In Support of "Difficult Data"

Jamaal Green

May 20, 2017

In an age where we are constantly reminded of the truly tremendous amounts of data produced on a daily basis (and not infrequently reminded that something like 80% of this data is spatial) what are the essential skills that we as GIS users need to have?

First, in terms of particular technicals skills, I think it's uncontroversial to say that we should be familiar with coding/programming, having mastery of at least one scripting language. We should also have a working knowldge of quantitative analysis techniques (whether Bayesian or frequentist) at least up to various forms of regression. Not only is this a requirement for many "data scientist" jobs, but understanding the statistics that go into our models allow us to be more critical readers and to more easily call bullshit on the many unscrupulous actors out there who take advantage of our country's unfortunate widespread innumeracy. Finally, we should be comfortable with a reasonable range of tools. Get semi-comfortable with the command line (both Unix-based systems and other), can you grab data from FTPs? Have you experimented with web scraping? These are the building blocks I'm finding are increasingly necessary for GIS users interested in social problems.

All of these skills are necessary, I believe, to do interesting, rigorous, beautiful

mapping and analysis, but we should remember that these are only the **tools** with which we create. But all of these tools mean next to nothing if we do not have interesting, or vital, questions that need to be explored. How many really cool maps have you seen that, in retrospect, were technically, or cartographically dazzling, but trivial? We have dozens of maps tracking cyclists on Strava or the best places to buy beer and for what? But the questions that motivate such projects are often shallow, masturbatory projects designed to either sell you some new product or to grow the creator's brand.

I believe the way to cut through this is to ask better questions and to embrace with mapping, and GIS, can do well– make the invisible, or the background, or the marginalized visible and to place these features, actors, or processes in relation to the world around us. This idea is, of course, not new. Countermapping exercises and radical cartographers like Bill Bunge were doing this work in the 60s. But as we look at our current time, and the rise of new tools and technical demands, we have to remind ourselves to still ask questions that are always critical. If we are to truly try and grasp the root of issues, then our questions should explore the unstated assumptions and ignored outcomes of a society built on equal parts capitalism, colonialism, racism and sexism.

There are some awesome recent examples of this kind of work. The good folks at the Anti-Eviction Mapping project are probably the prototypical example of this kind of contemporary work. Their project exposes and maps displacement, a notoriously difficult subject to capture qualitatively, statistically, or carto-graphically. Another project I found just yesterday from the New York Public Library was a webmap of attractions and stops from two years of the "green books", guides for Black travelers written in the mid-20th century America to inform Black folks where they could stay or eat. The map toggles between two years- 1947 and 1956. Out of curiosity I zoomed into Portland and found that

the green book identified only three places in the entire city that were safe for Black travelers say in 1956 and zero in 1947. The geography of Jim Crow, like the geography of slavery, is often assumed to be isolated to the south, but such maps remind us that Jim Crow, both past and in its present incarnations, has always been a *national* project and that the presence of Black people, and other people of color around the country, is often tied to intense violence, both in places where Blacks are especially concentrated or absent.

Both of these projects required collecting, organizing, and visualizing what I call "difficult" data. That is, this is data that is forgotten, deliberately obscured, or may even need to be coerced through tools such as FOIA. It is rare that the truly interesting data you need is already available in neat, tabular formats, or even electronic at all. This is doubly the case when dealing with historical data, but also often the case for data from smaller cities and towns today. If you don't believe me, ask a rural county for their taxlot or housing inspection data. Or look at the planners and engineers in Flint trying to track their sewage pipes with 40 year old faded index cards.

Have you, or are you willing, to spend time the archives? Can you navigate the hundreds of existing federal datasets, many of which are only ever used by government investigators and specialized trade analysts? This is the intellectual legwork we have to be willing to do because the questions we are asking require data and evidence that society largely finds marginalized already. Who cares about evictions data when it's predominately poor Black women who are evicted in neighborhoods we've long forgotten? Why would a state agriculture department keep its pesticide use data in good shape if it would primarily be used to track the health of farmworkers?

Our mission, then, is to ask these questions that challenge conventional logic, that are centered on the interests of the marginalized and oppressed, and help

make that data come alive through mapping. I don't want to overly romanticize our work or to make us heroes because we've been fortunate enough to learn these skills, but in this data-obsessed world, the ability to collect, analyze, and visualize data that challenges widespread injustice in parallel to and in support of social movements is incalculable.

In closing, remember to ask the hard questions. Wrestle with difficult data. Help plot a path to a better future.