

**Joining the Data Revolution:  
Big Data in Education and Social Science Research**  
June 15-19, 2020 | Ann Arbor, Michigan  
Institute for Social Research, Room 6080

### **Course Description**

This is an introductory-level course aimed at researchers in education and social science fields who are interested in gaining experience working with large-scale restricted-use data. No prior experience in quantitative data analysis is required. Datasets on topics including post-secondary educational opportunities, employment outcomes, and the makeup of research teams will be featured in the workshop.

In this hands-on class, participants will work in teams using large-scale datasets with the goal of achieving a better understanding of the research questions that can be answered with big data. Morning technical sessions will focus on working through well-documented examples of python code in Jupyter Notebook designed to address a shared research question using [UMETRICS](#) data. This session will focus on examining whether and how different types of diversity in scientific research teams influence the amount, character, and impact of research produced by those teams. This guided, step-by-step analytic work will help you develop skills in working with real administrative data in a privacy protected research environment. More substantive discussions during introductory lectures and working lunches will focus on literature and general approaches to questions that you will help define.

The overall goal of this work is to help you acquire or develop skills to support your own efforts to develop and articulate an individual research question and to frame that question in a fashion that might be suitable for you pursue external funding to support your research goals. Later workshop sessions will focus more heavily on the grant submission and review process at the National Science Foundation (NSF) with a lesser focus on other federal and non-federal funding sources. In addition to technical skills, participants will work with established investigators to develop and present ideas for projects that might eventually be submitted to NSF grant competitions. While examples will be drawn from UMETRICS data and you are welcome to use those data to help develop your research questions, your topics and datasets may focus on other sources.

Supported by the NSF as part of its [Harnessing the Data Revolution “Big Idea.”](#) the IRIS workshop is designed to help investigators from a wide range of backgrounds and disciplines acquire the tools and knowledge to secure grant funding for data-driven social science and education research.

## Course Goals & Guiding Principles

The overall goal of this workshop is to help you develop technical and substantive skills in large scale data analysis that you can turn to when developing or expanding your own research agendas. This opportunity is funded by the Education and Human Resources (EHR) division of the National Science Foundation (NSF). It is designed to help researchers bring their expertise and experience in educational and social science (ESS) research to bear on questions of interest to them using data science tools and methods. We hope to help you increase your ability to define and develop projects that will result in competitive proposals that might be submitted to NSF or other funders. We also hope to help support a community of researchers who can share information, tools, and insights to help strengthen research and teaching capabilities involving large scale data analysis in ESS fields.

### Workshop Themes:

1. Data exploration
2. Data visualization
3. Data linkage
4. Basic data analysis
5. Grant proposal writing

### We believe that:

1. data science tools and methods are best learnt through concrete, hands on work to address real research questions with real data.
2. you bring essential expertise and knowledge to this workshop. We have as much to learn from you as you do from us.
3. effective classroom support is necessary to help you over the technical challenges that inevitably accompany new tools.
4. all research is, at one level or another, collaborative and is best accomplished within a engaged intellectual community.
5. data science teaching and learning, as well as the data sets and tools we develop and share, must be responsive to the needs and interests of diverse research communities.
6. data science skill consists primarily in figuring out how to make creative use of imperfect data that often comes from multiple sources to address important questions.

---

## Instructional Materials

### Readings:

- Cook, L. & Chaleampong, K. (2010). "The Idea Gap in Pink and Black." (NBER Working Paper No. 16331). <https://doi.org/10.3386/w16331>
- Foster, I., Ghani, R., Jarmin, R., Kreuter, F., & Lane, J.(Eds.) (2017) *Big Data and Social Science: A Practical Guide to Methods and Tools*. CRC Press.
- Healy, K., Moody, J. (2014). Data Visualization in Sociology. *Annual Review of Sociology*, 40(1), 105-128. <https://doi.org/10.1146/annurev-soc-071312-145551>
- Lane, J., Bender, S., Nissenbaum, H., & Stodden, V. (Eds.) (2014) *Privacy, Big Data,*

*and the Public Good: Frameworks for Engagement*. Cambridge University Press.

- Leahey, E. (2016). "From Solo Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration." *Annual Review of Sociology*, 42: 81-100. <https://doi.org/10.1146/annurev-soc-081715-074219>
- Leahey, E., Beckman, C. & Stanko, T. (2017). "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists' Research." *Administrative Science Quarterly*, 62:1 105-139. <https://doi.org/10.1177/0001839216665364>
- Page, S. (2014). "Where Diversity Comes From and Why it Matters?" *European Journal of Social Psychology* 44(4), 267- 279. <https://doi.org/10.1002/ejsp.2016>
- Reagans, R., & Zuckerman, E. (2001). Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. *Organization Science*, 12(4), 393-521. <https://doi.org/10.1287/orsc.12.4.502.10637>
- Reagans, R., E. Zuckerman, and B. McEvily. (2004). "How to make the team? Social networks vs. demography as criteria for designing effective teams." *Administrative Science Quarterly*, 49: 101-133. DOI: 10.2307/4131457
- Uzzi, B., S. Mukerjee, M. Stringer & B. Jones (2013). "Atypical Combinations and Scientific Impact." *Science* 342: 468-472. <https://doi.org/10.1126/science.1240474>
- Wisnioski, M., et al. (2019) *Does America Need More Innovators?* The MIT Press. <https://doi.org/10.7551/mitpress/11344.001.0001>

### **Teaching Dataset:**

Workshop participants will be provided with the teaching dataset to use during the course. The dataset is constructed for teaching purposes based on both publicly available data and restricted-access data that IRIS curates for research and training. The public component of the teaching dataset includes two data sources: 1) Post-Secondary Employment Outcomes (PSEO)—the student-level data (from 47 campuses) about earnings and employment outcomes for college and university graduates by degree level, degree major, and post-secondary institution, compiled by and made available from the U.S. Census Bureau; and 2) Integrated Postsecondary Education Data System (IPEDS)—the aggregated-level data on enrollment, admission, graduation rate, degree, institutional characteristics, and finance from postsecondary institutions. The restricted access data (available only via the IRIS virtual data enclave) are a subset of the 2020 IRIS UMETRICS annual data release that includes transaction level information on nearly 400,000 sponsored projects that represent approximately \$100 billion of direct cost expenditures and employ more than 700,000 people at 33 universities. These data are linked to information on scientific outcomes at multiple levels of analysis.

### **Data and Linking Asset Worksheet:**

An ER diagram will be generated and provided to the class. This will help participants understand linking assets, select relevant variables, and build their own dataset for analysis and visualization.

---

**Thank you for joining us in this workshop.  
We look forward to learning with and from you.**