

Data Literacy as a Pathway to Data Science at Georgia Tech

Catherine Murray-Rust, Dean of Libraries, Georgia Tech

in collaboration with:

Ameet Doshi, Director, Service Experience and Program Design, Georgia Tech

Jay Forrest, Data and Statistical Analysis Manager Librarian, Georgia Tech

Ximin Mi, Data Visualization Librarian, Georgia Tech

Alison Valk, Multimedia Instruction Librarian and College of Computing Liaison, Georgia Tech

A 2019 Gartner report, *Design a Data and Analytics Strategy*, asserts that “data literacy is a growing challenge for most organizations. By 2020, 80% of organizations will initiate deliberate competency development in the field of data literacy, acknowledging their extreme deficiency.”¹ Campus collaborations increasingly rely on the ability to curate big data in support of the research community, and undergraduate and graduate student work increasingly requires basic skills in interpreting and presenting data. Librarians in their role as educators—serving the whole campus and the whole person—appreciate that data literacy, as distinct from data science in support of research, is fast becoming a required second language in a digital society.

One of the basic tenets of Georgia Tech Library’s strategic initiative, Library Next, is a continuing analysis, based on purposeful interactions with faculty and students, of service gaps that the library can fill. Through Library Next, Georgia Tech is creating a new version of a technologically focused research library for the 21st century. Begun several years ago as a library building renovation, Library Next is now a major transformation of the library’s vision. All components of a research library—inspirational physical and digital environments, curated scholarly content, outstanding services, and information

expertise—continue to change as the research and teaching goals and aspirations of Georgia Tech change. Data literacy, a pathway to data science, is one such service.

Mission Continuity

The Georgia Institute of Technology was founded in 1885 to help propel the agrarian South into the industrial economy. Today, the institute is a leading global research university, committed to improving the human condition through advanced science and technology. Georgia Tech's first library building, which opened in 1907, is one of 108 academic libraries (out of the total of 1,687 libraries) that Andrew Carnegie funded in the United States.² As the Georgia Tech Library redefines its role in the academic community for a new century, Carnegie's words continue to inspire new programs and services for a new generation of students and faculty: "A library outranks any other one thing a community can do to benefit its people. It is a never failing spring in the desert."³

Just as Carnegie libraries promoted reading and writing literacy in the 20th century, the Georgia Tech Library promotes data literacy in the 21st. For the past three years, Georgia Tech has been defining the changing role of higher education in the United States and continuing along its century-old path to becoming what Jan Youtie and Philip Shapira call "a knowledge hub to advance technological innovation and economic development in its region."⁴

The 2018 final report of the Georgia Tech Commission on Creating the Next in Education (CNE) includes a commitment to lifetime education: "a future for college not conceived solely...as a physical place one enters at a particular age and exits when a degree is completed but rather as a platform for an increasingly diverse population of learners."⁵ To make good on its promise, the commission states that innovation will be required to close knowledge gaps, develop and pilot new products and services, and build new technological infrastructure. Library Next developed in parallel to the CNE vision. The Georgia Tech Library worked with brightspot strategy on a master plan for

services that informed a transformative design of the physical spaces in the renovation of the two mid-century library buildings. Through this planning effort, library faculty and staff learned how to engage the campus community in discussions about their needs and wants for the future. The result was a *Playbook*⁶ that closely aligns institutional and user goals and aspirations. Read together, the CNE report and the *Playbook* connect the tradition of Georgia Tech and its library on a path forward that is both practical and inspirational. Data literacy and data science are core elements to Georgia Tech's future.

Learning Together

Supporting data science and data literacy has prompted the research library community to skill up. The Institute of Museum and Library Services (IMLS) has issued several recent grants to prepare librarians and other educators to teach data literacy, among them the ambitious Library Carpentry program expansion by the California Digital Library (CDL).⁷ Similarly, IMLS funded the University of Michigan, Duke University, and the University of Southern California to study data visualization support and usage in libraries, which will culminate in the Visualizing the Future Symposia in 2020.⁸ Earlier IMLS grants expanded the capacity of academic and research librarians to teach data literacy, such as at North Carolina State University Libraries, which supports a robust data visualization program on its campus. Additional IMLS-funded projects at Purdue University, Cornell University, University of Minnesota, and University of Oregon have all contributed to our collective insights and experience in developing strategies and techniques for preparing students for a data-centric future.

Most research universities have data science and data analytics programs directed toward students majoring in disciplines such as science, engineering, and business. Georgia Tech offers courses for undergraduates and graduate students in disciplinary data science, and recently added another master's degree in data analytics, on campus and online, organized into three tracks: analytical tools, business

analytics, and computational data. For the past year, Georgia Tech Professional Education has marketed online boot camps, lasting 24 weeks, in which students, who are generally working full time, build skills in several programming languages and tools. This program competes with the growing number of commercially offered boot camps, which are taught online and in person mainly in major cities in the US. These primarily post-graduate level programs are designed to meet the demand for data officers, data scientists, and related positions, primarily in industry.

On many campuses, however, students who are not in data-focused majors or intending to become data science subject-matter experts have few opportunities to learn such new tools and methodologies. Four Georgia Tech librarians, Ameet Doshi, Jay Forrest, Ximin Mi, and Alison Valk, are committed to filling this service gap at the institute. These librarians collaborate to offer non-credit courses, which the library markets to the campus community on its website. Although some classes are taught at the direct request of faculty, librarians create others based on feedback from the community.

Georgia Tech Offerings

In April 2019 alone, the four librarians leading this program offered 18 classes, some repeated, on a variety of data tools and technologies. In addition to Photoshop for Beginners, Illustrator and Adobe Creative Suite, Introduction to Zotero, and EndNoteX9, they offered Maya 3D Modeling, Introduction to R Studio, Grant Funding through Pivot, Python Twitter Scraping and Analysis, Intermediate R Studio: Visualization, Innovation Plus, and Using LaTeX for Advanced Mathematical Formulas. Some classes are online, and most are offered in the late afternoon or evening. Most require advance registration and all are free to participants.

Ximin Mi's goal is for students to effectively choose and apply the right data-visualization tools to tell the stories behind the data they are working with. With her, students learn through hands-on demonstrations with real-life data. In all "data viz" workshops, Ximin

and her graduate student assistants present interesting features of a data set, the right data format for a visualization design, and the right tool to format data. When moving into the visualization phase, Ximin emphasizes the logic behind the tool's functionality. Through design thinking and hands-on exercises, Ximin's expectation is that students will grow to understand effective design appropriate to the audience's social and cultural background.

Data viz-embedded classes are offered in several academic programs at Georgia Tech, including business, industrial and systems engineering, public policy, and computational media. Ximin points out that in embedded classes, instruction also features hands-on exercises to familiarize students with the tools. In these classes, the demonstration projects are designed with data the class instructor provides, or data about the topics of the class projects. Ximin and her team also offer lab hours and flexible follow-up consultation hours to review students' project designs and implementation.

After three semesters of instruction, the libraries' data viz program taught more than 1,500 learners, including students, faculty, and staff. Students are utilizing the tools they learned to create scholarship in innovative ways. In the spring 2019 semester, 30 students changed their class projects from print to digital. Not only did they save the cost of printing, but also their work is easily saved online for further study, sharing, or publishing in future. Ximin and her group keep detailed assessment of the effectiveness of their teaching so they continue to improve.

Ameet Doshi and Jay Forrest teach a popular coding workshop series on R and R Studio. R is an open source statistical software package commonly used for data analysis, visualization, and an increasing array of related scientific purposes. R Studio is an integrated development environment (IDE), which helps R be more user-friendly and expand its capacity. Because R is a popular open source tool, the breadth of R "packages," or niche system capabilities, as well as community support, continues to expand. These workshops have been well received by the

campus community, especially by the graduate students who comprise the majority of workshop attendees. Between fall 2018 and spring 2019, Ameet and Jay held 21 introductory or intermediate-level R workshops with approximately 450 total attendees. A post-workshop assessment indicates that 40% learned “a great deal” and 35% learned “a lot” by attending the R workshops.⁹ With 74% of attendees indicating they had a positive learning experience, the data suggest that attendees benefit from this new instructional offering by the library. Drawing from 54 qualitative comments, Ameet and Jay observe that students are eager for the library to offer additional advanced and intermediate-level R workshops, training with more tools and techniques (Python, MatLab, text mining), as well as more emphasis on the syntax of R commands. Many students remarked that the 90-minute session “flew by,” so a longer workshop may also be better aligned with graduate student needs.

R has become an increasingly prominent tool for data analysis across academic disciplines and within industry. Because it is open source software, the barrier to entry (cost and availability) is low, but many users still find an in-person introductory workshop helpful for building proficiency. The library has proactively recognized this trend and created a value-added service that reflects well upon the breadth of librarian capabilities, while also evolving the brand of the research library as crucial to academic and research success at Georgia Tech.

Alison Valk, whose specialty is teaching multimedia tools and techniques, believes that no matter the subject material or discipline, the ability to effectively communicate one’s research to a broad audience is vital. A skill set including multimedia-based hardware and software to enhance communication skills is no longer optional, but imperative, and interest in this kind of training has grown over the last six years.

Multimedia includes video editing, visual or graphic design, audio editing, and website development. The manner in which information is transmitted and the communication channels used have the potential

to either obscure or enhance the message. Students and faculty alike can effectively communicate their research or data through compelling storytelling in video production or strong visual designs. Distilling a concept down to its key components or main takeaway creates the building blocks for any visually rich project. For example, Alison says, one must consider such decisions as choice of typography or color palettes. With the right choice a presenter can positively emphasize information, or with the wrong one confuse and distract the audience.

Since 2013, Alison has taught more than 500 course-integrated workshops reaching more than 12,000 students from a variety of academic programs including literature, media and communication, biological sciences, and business, as well as drop-in workshops reaching an additional 1,000 students. In the pre/post tests in Alison's classes, on average, students increased their pre-score with no preparation from 50% correct, to 89% correct on the post-test. The pre/post tests were standardized across topics.

In keeping with the overarching goals of Library Next, Alison is working to enhance the library's role in curriculum development. One example is a computational media research section led by Alison and Ximin Mi. Computational Media is a joint program between the Ivan Allen College of Liberal Arts and the College of Computing. Students in the program are heavily involved in the user experience. In this course, students work with emerging technologies to propose a research initiative utilizing available library resources and showcase their work as the semester concludes. Students are challenged to think creatively and critically in concert with the instructors who assess their progress.

To meet the growing demand for data literacy, several librarians at Georgia Tech submitted three proposals for a new program called Minimester Classes. The Commission on Creating the Next in Education proposed a matrix of minimester classes, which are short courses that can overlay the regular semester calendar. The purpose is to create short courses—one credit for five weeks—that could be used in a variety of ways, including future faculty training, whole-

person education, partnerships, and experiments in pedagogy. The proposals submitted by the librarians range from an updated version of a research methods class, which includes statistical literacy, to open data with R within the framework of open science, to an introduction to video editing with the enticing title of “Scare Me, Teach Me, or Make Me Laugh.”

Conclusion

Within the context of Library Next, the data literacy program is expected to grow and change as the needs of students and faculty change. The biggest challenge the Georgia Tech Library faces is how to scale up the workshops and course offerings to meet the growing demand. This requires librarian-educators who are not only well versed in data literacy tools and technologies, but also skilled teachers and supervisors of teaching assistants. To develop data-literate students and faculty, the library will need to seek ways to broaden its reach. The library can do this through partnerships, generating and redirecting resources, and perhaps one day realizing its dream of having an intelligent agent, like Jill Watson, Georgia Tech’s virtual TA for the Online Master’s in Computer Science, who works alongside Ameet, Jay, Alison, and Ximin to train a new generation of data-literate students and faculty.

Endnotes

1. *Design a Data and Analytics Strategy*, ed. Andrew White (Stamford, CT: Gartner, 2019), 9, <https://www.gartner.com/en/publications/data-analytics-strategy>.
2. “A History of US Public Libraries,” Digital Public Library of America, accessed June 26, 2019, <https://dp.la/exhibitions/history-us-public-libraries/carnegie-libraries>.
3. Andrew Carnegie Birthplace Museum (@carnegiemuseum), “A library outranks any other one thing a community can do to benefit its people. It is a never failing spring in the desert’ - Andrew Carnegie,”

Twitter, October 8, 2018, <https://twitter.com/carnegiemuseum/status/1049269759060189189>.

4. Jan Youtie and Philip Shapira, “Building an Innovation Hub: A Case Study of the Transformation of University Roles in Regional Technological and Economic Development,” *Research Policy* 37, no. 8 (2008): 1188, <https://EconPapers.repec.org/RePEc:eee:respol:v:37:y:2008:i:8:p:1188-1204>.
5. Commission on Creating the Next in Education, *Deliberate Innovation, Lifetime Education: Final Report* (Atlanta, GA: Georgia Tech, April 2018): 5, http://www.provost.gatech.edu/sites/default/files/documents/deliberate_innovation_lifetime_education.pdf.
6. *User Research Project: Part 1: Research Report & Playbook* (Atlanta, GA: Georgia Tech, March 18, 2014), <https://www.library.gatech.edu/sites/default/files/2019-01/part1.pdf>.
7. John Chodacki, “Skills Training for Librarians: Expanding Library Carpentry,” University of California Curation Center (UC3), November 6, 2017, <https://uc3.cdlib.org/2017/11/06/skills-training-for-librarians-expanding-library-carpentry/>.
8. Visualizing the Future Symposia website, accessed June 13, 2019, <https://visualizingthefuture.github.io/>.
9. n=81 respondents; 18% survey response rate.

© 2019 Catherine Murray-Rust, Ameet Doshi, Jay Forrest, Ximin Mi, and Alison Valk



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

To cite this article: Catherine Murray-Rust, Ameet Doshi, Jay Forrest, Ximin Mi, and Alison Valk. “Data Literacy as a Pathway to Data Science at Georgia Tech.” *Research Library Issues*, no. 298 (2019): 6–15. <https://doi.org/10.29242/rli.298.2>.