# LADIES IN THE HOUSE: HOW STATE PARTY STRUCTURE AFFECTS WOMEN'S ELECTION TO CONGRESS

SI 544 Introduction to Statistics and Data Analysis
April 27, 2016

Daina Andries
Kate Barron
Vijeta Belandor
Micaela Hunter
Fiona Potter

# Table of Contents

# EXECUTIVE SUMMARY

It has been exactly one hundred years since the first woman was elected to the U.S. House of Representatives. In the last half century, women's increased access to traditional 'pipeline' jobs leading to federal government positions (such as business, law, and state politics) has aided a gradual increase in women's presence on Capitol Hill. Nevertheless, women still make up barely 20% of Congress.

In this study, we conducted a series of linear regressions to see (1) whether bias prevents the nomination of women (in state elected positions) for Congressional seats, and (2) how state party structure influences women's transition from state government to federal government. We hypothesized that there is some kind of bias against women transitioning from state legislatures to federal government: we checked for this bias by looking at the differences between the percentages of women in state legislatures and the mean probability of a woman being nominated to run for Congress. We also hypothesized that there would be more 'affirmative action' in support of women from the Democratic party, which we measured by introducing a dummy variable for party. Finally, we hypothesized that geography plays a significant role in determining the opportunity structure for women.

Our dataset consisted of 1849 observations. Each row contained the candidate name, state, congressional district, party, election cycle (2008-2016), gender, geographic region, and urban percentage of the district's population. We focused only on candidates running for 'open' seats in the House of Representatives (since there is more turnover in the House than in the Senate, and because incumbents are more likely to stay in office).

First, we used a simple regression model to examine the effects of party and representation of female state legislators on a woman's likelihood of running for office. Next, we incorporated dummy variables to investigate how women's political opportunities might differ by region. We interacted these variables with the party factor and the percentage of women in state legislatures for a more nuanced investigation of the effects of party structure are the state level. Finally, we ran a separate regression to examine the relative effects of urban and rural districts in different parts of the country.

Our results showed that while there are some instances of affirmative action for aspiring female Congress members, this effect is not limited to the Democratic party. Further, a party structure at the state level in which women are well-represented does not always translate to increased chances of women from that state moving up the political ladder. While more research could be done, particularly on the forthcoming 2016 General Election and on differences at the state level, our results provide evidence for the persistence of bias against women's nominations for higher political office.

# INTRODUCTION

The participation of women in the American political system is a relatively new phenomenon. In the country's 240 year history, women representatives have served only for the past 100 years. They still make up slightly less than 20% of the the seats in both houses of Congress.

In 1916, Jeanette Rankin (R-MT) became the first woman elected to Congress. Although not the case for Jeanette Rankin, many of the first women elected to Congress succeeded their deceased husbands' seats via appointment or special election. In fact, 47 women (out of of 252 women elected to Congress since 1916) have come to Congressional seats this way (O'Connor, 2010). Solowiej and Brunell (2003) refer to this phenomenon as the "widow effect." The authors assert that the widow effect has been a major force in reshaping the gender composition of Congress. Widows have a unique opportunity structure: their association with a deceased legislator (their husband) allows them to bypass the traditional candidate "pipeline": that is to say, widows needn't have advanced degrees or have held elected seats to run as legitimate candidates. Also, assuming risk aversion prevents women from running for Congress, widows' "celebrity" reduces the risk of loss (particularly if their husbands were powerful or well-respected legislators). Special elections or appointments of widows allows widows to "inherit" their husbands' incumbency.

O'Connor (2010) corroborates Solowiej and Brunell by describing professions in law or business, in addition to local or state government positions, as launchpads to Congress. Historically, women have not had access to these jobs, and as a result, couldn't run for Congress through the traditional pipeline. The landscape of state/local politics, however, is changing: in the 111th Congress, 55 out of 73 women in the House of Representatives had previously held elected offices. Women's increased access to pipeline jobs bodes well for increased representation in Congress, as voters tend to favor those who currently sit, or have sat, in publically elected positions. Consider the following three cases. California Senator Dianne Feinstein served for nine years as a San Francisco County Supervisor and nine years as San Francisco mayor before successfully running for the Senate in 1992 ("Dianne Feinstein…"). Alaska Senator Lisa Murkowski practiced commercial law in Anchorage and served three terms in the Alaska State House of Representatives before her successful election to Congress in 2004 ("United States Senator…"). Representative Alma Adams of North Carolina has held positions in the Greensboro City School Board, Greensboro City Council and North Carolina State assembly before entering Congress in 2014 ("United States Congresswoman…").

Women face additional challenges when competing with an incumbent. Incumbents tend to win re-elections and maintain their seats; further, they tend to serve in their positions for many years. This means that a woman has scant opportunities to run for an open seat. Because most incumbents are male, inertia makes it difficult for women to make headway in becoming incumbents themselves. Incumbency similarly challenges women's ascent to Committee chair or party leadership positions. Seniority is a significant factor in selection for these roles: many women legislators have not had their seats long enough to be earmarked for leadership role, which confounds women legislators' efforts to shape policy. For example, Senator Hattie Wyatt

Caraway (D-AR) chaired the Committee on Enrolled Bills in 1933, but another woman chair was not appointed until 1995, when Senator Nancy Kassebaum (R-KS) became chair of the Senate Committee on Labor and Human Resources. Similarly, between 1955 and 2007, only two women chaired committees in the House. Women had not made gains in party leadership until the 1980s and 1990s. Nancy Pelosi (D-CA) provides a choice, if not rare, example of a woman legislator ascending to both party and committee leadership, especially as Speaker of the House from 2007 to 2011(O'Connor, 2010).

Sanbonmatsu (2002) challenges her colleagues' traditional approach to assessing women politicians as a homogenous group. She argues that disparities in women's representation across state legislatures correlate with party: that is, women's paths to elected positions are determined differently according to party affiliation. Sanbonmatsu explains her thesis in two ways: first, the social eligibility pool for party recruitment differs between Republicans and Democrats (e.g. Democrat women tend to be in the workforce, while Republican women tend to be homemakers); second, the political opportunity structures that affect candidate emergence differ between parties (e.g. Democrat leadership tends to have tighter control over candidate selection than Republican leadership). Interestingly, Sanbonmatsu's statistical analysis suggests that Democratic party structure has a more negative effect on women candidates than does the Republican party structure.

This finding oddly contrasts Elder's (2008) discovery of a partisan gap between women legislators in the Congress (a gap which exists in both houses, and is not mirrored by male legislators who continue to be about equal between parties). Elder implicates Republican women's reduced numbers in Congress as the source of slowing growth in women's representation; she writes "slowed progress is a function of Republican women only" (2). Elder suggests that pipeline jobs (law, business and state elected office) have become increasingly important for nomination to Congress: Democratic women tend to hold these positions, while Republican women do not. Elder also points to regional realignment as a cause of reduced Republican women presence in Congress. Republicans have made gains in the South, but have suffered losses in other regions like the Northeast. This "loss of ground" in regions like the Northeast and West has affected Republican women far more strongly than Republican men; additionally, the Republican stronghold in the South is not as responsive (and perhaps even hostile) to women's candidacy. At the same time, Democrats have made gains in regions like the West, where the population happens to be receptive to women candidates. California in particular drives the partisan gap because ¼ of women legislators come from this state, and 23% of Congress' Democrats in total.

Elder (2008) refers to the increased success of minority women legislators as a final contributing factor to the partisan gap. Since 1987, women of color have made significant strides in obtaining House seats. Elder solidifies the significance of these lawmakers when she asserts that "since the late 1980s, women of color have achieved higher rates of representation among lawmakers of color than white women have among white lawmakers" (13). Elder explains that the "over-representation" of women of color applies to African American, Latina and Asian American women. In comparison, white women face steeper "under-representation" when compared to white male legislators. Further, women legislators of color are almost exclusively

Democrat. Elder concludes by referring to the success of non-white women legislators as a modest, but persistent, force in the partisan gap.

Lawless and Pearson (2008) take a somewhat unique approach in understanding women's under-representation in Congress. The authors contend that most studies about women's Congressional elections neglect the primaries process; that is, even though women tend to win Congressional elections at the same rate as men, no one had yet studied the implications of a gendered primary. To fill this dearth, Lawless and Pearson studied comprehensive data about House primary elections from 1958 to 2004. They hypothesized that three factors would negatively affect women's victories in primaries: 1) in the United States, candidates must be entrepreneurial and promote themselves, qualities which women are not socialized to develop; 2) women have less support from local political networks than male counterparts; and 3) primaries are low-turnout, low-profile events that are particularly vulnerable to stereotyping. Ultimately, Lawless and Pearson's results rejected their hypothesis ("female candidates...will face more challenging primary competition than will their male colleagues"(69)): women tended to win primaries at rates similar to men. However, in conjunction with Elder, Lawless and Pearson unveiled partisan differences for women running in primaries. No matter whether the woman is an incumbent, a challenger, or vying for an open seat, she will garner more competition in primaries than male colleagues. Republican women incumbents face more competition than Republican male incumbents in primaries. In addition, Democratic women benefit from party polarization: they are viewed as the most liberal candidates, and those most loyal to the party. The same is not true for Republican women, although they don't fare worse than male competitors. Nonetheless, Lawless and Pearson point out that the number of women competing in primaries is decreasing, and that a disproportionate number of female victories are in the Democratic party. Again, we hear that Republican women must increase in numbers in order to achieve women's equal representation in the Congress.

# RESEARCH QUESTIONS

The most common profession for Congressional representatives before their time in office is a career in the state legislatures of their respective states. 44% of senators and 51% of members of the House of Representatives have experience in their state legislatures prior to serving in Congress ("Former State Legislators in Congress and White House", 2014). With this in mind we want to investigate:

1. Whether a bias against female candidates exists between the state and federal levels of government
2. How parties might influence women's transition from local/state government to federal government

In order to investigate our question about the number of women elected officials moving from state legislature to Congress, we chose to focus our attention on women who are running for open seats in the House of Representatives. The bicameral federal legislative system ensures that each of the fifty states has both proportional and equal representation. Each state may send two representatives to the Senate, and a delegation that is proportionate to its population to the

House of Representatives. For example, the state of Montana has one representative for its entire population of 1.05 million people, and the state of New York has 27 Congressional representatives representing 19.3 million people. As the country's population increases and people migrate from one region to another, these district boundaries are redrawn and the number of seats alloted per state may shift from one state to another.

In order to build our model, gathering observations about candidates for the House of Representatives had some notable advantages:

- **More seats**: There are fifty seats in the Senate and 435 seats in the House. Incumbents in both houses tend to retain their seats for long periods of time, but members of the House of Representatives sometimes move on to the Senate or other offices.
- **Shorter terms**: Although Congress members can run for re-election an unlimited number of times, the two-year terms in the House as opposed to six years in the Senate mean that there is more turnover and more candidates to look at.

# HYPOTHESES

Given our prior research, we hypothesize that party will have a significant effect on the likelihood of female candidates being nominated to run for Congress. Due to formal mechanisms in the Democratic party to promote women, we predict that our results will show a stronger effect from the Democratic party. Secondly, we expect our regressions will show that there is a discrepancy between the percentages of women in politics at the state level and the probability of nomination of female candidates, suggesting the existence of a bias against women attempting to transition to the federal level of public service. Thirdly, we anticipate that geography will play a significant role in determining the opportunity structure for women to advance from state to federal government positions. Finally, we predict that congressional districts with stronger urban presence will be more likely to have female candidates running for election to Congress.

# GATHERING THE DATA

Our data came from four main sources online: the Federal Election Commission (FEC), the Center for American Women in Politics (CAWP), and the National Conference of State Legislatures (NCSL), and the US Census Bureau. We downloaded the FEC site's data for all candidates who had been registered to run for the last five election cycles (going back to 2008) in CSV format. Using a Python script, we extracted each candidate's name, state, congressional district, party, and election year. With the same Python script, we filtered these data to include only candidates running for open seats in the House of Representatives. The data from the FEC provided a base for building the remainder of our dataset. The party variable was coded as a dummy variable, with '1' meaning the candidate was a Democrat.

Before continuing to expand the dataset, we needed to classify all the candidates in the list as male or female. In order to do this, we took a Python dictionary available on GitHub, which was based on first names names taken from 2010 US Census Data. We wrote a second Python script

to loop through the data, split the candidates' names into first and last, and check if there was a match between each candidate's first name and the entries in the dictionary. The Python dictionary did contain some names such as 'Alex' which were listed as either male or female. We dropped these entries for ambiguous names from the dictionary before running the second Python script. After we ran the script on the data, approximately 130 of about 1860 observations remained unclassified. We looked up these 130 candidates individually and entered the data manually. Gender variables were coded as dummy variables, with '1' being female.

Our data for the share of women in each state legislature by year and party came from CAWP, with the exception of the 2014 election cycle. CAWP provided a PDF with tables going back to the 1970s with descriptive statistics on women in Congress, but the most recent year for this was 2012. CAWP's website provided a table with descriptive statistics for 2016. The data for 2014 came from the NCSL. Consequently, the data for 2014 and 2016 was scraped from two different sites. To import the data for 2008-2012 from the PDF into Python, we OCR'd the PDF, then ran the PDF through Tabula (a web application for converting PDF tables into tables in CSV format). We wrote a third Python script to clean and merge the data for all five election cycles into two data structures in Python. One data structure was a dictionary containing percentages of female Democrats in a state's legislature for a given year (marked with an identifier composed of state, party, year--e.g., 'NYREP2014'). The second data structure was the same, but contained percentages of female Republicans. We replicated the identifiers in our 'base' dataset and merged the two, so the entry for each candidate would include the share of women in the state legislature belonging to that candidate's political party for the same election year.

Our regional divisions followed the standards set by the US Census Bureau. Initially, we used nine regions, but after running some regressions decided to consolidate these into the four main Census regions of West, Midwest, South, and Northeast. These were coded as dummy variables.

**FIGURE 1: U.S. Census Bureau Regions**

| Region | States |
|---|---|
| Northeast | Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York and Pennsylvania |
| Midwest | Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota and South Dakota |
| South | Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia; Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma and Texas |
| West | Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon and Washington |

Lastly, we incorporated an 'urban' variable measuring the extent to which the candidate's district "consists of areas of high population density and urban land use resulting in a representation of the 'urban footprint'" (Census). The US Census Bureau provides 'urban percentage' tables (by state) for each congressional district in the USA. Our urban percentage variable indicates what percentage of the population is concentrated in these metropolitan areas of "high population density." We wrote a final Python script to scrape this data and merge it with our final dataset.

# EXPLORING THE DATASET

Of the resulting 1,849 candidate observations for open Congressional seats, 82% (1520) were men and 18% (329) were women. Figure 2 summarizes this information by election cycle and party affiliation. For the period observed, almost 50% of all candidates were Republican men, followed by Democrat men (33%), Democrat women (10%) and Republican women (7%). Republican women made the greatest gains in nominations during this time period: the number of candidates almost doubled from 12 in 2008 to 23 in 2016. Because these observations were collected before all states' filing deadlines, it is possible that the total number of Republican women candidates in 2016 will be even higher by the time of the general election in November. The group with the greatest decrease in candidates were male Democrats, who represented 38% of total nominees in 2008 and only 31% of total nominees as of April 2016. Although it is possible that additional male Democrats might be nominated, especially in populous California, it is unlikely that they will match the numbers represented by their Republican counterparts in November.

**FIGURE 2: Number of Candidates by Gender and Party Affiliation**

| | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| | Dem | | Rep | | Dem | | Rep | |
| | n | % total | n | % total | n | % total | n | % total |
| **2008** | 128 | 0.3787 | 159 | 0.4704 | 39 | 0.1154 | 12 | 0.0355 |
| **2010** | 121 | 0.2951 | 217 | 0.5293 | 35 | 0.0854 | 37 | 0.0902 |
| **2012** | 128 | 0.3360 | 183 | 0.4803 | 39 | 0.1024 | 31 | 0.0814 |
| **2014** | 133 | 0.3276 | 192 | 0.4729 | 45 | 0.1108 | 36 | 0.0887 |
| **2016*** | 96 | 0.3057 | 163 | 0.5191 | 32 | 0.1019 | 23 | 0.0732 |
| *Total* | 606 | 0.3277 | 914 | 0.4943 | 190 | 0.1028 | 139 | 0.0752 |

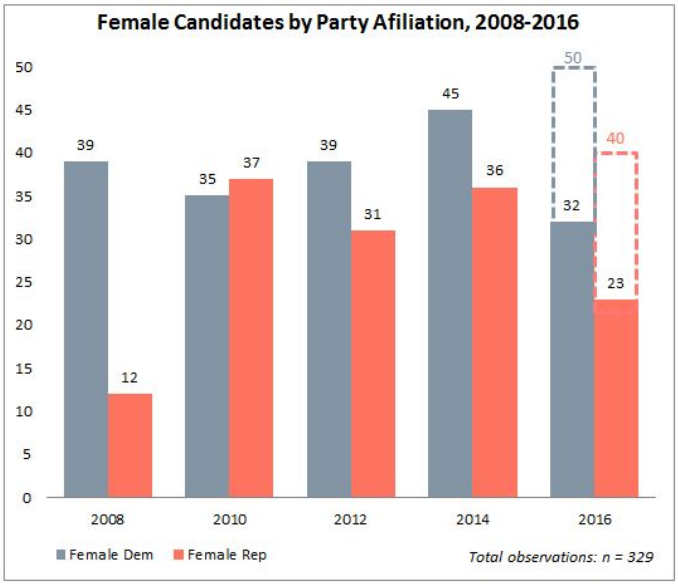
*Total observations: n = 1849*

** Candidate counts for November 2016 election as of April 2016*

Among female candidates, Democrats tended to outnumber Republicans. This observation is consistent with previous research about the mechanisms that promote women within the Democratic party (Elder, 2008; Sanbonmatsu, 2015). However, a notable exception to this can be seen in the 2010 general election, when slightly more female Republicans than female

Democrats were running for open seats, notably from the South and the Midwest. This is a continuation of the trend that Elder noticed in 2008. It could also be related to the Tea Party phenomenon which brought out many more Republican challengers, both men and women, seeking to fill open seats in Congress. Based on the observed election cycles, Figure 3 extrapolates the number of candidates for each party past the April 2016 cut-off for our scraped observations.

**FIGURE 3: Female Candidates by Party Affiliation**




Service in state or local political office is the most common profession of newly-elected Congressional representatives. In order to investigate the effect that local party structures might have on women's' chances of being nominated to run for Congress, we looked at the share of women legislators in each state for each election cycle ("PercentWomen"). We broke this data down further by party to test the effects of specific parties. Figure 4 presents the states that had the highest and the lowest percent of female legislators on average for the observation period. States with a high proportion of female state legislators tended to be in the West, Northeast, and Midwest. States with a low proportion of female state legislators tended to be from the Southern region. The interactions among region, party, and the representation of women in state politics are further explored in our regression model.

**FIGURE 4: Mean percent of women in state legislates by party, 2008-2016**

| | | Mean Dem % | Mean Rep % | Mean Total % |
|---|---|---|---|---|
| Top quartile | CO | 31.28 | 10.49 | 41.77 |
| | NH | 25.25 | 10.50 | 35.75 |
| | HI | 26.21 | 8.33 | 34.55 |
| | MN | 20.90 | 12.48 | 33.38 |
| | MD | 25.53 | 7.80 | 33.33 |
| | IL | 23.09 | 9.55 | 32.64 |
| | ME | 22.96 | 8.17 | 31.13 |
| | CT | 21.19 | 9.93 | 31.13 |
| | NJ | 19.93 | 11.19 | 31.12 |
| | AZ | 15.58 | 15.24 | 30.82 |
| | | | | |
| Bottom quartile | ID | 17.14 | -- | 17.14 |
| | MS | 13.32 | 3.69 | 17.01 |
| | VA | 11.67 | 4.78 | 16.44 |
| | PA | 6.71 | 9.56 | 16.27 |
| | UT | 10.67 | 5.33 | 16.00 |
| | ND | 3.19 | 12.77 | 15.96 |
| | SC | 7.26 | 7.59 | 14.84 |
| | LA | 9.12 | 4.89 | 14.02 |
| | OK | 4.05 | 9.80 | 13.85 |
| | AL | 9.59 | 3.98 | 13.57 |

In order to incorporate the potential effects of geography into our model, we assigned each of the candidates to one of four Census regions based on their state. Figure 5 (next page) summarizes the number of female candidates per state and Census region for each election cycle. The majority of the 329 female candidates came from southern states (113), followed by states in the West (99), the Midwest (72), and the Northeast (45). Forty-two of the female candidates came from California alone. In the regression model discussed below, we used region dummy variables to represent the effect of each of these regions, using the Southern region as a baseline.

**FIGURE 5: Number of women candidates per Census region and state**

SOUTH

| State | 2008 | 2010 | 2012 | 2014 | 2016 | Total |
|---|---|---|---|---|---|---|
| FL | 2 | 6 | 3 | 4 | 11 | **26** |
| TX | | | 8 | 2 | 3 | **13** |
| AL | 4 | 5 | | 2 | | **11** |
| TN | 1 | 9 | | | | **10** |
| MD | | | | | 8 | **8** |
| LA | 2 | 2 | | 2 | 1 | **7** |
| VA | 2 | | | 2 | 2 | **6** |
| GA | | 1 | | 5 | | **6** |
| NC | | | 3 | 3 | | **6** |
| AR | | 3 | 1 | 1 | | **5** |
| DE | | 2 | | | 2 | **4** |
| SC | | 2 | | 2 | | **4** |
| OK | | | 1 | 2 | | **3** |
| WV | | | | 2 | | **2** |
| MS | | | | | 1 | **1** |
| KY | | | 1 | | | **1** |
| **Total** | **11** | **30** | **17** | **27** | **28** | **113** |

WEST

| State | 2008 | 2010 | 2012 | 2014 | 2016 | Total |
|---|---|---|---|---|---|---|
| CA | 9 | 6 | 10 | 11 | 6 | **42** |
| AZ | 5 | 3 | 1 | 2 | 1 | **12** |
| NV | 1 | | 5 | | 4 | **10** |
| WA | | 3 | 3 | 1 | 1 | **8** |
| HI | | | 3 | 4 | | **7** |
| NM | 3 | | 2 | | | **5** |
| MT | | | 3 | 1 | | **4** |
| OR | 1 | | 2 | | | **3** |
| UT | | | 1 | 2 | | **3** |
| CO | 1 | | | 1 | | **2** |
| MO | 1 | | 1 | | | **2** |
| WY | 1 | | | | | **1** |
| **Total** | **22** | **12** | **31** | **22** | **12** | **99** |

MIDWEST

| State | 2008 | 2010 | 2012 | 2014 | 2016 | Total |
|---|---|---|---|---|---|---|
| IL | 3 | 6 | 2 | 5 | 1 | **17** |
| MI | | 4 | 2 | 7 | | **13** |
| OH | 7 | | 2 | | 1 | **10** |
| IN | | 3 | 3 | | 4 | **10** |
| KS | | 8 | | | | **8** |
| IA | | | | 5 | | **5** |
| WI | | 1 | 1 | 1 | | **3** |
| ND | | | 3 | | | **3** |
| MN | | | | 1 | 2 | **3** |
| **Total** | **10** | **22** | **13** | **19** | **8** | **72** |

NORTHEAST

| State | 2008 | 2010 | 2012 | 2014 | 2016 | Total |
|---|---|---|---|---|---|---|
| NY | 3 | 1 | 2 | 2 | 5 | **13** |
| NJ | 3 | | 2 | 6 | | **11** |
| PA | | 3 | | 3 | 2 | **8** |
| MA | 1 | | 3 | | | **4** |
| ME | 1 | | | 2 | | **3** |
| NH | | 3 | | | | **3** |
| CT | | | 2 | | | **2** |
| RI | | 1 | | | | **1** |
| **Total** | **8** | **8** | **9** | **13** | **7** | **45** |

Another aspect of geography that we incorporated into our model was the degree to which each district could be considered urban (as opposed to rural). These classifications were obtained from the U.S. Census Bureau and mapped to the observations in our data set. Figure 6 (next page) summarizes the average "urbanness" of the candidates' districts by region, gender, and party. Overall, our dataset corroborates findings by previous researchers that female candidates tend to come from more predominantly urban districts (Mean = 81.14%, S.D. = 15.30%) than male candidates (Mean = 76.83%, S.D. = 15%). Additionally, across all regions and for both genders, Democratic candidates' districts tended to be more urbanized than Republican candidates' districts. The effect of the urbanization of districts on female candidates will be further explored in the regression model below.

**FIGURE 6: Average percentage of urbanized districts by region**

| | Male | | | Female | | | |
|---|---|---|---|---|---|---|---|
| | Dem | Rep | All | Dem | Rep | All | Avg All |
| **ALL** | | | | | | | |
| *Avg Percent Urban* | 80.92 | 74.13 | 76.83 | 83.29 | 78.19 | 81.14 | 77.60 |
| *SD* | 15.73 | 15.20 | 15.00 | 14.86 | 17.02 | 15.30 | 15.03 |
| **WEST** | | | | | | | |
| *Avg Percent Urban* | 85.86 | 82.91 | 84.30 | 88.16 | 87.82 | 88.04 | 85.04 |
| *SD* | 13.14 | 13.11 | 12.92 | 12.47 | 13.99 | 13.08 | 12.99 |
| **SOUTH** | | | | | | | |
| *Avg Percent Urban* | 75.13 | 69.52 | 71.40 | 80.92 | 73.52 | 77.25 | 72.33 |
| *SD* | 17.29 | 14.70 | 15.31 | 13.00 | 16.27 | 14.12 | 15.21 |
| **MIDWEST** | | | | | | | |
| *Avg Percent Urban* | 80.91 | 72.21 | 75.60 | 80.38 | 76.79 | 78.74 | 76.18 |
| *SD* | 9.31 | 9.57 | 9.66 | 14.36 | 10.28 | 11.18 | 9.76 |
| **NORTHEAST** | | | | | | | |
| *Avg Percent Urban* | 83.22 | 77.17 | 79.93 | 81.44 | 75.38 | 79.55 | 79.86 |
| *SD* | 20.13 | 21.75 | 20.62 | 21.47 | 26.11 | 22.25 | 20.80 |

# BASE MODEL

The base regression model that we are using is a logistic regression:

*P(woman nominated to run for Congress) = α + β*share of women per party in that state + γ*party (0/1) + ε*

From our dataset, we are estimating the probability of a woman running for congress by analyzing the percentage share of women in the state legislature and party to which each candidate in the dataset belongs. Therefore we are finding the average probability on the estimate of *GenderVar,* the dummy variable that indicates if a candidate in the dataset is male or female.

**Coefficients**

Running the regression model is expected to identify biases that exist when it comes to estimating this probability. The coefficients are primary identifiers of different kinds of biases:
*α represents affirmative action offered by that state*
*β represents a bias that exists with respect to women representation per party for that state.*
*γ represents the effect of the Democratic party over the 4 cycles for all states.*

**An "Ideal" State**

An ideal state is one where there exists total equality i.e., no gender biases which leads to the lack of need of an affirmative action. Since there are no biases and everything is equal, biases with respect to women belong to different parties would also not exist. This translates to the following coefficient values:
$\alpha = 0$
$\beta = 1$
$\gamma = 0$
This means that the probability of any given woman X running for Congress is dependent on the share of women per woman X's party per woman X's state and not on affirmative action or party variable as equality of gender exists and party representation is irrelevant.

**Biases that exist due to party differences**

If a party invests more in its male candidates as compared to its female candidates, then we can say a bias exists.
$\alpha \neq 0$
$\beta < 1$
$\gamma > 0$

**Biases that exist due to random events**

If things such as career ceilings etc start to affect the female candidates' chances of running for Congress then these effects are calculated within the epsilon.
$\alpha \neq 0$
$\beta < 1$
$\gamma = 0$

In order to explore more variables that might affect the probability of women running for Congress, we decided to categorize each candidate's district according to the four main US Census regions: Northeast, South, Midwest, and West.

We also decided to explore the effect of urbanization on each region and conducted interactions between the percentage of women variable with the party variable as well as the various regions to understand the effect of the women composition per region per party.

# REGRESSION ANALYSES AND INTERPRETATIONS

## Base Regression

As explained above, the purpose of our base regression is to find the biases that occur when estimating the probability that a woman will be nominated to run for Congress by only looking at the percent of women running and party affiliation and no other factors. The results of the regression are presented below, where an analysis is shown for alpha and the variables, PercentWomen and PartyVar.

After running the base regression, we saw that **alpha** is .106258, and the p-value proves it to be significant. **PercentWomen** equals .003018 with a p-value of .111738, making it borderline significant. The coefficient for **PartyVar** is .082207 and is significant with a p-value of .000486.

## Regressions By Region

In order to understand how the variables PercentWomen and PartyVar differ across regions, we conducted four regressions, respectively subsetting them by Northeast, South, Midwest and West.

In the Northeast, **PercentWomen** is -0.004330, but the coefficient is not statistically significant. Further, this coefficient approaches zero. However, **alpha** and **PartyVar** are larger in the Northeast (0.139870 and 0.190580, respectively) than in any other region. As such, Northeast women's nominations to Congress are strongly informed by affirmative action; further, because PartyVar is a large, positive value, we know that affirmative action is stronger among Democrats.

In the South, **alpha** (0.074761) and **PartyVar** (0.073825) are statistically significant, although two to three times smaller than in the Northeast. Further, **PercentWomen** (0.005975) is borderline significant, with a p-value of 0.1014. The South is the only region in which PercentWomen approaches significance.

The Midwest's and West's coefficients are similar to the Northeast, withstanding party effect. In the West, **PartyVar** is 0.073234 and in the Midwest, **PartyVar** is 0.070424. Their **alphas** are

about equal, each approaching 0.12. Thus, both Republicans and Democrats use affirmative action (with slightly stronger affirmative action among Democrats). **PercentWomen** does not have statistical significance in either region.

We were able to compare regional differences more substantially in a subsequent regression which included interactions between Region and PercentWomen; between Region and PartyVar; and between Region/PartyVar/PercentWomen.

## Regression on Combined Interactions

We then decided to compare the interaction effects of our variables. Our observations were aimed at testing if a democratic party structure made a difference to the share of women in the legislature that aided the probability of a woman running for Congress. We also wanted to observe the effect of women's representation within the four regions and see if it is significant with respect to the mean probability of a woman running for Congress, followed by observing the effect of the Democratic Party in a region-by-region analysis and its effect on the mean probability of a woman running. Finally, we observed the effect of the Democratic Party and the resulting women's composition, region-by-region while estimating the mean probability.

**Model**

lm(formula = *GenderVar ~ PercentWomen + PartyVar + Northeast + Midwest + West + (PartyVar *PercentWomen) + (Northeast * PercentWomen)+(Midwest * PercentWomen) + (West* PercentWomen)+ (Northeast *PartyVar) + (Midwest * PartyVar) + (West * PartyVar) + (Northeast *PartyVar * PercentWomen) + (Midwest * PartyVar * PercentWomen) +(West * PartyVar * PercentWomen)*, data = women)

In this model, we interacted the following variables:

1. PercentWomen and PartyVar
2. PercentWomen and Regions (Northeast, Midwest, West)
3. PartyVar and Regions (Northeast, Midwest, West)
4. PartyVar, Regions and PercentWomen

Below are all the significant variables and their interpretations:

**Intercept**

The estimate 0.1041173 which has a p-value that is significant at the 5% level indicate the presence of affirmative action for both parties.

**PercentWomen:PartyVar:Midwest**

The estimate of -0.0331362 at a significance level of 5% suggests that the Republicans make more of an effort than Democrats to nominate women in state legislatures within the Midwest.

**PercentWomen:Midwest**

The estimate of 0.0209407 at a significance level of 10% indicates that woman composition in

the state legislatures of the Midwestern region affects the likelihood of a woman running for Congress.

**PartyVar:Northeast**

The estimate of 0.2969982 at a significance level of 10% shows that party structure is influential in the likelihood of a woman running for Congress in the Northeast in comparison to other regions (with the exception of the Midwest).

**PartyVar:Midwest**

The estimate of 0.3578729 at a significance level of 5% indicates that party structure influences the probability of a woman in the state legislature to run for congress especially from the Midwest (followed by the Northeast). **PartyVar:West** (0.0129050) is almost the same estimate as the influence of **PartyVar** on the South (0.0127278). With regard to these two estimates, we can infer that the West is liberal enough to not depend on the Democratic party to influence women running for Congress, while the South probably depends the least on the a Democratic party structure to influence women running for Congress, as the South tends to be more conservative and less likely to elect female candidates.

## Summary

In summary, both parties seem to invest in the prevention of any unlawful discrimination. However, the Midwest seems to be an interesting region under observation. The Midwest is significant in its representation of women in the state legislature, specifically by the Republican party having more of an impact on the likelihood of a woman running for Congress followed closely by the effect of party structure in the Northeast.

## Regressing on Urban Percentages

Finally, we were interested in understanding the effects of social and economic factors in each of the districts on women candidates' tendency to be nominated. Drawing on our background research, we were especially interested in the degree to which different districts could be considered "urbanized" based on the percentage of the total population living in more densely populated metropolitan areas. We pulled percentages from the US Census Bureau and created a 'percentage urban' variable (**PerUrban**). We interacted **PerUrban** with each of the region variables to see if there were notable differences in the effect of urban areas across regions.

**Model**

P(woman nominated to run for Congress) = $\alpha$ + $\beta$*share of women per party in that state + $\gamma$*party (0/1) + PerUrban + Midwest*PerUrban + Northeast*PerUrban + West*PerUrban+ $\varepsilon$

We found that **PartyVar** (0.0825493) was again statistically significant with a p-value of 3.481. **PerUrban** was also significant at a 5% level. Urbanized areas seemed to play a significant role in the South, our baseline region. Urban areas also appeared to play a positive, but not statistically significant role, in the West (0.0004070, p-value = 0.754749). In the Midwest, the

role of urban areas was slightly negative, but again not statistically significant (-0.0009112, p-value = 0.482304). In the Northeast, the role of urban areas approached 10% significance with a p-value of 0.1096. Here the effect was negative (-0.0020903): being from an urban area seems to decrease women's prospects for nomination.

# DISCUSSION

**Model Expansion**

We expanded our initial model based on different theories of what could be in the error term. From previous research, we were able to find different factors that affected our model. Two of these factors are whether a district is friendly towards women candidates and whether a woman candidate belongs to either the Democratic or Republican party. There are also other influences that determine if a particular district is "woman-friendly" and the difference in the amount of support given to women candidates depending on the political party they are affiliated with.

Women candidates are more reliant on party support than their male counterparts. Democratic women tend to receive more party support than Republican women. This support comes in various forms, including finance, recruitment, and media. Republican women are at a higher financial disadvantage because the Republican party contributes very little money to female candidates. Democratic women benefit from party support and PAC organizations granting financial support to Democratic candidates. PAC organizations also provide support in the form of recruitment, which equips candidates with encouragement, training, and other resources. These efforts are not as prevalent within the Republican Party (Sanbonmatsu, 2015). Party affiliation and geographic region are interconnected and also influence one another.

Aspiring candidates are also dependent on the geographic location and economic structures of their districts. Democratic women tend run in more "women-friendly districts" (these districts are more liberal, urban, educated, and diverse) which can affect the amount of support they receive. Republican women are more likely to represent districts that are more urban, but less diverse than their male counterparts (Sanbonmatsu, 2015). Our own model demonstrated the effects of geography on women's political opportunities. In both the Northeast and the Midwest, women seem to benefit from some measure of affirmative action from their respective parties. Even though Southern women did not seem to benefit from affirmative action, they did benefit from running in more urbanized districts. While our models did not indicate any statistically significant factors that might be helping women from the Western states seek seats in Congress, a basic survey of our data showed that they represented a large share (30%) of all female candidates and had a generally high number of women in state legislatures. Further study might be done at the state level to investigate what other factors are promoting women in this region.

**Limitations**

Our dataset had certain limitations which may have impacted the accuracy of our model's estimates. First, as mentioned earlier, additional candidates' names have likely been added to the FEC summaries after we downloaded the data for the 2016 election cycle, which is still ongoing at the writing of this paper. Although it remains to be seen at the time of the general election, there may be a significant number of female nominees for 2016 who were not included in our data.

Secondly, it was somewhat difficult to ensure that the urban percentages for each congressional district were matched correctly to the right election cycle, since several congressional districts were redrawn (and new districts created) in the last eight years. There was no clean and simple way to map the data from the Census tables to our existing data. In our dataset, time was defined by election cycles, whereas for the Census data, timeframes were divided by Congress terms, with certain terms omitted. We were able to map terms to election cycles, but it became a bit complicated for the 2012 election cycle, right before the redrawing of certain districts in 2013. This ambiguity introduced the possibility of some errors in our urban percentages for candidates' districts. There were likewise some ambiguous gaps in the urban percentage data provided from the US Census tables which had to be filled. Most notably, "districts at large" (states that are not divided into districts, but count as one district in themselves) were absent from the Census tables. We had to go to another source to obtain urban percentages for these districts (Decennial Census data provided by Iowa State University), but the data was provided according to decennial periods, so all districts-at-large only had one urban percentage entered for all five election cycles, whereas other districts did not.

# CONCLUSION

Despite the gains made by women over the last century, the share of male and female representatives in Congress continues to skew strongly toward men on both sides of the aisle. Our research shows that both political party membership and geography play a role in this divide by affecting the opportunity structure of women who are interested in running for a seat in the U.S. House of Representatives. Based on the candidate data that we collected and our regression models, it is clear that some form of bias against women still exists between the state government and federal levels of government. Further research is needed to identify other factors that might be contributing to these differences as well as opportunities to increase women's representation.

# REFERENCES

"Dianne Feinstein: United States Senator for California: Biography." Retrieved from http://www.feinstein.senate.gov/public/index.cfm/biography

Elder, L. (2008).Whither Republican Women: The Growing Partisan Gap among Women in Congress. *The Forum, 6(*1), 1-21.

"Former State Legislators in Congress and White House" (2014). *National Conference of State Legislatures*. Retrieved from http://www.ncsl.org/ncsl-in-dc/publications-and-resources/former-state-legislators-in-congress.aspx.

Lawless, JL & Pearson K. (2008). The Primary Reason for Women's Underrepresentation? Reevaluating the Conventional Wisdom. *The Journal of Politics. 70*(1) 67-82.

O'Connor, K. (2010). Gender and Women's Leadership: A Reference Handbook. Thousand Oaks, California: SAGE Publications, Inc.

Sanbonmatsu, K. (2002). Political Parties and the Recruitment of Women to State Legislatures. *The Journal of Politics, 64*(3), 791-809.

Sanbonmatsu, K. (2015) Women's Election to Congress. *Political Parity Research Inventory*. Retrived from https://www.politicalparity.org/research-inventory/

Solowiej, L & Brunell, TL.(2003). The Entrance of Women to the U.S. Congress: The Widow Effect. *Political Research Quarterly, 56*(3), 282-292.

"United States Senator Lisa Murkowski for the State of Alaska: Biography." Retrieved from http://www.murkowski.senate.gov/public/index.cfm/biography

"United States Congresswoman Alma Adams for the Twelfth District of North Carolina: Full Biography." Retrieved from http://adams.house.gov/about/full-biography

U.S. Census Bureau (n.d.). "Urban and Rural classification." Retrieved from https://www.census.gov/geo/reference/urban-rural.html

# DATA SOURCES

Federal Election Commission (Summaries for Registered Candidates):
http://www.fec.gov/data/CandidateSummary.do?format=html

Census regions/divisions:
http://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Urban census data:
https://www.census.gov/geo/maps-data/data/cd_state.html

Dictionary of First Names from 2010 US Census:
https://github.com/Bemmu/gender-from-name

Percentages of Women in State Legislatures in 2008-2012:
http://cawp.rutgers.edu/sites/default/files/resources/stleghist.pdf

Percentages of Women in State Legislatures in 2014:
http://www.ncsl.org/legislators-staff/legislators/womens-legislative-network/women-in-state-legislatures-for-2014.aspx

Percentages of Women in State Legislatures in 2016:
http://www.cawp.rutgers.edu/women-state-legislature-2016

Urban Percentages for Districts at Large:
"Urban Percentage of the Population for States, Historical"
http://www.icip.iastate.edu/tables/population/urban-pct-states

# APPENDIX I: Base Regression

```
Call:
lm(formula = GenderVar ~ PercentWomen + PartyVar, data = women)

Residuals:
      Min    1Q  Median   3Q    Max
-0.2860 -0.2247 -0.1364 -0.1206  0.8937

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.106258   0.019952   5.326 1.13e-07 ***
PercentWomen 0.003018   0.001896   1.591 0.111738
PartyVar     0.082207   0.023523   3.495 0.000486 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3788 on 1846 degrees of freedom
Multiple R-squared:  0.02042,    Adjusted R-squared:  0.01936
F-statistic: 19.24 on 2 and 1846 DF,  p-value: 5.357e-09
```

# APPENDIX II: Regression on Combined Interactions

```
Call:
lm(formula = GenderVar ~ PercentWomen + PartyVar + PartyVar *
      PercentWomen + Northeast + Midwest + West + Northeast * PercentWomen +
      Midwest * PercentWomen + West * PercentWomen + Northeast *
      PartyVar + Midwest * PartyVar + West * PartyVar + Northeast *
      PartyVar * PercentWomen + Midwest * PartyVar * PercentWomen +
      West * PartyVar * PercentWomen, data = women)

Residuals:
      Min    1Q  Median   3Q     Max
-0.3508 -0.2193 -0.1343 -0.1109  0.9394

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       0.1041173  0.0513850   2.026   0.0429 *
PercentWomen                      0.0023854  0.0058955   0.405   0.6858
PartyVar                          0.0127278  0.0837239   0.152   0.8792
Northeast                         -0.0226606  0.0884218  -0.256   0.7978
Midwest                           -0.1707772  0.1090565  -1.566   0.1175
West                              0.0282564  0.0773910   0.365   0.7151
PercentWomen:PartyVar             0.0061516  0.0077166   0.797   0.4254
PercentWomen:Northeast            0.0018198  0.0110072   0.165   0.8687
PercentWomen:Midwest              0.0209407  0.0117650   1.780   0.0753 .
PercentWomen:West                 -0.0009801  0.0080812  -0.121   0.9035
PartyVar:Northeast                0.2969982  0.1591026   1.867   0.0621 .
PartyVar:Midwest                  0.3578729  0.1650462   2.168   0.0303 *
PartyVar:West                     0.0129050  0.1437093   0.090   0.9285
PercentWomen:PartyVar:Northeast -0.0178021  0.0133222  -1.336   0.1816
PercentWomen:PartyVar:Midwest   -0.0331362  0.0140364  -2.361   0.0183 *
PercentWomen:PartyVar:West        -0.0027491  0.0108006  -0.255   0.7991
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.379 on 1833 degrees of freedom
Multiple R-squared:  0.02633,    Adjusted R-squared:  0.01836
F-statistic: 3.305 on 15 and 1833 DF,  p-value: 1.696e-05
```

# APPENDIX III: Urbanization Effect

```
Call:
lm(formula = GenderVar ~ PercentWomen + PartyVar + PerUrban +
      Midwest * PerUrban + Northeast * PerUrban + West * PerUrban,
      data = women)

Residuals:
      Min   1Q  Median  3Q    Max
-0.2954 -0.2165 -0.1435 -0.1008  0.9414
90   1.468 0.142209
West              -0.0381973  0.1084
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.0107697  0.0561614  -0.192 0.847948
PercentWomen      0.0020347  0.0020175   1.009 0.313346
PartyVar          0.0825493  0.0237164   3.481 0.000512 ***
PerUrban          0.0016675  0.0007458   2.236 0.025474 *
Midwest           0.0787771  0.0999837   0.788 0.430857
Northeast         0.1545457  0.10525735  -0.352 0.724778
PerUrban:Midwest   -0.0009112  0.0012967  -0.703 0.482304
PerUrban:Northeast -0.0020903  0.0013060  -1.601 0.109641
PerUrban:West     0.0004070  0.0013028   0.312 0.754749
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3785 on 1839 degrees of freedom
Multiple R-squared:  0.02569,    Adjusted R-squared:  0.02093
F-statistic: 5.389 on 9 and 1839 DF,  p-value: 2.598e-07
```