ACADEMIC SUCCESS FACTORS IN K-12 EDUCATION:
A QUANTITATIVE ANALYSIS

By

Brianna Chew

A capstone project submitted for
Graduation with University Honors

May 10, 2018

University Honors
University of California, Riverside

APPROVED

_____
Dr. Asish Satpathy
Department of Business

_____
Dr. Elaine Wong
Department of Business

_____
Dr. Richard Cardullo, Howard H Hays Jr. Chair and Faculty Director, University Honors
Interim Vice Provost, Undergraduate Education

Abstract

Standardized test scores are often used to measure students' academic success. Although factors that affect student success are teaching techniques and study skills, there are other factors outside the classroom that also influence student performance. Oftentimes, these factors are overlooked or easily deemed uncontrollable by school leaders. I develop a hypothesis that demographic and socioeconomic factors could play an important role in students' academic success. In this study, I will attempt to confirm or refute my hypothesis to understand key factors that could potentially be controlled to improve students' standardized test results. Some demographic and socioeconomic variables that I will investigate are household sizes, number of multigenerational households, average income and education obtainment within an area.

As time progresses and the academic achievement gap between the low and high socioeconomic communities widen, I hypothesize that family income could be correlated to standardized test results like the SAT. I also assume that there are particular demographic variables, such as the household size, that will have an effect on academic success for students. My overarching hypothesis is that the spending habits, socioeconomic trends, and demographic facts about an area all contribute to the current academic achievement gap.

This study uses a sample of ACT and SAT test scores from the California Department of Education to perform statistical tests to understand the influencing factors on test scores. I use industry software applications such as ArcGIS Online and Business Analyst Online to perform data visualization. Statistical software JMP Pro is used to perform statistical analysis. I test the hypothesis using a sample of average SAT scores collected for 1,070 schools located in California. I will then acquire demographic and lifestyle data of each of the school's surrounding neighborhoods. Based on my studies, I conclude that family income is indeed an influential factor of students' academic achievement. I also learn that single parent and multigenerational households have negative impacts on student test scores. I believe that this study will shed light on how to proactively take action to improve academic performance.

Acknowledgments

I would like to express my gratitude to my faculty mentors, Associate Dean Elaine Wong and Professor Asish Satpathy from the School of Business, for guiding and mentoring me throughout this research project. It has been a pleasure to have worked with and learn from such exceptional teachers. I would not be where I am today without my mentors' care, expertise, and dedication.

I would also like to express my gratitude to my family and friends for their support. They were always there for me when I needed that extra little push during those late nights of researching and writing.

Lastly, thank you to the Starbucks at Glen Mor for always conveniently providing Savannah Sprague and me with the large square tables to sit at as we attempted to diligently worked on our capstones.

Table of Contents

Introduction

Students from different demographic and socioeconomic statuses continue to show performance gaps in academic excellence. In fact, for the past fifty years, the academic performance gap has grown bigger as the gap between the rich and poor widens too (Reardon, 2011). For many decades, researchers have been attempting to find an explanation for this problem and so far, most studies conclude students with lower socioeconomic statuses tend to perform worse than those with higher socioeconomic statuses. Scholars have used U.S. administrative data and longitude surveys to discover that factors like median income, parental education, and family structure influence students' academic success (Ladd, 2012). Even though prior research has demonstrated that these general variables affect student success, we need to understand if this trend has changed recently thanks to the availability of various resources that promote academic excellence in underprivileged communities. Thus, the real question is, what household investments influence students' scholarly achievements? Understanding the differences between low and high academic performing communities will provide educators with more nuanced data, and in turn, enable them to create solutions for certain school areas. If low performing school districts can learn what high performing school districts are doing, we will be able to witness more of its pupils grow into well-educated young adults.

Instead of observing national trends and using panel studies, I will restrict this study to observe only the trends within a single state to produce a more personalized conclusion. The demographic and socioeconomic data will be gathered from Environmental Science Research Institute (ESRI), a geographic information systems (GIS) company that hosts its location-based demographic and lifestyle data from a combination of third-party resources, such as U.S. Census reports and American Community Surveys (www.ESRI.com). By using ESRI's ArcGIS Online

platform, I can collect demographic and population lifestyle data specific to each school location (ESRI's ArcGIS Online, 2018). I will also be able to visually compare and contrast the numerous gaps between academic success and variables such as income and family dynamic. With statistical software, JMP Pro (Version 13.0; Hodgson, 2016), I will be able to test all of the variables' data through various statistical methods and produce to create a prediction model for student academic success. Not only will I determine which demographic and socioeconomic variables contribute to the academic achievement gap, but also how much of a positive or negative impact they have on students' tests results.

Students' academic success can be measured by looking at their performances on standardized tests such as the Scholastic Assessment Test (SAT) and American College Test (ACT) (Camara & Echternacht, 2000). The purpose of these exams is to predict college freshmen grade point averages (GPA); thus, providing undergraduate admission boards with a common metric to use to compare applicants (The Princeton Review, 2018). In, "A Mathematician Reads the Newspaper," by John Allen Paulos, he examines SAT scores and argues that "the [SAT] test is 'biased,' but only toward the educationally prepared, the physically healthy, and psychologically receptive." He further notes, "Just as there are many dimensions of football ability that aren't measured by poundage, there are many dimensions of scholastic ability that aren't measured by the SAT" (Paulos, 2013). In conjunction with Paulos's argument, Olaf Jorgenson focuses on California Standards Test (CST) and California High School Exit Exam (CAHSEE) and expresses his concerns in his academic paper called, "What We Lose in Winning the Test Score Race." Jorgenson argues that standardized test scores do not measure a student's creative ability to "explain, debate, elaborate, present, rebut, or improvise" (Jorgenson, 2012). Tests like the CST, CAHSEE, and SAT do not look at the students' "perseverance, resiliency, and determination"

(Jorgenson, 2012). Because these standardized tests follow the multiple-choice format, they imply that there is only one right answer, which makes it hard to recognize students' innovative ways of thinking. Jorgenson elaborates that the traits that these tests do not measure, such as creativity and entrepreneurial leadership, are traits that companies look for in their future employees. Although Paulos and Jorgenson have a strong argument, traits like creativity and leadership are hard to measure. However, concepts like unconventional problem solving, basic math and English foundations can easily be measured through standardized tests like the SAT and ACT.

Both the SAT and the ACT have been proven to correlate with academic success in college. In 2000, College Entrance Examination Board attempted to verify how appropriate it is to use SAT scores and high school grade point averages (GPA) as predictors of college readiness. This paper concluded that SAT scores and high school GPAs do have a significantly high validity coefficient[1] of 0.65 (Camara & Echternacht, 2000). High school GPAs alone is a strong predictor of college GPAs; however, the combination of high school GPAs and SAT scores raised the validity coefficient by 0.1 (Camara & Echternacht, 2000). This implies that SAT scores do indeed add weight when determining student academic success. Furthermore, the College Entrance Examination Board discovered that the SAT predicts college readiness, with a strong validity coefficient equally amongst all ethnicities (Bridgeman, McCamley-Jenkins, & Ervin, 2000). However, these two predictor variables are more accurate for students from high socioeconomic backgrounds than low socioeconomic backgrounds and the same trend follows for parent education obtainment (Bridgeman, McCamley-Jenkins, & Ervin, 2000). Regardless, standardized tests like the SAT and ACT do play a role in predicting academic success.

[1] "Validity coefficient is a statistical index used to report evidence of validity for intended interpretations of test scores and defined as the magnitude of the correlation between test scores and a criterion variable" (Salkind, 2007).

As discussed above, previous studies have already demonstrated a relationship between SAT and ACT scores and academic success in post K-12 education. This study will focus on the determining which demographic and socioeconomic factors in California influence SAT results. Specifically, I will focus on family income, family size and living structure, and school location. The SAT and ACT will act as a metric to measure student success.

Previous studies have shown that students who read more books have the opportunities to learn and retain more advanced vocabulary words (Duff, 2015). With that being said, learning vocabulary words that are not used in day to day conversations could potentially better prepare students for the reading and writing sections of the SAT. If the amount of money spent on books per household is greater than the national average, it can be said that children within that area could be more exposed to reading leisurely and have the ability to score higher on the SAT than other students. Existing studies also lead us to predict that households that purchase more musical instruments and accessories perform better on the SAT, more specifically in the math section. After analyzing SAT scores over a ten-year period, a study concluded that students who participated in any art courses, such as orchestra or band, scored substantially higher than those students who did not participate in any art courses throughout their high school career (Vaughn, 2000). From this discussion, I predict:

> Hypothesis 1: Income can heavily influence how successful students are academically. Specifically, parents who invest in reading books and musical instruments for their children will increase their offspring's rates of success, more particularly in reading and mathematics.

As mentioned earlier, I hypothesize that certain family structures will have an impact on how well students perform on the SAT. There is evidence to presume that single parent households

affect students' academic performance negatively. Although the children who are part of a single parent family build characteristics like resiliency and independence, they tend to score lower on standardized tests, graduate from high school, and attend college at a lower rate (Barajas, 2010). I also predict that other family arrangements, such as multigenerational households, will have a negative relationship with SAT scores. Researchers have analyzed the effects of living in residences, similar to multigenerational households, and revealed that those particular households have fewer resources and are generally the ones the who need the most support (Augustine & Raley, 2013). Statistics also show that multigenerational households are typically families with lower income and only a high school education (Bethell, 2011).

Hypothesis 2: Larger household sizes or families living with extended family members, will have a negative effect on student performance in school.

This paper is split into two studies. Study 1 uses a small representative sample of schools, one from each county of California. Study two uses a larger sample of all public schools in California. These studies will observe the various spending habits, consumer behaviors, and household demographics at each location of the schools selected in the sample. To assess the research question, the first study will use linear regression and correlation analysis and due to the increase in the number of variables that will be analyzed in the second study, I will use stepwise regression to produce a prediction model. I will gather the predictor variables' data from ESRI using their cloud-based applications, such as ArcGIS Online and Office products as well as Business Analyst Online. With the data that I obtained, I will perform visualizations using ArcGIS Online and statistical analysis using JMP Pro.

Overview

In this paper, I deliver two studies. The first study uses a sample of 56 schools to predict which demographic and socioeconomic factors affect student success. I use regression and correlation methods to determine the results. The second study examines every public high school in California to predict which of the same factors studied in the previous experiment affect student success. I use stepwise regression analysis to test my hypothesis.

Study 1

Methods

Sample. One school from each of the 58 counties in California was chosen to represent the entire high school population in the state of California. Data on SAT and ACT scores were collected from the California Department of Education DataQuest. Alpine and Sierra county were the only counties that did not have any data on file, so those counties were not included in this study. Each of the 56 schools in the sample was selected based on the total population in grade 12. The highest populated high school in the largest school district of each county was chosen to be one of the 56 schools in the sample. Map 3 and 4 are a visual representation of the sample where each circle represents one of the 56 schools.

In this study, the SAT and ACT are used to measure student academic excellence. The 3-hour long SAT is divided into three subjects - math, reading, and writing. Until 2016, the full amount of points that could be scored was 2400. As of 2018, the maximum points for the SAT is 1600 (The Princeton Review, 2018). During the 2015-2016 academic school year, the average scores for math, reading, and writing in California were 494, 484, and 477 respectively (California Department of Education DataQuest, 2017). Students tend to perform slightly better in math than

reading and writing. The reading section of the exam focuses on the student's ability to make inferences, distinguish relationships, and synthesize graphics (College Board 1, 2018). The writing segment differs substantially from the reading portion because it measures the student's knowledge of punctuation, sentence structure, and ability to incorporate information (College Board 2, 2018). The final portion of the exam tests the student's intelligence in mathematical concepts relating to complexity, geometry, probability and statistics, and trigonometry (College Board 3, 2018).

The ACT is divided into four sections –English, math, reading and writing and science. The English, math, and reading and writing sections are quite similar to the SAT in terms of concepts. In the science section, students are tested primarily on their critical thinking skills and not necessarily on their knowledge of general science (ACT, 2018). This particular exam scores each of the sections out of 36 points, where the overall score is the mean of each of the sectional composite scores. During the 2015-2016 academic school year, the average score for all four subjects was 22 out of 36 (California Board of Education DataQuest, 2017).

Procedure and Design. The data collected from the California Department of Education database included these variables: total students in grade 12; number of students who took the SAT/ACT exam; the average score on reading, math, and writing; and the number of students who scored at least a 1500 out of 2400. After collecting data for each of the 56 schools in California, their latitudes and longitudes were imported into ESRI's ArcGIS Maps for Office (ESRI's ArcGIS Maps for Office, 2017). By using ArcGIS Maps for Office, I can retrieve appropriate demographic and socioeconomic data from a cloud of resources (ESRI 8, 2017). By default, ESRI supplies the data within a 1-mile radius of each specific location in the sample through a process that is defined as data enrichment. However, running analysis on a 1-mile radius may not be an accurate

representation of each local community. Thus, 1, 2, and 3-mile radii rings were set around each of the schools in the sample and data enrichment within each radius ring was conducted.

Through data enrichment, ArcGIS Online can help users receive data on spending habits, population, businesses, education, and much more. In this study, I focus on specific variables within the spending habits, socioeconomic, and demographic categories. Because previous research has already shown that family income has a major impact on how students perform academically, I wanted to look at what particular spending habits actually have an impact on SAT scores. I chose some of the most common activities that families spend money on, such as pets and online games. Refer to Appendix I for details on how average spending habits are calculated. As for demographic and socioeconomic variables, I chose those that oftentimes have negative connotations, such as number of divorced couples and number of households that fall below the poverty line. ESRI provides statistics on those variables by collecting U.S. Census and the American Community Survey results. After finding and retrieving data specific to each of the 56 locations, through data enrichment, maps were created in order to visualize the relationships between the exam scores and single attributes chosen from ESRI. All of the data were inputted into JMP Pro, where statistical analysis was completed.

This study will enrich each school with these variables:

Household Spending Habits	• Average Spending on Alcoholic Beverages • Average Spending on Online Entertainment/Game • Average Spending on Books • Average Spending on Health Insurance • Average Spending on Entertainment/Recreation - Pets • Average Spending on Cable and Satellite TV Services • Average Spending on Test Prep & Tutoring Services • Average Spending on Education (Annual) • Average Spending on Musical Instruments & Accessories
Socioeconomic Variables	• Average Household Income • Number of Households with Income Below Poverty Level • Education Attainment: Bachelor's Degree • Employed Civilian Population of Age 16+
Demographic Variables	• Average Household Size • Number of Owner Occupied Housing Units • Number of Multigenerational Households • Number of Husband-wife Family Households • Population of Divorced Couples • Dominant Tapestry Segment[2]

[2] Tapestry segments describes American neighborhoods based off of demographic and socioeconomic data.

Results

Before investigating specific demographic and lifestyle variables, I examined the dominant tapestry segments of the top scoring and bottom scoring schools in the sample to get a better idea of what type of neighborhoods these schools are located in. Then, I proceeded to enrich the sample with 18 different variables and import the data into JMP Pro to perform multilinear and simple linear regressions. The same analysis was done for ACT scores. Variables that did not maintain a *p*-value less than 0.05 were eliminated from the study and assumed to not affect students' academic success. Sometimes, running multilinear regression may not be sufficient because an independent variable's association with the dependent variable may be influenced by another independent variable. With that being said, it was important to run the simple linear regression between each of the variables against SAT and ACT scores. Each regression analysis was conducted at a 1, 2, and 3-mile ring radii around each school location. I also use ESRI ArcGIS Online to provide dominant tapestry segments for each of the neighborhoods in this sample.

To conclude the pilot study, the correlation was tested to determine the various *r*-values, which measures the strength of the linear association between two variables. All 18 variables were tested together to see individual correlations. To understand the correlations even further, a correlation analysis was conducted among the top 12 significant variables chosen from the multilinear regression model.

Dominant Tapestry Segment. ESRI has created a geodemographic system that assigns a tapestry segment for any given area based off of its demographic and socioeconomic factors (ESRI's ArcGIS Maps for Office). The tapestry segment summarizes information such as gender, age, income, types of jobs within the area, and much more. There are a total of 68 segments that describe American communities (ESRI 4, 2017). Figure 1 shows the dominant tapestry count in

California for all the 56 schools in this sample. The top tapestries within the sample are Up and Coming Families, Set to Impress, and Old and Newcomers (Figure 1). Up and Coming Families consists of hard-working young families that have a college education and earn a median household income of approximately $72,000 (ESRI 1, 2018). Set to Impress is a community mostly filled with singles living in apartments with a median household income of $32,800 (ESRI 2, 2018). Unemployment rates tend to higher around these areas and the number of educational degrees obtained tend to be lower. Lastly, Old and Newcomers, are areas with populations of around 2.5 million people who are either single or have a small family (ESRI 3, 2018). Their median household income can be around $44,900.

The top 3 performing schools in the sample, University High School, Leland High School, and Redwood High School, had the highest percentage of students scoring above a 1500 on the SAT, 88.29%, 88.09%, and 83.17% respectively. The tapestry segment of these 3 locations are Dorms to Diplomas, Top Tier, and Urban Chic, all of which have very similar traits such as higher income and a more luxurious family lifestyle. Their median income ranges from $17,000-$157,000 (ESRI 4, 2017). Two of the three communities are surrounded by higher education institutions, such as University of California, Irvine and San Jose State University, which may give students more of an incentive to perform well on the SAT. Universities like UC Irvine invite students from local middle and high schools to tour campus and host summer programs such as COSMOS to get younger students excited about becoming a college student. Students who live in these areas have the financial capability to enroll into these summer programs, acquire test preparation or other tutoring services, and even participate in extracurricular activities such as sports or music lessons to keep them proactive about learning. In addition, within their population, pursuing a degree in college is very common.

The lower performing schools in the sample, were Pacheco High School, Pacifica High School, and John F. Kennedy High school, where only about 17%-19% of their students scored above a 1500 on the SAT. The dominant tapestry segments for these lower performing areas are Up and Coming Families, Urban Villages, and Pleasantville. These three segments share the same type of attributes, such as low income, multigenerational households, and lower education attainment levels (ESRI 4, 2017). The median income around these schools ranged from $26,000-$58,000, which is considerably different from the top three school's median income (ESRI 4, 2017). The tapestries describe these neighborhoods as places that are just starting to become more developed. They may not have the same amenities, such as test preparation centers and extra tutoring programs. There are typically more multigenerational households and blue-collar workers within the area (ESRI 4, 2017). Students living in these neighborhoods may not have the same opportunities as other students to participate in summer programs, local sport leagues, or private music lessons. Lastly, graduating from college, let along high school is not a common trend among these communities.

Based on the tapestry segmentations of the top and bottom scoring schools, a trend was uncovered. The lower performing schools are located in lower socioeconomic areas and the higher performing schools are located in the higher socioeconomic areas. This analysis enabled me to infer that variables such as income and education obtainment do play a role in student's academic success.

Multilinear Regression at 1-Mile. The SAT scores were tested against all 18 variables at the same time. The model produced $R^2 = 0.80$ and $p < 0.0001$, meaning the model explains 80% of variability in the data. Although the 18 variables combined proved to be very significant, there were only four variables had an individual p-value less than 0.05. Those four variables are 2017

population of divorcees, 2017 owner occupied housing units, average amount spent on online entertainment and games, and average amount spent annually on education, all of which have more of an impact on SAT scores than other variables (Refer to Table 1 and Figure 4).

In regard to the ACT, the results did not digress much from the SAT sample. About 81% of the combined variables fit the linear model with a *p*-value of less than 0.0001. However, for this sample, the most significant variables occurred to be the 2017 population of divorcees, average amount of spending on online entertainment and games, 2017 owner occupied housing units, average spending on musical instruments and accessories, and the number of people with bachelor's degrees (Refer to Table 2 and Figure 5).

Multilinear Regression at 2-Mile Radius. The *p*-values noticeably changed when the multiple linear regression model was conducted at the 2-mile radius in regard to SAT scores. Less of the data points fit into the linear model as shown in Figure 6. The only significant variable was the population of divorcees in 2017 with a *p*-value of 0.04432. According to this particular model, 76% of the data can be explained by the linear model and the overall *p*-value still holds to be less than 0.0001, which is still highly significant (Refer to Table 3).

Multilinear Regression at 3-Mile. The *p*-values, once again, changed when the multiple linear regression model was conducted with the 3-mile radius data in regard to SAT scores. Even less of the data points fit into the linear model than the previous 2-mile test as shown in Figure 7. There were three significant variables, those being the population of divorcees in 2017, average spending on education annually, and the population of employed civilians over the age of 16. According to this regression model, 76% of the data fits the linear model and the overall *p*-value still holds to be less than 0.0001, which is highly significant (Refer to Table 4).

Simple Linear Regression at 1-Mile Radius. When each of the 18 attributes were individually tested against the SAT test scores, 12 of the 18 variables proved to be significant rather than five variables as seen in the multiple linear regression (Refer to Table 5). The 12 variables, from the most significant to the least significant, are:

- Average Household Size
- Average Spending on Education Annually
- Average Spending on Cable and Satellite TV Services
- Average Spending on Alcoholic Beverages
- Number of Multigenerational Households
- Average Spending on Health Insurance
- Average Spending on Pets
- Average Spending on Books
- 2017 Average Income
- Average Spending on Online Entertainment/Games
- Average Spending on Musical Instruments
- Average Spending on Test Prep & Tutoring

When each of the 18 attributes were individually tested against the ACT test scores, the same 12 variables that held a *p*-value less than 0.05 in the SAT sample proved to be significant in the ACT sample as well. However, the *p*-values do slightly differ (Refer to Table 6).

In both tables, the R^2 value tells you how much of the data is dependent on one another. The average household size and spending on education annually have R^2 values of 0.35 and 0.28 respectively. In other words, the number of students who score above a 1500 on the SAT is dependent on the average household size 34.6% of the time and approximately 28% of the data of SAT scores depends to spending on education annually.

Simple Linear Regression at 2-Mile Radius. As expected, the same variables that had a *p*-value less than 0.05 in the single variable regression at 1-mile radius proved to be significant again in the simple linear regression model at the 2-mile radius. The R^2 and *p*-values do vary; however, the order of impact of each variable still persists (Refer to Table 7).

Simple Linear Regression at 3-Mile. Once again, the same variables that had a *p*-value less than 0.05 in the 1 and 2-mile simple linear regression models proved to be significant in the 3-mile simple linear regression as well. However, their ranks of significance did change (Refer to Table 8).

I compared the R^2 and *p*-values between the 1, 2, and 3-mile radii and noticed that, at either radius, it showed good R^2 and *p*-values. However, the significance of each individual variable had changed. While running single regression tests on each of the variables at the different radii, I found that the same 12 variables continued to remain influential factors in determining SAT scores. With that being said, SAT results do depend on these 12 reoccurring factors.

Household Spending Habits	• Average Spending on Alcoholic Beverages • Average Spending on Online Entertainment/Game • Average Spending on Books • Average Spending on Health Insurance • Average Spending on Entertainment/Recreation - Pets • Average Spending on Cable and Satellite TV Services • Average Spending on Test Prep & Tutoring Services • Average Spending on Education (Annual) • Average Spending on Musical Instruments & Accessories
Socioeconomic Variables	• Average Household Income
Demographic Variables	• Average Household Size • Number of Multigenerational Households

All of the household spending habit variables survived each of the regression analysis, while only three of the demographic and socioeconomic variables combined survived. Based on off these results, there is support for hypothesis 1 and hypothesis 2. The model shows that average household spending on books and musical instruments and accessories have a significance when predicting SAT scores. It also shows that the size of a household plays a significant role in determining students' academic success. However, because each of the variables in the household spending habits remained significant, it suggests an issue with multicollinearity. I investigate further by analyzing the variables' correlation.

Correlation. To confirm whether there is an issue of multicollinearity or not, I tested the correlation between each of the 12 significant variables from the simple linear regression models. Figure 2 is a scatterplot matrix that illustrates the correlation coefficient between each of the variables. The narrower a single ellipse is, the closer the *r* value is to -1 or 1. If an ellipse appears to be wide, the correlation coefficient is closer to 0. Briefly glancing at the scatterplot matrix in Figure 2, one can see that majority of the variables have high correlations to one other. I compiled the ellipses with high *r* values and created another scatterplot matrix seen in Figure 3. In this matrix, all of the variables fall into the category of household spending habits. Thus, this proves that there is indeed an issue with multicollinearity within this model. When a model experiences multicollinearity, it becomes harder to assess how much of an effect the predictor variables have on the response variable, in this case the SAT scores.

SAT Scores vs. Average Income. To conclude the Study1, five variables were chosen for further examination. Map 4 features two layers, percentage of students who scored above 1500 on the SAT in orange and the average household income within a 1-mile radius of each school in purple. The majority of the higher income areas are centered around the San Francisco area. The

regression model for average income and number of students with SAT scores above 1500 shows an R^2 value of 0.18. Families that make a fair amount of money are able to give their children an advantage when it comes to academic success. Participating in music lessons, sports, community groups like Girl Scouts or Boy Scouts, can all encourage students to become critical thinkers; however, extra-curricular activities such as these are not affordable for every family.

SAT Scores vs. Average Spending on Musical Instruments. Music can be very beneficial to a student's academic success. In the more recent years, California has seen some school districts struggle to maintain their music programs. Through ESRI, I enriched all 56 points in the sample with the variable of average spending on musical instruments and accessories (Map 5). The R^2 value states that 12.4% of the SAT scores can be explained by the amount of money spent on musical instruments. The *p*-value of 0.0073, also proved spending on musical instruments to be fairly significant when determining student academic success. Furthermore, the average amount spent on instruments has an even stronger positive relationship and positive correlation with SAT and ACT Math, Reading, and Writing subject test scores individually according to the correlation analysis. Possible solutions to increase music participation are to ensure school districts continue to strive to fund music and art programs at all grade levels, put an equal amount of emphasis on the Arts in addition to STEM[3] subjects, and to have the local community conduct outreach programs to encourage students to explore the wonders of music.

SAT Scores vs. Average Spending on Books. An integral part of learning is reading. Studies have shown that students who read more frequently tend to be more intelligent in terms of vocabulary and verbal skills (Duff, 2015). Using ArcGIS Maps for Office, I enriched the sample with average amount spent on books per household within the selected area. The R^2 value states

[3] Type of education defined by the U.S. Department of Education to foster education for global leadership by supporting students and teachers in Science, Technology, Engineering, and Math

that 18.8% of the SAT scores can be explained by the amount of money spent on books. The recorded *p*-value, 0.00084, also proved to fairly significant, which means that average spending on books does have a linear relationship with the SAT scores. Furthermore, the average amount spent on books has an even stronger relationship and positive correlation with the SAT and ACT Reading and Writing subject test scores. It is likely that the more words students are exposed to through reading books and newspapers, the more words they will recognize on the SAT and ACT reading sections. Solutions to increase exposure to books in order to increase SAT scores could entail encouraging students to start reading at a younger age, implementing "Silent Sustained Reading" [4] teaching methods in classrooms, and assigning readings that are more likely to catch students' attention. Refer to Map 6 for a visual comparison between these two variables.

SAT Scores vs. Average Spending on Education. It has been shown that education spending has a positive impact on student success. Some households invest a significant amount of money on test preparation materials and courses in order to help students reach their full potential. When SAT scores were tested against average spending on education, the R^2 value stated that 27.8% of the SAT scores can be explained by this particular independent variable. The recorded *p*-value, 0.0003, also proved to be extremely significant, which means that average spending on education has a linear relationship with the SAT scores. Moreover, education spending has a correlation value of 0.53, which makes this particular relationship very strong. Solutions to increase education spending in order to increase SAT scores can mean guiding school districts to continue to strive to get more funds from the state, host information sessions for families to realize the importance of education and what materials they can invest in to increase student

[4] "Sustained silent reading is a time during which a class, or in some cases an entire school, reads quietly together. Students are allowed to choose their own reading materials and read independently during class time" (Gardiner, 2005).

success and create affordable SAT and ACT preparation courses. Refer to Map 7 for a visual comparison between these two variables.

SAT Scores vs. Number of Multigenerational Households. Multigenerational households have increasingly become more popular in the last 10 years. After running a regression test between the number of multigenerational households and SAT scores, the R^2 value states that 19.3% of the SAT scores can be explained by the number of multigenerational households within the area. The significant *p*-value, 0.00071, indicates that these two variables have a linear relationship. However, the number of multigenerational households have a negative correlation with SAT scores. Some reasons could be because educational priorities may not align, English is not the primary language used at home, or too many people living in the same vicinity which could make it harder to study at home. Several solutions for school districts with a high number of multigenerational households to increase SAT scores could be to enhance their English as a Second Language programs, create affordable after-school programs to help students focus more on their academics, and administer parent workshops to demonstrate the importance of education and provide tactics for parents on how they can help their children succeed. Refer to Map 8 for a visual comparison between these two variables.

Discussion

Through this initial study, it can be concluded that student success is based heavily on income related factors. Students who live in households that tend to spend more money on activities that exercise the brain such as books, test preparation, games, and music perform better on standardized tests like the SAT and ACT. However, there were variables that were not related to income that had a negative correlation with SAT scores such as the number of multigenerational homes within a community. Family demographic and socioeconomic factors all play a vital part

in a student's academic life. It is important that community leaders come together to create solutions to address the gap between the low and high performing schools.

Although this study found support for my hypotheses, there are some limitations. Limiting the sample to only one school per county makes the process of analyzing simpler; however, there is a bias by choosing the most populated school to represent a single county. Some counties in California oversee over 20 school districts, all of which could have very different identities. Future studies should have a sample that includes every public high school within California, rather than just one high school per county, in order to produce a more accurate prediction model. Also, a more in-depth statistical analysis can be conducted to produce a more detailed relationship between student success and demographic and socioeconomic variables. While linear regression and correlation can describe a lot about the variables, using a stepwise regression analysis may be able to more accurately identify which of the 18 variables influence SAT scores. When it comes to analyzing 18 predictor variables at once, using a stepwise regression can automatically eliminate insignificant variables one by one. Although I was able to establish which of the 18 variables in the study have an impact on SAT scores, the correlation analysis suggested that there may be a multi-collinearity problem that would need further investigation. At the end of this study, I realized that gathering data on the average dollar amount spent for some variables was not the most realistic way to compare these variables. Instead, in the next study, enriching locations with Spending Potential Indexes would be more appropriate because it accurately compares the local spending potential to the national spending potential.

Conclusion

SAT and ACT scores are heavily influenced by the community that a student comes from. Most of the visual data in this study shows that students who scored higher on the SAT and ACT

exams come from schools around the San Francisco Bay Area. Out of all the variables that were tested, the most significant ones turned out to be related to income, with the exception of household size and number of multigenerational households. These are the 12 independent variables that proved to have a notably positive or negative impact on student success.

Household Spending Habits	• Average Spending on Alcoholic Beverages • Average Spending on Online Entertainment/Game • Average Spending on Books • Average Spending on Health Insurance • Average Spending on Entertainment/Recreation - Pets • Average Spending on Cable and Satellite TV Services • Average Spending on Test Prep & Tutoring Services • Average Spending on Education (Annual) • Average Spending on Musical Instruments & Accessories
Socioeconomic Variables	• Average Household Income
Demographic Variables	• Average Household Size • Number of Multigenerational Households

Each of the spending habit variables have a positive correlation to SAT scores. The models show that the more money a household spends, whether it be on alcoholic beverages, test preparation materials, or health insurance, the better the SAT results are. Families with these financial expenditures are living in middle to high socioeconomic communities, so they have the resources available to help students perform better academically. All of the demographic and socioeconomic variables also have a positive correlation to SAT scores, with the exception of the number of multigeneration households. Multigenerational households was the only variable that

proved to affect student academic success negatively. This particular living arrangement is most common in lower socioeconomic areas where many of their resources are limited, such as after-school programs or proper school supplies.

As a response to this problem, schools across California should work together to create and fund programs that are easily accessible and financially feasible to help students perform better on exams and academics in general. In addition, making sure that students are in the proper mental and physical state to perform well in school should also be made a priority.

Study 2

Study 2 will address two limitations from the Study 1. In Study 1, I realized that the sample size of 56 schools may not accurately represent California as a whole. Thus, I will include every public high school in California into the sample for Study 2. It was also confirmed that the predictor variables in Study 1 had an issue with multicollinearity. To address the issue of multicollinearity, stepwise regression will be conducted. By using different methods and procedures from Study 1, this current study will be able to produce a better prediction model for academic success.

Methods

Sample. Unlike Study 1, this dataset focuses on every public and charter high school in California that have SAT records available on the California Department of Education Dataquest. There are a total 1,070 schools in this sample as seen in Map 9. For each school, I was able to gather data from the California Department of Education DataQuest on the number of students in grade 12, number of students who took the SAT during the 2015-2016 academic school year, the average score on each subject within the SAT – reading, math, and writing, and the number of students who scored above a 1500. With the average SAT score for each single subject, the overall average SAT score could be calculated for each high school.

Procedure and Design. The objective is find out what variables affect student academic success by looking into different demographic and socioeconomic factors and comparing them to different locations within California. This study begins by grouping the school locations into hot and cold spots to visualize where the academic performance gaps lie on the map. ESRI's ArcGIS Online enables users to perform a hotspot analysis, where the program will form clusters within the sample. The hot spots group together the areas that performed well together, and the cold spots group together the areas that did not perform as well. The rest of the points are outliers and are not considered in this analysis.

After performing the hotspot analysis, I will enrich the sample through ArcGIS Maps for Office. Different from Study 1, the Study 2 uses spending potential indexes (SPI) rather than average household spending amounts to conduct a more fair and accurate prediction model for not only this sample, but also the rest of the country. Utilizing the SPI instead of the average household spending amount is a better measurement of spending habits because it compares each household spending habits at a national level. ESRI calculates indexes by taking the average amount spent per household within the local area and comparing it to the average amount spent at a national rate (ESRI 5, 2017). The national average SPI is 100 (ESRI 5, 2017). For example, the SPI on test preparation and tutoring services within a 1-mile radius of Paloma Valley High School in Menifee, California is 127, which means the average spending by locals within that area is 27% above the national average.

The following are the variables that I enriched each school in the sample within a 1-mile radius:

Household Spending Habits	• SPI on Social/Recreation/Civic Clubs Member Fee • SPI on Education Annually • SPI on Test Prep & Tutoring Services • SPI on Health Insurance • SPI on Alcoholic Beverages • SPI on Online Entertainment/Games • SPI on Books • SPI on Cable and Satellite TV Services • SPI on Entertainment/Recreation - Pets • SPI on Musical Instruments & Accessories • SPI on Elementary/HS School Books/Supplies • SPI on School Meals
Socioeconomic Variables	• 2017 Average Household Income • Number of Households with Income Below Poverty Level • 2017 Education Obtainment: Bachelor's Degree
Demographic Variables	• 2010 Number of Husband-wife Families with Own Kids < 18 • Population aged 5-17 who speak Spanish/No English • 2017 Diversity Index • 2010 Number of Other Family/Male Householder with Own Kids < 18 (Single-Parent Households) • 2010 Number of Other Family/Female Householder with Own Kids <18 (Single-Parent Households) • 2017 Average Household Size • 2010 Number of Multigenerational Households
Businesses	• Number of Educational Services - Businesses (NAICS) • 2017 Number of Occupations: Education/Library

Results

Instead of performing multilinear regression, this study will utilize stepwise regression analysis. Stepwise regression is appropriate for this particular study because there are 24 possible predictor variables to test. The forward and backward stepwise function will automatically eliminate variables that do not impact SAT scores, formulating a prediction model. Furthermore, I will observe the variance inflation factor (*VIF*) to address multicollinearity within the predictor variables. To conclude this study, I will run a multivariate analysis to determine which variables have a positive or negative effect on SAT scores.

Hotspot Analysis. Instead of identifying the different tapestries that make up the sample as done in Study 1, I performed a hotspot analysis through ArcGIS Online. All 1,070 schools were split into subsets of hot spots, cold spots, and not significant as seen in Map 10. As stated before, the hot spots represent the group of schools that had high SAT results and the cold spots represent the group of schools that had low SAT results. Map 10 shows that majority of the hot spot clusters are centered around the San Francisco area, in addition to San Diego and Orange County in southern California. The cold spots are mostly in the southern and central parts of the state. By observing the visual representation of the hotspot analysis, it is evident that there indeed exists an academic performance gap between the northern and southern part of the state.

Stepwise Regression. After enriching the sample and importing the data into JMP, a forward direction was selected to perform the stepwise regression (Refer to Table 9). 12 out of the 24 variables appeared to be highly significant with *p*-values less than 0.05:

Household Spending Habits	• SPI on Education Annually • SPI on Online Entertainment/Games • SPI on Cable and Satellite TV Services • SPI on Entertainment/Recreation - Pets • SPI on Musical Instruments & Accessories
Socioeconomic Variables	• 2017 Education Obtainment: Bachelor's Degree
Demographic Variables	• 2010 Number of Husband-wife Families with Own Kids < 18 • 2017 Diversity Index • 2010 Number of Other Family/Male Householder with Own Kids < 18 (Single-Parent Households) • 2017 Average Household Size • 2010 Number of Multigenerational Households
Businesses	• Number of Educational Services - Businesses (NAICS)

However, as concluded in Study 1, many of the significant variables had a high correlation with one another, suggesting an issue with multicollinearity. A popular method used to detect and reduce multicollinearity is verifying the *VIF*. When a variable has a *VIF* greater than 5, it can be assumed that it has a high correlation with other variables within the model. Thus, I eliminated those variables with a *VIF* greater than 5 and produced the prediction model for SAT scores as seen in Table 10.

$$y = 1858.58 - 0.06x_1 + .01x_2 - .14x_3 - 5.35x_4$$

where x_1=Number of Multigenerational Households,
x_2=Number of Bachelor's Degrees Obtained,
x_3=Number of Other Family Male Householder with Kids<18, and
x_4=Diversity Index

This particular model can be used to predict which schools will most likely be able to produce high or low SAT scores. The model states that the number of multigenerational households and single father households within a community negatively impact SAT results. However, the diversity index around each school negatively impacts SAT scores to a greater effect than the household demographics. Lastly, the prediction model demonstrates that the number of bachelor's degrees obtained within a given area will have a positive impact on SAT scores. Maps 11, 12, 13, and 14 illustrate the relationship between each of the predictor variables and SAT scores through two-layered maps.

To verify the results, a backward stepwise regression was also performed, and the same exact results were produced as seen in Table 11.

Correlation. All predictor variables were subjected to a multivariate analysis where JMP was able to distinguish which variables had a positive or negative correlation to SAT scores. Table 12 describes each of the correlation values. The variables that positively affect students' academic success are the following:

Household Spending Habits	• SPI on Social/Recreation/Civic Clubs Member Fee • SPI on Education Annually • SPI on Test Prep & Tutoring Services • SPI on Health Insurance • SPI on Alcoholic Beverages • SPI on Online Entertainment/Games • SPI on Books • SPI on Cable and Satellite TV Services • SPI on Entertainment/Recreation - Pets • SPI on Musical Instruments & Accessories • SPI on Elementary/HS School Books/Supplies • SPI on School Meals
Socioeconomic Variables	• 2017 Average Household Income

These positively correlated variables are all related to income, just as shown in Study 1. It does not necessarily mean that spending more money on alcoholic beverages or online games will improve students' SAT results. The general idea is that the more money a family has, the higher their financial ability is to spend money on activities like health insurance, test preparation materials, and pets. However, there are other studies that show how some of these variables, such as spending money on musical instruments and additional reading books, can positively affect how well students perform in mathematics and reading (Santos-Luiz, 2007).

The variables that negatively affect students' academic success are the following:

Socioeconomic Variables	• Number of Households with Income Below Poverty Level
Demographic Variables	• 2017 Diversity Index • 2010 Multigenerational Households • 2010 Number of Other Family/Male Householder with Own Kids < 18 • 2010 Number of Other Family/Female Householder with Own Kids <18

The negative correlated variables all fall into to the demographic category. Diversity, in many ways, has positive effects on students' educational experience. Research shows how diversity within a classroom and even neighborhood can be useful in regard to education (Dronkers, 2013). However, there are a few aspects of it that can hinder students' academic results as well.

Discussion

For this particular study, I wanted to examine the various demographic and socioeconomic variables using a more accurate statistical analysis. Because I saw an issue with multicollinearity between the predictor variables in Study 1, I made sure to address this problem by performing a

stepwise regression and verifying each of the variable's *VIF* values. The stepwise regression analysis automatically removed variables that did not hold a significant *p*-value and eliminating the variables with *VIF*s greater than 5 produced the final prediction model of academic success. I found that diversified communities and multigenerational and single father households have a negative effect on SAT scores. I also discovered that the number of bachelor's degree within a 1-mile radius of each school had a positive effect on SAT scores.

Although most of the results that were produced in this study are quite intuitive, there were some variables that took me by surprise. While the number of single father households appeared in the prediction model, single mother households did not. I find this rather surprising because I would expect any type of single parent family dynamic to have a negative impact on how students perform in school. Another variable that produced an unexpected result was the diversity index. In today's society, we are constantly pushing and encouraging schools to become more diverse. Diversity is a beautiful concept and has its advantages. However, through this study, I learned that diversity greatly affects student academic success in a negative way.

Considering the results of this study, community leaders and school educators can be more proactive about helping students earlier in their academic careers. We should center our attention on creating solutions to minimize the academic achievement gap between low and high performing students. These solutions can be, but not limited to, creating new and improved academic intervention programs in K-12 education that are specifically catered to the lower socioeconomic communities to focus on increasing students' scores on standardized tests, such as the SAT. Lastly, public high schools and college campuses should build a stronger relationship with one another in order to increase students' curiosity and encourage them to work towards becoming a prepared college student.

Although this study addressed many of the concerns from Study 1, there are still some limitations. Because the diversity index appeared to be an influential variable in the prediction model, enriching the sample with other demographic variables can potentially reveal more detailed results about each of the communities surrounding each school. Investigating the variety of races nearby each school in the sample could create a completely different ending to this project. In the future, this study can be improved by looking at a different set of variables that cannot be acquired through ESRI. While ESRI provides data on demographic and lifestyle variables, the California Department of Education also provides data on truancy[5] rates and teachers' education obtainment. Investigating variables, such as truancy rates, could uncover more about the local demographics' educational priorities and help us understand the prediction models better.

Conclusion

As the academic achievement gap between low and high performing students continues to grow, there needs to be a discussion on how community leaders and school educators can address this issue. This paper investigates what demographic and socioeconomic variables that affect student academic success. Because previous studies have proven SAT scores to be a valid indicator of college readiness, SAT scores were used to measure students' academic achievement. I hypothesized that some demographic and socioeconomic variables, such as income and family structures, affect students' performance on the SAT. A sample of public high school from California was enriched with various data relating to average household spending habits, demographics, and socioeconomic statuses through ArcGIS Online. After importing the data into JMP Pro, a prediction model was produced. Through this study, it can be concluded that family

[5] Truancy is when a student is intentionally absent for more than 30 minutes of school with an unauthorized reason (California Department of Education, 2017).

structures such as single parent and multigenerational households have a negative correlation to SAT scores. The diversity index within a community is also negatively correlated to SAT results. However, there are variables, such as average household income and number of bachelor's degrees within a given area, that do positively affect students' SAT scores. Now that we know what factors influence how students perform on exams like the SAT, parents, teachers, and students can work together to overcome these demographic and socioeconomic barriers. It is also important to keep in mind that predictors of academic success are not just limited to demographics and socioeconomic statuses. More research can be done to determine what other type of variables affect how students perform, such as truancy rates. However, investigating which demographic and socioeconomic factors influence students' academic performance is a good start to addressing the academic achievement gap between low and high performing students.

References

ACT (2018). *About the ACT test*. Retrieved from http://www.act.org/content/act/en/products-and-services/the-act-educator/the-act-test.html#order-reg-materials

Augustine, J. M., & Raley, R. K. (2013). Multigenerational: households and the school readiness of children born to unmarried mothers. *Journal of Family Issues*, *34(4),* 431–459. http://doi.org/10.1177/0192513X12439177

Barajas, M.S. (2011). Academic achievement of children in single parent homes: A critical review. *The Hilltop Review, 5(4).* Retrieved from https://scholarworks.wmich.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1044&context=hilltopreview

Bethall, T. (2011). *Family matters: Multigenerational families in a volatile economy*. Retrieved from Generations United.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshmen and grade point average from the revised and recent recentered SAT I: Reasoning Test.* College Board Research Report No. 2000-1. New York: College Entrance Examination Board.

California Department of Education - Analysis, Measurement, and Accountability Reporting Division DataQuest. (2017). *2015-2016 SAT*. Retrieved from https://www.cde.ca.gov/ds/sp/ai/

California Department of Education. (2017). *Truancy*. Retrieved from https://www.cde.ca.gov/ls/ai/tr/

Camara, W.J & Echternacht, G (2000). *The SAT [R] I and high school grades: Utility in predicting success in college. Research notes.* Retrieved from College Entrance Examination Board.

College Board 1 (2018). *Reading content alignment*. Retrieved from https://collegereadiness.collegeboard.org/about/alignment/reading

College Board 2 (2018). *Writing and language content alignment.* Retrieved from https://collegereadiness.collegeboard.org/about/alignment/writing-language

College Board 3 (2018). *Math content alignment*. Retrieved from https://collegereadiness.collegeboard.org/about/alignment/mathThe Princeton Review (2018). *What is the SAT?* Retrieved from https://www.princetonreview.com/college/sat-information

Dronkers, J. & Velden R. (2013). Chapter 4: Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using PISA data. In M. Windzio (Eds), *Integration and inequality in educational institutions.* (71-98). New York, NY: Springer

Duff, D., Tomblin, J.B., & Catts, H. (2015). The influence of reading on vocabulary growth: A case for a Matthew Effect. *US National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4610292/

ESRI's ArcGIS Maps for Office. (2017). ArcGIS Maps [computer software]. United States of America.

ESRI's ArcGIS Online. (2018). ArcGIS Online [software] Available from https://www.arcgis.com/home/index.html

ESRI 1. (2018). *US tapestry: Up and coming families*. Retrieved from http://downloads.ESRI.com/ESRI_content_doc/dbl/us/tapestry/7A_UpAndComing_TapestryFlier_G79488_2-18.pdf

ESRI 2. (2018). *US tapestry: Set to impress.* Retrieved from http://downloads.ESRI.com/ESRI_content_doc/dbl/us/tapestry/11D_SetToImpress_TapestryFlier_G79488_2-18.pdf

ESRI 3. (2018). *US tapestry: Old and newcomers.* Retrieved from http://downloads.ESRI.com/ESRI_content_doc/dbl/us/tapestry/8F_OldAndNew_TapestryFlier_G79488_2-18.pdf

ESRI 4. (2017). *Tapestry segmentation: The fabrics of America's neighborhoods*. Retrieved from http://www.ESRI.com/library/brochures/tapestry-segmentation.pdf

ESRI 5, (2017). *ESRI consumer spending methodology 2017*. Retrieved from http://downloads.ESRI.com/ESRI_content_doc/dbl/us/J9945_2017_US_Consumer_Spending_Data.pdf

ESRI 6, (2018). Updated demographics. Retrieved from https://doc.arcgis.com/en/ESRI-demographics/data/updated-demographics.htm

ESRI 7, (2018). Census and ACS. Retrieved from https://doc.arcgis.com/en/ESRI-demographics/data/census-acs.htm

ESRI 8, (2017). Esri demographics. Retrieved from https://www.esri.com/data/esri_data

Gardiner, S. (2005). Building student literacy through sustained silent reading. Heatherton, Vic.: Hawker Brownlow Education.

Hodgson, N. (2016). JMP Pro 13. [computer software]. Cary, NC.

Jorgenson, O. (2012). What we lose in winning the test score race [Abstract]. *Principal*. 91(5). 12-15.

Ladd, H. F. (2012). Education and poverty: Confronting the evidence. *Journal of Policy Analysis and Management*, *00.* Retrieved from https://oied.ncsu.edu/selc/wp-content/uploads/2013/03/Ladd-Education-and-Poverty1.pdf

Paulos, J.A. (2013) *A mathematician reads the newspaper*. New York, NY: Basic Books.

Reardon, S.F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. New York: Russell Sage Foundation Press.

Salkind, N.J. (2007). In Encyclopedia of Measurements and Statistics online. Retrieved from http://dx.doi.org/10.4135/9781412952644.n470

Santos-Luiz, (2007). The learning of music as a means to improve mathematical skills. International Symposium on Performance Science. Retrieved from https://www.researchgate.net/publication/266318346_The_learning_of_music_as_a_means_to_improve_mathematical_skills

Vaughn, K., & Winner, E. (2000). SAT scores of students who study the arts: What we can and cannot conclude about the association. Journal of Aesthetic Education. Retrieved from http://www.jstor.org/stable/pdf/3333638.pdf?refreqid=excelsior:0c143f7977e9a3af692452680a47eac9

Appendix I

Enrichment data is provided by ESRI. ESRI has a data development team that consists of economists, statisticians, demographers, geographers, and analysts who all create the datasets which include annual demographic updates, Tapestry Segmentation, Consumer Spending, Market Potential, and Retail MarketPlace (ESRI 5, 2017).

Household spending habits represents the consumer spending within a given area. It is a mixture of surveys from the Bureau of Labor Statistics that focus on consumer expenditure (ESRI 5, 2017). The data is specific to different types of products of services. Study 1 uses the metric of average spending per household and Study 2 uses the Spending Potential Index (SPI). These are the specific average household spending habits observed within Study 1 and Study 2.

- Alcoholic Beverages
- Online Entertainment/Game
- Books
- Health Insurance
- Entertainment/Recreation - Pets
- Cable and Satellite TV Services
- Test Prep & Tutoring Services
- Education (Annual)
- Musical Instruments & Accessories
- Social/Recreation/Civic Clubs Member Fee
- Elementary/HS School Books/Supplies
- School Meals

Socioeconomic variables are related to income, education, and occupation. ESRI gathers socioeconomic data from U.S. Census and American Community Survey (ESRI 7, 2018). These are the socioeconomic variables this paper observed.

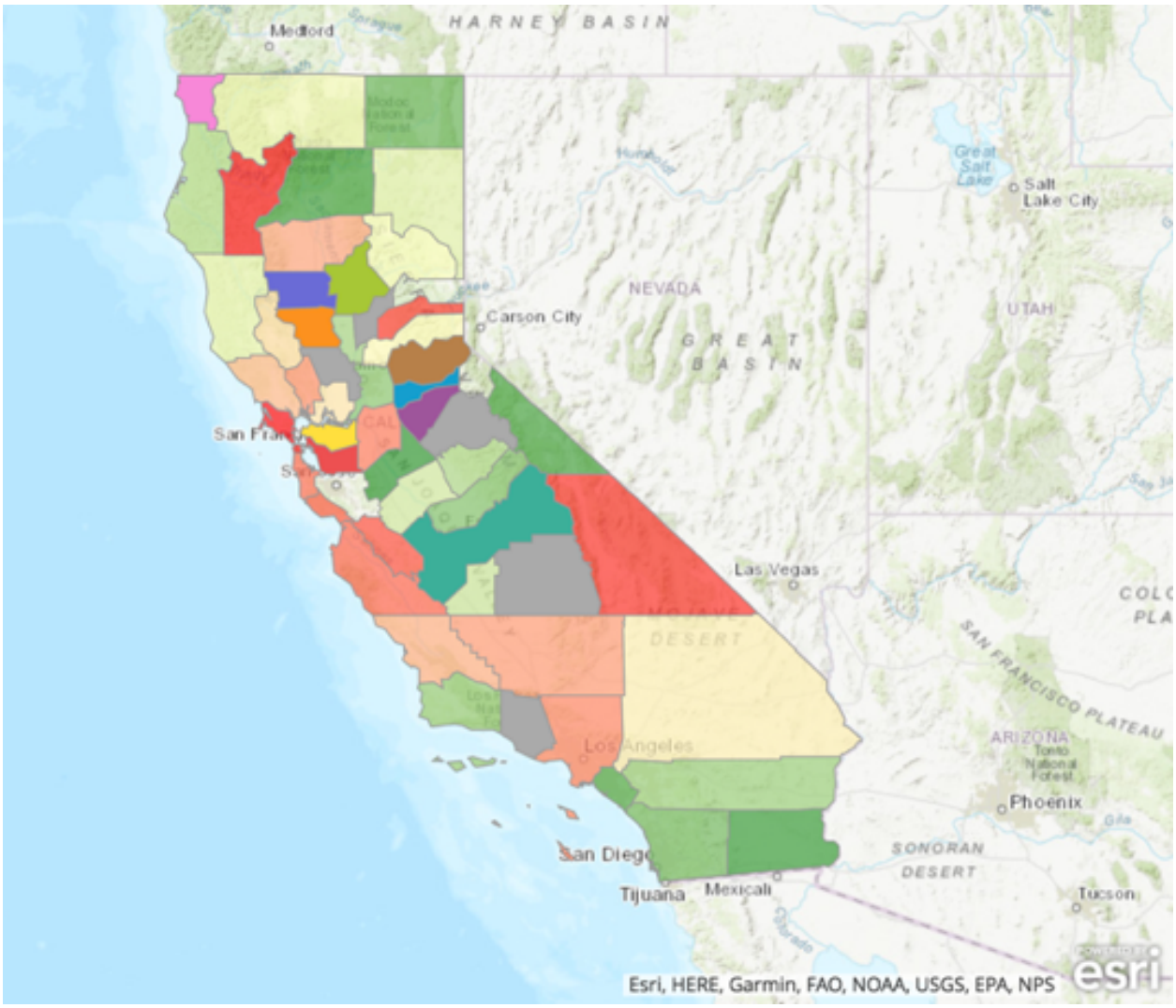
- Average Household Income
- Number of Households with Income Below Poverty Level
- Education Attainment: Bachelor's Degree
- Employed Civilian Population of Age 16+

Demographic variables are related to household size and family dynamic. ESRI constructs its own demographic data annually. The bases of their datasets are combined administrative records and other private sources (ESRI 6, 2018). These are the demographic variables observed in the paper.

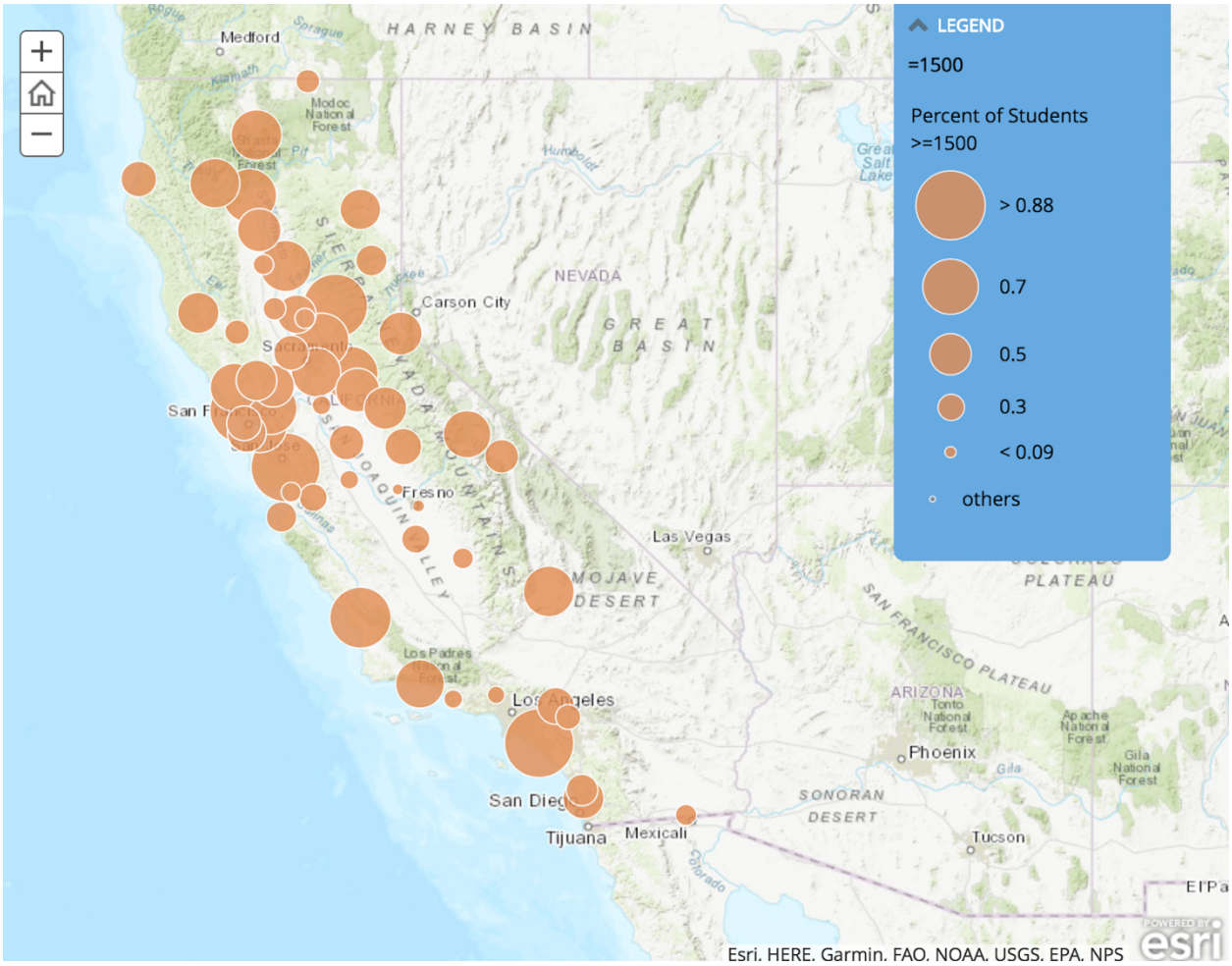
- Average Household Size
- Number of Owner Occupied Housing Units
- Number of Multigenerational Households
- Number of Husband-wife Family Households
- Population of Divorced Couples
- Number of Husband-wife Families with Own Kids < 18
- Population aged 5-17 who speak Spanish/No English
- Diversity Index
- Number of Other Family/Male Householder with Own Kids < 18 (Single-Parent Households)
- Number of Other Family/Female Householder with Own Kids <18 (Single-Parent Households)

Dominant Tapestry Segment describes American neighborhoods in to one of the 67 segments defined by ESRI, all of which are based off of demographic and socioeconomic data (ESRI 4, 2017).

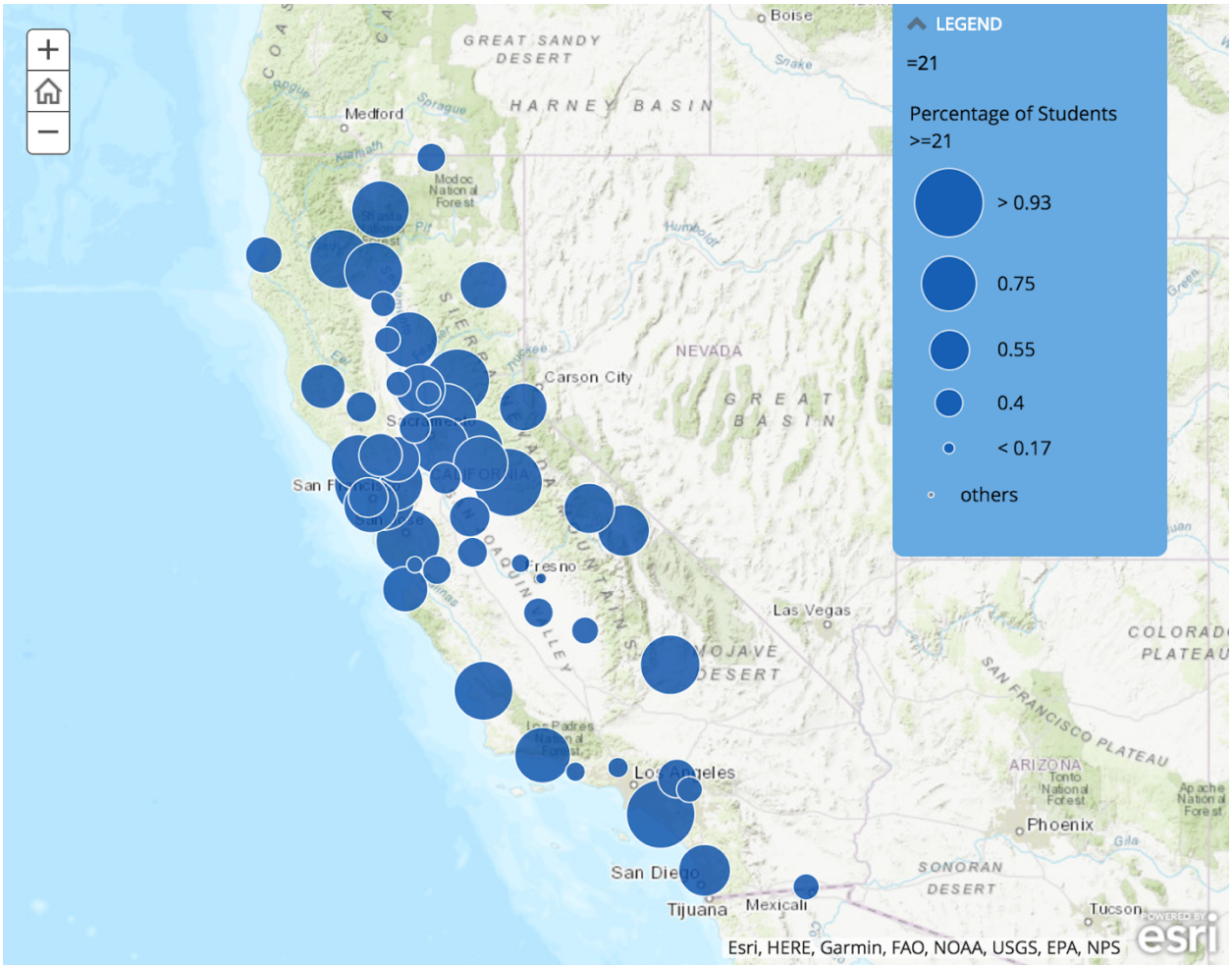
Appendix: Maps



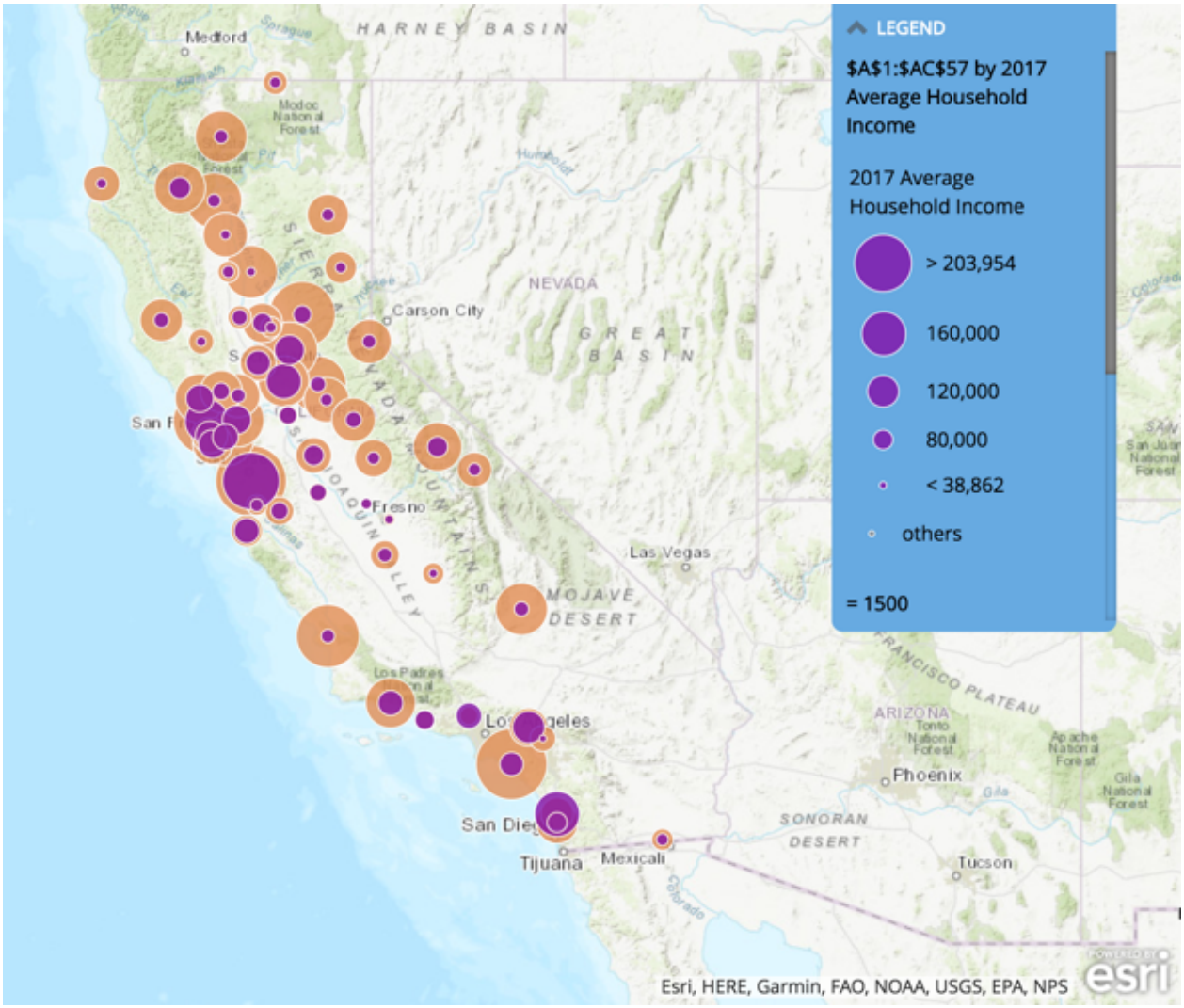
Map 1. County divisions in California



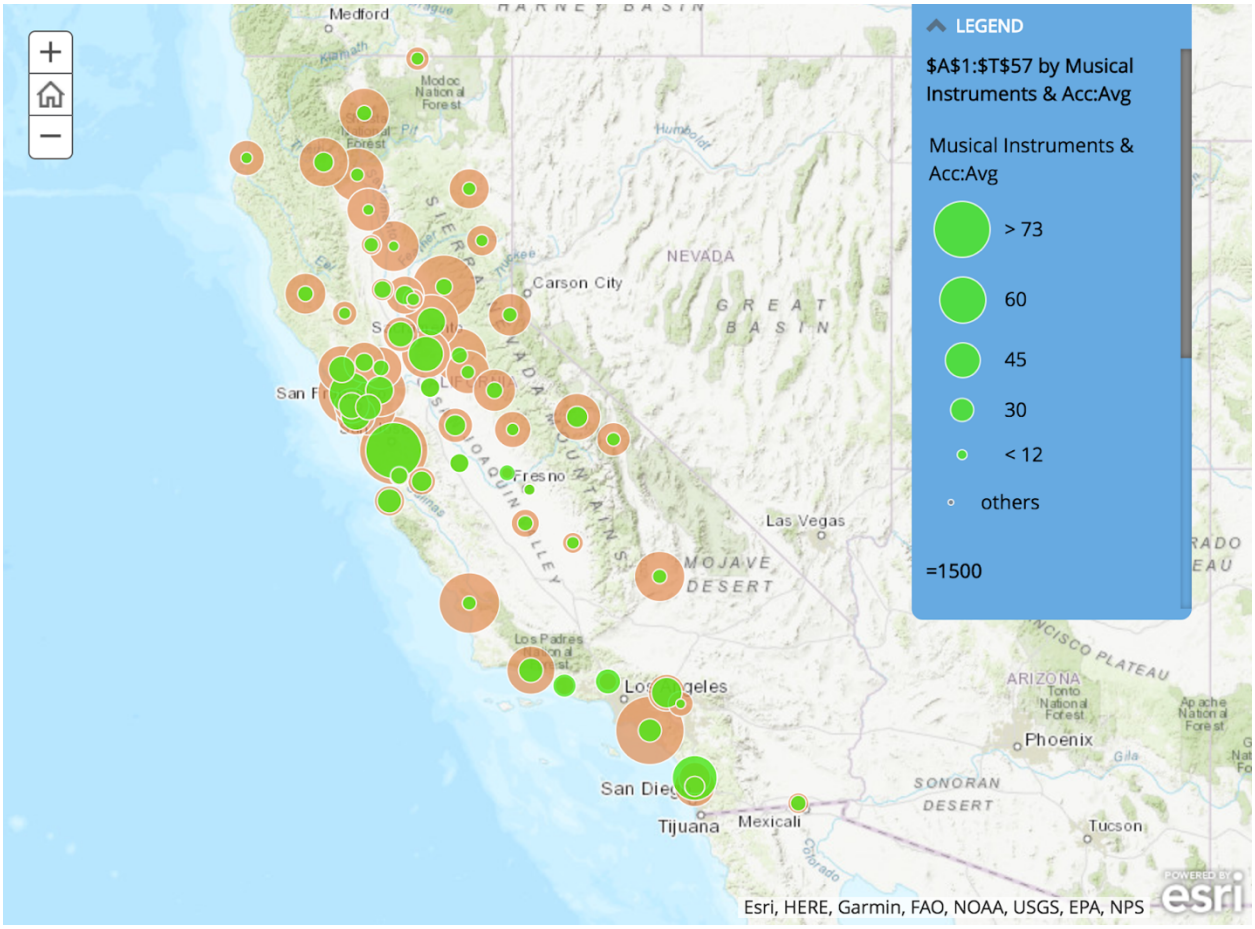
Map 2. Percent of students who scored above 1500 on SAT during 2015-2016 in California



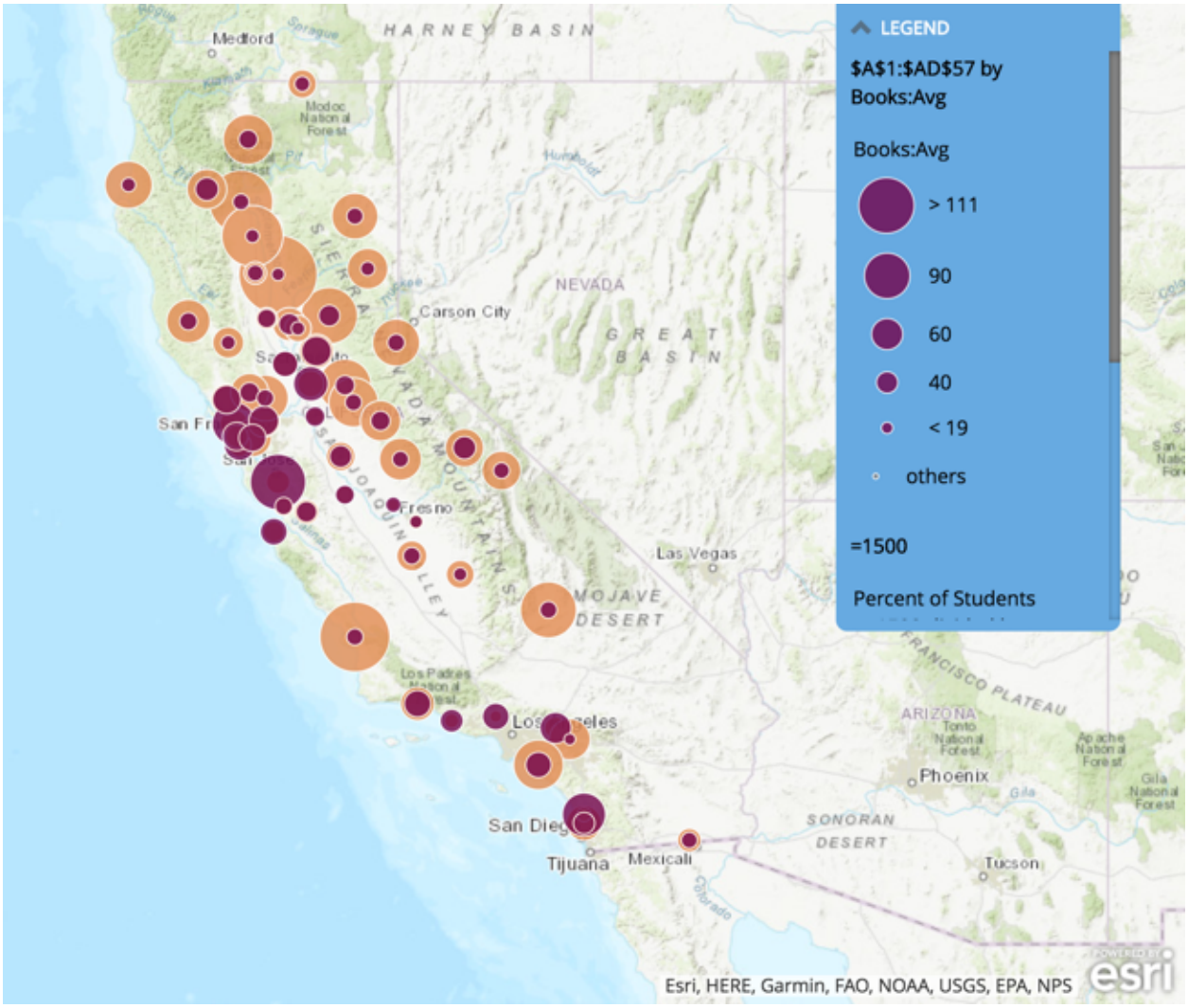
Map 3. Percent of students who scored above 21 on the ACT during 2015-2016 in California



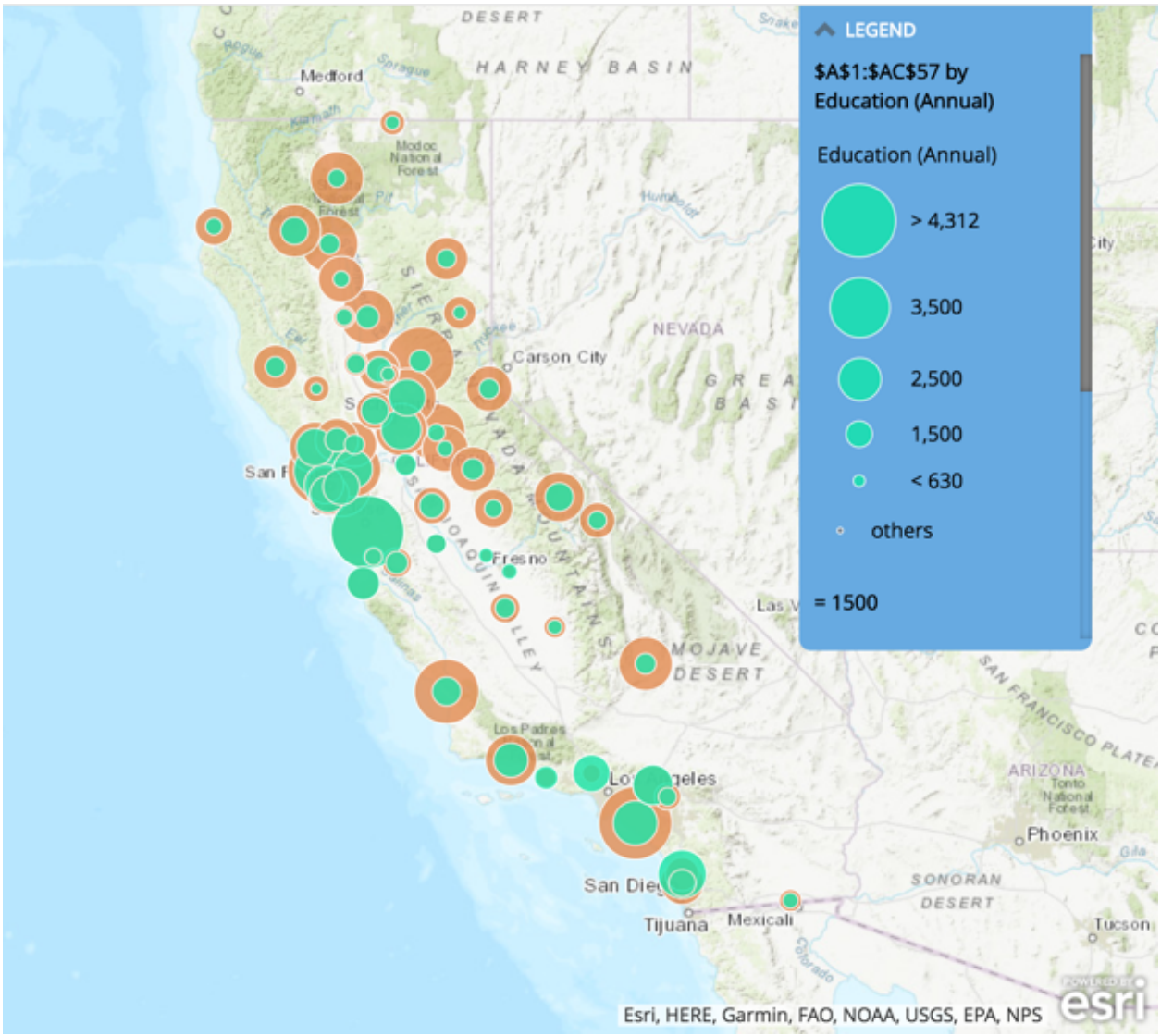
Map 4. 2017 Average household income vs. Percent of students who scored above 1500 on SAT during 2015-2016



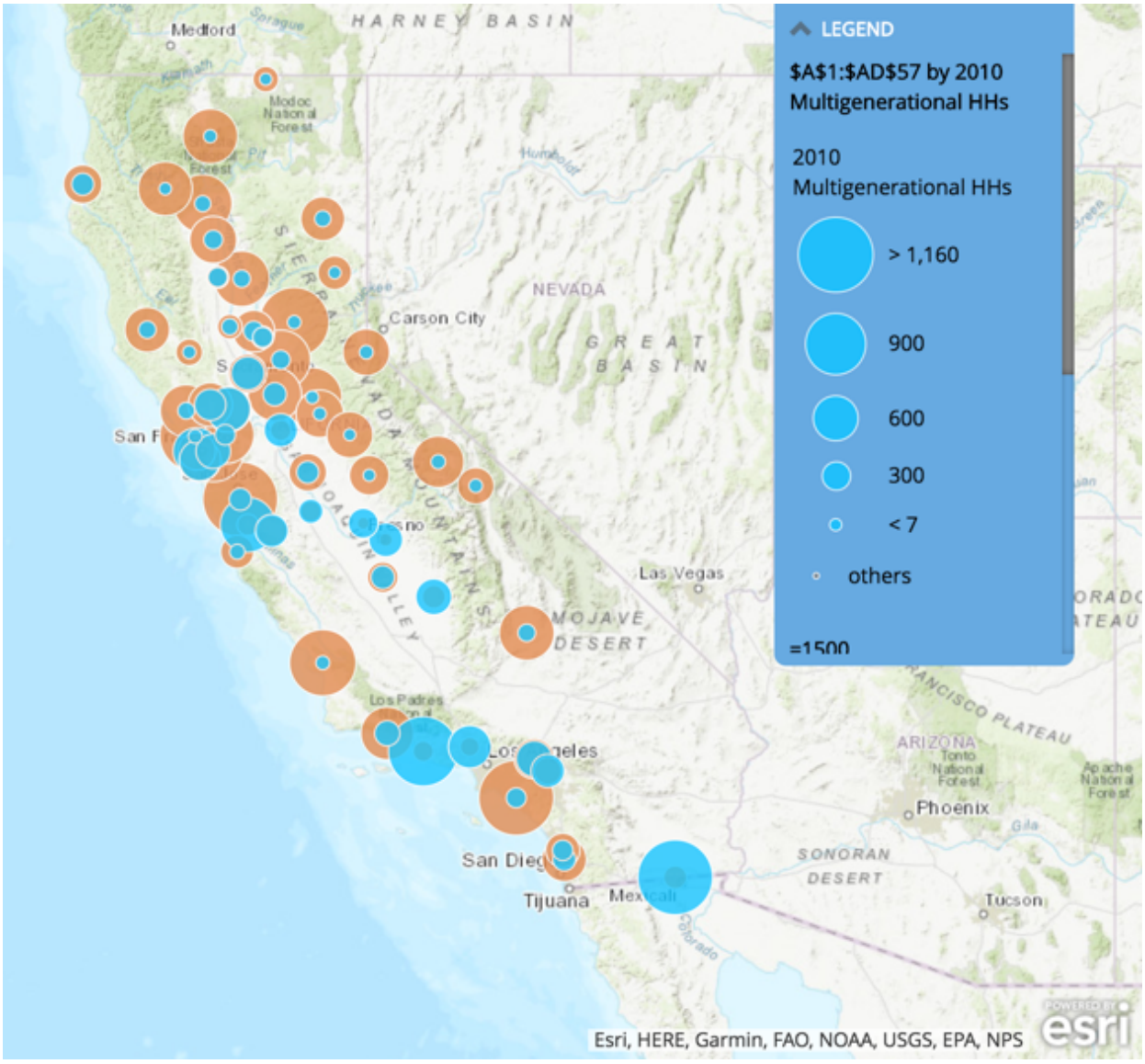
Map 5. Average spending on musical instruments and accessors vs. Percent of students who scored above 1500 on SAT during 2015-2016



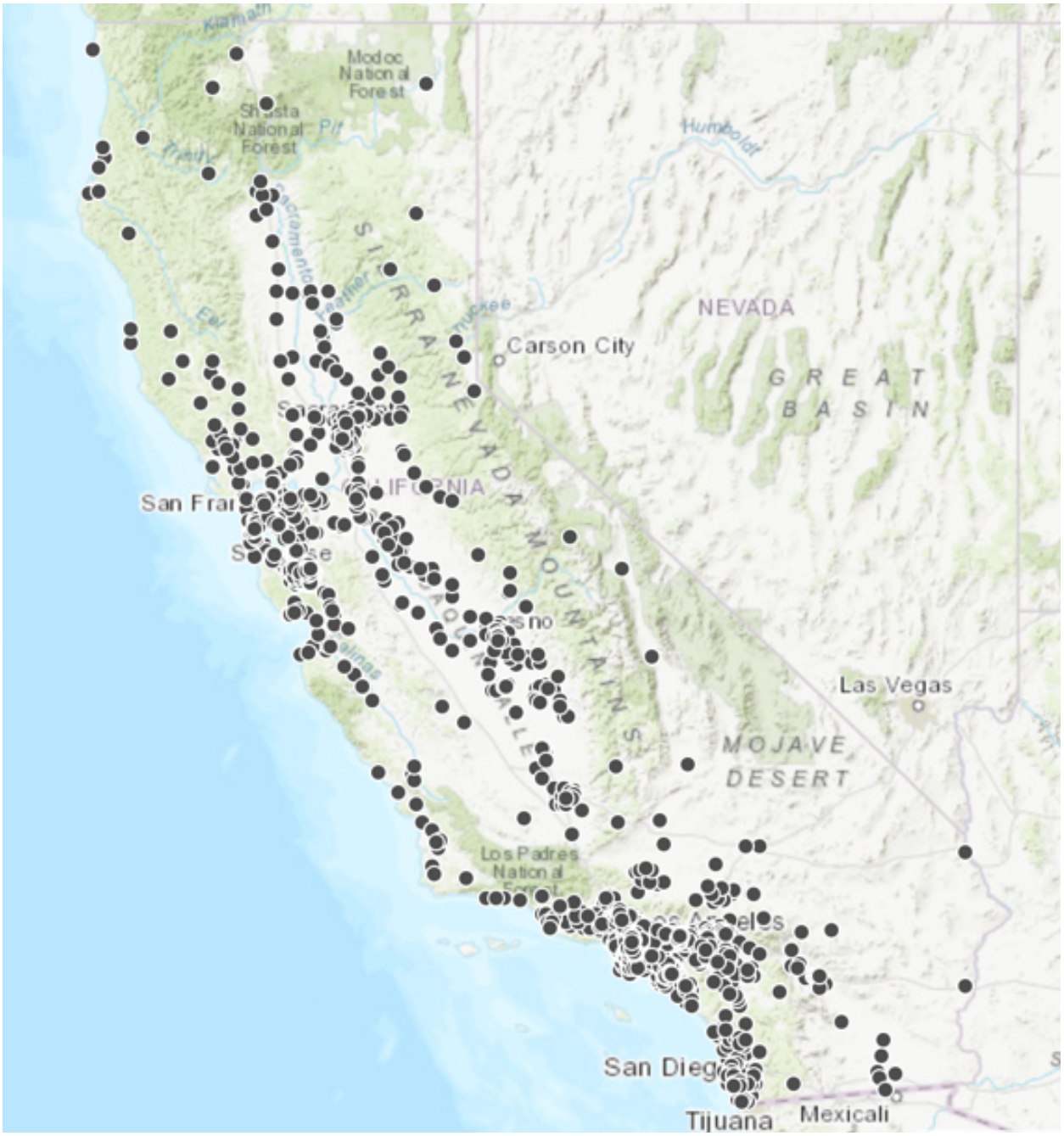
Map 6. 2017 Average spending on books vs. Percent of students who scored above 1500 on SAT during 2015-2016



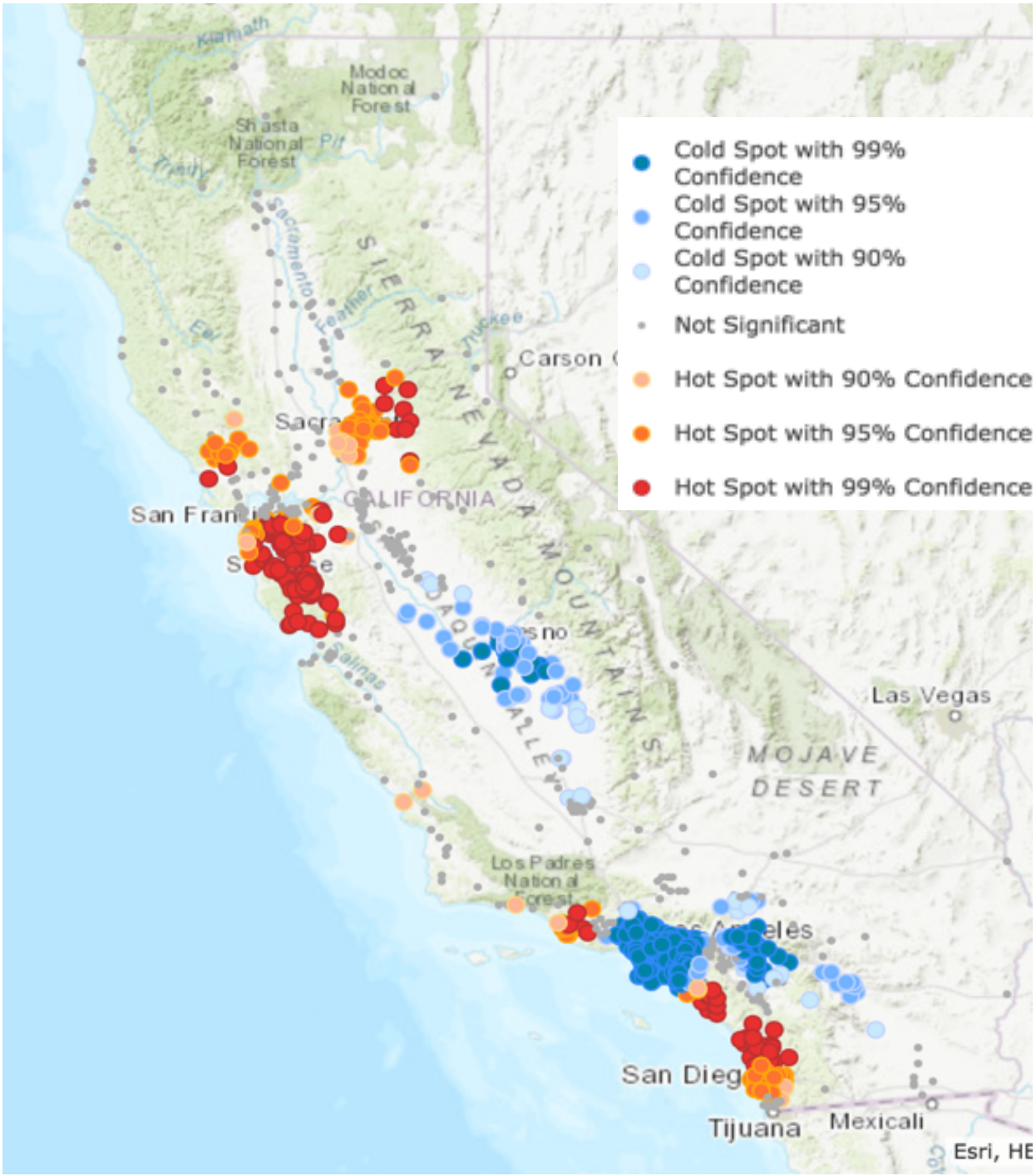
Map 7. Average spending on education (annual) vs. Percent of students who scored above 1500 on SAT during 2015-2016



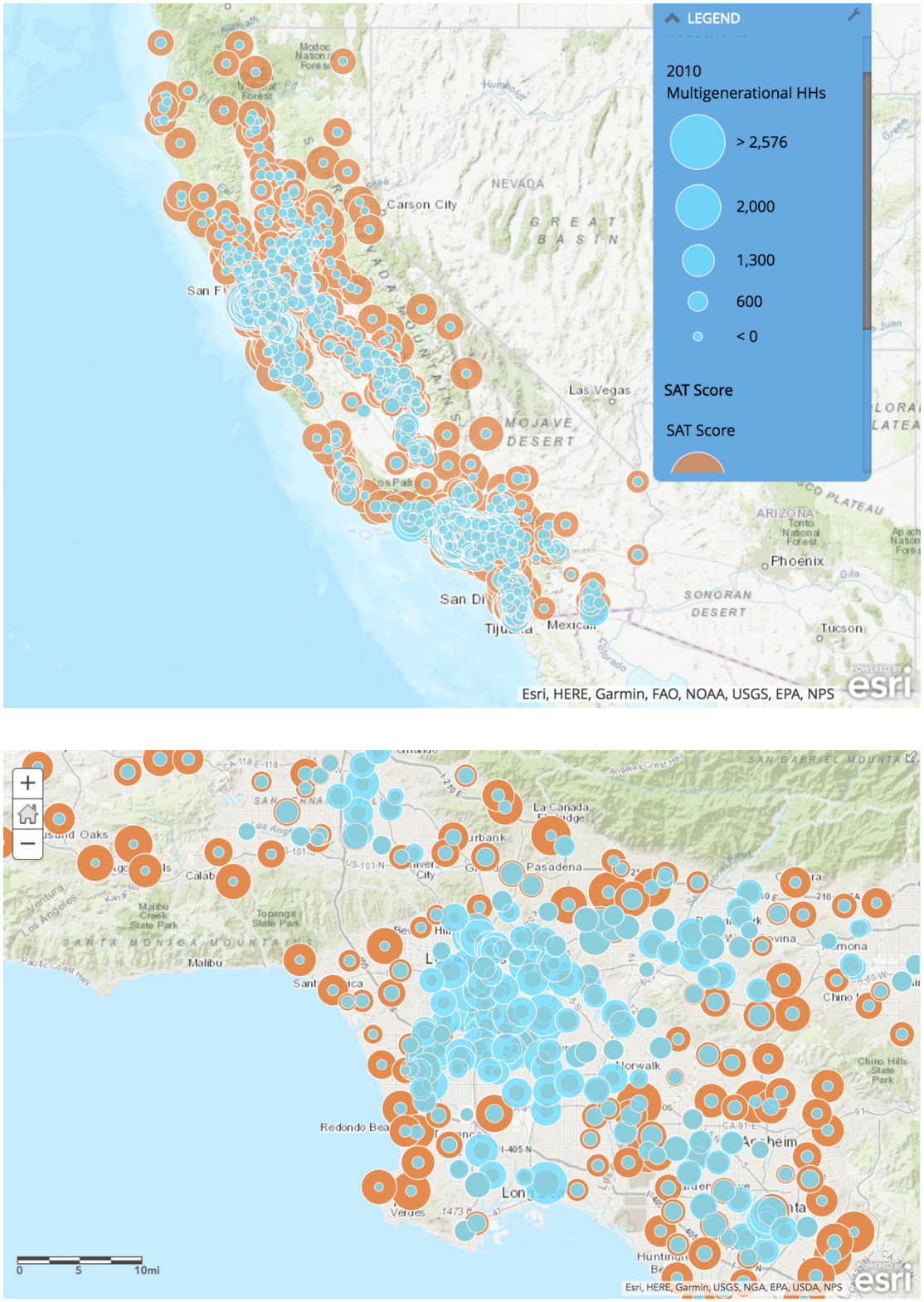
Map 8. Number of multigenerational households vs. Percent of students who scored above 1500 on SAT during 2015-2016



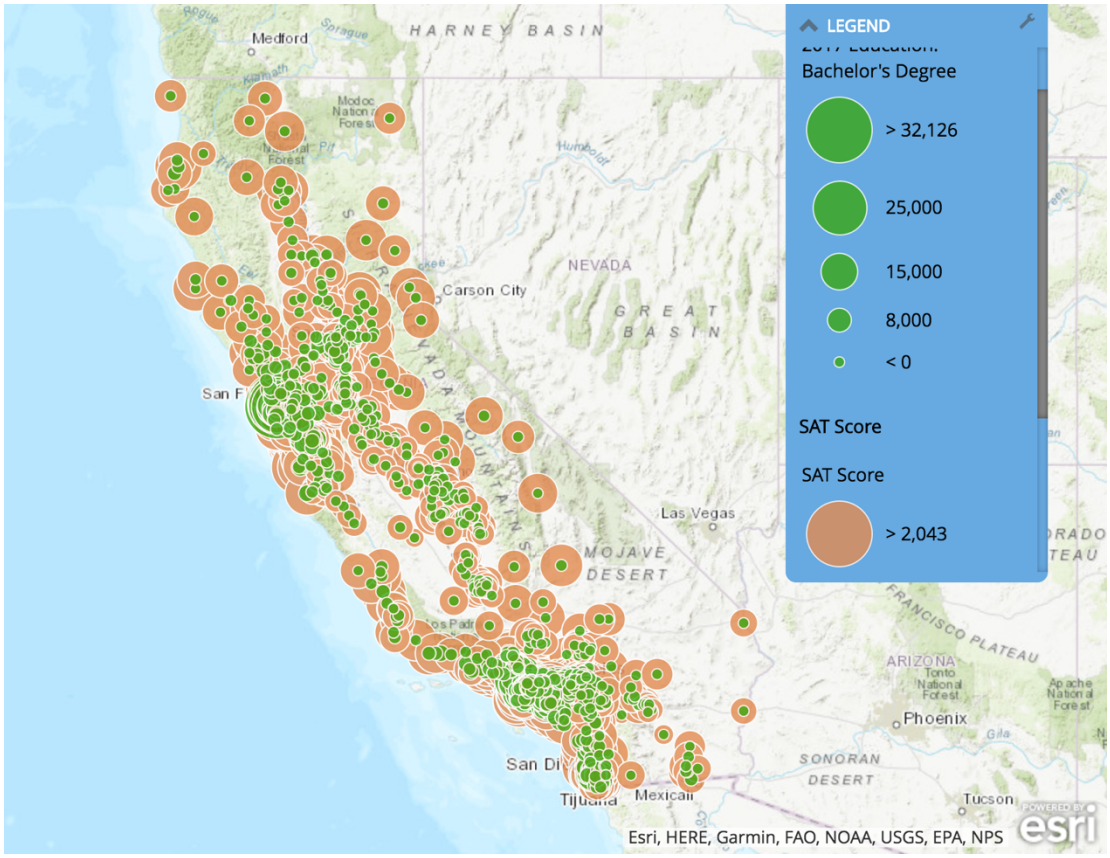
Map 9. Public and charter schools in California

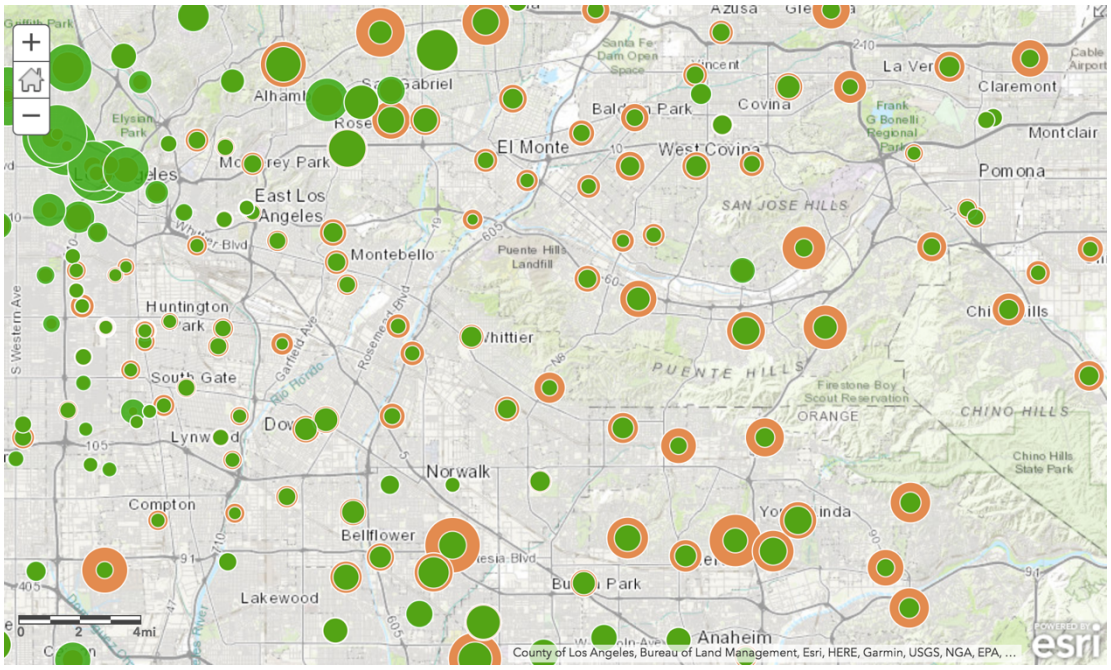


Map 10. Hot and cold spot clusters

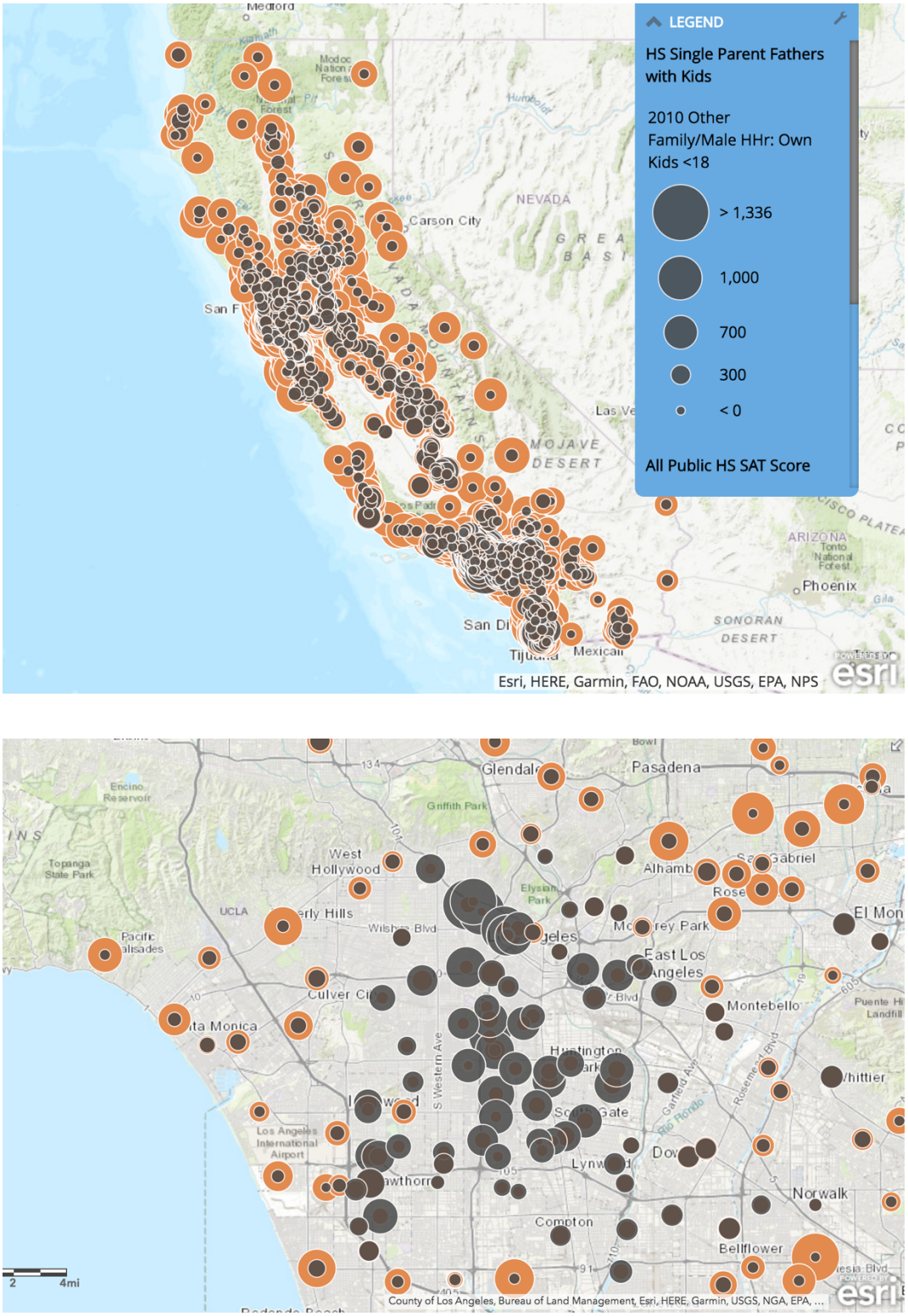


Map 11. Two layered map illustrating the negative relationship between SAT scores and number of multigenerational households

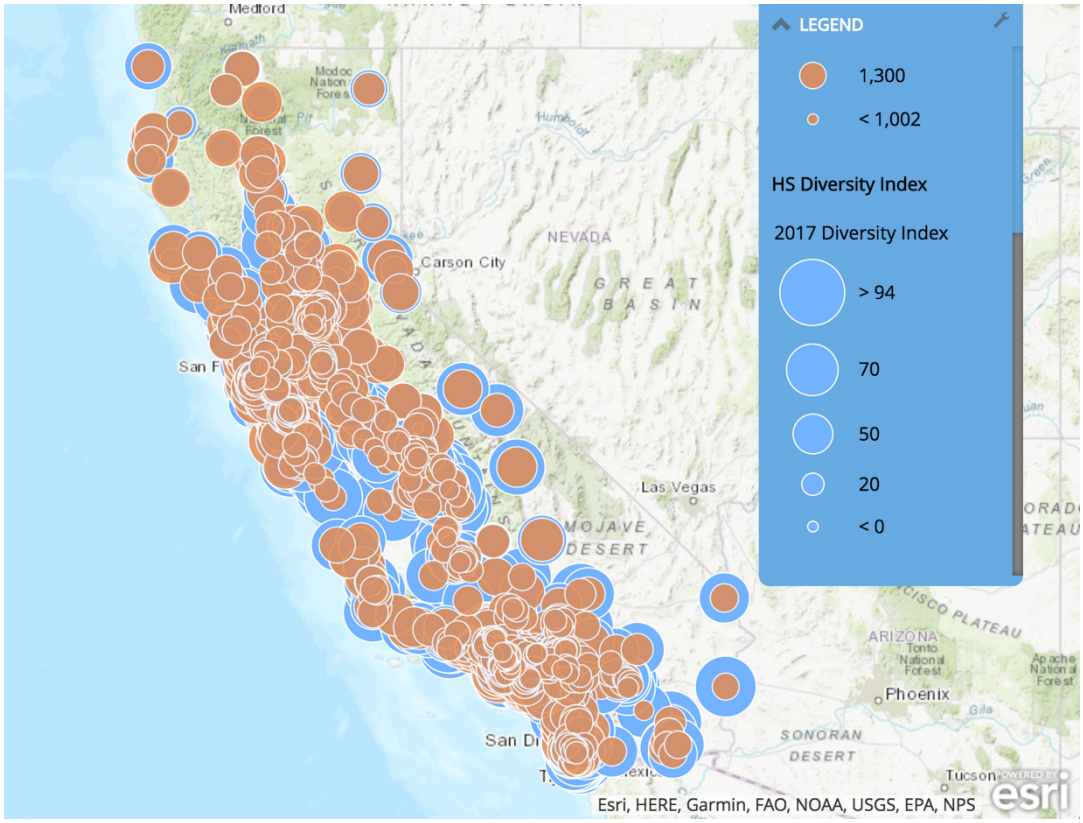


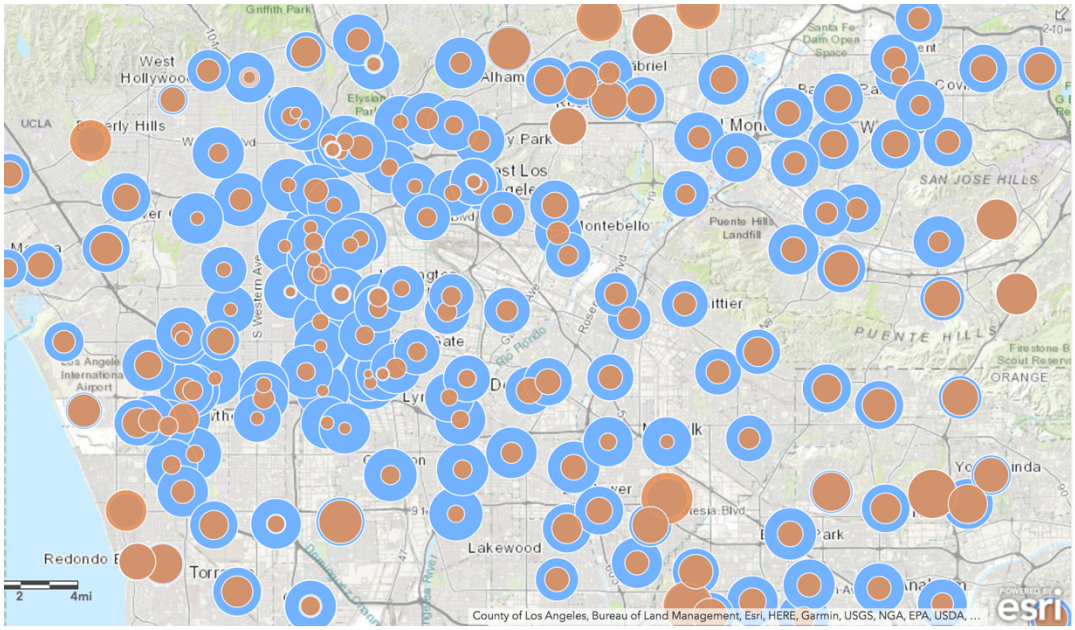


Map 12. Two layered map illustrating the positive relationship between SAT scores and number of bachelor's degrees



Map 13. Two layered map illustrating the negative relationship between SAT scores and number of single father households





Map 14. Two layered map illustrating the negative relationship between SAT scores and diversity index

Appendix: Figures

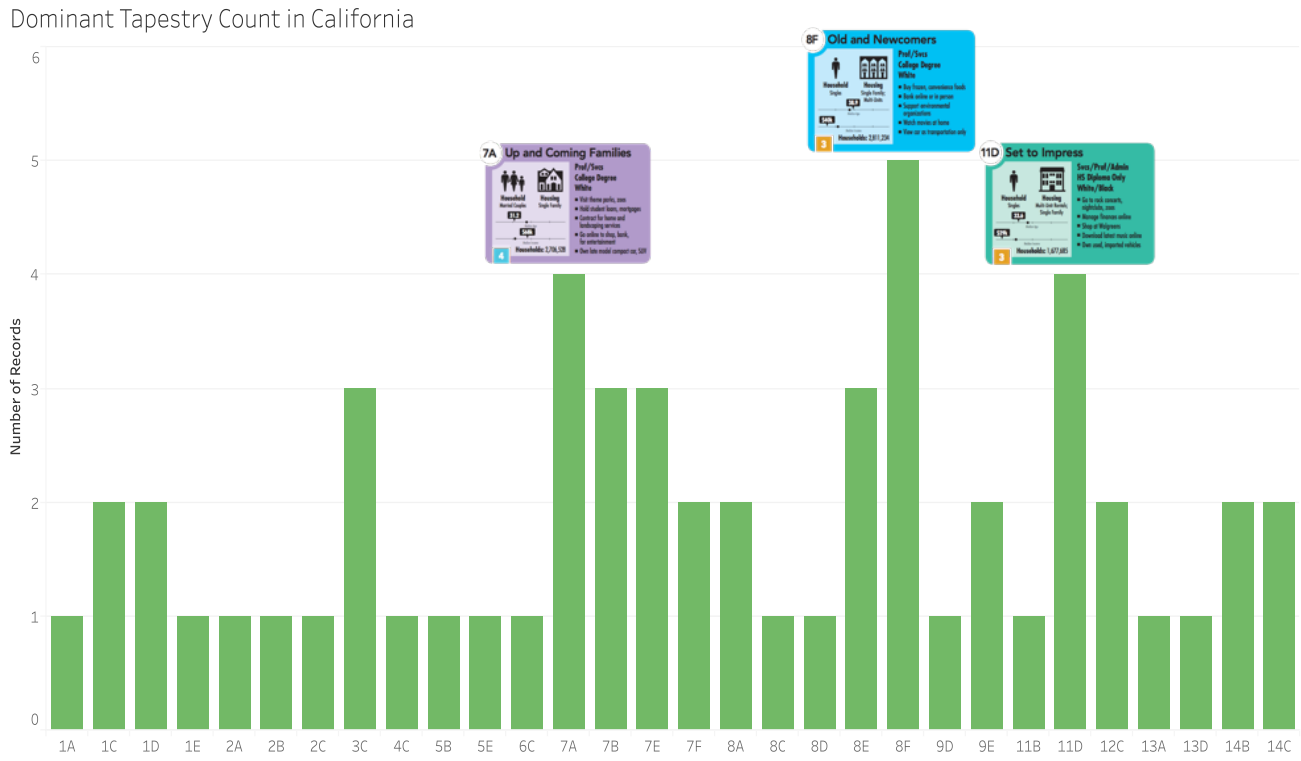


Figure 1. Dominant Tapestry Count in California

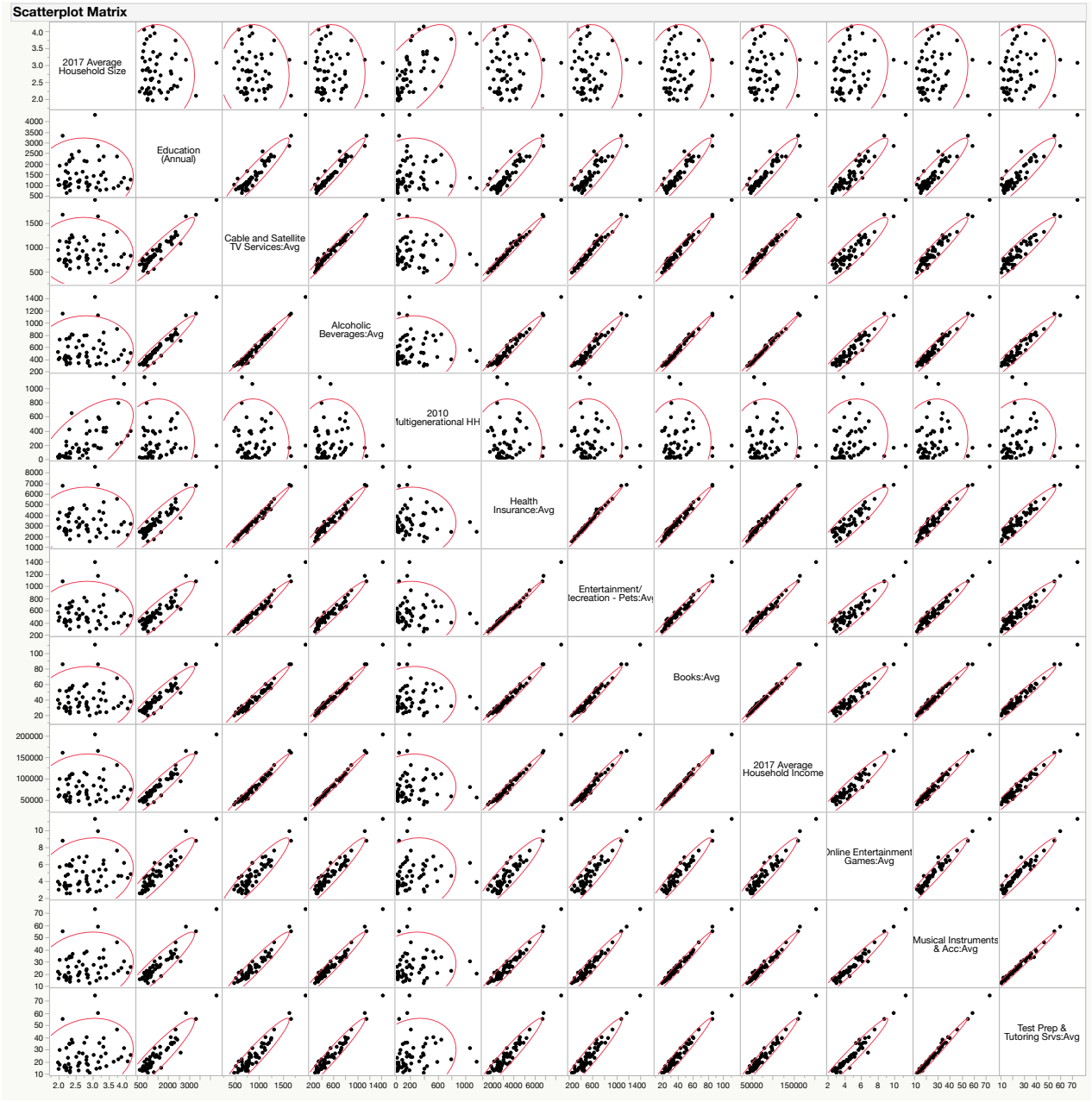


Figure 2. Scatterplot matrix of the correlations between top 12 significant variables

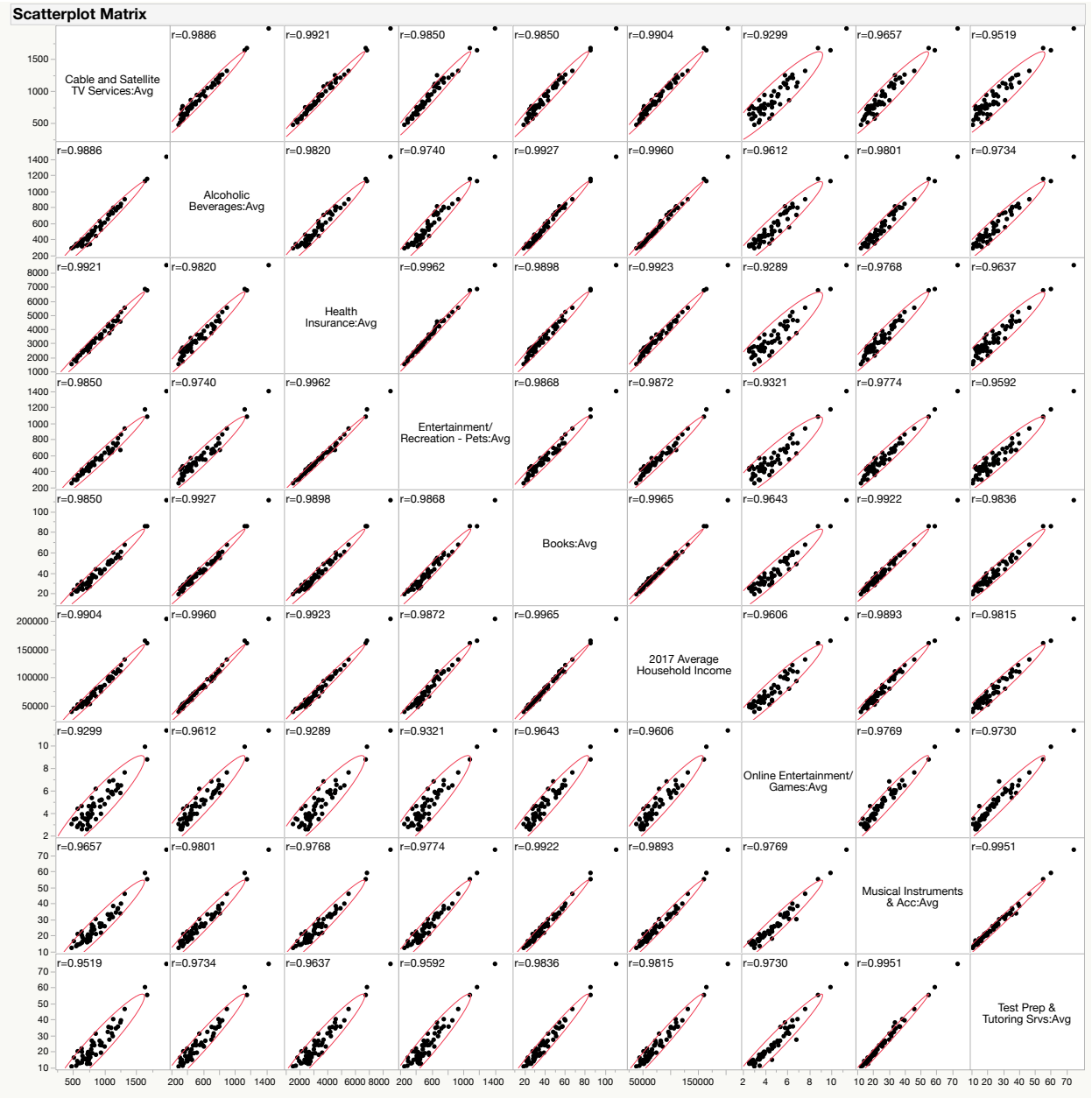


Figure 3. Scatterplot matrix of the correlations between variables that have high correlations to one another

Appendix: Tables

TABLE 1 SAT Scores vs. All 18 Variables		
Variable	***LogWorth***	***p***
2017 Population of Divorcees	2.55	0.00*
2017 Number of Owner Occupied Housing Units	1.39	0.04*
Average Spending on Online Entertainment/Games	1.31	0.04*
Average Spending Education Annually	1.30	0.05*
Average Spending on Musical Instruments	0.95	0.11
Employed Civilian Population of Ages 16+	0.92	0.12
Number of Husband-wife Family Households	0.85	0.14
Average Spending on Pets	0.84	0.14
Average Spending on Books	0.45	0.35
Average Spending on Cable and Satellite TV Services	0.43	0.37
Number of Multigenerational Households	0.32	0.48
Number of Households with Income Below Poverty Level	0.32	0.48
Education Attainment: Bachelor's Degree	0.24	0.58
Average Household Size	0.20	0.63
Average Spending on Alcoholic Beverages	0.18	0.66
Average Spending on Test Prep & Tutoring	0.14	0.73
2017 Average Household Income	0.06	0.88
Average Spending on Health Insurance	0.01	0.98
* $p \leq 0.05$		

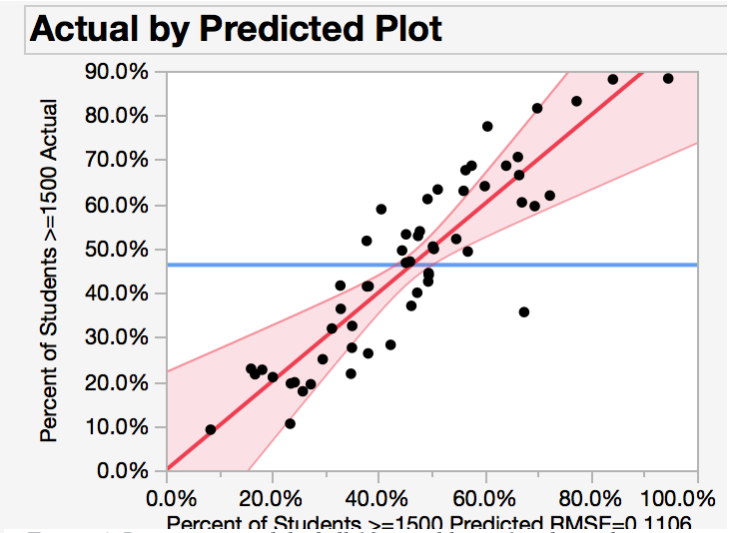


Figure 4. Regression model of all 18 variables at 1-miles radius versus SAT scores during 2015-2016

TABLE 2 ACT Scores vs. All 18 Variables		
Variable	***LogWorth***	***p***
Average Spending on Online Entertainment/Games	2.19	0.00*
2017 Number of Owner Occupied Housing Units	1.63	0.02*
2017 Population of Divorcees	1.62	0.02*
Average Spending on Musical Instruments	1.58	0.02*
Education Attainment: Bachelor's Degree	1.40	0.04*
Average Spending on Pets	1.16	0.07
Employed Civilian Population of Ages 16+	1.07	0.09
Average Spending on Cable and Satellite TV Services	0.74	0.18
Number of Households with Income Below Poverty Level	0.36	0.44
Average Spending on Books	0.28	0.53
Average Spending Education Annually	0.25	0.56
Average Household Size	0.23	0.60
2017 Average Household Income	0.18	0.66
Number of Husband-wife Family Households	0.11	0.78
Average Spending on Alcoholic Beverages	0.11	0.78
Average Spending on Test Prep & Tutoring	0.08	0.83
Average Spending on Health Insurance	0.05	0.89
Number of Multigenerational Households	0.03	0.94
* $p \leq 0.05$		

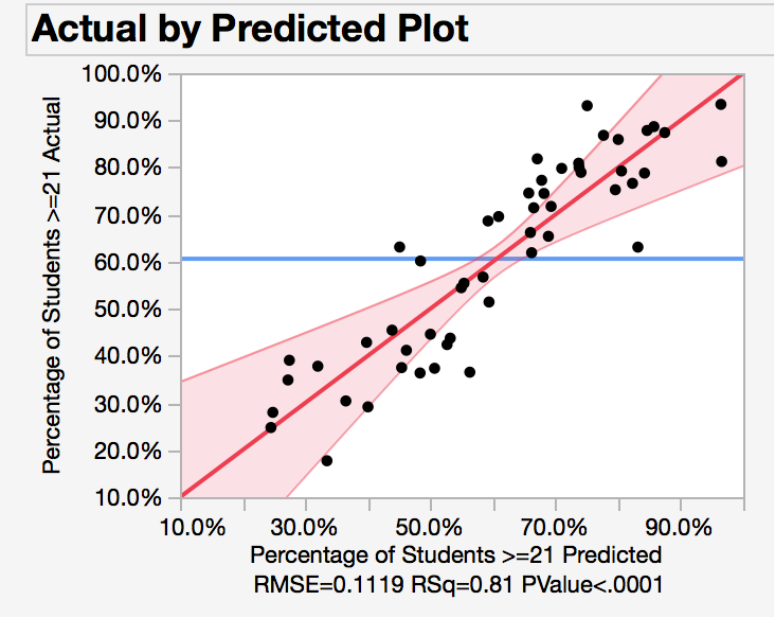


Figure 5. Regression model of all 18 variables at 1-mile radius versus ACT scores during 2015-2016

TABLE 3 SAT Scores vs. All 18 Variables at 2-Miles Radius		
X-Variable	***LogWorth***	***p***
2017 Population of Divorcees	1.35	0.04*
Average Spending on Education Annually	1.18	0.07
Employed Civilian Population of Ages 16+	0.81	0.15
Average Spending on Online Entertainment/Games	0.75	0.18
Number of Husband-Wife Family Households	0.58	0.26
Average Spending on Pets	0.57	0.27
2017 Number of Owner Occupied Housing Units	0.56	0.27
Average Spending on Musical Instruments	0.41	0.39
2017 Average Income	0.31	0.49
Education Attainment: Bachelor's Degree	0.31	0.49
Number of Multigenerational Households	0.29	0.51
Average Spending on Cable and Satellite TV Services	0.14	0.72
Average Spending on Alcoholic Beverages	0.13	0.74
Average Spending on Health Insurance	0.12	0.75
Average Spending on Books	0.11	0.78
Number of Household Incomes Below Poverty Level	0.10	0.80
Average Spending on Test Prep & Tutoring	0.09	0.81
Average Household Size	0.03	0.93
** $p \leq 0.05$ are considered significant*		

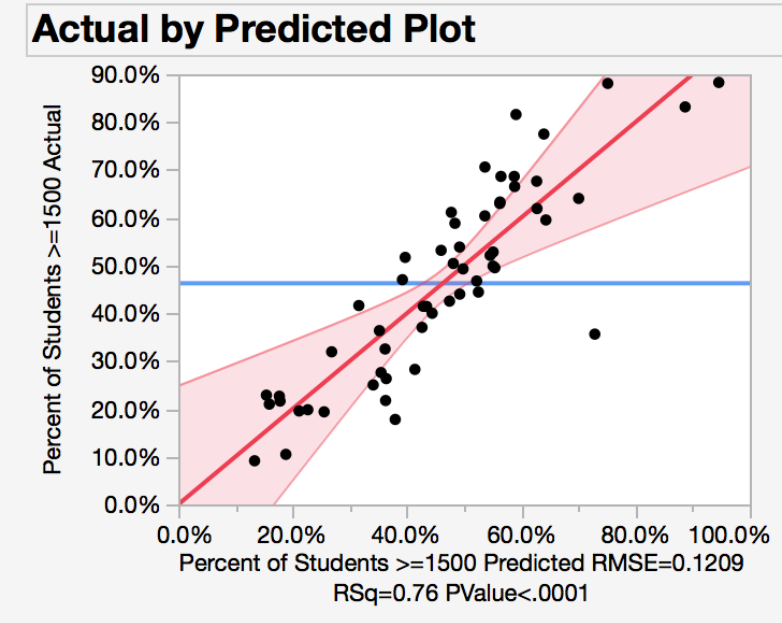


Figure 6. Regression model of all 18 variables at 2-mile radius versus SAT scores during 2015-2016

TABLE 4 SAT Scores vs. All 18 Variables at 3-Mile Radius		
Variable	***LogWorth***	***p***
2017 Population of Divorcees	2.55	0.00*
2017 Number of Owner Occupied Housing Units	1.39	0.04*
Average Spending on Online Entertainment/Games	1.31	0.04*
Average Spending Education Annually	1.30	0.05*
Average Spending on Musical Instruments	0.95	0.11
Employed Civilian Population of Age 16+	0.92	0.12
Number of Husband-wife Family Households	0.85	0.14
Average Spending on Pets	0.84	0.14
Average Spending on Books	0.45	0.35
Average Spending on Cable and Satellite TV Services	0.43	0.37
Number of Multigenerational Households	0.32	0.48
Number of Households with Income Below Poverty Level	0.32	0.48
Education Attainment: Bachelor's Degree	0.24	0.58
Average Household Size	0.20	0.63
Average Spending on Alcoholic Beverages	0.18	0.66
Average Spending on Test Prep & Tutoring	0.14	0.73
2017 Average Household Income	0.06	0.88
Average Spending on Health Insurance	0.01	0.98
* $p \leq 0.05$		

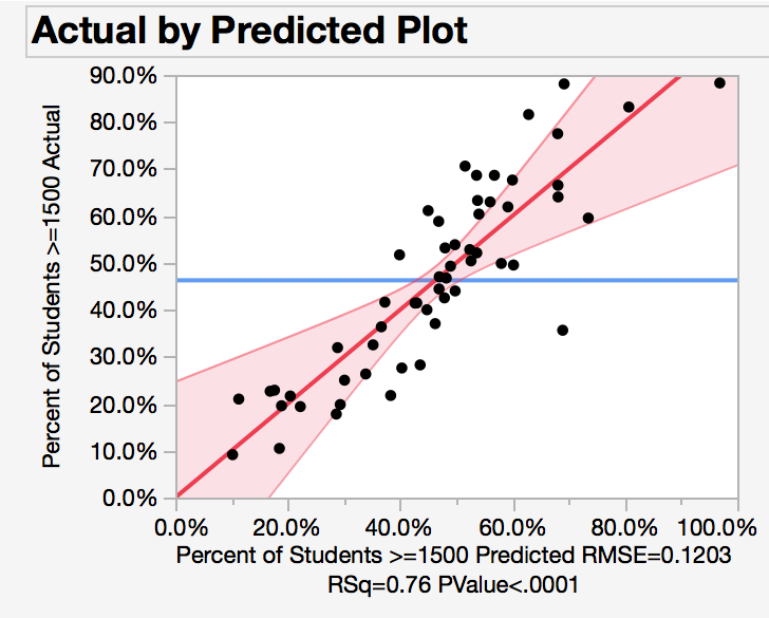


Figure 7. Regression model for all 18 variables at 3-mile radius versus SAT scores during 2015-2016

TABLE 5		
SAT Simple Linear Regressions at 1-Mile Radius		
X-Variable	***R-squared***	***p***
Average Household Size	0.35	0.00*
Average Spending on Education Annually	0.28	0.00*
Average Spending on Cable and Satellite TV Services	0.23	0.00*
Average Spending on Alcoholic Beverages	0.22	0.00*
Number of Multigenerational Households	0.19	0.00*
Average Spending on Health Insurance	0.19	0.00*
Average Spending on Pets	0.19	0.00*
Average Spending on Books	0.19	0.00*
2017 Average Income	0.18	0.00*
Average Spending on Online Entertainment/Games	0.14	0.00*
Average Spending on Musical Instruments	0.12	0.01*
Average Spending on Test Prep & Tutoring	0.09	0.02*
Education Attainment: Bachelor's Degree	0.06	0.07
Number of Household Incomes Below Poverty Level	0.01	0.55
2017 Population of Divorcees	0.00	0.63
2017 Number of Owner Occupied Housing Units	0.00	0.68
Number of Husband-Wife Family Households	0.00	0.90
Employed Civilian Population of Ages 16+	0.00	0.97
** $p \leq 0.05$ are considered significant*		

TABLE 6 ACT Simple Linear Regressions at 1-Mile Radius		
X-Variable	***R-squared***	***p***
Average Household Size	0.37	0.00*
Number of Multigenerational Households	0.26	0.00*
Average Spending on Cable and Satellite TV Services	0.26	0.00*
Average Spending on Education Annually	0.25	0.00*
Average Spending on Alcoholic Beverages	0.23	0.00*
Average Spending on Pets	0.22	0.00*
Average Spending on Health Insurance	0.21	0.00*
Average Spending on Books	0.20	0.00*
2017 Average Income	0.20	0.00*
Average Spending on Online Entertainment/Games	0.15	0.00*
Average Spending on Musical Instruments	0.13	0.01*
Average Spending on Test Prep & Tutoring	0.10	0.02*
Education Attainment: Bachelor's Degree	0.04	0.14
Number of Household Incomes Below Poverty Level	0.04	0.14
Number of Husband-Wife Family Households	0.01	0.44
Employed Civilian Population of Ages 16+	0.01	0.49
2017 Population of Divorcees	0.00	0.79
2017 Number of Owner Occupied Housing Units	0.00	0.86
** $p \leq 0.05$ are considered significant*		

TABLE 7 SAT Simple Linear Regressions at 2-Miles Radius		
X-Variable	***R-squared***	***p***
Average Household Size	0.39	0.00*
Average Spending on Education Annually	0.27	0.00*
Average Spending on Cable and Satellite TV Services	0.24	0.00*
Average Spending on Alcoholic Beverages	0.22	0.00*
Average Spending on Health Insurance	0.20	0.00*
Average Spending on Pets	0.20	0.00*
Average Spending on Books	0.19	0.00*
2017 Average Income	0.12	0.00*
Number of Multigenerational Households	0.13	0.01*
Average Spending on Musical Instruments	0.12	0.01*
Average Spending on Online Entertainment/Games	0.12	0.01*
Average Spending on Test Prep & Tutoring	0.09	0.02*
Education Attainment: Bachelor's Degree	0.06	0.07
2017 Population of Divorcees	0.02	0.36
2017 Number of Owner Occupied Housing Units	0.01	0.40
Number of Household Incomes Below Poverty Level	0.01	0.58
Number of Husband-Wife Family Households	0.00	0.64
Employed Civilian Population of Ages 16+	0.00	0.79
** $p \leq 0.05$ are considered significant*		

TABLE 8 SAT Simple Linear Regressions at 3-Miles Radius		
X-Variable	***R-squared***	***p***
Average Household Size	0.41	0.00*
Average Spending on Cable and Satellite TV Services	0.23	0.00*
Average Spending on Education Annually	0.23	0.00*
Average Spending on Pets	0.20	0.00*
Average Spending on Health Insurance	0.20	0.00*
Average Spending on Alcoholic Beverages	0.20	0.00*
Average Spending on Books	0.18	0.00*
2017 Average Income	0.18	0.00*
Average Spending on Musical Instruments	0.11	0.01*
Average Spending on Online Entertainment/Games	0.10	0.02*
Average Spending on Test Prep & Tutoring	0.09	0.03*
Number of Multigenerational Households	0.07	0.05*
Education Attainment: Bachelor's Degree	0.07	0.06
2017 Population of Divorcees	0.04	0.16
2017 Number of Owner Occupied Housing Units	0.02	0.28
Employed Civilian Population of Ages 16+	0.01	0.52
Number of Household Incomes Below Poverty Level	0.00	0.80
Number of Husband-Wife Family Households	0.00	0.96
** $p \leq 0.05$ are considered significant*		

TABLE 10

Forward Stepwise Regression Model of SAT Predictor Variables

Term	*Estimate*	*Std Error*	*t Ratio*	*Prob>|t|*	*VIF*
Intercept	1858.58	23.51	79.07	0.00	
2010 Multigenerational HHs	-0.06	0.02	-4.13	0.00	3.84
2017 Education: Bachelor's Degree	0.01	0.00	5.95	0.00	1.15
2010 Other Family/Male HHr: Own Kids <18	-0.14	0.04	-3.24	0.00	3.95
2017 Diversity Index	-5.35	0.33	-16.01	0.00	1.39
** VIF < 10*					

TABLE 11

Backward Stepwise Regression Model of SAT Predictor Variables

Term	*Estimate*	*Std Error*	*t Ratio*	*Prob>|t|*	*VIF*
Intercept	1858.58	23.51	79.07	0.00	
2010 Multigenerational HHs	-0.06	0.02	-4.13	0.00	3.84
2017 Education: Bachelor's Degree	0.01	0.00	5.95	0.00	1.15
2010 Other Family/Male HHr: Own Kids <18	-0.14	0.04	-3.24	0.00	3.95
2017 Diversity Index	-5.35	0.33	-16.01	0.00	1.39
** VIF < 10*					

TABLE 12 Correlation to SAT Scores	
Row	**SAT Score**
2017 Average Household Income	0.66
2017 Average Household Size	-0.46
Education:Ind	0.66
Test Prep & Tutoring Srvs:Ind	0.60
Health Insurance:Ind	0.67
Alcoholic Beverages:Ind	0.66
Online Entertainment/Games:Ind	0.59
Books:Ind	0.66
Cable and Satellite TV Services:Ind	0.68
Entertainment/Recreation - Pets:Ind	0.67
Musical Instruments & Acc:Ind	0.62
2010 Multigenerational HHs	-0.50
ACS HHs: Inc Below Poverty Level	-0.42
2017 Education: Bachelor's Degree	0.06
Elementary/HS School Books/Supplies:Ind	0.62
School Meals:Ind	0.63
2017 Occupation: Education/Library	0.00
2010 Other Family/Male HHr: Own Kids <18	-0.47
2010 Other Family/Female HHr: Own Kids <18	-0.46
2010 Spouse in Family HHs	-0.18
ACS Pop 5-17 speak Only English	0.00
Social/Recreation/Civic Clubs Member Fee:Ind	0.65
2010 Husband-wife Fam: Own Kids <18	-0.25
Educational Services - Businesses (NAICS)	-0.05
ACS Pop 5-17 speak Span/No English	-0.36
2017 Diversity Index	-0.60
SAT Score	1.00

TABLE 13 Final Predictors Correlation to SAT Scores	
Row	**SAT Score**
2010 Multigenerational HHs	-0.5
2017 Education: Bachelor's Degree	0.06
2010 Other Family/Male HHr: Own Kids <18	-0.47
2017 Diversity Index	-0.6
SAT Score	1