

Reviewing Assessment Tools for Measuring Country Statistical Capacity

Reproducibility Package for “Reviewing Assessment Tools for Measuring Country Statistical Capacity” paper

Overview

The code in this replication packages constructs the analysis files and tables and figures for Dang, Pullinger, Serajuddin, and Stacy (2024) using R. One main file runs all of the code to generate the data and figures. The file is located in 02_programs/SPI_literature_review_charts.qmd. The replicator should expect the code to run for around 20-30 minutes.

Directory Structure

1. 01_raw_data contains the raw data for the project for each indicator. This folder contains the raw data from the Statistical Performance Indicators (SPI), Open Data Inventory Index (ODIN), Open Data Barometer (ODB), Global Data Barometer (GDB), and other tools, as well as raw data from the World Bank World Development Indicators and several development indices used in the paper. A number of miscellaneous files are included as well that are used.
2. 02_programs contains the main replication file for the project, “SPI_literature_review_charts.qmd”. Execute this file to replicate the results. It also contains another file in ./02_programs/misc/spi_lit_review_data_preparation.Rmd. This file is used to pull the data from the WDI and other data sources and compile the data used to produce the tables. It is not necessary to run this file, as the data is already included in the repository. However, it is included for transparency purposes. If this file is executed, the replication code will no longer replicate, as the data will be overwritten.
3. 03_output_data. This folder contains a number of final output files in either csv or stata .dta format. The most important are SPI_Index_SDG_comparisons_data_2yr_avg.csv and SPI_regression_predictors.csv, which are used to generate the tables and figures in the paper. The other files are produced in the course of the data production, but are not used in the paper. Some of them were used as sensitivity checks, but the results were not included in the paper.

Instructions to Replicators

- Clone the repository to your local machine.
- Please run 02_programs/SPI_literature_review_charts.qmd to generate the data and figures. This file will run all of the code to generate the data and figures. The replicator should expect the code to run for around 20-30 minutes.

- There should be no need to change the working directory. The code should run as is, because the code is using the [here](#) package in R, which automatically handles file paths on local machines. Make sure the .here file is included when you clone the repository.
- This repository contains several files from the R package “renv”. The renv package helps manage specific package versions used to produce the results in this repository. Because package version conflicts can make code that runs on one system not run on another system, it is important to have a list of the specific package versions used and a workflow for accessing these specific packages. The renv package provides this. In order to use renv, see the renv documentation here (<https://rstudio.github.io/renv/articles/renv.html>). In general, the `renv::restore()` command should install all packages found in the `renv.lock` file in this repository, so that version conflicts do not cause errors.

License

The data are licensed under a Creative Commons/CC-BY-4.0 license.

Summary of Availability

- All data **are** publicly available.
- Some data **cannot be made** publicly available.
- **No data can be made** publicly available.

Data Sources

The data used in this analysis comes from a variety of sources. The main sources are the World Bank World Development Indicators (WDI), the Statistical Performance Indicators (SPI), the Open Data Inventory Index (ODIN), the Open Data Barometer (ODB), and the Global Data Barometer (GDB). The data are all publicly available. The data are all included in the repository, so there is no need to download the data from the original sources.

Data Description

There are two main files used in the analysis. Each file will be discussed in turn.

1. SPI_Index_SDG_comparisons_data_2yr_avg.csv
2. SPI_regression_predictors.csv

SPI_Index_SDG_comparisons_data_2yr_avg.csv

This file contains the data used to generate the tables and figures in the paper. It contains the SPI index, the SDG index, and the other variables used in the paper. Many of the indicators are pulled from the WDI. More details can be found in the `spi_lit_review_data_preparation.Rmd` file. The file is in csv format. The file contains the following variables:

Variable	Description
----------	-------------

Variable	Description
SI.POV.DDAY	Proportion of population living below the national poverty line
SI.POV.GINI	Gini index
GE.EST	Government Effectiveness
NY.GDP.PCAP.KD	GDP per capita (2015 constant \$)
HD.HCI.OVRL	Human Capital Index
SN.ITK.DEFC.ZS	Prevalence of undernourishment
SH.STA.MMRT	Maternal Mortality Ratio
SE.LPV.PRIM	Learning Poverty
SG.LAW.IND	Women, Business, Law Index
SH.H2O.SMDW.ZS	Safely Managed Water
EG.ELC.ACCS.ZS	Access to Electricity
NV.IND.MANF.ZS	Manufacturing value added (% of GDP)
EN.POP.SLUM.UR.ZS	Population in Slums
subsidies	Fossil Fuel Subsidies (% of GDP)
EN.ATM.GHGT.KT.CE	Greenhouse Gas Emissions
ER.MRN.PTMR.ZS	Marine protected areas
ER.LND.PTLD.ZS	Terrestrial Protected Areas
DT.TDS.DECT.EX.ZS	Total Debt Service
sdg_index_score	SDG Index Overall Score
hdi_value	Human Development Index
env_perform_index	Environmental Performance Index
eci_value	Economic Complexity Index
press_free_score	Press Freedom Index
better_life_index	OECD Better Life Index
SPI.INDEX	Statistical Performance Indicators Overall Score
SCI	Statistical Capacity Indicator Index
ODIN_score	Open Data Index
odb	Open Data Barometer
gdb	Global Data Barometer

[*SPI_regression_predictors.csv*](#)

This file contains the data used to generate the tables and figures in the paper. It contains severable variables used in the paper for regression analysis. Many of the indicators are pulled from the WDI. More details can be found in the `spi_lit_review_data_preparation.Rmd` file. The file is in csv format. The file contains the following variables:

Variable	Short Description
----------	-------------------

Variable	Short Description
SPI.INDEX	Statistical Performance Indicator (SPI) overall score
SCI	Statistical Capacity Indicator (SCI) overall score
ODIN	Open Data Index (ODIN) overall score
ODB	Open Data Barometer (ODB) overall score
GDB	Global Data Barometer (GDB) overall score
IIAG	Ibrahim Index of African Governance (IIAG) overall score
NY.GDP.PCAP.KD	GDP per capita (constant 2015 US\$)
NV.IND.MANF.ZS	Manufacturing, value added (% of GDP)
NV.AGR.TOTL.ZS	Agriculture, forestry, and fishing, value added (% of GDP)
NE.TRD.GNFS.ZS	Trade (% of GDP)
HD.HCI.OVRL	Human Capital Index (HCI) overall score
SE.PRM.ENRR	School enrollment, primary (% gross)
BN.CAB.XOKA.GD.ZS	Current account balance (% of GDP)
CC.EST	Control of Corruption
GE.EST	Government Effectiveness
PV.EST	Political Stability and Absence of Violence
RQ.EST	Regulatory Quality
RL.EST	Rule of Law
VA.EST	Voice and Accountability
BX.KLT.DINV.WD.GD.ZS	Foreign direct investment, net inflows (% of GDP)
SI.POV.DDAY	Poverty headcount ratio at \$2.15 a day (2017 PPP) (% of population)
SI.POV.GINI	GINI index (World Bank estimate)
sdg_index_score	Sustainable Development Goals (SDG) Index score
hdi_value	Human Development Index (HDI) value
env_perform_index	Environmental Performance Index (EPI) overall score
eci_value	Economic Complexity Index (ECI) value
press_free_score	World Press Freedom Index score
better_life_index	OECD Better Life Index (BLI) overall score
WGI.OVL	World Governance Indicator (WGI) overall score