The Political Leaders' Affiliation Database (PLAD)[1]

Codebook – Version Date October 2021

1. Introduction

The personal characteristics of political leaders shape economic and political decisions and settings (e.g., Franck and Rainer, 2012; Francois et al. 2015; De Luca et al., 2018). Yet, the lack of a publicly available, comprehensive dataset on political leaders' characteristics has prevented researchers from addressing important empirical questions. The Political Leaders' Affiliation Database (PLAD) contains information on the birthplaces and ethnicities of the effective leaders of 177 countries around the world in the 1989-2020 period. The dataset is at the political leader level and reports information on 1,109 effective leaders, in office for 1,266 distinct periods of time. It follows the definitions in the Archigos database on Political Leaders (Goemans et al. 2009). PLAD codes ethnicities and georeferences birthplaces of the effective leader, with highest precision being the village or city level.[2] In Section 2 of this codebook, we describe the dataset and its variables. Section 3 contains a detailed description of the precision codes and the data collection procedure.

Please cite these data as "Axel Dreher, Fuchs, Andreas, Kammerlander, Andreas, Kaplan, Lennart, Robert, Charlotte and Unfried, Kerstin (2021). Light of Their Lives: Country Leaders' Spouses and Regional Favoritism. Mimeo."

2. Variables and data set description

The following table describes the variables included in the dataset:

Variable	Description
idacr	Country code of the country that the leader governs based on the Correlates of War Project (source until 2015: Archigos)
leader	Name of the leader (source until 2015: Archigos)

[1] We are grateful for support from our excellent research assistants: Regina Dworschak, Erin Flannagan, Alexander Hudlet, Vivien Hülsen, Timo Kretschmer, Matthew Rudh, Dustin Stanizek, Levani Veshapidze, and Levin Wiebelt from the University of Göttingen; Aurora Brahimllari and Ramón Rey from the University of Freiburg; Giannis Hafizi and Franziska Hartung from the Helmut-Schmidt University Hamburg; Tim Deisemann, Yannik Hausmann, Tobias Hellmundt, and Felix Höfer from Heidelberg University. We thank Maria Perotta Berlin, Christian Bommer, Raj M. Desai, Richard Jong-A-Pin, Anders Olofsgård, Brad Parks, Marcello Pérez-Alvarez and Shu Yu for sharing their data.

[2] We link leaders' ethnicity to the definitions of the GeoEPR database (Vogt et al., 2015). A future version of this dataset will include biographical information on their spouses.

plad_id	Unique leader identification code
archigos_id	Unique identification code corresponding to Archigos variable “leadid” (available until 2015: source: Archigos)
startdate	Date of entry to office (source until 2015: Archigos)
enddate	Date of exit from office (source until 2015: Archigos)
startyear	Year of entry to office (source until 2015: Archigos)
endyear	Year of exit from office (source until 2015: Archigos). We code exit dates for leaders still in office on December 31, 2020 with this date.
adm0	Country in which the birthplace of the leader is located. Note: May differ from the *country* variable if the leader was born abroad. Countries are identified based on the GADM dataset (version 3.6, Hijmans et al., 2020).
adm1	First administrative division in which the birthplace of the leader is located. The first administrative division is identified based on the GADM dataset (version 3.6).
adm2	Second administrative division in which the birthplace of the leader is located. The second administrative division is identified based on the GADM dataset (version 3.6).
country	Name of the country that the leader governs
continent	Continent in which the country governed by the leader is located
latitude	Latitude of the leader birthplace
longitude	Longitude of the leader birthplace
geoname	Search term used in Geonames to geolocate birthplace[3]
geo_precision	Precision of birthplace information on a scale of 1 - 6 (see Section 3.2)
foreign_leader	Indicator variable that is one if the leader was not born in the country of government; zero otherwise

[3] When the place is not referenced on Geonames.org, we retrieved coordinates using Google Maps. In this case, geoname corresponds to the search term used to locate the birthplace on Google Maps. These rare cases are indicated in the comments. For some rare cases (Bosnia and Herzegovina, Eritrea) we adapted some of the administrative regions provided in the GADM manually, as their data did not correspond to the actual administrative regions.

ethnicity1	First ethnic group a leader belongs to as stated in the source
ethnicity2	Second ethnic group a leader belongs to as stated in the source
ethnicity_geoepr1	First ethnic group corresponding to the names of groups in the GeoEPR dataset
ethnicity_geoepr2	Second ethnic group corresponding to the names of groups in the GeoEPR dataset
ethnicitysource1	First source used to retrieve the leader's ethnicity
ethnicitysource2	Second source used to retrieve the leader's ethnicity
ethnicity_precision	Precision of the information on the leader's ethnicity on a scale of 1 - 4 (see Section 3.3)
entry	Type of entry (source: Archigos)
exit	Type of exit (source: Archigos)
gender	Leader's gender (source until 2015: Archigos)
yrborn	Year of leader birth (source until 2015: Archigos)
birthdate	Birthdate of leader (source until 2015: Archigos)
uid	Object ID from GADM dataset (version 3.6)[4]
id_0	Numerical ID for country from GADM dataset (version 3.6) (ADM0 layer)
id_1	Numerical ID for the first administrative division from GADM dataset (version 3.6) (ADM1 layer)
id_2	Numerical ID for the second administrative division from GADM dataset (version 3.6) (ADM2 layer)
gid_0	String ID for country from GADM dataset (version 3.6). ISO 3166-1 alpha-3 country code when available (ADM0 layer)
gid_1	String ID for the first administrative division from GADM dataset (version 3.6) (ADM1 layer)

[4] Object IDs are from the GADM polygons. We only provide them for leaders whose birth region we could identify at least at the second administrative level (precision code 3), as we merged our data with the GADM dataset at the second administrative level. We advise users who would like to merge the dataset on the first administrative layer via object IDs to carry out an individual geo-merge with the GADM data via the latitude and longitude data.

gid_2	String ID for the second administrative division from GADM dataset (version 3.6) (ADM2 layer)
edu_name	Name of the education degree/training obtained by the leader and the field of study
edu_r	Leader's level of education summarized in eight categories (source until 2012: Yu and Jong-A-Pin, 2020)
birthplace_comment	Notes on controversial cases and additional information on the leader birthplace
ethnicity_comment	Notes on controversial cases and additional information on the leader's ethnicity

3. Details on the data collection process

3.1 General information

The Political Leaders' Affiliation Database (PLAD) is a manually compiled dataset. We collected information by structured internet searches. As sources, we mainly rely on online libraries and databases such as CIA World Factbook, Munzinger, Encyclopaedia Britannica, and Ethnicity of Celebs, as well as on reports from popular news services such as BBC News, The Guardian, and The Washington Post.

The dataset provides information on the birthplace of a leader if at least two reliable sources report the same location of birth. In most cases, the determination of a leader's birthplace is clear and uncontroversial. As information on ethnicity is not reported as frequently as birthplaces, we provide a leader's ethnicity if at least one source is available and state the sources used in the dataset.

In order to ensure the accuracy of the data, we apply the following three steps:

(1) A second coder reviewed all information.
(2) We further checked for consistency by comparing our data with those of Berlin et al. (2019), Bommer et al. (2019) and Dreher et al. (2019).
(3) The lead researchers randomly selected and screened 25% of all leaders' entries again as quality assurance.

3.2 Information on leaders' birthplaces

This section describes the georeferencing of national leaders' birthplaces. The location of a leader's origin is determined by biographical information on the place where the leader was born. If sources stated contradictory places, we noted this in column *birthplace_comment*. For the identification of the birthplaces, we used the online encyclopedias Encyclopedia Britannica and Wikipedia[5] and complemented this information with resources focusing on bibliographic accounts, for example, CIDOB, The Famous People, and Munzinger. After identifying the birthplaces of all national leaders, we obtained the respective coordinates using Geonames (https://www.geonames.org/). We provide the search term used to retrieve the coordinates in the variable *geonames*.[6] The variables *longitude* and *latitude* contain the coordinates of the leaders' birthplace. As the quality of information, i.e., the accuracy of coordinates, differs substantially between different leaders, the variable *geo_precision* contains information on how accurate the coordinates are. The categories of *geo_precision* are based on Strandow et al. (2011):

[5] While we prefer other non-editable sources, Wikipedia offers comprehensive and well-structured information on leader characteristics.
[6] Oftentimes, the capital city has the same name as the administrative divisions in which it is located, on the first and second level. If we cannot identify whether leaders were born in the city or in the broader administrative division, we adopt a conservative approach and assume that they were born in the first or second order administrative division that contains the capital city and shares the same name.

1 = The coordinates correspond to an exact location, such as a populated place (villages, cities) or a hill. The code is also used for locations that join a location which is a line (such as a road or railroad). We coded the points that connect lines, but not the lines between points (all points that are mentioned in the source are coded).[7]
2 = The location is mentioned in the source as being "near," in the "area" of, or up to 25 km away from an exact location. The coordinates refer to that adjacent, exact, location.
3 = The location is, or is analogous to, a second order administrative division (ADM2), such as a district, municipality or commune.
4 = The location is, or is analogous to, a first order administrative division (ADM1), such as a province, state or governorate.
5 = The location can only be related to estimated coordinates, such as when a location lies between populated places; along rivers, roads and borders; more than 25 km away from a specific location; or when sources refer to parts of a country greater than ADM1 such as a National Park which spans across several provinces (e.g., Foret Classee de Gongon in Benin).
6 = The location can only be related to an independent political entity, meaning the pair of coordinates that represent a country. This includes leaders that were born in larger areas that cannot be geo-referenced at a more precise level.
7 = Unclear. No coordinates are entered to reflect that sub-country information is unavailable.

Furthermore, the names of the second and first administrative region can be found in *adm1 and adm2*, respectively. The variable *foreign_leader* shows if a national leader was born in a foreign country. It takes a value of one if the leader was not born in the country they govern and is zero otherwise. The variable *birthplace_comment* gives further information on the birthplace and explains decision making in controversial cases.

3.3 Information on leaders' ethnicity

The concept of ethnicities relates to a belief of a common culture and ancestry, being thus inherently subjective (Weber 1976). For our purpose, we draw on a commonly used classification by Vogt et al. (2015), who suggest a classification of such common culture and ancestry by features including (i) language, (ii) beliefs/religion, or (iii) phenotypical characteristics. The salience of those features may differ by world region, where ethnicities in Latin America separate by skin color, while African ethnicities are more separated along the lines of linguism. This also relates to the level of ethnic distinction, where some countries only know two ethnicities and in other countries ethnicity is a very complex concept with several sub-groups. We rely on the ethnic concepts as mentioned in the sources and the main groups given in Vogt et al. (2015). While the level of differentiation is not fixed across countries, it represents the salience within the country and is thus at the relevant level for questions in political economy.[8] Furthermore, similar ethnic

[7] Sometimes we could identify a precise birthplace, but the GADM data did not provide information on ADM2 regions. In this case, we add a note in the variable birthplace_comment.
[8] Due to differing relevance of ethnic affiliation across countries, sometimes it was harder to find data and we had to base information regarding ethnicity on related concepts (nationality, skin color or family ties).

groups may be named rather heterogeneously across countries, where, for instance, “Afro-American,” “Afro-Haitian” or “black” refers to people with African descent. In those cases we stick to the ethnicity name used by the source or by Vogt et al. (2015), to minimize interpretation biases. Our sources include previous databases on leader characteristics (e.g., Fearon et al., 2007; Parks, 2014), and encyclopedias such as CIDOB, The Famous People, Munzinger, Encyclopaedia Britannica, and Ethnicity of Celebs. If sources named contradictory ethnicities, we note this in column *ethnicity_comment*.

Note that the availability of information on leaders’ ethnicity depends on the country context. Information is less likely to be available in countries where ethnicity is less salient, e.g., in less ethnically fractionalized societies. What is more, data availability depends on whether ethnicity is determined by language, phenotypical factors, or religion. Therefore, the quality and quantity of sources differ strongly across contexts. For countries that are less represented on major encyclopedias, we draw on country-specific resources like books or webpages of the parliament but also country-specific webpages like Afghan Bios for Afghanistan, Banglapedia for Bangladesh, or BiographyBD for India and Pakistan. What is more, some sources just provide indirect information on the ethnicity of leaders based on the individual’s ancestry. Here, we also offer users the option to filter with the following precision codes (variable *ethnicity_precision*):

1 = Two sources state the ethnicity directly.
2 = Only one direct source or one of the sources is Wikipedia.
3 = No direct mentioning of ethnicity. Attribution via characteristics mentioned in the text or phenotypical factors in picture.

3.4 Information on leaders’ education

Data on leaders’ education level are taken from Yu and Jong-A-Pin (2020) until 2012 and updated to include leaders up to the year 2020. Leaders’ level of education is summarized in the categorical variable *edu_r* following the eight-way classification used in Ludwig (2002):

1 = Illiterate (no formal education)
2 = Literate (no formal education)
3 = Elementary/primary school education or tutors
4 = High/finishing/secondary/trade school
5 = Special training (beyond high school, such as mechanical, nursing, art, music, or military training)
6 = College-educated
7 = Qualifications from a graduate or professional school (e.g., master’s degree)
8 = Doctorates (e.g., Ph.D.)

Military training programs that do not lead to a bachelor degree are considered to be category 5. When it is known that a leader attended college, but there is uncertainty on whether they

We coded the *ethnicity_precision* in those cases as 3 and added in column *ethnicity_comment* “Assessment of ethnicity is based on nationality/skin color.” or provided a reference to the family’s linkages (e.g., “parents were farmers with long ancestry in the region”).

graduated or the level at which they graduated, we code their education level as category 6. However, if the leader is known to be a lawyer or a medical doctor, we rank the education level as category 7, as these professions require at least a master's degree in most countries.
We supplement the data from Yu and Jong-A-Pin (2020) with the variable *edu_name,* which reports the name of the highest degree obtained by the leader, complemented with their field of study, when such information is available.

4. Contact

We are continuously working on improving the quality of our data. Hence, we are happy to receive notes on inconsistencies and mistakes.

Contact email address: plad@listserv.dfn.de

Bibliography

Bommer, C., Dreher, A., & Perez-Alvarez, M. (2019). Home Bias in Humanitarian Aid: The Role of Regional Favoritism in the Allocation of International Disaster Relief. CEPR Discussion Paper No. DP13957.

Berlin, M.P., Desai, R.M. & A. Olofsgard (2019). Trading Favors? UN Security Council Membership and the Sub-National Allocation of Aid. Mimeo. Available via: https://www.dropbox.com/s/gkdiz8zy2nln6rx/BerlinDesaiOlofsgard_Sept_2019.pdf?dl=0 .

De Luca, G., Hodler, R., Raschky, P. A., & Valsecchi, M. (2018). Ethnic Favoritism: An Axiom of Politics? *Journal of Development Economics*, 132, 115–129.

Dreher, A., Fuchs, A., Hodler, R., Parks, B. C., Raschky, P. A., & Tierney, M. J. (2019). African Leaders and the Geography of China's Foreign Assistance. *Journal of Development Economics*, 140, 44–71.

Fearon, J., Kasara, K., & Laitin, D. (2007). Ethnic Minority Rule and Civil War Onset. *American Political Science Review*, 101(1), 187–193.

Franck, R., & Rainer, I. (2012). Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa. *American Political Science Review*, *106*(2), 294–325.

Francois, P., Rainer, I., & Trebbi, F. (2015). How Is Power Shared in Africa. *Econometrica*, *83*(2), 465-503.

Goemans, H. E., Gleditsch, K. S., & Chiozza, G. (2009). Introducing Archigos: A Dataset of Political Leaders. *Journal of Peace Research*, *46*(2), 269–283.

Hijmans, R., Garcia, N., Kapoor, J., Rala, A., Maunahan, A., & Wieczorek, J. (2020). Global Administrative Areas, Version 3.6.

Parks, B. C. (2014). *Brokering Development Policy Change: The Parallel Pursuit of Millennium Challenge Account Resources and Reform*. Ph.D. Dissertation. London, UK: London School of Economics and Political Science.

Strandow, D., Findley, M., Nielson, D., & Powell, J. (2011). *The UCDP and AidData Codebook on Georeferencing Aid: Version 1.1*. Uppsala, Sweden: Uppsala University Department of Peace and Conflict Research.

Yu, S., & Jong-A-Pin, R. (2020). Rich or alive? Political (in) stability, political leader selection and economic growth. *Journal of Comparative Economics*, 48(3), 561-577.

Vogt, M., Bormann, N. C., Rüegger, S., Cederman, L. E., Hunziker, P., & Girardin, L. (2015). Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. *Journal of Conflict Resolution*, *59*(7), 1327–1342.

Weber, M. (1976). Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie. 5th Ed. Tübingen, Germany: J.C.B. Mohr.