Research Note

Gender, Race, Age and Voting: A Research Note

Stephen Ansolabehere [1] and Eitan Hersh [2]

[1] Department of Government, Harvard University, 1737 Cambridge St., Cambridge, MA 02138, USA;
E-Mail: sda@gov.harvard.edu
[2] Department of Political Science, Yale University, 77 Prospect St., New Haven, CT 06520, USA;
E-Mail: eitan.hersh@yale.edu; Tel.: +1 2034369061

How to Cite this Article

Acknowledgement

About the Journal

Politics and Governance is an innovative new offering to the world of online publishing in the Political Sciences. An internationally peer-reviewed open access journal, *Politics and Governance* publishes significant, cutting-edge and multidisciplinary research drawn from all areas of Political Science.

www.cogitatiopress.com/politicsandgovernance

Editors-in-Chief

Professor Andrej J. Zwitter, Faculty of Law, University of Groningen, The Netherlands
Professor Amelia Hadfield, Department of Psychology, Politics and Sociology, Canterbury Christ Church University, UK

Managing Editor

Mr. António Vieira, Politics and Governance, Cogitatio Press, Portugal

Politics and Governance

Research Note

Gender, Race, Age and Voting: A Research Note

Stephen Ansolabehere[1] and Eitan Hersh[2,]*

[1] Department of Government, Harvard University, 1737 Cambridge St., Cambridge, MA 02138, USA;
E-Mail: sda@gov.harvard.edu
[2] Department of Political Science, Yale University, 77 Prospect St., New Haven, CT 06520, USA;
E-Mail: eitan.hersh@yale.edu; Tel.: +1 2034369061

* Corresponding author

Abstract: In this brief analysis, we use a new dataset of two million voter registration records to demonstrate that gender, race, and age do not correlate with political participation in the ways that previous research has shown. Among Blacks and Latinos, women participate at vastly higher rates than men; many Blacks participate at higher rates than Whites; and the relationship between age and participation is both not linear and varies by race and gender. Survey research is unable to capture the true relationship between demographics and participation on account of survey bias and, more importantly, the non-linearity of effects. As a result, theories of participation, like the dominant resources-based models, have been built on faulty premises and tested with inadequate data. Our evidence calls for a renewed effort to understand election participation by utilizing large datasets, by being attentive to linearity assumptions, and by returning to theory.

Keywords: big data; gender; race; SES; U.S. elections; voter registration; voting

1. Introduction

Using a sample of two million registration records, we demonstrate that the ways that race, gender, and age correlate with political participation in the United States are misunderstood. Political scientists have long used survey samples of 1,000 to 15,000 respondents to study political participation. Because of the predictors of self-reported behavior that routinely arise in such surveys, scholars have inferred that a certain set of social and political processes must drive participation. Specifically, arguments that people's "resources" drive participation dominate the field. These arguments predict that people who are disadvantaged or discriminated against in society are less likely to get involved (e.g [1,2]). Most notably, African Americans, as a group, Hispanics, as a group, and women, as a group, are claimed to have lower participation rates. Such inferences are usually drawn from surveys in which there are only a few hundred racial minorities, and the numbers of minorities within specific subgroups (such as low education) is extremely small.

We show that some of the most basic facts that led scholars to formulate the resource theories are wrong. Survey research finds that Whites participate in politics more than Blacks and men participate in elections

librello

at similar or, even higher, rates than women. According to Verba, Burns, and Schlozman, there is a "well known equivalence between men and women in electoral turnout" [3]. Participation is also thought to increase linearly or quadratically with age within each social group. Each of these findings and assumptions will be shown to be incorrect in ways that may call for a rethinking of the resource-based theories themselves. The actual voting behaviors of subgroups deviates from the received wisdom in ways that contradict the simple predictions of "resource models". Three striking facts stand out. First, there is no appreciable effect of age on registration for people older than 30 (see [4]). Second, women registered and voted in 2008 at higher rates than men, and for racial minorities the differences are substantial. Third, African American women registered at rates exceeding those for white men within nearly every age category, even after controlling for demographic factors. Such patterns are wholly surprising from a resource model perspective.

The results here suggest that inferences based on reported registration and turnout from relatively small sample surveys will typically fail to find statistically significant results even for the very substantial interaction effects between race and gender. In order to determine whether there are large differences between men and women within racial groups requires very large samples, and without large samples researchers may be prone to committing an inferential error akin to Simpson's Paradox.

2. Data

In 2010, we partnered with Catalist, a data vendor to Democratic politicians and liberal interest groups that collects voter registration records, cleans and updates them, and provides a database of all registered voters in the U.S. to its political clients (see [5] for details on the compilation of these data). For the present study, we use Catalist's national sample of 1% of all active registration records. This yields a sample of roughly 1.9 million registrants.

In most jurisdictions, age and gender are available on the voter file. When gender is unavailable, it can be predicted by registrants' first names. Voters in eight Southern states report their race in the public record. For voters in other states, Catalist makes a prediction of race based on voters' names and Census block contextual data. Catalist's precise method for predicting race is proprietary, but in a study in which we tested the accuracy of Catalist's racial predictions, we found that 91% of the time Catalist's prediction of race matches a voter's self-reported race. In this analysis, we use Catalist's racial predictions for the entire sample; however, if we were to restrict the sample to voters who report their race on the voter file, or to voters whose races are predicted with the highest degree of confidence, the results do not meaningfully change.

Catalist's records provide estimates of registrants

and voters. For this study, we focus on turnout in the 2008 general election as recorded on voter files by election officials [6]. However, the effects identified are not particular to 2008. We also utilize citizen population estimates as a baseline for participation. To create such estimates, we start with the Census Bureau's full population estimates by gender, race, and single year of age for 2009 [7]. We then use the American Community Survey (ACS) to generate the percentage of each age-gender-race cohort that are citizens, and we weight the population statistics accordingly [8]. The ACS has a sample size of three million residents, and is sufficiently large to provide accurate citizen estimates for all age-race-gender cohorts.

We analyze the registration and voting rates of each gender-racial group (e.g., white males or Hispanic females) within each age cohort. One reason for doing so is that minorities are younger, and younger people register and vote at lower rates. Hence, we control for age in the analysis in order to guard against finding a spurious difference that, in fact, reflects differences in ages. Another reason for presenting the data for each age cohort graphically is that Ansolabehere, Hersh, and Shepsle derive an alternative specification for the correlation between age and registration [4]. This specification is not linear or quadratic but logarithmic, and is driven entirely by residential mobility (not a causal connection between aging and registering). The analysis here is consistent with that model.

The Catalist data provide a distinctive advantage in the study of voting and registration. Other research projects have examined the relationship between gender, race, and participation using data from the Voting and Registration Supplement of the Current Population Survey. The CPS relies on reported registration and turnout and there are well-known biases with such data [9-11]. And, even with a sample of approximately 75,000 persons, the power of the CPS is limited by the sample sizes of the smallest groups in which comparisons are made, such as black men ages 18–24. The CPS only has the power to detect differences across race and gender within fairly coarse age groups.

One important caveat concerning the present study is in order. The 2008 election is distinctive in many ways. This was the first election in which one of the major parties nominated an African American for President, and Barack Obama won the election in no small part because of the sizable minority vote that year [12]. Turnout increased from 54% of the Voting Age Population in 2004 to 59% in 2008, and minorities and younger voters exhibited especially large increases in turnout, though all racial groups increased their participation rates according to national Exit Polls [13]. At the time of this writing (July 2013) the 2012 vote history data are not yet fully recorded in the Catalist system. Preliminary analysis of Catalist data for 2012 show patterns similar to those discussed here, albeit at a somewhat lower rate of participation across all groups. For example, in the state of Florida, black

women voted at the highest rate of any race-gender cohort, with 72% of registered individuals voting in 2012. In comparison, only 67% of registered white men participated in the 2012 election, according to records from Catalist.

3. Results and Discussion

Table 1 shows the rate of registration among citizens, the rate of voting among registrants, and the rate of voting among citizens, by gender and race. In all racial groups, women are registered to vote at higher rates than men. The gap differs by race: it is 2–3 points for Whites and Asians, but 14–16 percentage points for Hispanics and Blacks. Given that a person is registered, there is a separate effect of gender on voting, with women in all racial groups voting at higher rates. Again, these differences are small for Whites and Asians and large for Blacks and Hispanics. The two gender effects together (the effect on registration and the effect on turnout) contribute to the overall participation gap, as witnessed in the third section of the table. The largest gender gap is among Blacks: black women are 17 percentage points more likely to vote than black men. Apart from identifying substantially larger gender effects than previously found, the evidence in Table 1 also challenges the assumption that Whites participate more than Blacks. In fact, black women register at a higher rate than any other group, and overall black women vote at a higher rate than white men.

Figure 1 adds age into the analysis. For each race-gender cohort, the rate of registration among citizens (column 1), the rate of voting among registrants (column 2), and the rate of voting among citizens (column 3) are shown. The figures estimate voters in 3-year age cohorts, beginning with 18–21 year olds. The gender gap is somewhat higher among young people, especially for Blacks, but it is persistent across most age groups.

The relationship between age and participation is very non-linear, and it varies substantially by racial group. The evidence presented holds important methodological and substantive lessons. Methodologically, most social science surveys lack the statistical power to measure the non-linear relationships detected here. This paper estimates participation rates for twenty-one age cohorts, four racial groups, and two genders, or 168 cells. The results reveal a strong three-variable interaction. The 2008 National Election Study, which contained an over-sample of black respondents, included only 238 black men. Before even considering issues of vote misreporting and sample selection, the NES would not have the power to detect even sizeable interactions of race and gender. Similarly, Burns, Schlozman, and Verba report regression coefficients that in fact show a large interaction between race and gender, but they discount the results because of the standard error [14]. Larger samples like the Current Population Survey (CPS) can be more helpful, though these sur-

veys still face the problems of misreporting and sample selection [9], and even the CPS cannot measure with much accuracy the behavior of relatively small groups, such as Asians.

Without the standard modeling assumption that demographics like age, race, and gender affect election participation identically across groups, and having enough data to examine sub-groups, we need not turn to regression in this analysis. Even if our analysis observed just five age cohorts rather than twenty-one, a regression table interacting race, gender, and age would necessitate 40 coefficient estimates, and that would be before incorporating any other control variables that one might consider important.

Table 1. Mean voting and registration rates by race and gender.

Registration among Citizens		
	Male	Female
White	86.2	89.1
Black	75.1	90.9
Hispanic	67.4	81.4
Asian	50.7	52.6
Voting Among Registrants		
	Male	Female
White	70.9	72.7
Black	61.0	69.0
Hispanic	59.1	64.3
Asian	61.0	64.1
Voting Among Citizens		
	Male	Female
White	61.1	64.7
Black	45.9	62.7
Hispanic	39.8	52.3
Asian	30.9	33.8
Obs. (Registrants in Sample)		
	Male	Female
White	614,771	681,196
Black	87,409	123,441
Hispanic	65,493	81,370
Asian	16,843	19,151

Note: Voting, registration, race and gender statistics are from a 1% sample of active voter registration records, drawn in 2010 from Catalist. Racial identification is predicted from a voter's first and last names and Census block group characteristics when not available on the public registration file. 'Voting Among Registrants' statistics are estimated entirely from Catalist's data. For citizen estimates, we start with the Census full population estimates by race, gender, and age. For each race-gender-age combination, we use the American Community Survey to estimate the number of individuals who are citizens. By taking each comparable race-gender-age group in the Catalist 1% sample of registration records, multiplying by 100 and dividing by the Census counts, we arrive at estimates of citizen registration rates and citizen vote rates.
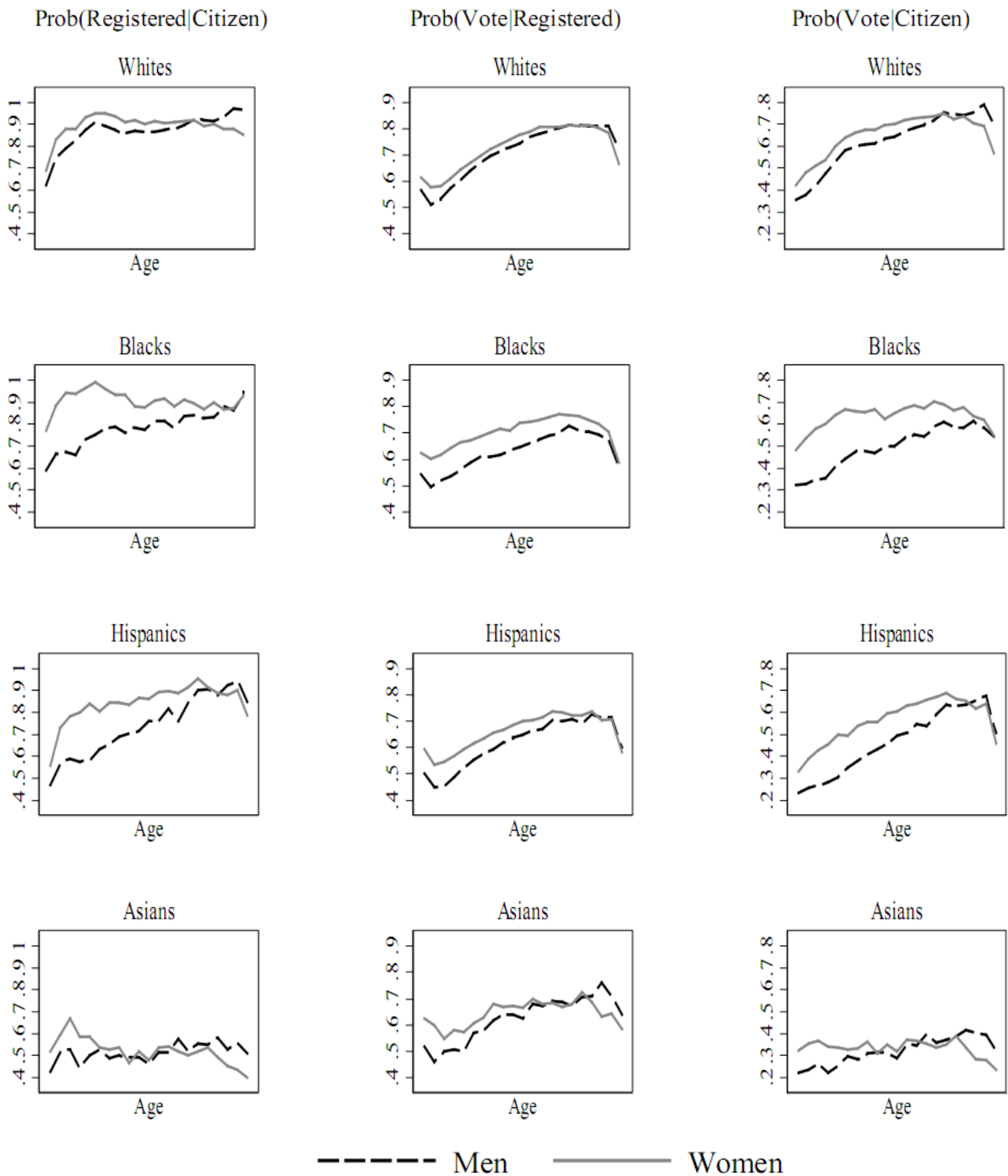
Prob(Registered|Citizen) · Prob(Vote|Registered) · Prob(Vote|Citizen)

Figure 1. Voting and registration by race, gender, and age.

The methodological take-away is not that scholars must estimate fully-interacted models or else their results are unreliable and, since most studies do not use data sources with two million observations, all hope is lost. Rather, the methodological lesson is that theory must guide the modeling process. When empirical tests of theoretical models are susceptible to non-linearities and group-size confounders, as they are in regard to the resources model of participation, then models should be estimated to capture the nuances of the relationship.

In substantive terms, the data show a surprising interaction between gender and race that runs contrary to most work on this subject. Most studies of gender and participation offer theories of why men participate in politics more than women, not the other way around.

Some writers suggest that men have a greater taste for politics than women (e.g. [3,14]). This blanket argument cannot explain why black and Hispanic women participate in elections at vastly higher rates than men. There is also some debate over the basic facts. Some scholars claim that women participate less, or that the participation rates by gender are equivalent; other scholars have found evidence in surveys that women participate slightly more than men (e.g. [15,16]), and a few scholars have noticed that black women report higher turnout than black men (e.g. [15-17]). Among Latinos, the previous evidence has been mixed about differential participation in voting between men and women (see [18,19]). Looking at a national sample of voting data, we see that black and Hispanic women vote a much higher rates than black and Hispanic men, and the difference is much greater than for Whites. Moreover, young black women participate at even higher rates than young white women. It is difficult to square these patterns with resource-based theories, and even more difficult to reconcile them with arguments that men are more attracted to politics than women.

There are several obvious potential explanations for the large gaps in participation between minority women and men. One suggestion is felony disenfranchisement laws, which disproportionately affect minority men. While these laws deserve further attention, a state- by-state analysis of the Catalist data reveals that in states with the most liberal felony disenfranchisement laws, the gender gap is even greater than in states with harsher laws [20,21]. Moreover, note that there is a gender gap among minorities even when the analysis is conditional on currently registered voters. Presumably, disenfranchised current and former felons are typically removed from registration rolls.

Socio-economic status and marriage rates hint at other hypotheses. Black and Hispanic men are less likely to be well-educated, wealthy, and married, and this in turn may lead to lower participation rates. Our own preliminary analyses discredit this view, as we find that at all socio-economic levels there is a substantial gender participation gap among minorities, at all socio-economic levels there is only a modest gender gap among Whites, and the gap exists among married as well as unmarried minorities [22]. A third potential explanation is the religious and community involvement of minority women. Since the late 1800s, black churches have been mostly female [23-25]. While church attendance and community participation may lead to political participation, such an explanation is complicated by endogeniety. We do not know whether community participation causes political participation or whether some other factor affects both, and we suspect the latter (see [26]).

Whatever the explanation, the results here point to a systematic problem in research on race, gender, and voting. The true relationship between race and gender and registration and voting as discovered in the Catalist database has strongly interactive effects. These interactions are sufficiently large that if they are not properly modeled, researchers are likely to make incorrect predictions and draw incorrect inferences (and not just about these variables but about other variables in a multivariate analysis). Most empirical researchers, however, ignore these interactions entirely and estimate linearly separable effects of race and of gender. This is not an intentional error; it is merely a problem of statistical power. The survey research tools that have been at the fingertips of most Political Science researchers for over half a century lack the power to detect even very large interactive effects between race and gender and across age groups.

We are at an historical moment in political science. We can study political participation not by administering surveys but by observing the full population based on official records and consumer profiles. Upon doing so, the patterns of participation complicate the conclusions drawn from surveys in ways that challenge long-standing arguments as to why some groups vote more than others. African-American women in their 30s are not the demographic group thought to be at the apex of politically-relevant resources, as defined in seminal works of participation. But, they are registered at higher rates than any other demographic, vote at rates nearly 25 percentage points higher than black men, and vote more than white men of similar age. The patterns here are just an initial step in a broader research agenda to rethink the nature of political participation based not on what people say they do but on what they actually do.

References and Notes

1. Rosenstone SJ, Hansen JM. Mobilization, Participation, and Democracy in America. New York, NY, USA: Macmillan; 1993.

2. Sidney V, Schlozman KL, Brady HE. Voice and Equality: Civic Voluntarism in American Politics. Cambridge, MA, USA: Harvard University Press; 1995.

3. Sidney V, Burns N, Schlozman KL. Knowing and Caring About Politics: Gender and Political Engagement. Journal of Politics. 1997;59(4):1051–1072.

4. Ansolabehere S, Hersh E, Shepsle K. Movers, Stayers, and Registration: Why Age is correlated with Registration in the U.S. Quarterly Journal of Political Science. 2012;7(4):333–363.

5. Ansolabehere S, Hersh E. The Quality of Voter Registration Records: A State-by-State Analysis. Caltech/MIT Voting Technology Project Report. Cambridge, MA, USA: Caltech/MIT; 14 July 2010.

6. In a study of election administration quality, Ansolabehere and Hersh [10] find turnout statistics on voter files to be highly accurate.

7. United States Census Bureau. NC-EST2008-alldata: Monthly Population Estimates by Age, Sex, Race

and Hispanic Origin for the United States. Suitland, MD, USA: United States Census Bureau; November 2009.

8. American Community Survey. 2008 1-year PUMS File. Suitland, MD, USA: United States Census Bureau; 2008.

9. Silver BD, Anderson BA, Abramson PR. Who overreports Voting? American Political Science Review. 1986;80(2):613–624.

10. Ansolabehere S, Hersh E. Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate. Political Analysis. 2012;20(4):437–459.

11. Comparing actual vote data with CPS data in states in which race is recorded on the registration rolls reveals that the CPS exaggerates the voting rates of Blacks slightly and of Hispanics substantially in these states. This analysis was performed by the authors for 2012.

12. Ansolabehere S, Persily N, Stewart III CH. Race, Religion, and Vote Choice in the 2008 Election: Implications for the Future of the Voting Rights Act. Harvard Law Review. 2009;123(6):1385–1436.

13. See, for example, the analysis offered in [12].

14. Burns N, Schlozman KH, Verba S. The Private Roots of Public Action. Cambridge, MA, USA: Harvard University Press; 2001.

15. Lien P-T. Does the Gender Gap in Political Attitudes and Behaviors Vary Across Racial Groups? Political Research Quarterly. 1998;51(4):869–894.

16. Seltzer RA, Newman J, Leighton M. Sex as a Political Variable: Women as Candidates and Voters in U.S. Elections. Boulder, CO, USA: Lynne Rienner Publishers; 1997.

17. Hornor LL. Black Americans: A Statistical Sourcebook. Palo Alto, CA, USA: Information Publications; 1999.

18. Montoya LJ, Hardy-Fanta C, Garcia S. Latina Politics: Gender, Participation, and Leadership. PS: Political Science and Politics. 2000;33(3):555–561.

19. Jones-Correa M. Different Paths: Gender, Immigration, and Political Participation. International Migration Review. 1998;32(2):326–349.

20. Miles, TJ. Felon Disenfranchisement and Voter Turnout. Journal of Legal Studies. 2004;33(1):85–129.

21. Miles [20] has currently the most sophisticated analysis of the effect of these laws on turnout, and he finds no effect.

22. Because the data come from public records, we measure socio-economic status with block-group level Census statistics and marriage based on household information in the voter file (e.g. many married voters can be identified by age, gender, and last name of individuals in the household). We generated a block group SES scale using factor analysis that incorporates measures of block group income, college education, single parent rates, public assistance rates, and percent living in poverty rates. These variables all scale onto a single factor. As block group SES increases, the gender gap for Blacks and Hispanics decreases, but it does not disappear or even approach levels as for Whites. For example, among Blacks and Hispanics in the highest SES block groups, the gender participation gap is four percentage points. At the lowest SES level, the gap was seven points for Hispanics and ten points for Blacks. For Whites at all levels of SES, the gender gap is approximately two points. A similar pattern exists in the difference between married and single people across races. Married Blacks have a smaller gender gap than unmarried Blacks, but there is still a substantially larger gender gap for married Blacks (and even high-SES married Blacks) than for comparable Whites.

23. Wiggens DC. Righteous Content. New York, NY, USA: New York University Press; 2005.

24. Lincoln CE, Mamiya L. The black Church in the African American Experience. Durham, NC, USA: Duke University Press; 1990.

25. Greenberg A. Race, Religiosity, and the Women's Vote. Feminist Studies. 2001;22(3):59–82.

26. Atkinson MD, Fowler A. Social Capital and Voter Turnout: Evidence from Saint's Day Fiestas in Mexico. British Journal of Political Science. Forthcoming.