Addressing SES Segregation in Charlottle-Meckelenburg Schools

A Genetic Programming Approach

Michael Robertson on the senior thesis of Lena Parker '17
Davidson College
mirobertson@davidson.edu; leparker17@alumni.davidson.edu

Abstract

Lena Parker's ('17) thesis addressed the problem of Socioeconomic segregation in Charlottle-Meckelenburg Schools by using a linear programming model to optimize a function which weighted the relative "good" of two opposing desired characteristics of a school: commute time for each student and diversity of the school. Her approach struggled to run in a reasonable amount of time given the size of the problem, so we attempt to solve the same problem using genetic programming. This method of solving the problem gives no guarantee of an optimal solution, as a linear programming approach theoretically does, but it can give a solution within any time frame, an advantage over Linear Programming.

Introduction

In contrast to Linear Programming, which attempts to find the values of the variables (here, the school which each student attends) which minimize or maximize and objective function, genetic programming attempts to harness the power of natural selection and reproduction in order to converge upon a solution with a desirable fitness value. To reframe the problem from Lena's linear programming approach to ours, we simply used her optimization functions as our fitness function, and created mechanisms to generate different assignments of children to schools, as well as ways for them to crossover, mutate, and compete for the ability to reproduce/survive.

Materials and Methods

Using data from the US Census American Community Survey (ACS), the CMS supplied Lina with a categorization of each census block group into one of three Socio-Economic-Statuses: Low, Medium, or High.

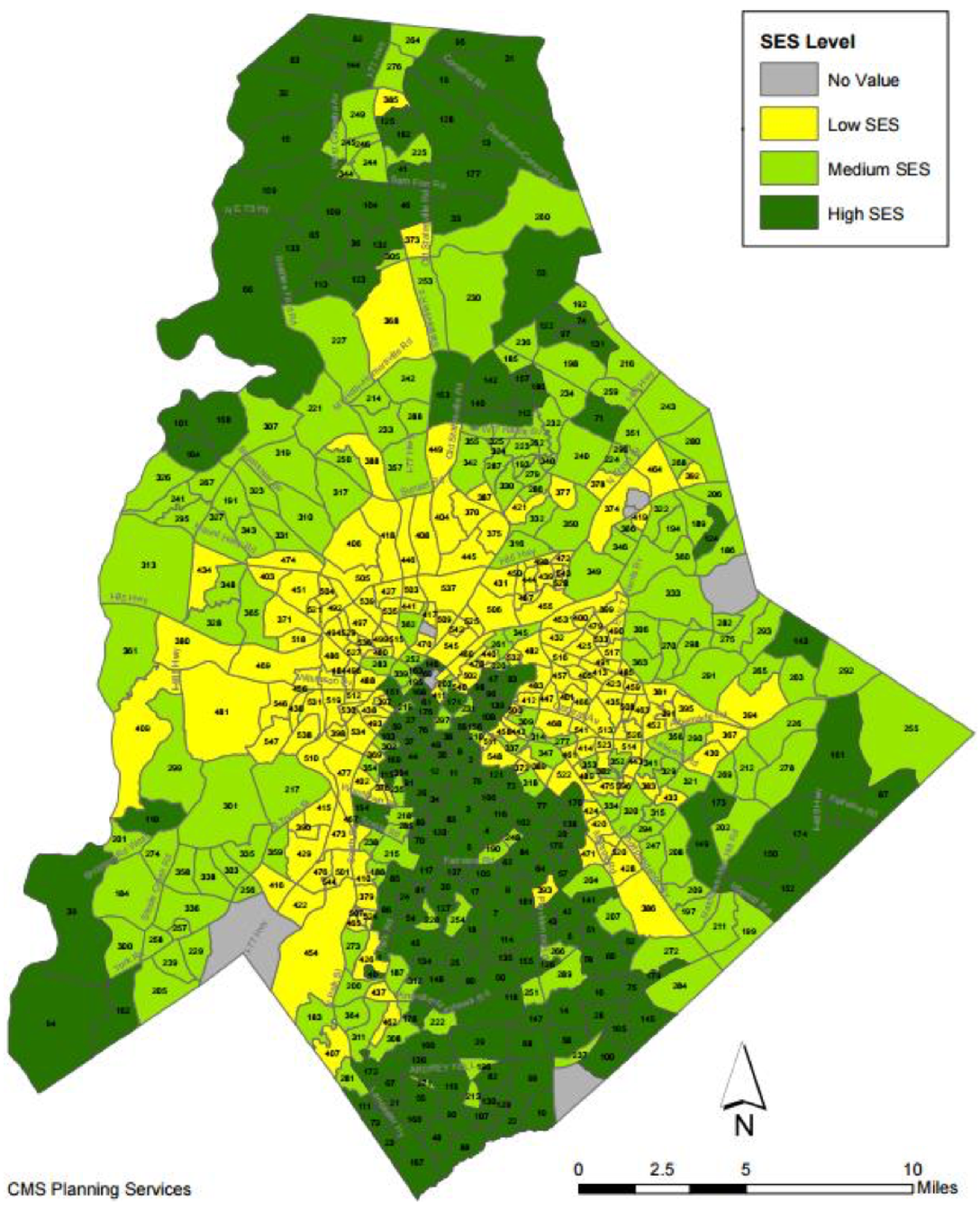

Figure 1: The SES measure across CMS census block groups.

She further used CMS data on school capacities, census estimations of each census block population, and a description of the roads in Mecklengburg from Open Mapping Mecklenburg to produce data estimating the population of school age children in each road segment as well as the distance to the closest five schools, and those student's SES.

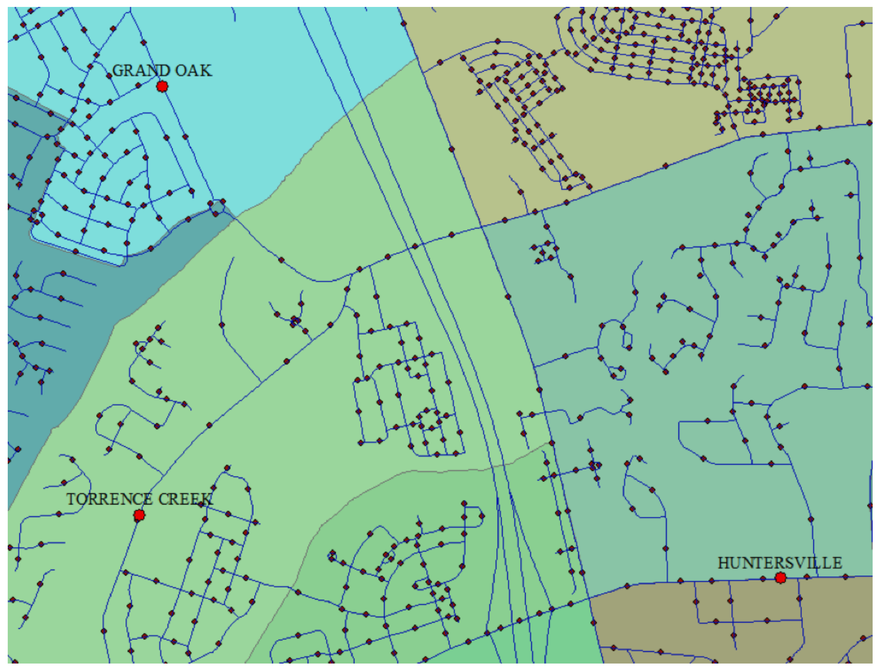

Figure 2: This map shows a section of Mecklenburg County with three schools, a network of roads and their midpoints. Schools and roads are overlaid on census block groups with varying populations and SES measures.

Mathematical Section

We create our fitness function

$$F(A) = \sum_{j=1}^{S} \sum_{i=1}^{R} T_{ij} P_i W(S_j)$$

Where S is the number of schools, R is the number of roads, T_{ij} is the commute time from road i to school j, P_i is the population of road i, S_j is school j, and $W(S) = \sum_{i=1}^{3} \left|(\frac{1}{3} - SES_i)\right| + 1$. We define SES_1 as the proportion of students in the school with a low SES status, SES_2 as the proportion of medium SES students, and SES_3 as the proportion of high SES students.

Results

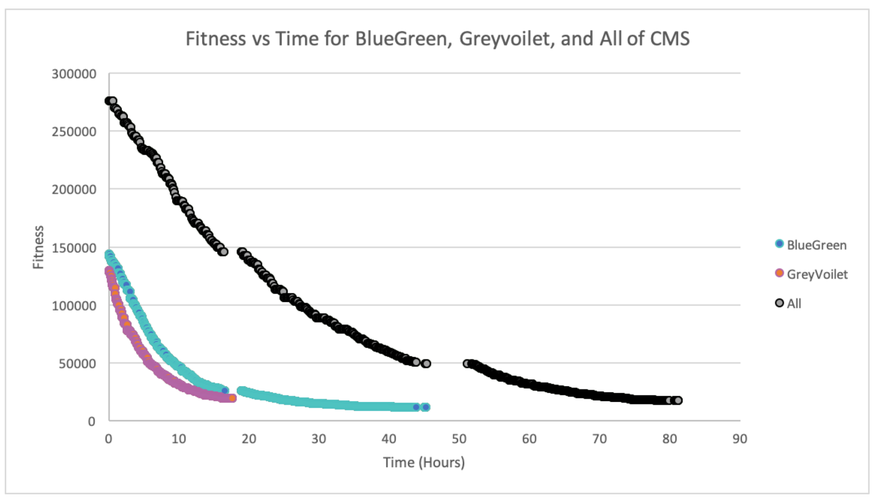

Figure 3: The graph shows 2,237 total data points representing 2,237 generations across three data sets over nearly 80 hours of runtime.

Our algorithm improved its fitness by over 80% on all three datasets, and shows promise in offering a efficient solution to problem.

References

[1] United States Census Bureau. Population by age (2015 american community survey).

[2] Mecklenburg Schools. 2016-17 school diversity report. http://www.cms.k12.nc.us/cmsdepartments/StudentPlacement/PlanningServices/Documents/PMR%20Month%2001bySchool_2016$_Diversity - final.pdf$.

Acknowledgements

We are grateful to the generous assistance of Lena Parker for providing her work freely and offering her time to explain it. We also thank Dr. Ramanujan for advising both Lena's and our own attempts to address this problem, as well as the US Census and CMS for supplying the relevant data.