Copernicus meets Coca-Cola:
What AI and big data mean for national statistics

Kenneth Cukier
Senior Editor, The Economist
Co-author “Big Data: A revolution that transforms how we live, work and think”
Statistiek Vlaanderen • Brussels • 17 December 2019

a story . . .

General Electric MRI scanner

morale: we embrace the imperfect because we’re comfortable with it, invested in it, optimized around it.
12/18/19

**Statistik**
“science of the state” (1749)
Gottfried Achenwall

**Physique sociale**
“la statistique morale (1835, 1848)
Adolphe Quetelet

“small data”

1930s
gdp (quarterly)
unemployment (estimated)

1950s
cpi (sample)
consumer confidence (survey)

what it measures, how it measures: different era
us gdp 4q 2008

first report (jan): -3.8%
revision (feb): -6.2%
final revision (july 2011): -8.9%

largest downward revision of GDP on record, more than twice as bad as first reported

“My contacts report that cutbacks in spending are widespread, especially for discretionary items. East Bay plastic surgeons and dentists note that patients are deferring elective procedures. [Laughter] Reservations are no longer necessary at many high-end restaurants…”

Janet Yellen
President, San Francisco Fed Reserve
FOMC meeting, 16 September 2008

“The city of Dublin appears to have more chimneys than Bristol, and consequently more people.”

– William Petty
Political Arithmetick circa 1676

“anecdotal”

proxy

methods
small data methods

- consumer sentiment
  50 questions: 500 people
- consumer confidence
  5 questions: 5,000 people

machine learning
how machine learning everywhere, everyday
machine learning
computer translation
speech recognition
self-driving car
etc.

hospitals activities
&
patients people

thinking of the world
as a platform for
the collection and
analysis of big data
technology
what does it lower the cost of?

computer → arithmetic
technology
what does it lower the cost of?

computer → arithmetic
ai → prediction
turn everything into prediction problems and solve them with ai
“This resolution is sufficient to delineate individual fields that are roughly 1 ha in size or greater. New commercial systems are delivering even higher spatial resolution … at costs that are approaching 1 USD$ per km2 (or $0.01 per ha).”


big data stats
more granular
more accurate
more timely

CargoMetrics
real-time tracking of “n=all” ships correlated with gdp

40.6

Key findings:
- Weekly index is calculated from 2010 to 2014
- Reporting status for the current week: 94% of all reporters have reported
- Reporting status for the previous week: 98% of all reporters have reported
big data stats
more granular
more accurate
more timely

Mastercard “SpendingPulse”
ex APT (2015): $1 of every $5 US retail dollars from 50,000 locations
correlated with official US monthly sales; reported daily by area, age, income, weather, etc

big data stats
more granular
more accurate
more timely

PriceStats
apprx 15 million prices reported daily
over 1,000 retailers from 60 countries

big data stats
more granular
more accurate
more timely

LinkedIn
ADP
The Economist
Google
UBER
Intuit
big data stats
more granular
more accurate
more timely
new: previously unmeasured

LinkedIn skills study: 500 mill. people, appx 130 mill. job transitions between over 4 mill. companies

stats startups
quandl
PREMISE

stats science fiction

stats science fiction
“investor sentiment”
based on sleep patterns, biochemistry, driving, diet and email read-response time
problem

hodgepodge sources
longevity uncertain
provenance unclear
quality inconsistent
biases rampant
idiosyncrasies unknown

google flu trends

ground truth?
- searches (media attention)
- visits to clinics (recession)

does this mean that the current version of GFT is not useful? No,
greater value can be obtained by combining GFT with other
near-real time health data.

hybrid approach

D. Lazer, R. Kennedy, C. King, A. Vespignani.
supplement not supplant

implications

change

people & organizations

change

concept of stats & society
thank you

KennethCukier@Economist.com

@kncukier