Reproducible Workflow: Coding Strategies and Software [75mins]

Introduction, Hands-on with Version Control (Github) and Dynamic Documents (RMarkdown)

Fernando Hoces de la Guardia
BITSS
-
Slides at https://goo.gl/aBQ3LR

Inter-American Development Bank Workshop, March 2018

File Management & Coding Suggestions [15 mins]

File Management [5 mins]

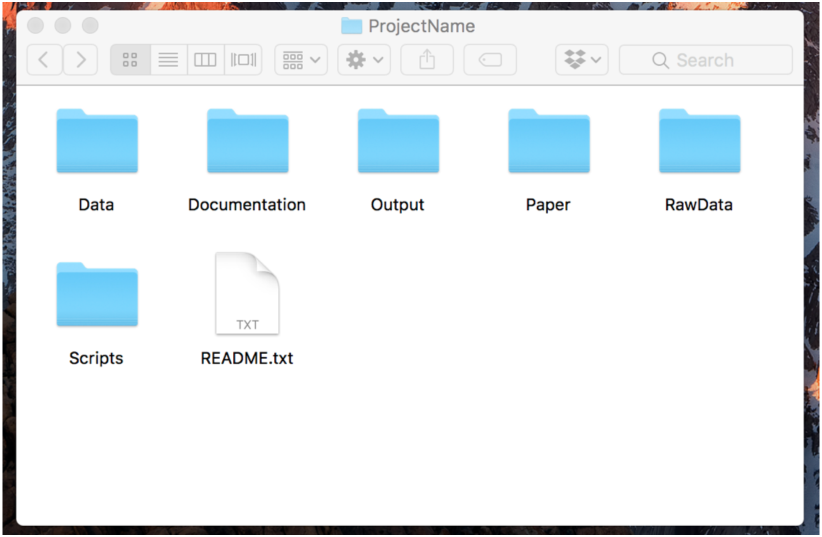


Figure 1:

Coding Suggestions [10 mins]

1. Make sure that your script files are self-contained. That is, don't write a program that only works if you run a group of other files previously in a specific order and then leave things hanging in a certain precarious way. You should be able to run intermediate steps of the workflow without disrupting things downstream.
2. Include tests in your code. This can alert you if output ever changes unexpectedly. For example, if merging intermediate datasets and dropping unmatched observations, you could write code to throw an error and alert you if the number of observations changes. (See the Stata example below.)
3. You can never comment your code too much. Comments should truly explain what the code is doing rather than merely transliterating: it is more useful to describe x<-1 with "initialize the population count to 1" than "set x equal to 1." Comments should also be checked to make sure they don't go out of date and convey inaccurate information.
4. Indent your code. (We are too smart to weigh in on the spaces

Version Control [30 mins]

Problem to avoid

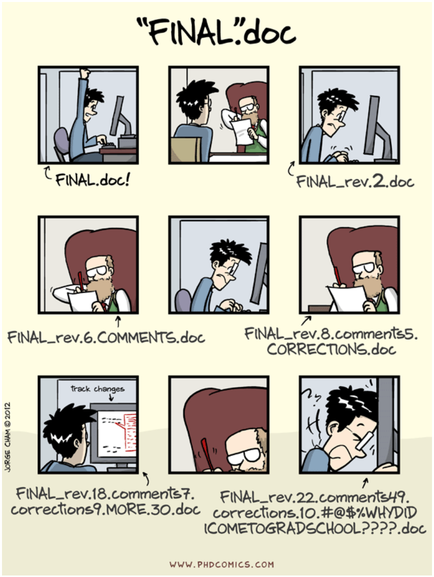

"FINAL".doc
FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.CORRECTIONS.doc
track changes
FINAL_rev.18.comments7.corrections9.MORE.30.doc
FINAL_rev.22.comments49.corrections.10.#@$%WHYDID ICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

Figure 2: http://www.phdcomics.com/comics/archive/phd101212s.gif

Managing expectations

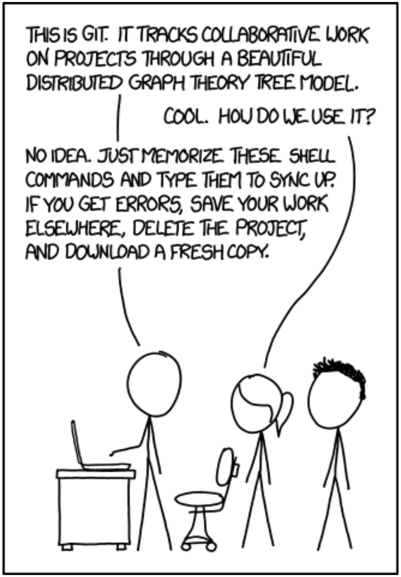
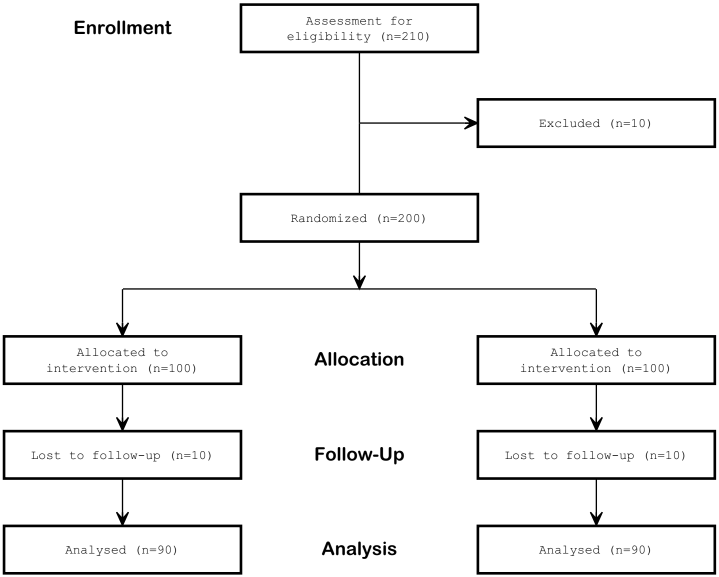

THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.
COOL. HOW DO WE USE IT?
NO IDEA. JUST MEMORIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.

Figure 3: Git xkcd comic

Dynamic Documents [30 mins]

CONSORT diagram of our little experiment



Balance of covariates

Estimated effect

The Stata version of all of the above: