Package 'pksR'

November 21, 2026

Title District-Level German Police Crime Statistics

Version 0.1.0

Description Access to the district (Kreis) tables of the German Police Crime Statistics ('Polizeiliche Kriminalstatistik', PKS) published by the Bundeskriminalamt, from 2003 onwards. Data come from versioned, validated releases of the 'pksR-data' pipeline, either as published or harmonized to a common territorial boundary and offence classification. Downloads are cached and checked against SHA-256 checksums; rates can be recomputed after harmonization, panels validated, and data releases cited.

License MIT + file LICENSE

URL https://github.com/ofurkancoban/pksR, https://ofurkancoban.github.io/pksR/, https://github.com/ofurkancoban/pksR-data

BugReports https://github.com/ofurkancoban/pksR/issues

Depends R (>= 4.1.0)

Imports cli, digest, dplyr, httr2, jsonlite, methods, readr, rlang, stats, tibble, tools, utils

Suggests knitr, rmarkdown, testthat (>= 3.0.0), withr

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-GB

RoxygenNote 7.3.3

NeedsCompilation no

Author Ömer Furkan Çoban [aut, cre, cph] (ORCID: <https://orcid.org/0009-0005-3623-7178>)

Maintainer Ömer Furkan Çoban <oemer.furkan.coban@uni-oldenburg.de>

Repository CRAN

Date/Publication 2026-11-21 19:20:02 UTC

Contents

pks_breaks . . . 2
pks_cache_clear . . . 3
pks_cache_info . . . 4
pks_cite . . . 4
pks_compare_weights . . . 5
pks_example . . . 6
pks_get . . . 7
pks_manifest . . . 9
pks_offences . . . 10
pks_rates . . . 11
pks_validate . . . 12
pks_versions . . . 13

Index **15**

`pks_breaks` *Known methodological breaks*

Description

Changes in the PKS that make values before and after a year incomparable or only partly comparable: new offence definitions, key changes, changes in the population denominator, boundary reforms and others. Check these before interpreting changes over time. Real shocks such as the Covid-19 pandemic (2020 and 2021) are not breaks: they change recorded crime, not how it is measured.

Usage

```
pks_breaks()
```

Details

The table is a snapshot of `known_breaks.csv` of pksR-data that ships with the package; each data release contains its own copy.

Value

A tibble with `break_id`, `first_year`, `offence_key` (keys separated by `;`, empty when all keys are affected), `variable`, `type`, `comparable` (`yes`, `likely`, `unknown`, `no`), `description`, `evidence` and `status`.

Deutsch

Bekannte methodische Brüche der PKS: neue Definitionen, geänderte Schlüssel, geänderte Bezugsbevölkerung, Gebietsreformen und andere. `comparable` gibt an, ob Werte vor und nach dem Jahr vergleichbar sind (`yes`, `likely`, `unknown`, `no`). Vor jeder Zeitreihe prüfen. Echte Schocks wie die Covid-19-Pandemie (2020 und 2021) sind keine Brüche: Sie verändern die erfasste Kriminalität, nicht ihre Messung.

See Also

Other metadata: pks_offences()

Examples

```
b <- pks_breaks()
b[b$comparable == "no", c("first_year", "offence_key", "description")]
```

---

pks_cache_clear *Clear the local cache*

Description

Deletes cached release files, for all data versions or only for the ones given. The files are downloaded again when they are next needed.

Usage

```
pks_cache_clear(version = NULL, ask = interactive())
```

Arguments

version	Character vector of data versions to remove, such as "2026.1", or NULL (default) for the whole cache.
ask	Ask for confirmation before deleting. Defaults to TRUE in interactive sessions.

Value

The removed files as a tibble like pks_cache_info(), invisibly.

Deutsch

Löscht Dateien aus dem Cache, für alle oder für die angegebenen Datenversionen. In interaktiven Sitzungen wird vorher nachgefragt.

See Also

Other cache: pks_cache_info()

Examples

```
# A temporary cache, so the example leaves your own cache alone
old <- Sys.getenv("R_USER_CACHE_DIR", NA)
Sys.setenv(R_USER_CACHE_DIR = tempfile())
pks_cache_clear(ask = FALSE)
if (is.na(old)) Sys.unsetenv("R_USER_CACHE_DIR") else Sys.setenv(R_USER_CACHE_DIR = old)
```

pks_cache_info *Inspect the local cache*

Description

Lists the release files that `pks_get()` and `pks_manifest()` have stored in the cache directory, `tools::R_user_dir("pksR", "cache")`.

Usage

```
pks_cache_info()
```

Value

A tibble with one row per cached file: `version`, `file`, `bytes` and `modified`. The cache directory is attached as attribute `path`.

Deutsch

Listet die im Cache (`tools::R_user_dir("pksR", "cache")`) gespeicherten Dateien, je Datenversion.

See Also

Other cache: `pks_cache_clear()`

Examples

```
pks_cache_info()
```

pks_cite *Cite pksR, the data release and the data sources*

Description

Work with these data should cite the package, the data release used (each release is versioned and never changes) and the original sources. The source data keep their terms of reuse: the PKS tables of the Bundeskriminalamt and the BBSR crosswalk are licensed under the Data licence Germany, attribution, version 2.0 (dl-de/by-2-0, https://www.govdata.de/dl-de/by-2-0), which requires naming the source and stating that the data were changed; the Destatis municipality directory may be reproduced with a source note.

Usage

```
pks_cite(version = "latest")
```

Arguments

`version`	Data version such as `"2026.1"`, or `"latest"` (default). Without internet access, only the package and the sources are cited.

Details

Each data release is archived on Zenodo. The data release entry carries the DOI of its version, read from the release notes; until Zenodo has archived a new version, it carries the concept DOI, which resolves to the latest version.

Value

A `utils::bibentry()` object with one entry each for the package, the data release, the PKS, the BBSR crosswalk and the Destatis municipality directory. Print it, or convert it with `format(x, "bibtex")`.

Deutsch

Zitierangaben für das Paket, die verwendete Datenveröffentlichung und die Quellen: die PKS des Bundeskriminalamts und der Umsteigeschlüssel des BBSR (beide Datenlizenz Deutschland, Namensnennung, Version 2.0, die die Nennung der Quelle und einen Hinweis auf Veränderungen verlangt) sowie das Gemeindeverzeichnis von Destatis. Jede Datenveröffentlichung ist auf Zenodo archiviert; der Eintrag trägt die DOI ihrer Version aus den Release-Notizen oder, solange Zenodo sie noch nicht archiviert hat, die Konzept-DOI. Ergebnis: ein `utils::bibentry()`-Objekt; `format(x, "bibtex")` liefert BibTeX.

See Also

Other releases: `pks_manifest()`, `pks_versions()`

Examples

```
pks_cite()
format(pks_cite(), "bibtex")
```

`pks_compare_weights` *Compare population and area weights*

Description

Districts that were split had to be divided between their successors, by population (default) or by area. `pks_compare_weights()` puts both versions of a harmonized panel side by side, so you can see where and how much a result depends on that choice. Merged and unchanged districts are the same under both weights.

Usage

```
pks_compare_weights(data, other = NULL)
```

Arguments

data	A harmonized tibble from pks_get() (not harmonize_to = NULL), with population or area weights.
other	NULL (default) to download the same years, offences, boundary and data version with the other weights, or that panel itself, for example when you have it already or work offline.

Value

A tibble with one row per year, district and offence key: year, ags, region_name, offence_key, estimated (split in at least one of the two panels), cases_population, cases_area, difference (area minus population), relative_difference (difference / cases_population), hz_population and hz_area (recomputed with pks_rates()). A message summarizes how many rows differ.

Deutsch

Stellt eine harmonisierte Tabelle mit Bevölkerungs- und mit Flächengewichten nebeneinander. Die Gewichte unterscheiden sich nur bei geteilten Kreisen; so lässt sich prüfen, wie stark ein Ergebnis von dieser Annahme abhängt. other kann die Tabelle mit den anderen Gewichten liefern, sonst wird sie mit denselben Jahren, Straftaten, demselben Gebietsstand und derselben Datenversion geladen.

See Also

Other analysis: pks_rates(), pks_validate()

Examples

```
d <- pks_get(years = 2005:2010, offence = "total")
w <- pks_compare_weights(d)
w[abs(w$relative_difference) > 0.01, ]
```

---

pks_example *A small example panel*

Description

A subset of the harmonized panel of pksR-data that ships with the package, for examples, tests and vignettes without internet access: five districts, two offence keys and the years 2019 to 2025, harmonized to the 2025 boundary with population weights, built from pksR-data commit 9b6985d (data-raw/example.R), not from a data release, so data_version is "example". Source: PKS Bundeskriminalamt, Datenlizenz Deutschland, Namensnennung, Version 2.0; changed (harmonized) by pksR-data.

Usage

```
pks_example()
```

Value

A tibble in the format of `pks_get()`.

Deutsch

Ein kleines Beispielpanel, das mit dem Paket ausgeliefert wird: fünf Kreise, zwei Straftatenschlüssel und die Jahre 2019 bis 2025, harmonisiert auf den Gebietsstand 2025. Es stammt aus einem lokalen Lauf von pksR-data, nicht aus einer Datenveröffentlichung (`data_version = "example"`). Quelle: PKS Bundeskriminalamt, Datenlizenz Deutschland, Namensnennung, Version 2.0; verändert (harmonisiert).

See Also

Other data: `pks_get()`

Examples

```
pks_example()
```

---

pks_get *Get district-level police crime statistics*

Description

Downloads the district (Kreis) tables of the German Police Crime Statistics (PKS) from a versioned data release, as published by the Bundeskriminalamt or harmonized to one territorial boundary and offence classification. Files are cached and checked against the SHA-256 checksums of the release manifest.

Usage

```
pks_get(
  years = NULL,
  offence = NULL,
  harmonize_to = "latest",
  weights = c("population", "area"),
  version = "latest",
  source = c("release", "bka"),
  cache = TRUE,
  lang = NULL
)
```

Arguments

years	Integer vector of reporting years, or NULL (default) for all years of the release.
offence	Offences to return, or NULL (default) for all. Each element can be an offence key ("435*00"), a short name in English or German ("residential_burglary", "wohnungseinbruch"), a full label ("Wohnungseinbruchdiebstahl") or a part of a label ("einbruch", which selects both burglary keys). An exact key, name or label wins over a part. Capitals, spaces and umlauts do not matter. See pks_offences() for all keys and names. For the data as published, the published keys that feed the selected keys are returned.
harmonize_to	Target boundary year (Gebietsstand). "latest" (default) is the boundary of the release, read as released; another year harmonizes the data as published with the crosswalk of the release (forward only, so years must not be later than the target); NULL returns the data as published.
weights	Weights for districts that were split: "population" (default) or "area". Ignored for harmonize_to = NULL.
version	Data version such as "2026.1", or "latest" (default).
source	"release" (default) reads a data release. "bka", which will rebuild the spreadsheet years from the original BKA files, is not available yet.
cache	Keep downloaded files in the cache (default TRUE). With FALSE, files go to a temporary directory and are removed with the R session.
lang	Language of the offence labels in column offence_label and of the messages: "de" or "en". NULL (default) uses options(pksR.lang), otherwise the language of the R session. Harmonized data get the label of the harmonized key; data as published keep the published German label, or get the English label of the matching harmonized key.

Value

A tibble with one row per year, district (ags) and offence key. Columns are described in vignette("pksR"). Attributes: data_version, retrieved_at, boundary_year, weights and source_urls.

Deutsch

Lädt die Kreistabellen der Polizeilichen Kriminalstatistik (PKS) aus einer versionierten Datenveröffentlichung, wie vom Bundeskriminalamt veröffentlicht oder auf einen Gebietsstand und eine Straftatenzuordnung harmonisiert. Jede Datei wird gegen die SHA-256-Prüfsumme des Manifests geprüft und im Cache abgelegt.

- years: Berichtsjahre, NULL für alle.
- offence: Straftaten als Schlüssel ("435*00"), Kurzname auf Deutsch oder Englisch ("wohnungseinbruch", "residential_burglary"), vollständige Bezeichnung oder Teil einer Bezeichnung ("einbruch"). Groß- und Kleinschreibung, Leerzeichen und Umlaute spielen keine Rolle; eine genaue Übereinstimmung hat Vorrang. Übersicht: pks_offences().
- harmonize_to: "latest" (Gebietsstand der Veröffentlichung), ein Jahr (Harmonisierung nur vorwärts) oder NULL (wie veröffentlicht).
- weights: Gewichtung geteilter Kreise, "population" oder "area".

- `version`: Datenversion, etwa `"2026.1"`, oder `"latest"`.
- `lang`: Sprache der Bezeichnungen und Meldungen, `"de"` oder `"en"`; Voreinstellung `options(pksR.lang)`, sonst die Sprache der R-Sitzung.

Ergebnis: ein Tibble mit einer Zeile je Jahr, Kreis und Straftatenschlüssel und den Attributen `data_version`, `retrieved_at`, `boundary_year`, `weights` und `source_urls`.

See Also

Other data: `pks_example()`

Examples

```
# Total offences, harmonized to the latest boundary
total <- pks_get(offence = "total")

# Theft in total and both burglary keys, German labels
theft <- pks_get(offence = c("diebstahl", "einbruch"), lang = "de")

# As published by the BKA, two years
published <- pks_get(years = 2024:2025, harmonize_to = NULL)
```

`pks_manifest` *Read the manifest of a data release*

Description

The manifest describes one data release: when and from which pipeline commit it was built, the years and target boundary year it covers, the validation result, and every file with its SHA-256, size and number of rows.

Usage

```
pks_manifest(version = "latest")
```

Arguments

`version`	Data version such as `"2026.1"`, or `"latest"`.

Value

A tibble with one row per release file (`name`, `sha256`, `bytes`, `rows`). The other manifest fields (`data_version`, `created_at`, `pipeline_commit`, `target_boundary_year`, `years`, `validation`, `notes`, `license`, `attribution`, ...) are attached as attribute `manifest`.

Deutsch

Liest das Manifest einer Datenveröffentlichung: wann und mit welchem Stand der Pipeline sie erstellt wurde, welche Jahre und welchen Gebietsstand sie abdeckt, das Ergebnis der Prüfung und jede Datei mit Prüfsumme, Größe und Zeilenzahl. Ergebnis: ein Tibble der Dateien, die übrigen Angaben im Attribut `manifest`.

See Also

Other releases: `pks_cite()`, `pks_versions()`

Examples

```
m <- pks_manifest()
attr(m, "manifest")$target_boundary_year
```

---

pks_offences *List the harmonized offence keys*

Description

The district tables contain a selected subset of offence keys (Straftatenschluessel). The harmonized keys used by `pks_get()` are listed here with a short name and a label in German or English, their parent key and the years they cover. Mappings and names marked `status = "draft"` have not been reviewed yet.

Usage

```
pks_offences(pattern = NULL, lang = NULL)
```

Arguments

`pattern`	Optional text to search for in the keys, short names and labels of both languages. Capitals, spaces and umlauts do not matter, so `"Diebstahl"`, `"theft"` and `"435*00"` all work.
`lang`	Language of the `name` and `label` columns: `"de"` or `"en"`. `NULL` (default) uses `options(pksR.lang)`, otherwise the language of the R session.

Details

The table is a snapshot of the offence map of pksR-data that ships with the package, so it works offline. Each data release also contains its own `offence_map.csv`, which `pks_get()` uses.

Value

A tibble with `offence_key`, `name` (the short name accepted by `pks_get(offence = )`), `label`, `parent_key`, `valid_from`, `valid_to`, `status` and `notes`.

Deutsch

Listet die harmonisierten Straftatenschlüssel mit Kurzname, Bezeichnung auf Deutsch oder Englisch, übergeordnetem Schlüssel und Gültigkeitsjahren. `pattern` durchsucht Schlüssel, Kurznamen und Bezeichnungen beider Sprachen; `lang` wählt die Sprache der Spalten `name` und `label`. Zuordnungen mit `status = "draft"` sind noch nicht geprüft.

See Also

Other metadata: `pks_breaks()`

Examples

```
pks_offences("Diebstahl", lang = "de")
pks_offences("theft", lang = "en")
```

---

pks_rates *Recompute rates*

Description

Rates cannot be harmonized: when districts are merged or split, the published frequency (Haeufigkeitszahl, cases per 100,000 inhabitants) and clearance rate of the old districts do not carry over. `pks_rates()` recomputes them from the counts, with the population of the data or one you supply.

Usage

```
pks_rates(data, population = NULL, per = 1e+05)
```

Arguments

`data`	A tibble from `pks_get()`.
`population`	`NULL` (default) to use the population behind the published HZ (published where available, otherwise derived from cases and HZ); `"corrected"` for the census-corrected population in column `population_corrected`; or a data frame with columns `year`, `ags` and `population`, for example from an official population source. Every row of `data` must find a population.
`per`	Rate denominator for column `rate` (default 100,000).

Details

Cases come from `cases`, or from `cases_derived` in the PDF years, where the BKA printed only rates. The clearance rate is 0 where cases are 0, as in the BKA tables.

The BKA divides by the population of 31 December of the year before, on the census base of its time: older registers until 2012, Zensus 2011 for 2013 to 2023, Zensus 2022 from 2024. Rates therefore jump in 2013 and 2024 without any change in crime (see `pks_breaks()`). With `population = "corrected"` every year uses the census-corrected population of the BBSR (INKAR), at the same date and on the same boundary, so rates are on one base over the whole period.

Value

`data` with `hz` (cases per 100,000) and `clearance_rate` recomputed, and a new column `rate` (cases per `per`). The population used is in column `rate_population`.

Deutsch

Berechnet Häufigkeitszahl (Fälle je 100.000 Einwohner) und Aufklärungsquote neu aus den Fallzahlen, etwa nach einer Harmonisierung, bei der veröffentlichte Quoten nicht übertragen werden können. `population` kann eine eigene Einwohnerzahl je Jahr und Kreis liefern; `per` setzt den Nenner der neuen Spalte `rate`. Das BKA teilt durch die Einwohnerzahl zum 31.12. des Vorjahres auf der Zensusbasis ihrer Zeit, daher springen die Quoten 2013 und 2024. `population = "corrected"` nutzt für alle Jahre die zensuskorrigierte Einwohnerzahl des BBSR (INKAR) und macht die Quoten über den ganzen Zeitraum vergleichbar.

See Also

Other analysis: `pks_compare_weights()`, `pks_validate()`

Examples

```
d <- pks_example()
pks_rates(d, per = 1000)[, c("year", "ags", "cases", "hz", "rate")]

# One census base for all years
pks_rates(d, population = "corrected")[, c("year", "ags", "hz", "rate_population")]
```

`pks_validate` *Validate a panel*

Description

Runs the consistency checks of the pksR-data validation gate that need only the panel itself: schema, uniqueness, coverage and plausibility. Use it after filtering, joining or editing data from `pks_get()`. The checks of national totals and conservation need the national tables and run in the pipeline before every release.

Usage

```
pks_validate(data)
```

Arguments

`data`	A tibble from `pks_get()`.

Details

Checks:

- `schema`: the columns of the data as published or harmonized, with their types.
- `uniqueness`: one row per `year`, `ags` and `offence_key`.
- `city_states`: Berlin (`11000`) and Hamburg (`02000`) present in every year.
- `coverage`: harmonized data have the same districts in every year (balanced panel).
- `counts`: no negative counts, and `attempts` not above `cases`.
- `population`: the median district population per year lies between 50,000 and 400,000 (a warning otherwise).

Value

A tibble with one row per check: `check`, `status` (`pass`, `warn` or `fail`), `problems` (number of offending rows or years) and `details`.

Deutsch

Prüft ein Panel auf Schema, Eindeutigkeit (eine Zeile je Jahr, Kreis und Schlüssel), Vorhandensein von Berlin und Hamburg in jedem Jahr, gleichen Kreisbestand in harmonisierten Daten, keine negativen Zahlen, nicht mehr Versuche als Fälle und plausible Einwohnerzahlen. Ergebnis: ein Tibble mit einer Zeile je Prüfung und dem Status `pass`, `warn` oder `fail`.

See Also

Other analysis: `pks_compare_weights()`, `pks_rates()`

Examples

```
pks_validate(pks_example())
```

`pks_versions` *List the available data releases*

Description

Data releases are versioned and never change; a correction is published as a new version. Versions are named `<year>.<n>`, for example `"2026.1"`.

Usage

```
pks_versions()
```

Value

A tibble with one row per data release, newest first: `version`, `tag`, `published_at`, `files`, `url` and `doi` (Zenodo DOI of the version, `NA` until Zenodo has archived it).

Deutsch

Listet die veröffentlichten Datenversionen, die neueste zuerst. Eine Veröffentlichung ändert sich nie; Korrekturen erscheinen als neue Version. Die Spalte `doi` enthält die Zenodo-DOI der Version.

See Also

Other releases: `pks_cite()`, `pks_manifest()`

Examples

```
pks_versions()
```

Index

∗ **analysis**
  pks_compare_weights, 5
  pks_rates, 11
  pks_validate, 12
∗ **cache**
  pks_cache_clear, 3
  pks_cache_info, 4
∗ **data**
  pks_example, 6
  pks_get, 7
∗ **metadata**
  pks_breaks, 2
  pks_offences, 10
∗ **releases**
  pks_cite, 4
  pks_manifest, 9
  pks_versions, 13

pks_breaks, 2, *11*
pks_breaks(), *11*
pks_cache_clear, 3, *4*
pks_cache_info, *3*, 4
pks_cache_info(), *3*
pks_cite, 4, *10*, *14*
pks_compare_weights, 5, *12, 13*
pks_example, 6, *9*
pks_get, *7*, 7
pks_get(), *4*, *6, 7*, *10–12*
pks_manifest, *5*, 9, *14*
pks_manifest(), *4*
pks_offences, *3*, 10
pks_offences(), *8*
pks_rates, *6*, 11, *13*
pks_rates(), *6*
pks_validate, *6*, *12*, 12
pks_versions, *5*, *10*, 13

utils::bibentry(), *5*