November 3, 2020

Readme file for "3G Internet and Confidence in Government"

by Sergei Guriev, Nikita Melnikov, and Ekaterina Zhuravskaya

The replication folder contains the datasets and codes, which can be used to (i) generate all the datasets from the raw data and (ii) replicate all the results, i.e., the tables and figures in the paper and the online appendix.

In order to generate the datasets used in the analysis and replicate all of the results in the paper, one needs to perform the following steps:

1. Download the replication materials from the repository. Note that to fully replicate the results you also need to obtain the proprietary data files from Gallup, Collins Bartholomew, International Telecommunication Union, and the World Wide Lightning Location Network (WWLLN). We provide detailed instructions on how to obtain the proprietary data files below.

2. Create a **replication** folder in which you will store the data files, dofiles, and the tables and figures with the results. Below, we provide description of each file. The replication folder should have the following structure of subfolders:

   - **original_data**—a subfolder with all the original nonproprietary data files and with a subsubfolder **original_proprietary_data**, where all original proprietary data sets need to be saved before replication can be executed.
   - **do_files**—a subfolder with all the replication dofiles.
   - **generated_data**—a subfolder where all the generated datasets are saved by the replication code.
   - **results**—a subfolder with two subsubfolders: **main_results** and **appendix**, where all the tables and figures will be saved by the replication code for the main text and for the appendix, respectively.
   - **auxiliary_data**—a subfolder with additional materials, such as maps, needed to generate data.

3. Open the **master_replicate_GMZ_3G_qje.do** dofile and re-define the path to the **replication** folder in line 3 of the file.

4. Execute the **master_replicate_GMZ_3G_qje.do** dofile from the root **replication** folder. This dofile calls on all other dofiles in the **do_files** subfolder in order to create the datasets (to be saved in **generated_data** subfolder) and then generate the results and save them in the **results** subfolder. The software that we use to execute this file is STATA 16.1.

The rest of this README file is organized as follows. We describe proprietary data in Section I. Section II describes all the data in open access, supplied with the replication package. Section III describes each of the dofiles. In Section IV, we describes the auxiliary data files, also supplied with the replication package.

I. Proprietary data files

The list of files in **original_proprietary_data** subsubfolder needed for replication to run is:

- **Gallup_original.dta**
- **brazil_3g_2018.dta**
- **3g_gwp_cy.dta**
- **3g_unw_gwp_2008_2018.dta**
- **2g_gwp.dta**
- **3g_populists.dta**
- **3g_gwp_2007_2018.dta**
- **lightnings_2005_2017_weighted.dta**
- **ambs_pc.csv**

The files with these names provided with the replication package have been modified because we cannot supply proprietary data with the replication package. In order to conduct replication, researchers need to replace these files with the files that contain real data. Below, in this section, we explain what these datasets are and how to obtain real data for each of them. For the available replication package, we made the following modifications to these files: (1) In the file **Gallup_original.dta**, we dropped all variables with the exception of two variables, WP5 (country name) YEAR_WAVE (year of the survey), that jointly indicate the country×years needed for replication, and the number of observations in the dataset. (2) In all other files from this list, we replaced all values with -99 in those variables that contain proprietary information, keeping the structure of each file intact, i.e., the number of observations, variables, and variable labels. These manipulations with the original proprietary data are made with **proprietary_data_not_available.do** dofile, which can be found in **auxiliary_data** folder.

We organize the description of these data and the instructions of how to get them by the data source: (1) Gallup World Poll (GWP) data; (2) 3G and 2G network coverage data; (3) Lightning strike frequency data; (4) ITU data.

- **Gallup World Poll (GWP) data:** the microdata from the original GWP surveys. These data are available from Gallup on a subscription basis. One can contact Gallup to purchase the data here: https://www.gallup.com/270188/contact-us-general.aspx (accessed November 3, 2020). When new surveys (i.e., data for new country×years) arrive, GWP appends the data for the new surveys to the end of the previous version of the dataset. Thus, to replicate our results, one needs to obtain GWP subscription that is sufficiently recent to include all surveys conducted in 2017. As mentioned above, the full list of country×years needed for replication is given in the file **Gallup_original.dta** provided in the replication package. These are all country×years, in which GWP conducted surveys, with full data for 2017 and all the previous years included. When the researcher obtains the GWP subscription, the data should be converted into Stata format and named **Gallup_original.dta**. (GWP also provides data in Stata format.) This dataset should replace the file under the same name in the **/original_data/original_proprietary_data** subfolder.

- **3G and 2G network coverage data:** the maps of worldwide 3G and 2G coverage. These data are available from Collins Bartholomew's Mobile Coverage Explorer on a subscription basis. One can find more information about the data and contact Collins Bartholomew to purchase a license here: https://www.collinsbartholomew.com/mobile-coverage-maps/mobile-coverage-explorer/
(accessed November 3, 2020). To fully replicate the results of the paper, a researcher needs to use Collins Bartholomew data to produce the following files:

  — **2g_gwp.dta**, the dataset with 2G coverage data for the subnational regions in the GWP, weighted by population density.

  — **3g_gwp_2007_2018.dta**, the dataset with 3G coverage data for the subnational regions in the GWP, weighted by population density.

  — **3g_unw_gwp_2008_2018.dta**, the dataset with 3G coverage data for the subnational regions in the GWP, not weighted by population density.

  — **3g_gwp_cy.dta**, the dataset with 3G coverage data for the countries in the GWP, weighted by population density.

  — **3g_populists.dta**, the dataset with 3G coverage data for the European subnational districts with the election results, weighted by population density.

  — **brazil_3g_2018.dta**, the dataset with 3G coverage data for Brazil's microregions, weighted by population density.

  Below we provide step-by-step instructions on how to produce each of these datasets from Collins Bartholomew's original data. Each of these files should be placed in the **/original_data/original_proprietary_data** subfolder.

2g_gwp.dta:

1. Convert the shapefile maps of 2G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 2G network coverage in year t is named Global_GSM_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_GSM_201712.shp represents 2G network coverage in 2018. The files that represent 2G network coverage for earlier years are as follows: Global_GSM_2009Q1.shp—2G coverage in 2010, Global_GSM_2008Q1.shp—2G coverage in 2009, Global_GSM_2007Q1.shp—2G coverage in 2008, and Global_GSM_2006Q3.shp—2G coverage in 2007. Note that in the resulting raster files areas without 2G coverage should be assigned the value of zero.
2. Open NASA's map of population density (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP, accessed November 3, 2020) and increase the resolution of the map to match the resolution of the 2G network coverage rasters.
3. For each grid cell on the map of 2G network coverage, multiply the binary value of 2G coverage in that grid cell by population density in the grid cell. The resulting raster illustrates the number of people with potential access to 2G per square kilometer.
4. Open the **level1_map.shp** and **level2_map.shp** files that can be found in **auxiliary_data** subfolder and, for each subnational region indexed by regionid_m (the ID of the subnational regions), calculate the average value of the file from step 3.
5. For each regionid_m, calculate the average value of population density.
6. Export the data from steps 4 and 5.
7. In the exported files, for each regionid_m, divide the value of population-weighted 2G network coverage by the value of population density. The resulting variable represents the share of the region's population with potential access to 2G.
8. Match the regionid_m variable (the ID of the subnational regions in the map) to the districtid variable (the ID of the subnational regions in the GWP data). To match the data, use the **map_to_gallup_level1.dta** file for the regionid_m values from **level1_map.shp** and the **map_to_gallup_level2.dta** file for the regionid_m values from **level2_map.shp**. All these files can be found in **auxiliary_data** subfolder.
9. Combine the level1 and level2 files from step 8 and drop the _merge and regionid_m variables.
10. Create the variable for the year, combine the data for different years, and save the file as **2g_gwp.dta** in the **/original_data/original_proprietary_data** directory.
11. Interpolate the data for 2011 as the average of the data in 2010 and 2012. Save.

3g_gwp_2007_2018.dta:

1. Convert the shapefile maps of 3G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 3G network coverage in year t is named Global_3G_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_3G_201712.shp represents 3G network coverage in 2018. The files that represent 3G network coverage for earlier years are as follows: Global_3G_2009Q1.shp—3G coverage in 2010, Global_3G_2008Q1.shp—3G coverage in 2009, Global_3G_2007Q1.shp—3G coverage in 2008, and Global_3G_2006Q3.shp—3G coverage in 2007. Note that in the resulting raster files areas without 3G coverage should be assigned the value of zero.
2. Open NASA's map of population density (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP , accessed November 3, 2020) and increase the resolution of the map to match the resolution of the 3G network coverage rasters.
3. For each grid cell on the map of 3G network coverage, multiply the binary value of 3G coverage in that grid cell by population density in the grid cell. The resulting raster illustrates the number of people with potential access to 3G per square kilometer.
4. Open the **level1_map.shp** and **level2_map.shp** files that can be found in **auxiliary_data** subfolder and, for each regionid_m (the ID of the subnational regions), calculate the average value of the file from step 3.
5. For each regionid_m, calculate the average value of population density.
6. Export the data from steps 4 and 5.
7. In the exported files, for each regionid_m, divide the value of population-weighted 3G network coverage by the value of population density. The resulting variable represents the share of the region's population with potential access to 3G.
8. Match the regionid_m variable (the ID of the subnational regions in the map) to the districtid variable (the ID of the subnational regions in the GWP data). To match the data, use the **map_to_gallup_level1.dta** file for the regionid_m values from **level1_map.shp** and the **map_to_gallup_level2.dta** file for the regionid_m values from **level2_map.shp**. All these files can be found in **auxiliary_data** subfolder.
9. Combine the level1 and level2 files from step 8 and drop the _merge and regionid_m variables.
10. Create the variable for the year, combine the data for all the years, and save the file as **3g_gwp_2007_2018.dta** in the **/original_data/original_proprietary_data** directory.
11. Interpolate the data for 2011 as the average of the data in 2010 and 2012. Save.

3g_unw_gwp_2008_2018.dta:

1. Convert the shapefile maps of 3G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 3G network coverage in year t is named Global_3G_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_3G_201712.shp represents 3G network coverage in 2018. The files that represent 3G network coverage for earlier years are as follows: Global_3G_2009Q1.shp—3G coverage in 2010, Global_3G_2008Q1.shp—3G coverage in 2009, Global_3G_2007Q1.shp—3G coverage in 2008, and Global_3G_2006Q3.shp—3G coverage in 2007. Note that in the resulting raster files areas without 3G coverage should be assigned the value of zero.
2. Open the **level1_map.shp** and **level2_map.shp** files that can be found in **auxiliary_data** subfolder and, for each regionid_m (the ID of the subnational regions), calculate the average value of the binary raster from step 1.
3. Export the data from steps 2.
4. Match the regionid_m variable (the ID of the subnational regions in the map) to the districtid variable (the ID of the subnational regions in the GWP data). To match the data, use the **map_to_gallup_level1.dta** file for the regionid_m values from **level1_map.shp** and the **map_to_gallup_level2.dta** file for the regionid_m values from **level2_map.shp**. All these files can be found in **auxiliary_data** subfolder.
5. Combine the level1 and level2 files from step 4 and drop the _merge and regionid_m variables.
6. Create the variable for the year, combine the data for different years, and save the file as **3g_unw_gwp_2008_2018.dta** in the **/original_data/original_proprietary_data** directory.
7. Interpolate the data for 2011 as the average of the data in 2010 and 2012. Save.

3g_gwp_cy.dta:

1. Convert the shapefile maps of 3G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 3G network coverage in year t is named Global_3G_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_3G_201712.shp represents 3G network coverage in 2018. The files that represent 3G network coverage for earlier years are as follows: Global_3G_2009Q1.shp—3G coverage in 2010, Global_3G_2008Q1.shp—3G coverage in 2009, Global_3G_2007Q1.shp—3G coverage in 2008, and Global_3G_2006Q3.shp—3G coverage in 2007. Note that in the resulting raster files areas without 3G coverage should be assigned the value of zero.
2. Open NASA's map of population density (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP) and increase the resolution of the map to match the resolution of the 3G network coverage rasters.
3. For each grid cell on the map of 3G network coverage, multiply the binary value of 3G coverage in that grid cell by population density in the grid cell. The resulting raster illustrates the number of people with potential access to 3G per square kilometer.
4. Open the **world_map_only_countries.shp** file located in the **auxiliary_data** subfolder and, for each ISO (the ID of the country), calculate the average value of the file from step 3.
5. For each ISO, calculate the average value of population density.
6. Export the data from steps 4 and 5.
7. In the exported files, for each ISO, divide the value of population-weighted 3G network coverage by the value of population density. The resulting variable represents the share of the country's population with potential access to 3G.
8. Create the variable for the year and combine the data for different years.
9. Drop the population density variable, rename "ISO" to "iso", and save the file as **3g_gwp_cy.dta**.
10. Interpolate the data for 2011 as the average of the data in 2010 and 2012.
11. Using the iso variable, merge the data with **3g_gwp_cy_without_3g.dta**. Save the file as **3g_gwp_cy.dta** in the **/original_data/original_proprietary_data** directory.

3g_populists.dta:

1. Convert the shapefile maps of 3G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 3G network coverage in year t is named Global_3G_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_3G_201712.shp represents 3G network coverage in 2018. The files that represent 3G network coverage for earlier years are as follows: Global_3G_2009Q1.shp—3G coverage in 2010, Global_3G_2008Q1.shp—3G coverage in 2009, Global_3G_2007Q1.shp—3G coverage in 2008, and Global_3G_2006Q3.shp—3G coverage in 2007. Note that in the resulting raster files areas without 3G coverage should be assigned the value of zero.
2. Open NASA's map of population density (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP) and increase the resolution of the map to match the resolution of the 3G network coverage rasters.
3. For each grid cell on the map of 3G network coverage, multiply the binary value of 3G coverage in that grid cell by population density in the grid cell. The resulting raster illustrates the number of people with potential access to 3G per square kilometer.
4. Open the **all_election_districts.shp** file located in the **auxiliary_data** subfolder and, for each locationid (the ID of the districts), calculate the average value of the file from step 3.
5. For each locationid, calculate the average value of population density.
6. Export the data from steps 4 and 5.
7. In the exported files, for each locationid, divide the value of population-weighted 3G network coverage by the value of population density. The resulting variable represents the share of the district's population with potential access to 3G.
8. Create the variable for the year and combine the data for different years.
9. Drop the population density variable, rename "locationid" to "nutsid", and save the file as **3g_populists.dta**.
10. Interpolate the data for 2011 as the average of the data in 2010 and 2012. Save in the **/original_data/original_proprietary_data** directory.

brazil_3g_2018.dta:

1. Convert the shapefile maps of 3G network coverage into binary raster files with the resolution of 1×1 kilometer. For 2012-2018, the shapefile that represents 3G network coverage in year t is named Global_3G_XXXX12.shp, where XXXX is the number of year $t-1$. For instance, Global_3G_201712.shp represents 3G network coverage in 2018. The files that represent 3G network coverage for earlier years are as follows: Global_3G_2009Q1.shp—3G coverage in 2010, Global_3G_2008Q1.shp—3G coverage in 2009, Global_3G_2007Q1.shp—3G coverage in 2008, and Global_3G_2006Q3.shp—3G coverage in 2007. Note that in the resulting raster files areas without 3G coverage should be assigned the value of zero.
2. Open NASA's map of population density (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP) and increase the resolution of the map to match the resolution of the 3G network coverage rasters.
3. For each grid cell on the map of 3G network coverage, multiply the binary value of 3G coverage in that grid cell by population density in the grid cell. The resulting raster illustrates the number of people with potential access to 3G per square kilometer.
4. Open the **brazil_microregions.shp** file located in the **auxiliary_data** subfolder and, for each microreg (the ID of the microregions), calculate the average value of the file from step 3.
5. For each microreg, calculate the average value of population density.
6. Export the data from steps 4 and 5.
7. In the exported files, for each locationid, divide the value of population-weighted 3G network coverage by the value of population density. The resulting variable represents the share of the district's population with potential access to 3G.
8. Drop the population density variable, and save the file as **brazil_3g_2018.dta** in the **/original_data/original_proprietary_data** directory.

- **Lightning strike frequency data:** the data on the locations of the lightning strikes. These data are available from the WWLLN on a subscription basis. One can find more information about the data and contact the WWLLN to purchase the data here: http://wwlln.net/. To fully replicate the data, a researcher needs to use the WWLLN data to produce the following file: lightnings_2005_2017_weighted.dta, a file containing the number of lightning strikes in the GWP subnational regions for each year in 2005-2017, weighted by population density. To produce the file, a researcher needs to perform the following steps:

  1. Create a folder with the WWLLN data. The folder should have the following structure. It should have four subfolders: admin, output, rawdata, tmp. The admin subfolder should have a subfolder 3G_gov containing the maps of the GWP subnational regions (**level1_map.shp** and **level2_map.shp**) and the raster with population density data (**population_density.tif**). (All these files are in the **auxiliary_data** subfolder.) The output subfolder should have the subfolder 3G_gov_weighted, where the results will be saved. The rawdata subfolder should have an Afiles subfolder with the raw data from the WWLLN (Afiles). tmp is a subfolder where the temporary files will be saved.
  2. Using Python, run the **lightning_frequency_gwp.ipynb** file located in the **auxiliary_data** subfolder. Before running the file, you need to set the path to the folder with the WWLLN data in line 23 of the code. Note that running the code requires commitment as it takes multiple days (or even weeks, depending on the computer specification) to run.
  3. The code from step 2 produces files for the number of individuals affected by lightning strikes in regionid_m on that day. The two variables of interest are regionid_m (the ID of the subnational regions in the map) and pop (the number of individuals affected by lightning strikes in regionid_m on that day); all other variables should be ignored. The researcher should then sum the data for each regionid_m for all the days in a year, calculating the number of individuals affected by lightning strikes in each year. Create the year variable and combine the data for all the years.
  4. Match the regionid_m variable (the ID of the subnational regions in the map) to the districtid variable (the ID of the subnational regions in the GWP data). To match the data, use the **map_to_gallup_level1.dta** and **map_to_gallup_level2.dta** files (the two files can be combined). Drop the regionid_m and _merge variables; save the file as **lightnings_2005_2017_weighted.dta** in the **/original_data/original_proprietary_data** directory.

- **International Telecommunication Union (ITU) data:** Panel country-level data on active mobile subscriptions per capita. To generate it, the following steps are needed:

  1. Purchase the ITU World Telecommunication/ICT Indicators (WTI) Database 2019 (December 2019 Edition). One can find more information about the data and contact ITU to purchase it here: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx (accessed November 3, 2020).
  2. Install this database as a software application ("ITU WTID") that allows exporting individual data series.
  3. After launching the application, use the "Time Series Query" feature and select the following indicator: i911mw ("Active mobile-broadband subscriptions per 100 inhabitants").
  3. Export the country-level data in 2007-2018 for all available countries as a CSV file. Save as **ambs_pc.csv** in the **/original_data/original_proprietary_data** directory.

II. Datasets in open access

The list below describes all the original nonproprietary datasets (located in **original_data** subfolder) and provides their sources. All datasets have all variables labelled. The online appendix provides more details.

- **brazil_2018_round2.dta**—the dataset with the results of the second round of the presidential election in Brazil in 2018, by microregion.
- **centroid_coord.dta**—the dataset with the latitude and longitude of the centroids of the GWP subnational regions.
- **collins_included.dta**—the dataset with a dummy variable for whether Collins Bartholomew provides 3G coverage data for the country.
- **corruption_incidents_IMF.dta**—the dataset with the Global Incidents of Corruption Index. The data are available from Furceri, Papageorgiou, and Ahir (2019).
- **country_gwp.dta**—the dataset with the code of the country in the GWP and the name of the country.
- **election_results_europe_stata_format.dta**—the dataset with the election results in Europe. The dataset also contains the classification of political parties.
- **elevation.dta**—the dataset with the elevation data for the GWP subnational regions.
- **fotn.xlsx**—the dataset with the Freedom on the Net data (i.e., internet censorship measures), by country-year.
- **fotp.dta**—the dataset with the Freedom of the Press data (i.e., measure of censorship of the traditional press), by country-year.
- **gwp_area_size.dta**—the dataset with the GWP subnational regions' area size. The dataset also contains variables for the share of territory covered by deserts and mountains.
- **gwp_leader_id.dta**—the dataset with the IDs of the country leaders. The IDs are used to match the GWP data with data on the education of the country leadership.
- **ITU_Key_2005-2019_ICT_data_with LDCs_28Oct2019_Final.xls**—the dataset with the information on fixed and mobile broadband subscriptions worldwide.
- **leaders_education.dta**—the dataset with the information on the education and occupation of the leadership of the world's countries (from Gerring et al., 2019).
- **luminosity_gwp.dta**—the dataset with the nighttime light density data for the GWP subnational regions.

- **luminosity_populists.dta**—the dataset with the nighttime light density data for the European subnational districts for which the election data are available.
- **medvedev_fbk.dta**—the dataset for the replication of the Russia case study results.
- **num_entities_panama.dta**—the dataset with the number of entities mentioned in the Panama papers, by country.
- **polity2.dta**—the dataset with the Polity2 score data, by country-year.
- **population_density_gwp.dta**—the dataset with the population density data for the subnational regions in the GWP.
- **smartphone-penetration-rate-by-country.dta**—the dataset with the data on smartphone penetration, by country-year.
- **world_bank_variables.dta**—the dataset with World bank data, by country-year.
- **world_bank_variables_income_groups.dta**—the dataset with World bank data on the country's income group.

III. STATA programs

The list below describes all the programs (STATA dofiles) that reproduce the results in the paper. Note that comments inside each dofile indicate which tables and figures they produce.

- **master_replicate_GMZ_3G_qje.do** is the main dofile that calls all other dofiles in order. This is the only dofile that needs to be executed to generate all the results in the paper. It is located in the root **replication** folder. All other dofiles are located in the **do_files** subfolder.

- **create_dataset_GWP_qje.do**—the dofile combines data from the original Gallup World Poll (GWP) dataset with data from other sources to produce the file that can be used to replicate the GWP-related results: **Gallup_merged_ready.dta**. (The dofile **create_dataset_GWP_qje.do** needs to be executed before running the following dofiles: **smartphones_qje.do** and **analysis_GWP_qje.do**, as indicated in **master_replicate_GMZ_3G_qje.do**.)

- **create_dataset_elections_qje.do**—the dofile combines data from the election results dataset with data from other sources to produce the file **3g_elections_europe.dta** that can be used to replicate the election-related results. (**create_dataset_elections_qje.do** needs to be executed before running the following dofiles: **analysis_elections_qje.do**, as indicated in **master_replicate_GMZ_3G_qje.do**.)

- **analysis_GWP_qje.do**—the dofile that reproduces the GWP-related results.

- **analysis_GWP_qje.do**—the dofile that reproduces the election-related results.

- **fig_a1_qje.do**—the dofile reproduces Figure A.1 of the paper.

- **smartphones_qje.do**—the dofile that reproduces the results related to smartphone penetration (Table A22; Figures A21 and A22).

- **case_studies_qje.do**—the dofile that reproduces the results related to the case studies (Table A25; Figures A23 and A24).

IV. Auxiliary data

The list below describes all the files in the **auxiliary data** subfolder.

- **3g_gwp_cy_without_3g.dta**—the dataset containing all the data from the **3g_gwp_cy.dta** file (described above) with the exception of the three_g_w variable.
- **all_election_districts.shp**—the map of the European districts for which the election results are available.
- **brazil_microregions.shp**—the map of Brazil's microregions for which the election results in 2018 are available.
- **districts_in_sample.dta**—the names and codes of the European districts for which the election results are available.
- **level1_map.shp**—the map of the GWP subnational regions that available at the first administrative division level.
- **level2_map.shp**—the map of the GWP subnational regions that available at the second administrative division level.
- **lightning_frequency_gwp.ipynb**—the Python workbook file that calculates the frequency of the lightning strikes in the GWP subnational regions.
- **map_to_gallup_level1.dta**—the file that matches the regionid_m variable (ID of the GWP subnational regions in the level1_map.shp map) to the districtid variable (ID of the GWP subnational regions in the GWP).
- **map_to_gallup_level2.dta**—the file that matches the regionid_m variable (ID of the GWP subnational regions in the level2_map.shp map) to the districtid variable (ID of the GWP subnational regions in the GWP).
- **population_density.tif**—the raster file with NASA's map of population density around the world (can be found here: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP accessed November 3, 2020).
- **proprietary_data_not_available.do**—the dofile that has been used to replace the values of the variables in the datasets using proprietary data.
- **world_map_only_countries.shp**—the shapefile with the borders of the countries around the world.