



Open your mind. LUT.

Lappeenranta **University of Technology**



# Viktor Mayer-Schönberger and Kenneth Cukier BIG DATA - Chapter 2 - MORE

Presentation by  
Markus Salminen

# Big data



Open your mind. LUT.  
Lappeenranta University of Technology

- Requires three shifts of mindset
  - **“The ability to analyze vast amounts of data about a topic rather than be forced to settle for smaller set”**
  - “Willingness to embrace data’s real-world messiness rather than privilege exactitude”
  - “Growing respect for correlations rather than a continuing quest for elusive causality”

# Why MORE



Open your mind. LUT.  
Lappeenranta University of Technology

- $n = x$  vs  $n = \text{all}$ 
  - Modern technology makes it possible to chew through ALL the data
  - Rapidly growing amount of data
  - Why settle for less if you can have more?
- Sample errors;  $n=1100 \rightarrow \sim 3\%$ 
  - Requires randomness
  - Randomness is not always possible (subgroups)
  - Sometimes resolution is not enough
- Sampling is like compressing (like an analog picture)
- Sometimes outliers are what you are looking for (bigger picture)
- What you are looking for is not always what you are looking for (sumo case)
- Humans have to use samples (not enough processing power)

# Big is not always **BIG**

- Google Flu Trends
  - Hundreds of millions of mathematical algorithms
  - Billions of data points
- Human DNA
  - three billion DNA base pairs
  
- Sumo wrestling
  - ~64 000 matches
  - Fewer bits than in a digital photo
- Like a fishing trip
  - Sample: how many
  - Big data: how many, where, how, what, who ->

