

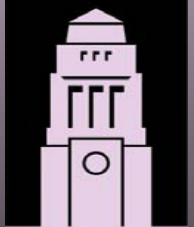
DIGITAL GEOGRAPHY *in a* WEB 2.0 WORLD



UCL

NCeSS

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL



UNIVERSITY OF
LEEDS

SPLINT
Spatial Literacy IN Teaching

WEDNESDAY 20 FEBRUARY 2008

WEDNESDAY 20 FEBRUARY 2008 @ The Barbican

The names projects: Onomap and publicprofiler/ WorldNames

Pablo Mateos, Alex Singleton, Maurizio Gibin,
Muhammad Adnan, Jakob Petersen & Paul Longley

Department of Geography and Centre for Advanced Spatial Analysis
University College London

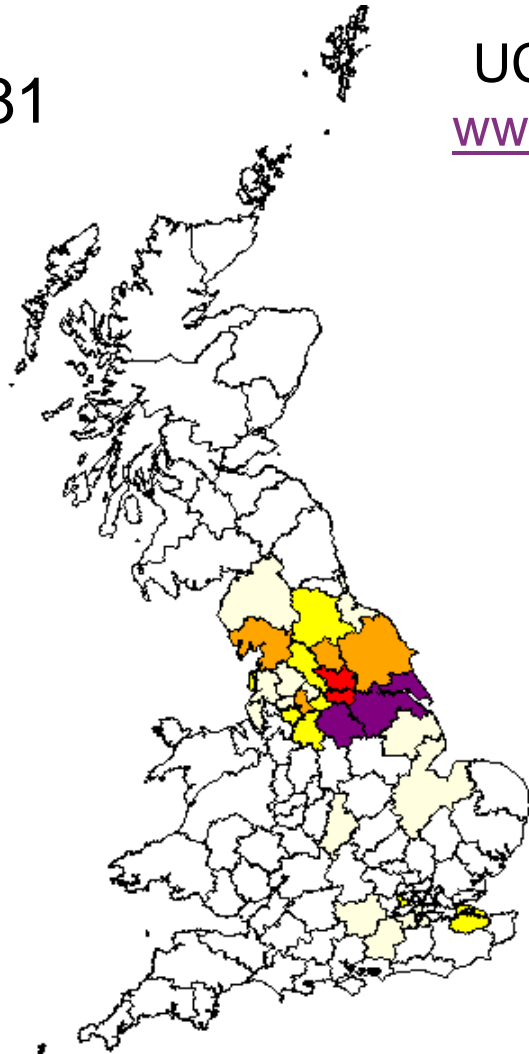
www.publicprofiler.org

Contents

1. Introduction: The geography of names
2. publicprofiler/ WorldNames
3. Onomap
4. Conclusions

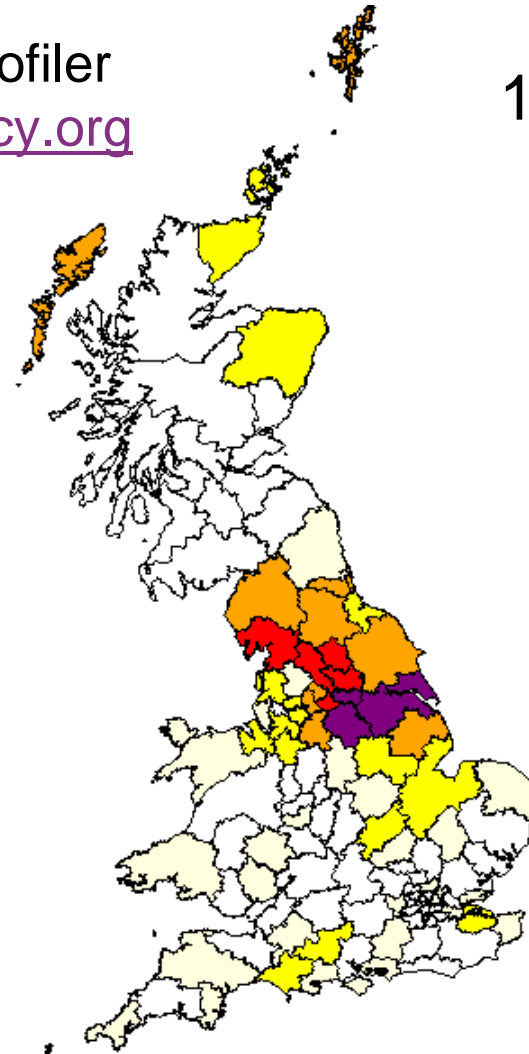
Distribution of 'Battys' 1881-1998

1881



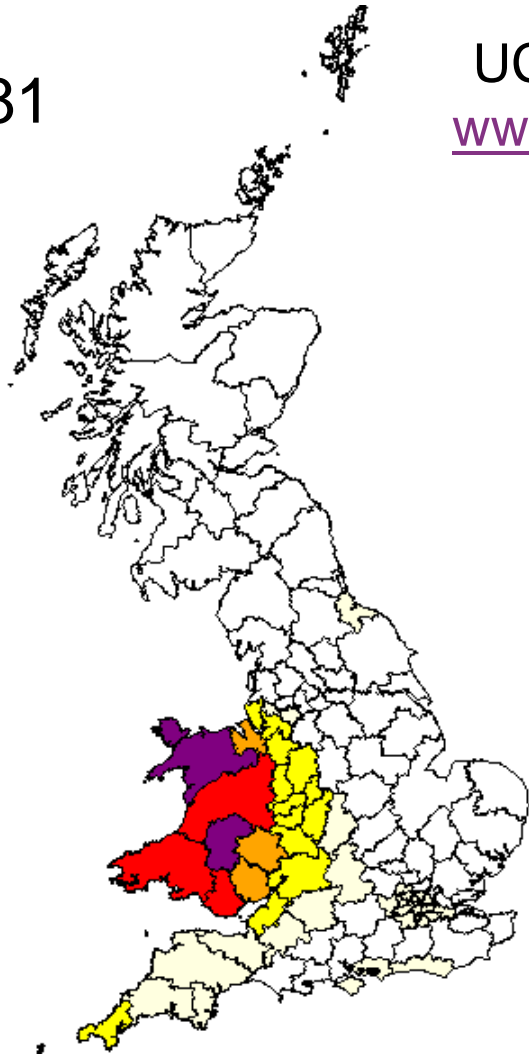
UCL Surname Profiler
www.spatial-literacy.org

1998



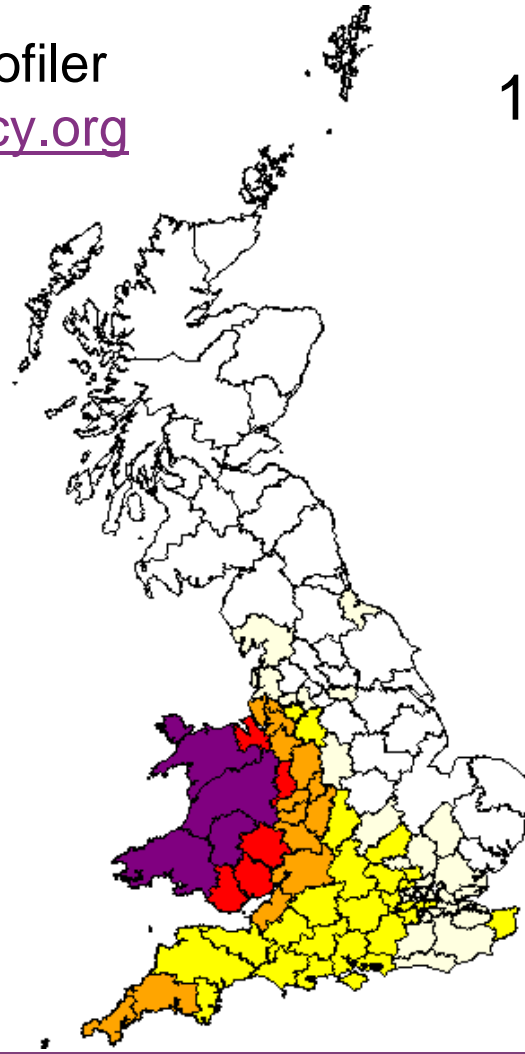
Welsh surnames 1881-1998

1881



UCL Surname Profiler
www.spatial-literacy.org

1998



Media exposure

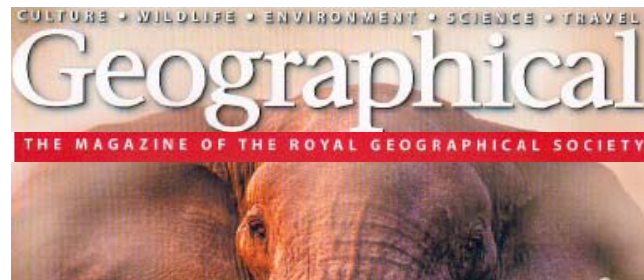


THE INDEPENDENT

WORLDWATCH



THE GLOBAL GEOGRAPHY OF SURNAMES



BBC NEWS

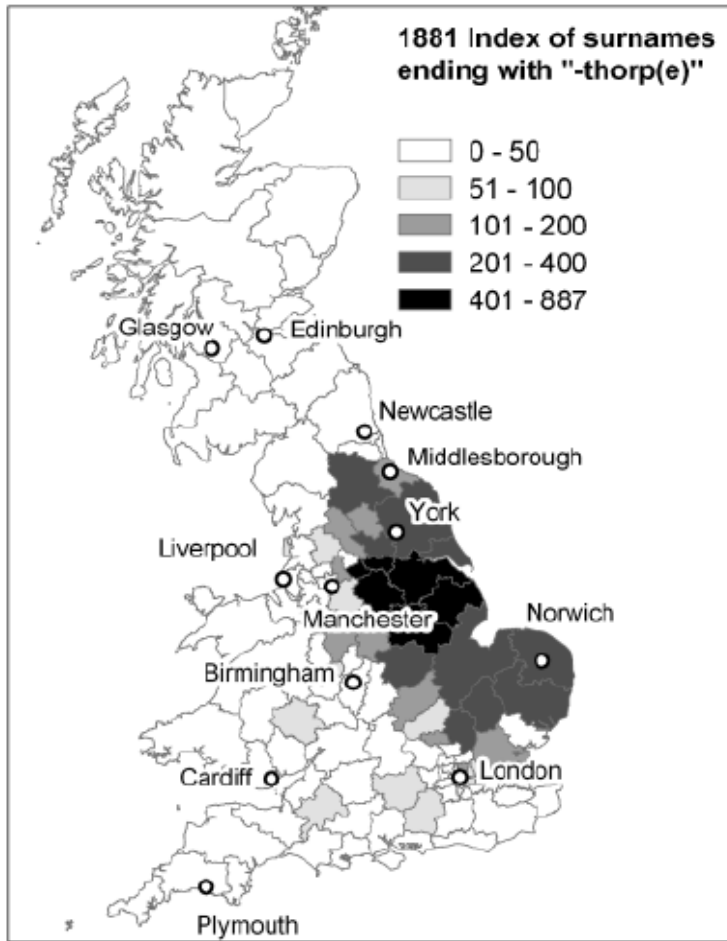
Most adventurous surnames mapped

The most adventurous Britons over the decades have been identified by academics analysing people's surnames.

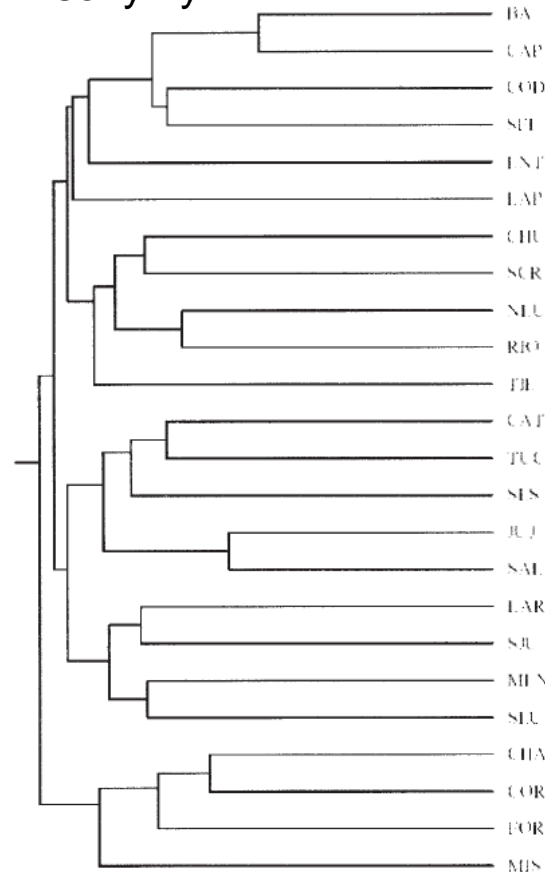
Geographers at the Centre for Advanced Spatial Analysis, at University College London, looked at the spread of 500,000 names across the world.

There are 50% more Blairs now than in 1881

Names and cultural regions



Clusters of Argentina's provinces according to isonymy

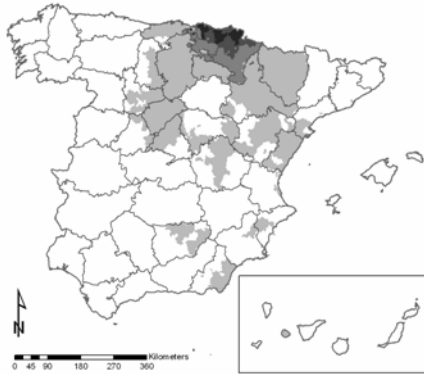


Actual geographical arrangement of provinces

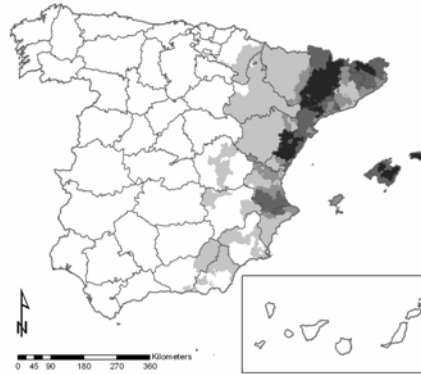


Surnames in Spain

Basque



Catalan & Valencian



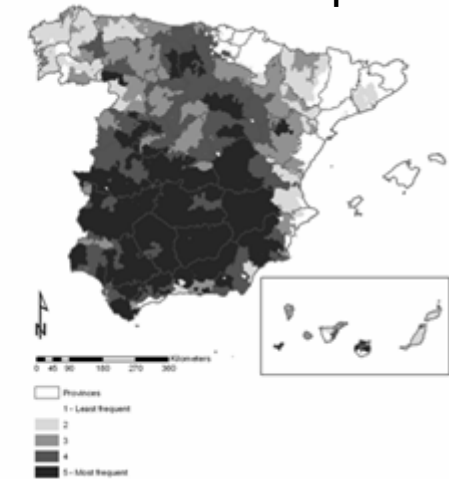
Galician



Castilian



Other Spanish



- Spanish population settlement history
- 2004 Telephone directory

Mateos (2008)

 **publicprofiler**
WorldNames

www.publicprofiler.org/worldnames

www.publicprofiler.org/worldnames

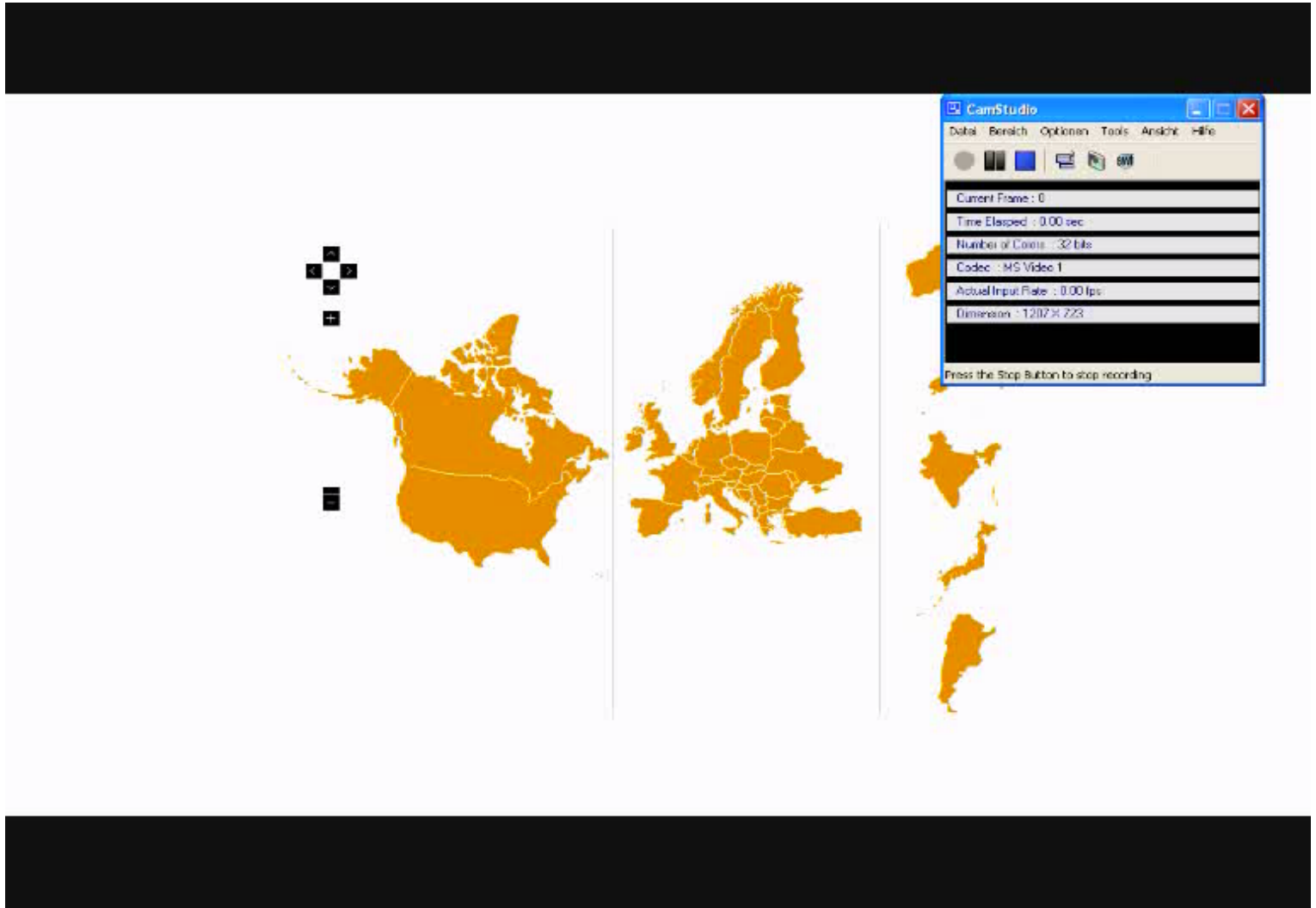
- 26 countries in America, Europe, Asia and Oceania
- Covering a total population of 1 billion people
- Individual level data for 300 million people (full name and address)
- 10.8 million unique surnames
- 6.5 million unique forenames
- Statistics at postcode level

Characteristics:

- Technical
 - Flash application on a standard web browser
- Contents
 - Over 1 million names (surnames and forenames)
 - Maps of name groupings (by ethnic origin or other features)
- Name Statistics
 - Top names per country, region, city, etc (queriable from map)
 - Top city, region, country, etc per name
 - Cultural, ethnic and linguistic origin of name
 - Most frequent forenames associated with a surname (and vice-versa)
 - Average altitude of a name. Others

WorldNames

Click on the picture to see a demo →



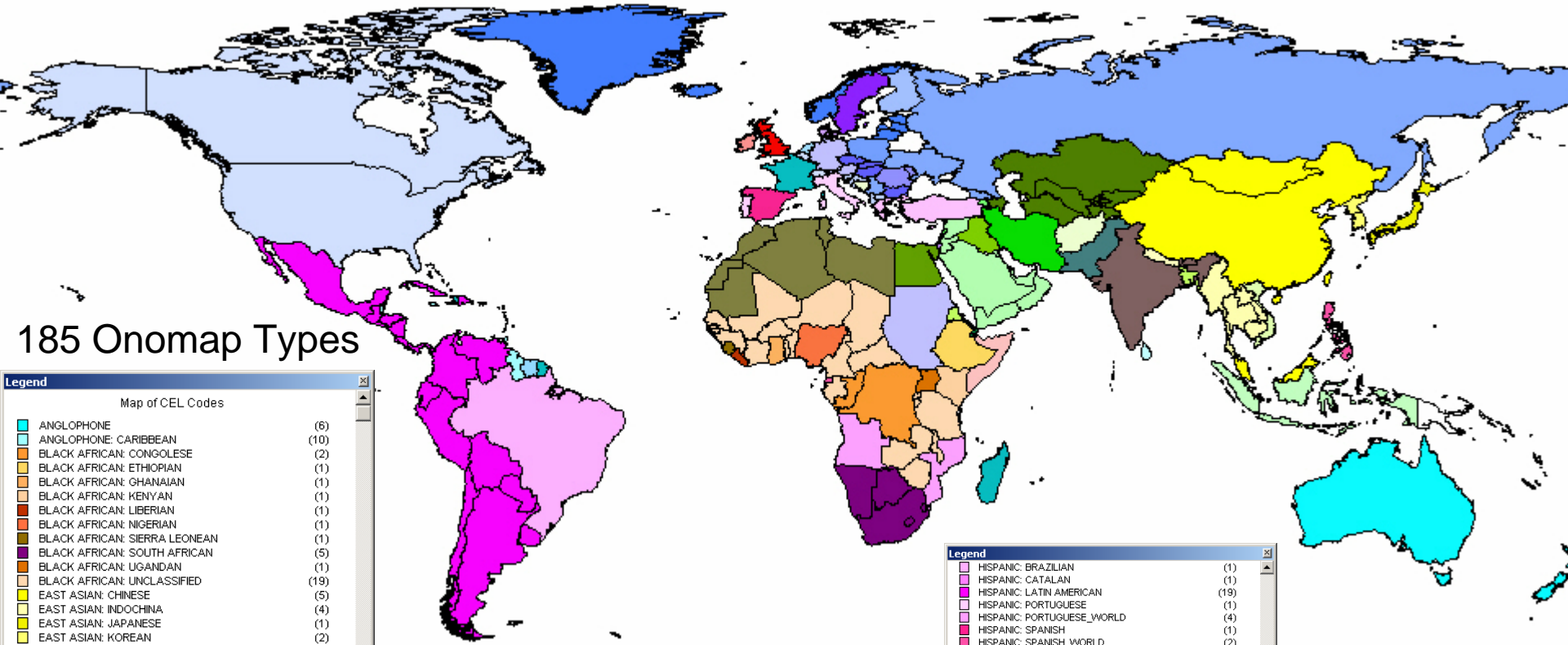


www.onomap.org

Names in Kreuzberg, Berlin



World map of Onomap categories



185 Onomap Types

Legend
Map of CEL Codes

ANGLOPHONE	(6)
ANGLOPHONE: CARIBBEAN	(10)
BLACK AFRICAN: CONGOLESE	(2)
BLACK AFRICAN: ETHIOPIAN	(1)
BLACK AFRICAN: GHANAIAN	(1)
BLACK AFRICAN: KENYAN	(1)
BLACK AFRICAN: LIBERIAN	(1)
BLACK AFRICAN: NIGERIAN	(1)
BLACK AFRICAN: SIERRA LEONEAN	(1)
BLACK AFRICAN: SOUTH AFRICAN	(5)
BLACK AFRICAN: UGANDAN	(1)
BLACK AFRICAN: UNCLASSIFIED	(19)
EAST ASIAN: CHINESE	(5)
EAST ASIAN: INDOCHINA	(4)
EAST ASIAN: JAPANESE	(1)
EAST ASIAN: KOREAN	(2)
EAST ASIAN: VIETNAMESE	(1)
EUROPEAN: BALKAN	(4)
EUROPEAN: BRITISH: UNCLASSIFIED	(1)
EUROPEAN: DANISH	(1)
EUROPEAN: DUTCH	(1)
EUROPEAN: DUTCH_WORLD	(1)
EUROPEAN: EASTERN EUROPE	(3)
EUROPEAN: FINNISH	(1)
EUROPEAN: FRENCH	(2)
EUROPEAN: FRENCH_WORLD	(8)
EUROPEAN: GERMAN	(3)
EUROPEAN: GREEK / GREEK CYPRIOT	(2)
EUROPEAN: HUNGARIAN	(1)
EUROPEAN: IRISH: UNCLASSIFIED	(1)
EUROPEAN: ITALIAN	(3)
EUROPEAN: NORDIC	(7)
EUROPEAN: OTHER	(5)
EUROPEAN: POLISH	(1)
EUROPEAN: ROMANIAN	(2)
EUROPEAN: SLAVIC	(4)
EUROPEAN: SWEDISH	(1)

Legend

HISPANIC: BRAZILIAN	(1)
HISPANIC: CATALAN	(1)
HISPANIC: LATIN AMERICAN	(19)
HISPANIC: PORTUGUESE	(1)
HISPANIC: PORTUGUESE_WORLD	(4)
HISPANIC: SPANISH	(1)
HISPANIC: SPANISH_WORLD	(2)
JEWISH	(1)
MUSLIM: AFGHAN	(1)
MUSLIM: ARAB	(5)
MUSLIM: ARMENIAN	(1)
MUSLIM: BALKANS	(1)
MUSLIM: BANGLADESHI	(1)
MUSLIM: BLACK AFRICAN OTHER	(1)
MUSLIM: EGYPTIAN	(1)
MUSLIM: ERITREAN	(1)
MUSLIM: EURASIA	(6)
MUSLIM: IRANIAN	(1)
MUSLIM: IRAQI	(1)
MUSLIM: LEBANESE	(1)
MUSLIM: MIDDLE EASTERN	(6)
MUSLIM: NORTH AFRICAN	(6)
MUSLIM: PAKISTANI	(3)
MUSLIM: SOMALI	(1)
MUSLIM: SOUTHEAST ASIA	(2)
MUSLIM: SUDANESE	(1)
MUSLIM: TURKISH	(1)
OTHER SOUTH ASIAN: NEPALESE	(1)
OTHER SOUTH ASIAN: SOUTH INDIAN & SRI LANKAN	(1)
SOUTH ASIAN: HINDI OR SIKH	(2)

Tucker's CELG Method

- The Dictionary of American Family Names, Oxford University Press
- CELG method developed by Ken Tucker and Patrick Hanks using telephone directories



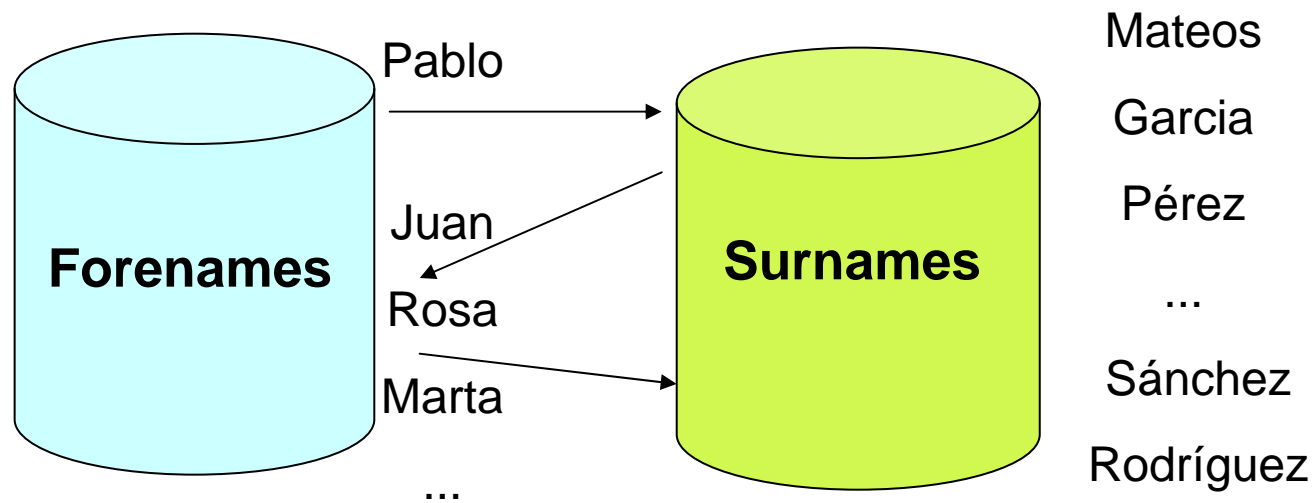
Hanks (2003)



Onomap classification

- Forename-Surname clustering

UK Electoral Roll



- Several iterations until self-contained cluster is exhausted
- Cluster assigned an Onomap type

Mateos *et al* (2007) CASA Working Paper 116

Correlations Onomap vs Census

Pearson Correlation Coefficient between:

2001 Census and 2004 GB Electoral Roll classified by CEL Types

<i>Ethnic Group</i>	<i>Geographical Unit of Comparison</i>			
	<i>OA</i>	<i>LSOA</i>	<i>WARD</i>	<i>LA</i>
A) White - British	0.88	0.93	0.93	0.95
B) White - Irish	0.32	0.37	0.42	0.46
C) White - Any other White background	0.74	0.85	0.88	0.93
H) Asian or Asian British - Indian	0.92	0.95	0.96	0.98
J) Asian or Asian British - Pakistani	0.90	0.93	0.93	0.91
K) Asian or Asian British - Bangladeshi	0.91	0.93	0.95	0.98
L) Asian or Asian British - Any other Asian background	-0.06	0.11	0.24	0.62
M) Black or Black British - Caribbean	0.32	0.77	0.91	0.98
N) Black or Black British - African	0.83	0.95	0.97	0.99
R) Other Ethnic Groups - Chinese	0.65	0.79	0.84	0.97
S) Other Ethnic Groups - Any other ethnic group	0.38	0.66	0.77	0.88
Number of Units valid for analysis	218,037	40,883	10,072	408

*GB = England, Wales and Scotland. Values over 0.75 are highlighted in **bold***

OA = Output Area, LSOA = Super Output Area, LA = Local Authorities

Evaluation at the individual level

- Hospital Episode Statistics (HES) in Camden & Islington (London) 1998-2006
- Onomap classification of names compared with self-reported ethnicity
- 343,068 patients matched to a unique ethnic code
- Problem of bad quality HES data
 - Inconsistent data (Aspinall and Jacobson, 2004)
 - Incomplete data. Only 75% of records contain ethnicity (London Health Observatory, 2005).

HES Evaluation Results

Predicted by CEL		Actual Ethnicity from HES data										
		0	1	2	3	4	5	6	7	8	9	Total
0	White	150,574	7,971	4,468	2,535	595	68	160	488	17,383	73,920	258,162
1	Black - Caribbean	92	226	21	32	3				69	197	640
2	Black - African	857	283	5,996	698	53	14	41	23	1,695	4,716	14,376
3	Black - Other											0
4	Indian	1,066	96	562	125	2,184	85	171	30	1,679	3,503	9,501
5	Pakistani	856	60	1,736	306	690	861	2,390	17	2,507	4,625	14,048
6	Bangladeshi	284	30	373	122	687	194	6,086	5	1,174	3,777	12,732
7	Chinese	227	39	72	21	11	2	7	1,473	531	1,088	3,471
8	Any other ethnic group	3,811	111	990	228	202	112	280	358	5,858	5,747	17,697
9	Unclassified	3,364	328	1,706	322	164	32	107	47	2,199	4,079	12,348
Total		161,131	9,144	15,924	4,389	4,589	1,368	9,242	2,441	33,095	101,652	342,975

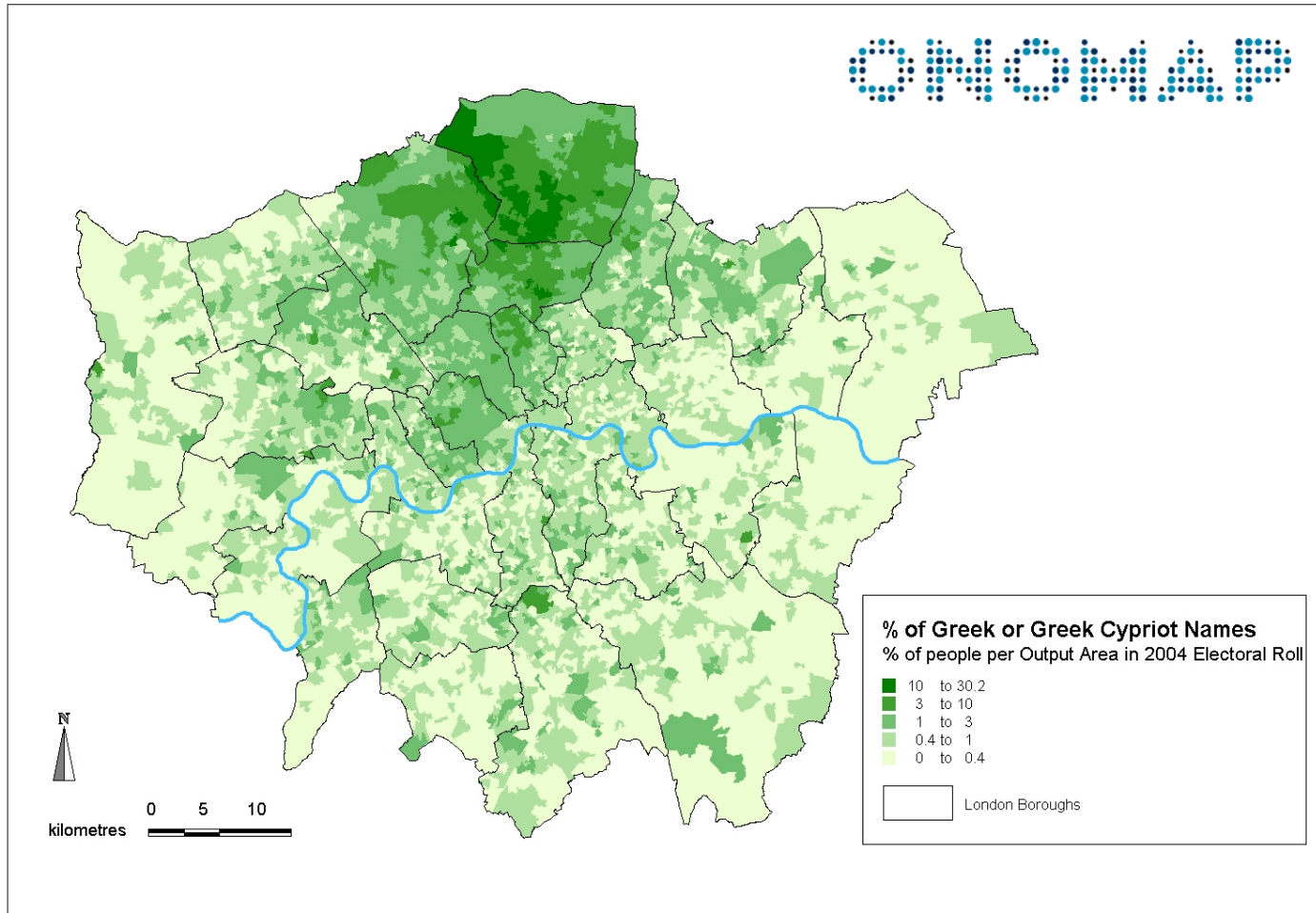
1991 Census Categories		Sensitivity	Specificity	PPV	NPV
0	White	0.93 - 0.98	0.58 - 0.62	0.82 - 0.90	0.82 - 0.89
1	Black - Caribbean	0.02 - 0.03	1.00 - 1.00	0.51 - 0.62	0.96 - 0.96
2	Black - African	0.38 - 0.45	0.98 - 0.99	0.62 - 0.76	0.96 - 0.96
3	Black - Other	n/a	n/a	n/a	n/a
4	Indian	0.48 - 0.52	0.98 - 0.99	0.36 - 0.50	0.99 - 0.99
5	Pakistani	0.63 - 0.70	0.96 - 0.97	0.09 - 0.12	1.00 - 1.00
6	Bangladeshi	0.66 - 0.69	0.99 - 0.99	0.68 - 0.79	0.99 - 0.98
7	Chinese	0.60 - 0.73	1.00 - 1.00	0.62 - 0.80	1.00 - 1.00
8	Any other ethnic group	0.18	0.97	0.49	0.88
9	Not Given	n/a	n/a	n/a	n/a



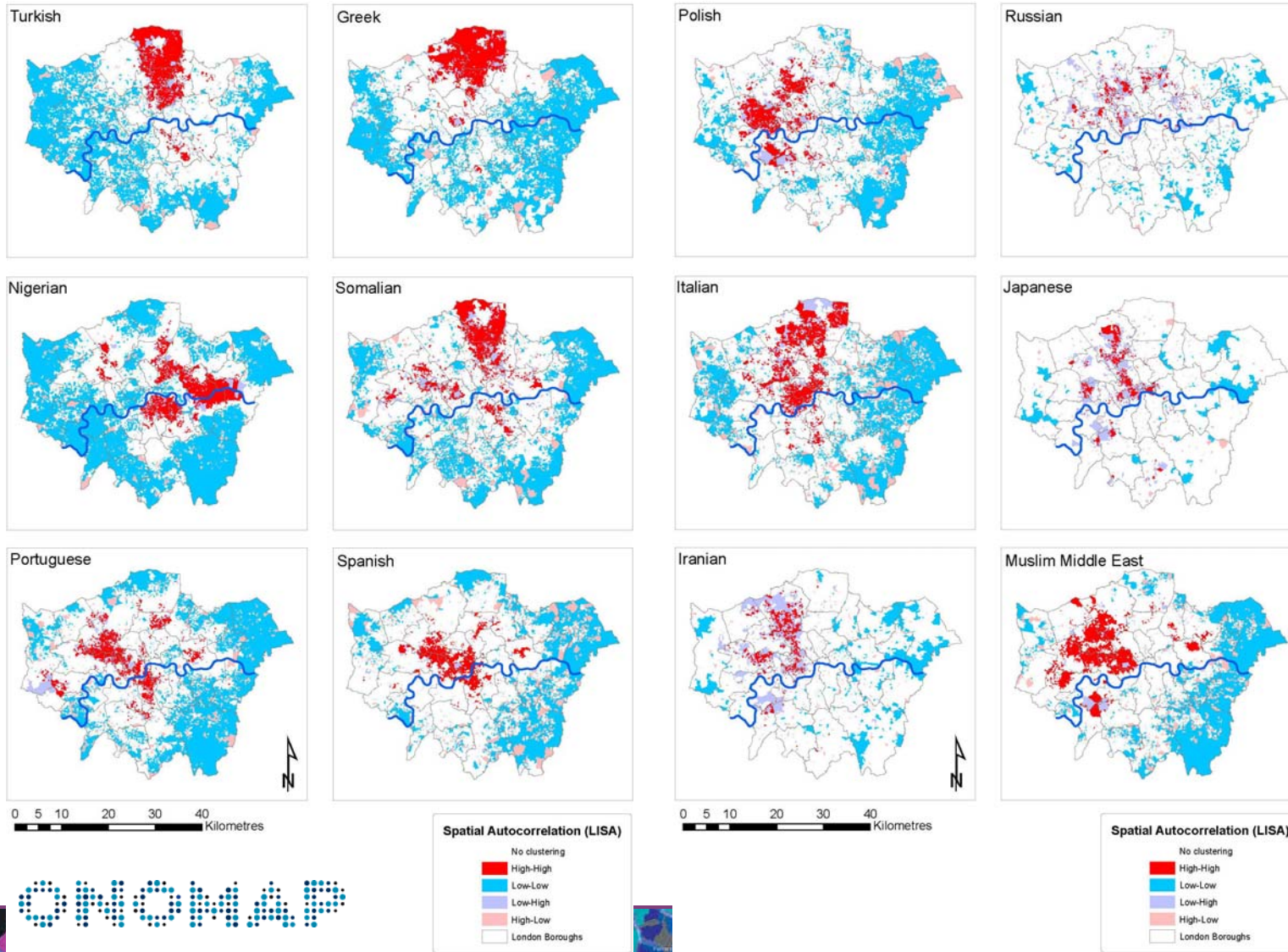
Characteristics:

- Technical:
 - Stand-alone Java application
 - Works without installation or internet connection
 - Preserves confidentiality of names data
 - Fast performance (13,000 names per second)
- Names profiling
 - 1 million names in encrypted dictionaries
 - Onomap Type assigned to individual based on probability scores
- Feedback & Statistics
 - Facility to report unclassified names and provide feedback
 - Statistics to compare datasets with base populations

Greek names in London (2004)



Spatial Autocorrelation (LISA)



Map of ethnic diversity in London

Multigroup Entropy Index (H) at Output Area level

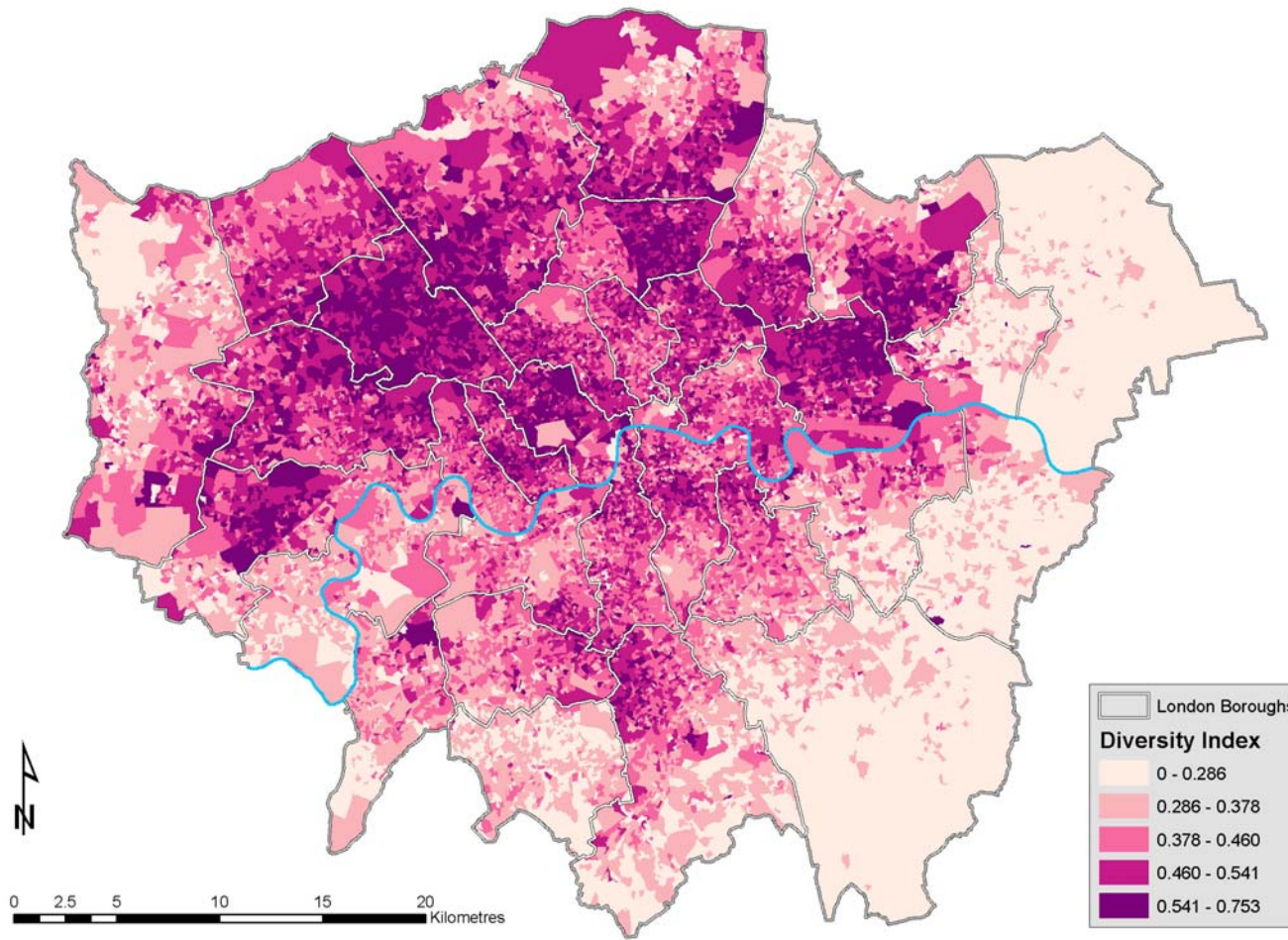
$$H = -\sum_{i=1}^n \frac{\left(\frac{P_{ij}}{P_j}\right) \ln\left(\frac{P_{ij}}{P_j}\right)}{\ln n}$$

where:

n = number of groups

P_{ij} = Population of group i in spatial unit j

P_j = Sum of population of all groups 1 to n in spatial unit j



ONOMAP

applications

- **Population Studies**
 - ethnic group population forecasting by small area
 - monitoring migration by small area
 - diagnosing possible Census undercount
 - study of residential segregation
- **Analysis of ethnic inequalities in health**
- **Improving public and private services to ethnic minorities**
- **Auditing accessibility to those services**
- **Evaluating equal opportunity policies**
 - (e.g. through profiling payroll with Onomap)
- **Geographical analysis of ethnicity and migration**



- Available soon
- For details keep an eye on www.onomap.org
- Or e-mail p.mateos@ucl.ac.uk

Thanks for listening!