Transportation Access & Student Success Metrics in California

Group 2 – Caiden Puma, Sidney Lin, Vicki Pu, Thomas Cronin

Topic & Research Question

Topic

- Parent Influence on Education/School Access and Demographics.
- Pivoted from national SAT scores as a metric to California district-specific data.

Final Research Question

Is there a statistically significant relationship between transportation access (measured by buses per 1000 students) and student success outcomes such as graduation rates, college enrollment, or chronic absenteeism?

Data Sources & Variables

Data Sources

- California Department of Education (2022) by the state of California.
- U.S. State-by-State Transportation Statistics (2021-2022) by School Bus Fleet Magazine.

Data Cleaning

- Merged datasets via overlapping districts.
- Removed observations with missing Entries.
- Converted all values to integer data types.
- Normalized response variables to student cohort size.

Variables

- Buses per 1000 Students (Explanatory)
- Absenteeism Rate
- College Going Rate
- Enrolled Out & In-State Rate
- Not Enrolled Rate
- HS Graduation Rate
- Adult Ed. Diploma & GED Rate
- Dropout Rate
- Still Enrolled Rate

Methodology

Initial Characterization

Summarized the dataset, graphed variables via scatter/pie/box plots, etc. to better understand the data.

Correlation Matrix

Created a heat map to further understand relationships between the variables to further explore.

OLS Regression

Used regression to quantify the strength and direction of the relationships between variables, as well as determine if they were significant.

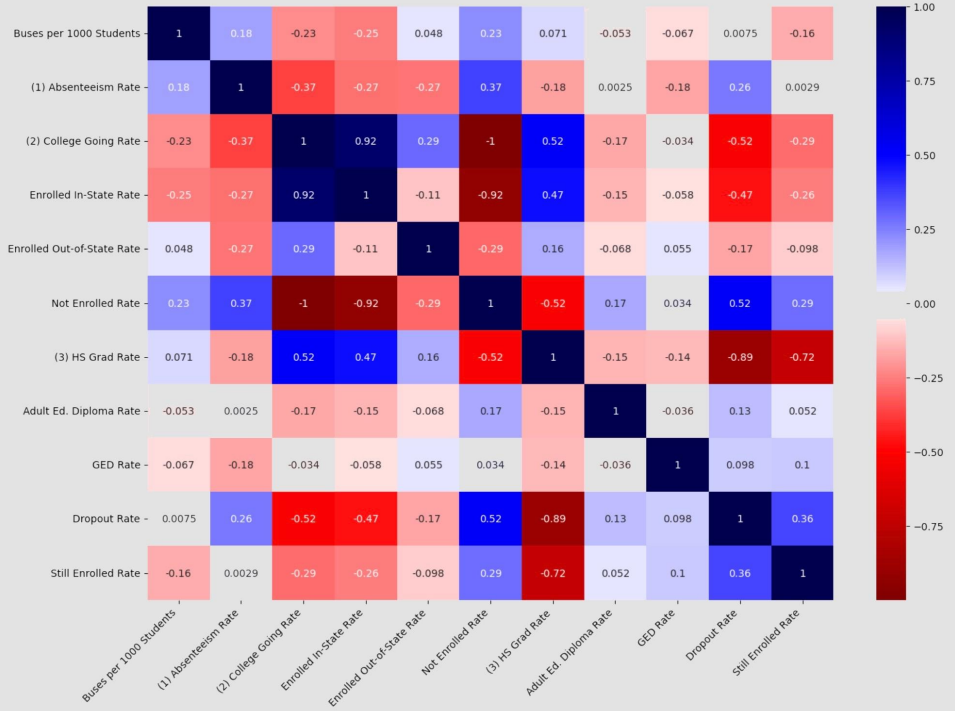
Assumption Checking

Used Shapiro-Wilk to check normality, Breusch-Pagan to check for equal variance/residuals, and Durbin-Watson to check for independence of residuals to ensure validity of the OLS modeling.

Correlation Results

Key Takeaways:

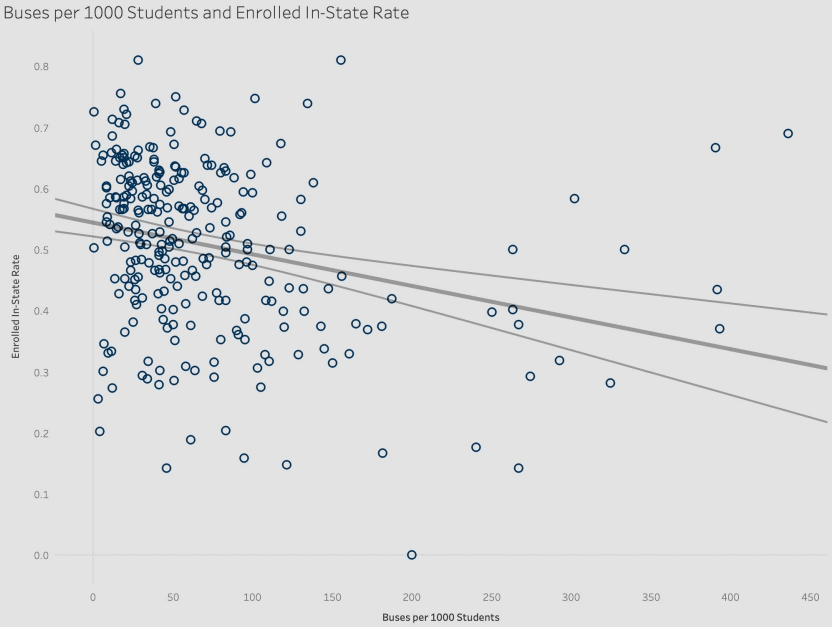
- Negative Correlations
  - College Going Rate (-0.23)
  - Enrolled In-State Rate (-0.25)
- Positive Correlations
  - Not Enrolled Rate (+0.23)
  - Absenteeism Rate (+0.18)
- Thus, districts with more buses per 1000 students seem to have decreased positive outcomes and increased negative outcomes.

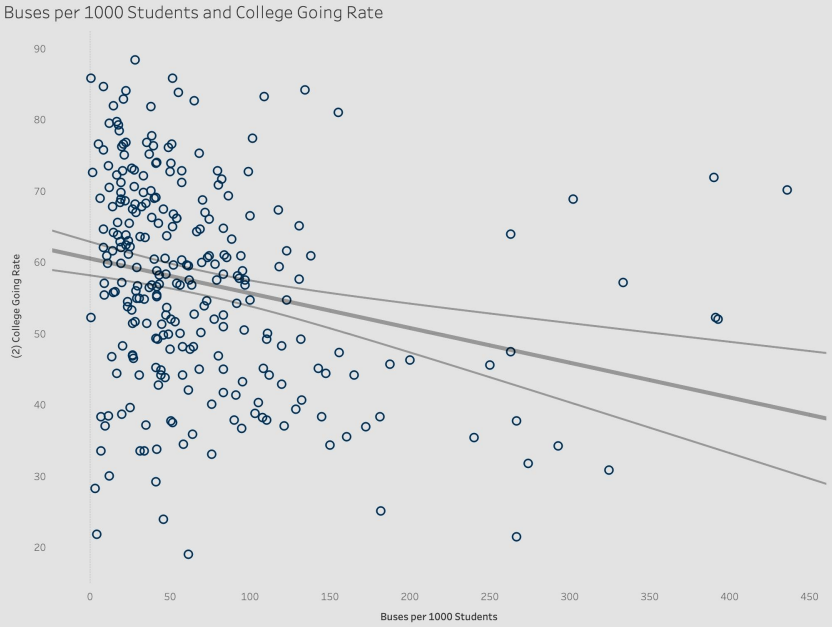


Significant OLS Regression Results

	P-Value	Adjusted R-Squared	Coefficient	Normality Satisfied	Residual Variance Satisfied	Independence Satisfied
Still Enrolled	0.00643	0.023	-0.001	No	Yes	Yes
College Going	0.000162	0.047	-0.0441	Yes	Yes	Yes
College Enrolled	0.000162	0.047	-0.0004	Yes	Yes	Yes
In-State College	1.95E-5	0.061	-0.0005	No	No	Yes
Not Enrolled	0.000162	0.047	0.0004	Yes	Yes	Yes
Chronic Absenteeism	0.00241	0.03	0.0294	No	Yes	Yes

Negative Correlations

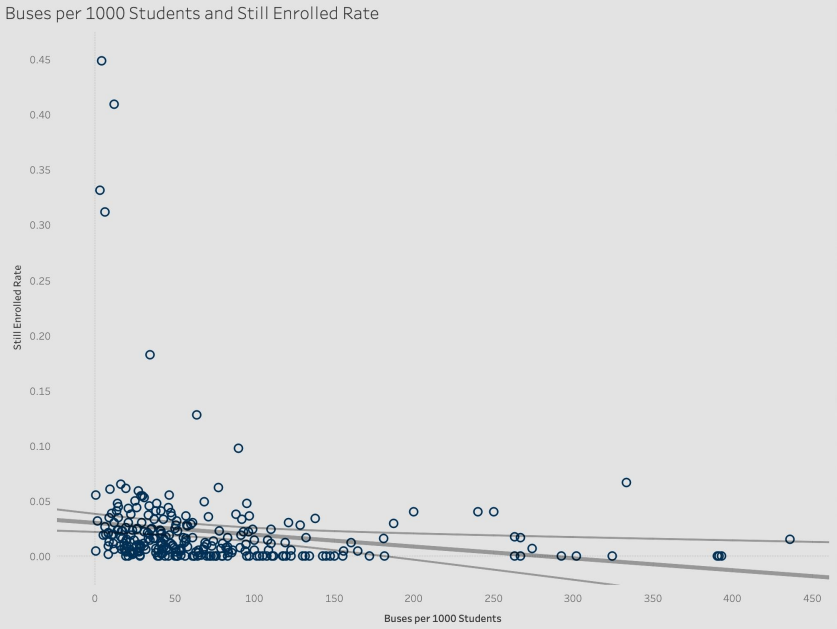




Enrolled In-State Rate (Decrease of 0.05%)

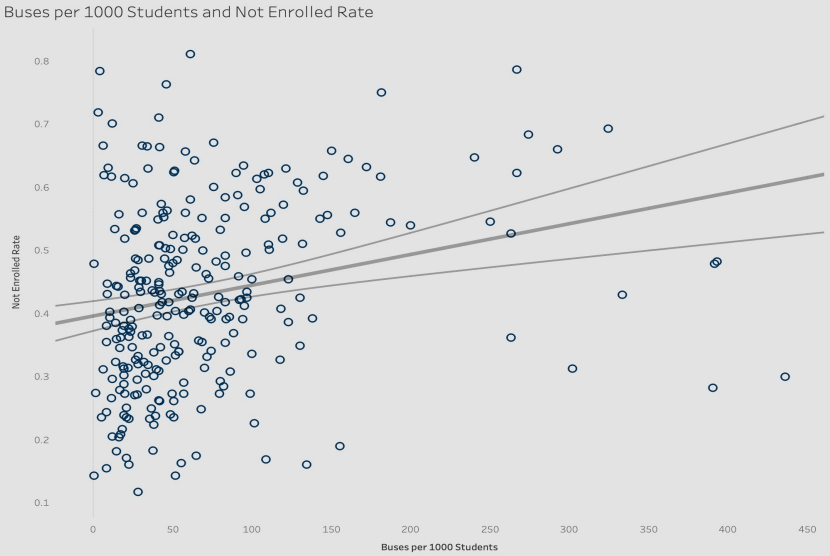
College Going Rate (Increase of 4.41%)

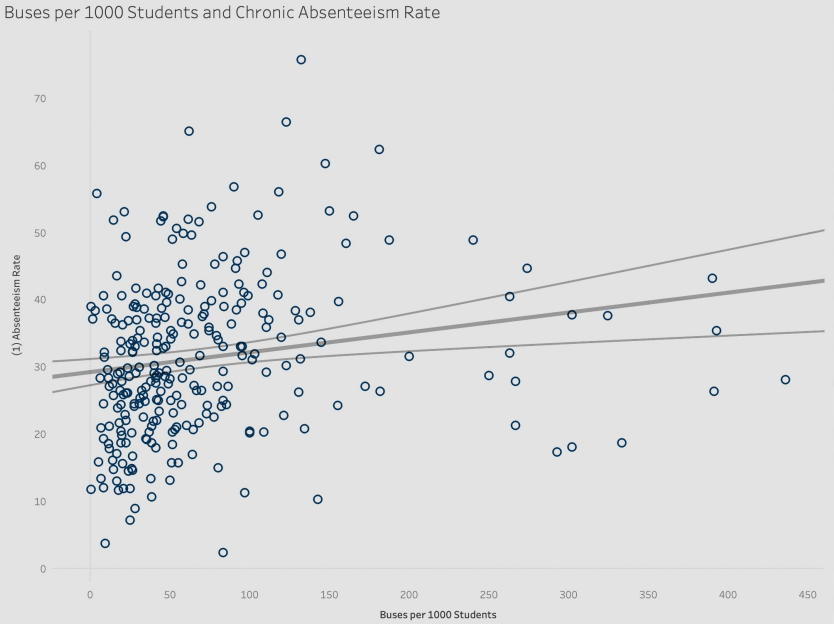
Negative Correlations Cont.



Stille Enrolled Rate (Decrease of 0.01%)

Positive Correlations





Not Enrolled Rate (Increase of 0.04%)

Chronic Absenteeism Rate (Increase of 2.94%)

Limitations

Demographics

No control for income, district funding, demographics, urban vs. rural classification, etc.

Limited Scope

Only used data from 2022, which is likely not representative of all years (i.e., during COVID).

Correlation ≠ Causation

While correlated, bus access is not necessarily the cause, especially due to low R-Squared values.

Failure of Some Assumptions

Not all of the assumptions were satisfied, so results are not 100% reliable.

Buses as a Proxy

Use of buses doesn't account for other forms of public transport, reliability of the system, distances, etc.

Reporting Gaps

Fleet size reported via districts, so sample contains only those who chose to participate.

Conclusions

Key Takeaway:

- School bus fleet size is likely an indicator of more systematic challenges (resource disparity, rurality, access to support, lack of infrastructure, demographics, etc.). The relationships seen are likely a symptom, rather than a root cause.

Next Steps:

- Explore confounding factors:
  - Family/District income
  - Geographic location
  - Resource access