



Geographic Variables in the NDB

CB EAGYE





Two Types of Geo Data in NDB

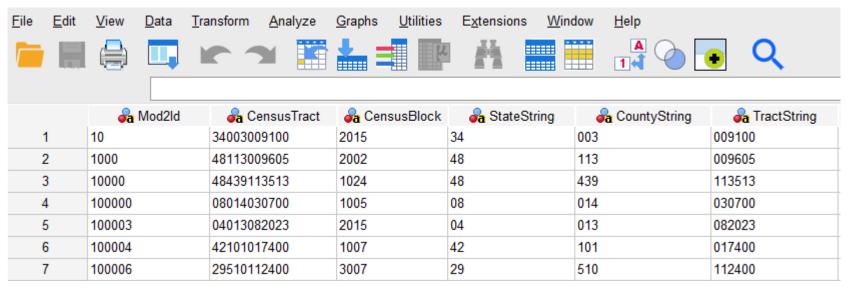
- Geo-identification (geo-id) variables aligned with an address (GeocodeResults)
 - Retrieved from U.S. Census website
 - CensusTract, CensusBlock
 - These variables are associated with an address in Form 2
 - An individual subject may have multiple addresses (collected at different follow-up time points) and therefore multiple census tracts/ census blocks
 - At each quarterly submission, new geo-id variables are added to correspond to newly-collected addresses
 - As of Q3 2021, there are 20,823 census tract records in the NDB (out of 72,455 total Form 2 records = 29%)
- Socioeconomic Status (SES) variables aligned with census tracts (GeoSESVariables)
 - Retrieved from IPUMS website that houses American Community Survey data
 - SESIndex2019, PercentUnemployed, PercentSingleHoH, PercentHoHSorGED, PercentBSorUp, PercentBelowPoverty, PercentSNAP, MedHHIncome, MedFamIncome
 - These are census tract-level variables
 - An individual subject with multiple census tracts will have multiple sets of associated SES variables
 - As of Q3 2021, there are 17,557 SES records





GeocodeResults Data Set

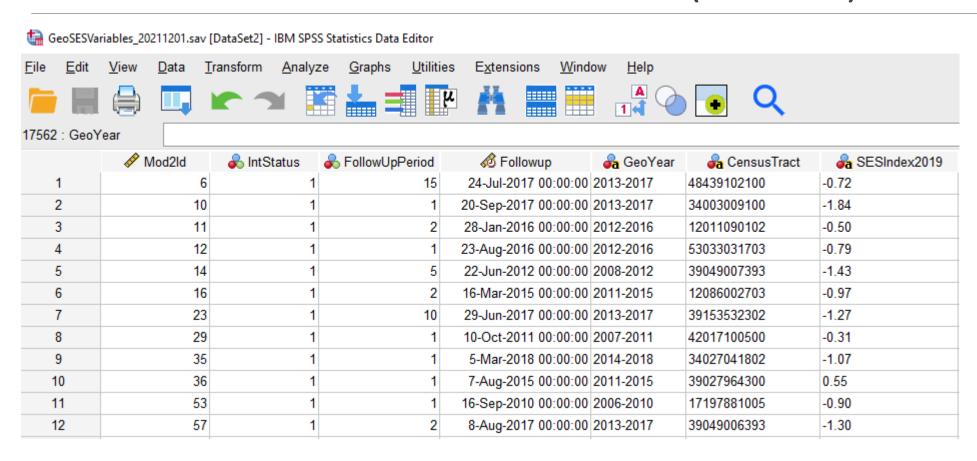








GeoSESVariables Data Set (1 of 4)







GeoSESVariables Data Set (2 of 4)

The most recent SES variables come from the 2015-2019 date range.

It is possible to have a Followup year that is more recent than the GeoYear.

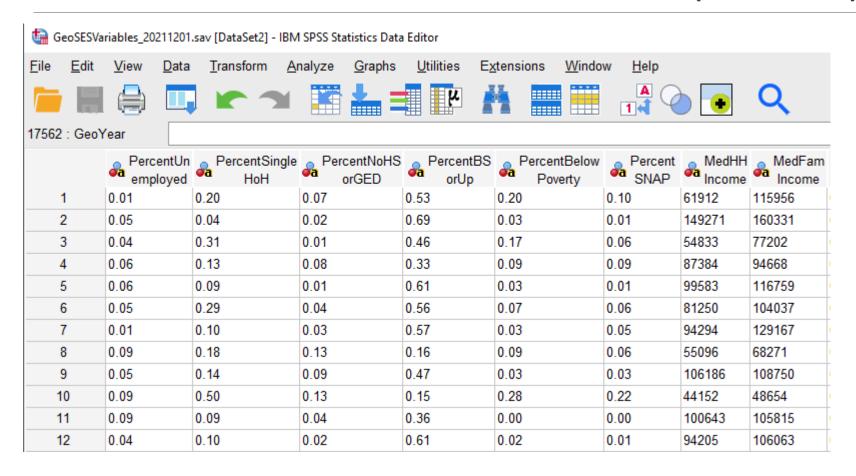
Need to include indicator variable (interim = lags address collection date; final = matches address collection date

Variable Name	geoYear	CensusTract	geoStatus	
	Aggregate Dates of	11-digit census tract,	Indicator of whether the date of the SES variables	
Definition	SES Variables from	matches	matches the date the subject's address was collected	
	IPUMS	GeocodeResults	(interim/final)	
Examples	2015-2019	01001020100	*Interim (Lags address collection date)	
	2012-2016	01003010300	**Final (Matches address collection date)	
			*Meaning the ending date in geoYear is 2019, but	
			the date the address was collected (interview date)	
			is greater than that (e.g., 2020)	
			**Meaning the ending date in geoYear is 2016,	
			which matches the date the address was collected	
			(interview year = 2016)	





GeoSESVariables Data Set (3 of 4)







GeoSESVariables Data Set (4 of 4)

GeoSESVariables_20211201.sav [DataSet1] - IBM SPSS Statistics Data Editor										
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> at	a <u>T</u> ransforr	n <u>A</u> nalyze	<u>G</u> raphs <u>U</u>	tilities E <u>x</u> tensio	ns <u>W</u> indow !	<u>H</u> elp			
	Percent White	Percent Black	Percent Hispanic	♣ STATEA	€ GeoState	€ COUNTYA	🔏 GeoCounty			
1	0.86	0.06	0.20	48	Texas	439	Tarrant County			
2	0.64	0.00	0.08	34	New Jersey	003	Bergen County			
3	0.96	0.04	0.18	12	Florida	011	Broward County			
4	0.77	0.05	0.12	53	Washington	033	King County			
5	0.83	0.13	0.00	39	Ohio	049	Franklin County			
6	0.87	0.06	0.48	12	Florida	086	Miami-Dade County			
7	0.96	0.01	0.01	39	Ohio	153	Summit County			
8	0.95	0.02	0.02	42	Pennsylvania	017	Bucks County			
9	0.62	0.00	0.13	34	New Jersey	027	Morris County			
10	0.97	0.00	0.02	39	Ohio	027	Clinton County			
11	0.98	0.00	0.04	17	Illinois	197	Will County			
12	0.82	0.05	0.03	39	Ohio	049	Franklin County			





Limited Data Set Restrictions

- Both the GeocodeResults and GeoSESVariables data sets include PHI. We can give them to you as part of a Limited Data Set.
- Per the HIPAA Privacy Rule, a Limited Data Set is information from which "facial" identifiers have been removed. For example, names, street addresses, telephone numbers, social security numbers, etc.
- Certain health information items may remain in a Limited Data Set:
 - Dates (such as birth, death, admission, discharge, injury, follow-up)
 - Age
 - Certain elements of geographic address smaller than a state (town or city, county, zip code)
 - Other geographic identifiers smaller than a state (census tract, census block)





How to Request Geo Data

- Internal Notification (Form 602bf) / Request for National Database (Form 602cf)
 - Revised in February 2021 to include requests of geo data and information on linking to external datasets
- Data Use Agreement (SOP 602a)
 - Revised in February 2021 to include policies surrounding the linking of NDB data to external dataset
 - Linking of the TBIMS NDB data to any other data is prohibited without the explicit description of the data to be linked and the linking procedures to be used in the request for TBIMS NDB data.
 - If allowed by the source of the data linked to the TBIMS NDB cases, the data linked to TBIMS cases may be redistributed by the original requestor along with the TBIMS NDB variables Mod1ld (a unique case ID assigned by the TBINDSC for Form 1 variables) and Mod2ld (a unique case ID assigned by the TBINDSC for Form 2 variables), but not any other TBIMS NDB data. [This is intended to prohibit the redistribution of the TBIMS NDB data, while still allowing others to re-link the non-TBIMS NDB data to TBIMS NDB data based on the Mod1ld and Mod2ld stored and distributed as a part of the non-TBIMS NDB data.]





Notification/Data Request

Do you intend to link the TBIMS NDB with any other datasets? If so, please describe the data to be linked to the TBIMS National Database or Module Database data, and the linking procedures/techniques to be used:

Do you require any linking variables in the limited dataset? If so, please check those required and include justification of why variable is needed for the purposes of your study:

- StateCode
- CountyCode
- X CensusTract
- CensusBlock
- Zip Code
 - ZipInj (zip code at injury)
 - ZipDis (zip code at discharge)
 - ZipF (zip code at follow-up)
- _X_ The full date of any date variable

Please indicate which dates: <u>Birth, Death, Injury</u>





The Data Generation Process

- At each follow-up period, ask for consent to collect an address
- Upon each quarterly data submission:
 - Retrieve census tract/block from U.S. Census website for each new address
 - Create dataset called GeocodeResults that contains geo-id variables
 - Update dataset called GeoSESVariables that contains the SES variables to include new census tracts
- Annually:
 - Download latest SES variables from IPUMS website
 - Create latest set of SES variables called GeoSESVariables
- Every 10 years:
 - Update SES Index variable
- Note: We are still working on putting accurate, thorough definitions in the Data Dictionary
 - Until this happens, feel free to contact the NDSC with any questions