

Codebook for Separate and Unequal Webpages: Metropolitan Regions and Counties

This codebook identifies the downloadable variables from US 2010's webpages for the "Separate and Unequal" report (<http://www.s4.brown.edu/us2010/SUC/default.aspx>).

There are hundreds of variables in these files. Most of them are exposure indices: the characteristics of the neighborhoods where a certain category of people live. "Neighborhoods" include their census tract plus all adjacent tracts. The categories of people are based on major racial/ethnic groupings (non-Hispanic white, black, Hispanic, and Asian). And these groupings are further broken down by the income level of households (low, medium and high) or by the nativity of individual persons (immigrants and natives).

Based on tract-level data, these exposure indices have been calculated for metropolitan regions (using 2010 definitions) and for individual counties. Data are included for the years 1990 and 2000 (from the decennial census sample counts) and 2005-2009 (from the American Community Survey).

It is not efficient to list all of the variable names and their definitions in the codebook. Instead we provide the naming conventions that we used. Every variable name can be "decoded" based on the following rules.

Every record is for a county or metropolitan region, identified as follows:

countyname or metroname= county or metro name
statename = state name
countyid or metroid=FIPS county or metro code in 2010

Other variables have a variable name with this form: **aXb_c**. That is, their names have four components:

- **X** is the year. It is a single digit, and can only take numerical values from {9, 0, 1}.
 - 9 = 1990 Census
 - 0 = 2000 Census
 - 1 = 2005-2009 ACS
- **b** is a character string between **X** and the underscore. It can take up to two characters. Its use depends on the situation. It can indicate whether data are for a county (**k**) or metropolitan region (**m**). It can also indicate whether data are a percentage of individuals (**p**) or a percentage of households (**ph**).
- **a** is the first component of the variable name, placed before **X**. It can take up to three characters. It indicates the racial/ethnic group that the variable refers to and (possibly) categories of persons within the racial/ethnic group.
- **c** is a character string after the underscore. It can take up to four characters. It identifies what we think of as the "substantive" variables, which are mostly characteristics of neighborhoods where people live.

I. Racial/ethnic groups and categories within them: a in aXb_c

When **a** is a single character, it identifies racial/ethnic groups using the following codes:

t = total
w = non-Hispanic white
b = black
h = Hispanic
a = Asian

When **a** has two characters, the first character indicates categories within the racial/ethnic group based on income (of households) or nativity (of individual persons). The second character indicates the racial/ethnic groups as discussed above.

l = low-income households
m = middle-income households
h = high-income households

n = native-born individuals
f = foreign-born individuals

When **a** has three characters, the first two characters indicate subgroup categories (by income or nativity) and racial/ethnic group. The third character is simply a reminder of whether the variable is based on households (when subgroups are by income) or individuals (when subgroups are by nativity).

h = household level
i = individual level

II. Substantive variables: c in aXb_c.

The main neighborhood-level outcome variables for exposure measures are identified in **c** following the underscore using the following codes:

md = median household income
pc = per capita income
pv = poverty
ce = college educated
pf = professional
ue = unemployed
vc = vacancy status
ow = owner-occupied housing unit
fb = foreign born
rc = recent immigrants
ol = language other than English
wh = concentration of white population
is = isolation

These neighborhood characteristics are incorporated into exposure indices for households (when people are categorized by household income, in which case the third digit of **a** is h) or for individuals (when

people are categorized by nativity, in which case the third digit of **a** is i). A more complete definition of these outcome variables can be found in the Separate and Unequal technical note:

<http://www.s4.brown.edu/us2010/SU2/technotes.htm>.

Some variables in the data file are not exposure indices. These include the following variables indicated by a special code for **c**. These are not neighborhood characteristics, but characteristics (medians or counts) of racial ethnic/groups or subcategories of those groups. These categories are identified by **a**.

minc = median household income
hh = total households
low = low-income households
mid = middle-income households
hi = high-income households
pop = total population
tot = total population for each racial group
nat = native-born individuals
for = foreign-born individuals

There are two other exceptional uses for **c**, which depend on values of **b** and **a**. Sometimes it means the variable is a percentage of households:

c = kt or mt and b = ph
c = kt or mt, b = p, a takes value from {lw, lb, lh, la, mw, mb, mh, ma, hw, hb, hh, ha}.

Sometimes it means the variable is a percentage of individuals:

c = kt or mt, b = p, and a takes value from {w, b, h, a, nw, nb, nh, na, fw, fb, fh, fa}.

Consider the following examples:

t9k_minc = county median household income in 1990
w9k_hh = county count of non-Hispanic white households in 1990
b9m_low = metro count of low-income black households in 1990
b9ph_kt = county percent of non-Hispanic black household in 1990
t9k_pop = county count of total population in 1990
w9k_tot = county count of total non-Hispanic white population in 1990
h9m_for = metro count of foreign-born Hispanic population in 1990
mw9p_kt = county percent of middle-income non-Hispanic white households in 1990
na9p_kt = county percent of native Asia population in 1990
b9p_mt = metro percent of black in 1990
twh9k_pv = county total non-Hispanic white households' exposure to poverty in 1990
fhi9m_ce = metro foreign-born Hispanic individuals' exposure to college educated in 1990