Forenames and Surnames in Spain in 2004

Pablo Mateos [1] and Ken Tucker [2]

[1] Department of Geography, University College London, UK - p.mateos@ucl.ac.uk
[2] Carleton University, Ottawa, Canada - posthaus@igs.net

Abstract

This paper quantifies the corpus of forenames and surnames in Spain in 2004 using the telephone directory. It describes their frequency patterns, major measurable characteristics, and gives some geographical distributions, international comparisons and historical explanations. The research presented here is set in a context of a broader study of the quantitative properties of the corpus of personal names in several countries undertaken by Tucker. Amongst the most significant findings are; a much more highly skewed distribution towards the most popular surnames than in other countries, the permanence of language regions since the Middle Ages, and important differences in top Hispanic names frequencies between five countries across the Atlantic. It is also suggested that the innovative techniques presented here, combining geographical and statistical analysis of names and their language of origin, opens up enormous possibilities for multidisciplinary work on onomastics.

Resumen extendido (extended abstract)

El presente artículo cuantifica el corpus de nombres y apellidos en España en 2004 utilizando el directorio telefónico. Se describen los patrones de sus frecuencias, sus principales características mensurables, y se ofrecen algunas distribuciones geográficas, comparaciones internacionales y explicaciones históricas para los patrones encontrados. Esta investigación se enmarca dentro de un proyecto más amplio que estudia las propiedades cuantitativas de los nombres personales en varios países dirigido por Tucker.

El ámbito principal de este trabajo son los nombres de España, si bien se hacen referencias a los nombres españoles en otros países de habla hispana. Se han encontrado una serie de aspectos comunes de los nombres españoles que pueden ser resumidos en los siguientes puntos. En primer lugar la actual distribución de frecuencias de los nombres de pila sigue un patrón similar a la de otros países, pese a que se sabe que ésta estuvo muy concentrada en pocos nombres de pila en la Edad Media, reducidos a un limitado santoral. En segundo lugar, la distribución de frecuencia de los apellidos presenta un mucho mayor sesgo positivo hacia los apellidos más comunes que el encontrado en otros países. Esta

peculiar distribución puede ser explicada por tres factores relacionados entre sí. En primer lugar, debido al patrón de repoblación cristiana de la Península Ibérica, desde un grupo inicialmente reducido de apellidos patronímicos en el norte hasta su expansión territorial y poblacional por toda la península. En segundo lugar, debido a un proceso de cambio forzado de apellido, por la presión de la Inquisición sobre los no cristianos y por la "castellanización" de apellidos a lo largo de los siglos. En tercer lugar, debido a un fenómeno de "endogamia de apellidos" (también conocido como "isonimia"), es decir, una alta frecuencia de matrimonios entre personas con el mismo apellido, lo cual en muchas zonas no ha sido contrarrestado por la "inmigración" de otros apellidos.

Uno de los rasgos característicos del sistema de apellidos españoles es la utilización de dos apellidos (paterno y materno). Sin embargo, la distribución de apellidos maternos y paternos es exactamente la misma, ya que al fin y al cabo todos son heredados vía paterna, aunque se tarda dos generaciones en perder el apellido materno mientras que en el sistema anglosajón éste desaparece en una (si bien desde hace unos años el orden de apellidos puede alterarse en España).

El artículo presenta algunos ejemplos de análisis geográfico a modo exploratorio, comparando la distribución geográfica de los apellidos agrupados según su origen en uno de los cuatro idiomas actuales presentes en España (Catalán-Valenciano, Castellano o Español, Euskera y Gallego). Los mapas resultantes revelan el proceso de poblamiento medieval de la península, aún fosilizado en la distribución actual de apellidos en el directorio telefónico, demostrando la interrelación permanente entre la estructura demográfica de la población y las zonas lingüístico-culturales de origen. Las técnicas presentadas demuestran el claro potencial del análisis cuantitativo y geográfico de los nombres para revelar procesos históricos demográficos y de asentamiento. Finalmente, mediante la comparación de las distribuciones de los apellidos más comunes en cinco países (España, Argentina, México, Venezuela, y Estados Unidos) se revela que los apellidos son un buen indicador de cómo distintas poblaciones se han asentado y mezclado entre países, así como la imposición de las prácticas de asignar nombres personales a las antiguas colonias ha empobrecido la riqueza onomástica de sus poblaciones.

Mediante este artículo esperamos haber desentrañado algunos de los "secretos cuantitativos" de los apellidos en España. A través de las técnicas presentadas, también intentamos introducir el análisis geográfico-temporal en el campo de la onomástica, para comprender mejor la estructura histórica y contemporánea de la población a través de su "geomorfología onomástica". En un esfuerzo común, lingüistas, historiadores, geógrafos, investigadores de genética, estadísticos y demógrafos entre otros deberían colaborar más cercanamente para revelar un poco más acerca del cómo vinimos a ser lo que somos hoy.

Introduction

The aim of this paper is to quantify the corpus of forenames and surnames in Spain in 2004 using a broad population register such as the telephone directory. It describes their frequency patterns, major measurable characteristics, and gives some geographical distributions and historical explanations. The paper does not seek to provide a history of forename and surname development in Spain and readers may like to read Kremer's brief review (Kremer, 2003). However, in order to understand some of the processes presented here, we need to describe a few major distinctive features of the historic linguistic context and naming conventions in the Iberian Peninsula.

Surnames started to be used in Spain in the 10th century and were well established by the 12th century (Kremer 2003), mainly as patronyms that changed with each generation, although they gradually became inherited surnames between the 13th and the 15th century (Faure et al, 2001). Surnames were introduced as a consequence of the reduction in the number of forenames in the Middle Ages, and the need to identify people in legal documents. For example, in the 10th century a study found 1.3 people per forename while a century later the same authors found 3 people per forename (Moll, 1982). This seemed to be due to the influence of religion, since most forenames were reduced to the most popular saints.

During these five centuries the Iberian Peninsula was settled by several population groups and languages, grouped into five more or less permanent kingdoms; four of them Christian and south-expanding: Galicia-Portugal, Castile-Leon, Basque Country-Navarre, and Aragon-Catalonia, and one Muslim and south-retreating: Al-Andalus. These groups spoke at least nine different languages (eight romance languages plus Arabic), and although only Castilian (Spanish), Catalan, Galician, Portuguese and Basque survive today; all have left traces in the surnames found in the Iberian Peninsula. To these we can add surnames from non-romance languages, Iberic & Germanic, previously spoken in the peninsula, Jewish surnames, and surnames brought from the native languages of former colonies in Latin America and the Philippines, for example: *Moztezuma* (Tibón, 2001).

With the growing expansion of Castille since the 14th and 15th century, and the political unification from the beginning of the 16th century of what now territorially constitutes Spain, the Castillian language, currently also known as Spanish, was imposed to all other kingdoms. Therefore, Castilianization of personal names over several centuries makes it difficult to ascribe all surnames to their original language and form (Kremer, 2003). Castilinization was combined with Christianization, enforced by the Spanish Inquisition since the 16th century, that forced many people change their Arab, Jewish, or 'foreign sounding' surnames to a mainstream Castilian one to avoid persecution. Some of the new Castilian surnames were adopted so frequently by religious converts that they have been identified as typical 'convert surnames'; these are amongst the most frequent surnames found today, as it will be explained later.

History has left a rich and diverse cultural sediment present in today's place names and personal names in all of the 21 Spanish speaking countries. However this paper is concerned only with the personal names of contemporary Spain. For simplicity we will use the overarching term of 'Spanish names' for these forenames and surnames.

The Spanish custom to use two surnames (father's and mother's surnames, see next section) seems to have started around the 16th century. According to Faure (2001) this was because the use of two surnames was associated with aristocratic families, and hence it became very fashionable amongst the popular classes or the upcoming bourgeois. In the 16th century those with a nobiliary title or their descendants did not have to pay taxes, and many people tried to claim they had aristocratic ancestors (the group of dispossessed "nobles", known as "hidalgos", was huge in the 17th century, and they were all registered due to tax exception reasons; "Padrón de Hidalguía"). Therefore, as Spain went into economic decline in the 17th century, it seems to have been important to keep both paternal and maternal surnames to identify individuals who might have some sort of "nobiliary rights" or just to distinguish themselves from the most common surnames. This explanation is similar to that of using double barrel names in the Anglo-Saxon naming system, of which there are also many examples with combination of Spanish most popular surnames.

Finally, in the 19th century this custom was institutionalized through a Civil Registration Act (1870) which made it mandatory to register births and always use two hereditary surnames, both father's and mother's surname. The act also forbid any change in the spelling of one's family name (Kremer, 2003). This made sure that paternity and maternity of a child was always clear, as well as to identify brothers and sisters of the same marriage. This had important implications in legal issues, for example in hereditary disputes. This was also the time when Spanish surnames were given to all the population of Philippines, then a Spanish colony, together with Cuba and Puerto Rico, before the 1898 war with the US. Today most people in Philippines carry Spanish names although only 3,000 people speaks Spanish in a country of 89 million people (CIA World Fact book).

The 1870 Act stopped the process of Castilianization of surnames, but that of forenames has continued, and specially reinforced during the Franco dictatorship, when Castilian was the only official language. The restoration of democracy in 1975 has brought back into official records the Galician, Catalan and Basque given name that many citizens were given at birth. The 40 year dictatorship had also a high impact on the surnames that migrated to Latin America from Spain as people emigrated to escape persecution. A high proportion of Catalan and Basque distinct surnames occur amongst these emigrants.

Finally, the return to democracy and the economic expansion in the last 20 years has seen Spain to shift from a net emigrant country in the 1960s and 70s to being now the country with the highest rate of

immigration in Europe (In 2005 the population grew by 2.1%, due to immigration, and in the period 2001-2006 by 9%) (Instituto Nacional de Estadistica, 2006). This process has brought surnames from all over the world to the Iberian Peninsula, especially to the major cities and the Mediterranean coast and the islands.

Most striking is the surge of rare Spanish surnames noticed in many population registers of Spanish cities (Instituto de Estadística de la Comunidad de Madrid, 2006), some of which were previously extinct in the peninsula (e.g. Simbaña, Armijos). These are surnames brought back from Latin American countries in a return journey after 500 years, having been preserved and disseminated across the Atlantic, closing a cycle of world-wide population migration and mixture. This is a fascinating journey that we are only starting to discover today by analyzing surname frequencies.

Main Features of the Spanish Naming System

Spanish Surname Structure

To an Anglophone, Spanish Surnames look quite complicated as most Spaniards, and people from Spanish-speaking countries, have two surnames. The nearest equivalent in the Anglophone world would be the hyphenated name such as Smith-Jones.

When a Spanish child is born it usually inherits as its first surname its father's first surname, and as its second surname its mother's first surname. For example say the father's names, in the *Forename- 1st Surname-2nd Surname* pattern is:

Esteban Martínez Muñoz

And his wife's name using the same pattern is

Pilar Ortiz Molina

Say they have a child and give her the forename *Ana*. Then her name using the same pattern would be

Ana Martínez Ortiz

If they had another child, and called him José, then his name would be:

José Martínez Ortiz

In this example all the children have the same surnames, but the mother and father have surnames that differ from each other and those of their children. For married couples, the tradition is not to change any surname, so groom and bride keep their three component name birth name. Every child will inherit the father's first surname and the mother's first surname, usually in that order, although the order can be now changed.

Spanish surnames are thus patrilinearly inherited, as are the Anglo-Saxon ones, but it takes two generations to loose surnames in the matrilineal linage, rather than just one in the Anglo-Saxon system. The advantage of the Spanish system is that one can trace a person to both of his or her parents, which helps researchers in different applications, such us in historic record linkage as well as pedigree reconstruction in genetic research (Rodriguez-Larralde et al, 2003).

For practical purposes, most people in Spanish-speaking countries just use their forename and first surname: e.g. *Pablo Mateos*, and they only use the full name in official documents or formal situations. Therefore, the second surname is used to avoid potential confusions when the purpose of uniquely identifying a person is important (e.g. Pablo Mateos Rodriguez), just as the middle name is used in the Anglo-Saxon system: e.g. George William Bush, vs. George Bush)

Other features of Spanish names

The most prominent feature of Spanish surnames is the presence of the ending *'-ez'*, which dominates most surnames, with 14 out of the top 20 most frequent surnames end in *'-ez'* or its derivatives. This ending is the patronymic form in old Castillian, and it was attached to the forename of the father (e.g. *Fernandez* was the son of *Fernando*). A variation of this ending in Galician and Portuguese is *'-es'*, what commonly serves to distinguish the origin of names between languages. However, name corruptions between '-ez' and '-es' endings, in both directions, are frequent in Spanish-speaking Latin America and even more in the U.S. (e.g. *Hernandes* instead of *Hernandez*, or *Valdez* or *Cortez* instead of *Valdes* or *Cortes*) since the letters 's' and 'z' are pronounced exactly the same in Latin America and the south of Spain. Hispanic surnames of this form in the U.S., other than a minority of Portuguese origin, are Anglicizations of the ending '-ez' into an '-es'.

Another important feature of Spanish names is the high frequent of toponyms, which Faure et al (2001) quantify as 58% of the surnames in their dictionary. Of these, a high proportion of toponyms come from Basque and Catalan place names (28% and 17% of the total surnames respectively), and while the former group's surnames are still located mostly in the Basque country, the latter's are present in parts of the south of Spain, explained by major re-population settlements in the south during the Middle Ages (Faure et al, 2001).

Quantitative Analysis of Forename & Surnames Frequencies

The data for this article were sourced from the Spanish 2004 telephone directory. The data represent 11.8 million telephone lines with up to two surname fields and one forename field; some entries had only one surname. Although the telephone directory contained 12.6 million residential telephone lines, 0.8 million opted out of the public version, which could introduce a small bias in this analysis. Assuming one person for every telephone line and given the Spanish population in 2004 to be 43.2 Million (Instituto Nacional de Estadistica, 2006), the data represent about one entry for every 3.6 people which are typical for telephone data, and represents a reasonable sample of over 27%.

The data were presented in either all upper-case, or in lower-case with upper case initial letters as in *MARIA DEL CARMEN* and *Maria Del Carmen*; all data were converted to the latter format.

The analysis follows the pattern established in Tucker (2001, 2002, 2007) for other national distributions of surnames and forenames. The top 100 forenames and surnames are listed. Graphs of percentage of population against percentage of names, and population against occupied frequency for both surnames and forenames are given.

Spanish Forenames

There were no gender indicators in the forename data and the assigned indicators for only the top 100 forenames have been added here. Table 1 gives the list of the top 100 forenames by count, with count, rank and gender (F/M). The top eight forenames are masculine and overall 64% of the names and 75.3% of the telephone subscribers in the top 100 forenames are masculine, which is common for such tables drawn from telephone data.

Table 1 Top 100 Spanish Forenames by Rank
Put Table 1 C:\vfp\spain\spanforconsol.doc hereabouts

Graph 1 shows the plot of percentage the population against percentage of forenames.

Graph 1 Percentage Population against Percentage of Forenames (logarithmic scale)

Put Graph 1 C:\vfp\spain\spanforconsol.xls Hereabouts

This graph is typical for national distributions of forenames.

Graph 2 shows the population against occupied frequency for forenames.

Graph 2 Population against Occupied Frequency for Forenames.

Put Graph 2 C:\vfp\spain\spanforeoccfreq.xls hereabouts

As there is no sex tag in the data supplied the graphs are for males and females combined. The forename occupied frequency graph exhibits the same shape and distortion of the First Quasi-Line (from x=6 to about x=160) as seen in the UK Electoral Roll graphs (Tucker, 2007). The overall form, however, is exactly as expected and shows a clear Zipfian distribution following a power law (Zipf, 1949).

Spanish Surnames

In this paper the two surnames will be analyzed as individual surnames as with any other country, but, additionally the surnames in the *first surname* group will be compared with those of the *second surname* group to see, what differences, there are, if any. For convenience we will call the first group surnames: *surname1* and the second group *surname2.*

Table 2 lists the top 100 surnames in the *first surname* group, including count and rank, by ascending rank order. The equivalent table for the *second surname* group is virtually a carbon copy of the first and not much would be added by reproducing it here. Rather, the difference in rank of the *second surname* group is included in Table 2. The first 18 entries are identical; *Romero* is ranked 19th in the first surname group; in the second surname group it is ranked 19th minus the difference which is 19-(-1) = 20th. Likewise *Gutierrez* is ranked in the second surname group 20-(+1)=19th.

Table 2 The Top 100 Surnames in the First Position

Put Table 2 Hereabouts
C:\vfp\spain\??????.doc

It is not surprising that the distributions for the first surname group and second surname group are very similar as both are drawn from the same population group. There are some differences between the two groups as 347,895 (2.9%) people in the telephone directory do not have two surnames, but just one, which is usually surname1. This occurs with foreign persons of non-Spanish speaking countries, due to privacy protection attitudes of certain subscribers, or just because of errors in the data collection process.

Table 3 List the top 100 Surnames, and count by Rank regardless whether the surname was used in the first or second surname position. The first 25 are identical to Table 1 and as expected there is very little difference between the two tables, and all 100 are Spanish surnames. *Garcia* is the most popular

surname by far (6.8% of the population bear it as either surname1 or surname2). Its etymological origin comes from the patronym *Garcia*, which has not been used as a forename since the 16th century, but it must have been a very popular forename in Spain during the Middle Ages (Faure et al, 2001)

Table 3 The Top 100 Surnames in Either Position

Put Table 3 Hereabouts
C:\vfp\spain\??????.doc

Graph 3 shows the plot of percentage of the population against the logarithm of percentage of surnames in either position.The graphs, not shown, for the first position surnames and that of the second position surnames are virtually identical and, closely mirror that of Graph 3. They run slightly below that of Graph 3: at 0.1% of the surnames in the first position the plot is about 50% of the population, whereas we will see that the combined plot is at 55%. Of course all graphs axiomatically reach the 100,100 point.

Graph 3 Percentage Population against Percentage of Surnames in Either Position

Put Graph 3 C:\vfp\spain\surconsol.xls Hereabouts

This graph is not quite typical for national distributions of surnames; particularly the bulge between 0.001% and 0.01% of surnames. The 1997 UK Electoral Roll (Tucker, 2003) has a more typical "S" curve also observed for the US, Canadian and UK surnames distributions. Table 4 shows a rough comparison between the UK plot and that of the Spanish Data.

Table 4 Partial Comparison of the Spanish & UK Surname Plots in Percentages

Surnames	Population	
	Spain	UK
0.001	10	5
0.01	32	20
0.1	55	47
1.0	80	80

The Spanish surname distribution is thus even more weighted to the popular surnames than that of the UK. Spanish society thus seeks comparatively more use of the popular surnames. This exception has been also found in a comparative studies by Scapoli et al (2007) who compared the surname frequency

9

distributions of 8 major European countries Austria, Belgium, France, Germany, Italy, Netherlands, Spain and Switzerland.

These authors found that the top 8 most popular 'European surnames' are all Spanish, and there are 39 Spanish surnames in the top 100 surnames of the countries studied, when its population only represents 13% of the study. This fact can be attributed to three major causes. The first one is the pattern of Christian re-population of Spain since the early Middle Ages, from relatively small Christian communities in the north to a territorial expansion thought the Iberian Peninsula and to America, thus expanding an originally small pool of local surnames. The second is the pressure of the Spanish Inquisition what forced Muslim and Jewish converts to adopt popular Castilian surnames (Faure et al, 2001), as well as the mentioned Castilinization of other surnames. The third is a phenomenon of 'surname drift' (propagation of the same local popular surnames) that has not been counteracted by enough internal migration, and therefore a symptom of surname inbreeding in many areas (that is, a high frequency of marriages between the same surnames, also known as isonymy) (Scapoli et al, 2007).

Graph 4 shows the Population by Occupied Frequency of the Surname1; again, the same style graph for Surname2 is virtually identical and is not shown.

Graph 4 Population against Occupied Frequency for Surname1

Put Graph 4 C:\vfp\spain\spansur1occfreq.xls hereabouts

Graph 5 shows the Population by Occupied Frequency of the combined surnames.

Graph 5 Population against Occupied Frequency for the combined Surnames

Put Graph 5 C:\vfp\spain\spansuroccfreq.xls hereabouts

Both Graphs 4 and 5 are typical for national surname distributions and are clearly Zipfian.

The Geographical Distribution of Surnames

The maps shown in Figure 1 show the geographical distribution of surnames in Spain according to their language region of origin; Basque, Catalan & Valencian, Galician, Castilian, and 'Other Spanish' which includes Spanish surnames with no particular regional or linguistic origin. The method to identify the language of origin for each Spanish surname is explained in Mateos (2007).The power of these five maps is that they summarize the history of the Middle Ages in Spain up to the 16th century, and how slowly

those population patterns have evolved since then. The striking fact that one needs to remember is that these maps do not come from a historical atlas, but they have been built from the surnames frequencies of the Spanish 2004 telephone directory.

Figure 1 The Distribution of Basque, Catalan & Valencian, Galician, Castilian, and 'Other Spanish' Surnames in Spain by Postal Area

Insert Figure 1 about here.

The maps show the five quintiles of the frequencies of each of the surname categories by postal area, from 1 (lower quintile) to 5 (highest quintile).

The history of the four languages of Spain, their origins in the north third of the country and their southwards expansion are explained by these maps. The maps show the southwards expansion of Castilian from its core in the north-northwest, of Catalan from the Northeast along the Mediterranean coast and islands (and south of Italy and Sardinia, should they also be on the map), the core of Galician language in the interior of Galicia (north-east) and onto Portugal (not seen on the map except for a few overspills along the border). They also show the uniqueness of Basque surnames and the relatively smaller interaction with their neighbors, a fact well studied in genetics (Cavalli-Sforza, 1997). Finally, a map of all other Spanish surnames reflects the inverse of the above maps, the re-population of the southern half of Spain through the Middle Ages, the generation of new surnames and the general spread of a much more mixed population that also went on to the Canary Islands (see map insets in Figure 1), and beyond on to Latin America. The maps also show the areas where more population mix and interaction have occurred, specially during the 20th century, eroding the 'bedrock' of local surnames established in Middle Ages. This is evident in the Catalonian coast and the province of Barcelona, leaving the highest incidence of Catalan surnames to the interior. A similar pattern is observed in the Balearic Islands, Valencia and Galicia. This is the power of unveiling a sort of 'demographic geomorphology', deposited during nearly 10 centuries, through the geography of contemporary surname frequencies

International Comparisons of Spanish Names

Graph 6 shows a comparison of the frequencies of the top 100 surnames in five Spanish-speaking countries; Spain, Argentina, Mexico, Venezuela, and the U.S. The sources for these countries are as follows: *Spain*, the telephone directory featured in this paper; *Mexico*, a list of the top 100 surnames from the 2006 Electoral Roll, supplied to the authors by the Mexican Electoral Commission (Instituto Federal Electoral) under a Freedom of Information Act request; *Argentina*, 100 most frequent surnames from the 2001 electoral roll (Cámara Nacional Electoral) published in Dipierri (2005); *Venezuela*, a list of the 40

most frequent surnames from the 1991 electoral roll (Consejo Supremo Electoral) published by Rodriguez-Larralde (2000); *U.S.*, a list of surname frequency data from the 1990 Census published on-line (US Census, 2006)

Graph 6 The Percentage Population Represented by the Top 100 Surnames in Spain, Mexico, Venezuela Argentina and the U.S.

Insert Graph 6 about here

There is a clear distinction between the slope of the frequency curves of Argentina and the U.S. distributions, much less steep than those of the other three countries. This is explained by a much higher rate of surname immigration from different countries and languages to both countries, whereas Spain, Mexico and Venezuela have not been substantially exposed to non-Spanish surnames. Secondly, the Mexican population is highly concentrated in a few surnames, 50% of them sharing just 74 surnames, and the top three surnames covering 9.5% (Hernandez, Garcia and Martinez). This must be due to the combined fact that a few surnames have been imposed to the pre-hispanic population by the Spanish colonial army and the church (Hernández means the son of Hernán, the forename of the conqueror of Mexico; Hernán Cortes), and that a low internal migration rate and low intermarriage between ethnic and socio-economic groups have produced high surname drift (propagation of the same popular surnames).

The top U.S. Hispanic surnames are not included in Graph 6 but if they were the line would run very close to the Mexican line. These data were derived from the list of Hispanic surnames and the overall surname frequency data from the 1990 Census published on-line (US Census, 2006). The close similarity between the Mexican and the U.S. Hispanic surname distributions point to a high proportion of the Hispanic surnames in the US having come from Mexico rather than the rest of Latin America, plus a reflection of the population settlements in the south western states prior to the 1848 US-Mexican border.

Summary

The characteristics of the forenames and surnames in Spain discussed in this paper can be summarized into a set of common features. The forename frequency distribution of the contemporary population of Spain follows a similar pattern as that of other countries studied by Tucker. The literature seems to indicate this was not the case in the Middle Ages, with most of the population shared just a small pool of religiously prescribed forenames, so there has been a phenomena of rapid expansion in the fore-naming practices. The surname frequency distribution presents a unique pattern, with a much higher concentration of the population in the most popular surnames than that found in other countries, an anomaly that other authors have also found (Scapoli et al, 2007). This peculiar surname frequency distribution could be explained by three combined processes. Firstly, the pattern of Christian re-population

of the Iberian Peninsula and population expansion since the early Middle Ages from a small pool of patronyms in the north. Secondly, due to an imposed process of name change, on the one side because of the pressure of the Spanish Inquisition against non-Christians, and on the other due to a process of Castilinization of surnames. Thirdly, a phenomenon of 'surname inbreeding', that is, a high frequency of marriages between the same surnames (also known as isonymy) not counteracted by migration in many areas. Amongst other features of Spanish names, the unique naming system of two surnames does not produce any substantial difference between the frequency distribution of parental and maternal surnames (both of them coming from males two generations up the genealogical chain), even when it better reflects both sexes in the population.

The exploratory geographical analysis of name groups classified by language of origin that has been presented here, does indicate the clear potential of using the quantitative analysis of names' geographical distributions to unveil historic population settlement and migration processes. In this case it reveals the original language regions of Spain in the Middle Ages and how these cultural regions are still structuring how populations mix today within still very confined interaction areas. Finally, a comparison of the frequency distribution the top surnames in five Spanish-speaking countries shows that surnames are a good indicator of how different populations have settled and have mixed between countries, as well as how the naming practices imposed upon former colonies have impoverished the naming heritage of their populations.

In this paper we hope to have unveiled some of the 'quantitative secrets' of Spanish names. Through the set of techniques presented here, we also aimed to introduce to the onomastic community a new field of spatio-temporal quantitative analysis of names to understand past and current population structures through 'name geomorphology'. We believe that linguists, historians, geographers, geneticists, statisticians and demographers should collaborate more closely to unveil a little bit more of how we came to be what we are today.

Appendix – Data Processing Issues

Data Hygiene - Errors in the Data

Typical for large data files but indicative of the poor data hygiene in the the gathering process; there is no reason why any of the following errors listed would not have been trapped by the simple expedient of only allowing symbols that appear in forenames & surnames.

Spaces before surnames & forenames
Non letter symbols other than hyphens eg: ".", "0" for "O" , "4", "%", " "(space), "AAA" (AAANAN), " _ " , "º " , " `".

Format Issues

The data sets were complicated by the fact that although *Garcia* looked liked *Garcia* they had been keyed in different forms such that *GARCIA*, 100 when added to *Garcia*, 99 did not always sum to *Garcia*, 199. The problem was that some of the data were in all capitals, and some were in initial capitals only, AND that converting the 'all capitals' format to 'initial capitals only' did not resolve the issue. This data problem was resolved, and the consolidated tables give the totals. However when comparing the first surnames with the second surnames the 'all capitals' data only has been used. In the case of Garcia for example this represents 91% of the total data so it is believed that the results are valid

References

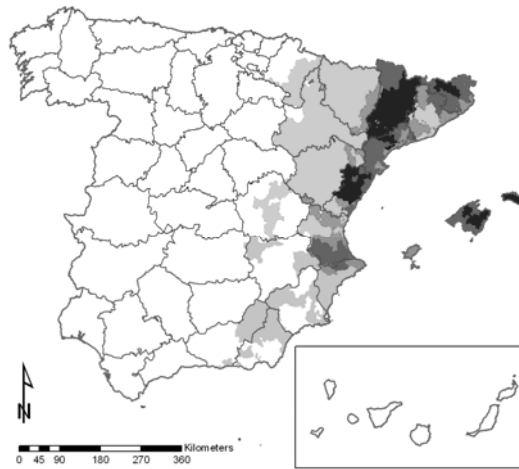Cavalli-Sforza LL. 1997. Genes, Peoples, and Languages. *Proceedings of the National Academy of Sciences* **94**(15): 7719-7724

Dipierri JE, Alfaro EL, Scapoli C, Mamolini E, Rodriguez-Larralde A, Barrai I. 2005. Surnames in Argentina: A Population Study through Isonymy. *American Journal of Physical Anthropology* **128**: 199-209

Faure R, Ribes MA, Garcia A. 2001. *Diccionario De Apellidos Españoles* Madrid: Espasa Calpe.

Instituto de Estadística de la Comunidad de Madrid. 2006. Guía De Nombres Y Primer Apellido De Los Residentes En La Comunidad De Madrid 1998-2005

Instituto Nacional de Estadistica. 2006. *Cifras Oficiales De Población [Official Population Figures].* Available at: http://www.ine.es/prensa/padron_tabla.htm. Accessed: 15/02/2007.

Kremer D. 2003. Spanish and Portuguese Family Names. In *Dictionary of American Family Names*, Hanks P (eds.), Oxford University Press: New York.

Mateos P. 2007. Segregación Residencial De Minorías Étnicas Y El Análisis Geográfico Del Origen De Nombres Y Apellidos [Residential Segregation of Ethnic Minorities and Geographic Analysis of Name Origins]. *Cuadernos Geograficos* **40**: (in press)

Moll FB. 1982. *Els Llinatges Catalans*. Mallorca: Moll.

Rodriguez-Larralde A, Gonzales-Martin A, Scapoli C, Barrai I. 2003. The Names of Spain: A Study of the Isonymy Structure of Spain. *American Journal of Physical Anthropology* **121**: 280-292

Rodriguez-Larralde A, Morales J, Barrai I. 2000. Surname Frequency and the Isonymy Structure of Venezuela. *American Journal Of Human Biology* **12**: 352-362

Scapoli C, Mamolini E, Carrieri A, Rodriguez-Larralde A, Barrai I. 2007. Surnames in Western Europe: A Comparison of the Subcontinental Populations through Isonymy. *Theoretical Population Biology* **71** 37-48

Tibón G. 2001. *Diccionario Etimológico Comparado De Los Apellidos Españoles, Hispanoamericanos Y Filipinos*. México D.F.: Fondo de Cultura Económica.

Tucker DK. 2001. Distribution of Forenames, Surnames, and Forename-Surname Pairs in the United States. *Names* **49**: 69-96

Tucker DK. 2002. Distribution of Forenames, Surnames, and Forename-Surname Pairs in Canada. *Names* **50**(2): 105-132

Tucker DK. 2003. An Analysis of the Forenames and Surnames of England and Wales Listed in the Uk Census Data. *Onoma* **38**: 181-216

Tucker DK. 2007. Surname Distribution Prints from the Uk 1998 Electoral Roll Compared with Those from Other Distributions. *Nomina* **30** In Press

US Census. 2006. *Us Census Bureau Geneaology Resources.* Available at: http://www.census.gov/genealogy/www/. Accessed: 12/05/2006.

Zipf GK. 1949. *Human Behavior and the Principle of Least Effort* Reading, MA: Addison-Wesley.

Basque

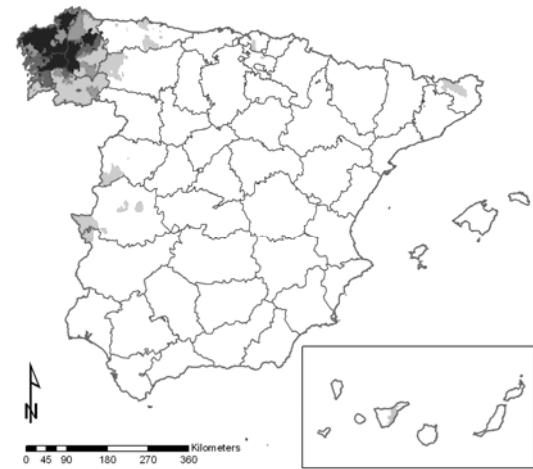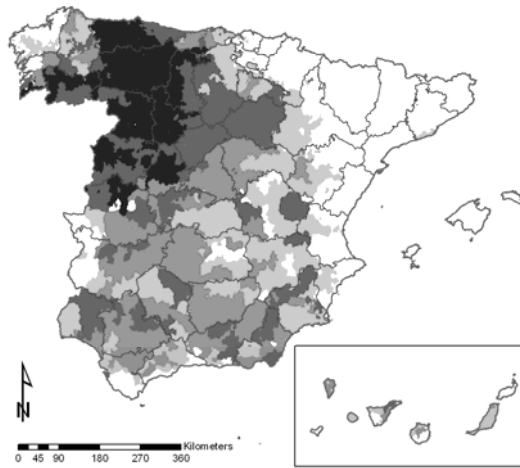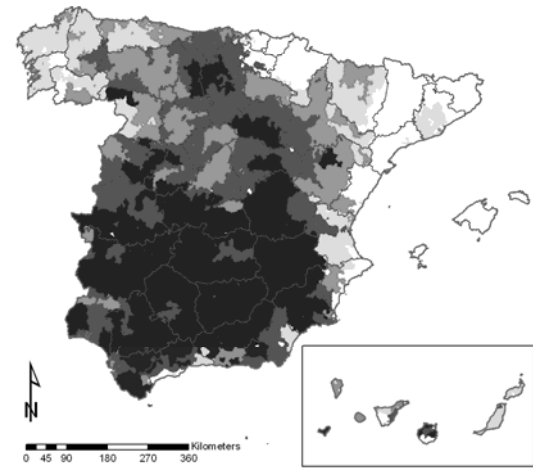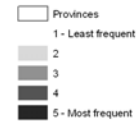## Catalan & Valencian
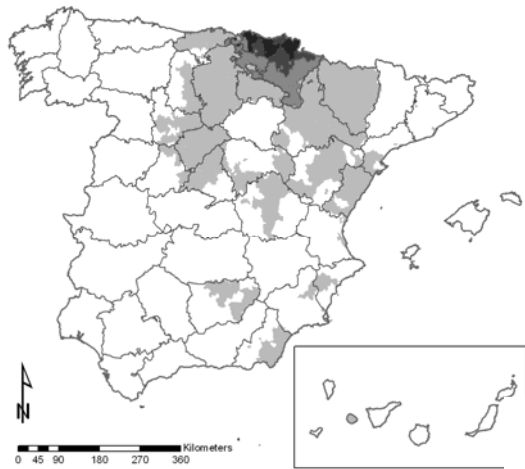
Galician

Context Map with area names here

Castilian

Other Spanish

Provinces
1 - Least frequent
2
3
4
5 - Most frequent
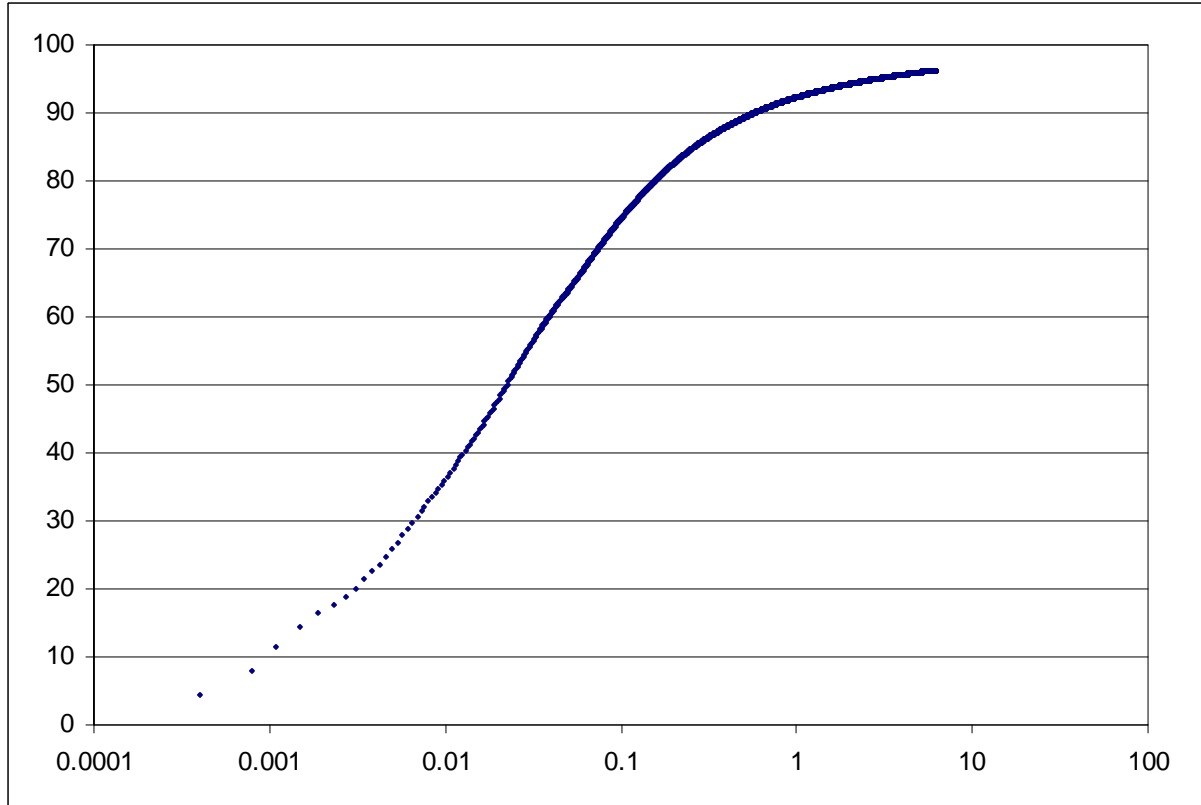
Graph 1 up – graph2 down
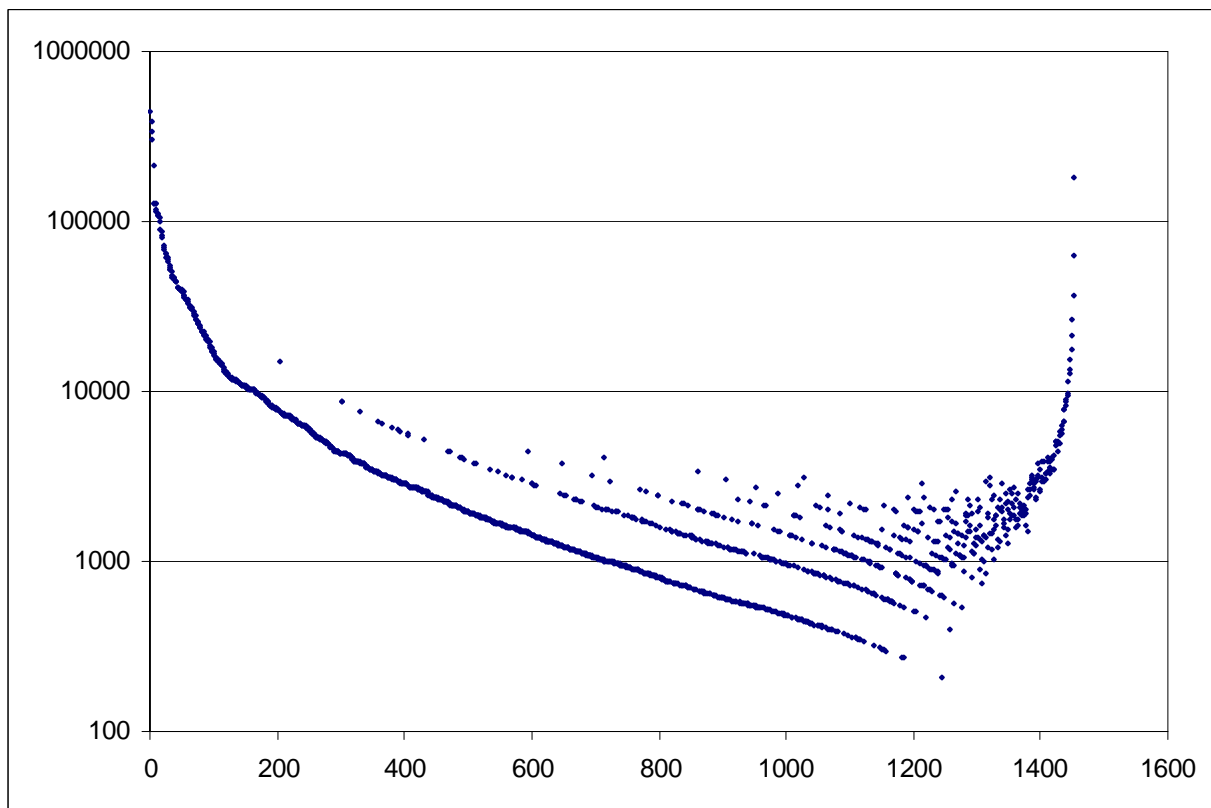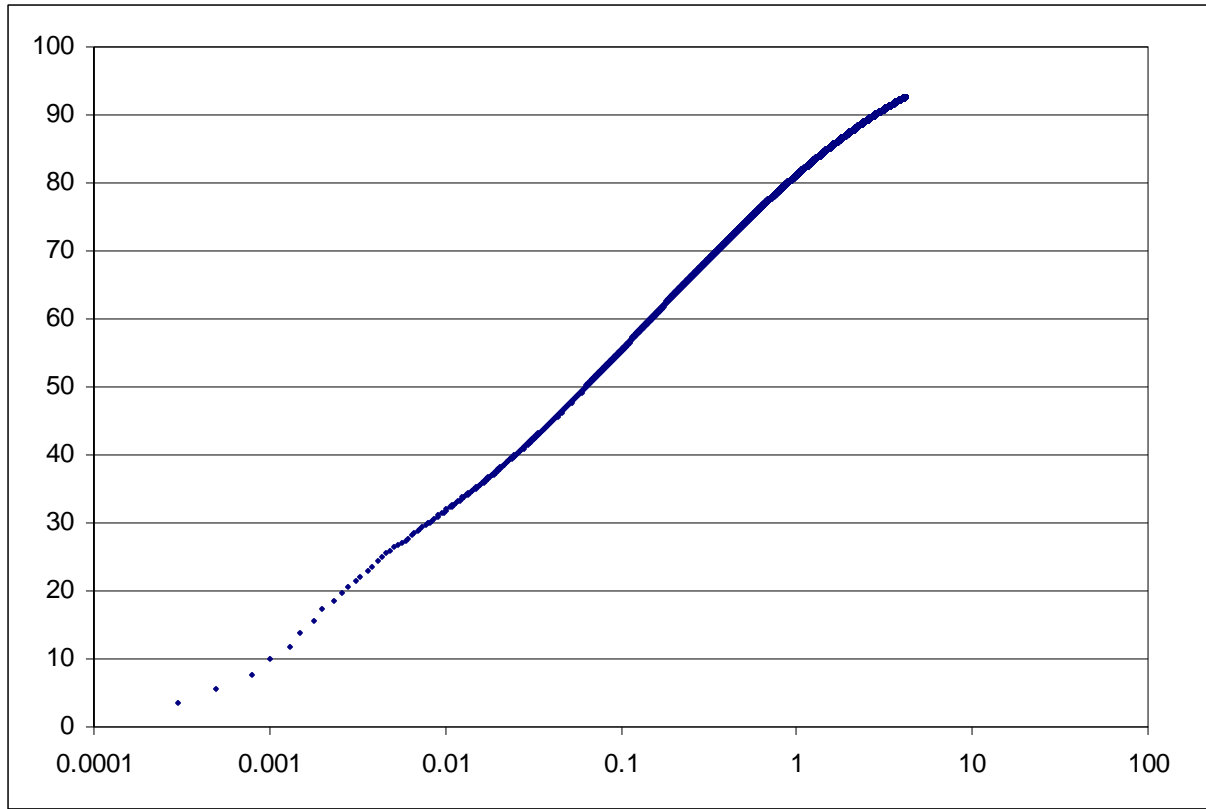
G3 up g4 down
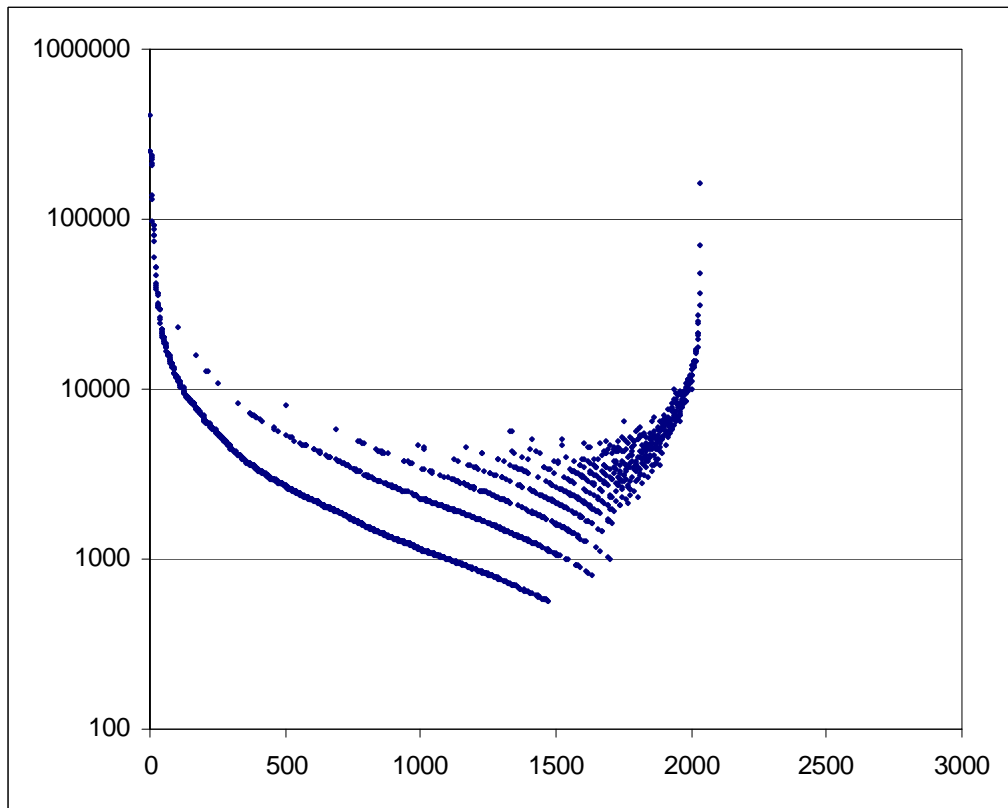
down

g5
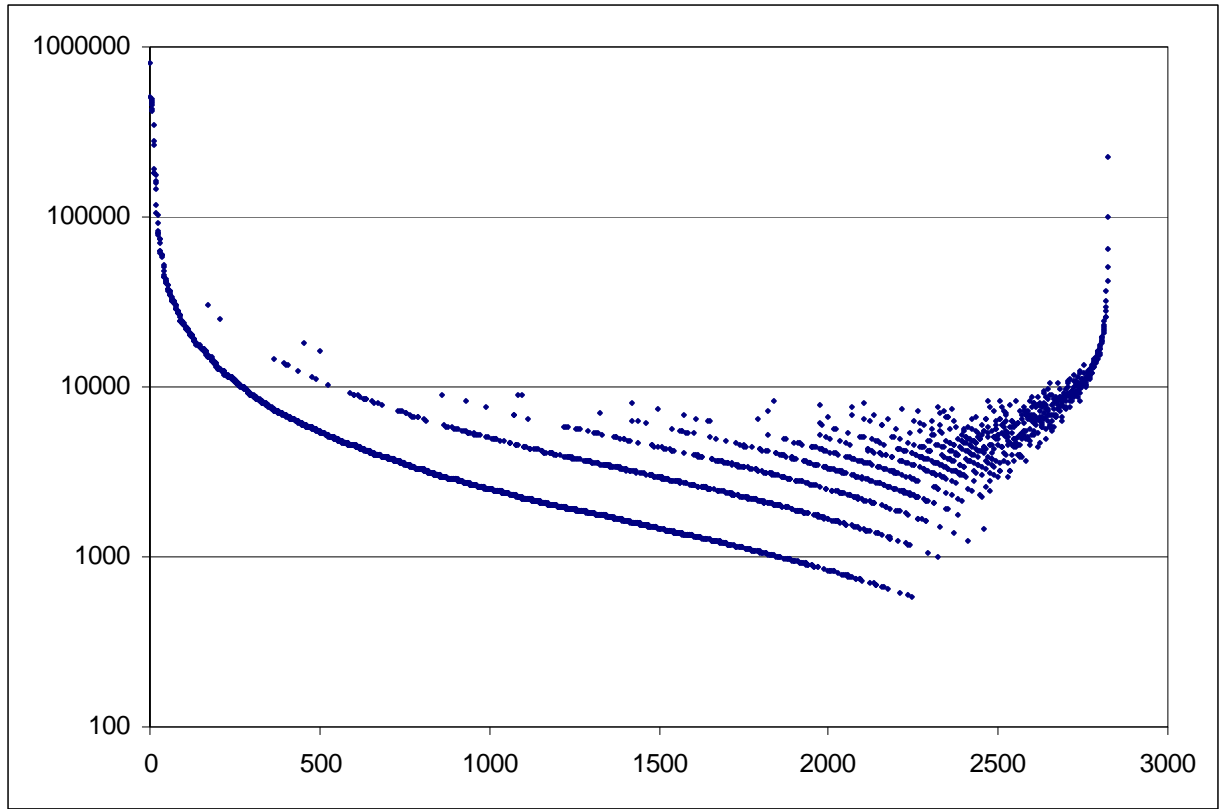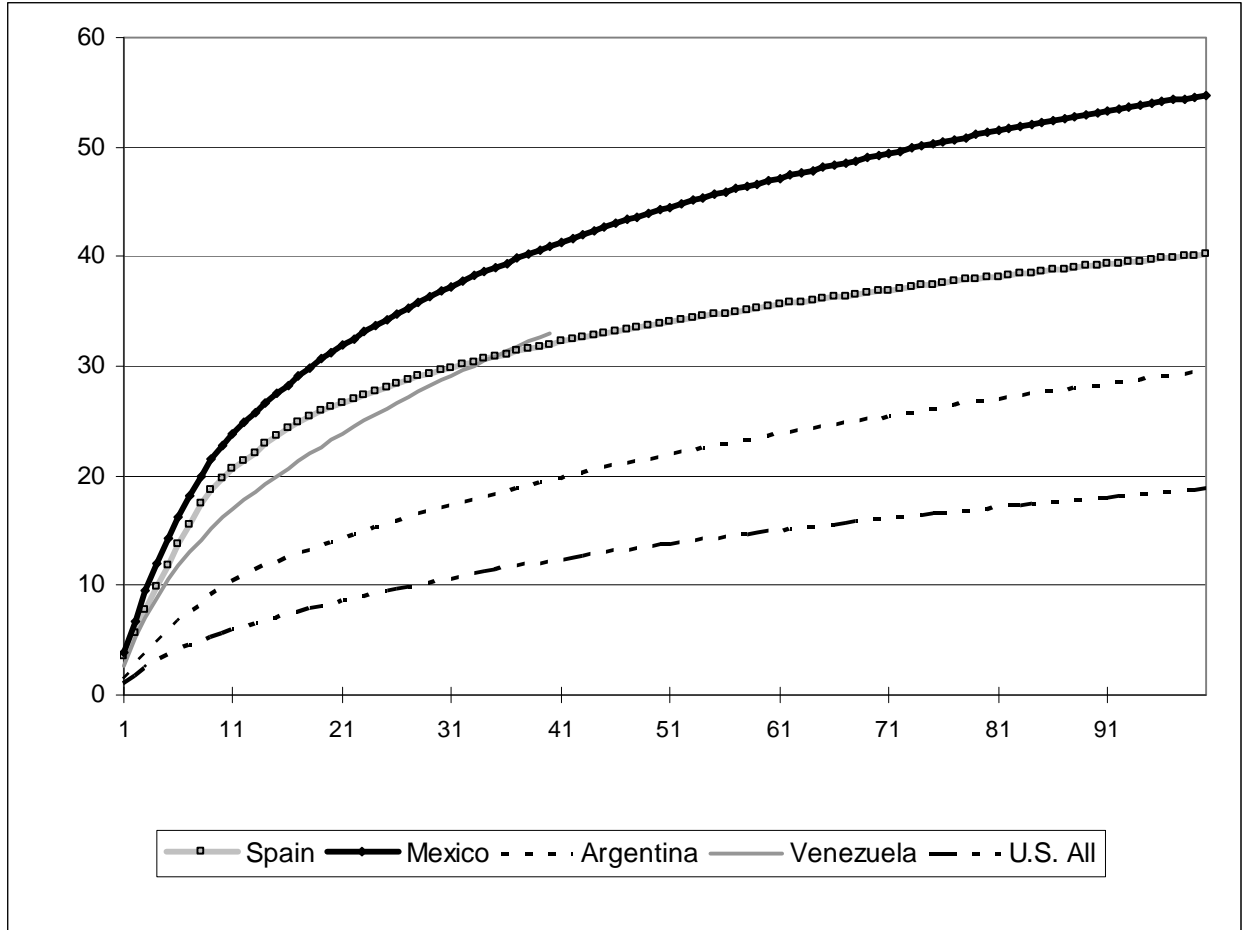
g6up

Name	Count	Rank	F-M
Jose	444066	1	M
Antonio	388249	2	M
Manuel	338917	3	M
Francisco	306218	4	M
Juan	214048	5	M
Pedro	129133	6	M
Jose Luis	128727	7	M
Jesus	127249	8	M
Maria	126576	9	F
Carmen	118138	10	F
Angel	114272	11	M
Luis	112671	12	M
Miguel	109741	13	M
Rafael	108028	14	M
Jose Antonio	105790	15	M
Jose Maria	100548	16	M
Maria Del Carmen	90275	17	F
Fernando	86526	18	M
Vicente	82888	19	M
Josefa	79591	20	F
Jose Manuel	72161	21	M
Ramon	70488	22	M
Carlos	69212	23	M
Isabel	65053	24	F
Francisco Javier	64296	25	M
Joaquin	62078	26	M
Enrique	60917	27	M
Dolores	59775	28	F
Francisca	57883	29	F
Antonia	55705	30	F
Miguel Angel	53394	31	M
Juan Jose	52305	32	M
Pilar	51501	33	F
Andres	50179	34	M
Maria Dolores	47544	35	F
Santiago	46720	36	M
Maria Teresa	46412	37	F
Emilio	46399	38	M
Javier	45559	39	M
Julian	44809	40	M
Concepcion	44753	41	F
Juan Antonio	44339	42	M

Name	Count	Rank	F-M
Julio	41148	43	M
Ana Maria	40855	44	F
Ana	40782	45	F
Felix	40119	46	M
Alfonso	40057	47	M
Juan Carlos	39600	48	M
Salvador	39387	49	M
Maria Luisa	39121	50	F
Mercedes	39009	51	F
Tomas	38489	52	M
Eduardo	36905	53	M
Agustin	36895	54	M
Manuela	35730	55	F
Mariano	34820	56	M
Juana	34698	57	F
Rosario	34685	58	F
Teresa	34443	59	F
Ricardo	34080	60	M
Pablo	32545	61	M
Alberto	32308	62	M
Juan Manuel	31417	63	M
Domingo	31314	64	M
Jaime	31090	65	M
Maria Jose	30201	66	F
Rosa	30119	67	F
Maria Isabel	29595	68	F
Encarnacion	29090	69	F
Ignacio	28080	70	M
Diego	27950	71	M
Maria Jesus	26556	72	F
Gregorio	26525	73	M
Alejandro	26158	74	M
Felipe	25154	75	M
Daniel	24910	76	M
David	24799	77	M
Maria Angeles	24269	78	F
Margarita	24039	79	F
Jose Ramon	23840	80	M
Jorge	22717	81	M
Angeles	22427	82	F
Maria Pilar	22275	83	F
Alfredo	22261	84	M

Name	Count	Rank	F-M
Julia	21908	85	F
Maria Del Pilar	21270	86	F
Rosa Maria	21230	87	F
Sebastian	20690	88	M
Amparo	20478	89	F
Eugenio	20342	90	M
Gabriel	20104	91	M
Lorenzo	19731	92	M
Roberto	19575	93	M
Maria Carmen	18530	94	F

Elena	18013	95	F
Consuelo	17992	96	F
Jose Miguel	17586	97	M
Guillermo	17157	98	M
Victor	17064	99	M
Francisco Jose	17014	100	M

Table 1 up

Surname	Count	Rank	Difference
Garcia	404150	1	0.0
Fernandez	249983	2	0.0
Gonzalez	248769	3	0.0
Rodriguez	241057	4	0.0
Lopez	233814	5	0.0
Martinez	224887	6	0.0
Sanchez	216267	7	0.0
Perez	209572	8	0.0
Martin	139762	9	0.0
Gomez	130565	10	0.0
Ruiz	96419	11	0.0
Hernandez	91153	12	0.0
Jimenez	91148	13	0.0
Diaz	88011	14	0.0
Alvarez	80681	15	0.0
Moreno	79530	16	0.0
Muḱoz	73569	17	0.0
Alonso	59837	18	0.0
Romero	52728	19	-1.0
Gutierrez	51601	20	1.0
Navarro	46638	21	-1.0
Torres	41576	22	-3.0
Dominguez	40779	23	0.0
Gil	40038	24	-2.0
Vazquez	38892	25	1.0
Ramos	37091	26	-1.0
Serrano	36863	27	-1.0
Blanco	35318	28	-1.0
Sanz	31671	29	-4.0
Suarez	31040	30	-1.0
Ortega	31038	31	-3.0
Castro	30290	32	-7.0
Molina	30201	33	-8.0
Ramirez	29846	34	-3.0
Rubio	29748	35	-1.0
Morales	29581	36	-2.0
Delgado	29502	37	2.0
Ortiz	26249	38	-7.0
Marin	25496	39	-9.0
Iglesias	24271	40	-7.0
Santos	22824	41	-9.0
Garrido	22436	42	-7.0
Castillo	22243	43	-15.0
Nuḱez	22134	44	-7.0
Calvo	21295	45	-7.0
Prieto	21069	46	-8.0
Lozano	20925	47	-8.0
Cruz	20415	48	-20.0
Medina	20309	49	-7.0
Vidal	20285	50	-11.0
Diez	20095	51	-2.0

Cano	19843	52	-5.0
Gallego	19347	53	-6.0
Pascual	18916	54	-12.0
Peña	18859	55	-7.0
Guerrero	18715	56	-4.0
Vega	18067	57	-8.0
Herrero	17875	58	-6.0
Mendez	17842	59	-4.0
Leon	17742	60	-7.0
Ferrer	17595	61	-22.0
Nieto	16628	62	-7.0
Fuentes	16475	63	-7.0
Marquez	16281	64	-15.0
Cortes	16172	65	-12.0
Ibañez	16080	66	-15.0
Campos	16032	67	-9.0
Vicente	15941	68	-10.0
Carrasco	15885	69	-11.0
Herrera	15849	70	-1.0
Caballero	15787	71	-4.0
Cabrera	15575	72	-2.0
Montero	15213	73	-9.0
Lorenzo	15053	74	1.0
Esteban	14486	75	-10.0
Aguilar	14481	76	-17.0
Gimenez	14369	77	-23.0
Crespo	14180	78	-10.0
Soler	14179	79	-43.0
Hidalgo	14153	80	-7.0
Pastor	14082	81	-9.0
Duran	13847	82	-14.0
Flores	13633	83	-11.0
Saez	13498	84	-5.0
Mora	13430	85	-13.0
Arias	13216	86	-5.0
Velasco	13029	87	-5.0
Santana	12585	88	4.0
Andres	12498	89	-16.0
Marti	12325	90	-55.0
Reyes	12294	91	-13.0
Merino	12289	92	-7.0
Moya	12202	93	-16.0
Izquierdo	12138	94	-8.0
Carmona	11998	95	-16.0
Bravo	11986	96	-1.0
Casado	11900	97	-18.0
Pardo	11703	98	-10.0
Soto	11700	99	-15.0
Miguel	11614	100	-32.0

Table2 up

Table 1 Top 100 Surnames by Rank

Surname	Count	Rank
Garcia	813257	1
Fernandez	503142	2
Gonzalez	499596	3
Rodriguez	482448	4
Lopez	467681	5
Martinez	449954	6
Sanchez	433030	7
Perez	421997	8
Martin	278261	9
Gomez	261776	10
Ruiz	193130	11
Hernandez	182808	12
Jimenez	181206	13
Diaz	176485	14
Alvarez	161674	15
Moreno	158435	16
Muñoz	145791	17
Alonso	119004	18
Romero	105603	19
Gutierrez	103776	20
Navarro	92302	21
Torres	82578	22
Dominguez	81473	23
Gil	79895	24
Vazquez	77755	25
Serrano	73552	26
Ramos	73544	27
Blanco	69810	28
Suarez	63257	29
Sanz	62803	30
Ortega	61904	31
Molina	60172	32
Rubio	59458	33
Ramirez	59370	34
Delgado	59145	35
Morales	59034	36
Castro	58945	37
Ortiz	52792	38
Marin	50412	39
Iglesias	48321	40
Garrido	45025	41
Santos	44190	42
Nuñez	43931	43
Calvo	42705	44
Lozano	42088	45
Castillo	41944	46
Prieto	41709	47
Diez	40607	48
Medina	40575	49
Vidal	40211	50
Cano	39535	51

Gallego	38130	52
Guerrero	37290	53
Pascual	36976	54
Cruz	36776	55
Peña	36538	56
Mendez	35490	57
Herrero	35020	58
Vega	34825	59
Ferrer	34815	60
Leon	34138	61
Nieto	32752	62
Fuentes	32670	63
Cortes	32434	64
Marquez	31973	65
Campos	31882	66
Caballero	31863	67
Ibañez	31793	68
Herrera	31534	69
Carrasco	31284	70
Vicente	31163	71
Cabrera	30876	72
Lorenzo	30112	73
Montero	30085	74
Gimenez	29429	75
Esteban	28648	76
Hidalgo	28460	77
Aguilar	28386	78
Pastor	27921	79
Soler	27776	80
Crespo	27638	81
Duran	27575	82
Flores	27414	83
Saez	27288	84
Arias	26803	85
Mora	26267	86
Velasco	26172	87
Santana	25689	88
Merino	24401	89
Izquierdo	24329	90
Moya	24287	91
Bravo	24234	92
Reyes	24067	93
Carmona	24050	94
Marti	23914	95
Andres	23895	96
Redondo	23542	97
Pardo	23209	98
Vila	23162	99
Casado	23062	100

Table 3 up