Efficient Computing for Social Scientists

Johannes Karreth

February 22, 2013

Why do you need a good workflow?

- Collaboration
- Save time
- Replication
- Changes
- Implement updates
- Reproduce your own work
- Expand work to other projects
- Learn from my many mistakes
  - Time lost
  - Data errors

Elements of a good workflow (today's outline)

- Backups
- File structure
- Bibliography management
- Note taking
- Mind mapping
- Word processing
- Presentations
- Text editors
- Statistics
- Qualitative analysis

Backups

- Time machine
- Carbonite
- Dropbox
- HDs on site / off site

File structure

- My example:
  - One folder for projects (papers, diss, etc.)
  - One folder for data (structured by topic & name)
  - One folder for articles & e-books (w/ master bib)
- Project-specific folder master structure

Johannes' project-specific file structure

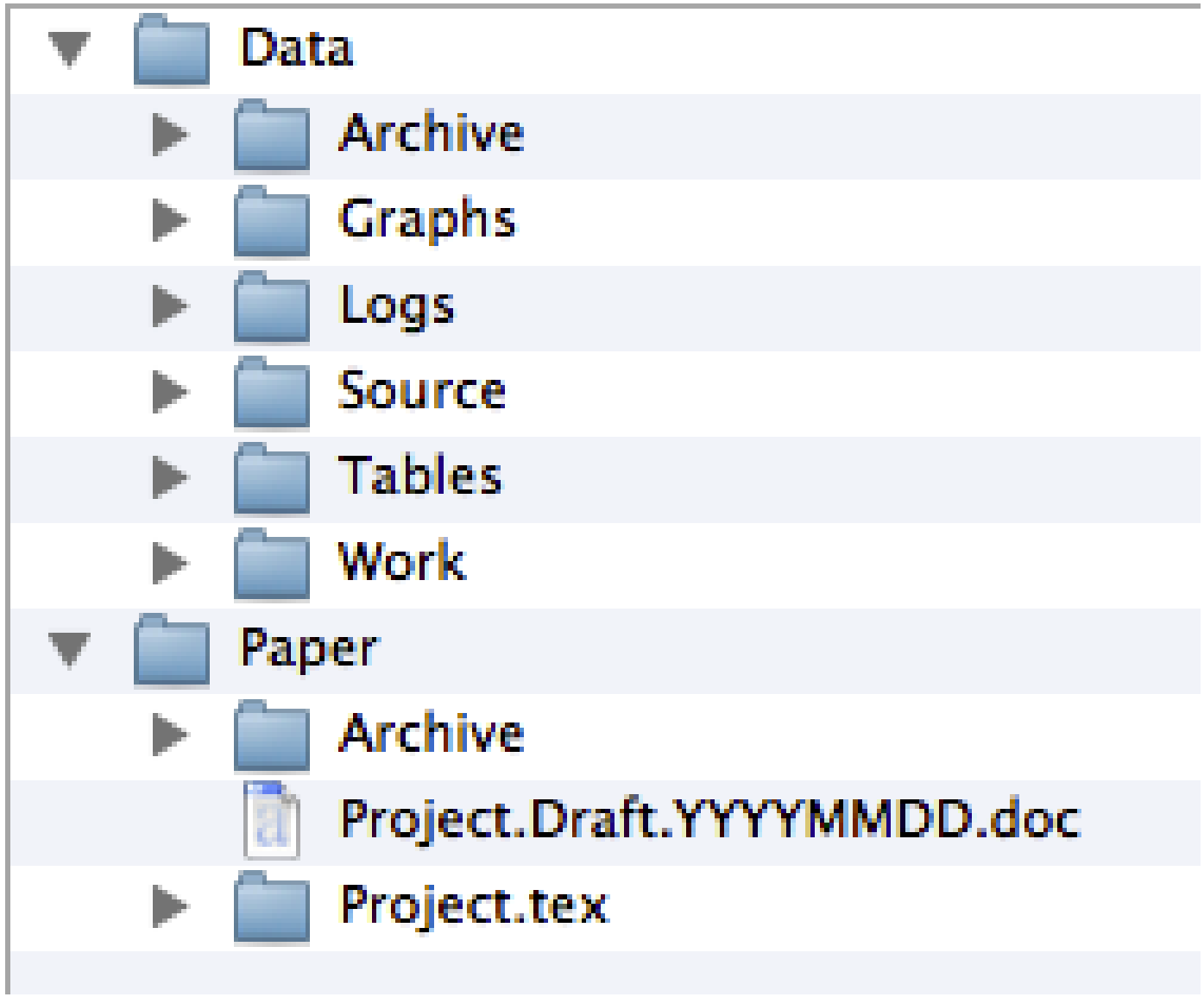

Figure : My folder structure

File structure

- My example:
  - One folder for projects
  - One folder for data
  - One folder for articles (w/ master bib)
- Project-specific folder structure
- Other examples?

Bibliography management

- Endnote (free at CU?)
- Papers (like iTunes, ~ $50)
- Bibdesk (free)
- Zotero (free)
- Integration with word processing (Word & LaTeX)
- Save articles in one master bibliography
- Use software to save notes where you can find them easily (for comps!!)

Note taking

- Simpler formatting is better
- You should have a consolidated place for notes, rather than files flying around
- Searchability & tagging are very important
  - Evernote works well for many, and also allows sharing & collaboration, also across platforms & devices
  - Simplenote
  - Other examples?

Mindmapping (hello theorists!)

- White/blackboards
- FreeMind (thanks to Matt Heller!)
- Mac: OmniGraffle (also for diagrams)

Word processing

- Word, Open Office, Pages: use headers (why?), what else?
- LaTeX (`http://spot.colorado.edu/~joka5204/latex.html`)

Presentations

- LaTeX Beamer (previous workshop)
- Cool option: Pandoc & MultiMarkdown
  - to PDF
  - to HTML

Pandoc: Source code for this presentation

workflow.pandoc.txt

```
% Efficient Computing for Social Scientists
% Department of Political Science, CU-Boulder
% February 22, 2013

# Why do you need a good workflow?

- Collaboration
- Save time
- Replication
- Changes
- Implement updates
- Reproduce your own work
- Expand work to other projects
- Learn from my many mistakes
    + Time lost
    + Data errors

# Elements of a good workflow (today's outline)

- Backups
```

Advantages of non-PPT

- Easy transfer from paper manuscript to slides
- You can always recover content

Text editors

- (In my view) necessary for statistical software and others...
- Syntax highlighting
- Balancing code elements (no more un-matched brackets)
- Windows: WinEdt, Notepad++
- Mac: Textmate(2), Textwrangler, Fraise, Emacs/ESS

Statistics software

- File structure. Separate:
  - Source data
  - Working (recoded) data
  - Recoding commands
  - Analysis commands
- MUST use script/do files (and log) files
- Nested script files
  - E.g., one master file calls recoding & analysis files
- Don't overwrite datasets unless you're certain that's what you want
- Useful version numbering
  - I use an archive for datasets, named by date (not ideal)
- *Look* at your data and summarize & plot it
  - My interpolation error: IGO memberships < 0
  - I didn't see it until someone else pointed this out

Statistics software: Resources

- Scott Long's book: *The Workflow of Data Analysis Using Stata*
- R equivalents?
  - `http://stackoverflow.com/questions/1429907/workflow-for-statistical-analysis-and-report-writing/`
  - `http://robjhyndman.com/researchtips/workflow-in-r/`
  - `https://github.com/johnmyleswhite/ProjectTemplate`

Qualitative analysis

- Evernote for storing notes, audio, and external files
- More complex software for text analysis
- QDAP/CAT (open source)
- Nvivo (not open source)
- WordFish (in R)
- RTextTools (also in R)

The #1 question you should ask yourself:

If you had to recreate all contents of a project, how long would it take you?

How clear and straightforward is this process?

Your life depends on it. . .

These slides will be posted at
`http://spot.colorado.edu/~joka5204/workflow.html`