



Age, gender, and... *ethnicity*?

How to segment populations by a slippery dimension in European multicultural geographies.

Pablo Mateos
Richard Webber

*Int'l Population Geographies
Conference
Liverpool
19-21 June 2006*

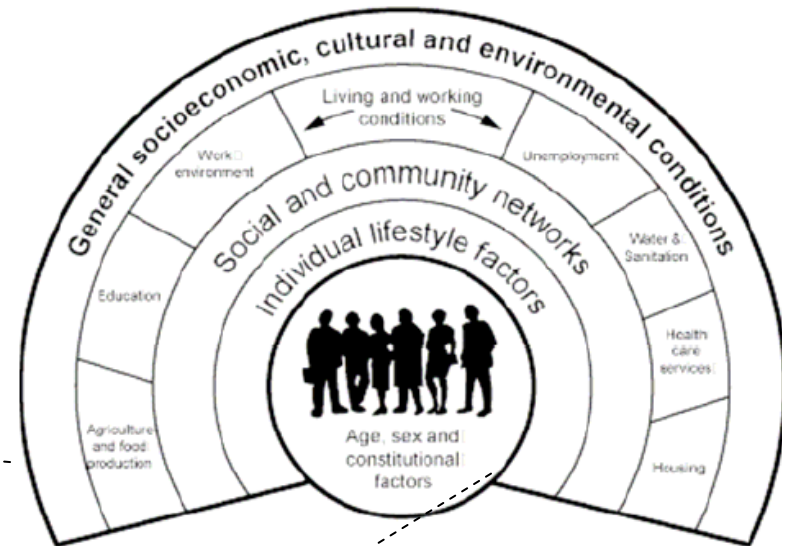
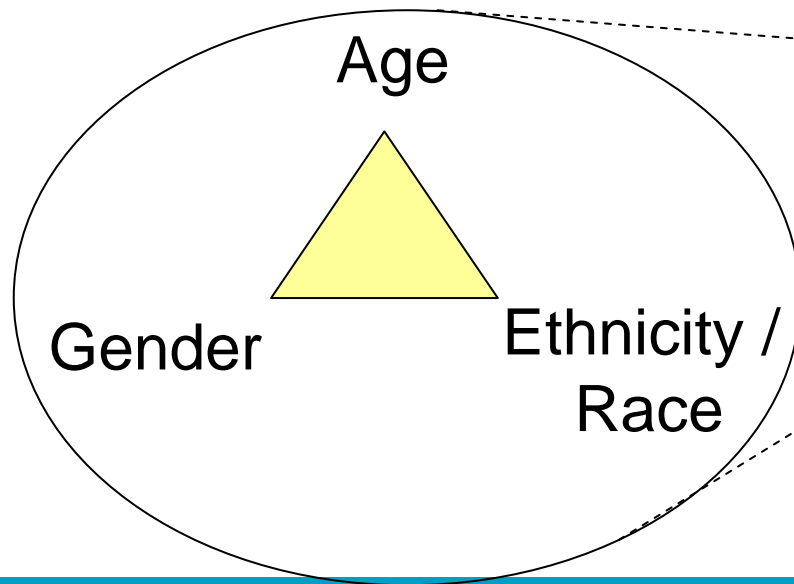
*Centre for Advanced Spatial Analysis (CASA)
Department of Geography
University College London
p.mateos@ucl.ac.uk*

Contents

1. Defining ethnicity
2. Measuring ethnicity
3. Name origin analysis
4. Applications & evaluation
5. Conclusions

The demographic triad

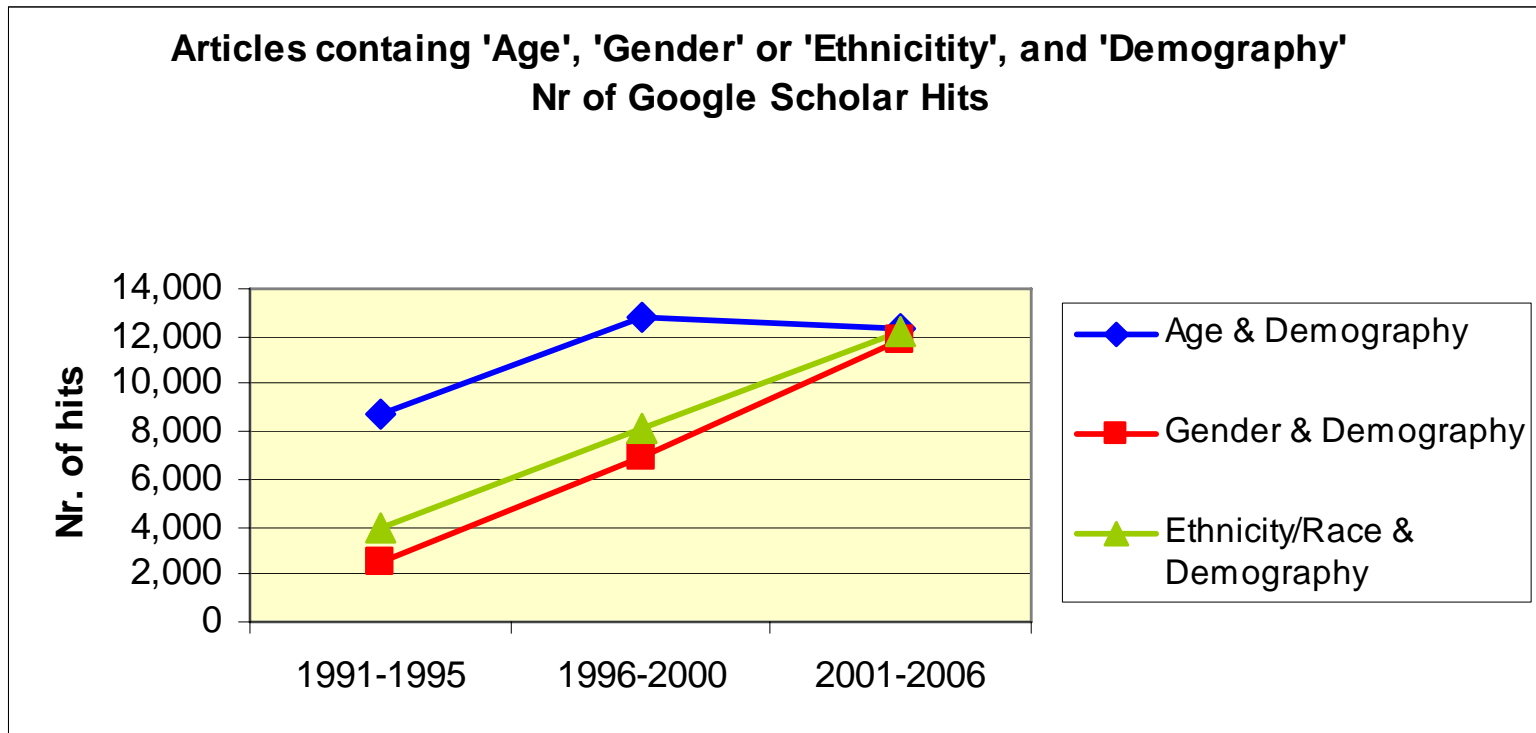
- Core constituents of a person
(conceived as unmutable over lifecourse)



*A model of the main determinants of health
(Whitehead, 1995)*

The demographic triad

- Gender & Ethnicity accompany Age in demographic research



1 – Defining ethnicity

Ethnicity & Race

Ethnicity \neq Race

Science
magazine

“125 big questions that face scientific inquiry over the next quarter-century”

What are human races, and how did they develop?

Anthropologists have long argued that race lacks biological reality. But our genetic makeup does vary with geographic origin and as such raises political and ethical as well as scientific questions.



Biological determinism

Mitchell's New Primary Geography.

THE SECOND BOOK OF THE SERIES.

THE NEW PRIMARY GEOGRAPHY:

ILLUSTRATED BY

TWENTY COLORED MAPS

AND

EMBELLISHED WITH A HUNDRED ENGRAVINGS.

The people in the world are divided into five principal races, named according to their color and residence.

What is known of the White race?

The White race is superior to the others, and is found in Europe and America.

Geography of Races
(Mitchell, 1868)

*An Eurocentric White man
view of the world*

OF THE PEOPLE WHO INHABIT THE EARTH.

How many people are there in the world?

There are upwards of a billion—(1,000,000,000.)

How are the people in the world divided?

The people in the world are divided into five principal races, named according to their color and residence.

Name the five races in the world.

The five races are the White or Caucasian; the Yellow or Mongolian; the Black or African; the Brown or Malay; and the Red or American.

What is known of the White race?

The White race is superior to the others, and is found in Europe and America.



THE WHITE RACE.

Of the Yellow race?

The Yellow race is found in Asia; the best specimens are in China and Japan.

Of the Black race?

The Black race is found in Africa, and is commonly called the Negro race.

Of the Brown race?

The Brown race inhabits the islands of the Pacific Ocean.



THE YELLOW RACE.
(A Chinese Laborer.)

THE BLACK RACE.
(An African Chief.)

Of the Red race?

The Red race includes the Indians of North and South America.



THE BROWN RACE.
(A New Zealand Chief.)

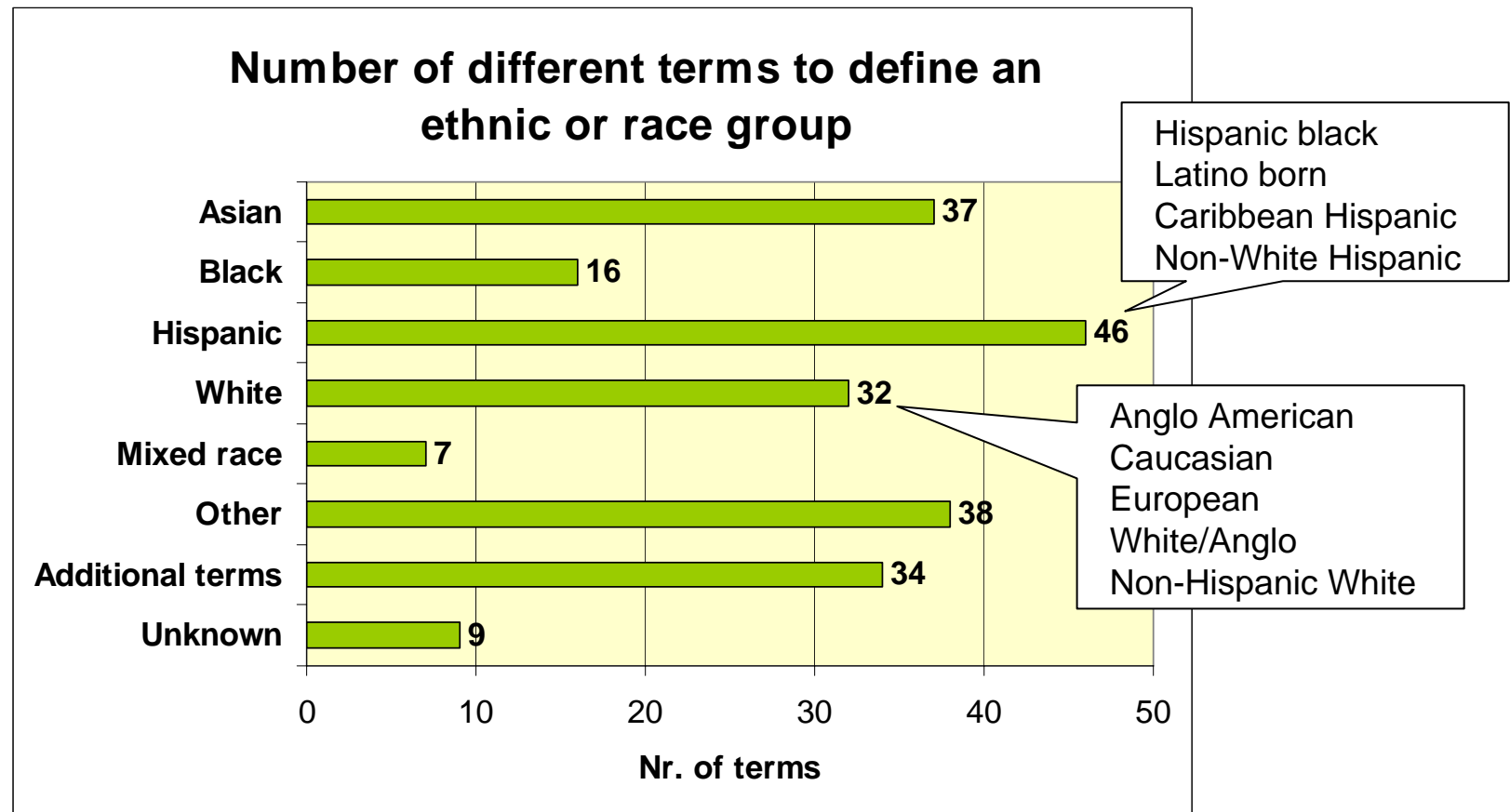
THE RED RACE.
(An Indian Chief.)

Modern concepts of Race & Ethnicity

- Consensus in that both concepts are socially constructed
- The word 'ethnicity' derives from the Greek word *ethnos*, meaning a nation. Thus, the basis of nationalism.
- Max Weber (1922)
 - **Race group**: *A group perceived as having common inherited and inheritable traits that derive from common descent*
 - **Ethnic groups**: *Those human groups that entertain a subjective belief in their common descent because of similarities of physical type or of customs or both, or because of memories of colonization and migration (...)*
- A firm belief in group's affinity is required for ethnic groups to be defined in opposition to other groups differently perceived and with whom contact is required (Eriksen, 2002)
- The characteristics that define ethnicity are not fixed or easily measured, so ethnicity is imprecise and fluid (Senior & Bhopal, 1994)

2 – Measuring ethnicity

Different terms, different ethnicities



219 terms for 8 'Ethnic Groups' in 1,198 articles published
in 2 American epidemiology journals 1996-99
(Comstock et al, 2004)

UK 2001 Census Ethnicity Classification

	UK	London
White		
British	87.5%	59.8%
Irish	1.2%	3.1%
Other White	2.6%	8.3%
Mixed		
White & Black Caribbean	0.5%	1.0%
White & Black African	0.2%	0.5%
White & Asian	0.4%	0.8%
Other Mixed	0.3%	0.9%
Black or Black-British		
Black-Caribbean	1.1%	4.8%
Black-African	0.9%	5.3%
Black-Other	0.2%	0.8%
Asian or Asian-British		
Indian	2.0%	6.1%
Pakistani	1.4%	2.0%
Bangladeshi	0.5%	2.1%
Any other Asian background	0.5%	1.9%
Chinese or other group		
Chinese	0.4%	1.1%
Any other ethnic group	0.4%	1.6%

	UK	London
Total Non- White British	12.5%	40.2%
Poorly Studied Groups	4.9%	18.8%

- 16 Categories
- Strongly based on a “skin colour problem”
- Confusing question

Source: ONS Census 2001

London 'non-16+ ethnic groups'

(1.2 million people stated 'other' ethnic identities in London 2001 Census)

Ethnic Group	Population
Other white European, European Mixed	185,690
Other white, white unspecified	171,744
English	154,203
Sri Lankan	53,307
Black British	46,348
Turkish	37,827
Italian	35,252
Other Mixed, Mixed unspecified	35,027
Any other group	29,469
Greek Cypriot	23,340
Middle Eastern (excluding Israeli, Iranian and 'Arab')	20,537
Arab	20,256
Filipino	19,669
Japanese	19,415
Other mixed white	19,239
Other Asian, Asian unspecified	18,334
Greek	17,888
Iranian	16,494
Multi-ethnic islands	15,952
Polish	15,928
South and Central American	15,607
British Asian	14,625
Turkish Cypriot	14,074

Ethnic Group	Population
Vietnamese	11,719
Commonwealth of (Russian) Independent States	11,606
North African	11,218
Kurdish	9,659
Latin American	9,188
Mixed Black	9,001
Jewish	8,912
Other Black, Black unspecified	8,344
Cypriot (part not stated)	7,360
Mixed: Irish and other white	7,071
Scottish	7,020
Kosovan	6,896
Welsh	6,895
Somali	6,172
East African Asian	5,328
Chinese and White	4,871
Tamil	4,758
Black and White	4,226
Moroccan	4,133
Caribbean Asian	4,070
Black and Asian	3,946
Malaysian	3,384
Albanian	3,226
Sikh	2,814

(.../...)

Source: 2001 Census GLA commissioned tables

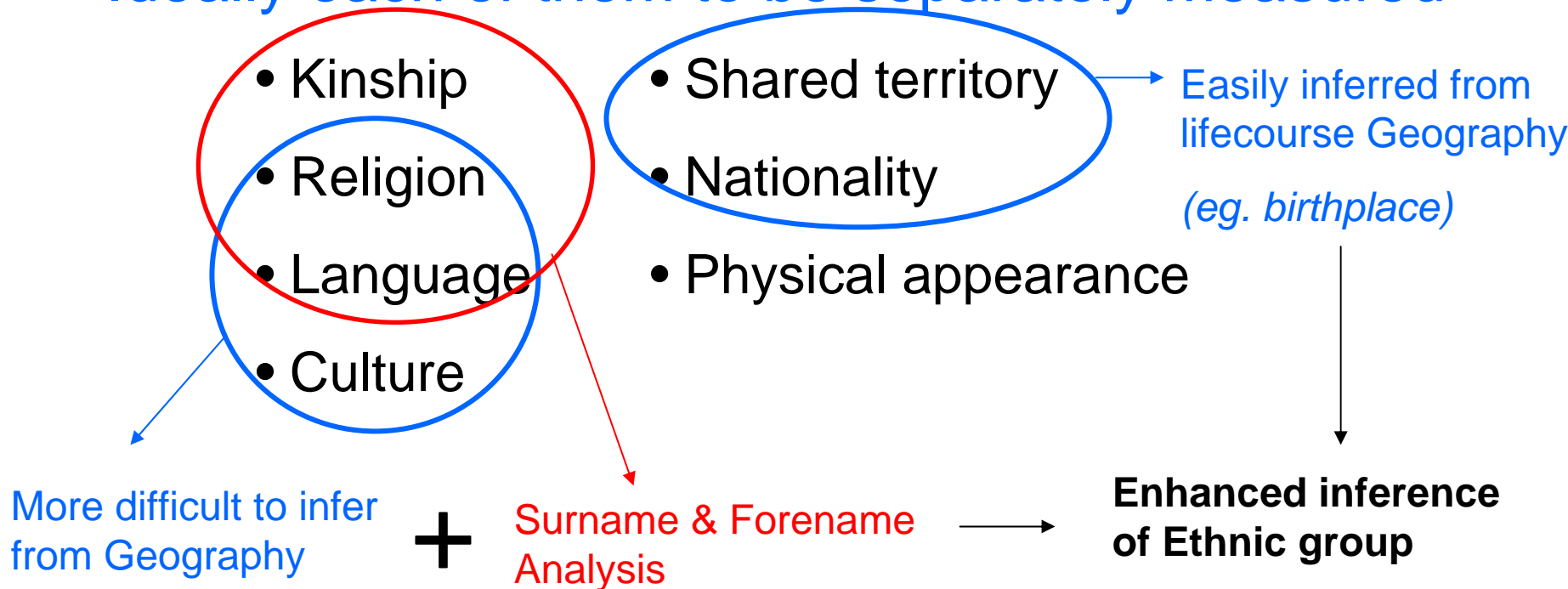
Sources of Ethnicity data

- Current information sources available (UK):
 - Census of Population (decennial, aggregated)
 - Official Surveys (few ethnic minorities represented)
 - Hospital Admissions (low quality)
- Problems of collecting ethnicity data:
 - Sensitive data – low accuracy, low coverage
 - Changing categorizations
 - Changing identities
 - Not always self-assessed (e.g. hospital, deaths)
 - Tries to measure too many things into one variable
- Result in a poor understanding of ethnicity

Multidimensionality of ethnicity

- **Ethnicity**: A multi-dimensional concept that encompasses different aspects of identity:

Ideally each of them to be separately measured



3– Name origin analysis

Names origins & Ethnicity

- Identity, though complex, can be encoded in a name (Seeman, 1980)
- Names can potentially provide information about:

<i>Aspect</i>	Etimology/ Onomastics	Space-time Distribution
Surname & Firstname	Language	Geographic Origin
	Religion	Migration flows
Firstname	Gender	Age

- Used since the 1950s in epidemiological and genetics studies to subdivide populations (Word & Perkins, 1996; Lasker, 1985)
- Hispanics, South Asians, Chinese, Muslim Names

Name analysis in genetic research

- Surnames generally adopted in the Middle Ages (Europe)
- Surnames in genetic studies dates back to 1875; George Darwin (son of Charles Darwin) used surname frequency to study population inbreeding
- Today surnames are used to study ancient patrilineal population structures (Manni et al 2005)

Assumptions:

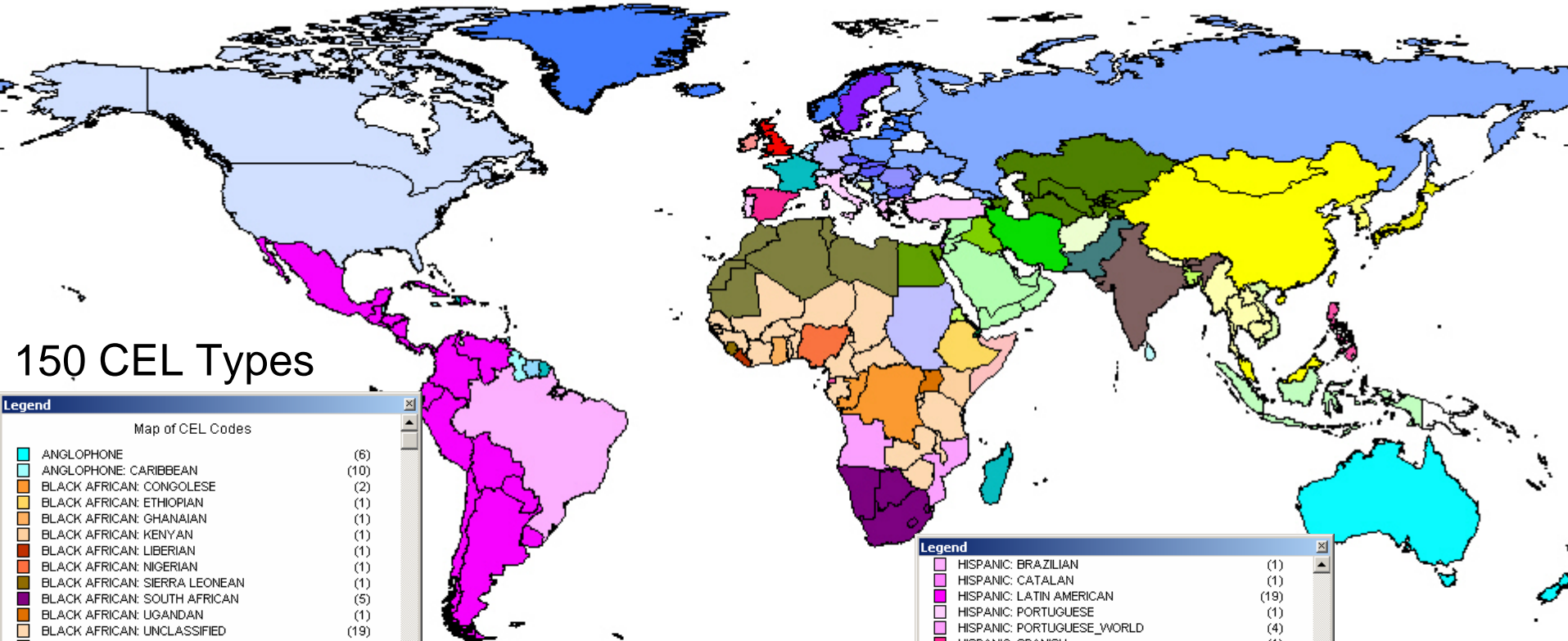
- Low intermarriage
- Low infidelity
- Common origin (monophyletic)
- Low name change rate

Cultural Ethnic Linguistic (CEL) classification

- 250,000 Family Names and 120,000 Personal Names coded by CEL Type
- +150 CEL Types aggregated into 15 CEL Groups

CEL GROUP	NR. CEL TYPES	FAMILY NAMES	PERSONS	PERSONS/ F. NAME
ENGLAND	5	86,289	30,856,110	358
CELTIC	5	19,356	10,548,055	545
EUROPEAN	32	41,035	973,590	24
MUSLIM	14	17,758	952,146	54
SOUTH ASIAN	12	8,904	467,455	52
SIKH	1	3,237	316,337	98
EAST ASIAN	11	1,218	170,032	140
HISPANIC	10	6,180	169,258	27
UNCLASSIFIED	2	13,696	155,330	11
AFRICAN	17	6,441	144,540	22
JEWISH AND ARMENIAN	5	1,960	118,099	60
GREEK ORTHODOX	3	6,609	98,958	15
VOID	4	811	89,872	111
NORDIC	6	3,416	46,536	14
JAPANESE	1	1,482	6,322	4
TOTAL	128	218,392	45,112,640	207

World map of CEL types



150 CEL 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)
JEVISH	(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	(4)
MUSLIM: NORTH AFRICAN	(6)
MUSLIM: PAKISTANI	(1)
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)

Main methods used to classify names

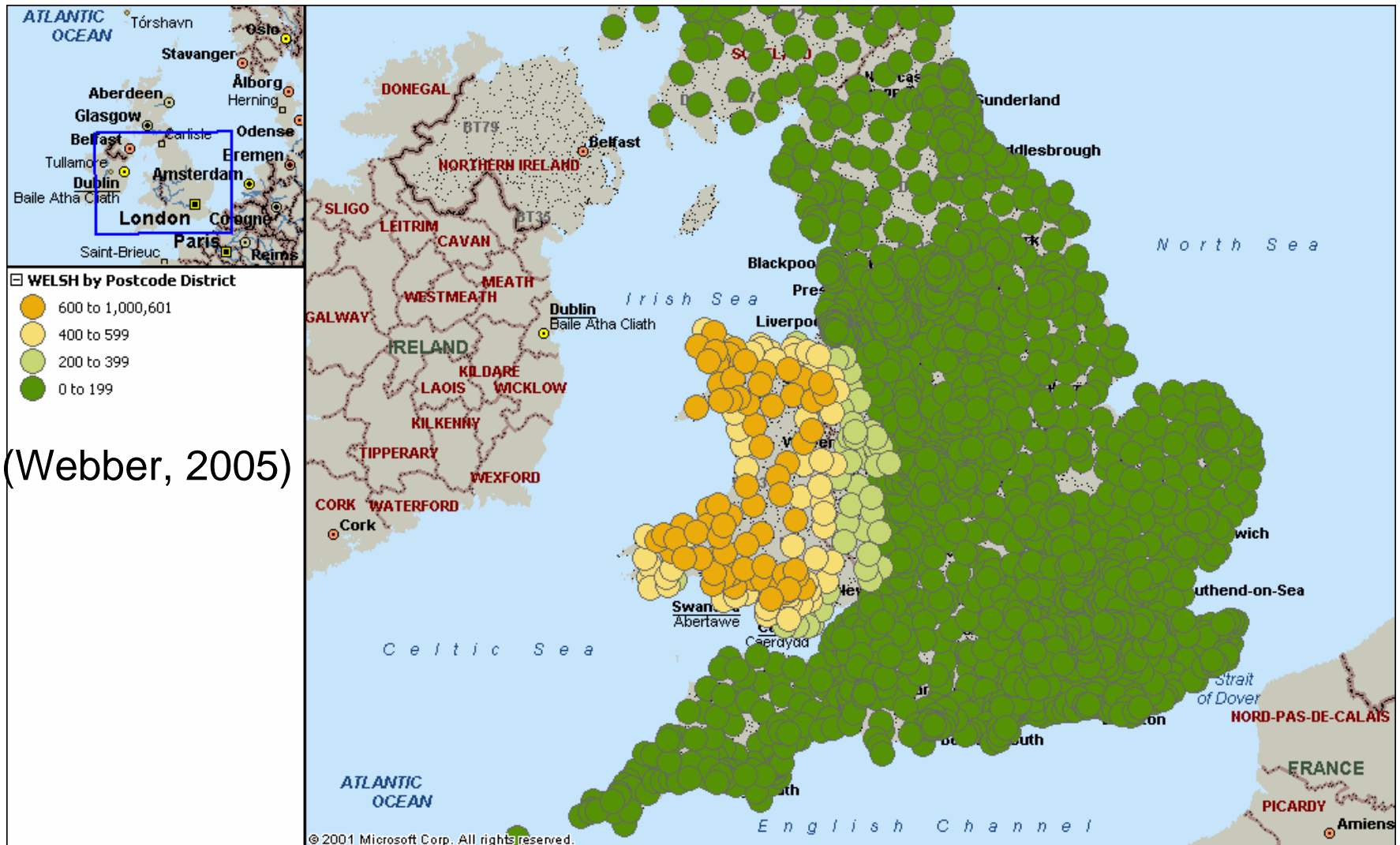
1. 'Correspondence analysis' between personal and family names
2. Census and Geodemographic area data
3. Geographical distribution & clustering
4. Text mining
5. Birthplaces & names
6. Lists of names by country
7. 'Googling' individual names

Issues with Names Analysis

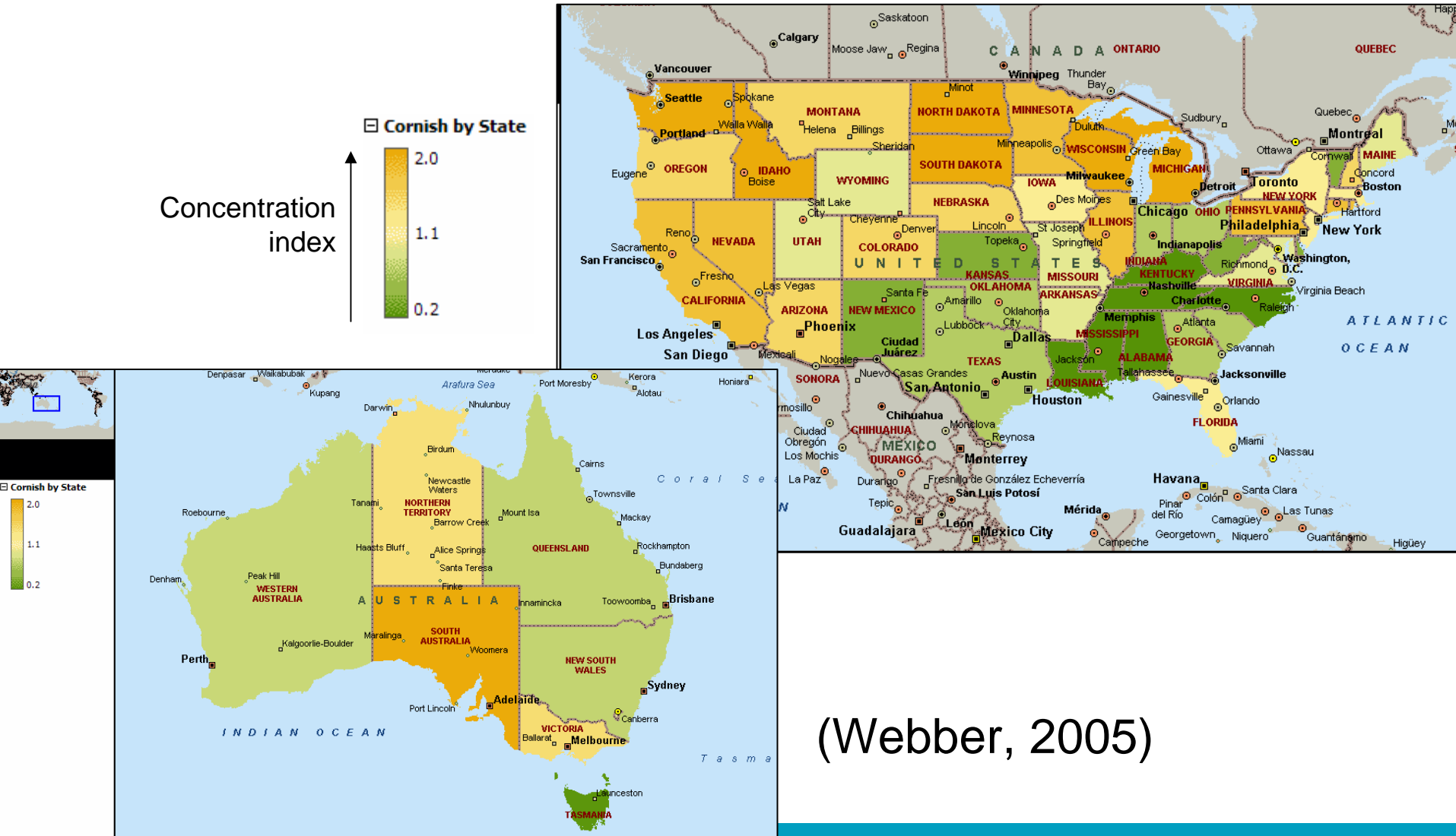
- Only reflects patrilineal heritage
- Different history of surname adoption, naming conventions & surname change
- Name normalisation is required
- Family/Household Autocorrelation
- Limited names lists, due to temporal & regional differences in name distribution
- Lack of consistency in self-conceived identity

(Senior & Bhopal, 1994; Martineau 1998, Word & Perkins, 1996; Jobling 2001)

2004 Electors with 'Welsh' surnames

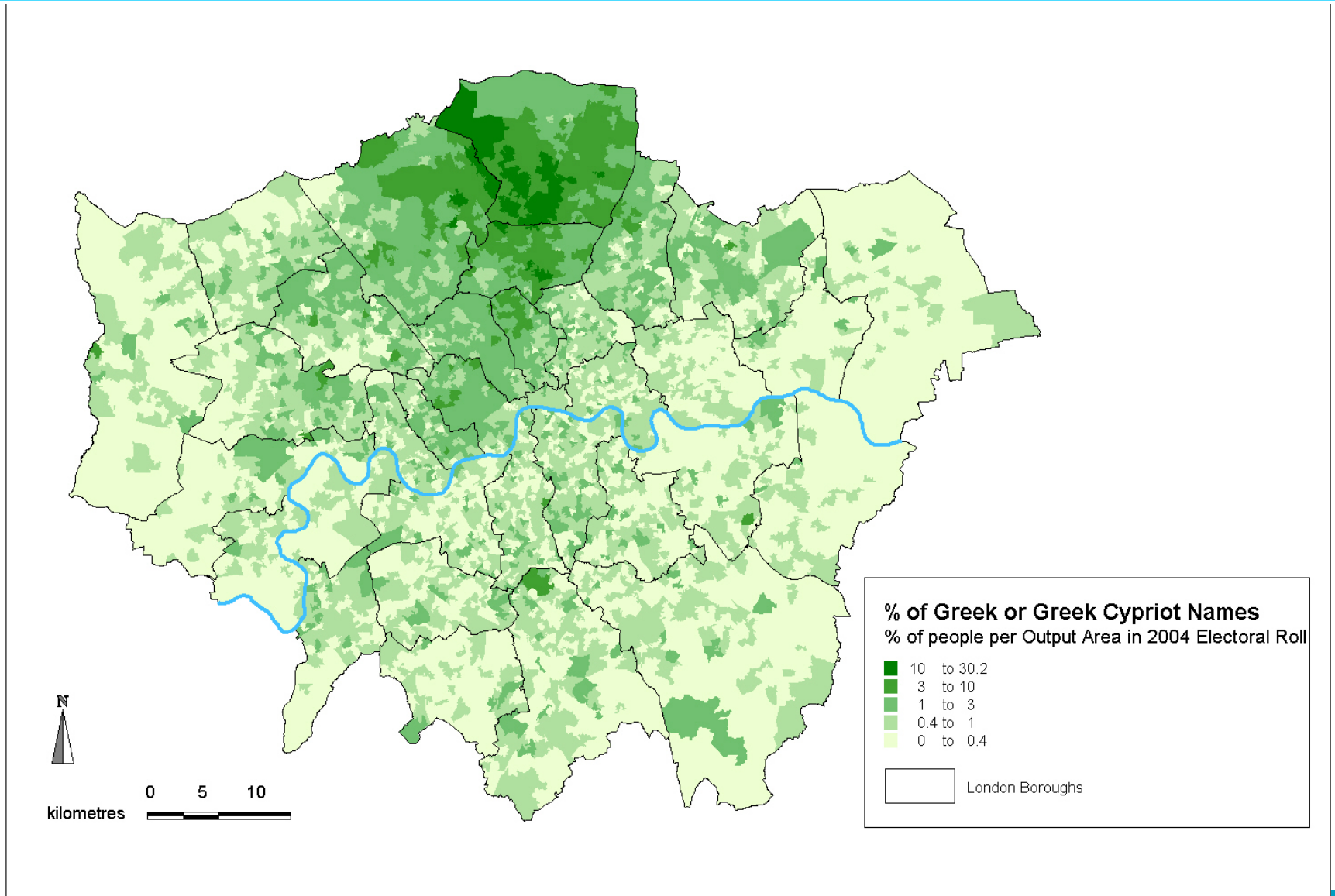


'Cornish' names & Anglosaxon diaspora

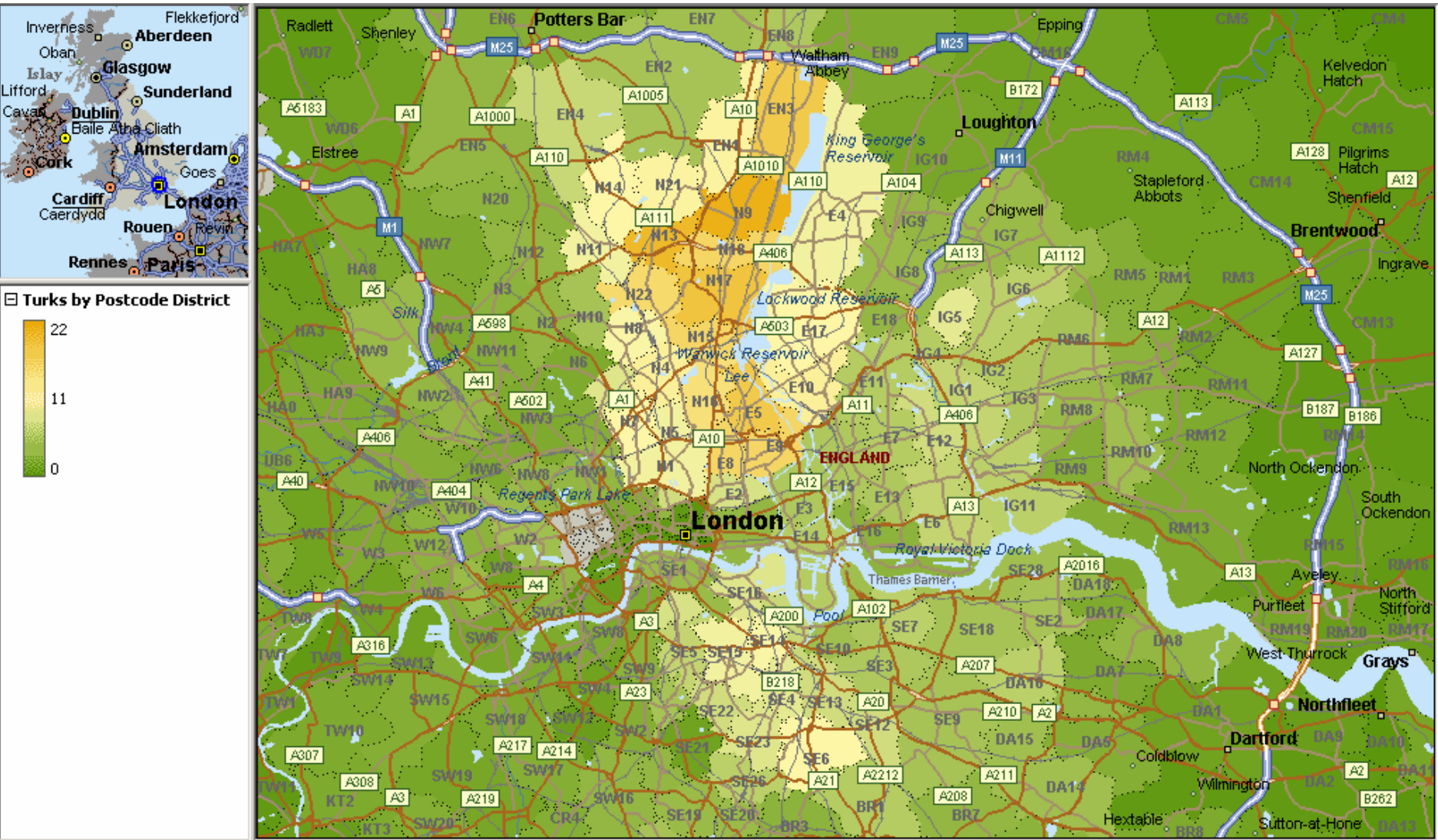


(Webber, 2005)

Greek & Greek Cypriot names in London



Turkish names in Greater London



4 – Applications & Evaluation

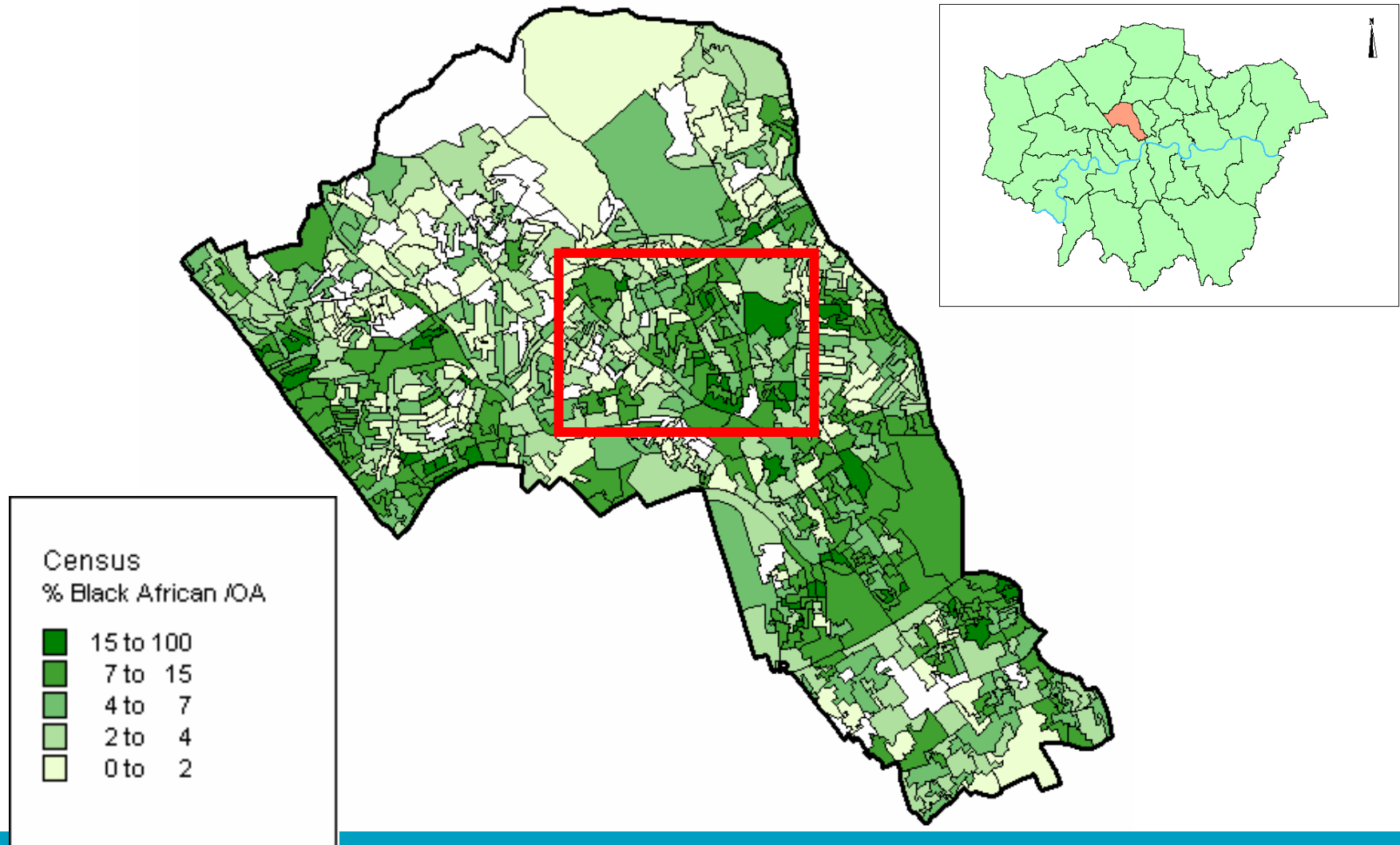
Applications of the CEL classification

- UCL analysis
 - Determining local associations of ethnic inequalities in health Camden PCT (London)
 - Classifying the UK 1881 Census, UK 2004 electoral roll, and 2004 Spanish Telephone directory.
 - Measuring ethnic residential segregation in London
- Other users in the public sector:

Application	User
Identifying new migrants in Emergency Admissions	Princess Royal Hospital, Telford
Evaluating uptake of breastfeeding promotion service by ethnic group	Heart of Birmingham PCT
Assessing the quality and coverage of social services population register by ethnic group	Hackney Council
Ethnic sample boost in a Household Survey of Entrepreneurship	IFF Research Ltd
Isolating genetic origin of new TB strains	Health Protection Agency
Completing and evaluating current patient profiling database	Liverpool PCT

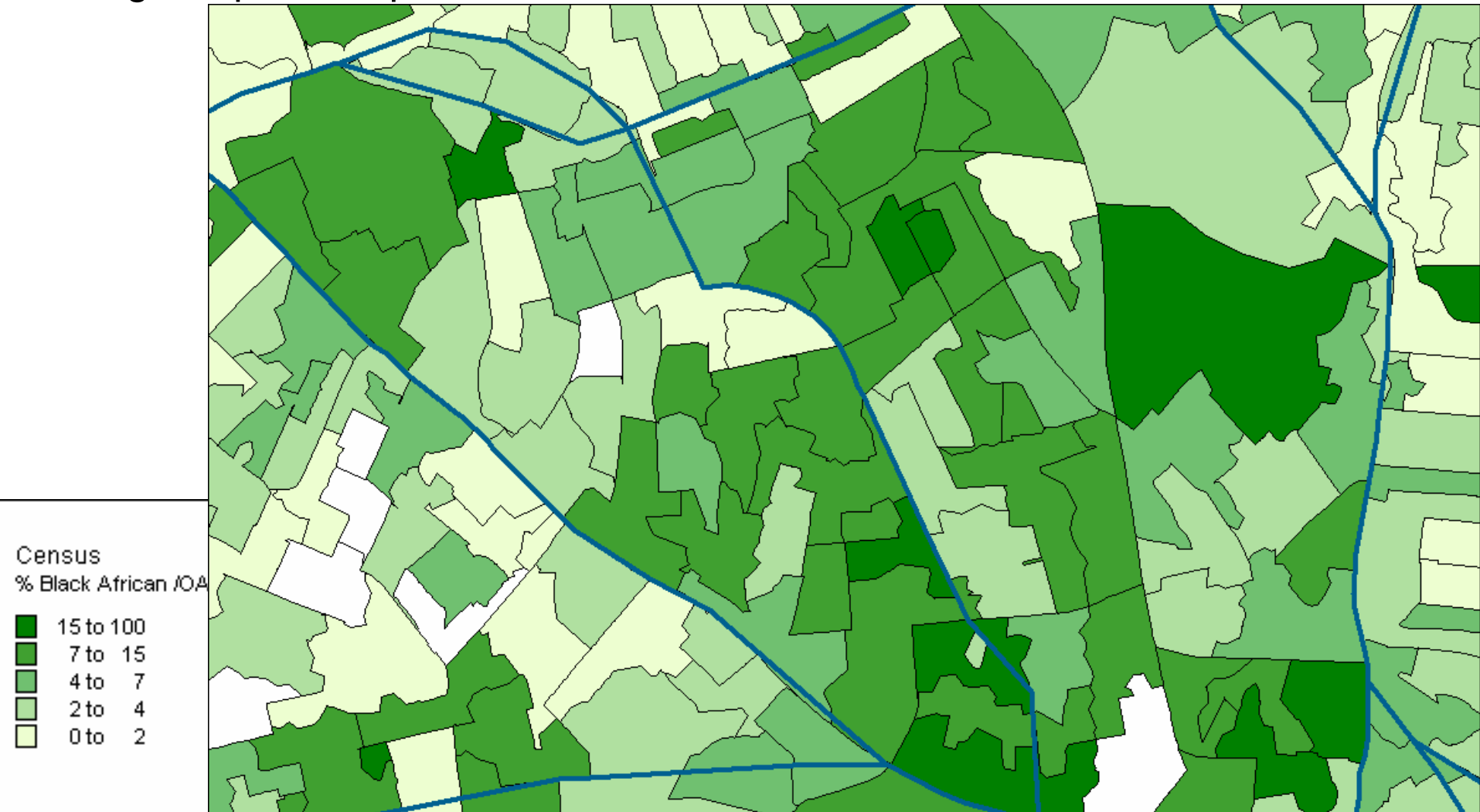
Census Vs CEL

Black African ethnicity in Camden



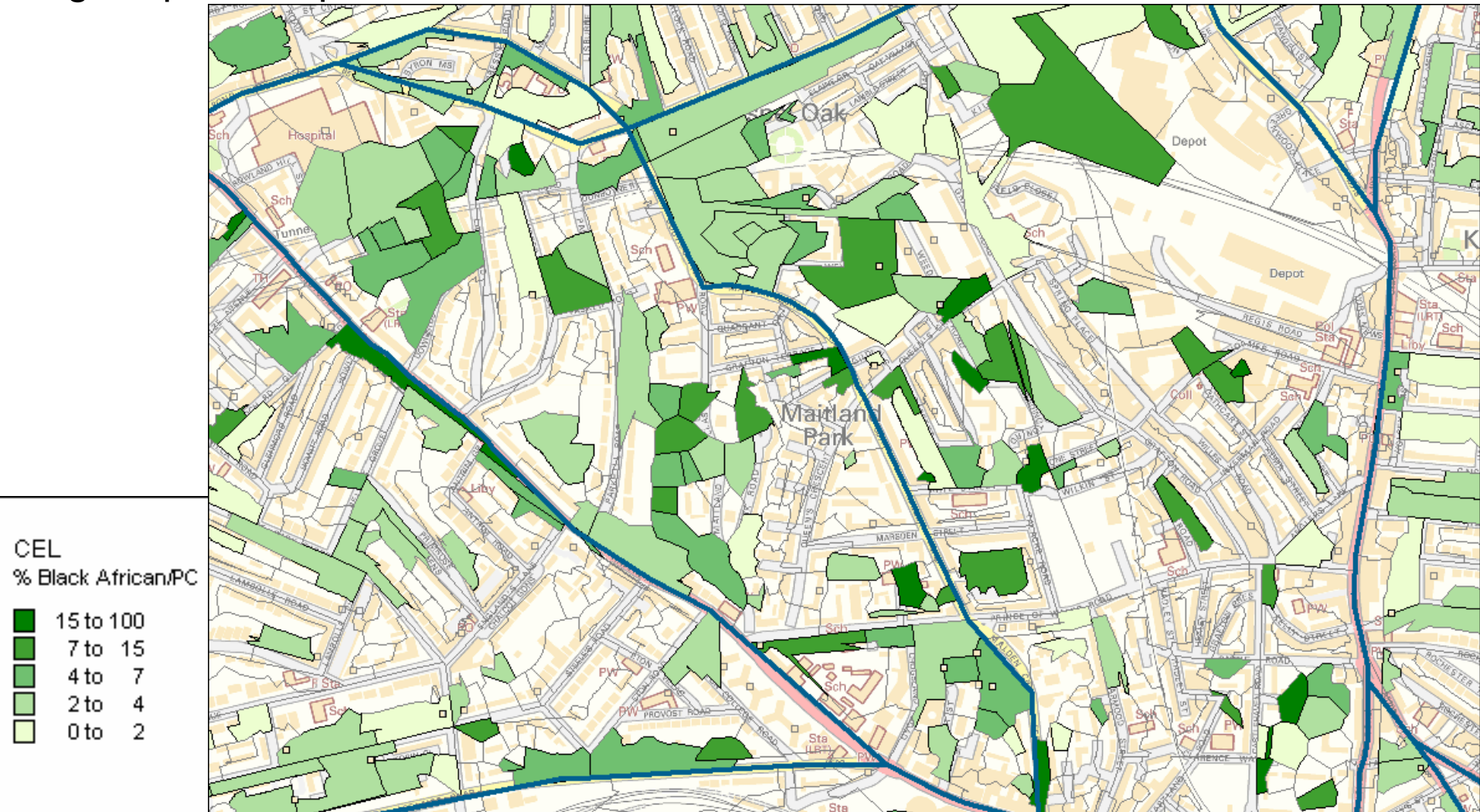
Census 'Black African' by Output Area (OA)

Average Population per OA: 285



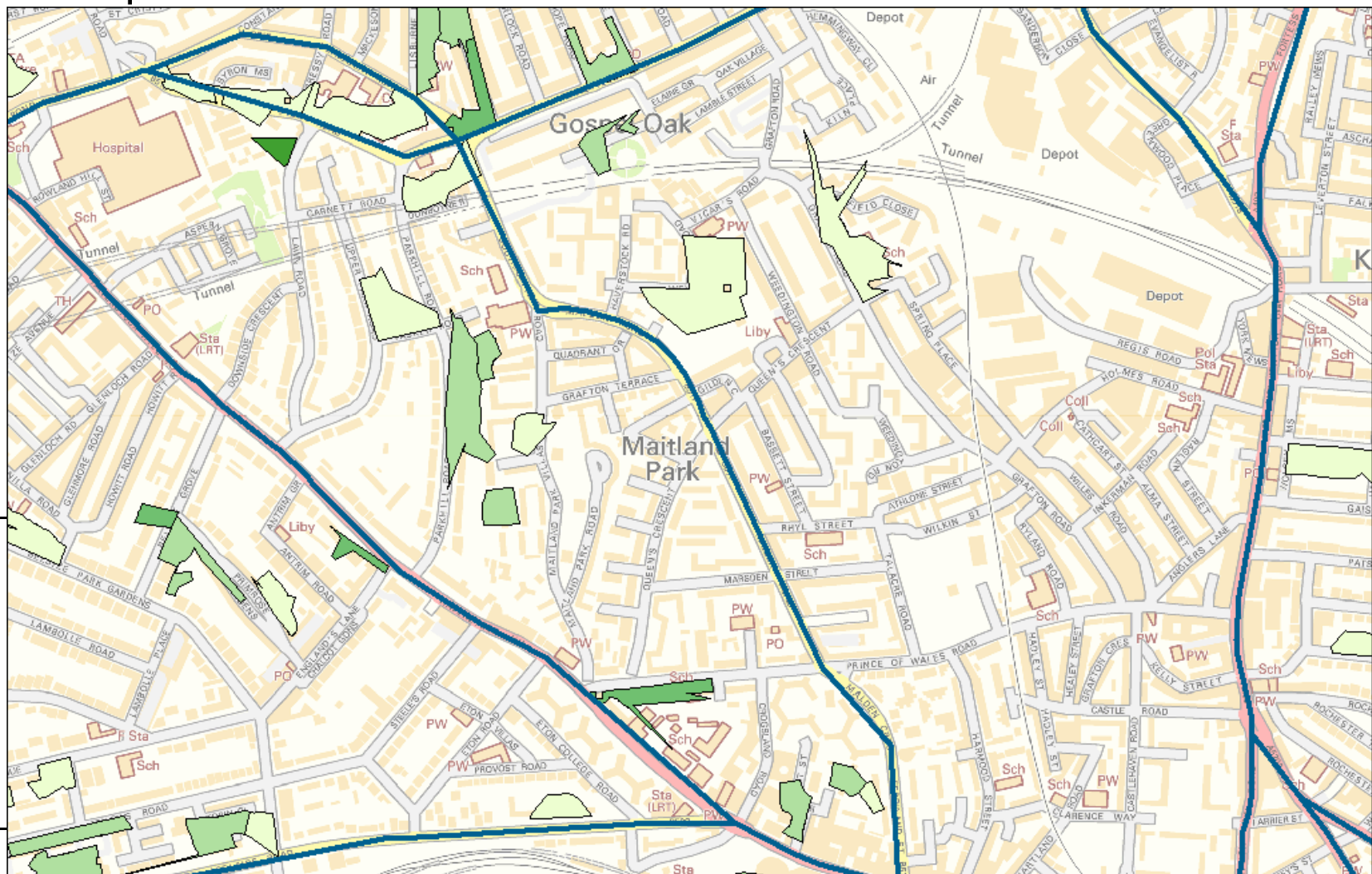
CEL 'Black African' by Postcode

Avg. Population per Postcode: 54



CEL 'Somali' by Postcode

Avg. Population per Postcode: 54



CEL
% Somali /PC

- 15 to 100
- 7 to 15
- 4 to 7
- 2 to 4
- 0 to 2

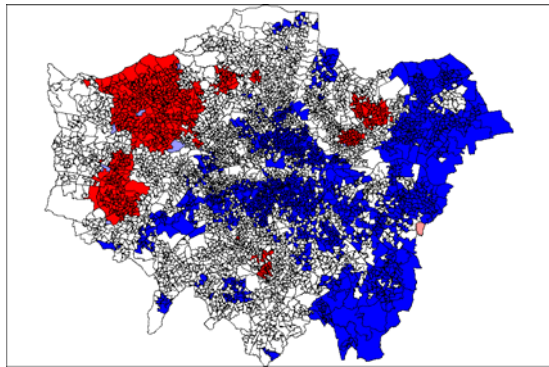
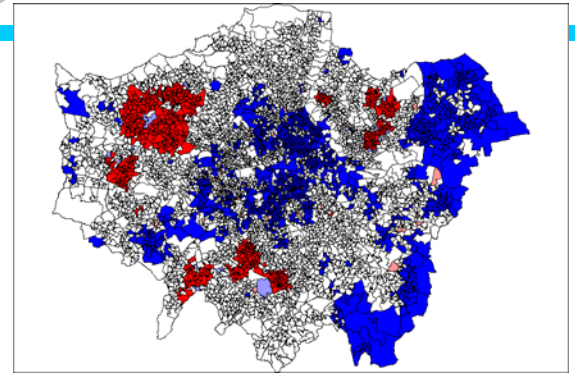
CEL Clusters in London by LSOA

(3) LISA Cluster Map (LSOA)

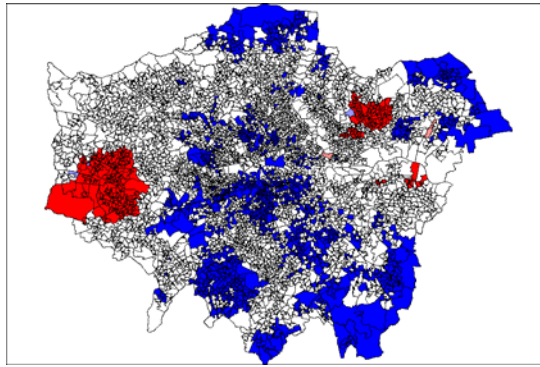
- Not Significant
- High-High
- Low-Low
- Low-High
- High-Low

Local Indicators of Spatial Association (LISA)
(Anselin, 1995) using
GeoDA

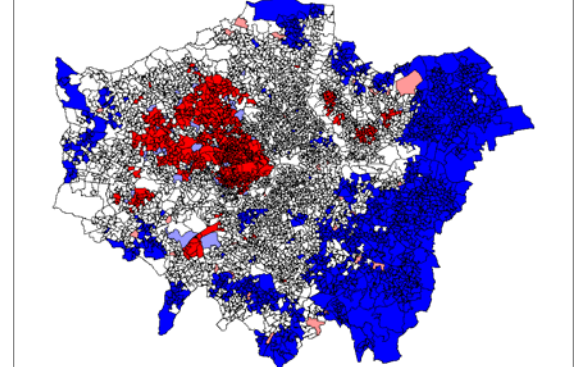
Somali



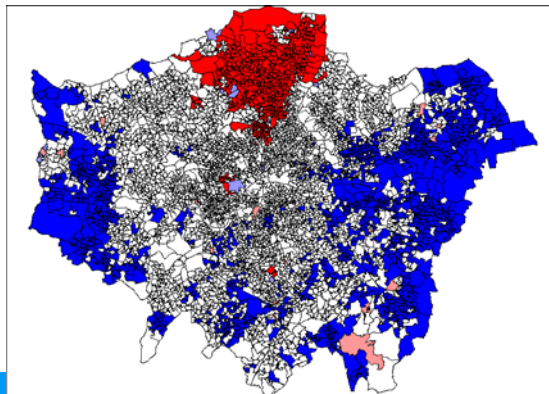
Hindu



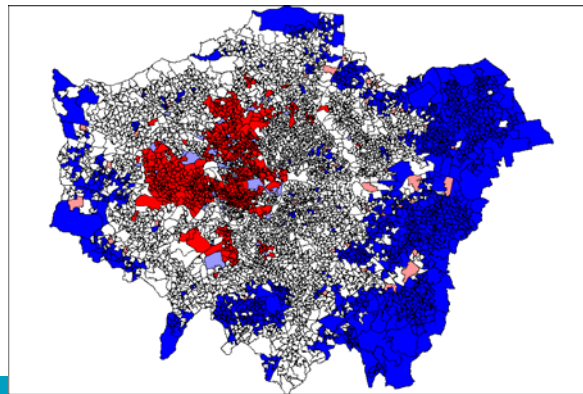
Sikh



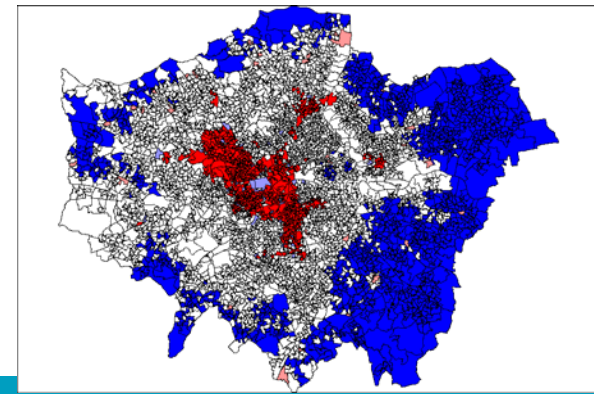
Other Muslim



Greek & G. Cypriot



Eastern Europe



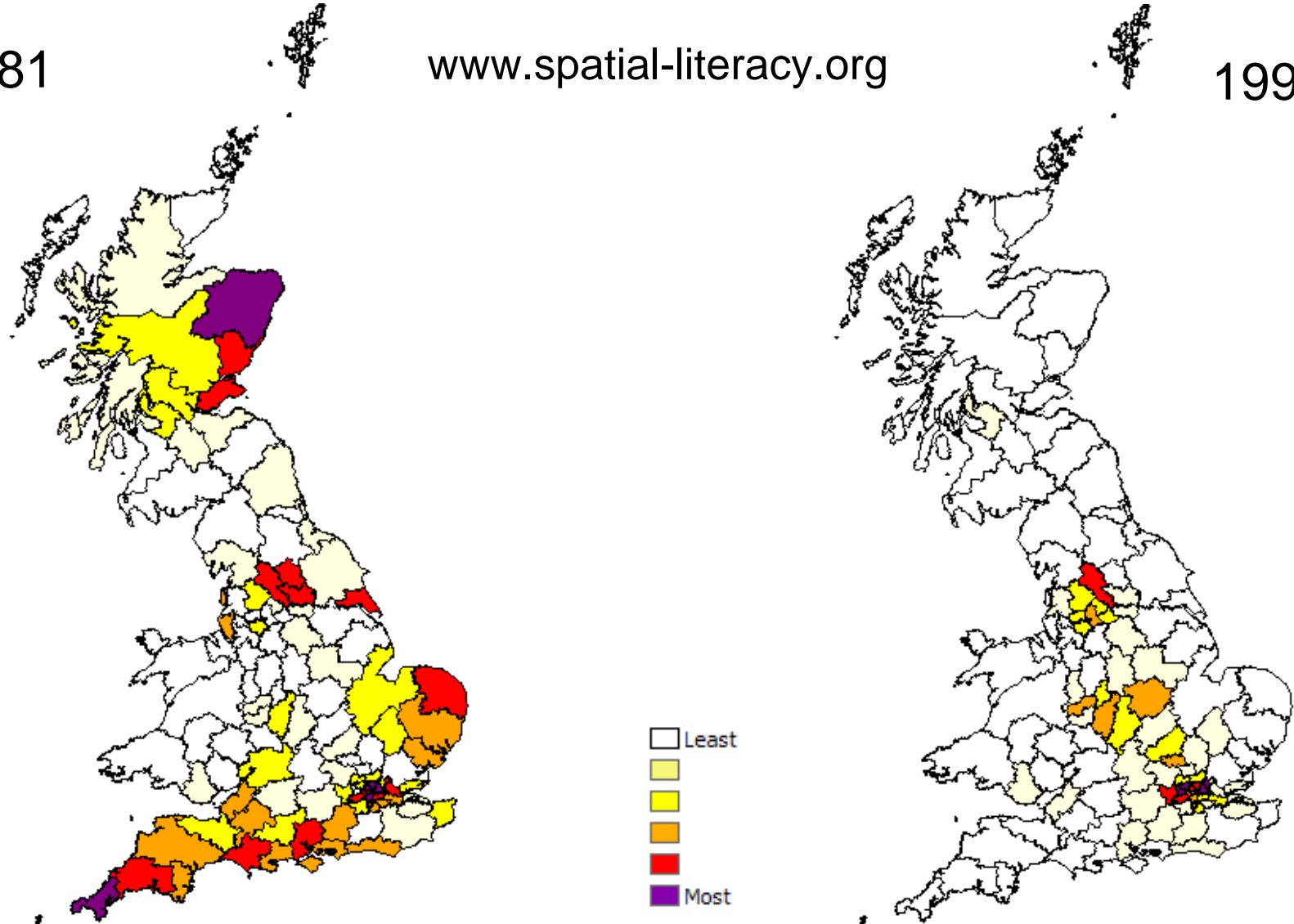
Hispanic

Distribution of Non-British Surnames 1881-1998

1881

www.spatial-literacy.org

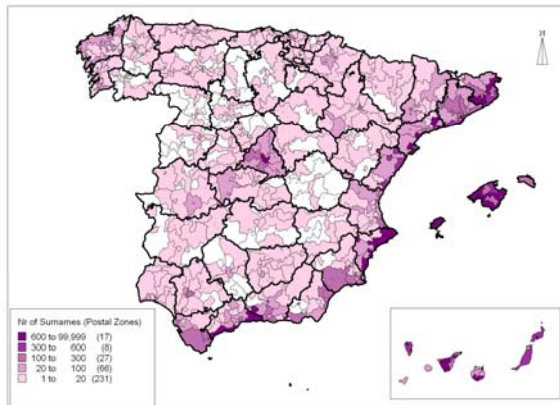
1998



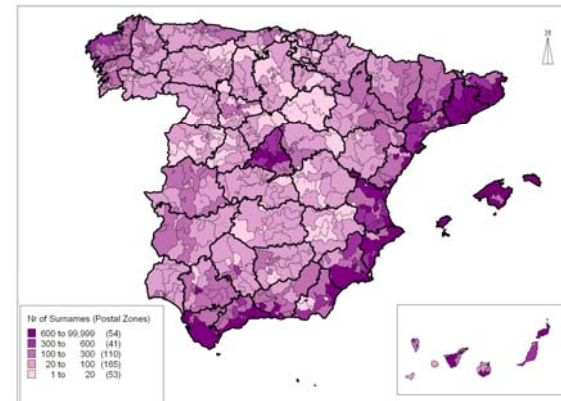
Ethnicity & Migration in Spain

- Name origins in the telephone directory

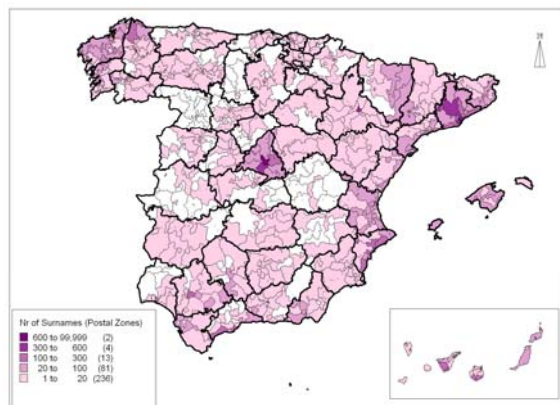
Germany & Austria



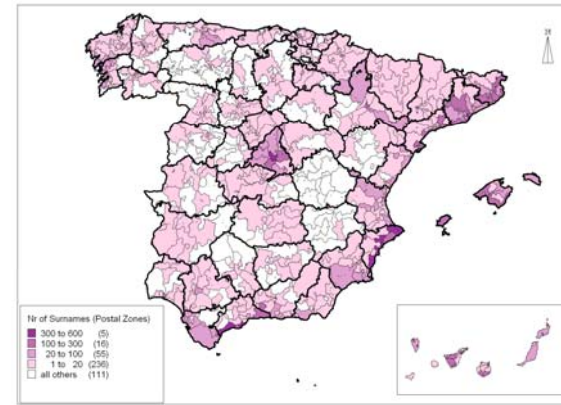
Britain & Ireland



China



Poland



Correlations CEL vs Census (London)

Census 2001 Ethnic Group	CEL Group Aggregation	Correlation at Geographical Levels			
		OA	LSOA	Ward	Borough
White British	British + Jewish	0.83	0.86	0.89	0.91
White Irish	Irish	0.01	0.05	0.16	0.34
Other White	W. & E. European + Hispanic + Greek & G. Cypriot	0.75	0.85	0.90	0.94
White and Black Caribbean	Black Caribbean	0.26	0.55	0.81	0.92
White and Black African	Black African + Somali	0.27	0.47	0.58	0.67
White and Asian	<i>Not Assigned</i>				
Other Mixed	<i>Not Assigned</i>				
Indian	Hindu + Sikh	0.96	0.98	0.99	1.00
Pakistani	Pakistani	0.06	0.09	0.12	0.06
Bangladeshi	Bangladeshi	0.94	0.98	0.99	1.00
Other Asian	Sri Lankan	-0.01	-0.04	-0.07	-0.17
Black Caribbean	Black Caribbean	0.47	0.75	0.94	0.99
Black African	Black African	0.78	0.87	0.89	0.93
Other Black	Avg. of Black Carib + African	0.40	0.64	0.80	0.87
Chinese	Chinese	0.64	0.78	0.89	0.90
Other ethnic group	Other Muslim + Japanese	0.34	0.43	0.56	0.76
Total Population (Ethnicity Question)	Total Adults (Persons in the CEL file)	0.63	0.73	0.90	0.99
	Avg. Persons / Geog. Unit	285	1,443	10,931	208,011
	Nr. of Geographical Units	24,100	4,758	628	33

Evaluation at the individual level

- Evaluation of the CEL classification through self-reported ethnicity from Hospital Episode Statistics
 - 40,714 patients (20% of total) matched to a unique true ethnic code (1991 Census categories)
 - Problem of bad quality HES data

Predicted by CEL		Actual Ethnicity from HES data									Total	Sensitivity	Specificity	PPV
		0	1	2	3	4	5	6	7	8				
0	White	24,656	624	652	331	88	23	388	46	2,499	29,307	0.92	0.67	0.84
1	Black - Caribbean	35	147	3	15	3			1	35	239	0.17	1.00	0.62
2	Black - African	385	44	1,948	174	47	11	22	5	438	3,074	0.67	0.97	0.63
3	Black - Other										0	0.00	1.00	
4	Indian	426	15	17	8	333	16	12	2	150	979	0.13	0.99	0.44
5	Pakistani	19	1	3		22	75	11		29	160	0.32	1.00	0.47
6	Bangladeshi	96	5	59	37	132	75	2,672	1	292	3,369	0.84	0.98	0.79
7	Chinese	126	2	12	2	6	1	1	272	94	516	0.73	0.99	0.53
8	Any other ethnic group	1,046	19	196	64	67	36	87	44	1,511	3,070	0.30	0.96	0.49
Total		26,789	857	2,890	631	698	237	3,193	371	5,048	40,714			

5 – Conclusions

Conclusions: Review of CEL methodology

- Advantages
 - Finer spatial, temporal, and nominal scales
 - Can be applied to Population & Patient Registers, Telephone Directories, etc.
 - Reveals segregation of very detailed groups in London, such as Sikh, Jewish, Greek, Japanese, or Somali
- Challenges
 - Improvements to some categories in the name classification
 - CEL overlap for some names
 - Different CEL allocation for a name in different countries
 - Mixed ethnicities, name change, etc

Thank you!
Any Questions?

www.casa.ucl.ac.uk/pablo
p.mateos@ucl.ac.uk