Demographic data 1-12-17

Memo

There are two tabs to the demographic data spreadsheet. The first shows data by address. You’ll notice that there are many near duplicates of addresses. This was to facilitate earlier data lookup methods where Excel or HazMatMapper would call on the demographics spreadsheet with the address spelling that was given in the raw data. To make sure that the spreadsheet responded, it was necessary to list all address spellings. This mostly isn’t all that consequential – most alternate spellings of the same address have the same data. But I’ve highlighted cases where a slightly different address spelling results in not having the same data, because I was unable to lookup some address spellings when I was finding the census tract each address was in. Bottomline: this isn’t a problem in HMM anymore because each site now only has one address associated with it throughout the span of the program. Also note that we simply don’t have data for some addresses.

How did I get this data? Data comes from the 2012 American Community Survey (ACS), downloaded from [American Fact Finder](https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml). I downloaded the “POVERTY STATUS IN THE PAST 12 MONTHS” dataset by zip code and then did a lookup based on the zipcodes given in our dataset. The “percent in poverty” number is used straight from the downloaded data. I also downloaded race data, and performed calculations to get it into the figure presented here. I downloaded the “HISPANIC OR LATINO ORIGIN BY RACE” table and subtracted the “Estimate; Not Hispanic or Latino: - White alone” number from the estimated total population to arrive at a figure representing percent nonwhite (including Hispanic or Latino).

I did the same for both kinds of demographic data at the census tract level. However, “census tract” is not an attribute of our original dataset. For some addresses, I was able to use a Census service online that looks up the tract containing the address. Every address or row in the first tab of the spreadsheet with a number that isn’t 0 in the “addressedTract” column is an address for which the Census service was able to provide a census tract number. For these, I simply then did a lookup on a dataset of all US census tracts by percent in poverty and by race, to get the figures presented. For the addresses without a census tract, I downloaded a file from the Census (<http://www2.census.gov/geo/docs/maps-data/data/rel/zcta_tract_rel_10.txt>) that lists the census tracts associated with each zip code. In many if not all cases a zip code contained or intersected with several census tracts. How did I choose? The lookup function returns the last census tract number associated with that zipcode, i.e. the one at the bottom of the list for the zipcode. While arbitrary, obviously this can affect our results.

We are only missing accurate census tract-level data for three sites. I am comfortable with everything as is for the moment because we provide many different measures in HazMatMapper – 1 mile radius, census tract, and zip – that can, to some extent, act as checks and balances against each other. We will keep looking into it though. If anything, the bigger question is our use of ACS data, for which the margins of error can be quite high.

**ADDRESSES**

receivingFacilityAddress – facility address as listed in the raw data

addressZip – zipcode, straight from the raw data

zippedTract – the census tract corresponding to the site zip code (chosen arbitrarily as noted above)

addressedTract – the census tract corresponding to the facility address. 0 if not found in Census crosswalk table

povTractFinal – the percent of individuals in poverty within a site’s census tract

datasetTractPov – takes all the census tracts in our data and calculates the average percent in poverty

raceTractFinal – percent of nonwhite individuals in a site’s census tract. This is not a figure the Census calculates. Includes those identifying as Hispanic.

datasetTractRace – takes all the census tracts in our data and calculates the average percent of nonwhite individuals

**ZIPCODE**

receivingfacilityzipcode – a site’s zipcode, as given in the raw data

zippedTract – see above

stateID - a state’s number a la <https://www.census.gov/geo/reference/ansi_statetables.html>

zipPov – the percent of individuals in poverty within the site’s zipcode

statePov – the percent in poverty for the state a site is in

ntlPov – the national poverty rate

datasetZipPov – across all zipcodes in our data, the average poverty rate

datasetStatePov – across all states with importers, the average poverty rate

zipRace – the percent of nonwhite individuals in a site’s zipcode

stateRace – percent minority of the state a site is in

ntlRace – percent nonwhite in the US

datasetZipRace – across all zipcodes in our data, the average minority rate

datasetStateRace – across all states with importers, the average minority rate