```

////////////////////////////////////////////////////////////////
//                   MCC FY21 Data Cleaning File              //
//                           S&E Team                         //
////////////////////////////////////////////////////////////////

/*
The following file provides a step by step demonstration of the
process used by MCC's S&E Team to clean all of the raw data files
from indicator institutions and create the scorecards.  These steps
are outlined in MCC's FY21 Guide to the Indicators
(https://www.mcc.gov/resources/doc/guide-to-the-indicators-fy-2021)
This file is included to provide complete clarity as to exactly how
these indicators are constructed.

To use this file, you need to download the appropriate files from the
indicator institutions as described below, and set the global macros
to the correct locations.  Note that many of MCC's indicator institutions
historically revise their data, so newly downloaded data may no longer match
the data recieved by MCC.  MCC does not historically revised scorecard data,
so the data presented in the Open Data is a record of the data produced by
that institution at the time of scorecard publication.

This file will output individual files for each indicator.  You can run
sections of the program individually, but the section on GNI must be run
for the rest to work.

The official final data files are contained on MCC's open data site.
The purpose of this file is to provide systematic instructions explaining
the methodology used to create these files. Due to methodological changes
and historical data revisions by indicator institutions or MCC, this program
is not designed to recreate previous year’s scorecards. Rather it is designed
to clearly show MCC’s methodology in FY21. Updates will be made annually as needed.
*/

****************************
**  Preliminary Settings  **
****************************
** The following settings set up the necessary environment for the do file to run
** This includes programs which are run repeatedly in the file.

set more off
clear
////////////////////
//  Global Macros //
////////////////////

//To use this file you need to set the "globals" to the correct file paths on your computer
//raw_data is where you are storing the raw data, (instructions for download are included in each section below)
//created_data is where you would like the program to store temporary files
//final_data is where you would like the program to output the final files
global raw_data "WRITE_RAW_FILE_PATH_HERE" //Your filepath will probably be something like C:\Users\YOURUSERNAME\Documents\RAWDATAFOLDERNAME\
global created_data "WRITE_CREATED_FILE_PATH_HERE" //Your filepath will probably be something like C:\Users\YOURUSERNAME\Documents\CREATEDDATAFOLDERNAME\
global final_data "WRITE_FINAL_FILE_PATH_HERE" //Your filepath will probably be something like C:\Users\YOURUSERNAME\Documents\FINALDATAFOLDERNAME\


////////////////
//  Programs  //
////////////////
//The following sub-programs are used throughout the main program
//to rename countries and calculate percentiles.


**WBNaming
***********
//The following program is used to rename countries to a common set of names
```

```
//So that they can easily be merged together.

capture program drop WBNaming
program WBNaming
args country_variable

replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Côte d'Ivoire"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "São Tomé and Principe"
replace `country_variable' = "Israeli Occupied Territories*" if `country_variable' == "Israeli Occupied Territories* " | `country_variable' == "Israeli-Occupied Territories*"
replace `country_variable' = "Kosovo" if `country_variable' == "Kosovo*"
replace `country_variable' = "Chechnya" if `country_variable' == "Chechnya*"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote dâ€™Ivoire"
replace `country_variable' = "Cyprus" if `country_variable' == "Cyprus "
replace `country_variable' = "Hong Kong*" if `country_variable' == "Hong Kong* "
replace `country_variable' = "Nagorno-Karabakh*" if `country_variable' == "Nagorno-Karabakh* "
replace `country_variable' = "Northern Cyprus" if `country_variable' == "Northern Cyprus*  "
replace `country_variable' = "Pakistani Kashmir" if `country_variable' == "Pakistani Kashmir*"
replace `country_variable' = "Palestinian Authority Administered Territories*" if `country_variable' == "Palestinian Authority Administered Territories* " | `country_variable' == "Palestinian Authority-Administered Territories*"
replace `country_variable' = "Puerto Rico" if `country_variable' == "Puerto Rico*"
replace `country_variable' = "South Ossetia" if `country_variable' == "South Ossetia*"
replace `country_variable' = "Transnistria" if `country_variable' == "Transnistria*"
replace `country_variable' = "Bahamas, The" if `country_variable' == "Bahamas"
replace `country_variable' = "Brunei Darussalam" if `country_variable' == "Brunei"
replace `country_variable' = "Cabo Verde" if `country_variable' == "Cape Verde"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo (Brazzaville)"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo (Kinshasa)"
replace `country_variable' = "Egypt, Arab Rep." if `country_variable' == "Egypt"
replace `country_variable' = "Iran, Islamic Rep." if `country_variable' == "Iran"
replace `country_variable' = "Kyrgyz Republic" if `country_variable' == "Kyrgyzstan"
replace `country_variable' = "Lao PDR" if `country_variable' == "Laos"
replace `country_variable' = "North Macedonia" if `country_variable' == "Macedonia"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Micronesia"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "North Korea"
replace `country_variable' = "Russian Federation" if `country_variable' == "Russia"
replace `country_variable' = "Slovak Republic" if `country_variable' == "Slovakia"
replace `country_variable' = "Korea, Rep." if `country_variable' == "South Korea"
replace `country_variable' = "Syrian Arab Republic" if `country_variable' == "Syria"
replace `country_variable' = "Gambia, The" if `country_variable' == "The Gambia"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela"
replace `country_variable' = "Yemen, Rep." if `country_variable' == "Yemen"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Côte d'Ivoire"
replace `country_variable' = "Bahamas, The" if `country_variable' == "Bahamas"
replace `country_variable' = "Bolivia" if `country_variable' == "Bolivia (Plurinational State of)"
replace `country_variable' = "Hong Kong SAR, China" if `country_variable' == "China, Hong Kong Special Administrative Region"
replace `country_variable' = "Macao SAR, China" if `country_variable' == "China, Macao Special Administrative Region"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo"
replace `country_variable' = "Czech Republic" if `country_variable' == "Czechia"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Democratic People's Republic of Korea"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic of the Congo"
replace `country_variable' = "Egypt, Arab Rep." if `country_variable' == "Egypt"
replace `country_variable' = "Gambia, The" if `country_variable' == "Gambia"
replace `country_variable' = "Iran, Islamic Rep." if `country_variable' == "Iran (Islamic Republic of)"
replace `country_variable' = "Kyrgyz Republic" if `country_variable' == "Kyrgyzstan"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao People's Democratic Republic"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Micronesia (Federated States of)"
replace `country_variable' = "Korea, Rep." if `country_variable' == "Republic of Korea"
replace `country_variable' = "Moldova" if `country_variable' == "Republic of Moldova"
replace `country_variable' = "St. Kitts and Nevis" if `country_variable' == "Saint Kitts and Nevis"
replace `country_variable' = "St. Lucia" if `country_variable' == "Saint Lucia"
replace `country_variable' = "St. Vincent and the Grenadines" if `country_variable' == "Saint Vincent and the Grenadines"
replace `country_variable' = "Slovak Republic" if `country_variable' == "Slovakia"
replace `country_variable' = "United Kingdom" if `country_variable' == "United Kingdom of Great Britain and Northern Ireland"
replace `country_variable' = "Tanzania" if `country_variable' == "United Republic of Tanzania"
replace `country_variable' = "United States" if `country_variable' == "United States of America"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela (Bolivarian Republic of)"
replace `country_variable' = "Vietnam" if `country_variable' == "Viet Nam"
replace `country_variable' = "Yemen, Rep." if `country_variable' == "Yemen"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "Palestine"
replace `country_variable' = "Eswatini" if `country_variable' == "Swaziland"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CÃƒÂ´te d'Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CÃ´te d'Ivoire"
replace `country_variable' = "Syrian Arab Republic" if `country_variable' == "Syrian Arab Republic (the)"
replace `country_variable' = "Bahamas, The" if `country_variable' == "Bahamas (the)"
replace `country_variable' = "United Kingdom" if `country_variable' == "United Kingdom of Great Britain and Northern Ireland (the)"
replace `country_variable' = "Central African Republic" if `country_variable' == "Central African Republic (the)"
replace `country_variable' = "Comoros" if `country_variable' == "Comoros (the)"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo (the)"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Democratic People's Republic of Korea (the)"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic of the Congo (the)"
replace `country_variable' = "Dominican Republic" if `country_variable' == "Dominican Republic (the)"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao People's Democratic Republic (the)"
replace `country_variable' = "Marshall Islands" if `country_variable' == "Marshall Islands (the)"
replace `country_variable' = "Netherlands" if `country_variable' == "Netherlands (the)"
replace `country_variable' = "Niger" if `country_variable' == "Niger (the)"
replace `country_variable' = "Philippines" if `country_variable' == "Philippines (the)"
replace `country_variable' = "Korea, Rep." if `country_variable' == "Republic of Korea (the)"
replace `country_variable' = "Moldova" if `country_variable' == "Republic of Moldova (the)"
replace `country_variable' = "Russian Federation" if `country_variable' == "Russian Federation (the)"
replace `country_variable' = "Sudan" if `country_variable' == "Sudan (the)"
replace `country_variable' = "United Arab Emirates" if `country_variable' == "United Arab Emirates (the)"
replace `country_variable' = "United States" if `country_variable' == "United States of America (the)"
replace `country_variable' = "Hong Kong SAR, China" if `country_variable' == "Hong Kong"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "São Tomé and Príncipe"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, Dem. Rep."
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, The Democratic Republic of the"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo, Rep. of"
replace `country_variable' = "Tanzania" if `country_variable' == "Tanzania, United Republic of"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, Democratic People's Republic of"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Micronesia, Federated States of"
replace `country_variable' = "Bolivia" if `country_variable' == "Bolivia, Plurinational State of"
replace `country_variable' = "Moldova" if `country_variable' == "Moldova, Republic of"
replace `country_variable' = "Bosnia and Herzegovina" if `country_variable' == "Bosnia Herzegovina"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao People's Dem. Rep."
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Democratic Republic"
replace `country_variable' = "Curaçao" if `country_variable' == "Curacao"
replace `country_variable' = "Timor-Leste" if `country_variable' == "East Timor"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Federated States of Micronesia"
replace `country_variable' = "Kazakhstan" if `country_variable' == "Kazakstan"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, Dem. People's Rep. of"
replace `country_variable' = "Libya" if `country_variable' == "Libyan Arab Jamahiriya"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "Palestinian Territory, Occupied"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "State of Palestine"
replace `country_variable' = "Reunion" if `country_variable' == "Réunion"
replace `country_variable' = "Saint-Barthelemy" if `country_variable' == "Saint-Barthélemy"
replace `country_variable' = "St. Martin (French part)" if `country_variable' == "Saint-Martin (French)"
replace `country_variable' = "St. Martin (French part)" if `country_variable' == "Saint-Martin (French part)"
replace `country_variable' = "Samoa" if `country_variable' == "Western Samoa"
replace `country_variable' = "North Macedonia" if `country_variable' == "The former Yugoslav Rep. of Macedonia"
replace `country_variable' = "Tanzania" if `country_variable' == "United Rep. of Tanzania"
replace `country_variable' = "Faroe Islands" if `country_variable' == "Faeroe Islands"
replace `country_variable' = "North Macedonia" if `country_variable' == "The former Yugoslav Republic of Macedonia"
replace `country_variable' = "Korea, Rep." if `country_variable' == "Korea, Republic of"
replace `country_variable' = "Macao SAR, China" if `country_variable' == "China Macao Special Administrative Region"
replace `country_variable' = "Hong Kong SAR, China" if `country_variable' == "China Hong Kong Special Administrative Region"
replace `country_variable' = "Virgin Islands (U.S.)" if `country_variable' == "United States Virgin Islands"
replace `country_variable' = "Reunion" if `country_variable' == "RÃ©union"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo, Republic of the"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Democratic Republic of"
replace `country_variable' = "Gambia, The" if `country_variable' == "Gambia (The)"
replace `country_variable' = "Central African Republic" if `country_variable' == "Central African Rep"
replace `country_variable' = "Cabo Verde" if `country_variable' == "Cabo Verde Republic of"
replace `country_variable' = "Bolivia" if `country_variable' == "Bolivia Plurinational States of "
replace `country_variable' = "Myanmar" if `country_variable' == "Burma"
replace `country_variable' = "West Bank and Gaza"  if `country_variable' ==  "Palestinian Authority Administered Territories*"
replace `country_variable' = "West Bank and Gaza"  if `country_variable' ==  "Palestinian Territories"
replace `country_variable' = "Caribbean Netherlands" if `country_variable' == "Bonaire Saint Eustatius and Saba"
replace `country_variable' = "South Sudan" if `country_variable' == "Republic of South Sudan"
replace `country_variable' = "Lebanon" if `country_variable' == "LebaNon"
replace `country_variable' = "Puerto Rico" if `country_variable' == "Puerto Rico (U.S.)"
replace `country_variable' = "North Macedonia" if `country_variable' == "Macedonia, FYR"
replace `country_variable' = "Taiwan" if `country_variable' == "Taiwan, China"
replace `country_variable' = "San Marino" if `country_variable' == "San MariNo"
replace `country_variable' = "Timor-Leste" if `country_variable' == "Timor Leste"
replace `country_variable' = "West Bank and Gaza"  if `country_variable' ==  "Occupied Palestinian Territory"
replace `country_variable' = "Czech Republic" if `country_variable' == "Czechia"
replace `country_variable' = "Guinea-Bissau" if `country_variable' == "Guinea Bissau"
replace `country_variable' = "Guinea-Bissau" if `country_variable' == "GuineaBissau"
replace `country_variable' = "North Macedonia" if `country_variable' == "The former Yugoslav Republic of Macedonia"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "State of Palestine"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "West Bank and Gaza Strip"
replace `country_variable' = "Timor-Leste" if `country_variable' == "TimorLeste"
replace `country_variable' = "United States" if `country_variable' == "United States of America"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela (Bolivarian Republic of)**"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela (Bolivarian Republic of)"
replace `country_variable' = "Northern Cyprus" if `country_variable' == "Northern Cyprus*"
replace `country_variable' = "Northern Cyprus" if `country_variable' == "Cyprus (Turkish)"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Ivory Coast"
replace `country_variable' = "Iran, Islamic Rep." if `country_variable' == "Islamic Republic of Iran"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "The Democratic Republic Of The Congo"
replace `country_variable' = "Northern Cyprus" if `country_variable' == "Cyprus North"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, North"
replace `country_variable' = "Korea, Rep." if `country_variable' == "Korea, South"
replace `country_variable' = "Macao SAR, China" if `country_variable' == "Macau"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Republic of the Congo (Brazzaville)"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Dem. Rep. of the Congo"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "São Tomé e Príncipe"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Dem. Rep. of Congo"
replace `country_variable' = "Antigua and Barbuda" if `country_variable' == "Antigua"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote d’Ivoire"
replace `country_variable' = "Tibet" if `country_variable' == "Tibet*"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, Dem. People’s Rep."
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Democratic Republic of the Congo"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo, Republic of"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, North "
replace `country_variable' = "Taiwan" if `country_variable' == "Taiwan "
replace `country_variable' = "Eswatini" if `country_variable' == "eSwatini"
replace `country_variable' = "Bosnia and Herzegovina" if `country_variable' == "Bosnia"
replace `country_variable' = "Bosnia and Herzegovina" if `country_variable' == "Bosnia & Herz."
replace `country_variable' = "Bosnia and Herzegovina" if `country_variable' == "Bosnia And Herzegovina"
replace `country_variable' = "Bosnia and Herzegovina" if `country_variable' == "Bosnia-Herzegovina"
replace `country_variable' = "Myanmar" if `country_variable' == "Burma (Myanmar)"
replace `country_variable' = "Myanmar" if `country_variable' == "Burma/Myanmar"
replace `country_variable' = "Central African Republic" if `country_variable' == "Central Afr. Rep."
replace `country_variable' = "China" if `country_variable' == "China Mainland"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Dem. Rep. (Zaire)"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Democratic Republic Of (Kinshasa)"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo, Rep"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo, Republic Of (Brazzaville)"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote D'Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote D’Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote d'Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote d’Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CôTe D'Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote d’Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Cote d’Ivoire"
replace `country_variable' = "Cyprus" if `country_variable' == "Cyprus (Greek)"
replace `country_variable' = "Czech Republic" if `country_variable' == "Czech Rep."
replace `country_variable' = "Czech Republic" if `country_variable' == "Czech Republic"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Dem. Rep. Congo"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic Of Congo"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic Of The Congo"
replace `country_variable' = "Vietnam" if `country_variable' == "Democratic Republic Of Vietnam"
replace `country_variable' = "Dominican Republic" if `country_variable' == "Dom. Rep."
replace `country_variable' = "Dominican Republic" if `country_variable' == "Dominican Rep."
replace `country_variable' = "Dominican Republic" if `country_variable' == "Dominican Republic"
replace `country_variable' = "Guinea-Bissau" if `country_variable' == "Guinea Bissao"
replace `country_variable' = "Hong Kong SAR, China" if `country_variable' == "Hong Kong Sar, China"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "Israeli-Occupied Territories And Palestinian Authority"
replace `country_variable' = "Korea, Rep." if `country_variable' == "Korea"
replace `country_variable' = "Kyrgyz Republic" if `country_variable' == "Kyrgyz Rep."
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao PDR"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao Pdr"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao, Pdr"
replace `country_variable' = "Macao SAR, China" if `country_variable' == "Macao"
replace `country_variable' = "North Macedonia" if `country_variable' == "Macedonia, Fyr"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Micronesia, Fs"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "Palestine/Gaza"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "Palestine/West Bank"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Republic Of The Congo"
replace `country_variable' = "Romania" if `country_variable' == "Roumania"
replace `country_variable' = "St. Vincent and the Grenadines" if `country_variable' == "Saint Vincent And The Grenadine"
replace `country_variable' = "St. Vincent and the Grenadines" if `country_variable' == "Saint Vincent And The Grenadines"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "Sao Tome And Pr."
replace `country_variable' = "Serbia & Montenegro" if `country_variable' == "Serbia And Montenegro"
replace `country_variable' = "St. Kitts and Nevis" if `country_variable' == "St. Kitts And Nev."
replace `country_variable' = "St. Vincent and the Grenadines" if `country_variable' == "St. Vincent And The Grenadines"
replace `country_variable' = "South Sudan" if `country_variable' == "Sudan South"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "SãO Tomé And Principe"
replace `country_variable' = "Trinidad And Tobago" if `country_variable' == "Trinidad And Tob."
replace `country_variable' = "United Arab Emirates" if `country_variable' == "UAE"
replace `country_variable' = "United Arab Emirates" if `country_variable' == "Uae"
replace `country_variable' = "Uraguay" if `country_variable' == "Uruguay *"
replace `country_variable' = "United States" if `country_variable' == "USA"
replace `country_variable' = "United States" if `country_variable' == "Usa"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela *"
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela, Rb"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "West Bank Gaza"
replace `country_variable' = "West Bank and Gaza" if `country_variable' == "West Bank And Gaza"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Republic of Congo"
replace `country_variable' = "North Macedonia" if `country_variable' == "FYR Macedonia"
replace `country_variable' = "Hong Kong SAR, China" if `country_variable' == "Hong Kong SAR"
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao P.D.R."
replace `country_variable' = "Taiwan" if `country_variable' == "Taiwan Province of China"
replace `country_variable' = "Bahamas, The" if `country_variable' == "The Bahamas"
replace `country_variable' = "Macao SAR, China" if `country_variable' == "Macao SAR"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "Côte d’Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CÔTe D'Ivoire"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo-Brazaville"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CÃƒÂ´te d'Ivoire"
replace `country_variable' = "Cote d'Ivoire" if `country_variable' == "CÃ´te d'Ivoire"
replace `country_variable' = "Curaçao" if `country_variable' == "CuraÃ§ao"
replace `country_variable' = "Afghanistan" if `country_variable' == "Afganistan"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo Dem. Rep"
replace `country_variable' = "Somalia" if `country_variable' == "Somalia "
replace `country_variable' = "Zimbabwe" if `country_variable' == "Zimbabwe "
replace `country_variable' = "Egypt, Arab Rep." if `country_variable' == "Egypt,Arab Rep."
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Dem. Rep. "
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo, Dem. Rep"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic of Congo"
replace `country_variable' = "Ghana" if `country_variable' == "Ghana "
replace `country_variable' = "Burundi" if `country_variable' == "Burndi"
replace `country_variable' = "Philippines" if `country_variable' == "Phillippines"
replace `country_variable' = "Bolivia" if `country_variable' == "Bolivua"
replace `country_variable' = "Congo, Rep." if `country_variable' == "Republic of the Congo "
replace `country_variable' = "Egypt, Arab Rep." if `country_variable' == "Egypt, Arab Rep"
replace `country_variable' = "Zambia" if `country_variable' == "Zambia "
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Micronesia,Fed. Sts."
replace `country_variable' = "Yemen, Rep." if `country_variable' == "Yemen Rep."
replace `country_variable' = "Lao PDR" if `country_variable' == "Lao PDR "
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea, Dem. Rep"
replace `country_variable' = "Ethiopia" if `country_variable' == "Ethopia"
replace `country_variable' = "Korea, Dem. People's Rep." if `country_variable' == "Korea,Dem.Rep"
replace `country_variable' = "Micronesia, Fed. Sts." if `country_variable' == "Mirconesia, Fed. Sts"
replace `country_variable' = "Senegal" if `country_variable' == "Sengal"
replace `country_variable' = "Serbia & Montenegro" if `country_variable' == "Serbia and Montenegro"
replace `country_variable' = "Tajikistan" if `country_variable' == "Taijikistan"
replace `country_variable' = "Yemen, Rep." if `country_variable' == "Yemen, Rep"
replace `country_variable' = "Cabo Verde" if `country_variable' == "Cape Verde "
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Congo,Dem.Rep."
replace `country_variable' = "Congo, Rep." if `country_variable' == "Congo,Rep."
replace `country_variable' = "India" if `country_variable' == "India "
replace `country_variable' = "Venezuela, RB" if `country_variable' == "Venezuela, Bolivarian Republic of"
replace `country_variable' = "Ethiopia" if `country_variable' == "Ethiopia "
replace `country_variable' = "Egypt, Arab Rep." if `country_variable' == "Egypt "
replace `country_variable' = "Yemen, Rep." if `country_variable' == "Yemen "
replace `country_variable' = "Iraq" if `country_variable' == "Iraq "
replace `country_variable' = "Benin" if `country_variable' == "Benin "
replace `country_variable' = "Syrian Arab Republic" if `country_variable' == "Syria "
replace `country_variable' = "Sri Lanka" if `country_variable' == "SriLanka"
replace `country_variable' = "Malawi" if `country_variable' == "Malawi "
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "DemocraticRepublicCongo"
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "DRC"
replace `country_variable' = "Pakistan" if `country_variable' == "Pakistan "
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "DR Congo"
replace `country_variable' = "Ukraine" if `country_variable' == "Ukraine "
replace `country_variable' = "Congo, Dem. Rep." if `country_variable' == "Democratic Republic of the Congo (DRC)"
replace `country_variable' = "Mali" if `country_variable' == "Mali "
replace `country_variable' = "Congo, Rep." if `country_variable' == "Republic of the Congo"
replace `country_variable' = "Philippines" if `country_variable' == "Phillipines"
replace `country_variable' = "North Macedonia" if `country_variable' == "The Republic of North Macedonia"
replace `country_variable' = "Sao Tome and Principe" if `country_variable' == "Sao Tomé and Principe"
end


**PercentRankInc
****************
//The following program mimics Excel's PercentRankInc function
//For indicators with a static cutoff (Civil Liberties, Political Rights,
//Immunization Rates for the higher scorecard pool and Inflation) this is done normally.
//However, for indicators with a median based cutoff, countries at the median are set to
//the 50th percentile and countries at the minimum are set to the 0th percentile
```

```

capture program drop PercentRankInc
program PercentRankInc
args vartest

generate over = 300  //Dummy value for now, this will be the number of countries scoring greater than or equal to a given country
generate under = 300  //Dummy value for now, this will be the number of countries scoring less than a given country
global count = _N //Number of observations
global vartest = "`vartest'"

forvalues line_change = 1(1)$count {  //This loops once for each country.
    global overcount = -1  //This counts the number of countries over the country in line_change.  It starts at -1 as we do not count the country itself as being greater than or equal to itself.
    forvalues line_test = 1(1)$count {  //This loops over all the countries again, comparing them each to the selected country from line_change
        if `=${vartest}[`line_change']' <= `=${vartest}[`line_test']' & `=${vartest}[`line_test']' != . {  //This if statement augments the overcount whenever a country from line_test is greater than or eqaul to the selected country from line_change
            global overcount = ${overcount} + 1
        }
    replace over = ${overcount} in `line_change' //This replaces the variable over with the overcount macro
    replace over = . in `line_change' if missing(`=${vartest}[`line_change']')  //This sets over to missing if the original value is missing.
    }
}

forvalues line_change = 1(1)$count { //This loops once for each country.
    global undercount = 0 //This counts the number of countries over the country in line_change. It starts at 0 since we are not testing less than or equal to, so we will not count the country against itself.
    forvalues line_test = 1(1)$count { //This loops over all the countries again, comparing them each to the selected country from line_change
        if `=${vartest}[`line_change']' > `=${vartest}[`line_test']' & `=${vartest}[`line_test']' != . {  //This if statement augments the undercount whenever a country from line_test is less than the selected country from line_change
            global undercount = ${undercount} + 1
        }
    replace under = ${undercount} in `line_change'  //This replaces the variable under with the undercount macro
    replace under = . in `line_change' if missing(`=${vartest}[`line_change']') //This sets over to missing if the original value is missing.
    }
}
gen percentile = under/(over+under)  //This calculates percentiles for all observations using the under and over variables.

//The if statement below forces countries at the median to the 50th percentile, at the max to the 100th percentile and at the min to the 0th percentile
//For all the median based indicators (see above for more information).
if "`vartest'" != "Civil_Liberties" & "`vartest'" != "Political_Rights" & "`vartest'" != "Inflation" & "`vartest'" != "Immunization_Rate_HISP" {
egen median = median(`vartest')
egen min = min(`vartest')
egen max = max(`vartest')
replace percentile = 0.5 if median == `vartest'
//replace percentile = 1 if max == `vartest' //In FY21 MCC stopped forcing countreis tied for the max to the 100th percentile
replace percentile = 0 if min == `vartest'
}
gen Score_In_FY21 = `vartest'
end

*****************************
**  Import/Clean Raw Data  **
*****************************
** The following sections take the necessary data files for MCC's Scorecard,
** import them, clean them and save the necessary files.  This section
** includes instructions for downloading these files as well as processing them.
** in comments before each section.

///////////
//  GNI  //
///////////
//Countries are divided into two scorecard pools based on GNI per capita.
//The following is the construction of that measure from World Bank data.

//Download Instructions:
/*
Download the xls file from:
https://datacatalog.worldbank.org/dataset/gni-capita-ranking-atlas-method-and-ppp-based
Check if columns and sheet names still match
Check if the footnotes match the income categories (When there is no GNI data
available for a country, in some cases the World Bank still provides a determination
of which income group that country would be in, there are captured in footnotes in the
GNIPC file.)

Save in the ${raw_data} folder as GNIPC.xls

Note that MCC uses the July release.  Sometimes the World Bank will change data
 in between the July and the September release.  The version from the July release
 is available on MCC's Open Data Portal
*/

import excel "${raw_data}\GNIPC July 1 2020 FINAL.xls", sheet("GNIPC") clear
//According the the World Bank, these countries fit MCC's definition of LISP and HISP
//Therefore the footnotes are translated into income categories
//D is country name, E is GNI, F is a footnote for countries missing GNI
//Sometimes the world bank will change which footnotes correspond to which income
categories, check this carefully.
//LISP stands for Lower Income Scorecard Pool
//HISP stands for Higher Income Scorecard Pool
//These are divided by the Historic IDA threshold.

replace E = "LISP" if F =="k"
//replace E = "HISP" if F =="l"
replace E = "UMIC" if F =="i"
replace E = "HIC" if F =="j"
keep D E
// The following are not listed as "Independent States" by the State Department,
// and therefore do not receive scorecards.  In the following sections, we limit the
// dataset by only keeping countries that match a country in this dataset.

drop if D == "" | D == "Economy" | D =="World" | D =="East Asia & Pacific" /*
*/| D =="Europe & Central Asia" | D =="Latin America & Caribbean" /*
*/ | D =="Middle East & North Africa" | D =="North America" | D =="South Asia" /*
*/| D =="Sub-Saharan Africa" | D =="Low income" | D =="Lower middle income" /*
*/| D =="Upper middle income" | D =="High income" | D=="West Bank and Gaza"

//These are placeholder values over the threshold for countries without data.
replace E = "15005" if E == "HIC"
replace E = "6005" if E == "UMIC"
replace E = "105" if E == "LISP"

rename D Country
rename E GNI

//Now that all of the values for GNI are numeric, we can create a string variable "Income"
Sorting them by threshold
//These threshold categories can be found in the following file labeled "Threshold"
//They can be different every year, and so will need to be changed for future iterations
destring, replace
generate Income = "LISP"
replace Income = "HISP" if GNI > 1945
replace Income = "UMIC" if GNI > 4045
replace Income = "HIC" if GNI > 12535

```

```
//This program standardizes all of the country names, so that they can be matched with other data files
WBNaming Country

save "${created_data}\GNI_FY21.dta", replace
export excel using "${final_data}\FY21_GNI.xlsx", firstrow(variables) replace

//////////////////
//  Thresholds  //
/////////////////
//Countries are divided into two scorecard income pools based on GNI per capita.
//The cutoff for these pools is the Historic IDA threshold from the World Bank
//The following section demonstrates how these are constructed from the World Bank Data.
//This is where the cutoffs for the Income variable in the previous section come from.

//Download Instructions:
/*
Download the historical classification by income in XLS format from
https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups
Check if columns and sheet names still match
Save in the ${raw_data} folder as OGHIST.xls
*/
import excel "${raw_data}\OGHIST.xls", sheet("Thresholds") clear
keep A S-AV
keep if A == "Date:" |A == "Bank's fiscal year:" | A == "II. IDA Eligibility /2  /7" | A == "Lower middle income" | A == "Upper middle income"
sort A
replace A = "Threshold" if A == "Bank's fiscal year:"
//You may need to install renvars. Enter "help renvars" and install the update.
//This command renames the variables with the first row of observations.
renvars, map (word(@[1],1))
drop in 1
foreach var of varlist _all {
replace `var' = strreverse(substr(strreverse(`var'),1,4)) in 1  //Takes the year (last 4 digits) only from each value in the first row
replace `var' = substr(`var',(strpos(`var'[2],"=")+2),.) in 2  //Isolates the cutoff value for the second row (the cutoff between the Lower Income Scorecard Pool and the Higher Income Scorecard Pool)
replace `var' = substr(`var',(strpos(`var'[3],"-")+1),.) in 3  //Isolates the cutoff value for the third row (the cutoff between LMICs and UMICs)
replace `var' = substr(`var',(strpos(`var'[4],"-")+1),.) in 4  //Isolates the cutoff value for the fourth row (the cutoff between UMICs and HICs)
replace `var' = subinstr(`var',",","",.)  //Removes commas from all numbers so they can be converted to numeric values.
}
destring, replace

xpose, clear varname  //Reshapes data
label variable v1 "Year"
label variable v2 "LISP"
label variable v3 "HISP"
label variable v4 "UMIC"
label variable _varname "Fiscal Year"
drop in 1
order _varname
save "${created_data}\Threshold_FY21.dta", replace
export excel using "${final_data}\FY21_Threshold.xlsx", firstrow(varlabels) replace

//////////////////
//  Population  //
//////////////////
//Population is not used to determine scorecard passage.
//However, it is included on the scorecard as a descriptive statistic in the upper left corner,
// so the method for constructing it is listed below

//Download Instructions:
/*
Go to https://data.worldbank.org/indicator/SP.POP.TOTL
```

```
Download the excel file
Save it as World Bank Population Data.xls in the raw data folder
*/
import excel "${raw_data}\World Bank Population Data.xls", sheet("Data") clear
//Dropping unnecessary observations with datasource information
drop in 3
drop in 2
drop in 1
//Formatting the first row of variables to use renvars
foreach var of varlist _all{
label var `var' "`=`var'[1]'"
replace `var' = "Y" +`var'[1] in 1
}
foreach var of varlist _all{
replace `var' = subinstr("`=`var'[1]'"," ", "_",.) in 1
}
renvars, map (word(@[1],1))
drop in 1
destring, replace
rename YCountry_Name Country
drop YCountry_Code YIndicator_Name YIndicator_Code
WBNaming Country
//This takes out all of the things that are not considered "Independent States" by the State Department
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keepusing() keep(3) //This removes those observations which were not included in the GNI data above.
drop GNI
keep Country Y2019
export excel using "${final_data}\FY21_Population.xlsx", firstrow(varlabels) replace


/////////////////////////////////////////
//  Democratic Rights Indicators (FH)  //
/////////////////////////////////////////
//The Democratic Rights Indicators are Political Rights and Civil Liberties
//They are both sourced from Freedom House's Freedom In The World report.
//Download Instructions
/*
Download the Aggregate Category and Subcategory Scores, 2003-2019 (Excel) file from
https://freedomhouse.org/report-types/freedom-world
Check if columns and sheet names still match the lines below (i.e. is the sheet still
named "FIW06-19" and the variable for country/territory still named CT)
Save in the ${raw_data} folder as 2020_Aggregate Category and Subcategory Scores FIW2003-2020.xlsx
*/
set more off
import excel "${raw_data}\2020_Aggregate_Category_and_Subcategory_Scores_FIW_2003-2020.xlsx", sheet("FIW06-20") firstrow clear
keep if CT == "c" //Only keeping countries
keep if Edition == 2020 //Only keeping current year
keep CountryTerritory PR CL //Only keeping country name and relevant data values (PR=Political Rights, CL=Civil Liberties)
rename CountryTerritory Country

**To Merge with GNI data
************************
WBNaming Country //Renaming countries to match GNI data
save "${created_data}\FH_FY21.dta", replace

merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)  //This only keeps countries included in the GNI data

keep Country Income CL PR
rename CL Civil_Liberties
rename PR Political_Rights
save "${created_data}\FH Full.dta", replace

//Percentile are not used for determining passage on this indicator, but they are included on the scorecard
//The following computes the percentile ranks for the Democratic Rights Indicators
```

```
//It is worthy of note that MCC's general rule of forcing countries at the median to the 50th percentile
// does not apply here since percentiles are not used in the calculation of passage

global FH_vars "Civil_Liberties Political_Rights"
foreach var of global FH_vars {
//Percentile Ranks for LISPs
preserve
keep if Income == "LISP"
PercentRankInc `var'  //This program is listed above, and calculates the percentile rank using MCC's Methodology
save "${created_data}\FH `var' LISP.dta", replace
restore

//Percentile Ranks for HISPs
preserve
keep if Income == "HISP"
PercentRankInc `var'

//Appending other scorecard pool data
append using "${created_data}\FH `var' LISP.dta"
keep Country percentile Score_In_FY21 Income
gen Indicator = "`var'"
rename percentile Percentile
if "`var'" == "Civil_Liberties" {  //We need to use an if statement here because we can't put a local macro at the beginning of the filename
save "${created_data}\Civil Liberties.dta", replace
}
else {
save "${created_data}\Political Rights.dta", replace
}

export excel "${final_data}\FY21_`var'.xlsx", firstrow(variables) replace
restore
}


////////////////////////////
//  Education Indicators  //
////////////////////////////

//Download Instructions:
/*
Go to http://data.uis.unesco.org/
On the left panel select National Monitoring > National Monitoring > National Monitoring Full Dataset
Click Customise > Selection > Indicator
Click Unselect Items > Unselect all then Collapse all
Click National > Indicator > Government expenditure on education as a percentage of GDP
 > Government expenditure on primary education as a percentage of GDP (%) (XGDP_1_FSGOV)


Then select Participation > Enrollment Ratios >
National > Indicator> Number of students and enrolment rates by level of education
> Enrolment ratios > Gross enrolment ratio by level of education > Gross enrolment ratio, lower secondary, female (%) (GER_2_F)

Click Time > Select latest data > Select last 50 years

Click View Data
Go to Export > Text File (CSV) > Download

Check if columns and sheet names still match
Save in the ${raw_data} folder as Ed Nat Data YEAR.csv

Completion is now with the SDG indicator:
Click Customise > Selection > Indicator
Click Unselect Items > Unselect all then Collapse all
Then select SDG > Sustainable Development Goal 4 >
Target 4.1: By 2030 ensure that all girls and boys complete free, equitable, and quality
```

```
primary and secondary education leading to relevant and effective learning outcomes
> 4.1.3 Gross intake ratio to the last grade (primary education, lower secondary education) >
Gross intake ratio to the last grade of primary education, female (%)

Click Time > Select latest data > Select last 50 years

Click View Data
Go to Export > Text File (CSV) > Download

Check if columns and sheet names still match
Save in the ${raw_data} folder as Ed SDG Data YEAR.csv
*/
set more off
import delimited "${raw_data}\Ed Nat Data 2020.csv", clear
rename ïnatmon_ind indcode
save "${created_data}\Ed Nat Data 2020.dta",replace
import delimited "${raw_data}\Ed SDG Data 2020.csv", clear
rename ïsdg_ind indcode
append using "${created_data}\Ed Nat Data 2020.dta"

drop v6
//These countries would be removed when the GNI data is merged anyway, but we need to drop them early due to the way the reshaping process handles duplicates
drop if country == "World" | country == "Latin America and the Caribbean" | country == "Small Island Developing States" /*
*/ | country == "Africa (Northern)" | country == "Africa (Sub-Saharan) " | country == "Arab States" /*
*/ | country == "Arab States" | country == "Asia (Central and Southern)" | country == "Asia (Central)" /*
*/ | country == "Asia (Eastern and South-eastern)" | country == "Asia (Eastern)" | country == "Asia (South-eastern)" /*
*/ | country == "Asia (Southern)" | country == "Asia (Western)" | country == "Africa (Sub-Saharan)" | country == "East Asia and the Pacific" /*
*/ | country == "Central Asia" | country == "Central and Eastern Europe" | country == "High income countries" | country == "Landlocked Developing Countries" /*
*/ | country == "Least Developed Countries" | country == "Low income countries" | country == "Lower middle income countries" | country == "Middle income countries" /*
*/ | country == "North America and Western Europe" | country == "Northern America" | country == "Northern America and Europe" | country == "Oceania" /*
*/ | country == "Oceania (Australia/New Zealand)" | country == "Oceania (excl. Australia/New Zealand)" | country == "South and West Asia" | country == "Sub-Saharan Africa" /*
*/| country == "Upper middle income countries" | country == "Western Asia and Northern Africa" | country == "Europe" | country == "Northern America" | country == "Northern America "



//Renaming countries for consistency
WBNaming country
rename time year

////Reshaping the data into a "wide" format so it is easier to use
**EdExpense
preserve
keep if indicator == "Government expenditure on primary education as a percentage of GDP (%)"
keep country year value
reshape wide value, i(country) j(year)
save "${created_data}\EdExpense FY21.dta" , replace
restore

**Primary
preserve
keep if indicator == "Gross intake ratio to the last grade of primary education, female (%)"
keep country year value
reshape wide value, i(country) j(year)
save "${created_data}\Primary FY21.dta" , replace
restore

**Secondary
```

```
keep if indicator == "Gross enrolment ratio, lower secondary, female (%)"
keep country year value
reshape wide value, i(country) j(year)
save "${created_data}\Secondary FY21.dta" , replace


**Merging
*********
//For these indicators, MCC takes the most recent data point from the last 7 years of data (including the current year, 2019)
//Other documentation states that we take the last 6 years of data, this is because only one country has data for 2019
//Therefore it is the 6 most recent years of data starting in 2018 or the 7 most recent starting in 2019
//The following code takes the most recent value for each indicator in the last 7 years (since 2013)
set more off
//Ed Expenditures
use "${created_data}\EdExpense FY21.dta", clear
keep country value2014-value2020
egen EdExpense = rowlast(value*)  //Using the last (most recent) value, note that the variables must be sorted chronologically for this to work
keep country EdExpense
//Primary Completion
merge 1:1 country using "${created_data}\Primary FY21.dta" , nogen
keep country EdExpense value2014-value2020 //Limiting to the last 7 years
egen Primary = rowlast(value*) //Taking the most recent data
keep country Primary EdExpense
//Secondary Enrollment
merge 1:1 country using "${created_data}\Secondary FY21.dta", nogen
replace country = "Cote d'Ivoire" if country == "CÃ´te d'Ivoire"
replace country = "Cote d'Ivoire" if country == "Côte d'Ivoire"
keep country Primary EdExpense value2014-value2020
egen Secondary = rowlast(value*)
keep country Secondary EdExpense Primary
rename country Country
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta" , nogen keep(3)

global EdIndicators "Secondary Primary EdExpense"
save "${created_data}\Ed Full.dta", replace

//The following sections calculate percentile ranks for each of the three education indicators and generate the files:
preserve
keep if Income == "LISP"
PercentRankInc EdExpense
save "${created_data}\Ed EdExpense LISP.dta", replace
restore
preserve
keep if Income == "HISP"
PercentRankInc EdExpense
append using "${created_data}\Ed EdExpense LISP.dta"
keep Country Income percentile Score_In_FY21
gen Indicator = "Primary_Education_Expenditures"
rename percentile Percentile
save "${created_data}\Primary Education Expenditures.dta", replace
export excel using "${final_data}\FY21_Primary_Education_Expenditures.xlsx", firstrow(variables) replace
restore

preserve
keep if Income == "LISP"
PercentRankInc Primary
keep Country Income percentile Score_In_FY21
gen Indicator = "Primary_Girls_Education_Completion"
rename percentile Percentile
save "${created_data}\Primary Girls Education Completion.dta", replace
export excel using "${final_data}\FY21_Primary_Girls_Education_Completion.xlsx", firstrow(variables) replace
```

```
restore

preserve
keep if Income == "HISP"
PercentRankInc Secondary
keep Country Income percentile Score_In_FY21
gen Indicator = "Secondary_Girls_Education_Enrollment"
rename percentile Percentile
save "${created_data}\Secondary Girls Education Enrollment.dta", replace
export excel using "${final_data}\FY21_Secondary_Girls_Education_Enrollment.xlsx", firstrow(variables) replace
restore


///////////////////////////////////
// Immunization Rate Indicator  //
///////////////////////////////////
//Download Instructions
/*
Download from https://www.who.int/immunization/monitoring_surveillance/data/en/
Under 4. Immunization coverage or administered doses, select:
4.6 WHO/UNICEF Estimates of National Immunization Coverage (excel)

Check if columns and sheet names still match (i.e. is the sheet still called MCV1 and the Country variable still called Cname)
Save in the ${raw_data} folder as WHO coverage_estimates_series.xls
*/
//This indicator is an average of two measures of immunization: MCV1 and DTP3
//The following code performs this calculation and outputs the final file.
set more off
import excel "${raw_data}\WHO coverage_estimates_series.xls", sheet("MCV1") clear

//MCV1
foreach var of varlist E-AR {  //This loop allows the first row to be renamed as Y{year} (e.g. Y2016) since Stata cannot take numbers as variable names
replace `var' = "Y" +`var'[1] in 1
}
renvars, map (word(@[1],1))
drop in 1
destring, replace
rename Cname Country
keep Country Y*
WBNaming Country
order _all, sequential
keep Country Y2019
save "${created_data}\Immunization MCV FY21.dta" , replace

//DPT3
import excel "${raw_data}\WHO coverage_estimates_series.xls", sheet("DTP3") clear
foreach var of varlist E-AR { //This loop allows the first row to be renamed as Y{year} (e.g. Y2016) since Stata cannot take numbers as variable names
replace `var' = "Y" +`var'[1] in 1
}
renvars, map (word(@[1],1))
drop in 1
destring, replace
rename Cname Country
keep Country Y*
WBNaming Country
order _all, sequential
keep Country Y2019
save "${created_data}\Immunization DTP FY21.dta" , replace

**Merging
*********
//Merging both files and then merging them with the GNI data
set more off
use "${created_data}\Immunization MCV FY21.dta", clear
keep Country Y2019
rename Y2019 MCV1
```

```
merge 1:1 Country using "${created_data}\Immunization DTP FY21.dta"
keep Country Y2019 MCV1
rename Y2019 DTP3
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

egen Score = rowmean(MCV1 DTP3)  //Taking the average of the two values

save "${created_data}\Imm Full.dta", replace

preserve
keep if Income == "LISP"
// Staring in FY21, when the scorecard pool median for immunization rates
// data is below 90% the data is treated as a median based indicator (judged only
// on if the value is above or below the median, ties for the median are
// forced to 50th percentile) by naming it Immunization_Rate_HISP.  If the median
// for the scorecard income pool is above 90% then the data is treated as an
// absolute threshold indicator (judged agains the threshold of 90%, ties for
// the median are NOT forced to the 50th percentile) by naming it
// Immunization_Rate_LISP.  The code below implements this change.  This naming
// structure corresponds to the rules in the PercentRankInc Program above.
qui sum Score, d
local Imm_Med = `r(p50)'
if `Imm_Med'>90 {
    //PercentRankInc Treats the variables named HISP as absolute and variables
    //named LISP as median based due to the old rules for this indicator.
    rename Score Immunization_Rate_HISP
    PercentRankInc Immunization_Rate_HISP
    rename Immunization_Rate_HISP Score

}
else {
    rename Score Immunization_Rate_LISP
    PercentRankInc Immunization_Rate_LISP
    rename Immunization_Rate_LISP Score
}
save "${created_data}\Imm LISP.dta", replace

restore
keep if Income == "HISP"
if `Imm_Med'>90 {
    //PercentRankInc Treats the variables named HISP as absolute and variables
    //named LISP as median based due to the old rules for this indicator.
    rename Score Immunization_Rate_HISP
    PercentRankInc Immunization_Rate_HISP
    rename Immunization_Rate_HISP Score

}
else {
    rename Score Immunization_Rate_LISP
    PercentRankInc Immunization_Rate_LISP
    rename Immunization_Rate_LISP Score
}

append using "${created_data}\Imm LISP.dta"
keep Country Income percentile Score_In_FY21
gen Indicator = "Immunization_Rate"
rename percentile Percentile
keep if Income == "LISP" | Income == "HISP"
save "${created_data}\Immunization Rate.dta", replace
export excel using "${final_data}\FY21_Immunization_Rate.xlsx", firstrow(variables) replace


///////////////////////////
//  Health Expenditures  //
///////////////////////////

//Download Instructions
/*
Go to http://apps.who.int/nha/database/Select/Indicators/en
```

```
Select Indicators > Financing Sources >
Domestic General Government Health Expenditure (GGHE-D) as % of Gross Domestic Product (GDP)
Click Next, select all countries, click next
Select all years, click next
Select % Gross domestic product (GDP)
click view data and build report
Click Excel Workbook

Save as NHA indicators.xlsx in the raw data folder
check to see if the columns and rows still match.
*/
set more off
import excel "${raw_data}\NHA indicators.xlsx", sheet("Table") clear
drop B C
drop in 2

foreach var of varlist D-U {  //replacing the first row with values that have characters in them so that variables can be renamed using renvars
replace `var' = "Y" +`var'[1] in 1
}
renvars, map (word(@[1],1))
drop in 1
rename Countries Country
WBNaming Country
destring, replace
save "${created_data}\Health FY21.dta", replace

**Merging
*********
keep Country Y2017  //Keep the most recent year
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)
save "${created_data}\Health Full.dta", replace

preserve
keep if Income == "LISP"
PercentRankInc Y2017
save "${created_data}\Health LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Y2017
save "${created_data}\Health HISP.dta", replace
append using "${created_data}\Health LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Health_Expenditures"
rename percentile Percentile
save "${created_data}\Health Expenditures.dta", replace
export excel using "${final_data}\FY21_Health_Expenditures.xlsx", firstrow(variables) replace


////////////////////////////
//  Gender in the Economy //
////////////////////////////


//Download Instructions:
/*

In 2020, Women, Business and the Law substantially changed their underlying data.  MCC is still able
to produce its Gender in the Economy indicator from the available data, however multiple files must be accessed.
Additionally, MCC must use the information provided by WBL in the "About" tab of the 50 Year Panel
About disaggregated values.  This information must be manually coded into the data by MCC.
The three required files can be found here:

50 Year Data File
The Data files used by MCC can be found here: https://wbl.worldbank.org/content/dam/wbl/data/WBL-Dataset-50years.dta?logActivity=true
```

```
or here https://web.archive.org/web/20200815143655/https://wbl.worldbank.org/content/dam/wbl/data/WBL-Dataset-50years.dta?logActivity=true
The current version of this data file can be found here under Women, Business and the Law Data for 1971-2021
(Download file (STATA)) https://wbl.worldbank.org/en/wbl-data
Name this file "WBL-Dataset-50years.dta" and save in the raw data folder

Additional Data File:
Can be found here https://wbl.worldbank.org/content/dam/sites/wbl/documents/2021/02/WBLAdditionalData_07July2020.xlsx or
http://pubdocs.worldbank.org/en/666161589301125247/WBL-Additional-Data.xlsx
Name this file "WBLAdditionalData.xlsx" and save in the raw data folder

2020 Data File Release
Can be found here: http://pubdocs.worldbank.org/en/625831591105047863/WBL-50YearPanelDetails-Web-01Jun2020.xlsx or
here https://web.archive.org/web/20210305161933/http://pubdocs.worldbank.org/en/625831591105047863/WBL-50YearPanelDetails-Web-01Jun2020.xlsx
Name this file "WBL50YearPanelDetailsWeb01Jun2020.xlsx" and save in the raw data folder

The Numbers below correspond to the following questions:
Accessing Institutions
1.  Can an unmarried woman obtain a national ID card in the same way as an unmarried man? (1/2 point)
2.  Can a married woman obtain a national ID card in the same way as a married man? (1/2 point)
3.  Can an unmarried woman travel outside the country in the same way as an unmarried man? (1/2 point)
4.  Can a married woman travel outside the country in the same way as a married man? (1/2 point)
5.  Can an unmarried woman travel outside her home in the same way as an unmarried man? (1/2 point)
6.  Can a married woman travel outside her home in the same way as a married man? (1/2 point)
7.  Can an unmarried woman get a job or pursue a trade or profession in the same way as an unmarried man? (1/2 point)
8.  Can a married woman get a job or pursue a trade or profession in the same way as a married man? (1/2 point)
9.  Can an unmarried woman sign a contract in the same way as an unmarried man? (1/2 point)
10. Can a married woman sign a contract in the same way as a married man? (1/2 point)
11. Can an unmarried woman register a business in the same way as an unmarried man? (1/2 point)
12. Can a married woman register a business in the same way as a married man? (1/2 point)
13. Can an unmarried woman open a bank account in the same way as an unmarried man? (1/2 point)
14. Can a married woman open a bank account in the same way as a married man? (1/2 point)
15. Can an unmarried woman choose where to live in the same way as an unmarried man? (1/2 point)
16. Can a married woman choose where to live in the same way as a married man? (1/2 point)
17. Can an unmarried woman confer citizenship to children in the same way as an unmarried man? (1/2 point)
18. Can a married woman confer citizenship to children in the same way as a married man? (1/2 point)
19. Can an unmarried woman be “head of household” in the same way as an unmarried man? (1/2 point)
20. Can a married woman be “head of household” in the same way as a married man? (1/2 point)
21. If customary law is recognized as a valid source of law under the constitution, is it invalid if it violates constitutional provisions on nondiscrimination or equality? (1 point)
22. Do married couples jointly share legal responsibility for financially maintaining the family’s expenses? (1 point)
Using Property
23. Does the law provide for valuation of nonmonetary contributions? (1 point)
24. Who administers marital property? (1 point for “husband”)
25. Do unmarried men and unmarried women have equal ownership rights to property? (1/2 point)
26. Do married men and married women have equal ownership rights to property? (1/2 point)
27. Do sons and daughters have equal rights to inherit assets from their parents? (1 point)
28. Do female and male surviving spouses have equal rights to inherit assets? (1 point)
```

```
Going to Court
29. Does a woman’s testimony carry the same evidentiary weight in court as a man’s? (1 point)
Getting a Job
30. Can nonpregnant and nonnursing women work the same night hours as men? (1 point)
31. Can nonpregnant and nonnursing women do the same jobs as men? (1 point)
Protecting Women from Violence
32. Is there domestic violence legislation? (1 point)
33. Are there clear criminal penalties for domestic violence? (1 point)
34. Is there a specialized court or procedure for cases of domestic violence? (1 point)
35. Is there legislation that specifically addresses sexual harassment? (1 point)
36. Are there criminal penalties for sexual harassment in employment? (1 point)
37. What is the legal age of marriage for girls? (1 point for ages < 18 or no data)
38. Are there any exceptions to the legal age of marriage? (1 point for “yes”)
39. Does the law prohibit or invalidate child or early marriage? (1 point)
40. Are there penalties in the law for authorizing or knowingly entering into child or
early marriage? (1 point)
*/


set more off
use "${raw_data}\WBL-Dataset-50years.dta", clear
keep Economy Code Incomegroup Year gr1_2trvlctrymrd gr1_3trvlhmmrd gr1_4whlivemrd
gr2_5profhmmrd gr2_7sexhrssemp gr2_8sexcomb gr3_10nprgeqnight gr3_11jobshazard
gr3_12industry gr4_14hhmrd gr4_15domleg gr6_23cntrcthmmrd gr6_24regbusmrd gr6_25bnkaccmrd
gr6_26disgend gr7_27prtyeqownmrdbth gr7_28prtyeqsondght gr7_29prtyeqsuvrspse
gr7_30prtylegadmin gr7_31valnonmntry
keep if Year == 2020  //Changed 9.4.20
rename Code EconomyCode
save "${created_data}\WBL Index Data.dta", replace
//20

import excel "${raw_data}\WBLAdditionalData.xlsx", clear firstrow sheet("2019")
//import excel "${raw_data}\Archive\WBLAdditionalData.xlsx", clear firstrow sheet("2019")
//7.17.20
keep Economy Year EconomyCode IncomeGroup Whatisthelegalageofmarriag
Arethereanyexceptionstothe Ismarriageunderthelegalage Aretherepenaltiesforauthoriz
Istherelegislationthatspecif Doesthelegislationestablishc Isthereaspecializedcourtor
Canawomanobtainanationalid /*Canawomanlegallyconfercitiz*/ Canamarriedwomanlegallyconf
Cananunmarriedwomanlegallyc Domarriedcouplesjointlyshare Doesawomanstestimonycarryt
Iscustomarylawavalidsource Ifcustomarylawisavalidsour  //7.17.20
rename IncomeGroup Incomegroup
save "${created_data}\WBL Additional Data.dta", replace
//14

import excel "${raw_data}\WBL50YearPanelDetailsWeb01Jun2020.xlsx", clear firstrow sheet(
"Workplace")

keep EconomyCode EconomyName N
rename N CriminalPenaltiesSH

gen Year = 2019
rename EconomyName Economy
replace Economy = "Puerto Rico (U.S.)" if EconomyCode == "PRI"
drop in 1
merge 1:1 Economy using "${created_data}\WBL Index Data.dta", nogen
merge 1:1 Economy using "${created_data}\WBL Additional Data.dta"

//1//  gr7_29prtyeqsuvrspse
//2//  gr7_28prtyeqsondght
//3//  gr7_31valnonmntry
//4//  gr4_15domleg
//5//  Istherelegislationthatspecif
//6//  gr7_30prtylegadmin
//7//  gr6_23cntrcthmmrd
/*8*/  gen nw_contract_unm = "Yes" //Yes is a good thing
//9//  gr1_2trvlctrymrd
/*10*/ gen nw_travel_abd_unm = "Yes" //Yes is a good thing
       replace nw_travel_abd_unm = "No" if Economy == "Sudan"  //No is a bad thing
//11// gr6_25bnkaccmrd
/*12*/ gen nw_bank_account_unm = "Yes" //Yes is a good thing
```

```
//13// gr6_24regbusmrd
/*14*/ gen nw_start_business_unm = "Yes" //Yes is a good thing
//15// gr1_3trvlhmmrd
/*16*/ gen nw_travel_home_unm = "Yes" //Yes is a good thing
//17// Isthereaspecializedcourtor
//18// Doesthelegislationestablishc
//19// gr2_5profhmmrd
/*20*/ gen nw_get_job_unm = "Yes" //Yes is good
//21// gr7_27prtyeqownmrdbth // FIXED 7.15.20 // This does not track.  There are many more countries listed with restrictions here than listed in Tea's email
/*22*/ gen nw_property_own_unm = "Yes" //Yes is good
       replace nw_property_own_unm = "No" if Economy == "Tonga" //| Economy == "Eswatini" //No is bad  //FIXED 7.15.20//This does not track.  See above.
//23// gr1_4whlivemrd
/*24*/ gen nw_where_live_unm = "Yes"
//25// gr4_14hhmrd
/*26*/ gen nw_HoH_unm = "Yes" //Yes is a good thing
       replace nw_HoH_unm = "No" if Economy == "Sudan"  //No is a bad thing
//27// gr3_10nprgeqnight
/*28*/ gen nw_job_type = (gr3_11jobshazard & gr3_12industry)
       tostring nw_job_type, force replace
       replace nw_job_type = "Yes" if nw_job_type == "1"
       replace nw_job_type = "No" if nw_job_type == "0"
//29// CriminalPenaltiesSH
//30// Canawomanobtainanationalid
/*31*/ gen nw_nat_id_unm = "Yes" //Yes is good
       replace nw_nat_id_unm = "No" if Economy == "Oman"
//32// Arethereanyexceptionstothe
//33// Aretherepenaltiesforauthoriz
/*34*/ gen nw_citizen_mar = Canamarriedwomanlegallyconf
/*35*/ gen nw_citizen_unm = Cananunmarriedwomanlegallyc
//36// Domarriedcouplesjointlyshare
//37// Doesawomanstestimonycarryt
//38// Ismarriageunderthelegalage
//39// Ifcustomarylawisavalidsour Iscustomarylawavalidsource
//40// Whatisthelegalageofmarriag

keep Economy Year gr7_29prtyeqsuvrspse gr7_28prtyeqsondght gr7_31valnonmntry gr4_15domleg Istherelegislationthatspecif gr7_30prtylegadmin /* //6
*/ Isthereaspecializedcourtor Doesthelegislationestablishc gr3_10nprgeqnight nw_job_type CriminalPenaltiesSH Arethereanyexceptionstothe /* //6
*/ Aretherepenaltiesforauthoriz Domarriedcouplesjointlyshare Doesawomanstestimonycarryt Ismarriageunderthelegalage /* //4
*/ Ifcustomarylawisavalidsour Iscustomarylawavalidsource Whatisthelegalageofmarriag /* //2
*/ gr6_23cntrcthmmrd nw_contract_unm gr1_2trvlctrymrd nw_travel_abd_unm gr6_25bnkaccmrd nw_bank_account_unm gr6_24regbusmrd /* //7
*/ nw_start_business_unm gr1_3trvlhmmrd nw_travel_home_unm gr2_5profhmmrd nw_get_job_unm gr7_27prtyeqownmrdbth /* //6
*/ nw_property_own_unm gr1_4whlivemrd nw_where_live_unm gr4_14hhmrd nw_HoH_unm Canawomanobtainanationalid nw_nat_id_unm /* //7
*/ nw_citizen_mar nw_citizen_unm //2


global coded "gr1_2trvlctrymrd gr1_3trvlhmmrd gr1_4whlivemrd gr2_5profhmmrd gr3_10nprgeqnight gr4_14hhmrd gr4_15domleg gr6_23cntrcthmmrd gr6_24regbusmrd gr6_25bnkaccmrd gr7_27prtyeqownmrdbth gr7_28prtyeqsondght gr7_29prtyeqsuvrspse gr7_30prtylegadmin gr7_31valnonmntry"

foreach var of global coded {
    decode `var' , gen(`var'_d)
    drop `var'
    rename `var'_d `var'
}
//Questions worth 1 point
global fullpoint_indicators "gr7_29prtyeqsuvrspse gr7_28prtyeqsondght gr7_31valnonmntry gr4_15domleg Istherelegislationthatspecif gr7_30prtylegadmin Isthereaspecializedcourtor Doesthelegislationestablishc gr3_10nprgeqnight nw_job_type CriminalPenaltiesSH Aretherepenaltiesforauthoriz Domarriedcouplesjointlyshare Doesawomanstestimonycarryt Ismarriageunderthelegalage Ifcustomarylawisavalidsour"

```

```
//Questions worth half a point
global halfpoint_indicators "gr6_23cntrcthmmrd nw_contract_unm gr1_2trvlctrymrd nw_travel_abd_unm gr6_25bnkaccmrd nw_bank_account_unm gr6_24regbusmrd nw_start_business_unm gr1_3trvlhmmrd nw_travel_home_unm gr2_5profhmmrd nw_get_job_unm gr7_27prtyeqownmrdbth nw_property_own_unm gr1_4whlivemrd nw_where_live_unm gr4_14hhmrd nw_HoH_unm Canawomanobtainanationalid nw_nat_id_unm nw_citizen_mar nw_citizen_unm"

//Assigning one point to the full point indicators
foreach fullpoint of global fullpoint_indicators {
replace `fullpoint' = "1" if `fullpoint' == "No"
replace `fullpoint' = "0" if `fullpoint' == "Yes"
replace `fullpoint' = "0" if `fullpoint' == "yes"
replace `fullpoint' = "0" if `fullpoint' == "Yes "
replace `fullpoint' = "1" if `fullpoint' == "Husband"
replace `fullpoint' = "1" if `fullpoint' == "husband"
replace `fullpoint' = "0" if `fullpoint' == "Other"
replace `fullpoint' = "0" if `fullpoint' == "other"
replace `fullpoint' = "0" if `fullpoint' == "Both must agree"
replace `fullpoint' = "0" if `fullpoint' == "both must agree"
replace `fullpoint' = "0" if `fullpoint' == "Original owner"
replace `fullpoint' = "0" if `fullpoint' == "original owner"
replace `fullpoint' = "0" if `fullpoint' == "Original Owner"
replace `fullpoint' = "0" if `fullpoint' == "Separate with spousal consent"
replace `fullpoint' = "0" if `fullpoint' == "separate with spousal consent"
replace `fullpoint' = "." if `fullpoint' == "N/A"
replace `fullpoint' = "1" if `fullpoint' == ".."
}
//Assigning a half point to the half point indicators
foreach halfpoint of global halfpoint_indicators {
replace `halfpoint' = "0.5" if `halfpoint' == "No"
replace `halfpoint' = "0" if `halfpoint' == "Yes"
replace `halfpoint' = "0" if `halfpoint' == "yes"
replace `halfpoint' = "." if `halfpoint' == "N/A"
replace `halfpoint' = "0.5" if `halfpoint' == ".."
}

//Assigning the correct point value to the remaining questions on child marraige and customary law
rename Arethereanyexceptionstothe any_exceptions_to_the_leg_agef

replace any_exceptions_to_the_leg_agef = "1" if any_exceptions_to_the_leg_agef == "Yes"
replace any_exceptions_to_the_leg_agef = "1" if any_exceptions_to_the_leg_agef == "yes"
replace any_exceptions_to_the_leg_agef = "0" if any_exceptions_to_the_leg_agef == "No"
replace any_exceptions_to_the_leg_agef = "0" if any_exceptions_to_the_leg_agef == "no"
replace any_exceptions_to_the_leg_agef = "1" if any_exceptions_to_the_leg_agef == ".."

rename Whatisthelegalageofmarriag leg_age_of_marr_girls
replace leg_age_of_marr_girls = "1" if leg_age_of_marr_girls == ".."
destring, replace

replace leg_age_of_marr_girls = 1 if leg_age_of_marr_girls < 18
replace leg_age_of_marr_girls = 0 if leg_age_of_marr_girls >= 18
drop Year Iscustomarylawavalidsource
unab Allvars : _all
unab exclude : Economy
local Addlist : list Allvars - exclude
di "`Addlist'"
preserve
collapse (sum) `Addlist'
restore
egen Gender_score = rowtotal(`Addlist')
rename Economy Country
keep Country Gender_score

label variable Gender_score "2019"
WBNaming Country
save "${created_data}\Gender FY21.dta", replace

**Merging
*********
```

```

use "${created_data}\Gender FY21.dta", clear
keep Country Gender_score
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Gender Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Gender_score
replace percentile = 1-percentile //Note this is different from other indicators due to the inverse scale
save "${created_data}\Gender LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Gender_score
replace percentile = 1-percentile //Note this is different from other indicators due to the inverse scale
save "${created_data}\Gender HISP.dta", replace
append using "${created_data}\Gender LISP.dta"


keep Country Income percentile Score_In_FY21
gen Indicator = "Gender_In_The_Economy"
rename percentile Percentile
save "${created_data}\Gender In The Economy.dta", replace
export excel using "${final_data}\FY21_Gender_In_The_Economy.xlsx", firstrow(variables) replace

//////////////////////////////////
//  Natural Resource Protection //
//////////////////////////////////

//Download Instructions:
/*
Go to CIESIN's Natural Resource Management Index http://www.ciesin.columbia.edu/data/nrpi-chi-2020/
If the link does not redirect you to SEDAC's webpage, select "NRPI and CHI data download". If it does redirect:
Click on Data Sets, and select the most recent release of the Natural resource protection and Child Health Indicators
Click on Data Download.  You may need to log into NASA's website to view the data.
Select the zip release

Note, some of this data has been historically revised.  MCC is working with CIESIN to make sure the original
data that was shared on the scorecards is published.

Save as the file referenced below in the raw data folder
*/


import excel "${raw_data}\NRPI_CHI.xlsx", sheet("NRPI_2020") firstrow clear
drop if NRPI_v2020_20 == "NA" //Duplicate Sudan
destring _all, replace
WBNaming CountryName
label var CountryName "Country"
keep CountryName NRPI*
save "${created_data}\NRP FY21.dta", replace

**Merging
*********

use "${created_data}\NRP FY21.dta", clear
rename CountryName Country
keep Country NRPI_v2020_20
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)
save "${created_data}\NRP Full.dta", replace
preserve
```

```
keep if Income == "LISP"
PercentRankInc NRPI_v2020_20
save "${created_data}\NRP LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc NRPI_v2020_20
save "${created_data}\NRP HISP.dta", replace
append using "${created_data}\NRP LISP.dta"


keep Country Income percentile Score_In_FY21
gen Indicator = "Natural_Resource_Protection"
rename percentile Percentile
save "${created_data}\Natural Resource Protection.dta", replace
export excel using "${final_data}\FY21_Natural_Resource_Protection.xlsx", firstrow(variables) replace

///////////////////////////
//      Child Health     //
///////////////////////////
//Download Instructions:
//This should be contained in the same download as the NRP data above

import excel "${raw_data}\NRPI_CHI.xlsx", sheet("CHI_2020") firstrow clear
rename CountryName Country
keep Country CHI_v2020_19 CHI_v2020_18 CHI_v2020_17 CHI_v2020_16 CHI_v2020_15  CHI_v2020_14 CHI_v2020_13 CHI_v2020_12 CHI_v2020_11 CHI_v2020_10
WBNaming Country

save "${created_data}\Child FY21.dta", replace

**Merging
*********
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Child Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc CHI_v2020_19
save "${created_data}\Child LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc CHI_v2020_19
save "${created_data}\Child HISP.dta", replace
append using "${created_data}\Child LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Child_Health"
rename percentile Percentile
save "${created_data}\Child Health.dta", replace
export excel using "${final_data}\FY21_Child_Health.xlsx", firstrow(variables) replace


///////////////////////////////////
// Land Rights and Access (IFAD) //
///////////////////////////////////

//Download Instructions
/*
Go to https://webapps.ifad.org/members/gc
click on "All Sessions" on the Right to go here https://webapps.ifad.org/members/gc/sessions
Find the most recent Session, and look for "Programme of work and budget of IFAD and its Office of Evaluation for YEAR"
If this is not present, check a previous Session
Scroll to the bottom of this document to the section titled "YEAR Rural Sector Performance Assessment (RSPA) scores"
Find the indicator for Access to Land (3.3 in 2018)
Copy all the numbers into a txt file, replace all spaces with commas.  Save as a .csv file
Open CSV in excel, the numbers should be in different columns.
Copy the data, use special paste to reformat it so that it is in a single column.
Copy the country labels in one group at a time.  Make sure you use the same order as the original data.
(You may need to open the file in Adobe Acrobat instead of a web browser to do this).
Delete any cells where the country title is on more than one line. Shift Cells up.
Check to make sure that all of the country data is lined up correctly.
Label the columns "Country" and "Data", Label the sheet IFAD Data

Save as a Excel file (not a CSV) named IFAD Data.xlsx


If data is not available in the current year for a country, MCC pulls forward data from the previous report (back to 2014 for FY21).
This report can be found by going to https://webapps.ifad.org/members/gc/sessions, selecting the next previous session (check for the 2015 RSPA at the bottom of the document
then selecting the agenda item with the "IFAD's 2017 results-based programme of work and regular and capital budgets, the IOE results-based work programme and budget for 2017 and indicative plan for 2018-2019, and the HIPC and PBAS progress reports"
Once again scroll to the bottom of the document and repeat the steps above for the indicator Access to Land (B.i in this document).  This should be the 2015 Rural Sector Performance Assessment
Label the columns "Country" and "Data", label the sheet IFAD Data
Save as IFAD_2015.xlsx

Repeat this process, for the 2015 session (38th) of the governing council.  These numbers should be from the 2014 RSPA.
Save as IFAD_2014.xlsx

*/

set more off
import excel "${raw_data}\IFAD Data.xlsx", clear firstrow sheet("IFAD Data")
WBNaming Country
rename Data Y2020
save "${created_data}\IFAD_Data.dta", replace

import excel "${raw_data}\IFAD_2015.xlsx", clear firstrow sheet("IFAD Data")
WBNaming Country
rename Data Y2015
merge 1:1 Country using "${created_data}\IFAD_Data.dta", nogen
replace Y2020 = Y2015 if missing(Y2020)
save "${created_data}\IFAD_Data.dta", replace

import excel "${raw_data}\IFAD_2014.xlsx", clear firstrow sheet("IFAD Data")
WBNaming Country
rename Data Y2014
merge 1:1 Country using "${created_data}\IFAD_Data.dta", nogen
replace Y2020 = Y2014 if missing(Y2020)


egen Max_Score2020 = max(Y2020)
egen Min_Score2020 = min(Y2020)
gen IFAD_Score2020 = 1-((Max_Score2020-Y2020)/(Max_Score2020-Min_Score2020))

//keep Country IFAD_Score*
save "${created_data}\IFAD_Score_Normalized.dta", replace


//This indicator is only half of the Land Rights and Access Indicator
//We need to add in the Doing Business Registering Property data in order to create the full indicator.



//////////////////////
//   Doing Business  //
//////////////////////

//Download Instructions:
/*
Note that MCC used old Doing Business Data in FY21 due to concerns explained here:
```

```
https://www.mcc.gov/who-we-select/indicators/doing-business-indicators-fy21

Go to https://www.doingbusiness.org/en/data and select historical data create a custom dataset
This should take you to: https://www.doingbusiness.org/en/custom-query
Choose all Economies
Choose Topics "Starting a Business" "Getting Credit" and "Registering Property"
Select years back to 2015
Click Create report
Click the excel button to download a report
Save as the file below
*/


use "${raw_data}\Historical-data.dta" , clear
keep startbustime startbuscost startbustimeW startbuscostW registertime registercost creditinformation8pnt5pr creditrights_new economy dbyear

rename startbustime Time_Men_days
rename startbuscost Cost_Men_pc_of_income_pera
rename startbustimeW Time_Women_days
rename startbuscostW Cost_Women_pc_of_income_pi
rename registertime Time_days
rename registercost Cost_pc_of_property_value
rename creditrights_new Strength_of_legal_rights_1
rename creditinformation8pnt5pr  Depth_of_credit_informatiD
rename dbyear Year
rename economy Economy
//Replacing Year to Match the actual year of data collection
replace Year = Year - 1

drop in 1

//Keeping only the necessary variables and observations
drop if Year < 2014
drop if missing(Year) & Economy == ""

//Setting variable names up so that they can be reshaped, and reshaping the data to a wide format
unab Allvars : _all
unab exclude : Economy Year
local Addlist1 : list Allvars - exclude

foreach var of varlist `Addlist1' {
rename `var' `var'_
local `var'lb : variable label `var'
}

unab Allvars : _all
unab exclude : Economy Year
local Addlist2 : list Allvars - exclude
reshape wide `Addlist2' , i(Economy) j(Year)

//Labeling Variables
foreach var of local Addlist1 {
forvalues k = 2014(1)2019 {
label var `var'_`k' "``var'lb' `k'"
}
}

//Generating a maximum for each variable, and putting it in place of the "No Practice" values
unab Allvars : _all
unab exclude : Economy
local Addlist3 : list Allvars - exclude

foreach var of varlist `Addlist3' {
egen x`var' = max(`var')
replace `var' = x`var' if `var' == -9999
drop x`var'
}
```

```

//Merging
WBNaming Economy
rename Economy Country
label var Country "Country"
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)
drop GNI
save "${created_data}\All_Doing_Business_Formatted.dta" , replace

//This file will be used in what follows to create three indicators: Business Start Up, Access to Credit, and (with IFAD) Land Rights and Access.


////////////////////////
//  Access to Credit  //
////////////////////////
use "${created_data}\All_Doing_Business_Formatted.dta", clear
keep Country Strength_of_legal_rights_1_20* Depth_of_credit_informatiD_20*
//The following computes access to credit scores for a given country by normalizing both components and then taking an average.
forvalues k=14(1)19 {
gen Norm_Depth_`k' = Depth_of_credit_informatiD_20`k' * 12
gen Norm_Legal_`k' = Strength_of_legal_rights_1_20`k' * 8
gen Access_score_`k' = (Norm_Depth_`k' + Norm_Legal_`k')/2
}
save "${created_data}\Access Calculated.dta" , replace


**Merging
*********

use "${created_data}\Access Calculated.dta", clear
keep Country Access_score_19
rename Access_score_19 Access_score
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Access Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Access_score
save "${created_data}\Access LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Access_score
save "${created_data}\Access HISP.dta", replace
append using "${created_data}\Access LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Access_To_Credit"
rename percentile Percentile
save "${created_data}\Access To Credit.dta", replace
export excel using "${final_data}\FY21_Access_To_Credit.xlsx", firstrow(variables) replace


/////////////////////////
//  Business Start-Up  //
/////////////////////////

use "${created_data}\All_Doing_Business_Formatted.dta", clear
keep Country Cost_Men_pc_of_income_pera_20*  Cost_Women_pc_of_income_pi_20*  Country Time_Men_days_20*  Time_Women_days_20*
forvalues k= 14(1)19 {
egen Time_Average`k' = rowmean(Time_Men_days_20`k'  Time_Women_days_20`k')
egen Cost_Average`k' = rowmean(Cost_Men_pc_of_income_pera_20`k' Cost_Women_pc_of_income_pi_20`k')
//The following finds the min and max of these variables
egen Max_Time`k' = max(Time_Average`k')
egen Max_Cost`k' = max(Cost_Average`k')
egen Min_Time`k' = min(Time_Average`k')
egen Min_Cost`k' = min(Cost_Average`k')
//The following normalizes these variables to a common scale using the min and max
gen Time_Score`k' = (Max_Time`k'-Time_Average`k')/(Max_Time`k'-Min_Time`k')
gen Cost_Score`k' = (Max_Cost`k'-Cost_Average`k')/(Max_Cost`k'-Min_Cost`k')
gen Business_score_`k' = (Time_Score`k' + Cost_Score`k')/2
}
keep Country Time_Score* Cost_Score* Business_score_*
save "${created_data}\Business Calculated.dta" , replace

**Merging
**********
use "${created_data}\Business Calculated.dta", clear
keep Country Business_score_19
rename Business_score_19 Business_score
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Business Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Business_score
save "${created_data}\Business LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Business_score
save "${created_data}\Business HISP.dta", replace
append using "${created_data}\Business LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Business_Start_Up"
rename percentile Percentile
save "${created_data}\Business Start Up.dta", replace
export excel using "${final_data}\FY21_Business_Start_Up.xlsx", firstrow(variables) replace


///////////////////////////////////
//  Land Rights and Access (WB)  //
///////////////////////////////////

set more off
use "${created_data}\All_Doing_Business_Formatted.dta", clear
keep Country Time_days_20* Cost_pc_of_property_value_20*

forvalues k= 14(1)19 {
egen Max_Time`k' = max(Time_days_20`k')
egen Max_Cost`k' = max(Cost_pc_of_property_value_20`k')
egen Min_Time`k' = min(Time_days_20`k')
egen Min_Cost`k' = min(Cost_pc_of_property_value_20`k')
gen Time_Score`k' = (Max_Time`k'-Time_days_20`k')/(Max_Time`k'-Min_Time`k')
gen Cost_Score`k' = (Max_Cost`k'-Cost_pc_of_property_value_20`k')/(Max_Cost`k'-Min_Cost`k')
gen Rg_Prp_score_`k' = (Time_Score`k' + Cost_Score`k')/2
}

**Merging
*********
merge 1:1 Country using  "${created_data}\IFAD_Score_Normalized.dta"
gen LRA_score = (Rg_Prp_score_19 + IFAD_Score2020)/2

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\LRA Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc LRA_score
save "${created_data}\LRA LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc LRA_score
save "${created_data}\LRA HISP.dta", replace
```

```
append using "${created_data}\LRA LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Land_Rights_And_Access"
rename percentile Percentile
save "${created_data}\Land Rights And Access.dta", replace
export excel using "${final_data}\FY21_Land_Rights_And_Access.xlsx", firstrow(variables) replace


///////////////////////////////////
//   WGI Control of Corruption   //
///////////////////////////////////
//The following four indicators are drawn from the World Governance Indicators.
//The code for each follows the same procedures
/*
//Download instructions:
Go to https://info.worldbank.org/governance/wgi/ and download the excel file
Name the file wgidataset.xlsx and save in the raw data folder
*/

import excel "${raw_data}\wgidataset.xlsx", sheet("ControlofCorruption") clear
forvalues k = 1(1)13 {  //Dropping extraneous header information
drop in 1
}
foreach var of varlist _all {
replace `var' = `var'[2] +`var'[1] in 1 //Combining the year and the type (Estimate, Standard Deviation, etc.) into a single variable
replace `var' = "." if `var' == "#N/A"  //Setting #N/A as missing
replace `var' = "Country" if `var' == "Country/Territory"
}
drop in 2  //Dropping type now that it is incorporated into the first row
renvars, map (word(@[1],1))
drop in 1
destring, replace
save "${created_data}\Control of Corruption Full.dta" , replace

**Merging
*********

keep Country Estimate2019
rename Estimate2019 Corruption_Score

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)
save "${created_data}\Corruption Full.dta", replace

//Calculating Percentile Ranks
preserve
keep if Income == "LISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(Corruption_Score)
replace Corruption_Score = Corruption_Score - median
drop median
PercentRankInc Corruption_Score
save "${created_data}\Corruption LISP.dta", replace
restore
keep if Income == "HISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(Corruption_Score)
replace Corruption_Score = Corruption_Score - median
drop median
PercentRankInc Corruption_Score
save "${created_data}\Corruption HISP.dta", replace
append using "${created_data}\Corruption LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Control_Of_Corruption"
rename percentile Percentile
save "${created_data}\Control Of Corruption.dta", replace
```

```
export excel using "${final_data}\FY21_Control_Of_Corruption.xlsx", firstrow(variables) replace


//////////////////////////////////
//  WGI Government Effectiveness //
//////////////////////////////////

import excel "${raw_data}\wgidataset.xlsx", sheet("GovernmentEffectiveness") clear
forvalues k = 1(1)13 {  //Dropping extraneous header information
drop in 1
}
foreach var of varlist _all {
replace `var' = `var'[2] +`var'[1] in 1 //Combining the year and the type (Estimate, Standard Deviation, etc.) into a single variable
replace `var' = "." if `var' == "#N/A"  //Setting #N/A as missing
replace `var' = "Country" if `var' == "Country/Territory"
}
drop in 2  //Dropping type now that it is incorporated into the first row
renvars, map (word(@[1],1))
drop in 1
destring, replace
save "${created_data}\Government Effectiveness Full.dta" , replace

**Merging
***********

keep Country Estimate2019
rename Estimate2019 Gov_Eff_Score

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

//Calculating Percentiles
preserve
keep if Income == "LISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(Gov_Eff_Score)
replace Gov_Eff_Score = Gov_Eff_Score - median
drop median
PercentRankInc Gov_Eff_Score
save "${created_data}\Gov_Eff LISP.dta", replace
restore
keep if Income == "HISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(Gov_Eff_Score)
replace Gov_Eff_Score = Gov_Eff_Score - median
drop median
PercentRankInc Gov_Eff_Score
save "${created_data}\Gov_Eff HISP.dta", replace
append using "${created_data}\Gov_Eff LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Government_Effectiveness"
rename percentile Percentile
save "${created_data}\Government Effectiveness.dta", replace
export excel using "${final_data}\FY21_Government_Effectiveness.xlsx", firstrow(variables) replace


//////////////////////////////////
//       WGI Rule of Law        //
//////////////////////////////////

import excel "${raw_data}\wgidataset.xlsx", sheet("RuleofLaw") clear
forvalues k = 1(1)13 {  //Dropping extraneous header information
drop in 1
}
foreach var of varlist _all {
replace `var' = `var'[2] +`var'[1] in 1 //Combining the year and the type (Estimate,
Standard Deviation, etc.) into a single variable
replace `var' = "." if `var' == "#N/A"  //Setting #N/A as missing
replace `var' = "Country" if `var' == "Country/Territory"
}
drop in 2  //Dropping type now that it is incorporated into the first row
renvars, map (word(@[1],1))
drop in 1
destring, replace
save "${created_data}\Rule of Law Full.dta" , replace

keep Country Estimate2019
rename Estimate2019 RuleofLaw_Score

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

//Calculate percentiles
preserve
keep if Income == "LISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(RuleofLaw_Score)
replace RuleofLaw_Score = RuleofLaw_Score - median
drop median
PercentRankInc RuleofLaw_Score
save "${created_data}\RuleofLaw LISP.dta", replace
restore
keep if Income == "HISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(RuleofLaw_Score)
replace RuleofLaw_Score = RuleofLaw_Score - median
drop median
PercentRankInc RuleofLaw_Score
save "${created_data}\RuleofLaw HISP.dta", replace
append using "${created_data}\RuleofLaw LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Rule_Of_Law"
rename percentile Percentile
save "${created_data}\Rule Of Law.dta", replace
export excel using "${final_data}\FY21_Rule_Of_Law.xlsx", firstrow(variables) replace

///////////////////////////////////
//    WGI Regulatory Quality     //
///////////////////////////////////

import excel "${raw_data}\wgidataset.xlsx", sheet("RegulatoryQuality") clear
forvalues k = 1(1)13 {  //Dropping extraneous header information
drop in 1
}
foreach var of varlist _all {
replace `var' = `var'[2] +`var'[1] in 1 //Combining the year and the type (Estimate,
Standard Deviation, etc.) into a single variable
replace `var' = "." if `var' == "#N/A"  //Setting #N/A as missing
replace `var' = "Country" if `var' == "Country/Territory"
}
drop in 2  //Dropping type now that it is incorporated into the first row
renvars, map (word(@[1],1))
drop in 1
destring, replace
save "${created_data}\Regulatory Quality Full.dta" , replace

keep Country Estimate2019
rename Estimate2019 RegQuality_Score

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

//Calculating Percentiles
save "${created_data}\RegQuality Full.dta", replace
preserve
```

```
keep if Income == "LISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(RegQuality_Score)
replace RegQuality_Score = RegQuality_Score - median
drop median
PercentRankInc RegQuality_Score
save "${created_data}\RegQuality LISP.dta", replace
restore
keep if Income == "HISP"
//For all the WGI indicators, we shift all values so that 0 is the median
egen median = median(RegQuality_Score)
replace RegQuality_Score = RegQuality_Score - median
drop median
PercentRankInc RegQuality_Score
drop median
save "${created_data}\RegQuality HISP.dta", replace
append using "${created_data}\RegQuality LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Regulatory_Quality"
rename percentile Percentile
save "${created_data}\Regulatory Quality.dta", replace
export excel using "${final_data}\FY21_Regulatory_Quality.xlsx", firstrow(variables) replace




///////////////////////////////////
//    Freedom of Information     //
///////////////////////////////////
// The Freedom of the Press indicator is a composite of three sub indicators
// The main sub indicator is Freedom of the Press, with two minor sub indicators focused on Right to Information, and Internet shutdowns
//The following code imports, cleans and combines these files to create scores for each country.

////////////////////////////
//  Freedom of the Press  //
////////////////////////////
// In previous years, the Freedom of the Press Indicator was drawn from Freedom House's Freedom of the Press Report
// However, Freedom House stopped producing this indicator in 2017.  This required MCC to find another source for these data
// This year, MCC began using the Reporters Without Borders (RSF) Press Freedom Index in place of Freedom House.
// However, RSF was missing data for several small island nations, therefore MCC uses percentile matching as shown below
// to impute RSF scores for these countries based on data from previous years of Freedom House.
//Download Instructions:
/*
Go to the RSP website.  Go to the Index details (https://rsf.org/en/ranking_table), and select download the dataset
Save as "RSF dataset 2019.csv"
For the historical files: the "index_format_upload 2017.csv" can be found at the bottom of the page here "https://rsf.org/en/2017-press-freedom-index-ever-darker-world-map"
And the classment2016 file can be found at the bottom of this page: "https://rsf.org/en/2016-world-press-freedom-index-leaders-paranoid-about-journalists"
Go to the Freedom House website.  Go to the Freedom of the Press Report (https://freedomhouse.org/report-types/freedom-press), and select "Freedom of the Press Data"
Save as FOTP1980-FOTP2017_Public-Data.xlsx
*/

// The Following Sections format the RSF files for the Press Freedom Index
//RSF Data Formatting
set more off
//RSF 2020
import delimited "${raw_data}\RSF Data\RSF Dataset 2020.csv", clear
keep en_country score2020
rename en_country Country
rename score2020 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2020.dta", replace

//RSF 2019
import delimited "${raw_data}\RSF Data\RSF Dataset 2019.csv", clear
keep en_country score2019
rename en_country Country
rename score2019 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2019.dta", replace

//RSF 2018
import delimited "${raw_data}\RSF Data\RSF Dataset 2019.csv", clear
keep en_country score2018
rename en_country Country
rename score2018 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2018.dta", replace

//RSF 2017
import delimited "${raw_data}\RSF Data\index_format_upload 2017.csv", delimiter(";") clear
//This file is mislabeled.  The data that is labeled as "overallscore2016" is actually 2017 data (according to the main map).  The score2015 data is actually 2016 data (as it matches data from the 2016 file).
keep en_country overallscore2016  //This is actually 2017 data
rename en_country Country
rename overallscore2016 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2017.dta", replace

//RSF 2016
import delimited "${raw_data}\RSF Data\index_format_upload 2017.csv", delimiter(";") clear
//This file is mislabeled.  The data that is labeled as "overallscore2016" is actually 2017 data (according to the main map).  The score2015 data is actually 2016 data (as it matches data from the 2016 file).
keep en_country score2015  //This is actually 2016 data
rename en_country Country
rename score2015 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2016.dta", replace

//RSF 2015
import delimited "${raw_data}\RSF Data\classement2016.csv", delimiter(";") clear
keep en_country score2015
rename en_country Country
rename score2015 RSFscore

replace RSFscore = subinstr(RSFscore,",",".",.)
destring, replace
WBNaming Country
save "${created_data}\RSF_Scores_2015.dta", replace


//The following section formats the Freedom House data
```

```
//FH Data Formatting
import excel "${raw_data}\FOTP1980-FOTP2017_Public-Data.xlsx", clear sheet(Data)
replace A = "Country" in 5
//The following section fills in missing year values for observations which were merged in the original document
qui ds  //List all variables
local variable_tally : word count `r(varlist)'  //Count the number of variables
local tally_minus = `variable_tally' - 1 //Create a macro which is one less than the full variable count
forvalues k = 1(1)`tally_minus' {
local left : word `k' of `r(varlist)'  //Finding the "left" value for the year observation
local right : word `=`k'+1' of `r(varlist)' //Finding the "right" value for the year observation
replace `right' = `left' if `right'[3] == "" in 3  //Filling forward the year values when the values are missing
}
//Dropping Statuses and Sub-Scores
foreach var of varlist _all {
if !(`var'[5] == "Print" | `var'[5] == "Broadcast" | `var'[5] == "Status" | `var'[5] == "Total Score" | `var'[5] == "Country")  {
  drop `var'
}
}

//Dropping extraneous headings
drop in 4
drop in 2
drop in 1
replace A = "Country" in 1
//The following section of code combines the year headings with the data type headings
foreach var of varlist _all {
label var `var' "`=`var'[1]'"
replace `var' = "Y`=`var'[1]'_`=`var'[2]'" in 1
replace `var' = subinstr("`=`var'[1]'","-","_",.) in 1
}
replace A = "Country" in 1

renvars, map (word(@[1],1))
drop in 1
drop in 1

destring, replace
WBNaming Country
save "${created_data}\Freedom_House_Press_Formatted.dta", replace


**Imputation
************
//Now that we have formatted the two files, we need to merge them and
//use the freedom house data to impute the new scores using percentile
//matching

set more off
//This loop takes each RSF year and uses the freedom house data to impute the correct scores.
forvalues RSF_year = 2015(1)2020 {

use "${created_data}\Freedom_House_Press_Formatted.dta", clear

merge 1:1 Country using "${created_data}\RSF_Scores_`RSF_year'.dta", nogen keep(1 3)
//Since the most recent Freedom House data is from 2017, when the RSF year is 2017, 2018, or 2019,
// we use that Freedom House data to impute the scores.  However in earlier years,
// we use matching year data:
if "`RSF_year'" == "2020" |"`RSF_year'" == "2019" | "`RSF_year'" == "2018" | "`RSF_year'" == "2017" {
rename Y2017_Total FHScore

}
else {
rename Y`RSF_year'_Total FHScore
}

replace FHScore = "." if FHScore == "-"
keep Country RSFscore FHScore
destring, replace
//In this line we are making the decision to only consider countries that are in the GNI dataset as described above.
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)
//In order to calculate the percentile matching scores, we need to calculate the number
//of Freedom House Scores in the full dataset.
egen n = count(FHScore)
//We also need to calculate a ranking (1+ the number of scores that are lower)
egen i = rank(FHScore), track
//The Percentile rank in the full Freedom House dataset is then calculated as the rank minus 1, divided by the count minus 1
gen pcrank_FH = (i - 1) / (n - 1)
drop i n
egen RSF_count = count(RSFscore)
//Now, we use the following calculations to determine where the rank of the freedom house percentile
//in the RSF dataset, and split it into its whole number and decimal part
gen FH_Rank_in_RSF = (pcrank_FH*(RSF_count-1))+1
gen FH_floor = floor(FH_Rank_in_RSF)
gen FH_decimal = FH_Rank_in_RSF - FH_floor

sort RSFscore
local RSF_cnt = `=RSF_count[1]'
//This code creates a set of global macros corresponding to the ranking of the RSF scores
qui forvalues k=1(1)`RSF_cnt' {
local k_plus = `k' + 1
global RSF_function_`k' = `=RSFscore[`k']'
noisily di "${RSF_function_`k'}  `k'"
}
gen function_x = .
gen function_x_plus = .
local count = _N
//The following code finds the RSF scores which rank on either side of
//The Freedom House ranking, and then adding the product of the decimal
//part of the Freedom House ranking and the difference of the two RSF scores
//on either side.
qui forvalues k = 1(1)`count' {
replace function_x = ${RSF_function_`=FH_floor[`k']'} in `k'
local floor_plus = `=FH_floor[`k']' + 1
local count_minus = `count' -1
if "`=FH_floor[`k']'" == "`RSF_cnt'" {
noi di "`k'"
local floor_plus = `=FH_floor[`k']'
noisily di "${RSF_function_`floor_plus'}"
}
replace function_x_plus = ${RSF_function_`floor_plus'} in `k'
}
gen New_RSF_Score = function_x + (FH_decimal*(function_x_plus-function_x))
//This replaces the missing RSF scores with the newly imputed scores based on the FH data
replace RSFscore = New_RSF_Score if missing(RSFscore)
drop New_RSF_Score function_x_plus function_x FH_decimal FH_floor FH_Rank_in_RSF RSF_count pcrank_FH GNI FHScore
rename RSFscore Y`RSF_year'_Total
label var Y`RSF_year'_Total "`RSF_year'"
gen Y`RSF_year'_Status = "NF"
label var Y`RSF_year'_Status "`RSF_year'"
WBNaming Country
save "${created_data}\Freedom_of_The_Press_Imputed_`RSF_year'.dta", replace
}
//Merging each of the year files together
use "${created_data}\Freedom_of_The_Press_Imputed_2015.dta", clear
forvalues RSF_year = 2016(1)2019 {
merge 1:1 Country using "${created_data}\Freedom_of_The_Press_Imputed_`RSF_year'.dta", nogen
}
save "${created_data}\FOP FY21.dta", replace

```

```
//Before we can create the final file, we need to process the Key Internet Controls and the Right to Information files.

//////////////////////////
//      Access Now      //
//////////////////////////
//Download instructions:
//Download from https://www.accessnow.org/keepiton/ by clicking "Download the STOP Data"

import delimited "${raw_data}\KeepItON-Internet-Shutdown-Data-STOP-Shutdown-Tracker-2019-Complete.csv", bindquote(strict) clear
WBNaming country
replace duration = "248"  if duration == "248 at the time of publication "
replace duration = "1" if duration == "0" | duration == "2 hours " |  /*
*/ duration == "For about two hours " | duration == "for a few hours "  // | duration == ""

rename end_date DateEnded

replace DateEnded = subinstr(DateEnded," ","",.)
replace DateEnded = start_date if DateEnded == "Unkown" | DateEnded == "Unko"
replace DateEnded = "12/31/2019" if DateEnded == "ongoing" | DateEnded == "Ongoing"
gen length = strlen(DateEnded)
gen first = strpos(DateEnded, "/")
gen second = strrpos(DateEnded, "/")
gen month = substr(DateEnded, 1, first-1)
gen day = substr(DateEnded, first+1, (second-first)-1)
gen year = substr(DateEnded, second+1, 4)
destring, replace
gen DateEnd = mdy(month, day, year)
format DateEnd %td
drop length first second month day year

rename start_date DateBegan
replace DateBegan = subinstr(DateBegan," ","",.)

gen length = strlen(DateBegan)
gen first = strpos(DateBegan, "/")
gen second = strrpos(DateBegan, "/")
gen month = substr(DateBegan, 1, first-1)
gen day = substr(DateBegan, first+1, (second-first)-1)
gen year = substr(DateBegan, second+1, 4)
destring, replace
gen DateStart = mdy(month, day, year)
format DateStart %td
gen duration_new = DateEnd - DateStart
replace duration_new = 1 if duration_new == 0 | duration_new == .
compare duration duration_new
replace duration = duration_new if missing(duration)
tab duration

drop if ordered_by == "Non-government"

destring, replace
gen count = 1
collapse (sum) count duration, by (country)

gen AN_Binary = 1
gen AN_Duration = duration
rename country Country
keep Country AN_Binary AN_Duration
gen Internet_Shutdowns = AN_Duration
replace Internet_Shutdowns = 9 if Internet_Shutdowns>9
label var Internet_Shutdowns "Internet Shutdowns (Duration)"
keep Country Internet_Shutdowns
save  "${created_data}\Freedom of Internet.dta", replace

//Before we can create the indicator, we need to process the Right to Information file.

///////////////////////////
```

```
//  Right to Information  //
////////////////////////////
//Download Instructions: Go to the Country data page: https://www.rti-rating.org/country-data/
// Click on "Excel" download.  Save as RTI Ratings in the raw data folder
import excel "${raw_data}\RTI Ratings.xlsx", sheet("Sheet1") firstrow clear
keep Country Total
WBNaming Country
save "${created_data}\Right_To_Information.dta", replace


**Freedom of Information Merging
***************************************
//Merging files together:
use "${created_data}\Right_To_Information.dta", clear
merge 1:1 Country using "${created_data}\Freedom of Internet.dta", nogen
merge 1:1 Country using "${created_data}\Freedom_of_The_Press_Imputed_2020.dta", nogen
keep Country Total Internet_Shutdowns Y2020_Total
gen RTI_score = 4 if !missing(Total)
replace Internet_Shutdowns = 9 if !missing(Internet_Shutdowns) & Internet_Shutdowns>9
recode  RTI_score Internet_Shutdowns (.=0)
gen FOI_score = (Y2020_Total - RTI_score) + Internet_Shutdowns

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\FOI Testing Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc FOI_score
replace percentile = 1-percentile  //Note this is different for FOI due to the inverted scale!
save "${created_data}\FOI Testing LISP.dta", replace
export excel using "${testing_data}\FOI Testing LISP.xls", firstrow(variables) replace
restore
keep if Income == "HISP"
PercentRankInc FOI_score
replace percentile = 1-percentile  //Note this is different for FOI due to the inverted scale!
save "${created_data}\FOI Testing HISP.dta", replace
export excel using "${testing_data}\FOI Testing HISP.xls", firstrow(variables) replace
append using "${created_data}\FOI Testing LISP.dta"
save "${created_data}\FOI Testing.dta", replace
gen Indicator = "Freedom_Of_Information"
rename percentile Percentile
keep Country Income Percentile Score_In_FY21 Indicator
export excel using "${final_data}\FY21_Freedom_Of_Information.xlsx", firstrow(variables) replace


/////////////////////////
//    Trade Policy     //
/////////////////////////
/* Download Instructions:
As of the publication of this do file, Heritage's scores for trade freedom are not public for all MCC candidate countries
However, when they become public they will be available here https://www.heritage.org/index/download
Select "Download Raw Data" and save the file as index2021_data.xls in the raw data folder
*/

set more off
import excel "${raw_data}\2021 Trade Freedom MCC.xlsx", clear firstrow
keep Country TradeFreedom
replace TradeFreedom = "." if TradeFreedom == "N/A"
destring, replace
rename TradeFreedom Trade_Policy_Score
WBNaming Country

**Merging
```

```
************
WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Trade Policy Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Trade_Policy_Score
save "${created_data}\Trade Policy LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Trade_Policy_Score
save "${created_data}\Trade Policy HISP.dta", replace
append using "${created_data}\Trade Policy LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Trade_Policy"
rename percentile Percentile
save "${created_data}\Trade Policy.dta", replace
export excel using "${final_data}\FY21_Trade_Policy.xlsx", firstrow(variables) replace


/////////////////////////
//    WEO/IMF Data     //
/////////////////////////
/*
Download instructions
Use the October 2019 release
Go to https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx
Select by countries
Select all countries
Select Inflation, average consumer prices (Percent change) and General government net 
lending/borrowing (Percent of GDP)
Select Start 1980, end 2024

*/
set more off
import delimited "${raw_data}\WEOOct2020all.xls", clear varnames(nonames)
drop v57
foreach var of varlist v10-v55 {

replace `var' = "Y`=`var'[1]'" in 1
}
foreach var of varlist _all {
replace `var' = subinstr("`=`var'[1]'"," ", "_",.) in 1
replace `var' = subinstr("`=`var'[1]'","/", "_",.) in 1
replace `var' = subinstr("`=`var'[1]'","-", "_",.) in 1

replace `var' = "." if `var' == "n/a"
replace `var' = "." if `var' == "--"
replace `var' = subinstr(`var',",","",.)
}
renvars, map (word(@[1],1))
drop in 1
destring, replace
drop Y2020 Y2021 Y2022 Y2023 Y2024 Y2025 Estimates_Start_After
foreach var of varlist Y* {
lab var `var' `=subinstr("`var'", "Y","",.)'
}
save "${created_data}\WEO_IMF_Data_Full.dta" , replace


//////////////////////
//  Fiscal Policy   //
//////////////////////
//This pulls directly from an IMF indicator
set more off
use "${created_data}\WEO_IMF_Data_Full.dta", clear
keep if Subject_Descriptor == "General government net lending/borrowing"
keep if Units == "Percent of GDP"
keep Country Y*
WBNaming Country

**Merging
*********
//Fiscal Policy is a three year rolling average
egen Fiscal_Policy_Score = rowmean(Y2017 Y2018 Y2019)
keep Country Fiscal_Policy_Score

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Fiscal Policy Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Fiscal_Policy_Score
save "${created_data}\Fiscal Policy LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Fiscal_Policy_Score
save "${created_data}\Fiscal Policy HISP.dta", replace
append using "${created_data}\Fiscal Policy LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Fiscal_Policy"
rename percentile Percentile
save "${created_data}\Fiscal Policy.dta", replace
export excel using "${final_data}\FY21_Fiscal_Policy.xlsx", firstrow(variables) replace

//////////////////////
//    Inflation     //
//////////////////////
//This pulls directly from an IMF indicator
set more off
use "${created_data}\WEO_IMF_Data_Full.dta", clear
keep if Subject_Descriptor == "Inflation average consumer prices"
keep if Units == "Percent change"
keep Country Y*
WBNaming Country

**Merging
*********
rename Y2019 Inflation_Rate
keep Country Inflation_Rate

WBNaming Country
merge 1:1 Country using "${created_data}\GNI_FY21.dta", nogen keep(3)

save "${created_data}\Inflation Full.dta", replace
preserve
keep if Income == "LISP"
PercentRankInc Inflation_Rate
replace percentile = 1-percentile  //Note this is different for Inflation due to the inverted scale
save "${created_data}\Inflation LISP.dta", replace
restore
keep if Income == "HISP"
PercentRankInc Inflation_Rate
replace percentile = 1-percentile  //Note this is different for Inflation due to the inverted scale
save "${created_data}\Inflation HISP.dta", replace
append using "${created_data}\Inflation LISP.dta"

keep Country Income percentile Score_In_FY21
gen Indicator = "Inflation"
rename percentile Percentile
save "${created_data}\Inflation.dta", replace
export excel using "${final_data}\FY21_Inflation.xlsx", firstrow(variables) replace


```