Title: GIS Analysis of University Enrollment Using Census Data and Geo-Coding*

Author: J. M. Pogodzinski**,** Dept of Economics (0114), San Jose State University, 1 Washington Square**,** San Jose, CA 95192-0114**,** USA**,** Tel: 408-924-5421
email: jmp@pogodzinski.net

Abstract
The purpose of this paper is to determine economic and demographic characteristics of three populations – applicants, students accepted to, and students who enroll, in an urban university in California. The addresses of applicants are geo-coded, distinguishing applicants who were admitted and applicants who enrolled. Using spatial joins these subgroups are identified with the census tract or block group of the address. The effect of economic and demographic variables on applications, admissions, and enrollment can then be analyzed. The project focuses on the nine counties of the San Francisco Bay Area from which the majority of applicants arise. This area consists of more than 4,400 census block groups. In addition to census data, regional authorities such as the Association of Bay Area Governments (ABAG) and the Metropolitan Transportation Commission (MTC), make available a large amount of data concerning the nine counties of the study.

I. Introduction

This paper examines the application-admission-enrollment (A-A-E) funnel for the First-Time Freshmen (FTF) at San Jose State University in Fall 2005. In addition, drop outs in the following two semesters are analyzed. The focus of this paper is on three variables: student credit units (SCUs) attempted in each of the semesters, the distance from campus, and a variable measuring exposure to college in the neighborhood (census tract) containing the student's address.

The analysis shows that drop outs have a "tell" – or an early warning signal – in the form of a significantly lower level of SCUs attempted in the semesters prior to dropping out. The decrease in SCUs is especially marked for students who do not continue their higher education at another institution, and is also especially marked for students who live closer to the campus.

As a general proposition, the A-A-E funnel filters on the exposure to college index: students who enroll at San Jose State come from neighborhoods where the "exposure to college" is lower than do students admitted to the University, and, likewise, those admitted to the University come from neighborhoods where the "exposure to college" is lower than do students who apply to the University. Students who enroll in the University do not, however, come from neighborhoods with the lowest "exposure to college" index. The average "exposure to college" index of enrollees from Santa Clara County is lower, but not markedly lower, than the average for the county. Although the "exposure to college" index differs significantly across Santa Clara County, the average "exposure to college" index is weighted by the target population (students in grades 9 through 12 reported in the 2000 Census). San Jose State enrollees have approximately the county average exposure, probably because the census tracts rated higher on the "exposure to college" index had fewer children in the target group.

II. Literature Review

Most economic and many demographic characteristics of applicants, accepted students, and enrolled students are not part of the student's record. For example, the household income of most students is not part of the information typically available to institutional researchers. Similarly, many characteristics of the students' broader environment – neighborhood characteristics like the level of vacant housing, the extent of poverty in the area, or the proportion of the population with exposure to college – are not available as part of the student record.

Ahlburg, McPherson, and Schapiro (1994) review a variety of models used for enrollment forecasting. Several of the studies reviewed bring in variables not included in

the student record, but these are for nationwide studies of college choice, suggesting that they refer mostly to universities whose market area is national. DesJardins, Ahlburg, and McCall (2006) develop a sophisticated model of the apply-accept-enroll decision process, tested on a large data set. Although they bring in a limited number of variables not included in the student record, these variables (e.g., state unemployment rates) are highly aggregated.

GIS has been applied previously (Marble, Mora, and Granados, (1997), Marble, Mora, and Herries (1995)) to parts of the apply-admit-enroll decision being studied in this paper. Marble, Mora, and Herries, (1995) examined census tract information, and correlated applications with school district data. Marble, Mora, and Granados (1997) used census block group data on household income to develop a decision-support model for targeting financial aid.

This paper seeks to extend the use of GIS to study subgroups (e.g., ethnic/racial groups, non-returning students) using census data at a low level of aggregation. Using GIS, it is possible to create a map (and an underlying database) which can display a variety of census variables such as median household income, percent of vacant housing, measures of poverty, etc., down to a fairly low level of aggregation – the census tract or census block group (U.S. Census Summary File 3 (SF3) (2000), U. S. Census, Census Tract and Census Block Group Boundary Files, (1990), U. S. Census, Census Tract and Census Block Group Boundary Files, (2000)). Census tracts usually contain about 5,000 people. Most census block groups consist of areas with population between 1,000 and 5,000 people. Sometimes, some variables are suppressed by the Census Bureau if the population of the block group is too small to preserve anonymity.

The street addresses of applicants, accepted students, and students who enroll can be geocoded (i.e., geographically identified with latitude and longitude); see Batch Geocode (n.d.). Then, by undertaking a "spatial join" operation (Gorr and Kurland (2005), Huxhold, Fowler, and Parr (2004), Arctur and Zeiler (2004)) it is possible to determine the census tract or block group containing the address. Statistics such as the percentage of students who have addresses in high-poverty areas, or areas with high housing vacancy rates, or areas consisting of people with low exposure to college, can be determined. Similarly, the median household income of the census block groups from which student applicants, accepted students, and enrolling students come can be determined. This is not the same as making a direct observation of the individual student's income, but it should shed some light on the significance of demographic and economic variables that are not usually available to institutional research.

This initial study is confined to the six counties of California from which the majority of FTF applicants to San Jose State University arise. This area consists of more than 3,000 census block groups.

III. A Model of the Application-Admission-Enrollment Funnel

As part of the 23-campus California State University system, applicants to San Jose State University use a standard system-wide application, and can apply to up to five campuses with a single application. There were 16,386 FTF applicants from California addresses in Fall 2005. Of these, 10,867 were accepted; 2,513 enrolled.

Given the ease of application, it is reasonable to expect that there will be greater geographic diversity in applications than in enrollment. It is also reasonable to assume that admissions will show a lower degree of geographic diversity than applications,

because prospective students who are less qualified are likely to think that they may not be accepted at a CSU campus closer to their home, and thus are more likely to apply and not be accepted at a campus farther away. The geographic diversity of enrollees is likely to be less than that of students admitted to the University because the total cost of attending the University is lower for those living at home. (San Jose State has been a traditional "commuter" campus, although the University has placed greater emphasis on increasing its regional, national, and international attractiveness.) The University does attract students from across the State of California, and has significant enrollment from other states and foreign countries.

IV. Data

The student data on FTF in Fall 2005 was provided by the Office of Institutional Research at San Jose State University. This data included the street address of each applicant, whether admitted, whether enrolled, SCUs attempted in Fall 2005, whether enrolled in Spring 2006, and, if so, SCUs attempted in Spring 2006, whether enrolled in Fall 2006, and, if so, SCUs attempted in Fall 2006. Additionally, the students' ethnic self-identification, age, and gender were included. For students admitted but not enrolled in Fall 2005, the higher education institution in which they enrolled, if any, was also provided.

Cartographic boundary files for California Counties, census tracts, and census block groups were downloaded from the U.S. Census Bureau website (http://www.census.gov/) and data about a variety of economic and demographic variables were downloaded from the Census Bureau's American Factfinder website (http://factfinder.census.gov/home/saff/main.html?_lang=en).

V. Method

The method for analyzing the data will now be briefly presented. The essential step was to geocode the 16,386 street addresses. This was done by submitting batches of 500 addresses per pass to batchgeocode.com (http://www.batchgeocode.com/). As an option on batchgeocode.com, the straight line distance from the first address given in a batch can be computed. Each batch was submitted with 129 S. 10th St, San Jose, CA 95192 (the street address of San Jose State University at the corner of S. 10th Street and San Fernando St in San Jose) as the first address. Of the 16,386 addresses submitted, 16,342 were successfully geocoded, with latitude and longitude provided. (Of the 10,867 admitted students, 10,833 were successfully geocoded; of the 2513 enrolled students, 2,508 were successfully geocoded.) This geocoding was the sole basis for determining the county, census tract, or distance to campus of any record.

The geocoded student data was then spatially joined with county and census tract cartographic boundary files, and the resultant data was joined with data downloaded from American Factfinder. These operations generated two types of data sets: student data with county and census tract information attached to each record, and county and census tract data with counts of students (parsed by various factors) appended to the county and census tract data.

VI. Overview of the Data

A general overview of the data will now be provided. As Table 1 shows, six California counties account for the bulk of the applications, admissions, and enrollment at San Jose State University.

Geocoded Addresses by County for Top Six Counties	Apply		Admit		Enroll	
Total observations	16,342	**Percent**	10,833	**Percent**	2,508	**Percent**
Alameda	2,173	13.30	1,473	13.60	360	14.35
Contra Costa	714	4.37	491	4.53	94	3.75
Los Angeles	2,062	12.62	997	9.20	80	3.19
San Francisco	1,061	6.49	649	5.99	100	3.99
San Mateo	1,025	6.27	727	6.71	164	6.54
Santa Clara	4,384	26.83	3,419	31.56	1,316	52.47
		69.88		71.60		84.29

Table 1.

San Jose State University is located in Santa Clara County; San Mateo County is adjacent to Santa Clara County, and Alameda County is the next closest, followed by San Francisco County, and Contra Costa County. Los Angeles County is the farthest away – about 400 miles. Alameda County (which contains the City of Oakland), San Francisco, and Los Angeles are major population centers in California. This fact accounts for their presence on each of the lists, although there are large campuses of the California State University system closer to residents in those counties. Southern California CSU campuses are especially heavily impacted by growth of the college-going population, so students from Los Angeles and other Southern California Counties can be expected to be an increasing component of applications, admissions, and enrollment at Northern California CSU campuses like San Jose State University.

The geocoded data can also be viewed in terms of ethnic breakdown as shown in Table 2.

Ethnic Breakdown										
	Apply		Average Distance (miles)	Admit		Average Distance (miles)	Enroll		Average Distance (miles)	SCUsF05
Total observations	16,342	Percent	107.03	10,833	Percent	93.93	2,508	Percent	43.02	13.57
American Indian	88	0.54	142.76	57	0.53	118.32	9	0.36	48.56	13.67
African American	1,740	10.65	164.44	803	7.41	153.24	201	8.01	92.46	13.26
Asian	4,095	25.06	55.80	2,972	27.43	48.82	700	27.91	22.55	13.75
Filipino	1,312	8.03	62.97	961	8.87	57.72	299	11.92	26.49	13.42
Pacific Islander	161	0.99	89.87	95	0.88	72.92	31	1.24	22.94	13.35
Hispanic	4,158	25.44	145.11	2,311	21.33	129.98	491	19.58	43.33	13.45
White	3,691	22.59	112.60	2,842	26.23	109.41	573	22.85	58.52	13.65
Other	1,097	6.71	96.47	792	7.31	87.02	204	8.13	47.26	13.58
		100.00			100.00			100.00		

Table 2.

There are several notable features of the table. Asians are the largest single group admitted and enrolling in the University. (The immediate neighborhood of the University is heavily Vietnamese.) Asian applicants, admitted students, and attendees also live, on average, the closest to the university. For all groups, applicants live farther away than admitted students, and than enrollees. Among all students, African American applicants, admitted students, and enrollees live the farthest from campus. The average distance of 92.46 miles for African American enrollees indicates that a greater proportion of these students have had to move away from home to attend the University. (The geocoded addresses are the addresses from which the application was submitted – the "home address." Data on which students reside in dormitories or other non-home housing is not available.)

VII. Drop Outs

The increased focus on student success nationwide makes it imperative to analyze drop outs. The student data allows us to distinguish two kinds of drop outs: those who leave San Jose State University and subsequently enroll in another institution of higher education, and those who leave San Jose State and do not enroll in another institution. These two groups differ in a couple of important characteristics.

Table 3A shows data for all drop outs from Fall 2005 to Spring 2006. (Data on some groups was withheld if there were fewer than 15 observations.) Table 3B shows data for drop outs who enroll in another institution of higher education. Table 3C shows data for drop outs who do not enroll in another institution of higher education.

Drop Outs F05-S06		Percent of total observations	Average Distance (miles)	SCUsF05
Total observations	159		59.43	12.71
American Indian	[redacted]	[redacted]	9.00	12.00
African American	19	11.95	85.84	12.79
Asian	31	19.50	16.58	13.35
Filipino	[redacted]	[redacted]	29.50	13.50
Pacific Islander	[redacted]	[redacted]	22.33	10.33
Hispanic	42	26.42	43.57	12.24
White	35	22.01	119.97	12.46
Other	16	10.06	52.81	13.06

Table 3A.

Drop Outs Fall 2005-Spring 2006 who subsequently enroll in another institution		Percent of drop outs	Average Distance (miles)	SCUsF05
Total observations	101		75.07	12.95
American Indian	0	0.00	na	na
African American	[redacted]	[redacted]	87.93	13.28
Asian	23	74.19	19.96	13.22
Filipino	[redacted]	[redacted]	45.14	13.43
Pacific Islander	[redacted]	33.33	44.00	14.00
Hispanic	22	[redacted]	55.27	12.63
White	22	62.86	164.86	12.68
Other	[redacted]	[redacted]	57.42	12.75

Table 3B.

Drop Outs Fall 2005-Spring 2006 who subsequently do not enroll in another institution		Percent of drop outs	Average Distance (miles)	SCUsF05
Total observations	58		32.19	12.29
American Indian			9.00	12.00
African American			80.00	11.40
Asian			6.87	13.75
Filipino			7.60	13.60
Pacific Islander			11.50	8.50
Hispanic	20	47.62	30.70	11.80
White			44.00	12.08
Other			39.00	14.00

Table 3C.

The main conclusions from this data are that drop outs generally take on fewer SCUs than the average for their group, and drop outs live closer to the campus than non-drop outs in the same ethnic group. The lower level of SCUs attempted and the closeness to the campus is more marked for students who do not continue their higher education. This suggests that an early warning sign or "marker" of the likelihood of dropping out entirely is closer than average distance for the subgroup and lower than average SCUs attempted. The significance of using SCUs attempted as a marker should be highlighted. Data about SCUs attempted is available very early in the semester – much earlier than mid-term grades. This kind of information may be useful as part of an early intervention plan.

VIII. Profile of the FTF Compared with the Community

As noted above, more than 50% of FTF live in Santa Clara County. Geocoding and spatially joining census tract cartographic boundary files allows one to identify the census tract associated with a particular student's address. Census data can then be associated with the data file. One way of doing this provides a count of applicants, admitted students, and enrolled students in each census tract. Although this can be done

for all 7,115 census tracts in California, this analysis is confined to the 341 census tracts in Santa Clara County.

A wide variety of variables of interest are included in the SF3 (Sample File 3) data set available from the U.S. Census. The SF3 file contains responses to the "long form" of the census, and therefore represents about a 17% sample (compared to the theoretical 100% sample associated with the "short form" of the census contained in the SF1 file). The SF3 file provides a greater number of variables of interest.

Of particular interest at an institution like San Jose State, where many students are the first in their family to attend college, is the impact of peer groups and the community on their success in college. The SF3 file contains two kinds of educational variables of interest: variables that proxy the "target market" of potential college students, and variables that measure the educational attainment of people who reside in the census tract. The 2000 U.S. Census long form asked how many people age 3 or more were enrolled in grades 9 through 12 (in December 1999). Roughly speaking, it is from this group that First Time Freshmen would arise. It is possible to compute from the census data, the proportion of potential FTF in each census tract. With the results of geocoding, it is possible to determine the proportion of FTF at San Jose State who come from a particular census tract. It can then be determined whether a particular census tract is "overrepresented" or "underrepresented" in the profile of FTF at San Jose State compared with a random draw from the census tract.

Data in the SF3 file provides great detail on educational attainment of persons 25 and older, including estimates of the number of people with some college but less than one year, some college but one year or more (but no degree), an Associate's degree, a

Bachelor's degree, a Master's degree, a professional degree, and a doctorate. Based on these census variables, it is possible to construct an "exposure index" – a measure of the concentration of people with some experience of college in the census tract in which the student lives. There are numerous potential confounding factors in using such a measure. People with the level of education ascribed to the census tract may have moved out shortly after the census, or the student may have moved in to the census tract shortly before applying to San Jose State. But if both kinds of mobility are modest, the exposure index mentioned above – the percentage of people 25 and older with some experience of college (in December 1999) – should serve as some indicator of community experience of college (the presence of some kind of informal network of information).

The analysis involving these variables is presented in the following three maps.

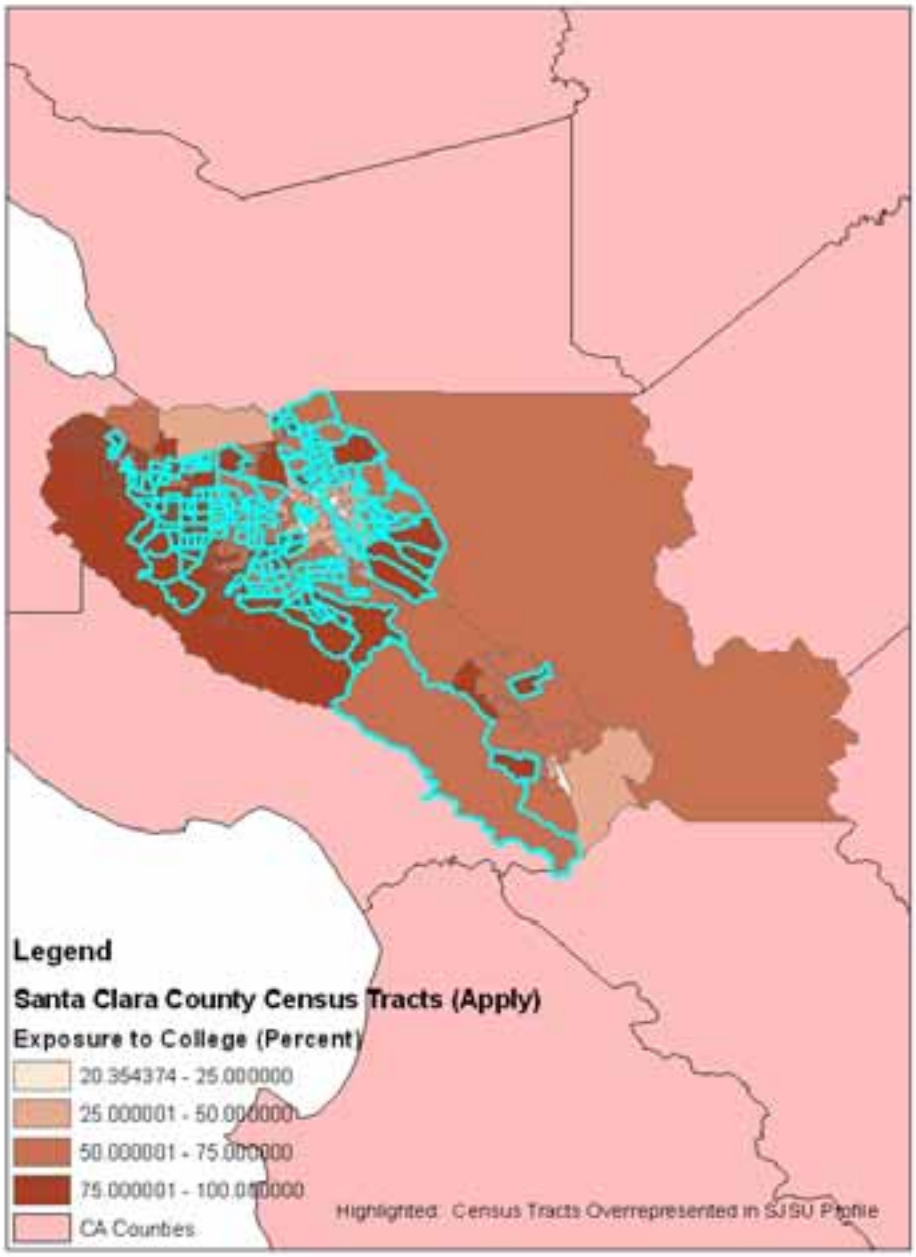
Legend
Santa Clara County Census Tracts (Apply)
Exposure to College (Percent)
20.354374 - 25.000000
25.000001 - 50.000000
50.000001 - 75.000000
75.000001 - 100.000000
CA Counties
Highlighted: Census Tracts Overrepresented in SJSU Profile

Figure 1.

Figure 1 shows the distribution of “exposure to college” across the census tracts of Santa Clara County. The highlighted census tracts are those from which a disproportionate number of San Jose State FTF applicants are drawn (given the target market calculations referred to above). One hundred and fifty of the 341 census tracts are overrepresented in the San Jose State profile of applicants, with 62.36% of applicants coming from such overrepresented census tracts. Generally, the census tracts fall in the middle of the “exposure to college” index. Many of the largest census tracts with high exposure to college ratings are those from which San Jose State draws relatively few students. However, San Jose State also draws relatively few students from census tracts with very low “exposure to college” ratings. (For applicants, admitted students, and enrollees, the average exposure to college index does not vary much, and is not very different from the County average exposure to college index. This is probably because of smaller family sizes in the census tracts with a higher exposure to college index.)

Figure 2 shows the same data for those admitted to San Jose State.

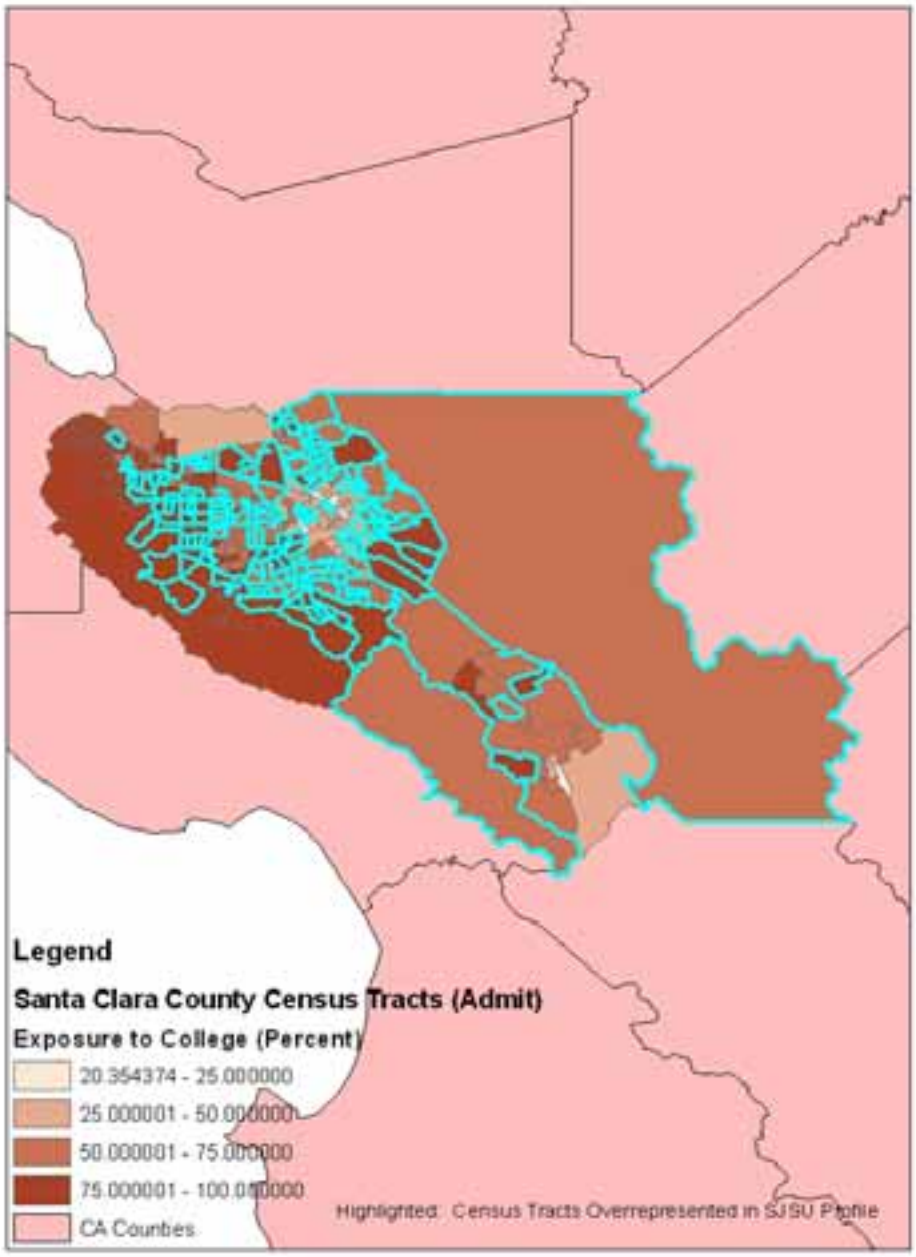
Legend
Santa Clara County Census Tracts (Admit)
Exposure to College (Percent)
20.354374 - 25.000000
25.000001 - 50.000000
50.000001 - 75.000000
75.000001 - 100.000000
CA Counties
Highlighted: Census Tracts Overrepresented in SJSU Profile

Figure 2.

For students admitted, there are 151 overrepresented census tracts, leaning slightly more toward greater exposure to college. More than 64% of students admitted to San Jose State come from these overrepresented census tracts.

Figure 3 shows the results for enrolled students.

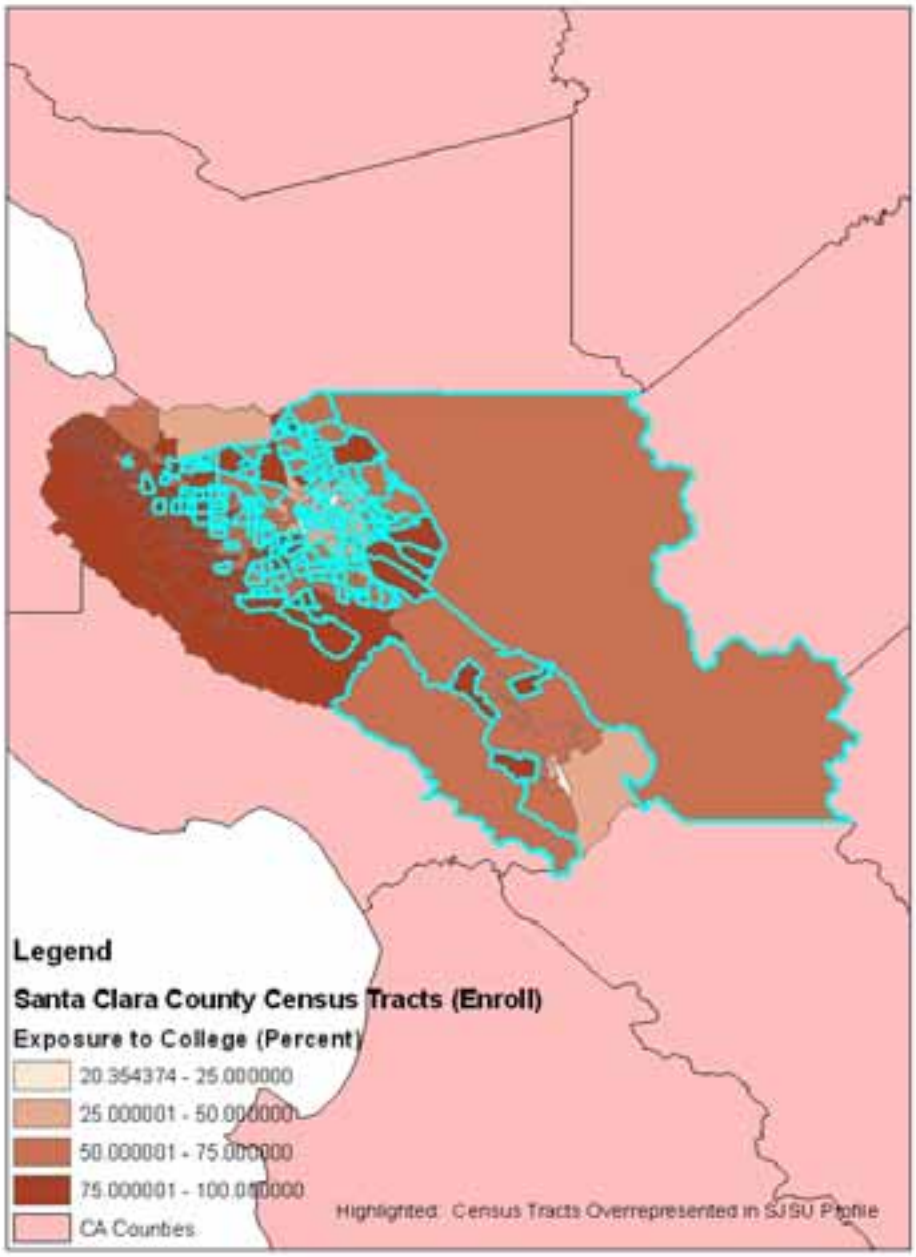
Legend
Santa Clara County Census Tracts (Enroll)
Exposure to College (Percent)
20.354374 - 25.000000
25.000001 - 50.000000
50.000001 - 75.000000
75.000001 - 100.000000
CA Counties
Highlighted: Census Tracts Overrepresented in SJSU Profile

Figure 3.

For enrolled students, there are 139 overrepresented census tracts, with a slightly lower average exposure to college than that for admitted students. Almost 71% of students from Santa Clara County come from these 139 overrepresented census tracts.

Similar methods can be applied in studying drop outs. Figure 4 give the distribution of drop outs across census tracts of Santa Clara County, with the drop outs who do not enroll in another institution of higher education highlighted. Figure 5 gives the spatial distribution of drop outs who do not enroll in another institution of higher education across census tracts of Santa Clara county by median income in the census tract.

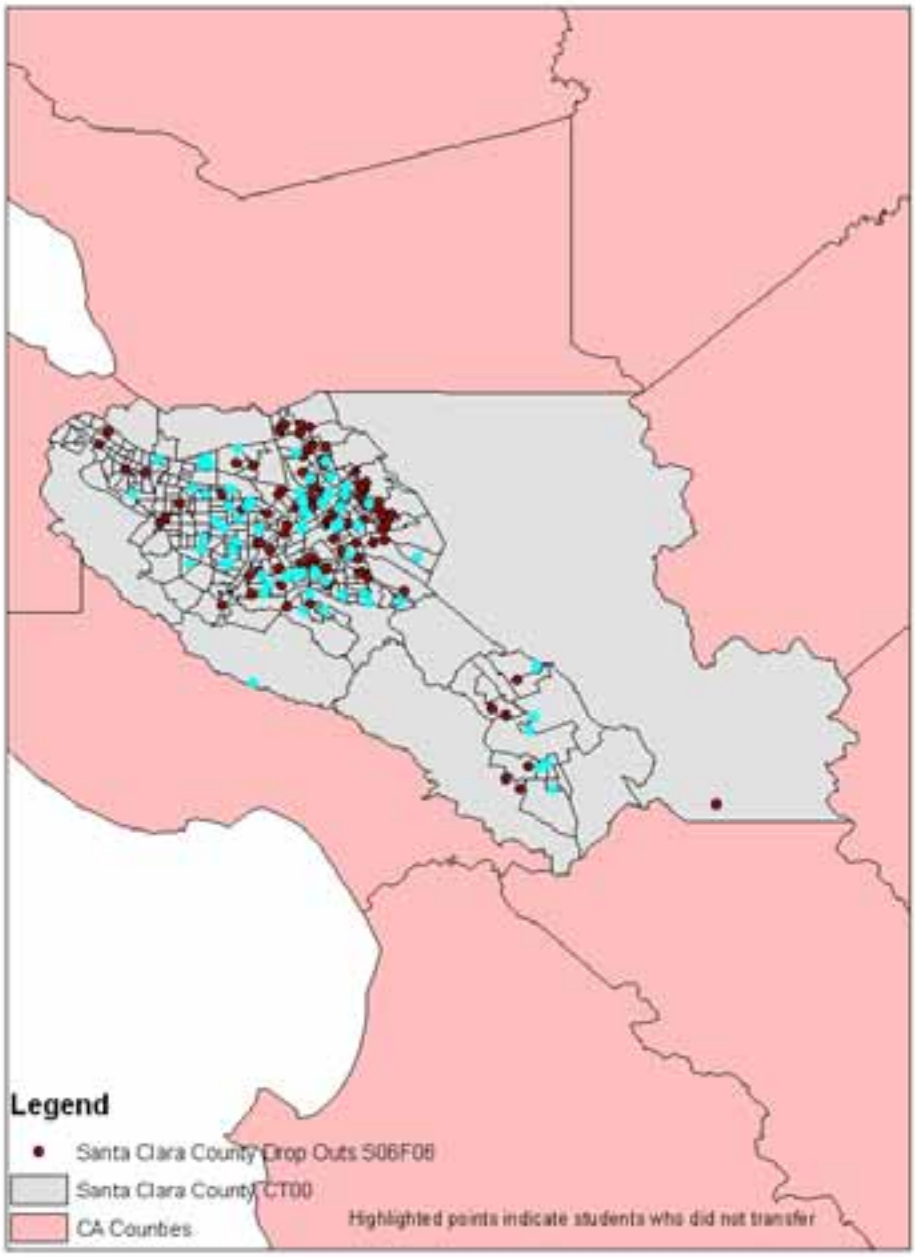
Legend
Santa Clara County Drop Outs S06F06
Santa Clara County CT00
CA Counties
Highlighted points indicate students who did not transfer

Figure 4.

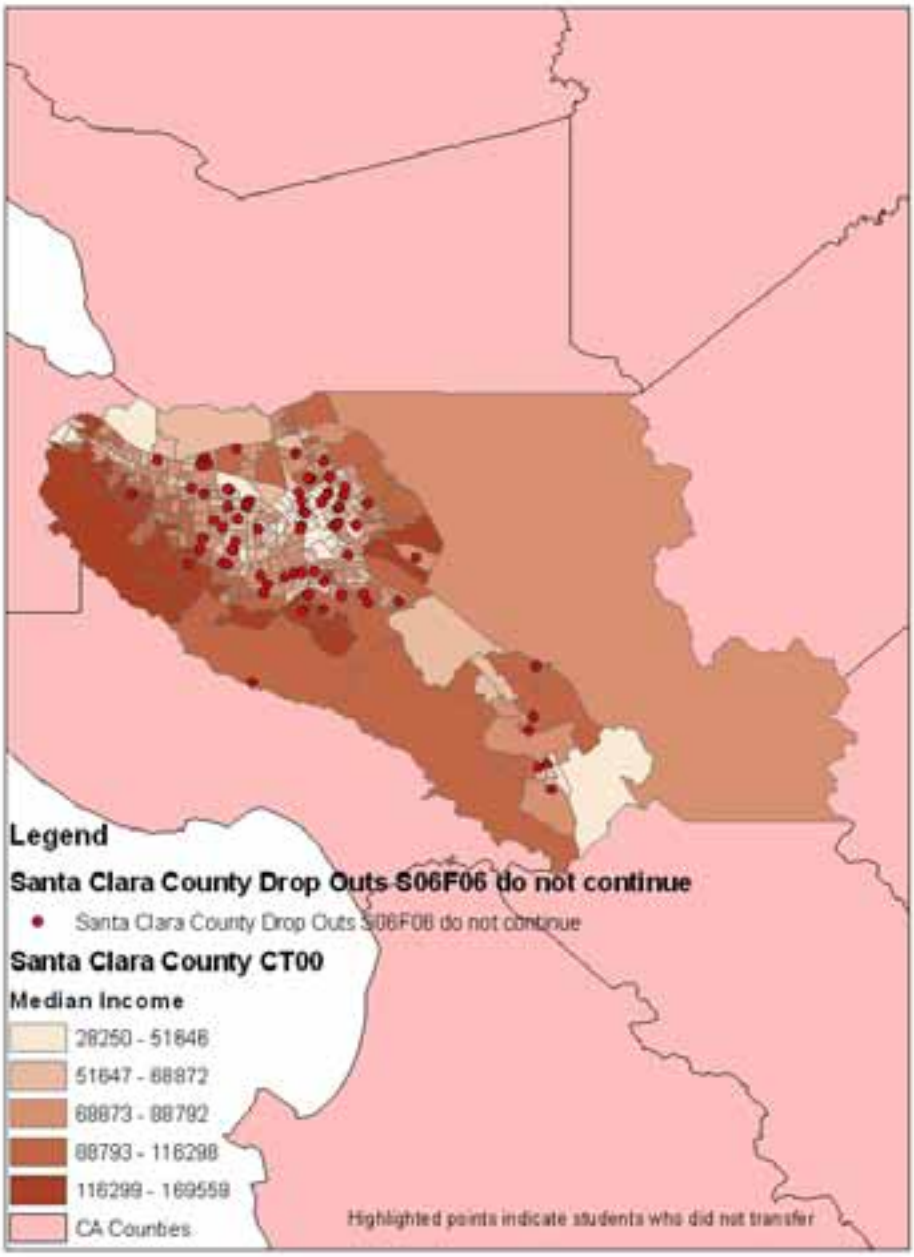
Legend
Santa Clara County Drop Outs S06F06 do not continue
Santa Clara County Drop Outs S06F06 do not continue
Santa Clara County CT00
Median Income
28250 - 51646
51647 - 68872
68873 - 88792
88793 - 116298
116299 - 169559
CA Counties
Highlighted points indicate students who did not transfer

Figure 5.

IX. Further Research

In future research, many other variables can be included. For example, Vasigh and Hamzaee (2004) suggest that tuition and fee payments may be a significant factor. Martin (2003) develops a model of optimal pricing. Another direction for further research would examine non-returning students. Toth and Montagna (2002) suggest that class size is an important factor in student achievement. This suggests that class size may be an important factor in determining whether students continue.

Future extensions of the project will examine additional counties in California encompassing areas from which virtually all applications arise (consisting of more than 12,000 census block groups). In addition to census data, regional authorities such as the Association of Bay Area Governments (ABAG) and the Metropolitan Transportation Commission (MTC), make available a large amount of data concerning the nine counties of the initial study (Association of Bay Area Governments (ABAG) (n.d.), Metropolitan Transportation Commission (n.d.)).

Bibliography

Ahlburg, D., McPherson, M., and Schapiro, M. (1994). Predicting Higher Education Enrollment in the United States: An Evaluation of Different Modeling Approaches. Williams Project on the Economics of Higher Education, Discussion Paper No. 26 (August)

Arctur, D. and Zeiler, M. (2004). *Designing Geodatabses: Case Studies in GIS Data Modeling.* ESRI Press.

Association of Bay Area Governments (ABAG), http://www.abag.ca.gov/

Batch Geocode (n.d.), http://www.batchgeocode.com/

DesJardins, S.,Ahlburg, D., and McCall, B. (2006). An Integrated Model of Application, Admission, Enrollment, and Financial Aid. *The Journal of Higher Education*, Vol. 77, No. 3 (May/June), pp. 381-429

Gorr, W. and Kurland, K. (2005). *GIS Tutorial: Workbook for ArcView 9*, ESRI Press.

Huxhold, W., Fowler, E., and Parr, B. (2004). *ArcGIS and the Digital City: A Hands-on Approach for Local Government.* ESRI Press

Marble, D., Mora, V. and Granados, M. (1997). Applying GIS Technology and Geodemographics to College and University Admissions Planning: Some Results From The Ohio State University, Proceedings of the ESRI International Users Conference. http://gis.esri.com/library/userconf/proc97/proc97/to550/pap501/p501.htm

Marble, D., Mora, V., and Herries, J. (1995). Applying GIS Technology to the Freshman Admissions Process at a Large University. Proceedings of the ESRI International Users Conference. http://gis.esri.com/library/userconf/proc95/to200/p182.html

Martin, R. (2003). Pricing and Enrollment Planning. *Planning for Higher Education*, vol. 31, no. 4, June-August, pp. 29-37

Metropolitan Transportation Commission. http://mtc.ca.gov/

Toth, L. and Montagna, L. (2002). Class Size and Achievement in Higher Education: A Summary of Current Research. *College Student Journal*, vol. 36 no2.

U. S. Census, Census Block Group Boundary Files, (1990). http://www.census.gov/geo/www/cob/bg1990.html

U. S. Census, Census Block Group Boundary Files, (2000). http://www.census.gov/geo/www/cob/bg2000.html

U.S. Census Summary File 3 (SF3) (2000). http://www.census.gov/Press-Release/www/2002/sumfile3.html

Vasigh, B. and Hamzaee, R. (2004). Testing Sensitivity of Student Enrollment With Respect to Tuition at an Institution of Higher Education. *IAER (International Atlantic Economic Review*, (May), Vol. 10, No. 2, pp. 133-149

* Thanks to Dr. Sutee Sujitparapitaya, Associate Vice President of Institutional Research at San Jose State University, for data, encouragement, and helpful comments. Thanks to an anonymous reviewer for thoughtful comments on the abstract.