Replication Package for:
How Partisan are U.S. Local Elections?
Evidence from 2020 Cast Vote Records*

American Political Science Review

Contents

1 Replication Instructions **1**

2 Dataset Structure and Table of Contents **1**
2.1 Universe of contests . . . 1
2.2 Wide (contest-level) datasets . . . 2
2.3 Grouped Long (voter × contest-level) datasets . . . 4
2.4 Auxilary Datasets . . . 4

3 List of What Produces What **5**
3.1 Each Script . . . 5
3.2 Each Table and Figure . . . 5

4 Other Specifications **6**
4.1 Hardware Specifications . . . 6
4.2 Software Dependencies . . . 6
4.3 Runtimes . . . 7
4.4 Random Seeds . . . 7

1 Replication Instructions

1. Download the Dataverse. Note: Download ingested data files as their original .dta, not .tab.
2. Change the filepath specifications in the Stata scripts to appropriate filepaths on your computer. These are defined at the top of each script, and are the local variables `PATH_TBL` for Tables, `PATH_FIG` for Figures, etc.. All scripts will run smoothly if there is a paper directory with the following subdirectories: `tables`, `figures`, and `numbers`. R scripts will work without editing the filepaths as long as the directory is set to this project directory.
3. Unzip (or uncompress) `tb7_long_local_ballot_measures_education_extract.dta.zip` and `tb7_long_state_partisan_races_education_extract.zip` into `.dta` files.
4. Run each script. They do need to be run in order, but the numbers roughly correspond to the order in which the results appear in the paper. Each particular Stata file need to be run all at once to export

*For further questions on the replication data, please contact shiro.kuriwaki@yale.edu or any of the other authors.

locals into tables and numbers.

2 Dataset Structure and Table of Contents

This project analyzes the vote choices of 50 million voters across their entire ballot, amounting to 1.3 billion voter $\times$ candidate combinations. All of the analyses in the main text can be done by much smaller grouped versions of this data.

2.1 Universe of contests

Contests are defined in the following metadata file, used in several scripts:

- `APSR_item_choice_info.dta`

The unit of observation in this file is a contest $\times$ candidate (choice) combination. This file includes the contests we define in the paper: All contests we can merge with a national partisanship and are not less than 100 votes.

In this file and elsewhere, a *contest* is keyed by the variable `item_id_2`, which is defined by the following attributes:

- `state`: State (AK, AZ, CA, etc..).
- `item`, the type of office (state vs. local; partisan vs. nonpartisan vs. measure).
- `dist`: a district for some contests. For example, for state legislative candidates, `item == ST_REP` for all districts and `dist` uniquely identifies a contest.
- `unexp_term`: a distinction for general elections (0) or special elections (1), which usually entailing filling a vacancy.
- `place`: Essentially the constituency of the contest. For example, a mayor's constituency may span multiple counties, or it may be only a subset of ballots in one county. In this case, a `place` is manually assigned separately from `county`. `county` only indicates where the ballot came from and does not help define a contest. If there are numerical districts within a place, use `dist` in addition.

Candidate attributes include `choice` (the name of the choice), `choice` (the nominal partisanship of the candidate on the ballot, or Yes or No). We provide vote counts for each choice for contest.[1] Other attributes of the item are documented later in this document.

2.2 Wide (contest-level) datasets

Datasets that start with `wide_` are small files where the unit of observation is a contest. That is, each row defines a contest (see above). These are the main datasets for this article. We partition the contests into one of six types[2]:

[1] `votes` is the total votes for that choice in that county in the CVR data, `votes_choice` is the total votes across counties for the same choice in the same contest, and `votes_total` are the total votes across choices for that contest.

[2] In the `APSR_item_choice_info.dta`, these types of contests can be constructed by a combination of `level` (N = national, S = state, L = local), `nonpartisan` (1 = nonpartisan, 0 = partisan, missing = measure), `measure` (1 = ballot measure, 0 = candidate contest).

1. wide_local_ballot_measures.dta: local ballot measures
2. wide_local_nonpartisan_races.dta: local offices that are nonpartisan
3. wide_local_partisan_races.dta: local offices that are partisan
4. wide_state_ballot_measures.dta: statewide ballot measures
5. wide_state_nonpartisan_races.dta: state offices that are (nominally) nonpartisan
6. wide_state_partisan_races.dta: state offices that are (nominally) partisan

The main variables in these datasets are of the form []_choice_[]_pid. These variables indicate the total number of voters cast for a particular choice among a particular national partisanship. The first letter indicates the choice of the voter in that particular contest:

- D: Democrat in a partisan contest
- R: Republican in a partisan contest
- A: Top vote getter in the nonpartisan contest
- B: Second vote-getter in the nonpartisan contest
- Y: Yes on a referendum
- N: No on a referendum
- U: Undervote in the contest (i.e., abstaining from that contest)
- X, X3, ..., X11: Votes for the third, fourth, ..., eleventh vote-getter. For example, in LA county community college board of trustees district 3 race, there were eleven candidates for one seat. X1 and X2 are often 0 and transferred instead to A and B.

Then, variables with D_pid indicate that the vote is only among national Democrats, and variables with R_pid indicate that the vote is only among national Republicans. Thus, these variables together represent the cells of a two-contest vote matrix:

Contest	National Partisanship		
Choice	D	R	S (not used)
D	D_choice_D_pid	D_choice_R_pid	D_choice_S_pid
R	R_choice_D_pid	R_choice_R_pid	R_choice_S_pid
U	U_choice_D_pid	U_choice_R_pid	U_choice_S_pid

The Absolute Partisan Gap we use throughout the paper is essentially abs(D_choice_D_pid / D_pid - D_choice_R_pid / R_pid) where D_pid = D_choice_D_pid + R_choice_D_pid

The wide datasets also have several metadata and covariates that are used in the paper. For ballot measures:

- T: The sum of four cells represented in the 2 by 2 subtable at the to-left of the table above (e.g., Yes and Nos and national Ds and national Rs, but not including national splitters or undervotes). We use this to enforce the constraint of more than 100 voters described in the paper.
- N: Total votes, including undervotes.

- `measure`: Indicator for ballot measures (all 1) or candidate contests (all 0)
- `yes_is_lib`: for ballot measures. Indicator for whether voting "Yes" on the referendum is for a liberal position. If 1, Y = Liberal. If 0, N = Liberal. If missing, then this is a non-ideological ballot measure. Used in Tables 3, 5, 6.
- `spending`: A binary variable for ballot measures. Whether the ballot measure is for spending (taxes and bonds). Used in Tables 5 and 6.
- `topic`: Our manual coding of the topic of the ballot measure. If `spending == 1`, then topic indicates what the spending is for. Note that a few topics are not mutually exclusive. Used in Table 6.

Finally, we derive the variable `competitive` in the script numbered `02`. This variable defines whether a contest is contested using the 95-5 two-party threshold discussed in the paper.

2.3 Grouped Long (voter × contest-level) datasets

These datasets are more granular than the main wide datasets. They are either granular microdata where each row defines a voter-choice combination, or a grouped version of that. These are the following datasets:

- `tb7_long_local_ballot_measures_education_extract.dta`: Long data only for local education ballot measures in Table 7. Unzip the `.dta.zip` before use.
- `tb7_long_state_partisan_races_education_extract.dta`: The state partisan offices for the same voters in the above dataset. Unzip the `.dta.zip` before use.
- `tbC6_michigan-grouped`: Grouped data for Michigan regression in Table C6. Grouped at contest × county × partisanship (a federal version and party lever version).
- `tbC7_long_data_extract.dta`: Grouped data for stacked regression in Table C7, at the contest × `pid` × vote choice level. The vote choice category is represented by `LEFT`. LEFT is coded following the description in Table 3 of the main text, such that LEFT = 1 indicates vote choices that are preferred by Democrats more than Republicans.

Each row in the original long dataset is uniquely defined by `state` × `county` × `cvr_id` × `item_id_2`. `cvr_id` is a unique ballot ID within each state-county combination. In most cases, each ID represents a voter, but sometimes represent ballot fragments.[3]

For each of these unique rows, `pid` is the voter's partisanship as measured by their votes for President and Congress (1 for Democrat, 0 for split, -1 for Republican), and `choice` is their vote choice on the context (see definitions of D, R, A, B, above).

We produce a *grouped* (or collapsed) format of these long datasets, that are a fraction of the size and provide the sufficient statistics for all our regressions, figures, tables. That is, summary statistics such as counts at the contest × partisanship level represent the same amount of information as the long data, for the purposes of our analysis.

2.4 Auxilary Datasets

Other one-off datasets that are used for some figures. These are not CVR data.

[3] Ballot fragments are rare in our dataset because a ballot fragment must match with a federal office.

- `APSR_counties_used.dta`: A list of counties used along with the total number of votes for Presidential votes cast in that county (CVR or not). Table A2-A3.
- `fg4_gallup_trust-in-govt.dta`: Figure 4, Gallup Poll Estimates. See article for citation.
- `tbA1_counties-demographics.dta`: All U.S. counties merged with Decennial Census variables. Counties with non-missing values for `st` and `county_name` are counties for which we do not have CVR data.
- `tbA2_nvoters_by-county.dta`: Pre-computed counts of unique voters for each county for each type of office, Table A2.
- `fgC5_county_race_seg.dta`: Racial segregation indices for Figure C5.

3 List of What Produces What

3.1 Each Script

- `02_analysis_all_politics_partisan_main.do`: Main Figures 2, 3; Tables 1, 2, 4, 5, 6; Appendix Figures B1, C2, C3; Appendix Tables B4.
- `02b_tbB5_alternative-metrics.do`: Makes Appendix Table B5.
- `03_tb3_rsq_regressions.do`: Makes Table 3.
- `04_tb7_local-education_structure.do`: Makes Table 7.
- `05_fg4_gallup_trust.do`: Makes Figure 4.
- `06_fg1_tbA2_sample-size_by-state.R`: [R script] Makes Figure 1, Tables A2, A3.
- `07_tbA1_counties-demographics.R`: [R script] makes Table A1.
- `08_tbC6_michigan-party-lever.R`: [R script] Makes Table C6.
- `09_fgC5_apg_race.do`: Makes Figure C5.
- `10_tbC7_stacked-regressions.do`: Makes Table C7.

There is no 01 script in this dataverse because these were preprocessing steps not relevant for this article.

3.2 Each Table and Figure

Here is a short list of where replicators will find the code needed to replicate all parts of the publication. Script names are abbreviated by their numbers.

Main Text:

- **Figure 1**: State of counts: Script `06_....R`
- **Table 1**: Absolute Partisan Gap: `02_....do`
- **Table 2**: Roll-Off: `02_....do`

- **Figure 2**: Histogram of Absolute Partisan Gap: `02_....do`
- **Figure 3**: APG vs. Democratic Percent: `02_....do`
- **Table 3**: Types of Local Offices: `02_....do`
- **Table 4**: Ideological and Spending-related Ballot Measures: `02_....do`
- **Table 5**: Issue Domains of Ballot Measures: `02_....do`
- **Table 6**: Within-Contest R-squareds from Regression: `03_....do`
- **Table 7**: Local Education Ballot Measures: `04_....do`
- **Figure 4**: Trust in Government: `05_....do`

Appendix:

- **Table A.1**: Sample Size of Voters by State and Office: `06_....R`
- **Table A.2**: Sample Size of Contests by State and Office: `06_....R`
- **Table A.3**: Census Characteristics of Counties Used: `07_....R`
- **Figure B.1**: Histogram of APG *as Odds Ratios*: `02_....do`
- **Table B.4**: Voting Patterns by Partisanship, Log Odds Measure: `02_....do`
- **Table B.5**: Other Measures Comparing Democratic and Republican Voters: `02b_....do`
- **Figure C.2**: APG vs. Margin of Victory: `02_....do`
- **Figure C.3**: Rolloff among Republicans and Democrats: `02_....do`
- **Table C.6**: Michigan Party Lever: `08_....do`
- **Figure C.4**: Absolute Partisan Gap vs. Number of Voters: `02_....do`
- **Figure C.5**: Racial Segregation: `09.....do`
- **Table C.7**: Stacked Regression: `10_....do`

4 Other Specifications

4.1 Hardware Specifications

All estimates were replicated by the authors on a Mac, but the code should run on any OS.

4.2 Software Dependencies

Estimates were reproduced on Stata 18.5 MP, and R 4.4.1.

Stata Scripts One custom Stata script that needs to be installed is `_gwtmean` for Script 03:

```
ssc install _gwtmean
```

R Scripts The R software dependencies are standard:

```
# Reading Stata files for Tables A1-A3, Figure C6
library(haven)

# shapefile map packages for Figure 1
library(urbnmapr)
library(sf)

# Tables
library(gt)
library(tidyverse)
```

`urbnmapr` is not on CRAN, and can be installed by Github via:

```
remotes::install_github("UrbanInstitute/urbnmapr")
```

Otherwise, the packages can be installed by CRAN.

4.3 Runtimes

`04_local-education_structure.do` takes about 2-3 minutes to run. All other scripts take less than 1 minute to run for each. Many scripts represents millions of observations, but the regressions are reprogrammed so that they run on sufficient statistics (i.e., group means).

4.4 Random Seeds

No random seeds are necessary.