

Constructing Social Vulnerability Indexes with Increased Data and Machine Learning Highlight the Importance of Wealth Across Global Contexts

Supplementary Methods

Autoencoder

We split the full dataset into training (two-thirds) and testing (one-third) sets. With the training set, we train a model using K-fold cross-validation ($K = 10$) to optimize the number of latent nodes in the latent layer and the number of hidden layers with the lowest loss in the test set. After that, we select the model with the lowest loss on the test set (0.53). The estimated model has 7 hidden layers and 13 latent dimensions. A Tanh activation function is used, as it allows for negative resultant values, which align with the results from the PCA approach.

To interpret the latent layer of the autoencoder, we apply agglomerative hierarchical clustering to cluster counties with similar latent representations corresponding to social vulnerability. The optimal number of clusters is determined by the DB score, which gives a measure of how similar clusters are to themselves compared to other clusters. Lower values of the DB index means that clusters are dense and well separated. Based on the DB score, the number of clusters is set to four. We also compute the number of counties that result in each cluster. A sensitivity analysis on the results using between 1 to 7 clusters shows that interpretation of the cluster with the highest number of counties (using the SHAP methodology to assess the 20 most important features) shows a similar concentration of variables in clusters of counties.

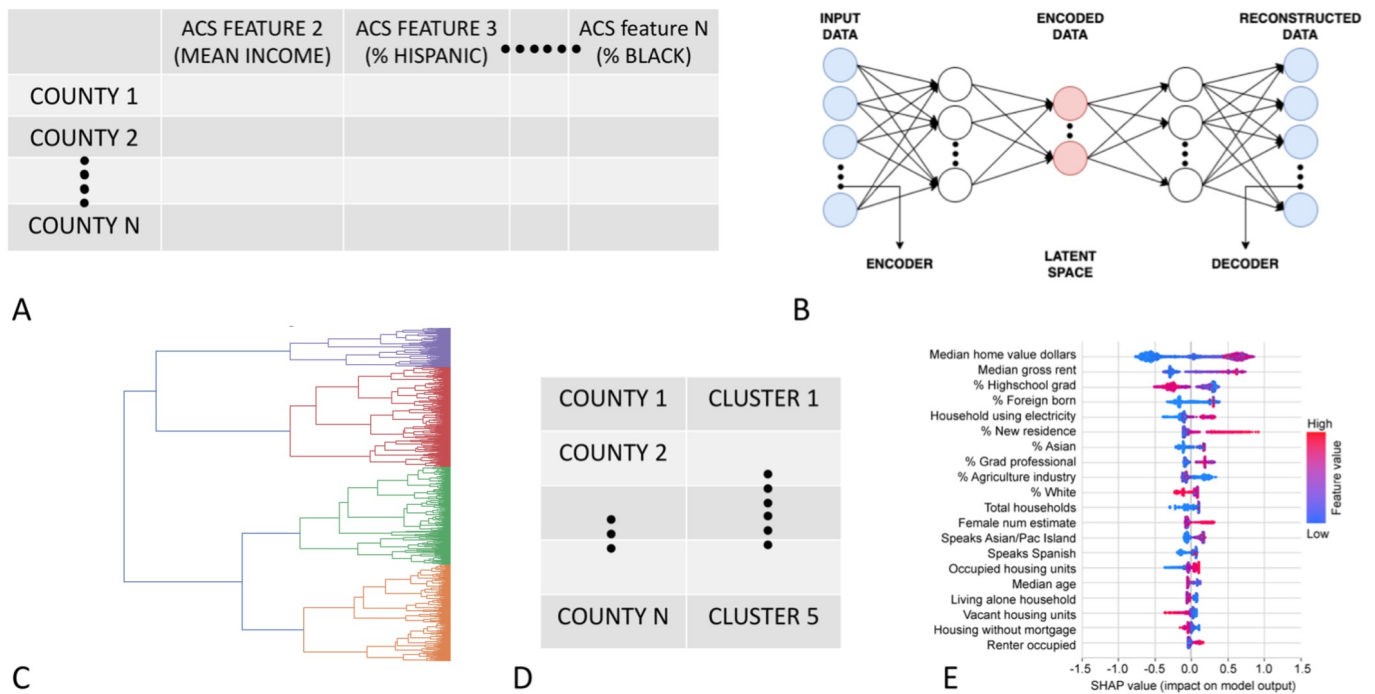


Figure 1. The process of autoencoder analysis: A. Selected variables from the ACS dataset; B. Autoencoder model training and encoded latent layer; C. Use of dendrogram on the latent layer to show the clustering of counties based on different social vulnerable statuses; D. Labeling of each county into the corresponding social vulnerable cluster; E. SHAP methodology applied to show which features are most important in determining social vulnerability clusters in a county.

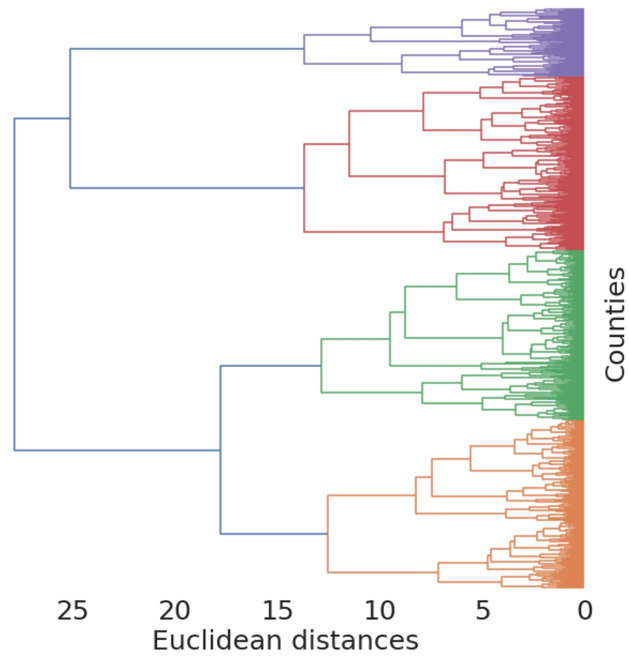


Figure 2. Dendrogram of US counties. The horizontal axis of the dendrogram represents the distance or dissimilarity between clusters. Based on the DB score, which gives a measure of how similar clusters are to themselves compared to other clusters, 4 clusters are ascertained.

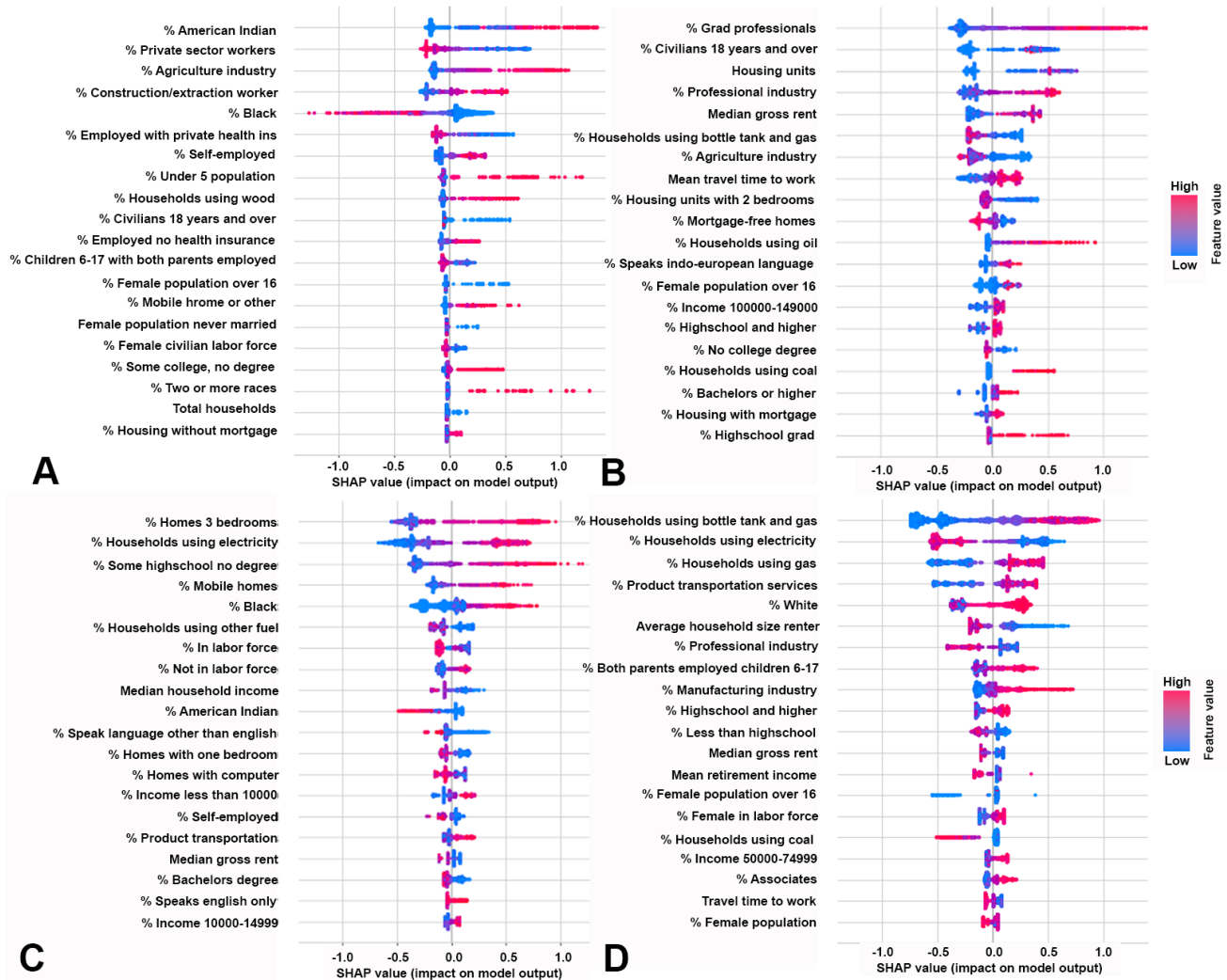


Figure 3. Results of the SHAP methodology showing the top 20 variables (4 clusters). The plot shows the importance (low to high) of a positive or negative value of each variable on the outcome (county groups by social vulnerability). The clusters above include the following proportions of counties: A: 24.8%, B: 30.2%, C: 22.0% and D: 22.9%. All variables are from the ACS data source and are demarcated with units (e.g., %, \$).

Table 1. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Cambodia).

Level 1 (103 Vars)	Labels	Percent Variance Explained	Level 2 (168 Vars)	Labels	Percent Variance Explained
PC1	Household assets	37.8	PC1	Household assets	30.4
PC2	Family composition	9.3	PC2	Family composition	7.9
PC3	Education, amenities and utilities	6.4	PC3	Education, amenities and utilities	5.1
PC4	Marital status and nativity	4.9	PC4	Household structure	3.9
PC5	Employment	4.3	PC5	Religion	3.7
PC6	Household structure	3.5	PC6	Marital status	3.0
PC7	Marital status and dwelling characteristics	3.1	PC7	Employment	2.7
PC8	Religion	2.6	PC8	Nativity, religion and disability status	2.3
			PC9	Size of dwelling	2.1
			PC10	Religion (Jewish and Muslim)	2.0

Table 2. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Costa Rica).

Level 1 (102 Vars)	Labels	Percent Variance Explained	Level 2 (189 Vars)	Labels	Percent Variance Explained
PC1	Household assets, education and employment	43.5	PC1	Household assets, education and employment	35.2
PC2	Family composition and structure	16.3	PC2	Family composition, nativity and dwelling ownership	13.7
PC3	Race and dwelling ownership	8.2	PC3	Household structure	7.3
PC4	Marital status, dwelling characteristics and utilities	5.5	PC4	Dwelling characteristics and marital status	5.2
			PC5	Household structure and age	4.8

Table 3. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Dominican Republic).

Level 1 (107 Vars)	Labels	Percent Variance Explained	Level 2 (195 Vars)	Labels	Percent Variance Explained
PC1	Household assets, education and occupation	38.9	PC1	Household assets, education and occupation	27.1
PC2	Age and marital status	12.9	PC2	Vulnerable population (age and disability)	13.5
PC3	Family composition and structure	11.7	PC3	Family composition and structure	9.3
PC4	Nativity and household structure	6.4	PC4	Household structure and nativity	6.9
PC5	Population size and medical facilities	4.6	PC5	Dwelling characteristics and population size	4.3
			PC6	Urban/rural populations	3.4

Table 4. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Morocco).

Level 1 (105 Vars)	Labels	Percent Variance Explained	Level 2 (195 Vars)	Labels	Percent Variance Explained
PC1	Household assets and education	35.0	PC1	Household assets and education	25.6
PC2	Family composition and population size	16.9	PC2	Family composition, population size and size of dwelling	17.0
PC3	Amenities and dwelling characteristics	8.8	PC3	Dwelling characteristics and household structure	8.0
PC4	Employment	7.5	PC4	Dwelling characteristics and amenities	5.5
PC5	Marital status	4.8	PC5	Language and Employment	4.3
			PC6	Vulnerable population (age and disability)	3.7
			PC7	Size of dwelling	3.1
			PC8	Age and marital status	2.7

Table 5. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Nepal).

Level 1 (103 Vars)	Labels	Percent Variance Explained	Level 2 (304 Vars)	Labels	Percent Variance Explained
PC1	Household assets, education and occupation	39.6	PC1	Ethnicity and religion	21.6
PC2	Dwelling characteristics and nativity	23.5	PC2	Household assets and education	17.0
PC3	Family composition, gender, age and migration	7.7	PC3	Religion and ethnicity	6.8
PC4	Religion	6.4	PC4	Family composition and household structure	5.5
			PC5	Ethnicity	4.3
			PC6	Dwelling characteristics	3.8
			PC7	Religion	2.8

Table 6. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Panama).

Level 1 (108 Vars)	Labels	Percent Variance Explained	Level 2 (219 Vars)	Labels	Percent Variance Explained
PC1	Household assets, occupation and education	57.0	PC1	Household assets, occupation and education	42.1
PC2	Indigenous status, marital status, education and dwelling characteristics	14.2	PC2	Age, marital status and dwelling characteristics	13.9
			PC3	Household structure, ethnicity and dwelling characteristics	7.9

Table 7. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (Senegal).

Level 1 (108 Vars)	Labels	Percent Variance Explained	Level 2 (250 Vars)	Labels	Percent Variance Explained
PC1	Household assets, occupation and education	44.9	PC1	Dwelling characteristics, household assets, age and occupation	31.7
PC2	Ethnicity, religion, gender, and family composition	16.9	PC2	Household structure, ethnicity, religion and education	14.4
PC3	Marital status, amenities, dwelling characteristics and employment	9.6	PC3	Ethnicity, dwelling characteristics and employment	10.0
			PC4	Size of dwelling and family composition	6.4

Table 8. Social vulnerability components including description and ranked by percent total variance explained for Level 1 and Level 2 methods (United States).

Level 1 (67 Vars)	Labels	Percent Variance Explained	Level 2 (164 Vars)	Labels	Percent Variance Explained
PC1	Wealth/Income	17.6	PC1	Wealth/Income	18.2
PC2	Hispanic (Language/Ethnicity)	14.9	PC2	Language/Ethnicity	11.3
PC3	Black (vulnerable black population)	6.6	PC3	Children and elderly	6.7
PC4	Elderly and children	6.3	PC4	occupation/unstable jobs?	5.2
PC5	insufficient housing	5.3	PC5	insufficient housing	4.5
PC6	Utilities	3.7	PC6	Industrial jobs	3.6
PC7	PI Housing	3.3	PC7	Inter-county migration	3.1
PC8	Property Ownership	3.1	PC8	Health insurance and marital status	2.9
PC9	Retail and agriculture	2.7	PC9	Asian household lacking infrastructure	2.0
PC10	Manufacturing	2.2	PC10	Population density	2.0
PC11	Foreign Born (Housing Density?)	1.9	PC11	Percent teenagers and teachers	1.8
PC12	Education	1.8	PC12	Racial diversity	1.6
PC13	Transportation	1.8	PC13	Households with children	1.5
			PC14	Migration and owner/renter	1.3
			PC15	Race and female-led household with child	1.2
			PC16	Military	1.1
			PC17	Household type/infrastructure	1.0
			PC18	Service and retail jobs (no labor jobs)	1.0
			PC19	Income	0.9
			PC20	Utilities	0.9
			PC21	Transportation	0.9
			PC22	Female	0.8

Table 9. Social vulnerability (Level 2, weighted method) correlation with child mortality by country.

Country	Correlation
Cambodia	0.051 $p < 0.05$
CR	0.73 $p < 0.05$
DR	0.51 $p < 0.05$
Morocco	0.71, $p < 0.05$
Nepal	-0.01, $p = 0.93$
Panama	0.48, $p < 0.05$
Senegal	0.68, $p < 0.05$

Table 10. Correlations for vulnerability by county across weighting schemes and Level 1 (L1)/Level 2 (L2) data selection approaches by country.

		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Cambodia	L1 Unweighted	1			
	L1 Weighted	.8899	1		
	L2 Unweighted	.2498	.3946	1	
	L2 Weighted	.6536	.8086	.8177	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Costa Rica	L1 Unweighted	1			
	L1 Weighted	.7739	1		
	L2 Unweighted	.8366	.7767	1	
	L2 Weighted	.6045	.9412	.7567	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Dominican Republic	L1 Unweighted	1			
	L1 Weighted	.8083	1		
	L2 Unweighted	.3061	.4532	1	
	L2 Weighted	.6089	.8695	.7942	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Morocco	L1 Unweighted	1			
	L1 Weighted	.8002	1		
	L2 Unweighted	.7145	.5234	1	
	L2 Weighted	.8682	.9011	.7500	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Nepal	L1 Unweighted	1			
	L1 Weighted	.8899	1		
	L2 Unweighted	.2498	.3946	1	
	L2 Weighted	.6536	.8086	.8177	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Panama	L1 Unweighted	1			
	L1 Weighted	.8570	1		
	L2 Unweighted	.2609	.4116	1	
	L2 Weighted	.5367	.8159	.8191	1
		L1 Unweighted	L1 Weighted	L2 Unweighted	L2 Weighted
Senegal	L1 Unweighted	1			
	L1 Weighted	.8423	1		
	L2 Unweighted	.5582	.7133	1	
	L2 Weighted	.7011	.9367	.8769	1

*All Coefficients are significant at $p < 0.05$

Table 11. Direction of vulnerability changes in standard deviation terms (proportion of counties in each change category) from the Level 1 to Level 2 analyses by country.

		< -2	-2 to -1	-1 to 1	1 to 2	> 2	Total Change
Cambodia	Decrease	0.0	0.33	0.02	0.6	1.0	0.05
	No Change	1.0	0.67	0.96	0.4	0.0	0.93
	Increase	0.0	0.0	0.02	0.0	0.0	0.02
Costa Rica	Decrease	NA	0.0	0.06	0.2	0.0	0.07
	No Change	NA	0.85	0.84	0.7	1.0	0.82
	Increase	NA	0.15	0.1	0.1	0.0	0.11
Dominican Republic	Decrease	NA	0.14	0.05	0.7	0.5	0.18
	No Change	NA	0.5	0.9	0.3	0.5	0.71
	Increase	NA	0.36	0.05	0.0	0.0	0.11
Morocco	Decrease	0.0	0.07	0.04	0.15	NA	0.6
	No Change	0.5	0.73	0.91	0.85	NA	0.86
	Increase	0.5	0.2	0.06	0.0	NA	0.08
Nepal	Decrease	0.0	0.0	0.03	0.0	NA	0.03
	No Change	0.75	1.0	0.95	0.78	NA	0.92
	Increase	0.25	0.0	0.02	0.22	NA	0.06
Panama	Decrease	0.0	0.0	0.00	0.5	NA	0.09
	No Change	1.0	0.67	0.82	0.5	NA	0.74
	Increase	0.0	0.33	0.18	0.0	NA	0.17
Senegal	Decrease	0.0	0.0	0.05	0.25	NA	0.07
	No Change	1.0	0.67	0.95	0.75	NA	0.89
	Increase	0.0	0.33	0.0	0.0	NA	0.04
United States	Decrease	0.0	0.1	0.05	0.41	0.44	0.1
	No Change	0.71	0.53	0.9	0.47	0.56	0.8
	Increase	0.29	0.37	0.05	0.12	0.0	0.1