Anyone Can Cook: Foundations of Programming and Data Science
Justification and Alignment with Ed & IS Guiding Principles

Ozan Jaquette

12/3/2020

1 "Anyone Can Cook" Sequence

1.1 Overview of Sequence

Anyone Can Cook: Foundations of Programming and Data Science is a two-course sequence that teaches foundational programming and data science skills. The sequence is explictly designed for students who do not have a programming background and perhaps never thought they would learn to write code.

The sequence primarily uses R, a free, open-source, object-oriented programming language. R is the industry standard programming language for statistical analysis in academic research. R is also one of the most popular languages for "data science." Eventually (e.g., 5, 10 years), when another programming language (e.g., *Python*) surpasses R as the most popular do-it-all tool for academic research, course materials will be revised to use that programming language but course learning goals will remain largely unchanged.

The first course in the sequence is EDUC XXX: Introduction to Programming and Data Management. The primary goals of this course are (1) to teach fundamental skills of "data management" (e.g., cleaning data, creating variables, merging datasets) which are important regardless of which programming language you use, and (2) to develop a strong foundation in the R programming language. The second course in the sequence is EDUC XXX: Fundamentals of Programming. The course teaches practical programming skills/concepts (e.g., using Git for version control and collaboration, writing functions, regular expressions) that are fundamental across all modern object-oriented programming languages.

1.2 Justification for Sequence

UCLA launched the "DataX" initiative in 2020. Provost Emily Carter and Vice Chancellor for Research Roger Wakimoto explained the motivation for DataX in their June 2020 call for white papers from academic units (Carter & Wakimoto, 2020):

> [DataX] emphasizes our conviction that the ongoing revolution in our ability to access and analyze data has transformative potential for all activities of the university. It can lead to innovative research, which often transcends academic units; new modalities of scholarship ...; timely critique of the social, political, and ethical implications of data; groundbreaking initiatives in undergraduate and graduate education; and novel pathways for community engagement.

Free, open-source, object-oriented programming languages such as R and *Python* are the engines of this revolution. These do-it-all programming languages fundamentally change the way we do research. Paul Jansen, CEO of Tiobe Software, which tracks programming language popularity, stated that "Universities and research institutes embrace Python and R for their statistical analyses ... [By contrast] the days of commercial statistical languages and packages such as SAS, Stata and SPSS are over" (Tung, 2020). These trends indicate that we are quickly approaching a time when researchers who cannot code will be excluded.

The data science revolution represents both an opportunity and a challenge for education research. The opportunity is that once our students master the basics of languages like *R* and *Python*, applications like social network analysis, interactive mapping, and web-scraping are easy to learn because they are just applications of basic programming concepts. The challenge is that learning how to code is hard! This is especially true for the majority of our students who do not have a programming background and may feel anxious about learning to code. I created the *Anyone Can Cook* sequence so that our students are systematically included in – rather than tracked away from – the opportunities created by the data science revolution.

Anyone Can Cook targets a niche not met by existing offerings across UCLA. Existing programming/data science course offerings tend to fall into one of two buckets. First, there are the math-heavy courses like those offered by the statistics department, which are too advanced and move too quickly for many of our students. Second, there are courses that provide a surface-level overview of many topics, but fail to help students develop competence in essential skills. By contrast, *Anyone Can Cook* is explicitly designed for students who do not have a programming background. The guiding philosophy is "deep and slow." That is, each 10-week course introduces a small number of core topics. We devote substantial time to each topic and students become proficient through weekly problem sets that enable them to practice and gain confidence.

In addition to strong enrollment demand from Education students, *Anyone Can Cook* is becoming popular with students from public policy, sociology, Anderson/business, political science, and other departments. The department of public policy has expressed interest in cross-listing *Anyone Can Cook* courses and including them in a new data analytics certificate they are creating. Over the next few years, I believe that *Anyone Can Cook* will become the go-to programming/data science sequence for UCLA social science and humanities students who do not have a programming background.

1.3   Alignment with Ed & IS Principles

Ed & IS Principles

>   GSE&IS is dedicated to inquiry, the advancement of knowledge, the improvement of professional practice, and service to the education and information professions. We develop future generations of scholars, teachers, information professionals, and institutional leaders. Our work is guided by the principles of individual responsibility, diversity, social justice, an ethic of caring, and commitment to the communities we serve.

***Anyone Can Cook* Philosophy and Principles**

- Our goal is to help students realize their full potential
- Anyone can cook and everyone is welcome
- Go slow and go deep
- Structure lectures and assignments so that students develop competence and confidence
    - Detailed lecture materials will include access to code and data so that students can run code on their own computer
    - Weekly problem sets will give students the opportunity to practice skills introduced in lecture materials
- Create an online community (using GitHub) where students help one another and feel comfortable asking for help

2   EDUC XXX: Introduction to Programming and Data Management

Links to syllabus and course website

- Syllabus: https://anyone-can-cook.github.io/rclass1/syllabus/rclass1_syllabus.html
- Course website: https://anyone-can-cook.github.io/rclass1/

2.1 Course Description

The primary goals of this course are (1) to teach fundamental skills of "data management," which are important regardless of which programming language you use, and (2) to develop a strong foundation in the R programming language. For goal (1), most statistics courses teach you how to analyze data that are ready for analysis. In real research projects, data management – the process of cleaning, manipulating, and integrating datasets in order to create analysis datasets – is often more challenging than conducting analyses. For goal (2), R is a free, open-source, object-oriented programming language. R is the most popular language for statistical analysis and one of the most popular languages for "data science." This course is designed for students who do not have a programming background. Students will become proficient in data management and R programming through weekly problem sets, which will be completed in groups. Course format consists of weekly asynchronous lectures and weekly synchronous workshop-style class sessions on Zoom.

3 EDUC XXX: Fundamentals of Programming

Links to syllabus and course website

- Syllabus: https://anyone-can-cook.github.io/rclass2/syllabus/rclass2_syllabus.html
- Course website: https://anyone-can-cook.github.io/rclass2/

3.1 Course Description

This course teaches practical programming skills and concepts that are important across all modern object-oriented programming languages (e.g., *Python*, *Javascript*). Course topics include: organizing files, folders, and scripts; reading (importing) and writing (exporting) data; using Git and GitHub for version control and collaboration; writing functions; iteration (e.g., "loops"); conditional execution; strings and regular expressions. These general programming skills are prerequisite for flashier data science applications (e.g., web-scraping, streaming and analyzing social media data, interactive maps). Students will become proficient in programming skills/concepts through weekly problem sets, which will be completed in groups. Course format consists of weekly asynchronous lectures and weekly synchronous workshop-style class sessions on Zoom.

References

Carter, E. A., & Wakimoto, R. (2020). Call for datax white papers. Retrieved from https://evcp.ucla.edu/announcements/2019-20/call-for-datax-white-papers/

Tung, L. (2020). Programming language rankings: R makes a comeback but there's debate about its rise. Retrieved from https://www.zdnet.com/article/programming-language-rankings-r-makes-a-comeback-but-theres-debate-about-its-rise/