





UCL University College London

Big Data, High Frequency and Low Frequency Cities

Michael Batty

Centre for Advanced Spatial Analysis
CASA-UCL

July 6th 2018

m.batty@ucl.ac.uk



@jmichaelbatty



SHUFE







上海财经大学
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS

Outline

- The High Frequency City: A Switch in Urban Thinking from the Long to the Short Term
- Networks & Flows: The Glue Keeps the City Together
- Linking Networks: The Theoretical Dilemma
- An Interlude About Big Data
- Low Frequency Cities: Scaling to Systems of Cities
- Mobility, Transit, & Real-Time Streaming: Examples of Urban Dynamics in the Very Short Term
- Learning about Mobility from the Data
- Challenges for Urban Informatics: Theory, Modelling, and Applications: Policy and Management

The High Frequency City: A Switch in Urban Thinking from the Long to the Short Term

- Most of our thinking about cities has been about how cities evolve in the long term: cities over years to decades to centuries to epochs
- Most control but not planning has been about the short term: cities over minutes, hours, days and years
- Sensing what is going on automatically in cities is clearly something that happens in real time and this focus is changing from long to the short term, from years to seconds-minutes-hours-days-months ……
- This we call the **high frequency city** and we will distinguish this thinking from the **low frequency city**.
- Frequency is in terms of time not volume

- When I was a student of planning in Manchester in the mid 1960s, we did a project looking at linkages between industries in the centre city. It was based on data collected manually by researchers through questionnaires about what firms related to what and in what manner – through production linkages, through labour supply and so on – what we now call industrial clusters
- But these linkages were timeless – if anything one supposed they would continue to exist certainly into the middle range future
- The notion that we could capture them day by day was fanciful even through we knew that things happened day by day – it simply wasn't possible due to our inability to collect the data.

All our analytics then were largely borne in a timeless universe. Economic models – input-output; gravity models; transportation assignment, location theory; urban rent models; all of urban economics; etc…..

Largely we ignored time because we could do little about it, not because we didn't think it important.

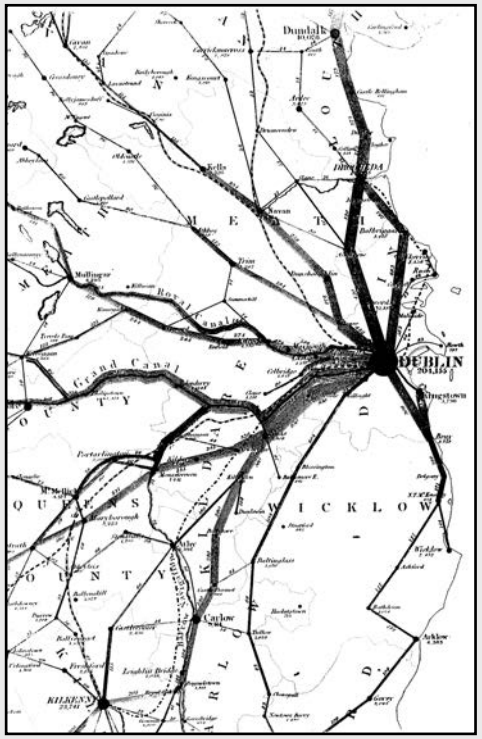
Cities looked as though they were in equilibrium – all our models were cross-sectional despite some brave attempts to try and model dynamics

But all this has changed. To an extent urban informatics and the smart city defined as *the embedding of computation into cities and the consequent control* has thrown time onto the agenda in a big way

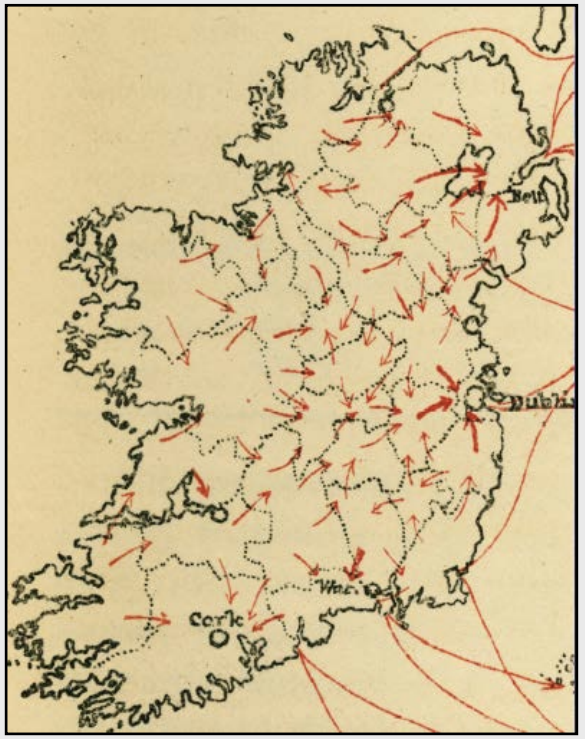
And along with this, has come the idea that this dynamics takes place on networks

Networks & Flows: Keeping the City Together

- The origins of cities as networks go back to Leonardo and before – largely to the low frequency city
- Networks and flows are different sides of the same coin and have been abstracted to visualise the form of cities for nearly 200 years.

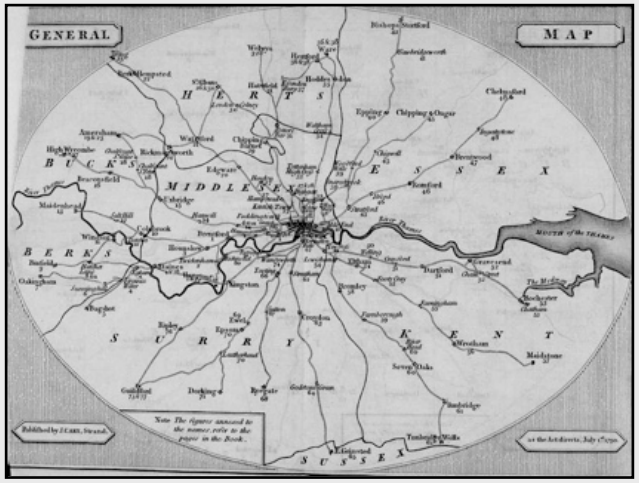


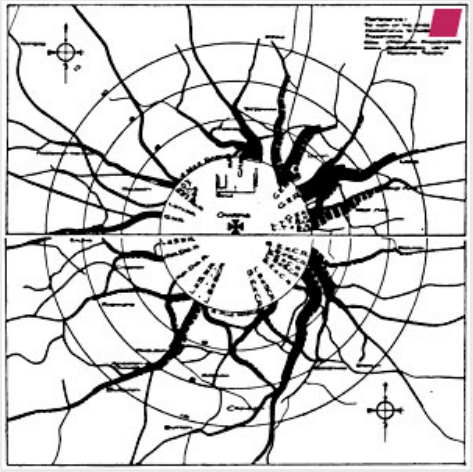
Harness, 1837

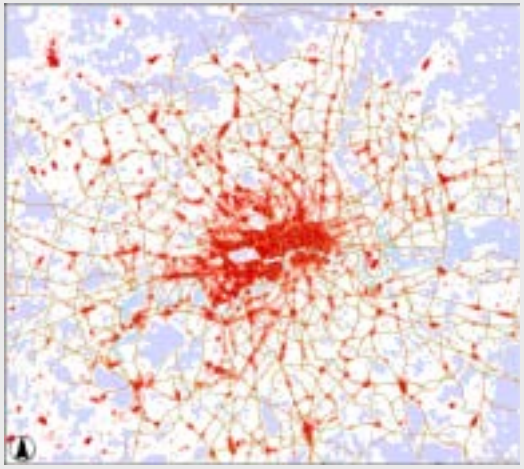


Ravenstein, 1888

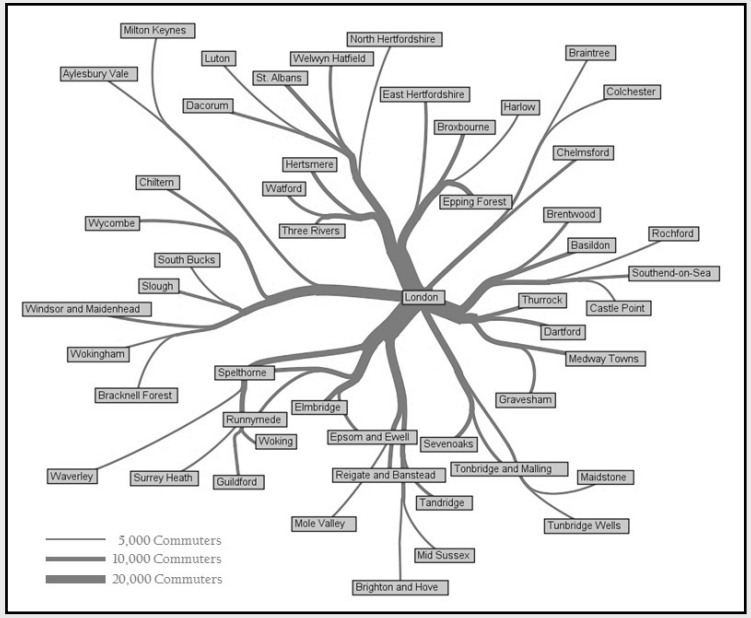
- There are many such flow patterns for the biggest cities – here are some for London



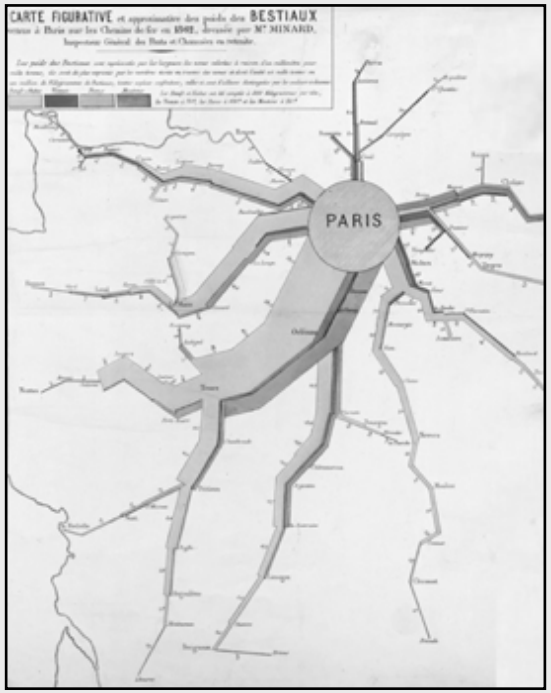


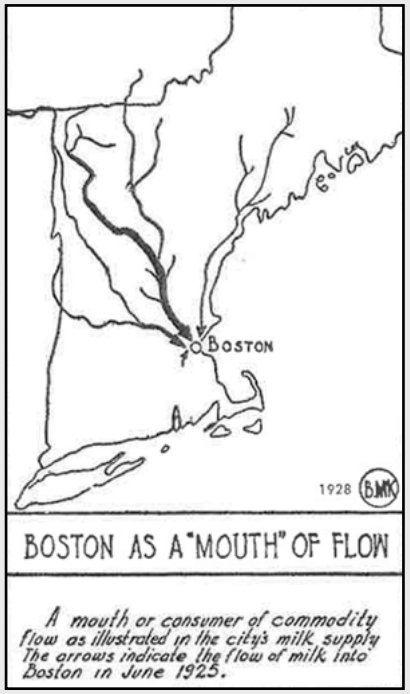


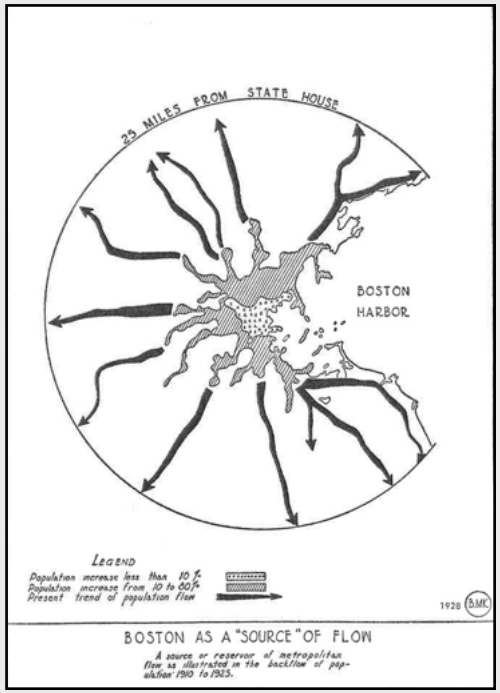


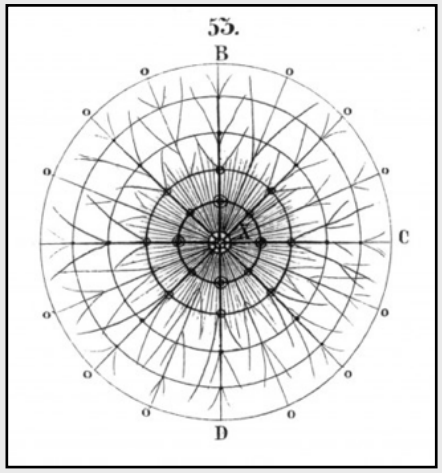


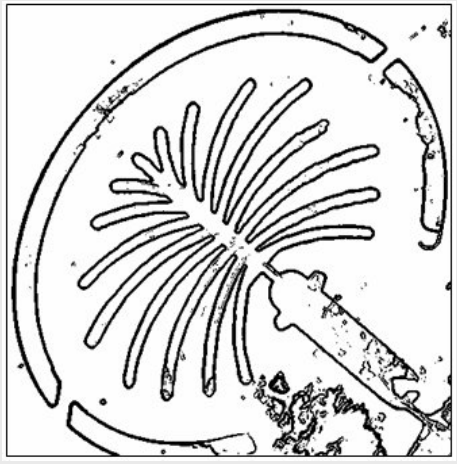
- Here are some more real and abstract flow maps

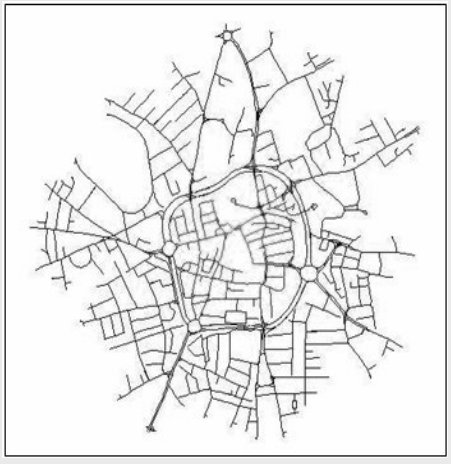












Linking Networks: The Theoretical Dilemma

- There are now many different kinds of networks, based on different kinds of channels and different kinds of flows. Urban Informatics is all about how very new kinds of flow – based on information as well as energy are beginning to dominate the city
- We do not have good theory to know how networks and markets determine how activities locate and change in cities.
- We don't have good ideas about how these networks are all interlinked.
- Although we now have good network science, we don't have good ways of connecting our models of flows to networks. And our models of flows are not that good. Moreover flows are all about dynamics.

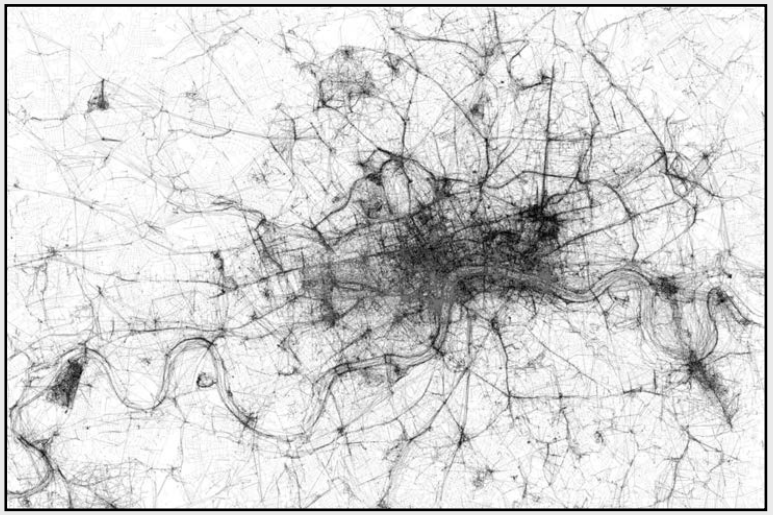
- We have lots of data now about networks which can be measured in terms of material flows and ethereal (digital flows) but how do we relate them.
- Most of those I have just illustrated deal with material – people, commodities and so on and we now have some many representations of digital flows from social and related media
- Here I will show some: ***Facebook*** – friendship(?) links



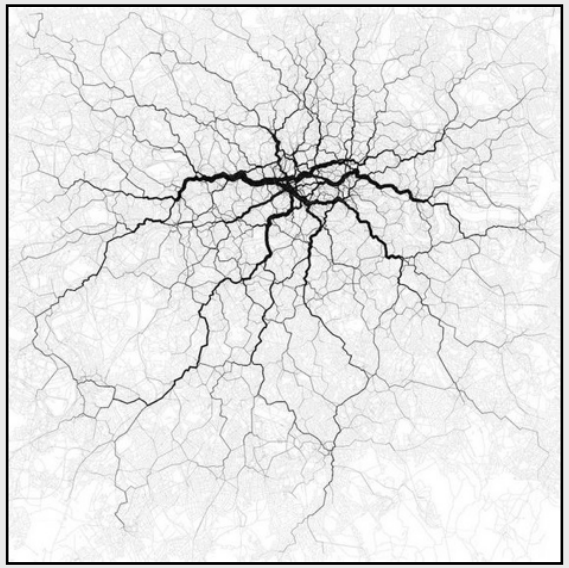
Nightlights – energy



Flickr locations: London:



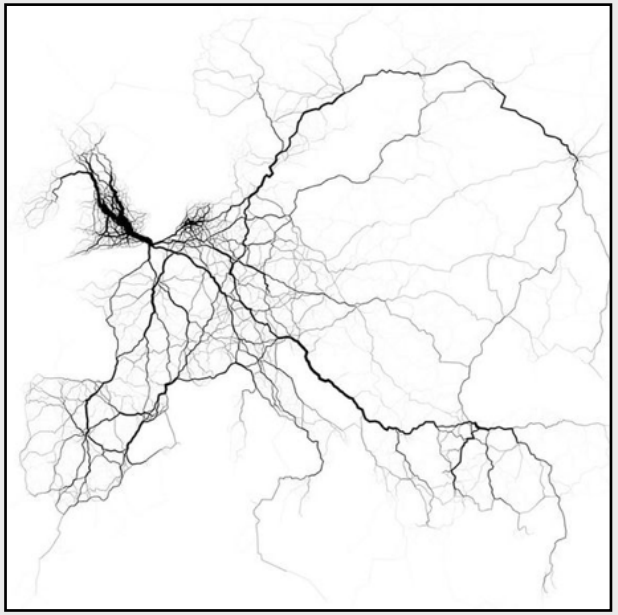
Tweet locations



Tweets

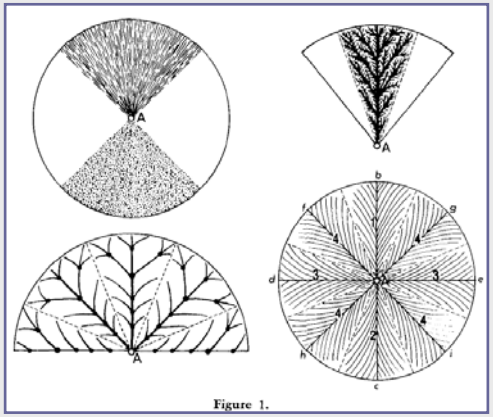
Across

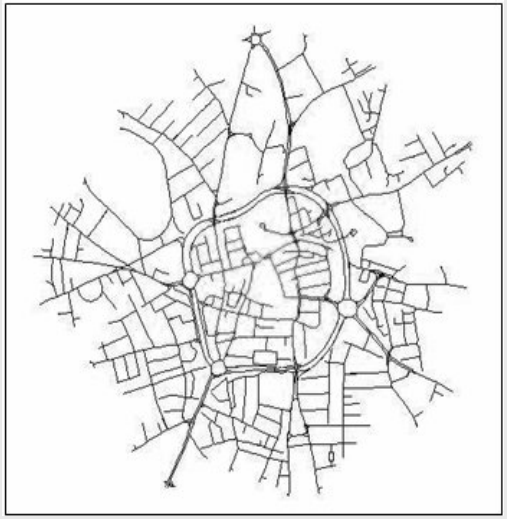
Europe

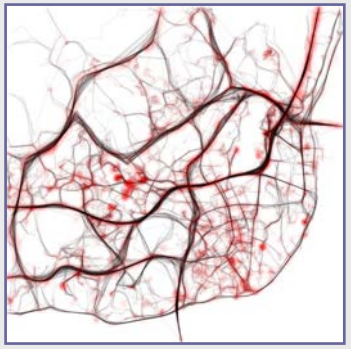


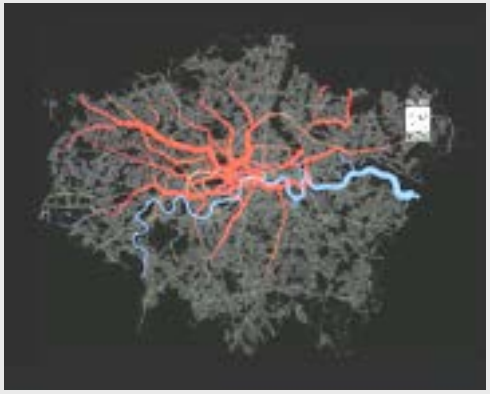
We also have some new methods for dealing with networks and flows – statistical methods based on ***scaling*** for example

And geometrical methods involving ***fractals***; but we don't have that much on the dynamics of networks – in fact we don't have that much on the dynamics of cities









- Ok: Urban Informatics is about **networks, flows and flow dynamics**; but a general theory of urban informatics is temporal and well as spatial but it is not restricted to the short term. However it is biased to the short term in my view.
- It connects up to much of what we have already been developing wrt to quant geography, spatial analysis, urban economics, cognitive psychology and so on but it is at the applied end of things
- Most of the connections have not yet been made – we urgently need to begin to make them and one of the great challenges is to define the field – generically and in terms of how it relates to other domains of academia and application

An Interlude About Big Data: Introducing Computation

Now these kinds of networks when they are of any size generate big data – there is a wonderful ***YouTube*** video about the early 1950s, where very very rudimentary computing was used to solve what was then a massive network problem

British Railways wanted to compute rates for shipping freight between their key stations of which then there were some n= 5500. The problem is of the order n^2

Then of course the matrix was quite large 5500 x 5500 having 30.25 million cells (~30,000K) – quite big if you only have 2K of RAM based on mercury delay lines and external paper tape, mag tape punched cards

How did they do it? I don't have time to tell you but ….

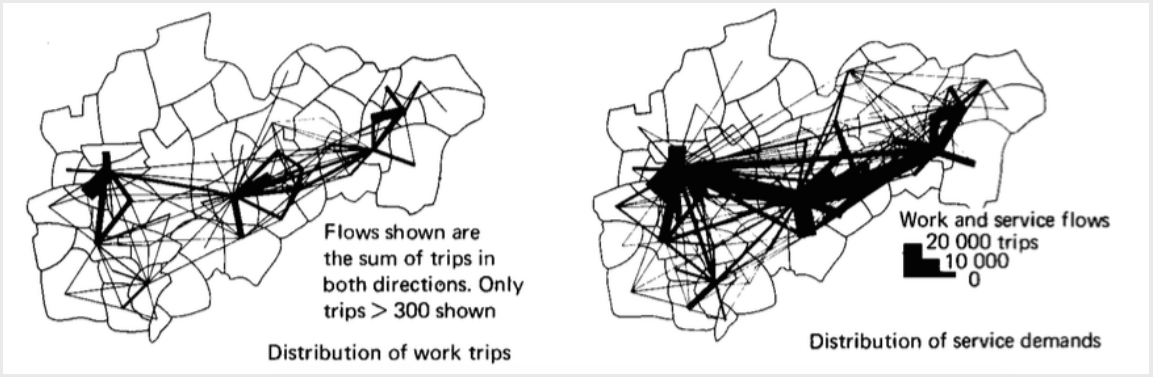


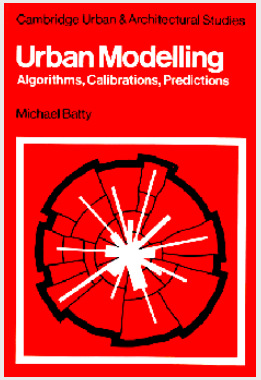




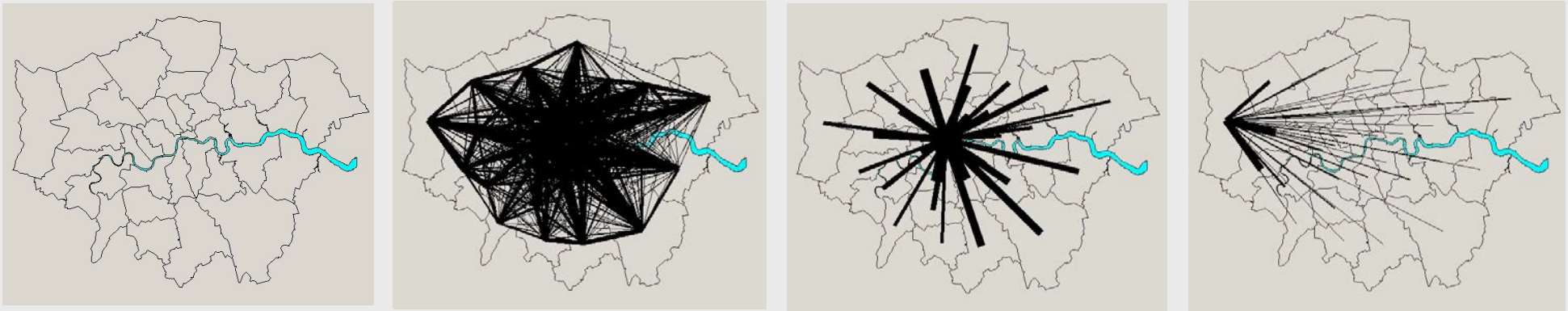
https://www.youtube.com/watch?v=pcBJfkE5UwU

I got involved in this world of networks a long time ago and we always dealt with 'big' data. An early model circa 1967-8 Central and NE Lancs

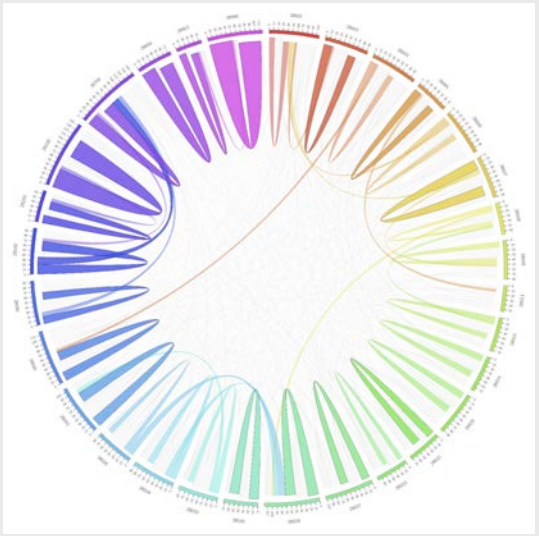




M. Batty (1976)
Urban Modelling
Cambridge UP

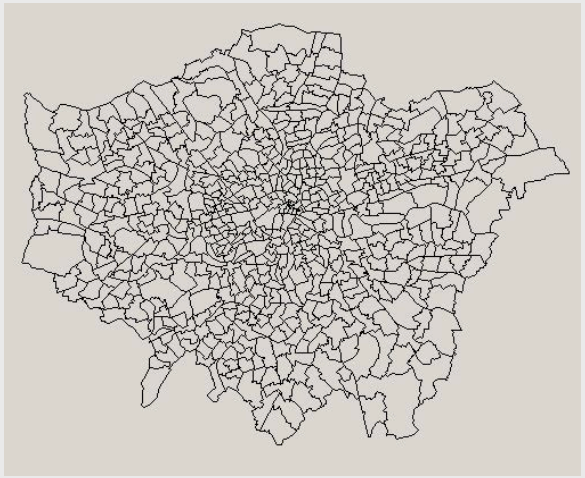


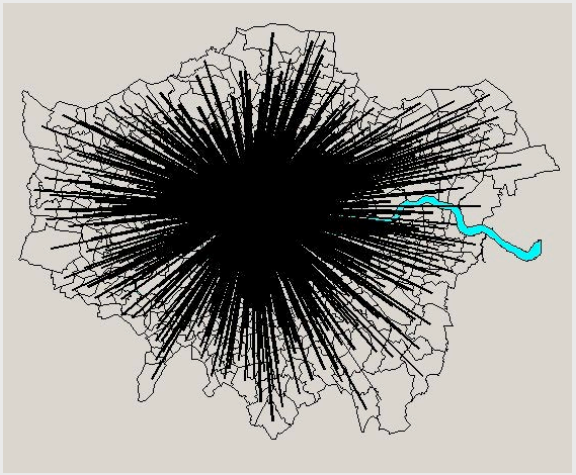
$n^2=33^2=1089$, not so big but hard to visualise

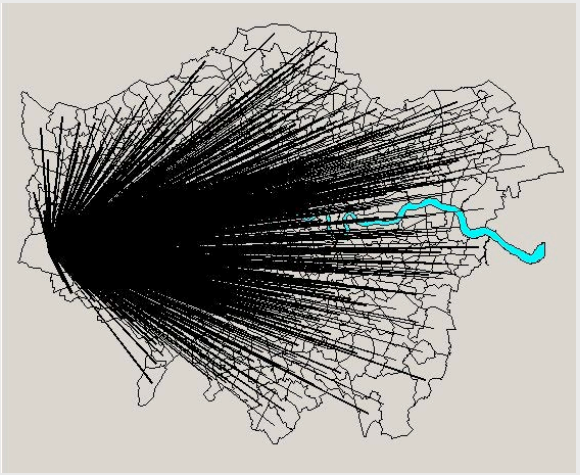




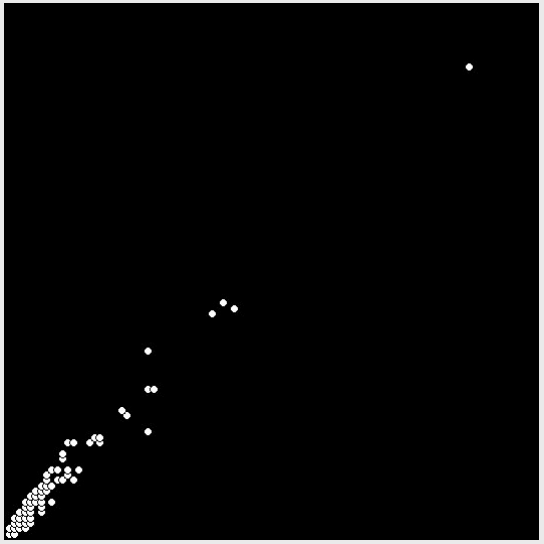
If we scale up to more reasonable levels of resolution – census block groups and bigger – tracts, then

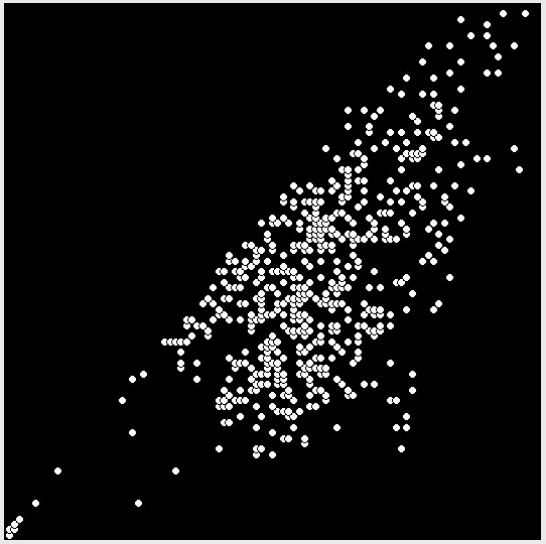


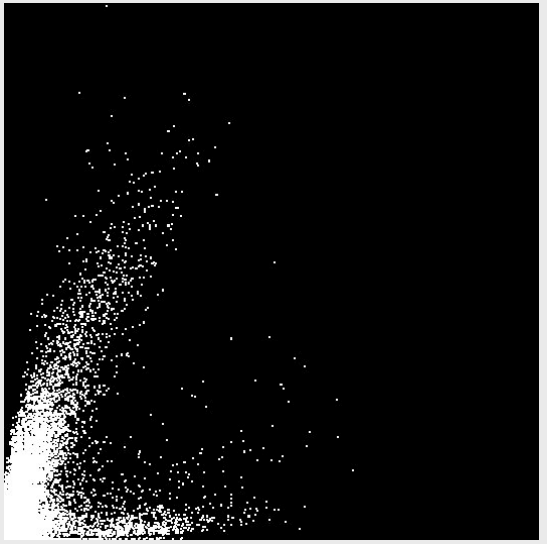




$n^2=633^2=400{,}689$, bigger but impossible to visualise









Even our statistics break down when we get large numbers like over several thousand as you can see on the left and above right for 400K data points where the pattern is highly convoluted. This is from a discrete choice model.

Low Frequency Cities: Scaling Networks to Systems of Cities

Ok so I now want to do two things – first of all scale up these traditional land use transport models to entire city systems – the megalopolis – in the UK this is the basically urban Great Britain

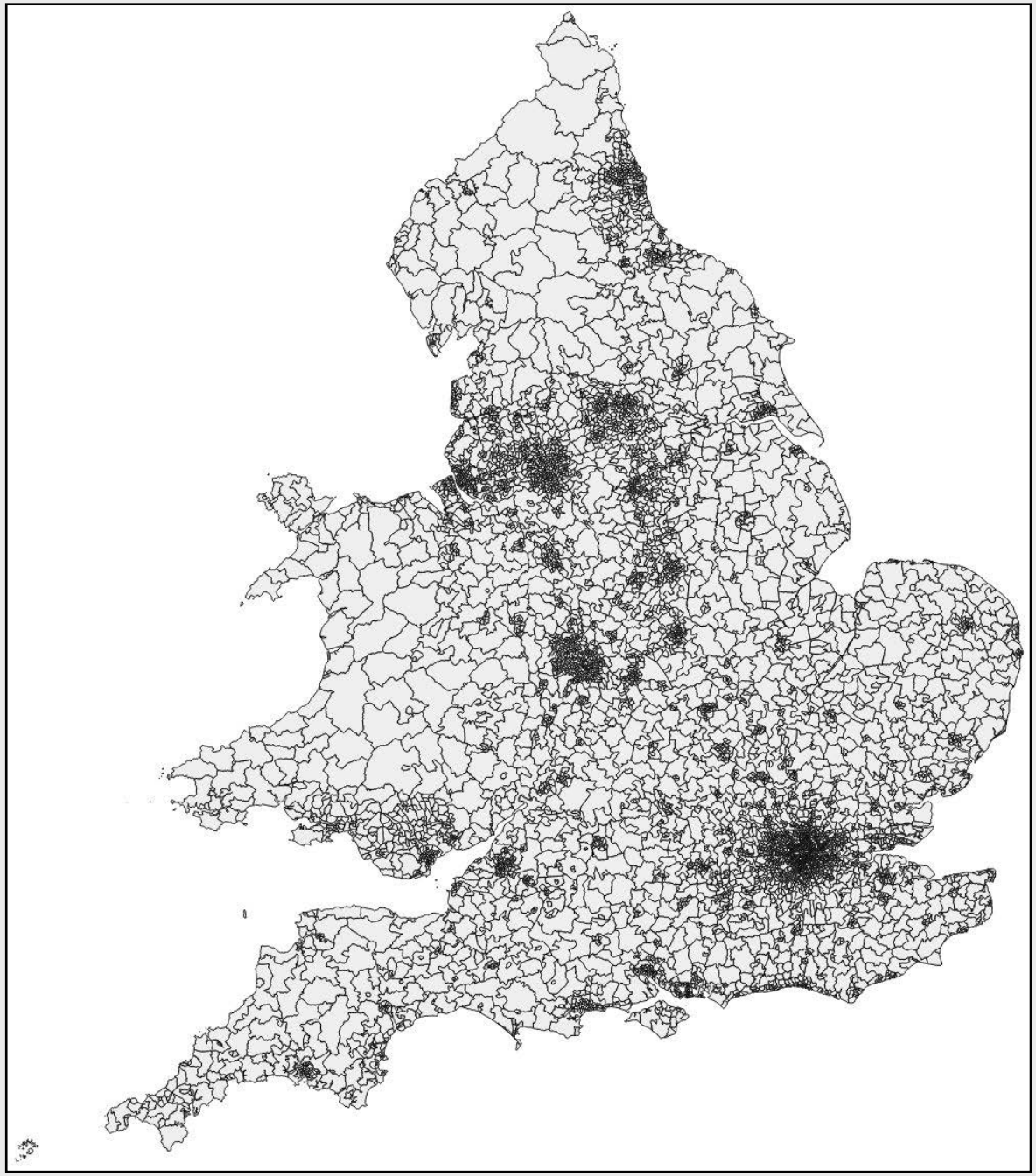
One of the consequences of ever bigger and faster computers and of computer networks is that we can handle bigger and bigger problems. So my first example is working with longer term change, scaling these models up to really big more open systems

The second example is at the very fine scale where we look at the dynamics of networks in the high frequency city – on a second by second basis – the other extreme in fact

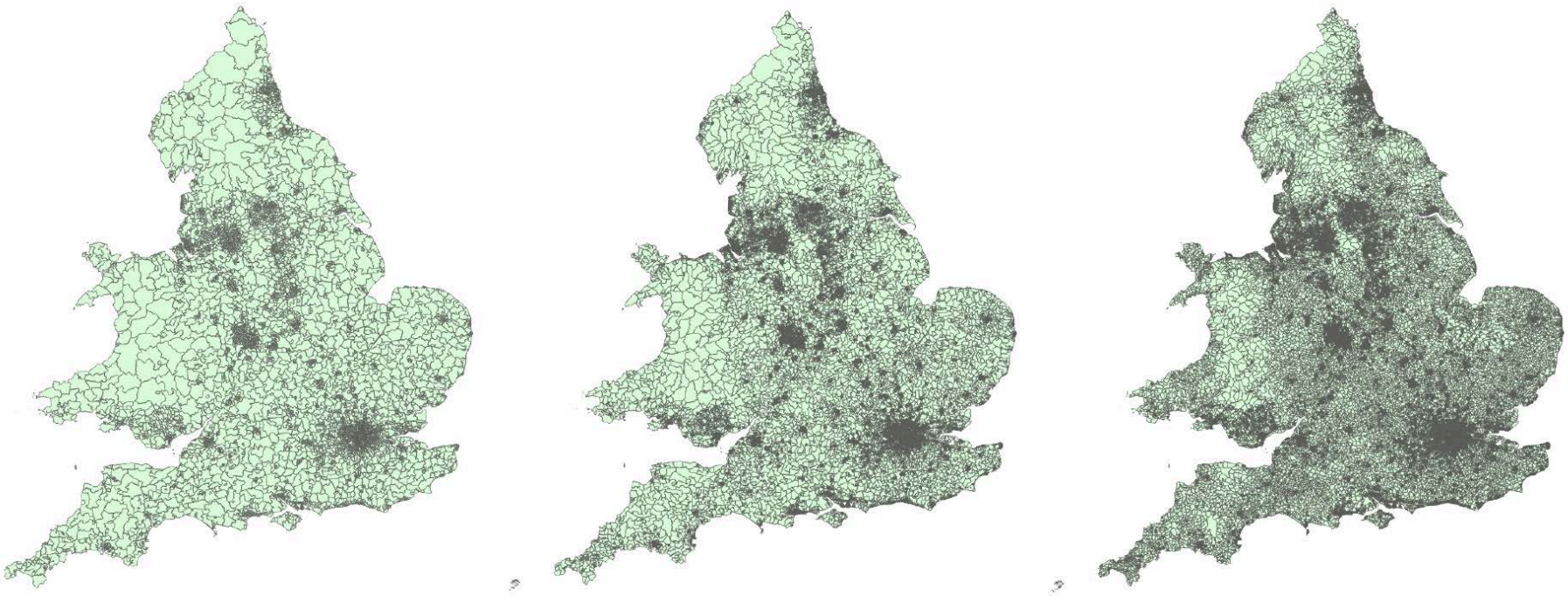
We will in fact scale up our models of flows to the level of spatial units called MSOAs of which there are 7201 in the UK – then there are $n^2=7201^2=51,854,401$ flows and the networks need to capture this – they are even finer scale in fact.

Circa 52 million points is an issue but our models run in a matter of seconds but that is a lot of data to store – ok it is sparse but sparsity isn't structured so we can't easily partition and in any case we want to compute any future possible flows between places that do not have any flow at present, say.

Here is the problem scaled up and this is what we are grappling with at present.



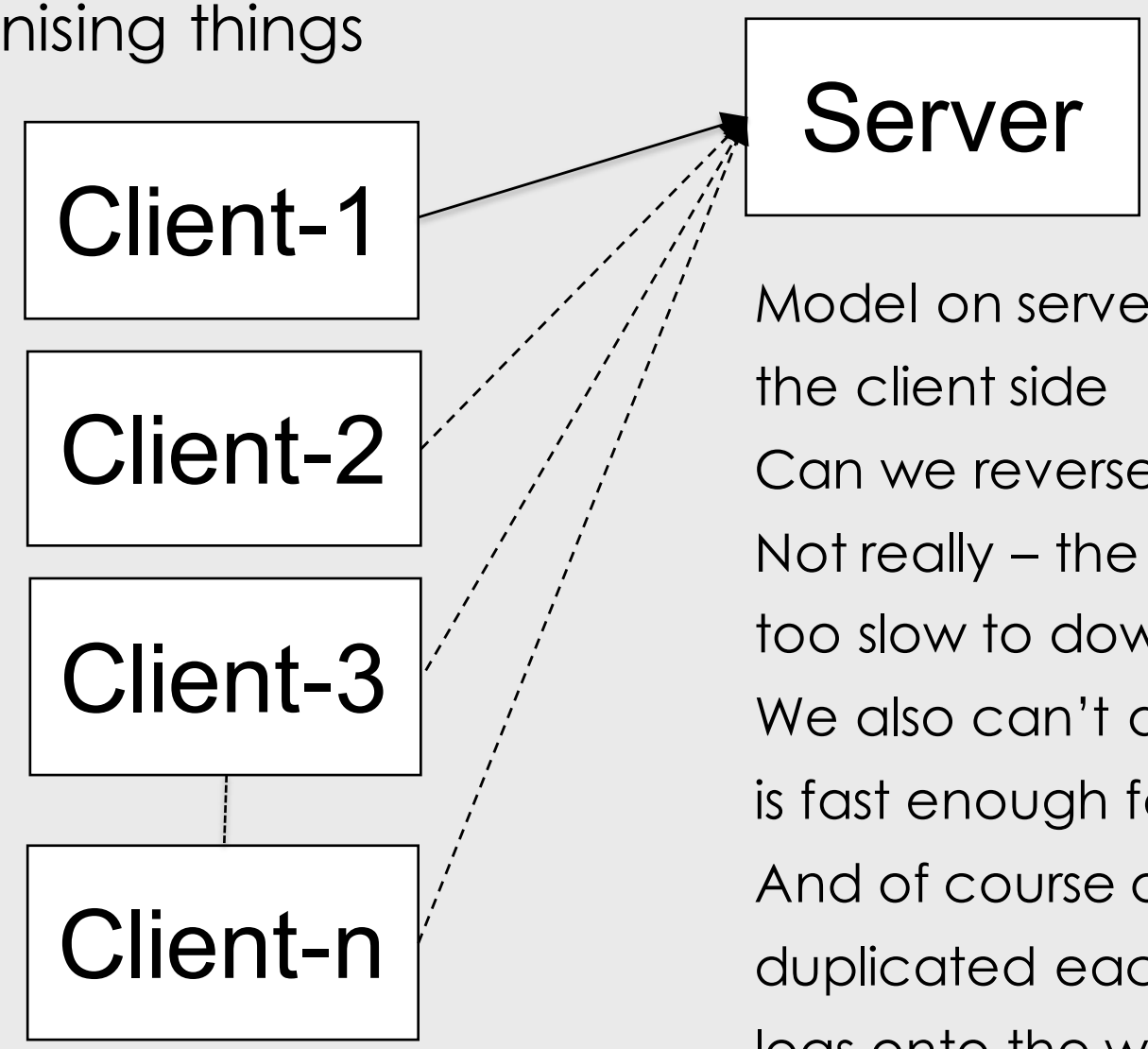
Big Data, High Frequency and Low Frequency Cities



(a) MSOA (A=7201) (b) LSOA (A=34753) (c) OA (A=181408)

Figure 8.2: ONS Geographies for MSOA, LSOA and OA levels.

Here is a block diagram of how we are currently organising things



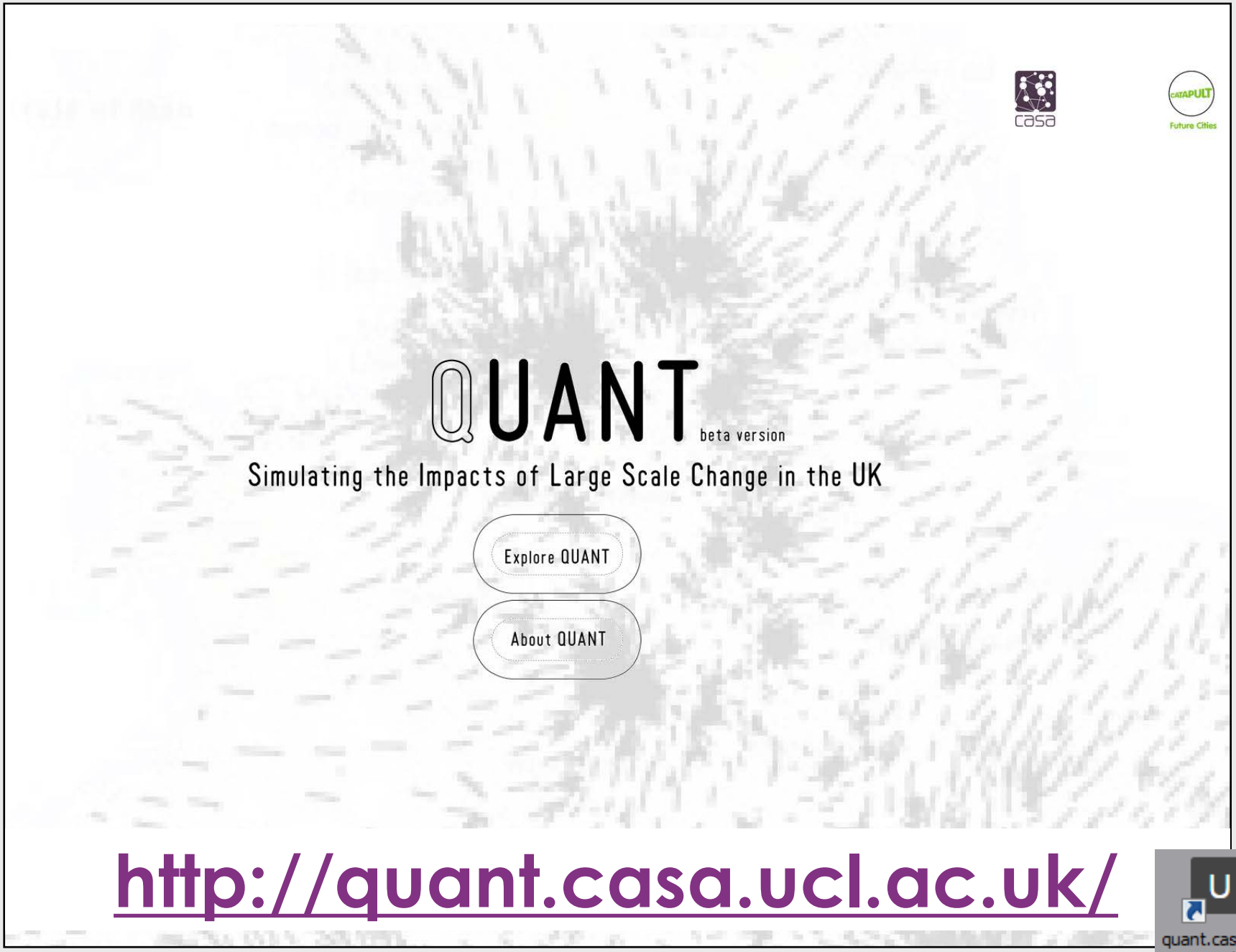
Model on server side; Maps on the client side

Can we reverse this?

Not really – the matrices are too slow to download to client?

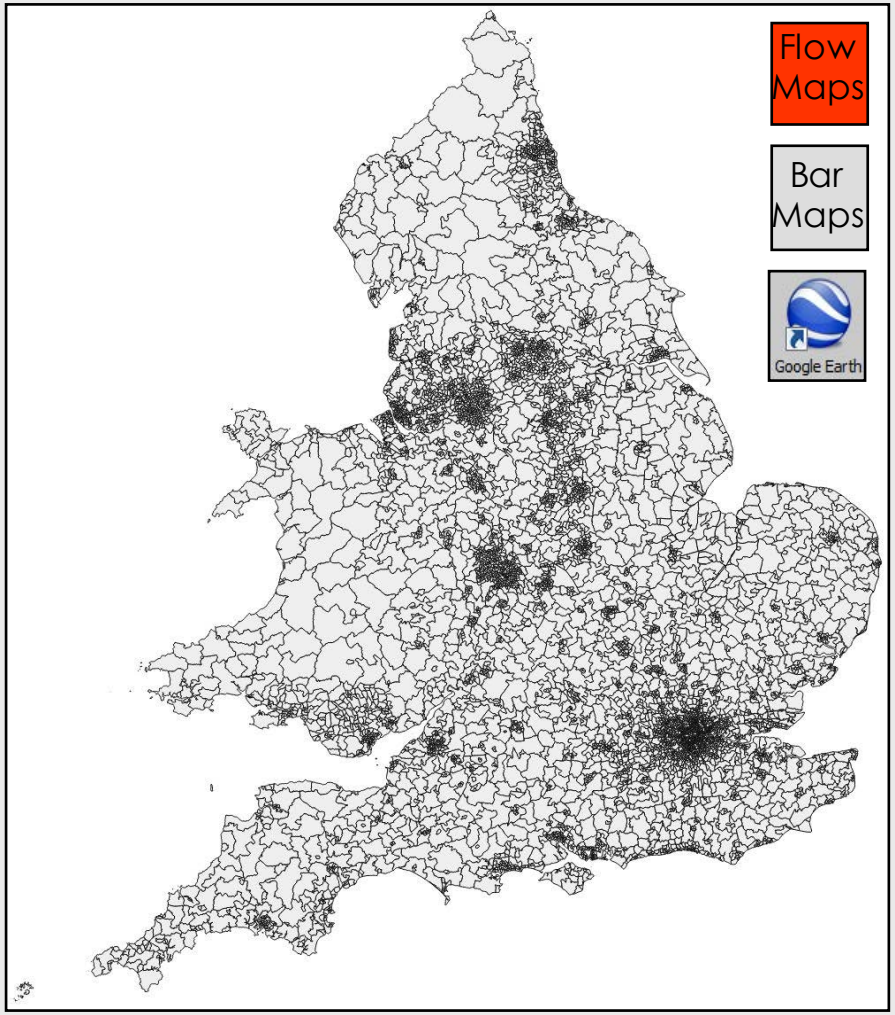
We also can't assume the client is fast enough for computation.

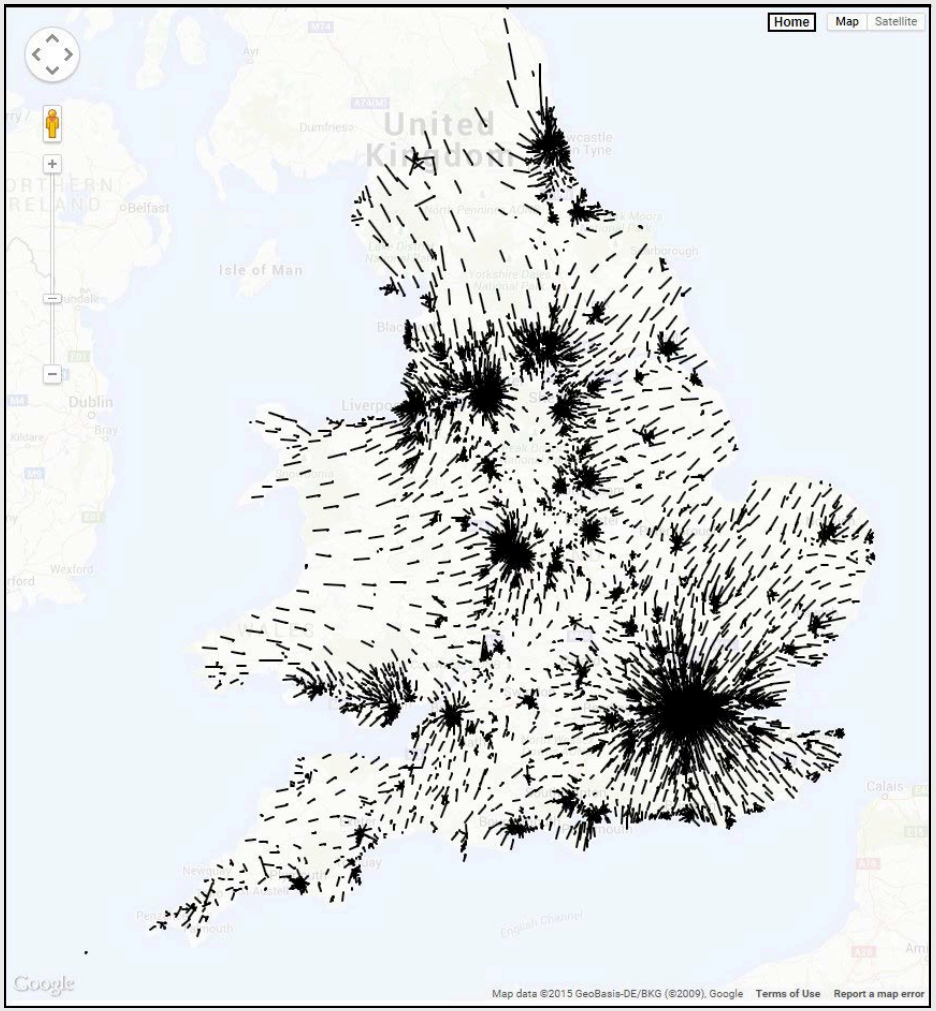
And of course our data is being duplicated each time a user logs onto the web interface



http://quant.casa.ucl.ac.uk/

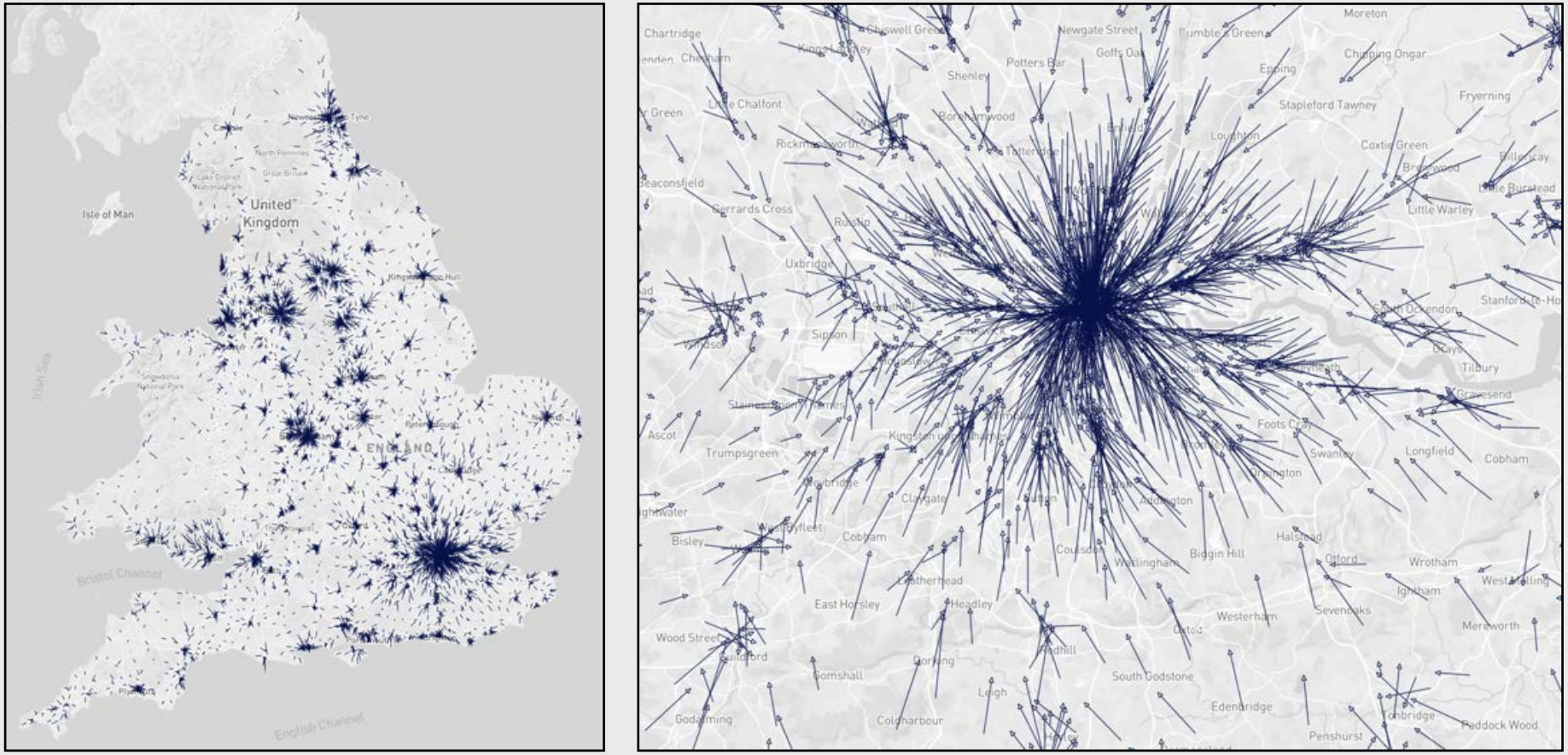
quant.casa.ucl.ac.uk





$$[x_i, y_i] = [[x_i, y_i], \left[\left[x_i + \frac{\sum_j T_{ij}\,[x_i - x_j]}{n}\right], \left[y_i + \frac{\sum_j T_{ij}\,[y_i - yy_j]}{n}\right]\right]$$

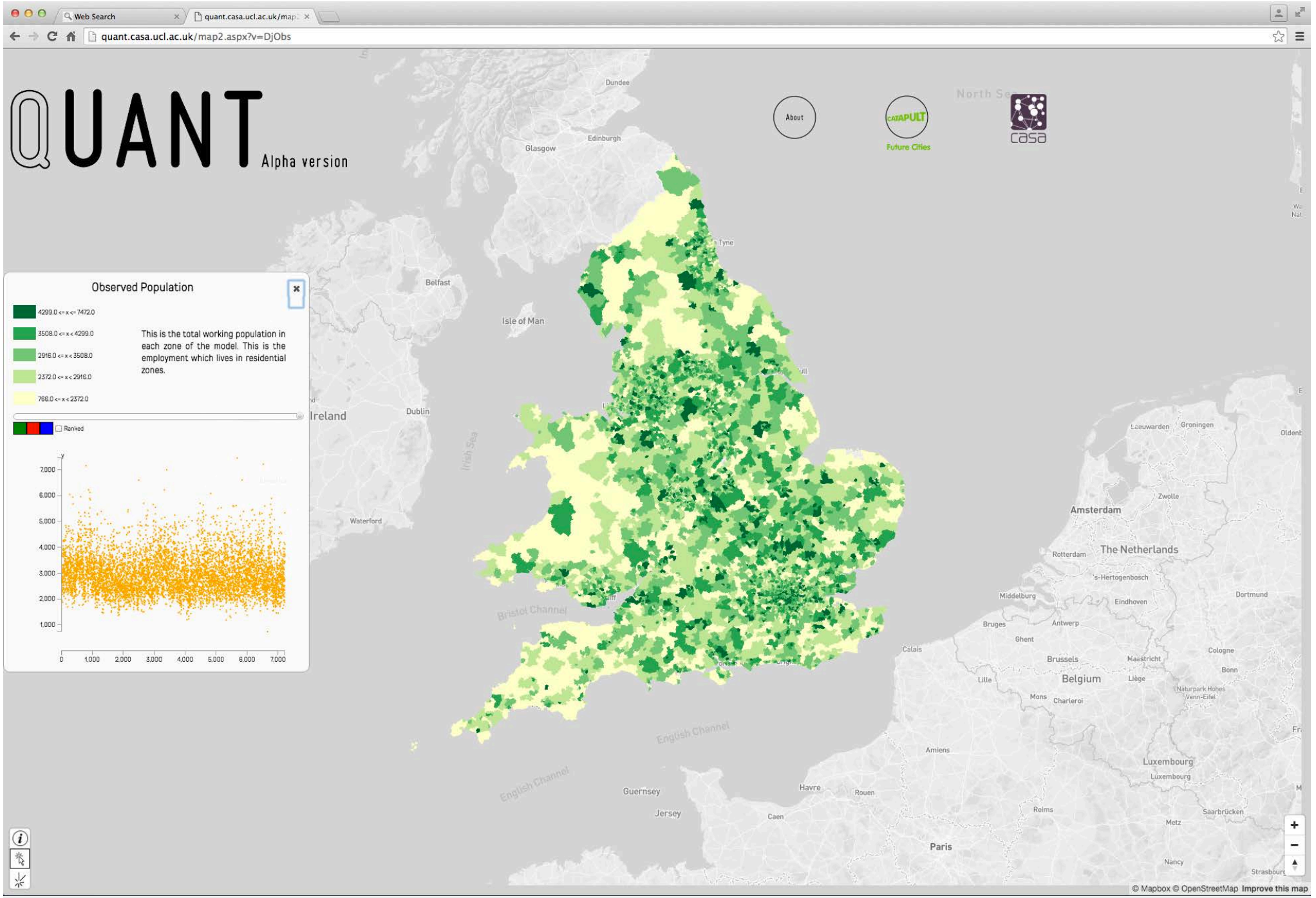
Big Data, High Frequency and Low Frequency Cities



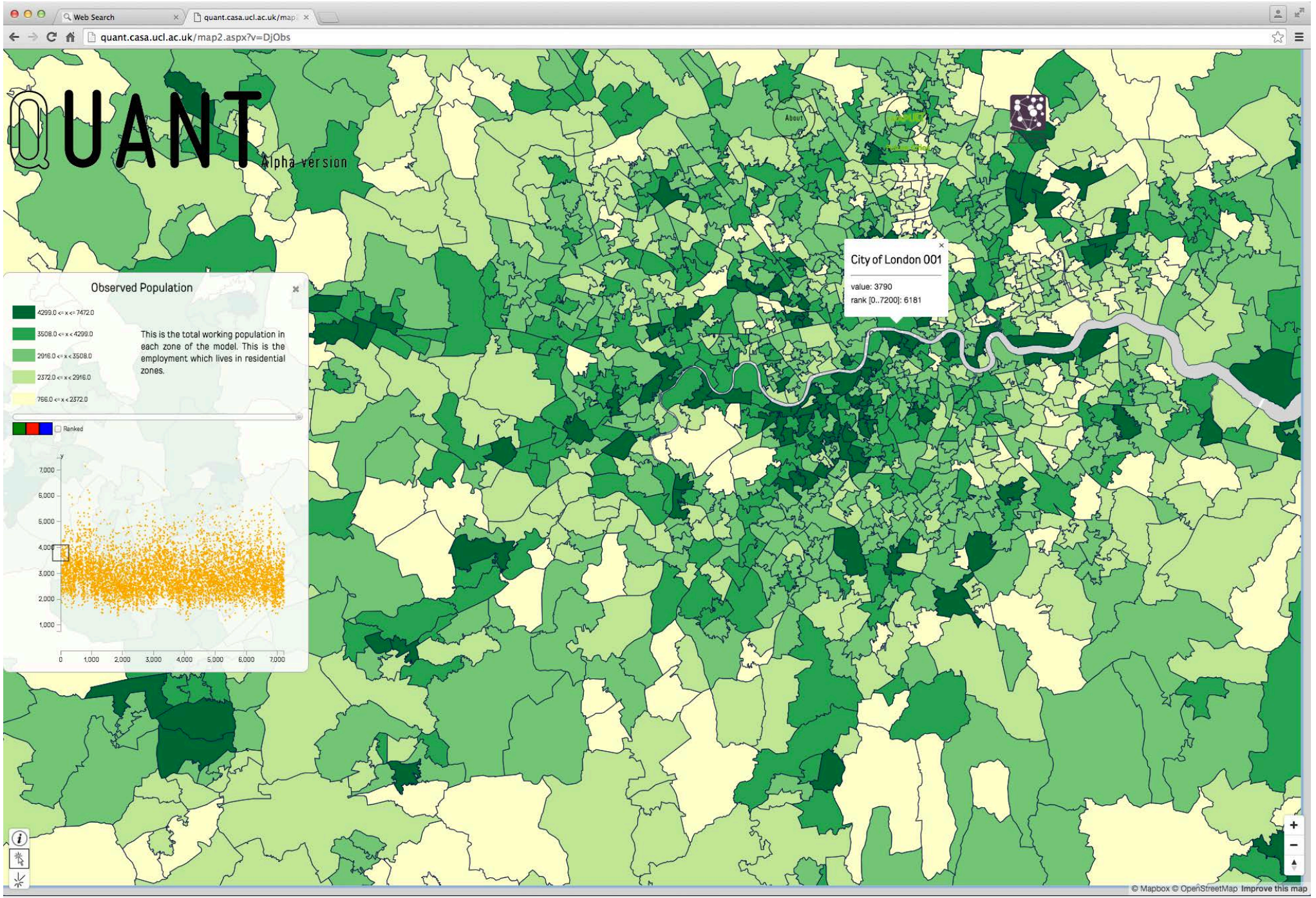
Big Data, High Frequency and Low Frequency Cities



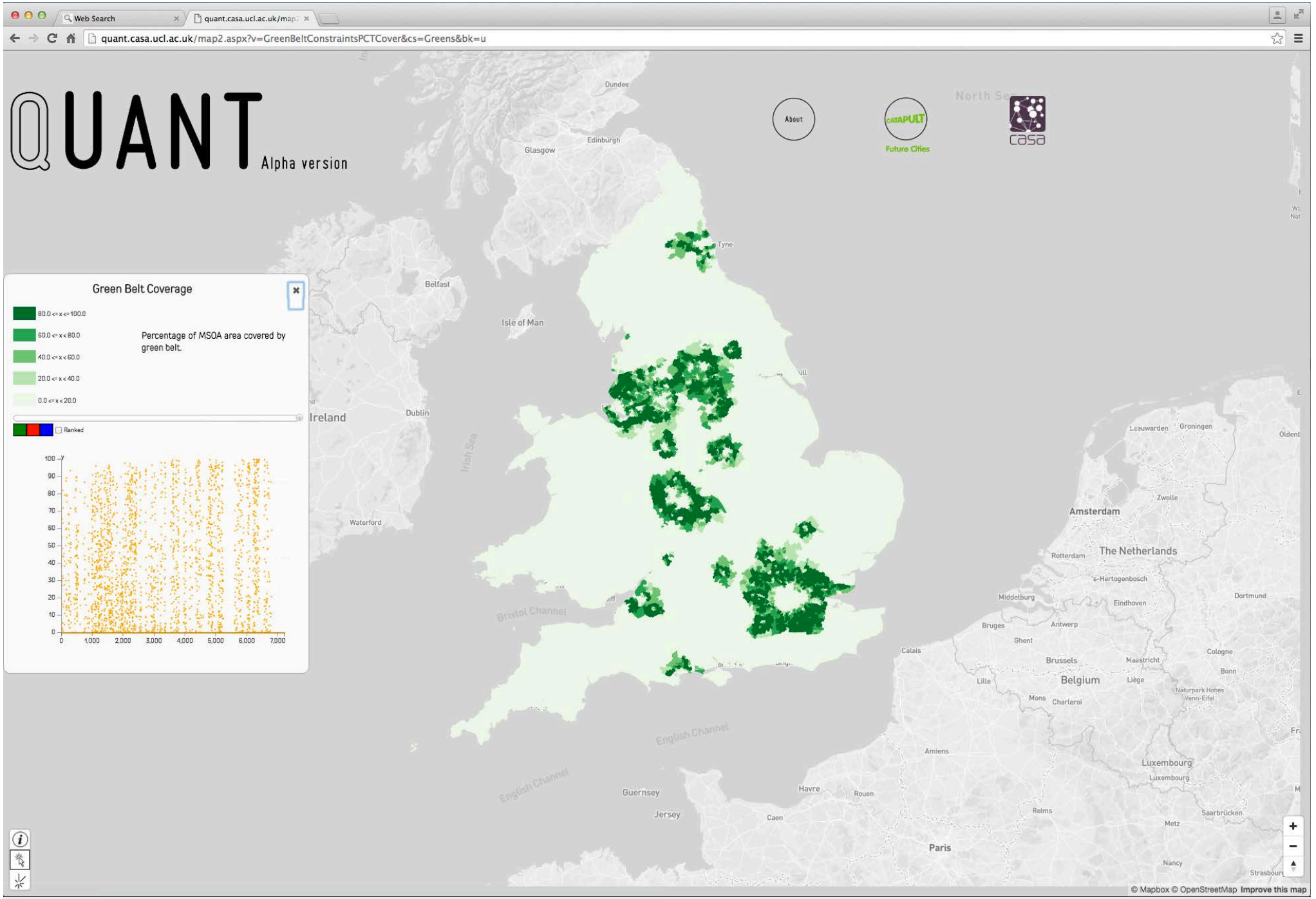
Big Data, High Frequency and Low Frequency Cities



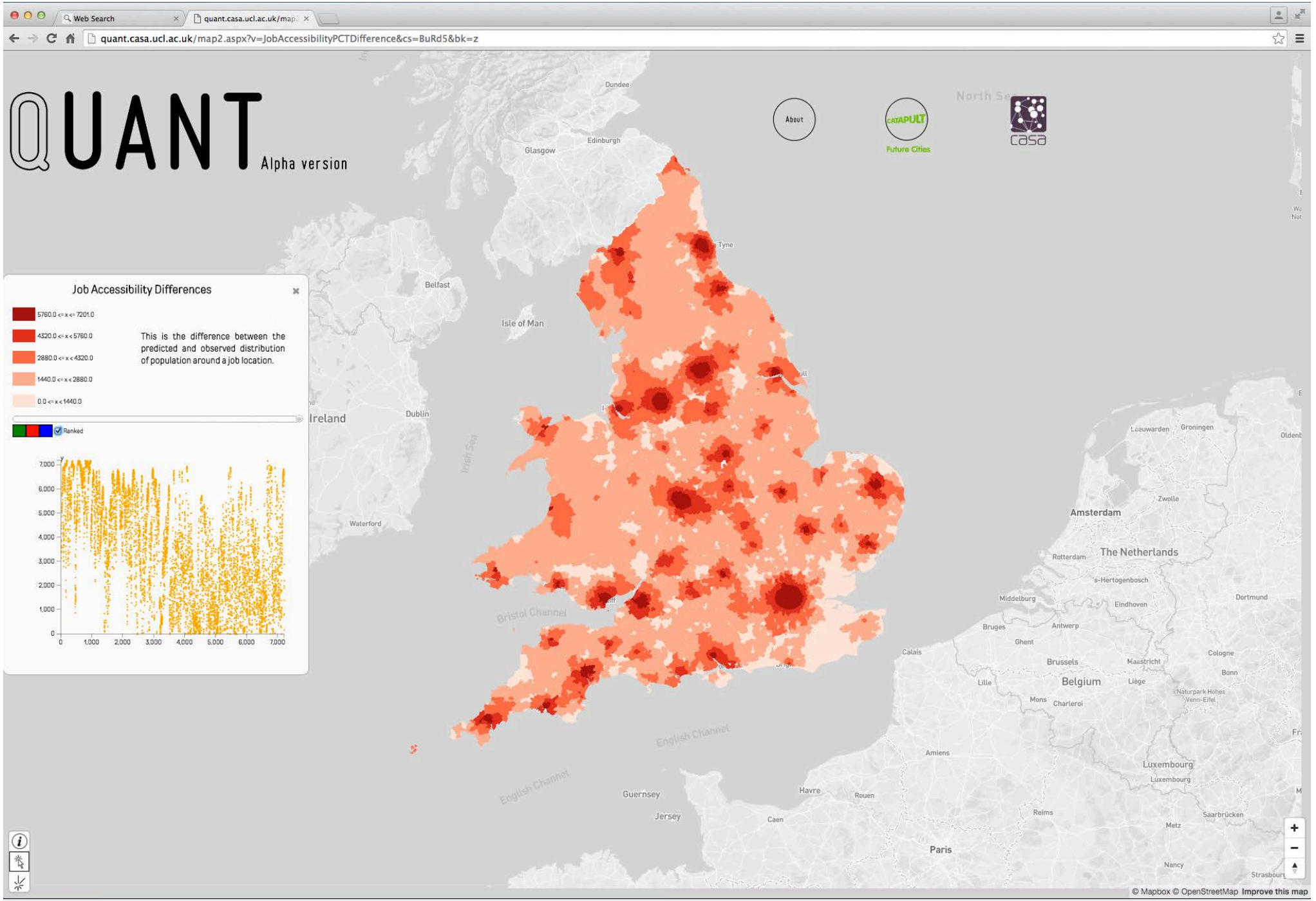
Big Data, High Frequency and Low Frequency Cities



Big Data, High Frequency and Low Frequency Cities



Big Data, High Frequency and Low Frequency Cities

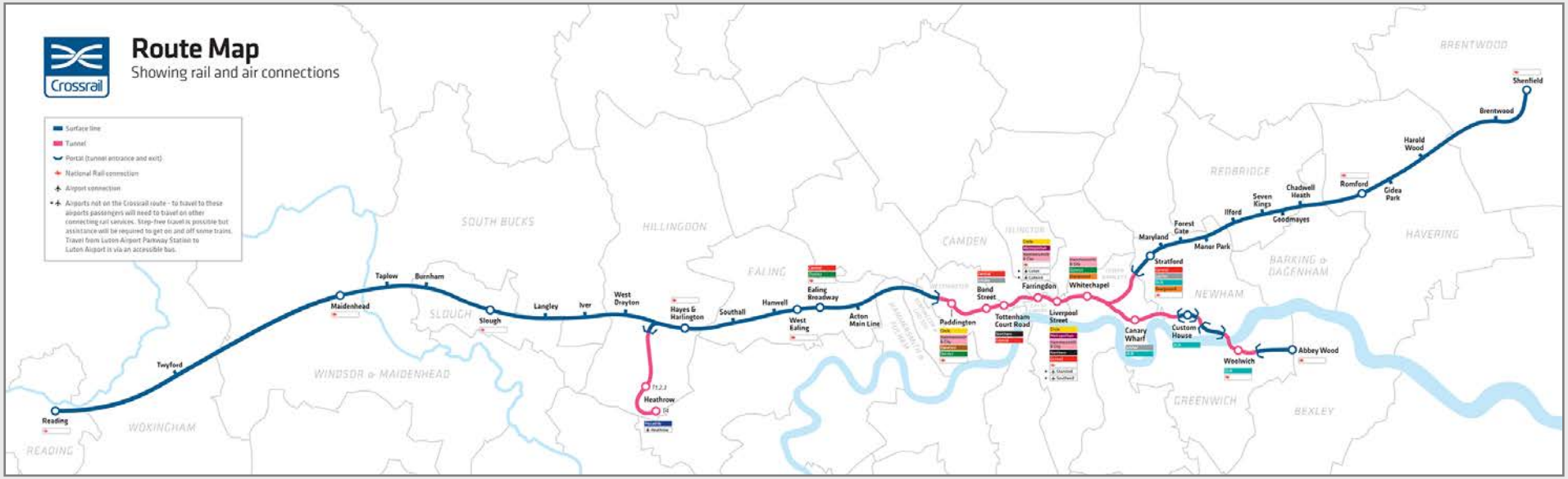


Big Data, High Frequency and Low Frequency Cities

Crossrail

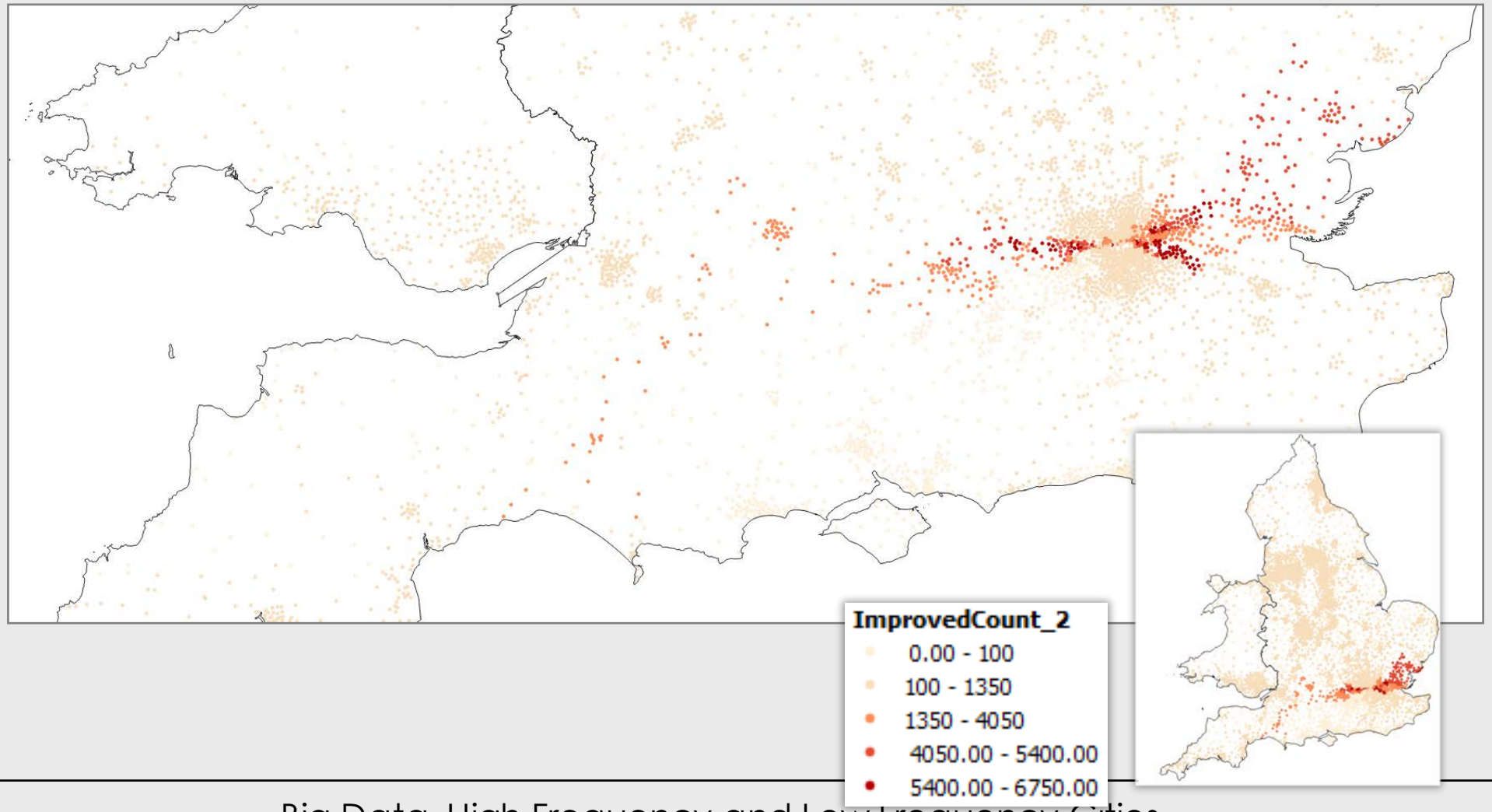
Reading, Heathrow, Shenfield, Abbey Wood





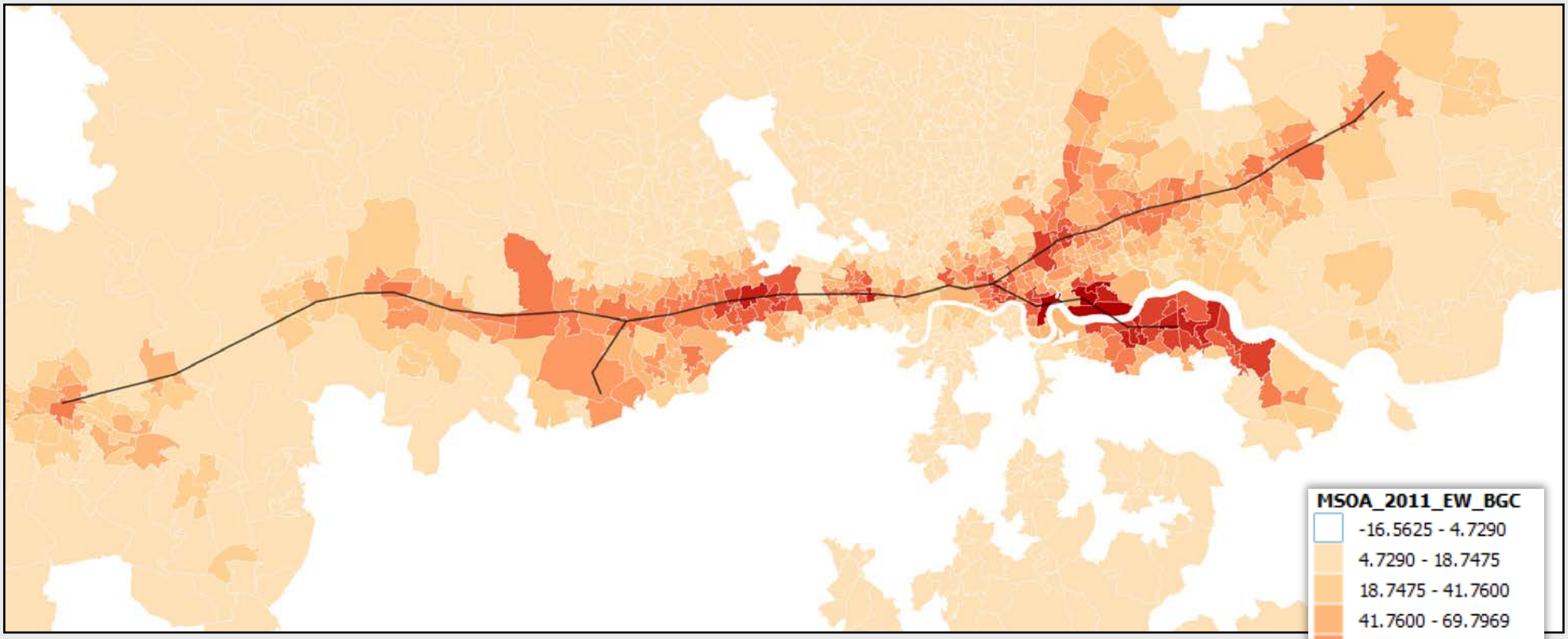
Crossrail

Number of Improved Journeys (n_i)



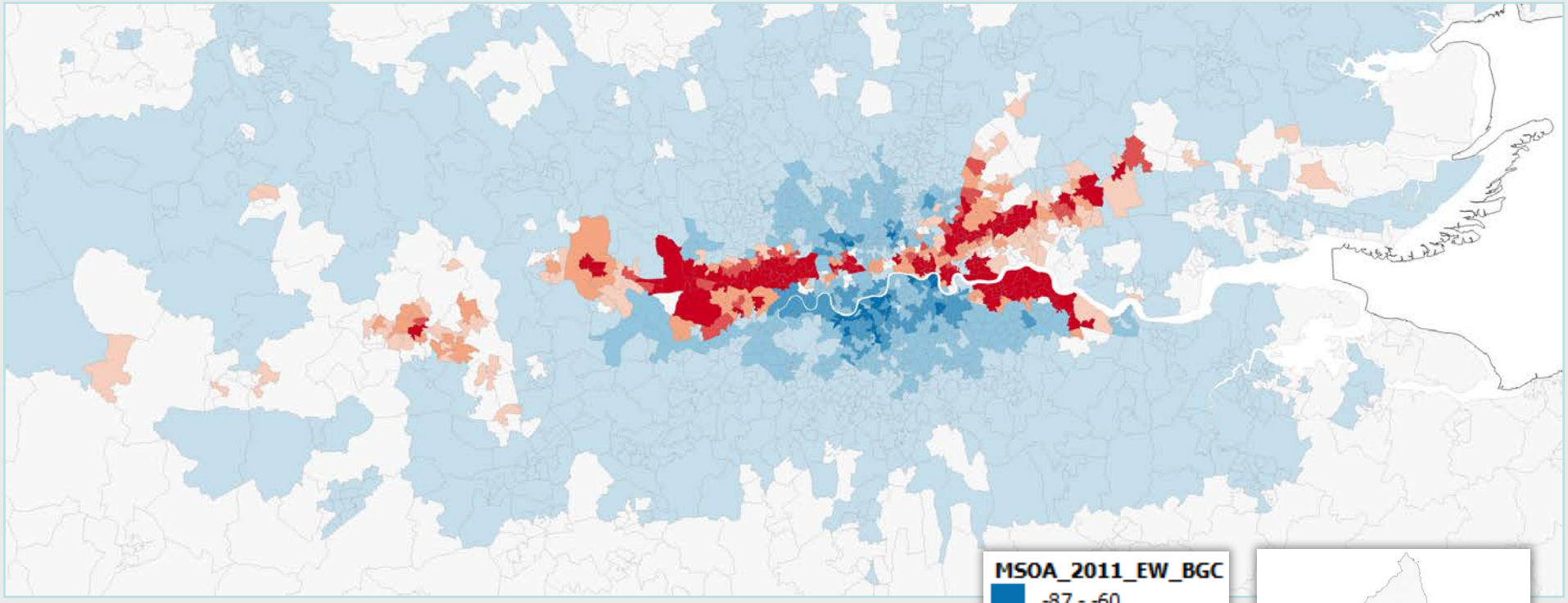
Crossrail

Population change (rail mode only)



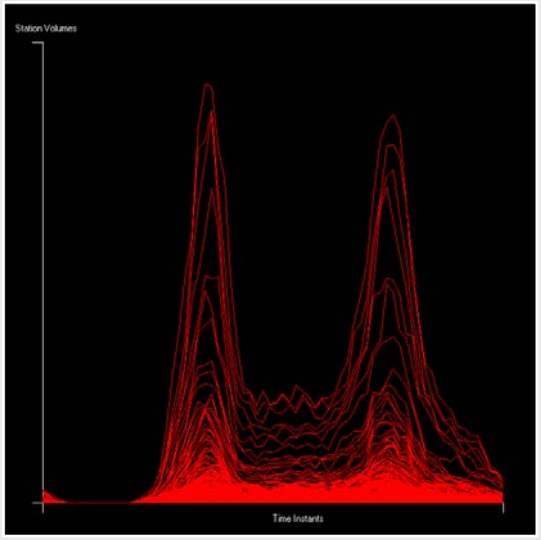
Crossrail

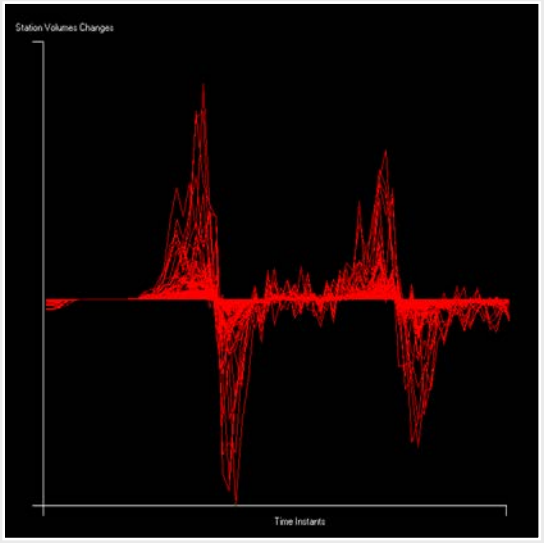
Population Change (all modes)

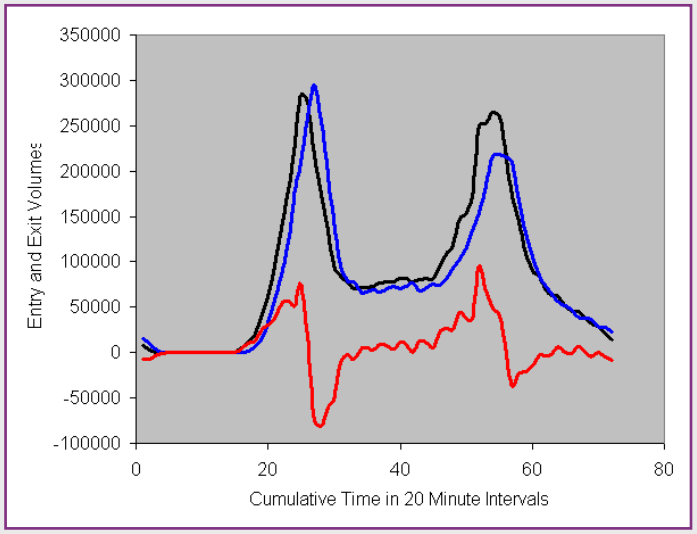


Mobility, Transit, & Real-Time Streaming: Examples of Urban Dynamics in the Very Short Term

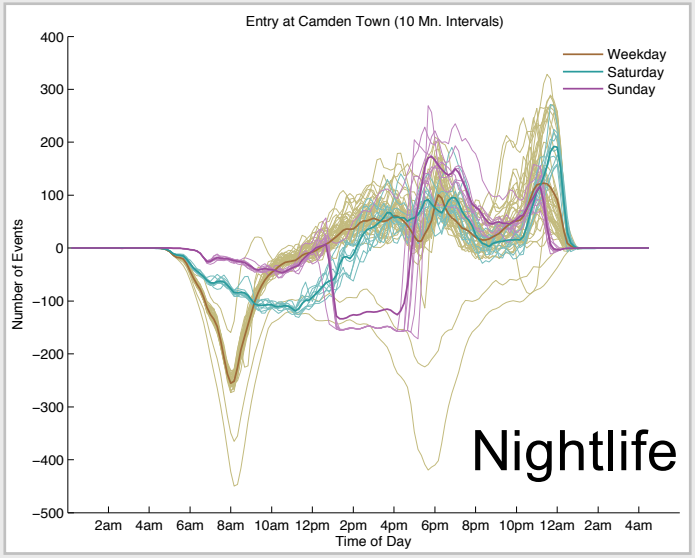
- Let me show you some dynamics – just pictures – first from our tube data in London captured by smart cards (RFID, Oyster, Apple Pay and so on)

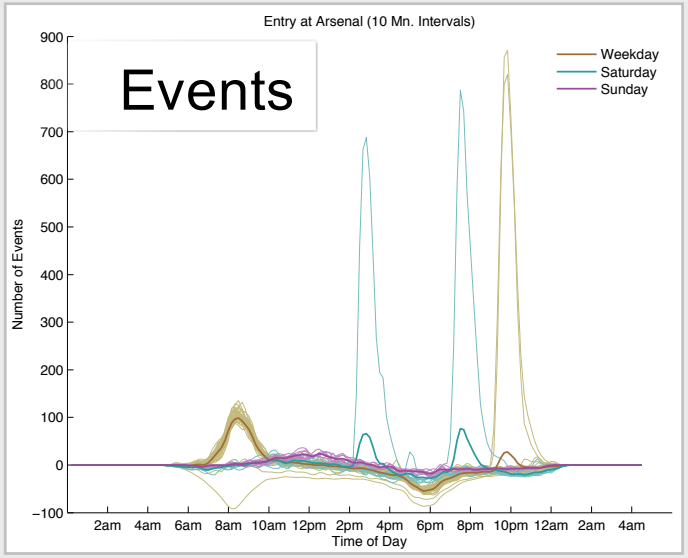


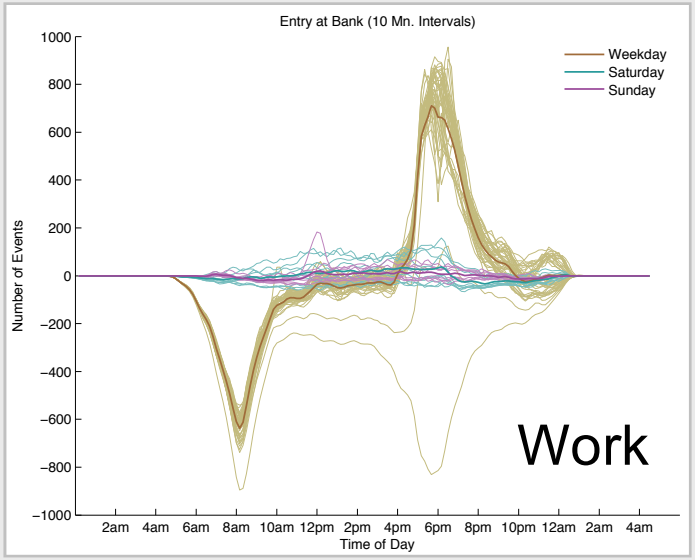


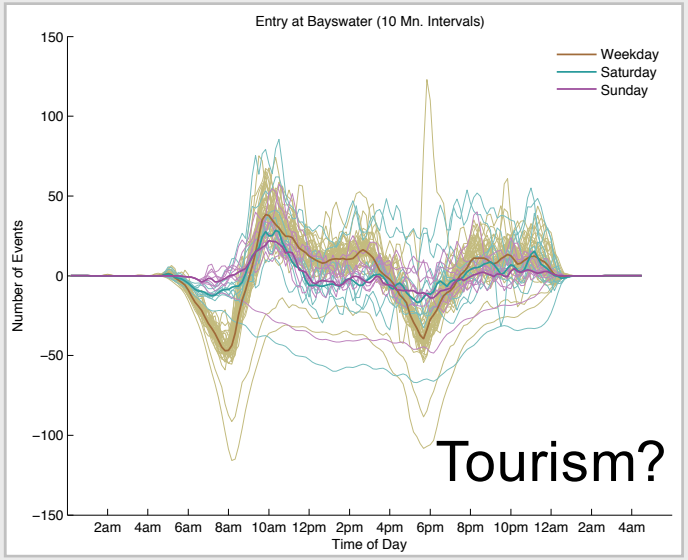


Particular Events: Weekdays, Saturdays and Sundays



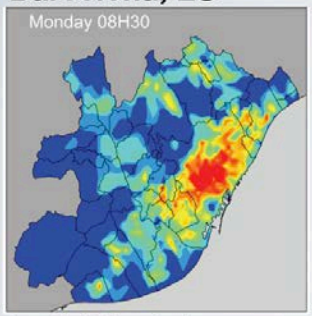






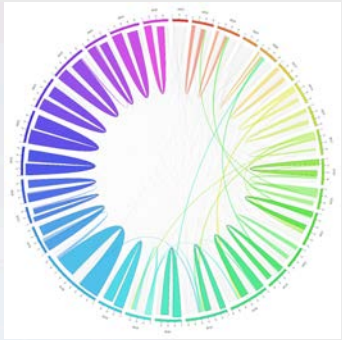
Here are some other data – journey to work, credit card flows and social media data – tweets – which are illustrative of many different rhythms

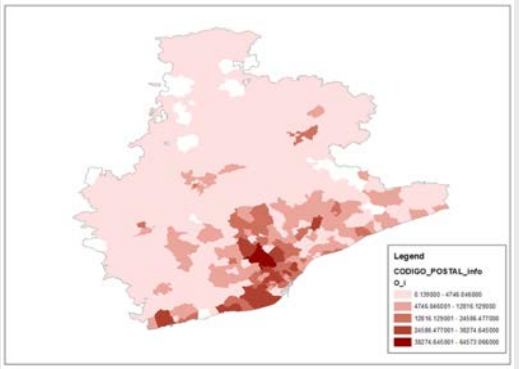
Barcelona, ES



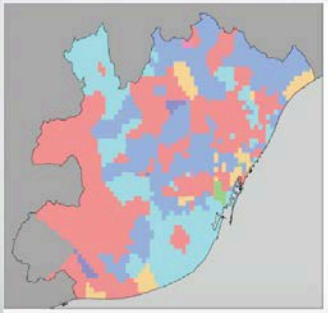
< Maps at different hours
Hotspots location
(Inter)national comparison of cities

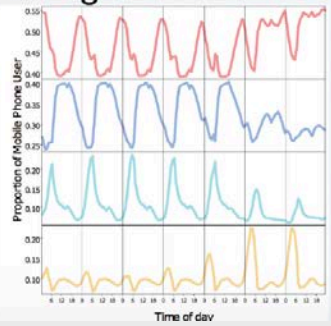
Louail et al. (2014)
Sci.rep.4:5276





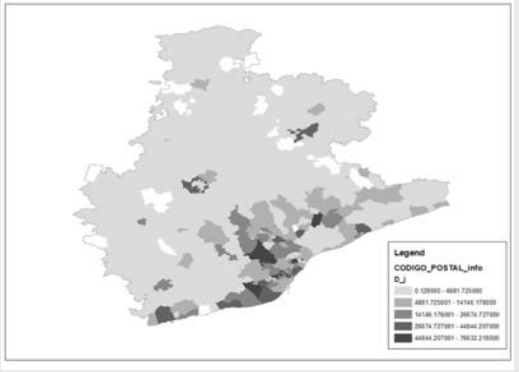
Individuals' most visited places
OD matrices / commuting distances >

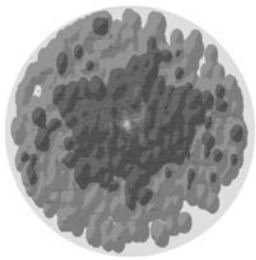


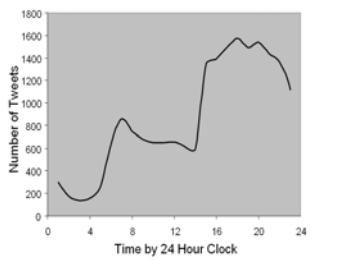


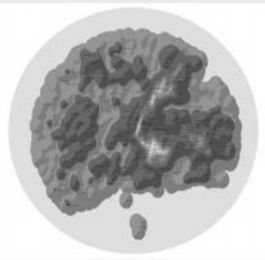
< Inferring land-use
from activity rhythms

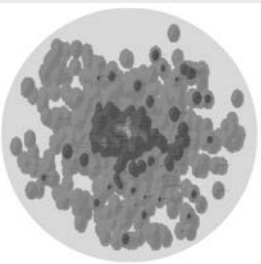
Lenormand et al. (2015)
R. Soc.Open.Sci. 2: 150449

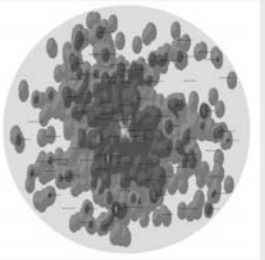




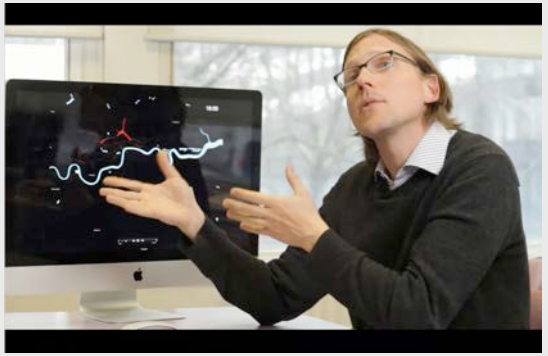






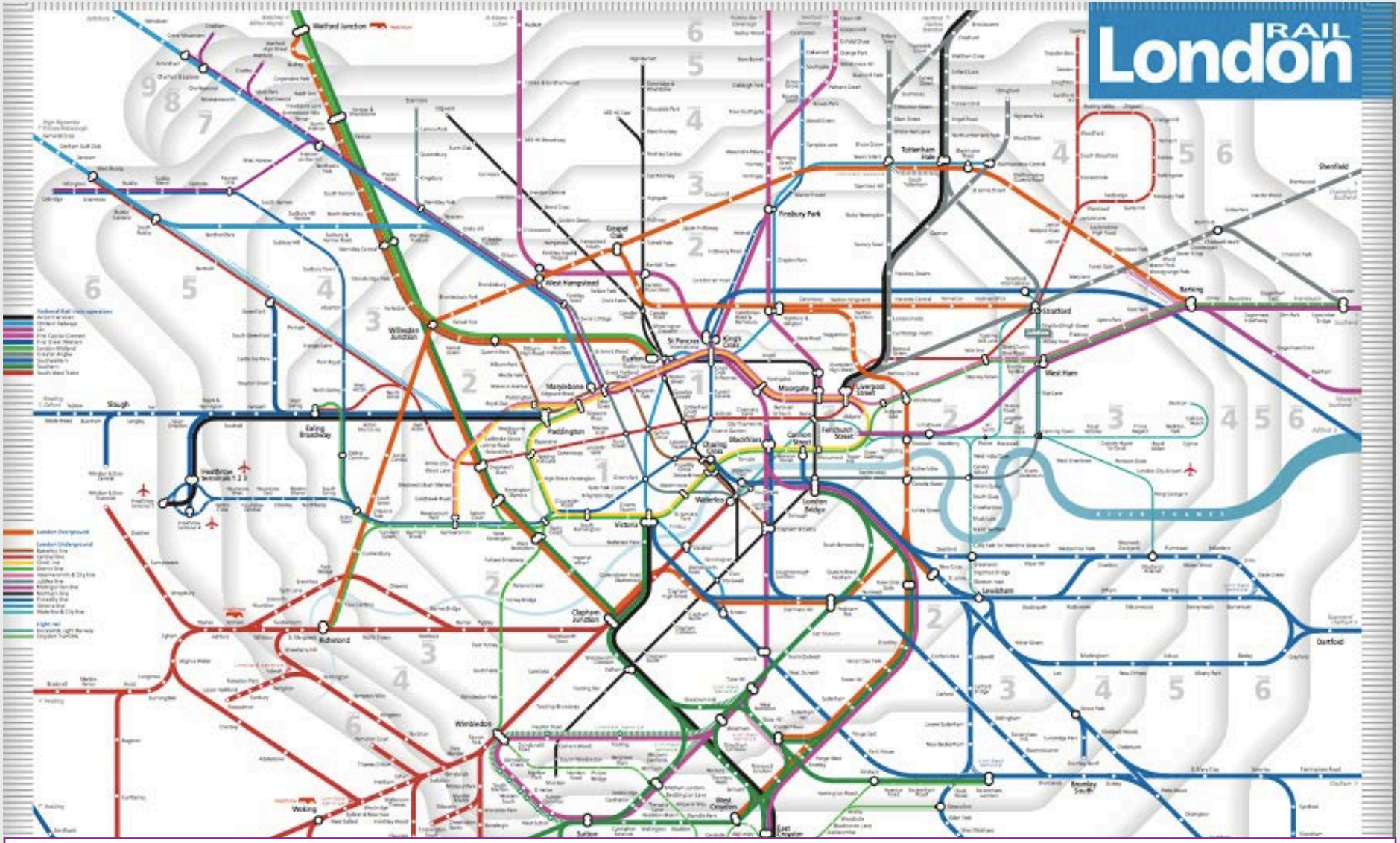




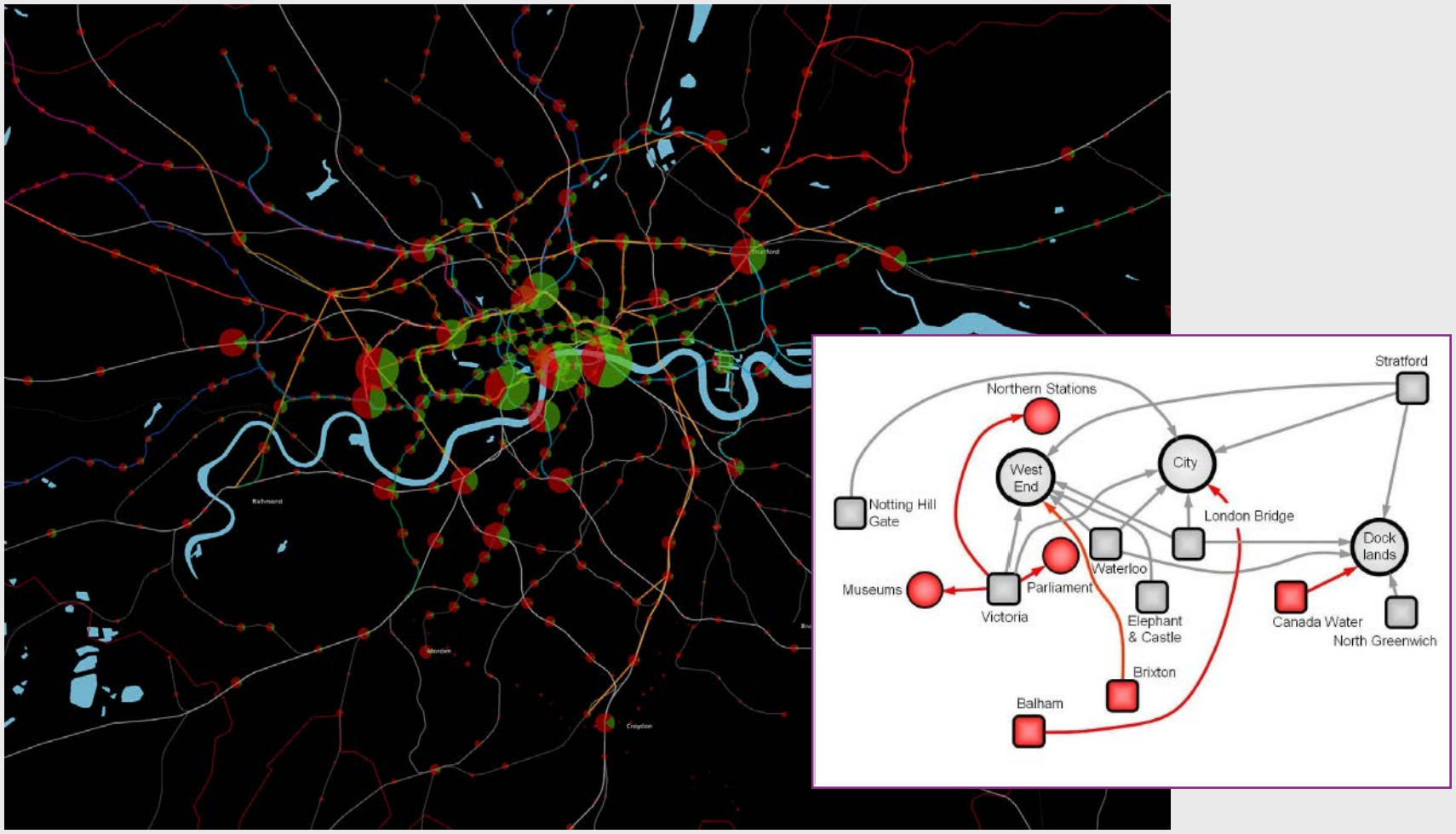



This of course was the thing that Lt Henry Harness did in Dublin in 1837 and what Minard et al. did a little later. In our LUTI models, this is an enormous problem as the scale of this assignment to networks is different



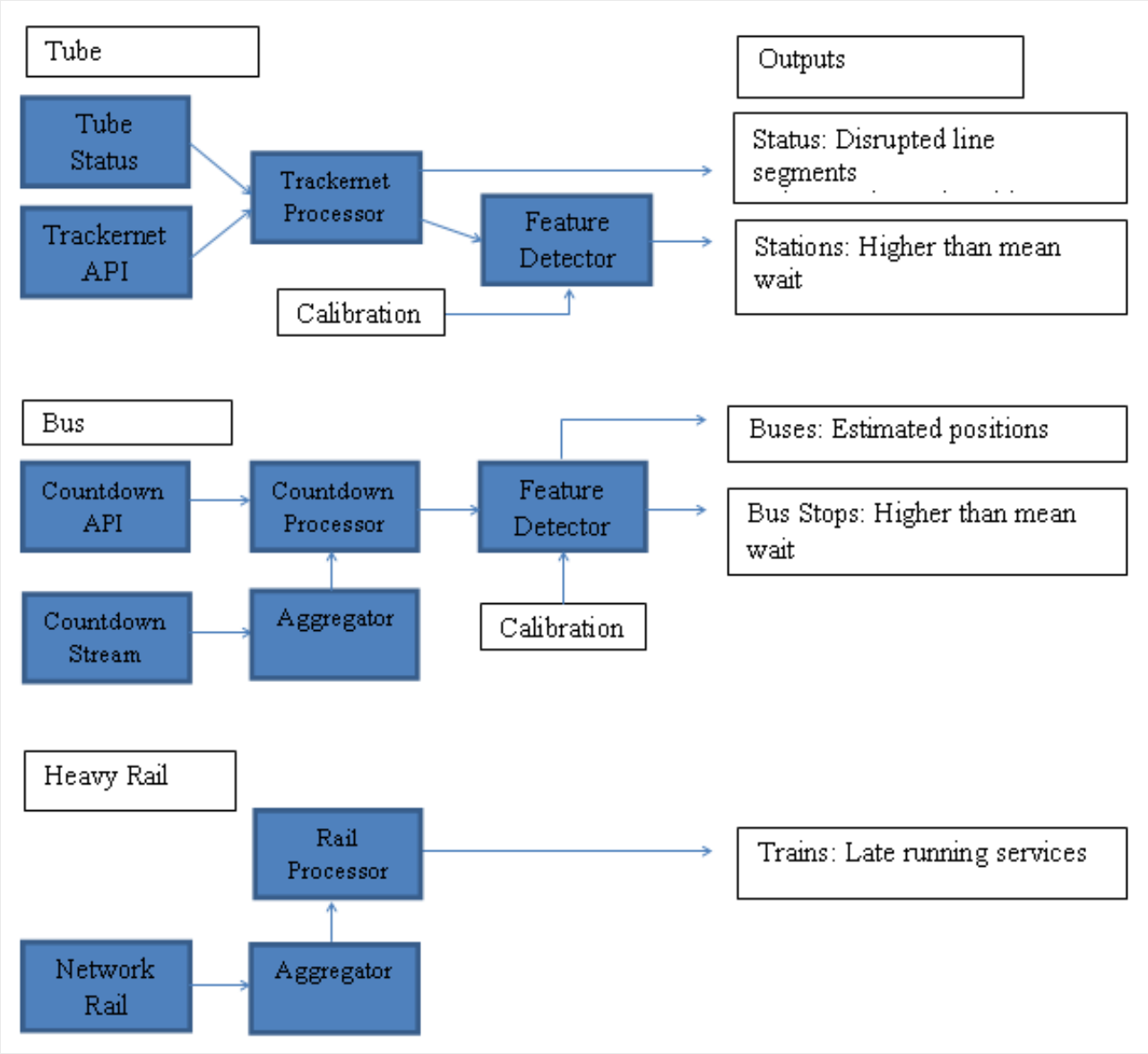
Tube, Overground and National Rail Networks in London where Oyster cards can be used

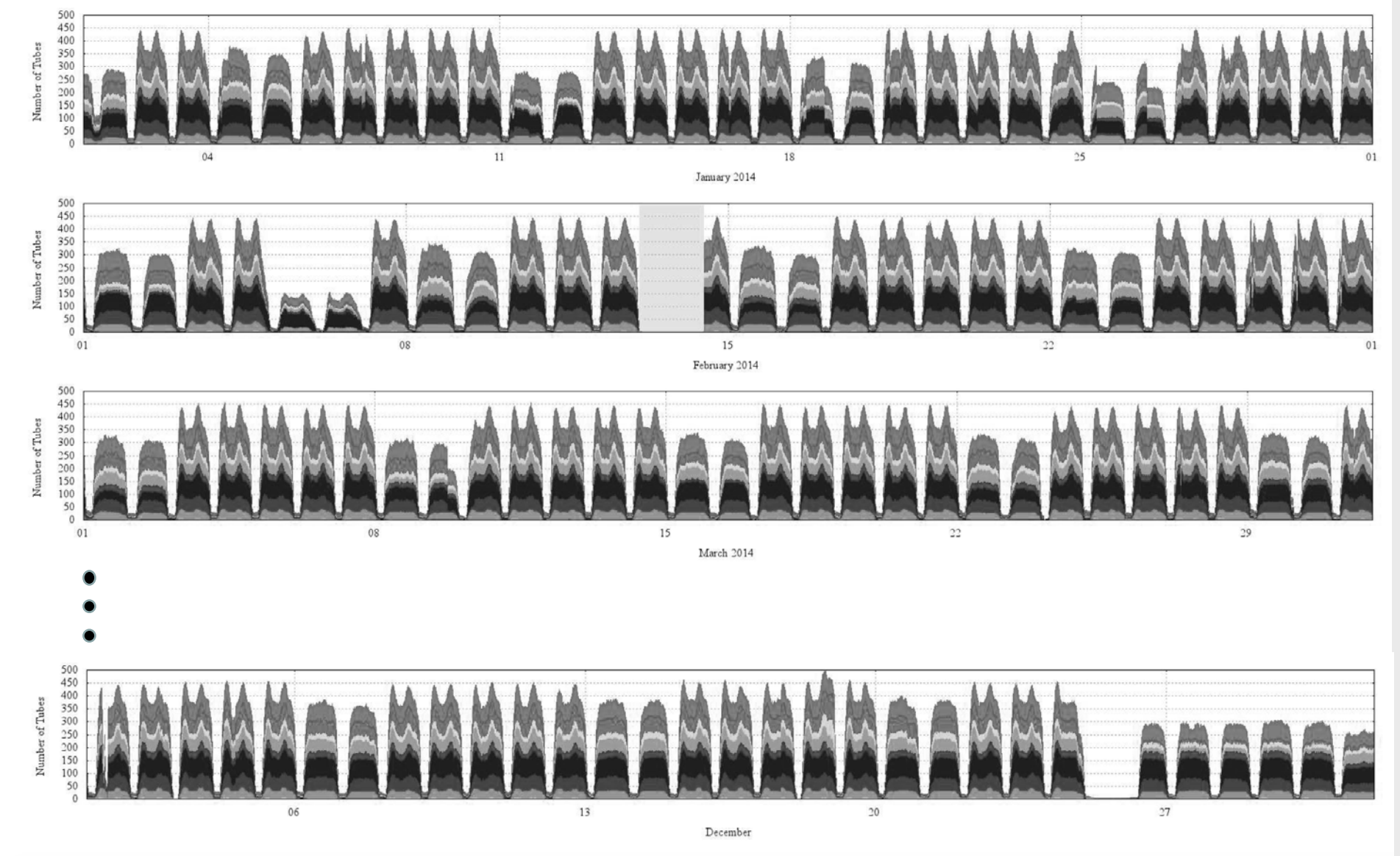
Oyster Card Data – interpreting urban structure, multitrips, etc.



Roth C., Kang S. M., Batty, M., and Barthelemy, M. (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. **PLoS ONE 6(1):** e15923. doi:10.1371/journal.pone.0015923

The Public Transport System in Terms of Vehicle Flows




Number of Tubes
500
450
400
350
300
250
200
150
100
50
0
04
11
18
25
01
January 2014
Number of Tubes
01
08
15
22
01
February 2014
Number of Tubes
01
08
15
22
29
March 2014
Number of Tubes
06
13
20
27
December

Learning about Mobility from the Data

Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

Table 1. Summary statistics of one-week of smart-card data (metro trips only)

	London	Singapore	Beijing
Monday	3,457,234	2,208,173	4,577,500
Tuesday	3,621,983	2,250,597	4,421,737
Wednesday	3,677,807	2,277,850	4,564,335
Thursday	3,667,126	2,276,408	4,582,144
Friday	3,762,336	2,409,600	4,880,267
Number of stations (1)	400	130	233
Number of tube line	13	4	17
Area (2)	1,572 km^2	718.3 km^2	2267 km^2
Total population (3)	8.63 million	5.3 million	21.15 million
Ridership of Metro	20%	35%	21%
Length of metro lines	402km	182km (MRT+LRT)	465 km

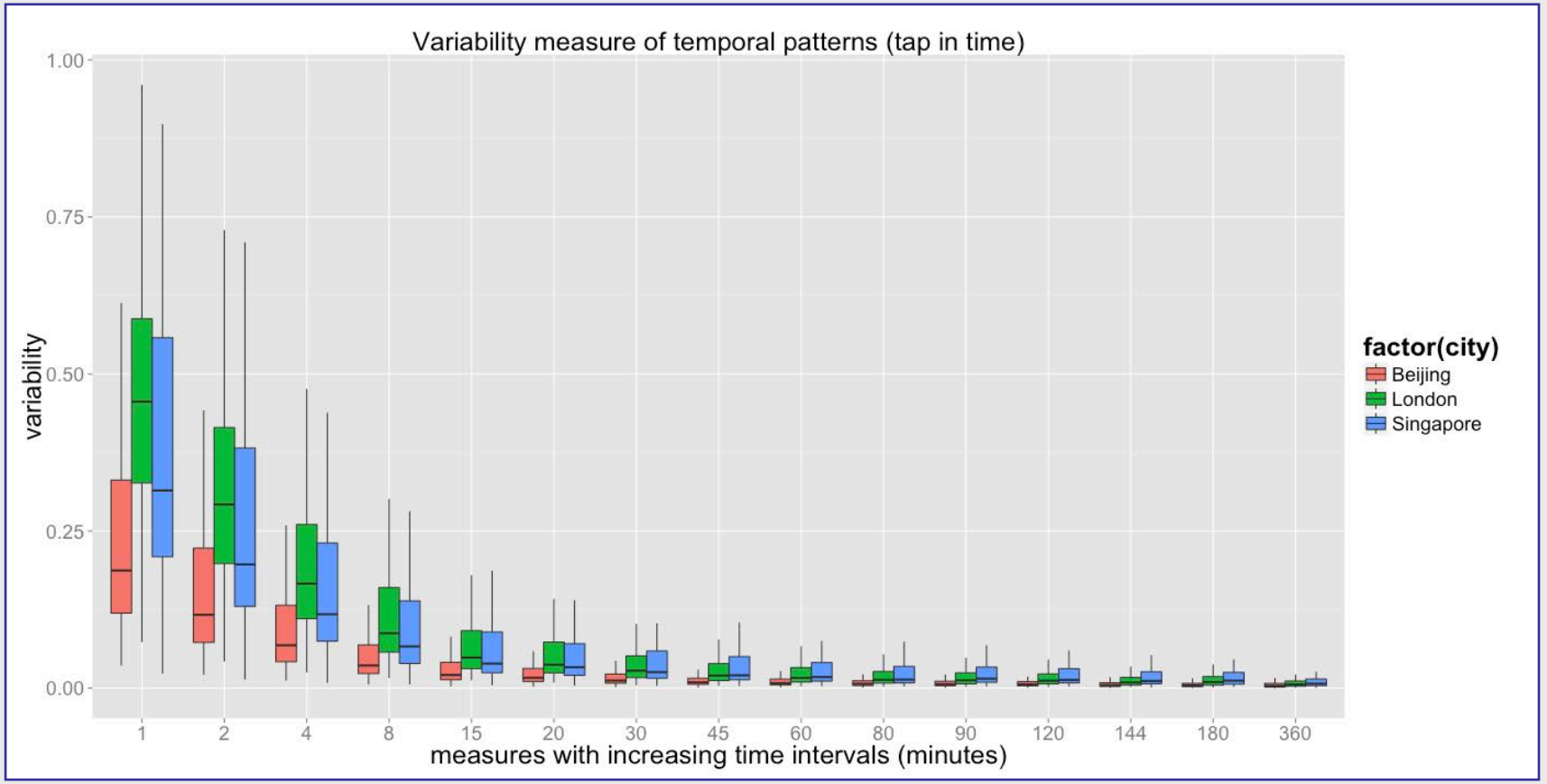
(1) Number of stations is the number of stations with smart-card records generated.

(2) The area of Beijing only counts the area enclosed by the 6th ring road for a fair comparison.

(3) From the World Population Review, http://worldpopulationreview.com/world-cities/ accessed 17 January 2016

Zhong, C., Batty, M., Manley, E., Wan, J., Wang, Z., Che, F., and Schmitt, G. (2016) Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing using Smart-Card Data., **PLOS One**, http://dx.doi.org/10.1371/journal.pone.0149222

From 1 minute intervals to the whole day



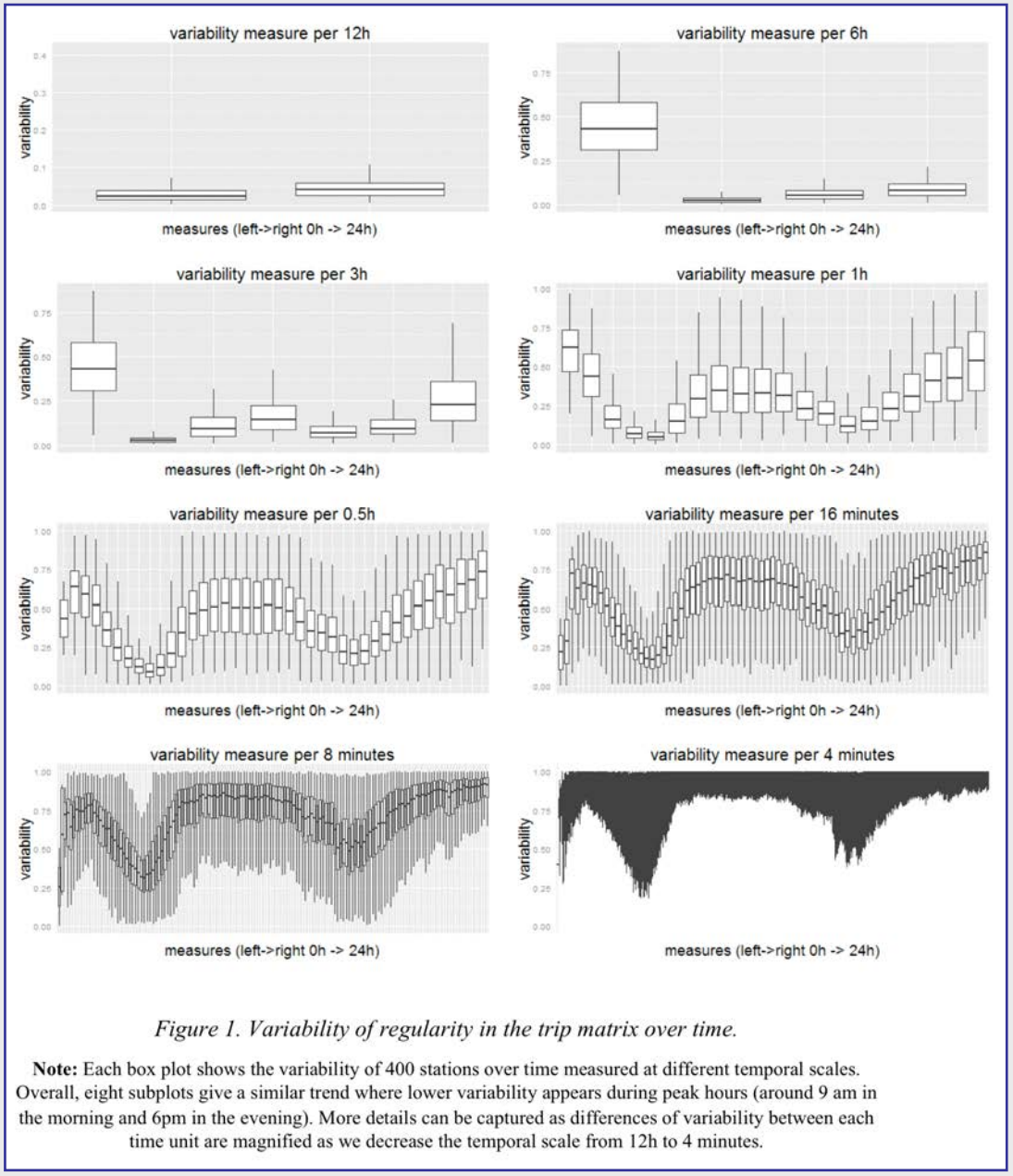
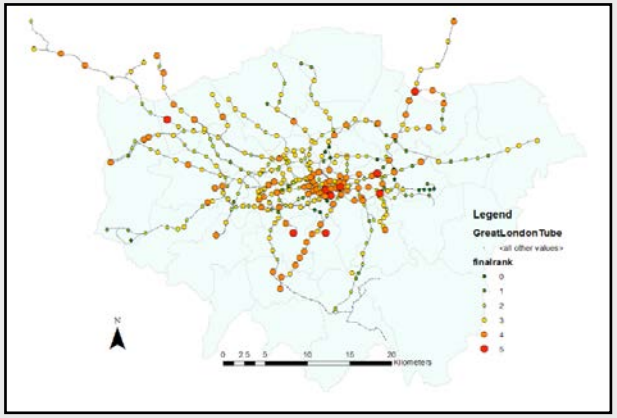


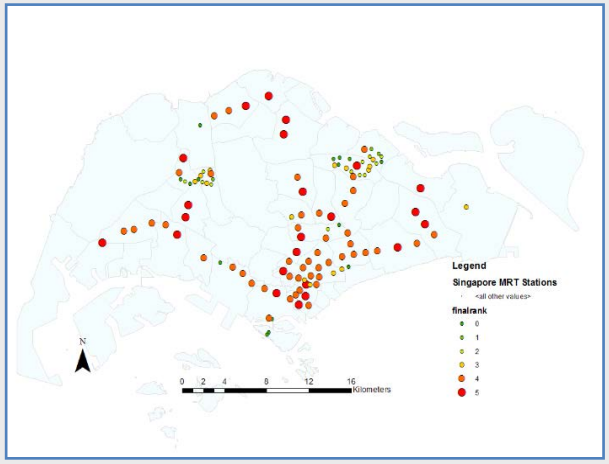
Figure 1. Variability of regularity in the trip matrix over time.

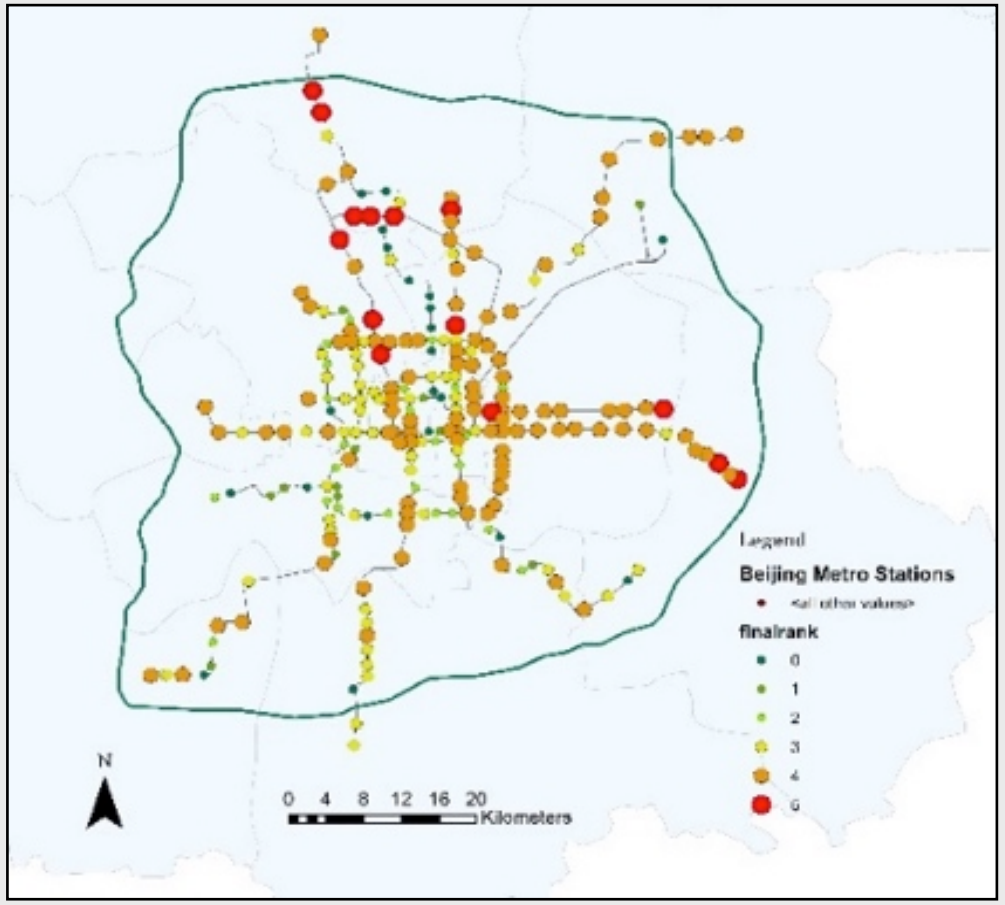
Note: Each box plot shows the variability of 400 stations over time measured at different temporal scales. Overall, eight subplots give a similar trend where lower variability appears during peak hours (around 9 am in the morning and 6pm in the evening). More details can be captured as differences of variability between each time unit are magnified as we decrease the temporal scale from 12h to 4 minutes.

Comparing Variability for different time Intervals over the day

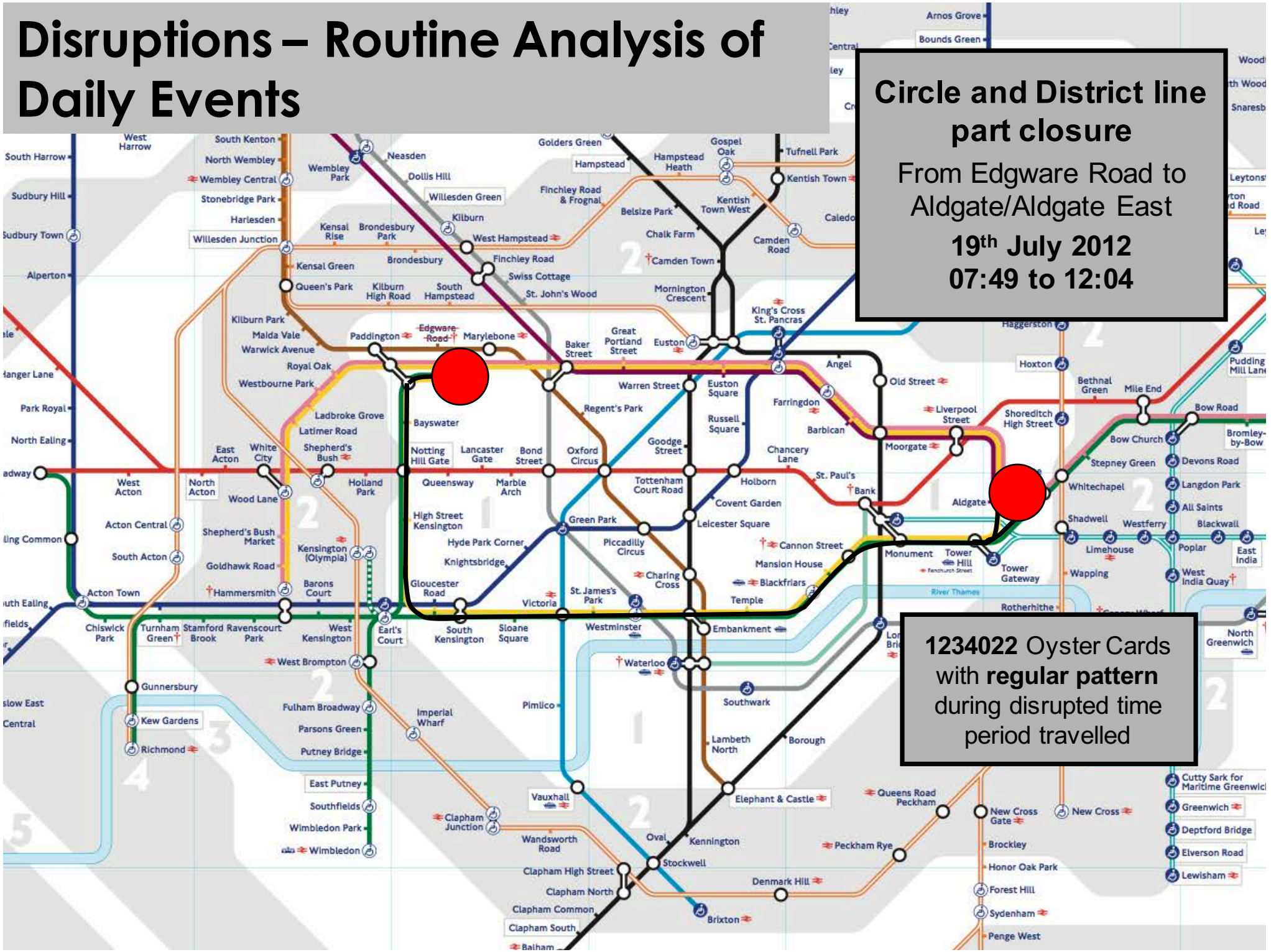
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

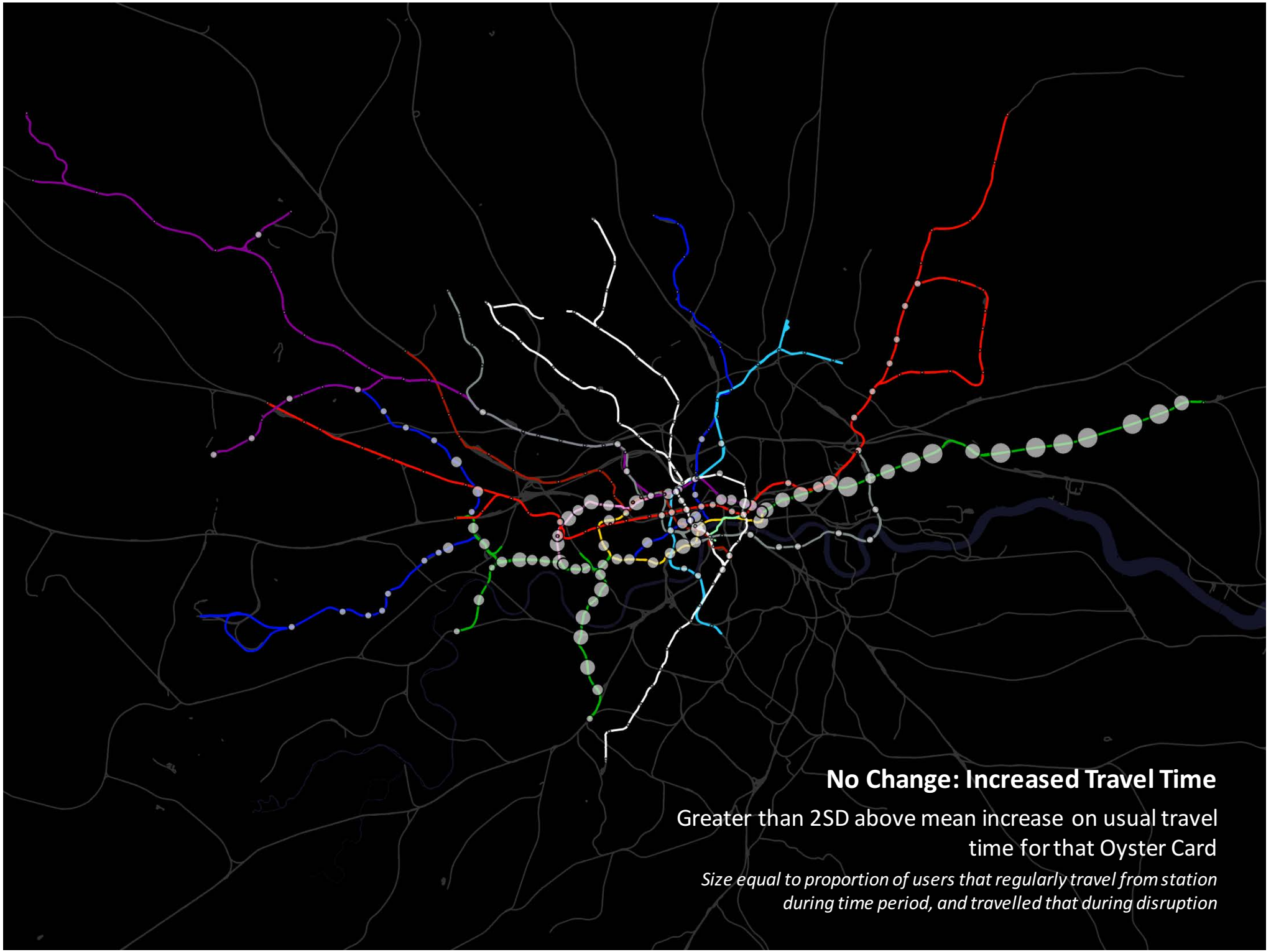


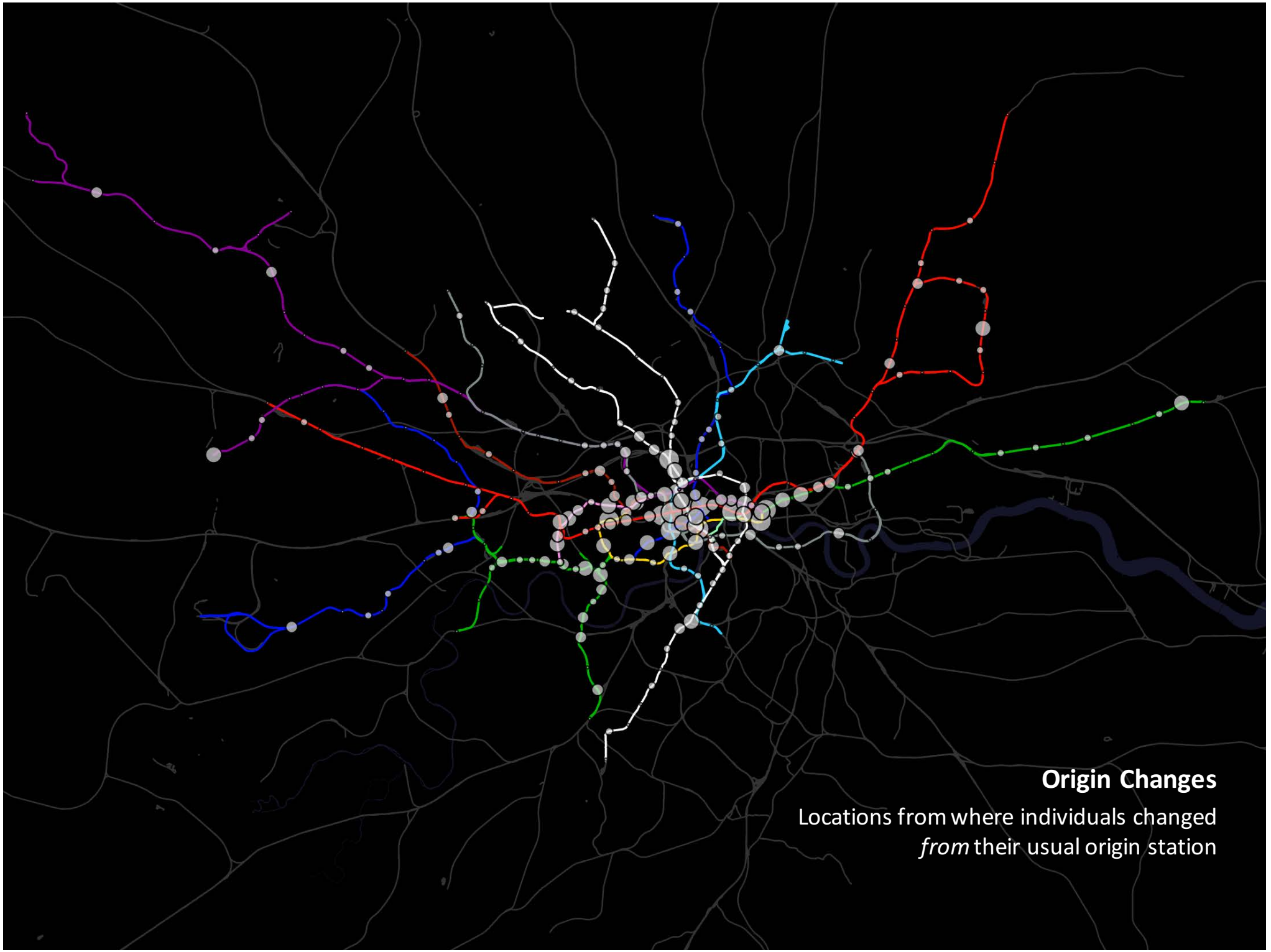


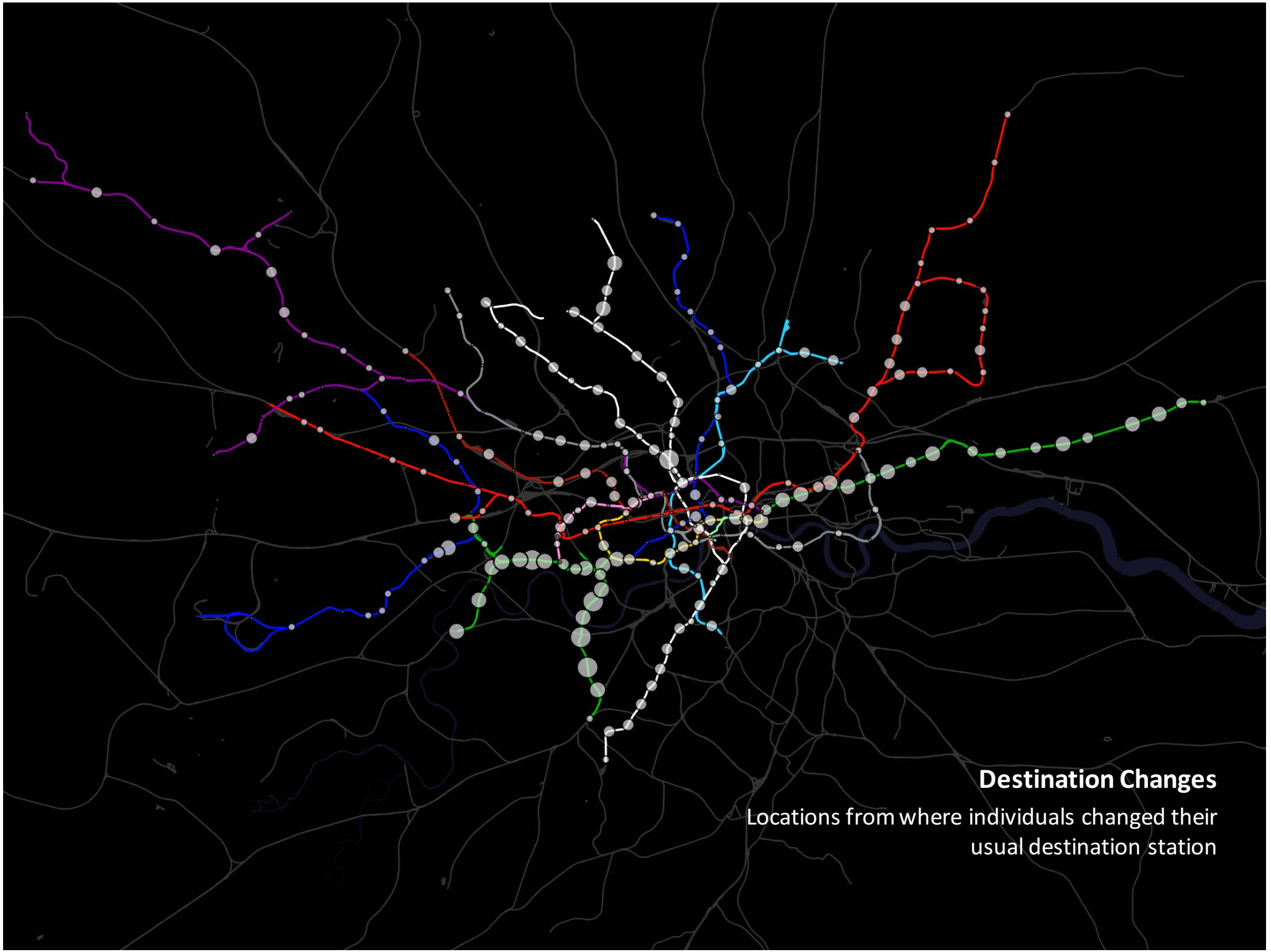


Disruptions – Routine Analysis of Daily Events




No Change: Increased Travel Time
Greater than 2SD above mean increase on usual travel time for that Oyster Card
Size equal to proportion of users that regularly travel from station during time period, and travelled that during disruption


Origin Changes
Locations from where individuals changed
from their usual origin station


Destination Changes
Locations from where individuals changed their usual destination station

Challenges for Urban Informatics: Theory, Modelling, and Applications: Policy and Management

We need a really well worked out landscape of how all this connects up

We need new theories of dynamics that integrate individual spatial behaviors with respect to spatial variations in the city

We need to tie flows and networks to dynamics and to locations. Chen Zhong my post doc, now lecturer at Kings (KCL) has done a lot of work – see IJGIS

International Journal of Geographical Information Science, 2014
http://dx.doi.org/10.1080/13658816.2014.914521



Detecting the dynamics of urban structure through spatial network analysis

Chen Zhong[a]*, Stefan Müller Arisona[a,b], Xianfeng Huang[c], Michael Batty[d] and Gerhard Schmitt[a]







University College London

Thanks

Acknowledgements

Melanie Bosredon, Gareth Simons, Roberto Murcio, Richard Milton, Oliver O'Brien, Stephen Gray, Fabian Neuhaus, Jon Reades, Ed Manley, Chen Zhong, Flora Roumpani & Stephan Hugel

http://www.complexcity.info/
http://www.spatialcomplexity.info/
http://blogs.casa.ucl.ac.uk/

m.batty@ucl.ac.uk
@jmichaelbatty

SHUFE





上海财经大学
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS