class wrap up

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Sunday 24th April, 2022 03:44

outline

misc

misc2

course summary (repetitive boilerplate, ONLY if we have time)

outline

misc

misc2

course summary (repetitive boilerplate, ONLY if we have time)

ps6

- do not overcomplicate!
  - better to have simple clean code that does the job
  - than messy complex fancy code that is wrong
- if parts of code take long time to run, say $> 10min$
  - optimize it, take random sample, etc
- again, it always must start from the very raw data!
- easy to make mistake:
  - think about it AND cross check; correctness is important!

ps6

- explain what you are doing!
- interpret things!
  - eg when you run descriptive stats, *and* find sth interesting, put a comment and say what you have found in few words
  - (don't comment output of every command)

ps6

- google things!
- before writing the code check if someone already wrote it
- and build on others work! ie copy and adapt and improve
- eg googling 'world development indicators stata'
  - yields `http://www.stata-journal.com/sjpdf.html?articlenum=dm0045`
  - "Using the world development indicators database for statistical analysis in Stata"
  - exacly what 2 of you are doing

Stata wins for soc sci data (opinion)

- sometimes I think Stata may be a dinosaur v R or Python, but it's not!
- actually I do think Stata is the winner for the foreseeable future v Python, say next 20 years
- I love Python and I use it for vis, os, APIs, and machine learning (TODO)
- But most stuff, say 80% is just way easier, faster, and better in Stata
- R, I think, is worse than Python: convoluted, weird, and no fun
- sas, excell, and spss are just bad software for most work

get into flow with programming!

https://en.wikipedia.org/wiki/Flow_(psychology)

got songs about stata, regressions, dissertations?

- https://vimeo.com/852637
- https://www.youtube.com/watch?v=42odMdLwZKI
- https://www.youtube.com/watch?v=Ga6pOiiTn0Q

publish or perish

- coding is fun
- but don't lose the end goal from your sight
- yes, you are an amateur software engineer–it helps a lot
  - it's actually necessary these days!
- but you are a professional soc sci
- your end goal is a publication
- 'keep article pipeline full'
  - at any time i have at least 3 articles under review
- how about you? submit paper right now! [stata song]
  - start simple, even just some des stats...but keep on submitting papers

this is super important! remember this!!

- publishing (and maybe conferences) is *the only way* to get in touch with people exactly in your area
- there's just a handful of them,
  - almost never at you university, sometimes at a conference
  - usually at a journal where you submit; (if you pick the right one, almost always at a journal)
- this is *the only way* to take your work to next level!!
  - it does take time; start now; otherwise you may never make it
- or even ssrn or arxiv etc

outline

misc

misc2

course summary (repetitive boilerplate, ONLY if we have time)

future research

- you've probably realized that i am into
  - Stata, Python, data
- and always happy to discuss them
- let's stay in touch!

keep your inner stata alive!

- http://stata.com
- http://www.stata.com/bookstore/books-on-stata/
- http://www.stata.com/links/resources1.html
- http://www.michaelnormanmitchell.com
- http://www.princeton.edu/~otorres/
- http://www.stata.com/links/

what next?

- if didn't yet, print out box1 from `https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745` and hang on your fridge, over your desk, etc
- make use of your data management skills...
- use it or lose it
- apply for postdocs (eg subscribe to POLMETH listserv)
- collaborate with faculty
  - faculty need your data management skills
  - they know much less about data mgmt than you!
  - and in general, just make $!

make $

- industry data jobs usually require SAS, SQL, Python, Java
- but there are many Stata jobs, eg:
- http://www.job-search-engine.com/keyword/stata/
- and a ton of data science jobs:
  - http://web.ccsu.edu/datamining/jobs.htm
  - http://www.kforce.com
  - http://www.gadball.com
  - http://hotlinuxjobs.com/
  - http://www.icrunchdata.com/
  - http://www.cybercoders.com/

outline

misc

misc2

course summary (repetitive boilerplate, ONLY if we have time)

see again

- theory.pdf:the golden rule
- intro_to_course.pdf