Geospatial analyses to determine academic success factors in California's K-12 education

Brianna Chew[a], Asish Satpathy[b] and Elaine Wong[c]

[a]University of California, Irvine, Paul Merage School of Business; [b]Arizona State University, W.P. Carey School of Business; [c]University of California, Riverside, School of Business

INTRODUCTION

As the geographical academic achievement gap continues to grow, the gap between the rich and poor widens too[1]. Researchers have attempted to understand this trend by using U.S. administrative data and longitudinal surveys to discover that factors including median income, parental education, and family structure influence students' academic success[2]. However, no studies have focused on an aggregated effect of location specific factors. We address this gap by studying various geospatial attributes such as consumer spending, household demographics, and socioeconomic factors at various school locations throughout California and their effects on students' SAT performance.

DATA SAMPLE

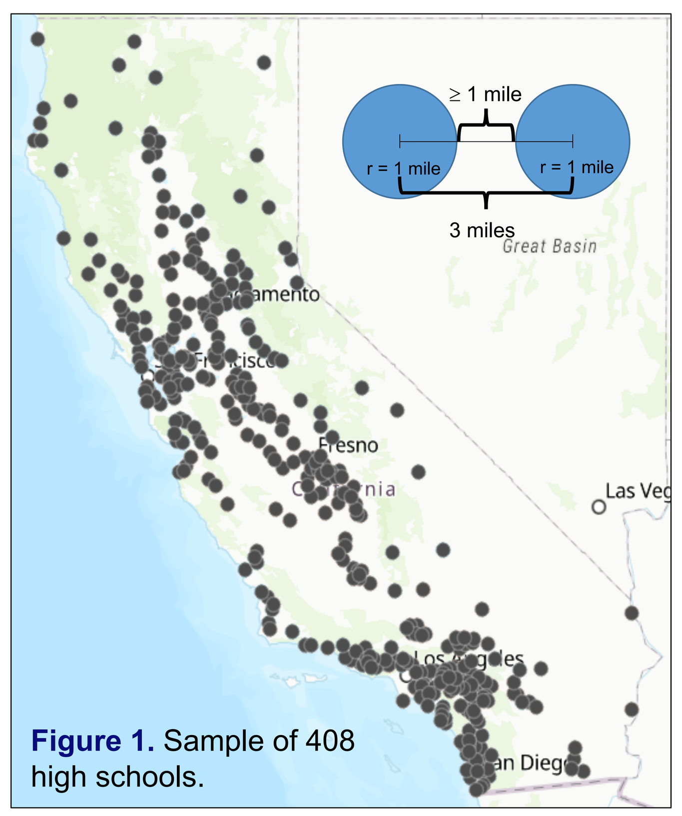

Figure 1. Sample of 408 high schools.

- Restricted study to all 1,070 charter, magnet, and public high schools in California
- Created one-mile radius buffer rings around each school and geo-enriched each ring with 23 consumer spending habits, demographic, and socioeconomic attributes
- To ensure the rings do not overlap, we only selected schools that are at least 3 miles apart from one another and removed all multivariate outliers
- Final sample size is 408 high schools (Figure 1).

METHODS

Exploratory analyses

- We use hot spot analysis to assess each feature and determine if the spatial clusters are statistically significant.
- Using Esri's Tapestry Segmentations, we analyze tapestry data from high and low performing schools to understand the characteristics defining each community.

Statistical analyses

- A stepwise regression model is developed to test the significance and effect of the 23 geo-enriched variables on students' SAT scores.
- A Geographically Weighted Regression (GWR) analysis is developed using the independent variables that survived in the stepwise model.

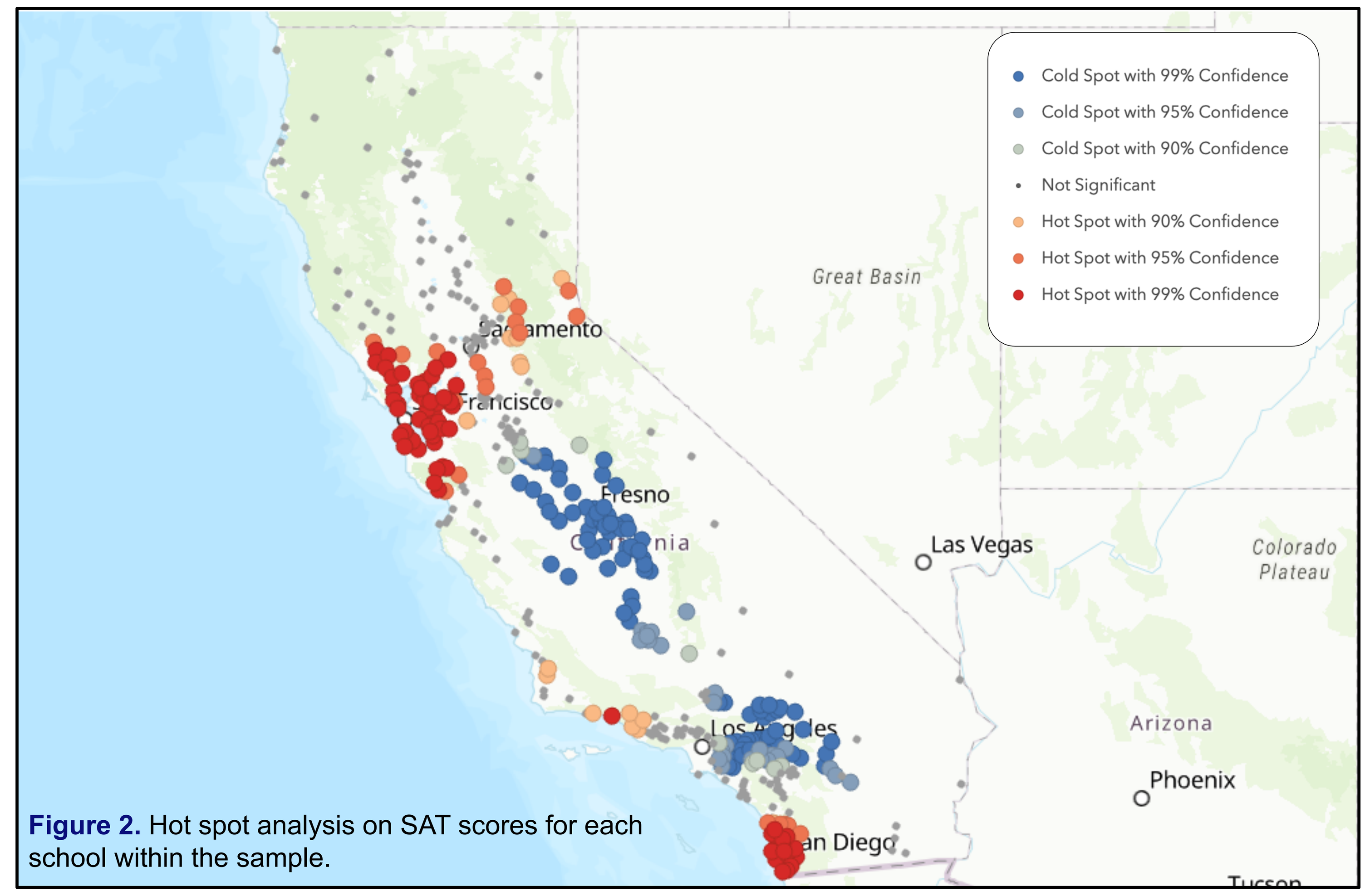

Figure 2. Hot spot analysis on SAT scores for each school within the sample.

Table 1. Stepwise multilinear regression model effects

Independent Variable	*Coefficient Estimate*	*p-value*	*VIF*
Intercept	1645.074		
Average household size	-71.050	0.000	2.332
Spending Potential Index on health insurance	1.367	0.000	1.789
Number of multigenerational households	-0.192	0.000	5.324
Number of husband-wife households	-0.062	0.001	1.664
Number of educational services	1.607	0.014	2.529
Diversity index	-2.165	0.000	
R^2 (adjusted)	0.752		

Table 2. GWR model diagnostics

Estimators	*AICc*
R-squared	0.708
Adjusted R-squared	0.699
AICc	4686.133
Sigma-Squared	5541.386
Sigma-Squared MLE	5364.300
Effective Degrees of Freedom	394.962

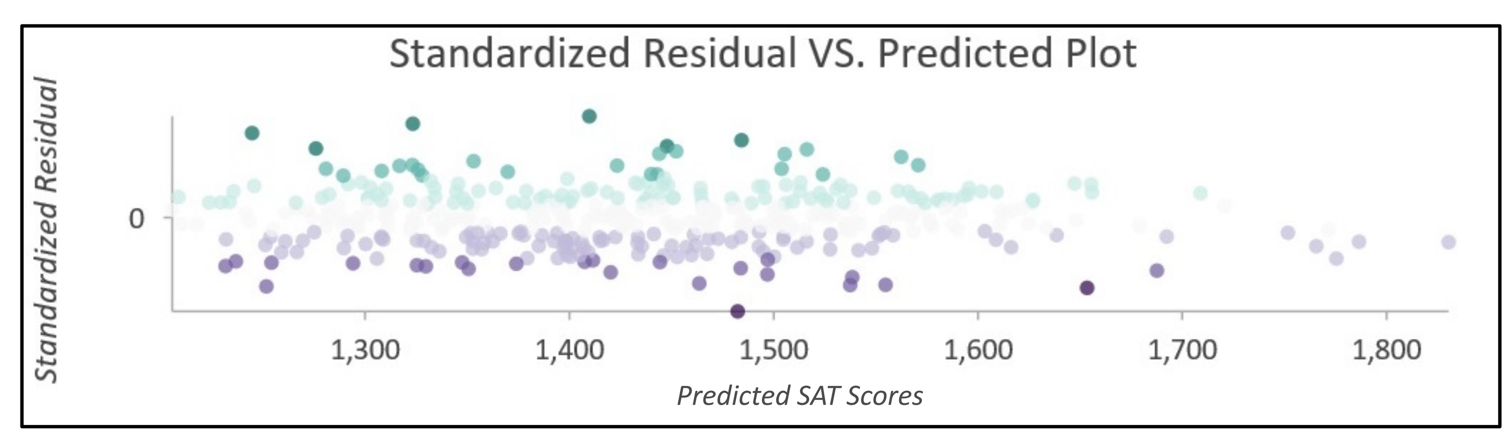

Figure 3. Standardized residual distribution from GWR model.

RESULTS

Exploratory analyses

- We learn that there is an evident academic performance gap between northern and southern California (Figure 2).
- Based on the hot spot and tapestry analyses, we found that low performing schools tend to be in lower socioeconomic areas and the higher performing schools tend to be in higher socioeconomic areas.
- We infer that variables such as income and parental education attainment play a role in students' academic success.

Statistical analyses

- The stepwise regression model consisted of six independent variables that significantly affect students' academic performance (Table 1).
- The GWR results showed consistent predictions of SAT scores (Table 2).
- The distribution of standardized residuals obtained from the GWR confirms the robustness of the stepwise regression model (Figure 3).
- Figures 4-9 show the positive and negative relationships featured in the regression model results.

CONCLUSION

- Location relevant metrics have a compelling effect on overall academic performances.
- Factors that demonstrated positive effects are spending on health insurance, two parent households, and access to educational entities, all of which are more common in high performing schools.
- Factors that demonstrated negative effects are household size, multigenerational living conditions, and higher diversity rates, all of which are more common in lower performing school communities.
- We believe our results will help develop better public policies to enhance overall academic performances. To bridge the academic performance gap, it is important for school educators and leaders to be more proactive about helping students earlier in their academic careers.

REFERENCES

[1]Reardon, S. F. 2011. "The Widening Academic Achievement Gap between the Rich and the Poor." (91-112).

[2]Ladd, H. F. 2012. Journal of Policy Analysis and Management 31: 2.

This study is now published in *Annals of GIS*. Access to paper: https://www.tandfonline.com/doi/full/10.1080/19475683.2020.1739141

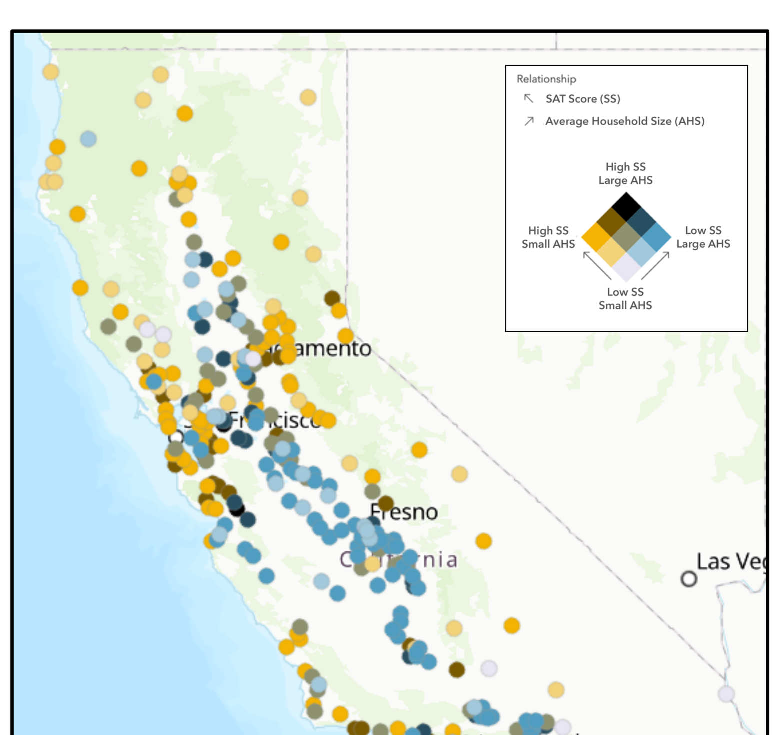

Figure 4. Negative relationship between average household size and SAT scores.

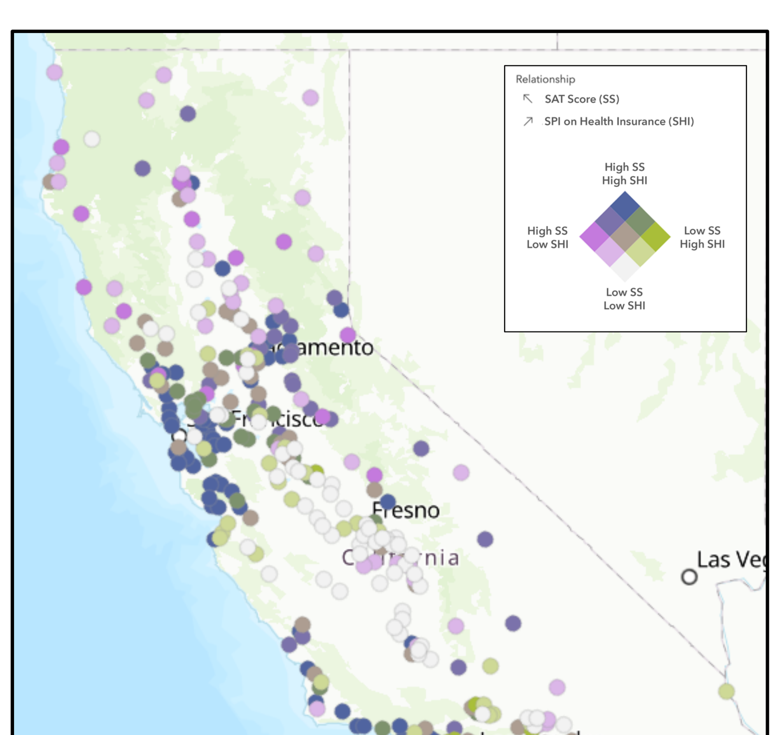

Figure 5. Positive relationship between spending on health insurance and SAT scores.

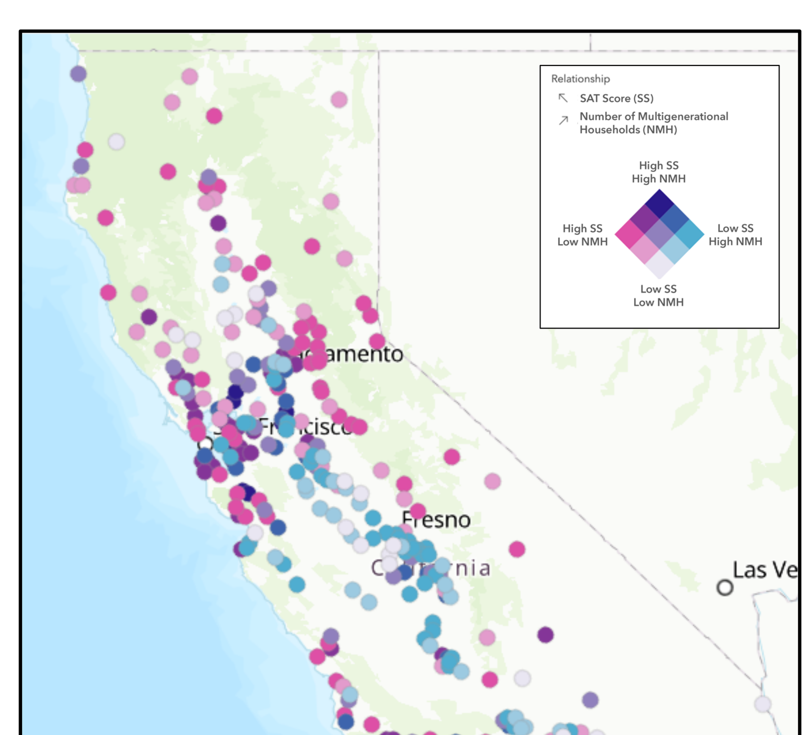

Figure 6. Negative relationship between number of multigenerational households and SAT scores.

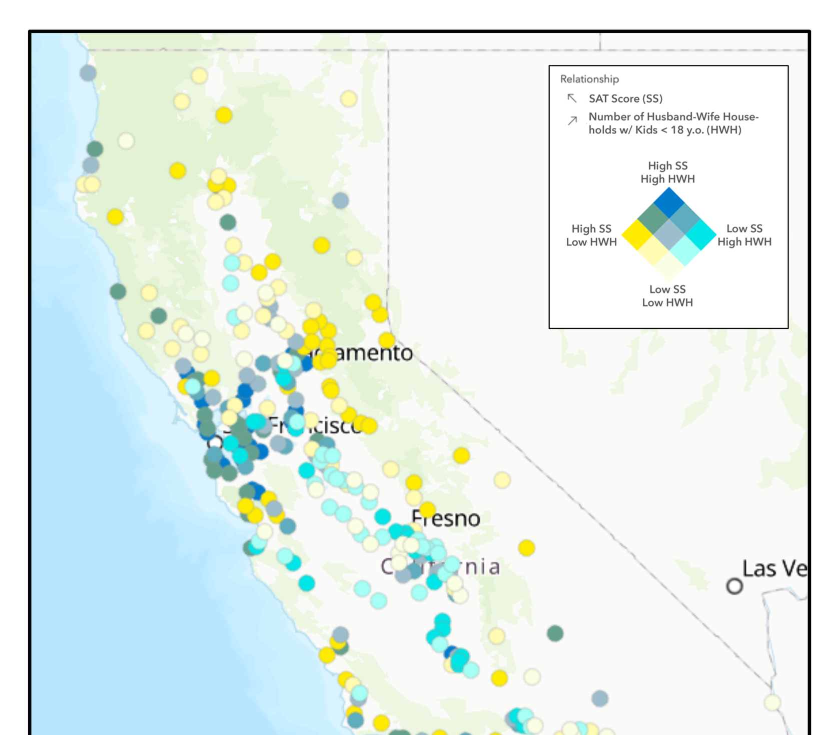

Figure 7 Positive relationship between number of husband-wife households and SAT scores.

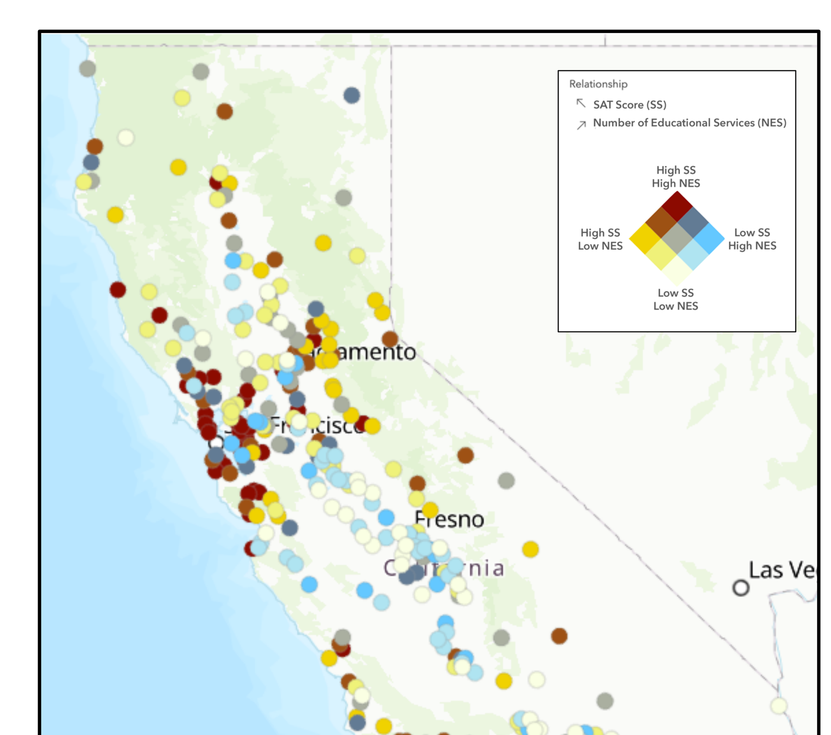

Figure 8. Positive relationship between number of educational services and SAT scores.

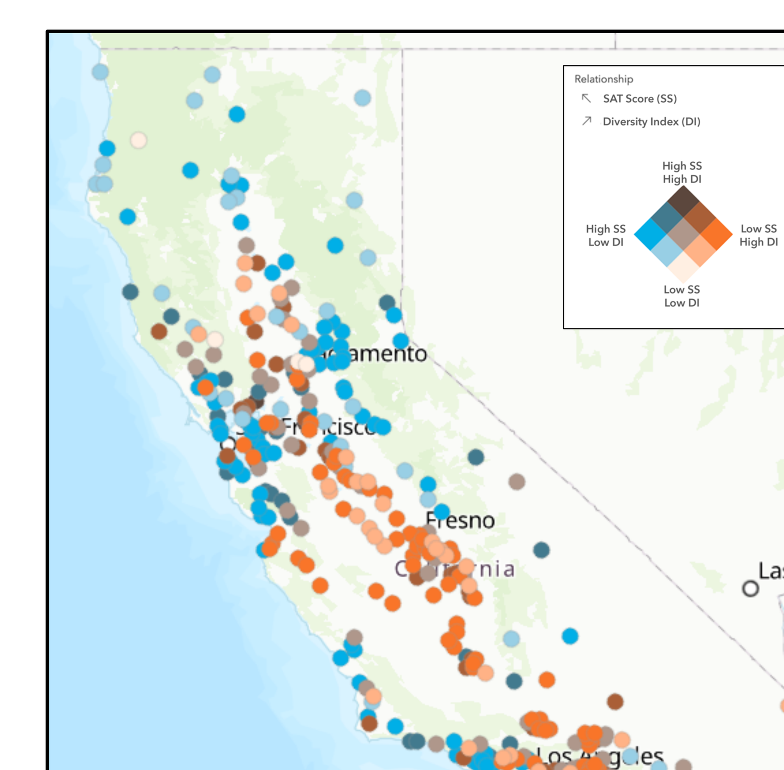

Figure 9. Negative relationship between diversity and SAT scores.