**Provided input files (place in root folder)**

Dashboard.csv (raw weekly DDC scrape; 2020-2023).

diarrhoea\_cases\_monthly.txt (monthly national totals; 2012-2019).

weekly.xlsx (weekly national totals; 2020-2023).

Diarrhoea\_cases\_2012-2019.csv (province-year totals, 2012-2019).

Diarrhoea\_cases\_2020-2023.csv (compiled after imputation).

Diarrhoea\_cases\_2012-2023.csv (cleaned and stitched file).

2021\_Data\_Comparison.csv (comparison of 2021 data across two sources)

Poverty.xlsx, Personnel.xlsx (two cleaned covariate files).

**Proprietary input files (place in root folder)**

(These are publicly available files to be obtained from sources as indicated in the “data availability” section of the essay)

Thailand shapefile: gadm41\_THA\_1.shp, together with .dbf/.shx [link: <https://gadm.org>]

Population: pop\_info\_2023.csv [link: <https://data.humdata.org/dataset/cod-ps-tha>]

**Run order**

**0. Obtain necessary propriety input files from the links provided.**

**1. Impute & clean weekly dashboard data** (XGBoost.R, Part A)

**2. Aggregate weekly data into annual data and stitch long series** (annual\_compile.R)

**3. Temporal analyses** (Temporal Analysis.R)

**4. Spatial mapping** (Choropleths.R)

**5. Equity analysis** (Equity Analysis.R)

**6. Association modelling** (Global Poisson Model.R)

**7. INLA model comparison** (INLA\_comparison.R)

**8. Feature ablation** (XGBoost.R, Part B)

Automated scraping of the DDC dashboard was done using **scraping\_script.py**