



Old and New Data

with Old and New Positives and Negatives

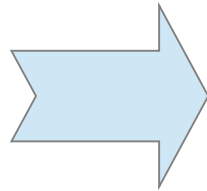
John Aston
University of Cambridge
& The Alan Turing Institute



Old and New Data

Parish of Collieston, Collieston

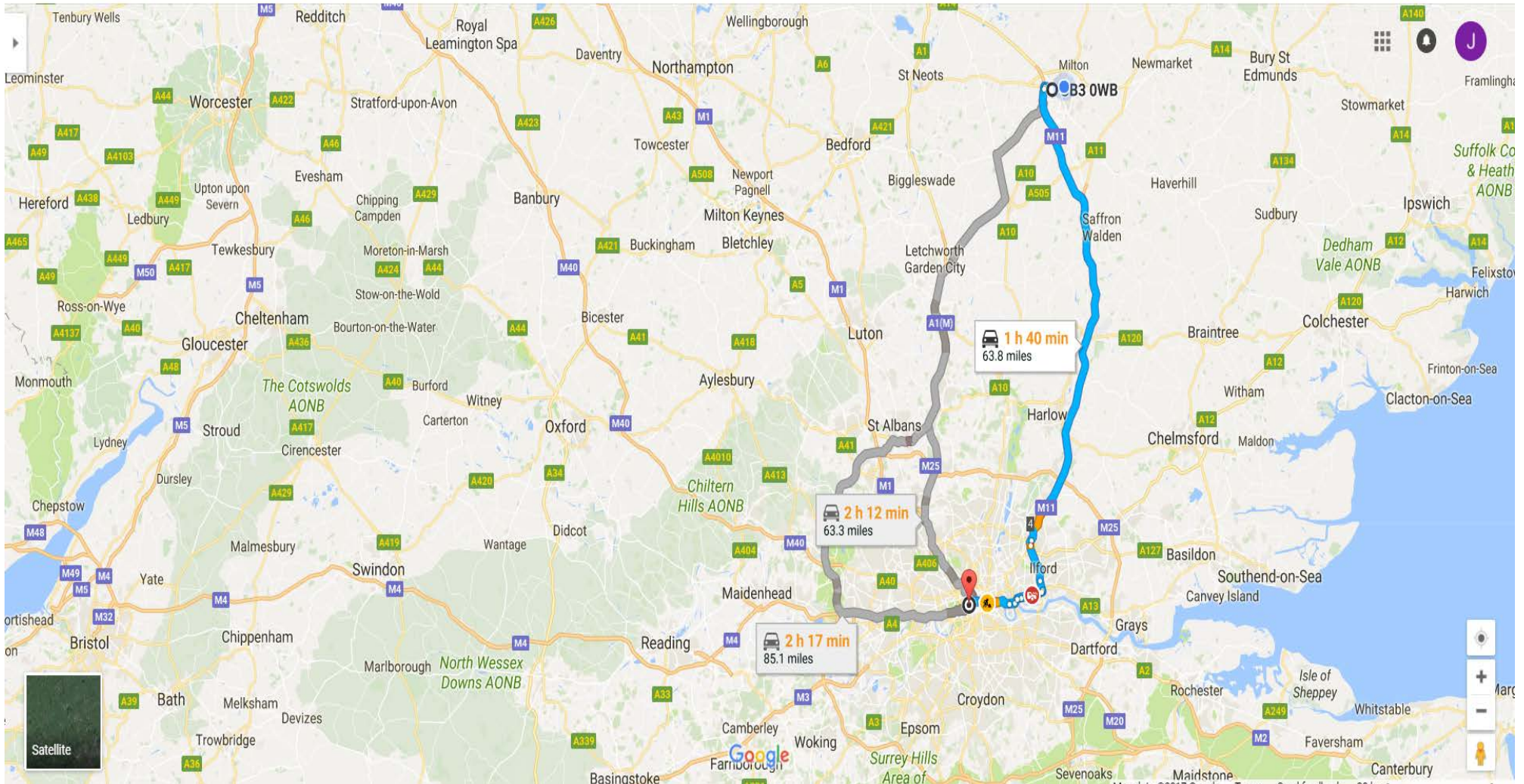
PLACE	HOUSE	NAME and SURNAME, SEX and AGE, of each Person who shall be such Person as to Part of 10 1841		OCCUPATION	WIFE'S BIRTH
		NAME and SURNAME	AGE		
In the town	1	John Kirkwood	61	Wagoner	20
		John do	20	do	20
		James do	11	do	20
		John do	9	Wagoner	20
		John do	62	Wagoner	20
		John do	71	Wagoner	20
		John do	65	Wagoner	20
		John do	22	Wagoner	20
		John do	21	Wagoner	20
		John do	20	Wagoner	20
		John do	11	Wagoner	20
		John do	66	Wagoner	20
		John do	70	Wagoner	20
		John do	15	Wagoner	20
		John do	61	Wagoner	20
		John do	12	Wagoner	20
		John do	20	Wagoner	20
		John do	4	Wagoner	20
	John do	1	Wagoner	20	
TOTAL		18	115		



Cambridge University's Data Centre

1841 Scottish Census Record
(Courtesy Scottish National
Records Office)

New Data – New Promise



UNIVERSITY OF
CAMBRIDGE

New data – old issues?

Adapted from Alan Turing Institute research by Wachter et al, 2017:

- It is taken for granted that making a system more interpretable will also make it less efficient. Is this true?
- Explanations may not always be technically feasible or practical. Alternative accountability mechanisms need to be explored for these cases, including certification schemes and auditing functions.
- Similar systems should be similarly regulated, AI, Robotics, Machine Learning, Autonomous Systems...



UNIVERSITY OF
CAMBRIDGE

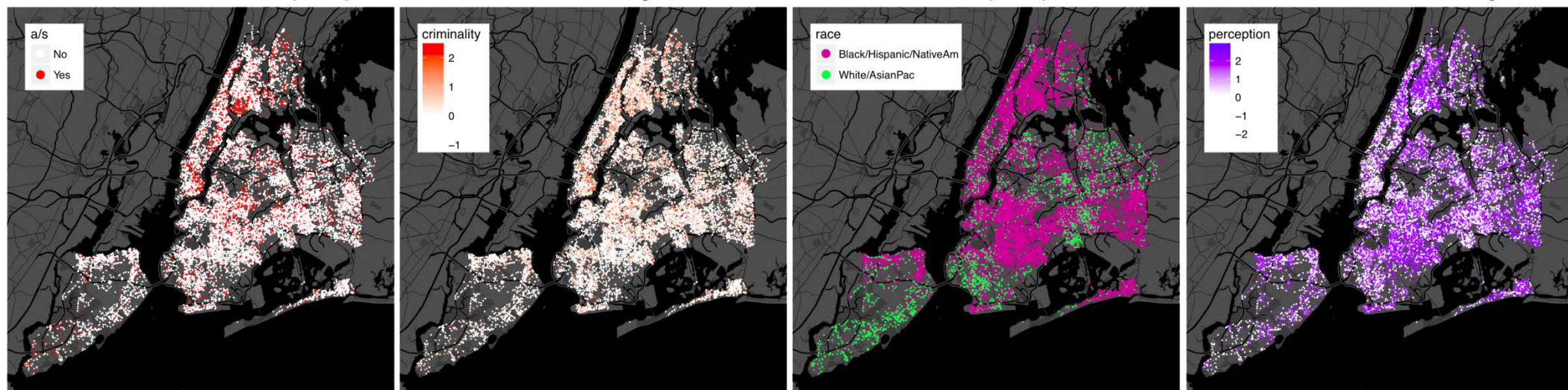
New data – new solutions?

Arrested / Summons (data)

Criminality

Race (data)

Perceived Criminality



Understanding criminality. The above maps show the decomposition of stop and search data in New York into factors based on perceived criminality (a race dependent variable) and latent criminality (a race neutral measure).

Research by Alan Turing Institute
(Kusner et al, 2017)



UNIVERSITY OF
CAMBRIDGE

New and Old Data

Bean Review into Economic Statistics addressed the burgeoning need for data integration

Independent Review of UK Economic Statistics

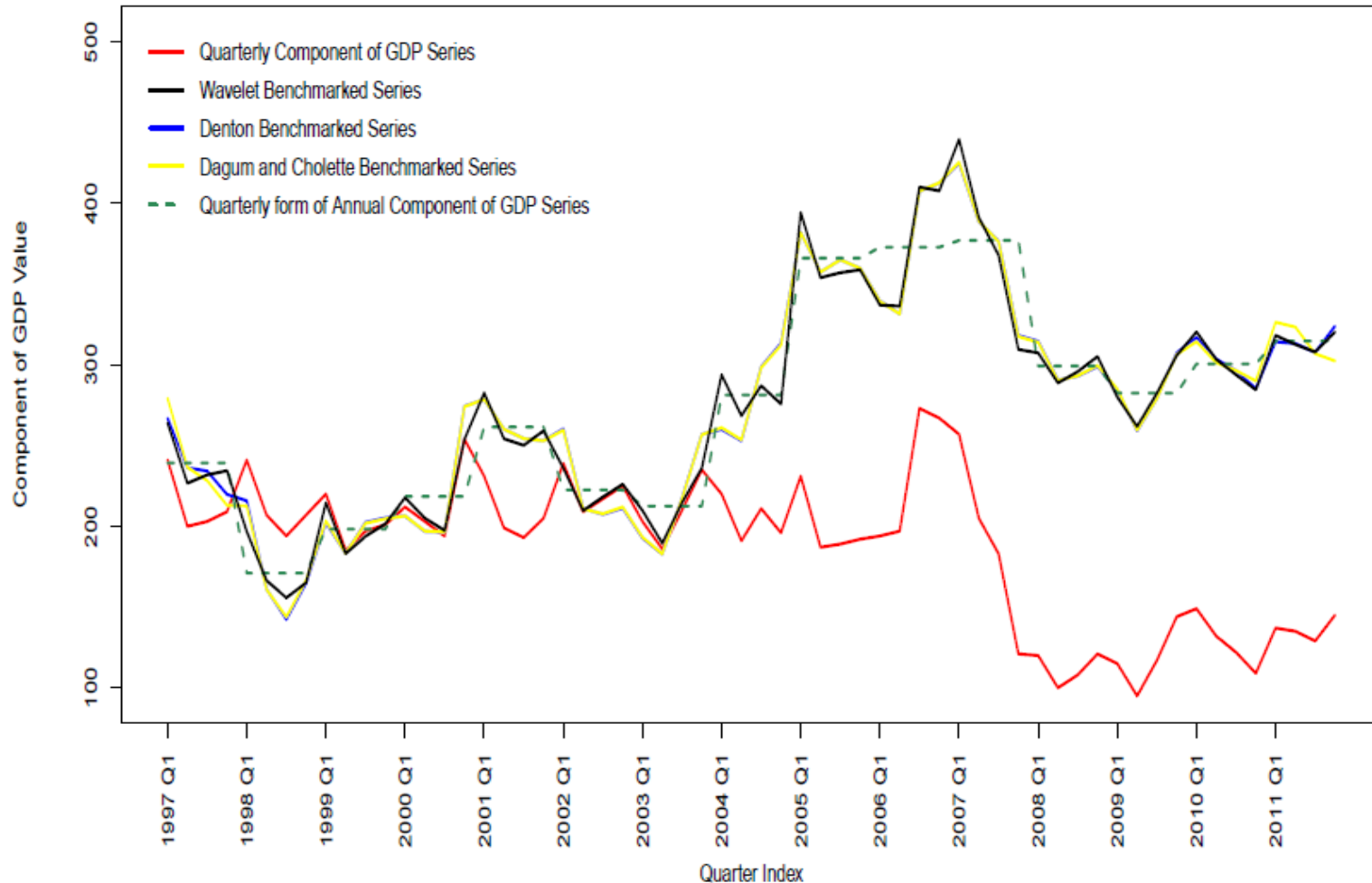
Professor Sir Charles Bean



UNIVERSITY OF
CAMBRIDGE

New and Old Data – Old Issues

Different Benchmarking Methods Applied to Component of UK GDP Data

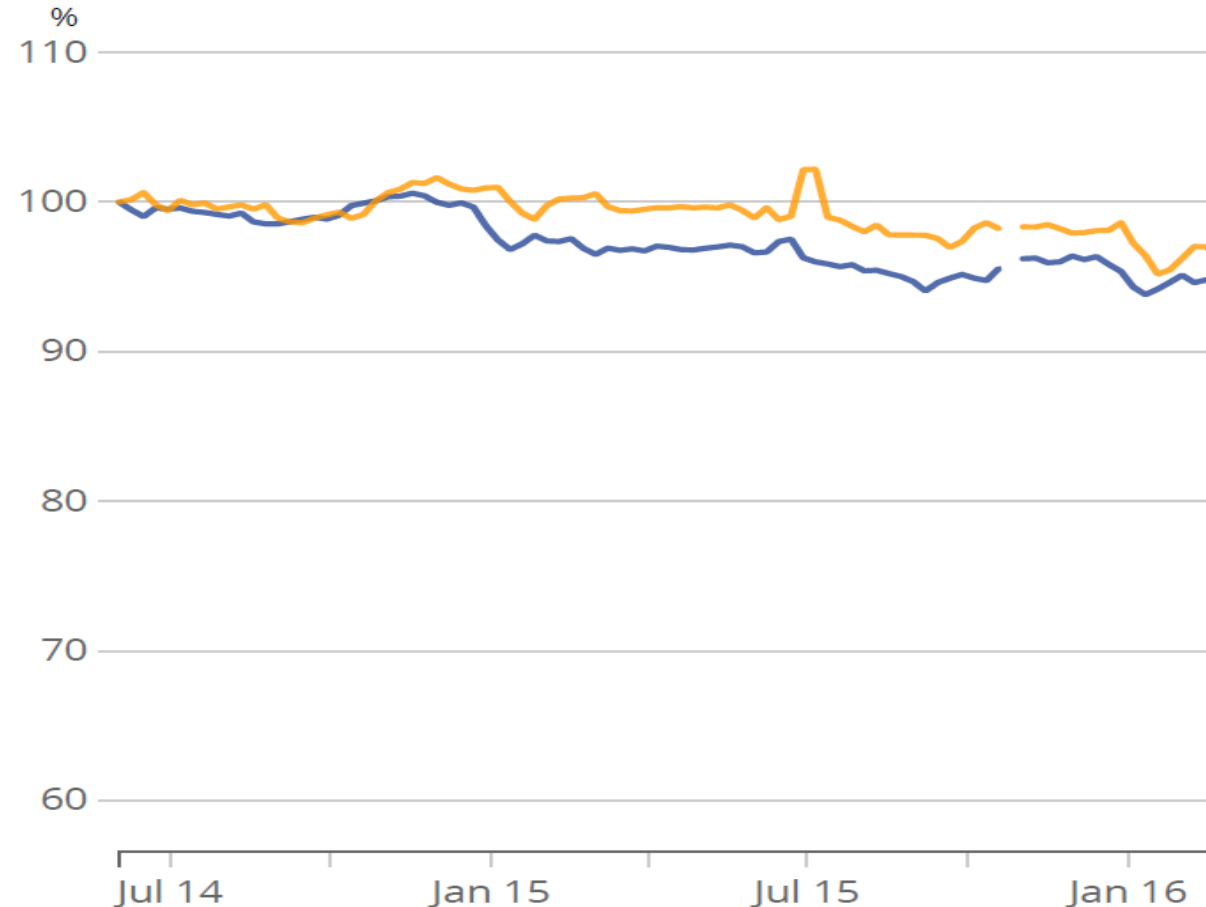


From Sayal et al, 2017

New and Old Data – New Promise

— GEKSJ Index — Unit Price Index

Food & Non-Alcoholic Beverages



From ONS, Research indices using web scraped price data: May 2016 update



UNIVERSITY OF
CAMBRIDGE

A Few Refs

- Wachter et al (2017) Transparent, explainable, and accountable AI for robotics, *Science Robotics*, 2:eaan6080
- Kusner et al (2017) Counterfactual Fairness, *arXiv*, 1703.06856
- Bean (2017) Independent Review of UK Economic Statistics
- Sayal et al (2017), An introduction to Wavelet Benchmarking with Seasonal Adjustment, *JRSSA*, 180: 863–889

