

# Package ‘spatialData’

May 9, 2026

**Title** Spatial Datasets for Ecological Modeling

**Version** 1.0.0

**Description** Provides spatial datasets ready to use for ecological modelling and raster companion data for prediction: Neanderthal presence during the Last Interglacial (Benito et al. 2017 <[doi:10.1111/jbi.12845](https://doi.org/10.1111/jbi.12845)>); Plant diversity metrics for the World’s Ecoregions (Maestre et al. 2021 <[doi:10.1111/nph.17398](https://doi.org/10.1111/nph.17398)>); tree richness across the Americas (Benito et al. 2013 <[doi:10.1111/2041-210X.12022](https://doi.org/10.1111/2041-210X.12022)>); plant communities from the Sierra Nevada (Spain) with future climate scenarios (Benito et al. 2013 <[doi:10.1111/2041-210X.12022](https://doi.org/10.1111/2041-210X.12022)>); butterfly-plant interaction data from Sierra Nevada (Spain) (Benito et al. 2011 <[doi:10.1007/s10584-010-0015-3](https://doi.org/10.1007/s10584-010-0015-3)>); plant species occurrences in Andalusia (Spain) (Benito et al. 2014 <[doi:10.1111/ddi.12148](https://doi.org/10.1111/ddi.12148)>); presence of the plant *Linaria nigricans* and greenhouses (Benito et al. 2009 <[doi:10.1007/s10531-009-9604-8](https://doi.org/10.1007/s10531-009-9604-8)>); global NDVI and environmental predictors, and European oak species occurrences. All datasets include pre-processed environmental predictors ready for statistical modelling.

**URL** <https://blasbenito.github.io/spatialData/>

**BugReports** <https://github.com/BlasBenito/spatialData/issues>

**License** CC BY 4.0

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**Imports** sf, terra

**Suggests** spelling, testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Language** en-US

**Depends** R (>= 4.1.0)

**LazyData** true

**LazyDataCompression** xz

**NeedsCompilation** no

**Author** Blas M. Benito [aut, cre]

**Maintainer** Blas M. Benito <[blasbenito@gmail.com](mailto:blasbenito@gmail.com)>

**Repository** CRAN

**Date/Publication** 2026-04-21 19:50:32 UTC

## Contents

andalusia . . . . .	3
andalusia_extra . . . . .	4
andalusia_predictors . . . . .	4
andalusia_responses . . . . .	5
communities . . . . .	6
communities_extra_2010 . . . . .	7
communities_extra_2050 . . . . .	8
communities_extra_2100 . . . . .	8
communities_predictors . . . . .	9
communities_responses . . . . .	9
interaction . . . . .	10
interaction_extra . . . . .	11
interaction_predictors . . . . .	12
interaction_responses . . . . .	12
linaria . . . . .	13
linaria_extra . . . . .	15
linaria_predictors . . . . .	15
linaria_responses . . . . .	16
neanderthal . . . . .	16
neanderthal_extra . . . . .	18
neanderthal_predictors . . . . .	19
neanderthal_response . . . . .	19
plantae . . . . .	20
plantae_east . . . . .	27
plantae_extra . . . . .	28
plantae_predictors . . . . .	28
plantae_responses . . . . .	29
plantae_west . . . . .	29
quercus . . . . .	30
quercus_extra . . . . .	32
quercus_predictors . . . . .	33
quercus_response . . . . .	33
trees . . . . .	34
trees_extra . . . . .	37
trees_predictors . . . . .	38
trees_response . . . . .	38
vi . . . . .	39
vi_extra . . . . .	42
vi_predictors . . . . .	43
vi_responses . . . . .	43
vi_smol . . . . .	44

**Index**

**45**

---

andalusia	<i>Presence records of 90 plant species and background points from Andalusia, Spain</i>
-----------	---

---

### Description

sf long format data frame with POINT geometry and CRS ETRS89 / UTM zone 30N (EPSG:25830), containing 37,773 presence records for 90 plant species and 8,692 background points (46,465 rows total) from Andalusia, Spain.

The dataset contains 3 columns (species, presence, geometry). Environmental predictors for each point can be extracted from the companion raster returned by `andalusia_extra()`. Predictor names are stored in `andalusia_predictors`.

### Usage

```
data(andalusia)
```

### Format

An sf data frame with 46,465 rows (presences and background points) and 3 columns:

- `species`: Character string (species name or "background"). Suitable for classification models.
- `presence`: Binary integer stored as integer (1 = confirmed species presence, 0 = background point).
- `geometry`: `sfc_POINT` column with coordinates in EPSG:25830.

### Source

#### Published study:

- Benito, B.M., Lorite, J., Pérez-Pérez, R., Gómez-Aparicio, L., & Peñas, J. (2014). Forecasting plant range collapse in a mediterranean hotspot: when dispersal uncertainties matter. *Diversity and Distributions*, 20(1), 72–83. doi:10.1111/ddi.12148

#### Landsat imagery:

- Nunes de Lima, M. V. (Ed.) (2005). *IMAGE2000 and CLC2000 - Products and methods*. Joint Research Centre, Institute for Environment and Sustainability, and European Environment Agency. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/84dd2bad-14d9-4a65-9b92-3b4507d09e44/language-en>

#### Climate variables:

- Ninyerola, M., Pons, X. & Roure, J.M. (2005). Atlas Climático Digital de la Península Ibérica: Metodología y aplicaciones en bioclimatología y geobotánica. Universidad Autónoma de Barcelona, Bellaterra.

#### Topography:

- Instituto Geográfico Nacional. Modelo Digital del Terreno (MDT25). <https://www.ign.es>

**See Also**

Other andalusia: [andalusia\\_extra\(\)](#), [andalusia\\_predictors](#), [andalusia\\_responses](#)

**Examples**

```
data(andalusia)
colnames(andalusia)
nrow(andalusia)
ncol(andalusia)
```

---

andalusia_extra	<i>Environmental raster for the dataset andalusia</i>
-----------------	---

---

**Description**

Downloads and reads the 20-band environmental raster associated with the [andalusia](#) dataset from the [spatialDataExtra](#) repository. The raster covers Andalusia, Spain, at 400 m resolution (EPSG:25830) and includes remote-sensing, climate, and topographic predictors (see [andalusia](#)).

**Usage**

```
andalusia_extra()
```

**Value**

SpatRaster object with 20 layers.

**See Also**

Other andalusia: [andalusia](#), [andalusia\\_predictors](#), [andalusia\\_responses](#)

---

andalusia_predictors	<i>Predictor names for the dataset andalusia</i>
----------------------	--

---

**Description**

Character vector of 20 predictor variable names corresponding to the layers of the environmental raster returned by [andalusia\\_extra\(\)](#), covering Landsat reflectance (7), rainfall (2), solar radiation (2), temperature (4), and topography (5). These are **not** columns in [andalusia](#); use `terra::extract()` to attach them to the point data.

**Usage**

```
data(andalusia_predictors)
```

**Format**

Character vector of length 20.

**See Also**

Other andalusia: [andalusia](#), [andalusia\\_extra\(\)](#), [andalusia\\_responses](#)

**Examples**

```
data(andalusia_predictors)
andalusia_predictors
```

---

andalusia_responses	<i>Response names for the dataset andalusia</i>
---------------------	---

---

**Description**

Character vector of length 2 containing the names of the response variables in [andalusia](#): "species" (character, species name or "background" for 90 species plus background points) and "presence" (binary integer, 1 = confirmed species presence, 0 = background point).

**Usage**

```
data(andalusia_responses)
```

**Format**

Character vector of length 2.

**See Also**

Other andalusia: [andalusia](#), [andalusia\\_extra\(\)](#), [andalusia\\_predictors](#)

**Examples**

```
data(andalusia_responses)
andalusia_responses
```

---

`communities`*Plant Communities of Sierra Nevada (Spain)*

---

### Description

`sf` data frame with POINT geometry containing 6,747 plant community records from the Sierra Nevada mountain range (SE Spain), with 6 response variables (see [communities\\_responses](#)) and 9 numeric predictors (see [communities\\_predictors](#)). Use [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2050\(\)](#), and [communities\\_extra\\_2100\(\)](#) to download the associated environmental rasters for the baseline (2010), 2050, and 2100 climate scenarios.

### Usage

```
data(communities)
```

### Format

An `sf` data frame with 6,747 rows and 16 columns:

#### Response variables (6):

- `community`: Factor column with 6 levels: "none" (no presence of target communities), "Pyrenean oak forests", "Juniper-broom shrublands", "Pinus forests", "Alpine pastures", "Holm oak forests".
- `pyrenean_oak`: Binary integer presence-absence (1/0) for Pyrenean oak forests.
- `juniper_shrubland`: Binary integer presence-absence (1/0) for juniper-broom shrublands.
- `pinus_forest`: Binary integer presence-absence (1/0) for Pinus forests.
- `alpine_pastures`: Binary integer presence-absence (1/0) for alpine pastures.
- `holm_oak`: Binary integer presence-absence (1/0) for holm oak forests.

#### Predictor variables:

- `max_temperature_summer`: Maximum summer temperature (degrees C).
- `max_temperature_winter`: Maximum winter temperature (degrees C).
- `min_temperature_summer`: Minimum summer temperature (degrees C).
- `min_temperature_winter`: Minimum winter temperature (degrees C).
- `rainfall_summer`: Summer rainfall (mm).
- `rainfall_winter`: Winter rainfall (mm).
- `northness`: Northness index (cosine of aspect, -1 to 1).
- `slope`: Terrain slope (degrees).
- `topographic_wetness_index`: Topographic wetness index.

#### Geometry:

- `geometry`: Point geometry (ETRS89 / UTM zone 30N, EPSG:25830).

### Source

- Benito, B., Lorite, J., & Peñas, J. (2011). Simulating potential effects of climatic warming on altitudinal patterns of key species in Mediterranean-alpine ecosystems. *Climatic Change*, 108, 471–483. doi:10.1007/s1058401000153

### See Also

Other communities: [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2050\(\)](#), [communities\\_extra\\_2100\(\)](#), [communities\\_predictors](#), [communities\\_responses](#)

### Examples

```
data(communities)
colnames(communities)
nrow(communities)
ncol(communities)
```

---

`communities_extra_2010`

*Download Environmental Raster for communities - 2010*

---

### Description

Downloads the baseline (2010) environmental raster associated with the [communities](#) dataset from the [spatialDataExtra](#) repository. Writes the file `communities_2010.tif` in the working directory and returns it as a `spatRaster` object.

### Usage

```
communities_extra_2010()
```

### Value

`spatRaster` object.

### See Also

Other communities: [communities](#), [communities\\_extra\\_2050\(\)](#), [communities\\_extra\\_2100\(\)](#), [communities\\_predictors](#), [communities\\_responses](#)

---

`communities_extra_2050`*Download Environmental Raster for communities - 2050*

---

**Description**

Downloads the future climate (2050) raster associated with the [communities](#) dataset from the [spatialDataExtra](#) repository. Writes the file `communities_2050.tif` in the working directory and returns it as a `spatRaster` object.

**Usage**

```
communities_extra_2050()
```

**Value**

`SpatRaster` object.

**See Also**

Other communities: [communities](#), [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2100\(\)](#), [communities\\_predictors](#), [communities\\_responses](#)

---

`communities_extra_2100`*Download Environmental Raster for communities - 2100*

---

**Description**

Downloads the future climate (2100) raster associated with the [communities](#) dataset from the [spatialDataExtra](#) repository. Writes the file `communities_2100.tif` in the working directory and returns it as a `spatRaster` object.

**Usage**

```
communities_extra_2100()
```

**Value**

`SpatRaster` object.

**See Also**

Other communities: [communities](#), [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2050\(\)](#), [communities\\_predictors](#), [communities\\_responses](#)

---

`communities_predictors`*Predictor variable names for the dataset communities*

---

**Description**

Character vector of 9 predictor variable names from [communities](#).

**Usage**

```
data(communities_predictors)
```

**Format**

A character vector of length 9.

**See Also**

Other communities: [communities](#), [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2050\(\)](#), [communities\\_extra\\_2100\(\)](#), [communities\\_responses](#)

**Examples**

```
data(communities_predictors)
communities_predictors
```

---

`communities_responses` *Response variable names for the dataset communities*

---

**Description**

Character vector of 6 response variable names from [communities](#).

**Usage**

```
data(communities_responses)
```

**Format**

A character vector of length 6.

**See Also**

Other communities: [communities](#), [communities\\_extra\\_2010\(\)](#), [communities\\_extra\\_2050\(\)](#), [communities\\_extra\\_2100\(\)](#), [communities\\_predictors](#)

**Examples**

```
data(communities_responses)
communities_responses
```

---

 interaction

*Butterfly and host plant presence in Sierra Nevada (SE Spain)*


---

**Description**

sf dataframe with co-occurrence records for a butterfly and its host plant in Sierra Nevada (SE Spain). Contains 3 response variables (see [interaction\\_responses](#)) and 10 numeric predictors at 100 m resolution (see [interaction\\_predictors](#)). Use [interaction\\_extra\(\)](#) to download the associated environmental raster.

**Usage**

```
data(interaction)
```

**Format**

An sf dataframe with 1,000 rows (presence and background points) and 14 columns:

**Response variables (3):**

- `butterfly`: Integer with three possible values: 1 (presence of *Agriades zullichi*), 0 (background), and NA (host plant observation site, where butterfly was not surveyed).
- `host_plant`: Integer with three possible values: 1 (presence of *Androsace vitaliana*), 0 (background), and NA (butterfly observation site, where the plant was not surveyed).
- `class`: Factor with three levels: "butterfly", "host\_plant", and "background", indicating the observation type of each record.

**Predictor variables:**

- `landsat_ndvi`: Normalized Difference Vegetation Index derived from Landsat imagery.
- `landsat_pca_bands_123`: First principal component of Landsat bands 1, 2, and 3 (visible).
- `landsat_pca_bands_457`: First principal component of Landsat bands 4, 5, and 7 (infrared).
- `rainfall_annual`: Mean annual rainfall (mm).
- `solar_radiation`: Mean annual solar radiation ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ).
- `temperature_annual_mean`: Mean annual temperature (degrees C).
- `temperature_summer_max`: Maximum summer temperature (degrees C).
- `temperature_winter_min`: Minimum winter temperature (degrees C).
- `topographic_complexity`: Index of terrain ruggedness and heterogeneity.
- `topographic_position`: Relative elevation of a cell compared to its surroundings.

**Geometry:**

- `geometry`: Point geometry (ETRS89 / UTM zone 30N, EPSG:25830).

## Source

### Species occurrences:

- Barea-Azcón, J.M., Benito, B.M., Olivares, F.J., Ruiz, H., Martín, J., García, A.L., & López, R. (2014). Distribution and conservation of the relict interaction between the butterfly *Agriades zullichi* and its larval foodplant (*Androsace vitaliana nevadensis*). *Biodiversity and Conservation*, 23(4), 927–944. doi:10.1007/s1053101406434

### Remote sensing data:

- Nunes de Lima, M. V. (Ed.) (2005). *IMAGE2000 and CLC2000 – Products and methods*. Joint Research Centre, Institute for Environment and Sustainability, and European Environment Agency. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/84dd2bad-14d9-4a65-9b92-3b4507d09e44/language-en>

### Climate and topographic variables:

- Benito, B., Lorite, J., & Peñas, J. (2011). Simulating potential effects of climatic warming on altitudinal patterns of key species in Mediterranean-alpine ecosystems. *Climatic Change*, 108, 471–483. doi:10.1007/s1058401000153

## See Also

Other interaction: [interaction\\_extra\(\)](#), [interaction\\_predictors](#), [interaction\\_responses](#)

## Examples

```
data(interaction)
colnames(interaction)
nrow(interaction)
ncol(interaction)
```

---

interaction\_extra      *Download Environmental Raster for interaction*

---

## Description

Downloads and reads the environmental raster associated with the [interaction](#) dataset from the [spatialDataExtra](#) repository. Requires the **terra** package. Writes the file `sierra_nevada_env.tif` to the working directory and returns a `spatRaster` object.

## Usage

```
interaction_extra()
```

## Value

`SpatRaster` object.

**See Also**

Other interaction: [interaction\(\)](#), [interaction\\_predictors](#), [interaction\\_responses](#)

---

interaction\_predictors

*Predictor variable names for interaction dataset*

---

**Description**

Character vector of 10 predictor variable names from [interaction](#).

**Usage**

```
data(interaction_predictors)
```

**Format**

A character vector of length 10.

**See Also**

Other interaction: [interaction\(\)](#), [interaction\\_extra\(\)](#), [interaction\\_responses](#)

**Examples**

```
data(interaction_predictors)
interaction_predictors
```

---

interaction\_responses *Response variable names for the dataset interaction*

---

**Description**

Character vector of 3 response variable names from [interaction](#).

**Usage**

```
data(interaction_responses)
```

**Format**

A character vector of length 3.

**See Also**

Other interaction: [interaction\(\)](#), [interaction\\_extra\(\)](#), [interaction\\_predictors](#)

## Examples

```
data(interaction_responses)
interaction_responses
```

---

linaria	<i>Presence of Linaria nigricans and greenhouses in Eastern Andalusia</i>
---------	---

---

## Description

sf data frame with POINT geometry containing presence records of the plant *Linaria nigricans*, greenhouses, and background points from Eastern Andalusia (Spain). The dataframe contains 2 response variables (see [linaria\\_responses](#)), and 20 numeric predictors (see [linaria\\_predictors](#)). Use [linaria\\_extra\(\)](#) to download the associated environmental raster.

The dataset combines species presence records, greenhouse presence records (representing a competing land use), and randomly sampled background points. Species presences and greenhouse presences were spatially thinned at 400 m to remove redundancy at the raster resolution. Background points were randomly sampled within the extent of the presence records. Environmental predictors were extracted from a Landsat/DEM-derived raster at 400 m resolution (EPSG:25830).

## Usage

```
data(linaria)
```

## Format

An sf data frame with 7386 rows (presences and background points) and 25 columns:

### Response variables:

- `linaria_nigricans`: Binary integer (1 = confirmed *Linaria nigricans* presence, 0 = greenhouse presence or background point).
- `greenhouses`: Binary integer (1 = greenhouse presence, 0 = *Linaria nigricans* presence or background point).

### Predictor variables:

- `landsat_band_1`: Landsat TM Band 1 — Blue (0.45–0.52  $\mu\text{m}$ ), surface reflectance.
- `landsat_band_2`: Landsat TM Band 2 — Green (0.52–0.60  $\mu\text{m}$ ), surface reflectance.
- `landsat_band_3`: Landsat TM Band 3 — Red (0.63–0.69  $\mu\text{m}$ ), surface reflectance.
- `landsat_band_4`: Landsat TM Band 4 — Near-infrared (0.76–0.90  $\mu\text{m}$ ), surface reflectance.
- `landsat_band_5`: Landsat TM Band 5 — Short-wave infrared 1 (1.55–1.75  $\mu\text{m}$ ), surface reflectance.
- `landsat_band_6`: Landsat TM Band 6 — Thermal infrared (10.4–12.5  $\mu\text{m}$ ), brightness temperature (K).
- `landsat_ndvi`: Normalized Difference Vegetation Index derived from Landsat bands 3 and 4.

- rainfall\_annual: Total annual rainfall (mm).
- rainfall\_summer: Total summer rainfall (mm, June–September).
- solar\_radiation\_summer: Mean daily solar radiation in summer (kJ m<sup>-2</sup> day<sup>-1</sup>).
- solar\_radiation\_winter: Mean daily solar radiation in winter (kJ m<sup>-2</sup> day<sup>-1</sup>).
- temperature\_summer\_max: Mean maximum temperature in summer (degrees C).
- temperature\_summer\_min: Mean minimum temperature in summer (degrees C).
- temperature\_winter\_max: Mean maximum temperature in winter (degrees C).
- temperature\_winter\_min: Mean minimum temperature in winter (degrees C).
- topography\_eastness: Eastward component of aspect (sin of aspect in radians).
- topography\_elevation: Elevation above sea level (m).
- topography\_northness: Northward component of aspect (cos of aspect in radians).
- topography\_position: Topographic position index (local elevation relative to neighbourhood mean).
- topography\_slope: Slope gradient (degrees).

#### Geometry:

- geometry: Point geometry (ETRS89 / UTM zone 30N, EPSG:25830).

#### Source

##### Published studies:

- Benito, B.M., Martínez-Ortega, M.M., Munoz, L.M., Lorite, J. & Penas, J. (2009). Assessing extinction-risk of endangered plants using species distribution models: a case study of habitat depletion caused by the spread of greenhouses. *Biodiversity and Conservation*, 18(9), 2509–2520. doi:10.1007/s1053100996048
- Peñas, J., Benito, B., Lorite, J., et al. (2011). Habitat fragmentation in arid zones: a case study of *Linaria nigricans* under land use changes (SE Spain). *Environmental Management*, 48, 168–176. doi:10.1007/s002670119663y

##### Landsat imagery:

- Nunes de Lima, M. V. (Ed.) (2005). *IMAGE2000 and CLC2000 – Products and methods*. Joint Research Centre, Institute for Environment and Sustainability, and European Environment Agency. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/84dd2bad-14d9-4a65-9b92-3b4507d09e44/language-en>

##### Climate variables:

- Ninyerola, M., Pons, X. & Roure, J.M. (2005). Atlas Climático Digital de la Península Ibérica: Metodología y aplicaciones en bioclimatología y geobotánica. Universidad Autónoma de Barcelona, Bellaterra.

##### Topography:

- Instituto Geográfico Nacional. Modelo Digital del Terreno (MDT25). <https://www.ign.es>

**See Also**

Other linaria: [linaria\\_extra\(\)](#), [linaria\\_predictors](#), [linaria\\_responses](#)

**Examples**

```
data(linaria)
colnames(linaria)
nrow(linaria)
ncol(linaria)
```

---

linaria_extra	<i>Download Environmental Raster for linaria</i>
---------------	--

---

**Description**

Downloads and reads the 20-band environmental raster associated with the [linaria](#) dataset from the [spatialDataExtra](#) repository. Writes the file `linaria_env.tif` in the working directory and returns it as a `spatRaster` object.

**Usage**

```
linaria_extra()
```

**Value**

`SpatRaster` object with 20 layers.

**See Also**

Other linaria: [linaria](#), [linaria\\_predictors](#), [linaria\\_responses](#)

---

linaria_predictors	<i>Predictor variable names for the dataset linaria</i>
--------------------	---

---

**Description**

Character vector of 20 predictor variable names from [linaria](#), covering Landsat reflectance (7), rainfall (2), solar radiation (2), temperature (4), and topography (5).

**Usage**

```
data(linaria_predictors)
```

**Format**

A character vector of length 20.

**See Also**

Other linaria: [linaria](#), [linaria\\_extra\(\)](#), [linaria\\_responses](#)

**Examples**

```
data(linaria_predictors)
linaria_predictors
```

---

linaria_responses	<i>Response variable names for the dataset linaria</i>
-------------------	--

---

**Description**

Character vector of length 2 containing the names of the response variables in [linaria](#).

**Usage**

```
data(linaria_responses)
```

**Format**

A character vector of length 2.

**See Also**

Other linaria: [linaria](#), [linaria\\_extra\(\)](#), [linaria\\_predictors](#)

**Examples**

```
data(linaria_responses)
linaria_responses
```

---

neanderthal	<i>Neanderthal presence in the Last Interglacial</i>
-------------	--

---

**Description**

sf data frame with POINT geometry containing 227 records of Neanderthal presence from Marine Isotope Stage 5e (Last Interglacial) in Europe and the Near East, 1 response variable (see [neanderthal\\_response](#)), and 25 predictors (see [neanderthal\\_predictors](#)). Use [neanderthal\\_extra\(\)](#) to download the associated environmental raster.

**Usage**

```
data(neanderthal)
```

**Format**

An sf data frame with 227 rows (presence and pseudo-absence sites) and 27 columns:

**Response variable (1):**

- presence: Binary integer (1 = Neanderthal presence site, 0 = pseudo-absence site).

**Predictor variables:**

Bioclimatic variables derived from a Last Interglacial GCM simulation (Otto-Bliesner et al. 2006), downscaled following the method of Hijmans et al. (2005). These are analogous to the standard WorldClim bioclimatic variables but represent Last Interglacial (MIS 5e) conditions rather than modern climate:

- bio1: Annual mean temperature (degrees C).
- bio2: Mean diurnal range (degrees C).
- bio3: Isothermality ( $\text{bio2}/\text{bio7} * 100$ ).
- bio4: Temperature seasonality (standard deviation \* 100).
- bio5: Max temperature of warmest month (degrees C).
- bio6: Min temperature of coldest month (degrees C).
- bio7: Temperature annual range (degrees C).
- bio8: Mean temperature of wettest quarter (degrees C).
- bio9: Mean temperature of driest quarter (degrees C).
- bio10: Mean temperature of warmest quarter (degrees C).
- bio11: Mean temperature of coldest quarter (degrees C).
- bio12: Annual precipitation (mm).
- bio13: Precipitation of wettest month (mm).
- bio14: Precipitation of driest month (mm).
- bio15: Precipitation seasonality (coefficient of variation).
- bio16: Precipitation of wettest quarter (mm).
- bio17: Precipitation of driest quarter (mm).
- bio18: Precipitation of warmest quarter (mm).
- bio19: Precipitation of coldest quarter (mm).
- topo\_aspect: Aspect in degrees.
- topo\_diversity\_local: Local topographic diversity.
- topo\_diversity: Regional topographic diversity.
- topo\_elev: Elevation in meters.
- topo\_slope: Slope in degrees.
- topo\_wetness: Topographic wetness index.

**Geometry:**

- geometry: Point geometry (WGS84, EPSG:4326).

**Source****Presence data:**

- Benito, B.M., et al. (2017). The ecological niche and distribution of Neanderthals during the Last Interglacial. *Journal of Biogeography*, 44, 51-61. doi:10.1111/jbi.12845
- Nielsen, T.K., Benito, B.M., Svenning, J.-C., Sandel, B., McKerracher, L., Riede, F., & Kjær-gaard, P.C. (2017). Investigating Neanderthal dispersal above 55°N in Europe during the Last Interglacial Complex. *Quaternary International*, 431, 88-103. doi:10.1016/j.quaint.2015.10.039

**Palaeoclimatic variables (GCM simulation):**

- Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H. & Hu, A. (2006). Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, 311, 1751-1753.

**Palaeoclimatic variables (interpolation):**

- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

**Elevation and topography:**

- Jarvis, A., Guevara, E., Reuter, H. I., & Nelson, A. D. (2008). Hole-filled SRTM for the globe: version 4, data grid. Web publication/site, CGIAR Consortium for Spatial Information. <https://srtm.csi.cgiar.org>

**See Also**

Other neanderthal: [neanderthal\\_extra\(\)](#), [neanderthal\\_predictors](#), [neanderthal\\_response](#)

**Examples**

```
data(neanderthal)
colnames(neanderthal)
nrow(neanderthal)
ncol(neanderthal)
```

---

neanderthal\_extra

*Download Environmental Raster for neanderthal*

---

**Description**

Downloads and reads the environmental raster associated with the [neanderthal](#) dataset from the [spatialDataExtra](#) repository. Writes the file `neanderthal_env.tif` to the working directory and returns it as a `spatRaster` object.

**Usage**

```
neanderthal_extra()
```

**Value**

SpatRaster object.

**See Also**

Other neanderthal: [neanderthal](#), [neanderthal\\_predictors](#), [neanderthal\\_response](#)

---

neanderthal\_predictors

*Predictor variable names for the dataset neanderthal*

---

**Description**

Character vector of 25 predictor variable names from [neanderthal](#).

**Usage**

```
data(neanderthal_predictors)
```

**Format**

A character vector of length 25.

**See Also**

Other neanderthal: [neanderthal](#), [neanderthal\\_extra\(\)](#), [neanderthal\\_response](#)

**Examples**

```
data(neanderthal_predictors)
neanderthal_predictors
```

---

neanderthal\_response *Response variable name for the dataset neanderthal*

---

**Description**

Character string with the name of the response variable in [neanderthal](#).

**Usage**

```
data(neanderthal_response)
```

**Format**

A character string of length 1.

**See Also**

Other neanderthal: [neanderthal](#), [neanderthal\\_extra\(\)](#), [neanderthal\\_predictors](#)

**Examples**

```
data(neanderthal_response)
neanderthal_response
```

---

plantae

*Plant diversity metrics for the World's Ecoregions*

---

**Description**

Plant diversity metrics (richness, rarity, beta diversity) obtained from GBIF Plantae records for the World's Ecoregions. Includes metrics for all plants, trees, and grasses at species, genus, and family taxonomic levels. Ecoregion boundaries are derived from [Ecoregions 2017](#). Original polygon geometries have been converted to point centroids to reduce file size while preserving spatial context. Use [plantae\\_extra\(\)](#) to download a full version with original polygon geometries. The datasets [plantae\\_west](#) and [plantae\\_east](#) are subsets of [plantae](#) focused on overall plant richness of the western and eastern hemispheres, respectively.

The GBIF download comprised **244,830,168 records from 4,741 datasets**, filtered to records with coordinates, no geospatial issues, and occurrence status "present".

**Tree species** were identified by cross-referencing GBIF records with the BGCI Global Tree Search database (BGCI 2020). **Grasses** were defined as members of family Poaceae.

**Rarity-weighted richness** was computed for each taxon as the inverse of its number of spatial presence records in GBIF, then scores are summed per ecoregion, while **mean rarity** is the mean of these inverse presence record counts per taxon in an ecoregion.

**Beta diversity** was computed between each ecoregion and its immediate neighboring ecoregions via Sorensen dissimilarity ( $B_{sor} = 1 - 2a / (2a + b + c)$ ) and Simpson dissimilarity ( $B_{sim} = \min(b, c) / (\min(b, c) + a)$ ), following Koleff et al. (2003).

**Fragmentation metrics** were computed with the R package `landscapemetrics` (Hesselbarth et al. 2019) at 5 km resolution in Lambert Azimuthal Equal-Area projection.

**Climate hypervolume** was computed using `hypervolume::hypervolume_svm()` from the climate predictors.

**Aridity** is computed as 1 minus the aridity index of Trabucco and Zomer (2019), so maximum aridity is coded as 1.

Environmental predictors were extracted as mean pixel values per ecoregion from rasters at 1 km resolution.

**Usage**

```
data(plantae)
```

**Format**

An sf data frame with 662 rows (ecoregions) and 143 columns:

**Identifier columns:**

- `ecoregion_id`: Unique ecoregion identifier.
- `ecoregion_name`: Ecoregion name.
- `ecoregion_biome`: Biome classification.
- `ecoregion_realm`: Biogeographic realm.
- `ecoregion_continent`: Continent name.

**Response variables - Richness (9):**

- `richness_species`: Number of plant species.
- `richness_genera`: Number of plant genera.
- `richness_families`: Number of plant families.
- `richness_classes`: Number of plant classes.
- `richness_species_trees`: Number of tree species.
- `richness_genera_trees`: Number of tree genera.
- `richness_families_trees`: Number of tree families.
- `richness_species_grasses`: Number of grass species.
- `richness_genera_grasses`: Number of grass genera.

**Response variables - Rarity-weighted richness (6):**

- `rarity_weighted_richness_species`: Rarity-weighted richness for species (sum of inverse spatial presence record counts per taxon; Williams et al. 1996).
- `rarity_weighted_richness_genera`: Rarity-weighted richness for genera (sum of inverse spatial presence record counts per taxon).
- `rarity_weighted_richness_species_trees`: Rarity-weighted richness for tree species (sum of inverse spatial presence record counts per taxon).
- `rarity_weighted_richness_genera_trees`: Rarity-weighted richness for tree genera (sum of inverse spatial presence record counts per taxon).
- `rarity_weighted_richness_species_grasses`: Rarity-weighted richness for grass species (sum of inverse spatial presence record counts per taxon).
- `rarity_weighted_richness_genera_grasses`: Rarity-weighted richness for grass genera (sum of inverse spatial presence record counts per taxon).

**Response variables - Mean rarity (6):**

- `mean_rarity_species`: Mean rarity index for species (mean of inverse spatial presence record counts per taxon).
- `mean_rarity_genera`: Mean rarity index for genera (mean of inverse spatial presence record counts per taxon).
- `mean_rarity_species_trees`: Mean rarity index for tree species (mean of inverse spatial presence record counts per taxon).

- `mean_rarity_genera_trees`: Mean rarity index for tree genera (mean of inverse spatial presence record counts per taxon).
- `mean_rarity_species_grasses`: Mean rarity index for grass species (mean of inverse spatial presence record counts per taxon).
- `mean_rarity_genera_grasses`: Mean rarity index for grass genera (mean of inverse spatial presence record counts per taxon).

**Response variables - Beta diversity R (absolute richness difference) (16):**

- `betadiversity_R_species`: Absolute richness difference between ecoregion and neighbors for species.
- `betadiversity_R_percent_species`: Absolute richness difference as percentage for species.
- `betadiversity_R_genera`: Absolute richness difference between ecoregion and neighbors for genera.
- `betadiversity_R_percent_genera`: Absolute richness difference as percentage for genera.
- `betadiversity_R_families`: Absolute richness difference between ecoregion and neighbors for families.
- `betadiversity_R_percent_families`: Absolute richness difference as percentage for families.
- `betadiversity_R_species_trees`: Absolute richness difference between ecoregion and neighbors for tree species.
- `betadiversity_R_percent_species_trees`: Absolute richness difference as percentage for tree species.
- `betadiversity_R_genera_trees`: Absolute richness difference between ecoregion and neighbors for tree genera.
- `betadiversity_R_percent_genera_trees`: Absolute richness difference as percentage for tree genera.
- `betadiversity_R_families_trees`: Absolute richness difference between ecoregion and neighbors for tree families.
- `betadiversity_R_percent_families_trees`: Absolute richness difference as percentage for tree families.
- `betadiversity_R_species_grasses`: Absolute richness difference between ecoregion and neighbors for grass species.
- `betadiversity_R_percent_species_grasses`: Absolute richness difference as percentage for grass species.
- `betadiversity_R_genera_grasses`: Absolute richness difference between ecoregion and neighbors for grass genera.
- `betadiversity_R_percent_genera_grasses`: Absolute richness difference as percentage for grass genera.

**Response variables - Beta diversity Sorensen (8):**

- `betadiversity_sorensen_species`: Sorensen dissimilarity for species ( $B_{sor} = 1 - 2a/(2a+b+c)$ ; Koleff et al. 2003).

- `betadiversity_sorensen_genera`: Sorensen dissimilarity for genera ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_families`: Sorensen dissimilarity for families ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_species_trees`: Sorensen dissimilarity for tree species ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_genera_trees`: Sorensen dissimilarity for tree genera ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_families_trees`: Sorensen dissimilarity for tree families ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_species_grasses`: Sorensen dissimilarity for grass species ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).
- `betadiversity_sorensen_genera_grasses`: Sorensen dissimilarity for grass genera ( $B_{sor} = 1 - 2a/(2a+b+c)$ ).

#### Response variables - Beta diversity Simpson (8):

- `betadiversity_simpson_species`: Simpson dissimilarity for species ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ; Koleff et al. 2003).
- `betadiversity_simpson_genera`: Simpson dissimilarity for genera ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_families`: Simpson dissimilarity for families ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_species_trees`: Simpson dissimilarity for tree species ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_genera_trees`: Simpson dissimilarity for tree genera ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_families_trees`: Simpson dissimilarity for tree families ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_species_grasses`: Simpson dissimilarity for grass species ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).
- `betadiversity_simpson_genera_grasses`: Simpson dissimilarity for grass genera ( $B_{sim} = \min(b,c)/(\min(b,c)+a)$ ).

#### Predictor variables:

- `bias_log_records`: Logarithm of the total GBIF records in ecoregion.
- `geo_neighbors_count`: Number of neighboring ecoregions.
- `geo_neighbors_area_km2`: Total area of neighboring ecoregions in square kilometers.
- `geo_neighbors_aridity_mean`: Mean aridity of neighboring ecoregions.
- `geo_area_km2`: Ecoregion area in square kilometers.
- `geo_polygons_count`: Number of polygons in multipolygon geometry.
- `geo_perimeter_km`: Ecoregion perimeter in kilometers.
- `geo_shared_perimeter_km`: Shared perimeter with neighbors in kilometers.
- `geo_shared_perimeter_fraction`: Fraction of perimeter shared with neighbors.
- `geo_distance_to_ocean`: Distance to nearest ocean in kilometers.
- `geo_elevation_mean`: Mean elevation in meters.

- human\_population: Total human population in ecoregion.
- human\_population\_density: Human population density per square kilometer.
- human\_footprint\_mean: Mean human footprint index.
- climate\_velocity\_lgm\_mean: Mean climate velocity since Last Glacial Maximum.
- climate\_hypervolume: Climate hypervolume (niche space size), computed with `hypervolume::hypervolume_svm()`.
- air\_humidity\_max: Maximum near-surface relative humidity (%).
- air\_humidity\_mean: Mean near-surface relative humidity (%).
- air\_humidity\_min: Minimum near-surface relative humidity (%).
- air\_humidity\_range: Near-surface relative humidity range (%).
- aridity\_mean: Mean aridity (1 minus aridity index; higher values indicate greater aridity).
- cloud\_cover\_max: Maximum cloud cover (%).
- cloud\_cover\_mean: Mean cloud cover (%).
- cloud\_cover\_min: Minimum cloud cover (%).
- cloud\_cover\_range: Cloud cover range (%).
- evapotranspiration\_max: Maximum potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_mean: Mean potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_min: Minimum potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_range: Potential evapotranspiration range (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- precipitation\_seasonality: Precipitation seasonality (coefficient of variation of monthly estimates; CHELSA bio15).
- precipitation\_total: Total annual precipitation (kg m<sup>-2</sup> year<sup>-1</sup>; CHELSA bio12).
- precipitation\_coldest\_quarter: Precipitation of coldest quarter (kg m<sup>-2</sup>; CHELSA bio19).
- precipitation\_driest\_month: Precipitation of driest month (kg m<sup>-2</sup>; CHELSA bio14).
- precipitation\_driest\_quarter: Precipitation of driest quarter (kg m<sup>-2</sup>; CHELSA bio17).
- precipitation\_warmest\_quarter: Precipitation of warmest quarter (kg m<sup>-2</sup>; CHELSA bio18).
- precipitation\_wettest\_month: Precipitation of wettest month (kg m<sup>-2</sup>; CHELSA bio13).
- precipitation\_wettest\_quarter: Precipitation of wettest quarter (kg m<sup>-2</sup>; CHELSA bio16).
- temperature\_isothermality: Isothermality: ratio of diurnal to annual temperature variation (degrees C; CHELSA bio3).
- temperature\_mean\_daily\_range: Mean diurnal temperature range (degrees C; CHELSA bio2).
- temperature\_mean: Mean annual temperature (degrees C; CHELSA bio1).
- temperature\_range: Annual temperature range (degrees C; CHELSA bio7).
- temperature\_seasonality: Temperature seasonality as standard deviation of monthly means (degrees C; CHELSA bio4).

- temperature\_coldest\_month: Minimum temperature of coldest month (degrees C; CHELSA bio6).
- temperature\_coldest\_quarter: Mean temperature of coldest quarter (degrees C; CHELSA bio11).
- temperature\_driest\_quarter: Mean temperature of driest quarter (degrees C; CHELSA bio9).
- temperature\_warmest\_month: Maximum temperature of warmest month (degrees C; CHELSA bio5).
- temperature\_warmest\_quarter: Mean temperature of warmest quarter (degrees C; CHELSA bio10).
- temperature\_wettest\_quarter: Mean temperature of wettest quarter (degrees C; CHELSA bio8).
- landcover\_bare\_percent\_mean: Mean percentage of bare ground.
- landcover\_herbs\_percent\_mean: Mean percentage of herbaceous vegetation.
- landcover\_trees\_percent\_mean: Mean percentage of tree cover.
- fragmentation\_ai: Aggregation index.
- fragmentation\_area\_mn: Mean patch area.
- fragmentation\_ca: Total class area.
- fragmentation\_clumpy: Clumpiness index.
- fragmentation\_cohesion: Patch cohesion index.
- fragmentation\_contig\_mn: Mean contiguity index.
- fragmentation\_core\_mn: Mean core area.
- fragmentation\_cpland: Core area percentage of landscape.
- fragmentation\_dcore\_mn: Mean disjunct core area.
- fragmentation\_division: Landscape division index.
- fragmentation\_ed: Edge density.
- fragmentation\_lsi: Landscape shape index.
- fragmentation\_mesh: Effective mesh size.
- fragmentation\_ndca: Number of disjunct core areas.
- fragmentation\_nlsi: Normalized landscape shape index.
- fragmentation\_np: Number of patches.
- fragmentation\_shape\_mn: Mean shape index.
- fragmentation\_tca: Total core area.
- fragmentation\_te: Total edge.
- soil\_clay: Soil clay content (%).
- soil\_nitrogen: Soil nitrogen content (%).
- soil\_organic\_carbon: Soil organic carbon content (%).
- soil\_ph: Soil pH.
- soil\_sand: Soil sand content (%).

- soil\_silt: Soil silt content (%).
- soil\_temperature\_max: Maximum soil temperature (degrees C).
- soil\_temperature\_mean: Mean soil temperature (degrees C).
- soil\_temperature\_min: Minimum soil temperature (degrees C).
- soil\_temperature\_range: Soil temperature range (degrees C).
- ndvi\_max: Maximum NDVI (1999-2019).
- ndvi\_mean: Mean NDVI (1999-2019).
- ndvi\_min: Minimum NDVI (1999-2019).
- ndvi\_range: NDVI range (1999-2019).

#### Geometry:

- geometry: Ecoregion centroids, POINT geometry (WGS84, EPSG:4326).

#### Source

##### Associated publications:

- Maestre, F.T., Benito, B.M., Berdugo, M., Concostrina-Zubiri, L., Delgado-Baquerizo, M., Eldridge, D.J., Guirado, E., Gross, N., Kefi, S., Le Bagousse-Pinguet, Y., et al. (2021). Biogeography of global drylands. *New Phytologist*, 231(2), 540–558. doi:10.1111/nph.17398
- GBIF Plantae Dataset (September 15, 2020). doi:10.15468/dl.xh5y5g
- Dinerstein, E., et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), 534-545. doi:10.1093/biosci/bix014
- Karger, D.N., et al. (2021). Climatologies at high resolution for the earth's land surface areas. *EnviDat*. doi:10.16904/enviDat.228.v2.1
- Hengl, T., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. doi:10.1371/journal.pone.0169748
- Lembrechts, J.J., et al. (2021). Mismatches between soil and air temperature. *Global Change Biology*. doi:10.1111/gcb.16060
- Copernicus Global Land Service: NDVI Long Term Statistics v3 (1999-2019). <https://land.copernicus.eu/en/products/vegetation>
- Buchhorn, M., et al. (2020). Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe. Zenodo. doi:10.5281/zenodo.3939050
- CGIAR-CSI SRTM 90m Digital Elevation Database. <https://srtm.csi.cgiar.org/>
- Trabucco, A. & Zomer, R.J. (2019). Global Aridity Index and Potential Evapotranspiration Climate Database v2. CGIAR-CSI. doi:10.6084/m9.figshare.7504448.v3
- BGCI (2020). GlobalTreeSearch online database. Botanic Gardens Conservation International. [https://tools.bgci.org/global\\_tree\\_search.php](https://tools.bgci.org/global_tree_search.php)
- Hesselbarth, M.H.K., et al. (2019). landscapemetrics: an open-source R tool to calculate landscape metrics. *Ecography*, 42(10), 1648-1657. doi:10.1111/ecog.04617
- Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, 72(3), 367-382. doi:10.1046/j.13652656.2003.00710.x

- Williams, P.H., et al. (1996). A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conservation Biology*, 10(1), 155-174.
- Venter, O., et al. (2016). Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data*, 3, 160067. doi:10.1038/sdata.2016.67

### See Also

Other plantae: [plantae\\_east](#), [plantae\\_extra\(\)](#), [plantae\\_predictors](#), [plantae\\_responses](#), [plantae\\_west](#)

### Examples

```
data(plantae)
colnames(plantae)
nrow(plantae)
ncol(plantae)
```

---

plantae_east	<i>Eastern Hemisphere subset of plantae</i>
--------------	---

---

### Description

Subset of the [plantae](#) dataset filtered to non-American ecoregions (`ecoregion_continent != "Americas"`) with `richness_species` (overall plant species richness) as the only response variable. All 84 predictor variables and identifier columns in [plantae](#) are retained.

### Usage

```
data(plantae_east)
```

### Format

An sf data frame with 434 rows and 91 columns.

### See Also

Other plantae: [plantae](#), [plantae\\_extra\(\)](#), [plantae\\_predictors](#), [plantae\\_responses](#), [plantae\\_west](#)

### Examples

```
data(plantae_east)
colnames(plantae_east)
nrow(plantae_east)
ncol(plantae_east)
```

---

plantae\_extra      *Download Extended plantae Dataset*

---

### Description

Downloads and reads the extended version of the [plantae](#) dataset with original polygon geometries instead of point centroids, from the [spatialDataExtra](#) repository. Writes the file `plantae.gpkg` to the working directory and returns it as an `sf` dataframe. See [plantae](#) for details on the response variables, predictors, and data sources.

### Usage

```
plantae_extra()
```

### Value

`sf` dataframe with 662 rows and 143 columns (MULTIPOLYGON geometry, WGS84).

### See Also

Other `plantae`: [plantae](#), [plantae\\_east](#), [plantae\\_predictors](#), [plantae\\_responses](#), [plantae\\_west](#)

---

plantae\_predictors      *Predictor variable names for the dataset plantae*

---

### Description

Character vector of 84 predictor variable names from `plantae`.

### Usage

```
data(plantae_predictors)
```

### Format

A character vector of length 84.

### See Also

Other `plantae`: [plantae](#), [plantae\\_east](#), [plantae\\_extra\(\)](#), [plantae\\_responses](#), [plantae\\_west](#)

### Examples

```
data(plantae_predictors)
plantae_predictors
```

---

plantae_responses	<i>Response variable names for the dataset plantae</i>
-------------------	--

---

**Description**

Character vector containing the names of the 53 response variables in plantae.

**Usage**

```
data(plantae_responses)
```

**Format**

A character vector of length 53.

**See Also**

Other plantae: [plantae](#), [plantae\\_east](#), [plantae\\_extra\(\)](#), [plantae\\_predictors](#), [plantae\\_west](#)

**Examples**

```
data(plantae_responses)
plantae_responses
```

---

plantae_west	<i>Western Hemisphere subset of plantae</i>
--------------	---

---

**Description**

Subset of the [plantae](#) dataset filtered to American ecoregions (`ecoregion_continent == "Americas"`) with `richness_species` (overall plant species richness) as the only response variable. All 84 predictor variables and identifier columns in [plantae](#) are retained.

**Usage**

```
data(plantae_west)
```

**Format**

An sf data frame with 228 rows and 91 columns.

**See Also**

Other plantae: [plantae](#), [plantae\\_east](#), [plantae\\_extra\(\)](#), [plantae\\_predictors](#), [plantae\\_responses](#)

**Examples**

```
data(plantae_west)
colnames(plantae_west)
nrow(plantae_west)
ncol(plantae_west)
```

---

quercus	<i>European Quercus (Oak) Species Distribution with Environmental Predictors</i>
---------	--

---

**Description**

sf data frame with POINT geometry containing 6,728 records of eight European Quercus (oak) species and absence points, 1 response variable (see [quercus\\_response](#)), and 31 numeric predictors (see [quercus\\_predictors](#)). Use [quercus\\_extra\(\)](#) to download the associated environmental raster.

**Usage**

```
data(quercus)
```

**Format**

An sf data frame with 6728 rows (species occurrences and absences) and 33 columns:

**Response variable:**

- **species:** Character column with 9 levels: "absence" (background absence points), "Quercus robur" (English oak), "Quercus petraea" (Sessile oak), "Quercus ilex" (Holm oak), "Quercus cerris" (Turkey oak), "Quercus faginea" (Portuguese oak), "Quercus pubescens" (Downy oak), "Quercus pyrenaica" (Pyrenean oak), "Quercus suber" (Cork oak).

**Predictor variables:**

WorldClim v2 bioclimatic variables (excludes bio8 and bio9):

- bio1: Annual mean temperature (degrees C).
- bio2: Mean diurnal range (degrees C).
- bio3: Isothermality (bio2/bio7 \* 100).
- bio4: Temperature seasonality (standard deviation \* 100).
- bio5: Max temperature of warmest month (degrees C).
- bio6: Min temperature of coldest month (degrees C).
- bio7: Temperature annual range (degrees C).
- bio10: Mean temperature of warmest quarter (degrees C).
- bio11: Mean temperature of coldest quarter (degrees C).
- bio12: Annual precipitation (mm).
- bio13: Precipitation of wettest month (mm).

- bio14: Precipitation of driest month (mm).
- bio15: Precipitation seasonality (coefficient of variation).
- bio16: Precipitation of wettest quarter (mm).
- bio17: Precipitation of driest quarter (mm).
- bio18: Precipitation of warmest quarter (mm).
- bio19: Precipitation of coldest quarter (mm).
- ndvi\_average: Average NDVI.
- ndvi\_maximum: Maximum NDVI.
- ndvi\_minimum: Minimum NDVI.
- ndvi\_range: NDVI range.
- sun\_rad\_average: Average solar radiation (kJ m<sup>-2</sup> day<sup>-1</sup>).
- sun\_rad\_maximum: Maximum solar radiation (kJ m<sup>-2</sup> day<sup>-1</sup>).
- sun\_rad\_minimum: Minimum solar radiation (kJ m<sup>-2</sup> day<sup>-1</sup>).
- sun\_rad\_range: Solar radiation range (kJ m<sup>-2</sup> day<sup>-1</sup>).
- landcover\_veg\_bare: Percentage of bare ground.
- landcover\_veg\_herb: Percentage of herbaceous vegetation.
- landcover\_veg\_tree: Percentage of tree cover.
- topographic\_diversity: Number of unique combinations of elevation, slope, and aspect classes within a neighborhood.
- topo\_slope: Topographic slope in degrees.
- human\_footprint: Human footprint index.

#### **Geometry:**

- geometry: Point geometry (WGS84, EPSG:4326).

#### **Source**

##### **Species occurrences:**

- GBIF.org. Global Biodiversity Information Facility. <https://www.gbif.org/>

##### **Bioclimatic variables and solar radiation:**

- Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302-4315. doi:10.1002/joc.5086

##### **NDVI:**

- Didan, K. (2015). MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006. NASA EOSDIS LP DAAC. doi:10.5067/MODIS/MOD13A2.006

##### **Land cover:**

- DiMiceli, C., et al. (2015). MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006. NASA EOSDIS LP DAAC. doi:[10.5067/MODIS/MOD44B.006](https://doi.org/10.5067/MODIS/MOD44B.006)

#### Elevation and topography:

- Jarvis, A., Guevara, E., Reuter, H. I., & Nelson, A. D. (2008). Hole-filled SRTM for the globe: version 4, data grid. Web publication/site, CGIAR Consortium for Spatial Information. <https://srtm.csi.cgiar.org>

#### Human footprint:

- Venter, O., et al. (2016). Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data*, 3, 160067. doi:[10.1038/sdata.2016.67](https://doi.org/10.1038/sdata.2016.67)

#### See Also

Other quercus: [quercus\\_extra\(\)](#), [quercus\\_predictors](#), [quercus\\_response](#)

#### Examples

```
data(quercus)
colnames(quercus)
nrow(quercus)
ncol(quercus)
```

---

quercus\_extra

*Download Environmental Raster for quercus*

---

#### Description

Downloads and reads the environmental raster associated with the [quercus](#) dataset from the [spatial-DataExtra](#) repository. Writes the file `quercus_env.tif` to the working directory, and returns it as a `spatRaster` object.

#### Usage

```
quercus_extra()
```

#### Value

`SpatRaster` object.

#### See Also

Other quercus: [quercus](#), [quercus\\_predictors](#), [quercus\\_response](#)

---

quercus_predictors	<i>Predictor variable names for for the dataset quercus</i>
--------------------	---

---

**Description**

Character vector of 31 predictor variable names from [quercus](#).

**Usage**

```
data(quercus_predictors)
```

**Format**

A character vector of length 31.

**See Also**

Other quercus: [quercus](#), [quercus\\_extra\(\)](#), [quercus\\_response](#)

**Examples**

```
data(quercus_predictors)
quercus_predictors
```

---

quercus_response	<i>Response variable name for the dataset quercus</i>
------------------	---

---

**Description**

Character string with the name of the response variable in [quercus](#).

**Usage**

```
data(quercus_response)
```

**Format**

A character string of length 1.

**See Also**

Other quercus: [quercus](#), [quercus\\_extra\(\)](#), [quercus\\_predictors](#)

**Examples**

```
data(quercus_response)
quercus_response
```

---

trees

*Mesoamerican tree species richness*

---

### Description

sf data frame with POLYGON geometry representing 3,373 hexagonal grid cells across the Americas, with 1 response variable encoding tree species richness and 50 numeric environmental predictors.

Tree species in this dataset does NOT represent total tree species counts! The dataset focuses on the tree species found in Mesoamerica according to the Tree Biodiversity Network (BIOTREE-NET; Cayuela et al. 2012). These tree species were later used as input for a search query at the Global Biodiversity Information Facility (GBIF). The resulting presence data and environmental data at 1km resolution were aggregated as a hexagonal grid.

The hexagonal grid was constructed using `sf::st_make_grid(..., cellsize = 1, square = FALSE)` at 1-degree resolution (WGS84, EPSG:4326), covering longitudes  $-125.3^{\circ}$  to  $-34.3^{\circ}$  and latitudes  $-34.4^{\circ}$  to  $49.9^{\circ}$ .

### Usage

```
data(trees)
```

### Format

An sf data frame with 3373 rows (hexagonal cells) and 53 columns:

#### Identifier (1):

- `cellid`: Integer row number identifying each hexagonal cell.

#### Response variable (1):

- `trees`: Integer count of tree species richness per hexagonal cell.

#### Predictor variables:

- `air_humidity_max`: Maximum monthly near-surface relative humidity (%).
- `air_humidity`: Mean annual near-surface relative humidity (%).
- `air_humidity_min`: Minimum monthly near-surface relative humidity (%).
- `air_humidity_range`: Annual near-surface relative humidity range (%).
- `aridity`: Mean aridity index (unitless ratio; higher values indicate wetter conditions).
- `cloud_cover_max`: Maximum monthly total cloud cover (%).
- `cloud_cover`: Mean annual total cloud cover (%).
- `cloud_cover_min`: Minimum monthly total cloud cover (%).
- `cloud_cover_range`: Annual total cloud cover range (%).
- `evapotranspiration_max`: Maximum monthly potential evapotranspiration ( $\text{kg m}^{-2} \text{ month}^{-1}$ ; Penman-Monteith).

- evapotranspiration: Mean annual potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_min: Minimum monthly potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_range: Annual potential evapotranspiration range (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- rainfall\_seasonality: Precipitation seasonality as coefficient of variation of monthly totals (CHELSA bio15).
- rainfall: Total annual precipitation (kg m<sup>-2</sup>; CHELSA bio12).
- rainfall\_coldest\_quarter: Precipitation of coldest quarter (kg m<sup>-2</sup>; CHELSA bio19).
- rainfall\_driest\_month: Precipitation of driest month (kg m<sup>-2</sup>; CHELSA bio14).
- rainfall\_driest\_quarter: Precipitation of driest quarter (kg m<sup>-2</sup>; CHELSA bio17).
- rainfall\_warmest\_quarter: Precipitation of warmest quarter (kg m<sup>-2</sup>; CHELSA bio18).
- rainfall\_wettest\_month: Precipitation of wettest month (kg m<sup>-2</sup>; CHELSA bio13).
- rainfall\_wettest\_quarter: Precipitation of wettest quarter (kg m<sup>-2</sup>; CHELSA bio16).
- temperature\_isothermality: Isothermality as ratio of mean daily range to annual range (unitless; CHELSA bio3).
- temperature\_mean\_daily\_range: Mean of monthly temperature ranges (degrees C; CHELSA bio2).
- temperature: Mean annual air temperature (degrees C; CHELSA bio1).
- temperature\_range: Annual air temperature range (degrees C; CHELSA bio7).
- temperature\_seasonality: Temperature seasonality as standard deviation of monthly means (degrees C; CHELSA bio4).
- temperature\_coldest\_month\_min: Minimum temperature of coldest month (degrees C; CHELSA bio6).
- temperature\_coldest\_quarter: Mean temperature of coldest quarter (degrees C; CHELSA bio11).
- temperature\_driest\_quarter: Mean temperature of driest quarter (degrees C; CHELSA bio9).
- temperature\_warmest\_month\_max: Maximum temperature of warmest month (degrees C; CHELSA bio5).
- temperature\_warmest\_quarter: Mean temperature of warmest quarter (degrees C; CHELSA bio10).
- temperature\_wettest\_quarter: Mean temperature of wettest quarter (degrees C; CHELSA bio8).
- distance\_to\_ocean: Distance to nearest ocean coastline (km).
- elevation: Elevation above sea level (m).
- latitude: Latitude of cell centroid (degrees).
- longitude: Longitude of cell centroid (degrees).
- soil\_clay: Soil clay content (%).

- soil\_nitrogen: Soil nitrogen content (g kg<sup>-1</sup>).
- soil\_organic\_carbon: Soil organic carbon content (g kg<sup>-1</sup>).
- soil\_ph: Soil pH in water.
- soil\_sand: Soil sand content (%).
- soil\_silt: Soil silt content (%).
- soil\_temperature\_max: Maximum annual land surface temperature (degrees C).
- soil\_temperature: Mean annual land surface temperature (degrees C).
- soil\_temperature\_min: Minimum annual land surface temperature (degrees C).
- soil\_temperature\_range: Annual land surface temperature range (degrees C).
- ndvi\_max: Maximum annual NDVI (unitless, 0-1).
- ndvi: Mean annual NDVI (unitless, 0-1).
- ndvi\_min: Minimum annual NDVI (unitless, 0-1).
- ndvi\_range: Annual NDVI range (unitless, 0-1).

#### Geometry:

- geometry: Hexagonal polygon geometry (WGS84, EPSG:4326).

#### Source

##### Dataset publication:

- Benito, B.M., Cayuela, L., & Albuquerque, F.S. (2013). The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: Guidelines to build better diversity models. *Methods in Ecology and Evolution*, 4(4), 327–335. doi:10.1111/2041210X.12022

##### Response variable (tree species richness):

- Cayuela, L., Gálvez-Bravo, L., Pérez Pérez, R., de Albuquerque, F.S., Golicher, D.J., Zahawi, R.A., et al. (2012). The Tree Biodiversity Network (BIOTREE-NET): prospects for biodiversity research and conservation in the Neotropics. *Biodiversity & Ecology*, 4, 211–224. doi:10.7809/be.00078
- GBIF: Global Biodiversity Information Facility. <https://www.gbif.org>

##### Climate predictors (temperature, precipitation, air humidity, cloud cover, evapotranspiration):

- Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L., & Karger, D.N. (2022). CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution. *EnviDat*. doi:10.16904/envidat.332

##### Aridity:

- Zomer, R.J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9, 409. doi:10.1038/s41597022014931

##### Soil properties:

- Hengl, T., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. doi:10.1371/journal.pone.0169748

**Soil temperature:**

- Wan, Z., Hook, S., & Hulley, G. (2015). MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. NASA EOSDIS LP DAAC. doi:10.5067/MODIS/MOD11A2.006

**NDVI:**

- Copernicus Land Monitoring Service. (2019). Normalised Difference Vegetation Index Statistics (Long Term 1999-2019), raster 1 km, global, version 3. European Commission, Joint Research Centre. doi:10.2909/290e81fb4c8442adae12f663312b0eda

**Elevation and geography:**

- Jarvis, A., Guevara, E., Reuter, H. I., & Nelson, A. D. (2008). Hole-filled SRTM for the globe: version 4, data grid. Web publication/site, CGIAR Consortium for Spatial Information. <https://srtm.csi.cgiar.org>

**See Also**

Other trees: [trees\\_extra\(\)](#), [trees\\_predictors](#), [trees\\_response](#)

**Examples**

```
data(trees)
colnames(trees)
nrow(trees)
ncol(trees)
```

---

trees\_extra

*Download Presence Records for trees*

---

**Description**

Downloads and reads an sf dataframe with the tree species presence records associated with the [trees](#) dataset from the [spatialDataExtra](#) repository. Writes the file `trees_presence.gpkg` to the working folder, and returns it as an sf dataframe.

**Usage**

```
trees_extra()
```

**Value**

sf data frame with POINT geometry (WGS84, EPSG:4326) and columns species and source.

**See Also**

Other trees: [trees](#), [trees\\_predictors](#), [trees\\_response](#)

---

trees_predictors	<i>Predictor variable names for the dataset trees</i>
------------------	---

---

**Description**

Character vector of 50 predictor variable names from [trees](#).

**Usage**

```
data(trees_predictors)
```

**Format**

A character vector of length 50.

**See Also**

Other trees: [trees](#), [trees\\_extra\(\)](#), [trees\\_response](#)

**Examples**

```
data(trees_predictors)
trees_predictors
```

---

trees_response	<i>Response variable name for the dataset trees</i>
----------------	---

---

**Description**

Character vector of length 1 containing the name of the response variable in [trees](#).

**Usage**

```
data(trees_response)
```

**Format**

A character vector of length 1.

**See Also**

Other trees: [trees](#), [trees\\_extra\(\)](#), [trees\\_predictors](#)

**Examples**

```
data(trees_response)
trees_response
```

## Description

sf data frame with POINT geometry representing 9,265 global locations with one response variable represented in five different encodings of the long-term average (1999-2019) of the Normalized Difference Vegetation Index (NDVI) and 58 environmental predictors (47 numeric, 11 categorical). Use `vi_extra()` to download an extended version with 30,000 rows. There is a smaller version of this dataset (580 rows) named `vi_smol`

NDVI values are derived from the Copernicus Global Land Service Long Term Statistics product (1999-2019) at 1 km resolution. Locations were spatially thinned to reduce spatial autocorrelation.

Environmental predictors were extracted as pixel values from normalized raster data at 1 km resolution.

## Usage

```
data(vi)
```

## Format

An sf data frame with 9265 rows (locations) and 64 columns:

### Response variables (5):

- `vi_numeric`: Continuous NDVI value (0-1).
- `vi_counts`: Integer count encoding of NDVI (`vi_numeric * 1000`).
- `vi_binomial`: Binary encoding of NDVI (1 if `vi_numeric > 0.5`, else 0).
- `vi_categorical`: Categorical encoding of NDVI ("very\_low", "low", "medium", "high", "very\_high").
- `vi_factor`: Factor encoding of NDVI (`vi_categorical` as factor).

### Predictor variables:

- `koppen_zone`: Koppen climate zone code (Beck et al. 2018).
- `koppen_group`: Koppen climate group name.
- `koppen_description`: Koppen climate description.
- `soil_type`: Soil classification type.
- `topo_slope`: Topographic slope in degrees.
- `topo_diversity`: Number of combinations of different elevations, slopes, and aspects in a 5 km radius around each 1 km cell.
- `topo_elevation`: Elevation in meters.
- `swi_mean`: Mean annual soil water index (unitless, 0-100 cm depth).
- `swi_max`: Maximum annual soil water index (unitless, 0-100 cm depth).

- swi\_min: Minimum annual soil water index (unitless, 0-100 cm depth).
- swi\_range: Annual soil water index range (unitless, 0-100 cm depth).
- soil\_temperature\_mean: Mean annual land surface temperature (degrees C).
- soil\_temperature\_max: Maximum annual land surface temperature (degrees C).
- soil\_temperature\_min: Minimum annual land surface temperature (degrees C).
- soil\_temperature\_range: Annual land surface temperature range (degrees C).
- soil\_sand: Soil sand content (%).
- soil\_clay: Soil clay content (%).
- soil\_silt: Soil silt content (%).
- soil\_ph: Soil pH.
- soil\_soc: Soil organic carbon content (%).
- soil\_nitrogen: Soil nitrogen content (%).
- solar\_rad\_mean: Mean annual solar radiation (kJ m<sup>-2</sup>).
- solar\_rad\_max: Maximum annual solar radiation (kJ m<sup>-2</sup>).
- solar\_rad\_min: Minimum annual solar radiation (kJ m<sup>-2</sup>).
- solar\_rad\_range: Annual solar radiation range (kJ m<sup>-2</sup>).
- growing\_season\_length: Length of the growing season (days).
- growing\_season\_temperature: Mean temperature of the growing season (degrees C).
- growing\_season\_rainfall: Accumulated precipitation of the growing season (kg m<sup>-2</sup>).
- growing\_degree\_days: Growing degree days above 0 degrees C accumulated over one year (degree-days).
- temperature\_mean: Mean annual air temperature (degrees C; CHELSA bio1).
- temperature\_max: Maximum temperature of warmest month (degrees C; CHELSA bio5).
- temperature\_min: Minimum temperature of coldest month (degrees C; CHELSA bio6).
- temperature\_range: Annual air temperature range (degrees C; CHELSA bio7).
- temperature\_seasonality: Temperature seasonality as standard deviation of monthly means (degrees C; CHELSA bio4).
- rainfall\_mean: Mean annual rainfall (kg m<sup>-2</sup>).
- rainfall\_min: Minimum monthly rainfall (kg m<sup>-2</sup>).
- rainfall\_max: Maximum monthly rainfall (kg m<sup>-2</sup>).
- rainfall\_range: Annual rainfall range (kg m<sup>-2</sup>).
- evapotranspiration\_mean: Mean annual potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_max: Maximum monthly potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_min: Minimum monthly potential evapotranspiration (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).
- evapotranspiration\_range: Annual potential evapotranspiration range (kg m<sup>-2</sup> month<sup>-1</sup>; Penman-Monteith).

- `cloud_cover_mean`: Mean annual total cloud cover (%).
- `cloud_cover_max`: Maximum monthly total cloud cover (%).
- `cloud_cover_min`: Minimum monthly total cloud cover (%).
- `cloud_cover_range`: Annual total cloud cover range (%).
- `aridity_index`: Mean aridity index (unitless ratio; higher values indicate wetter conditions).
- `humidity_mean`: Mean annual near-surface relative humidity (%).
- `humidity_max`: Maximum monthly near-surface relative humidity (%).
- `humidity_min`: Minimum monthly near-surface relative humidity (%).
- `humidity_range`: Annual near-surface relative humidity range (%).
- `biogeo_ecoregion`: Ecoregion name.
- `biogeo_biome`: Biome name.
- `biogeo_realm`: Ecological realm name.
- `country_name`: Country name.
- `continent`: Continent name.
- `region`: UN region name.
- `subregion`: UN sub-region name.

#### **Geometry:**

- `geometry`: Point geometry (WGS84, EPSG:4326).

#### **Source**

##### **Response variables (NDVI):**

- Copernicus Land Monitoring Service. (2019). Normalised Difference Vegetation Index Statistics (Long Term 1999-2019), raster 1 km, global, version 3. European Commission, Joint Research Centre. doi:10.2909/290e81fb4c8442adae12f663312b0eda

##### **Climate classification:**

- Beck, H.E., et al. (2018). Present and future Koppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5, 180214. doi:10.1038/sdata.2018.214

##### **Soil water index:**

- Copernicus Land Monitoring Service: Soil Water Index. doi:10.2909/290e81fb4c8442adae12f663312b0eda

##### **Climate predictors (temperature, rainfall, solar radiation, growing season, evapotranspiration, cloud cover, humidity):**

- Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L., & Karger, D.N. (2022). CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution. *EnviDat*. doi:10.16904/envidat.332

##### **Soil type and properties:**

- Hengl, T., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. doi:10.1371/journal.pone.0169748

#### Soil temperature:

- Wan, Z., Hook, S., & Hulley, G. (2015). MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. NASA EOSDIS LP DAAC. doi:10.5067/MODIS/MOD11A2.006

#### Ecoregions and biogeography:

- Dinerstein, E., et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), 534-545. doi:10.1093/biosci/bix014

#### Elevation and topography:

- Jarvis, A., Guevara, E., Reuter, H. I., & Nelson, A. D. (2008). Hole-filled SRTM for the globe: version 4, data grid. Web publication/site, CGIAR Consortium for Spatial Information. <https://srtm.csi.cgiar.org>

#### Aridity index:

- Zomer, R.J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9, 409. doi:10.1038/s41597022014931

#### Country, continent, region, and subregion:

- Natural Earth. Free vector and raster map data. <https://www.naturalearthdata.com/>

#### See Also

Other vi: [vi\\_extra\(\)](#), [vi\\_predictors](#), [vi\\_responses](#), [vi\\_smol](#)

#### Examples

```
data(vi)
colnames(vi)
nrow(vi)
ncol(vi)
```

---

vi\_extra

*Download extended vi dataset*

---

#### Description

Downloads and reads the extended version of the [vi](#) dataset (30,000 rows) from the [spatialDataExtra](#) repository. Writes the file `vi.gpkg` to the working directory, and returns it as an `sf` dataframe. See [vi](#) for details on the response variables, predictors, and data sources.

#### Usage

```
vi_extra()
```

**Value**

sf data.frame with 30,000 rows and 64 columns (POINT geometry, WGS84).

**See Also**

Other vi: [vi](#), [vi\\_predictors](#), [vi\\_responses](#), [vi\\_smol](#)

---

vi_predictors	<i>Predictor variable names for the dataset vi</i>
---------------	--

---

**Description**

Character vector of 58 predictor variable names from [vi](#).

**Usage**

```
data(vi_predictors)
```

**Format**

A character vector of length 58.

**See Also**

Other vi: [vi](#), [vi\\_extra\(\)](#), [vi\\_responses](#), [vi\\_smol](#)

**Examples**

```
data(vi_predictors)
vi_predictors
```

---

vi_responses	<i>Response variable names for the dataset vi</i>
--------------	---

---

**Description**

Character vector containing the names of the 5 response variables in [vi](#).

**Usage**

```
data(vi_responses)
```

**Format**

A character vector of length 5.

**See Also**

Other vi: [vi](#), [vi\\_extra\(\)](#), [vi\\_predictors](#), [vi\\_smol](#)

**Examples**

```
data(vi_responses)
vi_responses
```

---

vi\_smol

*Small version of vi*

---

**Description**

Same as dataset [vi](#), but with only 580 rows.

**Usage**

```
data(vi_smol)
```

**Format**

A data frame with 580 rows and 65 columns.

**See Also**

[vi](#)

Other vi: [vi](#), [vi\\_extra\(\)](#), [vi\\_predictors](#), [vi\\_responses](#)

**Examples**

```
data(vi_smol)
colnames(vi_smol)
nrow(vi_smol)
ncol(vi_smol)
```

# Index

- \* **andalusia**
  - andalusia, 3
  - andalusia\_extra, 4
  - andalusia\_predictors, 4
  - andalusia\_responses, 5
- \* **communities**
  - communities, 6
  - communities\_extra\_2010, 7
  - communities\_extra\_2050, 8
  - communities\_extra\_2100, 8
  - communities\_predictors, 9
  - communities\_responses, 9
- \* **datasets**
  - andalusia, 3
  - andalusia\_predictors, 4
  - andalusia\_responses, 5
  - communities, 6
  - communities\_predictors, 9
  - communities\_responses, 9
  - interaction, 10
  - interaction\_predictors, 12
  - interaction\_responses, 12
  - linaria, 13
  - linaria\_predictors, 15
  - linaria\_responses, 16
  - neanderthal, 16
  - neanderthal\_predictors, 19
  - neanderthal\_response, 19
  - plantae, 20
  - plantae\_east, 27
  - plantae\_predictors, 28
  - plantae\_responses, 29
  - plantae\_west, 29
  - quercus, 30
  - quercus\_predictors, 33
  - quercus\_response, 33
  - trees, 34
  - trees\_predictors, 38
  - trees\_response, 38
  - vi, 39
  - vi\_predictors, 43
  - vi\_responses, 43
  - vi\_smol, 44
- \* **interaction**
  - interaction, 10
  - interaction\_extra, 11
  - interaction\_predictors, 12
  - interaction\_responses, 12
- \* **linaria**
  - linaria, 13
  - linaria\_extra, 15
  - linaria\_predictors, 15
  - linaria\_responses, 16
- \* **neanderthal**
  - neanderthal, 16
  - neanderthal\_extra, 18
  - neanderthal\_predictors, 19
  - neanderthal\_response, 19
- \* **plantae**
  - plantae, 20
  - plantae\_east, 27
  - plantae\_extra, 28
  - plantae\_predictors, 28
  - plantae\_responses, 29
  - plantae\_west, 29
- \* **quercus**
  - quercus, 30
  - quercus\_extra, 32
  - quercus\_predictors, 33
  - quercus\_response, 33
- \* **trees**
  - trees, 34
  - trees\_extra, 37
  - trees\_predictors, 38
  - trees\_response, 38
- \* **vi**
  - vi, 39
  - vi\_extra, 42

- vi\_predictors, 43
- vi\_responses, 43
- vi\_smol, 44
  
- andalusia, 3, 4, 5
- andalusia\_extra, 4, 4, 5
- andalusia\_extra(), 3, 4
- andalusia\_predictors, 3, 4, 4, 5
- andalusia\_responses, 4, 5, 5
  
- communities, 6, 7–9
- communities\_extra\_2010, 7, 7, 8, 9
- communities\_extra\_2010(), 6
- communities\_extra\_2050, 7, 8, 8, 9
- communities\_extra\_2050(), 6
- communities\_extra\_2100, 7, 8, 8, 9
- communities\_extra\_2100(), 6
- communities\_predictors, 6–9, 9
- communities\_responses, 6–9, 9
  
- interaction, 10, 11, 12
- interaction\_extra, 11, 11, 12
- interaction\_extra(), 10
- interaction\_predictors, 10–12, 12
- interaction\_responses, 10–12, 12
  
- linaria, 13, 15, 16
- linaria\_extra, 15, 15, 16
- linaria\_extra(), 13
- linaria\_predictors, 13, 15, 15, 16
- linaria\_responses, 13, 15, 16, 16
  
- neanderthal, 16, 18–20
- neanderthal\_extra, 18, 18, 19, 20
- neanderthal\_extra(), 16
- neanderthal\_predictors, 16, 18, 19, 19, 20
- neanderthal\_response, 16, 18, 19, 19
  
- plantae, 20, 27–29
- plantae\_east, 20, 27, 27, 28, 29
- plantae\_extra, 27, 28, 28, 29
- plantae\_extra(), 20
- plantae\_predictors, 27, 28, 28, 29
- plantae\_responses, 27–29, 29
- plantae\_west, 20, 27–29, 29
  
- quercus, 30, 32, 33
- quercus\_extra, 32, 32, 33
- quercus\_extra(), 30
- quercus\_predictors, 30, 32, 33, 33
  
- quercus\_response, 30, 32, 33, 33
  
- trees, 34, 37, 38
- trees\_extra, 37, 37, 38
- trees\_predictors, 37, 38, 38
- trees\_response, 37, 38, 38
  
- vi, 39, 42–44
- vi\_extra, 42, 42, 43, 44
- vi\_extra(), 39
- vi\_predictors, 42, 43, 43, 44
- vi\_responses, 42, 43, 43, 44
- vi\_smol, 39, 42–44, 44