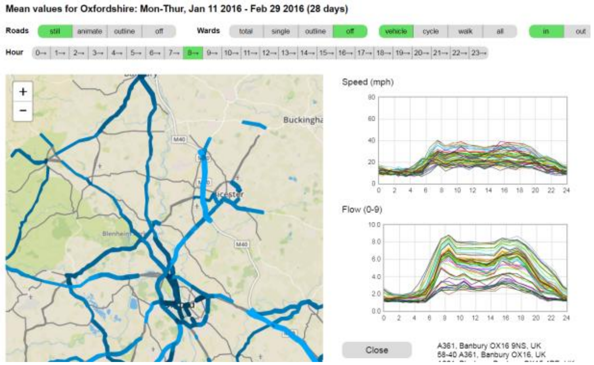
Estimating traffic disruption patterns with volunteer geographic information

Jonathan Bright
Oxford Internet Institute
@jonmbright
jonathan.bright@oii.ox.ac.uk

Background

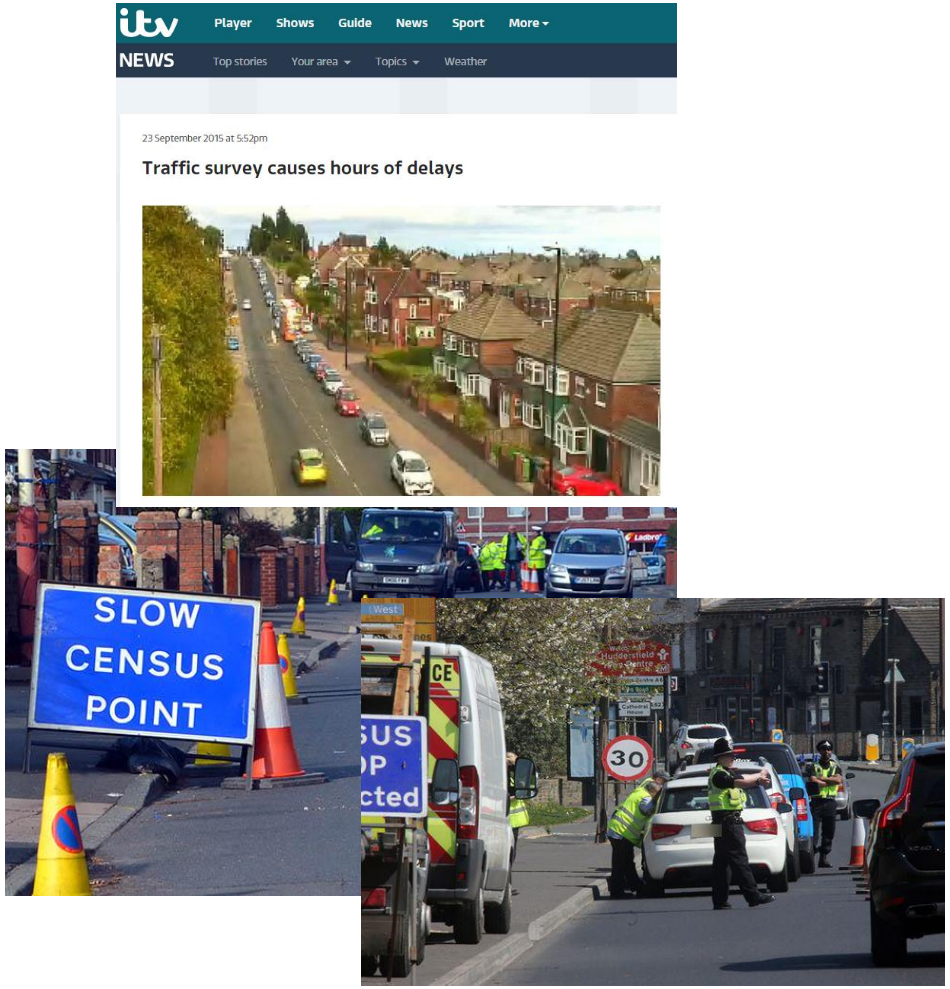
Introduction

- Oxford Internet Institute: dept at Oxford Uni studying impact of the internet on society
- We have a developing smart cities cluster (http://smartcities.oii.ox.ac.uk/) working on both practical and policy research
    - We are developing new data science technologies and building tools which local gov can use
    - We are also studying the impact of data science on government: winners and losers, power relations, etc.



- Particular interest in how (local) govt. can benefit from the 'data science' revolution

Why should local gov care? (1)



"…even when a recent census is available, crowded cities with large mobile populations are notoriously difficult to enumerate accurately with the result that urban populations are frequently under-enumerated" Cohen 2006.

Much of what we know about cities to date then has been gleaned from studies that are characterised by data scarcity (Kitchin, following Miller 2010 - The data avalanche is here. Shouldn't we be digging?).

Why should local gov care? (

- Local gov (in the UK) is facing a time of enormous financial difficulty
- Lack of money opens door to new ideas / thinking but also leaves very little room for manoeuvre
- Hence, interest in developing cheap/lightweight **‘smart proxies’**
    - Already seen loads of these in the last two days!
    - Our focus is on transport

Austerity

Britain’s local councils face financial crisis

Amid a painful fiscal squeeze, some authorities may soon be unable to meet their statutory obligations

News › Business › Business News

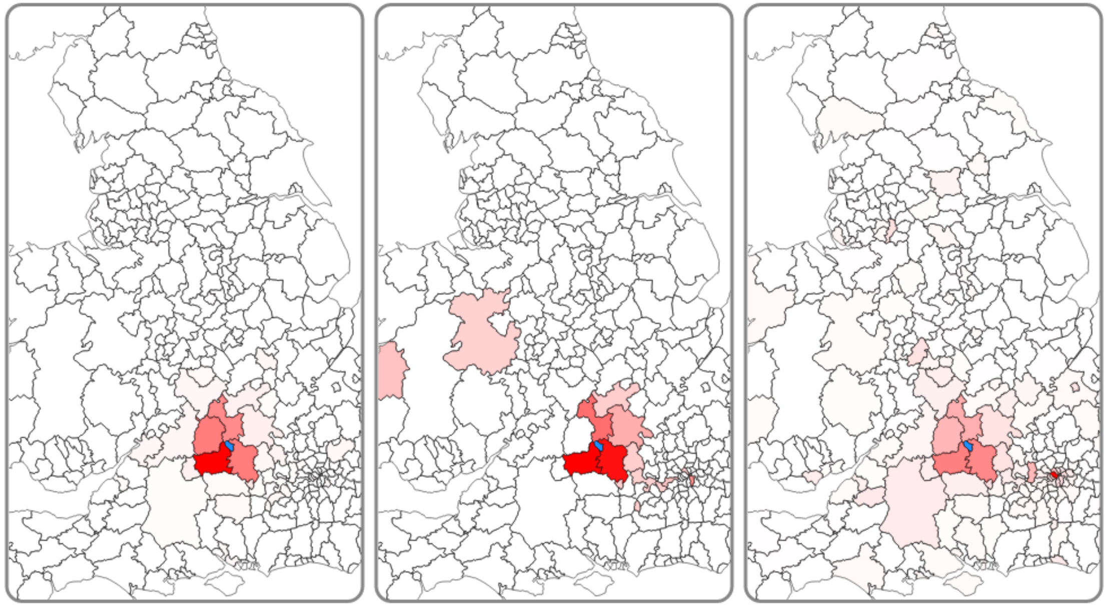
Local councils at financial breaking point due to austerity, warns National Audit Office

NAO estimates that if local authorities keep draining their reserves at the current rate, one in 10 will have exhausted them in just three years’ time

Tory county council runs out of cash to meet obligations

Northamptonshire is first to admit financial disaster for two decades in £10m shortfall

Just finished one paper on Twitter predictions of commuting



Work: Oxford

Live:

Oxford	54.9%
Vale of White Horse	13.0%
South Oxfordshire	6.6%
West Oxfordshire	6.6%

Live: Oxford

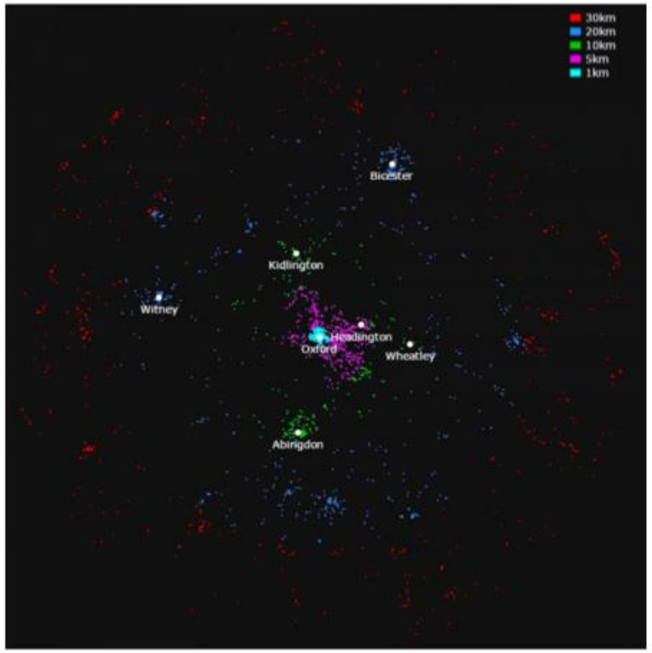
Work:

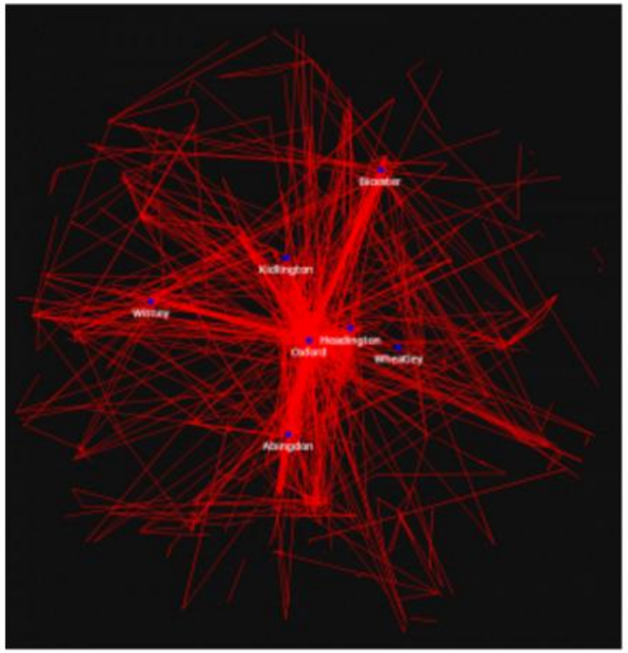
Oxford	80.9%
Vale of White Horse	3.7%
South Oxfordshire	3.4%
Cherwell	2.1%

Live - Twitter estimate: Oxford

Work - Twitter estimate:

Oxford	54.9%
Westminster	5.5%
Vale of White Horse	2.7%
South Oxfordshire	2.6%





McNeill et al. *EPJ Data Science* (2017) 6:24
DOI 10.1140/epjds/s13688-017-0120-x
EPJ.org
EPJ Data Science
a SpringerOpen Journal
REGULAR ARTICLE
Open Access

CrossMark

Estimating local commuting patterns from geolocated Twitter data

Graham McNeill, Jonathan Bright* and Scott A Hale

*Correspondence:
jonathan.bright@oii.ox.ac.uk
Oxford Internet Institute, University of Oxford, Oxford, United Kingdom

Abstract

The emergence of large stores of transactional data generated by increasing use of digital devices presents a huge opportunity for policymakers to improve their

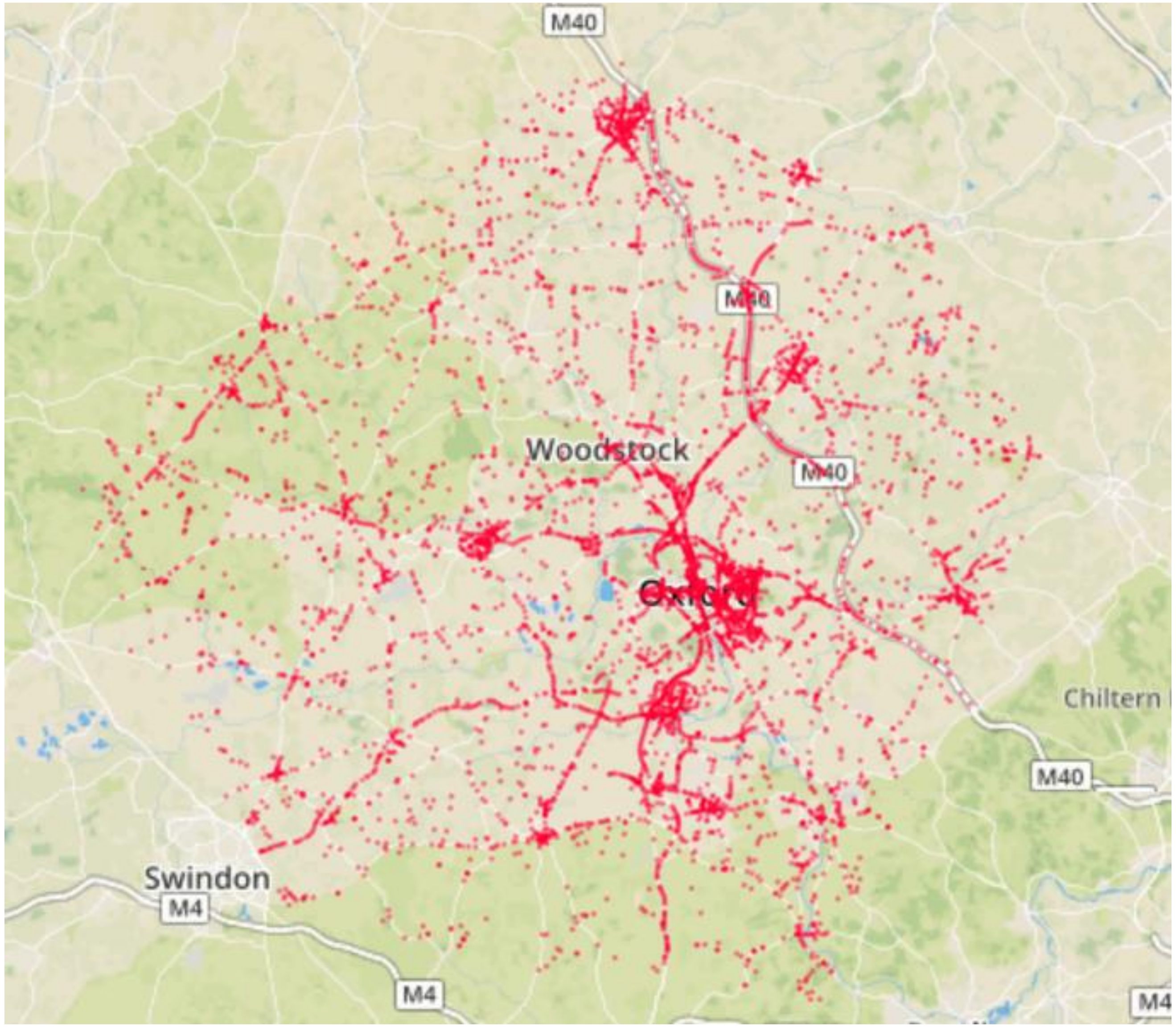
Project: to what extent can we estimate the volume of traffic disruptions using OpenStreetMap data?

Interest

- Most local gov. know where traffic blackspots are
- But only weak idea of how different types of land use alter these traffic dynamics
    - Of course, lots of companies offering traffic forecasting / ABMs, but £££
- Aim -> offer a *free* model which is of some use

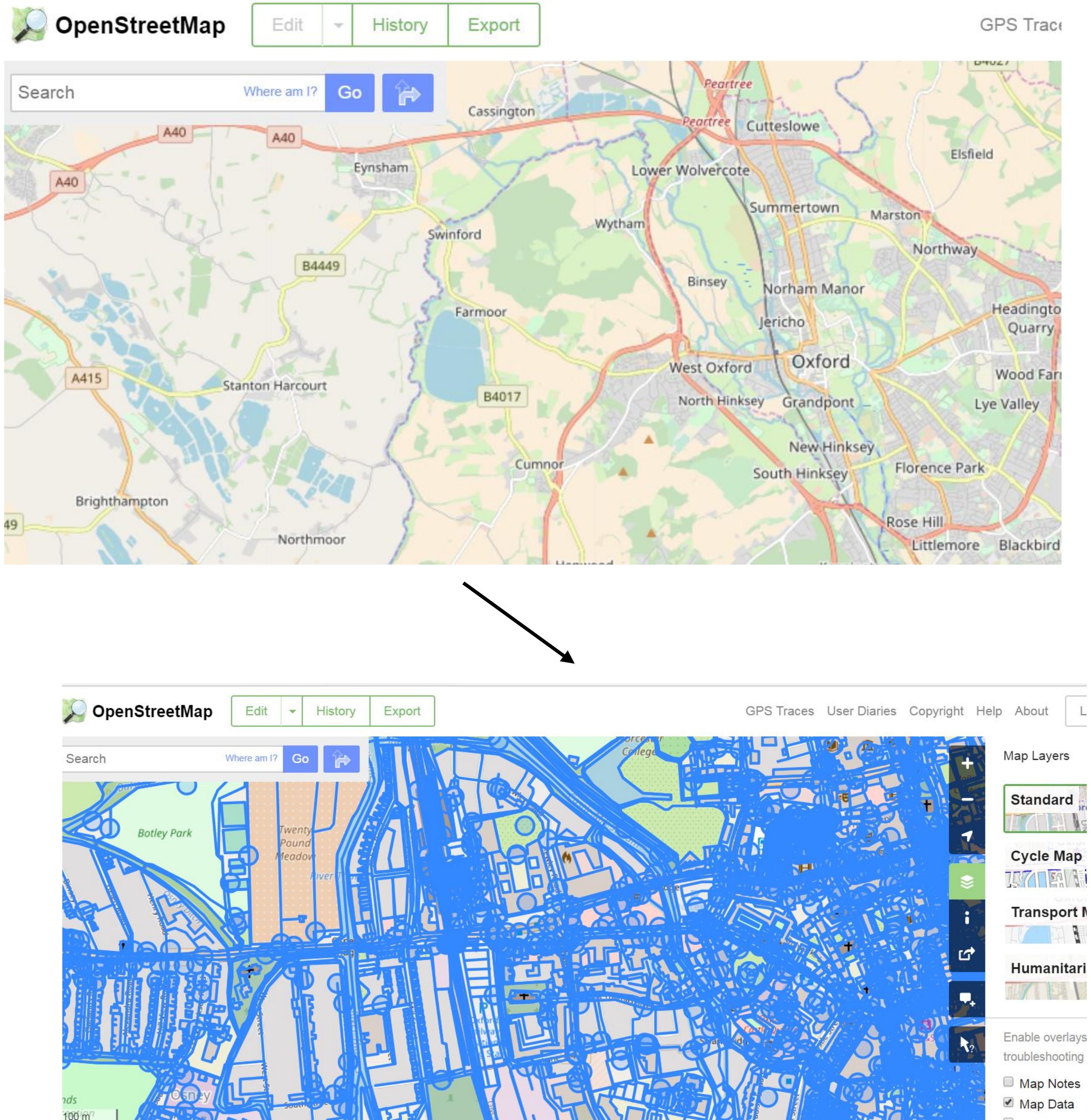
Data (1)

- Traffic disruption reports from major traffic mobile phone app for Oxfordshire
  - (?) not clear how representative it is or whether this makes a difference!
- ~6,500 observation points around Oxfordshire for one month
  - For each we have a count of number of incidences of disruption from 'smooth running'
- Over 1 million traffic jams ☺



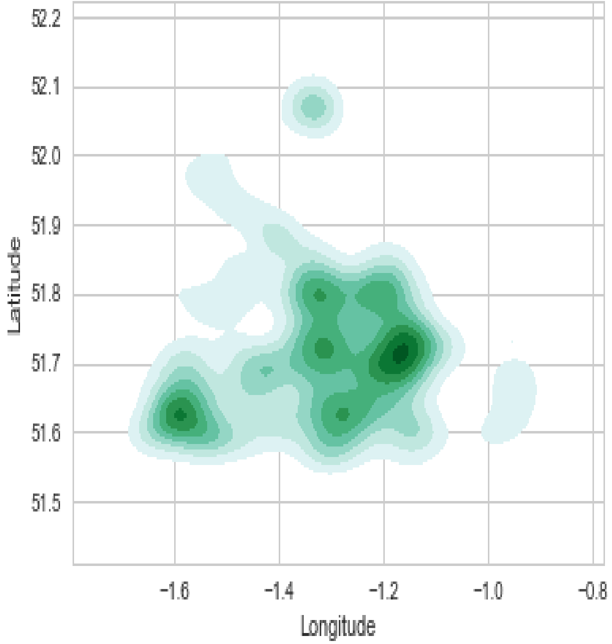
Data (2)

- OpenStreetMap – collaborative mapping project
- Contains rich ‘feature’ data
  - In another project we validated this as around 50% complete
  - Potentially extremely valuable and underexploited in local context
- We took all data relating to land use
  - Ignored transport features so far



Methods

- Point features are transformed into 2d kernel density estimates
  - Can estimate, for example, a 'bakery density' at any given map point
- Density estimates for each feature type generated at each point



Results (1)

- We first classify all OSM points of interest into six commonly used land use meta categories
  - Generates a kind of 'baseline' model which could be produced with existing institutional data
- Then run a simple OLS regression of count of jams (log) vs land use density at each point
- Results are a bit counter-intuitive and overall 'predictive' power of the model is low (adj. R2 = 0.11)

Variable	**Estimate**
Residential	-0.09
Industrial	-0.18
Recreational	-0.10
Institutional	0.14
Green Space	0.26
Commercial	0.32
Observations	6529
Adjusted R^2	0.11

Results (2)

- We then make use of more granular classification of 44 feature types
- Adj R2 jumps to 0.56 -> shows usefulness of granular data

house	0.0639	**bicycle_parking**	0.029	**telephone**	-0.0291	**apartments**	-0.0908
farmland	0.555	**garage**	-0.0146	**entrance**	-0.0453	**allotments**	0.0778
residential	0.6092	**place_of_worship**	0.0088	**restaurant**	0.0079	**convenience**	-0.0028
parking	0.2406	**pub**	-0.0425	**grave_yard**	-0.0497	**tennis**	0.071
meadow	0.1846	**bench**	0.0372	**garden**	-0.067	**soccer**	0.0937
forest	0.236	**playground**	-0.028	**pitch**	0.0811	**recreation_ground**	0.0269
farmyard	0.3094	**terrace**	-0.0426	**cafe**	-0.0727	**recycling**	-0.0028
post_box	0.0786	**school**	0.0422	**commercial**	0.0765	**clothes**	0.0659
grass	0.0241	**park**	-0.011	**fast_food**	0.0502	**reservoir**	0.1166
university	0.1179	**industrial**	0.1388	**retail**	0.0482	**supermarket**	-0.0689
atm	0.0161	**toilets**	0.0238	**farm**	0.2515		

Results (3)

- Model gives estimates of how things we might expect to cause local traffic jams vary with actual traffic jams

house	0.0639	**bicycle_parking**	0.029	**telephone**	-0.0291	**apartments**	-0.0908
farmland	0.555	**garage**	-0.0146	**entrance**	-0.0453	**allotments**	0.0778
residential	0.6092	**place_of_worship**	0.0088	**restaurant**	0.0079	**convenience**	-0.0028
parking	0.2406	**pub**	-0.0425	**grave_yard**	-0.0497	**tennis**	0.071
meadow	0.1846	**bench**	0.0372	**garden**	-0.067	**soccer**	0.0937
forest	0.236	**playground**	-0.028	**pitch**	0.0811	**recreation_ground**	0.0269
farmyard	0.3094	**terrace**	-0.0426	**cafe**	-0.0727	**recycling**	-0.0028
post_box	0.0786	**school**	0.0422	**commercial**	0.0765	**clothes**	0.0659
grass	0.0241	**park**	-0.011	**fast_food**	0.0502	**reservoir**	0.1166
university	0.1179	**industrial**	0.1388	**retail**	0.0482	**supermarket**	-0.0689
atm	0.0161	**toilets**	0.0238	**farm**	0.2515		

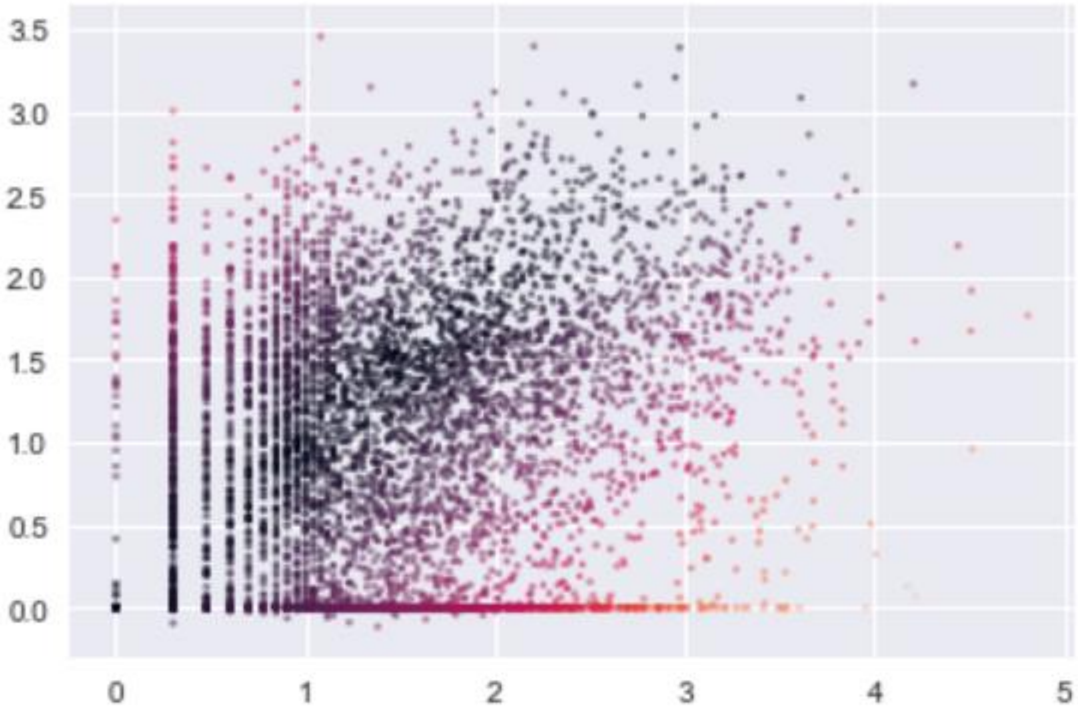
Results (4)

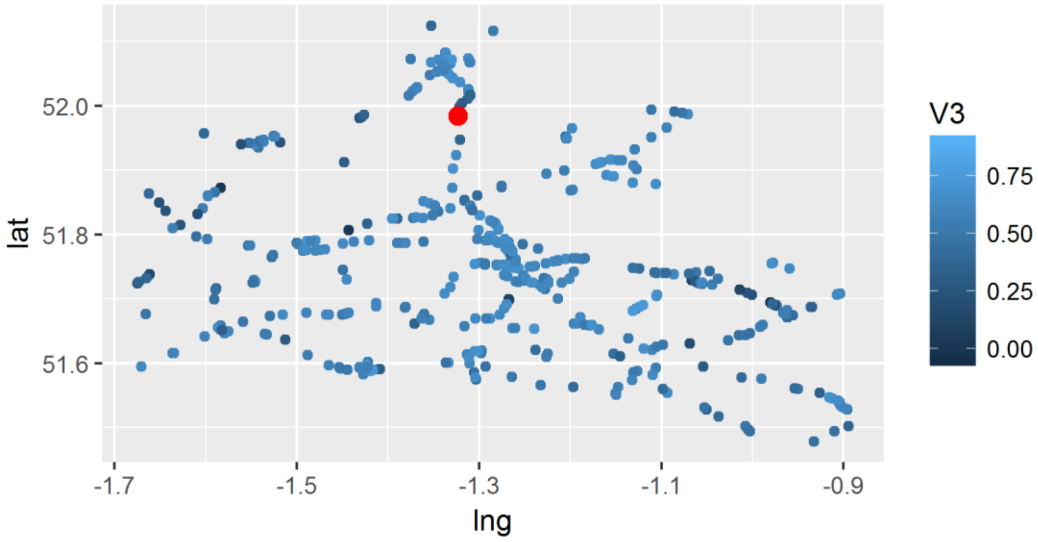
- Though correlation != causation of course

house	0.0639	bicycle_parking	0.029	telephone	-0.0291	apartments	-0.0908
farmland	0.555	garage	-0.0146	entrance	-0.0453	allotments	0.0778
residential	0.6092	place_of_worship	0.0088	restaurant	0.0079	convenience	-0.0028
parking	0.2406	pub	-0.0425	grave_yard	-0.0497	tennis	0.071
meadow	0.1846	bench	0.0372	garden	-0.067	soccer	0.0937
forest	0.236	playground	-0.028	pitch	0.0811	recreation_ground	0.0269
farmyard	0.3094	terrace	-0.0426	cafe	-0.0727	recycling	-0.0028
post_box	0.0786	school	0.0422	commercial	0.0765	clothes	0.0659
grass	0.0241	park	-0.011	fast_food	0.0502	reservoir	0.1166
university	0.1179	industrial	0.1388	retail	0.0482	supermarket	-0.0689
atm	0.0161	toilets	0.0238	farm	0.2515		

Results (3)

- Model + data also allows us to find outliers
  - i.e. unusually bad (or good) traffic
- Permits a more nuanced view of what a traffic blackspot is!





Discussion

- OSM provides (relatively) static features so unsurprisingly doesn’t explain all variation
  - But can provide a baseline model with some explanatory power
  - Hence can offer a rough answer to the question ‘what impact will placing another café at point a have on traffic jams?’
  - And the price is right!
- Next steps
  - Incorporate some sense of the road network structure: being between a populated place and an ‘interesting’ place is likely to generate jams
  - Would like to have some temporally varying data -> but Twitter is too coarse ☹

Thanks!

@jonmbright

jonathan.bright@oii.ox.ac.uk

OXFORD INTERNET INSTITUTE

UNIVERSITY OF OXFORD

Estimating local commuting patterns from geolocated Twitter data
G McNeill, J Bright, SA Hale, 2017,
EPJ Data Science 6 (1), 24

OpenStreetMap data for alcohol research: Reliability assessment and quality indicators
J Bright, S De Sabbata, S Lee,
B Ganesh, DK Humphreys, 2018
Health & Place 50, 130-136

The Lightweight Smart City and Biases in Repurposed Big Data
C Voigt, J Bright, 2016
Proceedings of HUSO 7