INTRODUCTION TO BIG DATA FOR SOCIAL SCIENCE
A two-days short course sponsored by the Joint Program in Survey Methodology

SEPTEMBER 23-24, 2019
Presented at RTI, Washington DC.

FRAUKE KREUTER
Professor, University of Maryland, University of Mannheim, IAB Nuremberg

CHRISTOPH KERN
Post-Doctoral Researcher, University of Mannheim

COURSE ABSTRACT
Social scientists and survey researchers are increasingly faced with integrating data from multiple data sources and are expanding their activities beyond experiments and surveys. This course will discuss these developments and provide practical guidance on combining methods and tools from computer science, statistics, and social science. In particular, techniques such as web scraping and machine learning applications will be discussed, but also the linkage of various records, and privacy issues and questions around informed consent. The course presents the key big data tools in a non-intimidating way to social and data scientists while not neglecting research questions and purposes. The course aims to illustrate social science and data science principles through real-world applications, and links computer science concepts to real social science as well as survey research. Portions of the course are based on parts of a recent textbook with the same title. Data and code to practice will be available on GitHub.

WHO SHOULD ATTEND
The course is designed for individuals in government, universities, private sector, and nonprofit organizations who are interested in understanding the fundamental concepts applying new analytical techniques to social science issues. This course gives practical guidance to those who have never performed their own data analytics project but who are planning to do so, and those who have performed data analytics projects but would like to learn more the underlying methodology. The instructors will operate under the assumption that students have no prior knowledge of the topic.

THE INSTRUCTORS

FRAUKE KREUTER is Professor in the Joint Program in Survey Methodology at the University of Maryland, Professor of Methods and Statistics at the University of Mannheim, and head of the statistical methods group at the German Institute for Employment Research in Nuremberg. Previously, she held positions in the Department of Statistics at the University of California, Los Angeles, and the Department of Statistics at the Ludwig Maximillian's University of Munich. Prof. Kreuter serves on several advisory boards for National Statistical

![Short Course Information and Schedules — The Joint Program in Survey Methodology]

Institutes around the world, and within the Federal Statistical System in the United States. She is also a Gertrude Cox Award winner, which recognizes statisticians in early- to mid-career who have made significant breakthroughs in statistical practice, winner of the inaugural Links Lecture Award and elected fellow of the American Statistical Association. Additionally, she is co-founder of the Coleridge Initiative, founder of the International Program in Survey and Data Science, and co-host of the digitalization podcast www.digdeep.de

CHRISTOPH KERN is a Post-Doctoral Researcher at the Professorship for Statistics and Methodology at the University of Mannheim, and Visiting Assistant Professor in the Joint Program in Survey Methodology at the University of Maryland. Christoph Kern graduated in 2011 (Dipl.-Soz.-Wiss.) from the University of Duisburg-Essen (UDE), where he studied sociology with a focus on methods for empirical social science research. In 2016 he completed his PhD (Dr. rer. pol.) at UDE. His current research focuses on the usage of machine learning methods in survey research.

COURSE MATERIALS AND MEALS
Registrants will be provided with a course lecture notebook. JPSM group continental breakfasts, lunches and refreshments are included in the course fee.

DAILY CHECK-IN
Registrants must check-in with the JPSM onsite each day of the course.

MONDAY: SEPTEMBER 23, 2019

8:00--9:00	Registrant Check-in and Continental Breakfast
9:00--9:45	Introduction to Big Data
9:45--10:30	Big Data Toolbox
10:30--10:45	Morning Break
10:45--11:15	Web-scraping and APIs
11:15--12:00	*Practical session: Web-scraping*
12:00--1:15	Lunch
1:15--2:45	Big Data Visualization
2:45--3:00	Afternoon Break

3:00--3:45	Foundations of Machine Learning
3:45--4:30	*Practical session: Machine Learning I*
4:30	Adjourn

TUESDAY: SEPTEMBER 24, 2019

8:00--9:00	Registrant Check-in and Continental Breakfast
9:00--9:45	Machine Learning Methods
9:45--10:30	*Practical session: Machine Learning II*
10:30--10:45	Morning Break
10:45--12:00	Record Linkage
12:00--1:15	Lunch
1:15--2:45	Text Analysis
2:45--3:00	Afternoon Break
3:00--4:30	Privacy and Confidentiality
4:30	Adjourn