

# 'Big Data' — Endlich die Antwort auf alles?

Prof. Dr. Diego Kuonen, CStat PStat CSci

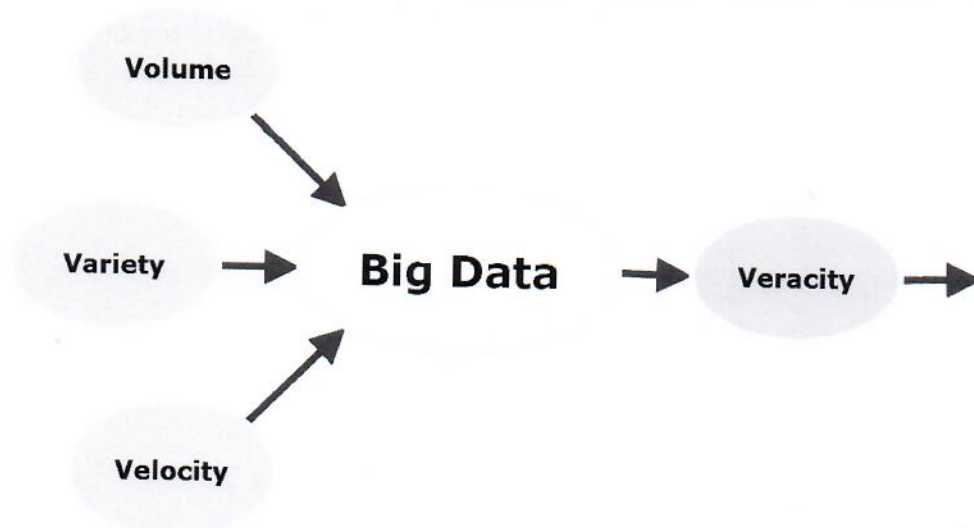
Statoo Consulting, Bern & Universität Genf

@DiegoKuonen + kuonen@statoo.com + www.statoo.info

VORTRAG für die VSÄG  
SMVS / VSÄG  
Société Médicale du Valais  
Walliser Ärztesgesellschaft  
JUBILÄUMSTÄGUNG  
175 Jahre VSÄG

Zentrum Sosta, Susten — 15. November 2018

## Entmystifizierung vom *Big Data* (Massendaten) Hype



→ Die **'Veracity'** (d. h. das Vertrauen in die Daten), inklusive Zuverlässigkeit ('Qualität über die Zeit'), Fähigkeit (der 'Messsysteme'), Validität, und die damit verbundene **Datenqualität** sind der Schlüssel!

---

# Entmystifizierung der zwei Vorgehensweisen der Analytik

---

## Statistik, Datenwissenschaft und deren Verbindung

---

◇ Die Statistik beschäftigt sich traditionell mit der Analyse von ('experimentellen') **Primärdaten**, welche für statistische Zwecke erhoben wurden, um die Gültigkeit von bestehenden Annahmen ('Ideen') anhand der Umsetzung theoretischer Konzepte zu erklären und zu überprüfen.

↪ Primäranalyse, d. h. erklärende und bestätigende Analytik.

↪ **'Ideenbewertung oder -prüfung'**.

↪ Analytik-Paradigma: **'deduktives Vorgehen'**, das von einer Idee (Theorie) ausgeht, d. h. '*idea (theory) first*'.

---

S+3+00

Copyright © 2001–2018, Statoc Consulting, CH. Alle Rechte vorbehalten.

◇ Datenwissenschaft (*Data Science*) — ein Rebranding von *Data Mining* — beschäftigt sich dagegen typischerweise mit der Analyse von ('beobachteten' oder 'gefundenen') **Sekundärdaten**, welche aus anderen Gründen erhoben wurden (oft nicht unter der Aufsicht des Ermittlers) und zum Entwickeln neuer Annahmen (Ideen) genutzt werden können.

↪ Sekundäranalyse, d. h. explorative und prädiktive Analytik.

↪ **'Ideengenerierung'**.

↪ Analytik-Paradigma: **'induktives Vorgehen'**, das von den Daten ausgeht, d. h. '*data first*'.

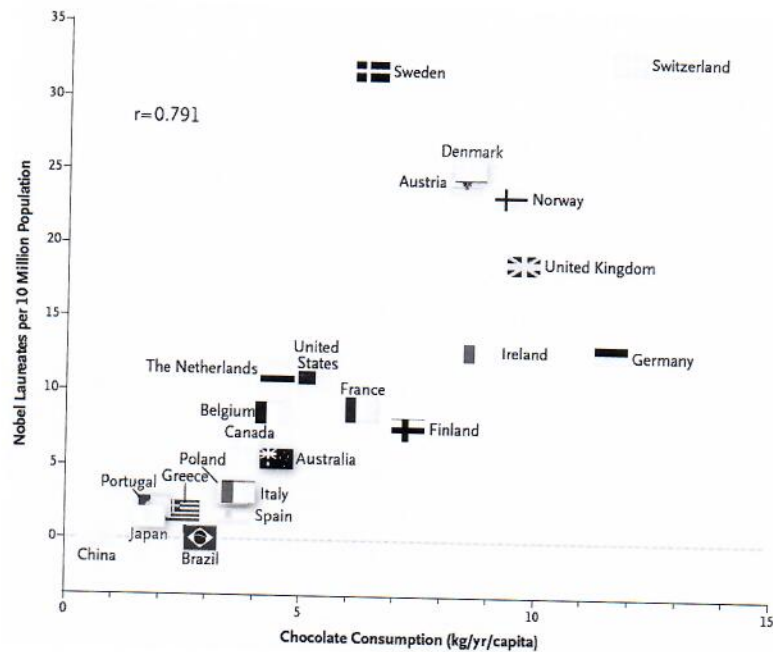
---

S+3+00

Copyright © 2001–2018, Statoc Consulting, CH. Alle Rechte vorbehalten.

19

## '(Zufällige) Korrelation bedeutet nicht Kausalität!'



Quelle: Messerli, F. H. (2012). 'Chocolate consumption, cognitive function, and Nobel laureates'. *The New England Journal of Medicine*. 367, 1562–1564.

s+a+oo

Copyright © 2001–2018, Statoo Consulting, CH. Alle Rechte vorbehalten.

11

'Jede Behauptung, die aus einer Sekundäranalyse stammt, ist höchstwahrscheinlich falsch.'

S. Stanley Young und Alan Karr, 2011

s+a+oo

Copyright © 2001–2018, Statoo Consulting, CH. Alle Rechte vorbehalten.

12