

Online exploration of cultural regions,  
migration and ethnicity using the  
geography of personal names

**Paul Longley**  
**Pablo Mateos**

Department of Geography and CASA  
University College London, UK



CASA Conference  
13<sup>th</sup> April 2010

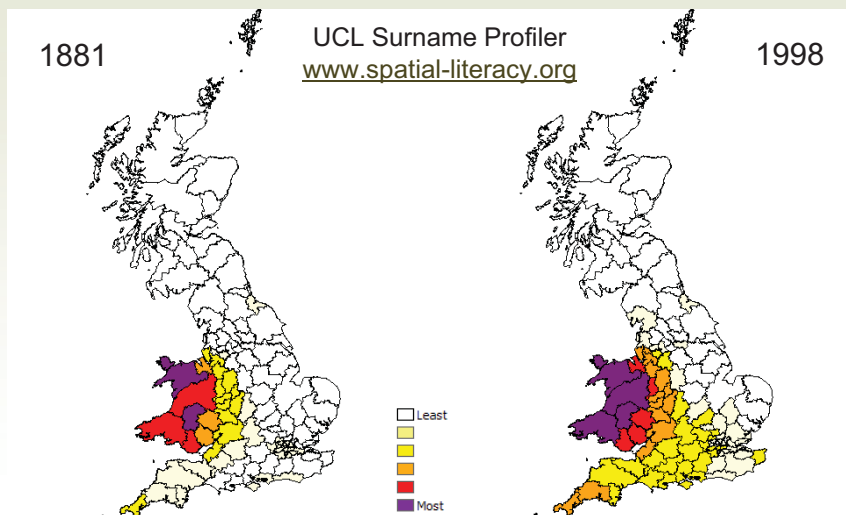
[www.onomap.org](http://www.onomap.org)  
[www.publicprofiler.org](http://www.publicprofiler.org)

**Contents**

1. Surname Geography
2. Ethnicity, Names and Networks

# Surname Geography

## Welsh Surnames 1881-1998



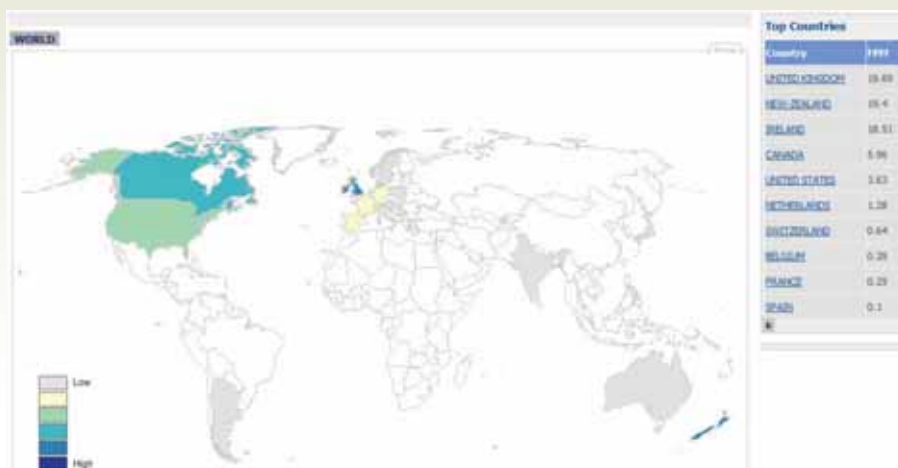
# WorldNames

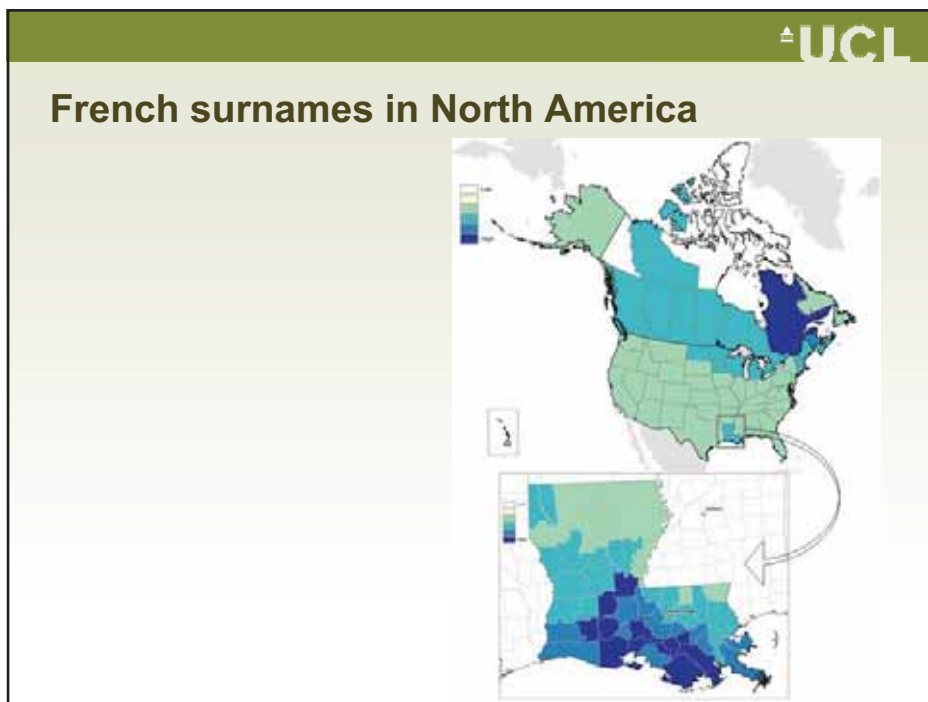
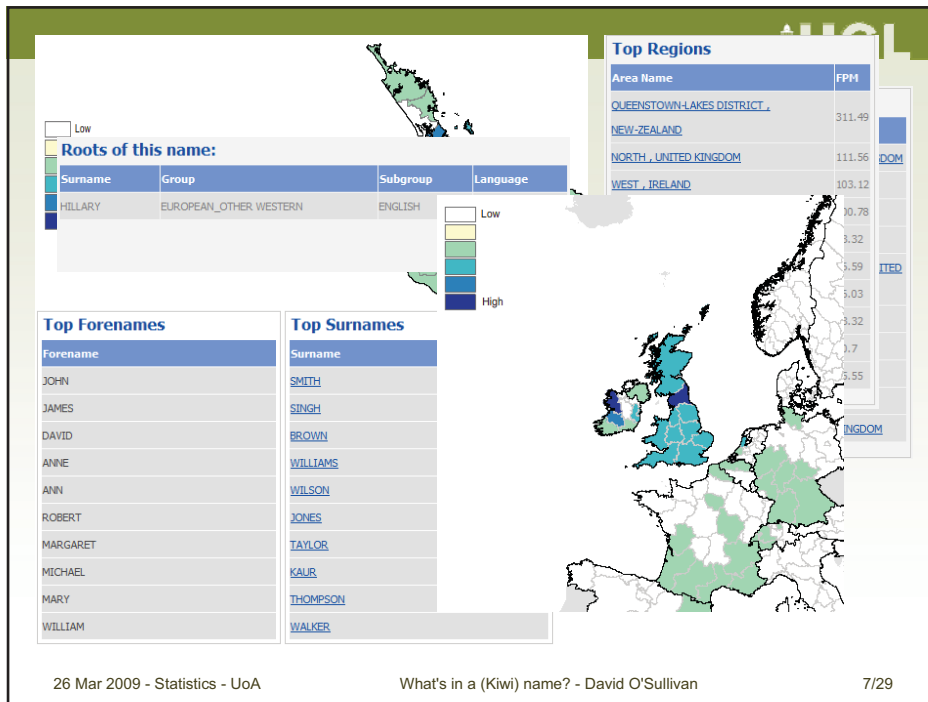
[www.publicprofiler.org/worldnames](http://www.publicprofiler.org/worldnames)

- 26 countries in America, Europe, Asia and Oceania
- Covering a total population of 1 billion people
- Individual level data for 300 million people (full name and address)
- 8 million unique surnames
- 5 million unique forenames
- Postcode geographies

## The world names project

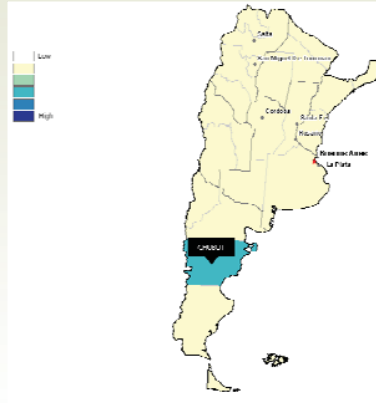
[www.publicprofiler.org/worldnames](http://www.publicprofiler.org/worldnames)





## Welsh surnames in Argentina

- Community of Welsh speakers in Patagonia
- Welsh migration and settlement in the 19th century.



## Mapping names

- The whole database is available on-line
- Mapping historic and contemporary migrations

 **publicprofiler worldnames**

[www.publicprofiler.org/worldnames](http://www.publicprofiler.org/worldnames)

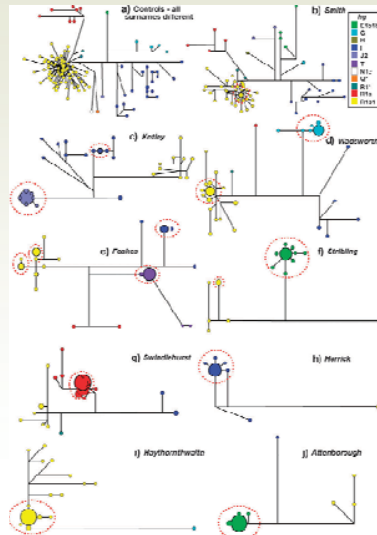
[www.onomap.org](http://www.onomap.org)

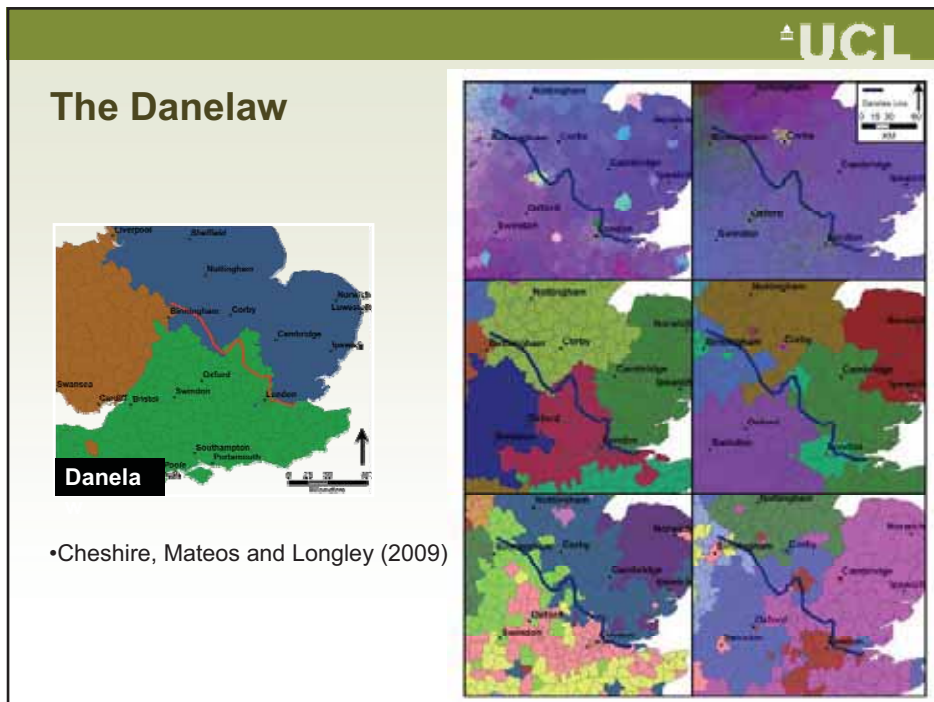
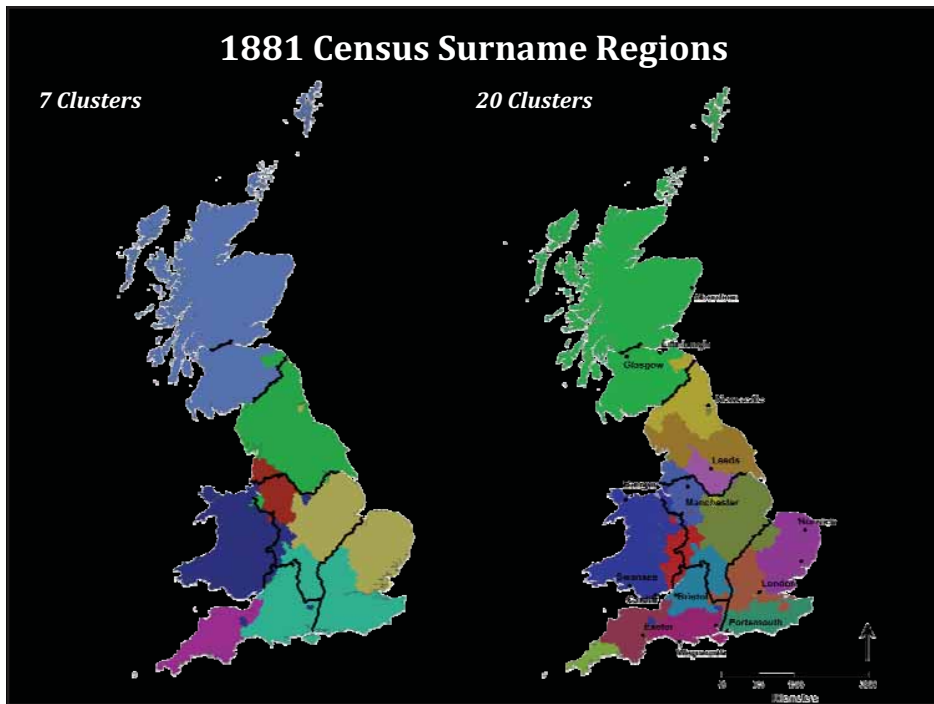




Thanks to James Cheshire for producing the Tag cloud using WorldNames data

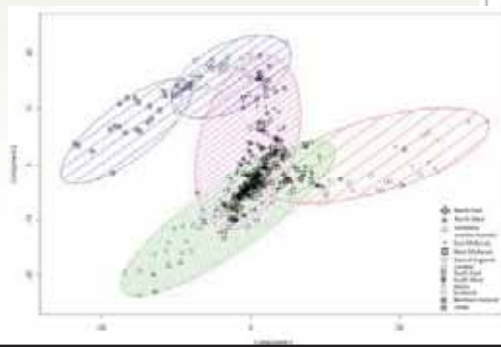
## Surnames and Genetics (Y-chromosome)



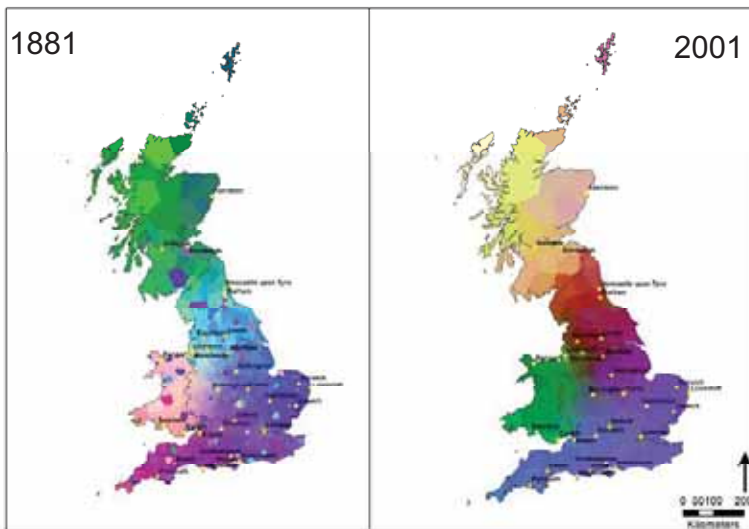


### GB Surname and cultural regions

- Multidimensional scaling (MDS) of name distances (isonymy) between districts
- Cheshire, Mateos & Longley 2009



### GB Surname Regions



•Cheshire, Mateos and Longley (2009)

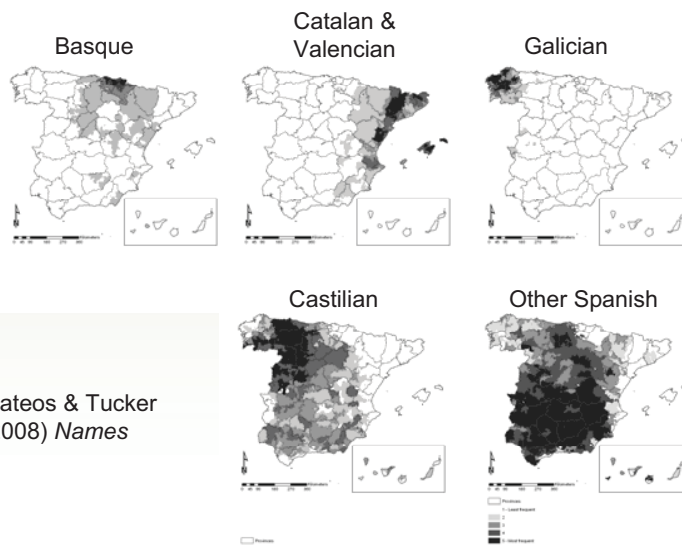


## Europe surname regions

- MDS of Lasker Distances between NUTS3 areas
- Cheshire (2009)



## Spain & Population Settlement History



Surname frequencies, grouped by CEL (2004 telephone directory)

Mateos & Tucker (2008) *Names*

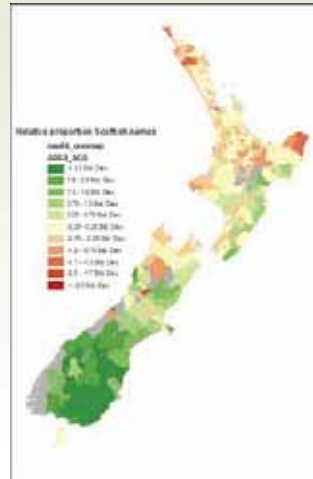
## New Zealand: Scottish names

Passenger arrivals at Port Chalmers, New Zealand, March 1848 - January 1851

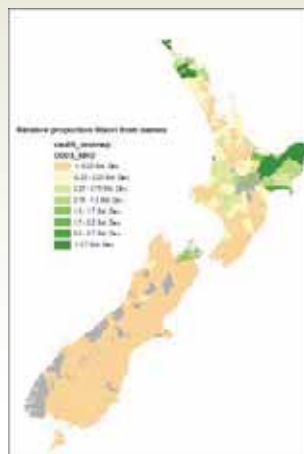
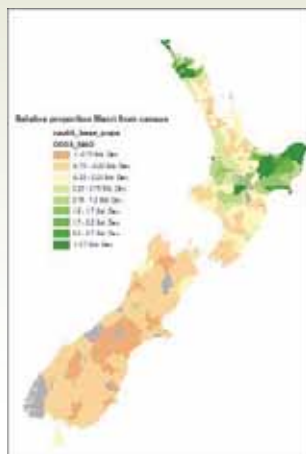
PHILIP LAING

PHILIP LAING, 47th York Street, Port Chalmers, on 22/11/1847, and arrived on 06/12/1847, and arrived at 18/12/1847

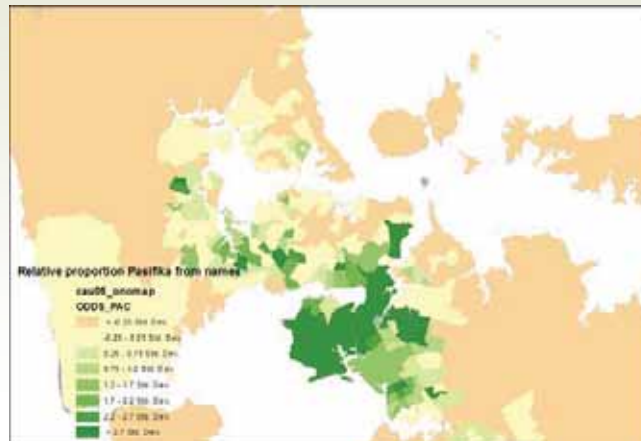
[Breakie, Hugh](#) (22) [unn.]  
[Brown, James](#) (23) [calico printer, ag.; Andersons Bay, d. 27/11/1891];  
[Brown, John](#) (36) [contractor]; wife (32).  
[Bruce, Alex, Hogg](#) (22) [schoolmaster; went to Wellington in 1848].  
[Buchanan, James](#) (44)  
[Buchanan, John](#) (46) [weaver, beadle; Great King St. d. 26/3/1880]; [Ma](#)  
[Buchanan, Thomas](#) (42) [weaver, farmer; West Taieri, d. 16/7/1876]; [Ma](#)  
18/1/1913]; [John](#) (8) [d. 31/5/1874]; [David](#) (4) [d. 11/8/1927]; [Jesse](#) (3) [[Burns, Thomas](#) Rev. (52) [first minister (Free Church of Scotland), d. 23  
[m. Rev William Bannerman; d. 9/10/1923]; [Anne](#) (8) [m. Alexander Robt  
[Callander, James](#) (41) [farmer; East Taieri]; [Jean](#) née Cuthbertson, wife  
[Cameron, Lewis](#) (23); [Margaret](#) née Martin, wife (20) [d. 10/1856]; [Ma](#)  
[Carnegie, John de la Comdaine](#) (19) [storekeeper, auctioneer; d. 13/8  
[Carrick, Robert](#) (14) [d. pre 1690].



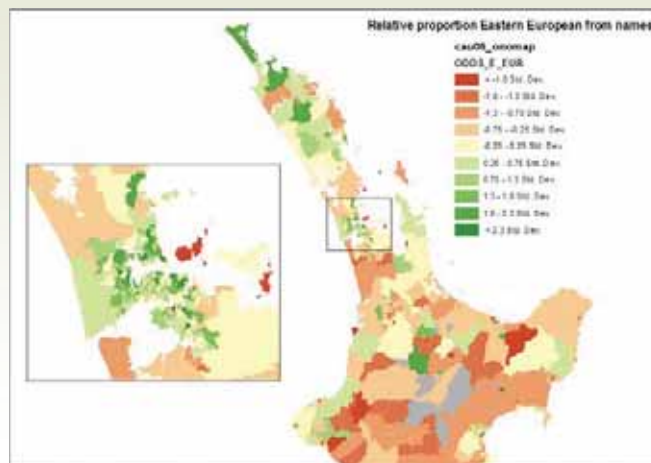
## New Zealand: Māori names



## Auckland, N: Pasifika names

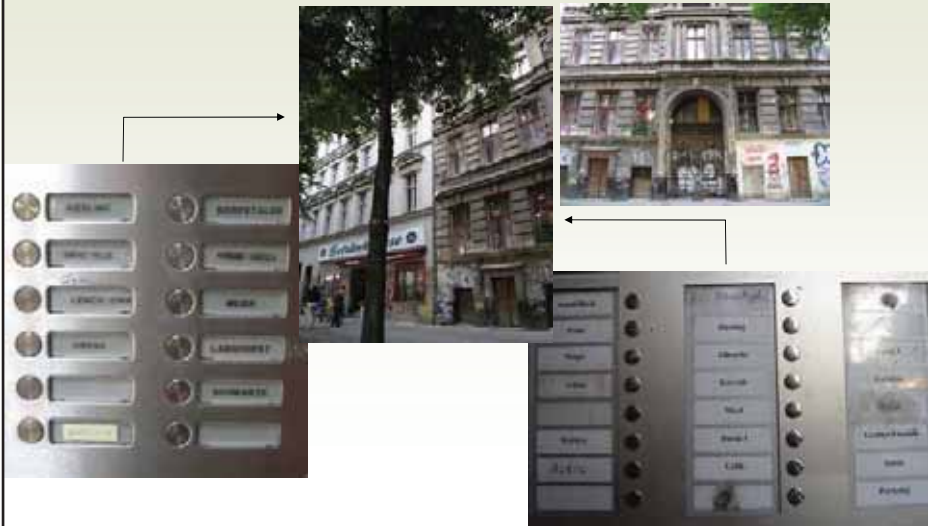


## New Zealand: 'Eastern European' names

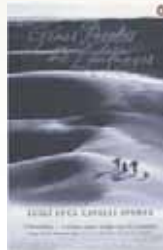
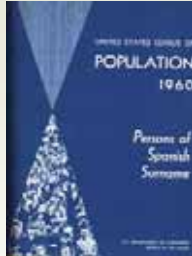


# Ethnicity Names & Networks

## Names in Kreuzberg, Berlin



## Research on names and identity



- **Demography and epidemiology;** subdivision of populations by ethnicity
  - US Census Hispanic names list (Passel and Word since 1950s)
  - Asian surnames in US (Lauderdale, 2004)
  - South Asian names in UK (Nam Pechan & SANGRA)
- **Genetics;** Population structure and geography, endogamy and gene mutations
- **Economics;** Name discrimination in labour, housing, and credit markets
- **Geography and Sociology;** cultural transmission, migration and spatial diffusion

## Decoding ethnicity from names

- Names can potentially provide information about:

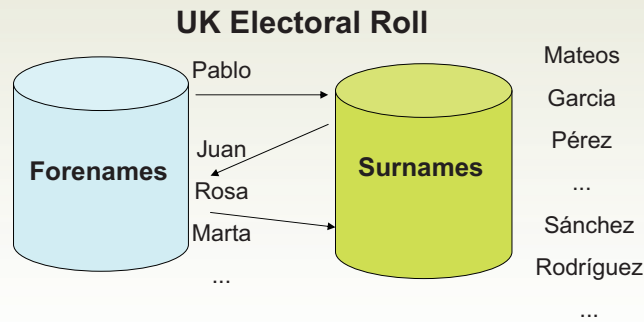
Aspect	Etymology/ Onomastics	Space-time Distribution
Surname & Forename	<b>Language</b>	<b>Geographic Origin</b>
	<b>Religion</b>	<b>Migration flows</b>
Forename	<b>Gender</b>	<b>Age...</b>

- Review paper of name-based classifications of ethnicity; Mateos (2007) *Population, Space and Place*
- Primarily public health applications
- Main groups: Hispanic, South Asian, Chinese, and Muslim



## Onomap classification

- Forename-Surname clustering (based on Hanks and Tucker, 2000)

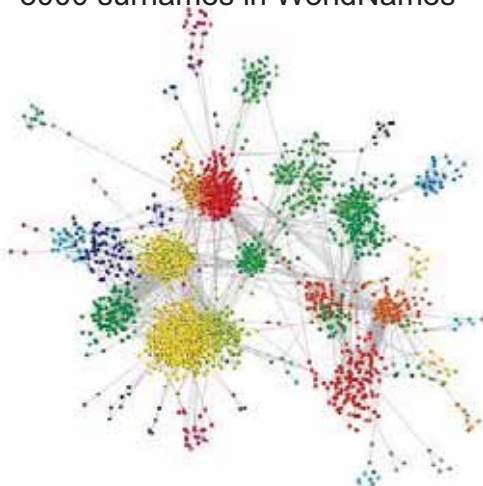


- Several iterations until self-contained cluster is exhausted
- Cluster assigned a cultural, ethnic & linguistic Onomap type
- Probability of ethnicity assigned to each name

Mateos et al (2007) CASA Working Paper 116

## Naming networks

- 5000 surnames in WorldNames



- **Nodes** = unique surnames
- **Edges** = common forenames between a pair of surnames
- **Weights** = relative importance of the forename to each surname (% of pop)
- **Colours** = clusters assigned by network clustering algorithms

• Mateos, Longley and O'Sullivan (forthcoming)

## Social networks parallel

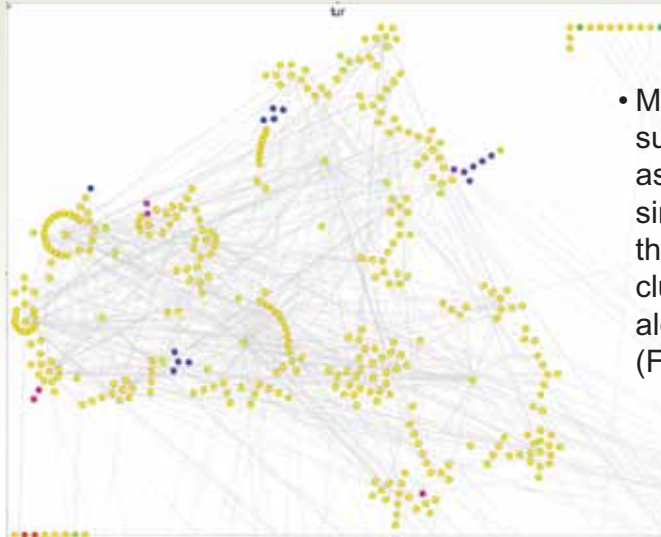
- Granovetter (1973) *The Strength of Weak Ties*
- Weak ties play an essential role in the diffusion of information and innovation
- Cliques of highly related names are separated by bridges or weak links (sparse links)
- Key: identify bridges removing 'the weak ties' through filtering on edge weights
- Network clustering algorithms (Newman)
- Forthcoming paper with David O'Sullivan on naming network clustering, using Fastcommunity (Clauset, Newman and Moore, 2004)

## WorldNames CEL clusters



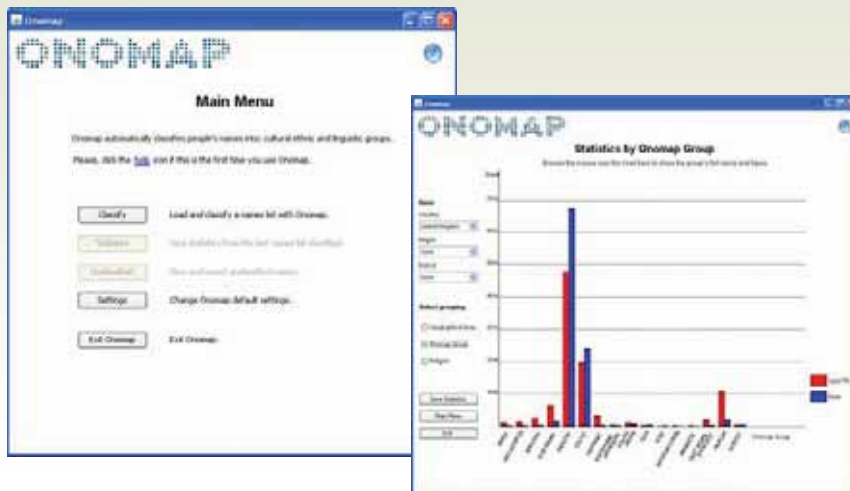


## Subnetwork of Turkish surnames



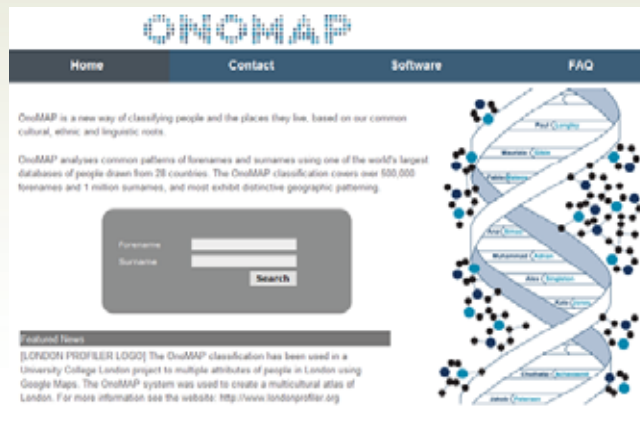
- Most Turkish surnames assigned to a single cluster by the automatic clustering algorithm (Fastcommunity)

## Onomap software



## Onomap online

- Individual names can be searched on-line  
[www.onomap.org](http://www.onomap.org)



## Evaluation of Onomap at the individual level

- Evaluated against individual's ethnicity (hospital admissions, birth register)
- On average: 80% accuracy
- Least indicated for:
  - Long established groups (i.e. Irish, Jewish)
  - Groups with non-unique names (i.e. Caribbeans)
- Highly indicated for all other groups

Mateos, Longley and Webber (2007)

Lahka, Gorman and Mateos (forthcoming)

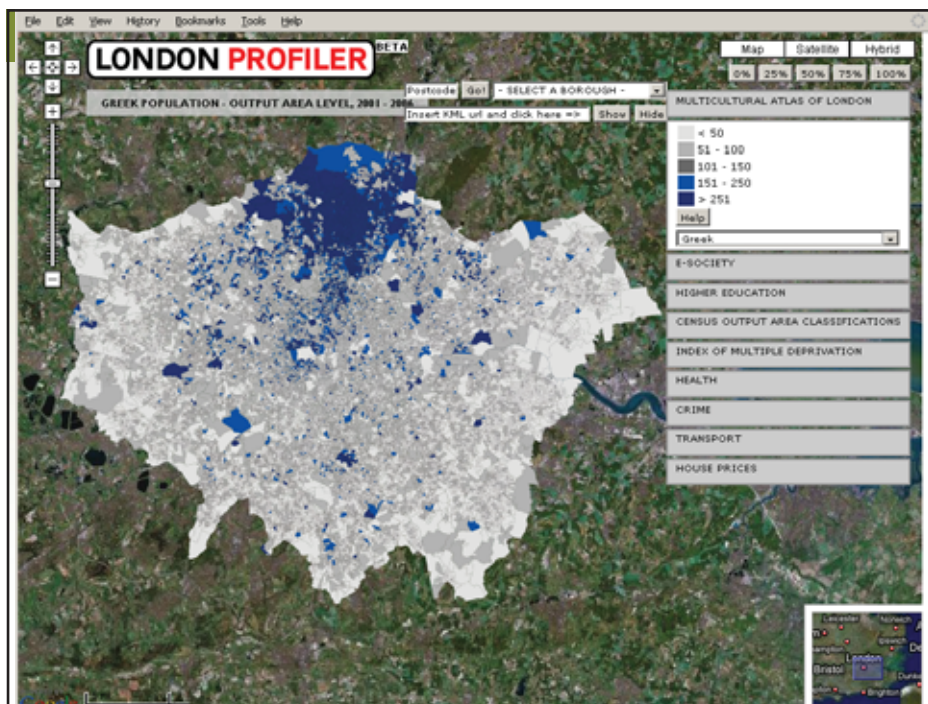
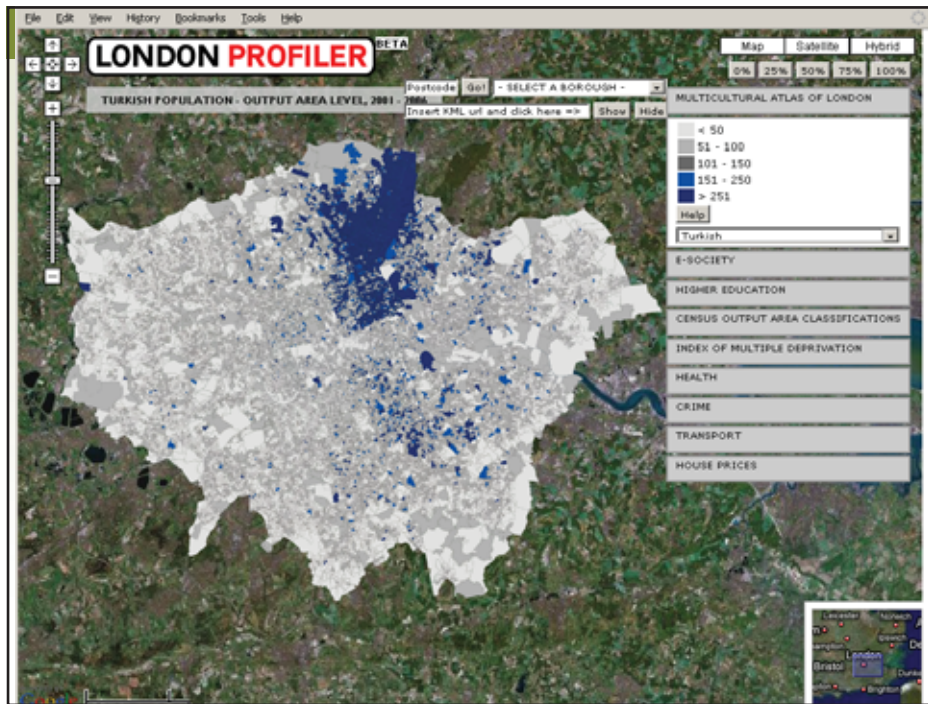
## Current applications of Onomap

- Cancer research
  - LSHTM (5 m. patients)
- Public Health
  - PCTs (Camden, Islington, Southwark)
  - University of Edinburgh & GROS (Onomap evaluation)
  - University of Essex
- Political party representation
  - ANU, Australia; Princeton Univ.
- Residential Segregation
  - Europe, US, UK

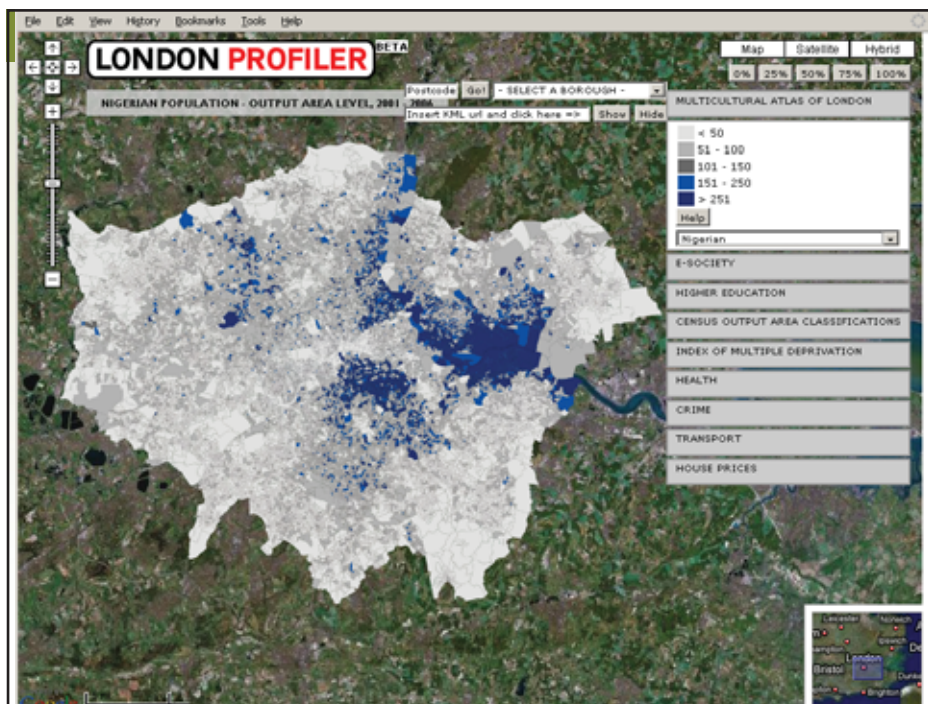
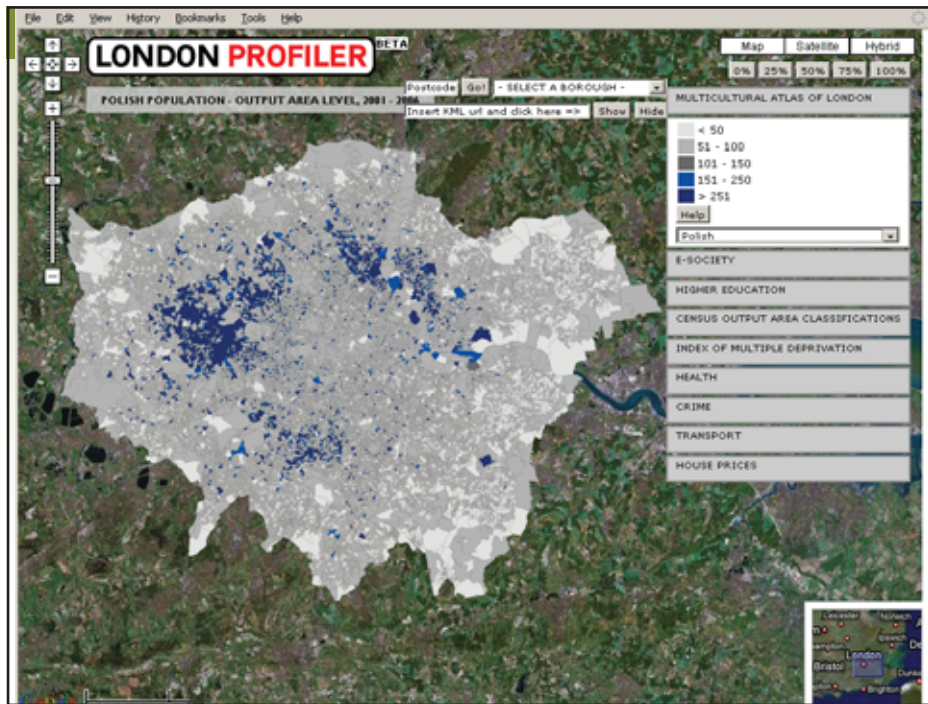
## Ethnic group distribution

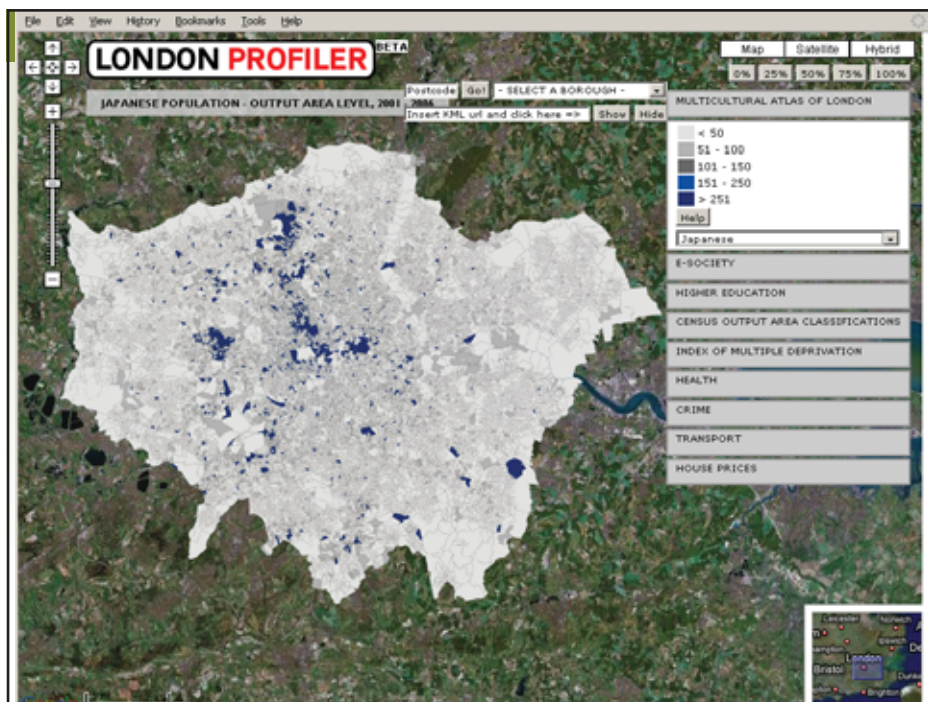
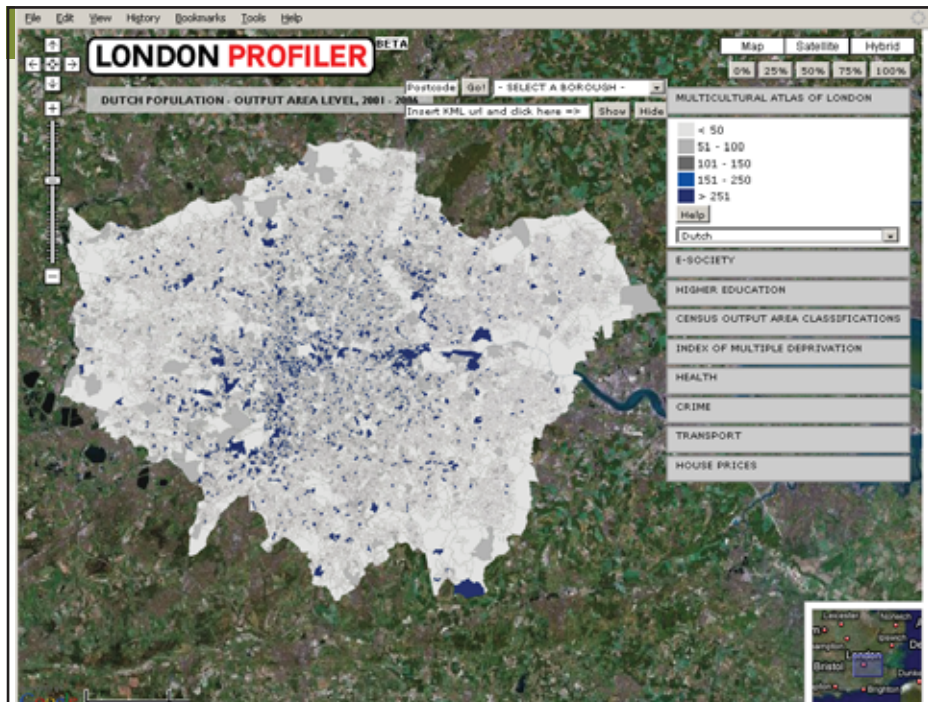
- E.g. Turkish names in Germany







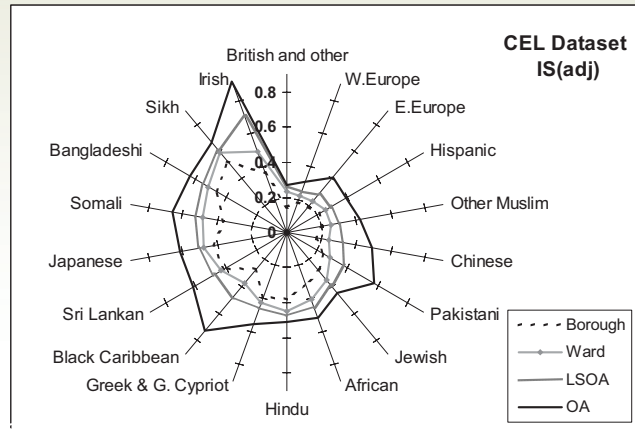




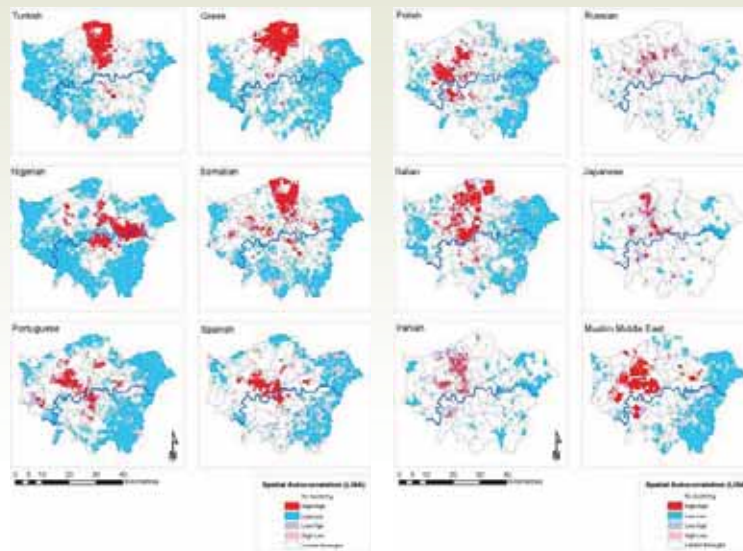


## Index of Dissimilarity – Onomap

- Index of Dissimilarity in London, using Onomap categories at four different geographical scales



## Spatial Autocorrelation (LISA) by Onomap type



## Lack of ethnicity data in Europe

- Incompatible definitions and measurements:
  - Migrant / Native
  - Foreigner / National
  - Migrant ancestry
  - Race / Skin colour
  - Cultural / National identity
- Lack of ethnicity data collection, or poor quality and comparability
- Lack of sufficient geographical granularity
- Low frequency of update
- Need for complementary methodologies to study segregation
  - There is a fragmented, nation-specific research perspective on spatial segregation in Europe (Musterd, 2005)

## European cities segregation analysis

- Names from telephone directories coded by Onomap Subgroup
- 31 largest cities selected
- Counts per Onomap Group and small area (postcodes) calculated within each city
- Index of Dissimilarity calculated for each city

$$D = \frac{1}{2} \sum_{i=1}^n \left| \frac{x_i}{X} - \frac{t_i - x_i}{T - X} \right| \quad (\text{Duncan and Duncan, 1955})$$

$x_i$  = Population of group X in area  $i$

$X$  = Total population of group X in the city

$t_i$  = Total population of area  $i$

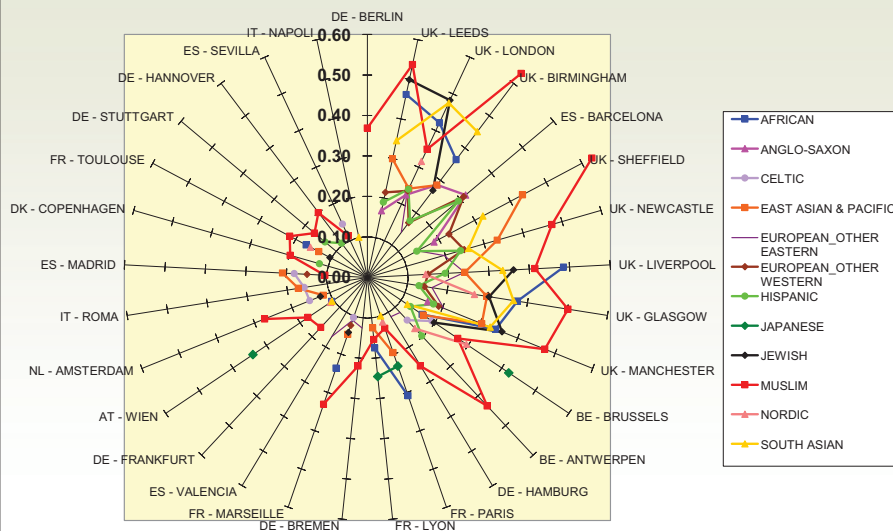
$T$  = Total population of the city



## Precautions taken

- To prevent small groups driving results
- Analysis done at Onomap Group level (14 groups) not Onomap Subgroup (80)
- Presentation of results filtered, after ID was calculated:
  - Remove Onomap Groups with average population per area in a city < 15 individuals
  - Remove Onomap Groups with ID < 0.1

## Results: Index of Dissimilarity per City



## Conclusions: European Segregation

- UK cities much more segregated than rest of countries
  - could be a result of finer small area geography
  - followed by cities in Belgium and France
- Muslim group most segregated across Europe
  - followed by South Asian, and African
  - at larger distance by Japanese, Jewish & East Asian
- Issues:
  - High variability between cities: population sizes, extent and number of areas
  - Differential population coverage (tel. dir. & elec. roll)
  - Onomap Group level may not reflect the full reality in each country (i.e. historic minority groups)
  - Dissimilarity index on its own not ideal (Simpson, 2005; Massey & Denton, 1988)

## Conclusions: Onomap methodology

- Advantages
  - Facilitates ethnicity analysis using finer spatial, temporal, and nominal granularity
  - Cost-efficient alternative when ethnicity data is missing/ low quality
  - Ethnicity categories can be re-aggregated in different ways
  - Probability scores; tailor classification to specific applications
- Disadvantages
  - Only reflects patrilineal heritage (problem of mixed ethnicity)
  - Different histories of surname adoption, naming conventions & name change rules in each language and country
  - Name normalisation decisions are required
  - Publicly available registers of names have biases
  - Not appropriate for reporting ethnicity at individual level
  - Ethical considerations and privacy issues

Thanks for listening

Paul Longley [plongley@geog.ucl.ac.uk](mailto:plongley@geog.ucl.ac.uk)

Pablo Mateos [p.mateos@ucl.ac.uk](mailto:p.mateos@ucl.ac.uk)

[www.onomap.org](http://www.onomap.org)

[www.publicprofiler.org/worldnames](http://www.publicprofiler.org/worldnames)

[www.londonprofiler.org](http://www.londonprofiler.org)