



International Program in Survey and Data Science

Frauke Kreuter

ICTOS 10, 07.07.2018



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:

Lilli Japec, Co-Chair, Statistics Sweden
Franke Kreuter, Co-Chair, JPSM at the U. of Maryland, U. of Mannheim & IAB
Marcus Berg, Stockholm University
Paul Biemer, RTI International
Paul Decker, Mathematica Policy Research
Cliff Lampe, School of Information at the University of Michigan
Julia Lane, American Institutes for Research
Cathy O'Neil, Johnson Research Labs
Abe Usher, HumanGeo Group

Acknowledgement: We are grateful for comments, feedback and editorial help from Eran Ben-Porath, Jason McMillan, and the AAPOR council members.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

REPORT

INNOVATIONS IN FEDERAL STATISTICS

Combining Data Sources While
Protecting Privacy

THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking

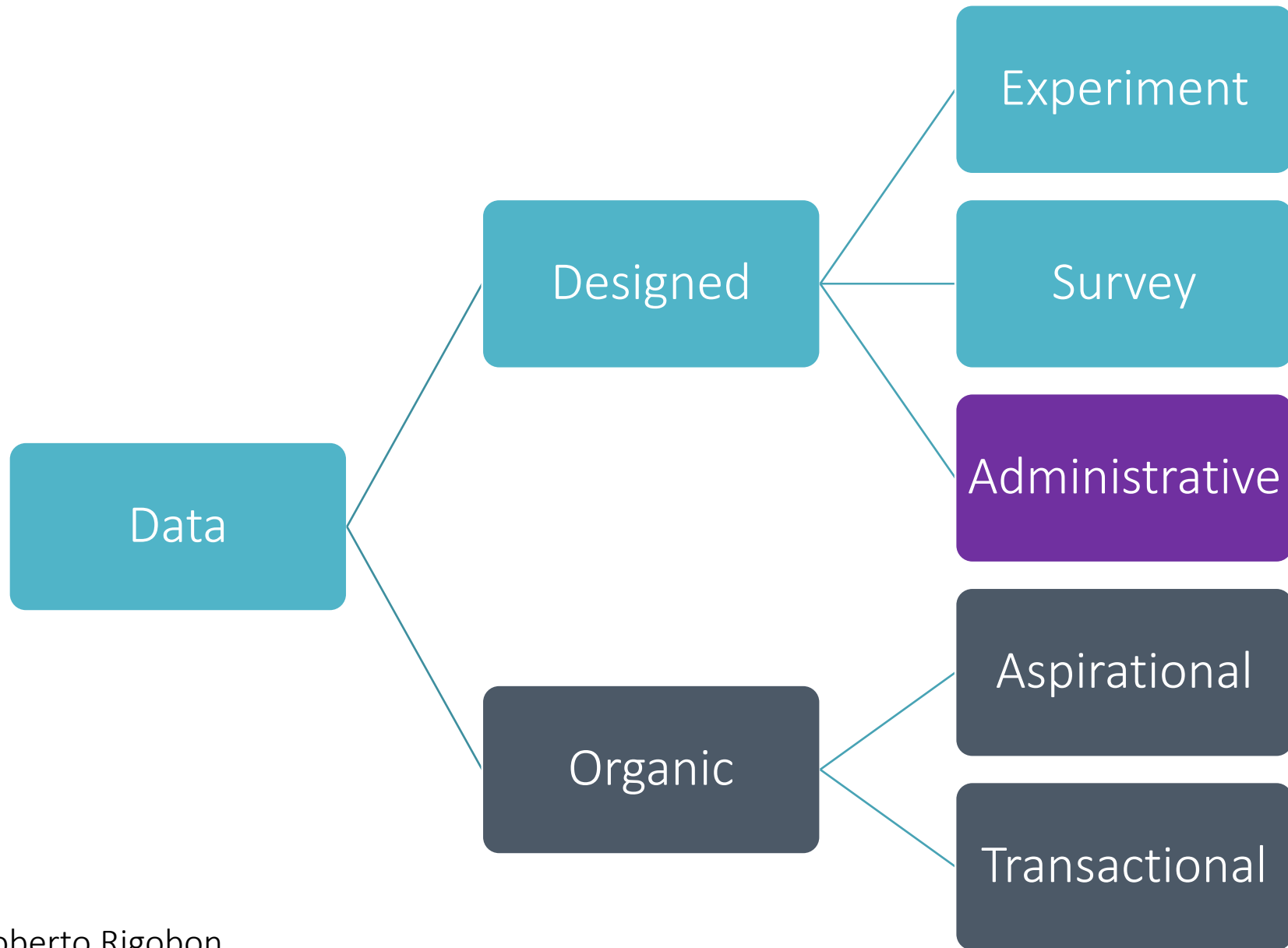


Economic Indicators

- Official CPI
- PriceStats Index



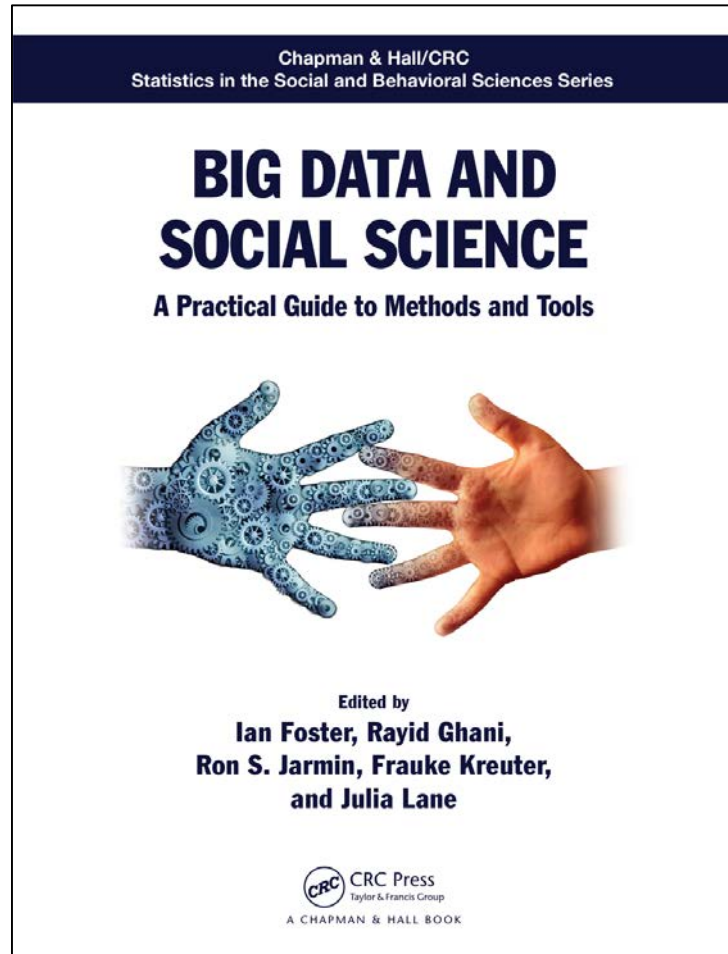
US Aggregated Inflation Series, Monthly Rate, PriceStats Index vs. Official CPI the PriceStats website. . 1/1/2015



Source: Roberto Rigobon

1st Example – Coleridge Initiative

Professional Training Workshops



Three Classes

- Different cohorts (ex-offenders, welfare recipients and veterans)
- Joined with housing, transportation and jobs data

Class Format

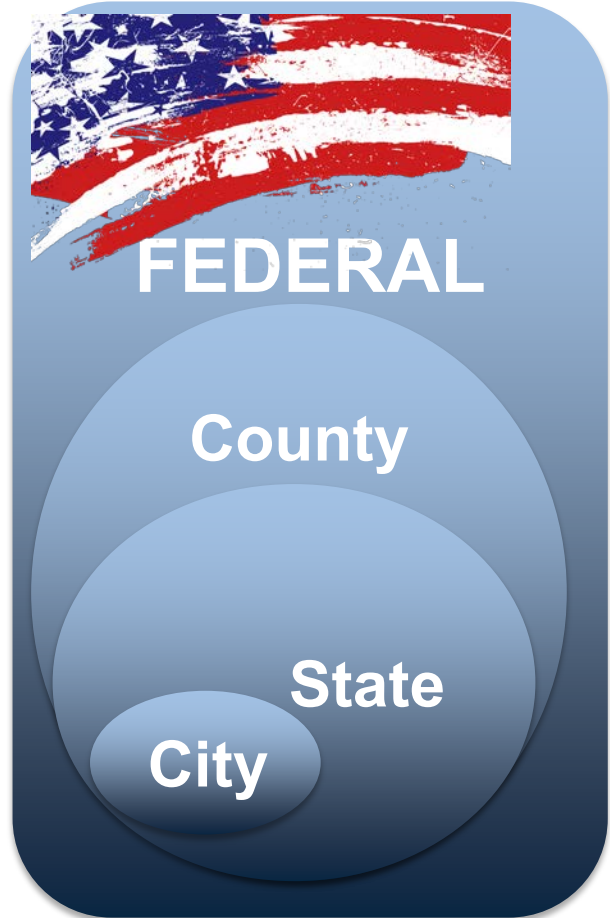
- Module 1: Foundations – Research Questions, Python, SQL
- Module 2: Data Acquisition – Web Scraping, API, Record Linkage
- Module 3: Data Analysis – Machine Learning, Networks, Text, Spatial
- Module 4: Visualization, Inference, Ethics, Privacy

Additional Information

- Final reports are all virtual
- Teaching Assistants and facilitators will be at each site for each module

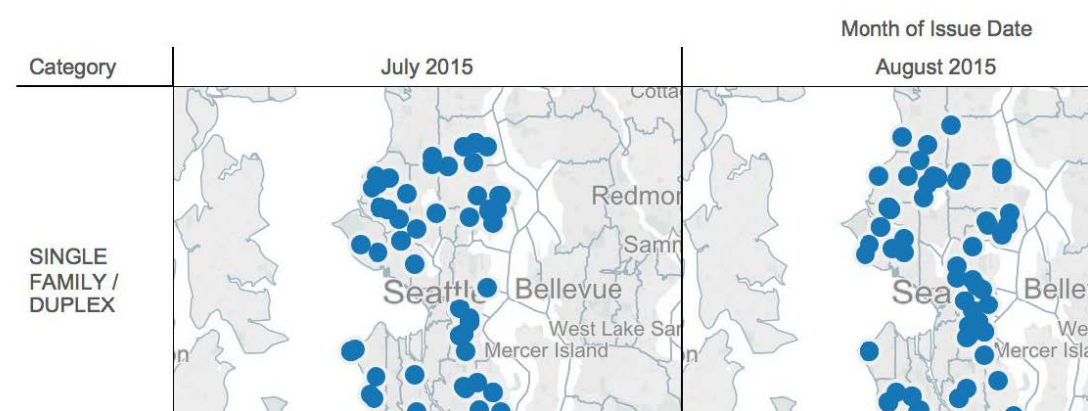
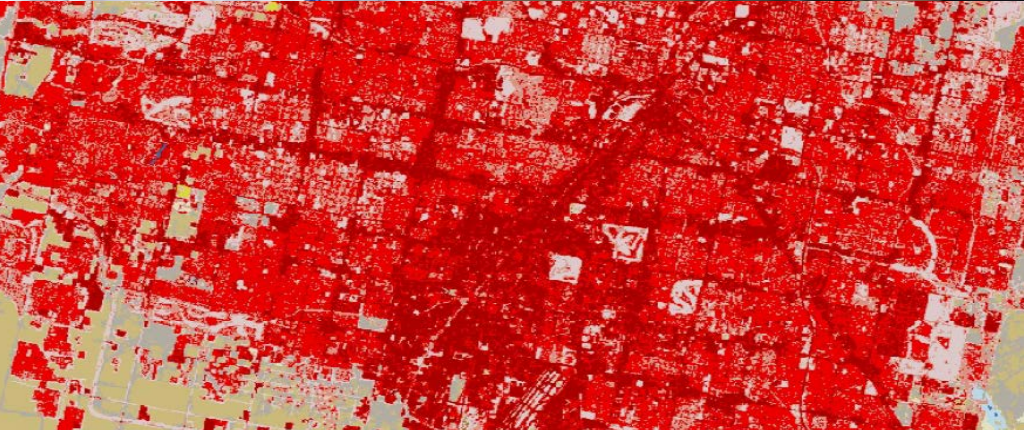
Networks: The first two classes brought together

~40 agencies from city, state, county and federal agencies

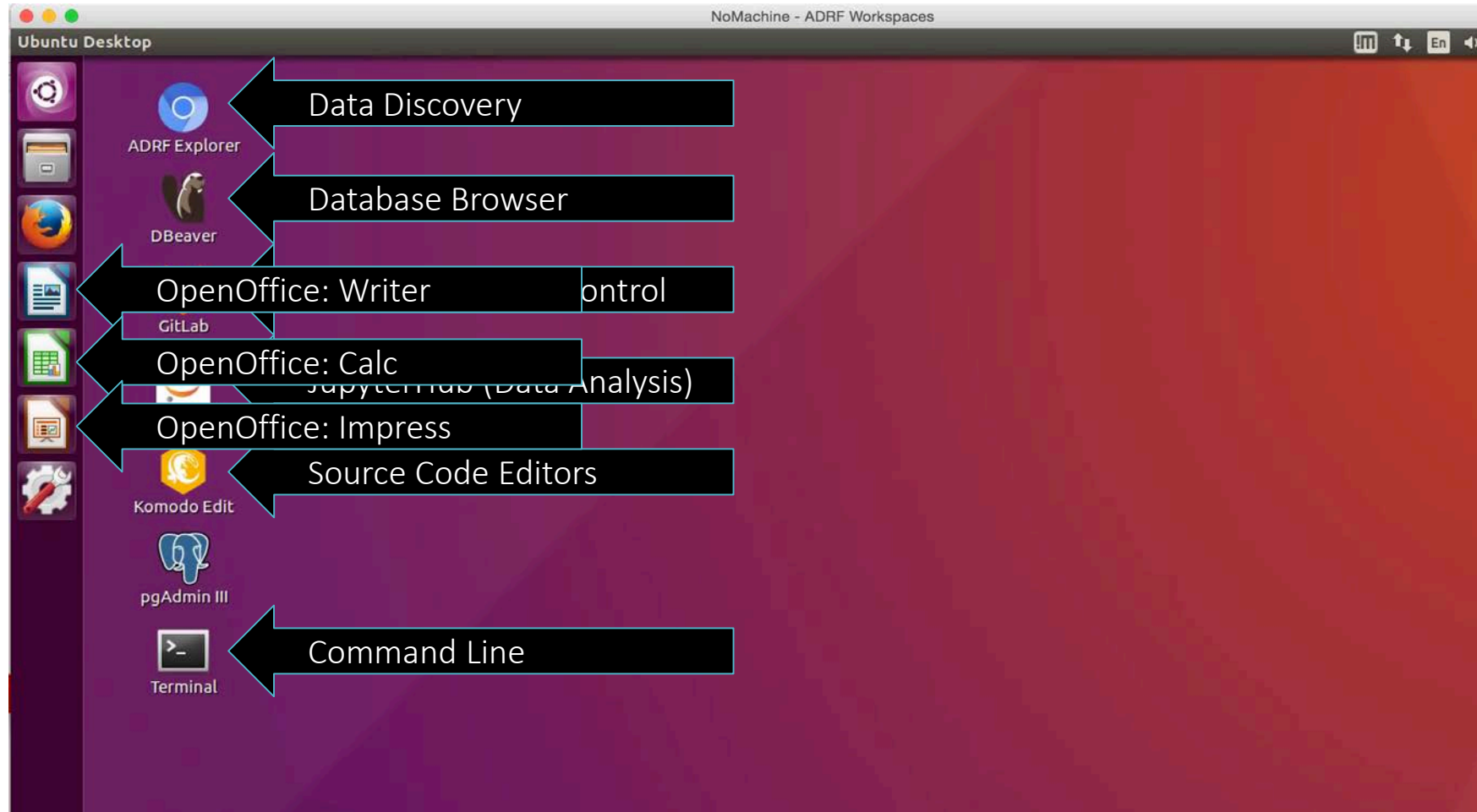


Approach





Collaborative secure environment



2nd Example –
International Program
in Survey and Data Science

[Home](#)[Program ▾](#)[Project ▾](#)[Contact](#)[FAQs](#)[Imprint](#)[Search](#)[APPLY NOW](#)

INTERNATIONAL PROGRAM IN SURVEY AND DATA SCIENCE

offered through the University of Mannheim and the Joint Program in Survey Methodology (Universities of Maryland and Michigan, Westat)

[BE PART OF IT](#)

survey-data-science.net

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

Partners and Funding

University Partners

- University of Maryland
- University of Mannheim

- Catholic University of Santiago de Chile
- Australian National University
- Beijing University
- NCAER -- India
- U. of Capetown (planned)



SPONSORED BY THE

Federal Ministry
of Education
and Research



Other Partners

- SRO - Michigan
- PEW
- German Record Linkage Center
- GESIS
- Bureau of Labour Statistics
- U.S. Census Bureau
- Statistics Netherlands

The project on which this report is based was promoted with funds from the Federal Ministry of Education and Research under the reference number [16OH22064]. Responsibility for the contents of this publication lies with the author.

Modules

Data Output/Access

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

Data Analysis

Learn a variety of analysis methods suited for different data types

Data Curation/Storage

Learn how to curate and manage data

Data Generating Process

Understand how to collect data yourself, and how data are generated through administrative and processes.

Research Question

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

Data Output/Access

min.
3 credits/
6 ECTS

Ethics
1 credit/2 ECTS

Data
Confidentiality and
Statistical
Disclosure Control
2 credits/4 ECTS

Visualization
2 credits/4 ECTS

Data Analysis

min.
6 credits/
12 ECTS

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

Data Curation/Storage

min.
3 credits/
6 ECTS

Database
Management
3 credits/6 ECTS

Data Munging I-III
1 credit/2 ECTS
each

Data Generating Process

min.
4 credits/
8 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

Research Question

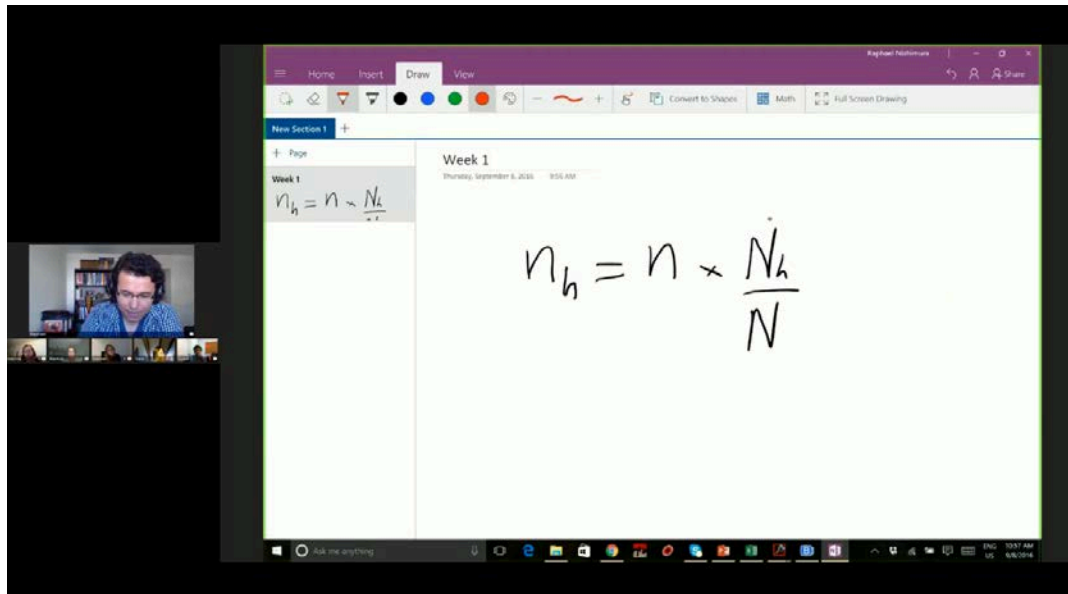
min.
3 credits/
6 ECTS

Fundamentals of
Survey and Data
Science
3 credits/6 ECTS



Research: How to interact?

Synchronous



- Reduces social isolation
- Questions answered right away

Asynchronous interaction



Privacy of self-administered modes when doing q's in public
by [redacted] - Tuesday, 7 June 2016, 3:48 PM

I was wondering if self-administered questionnaires when done in public (public transport, at the park, ...) can still be considered highly private? I'd rather assume that filling out a questionnaire in public leads to a feeling of low privacy and external factors like sex or race of the people surrounding the respondent are likely to alter his response behaviour.

[Permalink](#) | [Edit](#) | [Delete](#) | [Reply](#)



Re: Privacy of self-administered modes when doing q's in public
by [redacted] - Sunday, 12 June 2016, 7:25 PM

Great point. I agree that if people feel that questions are sensitive, they may just decide not to do the survey at all. I think self-administering the survey using an iPad would help. The respondent wouldn't have to say the answer out loud and the interviewer couldn't see the answers provided when finished (compared to if they were given a paper/pencil survey.)

[Permalink](#) | [Show parent](#) | [Edit](#) | [Split](#) | [Delete](#) | [Reply](#)

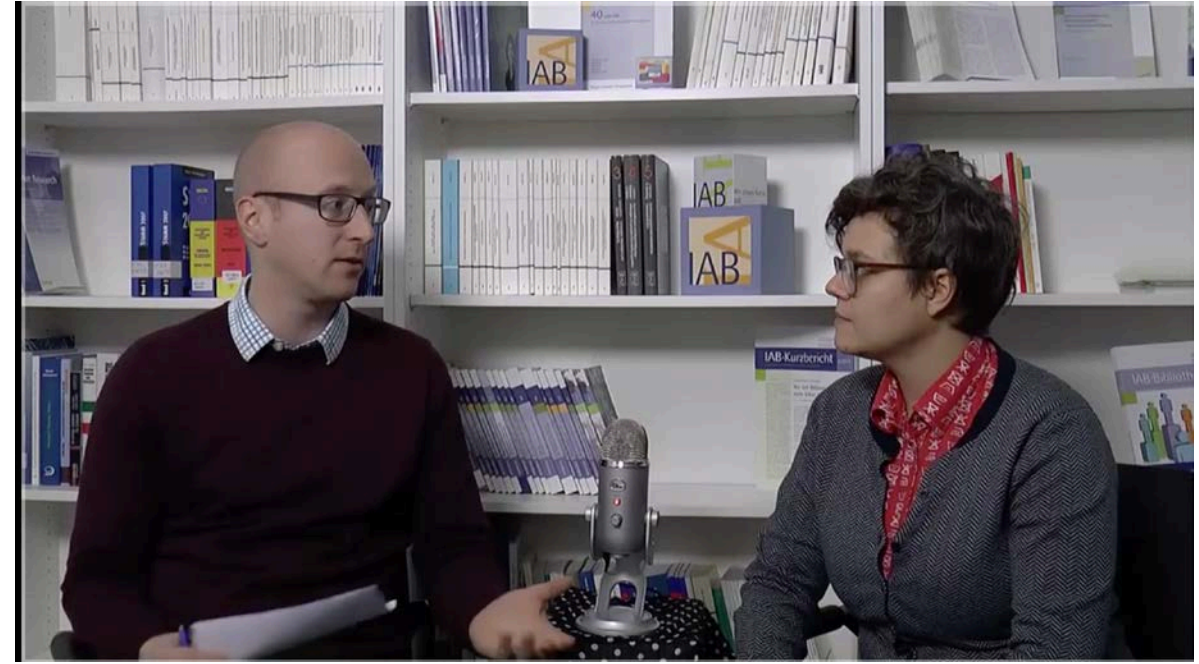
- Flexibility
- More time to think
(Hrastinski & Keller, 2007, p. 66)

Research: How to structure the material?

Video lectures



Lectures + Interviews



- Expert interviews can show practical application

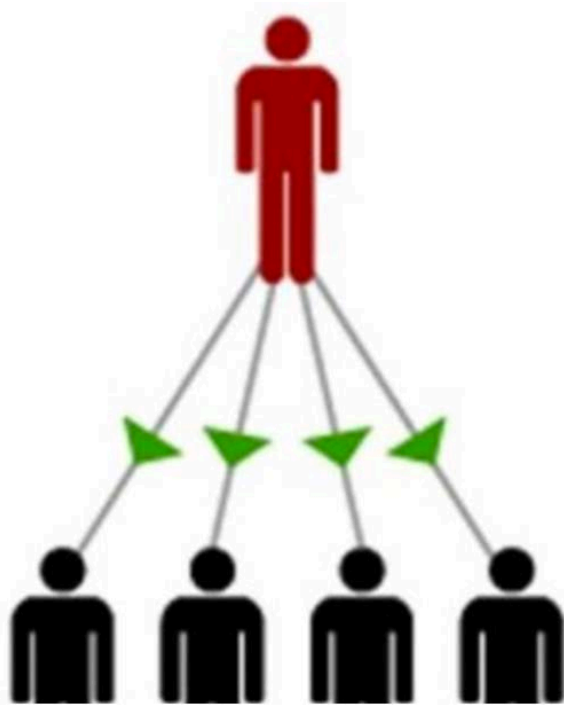
(Renninger & List, 2012)

Research: How much flexibility?

Instructor-paced:

11 week

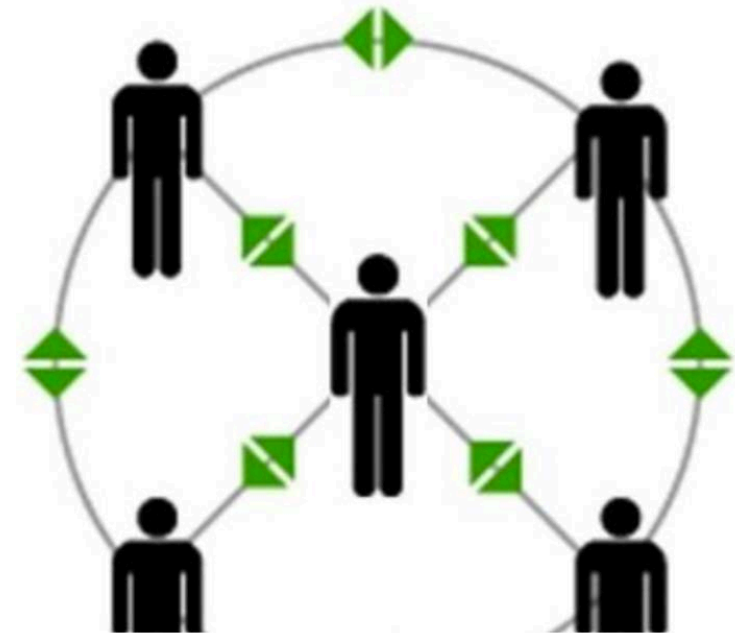
Deadlines for homework



Self-paced:

8 week

No deadlines – all due at the end



Lessons Learning

- Modular approach much appreciated by working professionals
- Learning with application at hand is key
- Participants from all sectors and disciplines
- Very high demand on graduates
- Privacy and confidentiality very important

- Hardest to learn and hardest to teach:
Asking the right question!

<http://coleridgeinitiative.org>
<http://survey-data-science.net/>

fkreuter@umd.edu