

PARTYPRESS datasets and variables*

Cornelius Erfort, Lukas F. Stoetzer, Heike Klüver

2022-10-05

Contents

Main dataset PARTYPRESS	2
Issue names dataset PARTYPRESS ISSUES	3
Raw texts dataset PARTYPRESS TEXTS	3
Aggregated datasets MONTHLY and WEEKLY AGENDAS	4

*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. We thank Leonie Fuchs, Paul Bochtler, Daniel Cruz, Johannes Lattmann, Felix Heimburger, Violeta Haas and Tim Wappenhans for valuable research assistance. Replication material can be found on <https://github.com/cornelius-erfort/scripts-issue-agendas>.

Main dataset PARTYPRESS

Load files `partypress.csv` (52MB) or `partypress.RData` (10MB).

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 (ParlGov)
<code>party</code>	[string]	Abbreviation of the party 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>	[yyyy-mm]	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 BERT using all labeled docs from all countries
<code>issue_ridge</code>	[integer, 1-23]	Issue category predicted from Ridge 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

Issue names dataset PARTYPRESS ISSUES

Load files `partypress_issues.csv` (1KB) or `partypress_issues.RData` (1KB).

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

Raw texts dataset PARTYPRESS TEXTS

Load files `partypress_issues.csv` (541MB) or `partypress_issues.RData` (189MB).

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

Aggregated datasets MONTHLY and WEEKLY AGENDAS

Load files `monthly_agendas.csv` (5MB)/`weekly_agendas.csv` (15MB) or `monthly_agendas.RData` (1MB)/`weekly_agendas.RData` (2MB).

Variables

Variable name	Format	Description
<code>country_name</code>	[string]	Name of the country in small caps
<code>parlgov_id</code>	[integer]	Party ID (ParlGov)
<code>party</code>	[string]	Abbreviation of the party name (ParlGov)
<code>month_start</code>	[date]	Date of the first day of the month, only in <code>monthly_agendas</code>
<code>month_end</code>	[date]	Date of the last day of the month, only in <code>monthly_agendas</code>
<code>week_start</code>	[date]	Date of previous Monday, only in <code>weekly_agendas</code>
<code>week_end</code>	[date]	Date of following Sunday, only in <code>weekly_agendas</code>
<code>issue_multi</code>	[integer]	Issue category predicted from multilingual BERT using all labeled docs from all countries
<code>issue_descr</code>	[string]	Issue category description
<code>attention</code>	[numeric]	Share of press releases of category <code>issue_multi</code> of all <code>party_sum</code>
<code>party_sum</code>	[integer]	Total number of press releases by this party in this month/week