

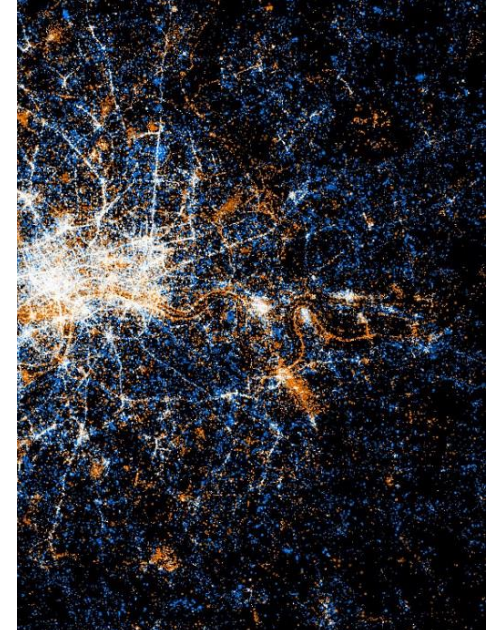
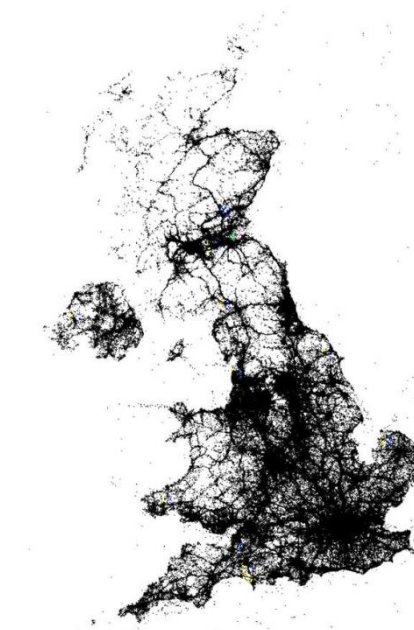
Using Social Media for Geodemographic Applications

Muhammad Adnan and Guy Lansley
Department of Geography, University College London

@gisandtech

@GuyLansley

Web: <http://www.uncertaintyofidentity.com>



Outline

1. Geodemographics and Social Media Geodemographics
2. Using Twitter Data for Geodemographics
3. Establishing Footfall Estimates across Cities
4. Identifying Bespoke Temporal Catchments of Social Media Users

Introduction

- Geodemographics
 - Analysis of **people** *by where they live*” [1]
 - Night time characteristics of the population
- Social Media Geodemographics
 - Moving beyond the night time geography
- **Who:** Ethnicity, Gender, and Age of social media users
- **When:** What time of day conversations happen
- **Where:** Where social media conversations happen

Twitter (www.twitter.com)

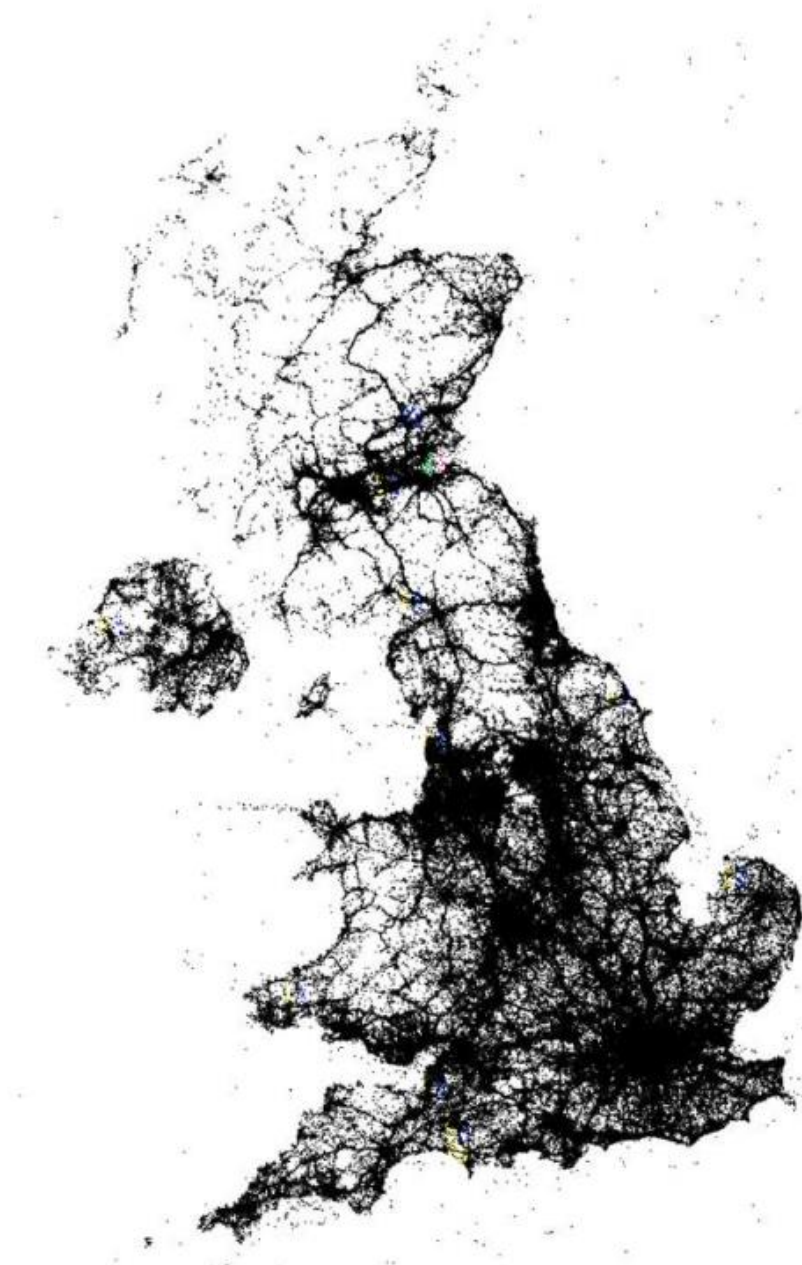
- Online social-networking and micro blogging service
 - Launched in 2006
- Users can send messages of 140 characters or less
- Approximately 200 million active users [2]
- 350 million tweets daily
- In 2012, UK and London were ranked 4th and 3rd, respectively, in terms of the number of posted tweets [3]

[2] Twitter. 2012. What is Twitter ?. Retrieved 31st December, 2012, from <https://business.twitter.com/basics/what-is-twitter/>.

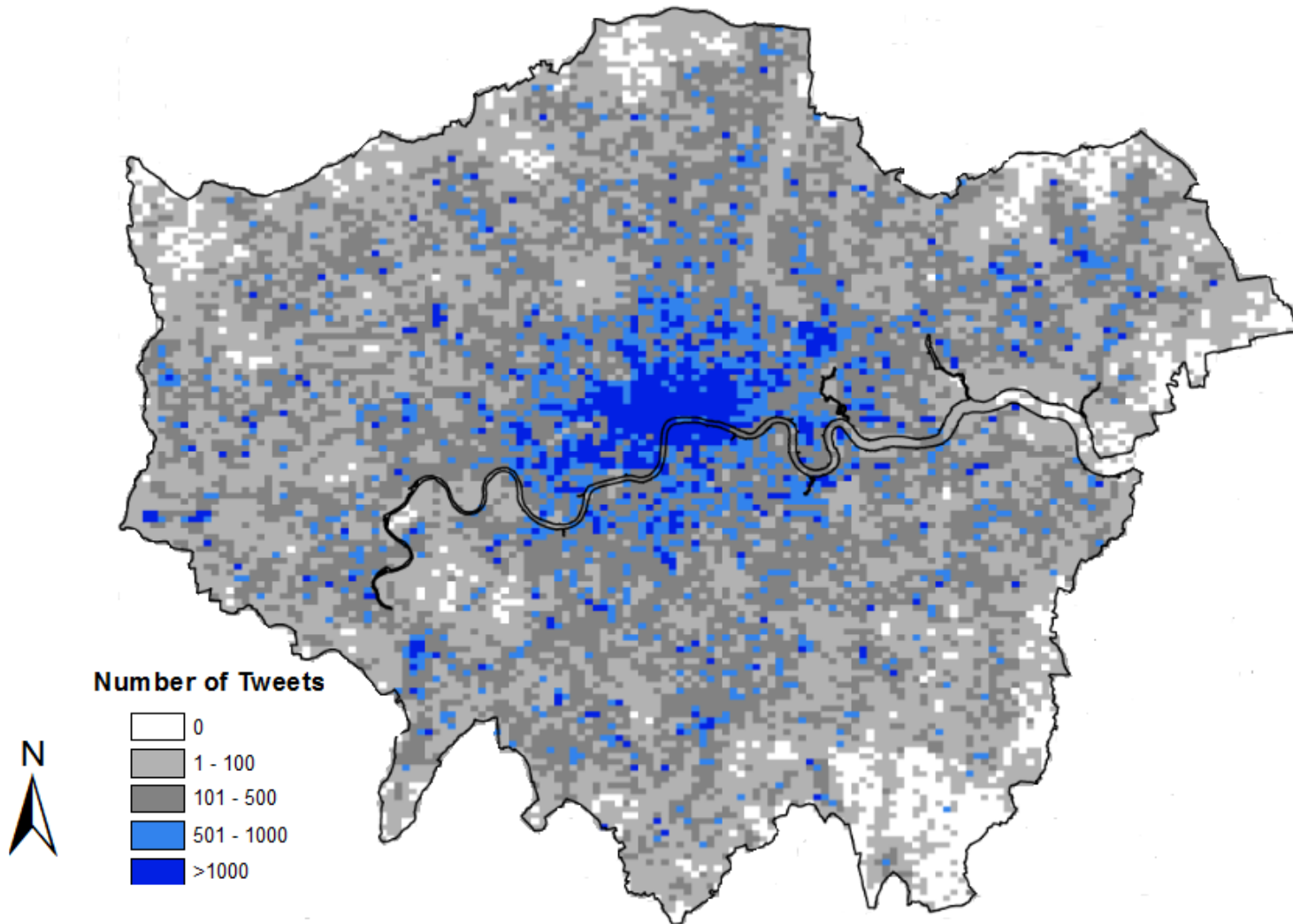
[3] Bennet, S. 2012. Revealed: The Top 20 Countries and Cities of Twitter [STATS]. Retrieved 31st December, 2012, from http://www.mediabistro.com/alltwitter/twitter-top-countries_b26726.

Data available through the Twitter API

- User Creation Date
- Followers
- Friends
- User ID
- Language
- Location
- Name
- Screen Name
- Time Zone
- Geo Enabled
- Latitude
- Longitude
- Tweet date and time
- Tweet text



- 47 million geo-tagged tweets (Sept 2012 – March 2013)



- 8 million geo-tagged tweets (Sept 2012 – Feb 2013)

Analysing Names on Twitter

- A family name is a statement of the bearer's cultural, ethnic, and linguistic identity [4]
- Some examples of NAME variations on Twitter

Real Names

Kevin Hodge
Andre Alves
Jose de Franco
Carolina Thomas, Dr.
Prof. Martha Del Val
Fabíola Sanchez Fernandes

Fake Names

Castor 5.
WHAT IS LOVE?
MysticMind
KIRILL_aka_KID
Vanessa
Justin Bieber Home

Classifying Twitter Names to Ethnic Origins

- Applied ONOMAP (www.onomap.org) on forename – surname pairs [5]
- Onomap was created by clustering names of 1 billion individuals around the world



Kevin Hodge (**English**)
Pablo Mateos (**Spanish**)

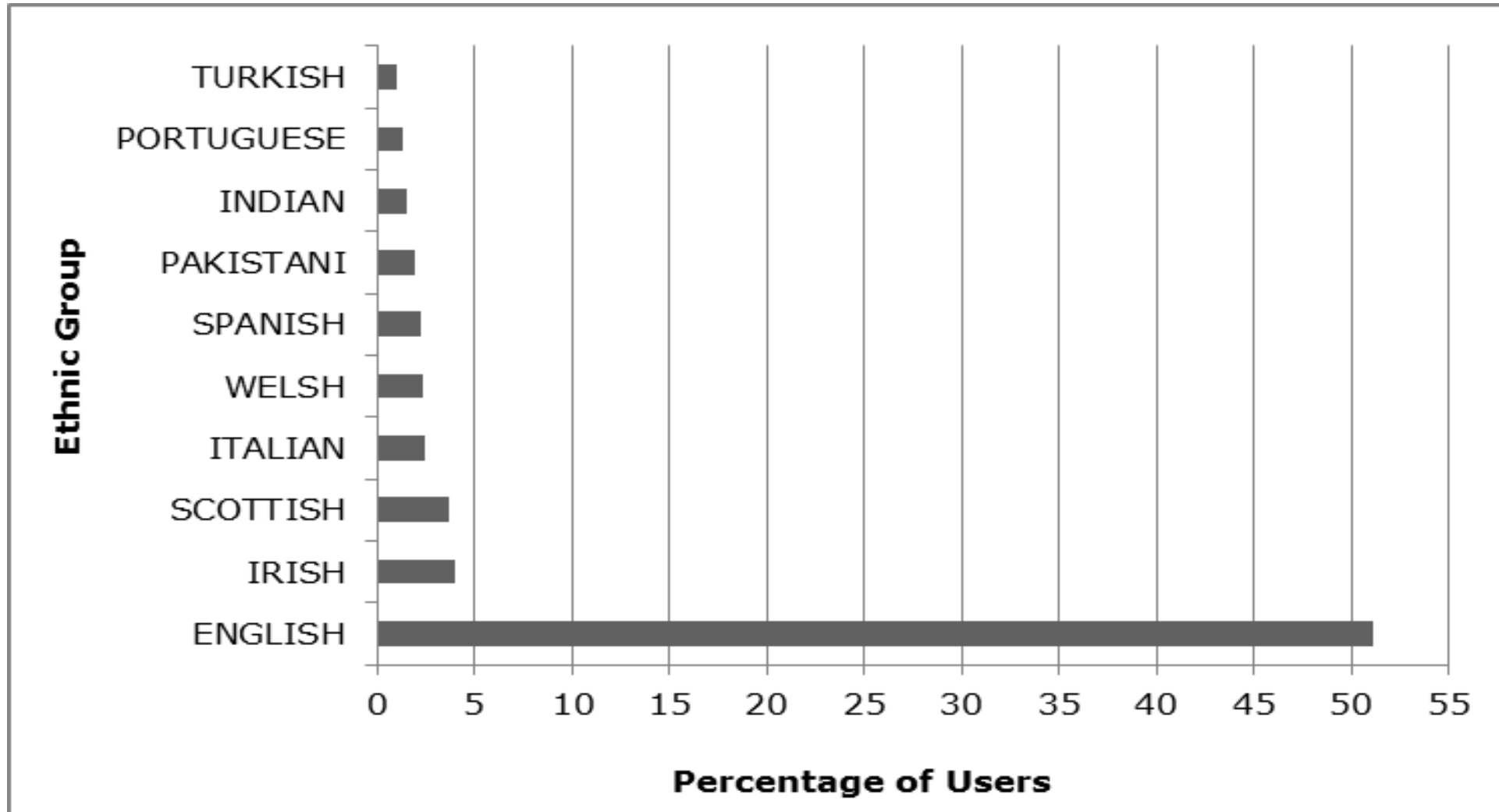
...

...

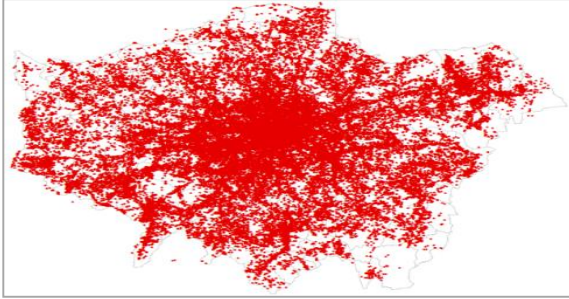
...

...

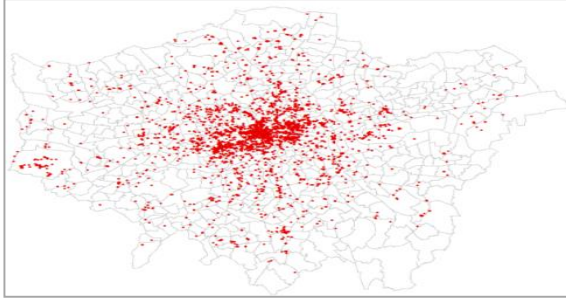
Top 10 Ethnic Groups of Twitter Users



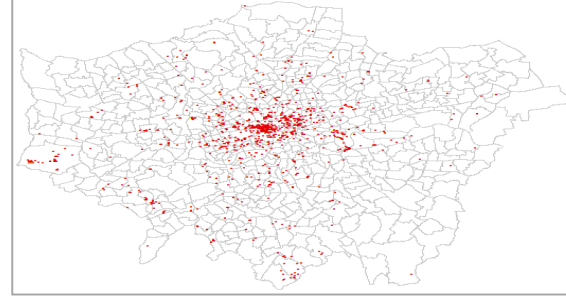
Twitter activity Patterns of various Ethnic Groups



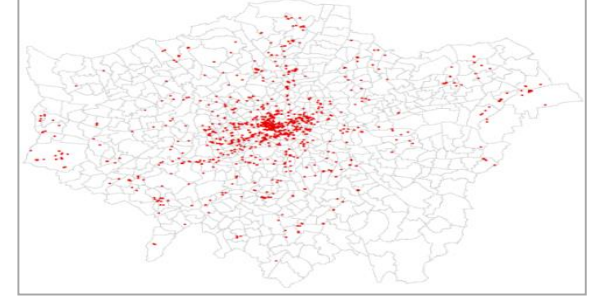
English



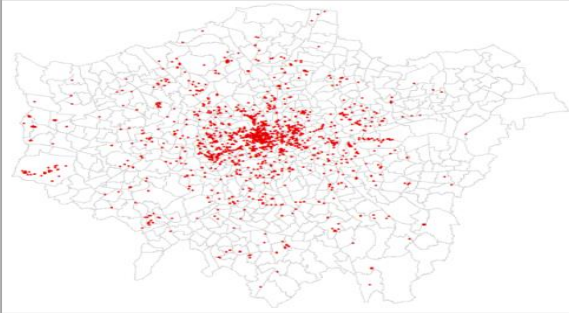
Italian



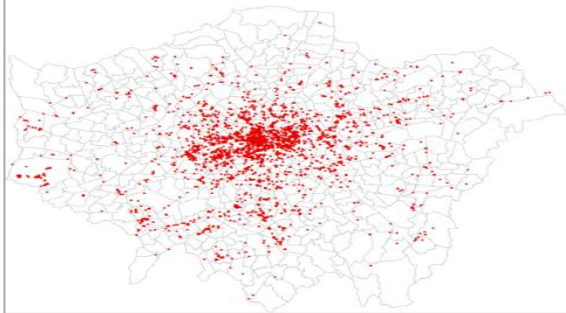
German



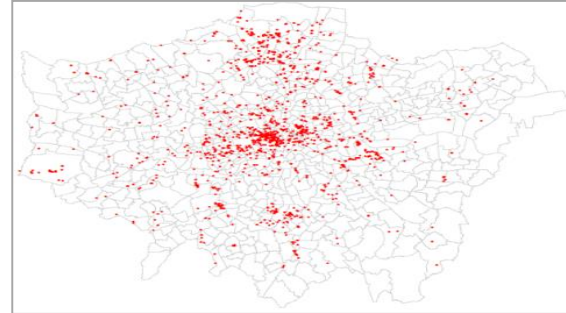
French



Portuguese



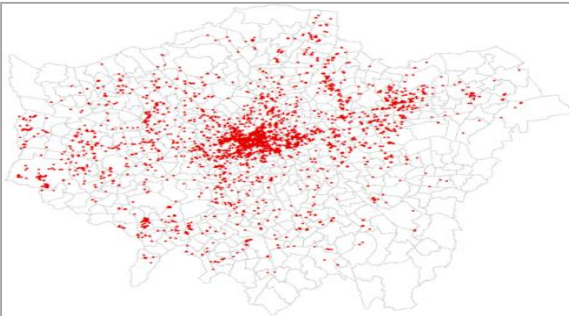
Spanish



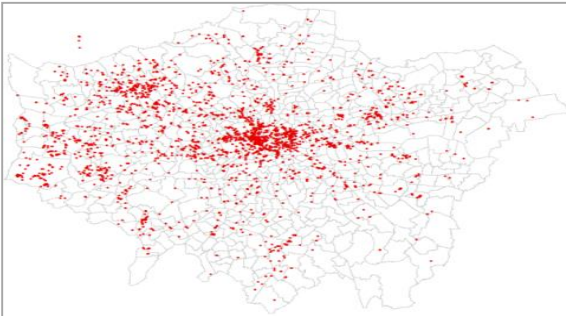
Greek



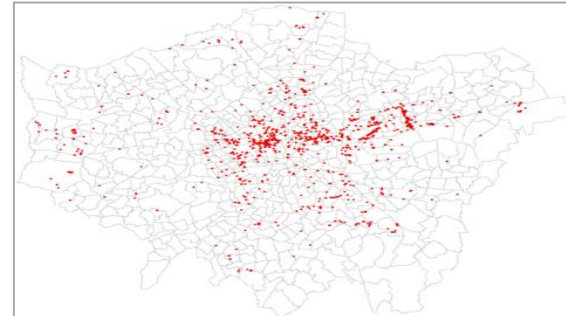
Turkish



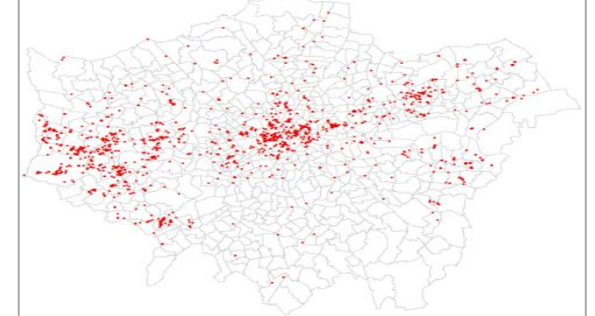
Pakistani



Indian



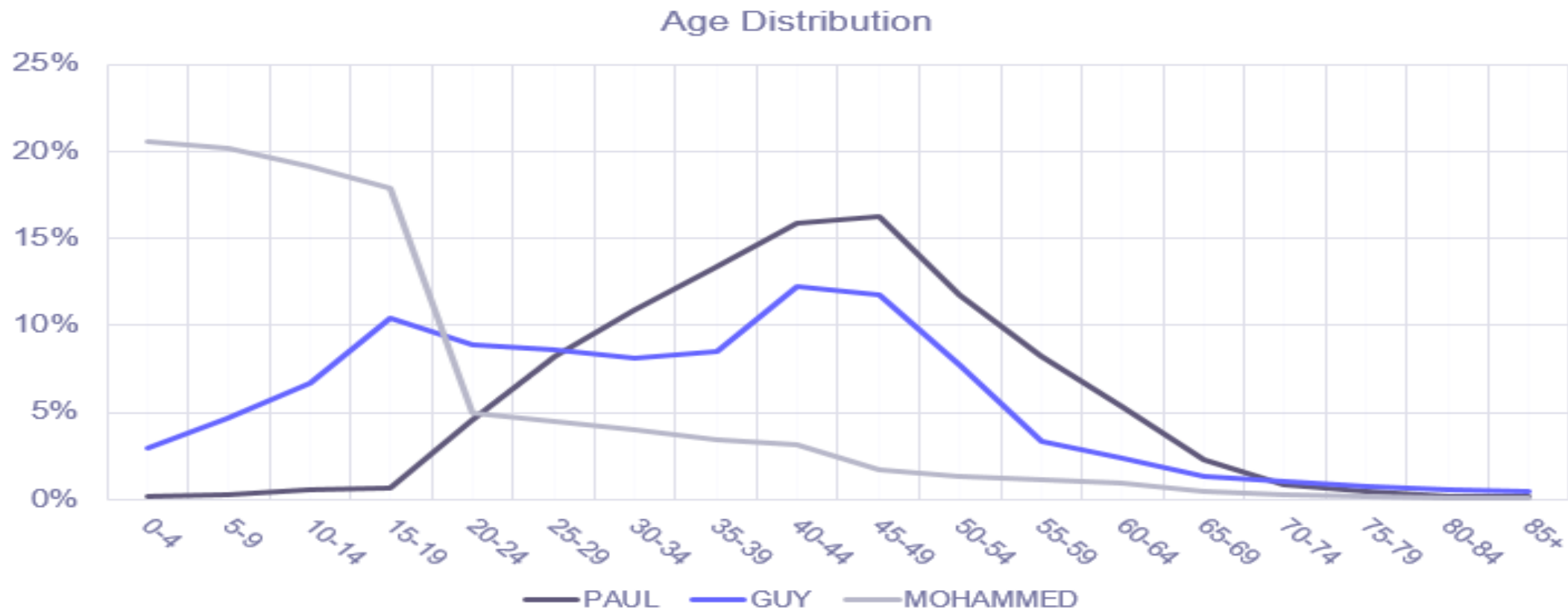
Bangladeshi



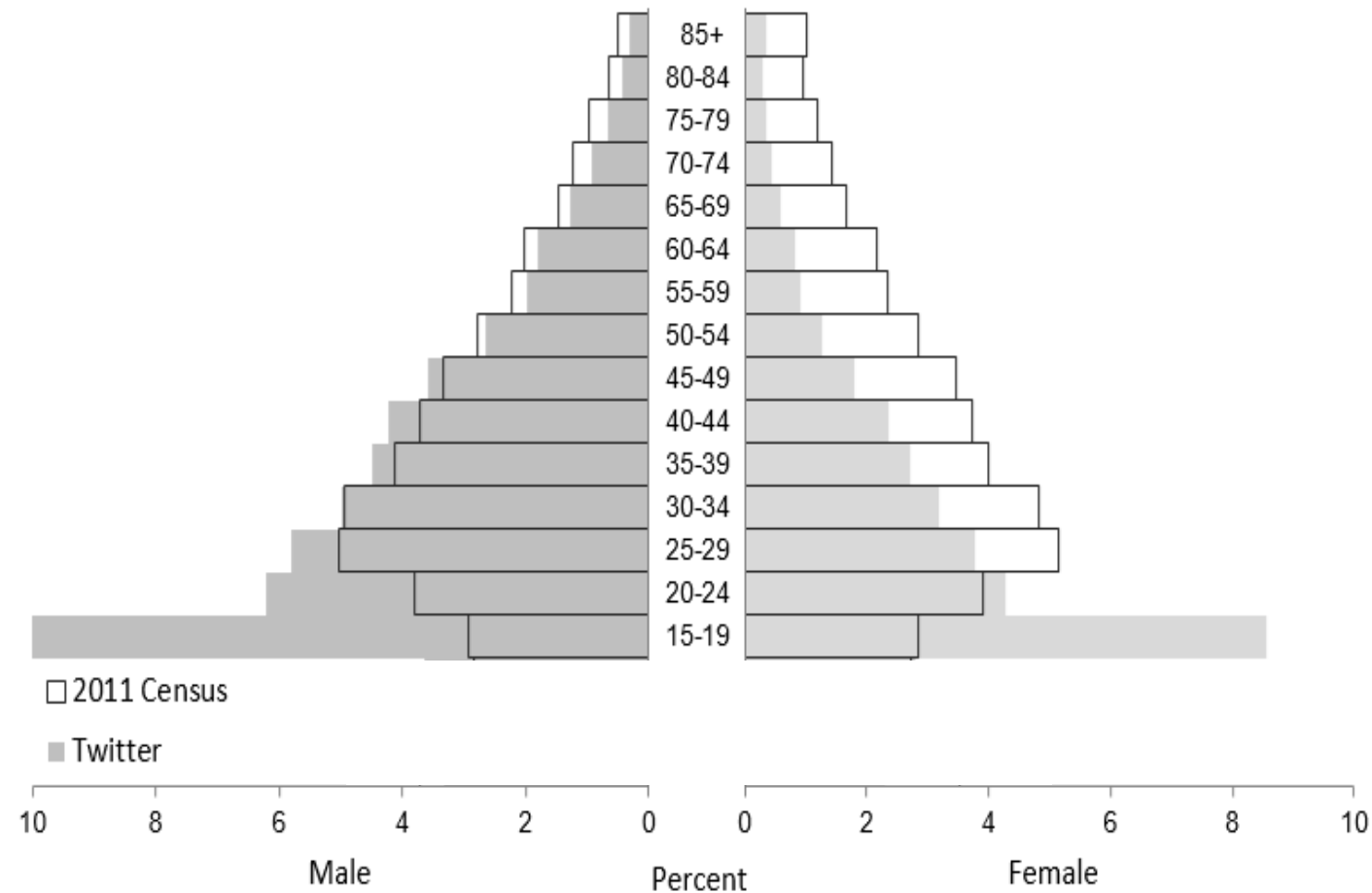
Sikh

Age estimation from 'forenames

- Monica dataset provided by CACI
- Supplemented with birth certificate records



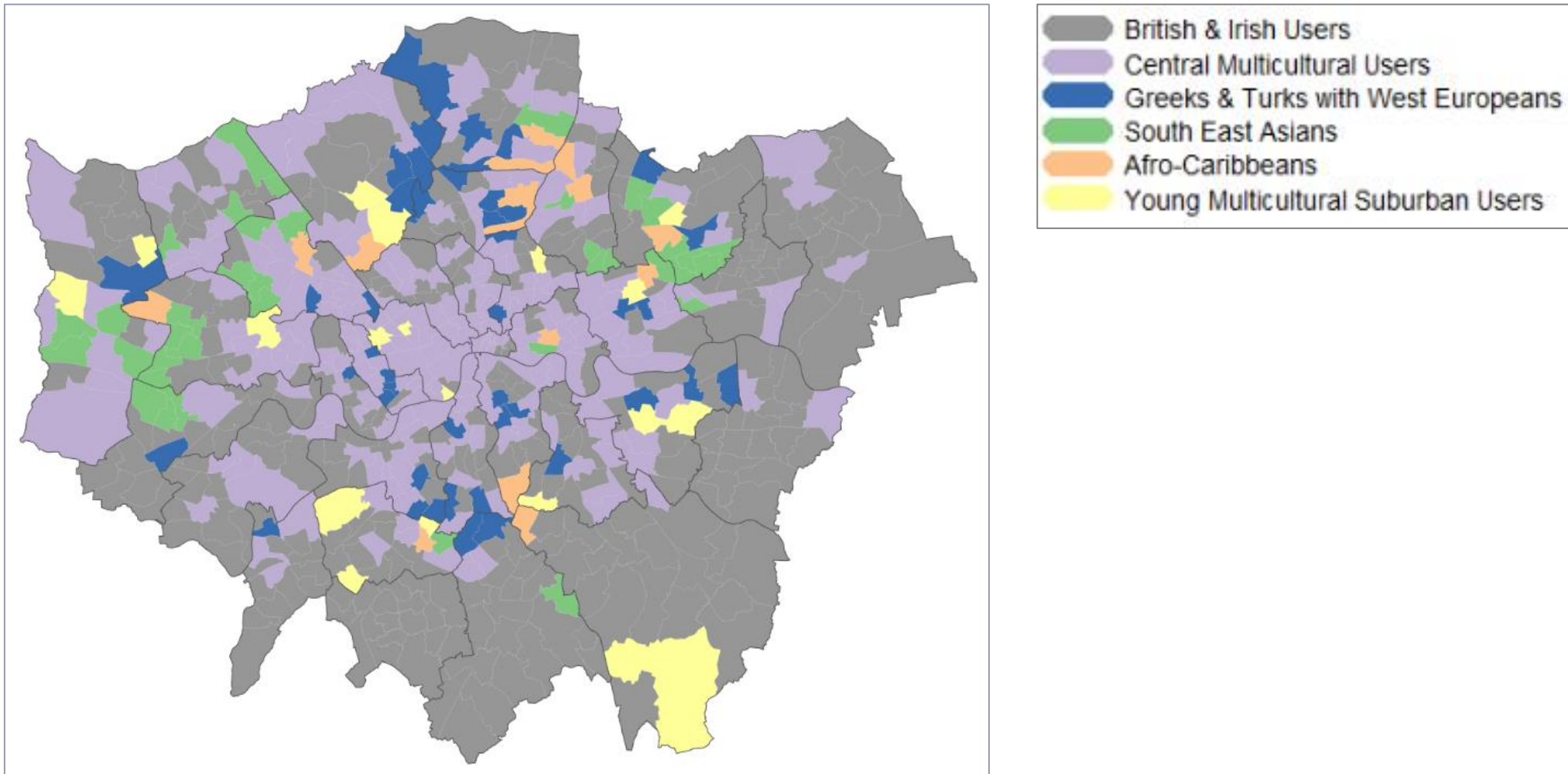
Age distribution of Twitter users



Twitter Users vs. 2011 Census (Greater London)

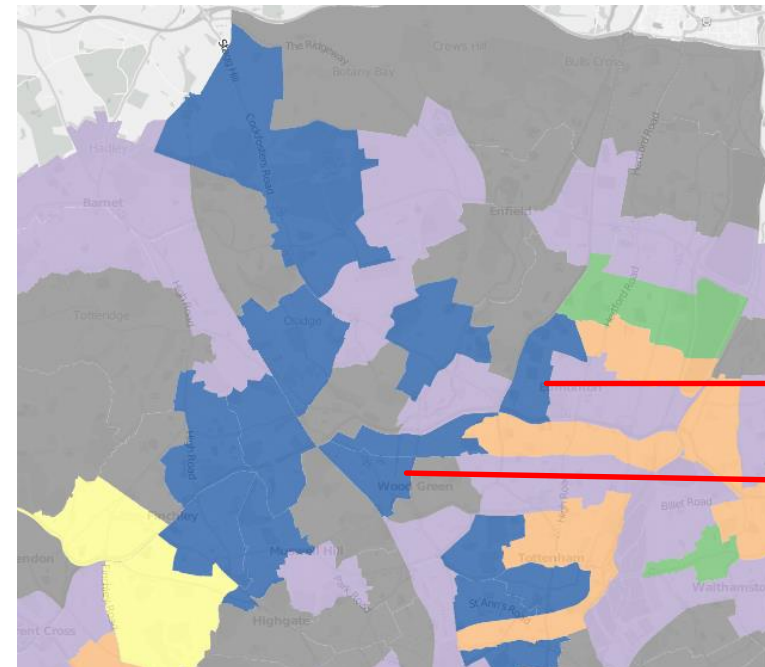
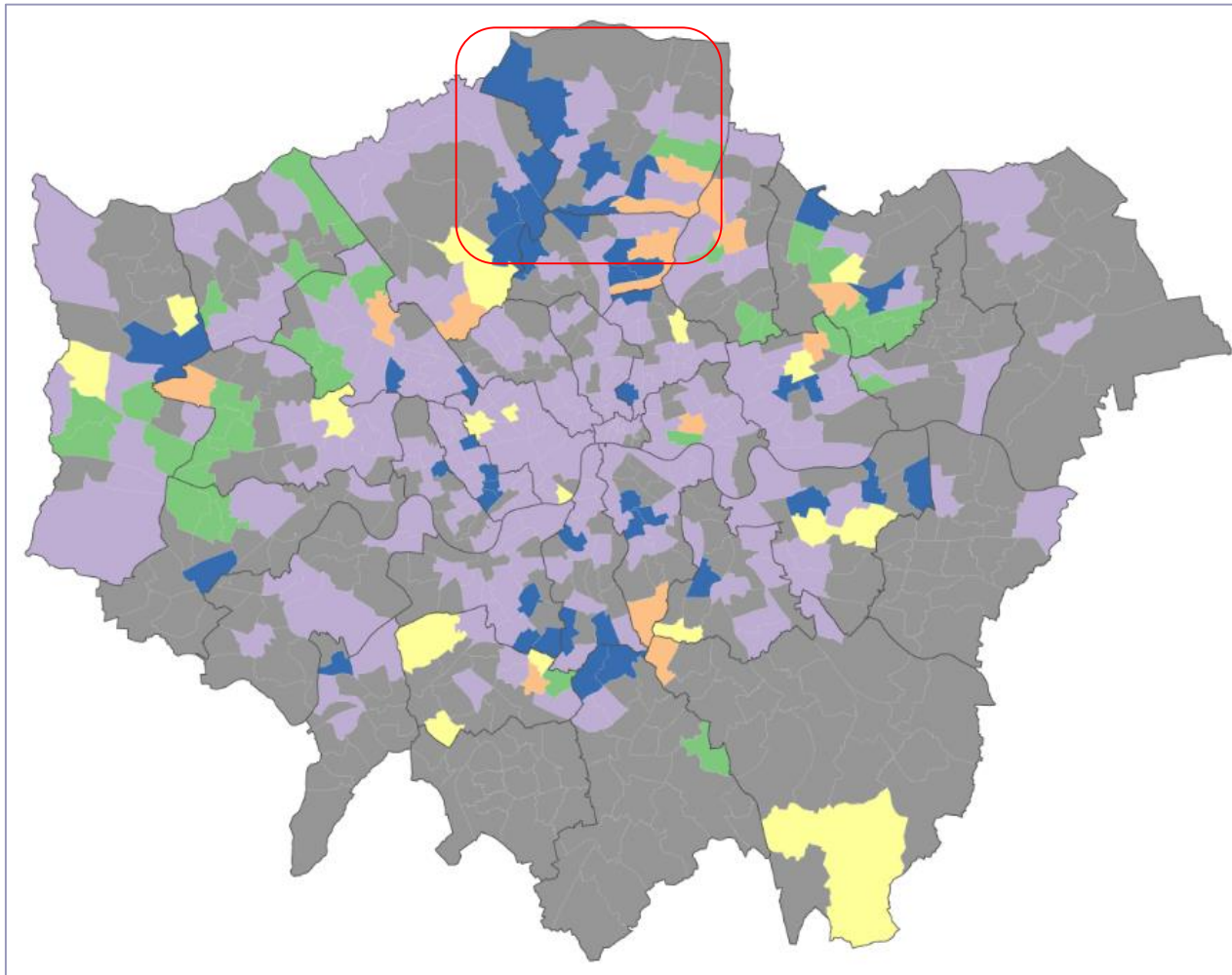
Twitter demographics (Week Days)

- 7 a.m. to 6 p.m. during week days



Twitter demographics (Week Days)

- 7 a.m. to 6 p.m. during week days

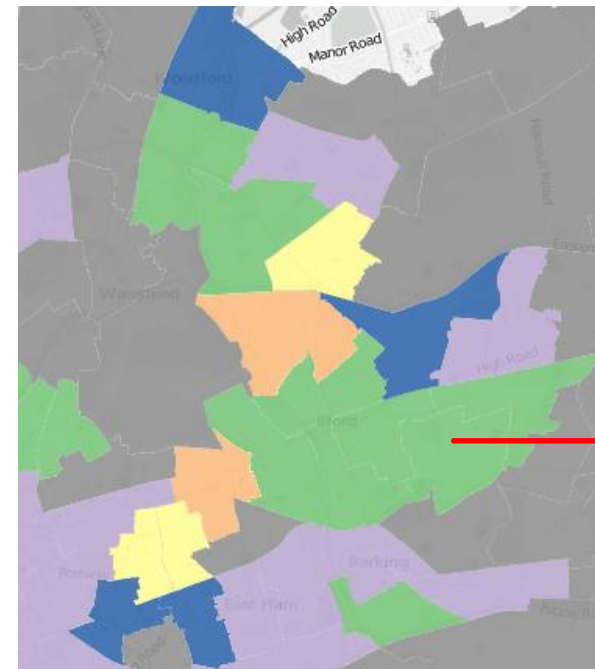
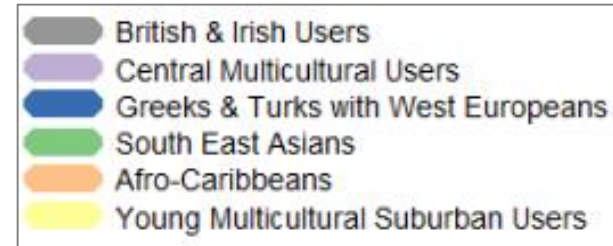
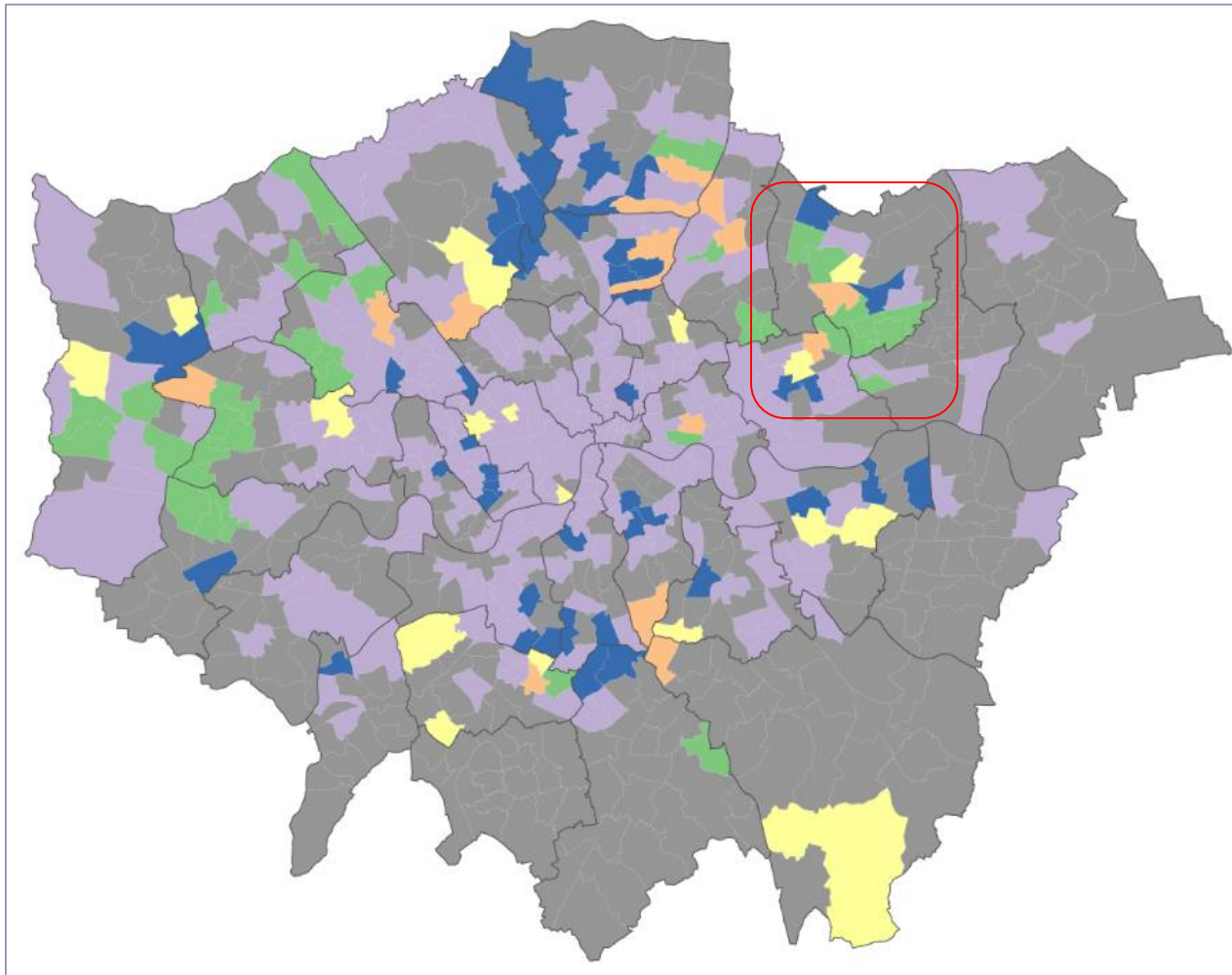


Edmonton

Wood Green

Twitter demographics (Week Days)

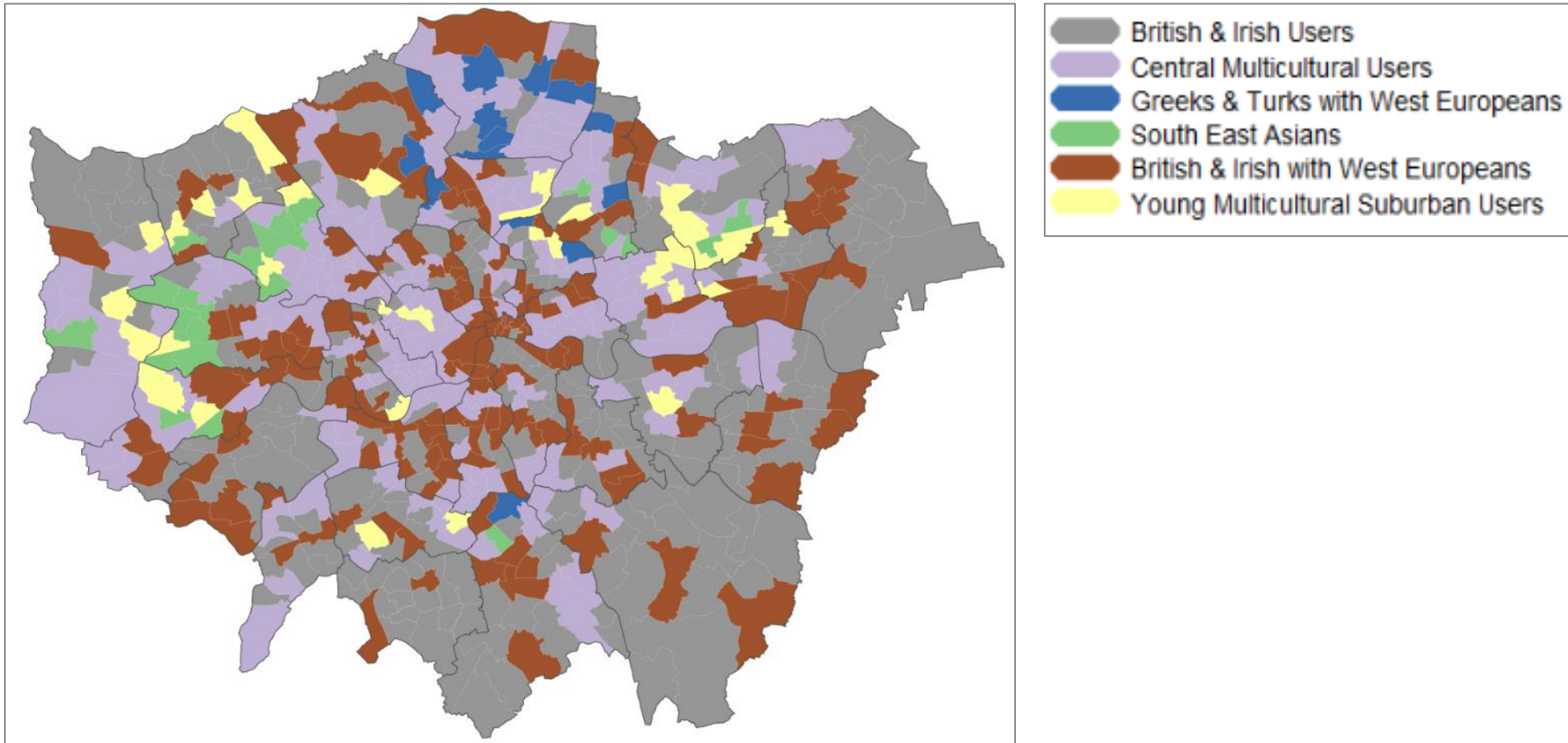
- 7 a.m. to 6 p.m. during week days

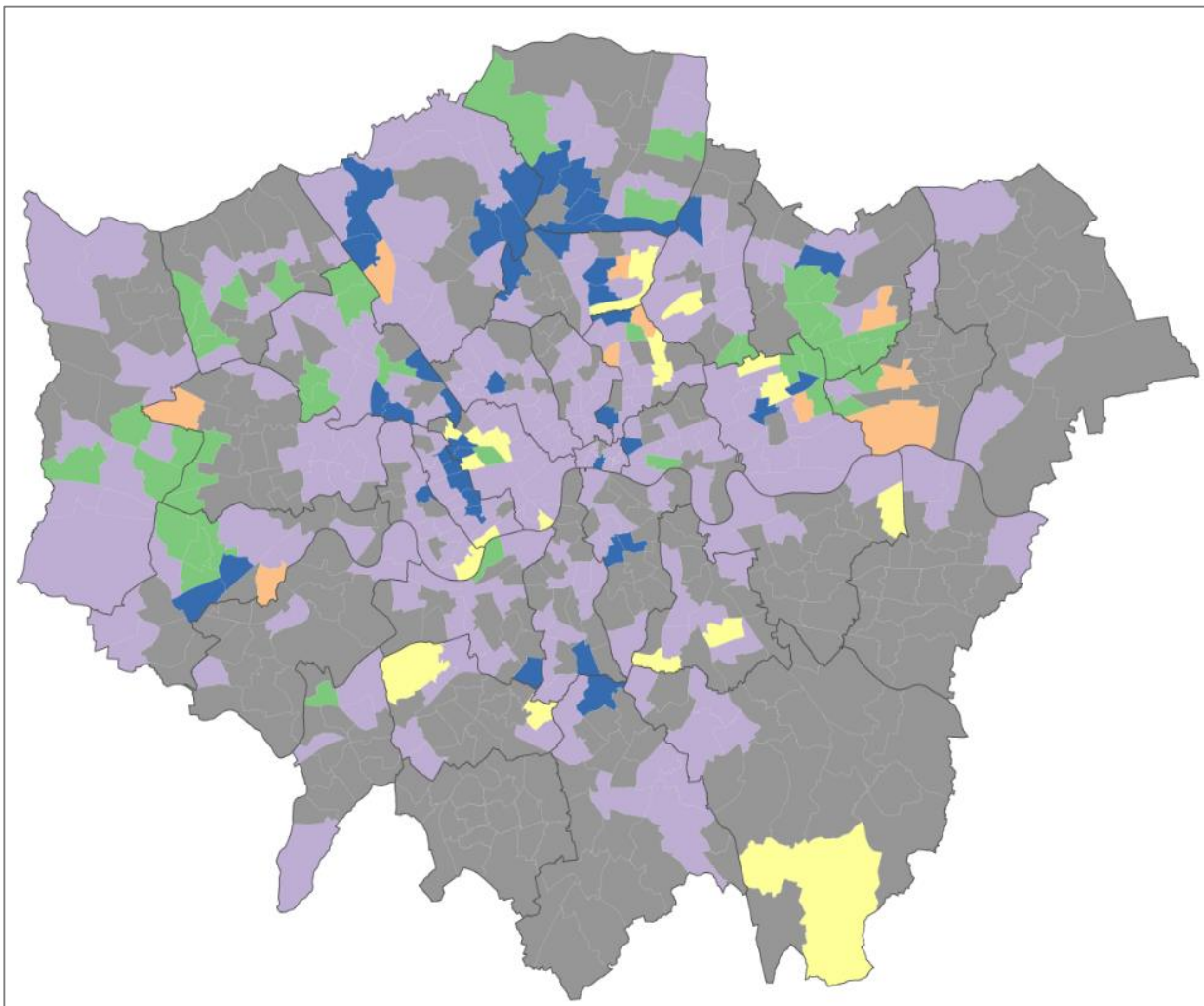


Ilford

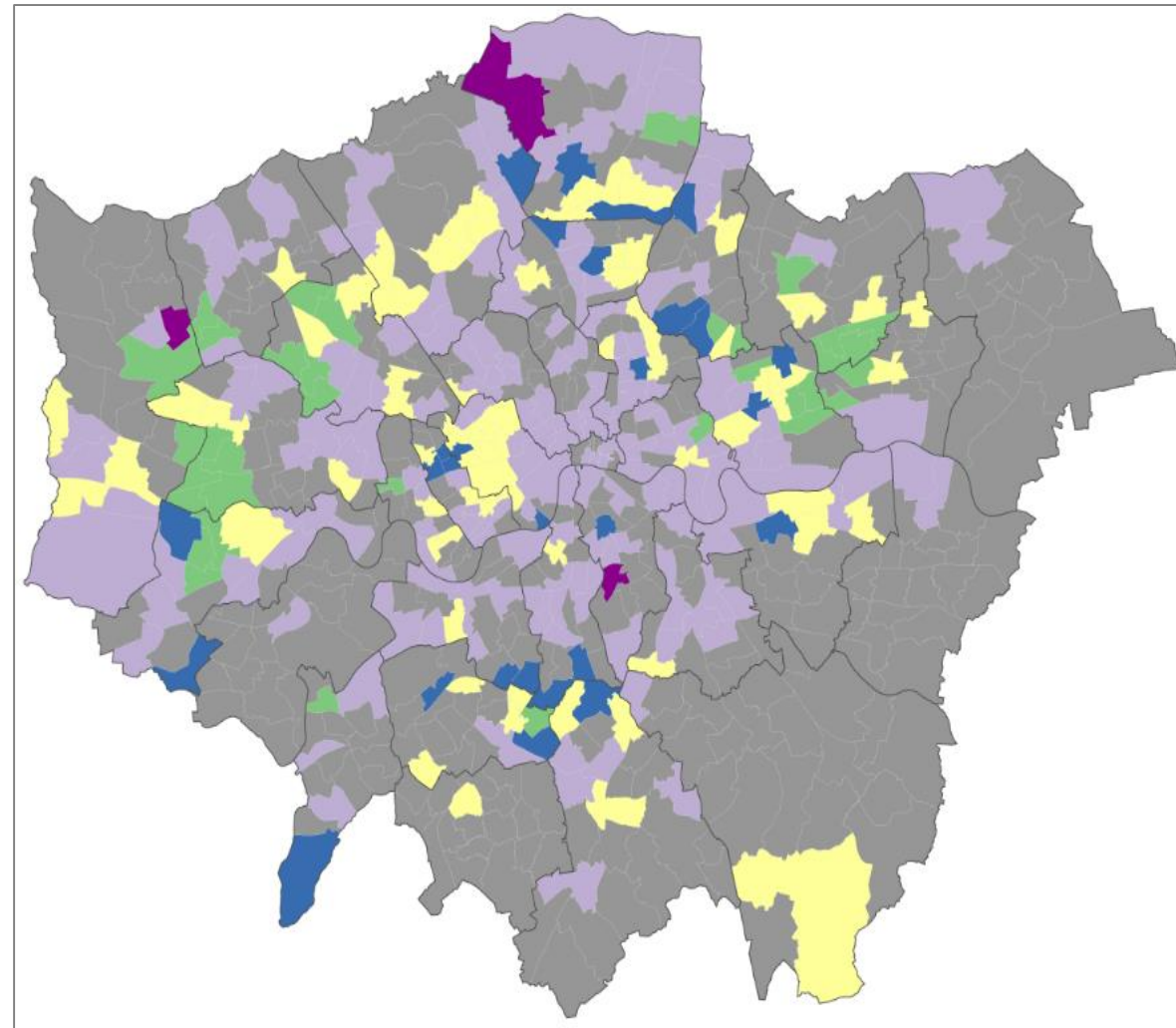
Twitter demographics (Week Evenings)

- 6 p.m. to 11 p.m. during week days





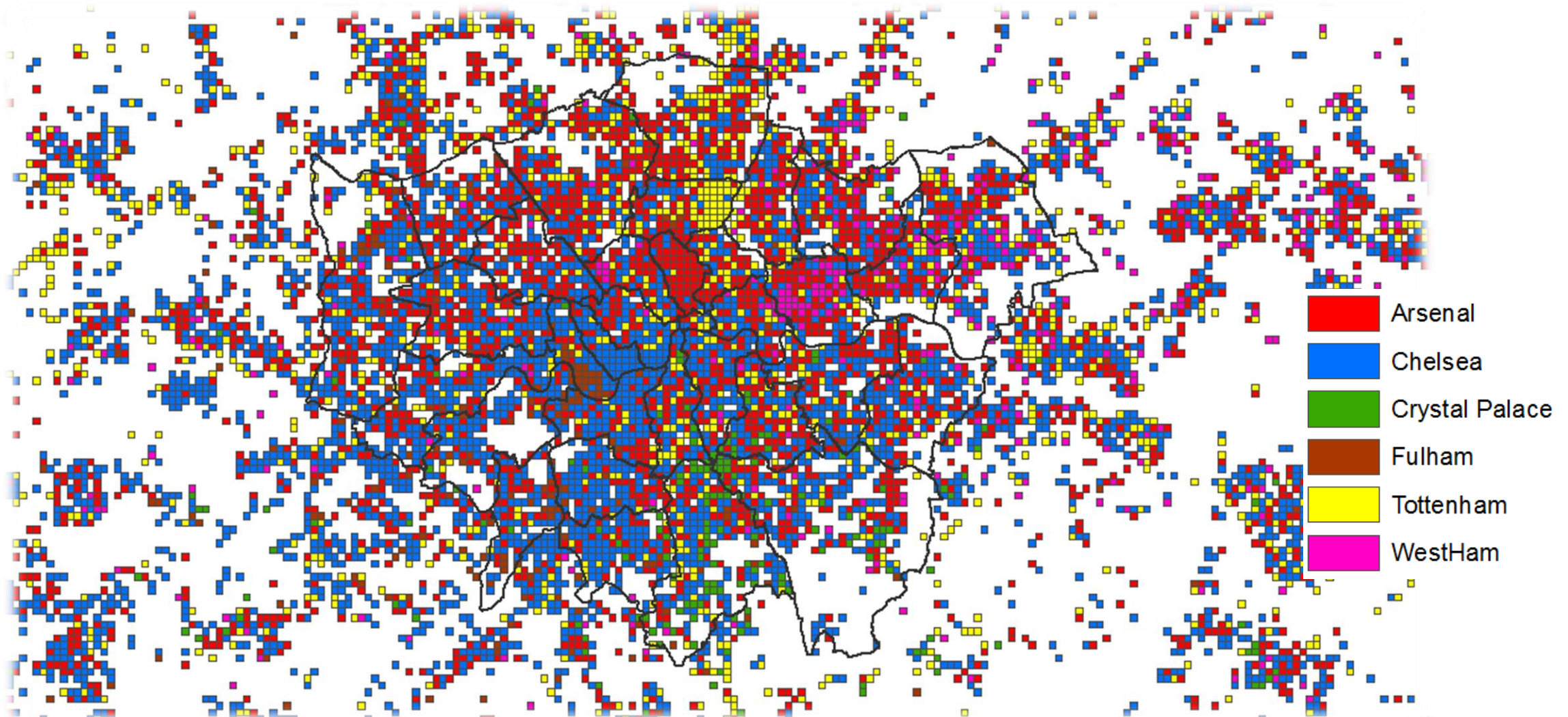
Weekend Days



Weekend Nights

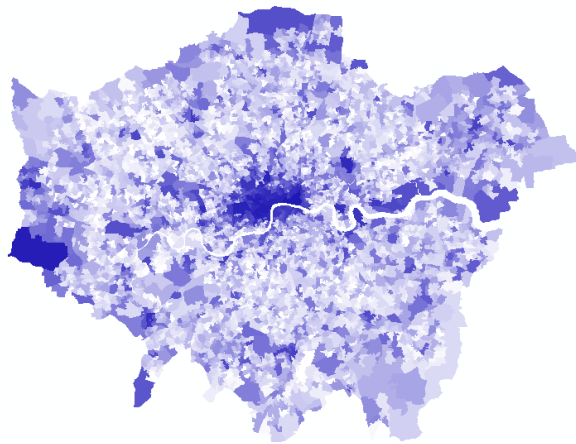


Twitter #Hashtags

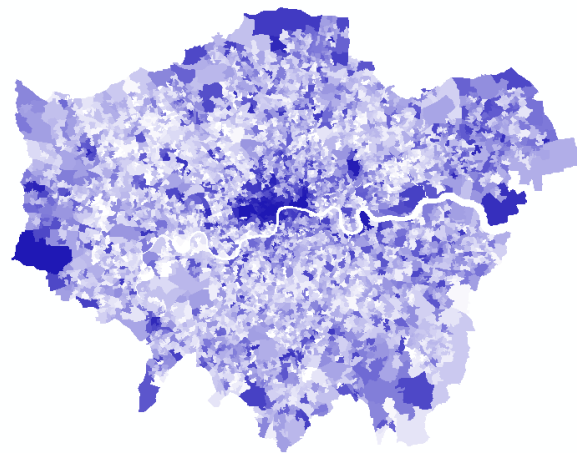


- Most common London football club hashtag

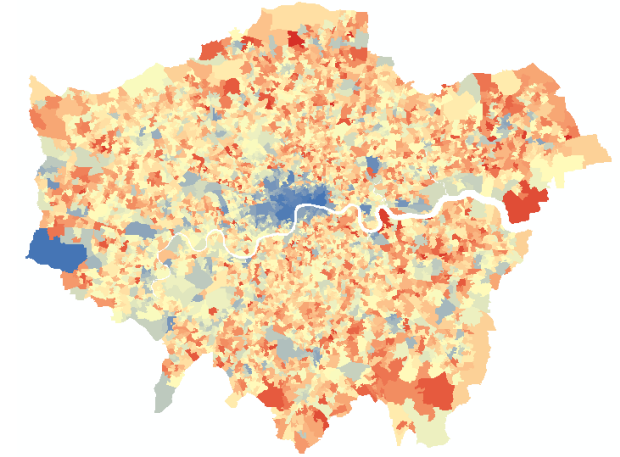
Dispersal of activity (LSOA)



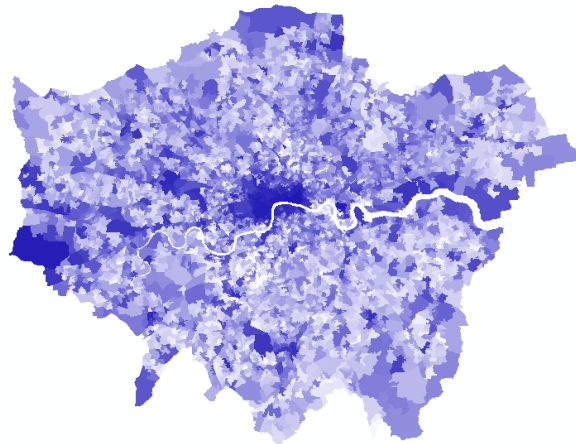
- 10:00 – 16:00 (Twitter)



- 20:00 – 0:00 (Twitter)



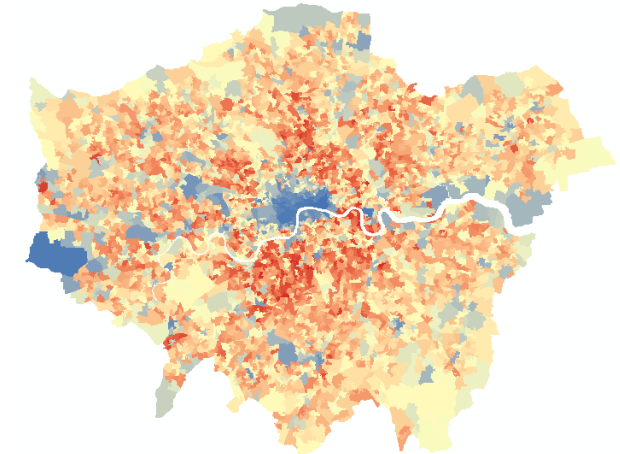
- Difference (Twitter 2013)



- Work day population (2011 Census)

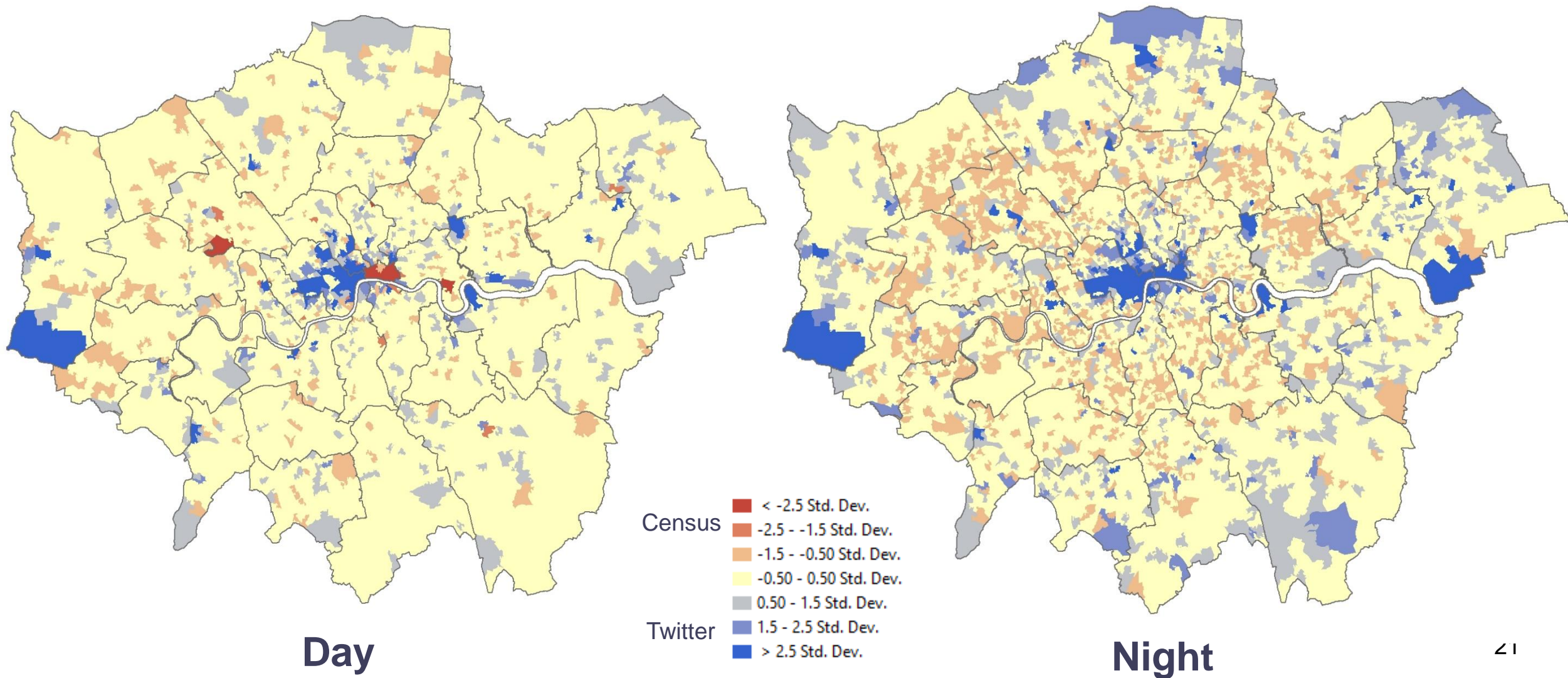


- Residential population aged 16 and above (2011 Census)



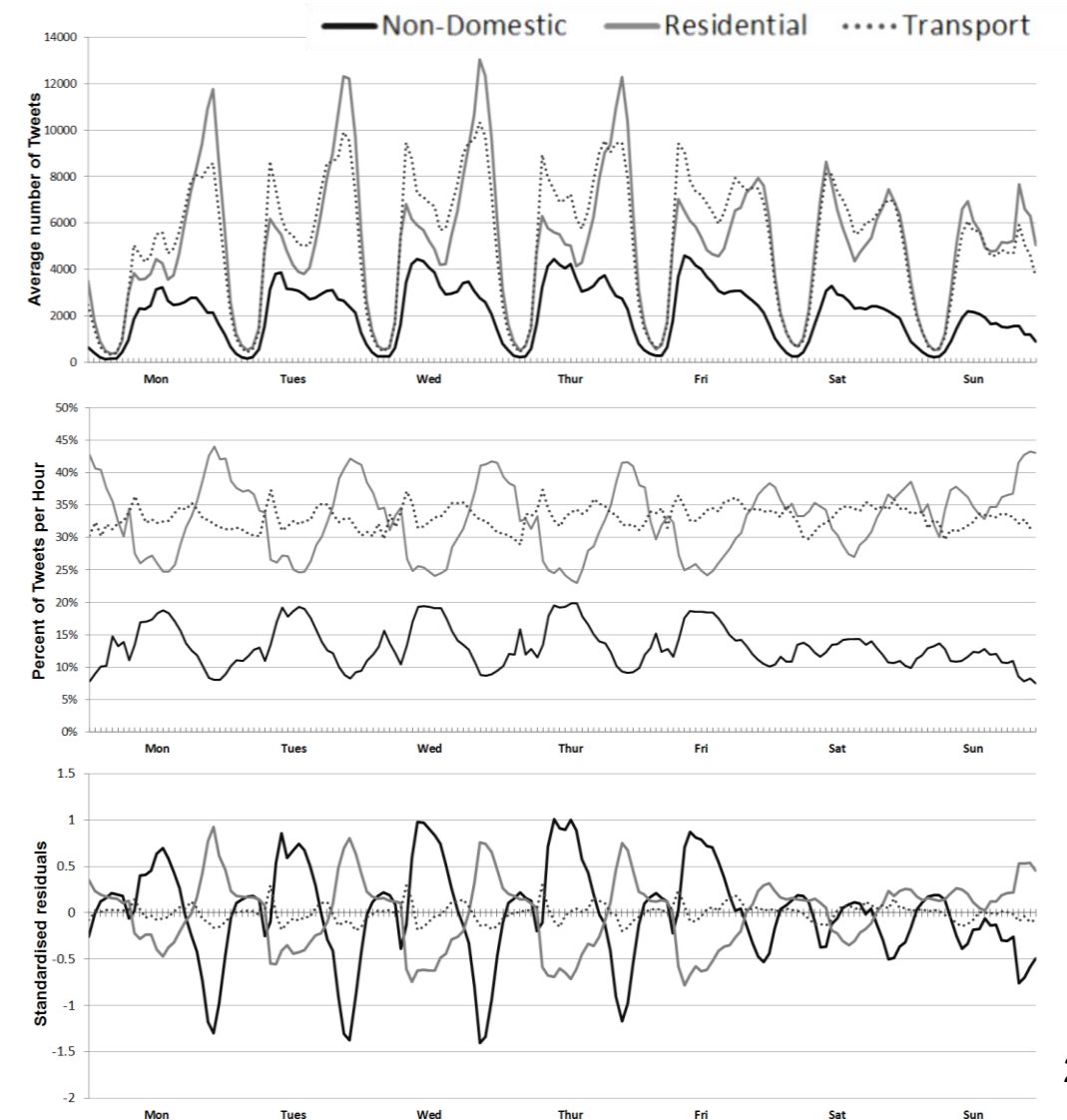
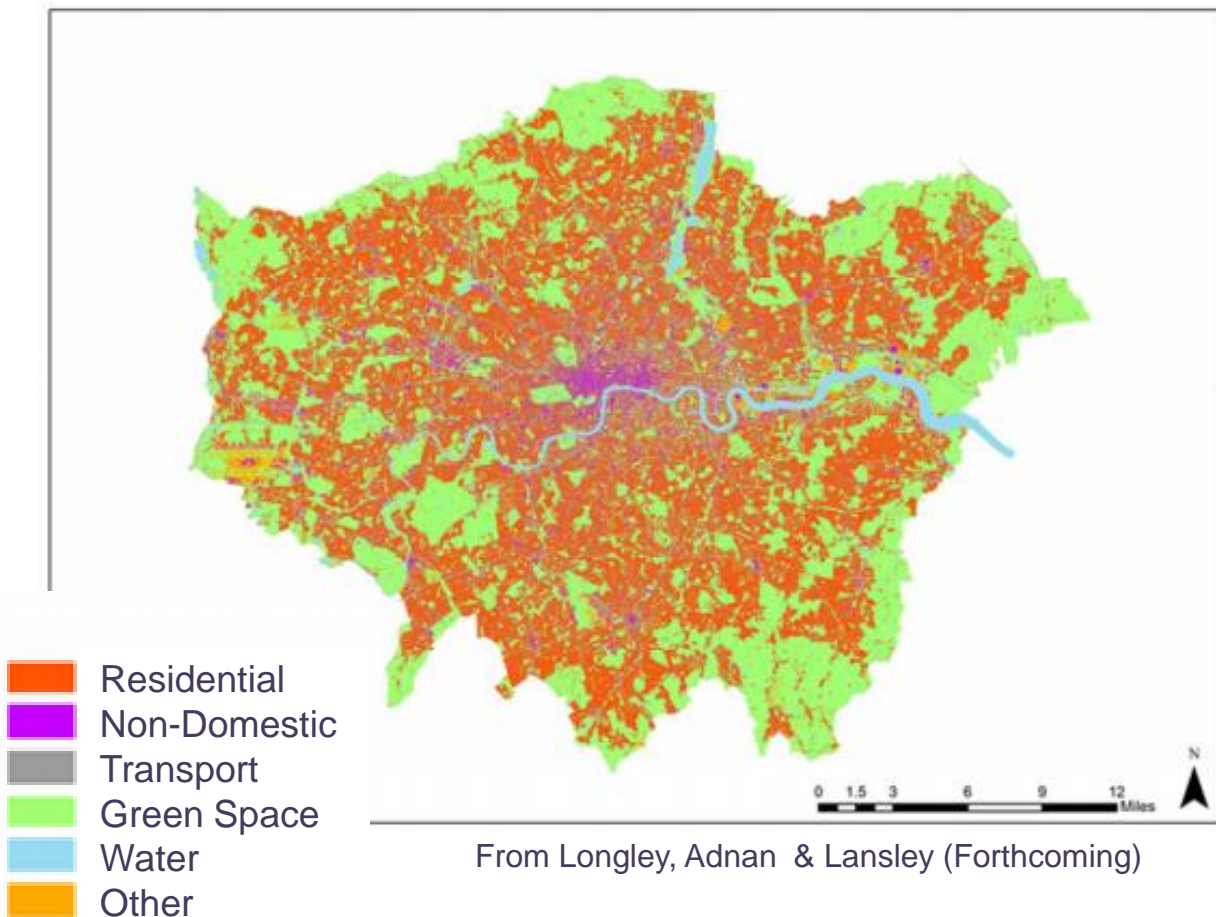
- Difference (Census 2011)

Twitter over-representation



Geo-located social network data

- Twitter activity by land use category
 - Generalised Land Use Database

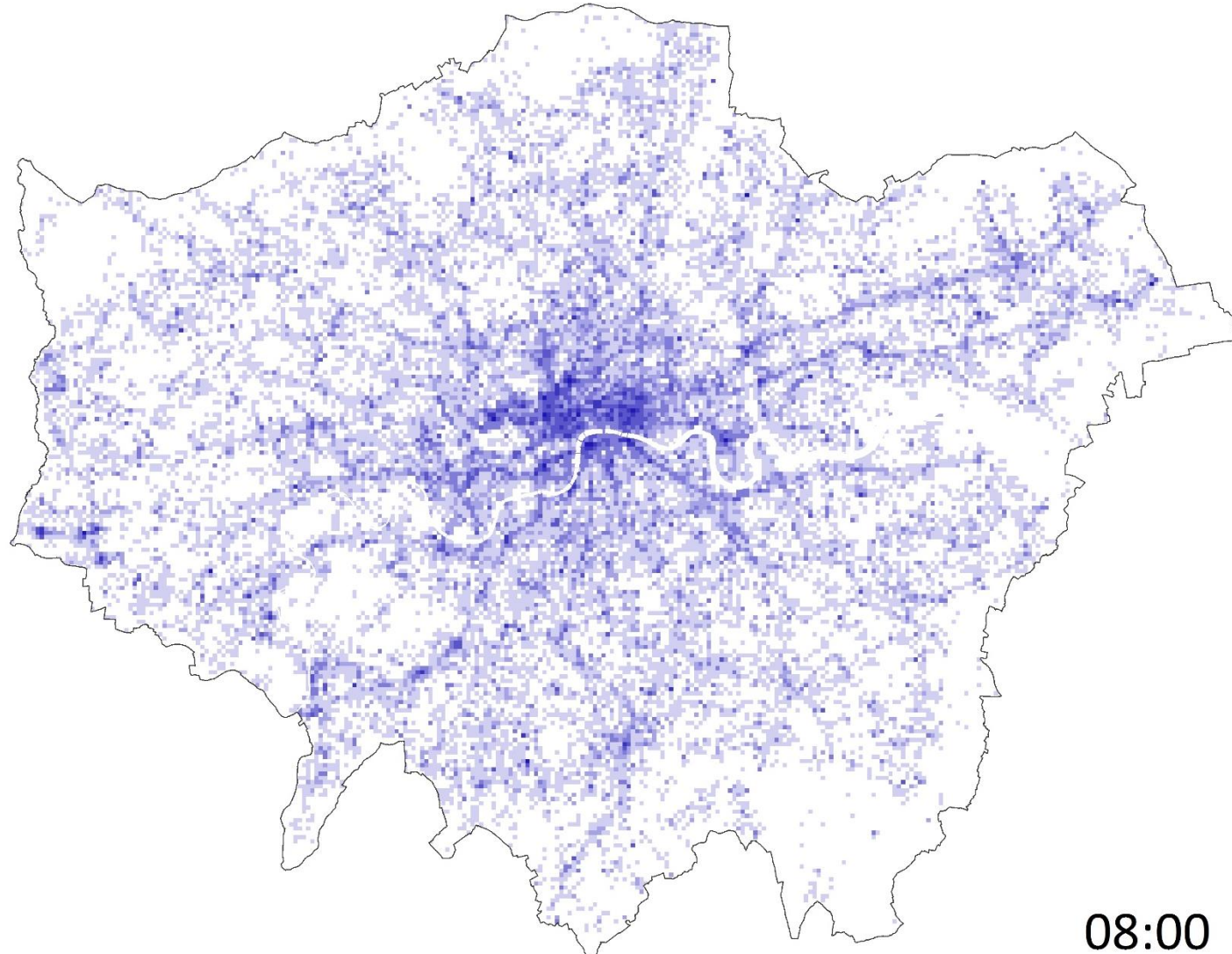


Interactive map of Twitter density (UK and Ireland)

- www.uncertaintyofindentity.com
- Online in a couple of weeks
- A display of aggregate geo-located tweets from a random selection of Tuesdays, Wednesdays and Thursdays throughout 2013
- 200x200m Grid
- Switch between hours



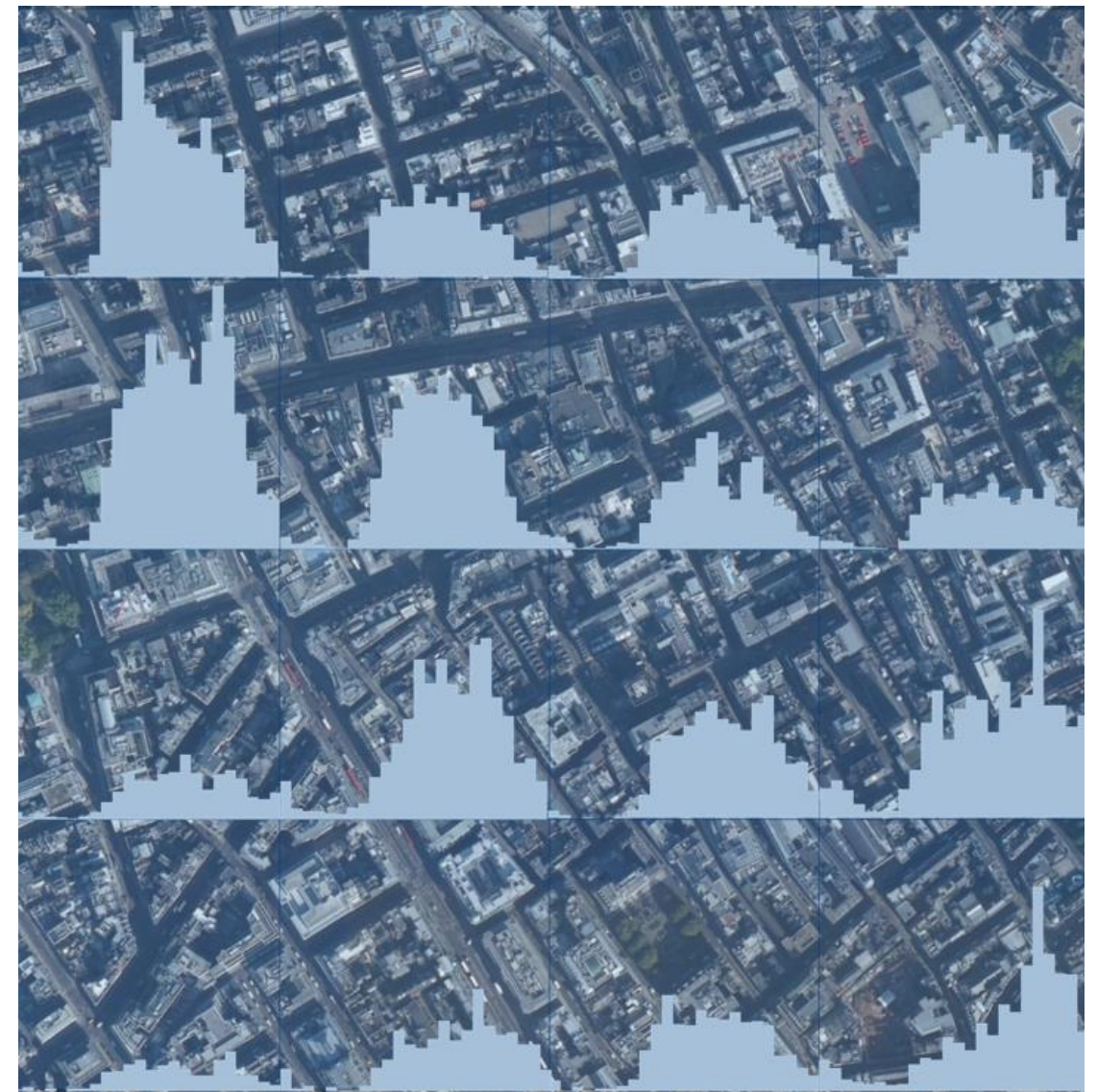
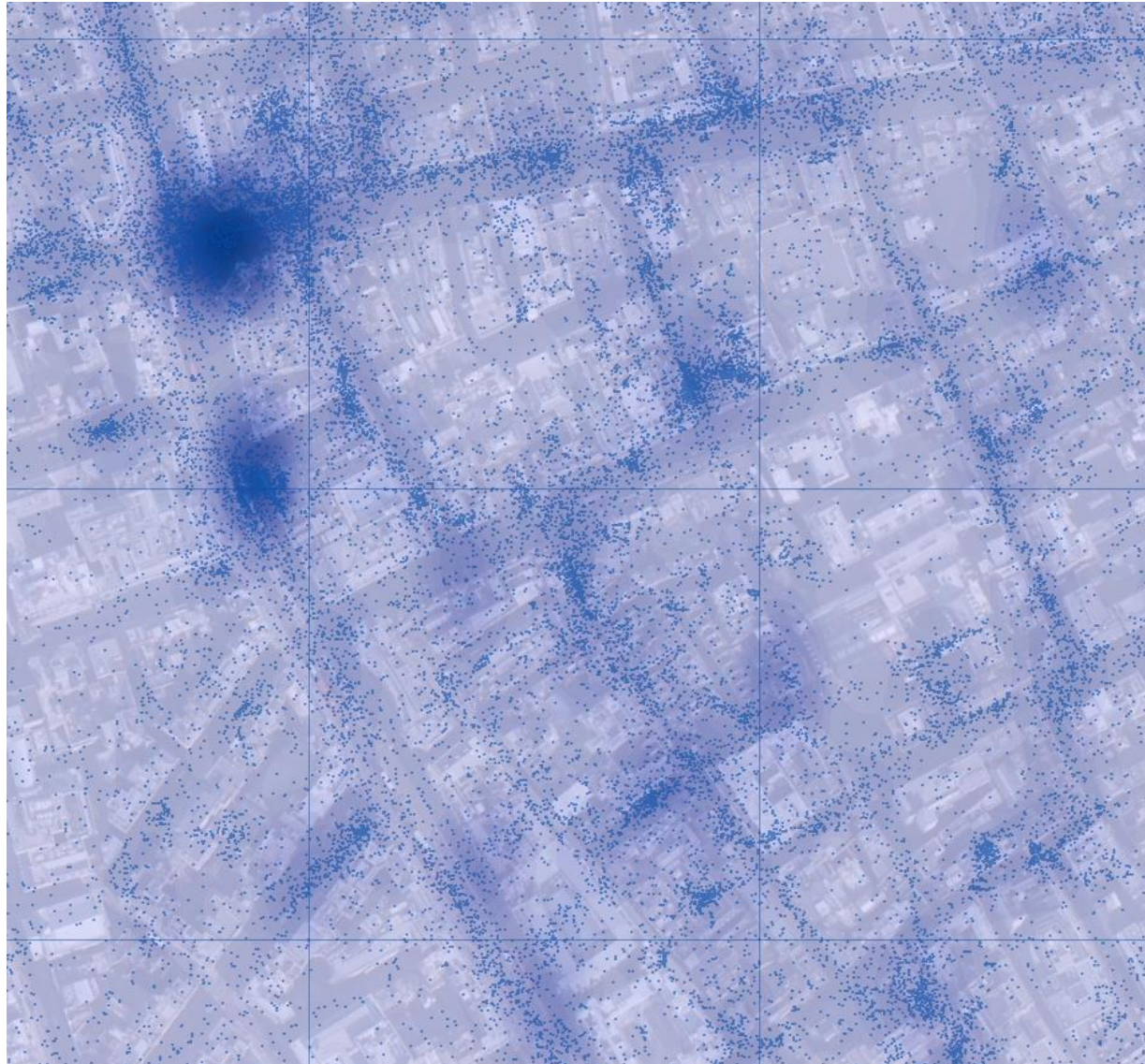
Twitter: Weekday activity in London



08:00

Temporal map of geo-located Tweets recorded on selected weekdays during the winter of 2012/13

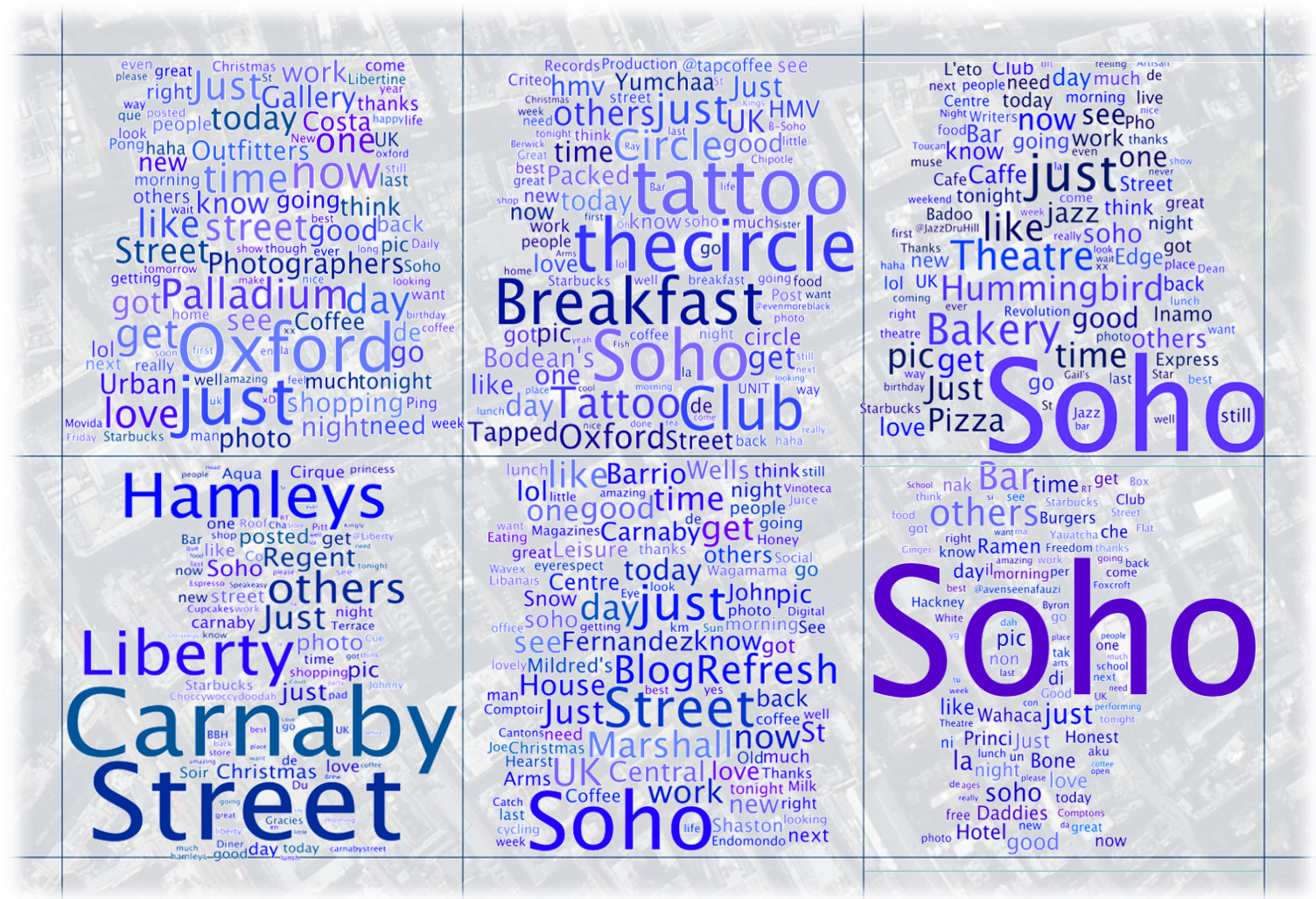




00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00

Tweet content

- Content reflects
 - Place
 - Land use
 - Activity
 - Sentiment
 - Language
- Content also reflect time and date
- Words can be aggregated to make a definitive classification of topics

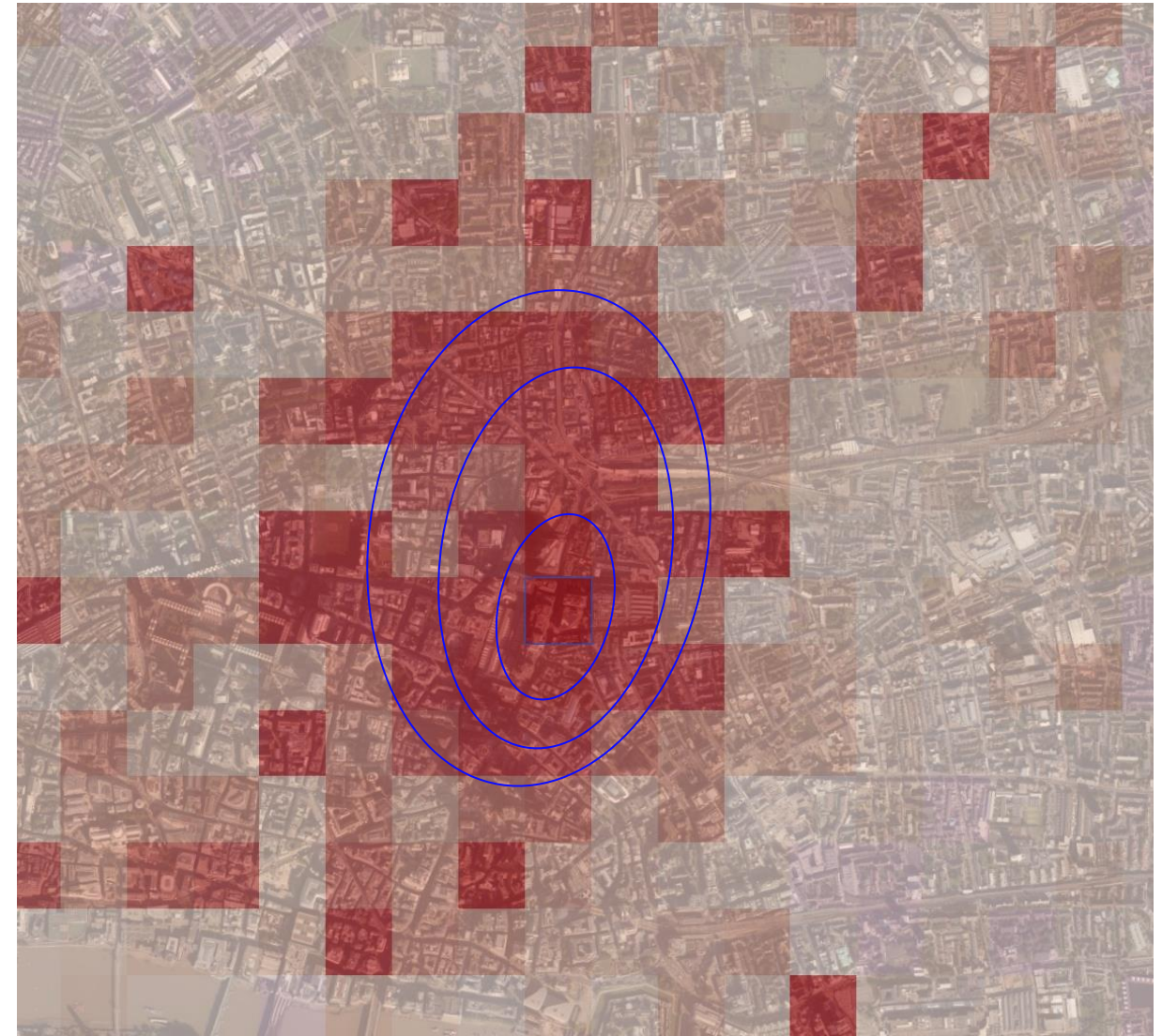
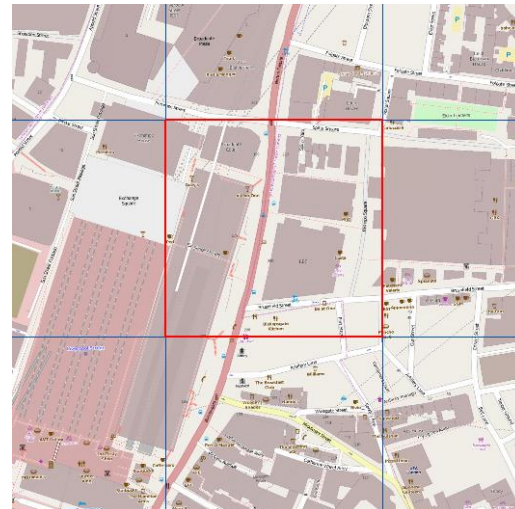
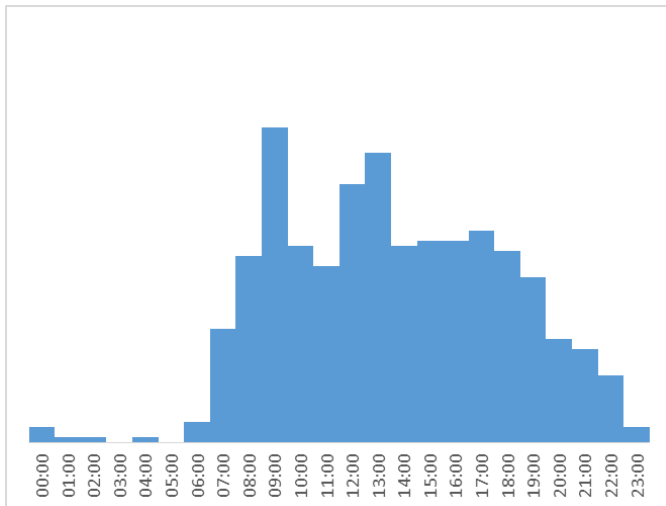


Note: the words “Greater” and “London” have been removed

Retail	Nightlife	Eating out	Entertainment	Outdoor	Tourism	Transport	Work	Home
140	170	156	103	73	163	63	45	23

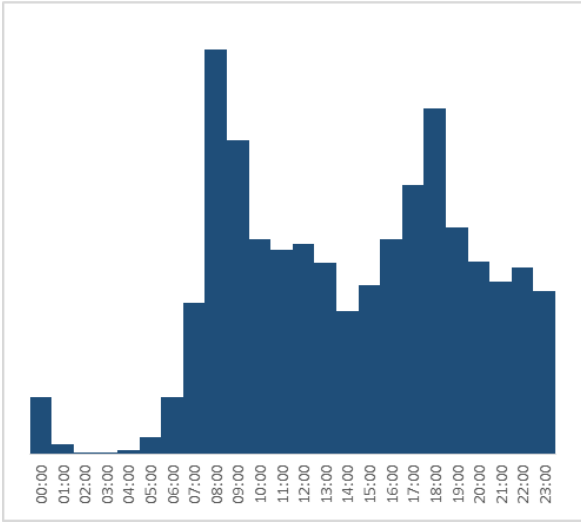
Specific time catchments

- E.g. Day-time catchment
 1. Identify the unique ID of users frequently transmitting from a particular location at a given time or date range
 2. Request their other activity through Twitter's API, filter by time/date
 3. Aggregate

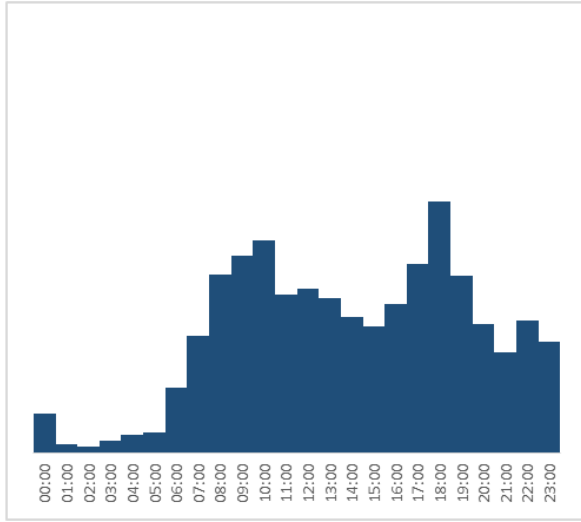


The Twitter work-day time catchment of Bishopsgate

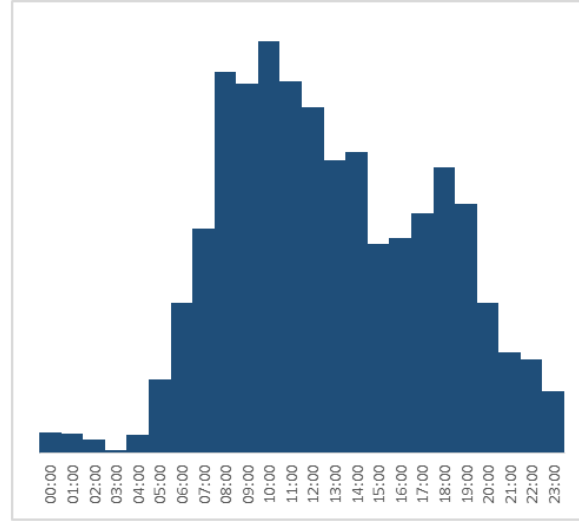
Waterloo



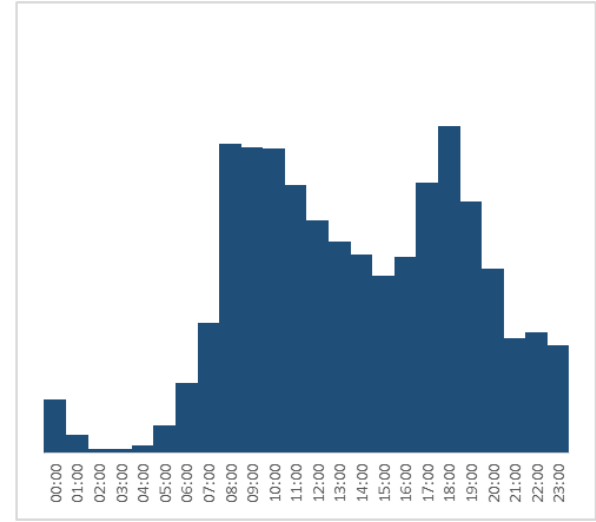
Victoria



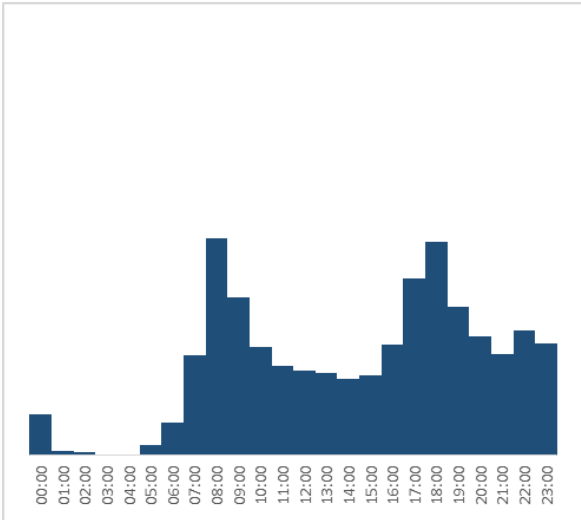
St Pancras



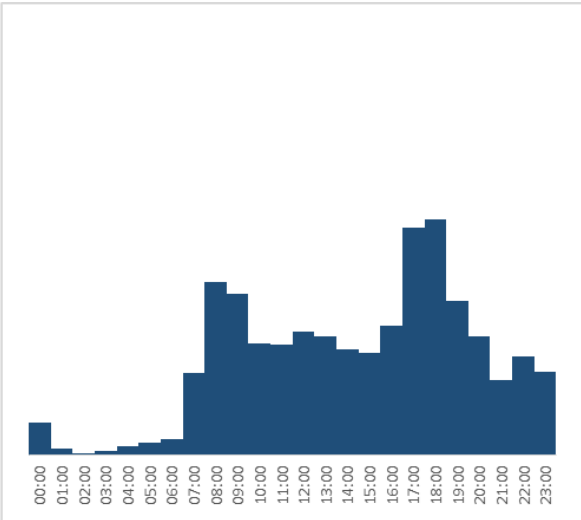
Paddington



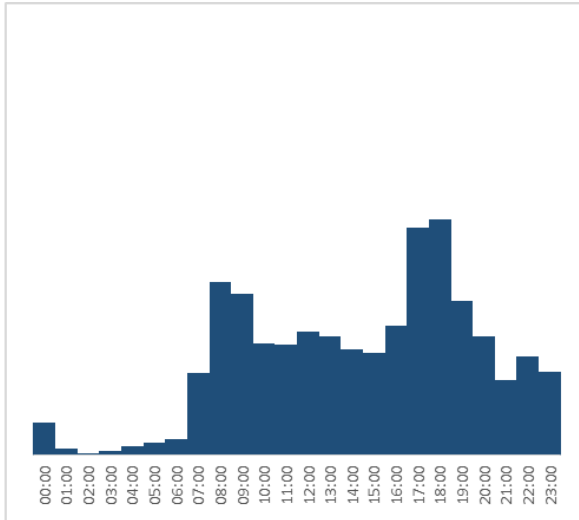
London Bridge



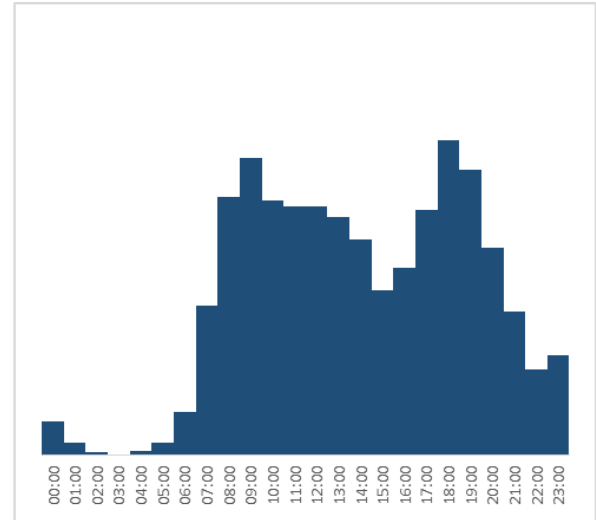
Liverpool Street



Kings Cross

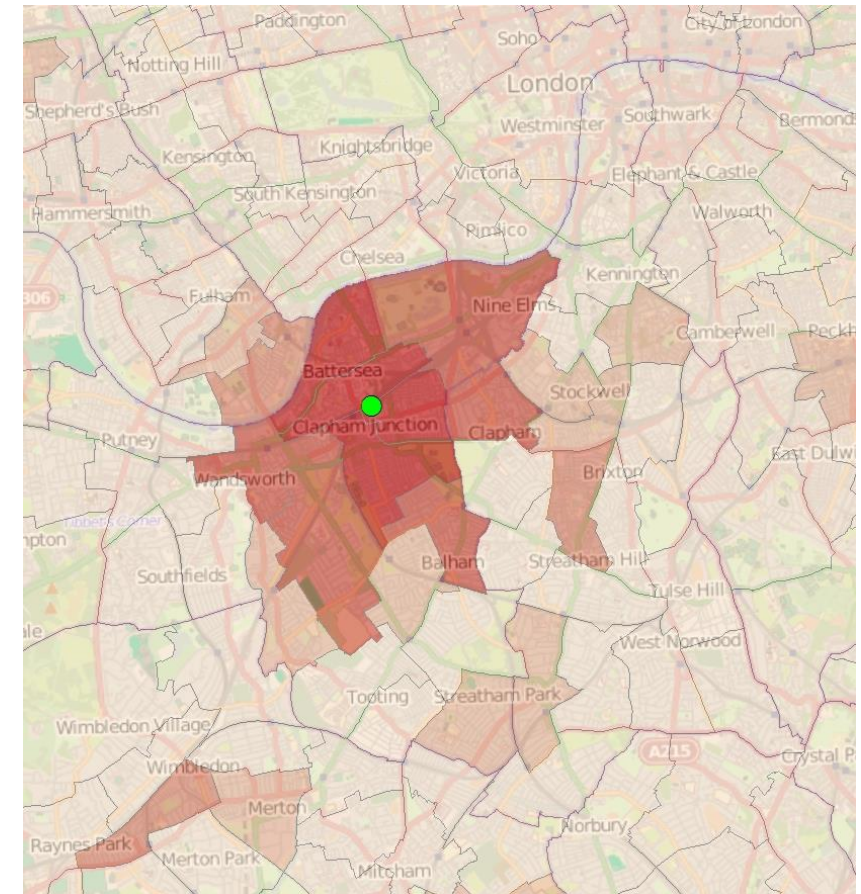
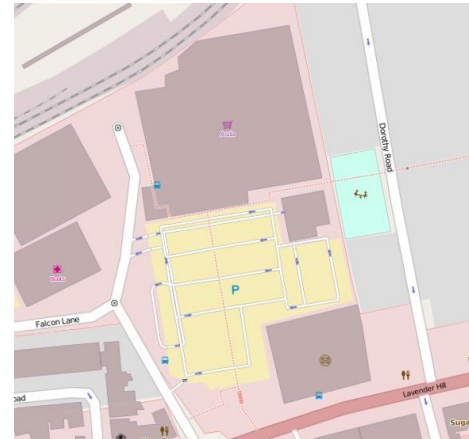


Euston



Inferring a residential catchment based on Twitter data

- First, extract the unique ID's of users have tweeted from inside the building
- Request these users' other Tweets for a given time/date range
- Create a customer catchment by identifying all Tweets sent from domestic land uses at a given time
- E.g. ASDA in Clapham Junction



The Twitter residential catchment of ASDA Supermarket at Clapham Junction

Limitations of Twitter Data

- Twitter users are not a representative sample of the British population
- Sample size
- Precision of geo-location varies between handheld devices
- Signal availability
- Tweets do not always reflect the place where they are transmitted
- Demographic characteristics & home address are not recorded
- Ethics

Twitter based MSc dissertations at UCL 2012/13

- **The validity of geo-referenced Twitter data for predicting footfall for retail**
 - Robert Lea (Sponsored by M&S)
- **Investigating the usefulness of Twitter data in characterising and classifying land use within small areas in Greater London**
 - Conrad Ow
- **An exploratory space-time analysis of abnormal Twitter events**
 - Thomas Wicks

Conclusion

- Social media dataset are a good resource for creating geo-temporal Geodemographics
 - Issues of representation
- An insight into the residential and travel geographies of individuals
- An insight into the online behaviours / usage

Thank you for Listening

Any Questions ?