



Analyzing International Educational Development Indicators Using K-Means Clustering

Directed Reading Program in Mathematics

Anya Sturm
 Advised by: Wai Ho Chak
 University of California, Davis, CA 95616



The Organization for Economic Cooperation and Development (OECD) compiled a set of over 20 teaching and learning indicators from 38 countries. These indicators measure teachers' and principals' satisfaction with their job, the way they are trained, and the way they behave in the workplace. The goal of this project is to use k-means clustering to analyze differences in these indicators across countries and to see how the differences translate to student learning.

The Data Set

We look at four indicators in this project, each with subcategories. Our indicators are teacher self efficacy, teaching students with special needs, working conditions, and classroom management.

The data set is taken from surveys where teachers and principals self-evaluate, and from statistics such as the amount of professional development and teacher and student demographics.

The K-Means Algorithm

Assignment: Given k number of groups, we randomize k numbers (or means). Each observation is placed in a group that contains a mean (such that the Euclidean distance from that mean to the average value of the observation is minimized).

Update and Iterate: Based on our assigned groups, update the mean of the group, and iterate until the mean doesn't change.

$$1) \quad S_i^{(t)} = \left\{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k \right\}$$

$$2) \quad m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

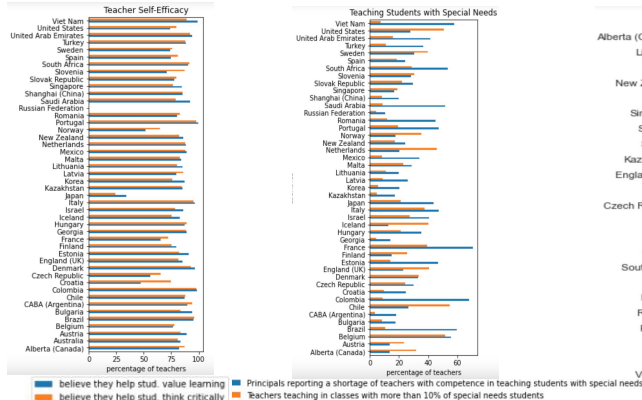
More Advanced/Further Questions

Patterns arise from our heatmap:

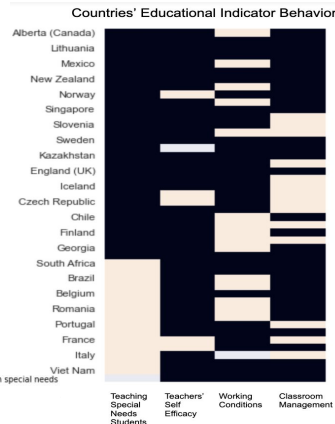
- Mexico, Norway, and Singapore behave similarly in all categories. (Let them be 1 group)
- Brazil, Romania, and Portugal are another group

Our project can be taken further by analyzing student data to see if the correlation on the educators' side matches that of test scores and student surveys.

Individual Indicators Examples



Differences in Country Behavior



We clustered countries by similar behavior for each indicator. In the bar chart, we see sub-indicators for each example, and in the k-means heatmap, we see the overall trend of each country.

Bibliography

Lloyd, Stuart P. "Least squares quantization in PCM." *Information Theory, IEEE Transactions on* 28.2 (1982): 129-137.
 Waskom, M. et al., 2017. *mwaskom/seaborn: v0.8.1 (September 2017)*, Zenodo. Available at: <https://doi.org/10.5281/zenodo.883859>
 Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3), pp.90-95.
 OECD (2022), "Data warehouse", *OECD.Stat* (database), <https://doi.org/10.1787/data-00900-en> (accessed on 07 March 2022).