

INTRODUCTION TO NEIGHBORHOOD SES DATA SETS

Data set description: One record per person with census tract information for participants address when they were 30-70 years old with up to 5 age points (age decades 30, 40, 50, 60, 70) over those years. Census range is 1960-2000.

Source of Address Data: Age 30 address was asked of everyone at the SES AFU, age 40 & 50 were asked of some depending on their age at the interview. The tract information for addresses of ages not asked at the interview was filled in using the persons Visit 1 – visit 4 tract information. For some of the older participants we were able to fill in age 60 & 70 tract information also.

Address data year gaps: The ideal would be to have address of residence at each age decade, and it fall on the calendar decade year.

Problem 1(addresses from visits only): the date of the visits being used to fill in addresses did not necessarily fall on a participants age decade, the visit with the age closest to the age decade being filled in was used & an age variable (agegap) included with the number of years between the actual age the age decade be represented.

Problem 2(all): the year of the address* did not usually fall on the census year so there is a variable (gapyr) that is the number of years between the date of the address and the census year.

* year of address from AFU form was year participant was 30, 40, 50, for addresses being filled in it was year of visit being used

Naming Convention & description

The variable names of information for age 30, 40, 50, 60, 70 are prefaced with a3, a4, a5, a6 & a7 respectively so a3unempl, a4unempl, and a7unempl are 'Percentage unemployed' at age 30, 40 & 70. Also each of the 5 points also has a variable CENSUS named a3census, a4census, etc. and gives the census year of the data used for that age as well as gapyr (a3gapyr, a4gapyr...) & those points with data taken from visit 1 to 4 data also have a nonmissing agegap (a4agegap, a5agegap, ..).

For the childhood data the variables are prefaced with a1. Childhood data is county based data & participants were asked 'where did you mostly live as a child'. The census closest to their age 10 was used. Census range is 1930-1950.

How useful is data to me?

It would be expected that gapyr would be 5 or less, that is address is within 5 years of the census data being used, however in order to reduce the amount of missing data we have assigned census data from 1970 tracts to people in areas that in 1960 had no available census data. And because we were using 1960 census data (the first year with data compiled by tracts) for people turning 30 in the early 50s, it is possible to have gapyr as large as 18 years. There is one person who turned 30 in 1952 & lived in an untraced area that has 1970 data for age 30 (a3gapyr=18), age 40 (a4gapyr=8) and age 50 (a5gapyr=2). We have left it to the investigator to decide what data to use.

Usually agegap would be only a few years, that is the address is within a few years of the age being represented. However since the range of the visits is only 9 years and some visits are missing address data, this number can be as high as 8(?). Again we have left it to the investigator to decide what data to use.