

Institutional Quality and Citizen Compliance- Data Cleaning Project for Programming Course

Rasheed Ibrahim

1 Data Cleaning Project for Programming Course

Description:

Data cleaning project for programming course. This is a data cleaning and merging process of three waves (waves 5, 6, and 7) of afrobarometer survey data for subsequent data analysis and then merged with liberal democracy indicator from vdem and corruption control data from the world bank.

Author:

Rasheed Ibrahim - University of Notre Dame - ribrahim@nd.edu

Last Updated:

May 10, 2023

2 COMPUTE ENVIRONMENT

Instructions: This project is completed in R. Thus, to run all the codes, you need to install R on your computer. Currently, these codes run on R version 4.2.2 which is the latest version of R available. The codes are mostly written for base R. However a few packages are still required, namely “wbstats” for worldbank api, “countrycodes” for converting country names to codes, and “tidyr” for cleaning and pivoting data.

R version R version 4.2.2 (2022-10-31)

Platform: aarch64-apple-darwin20

Running under: MacOS Ventura Version 13.1.

Attached Packages

- tidyr_1.3.0
- countrycode_1.4.0
- wbstats_1.0.4
- ggplot2_3.4.2
- dplyr_1.1.2
- RColorBrewer_1.1-3
- sf_1.0-9

3 FILE LIST

Data

1. “wave5_cleaned.csv” is the output file after cleaning wave 5, i.e. after running “wave 5 cleaning.R”
2. “wave6_cleaned.csv” is the output file after cleaning wave 6, i.e. after running “wave 6 cleaning.R”
3. “wave7_cleaned.csv” is the output file after cleaning wave 7, i.e. after running “wave 7 cleaning.R”
4. “vdem.csv” is a predownloaded data from varieties of democracy website.
5. “forAnalysis.csv” is the final dataset after merging everything, i.e. after running “merging data.R”
6. afshape is the folder containing African countries shape files used for creating geographical maps.

Code

1. “wave 5 cleaning.R” is the R script for scraping and cleaning wave 5 of the afrobarometer survey
2. “wave 6 cleaning.R” is the R script for scraping and cleaning wave 6 of the afrobarometer survey
3. “wave 7 cleaning.R” is the R script for scraping and cleaning wave 5 of the afrobarometer survey
4. “merging data.R” is the R script that binds all three afrobarometer data from waves 5, 6, 7, and merges them with liberal democracy and corruption control data.
5. “data-viz.R” is the R script for plotting the data

Outputs

1. “corruptionmap3.pdf” is a geospatial map generated using ggplot. The codes for this plot are in “data-viz.R”. The map spatially graphs corruption control by country.
2. “controlmap-baseR.pdf” is a geospatial map generated using ggplot. The codes for this plot are “data-viz.R” The map spatially graphs corruption control by country.

Documentation

1. README - File for instructions and description of the project

4 Project Update

As of 10th May, 2023, I have updated this readme file to account for all documents and scripts in this project. Note to colleagues: Please feel free to reach out to me if you spot any mistake or want to chat about my codes or the API that I used at ribrahim@nd.edu