Efficiency of Schools in the Visegrad Region, a StoNED Approach

BENNOUR Mohamed Hsin, University of Pécs

Abstract

The subject of school efficiency has attracted lots of literature attention, as school-level inefficiencies can have serious negative consequences on the development of human capital and the labor market. In this paper, we examine the efficiency of a sample of a combined 686 schools in the Visegrad countries using the 2018 PISA survey data. We employ an innovative nonparametric approach that combines both the benefits of Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA), namely the Stochastic Nonparametric Envelopment of Data (henceforth StoNED) developed in Kuosmanen and Kortelainen (2012). We construct four different models to control for return to scale, contextual (nondiscretionary) input variables, as well as the inclusion of undesirable variables, and accounting for multiple outputs. These models are applied country-wise and across all four countries to compare their efficiency frontiers and derive policy recommendations.

Introduction

In some empirical applications, researchers often seek to analyze the performance of certain Decision-Making Units (hereafter as DMU, eg: a school). These DMUs are often characterized by a production process that takes a certain level of resources (or inputs) and transforms them to a certain level of goods (or outputs). In this context researchers model productivity and efficiency as metrics of performance, by analyzing the production process which yields a production frontier. The frontier describes the feasible level of outputs for each level of inputs, this frontier reflects the state of technology for a given sector or industry.

Educational institutions can be considered (from an economic point of view) as production units (DMUs), and can be evaluated in terms of efficiency. Efficiency can be analyzed from various perspectives. School-level inefficiencies and mismanagement can have serious repercussions for human resource development and labor market outcomes as a recent study (**Bhutoria2022?**) highlights.

There are many papers that are measuring the efficiency within the sector of education, see for example the excellent review of (**Witte2017?**). To the best of our knowledge, no paper among these concentrates on the Visegrad (V4) countries, although the countries are historically linked and followed similar development paths and it seems useful to compare their education policies and performance (CITE: https://www.researchgate.net/publication/308994770_Education_policies_and_performance_of_the_Visegra
In this paper we focus on answering the question of what drives the efficiency of schools in the Visegrad countries in terms of standardized test performance? The efficiency analysis will be based on PISA 2018 data and the novel StoNED framework.

Methods

The early literature provided two major directions in performance analysis. The first is the "Stochastic Frontier Analysis" (SFA). The second direction is the non-parametric approach, leading to the development of "Data Envelopment Analysis" (DEA) and "Full Disposal Hull" (FDH). DEA and FDH gained a large theoretical and empirical literature grounds since they were introduced (DEA and FDH are considered the same technique, they are usually both referred to as DEA).

Both approaches have clear advantages and disadvantages. The main shortcoming of DEA is that all deviations from the frontier are treated as inefficiency. The strength of SFA is its probabilistic nature. It allows for decomposing deviations into an inefficiency term and noise term. However development on convex nonparametric least squares (CNLS) have led to the full integration of DEA and SFA into a unified framework of productivity analysis, which we refer to as Stochastic Nonparametric Envelopment of Data (StoNED). (**kous2015?**)

Formulation of a StoNED model requires identifying input, output and enviromnetal variables, and solving an optimization algorithm, depending on various possible assumptions.

SOME UGLY ASS EQUATIONS MAYBE?

Creators of the StoNED framework provide GAMS and MATLAB code originally, there exists a Python package (REFERENCE) for implementing different models.

In the next chapter we use nonparametric envelopment techniques to estimate a production (education) frontier for each Visegrad country, and for the region as a whole (metafrontier what he called it?).

Data

Data is extracted from the 2018 PISA survey, it contains information on schools (and their environments), teachers, and students (and parental engagement). Data analysis and exploration, visualisation was done in R, while the frontiers were estimated in Python.

Summary statistics of the data

Our sample is constituted of 623 schools decomposed per the table below.

CNT	n
CZE	187
HUN	99
POL	209
SVK	128

The variables we initially decided to use are the following:

- Students:
  - `total_sudents`
  - gender characteristics : `students_male (number of),``students_female (number of)`
  - Parents: `ap_parental_eng` (proportion of parents who engage with the school staff regarding their children education)
- Schools:
  - School characteristcs: `ratio_f2m (ratio of female to male students), ratio_pt2ft (part time to full time tachers), ratio_t2s (students per teacher), dropout_rate, funding (percentage of government funding)`
- Teachers:
  - `teachers_ft`(number of full timer teachers), `teachers_pt` (number of full timer teachers)
  - `total_teachers` : Total number of teachers
- Context:
  - Location `bol_locataion` : 1- school is in Urban area, 0- School is in Rural area
  - Extra activities : `bol_extra_acts` : 1- Yes there are official extra activities in the curricilum, 0- No
  - `bol_competition` : 1- There's at least another school in the area, 0- There's no other schools in the area
  - `bol_career_guidance` : 1- There's an official career guidance curriculum in the school, 0- There's not

Mean and standard deviation tables for each variable per country is depicted as follows:

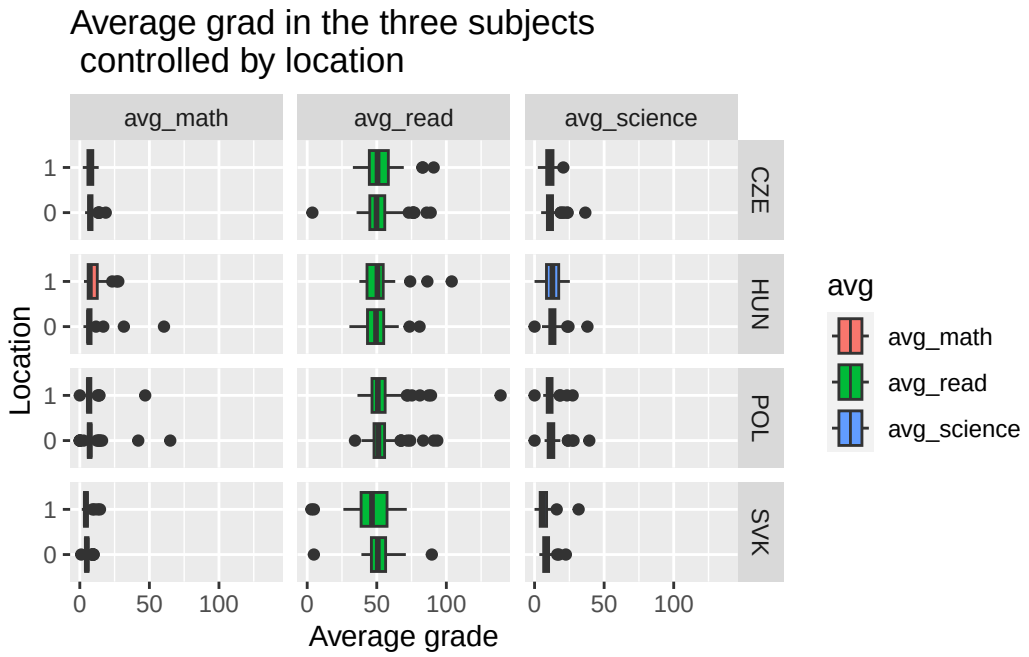
Table 1: Table of means

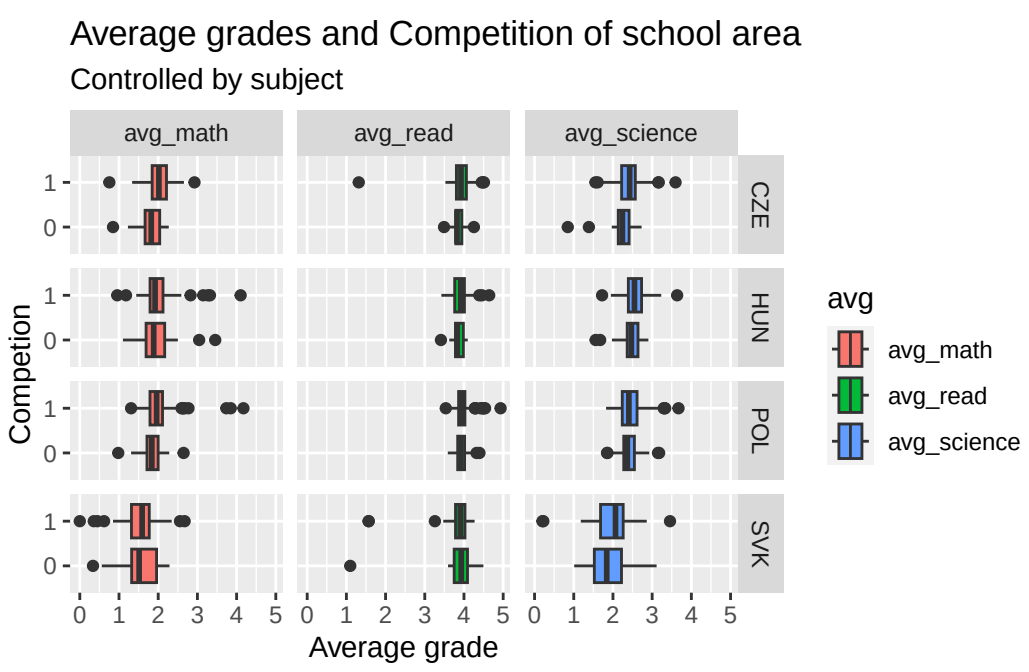
name	CZE	HUN	POL	SVK
bol_location	0.369	0.212	0.469	0.391
bol_extra_acts	0.989	0.990	1.000	1.000
bol_competition	0.882	0.788	0.732	0.859
bol_career_guidance	0.717	0.616	0.383	0.523
funding	0.946	0.974	0.971	0.969
students_male	243.717	270.212	181.134	196.930
students_female	208.695	244.970	173.330	206.289
teachers_ft	34.545	46.020	38.943	29.148
teachers_pt	8.481	6.586	9.191	5.445
total_students	452.412	515.182	354.464	403.219
total_teachers	43.027	52.606	48.134	34.594
ratio_f2m	1.301	1.321	1.078	1.301
ratio_pt2ft	0.342	0.176	0.780	0.217
ratio_ft2pt	7.843	16.056	8.542	8.678
ratio_t2s	0.102	0.119	0.194	0.092
ratio_s2t	10.932	9.620	7.126	11.756
ap_parental_eng	0.256	0.212	0.376	0.284
dropout_rate	0.052	0.033	0.006	0.013
avg_math	7.735	8.602	7.645	5.145
avg_read	52.010	50.511	53.278	49.529
avg_science	11.392	13.003	11.812	7.920

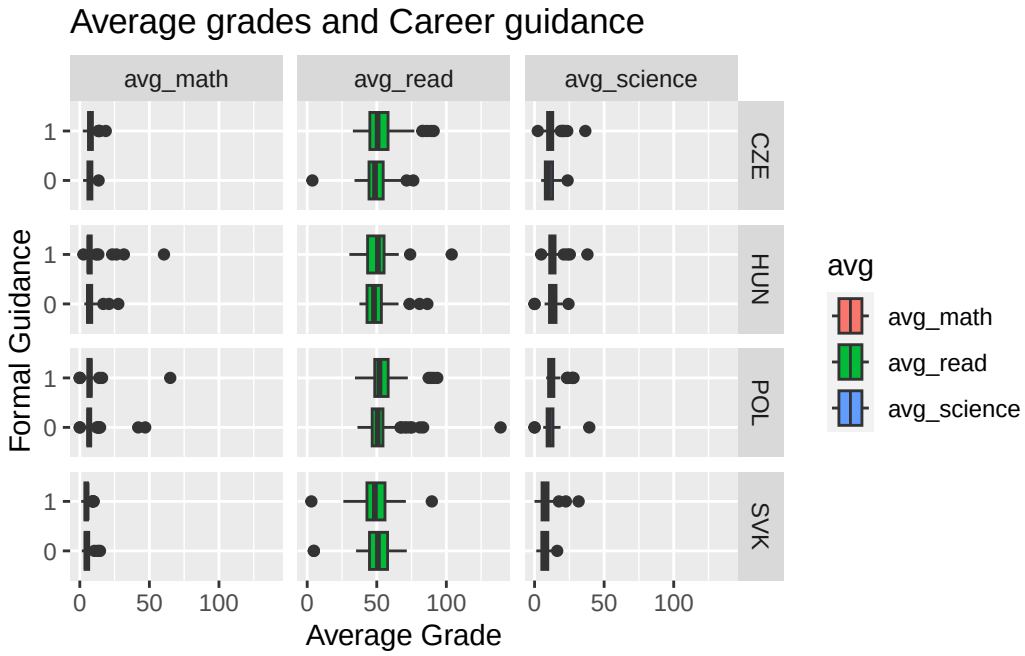
Table 2: Table of standard deviation

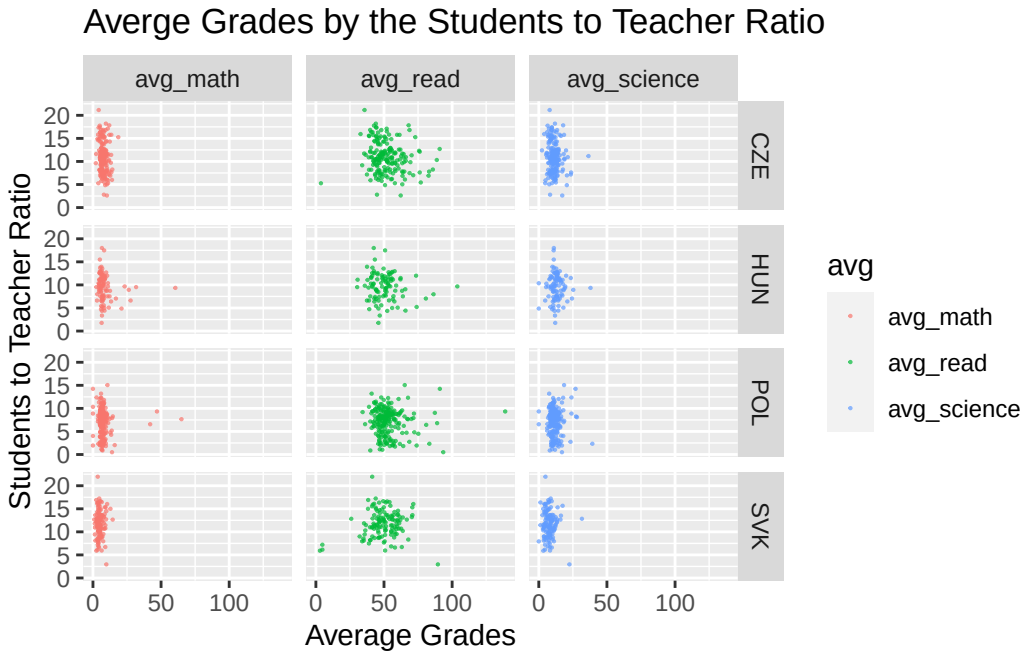
name	CZE	HUN	POL	SVK
bol_location	0.484	0.411	0.500	0.490
bol_extra_acts	0.103	0.101	0.000	0.000
bol_competition	0.323	0.411	0.444	0.349
bol_career_guidance	0.452	0.489	0.487	0.501
funding	0.101	0.077	0.090	0.068
students_male	183.576	230.992	141.259	117.215
students_female	134.320	170.437	130.440	128.833
teachers_ft	20.197	22.383	25.130	13.879
teachers_pt	8.922	6.830	7.776	4.913
total_students	232.565	290.519	261.044	196.683
total_teachers	22.588	24.300	25.650	16.051
ratio_f2m	1.337	1.214	0.936	1.106
ratio_pt2ft	0.576	0.227	2.622	0.184
ratio_ft2pt	8.047	17.792	11.112	7.800
ratio_t2s	0.042	0.062	0.192	0.032
ratio_s2t	3.313	2.897	2.844	2.883
ap_parental_eng	0.137	0.158	0.163	0.162
dropout_rate	0.072	0.057	0.035	0.036
avg_math	2.392	7.116	5.872	2.211
avg_read	11.115	10.893	11.015	11.774
avg_science	3.935	4.878	4.206	3.970

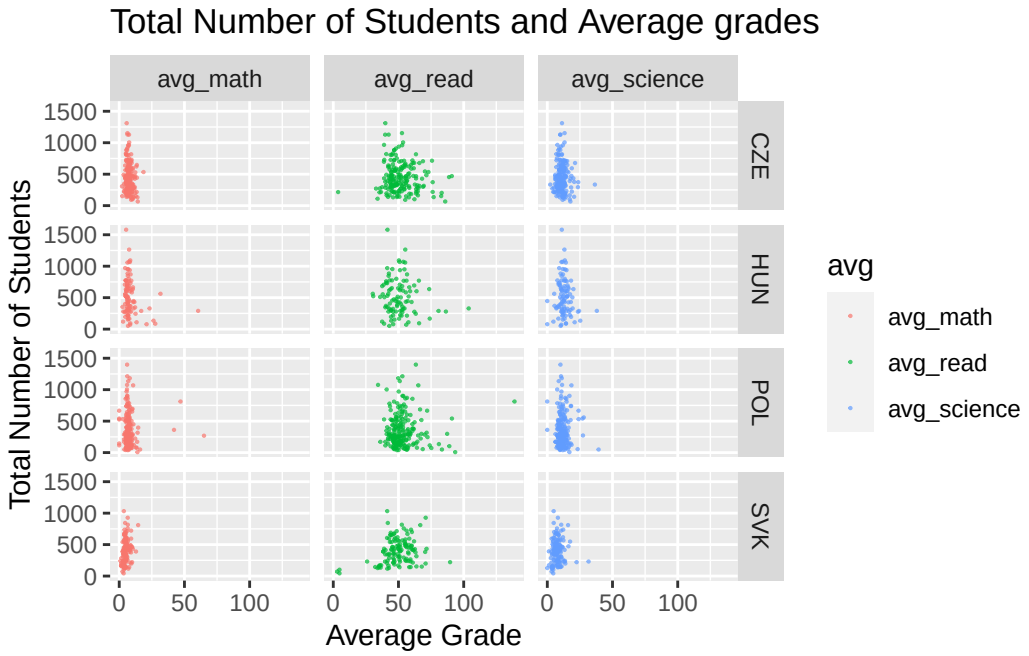
Visualisations

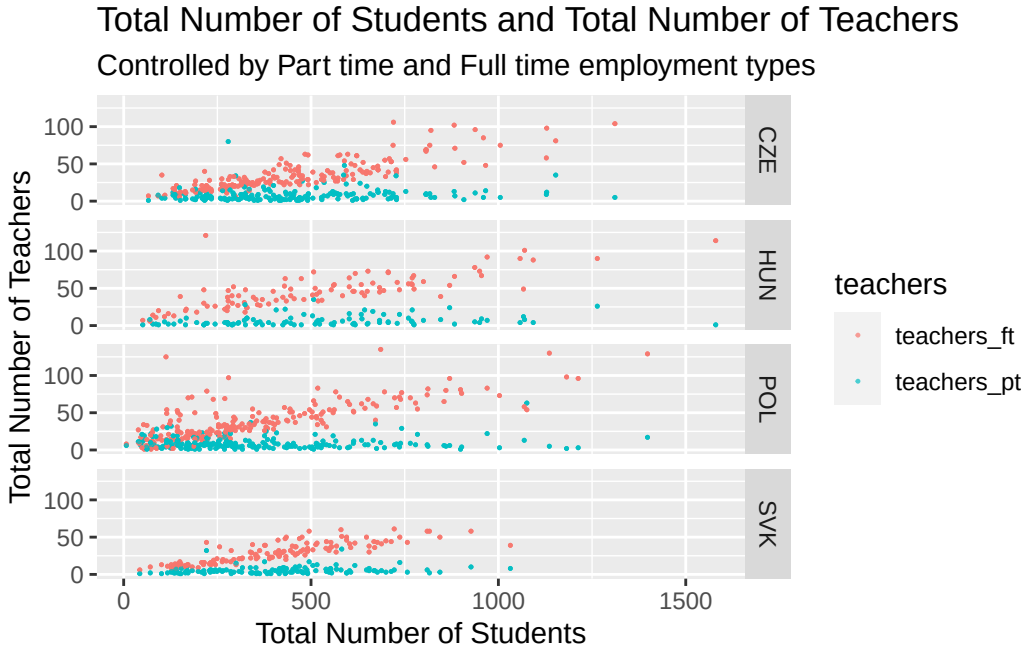


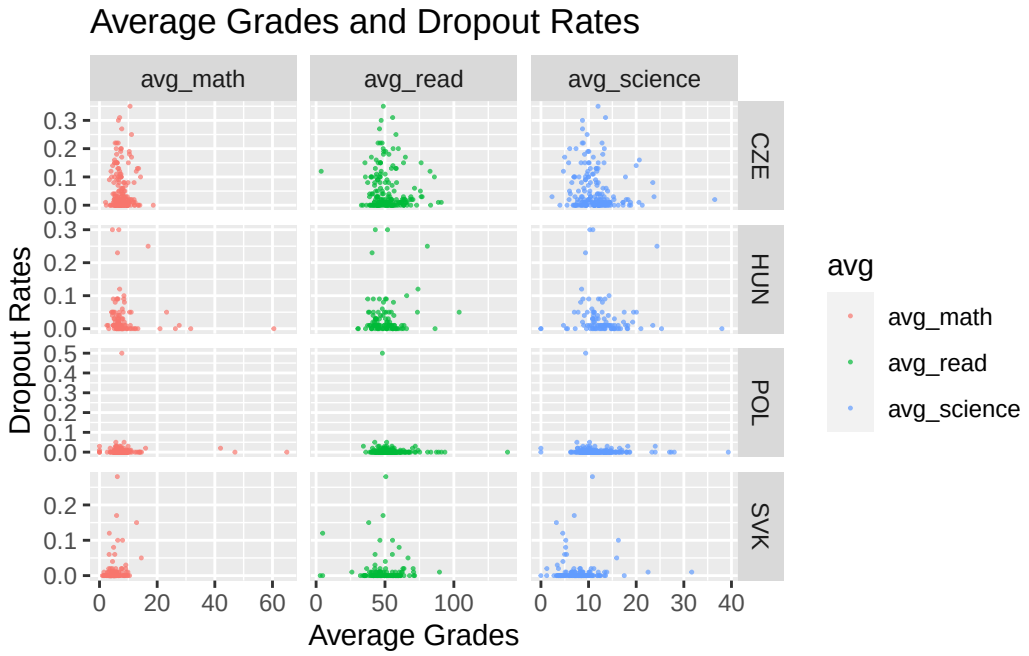


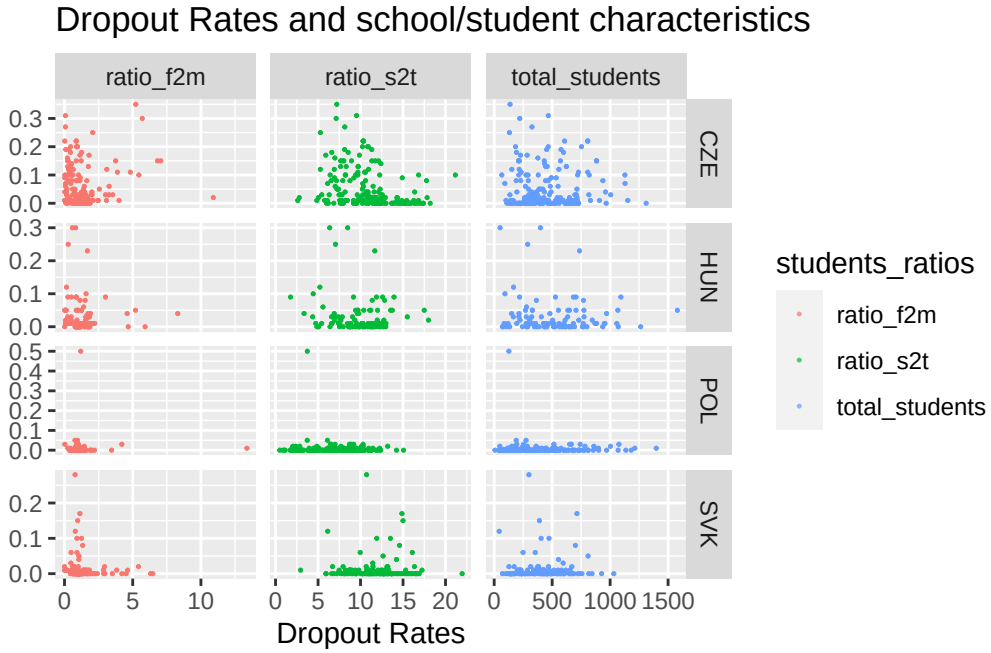

Average grades and Career guidance
avg_math
avg_read
avg_science
CZE
HUN
POL
SVK
Formal Guidance
1
0
0
50
100
Average Grade
avg
avg_math
avg_read
avg_science

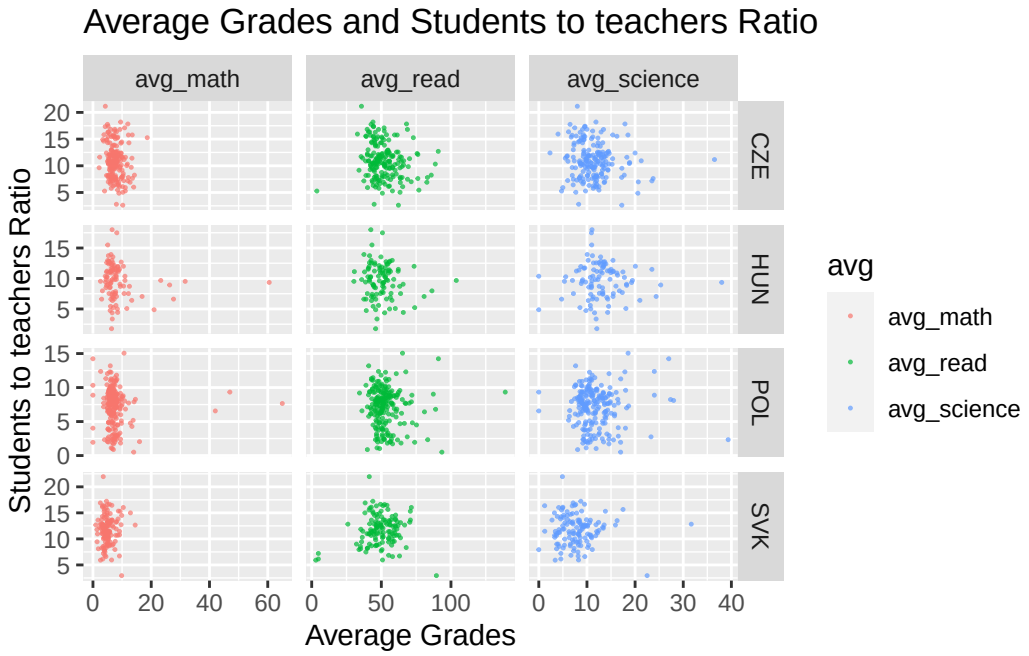

Averge Grades by the Students to Teacher Ratio
avg_math
avg_read
avg_science
CZE
HUN
POL
SVK
Students to Teacher Ratio
20
15
10
5
0
0
50
100
Average Grades
avg
avg_math
avg_read
avg_science


Total Number of Students and Average grades
avg_math
avg_read
avg_science
CZE
HUN
POL
SVK
Total Number of Students
Average Grade
avg
avg_math
avg_read
avg_science


Total Number of Students and Total Number of Teachers
Controlled by Part time and Full time employment types
CZE
HUN
POL
SVK
Total Number of Teachers
Total Number of Students
teachers
teachers_ft
teachers_pt


Average Grades and Dropout Rates
avg_math
avg_read
avg_science
CZE
HUN
POL
SVK
Dropout Rates
Average Grades
avg
avg_math
avg_read
avg_science


Dropout Rates and school/student characteristics
ratio_f2m
ratio_s2t
total_students
CZE
HUN
POL
SVK
Dropout Rates
students_ratios
ratio_f2m
ratio_s2t
total_students



Results and discussion

Kuosmanen, Timo, and Mika Kortelainen. 2012. "Stochastic Non-Smooth Envelopment of Data: Semi-Parametric Frontier Estimation Subject to Shape Constraints." *Journal of Productivity Analysis* 38: 11–28.