

September 2016

The SoVI® Recipe

1. Collect the input variables. SoVI® variables are derived primarily from the US Census Bureau using the Census Data Engine with some ancillary data from the Geographic Names Information System (GNIS). Alternate data sources may include City and County Databook or individual county offices.

2. Normalize all variables as either percentages, per capita values, or density functions (i.e. 'per square mile').

3. Verify accuracy of the dataset using descriptive statistics (i.e. min/max, mean, standard deviation). Missing values can be replaced by substituting the variable's mean value for each enumeration unit. The statistical procedure will not run properly with missing values. Census units with population values of zero should be omitted.

4. Standardize the input variables using z-score standardization: $z = \frac{\chi - \mu}{\sigma}$. This generates variables with a mean of 0 and a standard deviation of 1.

5. Perform the principal components analysis (PCA) using a varimax rotation (100 iterations) and Kaiser criterion (100 iterations) for component selection. This rotation reduces the tendency for a variable to load highly on more than one factor. Next set parameters for the extraction of factors. This can be aided by the examination of a scree plot for significant drops in Eigenvalue as the number of components included in the analysis increases. While some disjoints in the scree are anticipated (such as those that occur between the first few components) subsequent decreases in Eigenvalue indicate appropriate thresholds for factor extraction.

6. Examine the resulting factors. Determine the broad representation and influence on (i.e. increase or decrease) social vulnerability for each factor by scrutinizing the factor loadings (i.e. correlation between the individual variable and the entire factor) for each variable in each factor.

7. Factors are named via the choosing of variables with significant factor loadings (or correlation coefficients)--usually greater than .700 or less than -.700, but in some instances we also examine factor loading from greater than .500 and less than -.500 to help identify the dimension. Next, a directional adjustment (or cardinality) is applied to an entire factor to ensure that the signs of the subsequent defining variables are appropriately describing the tendency of the phenomena to increase or decrease vulnerability. Factor 1 below is an indicator of wealth which **decreases** vulnerability (negative cardinality). The important variables (median house value, percentage rich, median rent) have a positive correlation, so the factor scores on this dimension must be inverted. To adjust the sign of this factor so that those variables appropriately represent their tendency to decrease social vulnerability, a negative cardinality is applied, and the factor is multiplied by -1. Factor 2, labeled race and social status, has four dominant variables (percentage black, percentage female headed household percentage living in poverty, and percent of children not living in a 2-parent household) all of which **increase** vulnerability (positive cardinality), so Factor 2 is entered into the equation as + Factor 2.

8. Save the component scores as a separate file.

9. Calculate a new variable named “SoVI” by placing all the components with their directional (+,-) adjustments into an additive model to generate the overall SoVI® score for the place.

10. Map SoVI® scores using an objective classification (i.e. quantiles or standard deviations) with 3 or 5 divergent classes so illustrate area of high, medium, and low social vulnerability. Normally 1.5-2 standard deviations from the mean are used.

US County 2010-14 Social Vulnerability Component Summary					
Component	Cardinality	Name	% Variance Explained	Dominant Variables	Component Loading
1	-	Wealth	15.833	MHSEVAL	0.873906183
				QRICH200K	0.87161911
				MDGRENT	0.846370212
				PERCAP	0.817764954
				QASIAN	0.68335643
2	+	Race (Black) and Social Status	15.689	QBLACK	0.838618138
				QFHH	0.820467943
				QPOVTY	0.708101777
				QCVLUN	0.65510152
				QED12LES	0.589244914
				QNOAUTO	0.587784173
3	+	Age (Elderly)	12.483	QFAM	-0.792544127
				MEDAGE	0.840840732
				QSSBEN	0.791627081
				QAGEDEP	0.772990516
				QUNOCCHU	0.757050843
4	+	Ethnicity (Hispanic) and Lack of Health Insurance	10.094	QRENTER	-0.707241114
				QHISP	0.918217941
				QESL	0.89318902
5	+	Special Needs Populations	7.769	QNOHLTH	0.53952391
				QNRRES	0.722009603
				QHOSPTPC	0.65255876
6	+	Service Sector Employment	5.724	PPUNIT	-0.518618239
				QSERV	0.77226767
				QFEMLBR	0.590202961
7	+	Race (Native American)	5.554	QEXTRCT	-0.553394535
				QFEMALE	0.943608309
8	+	Gender (Female)	4.955	QFEMALE	0.943608309
Total Variance Explained			78.101		

SoVI® Score = (-Factor 1) + Factor 2 + Factor 3 + Factor 4 + Factor 5 + Factor 6 + Factor 7 + Factor 8