PARTYPRESS datasets and variables

2023-02-06

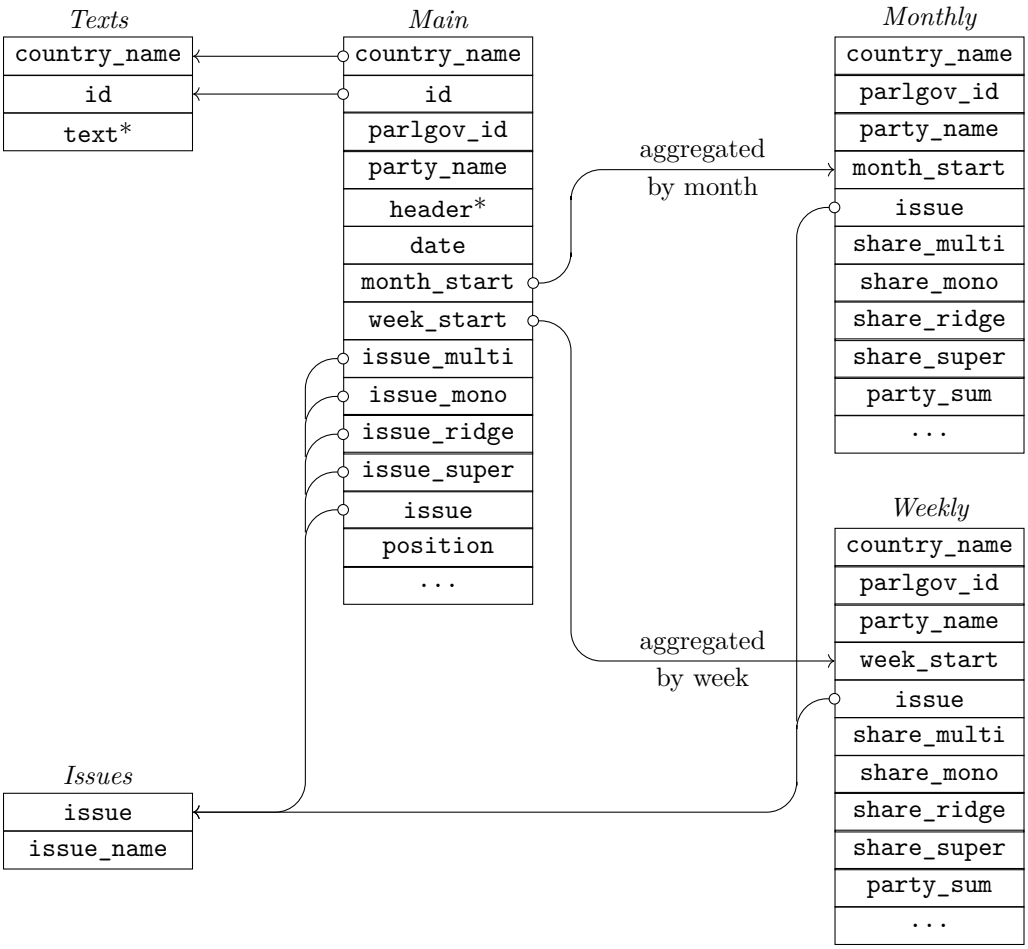
Contents

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



*The text and header of the press releases will be published after permission has been granted by the copyright holders.

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

*The text and header of the press releases will be published after permission has been granted by the copyright holders.

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
## 223999        spain 29924        902   PSOE 2012-07-01           3
## 246161           uk 16994       1556    Lab 2020-12-16         191
## 217621        spain 24144        902   PSOE 2018-04-01          20
## 172644  netherlands 21206        990   PvdD 2013-02-01           7
## 264217           uk  5316        467 Greens 2020-03-24           3
```

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
## 188307       poland  4093
## 265363           uk 26978
## 197672       poland 13458
## 200131       poland 15917
## 26385       austria 26385
## 259465           uk 21078
## 198197       poland 13983
## 108997      ireland  6562
## 252820           uk 14432
## 255040           uk 16652
##                                                                  text
## 188307 Dzięki Państwa zaangażowaniu, w ciągu ostatnich kilku dni ud(...)
## 265363 The UK government's plan to cap the number of times patients(...)
## 197672 PES Women zorganizowała w siedzibie Partii Europejskich Socj(...)
## 200131 przeciwnego zdania jest 9 proc. badanych - wynika z sondażu (...)
## 26385  Meinl-Reisinger: "Chance auf ein Haus der Geschichte muss je(...)
## 259465 The Liberal Democrats believe in a Britain that is open, mod(...)
## 198197 Kongres Lewicy zorganizowany dnia 16 czerwca 2013 roku na St(...)
## 108997 Speaking today about the closure of Mount Carmel Hospital, F(...)
## 252820 Vernon Coaker to set out Labour's vision for Britain's defen(...)
## 255040 Jo Stevens MP, Labour's Shadow Digital, Culture, Media and S(...)
```

*The text and header of the press releases will be published after permission has been granted by the copyright holders.

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
## 59040   netherlands        345   D66 201702   191  0.13333333
## 128185           uk       1284   SNP 201210     6  0.02666667
## 75091   netherlands       1501   PVV 201605    23  0.00000000
## 86519        poland       2602  Konf 201602    17  0.00000000
## 132787           uk       1556   Lab 201803     8  0.01503759
## 124724           uk        773   Con 202012    20  0.00000000
## 76892        poland        512    PO 201503     3  0.00000000
## 64282   netherlands        742  PvdA 201710    98  0.40000000
## 125798           uk       1272  UKIP 201304    12  0.00000000
## 94422         spain       2375    Cs 201309     7  0.00000000
```

```
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 [10]
##    country_name parlgov_id party    week_start issue share_multi
##    <chr>             <dbl> <chr>    <date>     <dbl>       <dbl>
##  1 netherlands         345 D66      2017-07-10    20       0
##  2 uk                 1556 Lab      2020-01-06    12       0
##  3 netherlands         742 PvdA     2011-04-25    23       0
##  4 sweden             1154 MP       2014-10-20     5       0
##  5 germany             791 Left     2012-10-22    17       0.0333
##  6 austria              50 FPÖ      2013-07-29    18       0
##  7 netherlands        1501 PVV      2017-11-06    14       0
##  8 poland             2600 Kukiz'15 2016-09-19   192       0
##  9 germany             543 FDP      2012-07-02     3       0
## 10 denmark             211 RV (B)   2015-05-25     1       0.2
```