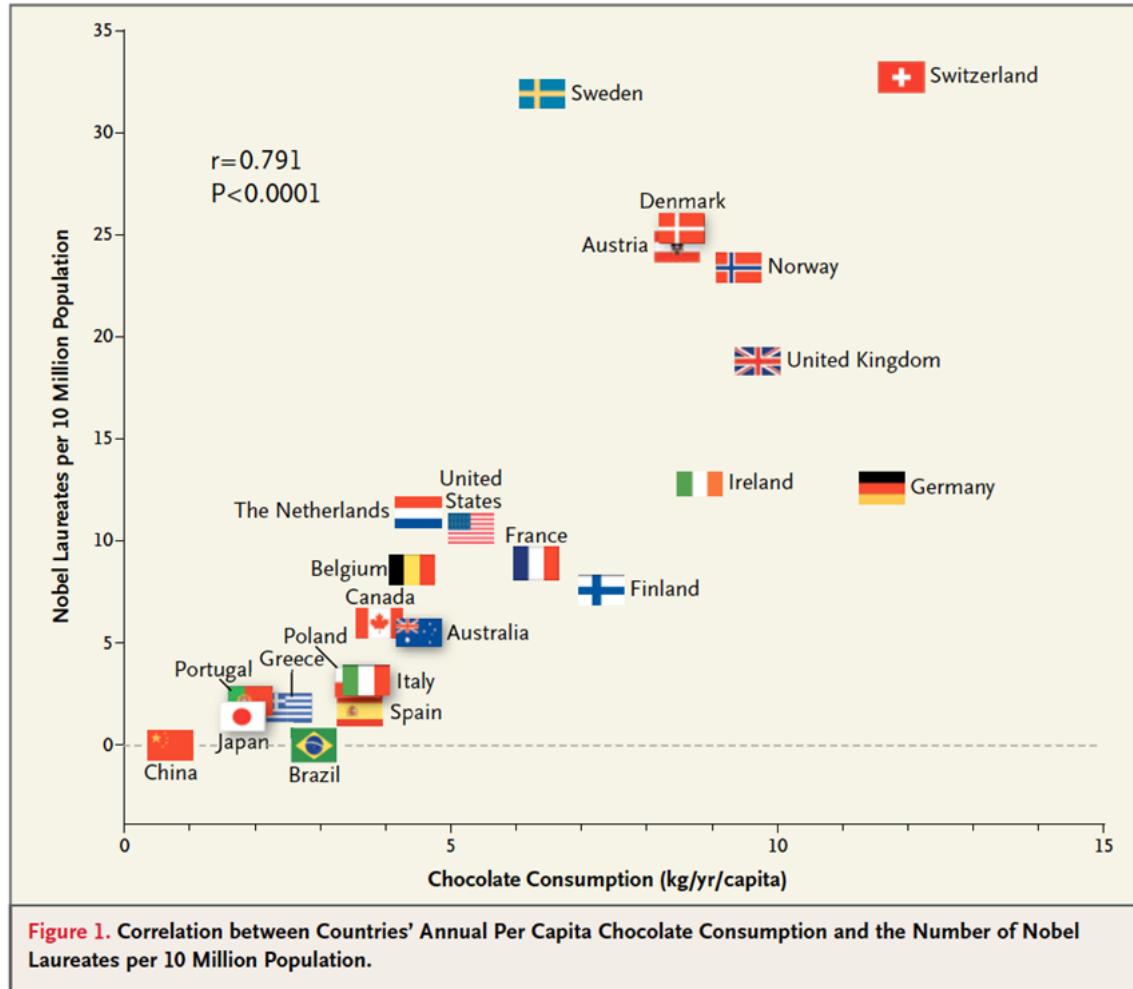


Peter Farago

## **Big Data in den Sozialwissenschaften**

Einleitende Bemerkungen

# Big Data – so nicht !



# 5 Schritte

1. Was sind Big Data ?
2. Wofür kann man Big Data in der Forschung einsetzen ?
3. Wie kann man Big Data in der Forschung einsetzen ?
4. Was sind bislang die Erkenntnisse und Erfahrungen ?
5. Wie weiter ?

# Was sind Big Data ?

## Eigenschaften:

- Umfangreich (*volume*)
- Schnell (*velocity*)
- Unübersichtlich (*variety or variability*)

(nach: Mick P. Couper: *Is the Sky Falling?* SRM 7/3, pp. 145-156, 2013)

# Was sind Big Data ?

## Quellen:

- Administrative Daten
- Transaktionsdaten
- Social Media-Daten

*(nach: Mick P. Couper: Is the Sky Falling? SRM 7/3, pp. 145-156, 2013)*

# **Wofür kann man Big Data in der Forschung einsetzen ?**

## **Einschränkende Rahmenbedingungen:**

- Nicht für Forschungszwecke generiert
- Beruhen weder auf einer wissenschaftlichen Zielsetzung noch auf wissenschaftlichen Methoden
- Oft unvollständig, fehlerhaft und/oder verzerrt
- Es ist fraglich, ob und inwieweit generalisierbare Schlüsse möglich sind
- Zugang erschwert, besonders wenn die Daten in privatem Besitz sind

# Wofür kann man Big Data in der Forschung einsetzen ?

## Grundsätzlich 2 Einsatzarten:

- Auswertungen und Analysen, die sich ausschliesslich auf Big Data stützen; Beispiel Twitter-Daten
  - Ergänzung anderer Datentypen; Beispiel: Survey-Daten
- ➔ In jedem Fall methodologisch anspruchsvoll, voraussetzungsreich und ressourcenintensiv

# Wie kann man Big Data in der Forschung einsetzen ?

- Auswahl: Ist Masse gleich Klasse ?
- Bereinigung, Verdichtung: Reduktion von Komplexität
- Verknüpfungen: record linkage
- Analyse: Komplexe Verfahren benötigen bei grossen Datenmengen hinreichende Rechenkapazität
- Persönlichkeits- und Datenschutz: weitreichende Anonymisierung muss gewährleistet sein

→ Kann Data Science weiterführen ?