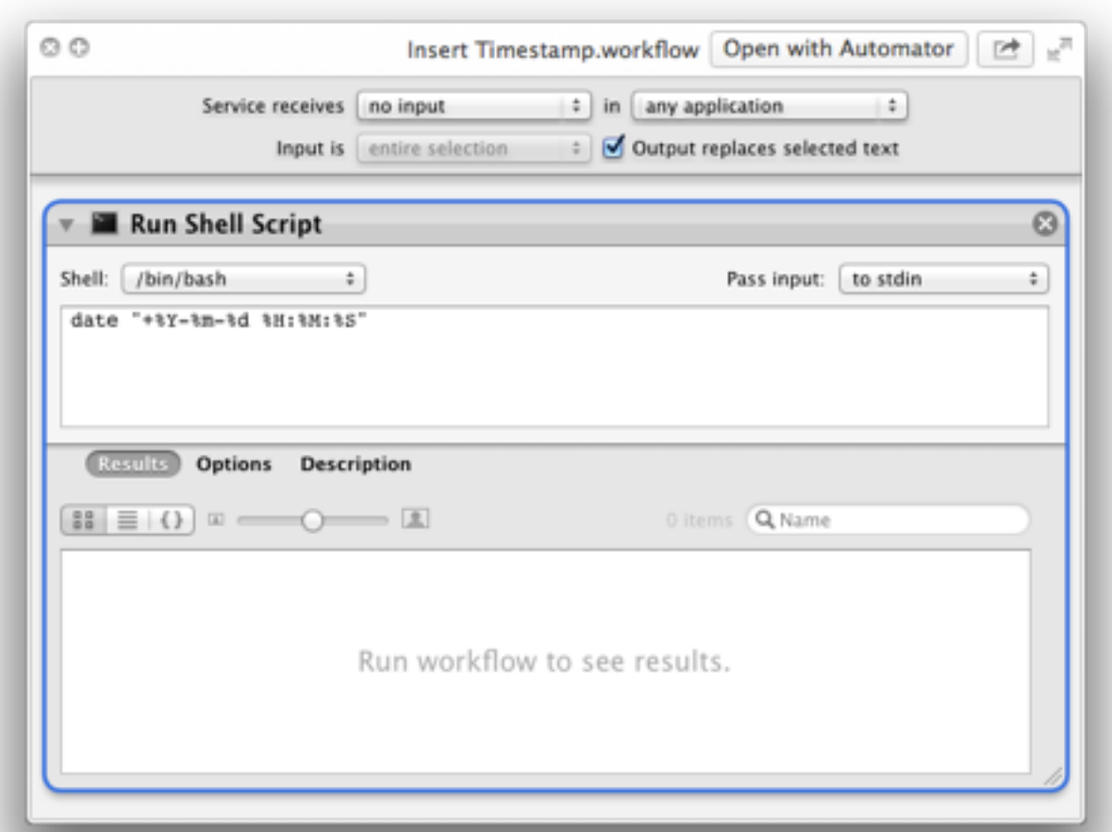
lab mtg

⇧^⌥⌘D = 2013-08-21 10:10:07

stupid computer tricks

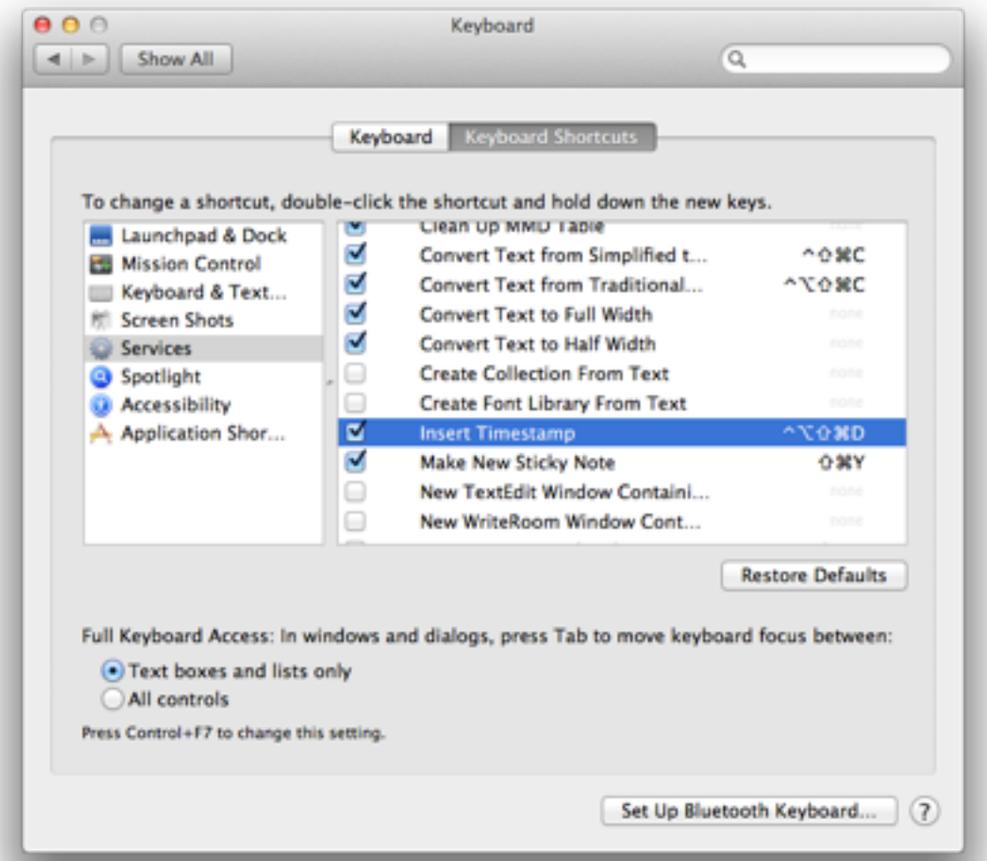
Automation!

```
date "+%Y-%m-%d %H:%M:%S"
```



script

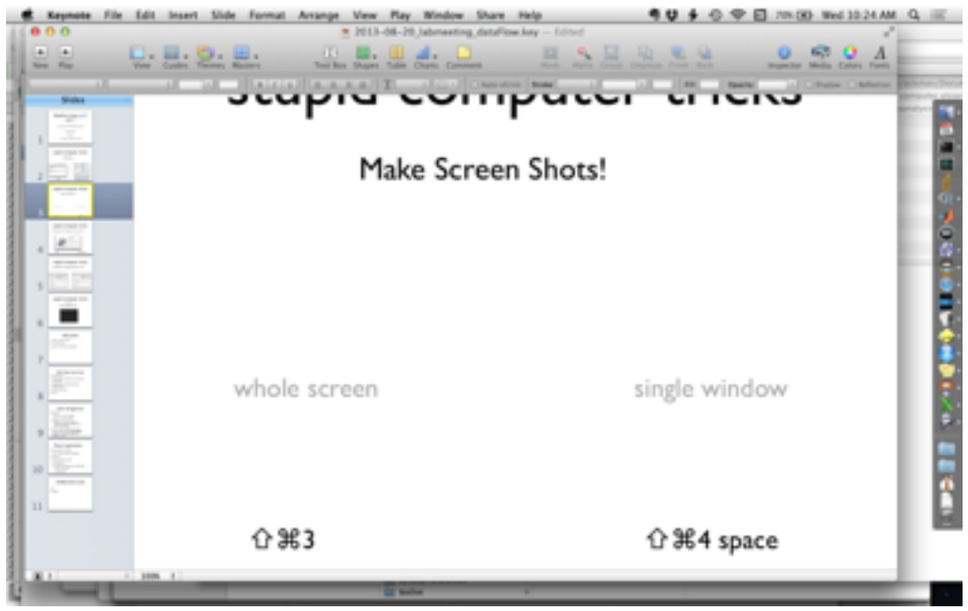
+



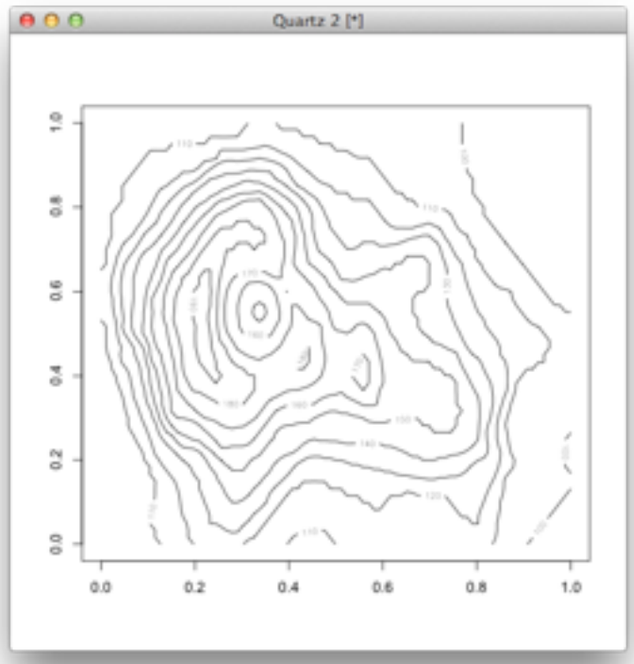
keyboard shortcut

stupid computer tricks

Make Screen Shots!



whole screen



single window

⇧⌘3

⇧⌘4 space

stupid computer tricks

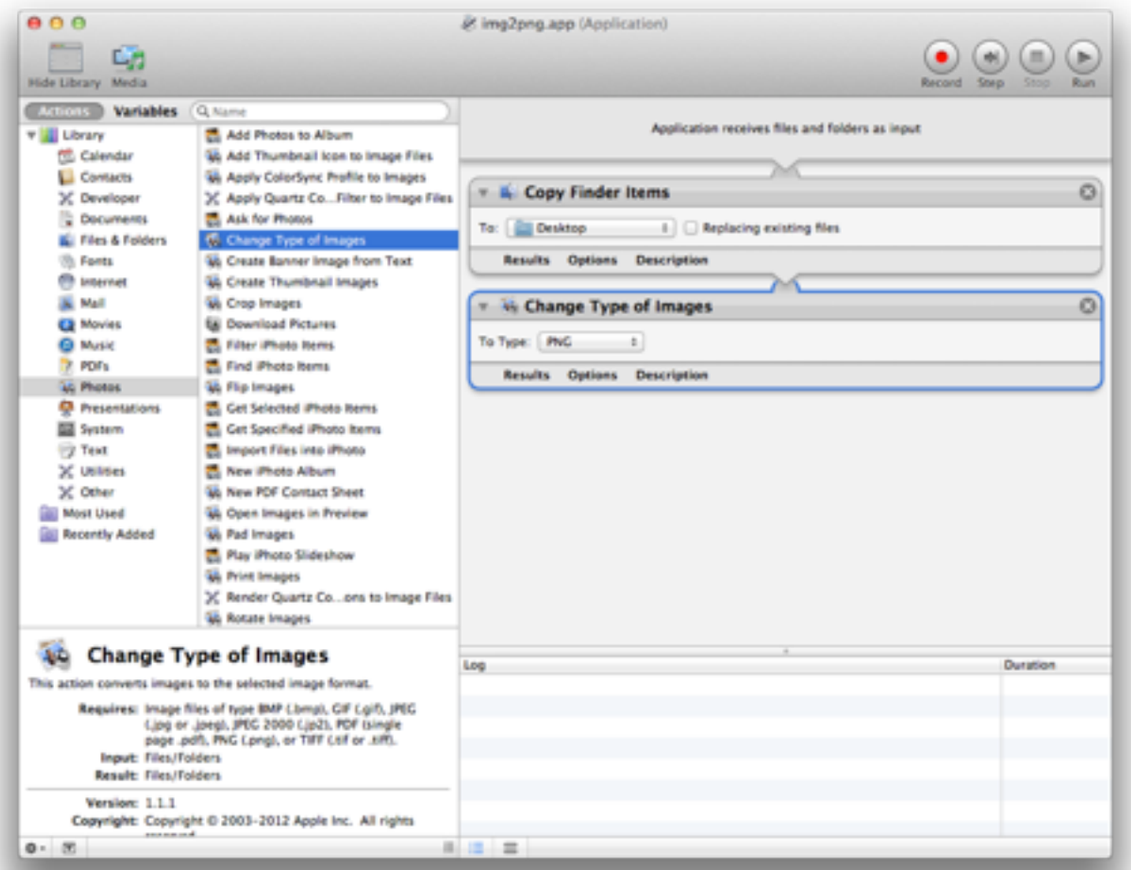
Drop App for changing filename and move

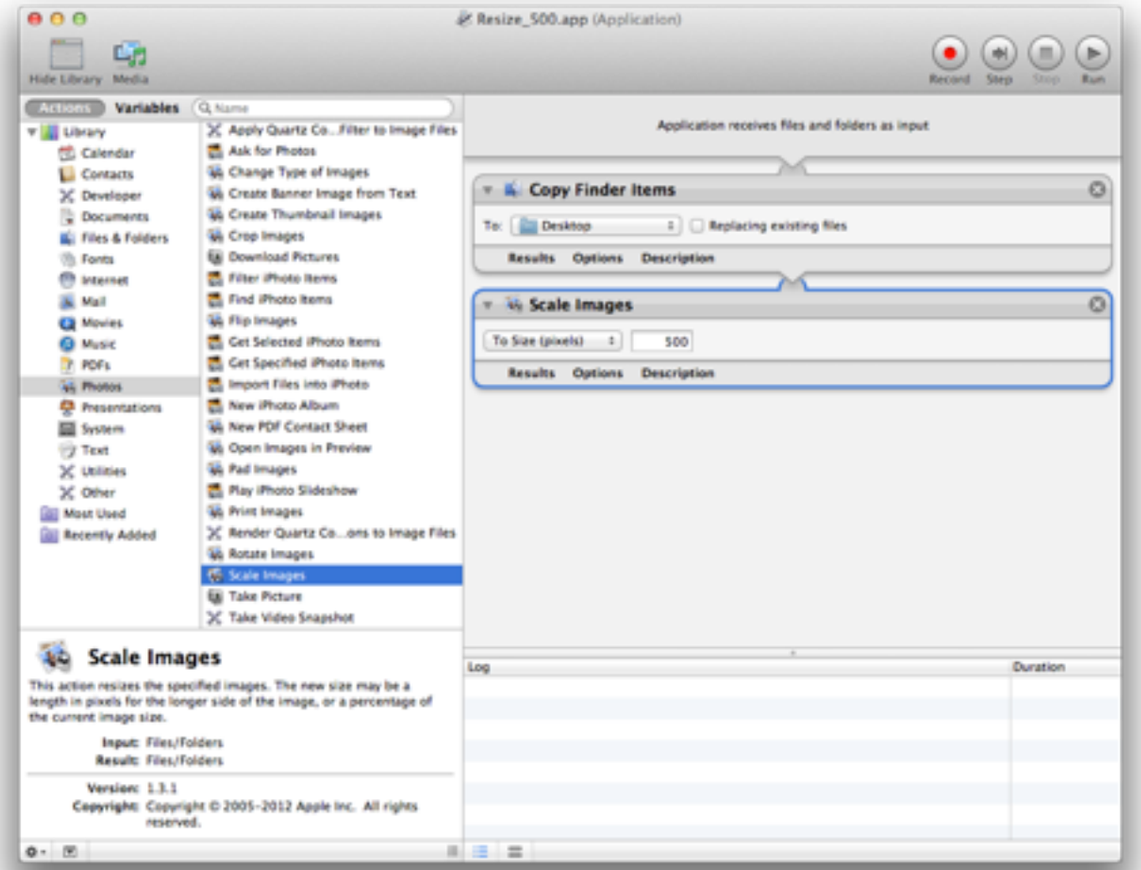
⇧⌘3

⇧⌘4 space

stupid computer tricks

Drop Apps for changing image type or size

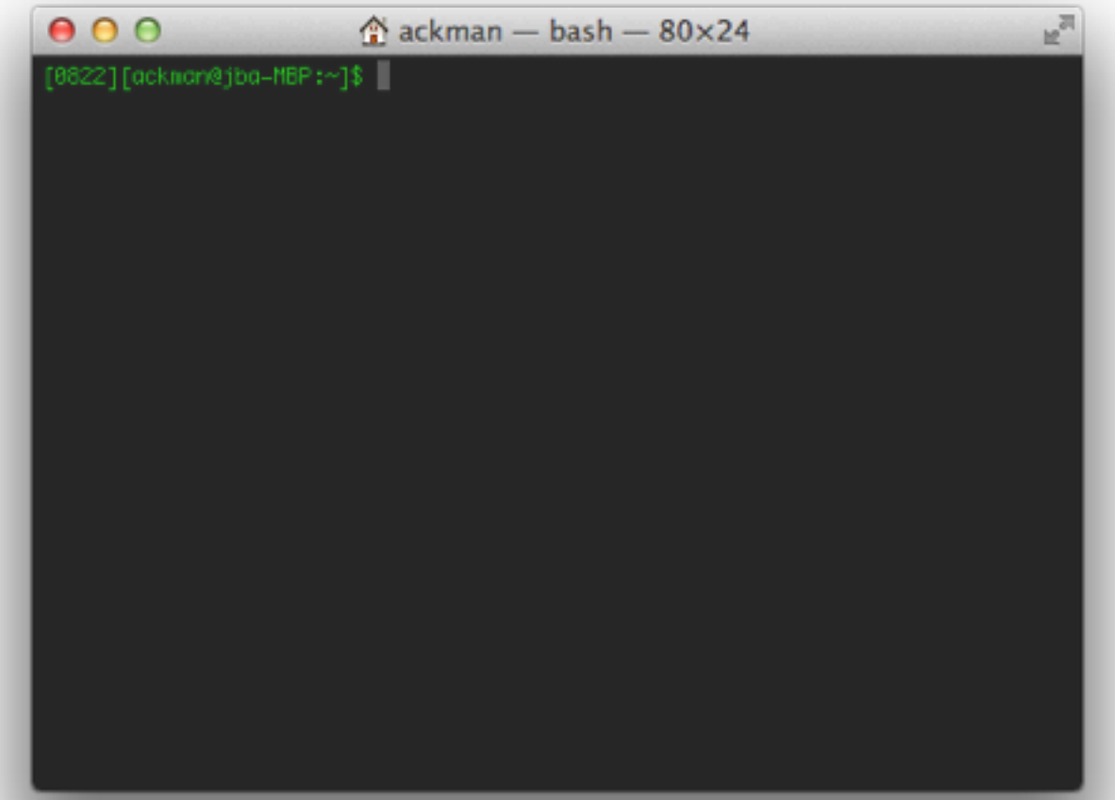




stupid computer tricks

Terminal?

What is this, the 1970s?



data jujitsu

- matlab – generate data
- R – explore data
- python – ditto, plus anything else

learning resources

- CodeAcademy
- Coursera (Intro to Data Science, R, python, matlab, SQL)
- Yale StatLab (R, matlab, python)
- Mathworks (matlab)
- Books!

data management

- Data sets
  - plain text – human readable
  - binary – machine readable
    - RDBMS (on hard disk: mySQL; in memory: R dataframe, Matlab struct, python lists & tuples
- tools – terminal consoles, rsync, grep; python, Microsoft Excel, **Text editors** (notepad++ on Windows, bbedit or TextMate on Mac)

Report generation

- Reproducibility in Science
  - 21st century scientific notebooks...
- Markdown
  - plain text (future proof)
  - multipurpose
  - can generate HTML5, pdfs, rtf, ODF (then to .doc), many others
    - Rmarkdown

Collaborative tools

- git
- GitHub