

Canadian Century Research Infrastructure Project

Nationality Background

Root variable: Nationality

Synopsis of code set construction:

Initially, the intention was to use the Country of Birth codes for the Nationality variable. However, there were two differences between the Country of Birth and Nationality questions. For one, the Country of Birth responses were given as geographic descriptors, while the Nationality question asked for a personal descriptor (Canadian instead of Canada, for example). Secondly, the Country of Birth codes included a number of codes for specific geographic places that were not used as personal descriptions of one's nationality. While the Country of Birth question could be answered with a reference to a specific territory (an island, town, city, or region for example), the Nationality question referred to one's belonging to a political entity.

Given these differences, a new code set was constructed based upon the Country of Birth codes (and consequently upon IPUMS Birthplace codes), in consultation with the micro-data.

Understanding the numeric code:

The Nationality code set is structured according to world geography, beginning with the United States of America. The first digit refers to each continent or region.

Continent (or sub-continent) of the Nationality Codes

| | |
|---------------------------------------------|--------------------------------|
| American (and citizens from US possessions) | 10000 to less than 12000000 |
| Canadian | 15000000 to less than 15200000 |
| Other North American | 15500000 to less than 20000000 |
| Central American | 20000000 to less than 30000000 |
| South American | 30000000 to less than 40000000 |
| European | 40000000 to less than 50000000 |
| Asian | 50000000 to less than 60000000 |
| African | 60000000 to less than 70000000 |
| Oceanian | 70000000 to less than 80000000 |
| Uncodeable | 80000000 and over |

Code set application notes:

- Dual responses: Unlike the Racial or Tribal Origin codes, the Nationality codes do not account for responses that include more than one nationality, such as "Canadian American", "Scots Irish", or "Russian Austrian." This is because the team decided to treat Nationality as a geo-political reference, while the Racial or Tribal Origin responses are seen as expressions of personal identity. Multiple responses in the Nationality data did not receive distinct codes.

Canadian Century Research Infrastructure Project

Nationality Background

When mapping these values, the first response was used. “Canadian American” was coded to “Canadian,” Scots Irish” to “Scottish”, and “Russian Austrian” to “Russian,” for example. The only exception to this rule is when a religious/ethnic qualifier was the first response. In these cases, the second value was used. For example, “Jewish, Russian” became “Russian.”

- Illegible responses: Data entry operators were instructed to enter a question mark for each illegible character in a response. These values were coded to “Illegible.”
- Certain regional distinctions were accounted for in this variable. For example “Bohemian” was mapped to a “Bohemian” code. However, if a country was given with the value, it would be mapped to that country. As such, “Galician Poland” was mapped to “Polish.”
- Those that responded with “Half Breed_____” were mapped to “Métis.”
- “Flemish” was mapped to “Belgian.”
- Hebrew was mapped to “Uncodeable,” because it was impossible to determine whether this was meant as a political statement, or as an indicator of a Palestinian or Israeli nationality.
- Aboriginal responses received their own codes if they were one of the language groups, such as Cree or Mohawk. Those who gave bands as responses were mapped to “First Nations n.e.c.”