Local Representation in the United States: A New Comprehensive Dataset of Elections

Christopher Warshaw[*]   Justin de Benedictis-Kessner[†]   Yamil R. Velez[‡]

August 3, 2022

Abstract

The study of urban and local politics in the United States has long been hindered by a lack of centralized sources of election data. We introduce a new dataset of nearly 55,000 electoral contests that encompasses races for seven distinct local political offices in most medium and large cities and counties in the U.S. over the last three decades. Our data provide partisan and demographic information about candidates in these races as well as electoral outcomes. We demonstrate the utility of these data with three applications: the descriptive representation of women and race/ethnic groups among candidates and office-holders, the partisan nationalization of local contests, and the match between district partisanship and local politicians' voting records in city councils. Together, our data provide a myriad of opportunities for future research on subnational politics and remove a significant barrier to the study of representation and elections in local governments.

Keywords: Local politics, representation, elections, gender, race and ethnic politics, nationalization

Word count: 3811

---

[*]Associate Professor, Department of Political Science, George Washington University, `warshaw@gwu.edu`

[†]Assistant Professor, John F. Kennedy School of Government, Harvard University, `jdbk@hks.harvard.edu`

[‡]Assistant Professor, Department of Political Science, Columbia University, `yrv2004@columbia.edu`

One of the most persistent challenges in the study of urban and local politics in the United States is the lack of information about local elections, candidates, and elected officials (Trounstine, 2010; Warshaw, 2019). As a result, there are only a handful of studies on the causes or consequences of local elections. Moreover, these studies tend to focus on a single time period (e.g., Sances, 2018), geographic unit (e.g., Arnold and Carnes, 2012), or office (e.g., Ferreira and Gyourko, 2009), rather than examining variation across time, geography, and offices.

In this paper, we describe a new comprehensive dataset of election returns from nearly 55,000 contests in over 1,600 cities, counties, and school districts from 1990-2021. It includes information about elections for mayors, city councils, county executives, county legislatures, sheriffs, prosecutors, and school boards. It also includes a host of supplemental data about candidates, including their partisanship, gender, race/ethnicity, and incumbency status. For many elections, it also includes information on the political characteristics of constituencies, such as their ideology and presidential voting patterns.

To demonstrate the utility of this dataset for expanding knowledge about democracy in subnational politics, we conduct three illustrative research applications with the dataset. We examine descriptive representation across types of local office, the nationalization of elections along partisan lines, and dyadic representation on city councils.

This new dataset will also enable scholars to study a host of other research questions. It enables examination of whether politicians represent the demographic, partisan, and ideological characteristics of their constituents (e.g., Holman, 2017; Schaffner, Rhodes, and La Raja, 2020; Tausanovitch and Warshaw, 2014). It also enables expanded work on the factors that affect local elections. For instance, do economic downturns hurt incumbents in local elections or candidates from the president's party (de Benedictis-Kessner and Warshaw, 2020a; Hopkins and Pettingill, 2018)? Moreover, it facilitates study of the incumbency advantage across election types, institutional contexts, and candidate characteristics (e.g., Bernhard and de Benedictis-Kessner, 2021; de Benedictis-Kessner, 2018; Trounstine, 2011). Finally,

this dataset enables scholars to expand the study of how elections shape a host of political outcomes such as policy (e.g., de Benedictis-Kessner and Warshaw, 2016, 2020*b*; Thompson, 2020), interest group activity (Anzia, 2019), and intergovernmental lobbying (Payson, 2020).

Data

We have assembled a novel dataset that aggregates and expands previous work on local elections. Our target universe was all cities with more than 50,000 people and counties with more than 75,000 people in the 2020 Census.[1] The dataset includes information on the vast majority of the cities and counties in our target universe. The foundation for this new dataset is previous work on mayoral elections (Ferreira and Gyourko, 2009; Gerber and Hopkins, 2011; de Benedictis-Kessner and Warshaw, 2016), county legislative elections (de Benedictis-Kessner and Warshaw, 2020*b*), sheriff elections (Thompson, 2020), prosecutor elections (Hessick and Morse, 2019; Krumholz, 2019), the MIT Election and Data Science's Lab's data on recent elections, and the California statewide election database (CEDA, 2020). We built upon these datasets using several approaches. First, we expanded both the types of offices covered and the temporal coverage of these datasets. We worked with a team of research assistants who coded results from thousands of local elections based on publicly obtainable information on city and county websites. In addition, we scraped data from the crowdsourced website OurCampaigns.com and statewide election websites where available. Finally, we obtained some unofficial returns from newspaper archives.

The resulting dataset of local election returns includes information on 54,689 contests and 77,105 unique candidates in 1,604 cities, counties, and school districts from 1990-2021 (Table 1 and Figure 1). It includes information about elections for mayors, city councils, county executives, county legislatures, sheriffs, prosecutors, and school boards. In many cases, we verified the validity of the election returns by cross-checking them across sources.

---

[1]There are 749 counties and 877 cities in our target universe. Our data collection for school boards was more opportunistic.

Table 1: Summary Information about Database

Office	Years Available	Geographic Units	Elections	% Contested	Unique Candidates
Mayor	1990–2022	573	4,238	81%	7,328
City Council	1990–2022	477	17,376	78%	30,715
County Executive	1990–2022	118	803	77%	1,112
County Legislature	1990–2022	542	22,581	64%	26,212
Sheriff	1990–2021	619	2,651	56%	2,892
Prosecutor	1990–2021	684	3,429	33%	2,689
School Board	1990–2021	221	3,467	92%	8,402

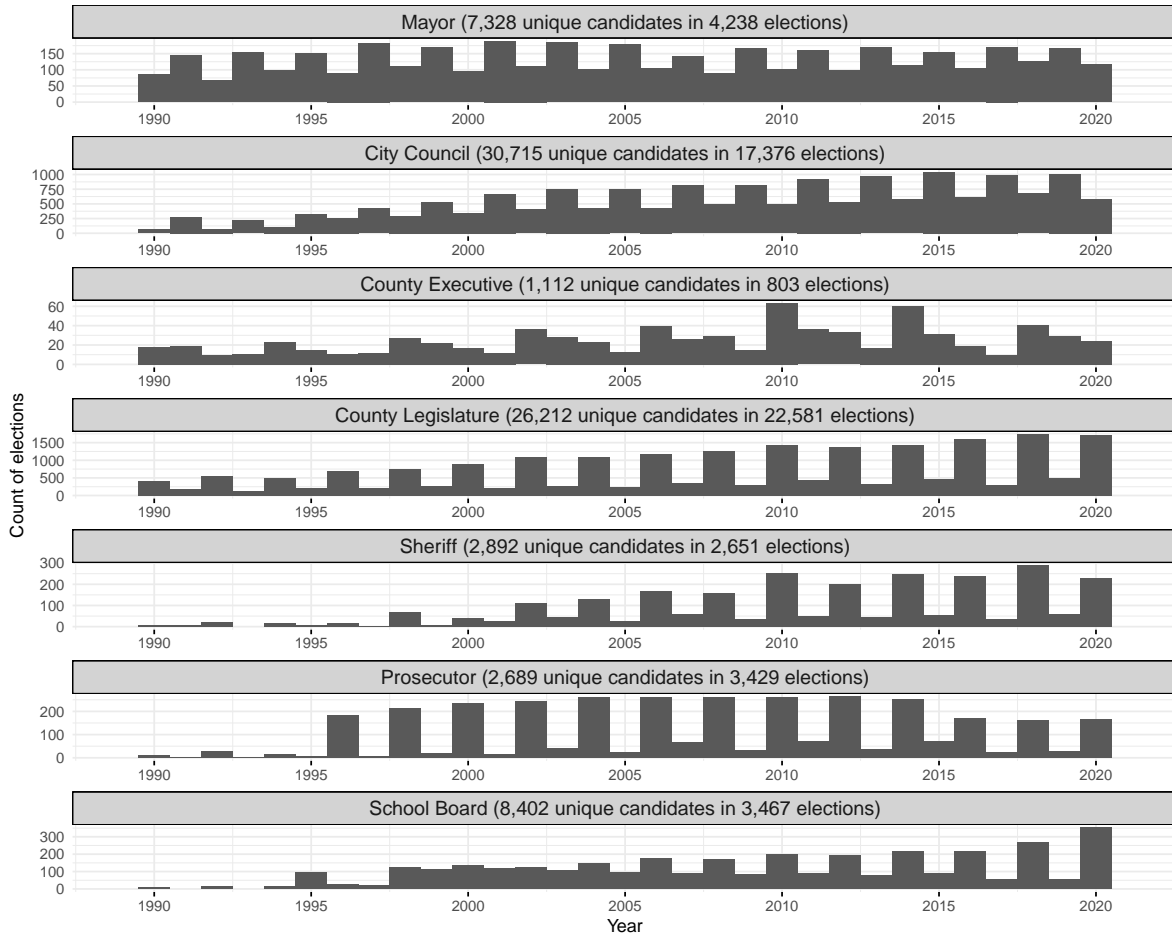Figure 1: Temporal Coverage of Elections Data

We then augmented the raw election returns with an array of supplementary information about individual candidates, including their partisanship (even in officially nonpartisan elections), gender, race/ethnicity, and incumbency status (see Appendix A). In order to do this, we matched the election returns with a wide range of auxiliary data. First, we sought to

match each candidate to a record in L2 and TargetSmart's national voter files by name and location. Second, we sought to match each candidate with campaign finance-based ideology scores (Bonica, 2014, 2019). Third, we matched candidates that served in Congress or state legislatures to determine their party and roll-call based ideal points (Shor and McCarty, 2011; Lewis et al., 2021). We also matched many candidates in recent elections with their Ballotpedia profiles and Twitter handles. In cases where gender was unavailable from other sources, we imputed it based on first names (Mullen, 2021). When race and ethnicity was unavailable from other sources, we imputed it based on a combination of candidates' pictures and names (Lee and Velez, 2022).[2]

We also augmented the election returns with a variety of information about most candidates' constituencies. We included information about the ideological preferences of each city and county in our dataset (Tausanovitch and Warshaw, 2013). We also included recent presidential election results for most cities and counties (Ansolabehere and Rodden, 2012; MIT Election Data and Science Lab, 2018; Voting and Election Science Team, 2020). In addition, we assembled a new collection of shapefiles for many city council and county legislative districts. This enabled us to estimate presidential election returns in many local governments' district-level constituencies by overlaying precinct-level presidential returns on top of the district shapefiles.

Applications

Here, we demonstrate the variety of substantive research questions on representation that our new dataset can help address using three illustrative applications.

---

[2]To estimate race/ethnicity, we combine features from two models: a Bayesian surname prediction model (Imai and Khanna, 2016) and a pre-trained convolutional neural network model that uses images of public officials to predict race (Parkhi, Vedaldi, and Zisserman, 2015). This approach produces more accurate predictions of race and ethnicity in our sample than existing methods that rely on names and/or geography alone (Lee and Velez, 2022).

Descriptive Representation Across Offices

An important question in the study of American politics is the under-representation of women at various levels of government (e.g. Clayton, O'Brien, and Piscopo, 2019; Barnes and Holman, 2020; Bernhard, Shames, and Teele, 2021; Lawless, 2015). Prior work suggests that women's under-representation in local governments mirrors their underrepresentation in Congress and state legislatures. Studies have found that women hold an average of only 25% of city council seats, 25% of county legislative seats, and 15-20% of mayoral seats (Center for Women in Politics, 2016; Ferreira and Gyourko, 2014; National League of Cities, 2016; McBrayer and Williams, 2022). Moreover, the representation of women in local government appears to have plateaued (Holman, 2017).

However, Holman (2017, 2) cautions that, "limited access to reliable information on women's representation ...[in] local offices limits the conclusions that we can draw about the barriers facing women in seeking political parity." Our dataset expands the previous available information about the gender of local candidates and elected officials by an order of magnitude. It contains information on the gender of 72,896 candidates (93% of those in our data) for seven types of local offices over the past three decades.
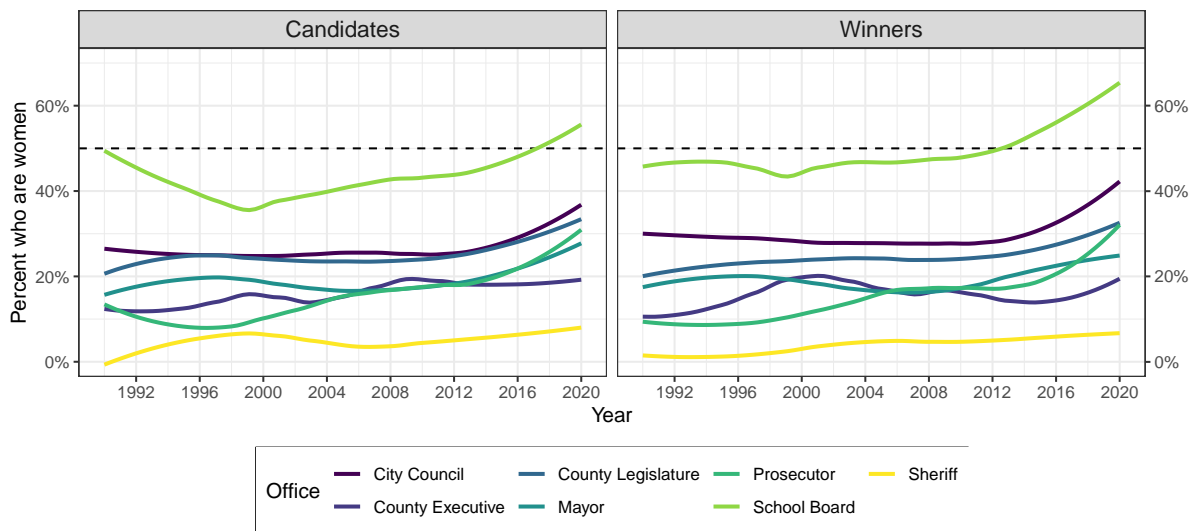

Figure 2: Percent of Candidates and Winners that are Women by Office

Figure 2 (above) shows the proportion of candidates and winners for mayor, city council, county executive, county legislature, sheriff, prosecutor, and school board that were women in each year between 1990-2020. It indicates that women remain woefully under-represented in the majority of local offices, with the percentage of winning candidates under 50% for all offices except school boards. However, the percentage of winning candidates that are women has trended slightly upward over time, especially in elections for school board and city council. Especially interesting are the dramatic differences in patterns across offices in women's representation. Women are most under-represented in sheriff elections, and tend to be best represented in school board elections, in line with recent work on California (Anzia and Bernhard, 2021).
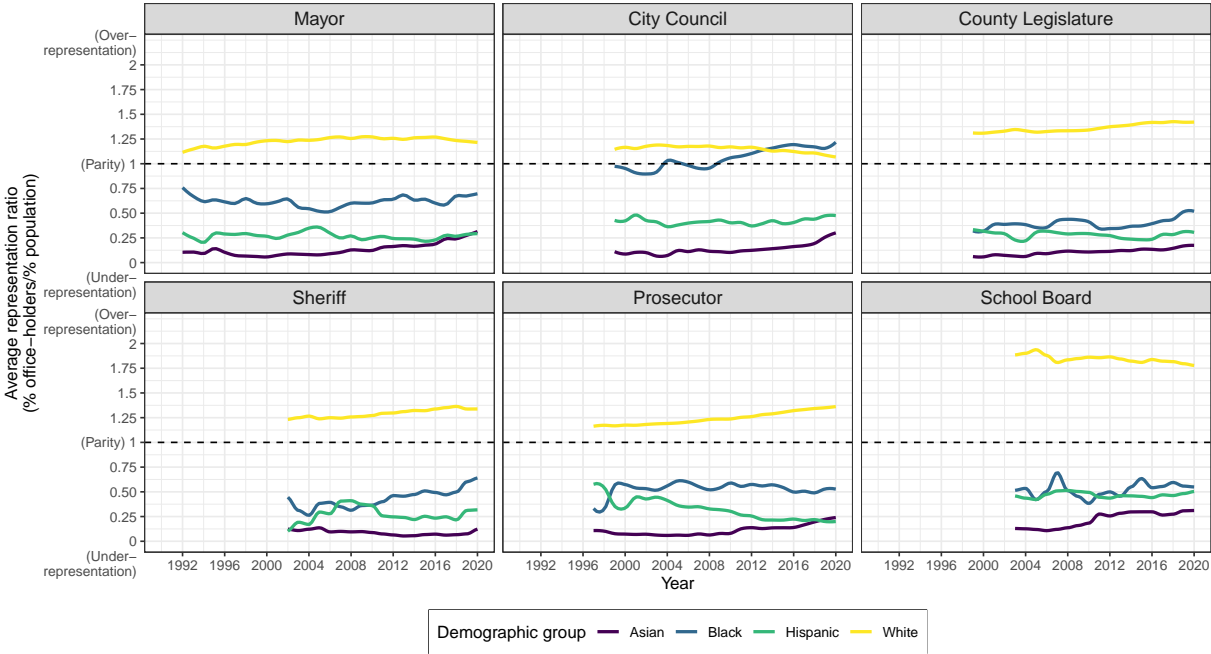
Figure 3: Racial representational by office. Lines indicate smoothed local averages of the ratio between the percent of officeholders and the percent of the population in each racial/ethnic group, and are plotted for years after which our data cover at least 20% of the total jurisdictions for which we have some composition data for that office.

Next, we examine the descriptive representation of racial and ethnic groups in local governments (Grofman and Davidson, 1994; Shah, Marschall, and Ruhil, 2013; Hajnal, 2020; Schaffner, Rhodes, and La Raja, 2020). We analyze the relative representation of Blacks,

Hispanics, and Asian-Americans based on the ratio between the composition of local office-holders of each type and their share of the population. Figure 3 shows this ratio across types of local government offices, and indicates that Blacks, Hispanics, and Asian-Americans are all under-represented in most local governments. Hispanics and Asian-Americans appear to be particularly under-represented. For instance, this ratio indicates that the Hispanic percentage of county legislators is less than a third of the Hispanic percentage of the population in those counties in 2020. Figure 3 also shows that descriptive representation is especially poor among sheriffs, but that city councils, in particular, consist of Black legislators at rates roughly proportional to population demographics.

Nationalization

Increasingly, elections at the state and local level feature policy debates that relate more to national politics (Abramowitz and Webster, 2016; Moskowitz, 2021). As a result, local elections appear to be increasingly nationalized (Hopkins, 2018). Moreover, recent evidence suggests that partisanship – a common feature of national politics – can shape local political communication (de Benedictis-Kessner, 2021) and policy (de Benedictis-Kessner and Warshaw, 2016, 2020b; Holman, Farris, and Sumner, 2020). Our dataset allows scholars to test both the causes and consequences of the nationalization across various levels of office.

One example of the nationalization of American politics is the close tie between partisan voting patterns in national and local elections (Hopkins, 2018; Kuriwaki, 2020). Figure 4 shows the correlations between presidential and local elections across levels of office between 2017 and 2020. As a benchmark, the top-left panel shows that the correlation between presidential results and state legislative elections is about 0.8. The correlations for other offices vary between 0.54 and 0.73. Figure 4 also shows that most local offices appear to have a large pro-Republican bias relative to presidential elections: in places with evenly split presidential results, the Republican candidate is likely to get about 60% of the vote in local elections. Future work might examine whether the pro-Republican bias in local elections

Figure 4: Nationalization in elections across levels of subnational governments, 2017-2020

is affected by institutional features, such as off-cycle elections (Anzia, 2011) or nonpartisan ballots.

Dyadic Representation in City Councils

Local politics scholars have long been hindered in their ability to study dyadic representation in city governments due to the unavailability of district-level data on the preferences of the mass public, the partisanship of city officials, or the outputs of local legislative processes. Scholars of representation have therefore tended to focus on the relationship between the mass public's preferences and the policy outputs of city governments (Tausanovitch and Warshaw, 2014; Einstein and Kogan, 2016). While valuable, this approach does little to unpack the micro-foundations of representation in city governments.

Here, we show how our data can be used to examine dyadic representation in one form of

(a) US House (2019-2020)   (b) State House (2017-2018)

(c) New York, NY   (d) Philadelphia, PA   (e) Phoenix, AZ

(f) Chicago, IL   (g) Boston, MA   (h) San Diego, CA

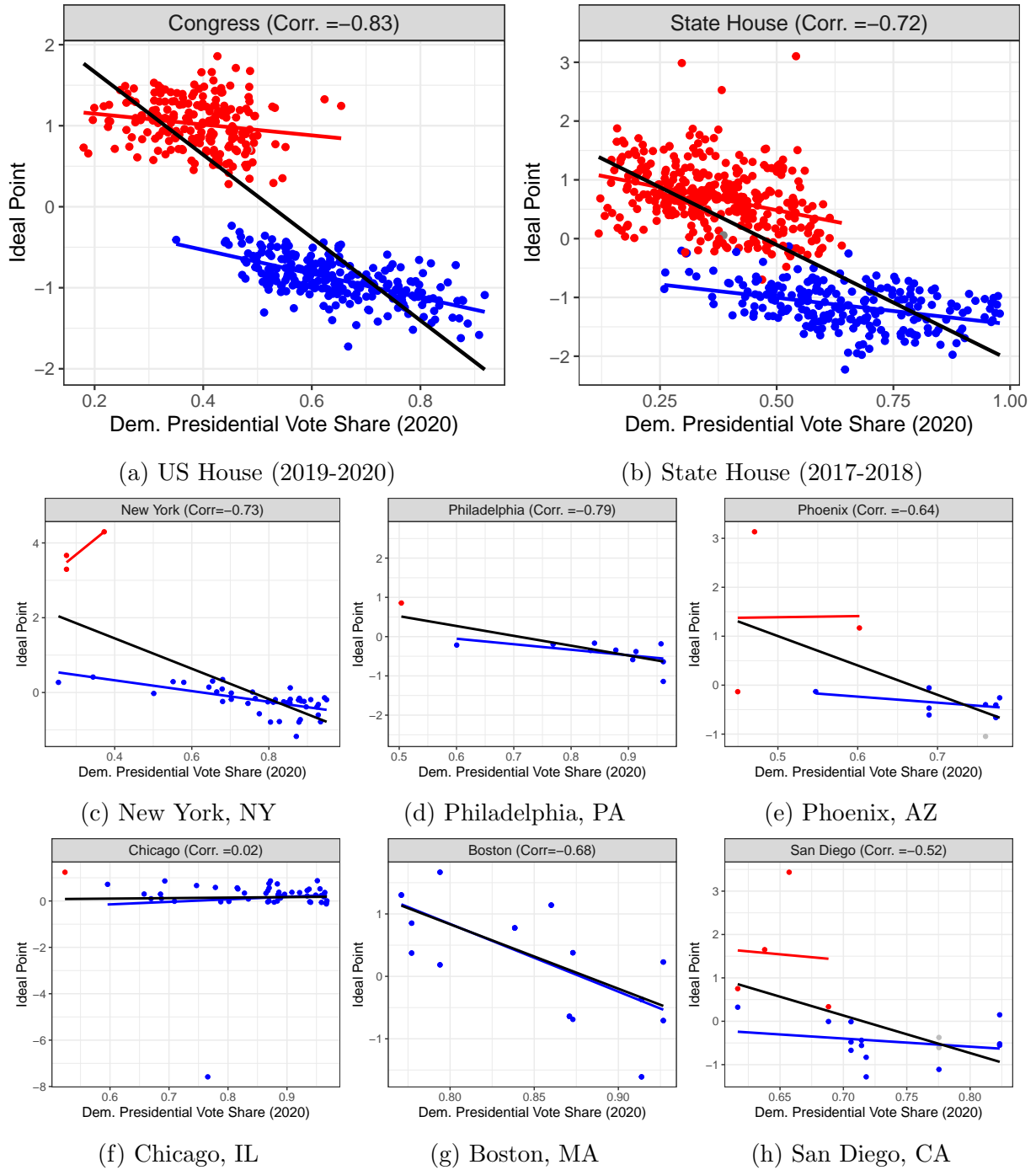Figure 5: Dyadic Representation in City Councils. Blue dots show Democratic legislators and red dots show Republican legislators. Lines represent linear regressions of the relationship between presidential vote and legislators' ideal points, for all legislators (black lines) and within-party (blue and red lines).

local legislatures. We merged data on presidential election returns at the city council district

level with estimates of city councilors' ideal points and partisanship in 6 large cities.[3] As a benchmark for representation in legislatures, the top row of Figure 5 shows the relationship between district preferences and the ideal points of members of Congress and state legislators (Lewis et al., 2021; Shor and McCarty, 2011). In these institutions, the correlation between presidential vote and legislators' ideal points ranges from 0.72 to 0.84. The graphs also show that the difference between the ideal points of Democratic and Republican legislators is about 1.5-2 standard deviations.

The other panels in Figure 5 show the relationship between district preferences and the ideal points of city councilors in New York City, Philadelphia, Phoenix, Chicago, Boston, and Dallas. In four of these cities, the correlation between district preferences and city councilors' ideal points is nearly as high as in Congress and state legislatures. In one city (Chicago), there no relationship between district preferences and councilors' ideal points. There is also about a 1.5-standard deviation difference between the ideal points of Democratic and Republican legislators in cities that elect multiple councilors from each party. The existence of stronger correlations between presidential vote share and councilors' ideal points in cities that elect councilors from both parties suggests that representation may be stronger in these places. Yet there is also a clear association between district preferences and councilors' ideal points in Boston, which has an all-Democratic city council. Overall, this analysis suggests that dyadic representation between the mass public and city councilors appears to be reasonably strong in many large cities.

This analysis also points toward a number of future pathways to expand the study of representation in city governments. As more data on the roll call voting patterns in city governments becomes available, it will be possible to use our data to examine the institutional factors that affect representation in city governments, such as partisan elections (Bucchianeri, 2020), at-large elections (Hankinson and Magazinnik, 2021), and off-cycle elections (Anzia, 2011).

---

[3]We obtained raw roll call data from Bucchianeri (2020) and collected supplementary data by scraping Legistar. We then scaled the one-dimensional ideal points using the `ideal` package in `R`.

Conclusion

This paper has introduced a new comprehensive dataset of local election returns in medium and large cities and counties. It includes information on 77,105 unique candidates from 54,689 contests in over 1,500 cities and counties. Moreover, it spans a wide variety of local elections, including mayors, city councilors, county legislators, and sheriffs.

The dataset has a bevy of potential applications for the study of democracy in the United States. We have illustrated three of these applications: descriptive representation, nationalization of elections, and dyadic representation in local legislatures. There are many other topics on local politics and elections that this dataset enables scholars to examine. It will facilitate a wide array of research on the factors that affect elections for local offices, as well as their similarities and differences with elections at other levels of governments. It will also enable scholars to study the consequences of these elections for the policymaking process. These new data, combined with future efforts, will help build a more holistic understanding of democratic representation in the United States – of which *local* elections are one important part.

References

Abramowitz, Alan I, and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of US Elections in the 21st Century." *Electoral Studies* 41: 12–22.

Ansolabehere, Stephen, and Jonathan Rodden. 2012. "Harvard Election Data Archive." Available at `http://projects.iq.harvard.edu/eda`.

Anzia, Sarah F. 2011. "Election Timing and the Electoral Influence of Interest Groups." *Journal of Politics* 73(2): 412–427.

Anzia, Sarah F. 2019. "Looking for Influence in All the Wrong Places: How Studying Subnational Policy Can Revive Research on Interest Groups." *Journal of Politics* 81(1): 343–351.

Anzia, Sarah F, and Rachel Bernhard. 2021. "Gender Stereotyping and the Electoral Success of Women Candidates: New Evidence from Local Elections in California." *British Journal of Political Science* (forthcoming).

Arnold, R Douglas, and Nicholas Carnes. 2012. "Holding Mayors Accountable: New York's Executives from Koch to Bloomberg." *American Journal of Political Science* 56(4): 949–963.

Barnes, Tiffany D., and Mirya R. Holman. 2020. "Gender Quotas, Women's Representation, and Legislative Diversity." *Journal of Politics* 82(4): 1271–1286.

Bernhard, Rachel, and Justin de Benedictis-Kessner. 2021. "Men and Women Candidates Are Similarly Persistent After Losing Elections." *Proceedings of the National Academy of Sciences* 118(26).

Bernhard, Rachel, Shauna Shames, and Dawn Langan Teele. 2021. "To Emerge? Breadwinning, Motherhood, and Women's Decisions to Run for Office." *American Political Science Review* 115(10): 379–394.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–386.

Bonica, Adam. 2019. "Database on Ideology, Money in Politics, and Elections: pre-release version 3.0.".

Bucchianeri, Peter. 2020. "Party Competition and Coalitional Stability: Evidence from American Local Government." *American Political Science Review* 114(4): 1055–1070.

CEDA. 2020. "California Elections Data Archive (CEDA)." `https://csu-csus.esploro.exlibrisgroup.com/esploro/outputs/99257830890201671`.

Center for Women in Politics. 2016. "Women in Elective Office 2016." Available at `http://www.cawp.rutgers.edu/women-elective-office-2016`.

Clayton, Amanda, Diana Z. O'Brien, and Jennifer M. Piscopo. 2019. "All Male Panels? Representation and Democratic Legitimacy." *American Journal of Political Science* 63(1): 113–129.

de Benedictis-Kessner, Justin. 2018. "Off-Cycle and Out of Office: Election Timing and the Incumbency Advantage." *Journal of Politics* 80(1): 119–132.

de Benedictis-Kessner, Justin. 2021. "Strategic Partisans: Electoral Motivations and Partisanship in Local Government Communication." *Journal of Political Institutions and Political Economy* 2(2): 227–248.

de Benedictis-Kessner, Justin, and Christopher Warshaw. 2016. "Mayoral Partisanship and Municipal Fiscal Policy." *Journal of Politics* 78(4): 1124–1138.

de Benedictis-Kessner, Justin, and Christopher Warshaw. 2020*a*. "Accountability for the Local Economy at All Levels of Government in United States Elections." *American Political Science Review* 114(3): 660–676.

de Benedictis-Kessner, Justin, and Christopher Warshaw. 2020*b*. "Politics in Forgotten Governments: The Partisan Composition of County Legislatures and County Fiscal Policies." *Journal of Politics* 82(2): 460–475.

Einstein, Katherine Levine, and Vladimir Kogan. 2016. "Pushing the City Limits: Policy Responsiveness in Municipal Government." *Urban Affairs Review* 52(1): 3–32.

Ferreira, Fernando, and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence from US Cities." *Quarterly Journal of Economics* 124(1): 399–422.

Ferreira, Fernando, and Joseph Gyourko. 2014. "Does Gender Matter for Political Leadership? The Case Of US Mayors." *Journal of Public Economics* 112: 24–39.

Gerber, Elisabeth R, and Daniel J Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55(2): 326–339.

Grofman, Bernard, and Chandler Davidson. 1994. "The Effect of Municipal Election Structure on Black Representation in Eight Southern States." In *Quiet Revolution in the South: The Impact of the Voting Rights Act, 1965-1990*, ed. Chandler Davidson, and Bernard Grofman. Princeton, NJ: Princeton University Press pp. 301–334.

Hajnal, Zoltan L. 2020. *Dangerously Divided: How Race and Class Shape Winning and Losing in American Politics.* New York: Cambridge University Press.

Hankinson, Michael, and Asya Magazinnik. 2021. "The Supply-Equity Trade-off: The Effect of Spatial Representation on the Local Housing Supply." *Journal of Politics* (forthcoming).

Hessick, Carissa Byrne, and Michael Morse. 2019. "Picking Prosecutors." *Iowa L. Rev.* 105: 1537.

Holman, Mirya R. 2017. "Women In Local Government: What We Know and Where We Go From Here." *State and Local Government Review* 49(4): 285–296.

Holman, Mirya R, Emily M Farris, and Jane Lawrence Sumner. 2020. "Local Political Institutions and First-Mover Policy Responses to COVID-19." *Journal of Political Institutions and Political Economy* 1(4): 523–541.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized.* Chicago, IL: University of Chicago Press.

Hopkins, Daniel J, and Lindsay M Pettingill. 2018. "Retrospective Voting in Big-City US Mayoral Elections." *Political Science Research and Methods* 6(4): 697–714.

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records." *Political Analysis* 24(2): 263–272.

Krumholz, Sam. 2019. "The Effect of District Attorneys on Local Criminal Justice Outcomes." Working paper. Online: `https://ssrn.com/abstract=3243162`.

Kuriwaki, Shiro. 2020. "Ticket Splitting in a Nationalized Era." Available at `https://osf.io/preprints/socarxiv/bvgz3/`.

Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18: 349–366.

Lee, Diana Da In, and Yamil Ricardo Velez. 2022. "Measuring inequality at scale: the use and misuse of ethnoracial predictions." *Poster presented at the 2022 Society for Political Methodology conference, St. Louis, Missouri.* .

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2021. "Voteview: Congressional Roll-Call Votes Database." Available at `https://voteview.com/`.

McBrayer, Markie, and Robert Lucas Williams. 2022. "The Second Sex in the Second District: The Policy Effects of Electing Women to County Government." *Political Research Quarterly* (forthcoming).

MIT Election Data and Science Lab. 2018. "County Presidential Election Returns 2000-2016.".

Moskowitz, Daniel J. 2021. "Local News, Information, and the Nationalization of US Elections." *American Political Science Review* 115(1): 114–129.

Mullen, Lincoln. 2021. "gender: Predict Gender from Names Using Historical Data." R package version 0.6.0, https://github.com/lmullen/gender.

National League of Cities. 2016. "National League of Cities: City Councils." Report available at `https://www.nlc.org/city-councils`.

Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep Face Recognition. In *Proceedings of the British Machine Vision Conference*, ed. Xianghua Xie, Mark W. Jones, and Gary K. L. Tam. pp. 41.1–41.12.

Payson, Julia A. 2020. "The Partisan Logic of City Mobilization: Evidence from State Lobbying Disclosures." *American Political Science Review* 114(3): 677–690.

Sances, Michael W. 2018. "Ideology and Vote Choice in US Mayoral Elections: Evidence from Facebook Surveys." *Political Behavior* 40(3): 737–762.

Schaffner, Brian F, Jesse H Rhodes, and Raymond J La Raja. 2020. *Hometown Inequality: Race, Class, and Representation in American Local Politics*. New York: Cambridge University Press.

Shah, Paru R, Melissa J Marschall, and Anirudh VS Ruhil. 2013. "Are We There Yet? The Voting Rights Act and Black Representation on City Councils, 1981–2006." *Journal of Politics* 75(4): 993–1008.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 205(3): 530–551.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–342.

Tausanovitch, Chris, and Christopher Warshaw. 2014. "Representation in Municipal Government." *American Political Science Review* 108(3): 605–641.

Thompson, Daniel M. 2020. "How Partisan is Local Law Enforcement? Evidence from Sheriff Cooperation with Immigration Authorities." *American Political Science Review* 114(1): 222–236.

Trounstine, Jessica. 2010. "Representation and Accountability in Cities." *Annual Review of Political Science* 13: 407–423.

Trounstine, Jessica. 2011. "Evidence of a Local Incumbency Advantage." *Legislative Studies Quarterly* 36(2): 255–280.

Voting and Election Science Team. 2020. "2020 Precinct-Level Election Results." Available at `https://doi.org/10.7910/DVN/K7760H`.

Warshaw, Christopher. 2019. "Local Elections and Representation in the United States." *Annual Review of Political Science* 22: 461–479.

Appendix for "Local Representation in the United States: A New Comprehensive Dataset of Elections"

August 4, 2022

A   Components of Elections Dataset

In this appendix, we describe the components of the elections dataset in more depth. The dataset includes two sets of files. First, we include candidate-level data. Second, we include constituency-level data at the level of cities, counties, city council districts, and county legislative districts.

1. **Candidate-contest level variables**:

   - **candid_fL**: Unique candidate identifier based on probabilistic name-matching within jurisdictions using `fastLink` as described below.

   - **Name**: Names are generally based on the official election returns.

   - **Votes**: The number of votes received by each candidate is based on a variety of sources, including administrative data from government websites, crowd-sourced data on OurCampaigns.org, California Election Data Archive (CEDA), and in a few cases, newspaper archives.

   - **Vote share**: The candidate's vote share in the election.

   - **Democratic vote share**: The Democratic vote share in the election.

   - **Number of winners**: The number of winners for each seat. In single-member districts (SMDs), this will be 1.

   - **Winner**: Whether the candidate won the election.

   - **CF-Score**: We match candidates to Adam Bonica's DIME database of contributors to federal and state election candidates based on name, city, and state (Bonica, 2014, 2019). We resolve duplicates primarily based on occupations (prioritizing people that indicate their occupation is an elected official or lawyer). We include their contributor CF-Score and Bonica CID. For matches whose CF-Score is projected (i.e. based on only a single contribution) we do not include the score.

- **Party**: In partisan elections, candidates' partisanship is based on official election returns. In non-partisan elections, we use a variety of sources to impute partisanship in the following order of priority. First, we use crowd-sourced data on OurCampaigns.org. Next, we match candidates' to the voter file and use their party of registration there. Third, we match candidates to their CF-Scores and assume that candidates with CF-Scores less than -.75 are Democrats and CF-Scores greater than .75 are Republicans. Fourth, we match to candidates' Ballotpedia pages. Finally, we try to match candidates to other partisan elections such as state legislative elections and use their party from that election.

- **Gender**: When possible, we use gender from voter file records. If gender is unavailable there, we impute gender using the `gender` package in `R`, as described in Blevins and Mullen (2015). We only use a gender imputation if there is at least a .95 probability that the assigned gender is correct.

- **Race/Ethnicity**: To estimate race and ethnicity, we combine features from both the voter file matches and two models: a Bayesian surname prediction model (Imai and Khanna, 2016) and a pre-trained convolutional neural network model that uses images of public officials from search engine results to predict race (Parkhi, Vedaldi, and Zisserman, 2015). This approach produces more accurate predictions of race and ethnicity in our sample than existing methods that rely on names and/or geography (Lee and Velez, 2022). We include a race/ethnicity variable from our voter file matches (`race`), and probability values for race/ethnicity categories from both image-based methods (`img.whi`, `img.bla`, etc.) and surname-based methods (`s.whi`, `s.bla`, etc.) as well as a combined final variable that we use for the analyses in the paper (`race_final`), which uses voter file matches and then uses the combined image- and surname-based method in cases where the race/ethnicity from voter file matches was unknown or missing and both image- and surname-based methods agreed. We plot the average composition of each office over time from this race-coding method in Appendix B. We also use predictions from a multinomial logit model trained on both the surname- and image-based probability values of crowdsource-validated race-codings of city councilors to further extend our race-coding following Lee and Velez (2022) and provide these predicted values for candidates (`mnl.whi`, `mnl.bla`, etc.). To estimate the prevalence of racial groups among all office-holders and candidates in the main manuscript, we also use probability values that take into account all of these methods of race-coding: in instances where the `race_final` variable has a value (i.e. either the voter file produced a match, or the surname and image-based methods agreed), we provide binary (1/0) values of racial group membership in the variables `prob_whi`, `prob_bla`, etc., and in instances where surname- and image-based methods disagree, we produce these probabilities directly from the multinomial logit predictions.

- **Incumbency Status**: Whether the candidate is an incumbent. We assign incumbency status by matching candidates' across contest-years within a given office and place (i.e. city, county, or school district) using a probabilistic name-matching process implemented using the `fastLink` package in `R`, as described in

Enamorado, Fifield, and Imai (2019). This variable is missing in the first 4 years in which we have election data in each individual place since we could not determine whether candidates were new (non-incumbents) vs. incumbents without a previous election cycle.

2. **Constituency-level variables**: Where possible, we include an array of constituency level variables. These are available for nearly all cities and counties. We also have them available at the city council district-level in about 150 cities and the county legislative district level in about 130 counties.

   - **Number of winners**: The number of winners for each seat. In single-member districts (SMDs), this will be 1.
   - **2020 Population**: Based on the 2020 Census.
   - **Percent Black (2019)**: Based on the 2019 5-year ACS.
   - **Percent Hispanic (2019)**: Based on the 2019 5-year ACS.
   - **Percent Asian-American (2019)**: Based on the 2019 5-year ACS.
   - **Mass Ideology**: A cross-sectional measure of mass ideology based on Tausanovitch and Warshaw (2013). Only available at the city and county-level.
   - **Presidential vote**: Based on precinct-level data on the 2008 presidential vote (Ansolabehere and Rodden, 2012), 2016 presidential vote, and 2020 presidential vote (MIT Election Data and Science Lab, 2018; Voting and Election Science Team, 2020). These are available for nearly all cities and counties. In addition, we assembled a new collection of GIS files for many city council and county legislative districts. We collected city council district and census designated place shape files for cities that hold ward based and at-large elections, respectively. This enabled us to estimate presidential election returns in many district-level constituencies of local governments.

B   Alternative Visualization of Racial Composition of Officeholders

In the main manuscript, we present estimates of the "representational ratio" of officeholders from various racial/ethnic groups relative to their percentage of the population. In this appendix we also present the raw percent of officeholders from each racial/ethnic group in Figure B-1 below.
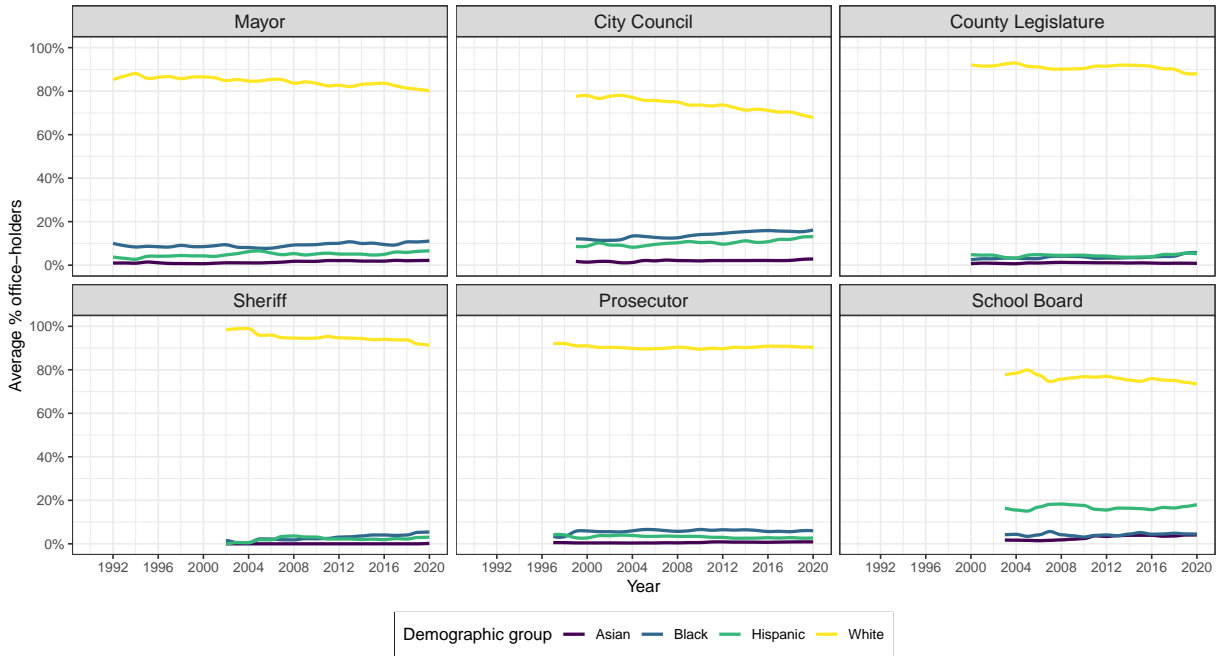

Figure B-1: Racial representation by Office. Lines indicate smoothed local averages of the percent of officeholders in each racial/ethnic group, and are plotted for years after which our data cover at least 20% of the total jurisdictions for which we have some composition data for that office.

References

Ansolabehere, Stephen, and Jonathan Rodden. 2012. "Harvard Election Data Archive." Available at `http://projects.iq.harvard.edu/eda`.

Blevins, Cameron, and Lincoln Mullen. 2015. "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction." *DHQ: Digital Humanities Quarterly* 9(3).

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–386.

Bonica, Adam. 2019. "Database on Ideology, Money in Politics, and Elections: pre-release version 3.0.".

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. 2019. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records." *American Political Science Review* 113(2): 353–371.

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records." *Political Analysis* 24(2): 263–272.

Lee, Diana Da In, and Yamil Ricardo Velez. 2022. "Measuring inequality at scale: the use and misuse of ethnoracial predictions." *Poster presented at the 2022 Society for Political Methodology conference, St. Louis, Missouri.* .

MIT Election Data and Science Lab. 2018. "County Presidential Election Returns 2000-2016.".

Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep Face Recognition. In *Proceedings of the British Machine Vision Conference*, ed. Xianghua Xie, Mark W. Jones, and Gary K. L. Tam. pp. 41.1–41.12.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–342.

Voting and Election Science Team. 2020. "2020 Precinct-Level Election Results." Available at `https://doi.org/10.7910/DVN/K7760H`.