Why Urban-Rural Political Cleavages Do Not Generalize*

Noah Dasanaike†

First Version: February 14, 2025

This Version: April 10, 2025

Click here for the latest version

Abstract

Are urban–rural political cleavages generalizable? Prior evidence from the West and particularly the United States suggests a strong contemporary urban–left versus rural–right divide. In the first large-scale "out-of-sample" test using a new dataset of granular, geocoded election results covering 106 countries and nearly 10 billion votes (SAGE), I find this pattern inconsistently appears beyond Western democracies. In many countries, urban-rural weakly predicts vote choice; and where such polarization does arise, it may even invert the typical ideological pattern. To partially explain these findings, I develop a theory in which urban–rural cleavages arise from the spatial clustering of distinct voter attributes that enable parties to craft geographically targeted appeals. In contrast, in contexts of "discordant composition" where such voter attributes are not spatially concentrated, the traditional urban–rural divide is muted.

Word Count: 12,227

*My thanks to Feyaad Allie, Stephen Ansolabehere, António Câmara, Grzegorz Ekiert, Ryan Enos, Meghan Ford, Torben Iversen, Gary King, Adam Przeworski, and audiences at Harvard for comments and discussion. Rahsaan Maxwell and Michael Peress also provided useful comments. This research was made possible by funding from the Harvard Center for European Studies, the Harvard Davis Center for Russian and Eurasian Studies, and the Victor and William Fung Foundation.

†PhD Candidate, Department of Government, Harvard University

1 Introduction

Urban–rural divisions in political attitudes have increasingly been portrayed as the defining electoral cleavage of the twenty-first century. Such place-based differences bear significant consequences for representation, accountability, redistribution, and more (Rodden 2010). Yet, we remain largely uncertain about where political preferences are spatially polarized in this way and why. Although recent studies in the United States and comparable Western democracies such as Canada and the United Kingdom (Rodden, 2019; Mettler and Brown, 2022; Armstrong et al. 2022; Taylor et al. 2024; Huijsmans and Rodden, 2024) well document undeniable patterns of urban–left and rural–right voting, the broader generalizability of these findings is unclear. Even within these settings, the underlying bases for such cleavages have not been fully elucidated. Do urban–rural cleavages strongly characterize party systems outside Western democracies, and if not, what explains such variation?

I begin by conducting the first global-level, "out-of-sample" analysis of urban-rural political cleavages, then propose a potential theory explaining cases where they either do not emerge at all, or do so in the opposite preference-ideological direction of convention (i.e, rural-left and urban-right). Contrary to conventional expectation, I find little evidence of a generalized urban–rural electoral cleavage. Across thousands of country–election–party combinations over the past two decades, a one standard deviation change in urbanicity is associated with only a one-tenth average standard deviation change in party vote share. The pronounced urban–rural differences in voting behavior previously and again here observed in select Western countries do not generalize broadly. Notably, the extent of this cleavage does not predicate simply on economic development or industrial activity; to the contrary, there are a number of countries in *both* the Eastern and Western hemispheres, including advanced post-industrial democracies and developing states, that fail to considerably manifest such cleavage structure.

To account for the considerable heterogeneity in urban–rural cleavages, I propose one potential explanation that reconceptualizes urban and rural not as inherent sources of polarization but as potential aggregators of distinct voter blocs. In this view, urban–rural cleavages arise when politically relevant socioeconomic and institutional features, such as ethnicity, income, economic structure (agrarianism, industrialization, knowledge production), and public expenditure on education, cluster spatially in a manner that enables parties to craft targeted appeals. This shift in perspective moves the focus away from a deterministic effect of population density *per se* and toward an examination of how these correlated structural characteristics can be mobilized electorally. The resulting framework yields clear predictions about when and where urban–rural divides should emerge: they are most pronounced in early modernizing societies where urbanization and industrialization induced large-scale population sorting and in contexts where a labor-intensive agrarian peasantry dominates the rural electorate, and are less apparent in settings characterized by cross-cutting ethnic, linguistic, or regional cleavages that diffuse the concentration of politically mobilizable attributes.

A review of the countries comprising this majority condition favors the theoretical framework advanced here: when politically salient attributes do not cluster neatly into urban and rural domains, the spatial cleavage is muted; conversely, when distinct voter blocs form within these domains, parties can deploy targeted appeals, leading to pronounced spatial polarization that may manifest in either the conventional (urban–left, rural–right) or an inverted (urban–right, rural–left) preference-ideological pattern. For example, in post-industrial, knowledge- and service-based economies with ethnically homogeneous and economically deprived rural populations, urban electorates tend to support different parties than rural ones, reflecting the first-order relationships among wealth, education, ethnicity, and policy preferences. The coalescence in distribution of these population and structural attributes facilitates bundled and spatially concentrated party appeals. Conversely, where rural economies comprise heterogeneous organizations ranging from agribusiness to institutions of higher education, and where the demographics of urban and rural areas are less respectively distinct, urban and rural in turn are less likely to constitute a significant cleavage in the national party system.

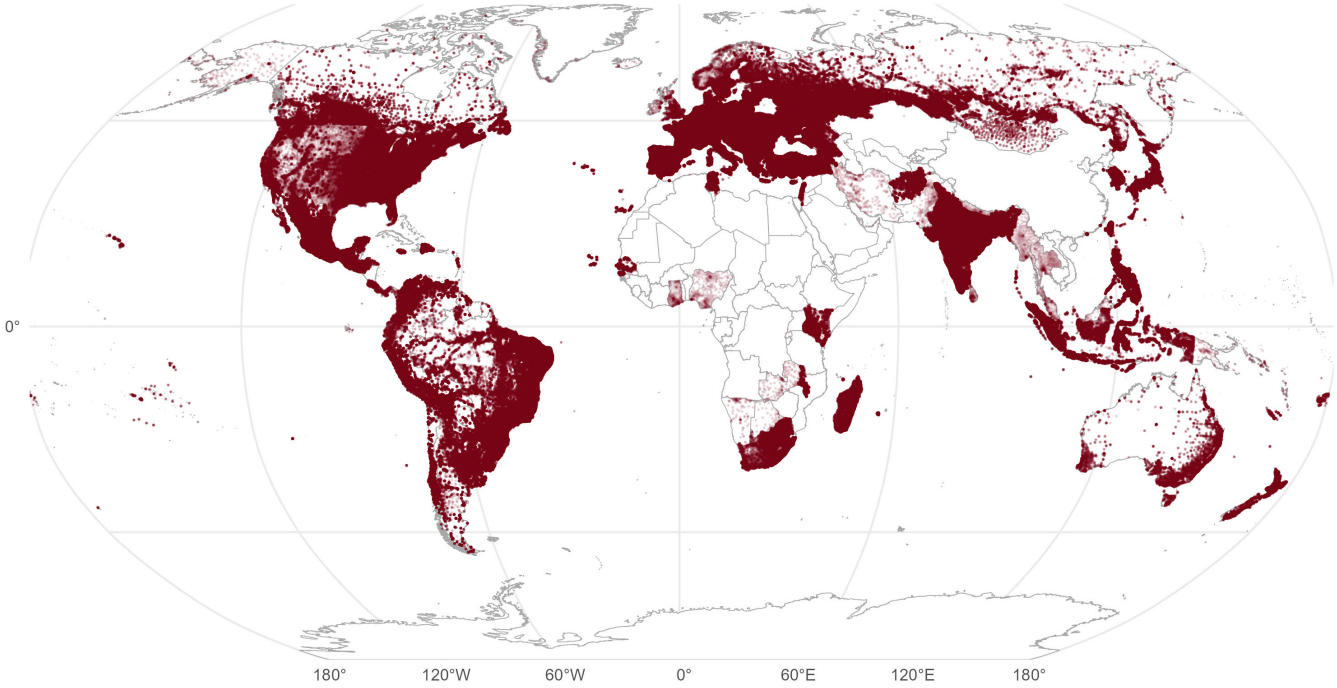
Urban–rural cleavage structure has attracted scholarly attention not only because political parties may garner differential support across these constituencies, but also because such spatial divisions are widely assumed to undergird contemporary left–right partisan polarization. To assess the generality of this claim, I develop novel measures of unidimensional left–right party ideology for thousands of legislative parties by matching Wikipedia-derived party tags with expert-assigned ideology scores. Using these measures, I confirm that many advanced, post-industrial democracies exhibit robust urban–rural cleavages in the conventional left-right ideological sense. Yet, I also identify countries ranging from Spain to Bolivia and Indonesia where urban areas vote more right-wing than their rural counterparts, a pattern that becomes especially pronounced when examining regions previously overlooked in the literature. In the majority of countries, urban and rural electoral preferences are only marginally distinct with relation to unidimensional ideological preferences.

These findings predicate on the ability to examine urban–rural cleavages (or the lack thereof) at scale. To do so, I present the Small-Area Global Elections archive, or SAGE, the most granular cross-national, geocoded elections dataset to date. SAGE encompasses lower house (or unicameral) parliamentary and presidential elections for 106 countries, representing nearly 10 billion votes cast across 8.2 million unique electoral units. For more than two-thirds of these countries, data are reported at the polling station level, and the average electoral unit contains fewer than 1,000 voters.[1] As Figure 1 illustrates, SAGE provides unprecedented spatial detail covering the entire spectrum of contemporary elections. The Economist Democracy Index reports, in the current year, 24 full democracies, 50 flawed democracies, and 34 hybrid regimes, or a total of 108 non-authoritarian regimes.

[1] As detailed in the Figure 1 notes, "electoral unit" in SAGE refers simply to the places for which votes are compiled. This ranges from the polling station, which is most common, to municipalities or formal voting "districts."

SAGE therefore covers, at a very high level of detail, national election results for the entire democratic world, as well as hybrid regimes and several current autocracies that previously held democratic elections.

Figure 1. Plot of all unique spatial points, or electoral areas, in SAGE.



Note: Each red point denotes the centroid of an electoral unit in the dataset. The meaning of "electoral unit" varies widely across settings: for most countries, the point denotes polling station addresses; in others, simply the centroid of a municipality, electoral district, or the equivalent. In some cases, like Pakistan or Mongolia, election results are collected at the polling station level but geocoded at a higher level of aggregation. Because of this latter category of cases, there are 6 million unique coordinates representing the 8.2 million unique voting "places."

Beyond its breadth, SAGE is often the only readily accessible source of small-area electoral data for many countries. In addition to vote shares, the dataset includes a wide range of election- and unit-level attributes such as indicators for early voting, snap elections, electoral system types (e.g., majoritarian, proportional representation, first-past-the-post, advanced voting), and provisions for absentee voting (e.g., from abroad). It also provides party and candidate names manually matched across multiple languages alongside measures of turnout. Perhaps most importantly, I geocode the majority of the data using either official administrative boundaries (when available) or, alternatively, algorithmically generated boundaries (for example, Thiessen polygons centered on polling sta-

tion addresses). Detailed documentation outlines country-specific nuances in data processing and geocoding, and every entry in SAGE adheres to a standardized, ready-to-use format. Appendix Table A1 offers comprehensive coverage details by country.

In addition to these detailed voting results, I incorporate data on more than 2.6 billion building footprints derived from satellite imagery to construct a standardized measure of the urban–rural continuum.[2] Original survey evidence indicates that the average distance between a building and its nearest neighbors correlates more strongly with individuals' perceptions of urbanicity than does conventional population density. Accordingly, I match these global building data to each of the 8.2 million electoral units in SAGE and compute a measure of urbanicity within each unit based on nearest-neighbor building distances. For every country–election–party combination in SAGE, I estimate the correlation between a party or presidential candidate's vote share and the level of urbanicity-rurality, thereby quantifying the extent to which voter bases are spatially polarized along urban–rural lines.

The paper proceeds as follows. I begin by reviewing the relevant literature and empirically test for the presence and intensity of urban–rural divisions across democracies using the SAGE dataset. Given substantial heterogeneity in the extent of urban-rural cleavages globally, I propose one possible explanation: an alternative theoretical framework that reconceptualizes urban–rural cleavages as outcomes of spatially clustered voter attributes rather than inevitable consequences of population density. Following this, I examine the mechanisms underpinning these patterns—with a particular focus on instances of "discordant composition," where the expected spatial clustering of politically salient factors is absent—and analyze the roles of economic activity, sociocultural structure, institutional legacies, and distinct modernization trajectories. I conclude by discussing the implications of these findings for our broader understanding of geographic political cleavages and by outlining promising directions for future research.

2 The Urban-Rural Political Divide

Since classical antiquity, Western societies have been characterized by a pronounced urban–rural polarization–a dynamic that scholars (e.g., Nathan, 2022; Huijsmans and Rodden, 2024; Rachman, 2018) have recently dubbed the "great global divider" of modern democracies. The electoral manifestation of this longstanding cleavage, arguably rooted in the democratic revolution and later formalized in discussions of center–periphery conflict (Lipset and Rokkan, 1967), remains central to debates over political polarization. Yet, while empirical studies in advanced industrial democracies frequently document distinct voting patterns between urban and rural constituencies (Ford and Jennings, 2020; Gimpel et al., 2020; Kenny and Luca, 2021), the extent to which these patterns generalize across diverse political settings is far from clear.

[2] "Footprints" refer simply to the birds-eye view boundary edges of buildings.

For example, Rodden (2019) contends that in the United States and other Western Anglophone contexts, leftist parties initially mobilized support among newly industrialized urban workers, and subsequently sustained their appeal among urban electorates in the post-industrial era. In a similar vein, Cramer (2012, 2016) posits that a "rural consciousness" emerged in response to urban-biased policymaking, thereby fueling rural support for right-wing populism. Gallego et al. (2019) and Maxwell (2019) argue that certain types of voters (British and Swiss, respectively) sort into cities, engendering a divide between cosmopolitan urbanites and nationalist rural voters. Rodríguez-Pose (2018) identifies "places that don't matter", or areas of considerable economic decline, as key sources of political backlash in the West; here, economic stagnation and population decline tend to characterize rural, rather than urban, constituencies.

Focusing more descriptively than mechanistically, various others demonstrate urban-rural cleavages in the same setting(s), that is to say almost exclusively the United States or its close cultural neighbors (e.g., Canada, the United Kingdom). Mettler and Brown (2022) attribute urban-rural polarization in part to American institutional arrangements (single-member districts, winner-take-all elections, federalism) that disproportionally grant political leverage to sparsely populated regions. Armstrong et al. (2022) identify several possible explanations for an emergent urban-rural divide in Canada, where agrarian socialist movements have historically been the focus of study; like Rodden (2019), these authors focus primarily on postwar urbanization and alignment between leftist parties and highly educated urban knowledge economy workers. Taylor et al. (2024) offer a broad overview of compositional and contextual explanations for urban-rural divides in Canada, Great Britain, and the United States, encompassing place-based identities that reinforce extant spatial polarization (Jacobs and Munis, 2020), differential exposure to ethnic diversity that shifts preferences on immigration (Sobolewska and Ford, 2020), and economic and cultural status anxieties induced by local conditions (Gidron and Hall, 2017).

Although these findings undoubtedly illuminate select origins of urban-rural political divisions, they remain confined to particular (Western) cultural and regional contexts. To rigorously test the generalizability of urban–rural cleavages, a more comprehensive, cross-national approach is needed. Doing so first requires overcoming the constraints imposed by previous high-level electoral data collection.

3 Data and Measurement

3.1 Small-Area Global Elections Archive

The study of elections is the study of mass political power in its most universal institutionalized form. Alongside survey data, elections are one of the few avenues through which political attitudes may be observed, and unlike survey data, elections are held regularly, in most countries, and at massive scale. Given

perceived anonymity in the ballot box, voters are more comfortable revealing their privately held preferences when compared with non-anonymous surveys (Gerber et al., 2013). However, systematic data for the analysis of cross-national elections remains limited. Previous efforts at cross-national election results compilation are made only at high levels of aggregation (for instance, Gerring et al., 2015; Kollman et al., 2024), and spatial information are either missing entirely or are difficult to merge in.

Put simply, elections are held nearly everywhere, they contain useful information on mass political attitudes, and they are ascribed substantial importance, yet their systematic and especially granular study has been inhibited by the lack of centralized data collection. In consequence, our theoretical understanding of diverse phenomena ranging from close-race elections (Eggers et al., 2015) and cleavage structures (Hooghe et al., 2002) to compulsory turnout (Fowler, 2013) and partisan realignment (Piketty et al., 2022) is biased by the settings in which detailed empirical analyses can be undertaken. Indeed, recent work has compiled polling-station election results for specific countries in well-studied regions like the United States or Spain (Pérez et al., 2021; Baltz et al., 2022; Benedictis-Kessner et al., 2023), but not cross-nationally, and never at the greatest scale possible. The consequences are myriad and include a systematic over-representation in the literature of countries for which election results are readily available, and specifically those countries with high-quality data environments.

The geographic and temporal scope of the dataset is determined by a simple procedure. Employing a minimalist, Schumpeterian conception of democracy, I examine whether each country in the world has held democratic, national-level elections in the previous several decades. For those that have, within reasonable limits of public contestation and electoral inclusivity (Dahl, 1971), I find the smallest unit of results ever reported at by each country's electoral agency, and then collect election results for all years where reporting is done at that level. For instance, given that Venezuela is currently a dictatorship, I collect polling station-level results for the 2013 presidential election where Nicolas Maduro first ascended to power, and collect polling station-level results for all national elections held in Russia since 2000—despite allegations of potential fraud in both settings. I also collect data for "special" territories like Greenland, Hong Kong, and Kosovo.

Appendix Table A1 provides brief details of country-by-country coverage. Appendix Section A2 describes how I collected and geocoded election results. To provide a more detailed example of the construction of SAGE, I then detail in Appendix Section A3 the process I undertook for digitizing the 2019 Lok Sabha elections in India, which with more than 900 million voters represented the largest election ever conducted at the time[3]. In summary, SAGE brings the availability of election results for the 2019 election down from an average of 2 million voters per each of 543 constituencies to 1,000 voters across nearly a million polling stations, and I match by hand nearly 10,000 candidates to their respective parties.

[3]Followed only by the 2024 Indian Lok Sabha elections.

Appendix B comprises the SAGE codebook, detailing for each country the data sources utilized, sample electoral units, aggregate election results by year, and maps of electoral geography.

3.2 Measuring the Urban-Rural Continuum

Testing for urban-rural cleavages globally requires standardized measures of urban and rural. Existing measures are inadequate for this task, a problem in part methodological and in part theoretical. Scott et al. (2007) note that more than 30 different definitions of "rural" are used across British government departments. Long et al. (2021) review operationalizations of rurality across federal agencies in the United States and identify examples such as "proximity to urbanicity" or the "sharing of borders with urban areas," circular definitions the advantages of which are difficult to conceive. According to the website of one federal agency (that which oversees health care access to "rural communities"),[4] "any area that is not urban is rural." At the individual level, Nemerever and Rogers (2021) and Trujillo (2024) demonstrate dissonance between urban-rural self-identification among survey respondents and common "objective" measures such as those derived from population density. An analysis conducted by Pew also reveals mismatches between the perceptions of European survey respondents and such measures: for instance, 59% of Greeks living in "suburban" area classifications and 7% of those in "rural" classifications derived from population density reported living in "urban" areas.

I argue instead that urban and rural are heuristically best understood as opposite ends of a scale defined by physical distances between markers of human activity. While population density therefore tends to work on average as an approximation of urban and rural, there are many cases where it fails: for instance, a city downtown without any residential zoning, where the "population" density is low but the degree of urbanicity would be commonly understood as high. In contrast, I further argue that the distances between buildings themselves map onto an intuitive "know it when you see it" conceptualization of urban-rural (Kaufman et al., 2021). Thus, to measure the degree of "urbanicity" and "rurality" of different places, I use global building polygons to estimate the distances between buildings (Nemerever and Rogers, 2021). To do so, I obtain all 2.6 billion global building footprints from Google's Open Buildings and Microsoft Maps and match them to the 8.2 million unique geocoded electoral units in SAGE. For a sample of buildings in each unit, I calculate the distance to $k \in \{30, 90, 270\}$ other buildings across the full country.[5] Appendix Section A3 formalizes the computational process and demonstrates the high internal consistency of measures across different values of k.

Given the predominance of population density as a proxy for urban and rural in the literature, I conduct a validation exercise comparing it against the

[4]The Health Resources and Services Administration: "Defining Rural Population".

[5]These values of k were selected arbitrarily but, as I show in Appendix Figure A2, there is strong correlation between estimates obtained from different k.

proposed measure. Since much of the related prior scholarship focuses on the United States, I asked 1,000 Americans in an original nationally representative survey to rank "the place [they] live" on a five-item scale from "very rural" to "very urban". I matched each respondent to voting areas using their postal code and averaged the building distance measure by self-reported urban-rural status. Appendix Figure A1 reveals that average nearest-neighbor building distances strongly correlate with individual perceptions of urbanicity-rurality (*criterion validity*). Buildings are statistically of different degrees of separation in the places where respondents who selected "somewhat rural" live compared to the "very rural" or "neither" respondents; the same holds true for "somewhat urban" compared to "very urban" respondents. This capacity to smoothly discriminate between each scale ranking starkly contrasts with population density, where only "very urban" differs markedly from the other options. Alongside SAGE, I release these global measures of nearest-neighbor building distances at the grid-cell and voting area levels to facilitate future research.

In addition to nearest-neighbor distances, I also estimate dyadic *between*-building distances as a robustness exercise. While this does tend on average to correspond to urban and rural, there are various land-use cases where it fails. For instance, a heavily agricultural area where each property comprises highly proximate residential and commercial (e.g., farm house and barn) buildings will have low between-building distances that increase only as you increase the size of an arbitrary spatial buffer intended to redress the very same problem. Conversely, a bounding box comprising urban sprawl will have buildings at each corner and thus the distribution of distances will approximate the normal with heavy tails. Figure A1 demonstrates that inter-building distances outperform alternatives at separating "very" and "somewhat" urban, while nearest-neighbor distances display stronger separation among types of rural respondents. Both measures more strongly correlate with self-reported urban-rural classifications than population density.[6]

4 Is There a Global Urban-Rural Cleavage?

To assess the extent to which urban–rural cleavages characterize electoral politics globally, I conduct a large-scale empirical test across diverse political contexts. Specifically, for each country–election–party combination in SAGE, I estimate a simple bivariate regression between vote shares at the small-area electoral-unit level and separate measures of urbanicity–rurality (both population density and nearest-neighbor building distances). The estimated coefficients capture the degree to which changes in urbanicity correlate with party vote shares. If urban–rural polarization is a robust and generalizable phenomenon, we would expect these correlations to be systematically large and consistently signed across a broad set of countries and elections.

[6]The respective absolute value correlations for inter-building distances, nearest-neighbor distances with $k = 90$, and population density with survey response categories are 41.8%, 33.5%, and 31.5%.

Figure 2. Correlations between urbanicity-rurality (building distances) and party vote-share.

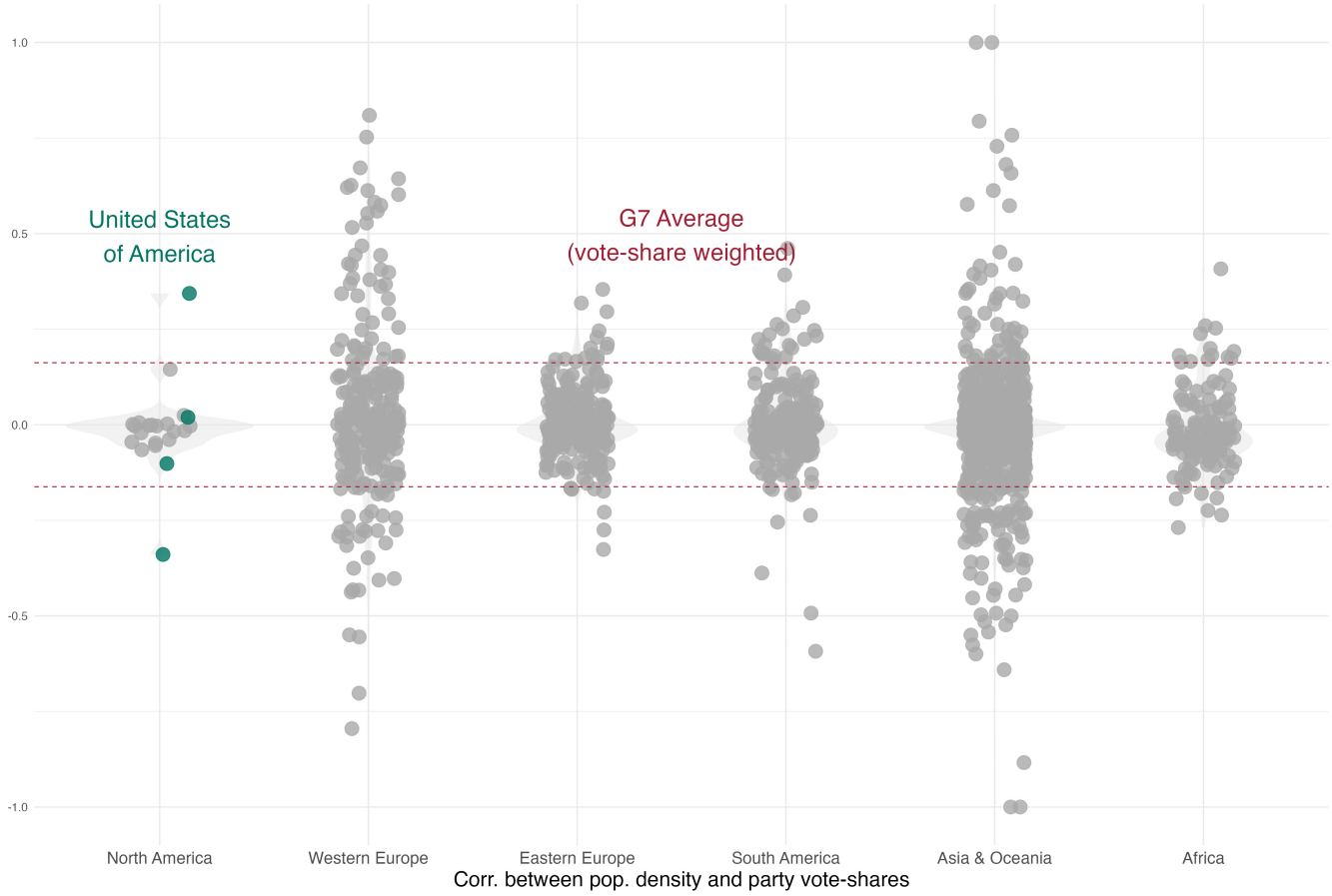
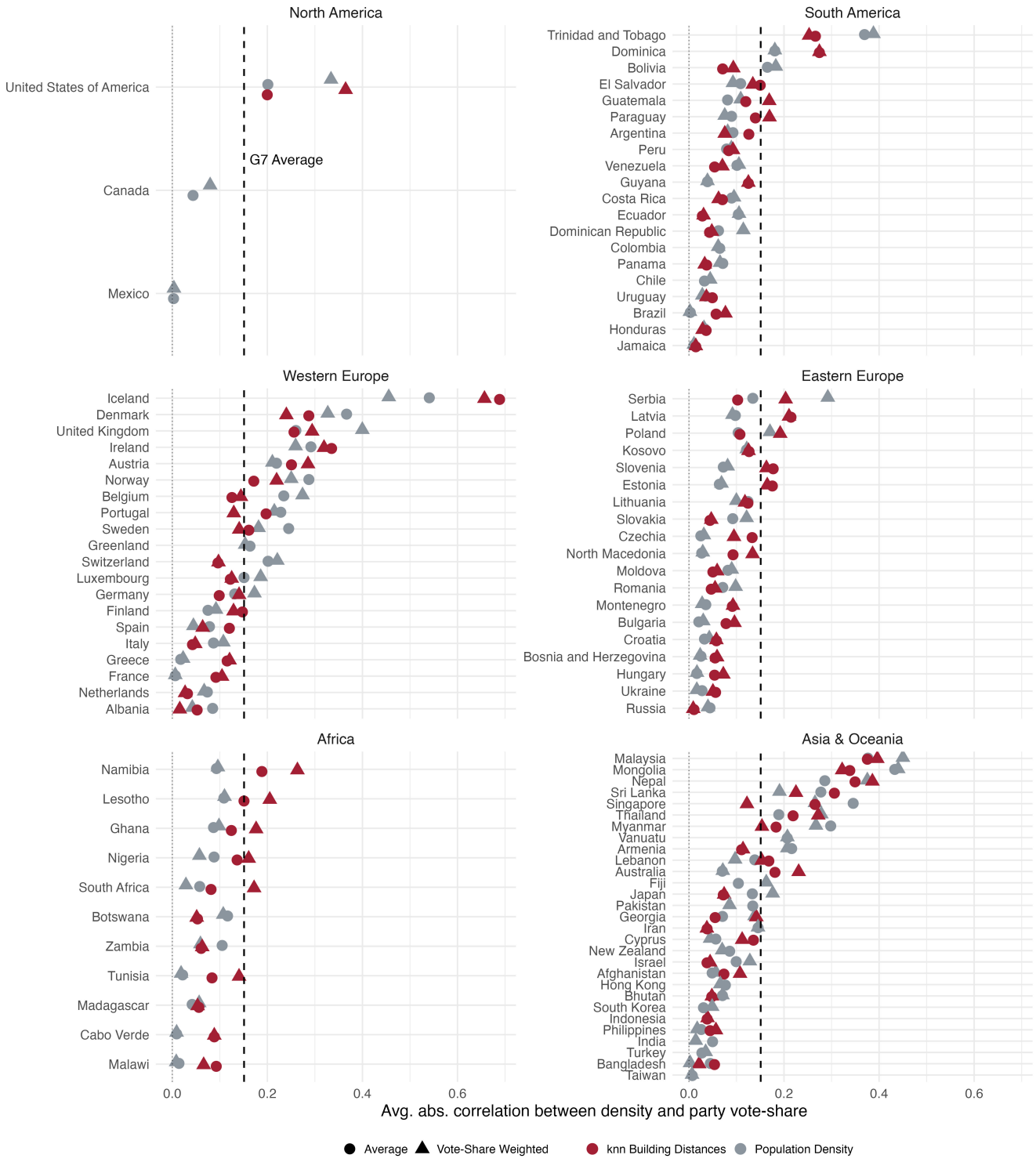


Figure 2 plots the correlation values at the party level by region for the most recent election in each country using population density. Figure 3 plots the average absolute values of these correlations by country, measure, and weighting scheme, providing a comparative snapshot of the extent to which urban–rural divides structure party competition across electoral democracies. The results reveal substantial cross-national variation. The G7 average, which offers a heuristic for wealthy Western democracies, far exceeds the extent of urban-rural divides elsewhere. While some countries exhibit strong urban–rural cleavages, in many cases the relationship between urbanicity and electoral behavior is weak or even absent. Neither does the magnitude of these cleavages systematically correspond to conventional indicators of economic development or industrialization. Advanced post-industrial democracies do not uniformly exhibit strong urban–rural divisions, nor do developing states consistently lack such polarization. Instead, urban–rural cleavages appear in a diverse array of contexts, while others–despite structural conditions that might suggest a natural divide–show little evidence of spatial polarization in voting behavior.

Figure 3. Absolute correlation of rurality-urbanicity (building proximity) and party vote-share.



Notably, the results do not significantly change given different measurements of urban-rural, nor weighting schema for aggregating across parties. As an additional robustness check, I also assess whether these correlations persist at higher levels of spatial aggregation. Specifically, I recursively grouped polling stations (or other small electoral units) at the next highest spatial level reported (for instance, from polling station to sub-region to region). At each level separately

by country, I recalculated both vote shares and the average k-nearest-neighbor building distances within those aggregates, and re-estimated the bivariate relationships. Across nearly all countries, the direction and approximate magnitude of the observed relationship between urbanicity and party vote shares remained stable, even as the sample size was drastically reduced at coarser levels of spatial aggregation. Moreover, there was no systematic decline in the share of cases with statistically significant coefficients at these larger aggregations, indicating that the finding of considerable cross-national heterogeneity in urban–rural cleavage strength is not an artifact of ultra-fine-scale data.[7]

Appendix Figure A3 presents trends in the extent of urban–rural electoral polarization over time by plotting changes in the strength of urbanicity–vote-share correlations across elections. Appendix C plots urban-rural party vote-share correlations over time by country. Interpreted cautiously given substantial variation in the range of cross-temporal coverage by country, I find that shifts in urbanicity-rurality correspond to an average tenth of a standard deviation (absolute) change in party vote-share for nationally contested elections. There is a marginal but largely indiscernible decrease in the size of this correlation over time, implying that there is *not* a global shift towards urban-rural electoral polarization. This result is surprising in relation to the extant scholarship: as low-income early modernizers developed and moved along the Kuznets curve, sectoral divisions increasingly formed along urban-rural lines as industrial blue-collar and then knowledge economy white-collar workers clustered in cities (Krugman, 1991; Iversen and Soskice, 2020). At least in this Western developmental story, shifts in economic inequality and subsequent class coalitions determined the focal political cleavages (Lipset and Rokkan, 1967; Esping-Anderson, 1990). However, that we fail to observe an upward trend in urban-rural electoral divisions here implies, firstly, that political change does not necessarily follow the form of preceding structural transformation; and secondly, that, akin to the distinct nature of their modernization (Gerschenkron, 1962), developing countries are likely forming political cleavages wholly distinct from those observed by Lipset and Rokkan (1967) in select advanced industrial democracies.

One important note should be made with respect to these findings. The results in Figure 3 and later in Figure 4 are country-level averages and in turn represent the extent of urban-rural cleavages in the respective country's party system *at large*. This does *not* mean that support for individual parties within a given setting do or do not themselves vary with density. The theory of discordant composition that I later present seeks to explain both party system- and party-level variation in electoral support among urban-rural lines. Towards this end, Appendix C

[7] For computational efficiency, I undertook this procedure for a sample of 70 countries. For 40 of these countries, I was able to aggregate once, and for 17, I aggregated twice. In the first case, there were 1,230 country-parties, out of which 538 had a statistically significant OLS regression coefficient regressing vote-share on average nearest-neighbor building distance with $k = 30$. 67.4% of those pairs, or 363 of the original 538, retained statistical significance when aggregated once. In the second case, there were 448 country-party pairs; 207 were statistically significant at the lowest level of aggregation, 132 (63.8%) when aggregated once, and 149 (72.0%) when aggregated twice.

plots urban-rural party vote-share correlations over time by country.

4.1 How Should We Think About Urban and Rural?

Why do we observe these results? Although a substantial body of scholarship documents the existence of an urban–rural cleavage in contemporary electoral politics, such work has focused almost exclusively on post-industrial knowledge economies. To address this gap, I introduce a simple probabilistic framework that yields clear expectations about why certain urban and rural areas vote differently and why these patterns may vary across contexts. I argue that urban and rural divisions emerge as a consequence of the spatial concentration of distinct clusters of voters, economic activities, and institutional arrangements that correlate with density. In turn, these spatial concentrations enable political parties to target specific constituencies, thereby generating discernible and distinct voter coalitions.

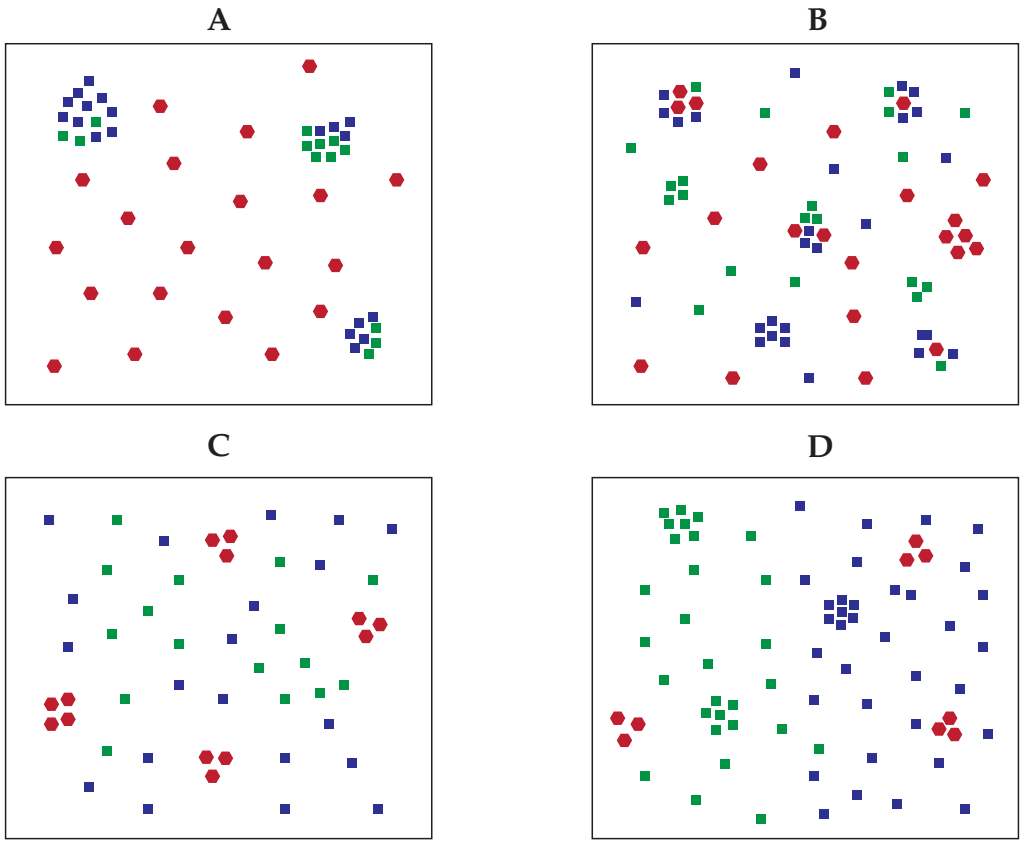
In advanced industrial democracies, the proposed framework appears almost self-evident. Modernization and industrialization in these settings have historically been accompanied by pronounced economic sorting, a process that unfolded concurrently with democratic expansion. In such contexts, key economic activities and demographic characteristics such as religiosity and ethnic composition are routinely bundled with population density. As logically concluded by Krugman (1991), industrial activities in these early modernizers *tend(ed)* to take place in urban areas, and agricultural activities in rural ones, for straightforward reasons of coordination: resources are extracted from the low-density areas where they naturally occur (e.g., lumber, coal, oil) or are grown (land-intensive farming) and are transported to, and processed in, centrally located areas. Over time, migration patterns have further reinforced these differences, as voters move from rural to urban areas in pursuit of more varied economic opportunities, specialized goods and services, and other associated benefits (Christaller 1933).

Even in analyses of latecomer industrializers (Gerschenkron 1962), the prevailing assumption is that during and after modernization, whenever it happens, the educated and white-collar cosmopolitans will sort into cities; agrarian peasants, whose labor is supplanted by technology- and capital-intensive production, will sort into cities and transform into urban workers; and those holding systematically disadvantaged ascriptive characteristics (Horowitz, 1985) will sort into cities, where diversity of economic opportunity offers the greatest labor market access. For example, in Great Britain, the least educated, low-income, and ethnically homogeneous segments of the population—often characterized as being "left behind" in the knowledge economy—are disproportionately found in low-density electoral districts. In such circumstances, rural-based parties are likely to emerge and subsequently encounter urban parties mobilizing a contrasting electorate. This coincidence has in much of the West made programmatic party appeals to spatially cleaved and demographically bundled coalitions of voters a low-cost option for consistent electoral results.

4.2 Discordant Composition

However, the logical sequence outlined above is disrupted when the key factors that form the basis of voter coalitions (such as education, economic activity, and demographics) do not neatly align with population density. I refer to this phenomenon as "discordant composition." In contexts of discordant composition, "urban" and "rural" differ only in terms of density rather than in the substantive attributes that typically shape electoral behavior. Consequently, if urban and rural areas do not also cluster distinct socioeconomic and institutional characteristics, a discernible urban–rural cleavage is unlikely to emerge within the party system. Figure 4 illustrates this point by contrasting voter distributions: Panels A and C depict scenarios in which spatial bundling of politically relevant attributes leads to a clear cleavage (with Panel A resembling the typical configuration in advanced Western democracies and Panel C representing an inverted pattern), whereas Panels B and D present hypothetical cases in which discordant composition precludes such polarization.

Figure 4. Illustrations of urban-rural bundling (A and C) versus discordant composition (B and D).



Examples abound of when discordant composition precludes the formation of an urban-rural voting cleavage. Regionally-concentrated ethnic parties, such as the Swedish-speakers' *Swedish People's Party of Finland* or the Kurdish *Peoples' Equality and Democracy Party* in Turkey, tend to proliferate across cities and countryside alike, because their appeals are not spatial in an urban or rural sense but spatial

across the territory of the ethnic group in question. So, too, do catch-all nationalists in India and Russia draw votes from across all types of constituencies, because their appeals are across the entire *nation*. Rural and urban areas alike with local economies centered around higher education and research, which sort their unwilling participants in such a way as is often contrary to those participants' locational preferences, tend to favor the same ideologically left-leaning parties. Improved mobility means that knowledge economy workers in the United States and Western Europe can live in low-density exurbs, work from home or in the high-density city centre, and vote for parties that tend to represent the latter types of constituencies (the consequences of which have yet to be fully realized). More broadly, when spatially concentrated appeals are inconvenient, owing to discordant composition, or misalign with party objectives (for instance, representing a particular region or ethnic group), urban-rural cleavages are unlikely to form.

Existing theories can be reframed to accommodate the notion of discordant composition. Maxwell (2019) characterizes urban-rural polarization in Switzerland as a "second-order manifestation" of compositional divisions that arise from self-sorting into areas of differing density. Cramer (2012) similarly contends that the political distinctiveness of rural voters stems not from an inherent rural identity but from the spatial clustering of inequalities: in particular, the concentration of wealth, access to public goods, and localized centers of political power in urban areas. Brown and Mettler (2022) document that in the United States, "areas with higher percentages of less-educated residents, a higher presence of evangelical congregations per capita, and higher levels of anti-Black racism, each more prevalent in rural areas than urban areas, shifted their support to Republicans." It can be intuited that if cities were instead less educated, more religious, and more ethnically nationalist, the Republican Party, or some other like it, would seek to capture the urban vote, and urban areas would tend more right-wing. If urban and rural areas were characterized by the same types of economic activity and levels of household income, education, religiosity, and ethnic nationalism, it is probable there would be no geographic cleavage at all.[8] In seminal early work on urban-rural cleavage structure, Tarrow (1971) argues that incomplete agrarian revolution in Italy resulted in a heterogeneous rural party system, because class "became a legitimate symbol around which to organize competing local groups." Conversely, in France, where the French Revolution homogenized rural land ownership among smallholding peasants, there was little capacity for *localized* political conflict on the basis of class. As earlier demonstrated (see Figure 3), the former case is today characterized by very little urban-rural left-right divide in part because of the persistence of discordant economic-territorial composition, whereas France, with a long-homogenized rurality, evinces substantial urban-rural left-right cleavage-preference alignment.

To summarize, when urban and rural areas each cluster voters, economic ac-

[8] Assuming an error term comprising assumed contextual effects of dense environments, wholly separate from population composition, economic activity, and institutional features as described here–though they have yet to be compellingly demonstrated.

tivities, and institutional features in a manner that allows political parties to mobilize distinct electorates along key dimensions of conflict, a discernible geographic cleavage is likely to emerge. This dynamic presupposes that political actors are engaged in national vote-seeking rather than pursuing region-specific or issue-specific representation. Absent such spatial clustering of politically relevant attributes, there is little reason, a priori, to expect significant urban–rural polarization in electoral preferences across the party system.

5 Left-Right Preference Structure

The theory I propose regarding the conditions under which urban–rural cleavages emerge generates straightforward predictions about the ideological preferences of voter coalitions. Simply put, the ideological appeals made by parties in the presence of an urban–rural cleavage should correspond to the underlying attributes that define urban and rural constituencies. If political and socioeconomic characteristics that predispose voters to support leftist parties are concentrated in rural areas, then rural electorates should lean ideologically left. To avoid extrapolating a broader theory of political preferences beyond urban–rural cleavages, I rely on well-established associations: religiosity and conservative preferences (e.g., Stegmueller, 2013), economic activity and voting behavior (e.g., Ebeid and Rodden, 2006), land tenure and redistribution preferences (Linz, 1976), and ethnic majoritarianism and minoritarianism (Wolfinger, 1965), among others. In this sense, I construct a strictly compositional theory of urban-rural cleavages, though one inclusive not only of population but economic activity, institutions, and other spatially distributed features that influence preference formation.

It is important to note that while I conceptualize ideological divisions explicitly along a unidimensional left–right spectrum, the theoretical framework extends to other dimensions of political conflict, provided that they correlate with spatial composition. For instance, in Indonesia, the legacy of anti-communist authoritarianism precludes the overt identification of major parties with leftist policies, making the core political divisions primarily religious rather than economic. In such a context, if secular voters predominantly reside in urban areas and Islamist voters in rural areas, urban–rural cleavages may instead manifest along a globalist-logistical or secular-religious axis rather than the conventional left–right spectrum. Similarly, in settings where environmental or cultural issues dominate political discourse, urban–rural cleavages may be structured along green-alternative-libertarian versus traditional-authority-nationalist (Hooghe and Marks, 2008) or technocratic-populist lines. The key insight is that urban–rural polarization does not emerge inherently from density alone but from the spatial clustering of politically salient voter attributes, with natural but necessary implications for left-right preference distribution.

5.1 Measuring Left-Right Party Ideology

Beyond measuring the existence of urban–rural cleavages, assessing their ideological direction–whether urban areas systematically lean left or right relative to rural areas–requires party-specific ideology scores across countries. However, existing datasets are limited in coverage. The Manifesto Project (Dinas and Gemenis 2010) assigns ideological placements to 856 parties across 67 countries, while the Global Party Survey (GPS) (Norris 2020) provides expert-coded ideology scores for 1,043 parties in 163 countries. By contrast, SAGE includes over 7,000 unique candidates and parties, necessitating a broader and scalable approach to ideological classification.

To estimate party positions along a unidimensional left–right scale, I employ a straightforward method leveraging existing data sources. Political parties on Wikipedia have "info boxes" that list, from a common set of labels (e.g., "right-wing populism," "liberalism"), the labels that characterize the dominant ideology of each party. I collect all labels for the approx. 90% of parties in SAGE with Wikipedia pages.[9] Given parties that also appear in the GPS, I calculate the average GPS ideology score (an average of the GPS "economic left–right" and "social liberalism–conservatism" scores) assigned to each label, then for each party average across the score of each assigned label to get a single measure of party ideology. This method allows me to assign ideology scores systematically across a far greater number of parties than existing expert-coded datasets. Examples of these ideology classifications, along with demonstration of high correlation with existing measures (V-Party, Chapel Hill Expert Survey, Manifesto Project), are provided in Appendix Sections A5 and A6.

5.2 Results

To formally test the relationship between urbanicity and ideological polarization, I next regress party vote shares on party ideology scores across all country–election–party observations in SAGE:

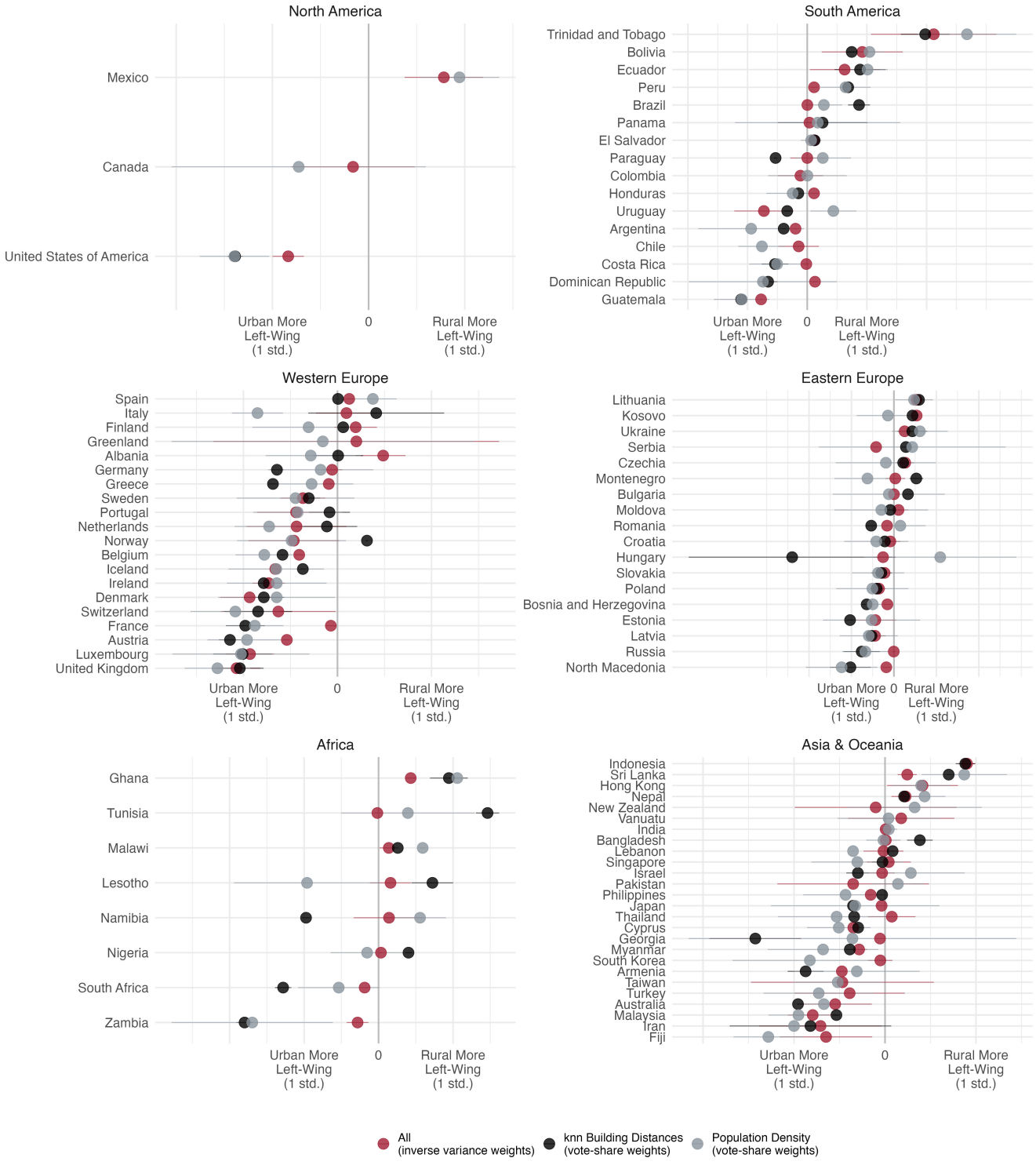
$$\text{Urban-Rural Vote Corr.}_{pe} = \alpha + \beta \text{Ideology}_{p} + \epsilon_{pe}, \text{ for all } c \in C. \tag{1}$$

The coefficient β captures whether a rightward shift in party ideology corresponds to greater electoral support in rural or urban areas. I estimate three separate models for each country. First, I regress party ideology on the bivariate regression estimates from Figure 3, weighting by overall party vote share, separately for population density and nearest-neighbor building distance measures of urbanicity. Next, I specify a model that incorporates both measures, weighting observations inversely by variance to adjust for uncertainty in the first-stage regression estimates. To mitigate bias from the time-invariant nature of the ideology measure, I restrict the analysis to the most recent election year for each

[9] This is similar to Herrmann and Döring (2023), which estimates party positions using Wikipedia-derived classifications.

country, separately for legislative and presidential elections. When appropriate, I include fixed effects for measurement approach and election year.

Figure 5. Average ideological direction of urban-rural political divide by country, across all years.



Note: Countries with high missingness in party ideology scores (fewer than 3 unique party-ideology matches), like Madagascar, are excluded. Iran is an exception given that

both presidential-contesting parties are measured. For joint lists or coalitions, I manually matched to the left-right score prediction of the dominant party(ies).

The results are presented in Figure 5, showing the average ideological direction of urban–rural political divides by country and region. The further left a point is, the more left-leaning urban areas are relative to rural areas, conditional on a unidimensional conceptualization of party ideology. With few exceptions, nearly all advanced democracies conform to the conventional expectation of urban areas voting more left-wing and rural areas more right-wing. However, while the inverse pattern–urban areas leaning more right-wing than rural ones–is less common, it is notably more prevalent in Latin America and Asia, regions where the literature has historically overlooked urban–rural ideological divisions.

6 Varieties of Urban-Rural Left-Right

The findings in Figure 5 suggest that while urban–rural differences in voting behavior are widespread, their ideological direction and intensity vary significantly across countries. In some cases, urban and rural areas exhibit strong and opposing partisan preferences, while in others, differences are muted or even reversed. To systematically categorize these patterns, I identify four distinct configurations of urban–rural left–right cleavages, as outlined in Table 1.

Table 1. Varieties of Urban-Rural Left-Right Cleavage-Preference Structure.

	Strong Urban-Rural	Weak Urban-Rural
High Avg. Corr. with Left-Right	I High Bundling Bolivia, United States	II Discordant Composition Bundling with Country-Specific Nuance France, South Korea
Low Avg. Corr. with Left-Right	III Discordant Composition Bundling with Country-Specific Nuance Japan, Nepal	IV Discordant Composition Low Bundling Norway, Philippines

The first category, Type I (High Bundling), includes cases such as the United States and Bolivia, where urban and rural voters exhibit strongly opposing ideological and party preferences, with little spatial overlap. In these cases, political parties have successfully mobilized spatially concentrated voter blocs, reinforcing stable urban-left and rural-right patterns (or, in Bolivia's case as elsewhere, rural-left and urban-right). These divisions emerge because urban and rural areas cluster voters with distinct economic structures, demographic compositions, and political preferences, enabling parties to make targeted geographic appeals. At the opposite end of the spectrum, Type IV (Discordant Composition with Low

Bundling) describes countries where urban and rural differences in voting behavior are either weak or ideologically inconsistent. Norway and the Philippines exemplify this category. In Norway, economic development and industrialization occurred in a decentralized manner, with rural areas hosting key industrial hubs (e.g., aluminum production and hydroelectric power), thereby reducing ideological polarization between urban and rural voters. Similarly, in the Philippines, political competition is shaped more by regional and personalistic factors than by urban–rural ideological conflict, limiting the emergence of a strong spatial cleavage.

These varieties of urban–rural left–right cleavages map onto the conceptual illustrations of bundling versus discordant composition in Figure 4. In many Western knowledge economies, political coalitions most closely resemble Type I-A. As Iversen and Soskice (2019) describe, the urbanizing young and college-educated form a "new middle class", whose economic and cultural preferences increasingly align with the urban working class. This coalition contrasts with the "old middle class", which once thrived in Fordist economies but has since been, or at least feels, economically and politically marginalized. Often ethnically homogeneous, these rural voters perceive redistributive policies and labor market competition as favoring urban, immigrant, or minority groups, fueling a rural-right backlash. In contrast, countries in Latin America and elsewhere, where agrarian laborers and indigenous communities dominate rural areas, often exhibit the opposite pattern (rural-left, urban-right), aligning more closely with Type I-C.

The two off-diagonal categories (Types II and III) are more complex, representing cases of discordant composition with localized nuances. Type II, exemplified by France and South Korea, describes cases where urban and rural areas display ideological polarization, but partisan competition is not exclusively structured along urban–rural lines. In France, for example, a strong urban-left and rural-right divide exists, yet considerable regionalism shaping electoral dynamics precludes a simple cross-national density-based cleavage. This is also the case in South Korea, where, despite urban areas on average leaning left and rural areas leaning right, electoral behavior is strongly influenced by regionalism, industrialization patterns, and the legacy of the developmental state. Type III, observed in Japan and Nepal, describes cases where urban and rural constituencies remain distinct but do not align clearly along a left–right ideological spectrum. In Japan, urban and rural areas favor different parties, but both are predominantly conservative, reflecting institutional legacies and the enduring dominance of nationalist and pro-business parties.[10] In Nepal, rural voters overwhelmingly support leftist and communist parties, while urban areas tend to favor left or centrist alternatives, a pattern shaped by agrarian political mobilization and the unique trajectory of Nepalese democratization.

[10]This is not simply the consequence of a predominant Liberal Democratic Party. Of the five largest parties currently sitting in the Japanese House of Representatives, only one–the Constitutional Democratic Party of Japan–might be characterized as left of center.

7 Mechanisms

Why do urban–rural cleavages manifest so strongly in some contexts yet remain weak or reversed in others? Building on the theoretical framework of "discordant composition," I argue that four central features–economic structure, sociocultural organization, institutional legacies, and the nature of modernization–shape whether urban and rural places cluster politically salient characteristics that parties can then mobilize. I illustrate these pathways, focusing primarily on cases in Asia, where patterns of urban–rural polarization diverge dramatically (see Figures 4 and 5). The same logic applies elsewhere, however, including Latin America (where rural-left patterns predominate) and Western Europe (where urban-left patterns are more common).

7.1 Economic Activity

The structure of agricultural production has historically figured prominently as a driver of political conflict (Kautsky, 1899). Linz (1976), for example, identifies distinct agricultural class strata across Europe ranging from large landowners to farm day laborers and demonstrates how, in 1950s Italy, conservative parties received the bulk of support from the former group, while Socialist-Communist parties gained strength in areas dominated by agrarian workers. In the cleavage-theory literature, large landowners favoring authoritarian or fascist institutions are often placed in contrast with an agrarian labor class oriented toward social modernization (Moore, 1966; Lipset and Rokkan, 1967; Ziblatt, 2008). Although these studies typically focus on periods of structural economic change, such as the transition from agrarian to industrial economies, recent research suggests that similar patterns persist in contemporary electoral contests. For instance, rural regions in Spain characterized by historical landholding inequality or high rates of agrarian labor employment appear more likely to support left-leaning parties (Domènech and Sánchez-Cuenca, 2021), whereas the ascendance of agribusiness in the American Midwest seems to have propelled more conservative electoral shifts, given that these capital-intensive, agglomerative industries oppose government regulation and taxation (Dasgupta and Ramirez, 2024).

This variation in agrarian structure helps account for rural-left versus urban-right voting patterns in parts of Latin America. In the regions of United States and Western Europe where the urban-rural cleavage is most pronounced, agricultural commodity production often centers on crops such as wheat, corn, soybeans, and livestock, sectors that favored adoption of large-scale mechanization. As a result, these rural areas typically feature low population density and employ relatively few agricultural laborers, with technology displacing the need for landless workers who would otherwise be more likely to vote for parties espousing wealth redistribution or land reforms. By contrast, in countries where agriculture continues to employ substantial portions of the workforce,[11] rural

[11]Such conditions may arise from various factors, including land-tenure regimes or specific forms of crop production.

regions comprise significant shares of voters whose preferences align with left-leaning policy agendas. Bolivia, Peru, and Ecuador exemplify this phenomenon, each exhibiting strong rural-left and urban-right voting patterns while also having some of the highest proportions of employment in agriculture in Latin America (29%, 28%, and 32%, respectively). But the theory of discordant composition as described requires *bundles* of features to be concentrated in urban and rural places for there to be any such cleavage. It is also the case that all three countries contain substantial indigenous populations, with histories of rural agrarian reform movements seeking native and small-holder land protections.

Previous scholarship also emphasizes how the industrial revolution contributed to the formation of urban-rural electoral cleavages. Industrial activities largely clustered within urban centers to minimize the costs of importing resources and exporting finished products (Krugman, 1991). As Rodden (2019) argues, parties that emerged to champion urban industrial workers subsequently appealed to broader urban constituencies even before industrial employment waned. In the United States and Western Europe, areas historically connected to transportation networks (e.g., railroads) that facilitated industrial clustering often remain bastions of the modern center-left, whereas regions specializing in agriculture or raw resource extraction (particularly coal) have become affiliated with "left-behind" cultural identities and tend to support conservative alternatives (Abramson and Esposito 2019). Because industrialization typically benefits from agglomeration economies, this economic transformation often correlates strongly with urbanization. However, when industrial activity is spatially dispersed, as in Norway, the result is less spatially distinctive cleavage formation, or the cleavages fail to align with left-right preference structure. In Norway, there is relatively less of a stable urban-rural left-right electoral cleavage (see Figure 4) because industrial production—such as Europe's largest aluminum plant in rural Sunndalsøra—is often located in more remote areas close to agrarian communities.

7.2 Sociocultural Structure

Elections in Taiwan, another advanced industrial democracy characterized by little urban-rural polarization, are starkly divided between the indigenous and Taiwanese Mandarin-speaking eastern half of the island, which supports the Kuomintang, and the Taiwanese-speaking western half the island, which largely favors the Taiwanese nationalist Democratic Progressive Party. Here, ethnolinguistic and political identities, coupled with legacies of forced assimilation, *do* permit political parties to rely on spatially clustered appeals–but these appeals hinge more on east–west differences than on urban–rural divides. Comparatively, countries with pronounced urban-rural divides tend to be characterized either by relatively homogeneous national ethnic and linguistic profiles—such as Japan or South Korea—or by distinct spatial segregation that enables parties to mobilize specific territorial bases, as in South Africa or Ghana (Nathan 2019). Immigration to Western democracies has similarly magnified sociocultural distinctions between urban and rural populations, fueling the salience of

populist or conservative backlash in rural localities. In India, which displays only marginally differentiated rural-left and urban-right voting patterns, the right-wing Bharatiya Janata Party (BJP) has broadened its support base to encompass disparate economic classes and castes.[12] While the BJP originated as an upper-class, urban-focused party, it embarked upon a process of "ruralization and proletarianization" that enhanced its appeal among farmers and the rural poor (Maiorano 2019). The fact that Hindus (comprising 79.8% of India's population according to the 2011 census) constitute the majority in much of India further reduces the likelihood that strong cross-spatial class or religious cleavages will overshadow pan-Hindu political mobilization.

7.3 Institutional Legacies

In Japan and Nepal, one observes strong urban-rural cleavages that do not align with standard left-right preferences, and these incongruities appear at least partially attributable to institutional histories. Japan's contemporary party system, for instance, is traceable to imperial structures that facilitated urbanization and industrialization without an equivalent transformation of the underlying social order (Moore, 1966). As a result of ethnic homogeneity, high rates of urban migration, and early 20th-century modernization policies, rural areas coalesced around parties that primarily target rural constituencies, while urban areas gravitated toward different, yet also generally right-leaning, political representation. Consequently, although Japan is marked by a salient urban-rural cleavage, multiple conservative or nationalist parties strive to capture different urban segments. The long-standing dominance of conservatism in Japanese national politics has also nudged many parties rightward, including the Liberal Democratic Party (strong in rural regions) and the center-right populist Japan Innovation Party (based in Osaka and promising decentralization and local revitalization; see Jou, 2015). Likewise, in Indonesia, the legacy of Sukarno's *pancasila* doctrine (1945) and the prohibition of communist parties (1966) continues to circumscribe the space available for leftist ideologies.

Nepal presents the opposite ideological configuration: though the country features a strong urban-rural cleavage, the major parties almost uniformly adhere to social democratic ideals. The 1990 People's Movement, which ended the absolute monarchy, was driven by a unified communist opposition, a legacy that endures in contemporary electoral politics. Multiple factions of the Nepal Communist Party collectively garnered a majority of the vote in the most recent national elections, and all enjoy particular strength in the country's predominantly rural areas.[13] Even the smaller Marxist-Leninist organizations, such as the Workers' and Peasants' Party and the Rastriya Janamorcha, perform well in rural constituencies. Nepal's few urban centers do tend to vote for other, more centrist alternatives, but none that are explicitly right-leaning. For instance, half of Kathmandu's 10 electoral districts supported the centrist Rastriya Swatantra Party in

[12] With the notable exception of Muslims.

[13] Nepal's urbanization rate was only 20.2% in 2019.

2022, whereas the only openly right-wing party of note, the Rastriya Prajatantra Party (which advocates a return to the monarchy) commanded just 5.6% of votes nationally. Thus, in Nepal, divergent patterns of urban and rural support endure, but the overarching continuity of communist or social-democratic discourse preempts the emergence of strong left-right polarization.

South Korea presents a similarly nuanced case. Here, modernization and industrialization proceeded rapidly under a developmental state but have not produced a pronounced urban-rural cleavage, even though cities and rural areas lean in different ideological directions. State interventions, including a series of Five-Year Plans begun under Park Chung Hee, propelled peasants into urban manufacturing and, later, service and knowledge industries (Chun and Kim, 2022). Public investments in urban redevelopment also transformed sprawling *panjachon* ("villages" of wooden-board houses) into modern apartment complexes capable of accommodating this emergent urban middle class (Shin and Kim, 2016). By 1985, South Korea's urbanization rate had soared to 77.3%, more than double its level in 1955. Despite the pronounced urban tendency to favor the liberal Democratic Party and the rural bias toward the conservative People Power Party, the salience of regionalism, shaped by autocratic legacies and democratization movements, blurs a straightforward urban-rural divide. The centrist-liberal Democratic Party today is favored in the population centres: Seoul (bundling of young, middle-class, well educated) in the northwest and increasingly the blue-collar manufacturers in Busan in the southeast. The right-wing People Power Party predominates in the rural, mountainous eastern half of the country. These conditions alone *would* produce an urban-rural divide if not for the autocratic and revolutionary legacies, and in turn regionalism, that disposes *most* of western South Korea, urban and rural alike, away from conservatism.[14]

7.4 Late or Distinct Modernization

Another possible explanation for these patterns may center on how timing and pathways of development structure urban-rural cleavages. Huntington (1968) notes the way in which early modernizers experienced a decisive separation between the countryside and the city, the latter which "becomes the locus of new economic activities, new social class, new culture and education." Where urbanization occurred simultaneously with industrialization and technological revolution, city and countryside became as if different nations;[15] those who moved to the former tended to receive higher-quality educations, assume cosmopolitan social and cultural identities, and secularize. Even for lower-class workers in manufacturing, their modernization at a minimum entailed exposure to mass media, entry into the market economy, and various capacities for political participation all of which shaped the preferences of this emergent urban electorate (Lerner,

[14]The *Minjung* movement, for instance, that ultimately brought about democratization flourished in the southwest.

[15]In Huntington's (1968) precise words, "the city and the countryside become different nations, different ways of life."

1958; Deutsch, 1961). Their rural counterparts were left in "a life of local isolation, traditionalism, and political apathy," fundamental differences which democratic enfranchisement elevated to the status of integral party cleavage structure. For the early modernizers that are today synonymous with the Western post-industrial democracies, it is clear the way in which urbanization constructed an urban electorate comprising certain bundles of features and a rural electorate comprising bundles of the opposite. These bundles have remained largely sticky through subsequent structural transformations and exacerbated by self-sorting on the basis of these shared characteristics.

Where non-Western and "late" modernizers were guided towards industrialization and urbanization, alongside other processes that led to spatial clustering along types of wealth, economic activity, education, ethnic identities, and other features of a "modern" society, urban and rural areas comprised conflicting electorates and an urban-rural structural cleavage is observed in contemporary political behavior. Among late modernizers, in the absence of a developmental state favorably disposed to urbanization and industrialization, there is less certainty that urban and rural areas will respectively comprise distinct electorates to which parties may differentially appeal. This situation characterizes many African nations, which post-independence lacked strong state-led industrialization. This resulted in urbanization without corresponding industrial development and therefore weak bundling (cross-cutting "over-urbanization"); this does not necessarily preclude spatially concentrated appeals, as in segregated-urban Ghana (Nathan, 2019); but confounds the capacity for, and benefits accrued by, doing so.

Not only late but distinct modernization can preclude the emergence of an urban-rural cleavage, or one that produces an aligned cleavage-preference structure. Norway has been discussed as a case where industrialization is situated in the rural countryside in order to leverage natural hydroelectric resources. During the 19th century, the crown and nobility crowded out peasant proprietors along the southwestern coast of the country, which contains the little available fertile land in the country. Today, this broader region, where the core urban centres are also situated, comprises the electoral stronghold of the Conservative Party, an ideologically right-leaning party associated with the traditional ("old") upper class. Writing on Norwegian modernization during industrialization, Semmingsen (1954) confirms that the dissolution of estate society caused agricultural laborers to gradually emigrate out of farming regions and into rural factories, leaving behind farms that "were separated as individual farms and became independently owned". Thus explains support for the elite-oriented Conservative Party in those few arable coastal regions, urban and rural alike; the dominance of the agrarian-traditional Centre Party among family farmers; and the Labor Party excelling in the heavily industrialized regions: rural and urban communities focused on aluminum production, hydroelectric energy generation, and oil and gas processing, among others, as well as cities in the agriculturally-unsuitable north.

8 Conclusions

Nearly four decades ago, Dix (1989) reflected that "the parties of integration (or class-mass parties) so familiar in European politics during its era of economic development and mass politics have, by and large, not appeared in Latin America," while Epstein (1967) similarly contended that "large-membership working-class parties"–such as contemporary left-wing urban parties in their earliest phases– "are a product occurring only at certain stages of social development in certain nations." Much of the prior work on urban-rural cleavage structure relies exactly on these historical conditions, an assumption often left unstated, in arguing for why urban areas almost "inevitably" conflict with rural ones. Indeed, while various scholarship has demonstrated urban and rural left-right cleavage-preference structure in the United States and other Western democracies (Rodden, 2019; Armstrong et al., 2022; Taylor et al., 2024), little work has been done empirically testing such relationships elsewhere. Nor have there been constructed generalizable and probabilistic theories about why such cleavages may or may not characterize contemporary party systems across the world.

Returning to the central question originally posed, where and why do urban-rural political cleavages manifest? The limited scope of previous studies and the (a-)granularity of existing cross-national elections data have constrained the identification of more universal patterns. I begin by detailing the construction of the most granular, large-scale elections dataset to date, termed the Small-Area Global Elections (SAGE) archive. SAGE comprises geocoded, small-area (polling station, neighborhood, municipality), national-level (presidential, lower-house or unicameral legislative) election results for 106 countries: every democracy, ranging from India to Greenland; most hybrid regimes, such as Bangladesh, Hong Kong, or Zambia; and the previous democratic elections of several currently authoritarian states, like Venezuela and El Salvador. In total, the data represent nearly 10 billion votes cast across 8.2 million unique places, cleaned and standardized to a ready-to-use format.

Drawing on SAGE and the $k-$nearest-neighbor distances among 2.6 billion buildings as a novel measure of urbanicity-rurality–a measure which I validate in an original, nationally-representative survey–I find little evidence of global urban-rural cleavages. A single standard deviation change in urbanicity-rurality corresponds, on average, to a one-tenth absolute standard deviation shift in party vote-share across hundreds of legislative and presidential elections (Figure A3). These results imply, firstly, that there is not a universal geographic division in contemporary electoral democracies; and secondly that party systems have not globally converged on a select few "frozen" party cleavages, such as those identified by Lipset and Rokkan (1967) in post-War Western Europe.

This marginal result in aggregate masks significant variation across settings (Figure 4). However, I find little immediately discernible basis for segregating parties with significant urban-rural divides from those without. I confirm that many of the advanced, post-industrial democracies where much of the prior theory-building has focused exhibit strong urban-rural cleavages in the expected urban-

left rural-right form (Figure 5); I also discover several such countries where urban-rural cleavages are unexpectedly muted, and others where urban areas vote more right-wing than their rural counterparts. In most countries, urban and rural differences in political behavior are too marginal to be of note.

To explain these patterns, I argue that urban-rural cleavages emerge most distinctly when politically relevant socioeconomic and institutional features (education, income, ethnicity, economic activity) are spatially clustered within, rather than cleaving across, urban and rural areas. This clustering enables parties to construct effective programmatic appeals to geographically-defined voter coalitions. The framework explains why advanced industrial democracies display strong urban-left and rural-right patterns (given spatial correspondence among politically salient factors), why several countries in Latin America evince rural-left urban-right voting (through indigenous organization, agrarian labor, and land tenure), and why others such as India (where Hindu nationalism cuts across urban and rural) show minimal urban-rural division. These last cases tend to evince patterns of what I call "discordant composition", which can in part be traced back to distinct paths of modernization: where developmental states guided simultaneous urbanization and industrialization among late modernizers, creating spatially concentrated bundles of features, urban-rural cleavages tend to emerge strongly. Where these processes occurred differently or independently, geographic divisions in voting behavior are less pronounced. In this way, the present paper works on extending the economic predictions of Gerschenkron (1962) to the implications of late industrialization for party systems.

I argue that discordant composition can also be attributed to institutional legacies, distinct sociocultural structure, and economic activity where organized orthogonal to urban and rural. In South Korea, autocratic and revolutionary legacies induce an east-west political divide that overcomes structural conditions otherwise encouraging urban-rural divisions. The same outcome is found in Taiwan, but for reasons of ethnolinguistic and political identity. Discordant composition can be overcome in ethnically fractionalized Ghana by neighborhood segregation (Nathan, 2016). Building on the work of Tarrow (1971), successful agrarian revolution likely contributed to the strong urban-left rural-right cleavages found in France today, while failed agrarian revolution precludes such divisions in Italy. Instances of "pluriactivity," that is, the diversification of economic activities in rural regions beyond exclusively agriculture (Marsden, 1990), may explain instances why urban and rural divisions fail to emerge in many countries not characterized by an urban-rural dual economy. These and other probabilistic expectations are encapsulated within the framework proposed here.

While this work demonstrates the varied manifestation of urban-rural cleavages globally, more research is needed to understand the precise mechanisms of voter bloc formation and spatial clustering. Several key areas warrant further investigation. First, which factors influence which combinations of voter characteristics tend to concentrate geographically, and how do parties adapt their programmatic appeals in response over time? The theory advanced here sug-

gests that parties will emphasize different dimensions of conflict depending on which, and how cleanly, voter features map onto urban and rural space, but additional empirical work is needed to test these dynamics causally. Relatedly, more research is needed on how different modernization pathways shape the emergence (or absence) of geographic cleavages. Are there more systematic differences in spatial bundling between countries that modernized through state direction versus market-led approaches? If so, why? While the developmental state experiences of countries like South Korea demonstrate how coordinated industrialization and urbanization can create distinct urban-rural voter blocs, we know less about how alternative development trajectories affect party systems and cleavage structures. Further scholarship is also needed examining how commuting and remote work, which allow the non-economic preference sorting of voters who would otherwise live in urban agglomerations, impact urban-rural divisions and the bases of party support particularly in knowledge economies. Various other questions are also left to future work.

Methodologically, SAGE enables detailed spatial analyses across every democracy and many semi-autocratic regimes, rather than only cases with more accessible data environments. It is also the hope that these data will facilitate cross-national and generalizable hypothesis testing, which is presently confounded by heterogeneity in data quality and accessibility. In the context of the substantive application examined here, SAGE can facilitate more rigorous study not only of the urban-rural cleavage but of electoral and political geographic questions more broadly. To date, potential scholarship in these fields has been limited by the necessity of granular and geocoded election results, conditions previously met almost exclusively in the United States and select Western European countries.

9 References

Abou-Chadi, Tarik, and Simon Hix. "Brahmin Left Versus Merchant Right? Education, Class, Multiparty Competition, and Redistribution in Western Europe." *The British Journal of Sociology* 72, no. 1 (2021): 79-92.

Adcock, Robert, and David Collier. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95, no. 3 (2001): 529-546.

Armstrong, David A., Jack Lucas, and Zack Taylor. "The Urban-Rural Divide in Canadian Federal Elections, 1896–2019." *Canadian Journal of Political Science/Revue Canadienne De Science Politique* 55, no. 1 (2022): 84-106.

Arora, Abhishek, and Melissa Dell. "LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models." arXiv preprint arXiv:2309.00789 (2023).

Baltz, Samuel, Alexander Agadjanian, Declan Chin, John Curiel, Kevin DeLuca, James Dunham, Jennifer Miranda et al. "American Election Results at the Precinct Level." *Scientific Data* 9, no. 1 (2022): 651.

de Benedictis-Kessner, Justin, Diana Da In Lee, Yamil R. Velez, and Christopher Warshaw. "American Local Government Elections Database." *Scientific Data* 10, no. 1 (2023): 912.

Chun, Yongwan, and Kamyoung Kim. "Temporal Changes in the Urban System in South Korea." Frontiers in Sustainable Cities 4 (2022): 1013465.

Cookman, Colin. Assessing Afghanistan's 2019 Presidential Election. *United States Institute of Peace*, 2020.

Cramer, Katherine J. "Putting Inequality in Its Place: Rural Consciousness and the Power of Perspective." *American Political Science Review* 106, no. 3 (2012): 517-532.

Cramer, Katherine J. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker*. University of Chicago Press, 2016.

Dahl, Robert A. *Polyarchy: Participation and Opposition*. Yale University Press, 1971.

Dasgupta, Aditya, and Elena Ramirez. "Explaining Rural Conservatism: Political Consequences of Technological Change in the Great Plains." *American Political Science Review* (2020): 1-23.

Deutsch, Karl W. "Social Mobilization and Political Development." American Political Science Review 55, no. 3 (1961): 493-514.

Dinas, Elias, and Kostas Gemenis. "Measuring Parties' Ideological Positions With Manifesto Data: A Critical Evaluation of the Competing Methods." *Party Politics* 16, no. 4 (2010): 427-450.

Dix, Robert H. "Cleavage Structures and Party Systems in Latin America." *Comparative Politics* 22, no. 1 (1989): 23-37.

Domènech, Jordi, and Ignacio Sánchez-Cuenca. "The Long Shadow of Agrarian Conflict: Agrarian Inequality and Voting in Spain." *British Journal of Political Science* 52, no. 4 (2022): 1668-1688.

Düpont, Nils, Berker Kavasoglu, Anna Lührmann, and Ora John Reuter. "Party Organizations Around The Globe. Introducing the Varieties of Party Identity and Organization Dataset (V-Party)." V-Dem Working Paper 124 (2021).

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder Jr. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from over 40,000 Close Races." *American Journal of Political Science* 59, no. 1 (2015): 259-274.

Ebeid, Michael, and Jonathan Rodden. "Economic Geography and Economic Voting: Evidence from The US States." *British Journal of Political Science* 36, no. 3 (2006): 527-547.

Epstein, Leon D. *Political Parties in Western Democracies*. 1967.

Esposito, Elena, and Scott F. Abramson. "The European Coal Curse." *Journal of Economic Growth* 26 (2021): 77-112.

Ford, Robert, and William Jennings. "The Changing Cleavage Politics of Western Europe." *Annual Review of Political Science* 23 (2020): 295-314.

Fotheringham, A. Stewart, and David WS Wong. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." Environment and Planning A 23, no. 7 (1991): 1025-1044.

Fowler, Anthony. "Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia." *Quarterly Journal of Political Science* 8, no. 2 (2013): 159-182.

Gallego, Aina, Franz Buscha, Patrick Sturgis, and Daniel Oberski. "Places and Preferences: A Longitudinal Analysis of Self-Selection and Contextual Effects." *British Journal of Political Science* 46, no. 3 (2016): 529-550.

Gerring, John, Maxwell Palmer, Jan Teorell, and Dominic Zarecki. "Demography and Democracy: A Global, District-Level Analysis of Electoral Contestation." *American Political Science Review* 109, no. 3 (2015): 574-591.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. "Is There a Secret Ballot? Ballot Secrecy Perceptions and Their Implications for Voting Behaviour." *British Journal of Political Science* 43, no. 1 (2013): 77-102.

Gerschenkron, Alexander. *Economic Backwardness in Historical Perspective*. 1962.

Gidron, Noam, and Peter A. Hall. "The Politics of Social Status: Economic and Cultural Roots of the Populist Right." The British Journal of Sociology 68 (2017): S57-S84.

Gimpel, James G., Nathan Lovin, Bryant Moy, and Andrew Reeves. "The Urban–Rural Gulf in American Political Behavior." *Political Behavior* 42 (2020): 1343-1368.

Herrmann, Michael, and Holger Döring. "Party Positions from Wikipedia Classifications of Party Ideology." Political Analysis 31, no. 1 (2023): 22-41.

Hooghe, Liesbet, Gary Marks, and Carole J. Wilson. "Does Left/Right Structure Party Positions on European Integration?." *Comparative Political Studies* 35, no. 8 (2002): 965-989.

Horowitz, Donald L. *Ethnic Groups in Conflict*. 1985.

Huijsmans, Twan, and Jonathan Rodden. "The Great Global Divider? A Com-

parison of Urban-Rural Partisan Polarization in Western Democracies." *Comparative Political Studies* (2024): 00104140241237458.

Huntington, Samuel. *Political Order in Changing Societies*. 1968.

Iversen, Torben, and David Soskice. *Democracy and Prosperity: Reinventing Capitalism Through a Turbulent Century*. Princeton University Press, 2020.

Jacobs, Nicholas F., and B. Kal Munis. "Staying in Place: Federalism and the Political Economy of Place Attachment." *Publius: The Journal of Federalism* 50, no. 4 (2020): 544-565.

Jolly, Seth, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. "Chapel Hill Expert Survey Trend File, 1999–2019." Electoral Studies 75 (2022): 102420.

Jou, Willy. "A Regional Party in a Centralized Country: The Case of One Osaka in Japan." Regional & Federal Studies 25, no. 2 (2015): 145-163.

Kaufman, Aaron R., Gary King, and Mayya Komisarchik. "How to Measure Legislative District Compactness If You Only Know It When You See It." American Journal of Political Science 65, no. 3 (2021): 533-550.

Kautsky, Karl. *The Agrarian Question*. 1899.

Kenny, Michael, and Davide Luca. "The Urban-Rural Polarisation of Political Disenchantment: An Investigation of Social and Political Attitudes in 30 European Countries." *Cambridge Journal of Regions, Economy and Society* 14, no. 3 (2021): 565-582.

Kollman, Ken, Allen Hicken, Daniele Caramani, David Backer, and David Lublin. *Constituency-Level Elections Archive.* Ann Arbor, MI: Center for Political Studies, University of Michigan [producer and distributor], 2024. Web. 19 April.

Krugman, Paul. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99, no. 3 (1991): 483-499.

Lerner, Daniel. *The Passing of Traditional Society: Modernizing the Middle East.* 1958.

Linz, Juan J. "Patterns of Land Tenure, Division of Labor, and Voting Behavior in

Europe." Comparative Politics 8, no. 3 (1976): 365-430.

Lipset, Seymour Martin. *Agrarian Socialism: The Cooperative Commonwealth Federation in Saskatchewan: A Study in Political Sociology*. Vol. 64. Univ of California Press, 1971.

Lipset, Seymour Martin, and Stein Rokkan. *Cleavage Structures, Party Systems, and Voter Alignments: An Introduction*. Vol. 2. New York: Free Press, 1967.

Long, Julianna C., Paul L. Delamater, and George M. Holmes. "Which Definition of Rurality Should I Use?: The Relative Performance of 8 Federal Rural Definitions in Identifying Rural-Urban Disparities." *Medical Care* 59 (2021): S413-S419.

Magar, Eric. "Gubernatorial Coattails in Mexican Congressional Elections." *The Journal of Politics* 74, no. 2 (2012): 383-399.

Maiorano, Diego. "The 2019 Indian Elections and the Ruralization of the BJP." Studies in Indian Politics 7, no. 2 (2019): 176-190.

Marsden, Terry. "Towards the Political Economy of Pluriactivity." *Journal of Rural Studies* 6, no. 4 (1990): 375-382.

Martínez-Gallardo, Cecilia, Nicolás de la Cerda, Jonathan Hartlyn, Liesbet Hooghe, Gary Marks, and Ryan Bakker. "Revisiting Party System Structuration in Latin America and Europe: Economic and Socio-Cultural Dimensions." Party Politics 29, no. 4 (2023): 780-792.

Maxwell, Rahsaan. "Cosmopolitan Immigration Attitudes in Large European Cities: Contextual or Compositional Effects?." *American Political Science Review* 113, no. 2 (2019): 456-474.

Mettler, Suzanne, and Trevor Brown. "The Growing Rural-Urban Political Divide and Democratic Vulnerability." *The Annals of the American Academy of Political and Social Science* 699, no. 1 (2022): 130-142.

Moore, Barrington. *Social Origins of Dictatorship and Democracy*. 1966.

Nathan, Noah L. "Local Ethnic Geography, Expectations of Favoritism, and Voting in Urban Ghana." *Comparative Political Studies* 49, no. 14 (2016): 1896-1929.

Nathan, Noah L. *Electoral Politics and Africa's Urban Transition: Class and Ethnicity in Ghana*. Cambridge University Press, 2019.

Nathan, Charles. "The Urban/Rural Divide in Athenian Political Thought." *American Political Science Review* 116, no. 4 (2022): 1490-1502.

Nemerever, Zoe, and Melissa Rogers. "Measuring the Rural Continuum in Political Science." Political Analysis 29, no. 3 (2021): 267-286.

Norris, Pippa. "Measuring Populism Worldwide." *Party Politics* 26, no. 6 (2020): 697-717.

Pérez, Virgilio, Cristina Aybar, and Jose M. Pavía. "Spanish Electoral Archive." *Scientific Data* 8, no. 1 (2021): 193.

Przeworski, Adam. "Defending Democracy." 2024.

Rachman, Gideon. "Urban-Rural Splits Have Become The Great Global Divider." Financial Times 30, no. 7 (2018): 18.

Rodden, Jonathan. "Federalism and Inter-Regional Redistribution." In *The Political Economy of Inter-Regional Fiscal Flows*. Edward Elgar Publishing, 2010.

Rodden, Jonathan A. *Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide*. Basic Books, 2019.

Rodríguez-Pose, Andrés. "The Revenge of the Places That Don't Matter (and What To Do About It)." *Cambridge Journal of Regions, Economy, and Society* 11, no. 1 (2018): 189-209.

Runfola, Daniel, Austin Anderson, Heather Baier, Matt Crittenden, Elizabeth Dowker, Sydney Fuhrig, Seth Goodman et al. "geoBoundaries: A Global Database of Political Administrative Boundaries." *PloS One* 15, no. 4 (2020): e0231866.

Scott, Alister, Alana Gilbert, and Ayele Gelan. *The Urban-Rural Divide: Myth or Reality?*. Aberdeen: Macaulay Institute, 2007.

Semmingsen, Ingrid. "The Dissolution of Estaste Society in Norway." *Scandinavian Economic History Review* 2, no. 2 (1954): 166-203.

Shin, Hyun Bang, and Soo-Hyun Kim. "The Developmental State, Specula-

tive Urbanisation and the Politics of Displacement in Gentrifying Seoul." Urban Studies 53, no. 3 (2016): 540-559.

Sobolewska, Maria, and Robert Ford. *Brexitland*. Cambridge University Press, 2020.

Stegmueller, Daniel. "Religion and Redistributive Voting in Western Europe." *The Journal of Politics* 75, no. 4 (2013): 1064-1076.

Tarrow, Sidney. "The Urban-Rural Cleavage in Political Involvement: The Case of France." *American Political Science Review* 65, no. 2 (1971): 341-357.

Taylor, Zack, Jack Lucas, David A. Armstrong, and Ryan Bakker. "The Development of the Urban-Rural Cleavage in Anglo-American Democracies." *Comparative Political Studies* (2024): 00104140231194060.

Trujillo, Kristin Lunz. "Feeling Out of Place: Who Are the Non-Rural Rural Identifiers, and Are They Unique Politically?" *Political Behavior* (2024): 1-25.

Wolfinger, Raymond E. "The Development and Persistence of Ethnic Voting." American Political Science Review 59, no. 4 (1965): 896-908.

Wood, Terence. "Papua New Guinea Election Results: Trends and Patterns 1972-2012." *Development Policy Centre Discussion Paper* 55 (2017).

Ziblatt, Daniel. "Does Landholding Inequality Block Democratization?: A Test of The "Bread and Democracy" Thesis and The Case of Prussia." World Politics 60, no. 4 (2008): 610-641.

A Appendix

Contents

A.1 Election Data Summary by Country

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Afghanistan	2018, 2019	All	Legislative	Polling Station	20,000	Colin Cookman	.999
Albania	2017, 2021	2021	Legislative	Polling Station	5,200		1
Argentina	2011 to 2023	2023	Legislative, Presidential	Polling Station	100,000		.98
Armenia	2012, 2013, 2018, 2021	All	Legislative, Presidential	Polling Station	2,000		.999
Australia	2004 to 2022	All	Legislative	Polling Station	8,000		1
Austria	1999 to 2024	$\geq$ 2013	Legislative	Municipality (gemeinde)	2,000		1
Bangladesh	2018	All	Legislative	Polling Station	40,000		.991
Belgium	2014, 2019, 2024	All	Legislative	Municipality (gemeente)	590		1
Bhutan	2018	All	Legislative	Polling Station	865		.999
Bolivia	2019, 2020	All	Legislative, Presidential	Polling Station	68,000 (geocode level: 6,600)		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Bosnia and Herzegovina	2018, 2022	All	Legislative	Polling Station	3,000		.999
Botswana	2014, 2019	All	Legislative	Parliamentary Constituency	57		1
Brazil	2014, 2018, 2022	All	Legislative, Presidential	Polling Station	93,000		.989
Bulgaria	2013 to 2023	2022, 2023	Legislative, Presidential	Polling Station	12,000		.999
Cabo Verde	2021	All	Presidential	Polling Station	1,000		.966
Canada	1997 to 2021	$\geq$ 2000	Legislative, Presidential	Polling Station	70,000		.990
Colombia	2018	All	Legislative	Polling Station	102,000 (ballot boxes; 11,000 unique places)		.993
Costa Rica	2018, 2022	All	Legislative, Presidential	Polling Station	2,101		.999

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Chile	2013, 2017, 2021	All	Legislative, Presidential	Polling Station	90,000 (geocode level: 7,000)		1
Croatia	2011 to 2024	All	Legislative, Presidential	Polling Station	6,100		1
Cyprus	2001 to 2023	All	Legislative, Presidential	Polling Station	1,000 (geocode leve: 400)		1
Czechia	2002 to 2021	2017, 2021	Legislative	Election Precinct (okrsek)	14,800		.989
Denmark	2011 to 2022	All	Legislative	Polling Station	1,300		1
Dominica	2019, 2022	All	Legislative	Polling Station	230		.986
Dominican Republic	2000 to 2024	!(2000, 2010, 2016)	Legislative, Presidential	Polling Station	12,000		.995
Ecuador	2002 to 2023	All	Legislative, Presidential	Parish	1,220		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
El Salvador	2014, 2018	All	Legislative, Presidential	Polling Station	1,600		1
Estonia	2015, 2019	2019	Legislative	Polling Station	500		1
Fiji	2022	All	Legislative	Polling Station	991		1
Finland	2011 to 2024	$\geq$ 2015	Legislative, Presidential	Voting Districts (2019), Municipality	1,900 (2019); 310 ($\geq$ 2015)		.996
France	2002 to 2024	All	Legislative, Presidential	Polling Station	70,000 ($\geq$ 2017); 35,000 ($<$ 2017)		.95 ($\geq$ 2022); .988 ($\leq$ 2017)
Georgia	2012 to 2024	All	Legislative	Polling Station	2,000		.985
Germany	1983 to 2021	$\geq$ 1998	Legislative	Polling Station	80,000 (geocode level: 11,000)		.989
Ghana	2012, 2016, 2020	All	Legislative, Presidential	Parliamentary Constituency	275		1
Greece	2012 to 2023	All	Legislative	Polling Station	20,000		.987

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Greenland	2002 to 2022	All	Legislative	Settlements	72		1
Guatemala	2023	All	Legislative	Polling Station	24,000 (geocode level: 3,500)		.928
Guyana	2015	All	Legislative	Polling Station	2,000		.999
Honduras	2021	All	Legislative, Presidential	Polling Station	18,300 (geocode level: 5,700)		1
Hong Kong	2016, 2021	All	Legislative	Polling Station	(2021: 650, 2016: 100)		1
Hungary	2014, 2018, 2022	All	Legislative	Polling Station	10,000		.999
Iceland	1959 to 2021	All	Legislative, Presidential	Parliamentary Constituency	($8 < 2003$, $6 \geq 2003$)		1
Indonesia	2019	All	Legislative, Presidential	Polling Station	800,000 (geocode level: 80,000)		.997
India	2019	All	Legislative	Polling Station	867,000		.944
Iran	2017	All	Presidential	City	380		1
Ireland	2002 to 2020	2016, 2020	Legislative	Parliamentary Constituency	40		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Israel	2006 to 2022	2020, 2021	Legislative	Polling Station	11,000		1
Italy	1953 to 2022	$\geq$ 2002	Legislative	Municipality (commune)	8,000		.96
Jamaica	2007 to 2020	All	Legislative	Polling Station	6,500		.965
Japan	2009 to 2024	All	Legislative	Municipality	2,000		1
Kenya	2022	All	Presidential	Polling Station	46,000		.996
Kosovo	2017, 2019, 2021	All	Legislative	Polling Station	2,500		1
Latvia	2014, 2018, 2022	All	Legislative	Polling Station	2,000		1
Lebanon	2018, 2022	All	Legislative	Polling Station	6,800		.998
Lesotho	2017, 2022	All	Legislative	Parliamentary Constituency	80 (geocode level: 10)		1
Lithuania	2016 to 2024	All	Legislative, Presidential	Precinct (apylinkės)	2,000		1
Luxembourg	2023	All	Legislative	Municipality (commune)	100		1
Madagascar	2018, 2023	All	Presidential	Polling Station	25,000		.997

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Malaysia	2008 to 2022	All	Legislative	Parliamentary Constituency	222		1
Malawi	2019	All	Legislative, Presidential	Polling Station	11,000 (geocode level: 5,000)		.997
Mexico	1991 to 2024	2006, 2009, 2015, 2018, 2024	Legislative, Presidential	Polling Station	150,000	Magar, 2019 ($\leq$ 2018)	.999
Moldova	2014 to 2024	2020, 2021, 2024	Legislative, Presidential	Polling Station	2,000		.999
Mongolia	2021	All	Presidential	Polling Station	1,700 (geocode level: 350)		.956
Montenegro	2023	All	Legislative	Polling Station	1,000		1
Myanmar	2010, 2015	All	Legislative	Parliamentary Constituency	320		.980
Namibia	2014, 2019	All	Legislative, Presidential	Parliamentary Constituency	120		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Nepal	2017, 2021	All	Legislative	Parliamentary Constituency	165		1
Netherlands	2010 to 2021	All	Legislative	Polling Station	400 (2017), 10,000 (others)		.999
New Zealand	1999 to 2023	All	Legislative	Polling Station	5,000		.999
Nigeria	2019	All	Legislative	Parliamentary Constituency	350		1
North Macedonia	2016, 2024	All	Legislative, Presidential	Polling Station	3,500		.999
Norway	2009 to 2021	2021	Legislative	Municipality ($\leq$ 2013), Electoral District ($\geq$ 2017)	226 ($\leq$ 2013), 1,250 ($\geq$ 2017)		.986
Pakistan	2018	All	Legislative	Polling Station	72,000 (geocode level: 250)	Colin Cookman	1
Panama	2004, 2009	All	Legislative	District (corregimiento)	620		.950
Papua New Guinea	1987 to 2017	All	Legislative	Electorate	100	Wood (2019)	1

Continued on next page

(Continued)

Country	Years		Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Paraguay	2003	to	All	Legislative, Presidential	Polling Station	17,000		.999
	2018							
Peru	2006	to	2021	Legislative, Presidential	Polling Station	150,000 ($\leq$ 2016), 83,000 (2021)		1
	2021							
Philippines	2022		All	Legislative, Presidential	Polling Station	104,000 (geocode level: 50,000)		1
Poland	1990	to	All	Legislative, Presidential	Polling Station	27,000		.999
	2023							
Portugal	1976	to	$\geq$ 2009	Legislative, Presidential	Parish	4,000		.997
	2024							
Romania	2014	to	$\neq$ 2016	Legislative, Presidential	Polling Station	18,500		1
	2024							
Russia	2000	to	$\geq$ 2012	Legislative, Presidential	Polling Station	95,000		1 (2012), .893 (2016, 2018)
	2024							

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Senegal	2024	All	Legislative	Polling Station	10,000 (incomplete data)		.996
Serbia	2000 to 2022	2017	Legislative, Presidential	Polling Station	8,000		.987
Singapore	2020	All	Legislative	Constituency	31		1
Slovakia	2016 to 2024	All	Legislative, Presidential	Polling Station	6,000 (geocode level: 1,500)		1
Slovenia	2012 to 2022	All	Legislative, Presidential	Polling Station	3,700		.995
Solomon Islands	2006 to 2019	All	Legislative	Parliamentary Constituency	50	Wood (2019)	1
South Africa	2004 to 2024	All	Legislative	Voting Districts	20,000		1
South Korea	2002 to 2024	$\geq$ 2007	Legislative, Presidential	Polling Station	13,000 to 34,000		.99
Spain	1982 to 2023	$\geq$ 2004	Legislative	Polling Station	36,000		.998
Sri Lanka	2020	All	Legislative	Polling Division	150		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
Sweden	2006 to 2022	All	Legislative	Electoral District	6,100		1
Switzerland	1971 to 2023	$\geq$ 2011	Legislative	Municipality	2,400		1
Taiwan	1996 to 2024	$\geq$ 2020	Legislative, Presidential	Polling Station	15,000		.989
Thailand	2023	All	Legislative	Parliamentary Constituency	400		1
Trinidad and Tobago	2015, 2020	All	Legislative	Parliamentary Constituency	40		1
Tunisia	2014	All	Presidential	Polling Station	19,000		.999
Turkey	2011 to 2023	All	Legislative, Presidential	Polling Station	190,000 (geocode level: 50,000)		1
Ukraine	2019	All	Legislative, Presidential	Polling Station	30,000		.996
United Kingdom	2005 to 2024	All	Legislative	Parliamentary Constituency	650		1

Continued on next page

(Continued)

Country	Years	Polygon Years	Election Types	Smallest Physical Unit (Data)	Units per Year (approximate average)	Additional Source	Geographic Coverage (non-missing years)
United States of America	2016, 2020	All	Presidential	Precinct	200,000	VEST	.999
Uruguay	2014, 2019, 2024	All	Legislative	Polling Station	7,200		.991
Vanuatu	2002 to 2020	All	Legislative	Parliamentary Constituency	17		1
Venezuela	2013	All	Presidential	Polling Station	40,000		.991
Zambia	2021	All	Presidential	Parliamentary Constituency	150		1

A.2 Constructing the Small-Area Global Elections (SAGE) archive

There is substantial variability in how I collected election results on a country-by-country basis. In some cases, I simply downloaded election results hosted by the respective election commission, such as for Taiwan or Peru; for others, like Turkey or the Philippines, the government only displays results for a single polling station after navigating through a series of online dropdowns, which I automated using bespoke web scrapers. Some cases, like Venezuela, required online archiving tools for viewing now-deleted results, while scraping elections in Russia—where foreign access to government sites is currently prohibited—entailed scraping data from behind a series of web proxies. In especially data-poor environments, such as Jamaica, Lebanon, or Sri Lanka, results tended to be provided in PDFs that required the use of tabular OCR (optical character recognition) methods. In these cases, I used proprietary tools, sometimes testing several for the same country, to obtain results of the highest possible quality.[16] Although nearly all the data represent my own original effort, there are several countries for which I relied on the work of others. This includes Magar (2019), Pérez et al. (2021), and Baltz et al. (2022), which collect polling station-level election results for Mexico, Spain, and the United States, respectively, and Wood (2019), which compiles election results for Papua New Guinea and the Solomon Islands.

Matching electoral units to spatial boundaries similarly entailed substantial variability across settings. Where I obtain election results "higher" than the polling station and for some precise, formally defined unit, such as the Portuguese parish or the Swiss commune, I find, download, and merge in administrative shapefiles matching that election year. For most countries with polling station-level results, I locate the exact voting addresses released for public use. I geocode these addresses using the Google Geocoding API or the ESRI ArcGIS World Geocoder, though preferred the former given its greater scope of granular coverage. I use the resulting coordinates to construct Thiessen polygons, or artificial boundaries where each point in a country space is matched to the closest address. There are several exceptions to this procedure. In some select data-scarce countries where there should otherwise be official boundaries–where results are reported above the polling station, but the respective shapefiles cannot be found or do not exist–I geocode the place names, such as those of villages, and then construct Thiessen polygons from the spatial centroids. In other cases, like Afghanistan, Germany, or Mongolia, I have polling station-level results without addresses; in these instances, I merge in shapefiles at a higher-level unit of aggregation. For instance, in the case of Germany, I match each polling station to the shapefile of the corresponding municipality.

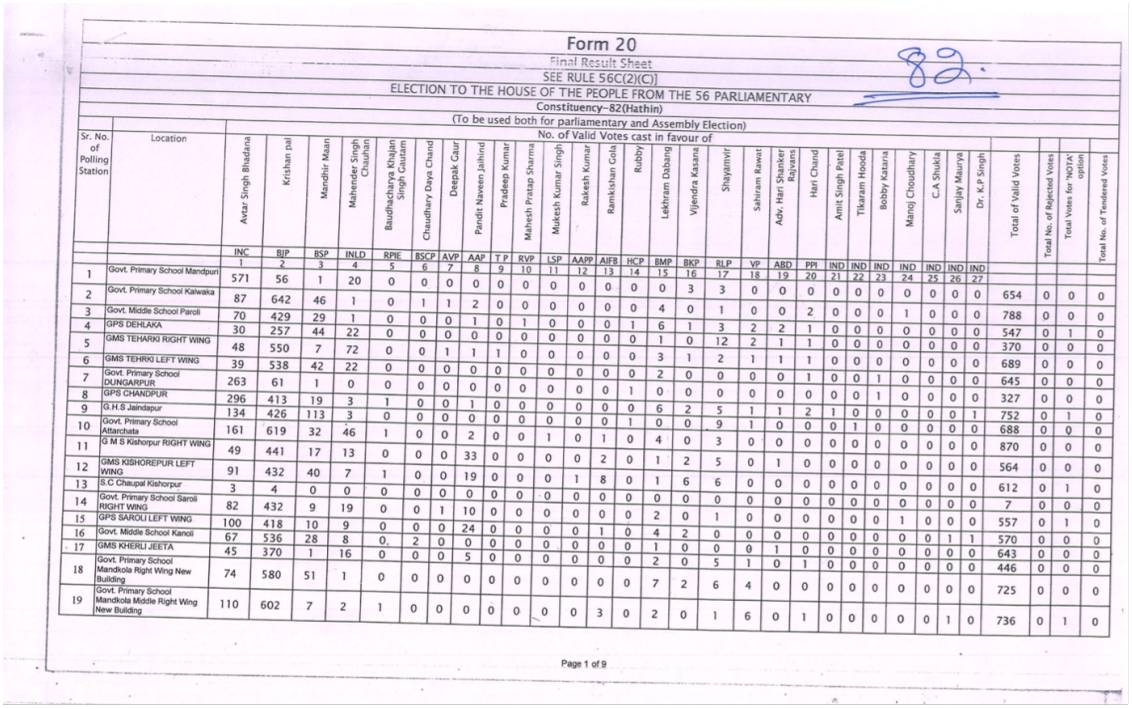
A.3 Example: the 2019 India Lok Sabha Election

Following the 2019 Lok Sabha elections in India, each of the 36 states and territories released "Form 20" PDF files containing polling-station level electoral

[16] I list each of the tools used in the following section on constructing electoral results in India

results, all in tabular format but with varying degrees of quality. Most of the files comprise pictures of printed tables in one of the official languages of India, and many of those pictures are blurry or non-vertical (e.g., tilted). The government only officially reported, and subsequent analyses have exclusively leveraged, constituency-level result–543 units, or nearly 2 million voters per constituency on average. To bring this down to an average of 1,000 voters across nearly a million polling stations, I first manually went through each region's electoral commission website and wrote scripts to download the Form 20 PDFs, totaling 2,896 files each containing anywhere from one to more than 100 pages each.

Example Form 20 polling booth election results.



Form 20

Final Result Sheet

SEE RULE 56C(2)(C)]

ELECTION TO THE HOUSE OF THE PEOPLE FROM THE 56 PARLIAMENTARY

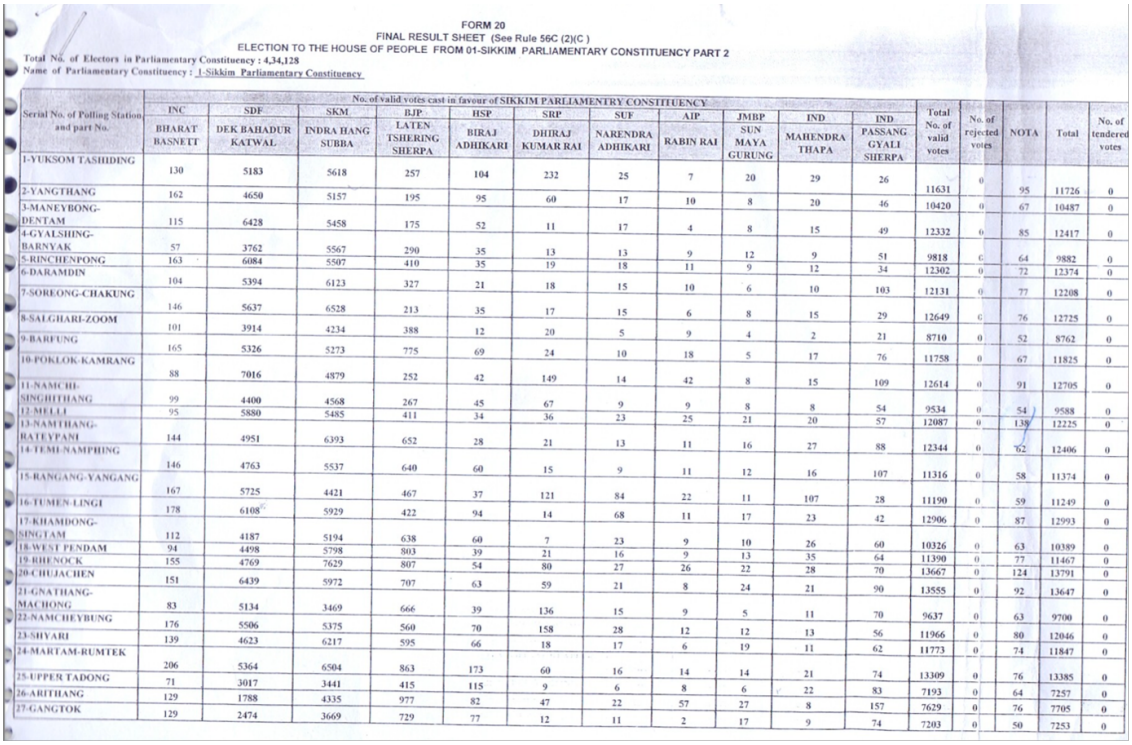
Constituency–82(Hathin)

(To be used both for parliamentary and Assembly Election)

No. of Valid Votes cast in favour of

Sr. No. of Polling Station	Location	Avtar Singh Bhadana	Krishan pal	Mandhir Maan	Mahender Singh Chauhan	Baudhacharya Khajan Singh Gautam	Chaudhary Daya Chand	Deepak Gaur	Pandit Naveen Jaihind	Pradeep Kumar	Mahesh Pratap Sharma	Mukesh Kumar Singh	Rakesh Kumar	Ramkishan Gola	Rubby	Lekhram Dabang	Vijender Kasana	Shayamvir	Sahiram Rawat	Adv. Hari Shanker Rajvanshi	Hari Chand	Amit Singh Patel	Tikaram Hooda	Bobby Kataria	Manoj Choudhary	C.A Shukla	Sanjay Maurya	Dr. K.P Singh	Total of Valid Votes	Total No. of Rejected Votes	Total Votes for 'NOTA' option	Total No. of Tendered Votes
		INC	BJP	BSP	INLD	RPIE	BSCP	AVP	AAP	T P	RVP	LSP	AAPP	AIFB	HCP	BMP	BKP	RLP	VP	ABD	PPI	IND	IND	IND	IND	IND	IND	IND				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27				
1	Govt. Primary School Mandpuri	571	56	1	20	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	654	0	0	0
2	Govt. Primary School Kalwaka	87	642	46	1	0	1	1	2	0	0	0	0	0	0	4	0	1	0	0	2	0	0	0	1	0	0	0	788	0	0	0
3	Govt. Middle School Paroli	70	429	29	1	0	0	0	1	0	1	0	0	0	1	6	1	3	2	2	1	0	0	0	0	0	0	0	547	0	1	0
4	GPS DEHLAKA	30	257	44	22	0	0	0	0	0	0	0	0	0	0	1	0	12	2	1	1	0	0	0	0	0	0	0	370	0	0	0
5	GMS TEHARKI RIGHT WING	48	550	7	72	0	0	1	1	1	0	0	0	0	0	3	1	2	1	1	1	0	0	0	0	0	0	0	689	0	0	0
6	GMS TEHRKI LEFT WING	39	538	42	22	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	645	0	0	0
7	Govt. Primary School DUNGARPUR	263	61	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	327	0	0	0
8	GPS CHANDPUR	296	413	19	3	1	0	0	1	0	0	0	0	0	0	6	2	5	1	1	2	1	0	0	0	0	1	0	752	0	1	0
9	G.H.S Jaindapur	134	426	113	3	0	0	0	0	0	0	0	0	0	1	0	0	9	1	0	0	0	1	0	0	0	0	0	688	0	0	0
10	Govt. Primary School Attarchata	161	619	32	46	1	0	0	2	0	0	1	0	1	0	4	0	3	0	0	0	0	0	0	0	0	0	0	870	0	0	0
11	G M S Kishorpur RIGHT WING	49	441	17	13	0	0	0	33	0	0	0	0	2	0	1	2	5	0	1	0	0	0	0	0	0	0	0	564	0	0	0
12	GMS KISHOREPUR LEFT WING	91	432	40	7	1	0	0	19	0	0	0	1	8	0	1	6	6	0	0	0	0	0	0	0	0	0	0	612	0	1	0
13	S.C Chaupal Kishorpur	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
14	Govt. Primary School Saroli RIGHT WING	82	432	9	19	0	0	1	10	0	0	0	0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	0	557	0	1	0
15	GPS SAROLI LEFT WING	100	418	10	9	0	0	0	24	0	0	0	0	1	0	4	2	0	0	0	0	0	0	0	0	1	1	1	570	0	0	0
16	Govt. Middle School Kanoi	67	536	28	8	0	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	643	0	0	0
17	GMS KHERLI JEETA	45	370	1	16	0	0	0	5	0	0	0	0	0	0	2	0	5	1	0	1	0	0	0	0	0	0	0	446	0	0	0
18	Govt. Primary School Mandkola Right Wing New Building	74	580	51	1	0	0	0	0	0	0	0	0	0	0	7	2	6	4	0	0	0	0	0	0	0	0	0	725	0	0	0
19	Govt. Primary School Mandkola Middle Right Wing New Building	110	602	7	2	1	0	0	0	0	0	0	0	3	0	2	0	1	6	0	1	0	0	0	0	0	1	0	736	0	1	0

Page 1 of 9



FORM 20
FINAL RESULT SHEET (See Rule 56C (2)(C)
ELECTION TO THE HOUSE OF PEOPLE FROM 01-SIKKIM PARLIAMENTARY CONSTITUENCY PART 2
Total No. of Electors in Parliamentary Constituency : 4,34,128
Name of Parliamentary Constituency : 1-Sikkim Parliamentary Constituency

Serial No. of Polling Station and part No.	No. of valid votes cast in favour of SIKKIM PARLIAMENTRY CONSTITUENCY											Total No. of valid votes	No. of rejected votes	NOTA	Total	No. of tendered votes
	INC BHARAT BASNETT	SDF DEK BAHADUR KATWAL	SKM INDRA HANG SUBBA	BJP LATEN TSHERING SHERPA	HSP BIRAJ ADHIKARI	SRP DHIRAJ KUMAR RAI	SUF NARENDRA ADHIKARI	AIP RABIN RAI	JMBP SUN MAYA GURUNG	IND MAHENDRA THAPA	IND PASSANG GYALI SHERPA					
1-YUKSOM TASHIDING	130	5183	5618	257	104	232	25	7	20	29	26	11631	0	95	11726	0
2-YANGTHANG	162	4650	5157	195	95	60	17	10	8	20	46	10420	0	67	10487	0
3-MANEYBONG-DENTAM	115	6428	5458	175	52	11	17	4	8	15	49	12332	0	85	12417	0
4-GYALSHING-BARNYAK	57	3762	5567	290	35	13	13	9	12	9	51	9818	0	64	9882	0
5-RINCHENPONG	163	6084	5507	410	35	19	18	11	9	12	34	12302	0	72	12374	0
6-DARAMDIN	104	5394	6123	327	21	18	15	10	6	10	103	12131	0	77	12208	0
7-SOREONG-CHAKUNG	146	5637	6528	213	35	17	15	6	8	15	29	12649	0	76	12725	0
8-SALGHARI-ZOOM	101	3914	4234	388	12	20	5	9	4	2	21	8710	0	52	8762	0
9-BARFUNG	165	5326	5273	775	69	24	10	18	5	17	76	11758	0	67	11825	0
10-POKLOK-KAMRANG	88	7016	4879	252	42	149	14	42	8	15	109	12614	0	91	12705	0
11-NAMCHI-SINGHITHANG	99	4400	4568	267	45	67	9	9	8	8	54	9534	0	54	9588	0
12-MELLI	95	5880	5485	411	34	36	23	25	21	20	57	12087	0	138	12225	0
13-NAMTHANG-RATEYPANI	144	4951	6393	652	28	21	13	11	16	27	88	12344	0	62	12406	0
14-TEMI-NAMPHING	146	4763	5537	640	60	15	9	11	12	16	107	11316	0	58	11374	0
15-RANGANG-YANGANG	167	5725	4421	467	37	121	84	22	11	107	28	11190	0	59	11249	0
16-TUMEN-LINGI	178	6108	5929	422	94	14	68	11	17	23	42	12906	0	87	12993	0
17-KHAMDONG-SINGTAM	112	4187	5194	638	60	7	23	9	10	26	60	10326	0	63	10389	0
18-WEST PENDAM	94	4498	5798	803	39	21	16	9	13	35	64	11390	0	77	11467	0
19-RHENOCK	155	4769	7629	807	54	80	27	26	22	28	70	13667	0	124	13791	0
20-CHUJACHEN	151	6439	5972	707	63	59	21	8	24	21	90	13555	0	92	13647	0
21-GNATHANG-MACHONG	83	5134	3469	666	39	136	15	9	5	11	70	9637	0	63	9700	0
22-NAMCHEYBUNG	176	5506	5375	560	70	158	28	12	12	13	56	11966	0	80	12046	0
23-SHYARI	139	4623	6217	595	66	18	17	6	19	11	62	11773	0	74	11847	0
24-MARTAM-RUMTEK	206	5364	6504	863	173	60	16	14	14	21	74	13309	0	76	13385	0
25-UPPER TADONG	71	3017	3441	415	115	9	6	8	6	22	83	7193	0	64	7257	0
26-ARITHANG	129	1788	4335	977	82	47	22	57	27	8	157	7629	0	76	7705	0
27-GANGTOK	129	2474	3669	729	77	12	11	2	17	9	74	7203	0	50	7253	0

Note: Examples of Form 20 results pages from the states of Haryana and Sikkim, respectively. Images are displayed in their original resolutions.

To process these PDF files, I wrote code that leveraged transformers-based deep learning models[17] to "cut out" the results table from each page, then ran the resultant images through proprietary OCR software.[18] For some of the smaller territories, I manually transcribed the voting returns for each polling station. This process produced tens of thousands of CSV files of varying quality and orientation. I then manually went through each state and territory separately and wrote code that cleaned and processed the data to a standard tabular format, resulting in vote-count tables by candidate for each constituency-polling station.

Because multilingual candidate names are more difficult for OCR software to read than vote counts, any single constituency had multiple strings corresponding to the same candidate, many of which were unintelligible. Additionally, while some constituencies reported both candidates and their respective parties—such as those in the above pictures—most only provided candidate names without a corresponding party. To fix this, I matched tens of thousands of unique candidate strings by hand to a "best-guess" name and party. This required referencing several different sources, such as the open-source webpage MyNeta, which pulls candidate information from the Electoral Commission of India, and

[17]Specifically, I used Microsoft's Table Transformer model.

[18]Because each state and territory released results of a wholly distinct format, different methods outperformed others in each unique setting. I differentially used, across all regions, Amazon Textract, Azure AI Document Intelligence, Adobe Acrobat, ABBYY FineReader, Tesseract, and PaddleOCR, among others.

the Indian Express daily newspaper, which reported candidate-specific vote counts by state. I used translation services when matching non-English candidate names and referenced the original PDF files where candidate names were OCR'd poorly. Finally, although the Indian government released a list of polling station numbers, addresses, and the respective coordinates, most of these coordinates are demonstrably incorrect—for instance, a single latitude-longitude value is assigned to every polling station in multiple states and territories. I therefore geocode all the addresses in the master file myself.

To validate the India results, the only country case with researcher-induced error in vote counts, I compare state- and territory-level party vote aggregates to government reporting using raw and log correlations and root mean-squared-error (RMSE). While there is a degree of error in vote counts at the polling station level, owing to OCR issues, and in the geocoding, from missing addresses in Google, the raw vote correlations range from .933 to .999; the logs from .935 to .999; and the RMSEs from no more than a hundred up to a million (in cases where whole constituencies are missing from the data because of missing files or illegible images). I also compare the country-level party totals, where the overall and log correlations with official reporting are .999 and .956, respectively. Nearly all (.944) of the 866,557 polling stations were successfully geocoded.

A.4 Nearest-Neighbor Buildings as Urban-Rural

Estimating k-nearest neighbor distances at different levels of k adjustments across, rather than within, electoral unit boundaries reduces bias from the modifiable areal unit problem, whereby estimates vary conditional on arbitrary levels of granularity (Fotheringham and Wong, 1991). Computationally, the process can be formalized as follows:

Let $B = \{b_1, b_2, ..., b_n\}$ be the set of all buildings in a country. For each voting area V, let S_V be a random sample of up to 30 buildings from B_V, the set of buildings within V.

For each building $b_i \in S_V$, I calculate:

$$D_k(b_i) = \frac{1}{k} \sum_{j=1}^{k} d(b_i, N_j(b_i)) \tag{2}$$

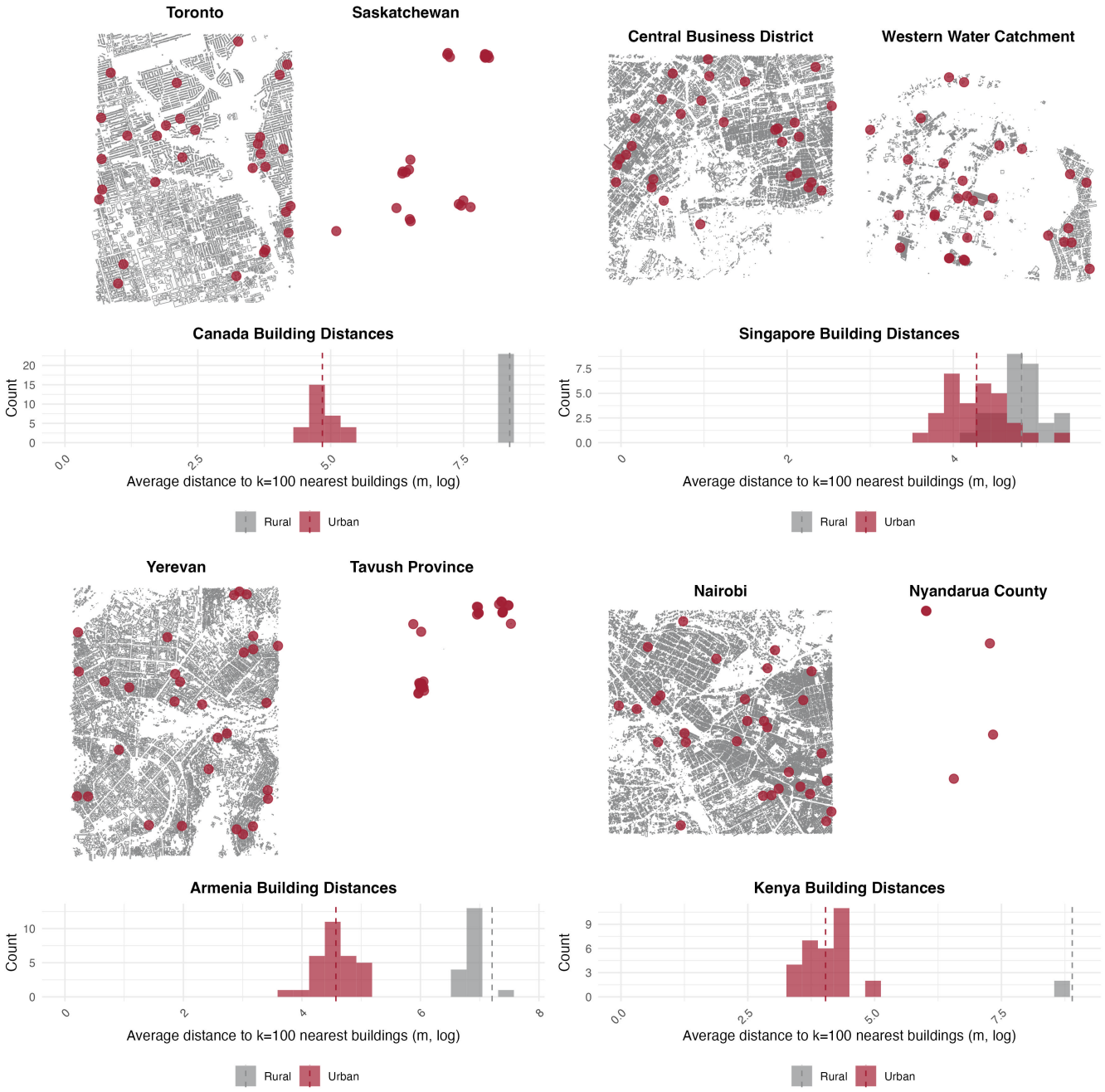
where $D_k(b_i)$ is the average distance to the k nearest neighbors of building b_i, $d(b_i, b_j)$ is the Euclidean distance between buildings b_i and b_j, and $N_j(b_i)$ is the j-th nearest neighbor of b_i in B.

For each voting area V, I then compute:

$$U_k(V) = \frac{1}{|S_V|} \sum_{b_i \in S_V} D_k(b_i) \tag{3}$$

where $U_k(V)$ is the urbanicity measure for voting area V using k nearest neighbors. This approach is implemented using a k-d tree data structure for efficient nearest neighbor searches, and I use parallel processing with batch operations to optimize computational efficiency.

Examples of urban and rural $k-$nearest-neighbor distances.



The above figure provides examples of more "urban" and "rural" places in a sample of countries and their respective nearest-neighbor building distances. To my knowledge, this is the first use of between-building distances as a measure of the urban-rural continuum, though I also use population density as a supplemental measure (the building footprints are not estimated across time). In contrast to population density, nearest-neighbor building distances match more closely, for instance, the transition from a "downtown" city centre, where buildings are closest together, to the suburbs, then the exurbs, then complete rurality, as the buildings gradually separate in turn. Nonetheless, a correlation heatmap (Appendix Figure A2) of the urban-rural vote-share estimates (see Fig-

ure 4) between those derived from voter density and those from each of the nearest-neighbor algorithms reveals strong correlation: for instance, estimates from $k = 30$ and voter density are 61% correlated. High correlations across k values indicate relatively low sensitivity to, and thus bias induced by, researcher choice.[19] Nearest-neighbor building distances are therefore internally consistent, inducing low estimate variability across values of k, and externally valid, demonstrating sufficiently high correlations with extant conceptualizations of urban-rural (Adcock and Collier, 2001; high *construct validity*).

[19]Estimates obtained from average nearest-neighbor building distances where $k = 270$ are correlated 84%, 84%, 90%, 85%, and 89% with mean $k = 30$, mean $k = 90$, and median $k = 30$, $k = 90$, and $k = 270$, respectively.

A.5 Party Ideology Examples

Table 2: Most Left and Right Tags

Most "Left" Tags	Score	Most "Right" Tags	Score
Eco-socialism	1.76	Far-right	8.27
Left-wing to far left	2.09	Christian right	8.11
Marxism-Leninism	2.55	Nationalism	7.57
Democratic socialism	2.64	National conservatism	7.04

Table 3: Example Parties

Example "Left" Parties	Ideology	Example "Right" Parties	Ideology
Sandinista National Liberation Front (Nicaragua)	1.78	Al-Islah (Yemen)	9.00
Progressive Party of the Working People (Cyprus)	2.09	Renua (Ireland)	8.51
Workers' Revolutionary Party (Spain)	2.13	Serbian Radical Party (Serbia)	8.19
Japanese Communist Party (Japan)	2.24	The Jewish Home (Israel)	8.10

A.6 Validating Party Ideology

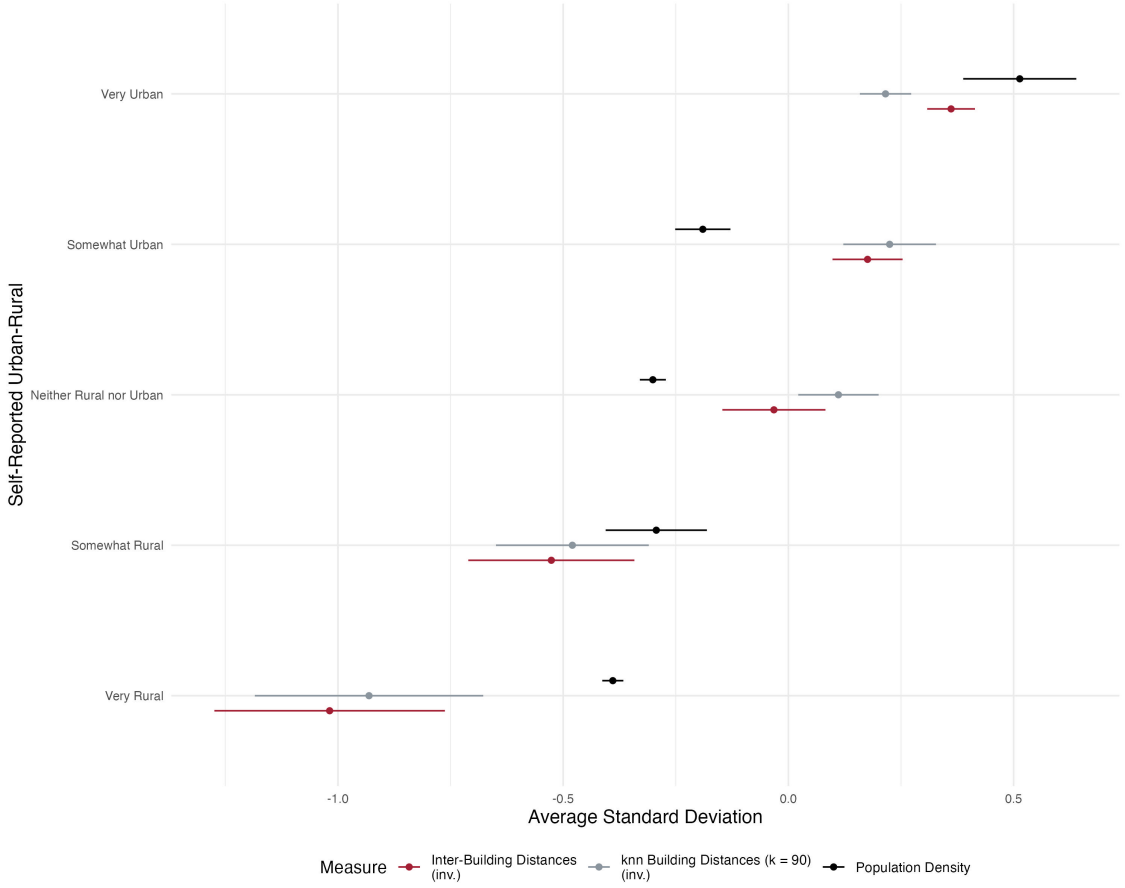
I extensively validate these unidimensional party ideology scores against existing measures. The matched country-party correlations with expert scores from the Global Party Survey is 94.7% (n = 828); 76.7% and 80.1% respectively with the economic left-right (n = 367) and social liberal-conservative (n = 106) scores from V-Party (Düpont et al., 2021); 91.4% and 76.0% with the CHES Europe left-right and post-materialist-traditional scores (n = 110; Jolly et al., 2022); 75.3% and 86.6% with the same dimensions from CHES Latin America (n = 110; Martínez-Gallardo et al., 2023); and upwards of 82.2% with the Manifesto Project.[20][21] Overall, this approach allows me to assign ideology scores to parties not conventionally represented in expert surveys, such as regionalist and minor local parties or parties in hybrid regimes. The main advantage of this measure is therefore its greater spatial coverage (3,428 unique parties) compared to alternatives, though at the cost of time invariance and the inability to score independent parties or presidential candidates. Thus, while the tests for cross-national urban-rural cleavages (independent of ideological direction) leverage both legislative and presidential voting results, the ideological tests draw exclusively on legislative elections.

[20] Because the measure proposed here is time-invariant, correlations with left-right scores from V-Party and the Manifesto Project are contingent on the time period of comparison. The correlation value should be higher where the time period compared is more recent, since the Wikipedia labels are scraped from the most recent iteration of the party encyclopedia entry.

[21] Because party names vary across datasets, I leverage deep-learning libraries to match greatest likelihood pairs with string embeddings (Arora and Dell, 2023).

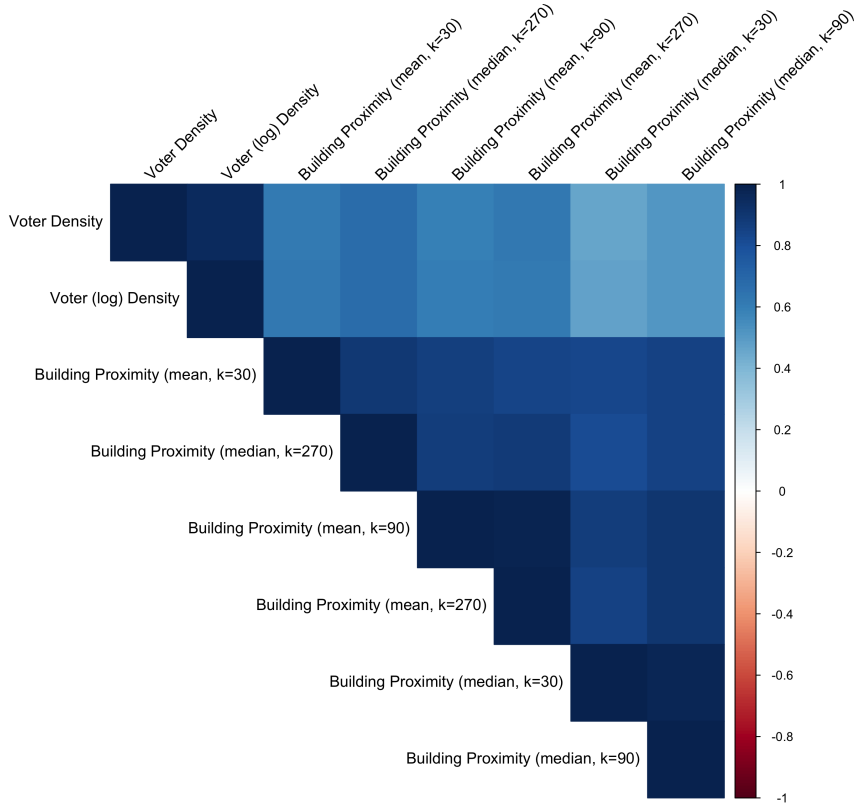
A.7 Building Survey Comparison

Appendix Figure A1. Correlations between measures, and self-reported status, of urban and rural among survey respondents.



A.8 Correlation Heatmap Between Urban-Rural Measures

Appendix Figure A2. Average correlation between $k-$ nearest neighbor building distances, inter-building dyadic distances, and population density with self-reported urban-rural classifications.



A.9 Urbanicity–Vote-Share Correlations Across Elections

Appendix Figure A3. Country-party-year point estimates for correlation with building proximity, by region (left) and over time (right).