Package: assemblykor (via r-universe)

October 29, 2026

Title Korean National Assembly Data for Political Science Education

Version 0.1.4

Description Provides ready-to-use datasets from the Korean National Assembly (assemblies 20 through 22, 2016-2026) for teaching quantitative methods in political science. Includes legislator metadata, bill proposals, roll call votes, asset declarations, and policy seminar records. Designed as a Korean politics counterpart to packages like 'palmerpenguins', enabling students to practice regression, panel data analysis, text analysis, and network analysis with real legislative data. Roll call vote data and spatial voting models are described in Poole and Rosenthal (1985) <doi:10.2307/2111172>. Legislative data is sourced from the Korean National Assembly Open API.

License MIT + file LICENSE

URL https://kyusik-yang.github.io/assemblykor/, https://github.com/kyusik-yang/assemblykor

BugReports https://github.com/kyusik-yang/assemblykor/issues

Depends R (>= 3.5.0)

Imports utils

Suggests arrow, broom, dplyr, fixest, ggplot2, htmltools, igraph, knitr, learnr, pkgdown, rmarkdown, scales, stringr, systemfonts, testthat (>= 3.0.0), tidyr, tidytext

Config/testthat/edition 3

Encoding UTF-8

LazyData true

LazyDataCompression xz

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

VignetteBuilder knitr

Repository https://kyusik-yang.r-universe.dev

Date/Publication 2026-09-29 01:48:12 UTC

RemoteUrl https://github.com/kyusik-yang/assemblykor

RemoteRef HEAD

RemoteSha 5ab934e778b6a11e7fc1b7abcb0b42bec217cb1a

Contents

assemblykor-package 2
bills 3
get_bill_texts 5
get_proposers 6
get_speech_tokens 7
legislators 9
list_tutorials 10
open_tutorial 11
path_to_file 12
roll_calls 12
run_tutorial 14
seminars 15
set_ko_font 16
speeches 17
votes 19
wealth 20

Index **22**

`assemblykor-package` *assemblykor: Korean National Assembly Data for Political Science Education*

Description

Provides ready-to-use datasets from the Korean National Assembly for teaching quantitative methods in political science. Includes seven built-in datasets covering legislator metadata, bills, asset declarations, policy seminars, committee speeches, plenary vote tallies, and member-level roll call votes.

Built-in datasets

- `legislators`: 963 MP records (20th-22nd assemblies)
- `bills`: 64,900 legislative bills
- `wealth`: 3,215 legislator-year asset declarations
- `seminars`: 5,962 legislator-year seminar records
- `speeches`: 15,795 speech records (22nd, Science & ICT Committee)
- `votes`: 8,611 plenary vote tallies (20th-22nd assemblies)
- `roll_calls`: 549,513 member-level roll call votes (22nd assembly)

Download functions

- get_bill_texts: 64,900 bill propose-reason texts
- get_proposers: 825,283 co-sponsorship records
- get_speech_tokens: 663,582 morpheme tokens for the speeches dataset (Kiwi morphological analyzer)

Data corrections and coverage

Version 0.1.4 corrects defects that releases 0.7.0 to 0.8.1 of the kna project (https://github.com/kyusik-yang/kna) found in the underlying Open Assembly records, among them the seniority and committees of legislators, the results of vetoed bills, the co-sponsorship records cut at 100 names per bill and the votes the API omits. It also extends the 22nd assembly to 2026-09-23 and the asset declarations to 2025, from kna 0.8.1. See the package NEWS for details.

Tutorials

Nine Korean-language tutorials covering tidyverse, visualization, regression, panel data, text analysis, network analysis, roll call analysis, bill success, and speech patterns. Use list_tutorials to see all tutorials, and open_tutorial to copy them to your working directory.

Author(s)

Maintainer: Kyusik Yang <kyusik.yang@nyu.edu>

See Also

Useful links:

- https://kyusik-yang.github.io/assemblykor/
- https://github.com/kyusik-yang/assemblykor
- Report bugs at https://github.com/kyusik-yang/assemblykor/issues

---

bills *Bills Proposed in the Korean National Assembly (20th-22nd)*

Description

Metadata for the 64,900 bills that members proposed during the 20th through 22nd Korean National Assembly (2016-2026), up to 2026-09-23.

Usage

```
bills
```

Format

A data frame with 64,900 rows and 11 variables:

bill_id Unique bill identifier from the National Assembly system

bill_no Numeric bill number

assembly Assembly number (20, 21, or 22)

bill_name Full bill title in Korean

committee Standing committee to which the bill was referred

propose_date Date the bill was formally proposed

result Legislative outcome in Korean. Common values include passed as-is, expired at term end, and incorporated into alternative bill. `NA` for bills pending on 2026-09-23. For a vetoed bill, the outcome after the veto (rejected on the re-vote, passed again, or expired at the end of the term). See `table(bills$result)` for all values.

proposer Name of the lead (primary) proposer

proposer_id MONA_CD of the lead proposer (links to `legislators$member_id`). Comma-separated for the few bills with joint lead proposers.

vetoed Logical: the President returned the bill to the Assembly for reconsideration (a presidential veto)

alt_vetoed Logical: the bill was incorporated into a committee alternative (`result` "incorporated into alternative bill") that was vetoed and not passed again, so its content never became law

Details

The Korean National Assembly has seen a dramatic increase in bill proposals: the 21st Assembly produced 23,655 bills versus 21,594 in the 20th. Most bills expire at the end of the assembly term (term expiry). Only about 5\ with amendments.

Up to version 0.1.3, 11 of the 12 vetoed bills, which were rejected on the re-vote or expired at the end of the term, kept the result of their first floor vote. Their `result` now follows the re-vote, as in release 0.7.0 of the kna project. To count bills whose content became law, exclude `alt_vetoed` bills as well.

Use `get_bill_texts()` to download the full propose-reason texts for text analysis, and `get_proposers()` for the complete co-sponsorship records (825,283 rows).

Source

Open National Assembly Information API (Republic of Korea), as corrected in kna 0.7.0 (20th and 21st) and kna 0.8.1 (22nd) (https://github.com/kyusik-yang/kna).

Examples

```
data(bills)

# Bills per assembly
table(bills$assembly)

# Top 10 committees
```

```
sort(table(bills$committee), decreasing = TRUE)[1:10]

# Distribution of legislative outcomes
head(sort(table(bills$result), decreasing = TRUE))
```

`get_bill_texts` *Download bill propose-reason texts*

Description

Downloads the full propose-reason texts (jean-iyu) for all 64,900 bills in `bills`. The file is approximately 26 MB and is cached locally after the first download. Requires the **arrow** package to read parquet files.

Usage

```
get_bill_texts(cache_dir = NULL, force_download = FALSE)
```

Arguments

`cache_dir`	Directory to cache downloaded files. Defaults to `tools::R_user_dir("assemblykor", "cache")`.
`force_download`	Logical. If `TRUE`, re-download even if cached.

Details

The texts are those of release 0.8.1 of the kna project (`https://github.com/kyusik-yang/kna`), for the bills in `bills`. Versions up to 0.1.3 downloaded the scraped texts of the korean-assembly-bills dataset, which cover the bills proposed by 2026-02-27 and are unchanged here. Record `source` in text analyses.

Value

A data frame with 64,900 rows and 4 variables, or `NULL` (invisibly) if the download fails (e.g., no internet connection):

bill_id Bill identifier (links to `bills$bill_id`)

propose_reason Full text of the propose-reason statement (Korean). `NA` for the 80 bills that have no text in the source.

scrape_status Status of the web collection of the text: "ok", "empty", "no_csrf", or "error". `NA` for texts from the API.

source "likms_scrape" for the texts collected from the Legislative Information System (bills proposed by 2026-02-27), or "BPMBILLSUMMARY" for the texts of the Open Assembly API, which begin with a heading that the scraped texts lack. `NA` for the eight bills without any text record.

Examples

```
if (requireNamespace("arrow", quietly = TRUE)) {
  texts <- get_bill_texts(cache_dir = tempdir())

  if (!is.null(texts)) {
    nchar_dist <- nchar(texts$propose_reason)
    hist(nchar_dist, breaks = 100, main = "Length of Propose-Reason Texts")
  }
}
```

`get_proposers` *Download bill co-sponsorship records*

Description

Downloads the complete proposer records (825,283 rows) listing every legislator who proposed, co-proposed or supported each of the 64,900 bills in `bills`. Requires the **arrow** package.

Usage

```
get_proposers(cache_dir = NULL, force_download = FALSE)
```

Arguments

`cache_dir`	Directory to cache downloaded files. Defaults to `tools::R_user_dir("assemblykor", "cache")`.
`force_download`	Logical. If `TRUE`, re-download even if cached.

Details

The records come from the official proposer list of each bill (BILLINFOPPSR endpoint), as rebuilt in the kna project (https://github.com/kyusik-yang/kna), release 0.8.1. Releases up to 0.1.3 of this package served an earlier file that stopped at 100 names per bill, which left out 7,447 records of the 208 bills with more than 100 proposers and supporters, and whose `is_lead` was `FALSE` for the lead proposer of 36 single-proposer bills. Bills with joint lead proposers have more than one row with `is_lead` = `TRUE`.

Value

A data frame with 825,283 rows and 9 variables, or `NULL` (invisibly) if the download fails (e.g., no internet connection):

bill_id Bill identifier (links to `bills$bill_id`)

bill_no Numeric bill number

bill_name Bill title in Korean

propose_date Proposal date

proposer_name Legislator name

proposer_party Party affiliation at the time of proposal

member_id Legislator identifier (links to `legislators$member_id`)

is_lead Logical: `TRUE` if lead (primary) proposer, `FALSE` if co-proposer or supporter (see `role`)

role Role on the bill in Korean, one of lead proposer (daepyo balui), co-proposer (gongdong balui) or supporter (chanseong). Supporters are the members counted in the "oe M in" part of the proposer text.

Examples

```
if (requireNamespace("arrow", quietly = TRUE) &&
    requireNamespace("dplyr", quietly = TRUE)) {
  props <- get_proposers(cache_dir = tempdir())

  if (!is.null(props)) {
    # Build co-sponsorship edgelist
    leads <- dplyr::select(
      dplyr::filter(props, is_lead), bill_id, lead = member_id
    )
    cosponsors <- dplyr::select(
      dplyr::filter(props, !is_lead), bill_id, cosponsor = member_id
    )
    edges <- dplyr::inner_join(
      leads, cosponsors,
      by = "bill_id", relationship = "many-to-many"
    )
  }
}
```

`get_speech_tokens` *Download morpheme tokens for committee speeches*

Description

Downloads a pre-tokenized version of the `speeches` dataset, produced with the Kiwi morphological analyzer (via `kiwipiepy`). Korean is an agglutinative language, so whitespace tokenization mixes particles and verb endings into the tokens; morphological analysis separates them and lemmatizes verbs and adjectives. This dataset lets students work with proper Korean tokens without installing a morphological analyzer. The file is approximately 1.3 MB and is cached locally after the first download. Requires the **arrow** package.

Usage

```
get_speech_tokens(cache_dir = NULL, force_download = FALSE)
```

Arguments

cache_dir	Directory to cache downloaded files. Defaults to `tools::R_user_dir("assemblykor", "cache")`.
force_download	Logical. If `TRUE`, re-download even if cached.

Details

Only content morphemes are included; particles (josa), verb endings (eomi), and punctuation are removed. Function words carry little topical meaning, so this is the usual starting point for keyword and topic analysis. For noun-based analysis, filter to `pos %in% c("NNG", "NNP")`.

Join back to `speeches` with `by = c("date", "speech_order")` to attach speaker metadata. Every speech has at least one token. Two meetings were held on 2024-06-25, so eight `speech_order` values of that date belong to two speeches each, and their tokens are pooled under the shared key. Up to version 0.1.3 the file also held a second copy of the tokens of 48 speeches that appeared twice in `speeches`.

The tokenization script is in the package source repository under `data-raw/tokenize_speeches.py`.

Value

A data frame with 663,582 rows and 4 variables, or `NULL` (invisibly) if the download fails (e.g., no internet connection):

date Date of the committee meeting (links to `speeches$date`)

speech_order Speech turn within the meeting (links to `speeches$speech_order`); `date` + `speech_order` identifies one speech, except on 2024-06-25 (see Details)

token Morpheme, in dictionary form. Verbs and adjectives are lemmatized (e.g., the stem plus `-da`)

pos Part-of-speech tag from the Sejong tagset: "NNG" (common noun), "NNP" (proper noun), "VV" (verb), "VA" (adjective), "MAG" (adverb), or "SL" (foreign word, e.g., "AI")

See Also

`speeches`

Examples

```
if (requireNamespace("arrow", quietly = TRUE)) {
  tokens <- get_speech_tokens(cache_dir = tempdir())

  if (!is.null(tokens)) {
    # Most frequent nouns
    nouns <- tokens[tokens$pos %in% c("NNG", "NNP"), ]
    head(sort(table(nouns$token), decreasing = TRUE), 20)
  }
}
```

`legislators` *Members of the Korean National Assembly (20th-22nd)*

Description

Biographical and political metadata for 963 records of legislators who served in the 20th (2016-2020), 21st (2020-2024), or 22nd (2024-2028) Korean National Assembly. Some legislators appear in multiple assemblies. The 22nd assembly covers the members seated by 2026-09-23, among them the winners of the by-elections of 2026-06-11.

Usage

```
legislators
```

Format

A data frame with 963 rows and 15 variables:

member_id Unique legislator identifier (MONA_CD from the National Assembly API)

assembly Assembly number (20, 21, or 22)

name Name in Korean (hangul)

name_hanja Name in Chinese characters (hanja)

name_eng Name in English (romanized)

party Party label that the official roster of that assembly records for the member. It reflects party mergers, renamings and switches during the term. For the 22nd assembly it is the party as of September 2026, or the last party recorded for a member who had left the Assembly by then.

party_elected Party at election, that is, the party on whose ticket or list the member was elected. For a successor to a proportional seat, the party of the list the seat came from. The source records ten such successors under the party that the list party had merged into by the time they took the seat (for example, the Democratic Party of Korea for the Democratic Alliance of Korea), as does release 0.7.0 of the kna project.

district Electoral district name, or party list position for proportional members

district_type Election type: "constituency" or "proportional"

committees Committees (standing and special) the member served on in that assembly, comma-separated in order of first assignment. Empty for a member with no assignment, such as the Speaker.

gender "M" (male) or "F" (female)

birth_date Date of birth

seniority Seniority at that assembly, that is, the number of terms served up to and including this one, counting terms before the 20th (1 = first-term)

n_bills Number of bills in `bills` the member proposed, co-proposed or supported (see `get_proposers`)

n_bills_lead Bills proposed as lead (primary) proposer

Details

672 unique legislators served across the three assemblies. `member_id` is consistent across assemblies, so legislators can be tracked over time. Party names may differ between `party` (mid-term) and `party_elected` (election day) due to party mergers and name changes, which are common in Korean politics. Some legislators share a name with another member of the same assembly (for example, two members named Kim Seong-tae in the 20th), so join on `member_id`, never on `name`.

Up to version 0.1.3, `seniority` held each member's lifetime number of terms at the time of data collection, so it overstated the seniority of 286 member-terms of the 20th and 21st assemblies, and `committees` came from a present-day string that did not match the assembly. Both now follow release 0.7.0 of the kna project, as do six values of `party_elected`, one district, two district types and the bill counts. The 22nd assembly is rebuilt from release 0.8.1 of the kna project.

Source

Open National Assembly Information API (Republic of Korea), as corrected in kna 0.7.0 (20th and 21st) and kna 0.8.1 (22nd) (https://github.com/kyusik-yang/kna). License: public domain (Korean government open data).

Examples

```
data(legislators)

# Party composition by assembly
table(legislators$assembly, legislators$party)

# Gender gap in bill production
tapply(legislators$n_bills_lead, legislators$gender, median)

# First-term vs senior legislators
boxplot(n_bills_lead ~ seniority, data = legislators,
        xlab = "Terms served", ylab = "Bills proposed (lead)")
```

`list_tutorials` *List available tutorials*

Description

Lists the tutorial R Markdown files included with the package. Tutorials are designed for classroom use in Korean political science methods courses. Each tutorial is available in two formats:

1. Plain Rmd for editing in RStudio (`open_tutorial`)
2. Interactive **learnr** format (`run_tutorial`)

Usage

```
list_tutorials()
```

Value

A character vector of tutorial file names (invisibly).

Examples

```
list_tutorials()
```

`open_tutorial` *Open a tutorial file*

Description

Copies a tutorial R Markdown file to the specified directory (default: current working directory) so students can edit and run it in RStudio.

Usage

```
open_tutorial(name, dest_dir = getwd())
```

Arguments

`name`	Tutorial name (with or without .Rmd extension), or a number corresponding to the tutorial order (1-9).
`dest_dir`	Directory to copy the file to. Defaults to the current working directory.

Value

The path to the copied file (invisibly).

See Also

`run_tutorial` for the interactive browser version.

Examples

```
if (interactive()) {
  # Copy by name
  open_tutorial("01-tidyverse-basics")

  # Copy by number
  open_tutorial(1)
}
```

path_to_file *Path to assemblykor CSV files*

Description

Returns the file path to CSV versions of the built-in datasets stored in `inst/extdata`. Useful for teaching file I/O with `read.csv()` or `readr::read_csv()`.

Usage

```
path_to_file(file = NULL)
```

Arguments

`file`	Name of the CSV file. One of `"legislators.csv"`, `"wealth.csv"`, or `"seminars.csv"`.

Value

A character string with the full file path.

Examples

```
# Read data from CSV (alternative to data())
path <- path_to_file("legislators.csv")
legislators_csv <- read.csv(path, fileEncoding = "UTF-8")
head(legislators_csv)
```

roll_calls *Member-Level Roll Call Votes (22nd Assembly)*

Description

Individual legislator voting records for all 1,847 bills that went to a recorded plenary vote in the 22nd Korean National Assembly from July 2024 to September 17, 2026. Each row represents one legislator's vote on one bill.

Usage

```
roll_calls
```

Format

A data frame with 549,513 rows and 9 variables:

bill_id Bill identifier (links to `votes$bill_id` and `bills$bill_id`)

assembly Assembly number (22)

member_name Legislator name in Korean

member_id Legislator identifier (MONA_CD, links to `legislators$member_id`)

party Party label reported by the API at data collection (September 2026). The API writes the member's party at that time onto every past vote, so a member who changed party during the term appears under the later party on all votes. For the votes taken from the LIKMS vote pages (see Source) it is the member's party in September 2026.

party_elected Party at election, as in `legislators$party_elected`. Members elected on the lists of the satellite parties (e.g., the People Future Party and the Democratic Alliance of Korea) carry the list party, although they sat with other parties. Three members who succeeded to proportional seats during the term carry the party into which the list party had merged, two the Democratic Party of Korea and one the People Power Party (see `legislators`).

district Electoral district at the 2024 election, or the proportional list, as in `legislators$district`

vote Vote cast in Korean: one of four values meaning yes, no, abstain, or absent

vote_date Date of the vote

Details

This dataset covers the 22nd assembly. The same API endpoint also has member-level votes of the 20th and 21st assemblies, which are left out to keep the package small. The kna project (`https://github.com/kyusik-yang/kna`) provides them. For the 20th and 21st assemblies, use the bill-level `votes` dataset.

Neither party column records the party at the time of each vote. Up to version 0.1.3, `party` was documented as the party at the time of the vote. Version 0.1.4 corrects that description, adds `party_elected`, and rebuilds the dataset from release 0.8.1 of the kna project, which adds the votes of the 16 members seated in 2026 that the API omits. Every member seated at a vote has a row for it, absent if the member did not vote.

This dataset enables ideal point estimation (e.g., W-NOMINATE), party unity scores, and analysis of legislative coalitions. Use `member_id` to link with `legislators` for biographical metadata.

Source

Open National Assembly Information API (Republic of Korea), endpoint `nojepdqqaweusdfbi`, as collected by release 0.8.1 of the kna project (`https://github.com/kyusik-yang/kna`) in September 2026. The votes of the 16 members the API omits come from the vote pages of the Legislative Information System (LIKMS), as in kna 0.8.0.

See Also

`votes`

Examples

```
data(roll_calls)

# Vote distribution
table(roll_calls$vote)

# Votes per party
head(sort(table(roll_calls$party), decreasing = TRUE))

# Number of unique legislators
length(unique(roll_calls$member_id))
```

`run_tutorial` *Run an interactive tutorial*

Description

Launches a **learnr** interactive tutorial in the browser. Students can type and run code directly in the browser with hints and solutions. Requires the **learnr** package.

Usage

```
run_tutorial(name)
```

Arguments

`name`	Tutorial name or number (1-9). Use `list_tutorials` to see available tutorials.

Value

No return value, called for the side effect of launching a **learnr** tutorial in the browser.

See Also

`open_tutorial` for the plain Rmd version.

Examples

```
if (interactive()) {
  run_tutorial(1)
}
```

`seminars` *Policy Seminar Activity by Legislator-Year (2004-2025)*

Description

Annual panel of policy seminar hosting activity for legislators in the 17th through 22nd Korean National Assembly. Policy seminars (jeongchaek semina) are informal legislative events where MPs invite experts, stakeholders, and colleagues from other parties to discuss policy issues.

Usage

```
seminars
```

Format

A data frame with 5,962 rows and 18 variables:

name Legislator name in Korean

member_id Legislator identifier (MONA_CD, links to `legislators$member_id`). Available for 5,696 rows (95.5\ and `NA` for unmatched or ambiguous (homonym) cases.

year Calendar year

assembly Assembly number (17-22), assigned from the calendar year (2004-2007 to the 17th, 2008-2011 to the 18th, and so on)

party Party affiliation

camp Political camp: "liberal", "conservative", "progressive", "centrist", or "other" (values are in Korean)

seniority Seniority at that assembly, that is, the number of terms served up to and including this one, counting terms before the 17th (1 = first-term)

n_seminars Number of policy seminars hosted that year

n_cross_party Number of seminars co-hosted with other-party legislators

cross_party_ratio Share of seminars that were cross-party (0-1)

avg_coalition_size Average number of co-hosts per seminar

is_governing Logical: belongs to the governing (presidential) party

is_female Logical: female legislator

is_proportional Logical: holds a proportional-representation seat in that assembly

is_seoul Logical: represents a Seoul district in that assembly

province Province or metropolitan city of the electoral district in that assembly, in Korean short form (e.g., Seoul, Gyeonggi). `NA` for proportional-representation members.

total_terms Total assembly terms served across the career, as of September 2026

n_bills_led Number of bills the legislator proposed as lead proposer in that assembly term (the same value in every year of the term)

Details

Policy seminars are a distinctive feature of the Korean National Assembly. Unlike floor speeches or committee hearings, seminars are voluntary and allow legislators to signal policy expertise and build cross-party ties. The `cross_party_ratio` variable captures how often a legislator cooperates across party lines in this informal arena.

The `is_governing` variable enables difference-in-differences designs: when a party transitions from opposition to governing (or vice versa), does its members' cross-party collaboration change?

The member attributes `seniority`, `total_terms`, `is_female`, `is_proportional`, `is_seoul` and `province` come from the member records of release 0.7.0 of the kna project, matched on `member_id` and `assembly`. Up to version 0.1.3 they were matched on the name alone and counted only the terms from the 17th assembly on. They are `NA` when a row has no `member_id`, and all but `total_terms` and `is_female` are `NA` when the legislator did not serve in that assembly.

The panel counts seminars by name and year. Because a new assembly begins on May 30 of an election year, the rows of an election year also hold the activity of members of the outgoing assembly, under the new assembly number. Legislators who share a name with another member of the same assembly have one row per year for both of them, with `member_id` `NA`.

Source

National Assembly Seminar Database, collected via API. Member attributes from kna 0.7.0 (https://github.com/kyusik-yang/kna).

Examples

```
data(seminars)

# Cross-party collaboration by governing status
tapply(seminars$cross_party_ratio, seminars$is_governing, mean, na.rm = TRUE)

# Seminar activity over time
agg <- aggregate(n_seminars ~ year, data = seminars, FUN = sum)
plot(agg, type = "b", main = "Total Policy Seminars by Year")

# Gender gap in seminar hosting
tapply(seminars$n_seminars, seminars$is_female, median, na.rm = TRUE)
```

`set_ko_font` *Set Korean font for ggplot2*

Description

Detects a Korean-compatible font on the current system and applies it to all ggplot2 plots via `theme_set()`. Call this once at the top of your script to avoid broken Korean text in plot titles and labels.

Usage

```
set_ko_font(font = NULL)
```

Arguments

font	Optional font family name to use directly. If NULL (default), auto-detects from common Korean fonts.

Value

The font family name used (invisibly).

Examples

```
if (interactive()) {
  library(ggplot2)
  set_ko_font()

  # Now Korean text renders correctly
  ggplot(data.frame(x = 1), aes(x, x)) +
    geom_point() +
    labs(title = "Korean Title Test")
}
```

---

speeches *Committee Speeches from the Science and ICT Committee (22nd Assembly)*

Description

Full corpus of 15,795 speech records from the Science, Technology, Information, Broadcasting and Communications Committee of the 22nd Korean National Assembly (2024). Standing committee meetings only.

Usage

```
speeches
```

Format

A data frame with 15,795 rows and 10 variables:

assembly Assembly number (22)

date Date of the committee meeting

committee Committee name in Korean

speaker Speaker label as it appears in the minutes (may include titles)

role Speaker role: "legislator", "chair", "minister", "vice_minister", "senior_bureaucrat", "agency_head", "witness", "expert_witness", "nominee", "minister_nominee", "testifier", "public_corp_head", "broadcasting", "committee_staff"

speaker_name Cleaned speaker name with titles removed

member_id Legislator identifier (MONA_CD, links to `legislators$member_id`) for the 12,060 speeches by members of the Assembly. `NA` for other speakers (ministers, witnesses, and the heads of other bodies).

speaker_id Numeric speaker identifier used in the committee minutes, which versions up to 0.1.3 stored in `member_id`. `NA` for speakers who are not members of the Assembly.

speech_order Order of the speech turn within the meeting

speech Full text of the speech in Korean

Details

This dataset contains the complete standing committee speech records (no sampling) for the Science and ICT Committee of the 22nd assembly (June-December 2024). Speeches shorter than 50 characters were excluded. `date` and `speech_order` identify a speech, except on 2024-06-25, when two meetings were held and eight `speech_order` values occur twice.

Up to version 0.1.3, `member_id` held the numeric speaker identifier of the minutes, so it did not link to `legislators$member_id`, and 48 speeches of 2024-08-14 appeared twice under two speaker identifiers. Version 0.1.4 attaches the MONA_CD through the member records of release 0.7.0 of the kna project and drops the duplicates.

The `role` variable distinguishes legislators from government officials, witnesses, and other participants. Filter to `role == "legislator"` for MP speeches only, or compare how legislators and ministers discuss the same agenda items.

This committee covers AI, telecommunications, broadcasting, space policy, and R&D governance, making it suitable for keyword analysis, topic modeling, and other text analysis exercises.

Source

National Assembly committee minutes via the Open National Assembly Information API.

Examples

```
data(speeches)

# Distribution of speech lengths
hist(nchar(speeches$speech), breaks = 100,
     main = "Speech Length Distribution", xlab = "Characters")

# Speaker roles
table(speeches$role)

# Most frequent legislator speakers
leg <- speeches[speeches$role == "legislator", ]
head(sort(table(leg$speaker_name), decreasing = TRUE), 10)

# Simple keyword search (example: AI-related speeches)
ai <- speeches[grepl("AI", speeches$speech), ]
nrow(ai)
```

votes *Plenary Vote Results in the Korean National Assembly (20th-22nd)*

Description

Bill-level vote tallies from plenary sessions of the 20th through 22nd Korean National Assembly (2016-2026). Each row represents one bill that went to a recorded floor vote.

Usage

```
votes
```

Format

A data frame with 8,611 rows and 13 variables:

bill_id Bill identifier (links to `bills$bill_id`). Unique except for bill 2000491 of the 20th assembly, which has two tally rows in the source.

bill_no Numeric bill number

bill_name Full bill title in Korean

assembly Assembly number (20, 21, or 22)

committee Standing committee to which the bill was referred

vote_date Date of the plenary vote

result Vote outcome in Korean (e.g., passed as-is, passed with amendments, rejected)

bill_type Type of bill (e.g., legislation, budget, resolution)

total_members Total number of assembly members at the time

voted Number of members who cast a vote

yes Number of yes votes

no Number of no votes

abstain Number of abstentions

Details

Not all bills go to a floor vote. Most bills are disposed of in committee or expire at the end of the assembly term. The `votes` dataset captures only those that reached the plenary floor for a recorded vote.

About 40\ because `bills` only contains legislator-proposed bills while `votes` also includes committee alternatives, budget bills, and resolutions that have separate identifiers.

See `roll_calls` for member-level voting records (22nd assembly), useful for ideal point estimation or party discipline analysis.

Source

Open National Assembly Information API (Republic of Korea), endpoint `ncocpgfiaoituanbr`. The 22nd assembly is the collection of release 0.8.1 of the kna project (https://github.com/kyusik-yang/kna), with votes up to 2026-09-17.

Examples

```
data(votes)

# Votes per assembly
table(votes$assembly)

# Pass rate
table(votes$result)

# Average yes rate
votes$yes_rate <- votes$yes / votes$voted
summary(votes$yes_rate)

# Contentious votes (yes rate < 70%)
contentious <- votes[votes$yes / votes$voted < 0.7, ]
nrow(contentious)
```

`wealth` *Legislator Asset Declarations (2015-2025)*

Description

Panel data of asset declarations for 776 Korean National Assembly members across 11 years (2015-2025). Derived from the mandatory annual public disclosures.

Usage

```
wealth
```

Format

A data frame with 3,215 rows and 14 variables:

member_id Legislator identifier (links to `legislators$member_id`)

year Year the declared wealth refers to (2015-2025). The declaration is published in March of the following year.

name Legislator name in Korean

total_assets Total declared assets, in thousands of KRW

total_debt Total declared liabilities, in thousands of KRW

net_worth Net worth (assets minus debt), in thousands of KRW

real_estate Total real estate value, in thousands of KRW

building Total building/structure value, in thousands of KRW

land Total land value, in thousands of KRW

deposits Total bank deposits, in thousands of KRW

stocks Total stock holdings, in thousands of KRW

n_properties Total number of properties disclosed

has_seoul_property Logical: owns property in Seoul

has_gangnam_property Logical: owns property in Gangnam (Seoul's wealthiest district)

Details

All monetary values are in thousands of KRW (1 unit = 1,000 won). To convert to billions of won, divide by 1,000,000. For example, a net_worth of 1,670,000 means 1.67 billion won (approximately USD 1.2 million).

Legislators are required by law to disclose their assets annually. Not all legislators appear in every year, as the panel is unbalanced (entries correspond to active service periods).

Source

2015-2024: OpenWatch (https://docs.openwatch.kr/data/national-assembly), CC BY-SA 4.0 license. 2025: the National Assembly Gazette (Gukhoe Gongbo) No. 2026-54 of 2026-03-26, the March 2026 regular disclosure. Both as compiled in release 0.8.1 of the kna project (https://github.com/kyusik-yang/kna).

Examples

```
data(wealth)

# Distribution of net worth (in billions of won)
hist(wealth$net_worth / 1e6, breaks = 50,
     main = "Legislator Net Worth", xlab = "Billion KRW")

# Real estate as share of total assets
wealth$re_share <- wealth$real_estate / wealth$total_assets
summary(wealth$re_share)

# Gangnam property owners vs others
tapply(wealth$net_worth / 1e6, wealth$has_gangnam_property, median, na.rm = TRUE)
```

Index

∗ **datasets**
bills, 3
legislators, 9
roll_calls, 12
seminars, 15
speeches, 17
votes, 19
wealth, 20

assemblykor *(*assemblykor-package*)*, 2
assemblykor-package, 2

bills, *2*, 3, *5, 6*, *9*

get_bill_texts, *3*, 5
get_proposers, *3*, 6, *9*
get_speech_tokens, *3*, 7

legislators, *2*, 9, *13*
list_tutorials, *3*, 10, *14*

open_tutorial, *3*, *10*, 11, *14*

path_to_file, 12

roll_calls, *2*, 12, *19*
run_tutorial, *10, 11*, 14

seminars, *2*, 15
set_ko_font, 16
speeches, *2*, *7, 8*, 17

votes, *2*, *13*, 19

wealth, *2*, 20