

Overview

This repository includes instructions and code to replicate estimates of the paper “**A New Global Subnational Poverty and Inequality Database Based on Harmonized Household Surveys**” using Stata. There are three types of data used: (1) subnational data from household surveys and related national/subnational data, (2) national and subnational boundaries.

Data Availability

Subnational database and boundaries

Access to subnational poverty and inequality datasets is publicly available in the PIPmaps [site](#). The data provided here is the same version published in Stata format. The data and the matched boundaries are also available in DDH [SPID](#) and [GSAP](#) links.

Intermediate data included in the replication package

Data.Name	Location	Provided	Citation	Accessed
1/SPID and GSAP datasets	02.input/	TRUE	World Bank (2025b,c)	Nov 2025
2/SPID and GSAP boundaries	05.maps/	TRUE	World Bank (2025b,c)	Nov 2025
3/PIP Data PIP_300_2021.dta	03.intermediate/	TRUE	World Bank (2025d)	Nov 2025
4/CLASS.dta	02.input/	TRUE	World Bank (2025e)	Nov 2025

1/SPID and GSAP datasets (Public)

Both SPID and GSAP datasets are downloaded from Development Data Hub (see reference above). The SPID data is from version 5, and GSAP is from version 7. The data files in Excel format are converted to the Stata format using “import excel”. The Stata format files are saved in the “02.input” folder.

2/SPID and GSAP boundaries (Public)

The original boundary packages (shp) are downloaded from SPID and GSAP respectively (see reference above). The shapefile boundary files are converted into Stata format. For example,

```
spshape2dta "..\am25_subnat_em.shp", replace saving(AM25_subnat_em)
```

```
spshape2dta "..\GSAP_AM25_2023.shp", replace saving(GSAP_AM25_2023)
```

All the boundaries in Stata format are provided in the “05.maps\boundaries” folder.

Those data files are: am25_admin0.dta, am25_admin0_shp.dta, AM25_subnat_em.dta, AM25_subnat_em_shp.dta, GSAP_AM25_2023.dta, and GSAP_AM25_2023_shp.dta.

3/PIP_300_2021.dta (Public)

The data is downloaded from Poverty and inequality Platform (PIP) using the Stata wrapper. The data vintage is November 2025.

```
pip, country(all) year(all) povline(3.00) ppp(2021) clear
```

```
save "\03.intermediate\PIP_300_2021", replace
```

4/CLASS.dta (Public)

The original Excel data is downloaded from:

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

and the Stata data is from:

<https://github.com/GPID-WB/Class/blob/master/OutputData/CLASS.dta>

5/data_coverage_poverty.csv (Public)

The authors compiled the listing data manually based on the availability of subnational data points or number of survey-representative areas for each country from the SPID database. The layer names are based on each country specific information of administrative divisions.

6/Additional inputs are not included in the package but are available publicly from the following sources:

- World Bank Official Boundaries “World Country Polygons - Very High Definition” (*WB_countries_Admin0_10m*) available from:
<https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Administrative-Boundaries?version=3>

The country world boundary (shapefile package) are converted into Stata format and saved into the “\05.maps” folder (WLD_ADM0.dta and WLD_ADM0_shp.dta).

```
spshape2dta "..\WB_GAD_ADM0.shp", replace saving(WLD_ADM0)
```

Instructions for Replicators

Replication from intermediate spatial data

Follow these steps to reproduce results from the intermediate spatial data:

1. Clone the reproducibility package (see folder structure below)
2. Open the Stata master do-file “01.dofile/SD paper tables.do”
3. Change the current directory on line 3 to the reproducibility package path
4. Run the master Stata do-file following the sequence described in the code

List of Exhibits

The provided code (SD paper tables.do) reproduces: all tables and maps. All codes are in the “01.dofile” folder. Map outputs are verified and edited by the World Bank’s Maps Department for publication. Output are saved under “04.output” folder.

Two Tableau files are provided. To replicate the figures, open the Tableau package file and load the corresponding csv data file.

- The “Fig 5 IDN charts.twbx” uses the “03.intermediate\IDN_subnat_national.csv”.
- The “Fig A2 Distribution of areas.twbx” use the “04.output\SPID_subnat_count.csv”.

Exhibit name	Output filename	Note
Figure 1		Authors’ compilation
Figure 2		Authors’ compilation
Figure 3a	Fig3a_WLD_ADM0_poor300.png	Line 125-132
Figure 3b	Fig3b_WLD_subnat_poor300.png	Line 51-59
Figure 4a	Fib4a_WLD_ADM0_gini.png	Line 135-142
Figure 4b	Fig4b_WLD_subnat_gini.png	Line 61-70
Figure 5	Fig 5 IDN charts.twbx	Tableau uses csv data from this code 03.intermediate\IDN_subnat_national.csv
Figure6a	Fig6A_SSA_ADM0_poor_change.png	Line 704-742
Figure6b	Fig6B_SSA_subnat_poor_change.png	Line 510-685
Figure6c	Fig6C_SSA_ADM0_gini_change.png	Line 704-759
Figure6d	Fig6D_SSA_subnat_gini_change.png	Line 510-701
Table 1		Authors’ own elaboration
Table A1	TableA1_SPID.xlsx	Line 175-266
Table A2	TableA2a.xlsx TableA2b.xlsx TableA2c2023.xlsx TableA2c2021.xlsx TableA2c2019.xlsx TableA2c2010.xlsx	Line 268-438
Figure A1	FigA1_WLD_2023_geo_source.png	Line 463-504
Figure A2	Fig A2 Distribution of areas.twbx	Tableau uses csv data from this code 04.output\SPID_subnat_count.csv
Figure A3	FigA3_WLD_subnat_chgini.png	Line 767-827

Requirements

Computational requirements

The Stata master script was run on a Windows (Windows 11 version 24H2, Intel(R) Core(TM) Ultra 7 165U 2.10 GHz, and RAM 32.0 GB RAM). The run time was 10 minutes for Stata.

Software requirements

Stata version 18

Packages

- PIP Stata wrapper – instruction to install with more details available here
net install github, from("https://haghish.github.io/github/")
github install worldbank/pip
- Other Stata packages available from SSC: spshape2dta, colorpalette, spmap

Tableau Desktop Professional Edition 2024 (license required)

Memory, Runtime and Storage Requirements

From intermediate spatial data:

Memory: at least 16 GB

Storage: at least 20 GB available

Run time: 10 minutes (Stata master do-file)

Code Description

Stata

Dofile	Description
SD paper tables.do	Master dofile for tables, maps

Folder structure

```
|— 01.dofile
|— 02.input
|— 03.intermediate
|— 04.output
```

└─ 05.maps
└─ boundaries

References

World Bank. (2025a). Geospatial Poverty Portal [Dataset]. World Bank.

<https://pipmaps.worldbank.org/en/data/datatopics/poverty-portal/home> Accessed on 2025-11-15

World Bank. (2025b). Subnational Poverty and Inequality Database (SPID) [Dataset]. World Bank.

https://datacatalog.worldbank.org/search/dataset/0064796/subnational_poverty_and_inequality_database_spid Accessed on 2025-11-15

World Bank. (2025c). Global Subnational Poverty Atlas (GSAP) [Dataset]. World Bank.

https://datacatalog.worldbank.org/search/dataset/0042041/global_subnational_poverty_atlas_gsap Accessed on 2025-11-15

World Bank (2025d), Poverty and Inequality Platform [data set]. pip.worldbank.org. Accessed on 2026-01-15

World Bank. (2025e). World Bank Country and Lending Groups [Dataset]. World Bank.

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> Accessed on 2025-09-08