## Reproducibility package

There are four scripts:

1. Tidying data
2. Cleaning data
3. Constructing indicators for analysis
4. Analysis

All these scripts can be run from the *Master\_script.R* withing the *Reproducibility\_package.RProj* R project.

## Dependencies

R packages required:

* "dplyr",
* “tidyr”
* “pacman”
* "ggplot2",
* "corrplot",
* "gt",
* "plm",
* "lmtest",
* "here",
* "labelled",
* "stringi",
* "Hmisc",
* "stringi",
* "stargazer"

## Data Availability Statement

The authors of the manuscript certify they have legitimate access to and permission to use the data used in this manuscript. All the data is publicly available. The authors are not responsible for updates to the data made at the source after the versions used in the analysis.

### Data sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Name | Data Files | Provided | | References |
| Open Street Maps | colombia\_infraestructure.csv | | TRUE | WITS (2023) |
| Ookla | colombia\_connectivity.csv | | TRUE | Ookla (2020) |
| Humanataria Data Exchange | No data file in package. It was merged to the previous data files. | | FALSE | HDX (2023) |
|  |  | |  |  |
|  |  | |  |  |

**Open Street Maps**

* The package uses data from Open Street Maps using the Python Overpass API**.** The package already provides the data, so you don’t need to consult the API, but if you need here is [how to do it.](https://towardsdatascience.com/loading-data-from-openstreetmap-with-python-and-the-overpass-api-513882a27fd0)

**Ookla**

* The package uses connectivity data from Ookla from the first and fourth quarter of 2020. The package already contains the data, but if you need to download it you can do it from [here](https://github.com/teamookla/ookla-open-data).

**Humanataria Data Exchange**

* Both datasets are combined with Colombia Admin level data. The package does not need it, but it is available [here.](https://data.humdata.org/)