**Class Reflections for UP 494**

Written by - Arpita Banerjee, for Tuesday Session No. 10 (04-06-2021)

Session Name: *The Geography of Evictions*

Reflection Prompts:

*What are our ethical responsibilities when analyzing information that is running in near real time? Are there different considerations for how we tell stories?*

*What is the burden of proof for neighborhood level analyses of two disparate kinds of information (for instance, covid infection rates and eviction rates)? What do we need to show to establish validity of our argument or mixing of data? How might we validate our analysis with “on the ground” examples?*

*How do we communicate effectively with non-technical audiences about near real time relationships in our data? What are some effective narrative vehicles or ways of storytelling that connect people to abstract concepts or information?*

Write Up:

I believe that when it comes to real time data , decisions are made very quickly. Real time data, invariably accompanies real time decisions. This means that the data that is collected and presented for analysis is presumed to be the best representation of reality, or rather the most correct description of reality. However, both personal experience and professional, coupled with existential ruminations often indicate otherwise, real time data can be misleading and often incomplete at best. It is also ill-equipped to handle situational biases or rationalize for inconsistencies that occur when data gets collected quickly, assembled, processed, visualized, and acted upon immediately. These ethical concerns are largely based on the type of concern and ways of representation/visualization we ultimately end up with. For example, if researchers want to investigate pollution data near an industrial site over a period their ethical concerns will be very different from a group of persons conducting a real time analysis of people’s sentiments on twitter. The sensitivity to stimuli will determine the nature of real time data and the fluctuations that it will accompany, however as researchers it is important to acknowledge these ethical dilemmas and provision for them prior to engaging in our projects.

When we look at two different kinds of analysis, for example Covid-19 infection rates and eviction data, our models of data collection, sampling and data handling methods will be very different. I would like to demonstrate this using my imagination of the burdens of proof that might be required. When it comes to clinical research which is dependent on quantitative data, it can be assumed to be of categorical nature – such as – infected/not infected, based on tested/not tested. Therefore, we have a basic typology of information that we can collect through a unified data portal that is accessed across all testing sites within a particular city. The categorical data can be fed as binary inputs and accurate data is collected, the proof being the very nature of definitive tests that are performed. However, when it comes to eviction data, particularly in times of Covid-19, it cannot be differentiated as evicted/not evicted, or reported/not reported very easily. This is because it is not a centralized system and involves a number of informal tactics, and affected population groups who may not necessarily resort to reporting of an eviction. The absence of a centralized system makes it very difficult to collect data, because data standards are missing, and data is highly disaggregating and difficult to locate. A number of proxy indicators are then adopted to validate the data.

Participatory models of information creation and dissemination can represent one of the most effective ways of communicating technical information to non-technical groups. Engaging the community in the generation of data can involve them into the process and enable them to take greater agency over the information being communicated. Visualizing the information through context sensitive maps – simple maps that also contain peripheral information can draw people in – social media can enable us to reach out to a broader audience. It is important to strike a healthy balance between simplification of concepts for public understanding, while retaining a few concepts of information validation and complexity to preserve the collective capacity of non-technical audiences to authenticate and examine technical information, and develop an appetite for the same.