Chiin-Rui Tan[1]

[1]Rho Zeta AI

February 17, 2021

Abstract

Many societal opportunities and challenges, both current and future, are either inter- or transdisciplinary in nature. Focus and action to cut across traditional academic boundaries has increased in research and, to a less extent, teaching. One successful collaboration has been the augmentation of fields within the Humanities, Social Sciences, and Arts by integrating complementary tools and methods originated from STEM. This trend is gradually materializing in formal undergraduate and secondary education.

The proven effectiveness of Jupyter notebooks for teaching and learning STEM practices gives rise to a nascent case for education seeking to replicate this interdisciplinary design to adopt notebook technology as the best pedagogical tool for this job. This article presents two sets of data to help argue this case.

The first set of data demonstrates the art of the possible. A sample of undergraduate and secondary level courses showcases existing or recent work of educational stakeholders in the US and UK who are already pioneering instruction where computational and data practices are integrated into the study of the Humanities, Social Sciences, and Arts, with Jupyter notebooks chosen as a central pedagogical tool. Supplementary data providing an overview of the types of technical material covered by each course syllabi further evidences what interdisciplinary education is perceived to be or is already feasible using this Jupyter technology with student audiences of these levels.

The second set of data provides more granular, concrete insight derived from user experiences of a handful of the courses from the sample. Four instructors and one student describe a range of pedagogical benefits and value they attribute to the use of Jupyter notebooks in their course(s).

In presenting this nascent case, the article aims to stimulate the development of Jupyter notebook-enabled, computational data-driven interdisciplinary education within undergraduate and secondary school programs.

Introduction

We live in an increasingly complex environment. At the time of writing, society is at the confluence of a devastating pandemic, unprecedented vaccine development, violent civil unrest fueled by rampant online disinformation, and more frequent and extreme weather, amongst a host of life-changing phenomena. No discipline truly exists in a vacuum, nor can researchers work in a silo to satisfactorily address real-world opportunities or challenges. Inter- and transdisciplinarity are better understood as fundamental approaches, rather than nice-to-have, in order to achieve a variety of objectives, whether academic, industrial, governmental, or beyond.

The research community has responded to the need to adapt. Collaboration across the boundaries of traditionally distinct disciplines is detectable simply from the new, enigmatically branded institutes, centers, and working groups that have been established worldwide over the last few decades. Fields such as Digital Humanities and Computational Social Science are now ubiquitous and reflect a wider societal trend of modernizing things historically considered the domain of the Humanities, Social Sciences, or Arts (HSSA) by weaving of contemporary computer-powered and data-centric (computational data) tools and methods into their fabric.

The influence of this trend to transform using computers and data has, however, trickled only gradually from research to education. In the UK the topic of interdisciplinary teaching has been less of a priority for universities [1]. This could explain the Royal Society having to draw attention in a report as recently as 2019 to the interdisciplinary nature of computational data skills and the need for further consideration around how universities can teach these effectively [2].

Indeed, when it comes to revamping HSSA curricula with computers and data, without first tackling the existing gap in STEM education itself, any ambition for higher complexity is futile. This impracticality is further compounded by the likelihood of myopic attitudes in earlier stages of education, as adulthood and the consideration of future-readiness are conceptually furthest away.

Despite these and other challenges, however, the signals of interdisciplinarity from both undergraduate and secondary education are strengthening. In London, UK, an Arts and Science BASc was launched back in 2012 by University College London, offering and requiring students to take a mix of non-science and science courses. From an original intake of 87, just 4 years later by 2016, this had grown to 450 students in steady state [1]. Additionally, at the secondary level one institution, North London Collegiate School, is piloting a Digital Humanities toolkit that will enable their students to undertake modern interdisciplinary learning activities, a project I am involved in.

Figure 1: Year 7 (6th Grade) students at North London Collegiate School using Jupyter notebooks in class to work on a computational sociology exercise in Python (Source: North London Collegiate School).

The purpose of this article is, therefore, to make the case for computational data-augmented HSSA education to adopt Jupyter notebooks as a central pedagogical tool, given the technology is widely held as the gold-standard for teaching and learning STEM tools and methods, hence interdisciplinary courses should take advantage of the many transferable benefits on offer. These utilities are well-documented in the official handbook [3].

In the next section, we present the first of two sets of data in support of this case, a regional sample of courses that represent some of the early adopters, i.e., computational data-transformed HSSA education at the undergraduate and secondary level designed to harness the pedagogical tooling power of Jupyter notebooks. This demonstrates what is considered achievable and desirable for interdisciplinary HSSA study for students at these two stages of education with this choice of educational technology.

The Art of the Possible: a sample of interdisciplinary HSSA courses using Jupyter notebooks

Using publicly available information supplemented with knowledge of my work, I compiled a sample of twenty courses that meet the interdisciplinary and tooling specification advocated by the article, within the regional scope of the US and UK:

Index	Instructor/s	Institution	Region	Education Level	Course Name	Department	Course Date (most recent)	Programming Language	Computational Data Practices							
									Programming Language Basics	Data Scraping	Data Cleaning & Transformation	Data Analysis	Data Visualization	Mathematical/Statistical Modeling	Machine Learning	Dynamic Optimization
0	Brian C. Jenkins	UC Irvine	US	Undergraduate	Computational Macroeconomics	Economics	Winter 2021	Python	Y		Y	Y	Y	Y		Y
1	Elisabeth Sadoulet	UC Berkeley	US	Undergraduate	Introductory Applied Econometrics	Agricultural & Resource Economics	Spring 2021 Fall 2020	R	Y		Y	Y	Y	Y		
2	Meredith Fowlie	UC Berkeley	US	Undergraduate	Regulation of Energy and the Environment	Agricultural & Resource Economics	Spring 2021	Python	Y		Y	Y	Y	Y		
3	James Millington Jon Reades	King's College London	UK	Undergraduate	Foundations of Spatial Data Science (previously Geocomputation)	Geography	Fall 2020	Python	Y		Y	Y	Y			
4	Jon Reades Naru Shiode	King's College London	UK	Undergraduate	Principles of Spatial Data Science (previously Spatial Analysis)	Geography	Winter 2021	Python			Y	Y	Y	Y		
5	James Millington	King's College London	UK	Undergraduate	Applications of Spatial Data Science (previously Applied Geocomputation and	Geography	Fall 2020	Python			Y	Y	Y	Y		

Table 1 represents simplified sample data. The full dataset with further columns providing course codes, all known course dates, and data sources is also available via a Zenodo archive [4].

Before any discussion of Table 1, I will make transparent the following caveats.

The courses discovered are likely a small sample of the true US and UK population given significant constraints and the individualistic, ad-hoc nature of my search and technology used, for example only encountering material that is online, unrestricted access, English-language, sufficiently keyword-matched and search engine-ranked. Furthermore, my sampling will be affected by selection bias. Courses with a more discoverable digital footprint, for a given specific search strategy, may correlate with variables such as 'success,' institutional resources, and or course, recency. Another caveat is the subjectivity of scope; courses were excluded on the basis of being judged to have either insufficient or too much computational data content, and/or an imbalance in terms of transferring computational data practices to HSSA vs. STEM contexts. A final caveat is the course data collected representing the (publicly stated) intentions, rather than, for example, actual execution, evaluation, etc.

With these caveats in mind, we examine Table 1, the first set of data in support of the article's case for Jupyter notebook adoption.

Of the twenty courses in the sample, 18 found are at undergraduate level with two from secondary schooling. 14 courses are US-based with the remaining six from the UK. In terms of the programming language on which the most recent course is based, Python is the dominant choice in this sample except for two courses: Introductory Applied Econometrics (Index 1) which uses R, and Interdisciplinary Computer Modeling (Index 17) using Julia.

The final eight columns under the heading "Computational Data Practices" provide more relevant insight into this sample. This matrix contains boolean data, manually created by reviewing the most granular instructional material available, that indicate which of eight categories of computational data tools and methods ubiquitous in STEM are covered within each course syllabus using Jupyter notebooks. Through this subset of the data we start to build a picture of interdisciplinary HSSA education Jupyter notebook use-cases that are considered possible for students at this level in these regions, simply by inspecting each record's combination of Course Name, Department, and the Computational Data Practices marked.

Referencing data outside of Table 1, I argue that many of the notebook use-cases can be seen as novel and pathbreaking, given the stage of education. At UC Berkeley the undergraduates taking Data, Law & Prediction (Index 14) are taught natural language processing techniques, then perform sentiment analysis to explore the question: "Did the way judges, prosecutors, and witnesses talk about moral culpability change after the Bloody Code was mostly repealed in 1827 (at the leading edge of a wave of legal reform in England)?" [5]. At Notting Hill & Ealing High School, the Year 10 (9th Grade) students taking my Code Art course (Index 19), after receiving minimal basic vector math instruction then started to customize animated elements for the GIFs they were programming.

The lack of a 'control' against which to compare even just syllabus design makes inference impossible. We cannot attribute the true educational effect of deploying Jupyter notebooks as the core pedagogical tool versus the alternatives. However, this first set of data does hint at what is considered feasible, actually executable, and, for the many sampled courses with some history of being offered [4], potentially successful by these pioneering educators seeking to provide interdisciplinary HSSA courses for undergraduate and secondary level students, leveraging notebook tooling.

With the art of the possible in mind, the next section helps to fill in a few gaps by presenting the second set of data in support of the article's case, namely user experiences of five stakeholders of a handful of the sampled courses in Table 1. These four instructors and one student provide more conclusive evidence of the pedagogical benefits related to their course experiences that they attribute to Jupyter notebooks.

User Experience insight: pedagogical benefits attributed by course stakeholders to the use of Jupyter notebook

Powerful, versatile, and more accessible tool for teaching

Reades, from the Department of Geography at King's College London (KCL), has written a paper "Teaching on Jupyter - Using notebooks to accelerate learning and curriculum development" [6], which reports learnings from methodological and pedagogical research the department undertook over a five-year period; the 3 courses in the sample (Index 3, 4, and 5), a suite of geocomputation modules, are a product of this research. Regarding their choice of Jupyter notebook technology for this suite, Reades states that it "removes significant barriers to teaching by providing a flexible and familiar interface that hides. . . some of the complexity of managing local programming language installations whilst also allowing instructors to provide rich media and contextual information next to the code where it is needed the most." This conclusion is especially credible given the department evaluated a variety of educational technologies in their quest for the best pedagogical tool for their geocomputation suite ambitions (see section "4 How we Reached Jupyter" [6]).

Jenkins, instructor of the Computational Macroeconomics course (Index 0) has also written a paper about his course, sharing valuable details about its development, delivery, and outcomes [7]. In the context of the course confirmed to have been a success, Jenkins stresses that "From a pedagogical perspective, the Jupyter Notebook is a fantastic tool that makes it easy to teach. . . The Notebook is a wonderful instructional tool because it allows me to write notes and instructions in HTML that students can read in advance and then we complete the Notebook together in class." [7]. So convincing has this experience been that Jenkin's in fact writes that his paper is intended to be a contribution to support better teaching by sharing how to take advantage of Jupyter notebook's versatility.

Nelson, whose course instruction includes both Digital Methods for Social Sciences and Humanities and Text Analysis for Digital Humanists and Social Scientists (Index 8 and 9), expresses similar sentiment on her website [8]. In the context of her overall passion for teaching programming to social scientists and humanities scholars, of which her courses in the sample can be assumed as examples of this activity, Nelson remarks that "I have found Jupyter Notebooks are, by far, the most powerful pedagogical tool to teach programming to students to any level, but particularly those with no programming background." Like Jenkins, she similarly recommends using this technology to teach applied programming.

Effective tool and reuseable reference resource for shareable, flexible learning

An undergraduate student who took the "Foundations of Spatial Data Science" course (Index 3), a module from the geocomputation suite previously referenced, is quoted on the KCL website describing the skillset gained as the most important learning of their entire university experience. Having since been employed in Data Science after completing their degree, they describe how "The Jupyter notebooks were fantastic and I regularly find myself referring back to them in my own work to find solutions." [9]. This user's successful learning retention, repeated reuse of course material, and proven ability to transfer knowledge from academic to industry context corroborates the conclusion made by KCL's Department of Geography that Jupyter notebooks were the best means to enable the students in their geocomputation program to develop the target computational data skills [6].

Walsh, instructor of the Introduction to Cultural Analytics: Data, Computation, and Culture course (Index 6), details in her blog about her pedagogical experience and thoughts on notebooks and the wider Jupyter ecosystem [10]. A general fan of this technology, she chose to create and publish most of the code for her classes as Jupyter notebooks. Walsh helpfully has a section in her course textbook, which she in fact created with Jupyter Book, entitled "Some of Jupyter's Nice Features", in relation to using notebooks for learning and teaching; here she highlights the technology's visually accessible data display, in-document visualization of data, and being able to mix executable code, text, images, and links [11].

In addition, Walsh explains the learning benefits for both course students but also beyond, enabled by the set-up she has put in place: "Publishing this code online allows students to reference and return to the material covered in class, but it also allows people who are not enrolled in the class... to learn and follow along, too. There's a big appetite for learning these computational skills, especially among people from humanities disciplines who have minimal or non-traditional programming education and experience. So these materials are an important resource for the whole community." [10]. There are evidently a variety of pedagogical advantages for Walsh that are made possible by her choice to use Jupyter notebooks as the core learning environment for her course material.

Conclusion

In this article, we considered two sets of data in support of the emerging case to adopt Jupyter notebooks as a pedagogical tool for interdisciplinary HSSA education.

The first set of data provided an overview of what educators already bought-in to this case consider feasible at the undergraduate and secondary education stage in the US and UK. From both the Computational Data Practices matrix and supplementary references we discovered example use-cases for Jupyter notebooks to enable novel and progressive syllabus content.

The second set of data helped to flesh out the initial insights from the first as we learned how highly four instructors and one student of some of these pioneering courses rated their Jupyter notebook experiences in that educational context, whether for teaching, learning, or transferable knowledge and skills.

This nascent case ultimately aims to stimulate Jupyter notebook-enabled, computational data-driven interdisciplinarity within undergraduate and secondary school programs, through whatever development pathway possible. Democratizing effective future-ready education to the next generation of citizens today is essential preparation for the society of tomorrow, and should be universally considered a high priority.

Chiin-Rui Tan is the founder and CEO of Rho Zeta AI, an EdTech startup in London, UK, with a mission to democratize opportunities for young people to develop modern computational data literacy. Working primarily with schools, she advises, develops, and delivers pioneering curricula, syllabi, instruction, and tools that introduce students and educators to contemporary scientific computing, data science, open-source Python software, and technology stacks. She has a particular interest in making technical learning accessible to harder-to-reach youth, especially those disengaged with STEM. She was the first Head of Data Science appointed to the Foreign & Commonwealth Office of the British Government, where she built the department's first artificial intelligence, and developed new open-source intelligence (OSINT) tools as per her Cyber & National Security remit. She received a B.A. degree in economics from the University of Cambridge, Cambridge, UK. An original founding member of the R-Ladies Global open-source community, she is a past Linux Foundation grant recipient who co-led the establishment of this network of R user groups worldwide. Contact her at info@rhozeta.io.

References

[1] "The British Academy", https://www.thebritishacademy.ac.uk/documents/213/crossing-paths.pdf.

[2] "The Royal Society (2019)". https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf?la=en-GB&hash=212DAE7D599B0A48687B372C90DC3FEA.

[3] *Teaching and Learning with Jupyter.* https://jupyter4edu.github.io/jupyter-edu-book/, 2019.

[4] "Tan".

[5]U. C. Berkeley, "ds-modules/Legalst-123", https://github.com/ds-modules/Legalst-123/blob/master/labs/24_Sentiment%20Morality/4-12_Morality_Sentiment_Analysis_SOLUTIONS.ipynb.

[6]J. Reades, "Teaching on Jupyter", *REGION*, vol. 7, no. 1, pp. 21–34, Mar. 2020, doi: 10.18335/region.v7i1.282.

[7]Jenkins, "A Python-Based Undergraduate Course in Computational Macroeconomics". .

[8]"Nelson".

[9]King's College London, "Spatial Data Science Pathway", https://kingsgeocomputation.org/teaching/sds/.

[10]"Introducing an Interactive Textbook for Cultural Analysis with Python". https://melaniewalsh.org/intro-ca-jupyter-book/.

[11]"How to Use Jupyter Notebooks". https://melaniewalsh.github.io/Intro-Cultural-Analytics/Python/How-to-Use-Jupyter-Notebooks.html.