

Package ‘urbstatdata’

September 1, 2026

Type Package

Title Seven Data Sets for Urban and Built-Environment Statistics

Version 0.1.0

Description Seven documented data sets from transport, traffic safety, urban planning, construction and architectural engineering. The package provides fixed, redistributable snapshots with consistent variable names. Each help page records the source, licence, unit of observation, transformations and limitations of its data set. Sources include Yeh (2018) <[doi:10.24432/C5J30W](https://doi.org/10.24432/C5J30W)>, Tsanas and Xifara (2012) <[doi:10.24432/C51307](https://doi.org/10.24432/C51307)>, Yeh (1998) <[doi:10.24432/C5PK67](https://doi.org/10.24432/C5PK67)>, Seoul Bike Sharing Demand (2020) <[doi:10.24432/C5F62R](https://doi.org/10.24432/C5F62R)>, and Singh and Chaudhari (2018) <[doi:10.24432/C5P605](https://doi.org/10.24432/C5P605)>.

URL <https://github.com/mariocools/urbstatdata>

BugReports <https://github.com/mariocools/urbstatdata/issues>

License MIT + file LICENSE

Copyright See inst/COPYRIGHTS.

Depends R (>= 3.5.0)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Encoding UTF-8

LazyData true

LazyDataCompression xz

NeedsCompilation no

Config/roxygen2/version 8.1.0

Author Mario Cools [aut, cre, cph] (ORCID:
<<https://orcid.org/0000-0003-3098-2693>>)

Maintainer Mario Cools <mario.cools@uliege.be>

Repository CRAN

Date/Publication 2026-09-01 08:50:02 UTC

Contents

urbstatdata-package	2
be_accidents	2
be_building_stock	3
building_energy	4
concrete_strength	5
room_occupancy	6
seoul_bikes	7
taipei_housing	8

Index	9
--------------	----------

urbstatdata-package	<i>urbstatdata: Seven Urban and Built-Environment Data Sets</i>
---------------------	---

Description

Provides seven fixed, documented data sets. Each data-set help page records the unit of observation, variables, source, licence, transformations and limitations. The package exports no functions.

be_accidents	<i>Grouped Belgian Road-Accident Victim Counts in 2022</i>
--------------	--

Description

Grouped victim counts by exact combinations of accident date, time, road, victim and administrative characteristics. Codes were preserved, French/Dutch labels retained, and English variable names added.

Usage

be_accidents

Format

A data frame with 82,876 rows and 45 variables. It includes:

date, hour Accident date and hour.

day_of_week_code, day_of_week_fr, day_of_week_nl Day code and French/Dutch labels.

victims_total Total victim count in the grouped cell.

uninjured_victims, slightly_injured, seriously_injured, dead_30_days Outcome counts.

severe_victims Derived count: seriously injured plus dead within 30 days.

nonsevere_victims Derived count: uninjured plus slightly injured.

built_up_area_code, built_up_area_fr, built_up_area_nl Built-up-area code and labels.

victim_type_code, victim_type_fr, victim_type_nl Victim-type code and labels.
road_user_type_code, road_user_type_fr, road_user_type_nl Road-user code and labels.
road_type_code, road_type_fr, road_type_nl Road-type code and labels.
light_condition_code, light_condition_fr, light_condition_nl Light-condition code and labels.
age_group_code, age_group_fr, age_group_nl Age-group code and labels.
municipality_code, municipality_fr, municipality_nl Municipality code and labels.
district_code, district_fr, district_nl Administrative-district code and labels.
province_code, province_fr, province_nl Province code and labels.
region_code, region_fr, region_nl Region code and labels.
sex_code, sex_fr, sex_nl Sex code and labels.

Details

A row is a grouped table cell, not necessarily one victim or one crash. Consequently, binary individual-level logistic regression is not appropriate. The data support contingency tables and grouped-binomial or count analyses when frequencies are used correctly. Outcome components sum to `victims_total`. Administrative codes are character strings to retain leading zeroes.

Source

Statbel and Belgian Federal Police, 2022 road-accident victim table, licensed CC BY 4.0. Stable source snapshot retrieved from the **rWSBIM1207** R package at Git commit 3561033e7a99182664a0a20f603382b5539941f0, <https://github.com/UCLouvain-CBIO/rWSBIM1207/tree/3561033e7a99182664a0a20f603382b5539941f0>.

be_building_stock	<i>Belgian Cadastral Building Stock in 2025</i>
-------------------	---

Description

Counts for nine cadastral building and dwelling characteristics, four building types, Belgium as a whole and the three regions. English labels were added while the French source labels were retained.

Usage

```
be_building_stock
```

Format

A data frame with 144 rows and 10 variables:

characteristic Stable English code for the reported characteristic.
characteristic_fr Original French characteristic label.
country, country_fr English and source-language country labels.

region, region_fr English and source-language region labels; blank source regions were labelled as the Belgium total.

year Reference year (2025).

building_type, building_type_fr English and source-language building-type labels.

count Reported count; missing where the source table does not provide a cell.

Details

Rows are aggregate table cells, not individual buildings. Analyses must avoid individual-level or causal interpretations. Missing counts are source-table omissions, primarily footprint categories for apartment buildings, and must not be recoded as zero.

Source

Statbel, Cadastral statistics of the building stock, official be.STAT API view 43d7cdce-3647-4f5c-86f1-a4e0c864f692, retrieved 10 August 2026. Licensed CC BY 4.0, <https://bestat.statbel.fgov.be/bestat/api/views/43d7cdce-3647-4f5c-86f1-a4e0c864f692/result/CSV>.

building_energy	<i>Simulated Heating and Cooling Loads for Residential Buildings</i>
-----------------	--

Description

Ecotect simulations for twelve residential-building forms under different orientations and glazing arrangements. Variable names and units were standardised; values were retained.

Usage

building_energy

Format

A data frame with 768 rows and 10 variables:

relative_compactness Dimensionless relative compactness.

surface_area_m2, wall_area_m2, roof_area_m2 Areas in square metres.

overall_height_m Overall height in metres.

orientation Coded orientation (2 through 5).

glazing_area_fraction Glazed-area fraction.

glazing_distribution Coded glazing-area distribution (0 through 5).

heating_load_kwh_m2, cooling_load_kwh_m2 Simulated loads in kWh per square metre.

Details

A row is a simulated design configuration, not an observed occupied building. Results should not be generalised beyond the simulation design and assumptions. Treat the orientation and glazing-distribution codes as factors when comparing their levels.

Source

Tsanas, A. and Xifara, A. (2012), Energy Efficiency, UCI Machine Learning Repository, doi:[10.24432/C51307](https://doi.org/10.24432/C51307). Licensed CC BY 4.0.

concrete_strength	<i>Concrete Mixes and Compressive Strength</i>
-------------------	--

Description

Concrete-mixture composition, test age and measured compressive strength. Variable names and units were standardised; values were retained.

Usage

concrete_strength

Format

A data frame with 1,030 rows and 9 variables:

cement_kg_m3 Cement content in kg per cubic metre.

blast_furnace_slag_kg_m3 Blast-furnace slag in kg per cubic metre.

fly_ash_kg_m3 Fly ash in kg per cubic metre.

water_kg_m3 Water in kg per cubic metre.

superplasticizer_kg_m3 Superplasticizer in kg per cubic metre.

coarse_aggregate_kg_m3 Coarse aggregate in kg per cubic metre.

fine_aggregate_kg_m3 Fine aggregate in kg per cubic metre.

age_days Age at testing, in days.

strength_mpa Compressive strength in megapascals.

Details

A row is one tested mix. Relationships are strongly nonlinear, so a single untransformed linear age effect is mainly useful as a diagnostic starting point. A comparison with a benchmark is not a certification or structural-safety decision.

Source

Yeh, I.-C. (1998), Concrete Compressive Strength, UCI Machine Learning Repository, doi:[10.24432/C5PK67](https://doi.org/10.24432/C5PK67). Licensed CC BY 4.0.

room_occupancy

Room-Sensor Measurements and Occupancy

Description

Time-ordered measurements from a controlled room-sensor experiment, paired with the observed number of occupants.

Usage

room_occupancy

Format

A data frame with 10,129 rows and 20 variables:

date, time Measurement date (ISO string) and clock time.

s1_temp_c, s2_temp_c, s3_temp_c, s4_temp_c Four temperature-sensor readings in degrees Celsius.

s1_light_lux, s2_light_lux, s3_light_lux, s4_light_lux Four light-sensor readings.

s1_sound_v, s2_sound_v, s3_sound_v, s4_sound_v Four sound-sensor readings.

s5_co2_ppm CO2 concentration in parts per million.

s5_co2_slope Estimated CO2 slope.

s6_pir, s7_pir Passive-infrared sensor indicators.

occupancy_count Observed number of room occupants, 0 through 3.

occupied Derived indicator: 1 when occupancy_count > 0, otherwise 0.

Details

A row is a measurement time, not an independent sampled room. The derived occupied field is provided as a convenient binary outcome. Serial dependence, sensor placement and the controlled setting limit ordinary independent-observation inference and external generalisation.

Source

Singh, A. P. and Chaudhari, S. (2018), Room Occupancy Estimation, UCI Machine Learning Repository, [doi:10.24432/C5P605](https://doi.org/10.24432/C5P605). Licensed CC BY 4.0.

seoul_bikes*Hourly Seoul Bicycle-Rental Demand*

Description

One year of hourly public bicycle-rental demand in Seoul with weather and calendar predictors. Dates were converted to ISO strings and variable names and units were standardised.

Usage

seoul_bikes

Format

A data frame with 8,760 rows and 14 variables:

date Calendar date as an ISO YYYY-MM-DD string.

rented_bikes Number of bicycles rented during the hour.

hour Hour of day, 0 through 23.

temperature_c, dew_point_c Temperatures in degrees Celsius.

humidity_pct Relative humidity in percent.

wind_speed_m_s Wind speed in metres per second.

visibility_10m Visibility measured in 10-metre units.

solar_radiation_mj_m2 Solar radiation in MJ per square metre.

rainfall_mm, snowfall_cm Rainfall in millimetres and snowfall in centimetres.

season Season label.

holiday Holiday-status label.

functioning_day Whether the rental system was functioning.

Details

Rows form a time series and are not independent random observations. Non-functioning periods can generate structural zeroes. Count analyses must therefore consider overdispersion, temporal dependence and service status.

Source

Seoul Bike Sharing Demand (2020), UCI Machine Learning Repository, [doi:10.24432/C5F62R](https://doi.org/10.24432/C5F62R). Licensed CC BY 4.0.

taipei_housing	<i>New Taipei City Real-Estate Valuation Data</i>
----------------	---

Description

Market-historical records collected from the Sindian District of New Taipei City, Taiwan. Variable names were standardised and units were made explicit; values were otherwise retained.

Usage

taipei_housing

Format

A data frame with 414 rows and 8 variables:

transaction_id Source row identifier.

transaction_year Transaction year expressed as a decimal year.

house_age_years Age of the house, in years.

mrt_distance_m Distance to the nearest MRT station, in metres.

nearby_stores Number of convenience stores in the living circle.

latitude, longitude WGS84 coordinates.

price_10000_twd_per_ping Unit-area price in 10,000 New Taiwan dollars per ping.

Details

A row is a transaction. The observational data support descriptive, associational and predictive analyses, but do not identify causal effects of MRT proximity or nearby services. A ping is approximately 3.306 square metres.

Source

Yeh, I.-C. (2018), Real Estate Valuation, UCI Machine Learning Repository, [doi:10.24432/C5J30W](https://doi.org/10.24432/C5J30W).
Licensed CC BY 4.0.

Index

* datasets

- be_accidents, [2](#)
- be_building_stock, [3](#)
- building_energy, [4](#)
- concrete_strength, [5](#)
- room_occupancy, [6](#)
- seoul_bikes, [7](#)
- taipei_housing, [8](#)
- urbstatdata-package, [2](#)

be_accidents, [2](#)

be_building_stock, [3](#)

building_energy, [4](#)

concrete_strength, [5](#)

room_occupancy, [6](#)

seoul_bikes, [7](#)

taipei_housing, [8](#)

urbstatdata (urbstatdata-package), [2](#)

urbstatdata-package, [2](#)