Teaching Philosophy

Karsten T. Maurer

We learn from doing. I find that the deepest learning is done when there is an important task that we don't initially know *how* to do, but we find the way to get it done right. In teaching statistics and data science, I identify mastery when a student can consistently rise to meet the new challenges, problems, and opportunities they encounter in the field. Mastery of a topic is not built by collecting up all the requisite bits of knowledge, then squirreling them away for use in the right time and place. Under the ever-evolving and expanding tent of data science and statistics, seeking mastery through learning every last concept, method and tool becomes Sisyphus' eternal boulder pushing. Rather, mastery requires several skill sets in order to consistently solve new problems and deliver actionable results. Naturally this begins with a foundational understanding of the discipline and generally applicable methods and tools; but also necessarily requires critical thinking skills, an adaptive and adoptive use of new methodologies and technologies, and effective communication skills.

My primary roles in teaching are to present new ideas, to support students as they explore these ideas, and foster the expansion of their problem-solving and critical thinking skills so that they can continue to grow, adapt and thrive while working as data professionals. This requires a larger emphasis on building foundational knowledge for students new to the field, and a more expansive and sophisticated development as students progress in their education. As such I address my teaching approach for introductory students separately from my approach for students pursuing careers in statistics and data science.

Introductory students may be starting their pursuit of a career in statistics, but often the course serves as the terminal education in data thinking for students outside the discipline. It is therefore important to teach the foundational concepts of working with and thinking about data in a way that supports both groups. I want my introductory students to learn the following facets of statistical literacy: Learn the principles for obtaining trustworthy data from surveys or experiments and the extent to which that data can inform them about a population; Learn to properly organize, store, display and summarize data; Learn to use simple statistical models; Learn to conduct statistical inference for simple situations. I find that the easiest way to convince introductory students that data thinking is worthy of their learning efforts is to convey clear value by discussing real world applications with real world data that has real world impact. For example, when we discuss multivariate relationships, I may choose to use data from The World Bank that explores the connection between the GDP per-capita, life expectancy, and population size over time from countries in each continent. Knowing that the data informs their worldview can help to engage students

1

of all backgrounds with the introductory course content. Conversely, I avoid or delay topics and technologies that disengage student and are not entirely necessary for fostering the core of data thinking. For instance, intensive statistical theory associated with probability distributions can be saved for more advanced courses that are designed for students majoring in statistics or mathematics. Many of the foundational inferential concepts can be carried well with a randomization-based inference approach. I teach the concepts of inference using simulation-based testing and bootstrap intervals using the StatKey software. I use several apps to help convey statistical concepts (Gapminder, Rossman/Chance Applets, Statkey), but in introductory courses that act as a gateway to subsequent statistics and data science courses, I additionally seek to give them exposure to computational tools that they will use in their future classes and careers. Open source software is the direction that many statistical and most data science professionals are heading, thus I introduce R as the programming language used for data cleaning, plotting and modeling done in my introductory courses. My introductory course pedagogy aligns with the principles laid out in the Guidelines for Assessment and Instruction in Statistics Education Report (GAISE, 2016) which provide a set of guidelines for efficiently and effectively teaching introductory statistics students to start viewing the world through a statistical lens. The six overarching recommended principles can be summarized as: (1) Teach statistical thinking through problem solving and multivariate problems, (2) Emphasize understanding concepts over mathematical memorization, (3) Use real data, (4) Foster active learning, (5) Use technology to explore both concepts and data and (6) Use assessments as a tool for student learning. To summarize, in my introductory course I integrate the GAISE principles, randomization-based inference methods and my previous experience to provide content and motivation to help my students attain statistical literacy.

I approach teaching the students in my more advanced courses in statistics and data science differently. All courses have official content and learning outcomes that dictate the minimal curricular coverage, but there remains flexibility to make important pedagogical decisions about how each course is shaped. At the onset of a course I try to gauge what domain knowledge and technical skills that my students already have by looking into curricular coverage in the prerequisite coursework, collecting start of semester surveys and having discussions with previous instructors of the course. I use this to help assess if there are academic or technological hurdles that need to be addressed for students to be successful throughout my course. The most common challenge is that many of my students lack the programming experience that would help them to comfortably approach data problems computationally. In some cases, it requires that we use a portion of the semester getting the entire class up to speed with coding for them to be successful; in other cases it may just require guiding a few

students to the supplementary materials that can help get them past minor hurdles. Building my students' comfort and confidence in their coding and data management skills is critically important in my courses where we often work with large data that has not been pre-cleaned while exploring new methods and concepts. I am a firm believer that if the real world applications of what is being taught are accomplished though coding, then the students need hands-on-keyboards inside and outside of the classroom. I encourage students to code along with me while I walk through coding examples in class, I provide in-class assignments where they code collaboratively with classmates to try new ideas, and I assign homework where well-documented code is a required deliverable. In addition to coding, I emphasize critical consideration of the mathematical and computational underpinnings of the methods taught in my courses. When statistical methods and predictive models are shallowly understood and treated as black-boxes, misuse and misinterpretation run rampant. For example, in my statistical learning course I will assign reading for the Lum et al. (2016) article titled *To Protect and Serve* on the racial bias found in Oakland's predictive policing procedures. I use this to discuss the real world implications of ignoring algorithm bias and the responsibility they have to consider data ethics in their professional work. Lastly, I focus on developing my students' ability to communicate their findings as they explore new topics. We work on identifying the audience, organizing and formatting results, documenting a reproducible workflow, developing data visualizations and simple examples to increase accessibility, and sharing the work through written and oral data narratives. This emphasis on data storytelling is consistently present through homework and larger projects where my grading always includes both points and feedback for clear communication and documentation of their work. I work to incorporate these elements into each of my advanced courses, because a student that can think critically, code competently and communicate clearly is equipped to effect real change in the world with their data skills.

An important component in my teaching philosophy is my approach to the day-to-day activity in my classroom. The default principle that guides the way I structure courses and interact with students is "the golden rule"; treat students as I would want to be treated if I were a student in my class. This mentality leads to the interplay of responsibilities that students and I hold. For instance, when I set deadlines for students to turn in assignments or projects, I owe it to them to provide prompt feedback. This is not only a fair exchange, but also ensures that the assessment is a part of the learning process and not solely for setting course grades. When I prepare lesson materials, I make sure to run through how I plan to explain the difficult topics so that the introduction of new ideas is expressed clearly and concisely to make the best use of students' time. In my lessons, I encourage students to ask questions and in turn I also ask students many questions throughout the class. This

forces students to remain engaged and ensures that they are free to ask for clarification. I hold consistent and generous office hours, and often schedule short meetings as needed. This accessibility comes with an explicit disclaimer to students that I value my time and that they need to have organized their thoughts and questions before the meeting. I also make a point to be inclusive and inviting in my course management to ensure that students from under-represented communities on campus understand that they are valued as part of the classroom community. A simple, yet imperative example of this is that I make immediate efforts for them to feel known as individuals; learning their preferred name and pronouns and how to pronounce their name properly. I enjoy watching the sudden and genuine increase in student engagement that happens early in the semester as students realize that I am addressing them individually, some seem quite surprised that I would have spent the time or effort to learn their names. Establishing a give-and-take relationship based on mutual respect opens the channels of communication through which I can help to more effectively guide student learning.

I help my data science and statistics students to develop mastery through the way I present new ideas, support them as they explore these ideas, and foster their problem-solving and critical thinking skills along the way. I best accomplish this through mindful planning of course structure and course materials that continuously challenge students as their understanding and abilities expand. I am also mindful to communicate clearly, respectfully and thoughtfully with students to keep their minds open to what I hope to teach them. This is the strategy that I will carry forward; however, with all well laid plans there remains room to evolve while new experience and pedagogical ideas continue to shape my teaching philosophy.