Lecture 10: Cell Phone Data

Friday 24 May 2024

Martina Kirchberger

Presentation Slides

Recording

Q: For apps based data user have the option to turning off their location, how big an issue is selection in this case? Is it similar with CDR or not the case as the source of data seems service providers?

A: Good question. I'm not aware of any source of information (survey-based or other) that tells us something about the specific characteristics of users who tend to disable location tracking disproportionately more. But this is effectively not an issue that we encounter with CDR data: users do not have control over the call records that are stored in the telco provider database.

Q: Missing tower locations can be a challenges, especially in Zimbabwe. Network coverage maybe an issue.

A: When you say 'missing tower locations,' do you mean that there are towers for which we do not have data on their location? Or that we may have data on towers that are not actually functioning? Those are two different things, both of them potentially problematic.

A: So this is very much a real issue... As Martina says, there isn't much to do about this except to study your data carefully to try to understand how much of an issue this is in the data. If there are some towers that are never in use, or that go out of use at certain times, you can identify this in the data and figure out how much you need to adjust your research question accordingly... or adjust your interpretation of the results.

In a sense, this is the meta-point of this whole course... Instead of just taking data sources like these and jumping into analysis, you really need to examine your data carefully and to think about the kinds of shortcomings of any data source.

Q: This sounds interesting indeed. More networks means more exposure to new opportunities. I am happy that in Zimbabwe a lot of people now have access to internet and use WhatsApp. It would be interesting to find out how access to networks can influence rural to urban migration. There are a lot of research areas for example how frequency of using certain apps is influenced by other location demographic features, migration etc.

A: Thanks for the comments, both of you. There are really exciting opportunities to use new data sources to think about how people are making choices such as rural-urban migration. These are rich data sources, and they are increasingly becoming available to researchers... so we ought to be able to learn a lot!!"

A: But you also raise the point that we need to be aware of different issues when we work with these data for developing countries than if we were using data from the US or European countries. We can't just take the code that someone has used from a paper about the USA and

apply it to data from Zimbabwe... The structure of the data is likely to be different, with more outages, and we need to pay attention to these differences.

Q: Do we have data on use time of phones as well? (i.e. if the phone is being used at each location or not?)

A: Different sources of phone data will allow us to see different variables. In some types of data, we can see measures that relate to the quantity of use, if that's what you mean. (The other meaning of 'time' here is easier: we often see the time of day at which phones are used.) But as Martina will say, both call data records and smartphone location data will have information about the frequency of use and thus the total amount of use.

A: To complement Doug's answer, a general principle to keep in mind when thinking about the cell phone data that we'll talk about today (i.e. CDR and smartphone app location data) is that users are only observed when they use their mobile phone, either by making calls, sending texts or using a particular application.

Q: Is phone data an example of big data?

A: Yes, you could think of phone data as one example of 'big data'.

Q: What is this mobility index in the paper, "Perpetual Motion: High-Frequency Human Mobility"? Is it at individual level or is it an aggregate index? Can this mobility index (if aggregated) be interpreted as a GDP proxy at local level?

A: A number of individual-level mobility metrics are produced (e.g. fraction of days away from home, average distance away from home etc.) and statistics on these metrics are provided in the paper. For instance, the paper shows the average fraction of days away from home across individuals belonging to distinct groups. I invite you to have a look at the paper for more details if you're interested. As to the second part of your question, mobility and GDP are most certainly related in many different ways but remain two distinct things so you cannot directly use mobility measures as proxy for GDP.