

ENHANCED GEOCODING APPROACH

Assigning Non-geocoded Case Records to Census Block Groups

To assign (or geocode) a case record to a Census Block Group a street address is usually required. Furthermore, the street address needs to be recognized and validated by the geocoding software. Some case records contain a PO Box as the address, with no street address, and can not be geocoded in the standard way. For still others that have a regular residential street address, the geocoding software cannot recognize the address and therefore cannot geocode it. In all these cases, the best geographic reference point is the ZIP Code of the case record.

The conventional approach to geocoding a ZIP Code to a Census Block Group is to use the centroid of the ZIP Code to place the ZIP Code into a single Block Group. This document describes an enhanced process for assigning case records to a Block Group based on the ZIP Code area instead of the ZIP Code centroid.

The Process

1) Identify the case records to be geocoded by this method.

The first step is to identify the collection of case records to be geocoded using this method. The collection consists of all case records with PO Box addresses, plus all those that have a regular residential street address but that could not be geocoded to a Census Block Group. This latter group can be identified using the Geocoding Result Code returned by the MapInfo MapMarker Plus software. Result codes of “Z1”, “S1”, or “M1” means that MapMarker validated the ZIP Code but not the street address.

2) Group the case records.

Next, the case records are grouped by ZIP Code. All the case records within a particular ZIP Code are treated as a single group and geocoded together, one group at a time. (The reason for this will become more apparent in later steps.)

3) Identify the reference ZIP Code for each group.

Normally the reference ZIP Code of each group is the ZIP Code recorded on the case records included in the group. This would in fact be true for standard residential ZIP Codes. Such ZIP Codes correspond to a set of streets, cover a geographic area, and can be represented as a boundary on a map.⁶ Some ZIP Codes though are used exclusively for PO Boxes. They do not cover a set of streets and therefore do not correspond to a geographic area. Since a boundary cannot be drawn for them, they are represented as points on a map; hence the name point ZIP Codes. Normally, point ZIP Codes are mapped to the location of the Postal facility serving the ZIP Code.

For point ZIP Codes, a proxy ZIP Code is substituted. It is defined as the regular residential ZIP Code that encloses the location of the point ZIP Code. In the following map (Map 6.1), ZIP Code 91729 is a point ZIP Code and is mapped to the location of the Postal facility at 10950 Arrow

⁶ Please note some standard ZIP Codes also serve PO Boxes in addition to residences and businesses.

Route, Rancho Cucamonga, CA. It is enclosed by the boundary ZIP Code 91730, which would be used as the reference ZIP Code, in place of 91729.



Map 6.1

4) Identify the set of candidate Census Block Groups for each group.

Candidate Block Groups are those Block Groups that overlap the reference ZIP Code for a group. A Block Group is considered to overlap the ZIP Code if at least 5 percent⁷ of its geographic area lies inside the ZIP Code’s boundary. Overlaps of less than 5 percent are considered incidental.

Referring to the map illustration (Map 2), ten Block Groups overlap ZIP Code 91701. Of these, nine overlap by at least 5 percent and therefore are considered candidates. For example, 77 percent of the geographic area of Block Group 002007-1⁸ (located in the eastern half of ZIP Code 91701) falls inside the ZIP Code, and therefore is considered a candidate. In contrast, less than one percent of the area of Block Group 009200-4 (which lies to the north of the ZIP Code) falls inside the ZIP Code. Therefore, it is not considered a candidate.

5) Determine the allocation of case records across the candidate Block Groups for each group.

Case records are allocated to a Block Group according to that Block Group’s share of the income-eligible population. The greater the Block Group’s share of the income eligible population, the greater the number of case records allocated to it. A Block Group’s share of the

⁷ Five percent is the conventional cut – off value for applications that use area weighting factors.

⁸ The full identifier of the Block Group is 06071 – 002007 – 1. It is in San Bernadino County.

income-eligible population is simply its proportion of this population among the entire set of candidate Block Groups.

To illustrate this concept, the total population from the 2000 Census is used as a proxy for the income-eligible population, as shown in Table 6.6. The total population across the entire set of candidate Block Groups for ZIP Code 91701 is 46,821. Block Group 002005-1 accounts for 9,723 of them, or 21 percent. Since its share of the population is 21 percent, it would get 21 percent of the case records.

Table 6.6

Block Group	Total Population ⁹	Share of Population
002005-1	9,723	21%
002003-2	7,218	15%
002003-1	5,937	13%
002007-3	5,521	12%
002006-1	5,489	12%
002004-1	4,755	10%
002007-2	3,392	7%
002007-1	3,118	7%
000812-1	1,668	4%
Total	46,821	100%

6) Allocate the case records in the group.

Next, case records are assigned to the candidate Block Groups. Case records (which have been randomly sorted) are allocated to each Block Group according to the Block Group’s share of the income eligible population.

For ZIP Code 91701 shown in the illustration in Map 2, 16 case records¹⁰ need to be allocated across the nine candidate Block Groups. Three case records, or 21 percent of the 16 case records, would be assigned to Block Group 002005-1. Two more cases would be assigned to Block Group 002003-2 because of its 15 percent share, another 2 to Block Group 002003-1 because of its 13 percent share, and so forth until all the cases are assigned. Tables 6.7 and 6.8 show the allocation.

⁹ The population values shown here are from the 2000 Census and do not reflect the actual income-eligible population. These values are used for illustration purposes only. In practice, the income-eligible population will be used.

¹⁰ Sixteen case records from the California grantees either have a MapMarker Result Code of Z1, S1, or M1, or had a PO Box address.

Table 6.7

Block Group	Case Count
002005-1	3
002003-2	2
002003-1	2
002007-3	2
002006-1	2
002004-1	2
002007-2	1
002007-1	1
000812-1	1
Total	16

Table 6.8

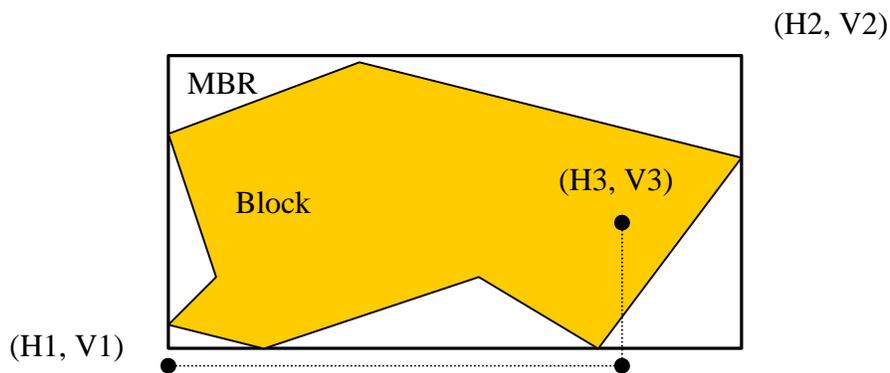
Case Number	Block Group
01E-2000740	002007-3
01E-2001043	002003-1
01E-25000131	002003-2
01E-25000217	002004-1
01E-25000251	002006-1
01E-25000285	002005-1
01E-25000488	002007-1
01E-25000524	002007-2
01E-7000503	002005-1
01E-7001087	002005-1
02E-30003485	002003-2
02E-7000355	000812-1
02E-7000912	002004-1
02E-7001020	002007-1
25-01-05-0522	002003-1
25-01-06-0529	002007-3

7) Edit the map coordinates of the case records.

Once a case record has been assigned to a new Block Group, its map coordinates need to be updated. The map coordinate for the case will be a randomly placed point within the Block Group.

The random placement is accomplished using the following algorithm.

- Calculate the “minimum bounding rectangle” (MBR) of the Block Group.
- Extract the coordinates of the lower left and upper right corners of the MBR. Identify them as $(H1, V1)$ and $(H2, V2)$, where H is the horizontal coordinate and V is the vertical coordinate.
- Generate two random numbers between 0 and 1. Treat the random numbers as a fraction of the distance along a side of the MBR. For example, a value of 0.5 represents the mid-point of the side line. A value of 0 or 1 represents one end of the side line.
- Apply one of the random numbers to the horizontal boundary line to compute a new H3; apply the other to the vertical boundary line to compute a new V3.
- The point defined by $(H3, V3)$ is the new coordinate for the case record. (If the new point happens to fall outside of the Block Group, then repeat the process to get an alternative.)



Map 6.2

8) Assign Geocoding Result Code.

In the final step, the case record will receive a meta-data tag in the Geocoding Pass field to signify how the census and coordinate assignment were made. A Geocoding Pass Code of “10” will denote Enhanced Geocoding Approach.

Implementation

In keeping with the goals of the mapping evaluation project, this entire process was captured in a “stored procedure” inside Microsoft SQL Server. This approach allows the process to be both completely automated and reused in future projects.