Dataverse Readme

Created by Ryan Jablonski

Last updated September 11, 2023

This dataverse contains all the data and scripts necessary to replicate the results in "What Politicians Don't Know Can Hurt You: The Effects of Information on Politicians' Spending Decisions." Additionally, you will find IRB letters, original survey scripts and full supplementary materials. As a condition of using these data, we ask that you do not attempt to identify respondents in this study or use any of these data in a manner that is inconsistent with the original study or IRB.

All code was originally run on RStudio Version RStudio.2022.07.1+554 with R 4.2.0 for Windows 10 x64. All packages were up-to-date as of 2023-04-01. To ensure that future packages do not break this replication code, we have often used the "groundhog" package. Groundhog creates local archived version of packages to ensure that the packages used in the original analysis remain consistent. See details and troubleshooting see here. If loading packages with groundhog throws errors, it may help to restart R and begin again with a fresh R environment.

You will also need to be able to produce rStudio rMarkdown documents. rMarkdown is a way to output latex formatted documents using R. See details on rMarkdown here. Installing the latest version of rStudio should provide this functionality.

For questions on any of these materials, you can email Ryan Jablonski at r.s.jablonski@lse.ac.uk and Brigitte Seim at bseim@ad.unc.edu.

A. The dataverse is structured as follows:

1. The "./" root folder contains the markdown file necessary to generate all results ("GenerateTablesFigures.rmd") and supplementary information. The "Replication Notes and Output.pdf" file contains the output of this rMarkdown file.
2. The "./Approval" folder contains IRB approval documentation.
3. The "./Data" folder contains the cleaned and merged files used to replicate the analysis. Note that these files will be replaced if you re-compile using the RunAllScripts.R file.
4. The "./Original Source Files" folder contains the original survey files and codebooks received from our technology partner (rtSurvey) or our enumeration team. This source files are un-edited except with the exception that we have replaced with anonymous hashes those values which are potentially identifying or especially sensitive. This folder also contains versions of the source files which have been merged with anonymous identifiers.
5. The "./Generate Survey Maps" contains the code necessary to generate all the treatment maps we used in this study, including those provided as sample maps in the text (Figure 5, Figures S20-S27).
6. The "./Sample Maps" folder contains the source files used to generate our sample maps in ArcGIS (Figure 4).
7. The "./Merge and Clean Data" folder contains the code necessary to merge the analysis files from original source files and treatment sample files and to code all variables for analysis.

B. Notes on anonymization

In accordance with our IRB, you will note that all variables which uniquely identify respondents, or about particularly sensitive subjects are placed with a secure and random key. Because of this anonymization procedure, some chunks of code cannot be run. For example, several merge

procedures require exact coordinates of respondents or schools. Where this is the case, we have included the original code for reference and separately generated the intermediate file that would have been created from the original code. We have attempted to carefully document our procedures in the source code.

C. Instructions for replicating tables and figures in the text (except maps)

In order to replicate the figures and tables in the main text and supplementary materials (except maps, as discussed below), you should run the rMarkdown file GenerateTablesFigures.rmd. This code is designed to be compiled with rMarkdown. This file will generate a pdf with a labelled list of the figures and tables. This file will also create a folder "./output/" which will contain formatted latex and pdf files for each figure and table.

In order to execute this file, you will need to first the "working_dir" variable with the location of the root replication folder on your local computer. Also note that the setup code will install new local versions of all needed packages. You may wish to do this in advance of running this code.

For further details on this file, including descriptions of the analysis source files in the "./Data" folder, please see the notes in the markdown output (Replication Notes and Output.pdf).

D. Instructions for rebuilding the analysis files from original source files

We have also included all the files necessary to merge from the original source survey and sampling files and create all variables for analysis. If you wish to do this, you will need to run "./Merge and Clean Data/RunAllScripts.R". This master script runs all the secondary scripts necessary to clean and merge the source data in the "input "folder and setup all analysis variables (with the exception of Figure S19 as noted in Section E below).

Before you run this file, you will need to set the "working_dir" variable with the location of the "Merge and Clean" folder on your local computer. You may also wish to confirm that you have loaded all the packages needed by the scripts listed below.

When you run this script, it will replace all the files in the "./Data" folder with the output from this script. This will allow you to regenerate the tables and figures using the refreshed data (see instructions in Section C).

This script executes the following files:

1. CleanSurveyData.R This script does some basic cleaning on the survey files and merges the survey files with the sample file containing treatment indicators. This file also creates some variables which are later used for analysis and setup.
2. CreateKnowledgeVariables.R This script codes responses to the post-treatment quiz.
3. CreatePopulationVariables.R This script merges in school-specific population variables.
4. CreatePovertyVariables.R This script merges in school-specific poverty variables.
5. CreateCovariates.R This script sets up most of the pre-specified variables used in the analysis and merges in some supplementary data on polling-station-level voting results. This script also implements our pre-specified imputation procedure for the covariates used as controls.

6. CreateDistanceVariables.R This script codes the distance between schools and politician hometowns. (Note that since no coordinate are included in the source files, much of this code is commented out. See discussion above and in the script.)
7. SetupCitizenSurveyData.R This script sets up some variables to generate Figure 2.
8. SetupQuizData.R This script generates a file (quiz_with_covariates.csv). This file is a list of all the schools included in the post-treatment quiz along with details on whether the respondent indicated knowledge of the school. We use this file to generate Table S34 in the supplementary materials.
9. DoPowerSimulations.R This script does the simulations necessary to generate Figure S1 and S2. *Note that this script does not run by default when you run the RunAllScripts.R file*. The reason for this is that these simulations take a *very* long time to run (in part because these simulations rely on real data and require considerable file manipulation).
10. DoFinalProcessing.R This file does some final processing, including coding some compliance variables, removing variables not needed for analysis, and generating treatment-specific variables.

E. Instructions for cleaning the head teacher file

The RunAllScripts.R will not run the code necessary to code variables for the head teacher survey (which we use for Figure S19). The code for this cleaning was programmed in Stata. In order to clean the original head teacher survey file (MALAWI_HEAD_TEACHER_BASELINE.dta), you will need to run the "Merge and Clean Data/ recodingteacherpriorities.do" script. This will copy a new version of the head teacher survey to the "./Data" folder with the variables necessary to generate Figure S19. This step is only necessary if you wish to replicate Figure S19 from the original source file. Prior to running this do file you will need to set the working directory on line 1 to the location of the root folder of the replication files on your local machine.

F. Instructions for generating treatment maps (including Figure 5 and Figures S20-S27)

The rMarkdown file does not generate the example treatment maps included in the text (Figure 5 and Figures S20-S27). These maps were generated as part of the original survey creation procedure which is somewhat involved. If you wish to regenerate these maps, you will need to re-generate the treatment messages from original sample and constituency boundary files. To do this, you can use "./Generate Survey Maps/CreateSurveyConstituencyMaps.R" (for Constituency maps) and "./Generate Survey Maps/CreateSurveyWardMaps.R" (for Ward maps).

In order to run these scripts, you must first set the "working_dir" variable in each of these scripts to the location of the "Generate Survey Maps" folder on your local computer. Additionally, and since the script relies on the google mapping API, you must set the "google_api_key" variable to a valid google mapping API key. You can view the details of how to setup this key at https://rdrr.io/cran/ggmap/man/register_google.html. This will require you to setup a google billing account, though, at the time of writing, google does not charge for the level of access required by these scripts.

Once you run these scripts, they will generate treatment maps for all sampled wards and constituencies in the "./Generate Survey Maps/maps/" folder. Additionally these will copy the example maps (Figure 5 and Figures S20-S27) to the "./Generate Survey Maps/maps."

Note that these scripts generate over a 1,000 maps and may take some time to run.

G. Instructions for generating sample maps (Figure 4)

Figure 4 was generated using ArcGIS. In the folder "./Sample Maps" you can see the source files which can be used to regenerate this figure in ArcGIS. Within this folder you will find a readme.txt file with exact instructions for how to compile this figure in ArcGIS.

H. Descriptions of source files:

ID	File	Description	Source
1	"./Original Survey Files/AID_MALAWI_COUNCILOR.dta"	The original councillor survey file produced by rtSurvey with identifying and sensitive variables removed.	rtSurvey (technology partner)
2	"./Original Survey Files/AID_MALAWI_COUNCILOR.pdf"	Codebook for councillor survey file	rtSurvey (technology partner)
3	"./Original Survey Files/AID_MALAWI_COUNCILOR_withids.dta"	A version of the councillor survey file containing anonymous keys for merging	rtSurvey (technology partner)
4	"./Original Survey Files/AID_MALAWI_MP.dta"	The original MP survey file produced by rtSurvey with identifying and sensitive variables removed.	rtSurvey (technology partner)
5	"./Original Survey Files/AID_MALAWI_MP.pdf"	Codebook for MP survey file	rtSurvey (technology partner)
6	"./Original Survey Files/AID_MALAWI_MP_withids.dta"	A version of the MP survey file containing anonymous keys for merging	rtSurvey (technology partner)
7	"./Original Survey Files/CouncilorFollowUpSurvey.csv"	Answers to a follow-up phone survey with councillors	Enumeration team
8	"./Original Survey Files/FollowUpSurvey Questionnaire.docx"	Questionaire for the follow-up phone survey	Principal investigators
9	"./Original Survey Files/MALAWI_CITIZEN_BASELINE.dta"	The original citizen survey file produced by rtSurvey with identifying and sensitive variables removed.	rtSurvey (technology partner)
10	"./Original Survey Files/MALAWI_CITIZEN_BASELINE.pdf"	Codebook for citizen survey file	rtSurvey (technology partner)
11	"./Original Survey Files/MALAWI_CITIZEN_BASELINE_withids.dta"	A version of the citizen survey file containing anonymous keys for merging	rtSurvey (technology partner)

12	"./Original Survey Files/MALAWI_HEAD_TEACHER_BASELINE.dta"	The original teacher survey file produced by rtSurvey with identifying and sensitive variables removed.	rtSurvey (technology partner)
13	"./Original Survey Files/MALAWI_HEAD_TEACHER_BASELINE.pdf"	Codebook for teacher survey file	rtSurvey (technology partner)
14	"./Merge and Clean Data/input/2014 councillor election results by ps.csv"	2014 polling station level election results for local councillor elections	Malawi Election Commission
15	"./Merge and Clean Data/input/2014 councillor election results by school.csv"	2014 polling station level election results for local councillor elections with school ids	Malawi Election Commission
16	"./Merge and Clean Data/input/2014 mp election results by ps.csv"	2014 polling station level election results for MP elections	Malawi Election Commission
17	"./Merge and Clean Data/input/2014 mp election results by school.csv"	2014 polling station level election results for MP elections with school ids	Malawi Election Commission, Ministry of Education
18	"./Merge and Clean Data/input/2017-02-22-randomized_const.csv"	Original sampling file containing treatment assignments for MP surveys	Principal investigators
19	"./Merge and Clean Data/input/2017-02-22local-randomized.csv"	Original sampling file containing treatment assignments for councillor surveys	Principal investigators
20	"./Merge and Clean Data/input/categorization_of_how_useful_question.csv"	A spreadsheet containing the manual catagorization of open-ended responses to a question about the usefulness of information in the survey	PI and RA coding
21	"./Merge and Clean Data/input/coding_of_enumerator_comments_about_schools.csv"	A spreadsheet containing the manual catagorization of open-ended responses to a question to enumerators about problems in the surveys mentioned by respondents	PI and RA coding
22	"./Merge and Clean Data/input/lc_children_school_details.csv"	A list of schools in which politicians have family based on comparison of survey responses to school names (councillor survey)	PI and RA coding
23	"./Merge and Clean Data/input/lc_results.csv"	Results of local councillor elections at the Ward level	Malawi Election Commission
24	"./Merge and Clean Data/input/map_h_schools_lc.csv"	List of all the schools included in the post-treatment quiz (councillor survey)	Survey sampling
25	"./Merge and Clean Data/input/map_h_schools_mp.csv"	List of all the schools included in the post-treatment quiz (MP survey)	Survey sampling

26	"./Merge and Clean Data/input/mp_children_school_details.csv"	A list of schools in which politicians have family based on comparison of survey responses to school names (MP survey)	PI and RA coding
27	"./Merge and Clean Data/input/mp_results.csv"	Results of MP elections at the constituency level	Malawi Election Commission
28	"./Merge and Clean Data/input/mp_results_wide.csv"	Results of MP elections at the polling station level	Malawi Election Commission
29	"./Merge and Clean Data/input/Past Aid Interventions in Malawi Primary Schools with id.csv"	List of schools with foreign aid projects	RA interviews with donor staff in Malawi
30	"./Merge and Clean Data/input/polling_station_data.csv"	Results of presidential election at the polling station level	Malawi Election Commission
31	"./Merge and Clean Data/input/school_home_distances.csv"	Distances between politician hometowns and schools in the survey	This contains output from the CreateDistanceVariables.R script which cannot be run due to the anonymization of school coordinates
32	"./Merge and Clean Data/input/school_home_distances_all_lc_schools.csv"	Distances between politician hometowns and schools for all schools (councillors only)	This contains output from the CreateDistanceVariables.R script which cannot be run due to the anonymization of school coordinates
33	"./Merge and Clean Data/input/school_home_distances_all_mp_schools.csv"	Distances between politician hometowns and schools for all schools (MPs only)	This contains output from the CreateDistanceVariables.R script which cannot be run due to the anonymization of school coordinates
34	"./Merge and Clean Data/input/school_treatment_information_lc.csv"	School details used to populate the LC treatment maps including only schools used in the treatment	Malawi Election Commission and Ministry of Education
35	"./Merge and Clean Data/input/school_treatment_information_lc_all_schools.csv"	School details used to populate the LC treatment maps including all eligible schools	Malawi Election Commission and Ministry of Education
36	"./Merge and Clean Data/input/school_treatment_information_mp.csv"	School details used to populate the MP treatment maps including only schools used in the treatment	Malawi Election Commission and Ministry of Education

37	"./Merge and Clean Data/input/school_treatment_information_mp_all_schools.csv"	School details used to populate the MP treatment maps including all eligible schools	Malawi Election Commission and Ministry of Education
40	"./Merge and Clean Data/input/worldpop_const.csv"	Constituency population statistics	WorldPop Project. http://worldpop.org
41	"./Merge and Clean Data/input/worldpop_poverty_school_lc.csv"	school-level poverty estimates for schools eligible for inclusion in local councillor survey	WorldPop Project. http://worldpop.org Note that this would be an output from the CreatePovertyVariables.R but the merge cannot be done dynamically due to anonymization of coordinates.
42	"./Merge and Clean Data/input/worldpop_poverty_school_mp.csv"	school-level poverty estimates for schools eligible for inclusion in MP survey	WorldPop Project. http://worldpop.org Note that this would be an output from the CreatePovertyVariables.R but the merge cannot be done dynamically due to anonymization of coordinates.
43	"./Merge and Clean Data/input/worldpop_school_lc.csv"	school-level population estimates for schools eligible for inclusion in local councillor survey	WorldPop Project. http://worldpop.org Note that this would be an output from the CreatePopulationVariables.R but the merge cannot be done dynamically due to anonymization of coordinates.
44	"./Merge and Clean Data/input/worldpop_school_mp.csv"	school-level population estimates for schools eligible for inclusion in MP survey	WorldPop Project. http://worldpop.org Note that this would be an output from the CreatePopulationVariables.R but the merge cannot be done dynamically due to anonymization of coordinates.
45	"./Merge and Clean Data/input/worldpop_ward.csv"	Ward population statistics	WorldPop Project http://worldpop.org
46	"./Generate Survey Maps/Constituency_Sampling02202017.csv"	Treatment sampling file	Principal investigators

47	"./Generate Survey Maps/gadm36_KEN_2_sp.rds"	Shapefile of Kenyan Constituencies (used to generate example maps shown in the survey)	https://gadm.org/
48	"./Generate Survey Maps/LC School Information Template 02202017b.csv"	School details used to populate the LC treatment maps including only schools used in the treatment	Malawi Election Commission and Ministry of Education
49	"./Generate Survey Maps/MP School Information Template 02202017b.csv"	School details used to populate the MP treatment maps including only schools used in the treatment	Malawi Election Commission and Ministry of Education
50	"./Generate Survey Maps/Name_corrections.xlsx"	Some corrections to politician names (mostly due to bi-elections)	PI and RA coding
51	"./Generate Survey Maps/shapefiles"	Constituency and ward shapefiles	Malawi Election Commission
52	"./Generate Survey Maps/Ward_Sampling02202017.csv"	Treatment sampling file	Principal investigators
53	"./Sample Maps\Source files/const_sample.txt"	A list of all sampled constituencies	Output of the enumeration process
54	"./Sample Maps\Source files/ward_sample.txt"	A list of all sampled wards	Output of the enumeration process
55	"./Sample Maps\Source files/CONST"	Constituency boundary shapefile	Malawi Election Commission
56	"./Sample Maps\Source files/gadm36_MWI_shp"	Administrative boundary shapefile	https://gadm.org/
57	"./Sample Maps\Source files/Malawi_2014_Ward_Boundary"	Ward shapefile	Malawi Election Commission