ETM 540 Group Project

Samira Akther
Jordan Hilton
Andey Nunes
Aparna Gandikota
Mohammed Sheikh

March 21, 2019

Executive Summary

Benchmarking is a non-parametric evaluation approach that allows organizations to evaluate their own performances and develop effective plans to improve accordingly. This paper aims to study the efficiency of education systems in the Organization for Economic Development (OECD) countries by using a benchmarking technique, data envelopement analysis (DEA).

DEA is a nonparametric linear programming method that compares relative efficiencies of Decision Making Units (DMUs) and provides a comprehensive measure of each DMU's performance. In our study, DMUs are the country's education system. The objective here is to find the efficiency of the education system in making educational achievements using the resources invested in the system, measured in time and money. Two DEA models, single input-single output and multiple input-multiple output, have been analyzed in fulfilling the objective.

To measure educational achievement, we took PISA scores as an output to the DEA model. PISA is widely recognized performance yardstick to assess the competency of a country's school system. In addition to this, tertiary graduation rate is introduced in multiple output model which refers to the estimated percentage of people who will graduate in tertiary education over their lifetime. For inputs, we considered public spending on education per person that covers direct expenditure on educational institutions as well as educational-related public subsidies given to households and administered by educational institutions. Another output, teaching hours, is also incorporated in the multiple output model . We followed output orientation with variable returns to scale for both the models.

The results of single input-single output model indicated that the education system of Canada, Japan, Greece and Mexico are relatively efficient in utilizing the education funding and thus achieving high PISA score. The countries that are relatively inefficient should adopt the spending policy of a single efficient country or a combination of two countries to achieve the desired efficiency. As we added one more input and output in addition to the ones already in single input-single output model, the multiple input-multiple output model shows improved average efficiency scores. Also, five more countries join the list of countries with efficiency score of 1.

This project is available on GitHub: https://github.com/AndeyNunes/educationDEA

Introduction

In the last few decades, Benchmarking, a non-parametric evaluation approach, has found applications in a variety of fields. The technique has equipped organizations with the means to evaluate their processes and compare it to the best practice of peer organizations. This has allowed them to gauge their own performance and enable them to learn and develop plans to improve aspects of their own performance and set future targets [9].

In the education sector, studies utilizing benchmarking techniques are numerous and vary widely in focus. The richness of the data collected on education far exceeds that which have been thoroughly studied and published. This makes comparative studies on education data an interesting endeavor as it could reveal interesting patterns and discussion points.

In this paper, our aim is to study the efficiency of the education systems of 25 countries from the Organization for Economic Development (OECD) using DEA technique. We have several motivations for working with OECD countries. All OECD countries dedicate a percentage of their Gross Domestic Product (GDP) to education spending and all participate in the Programme for International Student Assessment (PISA), a global educational assessment system. The OECD collects and publicly publishes data on several aspects relating to education such as public spending, enrollment, and graduation rates. And finally, this data is considered reliable as both PISA and OECD are independently run organizations with a track record of professionalism and trustworthy in reporting accurate data.

In literature, the list of studies utilizing DEA for analyzing efficiencies in the education sector are numerous. In their paper, Gavurova and co., summarize past studies utilizing the technique and their obtained results [4]. In an earlier study, afonso and aubyn, used data from 2000-2002 OECD countries and applied DEA using teacher per student ratio and time spent at school as their inputs and the PISA test scores as their output. The researchers used a two stage procedure to show that inefficiencies in spending correlate to GDP per capita and adult educational attainment [4]. Similarly, Agasisti and Dal bianco used 2006-2009 data from 20 European countries and used expenditure per student as the input and PISA test results as the output. The researchers extended their found relative efficiences and compared it to other variables such as GDP per capita, unemployment rate, etc... [4]. Gavurova and his team, used similar set up and relied on 2014 data from European countries. The team looked at government expenditure in secondary education as their input and PISA test results as the output. Their findings grouped countries into four categories. High spending with good results, high spending with low results, low spending with high results, and finally, low spending with low results [4].

Benchmarking with Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is just one of a few benchmarking techniques currently in use including stochastic frontier analysis (SFA). DEA is a widely used benchmarking technique that was first proposed by Charnes, Cooper, and Rhodes in 1978. The technique is a non-parametric linear programming method that compares relative efficiencies of Decision Making Units (DMUs) and provides a comprehensive measure of each DMU's performance. In a DEA analysis, efficiency is usually given as the weighted sum of outputs divided by the weighted sum of the input resources [2,3,6].

$$\text{Efficiency} = \text{Weighted} \sum \frac{outputs}{inputs}$$

DEA models can be categorized as either input oriented or output oriented. Input-oriented models minimize inputs while satisfying at least the given output levels, while output oriented models maximize outputs without requiring more of any observed input values [7].

DEA models can also be categorized according to the weight constraints. In this case, Models are subdivided into four categories. Constant returns to scale CRS, where the CCR model is built upon and has no constraints. Variable returns to scale VRS where the sum of all lambda is set to 1. Increasing returns to scale IRS where is the sum is set to greater than 1 and Decreasing returns to scale DRS having the sum being less than 1 [2].

Returns to Scale	Envelopment Constraint
CRS	No constraint needed
VRS	$\sum_{j=1}^{N^D} \lambda_j = 1$
IRS	$\sum_{j=1}^{N^D} \lambda_j \geq 1$
DRS	$\sum_{j=1}^{N^D} \lambda_j \leq 1$

In this paper, we specify an output oriented variable returns to scale model to look at the efficiences of the 25 OECD countries in terms of public education resources as the inputs and graduation rates along with the PISA test score as our outputs. The PISA test score used is the average of individual test scores in Science, Math and Reading. Our objectives is to compare different countries resource use against the academic achievements of students.

Methodology

Benchmarking using a DEA model involves measuring the relative efficiency of a Decision Making Unit (DMU), which is determined by comparing the studied unit's performance in producing output with a target. [2] In our study, DMUs are the country's education system. The objective here is to find the efficiency of this education system in making educational achievements, measured by the PISA test scores and college graduation rates, using the resources invested in the system, measured in time and money.

Public spending on education is one of the inputs in our two DEA models. It covers direct expenditure on educational institutions as well as educational-related public subsidies given to households and administered by educational institutions. [10] It comprises spending on primary, secondary and tertiary education. Spending is shown in USD per student as per student data will help us to eliminate the size effect of a country with large GDP and thus large budget for education.

In our second DEA model, we introduced another input, teaching hours. It refers to preparation time and statutory teaching time. Teaching time is the number of hours spent teaching a group or class of students according to the formal policy in the country; preparation time is the time spent preparing for teaching. [11]

As our goal is to evaluate the efficiency of the country's education system, we decided to use PISA score as measurement of success. PISA is widely recognized performance yardstick to assess the competency of a country's school system. PISA assesses the extent to which 15-year-old students, near the end of their compulsory education, have acquired key knowledge and skills that are essential for full participation in modern societies. [12] The assessment focuses on three core school subject: reading, mathematics and science. Another output, tertiary graduation rate, is incorporated in our second DEA model. Tertiary graduation rate represents the estimated percentage of people who will graduate in tertiary education over their lifetime. [13]

Our project looks first at a single input of Total Spending per student and single output of average PISA score. We then expand on the model by including additional input of Total Hours teachers spend along with the more granular PISA scores and an additional output, tertiary graduation rate, for a multiple input - multiple output model.

In each case, our DEA model is based on output-orientation because the focus of the model is on improving and optimizing the output based on the input as observed in studied units. We assumed that each country has a given fund for spending on education which is influenced by various factors including the GDP, education budgets, and political factors.

The basic DEA model is based on an assumption of constant returns to scale. This means that efficient output with a given input can be scaled up or down in a linear or constant rate. Prof.Anderson describes this as exhibiting no limit to how big an operation can get with the ratio of input, whereas often in reality, there are structural differences that inhibit such behavior [2]. The DEA model in our study is based on variable returns to scale (VRS) on the nature of the inputs and outputs. With VRS specified, the input can scale up or down, but the output may not scale up or down by the same amount. [3] In fact in the case of test scores, there is an upper limit on the output.

Consider n number of education systems DMU_j which consumes $x_{i,j}$ inputs and produce $y_{r,j}$ output. The education system operates in variable returns to scale. So this output oriented model can be expressed by a linear algebraic model where the objective will be to maximize the efficiency scores of j education system. The efficiency scores can be defined by ϕ

$$
\begin{aligned}
&\text{Max } \phi \\
&\text{subject to } \sum_{j=1}^{n} x_{i,j}\lambda_j \le x_{i,k} \;\forall\, i \\
&\sum_{j=1}^{n} y_{r,j}\lambda_j \ge \phi_{r,k} \;\forall\, r \\
&\lambda_j \ge 0 \;\forall\, j \\
&\sum_{j=1}^{n} \lambda_j = 1
\end{aligned}
$$

Here, a country's education system is considered efficient if the efficiency score ϕ is 1. The country with score of 1 lies in the efficient frontier which is a linear combination of all the countries with efficiency scores of 1. The countries with scores less than 1 will be considered as relatively inefficient in comparison to the target specified by λ. The vector λ is specific amount of a unit j used in setting the target for performance for studied unit k. To accommodate the variable returns to scale of our studied input and output, a constraint that λ sums up to 1 is added to the model. Also, as it is a output oriented model, the first (input) constraint will be satisfied while trying to exceed the second constraint (output) by as much possible.

The discussed methodology has been used to perform two DEA model and analysis for the 25 OECD countries for which we have all the data points.

The following OECD countries that were dropped from the analysis due to missing data:

```
##  [1] "Australia"      "Belgium"        "Denmark"        "Estonia"
##  [5] "France"         "Iceland"        "Ireland"        "Lithuania"
##  [9] "Sweden"         "Switzerland"    "United Kingdom" ""
```

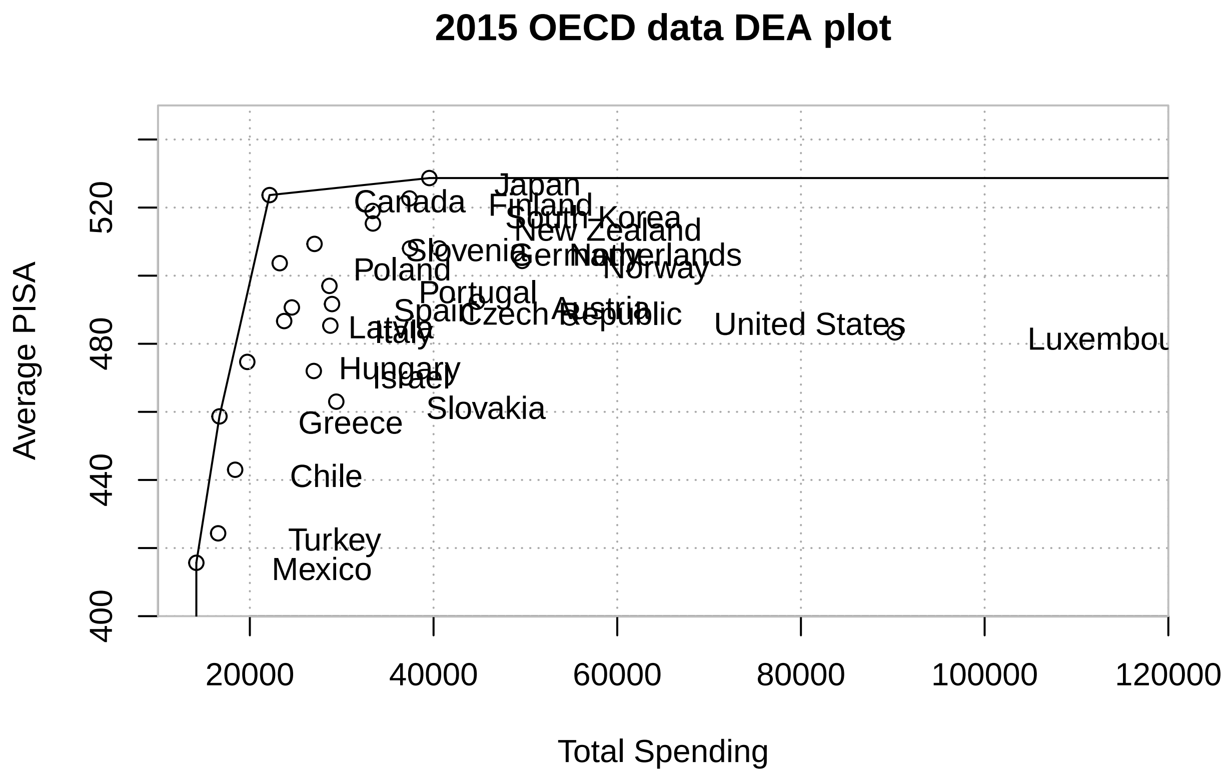
Data was obtained from secondary sources. Data on PISA scores and teaching hours were obtained from OECD website. Data on public education spending per student and tertiary graduation rate were obtained from UNESCO website [10-13].

Summary of the data and all modeling code chunks in the next several sections are included in the Appendix.

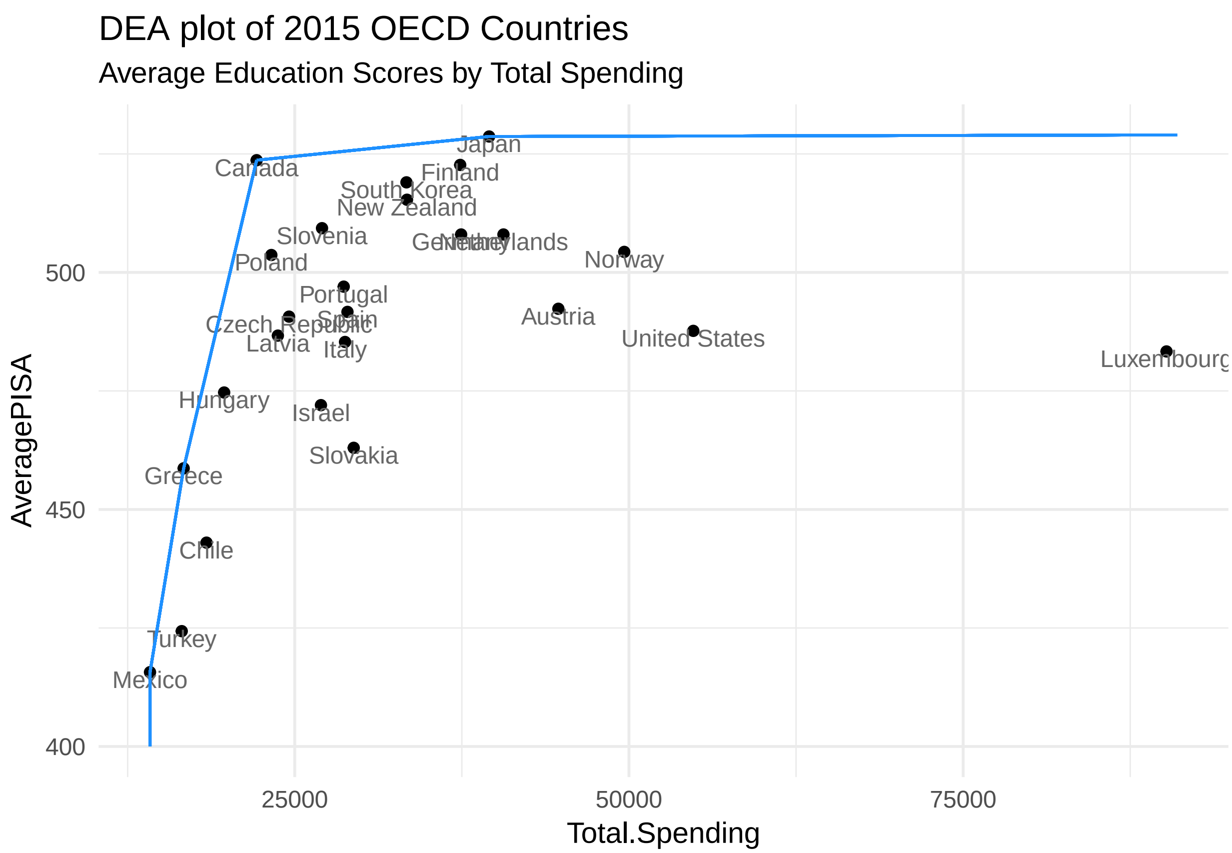
DEA Model

Part A: Single input, single output

For this first part, we will set up and solve a DEA model using total educational spending per student for the input and the aggregated PISA test score average for the single output. The input output diagram for this model is included in the Appendix.



This plot is not formatted for easy reading. After examining the `dea.plot()` function from Bogetoft & Otto's Benchmarking package, we were able to emulate the graphic in `ggplot`.

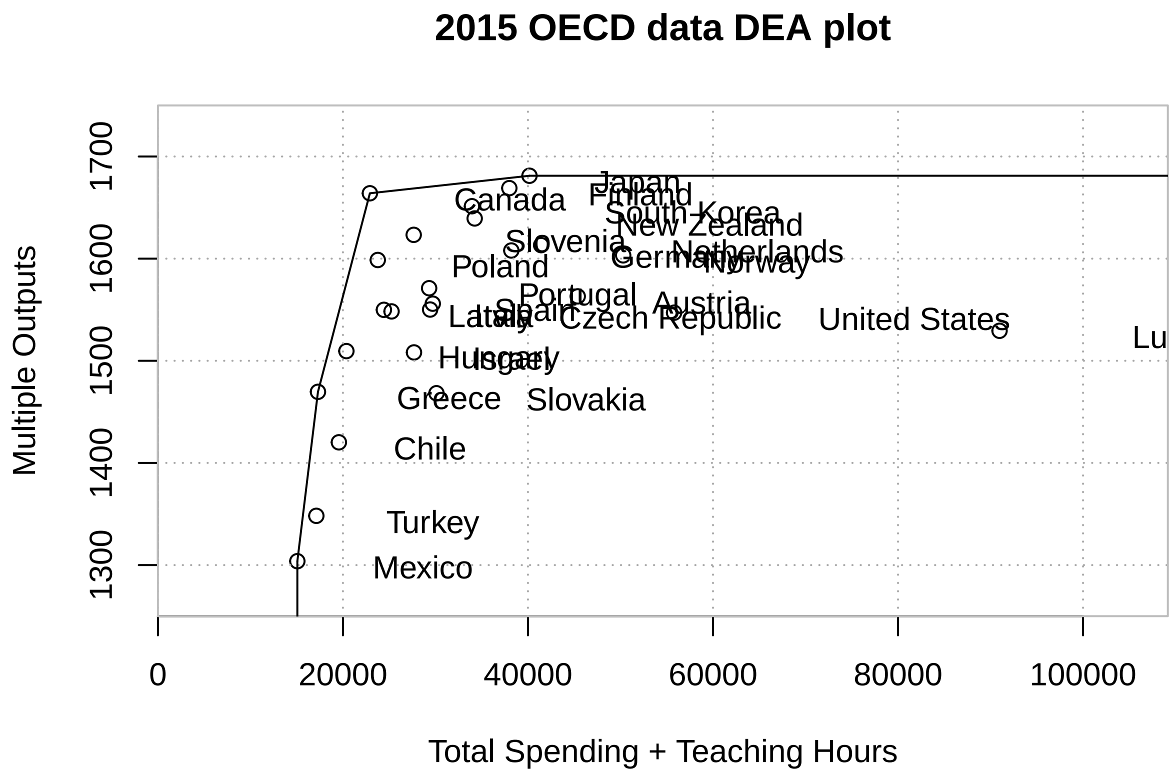


Part B: Multiple Inputs and Outputs

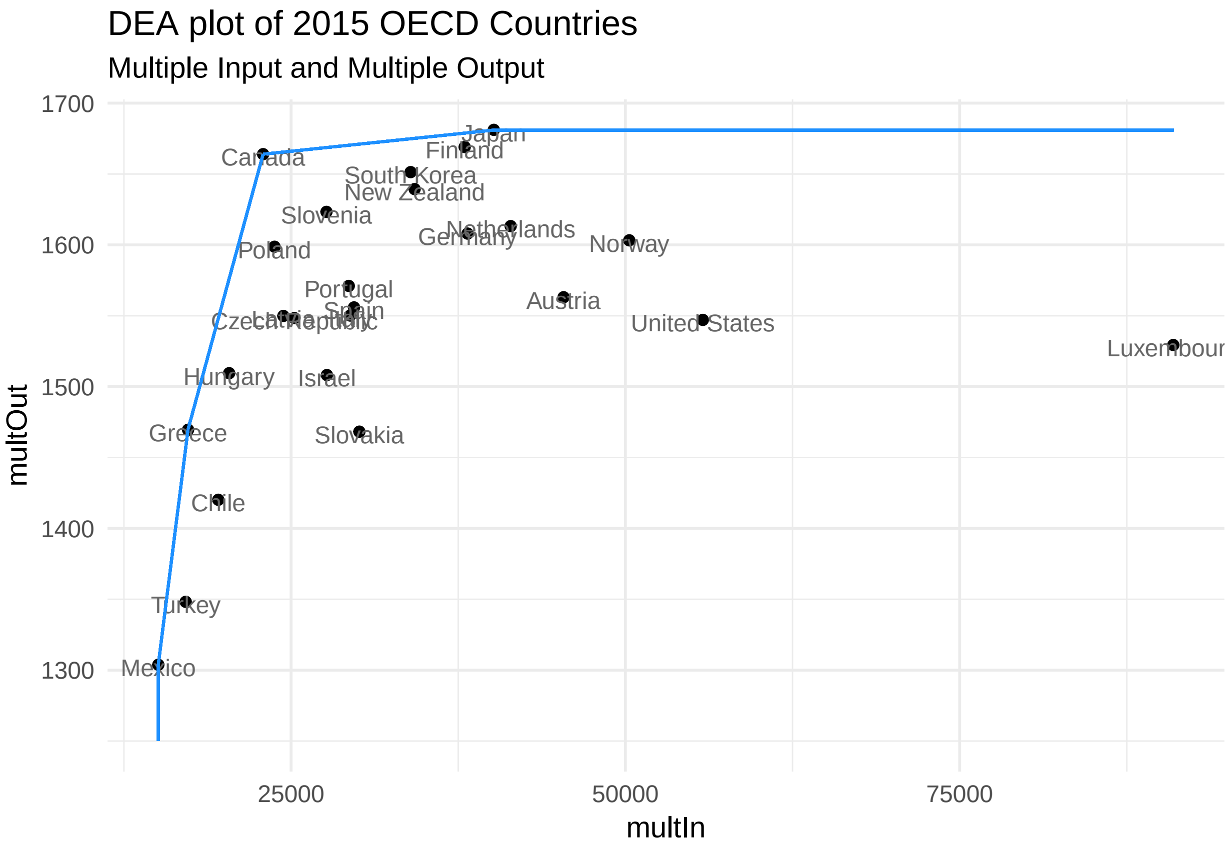
We are also interested in how many hours each country's teachers spend teaching per year as another input, and the tertiary graduation rate as another output. Additionally, we are also using the different science, math, and reading PISA scores that made up the `averagePISA` input from the single input single output model.

Here is a model with two inputs, total educational spending per student and average teacher hours per year, and four outputs: SciencePISA, ReadingPISA, MathematicsPISA, and tertiary graduation rate.

The multiple input multiple output DEA plot is shown here:



And just to show it more clearly, we recreate the DEA plot using ggplot.



Results and Discussion

Single input - output results

Our observation on our single input-single output are:

a. The average efficiency for the Total Spending yielding Average PISA score is 0.952. It indicates that the overall efficiency of the studied countries education system is overall high.

b. The table of efficiency and lambda scores refers that the following countries have the efficiency scores of 1: > Canada, Greece, Japan, Mexico

It indicates that education systems of these countries are relatively efficient in comparison to all studied education systems. So these DMUs lie in the efficient frontier as shown in the figure above. If we look at the PISA scores of all the countries, we can see that Canada and Japan achieved two of the highest scores of all subjects of PISA assessment. Interestingly, Mexico and Greece did not achieve good scores in comparison to other countries. But both of the countries spending on education per student are two of the lowest spending in our data set. So the input output ratios became higher and put them in the efficient frontier.

c. Slovakia has the minimum efficient scores indicating that Slovakia's education system is the least efficient of all the countries in achieving comparable PISA assessment.

d. The countries that have efficiency scores below 1 are benchmarked against a target. The target is formed by the countries with efficiency scores of 1 by scaling up the output. In our model, education system of Canada, Japan, Mexico and Greece are the benchmark for all the other countries education system.

e. Austria, Luxembourg, Netherlands, Norway and U.S. have been benchmarked against Japan. It indicates

Table 2: Results of Single input-Single output DEA displaying efficiency scores and positive lambda values

	Eff	Canada	Greece	Japan	Mexico
Austria	0.931	0.000	0.000	1.000	0.000
Canada	1.000	1.000	0.000	0.000	0.000
Chile	0.925	0.313	0.687	0.000	0.000
Czech Republic	0.936	0.861	0.000	0.139	0.000
Finland	0.990	0.124	0.000	0.876	0.000
Germany	0.962	0.120	0.000	0.880	0.000
Greece	1.000	0.000	1.000	0.000	0.000
Hungary	0.959	0.555	0.445	0.000	0.000
Israel	0.899	0.723	0.000	0.277	0.000
Italy	0.923	0.620	0.000	0.380	0.000
Japan	1.000	0.000	0.000	1.000	0.000
South Korea	0.985	0.356	0.000	0.644	0.000
Latvia	0.929	0.909	0.000	0.091	0.000
Luxembourg	0.914	0.000	0.000	1.000	0.000
Mexico	1.000	0.000	0.000	0.000	1.000
Netherlands	0.961	0.000	0.000	1.000	0.000
New Zealand	0.978	0.353	0.000	0.647	0.000
Norway	0.954	0.000	0.000	1.000	0.000
Poland	0.961	0.937	0.000	0.063	0.000
Portugal	0.946	0.625	0.000	0.375	0.000
Slovakia	0.881	0.582	0.000	0.418	0.000
Slovenia	0.970	0.719	0.000	0.281	0.000
Spain	0.935	0.609	0.000	0.391	0.000
Turkey	0.930	0.000	0.942	0.000	0.058
United States	0.922	0.000	0.000	1.000	0.000

Table 3: Results of Multiple Input Multiple Output DEA displaying efficiency scores and positive lambda values

	Eff	Canada	Finland	Greece	Japan	South Korea	Mexico	Poland	Slovenia	Turkey
Austria	0.938	0.119	0.000	0.000	0.881	0.000	0	0.000	0.000	0
Canada	1.000	1.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0
Chile	0.969	0.041	0.072	0.887	0.000	0.000	0	0.000	0.000	0
Czech Republic	0.955	0.602	0.000	0.000	0.000	0.197	0	0.202	0.000	0
Finland	1.000	0.000	1.000	0.000	0.000	0.000	0	0.000	0.000	0
Germany	0.972	0.707	0.000	0.000	0.293	0.000	0	0.000	0.000	0
Greece	1.000	0.000	0.000	1.000	0.000	0.000	0	0.000	0.000	0
Hungary	0.984	0.374	0.000	0.475	0.000	0.000	0	0.151	0.000	0
Israel	0.953	0.229	0.436	0.335	0.000	0.000	0	0.000	0.000	0
Italy	0.971	0.354	0.477	0.142	0.000	0.000	0	0.000	0.027	0
Japan	1.000	0.000	0.000	0.000	1.000	0.000	0	0.000	0.000	0
South Korea	1.000	0.000	0.000	0.000	0.000	1.000	0	0.000	0.000	0
Latvia	0.954	0.503	0.054	0.135	0.000	0.000	0	0.000	0.308	0
Luxembourg	0.924	0.392	0.000	0.000	0.608	0.000	0	0.000	0.000	0
Mexico	1.000	0.000	0.000	0.000	0.000	0.000	1	0.000	0.000	0
Netherlands	0.970	0.249	0.000	0.000	0.751	0.000	0	0.000	0.000	0
New Zealand	0.996	0.353	0.000	0.000	0.647	0.000	0	0.000	0.000	0
Norway	0.977	0.224	0.685	0.000	0.092	0.000	0	0.000	0.000	0
Poland	1.000	0.000	0.000	0.000	0.000	0.000	0	1.000	0.000	0
Portugal	0.956	0.399	0.246	0.000	0.114	0.051	0	0.190	0.000	0
Slovakia	0.909	0.565	0.000	0.000	0.386	0.049	0	0.000	0.000	0
Slovenia	1.000	0.000	0.000	0.000	0.000	0.000	0	0.000	1.000	0
Spain	0.941	0.986	0.000	0.000	0.014	0.000	0	0.000	0.000	0
Turkey	1.000	0.000	0.000	0.000	0.000	0.000	0	0.000	0.000	1
United States	0.943	1.000	0.000	0.000	0.000	0.000	0	0.000	0.000	0

that these five counties should achieve the average PISA scores of 528.67 (average PISA score of Japan) while maintaining the current spending on education. So these countries can become relatively efficient by following Japan's spending policy and schooling system.

f. The other relaively inefficient countries have been benchmarked against two countries, Canada and Greece. It indicates that by adopting a mixture of two countries spending policy and education system, the inefficient countries can improve their efficiency scores and thus become relatively efficient. This benchmarking process can be explained by taking example of one of the inefficient unit.[14] Consider the education system of Chile. It has average PISA of 443. The lambda values indicated that it should be benchmarked against Canada and Greece and the weights are 31.3% and 68.7% respectively. So the reference Average PISA score for Chile is 479.015 (31.3% of Average PISA of Canada 523.67)+ (68.7% of Average PISA of Greece 458.67). This reference score is higher than the actual average PISA scores of Chile by 36.015 points. So the DEA analysis infers that by following a combination of spending policy Canada and Greece, Chile can improve its average PISA score by 36.015 points and thus increase the efficiency score by (100-92.5) 7.5% points. Similar analysis can be done for all inefficient units.

Multiple input - output results

The average efficiency for the Total Spending & Total Hours yielding multiple PISA scores & tertiary graduation rates is 0.972.

The table of efficiency and lambda scores indicates the following countries are have efficiency scores of 1: > Canada, Finland, Greece, Japan, South Korea, Poland, Slovenia

The interpretation of the results of this multiple input-multiple output DEA will be similar as the above single input-single output DEA. But there is one interesting point to note here. If we look into the average efficiency score, we see that it is higher than the single input-single output DEA model. Again, in addition to the countries which have efficiency scores of 1 under single input-single output model, five more countries achieved efficiency score of 1. These results affirm the notion that adding more input and output variables increase the efficiency score. However, there is a limit to the usefulness of adding more inputs because it will always improve efficiency scores (but won't harm the poorer performers scores).

To wrap it all up, we looked to see if there is any trends or correlation between efficiences from the single and multiple input/output models and variables such as GDP, HDI, spending percentage per GDP. The results from this are inconclusive. For example, Japan, a country with high HDI and low spending as percent of GDP had an efficient system according to our models and similar, mexico with a high spending per GDP, and low HDI had an efficient system too. Overall, it seems like all the countries are operating near efficiency frontier. The least efficient country, slovaka, is operating at 88.1% and 90.9% of efficiency frontier in both single input/output and multiple input/output models respectively. OECD as an organization can be attirbuted to the low gap between member states as the organization has been collecting data and actively benchmarking between them to profilerate best known methods and shared learning.

Country	Per Student Spending	Teacher Hours	HDI	GDP	GDP%	Efficiency 1	Efficiency 2
Austria	44721	658	0.903	382.1	5.45%	0.931	0.938
Canada	22149	761	0.920	1560.0		1.000	1.000
Chile	18400	1157	0.840	244.0	4.87%	0.925	0.969
Czech Republic	24574	676	0.882	186.8	5.79%	0.936	0.955
Finland	37378	606	0.915	232.5	7.09%	0.990	1.000
Germany	37446	754	0.933	3376.0	4.81%	0.962	0.972
Greece	16691	607	0.866	195.5		1.000	1.000
Hungary	19720	650	0.834	122.9	4.58%	0.959	0.984
Israel	26961	718	0.901	299.1	5.88%	0.899	0.953
Italy	28762	661	0.876	1833.0	4.08%	0.923	0.971
Japan	39541	621	0.905	4395.0	3.59%	1.000	1.000
South Korea	33358	585	0.898	1383.0		0.985	1.000
Latvia	23739	685	0.841	27.0	5.34%	0.929	0.954
Luxembourg	90212	763	0.899	57.8	3.92%	0.914	0.924
Mexico	14173	898	0.767	1170.0	5.24%	1.000	1.000
Netherlands	40614	810	0.926	758.0	5.40%	0.961	0.970
New Zealand	33398	841	0.914	177.6	6.34%	0.978	0.996
Norway	49649	642	0.948	386.7	7.55%	0.954	0.977
Poland	23250	513	0.855	477.4	4.81%	0.961	1.000
Portugal	28664	651	0.842	199.4	4.88%	0.946	0.956
Slovakia	29411	698	0.851	87.5	4.65%	0.881	0.909
Slovenia	27040	608	0.889	43.1	4.91%	0.970	1.000
Spain	28945	762	0.885	1198.0	4.28%	0.935	0.941
Turkey	16546	576	0.783	859.8	4.29%	0.930	1.000
United States	54814	979	0.920	18120.0	4.99%	0.922	0.943

References

[1] Anderson, T. R. (2019) Operations Research Using R

[2] Anderson, T. R. (2019) Data Envelopment Analysis Using R

[3] Bogetoft, P., Otto, L. (2011) Benchmarking with DEA, SFS, and R. Springer.

[4] Gavurova, B., Kocisova, K., Belas, L., & Krajcik, V. (2017). Relative efficiency of government expenditure on secondary education. Journal of International Studies, 10(2), 329-343. doi:10.14254/2071-8330.2017/10-2/23

[5] Zhu, H. (2019) Create Awesome LaTeX Table with knitr::kable and kableExtra. https://haozhu233.github.io/kableExtra/awesome_table_in_pdf.pdf

[6] Lamb, J., Tee,K.H. (2012) Data envelopment analysis models of investment funds. European Journal of Operational Research, Volume 216, Issue 3, Pages 687-696

[7] Sherman, H. D., Zhu, J. (2006) Service Productivity Management: Improving Service Performance Using Data Envelopment Analysis (DEA). Springer Science, pages 49-89.

[8] Ji, Y., Lee, C. (2010) Data envelopment analysis. Stata Journal, 10, issue 2, p. 267-280, https://EconPapers.repec.org/RePEc:tsj:stataj:v:10:y:2010:i:2:p:267-280.

[9] Ruiz, J., Segura,J., Sirvent,I. (2015) Benchmarking and target setting with expert preferences: An application to the evaluation of educational performance of Spanish universities. European Journal of Operational Research,Volume 242, issue 2, p. 594-605, https://EconPapers.repec.org/RePEc:eee:ejores:v:242:y:2015:i:2:p:594-605

[10] OECD (2016), Public spending on education (indicator). doi: 10.1787/f99b45d0-en (Accessed on 20 March 2019)

[11] OECD (2018), Teaching hours (indicator). doi: 10.1787/af23ce9b-en (Accessed on 20 March 2019)

[12] OECD (2018), PISA 2015 Results in Focus. Retrieved from OECD website: www.oecd.org/pisa

[13] OECD (2019), Tertiary graduation rate (indicator). doi: 10.1787/15c523d3-en (Accessed on 20 March 2019)

[14] Sherman, D. H., Zhu, J.(2013). Analyzing Performance in Service Organizations. MITSloan Management Review, Vol.54, No. 4.

Appendix

Data Summary

```r
glimpse(data)
```

```
## Observations: 40
## Variables: 22
## $ Countries                     <chr> "Australia", "Austria", "Belgium...
## $ GDP                           <dbl> 1349.0, 382.1, 455.0, 1560.0, 24...
## $ SciencePISA                   <int> 510, 495, 502, 528, 447, 493, 50...
## $ ReadingPISA                   <int> 503, 485, 499, 527, 459, 487, 50...
## $ MathematicsPISA               <int> 494, 497, 507, 516, 423, 492, 51...
## $ AveragePISA                   <dbl> 502, 492, 503, 524, 443, 491, 50...
## $ HDI                           <dbl> 0.936, 0.903, 0.913, 0.920, 0.84...
## $ Primary.education             <chr> "9,546", "11,689", "10,211", "9,...
## $ Secondary.education           <chr> "12,303", "15,477", "13,070", "1...
## $ Tertiary.education            <chr> "20,344", "17,555", "17,320", "0...
## $ Total.Spending                <dbl> 42193, 44721, 40601, 22149, 1840...
## $ Tertiary.Graduation.Rate      <dbl> NA, 86.1, 31.4, 93.0, 91.2, 76.4...
## $ Bachelors.Graduation.rates    <dbl> 59.77, 25.01, 43.90, 37.59, 35.9...
## $ Masters                       <dbl> 20.53, 20.29, 26.76, 11.78, 10.1...
## $ Doctorates                    <dbl> 2.618, 1.862, 0.639, 1.559, 0.26...
## $ Spending.as.percentage.of.GDP <chr> "5.32%", "5.45%", "6.55%", "", "...
## $ Teaching.Hours                <dbl> 825, 658, NA, 761, 1157, 676, NA...
## $ X                             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ X.1                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ X.2                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
```

```
## $ X.3                              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ X.4                              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
```

Let's glance at the teaching hours and the graduation rate:

```
head(cbind(data$Countries, data$Teaching.Hours, data$Tertiary.Graduation.Rate))
```

```
##      [,1]             [,2]          [,3]
## [1,] "Australia"      "825.4093373" NA
## [2,] "Austria"        "658.2"       "86.079"
## [3,] "Belgium"        NA            "31.35"
## [4,] "Canada"         "760.623341"  "93.005"
## [5,] "Chile"          "1157.36"     "91.196"
## [6,] "Czech Republic" "676.3166667" "76.361"
```

DEA model R code

Single input single output

```
# Run a DEA with a single input (total spending) and a single output (Average PISA)
x <- DEAdata %>% select(Total.Spending) # input
y <- DEAdata %>% select(AveragePISA) # output

row.names(x) <- DEAdata$Countries # input labels
row.names(y) <- DEAdata$Countries # output labels

# output oriented model with variable returns to scale
ressingle <- DeaMultiplierModel(x, y, rts = "vrs", orientation = "output")
```

Table formatting using the `kable_styling()` function from the kableExtra package

```
df <- cbind(ressingle$Efficiency, ressingle$Lambda)

tempdf <- df[, colSums(df) != 0]

kable(tempdf, "latex", caption = "Results of Single input-Single output DEA
       displaying efficiency scores and positive lambda values",
      booktabs = T) %>%
kable_styling(latex_options = "striped", "repeat_header")
# code for kableExtra package from Zhu(2019)
```

Multiple input multiple output

```
# multiple input and multiple output
x <- DEAdata %>% select(Total.Spending, Teaching.Hours) ## input

y <- DEAdata %>% select(SciencePISA, ReadingPISA, MathematicsPISA, Tertiary.Graduation.Rate) ## output

row.names(x) <- DEAdata$Countries # input labels
row.names(y) <- DEAdata$Countries # output labels


resmult <- DeaMultiplierModel(x, y, rts = "vrs", orientation = "output")
```

Table formatting using the `kable_styling()` function from the kableExtra package

```
df <- cbind(resmult$Efficiency, resmult$Lambda)

tempdf <- df[, colSums(df) != 0]

kable(tempdf, "latex", caption = "Results of Multiple Input Multiple Output DEA
      displaying efficiency scores and positive lambda values", booktabs = T) %>%
kable_styling(latex_options = c("striped", "scale_down", "repeat_header"))
# code for kableExtra package from Zhu(2019)
```

Code for DEA Model Graphs

For some reason the DiagrammeR graphs won't render to pdf directly, only to html (then print to pdf). They can be previewed and saved as png in RStudio IDE, which is what we did for this report.

Single input single output DEA model

It turns out these grViz objects don't play nice rendering to pdf. They are much nicer looking/behaving in html.

```
grViz("
digraph nicegraph {

  # a 'graph' statement
  graph [overlap = true, fontsize = 10]

  # several 'node' statements
  node [shape = rectangle, fixedsize = true, width = 1,
        color = darkslategray]
  Eff

  node [shape = plaintext]
  TotalSpending; AveragePISA

  # 'edge' statements
  edge [color = grey]
  TotalSpending->Eff Eff->AveragePISA
}
")
```

Single input output dea plot

```
dea.plot(x = x, y = y, txt = dimnames(x)[[1]], GRID = T,
         xlab = "Total Spending", ylab = "Average PISA",
         xlim = c(10000,120000),
         ylim = c(400,550), main = "2015 OECD data DEA plot")
```

Single input output ggplot code

```
# warning, the geom_segment calls are not reproducible and must be hand specified
ggplot(DEAdata, aes(x = Total.Spending, y  = AveragePISA)) +
   geom_point() +
   geom_text(aes(label = Countries), size = 3,
             nudge_x = 2, nudge_y = -1.5, color = "gray40") +
   geom_segment(aes(x = 91000, y = 529, xend = 39541, yend = 528.67), color = "dodgerblue") +
   geom_segment(aes(x = 39541, y = 528.67, xend = 22149, yend = 523.67), color = "dodgerblue") +
   geom_segment(aes(x = 22149, y = 523.67, xend = 16691, yend = 458.67), color = "dodgerblue") +
```

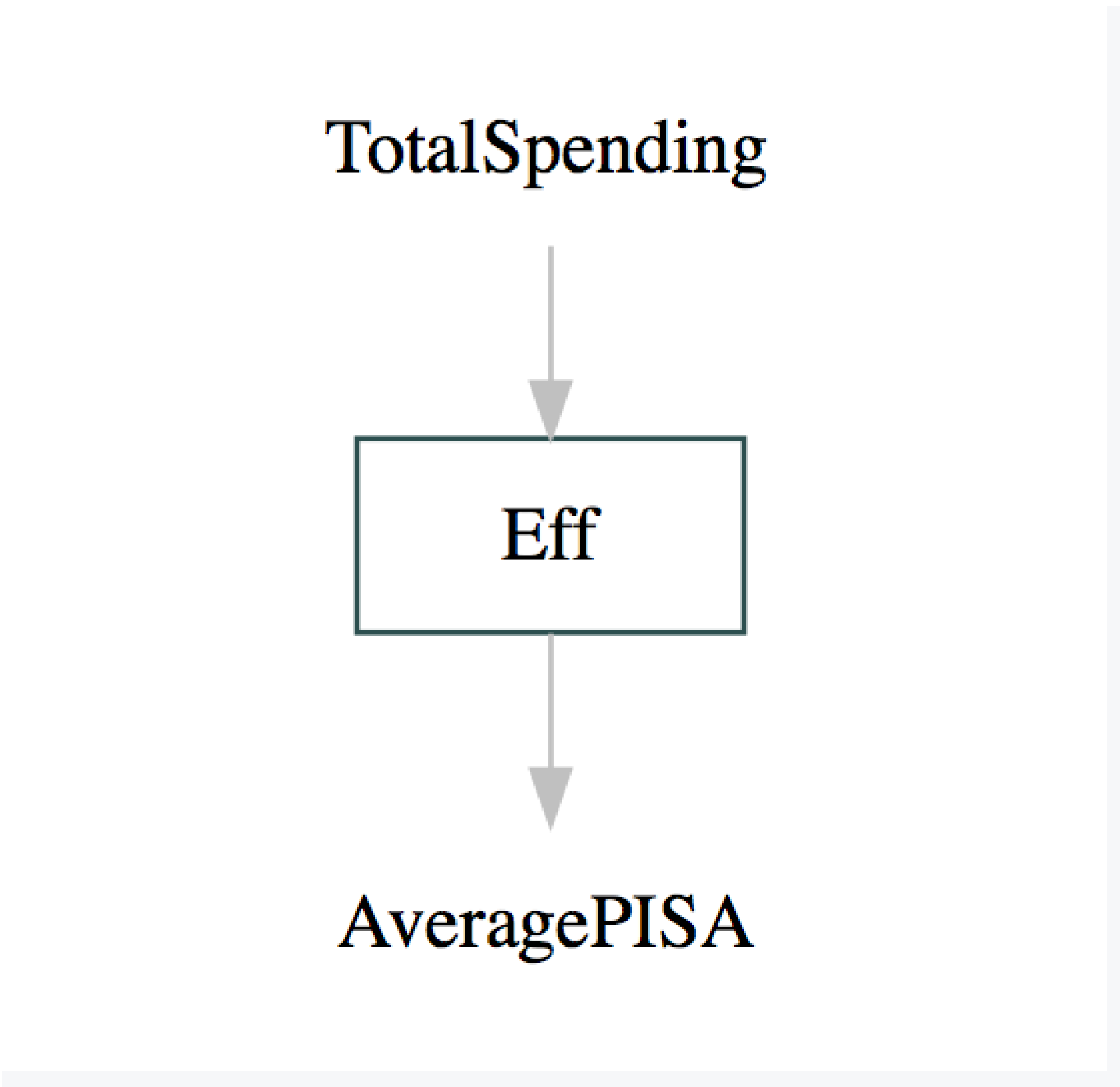


Figure 1: Single Input Single Output

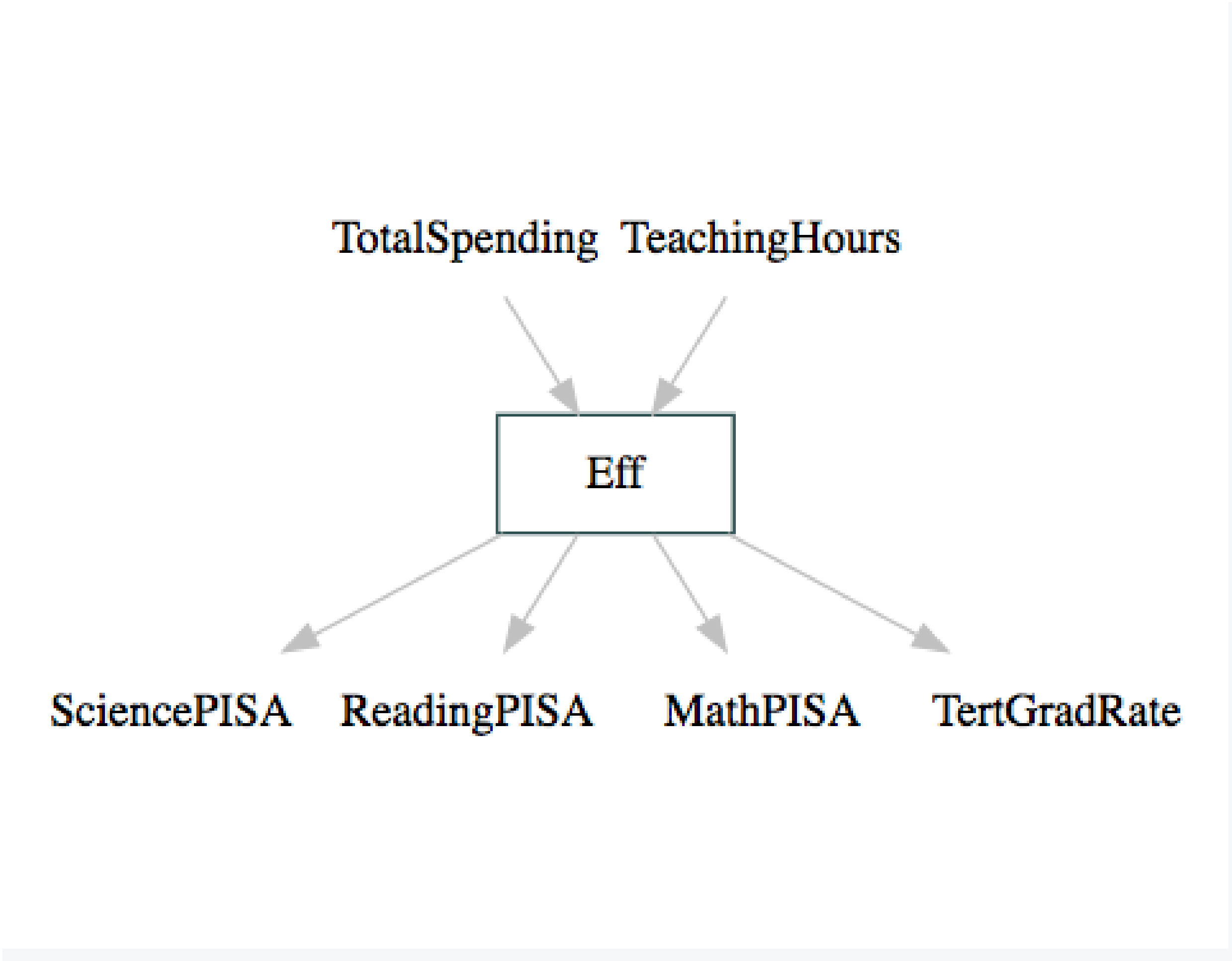


Figure 2: Multiple Input Multiple Output

```
  geom_segment(aes(x = 16691, y = 458.67, xend = 14173, yend = 415.67), color = "dodgerblue") +
  geom_segment(aes(x = 14173, y = 415.67, xend = 14173, yend = 400), color = "dodgerblue") +
  theme_minimal() +
  ggtitle("DEA plot of 2015 OECD Countries",
          subtitle = "Average Education Scores by Total Spending")
# ggsave("single output DEA.png") # uncomment this to update the png file for the presentation
```

Multiple input multiple output model

```
grViz("
digraph nicegraph {

  # a 'graph' statement
  graph [overlap = true, fontsize = 10]

  # several 'node' statements
  node [shape = rectangle, fixedsize = true, width = 1,
        color = darkslategray]
```

```
  Eff

  node [shape = plaintext]
  TotalSpending; TeachingHours;
  SciencePISA; ReadingPISA; MathPISA; TertGradRate

  # 'edge' statements
  edge [color = grey]
  TotalSpending->Eff TeachingHours->Eff Eff->TertGradRate
  Eff->SciencePISA Eff->ReadingPISA Eff->MathPISA
}
")
```

Multiple input multiple output DEA plot

```
dea.plot(x = x, y = y, txt = dimnames(x)[[1]], GRID = T,
         xlab = "Total Spending + Teaching Hours",
         ylab = "Multiple Outputs",
         ylim = c(1250,1750), main = "2015 OECD data DEA plot")
```

Multiple input multiple output ggplot code

```
# warning, the geom_segment calls are not reproducible and must be hand specified
DEAdata$multIn <- DEAdata$Total.Spending + DEAdata$Teaching.Hours

DEAdata$multOut <- DEAdata$SciencePISA + DEAdata$ReadingPISA + DEAdata$MathematicsPISA +
  DEAdata$Tertiary.Graduation.Rate


ggplot(DEAdata, aes(x = multIn, y  = multOut)) +
   geom_point() +
   geom_text(aes(label = Countries), size = 3,
             nudge_x = 2, nudge_y = -2, color = "gray40") +
   geom_segment(aes(x = 91000, y = 1681, xend = 40162, yend = 1681), color = "dodgerblue") +
   geom_segment(aes(x = 40162, y = 1681, xend = 22910, yend = 1664), color = "dodgerblue") +
   geom_segment(aes(x = 22910, y = 1664, xend = 17298, yend = 1470), color = "dodgerblue") +
   geom_segment(aes(x = 17298, y = 1470, xend = 15071, yend = 1304), color = "dodgerblue") +
   geom_segment(aes(x = 15071, y = 1304, xend = 15071, yend = 1250), color = "dodgerblue") +
   theme_minimal() +
   labs(xlab = "Total Spending + Teaching Hours",
        ylab = "SciencePISA + ReadingPISA + MathematicsPISA + Tertiary Graduation Rate") +
   ggtitle("DEA plot of 2015 OECD Countries",
           subtitle = "Multiple Input and Multiple Output")
#ggsave("multiple output DEA.png") # uncomment this to update the png file for the presentation
```

Additional code References

```
# Rich Iannone's DiagrammR package code example from
# http://rich-iannone.github.io/DiagrammeR/graphviz_and_mermaid.html#mermaid
grViz("
digraph boxes_and_circles {

  # a 'graph' statement
  graph [overlap = true, fontsize = 10]
```

```
  # several 'node' statements
  node [shape = box,
        fontname = Helvetica]
  A; B; C; D; E

  node [shape = circle,
        fixedsize = true,
        width = 0.9] // sets as circles
  1; 2; 3; 4; 5; 6; 7; 8

  # several 'edge' statements
  A->1 B->2 B->3 B->4 C->A
  1->D E->A 2->4 1->5
  E->6 4->6 5->7 6->7 3->8
}
")
```