

Tracking Trends in Minority Representation: Early Findings from the Congressional Candidate Dataset Project

Stephen Maynard Caliendo, North Central College
smcaliendo@noctrl.edu

Sidra Hamidi, North Central College

Charlton D. McIlwain, New York University

Abstract

Some of the most fundamental questions about the nature of representation in a democracy – including representation of women and racial minorities – can be answered by understanding the relationship amongst candidates, potential voters, and electoral outcomes. However, the lack of a comprehensive and accessible collection of relevant data has presented a challenge for political scientists who wish to explore these questions. We have been in the process of compiling a dataset that will include demographic information about each congressional candidate (date of birth, gender, race), electoral district/state (percentage of each racial designation), campaign finance and electoral outcome information for all U.S. House and Senate primary and general election contests since 1968. This paper features a preliminary descriptive account of the general election races for the U.S. House between 1968 and 2006, including correlations among some of the key variables that heretofore have not been available in one place. Though formal findings are not appropriate because all of the data has yet to be collected, we discuss trends over the time period and offer suggestions for further inquiry.



The authors would like to thank Dr. Caroline St. Clair and Alex Bestul for their ongoing help with the technical aspects of this project. They have been assisted by Matt Adorjan, Sean Gallagher, Tucker Graczkowski, Jeremy Lee, Tim Romel, and Victoria Wang. We would also like to thank Priscilla Martinez and Alexis Smith for their help with data collection and entry. They have been helped over the past year by a team of research assistants that include: Lauren Lolls Batten, Gwendolyn Bowman, Casey Brienza, Asha Carter, Jennifer Diaz, Marissa Evans, Michelle Fernandez, Alyssa Figueroa, Ann Fisher, Maria Hamidi, Jillian Hentschel, James Jack, Zakia Johnson, Zhang Lin, Ryan Kauffman, Anne Markel, Christine Meinders, Lisa Mueller, Mary Naset, Dori Paulino, Ryan Piers, Maria Qazi, Samantha Sullivan, Cindy Vincent, Yan Wang, Orlando Ward, and Jeremiah Wilson.

Tracking Trends in Minority Representation: Early Findings from the Congressional Candidate Dataset Project

The question about what determines minority candidates' electoral success has been generally seen as moot in the absence of substantive competition due, in part, to racial gerrymandering and the creation of majority-minority electoral districts. However, as minority candidates increasingly vie competitively for seats against other minority candidates, as well as against Whites, this is an outdated assumption. Additionally, though "competition" has, until recently, been low in majority-minority electoral districts, both the challengers and victors in these contests have contributed to the electoral dynamics of minority elections and the overall picture of minority elected officials. Similarly, questions about symbolic and substantive representation with respect to gender have interested political scientists for decades, but most scientific analyses are necessarily limited to individual years or a small collection of years.

Surprisingly, no list of candidates for federal office exists that indicates race or gender. We can easily find lists of women or African Americans who have won seats in Congress and complete lists of candidates who have lost (without race, age or gender identification), but women and racial minorities who have lost have not been identified in any systematic way – certainly not with any helpful accompanying data (such as campaign finance figures). This project involves the construction of a dataset that will offer scholars the ability to examine the role of minority candidates for federal office and contribute to the state of knowledge about race and gender politics in America.

Purpose and Significance of the Project

The primary purpose of this project is to compile a demographic profile of every candidate who has run for federal office in the United States since 1968. This database – which is currently about 20% complete – includes the following information for each candidate: race, gender, birth date/age, state/electoral district; office sought; electoral opponent; racial composition of the district or state at the time of election; total vote total and percentage of the vote received for each candidate; and amount of campaign funds raised and spent. In the future, we also expect to add the following: number of televised political ads produced and/or run in the campaign; number of newspaper stories written about the campaign; and photograph(s) of each candidate. Once we have completed an initial set of profiles, the database will be made public so that other researchers and members of the public will be able to access it and, potentially, add information or even categories of information to the database.

For scholars, journalists, and members of the public interested in social and political history, civil rights history, political science, gender studies, African American/Latino American history, etc. this project is highly significant. It fills a glaring void by providing demographic information about the men and women who have run for federal office.

Potential Usefulness

Currently, researchers lack a comprehensive tool that tracks important information about congressional candidates. Information regarding congressional candidates is available in a number of different sources and often forces researchers to invest much time and effort gathering data. Many studies that examine minority representation would have benefitted from the information contained in the database. For example, Caliendo and McIlwain (2006) examine racial minority candidates in the 2004 election cycle. Even if there were sufficient reason to center such a study on one election year, their analysis would have been enhanced by the ability to place those findings in a more systematic and broad historical context.

Several studies concerning minority representation have centered on the effectiveness of majority-minority districts and racial redistricting. For instance, Cameron, Epstein and O'Halloran (1996) critically examined the idea that majority-minority districts are the most effective means of increasing Black representation in Congress. The authors held that instead of increasing minority-sponsored legislation, the creation of majority-minority districts only serves to dilute minority populations in surrounding districts which results in less interest on the part of legislators to address the needs of the small number of minorities within their districts. They conclude that there is a trade-off between increasing the number of minority representatives through redistricting and also enhancing minority-oriented legislation. Further, they contend that minority candidates do have a "substantial chance of being elected from districts with less than 50% minority voters" (p. 794). A study such as this one will be enhanced by a database that

contains information about the racial composition of districts from which minorities have been elected over a number of decades. Through this information, the researchers will be able to determine, for instance, the degree to which minority candidates have been elected from majority-White districts.

Cameron, Epstein, and O'Halloran, like many other researchers, compiled data from Congressional Quarterly databases and the voting record provided by the Leadership Conference on Civil Rights. The authors then develop a number of equations to demonstrate the kind of redistricting technique that would best suit minority interests. The Race Project database will simplify the data gathering process so that researchers who examine these types of questions can download data into a spreadsheet and add such information such as voting record data in order to assess the efficacy of majority-minority districts. This might prove to be increasingly of interest as the debate surrounding redistricting and reauthorization of the Voting Rights Act of 1965 continues to be an area of study in race/ethnicity politics.

Another burgeoning area of study involves the connection between race and campaign finance. For example, Smith (2001) examines the racial implications of campaign finance reform and contends that efforts to reform campaign financing have ignored some of the difficulties that minority candidates may have in raising the funds necessary to run a successful campaign. Smith surmised that because there is substantial racial inequality in American society, minorities are already hindered from campaign success in comparison to their non-minority counterparts. Smith states, "greater equality through reform is only possible if citizens of color are

allowed to employ their full expressive resources in the political process” (p. 1461). While Smith’s evaluations are based on the general inequality faced by minorities, his hypothesis may be empirically tested with the data provided in the database. With this dataset, campaign finance data of minority candidates will be able to be analyzed in comparison to that of White candidates in order to understand the relation between campaign finance and race. Such a study could have policy implications in the form of differential treatment of White and Black voters and candidates.

Overton (2002) also examined the link between campaign financing and race. He notes that “existing frameworks fail to acknowledge that past state-mandated discrimination against racial minorities has shaped the current distribution of property, which in turn hinders the ability of many people of color to participate fully in a privately financed political system” (p. 989). Because of past inequalities, minority candidates are severely limited in their ability to run a successful campaign. The Race Project database will assist in composing an empirical study of the contended disparities in campaign financing in relation to the racial background of the candidates. If a link between race and campaign finance is found, further inquiry and speculation would be required in order to explain the gap. Such a study could have far-reaching implications for advocates of campaign finance reform. Additionally, other factors – such as party affiliation – might be helpful in explaining the gap in funding. Additionally, the data in the Race Project database can be used in conjunction with other sources of research, such as the data that track individual and group donors to congressional campaigns (such as that collected by The Center

for Responsive Politics and made available at www.opensecrets.org). The use of these data sources would allow for researchers to test these and other potential factors that relate to minority representation.

Another area of congressional research centers on the role of gender in electoral campaigns. By simple numerical estimations, and according to the notion of descriptive representation, Congress should consist of at least 50% women candidates. However, the lack of descriptive representation in Congress is a phenomenon that requires deeper analysis. Uhlaner and Schlozman (1986) conducted one of the first studies that centered on the role of gender in congressional campaigns. They examined some of the societal factors that prevent women from being successful in congressional campaigns, noting voter bias, prejudice of political influentials, and campaign financing gaps as reasons for the lack of descriptive representation in Congress. Uhlaner and Schlozman state, "On the one hand, women candidates might be victims of direct gender-based discrimination by contributors. On the other hand, women might simply command fewer of the resources, such as incumbency or committee chairmanships, that donors reward" (p. 31). This observation is an important one because it suggests the definite link between gender and campaign funds accumulation. The Race Project database will allow for the enhancement of such studies because the correlation between gender and campaign finance is relatively easy to test. Researchers could also compare trends in campaign finance over time. Additionally, it will be possible to test a number of independent variables, such as racial composition of the district and party affiliation, on the success of women candidates running for Congress.

Researchers can effectively measure if shifting gender roles are affecting congressional representation.

As with studies of the effects of race on campaign finance, the Race Project database will allow researchers to gather data relating the effects of gender on campaign finance. Additionally, it will also be revealing to determine trends in representation for women who are also members of racial minority groups. Comparing women of color to White women and examining factors such as party affiliation, campaign funds raised and spent, and racial composition of the districts in which women candidates run will provide some insight variables that are related to candidates' success.

Researchers will also be able to isolate certain candidates and conduct further research on the central campaign message and platform on which the candidate ran in order to understand which issues are employed by campaigns of racial minorities or women candidates to be successful. The many research possibilities that are provided by the accessibility and ease of use of the Race Project database will serve to enhance the study of representation in United States. It will provide insight into the extent to which congressional representation is reflective of the changing composition of the United States. Tracking the different trends in representation will continually enhance political science research.

Additionally, changes in trends of representation may be indicative of policy changes. For example, if there are any indicated changes in campaign funds since 2003, it may be attributable to the passing of the Bipartisan Campaign Reform Act (BCRA). Such trends can be used to gauge the effectiveness of policy reforms and

other political trends in relation to congressional campaigns. The database will allow for researchers to diversify the attribution variables that comprise candidate success.

Methods and Accessibility

The method by which this information is being retrieved varies depending on availability. The Congressional Quarterly Statistical Universe database contains the names, party, actual votes and vote percentage earned for each candidate who has run for Congress. Those data were uploaded via scripts that were constructed for that purpose; the remainder of the information must be retrieved from disparate sources. For some candidates – particularly those who won or who have run recently – is easily found. Search engines (e.g., Google) or specialized free databases (e.g., Congresspedia, The Political Graveyard, Politics1.com) are often excellent sources, but biographical information for many of the candidates who lost is often difficult to find, particularly for earlier years. When information cannot be found in these places, research assistants search proprietary databases such as CQ Almanac, Latino American Experience, ProQuest Historical Newspapers and Lexis-Nexis Academic Universe, with the hopes that a reference to a candidate's age, race or gender is made. Racial composition of each electoral district is readily available from the U.S. Census Bureau, and campaign finance data is available from the Federal Election Commission.

Far from compiling a simple list or even a spreadsheet with all of this information, our task is to make the data widely available and accessible. First, we

will make the data freely available via the RaceProject.org website. Scholars will be able to export the data in a variety of formats (e.g., tab- and/or comma-delineated, Excel, SPSS) for their own use and manipulation. Rather than simply exporting the entire dataset, however, researchers will have the option to conduct detailed searches to generate only the data in which they are interested, keep only the variables in which they are interested within those search criteria, and download a customized dataset for manipulation and analysis.

Second, the data will be available for searching via a user-friendly Web interface that permits detailed and specified searches with dropdown menus. Further, similar to the options to be made available to scholars, users will be able to choose what fields they would like to have displayed in the output of their search. For instance, perhaps a student working on a paper would want to know how many Latino candidates have run for office in the Deep South between 1980 and 1990. He or she would be able to search using those criteria and select an output that contains, for instance, candidates' names, opponents' names, race and gender, total votes received, and the racial composition of the electoral district. If the student is not interested in campaign finance data or knowing exactly in what states or districts the elections took place, that information need not be displayed. In short, the database will be free, accessible, and full customizable.

Early Results

There are 16,492 candidates who ran in general elections for the U.S. House of Representatives between 1968 and 2006. To date, we have collected and entered

roughly 40% of the demographic data for each of those candidates. Campaign finance data and information pertaining to the racial composition of the electoral district has been collected but not yet entered. Accordingly, this section provides some preliminary results to suggest the types of analyses that will be possible once the data is more complete. At this point, however, we test no hypotheses and make no claims as to the generalizability of these findings, as we cannot be certain that the missing data is random. In fact, with respect to examination of trends over time, we know that our data is much more complete for recent years than it is for earlier years.

The total number of general elections to the U.S. House from 1968 to 2006 (20 cycles) was 8,720.¹ The number of candidates running each electoral year ranges from 780 (in 1998) to 1,058 (in 2004), with a mean of 824.6 (N=16,492). With respect to party identification 47.8% of the ballot positions were held by Republican candidates; 50.5% have been Democratic; 1.7% by candidates running with another party label. With respect to incumbency status, 45.7% of ballot positions were held by incumbents seeking re-election, while 54.3% were challengers or candidates running in open seats.

¹ This is but one example of the data curiosities that must be examined and treated on a case-by-case basis. A general election contest for each of the 435 districts each year would yield 8,700 elections. Since special elections and runoffs are treated separately, and delegates from territories and the District of Columbia are not included, it is unclear at this time where the additional 20 elections in this dataset originate. The Congressional Quarterly dataset, from which the core information is drawn, contains numerous errors, such as omitting a middle initial from a candidate in some years but including it in others, failing to include “Jr.” and, thus, confusing the dataset as to the identity of a father and son who both served, etc.

Democrats won more frequently (56.2%) in general elections for the House during this period of time than Republicans (43.6%). Only one-tenth of one percent (13 contests) of independent or third party candidates who ran during this period were successful. Of the Republican candidates who ran, 48.3% (3,804) won; of the Democrats who ran, 58.8% (4,903) won; 0.04% (13) of independent and third party candidates were successful in their bids.²

To date, we have 44.9% of candidate gender data collected, of which men account for 89% of ballot positions. We have 38.1% of the racial data entered: 87.1% candidates on ballots were White, 6.8% were African American, 3.7% were Latino, 1.2% were Asian, and 1.2% identified as Native American or another race.

We wish to tread very carefully in providing a glimpse of the types of questions that will be able to be addressed with these data, but we offer Figures 1 through 7 to that end. Because of the incomplete nature of the dataset, we will not drill deeper into these numbers, but it should be apparent that there are a number of possibilities as to the types of research questions that these data will be able to help to address.

[Figures 1 through 7 here]

Challenges

We have encountered a number of different challenges throughout the process of data entry and design. As noted above, the Congressional Quarterly dataset, while tremendously helpful on the whole, contains a number of errors that

² There is some overlap on party wins and losses because some states (e.g., Louisiana) that permit more than one candidate from a political party to run in the general election and because all general election contests are not contested.

will need to be addressed. Other errors are those of our own research assistants. Some are a combination. For example, when we ran the initial age frequencies, we noticed that we had a number of teenagers in the dataset. We discovered that some of those errors were data entry mistakes, while at least one other was the result of our initial script not being able to identify the difference between two candidates with the same name. While a research assistant should have caught the error when he or she entered the birth date, he or she did not, and, thus, the same birth date was entered for two different candidates (who, in reality, were not at all close in age). When we subtracted the election year from the birth year to generate the approximate age, the older candidate, whose birth date was confused with the younger candidate, appeared to be 14 years old at the time of his candidacy.

One of the more substantive issues that surfaced towards the beginning of the data entry was the question of the different race delineations we chose to include. In the current database, the choices for race include African American, Asian, Caucasian, Latino, Native American, and Other. While this includes the major race categories (though the U.S. Census considers Latino to be an ethnicity rather than a race), we considered some serious questions about the inclusion of more specific ethnicities. For instance, the category "Asian" is somewhat deceptive because it is usually associated with individuals of East Asian descent (Chinese, Japanese, Korean, etc.). However, this is very different from a South Asian or Middle Eastern candidate running for congressional office. Additionally, a couple of research assistants expressed confusion as to the expressing the race of Hawaiian candidates. These individuals were listed as Asian but the question how accurate

that is still remains. We also did not include a “Mixed Race” category, which was problematic for some candidates, forcing us to look for cues as to the ways candidates identified themselves (such as participation in racial minority caucuses). As an obvious example, Barack Obama is listed as African American despite the fact that he is biologically both Black and White. Because race is primarily a social construct, we categorized him based on the race to which he is socially and culturally associated and to which he self-identifies. A similar process is used with other candidates of mixed-race heritage. Still, if we expect this dataset to be of use for years to come, we must be thoughtful about the myriad questions with which race scholars continue to struggle and provide the most forward-thinking, useful and accurate information possible.

Discussion

As is so often the case with conference papers that involve significant data collection and/or analysis, we had hoped to be further along in the process by the time this meeting came around, but we are not particularly surprised that we are not. After all, not having all of this information in one place is something that researchers have lamented for some time. Getting it all together and putting it into a format that will be useful for scholars and members of the public alike has been time consuming and, at times, frustrating. We are excited about the possibilities, but also realistic that completion will likely come in waves. We anticipate having a formal “launch” of the website by the American Political Science Association meetings in the fall of 2011 (the year marking the 10th anniversary of The Project on Race in

Political Communication), but only general election data is expected to be included in that initial release. We will add data from primary elections, special elections and runoff elections as the years go on, and we will continue to add new data biennially with each subsequent wave of federal elections.

References

- Caliendo, S. M., & McIlwain, C. D. (2006). Minority candidates, media framing, and racial cues in the 2004 election. *The Harvard International Journal of Press/Politics*, 11(4), 1-25.
- Cameron, C., Epstein, D., & O'Halloran, S. (1996). Do majority-minority districts maximize substantive Black representation in Congress? *The American Political Science Review*, 90, 794-812.
- Overton, S. (2002). But some are more equal: Race, exclusion, and campaign finance. *Texas Law Review*, 80, 987-1056.
- Smith, T. (2001). Race and money in politics. *North Carolina Law Review*, 79, 1469-1522.
- Uhlaner, C.J., & Schlozman, K.L. (1986). Candidate gender and congressional campaign receipts. *The Journal of Politics*, 48, 30-50.

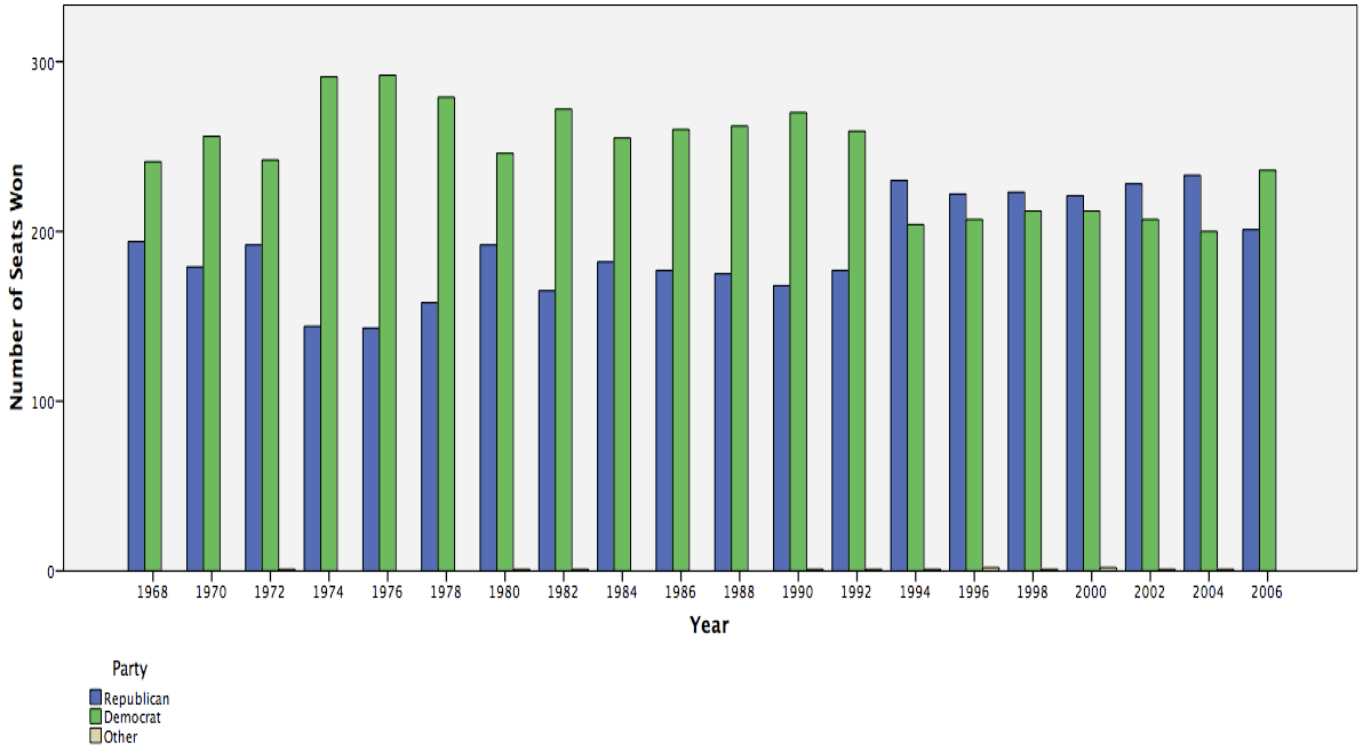


Figure 1a. Successful Candidates by Party

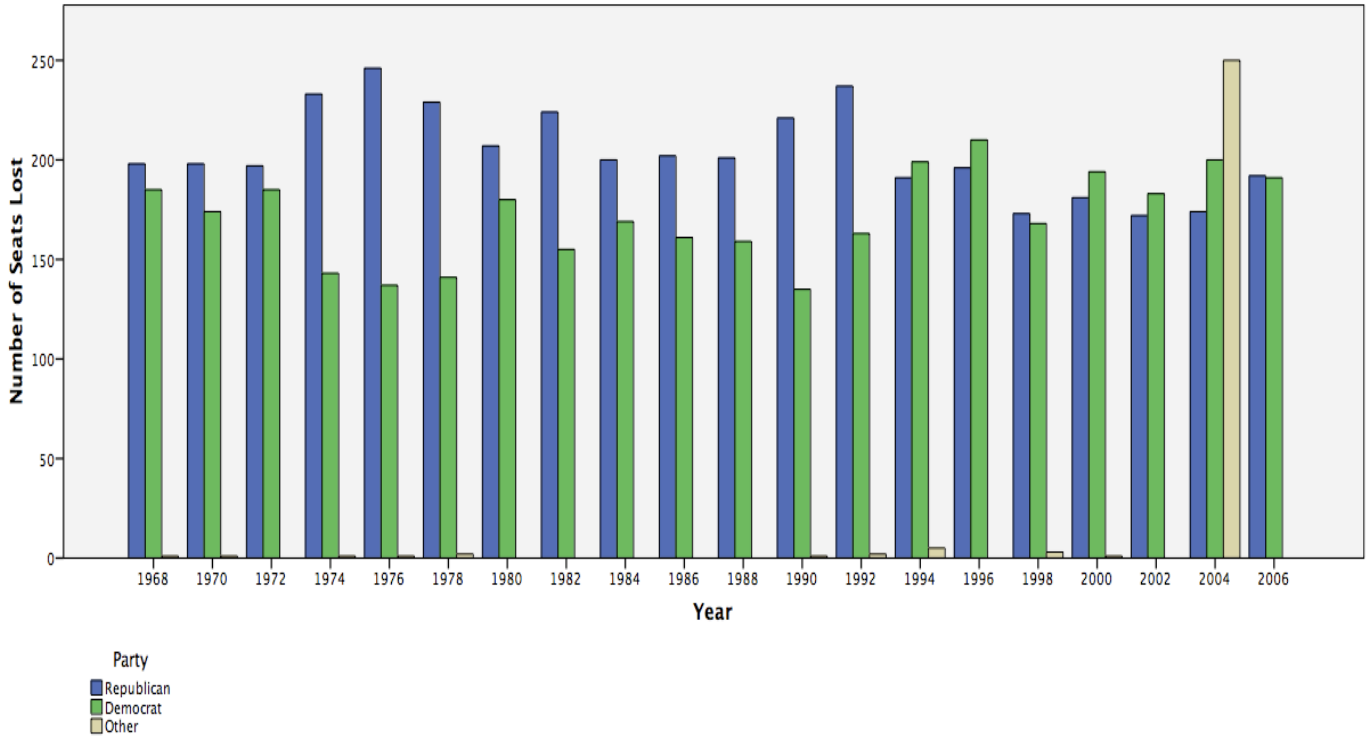


Figure 1b. Unsuccessful Candidates by Party

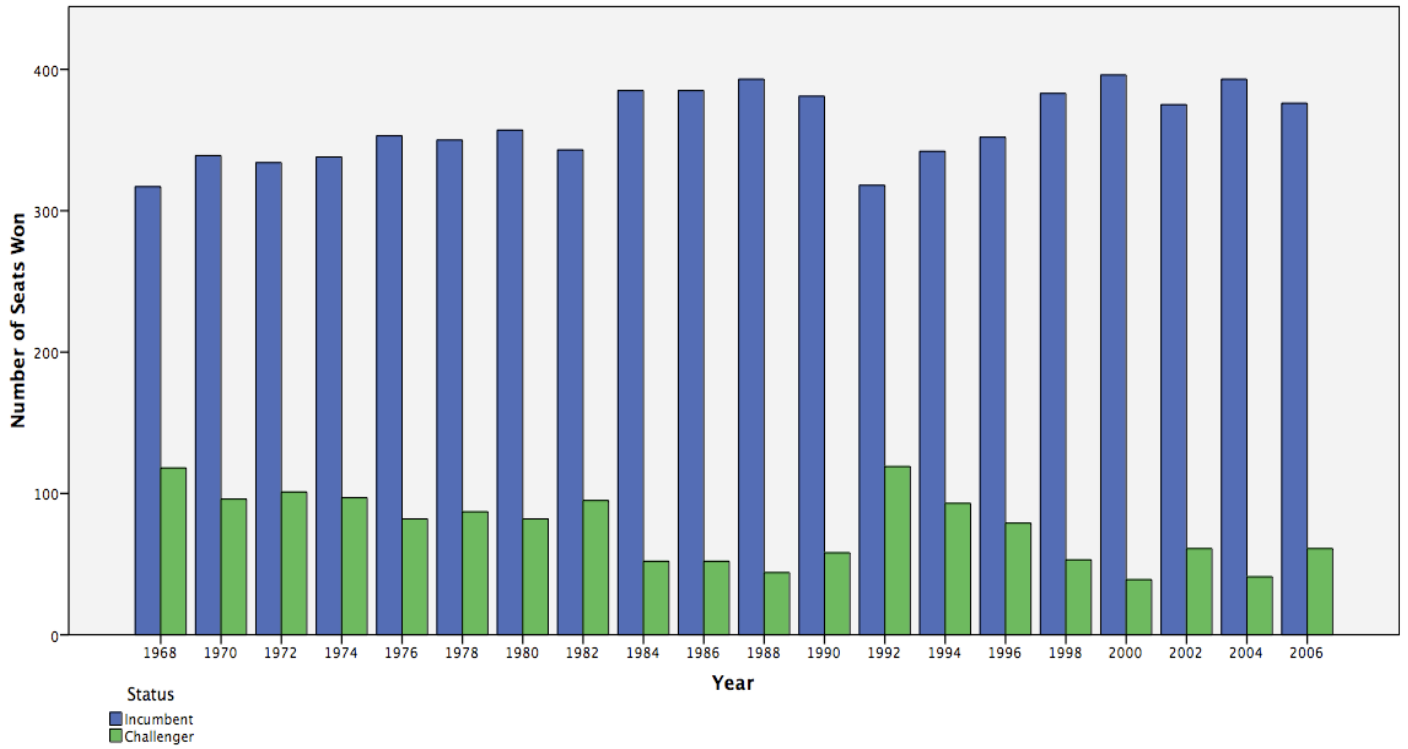


Figure 2a. Successful Candidates by Incumbency Status

Note: The “challenger” category includes candidates running for open seats.

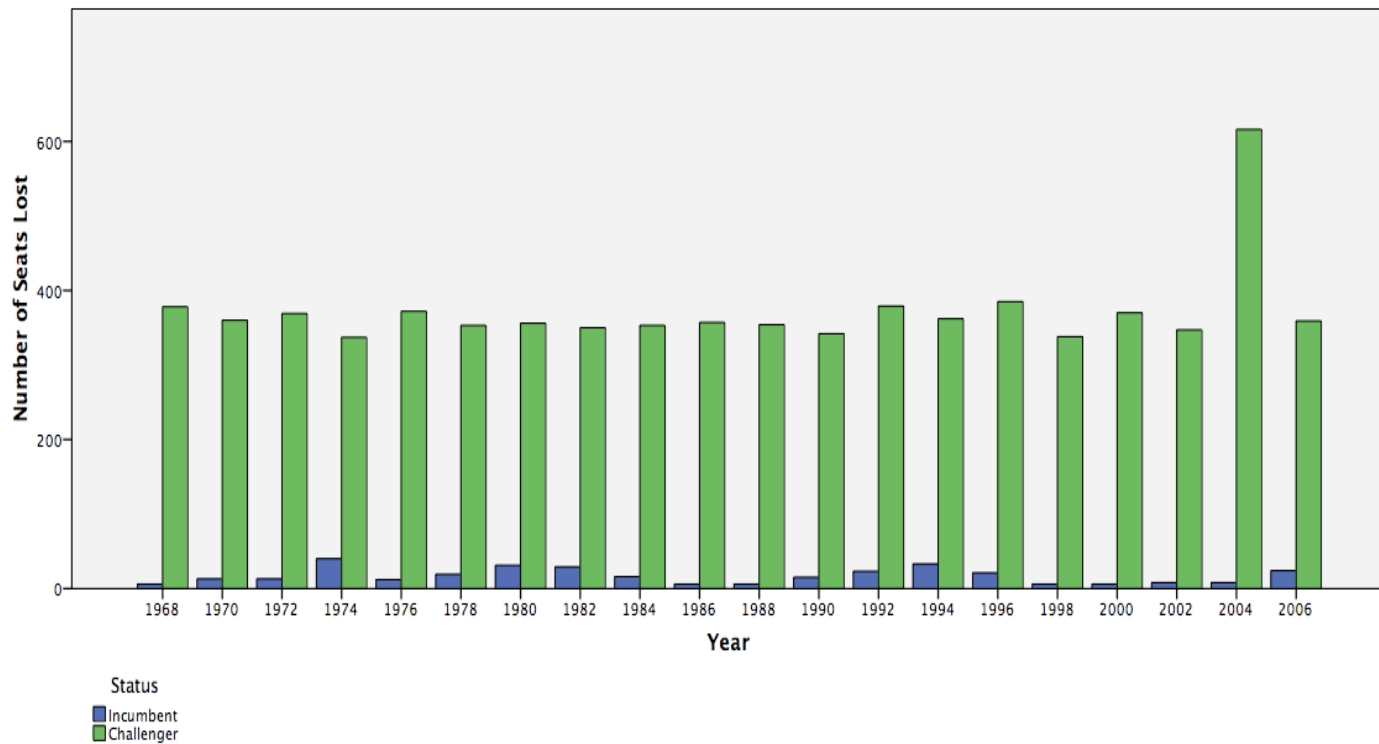


Figure 2b. Unsuccessful Candidates by Incumbency Status

Note: The “challenger” category includes candidates running for open seats.

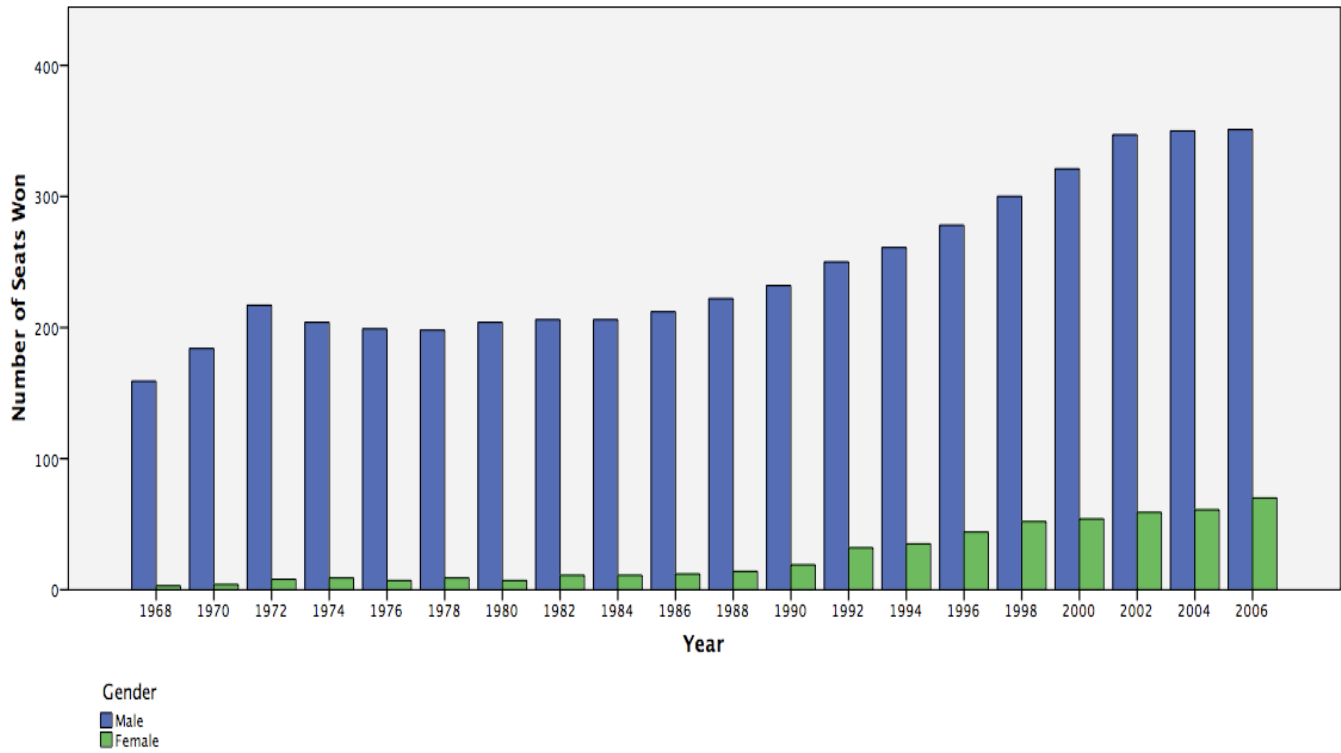


Figure 3a. Successful Candidates by Gender

Note: Data are incomplete. The Y-axis represents the number of all successful candidates from each gender. Only candidates whose gender is known are included.

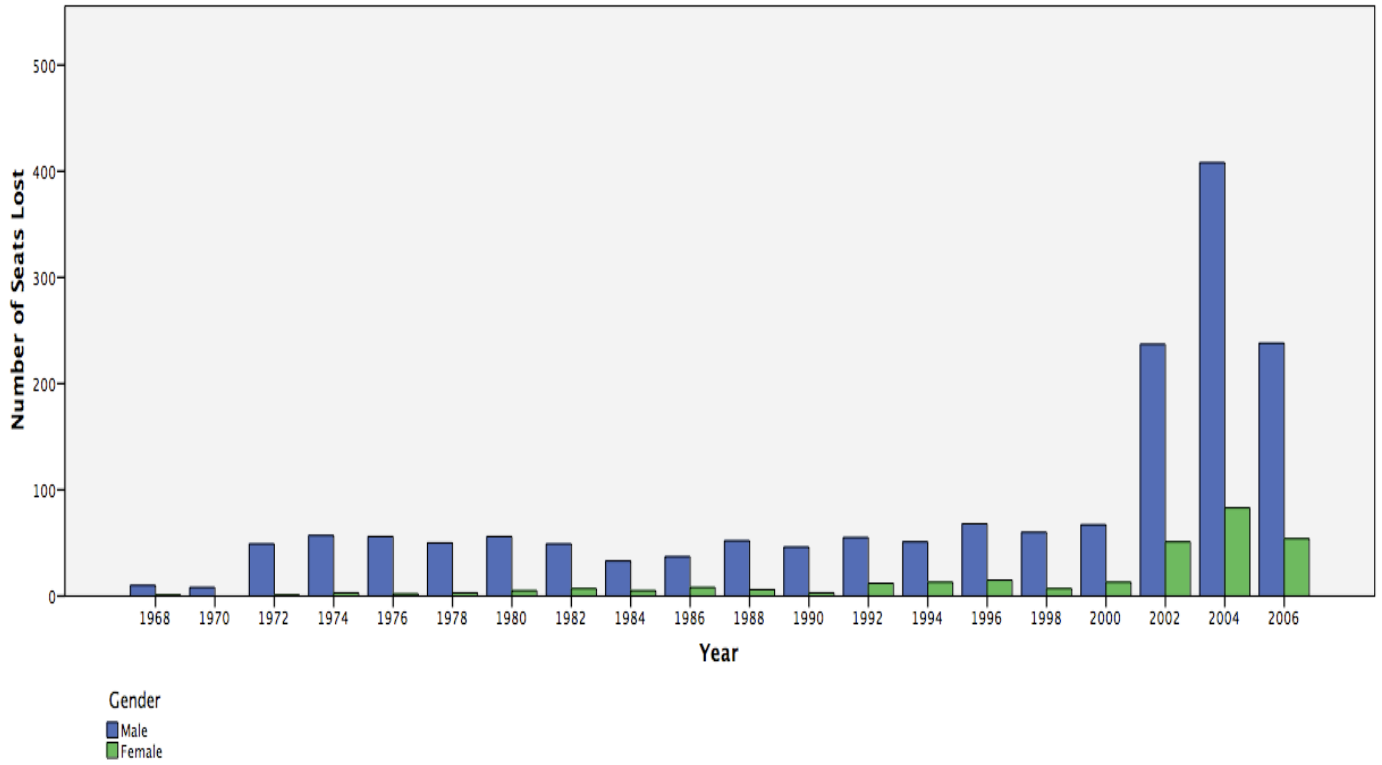


Figure 3b. Unsuccessful Candidates by Gender

Note: Data are incomplete. The Y-axis represents the number of all unsuccessful candidates from each gender. Only candidates whose gender is known are included.

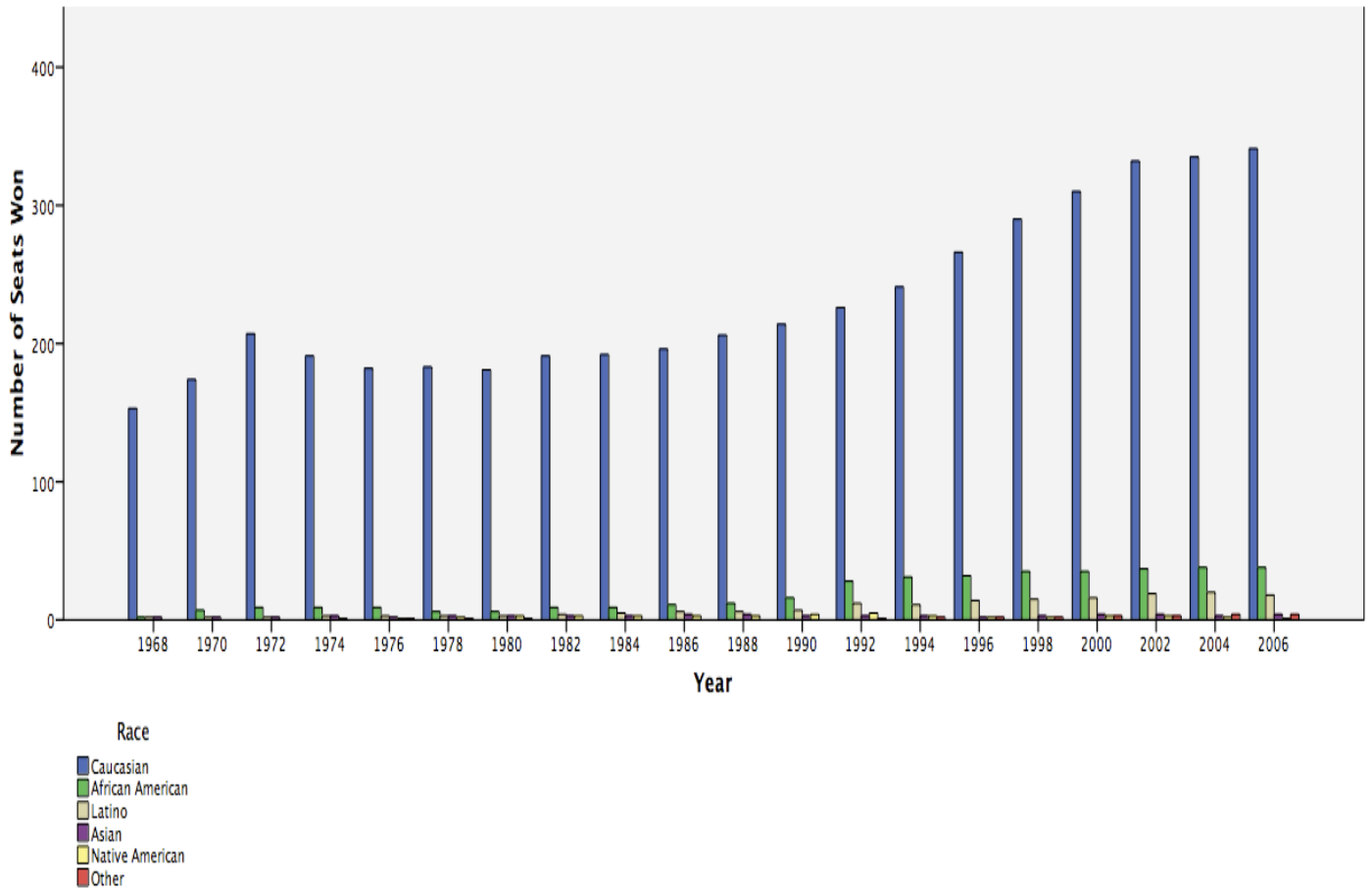


Figure 4a. Successful Candidates by Race

Note: Data are incomplete. The Y-axis represents the number of all successful candidates from each race. Only candidates whose race is known are included.

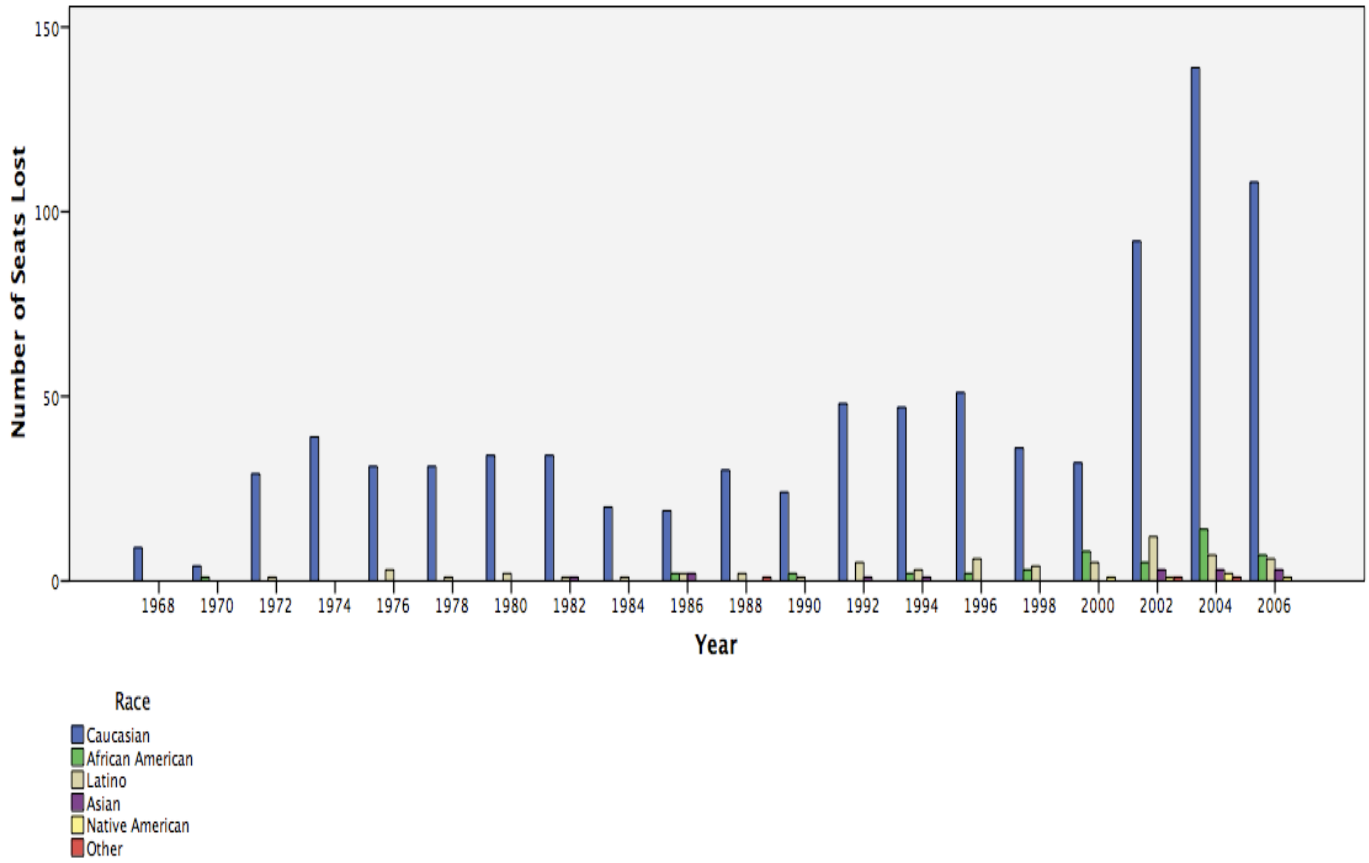


Figure 4b. Unsuccessful Candidates by Race

Note: Data is incomplete. The Y-axis represents the number of all unsuccessful candidates from each racial group. Only candidates whose race is known are included.

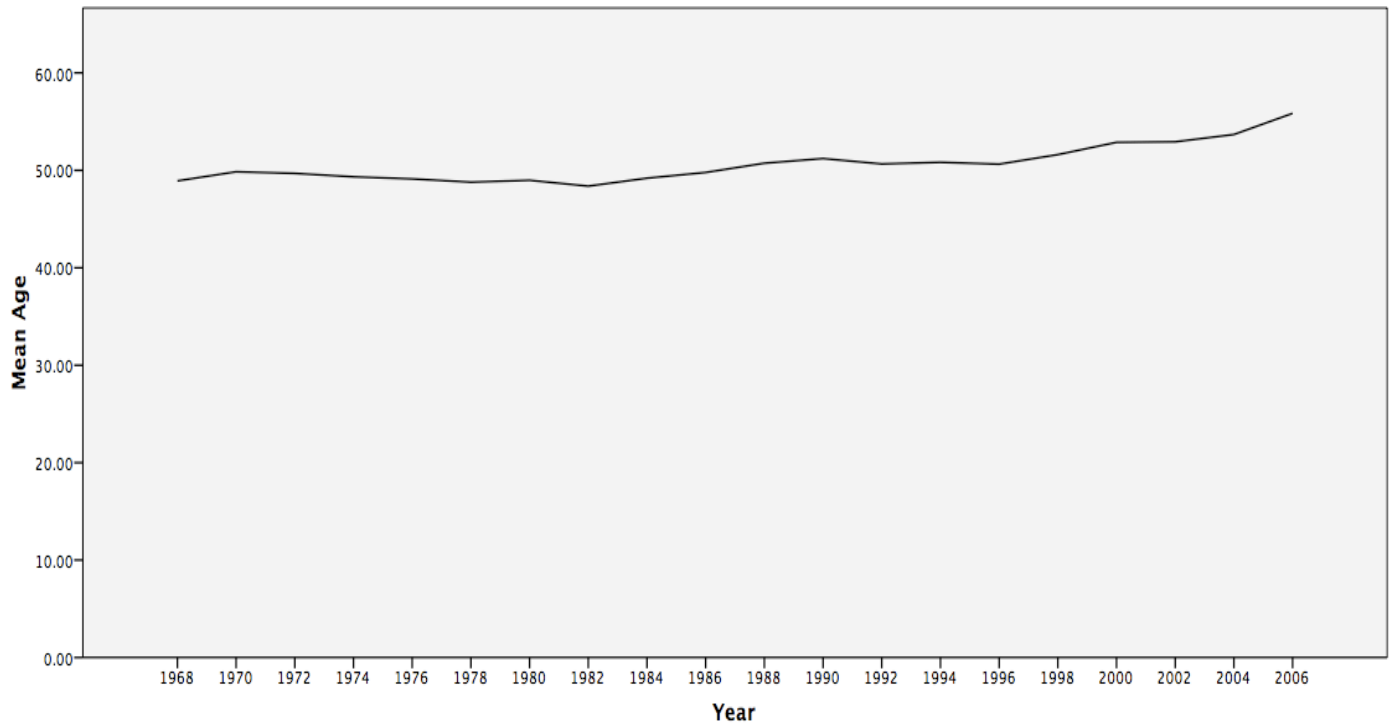


Figure 5a. Successful Candidates by Age

Note: Data are incomplete. Age is approximated (year of election minus birth year). Mean is calculated only for candidates whose birth date is known.

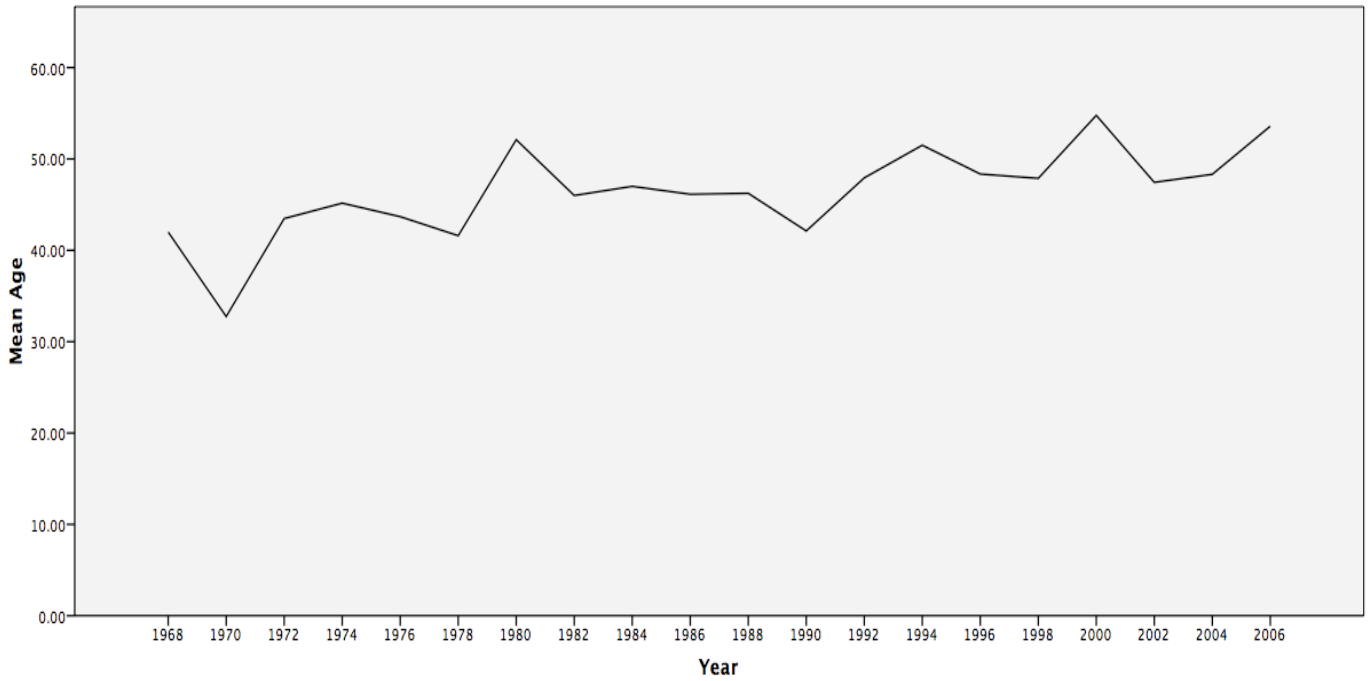


Figure 5b. Unsuccessful Candidates by Age

Note: Data are incomplete. Age is approximated (year of election minus birth year). Mean is calculated only for candidates whose birth date is known.

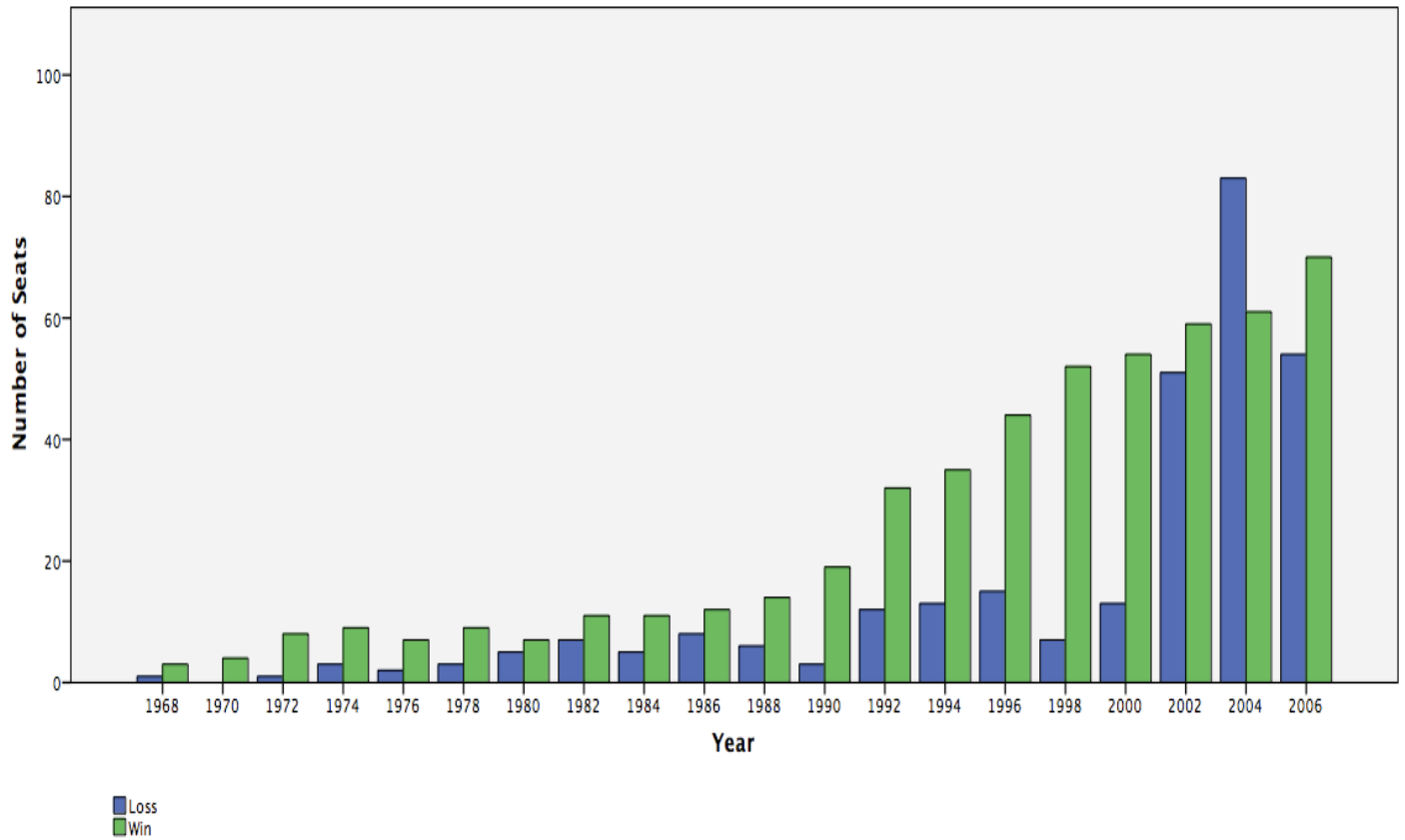


Figure 6. Success Rate for Women Candidates

Note: Data are incomplete. Only female candidates (for whom gender is known) are included.

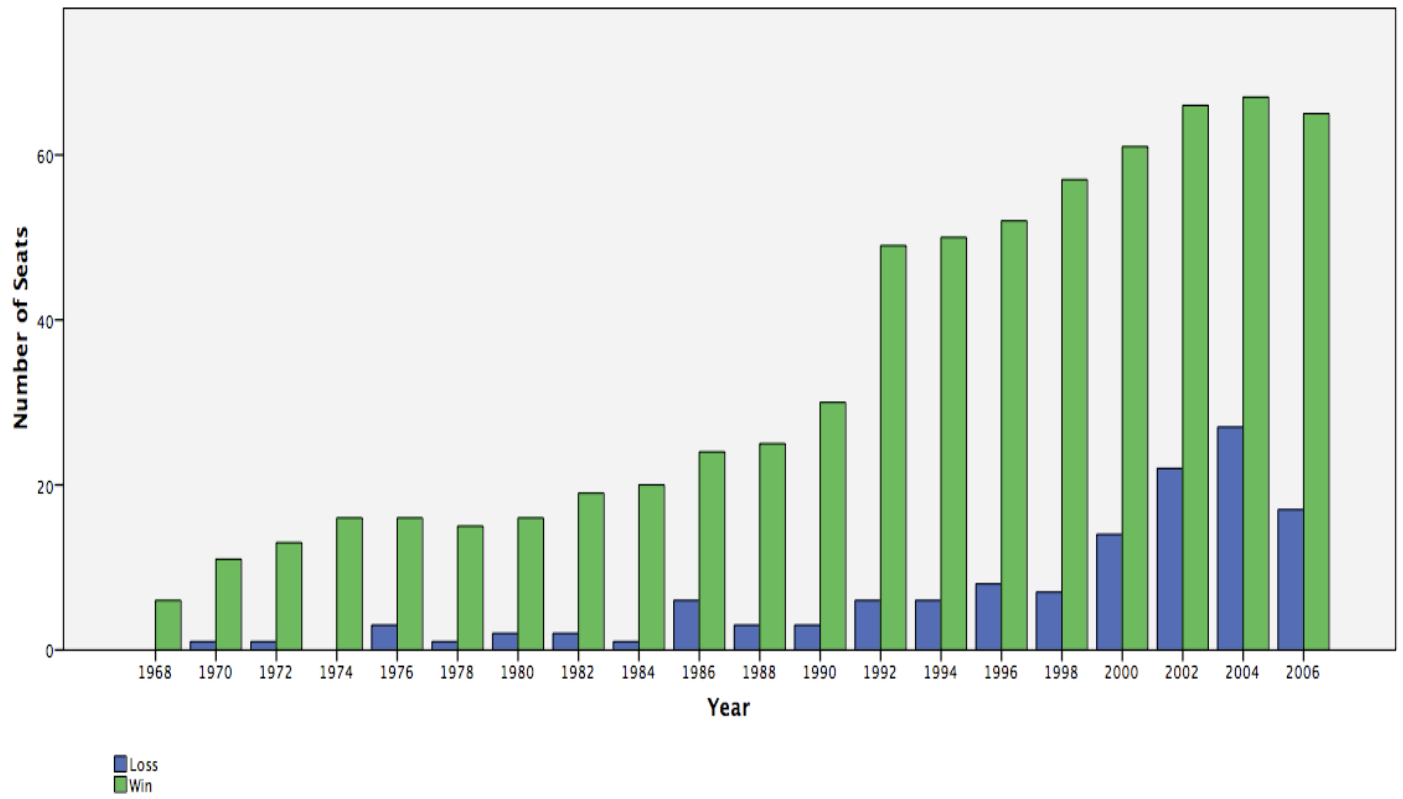


Figure 7. Success Rate for Non-White Candidates

Note: Data are incomplete. Only non-White candidates (for whom race is known) are included.