**Social Vulnerability Analysis Tool**

Scenario Planning Equity Analysis Project, University of Michigan

*Bonnie Wang, Robert Goodspeed, Sabiha Zainulbhai, Jacob Yan*

May 15, 2016 - Draft

**Overview**

This research will include developing a social vulnerability index that is suitable for urban policy decisions, a spreadsheet and R scripts that could be easily employed by the public, and a discussion on data quality and uncertainty.

**Literature Review**

The goal of this social vulnerability tool is to extend methods of social vulnerability analysis developed in the context of natural hazards to the arena of urban development. The concept of vulnerability is to describe states of susceptibility to harm, powerlessness, and marginality in both physical and social systems (Adger, 2006). Vulnerability analysis helps us clarify the sensitivity of people and places when facing risk and threats (Vogel et al., 2007). Past studies suggest that the measurement of vulnerability is both multi-faceted and dynamic, which might include considerations across multiple temporal and spatial scales (Cutter and Finch, 2008; Khan, 2012; Menoni et al., 2012; Turner et al., 2003). Although there is a large amount of literature dealing with vulnerability, most of this literature focuses on how vulnerability analysis can help with hazard recovery. The relationship between social vulnerability and the response to disruptive urban changes has not been rigorously examined, although scholars examining it have generally concluded similar socioeconomic factors make communities resilient for natural and human (planned) changes.

Developing a tool to measure social vulnerability not only help planners determine the sensitivity of certain populations to external effects, but also identify their relative capacity to respond to scenario changes (Brooks et al., 2005; Khan, 2012; Roger et al., 2007). The concept of social vulnerability emerged as a critique of natural hazards research which focused on physical hazards, but not social factors which influence the ability of communities to withstand and respond to crises (Bankoff, 2001; Bolin, 2006; Hewitt, 1983). Social inequalities not only affect the capacity of different populations to recover from disasters, they also illustrate different populations’ ability to adapt to human-caused changes.

Researchers have described several dimensions to the concept of vulnerability. Cutter proposed a model of *place vulnerability* showing interactions between nature, society, and technology, which included two main elements: (1) the geographic context that produce physical or biophysical vulnerability and (2) the social fabric that create social vulnerability (Adger et al., 2004; Birkmann, 2006; Cutter, 1996; European Commission, 2011; Tate et al., 2011). The physical and social vulnerability elements relate and produce overall vulnerability of a place. Social vulnerability emphasizes on the risk when certain populations are exposed to changes, including social, economic, political and institutional elements (Adger et al., 2004; Cutter, 1996; Cutter et al., 2003).

This paper based the social vulnerability index on Lee’s framework of integrated factors of social vulnerability (Lee, 2014). Previous research, such as Cutter, Boruff, and Shirley (2003) and Stephens et al. (1997), empirically validated their indicators of vulnerability by showing the correlation between social vulnerability and the capacity of certain populations to recover from natural hazards. However, they did not address how vulnerable populations might be affected by *planned changes*, such as redevelopment or land use change. Moreover, we did not find any validated metrics in the literature, so we proposed a new social vulnerability index which draws on the indicators that have not been independently validated from past studies of social vulnerability in natural hazards. Although Lee’s social vulnerability index is also used for evaluating the potential risk of a place to mitigate disaster and adapt to environmental changes, his index is the most inclusive and comprehensive as it is extensively supported by previous hazard studies (Table 1). Lee categorized the indicators into two major dimensions, *human capital* and *social capital*. Human capital captures the demographic background of the place, and social capital discusses community development and public infrastructure. This research will focus on human capital. As Lee focused more on hazard mitigation, we have omitted variables less relevant for planned change, such as the number of patients who were served by hospitals.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Types of capital** | **Description of factors** | **Indicators** | **Resources of references** | **Relations to social vulnerability** |
| Human capital | Demographic characteristics | Female population  Age | Morrow (1999), Cutter et al. (2000, 2003), Dwyer et al. (2004), Tate et al. (2010), Ojerio et al. (2011), Schmidtlein et al. (2011) Khan (2012) Cutter et al. (2000), Adger (2003), Cutter et al. (2003), Dwyer et al. (2004), Tate et al. (2010), Ojerio et al. (2011), Khan (2012) Menoni et al. (2012) | Positive  Positive |
|  | Social and economic characteristics | Population density  Birth rate  Infant mortality rate  Households with disabled members  Social and economic status  Poor population  Income  Percentage of population 25 years or older with lower than high school diploma  Rates of unemployment | de Oliveira Mendes (2009), Tate et al. (2010) Khan (2012) de Oliveira Mendes (2009) de Oliveira Mendes (2009) Cutter et al. (2003), Dwyer et al. (2004) Tunstall et al. (2007) Cutter et al. (2003), Schmidtlein et al. (2011), Khan (2012) Adger (2003), Cutter et al. (2000, 2003), Dwyer et al. (2004), Brooks et al. (2005), Tunstall et al. (2007), Polsky et al. (2007), Ojerio et al. (2011), Schmidtlein et al. (2011), Khan (2012) de Oliveira Mendes (2009), Schmidtlein et al. (2011)  Dwyer et al. (2004), de Oliveira Mendes (2009), Schmidtlein et al. (2011), Khan (2012) | Positive Positive Negative Positive Negative Positive Negative  Positive    Positive |
| Social capital | Community development | Working population in primary sector  strength of social network  Percentage of houses rented or seasonal houses | de Oliveira Mendes (2009), Schmidtlein et al. (2011), Khan (2012) Tunstall et al. (2007), Khan (2012)  de Oliveira Mendes (2009) | Positive Negative Positive |
| Public resource provision and public security | Public infrastructure and resources that belong to inhabitants and its safety | Public infrastructure and resources that belong to inhabitants  Quality and price of house  Percentage of old house | Polsky et al. (2007), de Oliveira Mendes (2009), Menoni et al. (2012)  Cutter et al. (2000), Adger et al. (2004) de Oliveira Mendes (2009) | Negative    Negative  Negative |

Table 1. Integrated factors of social vulnerability based on disaster response capacity (Lee, 2014)

**Proposed Social Vulnerability Index**

Lee’s framework provides a useful starting point; however, our research integrated more social variables from other studies shown in Table 1, including race, unemployment rate, poverty status, vehicle occupancy rate, food access, and housing conditions. In this research, we define social vulnerability as the diminished capacity of certain population to react, adjust, cope with and recover from man-made changes. Subsequent research could validate the relationship between these indicators and an outcome variable measuring resilience. Due to limited data availability, we have selected thirteen preliminary indicators for our index which are easily available for any place within the United States. As we target our study area at the neighborhood level, we will only use data that is available at the Census Tract level. Table 2 presents the factors we collected to use as indicators in our social vulnerability index.

|  |  |  |  |
| --- | --- | --- | --- |
| **Human Capital** | **Description** | **Variables** | **Source** |
| Demographic characteristics | Race | Percentage of non-white residents | ACS |
| Age | Percentage of population under age 18 and over age 65 |
| Social and economic characteristics | Rates of unemployment | Unemployment rate for civilian population in labor force 16 years and over | ACS |
| Vehicle ownership rate | Percentage of households with no vehicles available |
| Wealth and Inequality | Poverty | Percentage with income in the past 12 months below poverty level | ACS |
| Healthcare and food access | Health expenditure | Percentage of people without health insurance coverage | ACS |
| Disability | Percentage of population with disability | ACS |
| Food access | Food desert status (Yes = 1, No = 0) (more than 1 mile away from the nearest supermarket) | USDA |
| Education Attainment and Language | Education status | Percentage of population with less than regular high school diploma | ACS |
| Linguistically isolated household | Percentage of limited English speaking household | ACS |
| Housing | Occupancy status | Percentage of Vacant housing units | ACS, HUD/USPS |
| Housing cost burden | Percentage of households who pay more than 30 % of their income rent | ACS |
| Renter | Percentage of renter-occupied housing units | ACS |
| Data Sources:  ACS: http://www.census.gov/data/developers/data-sets/acs-survey-5-year-data.html  USDA: http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx | | | |

Table 2. Proposed Social Vulnerability Index

**Description of Proposed Index**

*Demographic Characteristics* — Demographic characteristics can be summarized as two main indicators: the percentage of non-white residents, and the percentage of population under age 18 and over age 65. Populations that include more non-white residents, children, or elderly people are more vulnerable because there is a higher possibility of these two groups of people may have less ability to rapidly learn about and react to changes.

*Social and Economic Characteristics* — The Social and Economic Characteristics includes percentage of unemployment and vehicle ownership rate. A higher unemployment rate usually indicates lower economic stability of a community. The vehicle ownership rate reflects the mobility of the population. A higher unemployment rate is positively correlated with social vulnerability not only because of reduced income, but also because of other effects such as the lack of work-based social ties.

*Wealth and Inequality* — The wealth and inequality section consists of the percentage of population living in poverty. This percentage is calculated using the population with income in the past 12 months below poverty level data. This group of people are economically disadvantaged and are less likely to have the resources to cope when changes happen.

*Healthcare and Food Access* — The health and nutrition conditions represents a population’s living conditions. The factors are determined by health expenditure, disability, and food access. Because of the limited access to data, we use health insurance coverage to identify the vulnerable population who faced difficulty to access healthcare services (Cohen and Yu, 2002). The percentage of population with disability indicates the potential cost burden for each family. Food access indicates whether a neighborhood lacks healthy food sources. The Food Access Research Atlas defines a *low-access census tract* as being one mile (urban) or ten miles (rural) away from the nearest supermarket, supercenter, or grocery store.

*Education Attainment and Language* — The education status also implies higher possibility of poverty, which is likely due to people’s lack of access to resources, knowledge, and information. Therefore, the populations that might be most vulnerable are those aged 25 years or older with education attainment of a high school diploma (GED) or less. A *linguistically isolated household* is defined as a household where all the members age 14 years old and over have at least some difficulty with English.

*Housing —* Housing characteristics reflect neighborhood stability. The factors used to measure neighborhood stability are occupancy status and percentage of houses rented. Occupancy status is calculated using the percentage of housing units that are vacant. The housing cost burden of a household indicates households who pay more than 30% of their income for rent. These variables assume neighborhoods with vacant properties are more likely to experience negative changes more rapidly, and that renters are more vulnerable to change than owners.

**Index Development Method**

In order to examine our social vulnerability index, we apply the index to the seven counties, identified by the Southeast Michigan Council of governments (SEMCOG), and create a map of social vulnerability. All of the data we use for our social vulnerability index has been recalculated. We created our own derived estimates to highlight vulnerable population from the summary data available in American FactFinder. To make our variables easier to read and compare, we recalculate all of the data into a consistent format. After identifying the “vulnerable groups” for each indicator, indicators showing higher percentages of population or household represent a higher possibility of social vulnerability. While all of the indicators are presented in the same way, each of them have their own scale and are not yet ready to be integrated as a single index. To solve this problem, we use standard score (z-score) to enable ourselves to compare and aggregate two datasets that have different normal distributions.

z = x – μ / σ

x = score

μ = mean

σ = standard deviation

Fig 1. Formula of calculating Z-score

Fig 1 shows the formula we used to calculate z-scores. Z-scores are expressed in terms of standard deviations from their means. These z-scores have a distribution with a mean of 0 and a standard deviation of 1, and it can tell us how many standard deviations from the mean our score is. After standardizing all of the data, we summarized the scores of all 13 variables for each of the census tract.

**Index Application in Metropolitan Detroit**

Figure 2 shows the social vulnerability scores of Southeast Michigan. The blank area is census tract that have either zero population or zero households. The most vulnerable neighborhoods are clustered in the City of Detroit and Pontiac, but areas of many surrounding municipalities are also vulnerable.

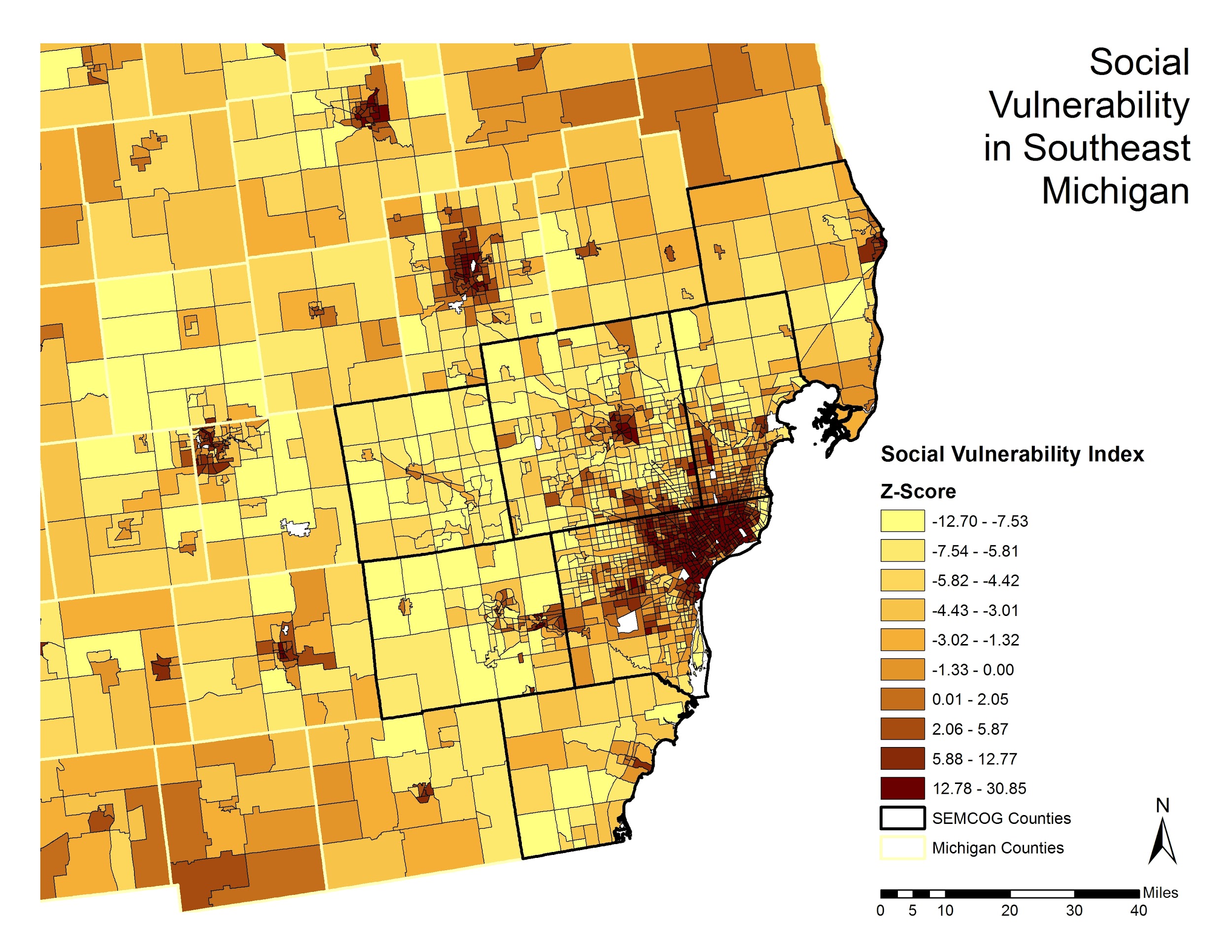


Fig 2. Social Vulnerability in Southeast Michigan

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Minimum | -12.70 | | Q1 | -5.20 | | Median | -1.50 | | Q3 | 3.62 | | Maximum | 30.85 | |  |

Fig. 3. Distribution of Index Scores for SEMCOG Region

**Data Quality and Uncertainty**

Although it is an often overlooked issue, since the index relies heavily on ACS data, we urge users to carefully consider the issues of data quality and uncertainty. This section contains an overview of how margins of error have been handled, and why they cannot be simply summed across all index components. In addition, it reports the results of our exploration of the extent to which the index variables are correlated at an individual level.

When something is *uncertain*, it means that something is unknown or cannot be known (Abbot, 2005). Uncertainty has always been part of the “ongoing processes of nature” (Dewey, 1929), and planning is all about controlling uncertainty by preparing actions to secure the future (Marris, 1987). “Uncertainty” occurs when information is inadequate. People have tried to find scientific ways to capture the mechanisms of places, but there are always gaps between the known and unknowns. American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people. It collected a large number of data that help us summarized the characteristics of a place, and has became the main source for planners to learn about the demographic and social characteristics of a place during decision making. However, because of the nature of statistical and sampling problem, the data never represents a definite true value, but instead lies within a range of uncertain values.

In order to address the uncertainty issue of using ACS census data, our research tries to include a margin of error (MOE) term from each indicator into the social vulnerability index. MOEs show the difference between an estimate and its upper and lower confidence bounds. The main purpose of MOEs is to protect against misinterpreting small or nonexistent difference as meaningful. The accuracy of data controls the reliability of estimates, because higher margin of error terms lead to higher variation of estimates. In our research, we try to incorporate margin of error to help us learn or even manage these “known unknowns.”

All of the data we included in our dataset has been recalculated, which means that the summary data are generated to derived estimates that are not available in the American FactFinder, including other kinds of aggregated estimates, proportions, or ratios. We also recalculate MOEs by using the formula shown in Table 2 to match our variables with new derived estimates. Although MOEs can be calculated for our derived statistics, the MOEs for all of the components of our index cannot be combined because the underlying correlations among variables is unknown.

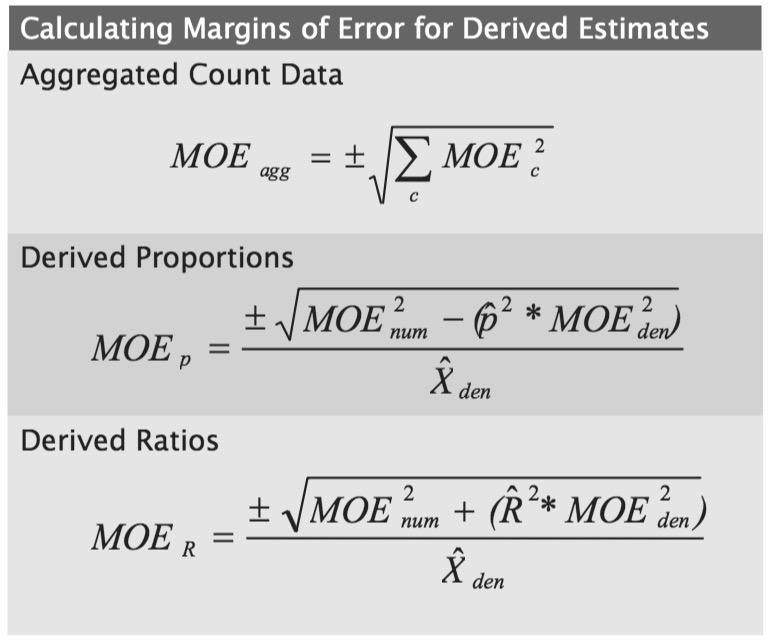


Table 2. Formula of Calculating Margins of Error for Derived Estimates   
(US Census Bureau, 2009)

ACS summary data is shown as aggregate data, so cannot be used it to check the correlation among variables at an individual level. The problem of using aggregate data to calculate correlation is ecological fallacy, which means that data about a group is used to conclude information about an individual. Therefore, it is inappropriate to assume that relationships at the aggregate level will also hold at the individual level. We need data from a single source to assess correlation among different variables. Therefore, we decided to use the ACS Public Use Microdata Sample to check the correlation of each variable from a single source.

The ACS Public Use Microdata Sample, or PUMS, is one of the ACS data products provided by the U.S Census Bureau. It is a sample of actual responses of the population and housing unit records from the American Community Survey. Different from ACS summary products that show data that have already been tabulated for specific geographic areas, PUMS files are a set of untabulated records showing information about individual persons or housing units for specific geographic areas. It includes individual response from self-reported data such as relationship, sex educational attainment, and employment status.

However, as required by federal law, the confidentiality of the ACS respondents is protected through a variety of means to ensure that it is impossible to identify the responses from any individuals. Therefore, some records are removed or switched by similar records from the nearby area to reduce the ability to identify individuals from their response. Because of this, PUMS data will not be as accurate as the ACS summary data, even after summarizing all of the PUMS records. The Census Bureau also protects confidentiality by limiting the geographic area code to only regions, divisions, states and Public Use Microdata Area (PUMAs). Different from ACS summary data, the smallest geographic level for PUMS data is PUMAs, which are defined to represent geographic areas with populations of at least 100,000. Therefore, PUMS data cannot satisfy this study’s need to compare demographics at the census tract level but allow us to see whether the underlying variables are correlated for individuals within a PUMA.

**Correlation Matrix Among Variables**

For the above reasons, PUMS cannot be used as the source data for our social vulnerability index. However, it is still helpful to use PUMS records to check the correlation between each of our variables to consider which might be eliminated. Correlation describes the degree of relationship between two variables. A correlation matrix is used to investigate the dependence between multiple variables at the same time by showing the correlation coefficients between each variable and others. There are different methods for correlation analysis. This study uses the Pearson Parametric Correlation Test to assess the relationship among 13 variables.

Table 3 shows the correlation matrix of PUMS person data of Michigan. It presents the value for Pearson’s r, a Sig. (2-tailed) value and a number (N) value. Pearson’s r shows the correlation between each variable. When Pearson’s r is close to 1, means that there is a strong relationship between two variables. When the value is close to 0, there is weak relationship between the two variables. When the value is positive, means that as one variable increases, the second variable also increase. When the value is negative, means that as one variable increases, the second variable decreases. Sig. (2-tailed) shows if there is a statistically significant correlation between two variables. If Sig. (2-tailed) value is greater than .05 means that there is no statistically significant correlation between two variables. In contrast, if the Sig. (2-tailed) value is less than or equal to .05 means that there is a statistically significant correlation between two variables.

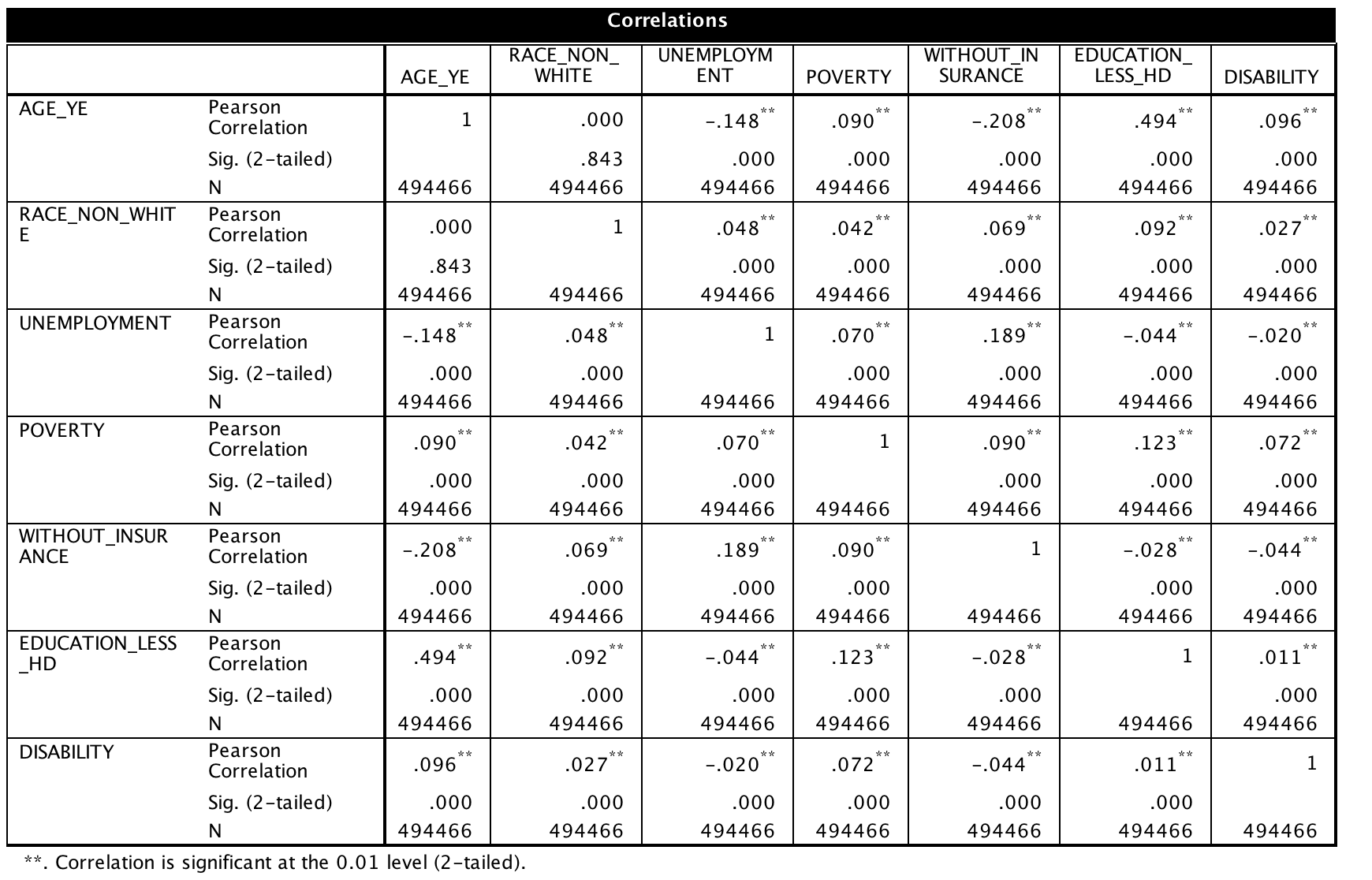


Table 3. Correlation Matrix of Person data

Overall, there is no significant correlation among these variables from person record. Only populations with less than high school diploma and populations aged less than 18 and over 65 years old have the highest positive correlation of 0.494 (correct?). It is because that people need more time to achieve to a higher education status. With above result, we decide keep all of the variables in our final index.

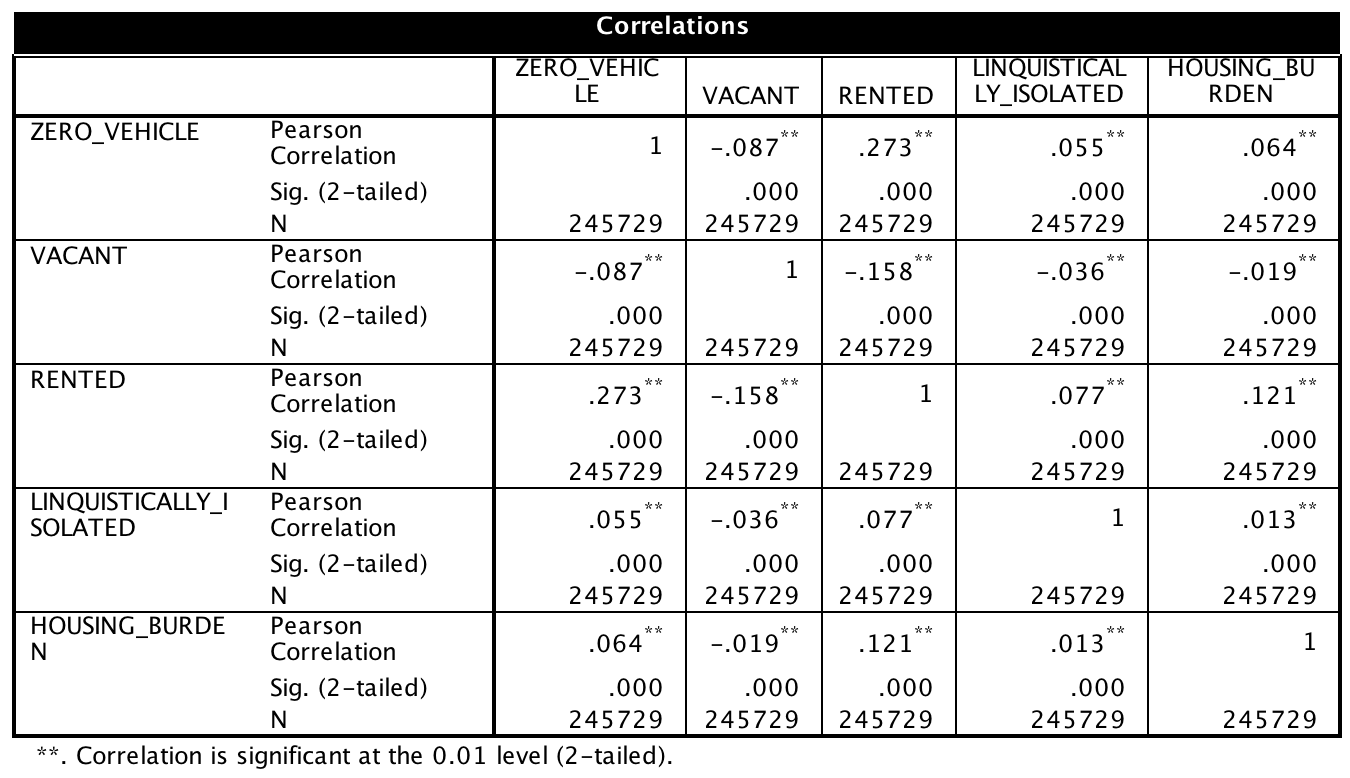


Table 4. Correlation Matrix of Housing data

Table 4 shows the correlation matrix of housing data from PUMS record. Similar to the person data, there is no significant correlation among these variables. Therefore, we did not eliminate any variables from the housing section.

**Technical Documentation**

Data Sources: American Community Survey

Year: 2010-2014

Geographic level: Census Tract (by State--County)

Dataset: 2014 American Community Survey: 5-Year Data (2010-2014, Block Groups & Larger Areas)

Data type(s): (E) Estimates and (M) Margins of error

Source: American FactFinder  
<http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?ref=geo&refresh=t&tab=map&src=bkmk>

Year: 2010

Geographic level: Census Tract (by State--County)

Dataset: 2010 USDA Food Access Research Atlas Data

Source: United States Department of Agriculture Economic Research Service

<http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx>

Variable Construction

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Human Capital** | **Description** | **ACS Source Code** | **Variables** | **Function** |
| Demographic | Race | B02001 | % non-white | (total pop – white alone) / total pop |
|  | Age | B01001 | % under 17 + % over 65 | (under 17 + over 65)  / total pop |
| Social / Economic | Employment | B23025 | % unemployment | In labor force: Civilian labor force: Unemployed / In labor force: Civilian labor force |
|  | Vehicle Occupancy | B25044 | % with 0 vehicle | (Owner occupied no vehicle + renter occupied no vehicle) / total housing units |
| Wealth | Poverty | C17002 | % poverty\* | Under 0.50 + 0.50 to 0.99 / total pop |
| Health | Insurance | B27010 | % without insurance | (Under 18 with no insurance + 18 to 34 with no insurance + 35 to 64 with no insurance + over 65 with no insurance) / total pop |
|  | Food access\* |  | food access  (1 or 0) | Low access tract at 1 mile for urban areas or 10 miles for rural areas |
| Education | Education | B15003 | % less than high school diploma | Sum(no schooling completed:12th grade, no diploma) / total pop |
| Housing | Occupied Housing | B25002 | % Vacant | Vacant / total housing units |
|  |  | B25003 | % Rented | Renter occupied / total housing units |
|  | Housing  cost burden | B25070  (Renter occupied)  B25091 (Owner occupied) | % Household pay more than 30% of their income | ((30.0 to 34.9 percent + 35.0 to 39.9 percent + 40.0 to 49.9 percent + 50.0 percent or more) / total renter-occupied housing units)  +  ((Housing units with a mortgage: 30.0 to 34.9 percent + Housing units with a mortgage: 35.0 to 39.9 percent + Housing units with a mortgage: 40.0 to 49.9 percent + Housing units with a mortgage: 50.0 percent or more + Housing units without a mortgage: 30.0 to 34.9 percent + Housing units without a mortgage: 35.0 to 39.9 percent + Housing units without a mortgage: 40.0 to 49.9 percent + Housing units without a mortgage: 50.0 percent or more) / total owner-occupied housing units) |
| Language | Linguistically isolated Household | B16002 | % limited English speaking household\* | (Spanish: Limited English speaking household + Other Indo-European languages: Limited English speaking household + Asian and Pacific Island languages: Limited English speaking + Other languages: Limited English speaking household) / total households |
| Disability | Disability | B18101 | % people more than 5 years old with a disability | Total disability / total pop |

Data Sources: Public Use Microdata Sample

Year: 2010-2014

Geographic level: PUMA

Dataset: 2010 – 2014 ACS PUMS DATA

Source: American FactFinder

<http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>

Variable Construction

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Human Capital** | **Description** | **PUMS Record** | **Variables** | **Detail** | **Command to dichotomous variables** |
| Demographic | Race | Person –  RAC1P | % non-white | Recoded detailed race code  1. White alone 2. Black or African American alone  3. American Indian alone  4. Alaska Native alone  5. American Indian and Alaska Native tribes specified; or American Indian or Alaska Native, not specified and no other races  6. Asian alone  7. Native Hawaiian and Other Pacific Islander alone  8. Some Other Race alone  9. Two or More Races | recode RAC1P  (1 = 0)  (else = 1) |
|  | Age | Person –  AGEP | % under 17 + % over 65 | 00. Under 1 year  01..99 .1 to 99 years (Top-coded\*\*\*) | recode AGEP  (00 = 1)  (0 thru 17 = 1)  (65 thru 99 = 1)  (else = 0) |
| Social / Economic | Employment | Person -  ESR | % unemployment | Employment status recode  b. N/A (less than 16 years old)  1. Civilian employed, at work  2. Civilian employed, with a job but not at work  3. Unemployed  4. Armed forces, at work  5. Armed forces, with a job but not at work  6. Not in labor force | recode ESR  (3 = 1)  (else = 0) |
|  | Vehicle Occupancy | Housing - VEH | % with 0 vehicle | Vehicles (1 ton or less) available  b. N/A (GQ/vacant)  0. No vehicles  1. 1 vehicle  2. 2 vehicles  3. 3 vehicles  4. 4 vehicles  5. 5 vehicles  6. 6 or more vehicles | recode VEH  (0 = 1)  (else = 0) |
| Wealth | Poverty | Person –  POVPIP | % poverty\* | Income-to-poverty ratio recode  bbb. N/A  000..500. Below 501 percent  501 .501 percent or more | recode POVPIP  (000 thru 500 = 1)  (else = 0) |
| Health | Insurance | Person -  HICOV | % without insurance | Health insurance coverage recode  1. With health insurance coverage  2. No health insurance coverage | recode HICOV  (2 = 1)  (1 = 0) |
|  | Food access\* |  | % low food access |  |  |
| Education | Education | Person –  SCHL | % less than high school diploma | Educational attainment  bb. N/A (less than 3 years old)  01. No schooling completed  02. Nursery school, preschool  03. Kindergarten  04. Grade 1  05. Grade 2  06. Grade 3  07. Grade 4  08. Grade 5  09. Grade 6  10. Grade 7  11. Grade 8  12. Grade 9  13. Grade 10  14. Grade 11  15. 12th grade - no diploma  16. Regular high school diploma  17. GED or alternative credential  18. Some college, but less than 1 year  19. 1 or more years of college credit, no degree  20. Associate's degree  21. Bachelor's degree  22. Master's degree  23. Professional degree beyond a bachelor's degree  24. Doctorate degree | recode SCHL  (01 thru 15 = 1)  (else = 0) |
| Housing | Occupied Housing | Housing –  NP | % Vacant | Number of person records following this housing record  00. Vacant unit  01. One person record (one person in household or any person in   group quarters)  02..20. Number of person records (number of persons in household) | recode NP  (00 = 1)  (else = 0) |
|  |  | Housing –  TEN | % Rented | Tenure  b. N/A (GQ/vacant)  1. Owned with mortgage or loan (include home equity loans)  2. Owned free and clear  3. Rented  4. Occupied without payment of rent | recode TEN  (3 = 1)  (else = 0) |
|  | Housing cost burden | Housing –  RNTP / HINCP | % Household pay more than 30% of their income | RNTP:  Monthly rent  bbbbb. N/A (GQ/not a rental unit/occupied without rent .payment)  00001..99999 .$1 to $99999 (Rounded and top-coded)  HINCP: Household income (past 12 months)  bbbbbbbb. N/A(GQ/vacant)  00000000. No household income  -0059999. Loss of $59,999 or more  -0059998..-0000001. Loss of $1 to $59,998  00000001. $1 or Break even  00000002..99999999. Total household income in dollars (Components are rounded) | HCB (Housing cost burden) RNTP / HINCP  recode HCB  (0.3 thru highest = 1)  (else = 0) |
| Language | Linguistically isolated Household | Housing –  LNGI | % limited English speaking household\* | Limited English speaking household  b. N/A (GQ/vacant)  1. At least one person in the household 14 and over speaks English only or speaks English 'very well’  2. No one in the household 14 and over speaks English only or speaks English 'very well' | recode LNGI  (2 = 1)  (else = 0) |
| Disability | Disability | Person –  DIS | % people more than 5 years old with a disability | Disability recode  1. With a disability  2. Without a disability | Recode DIS  (1 = 1)  (2 = 0) |

\*Ratio of income to poverty

People and families are classified as being in poverty if their income is less than their poverty threshold. If their income is less than half their poverty threshold, they are below 50% of poverty; less than the threshold itself, they are in poverty (below 100% of poverty); less than 1.25 times the threshold, below 125% of poverty, and so on. The greater the ratio of income to poverty, the more people fall under the category, because higher ratios include more people with higher incomes. Families and individuals with an income-to-poverty ratio of less than 100 percent are identified as in poverty.

\*Low food access at 1 mile for urban areas or 10 miles for rural areas   
A low-income tract with at least 500 people or 33 percent of the population living

more than 1 mile (urban areas) or more than 10 miles (rural areas) from the nearest supermarket, supercenter, or large grocery store.

\*A "limited English speaking household" is one in which no member 14 years old and over (1) speaks only English or (2) speaks a non-English language and speaks English "very well." In other words, all members 14 years old and over have at least some difficulty with English. By definition, English-only households cannot belong to this group. Previous Census Bureau data products have referred to these households as "linguistically isolated" and "Household where no one age 14 and over speaks English only or speaks English ‘very well.’"

**R Code and Instructions**

*These instructions are under development. We also plan to provide instructions utilizing RStudio for professionals intimidated by running code directly.*

**Part 1: Data Preparation**

**Step 1: Downloading the data**

1. Downloading the following ACS tables from American Factfinder:

B02001, B01001, B23025, B25044, C17002, B27010, B15003, B25002, B25003, B25070,

B25091, B16002, B18101

Depending on your project’s need, choose the appropriate dataset (eg. ACS 2010-2014 5-year estimates). These tables provide the data needed for calculating all indicators, except the food access indicator, required to compute our social vulnerability index score.

1. Downloading the “Food Access Research Atlas Data File.xls” Excel file from USDA’s website (<http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx>), in which it contains the data for the food access indicator.

Note the file downloaded from the website page as of May 13, 2016 is an Excel file named “DataDownload.xls” If it has a different file name, please change it to “DataDownload” to make it comparable with the R codes. Also, since R can only read CSV files, please convert the third spreadsheet (the “Food Access Research Atlas Data”) of this “DataDownload.xls” Excel file into a CSV file, and name it “Food Access Research Atlas Data.csv”

**Step 2: Set up the work space**

1. Create a folder in your desktop, name it “Social Vulnerability”, and extract the data downloaded into this folder.

[*Note: You are free to create the folder wherever you want and name it whatever your like, but make sure the record the folder name and its path correct. For example, Jane may create a folder in her laptop’s D Drive and name it “SocialVul”, and in this case the path for this folder is D:\SocialVul*]

2. Copy the “Food Access Research Atlas Data.csv” CSV file into this folder.

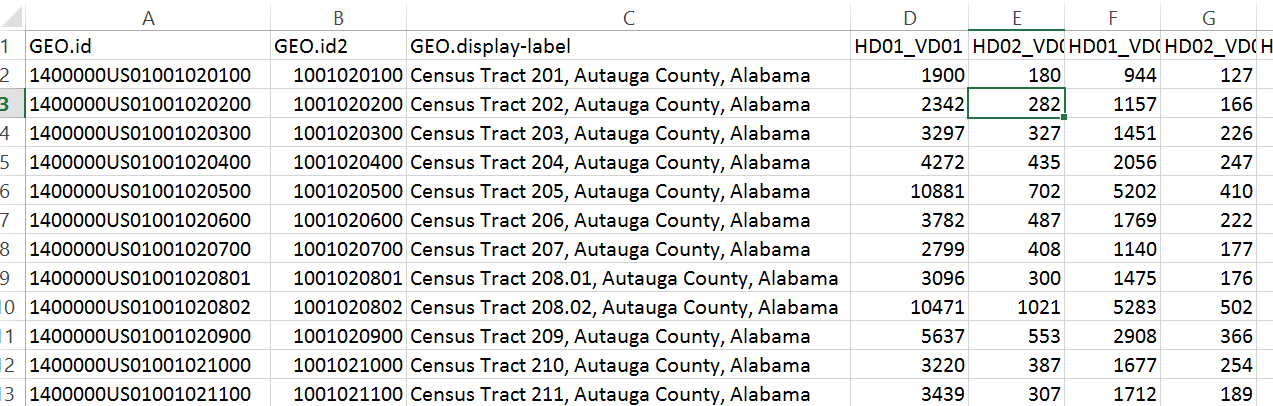
3. Exact all the ACS files into this folder. Open each of ACS tables (they should be CSV files with names like “ACS\_XX\_XXX\_XXXXX\_with\_ann.csv”) and delete the second row (the row starting with Id, Id2…).

**Step 3: Processing the Food Access Data**

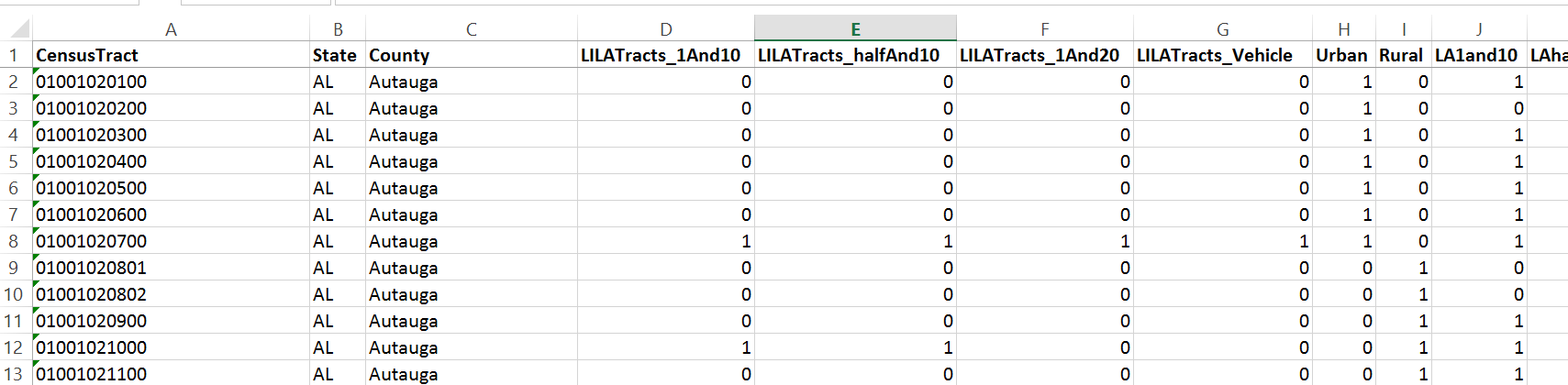
Since the Food Access Research Atlas Data file downloaded from USDA contains all US Census Tracts, (unless your project’s scale is the whole US) you need to reduce it to a file that contains only census tracts in your working area. You could do this either by using Microsoft Access’s Join function or simply through sorting in Excel.

When you have got rid of other irrelevant census tracts, make sure to sort the remaining census tracts the same as the ACS tables do.

For example, my ACS\_14\_5YR\_B01001\_with\_ann.csv and the “Food Access Research Atlas Data.csv” look like following after above steps (First row contains the headings of each table, and from the second row each table has the same census tract).



(ACS\_14\_5YR\_B01001\_with\_ann.csv)



(Food Access Research Atlas Data.csv)

**Part 2: Getting ready for running the R code**

If R is not installed on your computer yet, download R from the r-project website and install it on your computer.

[Note: you need to download the appropriate version for your operation system]

Create a CSV file and name it “Output.csv” (This is the table where the results will be stored). Copying the first row from any of the “ACS\_XX\_XXX\_XXXXX\_with\_ann.csv” files to be the first column of this file. Close the file and get ready to run the R codes!

**Part 3: Running the R code to calculate the Social Vulnerability Index Score**

#Set up the workspace (replace "jacobyan" below with your user name.)

setwd("C:\\Users\\Jacob Yan\\Desktop\\Social Vunerability")

#Loading Tables downloaded from AmericanFactfinder into R

B02001<-read.csv("ACS\_14\_5YR\_B02001\_with\_ann.csv")

B01001<-read.csv("ACS\_14\_5YR\_B01001\_with\_ann.csv")

B23025<-read.csv("ACS\_14\_5YR\_B23025\_with\_ann.csv")

B25044<-read.csv("ACS\_14\_5YR\_B25044\_with\_ann.csv")

C17002<-read.csv("ACS\_14\_5YR\_C17002\_with\_ann.csv")

B27010<-read.csv("ACS\_14\_5YR\_B27010\_with\_ann.csv")

B15003<-read.csv("ACS\_14\_5YR\_B15003\_with\_ann.csv")

B25002<-read.csv("ACS\_14\_5YR\_B25002\_with\_ann.csv")

B25003<-read.csv("ACS\_14\_5YR\_B25003\_with\_ann.csv")

B25070<-read.csv("ACS\_14\_5YR\_B25070\_with\_ann.csv")

B25091<-read.csv("ACS\_14\_5YR\_B25091\_with\_ann.csv")

B16002<-read.csv("ACS\_14\_5YR\_B16002\_with\_ann.csv")

B18101<-read.csv("ACS\_14\_5YR\_B18101\_with\_ann.csv")

FoodAcc <- read.csv("")

#Calculate the indicators

#Indicator 1: % non-white, noted as NonWhite

NonWhite <- (B02001$HD01\_VD01 - B02001$HD01\_VD02) / B02001$HD01\_VD01

#Indicator 2: % under 17 + % over 65 , noted as Under17\_Over65

Under17\_Over65 <- (B01001$HD01\_VD03 + B01001$HD01\_VD04 + B01001$HD01\_VD05 + B01001$HD01\_VD06 + B01001$HD01\_VD20 + B01001$HD01\_VD21 +

B01001$HD01\_VD22 + B01001$HD01\_VD23 + B01001$HD01\_VD24 + B01001$HD01\_VD25 + B01001$HD01\_VD27 + B01001$HD01\_VD28 +

B01001$HD01\_VD29 + B01001$HD01\_VD30 + B01001$HD01\_VD44 + B01001$HD01\_VD45 + B01001$HD01\_VD46 + B01001$HD01\_VD47 +

B01001$HD01\_VD48 + B01001$HD01\_VD49) / B01001$HD01\_VD01

#Indicator 3: % unemployment , noted as Unemployed

Unemployed <- B23025$HD01\_VD05 / B23025$HD01\_VD03

#Indicator 4: % with zero vehicle, noted as NoVehicle

NoVehicle <- (B25044$HD01\_VD03 + B25044$HD01\_VD10) / B25044$HD01\_VD01

#Indicator 5: % Population in poverty , noted as Poverty

Poverty <- (C17002$HD01\_VD02 + C17002$HD01\_VD03) / C17002$HD01\_VD01

#Indicator 6: % without health insurance , noted as NoHealthIns

NoHealthIns <- (B27010$HD01\_VD17 + B27010$HD01\_VD33 + B27010$HD01\_VD50 + B27010$HD01\_VD66) / B27010$HD01\_VD01

#Indicator 7: % less than high school diploma , noted as LessHighSch

LessHighSch <-(B15003$HD01\_VD02 + B15003$HD01\_VD03 + B15003$HD01\_VD04 + B15003$HD01\_VD05 + B15003$HD01\_VD06 + B15003$HD01\_VD07 +

B15003$HD01\_VD08 + B15003$HD01\_VD09 + B15003$HD01\_VD10 + B15003$HD01\_VD011 + B15003$HD01\_VD12 + B15003$HD01\_VD13 + B15003$HD01\_VD14 +

B15003$HD01\_VD15 + B15003$HD01\_VD16) / B15003$HD01\_VD01

#Indicator 8: % Vacant , noted as Vacant

Vacant <- B25002$HD01\_VD03 / B25002$HD01\_VD01

#Indicator 9: % Rental, noted as Rental

Rental <- B25003$HD01\_VD03 / B25003$HD01\_VD01

#Indicator 10: % Renter households paying more than 30% of their income, noted as Rental30Inc

Renter30Inc <-((B25070$HD02\_VD07 + B25070$HD02\_VD08 + B25070$HD02\_VD09 + B25070$HD02\_VD10) / B25070$HD01\_VD01)

#Indicator 11: % Owner households paying more than 30% of their income, noted as Owner30Inc

Owner30Inc <-((B25091$HD01\_VD08 + B25091$HD01\_VD09 + B25091$HD01\_VD10 + B25091$HD01\_VD11 + B25091$HD01\_VD19 + B25091$HD01\_VD20 +

B25091$HD01\_VD21 + B25091$HD01\_VD22) / B25091$HD01\_VD01)

#Indicator 12: % limited English Speaking Households, noted as LimEnglish

LimEnglish <-(B16002$HD01\_VD04 + B16002$HD01\_VD07 + B16002$HD01\_VD10 + B16002$HD01\_VD13) / B16002$HD01\_VD01

#Indicator 13: % pple more than 5 years old with a disability, noted as Disab\_Over5

Disab\_Over5 <-(B18101$HD01\_VD04 + B18101$HD01\_VD07 + B18101$HD01\_VD10 + B18101$HD01\_VD13 + B18101$HD01\_VD16 + B18101$HD01\_VD19 + B18101$HD01\_VD23 +

B18101$HD01\_VD26 + B18101$HD01\_VD29 + B18101$HD01\_VD32 + B18101$HD01\_VD35 + B18101$HD01\_VD38) / B18101$HD01\_VD01

#Indicator 14: Food Access,

FoodAcc

#Converting the values of indicators into Z-scores

Z\_NonWhite <- (NonWhite - mean(NonWhite))/sd(NonWhite)

Z\_Under17\_Over65 <- (Under17\_Over65 - mean(Under17\_Over65))/sd(Under17\_Over65)

Z\_Unemployed <- (Unemployed - mean(Unemployed))/sd(Unemployed)

Z\_NoVehicle <- (NoVehicle - mean(NoVehicle))/sd(NoVehicle)

Z\_Poverty <- (Poverty - mean(Poverty))/sd(Poverty)

Z\_NoHealthIns <- (NoHealthIns - mean(NoHealthIns))/sd(NoHealthIns)

Z\_LessHighSch <- (LessHighSch - mean(LessHighSch))/sd(LessHighSch)

Z\_Vacant <- (Vacant - mean(Vacant))/sd(Vacant)

Z\_Rental <- (Rental - mean(Rental))/sd(Rental)

Z\_Renter30Inc <- (Renter30Inc - mean(Renter30Inc))/sd(Renter30Inc)

Z\_Owner30Inc <- (Owner30Inc - mean(Owner30Inc))/sd(Owner30Inc)

Z\_LimEnglish <- (LimEnglish - mean(LimEnglish))/sd(LimEnglish)

Z\_Disab\_Over5 <- (Disab\_Over5 - mean(Disab\_Over5))/sd(Disab\_Over5)

#Calculate the Vunerability Index Score

[Weighting : two options. Option 1: assigning weight by changing the formula below]

Vul\_Score <- Z\_NonWhite + Z\_Under17\_Over65 + Z\_Unemployed + Z\_NoVehicle + Z\_Poverty + Z\_NoHealthIns + Z\_LessHighSch

+ Z\_Vacant + Z\_Rental + Z\_Renter30Inc + Z\_Owner30Inc + Z\_LimEnglish + Z\_Disab\_Over5

#Writing the intermediate results (indicators) and final result (vulnerability score) into an output table

Output <- read.csv("Output.csv")

Output$GEO.id <- B02001$GEO.id

Output$Z\_NonWhite <- Z\_NonWhite

Output$Z\_Under17\_Over65 <- Z\_Under17\_Over65

Output$Z\_Unemployed <- Z\_Unemployed

Output$Z\_NoVehicle <- Z\_NoVehicle

Output$Z\_Poverty <- Z\_Poverty

Output$Z\_NoHealthIns <- Z\_NoHealthIns

Output$Z\_LessHighSch <- Z\_LessHighSch

Output$Z\_Vacant <- Z\_Vacant

Output$Z\_Rental <- Z\_Rental

Output$Z\_Renter30Inc <- Z\_Renter30Inc

Output$Z\_Owner30Inc <- Z\_Owner30Inc

Output$Z\_LimEnglish <- Z\_LimEnglish

Output$Z\_Disab\_Over5 <- Z\_Disab\_Over5

write.csv(Output,"Result.csv")

**Works Cited**

Adger, W. Neil. "Vulnerability." *Global environmental change* 16, no. 3 (2006): 268-281.

Beard, John R. Beard, Magda Cerda, Shannon Blaney, Jennifer Ahern, David Vlahov, and Sandro Galea. 2009. “Neighborhood Characteristics and Change in Depressive Symptoms Among Older Residents of New York City.” *American Journal of Public Health* 99(7):1308-1314.

Berke, Ethan M., Laura M. Gottlieb, Anne Vernez Moudon and Eric B. Larson. 2007. “Protective Association Between Neighborhood Walkability and Depression in Older Men.” *J Am Geriatr Soc* 55:526–533. doi: 10.1111/j.1532-5415.2007.01108.x

Birkmann, Jorn. "Measuring vulnerability to promote disaster-resilient societies: Conceptual frameworks and definitions." In: *Measuring vulnerability to natural hazards: Towards disaster resilient societies* (2006): 9-54.

Brooks, Nick, W. Neil Adger, and P. Mick Kelly. "The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation." *Global environmental change* 15, no. 2 (2005): 151-163.

Brown, Scott C., Craig A. Mason, Joanna L. Lombard, Frank Martinez, Elizabeth Plater-Zyberk. “The Relationship of Built Environment to Perceived Social Support and Psychological Distress in Hispanic Elders: The Role of ‘Eyes on the Street.’” *Journal of Gerontology: Social Sciences* 64B(2): 234–246, doi:10.1093/geronb/gbn011.

Burton, Elizabeth. 2003. "Housing for an Urban Renaissance: Implications for Social Equity." *Housing Studies* 18 (4):537-562. doi: 10.1080/02673030304249.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2015. “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.” Working Paper. Harvard University and NBER.

Cohen, Deborah A., Sanae Inagami, and Brian Finch. 2007. “The built environment and collective efficacy.” *Health and Place* 14 (2): 198-208. doi:10.1016/j.healthplace.2007.06.001.

*Cohen, S. B., and W. Yu. "Agency for Healthcare Research and Quality; 2006." The Persistence in the Level of Health Care Expenditures over Time: estimates for the US Population 2003 (2002).*

Cutter, Susan. “Vulnerability to Environmental Hazards.” Progress in Human Geography 20 (1996): 529–39.

Cutter, Susan L., Jerry T. Mitchell, and Michael S. Scott. "Revealing the vulnerability of people and places: a case study of Georgetown County, South Carolina." *Annals of the Association of American Geographers* 90, no. 4 (2000): 713-737.

Cutter, Susan L., Bryan J. Boruff, and W. Lynn Shirley. "Social vulnerability to environmental hazards." *Social science quarterly* 84, no. 2 (2003): 242-261.

Cutter, Susan L., and Christina Finch. "Temporal and spatial changes in social vulnerability to natural hazards." *Proceedings of the National Academy of Sciences* 105, no. 7 (2008): 2301-2306.

Eicher, Caitlin and Ichiro Kawachi. 2011. “Social Capital and Community Design.” *Making Healthy Places.* Island Press/Center for Resource Economic. doi: 10.5822/978-1-61091-036-1\_8.

Ewing, Reid, and Shima Hamidi. 2014. *Measuring Urban Sprawl and Validating Sprawl Measure*s. Research Report. Metropolitan Research Center, University of Utah.

Ewing, Reid, Gail Meakins, Shima Hamidi, and Arthur C. Nelson. 2014. “Relationship between urban sprawl and physical activity, obesity and morbidity – Update and refinement.” *Health & Place* 26: 118-126. Doi: Ltd. <http://dx.doi.org/10.1016/j.healthplace.2013.12.008>.

Few, Roger, Katrina Brown, and Emma L. Tompkins. "Public participation and climate change adaptation: avoiding the illusion of inclusion." *Climate policy* 7, no. 1 (2007): 46-59.

Galea, Sandro, Jennifer Ahern, Sasha Rudenstine, Zachary Wallace, and David Vlahov. 2005. “Urban built environment and depression: a multilevel analysis.” *J Epidemiol Community Health* 59(10):822–827. doi: 10.1136/jech.2005.033084.

Gordon-Larsen, Penny, Melissa C. Nelson, Phil Page, Barry M. Popkin. 2006. “Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity.” *Pediatrics* 117(2): 417-424. doi: 10.1542/peds.2005-0058.

Grafova, Irina. 2008. “Overweight children: Assessing the contribution of the built environment.” *Preventative Medicine* 4 (3): 304-308. doi: 10.1016/j.ypmed.2008.04.012.

Jabareen, Yosef. "Planning the resilient city: Concepts and strategies for coping with climate change and environmental risk." *Cities* 31 (2013): 220-229.

Khan, Shabana. "Vulnerability assessments and their planning implications: a case study of the Hutt Valley, New Zealand." *Natural hazards* 64, no. 2 (2012): 1587-1607.

Kwan, Mei-Po. 2012. "The uncertain geographic context problem." *Annals of the Association of American Geographers* 102 (5):958-968.

Li, Fuzhong, K John Fisher, Ross C. Brownson, and Mark Bosworth. 2005. “Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults.” *J Epidemiol Community Health* 59:558–564. doi: 10.1136/jech.2004.028399.

Li, Fuzhong, Peter A. Harmer, Bradley J. Cardinal, Mark Bosworth, Alan Acock, Deborah Johnson-Shelton, and Jane M. Moore. 2008. “Built Environment, Adiposity, and Physical Activity in Adults Aged 50–75.” *Am J Prev Med* 35(1):38-46. doi:10.1016/j.amepre.2008.03.021

Lorenc, Theo, Mark Petticrew, Margaret Whitehead, David Neary, Stephen Clayton, Kath Wright, Hilary Thomson, Steven Cummins, Amanda Sowden and Adrian Renton. 2014. “Crime, fear of crime and mental health: synthesis of theory and systematic reviews of intervention and qualitative evidence.” *Public Health Research* 2 (2). doi: 10.3310/phr02020.

Lovasi, Gina S., Malo A. Hutson, Monica Guerra and Kathryn M. Neckerman. 2009. “Built Environments and Obesity in Disadvantaged Populations.” *Epidemiologic Reviews* 31 (1): 7-20. doi: 10.1093/epirev/mxp005.

Lund, H. 2003. “Testing the Claims of New Urbanism—Local Access, Pedestrian Travel, and Neighboring Behaviors.” *Journal of the American Planning Association* 69 (4): 414–29.

Mair, Christina F, Ana V Diez Roux, and Sandro Galea. 2008. "Are Neighborhood Characteristics Associated with Depressive Symptoms? A Critical Review." *Journal of Epidemiology and Community Health*. doi: 10.1136/jech.2007.066605.

Ojerio, Ryan, Cassandra Moseley, Kathy Lynn, and Neil Bania. "Limited involvement of socially vulnerable populations in federal programs to mitigate wildfire risk in Arizona." *Natural Hazards* Review (2010).

Papas, Mia A., Anthony J. Alberg, Reid Ewing, Kathy J. Helzlsouer, Tiffany L. Gary, and Ann C. Klassen. 2007. “The Built Environment and Obesity.” *Epidemiologic Reviews* 29 (1):129-143.

Patterson, Patricia and Nancy J. Chapman. 2004. “Urban Form and Older Residents’ Service Use, Walking, Driving, Quality of Life, and Neighborhood Satisfaction.” *American Journal of Health Promotion* 19(1):45-52.

Renalds, Arlene, Tracey H. Smith, and Patty J. Hale. 2010. “A Systematic Review of Built Environment and Health.” *Fam Community Health* 33 (1): 68-78.

Rundle, Andrew, Ana V. Diez Roux, Lance M. Freeman, Douglas Miller, Kathryn M. Neckerman, Christopher Weiss. 2007. “The Urban Built Environment and Obesity in New York City: A Multilevel Analysis.” *American Journal of Health Promotion* 21 (4): 326.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "Neighborhood Effects": Social Processes and New Directions in Research." *Annual Review of Sociology* 28:443-478. doi: 10.2307/3069249.

Smith, William R, Sharon Glave Frazee, and Elizabeth L Davison. 2000. "Furthering the integration of routine activity and social disorganization theories: Small units of analysis and the study of street robbery as a diffusion process." *Criminology* 38 (2):489-524.

Turner, Billie L., Roger E. Kasperson, Pamela A. Matson, James J. McCarthy, Robert W. Corell, Lindsey Christensen, Noelle Eckley et al. "A framework for vulnerability analysis in sustainability science." *Proceedings of the national academy of sciences* 100, no. 14 (2003): 8074-8079.

US Census Bureau. "A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know." (2009).

Vogel, Coleen, Susanne C. Moser, Roger E. Kasperson, and Geoffrey D. Dabelko. "Linking vulnerability, adaptation, and resilience science to practice: Pathways, players, and partnerships." *Global environmental change* 17, no. 3 (2007): 349-364.