**NNIPCamp Denver**

Session 1: Wednesday 10/22/2014, 2:15pm-3:00pm

Location: Theater

Session Title: Technology for ACS Indicators

Organizer: Eleanor Tutt

Primary Notetaker: Katya Abazajian

Participants:



Notes:

How do we all access ACS data?

* At the neighborhood level there is so much error that it’s not plausible to use anymore
	+ Equally probable that there are 12,000 to 15,000 households in poverty
	+ Randomator tool – share link with group
* General group message: Do we need an NNIP tool?
	+ Amy: hoping to get a tool to help make ACS data more efficient
	+ Spencer: Don’t use scripts, use templates with different functions basically models that create indicators (set models in Excel)
* Need to convert into a vertical array because they use Tableau
* Advantage is that it’s reusable from year to year – “old style”
* Group would like to see this tool shared ^^
* Eleanor uses scripts in R
* Erica Raleigh: Use pivot table template at very end that is already set up
	+ Dump into MySQL and use a rudimentary interface to bypass SQL scripting and allow new people to find the data that they need
	+ Do a lot of hand-matching because variable names often change
	+ Used pivot table after exporting from SQL to sum geographies and recalculate margins of error
* Haven’t been doing it for the last two five-year releases
	+ Partially because infrastructure building is not being funded
* Spencer: Variables from ACS are not even in the same order a lot of the time, but when they are, templates are great
* Kimberly: Use Census API, XML and get the table numbers they need, have their programmers aggregate data using API for them
* Eleanor: Do you have an internal doc you go into?
* Kimberly: Django admin database from raw data
* Erica: Is it easier to manage because you’re using the API?
* Kimberly: Yes, the fields match, and also the database caches -- [Profiles.provplan.org](http://profiles.provplan.org/)
* Package in R specify table name and it grabs it from the Census API [called ACS](http://cran.r-project.org/web/packages/acs/index.html)
* Spencer: Can you make new indicators using the scripts in R, is it easy to re-run?
	+ Eleanor: Yes, because that would either mean grabbing different fields in the scripts, or you can aggregate different geographies
* Amy: Uses an excel file for a set of indicators that they recalculate all the time
	+ Calculator of sorts that gathers data for 7 indicators
* April uses NEOCANDO on the front end, and then uses SAS to create tables on the back end
* Michael: Seems that everyone is solving the problems by themselves (subtext: is there something that NNIP can provide, even just a defined set of indicators that we can go by)
* PolicyMap was something everyone was excited about but they didn’t give enough info on how everything was calculated
	+ Would like for someone to be the starting org that can show how to calculate all of those profiles
* Kimberly: ProvPlan profiles can be a good starting point
* Michelle: Would really like to see open source info on ProvPlan profiles
* Kimberly: It is open source, so encourage your programmers to download open source and integrate that into your org
* Eleanor: Emerging best practice is to call AND cache API in case the gov’t shuts down again
* On policing data or making sure error/reliability is visible:
	+ Erica: Detroit has been relying on the ESRI reliability threshold, check ESRI before looking for yellows and greens
	+ Michael: Is there a UI connection to a statistician that could help us refine the rules / first set of formulas that the Census offers or who could help us affect some change (maybe as a consultant?)
	+ Eleanor: Ideally they could help us build one grand reliability tool
	+ Kimberly: ACS Users Group (went to a conference) is a really good resource of statistician/experts who can help understand ACS data
* John: Local cacheing is great, but what you need to do is rebuild your indicators on a local level without ACS … Things like community data collection are the key, and can start the local cacheing
* April also wants to talk to Erica about how infrastructural building is not something that is usually funded
* Spencer uses proprietary data from marketing firm, custom aggregations from their service area (not ESRI)
	+ Way easier to explain that this is year by year data rather than having to use summary data
* Michael: What about discrepancies between proprietary data sources? How many of those vendors of community data are taking advantage of the data sets that we just don’t see?
* Is there a point at which we just accept that we are on unequal footing because we can’t have access to that data?
* Spencer: We are outsourcing that to our provider because that’s what they sell, like people outsource their GIS needs, etc to us.
* Devin: Has anyone ever worked with the Facebook data? NYTimes uses it, how do they get it?
* Erica: Looked into Facebook, Twitter and other data, and there are some companies who sell it from each piece, but you can get some of it for free, but it’s very expensive
	+ 30-40k for 4 fields of individual-level data from Facebook
* Funders like to put their names on books, and the profit from putting out those books helps them fund their internal data systems
* Sean: Processing data is angled as a tool to teach grad students how to use large data, and pitch it that way to get funding
	+ Get interns to clean big data
* Josh: How do you get info on vacant lots because often Census surveyors go to empty lots and vacant houses
	+ Erica: They get the Master Address file from the city, so you would have to get your survey in there
	+ City gave D3 faulty building permits because they didn’t want to appear as if they were declining, so the next year when the 2010 came out they lost way too many housing units
* Eleanor: How have people maintained credibility and *messaged* that to the community?
	+ Devin: Counting the Somali population in Columbus is difficult because the survey and data collectors are not trusted and its hard to survey
	+ Huge discrepancy in what people would say and what surveys say
* Erica: Want a community data collector piece of program
	+ Susan: Use those kinds of issues to try to engage people in the community
		- Show them the faults with it, and ask people to engage with Austin regarding flaws so that they get involved and fill in blanks
	+ Works regarding neighborhood assetsbecasue those are the hardest to track and also most volatile
	+ Helps to both improve the data and move the action
* Maybe we should focus more on trend indicators because then we are just looking at levels
	+ Erica: Problem with that is that you can’t look at overlapping 5-year summaries
	+ Michael: There is an attribution code that is usable that we should look at when we’re talking about error terms (applicable to people who did answer the question)
* TAKEAWAYS:
* A working group around re-building indicators without ACS
	+ How can you use locally collected data to reduce the margin of error on ACS?
* Have a set of commonly used indicators using ACS through NNIP
* Are there alternative reliable providers who have *better statisticians* who have access to resources that we can use?
	+ Would like to understand better the reliability
* List of proprietary listings that we all know of and the accompanying costs