Analyzing and Contextualizing Social Vulnerability to Natural Disasters in Puerto Rico

Derya Ipek Eroglu
Virginia Tech
deryaipek@vt.edu

Duygu Pamukcu
Virginia Tech
duygu@vt.edu

Laura Szczyrba
Virginia Tech
lszczyrba@vt.edu

Yang Zhang
Virginia Tech
yang08@vt.edu

ABSTRACT

As the third hurricane the U.S. experienced in 2017, Hurricane María generated impacts that resulted in both short term and long term suffering in Puerto Rico. In this study, we aim to quantify the vulnerability of Puerto Ricans by taking region and society specific characteristics of the island into account. To do this, we follow Cutter et al.'s social vulnerability calculation, which is an inductive approach that aims to represent a society based on its characteristics. We adapted the Social Vulnerability Index (SoVI) for Puerto Rico by using data obtained from the U.S. Census Bureau. We analyzed the newly calculated SoVI for Puerto Rico and compared it with the existing deductive approach developed by the Center for Disease Control (CDC). Our findings show that the new index is able to capture some characteristics that the existing vulnerability index is unable to do.

Keywords

Data Analytics, Hurricane María, Principal Component Analysis, Social Vulnerability Index.

INTRODUCTION

Hurricane María made landfall in Puerto Rico on September 20, 2017 as a Category 4 storm. It has been over 80 years since Puerto Rico last experienced a storm of similar magnitude (Coto, 2017). Puerto Ricans faced hazardous storm surges, massive amounts of rainfall and riverine flooding, as well as many landslides on the island. The devastating impacts of Hurricane María were heightened since the federal government was struggling to respond to the two major hurricanes that affected the US prior to María. Hurricane Harvey, one of the costliest natural disasters in U.S. history, hit Houston a month before María. Then, Hurricane Irma passed through the north of Puerto Rico and weakened the island's infrastructure two weeks before María hit the landfall. The number of fatalities was officially recorded as 2,957 (Baldwin et al., 2018), and total damages were estimated to be $90 billion mostly in Puerto Rico (Pasch et al., 2018). The need for a simultaneous response to all the three disasters made the island more vulnerable; and Puerto Rico's distance from the mainland and high damage to the island's ports, airports and roadways worsened the aftermath of the disaster. In addition, the island's existing infrastructure was already vulnerable before Hurricane María hit due to the widespread prevalence of informal housing units, rapid urbanization efforts and late adaptation to strict building codes (Viglucci, 2018). More than a year after Hurricane María, many vulnerable people have no choice but to live in unsafe living conditions due to the poor state of structures prior to María as well as the insufficient response efforts.

Social vulnerability is defined as a multidimensional concept used to identify population characteristics and experiences that enable them to respond to and recover from environmental hazards (Cutter et al., 2003). The pioneering study of Cutter et al. focuses on a Social Vulnerability Index (SoVI) calculation for the United States. However, adaptations of this method are needed to fit this index to different cultural, socioeconomic and demographic characteristics of different regions (Aksha et al., 2018; Chen et al., 2013; Guillard-Gonçalves, 2015; de Loyola Hummell et al., 2016). Because of all mentioned above, there is a need to adapt the vulnerability index for Puerto Rico. In this research, we attempt to propose an augmented social vulnerability index in order to better capture regional and societal characteristics of the island.

In this study, we first investigated literature for existing vulnerability calculation methods, applications of vulnerability calculations and adaptations for different regions. Then, we acquired data for Puerto Rico from the U.S. Census Bureau which initially included roughly 900 variables. After performing data processing, we followed the factor analytic approach of Cutter et al. (2003) and calculated the Social Vulnerability Index for Puerto Rico. Then, we analyzed the newly calculated index in terms of the importance of variables and representative variable groups. Furthermore, we compared our index with an existing vulnerability index developed by the Center for Disease Control (CDC).

LITERATURE REVIEW

Severity and distribution of disaster impact depends on hazard characteristics, community exposure, and vulnerability of built environment (Yoon, 2012). Vulnerable groups are likely to be impacted more from hazardous events because they tend to reside in vulnerable structures and in hazard prone areas, and they have less resources to recover (Cutter, 2003). Therefore, measuring social vulnerability is important for understanding risk and for increasing resilience of the vulnerable groups (Birkmann, 2006).

Vulnerability measurement literature spans a variety of approaches to quantify social vulnerability such as factor analysis (Cutter et al., 2003; Aksha et al., 2018; Chen et al., 2013; Guillard-Gonçalves et al., 2015; Holand et al., 2011; de Loyola Hummell et al., 2016), analytical hierarchical process (Armas and Gavris, 2016; Fernandez et al., 2016) and survey-based measurement (Armas, 2008). Despite the success and theoretical strengths of these approaches in explaining the current vulnerability to environmental hazards in a specified region, updating an outdated index or quantification for a new region is a challenging issue because proposed methods require population-specific or region-specific modifications due to the cultural, socioeconomic, demographic or political characteristics, and most importantly, data availability. Because of this challenge, many adaptations of social vulnerability calculation exist in the literature, most of which are based on the social vulnerability index (SoVI) calculation method of Cutter et al. (2003), a leading, well-known and highly cited study. Aksha et al. (2018) suggest required modifications in SoVI calculation for Nepal in order to reflect the socioeconomic, physical and political context of the country. Similarly, modified SoVI approaches were proposed to reflect the social and cultural context of the Yangtze River Delta Region of China (Chen et al., 2013), Greater Lisbon in Portugal (Guillard-Gonçalves et al., 2015) and Brazil (de Loyola Hummell et al., 2016).

Some of these studies suggest deductive quantification methods that use the available and reliable small set of variables that are assumed to well-represent the population characteristics. Social Vulnerability Index (SVI) of the CDC is an example of the deductive approaches where a determined set of social factors grouped into four themes is used in the index calculation (Flanagan et al., 2011). Frigerio and Amicis (2016) use a small set of variables that represent socioeconomic conditions of Italians, and Gautam (2017) uses an available and reliable set of variables to calculate social vulnerability in Nepal.

As opposed to the deductive approaches, Cutter et al. (2003) follow an inductive variable selection approach where informative variables are reduced from a large set of variables collected. The SoVI of Cutter et al. (2003) is formed based on the Hazard-of-place model of Cutter et al. (1996) in order to help decision-makers to establish the factors that threaten the sustainability and stability of the community. The inductive approach uses a more systematic and exhaustive assessment of social vulnerability where all possible variants are considered at a time (Gautam, 2017). SoVI approach of Cutter et al. (2003) and the following modified versions, for example, of de Loyola Hummell et al. (2016) and Chen et al. (2013) apply factor analysis to reduce indicator variables to its principle components that are able to explain the majority of the total variance.

Our focus in this study is to calculate the Social Vulnerability Index of Puerto Rico that represents the existing characteristics of Puerto Rico. This computation is performed by following the inductive approach, and the newly calculated index is compared with the existing, deductive vulnerability index of the CDC. Details of the collected data and performed analysis are discussed in the following sections.

DATA DESCRIPTION

In this study, we collected two different datasets for two purposes: calculation of a new social vulnerability index, and comparison of this index with an existing vulnerability index.

The data used for the calculation of social vulnerability is collected from the U.S. Census Bureau. The dataset is a combination of ACS 5-year estimation from 2013-2017 and 2010 Census data. There are 945 different census tracts in Puerto Rico. The final dataset includes information for 895 census tracts after the elimination of missing entries.

For comparison purposes, we used an existing vulnerability index developed by the CDC) for Puerto Rico in 2016. The CDC's social vulnerability index, also known as SVI, uses ACS 5-year data from 2012-2016, and index calculation is based on the method of Flanagan et al. (2011). SVI gives a census tract-based vulnerability ranking based on 15 social factors grouped into four themes ("SVI 2016 Documentation," 2016).

COMPUTATION OF SOCIAL VULNERABILITY

The methodology we use is proposed by Cutter et al. (2003), known as the Social Vulnerability Index (SoVI), which is a well-accepted methodology in the literature. This methodology is primarily based on Principal Component Analysis (PCA), a dimensionality reduction technique to extract dominant patterns in the data and find representative variables (Wold et al., 1987). After extracting important principal components, which are the variable groups representative of these patterns, these principal components are used for SoVI calculation.

We started with preparing the dataset for index calculation. To begin with, we removed missing values and cleaned the raw data which contains 900 variables. Then, all variables are first transformed to percentage values from number values to eliminate dependency on population size in each census tract. We kept a small number of variables as is because they represent the size of tract; for example, the population in tract and number of structures. After this transformation, based on descriptive analysis, we removed variables that are not informative, that is, having a deficient range or standard deviation, and that are extremely skewed. As a result, a subset of around 125 variables was derived. The linear relationship between candidate variables was tested using a correlation matrix to eliminate redundant information. We removed variables with a correlation greater than 0.5 or less than -0.5; this range is selected because it corresponds to moderate to high correlation. After these cleaning, preparation and elimination processes, we obtained a final dataset of 49 variables.

Among different variations of PCA, we performed PCA based on covariance by using the statistical programming language R. All of the variables were centered and scaled to perform PCA. After performing PCA, in order to decide the number of principal components to use for index calculation, three rules were checked: principal components having (1) eigenvalue greater than 1, (2) elbow point in scree plot of eigenvalues, and (3) cumulatively explain at least 70% of the variance. Checking these rules, we found 19 principal components for 49 variables. As the last step of the calculation, we analyzed these principal components in terms of their contribution to vulnerability based on correlations (also known as loadings) of the original set of 49 variables with them, and we assigned sign of contribution to each principal component. Then, we calculated the adapted Cutter et al.'s SoVI (2003) for Puerto Rico. A more detailed analysis of variable groups and the calculated index is covered in the following section.

ANALYSIS AND COMPARISON

Analysis of Variable Groups

After we found principal components (or variable groups, interchangeably), we sorted these according to the magnitude of contribution, and we found the leading variables of each variable group as they represent their corresponding variable group. For the 19 variable groups selected, Table 1 describes the cumulative percentage of variance explained, the description of the leading variable and leading variable's contribution sign to the calculated index. We observed leading variables belonging to two main themes: structural and socioeconomic. The majority of the leading variables are related to socioeconomic themes. The group that explains the highest variance is lead by average household size, which increases our vulnerability index. Groups 2 and 18 are related to service and government worker populations, respectively. Group 3 is led by percentage of disabled population, and group 6 is represented by an ethnicity related variable, which is percentage of Hispanic or Latino population. Groups 9, 15 and 19 correspond to finance-related variables, which are percentage of people having income with public assistance, supplemental security, and percentage of people paying extra money for utilities, which possibly points us to the population with high income. Unlike these, a significant portion of 19 variable groups which consists of groups 4, 5, 7, 8, 10, 12, 13 and 14 is led by variables about the structural theme. The leading variable of group 4, which is percentage of housing units occupied by people, increases vulnerability. One important leading variable is one of group 14, which is percentage of structures built in 2010 or later. This variable is found important because it has two reasons: (1) structures are new which indicates higher resistance and lower vulnerability, and (2) structures are built following the new building code which went into effect after 2011 (Viglucci, 2018). In short, variable groups we found from our dataset are led by variables about household size, disabled population, ethnicity, income-related variables, and education enrollment, along with structural variables about resistance, the occupancy rate of housing units and variables describing structure type of housing units.

Table 1 Principal components and their leading variables

Variable Group Number	Cumulative % of Variance Explained	Leading Variable Description	Contribution to Vulnerability Index
1	13.40%	Average Household Size	+
2	21.60%	Percentage of Service Workers	+
3	26.90%	Percentage of Disabled Population	+
4	32.00%	Percentage of Occupied Housing Units	+
5	36.00%	Percentage of Structures with 2 Units	-
6	40.00%	Percentage of Hispanic or Latino Population	+
7	43.00%	Percentage of Mobile Homes	+
8	46.00%	Percentage of Homeowner Vacancy	-
9	48.50%	Percentage of Income with Public Assistance	+
10	51.20%	Percentage of Structures with 1 Attached Unit	-
11	53.80%	Percentage of Sales and Office Workers	-
12	56.30%	Number of Units with No Rent Paid	-
13	58.50%	Percentage of Boats, RVs and Vans	+
14	60.70%	Percentage of Structures Built in 2010 or Later	-
15	62.70%	Percentage of Income with Supplemental Security Income	-
16	64.80%	School Enrollment Population	-
17	66.70%	Percentage in Same House Over 1 Year	-
18	68.50%	Percentage of Government Workers	-
19	70.20%	Percentage Paying Extra Payment for Utilities	-

Comparison with the Existing Index

After the calculation of the vulnerability index, we compared this index with existing SVI developed by CDC to examine if these indices catch similar regions in terms of high and low vulnerability. We first analyzed both indices in terms of similarities and differences of variables they use. We then compared values for both indices. In order to do this comparison, we transformed our vulnerability index to ranking, which is the current format of SVI.

According to SVI documentation, 15 social variables used in SVI using the variables about "unemployment, minority status, and disability, and further groups them into four related themes" ("SVI 2016 Documentation," 2016). Compared to CDC's SVI, SoVI adaptation uses more variables, some of which stand out as representative variables of the variable groups we found in our calculation. The common variables of SVI and SoVI Adaptation are the variables about household size, disabled population size, number of units in structures, and percentage of mobile homes. Other variables we observed in our index also match with the variables used in CDC; however, because our dataset includes more specific variables, we believe that our SoVI Adaptation captures more details than SVI. Different from SVI's variables of Income, Population below Poverty and Unemployed Population, SoVI Adaptation provides us these representative variables: Percentage of Service Workers, Percentage of Government Workers, Percentage of Sales and Office Workers, and Percentage of Income with Supplemental Security Income and Percentage of Income with Public Assistance. Similarly, SVI uses variables Minority and Spoken Language while representative variables that stood out in SoVI Adaptation is Percentage of Hispanic and Latino Population. This shows that our SoVI Adaptation captures more details than SVI index which might help us measure social vulnerability more effectively. The reason SoVI Adaptation catches more details is that representative variables are specific to the dataset. This allows us to assign variable weights depending on the information they provide and change calculation parameters regarding the characteristics of the country or region.

For CDC's SVI, an overall ranking is calculated from four themes in percentile ranking format. As a flagging method based on percentile ranking, CDC uses the top 10% (90th percentile) for high vulnerability and bottom 10% for low vulnerability. After transforming our vulnerability index to percentile rank, we flagged both SVI and adapted SoVI as High, Medium and Low vulnerability. We flagged top quartile as High, bottom quartile as Low, and the rest as Medium. A total of 882 tracts are categorized for both indices and represented as a matrix in Table

2. Summation of diagonals shows overall consistency of two indices which shows that 50% of tracts are in the same categories in both SoVI Adaptation and CDC's SVI. There are 214 tracts that are found to have higher vulnerability according to our SoVI Adaptation compared to CDC, and there are 226 tracts that are found to have lower vulnerability according to our SoVI Adaptation compared to CDC.

Table 2 Comparison of SoVI adaptation and CDC's SVI

		SoVI Adaptation		
		Low	Medium	High
CDC	Low	109	90	23
	Medium	102	236	101
	High	9	115	97

The categorization made by using SoVI Adaptation and CDC's SVI is represented on maps in Figure 1. For both High and Low vulnerability, similarities could be observed in the maps. Both indices categorize northeast as Low vulnerability, roughly corresponding to regions San Juan and eastern La Ruta Panorámica Region. SoVI Adaptation categorizes most of the Eastern region as medium while capturing higher vulnerability. Western part of the island (Porta del Sol) is mostly categorized as Low vulnerability by CDC's SVI while SoVI adaptation categorizes more tracts as High vulnerability (especially some regions along the coastline). Also, some of the tracts of Vieques and Culebra islands which are at the east of Puerto Rico are categorized as Medium by SoVI Adaptation while they are categorized as Low by CDC's SVI. Overall, these analyses indicate that our SoVI Adaptation has some similarities with CDC's SVI, while there are differences deserving further attention and investigation. One future work we plan to do is to analyze the relationship between both indices with damage as a way to evaluate effectiveness of both in measuring vulnerability of Puerto Rico population.

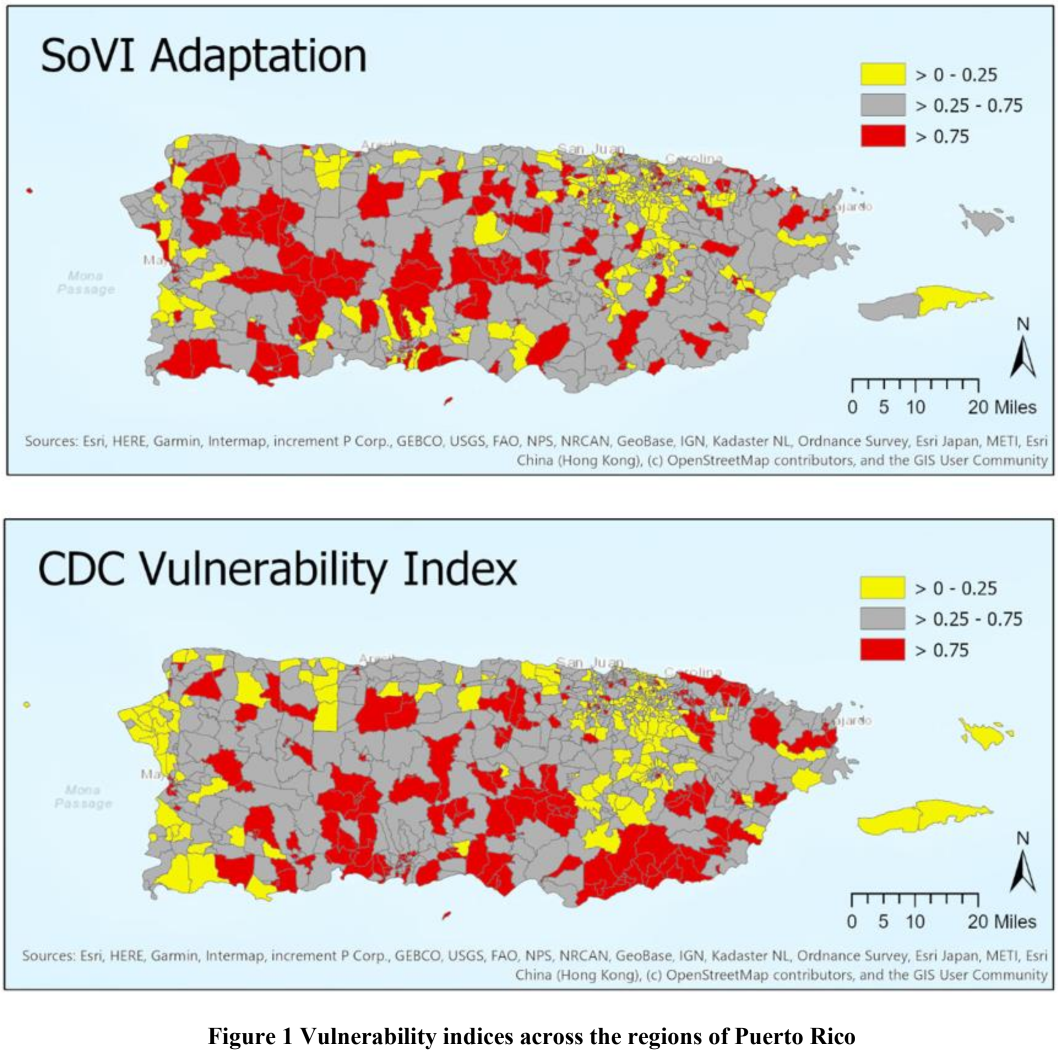

Figure 1 Vulnerability indices across the regions of Puerto Rico

CONCLUSION

In this study, we started with the aim of developing an effective index measuring the vulnerability of Puerto Rico. Existing literature focuses on different ways of measuring social vulnerability which incorporates unique characteristics of a specific population into vulnerability index calculation. After Hurricane María, Puerto Rico's characteristics and differences stemming from demographics, geographies and building conditions became more apparent, which motivated us to develop a vulnerability index that is tailored for Puerto Rico. We followed Cutter et al.'s SoVI (2003) calculation methodology, and we adapted this index by using Puerto Rico census tract data. The SoVI Adaptation is analyzed in terms of variable groups and leading variables. Additionally, a comparison is made to see differences between SoVI Adaptation and CDC's SVI, an existing vulnerability index for Puerto Rico. The preliminary findings show that the SoVI Adaptation highlights variables that might be of particular importance for measuring vulnerability of Puerto Rican population. This contextualization of the social vulnerability could give us a more effective way to measure vulnerability, and it deserves further investigation. An interesting finding is that our SoVI Adaptation has two main themes for 19 variable groups: socioeconomic and structural. Based on this finding, our future work includes separating structural and socioeconomic indices, calculating two vulnerability indices, and investigating the information they provide separately. A Socioeconomic Vulnerability Index will be calculated by using the variables representing the living conditions and population characteristics, and a Structural Vulnerability Index will be calculated based on the variables representing housing characteristics and the structural quality. As a future step, we desire to analyze the relationship between vulnerability indices with physical impact due to Hurricane María and use it as a way to compare the effectiveness of quantification variants in measuring Puerto Ricans' vulnerability to natural disasters. Another future work we are planning to do is to try different methods to find dominant patterns in the data. Despite being a very powerful method, PCA has a limited ability to capture different types of nonlinear relationships in the dataset. To overcome this problem, an alternative methodology could be developed. This might help to represent the vulnerability of a society more accurately and effectively.

REFERENCES

Aksha, S. K., Juran, L., Resler, L. M. and Zhang, Y. (2019) An Analysis of Social Vulnerability to Natural Hazards in Nepal Using a Modified Social Vulnerability Index, *International Journal of Disaster Risk Science*, 10(1), 103-116.

Armaş, I. (2008) Social vulnerability and seismic risk perception. Case study: the historic center of the Bucharest Municipality/Romania, *Natural hazards*, 47(3), 397-410.

Armaș, I., Gavriş, A. (2013) Social vulnerability assessment using spatial multi-criteria analysis (SEVI model) and the Social Vulnerability Index (SoVI model)–a case study for Bucharest, Romania, *Natural hazards and earth system sciences*, 13(6), 1481-1499.

Baldwin, S.L., Begnaud, D. (August 28, 2018) Hurricane Maria caused an estimated 2,975 deaths in Puerto Rico, new study finds, *CBS News*, Retrieved from https://www.cbsnews.com/news/hurricane-maria-death-toll-puerto-rico-2975-killed-by-storm-study-finds/

Birkmann, J. (2006). *Measuring vulnerability to natural hazards: towards disaster resilient societies* (No. Sirsi) i9789280811353).

Chen, W., Cutter, S. L., Emrich, C. T. and Shi, P. (2013) Measuring social vulnerability to natural hazards in the Yangtze River Delta region, China, *International Journal of Disaster Risk Science*, 4(4), 169-181.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003) Social vulnerability to environmental hazards, *Social science quarterly*, 84(2), 242-261.

de Loyola Hummell, B. M., Cutter, S. L. and Emrich, C. T. (2016) Social vulnerability to natural hazards in Brazil, *International Journal of Disaster Risk Science*, 7(2), 111-122.

Fernandez, P., Mourato, S. and Moreira, M. (2016) Social vulnerability assessment of flood risk using GIS-based multicriteria decision analysis. A case study of Vila Nova de Gaia (Portugal), *Geomatics, Natural Hazards and Risk*, 7(4), 1367-1389.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. and Lewis, B. (2011) A social vulnerability index for disaster management, *Journal of homeland security and emergency management*, 8(1).

Gautam, D. (2017) Assessment of social vulnerability to natural hazards in Nepal, *Natural Hazards and Earth System Sciences*, 17(12), 2313-2320.

Guillard-Gonçalves, C., Cutter, S. L., Emrich, C. T. and Zêzere, J. L. (2015) Application of Social Vulnerability Index (SoVI) and delineation of natural risk zones in Greater Lisbon, Portugal, *Journal of Risk Research*, 18(5), 651-674.

Holand, I. S., Lujala, P. and Rød, J. K. (2011), Social vulnerability assessment for Norway: A quantitative approach. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography*, 65(1), 1-17.

Coto, D. (September 21, 2017) Maria destroys homes, triggers flooding in Puerto Rico, *AP News*, Retrieved from https://apnews.com/5f2002103e2f42e4916efeda88d0e511/Maria-destroys-homes,-triggers-flooding-in-Puerto-Rico

Pasch, R.J., A.B. Penny, and R. Berg (2017) National Hurricane Center Tropical Cyclone Report: Hurricane Maria. *National Hurricane Center*, Retrieved from https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf

SVI 2016 Documentation CDC. (2016) Retrieved from https://svi.cdc.gov/Documents/Data/2014_SVI_Data/SVI2014Documentation.pdf

Viglucci, A. (2018, February 14) Half of Puerto Rico's housing was built illegally. Then came Hurricane Maria, *Miami Herald*, Retrieved from https://www.miamiherald.com/news/nation-world/world/americas/article199948699.html

Wold, S., Esbensen, K. and Geladi, P. (1987) Principal component analysis, *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.

Yoon, D. K. (2012). Assessment of social vulnerability to natural disasters: a comparative study. *Natural Hazards*, *63*(2), 823-843.