Visualizing Relationship Between Education and Corruption (2018)

Chang Kai Wen, Ho Zheng Wei, Jonathan Au Wai Chong, Lim Kei Yiang, Shyam Prasad, Tong Huiqing, Glennice (Palegoldenrod)

Introduction

Education and corruption represent critical dimensions in understanding the socioeconomic fabric of nations worldwide. A growing body of research has identified a robust correlation between levels of educational attainment and perceptions of corruption across countries. This relationship suggests that societies with higher average years of schooling tend to exhibit lower levels of corruption, as measured by indices such as Transparency International's Corruption Perception Index. Moreover, studies such as those by Glaeser and Saks (2006)[1] underscore the potential causal link between education and reduced corruption, illustrating how historical educational legacies can influence contemporary governance outcomes. This introduction sets the stage for exploring the nuanced interplay between education and corruption, emphasizing the multifaceted factors that contribute to these complex societal dynamics.
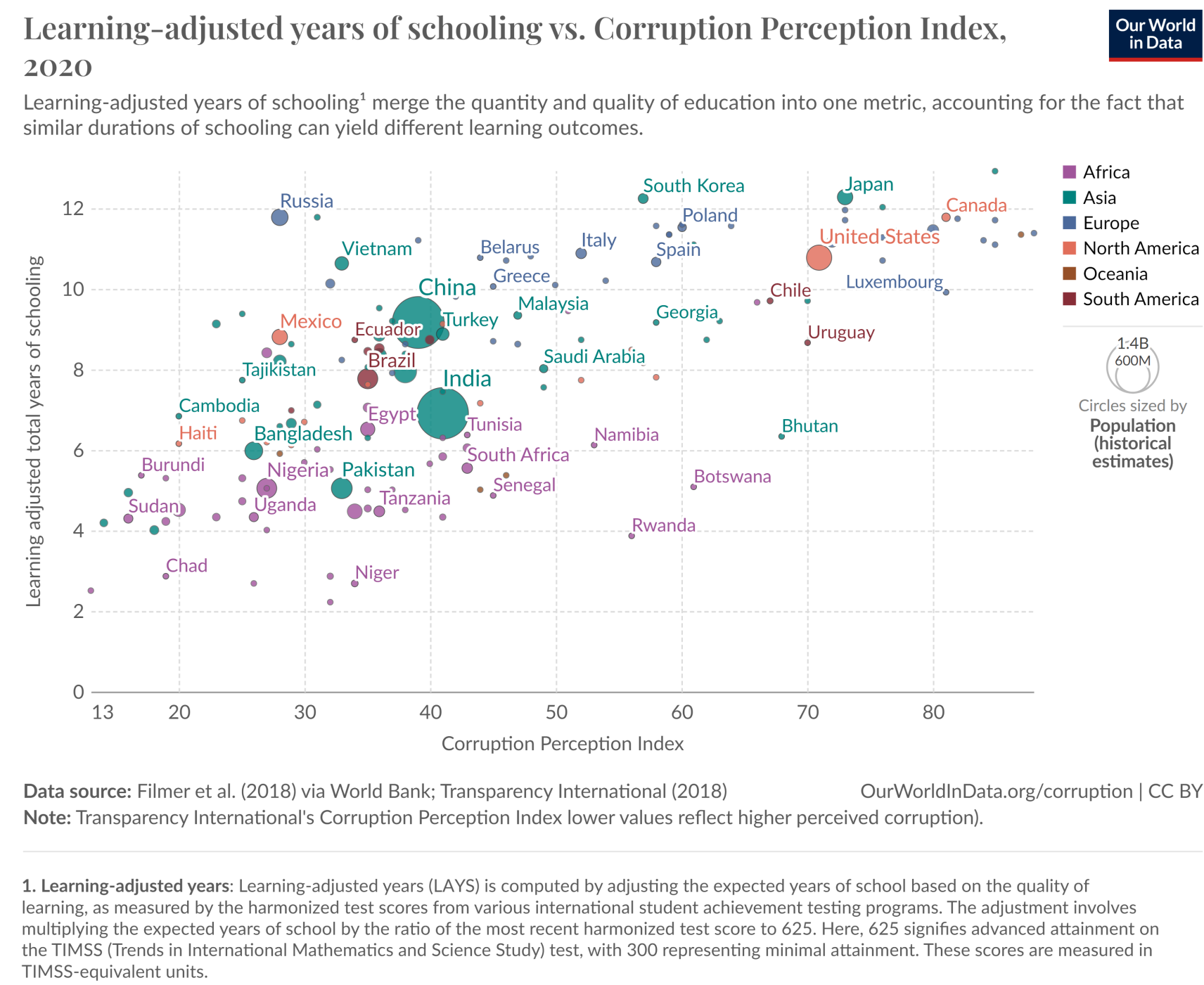
Previous Visualization

Figure 1: Learning-adjusted years of schooling vs. Corruption Perception Index, 2018

Strengths

- Visual representation: The scatter plot effectively visualizes the relationship between learning-adjusted years of schooling and the Corruption Perception Index across various countries and regions.

[1]Glaeser, E. L., & Saks, R. E. (2006). Corruption in America. Journal of public Economics, 90(6), 1053-1072.

- Color coding: The use of different colors for different regions or continents makes it easier to identify and distinguish countries belonging to the same geographical area.
- Clear labeling: The countries and regions are labeled directly on the plot, allowing for easy identification and reference.
- Bubble size: The size of the bubbles (circles) represents the population of each country, providing an additional dimension of information.

Suggested Improvements

1. *The inclusion of smoothing lines (trend lines).* It helps to visualize the general trends and relationships between the schooling index and the Corruption Perception Index within each global region.
2. *Use a discrete color palette.* Continuous palettes can make it challenging for humans to detect patterns with little color differences. Consider using a qualitative ColorBrewer palette instead[2].
3. *Add a legend for population size* providing context about the scale of each country's data.
4. *Divide into more sections* provides a clearer comparison of regions with different socio-economic backgrounds.
5. *Removing some of the lesser-known or less-popular countries* simplifies the visualization, reducing clutter and focusing attention on the more significant data points that are likely to be more familiar and relevant to the audience.
6. *Misleading Title* : Correct the header for the graph to accurately indicate the year the data is sourced from.

Implementation

Data

- The data compares learning-adjusted years of schooling and the 2018 Corruption Perception Index (CPI)[3]. Learning-adjusted years account for both education quality and quantity, recognizing that similar schooling durations can yield different outcomes. The CPI from Transparency International measures perceived corruption, with lower values indicating higher corruption.
- We used data from the top 50 most popular countries[4].
- We edited the schooling index and corruption index to 0-100 to spread out the distribution of the geom points..
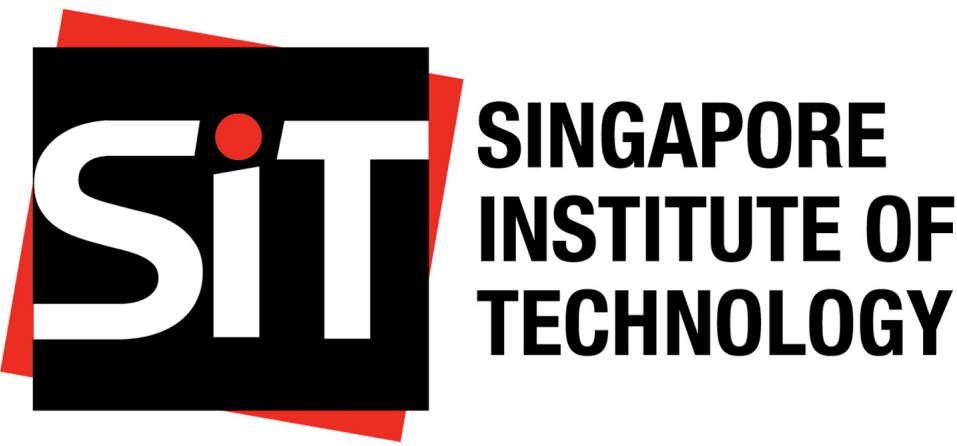
Software

We used the Quarto publication framework and the R programming language, along with the following third-party packages:

- *readxl* for data import
- *tidyverse* for data transformation, including *ggplot2* for visualization, *dplyr* for data manipulation tasks based on the grammar of graphics
- *ggreal* extension for ggplot2 that helps in preventing overlapping text labels in plots.
- *shadowtext* extension for ggplot2 that helps in adding shadow effects to text labels in plots.
- *countrycode* converting country names and codes into various coding schemes.
- *dbscan* for clustering data based on density.
- *FNN* for performing KNN classification.

[2]https://colorbrewer2.org/#type=sequential&scheme=Reds&n=5
[3]https://ourworldindata.org/grapher/average-years-of-schooling-vs-corruption-perception-index
[4]https://worldeducators.org/pages/top-100-countries-for-tourists
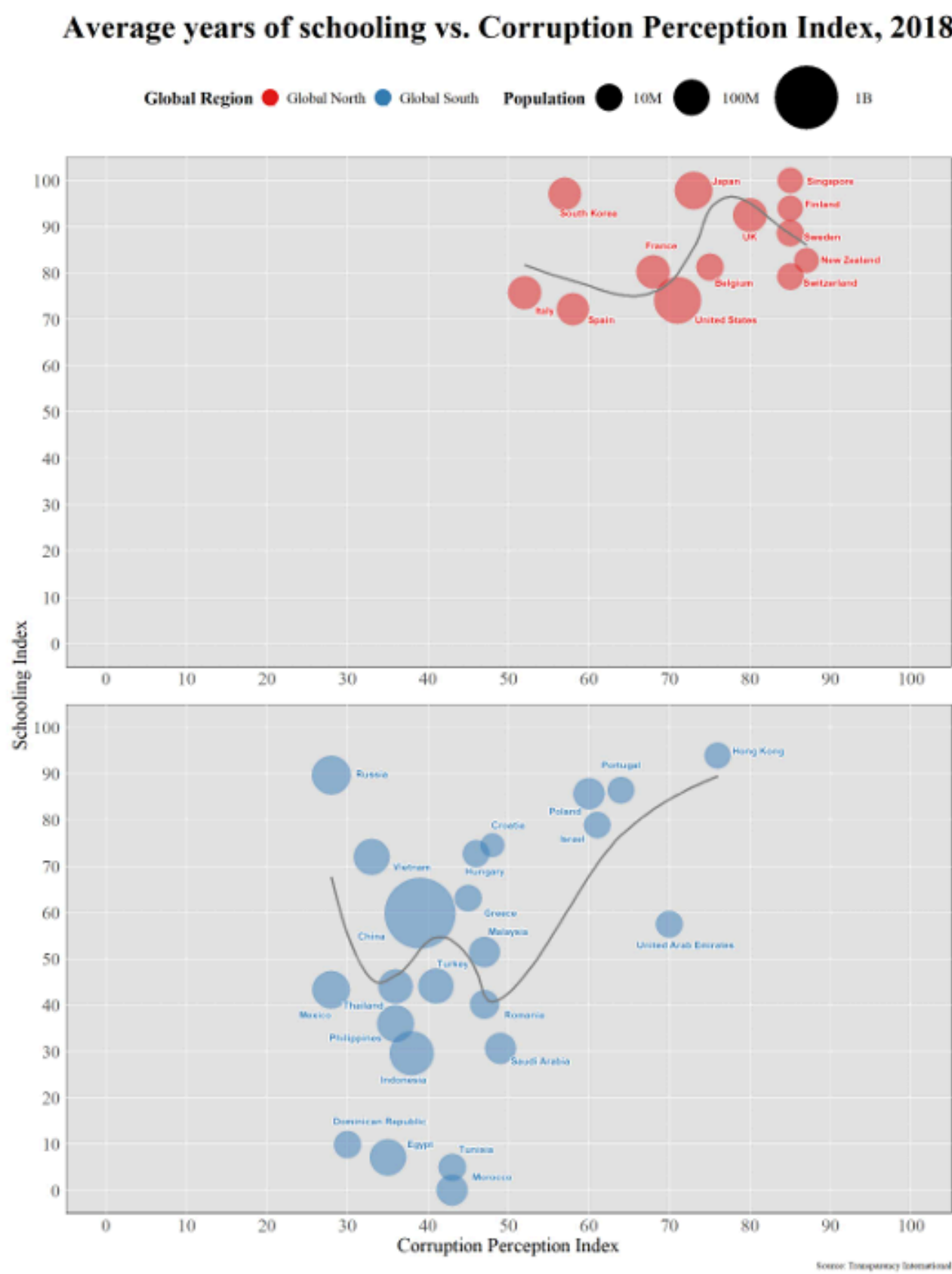
Improved Visualization

Figure 2: Learning-adjusted years of schooling vs. Corruption Perception Index, 2018

Further Suggestions for Interactivity

Because our visualization was intended for a poster, we did not implement any interactive features, including the infotip. However, if the data is visualized in an HTML document, interactive features can be achieved using the R packages such as *plotly*. In that case we recommend customizing the hover information to provide detailed insights when users interact with the data points. This can be done by setting the tooltip parameter in ggplotly().

Conclusion

We successfully implemented all suggested improvements for the non-interactive visualization. By separating global regions into distinct graphs and representing population sizes with bubble sizes, the plots effectively highlight the relationship between average years of schooling and the Corruption Perception Index across various countries.The visual distinction between the Global North and Global South, along with the clear labeling of key countries, allows for easy interpretation of the data.