

Socio-Economic “Factors” and Combination Methodology

January 9, 2004 Bruce Wicinas

In every trial the lottery is run on applicants divided into three classifications. In the current system the three classifications are racial. In the trial system the three classifications are socio-economic categories. Socio-economic classification is via a map, *not* via personal data. The socio-economic maps described herein is the latest in an evolutionary process of experiment aimed at deriving a most effective map from census and population data. The “weights” for combining data and the thresholds that divide three gradations of data are subjects of our experiments.

As in the current system, in the lottery sets aside fraction of seats at each school for each respective group. The seats not filled by the lottery gradually filled in the after-lottery phase.

The following “factors” were computed for each of the 445 “planning areas” – particles of division – of the city. **“What is a ‘planning area?’”** Since 1990 we have been using a scheme of geographic divisions which is much smaller than census tracts but larger than city blocks. Called “planning areas”, these were originally “drawn” around 1990 by a consultant, “Educational Data Systems” (EDS). Though EDS has long since left the picture their scheme of division has long seemed useful. Each is typically 4 – 8 city blocks. Our planning software has ever used it as a scheme of division. Since these do not coincide precisely with the U.S. Census schemes of division we have derived a mapping – a big set of linear equations – by which we map the block-group census data to our planning areas.

Computation of Socio-Economic Factors and Race Factor

1. **Socio-Economic Factor: “Household Income”** This is taken direct from Census 2000 matrix P52.
2. **Socio-Economic Factor: “Educational Average”** This is a weighted average computed from Census 2000, the populations contained in matrix P37, using the following methodology:

“Weight” per applied to education categories:

- 1 - Finished grade 8 or less;
- 2 - Did not finish high school;
- 3 - Finished high school;
- 3.5 - Some college or associate degree.
- 4 - Bachelor’s degree;
- 5 - Masters or professional degree;
- 6 - Doctorate.

Using the weights above the weighted “average” is computed as follows. The average yields a decimal number between 1.0 and 6.0.

“Education Average” =

$$\frac{\sum_{\text{over all the above categories}} (\text{Population of category} * \text{Weight per category})}{\text{Total population}}$$

3. **Race Factor: “Percent non-white”** For the purpose of combining a “race” factor with multiple other factors, a single-numeral measure of race is desired. In recent trials we represented racial diversity as a single percentage, “percent non-white.” The computation used BUSD K-5 enrollment data of several years. The data of several years was summed.

“Percent non-white” =

$$\frac{100 * \text{Sum of non-white population}}{\text{Total population.}}$$

Combining Factors to yield “uncorrected” Socio-Economic Map 101203

4. **Socio-Economic + Race Classification Map 101203, Unadjusted.** The three measures detailed above were combined to yield an integer “classification” code limited to values 1, 2 and 3. This three-code simplicity resembles our present three-race scheme, though of course the three codes no longer indicate race. Preserving a three-code scheme allows us to easily use current infrastructure (software and reports), is familiar to our staff and preserves our accumulated experience with student assignment methods.

“How do we combine this diverse data to get a simple 1, 2 or 3?”

Each of the three “factors” varies over a different data “space.” Average household income varies from 4900 to 153,500 (in Berkeley.) The Education average varies from 3.0 to 4.6. The “Percent non-white” varies from 0 to 100. To combine the three factors in a way that gives them equal weight, each factor must be “mapped” (linearly transformed) from these disparate “outcome spaces” (range) to a “common outcome” space – a decimal value between 1.0 and 3.9. When each factor has been transformed to a decimal value which varies identically then we can give each the same “weight.”

Below is the formula. The three components are the three “mapped” (linearly transformed) factors. To each mapped factor is applied a weight of .33 (33%.)

(Decimal) Socio-Economic Weighted Average =

$$.33 \times (2. + (\text{Household Income} - 34000) / (70000 - 34000)) +$$

$$.33 \times (2. + (\text{Education Average} - 3.4) / (4.1 - 3.4)) +$$

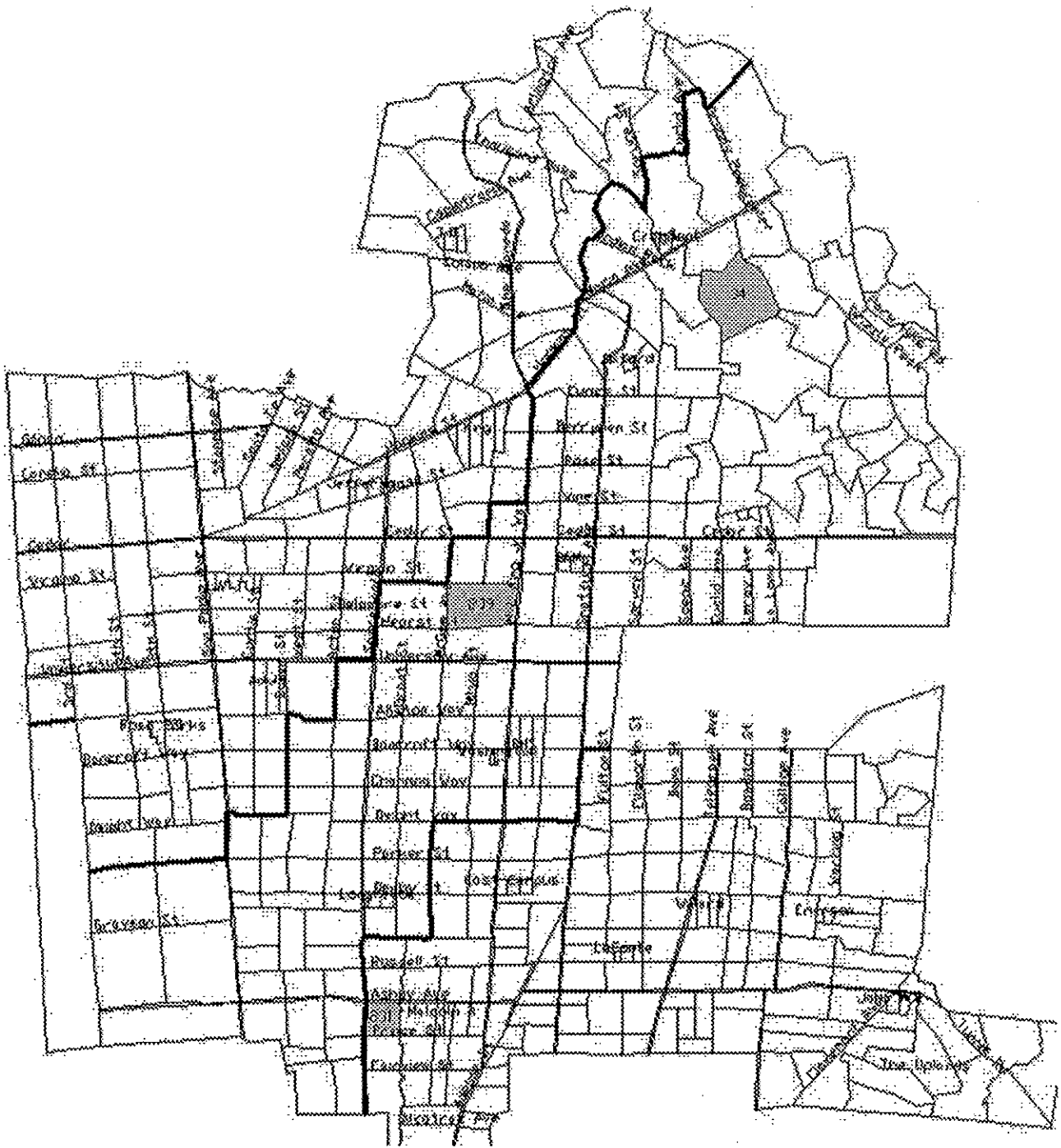
$$.33 \times (2. + (\text{Percent Non White} - 30) / (67 - 30))$$

The desired Socio-Economic Category - 1, 2 or 3 - is derived from this “Socio-Economic Weighted Average” by applying two thresholds or “break points” to the decimal value. The breakpoints were determined by experiment. The breakpoints were chosen simply to divide the city’s K-5 population into desired proportions – not too small, not too large.

Weighed Avg 0 to 2.2 →	Weighted Avg 2.2 to 3.0 →	Weighted Avg 3.0 and above →
SocioEc Category 1	SocioEc Category 2	SocioEc Category 3

The following is an example computation of this for three Planning Areas of the Berkeley map. The locations of areas 34, 231 and 239 can be seen on the map that follows.

Planning Area	Average Income	Average Education	Percent Non-white	Socio-Economic Weighted Average	Socio-ec Category
34	104753	4.5	10	3.66	3
231	36250	3.4	92	1.78	1
239	47574	4.2	29	2.82	2



Socio-Economic Map 101203, "Corrected"

5. **Socio-Economic + Race Classification Map 101203.** The constituents of the "final" map 101203 are the same as (4). But the map was "adjusted" by changing the category of the "outliers" – the isolated pieces. They were re-classified by hand to avoid a checkerboard arrangement. This map divides the K-5 population roughly 55% 25% 20%. That is, the "lower" socio-economic group is much larger than the others.