

SAT scores

...

Where you live

SAT Scores

Standardized test that most University required.

What impact SAT scores, is it related to where you live?

DATA

2019 California SAT Scores:

26 Columns: 9 needed.

2580 Records

CA School Districts Boundaries Shapes

Authoritative geographic data source for all school district area boundaries in California for the 2018-19 academic year

Connect By:

School District Code:CDCode

Data Import & Cleaning

...	NumERWBenchmark11	PctERWBenchmark11	NumMathBenchmark11	PctMathBe
...	NaN	NaN	NaN	
...	*	*	*	
...	42	24.14	12	

Data Import & Cleaning

	CDCode	RType	DName	Enroll12
1981	900000.0	C	NaN	2226.0
1982	1900000.0	C	NaN	114618.0
1983	2000000.0	C	NaN	2352.0
1984	5700000.0	C	NaN	2299.0
1985	1500000.0	C	NaN	14229.0

Data Import & Cleaning

	count	mean	std	min	25%	50%	75%	max
RType								
C	58.0	8442.241379	17344.164490	0.0	525.5	2325.5	7477.25	114618.0
D	539.0	907.730983	2165.403522	0.0	104.5	325.0	1063.50	41873.0
S	1981.0	232.673397	216.578502	0.0	53.0	141.0	398.00	1135.0
X	1.0	489650.000000	NaN	489650.0	489650.0	489650.0	489650.00	489650.0

Data Import & Cleaning

Feature	Type	Dataset	Description
code	float	sat_2019_ca	School District Code in CA, each district has a unique code
name	object	sat_2019_ca	School District Name
num_student	float	sat_2019_ca	Number of students enrolled in 12th grade in the county
num_taken	float	sat_2019_ca	Number of students taken SAT test in the county
erw_num	int	sat_2019_ca	Number of students who has taken SAT and exceed the writting benchmark
erw_per	float	sat_2019_ca	Percentge of student who exceed writting benchmark
math_num	int	sat_2019_ca	Number of students who has taken SAT and exceed the math benchmark
math_per	float	sat_2019_ca	Percentge of student who exceed math benchmark

Exploratory Data Analysis

Add another feature, percentage of student who taken SAT in 12th grade.

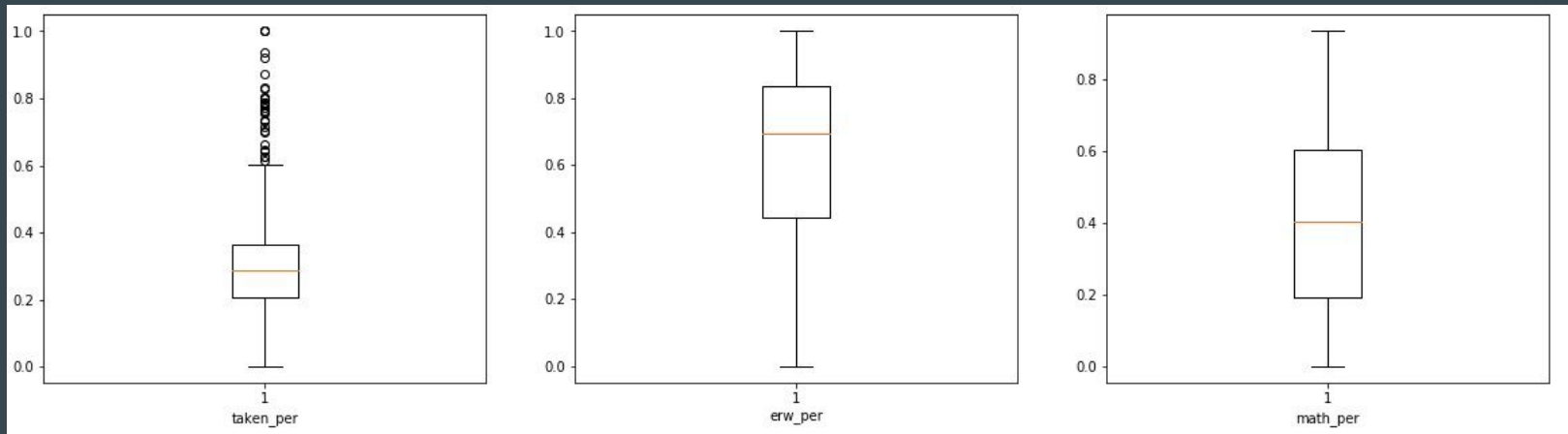
Exploratory Data Analysis

	count	mean	std	min	25%	50%	75%	max
code	501.0	2.974097e+06	1.548312e+06	110017.000000	1.864162e+06	3.066647e+06	4.168924e+06	5872769.00
num_student	501.0	9.720120e+02	2.232984e+03	2.000000	1.400000e+02	3.830000e+02	1.136000e+03	41873.00
num_taken	501.0	3.401277e+02	1.011470e+03	1.000000	2.700000e+01	9.200000e+01	3.560000e+02	19988.00
erw_num	501.0	2.316387e+02	5.661537e+02	0.000000	1.800000e+01	6.600000e+01	2.780000e+02	10448.00
erw_per	501.0	5.874543e+01	3.228995e+01	0.000000	4.450000e+01	6.955000e+01	8.340000e+01	100.00
math_num	501.0	1.623034e+02	3.591206e+02	0.000000	9.000000e+00	4.400000e+01	1.950000e+02	5981.00
math_per	501.0	3.983868e+01	2.682076e+01	0.000000	1.920000e+01	4.011000e+01	6.008000e+01	93.33
taken_per	501.0	2.992615e-01	1.704152e-01	0.002611	2.067039e-01	2.871012e-01	3.652531e-01	1.00

Exploratory Data Analysis

	code	name	num_student	num_taken	erw_num	erw_per	math_num	math_per	taken_per
296	1964733.0	Los Angeles Unified	41873.0	19988.0	10448	52.27	5981	29.92	0.477348

Visualize the Data

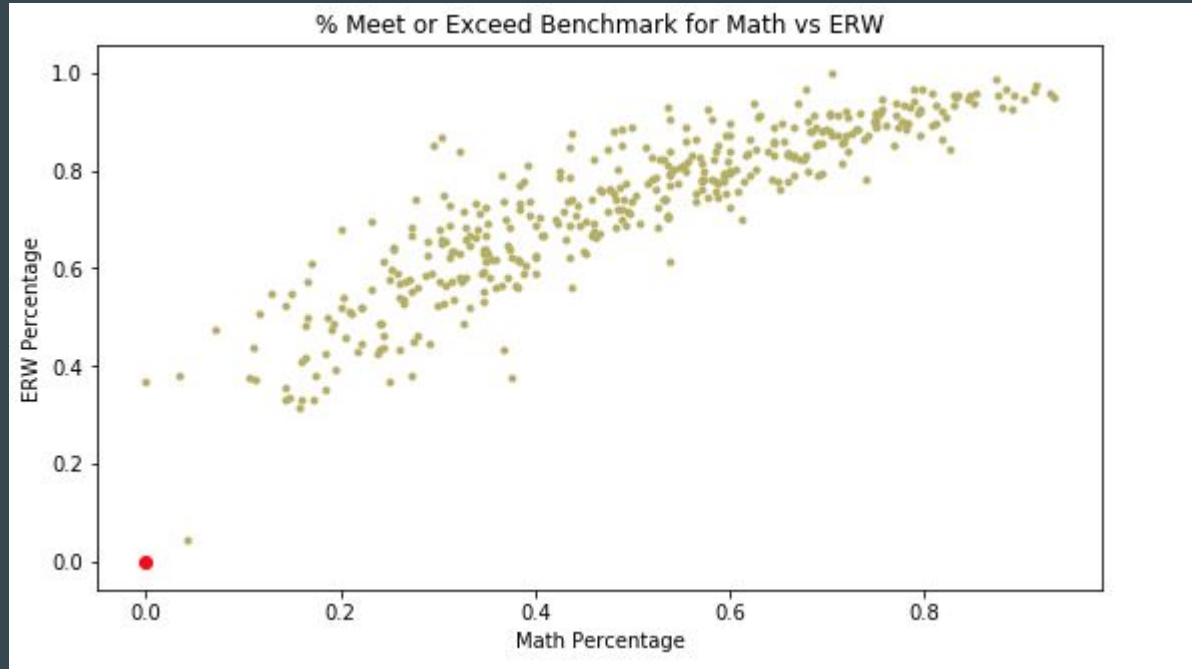


Visualize the Data

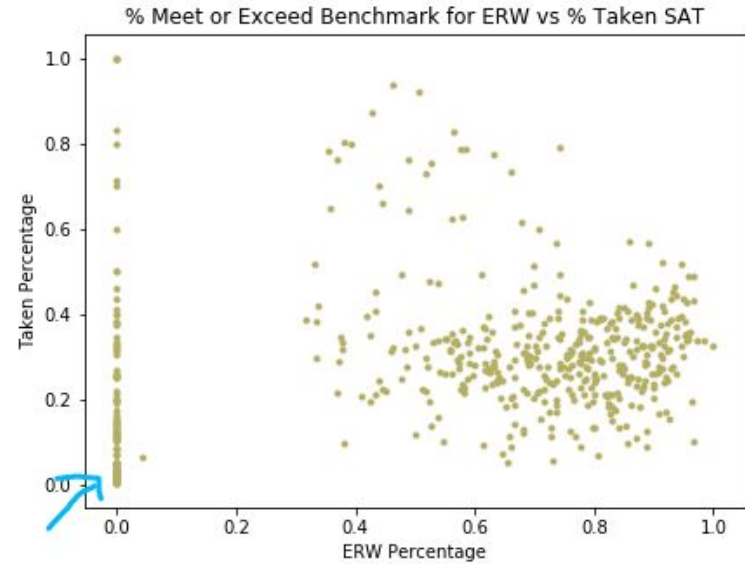
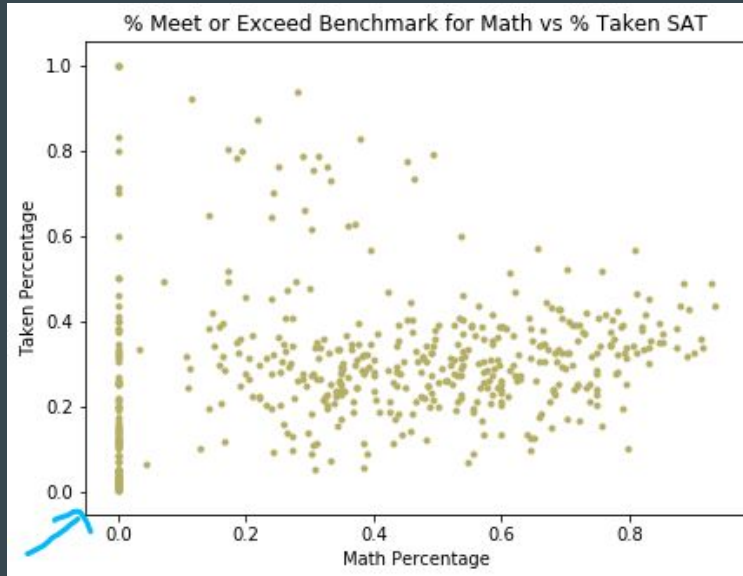
	num_student	num_taken	taken_per
221	2.0	2.0	1.000000
414	3.0	3.0	1.000000
64	3.0	3.0	1.000000
233	5.0	4.0	0.800000
13	6.0	5.0	0.833333
452	7.0	5.0	0.714286
428	10.0	7.0	0.700000

115	1858.0	1195.0	0.643165
112	2261.0	1870.0	0.827068
408	2434.0	1788.0	0.734593
82	2481.0	1992.0	0.802902
357	2520.0	1770.0	0.702381
185	4063.0	2690.0	0.662072
205	4128.0	3149.0	0.762839
400	5929.0	3705.0	0.624895

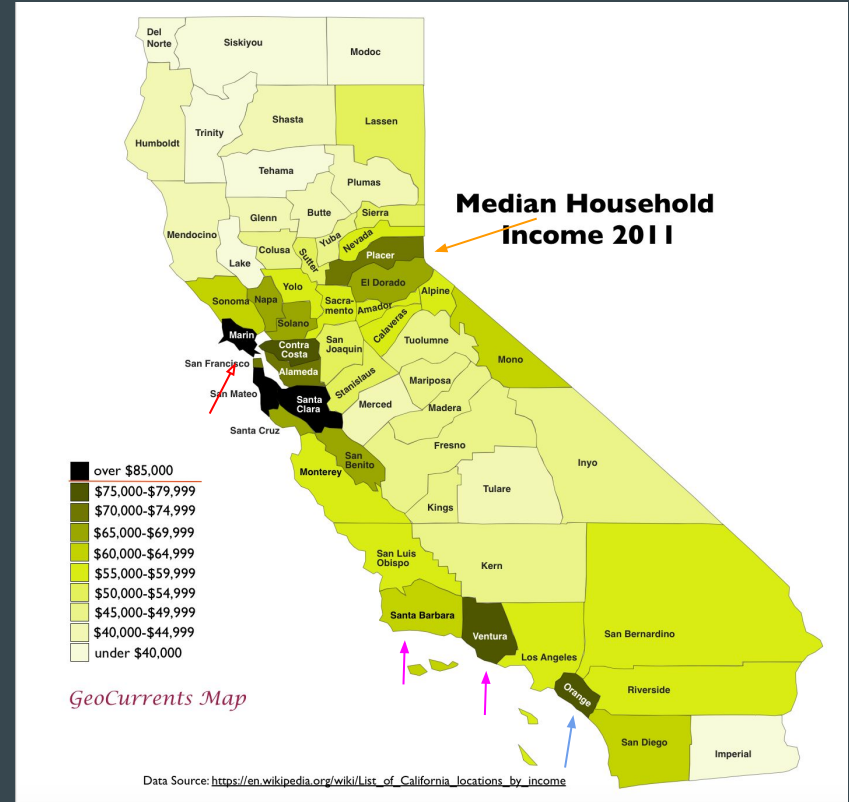
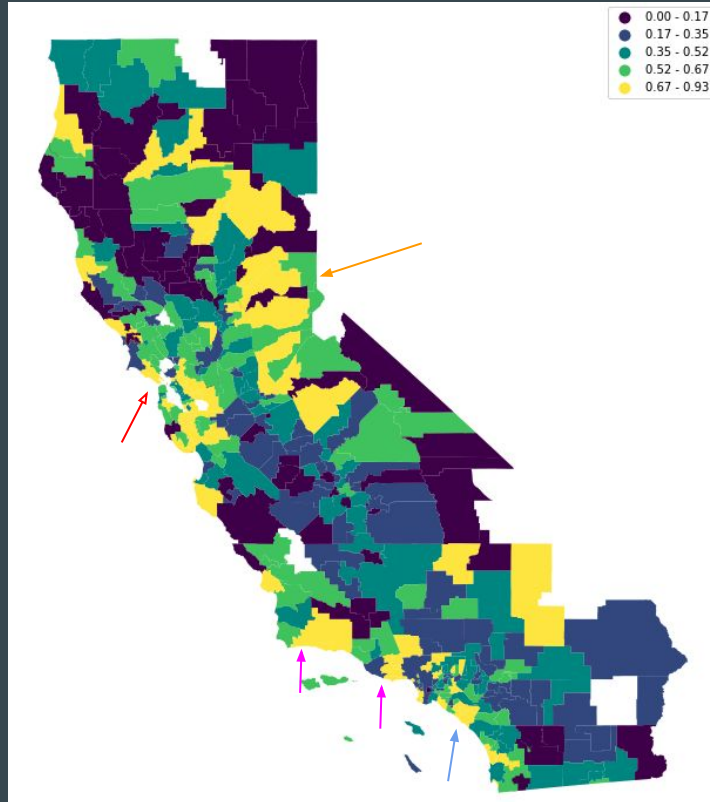
Visualize the Data



Visualize the Data



Visualize the Data



Conclusions

Students who lived high income areas tend to have higher SAT scores. But more data need to prove.