**On data**

I used a very sensitive data set on for my first paper in which we merged administrative juvenile crime data with a panel of the entire school records of students that attended the Chilean school system (from 1st to 12th grade) that I built.

Then I started working with Laura Guiliano in the construction of a similar panel for 2 counties in Florida.

In my fourth year, I receive funding from the Opportunity Lab to investigate how family background - parents' higher education decisions - influences their children's college and major choices. The plan was to leverage freshly digitalized data on admission records dating back to the 1970s. Despite being discouraged by another researcher who had complete access to the data, I decided to work - and we are still working on it - in collaboration with the public institution that owns the right to this data to ensure future access to this data for any researcher through a transparent and clear process.

Finally, for my JMP I had a hard time to be considered when I was requesting the data through the official channels. After much insistence, in June of this year I signed an agreement to use earnings data from the Department of Finance, which can be access only in-site and through an island computer (totally isolated). Given the sensitivity of the data, they have a strict protocol to add data to their computer.

My paper requires the combination of data identified by the Department of Education identifier, with the earning data, identified with the National identifier. I got the permit from the Department of Education to merge the data, however, they allow the merge of data with national identifier to their datasets only if it was done at their dependencies. Therefore, the data needs for my projects were *de facto* impossible to fulfill. I found a creative solution by aggregating the data.

In my paper all the data is identified with the department of education identifier. I worked with them at their computers to combine the information of the treatment with the national IDs. Then, I got the approval to add this data to the in-site computer at the department of Finance. This was possible only because it was impossible to back-up the student IDs from the shared information.

While this was happening, I started a second data request at the Department of Social Development. This is one of the only places that is allowed to combine information from different Department. Twice a year they open a transparent application process for Chilean research centers that want to develop research projects with their data. Unfortunately, it only works for Chilean institutions, and it requires signatures at all hierarchy levels inside the university (associate professor in charge, department chair, and Dean signature, as well as the ethical committee). I proposed an expansion of my JMP to a professor there to get his support in this application. Updates on this data should be coming soon. As in most of the cases, the access to this data can only be made in-person in-site. And similar to the delays that I observed with the Department of Finance, the COVID protocols made the waiting times significatively longer.

One possible project for me in the future also involve the access to very sensitive data that has never been used in Chile. From a former collaboration with the Justice Department, my coathor and I got access to data on child services. This data is identified at the child level with national ID. We have information on the family court cases, other cases reported later (it’s very common that children that sent to child centers as victims re-enter the center years later as offenders, and we evaluate that and possible alternatives of avoiding it), and judges IDs. The setting seems to be suitable to try a judge design.