

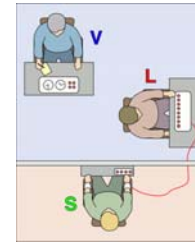
Frauke Kreuter

Data Science in den Sozialwissenschaften

Uni Mannheim & IAB & Uni Maryland @fraukolos

-
- Professor Universities Mannheim & Maryland
 - Head of statistical methods unit at Institute for Employment Research (IAB)
 - Founder: International Program in Survey and Data Science
 - Co-founder: Coleridge Initiative
 - Co-host: Podcast www.digdeep.de



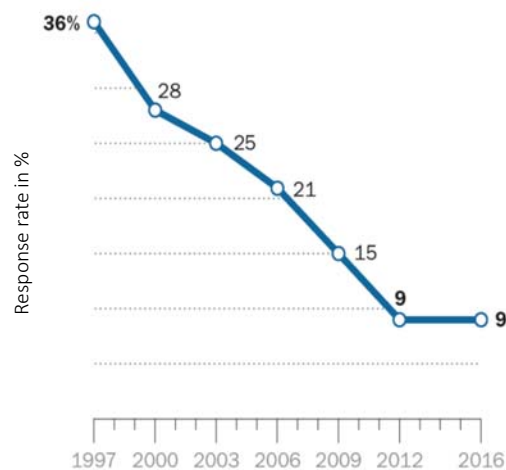


bisher

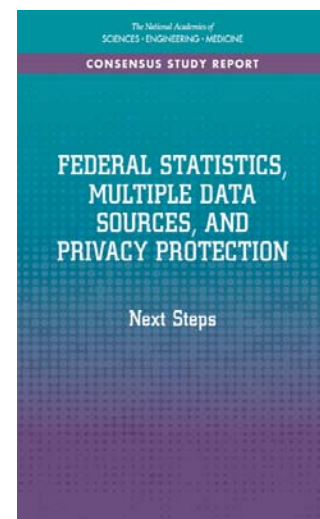
Drei Haupt-Datenquellen
in der quantitativen Forschung

2	Steuernummer	
3	*170k & Lohnsteuerbescheinigungen, sofern vorhanden	
4	*170k & weitere Lohnsteuerbescheinigungen	
5	Einkünfte aus nichtselbständiger Arbeit	
6	Bruttoarbeitslohn	110
7	Lohnsteuer	140
8	Sonderausgaben	150

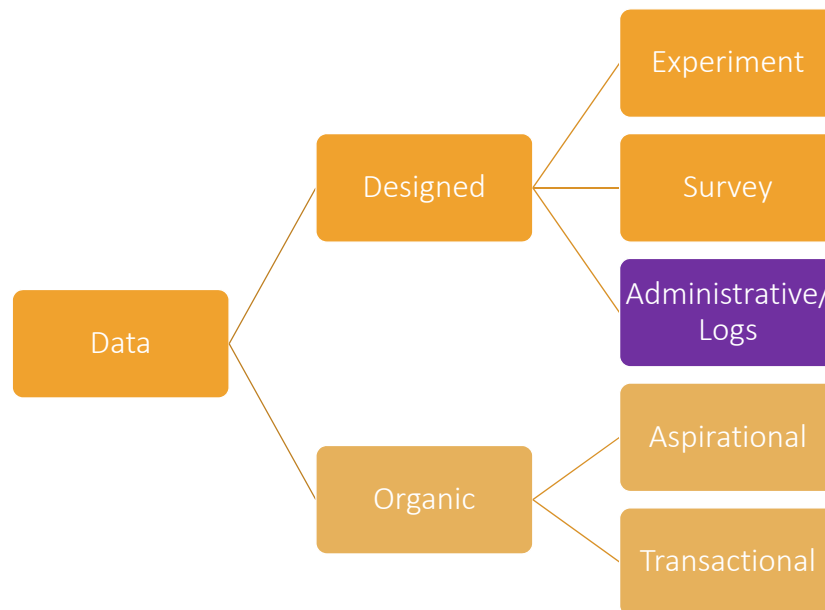
Sinkende Antwortraten



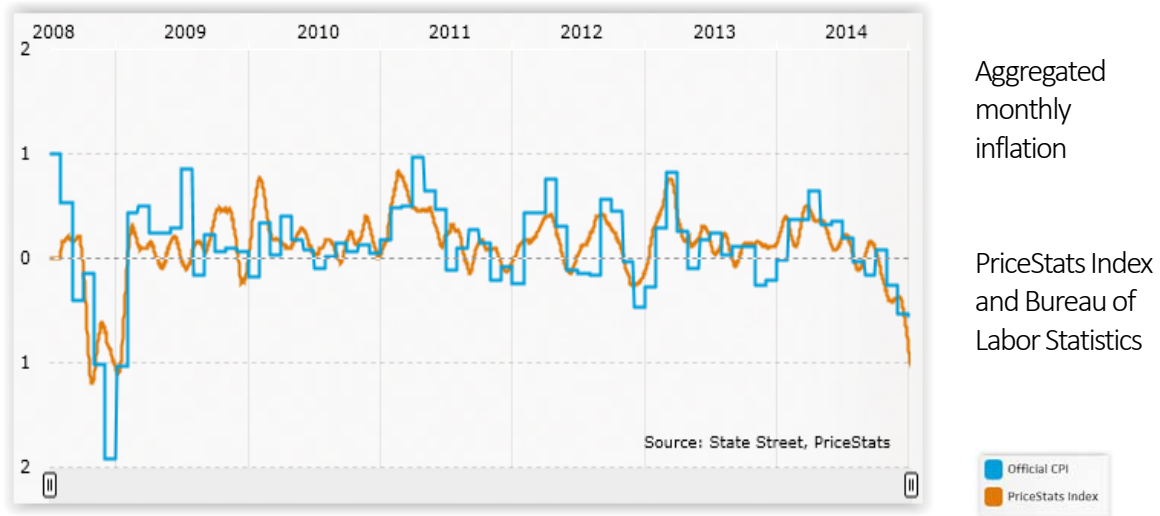
Source: Pew Research Center surveys conducted 1997 to 2016. "What Low Response Rates Mean for Telephone Surveys"



Hoffnung



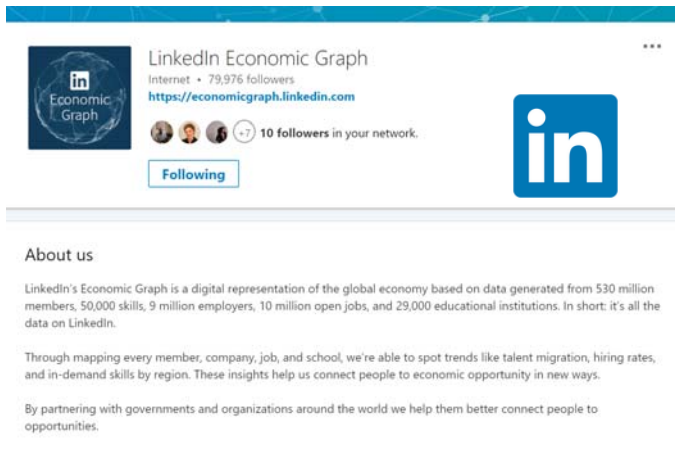
Source: Roberto Rigobon, BEA Expert Meeting on Exploiting Commercial Data for Official Economic Statistics 11/19/15



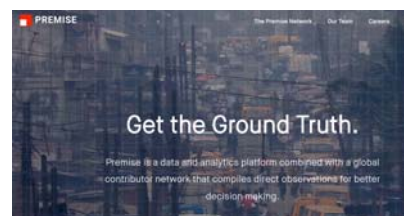
<https://www.pricestats.com/inflation-series?chart=1837>

Quellen

Employment Websites + Aggregators



Plattformen - Crowdsourcing



Beobachtungsdaten



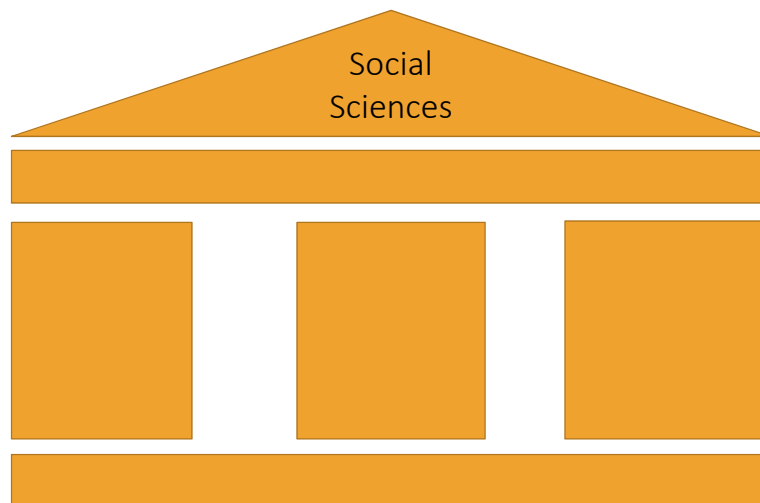
Herausforderungen

Wissenschaftlicher Wert?

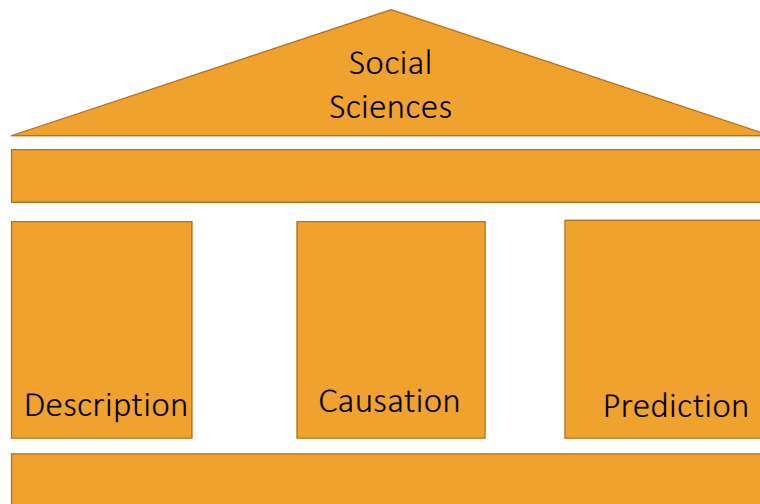
1. Messung

Proxies, wenig Variablen (Couper 2013 - SRM)

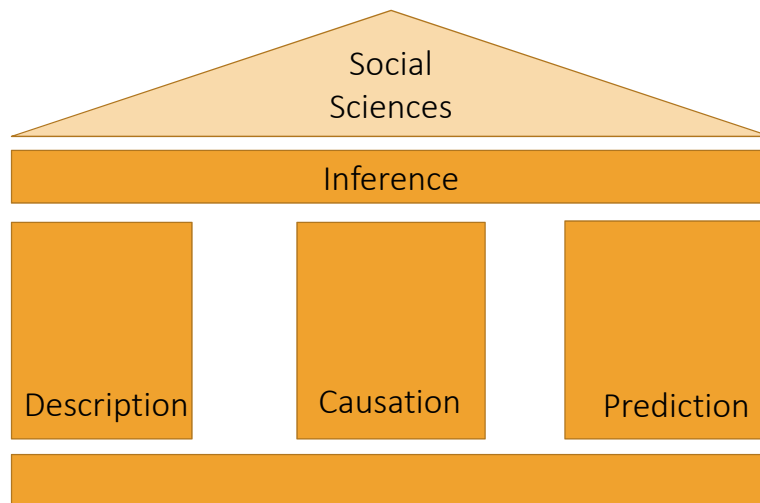
2. Inferenz



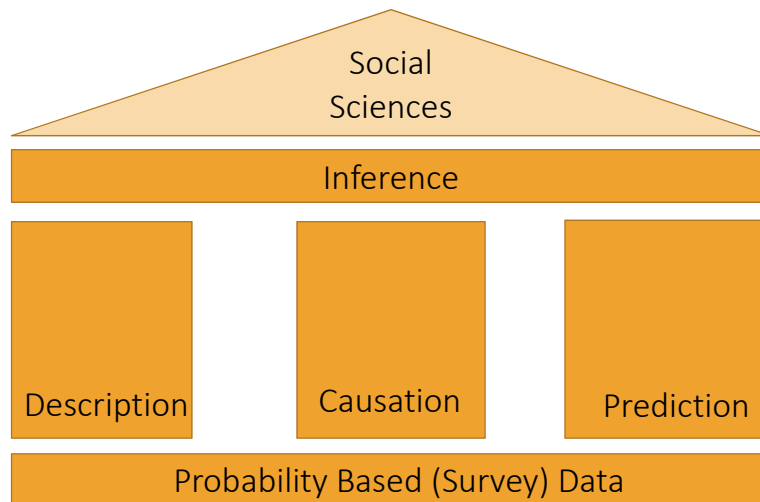
Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019



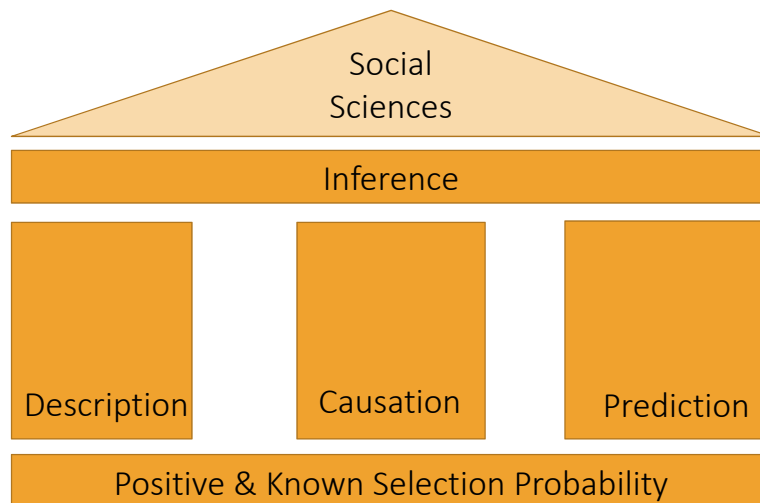
Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019



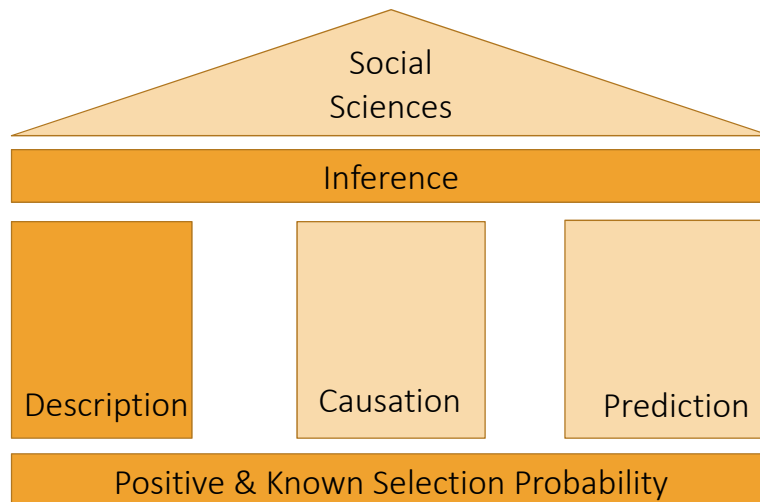
Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019



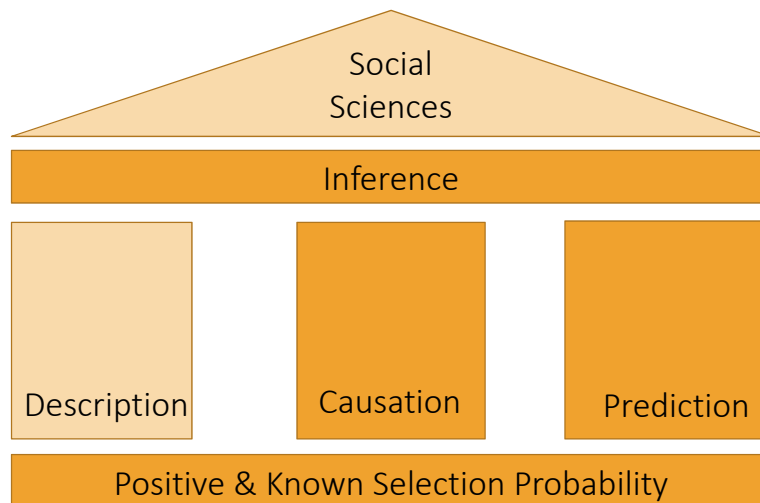
Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019



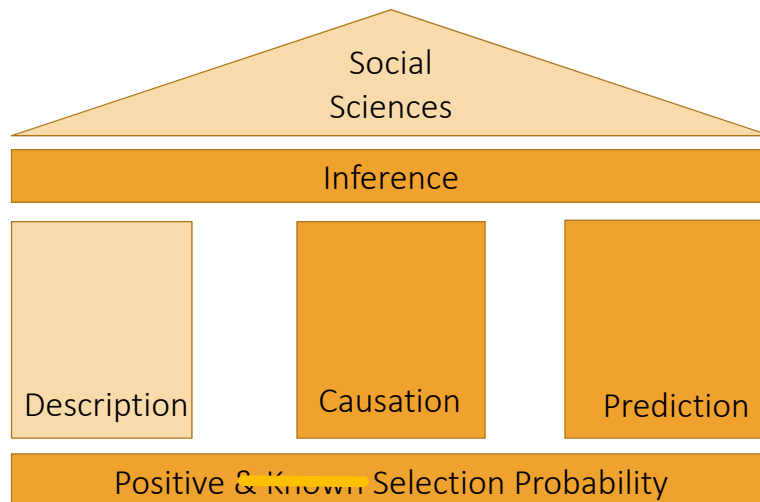
Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019



Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019

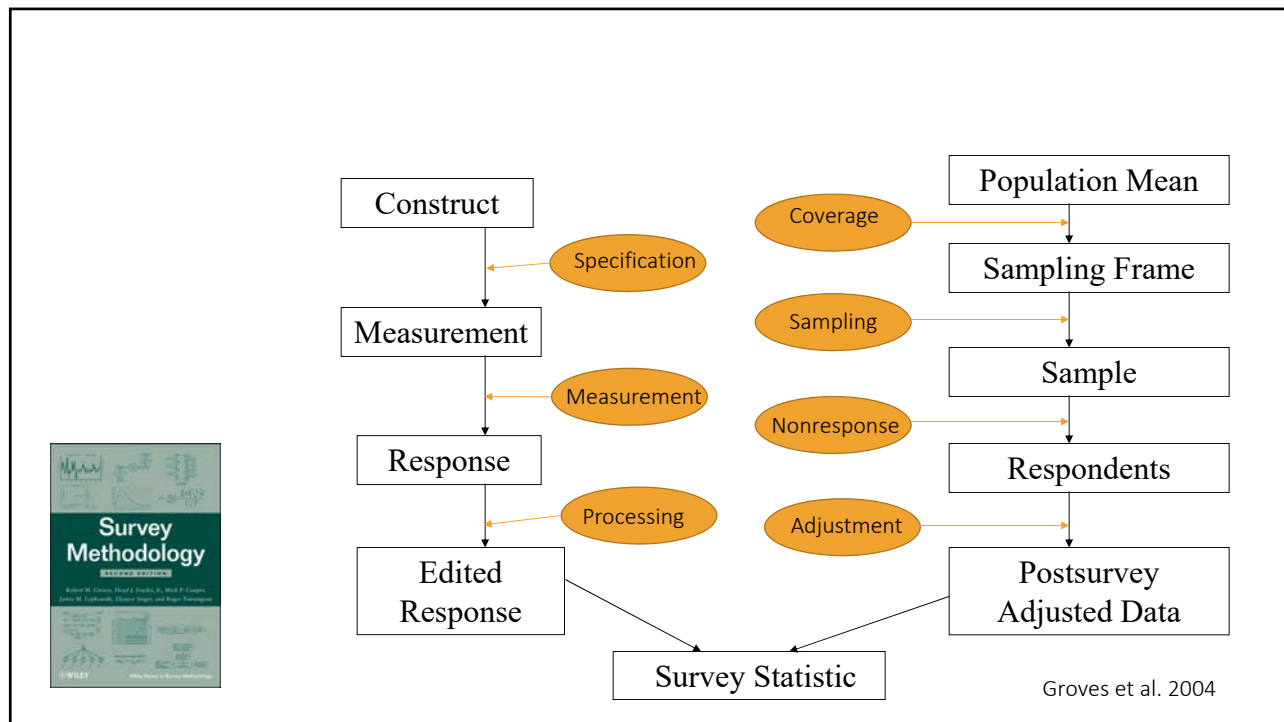


Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019

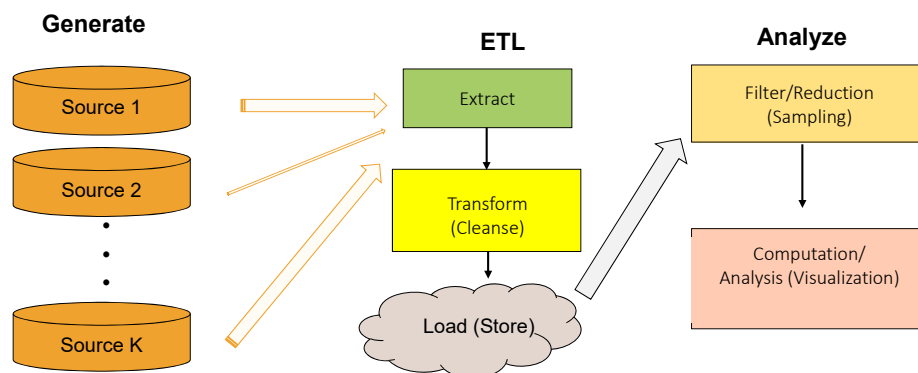


Kohler, Kreuter, Stuart (2018) Annual Review of Statistics and its Application, Vol 6.m March 2019

Daten generierende Prozess



Big Data Process Map



Source: Paul Biemer in Japac, Kreuter et al. 2015 – AAPOR Task Force Report



Google Image Search: June 3rd 2018

Search Term: "University Professor"

google translate  

[All](#) [Books](#) [News](#) [Shopping](#) [Images](#) [More](#) [Settings](#) [Tools](#)

About 557,000,000 results (0.34 seconds)

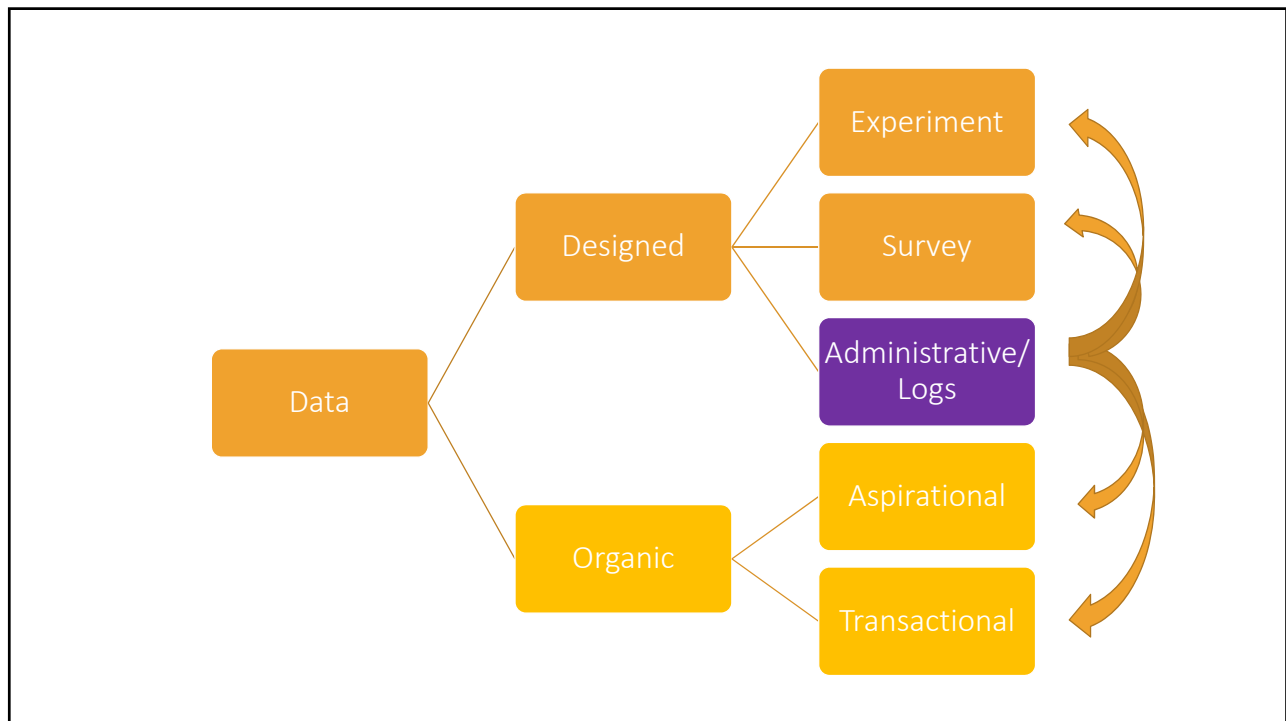
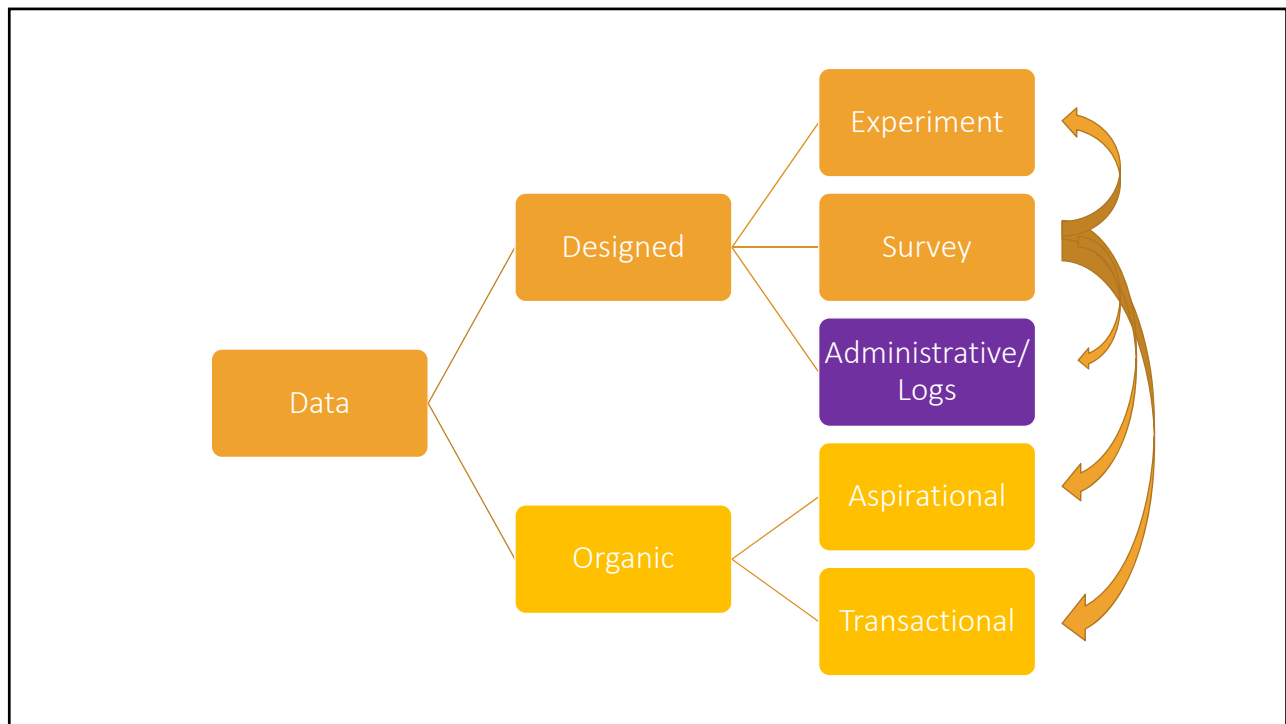
English - detected   

Spanish  

The professor is teaching the course

El profesor está impartiendo el curso.

[Open in Google Translate](#) [Feedback](#)



Beispiel – IAB: SMART

PASS – Panel (10 Jahre) + Administrative Daten

Stichprobe aus ALG II
Haushalten

Zufallsstichprobe
Haushalte (allgemein)

Auffrischung
jedes Jahr

Auffrischung
jedes Jahr

Survey jedes Jahr

Survey jedes Jahr



Meldung zur Sozialversicherung

Personalauswahl

Vericherungsnummer Personalauswahl (freiwillige Angabe)

Name Vorname Zusatz Titel

Vorname

Straße und Hausnummer (Anschrift nur bei Anmeldung und Anschriftenänderung)

Land Postleitzahl Wohnort

Grund der Abgabe Entgelt in Gleitzone Namensänderung ☐

Beschäftigungszeit
von bis Betriebsnummer des Arbeitgebers Personengruppe Mehrfachbeschäftigung ☐ Ost ☐ West ☐

Beitragsgruppen KV RV ALV PV Angaben zur Tätigkeit

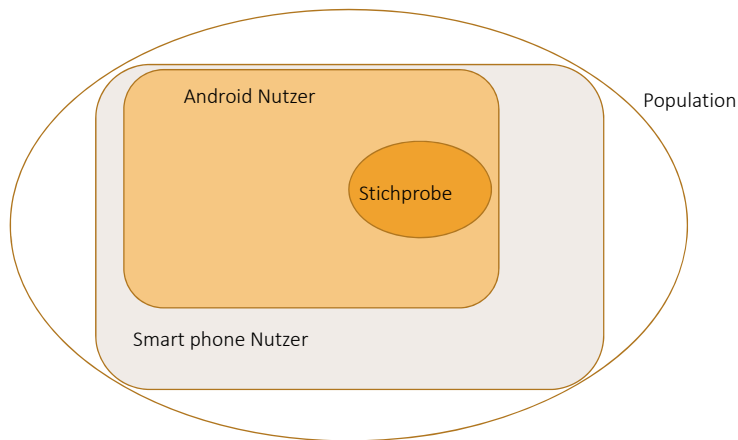
Beitragspflichtiges Bruttoarbeitsentgelt (in DM ohne Pfennige / Euro ohne Cent) DM Euro Statuskennzeichen

Wenn keine Versicherungsnummer angegeben werden kann:
Geburtsname Vorname Zusatz Geburtsort

Geburtsdatum Geschlecht männlich ☐ weiblich ☐

Trappmann M., Christoph B., Achatz J., Wenzig C. (2009) PASS: a new panel study for labour market research, Int. J. of Manpower, 30, 7, pp.765-770

Coverage // Selektion

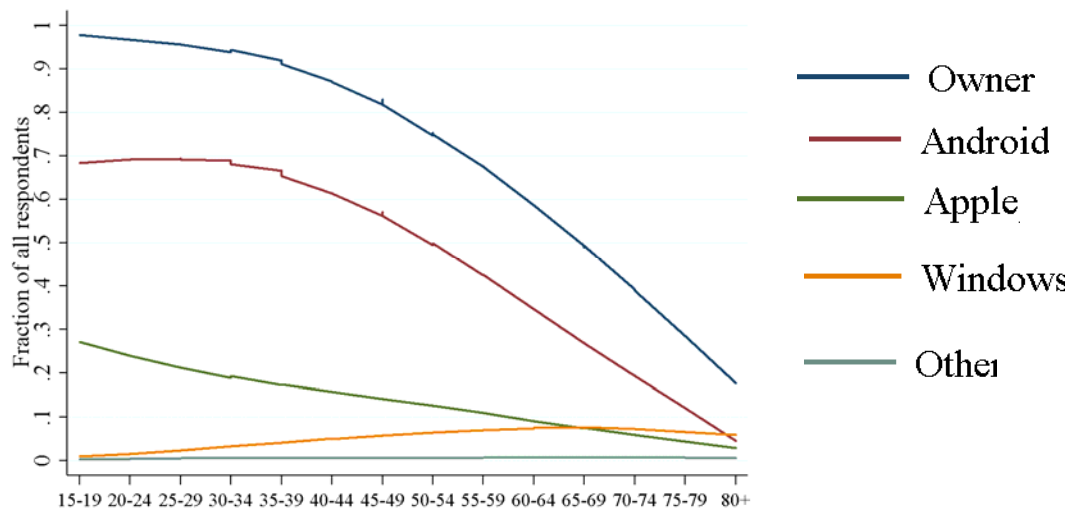


Pew Research estimates: 77% smart phone user in U.S. in 2016

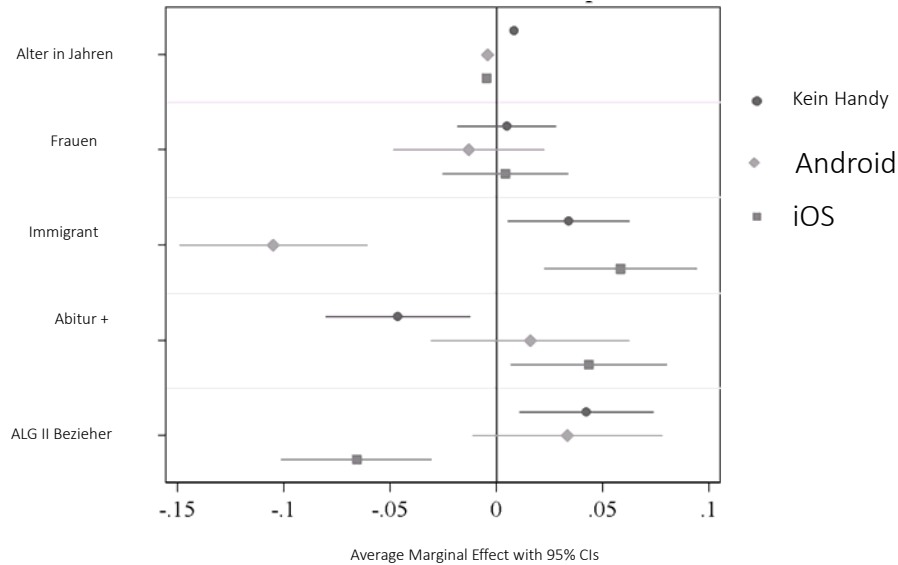
Source: Valliant R, Dever J, Kreuter F (2018): Practical Tools for Designing and Weighting Survey Samples. 2nd Edition. New York: Springer.

Example: IAB-SMART

Ownership by age groups (unweighted PASS estimates)



IAB-SMART: Predicting ownership and device type



IAB SMART: Informed Consent – Privacy - Technology

Bitte klicken Sie auf „AKZEPTIEREN“, um mit der Installation fortzufahren.

Damit alle Funktionen der App genutzt werden können, müssen Sie jeweils zustimmen:

Um die Erfassung von Nutzungsdaten zu erlauben, verschieben Sie den Schieber bitte **nach rechts**.

Der Google Play Store zeigt Ihnen nun, auf welche Daten unsere App Zugriff benötigt. Die Zustimmung zu diesen Berechtigungen ist eine technische Voraussetzung. Sie stimmen damit noch nicht der Datenerfassung zu! Die Zustimmung zur Datenerfassung erfolgt zu einem späteren Zeitpunkt innerhalb der App.

Klicken Sie auf die Pfeile, um mehr Informationen zu bekommen.

In diesem Bildschirm, die Funktionen angeben, die Datenerfassung verwenden noch Nutzungsdaten erfassen. Klicken Sie auf die Funktion, die Sie beschreiben. Klicken Sie auf das jeweilige Icon, um die Funktion zu aktivieren. Für jede Funktion, 30 Tage aktiviert, 100 Punkte. 500 Punkte können in einem Fünf-Euro-Gutschein umgewandelt werden.

Sie haben aber auch die Möglichkeit die App zu installieren und Punkte für die Beantwortung von Fragen zu erhalten, ohne den Nutzungsdatenzugriff zu erlauben.

13

14

Skills

Data Output/Access

Learn how to communicate results and distribute and store your data

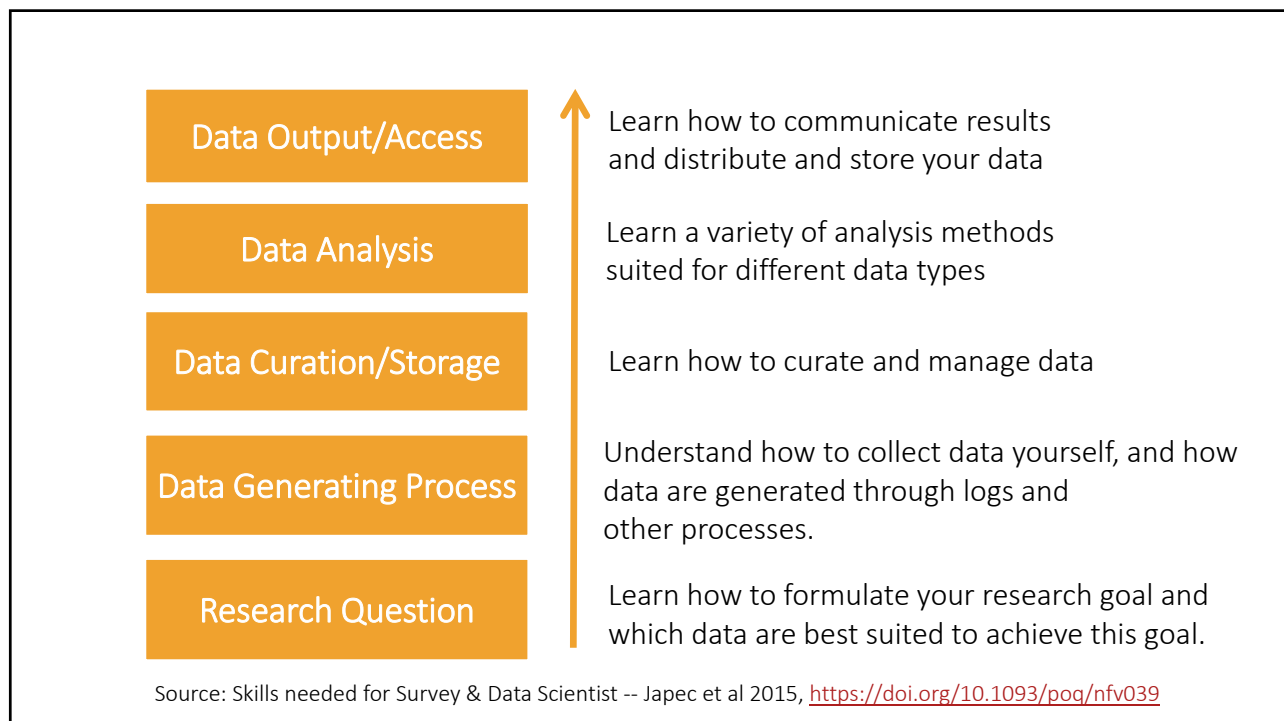
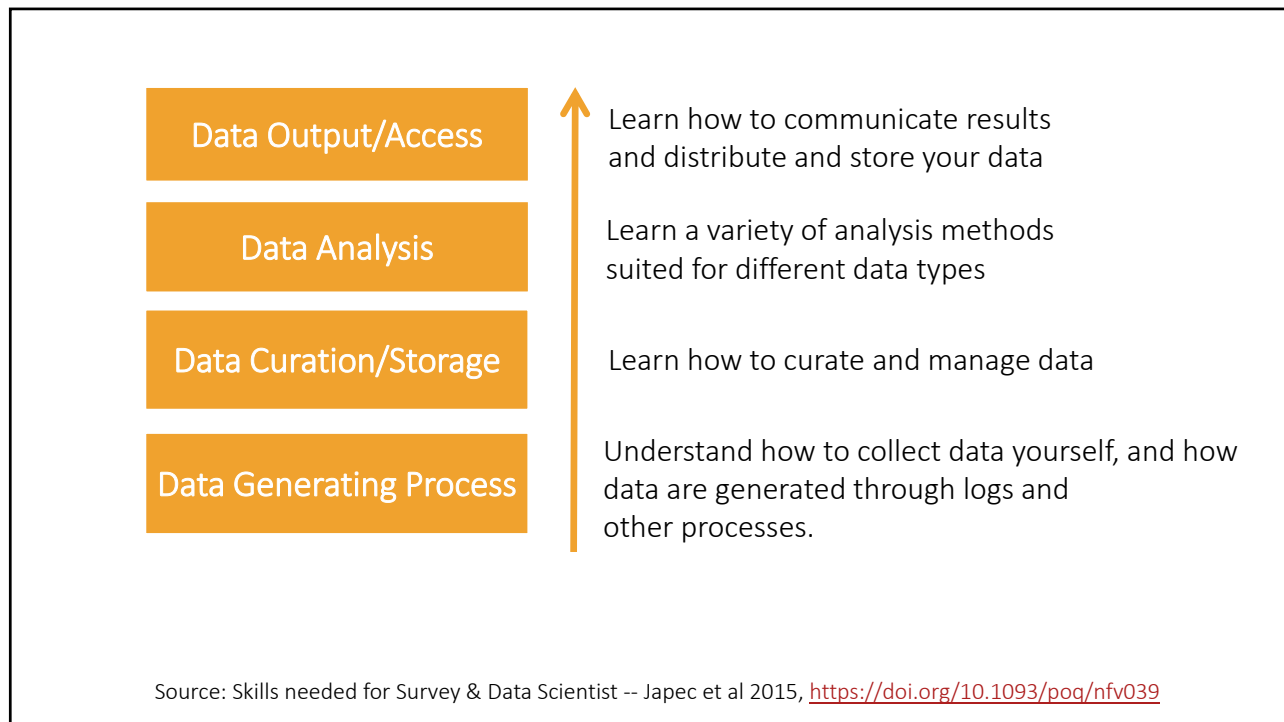
Data Analysis

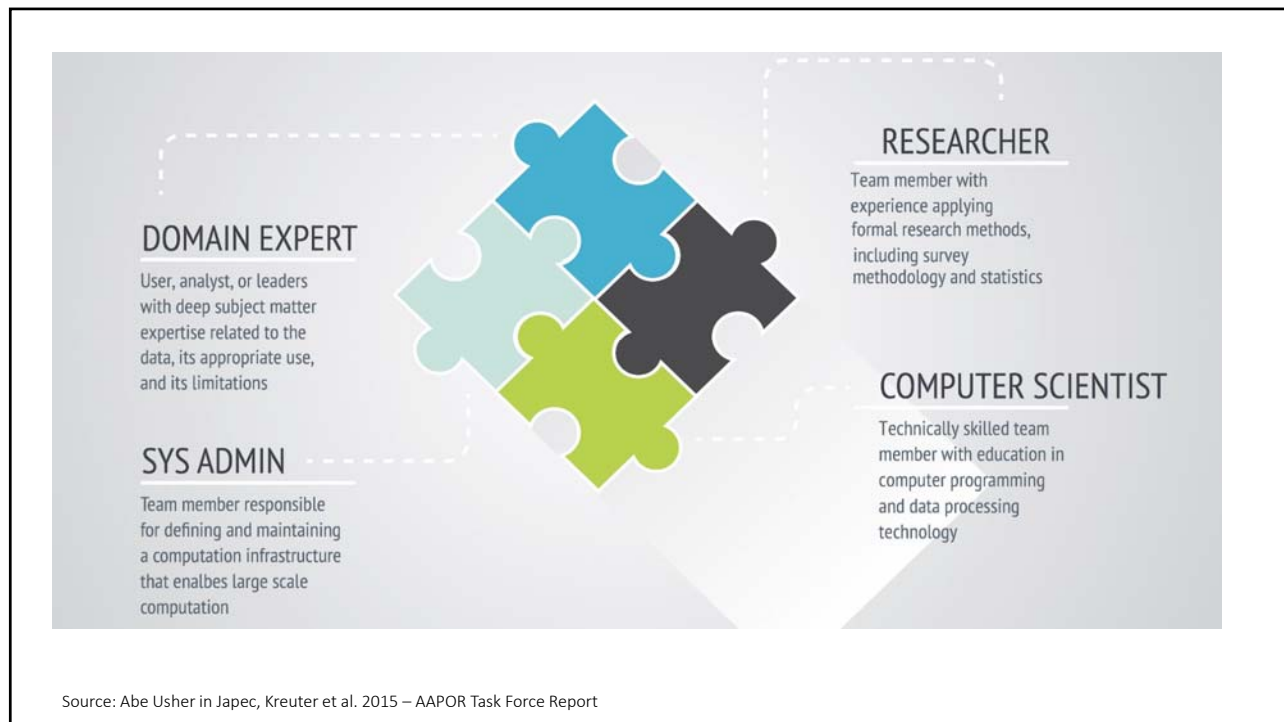
Learn a variety of analysis methods suited for different data types

Data Curation/Storage

Learn how to curate and manage data

Source: Skills needed for Survey & Data Scientist -- Japiec et al 2015, <https://doi.org/10.1093/poq/nfv039>





Zusammenfassung

1. Neue Forschungsfragen
2. Raum und Zeit im Detail
3. Seltene Populationen
4. Andere Datenformen

Kaum: weniger Kosten, schneller

Wertvoll wenn

1. ... der datengenerierende Prozess bekannt ist
2. ...Fehler untersucht werden (Framework)
3. ...konfundierendes modelliert/beseitigt wird
4. ...das Ziel der Inferenz klar ist

Danke!

fkreuter@umd.edu
www.digdeep.de

