Package: rmoriedata (via r-universe)

June 23, 2026

Type Package

Title Bundled Datasets for the rmorie Package

Version 0.1.1

Description Bundled open data fixtures used by the 'rmorie' package for examples, vignettes, and tests. Split out so 'rmorie' itself stays within CRAN's package-size soft cap. Contains snapshots of publicly-available datasets from CKAN/Socrata/Opendatasoft portals (Chicago, NYC, Toronto, Vancouver, etc.), Statistics Canada CCJS tables, and synthetic fixtures for unit tests. Also ships a small set of analyst-facing helpers for releasing aggregate statistics without re-identification risk: Laplace and Gaussian differential privacy mechanisms and k-anonymity, l-diversity, and cell suppression verifiers.

License AGPL (>= 3)

Encoding UTF-8

Depends R (>= 4.3.0)

Imports stats

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

URL `https://github.com/rootcoder007/rmoriedata`

BugReports `https://github.com/rootcoder007/rmoriedata/issues`

biocViews Software, ExperimentData

Repository https://rootcoder007.r-universe.dev

Date/Publication 2026-05-27 09:51:18 UTC

RemoteUrl https://github.com/rootcoder007/rmoriedata

RemoteRef main

RemoteSha 45deb38267166ada657e1009509d773b8881a96b

Contents

`morie_cell_suppress` *Cell suppression with optional complementary suppression*

Description

Standard StatCan / open-data complementary-suppression: identifies counts below `threshold`, suppresses them by setting to `NA`, and (if `return_complementary = TRUE`) also suppresses the smallest other count in each affected row and column so the suppressed value can't be reconstructed from marginal sums.

Usage

```
morie_cell_suppress(tbl, threshold = 5, return_complementary = TRUE)
```

Arguments

`tbl`	A numeric matrix or 2-D table of counts. Will be coerced to matrix; row/column names are preserved.
`threshold`	Minimum count to remain unsuppressed. Default 5.
`return_complementary`	Logical; if `TRUE` (default), apply complementary suppression so primary-suppressed cells can't be recovered from marginal sums.

Details

Only finite numeric cells are eligible for suppression. `NA` cells in the input pass through unchanged.

Value

A list with class `"morie_cell_suppress"`:

`suppressed` numeric matrix, suppressed cells set to NA.

`primary_mask` logical matrix, TRUE for primary suppressions.

`complementary_mask` logical matrix, TRUE for complementary suppressions (all FALSE when `return_complementary = FALSE`).

`n_primary` integer.

`n_complementary` integer.

`threshold` the threshold used.

Examples

```
tbl <- matrix(c(120, 3, 47, 88, 2, 99, 14, 51, 60), nrow = 3,
              dimnames = list(c("A","B","C"), c("X","Y","Z")))
morie_cell_suppress(tbl, threshold = 5)
```

`morie_dp_gaussian_mean` *Differentially-private mean via the Gaussian mechanism with bounded inputs*

Description

Releases an approximately (ϵ, δ)-DP mean of a bounded numeric vector. Sensitivity is derived from the user-asserted bounds: changing one record can shift the sum by at most `upper - lower`, so the mean's sensitivity is `(upper - lower) / length(x)`.

Usage

```
morie_dp_gaussian_mean(x, lower, upper, epsilon, delta = 1e-6)
```

Arguments

`x`	Numeric vector (no NAs).
`lower, upper`	Hard bounds on `x`. Caller must guarantee `all(x >= lower & x <= upper)`; the function clips defensively but emits a warning if clipping was necessary.
`epsilon, delta`	Privacy parameters. Standard recommendation: `delta < 1/length(x)`, `epsilon` in 0.1 to 5.0.

Details

The noise standard deviation follows the classical analytic-Gaussian calibration:

$$\sigma = \frac{\Delta \cdot \sqrt{2 \ln(1.25/\delta)}}{\epsilon}.$$

Value

A noised mean (single numeric).

Examples

```
set.seed(1)
x <- runif(1000, 0, 1)
morie_dp_gaussian_mean(x, lower = 0, upper = 1, epsilon = 1.0)
```

`morie_dp_laplace_count`

Differentially-private count via the Laplace mechanism

Description

Adds Laplace noise calibrated to sensitivity / epsilon. Use when releasing counts of records matching some predicate (e.g. number of UoF incidents in a division-year). Sensitivity is hardcoded to 1: one record entering or leaving the dataset changes the count by at most 1.

Usage

```
morie_dp_laplace_count(true_count, epsilon)
```

Arguments

`true_count`	Non-negative integer; the true count.
`epsilon`	Privacy budget (smaller = more noise = stronger privacy). Typical range: 0.1 to 5.0.

Details

Pure (ϵ, 0)-differentially-private under the standard add-or-remove-one neighbouring-databases definition.

Value

A noised count (numeric, may be fractional or negative). Caller should usually clip to a non-negative integer for display: `round(pmax(0, x))`.

Examples

```
set.seed(1)
morie_dp_laplace_count(true_count = 42, epsilon = 1.0)
```

`morie_dp_laplace_histogram`

Differentially-private histogram via the Laplace mechanism

Description

Adds independent Laplace(`1/epsilon`) noise to each bin count. Under the add-or-remove-one neighbouring-databases definition a single record participates in exactly one bin, so the per-bin sensitivity is 1 and the overall mechanism is (ϵ, 0)-DP.

Usage

```
morie_dp_laplace_histogram(counts, epsilon)
```

Arguments

`counts`	Integer vector of non-negative bin counts.
`epsilon`	Privacy budget (positive scalar).

Value

A numeric vector of the same length as `counts`. May contain fractional or negative values. Caller is responsible for any post-hoc non-negativity / rounding before display.

Examples

```
set.seed(1)
morie_dp_laplace_histogram(c(120, 45, 8, 230, 17), epsilon = 0.5)
```

`morie_k_anonymity_verify` *k-anonymity verification*

Description

Checks whether a data.frame satisfies k-anonymity over the supplied quasi-identifier columns. A dataset is k-anonymous if every combination of quasi-identifier values appears in at least `k` rows.

Usage

```
morie_k_anonymity_verify(data, quasi_identifiers, k = 5)
```

Arguments

`data`	data.frame.
`quasi_identifiers`	Character vector of column names.
`k`	Minimum equivalence-class size. Default 5 (a common public-health / open-data threshold).

Value

A list with class `"morie_k_anon"` containing:

`satisfies` logical, whether the dataset is k-anonymous.

`k` the threshold used.

`min_class_size` integer, size of the smallest class.

`n_classes` integer, total number of equivalence classes.

`n_violations` integer, number of classes below the threshold.

`violating_classes` data.frame of class keys plus their `.n` sizes (empty data.frame when none).

`summary` human-readable one-line summary.

Examples

```
df <- data.frame(
  age = c(25, 25, 25, 32, 32, 40),
  sex = c("F", "F", "F", "M", "M", "M")
)
morie_k_anonymity_verify(df, c("age", "sex"), k = 2)
```

`morie_l_diversity_verify`

l-diversity verification

Description

Checks whether a data.frame satisfies l-diversity: within each equivalence class defined by the quasi-identifiers, the sensitive attribute must take at least `l` distinct values.

Usage

```
morie_l_diversity_verify(data, quasi_identifiers, sensitive, l = 3)
```

Arguments

`data`	data.frame.
`quasi_identifiers`	Character vector of QI column names.
`sensitive`	Name of the sensitive-attribute column.
`l`	Minimum number of distinct sensitive values per class. Default 3.

Value

A list with class `"morie_l_div"` containing:

`satisfies` logical.

`l` the threshold used.

`min_diversity` integer, lowest per-class distinct count.

`n_classes` integer.

`n_violations` integer, classes below the threshold.

`violating_classes` data.frame of class keys plus their `.diversity` count.

`summary` human-readable.

Examples

```
df <- data.frame(
  age = c(25, 25, 25, 25, 32, 32, 32),
  sex = c("F", "F", "F", "F", "M", "M", "M"),
  dx  = c("A", "B", "C", "A", "X", "Y", "Z")
)
morie_l_diversity_verify(df, c("age", "sex"), "dx", l = 3)
```

Index