

ps0 (upto 5pts of extra credit)

[version: Monday 4th September, 2023 18:07]

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. load your data into Python, produce basic descriptive stats (as always interpret)
-
1. clarification: use as much data as possible; can use/reuse code i'm giving the class, just on your own data;

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 less feedback—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); and it is also easier to experiment on small datasets
 - keep it simple! at the beginning of your notebook 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 listserv 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
 - 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 learned techniques to answer interesting questions—say in few sentences (probably at the beginning) 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 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: say why you use data you are using, is it best, does it serve the purpose; also, feel free to ask us questions in comments
 - be prepared do present your code in class (if time), just briefly, key points, couple minutes
-
- if you work in a group of 2 or 3 people make it 2x or 3x better, eg If ps asks for joining 2 datasets, and there are 3 of you, then just join 6, etc, just do 3x more or better
 - always have a brief description of a map, say few sentences or a paragraph or max few; also list any problems you've encountered and can ask questions!
 - always have exact links to all of the source data (so that i could create the map myself); note: exact links, eg do not say census.gov, but give full url to the data—i must be able to find it; sometimes there is no generic URL—then give steps: what I need to click to get the data!; or upload data say to your dropbox or google drive and provide link to there; so there must be all raw gis data (e.g. .shp .dbf and .shx and others if any), and any other data (e.g., .csv, .xls) that you used to produce a map; in rare circumstances if you use confidential/private data that cannot be shared, please contact me (and still do provide links to all the other data you're using)