

BIG DATA

- eine Chance für die Sozialwissenschaften

Viktor Mayer-Schönberger, Oxford

Zahlensystem
Messmethoden
normierte Einheiten



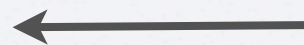
"Vermessung
der Wirklichkeit"

+

wissenschaftliche
Methode



Gesellschaft



empirische
Sozialwissenschaften

Note C.—Questions Nos. 13, 14, 22 and 23 are not to be asked in respect to persons under 10 years of age.

279

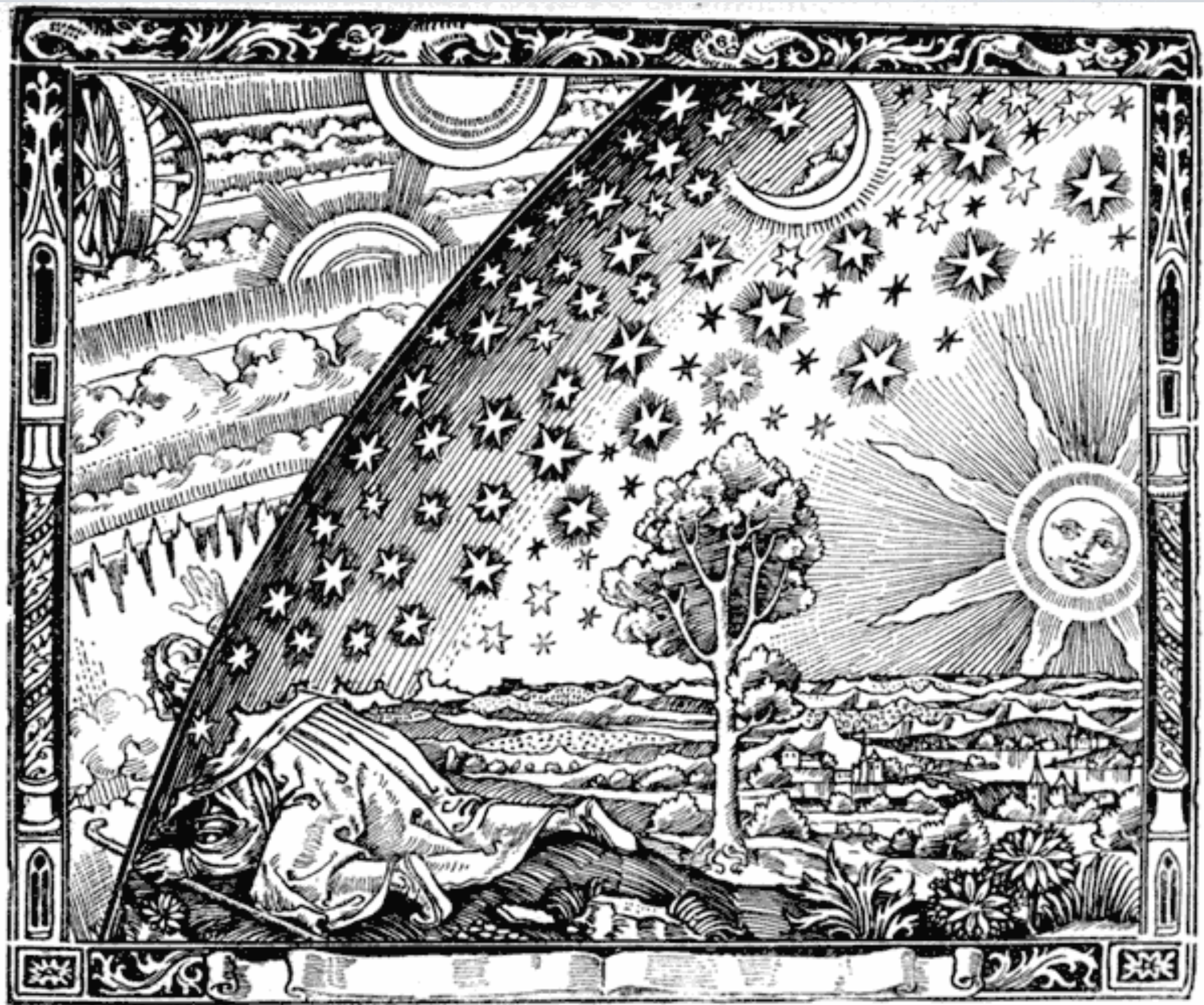
Enumerator

Friday, February 28, 14

Zufallsstichprobe

small data

- kleinstmögliche Datenmenge,
weil Sammeln und Analysieren kostspielig ist
- Betonung der Datenqualität,
weil wenig Daten gesammelt werden
- Verwendung für einen vorbestimmten Zweck,
weil Speichern kostspielig ist

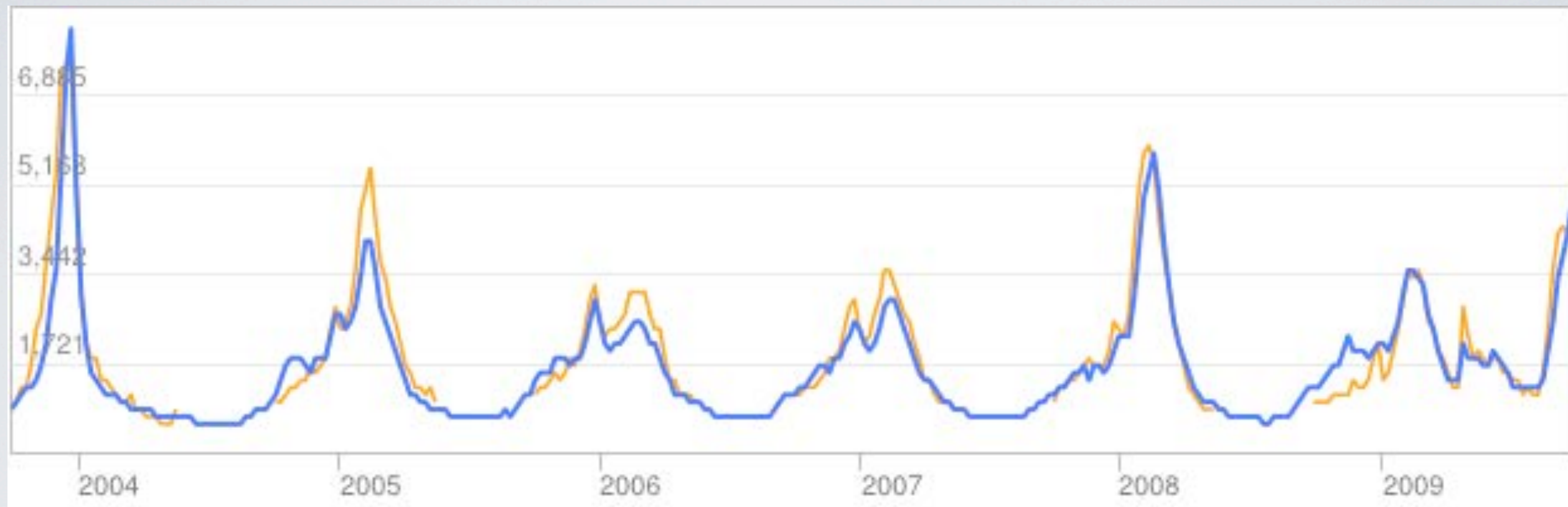


Un missionnaire du moyen âge raconte qu'il avait trouvé le point
où le ciel et la Terre se touchent...

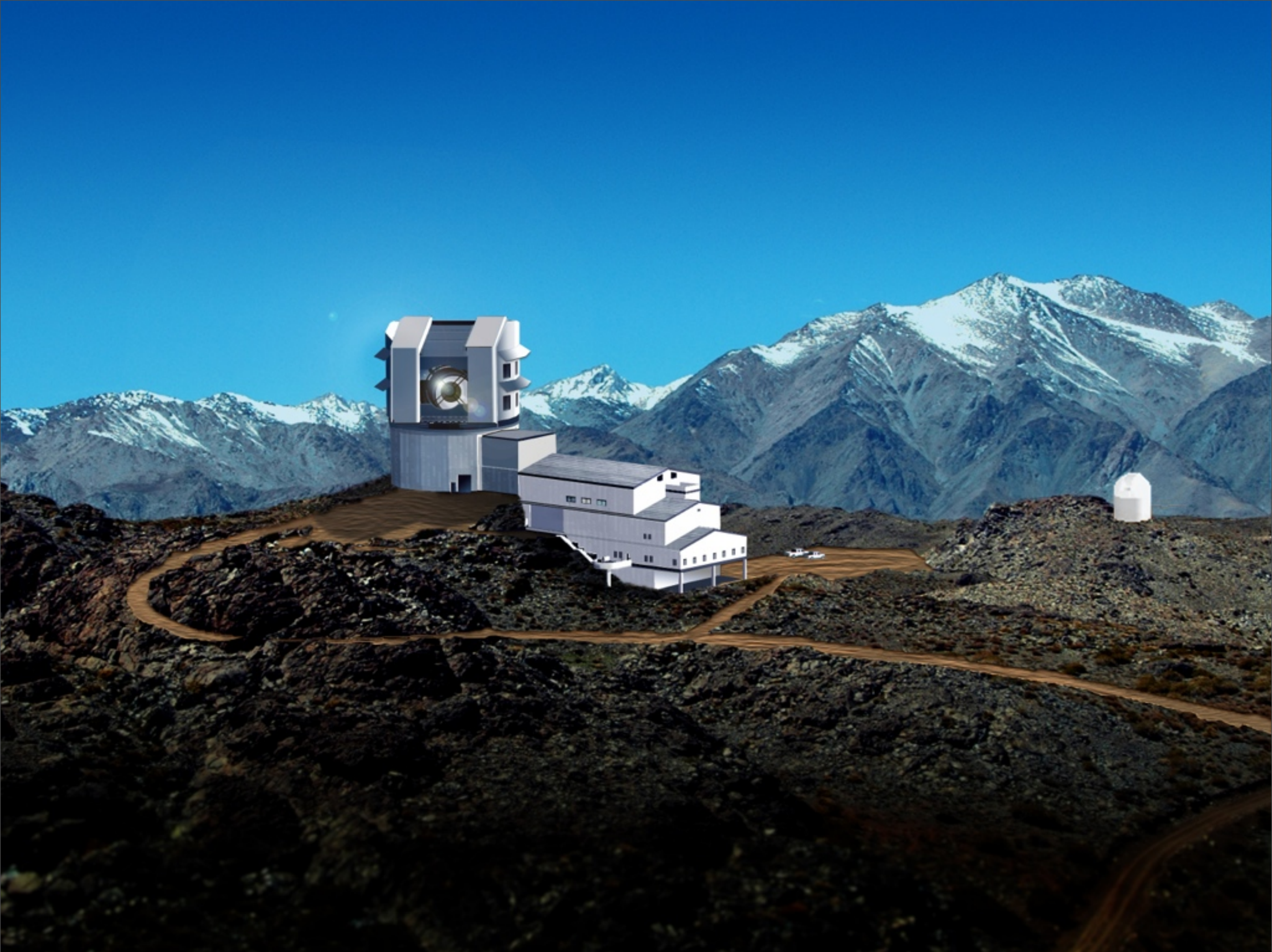
Grippe



Google Suchanfragen



Big Data





DAMALS HEUTE

2003

10 Jahre

\$1 Milliarde

2013

2-3 Tage

\$2000

Aufwand für die Sequenzierung der 3 Milliarden
Basenpaare der DNA eines Menschen

10 Millionen Fotos auf Facebook pro
Stunde hochgeladen

400 Millionen Tweets pro Tag

800 Millionen YouTube Nutzer / Monat
eine Stunde Video / Sekunde

1 Petabyte verarbeitet Google pro Tag

1987
Analog

2,62 Milliarden

Digital

0,02 Milliarden GB

Analog

Digital

2007

Analog

18,86 Milliarden GB

2007

Digital

276,12 Milliarden GB



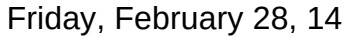
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR **MEHR** MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR
MEHR MEHR MEHR MEHR MEHR MEHR MEHR

“UNSCHARF”

maschinelle Übersetzung

Kausalität | Korrelation

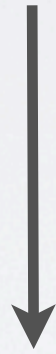
Demut gegenüber komplexer Wirklichkeit



Theorie

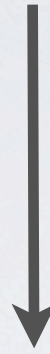


Hypothese

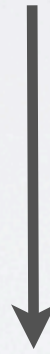


Daten

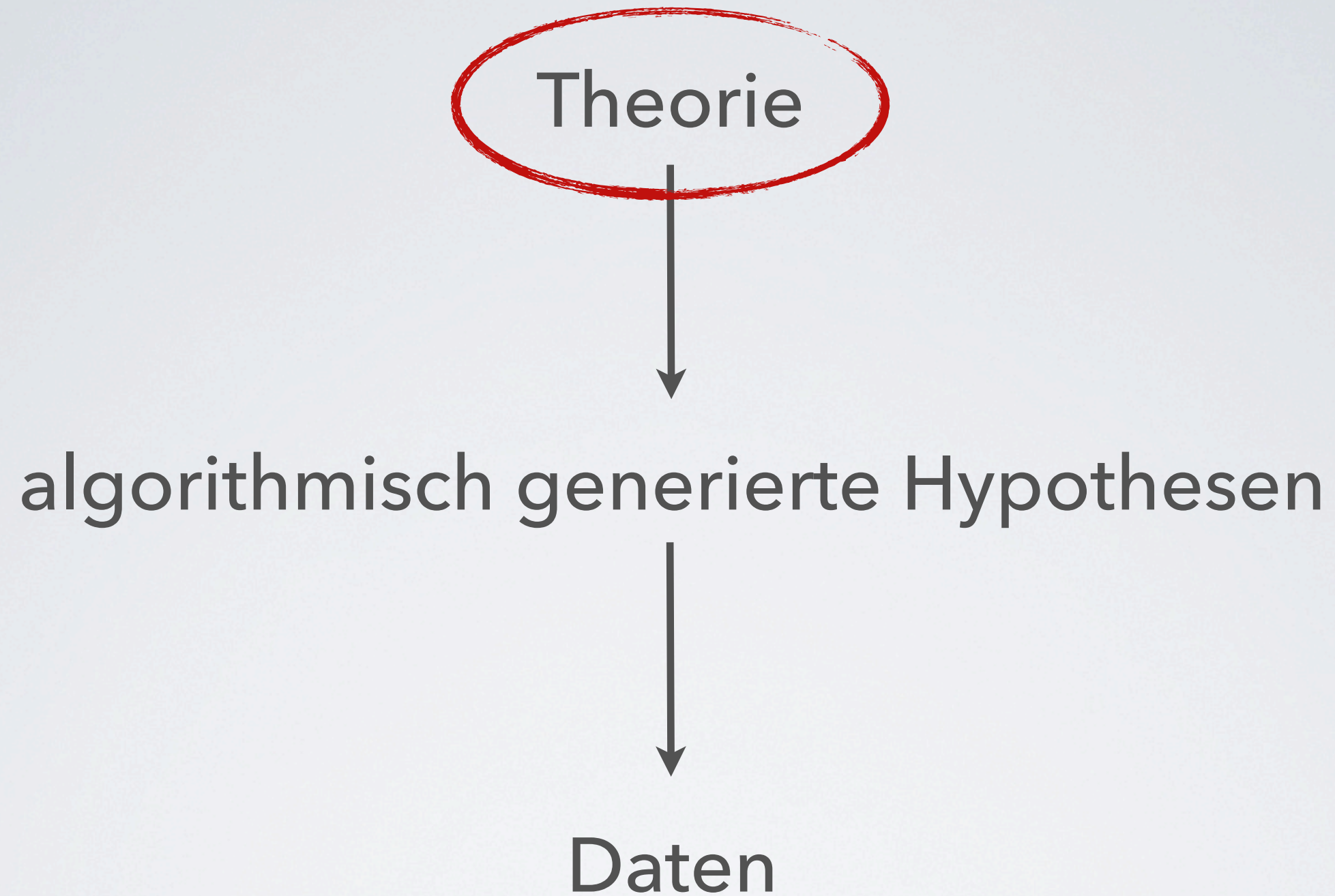
Theorie



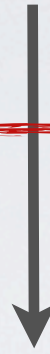
algorithmisch generierte Hypothesen



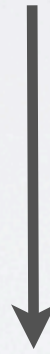
Daten



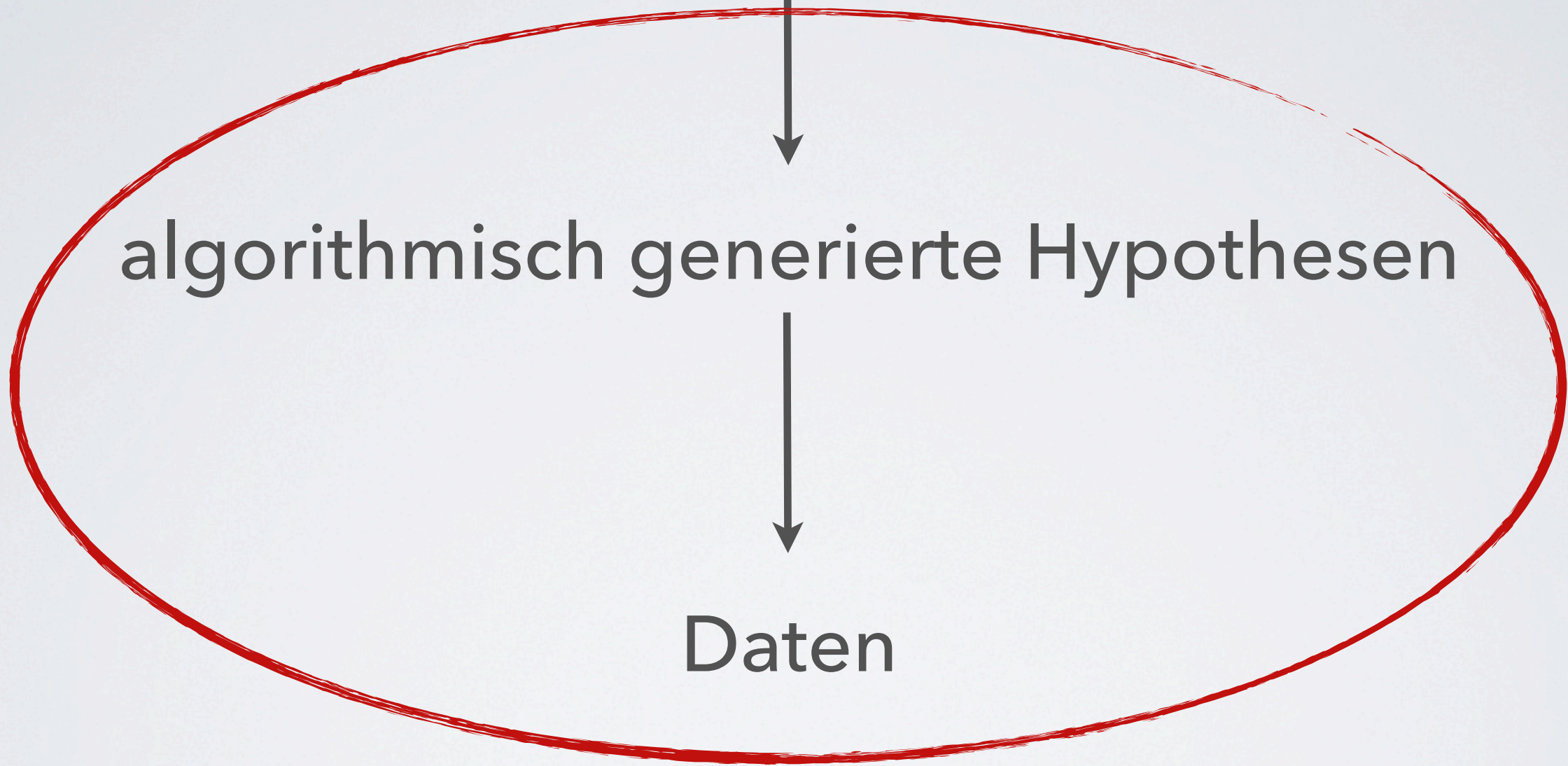
Theorie



algorithmisch generierte Hypothesen



Daten

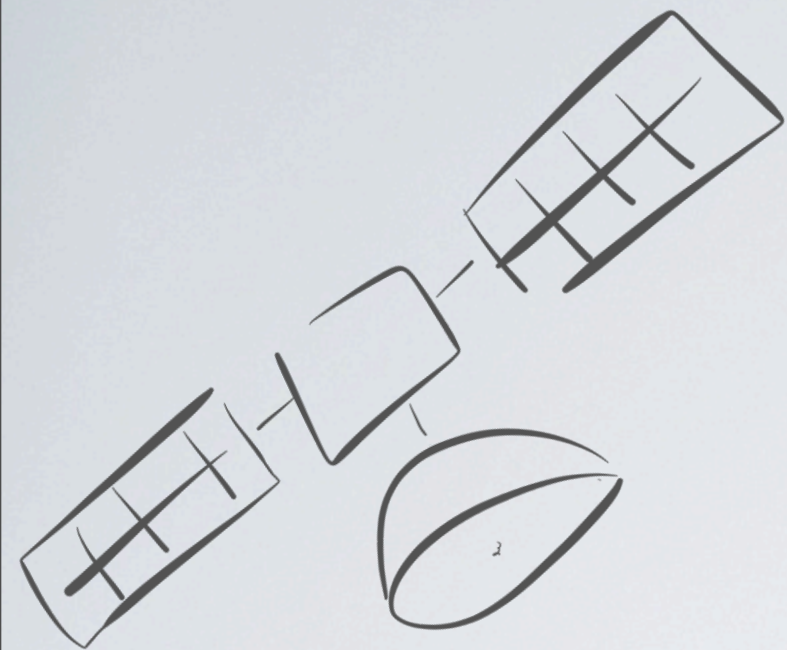


“spurious signals” vs Kausalitäts-Bias

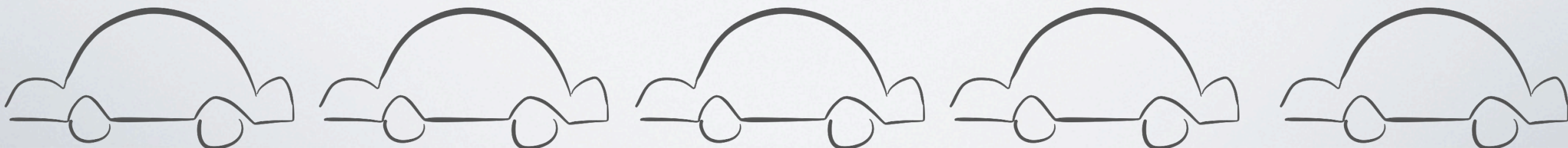
Trends

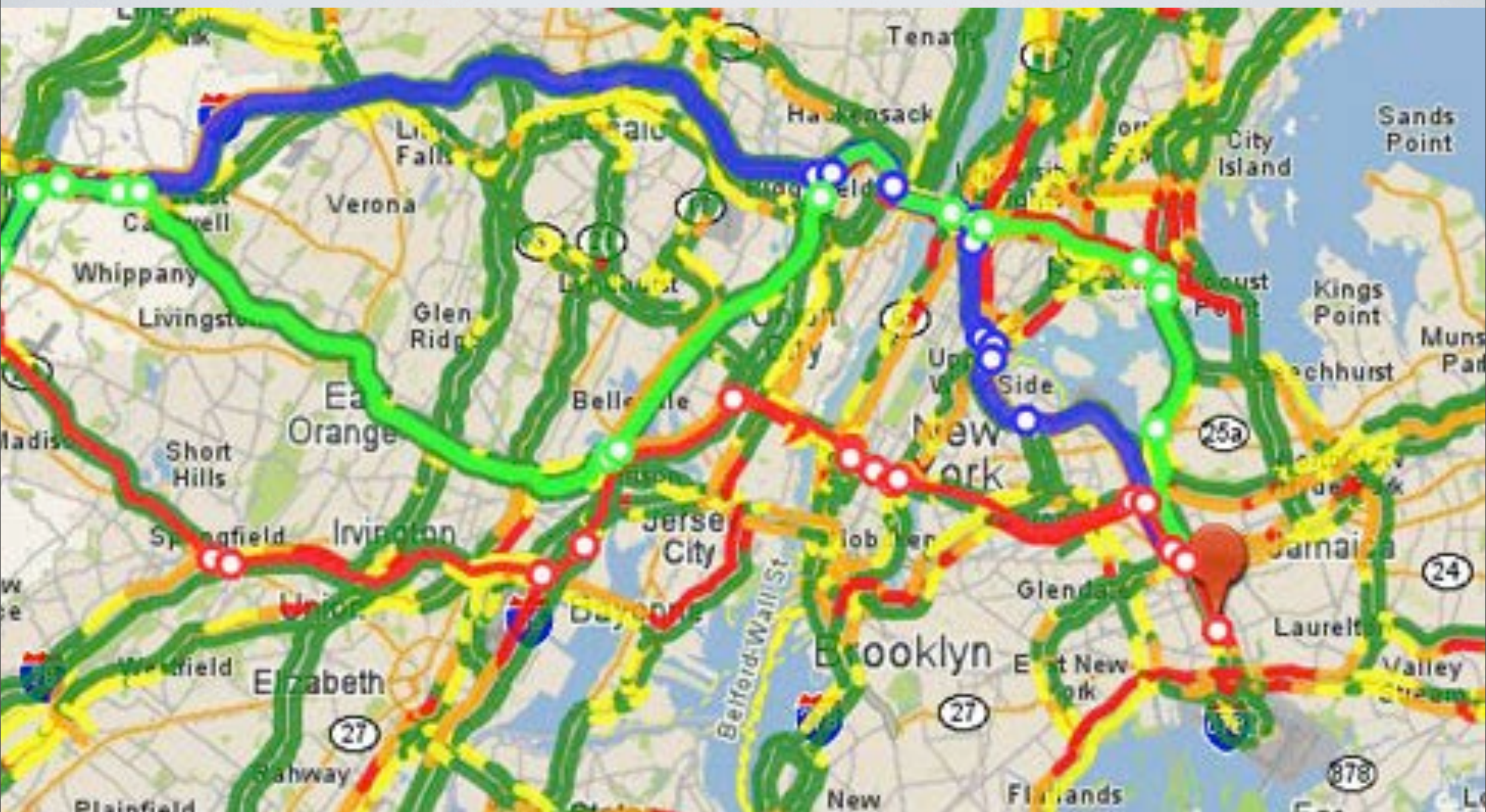
(1)

*Sozial*wissenschaften

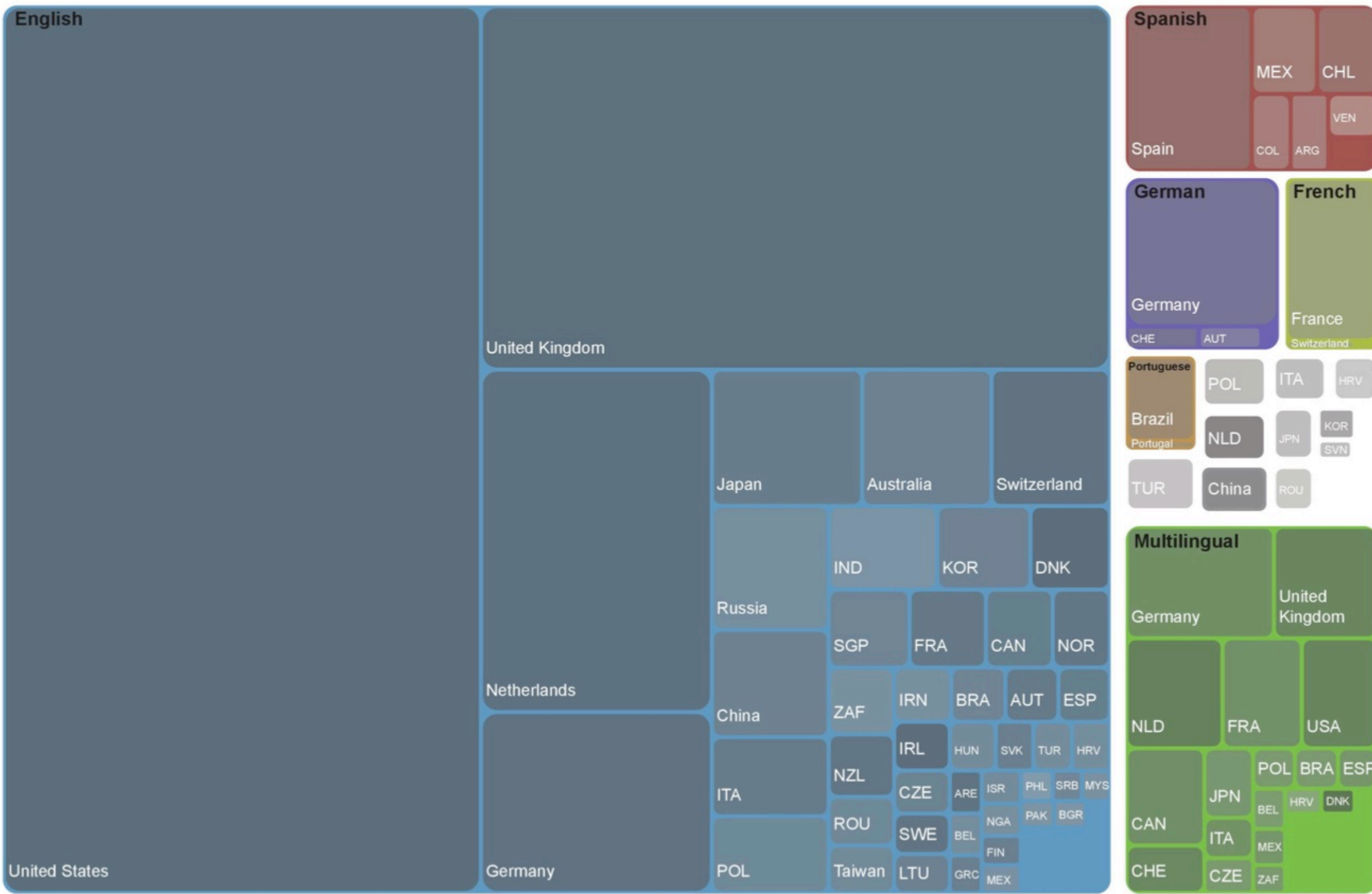


inrix

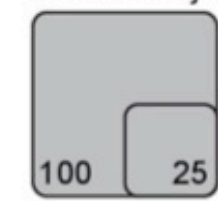




PriceStats



Number of journals



Impact factor



(2)

Naturwissen-
schaften



*Sozial*wissenschaften

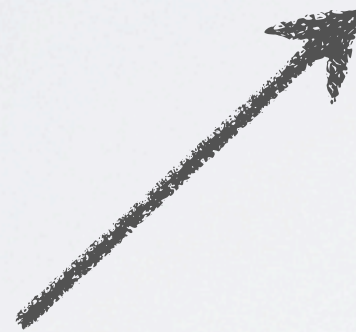
"Systems Biology"

(3)

Naturwissen-
schaften



*Sozial*wissenschaften



Geisteswissen-
schaften

(4)

Daten als Infrastruktur

Herausforderungen

- Datenschutz

Effektivität rechtlicher/technischer Mechanismen

- Grenzen der Vorhersage

"black swans" und "Minority Report"

- "Diktatur der Daten"

"spurious signals" und Kausalitätsmißbrauch



DANKE

Viktor Mayer-Schönberger

Viktor.MS@oii.ox.ac.uk

@Viktor_MS

big-data-book.com