

HazMatMapping – Socioeconomics data

2-14-17

There are two tabs in the socioeconomics data spreadsheet. The first organizes the data by zip code, the second by address. Notice that there are several near-duplicates of addresses. Ultimately this is a legacy of the different spellings that we coded from each manifest, but it serves the purpose of facilitating Excel table lookups - if you're interested in matching this data to the shipments data you may decide to do a lookup by address (given that each receiving facility is in a different city, we eventually found this attribute to be the most useful for lookups). Also note that we simply don't have socioeconomic data for some addresses.

How did we get this data? Data comes from the 2012 5-year American Community Survey (ACS), downloaded from American Fact Finder. We acquired 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. We also acquired race data, and performed some calculations to get it into the number presented in this spreadsheet. What we did was download the "HISPANIC OR LATINO ORIGIN BY RACE" table and subtract 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.

We did the same analysis for both kinds of socioeconomics data at the census tract level. However, "census tract" is not an attribute of our original dataset – the waste shipment manifests did not list receiving facility census tract. For some addresses, we were able to use a US Census service online that looks up the tract containing the address (Census Geocoder). Every address, or row, in the second tab of the spreadsheet with a number that isn't N/A in the "addressedTract" column is an address for which the Census service successfully provided a census tract number. For these, we then did a lookup on spreadsheets of all US census tracts by percent in poverty and by race, to get the figures presented. For the addresses without a census tract, we 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 we choose? The Excel 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.

However, we were able to check some census tract-zipcode associations against the Census Geocoder. We went through each of the addresses we didn't have addressed tracts and updated the zipped tracts where we could with results from the Census Geocoder. In short, we are only missing accurate census tract-level data for three sites, and these are highlighted in the spreadsheet (we are missing address-specific census-tract data from every address with an N/A in the addressed tract column). We provide many different measures in HazMatMapper – 1 mile radius, census tract, and zip code – that can, to some extent, act as checks and balances against each other. If anything, the more important uncertainties may be in the ACS data, for which the margins of error can be quite high.

ZIPCODE (first tab)

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

stateID - a state’s identifying 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 (“nonwhite” determined as noted above)

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

ADDRESSES (second tab)

receivingFacilityAddress – importer site address as listed in the raw data

addressZip – zipcode for the importer site address, straight from the raw data

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

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

povTractFinal – the percent of individuals in poverty within a site’s census tract (the number reflects the addressedTract if available. Otherwise, it reflects the zippedTract).

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 (“nonwhite” determined as noted above) (the number reflects the addressedTract if available. Otherwise, it reflects the zippedTract).

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