README for "Reexamining the Effect of Refugees on Civil Conflict: A Global Subnational Analysis"

Yang-Yang Zhou* and Andrew Shaver†

April 30, 2021

1 Instructions for Replicating All Findings

To replicate all findings, please follow the instructions below. Since the manuscript and appendix are `.Rnw` files with inline code, recompiling the documents also re-run all of the analyses in the manuscript and appendix. After installing the necessary packages, compilation and replication of all findings will take several hours.

1. **Check if the required packages are installed:** Run `install_required_packages.R` to check if all required packages are installed. If any are not installed, the script will automatically install them.

2. **Open main analysis file:** Open `RefugeesCivilWar_Paper_APSRfinal.Rnw` in RStudio.

   (a) **Select necessary `knitr` settings:** Once in RStudio, select RStudio -> Preferences -> Sweave from the toolbar. Then, set "Weave Rnw files using:" to "knitr" in the drop-down menu, and set "Typeset LaTeX int PDF using:" to "XeLaTeX" in the drop-down menu. Click "Apply" at the bottom of the window, and then 'OK".

3. **Compile main analysis file:** Click "Compile PDF" at the top of the RStudio window. This will run all analyses and compile the main manuscript. This should take about an hour. You may run into errors at the final LaTeX compilation stage having to do with citations. This is addressed by pressing "Compile PDF" two more times.

4. **Open appendix file:** Open `RefugeesCivilWar_SI_APSRfinal.Rnw` in RStudio.

5. **Compile appendix file:** Click "Compile PDF" at the top of the RStudio window. This will run all analyses and compile the main manuscript. **This will take several hours to run.** You may run into errors at the final LaTeX compilation stage having to do with citations. This is addressed by pressing "Compile PDF" two more times.[1]

*Assistant Professor, Department of Political Science, University of British Columbia, yangyang.zhou@ubc.ca, www.yangyangzhou.com

†Assistant Professor, Department of Political Science, University of California, Merced. Affiliate, CISAC, Stanford University. Email: ashaver@ucmerced.edu, www.andrewcshaver.com

[1]To get the cross-references working correctly, you may have to run Step (3) an additional time.

2 Structure of Replication Archive

In addition to the README, `install_required_packages.R`, and the manuscript `.Rnw` files, which are explained above, there are two folders in the top level of the replication archive: `Paper_Inputs`, and `Data Cleaning`. We provide more detail on the folder structure below.

1. **Paper_Inputs:** This folder contains several files called by the manuscript `.Rnw` files during compilation. These files include:

   - `RefCivFunctions.R`: Functions and libraries sourced by each manuscript file.
   - `world_admin_1998.*`: Global shapefile called for descriptive plots.

   The folder also includes cleaned datasets generated by `Data Cleaning/RefugeesConflict_DataPreparation_A` and pre-run regression models for especially large regression objects. We also include the code in the `.Rnw` files for you to re-run all of these regression models from scratch. Code chunks with these models have the option `eval = FALSE` set, so that compilation defaults to using the saved objects, but you can set them to `eval = TRUE` to rebuild the models from scratch.

2. **Data Cleaning:** This folder contains the code file (`RefugeesConflict_DataPreparation_APSR.R`) that intakes all of the raw data sources (more detail provided in Section 3), and outputs `.csv` files that we then use for analysis. It also includes the raw file provided by UNHCR with georeferenced refugee locations, in `wrl_ppl_poc_07_01_2020.xls` (more detail below).

3 Data Sources

The UNHCR georeferenced refugee locations dataset that we introduce in this paper can be accessed in our replication files. Specifically, the raw file from the UNHCR is `wrl_ppl_poc_07_01_2020.xls`, and the data that we additionally processed (see description in the manuscript) for use in the main analysis is `camps_settlements_processed.csv`.

Although we include our data cleaning code for transparency, we do not include within our replication files the following data that we use as inputs into the cleaning code, because they are either proprietary or already available publicly online. We detail how to find these data below.

UCDP GED

The Uppsala Conflict Data Program's (UCDP) Georeferenced Event Dataset (GED) (version 19.1) Sundberg and Melander (2013) can be accessed at:

ucdp.uu.se/downloads/olddw.html

UCDP's GED separate Syria dataset (version 652.1601.1911) can be accessed at:

ucdp.uu.se/downloads/olddw.html

PRIO Conflict Sites

The Peace Research Institute Oslo (PRIO) Conflict Site (1989-2008) data (version 3.0) Dittrich Hallberg (2012) can be accessed at:

prio.org/Data/Armed-Conflict/Conflict-Site/

wzoneData

The wzoneData: Zones of Armed Conflicts data (corresponding to UCDP GED version 19.1) developed by Kikuta (2020) are available at:

dataverse.harvard.edu/dataverse/kyosukekkt

CShapes Shapefile

The CShapes shapefile (version 0.6) (Weidmann, Kuse and Gleditsch, 2010) can be accessed at:

nils.weidmann.ws/projects/cshapes/shapefile.html

Administrative Unit Shapefile

The 1998 1st level administrative ("province") boundaries used in this project can be procured from ESRI:

www.esri.com

Universal Transverse Mercator Shapefile

The global Universal Transverse Mercator grid shapefile (updated as of 11/11/2018) can be access at:

hub.arcgis.com/datasets/esri::world-utm-grid

PRIO-GRID

PRIO-GRID (Tollefsen, Strand and Buhaug, 2012) data (version 2.0) can be accessed at:

grid.prio.org/download

The PRIO-GRID polygon (cell) shapefile can be accessed at:

grid.prio.org/extensions

The PRIO-GRID centroids shapefile no longer appears on PRIO-GRID website. Individuals interested in obtaining these data might reach out to the following contacts listed on the PRIO-GRID website: andreas@prio.org; marteg@prio.org.

Terrain Ruggedness

Data on terrain ruggedness (original version) from Shaver, Carter and Shawa (2016) can be accessed at:

`https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/WXUZBN`

4 OS, R, and Package Version Information

This analysis was compiled using R version 4.0.2, on a Mac OSX running macOS Big Sur 11.1. Additional information on the computing environment, and exact package versions, can be found below.

```
> sessionInfo()
R version 4.0.2 (2020-06-22)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS  10.16

Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      parallel  stats     graphics  grDevices utils
[7] datasets  methods   base

other attached packages:
 [1] ggmap_3.0.0        maptools_1.0-1     mapdata_2.3.0
 [4] maps_3.3.0         rgdal_1.5-19       sp_1.4-2
 [7] ggcorrplot_0.1.3   patchwork_1.0.1    lattice_0.20-41
[10] xtable_1.8-4       gridExtra_2.3      RColorBrewer_1.1-2
[13] PanelMatch_1.0.1   texreg_1.37.5      doMC_1.3.6
[16] iterators_1.0.13   foreach_1.5.1      forcats_0.5.0
[19] stringr_1.4.0      dplyr_1.0.5        purrr_0.3.4
[22] readr_1.3.1        tidyr_1.1.1        tibble_3.0.4
[25] ggplot2_3.3.2      tidyverse_1.3.0    estimatr_0.22.0
[28] reshape2_1.4.4     knitr_1.29         stargazer_5.2.2
[31] lubridate_1.7.9    arm_1.11-2         lme4_1.1-23
[34] Matrix_1.2-18      MASS_7.3-53

loaded via a namespace (and not attached):
 [1] minqa_1.2.4          colorspace_2.0-0    rjson_0.2.20
 [4] ellipsis_0.3.1       rsconnect_0.8.16    htmlTable_2.0.1
 [7] base64enc_0.1-3      fs_1.5.0            rstudioapi_0.13
[10] fansi_0.4.1          xml2_1.3.2          codetools_0.2-16
[13] splines_4.0.2        Formula_1.2-3       jsonlite_1.7.2
[16] nloptr_1.2.2.2       CBPS_0.21           broom_0.7.0
[19] cluster_2.1.0        dbplyr_1.4.4        png_0.1-7
[22] compiler_4.0.2       httr_1.4.2          backports_1.2.1
[25] assertthat_0.2.1     cli_2.2.0           acepack_1.4.1
[28] htmltools_0.5.0      tools_4.0.2         coda_0.19-3
[31] gtable_0.3.0         glue_1.4.2          Rcpp_1.0.5
[34] cellranger_1.1.0     raster_3.3-13       vctrs_0.3.5
[37] nlme_3.1-148         xfun_0.16           rvest_0.3.6
[40] lifecycle_1.0.0      statmod_1.4.34      scales_1.1.1
[43] hms_0.5.3            yaml_2.2.1          rpart_4.1-15
[46] latticeExtra_0.6-29  stringi_1.4.6       checkmate_2.0.0
[49] boot_1.3-25          shape_1.4.5         bitops_1.0-6
[52] RgoogleMaps_1.4.5.3  rlang_0.4.10        pkgconfig_2.0.3
[55] evaluate_0.14        htmlwidgets_1.5.1   tidyselect_1.1.0
```

```
[58] plyr_1.8.6          magrittr_2.0.1      R6_2.5.0
[61] generics_0.0.2      Hmisc_4.4-0         DBI_1.1.0
[64] pillar_1.4.7        haven_2.3.1         foreign_0.8-80
[67] withr_2.3.0         survival_3.1-12     abind_1.4-5
[70] nnet_7.3-14         modelr_0.1.8        crayon_1.3.4
[73] rmarkdown_2.3       jpeg_0.1-8.1        readxl_1.3.1
[76] data.table_1.13.4   blob_1.2.1          reprex_0.3.0
[79] digest_0.6.27       numDeriv_2016.8-1.1 MatchIt_4.0.1
[82] munsell_0.5.0       glmnet_4.0-2
```

References

Dittrich Hallberg, Johan. 2012. "PRIO Conflict Site 1989–2008: A Geo-Referenced Dataset on Armed Conflict." *Conflict Management and Peace Science* 29(2):219–232.

Kikuta, Kyosuke. 2020. "A new geography of civil war: a machine learning approach to measuring the zones of armed conflicts." *Political Science Research and Methods* pp. 1–19.

Shaver, Andrew, David B Carter and Tsering Wangyal Shawa. 2016. "Terrain Ruggedness and Land Cover: Improved Data for Most Research Designs." *Conflict Management and Peace Science* .

Sundberg, Ralph and Erik Melander. 2013. "Introducing the UCDP georeferenced event dataset." *Journal of Peace Research* 50(4):523–532.

Tollefsen, Andreas, Havard Strand and Halvard Buhaug. 2012. "PRIO-GRID: A Unified Spatial Data Structure." *Journal of Peace Research* 49(2):363–374.

Weidmann, Nils B, Doreen Kuse and Kristian Skrede Gleditsch. 2010. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36(1):86–106.