



L2's Ethnic Coding

Ethnic analysis is a complex art relying on our in-house databases of both surnames and given names. Our surname table currently contains nearly 300,000 unique surnames while our given name table numbers approximately 10,000. Each of these tables continues to be supplemented and refined over time.

Our analysis begins with a close look at the given name. A great many given names have strong ethnic affiliations and can provide clues to ethnicity even when, as in the case of marriage or adoption, surnames might be different. Following an analysis of the given name we check to see if the surname finds a match in our tables. Finally, the combination of the given and surnames is analyzed to check for the 'overlap groups', that is, those ethnic groups that tend to share common surnames. After taking these three analyses into consideration a final 'best guess' as to ethnicity is determined and assigned.

African-American Coding

The assigning of African-American coding is more complex and problematic. L2 uses a probability code to allow the user of the data to make his or her own judgments. In some states, African-American ethnicity is self-reported and in those cases we are able to assign a probability of 99%. These self-reported codes are unavailable in most areas.

A relatively small, but still significant, number of voters is of recent African descent and these voters will have true African surnames. To them we assign a 91% probability. Another, larger, group will find a match to our high probability African-American given name tables and will be assigned a 92% ranking.

For the remaining African-American voters our technique is as follows. We begin by identifying all voters who live in census blocks (a small geographical area) where 50% or more of the residents identified themselves as African-American in the Year 2000 Census. Within the census blocks that are 50% to 59% self-reported African-American we then analyze the surnames. If the surnames fall into one of eight high probability groups (Arab, Dutch, English/Welsh, French, Irish, Afghan, Pakistani and Scots) we assign a probability code of '5' followed by the surname type (e.g. 'E' for English/Welsh). Thus a '5E' voter would be a voter with an English surname living in a census block where 50% to 59% of the residents are African-American. We repeat the same process for the 60% to 69% group and each of the following deciles up to 90% to 99%. This coding technique eliminates those voters with low probability African-American surnames (e.g. East Asian or Slavic) even if they live in high

density African-American areas and allows the user of the data to decide at what level of confidence he or she wishes to make a selection. A higher level of confidence will generally correlate with a lower total number of voters selected.

Jewish-American Coding

Jewish coding is kept separate from ethnic coding so that different subsets of Jewish-American voters can be identified. Thus, a Jewish-Russian named voter might be seen as having characteristics different from a Jewish-German named voter or a Jewish-English voter. Ethnic coding is as much art as it is science and there will always be disagreements as to the ethnic derivation of particular names. Used with care, however, ethnic selections from the voter file can be a powerful campaign tool.

For more information contact:

Paul Westcott
Director, Marketing & Communications
L2 Political
646-457-1439
paul.westcott@l2political.com

