January 2011

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 standard deviation of 1.

5. Perform the principal components analysis (PCA) using a varimax rotation and Kaiser criterion 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 .500 or less that -.500. 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 class and poverty. As shown in the table, the dominant factors that theoretically **increase** vulnerability (people over age 25 w/o a diploma, percent in poverty) have a significant **positive** factor loading. Conversely, the other 2 dominant factors, while still being indicators of socioeconomic status (percent employment and per capita income), theoretically **decrease** vulnerability, and exhibit a **negative** factor loading. Thus, the cardinality of this factor remains positive (+) as the signs on the factor loadings for the individual variables is consistent with their tendency on social vulnerability.

Factor 2 is an indicator vulnerable age groups (i.e. the old and the young). As you can see, both the old and the young, as well as their proxies embody the dominant factors. In examining the variables' factor scores, we see that they exhibit both positive and negative factor loadings, but since all of the variables (i.e. kids under 5, elderly over 65, median age, and social security beneficiaries) have tendency to

increase vulnerability, we apply an absolute value to Factor 2 to dissolve the negative sign on the factors that increase vulnerability, and maintain the cardinality of the variables with non-negative loadings.

Alternatively, some factors may exhibit significant **positive** factor loadings on variables that theoretically **decrease** vulnerability. Factor 4 below is one such example, with positive loadings on mean rent, mean house value and percent rich. 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.

8. Save the component scores as a separate file.

9. Place all the components with their directional (+, -, ll) adjustments into an additive model and sum 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.

The following is an example of the 2000 County SoVI illustrating the factors loadings, naming of the factor, and the sign adjustment (cardinality), as well as the additive formula for the SoVI score. The SoVI score is computed for each enumeration unit (e.g. county, census tract, block group, etc.).

Sign Adjustment	Factor	Name	Dominant Variables	Factor Loading
			QED12LESS	0.873
			QPOVTY	0.867
			QCVLBR	-0.807
+	1	Class and Poverty	PERCAP	-0.776
			MEDAGE	-0.891
			QKIDS	0.836
			PPUNIT	0.829
			QSSBEN	-0.828
ǀǀ	2	Age	QPOP65O	-0.780
			QRFRM	0.795
			QAGRI	0.690
			HOSPPC	0.654
+	3	Rural, Special Needs	NRRESPC	0.520
			HODENT	0.682
			QASIAN	0.660
			MEANS_HSEVAL	0.579
			QRICH	0.514
-	4	Wealth	MC_RENT	0.507
			QFEMLBR	0.773
			QBLACK	0.703
			QFHH	0.556
+	5	Race and Gender	QSPANISH	-0.555
+	6	Female	QFEMALE	0.849
+	7	Service Workers	QSERV	0.782
		Ethnicity and	QINDIAN	0.861
+	8	Unemployment	QCVLUN	0.540
			QTRAN	-0.837
+	9	Migrants	MIGRA	0.502

SoVI Score = Factor 1 + ǀ(Factor 2)ǀ + Factor 3 - Factor 4 + Factor 5 + Factor 6 + Factor 7 + Factor 8 + Factor 9