Privacy vs Profit

Big Data's Sword of Damocles

Alexander Loffler
Principal Technology Architect
TELUS Security

November 2016 (v2.4-ubc)

Agenda

- Public vs Private Sectors
- The Lure of Big Data Analytics
- The Anonymization Problem
- Correlation vs Causation
- Conclusions

The Sword of Damocles

Public vs Private Sectors

	Primary Drivers	Controls
Public Sector	• Safety and Security of its Citizens • Forecasting / Planning	• Legislation • Transparency • Accountability • 1 Entity per region
Private Sector	• Customer Experience • Profit	• Legislation • Internal Governance • EULAs / Terms of Service • Large numbers usually spanning multiple geographies

The Lure of Big Data Analytics

- Big Data is the anti-thesis of Privacy
- Increased efficiency of product/service delivery
- More accurate supply/demand forecasting
- Personalized product/service offerings
- Improved responsiveness to unexpected change
- Monetizing existing datasets

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay.“ - Sherlock Holmes

But we have a choice…

The most popular EULA/ToS clauses today:

1. "We may change all or part of the EULA at any time. EULAs are subject to change without notice. Posting updated terms on our website will constitute your acceptance of the changed terms."
2. "Any user content or data gathered by any part of this Product/Service, automatically grants [the Company] an irrevocable, perpetual, non-exclusive, transferable, fully paid, worldwide license (with the right to sublicense) to use, copy, publicly perform, publicly display, reformat, translate, excerpt (in whole or in part) and distribute such data for any purpose, commercial, advertising, or otherwise..."

"If you are not paying for the product, you are the product." – Anonymous

The Anonymization Problem

- Tokenization
  - Substitution of sensitive data
- Generalization
  - Aggregation, Thresholding, Quantizing & Perturbation
- Context
  - What additional datasets will be combined?
  - How will the data be used?

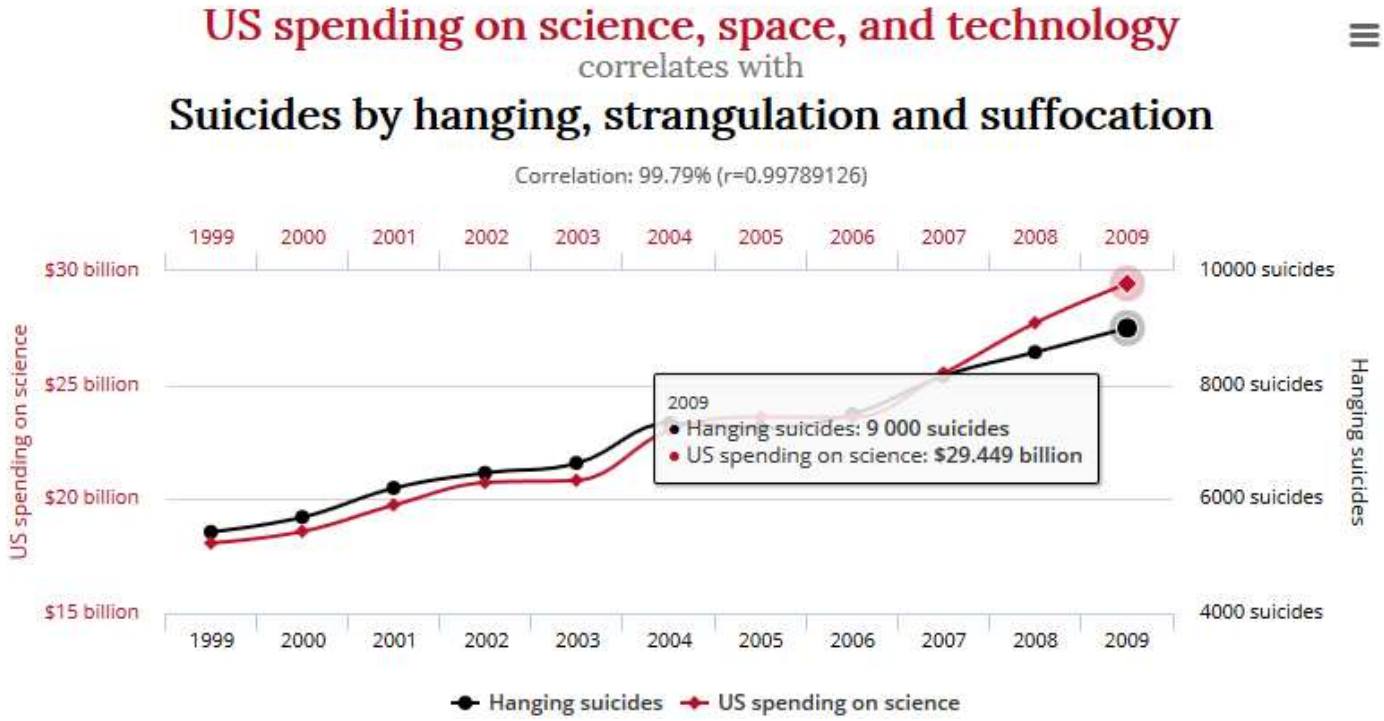
Correlation vs Causation

- **Correlation** – "*the measure of a relationship between two data sets.*" – Wikipedia

- **Causation** – "*the agency or efficacy that connects one process (the cause) with another (the effect) where the first is understood to be partly responsible for the second*" – Wikipedia

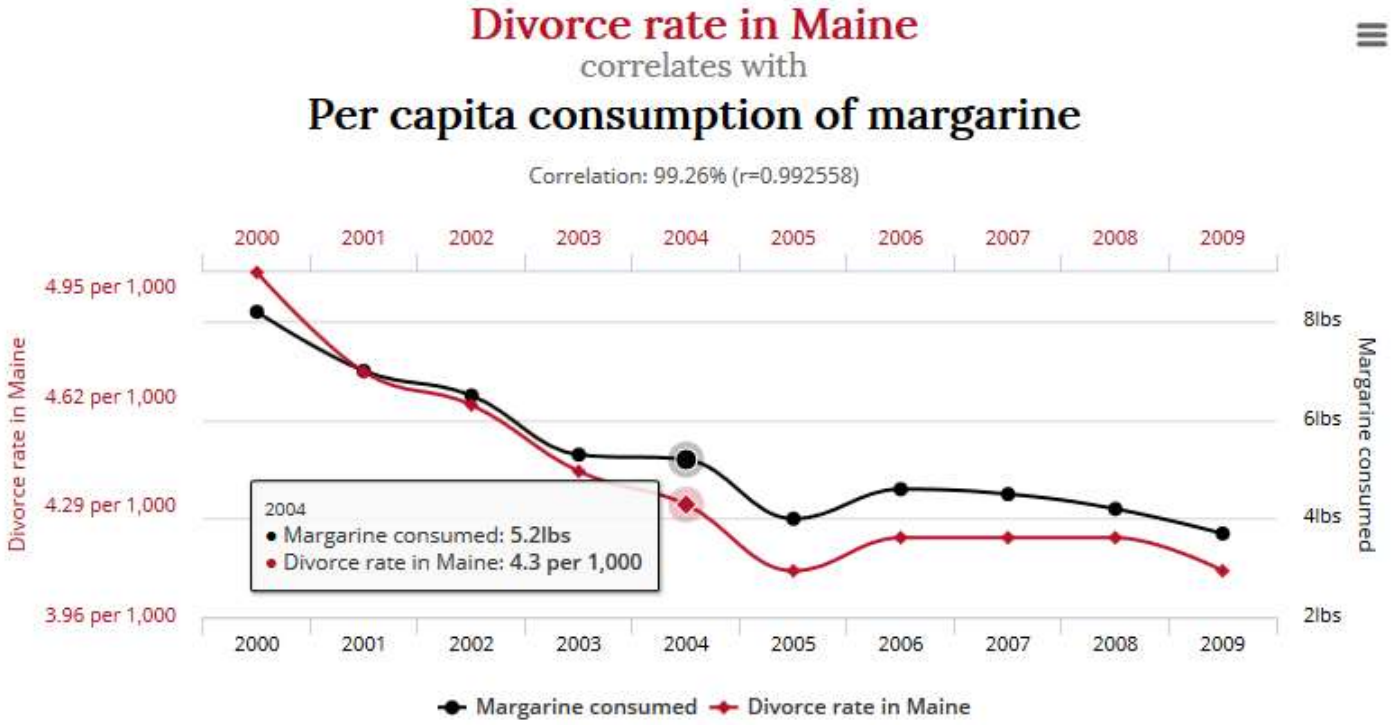
- We are **hardwired to assume** Correlation implies Causation
  - Pavlov's Dogs, Skinner Box's, etc.

Correlation #1



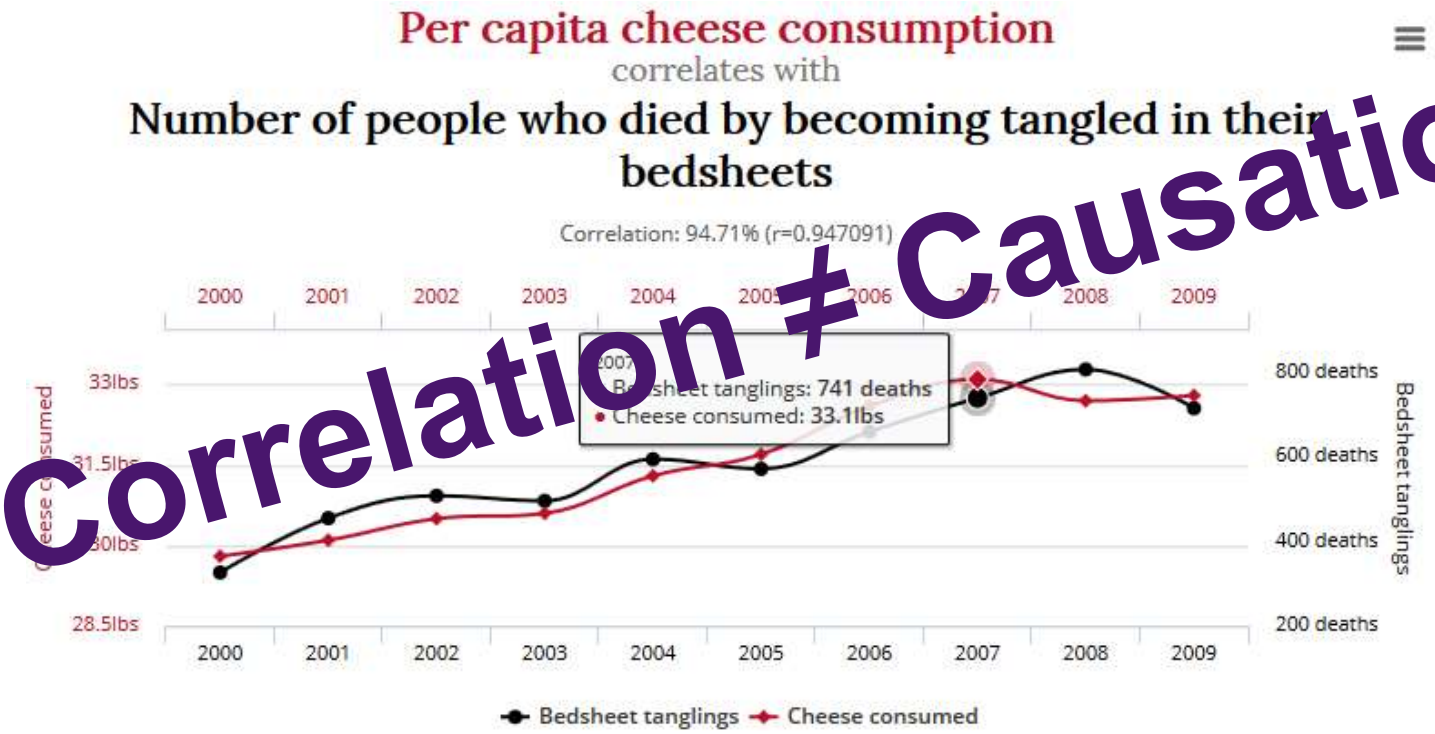
Source: Spurious-correlations

Correlation #2



Source: Spurious-correlations

Correlation #3



Correlation ≠ Causation

Source: Spurious-correlations

Conclusions

- Machine readable EULA's/ToS
  - Technical (vs legal) controls for enforcing data usage/sharing
- Context based data usage & anonymization controls
  - Better Data Lineage controls
  - Blockchain technologies for contract enforcement?
- Privacy as a Service
  - Anonymization Services
  - Profile Masquerading
  - Privacy Brokers

the future is friendly®