

The Incumbency Advantage in the U.S. House, 1946-2024

A vote-share premium that rose for thirty years, peaked in the 1980s, and has nearly vanished since

Democracy's Data

Last updated September 22, 2026 at 10:13a EDT

Every American who has followed a House race has heard some version of the same claim: incumbents almost always win, and it used to be an even bigger advantage than it is now. Gary Jacobson spent much of a career pricing exactly that – how many points of the vote a House seat is worth once you already hold it, and how that price has moved.¹ His own method, run on his own data through 2014, and on this chapter's own construction from 2016 on, gives 34 separate yearly estimates across 40 elections:

¹ Gary C. Jacobson, "It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections," *Journal of Politics* 77(3)(2015): 861-873.

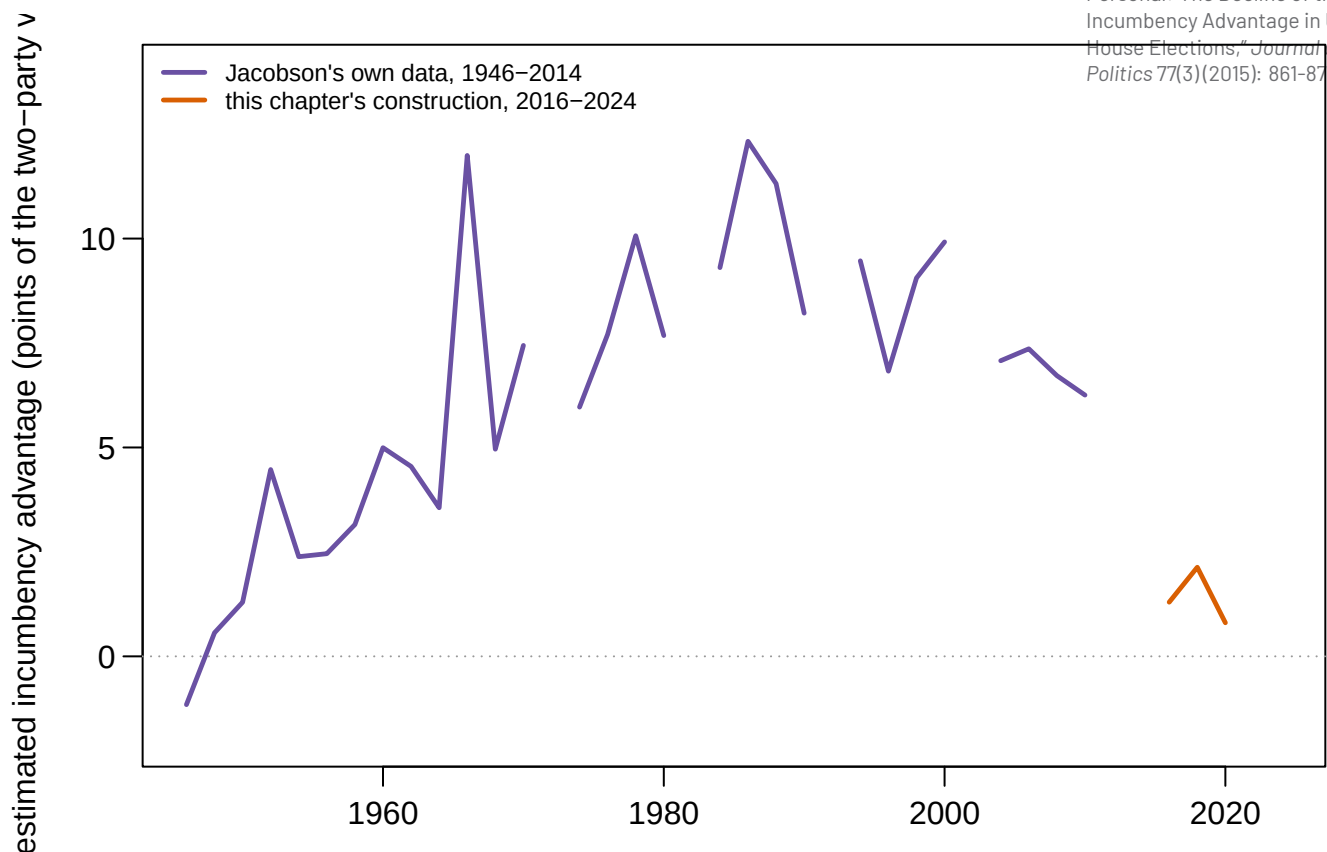


Figure 1. The estimated incumbency advantage, election by election, 1946-2024. Each point is a separate regression, one per year, of this election's Democratic two-party vote share on the same district's share two years earlier and an indicator for which party (if either) already held the seat; the indicator's coefficient is that year's estimate. Purple is Gary Jacobson's own data and his own documented method; orange is this chapter's

extension past where his file stops, built the same way from the MEDSL returns used throughout this book.

The advantage opens the series close to zero – -1.2 points in 1946, indistinguishable from no advantage at all – climbs through the 1950s and 60s, and is standing at a sustained 12.3 points by 1986, the high point of the whole series. Then it erodes, decade by decade, to 1.9 points by 2024: a House where having already won buys almost nothing.

01 The test

Two designs, not one. The lead measure above is the one Jacobson uses for his own Figure 1 – the Gelman-King regression, fit separately every election year, which needs no bandwidth and no restriction to close races because it adjusts for the previous vote share directly rather than by comparing only near-ties.² It uses every standard two-party race in a given year, which is what makes a clean *year-by-year* line possible at all – a bandwidth-restricted design would have too few close elections in any single year to say much.

The second design, built from scratch for this chapter and used from here on, sacrifices that yearly resolution for a cleaner causal claim. Two elections in the same district, two years apart: where the previous margin sits near 50-50, which party happened to cross the line is close to a coin flip, unrelated to anything about the district that would otherwise confound the comparison. That is David Lee's close-election design.³ It answers a narrower question than "does the incumbent win," and needs no name-matching at all – only which party crossed 50%. Pooled across 2016-2024, it gives essentially the same answer as the last points on Figure 1: see "What a held seat is worth."

Nowhere in the raw returns behind either design is a column called `incumbent`. Building one, and then asking what it is worth, is most of what the rest of this brief does.

02 The data

Two files, not one. Jacobson's file – his own replication data for the 2015 paper above – covers 1946-2014, carries a hand-built incumbency variable, and marks which *specific districts* were redrawn since the last election, not just which years.⁴ It is freely downloadable – no guestbook, no login – and fetched directly by this chapter's build script. It stops in 2014, so 2016-2024 is built the same way the rest of this chapter is: from the MEDSL candidate-level returns, 33,805 rows, one per candidate per party line per district-year.

The two constructions were checked against each other before being spliced, and they disagree more than raw vote totals ever do. For the 16 years both files cover, the underlying two-party vote shares agree almost exactly – but on which party's incumbent (if either) is running, they agree in only 90.7% of 4,895 compared races. California's 17th in 1994 is a clean example of why:

Year	Candidate	Party	Votes
1992	LEON E PANETTA	DEMOCRAT	151565
1992	BILL MCCAMPBELL	REPUBLICAN	49947

² Andrew Gelman and Gary King, "Estimating Incumbency Advantage Without Bias," *American Journal of Political Science* 34(4)(1990): 1142-1164.

³ David S. Lee, "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics* 142(2)(2008): 675-697, <https://doi.org/10.1016/j.jeconom.2007.05.004>; open copy at <https://www.princeton.edu/~davidlee/wp/RDrand.pdf>.

⁴ Gary Jacobson, 2015, "It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections" [replication data], <https://doi.org/10.7910/DVN/29559>, Harvard Dataverse, V1, file J0Prepfile1.tab.

Year	Candidate	Party	Votes
1992	MAUREEN SMITH	PEACE AND FREEDOM	4804
1992	JOHN D WILKES	LIBERTARIAN	4051
1994	SAM FARR	DEMOCRAT	87222
1994	BILL MCCAMPBELL	REPUBLICAN	74380
1994	E CRAIG COFFIN	GREEN	5591

Leon Panetta won the seat in 1992, then left mid-term to become Bill Clinton's OMB director; Sam Farr won the June 1993 *special* election to replace him, and was the sitting incumbent by the time this table's second row was cast. This file has no row for that special election at all – of the 90 special-election rows in the whole 1976-2024 file, none falls in an odd year, so a vacancy filled and resolved between regular general elections leaves no trace here. A construction built only from these returns reads Farr's 1994 race as an open seat, which understates the true incumbency effect a little in whatever years it happens. Jacobson's file, curated by hand against contemporary sources rather than assembled from returns alone, does not have this blind spot – which is why it is used for the whole 1946-2014 run instead of a version rebuilt from scratch here, and why Figure 1's two segments are drawn in different colors rather than joined as if they were measured the same way.

Building 2016-2024 from the raw returns, and the close-election design used for the rest of this brief, both start from the same table: one row per district-year, and three things that break a naive count of it before anything else can be computed.

Fusion voting. A candidate can appear on more than one party's line in the same race – 2,677 rows here carry `fusion_ticket = TRUE` – so a two-party share computed off the row marked DEMOCRAT alone undercounts anyone who also ran on, say, New York's Liberal line. Otis Pike's 1976 race in New York's 1st is a clean example:

Candidate	Party line	Votes on this line	Combined total
OTIS G PIKE	DEMOCRAT	129779	135528
OTIS G PIKE	LIBERAL	5749	135528
SALVATORE NICOSIA	REPUBLICAN	61671	61671
SETH C MORGAN	CONSERVATIVE	10269	10269

Read off the Democratic row alone, Pike drew 129,779 votes; his real total, the one voters actually gave him, is 135,528. Every candidate total in this brief is summed across party lines first, exactly as the codebook's own notes recommend.

The Minnesota trap. From 1975 to 1995, Minnesota Republicans ran under the party's legal state name, "Independent-Republican" – and the file records that string because it is what the ballot said.

District	Candidate	Party (as recorded)	Votes
1	TIMOTHY J PENNY	DEMOCRATIC-FARMER-LABOR	206369
1	TIMOTHY R DROOGSMA	INDEPENDENT-REPUBLICAN	72367

District	Candidate	Party (as recorded)	Votes
2	DAVID MINGE	DEMOCRATIC-FARMER-LABOR	132156
2	CAL R LUDEMAN	INDEPENDENT-REPUBLICAN	131587
2	STAN BENTZ	INDEPENDENT	12246
3	JIM RAMSTAD	INDEPENDENT-REPUBLICAN	200240
3	PAUL MANDELL	DEMOCRATIC-FARMER-LABOR	104606
3	DWIGHT FELLMAN	GRASSROOTS	9164
4	BRUCE F VENTO	DEMOCRATIC-FARMER-LABOR	159796
4	IAN MAITLAND	INDEPENDENT-REPUBLICAN	101744
4	JAMES L WILLESS	INDEPENDENT	6732
4	DAN R VACEK	GRASSROOTS	4418
4	LYNN MARVIN JOHNSON	NATURAL LAW	3602
4	JO ROTHENBERG	SOCIALIST WORKERS	1236
5	MARTIN OLAV SABO	DEMOCRATIC-FARMER-LABOR	174139
5	STEPHEN A MORIARTY	INDEPENDENT-REPUBLICAN	77093
5	RUSSELL B BENTLEY	GRASSROOTS	6786
5	SANDRA COLEMAN	NEW ALLIANCE	5927
5	MARY MELLEN	NATURAL LAW	5499
5	GLENN MESAROS	INDEPENDENT	4809
5	CHRISTOPHER NISAN	SOCIALIST WORKERS	2062
6	ROD GRAMS	INDEPENDENT-REPUBLICAN	133564
6	GERRY SIKORSKI	DEMOCRATIC-FARMER-LABOR	100016
6	DEAN BARKLEY	INDEPENDENT	48329
6	JAMES H PETERSON	INDEPENDENTS FOR PEROT	16411
6	TOM FIRNSTAHL	NATURAL LAW	2400
7	COLLIN C PETERSON	DEMOCRATIC-FARMER-LABOR	133886
7	BERNIE OMANN	INDEPENDENT-REPUBLICAN	130396
8	JAMES L OBERSTAR	DEMOCRATIC-FARMER-LABOR	167104
8	PHIL HERWIG	INDEPENDENT-REPUBLICAN	83823
8	HARRY ROBB WELTY	PEROT CHOICE	22619
8	FLOYD A HENSPETER	TERM LIMITS CANDIDATE	8602

A party == "REPUBLICAN" filter, the obvious thing to write, drops every one of these. All 8 of Minnesota's 1992 Republican nominees are recorded this way, and 8 of them are the *only* Republican name in their race: a naive filter manufactures a phantom Democratic sweep of Minnesota that never happened. Fixed by matching a short list of a party's known legal aliases rather than the bare string – North Dakota's Democratic-NPL is the same kind of case, in the other party.

Runoffs double as a second row, not a replacement row. Four district-years in this file went to a runoff: Texas's 8th, 9th and 25th in 1996 (Texas required an outright majority to win a House seat until the rule lapsed after that year), and Louisiana's 5th in 2002 (Louisiana's all-party "jungle primary" doubles as the November general, with a December

runoff when nobody clears 50% – the same construction the uncontested-races chapter flags for Louisiana generally). Both rounds are stored as ordinary rows for the same candidates:

Runoff round?	Candidate	Party	Votes
TRUE	GENE FONTENOT	REPUBLICAN	21004
TRUE	KEVIN BRADY	REPUBLICAN	30366

Sum candidatevotes across every row for Kevin Brady in that district-year, as the fusion fix above would otherwise do, and he gains a phantom 30,366 votes on top of his real total – the runoff result plus an election that already ended. Fixed by keeping only the runoff round wherever one exists, and discarding the round it replaced.

What “the same district” means across time. Comparing an election to the one two years earlier requires the district number to mean the same piece of land both times, and after a decennial census it does not – California’s 12th in 2010 and California’s 12th in 2012 are different places. Every pairing in this chapter’s own construction is restricted to elections inside the same redistricting cycle (new lines take effect in 1982, 1992, 2002, 2012 and 2022); a pairing that would straddle a remap is simply never made, which is exactly the gap Figure 1 shows rather than papers over. Jacobson’s own portion of that figure goes further: his `redist` flag marks individual *districts* redrawn mid-decade – Texas in 2004, for instance, or the mid-1960s “one person, one vote” wave that redrew hundreds of seats outside any census – and those specific districts are dropped from his estimate for that year even when most of the country is untouched. This chapter’s own 2016-2024 extension has no equivalent district-level flag and can only exclude whole redistricting-transition years; see “what this brief cannot tell you.”

Put together, the 33,805 raw rows collapse to 10,861 district-year races, 1976-2024, excluding the District of Columbia’s non-voting seat, special elections, and write-in-only rows:

Quantity	Count
general, non-special, 50-state, non-write-in candidate rows	30853
district-year races, 1976-2024 (50 states, general, non-special)	10861
standard contested two-party races	9425
uncontested races	1393
nonstandard same-party pairs (top-two / jungle primary)	110
valid within-cycle pairs	8237
previous winner on the ballot again	6556
reelected when running again	6268
Minnesota ‘Independent-Republican’ rows fixed	8
of those, the only Republican row in their race	8
runoff district-years resolved to their final round	4

110 of those races are excluded from every two-party comparison below: California’s and Washington’s top-two primaries, and Louisiana’s jungle primary, can put two Democrats

or two Republicans on the November ballot with nobody from the other party at all, and a two-party *share* has no meaning there. Another 1,393 had no major-party opponent to begin with. What is left, 9,425 standard, contested, two-party races, is what the rest of this brief works from.

Before you read on, write down a guess. Figure 1 shows the advantage falling for thirty straight years. Guess whether that fall is steady or lumpy, and guess how much of the *pooled* 1976-2024 average – a single number a careless reader might quote on its own – the most recent years actually resemble. Both halves of the guess are checkable below.

03 What a held seat is worth

Restricting the close-election design to 1976-2024 and pooling every bandwidth gives essentially the same figure Figure 1 already showed for these years:

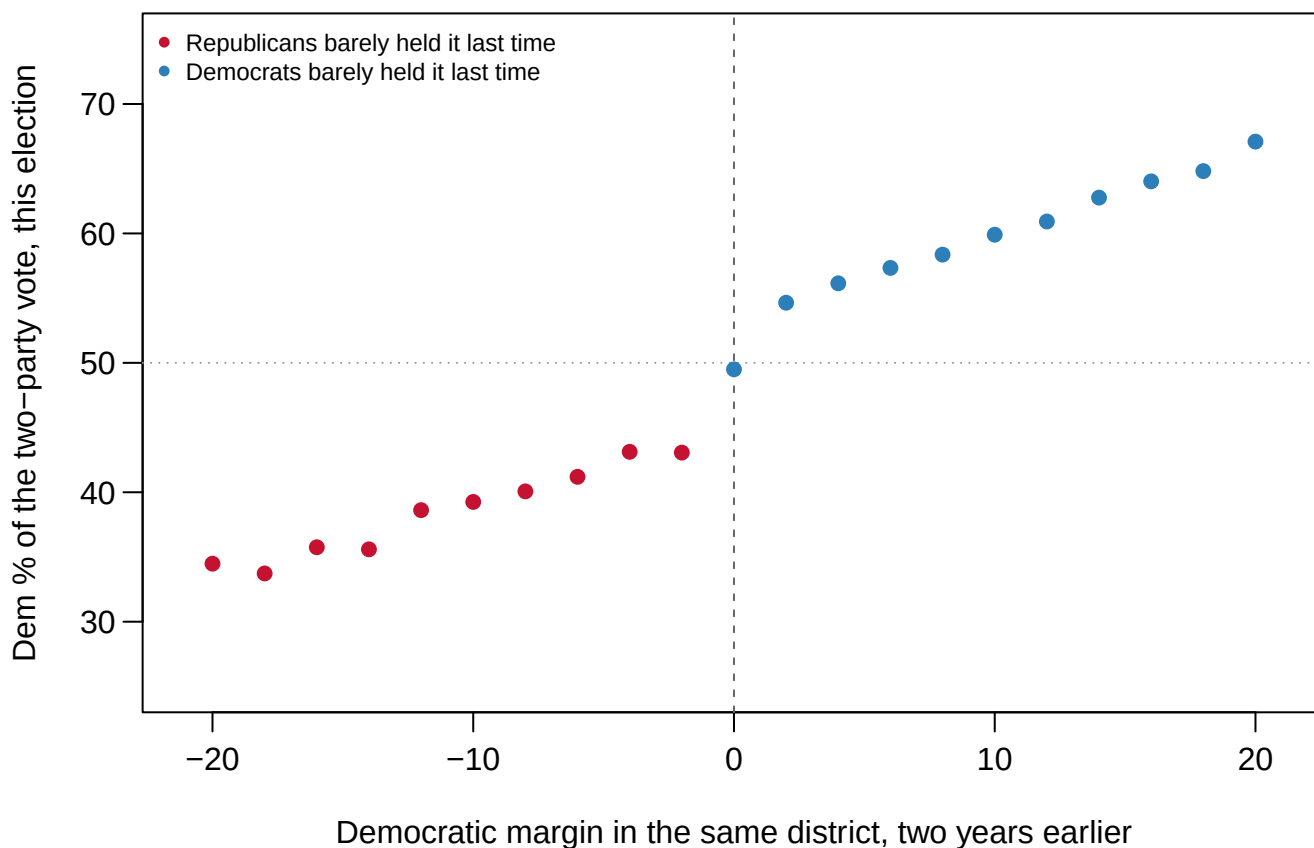


Figure 2. The two-party vote in 6,379 district-years, 1976-2024, against the same district's margin two elections earlier, binned every 2 points. A district where Democrats barely won last time (a hair to the right of the dashed line) opens this election several points bluer, on average, than a district where Republicans barely won (a hair to the left) – even though “barely won” is close to a coin flip on either side of the line. Fit as a local linear regression on each side of zero, the jump is 8.0 points at a 10-point bandwidth ($n = 2,216$, s.e. 0.65), and it barely moves at a 20-point bandwidth (8.1) or a 5-point one (8.1)

– stable across bandwidths, and close to Figure 1's own 1976-2024 average once the two methods are asked the same pooled question. Fit by decade instead of pooled, this same design falls from 12.3 points in the 1970s to 1.6 in the 2020s – a bandwidth-restricted echo of Figure 1's own decline, from a completely different design, built from a completely different regression.

Two independent designs agreeing is a check worth having, but neither answers who is on the ballot – only which party crossed 50%. That is a deliberate strength (no name-matching, no selection into running again) and also a limit: it cannot say how much of the jump is a specific person's name recognition and casework, as opposed to a party finding it easier to recruit a strong nominee, or scare one off, once it already holds the seat. Restricting the identical design to the pairs where `personal_incumbent_running` says the same person is on the ballot both times gets partway to an answer:

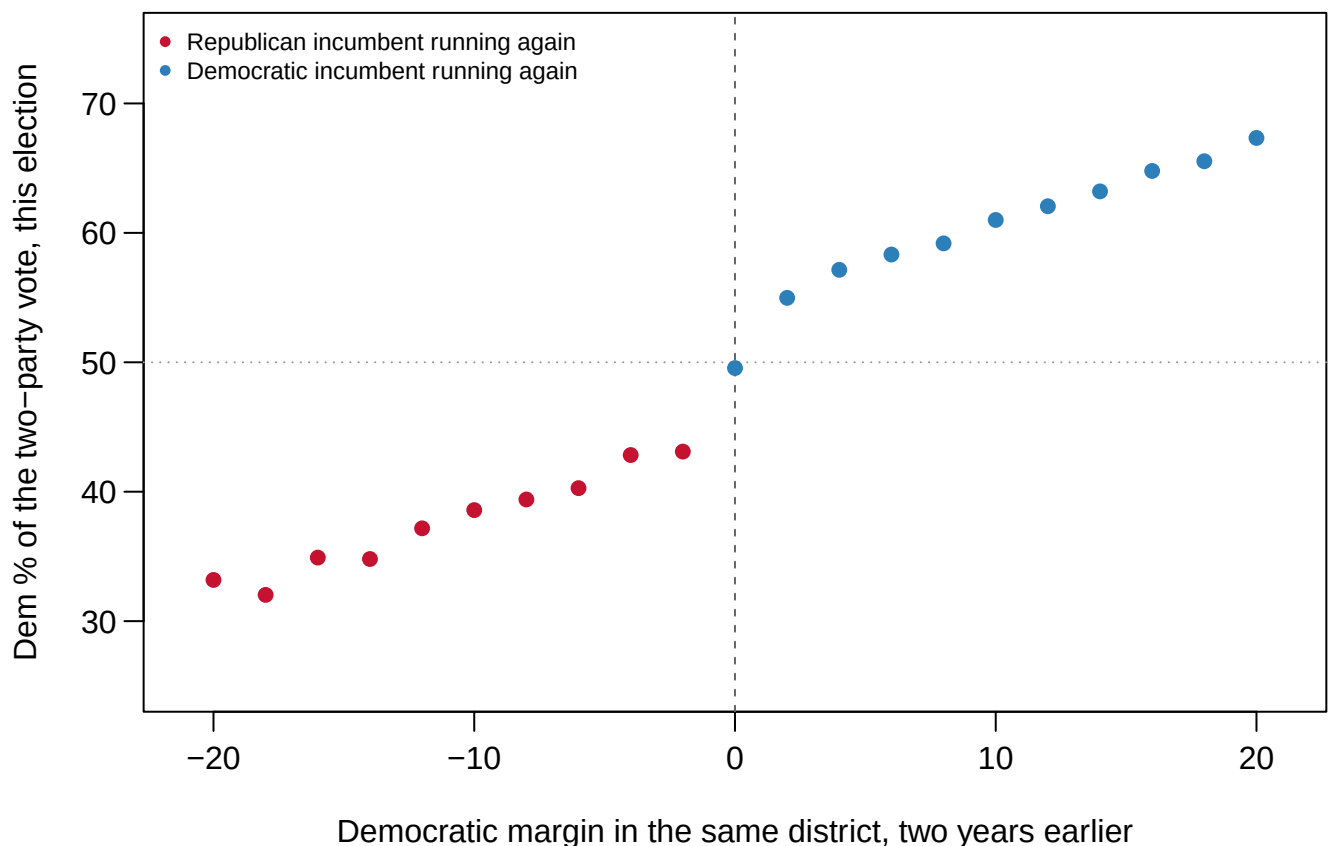


Figure 3. The same design, restricted to the 1,651 close elections (10-point bandwidth) where the same person who won last time is running again. The jump is 9.1 points (s.e. 0.71) – larger than Figure 2's 8.0, and just as stable across bandwidths (9.4 at 20 points, 9.6 at 5). Split the other way – the 565 close elections where the party held the seat but a *different* person is running – the jump falls to 5.3 points (s.e. 1.46, a much thinner slice of the data). The two numbers are not proof of a clean decomposition: choosing to run again is still a choice, made by someone who has just watched their own margin, and restricting the sample to that choice reintroduces a sliver of the same selection problem the reelection rate below runs into – only now confined to a band of already-close prior elections, where a member's decision has far less room to track district strength than it does across the full range of margins. Read cautiously, the roughly 3.8-point gap between the two is the closest this brief comes to separating the *person* from the *party*.

Members of the U.S. House who seek re-election win 96% of the time in this file – close to the number every American-politics textbook quotes. But 6,556 of 8,237 within-cycle pairs (80%) find the previous winner on the ballot again in the first place, and the reelection rate is not a causal estimate: among standard two-party races, seats where the previous winner runs again are decided by an average margin of 15.9 points against 12.4 for open seats, but the members who run again already held their seats by a wider average margin last time (14.9 points) than the ones who did not (13.2) – the wrong direction for comfort, since retirement is not a random draw from the caucus. See the retirements chapter for what the roster itself says about why members leave.

04 What this brief cannot tell you

It cannot cleanly separate a personal vote from a party-organizational one. Figure 3's split by `personal_incumbent_running` is a first pass, not a clean decomposition: restricting the sample to elections where the previous winner chose to run again reintroduces a sliver of selection, just confined to a band of already-close prior margins where that choice has much less room to operate.

Name-matching for Figure 3 and the reelection rate is exact-string matching, nothing more. A member who remarries, is recorded with a middle initial one year and without it the next, or is simply misspelled on a state's own returns will not match – landing, wrongly, on the “different person” side of Figure 3's split. So does a member seated by an off-cycle special election, the same gap discussed for Figure 1: this file has no row for such a special at all, so its winner shows up as a brand-new candidate two years later even when voters already knew exactly who they were re-electing. Figure 2's party-based design never needs a name at all, which is exactly why it is the more defensible of the two.

The 2016-2024 extension has a coarser redistricting flag than the 1946-2014 portion it continues. Jacobson's own data marks individual districts redrawn mid-decade; this chapter's own construction can only exclude entire redistricting-transition years (1982, 1992, 2002, 2012, 2022). A district redrawn mid-cycle by a court order – Texas in 2004, Pennsylvania in 2018, North Carolina several times over – keeps the same district number across a real change in the seat's boundaries in the years this chapter built itself, which will bias those years' estimates toward zero by an unknown amount.

It says nothing about California, Washington, or Louisiana. 110 district-years are excluded because their top-two or jungle-primary ballots can put two same-party candidates on the November ballot, where a two-party vote share is not defined. Whatever incumbency is worth in those states is not measured here.

05 What you should have learned

- The incumbency advantage is not a constant. It opened this 1946-2024 series near zero, climbed for thirty years to 12.3 points by 1986, and has fallen almost every decade since, to 1.9 points by 2024 – a single pooled average across the whole period describes almost none of it.
- Neither file in this brief has an `incumbent` column. Jacobson's hand-curated data catches things raw returns alone cannot – an off-cycle special election that seats a new incumbent invisibly to a file like this one's – which is why it anchors the 1946-2014

era rather than a version rebuilt here from scratch; the 2016–2024 extension inherits that blind spot along with three of its own (fusion voting, a party's legal aliases, runoffs stored as two rows).

- Two independently built, differently specified designs – a year-by-year regression using every race, and a close-election comparison using only the closest ones – land on close to the same 1976–2024 estimate, 8.0 points pooled.
- Restricted to the close elections where the same person runs again, the jump grows to 9.1 points; restricted to the ones where the party keeps the seat with someone new, it falls to 5.3. Neither restriction is a clean experiment on its own, but the gap between them is the closest this brief comes to pricing the *person* separately from the *party*.

06 Extensions

None of these is answered above, and every one can be answered from the tables the Sources section hands over.

- `by_year.csv` carries `n` for every year. Which years' estimates are thin enough (small `n`, wide implied uncertainty) that a single point should not be read too literally – and does the shape of the decline survive dropping them?
- `rdd.csv` carries `personal_incumbent_running`, so the personal-incumbency split behind Figure 3 could be fit by decade too, the way Figure 2 already is in “What a held seat is worth.” Has the *person's* share of the advantage collapsed at the same rate as the party's, or held up better?
- Split the close-election sample the way the uncontested-races chapter splits its own: South against the rest of the country. Does the advantage collapse everywhere at the same rate?
- `races.csv` flags every `nonstandard` race. For California since its top-two primary began, how often does the November runner-up (same party as the winner) come from a competitive primary versus an uncontested one?
- **Stretch.** The close-election design assumes close elections are “as good as random.” Plot the density of `lag_margin` in `rdd.csv` near zero (a McCrary-style check): does it look smooth through the cutoff, or is there a suspicious pile-up on one side?

07 Sources

Data

- MIT Election Data and Science Lab, 2017, “U.S. House 1976–2024,” <https://doi.org/10.7910/DVN/IGOUN2>, Harvard Dataverse, V15.
- Gary Jacobson, 2015, “It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections” [replication data], <https://doi.org/10.7910/DVN/29559>, Harvard Dataverse, V1.

The data itself

Every figure above is drawn from these tables. They are plain CSV, they open in a spreadsheet, and they are yours to take further.

`racess.csv`, `rdd.csv`, `by_year.csv`, `by_decade.csv`, `binned.csv`, `binned_personal.csv`,
`jacobson_validation.csv`, `headline.csv`

Academic literature

- Gelman, A. and King, G. (1990). "Estimating Incumbency Advantage Without Bias." *American Journal of Political Science* 34(4): 1142-1164.
- Jacobson, G. C. (2015). "It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections." *Journal of Politics* 77(3): 861-873.
- Mayhew, D. R. (1974). *Congress: The Electoral Connection*. Yale University Press. <https://yalebooks.yale.edu/book/9780300105872/congress/>
- Mayhew, D. R. (1974). "Congressional Elections: The Case of the Vanishing Marginals." *Polity* 6(3): 295-317.
- Lee, D. S. (2008). "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142(2): 675-697. <https://doi.org/10.1016/j.jeconom.2007.05.004>

WHY THIS DATA CANNOT BE FETCHED AGAIN

Two files behind this chapter. One refetches on its own; the other comes from a source no assistant can download for you – though you can download it yourself, once, in a browser, at no charge and with no login.

WHY NOT. The 1976–2024 returns are the MIT Election Data and Science Lab's "U.S. House 1976–2024," Harvard Dataverse, V15:

<https://doi.org/10.7910/DVN/IG0UN2>. It is openly licensed (CC0) and free. But Dataverse gates this particular file behind a guestbook – a name/email/purpose form a browser must submit before the download starts. A scripted request is not throttled or blocked quietly; it comes back immediately with HTTP 400 and the body

```
{"status":"ERROR","message":"You may not download this file without the required Guestbook response for guestbookID 458."}
```

The obstacle is a form, not a technical wall, but no script can fill out a form on a publisher's website on a stranger's behalf.

WHAT EXISTS INSTEAD, for that half. Open the DOI above in a browser, fill in the guestbook once, and download two files:

1976–2024-house.tab (the returns) and codebook-us-house-1976–2024.md (the variable list).

THE OTHER HALF FETCHES CLEANLY. Gary Jacobson's own replication data for "It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections" (*Journal of Politics*, 2015) sits at

<https://doi.org/10.7910/DVN/29559>, no guestbook, and this chapter's build script (build-jacobson-series.R) downloads J0Prepfile1.tab directly from

<https://dataverse.harvard.edu/api/access/datafile/:persistentId?persistentId=doi:10.7910/DVN/29559/19HFEF>

WHAT YOU CAN STILL CHECK. If you obtain both files, these numbers say whether you are holding the same objects:

- J0Prepfile1.tab is exactly 1,552,612 bytes: 30,005 observations, 15,228 of them real (the rest is padding with no year at all), 1946-2014
- 1976-2024-house.tab is exactly 4,156,562 bytes: 33,805 observations, 1 row per candidate per party line per district-year
- restricted to general elections, the 50 states, non-special and non-write-in rows, 30,853 rows remain
- collapsed to one row per district-year, the file yields 10,861 district-year races from 1976 through 2024
- of those, 9,425 are standard, contested, two-party races; 1,393 have no major-party opponent at all; 110 are top-two or jungle-primary ballots that put two candidates from the same party in the November race, where a two-party share has no meaning
- Minnesota's Republican nominees are recorded as "Independent-Republican" in 1992 (that was the party's legal name in the state from 1975 to 1995) - 8 rows, all 8 the only Republican candidate in their race
- four district-years went to a runoff and are stored as two rounds under the same candidates: Texas's 8th, 9th and 25th in 1996, and Louisiana's 5th in 2002
- over the 16 years both files cover, the two files' incumbency codings agree on 4,440 of 4,895 compared races (90.7%)