Causal Inference

Mixtape Session

Roadmap

- Hidden curriculum
  - Background
  - Empirical workflow
  - Hierarchical folder structure
  - Naming conventions
  - Version control
  - Soft skills

Folder structure and directories

- After my coding error with my job market paper, I researched how problems like these tended to happen
  1. **hierarchical folder structure**
  2. automation
  3. naming conventions
  4. version control
- Also see Gentzkow and Shapiro's 2014 "Code and Data for the Social Sciences: A Practitioner's Guide" `https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf`

REPORT

FILE NOT FOUND

A generation that grew up with Google is forcing professors to rethink their lesson plans

By Monica Chin | @mcsquared96 | Sep 22, 2021, 8:00am EDT

Illustrations by Micha Huigen

Catherine Garland, an astrophysicist, started seeing the problem in 2017. She was teaching an engineering course, and her students were using simulation software to model turbines for jet engines. She'd laid out the assignment clearly, but student after student was calling her over for help. They were all getting the same error message: The program couldn't find their files.

Garland thought it would be an easy fix. She asked each student where they'd saved their project. Could they be on the desktop? Perhaps in the shared drive? But over and over, she was met with confusion. "What are you talking about?" multiple students inquired. Not only did they not know where their files were saved — they didn't understand the question.

Gradually, Garland came to the same realization that many of her fellow educators have reached in the past four years: the concept of file folders and directories, essential to previous generations' understanding of computers, is gibberish to many modern students.

Helping your future self find what she needs

- Remember your future self is operating at the edge of her production possibilities frontier and has no more time left to screw around
- The typical applied micro project may have hundreds of files of various type and will take *years* just to finish not including time to publication
- So simply finding the files you need becomes more difficult if everything is stored in the same place
- Goal is to help her find what she needs quickly, accurately and without breaking anything else
- Hierarchical folder structures are the main way we do this

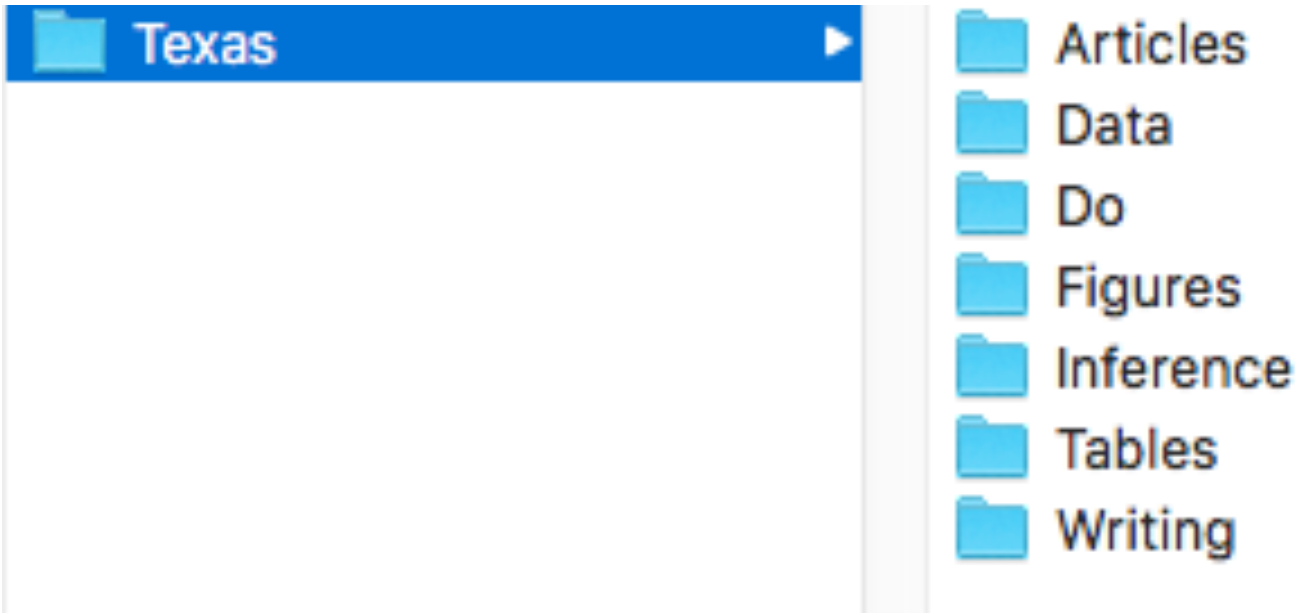
Finding files using folder structure

- Directory structure is a “hierarchical system of folders that modern computer operating systems use to arrange files” – Chin (2021)
- Directory structure connotes physical placement as in a file stored on a computer is *somewhere* on that computer in a specific and discrete location
- But it’s very easy in projects that takes years to accumulate so many files that you just give up and everything ends up in one directory
- Mental models have been changing (educators noticed it in 2017) as newer students tended to use more cloud-based storage, smartphone apps, and various forms of search, rather than hierarchical folders for organization
- Need to stick with hierarchical folder structure as much as possible

Directories

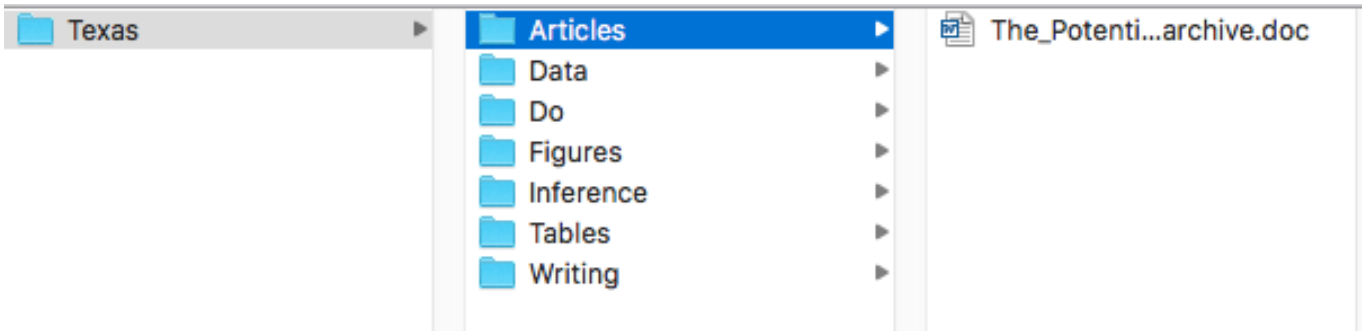
- But even well designed structure can be undermined by:
  1. Time: Multi-year projects lead to more files, and weird ones that don't quite fit in your pre-designed folders
  2. Files: More data you collect, the harder it becomes to store in logical places
  3. Coauthors: Because everyone tends to do things differently, shared dropbox with collaboration can create messes
- Hierarchical folder structures require constant vigilance because efficiency and accuracy both decrease with time, files and collaboration
- Consider the following relatively simple structure as a conceptual model only

Subdirectory organization



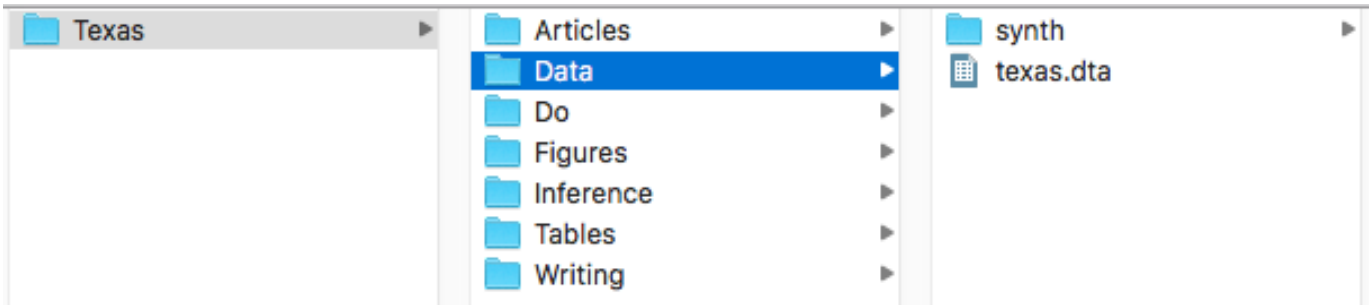
1) Name the project (“Texas”)

Subdirectory organization



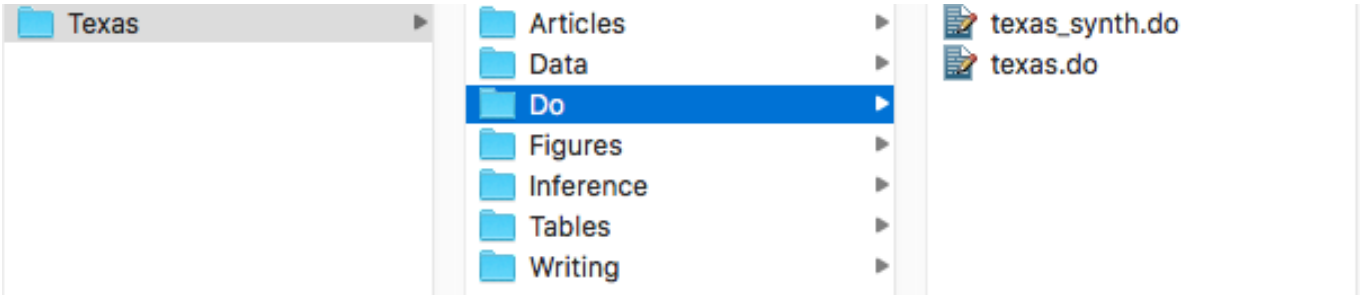
2) A subdirectory for all articles you cite in the paper

Subdirectory organization



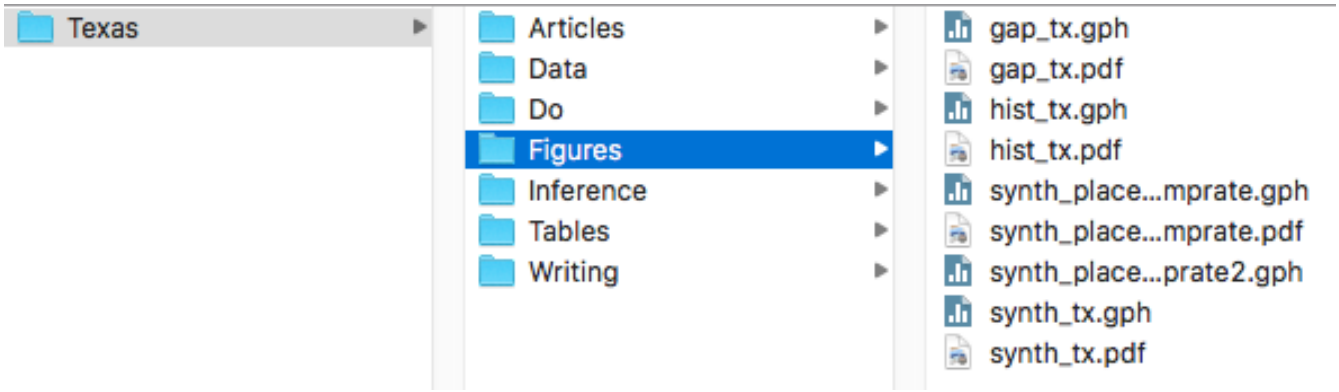
3) Data subdirectory containing all datasets

Subdirectory organization



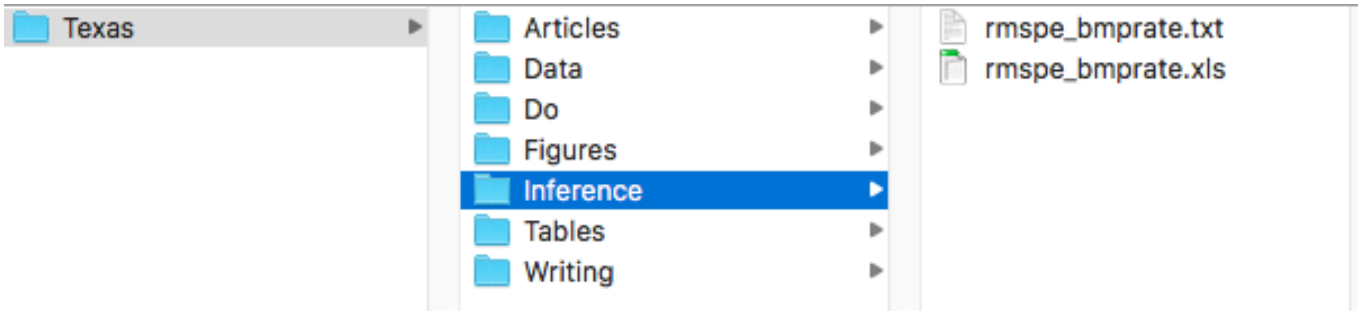
4) A subdirectory for all do files and log files

Subdirectory organization



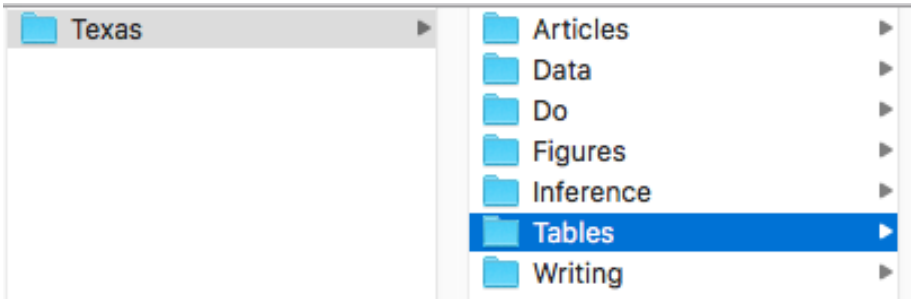
5) All figures produced by Stata or image files

Subdirectory organization



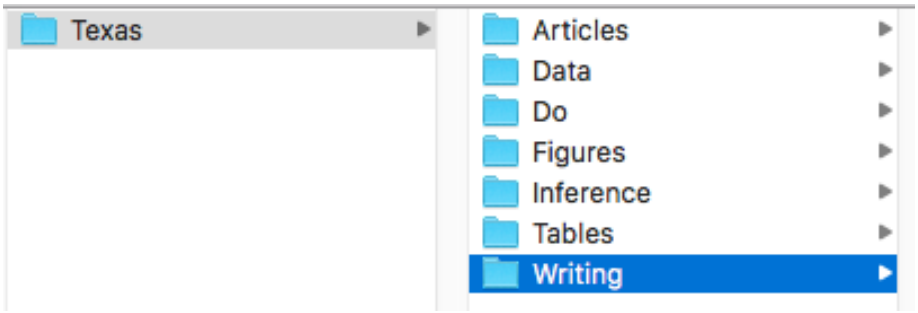
6) Project-specific heterogeneity (e.g., “Inference”, “Grants”, “Interview notes”, “Presentations”, “Misc”)

Subdirectory organization



7) All tables generated by Stata (e.g., .tex tables produced by -estout-)

Subdirectory organization



8) A subdirectory reserved only for writing