PARTYPRESS datasets and variables

2023-02-06

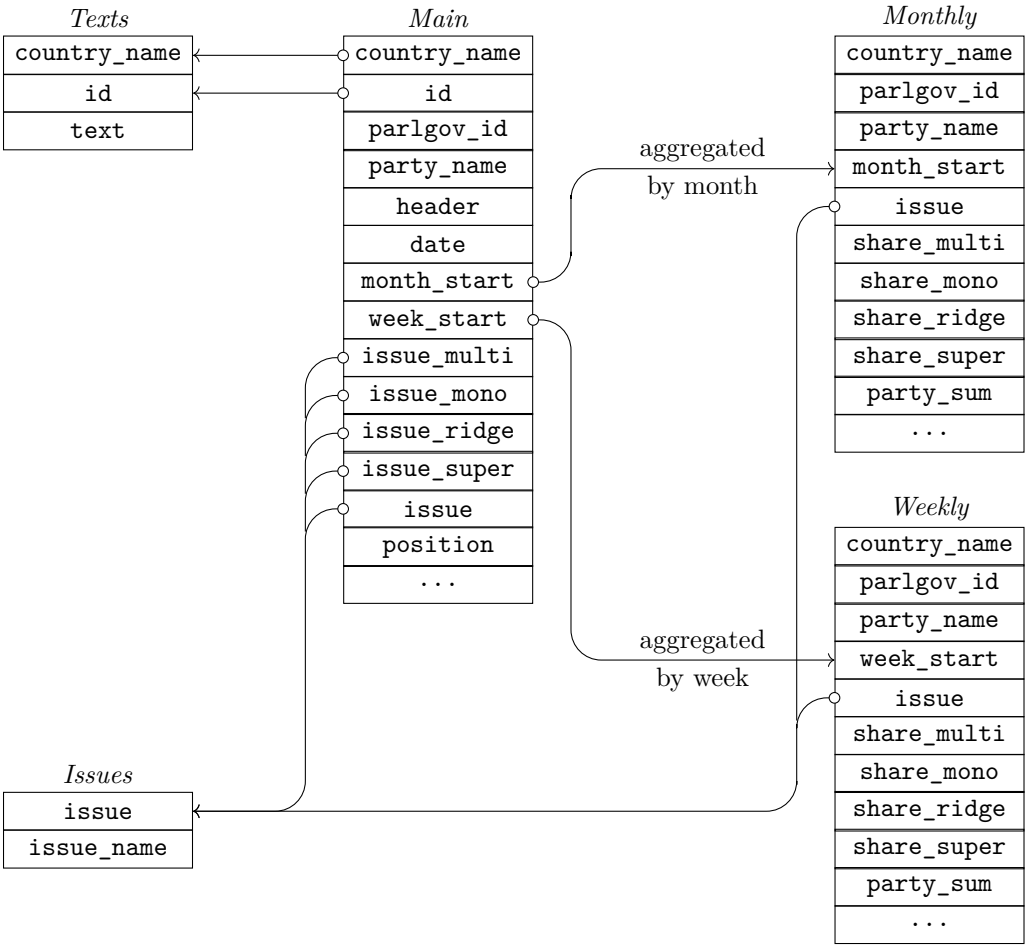
Contents

Overview

This document is designed to facilitate the use of the PARTYPRESS database. Figure 1 visualizes how the different datasets are connected. For each dataset the following sections provide an overview of variables with a description.

There are five different datasets. First, the main dataset *PARTYPRESS* that provides information about each individual press release in our database, e.g. issuing party or date. Second, *PARTYPRESS ISSUES* provides all issues used for classification; it can be merged to the other datasets. Third, *PARTYPRESS TEXTS* provides the full texts for each press release (the main dataset only contains the header); it can be merged to the main dataset. Fourth, *MONTHLY AGENDAS* provides data from the main dataset aggregated to calendar months. Fifth, *WEEKLY AGENDAS* provides data from the main dataset aggregated to calendar weeks.

Figure 1: Data structure



Main dataset PARTYPRESS

Load files `partypress.csv` (67MB) or `partypress.RDS` (9MB) or `partypress.dta` (600MB).

Variables

Variable name	Format	Description
`country_name`	[string]	Name of the country in small caps
`id`	[integer]	Press release identifier, unique within `country_name`
`parlgov_id`	[integer]	Party ID (merge with `party_id` in ParlGov)
`party`	[string]	Abbreviation of the party name (ParlGov)
`party_name`	[string]	Party name (ParlGov)
`party_name_english`	[string]	Party name in English (ParlGov)
`family_name`	[string]	Party family name (ParlGov)
`header`	[string]	Header of the press release in original language
`date`	[date]	Issue date of the press release
`month`	[yyyymm]	Issue month of the press release
`month_start`	[date]	Date of the first day of the month
`month_end`	[date]	Date of the last day of the month
`week_start`	[date]	Date of previous Monday
`week_end`	[date]	Date of following Sunday
`calendar_week`	[integer, 1-53]	Calendar week
`issue_multi`	[integer, 1-23]	Issue category predicted from Multilingual Transformer using all labeled docs from all countries
`issue_mono`	[integer, 1-23]	Issue category predicted from Monolingual Transformer using only labeled docs from one countries
`issue_ridge`	[integer, 1-23]	Issue category predicted from Ridge models, separate models by country
`issue_super`	[integer, 1-23]	Issue category predicted from SuperLearner models, separate models by country
`issue_coder`	[integer, 1-23]	Issue category hand coded by first or second country expert coder
`issue_coder2`	[integer, 1-23]	Issue category hand coded by second country expert coder, only for texts coded by both coders
`position_coder`	[integer, 1-2]	Position of press release, hand coded by one country expert coder, 1 pro, 2 anti
`position_coder2`	[integer, 1-2]	Position of press release, hand coded by 2nd country expert coder, only for texts coded by both coders, 1 pro, 2 anti
`cv_sample`	[integer]	Folds used in cross validation, $k = 1, ..., 5$, not used: -1

Sample

```
partypress <- readRDS("rds/partypress.RDS")
partypress[sample(1:nrow(partypress), 5), ] %>% select(country_name, id,
                                                       parlgov_id, party,
                                                       date, issue_multi)
```

```
##        country_name    id parlgov_id party       date issue_multi
## 26702       austria 34029       1013   ÖVP 2015-07-09           6
## 52718       austria  9244         50   FPÖ 2014-02-07         191
## 131129      ireland 35822       2217    SF 2015-03-24          13
## 210776        spain 17966        645    PP 2014-03-30          20
## 29013       austria 36109        973   SPÖ 2010-01-24           4
```

Issue names dataset PARTYPRESS ISSUES

Load files `partypress_issues.csv` (1KB) or `partypress_issues.RDS` (1KB) or `partypress_issues.dta` (3KB).

This dataset provides the issue category names for merging. Merge via one of the `issue_*` variables.

Variables

Variable name	Format	Description
`issue`	[integer, 1-23]	Issue category code
`issue_name`	[string]	Issue category description

Issue list

```
partypress_issues <- readRDS("rds/partypress_issues.RDS")
partypress_issues
```

```
##    issue            issue_name
## 1      1        Macroeconomics
## 2      2          Civil Rights
## 3      3                Health
## 4      4           Agriculture
## 5      5                 Labor
## 6      6             Education
## 7      7           Environment
## 8      8                Energy
## 9      9           Immigration
## 10    10        Transportation
## 11    12         Law and Crime
## 12    13        Social Welfare
## 13    14               Housing
## 14    15     Domestic Commerce
## 15    16               Defense
## 16    17            Technology
## 17    18         Foreign Trade
## 18   191 International Affairs
## 19   192  European Integration
## 20    20 Government Operations
## 21    23               Culture
## 22    98          Non-thematic
## 23    99                 Other
```

Raw texts dataset PARTYPRESS TEXTS

Load files `partypress_texts.csv` (561MB) or `partypress_texts.RDS` (193MB) (not available as `.dta`).

This dataset provides the raw texts of all press releases. The headers are contained in the main dataset. We provide the texts separately because they use a lot of disk space but some users may not be interested in raw text. Merge via `country_name` + `id`.

Variables

Variable name	Format	Description
`country_name`	[string]	Name of the country in small caps
`id`	[integer]	Press release identifier, unique within `country_name`
`text`	[string]	Text of press releases in original language

Sample

```
partypress_texts <- readRDS("rds/partypress_texts.RDS")
partypress_texts[sample(1:nrow(partypress_texts), 10), ] %>%
  mutate(text = substr(text, 1, 60) %>% str_c("(...)")) %>%
  select(country_name, id, text)
```

```
##        country_name    id
## 5932        austria  5932
## 146762      ireland 44327
## 109516      ireland  7081
## 36097       austria 36097
## 219404        spain 21964
## 4476        austria  4476
## 39984       austria 39984
## 53105       austria 53105
## 138812      ireland 36377
## 78592       germany 21107
##                                                                    text
## 5932    Seniorenplan birgt Chancen:\n "Der vom Seniorenrat vorgelegte(...)
## 146762   Sinn Féin spokesperson for Jobs, Enterprise and Innovation M(...)
## 109516   "Fianna Fáil's belated attempt to support small business rin(...)
## 36097    "Die heutige parlamentarische Enquete zeigt eines ganz deutl(...)
## 219404   "Los recortes económicos suponen  un riesgo para el mantenim(...)
## 4476   Bitte betrachten Sie die heutige Meldung OTS0144:  \r\nFPÖ-Dei(...)
## 39984    Fortschritte bei Reaktor-Stilllegung in Bulgarien, Litauen u(...)
## 53105    „Wir brauchen einen Schutzschirm für die ArbeitnehmerInnen u(...)
## 138812   Sinn Féin MEP Liadh Ní Riada will host a delegation on the s(...)
## 78592    „Der Wechsel der Financial Intelligence Unit (FIU) vom BKA z(...)
```

Aggregated datasets MONTHLY and WEEKLY AGENDAS

Load files `monthly_agendas.csv` (20MB)/`weekly_agendas.csv` (69MB) or `monthly_agendas.RDS` (1MB)/`weekly_agendas.RDS` (3MB) or `monthly_agendas.dta` (31MB)/`weekly_agendas.dta` (117MB).

Variables

Variable name	Format	Description
`country_name`	[string]	Name of the country in small caps
`parlgov_id`	[integer]	Party ID (merge with `party_id` in ParlGov)
`party`	[string]	Abbreviation of the party name (ParlGov)
`party_name`	[string]	Party name (ParlGov)
`party_name_english`	[string]	Party name in English (ParlGov)
`family_name`	[string]	Party family name (ParlGov)
only in `monthly_agendas`:		
`month`	[yyyymm]	Month
`month_start`	[date]	Date of the first day of the month
`month_end`	[date]	Date of the last day of the month
only in `weekly_agendas`:		
`calendar_week`	[integer, 1-53]	Calendar week
`week_start`	[date]	Date of previous Monday
`week_end`	[date]	Date of following Sunday
`issue`	[integer]	Issue category
`issue_descr`	[string]	Issue category description
`share_multi`	[numeric]	Share of press releases of category `issue` of all `party_sum` predicted from Multilingual Transformer using all labeled docs from all countries
`share_mono`	[numeric]	Share of press releases of category `issue` of all `party_sum` predicted from Monolingual Transformer using all labeled docs from each country in separate models
`share_ridge`	[numeric]	Share of press releases of category `issue` of all `party_sum` predicted from Ridge using all labeled docs from each country in separate models
`share_super`	[numeric]	Share of press releases of category `issue` of all `party_sum` predicted from SuperLearner using all labeled docs from each country in separate models
`party_sum`	[integer]	Total number of press releases by this party in this month/week

Samples

```
monthly_agendas <- readRDS("rds/monthly_agendas.RDS")
monthly_agendas[sample(1:nrow(monthly_agendas), 10), ] %>%
  select(country_name, parlgov_id, party, month, issue,
         share_multi)
```

```
##        country_name parlgov_id  party  month issue share_multi
## 14166       denmark        211 RV (B) 201805    99  0.00000000
## 127187           uk       1272   UKIP 202002    98  0.11111111
## 6217        austria       1013    ÖVP 201109     7  0.01904762
## 66041   netherlands        756     GL 201502     8  0.23684211
## 4886        austria        973    SPÖ 201704    10  0.08080808
## 48698       ireland       1573     GP 201607     7  0.33333333
## 77393        poland        512     PO 201612    99  0.00000000
## 10345       austria       1429  GRÜNE 201603    20  0.05555556
## 85076        poland        664    PSL 201912   191  0.00000000
## 71555   netherlands       1251    SGP 201606     2  0.14285714
```

```
weekly_agendas <- readRDS("rds/weekly_agendas.RDS")
weekly_agendas[sample(1:nrow(weekly_agendas), 10), ] %>%
  select(country_name, parlgov_id, party, week_start,
         issue, share_multi)
```

```
## # A tibble: 10 x 6
## # Groups:   parlgov_id [9]
##    country_name parlgov_id party  week_start issue share_multi
##    <chr>             <dbl> <chr>  <date>     <dbl>       <dbl>
##  1 ireland            1393 FG     2019-04-15    15       0
##  2 germany             772 Greens 2016-12-05     1       0
##  3 sweden             1546 SD     2017-09-04    98       0.667
##  4 denmark            1418 DF (O) 2014-05-26    16       0
##  5 ireland            1573 GP     2020-10-26     9       0
##  6 netherlands         357 SP     2011-03-07     8       0
##  7 germany             543 FDP    2018-02-05     7       0
##  8 poland              512 PO     2013-12-02     6       0.111
##  9 germany             543 FDP    2010-08-23     2       0.05
## 10 netherlands         742 PvdA   2013-07-29     7       0
```