Overview

This readme file reproduces the paper entitled "Where are Pakistan's Out-of-School Children? Mapping accessibility of schooling using satellite data and machine learning methods with an application to the 2022 floods". The latest version of the paper is dated 20 Jan 26. The latest version of this README file is dated 20 Jan 26.

The replication package submitted contains the Python scripts needed to replicate all the analytical exhibits (tables and figures) in the paper and those in the annex, as needed.

Data availability

Some of the data used in this paper cannot be made publicly available. The team received access to the following datasets directly from the relevant government agencies:

1. **Punjab School Education Department (SED) Census 2022-23,** School Education Department, Government of Punjab. This contains data on public school location as well as school type and enrollment. This data is not publicly available.
2. **Khyber Pakhtunkhwa Annual Public School Census Data for 2022-23,** Elementary and Secondary Education Department, Government of Khyber Pakhtunkhwa. This contains data on public school location as well as school type and enrollment. This data is not publicly available.
3. **Sindh Public School Census 2022**, School Education and Literacy Department, Government of Sindh. This contains data on public school location as well as school type and enrollment. This data is not publicly available.
4. **Balochistan Annual School Census 2022**; Government of Balochistan. This contains data on public school location as well as school type and enrollment. This data is not publicly available.

The following data was purchased and cannot be shared externally. It can be accessed by World Bank task teams using the internal World Bank Geospatial Library.

1. **Satellite Imagery** purchased by the World Bank from Maxar and Airbus. This shows images of public school locations before and after the 2022 floods. This data is not publicly available.

Datasets that are publicly available include:

1. **CHIRPS Daily Rainfall Dataset**, *CHIRPS v2.0: Climate Hazards Group InfraRed Precipitation with Stations* (dataset). https://www.chc.ucsb.edu/data/chirps (accessed 22 Aug 2025). This contains daily rainfall data and may be accessed by the URL above.
2. **Household ground-truthing survey**. The authors of the working paper collected this data. This contains information on the households and schools that were interviewed to ground-truth the findings reported in this paper. This will be made accessible via the World Bank's microdata catalog no later than February 1, 2026.

Instructions for replicators

Code Description

Each thematic folder includes relevant scripts to produce the respective analysis outputs. These scripts are designed to work independently. The user must execute the scripts one-by-one as there is no single master script.

Step-by-step order to run code files:

1. /2025/directory_creator.py
Run this script to create directories in following folders (change path to each folder separately in code):
- spatial_accessibility/results
- whitespace/results
- rainfall_analysis/results
- catchment/results

- catchment/voronoi/interim_results

2. /2025/whitespace/code
- run 0.9A.py
- run 0.9B.py
- run 0.9C.py
- run 0.9D.py
- run 0.9E.py
- then run summarize_whitespace_prov_lvl.ipynb

- then run get_whitespace_sec_loc_type.py

Note: to run '/0.9D.py', qgis needs to be installed because this script uses qgis python plugins which comes with qgis installation.

3. /2025/spatial_accessibility/code

- run 0.1A.py
- run 0.1B.py
- then run summarize_accessibility_prov_lvl.ipynb

4. /2025/catchment

- run /voronoi/voronoi_maker/main.py
- then run /voronoi/voronoi_maker/correct_multigeometry_catchments.py
- then run /code/main.py
- then run /code/summarize_enrollment_prov_lvl.ipynb

Note: to run '/voronoi/voronoi_maker/main.py', qgis needs to be installed because this script uses qgis python plugins which comes with qgis installation.

5. /2025/school_analysis/code

- run main.ipynb
- run plot.py

List of Exhibits

Tables

Exhibit name	Output filename	Script	Note
Table 1	**Published estimate. Reproducibility not required.**		
Table 2 and Table 3	**Private Schools** results\table2.csv **Public Schools** results\table3.csv	code\main.ipynb	Found in school_analysis
Table 4 but totals only	**Public Enrollment** results\table4.csv **Private Enrollment** results\table4_1.csv	code\main.ipynb	Found in school_analysis
Table 5	results\table5.csv	code\main.ipynb	Found in school_analysis
Table 6	results\table6.csv	code\main.ipynb	Found in school_analysis
Table 7 and additional breakdowns	**Catchment** results\all_province_variants\private_school_analysis_all_all_potential.json.gz results\all_province_variants\private_school_analysis_flooded_all_potential.json.gz **Whitespace** results\all_province_variants\private_school_analysis_all_all_whitespace_s	**Catchment** code\main.py code\ summarize_enrollment_prov_lvl. ipynb **Whitespace** - code\ 0.9A.py - code\ 0.9B.py - code\ 0.9C.py	Found in spatial_accessibility, whitespace, catchment and school_analysis.

	ummary.json.gz results\all_province_variants\private_school_analysis_flooded_all_whitespace_summary.json.gz **Accessibility** results\all_province_variants\private_school_analysis_all_all_accessibility.json results\all_province_variants\private_school_analysis_flooded_all_accessibility.json.gz **Breakdowns** results\table7.csv	- code\ 0.9D.py - code\0.9E.py code\ summarize_whitespace_prov_lvl. Ipynb **Accessibility** code\0.1A.py, code\0.1B.py code\ summarize_accessibility_prov_lvl. ipynb **Breakdowns** code\main.ipynb	
Table 8	**Catchment** results\all_province_variants\private_school_analysis_all_all_potential.json.gz **Whitespace** results\all_province_variants\private_school_analysis_all_all_whitespace_summary.json.gz **Accessibility** results\all_province_variants\private_school_analysis_all_all_accessibility.json	**Catchment** code\main.py code\ summarize_enrollment_prov_lvl. ipynb **Whitespace** - code\ 0.9A.py - code\ 0.9B.py - code\ 0.9C.py - code\ 0.9D.py - code\0.9E.py code\ summarize_whitespace_prov_lvl. Ipynb **Accessibility**	Found in spatial_accessibility, whitespace and catchment

		code\0.1A.py, code\0.1B.py code\ summarize_accessibility_prov_lvl. Ipynb	
Table 9 but additional breakdowns needed	**Whitespace** results\all_province_variants\private_school_analysis_all_all_whitespace_summary.json.gz results\all_province_variants\private_school_analysis_flooded_all_whitespace_summary.json.gz **Accessibility** results\all_province_variants\private_school_analysis_all_all_accessibility.json results\all_province_variants\private_school_analysis_flooded_all_accessibility.json.gz **Breakdowns** results\table9_accessiblity.csv results\table9_whitespace.csv	**Whitespace** - code\ 0.9A.py - code\ 0.9B.py - code\ 0.9C.py - code\ 0.9D.py - code\0.9E.py code\ summarize_whitespace_prov_lvl. Ipynb **Accessibility** code\0.1A.py, code\0.1B.py code\ summarize_accessibility_prov_lvl. Ipynb **Breakdowns** code\main.ipynb	Found in spatial_accessibility, whitespace, and school_analysis
Table 10	results\table10.csv	code\main.ipynb	Found in school_analysis
Table 11	results\table11.csv	code\main.ipynb	Found in school_analysis
Table 12	**Catchment** results\all_province_variants\private_school_analysis_all_all_potential.json.gz	Catchment code\main.py code\ summarize_enrollment_prov_lvl. ipynb	Found in spatial_accessibility, whitespace, catchment and school_analysis.

	results\all_province_variants\private_school_analysis_flooded_all_potential.json.gz **Whitespace** results\all_province_variants\private_school_analysis_all_all_whitespace_summary.json.gz results\all_province_variants\private_school_analysis_flooded_all_whitespace_summary.json.gz **Accessibility** results\all_province_variants\private_school_analysis_all_all_accessibility.json results\all_province_variants\private_school_analysis_flooded_all_accessibility.json.gz **Breakdowns** results\table12.csv	**Whitespace** - code\0.9A.py - code\0.9B.py - code\0.9C.py - code\0.9D.py - code\0.9E.py code\ summarize_whitespace_prov_lvl. Ipynb Accessibility code\0.1A.py, code\0.1B.py code\ summarize_accessibility_prov_lvl. ipynb Breakdowns code\main.ipynb	
Table 13	results\table13.csv	code\main.ipynb	Found in school_analysis
Table 14 TABLE A5.1 (Appendix 5) * TABLE A5.2 (Appendix 5) *		code\ ground_truth_tables	Found in Survey folder

TABLE A5.2 (Appendix 5)			
Appendix 6 Table 1	**Reproducibility not required**		
Appendix 6 Table 2	**Reproducibility not required**		

Figures

Exhibit name	Output filename	Script	Note
Figure 1	**Reproducibility not required.**		
Figure 2	**Reproducibility not required.**		
Figure 3	**Reproducibility not required.**		
Figure 4	→ using input files '/data/final_sindh_pvt_enrollment.csv' and '/data/Details of Private Schools District Wise.xlsx' and output figure is generated using '/school_analysis/code/main.ipynb'		
Figure 5	Developed using QGis (details below) (Lahore) → use '/catchment/voronoi/interim_results/punjab/lahore/tehsils/all/private_school_analysis/all_all_voronoi2.kml' file generated using '/catchment/voronoi/voronoi_maker/main.py' Faislabad) → use '/catchment/voronoi/interim_results/punjab/faisalabad/tehsils/all/private_school_analysis/all_all_voronoi2.kml' file generated using '/catchment/voronoi/voronoi_maker/main.py' Load these kml files separately in QGIS and use '/catchment/results/province_lvl_oosc/oosc_pk_all_all.csv' file to join data in kml file. (these files are exported in 'shapefiles' folder)		
Appendix 3 Figure 3	using '/data/enrollment_data.csv' as input file and output figure is generated using '/school_analysis/code/plot.py'		
Appendix 3 Figure 4a and 4b	→ using '/data/annex_fig4b_data.csv', '/data/annex_fig4a_pred.npy', '/data/annex_fig4a_gt.npy' as input files and output figure is generated using '/school_analysis/code/plot.py'		

Requirements

The project's code is executed on a Linux OS.

Required python packages

aceso==0.1.0
affine==2.4.0
fastkml==0.12
fiona==1.9.6
Fuzzy==1.2.2
geopandas==0.13.2
geopy==2.4.1
h5py==3.11.0
joblib==1.4.2
matplotlib==3.7.5
networkx==3.1
numba==0.58.0
opencv-python==4.10.0.84
openpyxl==3.1.5
pandas==1.5.3
protobuf==5.26.1
pyarrow==17.0.0
pyproj==3.5.0
rasterio==1.3.11
requests==2.32.3
Rtree==1.3.0
scikit-image==0.21.0
scikit-learn==1.3.2
scipy==1.10.1
Shapely==1.8.0
simplekml==1.3.6
tifffile==2023.7.10
tqdm==4.66.4
urllib3==1.26.20
xmltodict==0.14.2
gdal==3.6.2

Other software needed

Tableau
QGIS
Stata

Folder Structure

```
reproducibility/
├── spatial_accessibility/
│   ├── code/
│   │   ├── 0.1A.py
│   │   └── 0.1B.py
│   │   └── summarize_accessibility_prov_lvl.ipynb
│   └── result/  ← contains outputs
├── school_analysis/
│   ├── code/
│   │   └── main.ipynb
│   └── result/  ← contains .csv outputs
├── whitespace/
│   ├── code/
│   │   ├── 0.9A.py
│   │   ├── 0.9B.py
│   │   ├── 0.9C.py
│   │   ├── 0.9D.py
│   │   ├── 0.9E.py
│   │   └── summarize_whitespace_prov_lvl.ipynb
│   │   └── get_whitespace_sec_loc_type.py
│   └── result/  ← contains outputs
├── catchment/
│   ├── code/
│   │   ├── main.py
│   │   └── summarize_enrollment_prov_lvl.ipynb
│   └── result/  ← contains outputs
├── rainfall_analysis/
│   ├── code/
│   │   ├── 2A_rainfall_analysis.ipynb
│   │   └── 2C_rainfall_all_analysis_prov_lvl.ipynb
│   └── result/  ← contains outputs
```

Annex 1 – Steps take in QGIS to create maps for Figure 8.

Readme for FIGURE 5

OOSC figures

reproductibility/
├── catchment/
│ ├── figures/
│ │ ├── figure
│ │ ├── qgis_style
│ │ └── shapefiles
│

'figure' folder contains the output images for figure 8.

'shapefiles' folder contains the shapefiles required for all the figures. The name of the shapefiles contain fig no. as suffix. We have already assigned data to shapefiles using '/catchment/results/province_lvl_oosc/oosc_pk_all_all.csv' file.

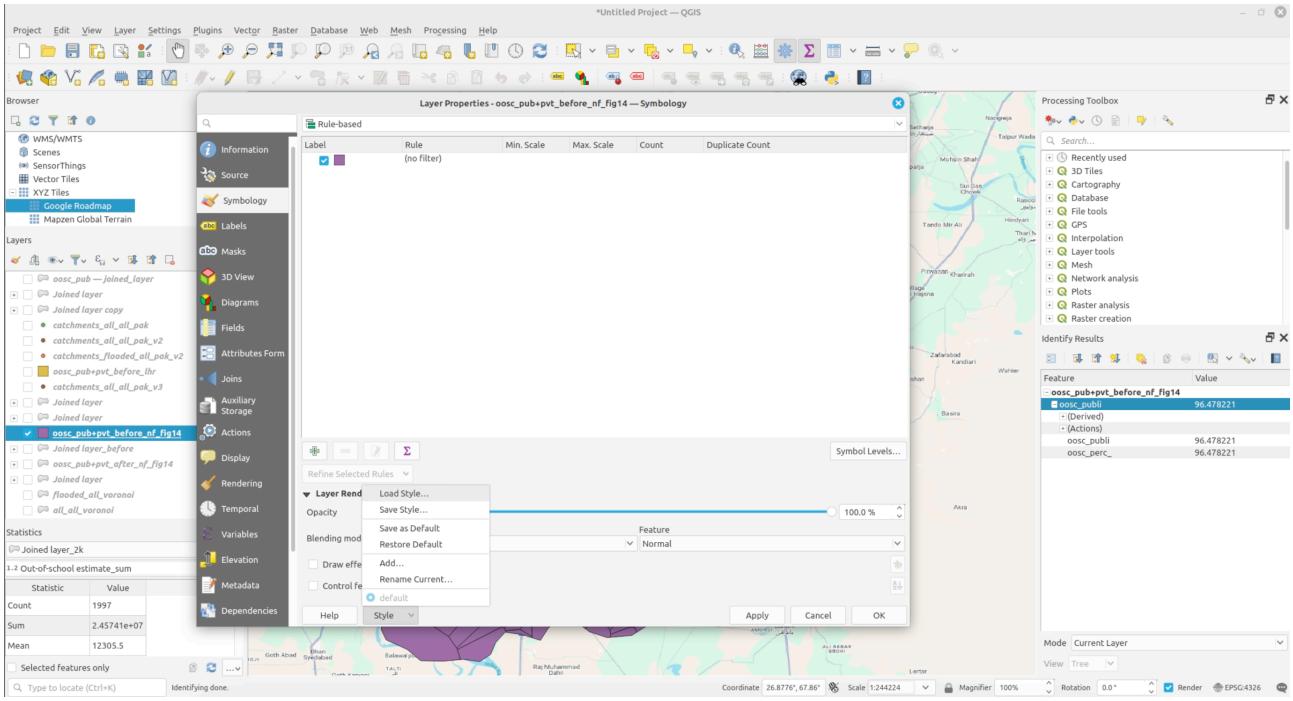
'qgis_style' contains the qgis style layers for each figure.

Step 1:

Open qgis. Import a shapefile(.shp file) of a figure from 'shapefiles' folder (or drag-drop) into qgis.

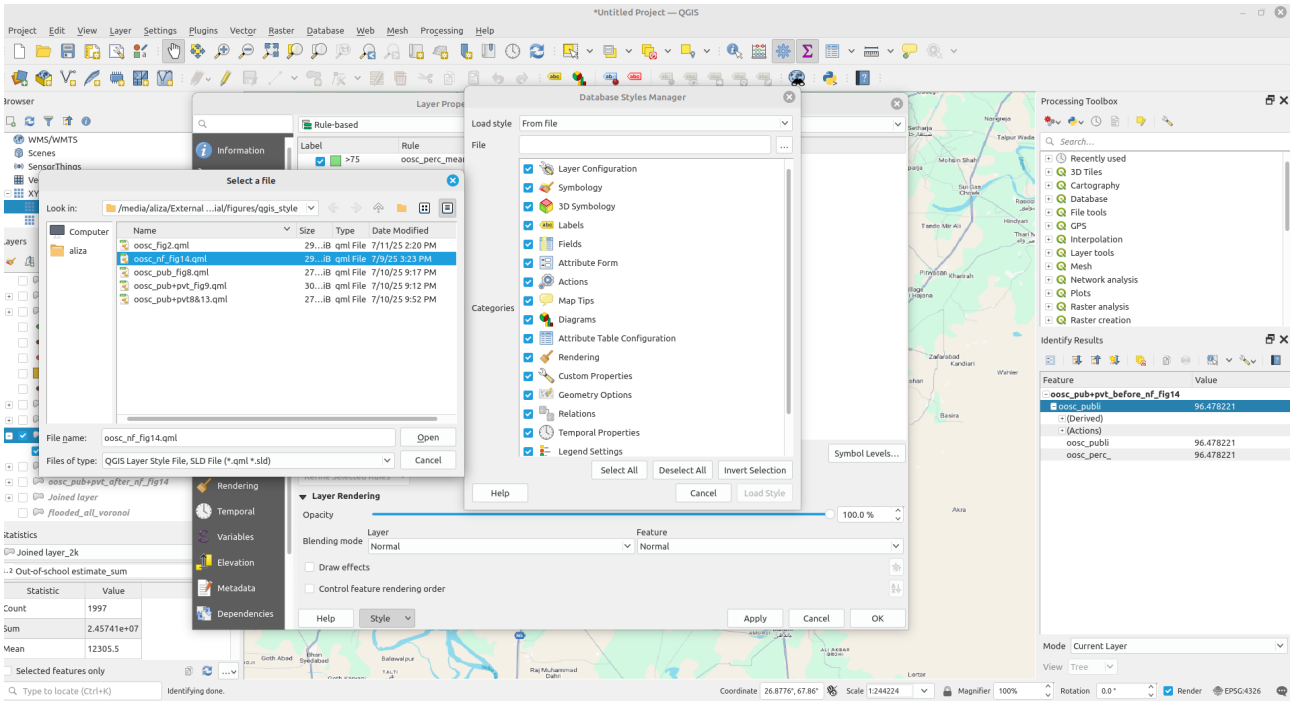
Step 2:

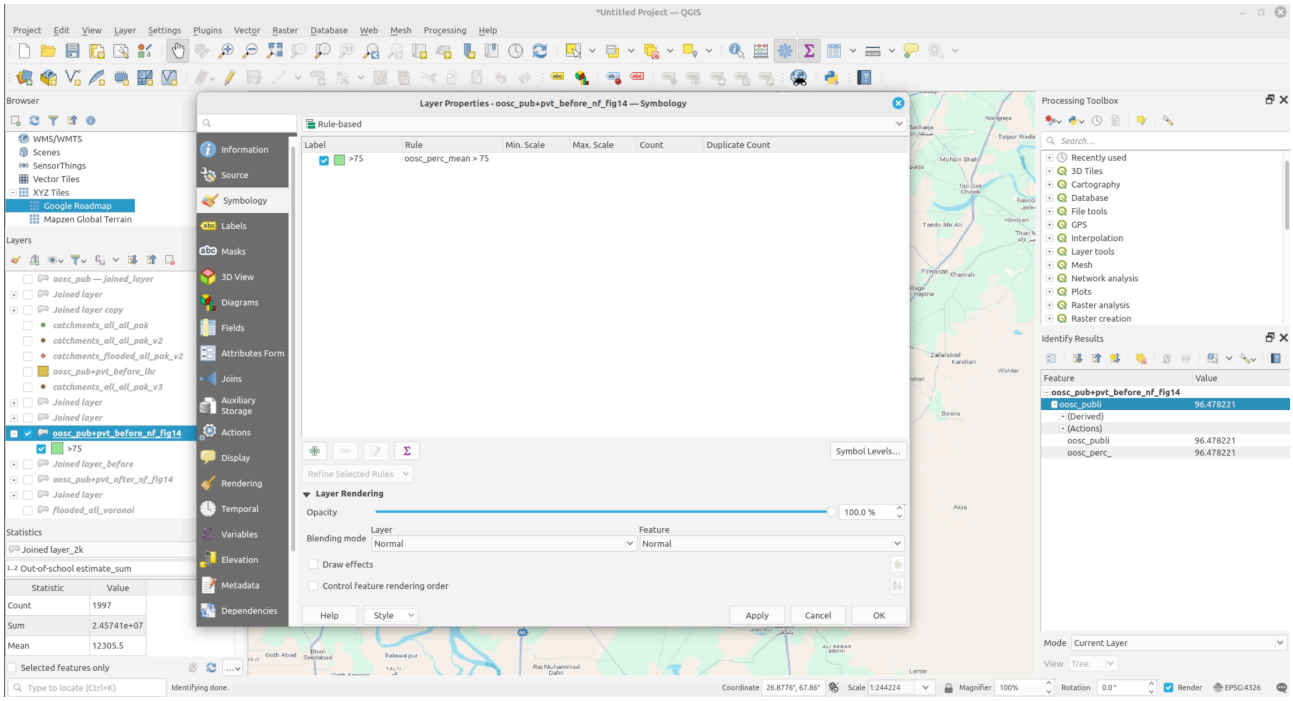
Right click on the shapefile in qgis and click Properties. Go to Symbology and select Rule-based from drop down.



Step 3:

Click on Load Style button at the bottom and load style for the required figure no. from 'qgis_style' folder. Click Ok.





Notes:

1. For each shapefile of a figure, there is a corresponding QGIS style file. All files within each folder include the figure number as a suffix in their names.

3. Both maps in Figure 5 use a single style file with the `_fig8` suffix located in the `qgis_style` folder.

Requirement:

QGIS