

# PARTYPRESS datasets and variables

2022-11-18

## Contents

Main dataset <b>PARTYPRESS</b>	<b>2</b>
Issue names dataset <b>PARTYPRESS ISSUES</b>	<b>3</b>
Raw texts dataset <b>PARTYPRESS TEXTS</b>	<b>4</b>
Aggregated datasets <b>MONTHLY</b> and <b>WEEKLY AGENDAS</b>	<b>5</b>

# 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 <code>country_name</code>
parlgov_id	[integer]	Party ID (merge with <code>party_id</code> 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, $k = 1, \dots, 5$ , 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
## 55659    denmark 57731      306  RG (Ø) 2013-02-09         191
## 82407    germany 32427      558   SPD 2014-02-27           2
## 78860    germany 29235      558   SPD 2019-03-22           2
## 161688 netherlands 11149      235   CDA 2017-06-13          12
## 153245    ireland 55727     2217    SF 2020-09-22          12
```

## 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
## 115833      ireland 13398
## 101510      germany 44025
## 207621       spain  3431
## 10488      austria 10488
## 225965       spain 28577
## 167064 netherlands  7205
## 214809       spain 17364
## 199195      poland 14981
## 183047 netherlands 23188
## 229050       spain 31667
##
##                                     text
## 115833 Following a briefing this afternoon by Justice Minister Alan(...)
## 101510 Am 15. Januar 1990 folgten tausende mutige Menschen dem Aufr(...)
## 207621 El delegado territorial de Ciutadans (C's) en Barcelona, Man(...)
## 10488  FPÜ bietet gerne der SPÖ Nachhilfe punkto Tourismus an:\n Für(...)
## 225965 El portavoz de Constitucional considera "lógico" que el juez(...)
## 167064 Dennis de Jong Op 1 juni vindt er weer een 'European Busine(...)
## 214809 El diputado popular y coordinador de Organización del Partid(...)
## 199195 Przedstawiamy Panstwu kalendarz wystapien medialnych polityk(...)
## 183047 Vragen van het lid De Roon (PVV) aan de minister van Veiligh(...)
## 229050 De La Vega destaca que la futura Ley de Mediación servirá pa(...)
```

## 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
## 98076      sweden      657     M 201901    18 0.000000000
## 62526 netherlands      742   PvdA 201106    17 0.004310345
## 129917      uk      1284    SNP 201903     6 0.000000000
## 56004 netherlands      235    CDA 201211     3 0.023255814
## 72697 netherlands     1409    VVD 201411     3 0.000000000
## 6036      austria     1013    ÖVP 201101    10 0.009493671
## 132772      uk      1556    Lab 201803     9 0.005639098
## 10863      austria     1429 GRÜNE 201802     7 0.000000000
## 35010      germany      791   Left 201311     8 0.007281553
## 94880      spain     2375     Cs 201505    15 0.002747253
```

```
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           773 Con    2020-10-05  191      0
## 2 spain        645 PP     2015-08-17  191      0
## 3 germany     1727 CDU/CSU 2013-11-11   1    0.107
## 4 netherlands  742 PvdA    2015-03-30  23      0
## 5 uk          1556 Lab     2013-08-05   6      0
## 6 uk          1556 Lab     2014-02-10   2    0.00610
## 7 sweden     1154 MP      2010-06-28  10      0
## 8 poland      629 SLD     2013-03-25  18      0
## 9 denmark    2567 A (Å)   2016-07-18   2      0
## 10 netherlands 1501 PVV    2012-01-02  16      0
```