

README: Replication Files for "*Cash Is Queen: Local Economy Effects of Cash Transfers to Women in West Africa*"

Overview

This repository contains the necessary files to replicate the analysis in "*Cash Is Queen: Local Economy Effects of Cash Transfers to Women in West Africa*" by Sreelakshmi Papineni, Paula Gonzalez, Markus Goldstein, and Jed Friedman.

Memory and Runtime Requirements:

- The Stata analysis code requires approximately 9 hours and 15 minutes to execute fully.
- The R analysis code takes about 1 minute to run.

Software Requirements: Stata (Tested on version 18)

Required Stata packages:

- ietoolkit (6.3)
- coefplot (1.8.3)
- outreg2
- estout
- schemepack
- colrspace
- swindex
- winsor
- destring
- split
- clonevar
- moransi
- acreg

Instructions for Replicators

1. **Prepare Environment:** Open the master do-file (MasterDofile_CashIsQueen.do). Ensure user settings, including unique user number and OneDrive paths are correctly configured.
2. **Install Packages:** If necessary, install user-written packages by modifying line 42 in the master do-file (change 0 to 1). This is a one-time set-up per computer.
3. **Set Repository Path:** Modify line 70 to reflect the path of the repository on your computer.
4. **Execute Analysis:** The master do-file (MasterDofile_CashIsQueen.do) is preconfigured for immediate execution. However, for a full analysis, you may need to set/check the flags (DataCleaningConstruction, DataAnalysis) to 1. After adjusting these flags, run the master do-file. For running individual sections, simply set the respective flags for those specific sections.
5. **Generate Outputs:** Running the master do-file will construct the main datasets for analysis and generate all tables and graphs for the working paper.

6. Output Location: Outputs will be stored in “Outputs\Raw/Figures” and “Outputs/Raw/Tables”. Main analysis outputs are in the 'Main' folder, and appendix outputs in the 'Appendix' folder. All outputs are formatted and compiled in two main files containing all tables.

Datasets Description

Primary Dataset

The primary dataset for this analysis is the Feed the Future Nigeria Livelihoods Project (FNLP) Household Survey, a comprehensive individual-level dataset collected in three rounds—baseline, midline, and endline—through surveys conducted with a primary woman in the household and her male spouse. The surveys were administered by TNS-RMS Kantar Public Survey Firm.

The baseline and endline datasets are publicly available in the World Bank Microdata Library, while the midline dataset is in the process of being made public. Below is a summary of each dataset, including collection dates and access information:

Project	Cycle	Dates of data collection	Data Access
Survey Data for Future Nigeria Livelihoods Project	Baseline	2015/05/01 - 2015/06/01	Microdata Library (Access date: 09/2015).
	Midline	2017/04/01- 2017/06/01	Not yet Public (Access date: 07/2017).
	Endline	2018/05/01 - 2018/07/30	Microdata Library (Access Date: 09/2018)

The dataset \$data/Raw/FNLP_HouseholdFullData.dta is the main source for this analysis. It compiles all three survey rounds and includes 3,976 unique household observations. The dataset spans from May 2015 (baseline) to July 2018 (endline).

Description of how HouseholdFullData was constructed from public datasets: The raw data collected at endline (dataset included surveys with both female and male respondents) was merged with a tracking sheet sent by the field team to ensure all data for completed surveys were received. A few duplicate observations were manually cleaned after consulting with the field team on the correct identifier to ensure observations in the dataset were uniquely identified. The endline dataset was appended to the midline dataset and then merged with the baseline dataset (the *time* variable indicates midline = 1 and endline = 2). While several variables were prepared for analysis and included in the cleaned data file this dataset also still includes all raw variables as they were originally collected. The clean dataset has been stripped of personally identifiable information (PII) including names and GPS locations as per data protection protocols. Note that the baseline dataset that is downloadable in the World Bank microdata library contains the same raw baseline variables from the 3,976 households plus some additional created variables. The same baseline data will be included in the data shared in the replication package. However, the data in the replication package includes the full set of variables.

Note that endline data for this paper will also be made available in the microdata library but it is currently embargoed. The data is expected to be made public after the paper is published in a journal.

After running the master data construction do-file, two main cleaned datasets are generated for household-level analysis:

- \$data/Final/FNLPPDataWideAnalysis.dta: The primary dataset for individual-level analysis in wide format. Variables are labeled as _bl (baseline), _f2 (midline), and _f3 (endline). This dataset is used for ANCOVA analyses or when outcome data is available for only one follow-up. The unique identifier is hhid.
- \$data/Final/FNLPPDataLongAnalysis.dta: A panel dataset that includes baseline, midline, and endline rounds. Key identifiers are wave and hhid_panel.

Secondary Datasets

Several secondary datasets stored in \$data/Raw/ provide supplementary data used in the analysis. These datasets cover census-based indicators, anthropometric data, program treatment details, village characteristics, and spatial variables. Below is a brief description of each dataset:

- FNLPP_Endline_ChildAnthroData: De-identified anthropometric data for children in surveyed households at endline (May 2018). The raw data containing children's (aged 5 years or under) anthropometric measurement (child name, gender, age, weight, height, MUAC) was cleaned (trimmed at the 1st and 99th percentiles) and merged with World Health Organization (WHO) growth charts. The WHO growth charts are used to calculate z-scores for height-for-age, weight-for-age, and weight-for-height (the datasets were created before the zanthro command was created in Stata). Mid-upper arm circumference, often shortened to MUAC, is a measurement that allows health workers to quickly determine if a patient is acutely malnourished. Indicators were created for values below the cut-offs of 125 mm and 115 mm to define moderate and severe acute malnutrition, respectively. Available upon request from the World Bank [Microdata Library](#) once the data is no longer under embargo.
- FNLPP_Treatment: Contains program-related information, including village Local Government Area (LGA), treatment status, and baseline household demographics. Derived from the baseline dataset, accessible via World Bank Microdata Library. Variables included: *gps_startlatitude gps_startlongitude gps_startaltitude hhid statecode_bl lgacode_bl wardcode_bl villagecode_bl hhhead_gender_bl hhhead_address_bl num_hh_members_bl treatment_bl vulnerability_category_ward_bl caseworker_b cashtransferassignment_bl age_spouse_bl village_new_bl ward_new_bl lga_new_bl polygamy_bl both_male_fem_bl female_adult_only_bl povertylinebelow_125_bl povertylinebelow_190_bl noland_bl farm_lessthan1hectare_bl farm_greaterthan1hectare_bl Monthly_fnlp_bl Monthly_nonfnlp_bl Lump_Sum_fnlp_bl Lump_Sum_nonfnlp_bl Cash_Transfer_fnlp_bl Cash_Transfer_nonfnlp_bl treatment_ct_bl farming_woman_bl empowered_new_bl marital_status_bl widowed_bl ward_vulnerable_bl Stratification_Dummies_bl*
- FNLPP_VillageLevel: Aggregated village-level data on labor force participation, infrastructure, and socio-economic indicators for 104 program villages. Accessible via baseline dataset request. Variables included from the raw data: *village_new_bl village_ineligible_ct CTvill d_lgacode_bl_1 RDD_18 FNLPP_treatment vulnerability_category age_hhhead_bl marital_status2_bl_1 marital_status2_bl_2 marital_status2_bl_3 marital_status2_bl_4 widowed_bl literac y_hhhead_bl num_hh_members_bl any_farm_hh_bl any_business_bl farming_work_w_bl farming_work_h_bl business_work_w_bl business_work_h_bl mean_dist_villmark min_distance_market min_distance_vCT min_distance_vFNLPP mean_dis tance min_distance max_distance num_vul_village primary_1 secondary_1 healthcenter_1 hospital_1 doctor_1 midwife_1 pharmacy_1 airtime_1 bus_1 road_1 bank_1 mfi_1 police_1 market_1 mosque_1 commcenter_1 primary_3 secondary_3 healthcenter_3 hospital_3 doctor_3 midwife_3 pharmacy_3 airtime_3 bus_3 road_3 bank_3 mfi_3 police_3 market_3 mosque_3 commcenter_3 primary_5 secondary_5 healthcenter_5 hospital_5 doctor_5 midwife_5 pharmacy_5 ai rtime_5 bus_5 road_5 bank_5 mfi_5 police_5 market_5 mosque_5 commcenter_5 primary_7 secondary_7 healthcenter_7 hospital_7 doctor_7 midwife_7 pharmacy_7*

airtime_7 bus_7 road_7 bank_7 mfi_7 police_7 market_7 mosque_7 com mcenter_7 primary_6 secondary_6 healthcenter_6 hospital_6 doctor_6 midwife_6 pharmacy_6 airtime_6 bus_6 road_6 bank_6 mfi_6 police_6 market_6 mosque_6 commcenter_6 group301_1 group302_1 group303_1 group304_1 group305_1 group306_1 group307_1 group308_1 group309_1 validation_score ppi_score kids_less5 hh_femaleheaded hh_elderlyheaded disabled_num meat_freq daily_wage_num any_wage_hh_bl score_1a score_1b score_1c score_1d score_2 score_3a score_3b score_3c score_4a score_4b score_4c score_4d score_5a score_5b score_6a score_6b score_7 score_8 score_9 score_10 total_num infrastructure_index est_vulnerable_crs d_vuln_1 d_vuln_2 d_vuln_3 villagecode_tr eat ppi_p50_merge intensity_cash_pop intensity_vul_pop est_ttl_pop_size est_ttl_pop_size_gr any_emp_proxy num_inf_ind num_com_ind

- **FNLP_PPI_CensusData:** De-identified 2015 Census Data, including progress-out-of-poverty index (PPI) scores for vulnerable households. Collected by TNS-RMS Kantar Public Survey Firm. Not publicly available. (Access date 09/2015).
- **HHSpatialLongData & HHSpatialWideData:** Household-level datasets measuring market conditions, including the number of CTs, eligible households, and total households within various distances (100m - 1000m).
 - Long format: Panel data
 - Wide format: Cross-sectional data
 - Not publicly available due to inclusion of GPS coordinates.
- **VillageDistancesHH:** Dataset for non-FNLP treatment villages, containing household intra-village distances (mean, max, min) and estimated village population sizes. Not publicly available due to GPS data.
- **VillagePovertyIntensityData:** Village-level dataset on population size and poverty intensity from pre-program estimates. Collected by TNS-RMS Kantar Public Survey Firm. Not publicly available due to sensitivity.
- **VulnerablesMatrixDatafromR:** Census-based household dataset computing the number of surrounding households within varying radii. Not publicly available due to GPS coordinates.
- **SUVTADistancesVillages:** Household-level dataset measuring distances between households across villages, categorized by treatment status. Not publicly available due to GPS coordinates.
- **DistancesHHtoVillageType:** Identifies the closest distances between households and villages based on treatment status. Primary identifier: hhid. Not publicly available due to household location data.
- **DistancesMarketData:** Household-level dataset containing distances to 17 identified markets in both LGAs. Collected by TNS-RMS Kantar Public Survey Firm. Not publicly available due to sensitivity.
- **DistancesWithinVillage:** Provides key intra-village household distance measures (mean, min, max). Constructed from information of households dataset at baseline using gps coordinates. Not publicly available due to GPS coordinates.
- **CensusData_VulnerableHH:** Census dataset providing PPI scores and estimates of vulnerable households per village, merged with GPS and market condition variables. Not publicly available due to household privacy concerns. Access data 09/2015.

The restricted datasets were constructed using household GPS coordinates from the **Feed the Future Nigeria Livelihoods Project baseline data**. The **public version** of this dataset does **not include GPS**

information. Due to **sensitivity concerns**, the GPS data cannot be disclosed and will **not be made publicly available**.

Access to the restricted version was granted **exclusively for reproducibility verification** by the **PRWP Reproducibility Team at the World Bank**.

For additional details about the files used, please refer to the documentation provided in the package. If you have further questions, you may contact the team at: pgonzalezmartine@worldbank.org.

Do-File Structure and Execution Order

1. **Master do-file (MasterDofile_CashIsQueen.do):** This script orchestrates the execution of all other do-files required for replicating the study "*Cash Is Queen: Local Economy Effects of Cash Transfers to Women in West Africa.*" Execute this master do-file first to ensure a full replication.

Main Sections of do-file:

Part 1: Data Cleaning and Construction

This section is responsible for preparing and structuring the dataset before analysis. It is divided into four parts:

- Part A: PartA_HouseholdDataMainEffects.do This do-file prepares and cleans the household-level dataset for the *Nigeria Feed the Future* study. It loads raw data, defines treatment groups, constructs key variables, and applies sample selection criteria. This script is Part A of the data cleaning and construction phase, ensuring the dataset is ready for analysis.
- PartB_CensusAndSpatialData.do This do-file processes census and spatial data for the *Nigeria Feed the Future* study. It loads raw census data, constructs village-level variables, and integrates spatial information, including neighbor and local Economy Characteristics. This script is Part B of the data cleaning and preparation phase, ensuring geographic and demographic data are ready for analysis.
- PartC_FNLWideData.do: This do-file constructs a wide dataset for the *Nigeria Feed the Future*. It integrates household, census, and neighborhood data, ensuring key variables are cleaned and structured for analysis. This script is Part C of the data preparation phase, focusing on generating a non-panel dataset with geographic and demographic variables
- PartD_FNLLongData.do: This do-file constructs a long-panel dataset for the *Nigeria Feed the Future* study. It loads the final wide dataset generate in Part C and defines variables necessary for panel analysis, including controls and time-varying factors. This script is Part D of the data preparation phase, ensuring the data is structured for longitudinal analysis.

Part 2: Data Analysis

This section executes all statistical and econometric analyses, including the production of tables and figures for the study.

Main Analysis

- RandomizationBalanceAndAttrition.do: Assesses balance across treatment groups.
- RegressionAnalysis.do: Estimate primary treatment effects using regression models and generate key tables for the main findings.

Appendix Analysis: This section provides additional analyses and robustness checks.

- Appendix A.do: Program Design and Data. Further details on program implementation and dataset construction.
- Appendix B: Spatial Modeling of Spillovers.
- Appendix C: Business Activities and Financial Access.
- Appendix D: Regression Discontinuity Design (RDD) additional details.
- Appendix E: Main Results Including Additional Household.
- Appendix F: Secondary Outcomes.

Additional Analysis: There are two Rscripts which need to be run separately to the Stata Analysis. However, this can be done at the end. The Rscripts provide two figures included in the appendix AppendixB_RScript and AppendixD_RScript.

Outputs:

All processed datasets, tables, and figures are stored in:

1. Raw Outputs: Outputs/Raw/Tables and Outputs/Raw/Figures.
2. Compiled Tables: Stored in Outputs/Final/CashIsQueen_MainTables_date` and Outputs/Final/CashIsQueen_AppendixTables_date.
3. Final Outputs: Use of the two compiled excel files saved in the final folder ready to be formatted for LaTeX integration.
4. Graph Formats: Graphs are saved as .gph and .png files for LaTeX integration.

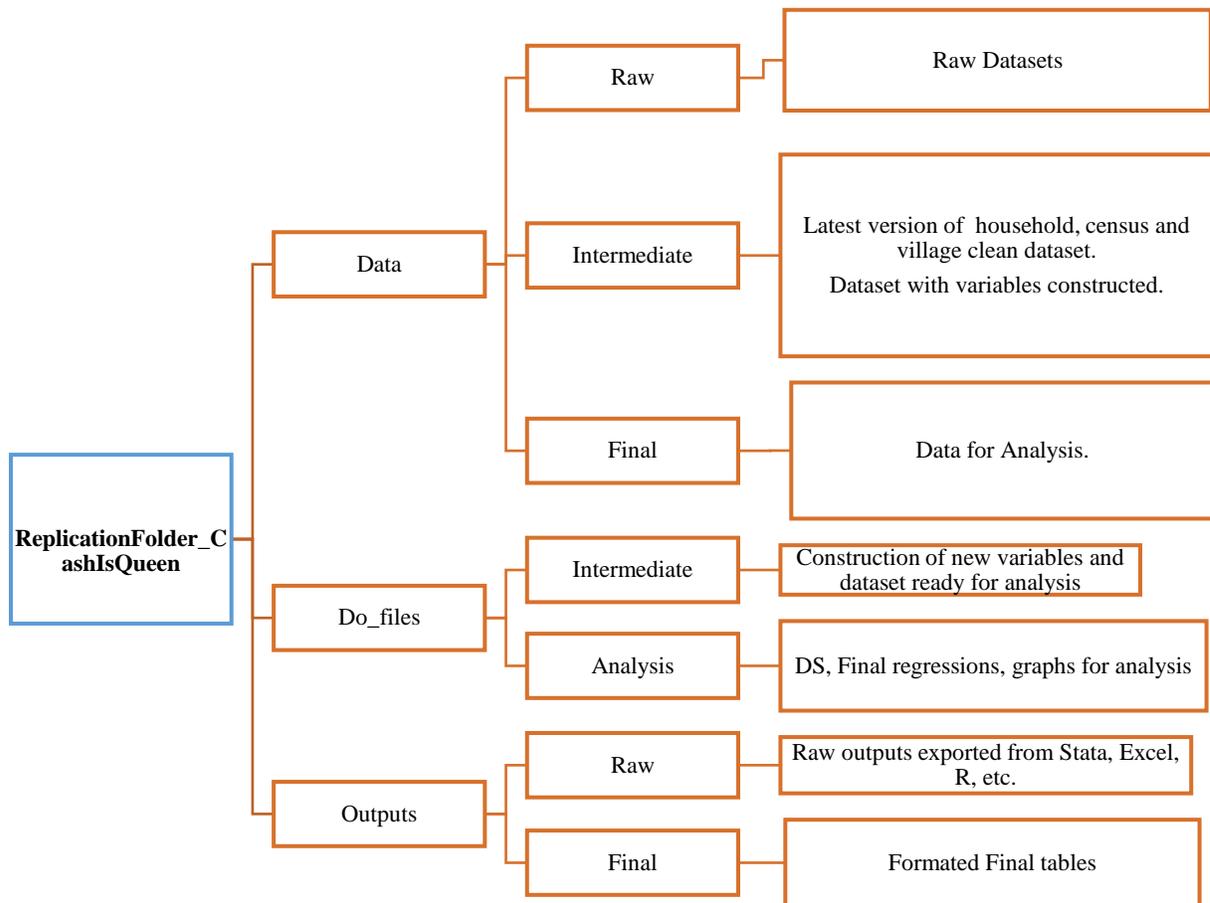
List of Analysis Codes and Outputs

All code can be executed from the Master.do script, providing a centralized approach for running the full analysis. Additionally, codes for individual exhibits are designed to be executed independently, with the condition that folder globals and custom programs are correctly set up in both the master do-file and the respective section do-files. This flexibility ensures the accurate reproduction of all numerical data presented in the paper. For ease of reference, a table is included which specifies the name of each output as mentioned in the paper, its location, the program responsible for its execution, and the individual output's file name in the folder

Main Analysis			
Folder path for individual dofiles: \$dofiles/Analysis/MainAnalysis			
Unique file for tables: \$outputs_raw/Tables\CashIsQueen_MainTables_`svdate'.xlsx			
Exhibit	Input dataset	Program	Individual Output file
Table 1	FNLpdataWideAnalysis.dta	RandomizationBalance	T1_CovariateBalanceFull
Table 1	FNLpdataWideAnalysis.dta	RandomizationBalance	T1_CovariateBalanceRDD
Table 2	FNLpdataLongAnalysis.dta	RegressionAnalysis.do	T2_NonFarmEnterprise
Table 3	FNLpdataWideAnalysis.dta	RegressionAnalysis.do	T3_AgricultureOutcomes
Table 4	FNLpdataLongAnalysis.dta	RegressionAnalysis.do	T4_FoodSecurityConsumption
Table 5	FNLpdataWideAnalysis.dta	RegressionAnalysis.do	T5_WomenEmpowermentDM
<i>Note: The elasticities computed at the bottom of each table follow Bellemare and Wichman (2020). These are exported to a new spreadsheet with the same name as the individual output file, appended with _Elas.</i>			
Appendix Analysis			
Folder path for individual dofiles: \$dofiles/Analysis/AppendixAnalysis			
Unique file for tables: \$outputs_raw/Tables\CashIsQueen_AppendixTables_`svdate'.xlsx			
Exhibit	Input dataset	Program	Individual Output file
Table A2 (PanelA)	FNLpdataWideAnalysis	AppendixA	TA2_PanelA
Table A2 (PanelB)	FNLpdataWideAnalysis	AppendixA	TA2_PanelB
Table A3	FNLpdataWideAnalysis	AppendixA	TA3_AttritionEV
Table A4	FNLpdataWideAnalysis	AppendixA	TA4_AttritionVV
Table A5	FNLP_VillageLevel	AppendixA	TA5_CovariateBalanceVillageFull
Table A5	FNLP_VillageLevel	AppendixA	TA5_CovariateBalanceVillageRDD
Table A6	FNLpdataLongAnalysis	AppendixA	TA6_NonFarmEnterpriseFull /TA6_NonFarmEnterpriseRDD
Table A7	FNLpdataWideAnalysis	AppendixA	TA6_AgricultureFull/TA6_Agriculturerdd
Table A8	FNLpdataLongAnalysis	AppendixA	TA6_FoodSecurityConsumptionFull /TA6_FoodSecurityConsumptionNRDD
Table A9	FNLpdataWideAnalysis	AppendixA	TA6_WomenEmpowermentFull /TA6_WomenEmpowermentRDD
Figure B2	FNLpdataWideAnalysis	Appendix B	FB2_MoranI_Profits /FB2_MoranI_Expenditures
Table B1	VillageDistancesHH	Appendix B	TB1_DS_VillageDistance.tex
Figure B3	VillageDistancesHH	Appendix B	FB3_HistogramDistVillage
Table B2	HHSpatialLongData	Appendix B	TB2_DSRradius_PanelA/ TB2_DSRradius_PanelP
Figure B5	HHSpatialLongData	AppendixB_Rscript	FB5_PET_CTVillages / FB5_PEV_ALLVillages /FB5_HH_ALLVillages
Figure B6	FNLpdataLongAnalysis	Appendix B	FB6_NonFarmEnterprise /FB6_BusinessProfits
Table B3	FNLpdataLongAnalysis	Appendix B	TB3_Business_Panel_Radius /
Table B4	FNLpdataWideAnalysis	Appendix B	TB4_SUVTA_PanelA /TB4_SUVTA_PanelB
Table C1	FNLpdataWideAnalysis	Appendix C	TC1_DSBusinessMid_Ttest / TC1_DSBusinessMid_Norm/ TC1_DSBusinessEnd_Ttest/ TC1_DSBusinessEnd_Norm
Table C2	FNLpdataWideAnalysis	Appendix C	TC2_BusinessDynamicsTtest / TC2_BusinessDynamicsNorm
Table C3	FNLpdataWideAnalysis	Appendix C	TC3_Agglomeration

Table C4	FNLpdataWideAnalysis	Appendix C	TC4_TypeBusiness
Table C5	FNLpdataWideAnalysis	Appendix C	TC5_AccessFinanceTtest
Figure D1	FNLpdataWideAnalysis	Appendix D	FD1_LocalRandomizationGraph
Table D1	FNLpdataWideAnalysis	Appendix D	TD1_CovariateBalanceRDDTtest /TD1_CovariateBalanceRDDNorm
Table D2	FNLpdataWideAnalysis	Appendix D	TD2_MultinomialCoeff
Table D2	FNLpdataWideAnalysis	Appendix D_Rscript	FD2_CovariateBalance
Table E1	FNLpdataLongAnalysis	Appendix E	TE1_BusinessPanel
Table E2	FNLpdataWideAnalysis	Appendix E	TE2_Multiplier (It exports main set of stats but the rest is computed as described in the appendix paper)
Table E3	FNLpdataLongAnalysis	Appendix E	TE3_CondProfits
Table E4	FNLpdataWideAnalysis	Appendix E	TE4_LocalPrices_Ttest/TE4_LocalPrices_Norm
Table F1	FNLpdataLongAnalysis	Appendix F	TF1_Business_Panel
Table F2	FNLpdataLongAnalysis	Appendix F	TF2_LFPWife
Table F3	FNLpdataLongAnalysis	Appendix F	TF3_LFPHusband
Table F4	FNLpdataWideAnalysis	Appendix F	TF4_WageEmployment
Table F5	FNLpdataWideAnalysis	Appendix F	TF5_AgOutcomes
Table F6	FNLpdataWideAnalysis	Appendix F	TF6_AgNonLabInp
Table F7	FNLpdataWideAnalysis	Appendix F	TF7_AgLabInp
Table F8	FNLpdataWideAnalysis	Appendix F	TF8_AgUse
Table F9	FNLpdataLongAnalysis	Appendix F	TF9_Consump_Panel
Table F10	FNLpdataWideAnalysis	Appendix F	TF10a_InvestmentsAppendix /F10b_InvestmentsAppendix
Table F11	FNLpdataWideAnalysis	Appendix F	TF11a_Assets /TF11b_Assets
Table F12	FNLpdataWideAnalysis	Appendix F	TF12_TimeUse
Table F13	FNLpdataWideAnalysis	Appendix F	TF13_NormsFemale
Table F14	FNLpdataWideAnalysis	Appendix F	FT14_NormsMale

Folder Structure: The diagram 1 provides a breakdown of the folder organization and displays the structure of the different folders.



Data folder structure

+---Final

+---Intermediate

+---Raw

- FNLP_Endline_ChildAnthroData.dta - Endline will be available after embargo is lifted
- FNLP_HouseholdFullData.dta - Includes baseline, midline, and endline rounds, will be available after embargo is lifted
- FNLP_Treatment.dta - Baseline - includes manual modifications, full list of variable selected and names in above. Will be included in the package when embargo is lifted.
- FNLP_VillageLevel.dta -Baseline - includes manual modifications, full list of variable selected and names in above. Will be included in the package when embargo is lifted.
- FNLP_PPI_CensusData.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.

- HHSpatialLongData.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- HHSpatialWideData.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- SUVTADistancesVillages.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- VillageDistancesHH.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- VillagePovertyIntensityData.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- VulnerablesMatrixDatafromR.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- DistancesMarketData.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- DistancesWithinVillage.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- CensusData_VulnerabilityHH.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.
- DistancesHHtoVillageType.dta - Restricted. Not publicly available due to inclusion of GPS coordinates.

Author Contact Information

- Author: Paula Gonzalez
- Email: pgonzalezmartine@worldbank.org