ETM 540 Group Project

Samira Akther
Jordan Hilton
Andey Nunes
Aparna Raghuram
Mohammed Sheikh

February 27, 2019

Efficiency of Public Expenditure on Education

In this section, the paper aims to study the efficiency of the public expenditures on education of OECD (Organization for Economic Development) countries. DEA (Data Envelopement Analysis), which is a vastly used tool for measuring efficiency, has been used to measure the efficiency of the public expenditures on education by the countries.The analysis provides an efficient frontier which is a linear combination of the data point with efficiency scores of 1. The analysis also allows us to know possible reduction in the inputs to generate the output of the countries with scores less than 1.

Optimization Model Definition

In our paper, we used input oriented model. Here, only one input is considered and defined as $x_{i,j}$ to be the amount of the i'th input used by unit j . Here j is the country we are analysing. Two outputs have been considered and defined as $y_{r,j}$ to be the amount of the r'th output produced by unit j. n is the number of countries we analysed.

The following optimization model has been solved to obtain the efficiency score while minimizing the value of

$$\begin{aligned} &\text{Min } \theta \\ \text{subject to } &\sum_{j=1}^{n} x_{i,j}\lambda_j \leq x_{i,k} \forall\, i \quad i = 1, \ldots, n \\ &\sum_{j=1}^{N^D} y_{r,j}\lambda_j \geq y_{r,k} \forall\, r \quad j = 1, \ldots, r \\ &\lambda_j \geq 0 \;\forall\, j \end{aligned}$$

Here, Theta provides the efficiency scores of the countries we analyzed. The country with score of 1 lies in the efficient frontier and shows that any input reduction is not possible for generating the output. The countries with scores less than 1 will be considered as inefficient within the efficient frontier. Vector Lambda he specific amount of a unit j used in setting the target for for performance for unit k.

Data Description

Data has been obtained from secondary sources. Data on PISA scores were obtained from OECD website. Data on public education expenditures as percentage of GDP were obtained from UNESCO website. We considered 36 countries of OECD. The input variable is the public education spending as percentage of GDP of all 36 countries for the year of 2015. The output variables are the average PISA score of all 36 countries and the graduation rates for the year of 2015.

Here is a summary of the data.

```
## load our data, downloaded from the google sheet
data <- read.csv("OECDdata.csv")

summary(data)
```

```
##           Countries       GDP            SciencePISA     ReadingPISA
##  Australia     : 1   Min.   :   16.94   Min.   :416.0   Min.   :423.0
##  Austria       : 1   1st Qu.:  184.50   1st Qu.:480.0   1st Qu.:481.8
##  Belgium       : 1   Median :  384.40   Median :497.0   Median :498.0
##  Canada        : 1   Mean   : 1298.73   Mean   :492.7   Mean   :492.0
##  Chile         : 1   3rd Qu.: 1235.75   3rd Qu.:509.0   3rd Qu.:506.8
##  Czech Republic: 1   Max.   :18120.00   Max.   :538.0   Max.   :527.0
##  (Other)       :30
##  MathematicsPISA  AveragePISA         HDI         Primary.education
##  Min.   :408.0   Min.   :415.7   Min.   :0.7670          : 3
##  1st Qu.:481.0   1st Qu.:482.7   1st Qu.:0.8605   10,211 : 1
##  Median :494.0   Median :498.0   Median :0.8990   10,853 : 1
##  Mean   :491.0   Mean   :491.9   Mean   :0.8887   11,047 : 1
##  3rd Qu.:510.2   3rd Qu.:508.0   3rd Qu.:0.9200   11,215 : 1
##  Max.   :532.0   Max.   :528.7   Max.   :0.9480   11,630 : 1
##                                  NA's   :1        (Other):28
##  Secondary.education Tertiary.education Total.Spending
##         : 3                  : 3         0      : 3
##  10,111 : 1          0      : 1         14,173 : 1
##  10,383 : 1          10,109 : 1         16,546 : 1
##  10,482 : 1          10,137 : 1         16,691 : 1
##  10,569 : 1          10,208 : 1         18,400 : 1
##  11,147 : 1          10,891 : 1         19,720 : 1
##  (Other):28          (Other):28         (Other):28
##  Tertiary.Graduation.Rate Bachelors.Graduation.rates    Masters
##  missing: 5               missing: 5                 missing: 5
##  101.045: 1               25.013 : 1                 10.181 : 1
##  31.35  : 1               25.523 : 1                 10.242 : 1
##  56.812 : 1               26.151 : 1                 11.777 : 1
##  75.206 : 1               28.395 : 1                 15.074 : 1
##  76.361 : 1               29.597 : 1                 15.28  : 1
##  (Other):26               (Other):26                 (Other):26
##    Doctorates
##  missing: 4
##  0.264  : 1
##  0.466  : 1
##  0.494  : 1
##  0.639  : 1
##  0.712  : 1
##  (Other):27
```

It appears that all but the first seven variables were read in as factors and not numeric values. Also, we can see that there are three observations missing from several of the population education level fields and three zero values for `Total.Spending`. These may need to be excluded from the analysis.

```
# convert factors to numeric
data$Primary.education <- as.numeric(data$Primary.education)
data$Secondary.education <- as.numeric(data$Secondary.education)
data$Tertiary.education <- as.numeric(data$Tertiary.education)
```

```r
data$Total.Spending <- as.numeric(data$Total.Spending)
data$Tertiary.Graduation.Rate <- as.numeric(data$Tertiary.Graduation.Rate)
data$Bachelors.Graduation.rates <- as.numeric(data$Bachelors.Graduation.rates)
data$Doctorates <- as.numeric(data$Doctorates)

# create a spending as a percentage of GDP value
data$spendingPctGDP <- data$Total.Spending/data$GDP*100
```

Other exploratory data information (such as correlations between numerical variables?) is 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 the single GDP data point for the input and the aggregated test score average for the single output.

Run a DEA with a single input (total education spending) and a single output (Average PISA score)

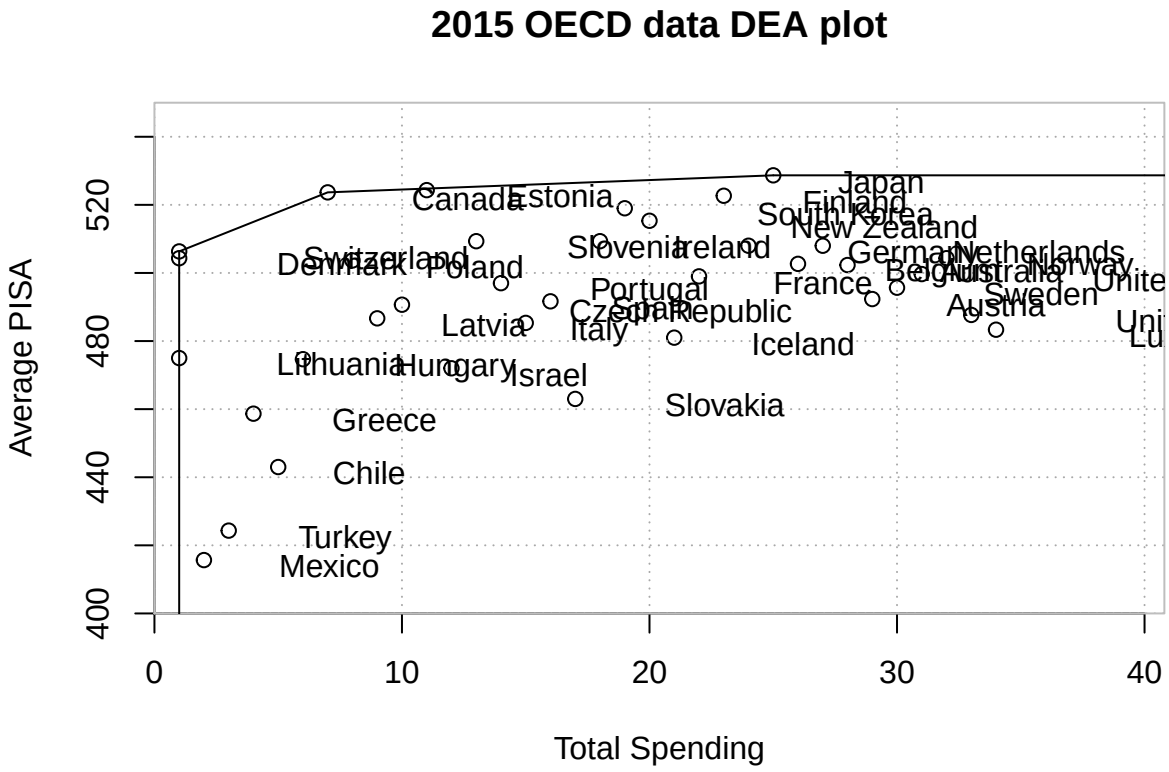
```r
#this is single input and single output
x1 <- data %>% select(Total.Spending) ## input
x2 <- data %>% select(spendingPctGDP) ## ratio input used in next section
y <- data %>% select(AveragePISA) ## output
row.names(x1) <- data[,1]
row.names(y) <- data[,1]


ressingle <- DeaMultiplierModel(x1, y, rts = "crs", orientation = "output")

efficiencyTable <- tibble(
   country = as.character(dimnames(ressingle$Efficiency)[[1]]),
   efficiency = as.numeric(ressingle$Efficiency)) %>%
   arrange(desc(efficiency))

pander(head(efficiencyTable), caption = "DEA Output Efficiency for Aggregate Scores")
```

Table 1: DEA Output Efficiency for Aggregate Scores

country	efficiency
Switzerland	1
Denmark	0.9961
Lithuania	0.9381
Mexico	0.4105
Turkey	0.2794
Greece	0.2265

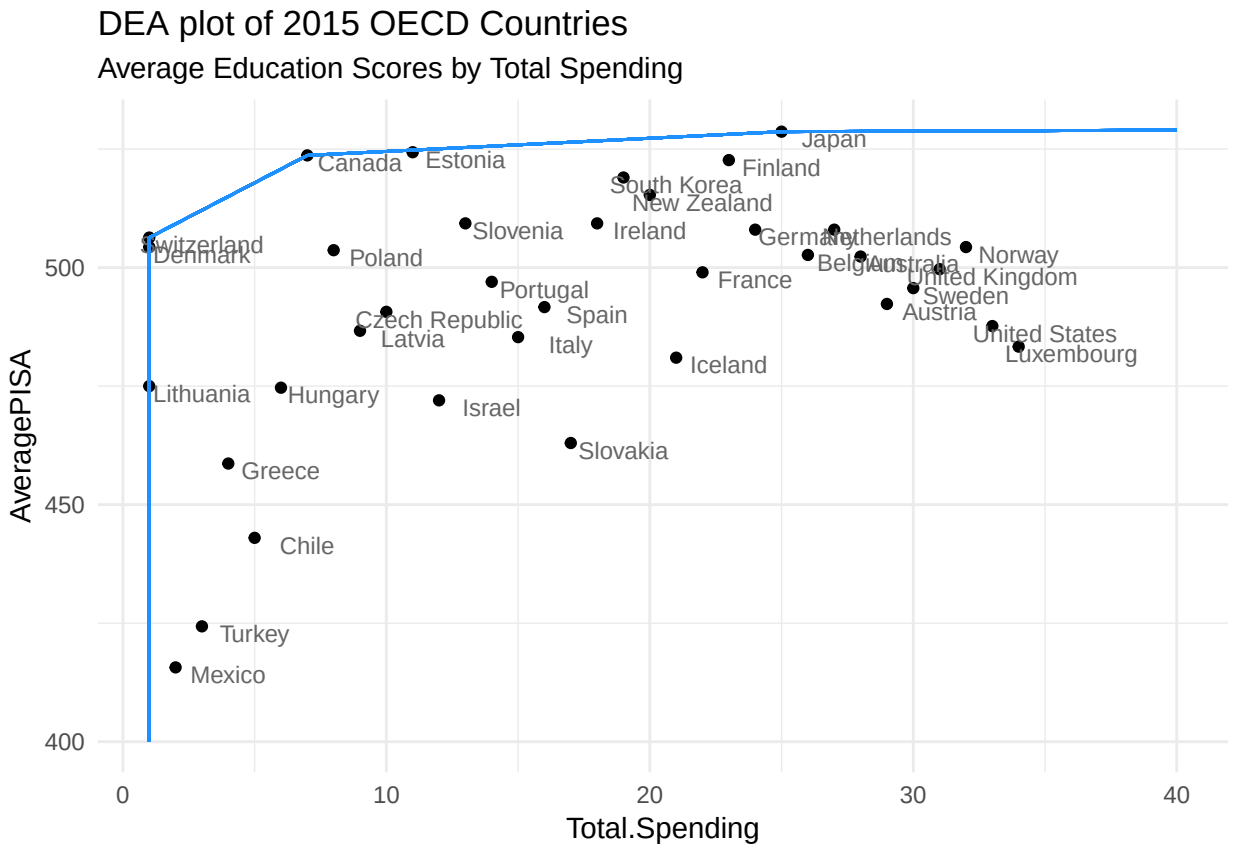
Looks like Switzerland is setting the benchmark for total spending efficiency as reflected by the Average test scores. Denmark is a very close runner up, but then the rest of the list is a bit surprising. Lets look at the graph.

```
dea.plot(x = x1, y = y, txt = dimnames(x1)[[1]], GRID = T,
         xlab = "Total Spending", ylab = "Average PISA",
         ylim = c(400,550), main = "2015 OECD data DEA plot")
```



The overall shape of the graph is somewhat like an anvil cloud, where there are several points that indicate the lower test scores and lower spending values, but quickly the points form a flatter layer what almost looks like a band stretching along the length of the spending axis where test values range from 450 up to the maximum indicated by the frontier boundary. This graph makes it a little harder to see some of the points, so we'll translate the information to a ggplot graph.

```
# warning, the geom_segment calls are not reproducible and must be hand specified
ggplot(data, 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 = 40, y = 529, xend = 25, yend = 528.67), color = "dodgerblue") +
   geom_segment(aes(x = 25, y = 528.67, xend = 7, yend = 523.67), color = "dodgerblue") +
   geom_segment(aes(x = 7, y = 523.67, xend = 1, yend = 506.33), color = "dodgerblue") +
   geom_segment(aes(x = 1, y = 506.33, xend = 1, yend = 504.33), color = "dodgerblue") +
   geom_segment(aes(x = 1, y = 504.33, xend = 1, yend = 400), color = "dodgerblue") +
   #geom_segment(aes(x = 11, y = 524.33, xend = 1, yend = 506.33), color = "lightblue") +
   theme_minimal() +
   ggtitle("DEA plot of 2015 OECD Countries",
           subtitle = "Average Education Scores by Total Spending")
```



Now its clearer to see that Switzerland and Denmark are on the edge of the efficiency frontier at the lower range of total spending. The curve indicates that there is an increase in test scores with an increase in total spending reflected by the positive slope of the curve between Switzerland and Canada and that this nearly flattens from there over Japan's maximum score. The ggplot graph also makes it a little easier to see how far the United States and Luxembourg are from their real world geographical neighbors.

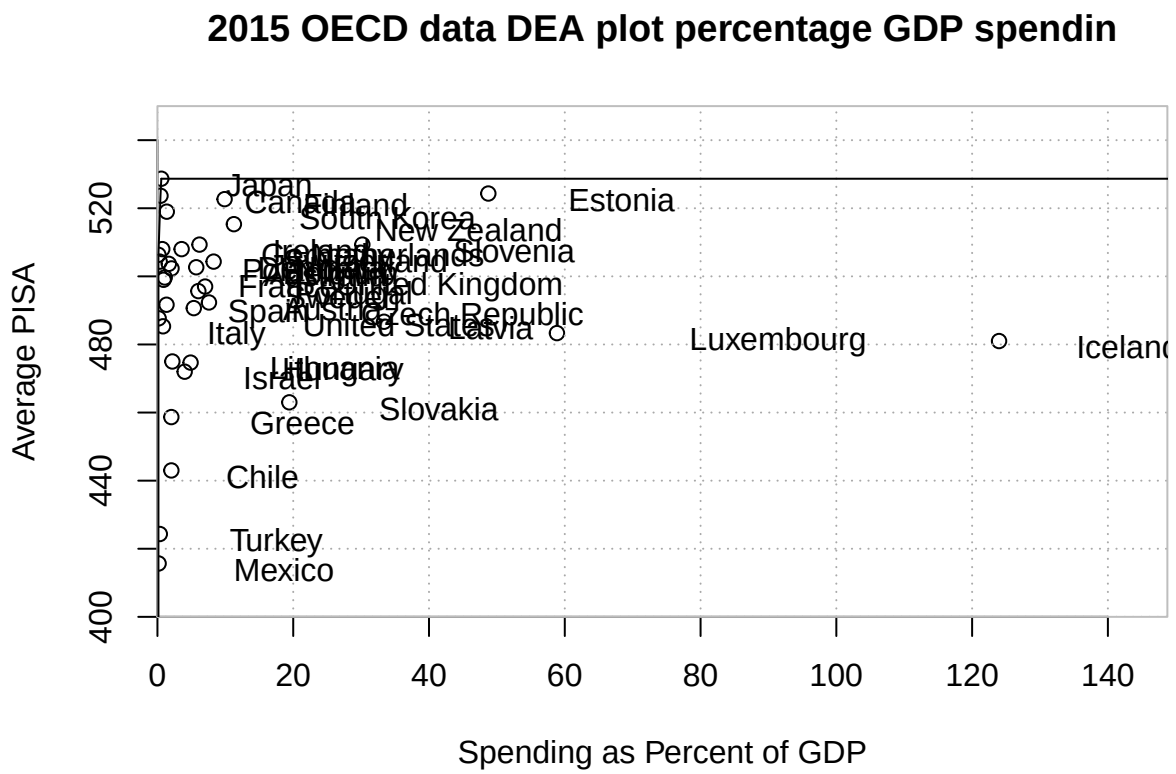
Efficiency of Total Spending as a Percentage of GDP

```
#this is single input and single output
# uses x2 as x and the same y from the previous section
row.names(x2) <- data[,1]

pctGDPressingle <- DeaMultiplierModel(x2, y, rts = "crs", orientation = "output")

pctGDPefficiencyTable <- tibble(
   country = as.character(dimnames(pctGDPressingle$Efficiency)[[1]]),
   efficiency = as.numeric(pctGDPressingle$Efficiency)) %>%
   arrange(desc(efficiency))

pander(head(pctGDPefficiencyTable), caption = "DEA Output Efficiency
       for Total Spending as a percentage of GDP on Aggregate Scores")
```

Table 2: DEA Output Efficiency for Total Spending as a percentage of GDP on Aggregate Scores

country	efficiency
Switzerland	1
United States	0.7785
Mexico	0.707
Denmark	0.4418
Turkey	0.3536
Canada	0.3393

```
dea.plot(x = x2, y = y, txt = dimnames(x2)[[1]], GRID = T,
         xlab = "Spending as Percent of GDP", ylab = "Average PISA",
         ylim = c(400,550), main = "2015 OECD data DEA plot percentage GDP spendin")
```



Something isn't right about this, why is Iceland spending 125% of their GDP on education? Can someone check those data values, is the currency the same?

Part B: Single input, multiple output

For this first part, we will set up and solve a DEA model using the single GDP data point for the input and multiple test scores for outputs.

Run a DEA with a single input, but multiple outputs:

```
#this is single input and multiple output
y <- data %>% select(SciencePISA, ReadingPISA, MathematicsPISA) ## output
row.names(y) <- data[,1]
```

```
resmult <- DeaMultiplierModel(x1, y, rts = "crs", orientation = "output")

multEfficiencyTable <- tibble(
   country = as.character(dimnames(resmult$Efficiency)[[1]]),
   efficiency = as.numeric(resmult$Efficiency)) %>%
   arrange(desc(efficiency))
```

```
pander(head(multEfficiencyTable, 10), caption = "DEA Output Efficency for Multiple Scores")
```

Table 3: DEA Output Efficency for Multiple Scores

country	efficiency
Denmark	1
Switzerland	1
Lithuania	0.9455
Mexico	0.423
Turkey	0.2853
Greece	0.2335
Chile	0.1836
Hungary	0.1578
Canada	0.1506
Poland	0.1265

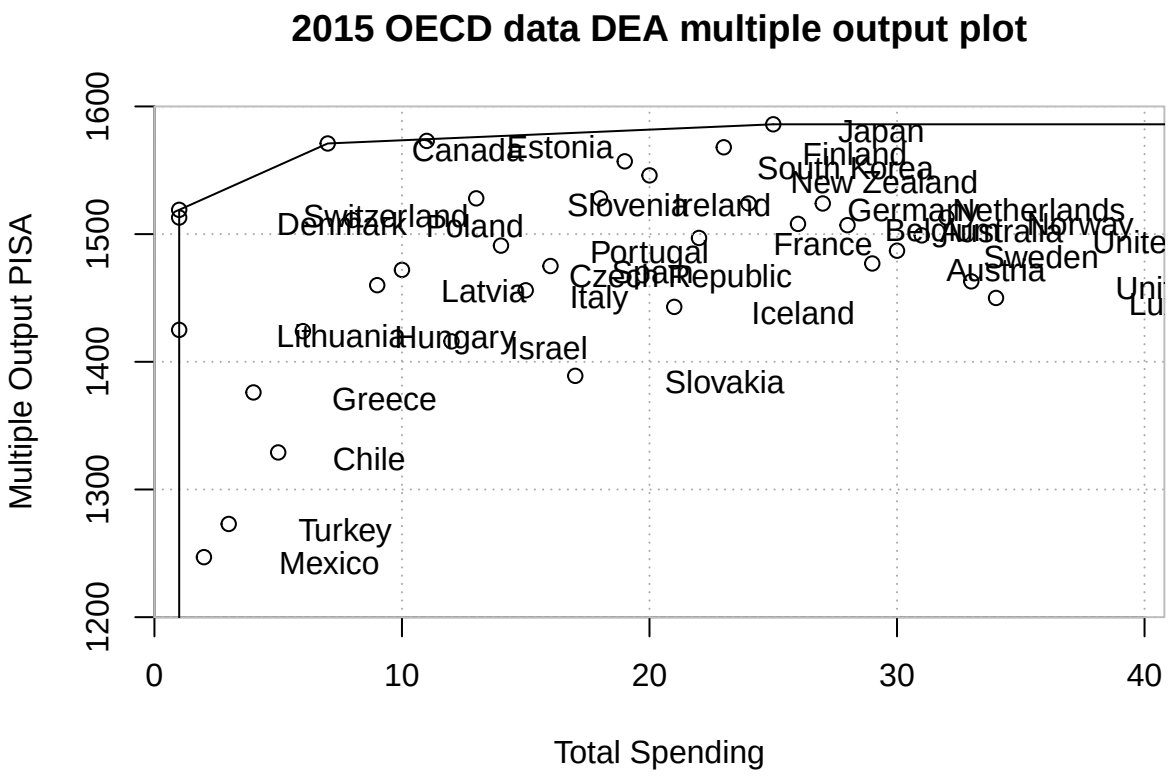
```
pander(head(resmult$uy), caption = "DEA Output Weights for Multiple Scores")
```

Table 4: DEA Output Weights for Multiple Scores

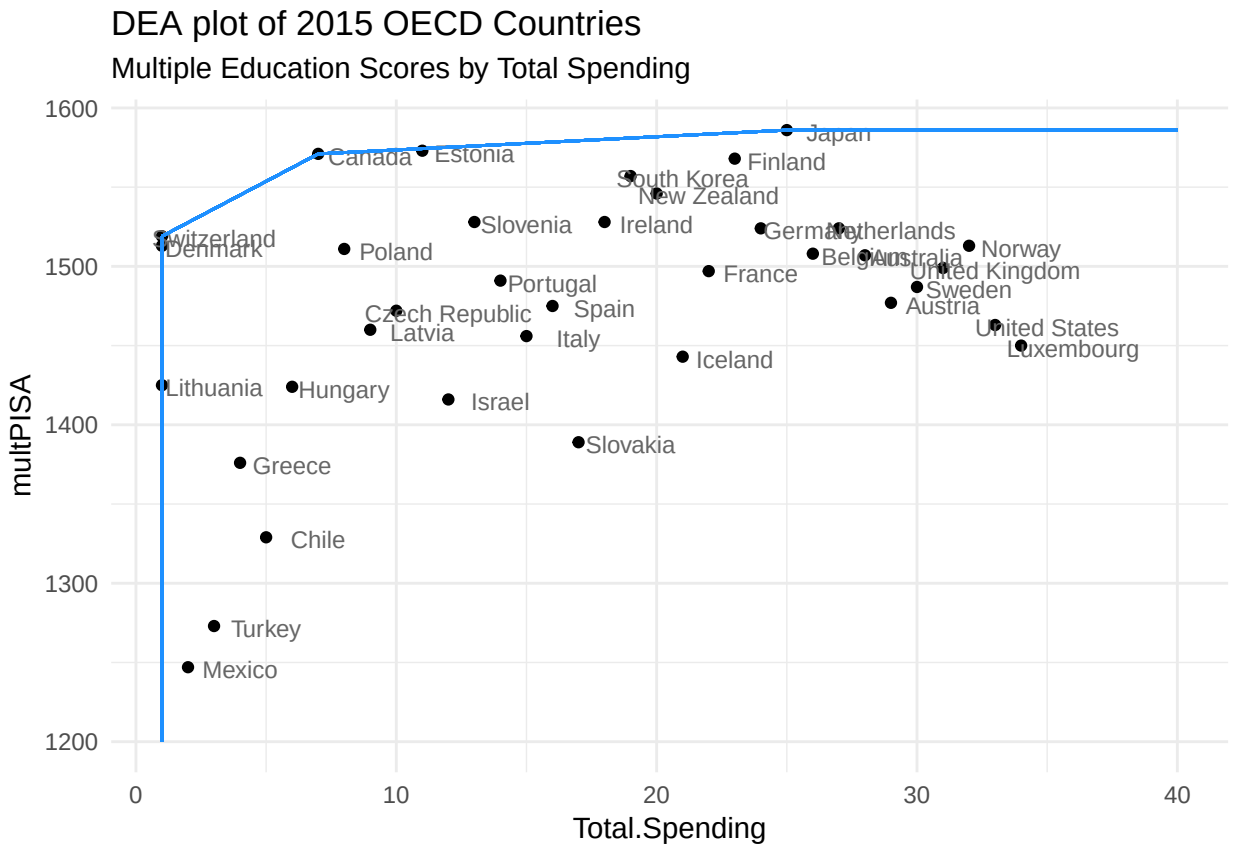
	SciencePISA	ReadingPISA	MathematicsPISA
Australia	0.001313	0.0006566	0
Austria	0.001356	0.000678	0
Belgium	0.001331	0.0006653	0
Canada	0	0.001898	0
Chile	0	0.002179	0
Czech Republic	0.001358	0.0006789	0

Here is the DEA plot for that multiple outputs model.

```
dea.plot(x = x1, y = y, txt = dimnames(x1)[[1]], GRID = T,
         xlab = "Total Spending", ylab = "Multiple Output PISA",
         ylim = c(1200,1600),
         main = "2015 OECD data DEA multiple output plot")
```




```
# warning, the geom_segment calls are not reproducible and must be hand specified
data$multPISA <- data$SciencePISA + data$ReadingPISA + data$MathematicsPISA
data$country <- data$Countries

ggplot(data, aes(x = Total.Spending, y  = multPISA)) +
   geom_point() +
   geom_text(aes(label = Countries), size = 3,
             nudge_x = 2, nudge_y = -1.5, color = "gray40") +
   geom_segment(aes(x = 40, y = 1586, xend = 25, yend = 1586), color = "dodgerblue") +
   geom_segment(aes(x = 25, y = 1586, xend = 7, yend = 1571), color = "dodgerblue") +
   geom_segment(aes(x = 7, y = 1571, xend = 1, yend = 1519), color = "dodgerblue") +
   geom_segment(aes(x = 1, y = 1519, xend = 1, yend = 1200), color = "dodgerblue") +
   theme_minimal() +
   ggtitle("DEA plot of 2015 OECD Countries",
           subtitle = "Multiple Education Scores by Total Spending")
```



Part C: Results and discussion

Some quick analysis of the results:

```
max(ressingle$Efficiency) ## the maximum efficiency for one output
```

```
## [1] 1
```

```
which.max(ressingle$Efficiency) ## the row value of the maximum efficiency
```

```
## [1] 33
```

```
data[33, ] ## the row from the data matching the answers above
```

```
##      Countries   GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA
## 33 Switzerland 679.3         506         492             521      506.33
##      HDI Primary.education Secondary.education Tertiary.education
## 33 0.942                 1                   1                  1
##    Total.Spending Tertiary.Graduation.Rate Bachelors.Graduation.rates
## 33              1                       32                         23
##    Masters Doctorates spendingPctGDP multPISA     country
## 33  18.201         32      0.1472104     1519 Switzerland
```

```
min(ressingle$Efficiency) ##same for mins
```

```
## [1] 0.02807574
```

```
which.min(ressingle$Efficiency)
```

```
## [1] 22
```

```
data[22,]
```

```
##     Countries   GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA
## 22 Luxembourg 57.78         483         481             486      483.33
##      HDI Primary.education Secondary.education Tertiary.education
## 22 0.899                11                  19                 29
##    Total.Spending Tertiary.Graduation.Rate Bachelors.Graduation.rates
## 22             34                       8                        31
##    Masters Doctorates spendingPctGDP multPISA    country
## 22    7.85          8       58.84389     1450 Luxembourg
```

```
ressingle$Efficiency["United States",] ## US efficiency
```

```
## [1] 0.02918626
```

References

Data Sources

insert links/doi to data sets/info here

Anderson, T. R. (2019) Operations Research in R

Anderson, T. R. (2019) Data Envelope Analysis in R

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

Appendix