Data Science for Economists | Project 3 Checkpoint 1

April 15th, 2024

Group composition

1. Johan Oelgaard, johan.oelgaard@berkeley.edu. Responsibility: Table 1
2. Gael Fonseca Gutierrez, gaelf.gutierrez@berkeley.edu. Responsibility: Table 5
3. Nabeel Rahman Qureshi, nabeelq@berkeley.edu. Responsibility: Table 6
4. Felipe Mach Gelbert Barreto, felipe_mach@berkeley.edu. Responsibility: Tables 2-4

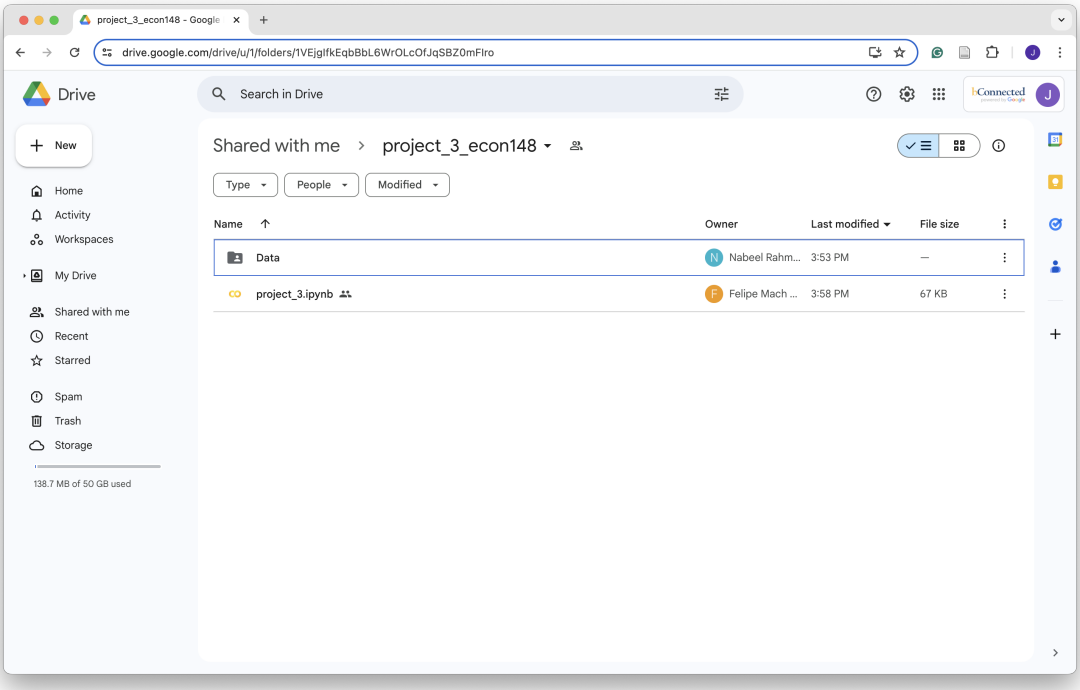
We have not given responsibility for creating any other tables yet.

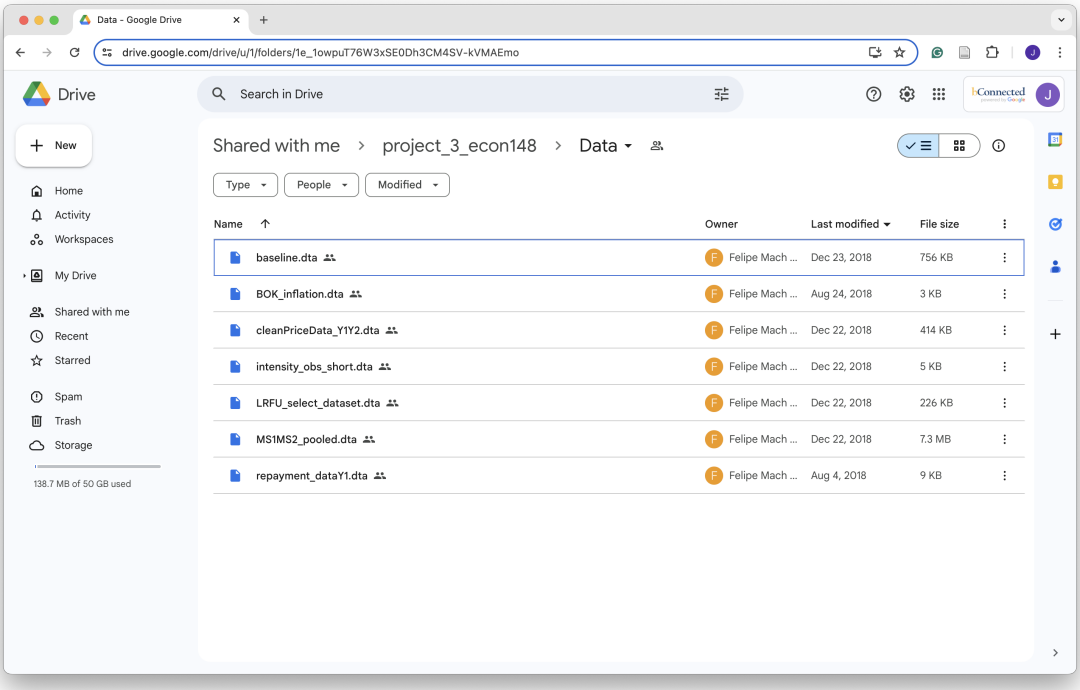
Chosen paper

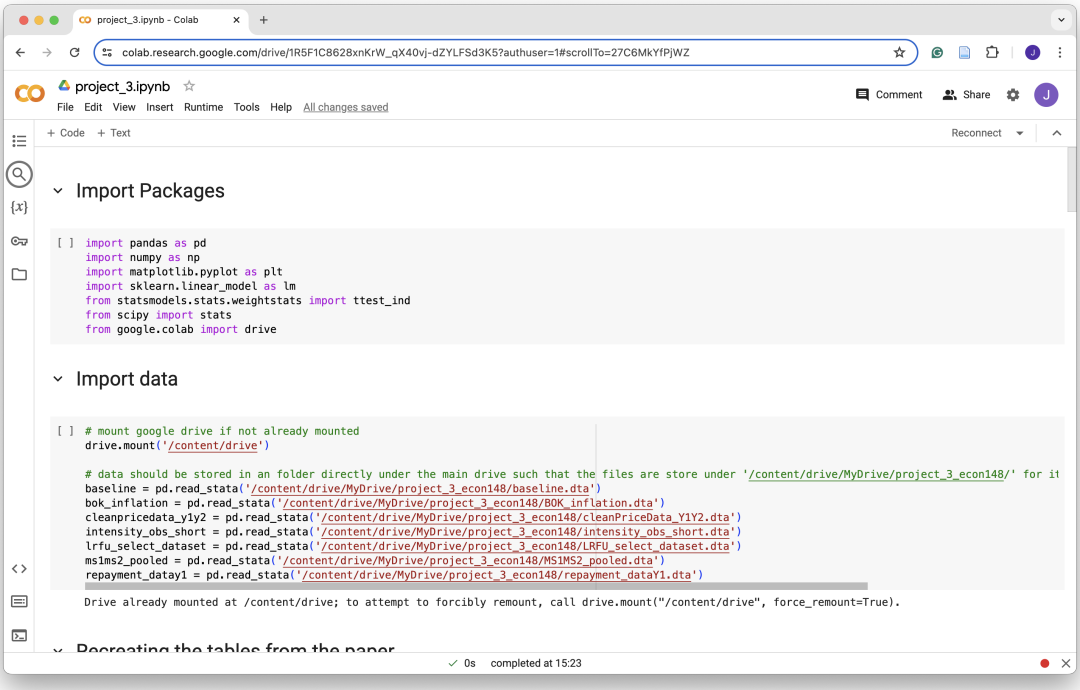
Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel, "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets," The Quarterly Journal of Economics, v.134 2, May 2019, 785-842

Screenshots of the project code organization

We have set up a Google Colab from which we will work from

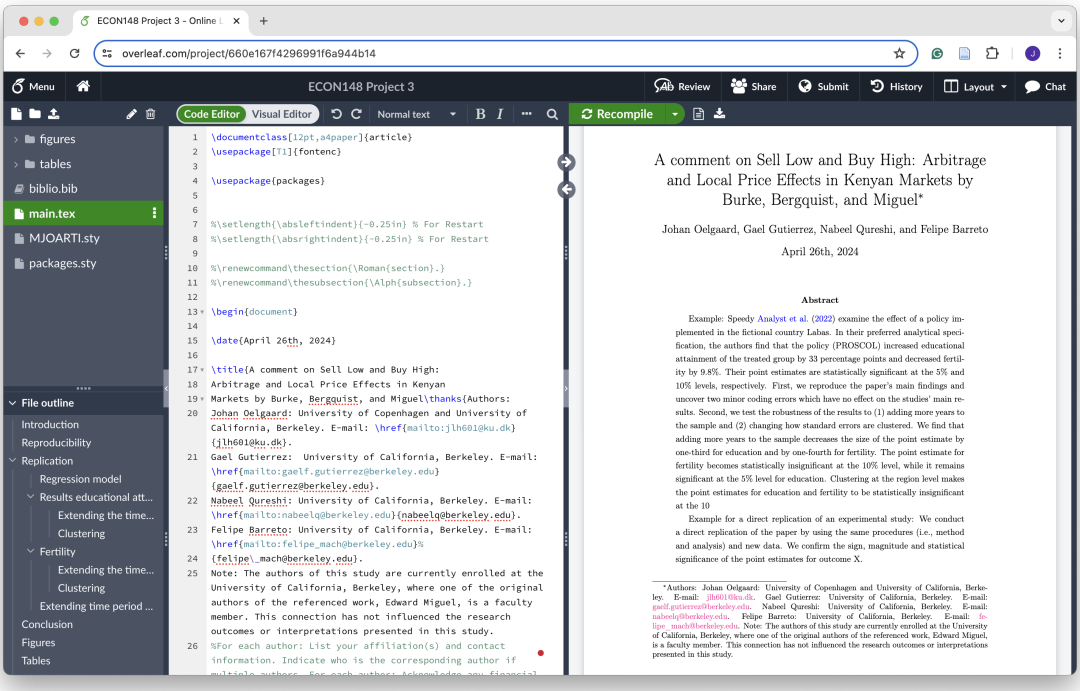






Screenshots of LaTeX project in Overleaf following the I4R template

So far, we have only imported the template and moved the import of the packages to a separate .sty file for a cleaner setup, as well as changing some formalities, e.g. adding our names.



Attempts so far

We have only attempted to recreate the first table, which describes some basic statistics about many of the variables.

Summary of successes and struggles

Successes: We have successfully recreated Table 1 from the paper and are working on the following tables.

Struggles: The data provided is not completely equivalent to what was used by the authors as some variables already have been renamed – a step we can see they do in the `.do` file. This tells us that the data has been altered from the original state before being submitted to the Harvard Dataverse, making it difficult to know if, e.g. in the `.do` we see the column `delta` is modified with `replace delta = 1 - delta` and the subsequently renamed to `delta_base`. When the only column we have is called `delta_base` – has this been modified, or why is the name already changed?

Additionally, there are parts of the code that we do not truly get why they are there – they seem obsolete

How course staff may help us succeed

In the link to the paper from the course website, the figures and appendices are missing (but can be found if we take the paper directly from Edward Miguel's website) – we assume this is because they are not as essential to the paper, and thus, we do not need to worry about this for our recreation – is this a correct assumption?