Using GeoData for School Reform
Bob Coulter

Abstract: The Center for Inquiry in Science Teaching and Learning (CISTL) is funded by the National Science Foundation to develop innovative collaborations between institutions of higher education, informal science institutions, and local school districts in the St. Louis, Missouri, area. Together, the partners are developing programs intended to build teachers' skills and boost student achievement from early childhood through high school. In this session, you will see how GIS is being used to analyze relevant demographic, economic, and public health data, and how this data is being linked to student achievement data to better inform project partners' efforts and to provide useful information to guide school reform efforts.

The St. Louis Center for Inquiry in Science Teaching and Learning (CISTL) is an effort funded by the National Science Foundation to improve the quality of science education in five participating school districts through better collaboration among various educational and cultural institutions in the area. Collectively, the program involves the five school districts, two universities and a multi-campus community college, and three informal science institutions.

In June 2004, several members of the CISTL team attended a meeting at the Chicago Consortium, with the goal of exploring ways in which demographic and geospatial data could be used to inform program decision making and to contribute to evaluation efforts. As a result of that meeting and other conversations prior to and subsequent to that meeting, the CISTL leadership initiated support for the development of the first phase of a database of relevant demographic factors and assessment results. Knowledge of this data is expected to contribute to the larger CISTL goals of improving the quality of science education afforded to students in the St. Louis region.

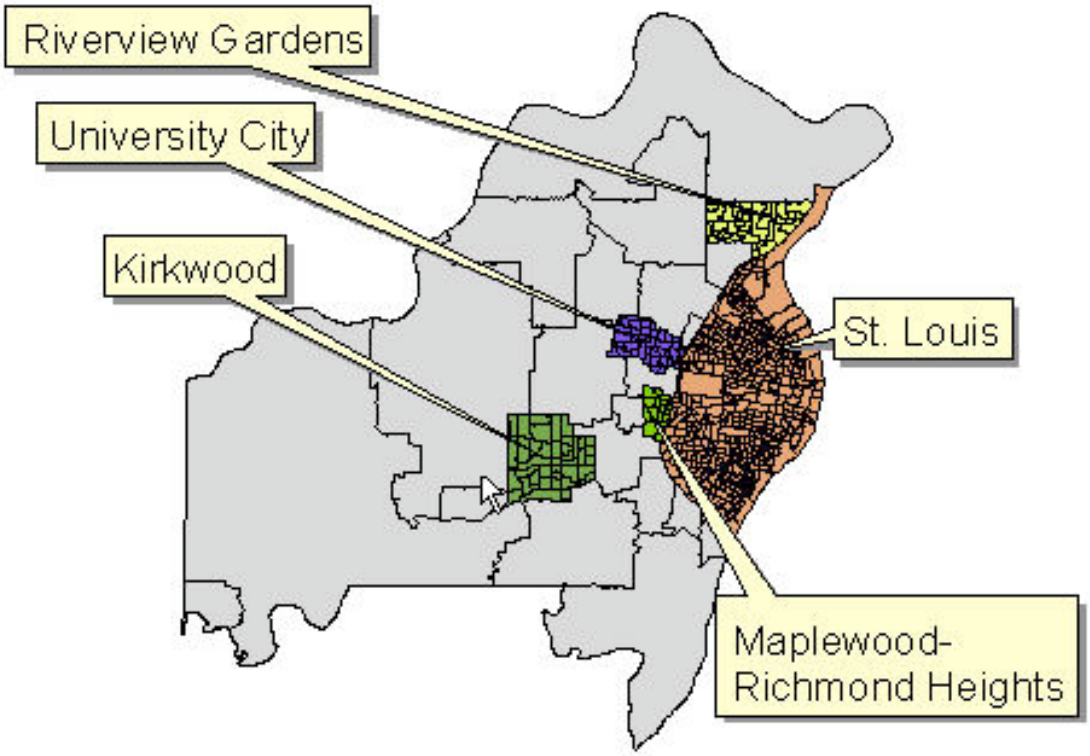
This paper describes the data tools being used and illustrates samples of baseline demographic analysis that can be conducted at the district level and school levels. In the second half of 2005, relevant demographic factors will be linked to student achievement data and teacher background data, supporting analysis at the both the district and school level. Unless otherwise restricted by the providers or necessitated for reasons of confidentiality, all of this data will be made available to researchers in standard quantitative and geospatial formats to facilitate further investigation.

Data Preparation Methodology

The major source of demographic data in this analysis is the CommunityInfo for Education product from ESRI. This DVD contains 1008 variables derived from the United States Census Bureau and other sources. This data is then aggregated at different levels of Census geography: Block Groups, Zip Codes, Core Based Statistical Areas, Census Tract, and State. For this analysis, the Block Group was chosen as being sufficiently fine-scale in geographic extent to

enable detailed analysis, while having a wider array of demographic variables than the Census offers for more localized groupings.

The complete CommunityInfo data set representing the United States was “clipped” spatially in ArcView 9 to St. Louis County and St. Louis City. It was then further clipped to contain only the geographic extent of the five CISTL districts (in aggregate and as individual districts). While not directly relevant to the CISTL districts, the county/city level data was taken as an intermediate step to enable future analysis of neighboring districts. Clipping of St. Louis



County school districts was achieved using a boundary theme provided by the St. Louis County Department of Planning.

Selection of block groups to be assigned to each school district proved to be more challenging, as the block groups are not intentionally contained within a given school district. Thus, many overlap two or more districts. To facilitate the appropriate assignment of block groups to the five participating school districts, a 4-part process was implemented:

1) For the St. Louis Public School District, there were no issues to address, as no block groups overlap the boundaries of the City of St. Louis. The district boundaries are coterminous with the city boundaries. (This occurs because the City of St. Louis is incorporated as as a separate county.)

2) For the four remaining districts (Kirkwood, Maplewood-Richmond Heights, University City, and Riverview Gardens), all block groups that were entirely or substantially contained within the school district boundaries were assigned to the district. Since there are no common boundaries among these four districts, there were no situations in which one block group was

assigned to more than one CISTL district, or where assignment to one district took precedence over another CISTL district. [1]

3) For situations where only a smaller portion of the block group is contained within a school district, aerial photography was used to assess the appropriateness of assigning the block group to the district. The decision was based on whether the overlap contained significant residential development or (as was true in some cases) only industrial/commercial or institutional land use was present. In these cases there would be no reason to include Census data as there was no residential presence within the portion of the block group overlapping the school district.

4) In cases where boundaries were essentially the same but not absolutely identical, the assumption was made that these slivers were cartographic artifacts arising from the mapmaking processes and/or the conversion of map projections between the school district boundaries and the Census data.

In all cases it should be remembered that absolute precision in these assignment matters is likely not feasible nor is it likely to improve the subsequent demographic analysis. Ultimately, it must be acknowledged that the analysis made possible with these tools is able at best only to capture broad-scale trends. Most of the demographic variables are from the 2000 decennial census, and thus are inherently somewhat dated. Also, given transfer of students among districts through normal mobility and programs such as the voluntary transfer (desegregation) program, as well as the uncertain correlation of community-wide demographics with the characteristics of district school-attending families, some acceptance of uncertainty is required. Finally, the ability of many parents to send their children to religiously-affiliated or independent schools skews the data as these are often the more affluent parents – a factor that correlates directly with education and race among other factors. Thus, public school enrollment is disproportionately poorer and composed of a greater proportion of racial minorities than is true of the community at large.

As noted in the concluding remarks, this tenuous link between Census data and school data limits the utility of this first phase of the research, except perhaps to document the variance that exist in many cases between the community demographics and student enrollment demographics. Knowing the community as a whole, however, does have potential value for school reform efforts, however, as it details the civic capacity that could be brought to bear on reform efforts.

Returning to the practical issues regarding the data, working with 1008 variables can be cumbersome and potentially intimidating to a novice data user. Thus, before public release of this data to project partners, project staff will “thin” the demographic attributes to facilitate access to those factors deemed most relevant for a typical user (e.g. a district administrator or grant writer), or to a subset of variables useful for specific research purposes. Also, in some cases,

---

[1] This issue of overlap does need to be addressed for analysis of the individual school attendance zones within a given district. Also, there is an unresolved issue of proportional weighting where one block group may be assigned to a district but only a portion of the residents in the assigned block group are actually residents of the school district. A future refinement of the database could use parcel-level analysis to provide an approximation of proportional weighting. For example, in the current database, all residents of a given block group are factored into the analysis. Parcel-level analysis could identify that a certain percentage of the homes are within a given district; from there, a proportion of that block group’s data could be used in compiling the aggregate data for the district. While this proportional weighting could improve accuracy, whether the data for the entire block group is representative of the portion contained within the school district is uncertain.

derived variables may be valuable, such as a grouping of different levels of education. For example, the discrete grade levels of education at less than a high school diploma (9th grade, 10th grade, etc.) could be aggregated into a “Less than high school education” variable. At all times, a separate, intact set of geospatial files based on the 1008 attributes in the CommunityInfo database will be maintained for those seeking to analyze more specific variables.

Sample District-Level Analyses

The intention of creating this geospatial analysis framework and related databases is to facilitate inquiry among interested parties. The analyses shown here are meant to be representative, and not in any way exhaustive of the capabilities of the tools.

As a general statement, two levels of analysis are possible: characteristics of and variability within a district, and comparisons among districts For simplicity, these latter analyses presented here include aggregated district-level and CISTL-wide statistics. As noted previously, there are presently unresolved issues deriving from the block group assignment process and from the uncertain correlation between the demographics of the residents within the district boundaries and those of the families who send their children to district schools.

To illustrate the capacities of the data tools at the district level, three sample analyses are presented here, with data presented quantitatively and spatially as appropriate:

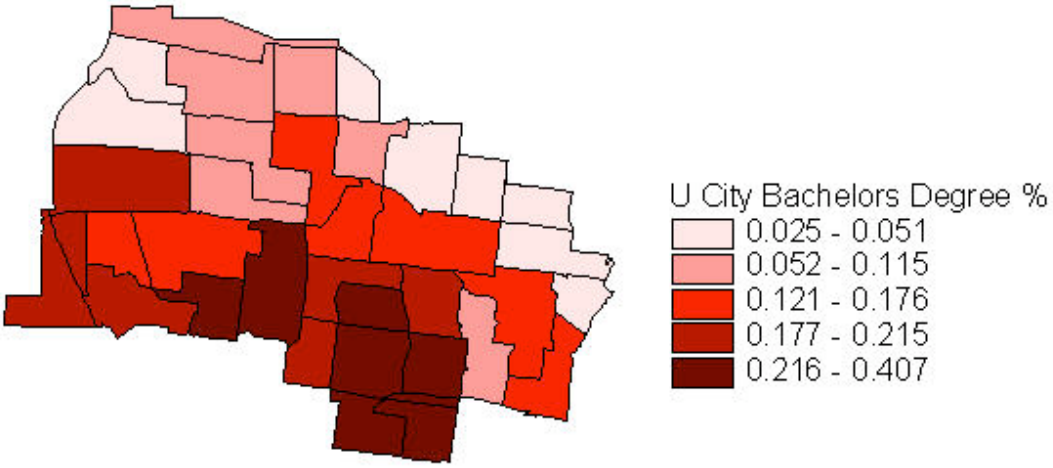
1. Education levels within and across CISTL districts
2. Racial groups within and across CISTL districts
3. Median income levels within and across CISTL districts

Education Levels Within and Across CISTL Districts

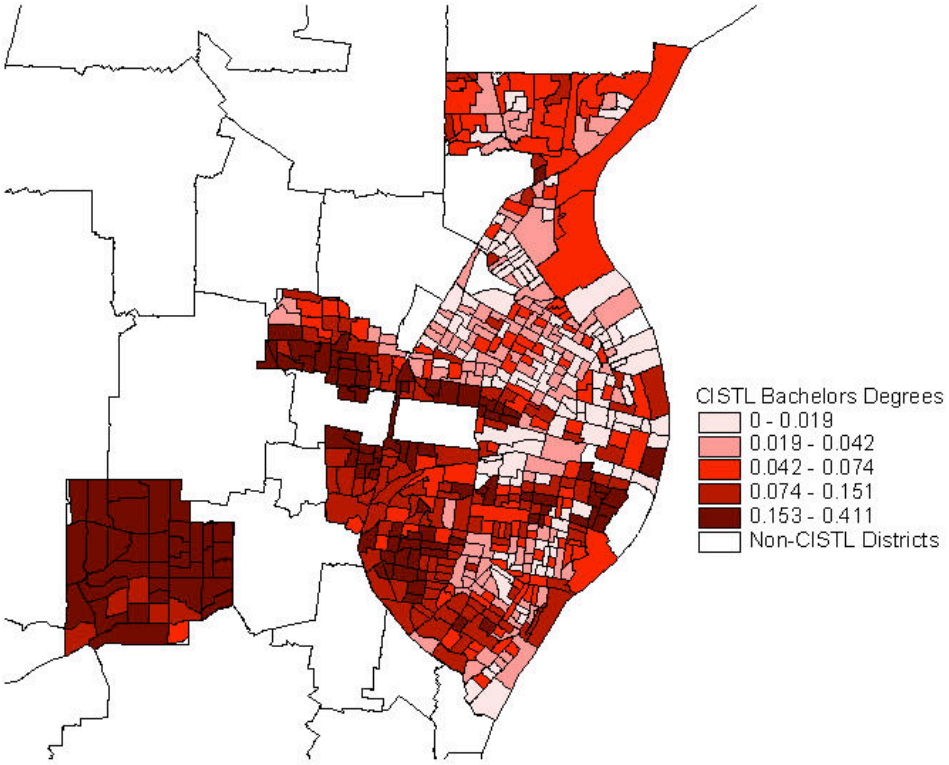
Given that parental education levels are a strong correlate of student achievement, knowing the education level within a community can be a strong predictor of student achievement. In this analysis, percentages of residents who are listed as holding a bachelors’ degree within and across districts is presented. Within each district there is considerable variation in education levels, with some areas (captured within Census block groups) having higher levels of formal education. Across the region, this variation is even more pronounced.

In the maps presented here, a sample district (University City) is presented showing the levels of college education by block group (defined as the number of identified residents in 2000 with bachelors’ degrees, divided by the total population that year within each block group). The map is constructed using a quantile breakdown representing the highest 20% within the area, second 20%, and so forth.

Notice that within University City, there are block groups with just over 40% of the residents holding bachelors’ degrees, while other block groups show as little as 2.5% of the residents holding a bachelors’ degree. In the second phase of this analysis, it will be interesting to see the extent to which this variation in educational attainment correlates with student achievement among the individual elementary schools in the University City school district. Also, note that in the quantile breakdown, the top 20% of block groups cover a wide range, from 21.6% to 40.7%. The remaining 80% of block groups represent a much more constrained range of 4-year college graduates:



Across all CISTL districts, a similar range is observed[2]:



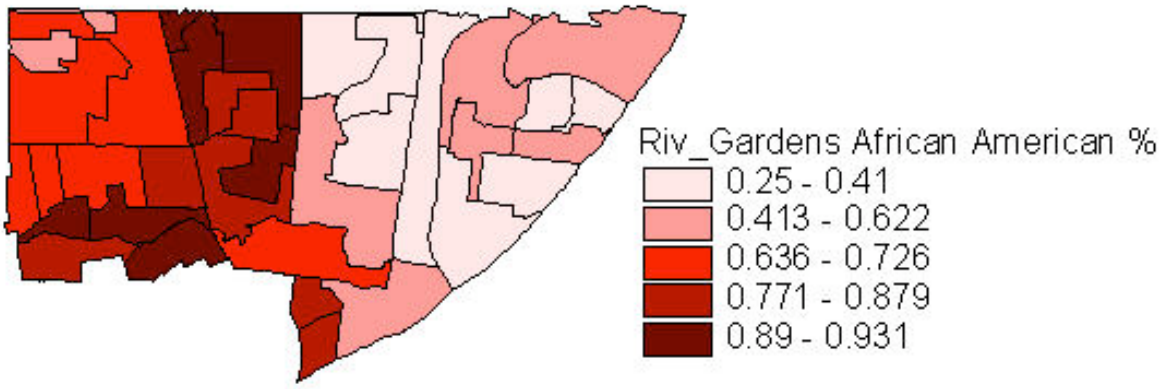
[2] Blank areas with no population in St. Louis (city) typically represent parks and/or industrial areas.

District and CISTL-wide data is summarized in the following table:

District	Bachelors' Degrees	Total Population	Percentage
Kirkwood	10,951	47,481	23.1
Maplewood-Richmond Heights	2,889	16,363	17.7
University City	5,578	39,117	14.3
St. Louis	25,453	350,953	7.3
Riverview Gardens	2,517	49,323	5.1
CISTL-wide	47,388	503,237	9.4

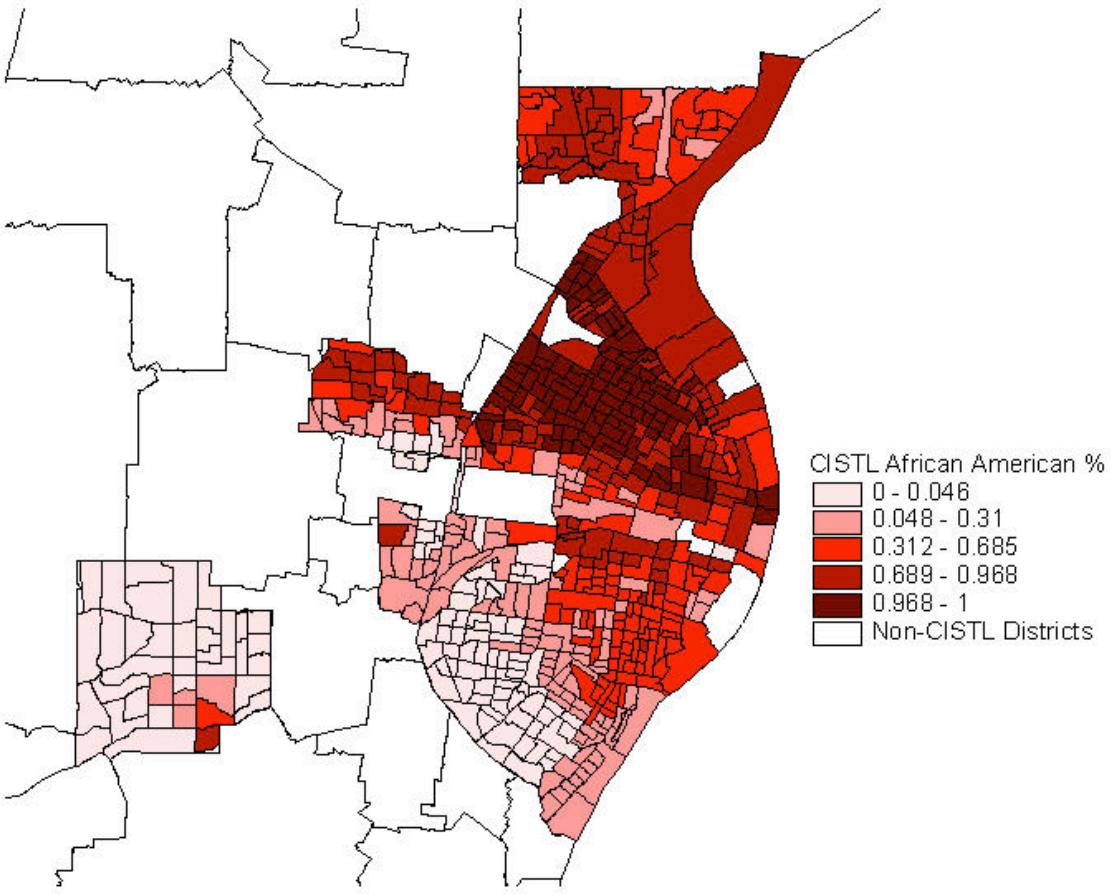
Racial Composition Within CISTL Districts

As with the analysis of educational levels, there exists among the five CISTL districts a range of racial diversity among specific block groups. Given the prevalence of desegregation concerns in the region, analysis of the racial balance in residential patterns will likely prove to be fruitful. The data presented here shows the African-American population ("Black" within the Census designations) as a percentage of the total population within a block group, calculated first for the Riverview Gardens district as a sample, and then for the CISTL service area as a whole:



Notice that even within a district such as Riverview Gardens that has a high percentage of African American population, there is considerable diversity, with block groups ranging from 25 to 93 % African American.

Viewed across the CISTL districts, the racial separations that divide the area can be seen clearly:

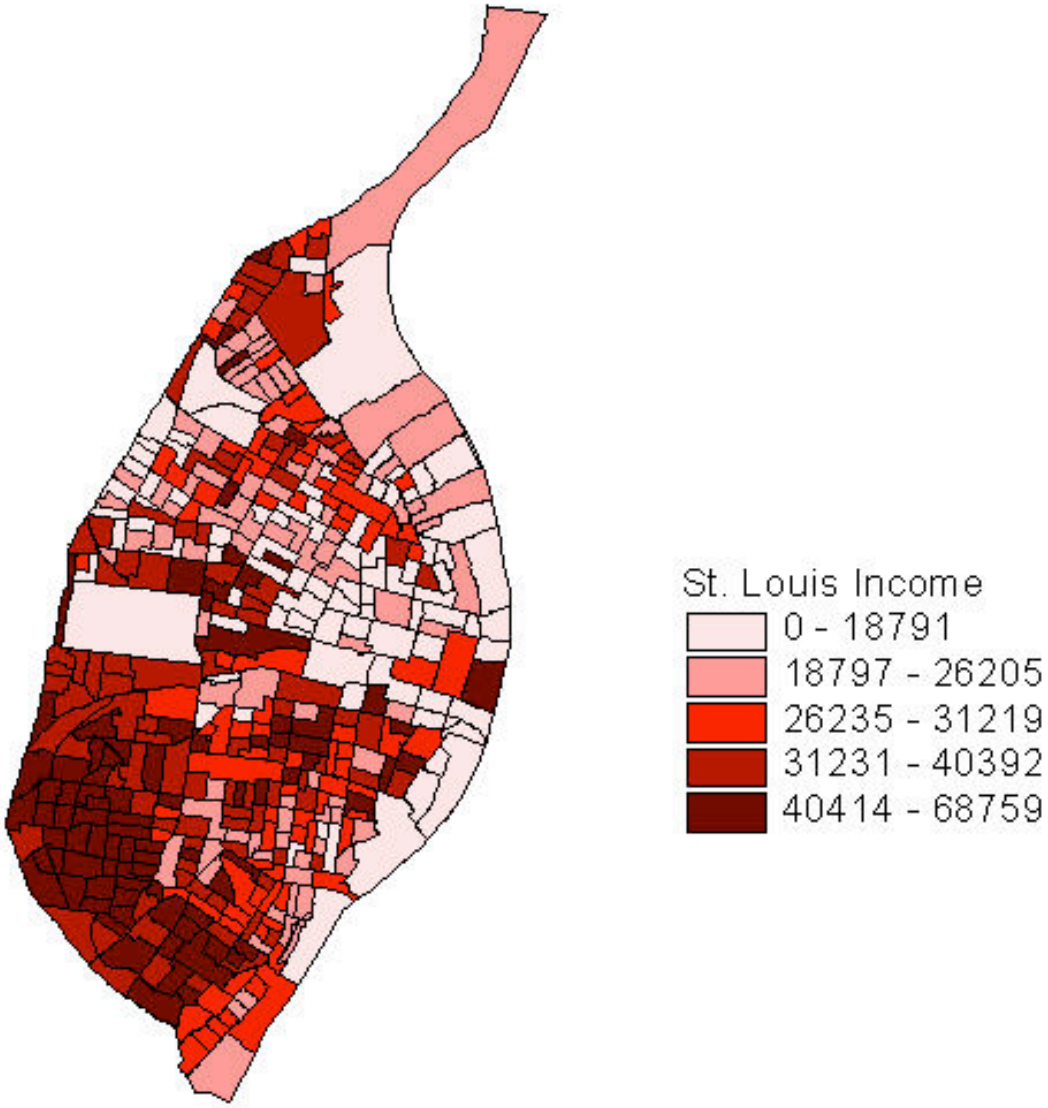


District and CISTL-wide data is summarized in the following table:

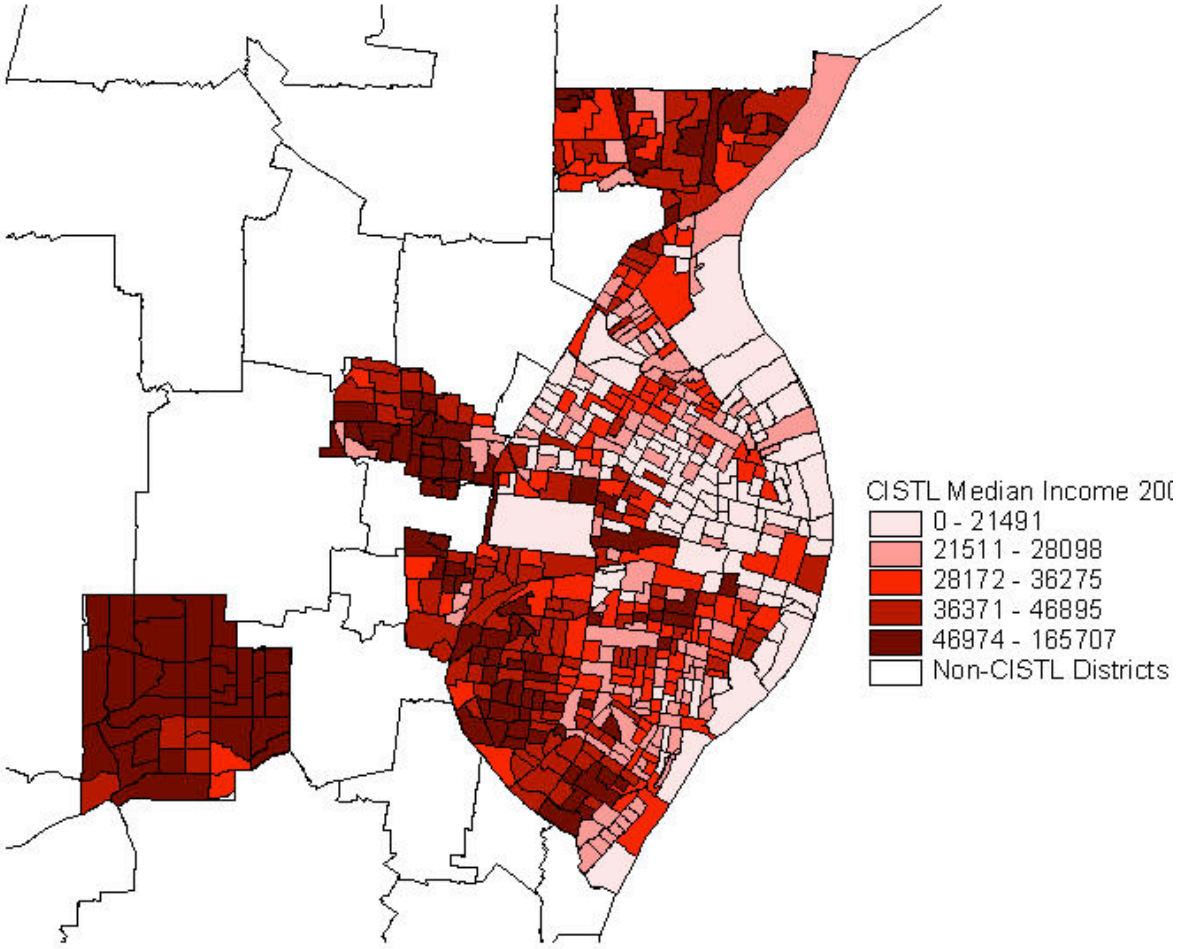
District	African-American Population	Total Population	Percentage
Riverview Gardens	32,063	49,323	65.0
St. Louis	178,406	350,953	50.8
University City	17,868	39,117	45.7
Maplewood-Richmond Heights	2,642	16,363	16.1
Kirkwood	2,114	47,481	4.4
CISTL-wide	233,093	503,237	48.2

Income Levels Within and Across CISTL Districts

As one might expect, there is also a wide range of incomes within and across the five CISTL districts. The Census data represents median income by block group; the last column in the table below represents the mean value of the median incomes for the block groups within the school district. Within the city of St. Louis there is a strong trend toward the more affluent zip codes being in the southwest portion of the city, with northern portions more prone to poverty:



Viewed across all five districts, this variation is even more apparent with substantial unevenness in income. Given the extent to which affluence is correlated with educational achievement, responding constructively to this variation is of considerable importance to the success of the CISTL program:

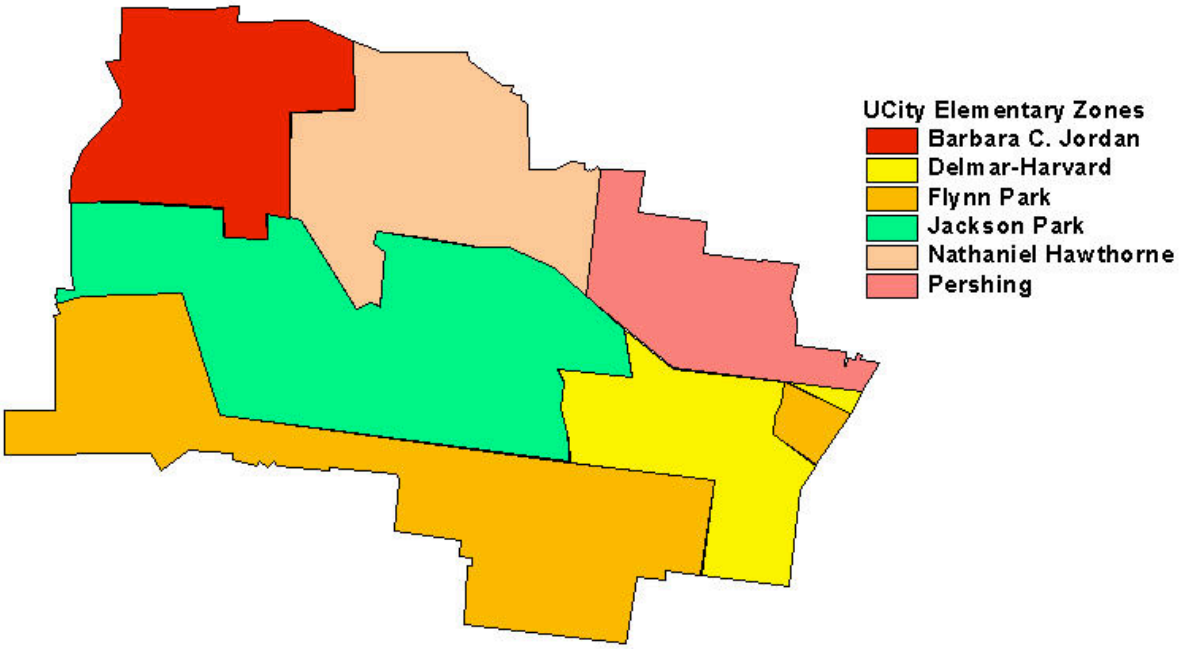


District	High Median	Low Median	Mean of Medians
Kirkwood	$165,707	$28,431	$81,711
University City	139,341	21,628	57,321
Riverview Gardens	61,862	22,664	41,160
Maplewood-R.H	87,554	21,532	46,688
St. Louis	68,759	7,308	30,599
CISTL-wide	165,707	7,308	36,947

Investigating Variability Within a District

Given the variability observed within the districts, it may be useful to analyze this variation further, seeking to link it to individual schools (and in turn, achievement within those schools). For each of the factors mapped in the previous analysis, there exists a range among the school attendance zones that comprise that district. Knowing the extent to which these variables correlate with student achievement will provide a finer grain to the overall analysis than would be possible if the analysis were conducted solely at a district level.

In the examples here, school attendance zones within the University City school district will be used. These were digitized from maps provided on the district web site (www.ucityschools.org). As the block groups overlap considerably, a "Select by Location" operation was conducted to choose the block groups with their centers in the school attendance zone. Admittedly, this creates an approximation at best of the demographic characteristics of the zone. As noted previously, however, there are inherent uncertainties in the data given the age of the Census data (5 years) and the fact that school enrollment is likely not a direct reflection of the overall demographics of the community. This lack of direct correlation is explored in the next section.



School	Median Income	Percent African-American	Percent with a Bachelor's Degree
Barbara C. Jordan	$36,581	72.9%	6.1%
Delmar-Harvard	$44,827	15.8%	24.1%
Flynn Park	$81,291	9.0%	20.6%
Jackson Park	$63,307	35.1%	20.8%
Nath. Hawthorne	$45,120	25.6%	7.1%
Pershing	$33,058	91.3%	3.9%

Connecting School and Community Data

As noted previously, the correlation between the Census data within a community and the demographics of those actually enrolled in a school or district is uncertain. After reviewing the data for individual schools as reported on the Missouri Department of Elementary and Secondary Education (DESE) web site[3], it appears that while there is some correlation, a disproportionate share of the more economically able essentially "escape" from the public schools. Comparing the same median income and racial data listed above with the data for those schools reported on the DESE site shows some striking gaps, particularly in the more affluent neighborhoods:

School	Median Income	Free/Reduced Lunch	Census African-American	DESE African-American
Barbara C Jordan	$36,581	71.2%	72.9%	99.3%
Delmar-Harvard	$44,827	61.3%	15.8%	83.3%
Flynn Park	$81,291	34.1%	9%	48.8%
Jackson Park	$63,307	39.8%	35.1%	72.6%
Nathaniel Hawthorne	$45,120	71%	25.6%	95.5%
Pershing	$33,058	75.2%	91.3%	99%

[3] State-mandated district "report cards" can be found at http://dese.mo.gov/planning/profile/096112.html

Concluding Thoughts

In the early stages of this inquiry, a number of questions can be raised. Note for instance that the Riverview Gardens median income appears to be solid. Given this, why is academic achievement consistently below the norm for the district? Several potential explanations are worth considering. Similarly, the citizens of the Maplewood-Richmond Heights district appear to be well-educated based on demographic data, but for years the school district's achievement lagged behind other districts. Recent promising curricular reforms initiated by the current administration appear to be reversing this trend, however. Investigating these issues in considerable detail will be made possible by the development of a robust data tool that integrates demographic and achievement data.

In this initial phase of the project, it has become apparent that the gap between the community demographics and the characteristics of those actually enrolled in the schools will need further documentation and analysis. It appears that based on the preliminary data analysis that school attendees are poorer and more likely to represent a racial minority than is true for the population as a whole. Thus, one might conclude that students within a school are more likely to be "at risk" than one would conclude based solely on community demographics. As the project moves forward, we anticipate being able to document and further analyze these issues both in the immediate case and over time, develop the capacity to document trends in the data.

Author Contact Information

Bob Coulter
Senior Manager – Education Division
Missouri Botanical Garden
PO Box 299
St. Louis, MO 63166-0299
(314) 577-0219
(314) 577-0847 fax
bob.coulter@mobot.org