Integrating the Use of Statistical Software into Undergraduate Political Methodology Courses[*]

August 9, 2021

Forthcoming at *PS: Political Science & Politics*

David S. Brown
Katherine V. Bryant
Andrew Q. Philips

Keywords: undergraduate teaching; pedagogy; statistical software; research methods

David S. Brown is the Divisional Dean and Professor, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333. David.S.Brown@colorado.edu.

Katherine V. Bryant is an Assistant Professor, Department of Political Science, Westmont College, 955 La Paz Road, Santa Barbara, CA 93108. kabryant@westmont.edu.

Andrew Q. Philips is an Assistant Professor, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333. andrew.philips@colorado.edu.

[*] We wish to thank our team of excellent research assistants: Samuel Beck, Yucong Li, Julie Mcnees, Frida Muedsam and Hannah Strassburger. We also acknowledge financial support from the CU Department of Political Science's STUDIO Undergraduate Research Lab.

It is widely recognized that statistical, programming, data science and analytic skills give students a strategic advantage on the job market and in their future careers. The American Statistical Association reports that by 2021, around 70% of business leaders will prefer candidates with data science skills.[1] This demand is expected to keep growing, with projections of 28% growth over the last five years.[2] Jobs that require coding skills pay on average $22,000 more per year than those that do not require them, and nearly half of all jobs paying over $58,000 require at least some level of coding.[3] In acknowledgement of this trend, higher education has added degrees, programs, and courses in data science, and emphasized a greater need for data literacy among their students.[4] Students also are recognizing the utility of these skills, since they are now often requirements for jobs beyond the business or physical science world, including education, healthcare, nonprofits, and even the wine industry.[5]

Yet for our discipline, teaching undergraduate political methodology courses remains a demanding task; faculty must instruct students to be writers, readers, and creators of political knowledge.[6] While such courses have existed for decades, the growing emphasis on data analysis and data literacy places a greater demand on research method courses to deliver these skills. Further complicating matters, there remains little common discussion among political scientists regarding the use of data software, skills, or approaches in these courses. As a discipline, while we recognize the growing need for greater data literacy among undergraduate students, we appear to have placed little emphasis on this trend in our own pedagogical discussions.

The criticism that research methodology lacks a strong pedagogical founding is not new. Previous studies recognize a dearth of discussion on the subject, despite acknowledging that methods courses are often some of the most academically challenging for students (Howard and Brady 2015). For instance, Wagner, Garner and Kawulich (2011, p. 2) state that research methods instructors, “appear not to cover pedagogical questions at all”. This is not to say that discussions of pedagogy in social science and political science methodology are nonexistent (c.f., Hubbell 1994; Payne and Williams 2011; Nind, Kilburn and Luff 2015). However, the fact remains that there is very little guidance for undergraduate political methodology instructors. As more students recognize the benefits of such courses, a deeper understanding of best practices in pedagogical approaches is urgently needed.

We fill this void by addressing two questions. First, *how* is statistical software incorporated into undergraduate political methodology courses? Second, *how* are these courses structured and taught? We answer these questions through two distinct approaches. First, a team of research assistants collected and coded course information from 93 syllabi from 78 separate institutions (some had two or more syllabi included) of undergraduate political methodology courses drawn from a cross-section of institutions, according to their rank in the *US News and World Report*. Of the 93 syllabi, 25 were from the South, 24 from the North, 9 from the East, and 35 from the West. Second, we rely on an anonymous survey conducted in January 2021---advertised on both the PolMeth email listserv as well as the “Political Scientists” private group page on Facebook---of instructors who have taught a political methodology course at the undergraduate level. We received over 140 responses, from a mix of R1 (53% of respondents), R2 (11%), and R3 institutions (9%), as well as liberal arts (26%) and community colleges (0.7%). Using these two original data sources, we provide a broad overview of the "state of the discipline" regarding not only how statistical software is incorporated into political methodology courses, but also how they are presented to students. We conclude with some advice for instructors of these courses.

Software

Do instructors use software in political methodology courses, and, if so, which types? We begin our analysis by asking whether methods courses are required for a political science major. There is clear evidence that political methodology classes are valued by departments and institutions; over 73% of survey respondents and 59% of syllabi coded indicated that their course is a major requirement. Our syllabi coding also suggest interesting variation depending on institution type and location. Figure 1 indicates that regional universities are far more likely to require a methods course, followed by liberal arts schools and national universities. Additionally, 62% of private institutions required a methods course, whereas public institutions lagged slightly behind at only 57%. Finally, institutions in the middle of the country are more likely to require these courses (64%) compared to their coastal counterparts (54%).

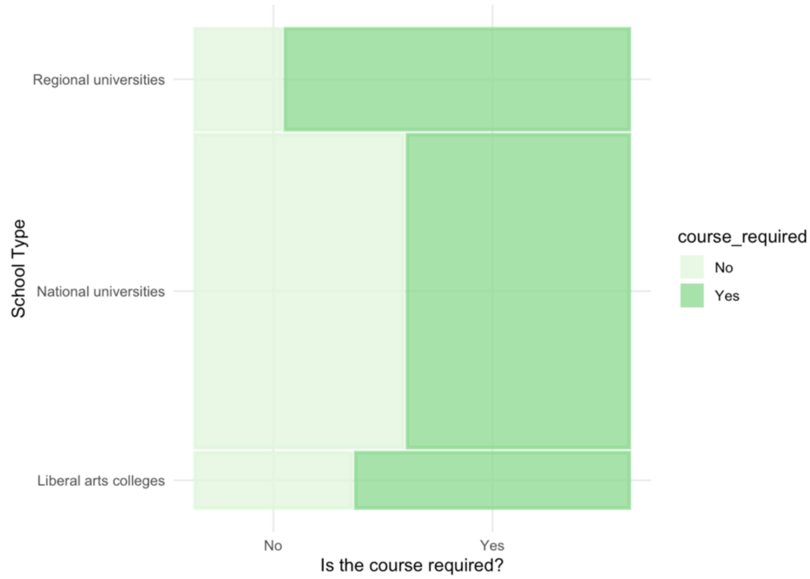

Figure 1

Table 1 summarizes in more detail the various types of software used. Our syllabi coding suggests that over 80% of courses use at least some statistical software. R and/or RStudio are by far most commonly used, with 29% of syllabi and 60% of survey respondents indicating they are used in their courses. One explanation for R's popularity may be cost; a substantial number of survey respondents remarked that this was one of the largest advantages of using R, as well as an active online community to provide help when needed (some using Excel also noted that it is often free to students through their university). Next most popular are Excel, Stata, then SPSS (according to the survey)---or SPSS, Stata and Excel (as per the syllabi)---with usage rates ranging between 9 and 30% of classrooms. In contrast, Python is seldom used by instructors in political methodology courses.

Table 1: Software patterns

	Syllabi (N = 93)	**Survey** (N = 133)		
	Usage (%)	Usage (%)	"Which software package would you guess is used most often in courses like yours?" (%)	% of users that would recommend their software to a colleague
Excel	9	26	14	77
Python	0	5	2	86
R/RStudio	29	60	30	89
SPSS	16	17	27	59
Stata	13	25	28	82
Other	5	8	0	80

Our survey also asked respondents which software they think is most often used in political methodology courses. As shown in Table 1, instructors were fairly accurate, although they appear to overestimate the use of SPSS in such courses while underestimating the use of R and Excel (at least according to the survey). We also asked respondents if they would recommend their chosen software to a colleague. Substantial majorities of respondents would recommend Excel, Python, R and Stata, although about 60 percent would recommend the use of SPSS. One common sentiment used by survey respondents to justify why a particular type of software was chosen is its potential for being useful not just in the classroom or academia, but in industry. This was most common in R, although others argued the same for Stata, SPSS, and Excel.

Despite its popularity, one of the oft-heard disadvantages of R (or Python) is the steep learning curve. Survey respondents called it "impenetrable", "intimidating", and "extremely daunting". Yet all software---excepting perhaps Excel---is likely alien to anyone who has not familiarized themselves with the interface. As one respondent colorfully put it, “my students hate Stata...but I think they would hate any statistical software program I assign.” In our experience, devoting the first few weeks of a course to orient students with downloading, installing, and showing basic functionalities such as setting a working directory, or how to open and save files, can work wonders in helping students become comfortable with the interface. Textbooks with embedded or included code can also help students.

We also found that nearly a third of survey respondents used multiple software in a single course. As one instructor remarked, this strategy, “allows students to use whatever they have around, guided by the same scientific knowledge.” Given that students may already know one type of statistical software, such an approach might be more flexible, and more realistic to the real world; as researchers, we certainly do not limit ourselves to a single software. Downsides include the need for the instructor to know all software competently, and issues with allocating class time to discussing multiple ways of doing the same thing across different software. Related to this, it is clear that familiarity is highly desirable. As one respondent remarked on their software of choice, “it's the one I'm most familiar with, so it's easy for me to problem solve when they get stuck”. Switching to another statistical program is likely challenging for instructors, given the time and effort involved in learning a new system.

Structure and Topics

How are undergraduate methods courses structured and taught? To address this question we examined the topics covered, course level and size, as well as the types of assessment used by instructors. As shown in Table 2, topics on research design, statistics, and data science are included in a sizeable majority of courses. However, substantial within-category variation often exists. For instance, according to our survey analysis 94% of courses discuss descriptive statistics, while only 18% cover analysis of variance (ANOVA/ANCOVA) techniques. Almost 60% of courses discuss how to manage datasets, and 88% discuss graphics, yet only 29% of methods courses explicitly include programming. Our syllabi analysis yields even lower percentages in a variety of categories, including general hypothesis testing, probability theory, and regression. However, the vast majority of syllabi cover elements of both research design and causal inference, and over half include discussions of epistemology.

Table 2: Included topics

Topic	***Survey*** *(N = 137)* ***Percentage***	***Syllabi*** *(N = 93)* ***Percentage***
Research Design/Causal Inference		
Research design	82	83
Causal inference	NA	76
Philosophy of science/epistemology	NA	59
Statistics		
Regression (bivariate/multivariate)	83	60
Regression assumptions	NA	16
Dichotomous dependent variables	31	18
Descriptive statistics	94	NA
Hypothesis tests	89	53
ANOVA	18	6
Probability theory/Central Limit Theorem	62	42
Data Science		
Data management	58	NA
Graphics	88	30
Programming	29	NA

Note: Not all categories were covered in both the survey and syllabi codings (marked as NA).

Our survey findings regarding course level and size are shown in Figure 2. A plurality of courses are intermediate undergraduate courses, although they are also taught at other levels. Courses also differed substantially in class size, ranging from a seminar-style eight-person course up to a 700 person course (with a median of 29). Large courses tend to occur in beginning undergraduate level courses, with a median of 37 students versus 20 for an advanced undergraduate course.

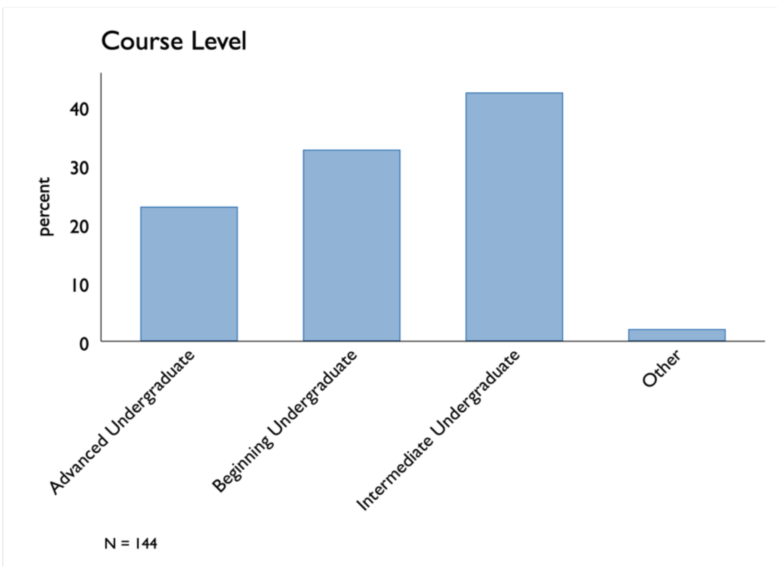

Figure 2

Finally, we examined various types of assessment used by instructors in these courses. Table 3 shows that the vast majority of courses include individual homework, labs or problem sets, as well as some form of exam. Research papers are also a common type of assessment. Less common are group assignments or problem sets, as well as presentations. While it is clear that individual homework or labs might be useful, our experiences suggest that group assignments---we prefer groups of three---are a good way to allow students to learn, often by helping one another. Weekly or bi-weekly group assignments that build on the previous week increase the knowledge base while still providing some familiarity with the software, given prior assignments. For instance, one week might involve bivariate visualizations (e.g., scatterplots) while the next week's assignment requires adding an additional variable (e.g., scatterplots with colors, shapes or sizes representing the additional dimension). We also believe that presentations are excellent ways for students to practice organizing and articulating their data findings, although these types of opportunities are likely to be limited for larger class sizes.

Table 3: Assessment types

Assessment Type	*Syllabi (N = 46)* *Percentage*	*Survey (N = 137)* *Percentage*
Individual problem sets/labs/homework	84	77
Group problem sets	18	23
Papers	45	66
Presentations	16	36
Exams (online/in-person)	81	69

Conclusion: Success with software

Just as it has greatly changed political science, statistical software has altered the classroom, in particular for methodology courses. We think this is for the better, despite the fact that incorporating these tools into courses is a challenging undertaking. The results of our analyses indicate that there is wide variation in the structure of these courses and the software that they use. We would also be remiss to not mention the substantial amount of variation---in terms of institutional support and resources---for these courses. While our survey indicated that many respondents had access to teaching assistants and lab sessions, the majority of syllabi we analyzed did not have those same resources. Given the high demands placed on faculty in teaching these courses, further investigation of this difference is warranted.

Drawn both from our own experience as well as from our colleagues, there are a few overarching pieces of advice or best practices to share regarding statistical software. First, there is a tension between learning a statistical program and learning statistics; they are not the same. Some students will relish learning the ins and outs of programming, others will abhor it. Still others do not want to be learning statistics in the first place---we often hear that they “didn't sign up for a math class!”---which makes the enterprise especially fraught with peril. Being clear about the goals and structure of the class early in the semester helps tremendously in this regard. Also be aware that different software packages or languages are best learned in different ways; some require higher upfront costs than others. The more open and upfront the instructor is about the advantages and disadvantages of a program or package, and why it was chosen, can help dampen both reluctance and frustration. We have often pursued the strategy of simply typing “jobs in R” into the search bar on the first day of class to assuage students that these *are* marketable skills worth investing in.

Second, while some students can read a mathematical expression and know exactly what it means, for a required class in political science this is probably the exception. Some students rely on simple, plain language that leads the student step-by-step through the process.

Others are more visually-oriented and can quickly learn commands, graphical techniques, and generate estimates only through example. Given such wide variation in how students learn, providing different ways of discovering how to use the software is important. In the same vein, it helps to have multiple resources available. Packages and languages that are well-supported on the internet, and actively maintained, are helpful. Also be very explicit that one of the course skills you would like them to develop is the ability to find their own answers; they do not all have to come from the same textbook or website. Showing students how to search for help online can be an asset for them and ease the burden on instructors.

Third, we are advocates of the flipped- or semi-flipped---i.e., active rather than passive---teaching modalities. Experimental evidence suggests that these promote learning, though at the cost of reduced student satisfaction (Deslauriers et al. 2019). While this requires condensing lectures into fewer course days (or moving some of them fully online as pre-recorded lectures), devoting course time for individuals or groups to work on assignments using software goes a long way to diffusing frustrations with small coding issues that can easily be answered by instructors. As an alternative---especially in smaller courses---research projects involving the entire class may be the best way to engage with students (Winn 1995).

Finally, students' ability to learn the software will depend on observation ("the code is not working!"), generating hypotheses ("probably just a typo"), and testing that hypothesis ("change the spelling and rerun the command"). This is an over-simplified example of how students learn and practice the scientific method as they work with software. While not the same as taking a spoonful of sugar with the medicine, if done correctly---e.g., making it explicit in learning goals---students will find that they sharpen their skills at problem-solving.

Despite the heterogeneity in undergraduate methods courses, a central prevailing theme is their growing demand. Given this, political science must lean into supporting instructors as they confront the challenges and complexities that accompany teaching political methodology courses. Greater pedagogical dialogue, sharing of best practices, and creating a repository of online resources are just some of the many steps forward in fostering a more comprehensive and supportive approach to these integral courses.

References

Brown, David S, Katherine V Bryant and Andrew Q Philips. 2021. "Replication Data for: Integrating the Use of Statistical Software into Undergraduate Political Methodology Courses." Harvard Dataverse. doi:10.7910/DVN/FENBA2.

Deslauriers, Louis, Logan S McCarty, Kelly Miller, Kristina Callaghan and Greg Kestin. 2019. "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom." *Proceedings of the National Academy of Sciences* 116(39): 19251–19257.

Howard, Cosmo and Michelle Brady. 2015. "Teaching social research methods after the critical turn: challenges and benefits of a constructivist pedagogy." *International Journal of Social Research Methodology* 18(5): 511–525.

Hubbell, Larry. 1994. "Teaching Research Methods: An Experiential and Heterodoxical Approach." *PS: Political Science & Politics* 27(1): 60–64.

Nind, Melanie, Daniel Kilburn and Rebekah Luff. 2015. "The teaching and learning of social research methods: developments in pedagogical knowledge." *International Journal of Social Research Methodology* 18(5): 1–9.

Payne, Geoff and Malcolm Williams. 2011. *Teaching quantitative methods: Getting the basics right*. London: Sage Publications.

Wagner, Claire, Mark Garner and Barbara Kawulich. 2011. "The state of the art of teaching research methods in the social sciences: Towards a pedagogical culture." *Studies in Higher Education* 36(1): 75–88.

Winn, Sandra. 1995. "Learning by doing: Teaching research methods through student participation in a commissioned research project." *Studies in Higher Education* 20(2): 203–214.

[1] Source: The American Statistical Association, https://www.amstat.org/ASA/News/New-Report-Highlights-Growing-Demand-for-Data-Science-Analytics-Talent.aspx. Accessed 1/20/21.

[2] Source: U.S. News and World Report, https://www.usnews.com/education/best-colleges/articles/why-more-colleges-are-offering-data-science-programs. Accessed 01/17/21.

[3] Source: https://www.fastcompany.com/3060883/why-coding-is-the-job-skill-of-the-future-for-everyone. Accessed 1/21/21.

[4] Source: Inside Higher Ed, https://www.insidehighered.com/digital-learning/article/2017/03/15/data-analytics-programs-taking-colleges. Accessed 01/17/21.

[5] Source: U.S. News and World Report, https://www.usnews.com/education/best-colleges/articles/why-more-colleges-are-offering-data-science-programs. Accessed 01/17/2021.

[6] For brevity, we refer to any research methods, statistics, or data science course as a “political methodology” or “methods” course.