



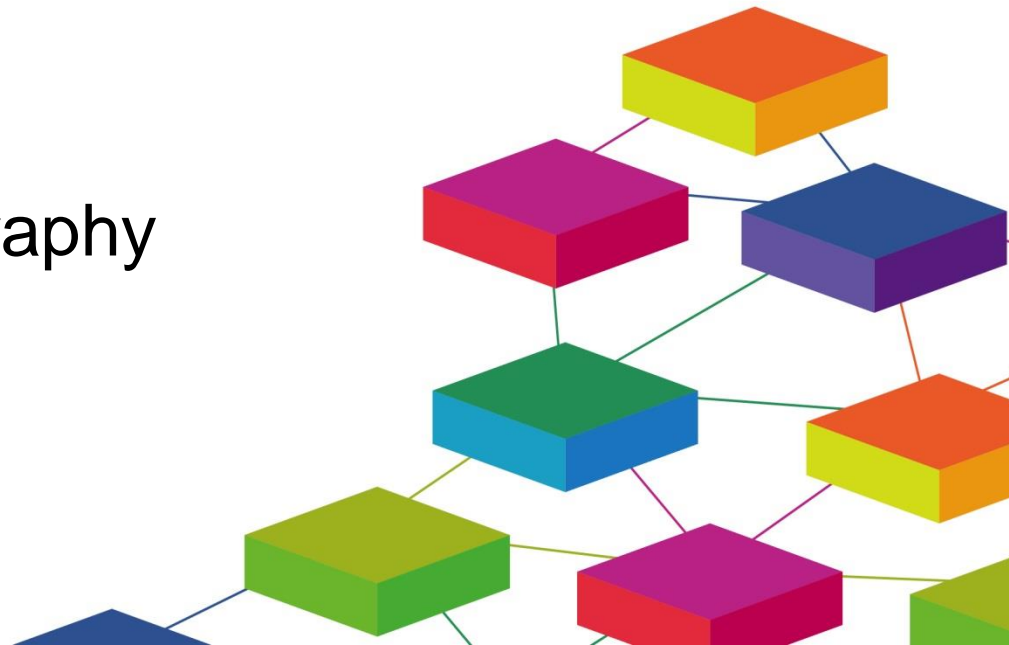
Consumer  
Data  
Research  
Centre

An ESRC Data  
Investment

# Big Data skills in the Social Sciences

## Transactional and consumer data

Dr Andy Newing  
Lecturer in Retail Geography  
School of Geography  
University of Leeds



- Lecturer in Retail Geography
- Consumer Data Research Centre (CDRC)

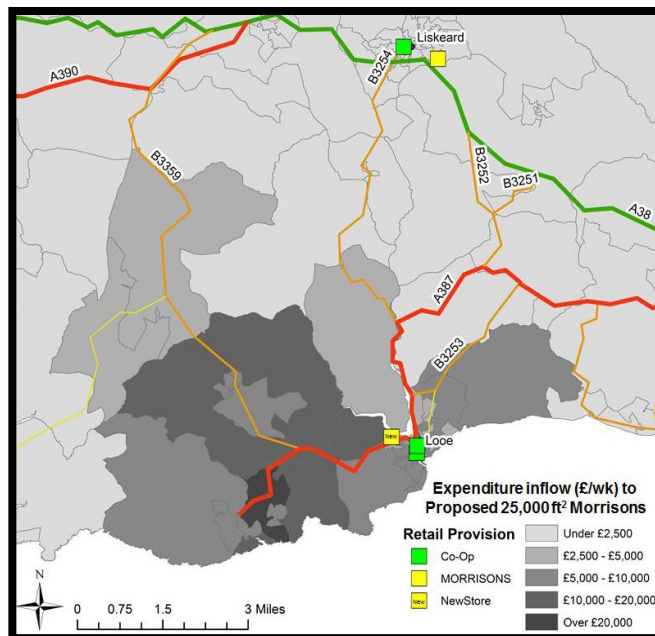
Sainsbury's  
*live well for less*





## Demand side

## Supply Side



- Electronic Point of Sale (EPOS) transactions/sales
- Loyalty cards
- GPS location data
- Mobile phone data – including apps (e.g. exercise)
- Transport (e.g. Oyster Cards)
- Footfall
- Mobilities
- Social media
- Utilities





## Using Big Data To Map A City's 'Heartbeat'

View the Tube Heartbeat, which shows the London Underground pulsing as passengers make their way around the city over the course of a typical weekday.

[Find out More](#)

---

## Welcome to the Consumer Data Research Centre (CDRC)

---

Vast amounts of UK consumer data are generated each day, providing valuable insight to help organisations operate more efficiently.

It's not just businesses that benefit. Researchers can utilise data to make the UK a better place.

Our aim is to work with organisations to open up their data to our trusted researchers so we can provide solutions that drive economic growth and improve our society.

[About our Data](#)

[About Us](#)



- Increasing recognition among businesses of the value of these data for wider reuse – e.g. ‘social good’.
- Awareness that their data covers only part of the market or a subset of consumers – skills required to link these data to other sources or to model missing data.
- Big data collected by these organisations is often underutilised.
- Innovation and collaboration.







Table A51

**Average weekly household expenditure by Output Area Classification<sup>1</sup> (OAC) supergroup**

UK, 2014

	Rural residents	Commuters	Ethnicity central	Metropolitan	Urbanites	Suburbanites	Constrained city dwellers	Hand pressed living	All households
	OAC Super-group 1	OAC Super-group 2	OAC Super-group 3	OAC Super-group 4	OAC Super-group 5	OAC Super-group 6	OAC Super-group 7	OAC Super-group 8	
Commodity or service	Average weekly household expenditure (£)								
1 Food and non-alcoholic drinks	66.70	59.70	51.60	59.60	60.40	64.30	42.80	54.40	58.80
2 Alcoholic drinks, tobacco and narcotics	14.20	11.70	8.30	9.40	11.60	13.30	13.30	13.00	12.30
3 Clothing and footwear	27.30	30.90	16.20	20.50	27.60	25.70	13.00	22.40	23.70
4 Housing (net) <sup>2</sup> , fuel and power	68.00	153.40	115.90	84.70	72.80	56.40	62.20	61.30	72.70
5 Household goods and services	42.90	30.80	22.50	33.30	39.90	43.20	24.40	28.50	35.40
6 Health	9.30	4.70	4.00	8.80	8.00	7.70	3.60	5.90	7.30
7 Transport	105.80	60.40	58.00	62.30	85.20	91.40	39.00	58.50	74.80
8 Communication	16.10	17.50	17.00	15.20	16.10	16.20	12.20	14.30	15.50
9 Recreation and culture	91.10	66.60	36.10	61.70	70.20	96.30	35.30	51.00	68.80
10 Education	16.10	[36.70]	[23.50]	10.90	8.20	10.00	[0.20]	1.00	9.80
11 Restaurants and hotels	48.60	63.90	45.70	39.70	44.90	48.70	25.40	32.90	42.50
12 Miscellaneous goods and services	49.20	45.20	29.00	39.60	46.40	45.80	20.80	31.70	40.00
1-12 All expenditure groups	555.20	581.60	427.40	445.80	491.50	518.80	292.30	374.90	461.20
13 Other expenditure items	77.60	63.30	52.70	66.70	98.10	98.50	32.20	45.40	70.10
<b>Total expenditure</b>	<b>632.80</b>	<b>644.70</b>	<b>480.10</b>	<b>512.60</b>	<b>589.50</b>	<b>605.40</b>	<b>324.50</b>	<b>420.30</b>	<b>531.30</b>
<b>Average weekly expenditure per person (£)</b>									
<b>Total expenditure</b>	<b>268.40</b>	<b>294.30</b>	<b>185.90</b>	<b>188.40</b>	<b>250.50</b>	<b>252.40</b>	<b>164.70</b>	<b>172.30</b>	<b>221.80</b>

Please see background notes for symbols and conventions used in this report.

<sup>1</sup> This table uses 2011 Output Area Classifications, replacing the 2001 classifications used in previous publications. 2001 and 2011 classifications are not comparable.

<sup>2</sup> Excluding mortgage interest payments, council tax and Northern Ireland rates.

**Household Questionnaire**  
England

Office for  
National Statistics

**Complete online**

[www.census.gov.uk](http://www.census.gov.uk)

Your personal internet access code is:

OR fill in this paper questionnaire and post it back using the pre-paid envelope supplied.

If your address is incorrect or missing, enter your correct address here:

Postcode

**Return to:**  
**FREEPOST 2011 Census,**  
**Processing Centre, UK**

**A message to everyone - act now**

Everyone should be included in the census - all people, households and overnight visitors.

It is used to help plan and fund services for your community - services like transport, education and health.

Please complete your census questionnaire on 27 March 2011, or as soon as possible afterwards. You can fill it in online or on paper.

Taking part in the census is very important and it's also compulsory. You could face a fine if you don't participate or if you supply false information.

**Your personal information is protected by law.** Census information is kept confidential for 100 years.

So help tomorrow take shape and be part of the 2011 Census.

*J. N. Matheson*  
National Statistician

**Declaration**

This questionnaire has been completed to the best of my knowledge and belief.

Signature

Date

Telephone number

We may contact you if we need to collect missing information.

If you have lost your envelope, please return to: FREEPOST 2011 Census, Processing Centre, UK

**Where can you get help?**

[www.census.gov.uk](http://www.census.gov.uk)

Census helpline 0300 0201 101

Text Relay 18001 0300 0201 160

**H1**

Help is available in large print and Braille



Office for  
National Statistics

20 December 2012

Issued by  
Office for National Statistics,  
Government Buildings, Cardiff  
Road, Newport, NP10 8XG

Media Office 0845 6041858

# Information paper

## Quality and Methodology Information

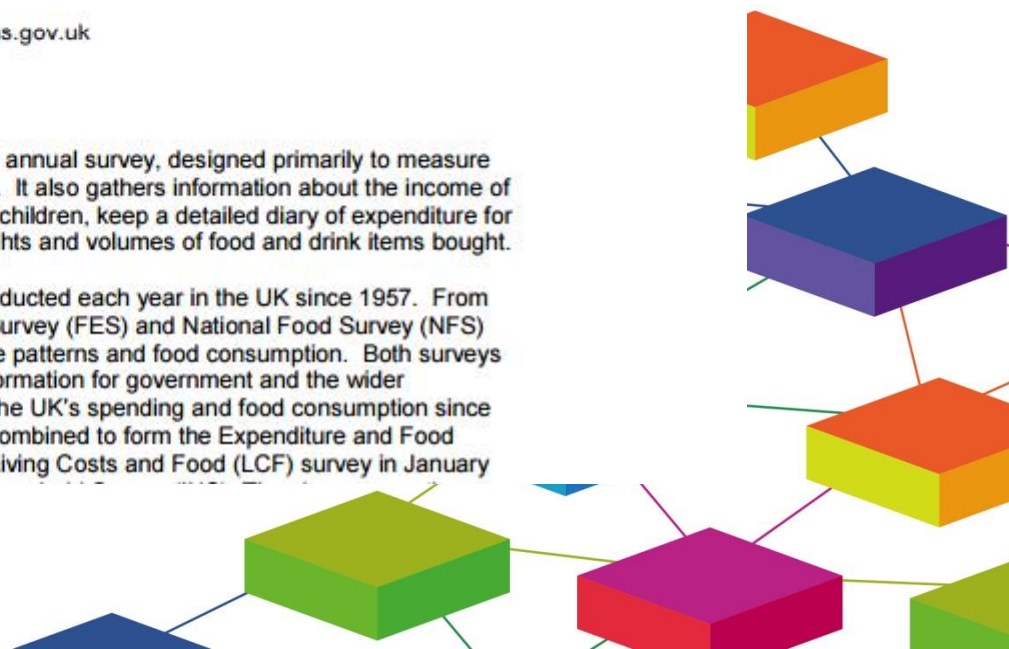
### General details

Title of output:	Living Costs and Food Survey
Abbreviated Title:	LCF
Designation:	National Statistics
Geographic Coverage:	UK
Date of last SQR or QMI*:	February 2008
Contact details:	<a href="mailto:SocialSurveys@ons.gov.uk">SocialSurveys@ons.gov.uk</a>

### Executive summary

The Living Costs and Food Survey (LCF) is an annual survey, designed primarily to measure household expenditure on goods and services. It also gathers information about the income of household members. Respondents, including children, keep a detailed diary of expenditure for two weeks. Respondents also record the weights and volumes of food and drink items bought.

A household expenditure survey has been conducted each year in the UK since 1957. From 1957 to March 2001, the Family Expenditure Survey (FES) and National Food Survey (NFS) provided information on household expenditure patterns and food consumption. Both surveys were well-established important sources of information for government and the wider community, charting changes and patterns in the UK's spending and food consumption since the 1950s. In April 2001 these surveys were combined to form the Expenditure and Food Survey (EFS). The EFS was renamed as the Living Costs and Food (LCF) survey in January







# Consumer Data Research Centre

An ESRC Data Investment

Have you been to any of the following in the last 12 months?  
Please select all that apply



City Centre Shopping Location

e.g. Oxford Circus



A Local High Street



A Leisure Park

Any standalone destination containing leisure activities such as a bowling alley or a cinema with some shops and restaurants



A Transport Hub

A train station/ airport with shops and restaurants

## Research Opinion Poll

Your views count. Manufacturers and service providers depend on your input to help them offer the right products in the right way at the right price. Please take a few minutes to complete this survey.

Just ignore any questions you would prefer not to answer. Thank you for helping us.

Your answers will be protected by the Data Protection Act and controlled by Acxiom Ltd; they will use this information for developing products & services, marketing, research, & updating and correcting databases. Acxiom Ltd may pass the information given to other reputable organisations, who will use it for their own market research and analysis purposes. They may send you, by mail or other media, details of their products and services. When answering questions for your partner, please check first that they are happy for you to do so. If you would rather that your details and/or your partners were not passed to these organisations then please tick the appropriate boxes you ( ) partner ( ). Certain questions have been included in the survey on behalf of: The Telegraph Group Plc: Newspapers Q1-5. Yorkshire Forward: (Regional Development Agency) various questions about Yorkshire, and You and Your Family Q5. RAC: Motoring Q4-5. Honda: Motoring Q6-7. Bupa: Family Health Q2.

Thank you very much for your help

PS: Don't forget the extra Digital Radio draw, to enter, simply return your survey within 7 days!

Beth Sanderson

### Groceries

#### 1 Where do you shop for groceries?

- |   |   |
|---|---|
| 01 <input type="checkbox"/> Aldi/Netto      | 10 <input type="checkbox"/> Morrisons                 |
| 02 <input type="checkbox"/> Asda            | 11 <input type="checkbox"/> Sainsbury's               |
| 03 <input type="checkbox"/> Budgens         | 12 <input type="checkbox"/> Sainsbury's Local/Central |
| 04 <input type="checkbox"/> Co-op           | 13 <input type="checkbox"/> Somerfield                |
| 05 <input type="checkbox"/> Iceland         | 14 <input type="checkbox"/> Tesco                     |
| 06 <input type="checkbox"/> Kwiksave        | 15 <input type="checkbox"/> Tesco Metro/Express       |
| 07 <input type="checkbox"/> Lidl            | 16 <input type="checkbox"/> Waitrose                  |
| 08 <input type="checkbox"/> Makro/Spar/VG   | 17 <input type="checkbox"/> Corner Shop/Other         |
| 09 <input type="checkbox"/> Marks & Spencer | 18 <input type="checkbox"/> Garage Forecourt          |

#### 2 From the list above please write the number of your MAIN supermarket:

Why do you shop where you do?

(Please tick a maximum of 2)

- |   |   |
|---|---|
| 1 <input type="checkbox"/> Distance/Convenience | 4 <input type="checkbox"/> Prices             |
| 2 <input type="checkbox"/> Quality of Products  | 5 <input type="checkbox"/> Store Loyalty Card |
| 3 <input type="checkbox"/> Parking Facilities   | 6 <input type="checkbox"/> Food Range         |

#### 4 What do you spend on groceries a week?

- |                                      |                                   |                                      |  |
|--------------------------------------|-----------------------------------|--------------------------------------|--|
| <b>Main Shopping</b>                 |                                   | <b>Top-Up Shopping</b>               |  |
| 1 <input type="checkbox"/> Up to £15 | 4 <input type="checkbox"/> £50-59 | 1 <input type="checkbox"/> Under £10 |  |
| 2 <input type="checkbox"/> £16-34    | 5 <input type="checkbox"/> £60-89 | 2 <input type="checkbox"/> £10-19    |  |
| 3 <input type="checkbox"/> £35-49    | 6 <input type="checkbox"/> £90+   | 3 <input type="checkbox"/> £20+      |  |

#### A Shopping Centre/Mall

An enclosed mall e.g. Westfield, The Bullring or Merry Hill

### Newspapers

#### 1 What newspapers do your family read?

- |   |   |
|---|---|
| <b>Daily Paper</b>                            | <b>Sunday Paper</b>                     |
| 01 <input type="checkbox"/> Sun               | 06 <input type="checkbox"/> Telegraph   |
| 02 <input type="checkbox"/> Mirror            | 09 <input type="checkbox"/> Times       |
| 03 <input type="checkbox"/> Mail              | 10 <input type="checkbox"/> Guardian    |
| 04 <input type="checkbox"/> Express           | 11 <input type="checkbox"/> Independent |
| 05 <input type="checkbox"/> Star              | 12 <input type="checkbox"/> FT          |
| 06 <input type="checkbox"/> Sport             | 13 <input type="checkbox"/> Other/Local |
| 07 <input type="checkbox"/> Daily Record      | 20 <input type="checkbox"/> Post        |
| 14 <input type="checkbox"/> News of the World | 15 <input type="checkbox"/> Mirror 21   |
| 16 <input type="checkbox"/> Mail              | 22 <input type="checkbox"/> Times       |
| 17 <input type="checkbox"/> Express 23        | 18 <input type="checkbox"/> Observer    |
| 19 <input type="checkbox"/> Sport             | 25 <input type="checkbox"/> Scot. Mail  |
| 26 <input type="checkbox"/> Other/Local       |   |

#### 2 Write in the number of your MAIN:

Daily Paper  Sunday Paper

#### 3 Is your Daily Paper delivered? 1 ☐ Yes 9 ☐ No

(If not everyday, please tick all that apply)

- |                                      |                                 |                                |                                      |
|--------------------------------------|---------------------------------|--------------------------------|--------------------------------------|
| 1 <input type="checkbox"/> Every day | 3 <input type="checkbox"/> Tues | 5 <input type="checkbox"/> Thu | 7 <input type="checkbox"/> Sat       |
| 2 <input type="checkbox"/> Mon       | 4 <input type="checkbox"/> Wed  | 6 <input type="checkbox"/> Fri | 9 <input type="checkbox"/> Don't Buy |

#### 5 How often do you buy your main

Sunday Paper? 1 ☐ 1-2 times a Month  
2 ☐ 3-4 times a Month 3 ☐ Less Often

### Home

#### 1 In which month do you renew the following insurance policies? e.g. May = 05

Home Contents  Buildings

- 2 Do you: Rent: 3 ☐ Private 4 ☐ Council  
1 ☐ Own your home 2 ☐ Live with parents 5 ☐ Housing Assoc.

#### 3 Is your home a:

- 1 ☐ Flat 3 ☐ Terraced 5 ☐ Detached  
2 ☐ Maisonette 4 ☐ Semi-Detached 6 ☐ Bungalow

#### 4 How many adults live in your home including yourself?

#### 5 How many bedrooms do you have?

#### 6 When did you move to this address?

#### 7 Please tell us the house number and postcode of your previous address:

#### 8 Are you planning to move in the next:

- 1 ☐ 0-3 mths 2 ☐ 4-6 mths 3 ☐ 7-12 mths 9 ☐ No



- “looks like the man drawer of stuff”  
(David Kemp, HP, Big Data Week Belfast 2015).
  - unstructured and undocumented
  - held by various teams
  - in various formats
  - within various systems
  - in numerous locations
  - some may be ‘digital exhaust’
  - considerable manual input to organise, extract and explain



- Data scarce to data rich – many social science methods traditionally designed for small datasets with well known properties.
- Spreadsheets to large databases
- How will major surveys (e.g. ONS LCF) keep pace?



- Social science has started developing capabilities to store and analyse transactional data.
- Much of these data are geographical in nature so spatial analytics increasingly important (Graham and Shelton, 2013).
- Segmentation – how do we reduce unwieldy data sets to manageable specific segments to analyse?
- Data mining
  - find unexpected relationships
  - Summarise data in meaningful ways
- Time series analysis and near real time insight
- Visualisation – communicating insight

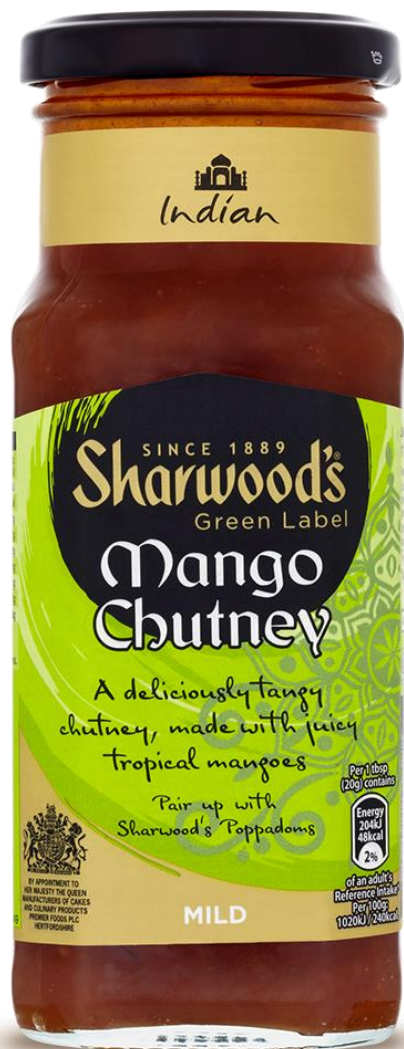






Consumer  
Data  
Research  
Centre

An ESRC Data  
Investment

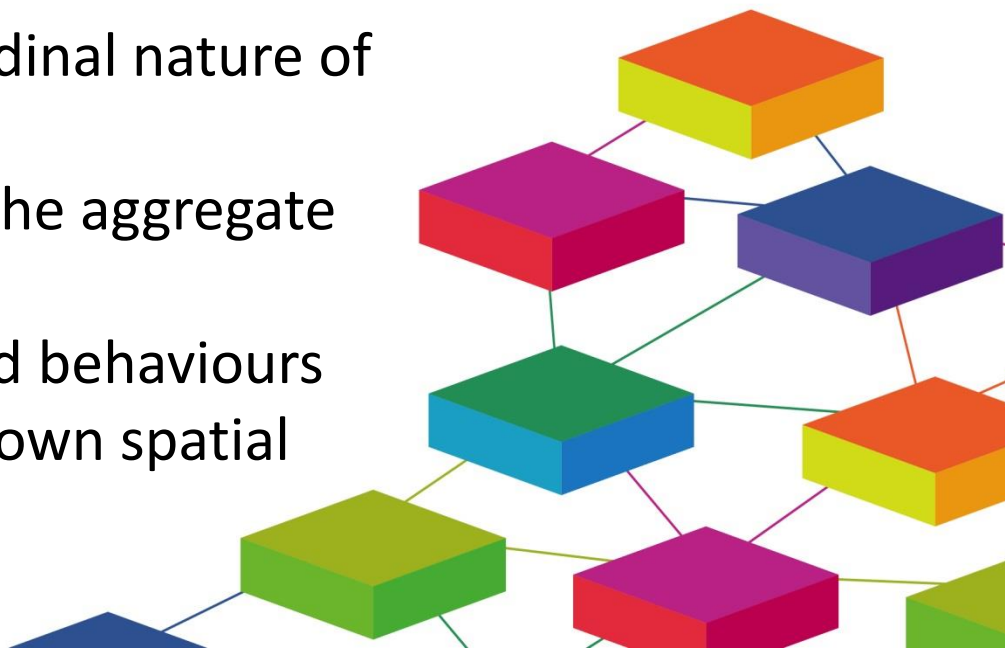




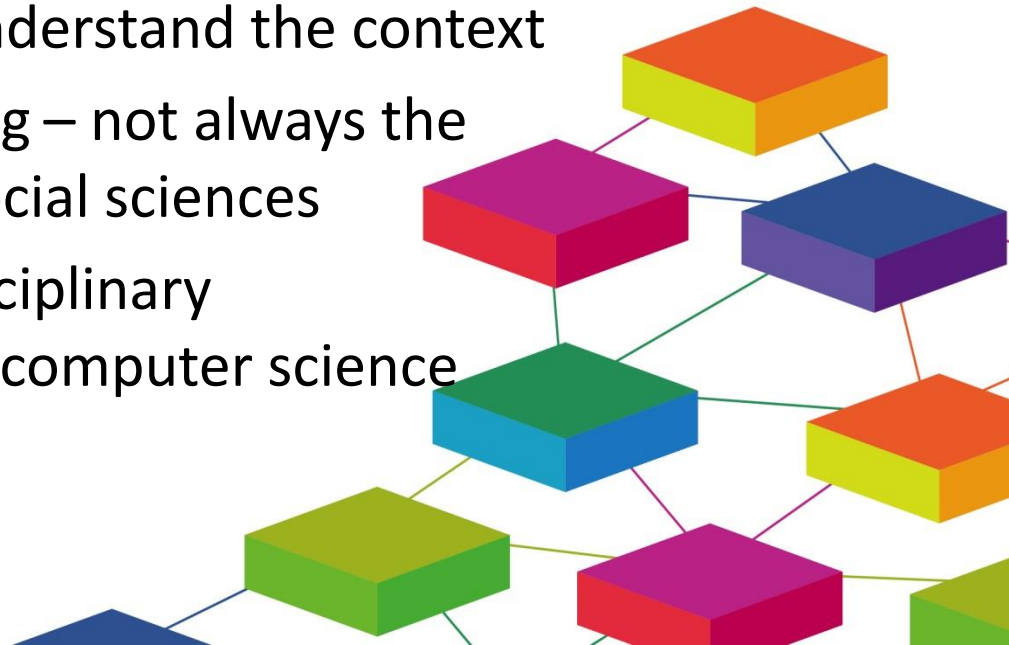
# “it is a customer problem, not a technological one”

Mick Yates, formerly Dunnhumby, International Big Data Symposium 2015

- Is our analysis data driven or problem driven?
- How do we exploit the longitudinal nature of these data?
- Do we want to understand at the aggregate or individual level?
- How do we link those observed behaviours to known individuals and/or known spatial origins?



- Collaborative research with commercial sector
  - Data access
  - Timeliness vs. quality
  - Publication potential
- Skills for dealing with the commercial sector – the academic way of doing things doesn't work!
- Commercial awareness – understand the context
- Huge overlap with marketing – not always the traditional the domain of social sciences
- Often requires new interdisciplinary collaborations – marketing, computer science

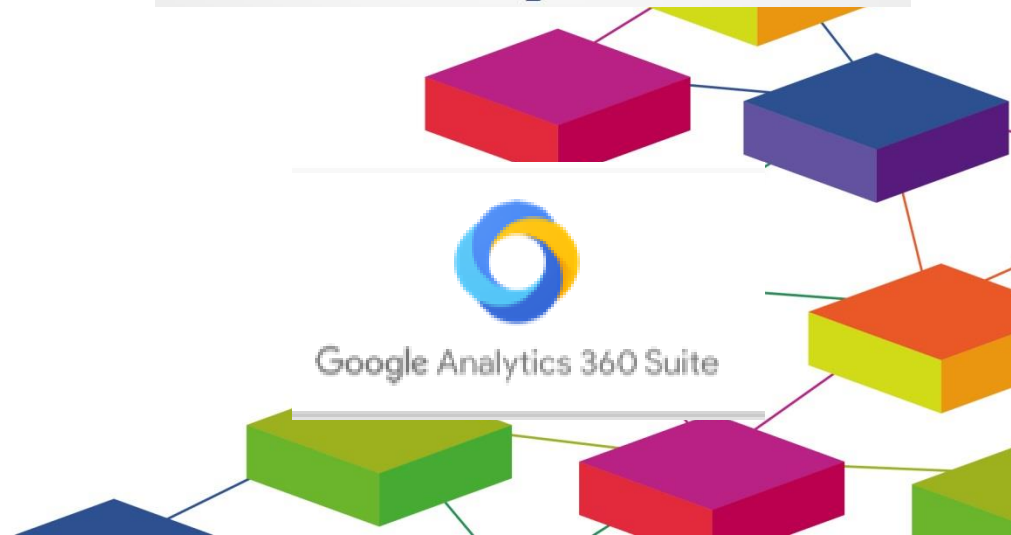
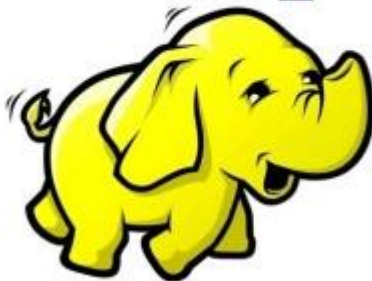




alteryx



**hadoop**



Alteryx Designer x64

File Edit View Tools Window Help

Search All Tools

Favorites In/Out Preparation Join Parse Transform Reporting Documentation Spatial Data Investigation Predictive Time Series Predictive Grouping Connectors

Browse Date Time Now Directory Input Map Input Output Text Input

Properties - Configuration - Output Data

Output Data Source

Options

Name	Value
1 Max Records Per File	
2 File Format	Alteryx database (*.yxdb)
3 No Spatial Index	<input type="checkbox"/>
4 Old (V.1) Spatial Index	<input type="checkbox"/>
5 Save Source & Description	<input checked="" type="checkbox"/>

☐ Take File/Table Name From Field

Append Suffix to File/Table Name

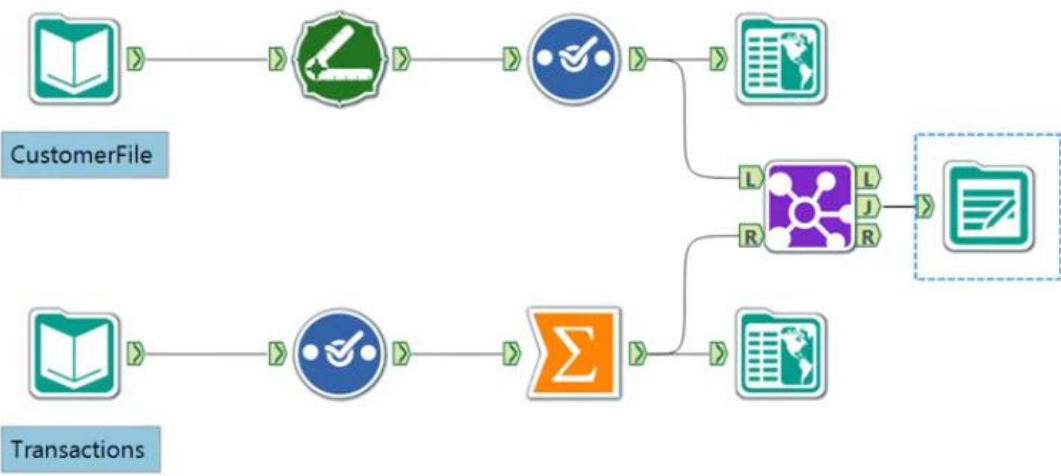
Field Containing File Name or Part of File Name

☒ Keep Field in Output

Campaign Response Analytics.yxmd\* x Resources

CustomerFile

Transactions

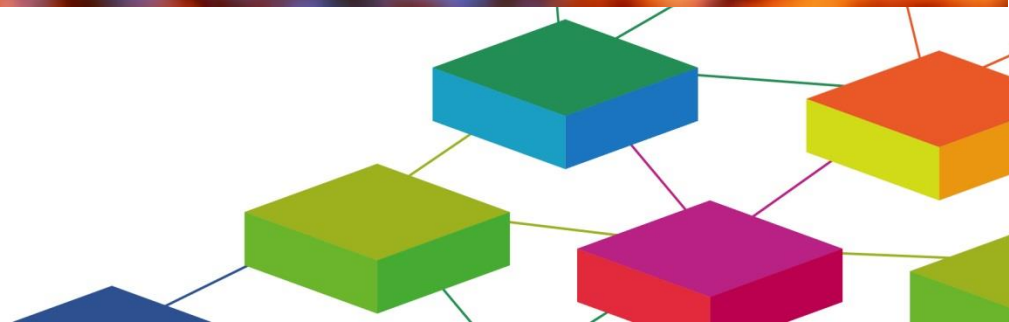




- CDRC Training and capacity building



- Or simply have a go ....







## Real-world data to put your theory into practice



### Breakfast at the Frat

Using sales and promotion info on pretzels, frozen pizza, boxed cereal, and mouthwash gathered from a sample of stores over 156 weeks, this dataset facilitates time series analyses in areas including promotional effectiveness and price sensitivity.

DOWNLOAD 



### Let's Get Sort-of-Real

The data's not real, but there sure is a lot of it. With 300M+ at-till transactions over 117 weeks, we've replicated the typical patterns found in real in-store sales data to help curious data scientists test their techniques and algorithms in a very real way.

*Multiple download options available. Click to view.*



### Carbo-Loading

Carbo-Loading contains household level transactions over a period of two years from four categories: Pasta, Pasta Sauce, Syrup, and Pancake Mix. These categories were chosen so that interactions between the categories can be detected and studied.

DOWNLOAD 

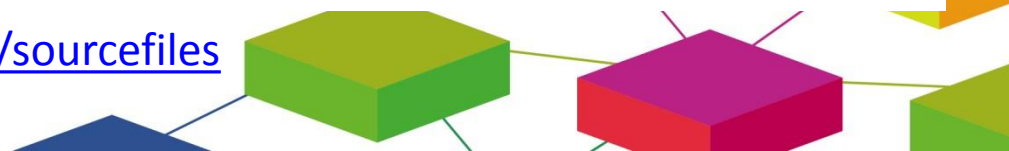


### The Complete Journey

This dataset contains household level transactions over two years from a group of 2,500 households who are frequent shoppers at a retailer. It contains all of each household's purchases, not just those from a limited number of categories.

DOWNLOAD 

<https://www.dunnhumby.com/sourcefiles>



## Using workplace population statistics to understand retail store performance

Tom Berry, Andy Newling<sup>a,b</sup>, Deborah Davies<sup>c</sup> and Kirsty Branch<sup>d</sup>

<sup>a</sup>Consumer Data Research Centre, University of Leeds, Leeds, UK; <sup>b</sup>School of Geography, University of Leeds, Leeds, UK; <sup>c</sup>The Co-operative Group, Manchester, UK

### ABSTRACT

We explore the value of recently released workplace geographies and accompanying census-based workplace zone statistics (WZS) and an associated classification of workplace zones (COWZ). We consider how these data could support retailers in their operational and strategic decision-making, including the evaluation of retail demand and retail store performance in localities where trade is driven by non-residential demand. In collaboration with major UK grocery retailer The Co-operative Group we explore the relationship between workplace population composition and store trading characteristics using a series of case study stores within Inner London. We use empirical store trading data to identify store and product category level temporal sales fluctuations attributable to workplace populations. We also use census-derived flow data to identify the spatial origins of workplace population inflow. We identify that store performance exhibits characteristics attributable to demand driven by these populations. We conclude that workplace population geographies, WZS and the COWZ afford considerable potential for understanding drivers of store performance, observed store trading patterns and evaluation of retail store performance. We suggest that the next step is to build these populations and their micro geography spatial and temporal characteristics into predictive models and evaluate their potential for store performance evaluation and location-based store and network decision-making within this sector.

### ARTICLE HISTORY

Received 17 February 2016  
Accepted 21 March 2016

### KEYWORDS

Workplace derived trade; workplace zone statistics; classification of workplace zones; Co-operative Food; Inner London

POPULATION, SPACE AND PLACE  
Popul. Space Place (2016)  
Published online in Wiley Online Library  
(wileyonlinelibrary.com) DOI: 10.1080/09595969.2016.1170066

## The Role of Digital Trace Data in Supporting the Collection of Population Statistics – the Case for Smart Metered Electricity Consumption Data

Andy Newling, Ben Anderson<sup>a</sup>, Abubakar Bahaj and Patrick James  
Sustainable Energy Research Group, Energy and Climate Change Division, Faculty of Engineering and Environment, University of Southampton, Southampton, UK

### ABSTRACT

Debates over the future of the UK's traditional decadal census have led to the exploration of supplementary data sources, which could support the provision of timely and enhanced statistics on population and housing in small areas. This paper reviews the potential value of a number of commercial datasets before focusing on high temporal resolution household electricity load data collected via smart metering. We suggest that such data could provide indicators of household characteristics that could then be aggregated at the census output area level to generate more frequent official small area statistics. These could directly supplement existing census indicators or even enable development of novel small area indicators. The paper explores this potential through preliminary analysis of a 'smart meter-like' dataset, and when set alongside the limited literature to date, the results suggest that aggregated household load profiles may reveal key household and household characteristics of interest to census users and national statistical organisations. The paper concludes that complete coverage, quasi-real time reporting, and household level detail of electricity consumption data in particular could support the delivery of population statistics and area-based social indicators, and we outline a research programme to address these opportunities. © 2015 The Authors. *Population, Space and Place* published by John Wiley & Sons Ltd.

\*Correspondence to: Ben Anderson, Sustainable Energy Research Group, Energy and Climate Change Division, Faculty of Engineering and Environment, University of Southampton, Southampton, UK.  
E-mail: B.Anderson@eecs.soton.ac.uk

Accepted 22 May 2015

**Keywords:** census; energy monitoring; small area statistics; digital trace data; big data

### INTRODUCTION

Provision of area-based population statistics in the UK is underpinned by the decadal census of housing and population as a crucial source of consistent baseline population estimates and robust local area statistics. Census data represent a backbone for commercial, academic and social research, widely used for in-depth analysis, policy making, and resource allocation (Eurostat, 2011), including allocation of billions of pounds of government and commercial investment at the local level. Key census outputs include estimates of usually resident population by age and sex reported using a hierarchy of output zones. However, the real value to policy and commercial analysis is derived from the publication of tables detailing the combinations of attributes of households and their usual residents at the small area level (ONS, 2014a). These provide information on indicators such as ethnic composition, education, socio-economic status, religion, and employment, and it is the combination of universal geographic coverage at the small area level coupled with this detailed attribute data that represents a major strength of the census as a tool for academic, policy, and commercial research as well as for public service resource allocation (Watson, 2009; ONS, 2013b). Given its importance, the cost of the census (£480m in 2011) represents exceptional value,

© 2015 The Authors. *Population, Space and Place* published by John Wiley & Sons Ltd.  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Abstracts

Alexander Bland (2015) Identifying fuel and poverty characteristics through e.on consumer records and geo-demographic segmentation data (Sponsor: e.on)  
[View abstract](#)

Thomas Berry (2015) Exploring the utility of the 2011 Work Place Statistics to help The Co-op better understand transient new store locations, worker flows and worker demographics (Sponsor: The co-operative food) – **ESRC Prize Winner**  
[View Abstract](#)

Brendon Edwards (2015) An investigation into the effects of a relaxation of the current Sunday Trading Legislation on the Co-operative's convenience stores (Sponsor: The co-operative food)  
[View Abstract](#)

Yiqiao Huang (2015) Shopping centre's turnover estimation using microsimulation: an exploratory research in Inverness (Sponsor: CACI)  
[View Abstract](#)

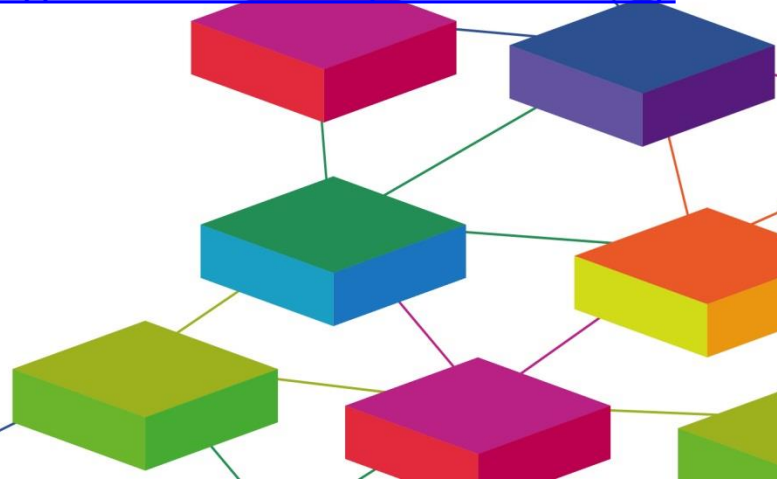
Radoslaw Kowalski (2015) Topic modelling online customer reviews (Sponsor: Argos)  
[View Abstract](#)

Karlo Lugomer (2015) Relations between structure and performance of retail centres in England and Wales and demographics of their catchment areas (Sponsor: LDC) – **ESRC Prize Winner**  
[View Abstract](#)

Anastasia Ushakova (2015) Can we identify vulnerable energy customers from smart meter data? (Sponsor: British Gas) – **ESRC Prize Winner**  
[View Abstract](#)

Clemens Zauchner (2015) Identifying the main drivers of customer satisfaction and dissatisfaction by mining customer verbatim feedback (Sponsor: easyjet) – **ESRC Prize Winner**  
[View Abstract](#)

<https://www.cdrc.ac.uk/retail-masters/>



<http://geog.leeds.ac.uk/people/a.newling>



Consumer  
Data  
Research  
Centre

An ESRC Data  
Investment

# Big Data skills in the Social Sciences

## Transactional and consumer data

Dr Andy Newing  
Lecturer in Retail Geography  
School of Geography  
University of Leeds

