README: PADIN Cost Analysis Reproducibility Package

This package contains the code and data for conducting the analysis presented in the paper
"Costs, Cost-Sharing, and Cost Drivers of a Home Visiting Program: The Case of PADIN in Brazil,"
authored by **Alaka Holla and Yilin Pan**.

Running the Scripts

To run the scripts, new users **only need to change the directory** in **line 134** of the `0_Master_script.R` file located in the `"code"` folder.

Input Files

Code Files (`./code/`)

The package includes three R scripts for data cleaning, variable construction, and analysis: 1. `1_Brazil_PADIN_data_cleaning.R` 2. `2_Brazil_PADIN_variable_construction.R` 3. `3_Brazil_PADIN_analysis.R`

Data Files

The package includes **seven datasets** stored in different subdirectories. The data will be published in the Microdata Library in the upcoming weeks.

De-identified Data (`./data/1_De-identified/`)

1. `survey_data_de-identified.csv`
    - Contains responses on time use, salaries of key personnel, transportation modes, and training frequencies.
    - **De-identified and included in the package.**
2. `monitoring_and_evaluation_data_de-identified.csv`
    - Includes Monitoring & Evaluation (M&E) data on the number of supervisors and home visitors by municipality.
    - **De-identified and included in the package.**
3. `state_budget_2018.csv`
    - Financial spreadsheets detailing PADIN personnel, training, and food costs from the state government.
    - **Aggregated data, included in the package.**
4. `municipality_budget_de-identified.csv`

- Monthly expenditure reports for office supplies and food from 20 municipalities.
- **De-identified and included in the package.**

5. `transportation_gps_de-identified.csv`
   - Simulated transportation profiles, including travel distances and estimated GIS-based travel times.
   - Based on a confidential dataset (not included in the package).
   - **De-identified and included in the package.**
6. `online_monitoring_de-identified.csv`
   - Records from the Online Monitoring System on supervisors, home visitors, families, children, visits, and meetings by municipality.
   - **De-identified and included in the package.**
7. `municipality_statistics_de-identified.csv`
   - Municipal statistics (area, population size, GDP per capita) with anonymized municipality labels.
   - **Used for analysis but NOT included in the package**
   - **Note**: The dataset is not included is because in all other datasets, municipalities are de-identified and labeled as M1, M2, M3, etc.. If a dataset with publicly available statistics such as population and GDP were shared with these anonymized labels, users might be able to reverse-engineer the identities of the municipalities

Output Files

Figures (`./output/1_Exhibits/`)

The package produces **five exhibits in the main text and six in Appendix A**:

1. **Figure 1:** Cost per student per month, by funding source (`figure1_total_cost_funding.png`)
2. **Figure 2:** Cost per child per month, by ingredient and funding source (`figure2_total_cost_ingredient.png`)
3. **Figure 3:** Municipal cost decomposition per child per month, by ingredient (`figure3_heatmap_m.png`)
4. **Figure 4:** Supervisors' gross salary and time dedicated to PADIN (`figure4_supervisor.png`)
5. **Figure 5:** Discrepancy between self-reported transportation costs and GIS-estimated costs (`figure5_dumbbell_transportation.png`)

Appendix A Figures:

6. **Figure A1:** Number of home visitors by data source (`figureA1_ADI_consistency.png`)
7. **Figure A2:** Number of children served by data source (`figureA2_children_consistency.png`)
8. **Figure A3:** State costs of PADIN by ingredient (% of total state costs) (`figureA3_decomposition_state_cost.png`)

9. **Figure A4:** Cost per child per month by funding source (% of total costs) (`figureA4_total_cost_percentage.png`)

10. **Figure A5:** Correlation between GDP per capita and municipal cost per child (`figureA5_corr_spending_wealth.png`) This figure won't be able to be reproduced by the code, as it is constructed using the not included dataset `municipality_statistics_de-identified.csv`
11. **Figure A6:** Cost estimation discrepancy using financial records vs. ingredients-based method (`figureA6_discrepancy.png`)

Tables (`./output/2_Tables/`)

1. **Table 3 (`table3.csv`)**
    - Cost per child by ingredient, by month and by year.
2. **Table 4A (`table4A.csv`)**
    - Monthly and annual quantities/prices of ingredients.
    - **Panel A:** State-level inputs for 5,135 children across 48 municipalities.
3. **Table 4B (`table4B.csv`)**