Final Project-PPD 631

Inrak Lee

Topic

What factors affect the scholastic achievement of schools?

Problem

The scholastic achievement of schools is likely to vary by the location of the school. For example, my town, Cypress is very popular for living, because the Cypress school district has lots of good schools including Oxford school. On the other hand, the downtown LA is usually said to be not good for growing children partly because of its low scholastic achievement of schools.

I was curious about which factors affect the scholastic achievement of schools. Wealth of parents, education of parents, ethnicity of students, or financial soundness of school district can be the examples of the factors. Furthermore, I wanted to confirm which factor is most related to the scholastic achievement of schools and double check them by using GIS.

The relationship between the scholastic achievement of schools and each factor can be identified by different colors or different sized dots by using GIS. The statistical approach (correlation) could be used also.

Data

At first, I had to decide what indicators are most appropriate to represent the factors. For example, which indicator is best for the measurement for the scholastic achievement of school or school district?

I set several candidate indicators for each factors and finally chose one indicator for one factor according to the availability of data. Regarding the analysis area, I decided

to analyze the public elementary schools in California by its school districts, because I thought that the number of California school districts, over 800, was enough to derive the credible conclusion.

Data collection

Factors	Indicators	Data sources
Scholastic achievement of schools	Academic Performance Index (from 1 to 10)	California dep. of education www.cde.ca.gov
Ethnicity of students	The percent of Hispanic(White, Asian) student	US Census Bureau www.census.gov
Financial soundness of school districts	The expenditure per student	California dep. of education www.cde.ca.gov
Education of parents	Parents education grade (from 1 to 5)	US Census Bureau www.census.gov
Wealth of parents	The percent of student in poverty	California dep. of education www.cde.ca.gov
California school districts name & shapefile		TIGER

Combining Data

The most difficult job was connecting the indicator data to their school district, because California department of education and US census bureau are using different school district codes. I had to sort the data by the name of school districts and had to compare them. (The function of Excel, 'vlookup' was also helpful)

Before connecting data

DNAME	Dcode	ST_RANK		Code	poverty %
ABC Unified	64212	6.3	ABC Unified School District	01620	16.6%
Ackerman Elementary	66761	7.0	Ackerman Elementary School District	01680	11.4%
Acton-Agua Dulce Unified	75309	6.5	Acton-Agua Dulce Unified School District	00001	10.5%
Adelanto Elementary	67587	3.6	Adelanto Elementary School District	01710	24.8%
Alameda City Unified	61119	7.9	Alameda City Unified School District	01770	16.9%

Albany City Unified	61127	9.7	Albany City Unified School District	01860	9.6%
Alexander Valley Union Element	70599	9.0	Alexander Valley Union Elementary School District	#N/A	13.4%
Alhambra Unified	75713	7.5	Alhambra Unified School District	00153	23.5%
Alisal Union	65961	2.1	Alisal Union Elementary School District	#N/A	38.6%
Allensworth Elementary	71795	1.0	Allensworth Elementary School District	01980	35.1%
Alpaugh Unified	71803	1.5	Alpaugh Unified School District	02010	33.7%

After connecting data

Dcode	DNAME	RANK	Asian %	Hispanic %	White %	Parents EDU	Poverty %
01620	ABC Unified	6.3	24.1	45.3	7.9	3.1	16.6%
01680	Ackerman Elementary	7.0	1.0	12.0	81.0	3.4	11.4%
00001	Acton-Agua Dulce Unified	6.5	1.0	32.0	61.5	3.0	10.5%
01710	Adelanto Elementary	3.6	1.5	62.6	10.2	2.4	24.8%
01770	Alameda City Unified	7.9	25.7	11.9	33.5	3.5	16.9%
01860	Albany City Unified	9.7	31.0	14.3	37.7	4.3	9.6%
01890	Alexander Valley Union Element	9.0	1.0	24.0	72.0	3.6	13.4%
00153	Alhambra Unified	7.5	46.8	45.4	4.7	2.9	23.5%
01950	Alisal Union	2.1	0.5	92.1	2.8	1.9	38.6%
01980	Allensworth Elementary	1.0	0.0	98.0	0.0	1.6	35.1%
02010	Alpaugh Unified	1.5	5.5	56.0	33.0	2.2	33.7%
02070	Alpine County Unified	5.0	0.0	0.0	52.0	3.1	26.4%
02100	Alpine Union Elementary	8.3	1.8	14.0	74.5	3.4	9.9%
02160	Alta Loma Elementary	8.4	5.0	35.8	47.3	3.4	10.6%
02220	Alta Vista Elementary	1.0	2.0	73.0	21.0	1.4	51.5%
02250	Alta-Dutch Flat Union Elementa	5.0	0.0	9.0	88.0	3.2	20.1%
02310	Alum Rock Union Elementary	3.6	9.9	78.7	2.2	2.2	24.1%
02360	Alview-Dairyland Union Element	8.0	0.0	66.0	34.0	2.1	31.7%

After I completed an Excel spreadsheet, I could connect those data to a shapefile. It was so easy, because the school district codes of shapefile are same as those of Excel spreadsheet and GIS software has powerful function to connect them.

Finally, I succeeded to make my own data frame which had several layers: the API rank (from 1 to 10, the average of all elementary school in the school district), the grade of parent education (from 1 to 5), the percent of Asian student, the percent of white student, the percent of Hispanic student and the percent of student in poverty.

Conclusion

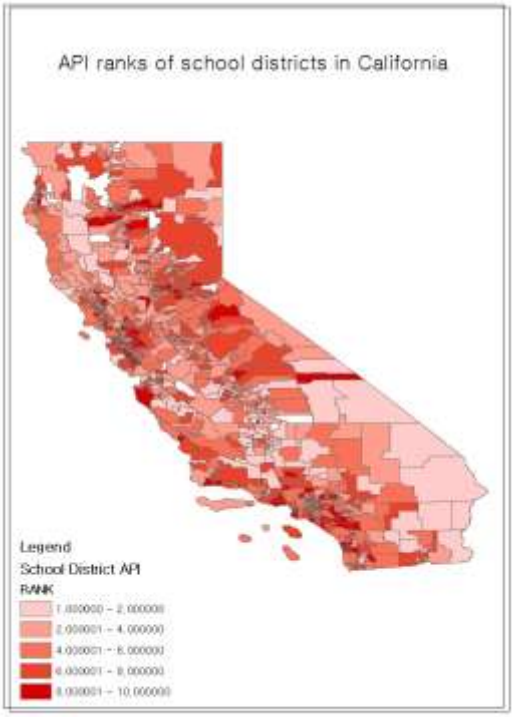
Before I confirmed the factors which might be related to the scholastic performance of school districts, I checked the correlation between each factor and the scholastic performance of school districts by using Excel. Finally, I confirmed five factors which have high correlation with the scholastic performance of school districts.

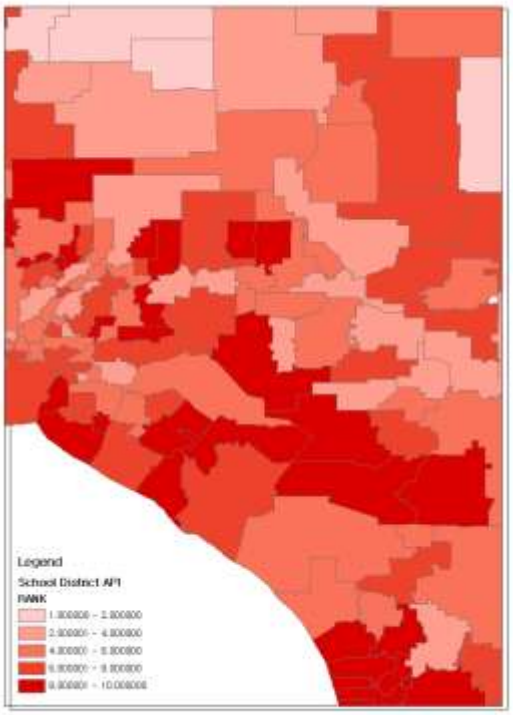
The correlations

	Parents education grade	The percent of student in poverty	The percent of Hispanic student	The percent of Asian student	The percent of White student
Correlation	0.6295541	-0.492093	-0.446366	0.3520769	0.3031958

Now, I want to display these relationships on maps and double check whether these relationships are real or not.

Firstly, I displayed the API Ranks of all school districts in California. Generally, schools in urban area have higher ranks than those in rural area.

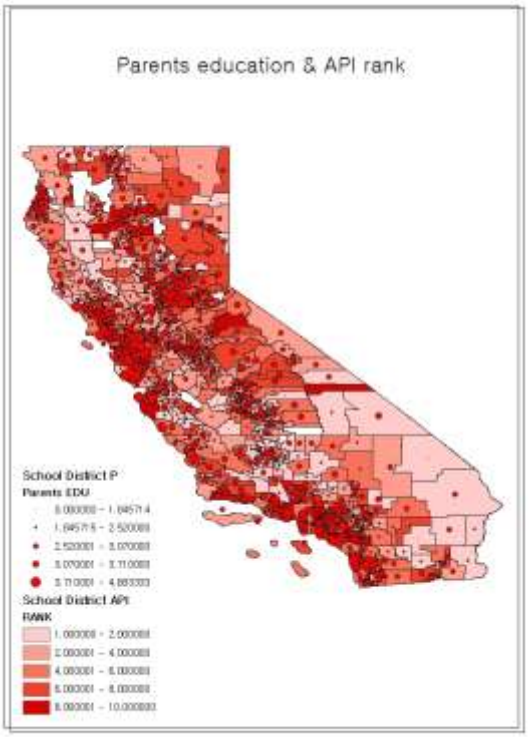


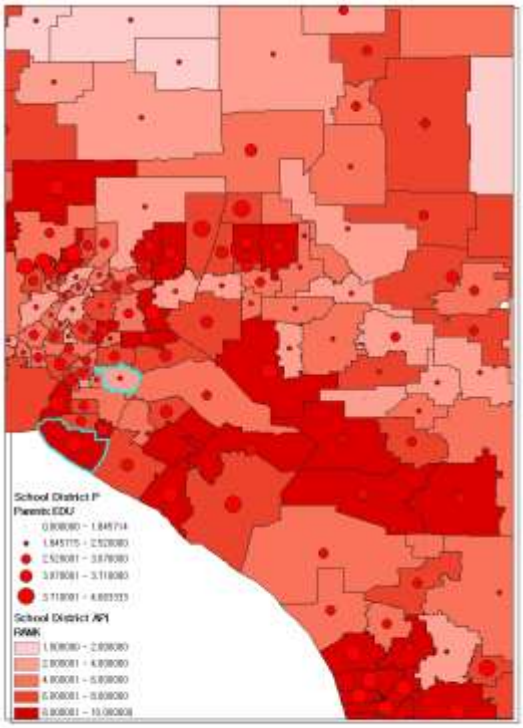


Secondly, I displayed the parent education grades* on this map. Generally speaking, we can double check that the school district which has high parent education grade has high API rank.

* 1: Not High School Graduate, 2: High School Graduate, 3: Some College, 4: College Graduate, 5: Graduate School

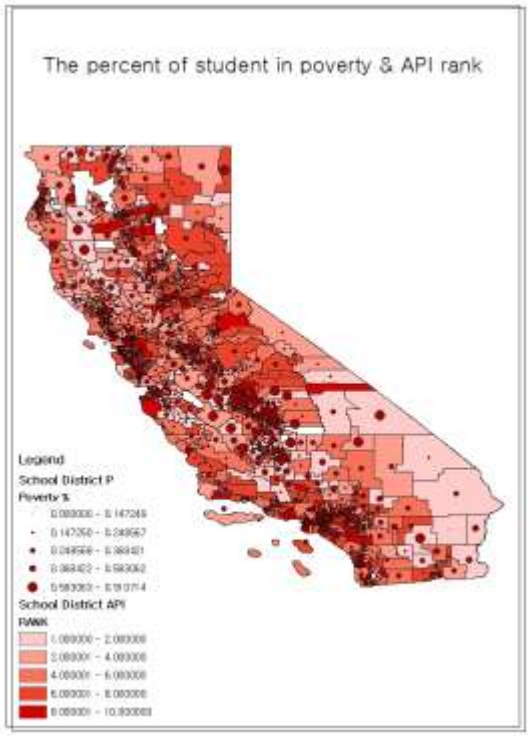
For example, the Huntington beach city elementary school district which has 3.8 average parent education grade gets 8.9 API rank. On the contrary, the Anaheim elementary school district which has 2.2 parent education grade gets 3.4 API rank.

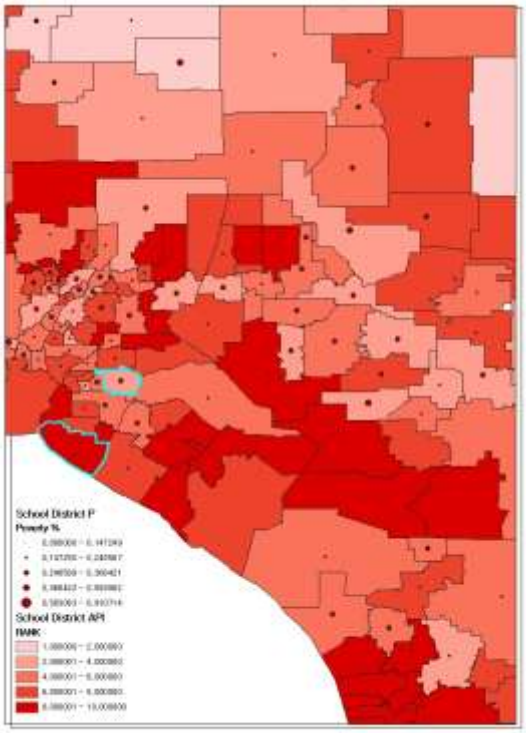




Thirdly, I displayed the percent of student in poverty on the map. Generally speaking, we can also double check that the school district which has high percent of student in poverty has low API rank. In other words, the financial situation of student can affect the academic achievement of the student. For example, the Huntington beach city elementary school district which has 9% (0.09) of students in poverty gets 8.9

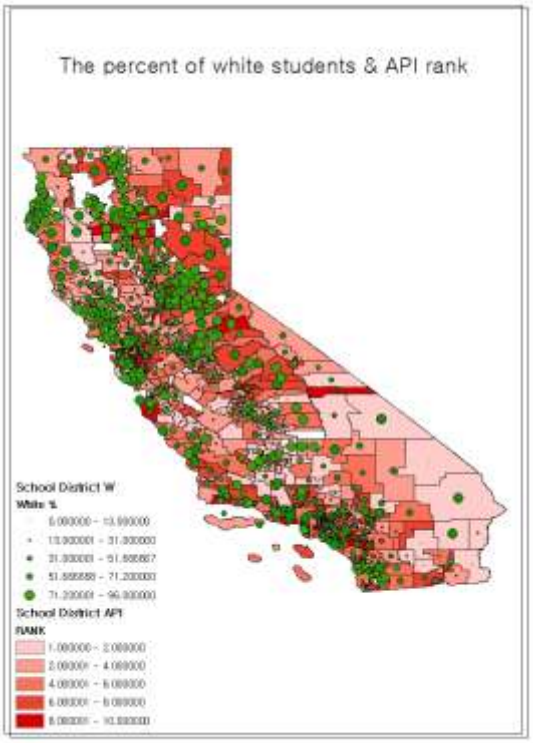
API rank. On the contrary, the Anaheim elementary school district which has 27.6% (0.276) of students in poverty gets 3.4 API rank.

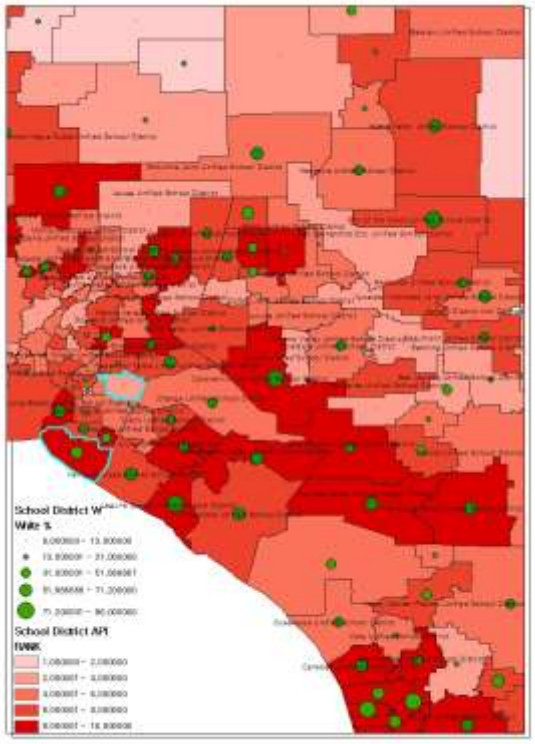




On the other hand, the below maps display the relationships between the percent of white students in a school district and the API rank of that district. Generally speaking, we can also double check that the school district which has high percent of white student has high API rank. For example, the Huntington beach city elementary school district which has 64% of white students gets 8.9 API rank. On the contrary, the Anaheim elementary school district which has 5% of white students gets 3.4 API rank.

The relationship between the percent of Hispanic students in a school district and the API rank of that district is contrary to the case of white students, but the relationship between the percent of Asian students is similar to the case of white students.





Data limitations

I experienced several data limitations in doing this project. For example, I would like to use income data or house price data for the wealth of parents, but the data I got were categorized not by school districts but by cities.

On the other hand, I think that the existence of other factors which cannot be imagined easily but have much relationship with the scholastic achievement of schools can impact my analysis much.

Epilogue

I could experience the tremendous usage of GIS by taking this class. Furthermore, it was so interesting to do the project by myself, even though collecting & modifying data was so hard working.

Someday, I want to make a GIS map which has important information about my company's work (financial industry in Korea), and present it to the colleagues of company.