Assignment

Coding for Economists, 2024

Task

Take 1 or 2 datasets and demonstrate your skills in manipulating and analyzing it. The topic of the project and the correctness of the analysis is not evaluated, only the coding skills you have learned.

Pick a sample dataset, or bring your own. You can also use a project or term paper for another course to satisfy this assignment.

Feel free to consult class notes, Google, StackOverflow, each other, but **type each line of code you write on your own**. No copy-pasting, please.

The deadline is 11pm (Vienna time) on Sunday, **November 10**.

You can access Stata via regular CEU ways (computer lab or VPN).

Show all your work in a public GitHub repository. (Please start a new one.) Submit the link to your repository to complete the assignment. (If your data is too large to add to GitHub or you want to keep it private, please give very specific instructions about how we can get it.)

Learning Outcomes to Demonstrate

1. Understand folder structure. Perform operations on files in different folders.
2. Automate repeating tasks using Python "for" loops.
3. Use Python "lists" and "dictionaries" appropriately. Demonstrate one of the two.
4. Break up work into smaller components using Stata .do files.
5. Read .csv data in. Fix common data quality errors (for example, string vs number, missing value). (in Stata and in Python as well)
6. Prepare a sample for analysis by filtering observations and variables and creating transformations of variables. Demonstrate all three. (in Stata and in Python as well)
7. Save data. (in Stata and in Python as well)
8. Create a summary statistics table. (in Stata and in Python as well)
9. Create a graph of any type. (in Stata and in Python as well)
10. Save graphs as files in Stata.
11. Commit files into git version control.
12. Name files, functions and variables appropriately.
13. Write helpful comments and commit messages. Demonstrate both.
14. Push a repository to GitHub.
15. Explain how to reproduce your work with short instructions or a self-documenting master script.

Grading

Class Participation

The maximum score for class participation is 40.

	Score
Attended 80% of classes, was active in most	40
Attended 80% of classes, was active in some	35
Attended 80% of classes, was inactive	30
Attended 50% of classes	25
Attended some classes	20

Coding Assignment

The maximum score for the coding assignment is 60. The score only depends on the number of learning outcomes demonstrated, not the quality of the analysis question and the project. Each of the 15 outcomes worths 4 points.

Grading

The course is Pass/Fail. You need a score of 60 or higher to pass.