The Distressed Community Index is based on how each zip code compares to every other zip code in America. If everybody were suddenly twice as rich, there would still be a bottom 20% who were categorized as “distressed.”

This means we cannot use a formula on the raw data (e.g. “percent without HS diploma”) for one zip code at a time. For each of seven metrics, we will:

1. Create a three variable data set with zip code, metric (e.g. “percent without HS diploma”), and position.
2. Sort the data set by metric. The sort is ascending or descending to put the best value of the metric in the first position.
3. Set the position values with 1 in the first row, 2, in the second row, … 26000 in the 26000th row.
4. Sort the data set by zip code.

Once we have seven data sets:

1. Copy columns into one data set of zip code and seven metrics.
2. Copy columns into one data set of zip code and seven positions.

In the positions data set:

1. Create a column with the mean of the seven positions.
2. Sort by the new column.
3. Create a column with a new position number, 1 to 26000.
4. Create a column for percentile, dividing position by 260 (or 1/100th of the number of rows).
5. Sort by zip code.

This is a crude algorithm, but if we save the results to a csv file, we only have to run it once. It could take longer to invent, write, test, and debug something more efficient.