

Teaching Data Science in Political Science: Integrating Methods with Substantive Curriculum

Unislawa Williams, *Spelman College*

Robert Brown, *Spelman College*

Marilyn Davis, *Spelman College*

Tinaz Pavri, *Spelman College*

Fatemeh Shafiei, *Spelman College*

ABSTRACT


The importance of data science in society today is undeniable, and now is the time to prepare data science talent (National Academies of Sciences, Engineering, and Medicine 2018). Data science demands collaboration, but collaboration within political science departments has been weak in teaching data science. Bridging substantive and methods courses can critically aid in teaching data science because it facilitates this collaboration. Our innovation is to integrate data science into both substantive and methods courses through a dedicated data science course and modules on data science topics taught in substantive courses. This approach allows not only for more opportunities for teaching and practice of data science methods but also helps students to understand how social, economic, and political biases and incentives can affect their data.

Data science is booming as an emerging academic field. Although it has infused political science research and teaching, it has done so largely in the methods curriculum. However, substantive political science and its students offer a critical and undervalued angle on data science. Our quest to answer “[w]ho [g]ets [w]hat, [w]hen, [h]ow” has led to a wealth of academic knowledge on social, economic, and political incentives and biases (Lasswell 1936). Understanding how these political incentives and biases can affect the data science workflow at every stage may provide insight into some of our most pressing societal questions.

Building data science modules into substantive political science courses can improve data science teaching because it compels students to think critically about the data life cycle. From data acquisition to data processing, analysis, and visualization, various social, economic, and political biases and incentives affect data. This type of teaching informs student learning on data science and provides a potential grounding in data ethics. However, few political science departments integrate data science training in the substantive curriculum, concentrating instead on the methods-related curriculum.

Teaching data science comprehensively is increasingly critical for several reasons: because of the growing importance of data science, because of the expanding job market for data scientists, and because data questions are now substantively interesting to political science students. As more undergraduate students express substantive interests in questions related to data’s socioeconomic and political aspects, political science can provide training that not only meets that demand but also is relevant and complementary to computer science and other data science programs.

We believe that our department has successfully created a unique curriculum integrating our data science teaching across methods and relevant substantive courses. As a political science department at a liberal arts college focused on educating women of

Unislawa Williams  is associate professor of political science at Spelman College. She can be reached at uwilliams@spelman.edu.

Robert Brown is associate professor of political science at Spelman College. He can be reached at rbrown61@spelman.edu.

Marilyn Davis is associate professor of political science at Spelman College. She can be reached at mdavis1@spelman.edu.

Tinaz Pavri is professor of political science at Spelman College. She can be reached at tpavri@spelman.edu.

Fatemeh Shafiei is associate professor of political science at Spelman College. She can be reached at fshafiei@spelman.edu.

African descent, our curriculum offers classes dealing with social, economic, and political biases. Courses include “Racism and the Law” and “Black Women: Status, Achievement, Impact.” However, political science can be broadly relevant to data science teaching, as we illustrate in this article.

By including data science topics in substantively driven courses that do not focus on methodology, our students can balance

curriculum requirements listed on department websites (Williams et al. 2020). Only two schools appear to integrate their offerings into both methods and substantive courses; the remaining programs do not have such intradepartmental collaboration.

The majority of top departments offer “methods” without reference to data science; less than half offer a methods track or concentration; and only two offer data-science-specific concen-

Collaboration is particularly important when teaching about data because working with data often is collaborative.

technical training with discussions of social, political, and economic incentives and biases stemming from the growing influence of data across politics and society. Importantly, if our experience is a guide, discussing data ethics and the range of social, economic, and political biases and incentives affecting data may effectively recruit minority women into data science—a field with severe underrepresentation of women and minorities.

A limitation of this study is that we are in the early stages of implementation and do not have a comprehensive assessment. Nonetheless, some students have gone on to pursue data science careers, and many others are interested in following suit.

Studies have found that collaboration focused on building the same skillset across different courses improves student learning (Warfield-Brown and Pontuso 2004). For example, writing across the curriculum is an example of a successful initiative that integrates multiple faculty members, teaching a set of skills across a range of courses (Warfield-Brown and Pontuso 2004). Another example of successful interdisciplinary change is the internationalization of the curriculum (Barber 2007; Bromley and Walker 2008; Cassell 2007; Ishiyama and Breuning 2006; Lantis 2011; Martin 2007; Ward 2007).

Collaboration is particularly important when teaching about data because working with data often is collaborative. First, data often require collaborative solutions, such as providing centralized access (McDermott 2010). Second, the practice of research is increasingly collaborative because articles increasingly have coauthors (McDermott and Hatemi 2010). Third, collaboration is integral to the rising data science discipline, requiring the convening of experts with different skillsets. “Data science” is itself an umbrella term for various activities across disciplines, including statistics, computer science, data visualization, information technology, and subject-specific expertise.

The more that teaching reflects what practitioners do with data, the more students learn. When students apply data-analysis skills outside of the classroom, student outcomes improve (Rosen 2018). Outcomes also improve when students apply data analysis to their own learning (Loepp 2019). In contrast, traditional approaches to teaching methods have disengaged students (McBride 1994). Therefore, Loepp (2019) called for the “creation of a network of teacher–scholars interested in developing, sharing, and refining best practices related to data-based teaching.”

However, in political science, integrating data-science-relevant teaching across the political science curriculum remains a rarity. Table 1 presents the extent of data science and analytics offerings in the top 25 political science programs (rankings are according to *US News and World Report* 2017), based on course descriptions and

trations (table 2). Some schools have a data science offering that includes formal theory (not algorithmic game theory), which traditionally is not a data science subject because it does not directly include data. The inclusion of formal theory with data science makes the data science offering appear larger, but it may not represent a real integration across the political science curriculum. In fact, nearly twice as many programs offer formal theory courses in their methods/data science concentrations than an actual data science course.

Despite extensive expertise in methods, few departments have successfully adapted their substantive curriculum to teach about data science. Developing collaboration among faculty members can be challenging (Lake 2010). “Mutual gain is the common objective;

Table 1
Data Science Integration in Substantive Coursework

Top 25 Political Science Departments		
	No.	Percentage
All Programs	28	100%
Methods Courses	27	96%
Formal Theory in Methods/Data Science Concentration	12	43%
Any Data Science Course	7	25%
Multiple Data Science Courses	5	18%
Data Science in Substantive Courses	2	7%

Note: Due to ties in the rankings, the top 25 actually represent 28 departments.

Table 2
Data Science Prevalence in Undergraduate Political Science Programs

Top 25 Political Science Departments		
	No.	Percentage
All Programs	28	100%
Methods Courses	27	96%
Methods Track/Concentration	13	46%
Data Science–Specific Track/Concentration	2	7%

Note: Due to ties in the rankings, the top 25 actually represent 28 departments.

the only difficulty lies in achieving it” (Wildavsky 1986). Hence, the question is not whether faculty collaboration can improve student outcomes but instead how to actualize it.

INTEGRATED TEACHING

Until recently, our department was not teaching data science. However, as data science was “catching fire” in the business world, academia, and federal agencies, we decided that this cutting-edge

are required to create an original data visualization and publicly present their results. The focus on communication—which also aligns with the data science workflow—has had an impact on students because they see their research as more immediately valuable not only for their own development and skill building but also for their community.

Second, for interested faculty, CPI sponsored more extensive training and support to develop data science courses. As a result, we

The majority of top departments offer “methods” without reference to data science; less than half offer a methods track or concentration; and only two offer data-science-specific concentrations.

skill must be imparted to our students. The college’s winning proposal for the Career Pathways Initiative (CPI) funded that work and was built around the infusion of career knowledge into the curriculum. To develop our curriculum, integrating data science across substantive and methods courses, we took the following steps.

First, we held an internal workshop that was designed around three goals: (1) providing data science training to faculty; (2) helping faculty develop data science teaching skills; and (3) creating short one-week modules that could be inserted into existing courses the following year (see appendix 2). Faculty members were encouraged to create modules best suited to their courses. We brought two outside experts in data science and social sciences from NORC at the University of Chicago to lead the workshop. We also included two internal faculty members as points of contact to support faculty in developing their course materials. Nearly all political science faculty participated.

The short modules contributed to the courses and furthered the students’ data science skills without significantly departing from existing substantive material. Some of our greatest successes have resulted from integrating data science topics into our marquis courses (see appendix 1 for a discussion of how data science informed student learning in “Black Women: Status, Achievement, Impact”). However, more traditional substantive courses were equally valuable.

An example is “Introduction to Asian Studies.” Although it is the only required course for the Asian Studies minor, it also constitutes a political science elective. Due to the COVID-19 pandemic, the instructor added a data science module that compares the pandemic responses of three countries (i.e., Japan, China,

created a new data science course that bridges algorithmic thinking, programming, and analysis with topics including social justice and politics. The data science course integrates students’ substantive interests with data science skills. To broaden awareness of data science, we opened the course to all majors and started with sophomores. The course was successful and will be offered again.

Third, we updated our methods curriculum. Although we already had methods’ requirements, programming now was emphasized. Within the course, we also encouraged students to learn free, open-source programs that they can download to their computer. Building that skillset also was combined with discussions about the broader applicability of data science. To make the point more relevant to students, we drew on datasets that are of substantive interest to them, including issues of poverty and education. Students discussed how analyses and presentations of data can be compelling in wide-ranging professional contexts beyond academia. In fact, within the major, we have updated our student-learning outcome to reflect a greater focus on career.

Among those faculty members who did not initially participate in the workshop, we saw increased interest later in connecting their past pedagogy to the current effort. For instance, in the late 1990s, national government, international relations (IR), and comparative politics courses started incorporating data modules by adopting the use of MicroCase statistical packages (Le Roy 2004). The IR text was not updated; therefore, its use was discontinued. However, faculty remembered that these concrete examples helped students to better understand the use of data as a method of inquiry and engage in lively discussions.

We can recruit more diverse students and we can more effectively and distinctly equip them by incorporating substantive political science into data science training.

and India) across different variables such as numbers of cases, deaths, and recoveries; extent and nature of lockdowns; and most-used treatments. The students examine both numeric and graphic data and then write a paper that analyzes and compares these responses, drawing conclusions based on their findings.

In the required senior seminar course, students traditionally researched a topic and wrote a paper summarizing their results. We have since changed this model to focus students’ attention on communication and data visualization. Instead of a paper, they now

Faculty members found that incorporating data science topics into methods courses was relatively seamless because those courses already focused on data. In substantive courses, access issues were more pronounced. However, across all courses, incorporating data relevant to students—especially data related to social justice, social bias, poverty, and racism—enhanced learning.

Although we do not have extensive data on student outcomes, we have early results. As anecdotal evidence, we have succeeded in placing our students at major tech companies and in graduate data

science programs. Students who had considered switching to different departments have continued with the major. On the faculty side, there is a sustained interest in data science, and we continue to offer the modules within the courses. Perhaps most important, nearly all majors have been exposed to some information about data science. Generally, we believe this points in a positive direction. Developing awareness and growing interest are particularly valuable at Historically Black Colleges and Universities, given the underrepresentation of women and minorities in data science.

CONCLUSIONS

In summary, our innovation centered on intradepartmental collaboration to integrate data science with substantive courses. Our experience indicates that short one-week modules make the integration feasible without significantly altering existing materials. Integrating data science teaching across the curriculum mirrors other successful pedagogical initiatives. Without this type of integration, it is difficult to argue that we effectively teach data science, given that technical and domain-specific collaboration is the cornerstone of the field. Teaching about data in silos is antithetical to data science practice.

Nonetheless, in many political science departments, data science and/or methods training remains siloed. Statistics departments faced similar criticisms. Some statisticians who viewed data science as rebranding were skeptical about the movement lasting. Although these statisticians were initially incredulous that their activities were now “bright, new, and carried out by...upstarts and strangers,” many now feel that the “train is leaving the station” (Donoho 2017). If statisticians consider staying siloed to be costly, then political science methodologists also may need to be mindful of the potential cost.

Because substantive political science examines social, economic, and political data, it can add to data science by teaching foundations in data ethics. Our accumulated knowledge of social, political, and economic incentives and biases provides an opportunity to lend substance and rigor to data science ethics questions. This is an opportunity for political science, which often borrows from but has a weaker record of lending to other fields (Box-Steffensmeier and Sokhey 2007).

Data ethics is not only an expertise needed in data science; it also attracts our students to data science careers. Recruiting minority women is vital, given the underrepresentation of women and minorities in data science. We can recruit more diverse students and we can more effectively and distinctly equip them by incorporating substantive political science into data science training.

ACKNOWLEDGMENTS

This research is supported by the United Negro College Fund Career Pathways Initiative Faculty Research Grant. We thank Shelby Lewis for her research assistance.

DATA AVAILABILITY STATEMENT

Replication materials are available on Harvard Dataverse at <https://doi.org/10.7910/DVN/JYIZ73>.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1049096520001687>. ■

REFERENCES

- Barber, Benjamin R. 2007. “Internationalizing the Undergraduate Curriculum: Opening Commentary.” *PS: Political Science & Politics* 40 (1): 105.
- Box-Steffensmeier, Janet M., and Anand E. Sokhey. 2007. “A Dynamic Labor Market: How Political Science Is Opening up to Methodologists, and How Methodologists Are Opening up Political Science.” *PS: Political Science & Politics* 40 (1): 125–27.
- Bromley, Pam, and Gloria Walker. 2008. “2008 APSA Teaching and Learning Track Summaries—Track Eleven: Internationalizing the Curriculum I.” *PS: Political Science & Politics* 41 (3): 625–26.
- Cassell, Mark. 2007. “Internationalizing Public Policy Courses.” *PS: Political Science & Politics* 40 (1): 106–8.
- Donoho, David. 2017. “50 Years of Data Science.” *Journal of Computational and Graphical Statistics* 26 (4): 745–66.
- Ishiyama, John, and Marijke Breuning. 2006. “How International Are Undergraduate Political Science Programs at Liberal Arts and Sciences Colleges and Universities in the Midwest?” *PS: Political Science & Politics* 39 (2): 327–33.
- Lake, David. 2010. “Who’s on First? Listing Authors by Relative Contribution Trumps the Alphabet.” *PS: Political Science & Politics* 43 (1): 43–47.
- Lantis, Jeffrey S. 2011. “2011 APSA Teaching and Learning Conference Track Summaries—Track: Internationalizing the Curriculum II: Study Abroad and Institutional Strategies.” *PS: Political Science & Politics* 44 (3): 662–63.
- Lasswell, Harold. 1936. *Politics: Who Gets What, When, How*. New York: Whittlesey House.
- Le Roy, Michael K. 2004. *Comparative Politics: An Introduction Using MicroCase Explorit*. Belmont, CA: Wadsworth Group/Thompson Learning.
- Loepp, Eric. 2019. “Data-Based Teaching: An Introduction and Call for Collaboration.” *PS: Political Science & Politics* 52 (3): 1–6.
- Martin, Pamela L. 2007. “Global Videoconferencing as a Tool for Internationalizing Our Classrooms.” *PS: Political Science & Politics* 40 (1): 116–17.
- McBride, Allen. 1994. “Teaching Research Methods Using Appropriate Technology.” *PS: Political Science & Politics* 27 (3): 553–57.
- McDermott, Rose. 2010. “Data Collection and Collaboration: Editor’s Introduction.” *PS: Political Science & Politics* 43 (1): 15–16.
- McDermott, Rose, and Peter Hatemi. 2010. “Emerging Models of Collaboration in Political Science.” *PS: Political Science & Politics* 43 (1): 49–58.
- National Academies of Sciences, Engineering, and Medicine. 2018. *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: National Academies Press.
- Rosen, Amanda. 2018. “The Best Breakfast in Town: A Comprehensive Research Methods Project.” *PS: Political Science & Politics* 51 (1): 173–77.
- US News and World Report*. 2017. “Best Political Science Schools.” Available at www.usnews.com/best-graduate-schools/top-humanities-schools/political-science-rankings (accessed July 15, 2019).
- Ward, Deborah E. 2007. “Internationalizing the American Politics Curriculum.” *PS: Political Science & Politics* 40 (1): 110–12.
- Warfield-Brown, Rosalind, and James Pontuso. 2004. “Clear Writing Means Clear Thinking.” *PS: Political Science & Politics* 37 (3): 455–58.
- Wildavsky, Aaron. 1986. “On Collaboration.” *PS: Political Science & Politics* 19 (2): 237–48.
- Williams, Unislawa, Robert Brown, Marilyn Davis, Tinaz Pavri, and Fatemeh Shafiei. 2020. “Replication Data for: Teaching Data Science in Political Science.” Harvard Dataverse. doi: 10.7910/DVN/JYIZ73.