Title: Political Scientists: A Profile of the Congressional Candidates with STEM Backgrounds

One Sentence Summary: Novel data profile the educational, political, and demographic backgrounds of nearly two-hundred STEM candidates who ran for Congress in 2018, describe whether or not they were successful, and shed light on their motivations for running.

Author: Matthew P. Motta, PhD

Affiliation: Annenberg Public Policy Center, University of Pennsylvania

ABSTRACT. Candidates with STEM backgrounds are running for office in record number in the 2018 Midterm Elections, leading some to dub the cycle the "year of scientists running for Congress." Journalistic profiles of several of these candidates have greatly advanced our understanding about who is running for Congress in 2018, and why. However, our understanding of this unique moment in American electoral history is incomplete, as few have attempted to study these candidates as a collective group. I build on previous research by offering a data-driven profile of the background (e.g., political affiliations; areas of scientific expertise), motivations, and electoral performances of STEM candidates who are running for Congress in 2018. I present three key sets of findings. First, of the nearly two hundred STEM candidates running for Congress this cycle, I find that most have advanced degrees, are affiliated with the Democratic party, and are disproportionately male. Second, I find that about a third of STEM candidates advanced to the general elections; cautioning that scientific expertise is no guarantee of electoral success. Third, I present evidence consistent with the idea that STEM candidates were mobilized to run for Congress in response to the presidency of Donald Trump. Nearly three-quarters of STEM Democrats ran for office for the first time during Trump's presidency. These findings not only improve our understanding of the role that STEM candidates are playing in the 2018 Midterms, but advance our knowledge of the motivations underlying scientific advocacy, and its potential effectiveness.

Main Text:

Congressional candidates with STEM backgrounds are running for office in record number in the 2018 Midterm Elections *(1–2)*, leading some to dub 2018 the "year of scientists running for Congress." (*3*) Journalistic profiles of individual candidates have greatly advanced our understanding of this unique moment in American electoral history. Still, few, have attempted to study Congressional candidates with scientific backgrounds *as a collective group.*

Systematically studying Congressional candidates with STEM backgrounds is important, as it can help us better understand the types of scientists who are running, what their motivations are, and whether or not they were successful. Similar efforts to quantify scientists' political involvement in the "Trump Era" – including participation in the March for Science *(4–5)* – have furthered our understanding of who participates in pro-science political advocacy, and how their political involvement shapes politics more broadly *(6).*

Consequently, systematically studying these candidacies, as a group, may not only advance our understanding of scientists' role in this Midterm election cycle, but for generating predictions about how scientists might fare if and when they run for office in the future. It can also be useful for highlighting areas in which scientists can run more effective and inclusive campaigns.

To facilitate the systematic study of these candidates, I assembled a new dataset profiling 194 candidates with STEM backgrounds who ran for Congress in 2018. These data are free to download at the [DATA REPOSITORY URL], and are structured such that rows list unique candidates, and columns correspond to various attributes about them (e.g., degree attainment, field of study), and their districts (e.g., the percentage voting for Donald Trump in 2016, the percentage who are college educated).

I collected these data by pulling a list of candidate names – as well as their corresponding Congressional Districts and fields of expertise – from `VoteSTEM.org`; a group that tracks all candidates with at least a Bachelors degree in a STEM field. I supplemented this list with publicly available data *(7)*, which profiled hundreds of Congressional candidates on various dimensions: including whether or not those candidates had STEM backgrounds. I cross-validated data from both sources by manually looking up each candidates' educational background and area of scientific expertise.

I then merged in data from a number of different sources to provide a more-detailed profile of the candidates' political, scientific, and district-level backgrounds: including from the Center for Responsive Politics (e.g., campaign contributions from pro-science groups), the US Census (e.g., district educational attainment level), `Ballotpoedia` (e.g., candidate party affiliation, election outcomes), the Cook Political Report (e.g., district-level partisan lean), and candidates' personal websites (e.g., gender, degree attainment). Full details on how the data were collected can be found in the Supplementary Materials.

Figure 1 summarizes several important descriptive findings about who the types of candidates with STEM backgrounds who ran or are running for Congress. First, concerning the candidates' political affiliations, I find that the vast majority (about 80%) are affiliated with the Democratic party (Panel A). However, it is important to note that the candidates were not *exclusively* Democrats. About 15% are affiliated with the Republican party, and 6% are third party candidates.

Additionally, Panels B and C in Figure 1 profile the candidates' STEM backgrounds. Interestingly, nearly half of these candidates have a doctoral degree (49%, see Panel B) – 43% of whom hold a doctoral degree in a STEM field, while an additional 6% hold at least a bachelors degree in a STEM field, with an doctoral degree in another field (e.g., someone who majored in a STEM field in college who went on to earn a Juris Doctorate; please consult the Supplementary Materials for more information). Moreover, Panel C shows that while the candidates' range of scientific expertise is broad, a majority come from just three fields; engineering (28%), medicine (25%), and computer science (11%).

Figure 1 also displays a prominent gender gap (Panel D). Less than one quarter of the STEM candidates who ran for Congress in 2018 were women (24%). This stands in notable contrast to the number of women seeking elected office more generally in 2018. For example, a recent study found that nearly half (48%) of all Democratic nominees for federal and gubernatorial races were women (*7*).

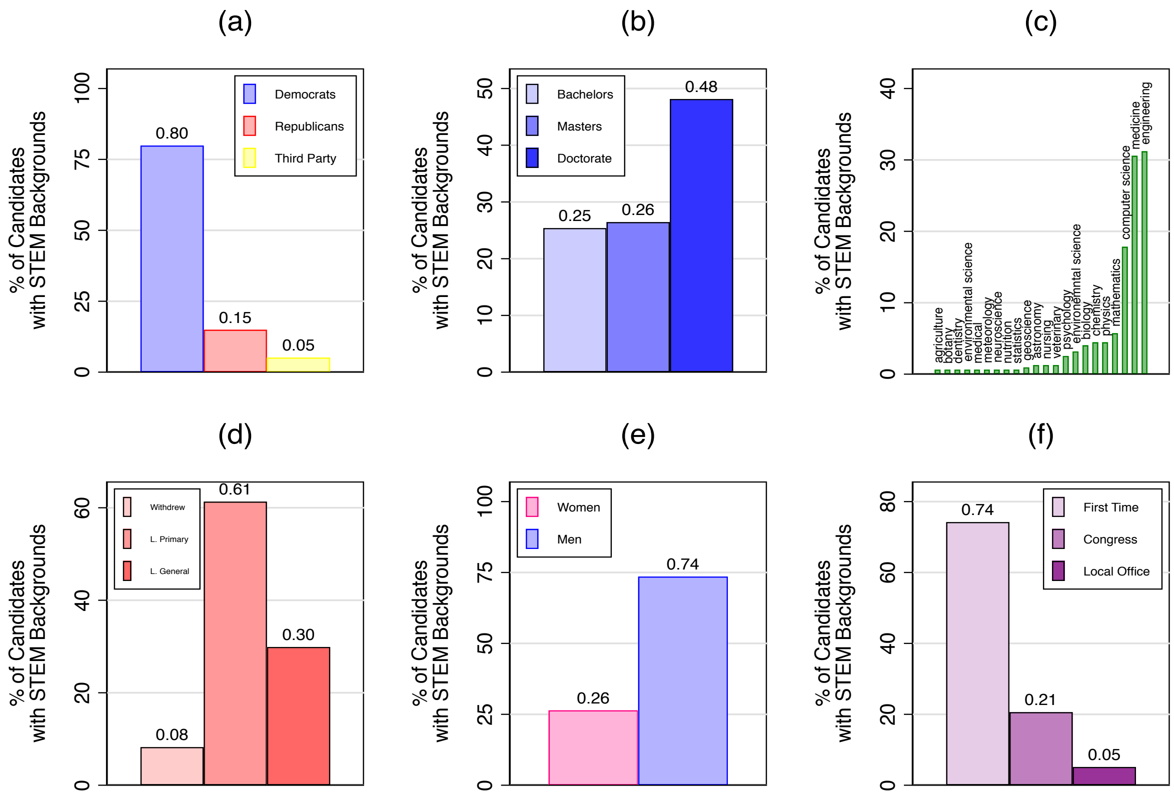

Figure 1. A Descriptive Profile of the Candidates with STEM Backgrounds who Ran for Congress N = 194. Please consult Table S1 in the Supplement for additional information about these variables.

In addition to describing the STEM candidates who ran for Congress in 2018, these data offer the opportunity to ask whether or not they were *sucessful*. Figure 1 (Panel E) shows that less than one third (30%) of the STEM candidates who ran for Congress in 2018 advanced beyond the primaries. While some (8%) were eliminated prior to primary contests, most lost in primary elections (61%). These data are consistent with the idea that candidates with scientific backgrounds are performing somewhat below the average candidate *(7)*, and caution that scientific expertise is no guarantee of electoral success.

Finally, the unprecedented number of candidates with STEM backgrounds who ran for Congress gives us the unique opportunity to ask *why* these candidates ran. Existing efforts in the popular press to profile some of these (mostly Democratic) candidates suggest that scientists ran for office as the result of dissatisfaction with President Trump's stance on science-relevant issues, and efforts to curtail the role that scientists play in the policymaking process *(2, 9)*. Additionally, given Democrats' strong polling performance leading up to the 2018 Midterms, the STEM candidates who ran in 2018 – who, as Figure 1 suggests, were predominately Democrats – may have simply felt that the political conditions for were amenable for their candidacies *(10)*.

If Donald Trump's presidency was indeed a mobilizing force for STEM candidates, we might expect that most who felt compelled to run in 2018 have never sought elected office in the past *(11)* – particularly amongst Democrats. If true, of course, this would not definitively demonstrate why these candidates chose to run. However, if most Democrats with STEM backgrounds are first-time Congressional candidates, this would at least provide information *consistent* with journalistic accounts of their motivations.

Figure 1 (Panel F) suggests that the overwhelming majority of STEM Democrats who ran in 2018 have never sought elected or Congressional office, prior to Trump's presidency. Three-fourths (75%) of the STEM Democrats who ran for office in 2018 have no previous experience with electoral politics, and about one fifth (20%) have previously run for Congressional or state-wide office in the past. An additional 5% previously ran for local offices in their cities and towns.

Overall, this analysis offers several important conclusions about the STEM candidates who ran for Congress in 2018. First, while these candidates were mostly Democrats, they are not *exclusively* Democrats; about a fifth belong to either the Republican party or a third party. Second, this analysis suggests that about a third of STEM candidates who ran in 2018 were able to advance to the general election. Although more media attention is likely to be given to the candidates who advanced to the general election, the data caution that scientific credentials are no guarantee of electoral success.

Third, these data help shed light on *why* candidates with STEM backgrounds ran for Congress. Consistent with the idea that STEM candidates ran in opposition to Donald Trump's performance in office, the overwhelming majority of STEM Democrats ran for Congressional office for the first time during Trump's presidency; consistent with journalistic evidence suggesting that STEM candidates were motivated to run in response to President Trump's tenure in office.

These data, of course, are not without limitations. As noted earlier (and expanded on in the Supplement), the process of determining who counts a STEM candidate, and profiling various dimensions of their scientific backgrounds, is – on some level – subjective.

Although I have attempted to provide transparent and consistent coding decisions, I hope scholars will view this data as less of a “final word” and more a blueprint for future research in this area.

This analysis offers several lessons to keep in mind for the 2020 general elections and beyond, as well as opportunities for additional research. For example, whether or not the candidates who ultimately may win seats in Congress take legislative action to advance pro-science causes is an open question. As I detail in the Supplementary Materials, various aspects of legislative performance data can be easily linked into this data. Moreover, I investigate only STEM candidates running for the nation’s highest legislative office. In the future, scholars may wish to conduct similar efforts to study those running for state and local offices, using procedures similar to those outlined here.

Finally, my analysis suggests a stark gender gap amongst the candidates with STEM backgrounds who ran in 2018. Future research should make an effort to better understand why this gender gap exists; e.g., due to the under-representation of women in STEM, whether or not STEM women are being given adequate incentives to run, and/or if STEM women are opting not to run for some other reason. Studying how parties and interest groups make efforts to train, mobilize, and recruit women with STEM backgrounds to run may be a particularly instructive area for future research on this subject.

References:

1. J. Mervis, Meet the scientists running to transform Congress in 2018. *Science.* doi:10.1126/science.aat3 (2018).

2. D. Rauf, For scientists runing for Congress, victory doesn’t depend on science. *Scientific American* (2018).

3. B. Guarino, L. McGinley, 2018 is the year of scientists running for Congress. *Washington Post* (2018)

4. D. Fisher, Scientists in the Resistance. *Sociological Forum.* **33** 247-250 (2017)

5. M. Motta, The polarizing effect of the March for Science on attitudes toward scientists. *PS: Political Science and Politics.* e1-6 (2018).

6. R. Brulle, Critical reflections on the March for Science. *Sociological Forum.* **33** 255-258 (2018).

7. M. Conroy, Primary candidates 2018. *FivetThirtyEight* (data repository). www.github.com/fivethirtyeight/data (2017).

8. J. Mervis, A house too far: two scientists abandon their bids for Congress. *Science* doi:10.1126/science.aau0136 (2018).

9. A. Sifferlin, Why more scientists are running for office in 2018. *Time Magazine* (2018).

10. J. Mervis, The science candidates: Kosper builds big tent after win in Texas. *Science* doi:10.1126/science.aau3328 (2018).

11. E. Yong, Here's how the scientists running for office are doing. *The Atlantic* (2018).

Acknowledgements: I think Asheley Landrum, Meredith Conroy, Robert Ralston, and Kathryn Pearson for their invaluable feedback. **Funding:** no funding to report. **Authors contributions:** MM contributed to all aspects of this project, including conceptualization, data collection, analysis, and writing. **Competing interests:** I declare no competing interests. **Data and materials availability**: All data used for this study (including raw data, codebooks, and other materials materials), as well as replication code can be found in the Supplementary Materials.

Supplementary Materials

Materials and Methods

STEM Candidate Identification Procedure

Building the dataset used in this manuscript entails identifying a list of Congressional candidates with scientific backgrounds, and subsequently merging in data from other sources (either automatically or manually) to provide additional information about their districts and personal backgrounds. As a reminder, all data is publicly available at [LINK TO REPOSITORY WILL BE INSERTED HERE IF ACCEPTED]. A summary of all variables provided in the dataset (even those not analyzed in this manuscript) can be found in Table S1.

I begin the data collection process by first pulling the **names**, **districts**, and **fields** of candidates with scientific backgrounds running for office, from the non-partisan group `VoteSTEM.org`. VoteSTEM considers a candidate to have a scientific background if that person has a least a bachelors degree in a STEM field. So, a candidate with a BS in chemistry would be listed; as would someone who got a BA in studio art, but then went on to get a PhD in electrical engineering. As noted in the paper, only about a quarter (28%) of candidates tracked had only a BS. These data (archived in August 2018) can be found here: https://web.archive.org/web/20180712092403/https://votestem.org/.

Recognizing that these data may be imperfect or incomplete, I supplement this list by appending data from FiveThirtyEight's list Democratic politicians running for Congress in 2018. These data were collected by Conroy and colleagues, and can be accessed on FiveThirtyEight's Github page (https://github.com/fivethirtyeight/data/tree/master/primary-candidates-2018). Although these data lacked some of the information that VoteSTEM provided (e.g., candidates' fields of expertise), it was nevertheless quite useful in ensuring the completeness of this dataset. This procedure identified an additional 37 candidates not reflected in the VoteSTEM data; meaning that this step increased the volume of candidates in the dataset by about 19%.

I considered candidates to have a STEM background if they (1) have at least a Bachelor's degree in a STEM field, (2) are employed or have employment experience in a STEM field (potentially without at least a Bachelor's degree), or (3) have an advanced degree in a STEM field (even if their Bachelor's degree is in something else). Individuals with degrees in social science or humanities fields were not included in this analysis.

One important caveat to note is that candidates who have bachelors degrees in scientific fields, but who go on to get advanced degrees in fields that are (potentially) unrelated to STEM, *are* classified as STEM candidates in these data. I choose to count these individuals for two reasons. First, although candidates with scientific backgrounds may go on to specialize in other fields, they may nevertheless use and campaign on their scientific experience. For example, Rep. Ted Leiu (D - CA) has an undergraduate degree in computer science, as well as a JD in law. Although his primary occupation prior to serving in Congress was as a lawyer, he frequently claims credit for his computer science degree; for example, he writes on his official US House webpage:

> "As one of only four computer science majors currently serving in Congress, Ted is frequently sought out for his insight on technology and innovation matters including cybersecurity, cloud computing and innovation as well as the sharing and creative economy." Source: https://lieu.house.gov/about/full-biography

Second, there are only 11 cases in this data (about 7% of the dataset) of people with scientific backgrounds who went on to specialize in fields not directly related to science. Erring on the side of inclusivity, I retain these people in the dataset.

The VoteSTEM data also include descriptions of candidates' fields of expertise; ascertained from their websites, Ballotpedia, and/or other public statements. The FiveThirtyEight data do not contain this information, so I pulled it manually (cross-validating the codes provided from VoteSTEM in the process). I make note of these in the dataset using the `field_detailed` variable. I also collapse these fields into similar groups (e.g., denoting electric engineering and aerospace engineering as "engineering") in the `field_broad`.

I want to caution, of course, that collapsing similar fields together is an inherently subjective process. Consequently, I urge users to consider both the detailed and broad versions of this variable when using the data.

District Level Variable Collection & Measurement

After identifying a list of candidates in Step 1, I then merge in district level variables pertaining to the **district's partisan lean** (known as "PVI;" via the non-partisan Cook Report), **percentage of the population that is college educated** (from the US Census), and the **percentage who voted for Donald Trump in the 2016 general presidential election** (via data scientist Stephen Wolf's publicly available data). Please refer to Table S1 for additional information about the distribution of these variables.

Typically, the common link between all three datasources of these datasources and my dataset is some combination of the candidate's state plus district information (e.g., MN01, MN-01, MN1). After arranging the state and district variable to match how they are presented in each of the aforementioned datasets, I (automatically) merge the relevant information into the list of candidates identified earlier. I perform this procedure for House districts only, as Senators represent entire states in Congress (and therefore do not have districts).

Note that PVI variable can, theoretically, range from -100 to 100 (although, as Table S1 demonstrates, the actual range is much shorter than that). PVI is a comparison of each district's two-party vote share – over the course of the two most recent election cycles (in this case 2012 and 2016) – to that of the national average. I code PVI such that positive values reflect more Democratic districts, and negative numbers reflect more Republican districts. A score of 0 indicates that the district voted in line with the national average.

Candidate Level Variable Collection & Measurement

In order to track whether candidates are **endorsed by or receive campaign contributions from 314 PAC**, I took the following two steps. First, I searched OpenSecrets' (run by the Center for Responsive Politics) campaign contribution list for 314 PAC in the 2018 cycle; which details every candidate who received money from the group. I then manually searched the candidate list generated in Step 1 for each of these individuals, and assigned them a code of 1 on the endorsement variable if they received any campaign cash from 314 PAC. Because some candidates might be officially endorsed without actually receiving any campaign donations, I then searched 314 PAC's website for its official endorsement list, and updated the codes produced in the previous step. In all, 314 PAC has endorsed 28 candidates – or 18% of all people with scientific backgrounds running for Congress.

I also incorporate information about each candidate's personal, academic, and political backgrounds into the dataset by manually coding each one on several dimensions. First, I determine candidates' **partisan affiliation** by searching for that information on their campaign websites. In the event that this information was not available on their site (e.g., the candidate dropped out of the race and deleted their campaign webpage), I instead gleaned this information from Ballotpedia (a website which tracks candidates and election results at all levels of government). I summarize this information in a variable called `party`, and create binary indicators of whether or not candidates are Democrats, Republicans, or a third party candidate. Note that the `party` variable includes information about what the *type* of party that third party candidates affiliate with.

I created measures of **candidates' performance in the election** by regularly assessing their status on Ballotpedia. Using a series of binary indicators, I coded candidates as either disqualified from the race (e.g., they failed to gain enough signatures to appear on the ballot); withdrawn (i.e., they announced their candidacy and perhaps even appeared on the ballot, but dropped out of the race); lost in a primary; or if they advanced to the general election.

Likewise, I recorded whether or not candidates had **previously run in state-wide or Congressional races** in past election cycles by referencing their electoral histories on Ballotpedia. I coded candidates into three groups: those who had never run for any office, those who had run for local office (e.g., mayor of a municipality, a state representative, etc.) but never for federal or state-wide office, and those who have run for federal or state-wide office in the past (e.g., governor, Congress, state attorney general, etc.).

I assessed candidates' **highest degree attained** by again searching their websites (or Ballotpedia, if the websites were taken down and not archived by a service like Internet Archive) and making a note of the highest degree listed. As I did in Step 1 with the "field" codes, I create two versions of this variable – one (`degree`) denoting the type of degree earned (e.g., MD, DMD) and another (`degree_broad`) collapsing them into three groups (Bachelor degrees, Master degrees, and Doctorates).

Finally, I determined candidates' `gender` by searching their campaign website and determining which gendered pronouns they use to refer to themselves (if written in the third person). If this information isn't readily available on the website, I consulted Ballotpedia to see which gendered pronouns they used to describe each candidate.

Supplemental Tables

Table S1. Variable List & Summary Statistics

Variable	Mean	SD	Min.	Max.	Valid N	Type
Candidate Name	-	-	-	-	-	Nominal
State	-	-	-	-	-	Nominal
District	-	-	-	-	-	Nominal[+]
State + District	-	-	-	-	-	Nominal
Field of Expertise (Specific)	-	-	-	-	-	Nominal
Field of Expertise (Broad)	-	-	-	-	-	Nominal
Bachelor Degree	0.251	0.435	0	1	191	Indicator
Master Degree	0.262	0.441	0	1	191 Indicator	
Doctorate (PhD/MD/JD)	0.487	0.501	0	1	191 Indicator	
Female	0.264	0.442	0	1	193	Indicator
Male	.736	0.442	0	1	193	Indicator
Advanced to General	–	–	0	1	157	Indicator
Lost Primary	–	–	0	1	157	Indicator
Withdrew/Disqualified	–	–	0	1	157	Indicator
Democrat	0.799	0.402	0	1	194	Indicator
Republican	0.149	0.357	0	1	194	Indicator
Third Party	0.052	0.222	0	1	194	Indicator
Endorsed by 314 PAC	0.139	0.347	0	1	194	Indicator
% Vote for Trump in District	48.885	12.355	5.4	76.7	183 *	Interval
% College Educated in District	33.384	10.623	9.121	62.429	183 *	Interval
District Ideology (PVI)	3.742	11.763	-43	28	182 *	Interval

Note. * Indicates that district-level variables are calculated for House candidates only (as Senators do not belong to districts). + Indicates that while this variable is theoretically nominal, it is listed in the dataset as a numeric value, encoded as a "string."