

Census

Geocoded demographics from the Census Bureau

1. Overview

The Census file is a combination of geocoded patient/enrollee address data and geography-based demographic information. Its is motivated by the relative dearth of good data on e.g., race/ethnicity and socioeconomic factors like income and education that was (is?) common among HMORN member organizations.

The most common source of demographic information is reported from the US Census Bureau, but there are also vendors who supply updated or more detailed information whose data may be used for this file.

This file should have one record per person (MRN). Include all people with geocodable addresses.

Variable name	Type (Length)	Description	Valid Values	Source File / Segment	Variables & Derivation	Comments
MRN	char (varies)	An arbitrary identifier unique to an individual within a site.	Any, so long as they uniquely identify individual people.	n/a	n/a	Used to link people across files within a site. May or may not contain the official local indigenous person identifier (e.g., "medical record number"). Regardless of whether it contains the official local identifier, this variable should never leave the site. Projects needing to move individual-level data should create a study-specific person identifier and substitute it for MRN on any data that is to move (See, e.g., the % DeIDDset() standard macro.)
geocode	char(15)	MapMarker+ friendly geocode--A numeric code signifying the geographic location of the address where the person was found.	2 to 15 numeric characters.	n/a	n/a	The first two characters signify the state, the next three signify the county, the next six signify the tract, the next 1 signifies the block group, and the final 3 signify the block. Implementers should attempt to geocode all the way down to block level, and include that level of detail. Users who don't need that level can truncate the values down to the

						level they need (though realize that the demographics are matched at the most specific level available--you may need to re-match people up with demographics).
state	char(2)	State Code--a two-character numeric code assigned to US states, districts, territories and protectorates.	Any valid state code, as used by the USCB.	n/a	n/a	
county	char(3)	County Code--a 3-character numeric code assigned to census counties.	Any valid county code, as used by the USCB.	n/a	n/a	
tract	char(6)	Tract Code--a six-character numeric code assigned to census tracts.	Any valid tract code, as used by the USCP.	n/a	n/a	
blockgp	char(1)	Block Group Code--a 1-character numeric code assigned to census block groups.	Any valid block group code, as used by the USCB.	n/a	n/a	
block	char(3)	Block code--a 3-character numeric code assigned to census blocks.	Any valid block code, as used by the USCB.	n/a	n/a	
zip	char(5)	Zip Code--a five-character numeric code assigned by the US Postal Service to various regions where it delivers mail.	Any valid zip code, as assigned by the US Postal Service	n/a	n/a	Note that this makes no portion of the geocode. Zip code regions can and do cross county lines, and change over time.
map_flag	char(1)	The level at which the address was mapped	B Block G Block group T Tract C County Z Zipcode U Unable to be mapped	n/a	n/a	See graph here.
match_strength	char(1)	Code indicating the type of match achieved	o	n/a	n/a	Part of the match metadata string output

		by the geocoding software.	No coordinates 1 Zip +0 centroid 2 Zip +2 centroid 3 Zip +4 centroid 4 Shape Path Centroid 5 Street address position 6 Point zip centroid X Street Intersection			by Map Marker Plus. Very seldom used.
education1	num(8)	Less than 9th grade.	Any real number between 0 and 1.	3/3	sum(P037003, P037004, P037005, P037006, P037020, P037021, P037022, P037023) / sum (P037002, P037019)	These variables give the proportion of people 25+ years old found in the geography, who have attained the education level given in the variable description. Note that these variables are not cumulative . That is, the value in e.g., education3 is not the proportion of people who have <i>at least</i> a high school education—it is the proportion that have <i>exactly</i> a high school education. To get the proportion of people with at least a high-school education you would have to sum the values in variables education3 through education7.
education2		9th -12th grade			Sum(P037007, P037008, P037009, P037010, P037024, P037025, P037026, P037027) /sum (P037002, P037019)	
education3		high school graduate			Sum(P037011, P037028) /sum (P037002, P037019)	
education4		some college, no degree			Sum(P037012, P037013, P037029, P037030) / sum (P037002, P037019)	
education5		associate degree			Sum(P037014, P037031) / sum (P037002, P037019)	
education6		bachelor degree			Sum(P037015, P037032) / sum (P037002, P037019)	
education7		graduate or professional degree			sum(P037016, P037017, P037018, P037033, P037034, P037035) / sum (P037002, P037019)	
medfamincome	num(8)	Median Family Income	Any integer.	3/7	P077001	The USCB does not report median family incomes > \$250k/year.

famincome1		less than \$10,000			P076002/P076001	
famincome2		\$10,000 - \$14,999			P076003/P076001	
famincome3		\$15,000-\$19,999			P076004/P076001	
famincome4		\$20,000-\$24,999			P076005/P076001	
famincome5		\$25,000-\$29,999			P076006/P076001	
famincome6		\$30,000-\$34,999			P076007/P076001	
famincome7		\$35,000-\$39,999			P076008/P076001	These variables give the proportion of families found in the geography whose annual income falls in the range given in the variable's description.
famincome8	num(8)	\$40,000-\$44,999	Any real number between 0 and 1.	3/7	P076009/P076001	
famincome9		\$45,000-\$49,999			P076010/P076001	
famincome10		\$50,000-\$59,999			P076011/P076001	
famincome11		\$60,000-\$74,999			P076012/P076001	
famincome12		\$75,000-\$99,999			P076013/P076001	
famincome13		\$100,000-\$124,999			P076014/P076001	
famincome14		\$125,000-\$149,999			P076015/P076001	
famincome15		\$150,000-\$199,999			P076016/P076001	
famincome16		\$200,000+			P076017/P076001	
fampoverity	num(8)	Proportion of family households in the geography with below-poverty level income.	Any real number between 0 and 1.	3/8	P092003/P013002	
medhousincome	num(8)	median household income	Any integer.	3/6	P053001	The USCB does not report household incomes > \$250k/year. See graph here.
housincome1	num(8)	< \$10,000	Any real number between 0 and 1.	3/6	P052002/P052001	These variables give the proportion of households found in the geography whose annual income falls in the range given in the variable's description.
housincome2		\$10,000-\$14,999			P052003/P052001	
housincome3		\$15,000-\$19,999			P052004/P052001	
housincome4		\$20,000-\$24,999			P052005/P052001	
housincome5		\$25,000-\$29,999			P052006/P052001	
housincome6		\$30,000-\$34,999			P052007/P052001	
housincome7		\$35,000-\$39,999			P052008/P052001	
housincome8		\$40,000-\$44,999			P052009/P052001	
housincome9		\$45,000-\$49,999			P052010/P052001	
housincome10					P052011/P052001	

		\$50,000-\$59,999				
housincome11		\$60,000-\$74,999			P052012/P052001	
housincome12		\$75,000-\$99,999			P052013/P052001	
housincome13		\$100,000-\$124,999			P052014/P052001	
housincome14		\$125,000-\$149,999			P052015/P052001	
housincome15		\$150,000-\$199,999			P052016/P052001	
housincome16		\$200,000+			P052017/P052001	
houspoverty	num(8)	%Households with below-poverty level income	Any real number between 0 and 1.	3/8	P092002/P092001	
pov_lt_50	num(8)	<50% of poverty level	Any real number between 0 and 1.	3/7	P088002/P088001	These variables give the proportion of people found in the geography whose 1999 income was X percent of the federal poverty level. X is given in the variable description.
pov_50_74		Between 50 and 74.99% of poverty level			P088003/P088001	
pov_75_99		Between 75 and 99.99% of poverty level			P088004/P088001	
pov_100_124		Between 100 and 124.99% of poverty level			P088005/P088001	
pov_125_149		Between 125 and 149.99% of poverty level			P088006/P088001	
pov_150_174		Between 150 and 174.99% of poverty level			P088007/P088001	
pov_175_184		Between 175 and 184.99% of poverty level			P088008/P088001	
pov_185_199		Between 185 and 199.99% of poverty level			P088009/P088001	
pov_gt_200		> 200% of poverty level			P088010/P088001	
ra_nhs_wh	num(8)	%white, non-hispanic	Any real number between 0 and 1.	1/2	P008003/ P008001	These variables give the proportion of people found in the geography whose ethnicity was not Hispanic, who indicated they were of the race indicated in the variable description. See graph here.
ra_nhs_bl		%black or african american, non-hispanic			P008004/ P008001	
ra_nhs_am		%american indian or alaska native, non-hispanic			P008005/ P008001	
ra_nhs_as		%asian, non-hispanic			P008006/ P008001	
ra_nhs_ha		%native hawaiian or other pacific islander, non-hispanic			P008007/ P008001	
ra_nhs_ot		%other, non-hispanic			P008008/ P008001	
ra_nhs_ml					P008009/ P008001	

		%two or more races, non-hispanic				
ra_his_wh	num(8)	%white, hispanic	Any real number between 0 and 1.	1/2	P008011/ P008001	These variables give the proportion of people found in the geography whose ethnicity was Hispanic, who indicated they were of the race indicated in the variable description.
ra_his_bl		%black or african american, hispanic			P008012/ P008001	
ra_his_am		%americian indian and alaska native, hispanic			P008013/ P008001	
ra_his_as		%asian, hispanic			P008014/ P008001	
ra_his_ha		%native hawaiian or other pacific islander, hispanic			P008015/ P008001	
ra_his_ot		%other, hispanic			P008016/ P008001	
ra_his_ml		%two or more races, hispanic			P008017/ P008001	
houses_n	num(8)	Number of housing units in geography	Any integer.	3/56	H006001	was H003001 before 14-aug-2009
houses_occupied	num(8)	Proportion of houses_n that are occupied.	Any real number between 0 and 1.	3/56	H006002/H006001	Was H004001/100 before 14-aug-2009
houses_own		Proportion of occupied housing units that are occupied by owners.			H007002/H007001	
houses_rent		Proportion of occupied housing units that are occupied by renters.			H007003/H007001	
houses_unocc_forrent		Proportion of unoccupied housing units that are for rent.			H008002/H008001	
houses_unocc_forsale		Proportion of unoccupied housing units that are for sale			H008003/H008001	
houses_unocc_rentsold		Proportion of unoccupied housing units that are rented or sold, but still unoccupied.			H008004/H008001	
houses_unocc_seasonal		Proportion of unoccupied housing units that are used seasonally or other occasional use.			H008005/H008001	

houses_unocc_migrant		Proportion of unoccupied housing units that are used for migrant workers.			Hoo8006/Hoo8001	
houses_unocc_other		Proportion of unoccupied housing units that are vacant for other reasons.			Hoo8007/Hoo8001	
race_appd_flag		Geographic level at which race data were appended from demographic statistics files.				
nonrace_appd_flag	char(1)	Geographic level at which Non-race data were appended from demographic statistics files.	Same as formap_flag.	n/a	n/a	

3. Notes

Note 1: File / Segment / Variable references are from Census 2000

The references to files, segments & variables in the data dictionary are from the public release datasets from Census 2000. At this writing the analogous files for the 2010 census have not been released.

Note 2: Reference datasets for all 50 states are available from HMORN

The Census workgroup has assembled files of the above variables (and selected others) for you to download and use in the construction of your Census file. Please see this page for details.

Note 3: Handling geocodes and demographic data at different levels of geography

A given address may be geocodeable at pretty much any of the five levels given in the map_flag variable. The USCB also reports different demographic statistics at multiple different levels of geographic specificity. In your geocoding, please preserve as much detail as possible for the file, and let users screen it out if it's not wanted. In other words, please geocode all the way down to the block level to the extent that you can (so--nobody expects you to get census blocks out of PO Box addresses, for instance). When attaching the demographic variables to each record, likewise use the ones from the most specific geographic level possible. For most USCB-sourced demographics, that is the block group level (the one exception being race vars, which are reported at the block level). The two "append flag" variables are there for you to document the levels of the race and non-race variables respectively.

The Census workgroup has put together a SAS program you can use to take a set of geocoded MRNs, and the abovementioned lookup datasets, and output a VDW standard file. Please see this page for details.

4. New V3.1 Census Specification

A new specification was approved at the VIG mid-year meeting in November 2012. Refer to the new data dictionary for the complete specification. (Updates were made to the Census Data Dictionary by Julie Liu after the workgroup call on February 11, 2013.)

V3.1 Census Data Dictionary

The target date for implementation of the V3.1 Census specification is found on the VIG Calendar page.

V3.1 Census Implementation Status by Site

Site Data Managers: please update with the following statuses.

- Not started
- Started
- Verification program results submitted
- Addressing issues (uncovered by program)
- Completed

Feel free to add additional comments and clarifications.


Site	Status	Date of last status update
EIRH	Location 93% (We have some problem addresses to work out) Demographics 30% (Have started programming don't yet have table)	5/23/2013
Fallon	Started	5/24/2013
Group Health	Started	7/22/2013
Geisinger	Started	2/26/2013
Henry Ford	Not Started	2/26/2013
Health Partners	Started	9/4/2013
Harvard Pilgrim	Not Started. We plan to do this when we will do our annual VDW update.	3/29/2013
KPCO	Variables still to be corrected	7/22/13
KPGA	Finished. Includes 2000 GA and all 50 states + DC for 2010 in the CENSUS_DEMOG file.	7/26/2013
KPH	Started.	4/8/13
KPMA	Not started. Just hired new analyst to do this.	7/22/2013
KPNC	Completed	8/28/13
KPNW	Started	5/24/13
KPSC	Not started	4/11/13
Marshfield	Started	4/11/13
SWH		
PAMFRI	Not Started	2/26/13


5. Contact Information


Please contact the VDW Census working group for questions and suggestions.

Related content

Making the CENSUS File at your site


 Bar chart of map_flag

 Boxplot of education level

 Boxplot of % white-not-hispanic

 Boxplot of median household income

Census Bureau Demographics files

 VDW Census Poster (HMORN 2008)

 2009 proposed Census data dictionary (DO NOT USE)

Census QA Workplan

Comments (1)

Liu, Julie Jan 10, 2011 04:19 PM

Hi,

Bhushan from Scott and White constructed the list below that I thought it might be useful to share:

- 1) MapMarker Plus (http://www.empower.com/pages/products_mapmarker.htm)
- 2) ArcGIS (<http://www.esri.com/software/arcgis/index.html>)
- 3) Trillium (<http://www.trilliumsoftware.com/home/index.aspx>)
- 4) Grass GIS (<http://grass.fbk.eu/>) - Open source
- 5) Geocoding using SAS