Teaching for a Data-Driven Future: Intentionally Building Foundational Computing Skills

Amy L. Johnson and Rebecca D. Gleit, Stanford University

Article forthcoming in *Teaching Sociology*

APPENDIX: DATA & METHODS

Content Analysis

To assess how software is currently taught in sociology departments, we conducted a content analysis of syllabi from undergraduate and graduate introductory quantitative methods courses at California colleges and universities.

For our content analysis of undergraduate sociology statistics training, we began by identifying all not-for-profit institutions of higher education (2-year, public 4-year, and private 4-year) located between San Francisco and San Jose, California (inclusive) that offer undergraduate degrees in Sociology (N=16). We chose this targeted area for two reasons. First, proximity to the authors' university allowed us to utilize connections to local colleges and universities and thus collect syllabi and course documents more easily. Second, because of Silicon Valley's longstanding emphasis on computing, schools in this area may be particularly likely to utilize statistical software with the goal of preparing students for jobs in the tech industry. We understand our sample as a conservative test of the claim that computing is underemphasized in sociology curricula: if courses in this area are not integrating computing explicitly, courses in other parts of the country are likely not either.

At each institution, we examined the requirements for the undergraduate sociology program and documented each instance of a statistics or research methods course. We located course descriptions and included any course that either explicitly mentioned software or included

phrases that implied the use of applied quantitative methods. 15 of the 16 institutions required at least one such course for their undergraduate sociology majors. In total, our sampling frame consisted of 32 different courses that fill undergraduate sociology majors' statistics requirement; some of these courses are housed in sociology departments, some are cross-listed with other social science departments such as psychology, and others are introductory courses offered in the math or statistics departments. We sought out syllabi from recent iterations of each of these 32 courses through online searches and direct outreach to instructors who have taught or will teach the course in either the 2020-2021 or 2019-2020 academic years. For community colleges, we also attempted to locate course outlines; many of these were publicly available online. In total, we received at least one syllabus or outline for 29 of the 32 total courses in the undergraduate sampling universe. For some courses, we obtained multiple syllabi and/or course outlines. While course outlines provided useful guidelines, syllabi for these courses often strayed from the outlines. Therefore, we relied on syllabi over the outlines for our analysis in cases where they differed. For the undergraduate analysis, we examined a total of 36 unique syllabi and 16 unique course outlines across nine community colleges, two public 4-year institutions, and three private 4-year institutions.

We systematically examined each syllabus and outline to assess several measures of software integration. First, are students required to utilize statistical software (broadly defined) in this course? If so, what statistical software is used? Second, does the syllabus or course outline explicitly include a statement of the learning goals? And if so, are there any learning goals explicitly related to computing and/or the use of software? And lastly, what textbook, if any, is used by the course? Syllabi were coded by hand and information was tracked and analyzed in a Google spreadsheet.

For our analysis of graduate statistics training, we identified all institutions that offer PhDs in Sociology in the state of California (N=11). We followed the same process as for undergraduate institutions, identifying each introductory required statistics or quantitative methodology course, locating syllabi through direct outreach to recent instructors, and analyzing the syllabi in the same manner. All 11 of the institutions require at least one statistics course for their Sociology PhD students. Many programs require multiple terms of statistics training; we generally included only the first course in statistics training in our analyses here. For two institutions with three-quarter methods sequences, we included the first two courses because both were described as introductory. In total, our sampling frame for graduate courses consisted of 13 required introductory statistics courses for Sociology PhD programs in California. We obtained syllabi for 12 of the 13 courses in our graduate sampling universe and analyzed these syllabi in the same way as the undergraduate syllabi in order to assess software integration.

Workshop

We ran our Introduction to Stata workshop twice, once for the introductory undergraduate data analysis course at Stanford University and once for the equivalent course at the graduate student level. For undergraduates, the workshop was offered to all students enrolled in the spring 2019 undergraduate-level introduction to data analysis course, which is required for sociology majors and counts towards a number of other social science majors. This course covers basic statistical analysis, such as hypothesis testing and linear regression, using Stata. The workshop was optional for students and taught by both authors in the second week of the term. One of the authors was the teaching assistant for this course and the other had been a teaching assistant for the course in a previous year. Students were recruited through an email from the teaching

assistant. Approximately two-thirds of the students enrolled in the course attended the optional workshop.

For graduate students, the workshop was offered to students enrolled in the fall 2019 introductory sociological methodology course that is required as part of the university's sociology PhD program. Like the undergraduate course, this class covers basic statistical analyses using Stata. Half of the students were incoming first-year sociology PhD students and half were graduate students from other social science disciplines. The two-hour workshop was part of an optional week-long "statistics bootcamp" taught by both authors prior to the start of the term. While the bootcamp was optional, all students enrolled in the fall term course opted to participate.

All students attending the workshop remained engaged throughout the first hour, though participation in the self-directed portion varied by iteration. While all participants in the iteration with graduate students stayed for the entirety of the self-directed portion, only a fraction of participants in the iteration with undergraduates did.

Participating students were surveyed three times: before the workshop, in the days following the workshop, and at the end of the 10-week term. Students who attended the workshop were emailed the day after the workshop and at the end of the term with the follow-up surveys. Completion of each survey and workshop for each iteration is detailed in Table B1. For the April 2019 iteration, 46 students took the pre-survey, 34 attended the workshop, and 21 completed the post-survey. In September 2019, nine students took the pre-survey, 10 attended the workshop, and 10 completed the post-survey. 12 students from the April 2019 iteration and seven from the September 2019 iteration completed our survey after the completion of the 10-week course. We do not highlight trends in the end-of-term survey in part due to attrition, but

also since any changes in self-rated abilities likely reflect an entire term spent utilizing Stata for course assignments, rather than the workshop itself. Surveys were anonymous, and results were linked across time points using anonymous identifiers known only to the students. Because linking identifiers were known only to the students, we did not (nor did we have the ability to) link the surveys to students' course outcomes.

[Table B1 about here]

Since students were recruited through courses, we were careful to ensure that they did not feel coerced to participate in our study. We received formal approval for research with students through our university's Institutional Review Board (IRB). We also explained clearly to students, through email and verbally at the beginning of the workshop, that their participation in the study was entirely optional, and that it would not affect their grade in the course in any way. As we describe above, student identities were kept anonymous during the research process, and this research study was entirely separate from any course assessments.

Table B1. Survey Completion by Workshop Iteration (Number of Observations)

Iteration	**Student level**	**Pre-workshop survey**	**Workshop attendance**	**Post-workshop survey**	**End of term survey**
April 2019	Undergraduate students	46	34	21	12
September 2019	Graduate students	9	10	10	7