# Package: cleaclean (via r-universe)

November 2, 2026

**Title** Clean and Classify Constituency-Level Election Results

**Version** 0.1.0

**Description** Provides cleaned constituency-level election results derived from the Constituency-Level Elections Archive and an analysis-ready subset of elections conducted under simple electoral systems. The repository also contains the auditable maintainer workflow used to construct the datasets.

**License** MIT + file LICENSE

**URL** https://github.com/jackobailey/clean_clea

**BugReports** https://github.com/jackobailey/clean_clea/issues

**Depends** R (>= 4.1)

**Imports** dplyr, tibble, tidyr

**Suggests** countrycode, digest, pkgload, readr, stringr, testthat (>= 3.0.0), vdemdata

**Remotes** vdeminstitute/vdemdata

**Config/testthat/edition** 3

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.3

**Config/pak/sysreqs** libicu-dev

**Repository** https://jackobailey.r-universe.dev

**Date/Publication** 2026-06-04 18:34:54 UTC

**RemoteUrl** https://github.com/jackobailey/clean_clea

**RemoteRef** HEAD

**RemoteSha** 11896c2e6eeb04260301a353addcd643b5bece6b

# Contents

clean_clea . . . . . 2
simple_systems . . . . . 4

**Index** **6**

---

`clean_clea` *Cleaned Constituency-Level Elections Archive*

---

## Description

A cleaned version of Release 18 of the Constituency-Level Elections Archive lower-chamber dataset. Election-specific corrections are documented in the package repository under `data-raw/corrections/`.

## Usage

```
clean_clea
```

## Format

A tibble with 1,296,813 rows and 33 variables:

**release** CLEA release identifier.

**id** CLEA election identifier.

**rg** CLEA region code.

**ctr_n** Country or territory name.

**ctr** Country or territory numeric code.

**yr** Election year.

**mn** Election month.

**sub** Subnational election identifier, where applicable.

**cst_n** Constituency name.

**cst** Constituency identifier.

**mag** Constituency district magnitude.

**pty_n** Party name.

**pty** Party identifier.

**can** Candidate name.

**pev1** First-election-period eligible voters.

**vot1** First-election-period voters.

**vv1** First-election-period valid votes.

**ivv1** First-election-period invalid votes.

**to1** First-election-period turnout.

**cv1** First-election-period candidate votes.

**cvs1** First-election-period candidate vote share.

**pv1** First-election-period party votes.

**pvs1** First-election-period party vote share.

**pev2** Second-election-period eligible voters.

**vot2** Second-election-period voters.

**vv2** Second-election-period valid votes.

**ivv2** Second-election-period invalid votes.

**to2** Second-election-period turnout.

**cv2** Second-election-period candidate votes.

**cvs2** Second-election-period candidate vote share.

**pv2** Second-election-period party votes.

**pvs2** Second-election-period party vote share.

**seat** Seats won.

## Details

The maintainer pipeline verifies the pinned source checksum before applying 610 election-specific correction scripts and tracked manual patches. These corrections can replace values, restructure results, add documented missing rows, or remove elections and rows whose source data cannot be reconciled.

## Known limitations

This dataset preserves CLEA sentinel values and 888 duplicate rows where no audited correction has been made. It inherits the coverage and coding limitations of CLEA and the cited correction sources.

## Source

CLEA Lower Chamber Elections Archive, Release 18 (October 15, 2025)

## References

Kollman, K., Hicken, A., Caramani, D., Backer, D. A., and Lublin, D. Constituency-Level Elections Archive.

`simple_systems` *Elections Conducted Under Simple Electoral Systems*

## Description

An analysis-ready subset of clean_clea containing elections classified as simple electoral systems using V-Dem classifications and documented manual corrections. Vote and seat shares and district-level summary statistics are included.

## Usage

```
simple_systems
```

## Format

A tibble with 280,569 rows and 31 variables:

**id** CLEA election identifier.

**iso3** Three-letter country or territory code.

**country** Country or territory name.

**subregion** United Nations geographic subregion.

**region** United Nations geographic region.

**yr** Election year.

**mn** Election month.

**cst_n** Constituency name.

**cst** Constituency identifier.

**m** District magnitude.

**pty_n** Party name.

**pty** Party identifier.

**can** Candidate name.

**c** Candidate votes in the source data.

**p** Party votes in the source data.

**v** Vote count used for analysis.

**s** Seats won.

**pv** District vote share.

**ps** District seat share.

**uncontested** Whether the source result was uncontested.

**electoral_system** V-Dem electoral-system classification.

**threshold** Whether a legal electoral threshold is recorded.

**nv0** Actual number of vote-winning parties or candidates.

**ns0** Actual number of seat-winning parties or candidates.

**nv2** Effective number of vote-winning parties or candidates.

**ns2** Effective number of seat-winning parties or candidates.

**d** Gallagher disproportionality index.

**w** Vote share cast for parties or candidates winning no seats.

**tx** Threshold of exclusion, calculated as `1 / (m + 1)`.

**tr** Threshold of representation, calculated as `1 / (m * nv2)`.

**tmin** Minimum of `tx` and `tr`.

## Details

The classifier combines V-Dem with tracked manual classifications. `simple_systems` excludes elections before 1900, systems not classified as simple, the United States and Panama, and elections that retain missing votes, missing seats, aggregate-party codes, or inconsistent vote and seat rankings. Uncontested results are assigned one analytical vote before shares and summary statistics are calculated.

## Known limitations

The data inherit the coverage and coding limitations of CLEA, V-Dem, and the tracked manual sources. Legal thresholds and complex allocation rules may not be fully represented by district-level summary statistics.

## Source

CLEA Lower Chamber Elections Archive and V-Dem.

# Index

∗ **datasets**
clean_clea, 2
simple_systems, 4

clean_clea, 2, *4*

simple_systems, 4