

Big Data Analysis

Frauke Kreuter¹ Christoph Kern Marcel Neunhoeffler Sebastian Sternberg

March 25, 2019

Mannheim Machine Learning Modules

frauke.kreuter@uni-mannheim.de

Background

Problems:

- Increased **nonresponse rates** → quality concerns
- Increased **cost** → changes in modes and procedures
- Increased demand for **timeliness** ← other information providers
- Increased demand for **fine grained** information

Hope lies in big data:

- Administrative data
- Commercial data
- Paradata

The Excitement



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:

Edith Iqbal, Co-Chair, Statistics Sweden
Frankie Braker, Co-Chair, IPSM at the U. of Maryland, U. of Minnesota & IAB
Morris Berg, Stockholm University
Paul Bremer, IIT International
Paul Decker, Mathematica Policy Research
Cheryl Lampe, School of Information at the University of Michigan
John Lene, American Institutes for Research
Carly O'Neil, Johnson Research Labs
Abe Usher, HumanGig Group

Acknowledgment: We are grateful for comments, feedback and editorial help from
Erin Ben-Pershi, Jesse McGillion, and the AAPOR council members.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

REPORT

INNOVATIONS IN FEDERAL STATISTICS

Combining Data Sources While
Protecting Privacy

Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series

BIG DATA AND SOCIAL SCIENCE

A Practical Guide to Methods and Tools

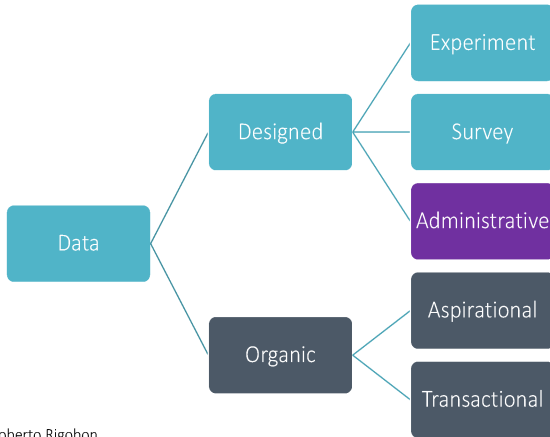


Edited by
**Ian Foster, Rayid Ghani,
Ron S. Jarmin, Frauke Kreuter,
and Julia Lane**

 **CRC Press**
Taylor & Francis Group
A CHAPMAN & HALL BOOK

The Characteristics

Characteristics of Big Data

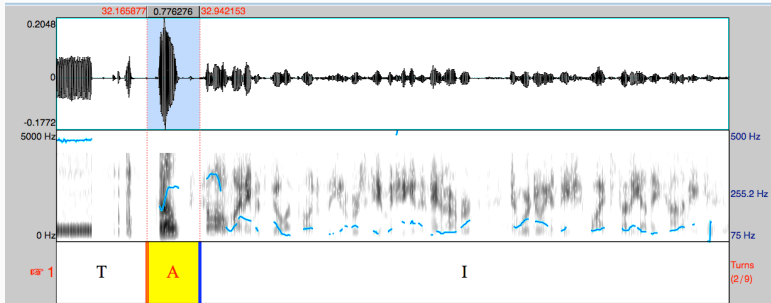


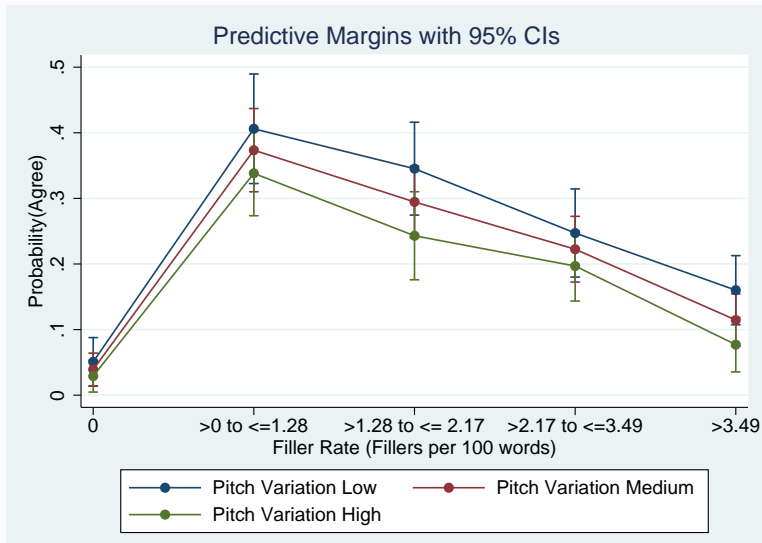
Source: Roberto Rigobon

Some Examples

Example: Vocal Attributes

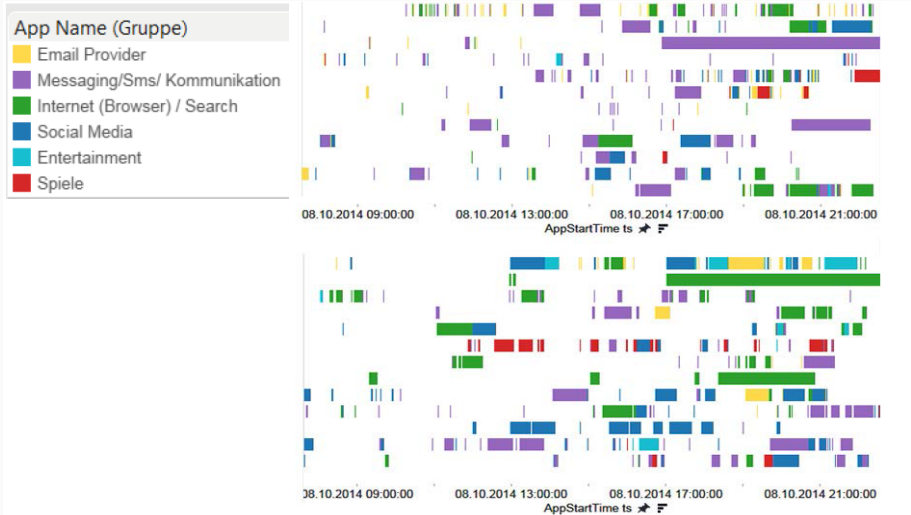
Conrad et al. JRSS-A 2013



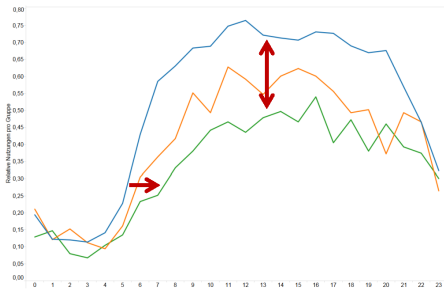


Example: App Usage

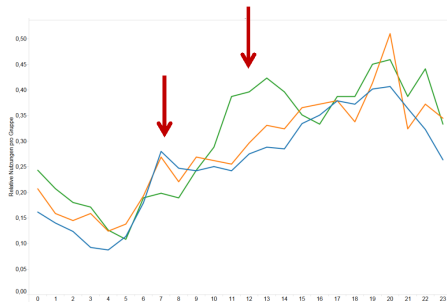
Schlickman et al. DataFest Mannheim

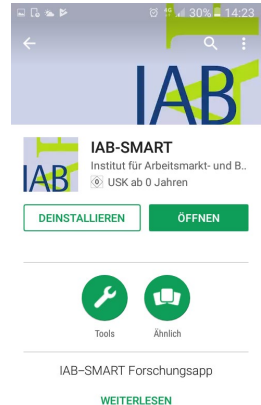
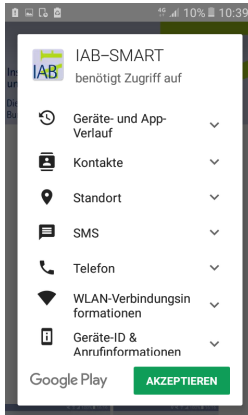


Email



Entertainment





Linkage to PASS

Vielen Dank, dass Sie die IAB-SMART App heruntergeladen haben und an unserem Forschungsprojekt teilnehmen.

Ein wichtiger Bestandteil der Studie ist es, die App-Daten gemeinsam mit den Daten aus der Befragung "Lebensqualität und soziale Sicherung" auswerten zu können. Dafür brauchen wir Ihre Zustimmung. Die Zustimmung können Sie jederzeit widerrufen.

Ich stimme zu, dass die Daten aus dieser App gemeinsam mit den Daten der Befragung

☐

Verlassen Weiter

Registration

teilnehmen.

Ein wichtiger Bestandteil der Studie ist es, die App-Daten gemeinsam mit den Daten aus der Befragung

Registration

Bitte geben Sie Ihren Registrierungscode ein.

Registrierungscode

ABBRECHEN OK

Verlassen Weiter

Data processing

Nutzungsbedingungen und Datenschutz

Namen und Adressen werden von den App-Daten strikt getrennt gehalten. Die Daten werden so ausgewertet, dass keine Rückschlüsse auf Ihre Person möglich sind.

Bitte lesen Sie sich die [Datenschutzhinweise](#) und die [Nutzungsbedingungen](#) aufmerksam durch und stimmen Sie der Datenverarbeitung zu.

Hiermit willige ich in die Datenverarbeitung ein und akzeptiere die Nutzungsbedingungen.

☐

Zurück Weiter

AWS Real Time Entry

The Consequences

The Consequences

- Data Acquisition
 - Web Scraping
 - API
 - Analysis
 - Machine Learning
 - Spatial Statistics
 - Networks
 - Ethics
 - Community
- Data Storage
 - SQL (MySQL)
 - NoSQL
 - Programming
 - Cloud computing
 - Hadoop ecosystem



References

- Conrad, F., Broome, J., Benkí, J., Kreuter, F., Groves, R., Vannette, D., & McClain, C. (2013). Interviewer Speech and the Success of Survey Invitations. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 176. 191-210. 10.2307/23355183.
- Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2018). Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. *Social Science Computer Review*. <https://doi.org/10.1177/0894439318816389>