COLLABORATORY
AT COLUMBIA UNIVERSITY

Collaboratory Projects: Aims & Impacts
Spring 2016 - Spring 2021

During spring 2021 the Collaboratory at Columbia checked in with the Collaboratory Fellows to take stock of all funded Collaboratory projects and to document snapshots of progress for each. The course descriptions and impact statements listed below are in chronological order, the projects listed first were funded first. To date, the Collaboratory has influenced 4,843 students.

What is a Book for the 21st Century?
Transforming Texts: Computational Approaches to Text Analysis & Visualization

Pamela Smith, GSAS, Department of History • Steven Feiner, SEAS, Department of Computer Science • Terence Catapano, Librarian and Digital Humanities • Tianna Uchacz, GSAS, Department of History • Dennis Tenen, Arts & Science, Department of English and Comparative Literature

Aims: The objectives of this collaboration between an historian, a computer scientist, a scholar of the digital humanities, and a literary and media theorist include creation of a series of co-taught Digital Humanities/Computer Science courses open to graduates and advanced undergraduates. The courses aim to develop baseline standards for digital literacy and sustainable methods of teaching in multidisciplinary teams. They bring together faculty, students, and scholars of the humanities and social sciences with those of the digital humanities and computer science to rethink the book as a scholarly object for the 21st century from the historical perspective of the early modern world. The courses equip students with the technical skills and conceptual methodologies to treat encoded text as data for analysis and visualization—acknowledging that words, texts, and documents continue to play a central role in modern society while presenting unique challenges for manipulation as data.

Impacts: This team has launched 4 courses, with 1 additional course planned for the future. This project also published a critical digital edition of a 16th century manuscript, "BNF Ms. Fr. 640," with help from students in their Collaboratory courses under the project title "Secrets of Craft and Nature in Renaissance France." This critical digital edition is one component of "The Making and Knowing Project," directed by Pamela Smith and involving additional team members. Other outcomes include the course "sandbox" for HIST/ENG4090. Their work has attracted attention, including an article in the Columbia Engineering Magazine. This team also leveraged Collaboratory funding to obtain a Provost's Interdisciplinary Teaching award to modify existing coursework to develop a new course in Spring 2023. This team also co-developed the Digital Literacy Competency Calculator, a tool made available to educators to "make competency mastery transparent for ourselves and our students." Columbia's Center for Teaching and Learning provides free consultation to Columbia faculty interested in using this tool, which at least one other Collaboratory project has utilized. Students impacted to date: 151 + auditors

What is a Book for the 21st Century? (continued)

Courses launched:

- HIST8975GR, "What is a Book in the 21st Century?," taught by Terry Catapano and Pamela Smith with guest lectures by Steven Feiner. Taught in Spring 2017 (6 students + auditors).
- COMS W4172, "3D User Interfaces and Augmented Reality," taught by Steven Feiner and Carmine Elvezio, Spring 2017 (46 students), 2019 (42 students), 2021 (33 students).
- ENGL84031/HISTGU4031/COMS4495, "Transforming Texts: Textual Analysis, Literary Modeling, and Visualization," taught by Pamela Smith, Dennis Tenen, and Steven Feiner, Spring 2019 (18 students + auditors).
- HISTGU4962, "Making and Knowing in early modern Europe: Hands-on History," taught by Pamela Smith, Summer A 2021 (6 students).

Courses yet to be launched:

- HIST/ENG4090, "Craft, Literature, and Data: Hands-on Material and Digital Methods for the Study of Recipes, How-to Texts, and Technical Instructions, taught by Pamela Smith and Dennis Tenen, Spring 2023.

Points Unknown:
New Frameworks for Investigation and Creative Expression Through Mapping

Juan Francisco Saldarriaga, GSAPP, Center for Spatial Research • Marguerite Holloway, School of Journalism, Department of Science and Environmental Journalism • Michael Krisch, Brown Institute, School of Journalism

Aims: This collaboration between the Brown Institute for Media Innovation at Journalism and the Center for Spatial Research at GSAPP aims to introduce mapping and spatial data analysis into the journalism curriculum and journalistic modes of interpreting spatial data analysis into the GSAPP curriculum. As data has found its way into the journalism curriculum, so should spatial data analysis and visualization be taught as emerging tools for reporting. For architecture, urban design, and planning, products from this program serve to translate methodologies of journalism and provide narratives and navigation for reimagining the city as an environment for design.

Impacts: This project developed a course in mapping, spatial analysis, and journalistic techniques for the architecture school offered each spring. It has been taught 4 times so far. The course receives high reviews and GSAPP continues to provide funding beyond the Collaboratory funding period. In the School of Journalism, developing a course was not realistic due to the tight Master's degree timeline and required coursework. Therefore, the team developed a 4-5 session module that is offered each semester to 20 Master's students. This module is 15 hours of voluntary, no-credit training on nights-and-weekends, yet each module has been filled to capacity, reaching 40% of journalism students – a remarkable success. Veronica Penny, who audited ARCH4063A and now works at the New York Times, recently commented to I do not think I would be asked to make maps if I didn't have the skills I learned in that class. Here are my latest maps."

Students impacted to date: 232 + auditors

COLLABORATORY AT COLUMBIA UNIVERSITY

Course/Workshop modules launched:

- ARCH4063A, "Points Unknown: New frameworks for investigation and creative expression through mapping" taught by Michael Krisch, Grga Basic, and Juan Saldarriaga, Spring 2018, 2019, 2020, 2021 (average of 18 students per semester)
- Mapping and Spatial Analysis Module for Journalism students, taught by Michael Krisch, Marguerite Holloway, and Juan Saldarriaga, Fall 2017, Spring 2018, Fall 2018, Spring 2019, Fall 2019, Spring 2020, Spring 2021, Summer 2021 (average of 20 students per module)

Computational Literacy for Public Policy (Computing in Context: Public Policy)

Dean *Merit E. Janow, SIPA • Dan McIntyre, Associate Dean for Academic Affairs, SIPA • Adam Cannon, SEAS, Computer Science • Gregory Falco, SIPA*

Aims: The goal of this project is a new cross-disciplinary course, developed and co-taught by a Senior Lecturer in Machine Learning and an Adjunct Lecturer at SIPA that shows students how Computer Science can be applied to policy issues. Topics include data collection, algorithmic thinking, and writing code as well as communication strategies for describing the methods and results of computer science analysis to a non-technical audience. In addition to developing the new course, SIPA will launch complementary capstone consulting projects that will provide practical opportunities for students to apply their new knowledge to real world problems.

Impacts: This team developed a new SIPA course offered each Fall (5 times so far) that combines a demanding introduction to computer science and coding with applications for public policy. Policy problems investigated include troubleshooting sources of environmental pollution, evaluating the effectiveness of public housing policy or determining the impact that local financial markets have on international healthcare or education. In spring 2017 and spring 2018 they also launched a series of SIPA Capstone projects in partnership with the following clients: Thomson Reuters, Wood Mackenzie, US Dept of State, Moody's Investors Service, Alto Data Analytics, and UN Women. Students have gone on to win tech/data science awards, including a SIPA award for best capstone project, and first prize in the 2018 Atos IT Challenge for a project called "Behind The Meter" that uses AI and chatbots to provide automated demand response applications for connected objects, reducing costs and environmental footprint.

Students impacted to date: 250

Courses/Capstones launched:

- INAF6006U, "Computing in Context: Public Policy," taught by Adam Cannon, Gregory Falco (2016, 2017), and Scarlett Swerdlow (2018, 2019, 2020), offered Fall 2016 (26 students), Fall 2017 (36 students), Fall 2018 (40 students), Fall 2019 (59 students), and Fall 2020 (54 students).
- 5 SIPA Capstone projects, Spring 2017 (21 students) and 2018 (14 students)

COLLABORATORY AT COLUMBIA UNIVERSITY

Programming, Technology and Analytics Curriculum for Columbia Business School

Dean Costis Maglaras, Columbia Graduate Business School • Hardeep Johar, Engineering and Applied Science, Department of Industrial Engineering and Operational Research • Wei Ke, Columbia Graduate Business School

Aims: Less than 30% of the 1,300 incoming Columbia Business students have programming experience yet 30-50% are interested in careers that require technical knowledge in startups, financial services firms, or "traditional" corporations that strive to leverage data and analytics to innovate business practices. The goal of this project is to develop a new set of courses and rich curriculum around technology, programming, and analytics for MBA students.

Impacts: Spearheaded by the dean of the Business School, this Collaboratory project rapidly jumpstarted the process of embedding data science throughout the curriculum. By 2018 this team had launched a slate of 9 reoccurring courses that reach approximately 1000 students each academic year. These courses include foundational courses in digital literacy, python programming, databases, and business analytics, as well as more advanced and case-specific courses in sports analytics, revenue analytics, and web app programming. These courses put Columbia's Business School at the forefront as a leader in data science education for future business leaders. This project's impact is showcased in the Collaboratory promotional video.

Students impacted to date: 3500+

Courses launched (see appendix for further details):

- Digital Literacy for Decision Makers (avg. 65 students/course)
- Web App Programming in Python (avg. 44 students/course)
- Introduction to Programming Using Python (avg. 78 students/course)
- Intro to Databases for Business Analytics (avg. 67 students/course)
- Sports Analytics (avg. 39 students/course)
- Business Analytics II (avg. 44 students/course)
- Analytics in Action: Master Class (avg. 41 students/course)
- Quantitative Pricing & Revenue Analytics (avg. 40 students/course)
- Data Analytics in Python (avg. 31 student/course)

COLLABORATORY AT COLUMBIA UNIVERSITY

Data: Past, Present, and Future

Chris Wiggins, Applied Physics and Applied Mathematics, Chief Data Scientist at New York Times
Matthew L. Jones, A&S, History

Aims: This cross-disciplinary course brings together undergraduates from SEAS and Columbia College to understand our civilization's relationship with data in order to become critical and effective participants in it. The civilization of our day is one saturated, improved, manipulated, and transformed through the accumulation and analysis of data. The course will combine traditional 'functional literacy' materials, those needed to explore and model data, with material sharpening critical and rhetorical literacy. With no prerequisites, the course promotes understanding the context and assumptions underlying data-driven experiences and narratives, as well as how to integrate analysis and modeling of data as part of contemporary discourse.

Impacts: This course is offered each spring, 4 times so far. During the development of this course, the Fellows' breakthrough realization was that data is a story of "truth and power." They discovered that the best way to highlight how truth and power impact data collection, analysis, and decision making was by focusing on historical case studies. This allowed them to truly integrate ethics into their data science curriculum, rather than appending it. They were also able to scale up the class from a small seminar to a lecture/lab course of 50-80 students using tools such as Python notebooks and Slack. They made their syllabus and course materials open source, and Chris Wiggins presented a lecture series about the course content at Princeton, Harvard, and Berkeley, so other educators may be implementing aspects of their work. The two Fellows also have a contract to write a book that grew out of the course.

Students impacted to date: 260

Course launched:

- HSAM2901W, "Data: Past, Present and Future," taught by Chris Wiggins and Matthew L. Jones, Spring 2017 (17 students), Spring 2018 (24 students), Spring 2019 (73 students), Spring 2020 (90 students), and Spring 2021 (56 students)

Interpreting Urban Environmental Data: New York City's changing landscapes

Zoe Crossland, Graduate School of Arts and Science, Department of Anthropology • Dorothy Peteet, Graduate School of Arts and Science, Earth & Environmental Science • Nan Rothschild, Lamont-Doherty Earth Observatory, Earth Institute • Jonathan Nichols, Lamont-Doherty Earth Observatory, Earth Institute

Aims: The goal of this project is to offer historically informed training in how to understand the source data that informs the public about human impact on past environments, including evidence for natural variability, landscape change, and the record of unregulated pollutants. In our polarized political world, models of climate change are contested and suspicion of scientific expertise is becoming common. Thoughtful and critical thinkers must understand how data are managed, stored, disseminated, and incorporated into various narratives of climate change. The course is designed to take students from hands-on environmental and archaeological sampling and interpretation through to the analysis of big palaeoecological datasets in large data archives.

Impacts: This project developed a course integrating two very different fields: paleoecology and historical archaeology. The course is cross-registered in the Department of Anthropology and the Department of Earth and Environmental Science. The project created new lines of interaction between Lamont-Doherty Earth Observatory and the main campus, including new interdepartmental relationships between both faculty and students. This led to collaborations not only in teaching but also in research. For example, one MFA student drew on the class in her final project, while other students developed research projects with Collaboratory Fellows that extended beyond the course. These Fellows also started a weekly volunteer lab that builds on the skills learned in the course and continues to offer research opportunities to students. An exhibition of research posters from the course was advertised widely, with good attendance from Anthropology and DEES. The Fellows plan to offer the course on a bi-annual schedule, but had to postpone the Fall 2020 iteration due to the incompatibility of the hands-on nature of the course with COVID-19 realities.

Students impacted to date: 8

Course launched:

- ANEE4522GU/W, "Environmental Histories of New York City," taught by Zoe Crossland, Nan Rothschild, Dorothy Peteet, and Jon Nichols, Fall 2018 (8 students)

DDS Squared: Digest of Data Science (DDS) for Doctors of Dental Surgery (DDS)

Letty Moss-Salentjin, Dental Medicine • Joseph Finkelstein, College of Dental Medicine, Center for Bioinformatics and Data Analytics in Oral Health • Ying Wei, Mailman School of Public Health, Department of Biostatistics

Aims: Though dental care practitioners are pivotal in advancing continuous knowledge discovery and improving quality of dental care delivery, comprehensive use of data science methodologies is still limited in dentistry. Thus, it is critical to equip dental professionals with hands-on understanding of major foundations of data science. Engaging dental students in data science projects early in their career will help create a new generation of oral health professionals fully equipped to lead precision dentistry initiatives. The goal of this project is to address this gap by developing a learner-centered introductory data science course fully integrated with the dental school curriculum.

Impacts: This project led to the development of an introductory online course in data management. In order to earn credit for the course, completion of a summer practicum or boot camp on a large data set was also required. The primary PI on this project left Columbia after the first iteration of the course, and the project has not continued nor the course offered subsequently. There is interest from leadership at the College of Dental Medicine to restart the project should an interested faculty member who can lead the project be identified.

Students impacted to date: 6

Course launched:

- TECH4500N, "Data Science for Clinicians," taught by Joseph Finkelstein, Fall 2018 (6 students)

COLLABORATORY AT COLUMBIA UNIVERSITY

The Search for Meaning in Big Data: Patterns, Representation and Empathy

Benjamin K. Holtzman, Lamont-Doherty Earth Observatory • Miya J. Masaoka, Sound Arts MFA program, Computer Music Center • John Paisley, Engineering and Applied Science, Data Science Institute • Arthur Paté, Lamont-Doherty Earth Observatory, Earth Institute

Aims: The goals of this project include a course (and concomitant tools and techniques) that would bridge the creative world of artistic expression with the rich data-world of scientific inquiry. The focus is on sonic and visual representation of complex data, the development of tools such as neural networks for revealing features in large data sets and artistic perspectives on the representation of these features in contemporary visual art and music. The course will be augmented by a "Data and Art Visiting Artist Lecture Series" to provide necessary perspectives on the work we develop in the class.

Impacts: Beyond developing a course that is offered each spring (4 times so far), this team developed a complementary "Data and Art Visiting Lecture Series" held in 2018, 2019, and 2020 with top practitioners bridging these fields. Using Collaboratory funds, Fellows also built a 32-channel Spatial Sound Lab at the Computer Music Center, coupled to virtual reality systems for data exploration and artistic purposes. Two students' class project grew into a film that won first prize in the American Geophysical Union's "Data Visualization and Storytelling" contest, while a paper on the methods developed for the project was submitted to MIT Press's Computer Music Journal. This project combines sonification and visualization of data from a geyser at Yellowstone National Park to convey the dynamics of geyser eruption cycles. This film and another student's film about hurricanes are frequently used to teach Seismology. Ben Holtzman plans to pull together class notes and programming methods from the course into a book.

Students impacted to date: approximately 36

Course launched:

- MUSI6602GR, "Sonic and Visual Representation of Data," taught by Ben Holtzman with guest lectures from John Paisley, Spring 2018, Spring 2019, Spring 2020, Spring 2021 (average of 9 per course)

COLLABORATORY AT COLUMBIA UNIVERSITY

In Vivo Magnetic Resonance Spectroscopy - from Data to Clinical Benefit

Christoph Juchem, Engineering and Applied Science, Depts. of Biomedical Engineering and Radiology • Lawrence S. Kegeles, CUIMC, Depts. of Psychiatry and Radiology

Aims: The goal of this project is to establish a course that comprises all aspects of in vivo magnetic resonance spectroscopy (MRS) from theory to experiment, from data acquisition to the derivation of metabolic signatures, and from study design to clinical interpretation. MRS allows the detection of chemical compounds from localized regions in living tissue, e.g., the brain, in a noninvasive fashion. The repertoire of measurable compounds along with the quantitative character of the derived information makes MRS a versatile tool for the identification of clinical conditions and for treatment control and monitoring of virtually all disorders with a metabolic signature. Before this project, Columbia did not offer a course on the topic.

Impacts: This project developed a course that is offered each fall (3 times so far), and also developed a spectroscopy workshop for biomedical engineering students for whom a full course would be unfeasible. The course brings engineers and physicists together with clinical researchers and practitioners. The course includes classroom teaching, a computer lab where students process data, and practical experience working with an MR scanner including subject handling, safety protocols, and IRB. Giving students experience with this machine, which costs $600 per hour to access, is a unique, hands-on learning experience for those who are interested in pursuing this field. This course serves undergraduate and graduate students, as well as faculty-level auditors, a number of whom have used the training to strengthen their research. To further incentivize participation of medical students, this team is working to establish CME credits for members of CUIMC.

Students impacted to date: 50
Course launched:

- BMEN E600, "In Vivo Magnetic Resonance Spectroscopy – from Data to Clinical Benefit," taught by Lawrence Kegeles and Christoph Juchem, Fall 2018 (21 students) was subsequently registered as BMEN E6410, "Principles and Practices of In Vivo Magnetic Resonance Spectroscopy," Fall 2019 (12 students + auditors), Fall 2020 (17 students)

Data Science for Social Good -- A Collaborative Course for Social Work and Data Science Students

Desmond U. Patton, School of Social Work, Data Science Institute • Tian Zheng, SEAS, Department of Statistics, Data Science Institute • Tara Batista, School of Social Work

Aims: The goal of this project is to develop a course designed to generate a new cadre of interdisciplinary social work students who will learn to utilize computational skills and deep contextual knowledge of the conditions and factors related to our most pressing social problems. Addressing social problems requires multifaceted solutions driven by research, experience, and an adaptive understanding of factors affecting people, institutions, communities, and systems. This course intends to pave a technical foundation for social work students to develop a greater appreciation of emerging technologies for social interventions.

COLLABORATORY AT COLUMBIA UNIVERSITY

Data Science for Social Good (continued)

Impacts: This course launched in Fall 2021 as a 4-credit course which included 7 School of Social Work students who also completed a python programming lab that teaches statistical analysis, worth 1 credit. The social work students in the first cohort shared very positive feedback on the course and the lab component. Based on word-of-mouth recommendations, during its second iteration in Spring 2021 the course was oversubscribed and was able to serve a total cohort of 43 students. The course will continue to be offered each semester.

Students impacted to date: 72

Course launched:

- UN1010, "Statistical Thinking for Data Science, Stat Thinking w. Python Labs," taught by Anthony Donoghue, Fall 2020 (29 students), Spring 2021 (43 students)

Neurogenomics

Rene Hen, Vagelos School of Physicians & Surgeons, Department of Neuroscience • Sergey Kalachikov, Engineering and Applied Science, Center for Genome Technology & Bimolecular Engineering • Irina Morozova, Computer Lab Instructor

Aims: The goal of this project is to create a new graduate-level 'Neurogenomics' course that integrates Neuroscience, Genomics, and Data Science. Neuroscience brings massive amounts of data into traditional scientific fields demanding a better understanding of Big Data from new experimentalists entering the field. The course will teach appreciation for the scientific rigor of data-driven research, practical skills needed to analyze data, and the interplay between the experimental design and data analysis techniques that are appropriate to the research question. Course topics include planning and analytical efforts of genomic experimentation, regulatory and ethical aspects of Big Data analysis, as well as evaluating the results, their limitations, and internal and external validity of research.

Impacts: This course, offered twice so far, serves graduate students in Biology, particularly those in Neuroscience and MD/PhD programs. The course utilizes existing analytical pipelines and creates new ones using RStudio, RNotebooks, and Github. The course meets students' identified need for a solid background in computational and analytical skills as well as reproducible analysis. The Fellows discovered students were very quickly able to apply knowledge obtained from the course to their own research. Students have gone on to use new skills in their labs, to produce research results used in presentations and posters, and to pass interviews for scholarships or to participate in external, advanced analytical courses with competitive enrollment such as the Cold Spring Harbor summer workshops. Inspired by the course, several students plan to take additional courses in statistics, algebra, and data science.

Students impacted to date: 33

Course launched:

- NBHV6050GR, "Neurogenomics," taught by Sergey Kalachikov, Rene Hen, and Irina Morozova, Fall 2019 (15 students), Fall 2020 (18 students)

COLLABORATORY AT COLUMBIA UNIVERSITY

Multilingual Technologies and Language Diversity: Innovations & Implications

Lydia H. Liu, Arts and Sciences, Institute for Comparative Literature and Society • Smaranda Muresan, Engineering and Applied Science, Computer Science • Isabelle Zaugg, Engineering and Applied Science, Data Science Institute, and Arts & Sciences, Institute for Comparative Literature and Society

Aims: Innovations in digital technologies have shown potential for both beneficial impact and troubling results. Evidence suggests that new technologies contribute to the predicted extinction of 50-90% of the world's languages this century. The objective of this project is to introduce a dual view on language diversity: 1) typology of language and 2) a resource-centric typology (low-resource vs. high-resource) regarding the development of computational models for language analysis. The course addresses the challenge of scaling natural language processing technologies developed mostly for English to the rich diversity of human languages.

Impacts: This course is offered in the spring and has been taught twice so far. The course is cross-registered between the Department of Computer Science (CS) and the Institute for Comparative Literature and Society (ICLS) and serves a mixed cohort of CS, humanities, and social science students. Students include upper-level undergraduates and graduate students. The course has received high reviews: 100% of survey respondents in 2020 stated they would recommend it to others. The team recently developed a course website in order to open-source the syllabus and showcase students' final projects. Top student essays from the course are published on the ICLS "Explorations in Language Justice" Blog. Recently, two humanities students were among the winners of the DSI Best Student Data Science Course Project 2021, awarded for their final project "The Columbia Language Justice Perspectives Project." Demonstrating this project's success in opening the field of data science to students from other disciplinary backgrounds, one of these students credits the course with inspiring her to complement her ICLS major with a second major in CS.

Students impacted to date: 58

Course launched:

- CPLS4315GU/C and COMS4995W, Sec. 7, "Topics in Computer Science (Multilingual Language Technologies and Language Diversity)", taught by Smaranda Muresan and Isabelle Zaugg, Spring 2020 (26 students), Spring 2021 (32 students)

COLLABORATORY AT COLUMBIA UNIVERSITY

Integrating Data Science into Environmental Health Sciences' Curricula

Andrea Baccarelli, Public Health, Dept. of Environmental Health Sciences • Jeff Goldsmith, Public Health, Dept. of Biostatistics • Nina Kulacki, Public Health, Dept. of Environmental Health Sciences • Tiffany Sanchez, Public Health, Dept. of Environmental Health Sciences

Aims: Environmental Health Sciences (EHS) examines how environmental exposures and climate change affect human health and disease. Compared to data sets typically generated for study in other disciplines, EHS data are infinitely complex and foundational to EHS research. It is the goal of this project to provide the next generation of EHS leaders with an ingrained understanding of data science and a higher level of quantitative fluency. Biostatistics faculty will develop a new intensive course, "Introduction to Data Science for Environmental Health" as well as a Master's-level program in Environmental Health Data Sciences.

Impacts: This project provided an impetus for the Department of Environmental Health Sciences to begin a focused conversation about the centrality of data science to the future of the field. The Fellows credit the Collaboratory with jumpstarting this process that would otherwise have taken an additional 5 years. The course launched, "Introduction to Data Science for Environmental Health," taught twice so far, will be a required course for students in the new Environmental Health Data Sciences Master's program. Finalizing the Master's program will require approval at the state level, and remains in development stages.

Students impacted to date: 55

Course launched:

- EHSC8321P, "Introduction to Data Science for Environmental Health," taught by Tiffany Sanchez with guest lectures from 5 professors, Spring 2020 (29 students), Spring 2021 (26 students)

Introduction to NYC Health Disparities Using Data Science

Mary Beth Terry, Public Health, Department of Epidemiology • Abigail Greenleaf, Public Health, Dept of Population, Family, and Reproductive Health • Samantha Garbers, Public Health, Dept. of Population, Family and Reproductive Health • Dana March, Public Health, Department of Epidemiology

Aims: The objective of the project is to use an experiential, place-based approach to teaching undergraduate students about health disparities in New York City. The course introduces a methodical approach to data science: writing a research question, obtaining data, data cleaning, data exploration, analysis, replication and validity evaluation, and finally presentation and summary. Intentionally accessible to students without foundational data science skills, the curriculum is aimed at groups traditionally under-represented in STEM. The course will also introduce students to the R programming language.

COLLABORATORY AT COLUMBIA UNIVERSITY

Introduction to NYC Health Disparities Using Data Science (continued)

Impacts: One of only a handful of public health courses for undergraduates, and launched for the first time at just ahead of the COVID-19 pandemic amid growing awareness of racial inequities in public services and health outcomes, the course could not be timelier. It hit its cap of 40 students in its first iteration, and received positive feedback in student evaluations. It is slated to be offered each Fall semester.

Students impacted to date: 40

Course launched:

- PUBH3400UN/PUBH3400C, "Data Science and Health Equity in New York City," taught by Abigail Greenleaf and Mary Beth Terry, Fall 2020 (40 students)

Building the Next Gen of Cognitive Neuroscientists: A Suite of Interdisciplinary Brain Imaging Courses

Alfredo Spagna, Arts & Sciences, Psychology • Xiaofu He, Vagelos School of Physicians & Surgeon, Data Science Institute • Lila Davachi, Arts and Sciences, Department of Psychology • Nikolaus Kriegeskorte, Arts and Sciences, Dept. of Psychology, Dept. of Neuroscience, Zuckerman Mind Brain Behavior Institute, and Vagelos School of Physicians & Surgeons • Paul Sajda, Engineering and Applied Science, Data Science Institute • Chris Baldassano, SEAS, Dept. of Psychology • Agnes Chang, Engineering and Applied Science, Dept of Computer Science

Aims: This project aims to educate the next generation of neuroscientists in the use of non-invasive neuroimaging techniques. Recent insights into the inner workings of the mind are supported by the development of refined statistical and computational models. Analytical approaches from data science have already had a significant impact in multiple subfields of cognitive neuroscience. Three courses are proposed: an undergraduate-level seminar followed by two graduate-level courses. Students gain an understanding of different techniques for measuring brain activity, dealing with handling big data sets, and how analysis pipelines can be verified, reproduced, and shared.

Impacts: Addressing cognitive neuroscience training's failure to keep pace with the increasing availability and scale of public data as well as the complexity of cutting-edge analysis pipelines, this project successfully launched its three courses in 2019 and 2020. The first course launched, "Tools for Reproducible & Collaborative Science," targets graduate students and will be offered every 2 years. Since the final project of the course involved applying new technical tools to the graduate students' ongoing research projects, many are now using these tools directly in their research and publications. Student evaluations for the course were excellent, rating it 4.5/5. "Tools for Reproducible & Collaborative Science," which is cross-listed in Pyschology and Neuroscience, also targets graduate students and will be offered every 2 years. "Fundamentals of Human Neuroimaging" targets undergraduates and will be offered every Fall.

Students impacted to date: 39

Courses launched:

- PSYC6060GR/G, "Human Brain Imaging for Cognitive Neuroscience: Data Acquisition, Analysis, and Sharing," taught by Nikolaus Kriegeskorte, Lila Davachi, Ruben van Bergen, Nir Yacoby, Spring 2020 (18 students)
- PSYCGU4930, "Fundamentals of Human Neuroimaging," Alfredo Spagna and Xiaofu He, Fall 2020 (5 students)
- GR6130, "Tools for Reproducible & Collaborative Science," Chris Baldassano & Agnes Chang, Fall 2019, 11 students

COLLABORATORY AT COLUMBIA UNIVERSITY

Data Science for Better Health Outcomes: A Nursing Perspective

Maxim Topaz, Nursing, Data Science Institute • Kathleen Mullen, School of Nursing • Kenrick Kato, Nursing

Aims: The objective of this project is to develop a course that exposes nursing students to data science. In the U.S., demand for nurses keeps rising while the profession is on the cusp of technological change. Omnipresent health IT (such as EHR – electronic health records) requires nurses to think differently about data and methods of data collection. Tailored to nursing, course topics include fundamental data science technologies (e.g., machine learning and text mining), discussion of ethical aspects of data science, and a hands-on data science project in collaboration with Columbia University Data Science Institute students.

Impacts: "Data Science for Better Health Outcomes, A Nursing Perspective," will be launched in the Summer B session, 2012, and offered every summer thereafter.

Course to be launched:

- NURS7150N, "Data Science for Better Health Outcomes: A Nursing Perspective," taught by Kathleen Mullen, Kenrick D. Cato, and Maxim Topaz, Summer B 2021 (7 students enrolled so far)

Programming, Analytics and Technology Curriculum for the GSAPP Real Estate Development Program

Patrice Derrington, Architecture Persevering and Planning, Real Estate Development • Hardeep Johar, Engineering and Applied Science, Industrial Engineering and Operations Research

Aims: Approximately 125 Real Estate Development students graduate annually and take up positions in investment banking, securities analysis, private equity, real estate development, and urban projects. These professions require a level of technical fluency that GSAPP has not previously offered. The objective of this project is to advance "PropTech" learning in the Real Estate program by exposing all students to technology and programming foundations, and offering a set of electives that explore these analytical tools. The courses will be designed to proceed in tandem with detailed instruction in the related domain content, initially with key courses and then followed by "lab" courses around specific applications or methodologies.

Impacts: Three courses were piloted this academic year, serving graduate students studying Urban Planning and Real Estate Development. "Intro to Python for Machine Learning" and "Machine Learning in Python" are 1.5 credit courses taken in tandem across the course of one semester. The goal of these courses is to develop the ability to use Python to extract knowledge and information from data. The course "Data Analytics for Real Estate," brings these skills into the specific realm of real estate.

Students impacted to date: 30

Courses launched:

- PLANA6845, "Intro to Python for Machine Learning," taught by Uday Menon, Fall A 2020 (19 students and 5 auditors)
- PLANA6846, "Machine Learning in Python," taught by Uday Menon, Fall B 2020 (same 19 students and 5 auditors)
- PLANA6868, "Data Analytics for Real Estate," taught by Uday Menon, Spring 2021 (6 students)

COLLABORATORY AT COLUMBIA UNIVERSITY

Data-Driven Decision-Making Modeling and Analytics

Yi Zhang, Engineering and Applied Science, Industrial Engineering and Operations Research
• Tony Dear, Engineering and Applied Science, Dept. of Computer Science

Aims: The objective of this project is to create an upper-level interdisciplinary course covering decision-making theory and associated real-world applications of artificial intelligence and business analytics. This addresses a crucial need in both the IEOR and CS departments for students to study and get practice in methods that are quickly becoming pervasive in industry. Students receive methodical training in decision-making theory, dynamic programming, reinforcement learning, and modeling estimation. Then students explore how these theories help data analysts model complicated decision-making processes and derive business values.

Impacts: This project is in the course development stage.

Neural Networks: Computational and Philosophical Perspectives

John Morrison, Barnard College, Dept. of Philosophy •
Christos Papadimitriou, Engineering and Applied Science, Dept. of Computer Science

Aims: Increasingly artificial neural networks make important decisions for us, but our understanding of them is fundamentally lacking and we typically don't know the logic behind their decisions. The objective of this course is to explore these issues from philosophical and computational perspectives and to provide students with the background necessary to engage with one of the most pressing issues in both machine learning and systems neuroscience.

Impacts: This project is in the course development stage.

Accessible and Inclusive Data Capture and Display: Systems for Multi-Sensory Data Engagement

Seth Cluett, Arts & Sciences, Dept. of Music, Computer Music Center • Mark Santolucito, Barnard, Computer Science • Brad Garton, Arts and Sciences, Dept. of Music • Benjamin K. Holtzman, Arts and Sciences, Dept. of Music, Computer Music Center, Lamont-Doherty Earth Observatory • Miya J. Masaoka, School of the Arts, Visual Arts

Aims: This project aims to launch a new upper-level computer science course at Barnard, and to redevelop existing courses from the Computer Music Center at the intersection of technology and art. These will form a cluster of courses that catalyze a community of practice in data science and art. The new course, “Creative Embedded Systems,” will teach design and development of hardware interfaces for data collection, presentation, and documentation, pushing students to consider the impact of their work on the community beyond the university’s walls. This project builds upon the Collaboratory project “The Search for Meaning in Big Data: Patterns, Representation, and Empathy” and engages the pre-existing courses “Instrument, Installation, Interface: Building Sound,” “Sonic and Visual Representation of Data,” and “Sound: Advanced Topics II (Virtual Reality).”

COLLABORATORY AT COLUMBIA UNIVERSITY

Accessible and Inclusive Data Capture and Display (continued)

Impacts: "Creative Embedded Systems" was piloted in Spring 2021, serving undergraduate CS majors at Barnard. It will be offered annually. The course syllabus and materials are available open source. Team members are working on publishing code libraries for data sonification in Python as well as videos of video artist lectures in a shared Lab/Project webspace. The new course and cluster were reported on in a recent Columbia Spectator article.

Students impacted to date: 23

Courses launched:

- COMS3930BC/X, "Creative Embedded Systems," taught by Mark Santolucito, Spring 2021 (23 students)

Interrogating Justice and Ethics in Digital Health

Noemie Elhadad, VP&S, Biomedical Informatics
Sandra Soo-Jin Lee, VP&S, Department of Medical Humanities and Ethics

Aims: The objective of this multi-disciplinary course includes teaching students to situate data technologies within their socio-political contexts and to examine the social life of data and its impact on society. Engaging with data has become a civic requirement. Technical expertise must include engagement with the ethical issues and policy implications related to emerging data-driven techniques. The biomedical, health, and clinical domains are going through in-depth changes as AI and data-driven thinking become inherent to routine processes.

Impacts: This project is in the course development stage.