

PARTYPRESS datasets and variables*

Cornelius Erfort, Lukas F. Stoetzer, Heike Klüver

2023-02-06

Contents

Overview	2
Main dataset PARTYPRESS	3
Issue names dataset PARTYPRESS ISSUES	5
Raw texts dataset PARTYPRESS TEXTS	6
Aggregated datasets MONTHLY and WEEKLY AGENDAS	7

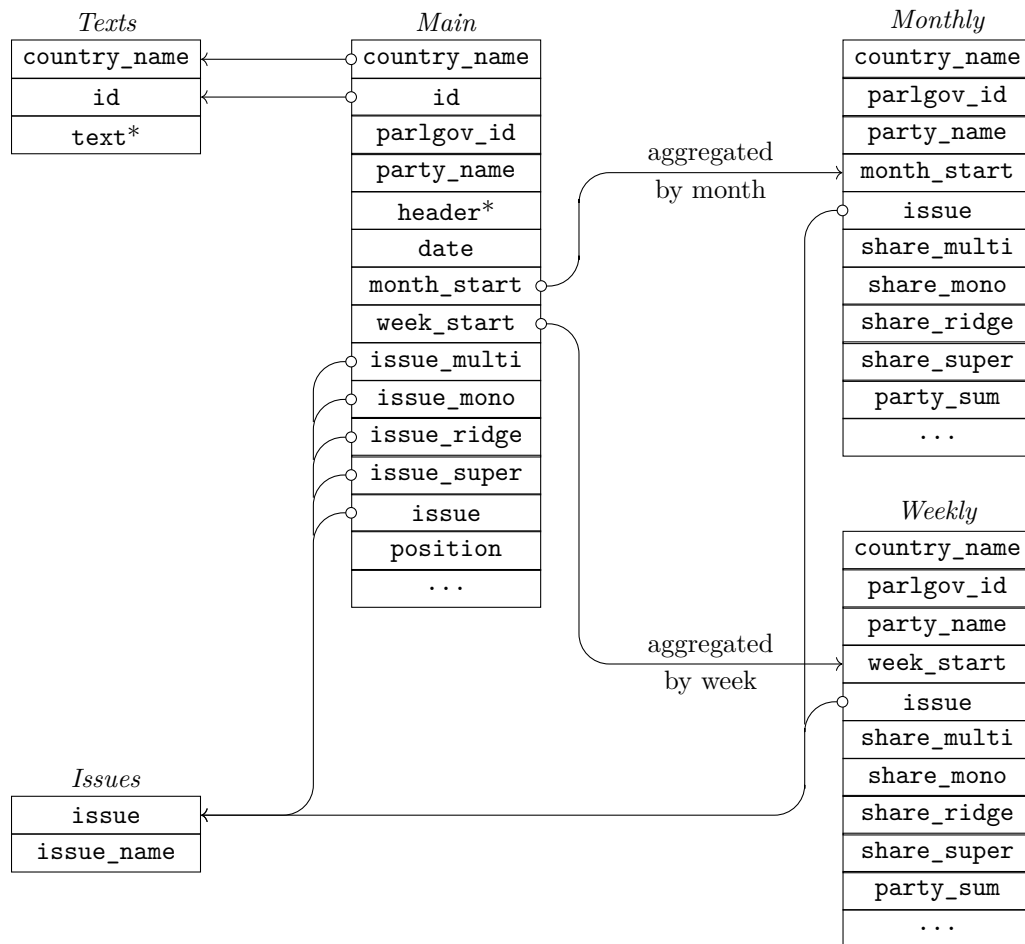
*We are grateful for valuable comments and suggestions by Eri Bertsou and excellent research assistance by Paul Bochtler, Daniel Cruz, Leonie Fuchs, Violeta Haas, Felix Heimbürger, Johannes Lattmann, and Tim Wappenhans. Research for this contribution is part of the Cluster of Excellence “Contestations of the Liberal Script” (EXC 2055, Project-ID: 390715649), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy. Cornelius Erfort is moreover grateful for generous funding provided by the DFG through the Research Training Group DYNAMICS (GRK 2458/1).

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
<code>country_name</code>	[string]	Name of the country in small caps
<code>id</code>	[integer]	Press release identifier, unique within <code>country_name</code>
<code>parlgov_id</code>	[integer]	Party ID (merge with <code>party_id</code> in ParlGov)
<code>party</code>	[string]	Abbreviation of the party name (ParlGov)
<code>party_name</code>	[string]	Party name (ParlGov)
<code>party_name_english</code>	[string]	Party name in English (ParlGov)
<code>family_name</code>	[string]	Party family name (ParlGov)
<code>header*</code>	[string]	Header of the press release in original language
<code>date</code>	[date]	Issue date of the press release
<code>month</code>	[yyyymm]	Issue month of the press release
<code>month_start</code>	[date]	Date of the first day of the month
<code>month_end</code>	[date]	Date of the last day of the month
<code>week_start</code>	[date]	Date of previous Monday
<code>week_end</code>	[date]	Date of following Sunday
<code>calendar_week</code>	[integer, 1-53]	Calendar week
<code>issue_multi</code>	[integer, 1-23]	Issue category predicted from Multilingual Transformer using all labeled docs from all countries
<code>issue_mono</code>	[integer, 1-23]	Issue category predicted from Monolingual Transformer using only labeled docs from one countries
<code>issue_ridge</code>	[integer, 1-23]	Issue category predicted from Ridge models, separate models by country
<code>issue_super</code>	[integer, 1-23]	Issue category predicted from SuperLearner models, separate models by country
<code>issue_coder</code>	[integer, 1-23]	Issue category hand coded by first or second country expert coder
<code>issue_coder2</code>	[integer, 1-23]	Issue category hand coded by second country expert coder, only for texts coded by both coders
<code>position_coder</code>	[integer, 1-2]	Position of press release, hand coded by one country expert coder, 1 pro, 2 anti
<code>position_coder2</code>	[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
<code>cv_sample</code>	[integer]	Folds used in cross validation, $k = 1, \dots, 5$, 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
## 27413	austria	3467	50	FPÖ	2011-04-01	10
## 39532	austria	45577	973	SPÖ	2015-05-21	8
## 16336	austria	2470	50	FPÖ	2010-11-19	5
## 225929	spain	31666	902	PSOE	2010-02-01	13
## 133868	ireland	38288	2217	SF	2015-11-17	14

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
<code>issue</code>	[integer, 1-23]	Issue category code
<code>issue_name</code>	[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
<code>country_name</code>	[string]	Name of the country in small caps
<code>id</code>	[integer]	Press release identifier, unique within <code>country_name</code>
<code>text*</code>	[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
## 71911      germany 14426
## 163602  netherlands 3743
## 182072  netherlands 22213
## 30828      austria 30828
## 180363  netherlands 20504
## 175520  netherlands 15661
## 264529      uk 26144
## 212759      spain 15313
## 5021      austria 5021
## 163107  netherlands 3248
##
##                                     text
## 71911  Zur Debatte über den Brexit erklärt der europapolitische Spr(...)
## 163602 Bezuinigen noodzakelijk? Tekst: Jola van Dijk Dat er 29 mil(...)
## 182072 Kamermeerderheid vóór aanmoediging handeldrijven met IsraëlH(...)
## 30828  ÖVP-Gesetzesentwurf bereits im März 2010 an Koalitionspartne(...)
## 180363 Wij willen daarom dat mensen eerder een vast contract krijge(...)
## 175520 Het nieuwe wetsvoorstel voor de luchtvaartsector van de Euro(...)
## 264529 More than 100 academics have written an open letter highligh(...)
## 212759 "Impulsar el talento, impulsar lo mejor que tiene nuestro pa(...)
## 5021   Kanzler ist völlig abgemeldet:\n FPÖ-Bundesparteioibmann HC St(...)
## 163107 Voormalig PvdA-minister Bos is als bewindspersoon in het laa(...)
```

* 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
<code>country_name</code>	[string]	Name of the country in small caps
<code>parlgov_id</code>	[integer]	Party ID (merge with <code>party_id</code> in ParlGov)
<code>party</code>	[string]	Abbreviation of the party name (ParlGov)
<code>party_name</code>	[string]	Party name (ParlGov)
<code>party_name_english</code>	[string]	Party name in English (ParlGov)
<code>family_name</code>	[string]	Party family name (ParlGov)
only in <code>monthly_agendas</code> :		
<code>month</code>	[yyyymm]	Month
<code>month_start</code>	[date]	Date of the first day of the month
<code>month_end</code>	[date]	Date of the last day of the month
only in <code>weekly_agendas</code> :		
<code>calendar_week</code>	[integer, 1-53]	Calendar week
<code>week_start</code>	[date]	Date of previous Monday
<code>week_end</code>	[date]	Date of following Sunday
<code>issue</code>	[integer]	Issue category
<code>issue_descr</code>	[string]	Issue category description
<code>share_multi</code>	[numeric]	Share of press releases of category <code>issue</code> of all <code>party_sum</code> predicted from Multilingual Transformer using all labeled docs from all countries
<code>share_mono</code>	[numeric]	Share of press releases of category <code>issue</code> of all <code>party_sum</code> predicted from Monolingual Transformer using all labeled docs from each country in separate models
<code>share_ridge</code>	[numeric]	Share of press releases of category <code>issue</code> of all <code>party_sum</code> predicted from Ridge using all labeled docs from each country in separate models
<code>share_super</code>	[numeric]	Share of press releases of category <code>issue</code> of all <code>party_sum</code> predicted from SuperLearner using all labeled docs from each country in separate models
<code>party_sum</code>	[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
## 113641      sweden      1546   SD 201606    99 0.00000000
## 52657      ireland      2217   SF 202001    10 0.02604167
## 113227      sweden      1546   SD 201412    99 0.00000000
## 115836        uk        311 Plaid 201406     8 0.00000000
## 63155  netherlands      742 PvdA 201309    98 0.03773585
## 58082  netherlands      345  D66 201309     7 0.07142857
## 29674      germany      558  SPD 201308     4 0.01562500
## 131119        uk      1556   Lab 201202    23 0.00000000
## 130084        uk      1284   SNP 201909    23 0.00000000
## 130102        uk      1284   SNP 201910    15 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 [9]
##   country_name parlgov_id party  week_start issue share_multi
##   <chr>         <dbl> <chr>    <date>    <dbl>    <dbl>
## 1 uk           467 Greens 2016-03-14     8      0
## 2 netherlands  357 SP    2013-01-21   192    0.0645
## 3 ireland      318 Labour 2010-12-27    14     0
## 4 sweden       892 L      2019-07-08     7     0
## 5 poland      2600 Kukiz'15 2016-04-18    23     0
## 6 uk           311 Plaid 2014-11-10    99     0
## 7 ireland     1573 GP     2015-04-13     1     0
## 8 poland      2600 Kukiz'15 2017-01-30     3     0
## 9 ireland     2776 S-PBP 2019-07-15     6     0
## 10 ireland     280 FF     2015-06-15    10    0.0385
```