TRANSPORT STUDIES UNIT
Centre for Research, Outreach and Graduate Studies in Transport

19TH JAN 2016

UNIVERSITY OF OXFORD

Improving Urban Mobility

Transit Systems, New Technologies & Smart Cities

Michael Batty

m.batty@ucl.ac.uk
@jmichaelbatty

http://www.complexcity.info/
http://www.spatialcomplexity.info/

Outline

- Some Ideas about Smart Cities and Big Data
- Real-Time Streaming: The Oyster Card Data Set
- Learning about Mobility from the Data

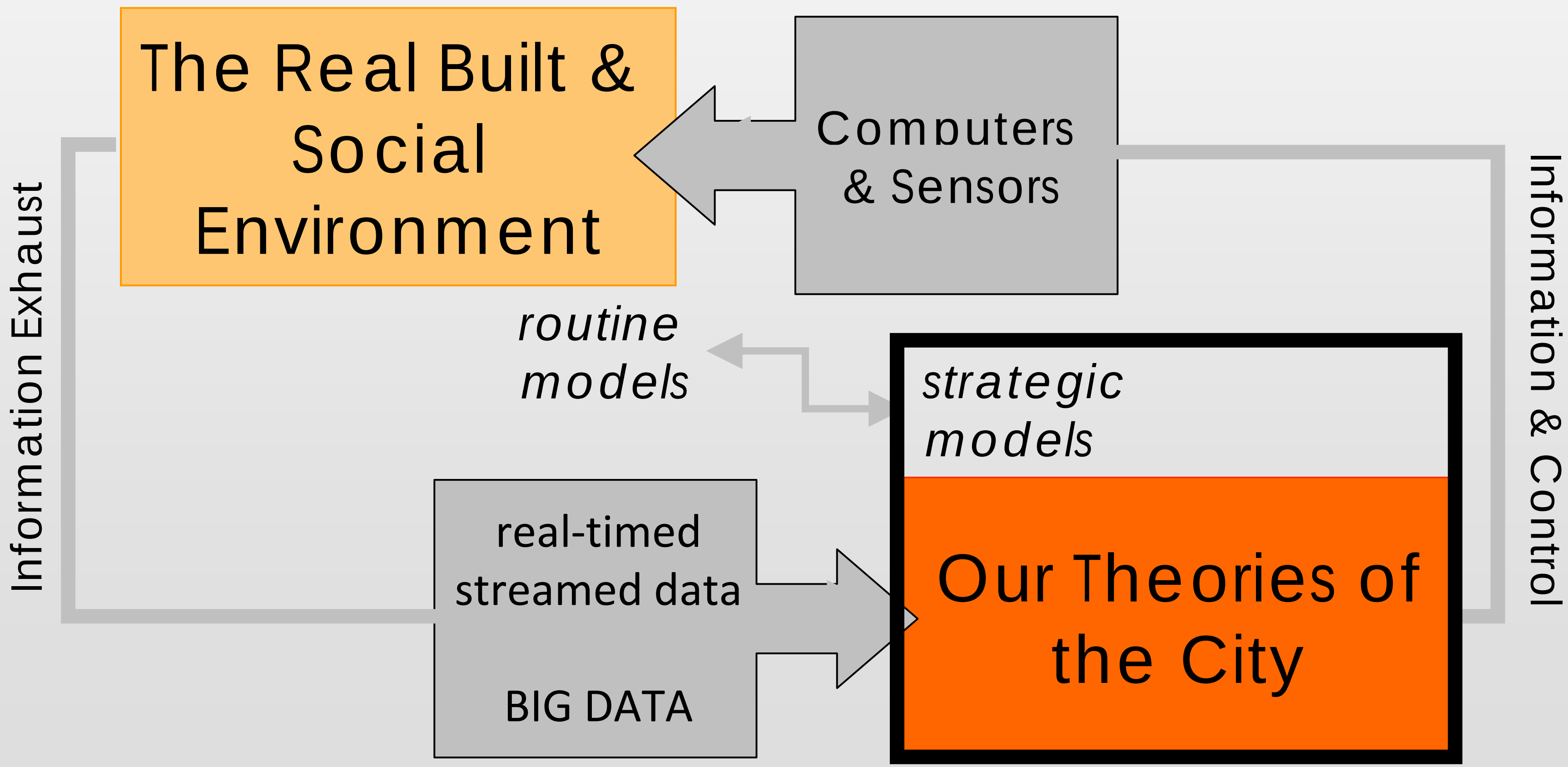
  Variabilities – Heterogeneity and Travel Profiles

  Disruptions – Signal Failures, Stalled Trains

  Variable Locational Dynamics of Demand
- Related Real -Time Data: Bikes, Social Media
- What Can We Learn: The Limits to Big Data

Some Ideas about Smart Cities and Big Data

The spreading out of computers into public places & the built environment and all their consequences



- The way we access the smart city is through technologies that let us generate and use data and its useful equivalent – **information** (data) is key
- Access through **mobile** and **fixed devices** like phones, smart cards, through fixed sensors
- These usually complement rather than substitute for data which we collected and used in the past. This data still essential and highly relevant.
- This has **introduced time into our thinking** – in the past most urban planning for future cities was timeless –garden cities, new towns, master plans
- This is all part and parcel of increasing complexity; more time scales, more opportunities, more diversity

Centre for Advanced Spatial Analysis

Real-Time Streaming: The Oyster Card Data Set

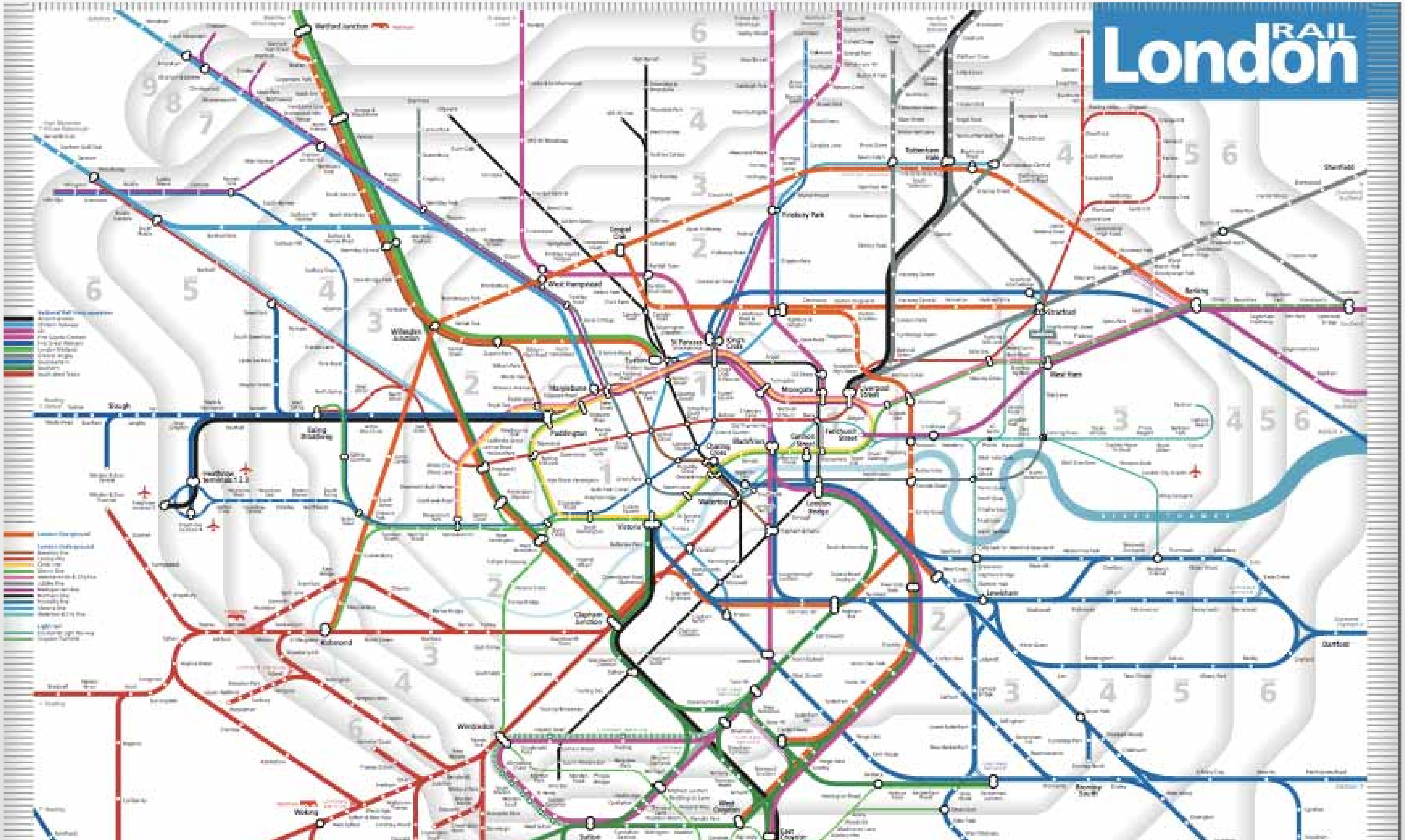
Tap at **start** and **end** of train journeys

Tap at **start only** on buses

Accepted at 695 Underground and rail stations, and on thousands of buses

Many Variants of the Data Sets

991 million Oyster Card taps over Summer 2012 – this is big data



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

UCL ENGINEERING
Change the world

UCL

Menu +

OYSTER GIVES UP PEARLS

How studying millions of Oyster Card journeys reveals London's 'polycentres'

Researchers from UCL have analysed millions of Oyster Card journeys in a bid to understand how, why and where we travel in London.

Professor Michael Batty (UCL Centre for Advanced Spatial Analysis) and Dr Soong Kang (UCL Management Science and Innovation) applied the techniques of statistical physics to their mountain of raw data.

The pair joined forces with a computational social scientist and a physicist, both based in Paris, to explore patterns of commuting by tube into central London.

They used Transport for London's database of 11 million records taken over one week from the Oyster Card electronic ticketing system.

Latest news from UCL Engineering

- New web privacy system could revolutionise the safety of surfing
- UCL host Google Girls Coding Programme with Generating Genius and University of West indies
- Professor Polina Bayvel to Give Royal Society Lecture

Twitter feed

RT @markmiodownik: Am giving a ENGins seminar today for @UCLEngineering @UCLENGins all UCL engineers welcome - Roberts G06, 6:30pm. http:/...
8:58am Thu 9th October 2014

RT @Centre4EngEdu: We're hiring! Multi-talented Centre Administrator required to help us launch and expand! bit.ly/ZsERSM
10:54am Wed 8th October 2014

Join our mailing list

[redacted]

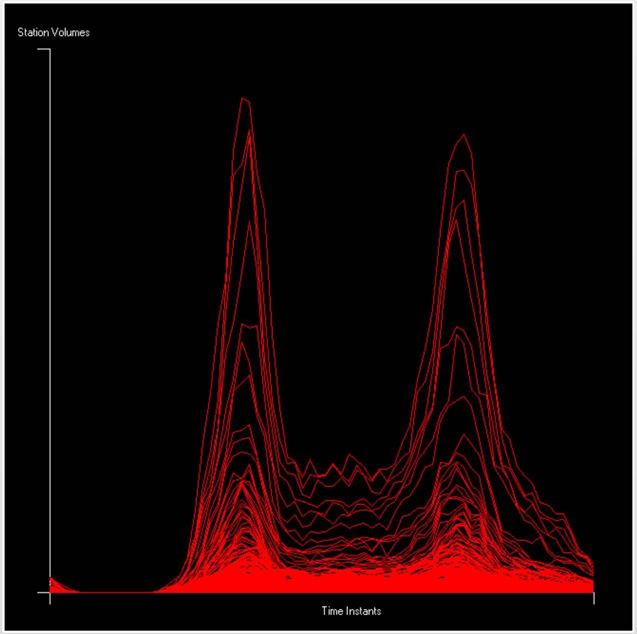
And how can we make sense of this

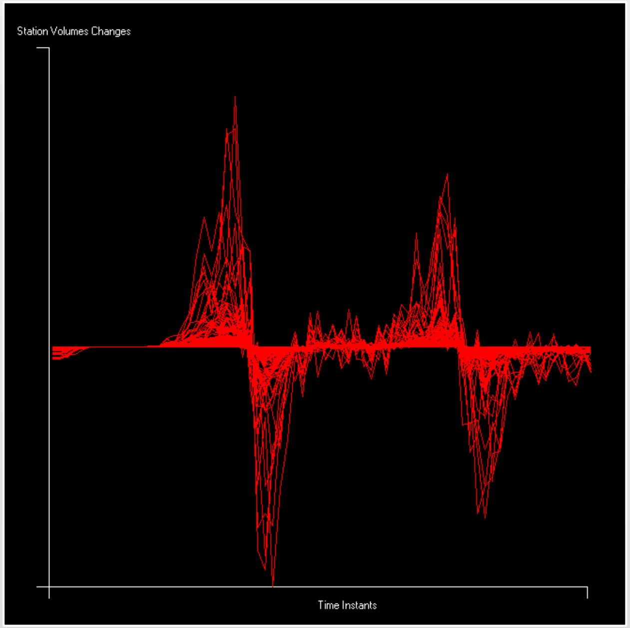
http://www.simulacra.info/

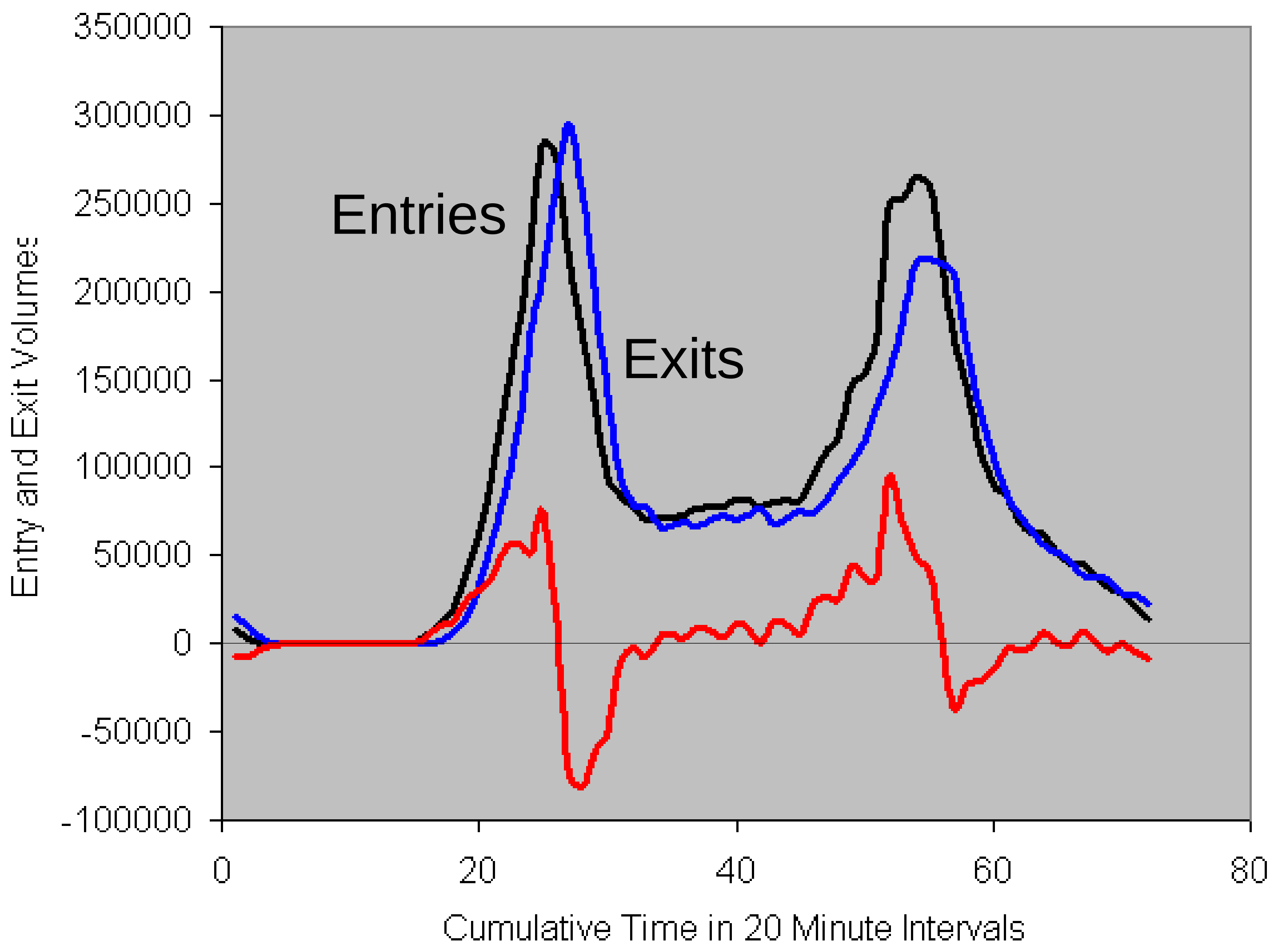
[redacted]

Variabilities – Heterogeneity and Travel Profiles

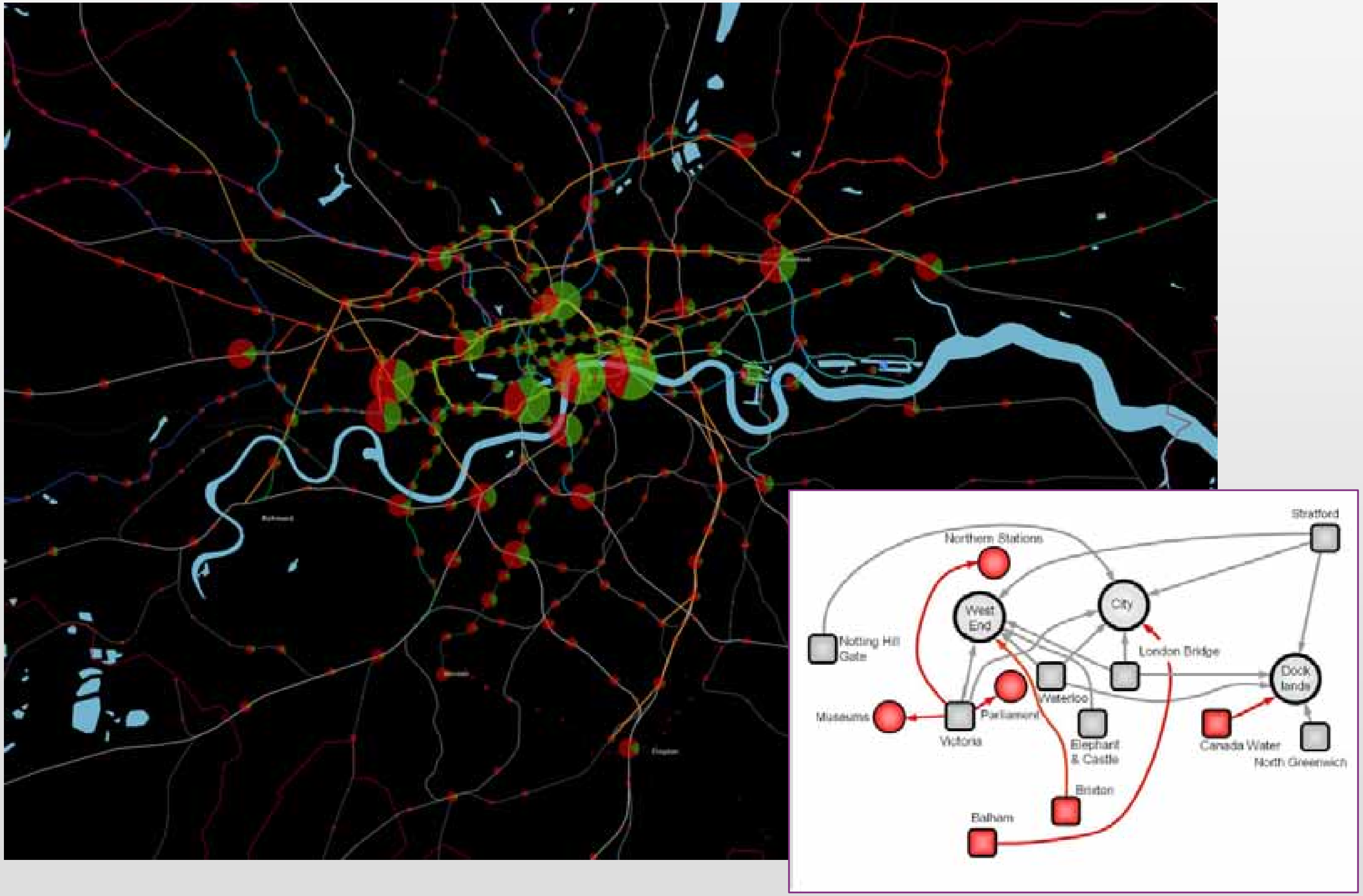
First we will look at some of the data and how it varies in terms of the diurnal flows usually morning and evening peaks, with a small blip (peak) around 10pm at night



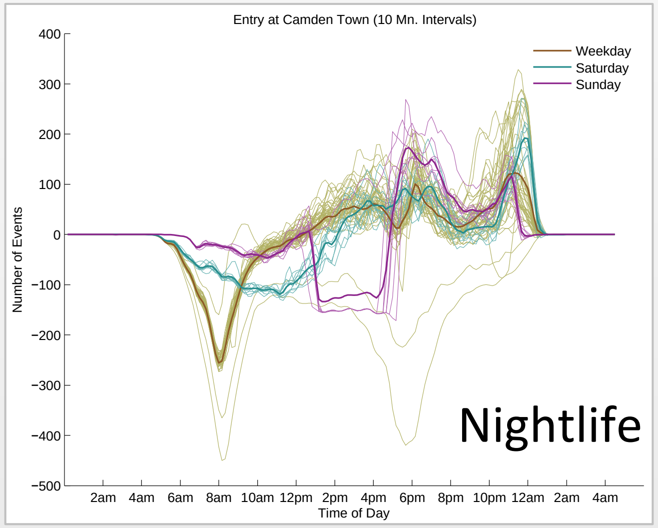


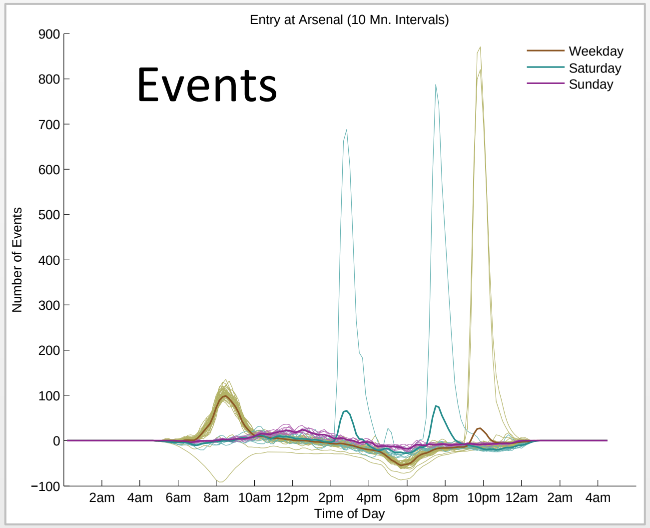

Entries
Exits
350000
300000
250000
200000
150000
100000
50000
0
-50000
-100000
0
20
40
60
80
Entry and Exit Volumes
Cumulative Time in 20 Minute Intervals

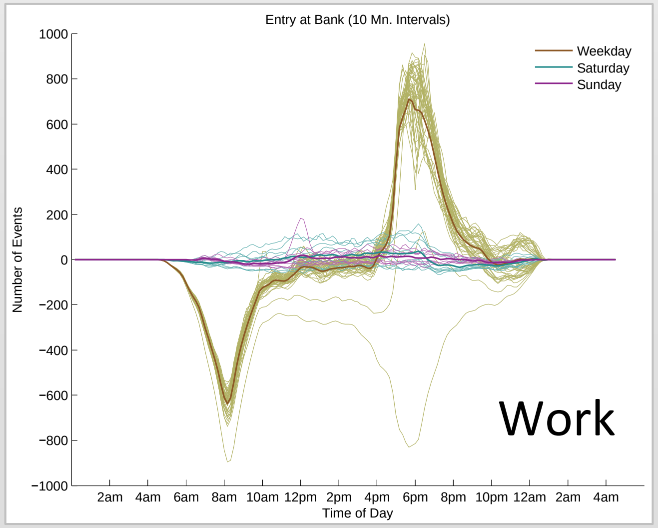
Oyster Card Data – interpreting urban structure, multitrips, etc.

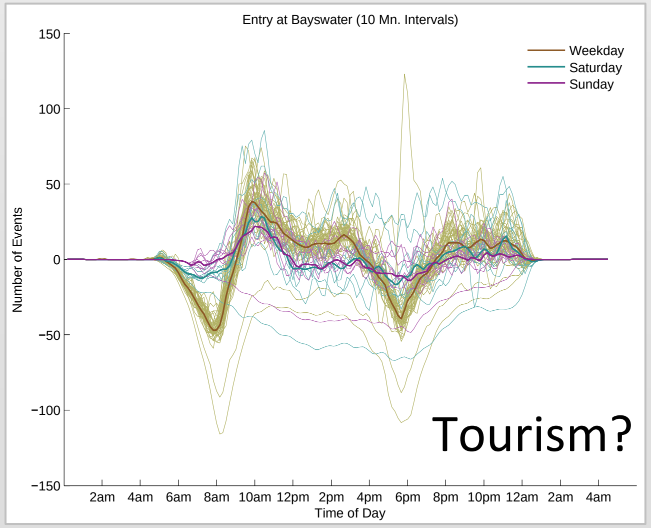


Particular Events: Weekdays, Saturdays and Sundays









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)

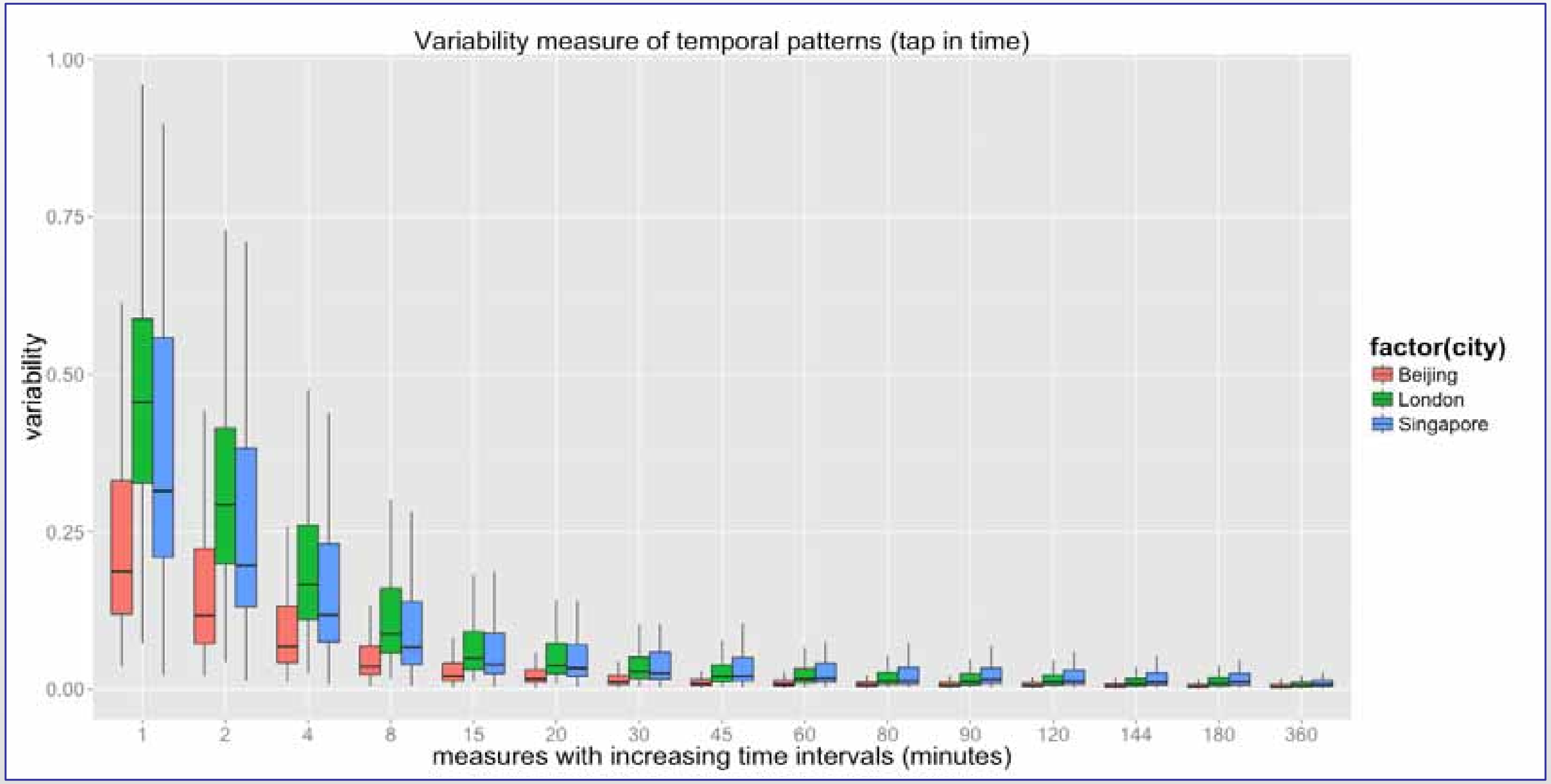
	London	Singapore	Beijing
Monday	3,457,234	2,208,173	4,577,500
Tuesday	3,621,983	2,250,597	4,421,737
Wednesday	3,677,807	2,277,850	4,564,335
Thursday	3,667,126	2,276,408	4,582,144
Friday	3,762,336	2,409,600	4,880,267
Number of stations (1)	400	130	233
Number of tube line	13	4	17
Area (2)	1,572 km^2	718.3 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

From 1 minute intervals to the whole day



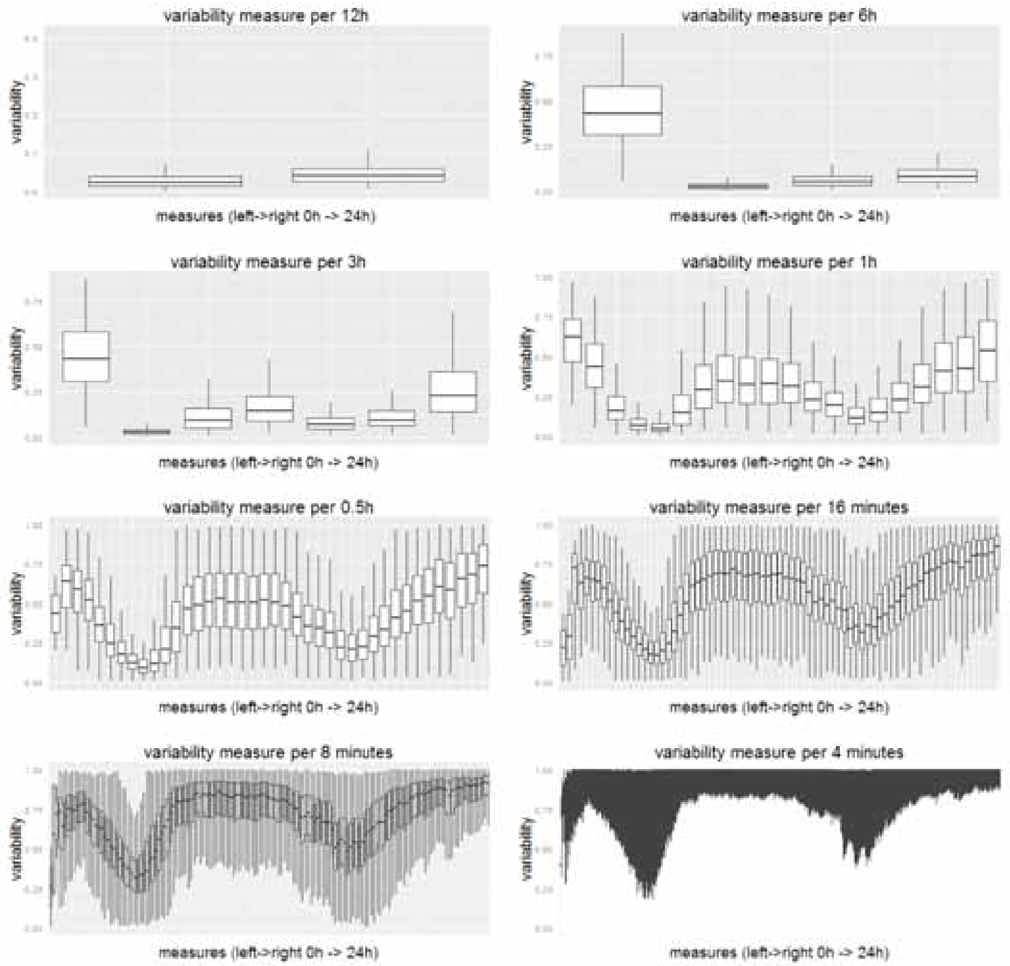
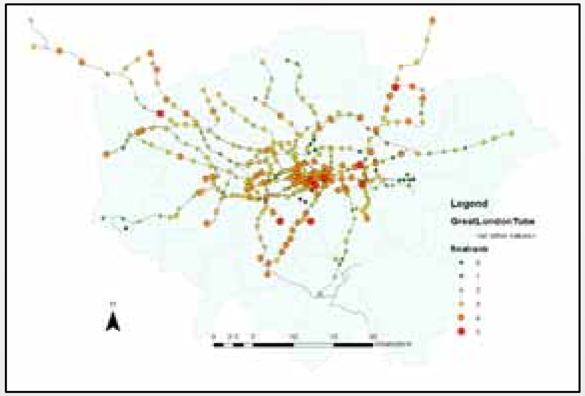


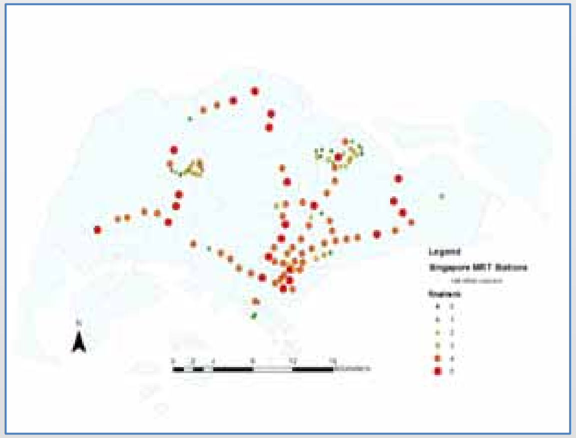
Figure 1. Variability of regularity in the trip matrix over time.

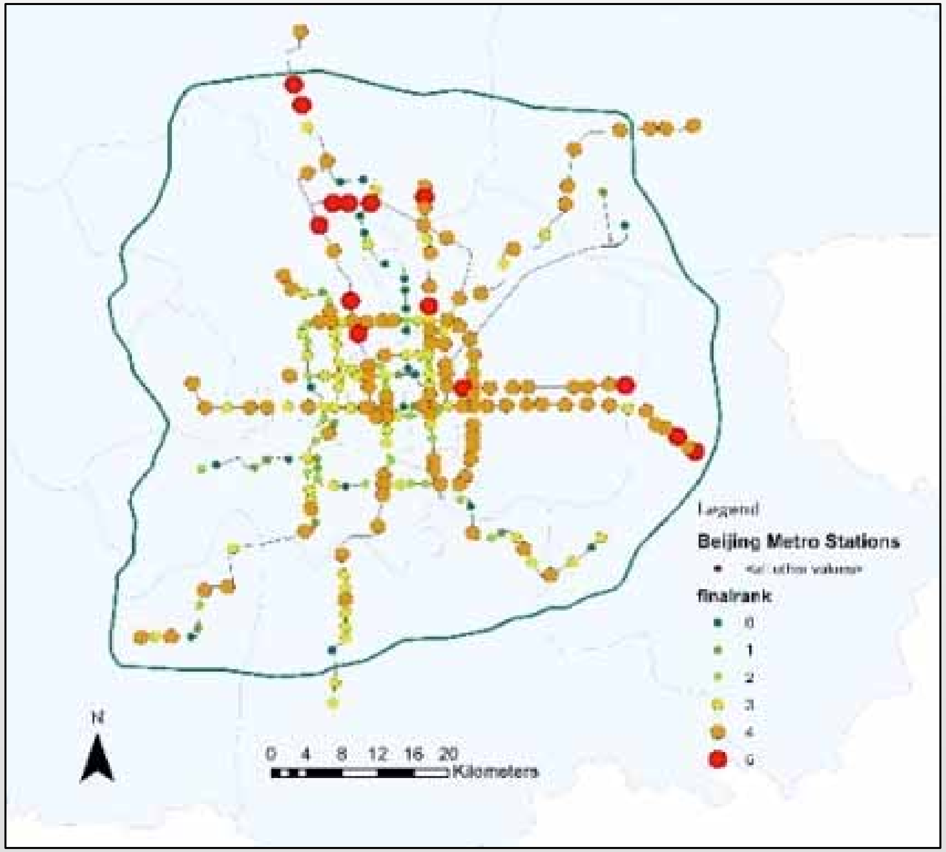
Note: Each box plot shows the variability of 400 stations over time measured at different temporal scales. Overall, eight subplots give a similar trend where lower variability appears during peak hours (around 9 am in the morning and 6pm in the evening). More details can be captured as differences of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

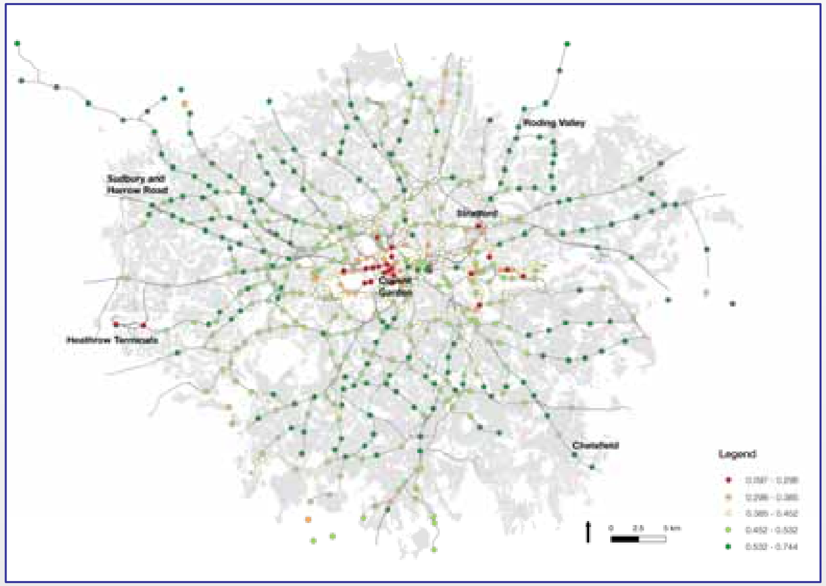
Comparing Variability for different time Intervals over the day

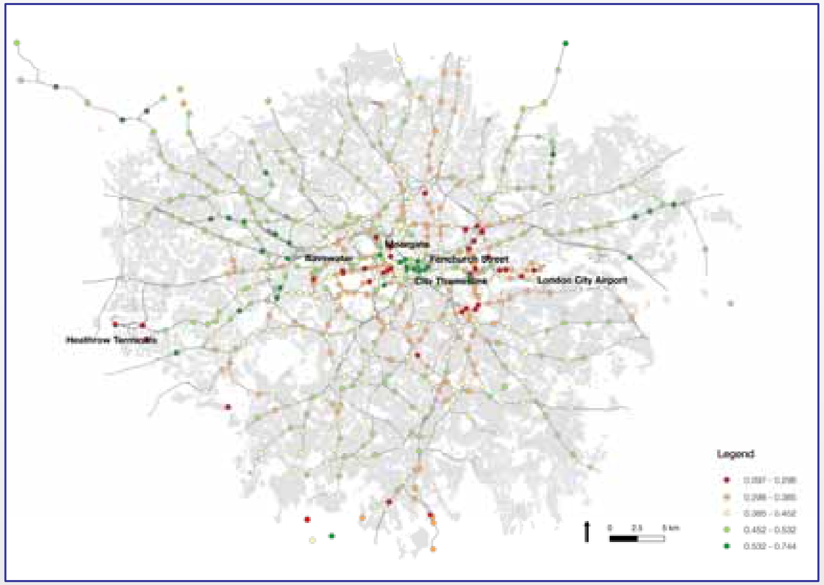
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

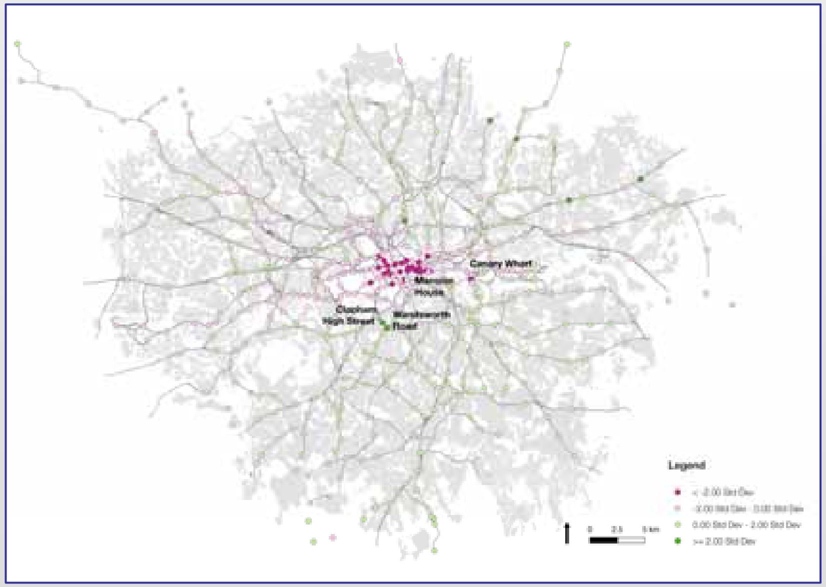












Maps of Underground and Rail stations in London visualised by the proportion of regular trips

originating at each location

ending at each location

starting and ending at each location

Disruptions – Signal Failures, Stalled Trains

- We will look at three disruptions – the Circle and District Lines which had a 4 hour stoppage on July 19th 2012

- And a Bus Strike in East London and how this shows up in the data

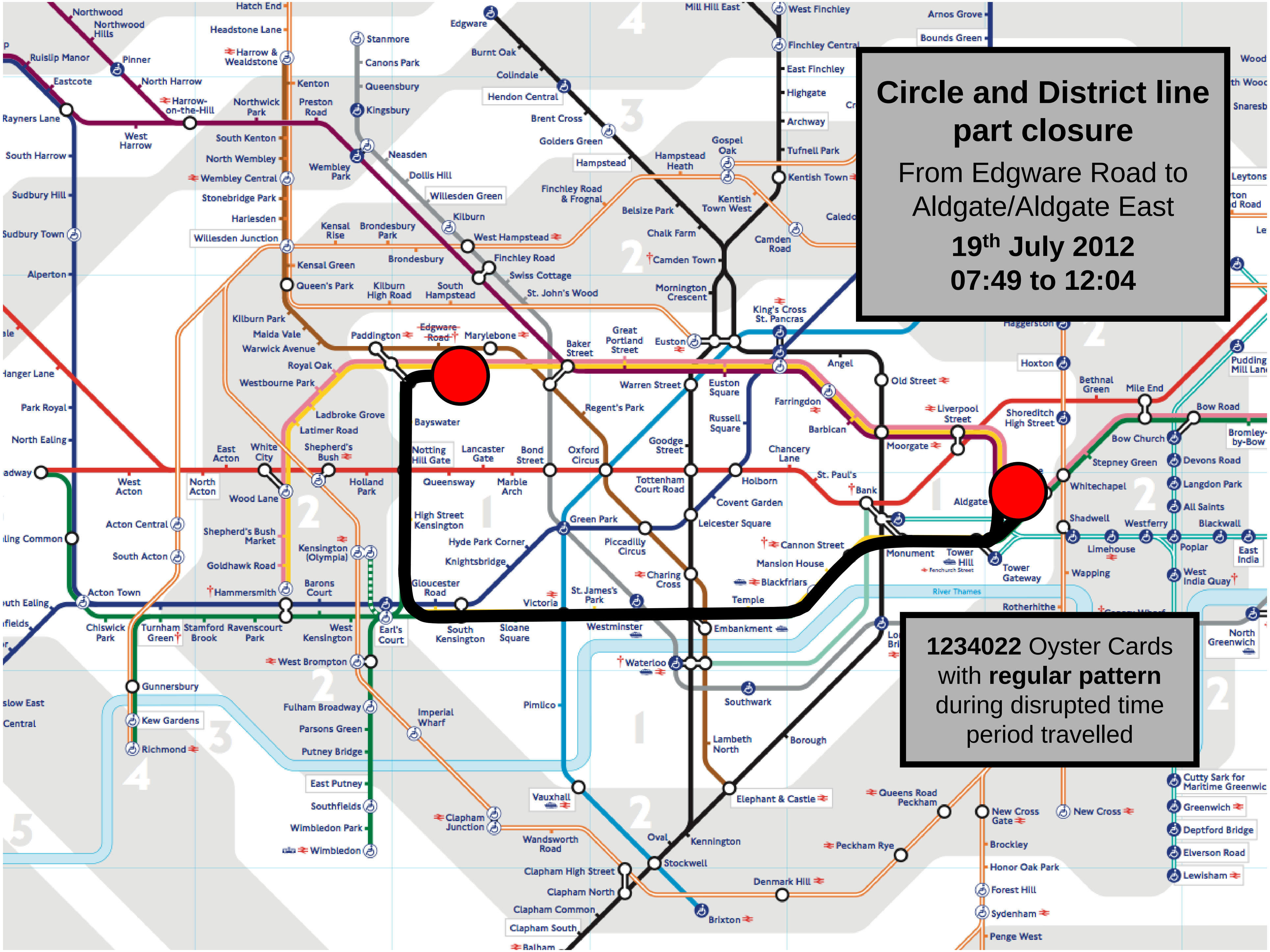
- And typical pattern of delay on all modes visualised for Greater London

Circle and District line part closure

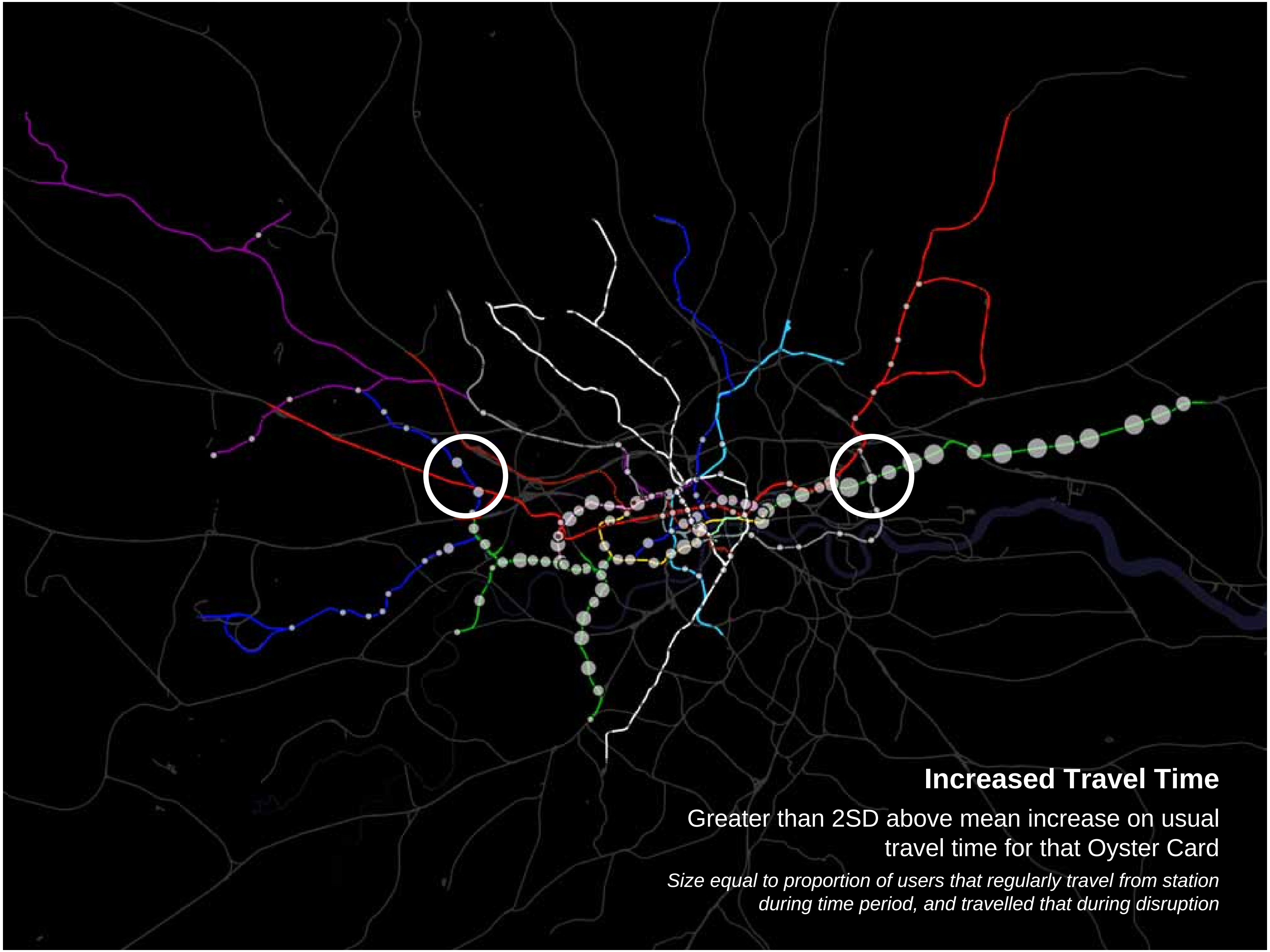
From Edgware Road to Aldgate/Aldgate East

19th July 2012

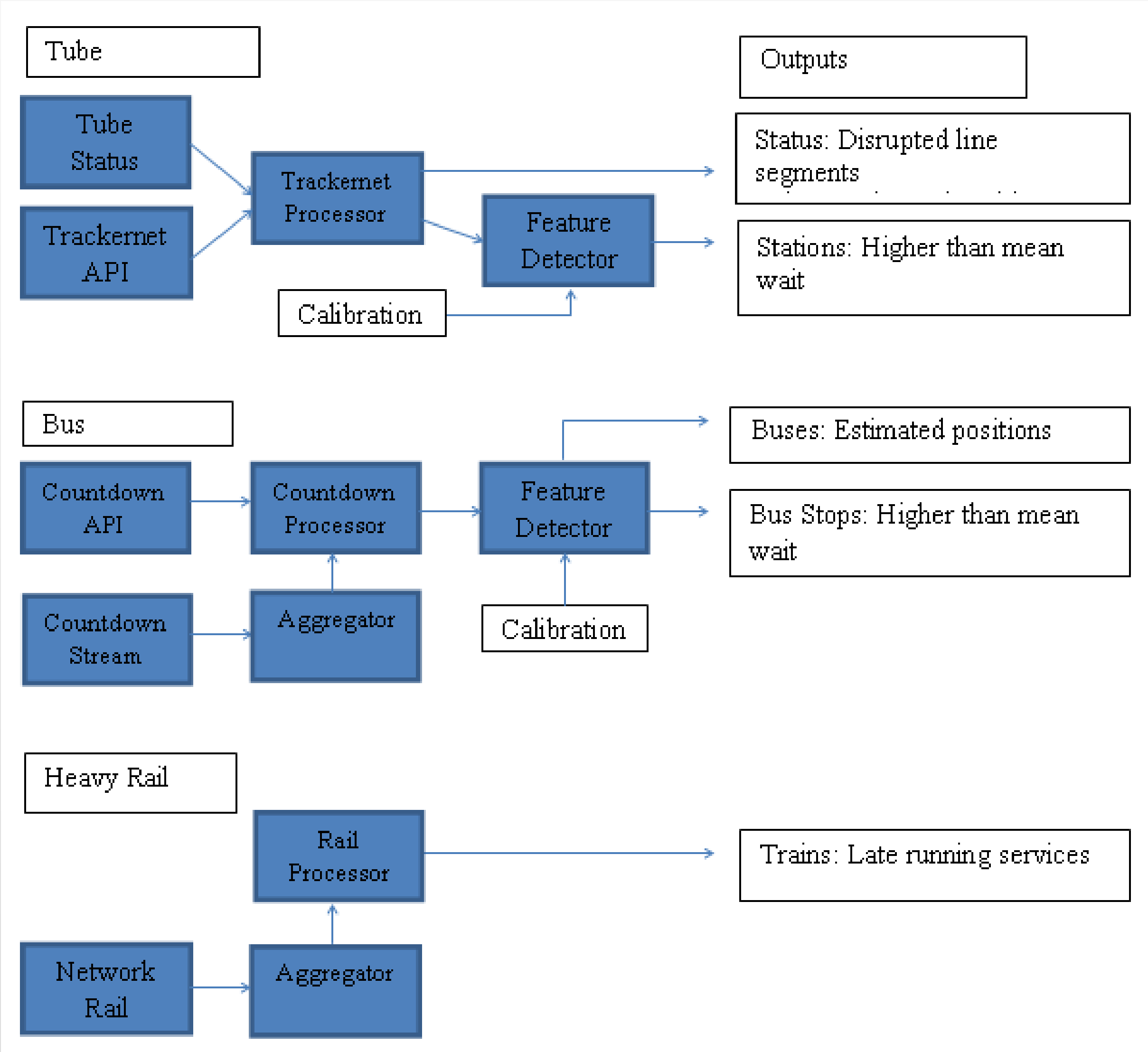
07:49 to 12:04

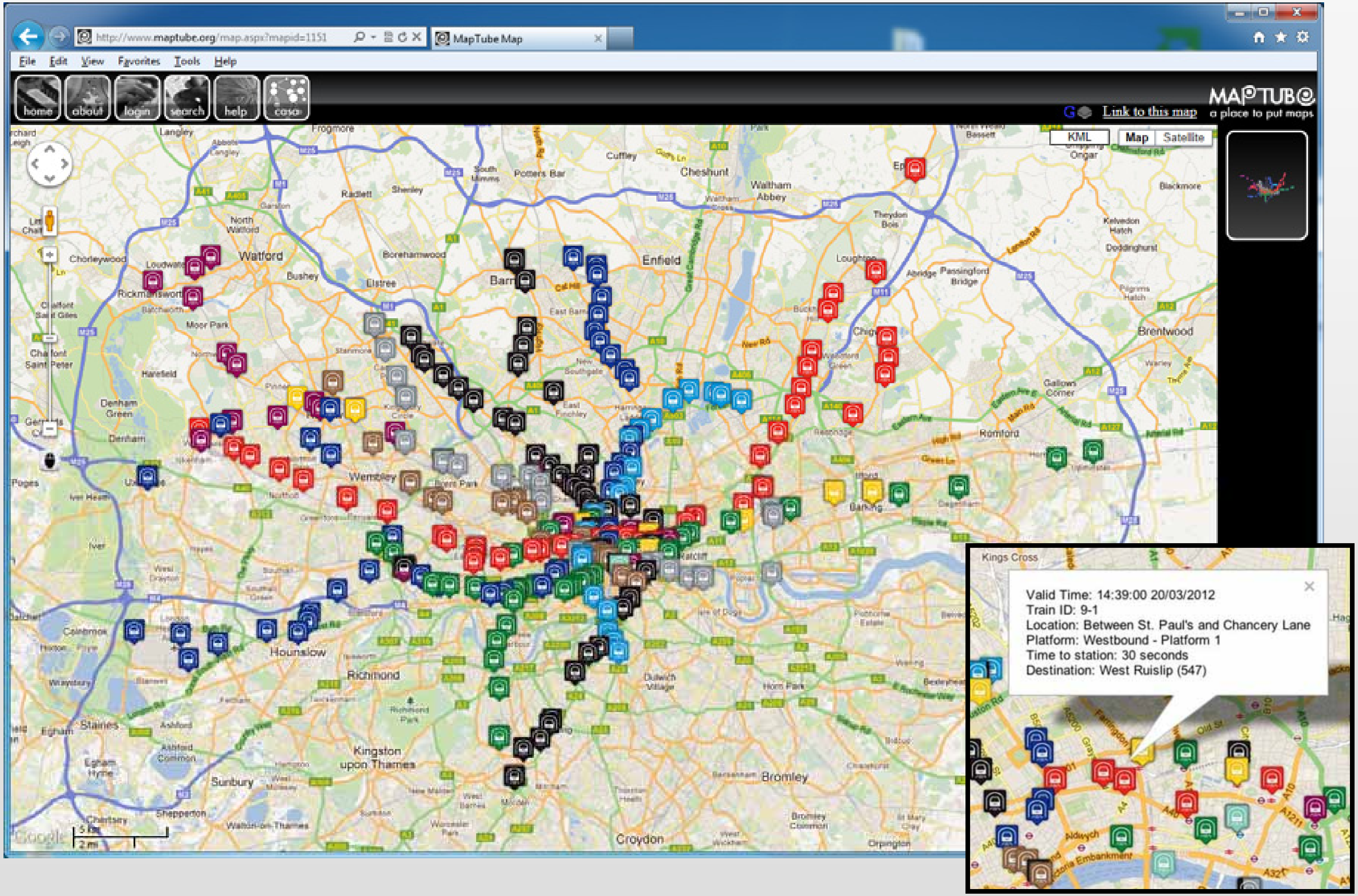


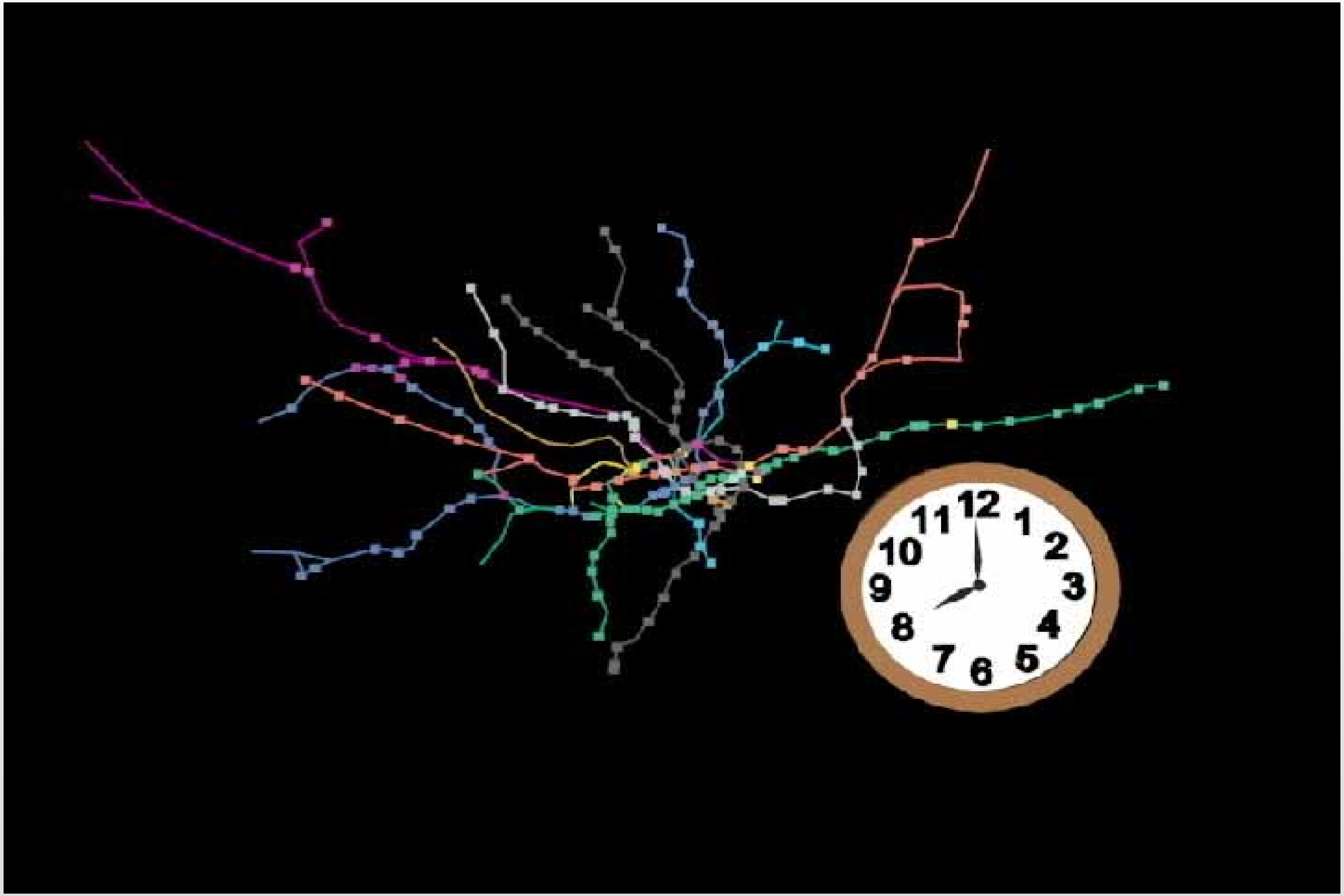
1234022 Oyster Cards with **regular pattern** during disrupted time period travelled


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

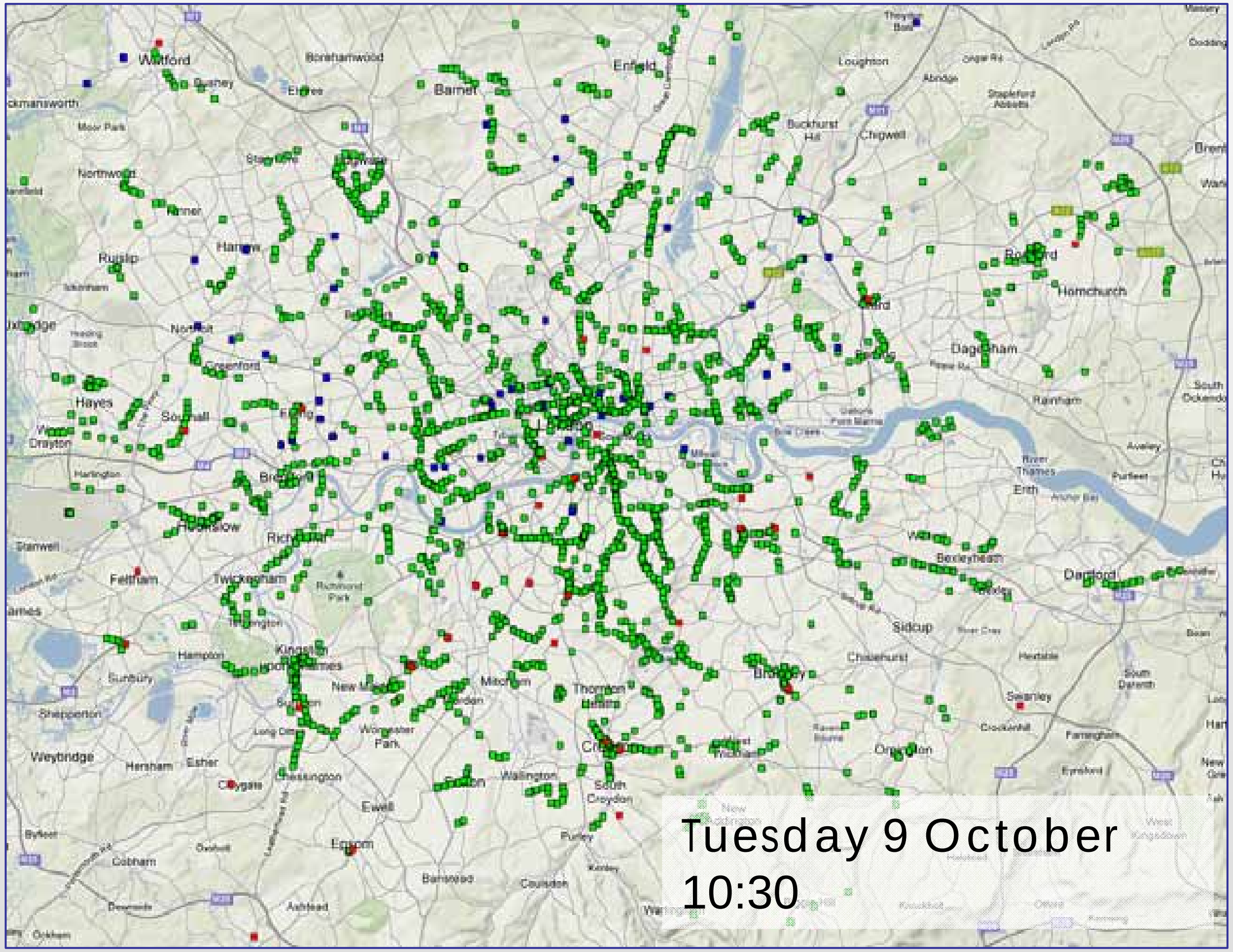
The Public Transport System in Terms of Vehicle Flows




http://www.maptube.org/map.aspx?mapid=1151
MapTube Map
File Edit View Favorites Tools Help
home about login search help casa
Link to this map
MAPTUBE
a place to put maps
KML Map Satellite
Kings Cross
Valid Time: 14:39:00 20/03/2012
Train ID: 9-1
Location: Between St. Paul's and Chancery Lane
Platform: Westbound - Platform 1
Time to station: 30 seconds
Destination: West Ruislip (547)


12
1
2
3
4
5
6
7
8
9
10
11

Delays from Tube, National Rail and Bus Fused



Key

- National Rail more than 5 minutes late
- Tube stations showing a wait time 15% above expected
- Bus stops showing a wait time 20% above expected

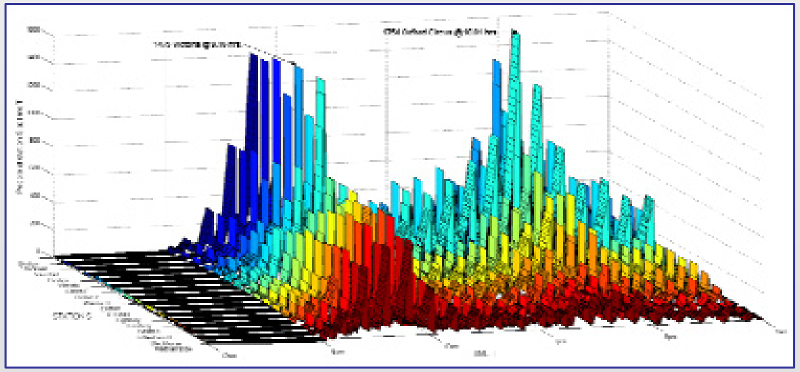
Tube delays from the TfL status feed are also plotted as lines

Locational Dynamics of Demand

We are currently using information theory to figure out how much information from trips is transmitted from station to station through time by working out how many passengers are in stations or on trains in stations over time. We are using the concept of **transfer entropy** to do this. I don't have time to say much about this but here is a picture about this for one line

Stations S — Euston @ 8.29am

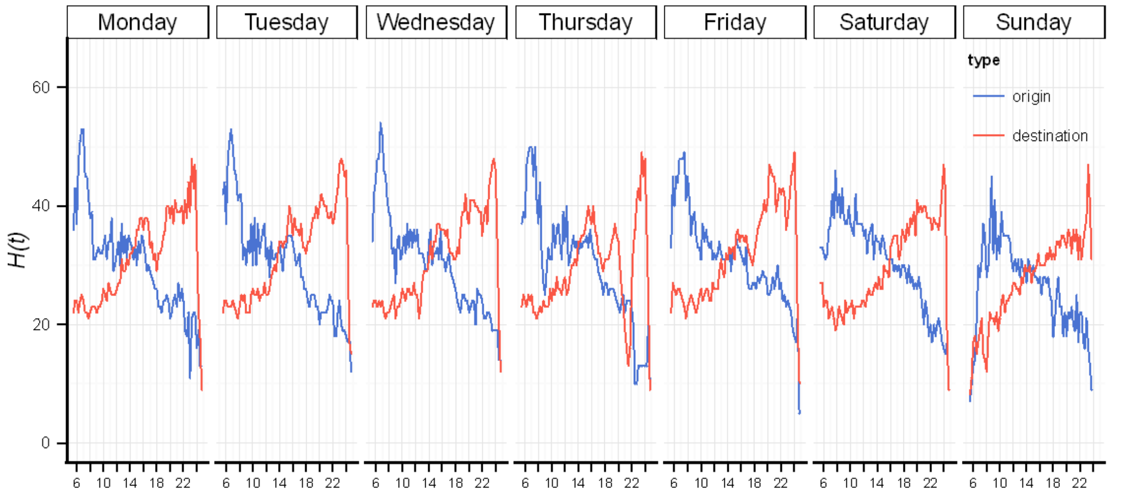
Time t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8.30am	196	61	211	132	400	152	488	358	252	252	188	170	128	180	144	159
	181	164	199	138	304	144	489	415	210	647	228	157	218	76	55	160
	178	165	222	179	288	188	395	280	604	274	211	137	178	178	111	137
	173	128	232	91	276	167	326	206	299	246	180	243	227	65	72	153
	183	114	217	132	322	279	342	562	244	278	143	153	111	135	148	130
	252	142	230	87	275	212	614	313	214	264	262	218	222	103	72	156
	171	85	216	162	375	203	469	257	251	155	136	183	82	201	140	128
	162	205	206	161	299	372	424	286	206	158	193	248	116	132	60	165
	124	170	215	176	366	265	422	301	[illegible]	312	181	200	106	241	121	158
	143	171	143	138	470	200	427	262	$\boxed{300}$	170	202	254	223	147	46	207
	206	117	236	161	381	315	431	190	178	241	201	133	142	213	120	186
	250	125	224	138	320	253	310	325	223	231	292	144	239	149	42	202
	224	86	277	163	367	221	467	216	202	281	173	123	170	209	112	149
	210	176	206	173	313	210	400	301	242	232	164	203	206	121	56	162
	209	224	216	178	297	273	447	266	192	361	138	177	176	200	164	94
	243	172	185	102	299	143	452	292	313	223	181	257	176	90	73	110
	256	189	225	127	348	216	520	263	220	218	163	211	95	179	162	112
	327	200	265	80	271	220	475	376	191	203	254	194	211	110	62	145
	271	118	198	148	312	239	549	270	198	221	213	179	119	208	119	142
	259	250	218	210	258	188	499	276	235	163	216	183	213	92	61	141
8.50am	310	204	263	163	338	286	508	285	170	288	252	135	162	176	97	171



$$T_{YX} = \sum_{t=1} p(y_{t+1}, y_t, x_t) \log \frac{p(y_{t+1}|y_t, x_t)}{p(y_{t+1}|y_t)}$$

Second we are working with the Oyster data again with Melanie Bosredon in out group and Marc Barthelemy in Paris on extracting clusters from the travel data using a new method of defining intensity. I will show this as a simple movie of origin and destination intensities as they change over time of day.

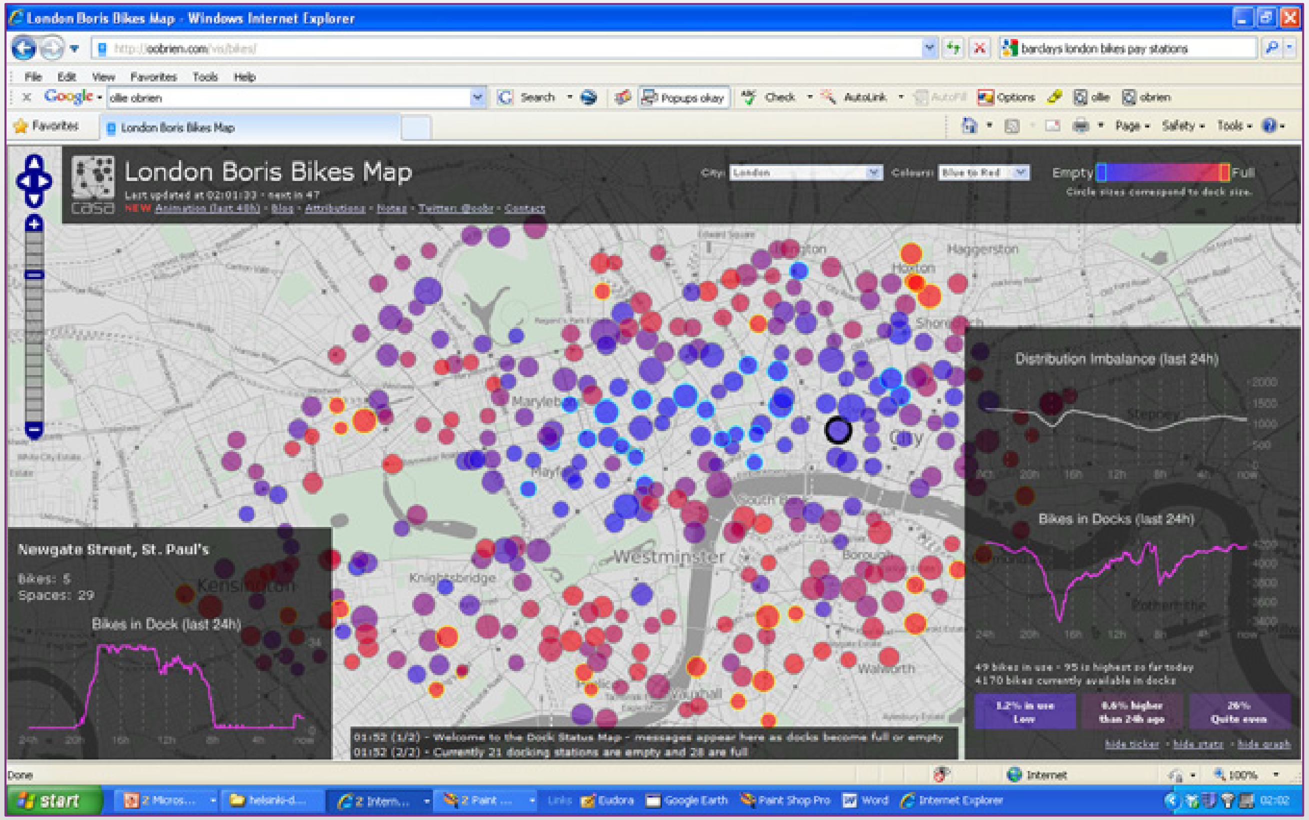
Number of hotspots (stations) vs. time



[redacted]

Related Real-Time Data: Bikes, Social Media

A lot of data is now coming online for travel and one of our group Oliver O'Brien has some 97 bike schemes world wide for which he has online data in real time - Bikes Data – 4200 bikes, started Nov 2010, all the data– everything – all trips, all times, all stations/docks

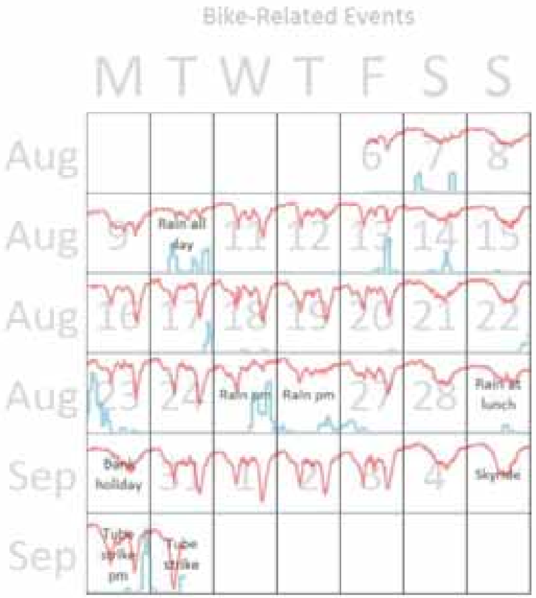


Animations of Public Bike Movements

Animations of Changes in the Bike Nodes: Docking

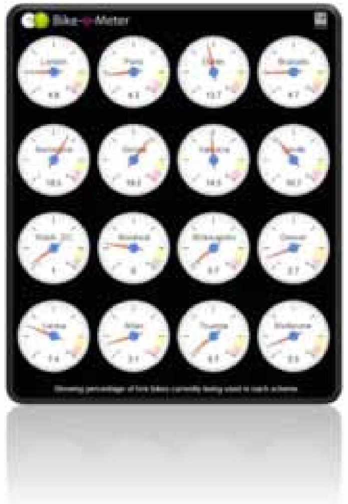
More Analysis

- **London**
- Graph shows number of bikes available to hire
- Effect of rain
  - Using the CASA weather station
- Effect of the tube strikes



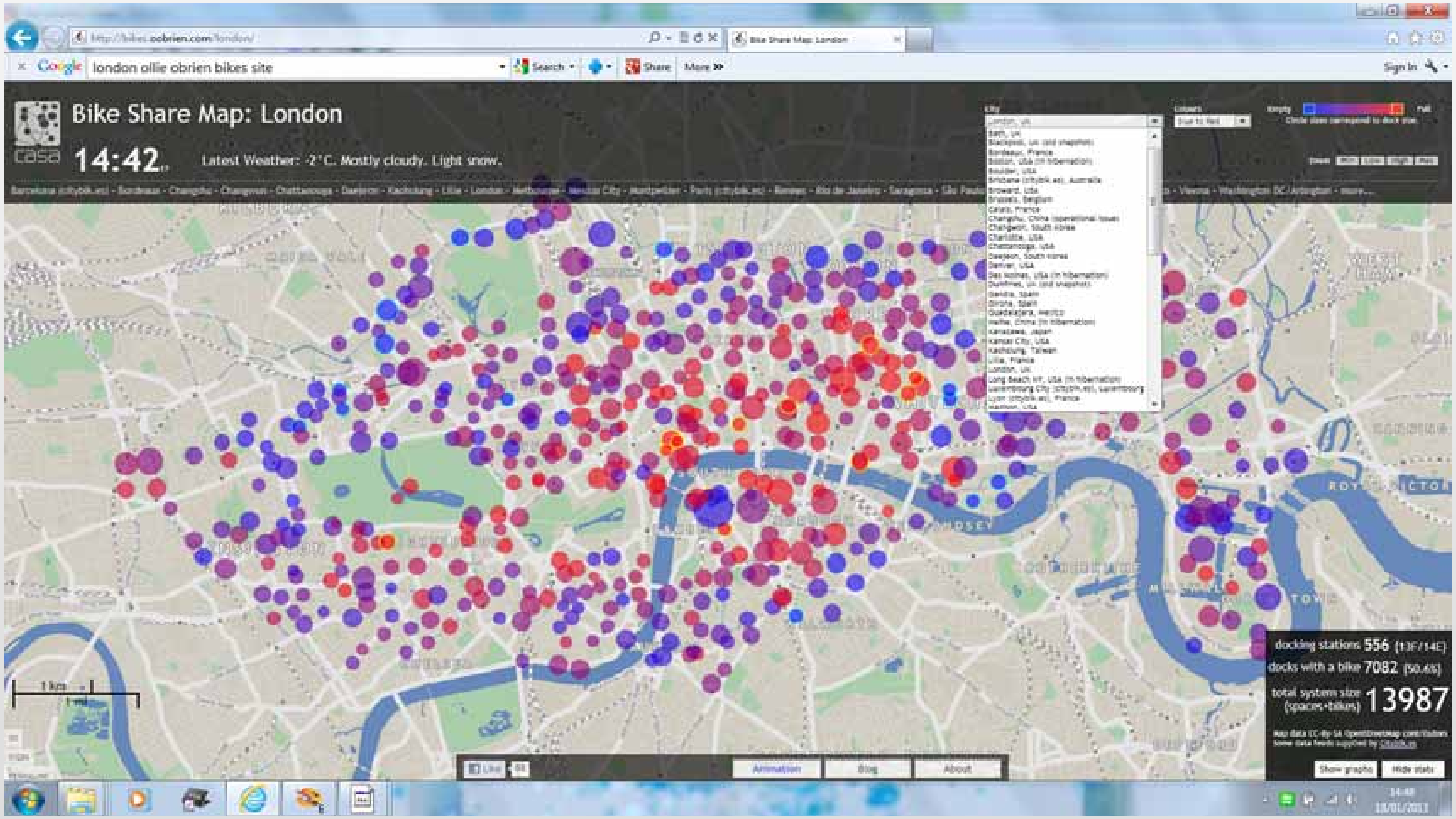
Bike-o-Meter casa.ucl.ac.uk/bom

- Tweet-o-Meter for bikes
  - Steven Gray (@frogo)
  - Using Google Gauges
- See the real life Tweet-o-Meters at the new British Library “Growing Knowledge” exhibition
  - Should be easy to hack to show the Bike-o-Meters instead ☺

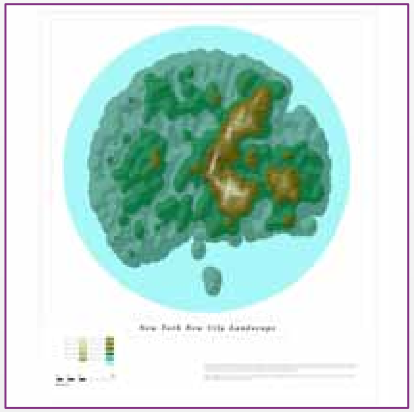


The Website: Real Time Visualisation of Origins and Destinations Activity

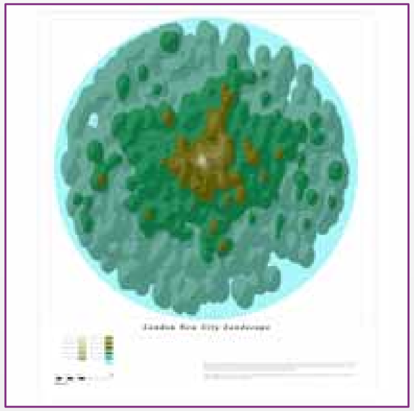
http://bikes.oobrien.com/london/



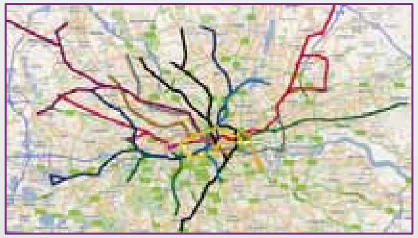

London Twitter Cloud

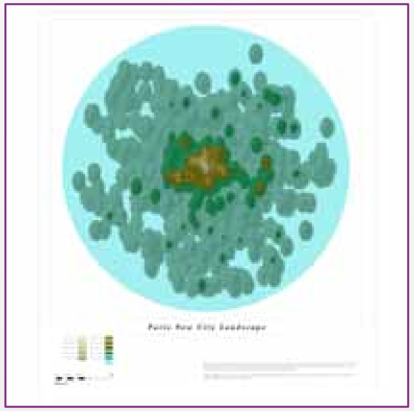


New York

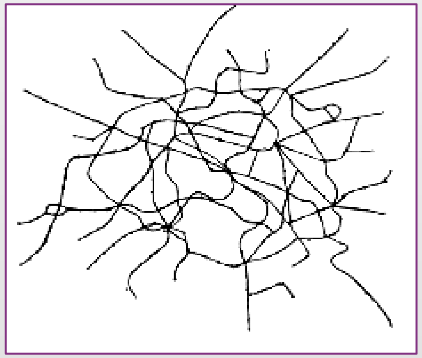


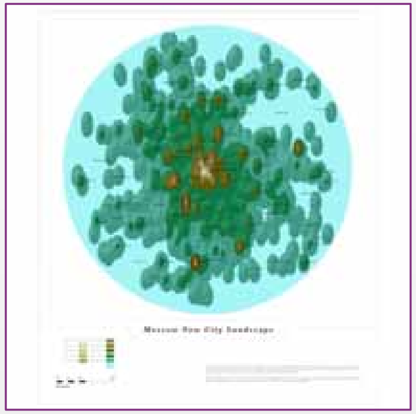
London



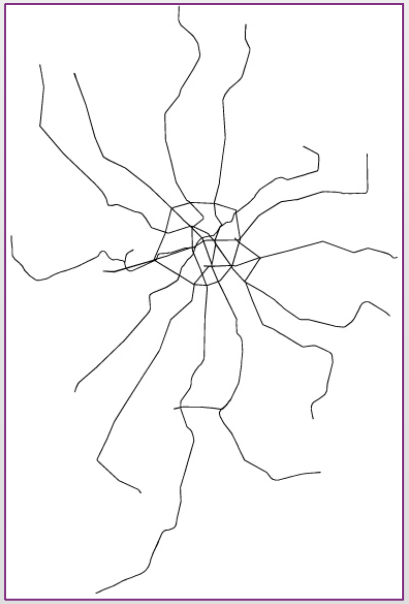


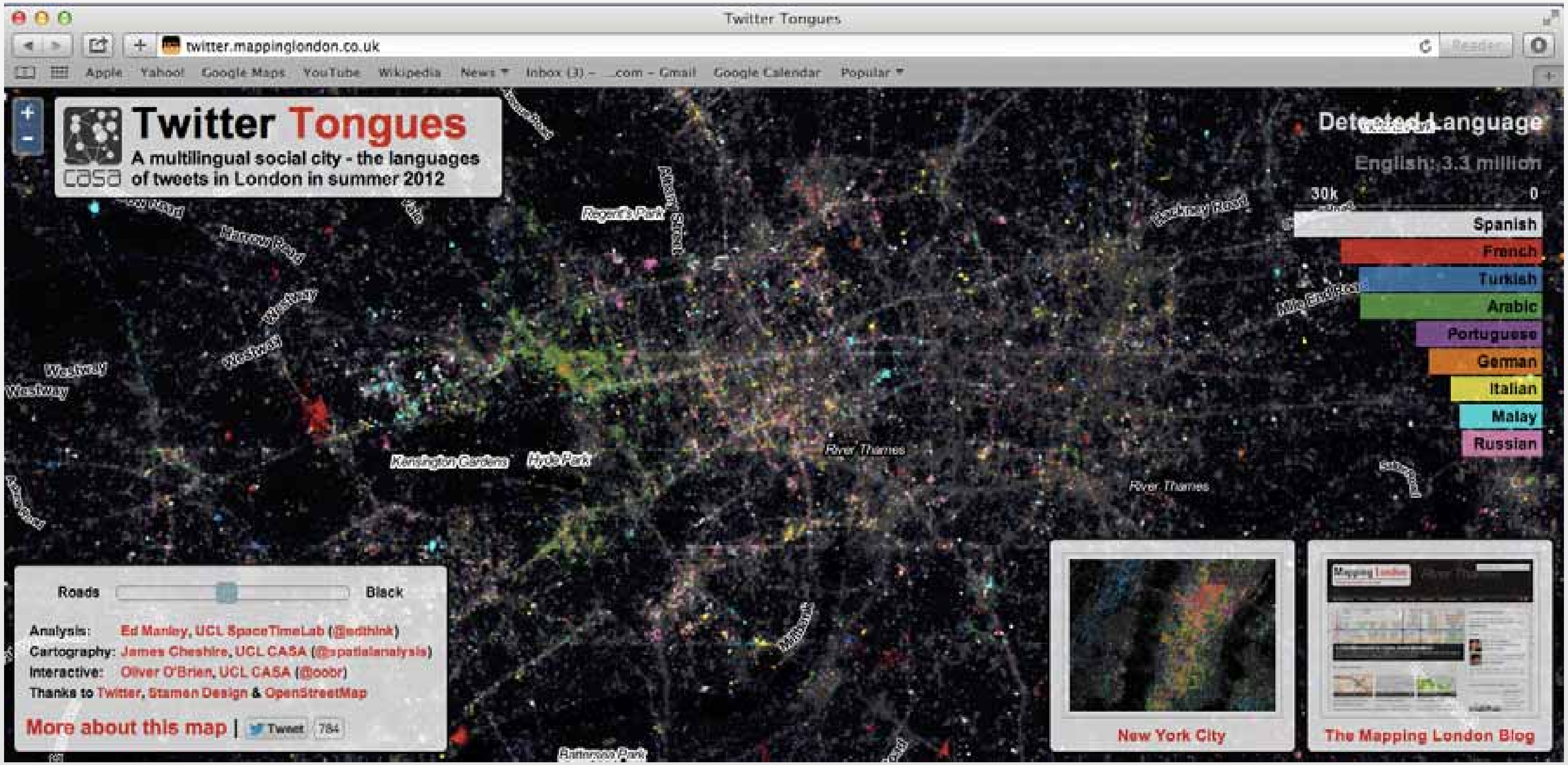
Paris



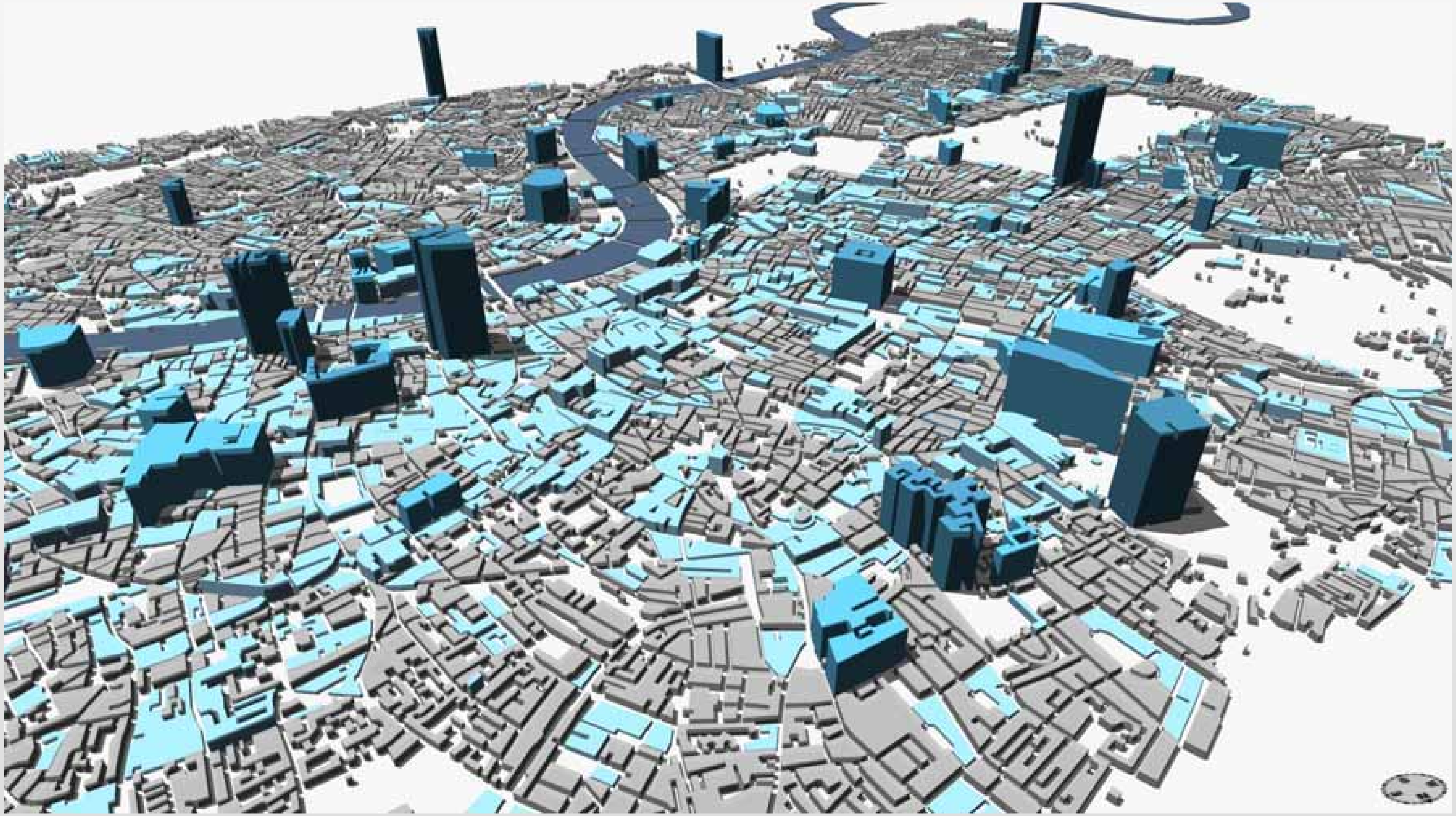


Moscow





Twitter Tongues
twitter.mappinglondon.co.uk
Twitter Tongues
A multilingual social city - the languages of tweets in London in summer 2012
Detected Language
English: 3.3 million
30k
0
Spanish
French
Turkish
Arabic
Portuguese
German
Italian
Malay
Russian
Regent's Park
Hyde Park
Kensington Gardens
River Thames
Westway
Roads
Black
Analysis: Ed Manley, UCL SpaceTimeLab (@edthink)
Cartography: James Cheshire, UCL CASA (@spatialanalysis)
Interactive: Oliver O'Brien, UCL CASA (@oobr)
Thanks to Twitter, Stamen Design & OpenStreetMap
More about this map
New York City
The Mapping London Blog



[redacted]

What Can We Learn: The Limits to Big Data

We need to add geo-demographics to this data – how – we barely have any possibility of doing this because of confidentiality

We only have a difference between young and old in terms of the card data

Chen Zhong my post doc has done a lot of work on this relating to extracting such data from related data sets producing synthetic results –our paper in IJGIS

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

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]

[redacted]

References

Manley, E., Chen, Z., and Batty, M. (2016) Spatiotemporal Variation in Travel Regularity through Transit User Profiling, to be submitted.

O'Brien, O, Cheshire, J. and Batty (2014) Mining Bicycle Sharing Data for Generating Insights in Sustainable Transport Systems, **Journal of Transport Geography**, **34**, 262–273

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

Zhong, C., Arisona, S. M., Huang, X., Schmitt, G. and Batty, M. (2014)) Detecting the Dynamics of Urban Structure through Spatial Network Analysis, **International Journal of Geographical Information Science**, http://dx.doi.org/10.1080/13658816.2014.914521

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**, in press

Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., and Batty, M. (2014) Inferring building functions from a probabilistic model using public transportation data, **Computers, Environment and Urban Systems, 48**, 124–137

Zhong, C., Manley, E., Stefan Muller Arisona, S., Batty, M.,and Schmitt, G. (2015) Measuring Variability of Mobility Patterns from Multiday Smart-card Data, **Journal of Computational Science**, doi.org/doi:10.1016/j.jocs.2015.04.021

Thanks

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

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