Coding and Writing Exercise
Abhishek Nagaraj, April 1 2022

The goal of the data task is to reproduce one table from an existing paper, plus generating a new figure providing additional insight. The following files are attached to these instructions:

1. Paper – Nagaraj_reimers_april21.pdf
2. Data – loans_merged.dta and LKUPborrowerstatus.csv

I need you to do a few things:

1. Process the loans data and make it into a balanced sample between 2003 and 2011, such that there are a total of 88006 observations (one per book) in each calendar year. Each row is one loan event in the raw data. Impute 0 loans when a book has not been loaned in a given calendar year. Clean this data to keep only relevant columns and call it "master.dta"
2. Using this master.dta data, replicate Table 5 columns 1 and 2 ("main effect") in the main paper and spit out table_5.tex. We do not care about the formatting of your output to match – just the numbers. The variable "year_scanned" indicates the year when a book was scanned, and this variable is missing if it was not scanned. The "location" var has the physical location of the book.
3. LKUPborrowerstatus.csv has codes for different types of borrowers – using this data generate additional results providing insight into how book digitization might have affected different types of borrowers differently focusing on a few interesting categories (students, faculty, staff for example). Why might effects be different across these groups? Represent your key result using a figure generated in stata and neatly labeled for presentation and provide a short discussion interpreting your results.
4. Create a "lastname_firstname_output.tex" file that has both the table and figure above and a short description of your approach and results.

Please avoid asking questions unless extremely necessary (we cannot guarantee a quick response!) – part of the goal of this exercise is to judge your ability to work independently. When you are done with this data task, please send us (nagaraj@berkeley.edu and mbstaha@gmail.com) a zip-file with your name (Last_First.zip) including all the following files:

1) The raw data that I sent you.
2) Your Stata .do file (please comment the code to make it easy to follow, please use Stata rather than R or python). I should be able to run your do-file in my computer and it should generate the final dataset and the files with the final tables/figures. Everything should be automatic (i.e. no need to copy-paste a number to excel or anything like that).
3) A short (say, 1-2 pages) written Latex document describing what you did and with the key figure and table. As before, I should be able to compile this latex document and it should generate the PDF document in my computer using the figure table that were generated with the do-file.
4) The final PDF output for the latex file above.

I expect this task to take approximately 4-6 hours to complete. My goal is to assess the current status of your coding skills related to empirical **economics** research, but a "perfect" score on this task is not a prerequisite to be considered for the job. With that in mind, please do this work yourself, without input from other people.