Final Project:

Public School System Human Resource Expenditure Allocation by State

Table of Contents

Introduction

In an age of tightening public budgets and heavily scrutinized capital expenditures, government agencies must be more careful than ever with how they spend the public's tax dollars and allocate resources to win back the confidence of the public. One such governmental agency that has been particularly subject to dwindling confidence is that of our public school system. A 2012 Gallop Poll showed that confidence in in America's school system had reached an all-time low with only 29% expressing "a great deal" or "quite a lot" of confidence in them (Jones, 2012).

It is not all bad news however as a 2014 PDK/Gallup poll showed that 64% of Americans have trust and confidence in public school teachers (PDK/Gallup, 2014). Also, the previously mentioned 2012 poll reflects general trends seen public perceptions of general distrust in governmental institutions in general. Much of this is driven by perceptions of wasteful spending which Americans believe amounts to 51 cents of every dollar of tax revenue federally, 42 cents at state level, and 37 cents in local government (Rifkin, 2014).

Unfortunately, teachers in certain locations may be feeling the brunt of this. My home state of North Carolina is ranked near the bottom in public school teacher pay and representatives from out of state school districts visit offering better pay and benefits to lure teachers away (Bonner & Hui, 2014). The issue has become a hot button topic in this year's mid-term election. In my county of residence, there is even a sales tax increase proposal that will go partially to public school funding if passed.

Research Question

In taking into account that schools systems must also have supporting personnel in addition to teachers, what are the expenditure characteristics by personnel types of states that have students performing at a high level? Specifically, do states whose public schools achieve certain interquartile performance levels as defined by combined 4th and 8th grade reading and math scores exhibit any common traits in per pupil expenditures in the categories of instructor wages, instructor benefits, pupil support, instructional staff support, general administration, and school administration expense?

Data Appraisal

Data Sources

The data utilized in this study come from multiple sources. Data on Public Elementary-Secondary Education Finance data for all 50 states was obtained from the United States Census Bureau. This information was obtained through the 2012 Census of Governments. This survey is completed every 5 years with 2012 being the most recent year data is available. The data utilized in this study covers the 2011-2012 school year in all states (U.S. Census Bureau, 2014) and is reported on a per pupil basis. The following relevant definitions for per pupil spending categories are shown below:

- Instructor (Teacher) Salary and Wages: Consists of gross compensation before deductions for withheld taxes, retirement contributions, or other purposes.
- Instruction (Teacher) Employee Benefits: Includes contributions on behalf of employees for retirement coverage, social security, group health and life insurance, tuition reimbursement, worker's Compensation, and unemployment compensation. (Note: This

was the only category where data splitting benefits from salaries and wages was available)

- Pupil Support: Expenditure for attendance record-keeping, social work, student accounting, counseling, student appraisal, record maintenance, and placement services. This category also includes medical, dental, nursing, psychological, and speech services.
- Instructional Staff Support: Expenditure for supervision of instructional service improvements, curriculum development, instructional staff training, and media, library, audiovisual, television, and computer assisted instruction services.
- General Administration: Expenditure for board of education and executive administration (office of the superintendent) services.
- School Administration: Expenditure for the office of principal services.

Data on student test scores for 4^{th} and 8^{th} grade reading performance for all 50 states was obtained from the National Center for Education Statistics and covers the 2012-2013 school year (NAEP, 2014). Unfortunately, data for the 2011-2012 school year was not available as it is published on a biannual basis. This data was chosen over 2010-2011 data from the standpoint that expenditures would exhibit a trailing effect over academic performance rather than the opposite. Additionally, 4^{th} and 8^{th} grade scores were chosen as these were the only two grade levels where data was available for all 50 states.

Finally, cost of living index or CLI data was obtained from the US Department of Commerce Bureau of Economic Analysis website (Bureau of Economic Analysis, 2012). CLI data is commonly used for the purpose for comparing living costs between countries, states, and metropolitan areas.

Data Preparation

The data from all sources listed above was combined in Excel 2013 and arranged in vertical columns by the categories of: “State”, “Instructor Salary and Wages”, “Instruction Employee Benefits”, “Pupil Support”, “Instructional Staff Support”, “General Administration”, “School Administration”, “COLI”, “Grade 4 Math”, “Grade 4 Reading”, “Grade 8 Math”, and “Grade 8 Reading”. Horizontal rows are arranged by the data for the state in question.

To normalize the expenditure data for cost of living differences between states and to eliminate the bias that this introduces, each expenditure variable was divided by the “COLI” (Cost of Living Index) variable in Excel to generate a series of new expenditure variables. These are listed in the Excel data file as “(expenditure variable name)/COLI” for all expenditure variables and are shown as “(expenditure variable name).COLI” in Rattle (see below). Additionally a new “Combined Performance” variable was created in Excel by summing the 4th and 8th grade math and reading scores for all states. The Combined performance scores ranged from a low of 964 to a high of 1063.

From here the data were loaded into the Rattle package within R. Using the transform function within Rattle, combined performance variables were binned into quartiles based on score to transform them into categorical variables. This new variable is listed as “BQ4_Combined.Performance” in Rattle. The bins generated organize the data by the following score ranges in order of performance: [964, 999], (999, 1020], (1020, 1030], and (1030, 1060]. These are also referred to as Performance Quartiles 1, 2, 3, and 4 respectively below.

Decision Tree Model

Using the Model function in Rattle, a decision tree (Rpart) was created from the cost of living index adjusted expense variables and total score performance bins to determine the quartile performance by COLI adjusted expense and shown below:

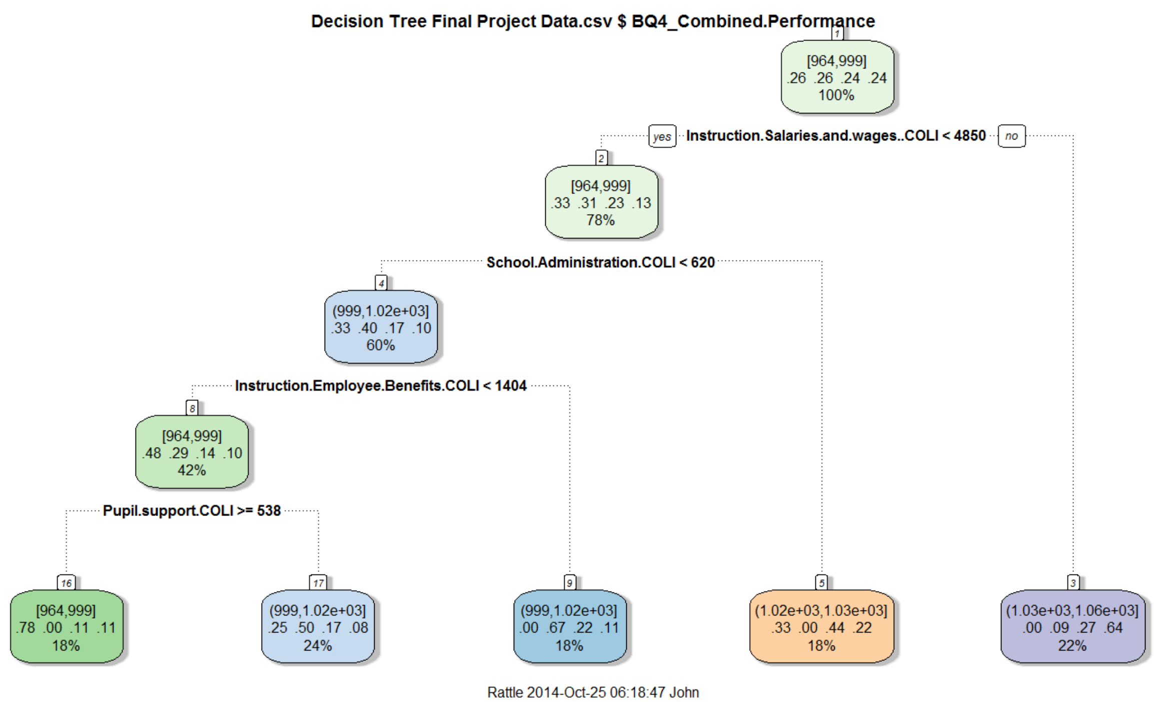

Figure 1. The decision tree model output from R using the Rattle data mining package. Within each box the top numbers represent the combined performance quartile variable bin range, the middle numbers represent the probability of interquartile combined performance values occurring, and the bottom number represents the proportion of the total sample at that node. For reference, the Performance Quartiles are: [964, 999] = 1, (999, 1020] = 2, (1020, 1030] = 3, and (1030, 1060] = 4.

From the top down (starting with node 1, the root node) the decision tree shows that for all 50 states, 26% were placed in either Performance Quartiles 1 or 2 and 24% were placed in Performance Quartiles 3 or 4. The root node was then split on Instruction Salaries and Wages/COLI at a value of less than (node 2) or greater than (node 3) $4,850 per student. 22% of states exceeded this amount and 78% did not. In terminal node 3, 64% of states were in Performance Quartile 4 (highest test scores), 27% were in Performance Quartile 2, 9% were in Performance Quartile 2, and 0% in Performance Quartile 1 (lowest test scores). While this distribution does not suggest certainty between high teacher pay and high student performance, it is reasonable to say that at state level that performance trends upward with higher teacher pay on a per pupil basis.

Non-terminal node 2, had 13% of states in Performance Quartile 4, 23% of states in Performance Quartile 3, 31% in Performance Quartile 2, and 33% in Performance Quartile 1. This node was split on School Administration (principal's office) per student expenditures of more (18% of states) or less than (60% of states) $620 per pupil. The states exceeding this amount are shown in terminal node 5 and are comprised of 22% states in Performance Quartile 4, 44% in Performance Quartile 3, 0% in Performance Quartile 2, and 33% in Performance Quartile 1. While this is more of a "mixed bag" than terminal node 3, the majority of states represented in this node are in the top two Performance Quartiles. This suggests that attracting and retaining School Administration with higher salary levels on a per pupil basis may assist with student performance as well. It is interesting to note that 33% of states in this node were in the lowest Performance Quartile.

The states with School Administration expenditures below $620 per student are shown in node 4. In this node, 10% of values are in Performance Quartile 4, 17% are in Performance

Quartile 3, 40% are in Performance Quartile 2, and 33% are in Performance Quartile 1. This node splits on Instructional Employee Benefits above or below $1,406. Going above this amount leads to terminal node 9 which contains 18% of total states. Of these, 11% of values are in Performance Quartile 4, 22% are in Performance Quartile 3, 67% are in Performance Quartile 2, and 0% are in Performance Quartile 1.

For the remaining states (42% of total) with Instructional Employee Benefits below $1,406, 10% of values are in Performance Quartile 4, 14% are in Performance Quartile 3, 29% are in Performance Quartile 2, and 48% are in Performance Quartile 1. This node then splits on Pupil Support of greater than or equal to $538 with 18% of total states greater than or equal to this (node 16) and 24% of states (node 17) less than this amount. Terminal node 17 contains 8% of its values in Performance Quartile 4, 17% in Performance Quartile 3, 50% in Performance Quartile 2, and 25% of its values in Performance Quartile 1. Nodes 9 and 17 include the same distribution in the absolute number of states by Performance Quartile with the exception that node 17 contains 3 states in the lowest quartile as where node 9 contains none (1/2/6/0 vs. 1/2/6/3).

Terminal node 16 contains 11% of its values in Performance Quartile 4, 11% in Performance Quartile 3, 0% in Performance Quartile 2, and 78% of its values in Performance Quartile 1. It would seem logical that more money has to be dedicated to Pupil Support and pulled away from Instructors and School Administrators in states with low student performance. No one disputes the value of parental involvement in a child’s education and it is reasonable to assume that a lack of parental involvement would lead lower school performance and also to an increased need for increased Pupil Support services to make up for deficiencies at home and their downstream impacts.

Conclusion

The model overall does an excellent job at predicting Performance Quartile 2, predicting 12 of 13 states correctly (see Figures 2 & 3 below) but performed worst on Performance Quartile 3 only predicting 4 of 12 states correctly. The model also predicted Performance Quartiles 1 & 4 correctly for 7 states each. It is reasonable to say that there is probably some relationship between student academic performance and Instructor/School Administration pay per pupil at the performance extremes, however it is far from a given in this model.

After completing the initial model above I introduced data on average class sizes per state in case of this biasing the per pupil pay amounts and reran the model but the variable was not used leading to the same outcome as above so I have not included it. Additionally, I had fears of this leading to confounding due to the relationship between the per pupil expense measurements. I also attempted to utilize a variable that combined Instructor Salary, Wages, and Benefits into one but this caused further misclassifications particularly in Performance Quartile 3 with not one state classified correctly.

Unfortunately, I was unable to find data that broke all of the information down to district levels in terms of student performance as well as the salary and benefits information. My expectation is that by having a much larger data set to work from that more accurate model results would be achieved than having a sample size of only 50 states to work with.

Table 1, Summary of Decision Tree for Classification (built using 'Rpart')

n= 50

node), split, n, loss, yval, (yprob)

* denotes terminal node

1) root 50 37 [964,999] (0.26000000 0.26000000 0.24000000 0.24000000)

2) Instruction.Salaries.and.wages..COLI< 4850 39 26 [964,999] (0.33333333 0.30769231 0.23076923 0.12820513)

4) School.Administration.COLI< 620.5 30 18 (999,1.02e+03] (0.33333333 0.40000000 0.16666667 0.10000000)

8) Instruction.Employee.Benefits.COLI< 1403.5 21 11 [964,999] (0.47619048 0.28571429 0.14285714 0.09523810)

16) Pupil.support.COLI>=538.5 9 2 [964,999] (0.77777778 0.00000000 0.11111111 0.11111111) *

17) Pupil.support.COLI< 538.5 12 6 (999,1.02e+03] (0.25000000 0.50000000 0.16666667 0.08333333) *

9) Instruction.Employee.Benefits.COLI>=1403.5 9 3 (999,1.02e+03] (0.00000000 0.66666667 0.22222222 0.11111111) *

5) School.Administration.COLI>=620.5 9 5 (1.02e+03,1.03e+03] (0.33333333 0.00000000 0.44444444 0.22222222) *

3) Instruction.Salaries.and.wages..COLI>=4850 11 4 (1.03e+03,1.06e+03] (0.00000000 0.09090909 0.27272727 0.63636364) *

Classification tree:

rpart(formula = BQ4_Combined.Performance ~ ., data = crs$dataset[,

c(crs$input, crs$target)], method = "class", parms = list(split = "information"),

control = rpart.control(usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:

[1] Instruction.Employee.Benefits.COLI

[2] Instruction.Salaries.and.wages..COLI

[3] Pupil.support.COLI

[4] School.Administration.COLI

Root node error: 37/50 = 0.74

n= 50

	CP	nsplit	rel error	xerror	xstd
1	0.189189	0	1.00000	1.18919	0.062103
2	0.094595	1	0.81081	1.02703	0.081620
3	0.081081	3	0.62162	1.02703	0.081620
4	0.010000	4	0.54054	0.97297	0.085808

Table 2, Tree as Rules

Rule number: 3 [BQ4_Combined.Performance=(1.03e+03,1.06e+03] cover=11 (22%) prob=7.00]

Instruction.Salaries.and.wages..COLI>=4850

Rule number: 5 [BQ4_Combined.Performance=(1.02e+03,1.03e+03] cover=9 (18%) prob=2.00]

Instruction.Salaries.and.wages..COLI< 4850

School.Administration.COLI>=620.5

Rule number: 9 [BQ4_Combined.Performance=(999,1.02e+03] cover=9 (18%) prob=1.00]

Instruction.Salaries.and.wages..COLI< 4850

School.Administration.COLI< 620.5

Instruction.Employee.Benefits.COLI>=1404

Rule number: 17 [BQ4_Combined.Performance=(999,1.02e+03] cover=12 (24%) prob=1.00]

Instruction.Salaries.and.wages..COLI< 4850

School.Administration.COLI< 620.5

Instruction.Employee.Benefits.COLI< 1404

Pupil.support.COLI< 538.5

Rule number: 16 [BQ4_Combined.Performance=[964,999] cover=9 (18%) prob=1.00]

Instruction.Salaries.and.wages..COLI< 4850

School.Administration.COLI< 620.5

Instruction.Employee.Benefits.COLI< 1404

Pupil.support.COLI>=538.5

Table 3, Error matrix for the Decision Tree model on Final Project Data Set.csv (counts):

Predicted	**[964,999]**	**(999,1.02e+03]**	**(1.02+03, 1.03+03]**	**(1.03+03, 1.06+03]**
Actual				
[964,999]	7	3	3	0
(999,1.02e+03]	0	12	0	1
(1.02+03, 1.03+03]	1	4	4	3
(1.03+03, 1.06+03]	1	2	2	7

Table 4, Error matrix for the Decision Tree model on Final Project Data Set.csv (%):

Predicted	**[964,999]**	**(999,1.02e+03]**	**(1.02+03, 1.03+03]**	**(1.03+03, 1.06+03]**
Actual				
[964,999]	14%	6%	6%	0%
(999,1.02e+03]	0%	24%	0%	2%
(1.02+03, 1.03+03]	2%	8%	8%	6%
(1.03+03, 1.06+03]	2%	4%	4%	14%

References

Bonner, L., & Hui, T. K. (2014, October 30). NC teacher turnover rate dips slightly, though more leave for out-of-state jobs | Education | NewsObserver.com. Retrieved from http://www.newsobserver.com/2014/10/30/4279054_teacher-turnover-rate-dips-slightly.html?sp=/99/102/110/&rh=1

Bureau of Economic Analysis. (2012). U.S. Bureau of Economic Analysis (BEA). Retrieved from http://www.bea.gov/index.htm

Jones, J. M. (2012). Confidence in U.S. Public Schools at New Low. Retrieved from http://www.gallup.com/poll/155258/confidence-public-schools-new-low.aspx

NAEP. (2014). NAEP State Profiles. Retrieved from http://nces.ed.gov/nationsreportcard/states/

PDK/Gallup. (2014, October). Home Copy - PDK International. Retrieved from http://pdkpoll.pdkintl.org/october/#1

Riffkin, R. (2014, September 17). Americans Say Federal Gov't Wastes 51 Cents on the Dollar. Retrieved from http://www.gallup.com/poll/176102/americans-say-federal-gov-wastes-cents-dollar.aspx

U.S. CENSUS BUREAU. (2014). Public Education Finances: 2012 (G12-CG-ASPEF). Retrieved from http://www2.census.gov/govs/school/12f33pub.pdf

Data Analysis Reference

Excel Data Set

Final Project Data

R Code

Final_Project_Data_Set_script.docx