Package 'SQIpro'

May 7, 2026

Type Package

Title Comprehensive Soil Quality Index Computation and Visualization

Version 0.1.0

Description Provides a comprehensive, modular framework for computing the Soil Quality Index (SQI) using six established methods: Linear Scoring (Doran and Parkin, 1994, <doi:10.2136/sssaspecpub35.c1>), Regression-based (Masto et al., 2008, <doi:10.1007/s10661-007-9697-z>), Principal Component Analysis-based (Andrews et al., 2004, <doi:10.2136/sssaj2004.1945>), Fuzzy Logic, Entropy Weighting (Shannon, 1948, <doi:10.1002/j.1538-7305.1948.tb01338.x>), and TOPSIS (Hwang and Yoon, 1981, <doi:10.1007/978-3-642-48318-9>). Implements four variable scoring functions: more-is-better, less-is-better, optimum-value, and trapezoidal, following Karlen and Stott (1994, <doi:10.2136/sssaspecpub35.c4>). Includes automated Minimum Data Set selection via Principal Component Analysis with Variance Inflation Factor filtering (Kaiser, 1960, <doi:10.1177/001316446002000116>), one-way ANOVA with Tukey HSD post-hoc tests, leave-one-out sensitivity analysis, and publication-quality visualization using 'ggplot2'.

License GPL (>= 3)

Encoding UTF-8

Language en-US

LazyData true

LazyDataCompression xz

RoxygenNote 7.3.3

Depends R (>= 4.0.0)

Imports dplyr (>= 1.0.0), tidyr (>= 1.1.0), ggplot2 (>= 3.3.0), FactoMineR (>= 2.4), factoextra (>= 1.0.7), stats, methods, rlang (>= 0.4.0), matrixStats (>= 0.61.0), glmnet (>= 4.1.0), car (>= 3.0.0)

Suggests openxlsx (>= 4.2.4), readxl (>= 1.3.1), ggpubr (>= 0.4.0), testthat (>= 3.0.0), knitr (>= 1.33), rmarkdown (>= 2.11), spelling (>= 2.2), covr (>= 3.5.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Sadikul Islam [aut, cre] (ORCID:
<https://orcid.org/0000-0003-2924-7122>),
Rajesh Kaushal [aut]

Maintainer Sadikul Islam <sadikul.islamiasri@gmail.com>

Repository CRAN

Date/Publication 2026-04-20 12:30:02 UTC

Contents

make_config *Build a Variable Configuration Table*

Description

Constructs a variable configuration data frame that specifies the scoring function type and relevant parameters for each soil indicator. This configuration table is the central object passed to all scoring and indexing functions in **SQIpro**.

Usage

```
make_config(
  variable,
  type,
  opt_low = rep(NA_real_, length(variable)),
  opt_high = rep(NA_real_, length(variable)),
  min_val = rep(NA_real_, length(variable)),
  max_val = rep(NA_real_, length(variable)),
  weight = rep(1, length(variable)),
  description = rep(NA_character_, length(variable))
)
```

Arguments

variable	Character vector of variable names (must match column names in the data).
type	Character vector of scoring types, one per variable. Must be one of: "more" Higher values are better (e.g., organic carbon, CEC, microbial biomass). "less" Lower values are better (e.g., bulk density, EC, heavy metals). "opt" A specific optimum value or range is best (e.g., pH, clay content). Requires opt_low and opt_high. "trap" A trapezoidal function with a flat optimum plateau and tapered shoulders. Requires all four boundary parameters. "custom" User-supplied scoring function via score_custom.
opt_low	Numeric vector. Lower bound of optimum range (required for "opt" and "trap" types; NA otherwise).
opt_high	Numeric vector. Upper bound of optimum range (required for "opt" and "trap" types; NA otherwise).
min_val	Numeric vector. Absolute minimum value (required for "trap"; NA otherwise). Values at or below this receive a score of 0.
max_val	Numeric vector. Absolute maximum value (required for "trap"; NA otherwise). Values at or above this receive a score of 0.
weight	Numeric vector of user-defined weights (0–1). Used only when method = "weighted" in sqi_linear. Defaults to 1 (equal weights).
description	Character vector. Optional human-readable description of each variable (units, rationale). Useful for automated reports.

Value

A data frame (class `sqi_config`) with one row per variable.

References

Doran, J.W., & Parkin, T.B. (1994). Defining and assessing soil quality. In J.W. Doran et al. (Eds.), *Defining Soil Quality for a Sustainable Environment*, pp. 1–21. SSSA Special Publication 35. doi:10.2136/sssaspecpub35.c1

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Examples

```
cfg <- make_config(
  variable    = c("pH",  "EC",  "BD",  "OC",  "MBC", "Clay"),
  type        = c("opt", "less","less","more","more","opt"),
  opt_low     = c(6.0,   NA,    NA,    NA,    NA,    20),
  opt_high    = c(7.0,   NA,    NA,    NA,    NA,    35),
  description = c("Soil pH (H2O)",
                  "Electrical Conductivity (dS/m)",
                  "Bulk Density (g/cm3)",
                  "Organic Carbon (%)",
                  "Microbial Biomass Carbon (mg/kg)",
                  "Clay content (%)")
)
print(cfg)
```

---

plot_pca_biplot *PCA Biplot of Soil Variables and Groups*

Description

Renders a PCA biplot showing both variable loadings and group centroids, using `factoextra::fviz_pca_biplot`. Useful for understanding variable relationships and group separation underlying MDS selection.

Usage

```
plot_pca_biplot(
  mds,
  scored,
  group_col = "LandUse",
  title = "PCA Biplot of Soil Quality Variables"
)
```

Arguments

mds	An object returned by select_mds.
scored	A scored data frame (for group colour coding).
group_col	Character. Column for group colours.
title	Character. Plot title.

Value

A ggplot object.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
mds    <- select_mds(scored, group_cols = c("LandUse","Depth"))
plot_pca_biplot(mds, scored, group_col = "LandUse")
```

---

plot_radar *Radar / Spider Chart of Variable Scores*

Description

Draws a radar (spider) chart comparing mean variable scores across groups. Useful for visualising the multidimensional soil quality profile of each land-use system.

Usage

```
plot_radar(
  scored,
  config,
  group_col,
  group_cols = group_col,
  vars = NULL,
  title = "Soil Quality Radar Profile",
  palette = c("#1b7837", "#762a83", "#d6604d", "#4393c3", "#f4a582")
)
```

Arguments

scored	A scored data frame from score_all.
config	A sqi_config object.
group_col	Character. Column used to define radar chart series.
group_cols	Character vector of all grouping columns.
vars	Character vector of variables to include. Defaults to all in config.
title	Character. Plot title.
palette	Character vector of colours for each group.

Value

Invisible NULL; the chart is rendered to the active graphics device.

References

Chambers, J.M., & Hastie, T.J. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
plot_radar(scored, cfg, group_col = "LandUse",
           group_cols = c("LandUse","Depth"))
```

---

plot_scores *Plot Individual Variable Scores as a Heatmap*

Description

Displays a heatmap of mean variable scores (0–1) per group, allowing rapid visual identification of which variables drive high or low SQI within each land-use system.

Usage

```
plot_scores(
  scored,
  config,
  group_cols = "LandUse",
  group_by = group_cols[1],
```

```
  facet_by = NULL,
  palette = "RdYlGn",
  title = "Mean Variable Scores by Group"
)
```

Arguments

scored	A scored data frame from score_all.
config	A sqi_config object.
group_cols	Character vector. Grouping columns.
group_by	Character. Which grouping column to display on the x-axis.
facet_by	Character or NULL. Optional column to facet by (e.g., "Depth").
palette	Character. Colour palette: "RdYlGn" (default), "Blues", or any RColorBrewer name.
title	Character. Plot title.

Value

A ggplot object.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
plot_scores(scored, cfg, group_cols = c("LandUse","Depth"),
            group_by = "LandUse", facet_by = "Depth")
```

---

plot_scoring_curves *Plot Scoring Curves for All Variables*

Description

Visualises the scoring function (0–1 transformation) for each variable in the configuration, overlaid on the observed data distribution. This plot is essential for verifying that the scoring configuration is biologically sensible before computing indices.

Usage

```
plot_scoring_curves(data, config, group_cols = "LandUse", ncol = 3)
```

Arguments

data	The raw (unscored) soil data frame.
config	A sqi_config object.
group_cols	Character vector of grouping columns to exclude.
ncol	Integer. Number of columns in the facet grid. Default 3.

Value

A ggplot object.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
plot_scoring_curves(soil_data, cfg,
                    group_cols = c("LandUse","Depth"))
```

---

plot_sensitivity *Sensitivity Tornado Chart*

Description

Visualises variable importance from sqi_sensitivity as a horizontal bar (tornado) chart, ordered from most to least sensitive.

Usage

```
plot_sensitivity(sa_result, title = "Variable Sensitivity to SQI")
```

Arguments

sa_result	Data frame from sqi_sensitivity.
title	Character. Plot title.

Value

A ggplot object.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
sa     <- sqi_sensitivity(scored, cfg, group_cols = c("LandUse","Depth"))
plot_sensitivity(sa)
```

`plot_sqi` *Plot Soil Quality Index Across Groups*

Description

Creates a grouped bar chart of SQI values per group, with optional error bars (standard deviation computed across replicate observations before indexing) and significance letters.

Usage

```
plot_sqi(
  sqi_result,
  sqi_col,
  group_col,
  fill_col = NULL,
  letters_df = NULL,
  title = "Soil Quality Index",
  y_label = "SQI (0-1)",
  palette = c("#2d6a4f", "#52b788", "#95d5b2", "#d8f3dc", "#b7e4c7")
)
```

Arguments

`sqi_result`	A data frame returned by any `sqi_*()` function.
`sqi_col`	Character. Name of the SQI column to plot.
`group_col`	Character. Grouping column for the x-axis.
`fill_col`	Character or `NULL`. Column for fill (e.g., `"Depth"` to produce side-by-side bars).
`letters_df`	Data frame with columns `Group` and `Letter` (from `sqi_anova`), or `NULL`.
`title`	Character. Plot title.
`y_label`	Character. Y-axis label.
`palette`	Character vector of colours.

Value

A ggplot object.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
res    <- sqi_linear(scored, cfg, group_cols = c("LandUse","Depth"))
plot_sqi(res, sqi_col = "SQI_linear", group_col = "LandUse",
         fill_col = "Depth")
```

---

score_all *Score All Variables Using a Configuration Table*

Description

Applies the appropriate scoring function to each soil variable according to a configuration table produced by make_config. This is the primary data-preparation step before computing any Soil Quality Index.

Usage

```
score_all(data, config, group_cols = "LandUse", custom_fns = list())
```

Arguments

data	A data frame containing the soil variables.
config	A sqi_config data frame (see make_config).
group_cols	Character vector of grouping column names to preserve unchanged. Default is "LandUse".
custom_fns	A named list of functions for variables with type = "custom". Names must match the variable column in config.

Value

A data frame with the same structure as data, but with each variable column replaced by its 0–1 score. Group columns are preserved unchanged.

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH",  "EC",   "BD",   "OC",   "MBC",  "Clay"),
  type     = c("opt", "less", "less", "more", "more", "opt"),
  opt_low  = c(6.0,   NA,     NA,     NA,     NA,     20),
  opt_high = c(7.0,   NA,     NA,     NA,     NA,     35)
)
scored <- score_all(soil_data, cfg,
                    group_cols = c("LandUse", "Depth"))
head(scored)
```

`score_custom` *Score a Variable With a User-Defined Function*

Description

Applies an arbitrary user-defined scoring function to a numeric vector. The function must accept a numeric vector and return a numeric vector of the same length with values in [0, 1].

Usage

```
score_custom(x, FUN, ...)
```

Arguments

`x`	Numeric vector of raw variable values.
`FUN`	A function with signature `function(x)` that returns numeric scores in [0, 1].
`...`	Additional arguments passed to `FUN`.

Value

Numeric vector of scores in [0, 1].

Examples

```
# Log-linear scoring for a skewed variable
mbc <- c(30, 80, 200, 400, 600)
score_custom(mbc, FUN = function(x) {
  s <- (log(x) - log(min(x))) / (log(max(x)) - log(min(x)))
  pmin(pmax(s, 0), 1)
})
```

score_less *Score a Variable Where Lower Values Are Better*

Description

Applies a "less is better" linear scoring function, transforming raw variable values to a 0–1 score. Suitable for soil indicators where lower values denote better soil quality, such as bulk density, electrical conductivity, or heavy metal concentrations (Andrews et al., 2004).

The score is computed as:

$$S_i = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}}$$

Usage

```
score_less(x, x_min = NULL, x_max = NULL)
```

Arguments

x	Numeric vector of raw variable values.
x_min	Numeric. Lower bound. Defaults to min(x).
x_max	Numeric. Upper bound. Defaults to max(x).

Value

Numeric vector of scores in [0, 1].

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Examples

```
bd <- c(0.9, 1.1, 1.3, 1.5, 1.7)   # Bulk Density (g/cm3)
score_less(bd)

# With domain bounds
score_less(bd, x_min = 0.8, x_max = 2.0)
```

score_more *Score a Variable Where Higher Values Are Better*

Description

Applies a "more is better" linear scoring function, transforming raw variable values to a 0–1 score. This is appropriate for soil indicators where higher values improve soil function, such as organic carbon, microbial biomass, or cation exchange capacity (Andrews et al., 2004; Karlen & Stott, 1994).

The score is computed as:

$$S_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

where $x_{\min}$ and $x_{\max}$ are taken from the observed data (or from user-supplied bounds).

Usage

```
score_more(x, x_min = NULL, x_max = NULL)
```

Arguments

x	Numeric vector of raw variable values.
x_min	Numeric. Lower bound for scoring. Defaults to `min(x, na.rm = TRUE)`.
x_max	Numeric. Upper bound for scoring. Defaults to `max(x, na.rm = TRUE)`.

Value

Numeric vector of scores in [0, 1].

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Karlen, D.L., & Stott, D.E. (1994). A framework for evaluating physical and chemical indicators of soil quality. In J.W. Doran et al. (Eds.), *Defining Soil Quality for a Sustainable Environment*, pp. 53–72. SSSA Special Publication 35. doi:10.2136/sssaspecpub35.c4

Examples

```
oc <- c(0.5, 1.2, 2.1, 3.4, 4.5)   # Organic Carbon (%)
score_more(oc)

# With user-defined bounds (e.g., 0 to 5%)
score_more(oc, x_min = 0, x_max = 5)
```

`score_optimum` *Score a Variable With an Optimum Value or Range (Bell Curve)*

Description

Applies a bell-shaped (peaked) scoring function appropriate for soil variables that have an optimum range, beyond which both higher and lower values reduce soil quality. Classic examples include pH (optimal 6.0–7.0 for most crops) and clay content (Liebig et al., 1996; Karlen & Stott, 1994).

The scoring rules are:

- $S = 1$ if $x \in [\texttt{opt\_low}, \texttt{opt\_high}]$
- $S = (x - x_{\min})/(\texttt{opt\_low} - x_{\min})$ if $x < \texttt{opt\_low}$
- $S = (x_{\max} - x)/(x_{\max} - \texttt{opt\_high})$ if $x > \texttt{opt\_high}$

Usage

```
score_optimum(x, opt_low, opt_high, x_min = NULL, x_max = NULL)
```

Arguments

`x`	Numeric vector of raw variable values.
`opt_low`	Numeric. Lower bound of the optimum range.
`opt_high`	Numeric. Upper bound of the optimum range.
`x_min`	Numeric. Absolute minimum (score = 0). Defaults to `min(x)`.
`x_max`	Numeric. Absolute maximum (score = 0). Defaults to `max(x)`.

Value

Numeric vector of scores in [0, 1].

References

Karlen, D.L., & Stott, D.E. (1994). A framework for evaluating physical and chemical indicators of soil quality. In J.W. Doran et al. (Eds.), *Defining Soil Quality for a Sustainable Environment*, pp. 53–72. SSSA Special Publication 35. doi:10.2136/sssaspecpub35.c4

Liebig, M.A., Varvel, G., & Doran, J.W. (1996). A simple performance- based index for assessing multiple agroecosystem functions. *Agronomy Journal*, 88, 739–745. doi:10.2134/agronj1996.00021962008800050011x

Examples

```
ph <- c(4.5, 5.5, 6.2, 6.8, 7.0, 7.5, 8.2)
score_optimum(ph, opt_low = 6.0, opt_high = 7.0)

clay <- c(10, 18, 25, 32, 45, 60)
score_optimum(clay, opt_low = 20, opt_high = 35)
```

score_trapezoid *Score a Variable With a Trapezoidal Function*

Description

Applies a trapezoidal scoring function where scores are 1 within an ideal plateau [opt_low, opt_high], rise linearly from 0 at min_val to 1 at opt_low, and fall linearly from 1 at opt_high to 0 at max_val. Values outside [min_val, max_val] receive a score of 0.

This function is more flexible than score_optimum because the user explicitly controls the zero-score boundaries, making it suitable for variables with well-established critical thresholds.

Usage

```
score_trapezoid(x, min_val, opt_low, opt_high, max_val)
```

Arguments

x	Numeric vector of raw variable values.
min_val	Numeric. Value at which score becomes 0 on the low side.
opt_low	Numeric. Lower bound of the plateau (score = 1).
opt_high	Numeric. Upper bound of the plateau (score = 1).
max_val	Numeric. Value at which score becomes 0 on the high side.

Value

Numeric vector of scores in [0, 1].

References

Wymore, A.W. (1993). *Model-Based Systems Engineering*. CRC Press, Boca Raton, FL.

Buse, R., & Lele, S. (2003). Trapezoidal membership functions in fuzzy soil quality assessment. *Geoderma*, 114, 177–196.

Examples

```
ph <- c(3.5, 5.0, 6.5, 7.0, 7.8, 8.5, 9.5)
# pH: absolute zero below 4 and above 9; ideal 6.0-7.0
score_trapezoid(ph, min_val = 4.0, opt_low = 6.0,
                opt_high = 7.0, max_val = 9.0)
```

`select_mds` *Select a Minimum Data Set (MDS) of Soil Quality Indicators*

Description

Identifies the most informative subset of soil variables (the Minimum Data Set, MDS) using Principal Component Analysis (PCA). Only variables with high factor loadings on principal components explaining eigenvalue > 1 (Kaiser criterion) are retained. Where multiple variables load highly on the same component, the one with the highest correlation to others in that component is selected to minimise redundancy.

This approach follows the widely cited method of Andrews et al. (2004) and Sharma et al. (2008), and is equivalent to the `PCAIndex` algorithm in Wani et al. (2023).

Usage

```
select_mds(
  data,
  group_cols = "LandUse",
  load_threshold = 0.5,
  vif_threshold = 10,
  n_pc = "auto",
  verbose = TRUE
)
```

Arguments

`data`	A data frame of scored or raw soil variables (numeric columns only, or with group columns specified in `group_cols`).
`group_cols`	Character vector of grouping columns to exclude from the analysis. Default: `"LandUse"`.
`load_threshold`	Numeric in (0, 1). Minimum absolute factor loading for a variable to be considered for MDS membership. Default: `0.6` (Andrews et al., 2004).
`vif_threshold`	Numeric. Maximum allowable Variance Inflation Factor among MDS variables. Variables exceeding this are iteratively removed. Set to `Inf` to skip VIF filtering. Default: `10`.
`n_pc`	Integer or `"auto"`. Number of principal components to consider. `"auto"` (default) uses the Kaiser criterion (eigenvalue > 1).
`verbose`	Logical. Print MDS selection summary. Default `TRUE`.

Details

Algorithm steps:

1. Standardise all numeric variables (mean = 0, sd = 1).
2. Perform PCA; retain components with eigenvalue > 1.

3. For each retained component, identify variables with absolute loading $\geq$ `load_threshold`.
4. Among those, select the variable with the highest sum of absolute Pearson correlations to all others in the set (i.e., the most correlated, least redundant variable).
5. Optionally, remove variables with high Variance Inflation Factor (VIF > `vif_threshold`) among the MDS candidates.

Value

A list of class `sqi_mds` with:

mds_vars Character vector of selected MDS variable names.

all_vars Character vector of all candidate variable names.

pca The PCA result object.

loadings Matrix of factor loadings.

eigenvalues Numeric vector of eigenvalues.

var_explained Numeric vector of variance explained (%) per component.

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework: A quantitative soil quality evaluation method. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. doi:10.1177/001316446002000116

Sharma, K.L., et al. (2008). Long-term soil management effects on soil quality indices. *Geoderma*, 144, 290–300. doi:10.1016/j.geoderma.2007.11.019

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","PMN","Clay","WHC","DEH","AP","TN"),
  type     = c("opt","less","less","more","more","more",
               "opt","more","more","more","more"),
  opt_low  = c(6.0, NA, NA, NA, NA, NA, 20, NA, NA, NA, NA),
  opt_high = c(7.0, NA, NA, NA, NA, NA, 35, NA, NA, NA, NA)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
mds    <- select_mds(scored, group_cols = c("LandUse","Depth"))
mds$mds_vars
```

`soil_data` *Hypothetical Soil Quality Dataset*

Description

A hypothetical dataset representing soil physicochemical and biological properties across five land-use systems and two soil depths, generated using realistic parameter ranges reported in the soil quality literature. This dataset is intended for demonstrating the functions in the **SQIpro** package and for pedagogical purposes.

Usage

```
soil_data
```

Format

A data frame with 100 rows and 14 variables:

LandUse Character. Land-use system: `Natural_Forest`, `Agroforestry`, `Cropland`, `Grassland`, `Degraded_Land`.

Depth Character. Soil depth: `Surface_0_15cm` or `Subsurface_15_30cm`.

pH Numeric. Soil pH (1:2.5 water suspension). Unitless. Optimal range 6.0–7.0 for most crops (Brady & Weil, 2008).

EC Numeric. Electrical Conductivity (dS m^{-1}). Lower values indicate less salinity stress; <0.2 dS m^{-1} considered non-saline (Richards, 1954).

BD Numeric. Bulk Density (g cm^{-3}). Lower values indicate better soil structure and aeration; >1.6 g cm^{-3} restricts root growth (Arshad et al., 1996).

CEC Numeric. Cation Exchange Capacity (cmol(+) kg^{-1}). Higher values indicate greater nutrient-holding capacity (Sparks, 2003).

OC Numeric. Organic Carbon (%). Higher values indicate greater soil organic matter, a key indicator of soil health (Doran & Parkin, 1994).

MBC Numeric. Microbial Biomass Carbon (mg kg^{-1}). Indicator of soil biological activity (Brookes, 1995).

PMN Numeric. Potentially Mineralizable Nitrogen (mg kg^{-1}). Indicates N-supplying capacity of soil (Stanford & Smith, 1972).

Clay Numeric. Clay content (%). Optimal range 20–35% for water and nutrient retention (Arshad et al., 1996).

WHC Numeric. Water Holding Capacity (%). Higher values indicate better moisture retention (Reynolds et al., 2009).

DEH Numeric. Dehydrogenase Enzyme Activity (μg TPF g^{-1} day^{-1}). Indicator of overall microbial metabolic activity (Casida et al., 1964).

AP Numeric. Available Phosphorus (mg kg^{-1}). Higher values indicate better P availability for plants (Olsen & Sommers, 1982).

TN Numeric. Total Nitrogen (%). Higher values indicate greater N reserves (Bremner, 1996).

Details

Parameter ranges were informed by values reported in:

- Doran and Parkin (1994) for biological indicators
- Andrews et al. (2004) for MDS indicator ranges
- Masto et al. (2008) for land-use comparison ranges

The dataset is entirely synthetic and does not represent any specific geographic location.

Source

Synthetically generated for the **SQIpro** package.

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework: A quantitative soil quality evaluation method. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Arshad, M.A., Lowery, B., & Grossman, B. (1996). Physical tests for monitoring soil quality. In J.W. Doran & A.J. Jones (Eds.), *Methods for Assessing Soil Quality*, pp. 123–141. SSSA Special Publication 49. doi:10.2136/sssaspecpub49.c7

Brady, N.C., & Weil, R.R. (2008). *The Nature and Properties of Soils* (14th ed.). Prentice Hall, New Jersey.

Doran, J.W., & Parkin, T.B. (1994). Defining and assessing soil quality. In J.W. Doran et al. (Eds.), *Defining Soil Quality for a Sustainable Environment*, pp. 1–21. SSSA Special Publication 35. doi:10.2136/sssaspecpub35.c1

Masto, R.E., Chhonkar, P.K., Singh, D., & Patra, A.K. (2008). Alternative soil quality indices for evaluating the effect of intensive cropping, fertilisation and manuring for 31 years in the semi-arid soils of India. *Environmental Monitoring and Assessment*, 136, 419–435. doi:10.1007/s10661007-9697z

Examples

```
data(soil_data)
head(soil_data)
summary(soil_data)
table(soil_data$LandUse, soil_data$Depth)
```

---

sqi_anova *One-Way ANOVA and Tukey HSD Post-Hoc Test for SQI*

Description

Performs a one-way ANOVA to test whether Soil Quality Index values differ significantly across land-use groups, followed by Tukey's Honest Significant Difference (HSD) test for pairwise comparisons.

Usage

```
sqi_anova(scored, sqi_col, group_col, alpha = 0.05)
```

Arguments

scored	A scored data frame from score_all.
sqi_col	Character. Name of the SQI column to test (e.g., "SQI_linear"). This must be a column in scored or returned by one of the indexing functions joined back to the data. Alternatively pass the output of sqi_compare with individual observations merged in.
group_col	Character. Grouping variable column name (e.g., "LandUse").
alpha	Numeric. Significance level for the ANOVA. Default 0.05.

Value

A list with:

anova_table An anova object.

tukey A TukeyHSD object.

significant Logical. Whether the ANOVA is significant at alpha.

compact_letters Data frame of compact letter display for plotting.

References

Tukey, J.W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99–114. doi:10.2307/3001913

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
# Compute per-observation linear SQI for ANOVA
scored$SQI_obs <- rowMeans(scored[, cfg$variable], na.rm = TRUE)
aov_result <- sqi_anova(scored, sqi_col = "SQI_obs",
                        group_col = "LandUse")
print(aov_result$tukey)
```

`sqi_compare` *Compare All SQI Methods*

Description

Runs all six SQI methods (`sqi_linear`, `sqi_regression`, `sqi_pca`, `sqi_fuzzy`, `sqi_entropy`, `sqi_topsis`) on the same scored dataset and returns a combined results table for method comparison.

Usage

```
sqi_compare(scored, config, group_cols = "LandUse", dep_var = NULL, mds = NULL)
```

Arguments

`scored`	A scored data frame from `score_all`.
`config`	A `sqi_config` object.
`group_cols`	Character vector of grouping column names.
`dep_var`	Character. Dependent variable for `sqi_regression`. If `NULL`, the regression method is skipped.
`mds`	Object from `select_mds`, or `NULL` to compute automatically.

Value

A data frame with one row per group and columns for each SQI method. Also includes `Mean_SQI` and `Rank` columns.

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored  <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
results <- sqi_compare(scored, cfg, group_cols = c("LandUse","Depth"),
                       dep_var = "OC")
print(results)
```

`sqi_entropy` *Soil Quality Index: Entropy Weighting Method*

Description

Computes SQI using Shannon entropy to derive objective weights for each variable. Variables with higher information entropy (greater discriminating power among groups) receive higher weights. This removes subjectivity from weight assignment.

The entropy weight for variable j is:

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln(p_{ij})$$

$$w_j = \frac{1 - e_j}{\sum_k (1 - e_k)}$$

where $p_{ij} = \bar{S}_{ij} / \sum_i \bar{S}_{ij}$.

Usage

```
sqi_entropy(scored, config, group_cols = "LandUse", mds_vars = NULL)
```

Arguments

`scored`	A scored data frame from `score_all`.
`config`	A `sqi_config` object.
`group_cols`	Character vector of grouping column names.
`mds_vars`	Character vector of MDS variable names.

Value

A data frame with group columns, `SQI_entropy`, and attribute `entropy_weights` (named numeric vector).

References

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. doi:10.1002/j.15387305.1948.tb01338.x

Li, P., Qian, H., & Wu, J. (2010). Groundwater quality assessment based on improved water quality index in Pengyang County, Ningxia, Northwest China. *E-Journal of Chemistry*, 7, 209–216. doi:10.1155/2010/451304

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
result <- sqi_entropy(scored, cfg, group_cols = c("LandUse","Depth"))
attr(result, "entropy_weights")
```

---

sqi_fuzzy *Soil Quality Index: Fuzzy Logic Method*

Description

Computes SQI using a fuzzy membership aggregation approach. Each scored variable (already 0–1) is treated as a fuzzy membership value, and groups are aggregated using either the arithmetic mean (equivalent to the linear method) or the fuzzy weighted average operator.

This approach is appropriate when variable importance is uncertain or when expert-elicited weights are available (Zhu et al., 2006; Torbert & Wood, 1992).

Usage

```
sqi_fuzzy(
  scored,
  config,
  group_cols = "LandUse",
  mds_vars = NULL,
  fuzzy_weights = NULL,
  operator = c("mean", "geometric")
)
```

Arguments

`scored`	A scored data frame from `score_all`.
`config`	A `sqi_config` object.
`group_cols`	Character vector of grouping column names.
`mds_vars`	Character vector of MDS variable names.
`fuzzy_weights`	Named numeric vector of fuzzy importance weights (sum need not equal 1; they are normalised internally). Defaults to equal weights.
`operator`	Character. Aggregation operator: `"mean"` (default) or `"geometric"` (product-based, penalises low scores on any variable).

Value

A data frame with group columns and `SQI_fuzzy` (0–1).

References

Zhu, A.X., Liu, F., Li, B., Pei, T., Qin, C., Liu, G., Wang, Y., Chen, Y., Ma, X., Qi, F., & Li, R. (2010). Differentiation of soil conditions over flat areas using land surface feedback dynamic patterns extracted from MODIS. *Soil Science Society of America Journal*, 74(1), 861–869.

Torbert, H.A., & Wood, C.W. (1992). Effects of soil compaction and water-filled pore space on soil microbial activity and N losses. *Communications in Soil Science and Plant Analysis*, 23, 1321–1331. doi:10.1080/00103629209368668

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
result <- sqi_fuzzy(scored, cfg, group_cols = c("LandUse","Depth"))
print(result)
```

---

sqi_linear *Soil Quality Index: Linear Scoring Method*

Description

Computes the Soil Quality Index (SQI) using the linear additive scoring method of Doran & Parkin (1994) and Andrews et al. (2004). Each variable score (0–1, from `score_all`) is averaged across replicates within each group, optionally weighted, and then min-max normalised to produce the final index.

$$SQI_g = \frac{\sum_{j=1}^{p} w_j \bar{S}_{gj}}{\sum_{j=1}^{p} w_j}$$

where $\bar{S}_{gj}$ is the mean score of variable j in group g and w_j is the weight of variable j.

Usage

```
sqi_linear(
  scored,
  config,
  group_cols = "LandUse",
```

```
  mds_vars = NULL,
  weights = NULL
)
```

Arguments

scored	A scored data frame from score_all.
config	A sqi_config object (see make_config).
group_cols	Character vector of grouping column names.
mds_vars	Character vector. If supplied, only these variables are used. Otherwise all numeric variables in config are used.
weights	Named numeric vector of variable weights. Defaults to equal weights (1 for all). Names must match variable names.

Value

A data frame with group columns plus:

SQI_linear Final normalised Soil Quality Index (0–1).

Raw_score Weighted mean score before normalisation.

References

Doran, J.W., & Parkin, T.B. (1994). Defining and assessing soil quality. In J.W. Doran et al. (Eds.), *Defining Soil Quality for a Sustainable Environment*, pp. 1–21. SSSA Special Publication 35. doi:10.2136/sssaspecpub35.c1

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
result <- sqi_linear(scored, cfg, group_cols = c("LandUse","Depth"))
print(result)
```

sqi_pca *Soil Quality Index: PCA-Based Method*

Description

Computes SQI using Principal Component Analysis, weighting selected MDS variables by the proportion of variance their component explains. This is the most widely cited data-driven approach in soil quality research (Andrews et al., 2004; Bastida et al., 2008).

$$SQI_{PCA} = \sum_{k=1}^{m} \frac{V_k}{\sum V} \bar{S}_{g,j_k}$$

where V_k is the variance explained by component k, j_k is the MDS variable selected from component k, and $\bar{S}_{g,j_k}$ is the group mean score of that variable.

Usage

```
sqi_pca(scored, config, group_cols = "LandUse", mds = NULL)
```

Arguments

scored	A scored data frame from score_all.
config	A sqi_config object.
group_cols	Character vector of grouping column names.
mds	Object returned by select_mds. If NULL, select_mds is run internally with default parameters.

Value

A data frame with group columns and SQI_pca (0–1).

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Bastida, F., Zsolnay, A., Hernandez, T., & Garcia, C. (2008). Past, present and future of soil quality indices: A biological perspective. *Geoderma*, 147(3–4), 159–171. doi:10.1016/j.geoderma.2008.08.007

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
```

```
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
result <- sqi_pca(scored, cfg, group_cols = c("LandUse","Depth"))
print(result)
```

`sqi_regression` *Soil Quality Index: Regression-Based Method*

Description

Computes the SQI using stepwise multiple linear regression to identify and weight the most predictive soil variables. The dependent variable (e.g., crop yield, total biomass) determines which variables enter the model. Regression coefficients serve as weights in the index.

This follows the method described by Masto et al. (2008) and Mukherjee & Lal (2014).

Usage

```
sqi_regression(
  scored,
  config,
  dep_var,
  group_cols = "LandUse",
  mds_vars = NULL,
  direction = "both"
)
```

Arguments

`scored`	A scored data frame from `score_all`.
`config`	A `sqi_config` object.
`dep_var`	Character. Name of the dependent variable column in `scored`.
`group_cols`	Character vector of grouping column names.
`mds_vars`	Character vector of candidate predictor variable names. If `NULL`, all variables in `config` are used.
`direction`	Character. Direction for stepwise selection: `"both"` (default), `"forward"`, or `"backward"`.

Value

A data frame with group columns plus:

SQI_regression Normalised SQI (0–1).

selected_vars (attribute) Character vector of selected predictors.

References

Masto, R.E., Chhonkar, P.K., Singh, D., & Patra, A.K. (2008). Alternative soil quality indices. *Environmental Monitoring and Assessment*, 136, 419–435. doi:10.1007/s106610079697z

Mukherjee, A., & Lal, R. (2014). Comparison of soil quality index using three methods. *PLOS ONE*, 9(8), e105981. doi:10.1371/journal.pone.0105981

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
# OC used as surrogate dependent variable
result <- sqi_regression(scored, cfg, dep_var = "OC",
                         group_cols = c("LandUse","Depth"))
print(result)
```

`sqi_sensitivity` *Sensitivity Analysis for SQI*

Description

Quantifies the contribution of each soil variable to the overall Soil Quality Index by a leave-one-out approach: each variable is removed in turn and the resulting index is compared to the full index. A larger change indicates a higher-sensitivity (more important) variable.

Usage

```
sqi_sensitivity(
  scored,
  config,
  group_cols = "LandUse",
  method = c("linear", "fuzzy", "entropy", "topsis"),
  mds_vars = NULL
)
```

Arguments

`scored`	A scored data frame from `score_all`.
`config`	A `sqi_config` object.
`group_cols`	Character vector of grouping columns.

`method`	Character. Which indexing method to use for sensitivity analysis: `"linear"` (default), `"fuzzy"`, `"entropy"`, or `"topsis"`.
`mds_vars`	Character vector of MDS variable names. If `NULL`, all config variables are used.

Value

A data frame with columns `variable`, `mean_change` (mean absolute change in SQI when variable is removed), `sd_change`, and `relative_importance` (0–1, normalised).

References

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, Chichester. doi:10.1002/9780470725184

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored  <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
sa      <- sqi_sensitivity(scored, cfg, group_cols = c("LandUse","Depth"))
print(sa)
```

---

sqi_topsis *Soil Quality Index: TOPSIS Method*

Description

Computes SQI using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), a multi-criteria decision analysis method. Each group is ranked by its Euclidean distance to the positive ideal solution (all scores = 1) and negative ideal solution (all scores = 0).

$$C_i^* = \frac{d_i^-}{d_i^+ + d_i^-}$$

where d_i^+ and d_i^- are distances to the positive and negative ideal solutions. $C_i^* \in [0, 1]$ with higher values indicating better soil quality.

Usage

```
sqi_topsis(
  scored,
  config,
  group_cols = "LandUse",
  mds_vars = NULL,
  weights = NULL
)
```

Arguments

scored	A scored data frame from score_all.
config	A sqi_config object.
group_cols	Character vector of grouping column names.
mds_vars	Character vector of MDS variable names.
weights	Named numeric vector of criteria weights. Defaults to equal weights.

Value

A data frame with group columns and SQI_topsis (0–1).

References

Hwang, C.L., & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Springer, Berlin. doi:10.1007/9783642483189

Yoon, K. (1987). A reconciliation among discrete compromise solutions. *Journal of the Operational Research Society*, 38, 277–286. doi:10.1057/jors.1987.44

Examples

```
data(soil_data)
cfg <- make_config(
  variable = c("pH","EC","BD","OC","MBC","Clay"),
  type     = c("opt","less","less","more","more","opt"),
  opt_low  = c(6.0, NA, NA, NA, NA, 20),
  opt_high = c(7.0, NA, NA, NA, NA, 35)
)
scored <- score_all(soil_data, cfg, group_cols = c("LandUse","Depth"))
result <- sqi_topsis(scored, cfg, group_cols = c("LandUse","Depth"))
print(result)
```

validate_data *Validate Input Data for SQI Analysis*

Description

Checks that a data frame meets requirements for Soil Quality Index (SQI) computation: correct column types, sufficient sample sizes, absence of infinite values, and appropriate variable configuration.

Usage

```
validate_data(
  data,
  group_cols = NULL,
  config = NULL,
  min_n = 3,
  verbose = TRUE
)
```

Arguments

data	A data frame. The first column(s) should be grouping factors (character or factor); remaining columns should be numeric soil variables.
group_cols	Character vector. Names of grouping columns (e.g., c("LandUse", "Depth")). Defaults to the first column.
config	A data frame produced by make_config or manually created, with columns variable, type, opt_low, opt_high, min_val, max_val. If NULL, only basic data checks are performed.
min_n	Integer. Minimum number of observations per group. Default is 3.
verbose	Logical. If TRUE (default), prints a validation summary to the console.

Value

Invisibly returns a list with components:

valid Logical. TRUE if all checks pass.

messages Character vector of warning/info messages.

n_per_group Data frame of group sizes.

References

Andrews, S.S., Karlen, D.L., & Cambardella, C.A. (2004). The soil management assessment framework: A quantitative soil quality evaluation method. *Soil Science Society of America Journal*, 68(6), 1945–1962. doi:10.2136/sssaj2004.1945

Examples

```
data(soil_data)
result <- validate_data(soil_data, group_cols = c("LandUse", "Depth"))
result$valid
result$n_per_group
```

Index