CODEBOOK · EDITION 1.0

Multi-Level Political Knowledge

Nicholas Clark, Susquehanna University
Todd Makse, Florida International University

September 9, 2026
Edition 1.0 — figures computed from the replication file

About this codebook

This codebook documents a survey of political knowledge across three levels of government, fielded in two waves and covering 3,171 respondents in 20 American cities. Its distinguishing feature is that the local knowledge questions are written specifically for the city each respondent lives in, so that local knowledge can be measured with the same kind of factual question conventionally used for national politics rather than by accepting any name a respondent offers.

The document is organized so that a reader who has never seen the study can use the data without asking the authors anything, and so that a reader who intends to publish from it knows precisely where it is thin. Section 8 is the honest account of the limits, and it should be read before the data are used, not after.

1 What this dataset is

Research on political knowledge has overwhelmingly asked what citizens know about national government. That imbalance is not an oversight so much as a measurement problem: asking a factual question about local politics requires content tailored to every respondent's own community, which is impractical in a nationally representative sample drawn from thousands of places. The usual workaround — asking respondents to name their mayor and treating any name as correct — measures something, but not knowledge.

This study takes the other route. By concentrating the sample in 20 cities, it becomes possible to write genuine multiple-choice knowledge questions for each city: who the mayor is, who the police chief is, which office sets the property tax rate, how long a mayoral term runs and whether it is limited. Parallel four-item batteries cover national and, in the second wave, state politics, so knowledge at different levels is compared within the same respondent rather than across different studies with different instruments.

The released file holds 3,171 respondents and 49 variables: 1,558 from the first wave and 1,613 from the second. Alongside the knowledge items it carries demographics, partisanship and ideology, residential tenure and attachment, media consumption, survey weights, and thirteen measures of the respondent's neighborhood drawn from the American Community Survey.

2 How the survey was fielded

Both waves were administered online through the Lucid platform. Respondents were screened to those living inside the city limits of a sampled city, using self-reported ZIP code, because a knowledge question about a city's mayor is only meaningful for people that mayor governs — suburban residents of the same metropolitan area were therefore excluded rather than absorbed.

Wave 1 was fielded in April 2022 and wave 2 in June 2023. The sampled cities were selected from American cities of at least 250,000 residents, chosen to vary in size, region, and municipal institutional form so that the set is collectively representative of medium-sized and large American cities. The study does not claim to represent Americans generally, and in particular says nothing about suburban or rural residents.

June and April were both chosen to sit outside periods of heavy campaign attention, so that measured knowledge reflects a normal information environment rather than the tail of a campaign.

3 The cities

Code	City	Wave	N
14	Buffalo	2	124
19	Philadelphia	2	225
28	Charlotte	2	147
32	Jacksonville	1	185
38	Memphis	1	130
40	Louisville	1	187
43	Columbus	1	194
46	Indianapolis	1	188
53	Milwaukee	2	145
60	Chicago	2	153
67	Wichita	2	133
68	Omaha	2	135
70	New Orleans	2	126
74	Tulsa	1	149
78	Austin	1	151
80	Colorado Springs	2	131
85	Tucson	1	188
89	Las Vegas	2	152
92	San Diego	1	186
97	Portland	2	142

TABLE 3.1. *The 20 cities, their citycode values, the wave in which each was fielded, and the number of respondents. Every city appears in exactly one wave.*

The two waves cover entirely different cities: the 9 cities in wave 1 and the 11 in wave 2 do not overlap at all. This is a design fact with real consequences for analysis, and section 8 returns to it.

4 Weights

The variable weight carries a survey weight built from city-level Census data, so that each city sample is brought into line with that city's population on the characteristics used to construct it. Weights run from 0.054 to 6.303 and average 1.00 by construction. They are missing for 4.5% of respondents.

Analyses reported in published work using these data are weighted. Unweighted estimates will differ, and the difference is not negligible in the cities where the achieved sample departed furthest from the population.

5 Variables

Variables are grouped below by what they measure. Codings are given exactly as they appear in the released file; where a variable is a dummy it takes 1 for the condition named and 0 otherwise.

5.1 Identifiers and administrative

Variable	Description	Coding
id	Survey unique identifier	
wave	Survey wave	1 = April 2022; 2 = June 2023
weight	Survey weight, from city-level Census data	
duration	Length of survey in seconds	
citycode	City of residence	See Table 3.1
zip	Postal code at respondent's residence	See section 8
philly_error	Flag: Philadelphia knowledge item unusable owing to a programming error	1 = affected
polknow_search	Flag: respondent likely searched during the knowledge questions	1 = likely searched

5.2 Demographics

Variable	Description	Coding
age	Age in years	
female	Dummy: respondent is female	
educ4	Education	0 = HS or less; 1 = Some college; 2 = College degree; 3 = Graduate degree
income4	Household income	0 = Under $25K; 1 = $25–50K; 2 = $50–100K; 3 = $100K+
ethnicity	Racial and ethnic self-identification	1 = Asian-American; 2 = Black; 3 = Hispanic; 4 = White; 5 = Other
pid7	Party identification, 7-point	0 = Strong Democrat … 6 = Strong Republican (not asked in wave 1)
pid3a	Partisanship, 3 categories	0 = Democrat; 1 = Republican; 2 = Independent
ideol	Ideological self-identification, 7-point	0 = Extremely liberal … 6 = Extremely conservative

5.3 Residence and attachment

Variable	Description	Coding
urban	Place type	0 = Rural; 1 = Suburban; 2 = Urban
homeown	Dummy: respondent owns their home	
lived_res	Time in current residence	0 = Under 1 year; 1 = 1–5; 2 = 5–10; 3 = Over 10 years
lived_city	Time in city	0 = Under 1 year; 1 = 1–5; 2 = 5–10; 3 = Over 10 years
soc_moveout	Agreement that they would like to move out of the city	0 = Strongly disagree … 4 = Strongly agree

5.4 Media consumption

Variable	Description	Coding
mc_loctv	Days per week consuming local television	0–7
mc_locpaper	Days per week consuming a local newspaper	0–7
mc_cable	Days per week consuming cable television	0–7
mc_natpaper	Days per week consuming a national newspaper	0–7
loc_media	Local minus national media consumption	(mc_loctv + mc_locpaper) − (mc_cable + mc_natpaper)

6 The knowledge batteries

Each battery is four multiple-choice items. The close-ended format was chosen deliberately: it holds response rates up and avoids the coder-judgment problem that open-ended knowledge questions create, at the cost of some guessing, which the summary scores do not correct for.

Variable	Level	Item	% correct
npk1	National	Correctly identified the Speaker of the House	66.8%
npk2	National	Correctly identified Antony Blinken	45.8%
npk3	National	Correctly identified the majority required to override a presidential veto	64.5%
npk4	National	Correctly identified who decides whether a law is constitutional	67.6%
lpk1	Local	Correctly identified the mayor of their city	78.5%
lpk2	Local	Correctly identified their city's police chief	52.9%
lpk3	Local	Correctly identified the local office that sets the property tax rate	69.7%
lpk4	Local	Correctly identified the length and term limits of the mayoralty	40.8%

TABLE 6.1. *The eight knowledge items and the unweighted share answering each correctly across the full file.*

The two summary scores, polknow_nat and polknow_loc, count correct answers out of four. National knowledge averages 2.45 and local knowledge 2.42, and the two correlate at r = 0.43 — related, but far from the same thing, which is the finding that motivates measuring them separately.

Two flags qualify these scores. polknow_search marks respondents who answered a catch question correctly in a way suggesting they looked answers up (15 respondents). philly_error marks Philadelphia respondents for whom a programming error made one item unusable (91 respondents). Neither is dropped from the released file; both should be handled explicitly.

7 Neighborhood context

Thirteen variables describe the respondent's ZIP code tabulation area, drawn from the American Community Survey five-year estimates ending 2020 and, for the vote measures, from precinct returns aggregated to the ZCTA. They are attached by the respondent's ZIP code and are properties of the area, not of the person.

Variable	Description	% missing
zcta_medinc	Median household income, logged	0.1%
zcta_dens	Population density, logged	0.0%
zcta_drive	Share commuting to work by car	0.0%
zcta_homeage	Median age of the housing stock	0.0%
zcta_herf	Racial diversity, Herfindahl index	0.0%
zcta_biden	Share of the two-party vote for Biden	0.0%
zcta_inparty	Share voting for the respondent's own preferred party	22.0%
zcta_baplus	Share holding a BA or higher	0.0%
zcta_white	Share White	0.0%
zcta_black	Share Black	0.0%
zcta_asian	Share Asian	0.0%
zcta_hisp	Share Hispanic	0.0%
zcta_sameeth	Share of the same race or ethnicity as the respondent	9.4%

TABLE 7.1. *Neighborhood measures. zcta_inparty and zcta_sameeth are missing more often than the rest because each depends on a respondent characteristic — party and ethnicity — that is itself sometimes missing.*

8 Limits and cautions

8.1 The waves are different cities, not the same cities twice

The 9 cities in wave 1 and the 11 in wave 2 do not overlap. Wave is therefore perfectly confounded with city: any apparent difference between 2022 and 2023 is equally a difference between one set of cities and another, and no analysis of these data can separate the two. Pooling the waves is reasonable for individual-level questions; treating wave as a time comparison is not.

8.2 State knowledge exists only in wave 2

The state battery was added in the second wave. Any analysis comparing all three levels of government is therefore restricted to the wave 2 cities, and the full replication file for the published local-knowledge analysis does not carry the state items.

8.3 ZIP code is in the file

The file carries the respondent's ZIP code alongside age, gender, race, education, income, homeownership and tenure. Across 514 ZIP codes and 3,171 respondents, that combination is close to unique for most of the sample, and it is what makes re-identification against an external list conceivable. The neighborhood measures in section 7 are aggregates and carry no such risk. Anyone circulating or depositing these data should think about whether the ZIP variable needs to travel with them; the published analyses do not use it directly.

8.4 Guessing, and what the scores do not correct for

The knowledge scores are counts of correct answers on close-ended items, with no correction for guessing and no separate "don't know" option modeled. A respondent who guesses well is scored as knowing. This is a deliberate trade against the alternative of open-ended items, but it means the absolute levels should be read as upper bounds and the comparisons across levels of government — which share the format — trusted more than the levels themselves.

8.5 Who the sample represents

These are residents of medium-sized and large American cities, screened to inside the city limits. The set of cities is chosen to be collectively representative of that universe, not of the country. Nothing here speaks to suburban, small-town or rural political knowledge, and the exclusion of suburban residents of the same metropolitan areas is deliberate rather than incidental.

APPENDIX A Complete variable list

All 49 variables in the released file, in file order, with the share of respondents missing each.

Variable	Missing	Variable	Missing
id	0.0%	polknow_search	0.0%
wave	0.0%	polknow_nat	0.0%
weight	4.5%	polknow_loc	0.0%
duration	0.0%	npk1	0.0%
citycode	0.0%	npk2	0.0%
philly_error	49.1%	npk3	0.0%
zip	0.0%	npk4	0.0%
age	0.0%	lpk1	0.0%
income4	0.0%	lpk2	0.0%
educ4	0.5%	lpk3	0.0%
female	0.6%	lpk4	0.0%
urban	0.0%	zcta_medinc	0.1%
pid7	49.4%	zcta_dens	0.0%
pid3a	0.3%	zcta_drive	0.0%
ideol	12.3%	zcta_homeage	0.0%
ethnicity	1.9%	zcta_herf	0.0%
homeown	0.0%	zcta_biden	0.0%
lived_res	0.3%	zcta_inparty	22.0%
lived_city	0.9%	zcta_baplus	0.0%
soc_moveout	0.9%	zcta_white	0.0%
mc_loctv	0.0%	zcta_black	0.0%
mc_locpaper	0.0%	zcta_asian	0.0%
mc_cable	0.0%	zcta_hisp	0.0%
mc_natpaper	0.0%	zcta_sameeth	9.4%
loc_media	0.0%		

TABLE A.1. *Variables and missingness, computed from the released file (3,171 respondents).*