Replication Codebook and Commentary

J. Andrew Sinclair
Ian O'Grady
Bryn Miller
Catherine Murphy

"Participation and Competition in Top-Two Elections: Tradeoffs in Election Reform"

for

State Politics and Policy Quarterly, 2025

Version August 1, 2025

For questions:
asinclair@cmc.edu

This file provides the replication information for our paper. The included replication materials are:

1. Sinclair_etal_2025_SPPQ_README.txt (Plain Text Read Me).
2. Sinclair_etal_2025_SPPQ_Replication_Commentary.pdf. (This document.)
3. Sinclair_etal_2025_SPPQ_ReplicationFile_Final.do (The Stata DO file.)
4. TeamPaper_RegularVersion_2025_0613_replicationset_FINAL.dta (The main data.)
5. Search_AdrinNazarian.xlsx (Supplemental Google Search Data.)

This document provides a guide for carrying out the replication (and may be a bit easier to read than a plain-text "README").

The Complete Code

We used Stata 18 for Windows to prepare the replication materials. You will need access to at least Stata 18 to run the .do file, although it should run in later versions of Stata as well. Once you have downloaded the files (above) and modified the file paths, the findings should replicate without any further action. Graphs have .pdf and, for Windows users, also .emf file formats.

Modifying the File Paths

As the header of the .do file indicates, we had organized the project with separate folders for the code, data, and paper drafts. The "SPPQ Replication Final" folder was a subfolder of our draft folder. For the replication, everything came from and went into that folder for simplicity. You can create file folders that mirror this structure or modify the file paths in the code to suit your own design.

You must modify line 67 to set the file path for your computer, and to make sure the rest of the filepaths (the log, data, and table/figure outputs) match your own file folders.

Installing Packages

The code is set up to install, if you do not have them already, three packages: scheme-burd, estout, and labutil.

Organization of the Do-file

After a few preliminary steps to get everything set up, the file makes a few modifications to the data, produces the tables and figures for the main paper, and then produces everything for the appendix.

```
/* These Comments Indicate New Sections of the Code */

* These comments indicate the next major item

** These comments are sub-items
```

The Data

The data for this project is mostly just election records from the California Secretary of State's office, available online. One figure in the appendix also uses Google Trends data.

The Main Elections Data

Each row in the main dataset represents a *candidate-office-year*: for example, Adrin Nazarian, running for the CA Assembly in 2012. Candidates were included in the data if they appeared in those records, no matter how few votes they obtained.[1] In total this includes 5,470 CA Assembly, CA Senate, and U.S. House candidates from 2002-2020, of whom 1,530 won their elections and 3,940 lost.

The main data is included in Stata format (.dta).

In the main data, the key variables are:

- "elecyear" Election year
- "level" Level of office (1=Assembly, 2=State Senate, 3=US House)
- "distnum" District number
- "name" Candidate name
- "incumb" Marked as an incumbent
- "ctst_wi_finalist" Contest has a write-in finalist
- "ctst_wi_finalist_majorparty" As above, but from the largest party
- "party" Candidate party
- "pvoten" Number of primary votes
- "pctpvote" Percent of primary election vote
- "pre_pctpvote" Percent of primary election vote within party[2]
- "porder" Rank order of candidates by primary votes (1=most votes)
- "gvoten" Number of general votes
- "pctgvote" Percent of general election vote
- "gorder" Rank order of candidates by general votes (1=most votes)
- "gwin" Identifies general election winner (=0 not, =1 winner)
- "wind" Democratic general election winner (=1; =0 otherwise)

[1] A personal favorite: Joseph Napolitano, a 2018 write-in U.S. House candidate for CD8. Napolitano was an official write-in candidate for CD8, and thus appeared in the SOV. Yet, he apparently did not vote for himself, and obtained zero votes. For confirmation that he was certified, see: https://elections.cdn.sos.ca.gov/ccrov/pdf/2018/may/18137wk.pdf. For confirmation that he obtained zero votes, see: https://elections.cdn.sos.ca.gov/sov/2018-primary/sov/82-congress.pdf (p. 85).

[2] Note that this lumps all third-party candidates together, even if they did not actually compete against each other under the old rules.

- "copartisan" Indicates a same-party general election (=1 if yes; =0 if no)
- "ptotvotes" All primary votes in the contest
- "pre_ptotvotes" All primary votes in the *partisan* primary of the candidate's party
- "gtotvotes" All general election votes in the contest
- "ctst_dcands" Number of Democratic Party candidates in contest
- "ncands" Number of all candidates in contest
- "rpvs" Republican presidential votes share (2008 for 02-10, 2012 for 12-20)
- "safeDcutoff" Considered safe Democratic (=1) or not (=0)
- "totalreg" Total number of registered voters in the district
- "topticket_vote" Total votes cast for major candidates
- "topticket_loss_percent" Decrease in participation
- "topticket_rep_twoparty" Republican % of the top-ticket two-party vote
- "ofreg_pctpvote" Primary vote as a percent of total registration
- "ofreg_pctgvote" General vote as a percent of total registration
- "cats_pre_dprimary" Democratic vote among Democratic primary candidates (bins)[3]
- "pseudorule" Election configuration categories (a convenience variable)
- "ctst_gvote_ofrv" All general election votes out of all RVs
- "ctst_prepvote_ofrv" All votes in winner's primary[4] out of total RVs

A few of these variables are produced in the replication code, since they use information available in other variables and involve cutoffs that someone might wish to modify (for example, safeDcutoff, cats_pre_dprimary).

In addition, the dataset includes the following variables, most of which have been left in the dataset for the sake of convenience—including alternative labels or helping users connect this data to other data sets of their own:

- "state" State (here, all in CA, but used for display purposes)
- "CA" A California indicator (left here for display purposes)
- "ca" A different California indicator (again, for display purposes)
- "district" Equivalent to the "distnum" variable, but kept for convenience
- "period" =1 if 2002-2010; =2 if 2012-2020
- "distperiod" Equivalent to "period" with different text labels for convenience
- "toptwo" Identifies which period used top-two rules (=1 if toptwo, 0 otherwise)
- "legrace" =1 for districted legislative seats (state leg. or US House).

[3] This is sorted into categories and then displayed in Figure 4. In 2012's AD50, Richard Bloom had 25.6 of the total primary vote out of a field with three Democrats and one Republican, roughly split evenly. That gives him "under 45%" because he is at about 1/3 of the *Democratic* primary vote. In 2014, he gets 100% of the *Democratic* vote by winning 73.4% of the primary vote in a field only including one Republican.

[4] Winner's primary here is the party primary (before the top-two) or among all of the candidates in the primary (top-two era).

Google Trends Data

One figure in the supplemental appendix also uses Google Trends data. This is freely available and can be downloaded as an xlsx file: https://trends.google.com/trends/. The data included here preserves the result of our original search from August 2023, in case Google changes its formula or makes the data unavailable later.

Note that you must use the same time frame to access the same data, since Google will set peak search for that term during the required period to 100, and produce values that are relative to peak search.

The variables are:

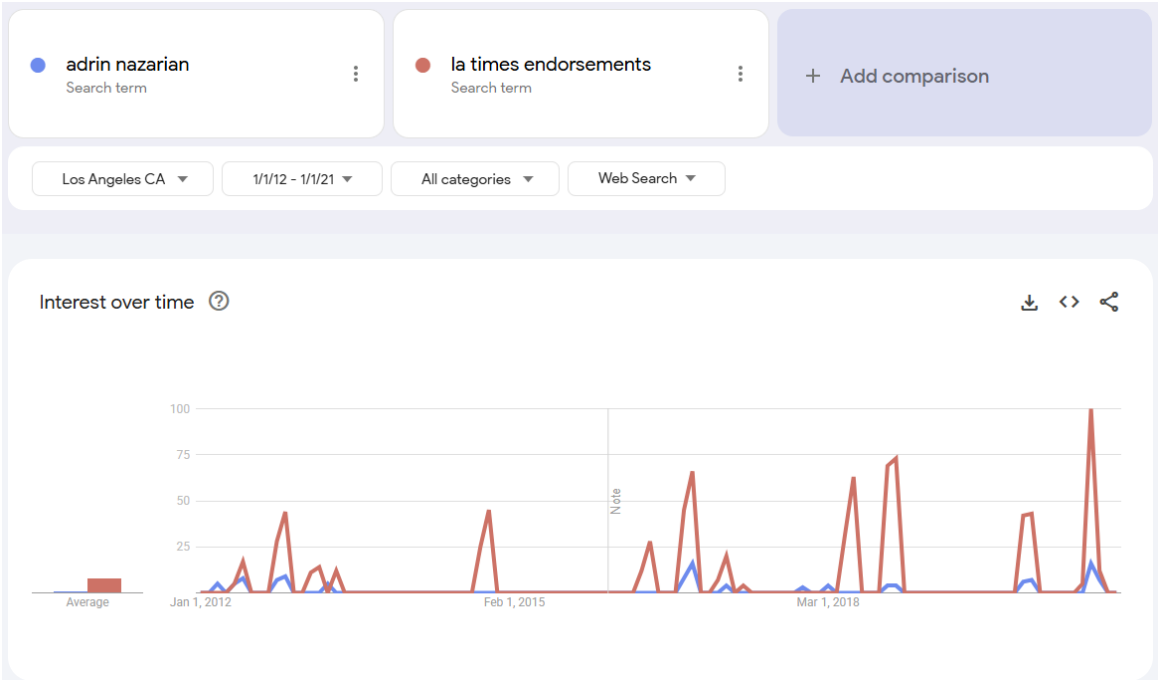
- "Date" This is the month included in the downloaded Google Trends data
- "Year" This is the year, disaggregated from the full date
- "Month" This is the month, disaggregated from the full date
- "oid" This is an observation identification number
- "Election" This labels the month and type of each election
- "Search_AdrinNazarian" This is the reported relative search

The replicability of Google Trends data over time can be a problem for several reasons. Those reasons can include: Google samples from its data, instead of pulling information from all of it, and Google sometimes makes poorly documented changes to what it reports. See:

Cebrian, Eduardo and Josep Domenech. 2024. "Addressing Google Trends inconsistencies." Technological Forecasting and Social Change, 202 (123318): 1-9. https://doi.org/10.1016/j.techfore.2024.123318.

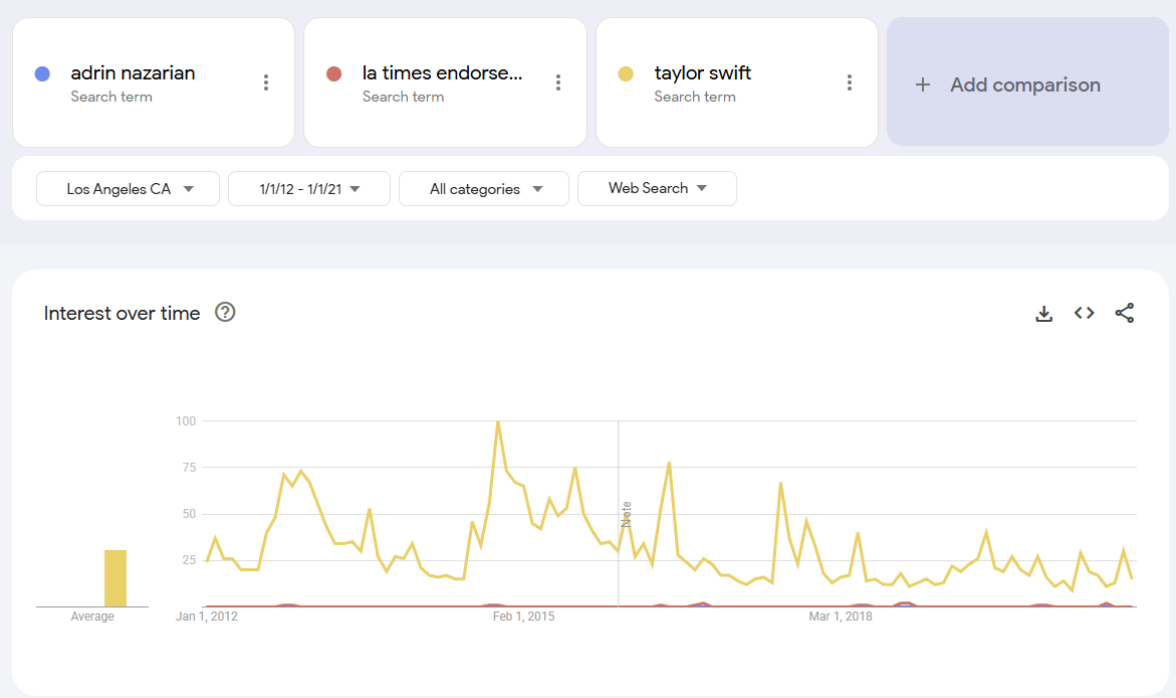
The data in the paper is from the August 2023 search. One reason that it may be somewhat unstable is that there are just not very many searches, so the data Google retrieves will include more noise. Google only provides *relative* search volume, not *total* search. This can be visualized by comparing one search term to another.

Below is a screenshot from June 2025 of the same search request, but in comparison to the more generic "la times endorsements." You can see that search for Adrin Nazarian peaks at roughly even levels in 2016 and 2020, with slightly less (in two bumps) in 2012—and with almost no search at all in 2014 or 2018. In the years of heightened search, you can see that it is occurring during the periods in which there is also search for the *LA Times* endorsements—so this looks to be capturing real behavior, if not very much of it.



https://trends.google.com/trends/explore?date=2012-01-01%202021-01-01&geo=US-CA-803&q=adrin%20nazarian,la%20times%20endorsements

To get a sense of the scale of elections-related searching, it's also helpful to have something as a point of comparison that is not elections-related. Below, for example, you can see the comparison between these search terms and "taylor swift." Swift searches obliterate the elections-related searches, which barely even register in relative volume.



Successful replication of capturing the Google Trends data, then, should show the most "Adrin Nazarian" search activity in 2016 and 2020, the elections with the same-party opponents, and the least in 2014 and 2018.

That there is not enough search activity to produce precisely replicable search results demonstrates what is in Figure 5 of the main paper—that most people are just picking someone with the same party label, and many are not searching or evaluating at all.

First Data Example: CA Assembly District 50, 2010

The main text of the paper shows results from AD50 for 2012 (Table 1). Let's look at 2010 for the same district number (recall: not the same district boundaries, though).

The California Secretary of State's office publishes a "Statement of Vote" (SOV), the official record of the outcome of an election. For each election, there will be both a primary and general election SOV. Most of the data in the paper is coming directly from the SOVs.

These totals may differ from totals reported on election night or shortly after the election in newspapers. For some other commonly-referenced online sources (the Wikipedia or Ballotpedia entries), occasionally earlier-reported numbers may be used—those may be similar, if not precisely identical, to what we have here. It can take a long time for the official SOV to be published after the election.

From the 2010 California Statement of Vote, in the primary:

50th Assembly District

	Ricardo Lara DEM	Art Olivier DEM	Carmen Avalos DEM	Luis Marquez DEM	P.J. Mellana REP	Gladys O. Miller REP
Los Angeles	6,314	1,695	4,608	2,174	1,805	2,828
District Totals	6,314	1,695	4,608	2,174	1,805	2,828
Percent	42.7%	11.4%	31.2%	14.7%	38.9%	61.1%

https://elections.cdn.sos.ca.gov/sov/2010-primary/pdf/101-118-ad.pdf

From the 2010 California Statement of Vote, in the general:

50th Assembly District

	Ricardo Lara DEM	Gladys O. Miller REP
Los Angeles	46,676	13,452
District Totals	46,676	13,452
Percent	77.7%	22.3%

https://elections.cdn.sos.ca.gov/sov/2010-general/73-state-assembly.pdf

You can see here that Ricardo Lara, the ultimate winner, had 6,314 votes in the primary (the variable "pvoten" in our data, for "number of primary votes") and 46,676 votes in the general election (the variable "gvoten" in our data).

Lara turned those 6,314 primary votes in 2010 into a career including the State Senate (2012, 2016) and a statewide office, Insurance Commissioner (2018, 2022).

From the last general election 15-day registration report:

Report of Registration as of October 18, 2010
Registration by State Assembly District

	Total Registered	Democratic	Republican	American Independent	Green
State Assembly 46					
Los Angeles	112,549	72,299	12,547	1,982	642
District Total	112,549	72,299	12,547	1,982	642
Percent		**64.24%**	**11.15%**	**1.76%**	**0.57%**
State Assembly 47					
Los Angeles	234,787	151,193	26,557	4,091	1,326
District Total	234,787	151,193	26,557	4,091	1,326
Percent		**64.40%**	**11.31%**	**1.74%**	**0.56%**
State Assembly 48					
Los Angeles	150,892	104,898	13,383	2,552	560
District Total	150,892	104,898	13,383	2,552	560
Percent		**69.52%**	**8.87%**	**1.69%**	**0.37%**
State Assembly 49					
Los Angeles	161,087	73,922	36,048	2,883	612
District Total	161,087	73,922	36,048	2,883	612
Percent		**45.89%**	**22.38%**	**1.79%**	**0.38%**
State Assembly 50					
Los Angeles	144,764	87,945	24,429	2,659	503
District Total	144,764	87,945	24,429	2,659	503
Percent		**60.75%**	**16.88%**	**1.84%**	**0.35%**

https://elections.cdn.sos.ca.gov/ror/ror-pages/15day-gen-10/assembly.pdf

In the total registration column for California State Assembly District 50, which was wholly contained in Los Angeles County, you can see 144,764 registered voters. This is the number used for total registration (variable "totalreg"). These registration reports can be found on the Secretary of State's website. We have attempted to use the last registration report before the general election as our source for registration numbers. Typically, these do not change very much between registration reports, but other datasets may pull from different registration reports and report slightly different numbers.

Note that in Figure 3 in the paper, we look at the winning Democrat's vote share facing the "last Democrat." For Lara in 2010, that's the primary election—which Lara won with 6,314 votes. So the winner's vote out of total registered voters is 6,314 in 144,764 = 4.4% (variable "ofreg_pctpvote"). Since winning the primary was effectively winning the election here, Lara did not need a very large amount of support among the registered voters to do so.

Table 1 and Figure 2 both include roll-off estimates.

The 2010 general election "top of the ticket" vote, also reported in the SOV:

Supplement to the Statement of Vote
Counties by Assembly Districts for Governor

	Jerry Brown DEM	Meg Whitman REP	Chelene Nightingale AI	Laura Wells GRN	Dale F. Ogden LIB	Carlos Alvarez PF
State Assembly District 50						
Los Angeles	46,636	13,773	858	736	781	1,313
District Totals	46,636	13,773	858	736	781	1,313
Percent	72.8%	21.5%	1.3%	1.1%	1.2%	2.0%

https://elections.cdn.sos.ca.gov/sov/2010-general/ssov/governor-assembly.pdf

If you add across the row, that is 64,097 votes cast for candidates listed in the Statement of Vote, the number used for the "top of the ticket" calculation in this district in 2010 (the variable "topticket_vote").

And, in terms of the roll-off: there were 60,128 votes cast in the general election for Lara and Miller combined, with Lara getting most of them—and, in fact, more votes than Brown did: 46,676. Yet, if we sum the top row here, there were 64,097 votes cast overall in the gubernatorial election in this district, including 1,313 for a Peace & Freedom Party candidate and 736 for a Green Party candidate (left-leaning parties).

The roll-off computation compares the total votes for the Assembly contest (60,128) to the total votes for the gubernatorial contest (64,097).

$$[\ (64{,}097-60{,}128) / 64{,}097\ ] \times 100 = 6.2$$

This is the "topticket_loss_percent" variable.

Figure 2 only uses the 2012-2020 elections, but a 6.2% roll-off would be fairly normal for cross-party elections in that period, too.

To look at the top-ticket *Democratic* votes out of the two highest-voting getting candidates—done this way to get something comparable to the top-two rules (before 2012, or in the top-two era's presidential elections)—take Brown's percent out of Brown+Whitman:

$$[\ 46{,}636 / (46{,}636+13{,}773)\ ] \times 100 = 77.2$$

This is the "topticket_dem_twoparty" variable.

In addition, Figure 1 requires presidential vote share, as do Figures 2 and 3.

Presidential vote share: for 2010, the previous election with these district boundaries was in 2008:

Supplement to the Statement of Vote
Counties by Assembly Districts for United States President

	Barack Obama DEM	John McCain REP	Alan Keyes AI	Cynthia McKinney GRN	Bob Barr LIB	Ralph Nader PF
State Assembly District 49						
Los Angeles	72,338	36,306	294	400	449	966
District Totals	72,338	36,306	294	400	449	966
Percent	65.3%	32.8%	0.3%	0.4%	0.4%	0.9%
State Assembly District 50						
Los Angeles	74,989	20,523	314	615	387	693
District Totals	74,989	20,523	314	615	387	693
Percent	76.9%	21.0%	0.3%	0.6%	0.4%	0.7%

https://elections.cdn.sos.ca.gov/sov/2008-general/ssov/8-pres-by-assembly.pdf

There are several possible approaches for computing presidential vote share for each district. For our purposes, it seemed reasonable to use 2008 for the period between 2002-2010 and to use 2012 for the period between 2012-2020: that meant they were both Obama victories, keeping some things relatively similar.

For AD50 (2002-2010), then, the Republican presidential vote share (variable "rpvs") is the two-party share:

$$[\ 20{,}523 / (20{,}523 + 74{,}989)\ ] \times 100 = 21\%$$

We use *Republican* presidential vote share so that districts which are *more Republican* appear on the right side of the graph in Figure 1. In Figure 2, we are looking at the difference between Democratic two-party presidential vote share (simply 100-rpvs) and the Democratic legislative candidate's performance. In this case Lara's approximately 78% of the vote is very similar to Obama's approximately 79% of the vote.

The point is that in the top-two era, districts like this sometimes end up with candidates getting closer to 50% of the vote in these types of districts.

And, now, AD50 in 2010 as an entry in the dataset:

```
list elecyear dpvs name party pvoten pctpvote gvoten pctgvote ///
        topticket_dem_twoparty gtotvotes topticket_vote topticket_loss_percent ///
        if CA==1 & level==1 & distnum==50 & elecyear==2010
```

elecyear	dpvs	name	party	pvoten	pctpvote	gvoten	pctgvote	t~dem_~y	gtotvo~s	toptic~e	toptic~t
2010	79	Ricardo Lara	DEM.	6314	32.5	46676	77.6	77.2	60128	64097	6.2
2010	79	Gladys O. Miller	REP.	2828	14.6	13452	22.4	77.2	60128	64097	6.2
2010	79	Carmen Avalos	DEM.	4608	23.7	.	.	77.2	60128	64097	6.2
2010	79	Luis Marquez	DEM.	2174	11.2	.	.	77.2	60128	64097	6.2
2010	79	P.J. Mellana	REP.	1805	9.3	.	.	77.2	60128	64097	6.2
2010	79	Art Olivier	DEM.	1695	8.7	.	.	77.2	60128	64097	6.2

```
list elecyear dpvs name party pvoten pctpvote gvoten pctgvote ///
        rpvs totalreg ofreg_pctpvote ofreg_pctgvote ///
        if CA==1 & level==1 & distnum==50 & elecyear==2010
```

elecyear	dpvs	name	party	pvoten	pctpvote	gvoten	pctgvote	rpvs	totalreg	of~pvote	of~gvote
2010	79	Ricardo Lara	DEM.	6314	32.5	46676	77.6	21	144764	4.4	32.2
2010	79	Gladys O. Miller	REP.	2828	14.6	13452	22.4	21	144764	2	9.3
2010	79	Carmen Avalos	DEM.	4608	23.7	.	.	21	144764	3.2	.
2010	79	Luis Marquez	DEM.	2174	11.2	.	.	21	144764	1.5	.
2010	79	P.J. Mellana	REP.	1805	9.3	.	.	21	144764	1.2	.
2010	79	Art Olivier	DEM.	1695	8.7	.	.	21	144764	1.2	.

Second Data Example: CA CD31, 2012 & Related

Congressional District 31 in 2012 is an oft-referenced California election for the top-two era, notable for resulting in a Republican-on-Republican general election in a district President Obama won. While this is a very rare occurrence, as clearly demonstrated in Figure 1 of the paper, it is an interesting case.

The 2012 primary data is here:

United States Representative

31st Congressional District

	Pete Aguilar DEM	Justin Kim DEM	Rita Ramirez-Dean DEM	Renea Wickman DEM	Bob Dutton REP	Gary G. Miller* REP
San Bernardino	14,181	8,487	3,546	4,188	15,557	16,708
District Totals	14,181	8,487	3,546	4,188	15,557	16,708
Percent	22.6%	13.5%	5.7%	6.7%	24.8%	26.7%

https://www.sos.ca.gov/elections/prior-elections/statewide-election-results/presidential-primary-election-june-5-2012/statement-vote

Gary Miller is marked as an incumbent (the *); he had represented the old CD31. Bob Dutton was a member of the California State Senate, but not an incumbent U.S. House member.

The 2012 general data:

31st Congressional District

	Bob Dutton REP	Gary G. Miller* REP
San Bernardino	72,255	88,964
District Totals	72,255	88,964
Percent	44.8%	55.2%

https://www.sos.ca.gov/elections/prior-elections/statewide-election-results/general-election-november-6-2012/statement-vote

And here is the presidential result:

Supplement to the Statement of Vote
Counties by Congressional Districts for President

	Barack Obama* DEM	Mitt Romney REP	Thomas Hoefling AI	Jill Stein GRN	Gary Johnson LIB	Roseanne Barr PF
Congressional District 29						
Los Angeles	129,323	34,454	681	1,121	1,466	844
District Totals	129,323	34,454	681	1,121	1,466	844
Percent	77.0%	20.5%	0.4%	0.7%	0.9%	0.5%
Congressional District 30						
Los Angeles	185,773	91,116	888	1,946	3,298	1,095
Ventura	528	564	1	7	9	1
District Totals	186,301	91,680	889	1,953	3,307	1,096
Percent	65.3%	32.1%	0.3%	0.7%	1.2%	0.4%
Congressional District 31						
San Bernardino	118,043	83,822	721	807	2,027	822
District Totals	118,043	83,822	721	807	2,027	822
Percent	57.2%	40.6%	0.3%	0.4%	1.0%	0.4%

https://elections.cdn.sos.ca.gov/sov/2012-general/ssov/pres-by-congress.pdf

That is 206,242 presidential election votes. As you can see, above, there were only 161,219 U.S. House votes cast in the general election for the combined Miller and Dutton. That is, as represented below, a 21.8% roll-off (the computation in the "topticket_loss_percent" variable).

Miller-Dutton from our data (in part):

name	party	pvoten	gvoten	incumb	rpvs	totalreg	of~pvote	of~gvote	toptic~t
Gary G. Miller	REP.	16708	88964	1	42	307575	5.4	28.9	21.8
Bob Dutton	REP.	15557	72255	0	42	307575	5.1	23.5	21.8
Pete Aguilar	DEM.	14181	.	0	42	307575	4.6	.	21.8
Justin Kim	DEM.	8487	.	0	42	307575	2.8	.	21.8
Renea Wickman	DEM.	4188	.	0	42	307575	1.4	.	21.8
Rita Ramirez-Dean	DEM.	3546	.	0	42	307575	1.2	.	21.8

Since the cutoff we used for "safe Democratic seat" was a Republican presidential vote share (rpvs) below 40, this election is not included (just on the other side of the boundary) in the "safe seat" calculations.

How competitive was it? With a more typical cross-party election, with an RPVS of 42, we would expect the winner to be a Democrat and somewhere close to 58% of the vote. Miller wins here with 88,964 out of 161,219 votes, or 55%.

So, in the example often cited as the most pathological case for the top-two election system, a large roll-off is associated with a very modest competitiveness improvement. AD50 in 2012, Table 1 in the paper, is a more favorable outcome from the perspective of advocates for the top-two.

Yet, the paper is making an additional argument here. The election does not take place within the same district boundaries as the old CD42, Miller's previous district, but it remains notable that he won an election with 62% of the vote in 2010, an easy victory in a safe seat. Three Republicans ran against him in the 2010 primary, and Miller won handily with 48.8% of the primary vote—32,669 votes.

42nd Congressional District

	Michael Williamson DEM	Gary G. Miller* REP	Phil Liberatore REP	Lee McGroarty REP	David Su REP	Mark Lambert LIB
Los Angeles	4,708	5,291	3,143	660	524	26
Orange	14,472	22,726	18,724	4,971	1,248	294
San Bernardino	4,186	4,652	3,314	1,482	269	51
District Totals	23,366	32,669	25,181	7,113	2,041	371
Percent	100.0%	48.8%	37.6%	10.6%	3.0%	100.0%

https://elections.cdn.sos.ca.gov/sov/2010-primary/pdf/85-95-cd.pdf

While the paper computes this for the "last Democrat" elections (usually the more relevant case in California), you can see the top-two effect in the "last Republican" case here. In 2010 Miller was effectively elected with 32,669 primary votes, and could have won with 25,182. In 2012, facing another Republican in the general election, Miller won 88,964 and needed at least 72,256.

Here is Miller's career, in the dataset:

elecyear	level	distnum	rpvs	name	party	incumb	pvoten	pctpvote	pre_pc~e	gvoten	pctgvote	totalreg
2012	US House	31	42	Gary G. Miller	REP.	Incumbent	16708	26.7	51.8	88964	55.2	307575
2010	US House	42	54	Gary G. Miller	REP.	Incumbent	32669	36	48.8	127161	62.2	373854
2008	US House	42	54	Gary G. Miller	REP.	Incumbent	39168	67.7	100	158404	60.2	372054
2006	US House	42	54	Gary Miller	REP.	Incumbent	43813	99.4	100	129720	100	347162
2004	US House	42	54	Gary Miller	REP.	Incumbent	65558	68.9	100	167632	68.1	348708
2002	US House	42	54	Gary Miller	REP.	Incumbent	59065	69.9	100	98476	67.8	310877

Note that Miller's percent of the *total* primary vote was only 36.0% in 2010, although within the Republican primary was 48.8%.

Miller retired after 2012. Pete Aguilar, the losing Democrat in the 2012 primary, would fight his way through to win in 2014. By 2025 he would become the Chair of the House Democratic Caucus.

Other Data Notes

In addition to spot-checking the data for errors, we looked at some notable outliers for primary and general election vote totals at each level.

State Assembly Outliers

Yes, Buffy Wicks did get 135,623 votes in the 2020 Assembly primary for AD15 and 204,108 votes in the general election. Her primary total is much larger than the average election winner's primary vote for 2002-2020—but *every* winning Assembly candidate with more than 100,000 primary votes obtained that total in either the 2016 or 2020 primary election: high turnout affairs in the top-two era. Not all were Democrats.

Republican Kevin Kiley (in AD6) had 104,412 primary votes in 2020. This was a preview of the general election: there were only two candidates, and Kiley ran ahead of the Democrat 104,412-75,557 in the primary on his way to winning the general election 178,559-124,294. Kiley would move up to the US House in the 2022 election (outside the data for this paper).

In the other direction, compare Patty Lopez in AD39 from 2014 (RPVS=24) with John Perez in AD46 from 2008 (RPVS=14). Lopez won 4,940 primary votes—the second fewest by an ultimate general election winner in 2002-2020. Perez won the fewest: 4,905.

- Perez: The 4,905 votes came in a 4-way divided Democratic primary field, although he won easily; this came to 54.3%, with the three other challengers (including Ricardo Lara! See Data Example 1, above.) splitting the remaining vote. The story is that the other candidates agreed, after a meeting, not to really run.[5]

- Lopez: Came in 2nd with 4,940 votes (23.6%) to incumbent Democrat Raul Bocanegra (62.5%) in a field of four candidates. This created a Democrat-on-Democrat general election, which Lopez narrowly won—a rare hostile takeover by a challenger against an incumbent. Bocanegra was back in 2016 to reclaim the seat, and forced Lopez out again—only to be forced to resign himself after being accused of sexual misconduct.[6]

[5] This appears to have originated in the now-defunct LA Weekly: https://web.archive.org/web/20100608022602/https://www.laweekly.com/2010-06-03/news/the-chosen-one/.
[6] See here: https://tinyurl.com/2tj7493t

	46th Assembly District				
	John A. Perez	Arturo Chavez	Michael Aldapa	Ricardo Lara	Manuel "Manny" Aldana Jr.
	DEM	DEM	DEM	DEM	REP
Los Angeles	4,905	1,507	1,304	1,326	1,084
District Totals	4,905	1,507	1,304	1,326	1,084
Percent	54.3%	16.7%	14.4%	14.6%	100.0%

https://elections.cdn.sos.ca.gov/sov/2008-statewide-direct-primary/stassem08primary.pdf

	State Assemblymember 39th Assembly District			
	Raul Bocanegra* DEM	Patty Lopez DEM	Kevin J. Suscavage DEM	Michael B. Boyd REP (W/I)
Los Angeles	13,069	4,940	2,876	36
District Totals	13,069	4,940	2,876	36
Percent	62.5%	23.6%	13.7%	0.2%

https://elections.cdn.sos.ca.gov/sov/2014-primary/pdf/84-state-assemblymember.pdf

In any event, neither are errors in the data, as you can see above.

In the general election, the winner with the fewest votes was Miguel Santiago, the winner of a Democrat-on-Democrat general election in AD53 (RPVS=13) in 2014. Santiago defeated Sandra Mendoza 20,472 to 11,753. That made it marginally more competitive than it would have been as a cross-party election, but not a particularly close election.

State Senate Outliers

Nancy Skinner indeed won 263,751 votes in the SD9 primary in 2020—50,000 primary votes more than any other State Senate winner during this period. Her opponent? She only faced write-in candidate Jamie Dluzak, a Libertarian who garnered 126 votes. Our results suggest that Dluzak might have done better to run as a Democrat, although by making it to the top-two Dluzak still got 53,246 votes in the general election (good for almost 12%). Skinner had 404,455, though—the highest general election total in the 2002-2020 period.

The lowest State Senate primary total came in 2014 in SD20. In a Democratic-leaning district, one Republican and four Democrats ran in the common primary. The Republican advanced with 33% of the vote (14,124 votes). Democrat Connie M. Leyva advanced after placing second—but, of course, ahead of all of the other Democrats—with 9,096 votes. (The next closest Democrat had 7,958). It was a low-turnout general election, with Leyva winning easily (62.4%, 56,943 votes).

The lowest State Senate general election total also came in 2014—broadly, a bad year for turnout, with not much going on statewide to draw in interest. Republican Andy Vidak managed to defeat Democratic Luis Chavez 54,251-46,035 in November. Note that this falls just on the other side of our "safe Democratic seat" threshold (RPVS=41).

The lowest State Senate total in a general election for the 2002-2010 period, for context, was Democrat Lou Correa winning SD34 in 2006. This district also falls just outside the "safe seat" threshold (RPVS=41, also), and the election was competitive. Correa got 56,534 to the Republican candidate's 55,142.

U.S. House Outliers

Barbara Lee did, in fact, get 230,482 primary votes in CD13 in 2020—the only House member to get over 200,000 in this period. It is a good example of how lopsided some of these Bay Area districts can be: in 2012, the Republican two-party vote share was only 9%. There *was* a Republican opponent in this two-candidate race, Nikka Piterman, but he only won 7.4% in the primary and 9.6% in the general election.

A Green Party candidate had done hardly better against Lee in 2018, and Lee did not face a Democrat in top-two era analyzed here (2012-2020). Lee was, famously, the only House member to vote against the Bush-era authorization of force after 9/11. She ran for U.S. Senate in 2024 but did not make the general election, amid some controversy about whether the ultimate winner, Adam Schiff, had supported a Republican in the primary[7] to help ensure that he would face the Republican in the general election.

The House winner with the lowest primary vote total was Pete Aguilar—the loser of the Miller-Dutton primary (Data Example #2 for 2012), trying again in 2014. A crowded Democratic field nearly split the district again! But Aguilar prevailed over the third-place candidate by about 200 votes, and then won a very close general election (51.7%).

The lowest general election House total came in 2014 in CD35 in a Democrat-on-Democrat race. Norma Torres defeated another Democrat 39,502-22,753. This was not a consequence of massive roll-off in the district as much as it was low turnout overall. The gubernatorial candidates combined for 68,438 votes in this district relative to the 62,255 votes in the House contest. Turnout in 2014, as many of these examples illustrate, was very low.

Odds and Ends

In 2002, the PDF version of the Statement of Vote for AD10 has Alan Nakanishi with 22,549 votes. The Excel version, provided adjacent to the PDF version on the Secretary of State's website, has Alan Nakanishi with 22,599 votes. We have used the 22,599 number—and clearly of no consequence one way or the other.

[7] See here: https://www.sacbee.com/opinion/article285944706.html.

Our totals also occasionally differ from the reported results when there are potential procedural issues arising with write-in candidates. For example, in 2004, Pat Driscoll obtained 1,041 votes as the on-ballot Green Party primary candidate and 11 votes as a Democratic Party write-in candidate in CD5. We disregarded the 11 write-in votes.