

Data and the State

PUBPOL 2130 / INFO 3130



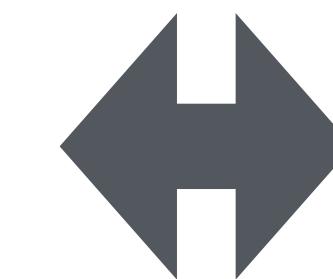
Elections and Election Data 2

Lecture 10, Thursday Feb 27

Let's take a look at some data curation

- Example: Massachusetts
- In principle, the state makes it easy to get election results, including both tabular data (from the Secretary of the Commonwealth) and shapefiles (from MassGIS).
- However they don't quite match.

The screenshot shows the homepage of the Massachusetts Secretary of the Commonwealth's election statistics website. At the top, there's a portrait of William Francis Galvin and a navigation bar with links like 'Popular Features', 'Departments', 'News & Events', 'Records', 'Publications', 'Services', 'Learning', and 'About Us'. Below that is a secondary navigation bar for 'Elections' with links to 'Popular Features', 'Voter Resources', 'Voting Information', 'Getting on the Ballot', 'Recent Updates', 'Research & Statistics', and 'Publications'. A large banner in the center says 'Every election at your fingertips.' and features a link to 'PD43+'. Below the banner is a section titled 'Explore the Data' with a query tool for selecting years (1970 to 2024) and offices (All Offices). There's also a small graphic of a map with red and blue areas.



The screenshot shows the 'MassGIS Data Layers' page from the Mass.gov website. The header includes the Mass.gov logo and a 'Menu' button. Below the header, it says 'OFFERED BY MassGIS (Bureau of Geographic Information)'. The main content area is titled 'MassGIS Data Layers' and contains the following text: 'Each digital dataset name below links to a complete data layer description. On each page you will find metadata and links to free data downloads.' It also notes that the date below each dataset name represents the month and year of the most recent update. A 'Learn more' link is provided for attribution information. At the bottom, there are links for 'Request to download statewide vector dataset' and 'Contact Us to receive Data Update notices via email, or to let us know about an error or submit questions on MassGIS'. The MassGIS logo is located in the bottom right corner.

MA data curation

- The shapefile's attribute table shows 2152 precincts. The election results show 2175.
- We will have to some work to figure out how to make them agree.
- One kind of issue: N. Adams versus North Adams. Easily cleaned.
- Another kind of issue: the shapefile has “Warren Precinct A” while the election results have “Warren Precinct A” and “Warren Precinct B.” OK, combine.
- But many problems like this are best addressed by talking to local officials.

	TOWN	WARD	PRECINCT	VTD	geometry
0	Braintree	None	5B	5B	POLYGON ((242364.517 883741.979, 242359.642 88...
1	Braintree	None	6A	6A	POLYGON ((243479.798 881994.897, 243477.538 88...
2	Braintree	None	6B	6B	POLYGON ((241614.364 881551.551, 241602.936 88...
3	Chelsea	1	1	1-1	POLYGON ((239281.036 904133.287, 239277.513 90...
4	Chelsea	1	2	1-2	POLYGON ((238408.578 904480.138, 238393.140 90...
...
2147	Boston	7	8	7-8	POLYGON ((236328.257 897806.825, 236339.793 89...
2148	Boston	12	3	12-3	POLYGON ((234387.165 896770.015, 234357.873 89...
2149	Boston	12	2	12-2	POLYGON ((234109.165 896499.623, 234101.151 89...
2150	Boston	11	2	11-2	POLYGON ((233908.714 896981.859, 233743.883 89...
2151	Boston	19	4	19-4	POLYGON ((232645.503 896432.326, 232621.204 89...

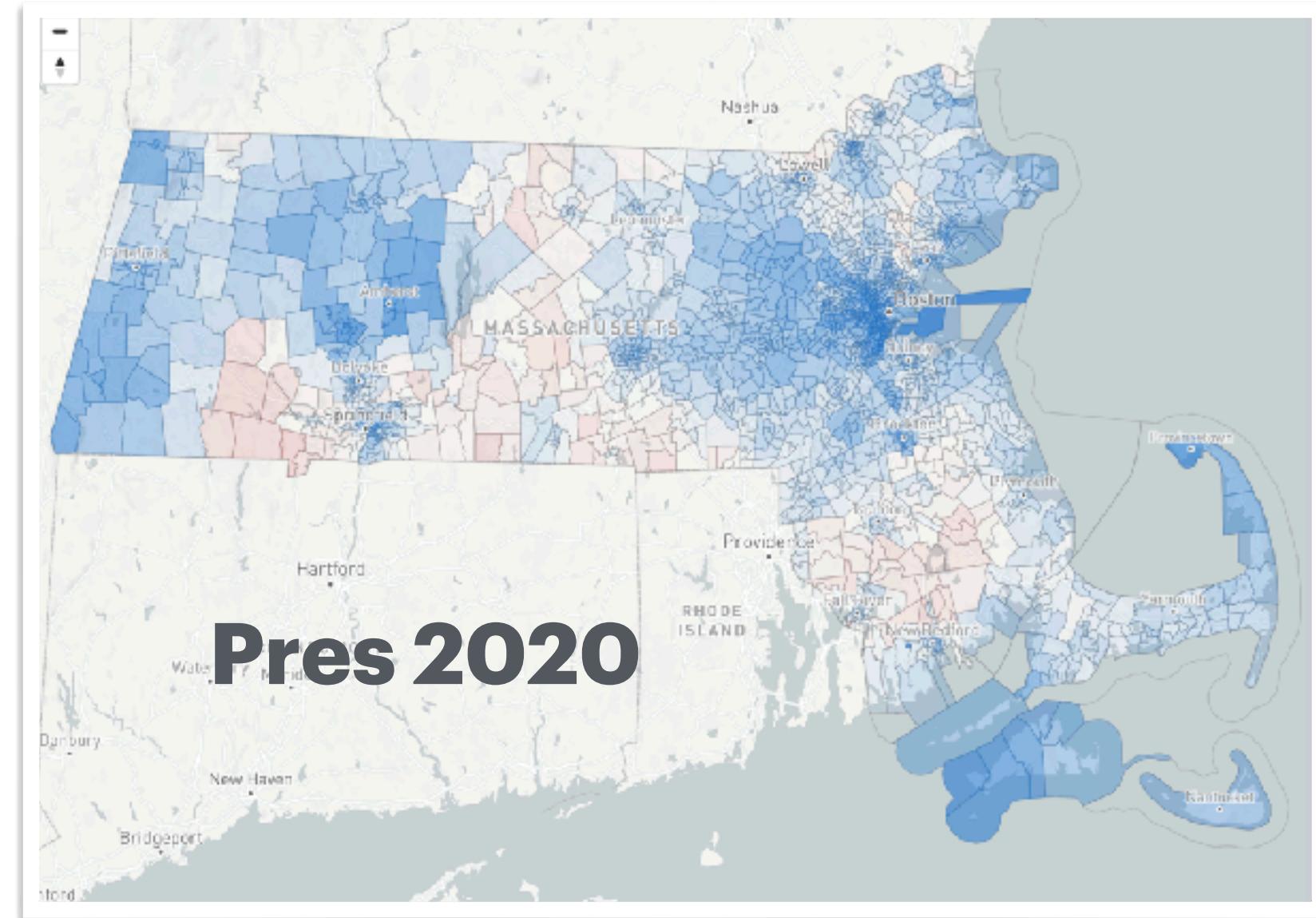
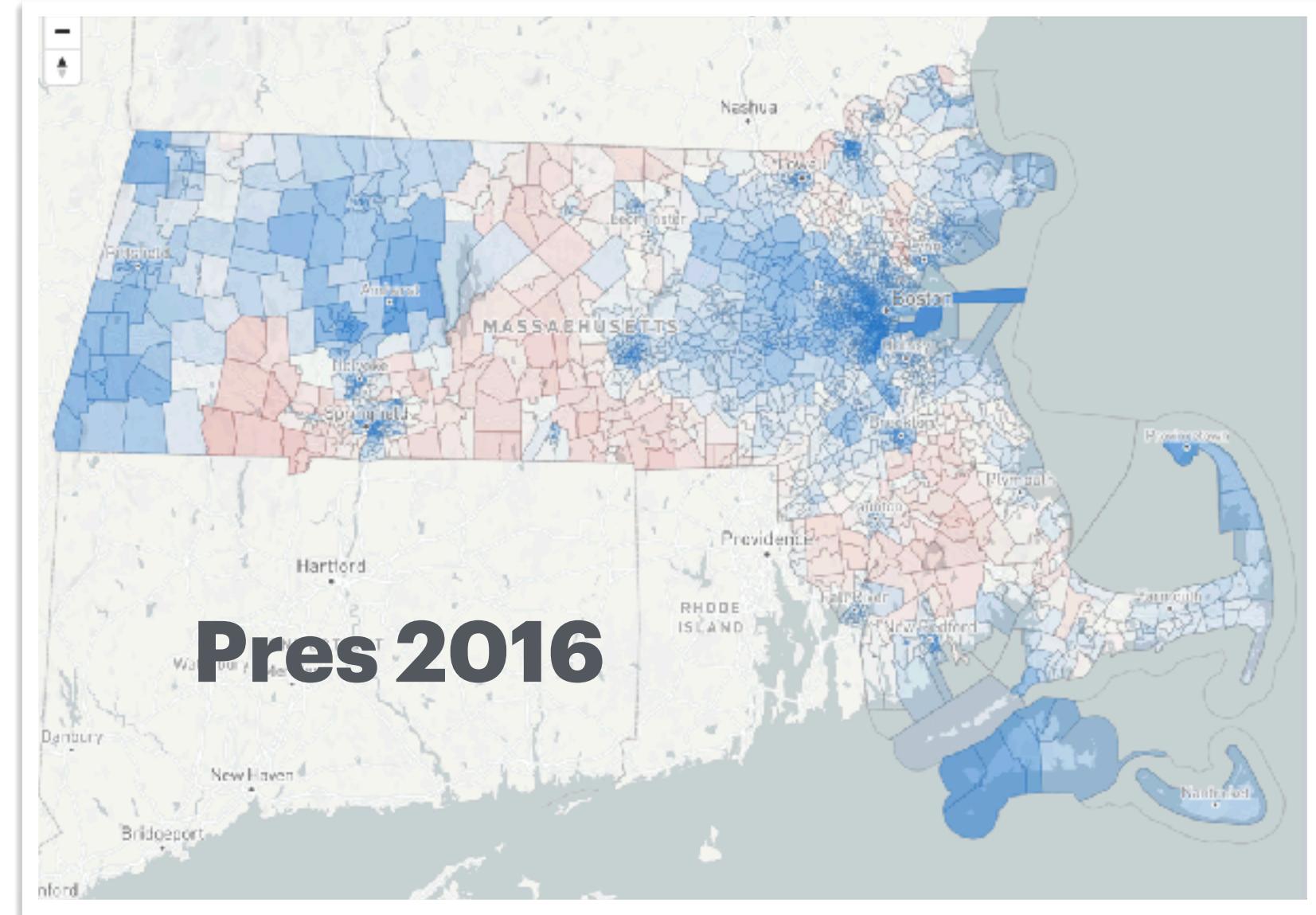
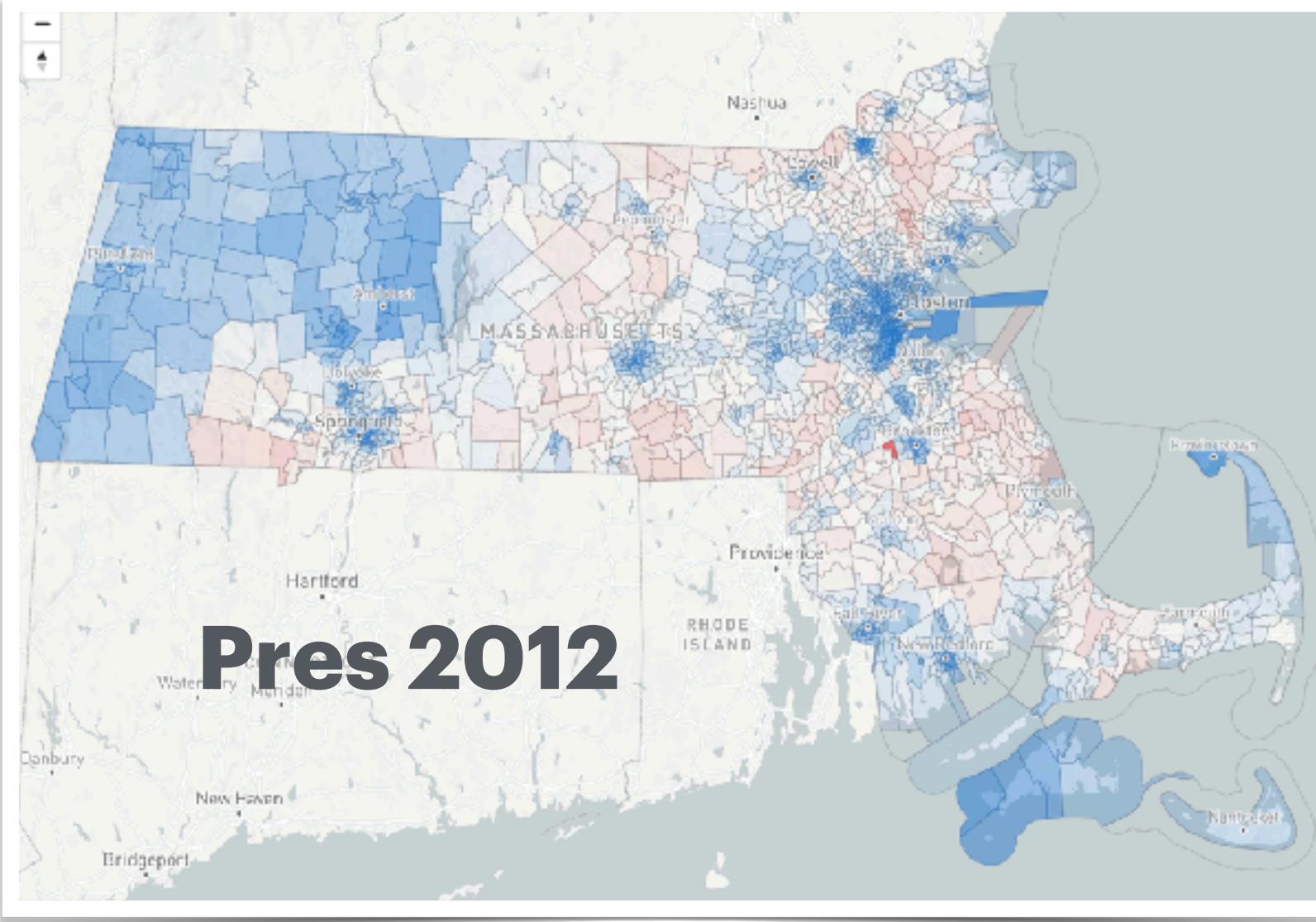
2152 rows × 5 columns

	City/Town	Ward	Pct	Obama/ Biden	Romney/ Ryan	Johnson/ Gray	Stein/ Honkala	All Others	No Preference	Blanks	Total Votes Cast
0	NaN	NaN	NaN	Democratic	Republican	Libertarian	Green-Rainbow	NaN	NaN	NaN	NaN
1	Abington	-	1	844	705	6	7	4	0.0	5	1,571
2	Abington	-	2	742	741	18	8	4	0.0	7	1,520
3	Abington	-	3	815	801	13	11	2	0.0	6	1,648
4	Abington	-	4	899	904	8	9	5	0.0	5	1,830
...
2171	Yarmouth	-	4	1,056	909	10	9	2	0.0	15	2,001
2172	Yarmouth	-	5	882	810	11	8	0	0.0	8	1,719
2173	Yarmouth	-	6	924	826	13	11	2	0.0	7	1,783
2174	Yarmouth	-	7	1,169	1,269	13	11	1	0.0	6	2,469
2175	TOTALS	NaN	NaN	1,921,290	1,188,314	30,920	20,691	6,552	0.0	16,429	3,184,196

2176 rows × 11 columns

Research application: Republicans in MA

- Massachusetts seems to have an unusual political geography: It's not a sea of red.
- In fact, it looks much less differentiated between red and blue than most states.
- Will it be hard to draw Republican districts?

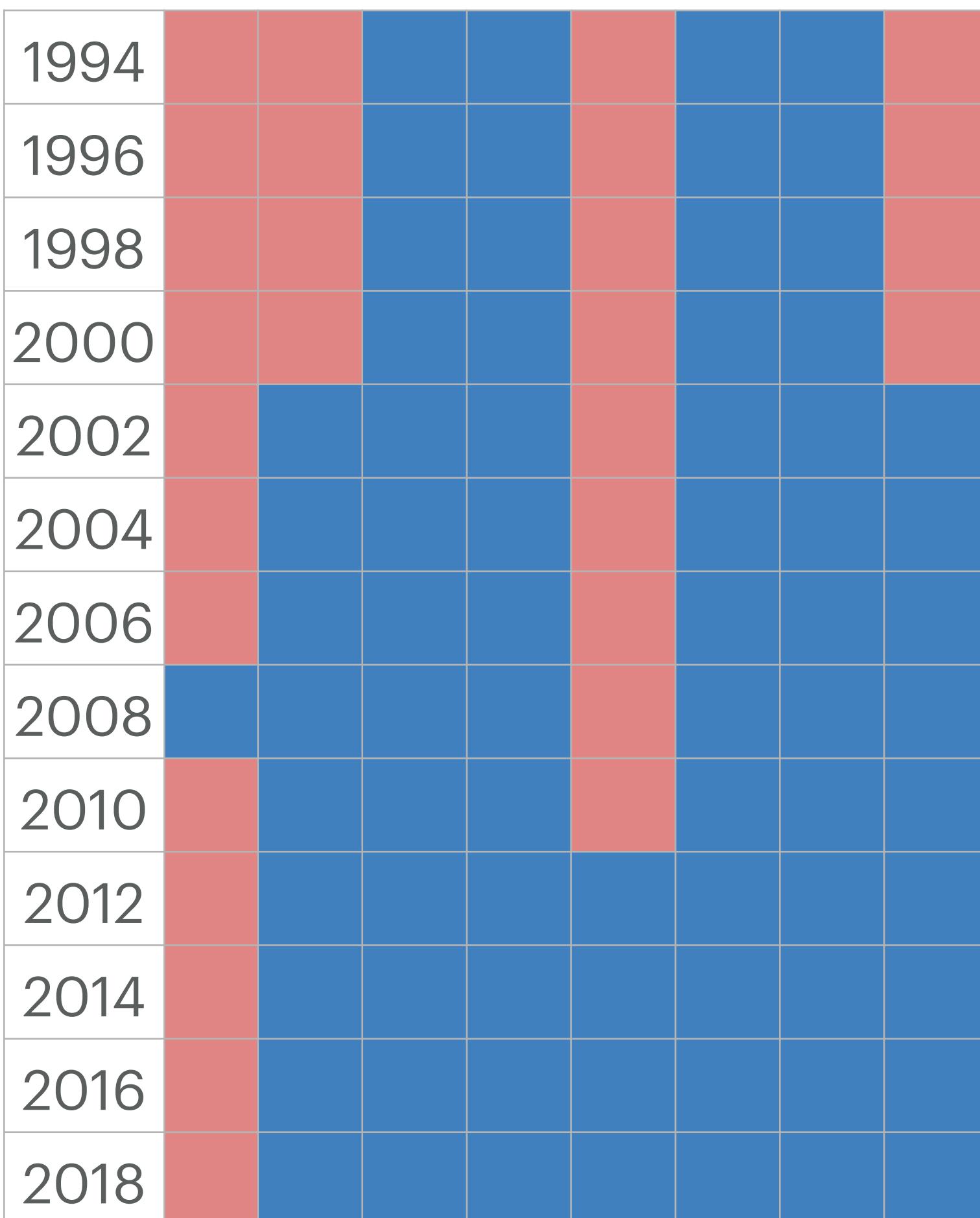


HOW BLUE IS...



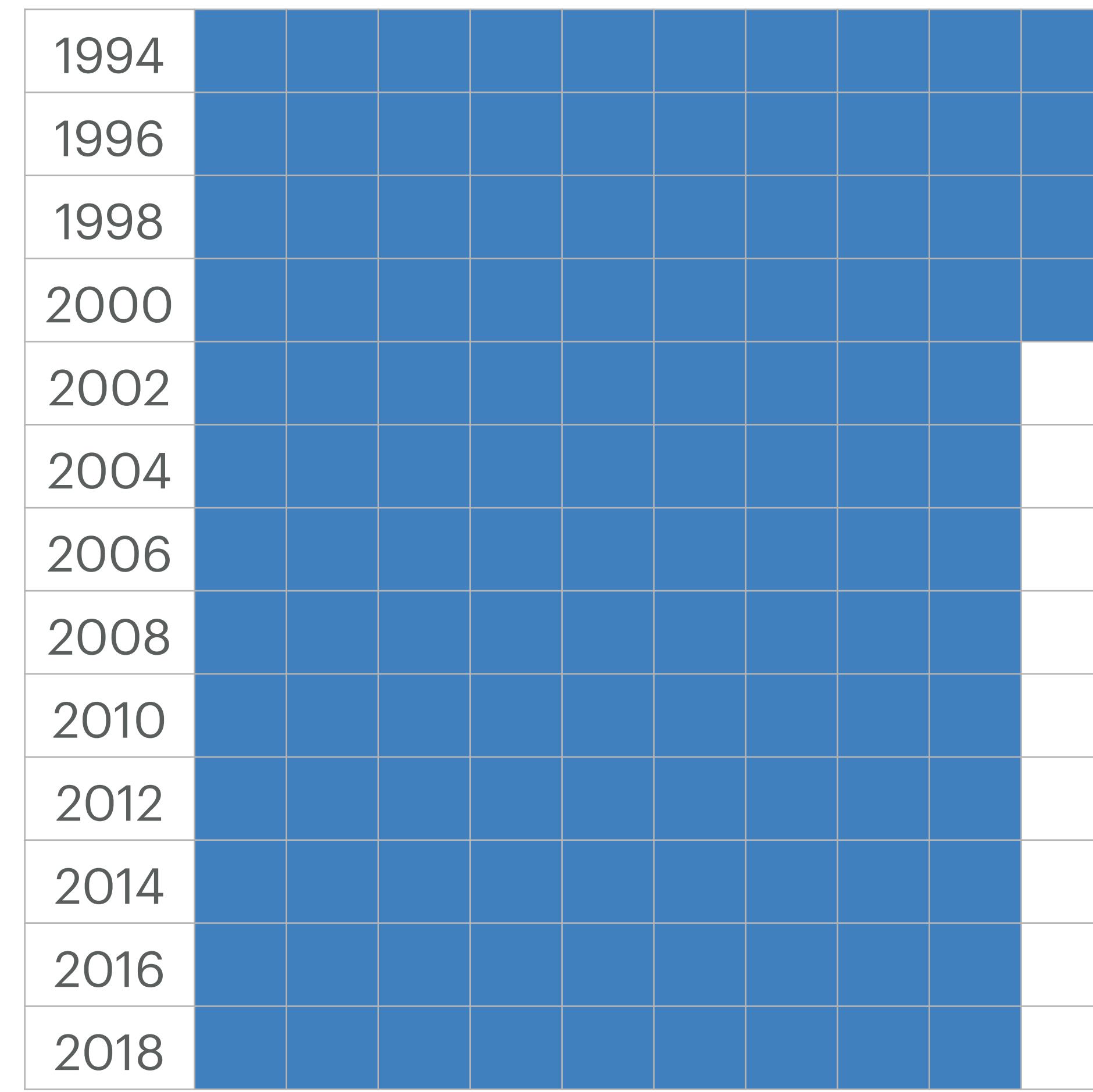
Maryland

36.9% R votes 28% R reps

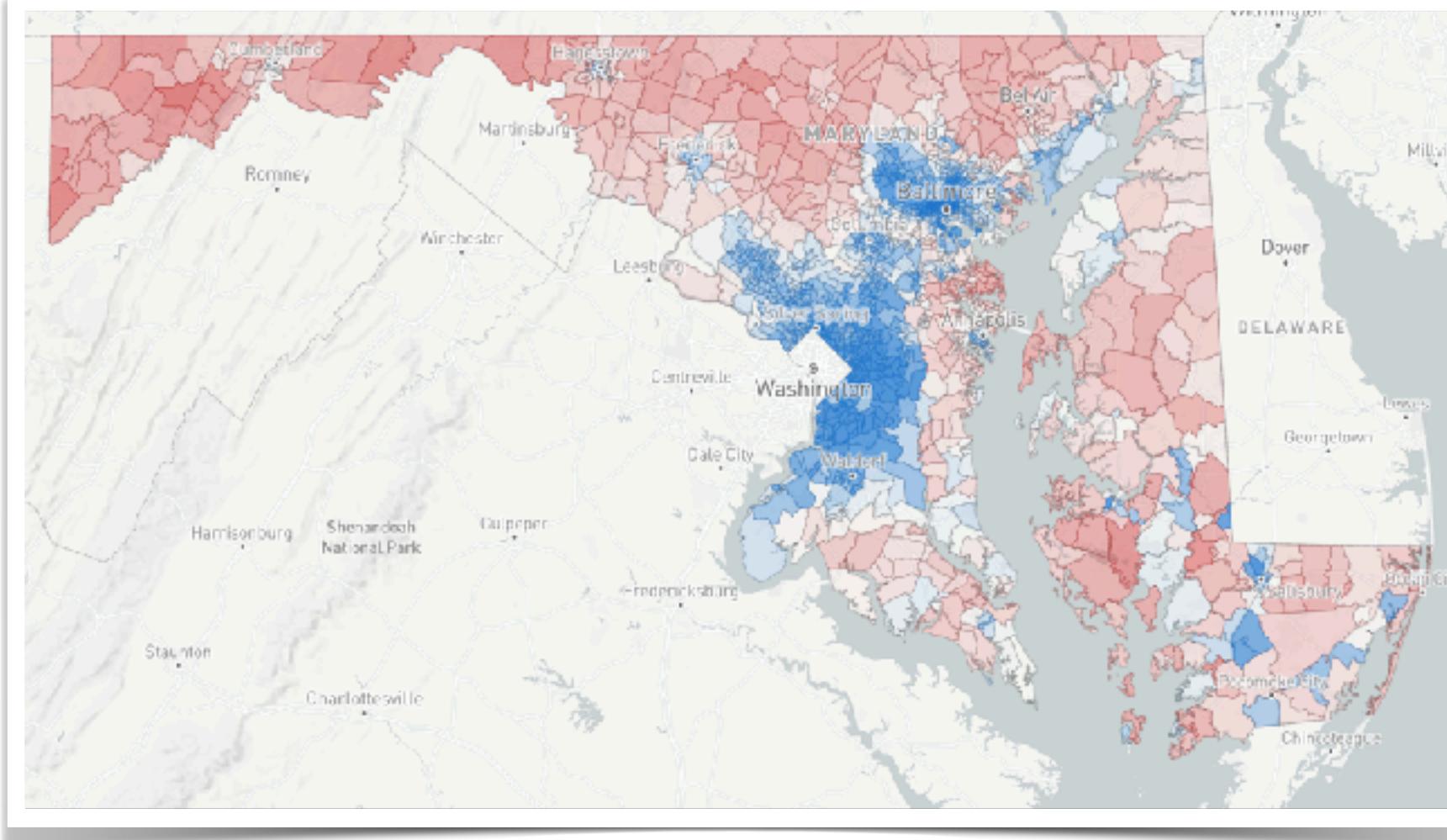


Massachusetts

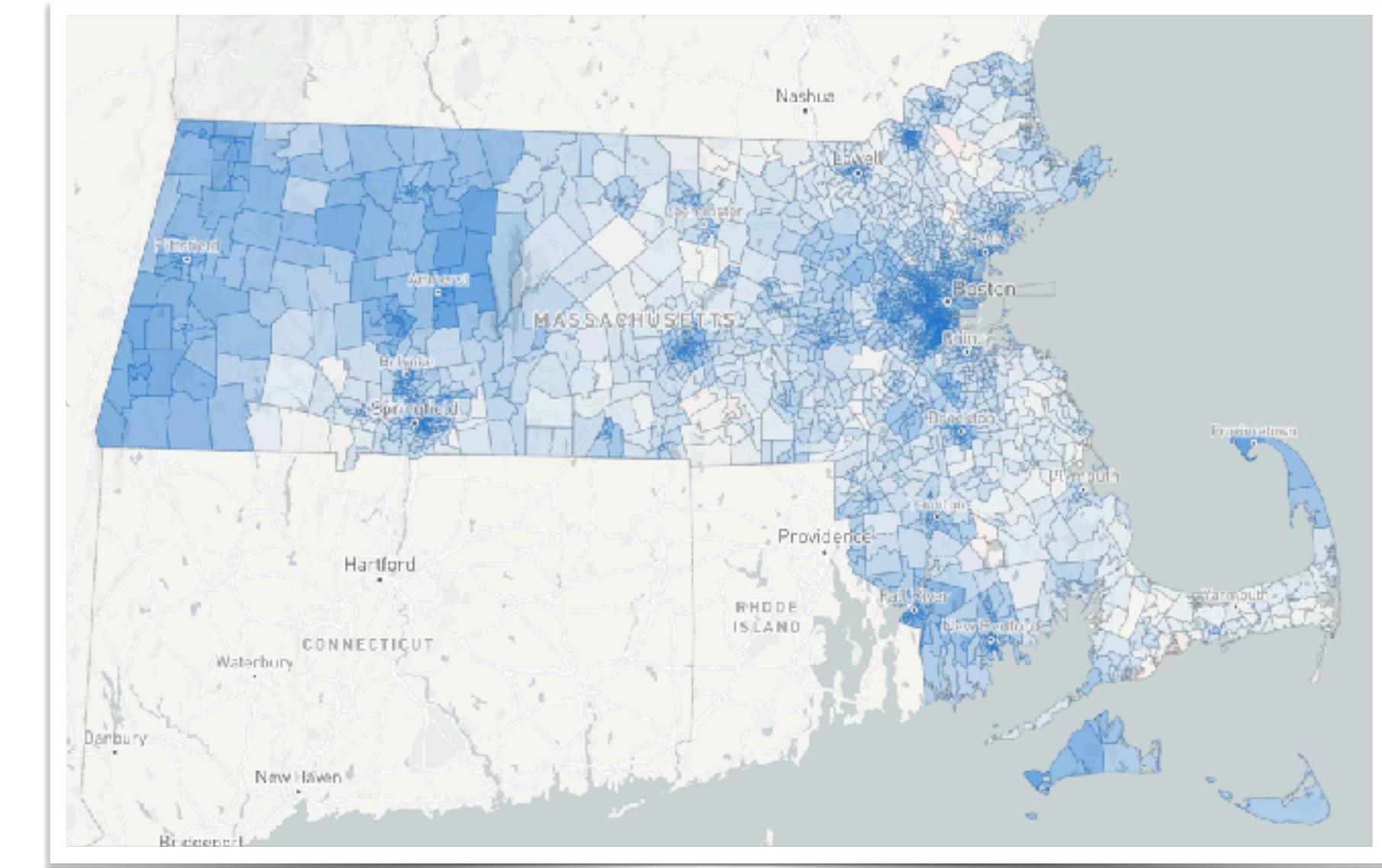
36.7% R votes 0% R reps



Gerrymandering? or Geometry?



this election has 34% R



this election has 32% R

Republican Futility Theorem

Suppose you have a list of units with corresponding populations p_i and R margins $\delta_i = r_i - d_i$; the number of R votes minus the number of D votes. Re-index so that they are ordered from greatest to least by margin per capita:

$$\delta_1/p_1 \geq \delta_2/p_2 \geq \dots \geq \delta_n/p_n.$$

We will call a collection of units S a *grouping*, and let $p(S)$ and $\delta(S)$ be its population and R margin, found by summing the p_i and δ_i for its units. Let D_k be the grouping indexed by $\{1, \dots, k\}$. Let K be the smallest integer k for which $\delta(D_k) \leq 0$. This means that D_{K-1} has a collective R majority, but if you add the K th unit you get a grouping D_K that fails to have an R majority.

Theorem 1. *With the notation above, let M be any positive integer.*

Case 1. $M \leq p(D_{K-1})$. There exists an R-majority grouping of size at least M .

Case 2. $p(D_{K-1}) < M \leq p(D_K)$. Inconclusive: such a grouping may or may not exist.

Case 3. $p(D_K) < M$. There does not exist an R-majority grouping of size at least M .

Proof. In Case 1, it is clear that a Republican grouping can be created, because D_{K-1} is a Republican-majority grouping of sufficient size.

We present examples to illustrate that Case 2 is inconclusive.

i	r_i	d_i	p_i	δ_i/p_i
1	8	0	8	1
2	1	9	10	-4/5
3	0	5	5	-1

i	r_i	d_i	p_i	δ_i/p_i
1	8	0	8	1
2	1	9	10	-4/5
3	0	8	8	-1

For both examples, fix $M = 13$. We have $K = 2$ in both examples because $\delta(D_1) = 8 > 0$ and $\delta(D_2) = 0$. Both fall under Case 2 because $p(D_1) = 8$ and $p(D_2) = 18$, while $M = 13$. In the left-hand example there exists an R-majority grouping, made by putting together units 1 and 3 to form a grouping with $\delta = 3$ and population 13. But in the right-hand example there is none, which is easily confirmed by considering all of the combinations.

Finally, in Case 3, we have $p(D_K) < M$.

Claim. *Let $S = D_K$ and suppose that $p(S) < M$. Then for any $S' \subseteq \{1, \dots, n\}$,*

$$p(S') > p(S) \implies \delta(S') < \delta(S).$$

The claim asserts that D_K has the optimal R margin among all groupings with at least as much population. Since we seek a grouping larger than $p(D_K)$ and since $\delta(D_K) \leq 0$, this implies that a R-majority grouping cannot be formed. So it just remains to prove the claim.

Let $A = S' \setminus S$ and $R = S \setminus S'$ denote the sets of indices added to and removed from S , respectively, to make S' . Since A and R are disjoint, and we have assumed that $p(S') > p(S)$, it follows that $p(A) > p(R)$. Let $\mu = \max\{\frac{\delta_i}{p_i} \mid i \in A\}$ and let $\mu' = \min\{\frac{\delta_i}{p_i} \mid i \in R\}$. Note that, since $R \subseteq S = \{1, \dots, K\}$ and $A \subseteq S^c = \{K+1, \dots, n\}$ and the $\frac{\delta_i}{p_i}$ are non-increasing, we have $\mu \leq \mu'$.

Note that every unit $i \notin S$ has a Democratic majority ($\delta_i < 0$). This is because Republican-majority units are added to S in decreasing order of $\frac{\delta_i}{p_i}$ until the overall margin satisfies $\delta \leq 0$, so by construction every unit with a Republican majority is in S . It follows, since $A \subseteq S^c$, that $\mu < 0$.

We have $\mu \cdot p(R) > \mu \cdot p(A)$ because $p(R) < p(A)$ and $\mu < 0$. Also, $\mu' \cdot p(R) \geq \mu \cdot p(R)$. So, transitively, $\mu' \cdot p(R) > \mu \cdot p(A)$.

Note that

$$\mu' \cdot p(R) = \sum_{i \in R} \mu' \cdot p_i \leq \sum_{i \in R} \frac{\delta_i}{p_i} \cdot p_i = \delta(R).$$

There are more ways to redistrict MA than particles in the galaxy... and every possible plan gives a 9-0 Democratic sweep!

see "Locating the Representational Baseline" in Election Law Journal, 2019

([LINK](#))

Uniformity

- We showed that, in recent history, Massachusetts was just incredibly uniform in its political distribution.
- In the 2000 presidential election, Bush got 35.2% of the vote statewide – but he almost never cracked 50% in any location.
- By the year 2016, Trump performs very similarly statewide (35.3%) but now the state sees a few more extremes—the histogram is fatter.

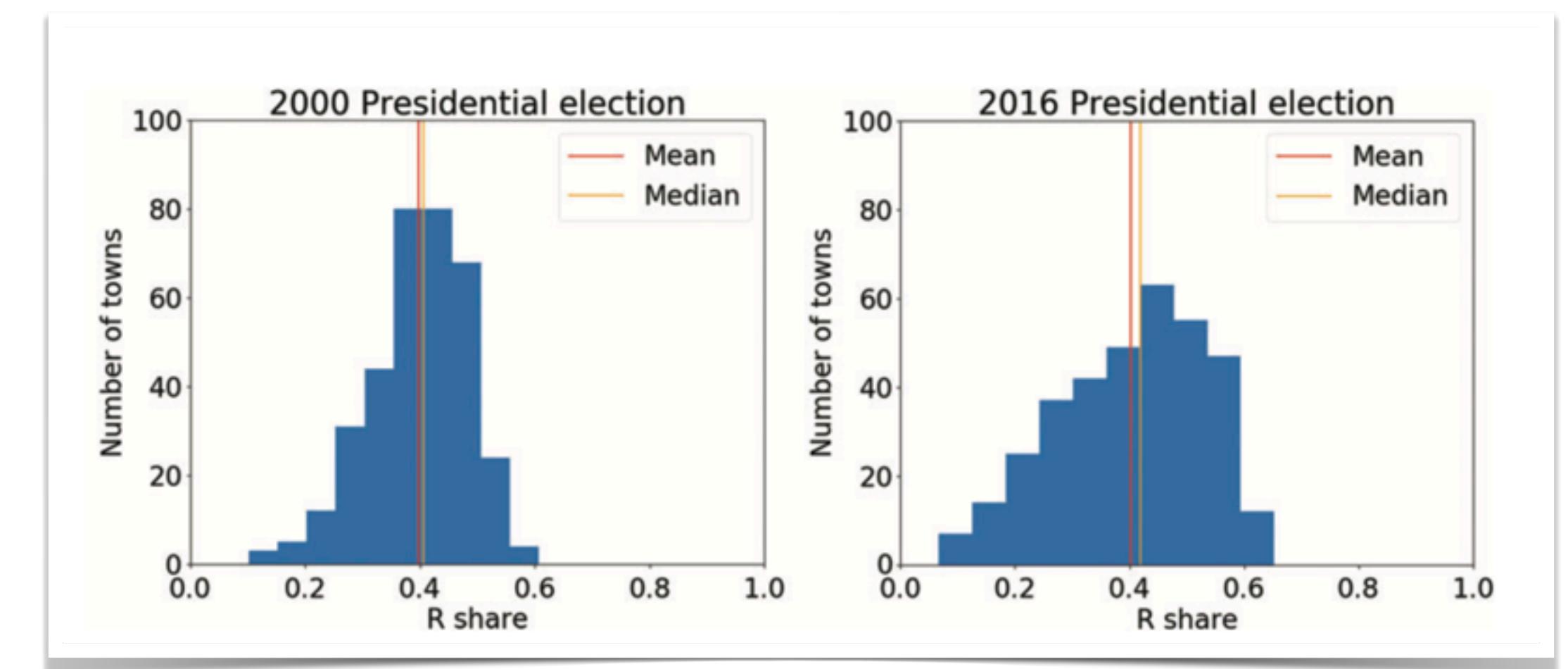


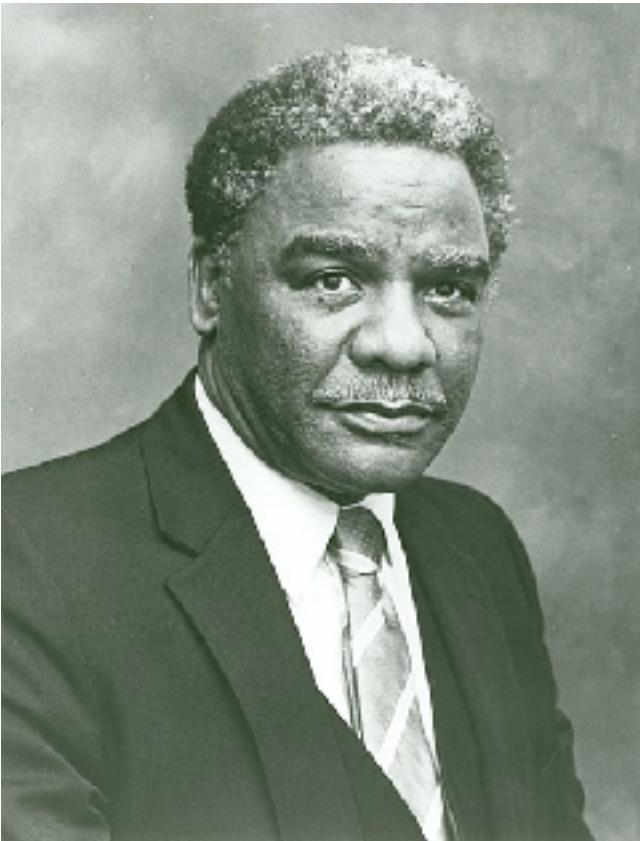
TABLE 2. IF DISTRICTS WERE TO BE MADE OUT OF TOWNS OR OUT OF PRECINCTS, WITH NO REGARD TO SHAPE OR EVEN CONNECTEDNESS, HOW MANY R OR D DISTRICTS COULD BE FORMED? FEASIBILITY AND INFEASIBILITY BOUNDS ARE SHOWN IN THIS TABLE

Election	<i>D</i> candidate– <i>R</i> candidate	R share	Seat quota (9 seats)	<i>R</i> feas/infeas		<i>D</i> feas/infeas	
				Town	Prec	Town	Prec
Pres 2000	Gore–Bush	35.2%	3.2	0/1	—	9/-	—
Sen 2000	Kennedy–Robinson/Howell	25.3%*	2.3	0/1	—	9/-	—
Sen 2002	Kerry–Cloud	18.7%	1.7	0/1	0/1	9/-	9/-
Pres 2004	Kerry–Bush	37.2%	3.4	1/2	1/2	9/-	9/-
Sen 2006	Kennedy–Chase	30.5%	2.8	0/1	0/1	9/-	9/-
Pres 2008	Obama–McCain	36.8%	3.3	1/2	1/2	9/-	9/-
Sen 2008	Kerry–Beatty	31.9%	2.9	0/1	0/1	9/-	9/-
Sen 2010	Coakley–Brown	52.4%	4.7	9/-	9/-	8/9	8/9
Pres 2012	Obama–Romney	38.2%	3.4	3/4	3/4	9/-	9/-
Sen 2012	Warren–Brown	46.2%	4.2	7/9	7/8	9/-	9/-
Sen 2013	Markey–Gomez	44.8%	4.0	7/9	7/8	9/-	9/-
Sen 2014	Markey–Herr	38.0%	3.4	3/4	3/4	9/-	9/-
Pres 2016	Clinton–Trump	35.3%	3.2	2/3	3/4	9/-	9/-

Lower-variance elections (see previous table) are marked in gray. Election winners shown in boldface; R share is with respect to two-way vote; seat quotas are proportional share of nine seats.

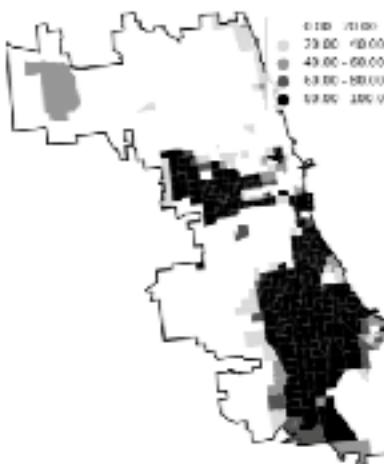
*Libertarian vote share included with R in 2000 Senate race.

Application: Polarization in Chicago

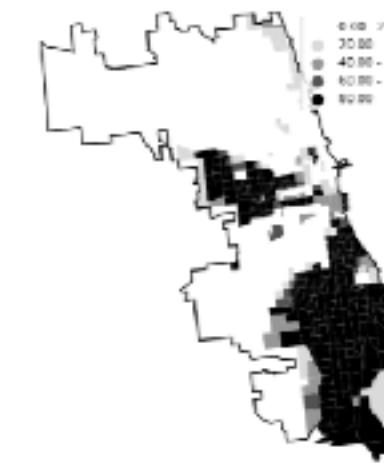


Mayor Harold Washington

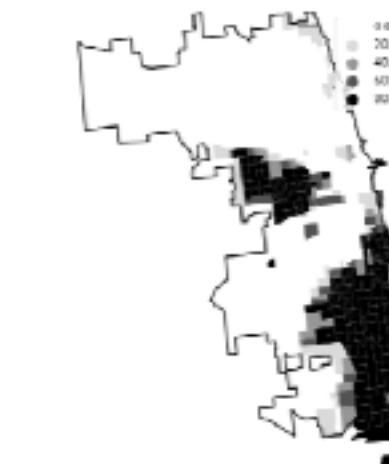
more than half of W Dems crossed
over to support the Repub rather
than vote for him in 1983



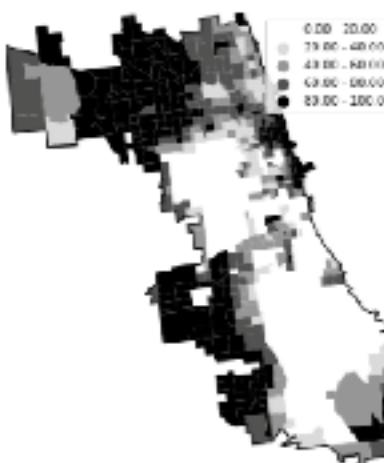
(A) 1990, Black



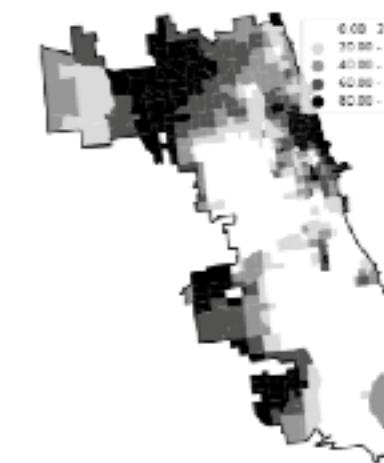
(B) 2000, Black



(C) 2010, Black



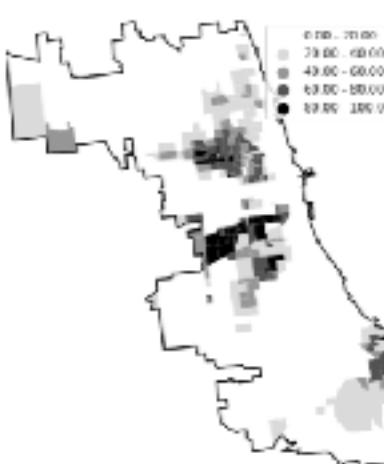
(D) 1990, White



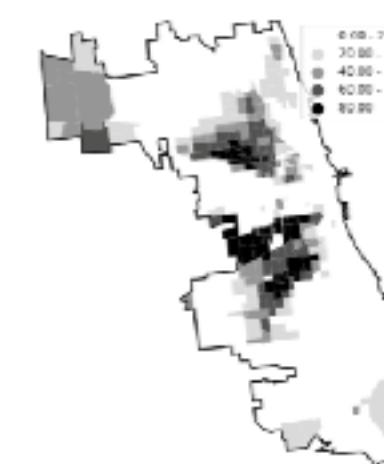
(E) 2000, White



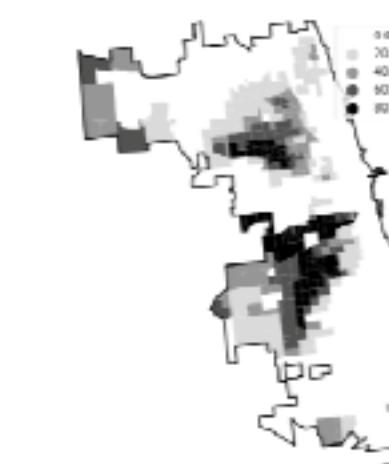
(F) 2010, White



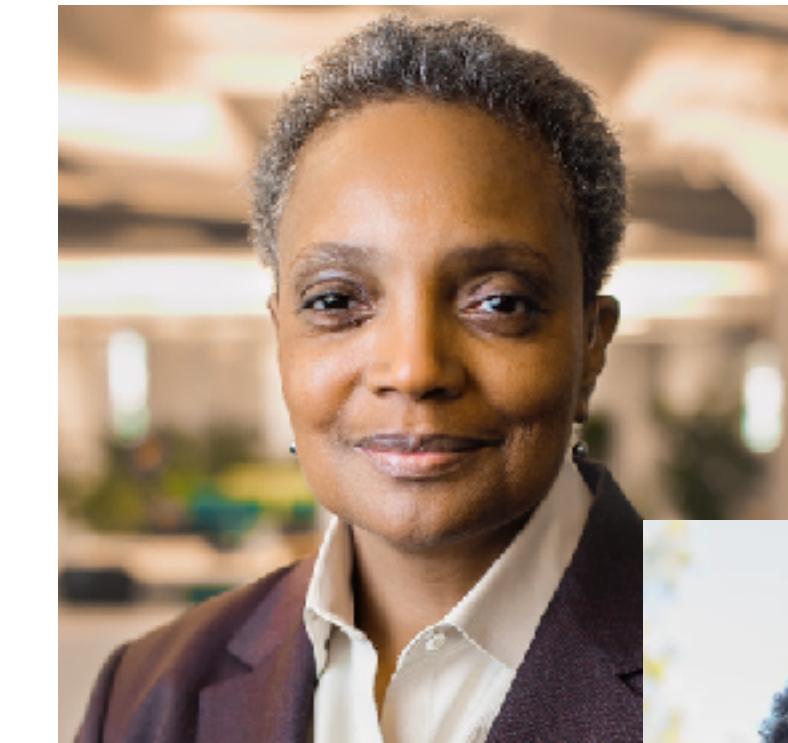
(G) 1990, Hispanic



(H) 2000, Hispanic



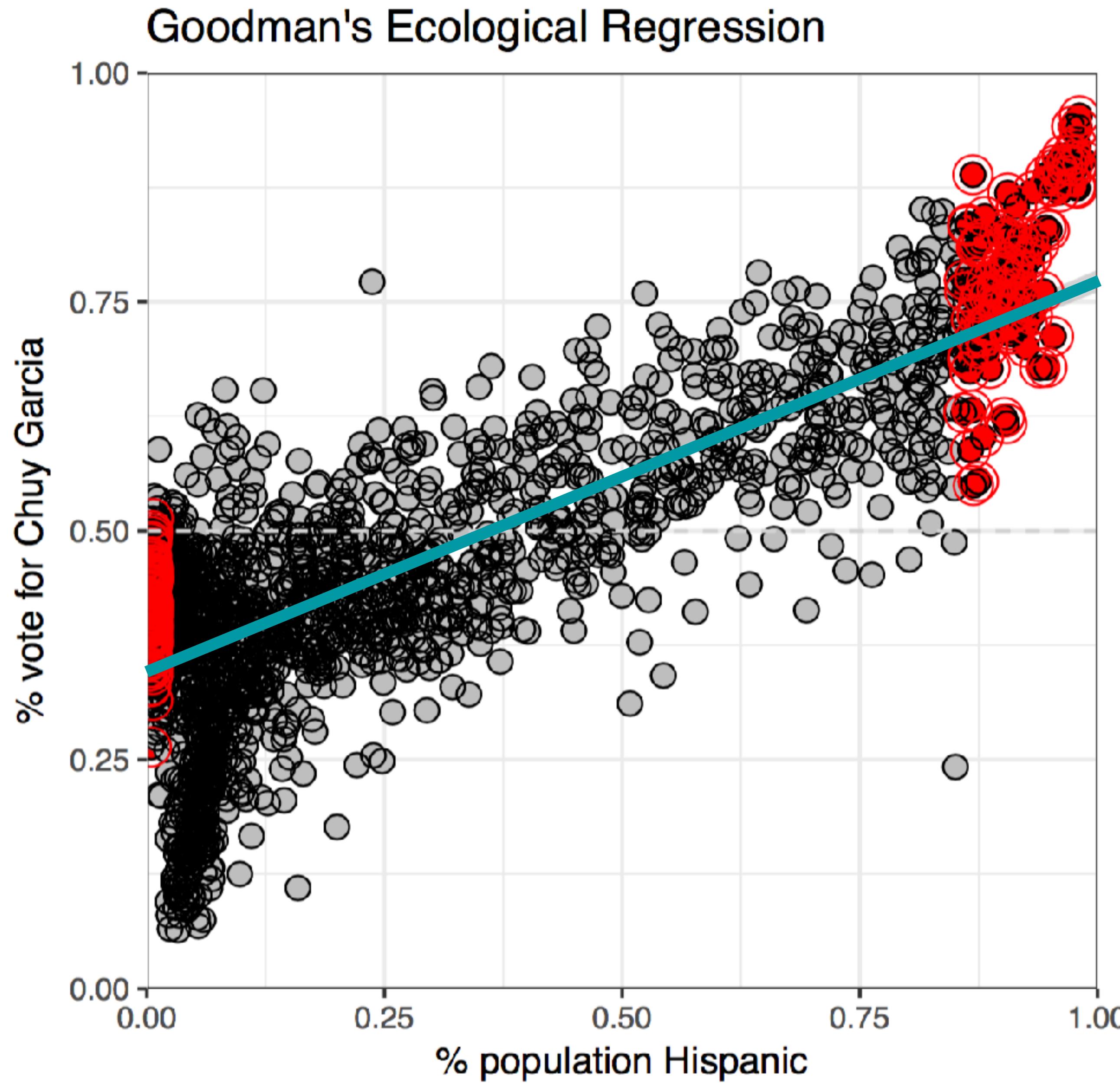
(I) 2010, Hispanic



Lightfoot, Johnson

now the last two mayors are Black

Chicago 2015 mayoral runoff (Rahm–Chuy)



Ecological regression:
fit a line! Intercepts
with $x=0$ and $x=1$ are
estimates of vote for
non-Hisp and Hisp
voters, respectively.

