getting started; 3pt of extCre and important to get going; may get behind on ps1

1. find datasets you would like to use, the sooner you find them, the better! very important! list them and in few sentences explain why these data: like a mini few sentence research abstract **this class is as much about stata as about data! ask q about data!**

2. And if you have not used stata, familiarize yourself with it: see above links, especially ucla website: do basics like descriptive stats and play with your own datasets

3. and push the ps0 dofile into github, will present next week

1. clarification: use as much data as possible; can use/reuse code i'm giving the class, just on your own data;

2. tip: push it to github early and email and i'll give you comments :)

---

general directions (always the same):

- i will show your code in class and possibly post some of your code or link to it–again, as per our core values–opensource, transparency, sharing; but if you'd like to keep your code private, that's fine–just let me know, and i will keep your code secret (no penalty, except that you may get little less feedback–usually if we discuss your code in the class, you will benefit from it!)

- you must submit all the code that was executed from the very beginning starting with the very raw data as per replication principle; unless data is too big to fit online, then just start with a comment, eg "to fit data online i had to take a random sample of 10perc"

- all ps are mostly cumulative–you can, and should, include much of previous code you've written for this class; can also use code you've written outside of this class (other classes, projects, etc)–but you have to clearly mark the code that has not been written for this class–otherwise, scholastic dishonesty!

- use your own dataset; again if you do not have a dataset, ask for help finding it

- because you are only submitting code, it must load data from Internet–just put your data onto your own website, wordpress, google drive, etc; (when you put data into any public space, try not to violate data copyrights… I haven't heard of anyone having problems with that, but be careful–for instance you may subset dataset to few vars and smaller sample using **sample**); and it is also easier to experiment on small datasets

- keep it simple! at the beginning of your dofile drop unnecessary vars; and even retain only certain, say most important, observations; keep it manageable; it is much easier to learn using simple data; can always complicate later!; much better to do it right using simple data than do it wrong using complex data!

- have nice structure in your file: sections, subsections, etc; may also have multiple files

- great idea to submit ps as early as possible–we will probably give you some comments; if not, email us and ask for comments!

- it is great to copy code from others; again, one of the rules for this class is 'be lazy': don't reinvent the wheel, whatever you are coding, it has already been done, google things often; but of course you cannot submit 100% code by someone's else.

- if you do something extra/fancy that is relevant and closely related to the assignment questions, it will be extra credit

- use coding rules that we've learned so far

- submit (only) the code into git repo; ps are due by the beginning of the next class unless indicated otherwise, eg "due in 2 weeks"; late ps are not accepted; NOTE: push to github early and send email to listerv with the link to your submission and ask for comments and ask any questions–the surest way to get the ps right!

- we are on the way to developing the final project with these ps: as we progress, your ps should start resembling a coherent and logical project where you use data management techniques to build new a dataset that can be used to answer interesting questions– say in few sentences (as a comment) why are you doing what you are doing–that is, answer the "so what question": "ok, you're gonna run all that code, and so what?" what's the goal of all that, why are you doing this? you need a compelling justification for what you are doing; typically: to develop a new dataset (that has not existed before) that can be used to answer some exciting questions: say what are those questions you want to answer; be brief, say couple sentences, and definitely not more than say 100 lines, typically 10-50 lines is enough; related: even at the beginning, already in ps1, say why you use data you are using, is it best, does it serve the purpose; also, feel free to ask me questions in comments

- be prepared do present your code in class (if time), just briefly, key points, couple minutes