

Le 'Big Data' — Enfin une réponse à tout?

Prof. Dr. Diego Kuonen, CStat PStat CSci

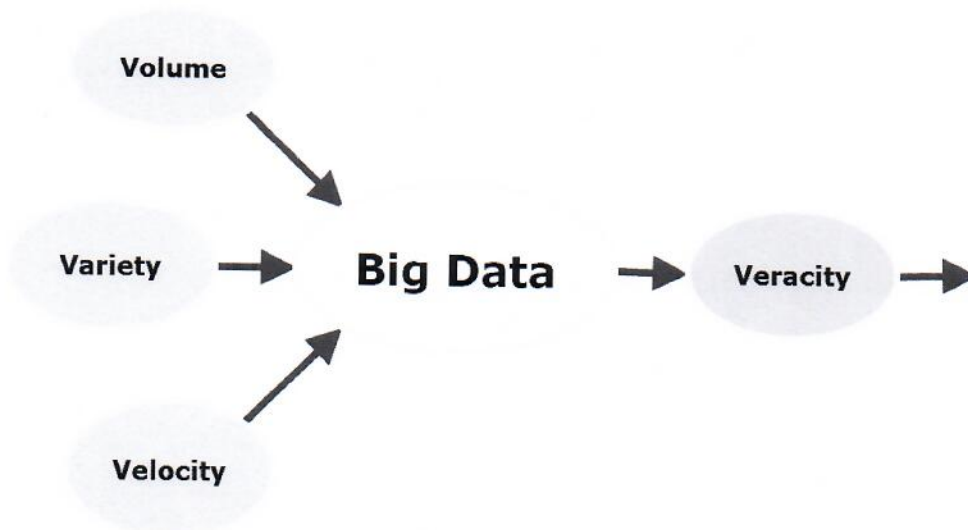
Statoo Consulting, Berne & Université de Genève

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

Conférence donnée pour SMVS
SMVS / VSÄG
Société Médicale du Valais
Walliser Ärztesgesellschaft
à l'occasion des 175 ans SMVS

L2 Souste, Susten — 15 novembre 2018

Démystification du 'hype' *Big Data* (mégadonnées)



↪ La **'Veracity'** (c.-à-d. la confiance dans les données), qui comprend la fiabilité ('qualité au fil du temps'), capabilité (des 'systèmes de mesure'), validité, et la **qualité des données** associée sont la clé!

Démystification des deux approches de l'analyse

Statistique, science des données et leur connexion

◇ La statistique est concerné traditionnellement par l'analyse des **données primaires** ('expérimentales') qui étaient collectées à des fins statistiques pour expliquer et vérifier la validité d'hypothèses ('idées'), c.-à-d. par l'opérationnalisation d'une théorie.

↪ Analyse primaire, c.-à-d. analyse explicative et confirmatoire.

↪ **'Évaluation ou vérification d'idées'**.

↪ Paradigme d'analyse: **'raisonnement déductif'** à partir d'une idée (théorie), c.-à-d. '*idea (theory) first*'.

S+B+00

Copyright (c) 2001–2019, Statae Consulting, CH. Tous droits réservés.

◇ La science des données (*Data Science*) — une nouvelle appellation de *Data Mining* — s'intéresse typiquement à l'analyse des **données secondaires** ('observées' ou 'trouvées') qui étaient collectées à d'autres fins (et qui souvent ne sont pas sous le contrôle de l'enquêteur) et qui peuvent être utilisées pour développer des nouvelles hypothèses (idées).

↪ Analyse secondaire, c.-à-d. analyse exploratoire et prédictive.

↪ **'Génération d'idées'**.

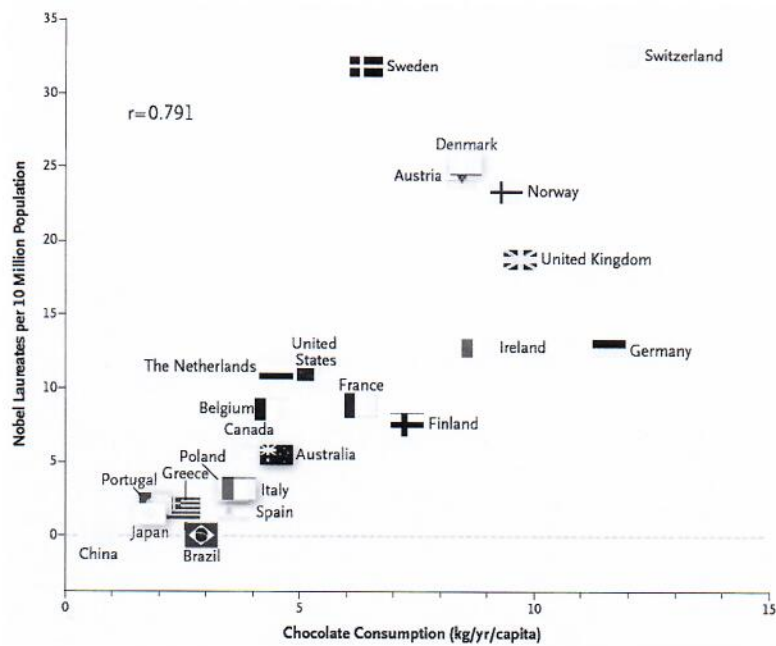
↪ Paradigme d'analyse: **'raisonnement inductif'**, dans lequel les données sont premières, c.-à-d. '*data first*'.

S+B+00

Copyright (c) 2001–2019, Statae Consulting, CH. Tous droits réservés.

10

'Corrélation (aléatoire) ne signifie pas causalité!'



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

S+a+00

Copyright © 2001–2018, Stattoo Consulting, CH. Tous droits réservés.

11

'Toute affirmation provenant d'une analyse secondaire est probablement fausse.'

S. Stanley Young et Alan Karr, 2011

S+a+00

Copyright © 2001–2018, Stattoo Consulting, CH. Tous droits réservés.

12