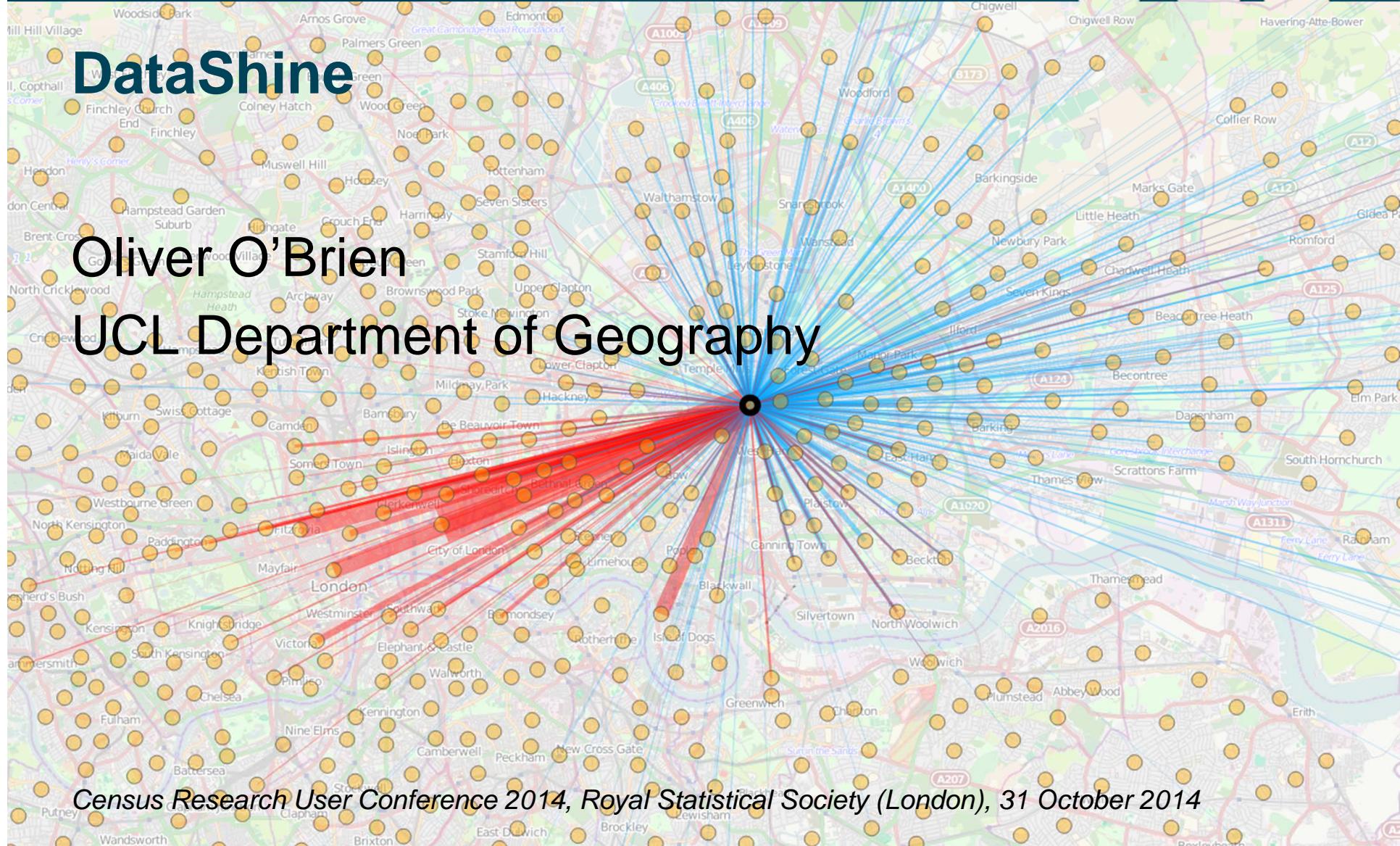


DataShine

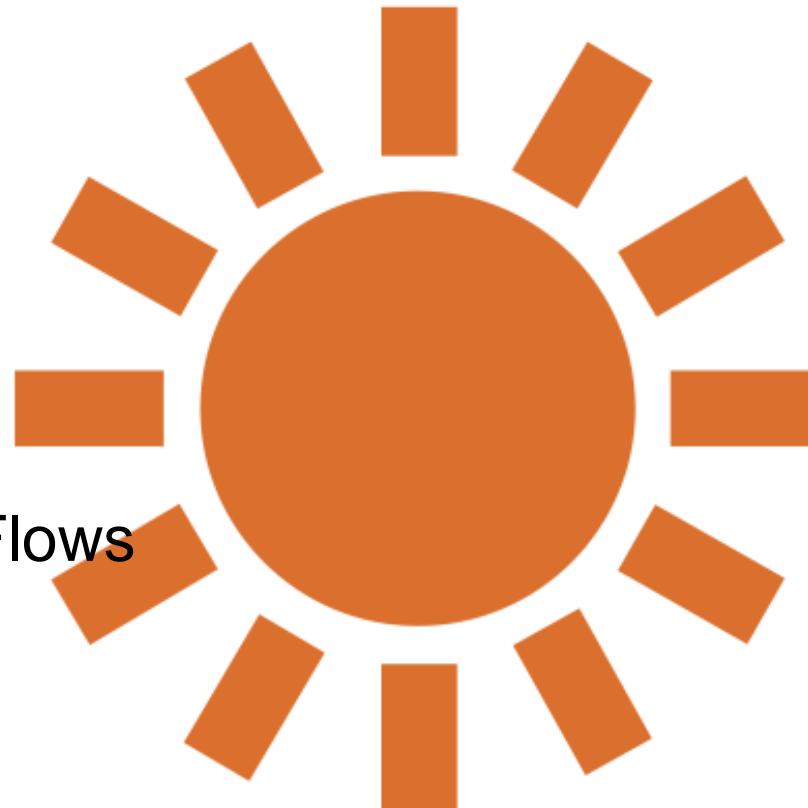
Oliver O'Brien

UCL Department of Geography



Contents

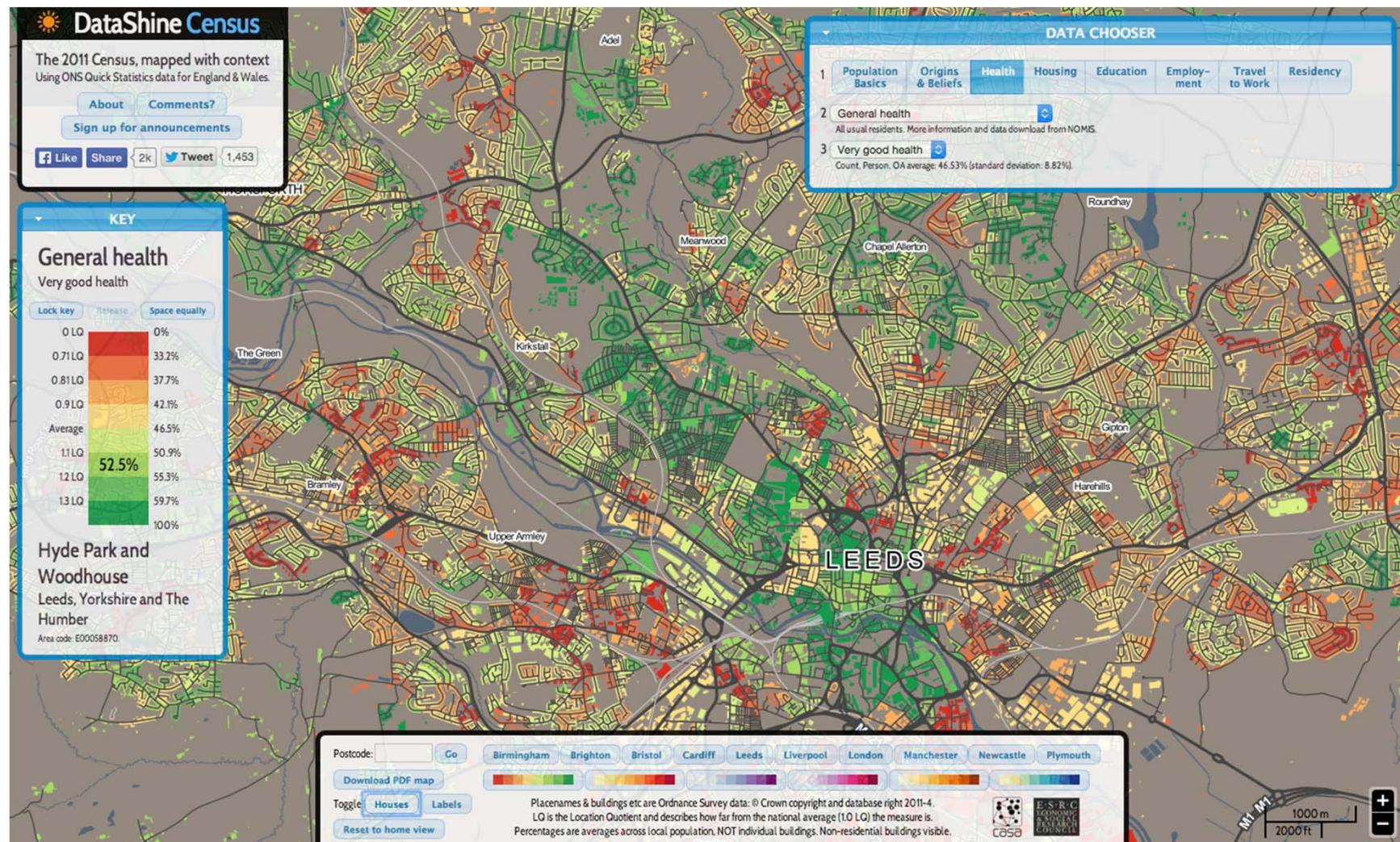
- The BODMAS Project
- Sourcing the Data
- DataShine: Census
- DataShine: Travel to Work Flows
- DataShine: OAC and LOAC
- Launching Today
- Next Steps
- Contact



The BODMAS Project

- + - x /
- Big, Open Data Mining, Analysis & Synthesis
 - Using large social science datasets to discover and visualise information
 - Creating tools (initially internally)
- Funded by the ESRC Future Research Leaders program
- 2014-2016+
- Starting with the Census 2011 datasets

DataShine: Census (<http://datashine.org.uk/>)

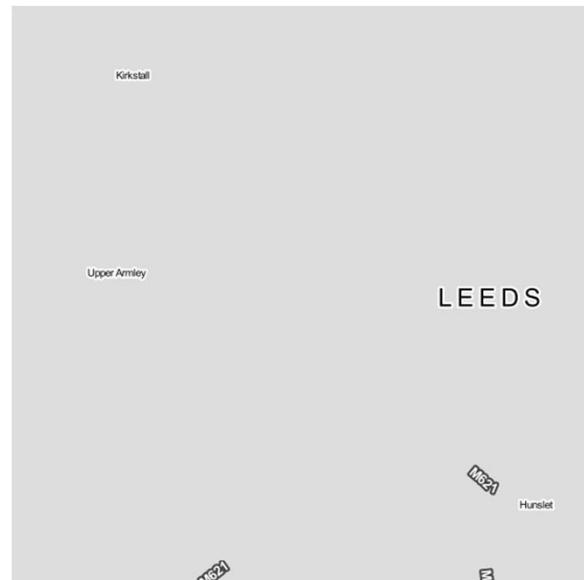
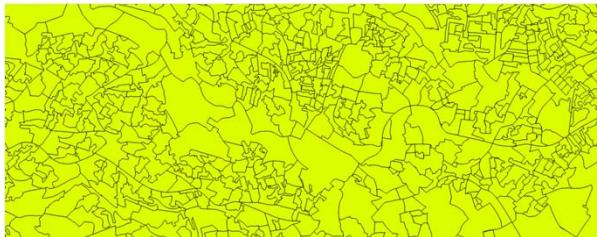


DataShine: Census – Data

- 2011 Census “Quick Statistics”
 - OAs and Wards
 - England/Wales only, for the moment
- Ordnance Survey Open Data
 - Large Scale: VectorMap District (building blocks)
 - Small Scale: Meridian 2 (urban extent)
 - Labels, water and other contextual features

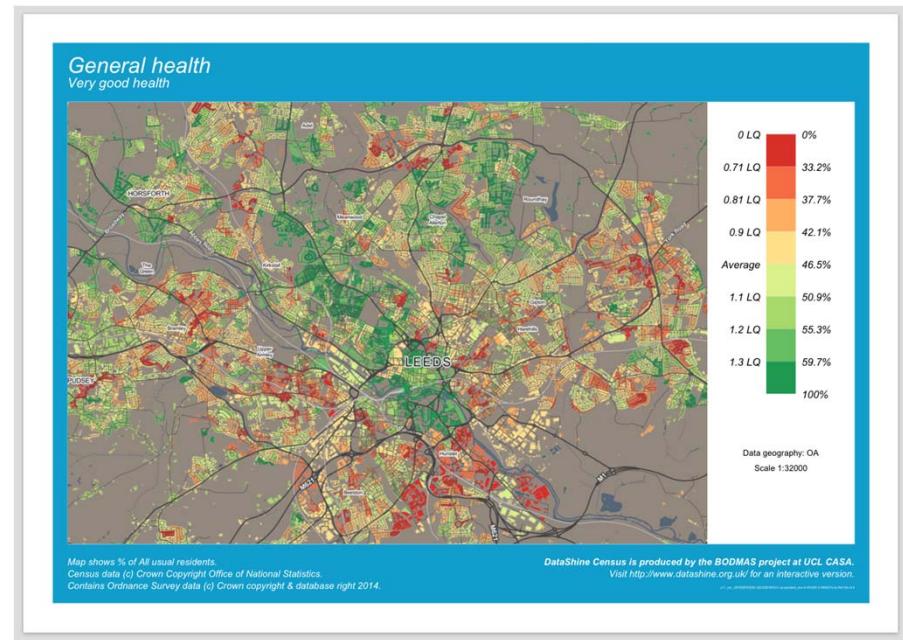
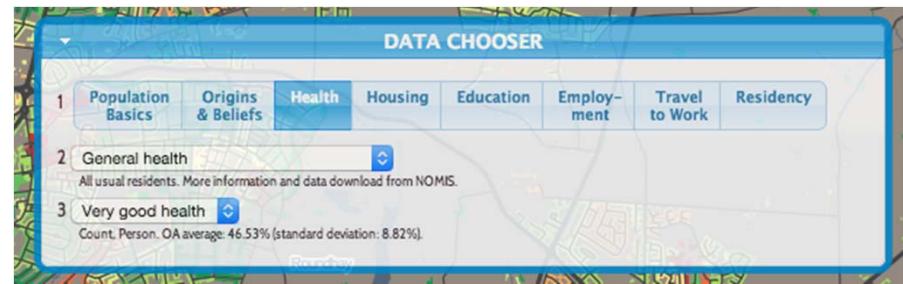
DataShine: Census – Data

A	B	C	D	E	F	G
OA Code	Ward Name	LA Name	Region Name	QS302EW0002	QS302EW0001 (base)	OA Percentage
E00057146	Gipton and Harehills	Leeds	Yorkshire and The Humber	50	208	24.04
E00058705	Roundhay	Leeds	Yorkshire and The Humber	84	163	51.53
E00169796	City and Hunslet	Leeds	Yorkshire and The Humber	112	192	58.33
E00059067	Cross Gates and Whinmoor	Leeds	Yorkshire and The Humber	101	232	43.53
E00058048	Morley North	Leeds	Yorkshire and The Humber	159	333	47.75
E00057088	Bramley and Stanningley	Leeds	Yorkshire and The Humber	146	312	46.79
E00059113	Farnley and Wortley	Leeds	Yorkshire and The Humber	121	276	43.84
E00056911	Armley	Leeds	Yorkshire and The Humber	84	220	38.18
E00057564	Gipton and Harehills	Leeds	Yorkshire and The Humber	133	321	41.43
E00058508	Pudsey	Leeds	Yorkshire and The Humber	126	284	44.37
E00058728	Roundhay	Leeds	Yorkshire and The Humber	85	179	47.49
E001170604	City and Hunslet	Leeds	Yorkshire and The Humber	147	337	58.46



DataShine: Census – Toolstack

- Mapnik
 - to create & composite the maps
- JavaScript Libraries
 - OpenLayers 2
 - Jquery(UI)
- PyCairo
 - PDF Export
- PostgreSQL/PostGIS

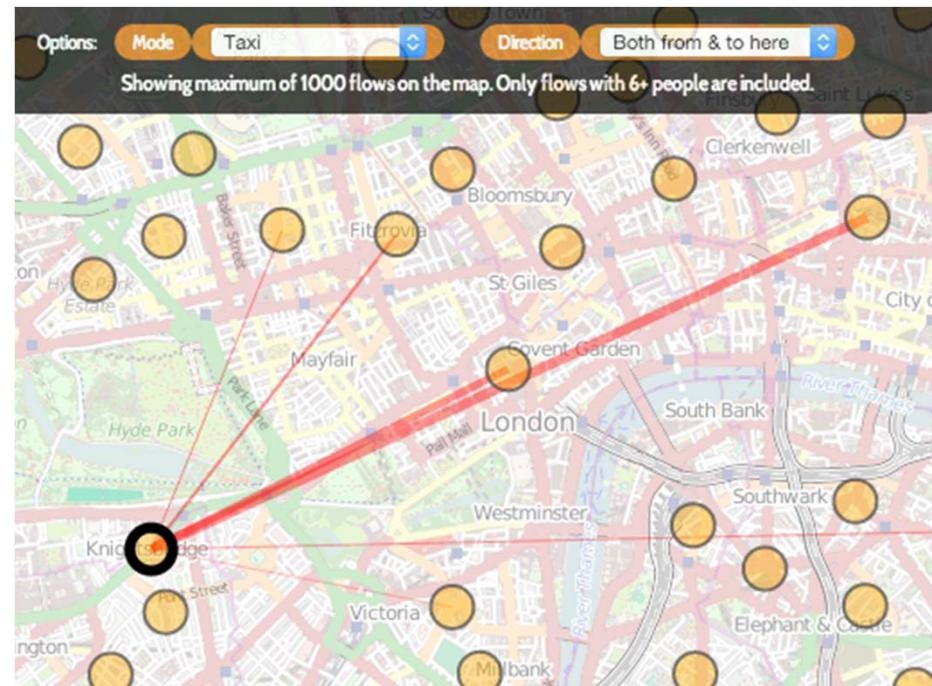
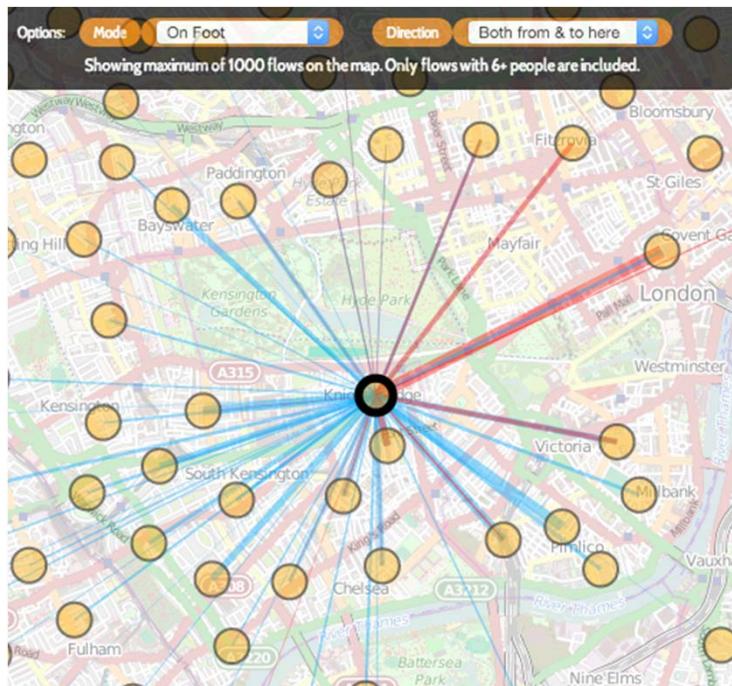


DataShine: Travel to Work Flows

- Blue = commute to here
- Red = commute from here



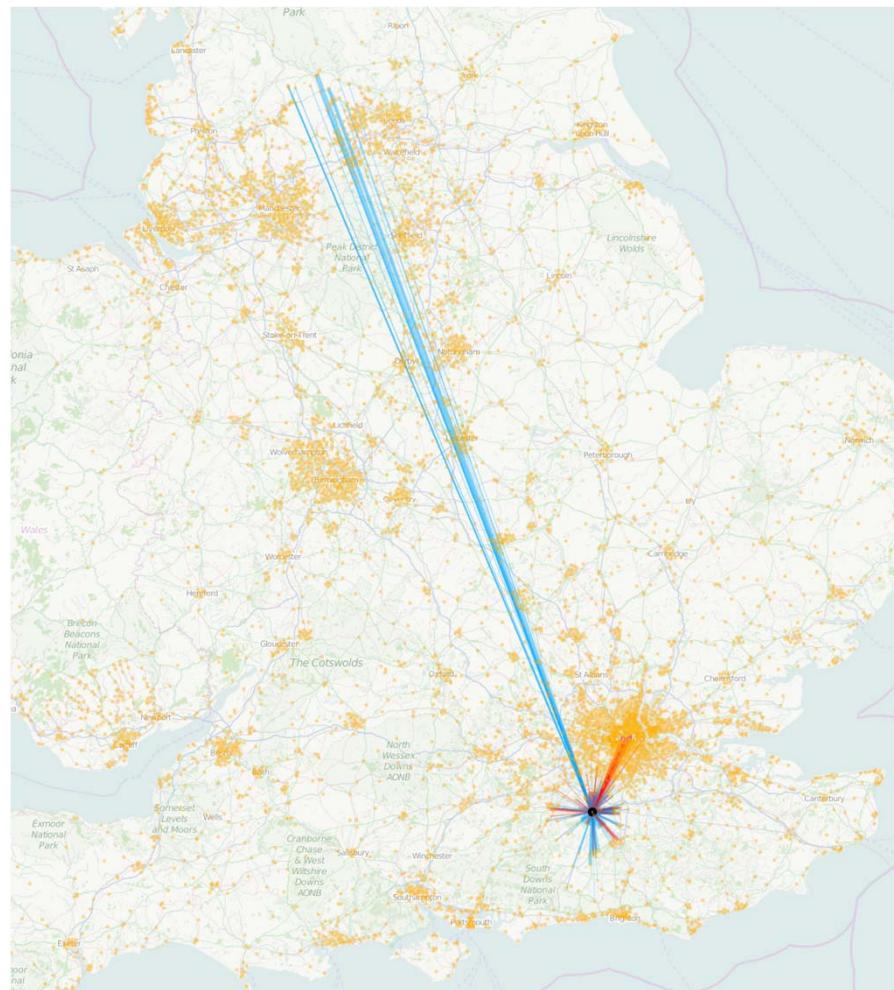
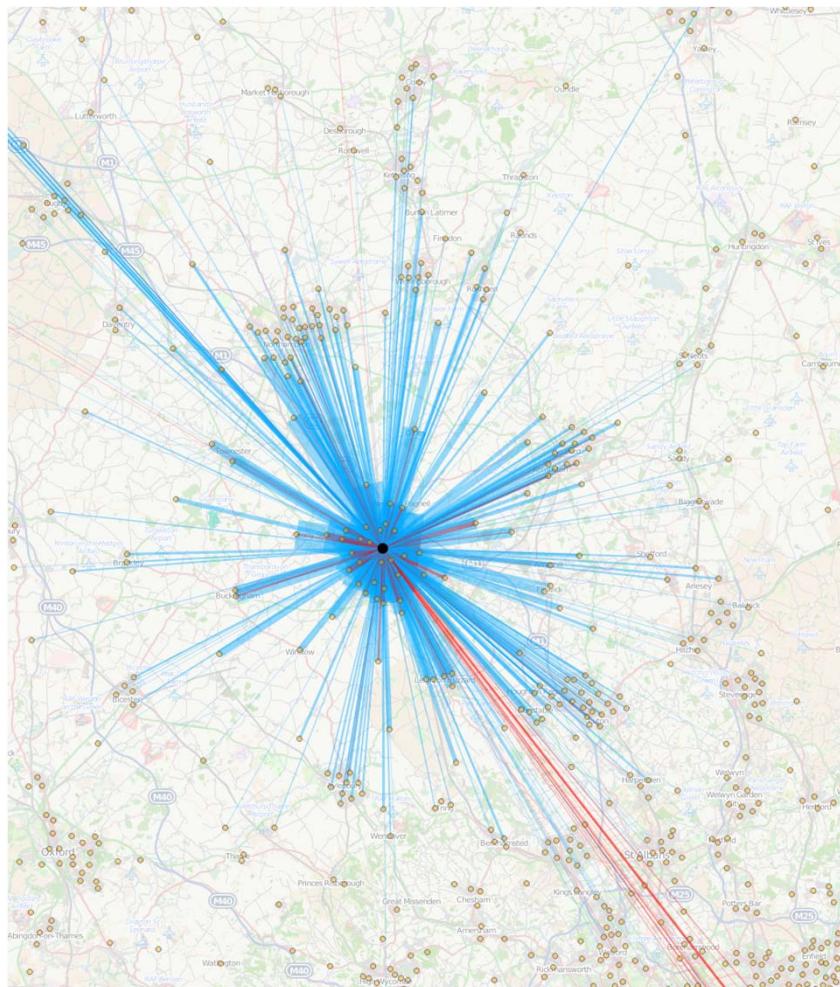
DataShine: Travel to Work Flows



DataShine: Travel to Work Flows

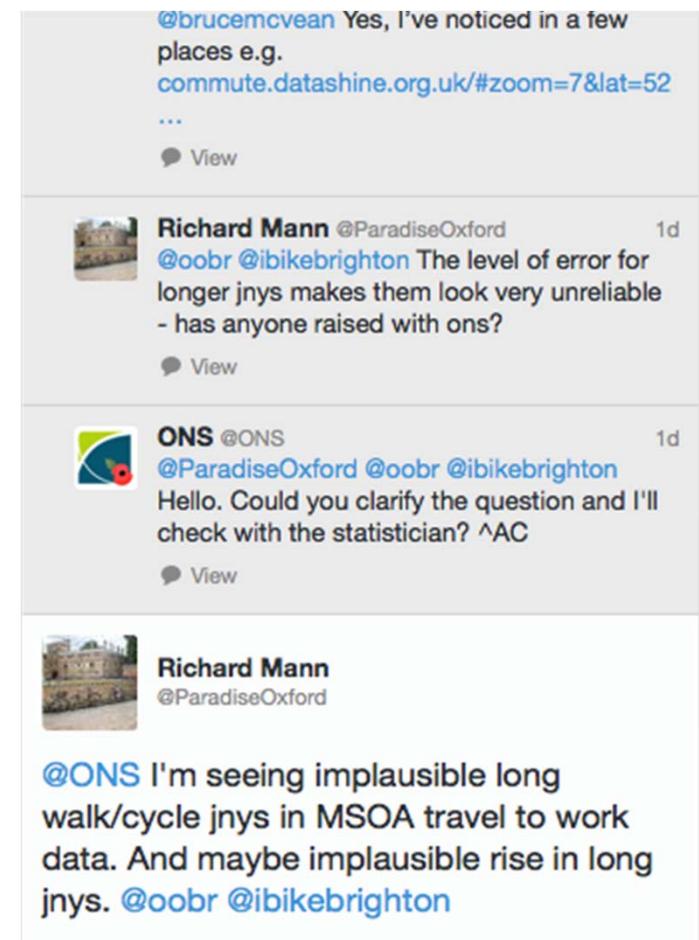
- MSOA population weighted centroids (ONS)
- Origin-Destination aggregate table for the 2011 census, split by mode of travel.
- OpenStreetMap (background map)
- OpenLayers 2
- PostgreSQL/PostGIS querying

DataShine: Travel to Work Flows



DataShine: Travel to Work Flows

- Some anomalous data (e.g. “walking” 200 miles)
 - People reporting their head office address as their workplace address, but actually work at a branch office
- Filtering out journeys with <6 people to minimise this problem, reduce “noise” and mitigate any low-level statistical record swapping

A screenshot of a Twitter thread from the DataShine account (@commute.datashine.org.uk). The thread consists of four tweets from different users, each with a small profile picture and a timestamp of '1d' ago.

@brucemcvean Yes, I've noticed in a few places e.g. commute.datashine.org.uk/#zoom=7&lat=52 ...

 **Richard Mann** @ParadiseOxford 1d
@oobr @ibikebrighton The level of error for longer jnys makes them look very unreliable - has anyone raised with ons?

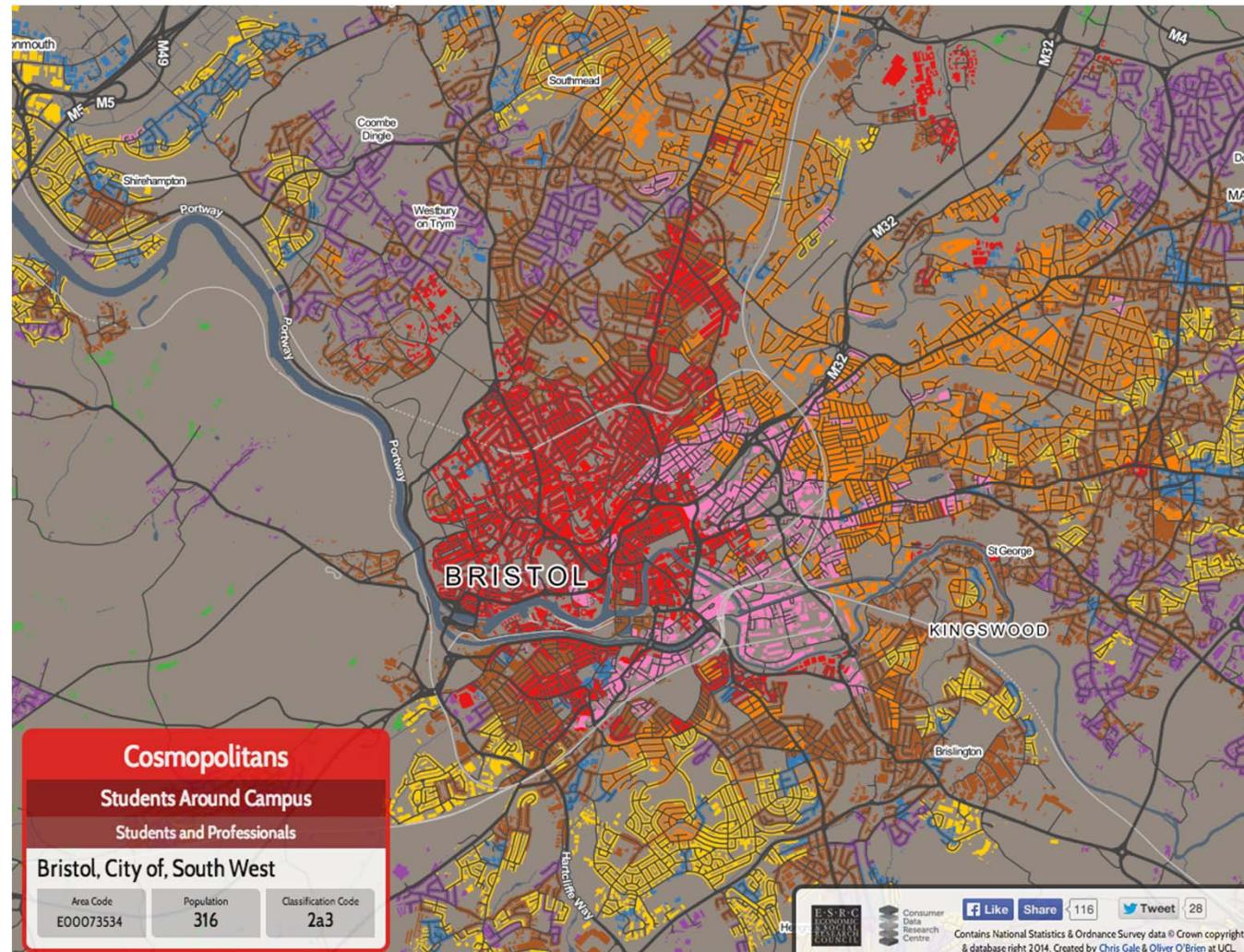
 **ONS** @ONS 1d
@ParadiseOxford @oobr @ibikebrighton Hello. Could you clarify the question and I'll check with the statistician? ^AC

 **Richard Mann** @ParadiseOxford
@ONS I'm seeing implausible long walk/cycle jnys in MSOA travel to work data. And maybe implausible rise in long jnys. @oobr @ibikebrighton

DataShine: OAC and LOAC

- Output Area-level geodemographic classifications
- Both using the DataShine “platform” to combine choropleth map with building outlines
 - But website much simpler than DataShine: Census
- 2011 Area Classification for Output Areas “OAC”
 - Produced by Dr Gale at UCL
 - Notably better for classifying London
 - 8 Supergroups, 26 groups, 76 subgroups, all with names
- <http://oac.datashine.org.uk/>

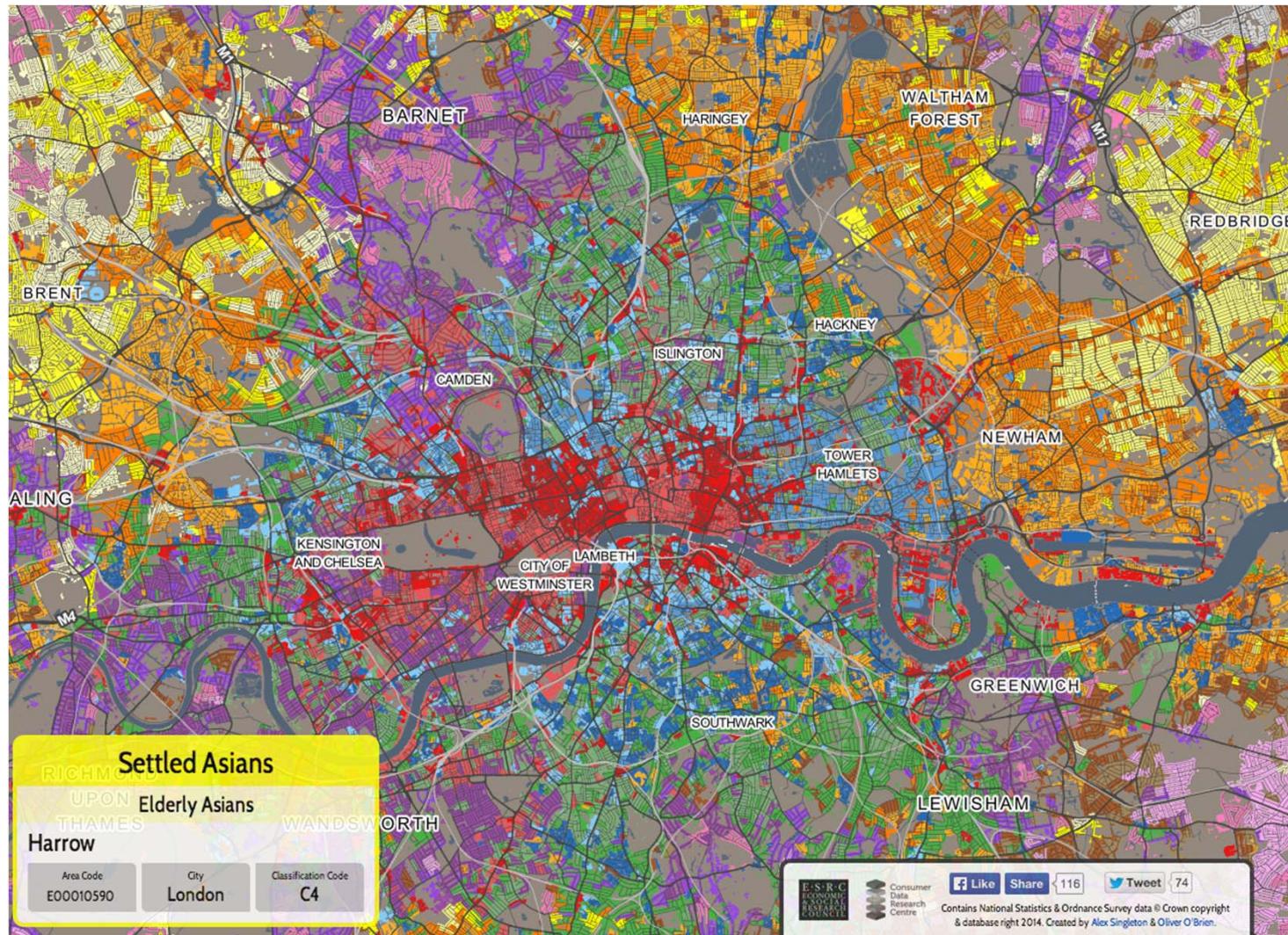
2011 Area Classification for Output Areas



London Output Area Classification

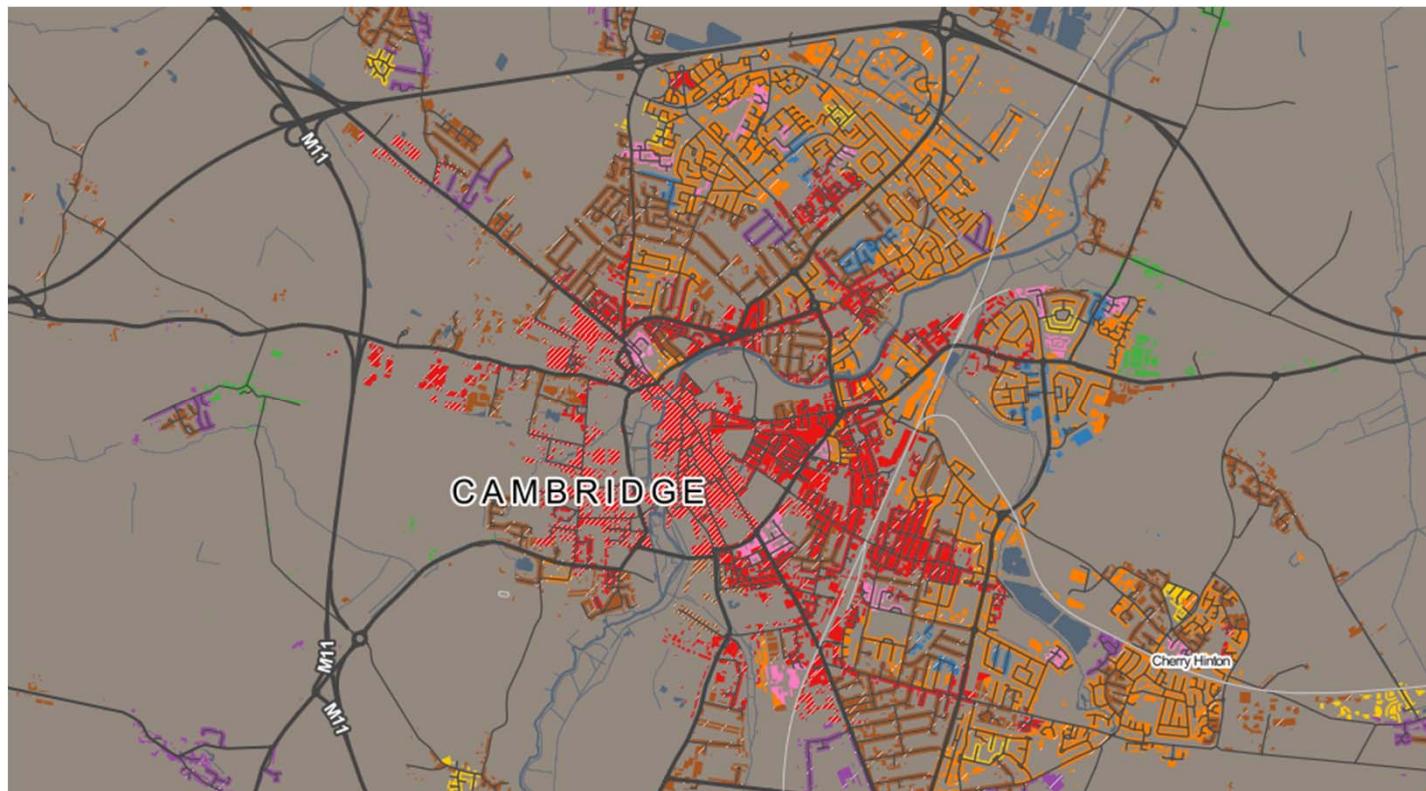
- “LOAC”
 - Focused entirely on London
 - Commissioned by the GLA
 - Produced by Dr Singleton at Liverpool University
 - Also has 8 “supergroups”
- <http://loac.datashine.org.uk/>
- The LOAC and OAC websites have individual supergroup toggle buttons

London Output Area Classification



Some Ongoing Research

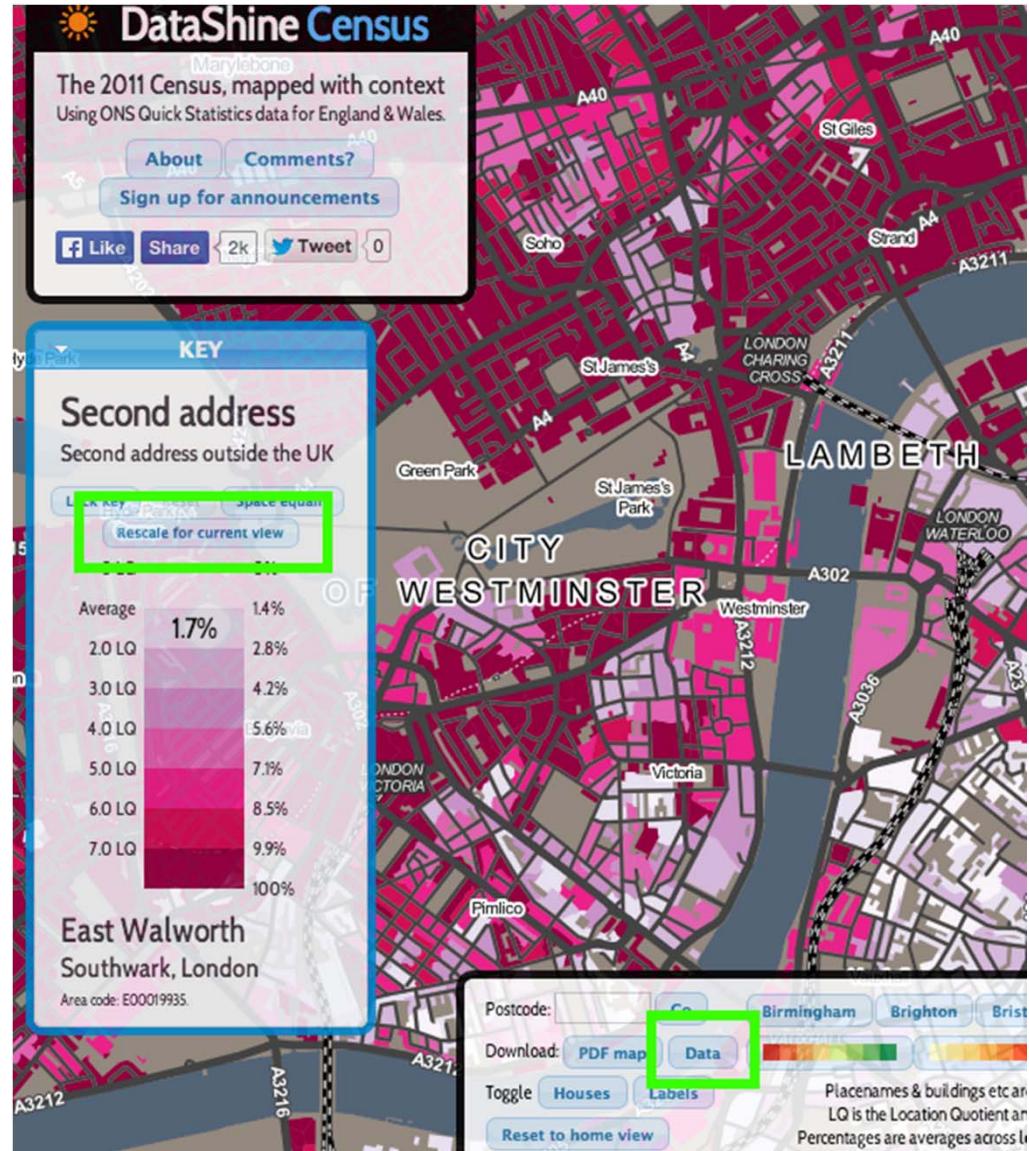
- Visualising uncertainty in the classification



Launching Today

- 2 “Power User” features for DataShine: Census
 - Local Area Rescaling
 - Local Area Data Download
- Available from today at the regular website:
<http://datashine.org.uk/>

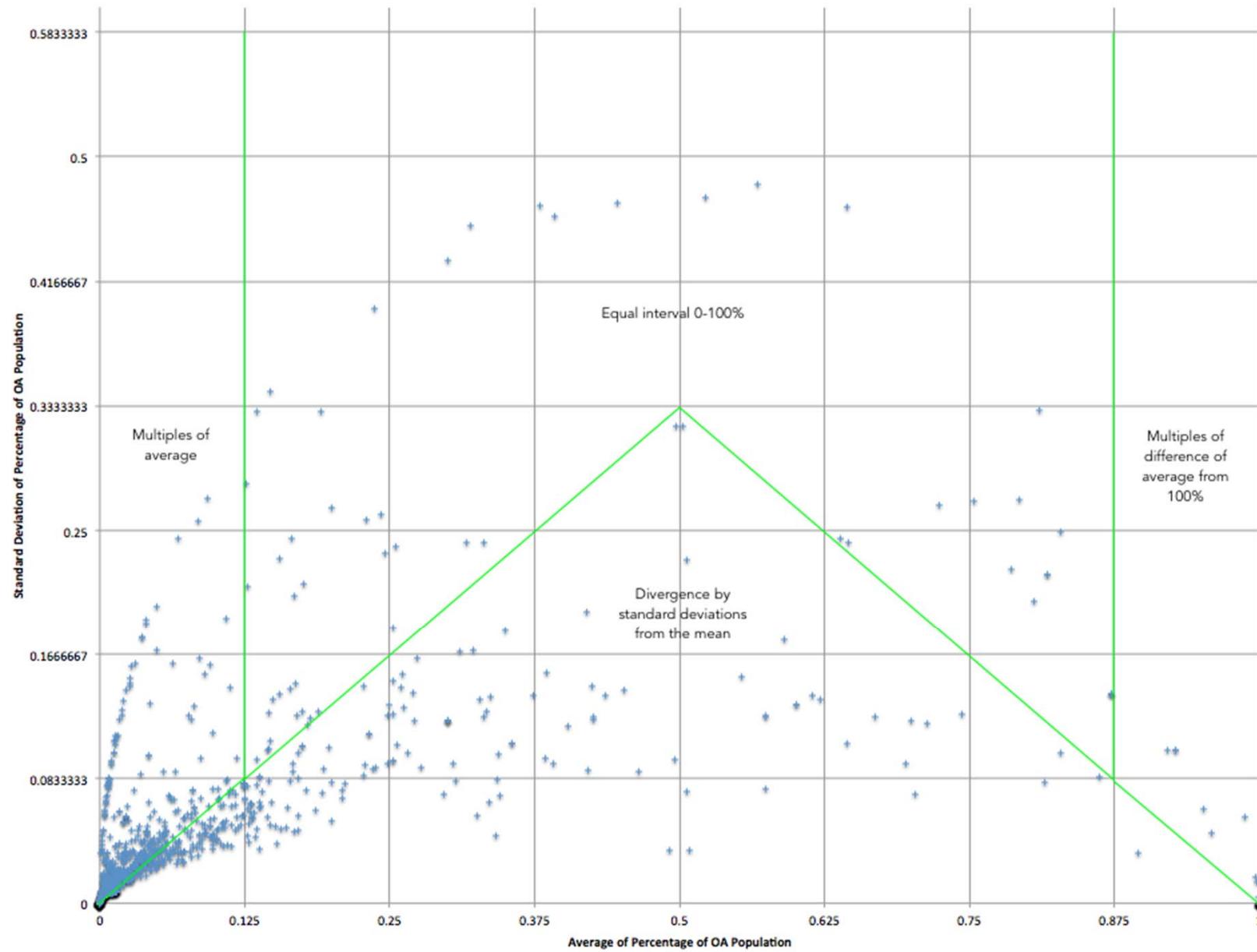
Launching Today



Local Area Rescaling

- Useful particularly outside of London as the city's unusual composition (w.r.t rest of England/Wales) tends to have undue influence on many measures.
- Measures local average (of geographical units) and standard deviation, and then rescales based on these measurements
 - The “binning strategy” can also change as it is based on thresholds of the average/standard deviation.

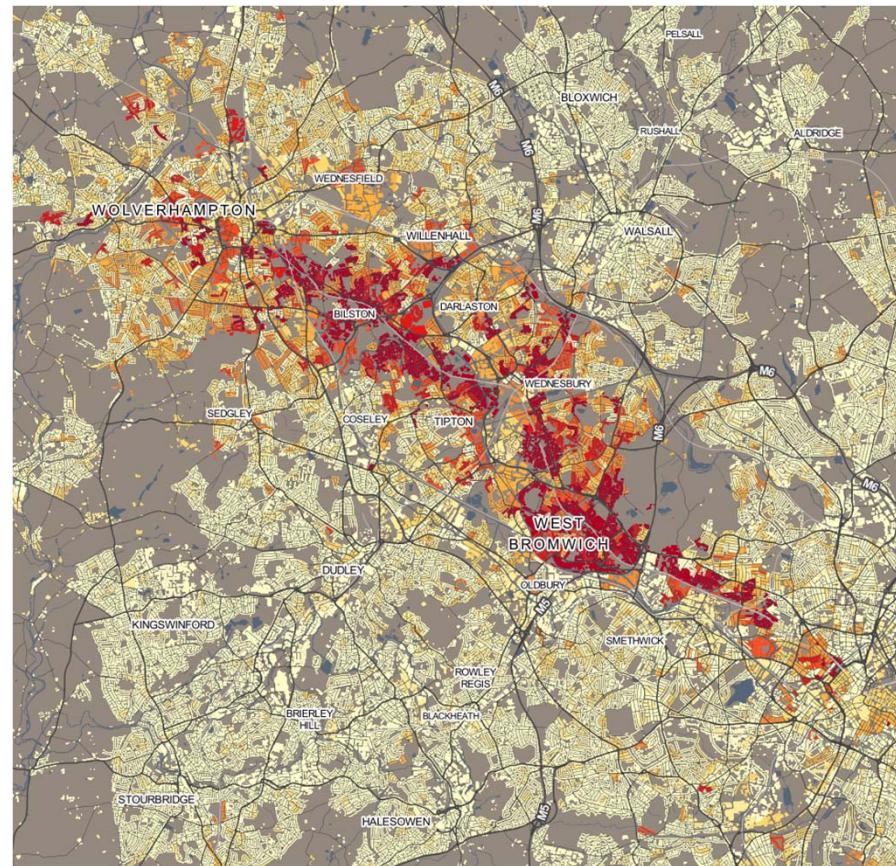
2011 Census Quick Statistics (Percentages) for England/Wales



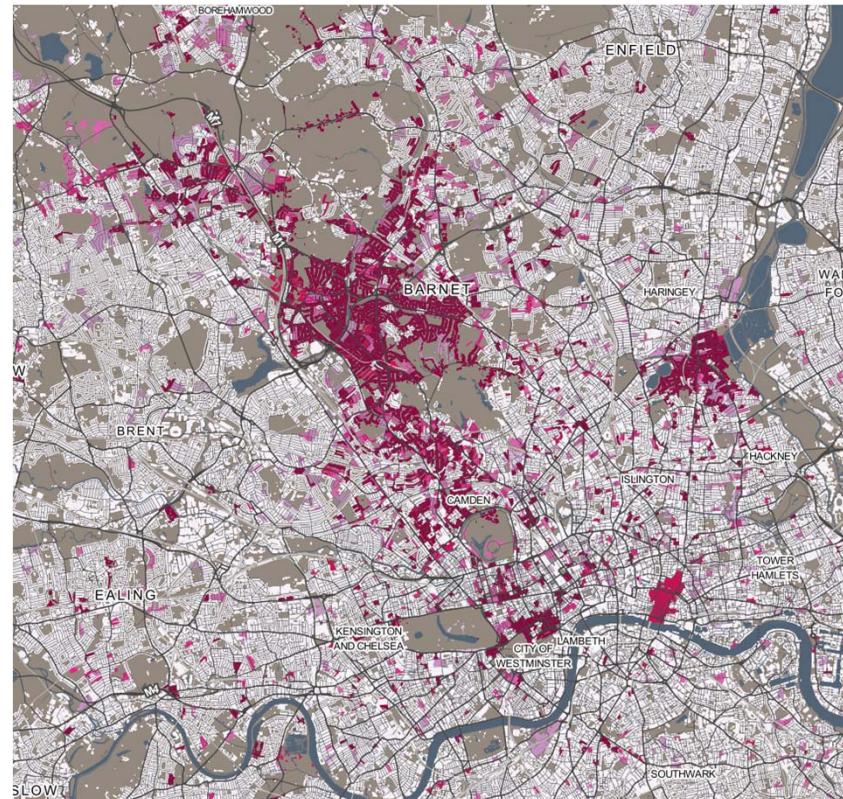
Local Area Rescaling: Examples

- Allows detail to come out for metrics with low local averages but still significant populations.
 - Detail of Birmingham metro usage (London underground skews the national picture)
- Reduces speckle/noise
 - Hebrew speakers in north London

Local Area Rescaling: Examples



Local Area Rescaling: Examples



Data Download

- Zoom to an area of interest, then click “Data”
 - CSV download, containing the metric and base population, percentage, area names and geography IDs
 - No hassle, very fast
- Link with geography files from ONS, in a GIS (e.g. QGIS) to rebuild a map
 - Filter by local authority (as LA & region names included)

Next Steps

- Next for BODMAS
 - Developing and opening the tools for DataShine to allow reuse.
- Next for DataShine
 - Including non-census social media datasets
 - IMD/IDACI
 - Land Registry (house prices/transactions)
 - Other datasets on data.gov.uk & London Data Store 2
 - Add synthesised datasets (e.g. multiple census variables)
 - Add Census 2011 data for Scotland & N.I.

Contact

- The DataShine Blog
 - <http://blog.datashine.org.uk/>
 - <http://blog.datashine.org.uk/comments/>
- Twitter
 - @oobr (me)
 - @spatialanalysis (James)
- Email
 - o.obrien [at] ucl.ac.uk