

Datasheet for Dataset: Global Education Spending and PISA Scores

David James Dimalanta

2024-04-23

This datasheet accompanies the dataset used in analyzing the correlation between global education spending and PISA scores across various countries including Chile, Finland, Canada, and Hong Kong. The focus is on the impact of government spending on education outcomes, particularly in standardized testing contexts such as PISA.

Dataset Description

Motivation

- **Purpose:** The dataset was created to analyze the relationship between government spending on education and academic performance across multiple countries, aimed at understanding how financial investments influence educational outcomes in global standardized tests such as PISA.
- **Creators:** The dataset was compiled by the research team led by David James Dimalanta, associated with the Bachelor of Information at UofT.
- **Funding:** The creation of the dataset was not supported by funding.

Composition

- **Data Instances:** The dataset comprises aggregated educational spending data and PISA scores.
- **Volume:** It contains over 1,000 records spanning multiple assessment years and financial years.
- **Type and Source:** Data includes government spending percentages (sourced from the World Bank) and academic performance metrics (sourced from PISA results).

- **Sampling:** The dataset is a comprehensive aggregation, not a sample. It aims to cover all available data from 2000 onwards for the included countries.
- **Labels:** Each instance is labeled with the country, year, spending percentage, and PISA scores in reading, mathematics, and science.

Collection Process

- **Data Acquisition:** Data was obtained directly from the World Bank databases and PISA's official publications.
- **Collection Methods:** Automated scripts were used to collect and aggregate data annually.
- **Validation:** Data integrity was checked through consistency validations against previous years' datasets.

Preprocessing/cleaning/labeling

- **Preprocessing Steps:** Data was cleaned to remove any incomplete records and normalized to ensure consistency across different years.
- **Software Used:** R and Python scripts, utilizing libraries such as `tidyverse` and `pandas`, were employed for preprocessing.

Uses

- **Primary Uses:** The dataset was primarily used to analyze trends and correlations in education spending and academic outcomes, as detailed in the associated research paper.
- **Intended Uses:** Recommended for academic research and policy analysis in education economics.
- **Restrictions:** Not recommended for commercial use without prior consent from the data curators.

Distribution

- **Access:** The dataset is available upon request from the corresponding author and will be hosted on a dedicated GitHub repository.
- **License:** Distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Maintenance

- **Dataset Maintenance:** Maintained by David James Dimalanta's research team.
- **Contact Information:** For queries or further information, contact david.dimalanta@mail.utoronto.ca.
- **Update Plans:** The dataset will not be updated biennially with new data from subsequent PISA assessments and financial reports.

Legal and Ethical Considerations

- **Compliance:** All data collection and dissemination practices comply with applicable laws and university ethical standards for research.
- **Data Anonymity:** Ensures no individual student or specific schools can be identified from the PISA scores.

This datasheet reflects a commitment to transparency and rigorous academic standards in handling educational data for research purposes.