

Finding a Place in Political Data Science

Andrew Therriault, *Democratic National Committee*

In the summer of 2004, I decided to join my first political campaign. Though it was relatively late in the cycle, I found an organization which was mobilizing students on college campuses, so I packed up my car and headed for Cleveland. By day I registered voters and recruited student organizers, at night I crashed on a friend's couch, and in between I delivered pizza so that I would have gas money to get home afterward. I did little else but work for the time I was there, and once the voter registration deadline had passed, I left to spend most of October on Get-Out-The-Vote programs in New Hampshire and Pennsylvania.

That election taught me two things. First, I had the political bug, and wanted to find a way to keep doing it. And second, I never, ever wanted to carry a clipboard again. As someone who has never been mistaken for an extrovert, starting conversations with strangers in public or at their doors was painfully difficult. I had an idea of the job I really wanted, something in which I could combine my science, math, and computing skills to win campaigns. But I did not know what that job was at the time—in hindsight, because it did not really exist yet—so I decided to go to grad school instead. I sent in my applications in the fall of 2005, and started at NYU the following September.

Over the next five years, I warmed to academic life, and went on the poli-sci job market in 2011. It did not go as planned. By January I found myself in the second semester of a one-year post-doc, in a city I had no other reason to be in, and decided to look at getting into campaigns again. I received a job offer from a prominent polling firm, and was tempted to go to DC and take my place among the *Politico*-reading class, but instead I joined a political analytics start-up founded by two political scientists in California. Their firm had just landed a major client for 2012, a coalition of labor unions fighting an anti-union ballot measure, and I was brought on as the firm's first hire.

Having a diverse methods background proved quite valuable in that year's campaign. I wrote and analyzed surveys, designed message-testing experiments, guided the campaign in the use of microtargeting models, and assessed the impact of voter outreach programs. Though there was certainly plenty to learn along the way, I had focused my academic studies on areas that were directly relevant to this work—survey research, voter behavior, campaign strategy, and every methodological approach my program offered—and that training provided a great starting point for using data in a real political campaign.

As often happens in a start-up, I chose my own title for that first position: "Strategic Model Developer." After the

election, I took on a bigger role and became "Director of Research and Business Development," but neither of those titles quite fit what I actually did day-to-day. A more fitting title for this role was "Data Scientist," but I still was not really sure I was one of those until my next employer (an opinion research and strategy firm) printed it on my business cards. By the time the Democratic National Committee asked me to lead its data science team a year later, I had finally come to really understand what it meant to be a data scientist, in all the ambiguity that term entails.

Though the role of data scientist has only been around a few years, it is really more of a catch-all term for anyone who does complex things with large amounts of data to solve real-world problems. Many of the jobs that now fall under the heading of data scientist are not themselves new. Five years ago, some data scientists were programmers, some were data analysts, and others had more specialized titles. But the distinguishing characteristic of a data scientist comes from not being confined to a single position. Instead, a data scientist often bridges fields and brings a wide range of skills when tackling a problem, so a more specific title would be too narrow.

What does it mean to be a data scientist in politics? There is no single answer, but there are some common themes across roles. A classic definition of a data scientist (which, in a field this new, means a definition that has been around since 2012) is someone who is better at statistics than any software engineer, and better at software engineering than any statistician.¹ But there is also a third cornerstone to the qualifications of a data scientist—substantive knowledge—and in politics that is a subject one can easily spend their whole career on.

In the broadest sense, the role of a political data scientist is to make the work of our campaigns and organizations more efficient and effective. To make things more efficient, we typically employ some form of targeting: choosing the most persuadable voters to contact, finding the best TV channels to run ads on, or identifying the best potential donors to send fundraising mail, for example. And to make things more effective, we use testing and analysis to determine which messages, tactics, or strategies will have the greatest impact. An important corollary to all this, though, is that we do not ever accomplish anything directly: data scientists are not the ones reaching out to potential voters, creating ad messages, or collecting donations. So to be successful in our own work, we need to understand the work that others around us are doing, and find ways to contribute to *their* success. After all, the most ingenious model ever made would be worthless if nobody knew how to use its predictions.

In more specific terms, the methodologies we employ include predictive modeling, survey and field experiments, public opinion research, and observational analysis. Less glamorous but equally important, though, is the more fundamental work of collecting, organizing, and transforming large amounts of messy data. Part of this skillset involves the use of database systems and data processing tools (most notably SQL, Hadoop, and more recently Spark) to work with datasets that

Most people who work in political data have an undergrad's level of stats knowledge at best, while their more technical skills have been learned mainly on the job. By contrast, having spent the better part of a decade studying these topics at a deeper level, I was able to start at a high level and advance quickly from there. To be sure, I did not know everything I needed to know when I started—far from it! But I did have a solid base of technical skills to start from, and enough

We do not have the luxury of choosing to work on problems that have clean solutions, or the time to spend years refining answers—in this field, a dissertation is a great example of what not to do.

are far too large and complex for Stata or R.² But there is also a philosophical element to this work. Opportunities to practice ideal research design are incredibly rare, so making the most of whatever data you have available is key to being productive, and the best data scientists are those who are comfortable making compromises while consciously working to mitigate their downsides.

This is one aspect in which political data science differs greatly from academic political science: the relative weighting of pragmatism and perfectionism. And it is one that can often hinder academics who make the leap to politics, as it requires a retraining of the instincts learned through many years of graduate school. The things most of us work on every day involve a degree of compromise that would get us thrown out of an academic job talk, because the obsessive approach required to prepare for that event would never work in this context. We do not have the luxury of choosing to work on problems that have clean solutions, or the time to spend years refining our answers—in this field, a dissertation is a great example of what *not* to do. Instead, our mission is to find the best solution to a problem given the time and resources available. The value of our expertise is often in knowing the limitations of what we can do, then coming as close to that line as possible.

I have always enjoyed working on messy problems more than clean ones, and perhaps that is why I am a better data scientist than I was an academic. But my transformation from job market leftover to lead data scientist for a national political party was surprisingly fast, which speaks to the unique characteristics of this field. For one, the political workforce turns over at a much faster rate than almost any other, so it is possible to advance quickly with the right mix of hard work and good luck. And my timing was ideal: the kind of work I do was almost unheard of in 2008, but by 2012 it had become common to many campaigns (on our side, at least), and in 2016 its domain will expand even further. This is in part a function of knowledge and cultural development, but it is also a function of technological capacity, as our ability to work with data goes hand in hand with our ability to learn from it. With this growth in the scope and scale of political data science, there was an ever-growing need for people with the sort of skills that I had gained in my political science training.

knowledge of both research and politics to build on going forward.

While the field of political data science has grown and matured over the past few years—and I am sure there will be much tougher competition for my job by the time I move on—there will still be opportunities for technically-skilled political scientists for many years to come. Every election cycle brings in a new crop of talent, who often come with greater technical ability than the people hiring them. There will always be a need for those who can bridge the substantive and technical sides of politics, and among that set of individuals, quantitatively-trained political scientists are well-positioned to continue driving the campaign industry forward. Field and survey experiments, for instance, are becoming a standard part of many campaigns' messaging, voter contact, and fundraising strategies, and a political scientist's training could help to not only design those tests but also to determine what should be tested in the first place.

For students thinking of working in “applied politics,” the choices you make now—which classes to take, which methods to learn, which research areas to focus on—can have a big impact on both your immediate value and your long-term potential as a data scientist. From a substantive standpoint, a background in electoral politics and public opinion are essential. But just as important is a wide-ranging methodological training, particularly in statistical modeling, experimental research, and survey methods. Students would also be well-advised to look outside of the traditional political science curriculum to build up their computing and research skills:

- Conferences and meetings which bring together academics and practitioners, such as those hosted by the American Association for Public Opinion Research and its regional chapters, are a great opportunity to learn more about what applied research looks like in practice.
- For those in departments with limited methodological offerings, there are plenty of opportunities to learn through books and blog tutorials, free online courses from sites like Coursera and Udacity, and industry conferences like PyData and Spark Summit (both of which offer free video archives of past events).

-
- The field of machine learning should be high on any future data scientist's list, as this approach to modeling and analysis is quickly overtaking the more traditional models of statistics and econometrics in popularity.

recent grad student. Much like academia, success in this field requires a level of drive and passion that borders on pathological, because the tangible rewards are seldom in proportion to the amount of work put in.

To be sure, I did not know everything I needed to know when I started—far from it! But I did have a solid base of technical skills to start from, and enough knowledge of both research and politics to build on going forward.

Books like *Machine Learning* by Peter Flach and *Python Machine Learning* by Sebastian Raschka provide introductions to the field that should be very accessible to anyone with a semester or two of quantitative methods training.

More broadly, knowledge of general-purpose programming languages (especially Python), database principles and usage, and software development processes are all key to becoming a successful data scientist, so a few well-timed computer science courses could really give students a head start.

You should also be prepared for the many challenges you will face in making the transition from academia to politics. As mentioned above, the pragmatic approach that is inherent to political data science will be an inevitable culture shock, no matter how much you try to prepare. You will need to learn to work more quickly than you are comfortable with, and you will have to call things complete without a chance to polish them, and you will sometimes make mistakes because of this. You will also need to swallow pride on occasion. Though you may have “Dr.” in front of your name, when you start, there will be plenty of things you cannot do that your eighteen-year-old interns can. You will not last very long unless you are able to listen, take feedback, and ask for help when you need it.

As a final dose of reality, I should note that the political job market does bear some unfortunate similarity to the academic one. The number of jobs is limited, and no matter how great your qualifications, your chance of getting a good one often comes down to timing, connections, and outright luck. And if you do manage to find such a position, the long hours and modest compensation will seem familiar to any

Or to put it another way: political data science is like any other job in politics. Nobody gets into this industry for the money or quality of life, and not many stay in it for more than two or three campaign cycles. But on the plus side, the broader data science field is booming, and the training and experience you get in politics is among the best you could hope for. The pace, variety, and autonomy in a political data scientist's workload is far beyond that found in any other industry, and the people you will work with are some of the smartest you will ever meet. All of that ultimately makes you a much better all-around data scientist. So even if you do wind up working in the private sector after a few years, the experience is great preparation for whatever you choose to do next. And you do get a chance to change the world in the meantime—I would say that counts for something.

ACKNOWLEDGMENTS

The author would like to thank Jonathan Nagler, Josh Tucker, Pat Egan, John Geer, Mike Alvarez, Peter Foley, Elizabeth Sena, Anna Greenberg, Andrew Brown, Hilary Keller, and countless others for their support throughout the process that led to this article. ■

NOTES

1. Credit for this definition goes to Josh Wills. Other proposed definitions include “a statistician who lives in San Francisco,” “a data analyst who uses a Mac,” and “someone who is worse at statistics than any statistician, and worse at software engineering than any software engineer.” I prefer Wills' definition for obvious reasons.
2. A nationwide voter list, for example, contains roughly 200 million rows and dozens or hundreds of variables, making it several orders of magnitude greater than even the largest datasets you are likely to see in grad school.

Navigating the Profession: Sage Advice from the Pages of PS

This *first* virtual issue of *PS: Political Science & Politics* brings together some of the journal’s most highly cited articles about the profession from more than a dozen past issues. This practical material spans numerous topics that especially apply to young scholars. From “ideas and debates” of political science as a vocation, to the “nuts and bolts” of preparing a literature review and publishing as a graduate student, this virtual issue boasts something both for scholars just beginning their studies and for academics early in their careers. Other topics include:

- Restructuring the social sciences
- Navigating the challenges of fieldwork and research design
- Harnessing the potential in blogging and other outreach in the profession and the public
- Becoming a reviewer
- Writing more effectively
- And more!

APSA is excited to offer this collection of 24 *PS* articles in this virtual issue *Navigating the Profession*. The association looks forward to producing more virtual issues to highlight excellent research and make this research more accessible to members. Read the virtual issue today at journals.cambridge.org/PS/virtualissue/1.



“Mentoring the profession has been and should continue to be one of the most important missions of *PS*.”

**–Robert J-P. Hauck,
Editor, *PS*, 1987–2014**

**Questions? Comments? Let us know at
publications@apsanet.org.**