

Teaching Computational Social Science Skills to Psychology Students: An Undergraduate Research Lab Case Study

Brian A. Eiler, *Davidson College*
Patrick C. Doyle, Rosemary L. Al-Kire, Heidi A. Wayment, *Northern Arizona University*

Abstract

Data science methods increasingly are utilized to analyze theoretically derived psychological research questions. This article provides a case study of a student-focused research experience that introduced basic data science skills and their utility for psychological research, providing practical learning experiences for students interested in learning computational social science skills. Skills included programming; acquiring, visualizing, and managing data; performing specialized analyses; and building knowledge about open-science practices. Using examples from their teaching experiences, the authors describe how these skills can be incorporated into an active and engaging student learning experience that culminates in computational social science projects and presentations.

Keywords: *big data, computational social science, network analysis, social media data, text analysis, undergraduate research*

doi: 10.18833/spur/4/1/5

Exposure to research is a key aspect of social science education (Kierniesky 2005). By conducting research, students have an opportunity to think about behavior scientifically (Brewer et al. 1993). In the psychological sciences, this educational tradition is especially important for undergraduate students and can be provided through research opportunities (Wayment and Dickson 2008). Ideally, student research-related learning experiences should reflect current trends in the discipline, both theoretical and methodological. Student mentoring also provides opportunities for more experienced students, graduate students, postdoctoral students, and faculty (Linn et al. 2015). Yet

the time faculty need to devote to incorporating new ideas and methodologies into their existing research programs often competes with teaching and service activities.

This article explores data science skills, a fast-growing area of interest in the social and psychological sciences. Referred to as computational social science (CSS), these methods offer several distinct advantages for students and faculty. First, large data sets afford a number of benefits, including relatively quick data collection (as with publicly available data sets and open-science framework projects); IRB-exempt status; high-powered analyses; and topics that may be strongly tied to student and faculty interests and related to current events or social media (Chen and Wocjik 2016). Second, the skills required to obtain, manage, and analyze such data sets allow opportunities for faculty and students to cultivate interdisciplinary skills related to computer science, data science, and novel analytics (such as R), which are rapidly becoming a part of the psychological research landscape. Analysis of large social media data sets requires the acquisition and development of these new computational skills.

This article seeks to provide an introduction and practical information for faculty to implement the teaching of basic CSS skills in their own research labs or research methods courses. First, basic data science methods and relevant learning objectives are described. Next, a case study of an actual undergraduate research lab that incorporated CSS learning goals, outcomes, and objectives is provided. Additionally, some sample research activities used to introduce and develop these skills during a one-year lab experience for undergraduates are described.

Computational Data Science Skills in Psychological Research

The unprecedented availability of “big data” is very relevant for the social sciences (Foster et al. 2017). In fact, new graduate programs in computational social science are on the rise, and existing programs have begun to value these skills in both incoming students and faculty searches. CSS projects generally leverage large data sets of “found” data from social media, digital, and other electronic databases and employ computational and visualization techniques to illuminate patterns and make inferences from these data, as they relate to theory. Although big data has been criticized, these computational analyses complement traditional quantitative social science methods (that also have their limitations) to confirm key findings and theory (Shah, Capella, and Neuman 2015). Moreover, by using this approach, researchers gain access to large amounts of electronic data that are of interest to psychologists, such as measures of discrete behavior, personal and social relationship networks, and expressed attitudes. This information can be used to complement other types of data social scientists employ to test hypotheses about a range of social phenomena and social influence processes. CSS skills also are important for business analytics, marketing, and survey research (Callegaro and Yang 2018) and help students and faculty increase their employment opportunities.

Big data—often described in terms of five Vs: *volume*, *velocity*, *variety*, *veracity*, and *value*—is typically free, can be collected quickly, and includes content with real-world applicability (Kitchin and McArdle 2016). There are many types of big data that are useful in an undergraduate research setting. Some data, like social media data, must be scraped and formatted, but there are many other publicly available data sets that are available in readable file formats (complete with codebooks) that do not require data cleaning. There are large data sets of American health practices, such as those available from the National Center for Health Statistics (2020), as well as many social science–related measures collected from adults from 31 countries in Europe (European Social Survey 2020). Some data sets require some knowledge of probabilistic samples (as in complex survey analysis). Notwithstanding the challenges associated with big data (including issues of generalizability, ethics, and theory), key data science skills are required to harness the wealth of newly available data for the purposes of social science research. Generally speaking, these skills include data acquisition (collecting, storing, and processing processes), data management (including collaboration and open-science practices), and data analysis. More specifically, key data science skills include programming skills (such as R, Python, and SQL); record linkage and database management; web scraping and the use of application programming interfaces (APIs); specialized forms of analysis (basic machine learning, text analysis, and analysis of network and geospatial data); and

the ability to think through and articulate issues regarding inference, data privacy, and ethics (Foster et al. 2017).

Insights from Teaching Experiences

This article provides some examples of curriculum related to the five basic CSS skill sets that were implemented during a pilot one-year course at a medium-size state university. These skill sets included programming, data acquisition and management (data wrangling), data visualization, specialized analyses, and open-science practices. Below, each skill set is defined, and a few relevant examples helpful for curriculum design are provided, with the goal of helping students progress to an initial level of CSS competency.

Programming

CSS requires some knowledge of computer science, specifically an understanding of programming. Although there are several programming languages and interfaces, R appears to be the most highly adopted among psychological scientists, perhaps because of its enhanced capabilities for statistical analyses, data manipulation, and visualization. It also seems to have the shortest learning curve and maximum capability for CSS. Furthermore, there exist a plethora of resources due to the open-source nature of R. See *Cookbook for R* (2020) for an introduction to R and free beginner tutorials that also illustrate best practices (e.g., annotated code) to enhance replicability and transparency. Swirl is an effective, free, and interactive online tool for learning R basics, available within the publicly available R console (Swirl 2020). Faculty researchers more familiar with statistical packages such as SPSS or SAS will find resources such as Muenchen’s book (2011) valuable in that it leverages existing knowledge of SPSS and SAS syntax to learn how to code in R. One practical exercise for SPSS and SAS users is to rerun previous analyses using R. If users generalize code so that it can be repurposed across disparate data sets, utility over time is increased because analyses can be highly automated. It is important to acknowledge that all statistical software packages have advantages and disadvantages, and ultimately use depends on individual preferences. However, the flexibility, control, and enhanced capabilities of R may be appealing to those who wish to highlight quantitative skills. One practical advantage of using R for students (especially those who are pursuing quantitative concentrations or minors), is that it is free and open source, minimizing issues of access and inclusion.

Data Acquisition and Management

CSS requires data, often in large quantities, and an important part of data acquisition is the management of these data, also called “data wrangling” (i.e., retrieving, formatting, cleaning, and readying for analysis). Although there are many types of secondary data sets available for use in

a CSS lab, undergraduates in this case were very interested in analyzing social media data. Considerable time was spent working with them to identify appropriate sources of primarily social media data for a research question and teaching them how to retrieve it and how to clean and format the data for analysis. Prior to this, students were taught about the ethics of using online data. For example, students learned of constraints on gathering data from a website and that a robots.txt file could tell them whether data scraping (automated data collection by reading HTML or extracting information directly) was permitted. Many social media sites provide an API that specifies how to access website content via software. For instance, the R package *rtweet* (Kearney 2018) allows users to build software in R that interacts with Twitter's API to download various types of information in line with the guidelines established by Twitter. Although the specifics differ for each site, there are free R packages for many social media sites. These are complete with documentation and example code and are available from the Comprehensive R Archive Network (2020), which also is embedded in R Studio. Once data were downloaded, students imported text data into Excel files and cleaned the data, removing symbols and hashtags as well as doing other types of data cleaning. These processes take a significant amount of time (up to 80 percent of the total data analysis time).

Data Visualization

One key feature of a CSS approach is the emphasis on data visualization. Often in the social sciences data visualization does not enter the process until the end of the project, serving as a way to summarize, rather than explore, patterns in the data. In CSS, data visualization occurs early. Understanding trends in hundreds of thousands of datapoints is greatly facilitated by visualization. Although Microsoft Excel can be useful for simple visualizations and graphs, especially when data are extracted into a .csv file, R is well known for its excellent graphics. Many data science capabilities reside in the library of tidyverse, and a myriad of learning materials can be found at that website (Tidyverse 2020). Readers are referred to a freely available online book (Wickham and Grolemund 2017) that introduces readers to data science using R and includes excellent sections on visualization.

Specialized Analyses

Students were introduced to two specialized analyses that work well with Twitter data. The first was text analysis using a well-established analysis software called Linguistic Inquiry and Word Count (LIWC 2020; Pennebaker et al. 2015). For a nominal fee, a researcher can utilize LIWC to garner insight into individuals' thoughts, feelings, motivations, and drives. These markers can then be

used to test hypotheses regarding differential language use as a function of theoretically relevant independent variables. Also available is a free natural language processing software (a machine-learning approach that identifies relevant indicators of sentiment in language) called the Sentiment Analysis and Cognition Engine (SEANCE; Crossley, Kyle, and McNamara 2017). This program provides 254 core indices and 20 component indices of sentiment, supplies a negation feature (i.e., it can differentiate between "happy" and "not happy"), and has high validity due to the number of domain-independent dictionaries it employs. If a researcher is interested in signals of emotionality, valence, and extremity in evaluative contexts, a recently developed tool, the Evaluative Lexicon, can be used to quantify language in this way for free (Evaluative Lexicon 2020; Rocklage and Fazio 2015; Rocklage, Rucker, and Nordgren 2018). Finally, latent semantic analysis can be useful for identifying thematic content in a body of text. For an introduction to this technique, relevant theory, and annotated example R code, a highly digestible paper by Gefen and colleagues (2017) may be useful. Finally, the Meaning Extraction Helper (Boyd 2018), especially in combination with machine-learning techniques such as latent Dirichlet allocation and traditional data reduction techniques such as principal components analysis, provides accessible and robust CSS techniques for topic modeling (for tutorials and R scripts, see Boyd 2020).

Students also were introduced to network analysis, which has a long history in the social sciences. This technique primarily has been used to map the relationships between entities or individuals as a function of their social relationships (e.g., Borgatti et al. 2009). Network analysis is useful for visually depicting a network of Twitter users who follow one another, a network of individuals who share the same content online, or connections between those who belong to similar online forums. An additional use of network analysis involves representing a correlation matrix as a network and using network metrics to identify variables that are highly influential within a network (e.g., Eiler et al. 2019; Montazeri et al. 2019). When combined with simulation techniques and time series methods by the more advanced student, networks also can function as explanatory models across a range of network instantiations (for relevant publications and source code, see Fried 2020). There are exceptional resources available regarding social network analysis (see Butts 2008; Hanneman and Riddle 2005). Additionally, there are minimally priced point-and-click programs for network analysis (UCINET; Borgatti, Everett, and Freeman 1999) and visualization (NetDraw; Borgatti 2002). R's *statnet* package for analytics and visualization is increasingly used; it includes more advanced network analysis such as QAP, ERGM, and SIENNA (Handcock et al. 2003).

Open-Science Practices

Another advantage of integrating data science methods into the research process is the normative inclusion of open-science practices. The field of psychology has been recently plagued by ethical concerns and questions of reproducibility (Nosek et al. 2015). Psychological scientists have outlined a set of standards to improve the psychology research field that are being rapidly adopted by academic journals (more than 1,000 publications to date). These standards include (1) data transparency, code sharing, and transparent design and analysis; (2) study and analysis preregistration; (3) replication; and (4) more stringent citation standards for data, code, and materials (Blaine 2019; Nosek et al. 2015). Platforms such as the Open Science Framework (2020), AsPredicted (Wharton Credibility Lab 2020), and GitHub (2020) are repositories that allow researchers to share and organize files, code, and workflow as well as preregister studies. These tools also let researchers credit authors for nontraditional forms of science such as the development of source code and analytic strategies that do not show up as easily in the current academic peer review system. These platforms are intuitive for students to use and provide templates for users who may be less familiar with data sharing and open-science processes. In addition, many websites have large global communities that help users navigate and troubleshoot issues with everything from coding across multiple statistics languages (Stack Overflow 2020), to utilizing R (R-bloggers 2020; Nabble 2020), to implementing open-science best practices (Center for Open Science 2020).

One-Year Research Experience in Computational Social Psychology

During the 2017–2018 academic year, for the reasons articulated earlier, data science methods were introduced to a research lab, largely inspired by curiosity following participation in a number of online courses in data science. The CSS lab was structured as a two-hour weekly seminar. Eight students made a one-year commitment to the CSS lab experience. Participants included the lab supervisor (a professor in the psychological sciences department), who designed and oversaw implementation of all aspects of the course; a postdoctoral fellow with experience in CSS theory and analytics; two advanced master's-level students with interest in CSS and peer mentoring opportunities; two first-year master's-level students with intersecting interests in CSS and social psychology; and four undergraduate students with no psychological research experience. Time was devoted to instruction, discussion, and practice. Additionally, significant one-on-one mentorship outside of lab time was provided by the lab supervisor, postdoctoral fellow, and advanced graduate students.

Lab Structure and Student Learning Objectives

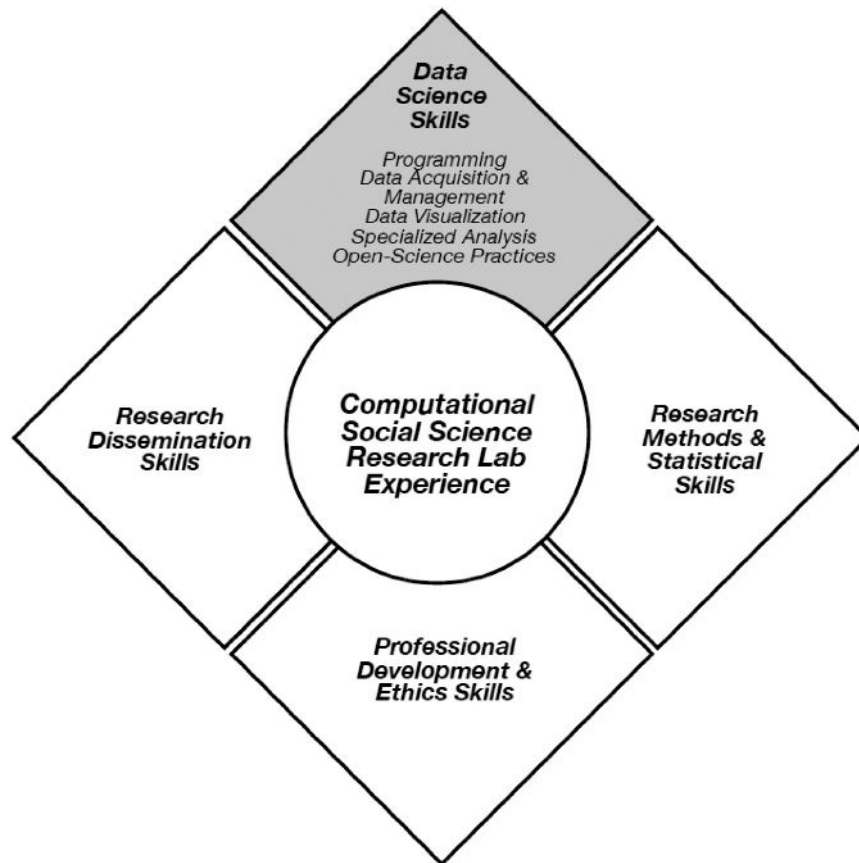
The one-year research course had learning goals, learning outcomes, and activities designed to meet the learning outcomes built into its structure. Some students took the lab for course credit, whereas others simply participated on a voluntary basis to gain research experience. At the beginning of the year, it was agreed that the CSS skills learned in the lab would be used to conduct research investigations that would culminate in two professional development opportunities: group presentations at a regional conference symposium and poster presentations at the university's undergraduate research symposium.

In alignment with local departmental learning objectives, there were four learning goals that centered on data science skill acquisition and three areas of expertise cultivated through previous undergraduate courses. As depicted in Figure 1, the primary learning goal was the acquisition of *data science skills* as described in this article. The remaining three goals served to utilize and strengthen existing research-related skills: *research methods and statistical skills* (the ability to use psychological theory as a foundation to formulate justifiable hypotheses and to design, conduct, statistically test, and interpret sound research projects); *research dissemination skills* (demonstrated competence in written, visual, and oral presentation skills); and *career professional development and ethics skills* (demonstration of socially responsible, ethical, and professional behavior in all aspects of the research process). Each learning goal was divided into specific learning outcomes onto which course activities were mapped and scheduled over the course of the academic year. The full syllabus, with further description of learning goals, is available.

Activities and Exercises

One enriching aspect of research is collaboration. Further, learning, sharing, and documenting programming skills (for data acquisition, management, analysis, and visualization) form an important set of CSS skills, as previously described. The CSS lab provided an opportunity for students to participate in an enriching nontraditional collaborative research experience. Their participation deviated from typical undergraduate research activities, in which students are confined to running participants' data or inputting data and are not involved in designing, analyzing, or disseminating project results. To facilitate this experience, a collaborative, experiential learning atmosphere was created. In contrast to a lecture-based learning approach, a range of educational strategies was employed, including active learning, self-directed learning, peer mentoring, guest lecturing, in-lab presentations, and professional presentations. In-class presentations included introductory materials for all topics and techniques. Out-of-class assign-

FIGURE 1. Learning Goals



ments were used to practice and refine skills, with the goal of conducting two to three small research projects during the one-year course. Students had opportunities to work on both individual and group projects. Regardless of the topic or project, students worked with each other outside of lab meetings to accomplish the work and relied on the guidance and leadership of more advanced lab members as well as information they could gather from online resources (for additional resources, see Eiler 2020). Twitter was selected as a source of social media data due to its applicability to many research interests, textual nature, and amenability to network analysis. The course materials and activities were structured to provide students with ample experience in the scientific process, from design to dissemination.

Briefly, there were eight major course requirements. These included (in chronological order): (1) and (2) a Twitter-related research article review and presentation (to learn relevant terminology and refresh research methods skills); (3) and (4) two Twitter research exercises (for which students downloaded Twitter data with R; managed Excel data; analyzed data using LIWC, SENACE, or Evaluative Lexicon; conducted simple data analyses and visualization; and made a presentation); (5) a major research project (for semester- and year-end presentations); (6) a network analysis (for which students collected, utilized, and drew

networks using UCINET and NetDraw); (7) a tutorial project (for which students wrote a brief “how to” on a specific skill learned in class and posted this to a lab-based website that students also shared with conference attendees); and (8) a final conference presentation (Eiler 2020).

The lab course was successful, and learning goals were met. Students provided positive feedback regarding their experiences and successfully presented their CSS-based research at local and regional conferences. One project assessed the frequency with which different online companies used language appealing to the millennial generation. Another examined differential language use as a function of celebrity fandom on Twitter, and another examined friendship networks in a community church. Graduate student presentations included an event-based analysis of tweets related to the 2017 Las Vegas shooting, an analysis of responses to tweets from Planned Parenthood, and a linguistic analysis of weight loss posts on Reddit. Beyond the products that directly resulted from work done during the academic year, several students continued to work on projects that have since been published (e.g., text analysis and network analysis were used to examine experiences of sexual violence in a sports context and identified power and trust dynamics as important contributors to sustained and systemic sexual assault in elite sports; Eiler et al.

2019). Additionally, three of the four graduate students utilized skills learned in the lab for their thesis work (all of whom have submitted their work to peer-reviewed journals at the time of writing). Students also provided feedback demonstrating the value of learning CSS skills:

“Working on our own project and honing novel skills . . . made it very enjoyable and different from other labs . . . not only was I able to increase my confidence and proficiency with using R . . . but presenting . . . was also great for my presentation skills.”

“. . . seeing that I was capable of so much more than I thought. Also, creating a project from start to finish that was all my, and my peer’s, doing.”

“I feel more comfortable and confident that I could successfully apply these techniques to other projects . . . the skills we learned are very marketable for a PhD program or industry.”

Beginning Computational Social Psychology Research

The paragraphs below describe the conditions that facilitate the ability to run a CSS lab. Because such resources may not be universally available in all educational settings, other options are presented. This work conducted with students was completed at an R2 institution in a department offering both undergraduate and master’s degrees. The existing curriculum structure allowed undergraduate students to participate in two consecutive semesters of research experience. Furthermore, students already had some background, which included an introductory statistics course. The inclusion of more advanced students was critically important because they were able to model behaviors that facilitated learning and had initial experience with R. The assistance provided by the department’s first ever postdoctoral fellow, who also was interested in mentorship experiences, was invaluable. Finally, the senior faculty member’s organization of the CSS research lab experience was structurally facilitated by the lab being counted as a course in her teaching load. Institutional and departmental support is important; documenting the program’s positive impact on students is an important rationale for such support.

For those faculty interested in incorporating CSS skills in their work with undergraduate- and master’s-level students, there are some challenges to consider. First, learning new skills takes time, patience, and resources, and faculty must be committed to and comfortable with learning alongside students. Students must be self-directed learners, willing to learn from their mistakes, and willing to help their peers. As described in the literature on active learning and engagement, pedagogical best practices can facilitate this directive. For those faculty who may have little or

no access to graduate students or postdoctoral students to assist in the teaching and mentoring of CSS skills, there are other ways to include more experienced students. As an example, there may be students with data science skills (usually in mathematics, computer science, or other natural science departments) who would be interested in applying and teaching these skills in the context of psychology and could be invited to be part of a CSS lab or teaching curriculum. Additionally, departmental momentum surrounding student preparation for postbaccalaureate opportunities may be leveraged to modify curriculum. In one example, a tenure-track assistant professor at a small, highly selective liberal arts institution has implemented and extended the current project into an upper-level interdisciplinary seminar: the course counts toward a psychology major, an interdisciplinary science major, and data science and applied math minors. This strategy requires additional resources that do not exist at every institution. Faculty will need to determine how to best modify the strategies presented here to accommodate the resources available to them.

There also are other types of resources available to faculty members or student researchers interested in CSS skills. There are several good books on these topics. For instance, Foster and associates (2017) have published a comprehensive selection of chapters with practical guidance on how to apply data science to real-world research, with applied problems from computer science, statistics, and social science. This resource explains how specific skills can be used to identify and capture data, how to apply data science models and methods to that data, and how to recognize and respond to data errors and limitations. A 2016 special issue of *Psychological Methods* focused on data science methods. Here, Kosinski and colleagues (2016) describe opportunities for using big data in the prediction of real-world outcomes and provide a tutorial about applying CSS to text. In another article, Kern and associates (2016) discuss methodological challenges to gaining insights from language posted on social media sites. Chen and Wojcik (2016) provide a practical guide to using big data in psychological research. Landers and colleagues (2016) describe Python-based approaches to data scraping. An excellent book on advanced topics related to computational and dynamic approaches in social psychology has been written by Vallacher, Read, and Nowak (2017). Second, there are excellent online courses, including a suite of data science courses offered by SAGE Publishing (SAGE Campus 2020). The courses are designed specifically for social scientists with a focus on practical skills that can be applied directly to relevant research topics in the social sciences. The popular online educational platform Coursera offers hundreds of R- and Python-related courses at beginning, intermediate, and advanced levels, such as *The R Programming Environment*; *Data Science: Foundations Using R*; *Python for Everybody*; and *Applied Data Science with Python*. R and Python also have website resources

(e.g., RStudio Support 2020), and there are many publicly available teaching resources online such as Twitter-based groups (#rstats). Finally, many institutions subscribe to additional learning platforms (e.g., LinkedIn Learning, DataCamp), which exponentially increase the available learning resources for faculty and students alike.

Finally, faculty may have concerns about personal development of these skills above and beyond existing workloads and service responsibilities. Although there is no one way to acquire a working knowledge of R and recruit students from within or outside the department to create a shared-learning experience, some ideas include (1) teaming up with a faculty member in another department to teach an introductory interdisciplinary course in computational social science or conduct a shared lab experience, (2) seeking out modest resources (such as professional development funds or special funding at the institutional level) to take data science courses or workshops, (3) sitting in on courses taught within the institution to learn basic R skills, or (4) enrolling in mini-workshops or talks offered through one's institution or at professional conferences (e.g., the Society for Personality and Social Psychology offered a big data workshop in 2019). A certain amount of flexibility and patience is helpful for incorporating CSS skills into an undergraduate research experience, either lab-based or in the classroom. Consistent with an open-science framework, CSS tools are updated and upgraded frequently on shared code repositories (e.g., GitHub), adding to the need for patience and a motivation to learn. However, experience has shown that undergraduate students adapt quickly to this learning environment and especially enjoy being able to work alongside faculty and develop stronger bonds that help them persist in STEM fields.

Discussion

To capitalize on the growing interest in and utility of learning and applying computational methods to psychological research, an innovative one-year research experience for undergraduate- and master's-level students was implemented at a medium-size university in the southwestern United States. Over the course of two semesters, five data science skills were taught and practiced, resulting in professional presentations. Information is provided for faculty who may be interested in implementing this type of experience in their own labs or departmental curriculum.

The implications of the work are threefold. First, the incorporation of CSS skills provides students with a glimpse of skills believed to be vital in the twenty-first century. These skills are not always offered in the existing curriculum (although R is increasingly added to undergraduate statistics courses), and students who partake in these activities may have a competitive advantage following graduation. Second, by incorporating CSS skills into undergraduate

research experiences, students remain engaged in active learning because they can design research projects that are not only relevant to their interests but that can be completed within a semester (from conceptualization to presentation). In this way these students engage in pedagogical best practices for mastery. Third, by their very nature, CSS skills (whether being learned, used, or shared) require an environment that supports open-science principles and practices such as data sharing, code sharing, and collaboration—practices that are in line with active learning and accessibility principles.

This approach adds a proverbial tool to the academic psychologist's teaching toolbox and, in so doing, helps maximize the number of psychology students who participate in research experiences. These activities also increase opportunities to work with more advanced students. Only 67 percent of institutions offering master's degrees have opportunities for students to work with faculty, and only 15 percent of these students work with graduate students (Perlman and McCann 2005). Given that these types of experiences are beneficial (McConnell, Albert, and Marton 2008; Wayment and Dickson 2008), the incorporation of computational methods also may indirectly assist faculty at institutions with larger teaching loads. For these faculty, supervising student research experiences may facilitate a sustained, low-cost research program. Furthermore, at small institutions that do not have large participant pools, these projects can supplement projects that require larger samples (i.e., more high-powered studies) and minimize risk by amplifying productivity. Indeed, integrating many students into various projects at differing stages of the research process may help faculty build and sustain a research program by allowing faculty to spend more time on activities appropriate to their experience levels (e.g., research question generation, research design, or complex analytics; Forrest, Stastny, and Bruns 2008). This process also may benefit early-career faculty who have yet to develop internal collaboration networks or have an established lab by off-loading tasks that can be delegated appropriately. Finally, mentorship opportunities are not limited to faculty. Many questions from undergraduates may be adequately answered by second-year master's program students, and anecdotally often these interactions result in dynamic changes in understanding (van Geert 1998). These mentorship opportunities are not limited to undergraduate-graduate student interaction but may be beneficial in any situation in which students are minimally separated in understanding; these interactions scale because they leverage the zone of proximal development.

Conclusion

Although challenging to implement, learning of CSS skills provides myriad benefits to faculty and to undergraduates who incorporate these skills in research experiences. Engaging in CSS research requires a commitment from

faculty to be open to new ideas and a willingness to engage with a community of similar-minded researchers. It is a challenge but also an opportunity—to reach out to others on campus as well as be available to serve as a resource for others. The authors believe that any exposure to CSS skills will benefit undergraduates during their education and following graduation. Students live with, interact with, and create forms of big data each time they browse the web or interact with their friends on social media. Leveraging these everyday experiences to teach them skills associated with the analysis of big data benefits students in disproportionate ways. This article describes the authors' own experiences and offers resources and ideas about how others might begin. The hope is that other faculty will be inspired to expose themselves and their students to CSS.

Acknowledgments

The authors would like to thank the student participants in the 2017–2018 Computational Social Psychology Lab for their dedication and interest in this emerging area of psychological science.

References

- Blaine, Bruce Evan. 2019. "Toward Greater Reproducibility of Undergraduate Behavioral Science Research." *Scholarship and Practice of Undergraduate Research* 3(1): 3–10. doi: 10.18833/spur/3/1/2
- Borgatti, Stephen P. 2002. *NetDraw Software for Network Visualization*. Lexington, KY: Analytic Technologies.
- Borgatti, Stephen P., Martin G. Everett, and Linton C. Freeman. 1999. *UCINET for Windows: Software Package for Social Network Analysis*. Natick, MA: Analytic Technologies. doi: 10.1007/978-1-4939-7131-2_316
- Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323: 892–895. doi: 10.1126/science.1165821
- Boyd, Ryan L. 2019. "Meaning Extraction Helper." Accessed October 12, 2020. <https://www.ryanboyd.io/software/meh>
- Boyd, Ryan L. 2020. "Home." Accessed October 12, 2020. <https://meh.ryanb.cc>
- Brewer, Charles L., J. Roy Hopkins, Gregory A. Kimble, Margaret W. Matlin, Lee I. McCann, Ogretta V. McNeil, Barbara F. Nodine, et al. 1993. "Curriculum." In *Handbook for Enhancing Undergraduate Education in Psychology*, ed. Thomas V. McGovern, 161–182. Washington, DC: American Psychological Association. doi: 10.1037/10126-006
- Butts, Carter T. 2008. "Social Network Analysis: A Methodological Introduction." *Asian Journal of Social Psychology* 11: 13–41. doi: 10.1111/j.1467-839X.2007.00241.x
- Callegaro, Mario, and Yongwei Yang. 2018. "The Role of Surveys in the Era of 'Big Data.'" In *The Palgrave Handbook of Survey Research*, ed. David Vannette and Jon Krosnick, 175–192. Cham, Switzerland: Palgrave Macmillan. doi: 10.1007/978-3-319-54395-6_23
- Center for Open Science. 2020. "Home." Accessed October 12, 2020. <https://www.cos.io>
- Chen, Eric E., and Sean P. Wojcik. 2016. "A Practical Guide to Big Data Research in Psychology." *Psychological Methods* 21: 458–474. doi: 10.1037/met0000111
- The Comprehensive R Archive Network. 2020. "Home." Accessed October 12, 2020. <https://cran.r-project.org>
- Cookbook for R. 2020. "Home." Accessed October 12, 2020. <https://www.cookbook-r.com>
- Crossley, Scott A., Kristopher Kyle, and Danielle S. McNamara. 2017. "Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social-Order Analysis." *Behavior Research Methods* 49: 803–821. doi: 10.3758/s13428-016-0743-z
- Eiler, Brian. 2020. "Resources." Accessed October 12, 2020. <https://brianeiler.com/resources>
- Eiler, Brian A., Rosemary L. Al-Kire, Patrick C. Doyle, and Heidi A. Wayment. 2019. "Power and Trust Dynamics of Sexual Violence: A Textual Analysis of Nassar Victim Impact Statements and #MeToo Disclosures on Twitter." *Journal of Clinical Sport Psychology* 13: 290–310. doi: 10.1123/jcsp.2018-0056
- European Social Survey. 2020. "Home." Accessed October 12, 2020. <http://www.europeansocialsurvey.org>
- Evaluative Lexicon. 2020. "Home." Accessed October 12, 2020. <http://www.evaluativelexicon.com>
- Forrest, Krista D., Bradley J. Stastny, and Jennifer A. Bruns. 2008. "Supervising Undergraduate Students in a Faculty Driven Research Program." In *Developing, Promoting, and Sustaining the Undergraduate Research Experience in Psychology*, ed. Richard L. Miller, Robert F. Rycek, Emily Balcetis, Steve T. Barney, Bernard C. Beins, Susan R. Burns, Roy Smith, et al., 224–229. Washington, DC: Society for the Teaching of Psychology, Division 2, American Psychological Association.
- Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane (Eds.). 2017. *Big Data and Social Science: A Practical Guide to Methods and Tools*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Boca Raton, FL: Taylor and Francis.
- Fried, Eiko. 2020. "Publications." Accessed October 12, 2020. <https://eiko-fried.com/publications>
- Gefen, David, James E. Endicott, Jorge E. Fresneda, Jacob Miller, and Kai R. Larsen. 2017. "A Guide to Text Analysis with Latent Semantic Analysis in R with Annotated Code: Studying Online Reviews and the Stack Exchange Community." *Communications of the Association for Information Systems* 41: 450–496. doi: 10.17705/1CAIS.04121
- GitHub. 2020. "Home." Accessed October 12, 2020. <https://github.com>
- Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2003. "*Statnet*: Software Tools for the Statistical Modeling of Network Data." Accessed August 1, 2018. <http://statnetproject.org>

- Hanneman, Robert A., and Mark Riddle. 2005. *Introduction to Social Network Methods*. Riverside: University of California, Riverside. <http://faculty.ucr.edu/~hanneman/nettext>
- Kearney, Michael W. 2018. "rtweet: Collecting Twitter Data. R Package Version 0.6.7." Accessed August 1, 2018. <https://cran.r-project.org/package=rtweet>
- Kern, Margaret L., Gregory Park, Johannes C. Eichstaedt, H. Andrew Schwartz, Maarten Sap, Laura K. Smith, and Lyle H. Ungar. 2016. "Gaining Insights from Social Media Language: Methodologies and Challenges." *Psychological Methods* 21: 507–525. doi: 10.1037/met0000091
- Kierniesky, Nicholas C. 2005. "Undergraduate Research in Small Psychology Departments: Two Decades Later." *Teaching of Psychology* 32: 84–90. doi: 10.1207/s15328023top3202_1
- Kitchin, Rob, and Gavin McArdle. 2016. "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets." *Big Data & Society* 3(1): 1–10. doi: 10.1177/2053951716631130
- Kosinski, Michal, Yilum Wang, Himabindu Lakkaraju, and Jure Leskovec. 2016. "Mining Big Data to Extract Patterns and Predict Real-Life Outcomes." *Psychological Methods* 21: 493–506. doi: 10.1037/met0000105
- Landers, Richard N., Robert C. Brusso, Katelyn J. Cavanaugh, and Andrew B. Collmus. 2016. "A Primer on Theory-Driven Web Scraping: Automatic Extraction of Big Data from the Internet for Use in Psychological Research." *Psychological Methods* 21: 475–492. doi: 10.1037/met0000081
- Linn, Marcia C., Erin Palmer, Anne Baranger, Elizabeth Gerard, and Elisa Stone. 2015. "Undergraduate Research Experiences: Impacts and Opportunities." *Science* 347: 1261757. doi: 10.1126/science.1261757
- LIWC. 2020. "Home." Accessed October 12, 2020. <http://liwc.wpengin.com>
- McConnell, William, Roger G. Albert, and John P. Marton. 2008. "Involving College Students in Social Science Research." *Teaching & Learning Journal* 2(1): 1–8. [http://www.kpu.ca/sites/default/files/Teaching and Learning/TD.2.1_McConnell_etal_Students_in_Social_Science_Research.pdf](http://www.kpu.ca/sites/default/files/Teaching_and_Learning/TD.2.1_McConnell_etal_Students_in_Social_Science_Research.pdf)
- Montazeri, Farhad, Annelies de Bildt, Vera Dekker, and George M. Anderson. 2019. "Network Analysis of Anxiety in the Autism Realm." *Journal of Autism and Developmental Disorders* 49: 2219–2230. doi: 10.1007/s10803-018-3474-4
- Muenchen, Robert A. 2011. *R for SAS and SPSS Users*. 2nd ed. New York: Springer Verlag.
- Nabble. 2020. "Home." Accessed October 12, 2020. <https://www.Nabble.com>
- National Center for Health Statistics. 2020. "Home." Centers for Disease Control and Prevention. <https://www.cdc.gov/nchs/nhanes.htm>
- Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. "Promoting an Open Research Culture." *Science* 348: 1422–1425. doi: 10.1126/science.aab2374
- Open Science Framework. 2020. "Home." Center for Open Science. <https://osf.io>
- Pennebaker, James W., Roger J. Booth, Ryan L. Boyd, and Martha E. Francis. 2015. *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.
- Perlman, Baron, and Lee I. McCann. 2005. "Undergraduate Research Experiences in Psychology: A National Study of Courses and Curricula." *Teaching of Psychology* 32: 5–14. doi: 10.1207/s15328023top3201_2
- R-bloggers. 2020. "Home." Accessed October 12, 2020. <https://www.r-bloggers.com>
- Rocklage, Matthew D., and Russell H. Fazio. 2015. "The Evaluative Lexicon: Adjective Use as a Means of Assessing and Distinguishing Attitude Valence, Extremity, and Emotionality." *Journal of Experimental Social Psychology* 56: 214–227. doi: 10.1016/j.jesp.2014.10.005
- Rocklage, Matthew D., Derek D. Rucker, and Loran F. Nordgren. 2018. "The Evaluative Lexicon 2.0: The Measurement of Emotionality, Extremity, and Valence in Language." *Behavior Research Methods* 50: 1327–1344. doi: 10.3758/s13428-017-0975-6
- The R Project for Statistical Computing. 2020. "Home." Accessed October 12, 2020. <https://www.r-project.org>
- RStudio Support. 2020. "Home." Accessed October 12, 2020. <https://support.rstudio.com/hc/en-us>
- SAGE Campus. 2020. "Home." Accessed October 12, 2020. <https://campus.sagepub.com/#home/courses>
- Shah, Dhavan V., Joseph N. Capella, and W. Russell Neuman. 2015. "Big Data, Digital Media, and Computational Social Science: Possibilities and Perils." *Annals of the American Academy of Political and Social Science* 659: 6–13. doi: 10.1177/0002716215572084
- Stack Overflow. 2020. "Home." Accessed October 12, 2020. <https://stackoverflow.com>
- Swirl. 2020. "Home." Accessed October 12, 2020. <https://swirl-stats.com>
- Tidyverse. 2020. "Home." Accessed October 12, 2020. <https://www.tidyverse.org>
- Vallacher, Robin R., Stephen J. Read, and Andrzej Nowak. 2017. *Computational Social Psychology*. New York: Routledge.
- Van Geert, Paul. 1998. "A Dynamic Systems Model of Basic Developmental Mechanisms: Piaget, Vygotsky, and Beyond." *Psychological Review* 105: 634–677. doi: 10.1037/0033-295X.105.4.634-677
- Wayment, Heidi A., and K. Laurie Dickson. 2008. "Increasing Student Participation in Undergraduate Research Benefits Students, Faculty, and Department." *Teaching of Psychology* 35: 194–197. doi: 10.1080/00986280802189213

Wharton Credibility Lab. 2020. "AsPredicted." University of Pennsylvania. Accessed October 12, 2020. <https://aspredicted.org>

Wickham, Hadley, and Garrett Grolemund. 2017. *R for Data Science*. Sebastopol, Canada: O'Reilly Media. <https://r4ds.had.co.nz>

Heidi A. Wayment

Northern Arizona University, heidi.wayment@nau.edu

Brian A. Eiler is an assistant professor of quantitative psychology at Davidson College. His research and teaching synthesize concepts from psychology, philosophy, complexity science, artificial intelligence, and data science. Eiler's primary area of expertise is in matching methodological and mathematical concepts to research questions.

Patrick C. Doyle earned his master's degree in the Department of Psychological Sciences at Northern Arizona University and is currently a PhD candidate in social and personality psychology at the University of Georgia. Doyle served as a lead teaching assistant in the course described in this article and incorporates many of these practices in his current teaching and mentoring,

focusing on preparing psychology undergraduates for careers in industry research.

Rosemary L. (Marah) Al-Kire received her master's degree from the Department of Psychological Sciences at Northern Arizona University and is currently a PhD student in social psychology at Baylor University. Al-Kire served as a lead teaching assistant in the course described in this article and continues to incorporate these methods in her doctoral studies. Her research focuses primarily on the interplay between social identities and intergroup relations, particularly within the context of religion.

Heidi A. Wayment is a professor of social psychology and chair of the Department of Psychological Sciences at Northern Arizona University. Her research and teaching efforts lie at the intersection of self-related processes and well-being. She began incorporating computational data science skills into her professional activities in 2010 and is completing an MS degree in applied data science and measurement at the University of Mannheim's School of Business.

CUR's *Excellence in Mentoring* book



Excellence in Mentoring Undergraduate Research incorporates diverse perspectives on mentoring undergraduate research, including work from scholars at many different types of academic institutions in Australia, Canada, the United Kingdom, and the United States. It strives to extend the conversation on mentoring undergraduate research to enable scholars in all disciplines and a variety of institutional contexts to critically examine mentoring practices and the role of mentored undergraduate research in higher education.

To order, visit the CUR Bookstore (<https://bit.ly/CURbkstore>)