The Comparative Legislators Database

Online Supplementary Information

Published in *British Journal of Political Science*

Sascha Göbel (corresponding author)
University of Konstanz
Department of Politics and Public Administration
Universitätsstraße 10
D-78464 Konstanz, Germany
sascha.goebel@uni-konstanz.de

Simon Munzert
Hertie School
Friedrichstr. 180
D-10117 Berlin, Germany
munzert@hertie-school.org

Contents

Appendix A Literature Review

The survey was conducted with the support of three research assistants. We conducted a search in the archives of four flagship general-interest journals of political science, the *American Political Science Review*, the *American Journal of Political Science*, the *Journal of Politics*, and the *British Journal of Political Science*, as well as the official journal of the Legislative Studies Section of the American Political Science Association, *Legislative Studies Quarterly*. Using JSTOR's search mask, we used the query "`(MP OR legislator) AND data`" for all fields, filtered for articles only as item type, and restricted the search to articles published in a period of ten years, between 2009 and 2018. For periods not covered by JSTOR, we referred to the search function for the respective archive hosted by the journal's publisher. Overall, this search identified 535 articles. In the next step, we screened the abstract of every article to identify studies that potentially used individual-level data of members of national-level legislative bodies. This resulted in 225 articles. In the third step, three research assistants inspected the body of the articles and collected information on the data sources used by the studies as well as their type (whether the data was provided by governmental sources or non-governmental institutions, journalistic data sources, databases provided by other scholars, or whether the study involved an original data collection effort), countries and years covered, and the types of features that were used. Here, we distinguished between sociodemographics (e.g. age, gender, ethnicity), political behavior (e.g. roll call votes, ideology measures), office-related variables (e.g. tenure, committee memberships, leadership positions), and constituency-specific variables (e.g. election results of legislator's district, district public opinion, district sociodemographics). 16 studies turned out to be irrelevant in the last coding step.

Figure A1: Usage of data on national political representatives over time.

Note: Based on a survey of articles published in five top political science journals between 2009 and 2018.

Appendix B Details on Data Collection

After identifying parliamentary members' index sites on Wikipedia, we downloaded the respective HTML files. Next, we created legislature-specific scraper functions with the XPath query language. These functions were used on the HTML files to automatically extract basic data (name, party, constituency, period of service, and session) on representatives and URLs to their personal Wikipedia pages. The URLs were then applied to gain access to the individual Wikipedia page and Wikidata IDs via Wikipedia's API. Using several custom-built API bindings, we finally fed the collected IDs to the APIs of Wikipedia, Wikimedia, and Wikidata to automatically extract all remaining data. For the Wikipedia and Wikimedia APIs, we simply set the parameters to include the desired information (traffic, revision histories, and portraits). For the Wikidata API, we tapped content-specific properties pre-specified by Wikidata (IDs, offices, professions, social media, and all remaining sociodemographic data).

Following the automated data extraction on Wikipedia and Wikidata, we identified other sources to fill in some gaps in the data. For several countries, we filled missing religious affiliation by gathering information from governmental databases or other Wikipedia entries

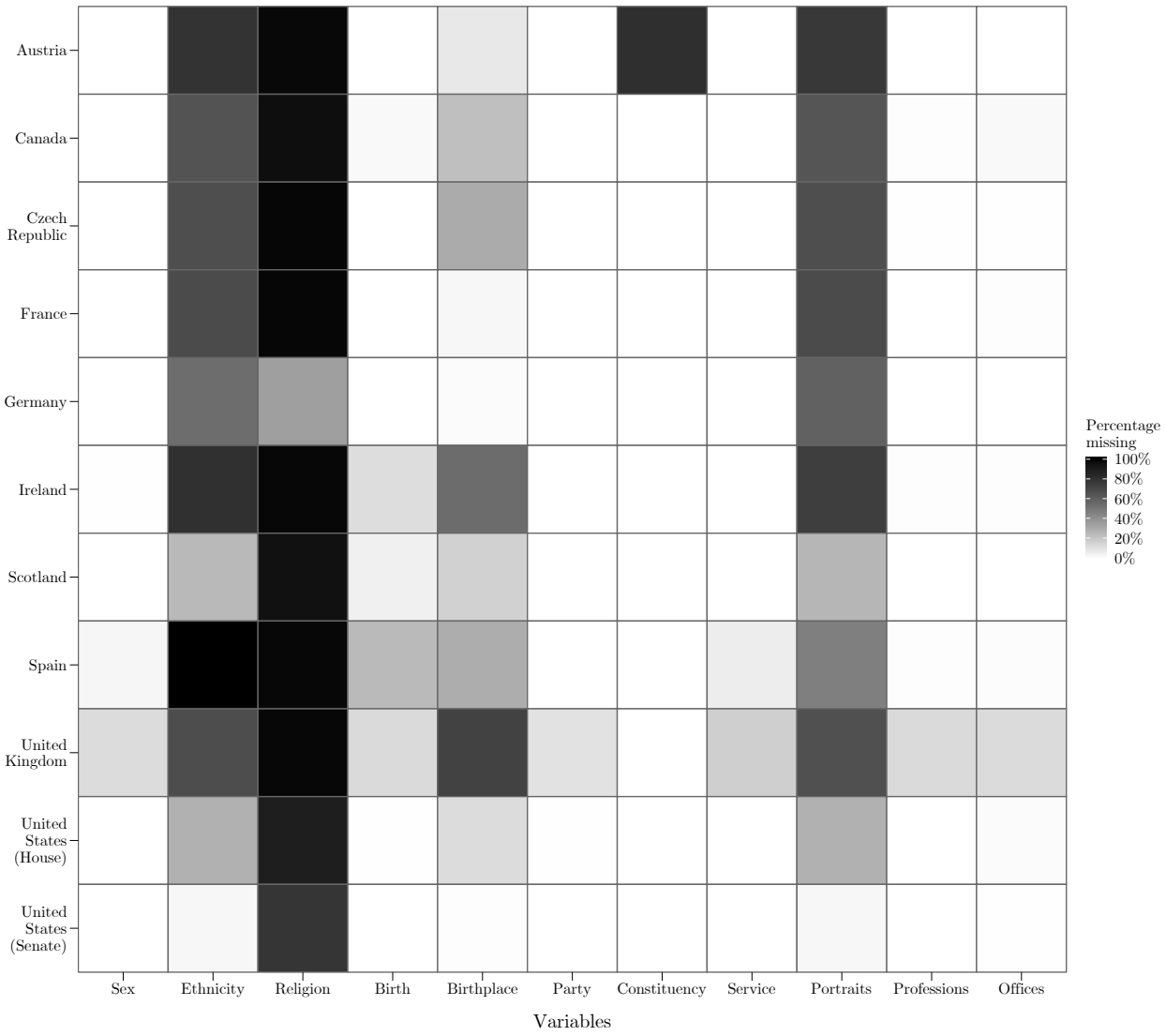
and websites. We proceeded, as above, by downloading HTML or XML files and building custom scraper functions to automatically extract the required information. Missing periods of service and social media data were gathered the same way. Period of service data was, in part, also manually extracted by the research assistants from notes on the parliamentary members' index sites. Social media data was, in part, shared with us by colleagues. Missing ethnicity information was partially coded by applying facial recognition software (Face++ API) to collected portraits. If the estimated accuracy was below 95 percent, we recoded ethnicity manually. Government membership and leader positions were manually collected by the research assistants in the form of relevant Wikidata IDs from the index and election-specific Wikipedia sites.

IDs for political science datasets were not collected via Wikipedia or Wikidata. Instead, we and the research assistants wrote scripts to match these datasets to our database. A combination of variables, such as session, name, date of birth, and party affiliation was used to uniquely link politicians. For politicians that could not be matched this way, the research assistants tried to achieve a match via manual research.

The formatting and harmonization of variables within and across legislatures was conducted by applying regular expressions and by grouping information. For instance, religious affiliation was manually grouped into major religious groups and Christian denominations. Messy party affiliations were cleaned and aligned by removing and replacing character sequences using regular expressions.

Appendix C Data Coverage, Quality, and Verification

Figure C1: Percentage of missings for selected variables by country.



Note: Country, pageid, wikidataid, wikititle, name, session, session start, session end, wikipedia traffic and wikipedia history are complete by definition or due to the way we approach data collection, only very few do not have a pageid, even fewer do not have a wikidataid. Variables on which missingness is not necessarily due to lack of data availability were omitted from the assessment of missings (e.g. death and deathplace, government membership, leadership positions, constituency affiliation, and social media profiles.)

Table C1: Missing data analysis for selected countries.

United States (House)				
	Birthplace	Ethnicity	Portrait	Religion
Female	1.5%	1.2%	1.2%	57.8%
Male	12.4%	28.7%	28.7%	89.8%
Black	1.4%	–	0%	52.1%
Hispanic	3.3%	–	0%	62.5%
Other	2.9%	–	0%	58.8%
White	8.6%	–	0%	86.4%
Recency	0.84	0.74	0.74	0.57
United Kingdom				
	Birthplace	Ethnicity	Portrait	Religion
Female	20%	29.1%	29.3%	93.5%
Male	70.6%	63.7%	63.5%	97.9%
Asian	10.4%	–	12.5%	58.3%
Black	9.1%	–	9.1%	95.5%
White	46.7%	–	0%	96.6%
Recency	0.73	0.89	0.89	0.74
France				
	Birthplace	Ethnicity	Portrait	Religion
Female	1.2%	46%	46%	99.0%
Male	2.9%	71.8%	71.8%	98.4%
Black	4.5%	–	0%	81.8%
Other	0%	–	0%	100%
White	0.3%	–	0%	96.5%
Recency	0.83	0.76	0.76	1.1

Note: For the discrete variables sex and ethnicity the group-specific share of missings with reference to the column variables is reported. The continuous variable recency measures the time difference in years between legislators' birthdate and the current time at the date of computation. Reported is the ratio of recency means for missing and observed data with reference to the column variables. A ratio smaller than 1 indicates recency bias.

For many countries Wikipedia seems to offer comprehensive lists of members of parliament. To evaluate this comprehensiveness we compared the CLD to two integrated datasets that are based on hand-coding of official and comprehensive administrative records, the German BTVote dataset by Sieberer et al. (2020) and the UK dataset by Eggers and Spirling (2014). Focusing on the periods of observation that match between datasets, BTVote includes 99.9 percent (we integrated 99.9 percent of BTVote) and the Eggers and Spirling data includes 98.9 percent (we integrated 99.1 percent of the Eggers and Spirling dataset) of the legislators in the CLD. We additionally compare the CLD to Public Whip, another more topic specific single-country dataset. Here, we find that focusing only on the period covered by Public Whip, the number of legislators in the project amounts to 92 percent of the MPs listed in the CLD. These comparisons highlight the benefit of using Wikidata and Wikipedia as data source for full coverage of legislators.

Our main data source, Wikipedia, is among the ten most-visited websites worldwide (See https://www.alexa.com/siteinfo/wikipedia.org. Last accessed February 2020), which emphasizes its importance as a source of information. The open-collaboration platform is a long-standing online encyclopedia, where information can be created and revised by anyone at any time. Wikipedia asks its editors to maintain a neutral point of view when editing content and employs various bots for picking up and correcting biased content (See https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view. Last accessed February 2020). Indeed, studies confirm the accuracy of information offered on Wikipedia in comparison to traditional encyclopedias (Giles, 2005). While this encourages trust in the CLD, it does not provide explicit verification of the very specific information the CLD offers. Furthermore, several studies indicate that information on Wikipedia can fall prey to political biases (Göbel and Munzert, 2018; Kalla and Aronow, 2015). For this reason, we conducted several verification checks, detailed in the following paragraphs.

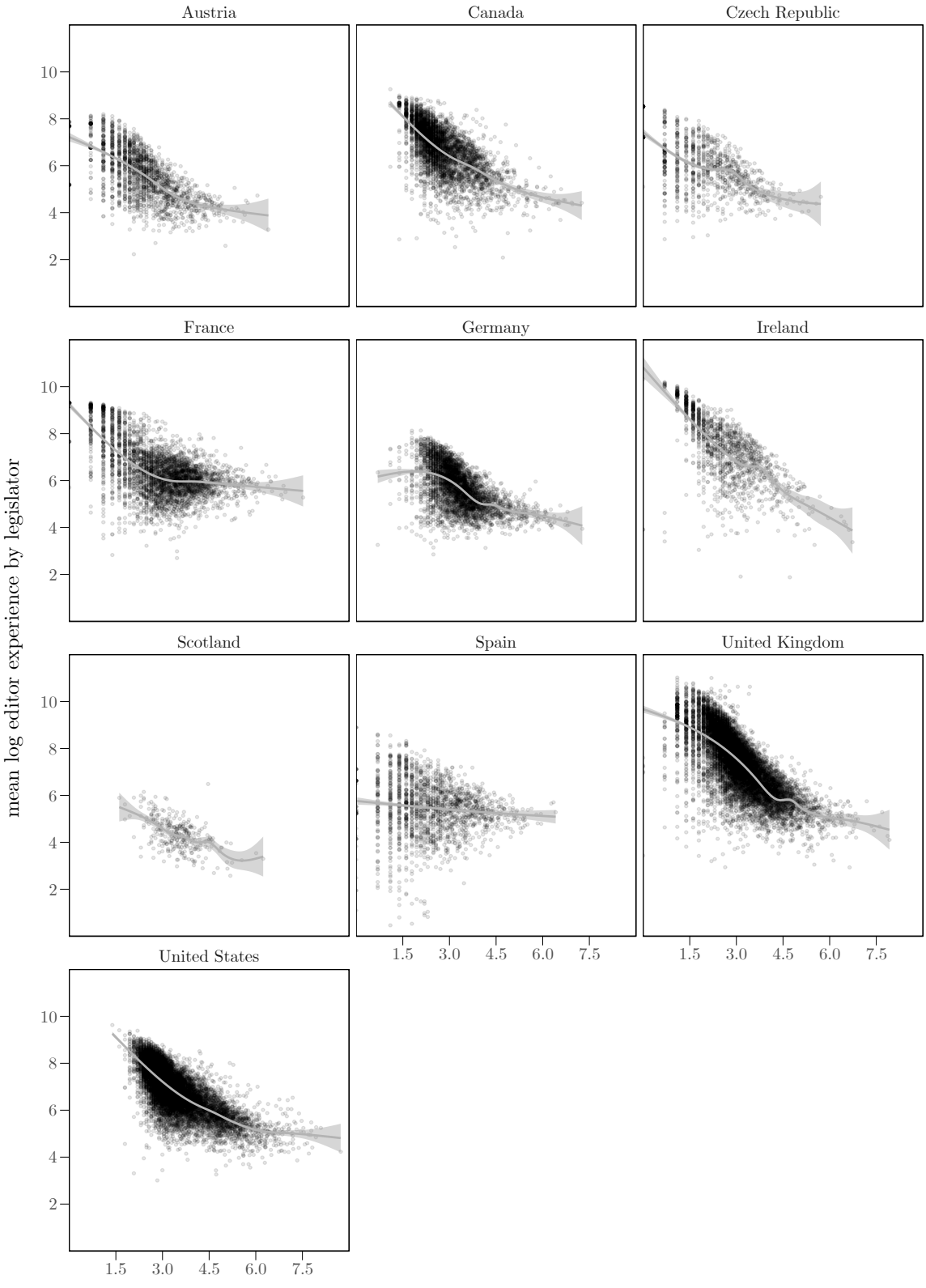
Recent research suggests that both the number of editors and the amount of editorial experience are related to the degree of biased information on the encyclopedia (Greenstein, Gu and Zhu, 2017). Contributors were found to target content that was especially slanted way from their viewpoint and enter into dialogue with each other during editing. This indicates heterogeneous conversations that drive a bias-reducing environment. Naturally,

the more participants involved in editing an article, the more this collective intelligence can thrive, and the more we can trust the information it generates. This lends skepticism towards articles that attract only a few editors. However, the study found that editorial experience plays an important role as well. It shows that contributors grow more moderate in their editing behavior, i.e. the information they produce is less slanted and biased, the more articles they edit.

Figure C2 shows the amount of editorial experience and the number of editors for each legislator's Wikipedia page in the database. We measure domain-specific editorial experience as the amount of legislators' Wikipedia pages an editor contributed to within the respective legislature. We further weigh this amount by a contributor's mean number of edits. Plotting the measures against each other, we find that legislators' Wikipedia pages with few editors usually do attract more experienced contributors. As the number of editors on an article increases, the average amount of editorial experience decreases. Hence, editorial experience steps in when collective intelligence does not and vice versa.

To further verify the quality and accuracy of the specific information we offer, we compared part of the database against information available in other data projects, which relied on human coders. We picked two projects integrated with the CLD: (1) the Bundestag Roll Call Vote Data (BTVote) (Sieberer et al., 2020), which includes roll call votes taken in the German Bundestag from 1949 to 2013, together with several personal characteristics of the respective legislators, and (2) Voteview (Lewis et al., 2019), which provides roll call votes and basic biographical information of legislators for all chambers of the United States Congress. We chose these datasets for comparison because of their longitudinal scope, and because data on roll call votes are among the most prominent and frequently used in political science. In addition, their coder-based mode of data collection is the gold standard in terms of data accuracy. A student assistant joined both datasets with the CLD using the identifiers offered in the CLD's *IDs* table, harmonized variables available in both the CLD and the datasets to facilitate comparison, and filtered information where the CLD and the datasets disagree. In a case of disagreement, the student assistant searched for official information and noted whether the value in the CLD or the value in the other datasets is correct. Figure C3 presents results from the verification exercise. On the compared variables, the CLD

Figure C2: Editor experience and number of editors on legislators' Wikipedia pages.



shows very high agreement with information offered in BTVote and Voteview. Disagreement is either resolved in favor of the CLD or tied between the CLD and the other datasets. This shows that our crowd-sourced approach mirrors the human gold standard in terms of accuracy of information.

Finally, a student assistant conducted a Google search of official sources for a random selection of 500 legislators from the CLD to assess both the availability and accuracy of information. The assistant was instructed to search for "[legislator's name] member of [country] parliament" on Google and screen the first landing page for official information. We considered pages such as personal, parliamentary, governmental, or campaign websites as official. Other sources such as Wikipedia and Wikidata were omitted. In case no official page was available for a legislator, the assistant resorted to the Google info box to verify information. However, this happened only for a handful of legislators. Usually, there was either an official source or none at all. Once official sources were located on the first landing page, the student assistant searched those pages for a set of individual information, including sex, date of birth and death, ethnicity, religion, birthplace, place of death, legislative session, party, constituency, period of service. If the respective information was traceable and included in the CLD, the student assistant noted whether it confirmed (C) or disconfirmed (D) information in the CLD. If the information was traceable but not included in the CLD, it was similarly categorized as disconfirmed. If information was not traceable through official sources on the first landing page but available in the CLD, it was categorized as not found (NF). Figure C4 shows that for most variables, information in the CLD is confirmed by official sources. Interestingly, the share of information in the CLD that was disconfirmed by official sources is negligibly small, except for places of birth and death, and periods of service. However, the bulk of disconfirmed information for these variables arises because the information was traceable via official sources but not available in the CLD. For the other variables, if information could not be verified it was primarily because it was not traceable via official sources. This supports both the accuracy and coverage of the CLD.

Figure C3: Comparison of CLD to human gold standards on selected variables.

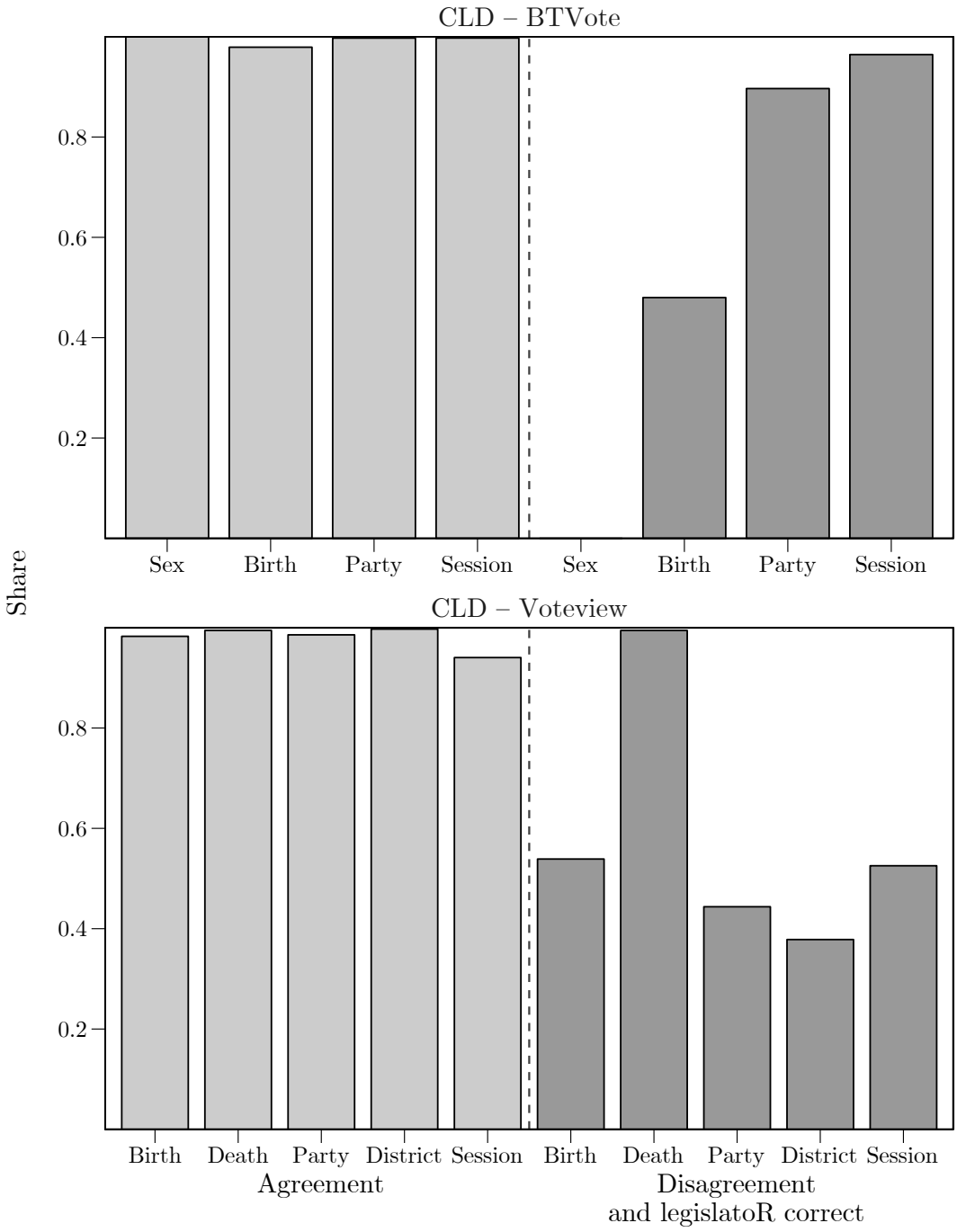
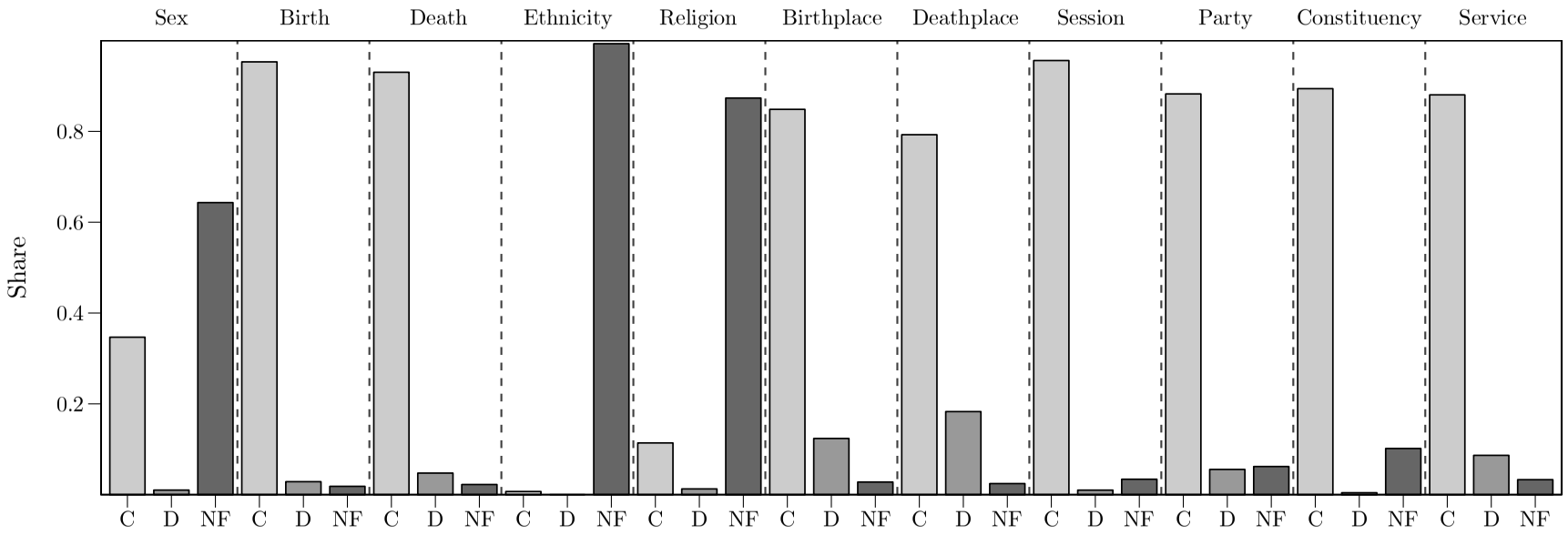


Figure C4: Google search for information on 500 randomly drawn legislators in CLD.



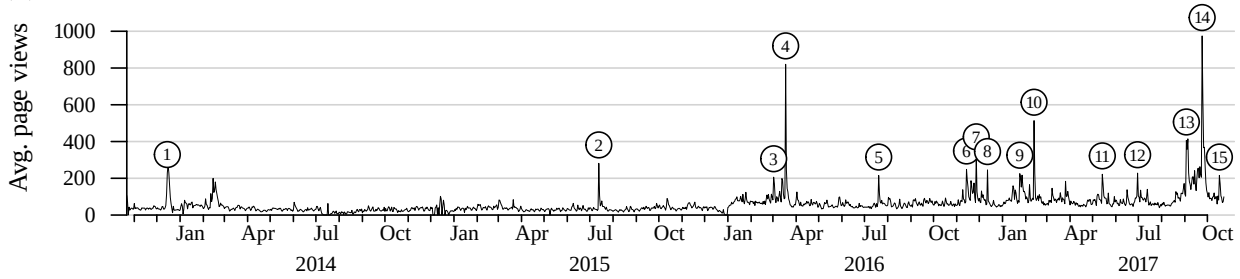
Note: C = Confirmed, D = Disconfirmed, NF = Not found. Information not available in CLD but found through a Google Search was categorized as disconfirmed before computing shares. Otherwise, information not available in CLD is excluded here.

Appendix D Application: Tracking Public Attention Paid to Legislators with Wikipedia Page Views: Additional materials

In this section we provide additional materials of the page views analysis reported in the main text. Figures D1 to D8 provide replications for the most recent completed sessions in the other legislatures provided by the CLD.

Figure D1: Descriptive statistics and predictive model of Wikipedia page views of members of the 17th German Bundestag.

(a) Time series of aggregated average of Wikipedia page views.



1. Appointed first female defence minister (Ursula von der Leyen)
2. Deceased (Philipp Mißfelder)
3. Drug affair (Volker Beck)
4. Deceased (Guido Westerwelle)
5. Bullying affair (Petra Hinz)
6. Candidacy for presidency (Frank−Walter Steinmeier)
7. Deceased (Peter Hintze)
8. Unidentifiable cause (Thomas Strobl)
9. Chancellorchip announcement (Sigmar Gabriel)
10. Elected federal president (Frank−Walter Steinmeier)
11. State election North−Rhine Westphalia (Christian Lindner)
12. Policy success (Volker Beck)
13. TV debate (Angela Merkel)
14. General election (Christian Lindner)
15. Threat to resign (Sahra Wagenknecht)

(b) Top/bottom 10 mean daily page views for members of the 17th German Bundestag (2013–2017).

Rank	Legislator	Mean	Maximum
1	Angela Merkel	3962	87232
2	Sahra Wagenknecht	1899	71415
3	Ursula von der Leyen	1232	53293
4	Frank-Walter Steinmeier	1103	189365
5	Sigmar Gabriel	1051	56427
6	Wolfgang Schäuble	916	23966
7	Gregor Gysi	756	16593
8	Thomas de Maizière	688	10534
9	Andrea Nahles	619	54492
10	Katrin Göring-Eckardt	613	36915
647	Dirk Vöpel	3	91
648	Marina Kermer	3	43
649	Alexander Funk	3	66
650	Karin Thissen	3	82
651	Udo Schiefner	3	60
652	Marion Herdan	2	177
653	Iris Ripsam	1	268
654	Rainer Hajek	1	64
655	Thomas Jepsen	0	28
656	Karl-Heinz Wange	0	59

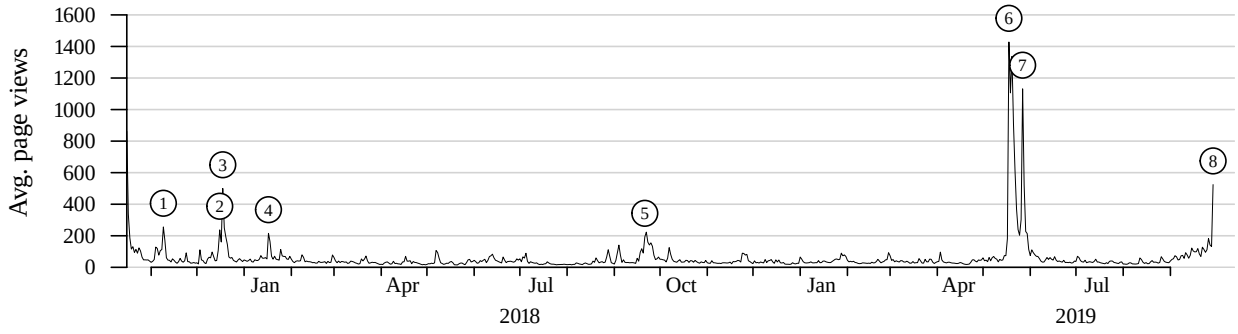
(c) OLS estimates of log page views.

	Log page views
Sessions served	5.81 (0.75)***
Party: CDU	0.78 (0.07)***
Party: CSU	−0.48 (0.13)***
Party: Left	−0.32 (0.17)
Party: None	−0.13 (0.16)
Party: SPD	2.34 (0.93)*
Office: Secretary	−0.60 (0.13)***
Office: Mayor	2.38 (0.20)***
Office: Party Chairman	0.04 (0.15)
Male	1.89 (0.35)***
Dead	−0.12 (0.08)
Age	1.52 (0.42)***
(Intercept)	−0.92 (0.19)***
R^2	0.44
Adj. R^2	0.43
Num. obs.	655

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D2: Descriptive statistics and predictive model of Wikipedia page views of the 26th National Council of Austria.

(a) Time series of aggregated average of Wikipedia page views.



1. Inauguration (President of the National Council) (Köstinger Elisabeth)
2. Inauguration (Chancellor) (Kurz Sebastian)
3. Inauguration of government (Kurz Sebastian)
4. Press conference with Angela Merkel (Kurz Sebastian)
5. Announcement of party leadership (Rendi−Wagner Pamela)
6. Ibiza affair (Strache Heinz−Christian)
7. Motion of no confidence (Kurz Sebastian)
8. Federal election (Kurz Sebastian)

(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Kurz Sebastian	2976	104659
2	Strache Heinz-Christian	1411	126748
3	Rendi-Wagner Pamela	844	34332
4	Kickl Herbert	722	36444
5	Gudenus Johann	675	68013
6	Kern Christian	404	11464
7	Hofer Norbert	390	19902
8	Köstinger Elisabeth	384	24718
9	Meinl-Reisinger Beate	326	8806
10	Pilz Peter	267	15566
201	Sandler Birgit	3	41
202	Keck Dietmar	3	53
203	Greiner Karin	2	38
204	Smodics-Neumann Maria	2	65
205	Hofinger Manfred	2	24
206	Wimmer Petra	2	51
207	Unterrainer Maximilian	2	26
208	Singer Johann	2	12
209	Bacher Walter	2	17
210	Yilmaz Nurten	1	38

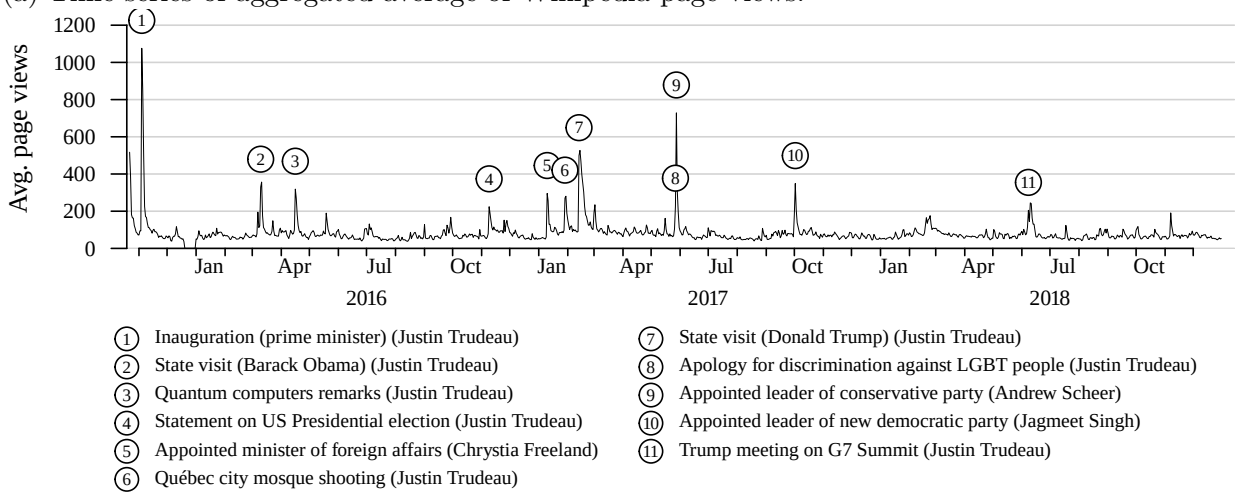
(c) OLS estimates of log page views.

	Log page views
Sessions served	5.34 (1.44)***
Party: NEOS	0.19 (0.15)
Party: ÖVP	1.12 (0.37)**
Party: PILZ	−0.24 (0.20)
Party: SPÖ	1.17 (0.42)**
Office: Secretary	−0.45 (0.22)*
Office: Mayor	2.74 (0.41)***
Office: Party Chairman	−0.87 (0.83)
Male	3.75 (0.86)***
Age	−0.04 (0.18)
(Intercept)	−0.79 (0.38)*
R^2	0.37
Adj. R^2	0.34
Num. obs.	210

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D3: Descriptive statistics and predictive model of Wikipedia page views of members of the 42nd Canadian Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Justin Trudeau	12096	1047864
2	Stephen Harper	1959	116704
3	Jagmeet Singh	911	79456
4	Andrew Scheer	868	175599
5	Chrystia Freeland	739	26486
6	Harjit Sajjan	720	40104
7	Rona Ambrose	553	22226
8	Maxime Bernier	392	15432
9	Jason Kenney	301	7367
10	Bill Morneau	278	21717
345	Rosemarie Falk	8	984
346	François Choquette	8	118
347	Pierre Breton	7	90
348	Stéphane Lauzon	7	66
349	Rémi Massé	7	95
350	Steven MacKinnon	6	76
351	Richard Hébert	5	795
352	Michael Barrett	5	807
353	Michel Picard	4	144
354	Ron McKinnon	2	35

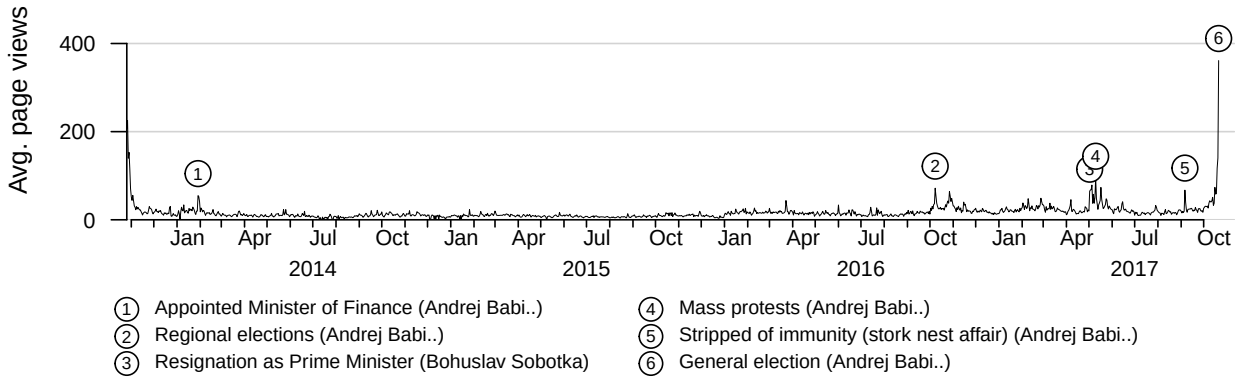
(c) OLS estimates of log page views.

	Log page views
Sessions served	5.12 (0.84)***
Party: Conservative	0.34 (0.09)***
Party: Green	0.50 (0.30)
Party: Liberal	2.87 (0.86)***
Party: New Democratic	0.66 (0.29)*
Office: Secretary	0.72 (0.31)*
Office: Mayor	1.48 (0.15)***
Office: Party Chairman	−0.03 (0.41)
Male	1.11 (0.27)***
Age	−0.24 (0.11)*
(Intercept)	−0.72 (0.20)***
R^2	0.46
Adj. R^2	0.44
Num. obs.	319

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D4: Descriptive statistics and predictive model of Wikipedia page views of members of the 7th Czech Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Andrej Babiš	492	17658
2	Bohuslav Sobotka	200	5265
3	Karel Schwarzenberg	169	5327
4	Martin Stropnický	158	4039
5	Miroslav Kalousek	152	5929
6	Tomio Okamura	138	17044
7	Milan Chovanec	85	1878
8	Daniel Herman	83	8910
9	Věra Jourová	78	2720
10	Petr Fiala	71	3266
203	Igor Nykl	1	20
204	Pavel Šrámek	1	68
205	Josef Nekl	1	12
206	Karel Šidlo	1	69
207	Pavel Volčík	1	13
208	Pavel Čihák	1	35
209	Šenfeld Josef	1	15
210	Karel Černý	1	12
211	Miroslava Strnadlová	1	12
212	René Číp	1	47

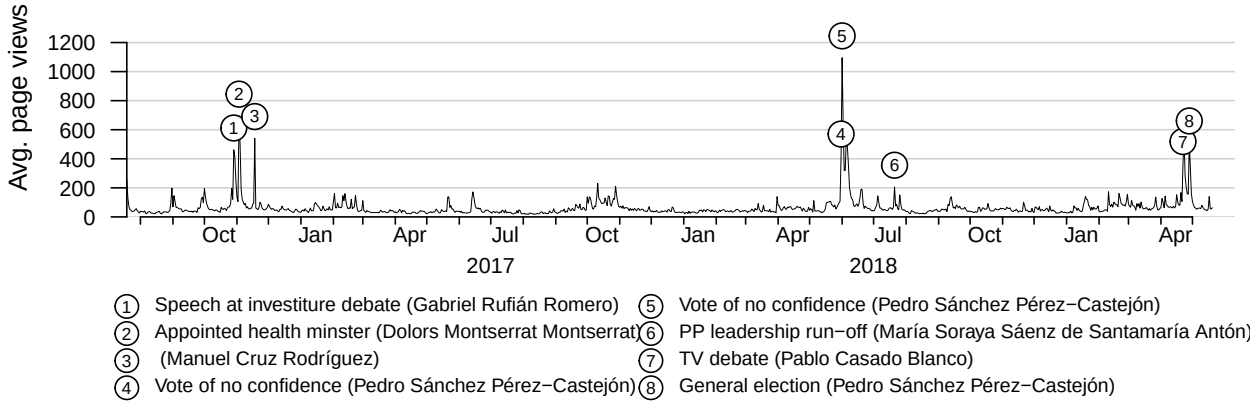
(c) OLS estimates of log page views.

	Log page views
Sessions served	4.06 (1.28)**
Party: CSSD	0.70 (0.13)***
Party: KDU-CSL	−0.45 (0.18)*
Party: KSCM	0.15 (0.27)
Party: ODS	−0.65 (0.22)**
Party: TOP 09	0.42 (0.26)
Party: Usvit	0.09 (0.22)
Office: Secretary	0.51 (0.27)
Office: Mayor	2.59 (0.25)***
Male	0.03 (0.19)
Age	0.05 (0.16)
(Intercept)	−0.70 (0.32)*
R^2	0.47
Adj. R^2	0.44
Num. obs.	212

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D5: Descriptive statistics and predictive model of Wikipedia page views of members of the 12th Spanitsh Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Mariano Rajoy Brey	1964	94603
2	Pedro Sánchez Pérez-Castejón	1898	156559
3	Pablo Iglesias Turrión	1326	35121
4	Albert Rivera Díaz	1106	45767
5	María Soraya Sáenz de Santamaría Antón	999	30793
6	Irene María Montero Gil	946	28056
7	Íñigo Errejón Galván	946	24663
8	Gabriel Rufián Romero	915	35124
9	María Dolores De Cospedal García	625	21777
10	Pablo Casado Blanco	583	45494
362	Sebastián Franquis Vera	1	60
363	Soledad Amada Velasco Baides	1	12
364	María Luz Bajo Prieto	1	35
365	Juan Jiménez Tortosa	1	49
366	Jesús Postigo Quintana	1	13
367	Alejandro Ramírez del Molino Morán	1	48
368	María Dolores Bolarín Sánchez	1	16
369	Luis Miguel Salvador García	1	6
370	Carlota Merchán Mesón	0	8
371	Sergi Miquel i Valentí	0	2

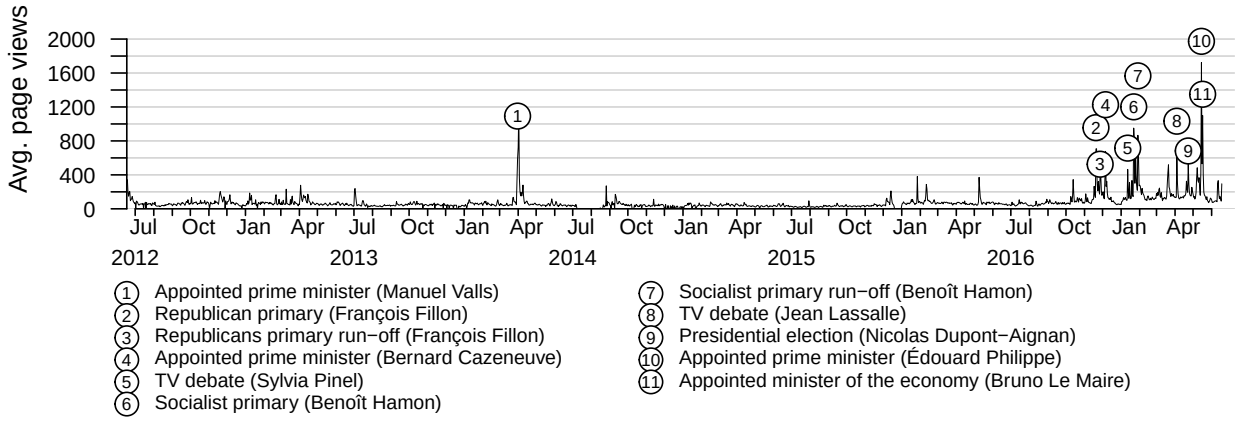
(c) OLS estimates of log page views.

	Log page views
Sessions served	3.78 (1.69)*
Party: Cs	0.97 (0.16)***
Party: PPL	0.25 (0.33)
Party: PSOE	−0.35 (0.22)
Party: UP	−0.43 (0.25)
Office: Secretary	0.82 (0.31)**
Office: Mayor	2.86 (0.34)***
Male	0.26 (0.26)
Age	0.20 (0.17)
(Intercept)	−0.75 (0.44)
R^2	0.28
Adj. R^2	0.27
Num. obs.	371

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D6: Descriptive statistics and predictive model of Wikipedia page views of members of the 14th French Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Manuel Valls	2238	159999
2	François Fillon	2181	328466
3	Benoît Hamon	1702	214296
4	Nathalie Kosciusko-Morizet	1049	32238
5	Nicolas Dupont-Aignan	964	102560
6	Jean-François Copé	939	48645
7	François Baroin	923	52777
8	Édouard Philippe	912	892494
9	Bruno Le Maire	781	159066
10	Jean Lassalle	744	225165
600	Romain Joron	3	597
601	Émeric Bréhier	3	94
602	Élisabeth Pochon	3	78
603	David Comet	2	144
604	Guy Bailliart	2	125
605	Monique Lubin	2	498
606	Marion Maréchal-Le Pen	2	43
607	Dominique Potier	1	47
608	Napole Polutélé	1	80
609	Pierre Morel-A-L'Huissier	0	3

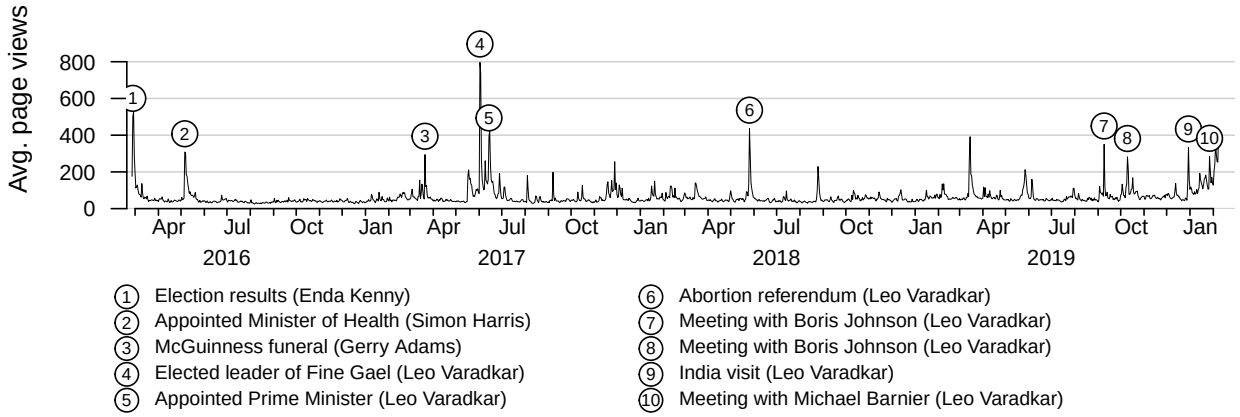
(c) OLS estimates of log page views.

	Log page views
Sessions served	10.20 (1.03)***
Party: GDR	0.68 (0.08)***
Party: LR	−1.02 (0.31)**
Party: PS	−1.09 (0.19)***
Party: RRDP	−0.41 (0.33)
Party: SER	−1.13 (0.29)***
Party: SRC	−1.26 (0.18)***
Party: UDI	−0.24 (0.48)
Office: Secretary	−1.08 (0.25)***
Office: Mayor	2.26 (0.21)***
Male	−0.04 (0.36)
Age	0.11 (0.10)
(Intercept)	−1.71 (0.26)***
R^2	0.39
Adj. R^2	0.37
Num. obs.	608

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D7: Descriptive statistics and predictive model of Wikipedia page views of members of the 32nd Irish Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Leo Varadkar	2511	100082
2	Gerry Adams	1218	34861
3	Enda Kenny	799	12980
4	Mary Lou McDonald	393	15338
5	Richard Boyd Barrett	327	4197
6	Simon Coveney	216	7136
7	Simon Harris	184	11967
8	Eoghan Murphy	182	3868
9	Micheál Martin	156	4847
10	Frances Fitzgerald	144	16390
153	James Browne	8	69
154	Fiona O'Loughlin	8	68
155	Pat Casey	8	180
156	Eugene Murphy	8	66
157	Shane Cassells	8	155
158	Jackie Cahill	8	97
159	Aindrias Moynihan	8	97
160	Frank O'Rourke	8	146
161	Kevin O'Keeffe	7	170
162	Michael McGrath	4	113

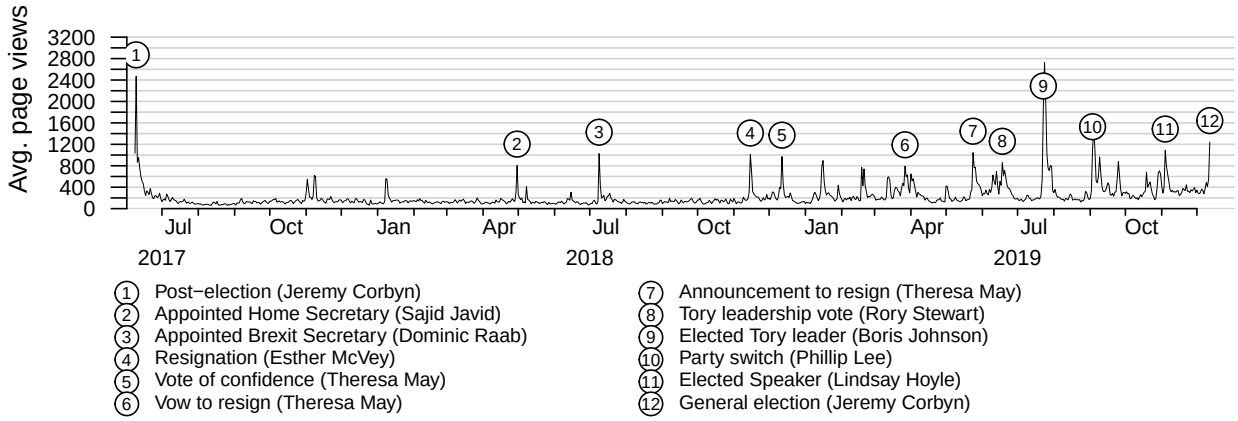
(c) OLS estimates of log page views.

	Log page views
Sessions served	6.59 (1.28)***
Party: Fianna Fáil	0.19 (0.12)
Party: Fine Gael	−0.66 (0.22)**
Party: Labour	−0.32 (0.22)
Party: Sinn Féin	−1.02 (0.34)**
Party: Independent	−0.11 (0.24)
Office: Secretary	0.13 (0.25)
Office: Mayor	0.68 (0.17)***
Office: Party Leader	1.06 (0.54)
Male	1.71 (0.28)***
Age	−0.27 (0.15)
(Intercept)	−0.83 (0.32)*
R^2	0.49
Adj. R^2	0.45
Num. obs.	159

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Figure D8: Descriptive statistics and predictive model of Wikipedia page views of members of the 57th UK Parliament.

(a) Time series of aggregated average of Wikipedia page views.



(b) Top/bottom 10 mean daily page views.

Rank	Legislator	Mean	Maximum
1	Boris Johnson	12375	733401
2	Theresa May	9345	233660
3	Jeremy Corbyn	6076	265000
4	Jacob Rees-Mogg	5069	86738
5	John Bercow	4206	191713
6	Priti Patel	2444	274782
7	Rory Stewart	2388	184497
8	Sajid Javid	2337	213938
9	Jo Swinson	2210	108260
10	Jeremy Hunt	2017	103463
646	Marcus Jones	21	368
647	Ronnie Cowan	20	189
648	Pat McFadden	20	985
649	John McNally	19	441
650	Gordon Henderson	19	344
651	Steve McCabe	18	1627
652	Christina Rees	16	1630
653	Nigel Mills	15	191
654	Colleen Fletcher	15	148
655	Susan Elan Jones	12	95

(c) OLS estimates of log page views.

	Log page views
Sessions served	7.52 (0.75)***
Party: Conservative	0.22 (0.10)*
Party: Labour	0.30 (0.13)*
Party: LibDem	−0.07 (0.14)
Party: SNP	0.77 (0.29)**
Office: Secretary	−0.46 (0.20)*
Office: Mayor	0.96 (0.11)***
Office: Party Leader	0.83 (0.47)
Male	0.72 (0.17)***
Age	−0.34 (0.08)***
(Intercept)	−0.86 (0.21)***
R^2	0.28
Adj. R^2	0.27
Num. obs.	651

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Appendix E Application: Tracking Women's Descriptive Representation and Network Centrality in Parliaments

In this section, we present another application using data from the CLD to provide evidence on the descriptive and substantive representation of women in parliaments. Research on female political representation has highlighted discrepancies between descriptive and substantive representation, i.e. the difference between a mere existence of women in parliament and their actual effectiveness (Wängnerud, 2009). Presence in parliament poses a pre-condition of substantive representation of women that would also reflect in bringing in women's interests and actual influence in policymaking by female representatives. Previous studies have used a diverse set of indicators for substantive representation, including but not limited to data on committee membership, bill co-sponsorship, voting behavior, and parliamentary speeches (see Kroeber, 2018, for an overview).

Here, we suggest an alternative proxy to approach substantive representation in parliament using some of the data that is made easily accessible with the CLD. In particular, we draw on the valuable feature that content is heavily linked on the Wikipedia. Links between articles are primarily manually defined by human editors contributing to the project. These links are not set randomly but indicate a meaningful relationship between entities or articles.[1] The link structure has been exploited earlier to derive measures of centrality for all nodes (articles) in the network (see, e.g., Eom and Shepelyansky, 2013; Skiena and Ward, 2013).

The underlying intuition is that more important pages have more incoming links (i.e. other pages referring to this page) than less important pages. How does this speak to the concept of substantive representation? Actual political influence should manifest in a vast number of real-world outcomes, such as the record in law-making or holding high-ranking political offices. However, politics is a social business: Legislative influence requires regular collaboration with peers (Bratton and Rouse, 2011; Fowler, 2006; Kirkland and Gross, 2014),

[1] The Manual of Style for the platform states that "[i]nternal links bind the project together into an interconnected whole. (...) Whenever writing or editing an article, it is important to consider not only what to put in the article, but what links to include to help the reader find related information, as well as which other pages should carry links to the article" (Wikipedia Manual of Style, 2017).

party networks provide access to resources and power (Cohen et al., 2009; Norris, 1997), and politicians are immediately linked to each other through their positions, constituencies, and actions. The linkage structure provided by a network of Wikipedia articles on politicians certainly does not provide a comprehensive real-world representation of the formal and informal networks. However, we argue that the centrality of actors within this network is already revealing about their substantive role in the political arena.

In order to make sense of the linkage structure, we follow an established practice and implement the PageRank algorithm, which has originally been developed to rank webpages by Google search engine importance (Brin and Page, 1998). Technically, the PageRank of a page p_i that is part of a universe of $p_1, \ldots, p_N$ pages, is given by

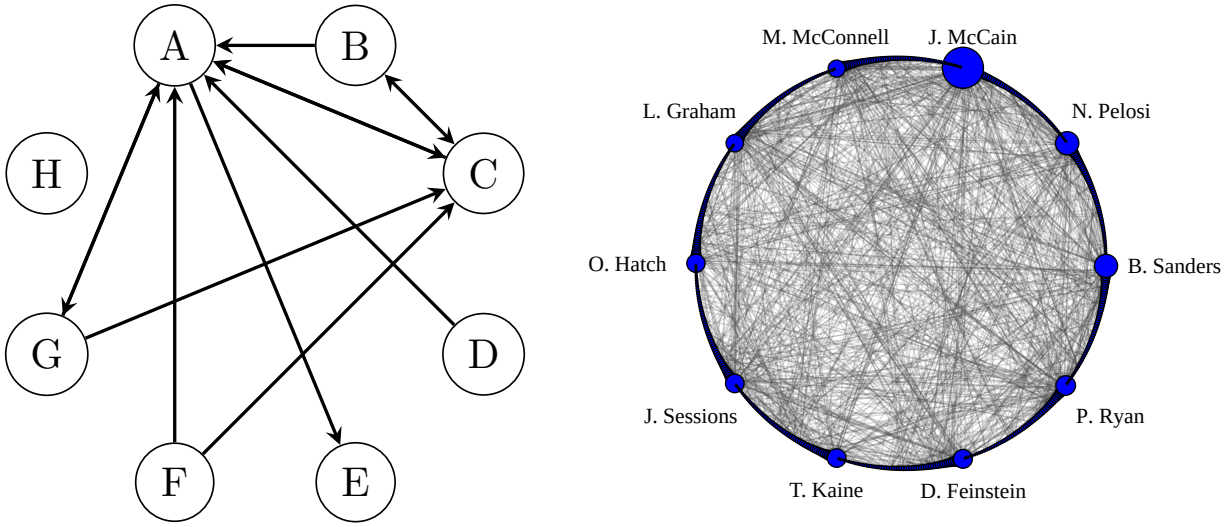
$$\text{PageRank}(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{\text{PageRank}(p_j)}{L(p_j)}, \tag{1}$$

where $M(p_i)$ represents the set of pages that link to p_i and L_{p_j} is the number of outgoing links on page p_j. In other words, a page's PageRank is higher the more pages refer to it, which consequently contributes more weight. The output represents the likelihood of a random click on a link in the network leading to a particular page. d is then the so-called damping factor—the probability that any user clicking through this network will stop at any step— and is commonly set to 0.85 (Brin and Page, 1998, 4).[2]

Figure E1 illustrates the logic of the PageRank algorithm. Consider the directed graph as provided in Panel E1a. A is a very prominent node in the network, as many other nodes point to it. E has only one incoming link, but the edge comes from A, resulting in a portion of the prominence of A reflecting on E. Node F is connected by two edges, both of which are outgoing. Therefore F will be described as fairly unimportant. The right-hand panel E1b applies the PageRank on the members of the 115th United States Congress. Larger circles represent higher PageRank values. We see that very few politicians dominate this ranking: Highest scoring is John McCain, who did not hold a formal leadership position at that point but can be considered one of the most important politicians in that legislature, running for President as the Republican nominee in 2008 and holding House and Senate

[2]Therefore, a PageRank of 0.1 implies a 10% probability that a click on a random link leads to a particular page with that PageRank. In real-world examples, these PageRank values are generally much smaller.

Figure E1: Illustration of the PageRank algorithm



(a) Schematic representation of directed graph (b) PageRank graph of 115th Congress

positions between 1982 and 2018. Other high-ranking politicians include House Minority Leader Nancy Pelosi and her Republican counterpart, Paul Ryan, but additionally people with high seniority, such as Dianne Feinstein (Senate member since 1992) or Orrin Hatch (Senate member 1977–2019).

Finally, note that focusing on incoming instead of outgoing links makes the measure robust towards self-serving edits, a phenomenon that has been described before (Göbel and Munzert, 2018): In order to boost a politician's score, an editor would have to smuggle a link to her article from a relatively more central person or institution, thus, edit a more central article. Irrelevant links (e.g., a link from a rank-and-file member's entry to the President's entry, or the other way round, reporting a one-time encounter during a campaign rally) should therefore not be a major issue.

Using the Wikipedia page IDs of all legislators for the language edition that was used in the database, we downloaded snapshots of their articles on October 23, 2020. These snapshots represented the raw HTML of each article, retaining the hyperlinks embedded in the article bodies. We then used these raw HTML files to construct a directed graph of links between legislator entries. Based on this graph the PageRank values are computed within the entire set of legislators per legislature.

Figure E2: Ratio of female legislators in legislatures and among legislators ranking among the top 25% on the PageRank centrality metric based on Wikipedia legislator article graphs.

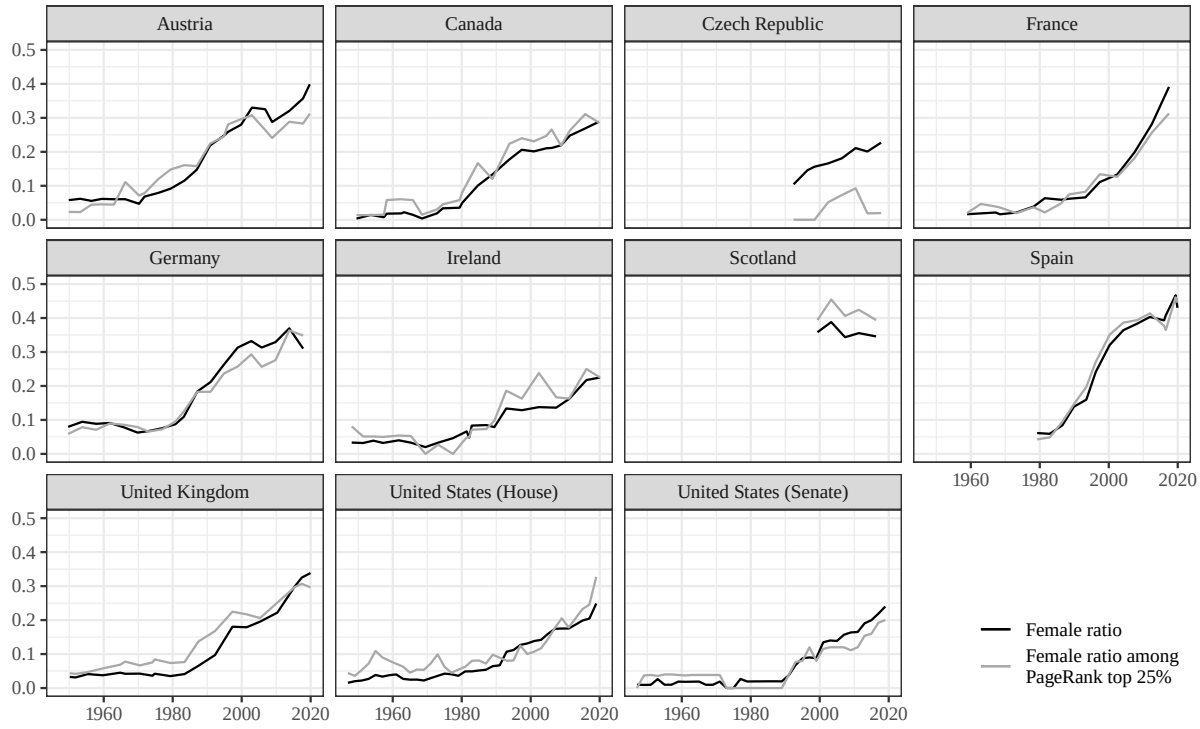


Figure E2 compares two metrics of female representatives in parliament by legislature. First, it shows the ratio of female legislators over the complete session history (given by the black lines). Second, it provides the ratio of female legislators among legislators ranking among the top 25% on the PageRank centrality metric (grey lines). The patterns are similar when other thresholds (such as 10%, 40%) are used. By focusing on the female ratio among more central actors in the system, the computed statistics remain easily comparable to the unconditional gender ratio.

Somewhat surprisingly, both trajectories follow each other closely in almost all legislatures, with the exception of the Czech Parliament, where the fraction of women among the top 25% legislators as measured by the PageRank metric trails behind the mere female ratio. This indicates that women entering parliament take, on average, similarly central roles as male representatives, at least according to our measure. Obviously, these are preliminary findings and more research is needed to evaluate the validity of the measure.

Appendix F Introduction to the R Package

Access to the database is managed via the legislatoR R package. The package can be installed from GitHub as follows:

```
# install from GitHub
install.packages("devtools")
library(devtools)
install_github("legislatoR")
```

After having installed the package, it must be loaded via:

```
# load library
library(legislatoR)
```

The package provides dataset-specific function calls named after the datasets and preceded by "get_". Each of these functions take only one single argument, the ISO 3166-1 alpha-3 code of the country for which a dataset shall be fetched from the server. Below are lists of datasets with corresponding functions as well as countries with corresponding codes.

Dataset	Function
Core	get_core()
IDs	get_ids()
Offices	get_office()
Political	get_political()
Portraits	get_portrait()
Professions	get_profession()
Social	get_social()
Wikipedia History	get_history()
Wikipedia Traffic	get_traffic()

Country	Function Argument
Austria	"aut"
Canada	"can"
Czech Republic	"cze"
France	"fra"
Germany	"deu"
Ireland	"irl"
Scotland	"sco"
Spain	"esp"
United Kingdom	"gbr"
United States Congress	"usa_house/_senate"

Datasets can be accessed for specific countries the following way:

```
# access Core dataset for Canada
can_core <- get_core("can")


# access Traffic dataset for Germany
deu_traffic <- get_traffic("deu")
```

```
# access Political dataset for United States Senate
usa_senate_political <- get_political("usa_senate")
```

These commands download a whole dataset for a country and assign it to the R environment under the user-specified name provided to the left of the assignment arrow. Datasets are stored in the R environment as data.frame objects.

Usually, we wish for somewhat more targeted access, i.e., sociodemographic data about politicians from a specific legislative session or social media handles for politicians of a specific party. The relational database was designed to achieve this via filtering joins. To allow for such joins, the dplyr package is required.

```
# install the dplyr package
install.packages("dplyr")

# load the dplyr package
library(dplyr)
```

Extracting sociodemographic data for politicians from a specific session is fairly easy. Since the *Core* and *Political* datasets are directly linked via the Wikipedia page ID, this requires only one semi-join. A semi-join returns all rows from a dataset in X where there are matching values of a key from a dataset in Y, keeping only columns from X.

```
# extract sociodemographic data about United Kingdom politicians
# from the 42nd legislative session
uk_core_sub <- semi_join(x = get_core("gbr"),
                         y = filter(get_political("gbr"),
                                    session == 42),
                         by = "pageid")
```

This chunk of code downloads the *Core* and *Political* datasets for the UK. It then filters the *Political* dataset to only include members from the 42nd legislative session. Lastly, it extracts *Core* data of members appearing in the 42nd session *Political* dataset and assigns it to the environment. This process takes less than a second. It assigns only a data frame of size 239.9 KB with 12 variables for 691 politicians into the environment, not the full *Core* and *Political* datasets called in the process and amounting to 6.9 MB.

Extracting social media handles for politicians who are members of a specifc party is slightly more complicated. Since the *Social* and *Political* datasets are not directly linked via a common key, a detour through the *Core* dataset is required. In essence, this simply means that we need two instead of one semi joins.

```
# extract social media handles for German politicians who are members
# of the SPD party
deu_spd <- semi_join(x = get_core("deu"),
                     y = filter(get_political("deu"),
                                party == "SPD"),
                         by = "pageid")
deu_social_spd <- semi_join(x = get_social("deu"),
                            y = deu_spd,
                            by = "wikidataid")
```

The first chunk works similar as before. It first downloads the *Core* and *Political* datasets for Germany, filters the latter by SPD party membership, and assigns only *Core* data of respective politicians to the environment. The second call now downloads the *Social* dataset for Germany but keeps only politicians included in the previously extracted *Core* data for SPD members. To achieve the same with even less code and having to assign only the desired result into the environment, we can use the pipe operator from the magrittr package.

```
# install the magrittr package
install.packages("magrittr")

# load the magrittr package
library(magrittr)

# extract social media handles for German politicians who are members
# of the SPD party
deu_social_spd <- semi_join(x = get_core("deu"),
                            y = filter(get_political("deu"),
                                       party == "SPD"),
                            by = "pageid") %>%
                  semi_join(x = get_social("deu"),
```

```
                          y = .,
                          by = "wikidataid")
```

The pipe operator %>% takes the output of a function call and passes it on to the next function call. The output of the previous call is referred to via a dot.

In addition to conditional subsets of datasets, we might also want to combine information from different datasets. This can be achieved via mutating joins, which are also implemented using the dplyr package. Lets say we wish to combine sociodemographic and political data.

```
# extract combined sociodemographic and political data for the Czech
# parliament
cze_soc_pol <- left_join(x = get_core("cze"),
                         y = get_political("cze"),
                         by = "pageid")
```

The left join that is used here appends the information stored in the *Political* dataset to politicians and information in the *Core* datset. This results in some kind of panel data, where politicians are repeated the number of sessions they are observed in. A format like this can be used immediatly to study individual party change between legislative sessions.

For those who wish to conduct their analyses using other software, such as EXCEL, SAS, STATA, or SPSS, the following code can be used to export the extracted data in the desired format.

```
# save data frame as .csv for use with Excel
write.csv(cze_soc_pol, "cze_soc_pol.csv")

# install haven package for export into other formats
install.packages("haven")

# load haven package
library(haven)

# save data frame as .sas for use with SAS
write_sas(deu_social_spd, "deu_social_spd.sas")
```

```
# save data frame as .dta for use with STATA
write_dta(can_core, "can_core.dta")

# save data frame as .sav for use with SPSS
write_sav(deu_traffic, "deu_traffic.sav")
```

We offer further specialized tutorials online at https://cran.r-project.org/web/packages/legislatoR/vignettes/legislatoR.html. There, we illustrate how to extract and combine data with external datasets, construct cross-country panels, loop over countries in the database, or combine filtering and mutating joins for even more targeted data access.

Appendix G Software Statement

All scripts were run under Windows 10 x86-64 using R version 3.4.2. The following R packages were used for assembling the database and for analyses reported in this paper.

Table G1: R packages.

coefplot (Lander, 2018)	pageviews (Keyes and Lewis, 2019)
cowplot (Wilke, 2019)	plyr (Wickham, 2011)
crayon (Csárdi and Gaslam, 2017)	purrr (Henry, Wickham and RStudio, 2019)
data.table (Dowle et al., 2017)	R.utils (Bengtsson, 2019)
dplyr (Wickham et al., 2020)	readxl (Wickham et al., 2019)
eeptools (Becker and Knowles, 2019)	reshape2 (Wickham, 2017*a*)
extrafont (Chang, 2014)	rvest (Wickham, 2016)
finalfit (Harrison, Drake and Ots, 2019)	stargazer (Hlavac, 2018)
ggplot2 (Wickham, 2009)	stringr (Wickham, 2017*b*)
gtools (Warnes, Bolker and Lumley, 2018)	tibble (Müller and Wickham, 2019)
haven (Wickham and Miller, 2019)	tidyr (Wickham, Henry and RStudio, 2019)
httr (Wickham, 2019)	tidyselect (Henry and Wickham, 2018)
jsonlite Ooms, Lang and Hilaiel (2018)	toOrdinal (Betebenner, Martin and Erickson, 2019)
lubridate (Grolemund and Wickham, 2011)	vroom (Hester et al., 2020)
magrittr (Bache and Wickham, 2014)	WikidataR (Keyes et al., 2017s)
mpoly (Kahle, 2019)	Wikipediatrend (Meissner, 2017s)
pacman (Rinker et al., 2017)	xtable (Dahl et al., 2019)
padr (Thoen, 2019)	zoo (Zeileis and Grothendieck, 2005)

References

Bache, Stefan M. and Hadley Wickham. 2014. *magrittr: A Forward-Pipe Operator for R.* R package version 1.5.
URL: *https://CRAN.R-project.org/package=magrittr*

Becker, Jason P. and Jared E. Knowles. 2019. *eeptools: Convenience functions for education data.* R package version 1.2.2.
URL: *https://CRAN.R-project.org/package=stringr*

Bengtsson, Henrik. 2019. *R.utils: Various programming utilities.* R package version 2.9.0.
URL: *https://CRAN.R-project.org/package=R.utils*

Betebenner, Damian W., Andrew Martin and Jeff Erickson. 2019. *toOrdinal: Cardinal to ordinal number and date conversion.* R package version 1.1.
URL: *https://CRAN.R-project.org/package=toOrdinal*

Bratton, Kathleen A. and Stella M. Rouse. 2011. "Networks in the Legislative Arena: How Group Dynamics Affect Cosponsorship." *Legislative Studies Quarterly* 36(3):423–460.

Brin, Sergey and Larry Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the 7th World-Wide Web Conference*. Brisbane, Australia: pp. 1–20.

Chang, Winston. 2014. *extrafont: Tools for using fonts.* R package version 0.17.
URL: *https://CRAN.R-project.org/package=extrafont*

Cohen, Marty, David Karol, Hans Noel and John Zaller. 2009. *The party decides: Presidential nominations before and after reform.* University of Chicago Press.

Csárdi, Gábor and Brodie Gaslam. 2017. *crayon. Colored terminal output.* R package version 1.3.4.
URL: *https://CRAN.R-project.org/package=crayon*

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, Jonathan Swinton, Ajay Shah, Arne Henningsen, Benno Puetz, Bernhard Pfaff, Claudio Agostinelli, Claudius

Loehnert, David Mitchell, David Whiting, Fernando da Rosa, Guido Gay, Guido Schulz, Ian Fellows, Jeff Laake, John Walker, Jun Yan, Liviu Andronic, Markus Loecher, Martin Gubri, Matthieu Stigler, Robert Castelo, Seth Falcon, Stefan Edwards, Sven Garbade and Uwe Ligges. 2019. *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.
URL: *https://CRAN.R-project.org/package=xtable*

Dowle, Matt, Arun Srinivasan, Jan Gorecki, Tom Short, Steve Lianoglou and Eduard Antonyan. 2017. *data.table: Extension of 'data.frame'*. R package version 1.10.4.
URL: *https://CRAN.R-project.org/package=data.table*

Eggers, Andrew and Arthur Spirling. 2014. "Electoral Security as a Determinant of Legislator Activity, 1832–1918: New Data and Methods for Analyzing British Political Development." *Legislative Studies Quartely* 39(4):593–620.

Eom, Young-Ho and Dima L Shepelyansky. 2013. "Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles." *PloS one* 8(10):e74554.

Fowler, James H. 2006. "Legislative cosponsorship networks in the US House and Senate." *Social Networks* 28(4):454 – 465.

Giles, Jim. 2005. "Internet encyclopaedias go head to head." *Nature* 438:900–901.

Göbel, Sascha and Simon Munzert. 2018. "Political Advertising on the Wikipedia Marketplace of Information." *Social Science Computer Review* 36(2):157–175.

Greenstein, Shane, Yuan Gu and Feng Zhu. 2017. "Ideological segregation among online collaborators. Evidence from Wikipedians." *NBER Working Papers* .

Grolemund, Garrett and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40(3).
URL: *https://www.jstatsoft.org/article/view/v040i03*

Harrison, Ewen, Tom Drake and Riinu Ots. 2019. *finalfit: Quickly create elegant regression results tables and plots when modelling*. R package version 0.9.4.
URL: *https://CRAN.R-project.org/package=finalfit*

Henry, Lionel and Hadley Wickham. 2018. *tidyselect: Select from a set of strings*. R package version 0.2.5.
URL: *https://CRAN.R-project.org/package=tidyselect*

Henry, Lionel, Hadley Wickham and RStudio. 2019. *purrr. Functional programming tools*. R package version 0.3.2.
URL: *https://CRAN.R-project.org/package=purrr*

Hester, Jim, Hadley Wickham, Jukka Jylänki, Mikkel Jørgensen and RStudio. 2020. *vroom: Read and Write Rectangular Text Data Quickly*. R package version 1.2.0.
URL: *https://CRAN.R-project.org/package=vroom*

Hlavac, Marek. 2018. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.2.
URL: *https://CRAN.R-project.org/package=stargazer*

Kahle, David. 2019. *mpoly: Symbolic computation and more with multivariate polynomials*. R package version 1.1.0.
URL: *https://CRAN.R-project.org/package=mpoly*

Kalla, Joshua L. and Peter M. Aronow. 2015. "Editorial bias in crowd-sourced political information." *Plos One* 10.

Keyes, Oliver and Jeremiah Lewis. 2019. *pageviews: An API client for Wikimedia traffic data*. R package version 0.5.0.
URL: *https://CRAN.R-project.org/package=padr*

Keyes, Oliver, Serena Signorelli, Christian Graul and Mikhail Popov. 2017s. *WikidataR: API client library for 'Wikidata'*. R package version 1.4.0.
URL: *https://CRAN.R-project.org/package=WikidataR*

Kirkland, Justin H and Justin H Gross. 2014. "Measurement and theory in legislative networks: The evolving topology of Congressional collaboration." *Social Networks* 36:97–109.

Kroeber, Corinna. 2018. "How to measure the substantive representation of traditionally excluded groups in comparative research: a literature review and new data." *Representation* 54(3):241–259.
URL: *https://doi.org/10.1080/00344893.2018.1504112*

Lander, Jared P. 2018. *coefplot: Plots Coefficients from Fitted Models*. R package version 1.2.6.
URL: *https://CRAN.R-project.org/package=coefplot*

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2019. *Voteview. Congressional roll-call votes database*.
URL: *https://voteview.com/data*

Meissner, Peter. 2017s. *wikipediatrend: Public subject attention via Wikipedia page view statistics*. R package version 1.1.14.
URL: *https://CRAN.R-project.org/package=wikipediatrend*

Müller, Kirill and Hadley Wickham. 2019. *tibble: Simple Data Frames*. R package version 2.1.3.
URL: *https://CRAN.R-project.org/package=tibble*

Norris, Pippa. 1997. *Passages to Power: Legislative Recruitment in Advanced Democracies*. Cambridge: Cambridge University Press.

Ooms, Jeroen, Duncan T. Lang and Lloyd Hilaiel. 2018. *jsonlite: A robust, high performance JSON parser and generator for R*. R package version 1.6.
URL: *https://CRAN.R-project.org/package=jsonlite*

Rinker, Tyler, Dason Kurkiewicz, Keith Hughitt, Albert Wang and Jim Hester. 2017. *pacman: Package Management Tool*. R package version 0.4.6.
URL: *https://CRAN.R-project.org/package=pacman*

Sieberer, Ulrich, Thomas Saalfeld, Tamaki Ohmura, Henning Bergmann and Stefanie Bailer. 2020. "Roll-call votes in the German Bundestag. A new dataset, 1949–2013." *British Journal of Political Science* 50(3):1137–1145.

Skiena, Steven S. and Charles B. Ward. 2013. *Who's Bigger?: Where Historical Figures Really Rank*. Cambridge University Press.

Thoen, Edwin. 2019. *padr: Quickly get datetime data ready for analysis*. R package version 0.5.0.
URL: *https://CRAN.R-project.org/package=padr*

Wängnerud, Lena. 2009. "Women in parliaments: Descriptive and substantive representation." *Annual Review of Political Science* 12:51–69.

Warnes, Gregory R., Ben Bolker and Thomas Lumley. 2018. *gtools: Various R programming tools*. R package version 3.8.1.
URL: *https://CRAN.R-project.org/package=gtools*

Wickham, Hadley. 2009. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
URL: *http://ggplot2.org/*

Wickham, Hadley. 2011. "The split-apply-combine strategy for data analysis." *Journal of Statistical Software* 40(1):1–29.
URL: *https://www.jstatsoft.org/article/view/v040i01*

Wickham, Hadley. 2016. *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.3.2.
URL: *https://CRAN.R-project.org/package=rvest*

Wickham, Hadley. 2017*a*. *reshape2: Flexibly reshape data. A reboot of the reshape package*. R package version 1.4.3.
URL: *https://CRAN.R-project.org/package=reshape2*

Wickham, Hadley. 2017*b*. *stringr: Simple, consistent wrappers for common string operations*. R package version 1.2.0.
URL: *https://CRAN.R-project.org/package=stringr*

Wickham, Hadley. 2019. *httr: Tools for working with URLs and HTTP*. R package version 1.4.1.
URL: *https://CRAN.R-project.org/package=httr*

Wickham, Hadley and Evan Miller. 2019. *haven: Import and export 'SPSS', 'Stata' and 'SAS' files*. R package version 2.1.1.
URL: *https://CRAN.R-project.org/package=haven*

Wickham, Hadley, Jennifer Bryan, Marcin Kalicinski, Komarov Valery, Christophe Leitienne, Bob Colbert, David Hoerl and Evan Miller. 2019. *readxl: Read Excel files*. R package version 1.3.1.
URL: *https://CRAN.R-project.org/package=readxl*

Wickham, Hadley, Lionel Henry and RStudio. 2019. *tidyr. Tidy messy data*. R package version 1.0.0.
URL: *https://CRAN.R-project.org/package=tidyr*

Wickham, Hadley, Romain François, Lionel Henry and Kirill Müller. 2020. *dplyr: A Grammar of Data Manipulation*. R package version 0.8.3.
URL: *https://CRAN.R-project.org/package=dplyr*

Wikipedia Manual of Style. 2017.
URL: *https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style/Linking&oldid=791759470*

Wilke, Claus O. 2019. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.0.0.
URL: *https://CRAN.R-project.org/package=cowplot*

Zeileis, Achim and Gabor Grothendieck. 2005. "zoo: S3 infrastructure for regular and irregular time series." *Journal of Statistical Software* 14(6).