Estimating Extinction Threats with Species Occurrence Data from the Global Biodiversity Information Facility

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 11 figures and 4 tables in the World Bank Policy Research Working Paper entitled, "Estimating Extinction Threats with Species Occurrence Data from the Global Biodiversity Information Facility" (No. 10822). The replicator should expect the code to run for about 2-8 hours and require at least 40GB of data.

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

The data are licensed under a Creative Commons/CC-BY-NC license.

Summary of Availability

- Some data are not included in the reproducibility package, but they're available and the code automates its download. The user must run the code to download data from the IUCN Red List focusing on species in five IUCN Red List categories: Least Critical (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN) and Critically Endangered (CR): https://www.iucnredlist.org/resources/data-repository

Dataset list

Please note that the code downloads all these files automatically, except for the coastlines and boundary lines from the Wrold Bank.

Data file	Source	Notes	Provided
iucn-latest.zip	IUCN	Public	No
GBIF_Merged_IUCN_Risks_Tab.csv	GBIF/ Authors	Public	No
GBIF_IUCN_Red_List2022-2.csv	GBIF/ Authors	Public	No
global_biod_species_grid.csv	Authors	Public	No
global_biod_species_level_grid.rar	Authors	Public	No
ne_50m_WB_admin_0_boundary_ lines.shp	World Bank	Public	No
ne_50m_WB_admin_0_boundary_ lines_disputed.shp	World Bank	Public	No
ne_50m_WB_coastline.shp	World Bank	Public	No

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 [1] dplyr_1.1.3 stars_0.6-4 abind_1.4-5 gt_0.10.0 xfun_0.40 modelsummary_1.4.5
  [7] ordinal_2023.12-4 MASS_7.3-60 terra_1.7-55 archive_1.1.8 RColorBrewer_1.1-3 R.utils_2.12.2
  [13] R.oo_1.25.0 R.methodsS3_1.8.2 ggplot2_3.5.1 renv_1.0.5 tidyr_1.3.0 haven_2.5.3
  [19] doParallel_1.0.17 iterators_1.0.14 foreach_1.5.2 data.table_1.14.8 lubridate_1.9.3 ymd_0.1.0
  [25] readr_2.1.4 exactextractr_0.10.0 Hmisc_5.1-2 stringr_1.5.1 httr_1.4.7 sf_1.0-14

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: - 2-8 hours

Approximate storage space needed: - 25 GB - 250 GB

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_threats_wp___main.R is the main script
- Scripts starting with biod_threats_wp_load load the data
- Scripts starting with biod_threats_wp_fig construct the figures
- Scripts starting with biod_threats_wp_tbl construct the tables

List of tables and programs

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

Figure/Table #	Program	Output file	Note
Table 1	tbl01_corr_vars... .R	Table01_correlations... .csv	
Table 2a	tbl01_corr_vars... .R	table02a_ordered_logit... .csv	
Table 2b	tbl01_corr_vars... .R	table02b_ordered_logit... .csv	
Table 3	tbl03_model_predicted .R	Table03_model_predicted... .csv	
Table 4	tbl04_range_sizes .R	Table04_Range_size... .csv	
Figure 1	fig01_terrestrial .R	fig_01_terrestrial... .png	
Figure 2	fig02_plants_and_ .R	fig_02_GBIF_Plants... .png	
Figure 3	fig03_plants_and_ .R	fig_03_Threat_Status... .png	
Figure 4	fig04_plants_and_ .R	fig_04_Global_cell_totals... .png	
Figure 5	fig05_Normalized_ .R	fig_05_Normalized_GBIF... .png	
Figure 6	fig06_LC_assignment .R	fig_06_LC_assignment... .png	
Figure 7	fig07_plants_and_ .R	fig_07_Tprob_by_threat... .png	
Figure 8	fig08_plants_and_ .R	fig_08_Tprob_stacked... .png	
Figure 9	fig09_plants_and_ .R	fig_09_GBIF_Plants... .png	

Figure/Table #	Program	Output file	Note
Figure 10	fig10_GBIF_Plants_.R	fig_10_GBIF_dif_Tprob....png	
Figure 11	fig11_GBIF_Plants_.R	fig_11_plants_and_vert....png	

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.

IUCN (2022). The IUCN Red List of Threatened Species. Version 2022-2. https://www.iucnredlist.org. Downloaded on 2023-05-09. https://doi.org/10.15468/0qnb58 accessed via GBIF.org on 2023-11-17. accessed via GBIF.org on 2024-02-29.

Acknowledgements

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