

Visualizing subnational democracy in Colombia using Varieties of Democracy (V-Dem)

Merging V-Dem responses with municipal-level data

Patrick McQuestion*, Michael Coppedge†, Matthew Sisk‡

First version: February 4, 2024. This version: 29 March 2024.

Introduction

The objective of this methodological paper is to introduce geo-spatial mapping of democracy at the sub-national level using Varieties of Democracy (V-Dem) data. This effort in democracy studies is data-driven and descriptive (Coppedge (2012), 33-36). We describe the measurement process applied to one country (Colombia) in order to stimulate discussion on the methodological potential for V-Dem data to be used to measure democracy within countries.

A main contribution of this project is a systematic and applied conceptualization of V-Dem data that can provide a granular-level measure of subnational democracy.

An important limitation to mapping subnational V-Dem data is that expert interpretations of background concepts are themselves unobservable. The dichotomous format of the questions included in the survey are simplifications of much more contested concepts, such as development or illicit activity. Nonetheless, mapping provides a unique method for creating composite images of these concepts and their salience for country experts.

This paper introduces a process for visualizing subnational V-Dem data starting with the systematic conceptualization of the background concepts contained in the survey, assigning measures to those concepts on

*Department of Political Science & Kroc Institute, University of Notre Dame. Email: pmcquest@nd.edu

†Department of Political Science, University of Notre Dame. Email: mcoppedg@nd.edu

‡Department of Applied and Computational Mathematics and Statistics, University of Notre Dame. Email: msisk1@nd.edu

a geographic scale, and then identifying subnational scores. We demonstrate this process for a portion of the 21 response items in the survey.

The conclusion will discuss how these measurements could be combined or aggregated to provide estimations of the latent concepts of electoral fairness or civil liberty strength at the subnational level. We also discuss future exercises in validation (e.g., expert surveys, or comparison with other measures of subnational democracy) and replication (e.g., applying the same process to other countries). Finally, we suggest further exploring the causal relationships of first or second order complexity (e.g., the correlation of subnational democracy and development).

Relevance

The visualization of V-Dem scores has many potential benefits for comparative political analysis. First, this process provides a systematic method for integrating quantitative survey data into subnational research (Giraudy, Moncada, and Snyder (2019)) while avoiding typical spatial dependence measurement issues (Harbers and Ingram (2017)). Rather than spatial modeling, this inductive approach favors descriptive over explanatory inference, leveraging the systematic data generating process in V-Dem. V-Dem's data generation process captures the level of agreement among expert coders on the relevance of factors affecting subnational democracy. These levels of agreement produce proportional scores for the relative importance of a given variable, however it is conceptualized. When these proportions are interacted with the measurement of a systematized concept in geographic space, such as GDP per capita measures for economic development in a given municipality or region, this produces unique observational data. When mapped, this data produces a geographic visualization of a concept's salience for subnational democracy, according to country experts.

This inductive approach has benefits and limitations.

- On the one hand, the interaction of salience with geographic data is descriptive rather than explanatory.
- On the other hand, the estimations are heavily influenced by coder interpretations of complex concepts.

Nonetheless, this procedure represents a meaningful and transparent approach to measuring subnational democracy. The transparency behind this visualization procedure provides for validation and replication exercises. Once validated, these measures of subnational democracy can be used to test traditional hypotheses, such as whether development outcomes in subnational units are correlated with free and fair elections, or with the protection of civil liberties. We can use this data to examine the degree to which democracy is

weaker or stronger in areas with illicit activity and economic productivity, and how tightly these variables are linked together.

Background: V-Dem data

In the V-Dem dataset, both Elections and Civil Liberties surveys contain four questions related to subnational democracy¹:

1. *How would you describe the areas of the country in which elections are significantly less free and fair?*
2. *How would you describe the areas of the country in which elections are significantly more free and fair?*
3. *How would you describe the areas of the country where government officials' respect for civil liberties is significantly stronger?*
4. *How would you describe the areas of the country where government officials' respect for civil liberties is significantly weaker?*

On average, a total of five country experts respond to these questions by making dichotomous selections (0 = no, 1 = yes) across 21 individual variables. These refer to levels of rurality, economic development, urbanization, regional locations, civil unrest, illicit activity, sparse populations, difficult access, indigenous areas, support for national ruling parties, and colonial rule. Experts may also provide text comments that clarify their selections.

The selections are then aggregated by coders to calculate the proportion of expert responses that selected “1” for a given variable. The results are country-year proportional scores for each variable between 0-1. These proportional scores reflect the degree of agreement or disagreement around the relevance of some background concepts conditioning subnational elections or civil liberties. For example, a proportional score of 0.8 suggests that 8 out of 10 experts agree that more economically developed areas have stronger civil liberties.

Uncertainty

V-Dem’s subnational survey items encompass a wide range of background concepts, from geographic features (e.g., regions) to social phenomena (e.g., illicit activities), that may factor into or characterize subnational

¹In the dataset, these questions are labeled: v2elsnlfc, v2elsnmrhc, v2clrgstch, and v2clrgwkch, respectively. Note that while these questions are complementary, they are framed to elicit an open range of responses.

democracy. The complexity of these concepts produces a high degree of uncertainty around experts' interpretations and how they apply to the country in question. Moreover, some dichotomized concepts in the survey, such as "high/low economic development" or "strong/weak support for ruling parties", are best measured by continuous variables, such as GDP per capita or margin of victory. Respondents, however, must apply arbitrary thresholds to dichotomize these concepts, producing simplified generalizations.

Fundamentally, mapping subnational democracy faces many issues of measurement validity, what Adcock and Collier (2001) characterize as the movement between background concepts and observations.

- In mapping these concepts, the procedure outlined in this paper uses simple definitions and systematizes them by selecting one or two principal indicators for each response.
- Rather than discount alternative systematizations, our objective is to visualize the layering of these indicators that characterize subnational democracy in a given context, and therefore prioritize reliable over valid measurements.

Concepts, measures, and scores

This mapping procedure involves multiple stages, from the systematization of concepts with explicit definitions, to the assignment of measures based on coder responses, and finally the identification of subnational "scores" based on these measures.

Systematizing concepts

First, subnational V-Dem variables must be defined at "zero order complexity" (Coppedge (1999)) for country contexts, balancing cross-national and context-based interpretations.

- On the one hand, the cross-national nature of V-Dem surveys require the use of general rather than specific background concepts for measurement. For example, to assess the quality of subnational elections, the expert survey respondent is asked to consider the impact of "economic development" writ large across territories within the country. The question does not specify a threshold or type of activity. To ameliorate respondent bias, the final country-year score for this indicator is the proportion of respondents that made the same assessment. Without this step in generalization, V-Dem data would be unable to produce cross-national comparative data.

- On the other hand, country context is important for identifying indicators for social phenomena, particularly civil unrest and illicit activity. These are clearly **contested concepts**, and may even be highly **correlated**.

After V-Dem responses have been conceptualized for the country in question, the operationalization of these indicators requires the use of external (within-country) data. This may come from a wide array of sources and formats. In Colombia, we consider the indicators as shown in Table 1. All sources in parentheses are official government sources, although alternative indicators and sources certainly exist.

Table 1: Operationalization of V-Dem responses in Colombia

ID Response	Indicator
0 Rural	Rural Index (DANE)
1 Urban	
2 Less econ devt	Mun. Value Added / Dept. GDP (DNP)
3 More econ devt	
4 Inside capital	Geographic distance from center (DANE)
5 Outside capital	
6 North	
7 South	Six macro-regions (DANE)
8 West	
9 East	
10 Civil unrest	Illegally armed groups (Min. Defense)
11 Illicit activity	Coca cultivation (Min. Defense/UNODC)
12 Sparse pop.	Population density (DANE)
13 Remote	Access (Min. Transportation/Health)

Table 1: Operationalization of V-Dem responses in Colombia

ID Response	Indicator
14 Indigenous	Ethnic population (DANE)
15 Ruling party strong	Pres. margin of victory (RNEC)
16 Ruling party weak	

Assigning measures

Second, using geo-spatial data, these parameters are transformed into map polygons.

- layering of the subnational response variables revealing unevenness below the regional or departmental-level
- compatibility with variables beyond the V-Dem indicators for subnational democratic unevenness

Subnational democracy “scores”

Third, these concepts and measures produce a score of the unit under analysis (“map indicator”). Here, composite scores will provide an initial measure of operationalized concepts, revealing geographically located “pockets” of stronger or weaker democratization in Colombia.

Scoring is based on the aggregation of response values across all indicators.

Mapping methodology

General Pipeline

First, we decide on the level of analysis for our study. In this case, we observe that Colombia has fairly complete municipal-level data dating back to at least the 1950s, at least for a relevant amount of time-varying variables (e.g., population density, economic development, voting behavior, etc.).

Second, we collect and format data for mapping purposes. This is a multi-step process that involves grouping responses in some cases (e.g., responses 6-9 that ask to identify 4 cardinal directions). Grouping is necessary for some variables, but not all. Once grouped, the response categories themselves are “translated” into locally relevant information, which requires a degree of subjective decision-making. For example, how many votes or proportion of votes are necessary to measure “strong” support for the ruling party (question 15)? Or, which municipalities are considered part of the “North” (question 6) versus the “West” (question 8)? We privilege the use of granular data that will allow researchers flexible interpretations of these questions, as well as subsequent validation from country experts.

Third, we export the response grouping data-frames to `.shp` or `.csv` formats that can be added as data in Geographic Information System (GIS) software. In ArcMap, this can be done by joining data to the attribute table for the base layer (in this case, the shape-file provided by DANE). We then create and format layers based on the response groupings; for each response group, we visualize the locations where free and fair elections are held (and not held), and where civil liberties are stronger (or weaker).

Wrangling data

We undertake a series of steps to wrangle data from V-Dem and external sources. In order to facilitate merging tables to the shape-file maps of Colombia in ArcMap, we use a key ID (either the official DANE code for each municipality, or the ArcMap FID number). In the case we prefer to use the `sf` package, we can proceed to merge along the DANE code.

Note: This code below concentrates on cross-sectional data, but it will expand to time-series cross-sectional (~1900-2022).

Exercise: 2018 cross-section

This exercise is designed test the viability of mapping cross-sectional country-level data before expanding the dataset into panel format. Therefore, internal data from V-Dem is subset to 2018, and last five subnational responses are removed because they are not relevant for Colombia.

Table

External data from Colombia is drawn from a variety of sources. We merge this data at the municipal level using DANE codes. These codes are also used in `.dbf` tables, such as Colombia’s 2018 Census data

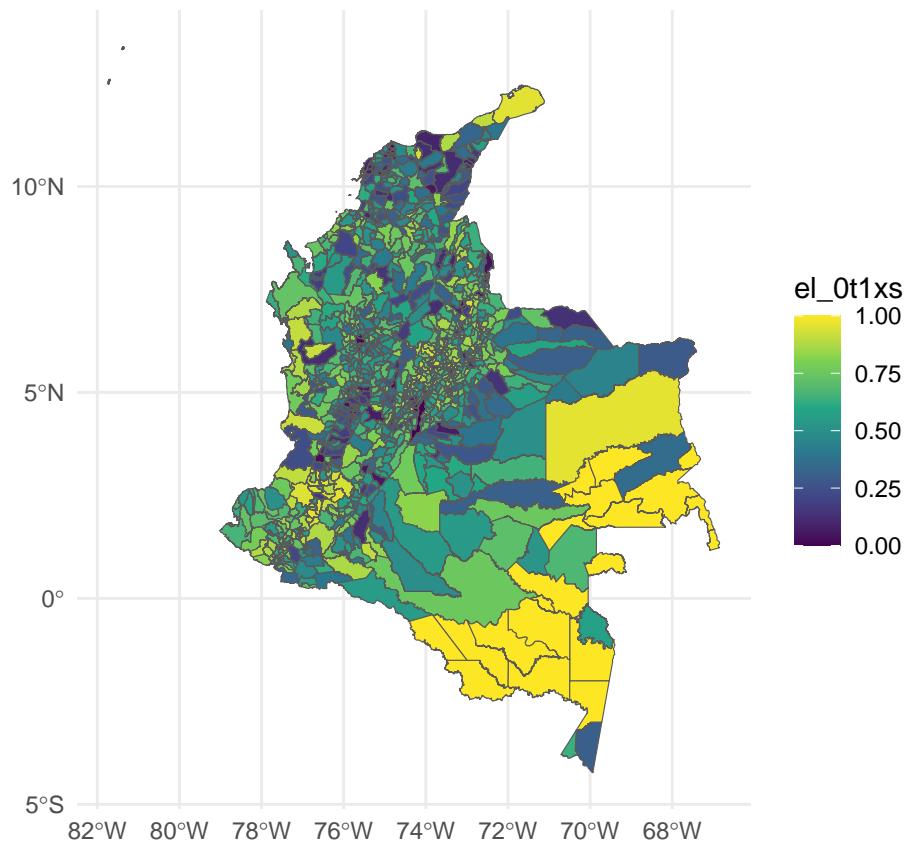
from DANE. The Census data Shapefile is useful for initial mapping visualization, containing not only the administrative boundaries but also contextual information. The [dictionary](#) is in the repository.

Note: To facilitate integration of the datasets across software, we export the Census 2018 attribute table from ArcMaps to .csv format, then load this data into R, preserving only relevant variables for merging purposes. While we use DANE codes to merge external data using R, we use the 2018 Census variable “FID” to merge .csv files into ArcMap. This is because ArcMaps automatically translates some string variables (such as the DANE codes) into a numeric ones. For this reason, we integrate the FID variable for the individual .csv files we write.

0-1: Rurality

The survey asks about elections and civil liberties in rural or urban locations. Rurality or urbanization can be measured in many ways. For example, the World Bank’s [Rural Access Index \(RAI\)](#) or European Commission’s [Global Human Settlement Layer \(GHSL\)](#) are exemplary. Similarly, [Waldorf and Kim \(2015\)](#) create an index that considers population size, density, remoteness, and built-up areas. [DANE \(2015\)](#) uses a classification based on OECD criteria that defines rural territories as those with 150 or fewer persons per kilometer squared, finding that over 75% of Colombian municipalities do not meet the threshold of 100 persons per km². From another perspective, DANE estimates that close to 4/5 of the population lives in urban centers. In this exercise, we take a demographic approach that considers the proportion of the municipal rural population over the total municipal population. This basic “Rurality Index” measure created by CEDE uses Census data (DANE).

The survey asks two questions, however, the systematized concept can take many forms. In this case, our V-Dem scores reflect the proportion of respondents who selected rural, urban, or both categories to describe election fairness and civil liberties. We can interact these proportions per question with the rurality index from CEDE (also a proportional score) to provide a more granular view of democracy at the municipal level based on our particular systematization. Another option is to create a dichotomous threshold between urban and rural variables and assign V-Dem scores accordingly. We can take the mid-point as a threshold for rural and urban municipalities: if more than half of the population is considered rural inhabitants, then the municipality is considered rural. This can be adjusted later if necessary. - As a first cut a good way to start -



2-3: Economic development

DANE provides data for Gross Domestic Product (GDP), known as Producto Bruto Interno (PBI) per capita. They offer a historical “retropolation” of department-level PBI dating back to 1980, and a municipal-level measure of “value added” dating back to 2011 (methodological documentation is included in the repository). Values are in Colombian pesos (COP) and scaled to \$1,000,000,000.

The metropolitan data extends farther back in time, making it more amenable to panel data analysis. Also, there may also be reduced measurement error given the difficulty of isolating economic productivity to municipalities.

Most departments are covered by the retropolation measure, but newer departments (many of which were inducted in 1991) are grouped into one unit, and therefore cannot be mapped based on DANE’s Departmental Code system without data intervention. These are: Amazonas (91), Arauca (81), Casanare (85), Guainía (94), Guaviare (95), Putumayo (86), San Andrés, Providencia y Santa Catalina (Archipiélago) (88), Vaupés (97) y Vichada (99).

Nearly all municipalities are measured by the Municipal Value Added (VAM) measure, however, data only

exists from 2011 onward.

For experimental purposes, we try both measures for 2018.

After creating the data frame for economic development at municipal and departmental levels, we must conceptualize “less” versus “more” developed areas before merging with V-Dem data. A simple option could be to dichotomize economic development based on median values. A more complex method would interact these municipal measures with V-Dem responses, as we did with rurality, but this would involve scaling and other steps beforehand.

Unclear how to set up the interaction option.

4-5: Inside or outside of capital city

The survey asks about elections and civil liberties “Inside / Outside the capital city”. In the context of V-Dem this likely refers to the national capital (Bogota), but it could also refer to municipal districts that are also department capitals. CEDE contains 2 variables of interest for systematization of this concept: distance from Bogota and distance from department capital. Even though we can assume respondents were considering the national capital, We will keep both in case we consider the alternative interpretation. Note: It may be possible to calculate this variable in ArcMaps using the field calculator, but this requires advanced techniques.

Although the survey asks two questions, technically these responses can be grouped into one dichotomous variable. We will therefore create one variable for each survey pair of survey questions that assign dichotomous values (0=capital, >0=non-capital). Nonetheless, we can also consider the relevance of proximity following a continuous scale. We therefore calculate this scale for the national capital distance using the range.

6-9: Cardinal directions

In order to map the four cardinal regions in Colombia, we are guided by external data from DANE who recently classified the country into 6 [macro-regions](#). We find a public-access table of municipalities classified into these regions [here](#). There are other divisions, such as the 5 regions used by CEDE studies (Andina, Caribe, Pacifica, Orinoquia, Amazonia), but for now we go with the prior classification.

We decide to group the regions in the following manner (see code below), creating a new variable called “Cardinal”. Then, we merge our V-Dem data related to responses #6-9 for each subnational survey question.

10: Civil unrest

11: Illicit activity

12: Sparse population density

- will likely covary with rurality

13: Remoteness

- will likely covary with rurality

14: Indigenous

15-16: National ruling party

The questions ask “Areas where national ruling party or group is strong / weak.” The most straightforward measure is a dichotomous variable where president receives a majority of the vote or not. A more variegated measure can leverage the percentage of votes cast for the president-elect vs. runner-up, for example, however this measure may need to account for blank votes (which may signify protest votes).

The measure for ruling party support used here will be: % of vote that supports ruling party. For first wrangle, I will choose 2018 runoff election (second round) between Petro and Duque (Duque won, so his is ruling party). This makes things easier because there are only 2 candidates, the winner and runner-up. Elections without second rounds (prior to 1994, as well as 2002 and 2006) may require additional coding.

The voting data requires more steps for cleaning. In this exercise, we will take all possible vote entries: top 2 candidates, as well as no mark, blank, or null ballots. We see that Duque won by high margin (over 26% on average) in 2018.

After creating the data frame for voting behavior at the municipal level, we must decide on the threshold for “strong” versus “weak” support of the ruling party. We could take one standard deviation or higher as the threshold, but even so: what is the criteria? In the interest of simplicity, we will create a dichotomous measure, with margin of victory (MOV) above 10 percentage points, applied in either direction (positive for winner Duque; negative for loser Petro). This seems like a standard difference, but further research could guide this threshold setting. - ideally, don’t dichotomize. Use a continuous scale. - cl stronger where RP is stronger and experts agree that's

Exporting data

Mapping data can be exported to a shapefile then imported into `sf` or ArcMap

Conclusion

- Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–546. https://www.cambridge.org/core/product/identifier/S0003055401003100/type/journal_article (Accessed April 15, 2022).
- Coppedge, Michael. 2012. *Democratization and research methods*. Cambridge ; New York: Cambridge University Press.
- Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics." *Comparative Politics* 31(4): 465. <https://www.jstor.org/stable/422240?origin=crossref> (Accessed November 6, 2023).
- Graudy, Agustina, Eduardo Moncada, and Richard Snyder, eds. 2019. *Inside Countries: Subnational Research in Comparative Politics*. 1st ed. Cambridge University Press. <https://www.cambridge.org/core/product/identifier/9781108678384/type/book> (Accessed January 29, 2024).
- Harbers, Imke, and Matthew C. Ingram. 2017. "Geo-Nested Analysis: Mixed-Methods Research with Spatially Dependent Data." *Political Analysis* 25(3): 289–307. https://www.cambridge.org/core/product/identifier/S1047198717000043/type/journal_article (Accessed February 16, 2024).