Package 'assemblykor'

April 7, 2026

Title Korean National Assembly Data for Political Science Education

Version 0.1.1

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

RoxygenNote 7.3.3

VignetteBuilder knitr

NeedsCompilation no

Author Kyusik Yang [aut, cre]

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

Repository CRAN

Date/Publication 2026-04-07 07:30:02 UTC

Contents

`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 five built-in datasets covering legislator metadata, bills, asset declarations, policy seminars, and committee speeches.

Built-in datasets

- `legislators`: 947 MP records (20th-22nd assemblies)
- `bills`: 60,925 legislative bills
- `wealth`: 2,928 legislator-year asset declarations
- `seminars`: 5,962 legislator-year seminar records
- `speeches`: 15,843 speech records (22nd, Science & ICT Committee)
- `votes`: 7,997 plenary vote tallies (20th-22nd assemblies)
- `roll_calls`: 368,210 member-level roll call votes (22nd assembly)

Download functions

- `get_bill_texts`: 60,925 bill propose-reason texts
- `get_proposers`: 769,773 co-sponsorship records

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 60,925 legislative bills proposed during the 20th through 22nd Korean National Assembly (2016-2026).

Usage

```
bills
```

Format

A data frame with 60,925 rows and 9 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. 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`)

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\

Use `get_bill_texts()` to download the full propose-reason texts for text analysis, and `get_proposers()` for the complete co-sponsorship records (769,773 rows).

Source

Open National Assembly Information API (Republic of Korea).

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 60,925 bills. The file is approximately 40 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.

Value

A data frame with 60,925 rows and 3 variables:

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

propose_reason Full text of the propose-reason statement (Korean)

scrape_status Data collection status: "ok", "empty", "no_csrf", or "error"

Examples

```
texts <- get_bill_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 (769,773 rows) listing every legislator who co-sponsored each bill. 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.

Value

A data frame with 769,773 rows and 8 variables:

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 co-sponsorship

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

is_lead Logical: `TRUE` if lead (primary) proposer, `FALSE` if co-sponsor

Examples

```
props <- get_proposers()

# Build co-sponsorship edgelist
library(dplyr)
leads <- props %>% filter(is_lead) %>% select(bill_id, lead = member_id)
cosponsors <- props %>% filter(!is_lead) %>% select(bill_id, cosponsor = member_id)
edges <- inner_join(leads, cosponsors, by = "bill_id")
```

---

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

Description

Biographical and political metadata for 947 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.

Usage

```
legislators
```

Format

A data frame with 947 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 affiliation during the assembly term

party_elected Party at the time of election

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

district_type Election type: "constituency" or "proportional"

committees Standing committee assignments (comma-separated)

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

birth_date Date of birth

seniority Number of terms served, including current (1 = first-term)

n_bills Total bills participated in (as lead proposer or co-sponsor)

n_bills_lead Bills proposed as lead (primary) proposer

Details

661 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.

Source

Open National Assembly Information API (Republic of Korea). 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 1,233 bills that went to a recorded plenary vote in the 22nd Korean National Assembly (2024-2026). Each row represents one legislator's vote on one bill.

Usage

```
roll_calls
```

Format

A data frame with 368,210 rows and 8 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 affiliation at time of vote

district Electoral district or proportional list position

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

vote_date Date of the vote

Details

The member-level roll call API is only available for the 22nd assembly. For the 20th and 21st assemblies, use the bill-level votes dataset.

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`.

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 (2000-2025)*

Description

Annual panel of policy seminar hosting activity for legislators in the 16th 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 ~95\ `NA` for unmatched or ambiguous (homonym) cases.

year Calendar year

assembly Assembly number (17-22)

party Party affiliation

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

seniority Number of terms served

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: proportional-representation member

is_seoul Logical: represents a Seoul district

province Province/metro area of electoral district

total_terms Total assembly terms served across career

n_bills_led Number of bills proposed as lead proposer that year

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?

Source

National Assembly Seminar Database, collected via API.

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 15,843 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 15,843 rows and 9 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`). Available for all rows; however, non-legislator speakers (ministers, witnesses, etc.) will not match entries in `legislators`.

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.

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 7,997 rows and 13 variables:

bill_id Unique bill identifier (links to `bills$bill_id`)

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`.

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 773 Korean National Assembly members across 13 reporting periods (2015-2025). Derived from mandatory public disclosures via the OpenWatch project.

Usage

```
wealth
```

Format

A data frame with 2,928 rows and 14 variables:

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

year Disclosure year (2015-2025)

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

OpenWatch (https://docs.openwatch.kr/data/national-assembly), CC BY-SA 4.0 license.

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