Bridging Conflicts and Biodiversity Protection: The Critical Role of Reliable and Comparable Data

Overview

The code in this replication package constructs the analysis file from the data sources on the World Bank Development Data Hub and local data using R. A main script run all of the code to generate the data for the figures and tables in the World Bank Policy Research Working Paper entitled, "Bridging Conflicts and Biodiversity Protection: The Critical Role of Reliable and Comparable Data" (No. X). The replicator should expect the code to run for about <2 hours.

Data Availability and Provenance Statements

Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

License for Data

Summary of Availability

- Some data **cannot be made** publicly available.

Dataset list

Below is a comprehensive list of datasets from `local/geo_data` and `local/tab_data`, organized by source and grouped by access level.

Geographic Data

Source	Dataset Name	Files	Access	Details
TFDD	Transboundary Freshwater Diplomacy Database	`geo_data/TFDD_SpatialData_Public202203/BasinMaster311_20220224.*geo_data/TFDD_SpatialData_Public202203/BasinMaster313_20240807.*`	Public	Access here
IBGE via HDX	Brazil Admin	`geo_data/BRA_adm1.*`	Public	Shapefile with metadata.

Source	Dataset Name	Files	Access	Details
	Level 1 Boundaries			Available here
World Bank (Restricted)	Admin0 Disputed Areas (10m)	`geo_data/ne_10m_WB2019_admin_0_disputed.*`	Restricted	Provided by the WB Cartography Unit. Not publicly available.
World Bank (Restricted)	GAD Disputes	`geo_data/WB_GAD_Disputes.*`	Restricted	Provided by the WB Cartography Unit. Not publicly available.
World Bank (Restricted)	EEZ + Disputed Merged	`geo_data/World_Bank_Countries_Disputed_merged_EEZ_land_and_ocean2024-05-09.*`	Restricted	Derived using restricted data from WB Cartography Unit. Not publicly available.
World Bank (Restricted)	Admin0 Boundaries (50m, incl. disputes & coastlines)	`geo_data/Boundary_Data_50mil/ne_50m_WB2019_*.shp`and associated `.shx`, `.dbf`, `.prj`, `.xml`, etc.	Restricted	Provided by the WB Cartography Unit. Not publicly available.
WorldPop	Aggregated Population 2020 (1km)	`geo_data/pop/ppp_2020_1km_Aggregated.tif`	Public	API access. Code included. Visit WorldPop
Authors	Species Occurrence Region Examples	`geo_data/species_occurrence_region_examples.*`	Authors	Will be made available on World Bank Microdata Library (DDH)

📊 Tabular Data

Source	Dataset Name	Files	Access	Details
World Bank	FCV Country Classification	`tab_data/fcv.xlsx`	Public	Included. See full list (FY24)
Authors	Master Wheel Diagram Data	`tab_data/master_data_for_wheel_diagram.xlsx`	Public	Included in this package.
Authors	Source Data for Figure 07	`tab_data/master_fig07.xlsx`	Public	Included in this package.
Authors	Species x Country Matrix	`tab_data/master_species_x_country2024-07-22.dta`	Restricted	Derived from restricted data. Not publicly available.

🛠️ How to Run the Code

To run the package, follow these steps:

1. Open the `.Rproj` file.
2. In the R console, run:

```
renv::restore()
```

3. Open and run the main script:

```
biod_intl_wb__main.R
```

💡 **Note:**
The repository includes a file named `folder_structure` that outlines the full structure of all datasets used in this project, including both public and restricted files. This allows users to understand the organization and flow of data even if access to some files is limited.

Computational requirements

Software Requirements

- The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.

- R 4.3.1 attached base packages: [1] grid parallel tools stats graphics grDevices utils datasets methods
  [10] base

other attached packages: [1] ggvenn_0.1.10 fasterize_1.0.5 XML_3.99-0.14 dplyr_1.1.3
[5] stars_0.6-4 abind_1.4-5 gt_0.10.0 xfun_0.40
[9] modelsummary_1.4.5 ordinal_2023.12-4 terra_1.7-55 archive_1.1.8
[13] RColorBrewer_1.1-3 R.utils_2.12.2 R.oo_1.25.0 R.methodsS3_1.8.2
[17] ggplot2_3.5.1 renv_1.0.5 tidyr_1.3.0 haven_2.5.3
[21] doParallel_1.0.17 iterators_1.0.14 foreach_1.5.2 data.table_1.14.8
[25] lubridate_1.9.3 ymd_0.1.0 readr_2.1.4 exactextractr_0.10.0 [29] Hmisc_5.1-2
stringr_1.5.1 httr_1.4.7 sf_1.0-19

loaded via a namespace (and not attached): [1] tidyselect_1.2.0 fastmap_1.1.1 digest_0.6.33
rpart_4.1.19
[5] timechange_0.2.0 lifecycle_1.0.4 cluster_2.1.4 magrittr_2.0.3
[9] compiler_4.3.1 rlang_1.1.5 utf8_1.2.3 knitr_1.44
[13] htmlwidgets_1.6.2 sp_2.1-2 classInt_0.4-11 xml2_1.3.5
[17] KernSmooth_2.23-21 numDeriv_2016.8-1.1 foreign_0.8-84 withr_3.0.0
[21] purrr_1.0.2 nnet_7.3-19 fansi_1.0.4 e1071_1.7-16
[25] colorspace_2.1-0 MASS_7.3-60 scales_1.3.0 insight_0.19.10
[29] cli_3.6.1 rmarkdown_2.25 generics_0.1.3 rstudioapi_0.15.0
[33] tzdb_0.4.0 DBI_1.2.3 proxy_0.4-27 base64enc_0.1-3
[37] vctrs_0.6.3 Matrix_1.6-1.1 hms_1.1.3 Formula_1.2-5
[41] htmlTable_2.4.2 units_0.8-5 glue_1.6.2 codetools_0.2-19
[45] stringi_1.7.12 gtable_0.3.4 raster_3.6-26 tables_0.9.17
[49] munsell_0.5.0 tibble_3.2.1 pillar_1.9.0 htmltools_0.5.6
[53] R6_2.5.1 ucminf_1.2.1 evaluate_0.21 lattice_0.22-5
[57] backports_1.4.1 class_7.3-22 Rcpp_1.0.11 nlme_3.1-162
[61] gridExtra_2.3 checkmate_2.3.1 forcats_1.0.0 pkgconfig_2.0.3

Controlled Randomness

- No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary

Approximate time needed to reproduce the analyses on a standard 2024 server machine: - 2 hours

Approximate storage space needed: - < 2GB

Details

Portions of the code were last run on a 32-core Intel server with 256 GB of RAM, 100 GB of network storage.

Description of programs/code

- Script biod_intl_wp__main.R is the main script
- Script biod_intl_wp_global_libraries.R loads the libraries for R
- Scripts starting with biod_intl_wp_load load the data
- Scripts starting with biod_intl_wp_fig construct the figures
- Scripts starting with biod_intl_wp_tbl construct the tables

List of tables and programs

The provided code reproduces: - All tables and figures in the paper

Figures and Tables – Code and Output Mapping

Figure/Table	Program / Dataset	Output File	Note
Figure 1	`master_data_for.xlsx`	Embedded in Word document	
	`species_info_all.R`	`fig_01_species_info_by.csv`	
Figure 2	`fig02_species_map.R`	`fig_02_species_occurrence.png`	
Figure 3	`fig03_country_in_fcs.R`	`fig_03_countries_fcs.png`	
Figure 4	`fig04_human_pop_in.R`	`fig_04_popden_in_trans.png`	
Figure 5	`fig05_venn_diagram_GOA.R`	`fig_05_species_GOA_venn_diag.png`	
Figure 6	`fig06_venn_diagram_oth.R`	`fig_06_species_GOA_or_not.png`	
Figure 7	`master_fig07.xlsx`	Embedded in Word document	
Figure 7 - NDLS	`...ndls_territories.R`	`disputed_species_all_summary.csv`	
Figure 7 - FCS	`...fcv_territories.R`	`fcv_species_all_summary.csv`	
Figure 7 - MJR	`...joint_territories.R`	`joint_species_all_summary.csv`	
Figure 7 - TRB	`...intl..._river_basins.R`	`tfddbasins_species_all_summary.csv`	
Figure 8	`master_fig07.xlsx`	Embedded in Word document	
Figure 8 - GLOBAL	`geographic_union_groups.R`	`global_species_all_summary.dta`	
Table 1	`tbl01_pop_and_area_and.R`	`tbl01_sensitive_groups_pop.csv`	
Table 2	`tbl02_spp_info_by_geo.R`	`Tbl02_groups_species_counts.csv`	
Table 3	`tbl03_spp_info_by_basin.`	`Tbl03_species_by_treaties.csv`	

Figure/Table	Program / Dataset	Output File	Note
	R		

References

Dasgupta, S., Blankespoor, B., & Wheeler, D. (2024a). Revisiting Global Biodiversity: A Spatial Analysis of Species Occurrence Data from the Global Biodiversity Information Facility (No. 10821). The World Bank.

Dasgupta, S., Blankespoor, B., & Wheeler, D. (2024b). Estimating Extinction Threats with Species Occurrence Data from the Global Biodiversity Information Facility (No. 10822). The World Bank.

Acknowledgements

The authors gratefully acknowledge financial support from the Global Environment Facility.