

International Program in Survey and Data Science

Frauke Kreuter

JPSM – Uni Mannheim – IAB

ASI 29.6.2017

[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](#)

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.

survey-data-science.net

Koordination und Finanzierung



SPONSORED BY THE

Federal Ministry
of Education
and Research



Neuer Studiengang -

Warum?



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 Japac, Co-Chair, Statistics Sweden
Frauke 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

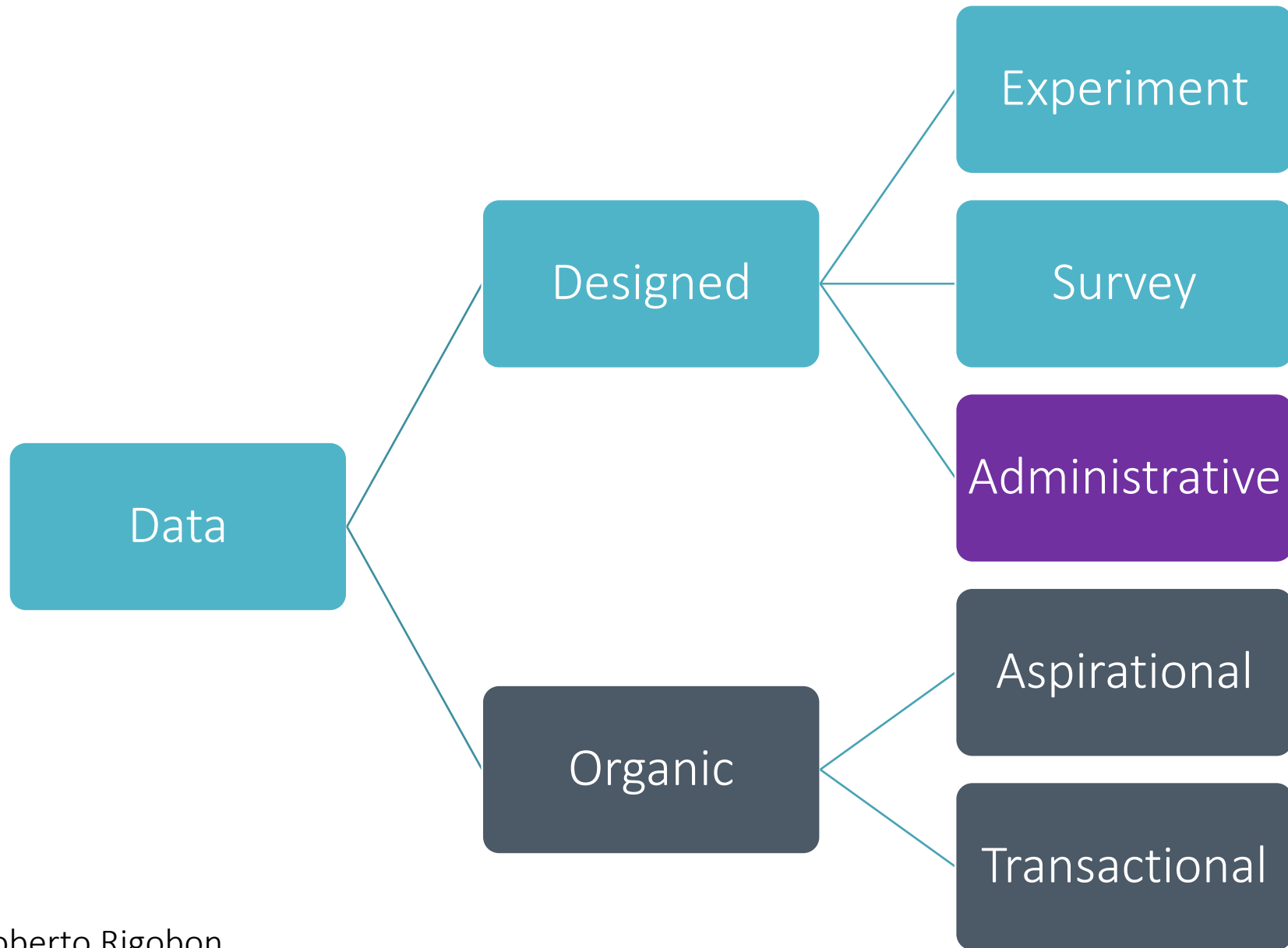
INNOVATIONS IN FEDERAL STATISTICS

Combining Data Sources While
Protecting Privacy

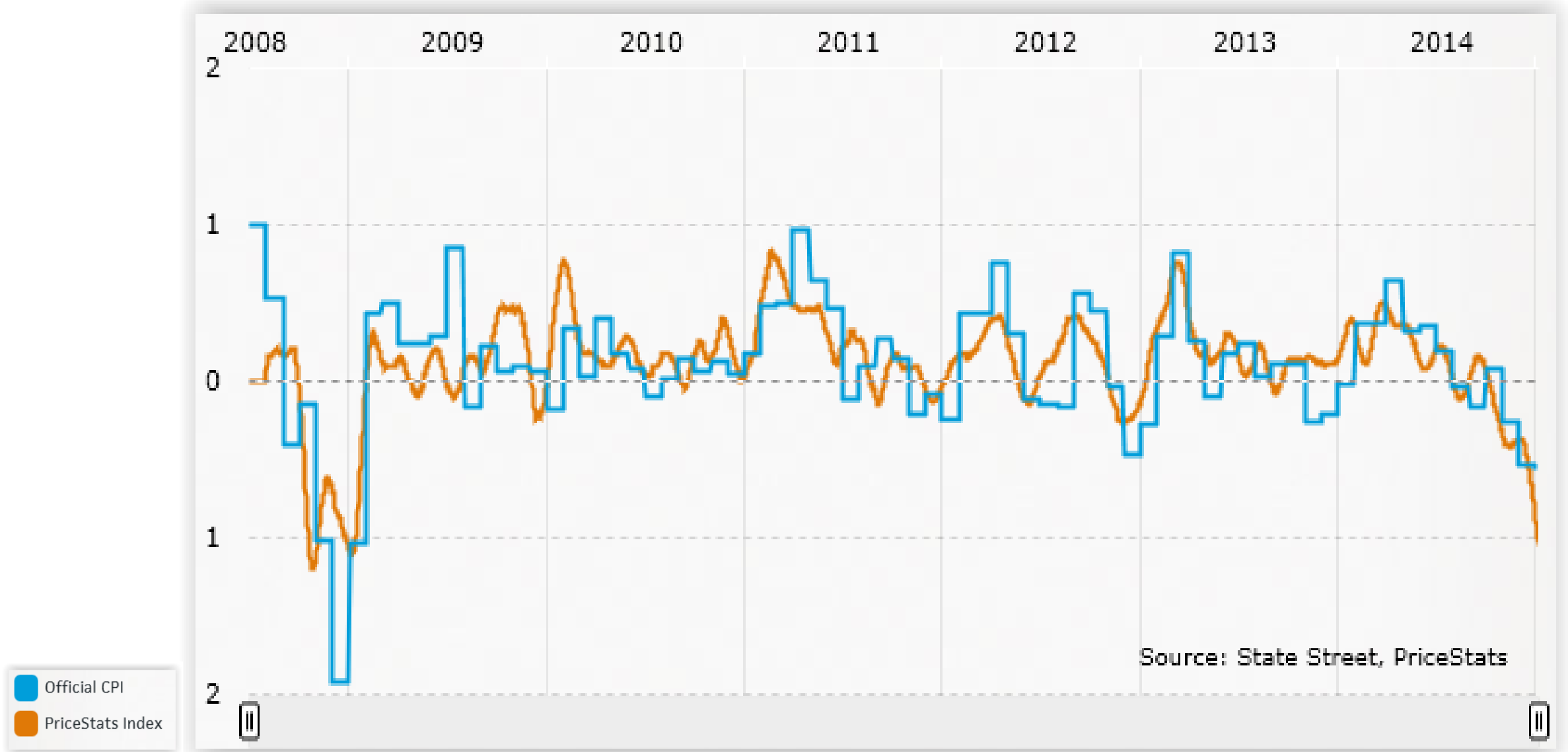


Edited by
**Ian Foster, Rayid Ghani,
Ron S. Jarmin, Frauke Kreuter,
and Julia Lane**

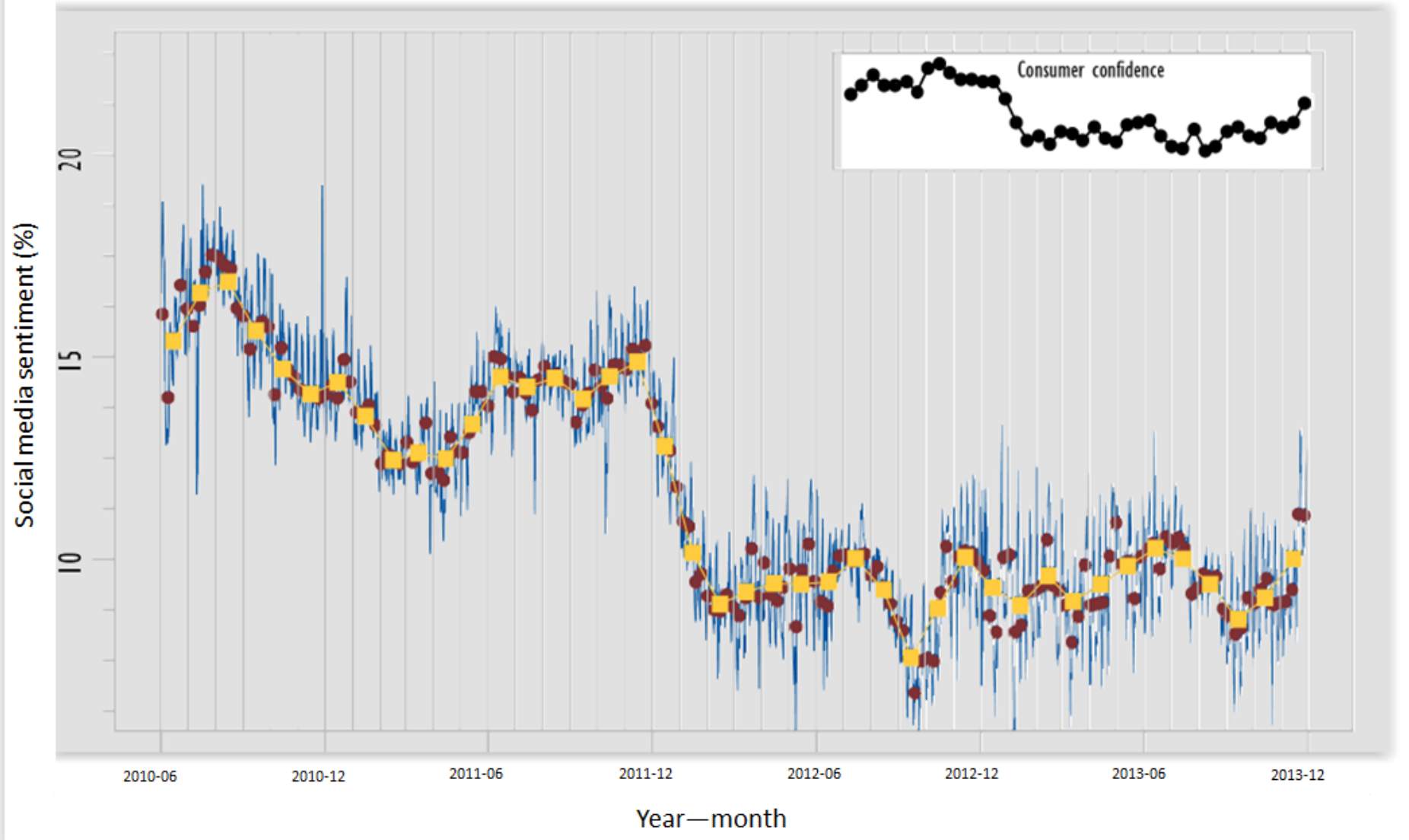
 CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK



Source: Roberto Rigobon



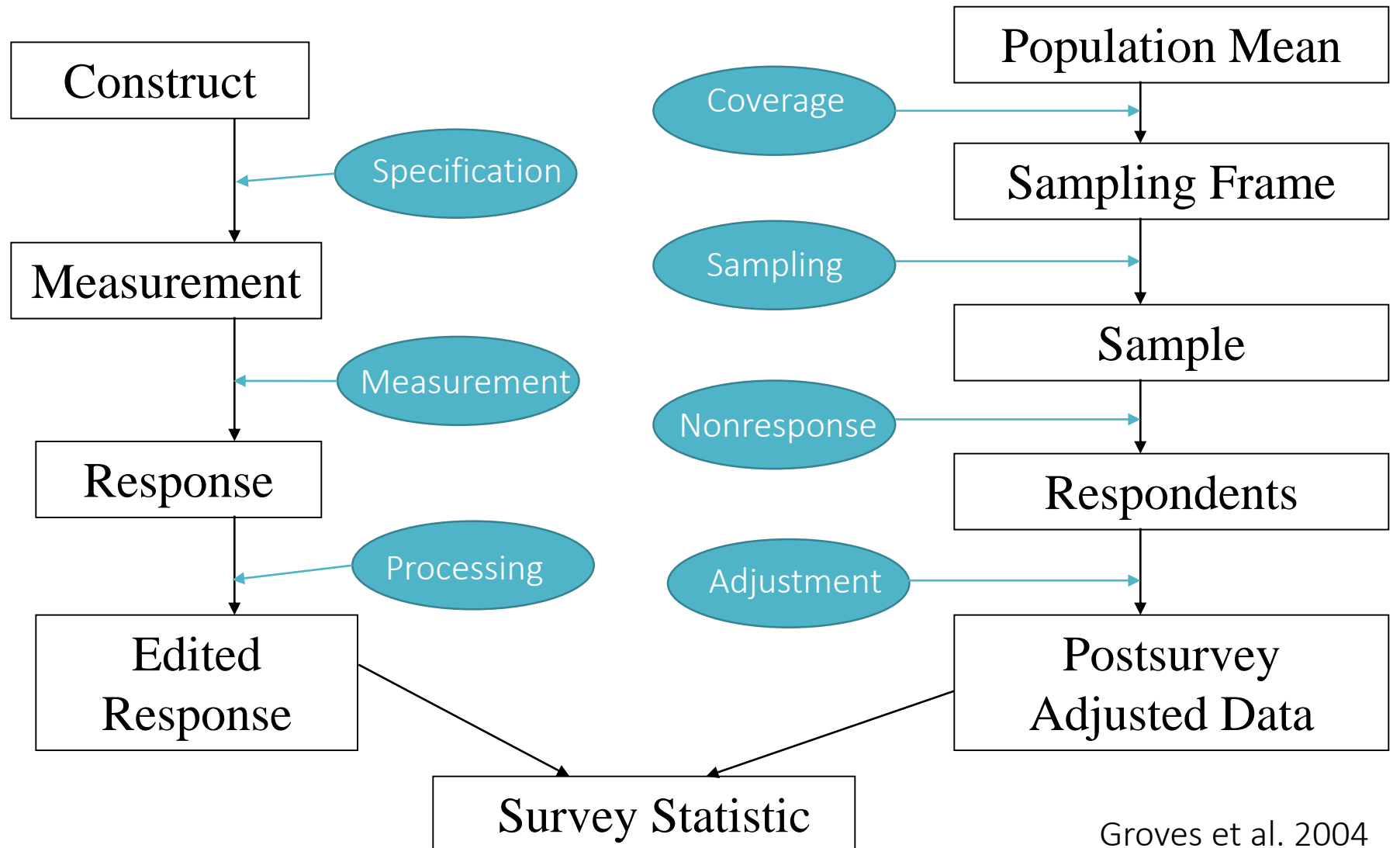
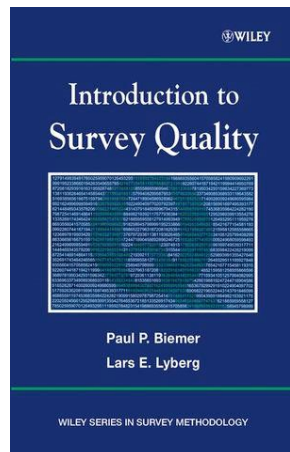
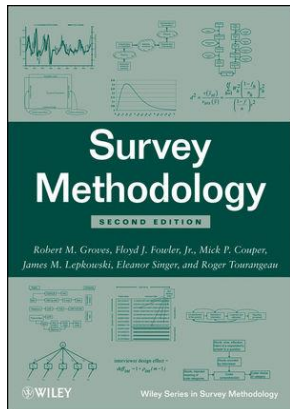
US Aggregated Inflation Series, Monthly Rate, PriceStats Index vs. Official CPI.
Accessed January 18, 2015 from the PriceStats website.



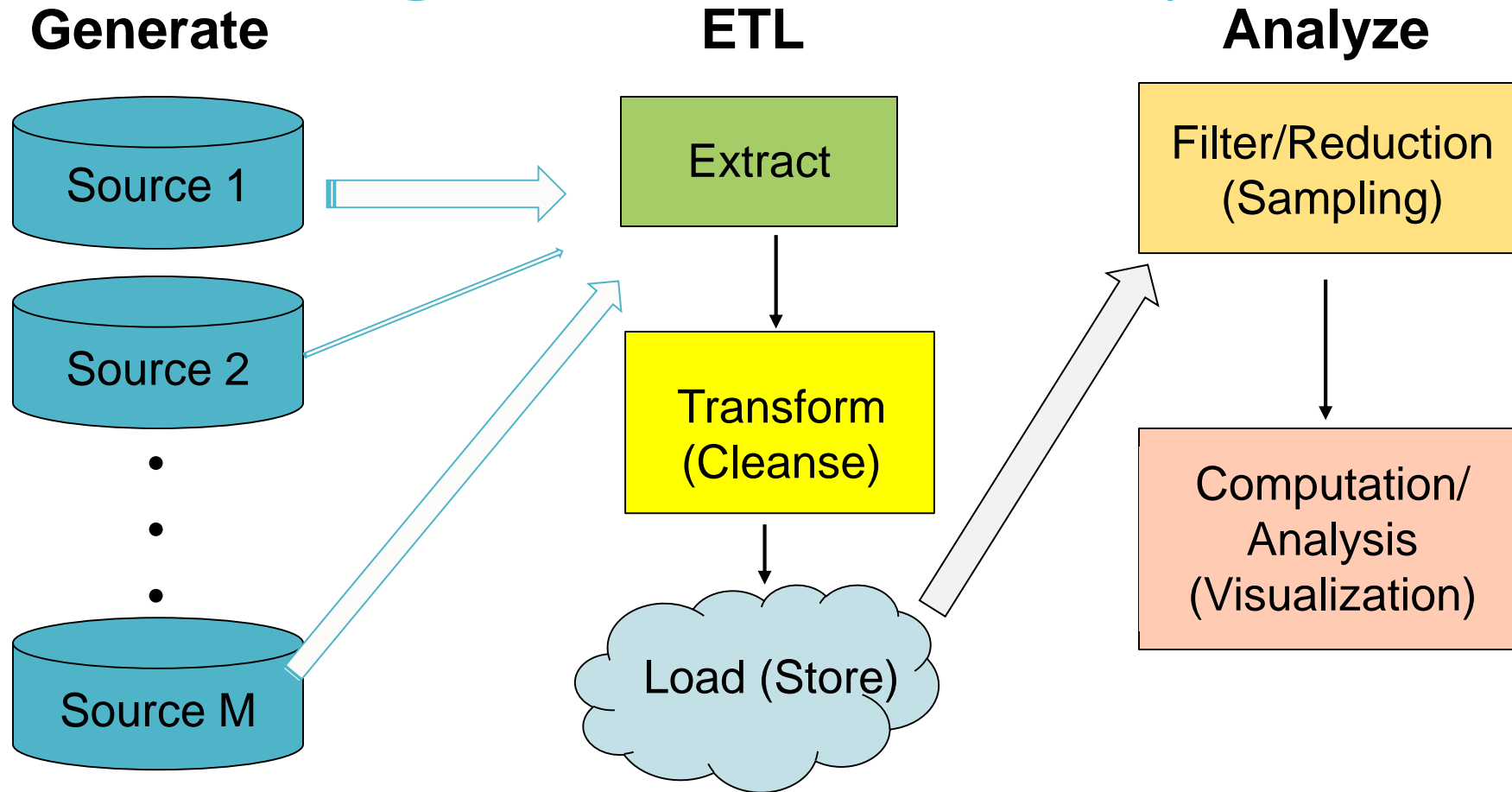
Social media sentiment (daily, weekly and monthly) in the Netherlands, June 2010 - November 2013. Consumer confidence for the same period is shown in the insert (Daas and Puts 2014).

Die Zweifel

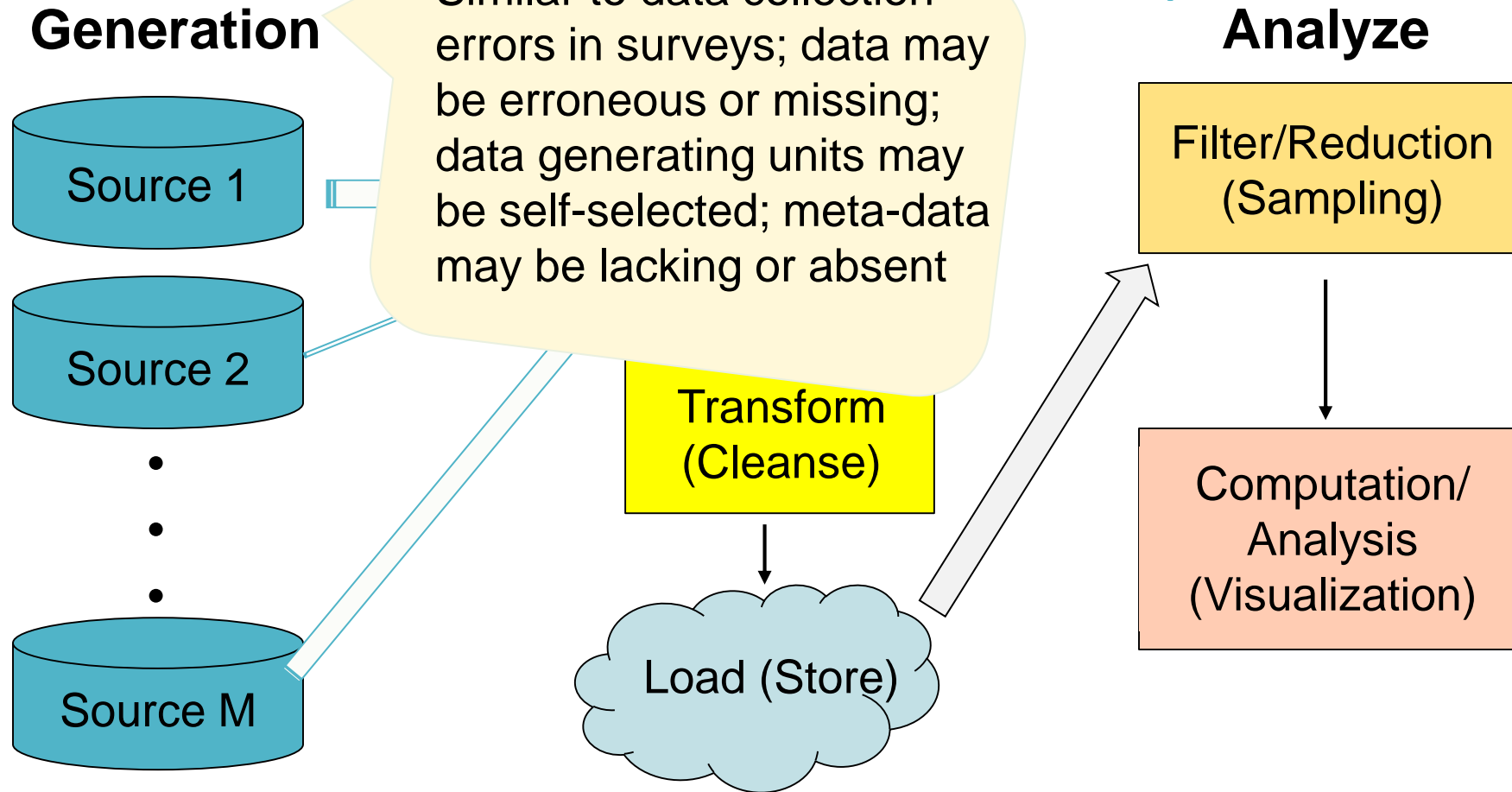
Daten generierender Prozess



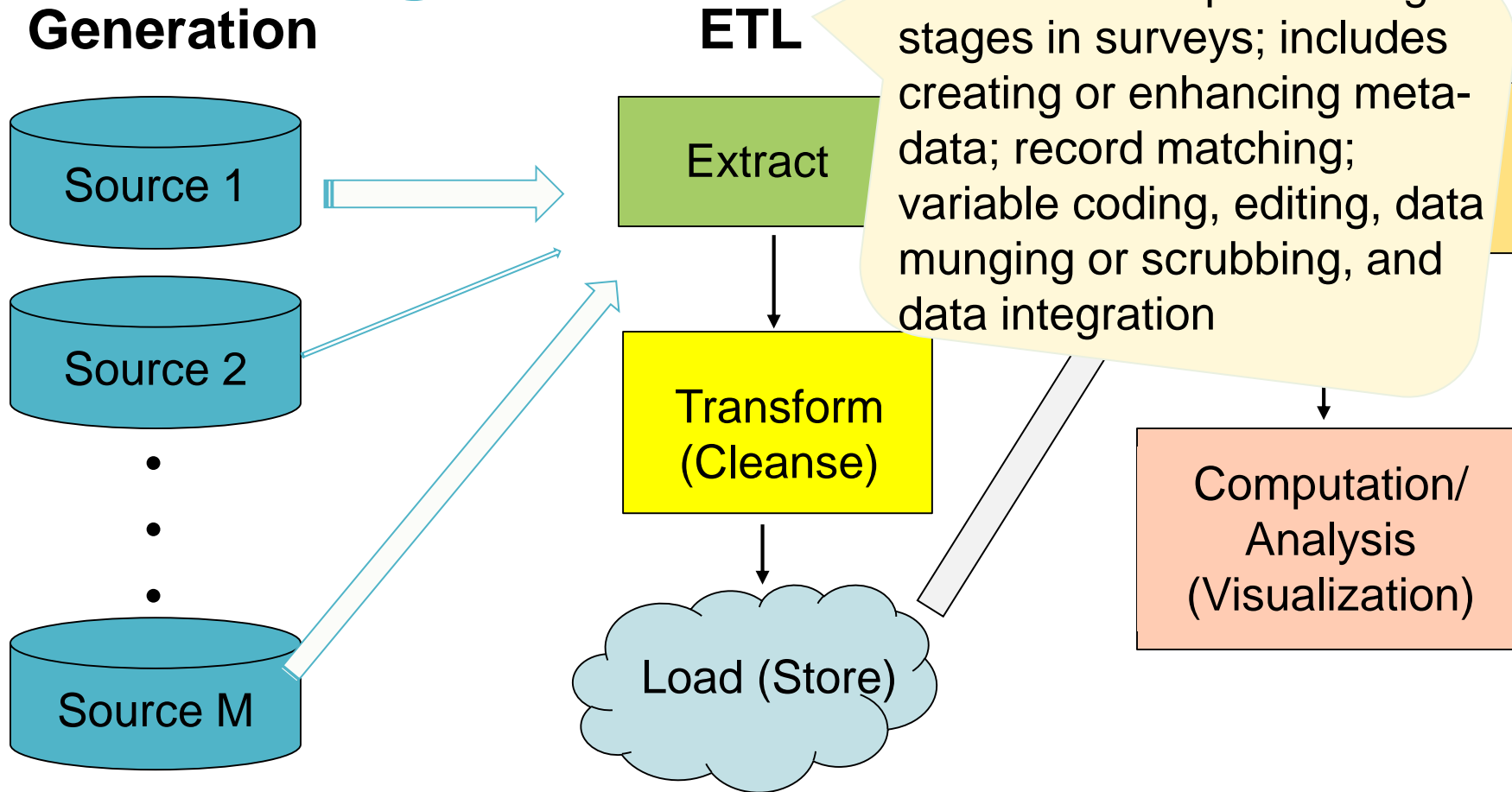
Big Data Process Map



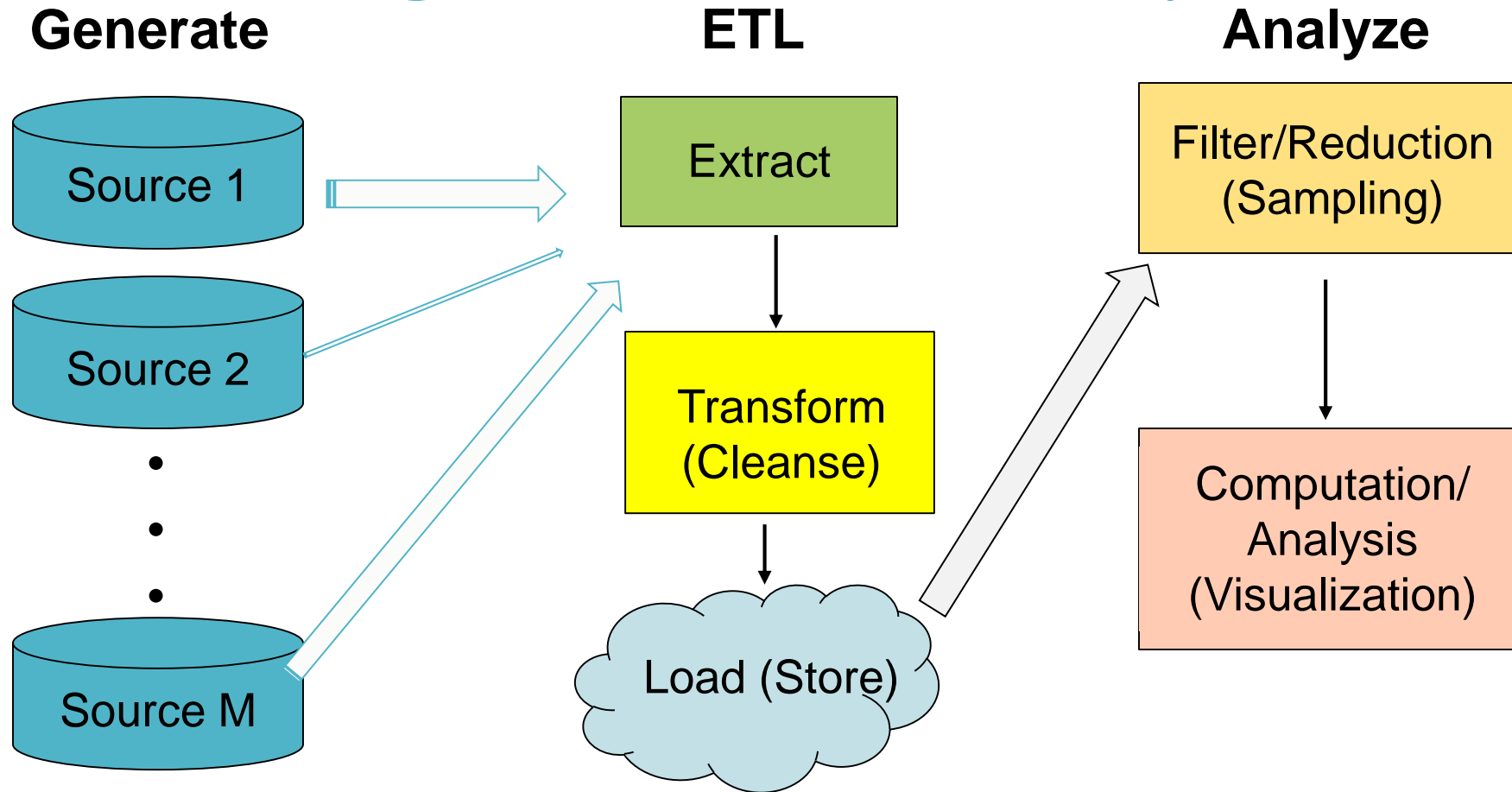
Big Data Process Map



Big Data Process



Big Data Process Map



Lerninhalte

Modules

Data Output/Access

Data Analysis

Data Curation/Storage

Data Generating Process

Research Question

Learn how to communicate results and distribute and store your data

Learn a variety of analysis methods suited for different data types

Learn how to curate and manage data

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

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

Content key words

Data Output/Access

Visualization, disclosure control, ethics, privacy

Data Analysis

Statistical methods, machine learning, Bayesian, hierarchical, small area estimation

Data Curation/Storage

Practical training in data base management, SQL, editing, coding, imputation, etc.

Data Generating Process

Designed (survey and admin) and organic data (transaction and aspirational), linkage, matching

Research Questions

Economics, public policy, criminology, journalism, public health, sociology, etc.

IPSDS – im Detail

Partner

Uni Partner

- University of Maryland
- University of Mannheim

- Catholic University of Santiago de Chile
- Australian National University
- Beijing University
- Ashoka University (expressed interest)
- U. of Capetown (planned)

Andere Partner

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

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

Generalized Linear
Models
2 credits/3 ECTS

Analysis of
Complex Data I-III
1 credits/2 ECTS
each

Propensity
Score/Statistical
Matching
2 credits/4 ECTS

Machine Learning
I-III
1 credit/2 ECTS
each

Data Curation/Storage

min.
3 credits/
6 ECTS

Database
Management I-III
1 credits/2 ECTS
each

Data Munging I-III
1 credit/2 ECTS
each

Data Generating Process

min.
4 credits/
8 ECTS

Data Collection
Courses
1 credits/2 ECTS
each

Record Linkage
1 credit/2 ECTS

Practical Tools for
Sampling and
Weighting
3 credits/6 ECTS

Applied Sampling
I-III
1 credits/2 ECTS
each

Experimental
Design
2 credits/4 ECTS

Research Question

min.
3 credits/
6 ECTS

Fundamentals of
Survey and Data
Science
3 credits/6 ECTS

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

Total:
30 credits
Min 36 credits as dual degree

Master Thesis

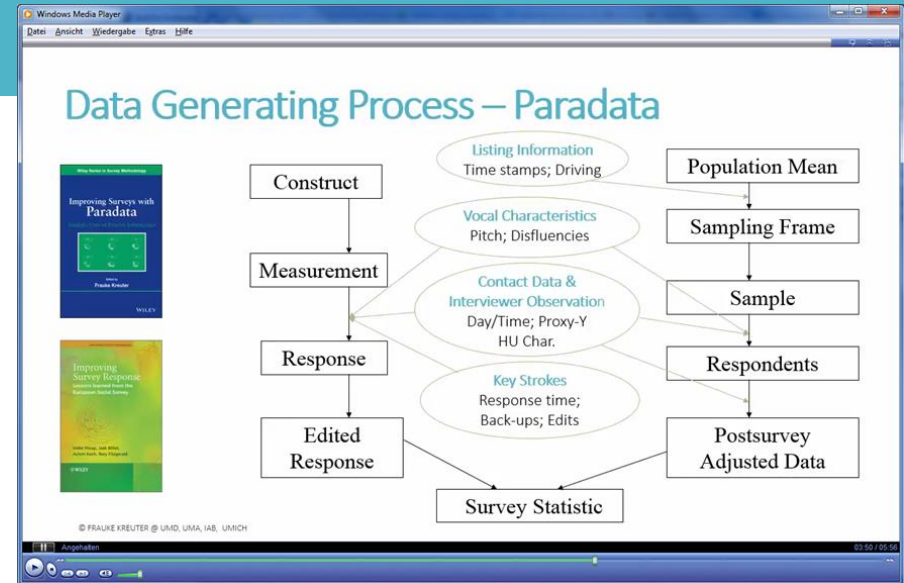
Format

Jede Woche Videos (pre-recorded)
Etwa 1.5-2h pro Woche
In 5-10 minütigen Portionen

Eigene Zeiteinteilung

Wöchentliche Tests und Aufgaben

Wöchentliche Diskussionsrunde



IPSDS Struktur



2. Model Eval_Validation
January 12, 2016 Mediasite Presenter

3. K-Means Clustering
January 12, 2016 Mediasite Presenter

Homework Assignment 1
data file for homework n
Tasks for Homework Num

Quiz 2
HW Number 1 Solutions

This is a .R file that can be opened using Notepad or other text editor (or Web tasks of HW 1.

Week 3
Bluejeans Join Meeting [Tuesday, 02/16/

4. K-Nearest Neighbors
January 12, 2016 Media: Switch to Active Speaker

5. CARTS
January 12, 2016 Media:
HW 2 Assignment
Tasks for
Datasets

Quiz 3
HW 2 Solutions

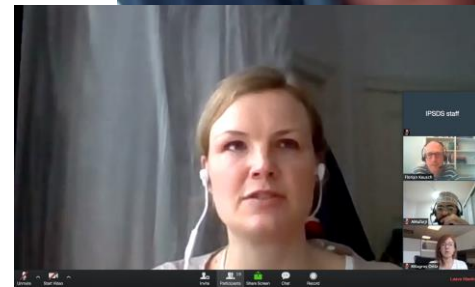
Here is the R script file co

3. K-Means Clustering
www.jpmsclasses.umd.edu/Mediasite/Play

Machine Learning Methods/Techniques

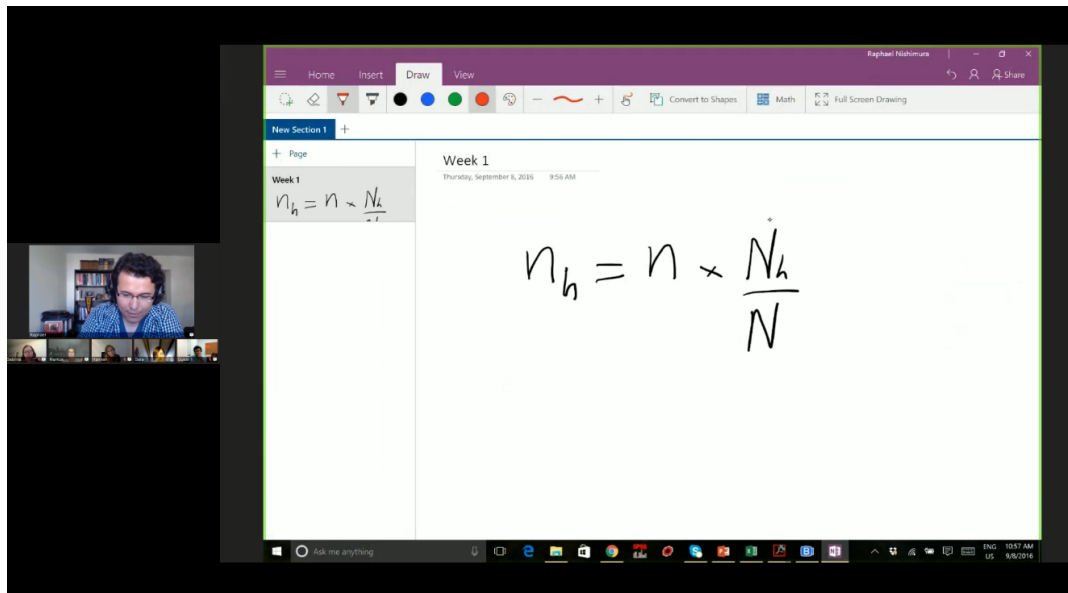
- There are many different machine learning methods available
- Many are non-parametric in nature and while a functional form can be specified, it is generally not a natural byproduct of the method
- Wu et al. (2008) provide an overview of ten of the top machine learning algorithms including (see <http://bit.ly:/liwTir>):

- K-means Clustering
- PageRank
- K-nearest neighbors



BMBF - Studie 1: Art der Interaktion

Synchrone Interaktion



The screenshot shows a Zoom meeting interface. On the left, there is a video thumbnail of a participant. The main area is a whiteboard with a purple header bar. The whiteboard contains the following text:

Home Insert Draw View
Convert to Shapes Math Full Screen Drawing

New Section 1 +

Page

Week 1
Thursday, September 8, 2016 9:56 AM

$$n_h = n \times \frac{N_h}{N}$$

Ask me anything

- verhindert soziale Isolation
- Fragen werden sofort beantwortet

Asynchrone Interaktion



Privacy of self-administered modes when doing q's in public
by [Name] - 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 [Name] - 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](#)

- Flexibilität
- Mehr Zeit zum Nachdenken/Reflexion (Hrastinski & Keller, 2007, p. 66)

BMBF- Studie 2: Video-Lehrinhalte

Video-Vorlesungen



+ Experten-Interviews



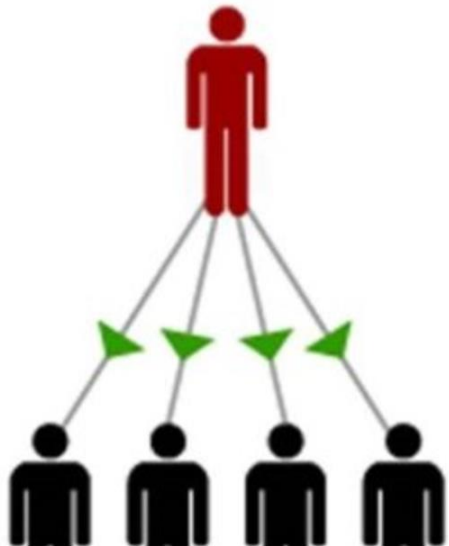
- Experten-Interviews betonen Nutzen der Lehrinhalte
(Renninger & List, 2012)
- Kurs: **12 Wochen** ab Feb. 2016
- Within-subject Design: 16 ProbandInnen
- Experten-Interviews in Wochen: **3, 4, 5, 8**

BMBF - Studie 3: Flexibilität der Abgabefristen für Hausaufgaben

Instructor-Paced Format:

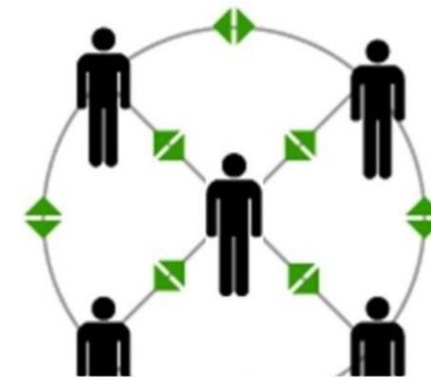
11 Wochen

Friste für alle Hausaufgaben

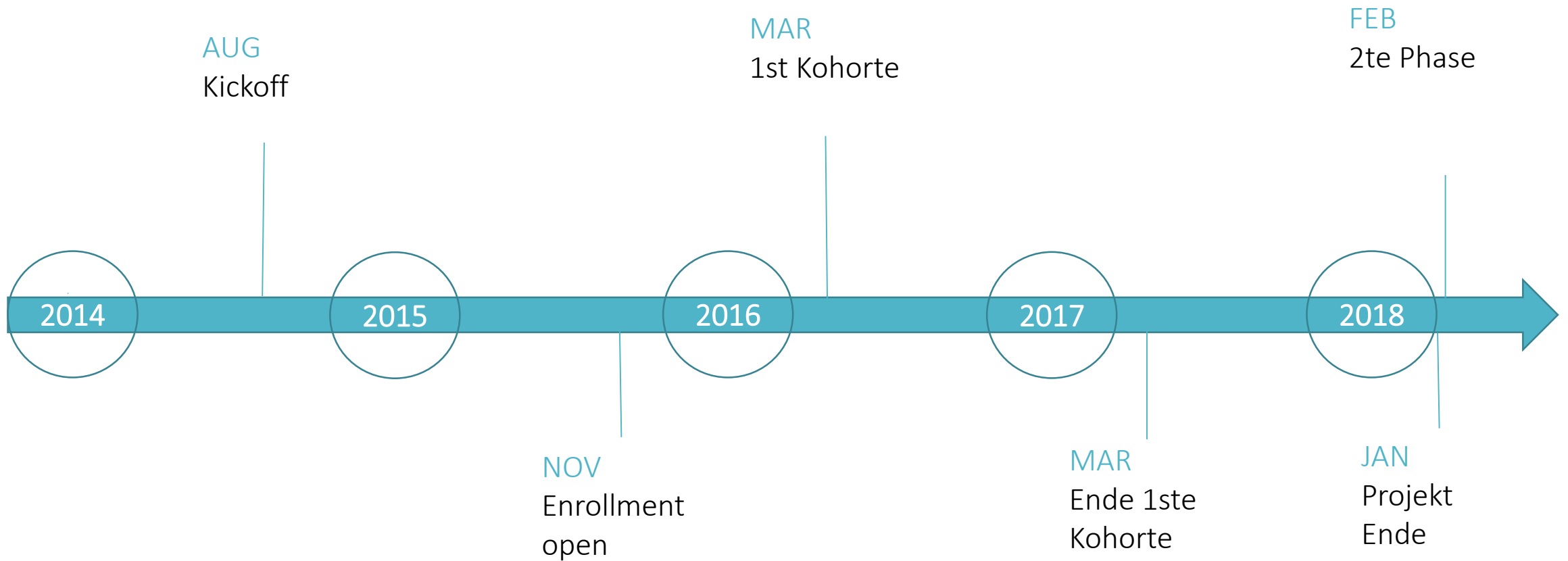


Self-Paced Format:

8 Wochen, Keine Fristen (alle Hausaufgaben mussten bis Ende des Kurses abgegeben werden)

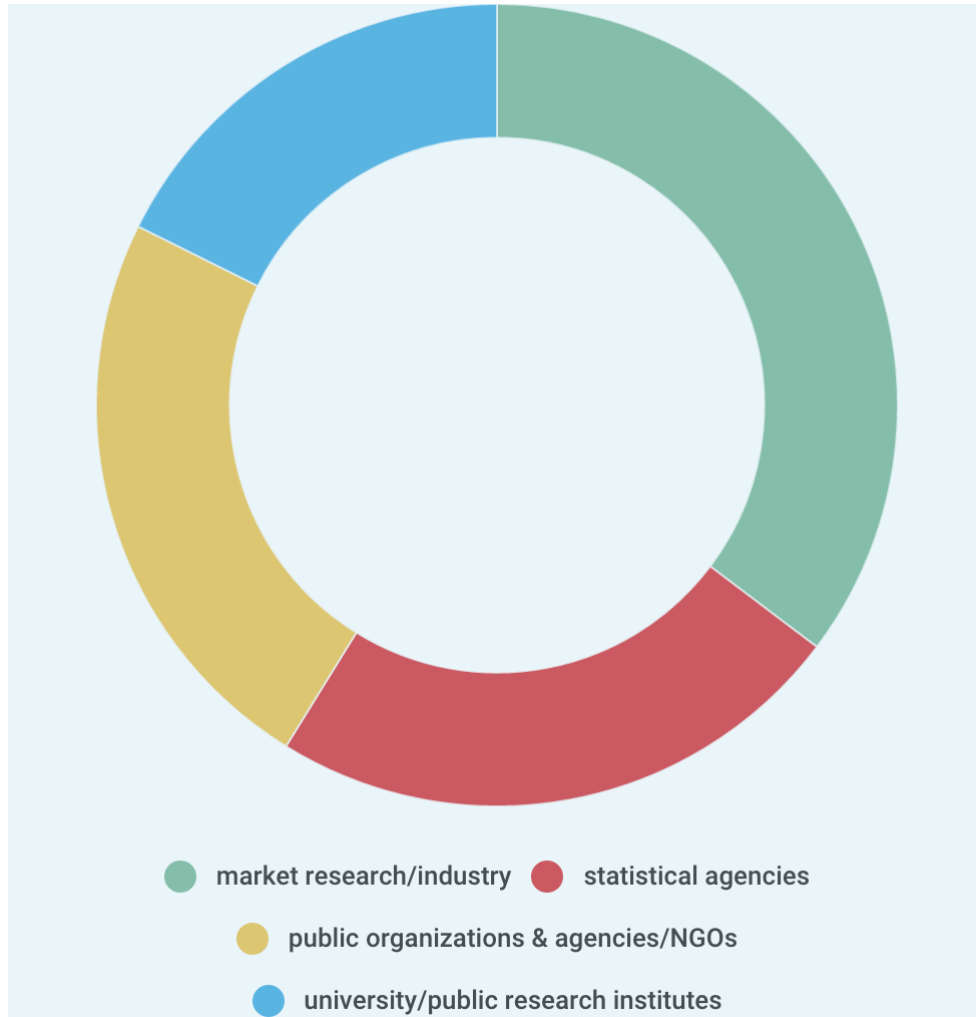


Zeitstrahl der Testphase in Mannheim

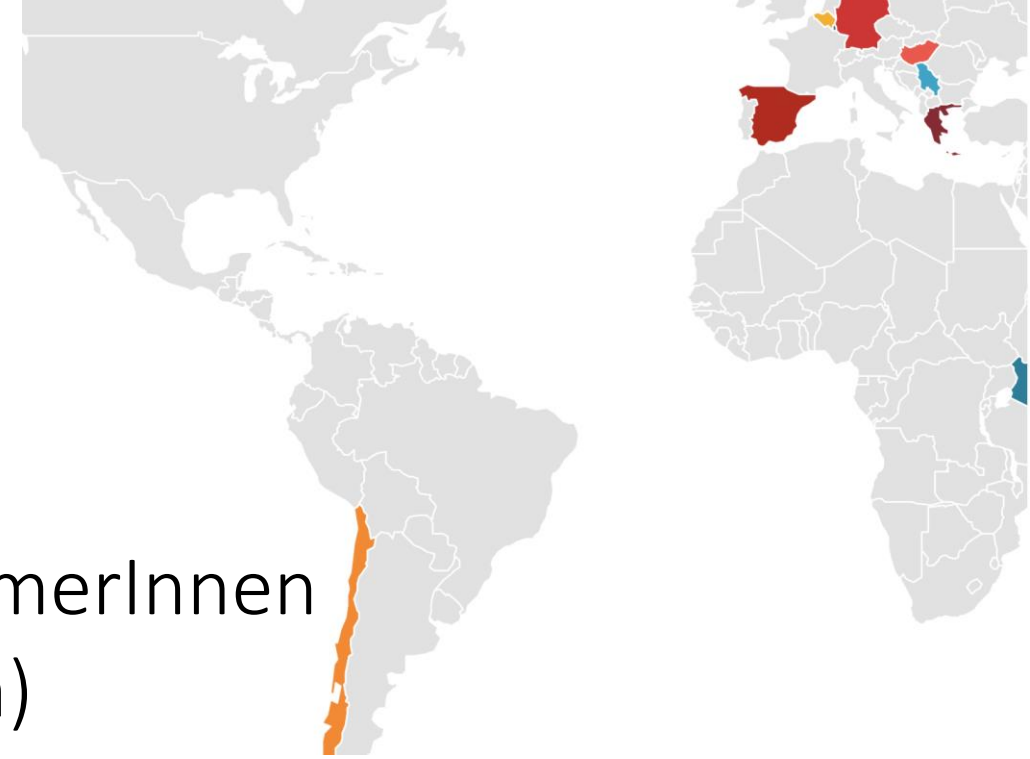


Interesse ist groß

IPSDS 1ste Test-Kohorte



- **16** TeilnehmerInnen (10 f + 6 m)
- Medianalter **29,5** Jahre (Min-22; Max-55)
- Alle haben mindestens B.A.
- **40,76** Arbeitsstunden/Woche (Min-35; Max-55)



Hohe Nachfrage auf Arbeitgeber Seite



