

# What's in a (Kiwi) name?

David O'Sullivan

School of Geography, Geology and Environmental Science  
University of Auckland – Te Whare Wānanga o Tāmaki Makaurau  
[d.osullivan@auckland.ac.nz](mailto:d.osullivan@auckland.ac.nz)

# Outline

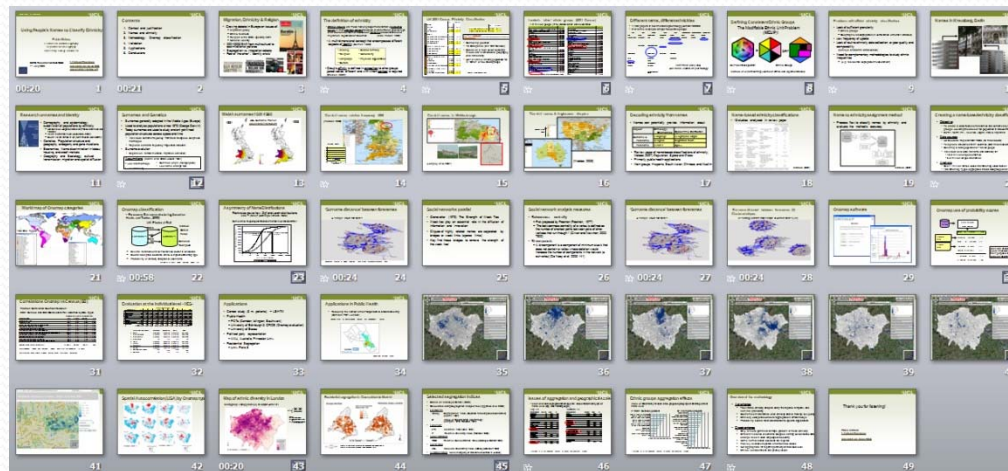
- Motivation for research
- Some background on the WorldNames and Onomap projects at University College London
- Basic characteristics of the New Zealand data
- Challenges of incorporating New Zealand data
- Some preliminary findings
- Future research directions

# Motivation and background: me

- I'm from Belfast...
  - where names (still) matter
- I emigrated to London
  - then to the US
  - then to Aotearoa-New Zealand
- I have children called
  - Fintan and Malachy, yet
  - my sister (in Belfast) has children called Matthew, Peter and Thomas,
  - while my brother (in England) has Max and Jamie

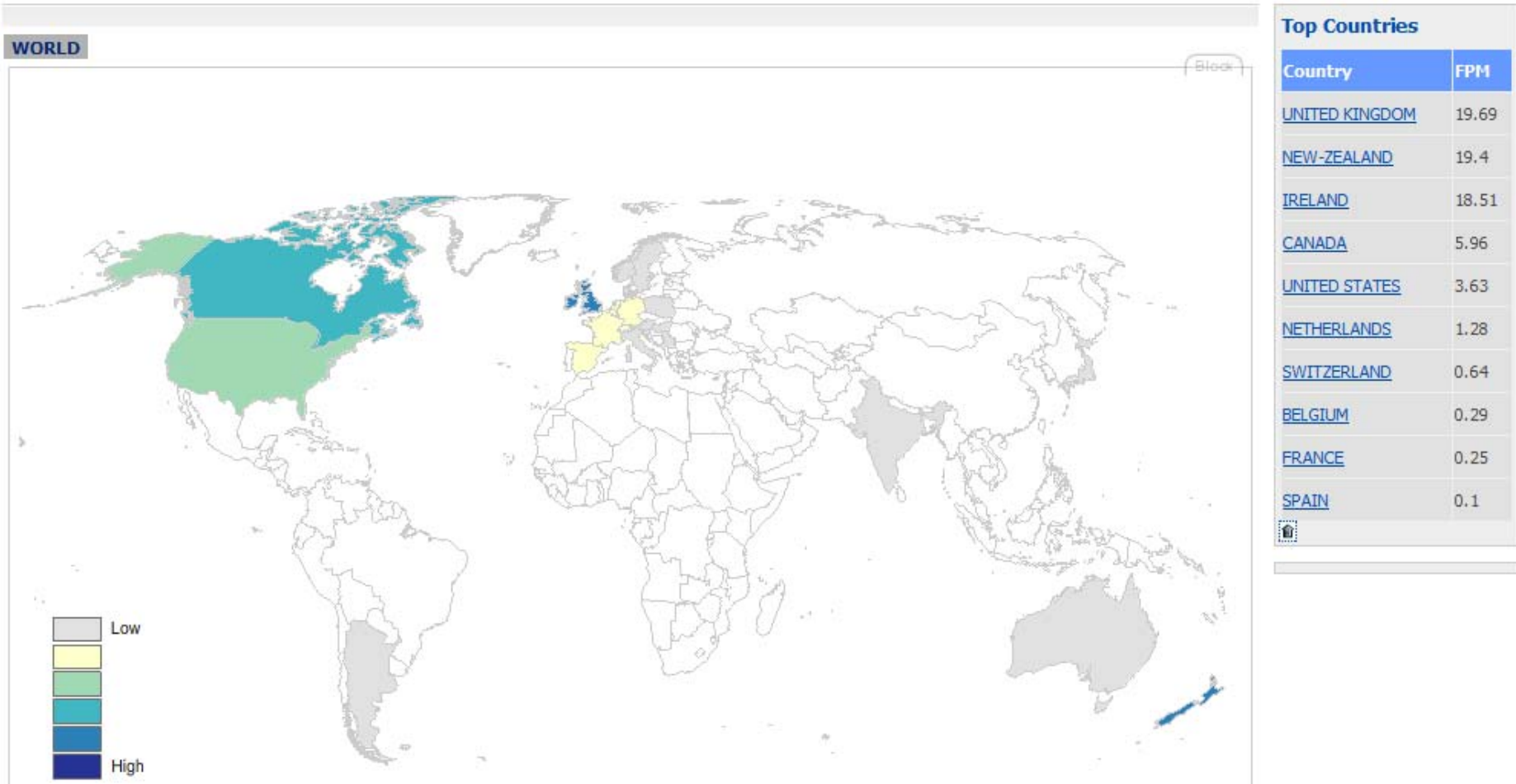
# Motivation and background: the 'real' reason

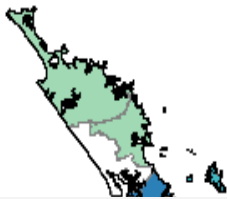
- Encountered Worldnames and Onomap work (Mateos and Longley) at a meeting in Boston, April 2008
- High relevance in New Zealand (esp. Auckland) setting



# The world names project

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





Low

### Roots of this name:

Surname	Group	Subgroup	Language
HILLARY	EUROPEAN_OTHER WESTERN	ENGLISH	

Low

High

### Top Regions

Area Name	FPM
<a href="#">QUEENSTOWN-LAKES DISTRICT, NEW-ZEALAND</a>	311.49
<a href="#">NORTH, UNITED KINGDOM</a>	111.56
<a href="#">WEST, IRELAND</a>	103.12

### Top Forenames

Forename
JOHN
JAMES
DAVID
ANNE
ANN
ROBERT
MARGARET
MICHAEL
MARY
WILLIAM

### Top Surnames

[WALKER](#)

Statistics for : **HENARE**

false true NEW-ZEALAND

### Roots of this name:

Surname	Group	Subgroup	Language
HENARE	UNCLASSIFIED	UNCLASSIFIED	MISC.

Do you agree? [Yes](#) [No](#)

Powered by [OnoMap](#)

Statistics for : **TAUFA**

false true NEW-ZEALAND

### Roots of this name:

Sorry we don't have any information on the roots of this name.

Powered by [OnoMap](#)

# Problems with ethnicity data

## Europe (Mateos, 2008)

- 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 granularity
- Low frequency of update

## New Zealand

- YES: but not between countries, *over time*
  - ‘New Zealander’ problem
- NO, but inconsistencies in data availability at different scales
- High granularity, but... see previous
- Not so bad (5 years)

# The 'New Zealander' problem

- In recent years there has been a campaign in the 5-yearly census to write in 'New Zealander' in response to the ethnicity question

11 Which ethnic group do you belong to?  
Mark the space or spaces which apply to you

The New Zealand Herald  
nzherald.co.nz WEDNESDAY OCTOBER 8 12:54AM NZT

News Business Election 08 Sport Technology Enter  
National World Weather Politics Crime Health Env

National ShareThis Print

## Email urges 'New Zealander' for

By Julie Middleton

A fast-spreading email appeal is urging people to state their ethnicity as "New Zealander" in next Tuesday's Census.

The email, which came to the Herald from several sources, reads: "Maybe we can get the powers-that-be to sit up and recognise that we are proud of who we are and that we are recognised as such, not divided into sub-categories and all treated as foreigners in our own country."

Census  
→ Make your vote count  
→ One in three working hours a week - Census

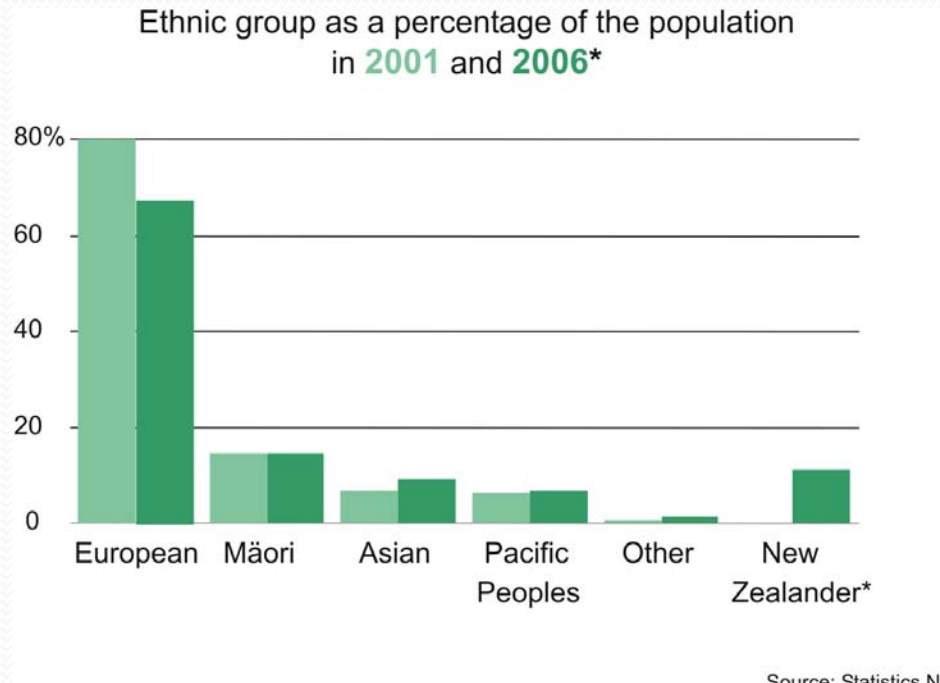
"The email appeal seems to be "a statement of independence from Europe or the mother country", says the president of the Population

Association. "Following an email campaign at the time of the last Census, 53,715 people nominated their religion as "Jedi", the fictional faith of *Star Wars* movie characters."

"As "New Zealander" is a nationality, not an ethnicity, says the Auckland University geographer, and the call is "a tragedy for the data" which will hinder social analysis."



# The extent of the problem?





# Other questions of interest

- Aotearoa/New Zealand is a good context for looking at migration:
  - Small, yet diverse population
- Historical sequencing of immigrations from British Isles
  - Scottish in the south
  - English elsewhere
  - Irish??
- Other questions:
  - Māori migration to the cities: are Iwi affiliations relevant to inner urban communities, and to what extent do names help us understand them?
  - Similar questions with respect to Pacific Island migrations
  - Persistence of historical patterns of early European settlement

# Previous work: Onomap

- The Onomap methodology
- This is work by Pablo Mateos, Paul Longley and others at University College London – Centre for Advanced Spatial Analysis

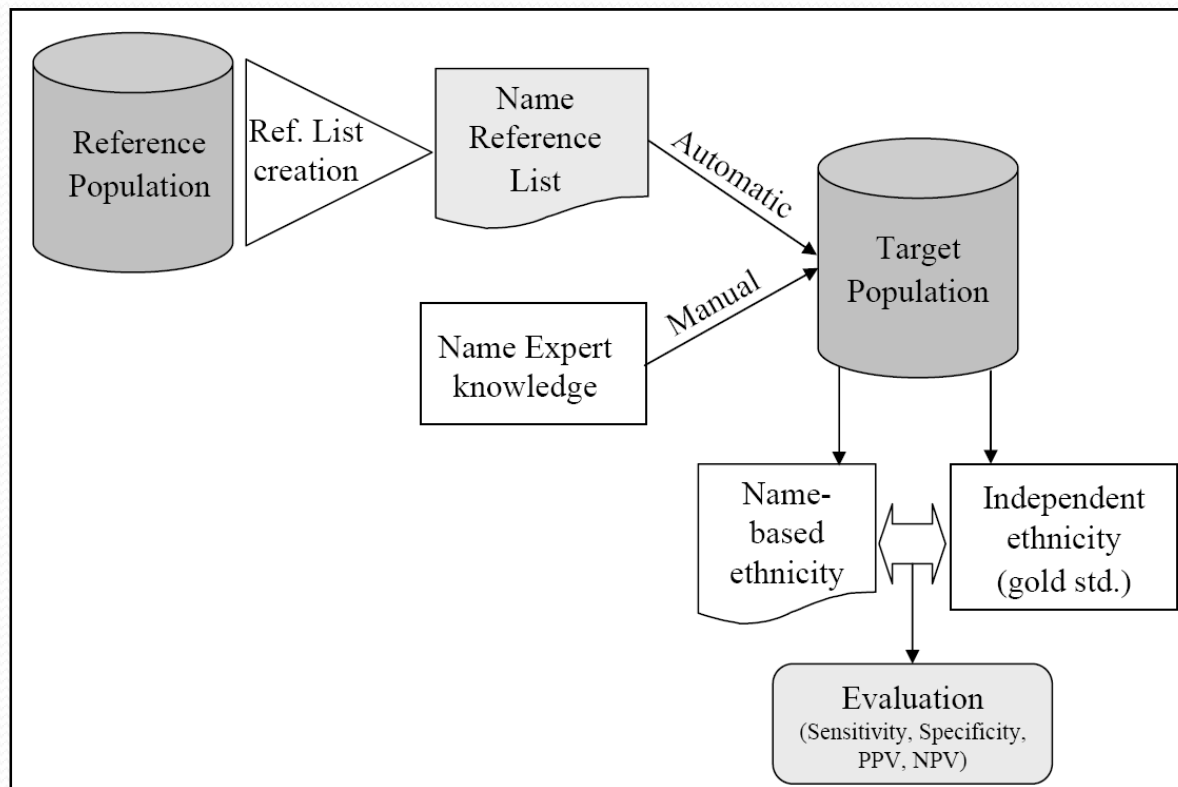
# Name-based ethnicity classifications

- 13 studies analysed in review paper: Mateos, P. 2007. *Population, Place and Space* 13(4), 243-63.

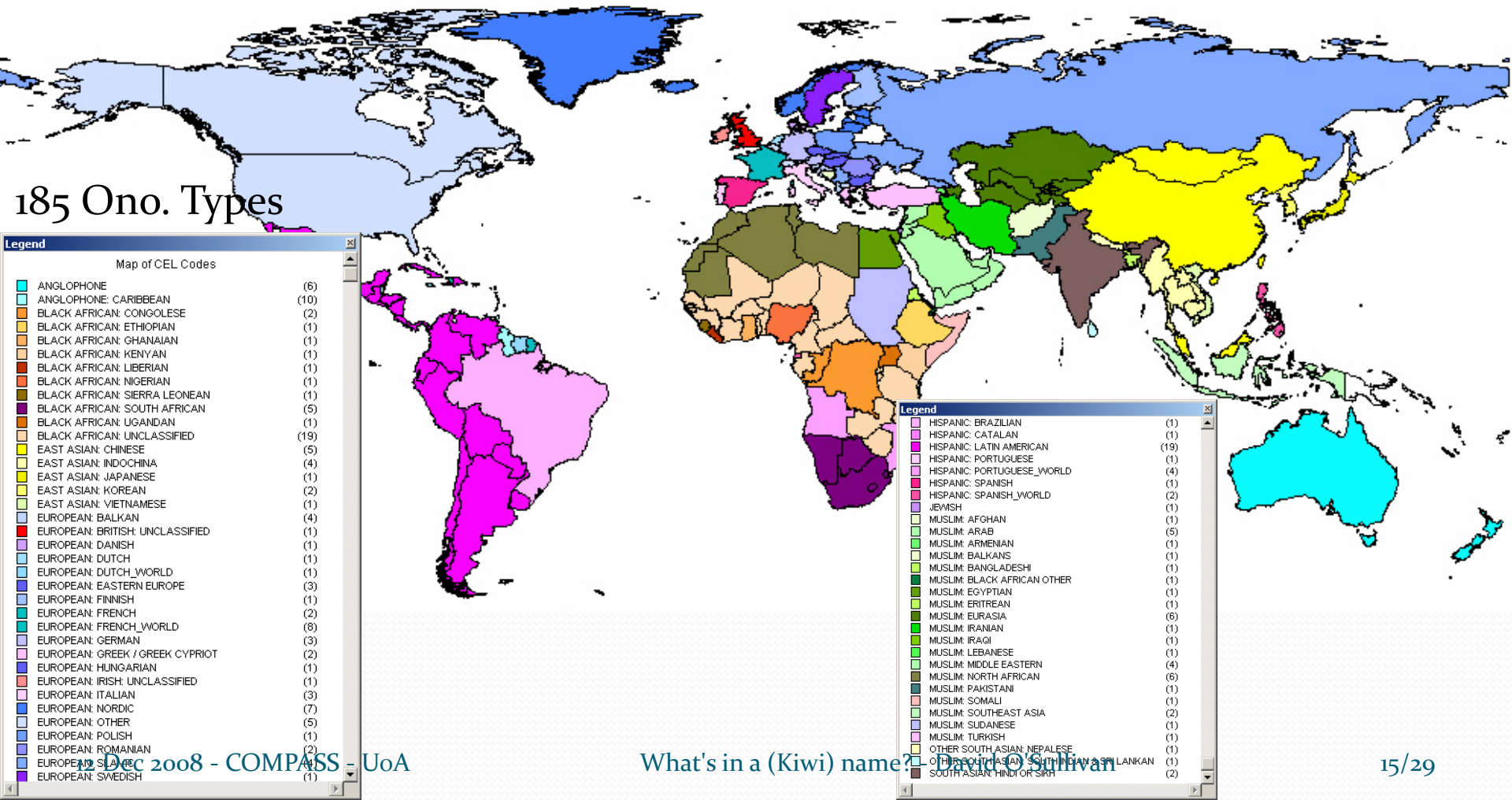
Paper reference	Geographical area of study: country and (region)	Ethnic minorities (EM) classified	Name to ethnicity assignment	
			Method: (Automatic Manual)	Name components: (Surname Forename Middle name)
Choi <i>et al.</i> (1993)	Canada (Ontario)	Chinese	A	S
Coldman <i>et al.</i> (1988)	Canada (British Columbia)	Chinese	A	F, S, M
Lauderdale and Kestenbaum (2000)	US (national)	Chinese, Japanese, Filipino, Korean, Indian and Vietnamese	A	S
Razum <i>et al.</i> (2001)	Germany (Rhineland Palatinate & Saarland)	Turkish	A	F, S
Word and Perkins (1996)	US (national)	Hispanic	A	S
Harding <i>et al.</i> (1999)	UK (Bradford & Coventry)	South Asian and Hindu, Muslim and Sikh	A	F, S
Cummins <i>et al.</i> (1999)	UK (Thames, Trent, W.Midlands & Yorkshire)	South Asian	A	F, S
Nanchahal <i>et al.</i> (2001)	UK (London, W. Midlands, Glasgow)	South Asian	A	F, S, M
Sheth <i>et al.</i> (1999)	Canada (national)	South Asian and Chinese	A/M	S
Martineau and White (1998)	UK (Newcastle; four general practices)	Bangladeshi, Pakistani, Indian Muslims, nonSouth Asian Muslims, Sikh, Hindu, White, Other	M	F, S and Gender
Bouwhuis and Moll (2003)	Netherlands (Rotterdam; one hospital)	Turkish, Moroccan, Surinamese	M	F, S
Nicoll <i>et al.</i> (1986)	UK (selected areas)	South Asian	M	F, S
Harland <i>et al.</i> (1997)	UK (Newcastle)	Chinese	M	F, S

Method of name to ethnicity assignment: A, Automatic; M, Manual. Name components used in the classification: S, Surname; F, Forename; M, Middle Name.

# Name to ethnicity assignment method



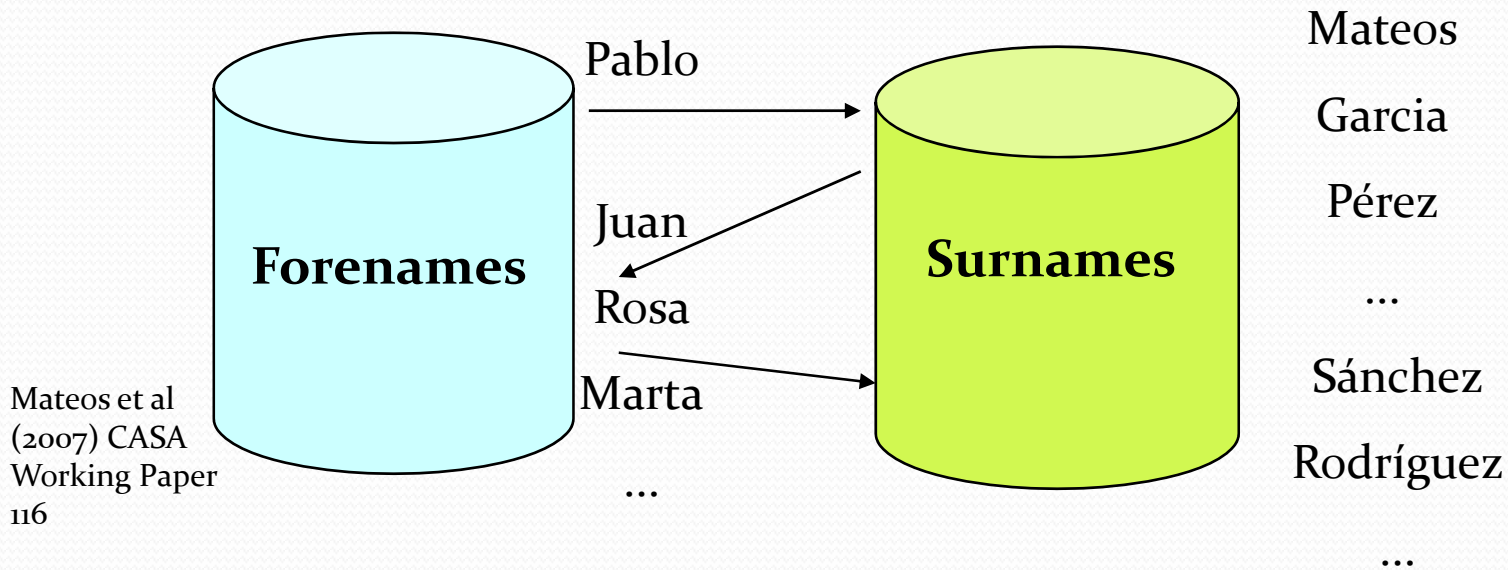
# World map of Onomap categories



# Onomap classification

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

## UK Electoral Roll



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



# Characteristics of NZ Data

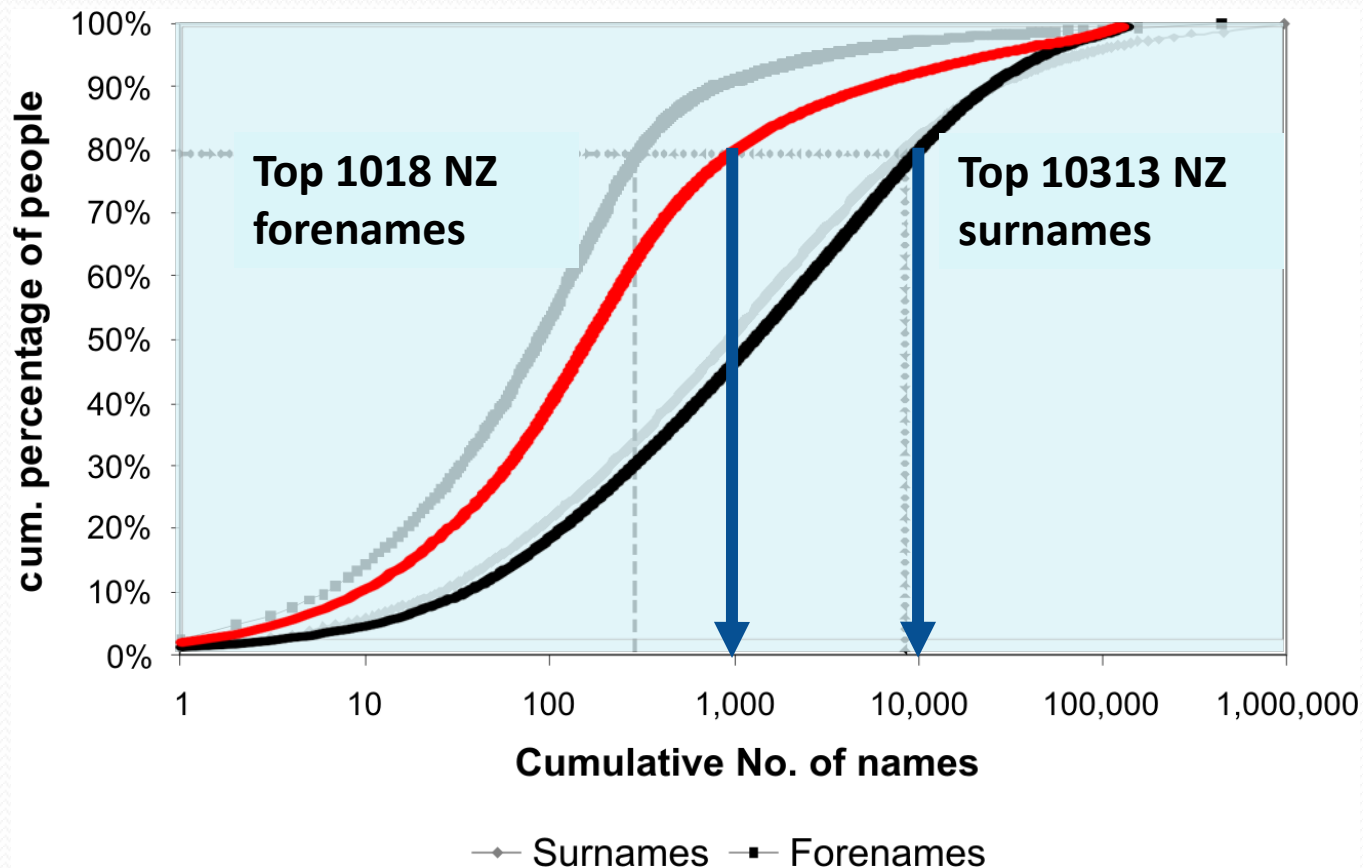
- Electoral roll
  - 2.85 million records including all forenames, surname
  - Ancillary data
    - Age indicator
    - Māori descent indicator
    - Full postal address
    - ID of smallest output unit from the census
    - Occupation

# The names themselves

- *Very high diversity in the names*
- 132,000 forenames and 142,000 surnames in the 2.8 million records

SURNAME	FREQ	FORENAME	FREQ
1. SMITH	25543	1. JOHN	46155
2. WILSON	14578	2. DAVID	41312
3. BROWN	13637	3. MICHAEL	34096
4. WILLIAMS	13517	4. PETER	31911
5. TAYLOR	12706	5. ROBERT	26617
6. JONES	11347	6. MARGARET	24144
7. ANDERSON	9287	7. PAUL	22278
8. THOMPSON	8810	8. JAMES	21860
9. WALKER	8126	9. ANDREW	20308
10. LEE	7494	10. WILLIAM	19522
...		...	
25. SINGH	5465	185. TE	3228
...		...	
85. LI	3131	217. NGAI RE	2795
...		...	
648. HENARE	587	485. SIONE	871

# Skewness of the distribution



# Challenges

- The manual process of building up a set of ‘seed’ names is difficult and error prone
  - For Māori names the ‘Māori descent’ flag in the data is a big help
  - For other ‘new’ groups it remains a significant problem – particularly Pacific Island names
  - Skew nature of the distribution helps *and* hinders

# Preliminary results

- Overall numbers of people by ethnicity

FINAL_ONO_GROUP	COUNT	PERCENT		CENSUS
ENGLISH	1460158	51.5%		
SCOTTISH	397913	14.0%		
EUROPEAN_OTHER WESTERN	221443	7.8%		
IRISH	182300	6.4%		
WELSH	105524	3.7%		
EUROPEAN_OTHER EASTERN	40049	1.4%		
CELTIC	14897	0.5%	85.4%	61.3%
MAORI	104132	3.7%	3.7%	12.2%
PACIFIC	3302	0.1%		
FIJI	2397	0.1%	0.2%	6.3%
CHINESE	100449	3.5%		
SOUTH ASIAN	78805	2.8%		
EAST ASIAN	18424	0.6%	7.0%	8.3%
MUSLIM	38642	1.4%		
AFRICAN	7308	0.3%	1.6%	0.8%
UNCLASSIFIED	53931	1.9%		
OTHER	5676	0.2%	2.1%	10.1%

Lack of data

'NZ-er's

# The problem

- Large numbers of 'unclassified' names in the forename and surname lists considered separately (both around 7.5%)
- Many of these get classified based on the other name
- Most common names are in the British/European groups, and the ones 'unfamiliar' to Onomap are in the Maori and Pacific groups

FORENAME_COMP	MAO_FREQ	TOTFREQ	PERC_MA
JOHN	6190.0	71181.3	8.7
JAMES	4923.1	43492.6	11.3
WILLIAM	3748.9	30227.0	12.4
TE	3159.9	3264.7	96.8
MARIE	3155.8	21733.5	14.5
DAVID	3015.7	36391.0	8.3
ANNE	2986.5	33738.9	8.9
MICHAEL	2819.2	26608.6	10.6
ROBERT	2650.7	26553.2	10.0
MARY	2475.8	32949.7	7.5
ANN	2429.2	28111.1	8.6
ELIZABETH	2397.3	29352.1	8.2
PETER	2148.5	24125.2	8.9
JOSEPH	2140.5	10127.0	21.1
PAUL	2120.9	19490.9	10.9
MARGARET	2088.8	33716.9	6.2
JANE	2064.2	23752.4	8.7
RICHARD	1934.7	16857.1	11.5
THOMAS	1870.7	12968.6	14.4
LEE	1739.2	8999.4	19.3
MARIA	1699.1	8886.7	19.1
GEORGE	1688.6	12904.7	13.1
EDWARD	1636.3	11311.4	14.5
ANTHONY	1611.1	12674.3	12.7
LOUISE	1590.1	15826.9	10.0
PATRICIA	1503.1	15740.9	9.5
CHARLES	1502.4	12287.3	12.2
MAREE	1495.0	11860.0	12.6
AROHA	1433.9	1545.1	92.8

# Some more encouraging results: Scottish names

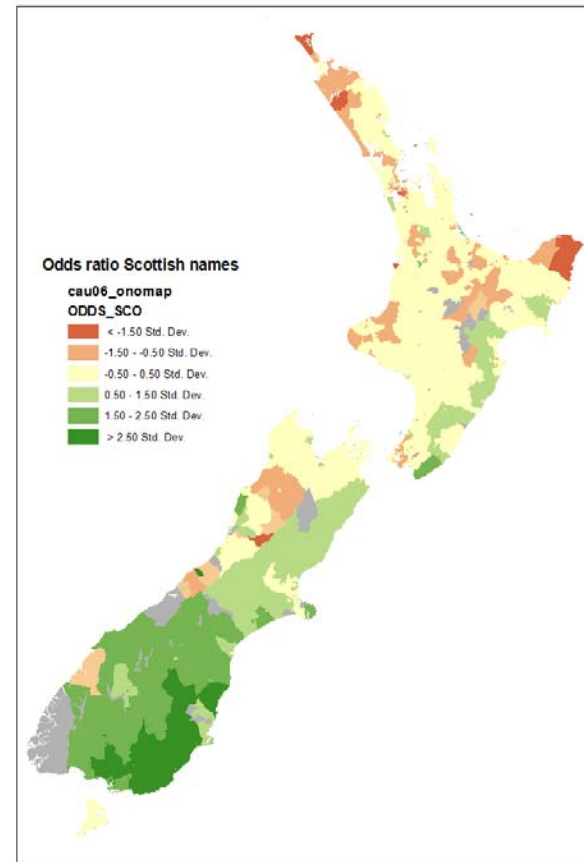
## Passenger arrivals at Port Chalme New Zealand, March 1848 - January

(Expanded and corrected from Dr Hocken's 1898 lists)

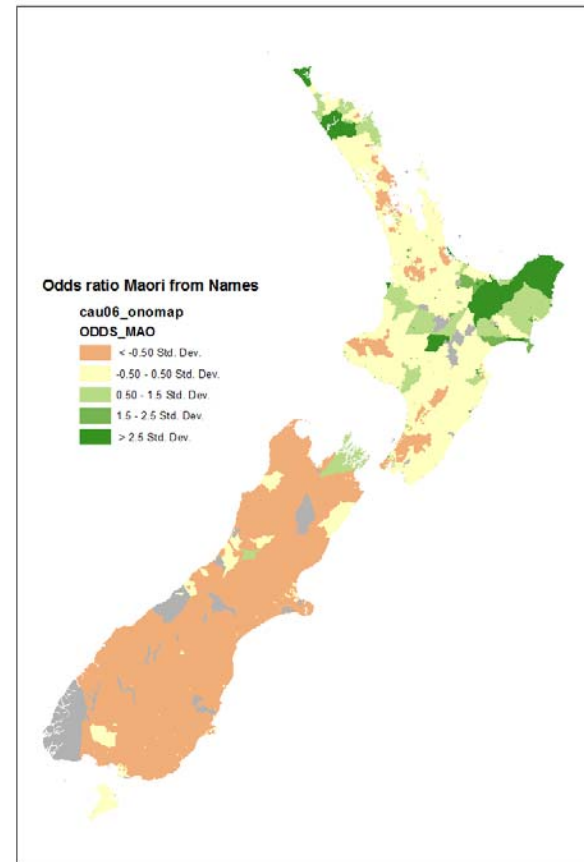
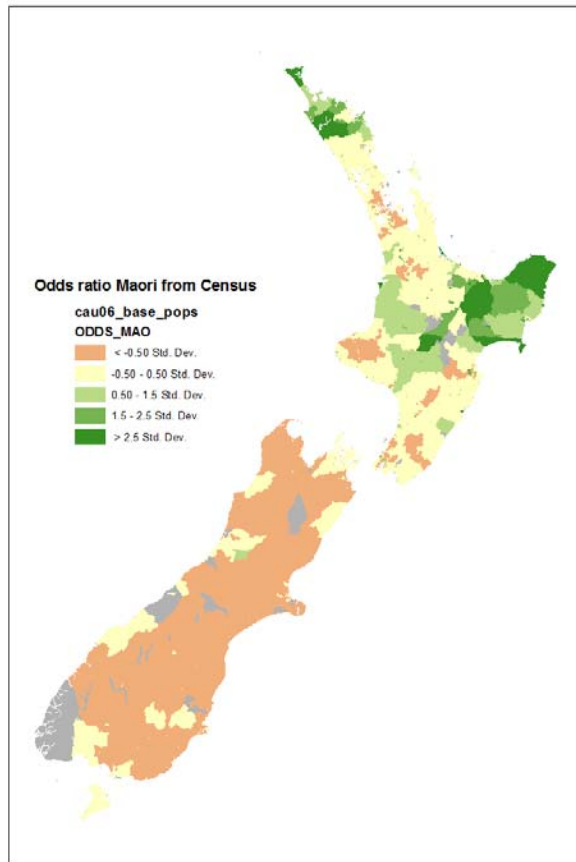
### PHILIP LAING

PHILIP LAING, 459 tons, sailed from Greenock on 23/11/1847, and Milford Haven on Robert Ramsay, surgeon.

**Adam, James** (25) [boatbuilder, precentor, emigration agent, farmer; Tokomairiro; d. 27/3/1908]; **Marc** Tokomairiro; d. 23/5/1917];  
**Aitken, George** (24) [shopkeeper; went to Nelson]; wife (21).  
**Barr, John**, (38) [weaver, 1st gaoler, poundkeeper; Little Paisley; d. 25/11/1873]; **Elizabeth** née Stewa (14) [butcher; d. Te Houka, 1876]; **Frances** (12) [m. Donald Reid (ex Mary) of West Taieri; d. 4/11/1863]; **Elizabeth** (1) [d. 2/7/1928].  
**Bell, John** (25) [ag., farmer; W. Taieri; d. 4/5/1887].  
**Blackie, James** (21) [schoolmaster; d. Sydney, 25/1/1851].  
**Bower, David** (28) [gardener; Balclutha; d. 29/5/1897]; **Jean** née Cuddie, wife (22) [d. 16/5/1894].  
**Breakie, Hugh** (22) [unm.]  
**Brown, James** (23) [calico printer, ag.; Andersons Bay; d. 27/11/1891]; **Hannah** née Renfrew, wife (2) **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]; **Margaret** née McCulloch, wife (4) Stevenson, d. 16/2/1926].  
**Buchanan, Thomas** (42) [weaver, farmer; West Taieri; d. 16/7/1876]; **Margaret** née Graham, wife (40) West Taieri; d. 24/2/1913]; **James** (15) [ag.; never married; West Taieri; d. 28/10/1913]; **Ellen** (9) [d. 1 27/3/1876].  
**Burns, Thomas** Rev. (52) [first minister (Free Church of Scotland), d. 23/1/1871]; **Clementina** née Gra Dickson; Mosgiel; d. 15/9/1901]; **Clementina** (15) [m. Capt. Andrew Jamieson Elles; d. 18/9/1862]; **Ja** Robert Livingston; d. 21/11/1918]; **Frances** (6) [m. Henry Livingston; d. 28/7/1922]; **Agnes** (1) [unm.; c  
**Callander, James** (41) [farmer; East Taieri]; **Jean** née Cuthbertson, wife (37) [d. 8/1900]; **James** (16) née Law (ex Blundell) on 26/12/1862; Scrogg's Hill; d. 25/12/1909]; **Rebecca** (3) [died young].  
**Compton, James** (22); **Margaret** (15) [m. 18/1862]; **Henry** (20); **John** (14) [d. 6/1846].



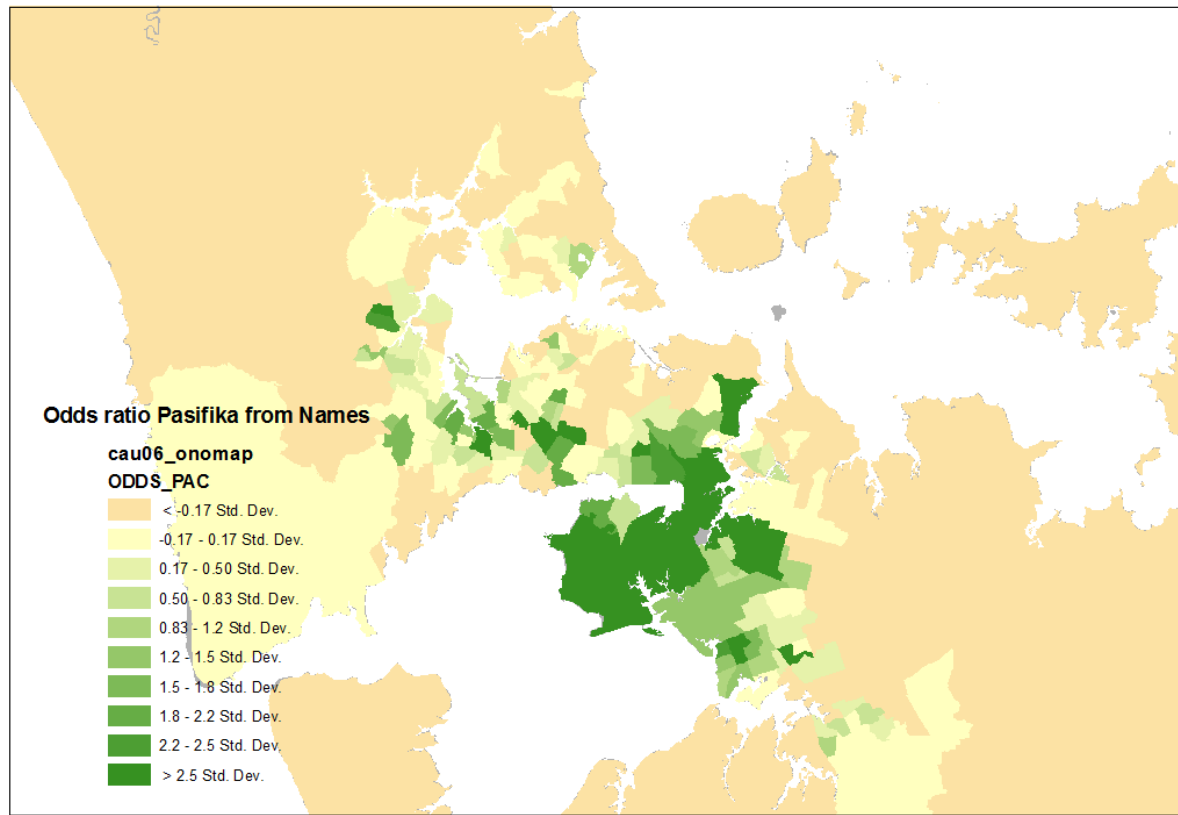
# Some more encouraging results: Māori names



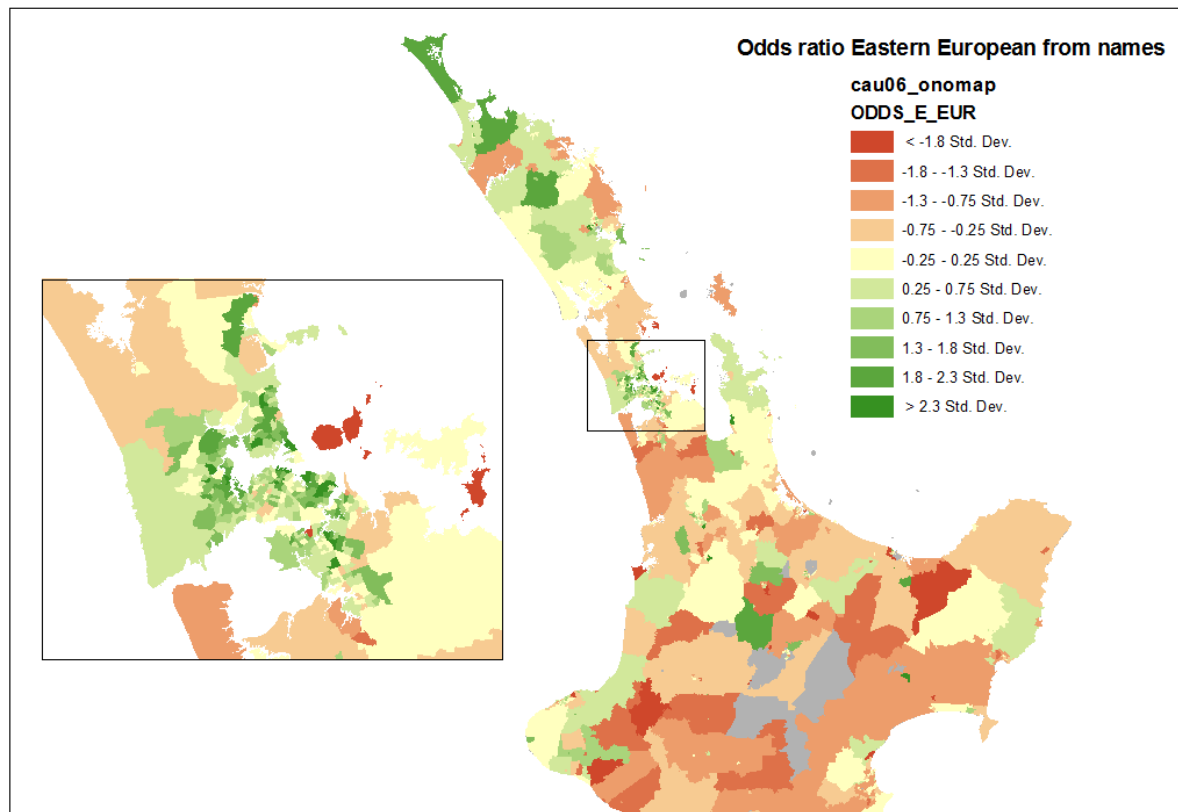


# And focusing on Auckland...

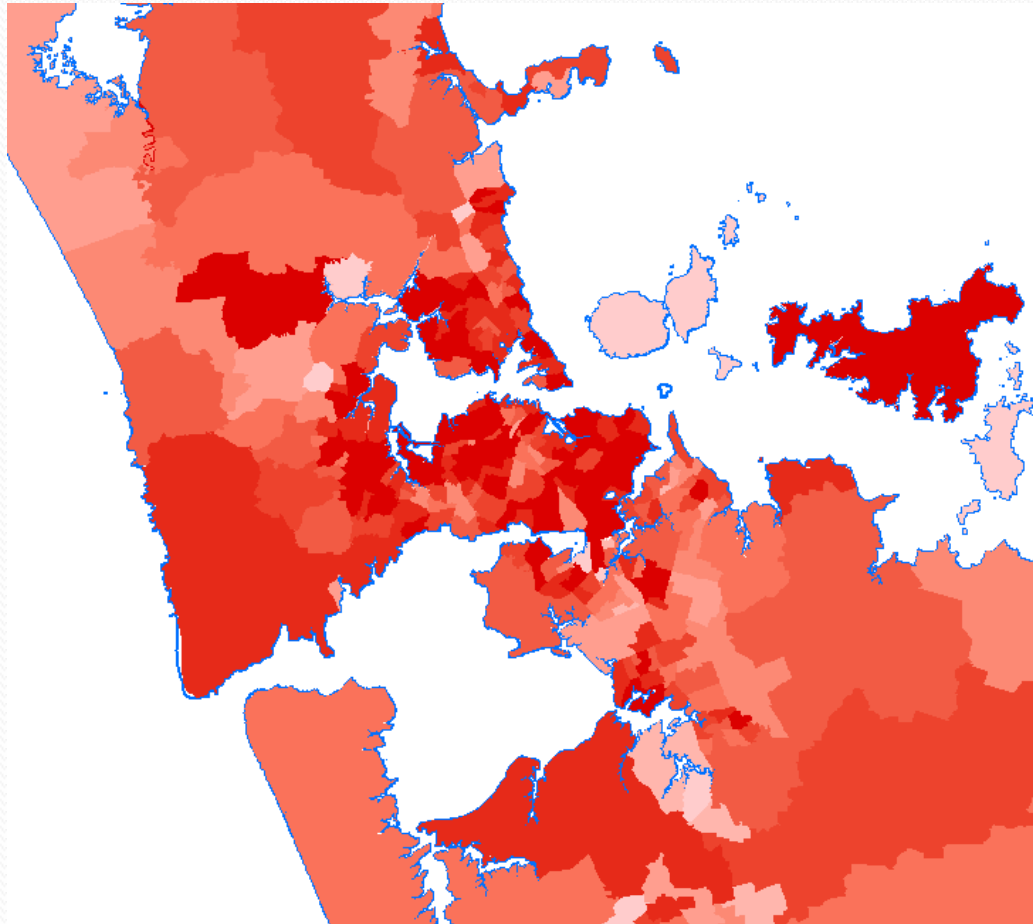
## Pasifika names



# 'Eastern Europeans' in the north



# Measuring diversity using names



# Future directions

- Priority is to improve the Onomap classification to match NZ census data better – at aggregate level to begin with
  - Change how ethnicity is assigned in ambiguous cases
  - Automate the identification of names from groups new to the names-based classifier
  - Automate the process of growing clusters of ‘new’ names
- Simulation modelling of how names distributions end up the way they do

# Acknowledgments

- Pablo Mateos
- Paul Longley
- Alex Singleton
- Muhammad Adnan
- All at CASA / UCL Geography / SPLINT
  
- Funding from University of Auckland, Study Leave Grant in Aid, and a SPLINT Fellowship
  
- QUESTIONS?