Data Science Education With MOOCs and Active Learning

Rafael Irizarry, Harvard University, Dana-Farber Cancer Institute; *Stephanie Hicks, Dana-Farber Cancer Institute*

Educational institutions across the world are responding to the unprecedented demand of training in statistics and data science by the creation of new courses, curriculums and degrees in applied statistics and data science. We have participated in two data science courses taught at Harvard and the creation of an online course of data analysis for the life sciences. In this presentation, I will discuss our approach to developing a MOOC based almost exclusively on real-world examples and how our lecturers revolved around dozens of exercises that required R programming to answer. We taught a total of seven different MOOCs in topics ranging from basic statistics to the analysis of RNA-Seq data. I will also discuss how the experience of teaching a MOOC changed the way I teach in the classroom. Specifically I will describe how we transformed the usual classroom by using active learning and collaborative techniques to teach concepts in statistics and data science. Examples include motivating real-world problems with data and code instead giving of traditional lectures, using Google Polls to get live feedback, and teaching the importance of reproducible research and collaborative practices with git/GitHub. In every assignment in the classroom, students performed a complete data analysis integrated programming skills with statistical analyses. As a final project, students analyzed a dataset of their choice and created a website and two minute video summarizing their results. This led to many students successfully obtaining jobs by discussing their homework and final projects in job interviews.