Teaching Survey and Data Science Outside the Regular Classroom

Frauke Kreuter

University of Maryland, U.of Mannheim

Atlanta 3/26/18
Project coordinators and funding

Laura & John Arnold Foundation

U.S. National Science Foundation

The project on which this report is based was funded by the Federal Ministry of Education and Research under the number [16OH22064]. Responsibility for the contents of this publication lies with the author.
AAPOR Report on Big Data

AAPOR Big Data Task Force
February 12, 2015

Prepared for AAPOR Council by the Task Force, with Task Force members including:

Lila Rogers, Co-Chair, Statistics Sweden
Freddie Bryant, Co-Chair, IPUMS at the U. of Maryland, U. of Minnesota & SUH
Marcella Berg, Stockholm University
Paul Barsky, ICF International
Paul Becker, Mathematica Policy Research
Cliff Lampe, School of Information at the University of Michigan
Julie Lott, American Institutes for Research
Carly O’Neil, Johnson Research Labs
Abi Ober, Hixon/Ober Group

Acknowledgement: We are grateful for comments, feedback and editorial help from Evan Ben-Porath, James McMillen, and the AAPOR council members.
1st Example – ASA DataFest
- Teams of 3-5 undergraduates
- Friday evening – Sunday afternoon
- One (unknown) data set
- Three winning categories:
  - Best insights
  - Best visualization
  - Best use of outside data

- Best educational experience ever !
2nd Example – Coleridge Initiative
COLERIDGE INITIATIVE

Building the capacity needed to accelerate the effective use of new data.

“Data, data everywhere, we have to stop and think

with apologies to the Rime of the Ancient Mariner”

UPCOMING TRAINING PROGRAMS

Spring 2018
Kansas City, MO
Application Closed

Summer 2018

Program directors Rayid Ghani, Frauke Kreuter, and Julia Lane are also co-editors of “Big Data and Social Science: A Practical Guide to Methods and Tools,” the textbook for the class.
Approach: hands-on with real microdata

- Data on ex-offenders, welfare recipients
- Data on housing and transportation
- Joined Up Datasets
- Trained Staff
- New Products
- New Networks
The first classes brought together ~40 agencies from city, state, county and federal agencies.
Understand how to collect data, and how data are generated through administrative and processes.

Learn how to formulate the research goal, and which data are best suited to achieve this goal.

Apply machine learning, text and network analysis.

Learn how to curate, manage and link complex data.

Learn how to communicate results, distribute and store your data; Ethics.

Learn how to curate, manage and link complex data.
Collaborative and Secure Environment

- Data Discovery
- Database Browser
- Software Version Control
- JupyterHub (Data Analysis)
- Source Code Editors
- Command Line
What our participants say about the program

"Love the Jupyter notebooks!! ... I love how the code snippets and explanations are set up in the Jupyter notebooks. The format of going through it individually and discussing questions/challenges in your group, with the experts available when needed, worked really well for my learning style."

I could see our agency benefiting potentially from something like this in that, as the system builds out and collects additional resources/datasets that impact criminal justice system practices, this may be an option for a place for us to look for the results of studies using evidence based practices.
3rd Example – International Program in Survey and Data Science
We are pleased to announce the launch of the International Program in Survey and Data Science (IPSDS). Fundamental changes in the nature of data, their availability, the way in which they are collected, integrated, and disseminated are a big challenge for all those working with designed data from surveys as well as organic data. IPSDS was developed in response to the increasing demand from researchers and practitioners for the appropriate methods and right tools to face these changes. We offer a multidisciplinary curriculum, world-class faculty, and a web-based learning environment that allows you to take courses from anywhere in the world.
Problem we tried to solve – In brief

- Allow for multidisciplinary curriculum
- Modularized – adapt to prior skills and work needs
- Relevant methods and tools
- Mix of faculty from academia and industry

Key elements:
- Flexible web-based learning environment
- Live (video) interaction with faculty and students
- Face-to-face networking meetings
Mix of asynchronous and synchronous formats

- Pre-recorded lectures (small video units)
- (Bi)weekly assignments
- Discussion forums

- Small virtual classrooms
- Weekly 50-minute discussions with instructor
- Obligatory component
Data Generating Process

- Fundamentals of Survey and Data Science: 3 credits/6 ECTS
- Data Collection: 3 credits/6 ECTS
- Record Linkage: 1 credit/2 ECTS
- Practical Tools for Sampling and Weighting: 3 credits/6 ECTS
- Applied Sampling: 3 credits/6 ECTS
- Experimental Design: 3 credits/6 ECTS

Data Curation/Storage

- Database Management: 3 credits/6 ECTS
- Data Munging I-III: 1 credit/2 ECTS each

Data Analysis

- GLM: 3 credits/6 ECTS
- Analysis of Complex Data: 3 credits/6 ECTS
- Propensity Score/Statistical Matching: 3 credits/6 ECTS
- Machine Learning I-III: 1 credit/2 ECTS each
- Text Analysis: 1 credit/2 ECTS

Research Question

- Fundamentals of Survey and Data Science: 3 credits/6 ECTS

Data Output/Access

- Ethics: 1 credit/2 ECTS
- Data Confidentiality and Statistical Disclosure Control: 2 credits/4 ECTS
- Visualization: 2 credits/4 ECTS

Research Question

- Ethics: 1 credit/2 ECTS
- Data Confidentiality and Statistical Disclosure Control: 2 credits/4 ECTS
- Visualization: 2 credits/4 ECTS
Lessons Learned

- Learning with application at hand is key
- Teams can quickly overcome shortcomings
- Modular approach much appreciated by working professionals
- Privacy and confidentiality very important

- Hardest to learn and hardest to teach: Asking the right question!
https://ww2.amstat.org/education/datafest/

http://coleridgeinitiative.org

http://survey-data-science.net/

fkreuter@umd.edu

Contact me if you want to host locally or become a partner