

# Preface

The research which conforms this book started in autumn 2004 in London, UK and its writing ended in early 2013 in Guadalajara, Mexico. In 2004 I used to live and work in the London Borough of Camden where this book's gestation took place. Walking a few hundred meters around Camden one seemed to be travelling around the world, given its diversity of people, smells, foods, clothes, or music, switching continents as you turn a street corner and listen to tens of languages in every bus ride. However, ethnic diversity in Camden also reflects stark wealth inequalities, home to some of the richest postcodes in the country lying next door to the poorest neighbourhoods in national rankings.

At the time I was analysing health inequalities by ethnic group in Camden, as part of a research project between geographers at University College London and epidemiologists at Camden National Health Service (NHS). It was then that I quickly became dissatisfied with the UK Census ethnic group classification, commonly used to produce all sorts of official statistics by population researchers in the UK. It was, and still is, a broad-brush classification of humanity into eight major groups of "ethnic origin". Its simplicity clearly fails to represent the wide range of ethnic groups present in Inner London. These are by all means no small population groups. For example, 40 % of pupils in London schools speak a language at home other than English, covering a total of 322 different languages (Von Ahn et al. 2010). The wide-spread use of the census ethnic group classification reified deeply-rooted stereotypes and expectations of ethnic disadvantage in British society. For example, a wealth of evidence in population studies points at Bangladeshis as the poorest, most segregated group, presenting the worst health outcomes in London. Because Bangladeshis comprises an ethnic group on their own in the census form, they get all the good and bad attention in academia and public policy. Meanwhile in Camden an equally sized group, the Somalis, complained of not getting the same level of resources because they were 'statistically hidden' under the all-encompassing 'Black African' group. As in-depth analyses of the 2001 Census unfolded at the time, the intrinsic characteristics and needs of tens if not hundreds of ethnic, national, linguistic, religious, and geographic groups were being ignored in London.

It was in this context that I turned to name-origin analysis, as I searched for unconventional ways to group finely grained sub-population ‘labels’ into alternative configurations of ethnicity that better reflected the diversity of London’s population. This was not a gratuitous pastime, stark social inequalities by ethnic group that were clearly observable at the general practices and hospitals of the National Health Service, could not be addressed using population research methods because of a lack of statistics on those very same ethnic groups. However, I noticed that administrative datasets such as population and health registers contained millions of records of people’s names and household addresses, partly reflecting that very rich cultural diversity that I was trying to grasp, and to map. “If only I could find a bunch of linguistic experts that could read all those hundreds of thousands of names and code them by ethnic origin. . .”, I kept thinking at the time. I was surprised to find out that others before me have had that very same idea in various countries since at least the early twentieth century. I soon began conducting a literature review that formed part of what was then to become my PhD thesis. “There were ‘tried and tested’ ways to do this automatically and reliably”, I found to my relief! Little did I know at the time that 9 years later I was going to be wrapping up a whole book on the subject.

Various projects and applications spun from that initial interest in mapping ethnic inequalities in health in Inner London. Within a year I managed to enthuse several other researchers at University College London’s (UCL) Department of Geography in unveiling this intriguing and fascinating world of name analysis. The core team was comprised by Paul Longley, James Cheshire, Alex Singleton, Muhammad Adnan, Maurizio Gibin and myself. I must thank them all for their support, advice, guidance and co-authorship throughout these years, and more concretely for the materials they kindly gave me permission to publish in this book. Especial thanks go to Paul Longley, whose sense of humour and leadership was central to create the team spirit that made this research possible as well as the most enjoyable experience at UCL during almost a decade. Other researchers joined this team for specific projects and publications, weaving a world-wide network of researchers that over these 9 years has managed to compile and analyse name frequency statistics for the whole population of over 30 countries in four continents, representing at least a third of the World’s population. I must first thank Richard Webber, at Experian, whose mentoring over my PhD research years was invaluable and to whom I owe a great deal of inspiration to bring name analysis to the next level. Special thanks must go to Michael Batty and colleagues at the Centre for Advanced Spatial Analysis (CASA), my intellectual home in London, at whose Wednesday seminars (which I never missed for at least 6 years) many of the ideas and collaborations in this book came to fruition. Thanks to Andrew Crooks, Junior Sinesio Alves (†), Oliver O’Brien, Jakob Petersen, Alan Wilson, Yi Gong, Richard Milton, Peter Wood, Adam Dennett, Juliana Cipa, Paloma Rojas, and many other researchers at CASA for their support and encouragement. All of these geographers, computer scientists, architects, planners, epidemiologists, and urban sociologists provided hints and knowledge that are somewhere present in this book.

Outside my *alma mater* at UCL, I was very fortunate to come across a number of researchers from various disciplines around the world that helped me mature the central idea of this book; linking forenames and surnames frequencies into groups of common origin using a network approach. I must specially thank Ken Tucker (computer scientist), at Carleton University, Ottawa Canada, a pioneer in the forename-surname linking approach, and David O’Sullivan (geographer), at the University of Auckland, New Zealand, for his mathematical ability to crack the clustering of huge networks and his key co-authorship that forms part of the materials in Chap. 7 of the book. Thanks must also go to Mario Cortina-Borja (statistician), at the Institute of Child Health, for very valuable insights into name frequency distributions, Franz Manni (geneticist), at the Musée de l’Homme in Paris, for his unorthodox thinking in applying names to population genetics, and to Ludi Simpson (demographer) at the University of Manchester, who read and commented many of my papers on this topic. I am indebted to Andres Moreno and Karla Sandoval (geneticists), at Stanford University, for furnishing my scarce knowledge on human population genetics, and opening up a whole new world that largely facilitated the transdisciplinary research accomplished in the book. My work in this book certainly stands upon these giants’ shoulders who preceded me or walked alongside me in moving the research frontier in ethnicity classifications a little further.

Furthermore, my two PhD examiners, Paul Boyle, University of St. Andrews/ESRC Chief executive, and John Stillwell, University of Leeds, provided very useful feedback that allowed me to transform a dull PhD thesis into a much more amenable and coherent book monograph. Many others helped me to source valuable names data across the world or hints on how to get it in different languages, too many to mention here. I am indebted to every one of them.

This book also benefited from extended research visits to a number of institutions. I am indebted to Douglas Massey at Princeton University for hosting me over several months at the Office of Population Research, whose immense library allowed me to locate part of the evidence collected in this book. I must also thank the University of Auckland and the Royal Society for funding an academic stay at Auckland, New Zealand, forging the aforementioned crucial co-authorship work with David O’Sullivan. Repeated academic visits to the Geography Department at the Autonomous University of Madrid, Spain, facilitated by Antonio Moreno, as well as to the Institute of Geography at the National Autonomous University of Mexico, organised by Adrian Aguilar, provided another avenue of interesting research and teaching interactions with urban and population geographers, as well as the time and space to complete parts of this book.

A number of research grants funded the work that underlies the book. Various grants from the UK Economic and Social Research Council (ESRC) (Knowledge Transfer Partnership KTP-037) (RES-348-25-0015) (PTA-026-27-1521) (RES-172-25-0019), the Engineering and Physical Sciences Research Council (EPSRC), the Spanish Ministry of Science and Innovation Base Research Projects (CSO2011-26177), the Royal Society, and the UCL Graduate School, provided the funds for academic mobility and to conduct primary research that partially led to this book.

Finally, but most importantly, I would like to thank my family. My siblings and specially my parents, Manoli and Jose Maria, who laid the foundations that made me an inquisitive person, encouraging me to question all the time the world that surrounds me. Only my wife Brenda, has borne the burden side of writing this book. Without her continuous support, patience, sacrifices and enthusiasm to accomplish and finish this book, not a single chapter would have come to fruition. She also knows that our small children, Blas and Julian, both with multiple passports from birth, will probably laugh at their father's rather simplistic views on ethnicity in this book if they ever get to read it over the next decades. I only hope that in their adulthood, the importance currently placed on ethnicity as an essentialist dimension of a person's identity fades away, and people stop being judged by their physical appearance, accent, religion, kinship, place of birth or colour of their passport. May the reader take this book as a contribution to improving how we monitor progress towards this end.

Guadalajara, Mexico  
30 June 2013

Pablo Mateos

## Reference

Von Ahn M, Lupton R, Greenwood C, Wiggins D (2010) Languages, ethnicity, and education in London. UPTAP. Working Papers 10–12. Institute of Education, University of London



<http://www.springer.com/978-3-642-45412-7>

Names, Ethnicity and Populations

Tracing Identity in Space

Mateos, P.

2014, XX, 269 p. 55 illus., 27 illus. in color., Hardcover

ISBN: 978-3-642-45412-7