Education Analysis

Jared Cummings and Carly Lundgreen

Problem Statement and Data

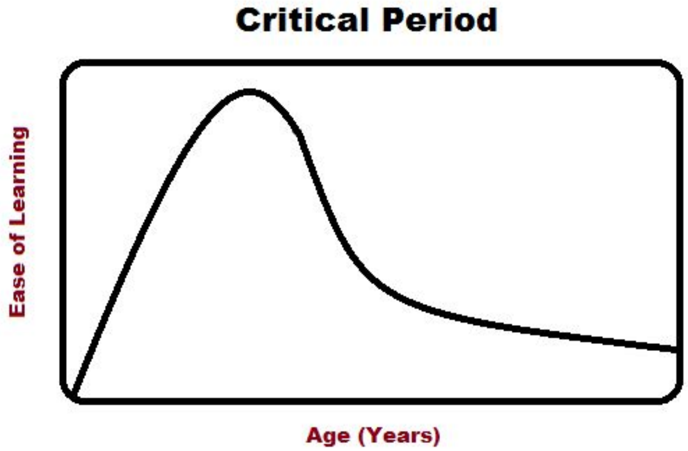
Problem Background

The “Critical Period” Hypothesis:

People learn languages easiest and fastest during early childhood.

A similar effect is found when learning musical instruments!

The brain learns very quickly and efficiently during our elementary school years



Problem Background, continued

- Early elementary education results are a major predictor of later life outcomes because of the brain's neuroplasticity in this stage of life (i.e its ability to LEARN).
- It's clear that early education is important.

Goal: Can we identify any factors that affect learning in elementary school to help inform changes in policy/practice that may improve student learning?

- Note that we measure students' learning by their scores on a standardized test*

*Standardized tests come with their own set of problems, but this is what we have in the data and what schools have used as measures of student learning for years. Alternatives could include performance/portfolio based assessments or increased use of adaptive testing.

Research Questions

1. “Income” is generally a measure of how much money a school has to spend on extracurricular activities (as opposed to expenditures which is how much spent per student in the classroom). Is there evidence of diminishing returns on extracurricular activities in terms of student learning?

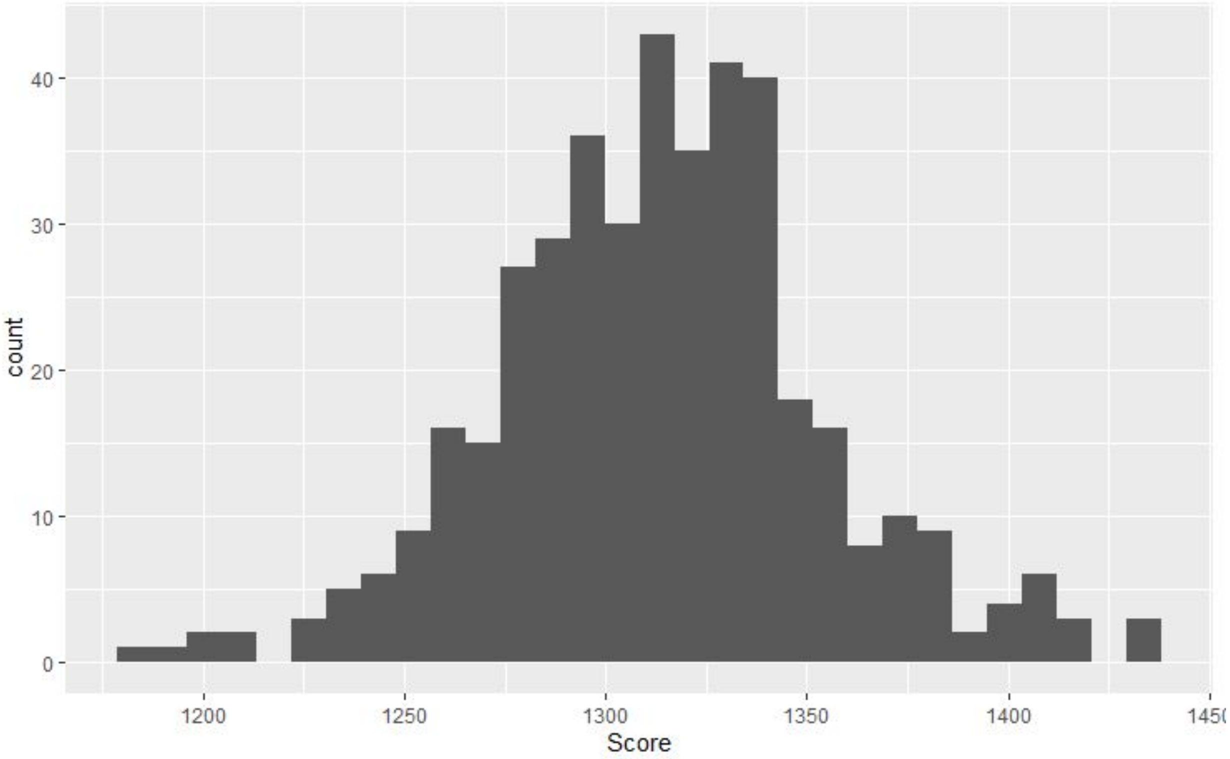
2. Is English as a second language a barrier to student learning?

3. In your opinion and based on the data, what can be done to increase student learning?

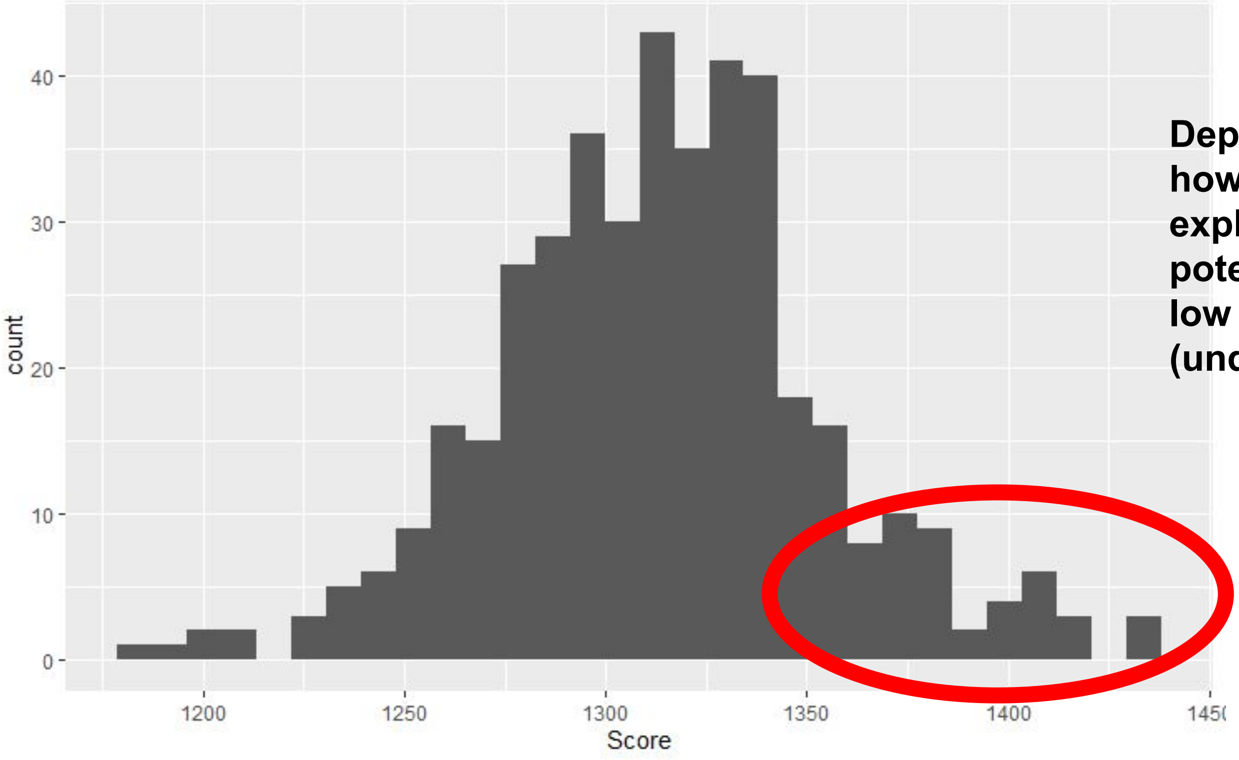
Data: Description

Variable Name	Description
Score	Average cumulative Score on the Stanford 9 standardized test (out of 1600)
Lunch	Percent qualifying for reduced-price lunch
Computer	Number of Computers
Expenditure	Expenditure per student
Income	District average income (in USD 1,000)
English	Percent of English learners
STratio	Student-to-teacher ratio

Data: Distribution of Score

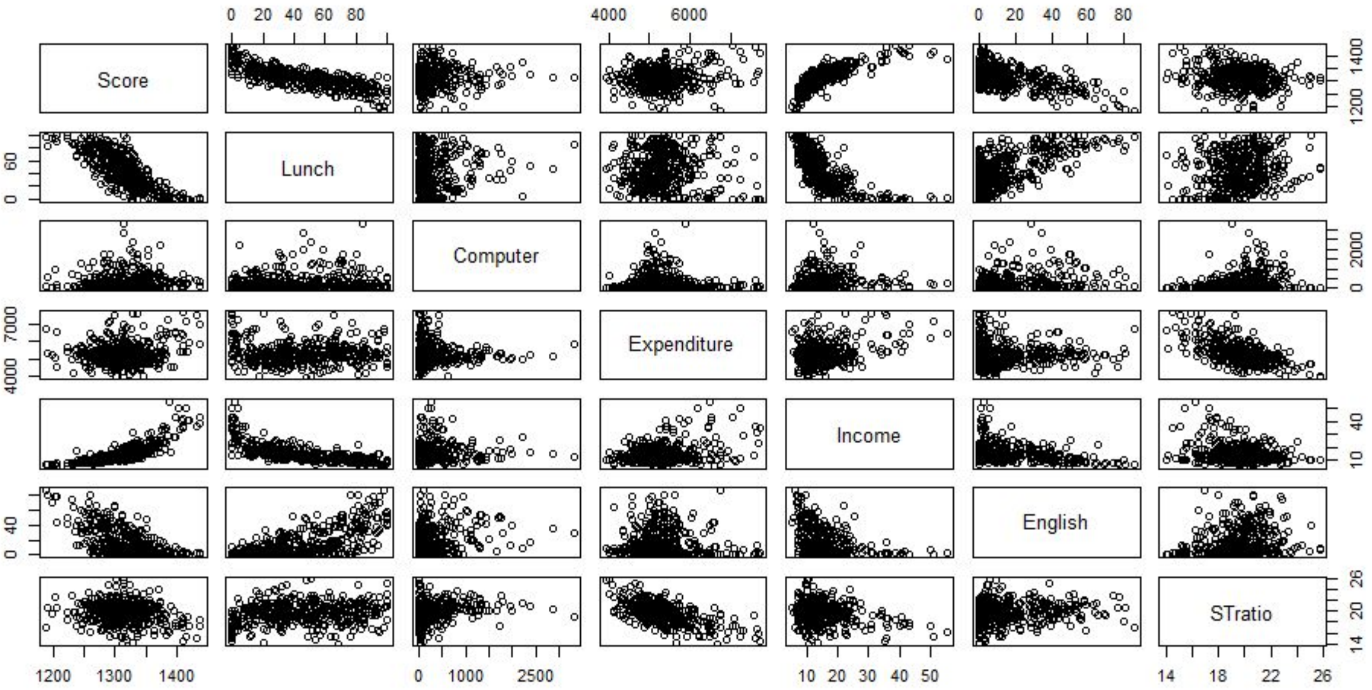


Data: Distribution of Score

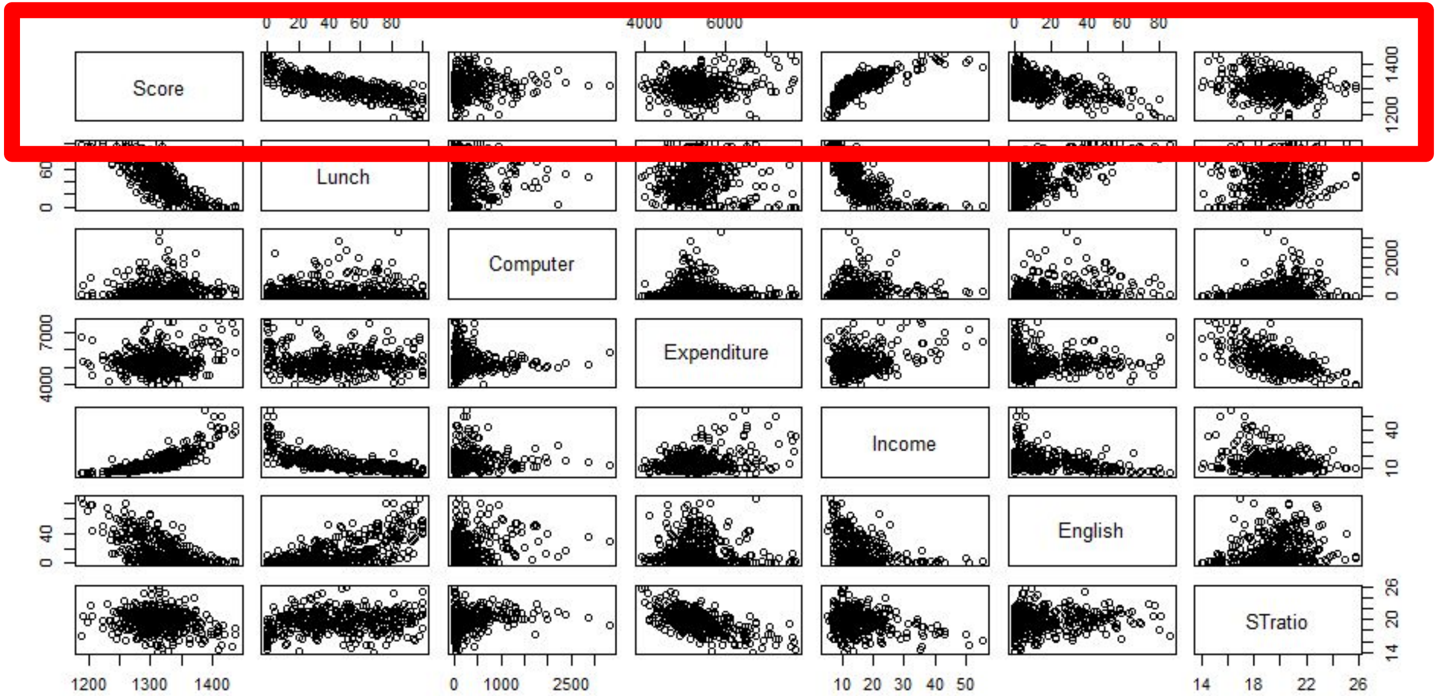


Depending on how variation is explained by data, potential for some low residuals (underprediction)

Data: Pairs plots

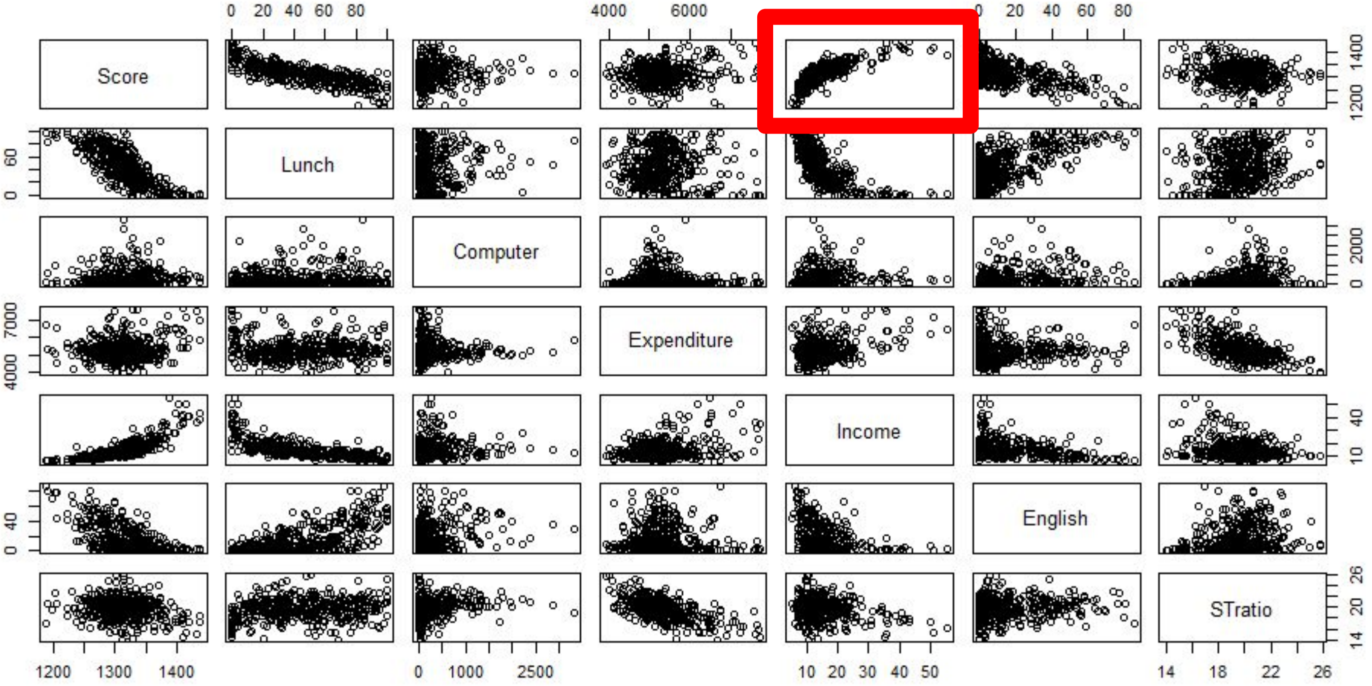


Data: Non-linearity



Data: Decreasing returns

Concavity of non-linearity suggests decreasing returns of extracurriculars to student learning



Data: Spearman Correlations

Data: Improving student learning

Expenditures are strongly related to the number of teachers on the payroll relative to the student body size.

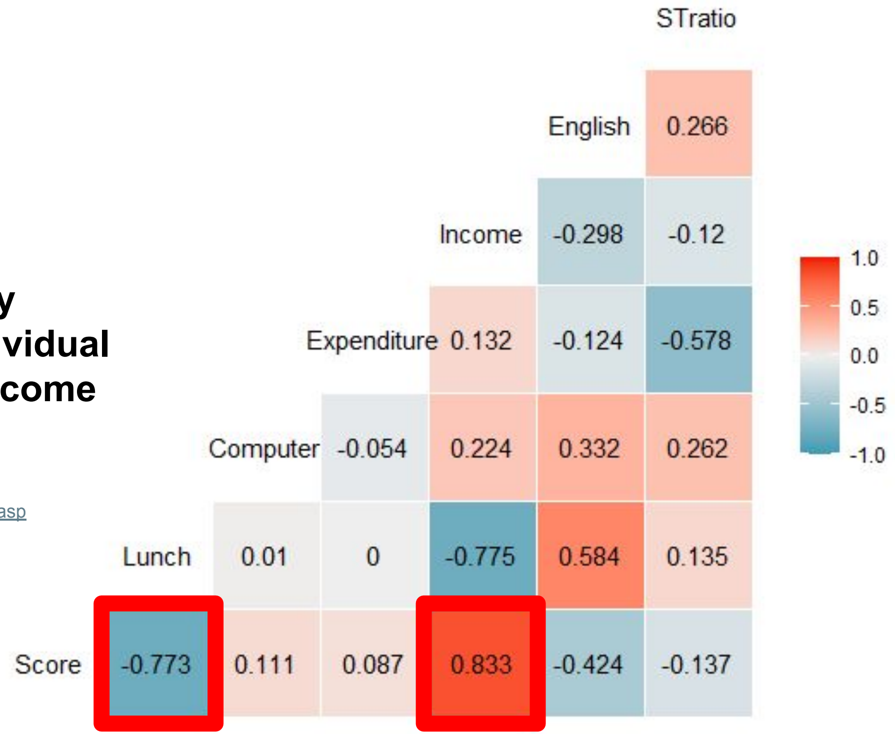
Both have weak effect on score.

Data: Improving student learning

Student learning most strongly associated with wealth metrics.

In 2021, qualifying for reduced-price lunch in California means a family income of ~$120 per individual per week; this adjusts income for family size

Source: https://www.cde.ca.gov/ls/nu/rs/scales2021.asp

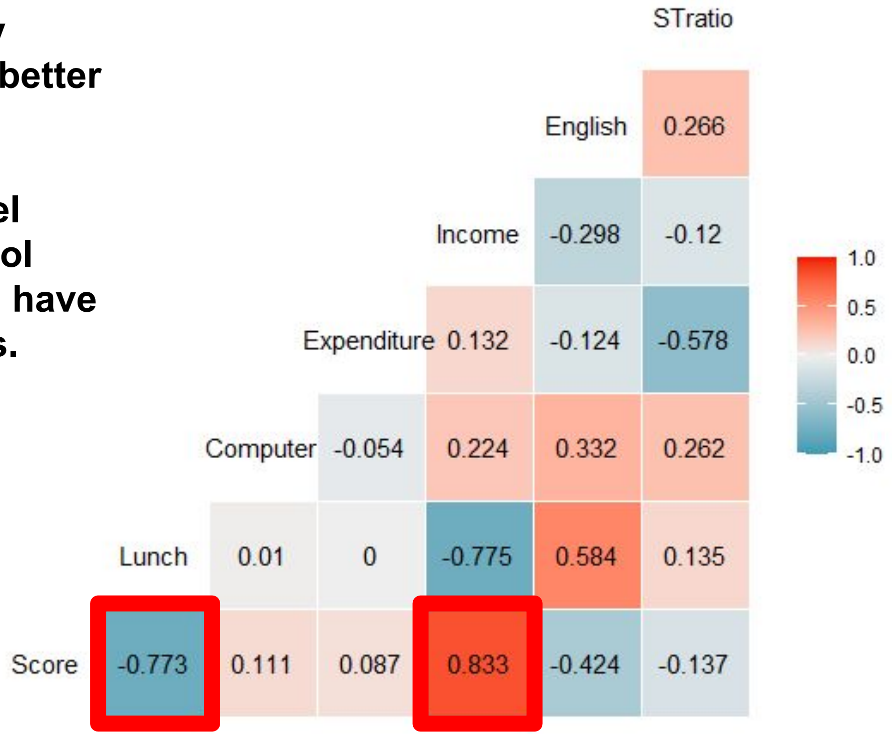


Data: Improving student learning

Fact: Children of wealthy families tend to perform better on standardized tests.

Theory: Children who feel “successful” in the school institution and programs have better learning outcomes.

Source: *Disrupting Class* by Clayton M. Christensen



Data: English barrier

English as a Second Language (ESL) students tend to perform worse.

ESL performance strongly related to family wealth.

Theory: ESL students may have a harder time feeling “successful” in the traditional school environment.

Methods

General Additive Models (GAMs)

Performing nonlinear regression in a multivariate framework is easily accomplished using GAMs, where separate functions are fit for each covariate, depending on the univariate relationship between that variable and the response.

$$y_i = \beta_0 + \sum_{p=1}^{P} f_p(x_{ip}) + \epsilon_i$$

$f_p(x_{ip})$: Function for p^{th} variable

GAMs, continued

The response variable, test score, will depend linearly on smooth functions of our predictors.

Selecting a GAM model includes a method of “function selection”, where models with and without the selected function are compared, via BIC/ ANOVA/ MSE/ etc, and the optimal model between the two is selected.

There is the assumption of linearity in beta (ie the parameters), but not in x. The ‘generalized’ part of GAMs means that we are generalizing the typical MLR scenario to include nonlinear covariates.

GAM Assumptions

Observations independent?

Note for this analysis, we assume that an individual school district's average test score is independent of other school districts.

Homoscedasticity?

We will look at fitted values vs. residual plots.

Normality?

Histogram of residuals

Smoothing Splines

Goal: identify a (nonlinear and smooth) function $g(x_i)$ where the quantity

$\sum_{i=1}^{n}(y_i - g(x_i)^2)$ is small.

The smoothing spline will identify the function that minimizes the penalized residual sum of squares:

$$\underbrace{\sum_{i=1}^{n}(y_i - g(x_i))^2}_{\text{Loss}} + \underbrace{\lambda \int g''(x)^2 dx}_{\text{Smoothness Penalty}}$$

Where the second derivative of the function is penalized so that the function is not too 'wiggly' (ie overfit to the data). We don’t want the function to be overly erratic.

Smoothing splines, continued

The smoothing parameter, λ, is chosen via cross validation. λ will decide the extent to which the second derivative is penalized. (As λ increases, we are 'smoothing' away the 'wigglier' basis functions)

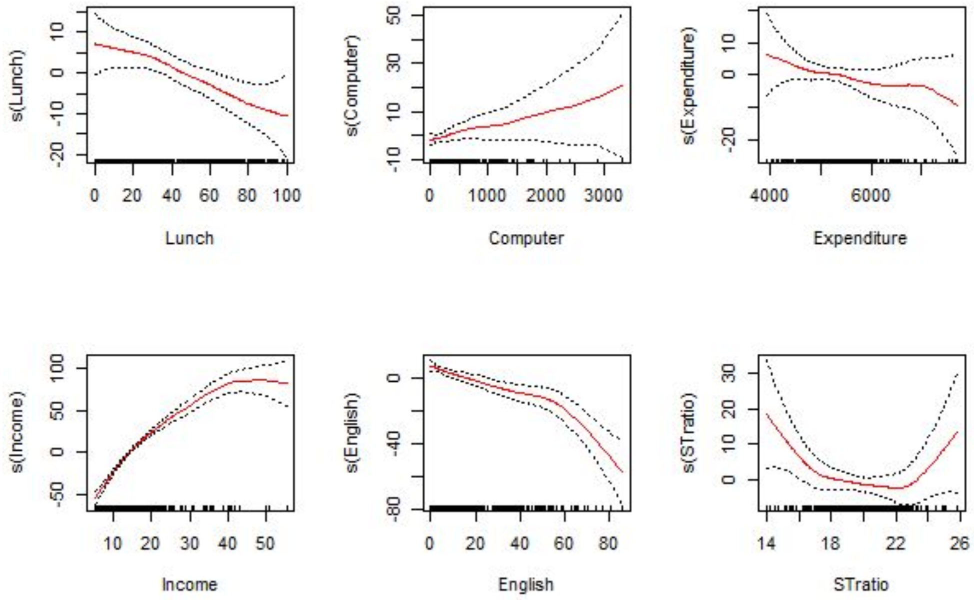
PROS: Don't need to select knot points because each individual x value that is input into the function is used as a 'knot'

CONS: You don't get nice, interpretable coefficients; you also have another parameter that has to be estimated (λ).

Initial "comparison" model (using splines)

Smoothing splines on all predictors:

$$y_i = \beta_0 + \sum_{p=1}^{P} s(x_{ip}) + \epsilon_i$$



We can see definite nonlinearity in the Income, English, and STratio variables. The other variables appear to be linear.

We decided with model comparisons via ANOVA (actually Likelihood Ratio Test -- comparison of "full and "null" models)

Function Selection Process and Chosen Model

We compared the 'full' model from the previous slide with models that:

i) excluded the variable in question entirely

ii) included the variable linearly (ie without a smoothing spline function applied)

If p-value < 0.05 (our chosen significance level), we reject the reduced/'null' model in favor of the full model.

This process was repeated for all variables in the dataset.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \sum_{i=3}^{4} s(x_i) + \epsilon_i$$

Where x1 corresponds to 'Lunch' and x2 corresponds to 'Computer', and x3 and x4 correspond to Income and English, respectively.

B0 → expected average cumulative score when all covariates are zero (ie 0% English speakers, 0 computers, etc.)

β1 → for every 1% increase in students who qualify for reduced lunch, we expect a 1-unit increase in average cumulative score on the Stanford 9 standardized test. (holding all else constant)

β2 → for every additional computer in a California school district, we expect a 1-unit increase in average cumulative score on the Stanford 9 standardized test. (holding all else constant)

Local Regression

- Model flexibility controlled by span, s, the proportion of observations used in prediction
    - Higher $s \Rightarrow$ less flexible, “smoother”, global fit
    - Lower $s \Rightarrow$ more flexible, “wiggly”, local fit
    - selected via cross-validation
- Algorithm Outline:
    - Select the $100s$% of observations closest to prediction point, x_0
    - Weight the observations via weight function $K_{i0} = K(x_i, x_0)$; closer observations have higher weight
    - Obtain estimate of coefficients with weighted least squares
    - Generate prediction for x_0 with estimated coefficients

Local Regression

$$\min_{\hat{\beta}_0(x_0),\hat{\beta}_1(x_0)} \sum_{i=1}^{n} K_\delta(x_0, x_i)(y_i - \hat{\beta}_0(x_0) - \hat{\beta}_1(x_0)x_i)^2$$

- For our model, we used the tri-cube weight function

  $$K_{i0} = (1 - |d|^3)^3$$

  where *d* is the scaled (0 to 1) distance from the point x_i from the fitted line

- Span allowed to vary for different variables (*varying coefficients model*)
- Computational concessions:
  - Considered only linear local regression
  - Span selected by hold-out validation (rather than k-fold, LOO)

Model Justification

Model comparison

Model	**RMSE**	**BIC**	**R^2**
Local Regression 1	14.301	3773.21	0.876
Local Regression 2	1.02	3289.417	0.999
Smoothing Spline	17.944	3709.397	0.804

Local Regression comparison

Variable Name	Description	Span Model 1	Span Model 2
Score	Average cumulative Score on the Stanford 9 standardized test (out of 1600)	—	—
Lunch	Percent qualifying for reduced-price lunch	0.3875	0.1131579
Computer	Number of Computers	0.05	0.01
Expenditure	Expenditure per student	0.5	0.06157895
Income	District average income (in USD 1,000)	0.10625	0.01
English	Percent of English learners	0.05	0.01
STratio	Student-to-teacher ratio	0.05	0.01

Local Regression variable selection

- Included all variables' parametric effects
  - Because of significant contribution to model fit (all but *STratio*)
  - *STratio* kept for contribution to interpretation
- Included all variables' non-parametric effects
  - Prediction not a task of analysis; risk of overfitting "worth it"

Variable	**ANOVA p-value for parametric effect**
Lunch	< 0.0001
Computer	~ 0.0001
Expenditure	< 0.0001
Income	< 0.0001
English	< 0.0001
STratio	0.5564

Smoothing Splines variable selection

Variable	ANOVA p-value for smoothing spline vs. linear effect	ANOVA p-value for smoothing spline vs. excluding variable	
Lunch	0.8187	0.0075	← Lunch should be kept with a linear effect
Computer	0.8786	0.0217	← Computer should be kept with a linear effect
Expenditure	0.5929	0.1265	← A model without Expenditure performs better
Income	6.31e-11	< 2.2e-16	← Income should be kept w/ a smoothing spline
English	0.01675	2.21e-09	← English should be kept w/ a smoothing spline
STratio	0.0165	0.2399	← A model without STratio performs better

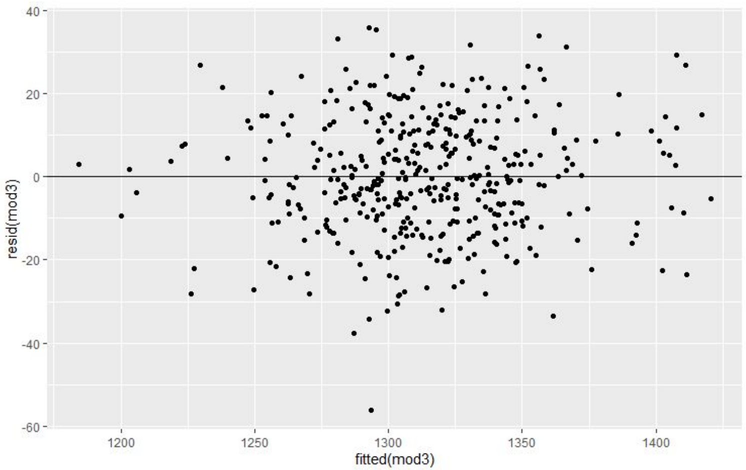
Smoothing Parameter Estimates

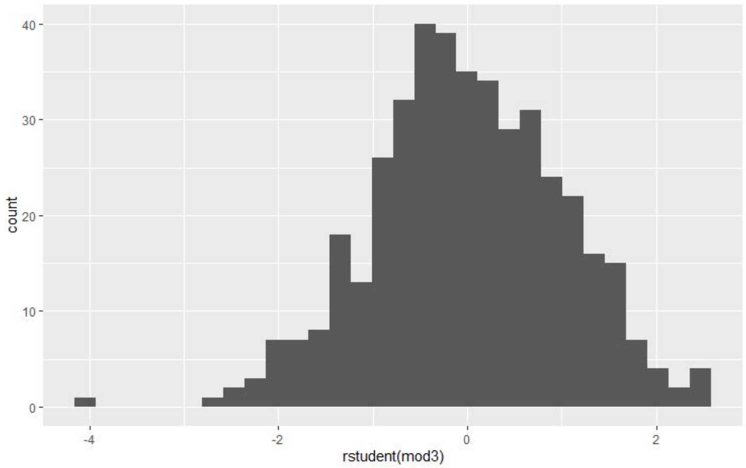
Income	English
0.00512	0.30363

Income is penalizing “wiggliness” much less than English. The smoothing parameter estimates are small which means we are definitely supposed to have nonlinear functions on these variables.

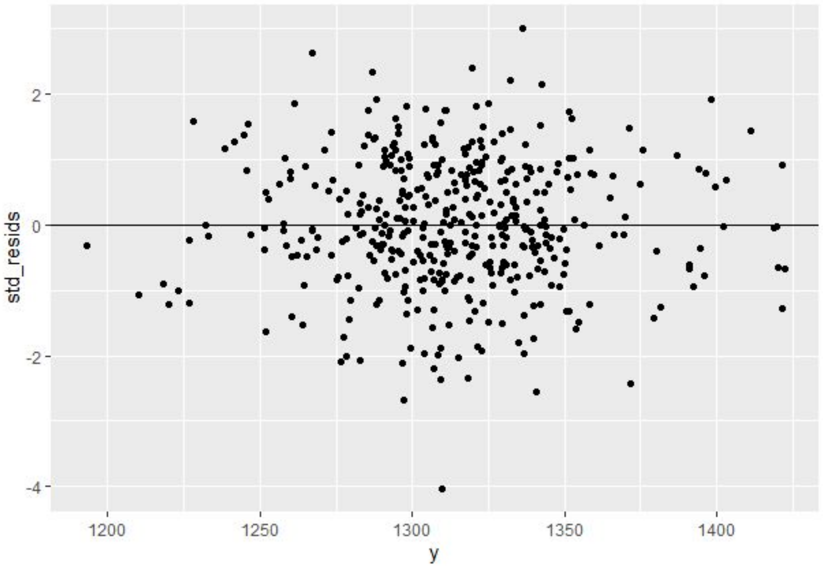
Local Regression Assumptions

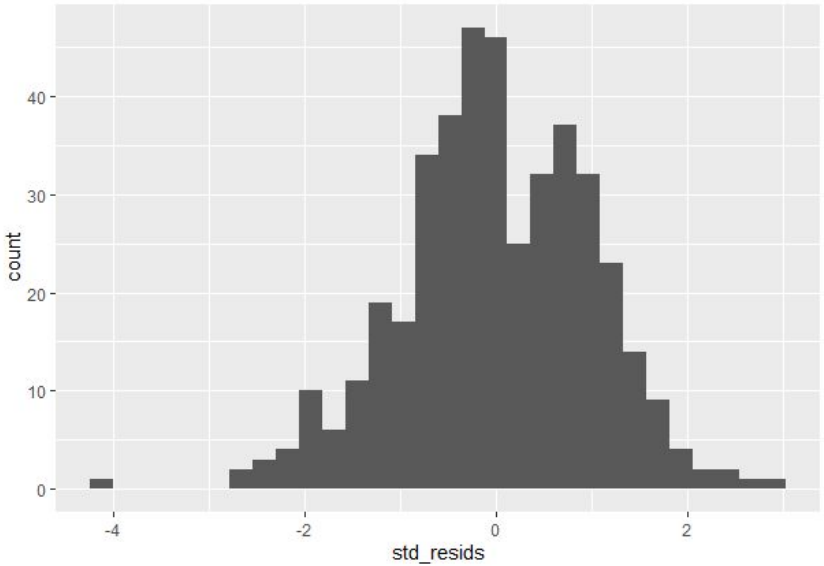
- Data is “dense” enough for valid inference
  - Only 1-D neighbors considered
- Independent, identically distributed errors
  - Tendency to underpredict
  - Relatively symmetric
- Additive effects
  - Essential assumption for methods
  - No way to validate without understanding process



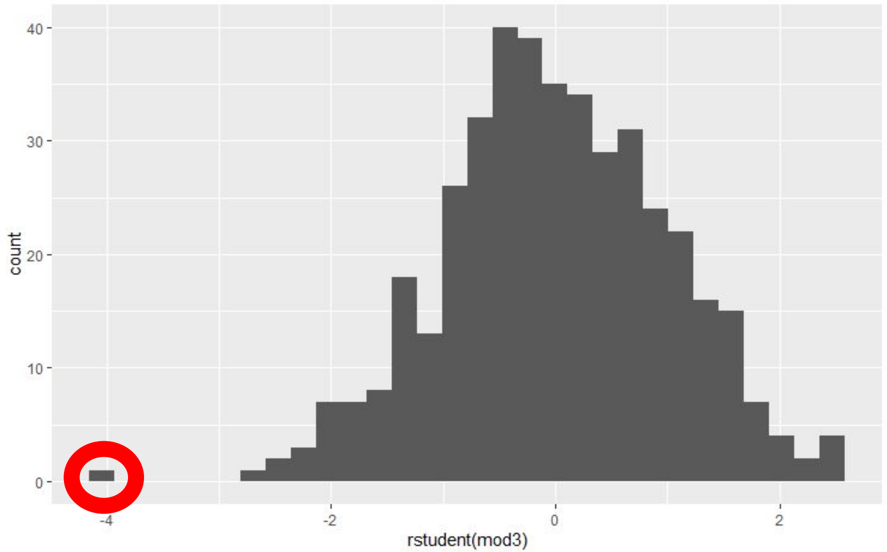


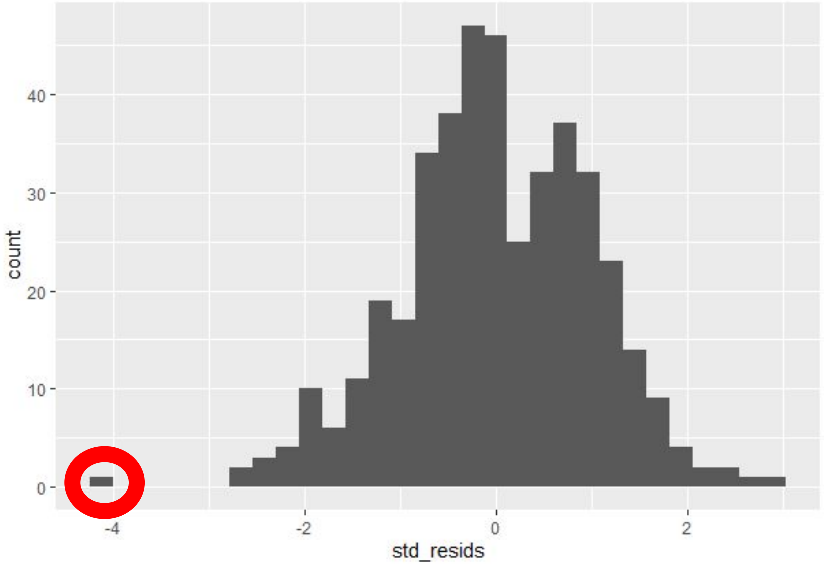
Smoothing Spline Model Assumptions





Residual comparison

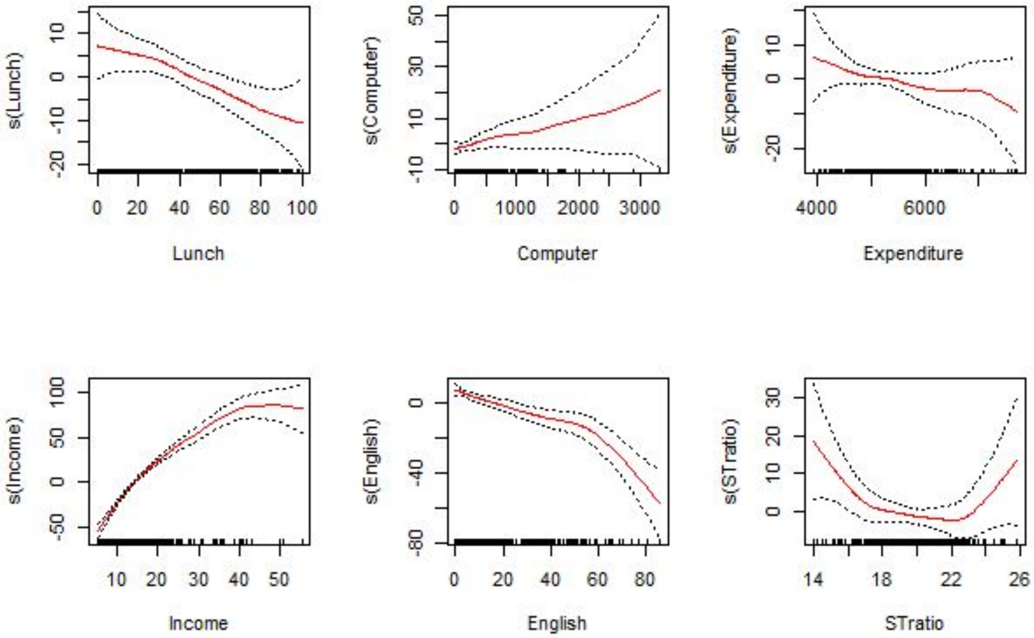




Score	Lunch	Computer	Expenditure	Income	English	STratio
1237.29	67.23	51	5621.69	16.62	39.57	17.15

Smoothing Splines Model

The ANOVA results reflect what we expected from plots: linear functions are more appropriate for Lunch, Computer, and Expenditure, and nonlinear functions are more appropriate for the other variables.



Results

We chose the Local Regression Method

In comparison of RMSEs and R^2, it performed much better. Though the goal is not prediction, if we were going to predict, we would be more accurate in our predictions with local regression.

The higher R^2 implies that the GAM with local regression functions fits the data better than the GAM using smoothing splines.

Questions

1. “Income” is generally a measure of how much money a school has to spend on extracurricular activities (as opposed to expenditures which is how much spent per student in the class room). Is there evidence of diminishing returns on extracurricular activities in terms of student learning?

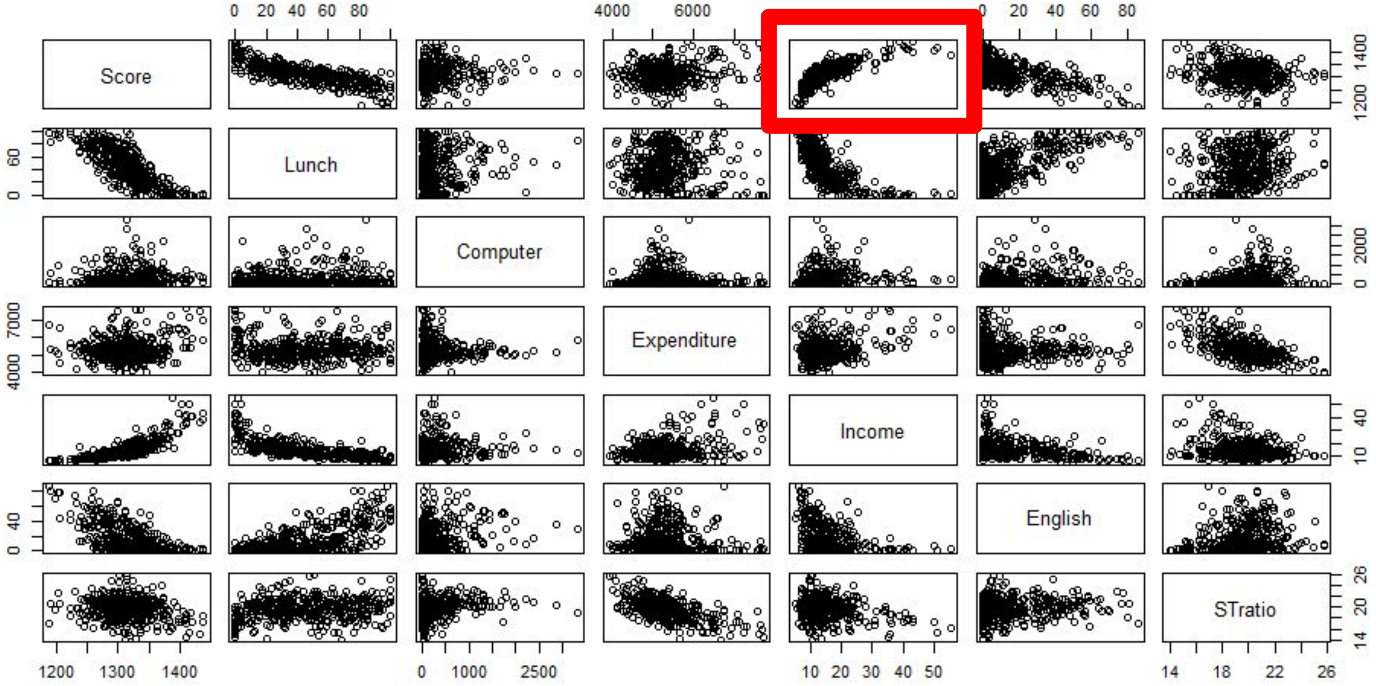
2. Is English as a second language a barrier to student learning?

3. In your opinion and based on the data, what can be done to increase student learning?

Question 1

Data: Decreasing returns

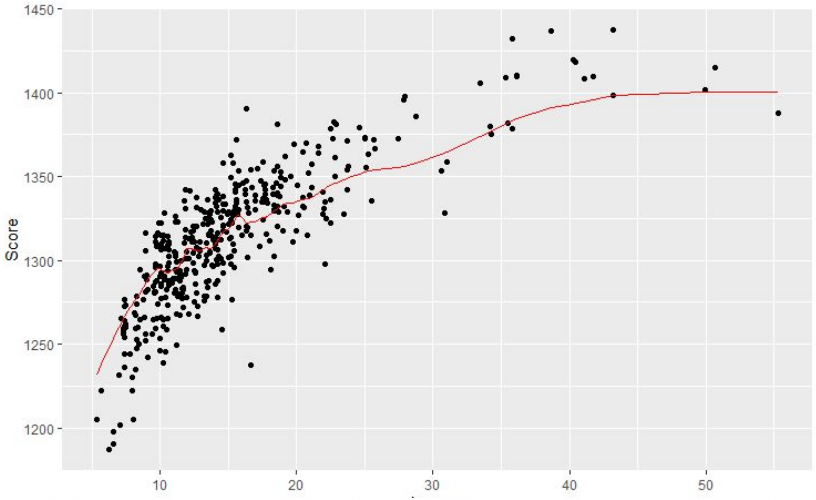
Concavity of non-linearity suggests decreasing returns of extracurriculars to student learning

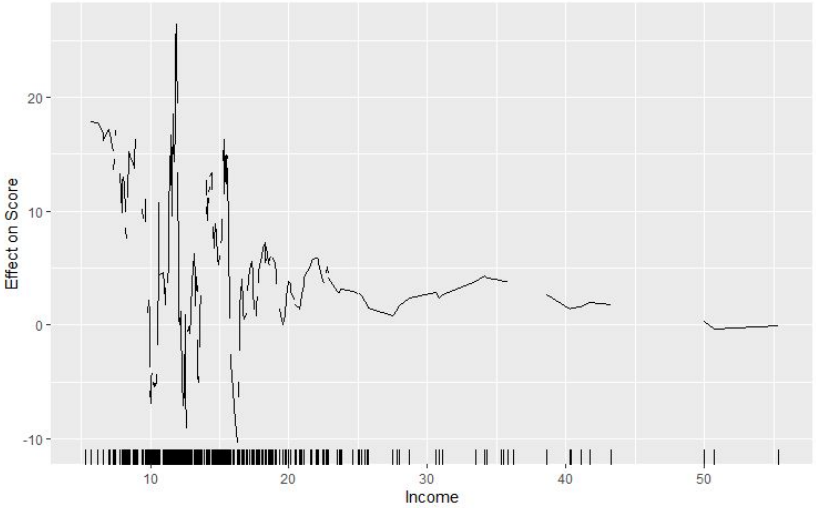


Question 1

Results: Decreasing returns

- Significant Parametric effect *(p-value < 0.0001)*
- Significant Nonparametric effect *(p-value < 0.0001)*

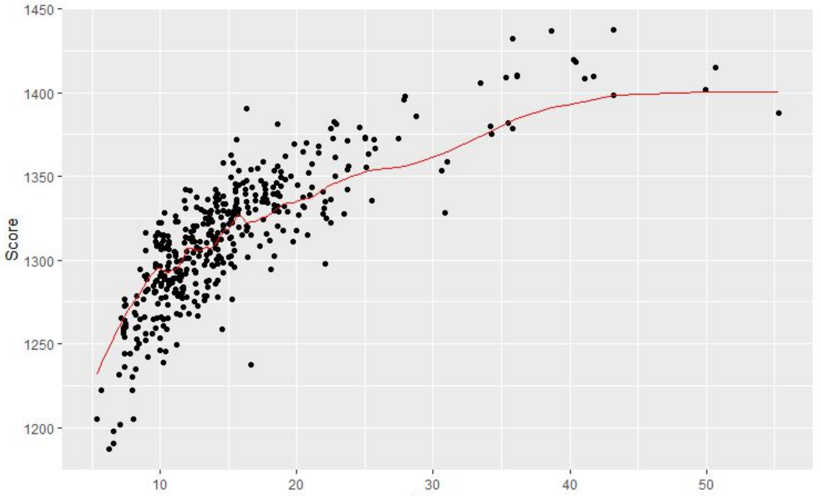


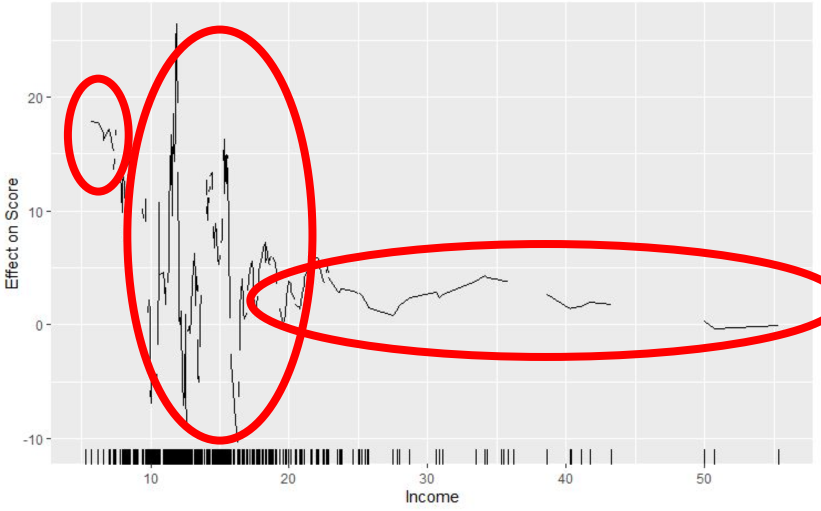


Question 1

Results: Decreasing returns

- Very low-income = higher return
- For the “average” school, effects are mixed but decreasing
- Beyond a certain point, fairly flat





Data: English barrier

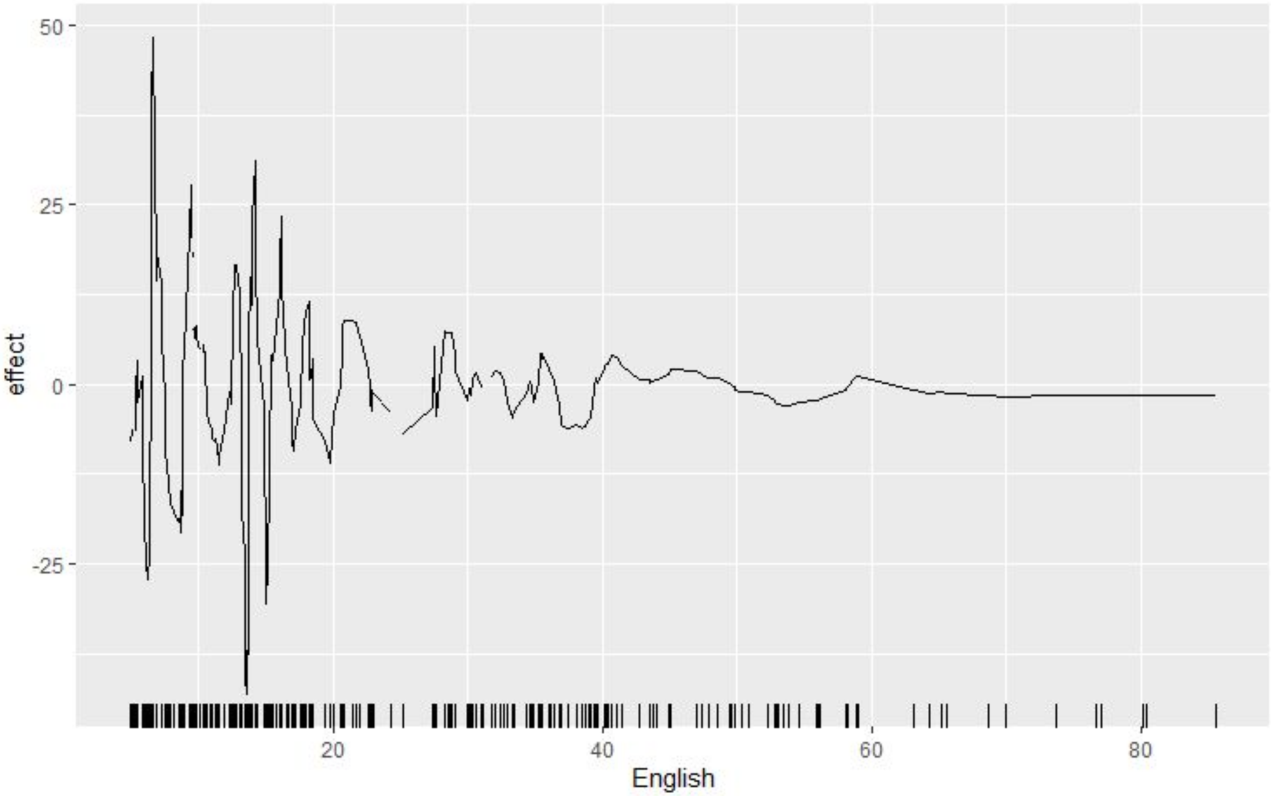
English as a Second Language (ESL) students tend to perform worse.

ESL performance strongly related to family wealth.

Theory: ESL students may have a harder time feeling “successful” in the traditional school environment.

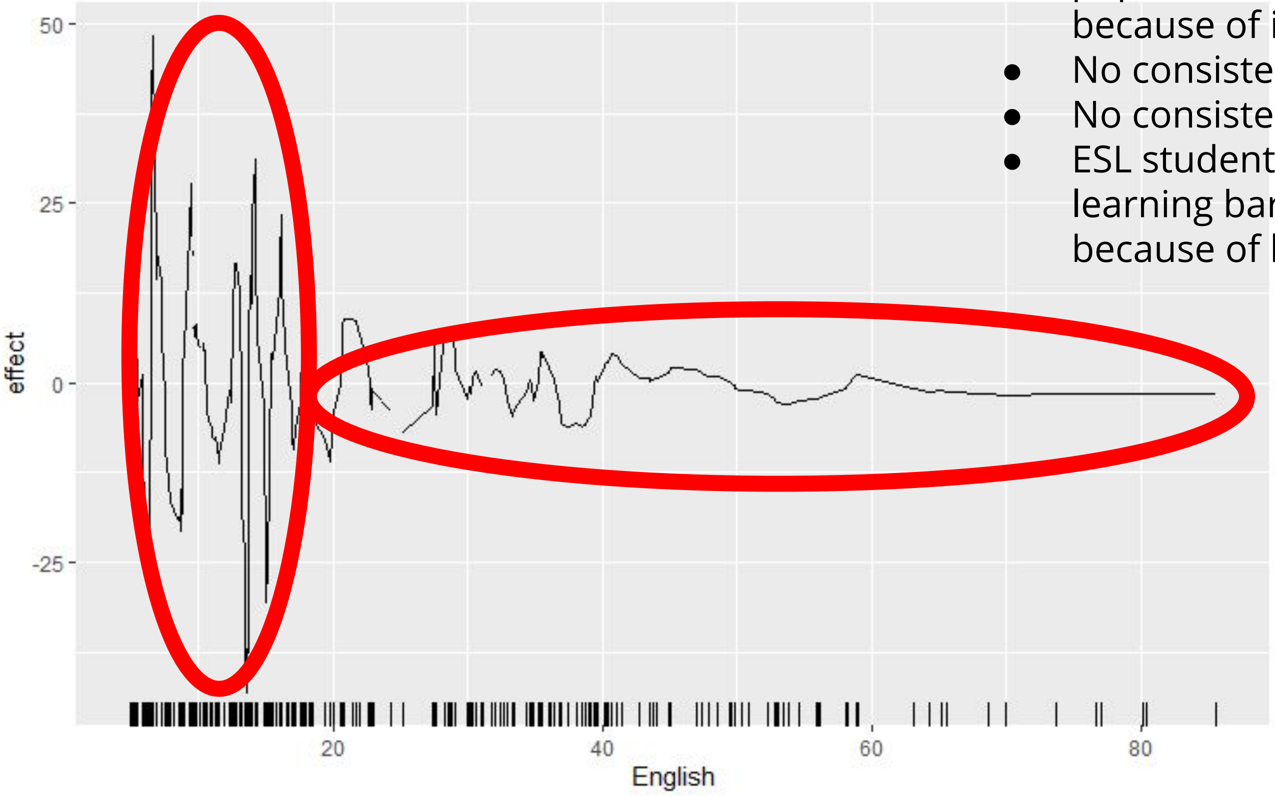
Question 2

Results: English barrier



Question 2

Results: English barrier



Holding everything else constant:

- Schools with a small ESL population can struggle because of it
- No consistent non-zero effect
- No consistent trend in effect
- ESL students experience a learning barrier, but not because of learning English

Question 3

Data: Improving student learning

Expenditures are strongly related to the number of teachers on the payroll relative to the student body size.

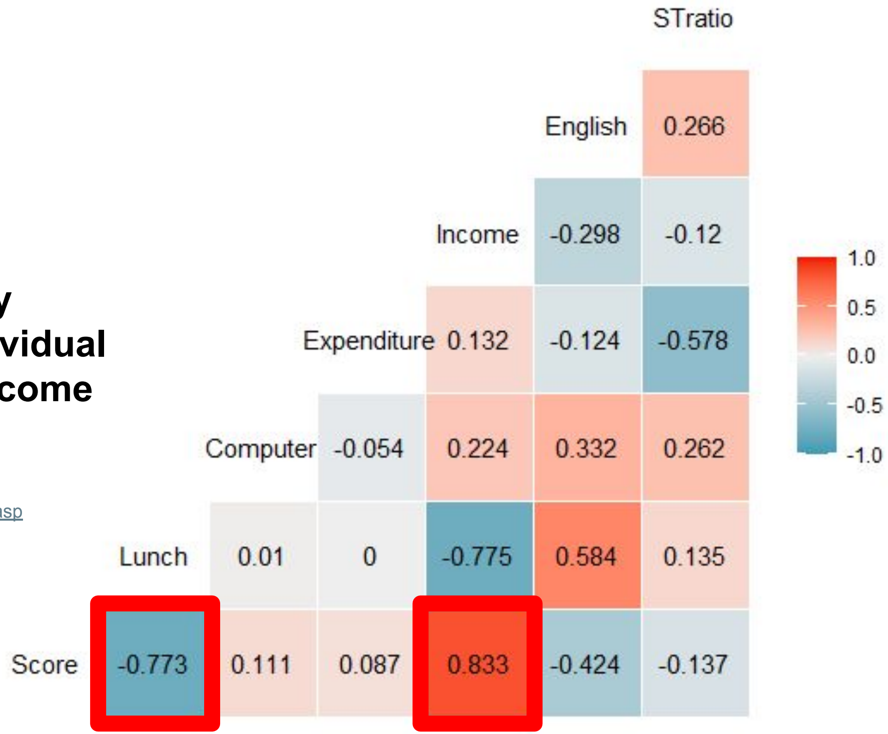
Both have weak effect on score.

Data: Improving student learning

Student learning most strongly associated with wealth metrics.

In 2021, qualifying for reduced-price lunch in California means a family income of ~$120 per individual per week; this adjusts income for family size

Source: https://www.cde.ca.gov/ls/nu/rs/scales2021.asp



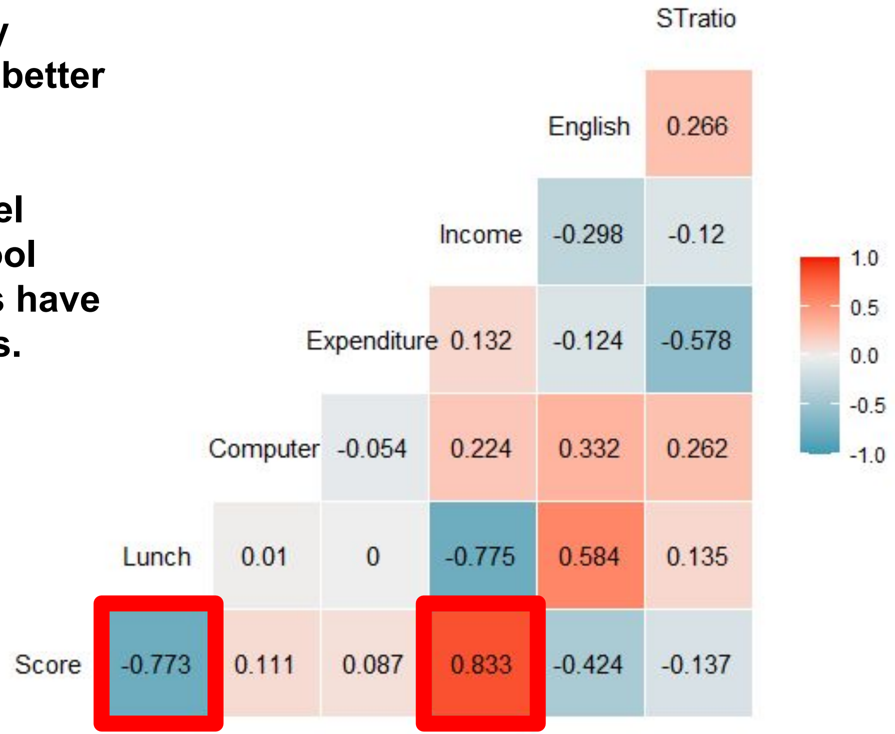
Question 3

Data: Improving student learning

Fact: Children of wealthy families tend to perform better on standardized tests.

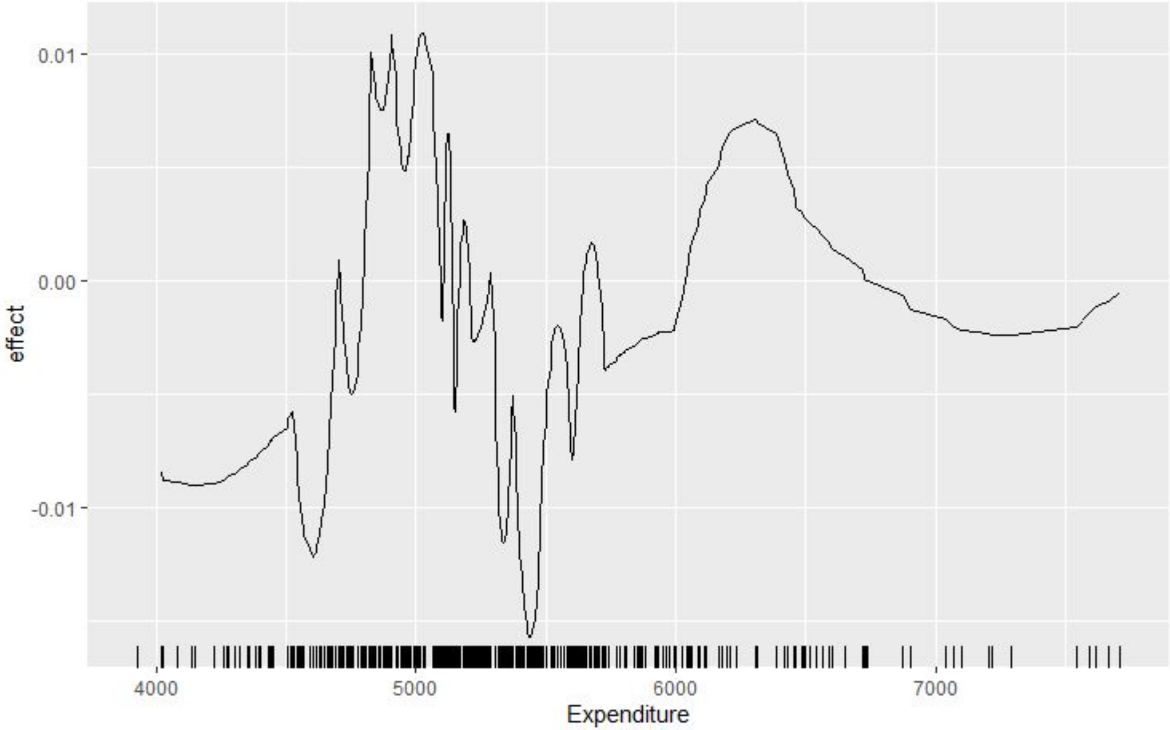
Theory: Children who feel “successful” in the school institution and programs have better learning outcomes.

Source: *Disrupting Class* by Clayton M. Christensen



Question 3

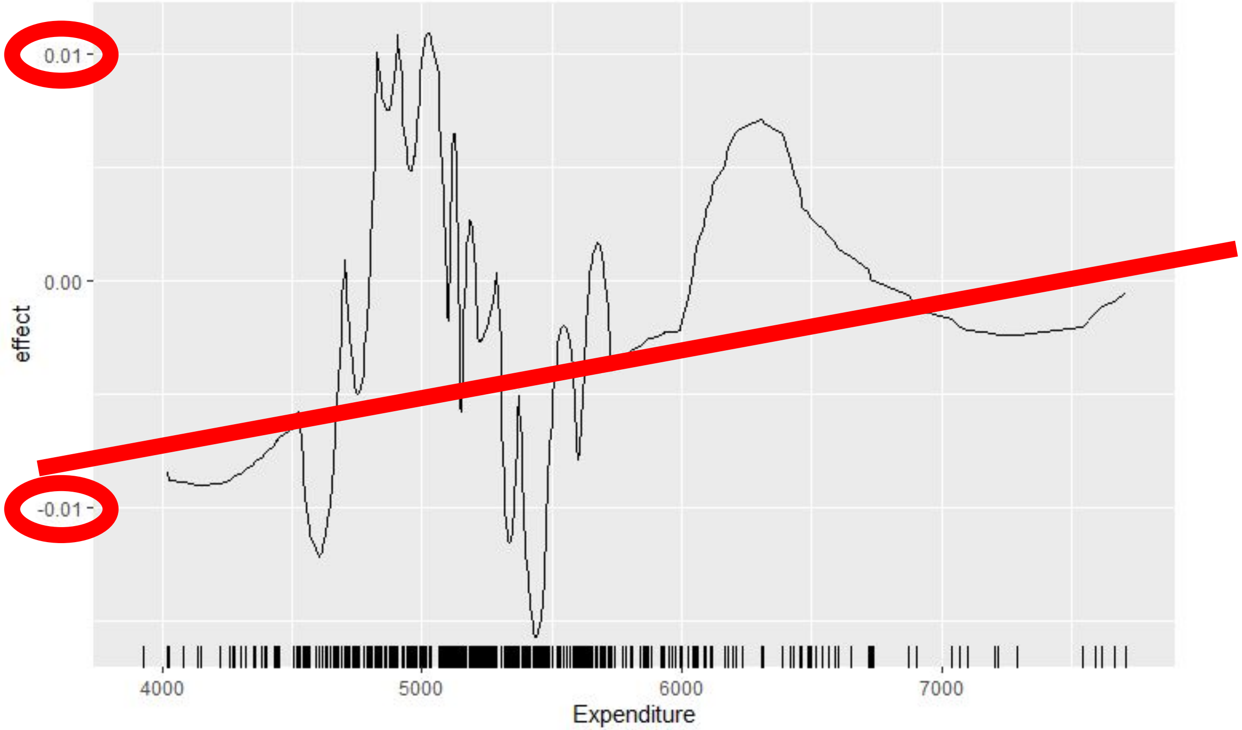
Results: Improving student learning



Span: 0.5

Question 3

Results: Improving student learning

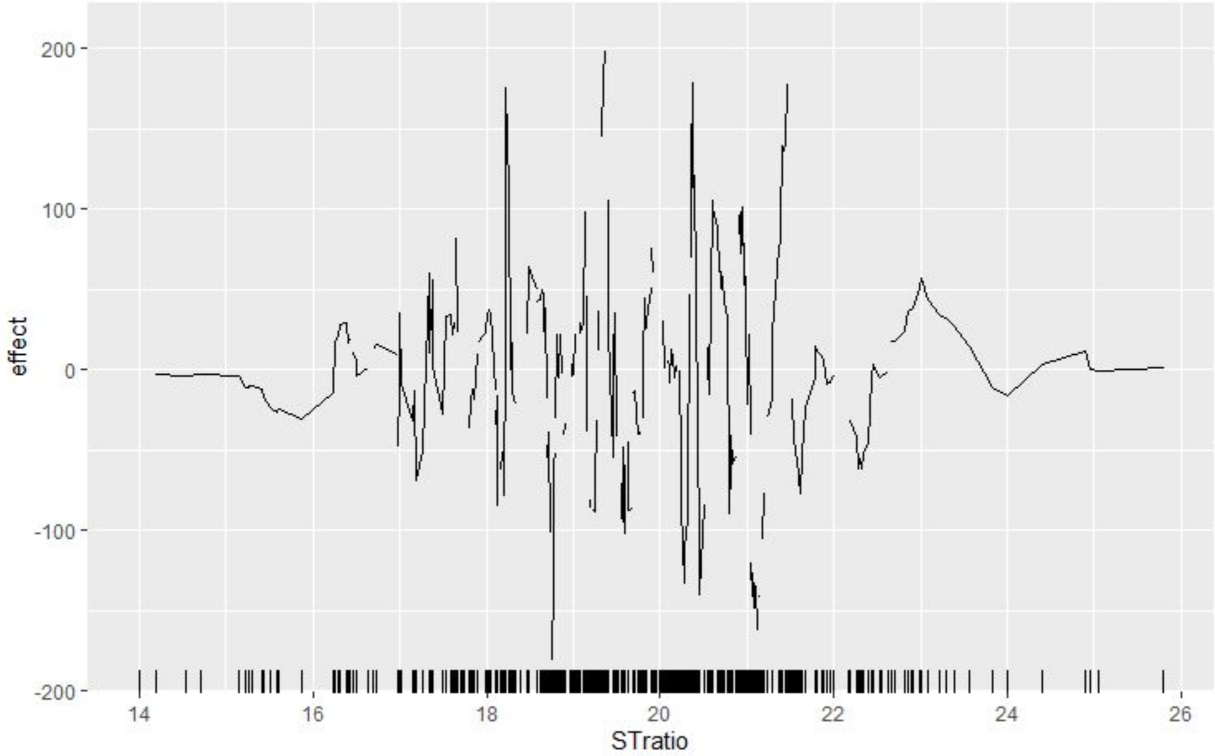


Span: 0.5

More spending per student does increase score…but not really

Question 3

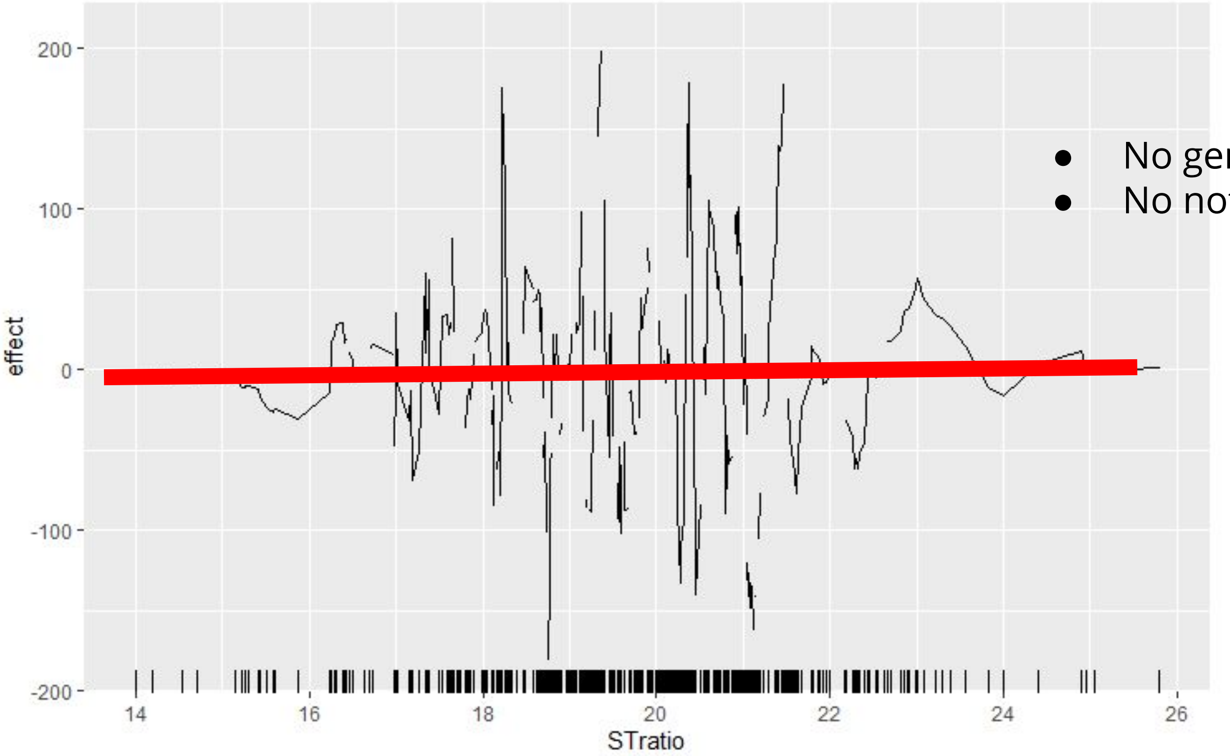
Results: Improving student learning



Span: 0.05

Question 3

Results: Improving student learning

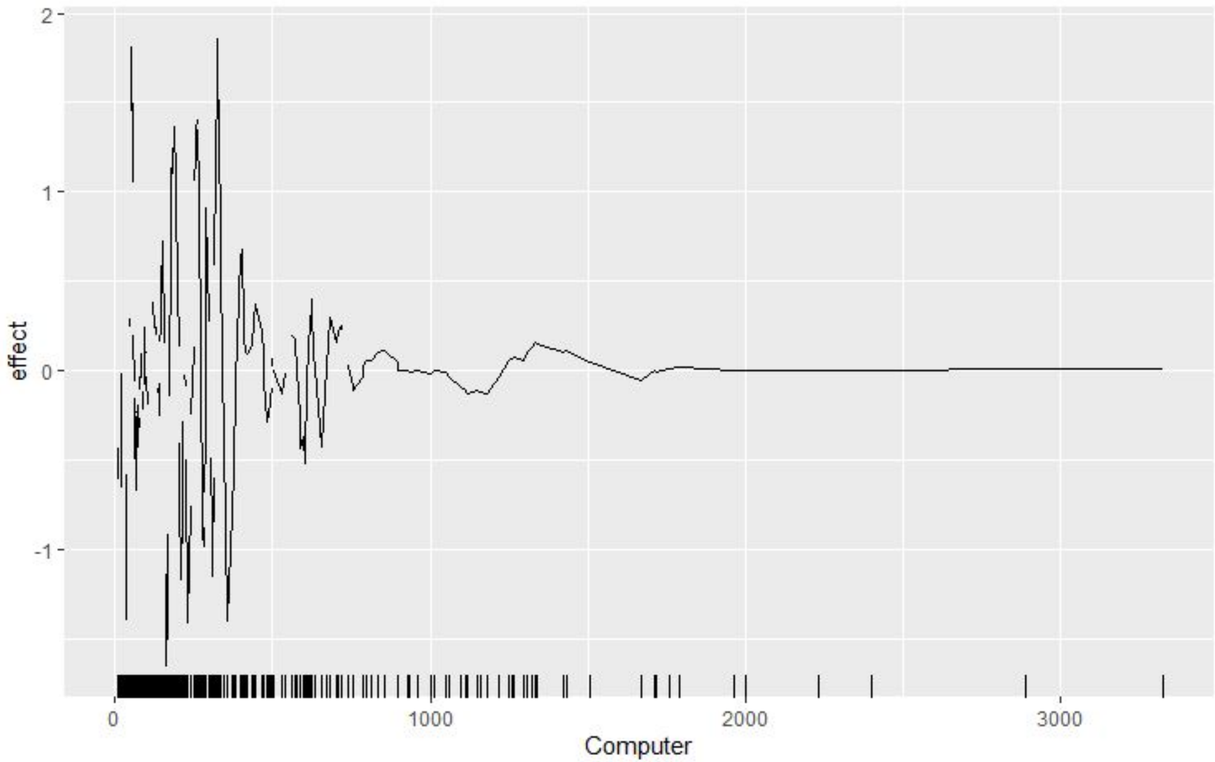


- No generalizable “rule”
- No noticeable effect

Span: 0.05

Question 3

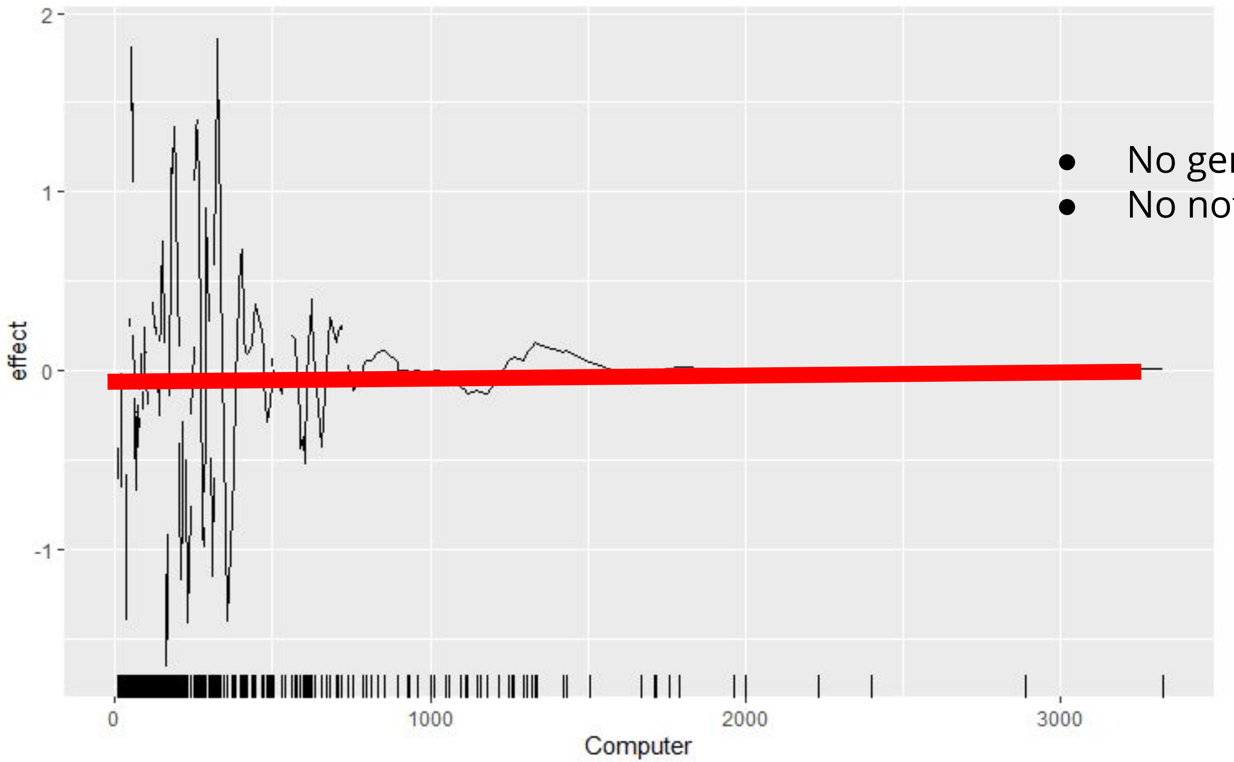
Results: Improving student learning



Span: 0.05

Question 3

Results: Improving student learning

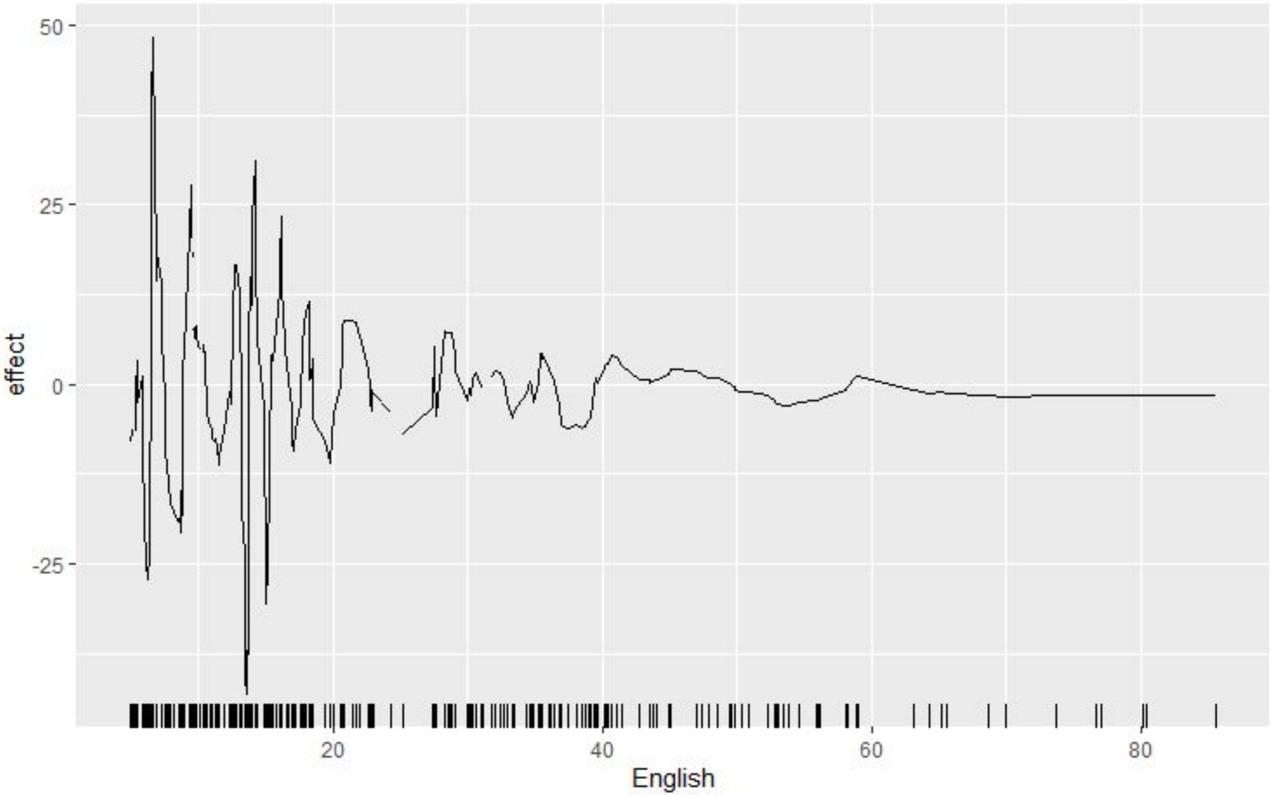


- No generalizable “rule”
- No noticeable effect

Span: 0.05

Question 3

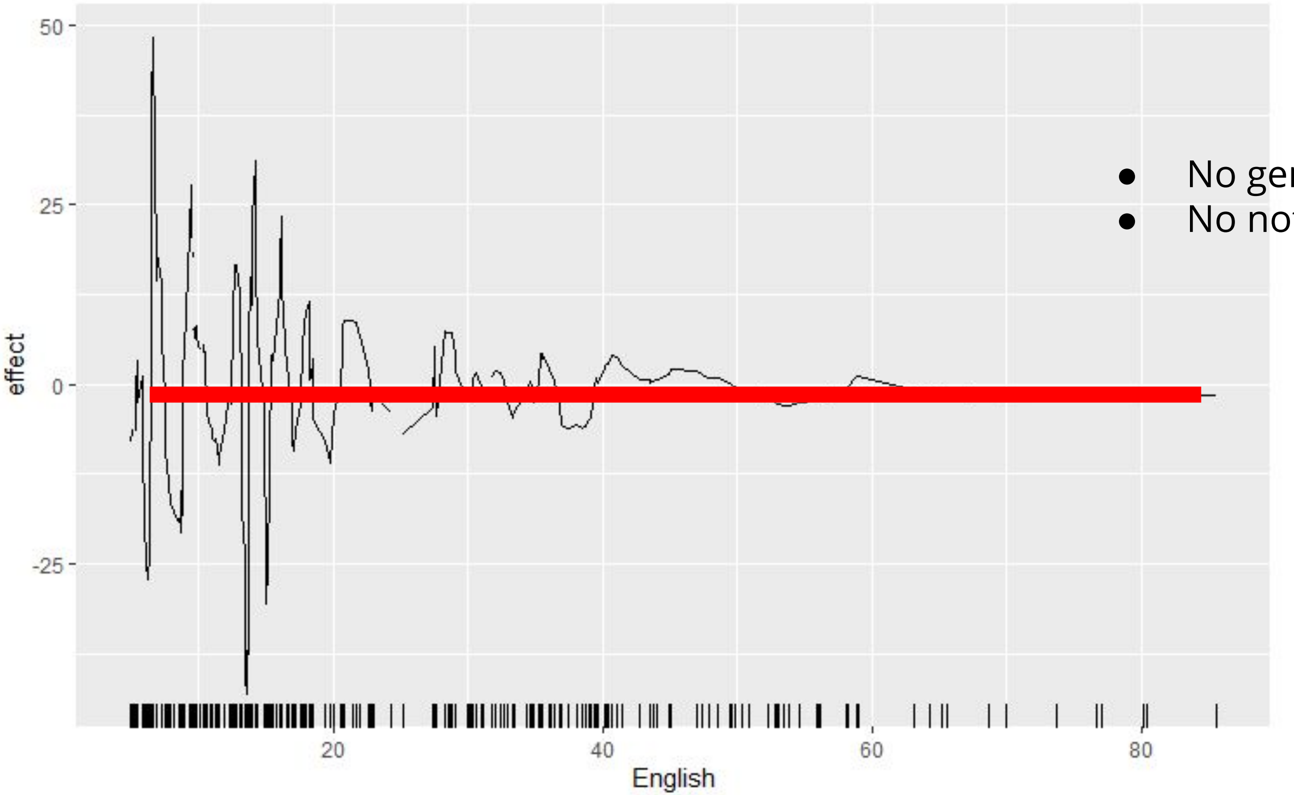
Results: Improving student learning



Span: 0.05

Question 3

Results: Improving student learning

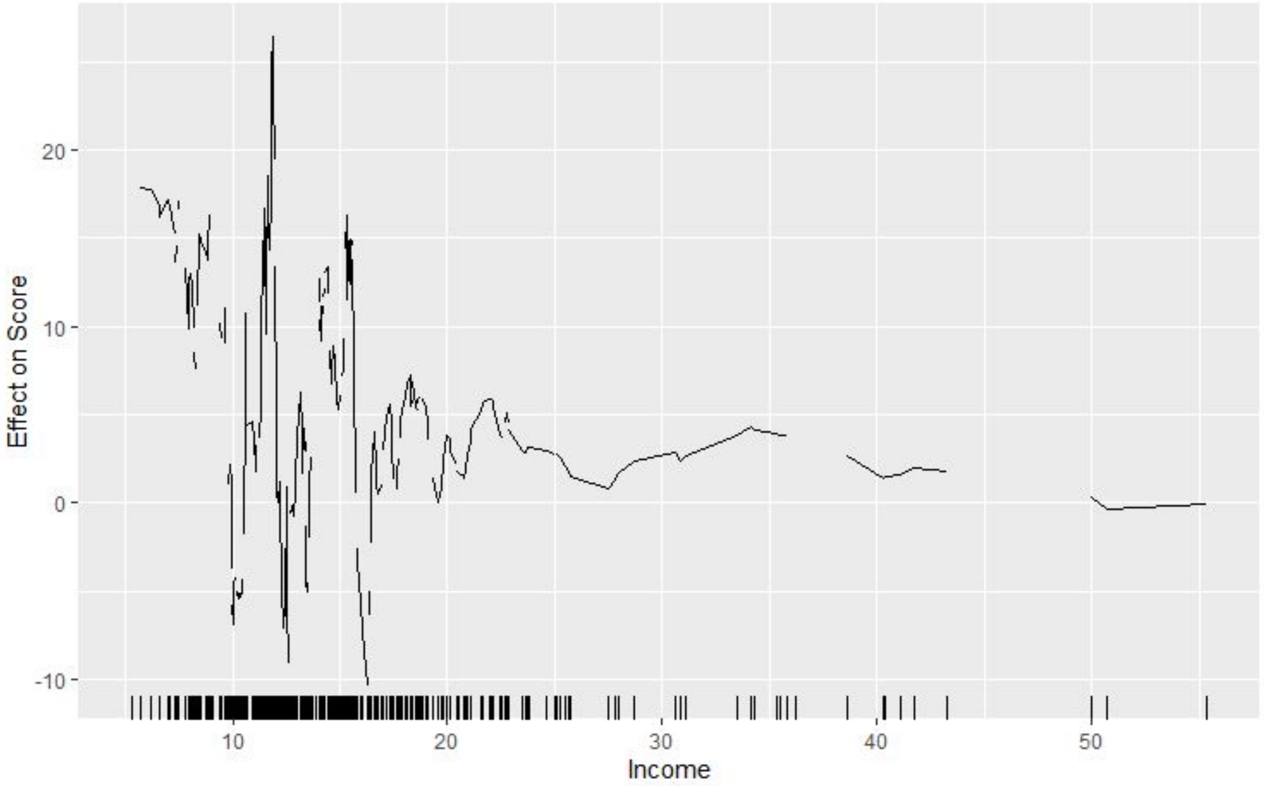


- No generalizable “rule”
- No noticeable effect

Span: 0.05

Question 3

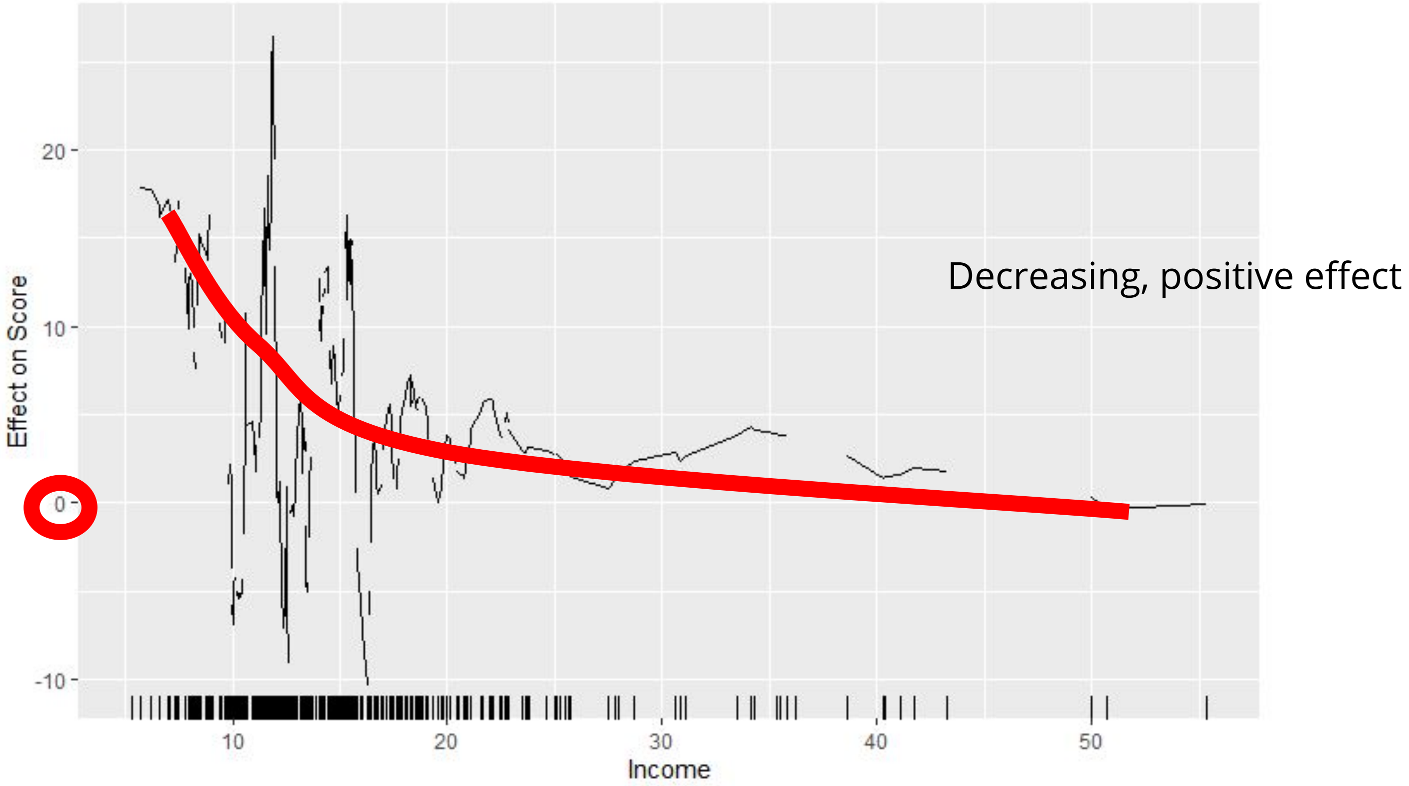
Results: Improving student learning



Span: 0.10625

Question 3

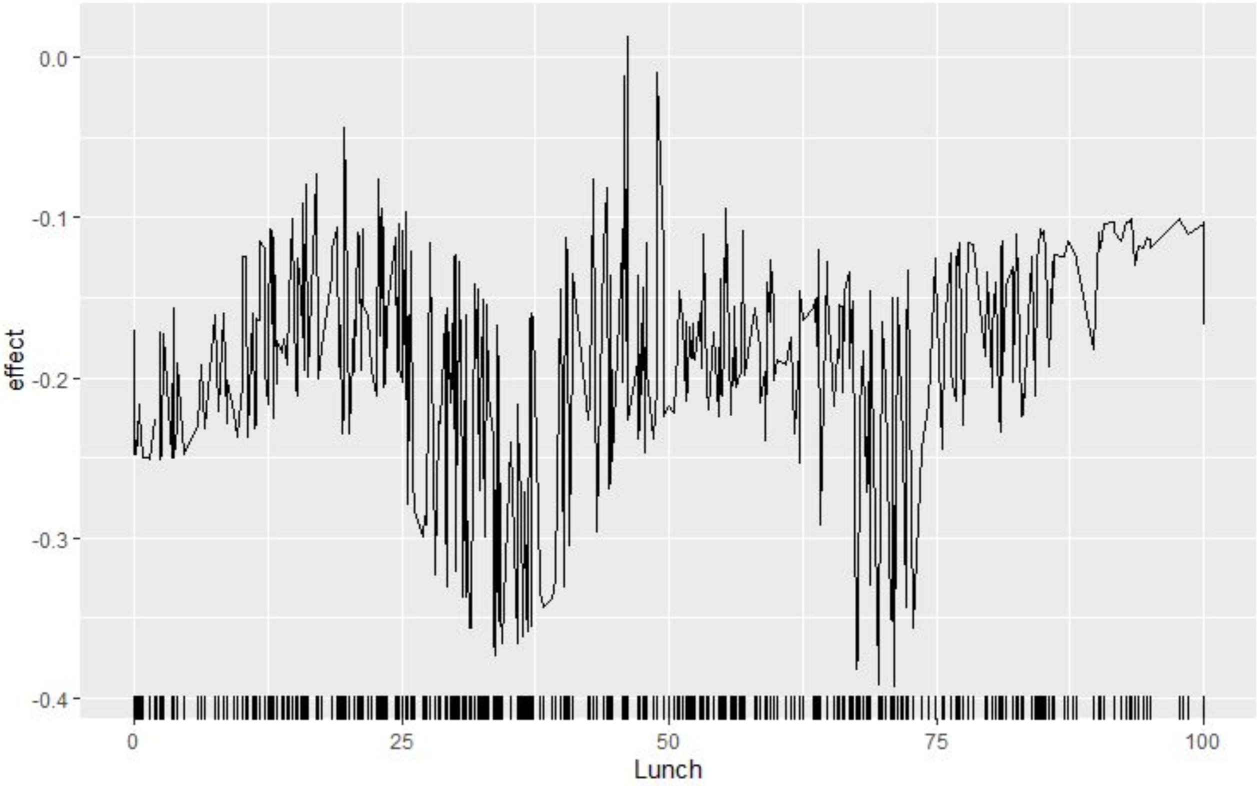
Results: Improving student learning



Span: 0.10625

Question 3

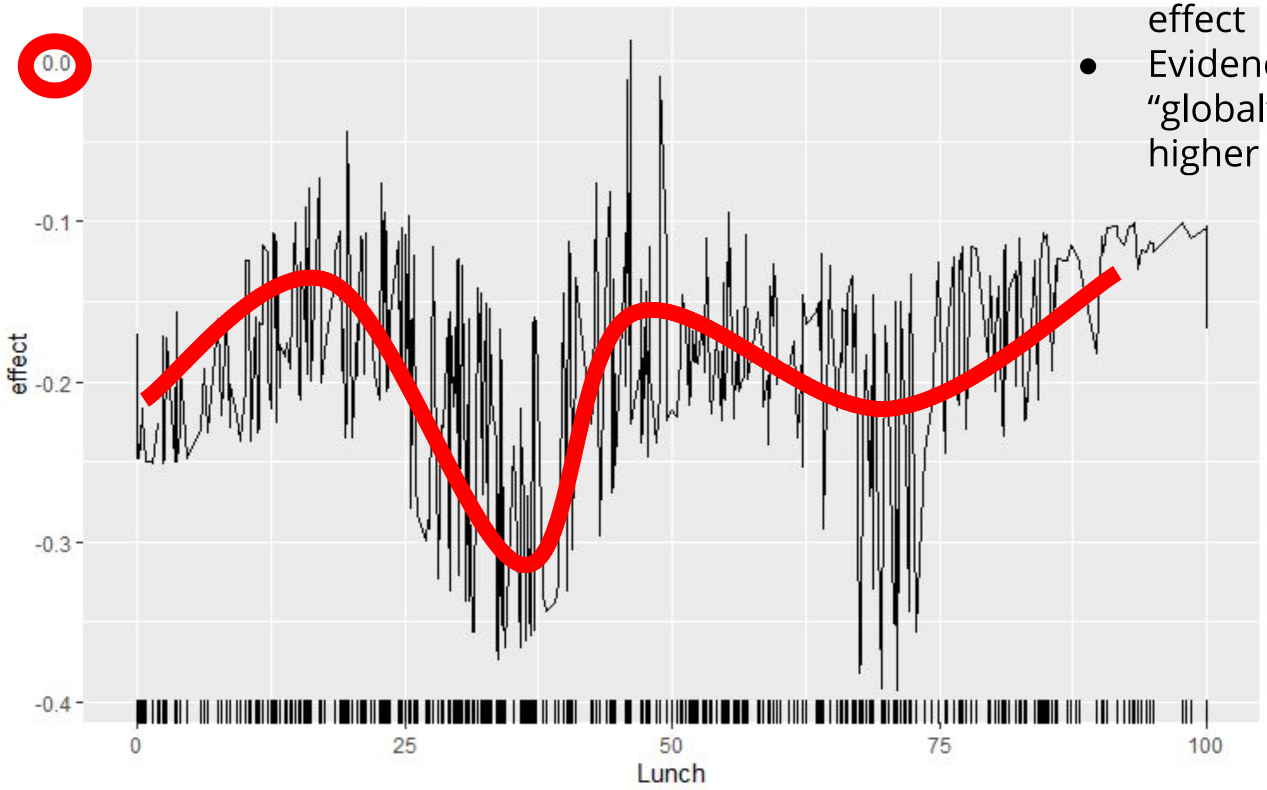
Results: Improving student learning



Span: 0.3875

Question 3

Results: Improving student learning



- Persistent negative effect
- Evidence of more "global" influence from higher span

Span: 0.3875

Question 3

Results: Improving student learning

- The results show that increased spending on extracurriculars (through the *Income* variable) has a very positive effect on student learning (measured by standardized test score).
- Giving students more opportunities to feel successful (through varied arts/sports/academics/service opportunities/etc) will likely increase student learning and success in elementary school.
- Potential policy recommendations:
  - Budget increase/redistribution to provide more funds to extracurriculars
  - Cost-benefit analysis — redistribution of local property tax incomes to smooth extracurricular spending
  - Current system grew out of needs of agrarian society; needs an urban/suburban overhaul
  - Changing the learning “yardstick” to not penalize blue-collar labor
  - Put the means of production in the hands of the proletariat? #marx

Conclusion

Shortcomings?

The model is intentionally overfit to the data that we have, so it lacks generalizability to other school districts/states/etc. Again, interpretability comes mostly down to looking at plots (consequence of *memory-based procedure*).

Next Steps

It would be interesting to use the predictive capabilities of the local regression method in order to predict/forecast scores for various school districts.

Another interesting direction may be a GAM that combines both smoothing splines and local regression (depending on the values of function selection metrics such as p-values/MSE/AIC/BIC).