Calculating Distressed Communities Index

The Economic Innovation Group calculates the Distressed Communities Index (DCI) from U.S. Census Bureau data to compare the economic well-being throughout the United States. EIG has created DCI at several grouping levels: zip code, city, county, and congressional district. In June 2018, the City of San Antonio asked the students of the UTSA Data Analytics program to assist in interpreting the DCI.

This document concerns that project’s first step, which is to replicate the work that EIG did to create the DCI at the zip code level. The purpose of this first step is (1) to understand in detail EIG’s methodology, (2) to create a data set for analysis, and (3) to confirm the accuracy of EIG’s published results.

The DCI is based on three surveys by the U.S. Census Bureau:

* American Community Survey 5-Year Estimates for 2011-2015
* Business Patterns of 2011
* Business Patterns of 2015

The American Community Survey is not an enumeration (a complete count like the decennial census); it uses statistical sampling and provides estimates with margins of error. The Business Patterns are based on mandatory reporting to government agencies.

There are seven components of the DCI:

* Percent of the population 25 years and older without a high school diploma or equivalent
* Percent of habitable housing that is unoccupied but is not for seasonal, recreational, or occasional use
* Percent of the population 25 – 64 years old that is unemployed
* Percent of the population living under the poverty line
* The ratio of the median income of the zip code to the median income of the state
* The percent change in the number of jobs from 2011 to 2015
* The percent change in the number of businesses from 2011 to 2015

Here is our methodology:

1. For each of the seven DCI components:

* Download the component’s raw value from the U.S. Census Bureau, or download multiple statistics and use them to calculate the component’s raw value
* Sort the zip codes by the raw value of the component
* Rank the zip codes, with “1” being the zip code with the most desirable raw value

1. Afterwards:

* Average the DCI component ranks for each zip code
* Sort the zip codes by their averages
* Rank the zip codes, with “1” being the lowest average
* With N equal to the number of zip codes, divide each rank by N, multiply by 100, and round to one decimal place, to calculate the DCI percentile

The American Community Survey includes 33,120 zip codes and the Business Patterns each include over 38,000 zip codes. EIG reports on 26,126 zip codes. UTSA created a data set with 26,125 zip codes as follows:

* Eliminate all zip codes with populations under 500 people
* Eliminate all zip codes that have missing data for six of the seven DCI components:
  + No high school diploma
  + Housing vacancy rate
  + Adults not working
  + Poverty rate
  + Median income ratio
  + ~~Change in employment~~
  + Change in businesses
* If a zip code does not have a value for “Change in employment,” set it to the “Change in businesses” for that zip code

Our spot checks have compared the raw values and rankings of hundreds of zip codes. Our numbers match EIG numbers with a few exceptions. Our “Adults not working” percentages vary from EIG’s by up to 0.1 (for example, we might have 10.1 and EIG has 10.2). Two out of 243 zip codes have completely different values for change in employment. Dr. Campbell informed us that the U.S. Census Bureau sometimes discovers and fixes errors in its data set. If EIG downloaded data in 2017 and we downloaded data in 2018, our source data may be different.